THE UNIVERSITY OF CHICAGO

ENGINEERING AND EVOLVING DIRHODIUM ARTIFICIAL METALLOENZYMES FOR CARBENE TRANSFER REACTIONS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

DAVID M. UPP

CHICAGO, ILLINOIS

DECEMBER 2021

To my family, whose constant love and support made this possible.

And to Lucía, te amo.

In memory of Alan Swartz

1987-2018

Table of Contents

List of Tables vii
List of Figures viii
List of Schemes ix
Acknowledgements x
Abstract xiii
Chapter 1: An Introduction to Artificial Metalloenzyme Catalysis 1
1.1 Approaches to Selective Catalysis 1
1.1.1 Catalysis and Selectivity 1
1.1.2 Selective Transition Metal Catalysts 2
1.1.3 Selective Catalysis by Metalloenzymes 4
1.1.4 Expanding the Scope of Metalloenzyme Catalysis 6
1.2 Artificial Metalloenzymes 8
1.2.1 Introduction to Artificial Metalloenzymes 8
1.2.2 Nitrobindin ArMs 11
1.2.3 Streptavidin ArMs 17
1.2.4 Myoglobin ArMs 23
1.3 Prolyl Oligopeptidase Scaffold for Artificial Metalloenzymes 29
1.3.1 POP Scaffold and ArM Formation 29
1.3.3 Dirhodium-POP ArMs 35
1.4 Conclusion 38
1.5 References 40
Chapter 2: Exploration of the Substrate Scope of Dirhodium Artificial Metalloenzymes 47
2.1 Introduction 47
2.1.1 Random mutagenesis for improved POP-Rh_2 ArM catalysis 47
2.1.2 Dirhodium carbenoid insertion mechanisms 50
2.1.3 ArM-catalyzed carbenoid insertion reactions 51
2.2 Results and Discussion 52
2.2.1 Initial Exploration of N-H Functionalization 52
2.2.2 Initial Exploration of S-H Functionalization 54
2.2.3 Initial Exploration of Si-H Functionalization 59
2.2.4 Initial Exploration of C-H Functionalization 61
2.2.5 High-Throughput Methodology Optimization for C-H Functionalization Screening . 65
2.3 Conclusion 68
2.4 Experimental 68
2.5 Acknowledgements 79
2.6 References 79

Chapter 3: Engineering Dirhodium Artificial Metalloenzymes for Diazo Coupling Cascade Reactions 81
3.1 Introduction 81
3.1.1 Chemo- and Site-Selectivity in Catalysis 81
3.1.2 Artificial Metalloenzymes in Biosynthetic Cascade Reactions 82
3.1.3 Dirhodium ArM-Catalyzed Diazo Coupling 83
3.2 Results and Discussion 85
3.2.1 Initial Investigation and Optimization of ArM-Catalyzed Diazo Coupling 85
3.2.2 Directed Evolution of ArM Scaffold for Diazo Coupling 88
3.2.3 Diazo Coupling Substrate Scope 90
3.2.4 Cascade Reaction and Chemoselectivity of Dirhodium ArMs 91
3.2.5 Molecular Dynamics of Dirhodium ArMs 97
3.3 Conclusion 99
3.4 Experimental 101
3.4.1 General materials and methods 101
3.4.2 Synthetic Procedures 102
3.4.3 Overexpression and purification of proteins 117
3.4.2 ArM Bioconjugation 118
3.4.5 ArM Biocatalysis Method 121
3.4.6 Library Cloning and Screening Method 123
3.4.7 ER and Glucose Modfication 130
3.4.8 Computational Procedures 134
3.4.10 Experimental References 137
3.5 Acknowledgements 138
3.6 References 139
Chapter 4: Engineering Dirhodium ArMs for N-H Functionalization Reactions 142
4.1 Introduction 142
4.2 Results and Discussion 146
4.2.1 N-H Functionalization of Anilines 146
4.2.1.1 High-Throughput Method Optimization and Parent Selection 146
4.2.1.2 Directed Evolution 148
4.2.1.2 Reaction Optimization and pH Dependence of Enantioselectivity 152
4.2.1.3 Structural Modeling and Mechanism 155
4.2.1.4 Aniline Substrate Scope 161
4.2.2 N-H Functionalization of Alkyl Amines 162
4.2.2.1 Initial Reaction Exploration and Optimization 162
4.2.2.2 Amine Substrate Scope 165
4.3 Conclusion 166
4.4 Experimental 167
4.5 Acknowledgements 172
4.6 References 172

Chapter 5: Engineering Dirhodium ArMs for Concerted Carbene Insertion Reactions 175
5.1 Introduction 175
5.2 Results and Discussion 178
5.2.1 Engineering ArMs for Si-H Functionalization 178
5.2.1.1 High Throughput Screening Validation 178
5.2.1.2 Parent Selection and Reaction Optimization 181
5.2.1.3 Directed Evolution of Si-H Chemoselectivity 185
5.2.1.4 Substrate Scope 186
5.2.1.5 Alternate Approaches to Improving Si-H Functionalization Yield 187
5.2.2 Alternate Approaches Towards C-H Functionalization 191
5.3 Conclusion 197
5.4 Experimental 198
5.5 Acknowledgements 203
5.6 References 204
Appendix – Supplemental Spectra 207

List of Tables

Table 2.1	Evaluation of the cyclopropanation lineage for the N-H functionalization of THQ....	53
Table 2.2	S-H functionalization catalyzed by the ArM variant POP-ZA$_4$-H328.....	56
Table 2.3	Diazo substrate scope of the S-H functionalization of thiobenzene.....	57
Table 2.4	Evaluation of the cyclopropanation lineage on S-H functionalization.....	58
Table 2.5	Silane scope in the Si-H functionalization reaction catalyzed by 3-VRVH.....	60
Table 2.6	Si-H functionalization catalyzed by the cyclopropanation of lineage.....	61
Table 2.7	C-H functionalization of phthalan catalyzed by POP-ArMs.....	63
Table 2.8	C-H functionalization of activated substrates with a trifluoroethyl ester diazo.....	64
Table 3.1	Scaffold selection and reaction optimization.....	86
Table 3.2	Optimization of ArM concentration for diazo coupling.....	87
Table 3.3	Directed evolution of ArM for diazo coupling.....	89
Table 3.4	Substrate scope of ArM-catalyzed diazo coupling.....	91
Table 3.5	Ene-reductase substrate selection.....	92
Table 3.6	Substrate scope of ArM/ER cascade reactions.....	94
Table 3.7	Modification of additives by dirhodium catalysts in aqueous buffer.....	95
Table 4.1	N-H functionalization catalyzed by the cyclopropanation lineage.....	145
Table 4.2	Optimization of N-H functionalization screening conditions.....	147
Table 4.3	Parent screening for the N-H functionalization of THQ.....	148
Table 4.4	Mutation deconvolution of D7 hit.....	150
Table 4.5	Summary of the N-H Functionalization Lineage.....	151
Table 4.6	Optimization of reaction conditions.....	153
Table 4.7	The effect of phosphate buffer on product enantioselectivity.....	157
Table 4.8	Benzylamine substrate scope.....	163
Table 5.1	Si-H functionalization catalyzed by the cyclopropanation lineage.....	178
Table 5.2	Ni-NTA method with controlled ArM loading.....	179
Table 5.3	Validation and optimization of Ni-NTA screening methodology.....	180
Table 5.4	Exploration of a panel of ArMs for Si-H functionalization.....	182
Table 5.5	Further optimization of reaction conditions.....	185
Table 5.6	Summary of directed evolution lineage.....	186
Table 5.7	Rotation compared to shaking in Si-H functionalization.....	188
Table 5.8	Use of surfactants in ArM-catalyzed Si-H functionalization.....	189
Table 5.9	Effect of aniline inhibitors on Si-H functionalization.....	190
Table 5.10	POP WT-RHP activity in toluene.....	194
Table 5.11	Cyclopropanation catalyzed ArM-RHP mixtures in organic solvents.....	195
Table 5.12	ArM-catalyzed cyclopropanation in an ionic liquid.....	196

List of Figures

Figure 1.1	General directed evolution methodology	6
Figure 1.2	Metal incorporation strategies	10
Figure 1.3	Nitrobindin cofactors and Nb4-1 crystal structure	12
Figure 1.4	Structure of Sav-5-Cp*Ir and cofactors	18
Figure 1.5	Structure of Mb and cofactors	26
Figure 1.6	Concerted func. mechanism catalyzed by dirhodium tetracarboxylate catalysts	32
Figure 1.7	Mechanism of dirhodium carbenoid reaction with polar nucleophiles	33
Figure 1.8	A model of the POP variant 5-G with cofactor bound	36
Figure 2.1	Directed evolution of the POP-ArM for cyclopropanation	48
Figure 2.2	Directed evolution of dirhodium ArMs using Ni-NTA immobilization	50
Figure 2.3	C-H functionalization dehydrobenzene in high throughput screening conditions	67
Figure 3.1	Dirhodium ArM formation and catalysis	84
Figure 3.2	Mass spectra analyzing the modification of additives by dirhodium catalysts	96
Figure 3.3	Domain dynamics of apo-POP and POP-**1** ArMs	99
Figure 4.1	pH dependence of N-H functionalization	152
Figure 4.2	Substrate inhibition by tetrahydroquinoline	156
Figure 4.3	5-G and 3-E docking of THQ-enamine	159
Figure 4.4	Effect of pH on benzylamine N-H functionalization	164
Figure 5.1	The effect of pH on yield and enantioselectivity of Si-H functionalization	183

List of Schemes

Scheme 2.1 O-H functionalization of phenol..... 59
Scheme 2.2 Intramolecular C-H functionalization reactions attempted..... 65
Scheme 4.1 Aniline substrate scope..... 162
Scheme 4.2 Substrate Scope of Alkyl N-H Functionalization..... 165
Scheme 5.1 Silane substrate scope..... 187
Scheme 5.2 C-H functionalization of rhodium-binding substrates..... 192

Acknowledgements

While this document is the cumulation of my work over the last six years, it would not have been possible without the contributions of many people. I would first like to thank my advisor, Prof. Jared Lewis, whose guidance and encouragement have been essential both to the success of the projects described here and to my growth as an independent scientist. His direction in the small things, like planning and executing an experiment, and the big things, such as long-term project directions, have shaped the way I approach scientific problems. I am grateful for his mentorship and support in my future endeavors.

There are two other individuals who deserve particular recognition for their scientific contributions. I sincerely appreciate the work that Dr. Rui Huang has put into the projects described here. Rui's strengths, such as his attention to detail and systematic approach to problems, compliment my weaknesses well and have made us an effective duo. I would also like to specially thank Max Bultman. Over his two and a half years in the group I had the pleasure of mentoring him and witnessing his development into a capable synthetic chemist. His assistance in chemical synthesis was crucial in the progression of many of these projects.

I also want to thank the members of the Lewis group, past and present. During my first three years at the University of Chicago I received invaluable training and mentoring from Dr. Chen Zhang, Dr. Hao Yang, Dr. Hyun June Park, and Alan Swartz. Their guidance was a critical in my progression as an independent researcher. While we may not have directly interacted on projects, I would also like to thank Dr. Ketaki Belsare, Dr. Joe Gair, Dr. Mary Andorfer, and Dr. Krysten Jones for their support and leadership during my time at UChicago.

The transition to Indiana University was challenging in many ways but it would have been far more difficult without Natalie Chan, Christian Gomez, Harrison Snodgrass, Dr. Brian

Fisher, and Atreyi Bhattacharya who also moved to Bloomington. I sincerely appreciate the scientific, logistic, and social support that they provided. I would also like to thank those at Indiana University I have worked closely with, particularly Yasmine Zubi and Dr. Bingqing Liu, who have provided valuable insight and new ideas on the projects we have worked together on. I am grateful to our external computational collaborator Dr. Ying Li for her work on a number of projects relating to this work. I would also like to thank the more recent members of the Lewis group, Dr. Dibyendu Mondal, Dr. Xinhang Yang, Yuhua Jiang, Amardeep Kumar, Jerry Lin, Qian Du, and Prabir Saha for their helpful discussions and support. Finally, I would like to thank my committee, Profs. John Anderson and Bryan Dickinson, for their helpful feedback and support throughout the process of writing this dissertation.

Graduate school is a challenging experience and I would not have succeeded without the scientific contributions of those listed above. The personal support I have received, however, has been equally important, especially in overcoming the unforeseen challenges of moving and COVID-19. I am deeply grateful to my friends Brian Meyer, Eric Gould, Ben Slaw, and Jon Kusner, who have been there through the good times and bad. Additionally, I want to thank everyone who has brought joy to my life through inner tube water polo, dodgeball, volleyball, D&D, climbing, poker, or just going out for a beer.

A constant through this adventure has been the support of my family. I am very thankful to my parents, Karen and Steve, who have been incredibly encouraging and have been a grounding force when times got tough. I also very much appreciate my brothers, Nate and Zach, for their ability to help me slow down and enjoy life. I would also like to thank Sue, Joe, Jordan, and Taylor who have made holidays very special. Finally, I would like to thank the person who has been the most important part of my success during the last few years, Lucía Cores Sarría. I

am indebted to her for the constant love and support, even through some of the most turbulent and stressful periods of graduate school. This work would not have been possible without these people.

Abstract

This dissertation describes the engineering and evolution of dirhodium artificial metalloenzymes (ArMs) for selective carbene transfer catalysis. The ArM hybrid catalyst is formed by covalent attachment of a dirhodium tetracarboxylate cofactor a prolyl oligopeptidase scaffold, which is shown to control the environment around the metal center to enable selectivity. The process of improving this scaffold-based control through multiple carbene transfer reactions has improved our understanding of this system and ArMs more broadly. Chapter 1 begins with an introduction to selective catalysis and the inspirations for ArMs. Case studies of four well-established ArM systems are then used to identify common factors that enable control over diverse cofactor reactivities. Three factors: localization of the metal center, encapsulation of the cofactor, and evolvability of the scaffold are found to be critical in this regard. Chapter 2 describes an exploration of the reactivity and selectivity enabled by the dirhodium POP ArM. A diverse set of insertion reactions are examined, including N-H, S-H, Si-H and C-H bond functionalizations, which demonstrate the scaffold's capacity to control dirhodium catalysis. The reactions analyzed here lay the groundwork for engineering efforts in later chapters. Chapter 3 describes the directed evolution of the ArM for improved diazo coupling. The improved variants were used in a multistep, biocatalytic cascade that was possible due to scaffold-controlled chemoselectivity. Chapter 4 contains progress on the engineering of improved variants for N-H functionalization. This reaction involves an enantiodetermining proton transfer, which the ArM scaffold is found to facilitate. Finally, Chapter 5 details directed evolution and reaction engineering efforts towards improved ArM-catalyzed Si-H and C-H functionalization. The work described in each of these chapters has produced a number of highly selective ArMs and has furthered our understanding of scaffold-based control over non-native metal catalysts.

Chapter 1: An Introduction to Artificial Metalloenzyme Catalysis

1.1 Approaches to Selective Catalysis

1.1.1 Catalysis and Selectivity

Chemical catalysis enables rapid synthesis of materials that are important in all facets of life. Catalysts function by binding and stabilizing the transition state of a chemical reaction, thereby increasing the rate of the reaction.[1] As chemical understanding has progressed over the last few centuries, an abundance of catalysts for chemical reactions were developed that enable diverse transformations of functional groups.[2] Particularly impactful examples, such as the Haber-Bosch process,[3] which produces ammonia from dinitrogen, and the Ziegler-Natta catalysts[4] that enable the production of valuable hydrocarbon plastics from simple precursors, have changed the way humans live and interact with our environment. In these cases, the chemical feedstocks are simple molecules containing only a single functional group, so they require catalysts that focus primarily on decreasing reaction barriers.[5]

Functional groups imbue molecules with interesting and unique chemical properties, so synthesizing complex molecules necessitates methods to catalyze reactions on substrates containing numerous functional groups.[6] This challenge requires the differentiation of multiple potentially reactive positions on a molecule, i.e. selectivity. There are many types of selectivity that may be required in the context of catalyst design. Chemoselectivity, reacting with a desired functional group, and site-selectivity, reacting at the desired position among many comparable ones, require the energetic differentiation of similar but distinct functionalities. These selectivity concerns are typically resolved through appropriate design of the synthetic sequence and the use of protecting groups to avoid cases where there are multiple potentially reactive functional groups.

Enantioselectivity, on the other hand, necessitates the differentiation of enantiotopic groups of a single functional group through preferential binding. Catalysts with poor selectivity will likely form the desired product, but multiple side products caused by reaction at other positions, as well as over-functionalization, will result in low yields and complicated purification. Highly selective catalysts minimize these concerns, making new approaches to achieve selective catalysis a valuable endeavor.

1.1.2 Selective Transition Metal Catalysts

The unique properties of transition metals have driven their use as homogeneous catalysts for an abundance of chemical transformations, such as oxidation,[7] reduction,[8] cross-coupling,[9] olefin metathesis,[10] polymerization,[4] and C-H functionalization.[11] The wide scope of reactivity available through transition metal catalysts arises from the fact that the distinct electronic properties of each metal can be altered by ligands, which have their own electronic and steric properties. By changing the ligands directly coordinated to the metal (the primary coordination sphere) and the functionality on these ligands distal to the metal (the secondary coordination sphere), improvements of a metal complex for desired selectivity and reactivity can be achieved.

Catalyst-controlled stereoselectivity is frequently a goal of this catalyst design process. When appropriate chiral ligands are bound to the catalytic metal, a transfer of stereochemical information between the ligands and approaching substrate can occur.[1] This can take the form of attractive interactions, such as secondary sphere hydrogen bonding or electrostatics,[12–14] or, more commonly, repulsive steric interactions.[15,16] These interactions have enabled catalyst-controlled asymmetric induction for a variety of chemical transformations. A few of these, including Noyori

asymmetric hydrogenation,[17] Sharpless epoxidation,[7] and oxidation using Jacobsen's catalyst,[18] have been particularly valuable in the synthesis of chiral compounds,.

The development of catalysts that control site-selectivity, however, has proven far more challenging.[19,20] In these cases, the catalyst must differentiate similar functional groups that may be located in sterically similar but spatially distinct locations in a molecule. This situation is typically avoided altogether by masking potentially reactive functionality with protecting groups. Masking of functionality is not always desirable or possible, however, as is the case in C-H functionalization chemistry. This field has therefore led the way in the development[21,22] and utilization[23,24] of catalysts focused on site selectivity. A common strategy to achieve this type of selectivity is directing metal activity towards a single position through catalyst coordination to a proximal functional group.[25] This directing group approach has proven effective at enabling site selectivity for C-H functionalization reactions, though installing and removing directing functionality is reminiscent of the protecting groups otherwise used. An alternative approach relies simply on steric repulsion between the substrate and catalyst.[26] In a particularly powerful example of this approach, an iridium catalyst can be used to borylate the least hindered C-H bond(s) of an aromatic compound.[27] More complex interactions are necessary in the functionalization of simple alkanes, however. Davies has developed a series of bulky tetracarboxylate ligands for dirhodium catalyzed C-H functionalization reaction on simple alkyl chains that achieve very high levels of site- and enantioselectivity.[28] Slight variations to these ligands formed catalysts that were capable of selective functionalization of alkyl C-H bonds in the presence of far more activated benzyl ones.[29] The considerable bulk of these ligands formed a deep binding pocket that prevented the entry of the bulky benzyl substituents to enable selective functionalization of the smaller alkyl groups.[19] Compared with the wealth of catalysts for stereospecific chemistries, the few examples

of high levels of catalyst control over site selectivity highlights the challenges associated in developing small molecule catalysts for this chemistry.

1.1.3 Selective Catalysis by Metalloenzymes

1.1.3.1 Introduction to Metalloenzymes

The examples of selectivity outlined above are impressive, but perhaps the most remarkable examples of selective catalysis can be found in natural metalloenzymes.[30,31] As cells contain an extraordinarily complex chemical environment, life has had to evolve enzymes that react only on the desired substrate(s). It is necessary to have a broad substrate scope in some cases, such as in human cytochrome P450s that are involved in the oxidation and breakdown of organic compounds.[32–34] In many other cases, metalloenzymes demonstrate exquisite selectivity by using the protein scaffold to bind and orient a desired substrate, a feature termed molecular recognition.[30,31] An impressive example of molecular recognition can be found in the biosynthesis of taxol by the yew tree. From the initial diterpenoid carbon skeleton, nine site- and diastereoselective C-H bond functionalization reactions are catalyzed by specialized cytochrome P450s. This example is one of many that demonstrates the power of the protein scaffold for controlling the arrangement of the substrate and metal catalyst in three-dimensional space to promote reaction at a single position.

Though proteins are made up of only 20 canonical amino acids, the order (primary structure) and orientation of these in space (secondary and tertiary structure) enable a diverse set of functions. In metalloenzymes, the protein binds the metal or metal cofactor and tunes the electronics for the desired reaction. These direct interactions are in the area immediately around the metal center in the primary coordination sphere, similar to the small molecule ligands discussed

above. The true power of the enzyme scaffold is in its ability to modify the environment outside of this immediate area, called the secondary sphere. These interactions play a significant role in the high levels of selectivity that metalloenzymes can achieve by binding and positioning substrates within the active site. These remarkable levels of molecular recognition have inspired the use of metalloenzymes for the synthesis of organic compounds.

In addition to molecular recognition, there are numerous advantages to the use of biocatalysts that have led to their adoption as catalysts for chemical synthesis.[35,36] Enzymatic reactions are typically conducted in water, which reduces the use of organic solvents.[37] Additionally, enzymes are produced *in vivo*, which can reduce catalyst cost relative to many transition metal catalysts. This is done by introducing exogenous DNA containing the heterologous gene into a cellular host, such as *Escherichia coli* or *Saccharomyces cerevisiae*, followed by gene expression.[38,39]

1.1.3.2 Directed Evolution

The characteristics of a biocatalyst that are required for their use in organic synthesis, such as tolerance to elevated temperatures, organics solvents, or high substrate concentrations, may not be present in a wild-type enzyme.[40] However, pioneering work from the group of Prof. Frances Arnold[41–43] demonstrated that enzymes can be engineered for improved catalysis through an approach termed directed evolution. As shown in Figure 1.1, this process begins with the generation of variants of the protein of interest. The activities of these variants are evaluated and a variant with desired improvements is selected as the parent for another round of mutagenesis and screening. This iterative process has been used extensively to improve biocatalysts for desired catalytic characteristics.[36,44,45] While the scope of natural metalloenzyme catalysis is extensive

and can be guided towards new selectivity or substrate preference using directed evolution, there are reactions of great synthetic importance that are not performed in nature.

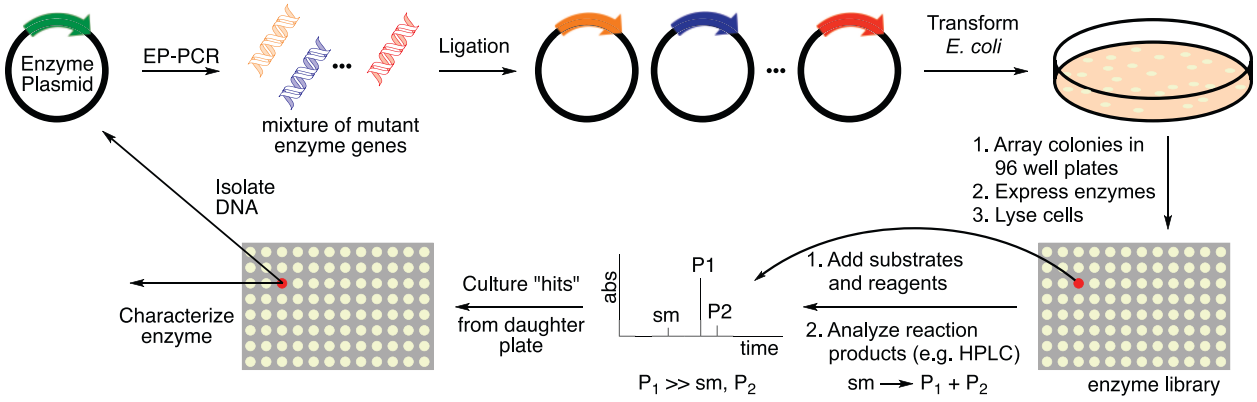


Figure 1.1 General directed evolution methodology.

1.1.4 Expanding the Scope of Metalloenzyme Catalysis

Repurposing natural metalloenzymes for non-native chemistry is one way that the limited scope of reactivity has been overcome.[46–48] Enzyme repurposing can take many forms, from supplying a natural metalloenzyme with unnatural substrates to mutating the ligating residues to alter the electronics and reactivity the metal center. While there are numerous examples of metalloenzyme repurposing, engineering heme-binding proteins for non-native chemistry is demonstrative of the potential of this approach for expanding the reactivity enabled by natural metalloenzymes.[46,49] This class of proteins has been the target of many of repurposing efforts due to the impressive oxidation chemistry the heme cofactor can catalyze.

Heme has many biological functions. In hemoglobin and myoglobin, its purpose is the binding and transport of oxygen. In these cases, a histidine binds the iron in the axial position, tuning the electronics to favor reversible binding of oxygen. In cytochromes P450 where the axial residue is a cysteine, the function is very different.[32] These enzymes catalyze the oxygenation of

a number of organic molecules through the formation of Compound I, an Fe^{IV}-oxo radical cation. It follows that the comparable Fe-nitrene or Fe-carbenes could have similarly impressive activity to the oxo. Indeed, Dawson and Breslow established more than thirty years ago that cytochromes P450 can catalyze nitrene insertion reactions by providing iminoiodinane substrates to purified microsomal enzymes.[50,51] While this work received limited attention in the years that followed, this capacity for non-native activity was eventually pursued in repurposing efforts, leading to the development of numerous heme-binding metalloenzymes capable of catalyzing nitrene[52,53] and carbene transfer[54–56] reactions. A critical finding in the engineering of these repurposed metalloenzymes was the mutation of the axial ligand from cysteine to serine, shifting λ_{max} from 450 nm to 411 nm and leading to their designation as P411s.[57] These P411s have shown to be evolvable for a number of insertion reactions with excellent selectivities and total turnovers.[46,56] In addition to standard directed evolution methods, the activity and selectivity of these repurposed metalloenzymes have been altered by the introduction of non-canonical amino acids, further expanding the possibilities for non-native catalysis.[58] While these examples show that metalloenzyme repurposing can expand the scope of reactivity by considerable amounts, these enzymes are still limited by the inherent properties of the metal ion, typically ones that are abundant in Earth's crust.

1.2 Artificial Metalloenzymes

1.2.1 Introduction to Artificial Metalloenzymes

In addition to continuing to expand natural reactivity through enzyme repurposing, metal cofactors with the desired reactivity can instead be incorporated into suitable protein scaffolds. The resulting artificial metalloenzymes (ArMs) can then serve as starting points for subsequent evolution efforts to improve the activity and selectivity of transition metal catalysts.[61] The profound influence that protein scaffolds have over the reactivity of cofactors in native metalloenzymes,[31] as discussed above, suggests that similar levels of control over synthetic catalysts could potentially be achieved. Though some of the earliest investigations into ArMs occurred more than 40 years ago,[62] interest in this area has accelerated due to advances in directed evolution and methods for the construction of ArMs.[61]

There are limitless combinations of metals and protein scaffolds that can be used to form an ArM but, as we have begun to explore the breadth of possibilities, it has become clear that decisions made early in the design of an ArM system are critical for successful, scaffold-controlled catalysis in later stages. After one selects a metal cofactor of interest, the two most important factors for ArM construction are 1) how it will be attached to the protein scaffold, and 2) what that protein scaffold will be. These factors will eventually apply limitations to the system, so it is necessary to understand the advantages and drawbacks of the numerous options available.

Four methods have been broadly used to bioconjugate metal cofactors to protein scaffolds: i) covalent, ii) supramolecular, iii) dative, iv) metal substitution.[61] As shown in Figure 1.2, covalent bioconjugation proceeds through an irreversible reaction between the metal-containing cofactor and an amino acid in the protein scaffold. Both canonical, such as cysteine,[63] and noncanonical amino acids, including azidophenylalanine,[64] have been used to form the covalent

protein-cofactor bond. Supramolecular anchoring takes advantage of high affinity binding of certain small molecules by proteins to non-covalently anchor a metal-containing cofactor, exemplified in the streptavidin-biotin ArM system that has proven very flexible.[65] It is also possible to ligate the metal directly to residues in the scaffold.[66] This method of dative anchoring has been used to generate ArMs without a synthetic primary coordination sphere. The final approach, metal substitution in a naturally-incorporated cofactor, enables non-native metal reactivity within a metalloenzyme scaffold.[67] These approaches provide a diverse set of routes for the incorporation of a metal within a protein scaffold, though there is significant interconnectivity between these aspects. Covalent bioconjugation can be used with any scaffold while supramolecular anchoring and metal-replacement strategies require scaffolds that tightly bind small molecule ligands. Metal-binding sites are extremely common in proteins, but the active site geometry required by an unnatural metal may considerably reduce the number of scaffolds available for dative anchoring.

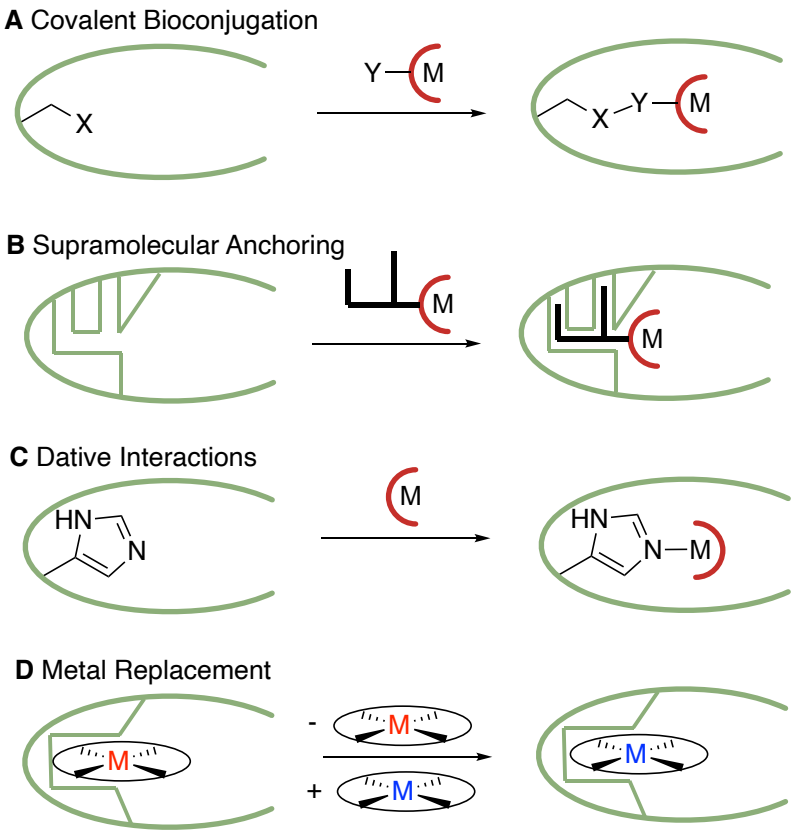


Figure 1.2 Metal incorporation strategies.

Within the constraints of the bioconjugation strategy, selection of a scaffold is critical for successful ArM catalysis. There are an expansive set of scaffolds that have been pursued in the formation and engineering of ArMs,[61] but only a limited number of these have proven capable of improving metal catalysis. We are particularly interested in cases where a single ArM scaffold is engineered to control the reactivity and selectivity of multiple catalytic reactions. This "scaffold promiscuity" can take many forms, but in all cases it requires control over the primary and secondary coordination spheres of the metal catalyst, one of the primary design goals of ArMs. As this level of control is not universal, gaining an understanding of the factors that impact promiscuity in a scaffold would aid in the design of highly selective and active ArMs.

An analysis of four ArM systems was performed and, as detailed in the following sections, three characteristics were often found to be present in promiscuous scaffolds. First, localization of the cofactor, defined here as maintaining a single position and orientation of the metal catalyst within the scaffold, is important for preserving a consistent set of scaffold-cofactor interactions throughout catalysis. There are many aspects of ArM design that impact localization, including bioconjugation strategy, cofactor design and scaffold selection, which are highlighted throughout these case studies. Additionally, the extent to which the scaffold encapsulates or surrounds the cofactor impacts its ability to influence the chemical environment around the metal. Examples of both high and low levels of encapsulation demonstrate the effects this can have on catalysis. Finally, the evolvability of a scaffold is critical for optimization of desired reactivity. Improvements to ArM catalysis through evolution are discussed in each of the case studies below, as are the equally important cases of when directed evolution *cannot* solve a particular challenge due to inherent ArM characteristics. Each of these features are considered in the context of four case studies with the goal of guiding future ArM design efforts.

1.2.2 Nitrobindin ArMs

Nitrobindin (Nb) from *Arabidopsis thaliana* is a small (20 kDa) β-barrel protein that, in its native role, contains a ferric heme cofactor to bind and transport nitrous oxide. After removal of heme, the interior of the protein is roughly 8 nm in diameter, which provides ample space to host and control the primary and secondary coordination spheres of an unnatural metal catalyst (Figure 1.3).[68] Nb has been used as the scaffold for a diverse set of metal cofactors, and varying levels of selectivity for Nb ArM-catalyzed reactions have been reported. Both independently and as a team, the Hayashi, Schwaneburg, and Okuda groups have used this scaffold for the formation of ArMs,

using cysteine bioconjugation to covalently attach a variety of maleimide-linked metal cofactors (Figure 1.3).[68]

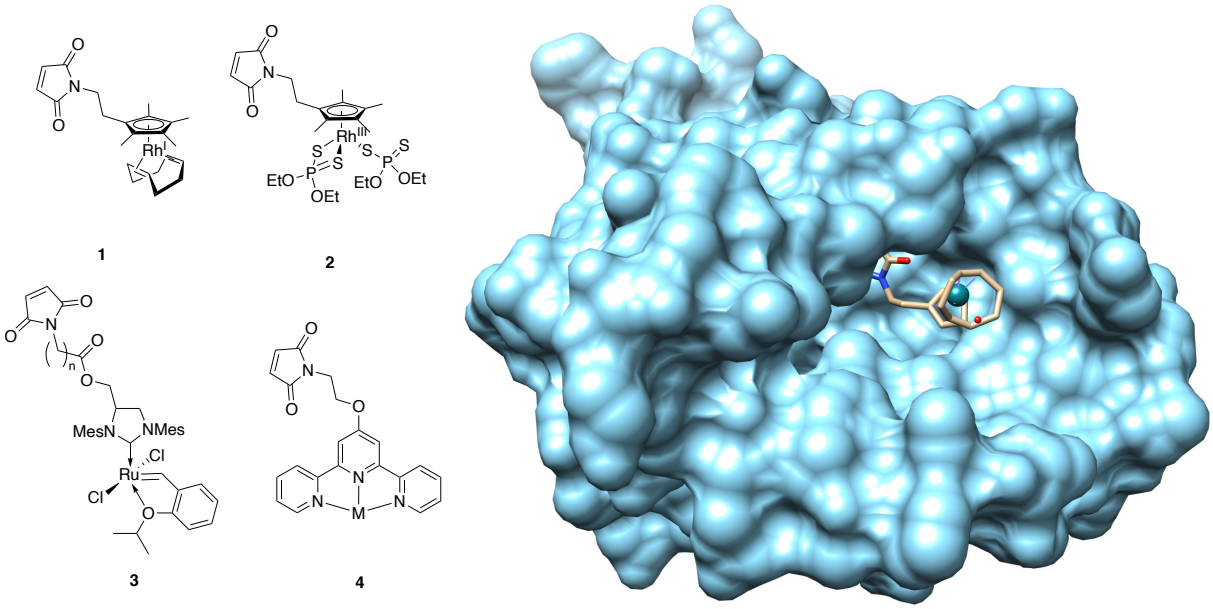


Figure 1.3 Nitrobindin cofactors and Nb4-**1** crystal structure (PDB: 3WJC). The scaffold is shown as a blue surface with the cofactor shown in beige sticks.

The first cofactor investigated using this scaffold utilized a Cp*Rh^{I} cofactor for phenylacetylene polymerization.[69] Initial studies compared the polymerization reaction between three catalysts: the Cp*Rh^{I} catalyst alone in organic solvent, an ArM with a surface-exposed cofactor (Nb-A125C-**1**), and an ArM with cofactor linked near the mouth of the β-barrel (Nb-Q96C-**1**). The trans/cis ratio of the polymer product was identical (7:93) for the small molecule catalyst and Nb-A125C-**1**, confirming that, when lacking scaffold interactions, the cofactor performs similarly in aqueous buffer and organic solvent. In Nb-Q96C-**1**, however, the cofactor is

linked in a position surrounded by interior residues that can modify the environment around Rh^{I} during catalysis. These interactions were found to change the selectivity of polymerization, with this ArM achieving 53:47 trans:cis selectivity.

Scaffold engineering was performed using a structure-guided approach[70] based on modeling of Nb-Q96C-**1** that highlighted eight amino acids in close proximity to the potential cofactor. Polymerization catalyzed by ten variants containing combinations of mutations at these positions were examined. These variants were constructed with the goal of altering the active site volume, but no significant increases to the trans:cis ratio was observed. The best variant was found to be Nb-H76L-Q96C-H158L-**1** (Nb4-**1**), which achieved a ratio of 82:15 with two histidine to leucine mutations. A crystal structure of this variant was obtained, which shows the Cp* ligand positioned in a hydrophobic pocket near the H158L mutation with high occupancy. The high localization of cofactor in this position suggests that, at least in the static environment of a crystal lattice, the ArM scaffold controls the position and orientation of the Cp*-Rh^{I} ligand. The authors suggest that this effect may be responsible for the improved trans-selectivity, but without crystal structures of other less selective variants, the mechanism of improvement cannot be determined with confidence.

With this Nb-**1** ArM system established, the Hayashi group explored reactions performed by analogous catalysts, such as the cycloaddition of alkynes to acetophenone oximes catalyzed by the Cp*Rh^{III} cofactor **2**.[71,72] After activation of this catalytic center by ligand exchange with Ag^{I}, this cycloaddition reaction was found to be catalyzed with low yields by Nb-Q96C-**2**. Interestingly, introducing carboxylate mutations near the Rh^{I} cofactor was found to increase yields, which the authors attribute to increased access of the Ag^{I} ion into the hydrophobic NB cavity due to electrostatics. Though not addressed in the initial report, it should be noted that Cp*Rh^{III}

complexes are known to bind to carboxylates. This type of coordination has been found to be beneficial in ArM systems based on similar cofactors[73] so a more thorough investigation into this potential primary sphere effect would aid in understanding how the scaffold is modulating catalysis.

After demonstrating this reaction was successfully catalyzed by activated Nb-**2** ArMs, the authors turned to directed evolution to improve the activity of the catalyst. Development of high-throughput screening methods for the directed evolution of ArMs is frequently challenging due to background activity catalyzed by free cofactor, which the authors overcome using an adapted resin-purification approach[74]. Twenty-three positions at the mouth of the β-barrel near the metal center were explored using site-saturation mutagenesis (SSM). The final variant found after 3 rounds of directed evolution, Nb-Q96C-T98H-L100K-K127E-**2**, demonstrated 4.9-fold improvement in activity compared to parent.[72] The location and identity of these mutations can once again affect the rate of Ag^{I} activation or alter the environment of the Cp^*Rh^{III} cofactor through primary or secondary sphere interactions.

In addition to Cp^*Rh^{I} catalysis, a variety of other cofactors have been investigated. A maleimide-linked diiron dithiolate complex mimicking the active site of [FeFe]-hydrogenases prepared and bioconjugated to the Nb-Q96C scaffold.[75] In the presence of a $[Ru(bpy)_3]^{2+}$ photosensitizer and visible light, the resulting ArM was found to catalyze the desired proton reduction. Unfortunately, the rate of hydrogen evolution was found to be lower using the Nb scaffold than the free cofactor alone, which could be due to the Nb scaffold decreasing access of the bulky photosensitizer to the diiron center. No engineering of the scaffold was performed in this work, so the extent to which the Nb scaffold can control this cofactor is not clear.

A third reaction platform, ruthenium-catalyzed olefin metathesis, was developed using the nitrobindin scaffold.[76] Initial attempts at bioconjugation of a maleimide-linked ruthenium cofactor **3-C3** to an Nb scaffold with a small cavity size resulted in poor bioconjugation efficiency of less than 25%. A variant with an expanded cavity volume, Nb-L75A-H76L-Q96C-M148L-H158A-**3-C3**, was found to achieve much better bioconjugation (89%) and was, therefore, used for an exploration of activity. Ring-closing metathesis (RCM) on a hexadiene was quantitatively catalyzed by the ArM using 1 mol% loading and to 57% yield with 0.3 mol%. Ring-opening metathesis-polymerization (ROMP) was also investigated, with an impressive 9,900 total turnovers catalyzed by the ArM compared to the 4100 by the cofactor in organic solvent. The scaffold was also found to modify the selectivity of the product polymer, with an ArM cis:trans ratio of 48:52 compared to the 70:30 for the cofactor alone. Though no directed evolution was performed as a part of this work, the authors approached the engineering of the Nb scaffold from a different direction.

Increasing the chain length and cavity volume were found to be necessary for bioconjugation of the bulky ruthenium cofactor **3**, but these compromises may reduce the interactions between the metal center and protein scaffold by extending the cofactor into solution. To avoid this, the interior of Nb scaffold was expanded by incorporating two additional β-strands into the 10-stranded Nb β-barrel. The new strands, 29 amino acids in total, were duplicates from another section of the Nb4 scaffold and resulted in a significant change to the internal volume, calculated to increase from 855 $Å^3$ to 1389 $Å^3$ in the expanded Nb4 (Nb4exp). This engineered scaffold was slightly less thermostable, with a decrease in T_m from 52 °C for Nb4 to 45 °C but was similarly stable under the reaction conditions. Consistent with the hypothesis that the Nb4 active site was too restricted for the bioconjugation of short (**3-C1** and **3-C2**) linkers, bioconjugation of

3-C1, **3-C2**, and **3-C3** to Nb4exp proceeded in greater than 90% efficiency. The expanded scaffold was used for RCM and ROMP and cross metathesis (CM) reactions, where it was found to catalyze the reaction in higher total turnovers than Nb4 in each case. While no mutagenesis was performed to tune the expanded scaffold for metathesis, this example demonstrates that improving the encapsulation of the cofactor by the scaffold increases its ability to alter the primary and secondary coordination spheres of the metal.

The importance of encapsulation is further exemplified in the final class of nitrobindin-based ArMs. Maleimide-linked terpyridine cofactors **4** containing Cu^{2+}, Zn^{2+}, and Co^{2+} have been linked to Nb-Q96C and used as catalysts for Diels-Alder reactions.[77] Unfortunately, the Nb scaffold was not found to have a significant impact on the reaction with any of the three metals, as the yields and endo/exo selectivities were similar to the small molecule catalysts alone and no enantioselectivity was observed. This system was also used in a separate effort where a Mn^{2+}-terpyridine cofactor was used to catalyze the oxygenation of C-H bonds and epoxidation of olefins,[78] but once again, minimal scaffold-controlled selectivity was observed. Modelling of the cofactor **4** bound to Nb-Q96C was performed in these reports and suggested that the cofactor protruded into solution, positioning the catalytic metal away from influence of the scaffold.

Notably, Nb-Q96C-**4** ArMs are the only nitrobindin-based systems that produce a small molecule chiral product. The absence of either enantio- or geometric-selectivity indicates that the scaffold is not able to sufficiently control the primary and secondary coordination spheres in these cases. Modeling of these systems suggests that the metal is exposed to solution where limited interactions with the scaffold can occur. These interactions are critical for controlling the positioning of small molecule reactants during catalysis, the lack of which explains why no selectivity is observed. Though geometric selectivity is observed in both ROMP and

phenylacetylene polymerization reactions, the considerable bulk of the growing polymer may be sufficient to interact with the scaffold to alter the selectivity. If this were the case, it is unlikely that directed evolution could be used to overcome this challenge, as a lack of scaffold-cofactor interactions would make mutagenesis ineffective. Engineering efforts outside of simple mutagenesis, such as the use of the expanded nitrobindin scaffold Nb4exp, may improve encapsulation of the cofactor and result in scaffold-based control over a variety of artificial metalloenzyme reactions, but this remains to be seen.

1.2.3 Streptavidin ArMs

The streptavidin (Sav) scaffold is likely the best explored scaffold for ArM formation. This tetrameric protein binds biotin with sub-picomolar affinity, a property that has been exploited in the formation of ArMs from their inception. Indeed, the first example of an ArM prepared from a non-metalloenzyme scaffold was in 1976 by Wilson and Whitesides,[62] who used a biotinylated Rh^{I}-diphosphine complex to anchor the catalyst in avidin. The resulting ArM catalyzed the hydrogenation of a dehydroamino acid derivative with 41% e.e. and 500 turnovers, clearly demonstrating that a protein scaffold can be used to control the environment of an unnatural metal catalyst. Since this seminal work, Sav has been employed as the scaffold for numerous cofactors by Ward and others, which have made benefits and limitations of this scaffold evident. This section will discuss a few of these in systems in depth, with a focus on selectivity and control over the coordination spheres.

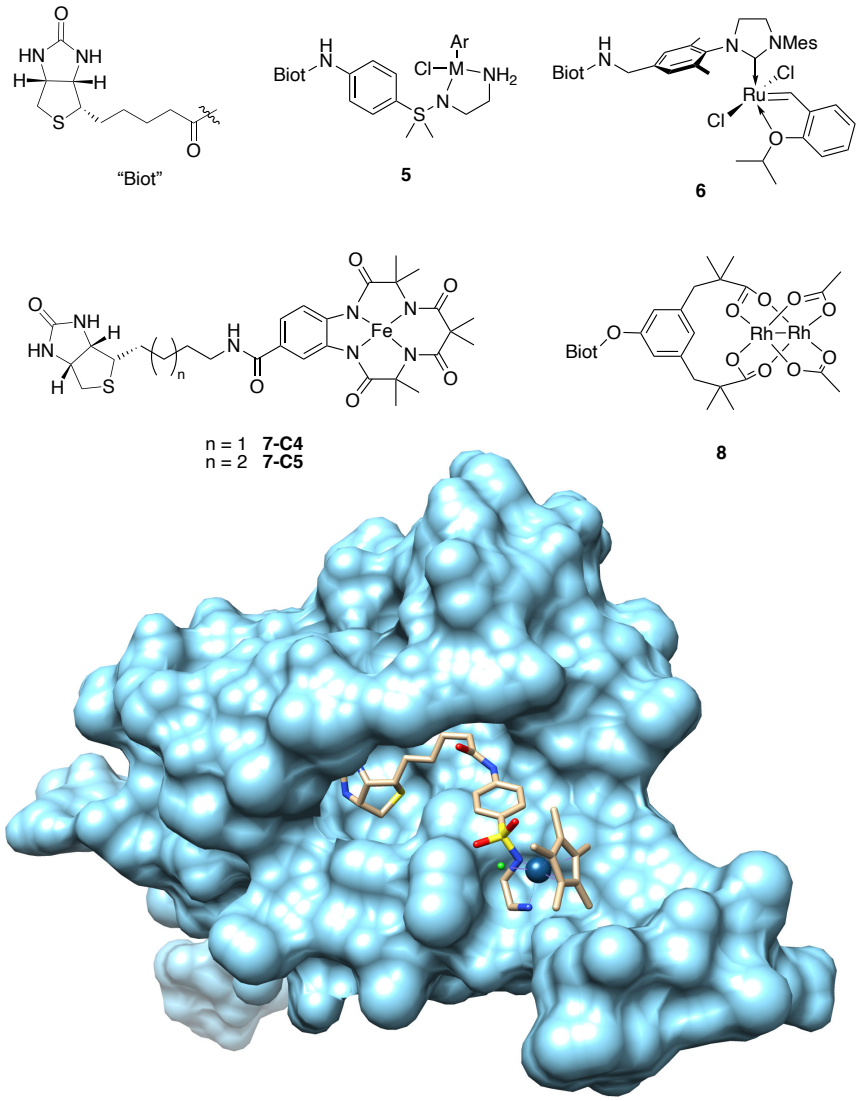


Figure 1.4 Crystal structure of Sav-**5**-**Cp*Ir** (PDB: 3PK2) showing only a single monomer and cofactors discussed in this section.

The streptavidin scaffold has been exploited to a great extent for transfer hydrogenation chemistry (Sav-ATHases). Early development of these ATHases explored the reactivity and selectivity of three transition metal catalysts known for this reaction, the piano-stool complexes Cp*Rh^{III}, Cp*Ir^{III}, and Ar*Ru^{II}.[79,80] Each of these was biotinylated through a sulfonamide-ethylenediamine linking ligand, the features of the which, including the linker length and arene

geometry, were found to substantially alter the selectivity of the reaction. These features were analyzed combinatorially, and the best was found to be the Ar*Ru^{II} complex of cofactor **5** (Figure 1.4), which catalyzed the transfer hydrogenation of acetophenone to give the R-enantiomer of the desired product in 91% e.e. The other metals were found to be less selective, though Cp*Rh^{III} formed the primarily the opposite enantiomer. The authors suggested that the origin of selectivity of selectivity for Sav-**5**-**Ru** is through primary coordination sphere control over of the metal and piano stool ligand.[80] Though the metal center can racemize readily at room temperature, the Sav scaffold exerts a strong preference for one of the orientations, effectively enforcing metal chirality. The chiral Ar*Ru^{II} center then transmits that information during hydrogenation. In agreement with this notion is a crystal structure of the Sav-S112K-**5**-**Ru** ArM, which, though it has low cofactor occupancy, contains exclusively one orientation of the piano stool arene.[81]

Additionally, the authors sought to improve the Sav scaffold for selective transfer hydrogenation. A docking analysis of Sav-**5**-**Ru** found that S112 points directly at the metal center,[80] indicating it may be influencing the primary and secondary coordination sphere. This position was mutated to each of the 19 amino acids, with S112Y proving to be the most (R)-selective variant (97%). Interestingly, the variant S112K favors the opposite enantiomer with 20% (S) enantioselectivity, highlighting the ability of the scaffold to control and alter the environment of the catalytic metal. These initial investigations into the development of selective Sav-ATHases demonstrate the ability of the scaffold to control the primary coordination sphere of the metal catalyst. This system has progressed in many exciting directions which cannot be described in detail here, but there are a number of excellent reviews that detail these advances.[65,81,82]

The Sav scaffold has also been used for ArM-catalyzed metathesis chemistry.[83] The biotinylated Grubbs-type ruthenium cofactor **6** was found to bind to Sav and catalyze an

intramolecular metathesis reaction to form the fluorescent hydroxycumene. The ArM was completely inhibited in the presence of the glutathione, indicating that the ArM is not capable of controlling substrate access to the metal center. Glutathione is an antioxidant present in the cytoplasm of *E. coli*, though it is not present in the periplasmic space. Thus, when the Sav scaffold was tagged to facilitate localization in the periplasm, the cofactor could enter the cell and bind to the scaffold in this region while maintaining activity. This enabled the development of an *in vivo* screen of Sav-**6** variant activity, which was used to evolve increased metathesis activity. Over 5 rounds of directed evolution, the final variant exhibited V_{max} and K_{cat}/K_M values double that of the parent. Crystal structures of the parent and final ArM variants were obtained, which suggested that a loop between the biotin binding pocket and the protein-protein interface is significantly more flexible in the final variant. This loop between residues 112-122 would be impacted by the mutations T114Q, A119G and V47A (at the protein-protein interface), and the authors posit that the increase in flexibility may be responsible for the observed increase in selectivity. Though the mechanism of improved activity is not known, the distance between this loop and the cofactor suggests that subtle alterations to the secondary coordination sphere may be occurring.

Recently, the Ward group reported an Fe^{III}-based ArM for C-H functionalization utilizing the Sav scaffold.[84] The biotinylated Fe(TAML) cofactors **7-C4** or **7-C5** were found to catalyze the oxidation of sp^3 C-H bonds of ethylbenzene in the presence of H_2O_2. An initial screen of Sav variants containing mutations at positions 112 and 121 with both cofactors found that Sav-S112R-**7-C4** catalyzed the reaction with the highest enantioselectivity (32%) but only 57 turnovers. These cofactors differ by only a single carbon in the linker but of the 32 variants tested, **7-C4** had non-zero enantioselectivity in 13 of the variants, while this was only true with two variants using **7-C5**. Linker length is frequently observed to be impactful in Sav-based ArMs, but the considerable

differences in selectivity between the C4 and C5 linkers in this case highlight the importance of localization of the metal center. The extra carbon in **7-C5** apparently projects the cofactor away from the scaffold, minimizing opportunities for scaffold to influence catalysis and decreasing localization of the metal center. In support of this notion, the crystal structure of Sav-**7**-**C5** shows 100% occupancy of the biotin anchor, but the linked Fe^{III}(TAML) is disordered and cannot be identified.

Overoxidation of the alcohol, a known problem in small molecule C-H oxidation chemistry, is observed to a significant extent here. The Sav variants that were analyzed in depth did not demonstrate significant chemoselectivity for ethylbenzene over the alcohol. While the mechanism of Fe^{III}(TAML) C-H oxidation is substrate dependent, the homolytic cleavage of the substrate C-H bond necessarily occurs.[85] The low levels of enantioselectivity observed using Sav-S112R-**7**-**C4** suggests that the scaffold is capable of controlling the position of the substrate during C-H cleavage and rebound, though to only a limited extent.

A final Sav-ArM system worth examining utilizes Rh_2^{II}(tetracarboxylate) cofactors for carbene transfer catalysis.[86] Six biotinylated cofactors containing different carboxylate ligands and linkers of varying construction and lengths were prepared, such as the representative cofactor **8**. These were anchored in the WT Sav scaffold and the resulting ArMs used to catalyze the cyclopropanation of styrene with methyl phenyldiazoacetate. Despite the diversity in cofactor construction, turnover numbers to the product were consistently lower than the cofactors in the absence of Sav and no enantioselectivity was observed in any case. As was done in the previous example, a panel of variants with mutations at positions 112 and 121 were screened using cofactor **8**, but no enantioselective variants were found. In order to probe reasons for this lack of reactivity, the authors docked the **8** into the Sav scaffold. This model showed that the cofactor was positioned

a considerable distance from the scaffold, with the shortest Cα–Rh distances of 6.1 Å, 6.0 Å, 7.4 Å to the closest amino acids, S112, K121 and L124, respectively. As shown below in Figure 1.6, dirhodium catalyzed cyclopropanation proceeds through a single, concerted step without attractive interactions between the styrene and the dirhodium cofactor. Achieving selectivity in this reaction mechanism requires control over the secondary coordination sphere to properly orient the styrene before reaction at the dirhodium carbenoid occurs.

These two examples indicate that the partially exposed active site of Sav-based ArMs has inherent limitations in its capacity to control secondary coordination sphere interactions. In both Fe^{II}- and Rh_2^{II}-Sav-ArMs, these interactions are required to position a substrate with respect to a reactive intermediate in order to attain selectivity. The lack of complete encapsulation limits the ability of the Sav scaffold to achieve robust control over the secondary coordination sphere, but, as was the case in the engineering of an expanded nitrobindin scaffold, engineering of Sav to overcome this limitation has been pursued. Multiple reports have engineered additional protein structures in the space surrounding the active site with the goal improving scaffold-based control over the secondary coordination sphere. An initial effort introduced new protein motifs with sizes that varied from a 35 amino α-helix loop to a 65 amino acid small domain.[87] The chimera ArMs were used to catalyze transfer hydrogenase and metathesis reactions where the chimeras frequently outperformed Sav. It is difficult to interpret the extent to which these additional structural motifs impact the secondary coordination sphere, however, as the reactions investigated do not require high levels of control, as described above. A recent report[88] refined this approach with the introduction of SOD domains, which are known to dimerize. This interaction between neighboring SOD domains formed a closed structure around the active site, which proved vital for favoring the desired reaction pathway in the multi-cofactor hydroaminase chemistry performed by the ArMs.

The examples of Sav-based ArMs discussed in this section show that it is a scaffold capable of controlling a number of reactions. The supramolecular bioconjugation of a biotinylated cofactor is a straightforward and highly efficient method, as long as biotin and its installation is tolerated by the metal cofactor. This method also provides an opportunity to tune the position of the metal in the active site by increasing or decreasing the linker length. An interesting outcome of this bioconjugation strategy is the near-identical location of catalytic metals across many unrelated ArMs. This results in two amino acids positioned in this area, S112 and K121, frequently impacting the selectivity and activity of an ArM. This knowledge greatly decreases the effort required in finding improved variants, but it may also hinder the exploration of potentially impactful mutations in other areas of the scaffold, such as the 112-122 loop that was found to have a far greater impact on metathesis than mutations at positions 112 and 121 alone. A combination of cofactor design, linker length optimization, and mutations around the scaffold have been used to engineer numerous highly selective Sav-based ArMs, especially for reactions where only control over the primary coordination sphere is required. The partially exposed active site of Sav, however, has been found to be insufficient in its influence over the secondary coordination sphere when a reaction requires a complex arrangement of unbound substrates. The Sav-chimera engineering efforts are an exciting development that could prove impactful in this regard, though demonstration of these scaffolds in reactions requiring secondary sphere organization will be required.

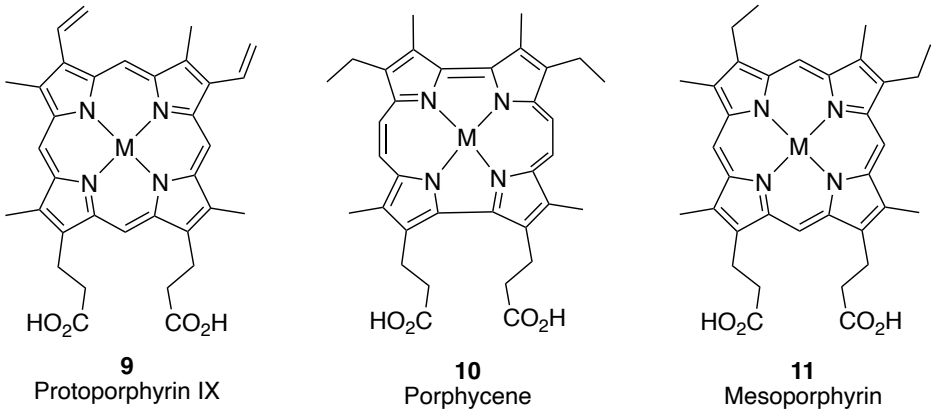
1.2.4 Myoglobin ArMs

Myoglobin (Mb) is a monomeric heme-binding protein that binds oxygen for short-term storage.[68] Its globular structure consists of 8 α-helices that form a shallow pocket where heme **9-Fe** is bound and contains a histidine at position 93 that axially coordinates to iron. The identity of

this axial residue plays a significant role in tuning the electronics of the heme-iron in different heme enzymes. Oxygen-binding proteins contain the histidine observed in WT Mb, while the iron in oxygen-reducing cytochromes P450s (CYPs), for example, is ligated by an electron-donating cysteine. In addition to the axial residue, there are also significant differences in the size and functions of the active sites of between these classes of proteins, with Mb containing a compact scaffold environment near the heme-iron that is ideal for O_2 binding and CYPs having relatively spacious active sites to accommodate various large organic substrates.

As discussed in section 1.1.3.1, both Mb[89] and CYPs[49,90,91] containing the natural heme cofactor have been repurposed for a variety of non-native reactions. For ArMs, however, other than a recent report using a CYP scaffold for ArM catalysis,[92] the majority of ArM engineering efforts taking advantage of heme-substitution for anchoring the non-native cofactor use Mb from *Physeter macrocephalus*, the sperm whale (*Pma* Mb), or *Equus caballus*, the horse (*Eca* Mb). Early investigations into cofactor replacement of Mb introduced modified-heme cofactors with substituted arenes attached to the heme propionates.[93] This resulted in a turn-on of peroxidase activity, but the origin of this effect was never conclusively determined. The first non-iron Mb-ArM was prepared by Hayashi and coworkers, who incorporated the Mn^{III}-substituted porphyrin **9-Mn** and structural analog porphycene **10-Mn** into *Eca* Mb.[94] The porphycene-containing ArM Mb-**10**-**Mn** displayed sp^3 C-H bond hydroxylation activity on ethylbenzene in the presence of H_2O_2, catalyzing 13 total turnovers with no enantioselectivity. Interestingly, metal-substituted Mb-**9**-**Mn** displays no oxidase activity, suggesting that the electronic and steric perturbations caused by the **10**-**Mn** cofactor are critical for activity. While oxidation of thioanisole is observed with the WT Mb-**9**-**Fe** and Mb-**10**-**Fe**, no C-H functionalization activity is in these cases either. A follow up study[95] confirmed the hypothesis that the mechanism of Mb-**9**-**Mn** oxidation proceeds through

a Mn^{V}-oxo intermediate with subsequent rebound. As this was the first example of ArM catalyzed sp^3 C-H oxidation, this work is certainly impressive, but the lack of selectivity and low turnover numbers make the scaffold's role unclear. It is certainly enabling the reactivity of **9-Mn** and protecting the Mn^{V}-oxo intermediate from undesired side reactions, but without a deeper investigation into the scaffold's role or improvement of the scaffold through evolution, it is not clear to what extent the Mb scaffold is controlling the primary and secondary coordination sphere during catalysis.



9
Protoporphyrin IX

10
Porphycene

11
Mesoporphyrin

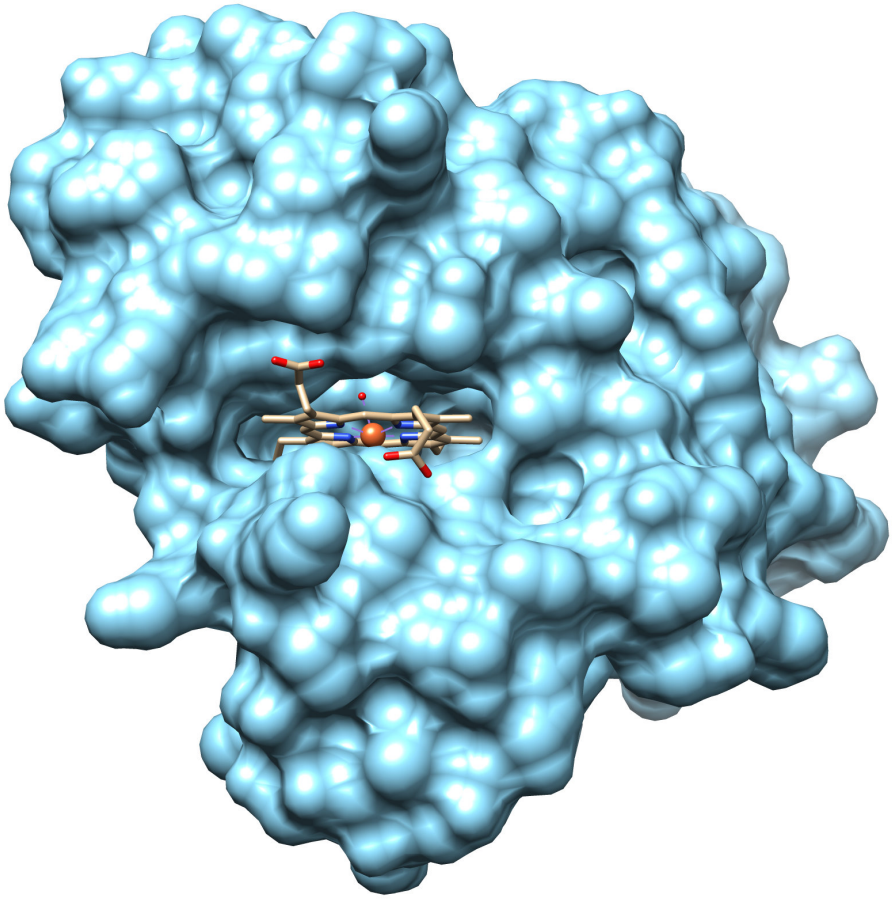


Figure 1.5 Structure of the cofactors used with Mb. The Mb crystal structure (PDB: 1MBN) shows heme (**9-Fe**) bound in the active site.

Fortunately, there are other Mb-ArMs where these effects have been investigated, such as in the carbene transfer activity of a *Pma* Mb-based ArM containing the Ru^{II} mesoporphyrin **11-Ru**.[96] In this work, the authors explore the N-H functionalization of anilines and cyclopropanation of styrenes catalyzed by the Mb-**11-Ru** in the presence of EDA. A significant background reaction was catalyzed by the **11-Ru** cofactor, so all Mb-ArMs were purified by size-exclusion chromatography before use. The N-H functionalization of aniline was catalyzed by four Mb variants containing different residues at position H64, the distal histidine. The free cofactor in solution yielded higher amounts of product (48%) in all cases except the variant Mb-H64A-**11-Ru** which resulted in a 52% yield at pH 7. In the cyclopropanation of styrene, the ArMs yield an unexpectedly low amount of product and with no significant change in selectivity compared to the cofactor alone. The authors found that in this case, scaffold modification by the carbenoid intermediate was occurring during catalysis, with 100% of the ArM containing at least one addition of EDA. Interestingly, treatment of WT Mb-**9-Fe** to similar conditions also resulted in modification, but only ~50% of the enzyme with a single modification. Scaffold-based control over the reaction is poor in this system, but rather than too little interaction with the cofactor as has been the case in other systems, the Mb scaffold interacts too closely with the catalyst. Varying H93 did have some effect on yield for N-H functionalization, though in no cases was the ArM scaffold able to improve activity compared to the cofactor alone.

Outside of Mb-ArMs formed by the introduction of iron-containing porphyrin analogues,[97] the only example of Mb-controlled selectivity over ArM catalysis is with the use of the cofactor **9-Ir(Me)**.[98] In this investigation by Hartwig and coworkers, a set of eight non-native metals were incorporated into protoporphyrin IX, **9**. The activity of each of these towards intramolecular C-H functionalization and olefin cyclopropanation was screened using a panel of

different *Pma* Mb variants with mutations to the axial H93 ligand. The cofactor **9-Ir(Me)** was far superior to any of the other cofactors examined in both reactions, regardless of the axial ligand. The intramolecular C-H functionalization was catalyzed with no enantioselectivity by WT Mb-**9-Ir(Me)**, but over the course of 5 rounds of directed evolution, a panel of variants were developed for that achieved good selectivity on a number of substrates, reaching 84% e.e. with around one hundred turnovers under typical reaction conditions. Additionally, intermolecular cyclopropanation, which requires control over the orientation of the olefin as it approaches the carbene, was catalyzed with 82% e.e. using the best variant. The improvements observed through engineering demonstration of the ability of a scaffold to control the secondary coordination sphere to alter selectivity.

It is worth noting that continued development of ArMs containing **9-Ir(Me)** has moved away from the Mb scaffold to a cytochrome P450, CYP119.[92,99] The larger active site in CYP119, which is known to have WT H_2O_2-dependent peroxidase activity on a number of small molecule organic substrates, was found to enable a much higher rate than was obtained using Mb. While the Mb scaffold was engineered to make use of the compact environment around the metal center to control selectivity, it was apparently at the cost of activity. An additional drawback to this system is a product of the anchoring approach, the necessity of a porphyrin-like cofactor. Variation of the porphyrin has been shown to be an effective method to alter activity,[94] but the inherent geometric constraints of a planar, 4-coordinate ligand limits the possible catalysts one can incorporate. Indeed, the Mb-ArMs discussed proceed through intermediates analogous to those in heme metalloenzymes. The Mn^{V}-oxo is comparable to Fe^{IV}-oxo, and the Ru^{IV}-carbene and Ir^{III}-carbene are similar in structure to the Fe^{IV}-carbene formed in the comparable reaction with EDA. Though Mb-ArMs benefit from a straightforward bioconjugation route and a compact active site, with only

one example of selectivity resulting from similar intermediates, it apparent that the myoglobin scaffold does not offer the "promiscuity" that is sought in an ArM scaffold.

1.3 Prolyl Oligopeptidase Scaffold for Artificial Metalloenzymes

1.3.1 POP Scaffold and ArM Formation

Prolyl oligopeptidase (POP) from the hyperthermophile *Pyrococcus furiosus* is a hydrolase that cleaves small polypeptide substrates after proline residues.[100] A crystal structure of the WT enzyme showed that, like the rest of the POP family, *Pfu* POP is comprised of two domains, a ~30 kDa peptidase domain containing a Ser-His-Asp catalytic triad to which a ~40 kDa seven-bladed β-propeller domain is linked.[101] Experimental and computational results from our laboratory and others have shown that this domain restricts substrate access to the large active site cavity with large-scale interdomain conformational changes regulating this process. The structure opens to allow substrate binding in the active site, after which the two domains close to form the catalytic triad and hydrolyze the substrate.[101]

Though these interdomain dynamics were not understood at the time, we believed that the large internal volume in the β-propeller domain would make this protein an excellent scaffold for localizing and encapsulating a metal cofactor. As this protein scaffold does not have a metal binding site or high affinity for small molecules, covalent bioconjugation was the only strategy available for linking the cofactor to the scaffold. Azide-alkyne cycloaddition is an approach that had previously been used by our group to covalently bioconjugate a metal containing cofactor with good selectivity and efficiency.[64] Briefly, this process involves genetically encoding azidophenylalanine into the protein structure using well-established amber stop codon suppression

technology,[102] and then using strain-promoted azide-alkyne cycloaddition (SPAAC) [103] to covalently link the cyclooctyne-containing cofactor **12** (Figure 1.8) to the scaffold.

In principle, this bioconjugation strategy enables the linkage of any cofactor to POP. An organic acridinium photosensitizer was bioconjugated and used for visible light-induced sulfoxidation chemistry, though no selectivity was observed.[104] More recently, Yasmine Zubi and Bingqing Liu in the Lewis group have incorporated ruthenium and iridium-based photosensitizers into the POP scaffold which altered the fluorescent lifetimes of the cofactors and imparted modest selectivity in cycloaddition reactions (unpublished). The majority of work using this scaffold, however, has been with the dirhodium tetracarboxylate cofactor **12**.[105] The diverse carbene transfer chemistry catalyzed by this class of molecules made this a promising catalyst to demonstrate scaffold-controlled reactivity and selectivity.[106–110]

1.3.2 Dirhodium Tetracarboxylate Catalysis

The mechanism of dirhodium carbenoid insertion chemistry is important in considering how the POP scaffold could be engineered to achieve control over the reactivity and selectivity of the cofactor. As shown in Figure 1.6, the dirhodium catalyst reacts with a diazo substrate and, after the rate-limiting loss of dinitrogen, forms a dirhodium carbenoid intermediate. This intermediate is very electrophilic and can react with diverse nucleophiles in a number of ways, including olefins to form cyclopropanes and another diazo compound to form alkenes. Additionally, the carbenoid can undergo a formal insertion into N-H, S-H, and O-H bonds as well as a concerted insertion into Si-H and C-H bonds. The broad reactivity of this catalyst would allow the ArM to control both the chemoselectivity and stereoselectivity of the carbene transfer. Moreover, water is a capable nucleophile in this reaction and, as it is the solvent in these reactions,

high levels of scaffold-controlled chemoselectivity to favor reaction with the desired nucleophiles over the side reaction with water would be required. While the dirhodium carbenoid formation occurs via the same mechanism in all cases, the nature of the X-H insertion changes drastically depending on the nucleophile.

Reactions with non-polar nucleophiles, such as the insertion into C-H or Si-H bonds and the cyclopropanation of olefins, were proposed to proceed through a three-center, concerted asynchronous transition state (Figure 1.6) in a theoretical investigation by Davies and coworkers.[111] They found that the activation barrier for the insertion into the electronically-activated C-H bonds of dihydrobenzene (6.5 kcal/mol) was far lower than for the unactivated C-H bonds of cyclopentane (17.4 kcal/mol), in agreement with the considerable difference in rates observed for these reactions. Interestingly, the calculated barrier to nitrogen extrusion is 11.4 kcal/mol, suggesting that in the rate-limiting step in the reaction with unactivated C-H bonds is the concerted insertion rather than the loss of dinitrogen. Overcoming this considerable barrier requires precise positioning of the substrate C-H bond with respect to the rhodium carbenoid without any attractive forces between the substrate and catalyst, making selectivity a significant challenge that can only be achieved through catalyst-controlled substrate positioning in the secondary coordination sphere. As described in section 1.1.2, exceptionally bulky tetracarboxylate ligands on dirhodium achieved this level of control, enabling the functionalization of simple alkanes with remarkable selectivity.[28]

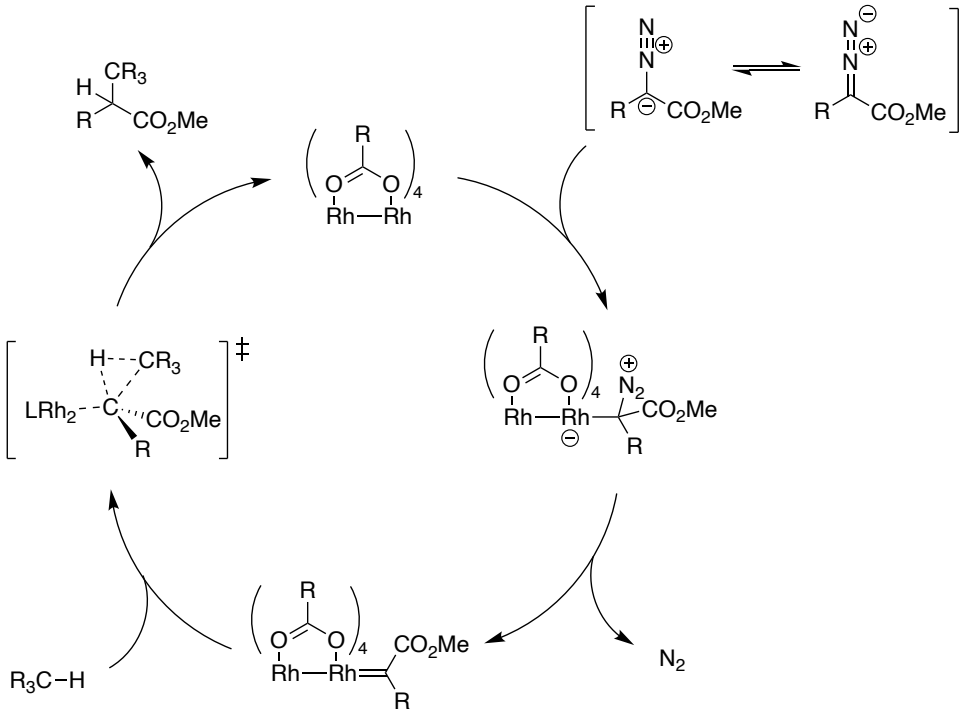


Figure 1.6 Concerted functionalization mechanism catalyzed by dirhodium tetracarboxylate catalysts.

When substrates containing a nucleophilic lone pair are used, however, the reaction proceeds in multiple steps. As exemplified by N-H functionalization in Figure 1.7, the dirhodium carbenoid is attacked by the nucleophile, forming a zwitterionic adduct. An initial proton transfer then occurs to form an enamine that remains bound to dirhodium. The proton-shuttling steps that would result in the chiral amine product, however, are found to be most favorable after this enamine dissociates from dirhodium. This presents a significant challenge for the development of chiral methods to this reaction, as chiral ligands on the dirhodium cannot affect a proton transfer that is dissociated from the catalyst. Indeed, initial attempts at N-H functionalization using chiral dirhodium ligands resulted in poor levels (<50%) of enantioinduction.

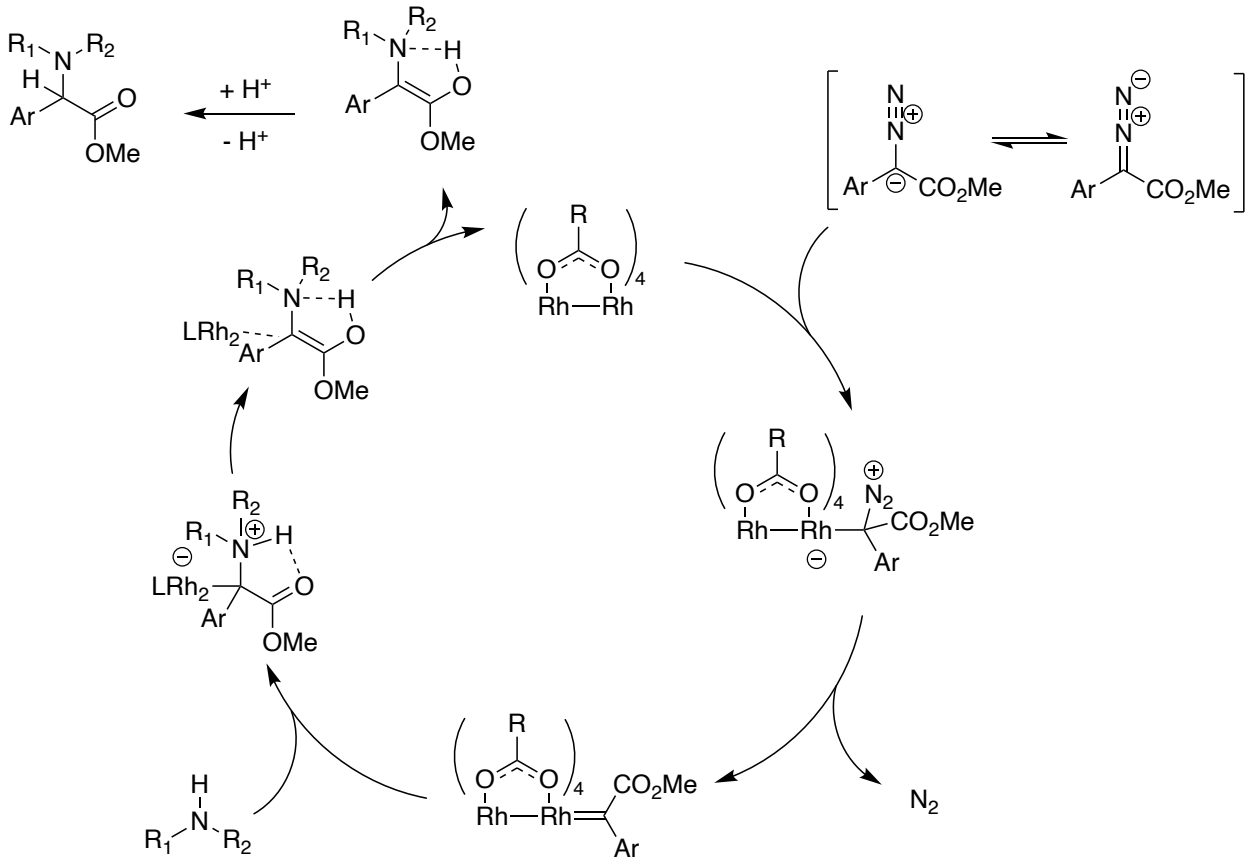


Figure 1.7 Mechanism of dirhodium carbenoid reaction with polar nucleophiles.

This mechanism has been explored in numerous theoretical investigations into N-H, S-H, and O-H functionalization reactions,[112–116] all of which have found that the most energetically favored pathway is through the stepwise process in which the planar intermediate dissociates from dirhodium as shown above. A comparative analysis of the O-H functionalization reaction pathway details the differences between dirhodium (II) tetracarboxylate and copper (I) box catalysts.[116] The copper-catalyzed reaction, which provides excellent levels of enantioselectivity with chiral ligands, proceeds through a copper-bound intermediate during the enantiodetermining proton transfer. Conversely, in the dirhodium-catalyzed reaction, proton transfer was found to be more

favorable by ~6 kcal/mol when dissociated from the metal catalyst, explaining the lack of enantioinduction observed when chiral ligands were used in dirhodium catalysis.[117]

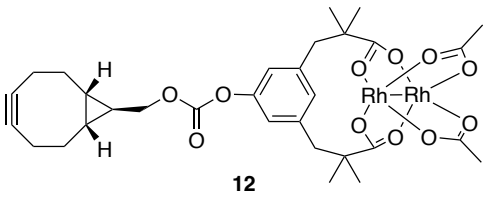
Understanding of the stepwise nature of this reaction inspired new approaches to overcome this problem. Chiral proton transfer catalysts were used in addition to achiral dirhodium, which enabled high levels of enantiocontrol over the proton transfer. Enantioselective N-H functionalization of anilines,[108] amides, and carbamates[118,119] was achieved with good-to-excellent enantioselectivities using chiral cocatalysts for proton transfer. The S-H functionalization of thiols was also found to proceed with high levels of enantioselectivity using this approach.[113] Recently, the asymmetric O-H functionalization of water using chiral phosphoric acids was reported. This study also provided a combined experimental/computational analysis of the reaction mechanism. This analysis showed that initial proton transfer and dissociation of the enol intermediate from rhodium is followed by phosphoric acid-facilitated proton transfer to produce the formal O-H insertion product with a lower barrier than any of the water-facilitated processes they examined. These results explain the origin of the high levels of enantioinduction observed in water and suggests that the ArM scaffold may be able to lower the barrier to proton transfer compared the same process in bulk water.

The differences between the concerted and stepwise mechanisms are critical for approaching these reactions within the context of ArMs based on dirhodium tetracarboxylate catalysts. In the case of concerted insertion reactions, the scaffold will need to control the positioning of the substrate while also excluding water. The chemoselectivity-determining step in this mechanism is the same as the enantiodetermining one, so high levels of control over the secondary coordination sphere during this step would result in good enantio- and chemoselectivity. The stepwise mechanism operative in N-H, S-H, and O-H functionalization requires an ArM with

a different set of catalytic properties, however. In these cases, achieving enantioselectivity would require binding of the prochiral enol intermediate to enable selective proton transfer. For this to be achieved with high levels of selectivity, the secondary and tertiary coordination spheres would need to be intricately controlled by the protein scaffold. These possibilities for scaffold-based control over the diverse chemistry catalyzed by dirhodium tetracarboxylate complexes makes this catalyst an excellent target for use in ArMs.

1.3.3 Dirhodium-POP ArMs

When the dirhodium-BCN cofactor **12** is linked to the POP-Z scaffold as discussed above, a hybrid catalyst capable of carbene transfer is formed. Initial work by Hao Yang and Poonam Srivastava found that the introduction of 4 alanine residues at the top of the β-propeller domain assisted in bioconjugation of the cofactor.[105] These mutations along with azidophenylalanine at position 477 were the only mutations in the initial ArM investigated, named ZA_4. This hybrid catalyst was found to catalyze the cyclopropanation of styrene with modest yields (25%) and enantioselectivities (38%) under optimized conditions. This was the first example of scaffold-controlled carbene insertion in the ArM literature and, as discussed above, enables the use of evolution to improve the scaffold.



12

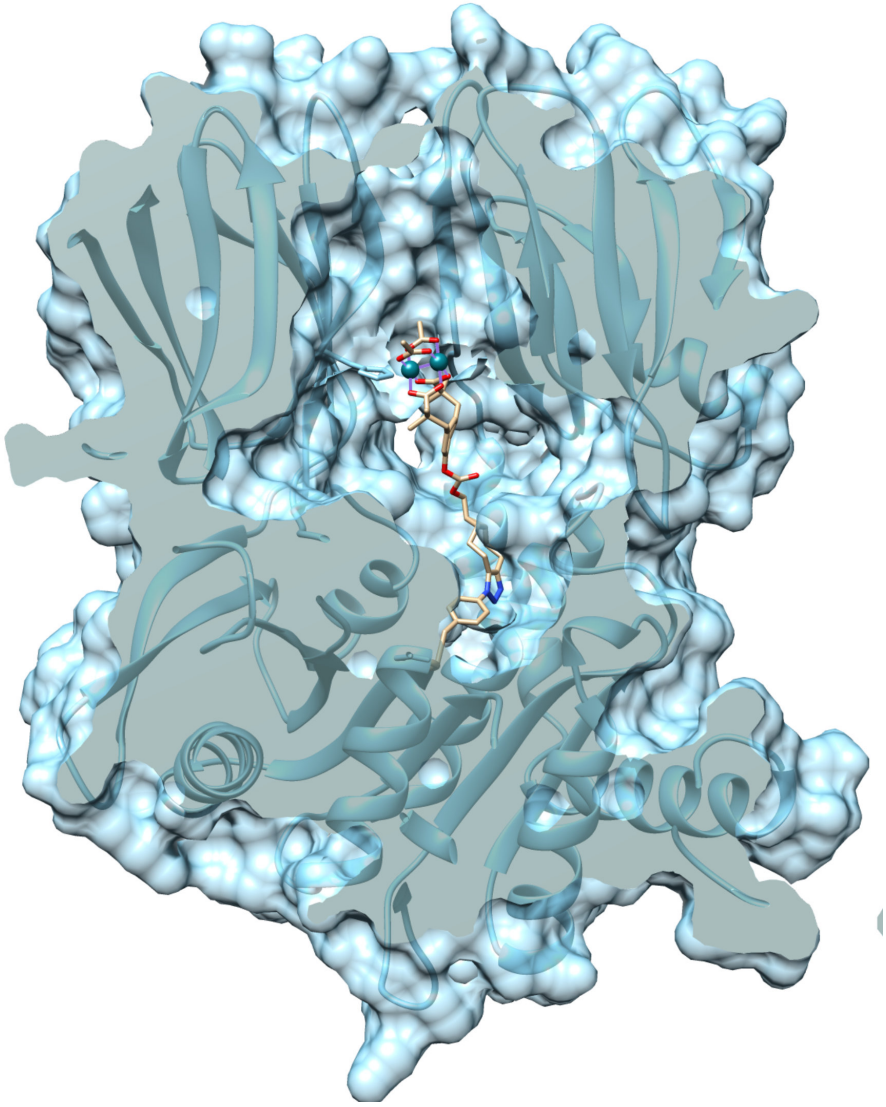


Figure 1.8 A model of the POP variant 5-G with cofactor bound. A Rh-His326 bond has been modelled into the structure (see Section 3.2.5 for details)

Initial efforts to towards this end used a rational design approach to introduce targeted mutations to the protein scaffold. A model of the ArM was constructed by fusing a geometry optimized structure of the cofactor within the β-propeller domain of a homology model of the *Pfu* POP scaffold. This model indicated that there was considerable empty space around the cofactor in the interior cavity that would result in significant freedom for the cofactor to sample different positions during catalysis, i.e. poor localization. Additionally, that space would be filled with water, which could react to give the O-H functionalization side product. The model was used to predict residues near the cofactor location that could influence the primary and secondary coordination spheres and a series of targeted mutations were made with two goals in mind: improving the localization of the dirhodium cofactor and increasing the hydrophobicity of the secondary coordination sphere to exclude water from the active site.

To improve localization of the cofactor, Lewis basic residues such as methionine or histidine were introduced in the β-propeller domain. Previous work from the Ball group in the development of dirhodium tetracarboxylate catalyst based on short polypeptides showed that incorporation of a histidine ligand bound the only open coordination site and inhibited the reaction.[121,122] It was hypothesized that by introducing residues capable of binding to dirhodium into the POP scaffold, one of the axial positions would anchor the cofactor in place while allowing the other rhodium to catalyze carbene transfer in a specific region of the POP active site. As the model suggested that residue 328 was well-positioned for rhodium binding, L328C, L328M, and L328H mutations were made. In each case, the enantioselectivity was significantly improved with a single mutation, making this an excellent example of the effect that localization can have on ArM catalysis. Though it was not understood at the time, the Rh-His interaction also altered the interdomain dynamics and improved encapsulation of the cofactor, discussed in depth in section

3.2.5. The best of these variants was found to be POP-ZA$_4$-L328H-**12** (61% yield, 85% e.e.), which was used as the parent for further evolution. The other goal of this design effort was the exclusion of water from the active site, which was performed by introducing phenylalanine mutations at positions near the cofactor. The best of the variants from this effort was ZA$_4$-F99-L328H-F594-**12** (HFF), which yielded 74% of the desired cyclopropane with 92% enantiomeric excess. The impact that these mutations had on catalysis indicate the scaffold is capable of controlling the primary coordination sphere through Rh-His localization as well as the secondary sphere, with the incorporation of phenylalanine mutations that result in improved chemo- and enantioselectivity. This work showed that rational design could be used to improve the POP scaffold for cyclopropanation. In cases where one does not have structural information about potential cofactor positioning, however, point mutations and large-scale expression of the scaffold is not an efficient approach. Directed evolution using high throughput screening is the only option in these cases. To demonstrate that the POP scaffold could be improved in this way, directed evolution using random mutagenesis was performed using this system (Section 2.1.1) and the resulting variants were used as the starting point for the work described in this dissertation.

1.4 Conclusion

Protein scaffolds can exhibit remarkable control over the activity and selectivity of natural cofactors through tuning of the primary and secondary coordination spheres. The desire to achieve similar control over reactions that don't exist in nature has inspired the development of ArMs. A number of methods exist for the development of these hybrid catalyst, but not all systems can achieve this vision of promiscuous scaffold-based control over any reaction. In the ArM case

studies discussed above, three characteristics were observed in systems that approached this notion of promiscuity.

1) Localization of the metal center is necessary to maintain a continuous chemical environment during catalysis. Cofactors with significant rotational freedom or those that protrude into solution frequently cannot be controlled by the scaffold. This challenge takes different forms depending on the bioconjugation method. In general, inherent flexibility of the linkers makes most challenging for covalent bioconjugation methods,[64] though primary coordination sphere linking to catalyst was found to provide a resolution this case.[74,105] Localization can also be an issue for supramolecular anchoring methods, though optimization of the linker and ligand frequently resolve this issue. [84] The direct ligation of the metal to the protein structure minimizes this issue in dative and metal replacement anchoring.

2) Encapsulation of the metal cofactor is necessary to enable the scaffold to affect the entire secondary coordination sphere. While this may not be prohibitive to selective ArM catalysis in reactions where selectivity is controlled through interactions with the metal, in cases without these interactions the scaffold may not be able to appropriately position the substrates in space. Though issues have been reported where the metal cofactor was too large to fit in the active site, this characteristic is more frequently inherent to the scaffold from the start. Efforts to engineer this quality into scaffolds have been reported, which provide a novel way to approach this challenge though continued progress will be necessary to achieve the control over catalysis that is desired.

3) The final characteristic results from the use of the protein scaffold: evolvability. In ArMs with good localization and encapsulation, the scaffold can likely be evolved to improve the chemical environment around the metal for catalysis. However, in cases without robust localization and encapsulation, there may be limitations in the system that cannot be overcome

through directed evolution. It should be noted, however, that directed evolution absolutely can result in improvements in localization and encapsulation.

Dirhodium ArMs based on the POP scaffold have shown promise in each of these respects. In this dissertation, I describe my doctoral research on improving these prolyl oliogopeptidase-based dirhodium artificial metalloenzymes for carbene transfer chemistry. Chapter 2 explores the scope of reactions catalyzed by early iterations of the ArM. Chapters 3-5 describe the process of engineering improved scaffolds for diazo coupling, N-H functionalization, and Si-H functionalization, respectively.

1.5 References

[1] E. V. Anslyn, D. A. Dougherty, *Modern Physical Organic Chemistry*, University Science, 2005.

[2] H. A. Wolfgang, C. Boy, *Angewandte Chemie Int Ed Engl* 1997, *36*, 1048–1067.

[3] J. W. Erisman, M. A. Sutton, J. Galloway, Z. Klimont, W. Winiwarter, *Nat Geosci* 2008, *1*, 636–639.

[4] H. H. Brintzinger, D. Fischer, R. Mülhaupt, B. Rieger, R. M. Waymouth, *Angewandte Chemie Int Ed Engl* 1995, *34*, 1143–1170.

[5] S. Robert, *Angewandte Chemie Int Ed* 2003, *42*, 2004–2008.

[6] T. Cernak, K. D. Dykstra, S. Tyagarajan, P. Vachal, S. W. Krska, *Chem Soc Rev* 2016, *45*, 546–576.

[7] E. M. McGarrigle, D. G. Gilheany, *Chem Rev* 2005, *105*, 1563–1602.

[8] J.-H. Xie, S.-F. Zhu, Q.-L. Zhou, *Chem Rev* 2011, *111*, 1713–1760.

[9] Norio. Miyaura, Akira. Suzuki, *Chem Rev* 1995, *95*, 2457–2483.

[10] T. M. Trnka, R. H. Grubbs, *Acc. Chem. Res.* 2001, *34*, 18–29.

[11] H. M. L. Davies, D. Morton, *J Org Chem* 2016, *81*, 343–350.

[12] H. J. Davis, R. J. Phipps, *Chem Sci* 2017, *8*, 864–877.

[13] J. Trouvé, R. Gramage-Doria, *Chem Soc Rev* 2021, *50*, 3565–3584.

[14] A. J. Tyler, S. G. Andrew, M. E. Anita, *Eur J Org Chem* 2014, *2014*, 2633–2646.

[15] G. J. David, S. D. Benjamin, T. Dean F., *Chem Rev* 2008, *108*, 3351–3378.

[16] S. M. Dennis J., van K. Gerard, K. G. M. Robertus J., *Chem - European J* 2011, *17*, 42–57.

[17] C. A. Sandoval, T. Ohkuma, K. Muñiz, R. Noyori, *J Am Chem Soc* 2003, *125*, 13490–13503.

[18] W. Zhang, J. L. Loebach, S. R. Wilson, E. N. Jacobsen, *J Am Chem Soc* 1990, *112*, 2801–2803.

[19] H. M. L. Davies, K. Liao, *Nat Rev Chem* 2019, *3*, 347–360.

[20] H. M. L. Davies, J. R. Manning, *Nature* 2008, *451*, 417–24.

[21] J. Wencel-Delord, F. Glorius, *Nat Chem* 2013, *5*, 369–375.

[22] J. Wencel-Delord, T. Dröge, F. Liu, F. Glorius, *Chem Soc Rev* 2011, *40*, 4740–4761.

[23] W. R. Gutekunst, P. S. Baran, *Chem Soc Rev* 2011, *40*, 1976–1991.

[24] T. Brückl, R. D. Baxter, Y. Ishihara, P. S. Baran, *Accounts Chem Res* 2012, *45*, 826–839.

[25] K. M. Engle, T. S. Mei, M. Wasa, J. Q. Yu, *Accounts Chem Res* 2012, *45*, 788–802.

[26] J. F. Hartwig, M. A. Larsen, *Acs Central Sci* 2016, *2*, 281–292.

[27] J. F. Hartwig, *Accounts Chem Res* 2012, *45*, 864–873.

[28] K. Liao, S. Negretti, D. G. Musaev, J. Bacsa, H. M. L. Davies, *Nature* 2016, *533*, 230–234.

[29] W. Liu, Z. Ren, A. T. Bosse, K. Liao, E. L. Goldstein, J. Bacsa, D. G. Musaev, B. M. Stoltz, H. M. L. Davies, *J Am Chem Soc* 2018, *140*, jacs.8b07534.

[30] J. C. Lewis, P. S. Coelho, F. H. Arnold, *Chem Soc Rev* 2010, *40*, 2003–2021.

[31] K. M. Lancaster, in *Molecular Electronic Structures of Transition Metal Complexes I. Structure and Bonding* (Eds.: M.D.D. P., D. J., D. P.), 2011.

[32] U. B. Vlada, G. Marco, *Trends Biotechnol* 2012, *30*, 26–36.

[33] R. Kaspera, R. Croteau, *Phytochem Rev* 2006, *5*, 433–444.

[34] X. Ren, J. A. Yorke, E. Taylor, T. Zhang, W. Zhou, L. L. Wong, *Chem - European J* 2015, *21*, 15039–15047.

[35] C. Jung-Min, H. Sang-Soo, K. Hak-Sung, *Biotechnol Adv* 2015, *33*, 1443–1454.

[36] E. L. Bell, W. Finnigan, S. P. France, A. P. Green, M. A. Hayes, L. J. Hepworth, S. L. Lovelock, H. Niikura, S. Osuna, E. Romero, K. S. Ryan, N. J. Turner, S. L. Flitsch, *Nat Rev Methods Primers* 2021, *1*, 46.

[37] W. Roland, *Curr Opin Biotech* 2010, *21*, 713–724.

[38] P. Ophry, I. Masayori, *Mol Biotechnol* 1999, *12*, 25–34.

[39] K. Jashandeep, K. Arbind, K. Jagdeep, *Int J Biol Macromol* 2018, *106*, 803–822.

[40] B. T. U., H. W. G., K. J. R., L. S., M. C. J., R. K., *Nature* 2012, *485*, 185–194.

[41] A. H. Frances, V. A. Alexander, *Curr Opin Chem Biol* 1999, *3*, 54–59.

[42] P. A. Romero, F. H. Arnold, *Nat Rev Mol Cell Bio* 2009, *10*, 866–76.

[43] J. D. Bloom, F. H. Arnold, *Proc National Acad Sci* 2009, *106*, 9995–10000.

[44] M. T. Reetz, *Angewandte Chemie Int Ed* 2011, *50*, 138–174.

[45] F. H. Arnold, *Q Rev Biophys* 2015, *48*, 404–410.

[46] Y. Yang, F. H. Arnold, *Accounts Chem Res* 2021, *54*, 1209–1225.

[47] T. Heinisch, T. R. Ward, *Eur J Inorg Chem* 2015, *2015*, 3406–3418.

[48] D. M. Upp, J. C. Lewis, *Curr Opin Chem Biol* 2017, *37*, 48–55.

[49] K. Chen, F. H. Arnold, *Nat Catal* 2020, *3*, 203–213.

[50] E. W. Svastits, J. H. Dawson, R. Breslow, S. H. Gellman, *J Am Chem Soc* 1985, *107*, 6427–6428.

[51] R. Breslow, *Accounts Chem Res* 1995, *28*, 146–153.

[52] O. F. Brandenberg, R. Fasan, F. H. Arnold, *Curr Opin Biotech* 2017, *38*, in press.

[53] R. Singh, J. N. Kolev, P. A. Sutera, R. Fasan, *Acs Catal* 2015, *5*, 1685–1691.

[54] S. B. J. Kan, R. D. Lewis, K. Chen, F. H. Arnold, *Science* 2016, *354*, 1048–1051.

[55] O. F. Brandenberg, K. Chen, F. H. Arnold, *J Am Chem Soc* 2019, *141*, 8989–8995.

[56] A. M. Knight, S. B. J. Kan, R. D. Lewis, O. F. Brandenberg, K. Chen, F. H. Arnold, *Acs Central Sci* 2018, *4*, 372–377.

[57] T. K. Hyster, F. H. Arnold, *Israel J Chem* 2015, *55*, 14–20.

[58] T. Hayashi, M. Tinzl, T. Mori, U. Krengel, J. Proppe, J. Soetbeer, D. Klose, G. Jeschke, M. Reiher, D. Hilvert, *Nat Catal* 2018, *1*, 578–584.

[59] M. L. Matthews, C. S. Neumann, L. A. Miles, T. L. Grove, S. J. Booker, C. Krebs, C. T. Walsh, J. M. Bollinger, *Proc National Acad Sci* 2009, *106*, 17723–17728.

[60] M. L. Matthews, W. Chang, A. P. Layne, L. A. Miles, C. Krebs, J. M. Bollinger, *Nat Chem Biol* 2014, *10*, 209–15.

[61] F. Schwizer, Y. Okamoto, T. Heinisch, Y. Gu, M. M. Pellizzoni, V. Lebrun, R. Reuter, V. Köhler, J. C. Lewis, T. R. Ward, *Chem Rev* 2018, *118*, 142–231.

[62] M. E. Wilson, G. M. Whitesides, *J Am Chem Soc* 1978, *100*, 306–307.

[63] J. C. Lewis, *Acs Catal* 2013, *3*, 2954–2975.

[64] H. Yang, P. Srivastava, C. Zhang, J. C. Lewis, *Chembiochem* 2014, *15*, 223–227.

[65] A. D. Liang, J. Serrano-Plana, R. L. Peterson, T. R. Ward, *Accounts Chem Res* 2019, *52*, 585–595.

[66] Y. Sano, A. Onoda, T. Hayashi, *Chem Commun* 2011, *47*, 8229–8231.

[67] S. N. Natoli, J. F. Hartwig, *Accounts Chem Res* 2019, *52*, 326–335.

[68] K. Oohora, A. Onoda, T. Hayashi, *Accounts Chem Res* 2019, *52*, 945–954.

[69] A. Onoda, K. Fukumoto, M. Arlt, M. Bocola, U. Schwaneberg, T. Hayashi, *Chem Commun* 2012, *48*, 9756–9758.

[70] K. Fukumoto, A. Onoda, E. Mizohata, M. Bocola, T. Inoue, U. Schwaneberg, T. Hayashi, *Chemcatchem* 2014, *6*, 1229–1235.

[71] S. Kato, A. Onoda, A. R. Grimm, K. Tachikawa, U. Schwaneberg, T. Hayashi, *Inorg Chem* 2020, *59*, 14457–14463.

[72] S. Kato, A. Onoda, N. Taniguchi, U. Schwaneberg, T. Hayashi, *Chembiochem* 2021, *22*, 679–685.

[73] T. K. Hyster, L. Knörr, T. R. Ward, T. Rovis, *Science* 2012, *338*, 500–503.

[74] H. Yang, A. M. Swartz, H. J. Park, P. Srivastava, K. Ellis-Guardiola, D. M. Upp, G. Lee, K. Belsare, Y. Gu, C. Zhang, R. E. Moellering, J. C. Lewis, *Nat Chem* 2018, *10*, 318–324.

[75] A. Onoda, Y. Kihara, K. Fukumoto, Y. Sano, T. Hayashi, *Acs Catal* 2014, *4*, 2645–2648.

[76] D. F. Sauer, T. Himiyama, K. Tachikawa, K. Fukumoto, A. Onoda, E. Mizohata, T. Inoue, M. Bocola, U. Schwaneberg, T. Hayashi, J. Okuda, *Acs Catal* 2015, *5*, 7519–7522.

[77] T. Himiyama, D. F. Sauer, A. Onoda, T. P. Spaniol, J. Okuda, T. Hayashi, *J Inorg Biochem* 2016, *158*, 55–61.

[78] C. Zhang, P. Srivastava, K. Ellis-Guardiola, J. C. Lewis, *Tetrahedron* 2014, *70*, 4245–4249.

[79] C. Letondor, N. Humbert, T. R. Ward, *P Natl Acad Sci Usa* 2005, *102*, 4683–4687.

[80] C. Letondor, A. Pordea, N. Humbert, A. Ivanova, S. Mazurek, M. Novic, T. R. Ward, *J Am Chem Soc* 2006, *128*, 8320–8328.

[81] T. R. Ward, *Accounts Chem Res* 2011, *44*, 47–57.

[82] H. J. Davis, T. R. Ward, *Acs Central Sci* 2019, *5*, 1120–1136.

[83] M. Jeschek, R. Reuter, T. Heinisch, C. Trindler, J. Klehr, S. Panke, T. R. Ward, *Nature* 2016, *537*, 661–665.

[84] J. Serrano-Plana, C. Rumo, J. G. Rebelein, R. L. Peterson, M. Barnet, T. R. Ward, *J Am Chem Soc* 2020, *142*, 10617–10623.

[85] T. J. Collins, A. D. Ryabov, *Chem Rev* 2017, *117*, 9140–9162.

[86] J. Zhao, D. G. Bachmann, M. Lenz, D. G. Gillingham, T. R. Ward, *Catal Sci Technol* 2018, *8*, 2294–2298.

[87] M. M. Pellizzoni, F. Schwizer, C. W. Wood, V. Sabatino, Y. Cotelle, S. Matile, D. N. Woolfson, T. R. Ward, *Acs Catal* 2018, *8*, 1476–1484.

[88] F. Christoffel, N. V. Igareta, M. M. Pellizzoni, L. Tiessler-Sala, B. Lozhkin, D. C. Spiess, A. Lledós, J.-D. Maréchal, R. L. Peterson, T. R. Ward, *Nat Catal* 2021, *4*, 643–653.

[89] M. Bordeaux, V. Tyagi, R. Fasan, *Angewandte Chemie Int Ed* 2015, *54*, 1744–1748.

[90] J. A. McIntosh, P. S. Coelho, C. C. Farwell, Z. J. Wang, J. C. Lewis, T. R. Brown, F. H. Arnold, *Angewandte Chemie Int Ed* 2013, *52*, 9309–9312.

[91] T. K. Hyster, C. C. Farwell, A. R. Buller, J. A. McIntosh, F. H. Arnold, *J Am Chem Soc* 2014, *136*, 15505–15508.

[92] P. Dydio, H. M. Key, A. Nazarenko, J. Y.-E. Rha, V. Seyedkazemi, D. S. Clark, J. F. Hartwig, *Science* 2016, *354*, 102–106.

[93] H. Takashi, H. Yutaka, A. Tsutomu, M. Tadashi, H. Yoshio, K. Susumu, O. Hisanobu, *J Am Chem Soc* 1999, *121*, 7747–7750.

[94] K. Oohora, Y. Kihira, E. Mizohata, T. Inoue, T. Hayashi, *J Am Chem Soc* 2013, *135*, 17282–17285.

[95] K. Oohora, H. Meichin, Y. Kihira, H. Sugimoto, Y. Shiro, T. Hayashi, *J Am Chem Soc* 2017, *139*, 18460–18463.

[96] M. W. Wolf, D. A. Vargas, N. Lehnert, *Inorg Chem* 2017, *56*, 5623–5635.

[97] K. Oohora, H. Meichin, L. Zhao, M. W. Wolf, A. Nakayama, J. Hasegawa, N. Lehnert, T. Hayashi, *J Am Chem Soc* 2017, *139*, 17265–17268.

[98] H. M. Key, P. Dydio, D. S. Clark, J. F. Hartwig, *Nature* 2016, *534*, 534–537.

[99] J. Huang, Z. Liu, B. J. Bloomer, D. S. Clark, A. Mukhopadhyay, J. D. Keasling, J. F. Hartwig, *Nat Chem* 2021, 1–6.

[100] M. N. Harris, J. D. Madura, L.-J. Ming, V. J. Harwood, *J Biol Chem* 2001, *276*, 19310–19317.

[101] K. Ellis-Guardiola, H. Rui, R. L. Beckner, P. Srivastava, N. Sukumar, B. Roux, J. C. Lewis, *Biochemistry-us* 2019, *58*, 1616–1626.

[102] J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *J Am Chem Soc* 2002, *124*, 9026–9027.

[103] N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J Am Chem Soc* 2004, *126*, 15046–15047.

[104] Y. Gu, K. Ellis-Guardiola, P. Srivastava, J. C. Lewis, *Chembiochem* 2015, *60637*, n/a-n/a.

[105] P. Srivastava, H. Yang, K. Ellis-Guardiola, J. C. Lewis, *Nat Commun* 2015, *6*, 7789.

[106] J. R. Denton, H. M. L. Davies, *Org Lett* 2009, *11*, 787–790.

[107] B. Xu, S.-F. Zhu, Z.-C. Zhang, Z.-X. Yu, Y. Ma, Q.-L. Zhou, *Chem Sci* 2014, *5*, 1442.

[108] H. Saito, T. Uchiyama, M. Miyake, M. Anada, S. Hashimoto, T. Takabatake, S. Miyairi, *Heterocycles* 2010, *81*, 1149.

[109] H. M. L. Davies, T. Hansen, M. R. Churchill, *J Am Chem Soc* 2000, *122*, 3063–3070.

[110] B. Xu, S. F. Zhu, X. D. Zuo, Z. C. Zhang, Q. L. Zhou, *Angewandte Chemie Int Ed* 2014, *53*, 3913–3916.

[111] J. Hansen, J. Autschbach, H. M. L. Davies, *J Org Chem* 2009, *74*, 6555–6563.

[112] D. Gillingham, N. Fei, *Chem Soc Rev* 2013, *42*, 4918–4931.

[113] Y.-Y. Ren, S.-F. Zhu, Q.-L. Zhou, *Org Biomol Chem* 2018, *16*, 3087–3094.

[114] X.-C. Wang, X.-S. Song, L.-P. Guo, D. Qu, Z.-Z. Xie, F. Verpoort, J. Cao, *Organometallics* 2014, *33*, 4042–4050.

[115] Y. Zhang, Y. Yao, L. He, Y. Liu, L. Shi, *Adv Synthesis Amp Catal* 2017, *359*, 2754–2761.

[116] Y. Liang, H. Zhou, Z.-X. Yu, *J Am Chem Soc* 2009, *131*, 17783–17785.

[117] R. T. Buck, C. J. Moody*, A. G. Pepper, *ARKIVOC* 2002, 16–33.

[118] B. Xu, S. Zhu, X. Xie, J. Shen, Q. Zhou, *Angewandte Chemie Int Ed* 2011, *50*, 11483–11486.

[119] B. Xu, S. Zhu, X. Zuo, Z. Zhang, Q. Zhou, *Angewandte Chemie Int Ed* 2014, *53*, 3913–3916.

[120] Y. Li, Y.-T. Zhao, T. Zhou, M.-Q. Chen, Y.-P. Li, M.-Y. Huang, Z.-C. Xu, S.-F. Zhu, Q.-L. Zhou, *J Am Chem Soc* 2020, *142*, 10557–10566.

[121] B. V. Popp, Z. T. Ball, *Chem Sci* 2011, *2*, 690–695.

[122] Z. T. Ball, *Accounts Chem Res* 2013, *46*, 560–570.

Chapter 2: Exploration of the Substrate Scope of Dirhodium Artificial Metalloenzymes

2.1 Introduction

2.1.1 Random mutagenesis for improved POP-Rh_2 ArM catalysis

As discussed in depth in Chapter 1, dirhodium-ArMs based on the prolyl oligopeptidase[1,2] scaffold were established based on covalent bioconjugation of a cyclooctyne-linked dirhodium tetracarboxylate cofactor to a genetically incorporated azidophenylalanine residue[3] (abbreviated as Z). The initial scaffold with the Z mutation at position 477 and four alanine mutations found to aid in bioconjugation, referred to as ZA_4, catalyzed the cyclopropanation of styrenes with 38% e.e. and 25% yield.[2] Initial efforts were made to rationally improve the scaffold by introducing rhodium-ligating residues, such as histidine and methionine, and increasing the hydrophobicity of the interior cavity by adding phenylalanine residues. The best variant from this work, POP-ZA_4-H328-F99-F594, had substantially improved enantioselectivity (92%) and yield (74%) compared to the parent and demonstrated that mutations to the scaffold were capable of significantly controlling the primary and secondary coordination sphere around the dirhodium cofactor during catalysis.[2]

While rational design was shown to be an effective method to improve the scaffold in our case and many others, proteins in nature evolve via random mutations. Artificial metalloenzymes had not been evolved using a random approach, so a directed evolution campaign aimed at using random mutagenesis to achieve the similar improvements on cyclopropanation was performed by Hao Yang.[4] Starting from the same parent as was used in the rational design project, 0-ZA_4, 3 rounds of directed evolution using error-prone PCR for gene diversification resulted in the variant

3-VRVH, that had improved enantioselectivity (93%) and yield (75%) (Figure 2.1a). Interestingly, hits were selected based on improved enantioselectivity, but the rate, yield and **3**/**4** ratio of each hit were found to increase as well (Figure 2.1b). We hypothesize that the improved localization of the dirhodium center enabled more consistent scaffold-based control over substrate positioning, for improved enantioselectivity, and water exclusion, for improved chemoselectivity.

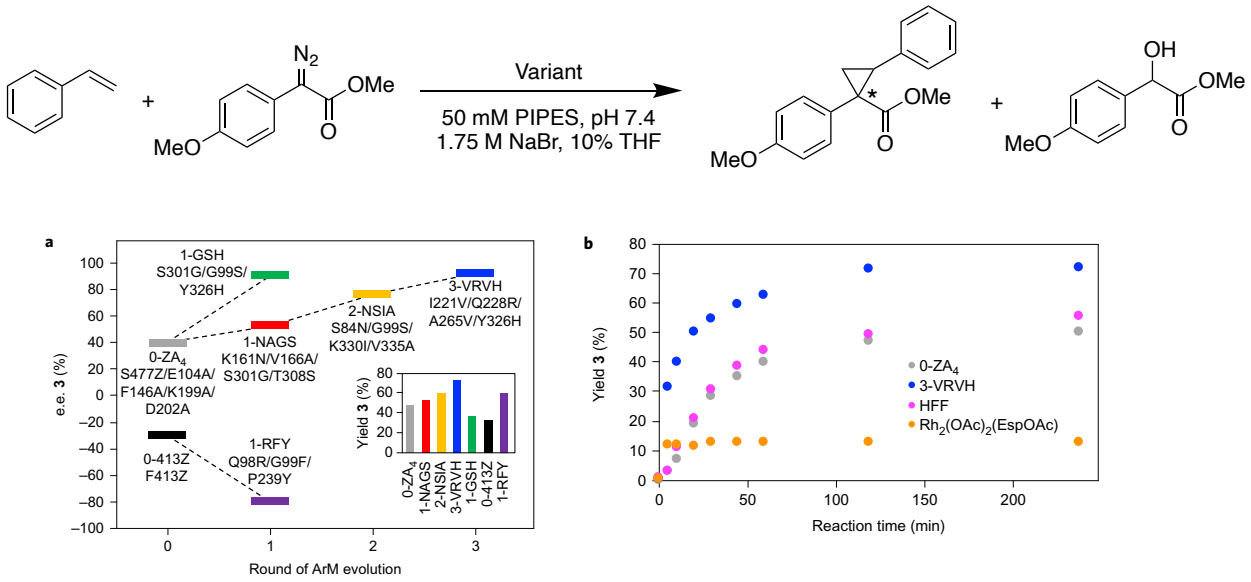


Figure 2.1 Directed evolution of the POP-ArM for cyclopropanation.

Mutations were made throughout the scaffold in each of the hits, but the three that were found to impact selectivity (S301G in 1-NAGS, G99S in 2-NSIA, Y326H in 3-VRVH) were located in the interior of the β-propellor domain. The identification of the Y326H mutation in 3-VRVH was found to increase the enantioselectivity from 77% in 2-NSIA to 92%. As was hypothesized in the rational design of the POP scaffold for dirhodium catalysis, the introduction of the active site histidine can bind to one of the axial positions of the dirhodium cofactor, anchoring it in place for a more consistent active site environment.

The discovery of these mutants was possible by using a high-throughput screening method, developed by Hao Yang, for testing cyclopropanation performance of ArM variants in 96-well plates.[4] This method involved expression of the scaffolds in 1.5 mL of media, followed by lysis of the cells and addition of cofactor for bioconjugation in lysate. Because there was a significant excess of cofactor, 50 μL of azide-agarose resin had to be added to each well to scavenge the non-bioconjugated cofactor before catalysis could be performed. While this high-throughput method enabled reliable detection of yield and enantioselectivity of the cyclopropanation reaction, the ~5 mL of azide-agarose resin required per 96-well plate was not ideal due to its high cost.

An alternate strategy for the elimination of excess cofactor via ArM immobilization was developed by Hyun June Park in the Lewis group, as shown in Figure 2.2. In this method, the lysate is loaded into a filter plate, where bioconjugation of the scaffold is once again performed in lysate using excess cofactor. Ni-NTA agarose is then added to each well and non-bound cofactor is removed using repeated washing of the resin. To demonstrate that this new method could be used for reliable detection of ArM activity in high-throughput, the cyclopropanation activity of a POP-ArM scaffold with the AzF anchor switched to position 413 from 477. Interestingly, this variant catalyzes cyclopropanation to form the opposite enantiomer product with modest enantioselectivity (0-413Z, Figure 2.1a). A single round of evolution using combinatorial codon mutagenesis resulted, resulting in variant 1-RFY that achieved -80% e.e.

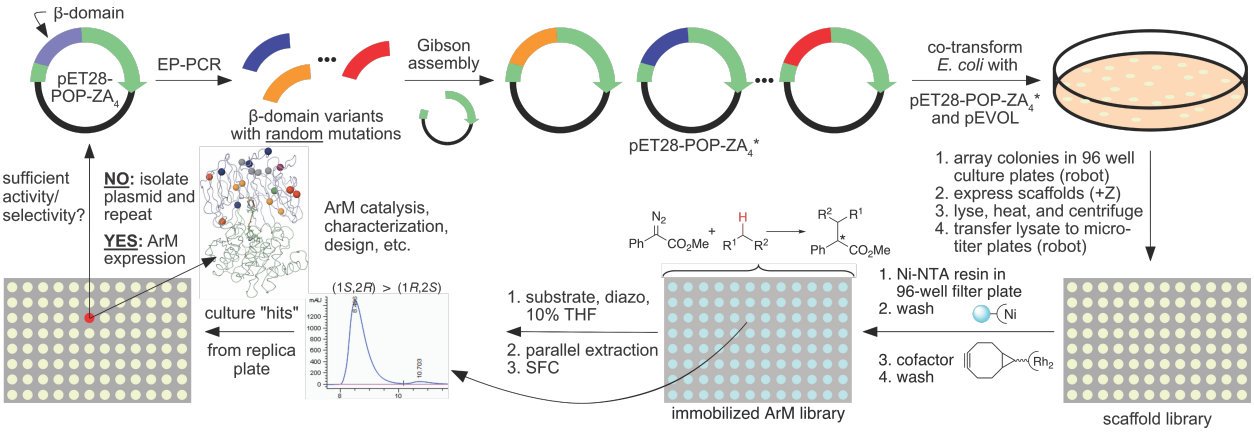


Figure 2.2 Directed evolution of dirhodium ArMs using Ni-NTA immobilization.

2.1.2 Dirhodium carbenoid insertion mechanisms

As discussed in depth in Chapter 1.3.2, the mechanism by which the dirhodium carbenoid inserts into X-H bonds depends on the nucleophile.[5–7] In the case of nucleophiles without a lone pair of electrons, such as C-H insertion, Si-H insertion, or olefin cyclopropanation,[5] the reaction proceeds through an asynchronous concerted transition state where the bond breaking and forming occurs in a single, multi-atom step (Figure 1.6). Where there is a lone pair on the nucleophile, as is the case in N-H,[6] S-H,[8] and O-H[7] functionalization, a different mechanism is operative. A number of computational studies have suggested that the nucleophile attacks the carbenoid and dissociates from the dirhodium as the corresponding planar enolate. Subsequent proton transfer to the α-carbon is the enantio-determining step, which presents a unique challenge for controlling the enantioselectivity since chiral ligands on the dirhodium catalyst would not be expected to influence this step. To overcome this issue, chiral proton transfer catalysts that bind to the enolate intermediate have been utilized to achieve highly selective transformations.[8–12]

2.1.3 ArM-catalyzed carbenoid insertion reactions

The improvements to enantioselectivity and **3**/**4** ratio achieved in olefin cyclopropanation using both rational design and random mutagenesis suggest that the ArM-scaffolds of the final variants can control the local environment around the cofactor better than parent, 0-ZA_4. The concerted mechanism that this model reaction proceeds through requires positioning of the olefin substrate and exclusion of water from the area surrounding the carbenoid. We hypothesized that these types of improvements to the scaffold would be generally beneficial to other carbene transfer reactions. To this end, I explored a variety of carbene transfer reactions, including nucleophiles that undergo stepwise insertion, such as N-H, S-H and O-H (on organic substrates), as well as concerted insertion, operative in Si-H and C-H functionalization.

In water, dirhodium tetracarboxylate catalysts nearly exclusively catalyze the O-H functionalization of solvent,[2] so a dirhodium-ArM capable of forming these bonds in a selective fashion would require a scaffold capable of controlling the primary and secondary coordination sphere of the metal catalyst, through substrate positioning and water exclusion. This would clearly demonstrate the unique properties of POP as a promiscuous scaffold for ArMs and further show the control that protein scaffolds can have over small molecule catalysts. In addition, each of these reaction types are interesting from a synthetic perspective, as the formation of C-X bonds in a selective manner are of great value in the synthesis of complex chiral molecules. My contributions to this work were in the exploration of these scaffolds for the diverse reactivity explored in this chapter and were intended to demonstrate the improved catalytic qualities of variants found in the directed evolution of cyclopropanation and as starting points for future evolution work (Chapters 4 and 5).

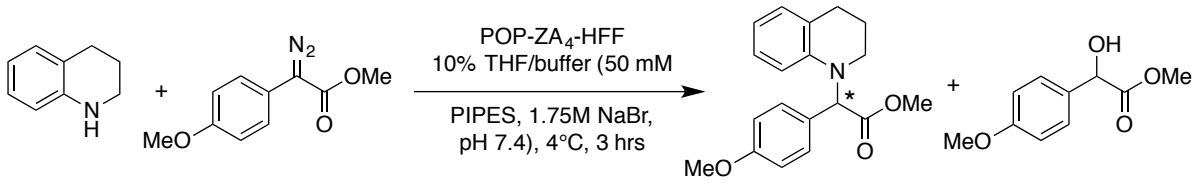
2.2 Results and Discussion

2.2.1 Initial Exploration of N-H Functionalization

The formation of C-N bonds is of great synthetic interest in organic chemistry due to their ubiquity in bioactive molecules including pharmaceuticals and natural products.[13,14] Numerous catalysts for transition metal-catalyzed carbenoid insertion into N–H bonds have been developed, and dirhodium (II) tetracarboxylate complexes in combination with chiral H-bond catalysts have received significant attention in this regard.[8,9,11,15] Excellent yields and enantioselectivity have been reported for anilines[11] and amides[9,15] using non-chiral dirhodium catalysts in combination with chiral phosphoric acids or chiral amine H-bonding catalysts. We therefore envisioned using dirhodium-POP ArMs as dual-role catalysts, both catalyzing the addition of the N-H substrate to the carbenoid and binding the planar intermediate to enable enantiospecific proton transfer.

An initial evaluation of N-H functionalization activity catalyzed by dirhodium-ArMs was performed by a previous graduate student in the Lewis group, Chen Zhang. He found that a scaffold from the rational design project, ZA_4-HFF, catalyzed the N-H functionalization of a number electron-rich secondary anilines with modest selectivity (~30% e.e.) and decent yields (50-70%). The best of these substrates was found to be 1,2,3,4-tetrahydroquinoline (THQ), which I selected for my investigation into the activity and selectivity of the cyclopropanation lineage for N-H functionalization.

Table 2.1 Evaluation of the cyclopropanation lineage for the N-H functionalization of THQ.



Entry	Variant	Lineage	Yield (%)	e.e. (%)
1	0-ZA4	Z477	55	8
2	1-NAGS	Z477	77	6
3	3-VRVH	Z477	73	40
4	1-RFY	Z413	87	-12

After a brief optimization of the reaction conditions, the yields and selectivity for the desired N-H functionalized product were found to increase over the lineage, as shown in Table 2.1. The parent 0-ZA_4 catalyzed the reaction with a modest yield but very low enantioselectivity (8%) while the final variant, 3-VRVH, was considerably improved, forming the desired product in 73% yield with 40% e.e.. The improvements observed here are consistent with both improved water exclusion, leading to a higher yield, and improved binding of the prochiral intermediate, resulting in higher enantioselectivity. These data were included in the publication detailing the use of random mutagenesis for directed evolution of ArMs.

In addition to the N-H functionalization of 1,2,3,4-tetrahydroquinoline using (4-methoxyphenyl)diazoacetate, a number of electronically distinct diazos were also examined as substrates for N-H functionalization reactions. Compared with the previously examined donor-acceptor diazos, containing phenyl and ester substituents adjacent to the diazo carbon, the carbenoids formed from acceptor-only and acceptor-acceptor diazos are much more electrophilic, and therefore should react more quickly when the carbenoid is formed. Acceptor-acceptor diazos, such as ethyl acetodiazoacetonate, were used in the N-H functionalization of THQ. With both 3-

VRVH and HFF, however, no desired product was observed. If the reaction was occurring too rapidly for the N-H substrate to approach the rhodium carbenoid, we hypothesized that a substrate containing an intramolecular N-H bond might localize the desired functionality and allow for reaction. An acceptor-acceptor diazo with this functionality was synthesized and used in an ArM catalyzed N-H functionalization reaction. Once again, no desired product was observed. There could be many reasons why this was found to be the case. Active site constraints could preclude the bulky substrate from adopting the necessary orientation to react with the aniline, or the carbenoid may be too reactive for any type of chemoselectivity over water to be achieved. Tuning down the electrophilicity was therefore attempted by using ethyl diazoacetate as a substrate. Once again, however, very little product was observed in the reaction with THQ. Taken together, these results suggest that the lower carbenoid reactivity of donor-acceptor carbenoids is beneficial or even necessary for achieving N-H functionalization. When these are used, modest enantioselectivity with good yields are achieved, making this and excellent starting point for future evolution. Potential binding positions that could enable the enantioselective protonation along with the directed evolution of this reaction are described in Chapter 4.

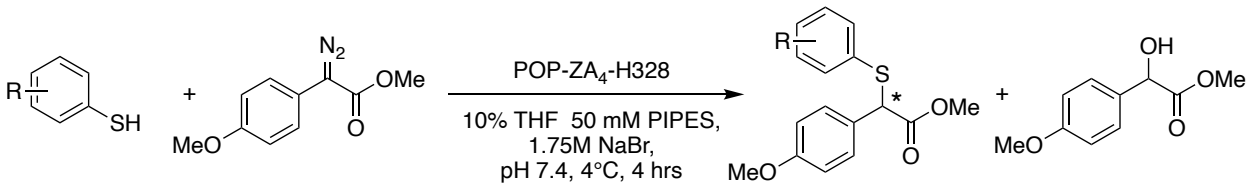
2.2.2 Initial Exploration of S-H Functionalization

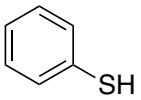
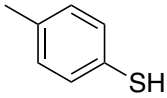
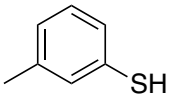
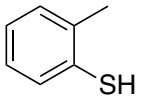
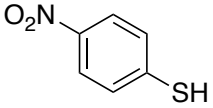
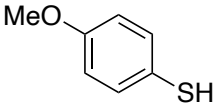
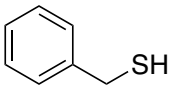
The stepwise mechanism controlled by the protein scaffold in the N-H functionalization reactions suggested that other reactions where selectivity is achieved through a chiral proton transfer, such as the insertion into S-H bonds may be similarly affected. While chiral S-C bonds are less common than the N-C bonds previously discussed, there are numerous examples in antibiotics and other bioactive molecules that make this reaction of potential interest.[16–18] Chiral methods to form these bonds have been developed using dirhodium tetracarboxylate catalysts in

combination with chiral phosphoric acids,[8,17] achieving high yields and excellent enantioselectivity, which suggested that dirhodium POP-ArMs may be able to control the reaction in a similar way to N-H functionalization.

While work had been done to determine the substrate scope of N-H functionalization before my contributions began, this was not the case for thiol nucleophiles. Encouragingly, initial experiments found that dirhodium ArMs catalyzed the formal S-H functionalization of benzenethiol with good yield, though low enantioselectivity (Table 2.2). The formation of significant amounts of product suggests that the nucleophilicity of relatively acidic benzenethiol, which has a pK_a of 6.6 in water and will thus be primarily deprotonated under the reaction conditions at pH 7.4, is sufficient to achieve the desired reaction. In order to determine the ideal thiol and diazo pair to demonstrate improvement over the cyclopropanation lineage, a number of substrates were explored in detail using a scaffold containing only the H328 mutation from the rational design project. Methylation of benzenethiol at the *ortho*-, *meta*-, or *para*-positions did not result in a significant change in activity or selectivity, but altering the electronics of the aryl ring was found to do so. Interestingly, an electron-withdrawing nitro group at the *para*-position decreased product formation substantially, presumably due to decreased nucleophilicity, while *para*-methoxy substitution eliminated activity altogether. This may be due to the increased pK_a of this thiol, resulting in a lower concentration of the deprotonated substrate in solution. Supporting this conclusion, a similar lack of product formation was observed using benzyl mercaptan which would remain protonated in the reaction mixture.

Table 2.2 S-H functionalization catalyzed by the ArM variant POP-ZA$_4$-H328.

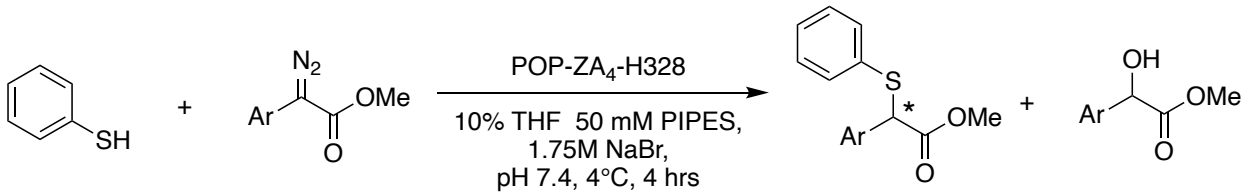


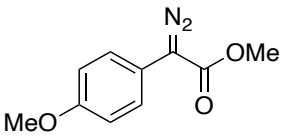
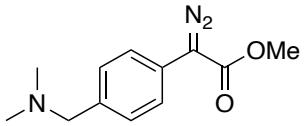
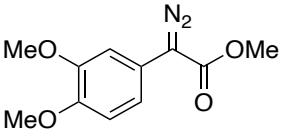
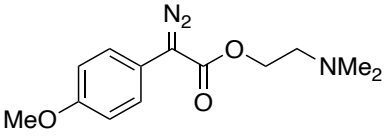
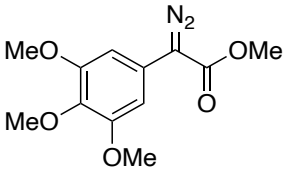
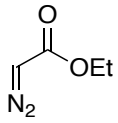
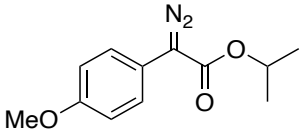
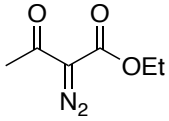
Entry	Substrate	Yield (%)	e.e. (%)
1		77	16
2		70	16
3		72	10
4		65	17
5		17	12
6		0	N/A
7		0	N/A

With the best thiol determined to be the simple benzenethiol, the diazo substrate was then varied using different aryl and ester groups (Table 2.3). As methyl (4-methoxyphenyl)diazoacetate was found to be the highest yielding diazo in cyclopropanation reactions, substrates with additional methoxy substitution were examined. Modest changes to the enantioselectivity and significantly lower yields were observed. Increasing the steric bulk of the ester also resulted in lowered yields.

Additionally, using acceptor-acceptor or acceptor-only diazos resulted in no product formation, as was the case in N-H functionalization.

Table 2.3 Diazo substrate scope of the S-H functionalization of thiobenzene.

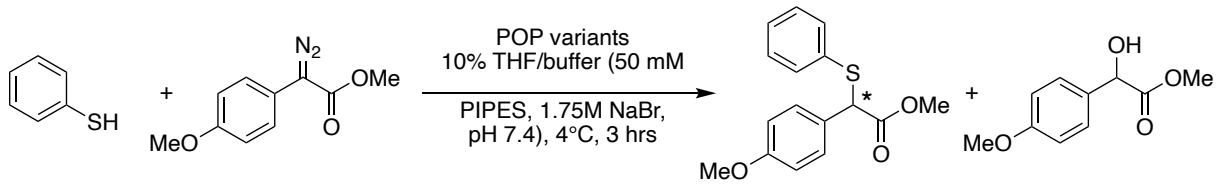


Entry	Substrate	Yield (%)	e.e. (%)	Entry	Substrate	Yield (%)
1		77	16	5		0
2		51	18	6		0
3		34	33	7		0
4		41	10	8		0

As the highest yield and selectivity were observed using benzenethiol and methyl (4-methoxyphenyl)diazoacetate, these substrates were used in the examination of S-H functionalization catalysis by the cyclopropanation lineage (Table 2.4). As was the case for N-H functionalization, a clear trend of higher yields and improved selectivities were observed along the

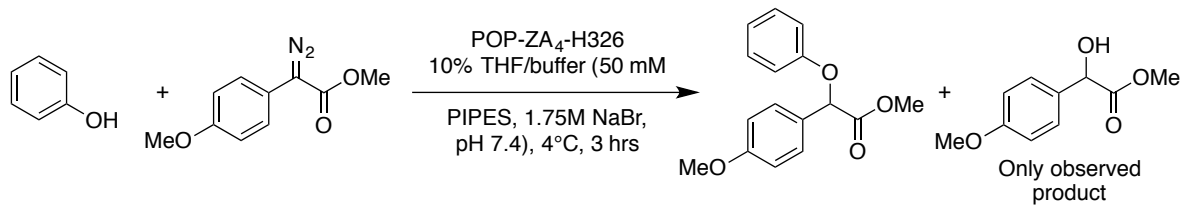
lineage. The binding of the prochiral intermediate to enable enantiospecific proton transfer is controlled by the scaffold and affected by mutations over the course of the lineage.

Table 2.4 Evaluation of the cyclopropanation lineage on S-H functionalization.



Entry	Variant	Lineage	Yield (%)	e.e. (%)
1	0-ZA4	Z477	51	6
2	1-NAGS	Z477	66	10
3	3-VRVH	Z477	64	32
4	1'-RFY	Z413	78	-8

With the best substrate pair for yield and selectivity determined to be benzenethiol and methyl (4-methoxyphenyl)diazoacetate, S-H functionalization catalysis by the cyclopropanation lineage was examined. As was the case for N-H functionalization, a clear trend was observed with higher yields and improved selectivity achieved by 3-VRVH compared to earlier variants. The binding of the prochiral intermediate to enable enantiospecific proton transfer is controlled by the scaffold and affected by mutations over the course of the lineage. As the major side product in each of these reactions is the O-H functionalization of water, we then wondered if the activity observed in these scaffolds would extend to less nucleophilic substrates, such an analogous O-H functionalization reaction with phenol. With a pK_a of 9.9, this substrate would not be deprotonated to a significant extent under the reaction conditions, so it would be considerably less nucleophilic than benzenethiol. Indeed, the dirhodium ArM reaction involving this substrate resulted exclusively in the O-H insertion side product, as shown in Scheme 2.1.



Scheme 2.1 O-H functionalization of phenol.

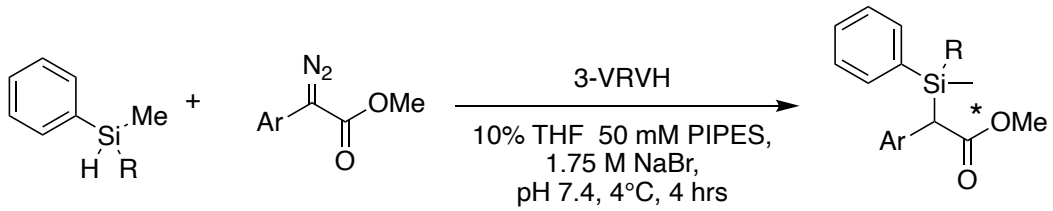
The results presented here further demonstrate the POP scaffold's ability to both catalyze the formal insertion of the carbenoid into S-H bonds as well as bind the prochiral enolate intermediate for enantiospecific proton transfer. The only chiral dirhodium-catalyzed examples of this reaction require a chiral phosphoric acid catalyst to bind the intermediate, so the fact that a single ArM can achieve selectivity, albeit with low e.e., highlights the potential utility of this class of catalysts. S-H functionalization catalyzed by dirhodium-ArMs was used to demonstrate these features, but due to the relatively poor enantioselectivity and low synthetic utility for this class of transformations, N-H functionalization was selected as the primary future direction for development and directed evolution.

2.2.3 Initial Exploration of Si-H Functionalization

The stepwise mechanism operative in the previous reactions requires binding of a planar intermediate to the protein scaffold to enable chiral proton transfer. The cyclopropanation reaction that this lineage was evolved for, on the other hand, proceeds through an asynchronous concerted mechanism,[5] where the enantiodetermining step occurs in a rhodium-bound transition state. This mechanism is analogous to the concerted carbenoid insertion into Si-H and C-H bonds, which suggests a variant with improved cyclopropanation activity should also be improved for these reactions. During the initial development of dirhodium-BCN-based ArMs, the tHisF scaffold had

been used to catalyze cyclopropanation and Si-H functionalization.[1] No enantioselectivity had been observed for either reaction in that effort, presumably due to poor encapsulation of the bulky cofactor in the small tHisF scaffold. In the directed evolution of POP-based ArMs[4] however, good enantioselectivity for cyclopropanation was observed, which suggests that Si-H functionalization may also be carried out with similar control.

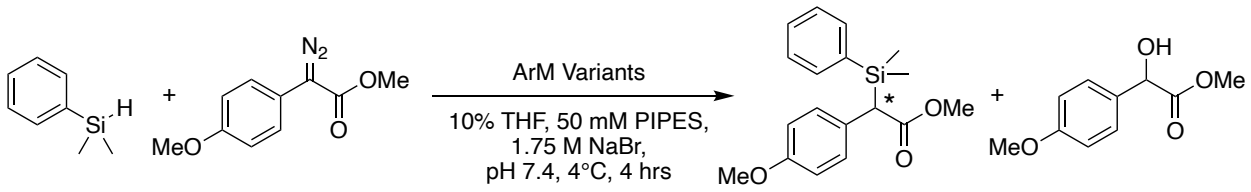
Table 2.5 Silane scope in the Si-H functionalization reaction catalyzed by 3-VRVH.



Entry	R	Ar	Yield (%)	e.e. (%)
1	Me	4-Methoxyphenyl	35	64
2	H	4-Methoxyphenyl	28	58
3	Ph	4-Methoxyphenyl	15	16
4	Me	3,4-Dimethoxyphenyl	21	28
5	Me	3,4,5-Trimethoxyphenyl	20	58

To test this hypothesis, the Si-H functionalization of several aryl silanes was examined using the final variant from the cyclopropanation lineage, 3-VRVH, as the catalyst (Table 2.5). Satisfyingly, multiple silanes yielded the desired product with good enantioselectivity, though yields were generally low. The best substrate was found to be dimethylphenylsilane, yielding 35% of the desired product with an e.e. of 64%. This silane substrate was then used with a number of diazos, though it was found that the initial methyl (4-methoxylphenyl)diazoacetate performed best in both yield and selectivity.

Table 2.6 Si-H functionalization catalyzed by the cyclopropanation of lineage.



Entry	Variant	Lineage	Yield (%)	e.e. (%)
1	0-ZA4	Z477	45	6
2	1-NAGS	Z477	43	14
3	3-VRVH	Z477	35	64
4	1'-RFY	Z413	40	-30

These substrates were then used to examine the performance of the cyclopropanation lineage (Table 2.6). While the e.e. was found to increase along the lineage as expected, the yield of Si-H functionalized product decreased slightly. This differs from all other previous observations, where variants with improved enantioselectivity for cyclopropanation also have improved yield. Up to this point it appeared that mutations were generally improving the scaffold, likely by improving localization of the cofactor and substrate positioning, as well as decreasing access of water to the carbenoid intermediate. However, it appears that these effects are not operating in the same way for the silane in Si-H functionalization, suggesting that there may be substrate-specific binding interactions. These experiments lay the groundwork for the evolution of Si-H functionalization catalyzed by dirhodium ArMs, described in depth in Chapter 5.

2.2.4 Initial Exploration of C-H Functionalization

The functionalization of C-H bonds is one of the foremost challenges in synthetic chemistry as their ubiquity in organic molecules makes them an excellent target for rapid, late-stage transformations.[19,20] The universality of these bonds presents a significant selectivity challenge,

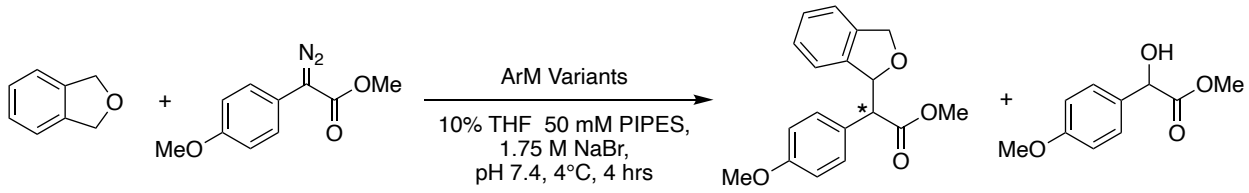
where reacting at a single position among many similar ones is necessary for the practicality of this chemistry. Significant recent progress toward this end has focused on organometallic complexes that coordinate to proximal functional groups on the substrate and direct catalysts to a single position. Selective C-H functionalization in the absence of directing groups is also possible using catalysts that can discriminate subtle steric, electronic, or stereoelectronic properties of different C-H bonds, including some excellent examples from the Davies group using complex dirhodium tetracarboxylate catalysts.[21,22]

Despite impressive advances in catalytic C-H bond functionalization, controlling catalyst selectivity, and perhaps more importantly tuning the selectivity of a given catalyst to functionalize different C-H bonds, remains challenging. The synthetic power of C-H functionalization thus remains best illustrated by natural product biosyntheses, which frequently involve enzyme-catalyzed functionalization of unactivated C-H bonds on both simple and complex molecular frameworks.[23] Based on the precedent for this type of control by enzymes and the successful Si-H functionalization of aryl silanes by members of the ArM cyclopropanation lineage described above, we hypothesized the ArM scaffold could be used to control C-H functionalization by a dirhodium cofactor in a similar way.

Highly activated C-H bonds were targeted for initial exploration of this type of reactivity to maximize the possibility of successful catalysis given the problems associated with O-H insertion even for facile reactions like N-H insertion. Gratifyingly, the benzyl ether C-H bonds of phthalan were successfully functionalized by 3-VRVH (Table 2.7), though low yield and poor selectivity were observed. The diastereoselectivity was altered compared to same reaction catalyzed by dirhodium tetraacetate in organic solvent, which suggests that the ArM was affecting the selectivity to some extent. Indeed, 0-ZA_4 was found to yield only trace product meaning that

the mutations in 3-VRVH improved the scaffold for C-H functionalization. This was very exciting, but the very low yields prohibited further development of the reaction.

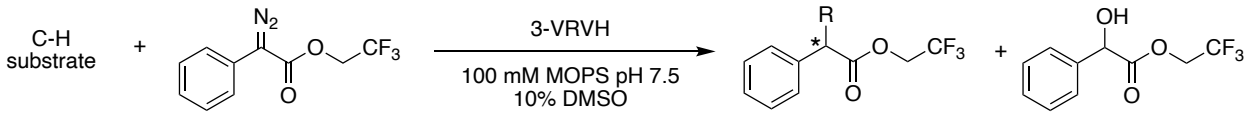
Table 2.7 C-H functionalization of phthalan catalyzed by POP-ArMs.

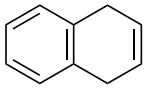
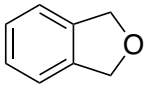


Entry	Variant	Yield (%)	e.e. (%)	d.r.
1	3-VRVH	2.1	16	1.2:1
2	HFF	2.4	12	1.2:1
3	Rh2(OAc)4	-	0	4.2:1

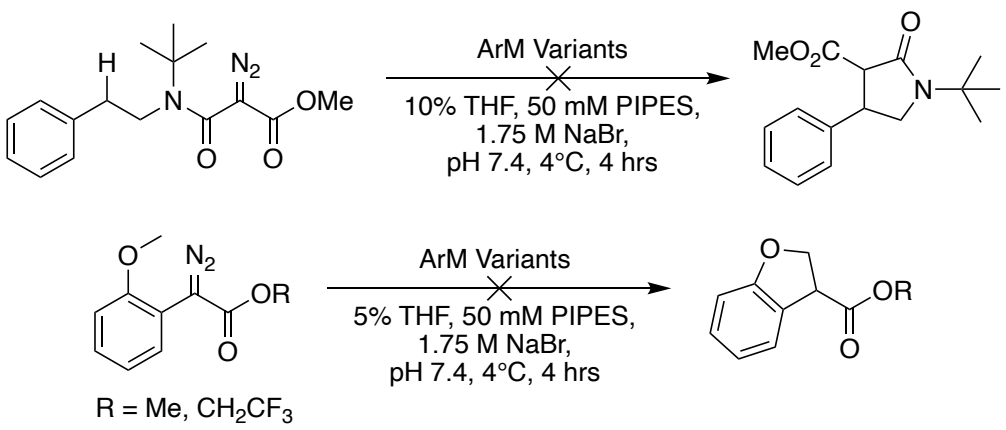
In order to increase the yield of C-H functionalized product, several approaches were explored. There have been many examples of dirhodium catalyzed C-H functionalization that make use of electron-withdrawing esters,[24] such as trifluoro- or trichloroethyl esters, to achieve high levels of activity and selectivity on many types of C-H bonds. It was also reported that streptavidin-based dirhodium ArMs catalyze C-H functionalization using trifluoroethyl phenyldiazoacetate. Though no selectivity was achieved in that investigation, as shown in Table 2.8, we used the reported substrates and conditions for testing the activity of our ArM. Modest yields were observed using the trifluoroethyl ester with three C-H substrates but unfortunately no enantioselectivity was observed in any case. It is likely that scaffold cannot differentiate between subtle differences in binding with the increased reactivity of this ester.

Table 2.8 C-H functionalization of activated substrates with a trifluoroethyl ester diazo.



Entry	Substrate	Yield (%)	e.e. (%)
1		35	0
2		13	0
3		8	0

While these highly activated substrates produced product, there were also several substrates with less activated C-H bonds that were tested but did not yield greater than trace product. Among these were the benzyl C-H bonds of cumene, ethylbenzene, and indane. We hypothesized that either that the necessary transition state geometry for C-H functionalization was not being accessed by these substrates or the local concentration of substrate was not sufficient to out-compete active site water. To overcome the latter issue, we turned to intramolecular C-H functionalization reactions.



Scheme 2.2 Intramolecular C-H functionalization reactions attempted.

A shown in Scheme 2.2, an initial attempt using the more reactive acceptor-acceptor diazo resulted in no C-H functionalization of the benzyl position. Both methyl and trifluoroethyl esters were used as the accepter moiety in the intramolecular C-H functionalization of 2-methoxylphenyldiazoacetates, but in both cases no desired product was formed. Based on the results throughout the initial exploration of C-H functionalization, it was clear that substrate selection and optimization was not going to be sufficient to achieve the desired activity and selectivity.

2.2.5 High-Throughput Methodology Optimization for C-H Functionalization Screening

Finally, we explored the possibility of using directed evolution to improve the yield and selectivity of ArM-catalyzed C-H functionalization. Formation of enantioenriched product from the reaction of phthalan with methyl phenyldiazoacetate (Table 2.7) indicates that the ArM scaffold was controlling the reaction to some extent and could therefore be improved. The high-throughput Ni-NTA screening method developed by Hyun June Park was a significant improvement for cyclopropanation directed evolution, so we started by examining this method for

C-H functionalization. Despite significant optimization of the library screening method (see Section 4.2.1.1 for more details), including expression conditions, lysis and wash procedure, as well as reaction conditions, such as buffer, sodium halide, and cosolvent, we were unable to reliably determine the yield and selectivity of variants in high-throughput. The very low product yield meant that the error associated with the screening methodology was far too high to trust any hit that might have been found in screening.

We had found that using a trifluoroethyl ester diazo substrate produced significantly higher product yield, even though no enantioselectivity was found. Because the yield with the methyl ester was simply too low for reliable results, we thought that using this substrate to evolve higher yields could enable us to find generally improved variants for C-H functionalization, then use further evolution to obtain enantioselective variants in the future. Many of the improvements found to improve screening for the methyl ester carried over for the trifluoroethyl ester. Unfortunately, despite these improvements, there was a significant background reaction that was observed when using this diazo. Wells containing no lysate (None, Figure 2.3) resulted in significantly higher product yields than those with non-bioconjugating scaffold (POP WT) or ArM (3-VRVH), meaning that evolution was not possible using this approach.

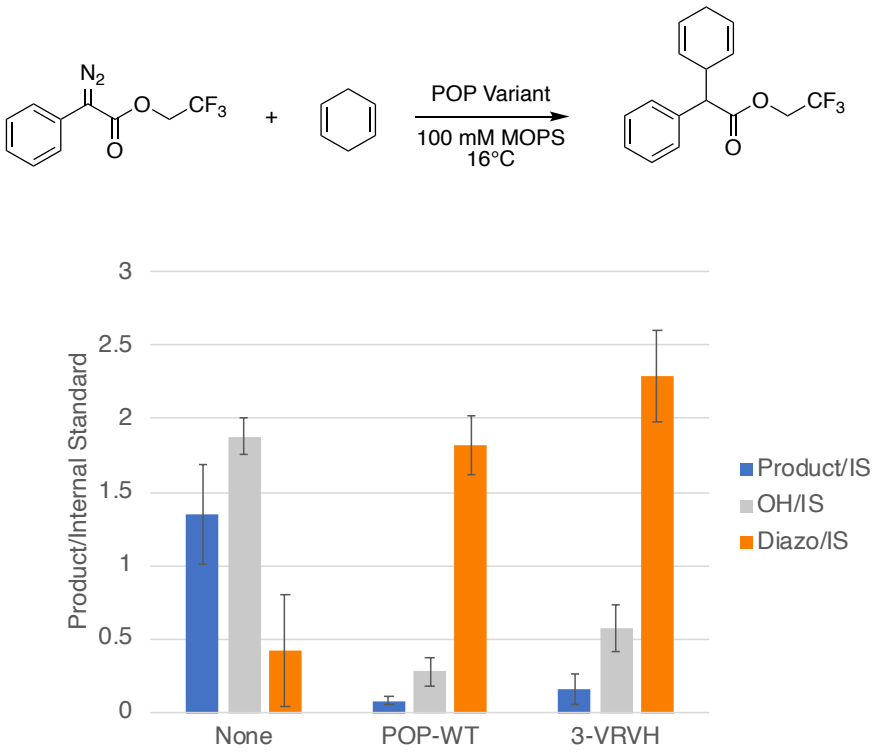


Figure 2.3 C-H functionalization dehydrobenzene in high throughput screening conditions.

Though the screening optimization that was performed here did not enable reliable screening of C-H functionalization, the improvements to the method carried over to the directed evolution campaigns described in Chapters 3, 4 and 5, enabling reliable determination of the yields and selectivity for diazo coupling, N-H functionalization and Si-H functionalization reactions. Through Si-H functionalization in particular, we envisioned an alternate path to selective C-H functionalization. The mechanistic similarities between Si-H and C-H functionalization suggest that a scaffold capable of high levels of control over Si-H bonds functionalization should also be better for C-H bonds. We therefore planned to use the concept of substrate walking in a mechanistic context, where improving Si-H functionalization can lead us to variants with superior C-H functionalization as well.

2.3 Conclusion

In each of these reactions, the final variant, 3-VRVH, was better yielding and more selective than the parent. The mutations accumulated throughout the lineage seem to generally improve the scaffold for dirhodium catalysis. The improved enantioselectivities suggest that the cofactor is being localized better by 3-VRVH than 0-ZA_4, but in the cases of N-H and S-H functionalization, the mutations accumulated seem to also play a role in binding intermediates to enable chiral proton transfer. The His326 mutation in 3-VRVH accounts for much of this, but the improvements found in the prior variants suggest that ligating residues are not the only way to accomplish this. Additionally, the higher yields and product/alcohol ratio observed over the lineage suggests that the scaffold is improved at excluding water from the active site.

In the case of N-H and Si-H functionalization reactions, directed evolution campaigns are underway to further improve the yields and selectivities of these reactions (Chapters 4 and 5, respectively). It may be possible that improved variants in N-H functionalization are also more selective in S-H insertion chemistry, but that remains to be seen. For C-H functionalization, we determined that our current methods are not sufficient for reliable for screening this reaction. We do, however, plan to use the variants with improved Si-H functionalization yields to catalyze this reaction and eventually move towards C-H functionalization reactions catalyzed by these dirhodium ArMs.

2.4 Experimental

General materials:

Unless otherwise noted, all reagents were obtained from commercial suppliers and used without further purification. Deuterated solvents were obtained from Cambridge Isotope Laboratories, Inc

(Tewksbury, MA). Silicycle silica gel plates (250 mm, 60 F254) were used for analytical TLC, and preparative chromatography was performed using SiliCycle (Quebec City, QC) SiliaFlash silica gel (230-400 mesh). Azide Agarose (Cat# 1038-25) was purchased from Click Chemistry Tools LLC. Labquake™ Tube Shaker/Rotators was purchased from Thermo Scientific (Catalog# 4002110Q). Oligonucleotides were purchased from Integrated DNA Technologies (San Diego, CA). pEVOL-pAzF was a gift from Peter Schultz (Addgene plasmid# 31186) of the Scripps Research Institute, CA[1]. *E. coli* DH5α and BL21 (DE3) cells were purchased from Invitrogen (Carlsbad, CA). NcoI-HF (Cat# R3193) and XhoI (Cat# R0146) restriction enzymes, T4 DNA Ligase (Cat# M0202), Taq DNA Ligase (Cat# M0208) and Phusion® High-Fidelity DNA Polymerase (Cat# M0530) were purchased from New England Biolabs®, Inc (Ipswitch, MA). Jacquetonitrile was purchased from Zephyr Labs. Luria broth (LB; Cat# L24040), rich medium (2YT; Cat# X15600) and Agar (Cat# A20020) were purchased from Research Products International, Corp (Mt. Prospect, IL). QIAquick Gel Extraction Kit (Cat# 28706) and QIAprep Spin Miniprep Kit (Cat# 27106) were purchased from QIAGEN Inc. (Valencia, CA) and used according to the manufacturer's instructions. DNA Clean and Concentrator™-5 (Cat# D4004) was purchased from Zymo research (Irvine, CA) and used according to the manufacturer's instructions. Library colonies were picked using an automated colony picker (Norgren Systems CP7200). All genes were confirmed by sequencing at the University of Chicago Comprehensive Cancer Center DNA Sequencing & Genotyping Facility (900 E. 57th Street, Room 1230H, Chicago, IL 60637). Electroporation was carried out on a Bio-Rad MicroPulser™ using method Ec2. Nickel (II) nitrilotriacetic acid (Ni-NTA) resin (Cat# 88223) and Pierce® BCA Protein Assay Kits (Cat #23222) were purchased from Fisher Scientific International, Inc. (Hampton, NH), and used according to the manufacturer's recommendations (for Ni-NTA resin, 5 mL of resin was used,

with buffers delivered by a peristaltic pump at a rate of 1 mL/min, in a 4 °C cold cabinet in the dark under a red safelight[1]). Amicon® Ultra-15 Centrifugal Filter Units with Ultracel-30 membrane (Cat# UFC903024) were purchased from EMD Millipore (Billerica, MA) and used according to manufacturer's recommendations. Biotage reverse phase columns (Cat# SNAPKP-C18-HS) were purchased from Biotage (Uppsala, Sweden). 96-well filter microplates (Cat# 201009-100) were purchased from Seahorse Bioscience (North Billerica, MA).

General Experimental

Unless otherwise specified, all reactions were prepared in flame or oven-dried glassware under an inert N_2 atmosphere using either syringe or cannula techniques. TLC plates were visualized using 254 nm ultraviolet light. Flash column chromatography was carried out using Silicycle 230-400 mesh silica gel. ^{1}H and ^{13}C NMR spectra were recorded at 500 MHz and 126 MHz, respectively, on a Bruker DMX-500 or DRX-500 spectrometer, and chemical shifts are reported relative to residual solvent peaks. Chemical shifts are reported in ppm and coupling constants are reported in Hz. Yields were determined by HPLC with 1,2,4-trimethoxybenzene or 1,3-benzodioxole as the internal standard and reported as the average of three trials from the same batch of ArM set up in parallel. High resolution ESI mass spectra were obtained using an Agilent Technologies 6224 TOF LC/MS or an Agilent Technologies 6540 Q-TOF MS-MS. Low resolution ESI mass spectra were obtained using an Agilent Technologies 6130 LC-MS. LC-MS/MS experiments were performed with an Easy-nLC 1000 ultra-high pressure LC system (ThermoFisher) using a PepMap RSLC C18 column (column: 75 μm x 15 cm; 3 μm, 100 Å) coupled to a Q Exactive HF Orbitrap and Easy-Spray Nanosource (ThermoFisher). Amicon® Ultra-15 Centrifugal Filter Units with Ultracel-30 membrane (50 mL volume, 30 kDa cutoff) were used to concentrate or wash protein solutions.

Protein concentrations were measured using the Pierce® BCA Protein Assay Kit and protein stocks were then stored at -80 °C until use.

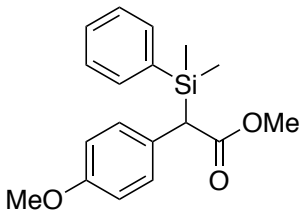
Synthesis of aryldiazoacetates and cyclopropanes

Aryldiazoacetates were synthesized as previously reported.[1]

Synthesis of racemic standards

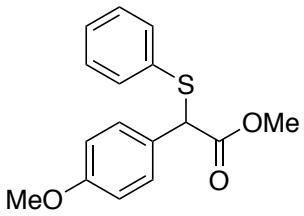
The standards for heteroatom insertions were synthesized using the following general method: 113.4 mg (0.55 mmol) methyl (4-methoxy)phenyldiazoacetate in 4 mL anhydrous dichloromethane was added dropwise to a stirring solution of 0.60 mmol substrate and 2.7 mg (1 mol%) $Rh_2(OAc)_4$ in 8 mL anhydrous dichloromethane under N_2. The resulting green solution was refluxed with stirring for 3 hours. The solvent was evaporated and the resulting oil was purified *via* flash chromatography (10% ethyl acetate/hexanes) to give the product.

- Methyl (1-phenyl-*p*-methoxy)-2-(dimethylphenylsilyl)acetate



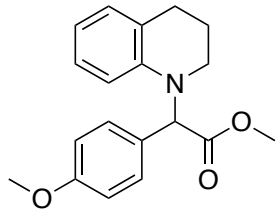
General procedure using dimethylphenylsilane as the substrate resulted in a colorless oil product. Spectra match previous reports.[4]

- Methyl 2-(4-methoxyphenyl)-2-(phenylthio)acetate



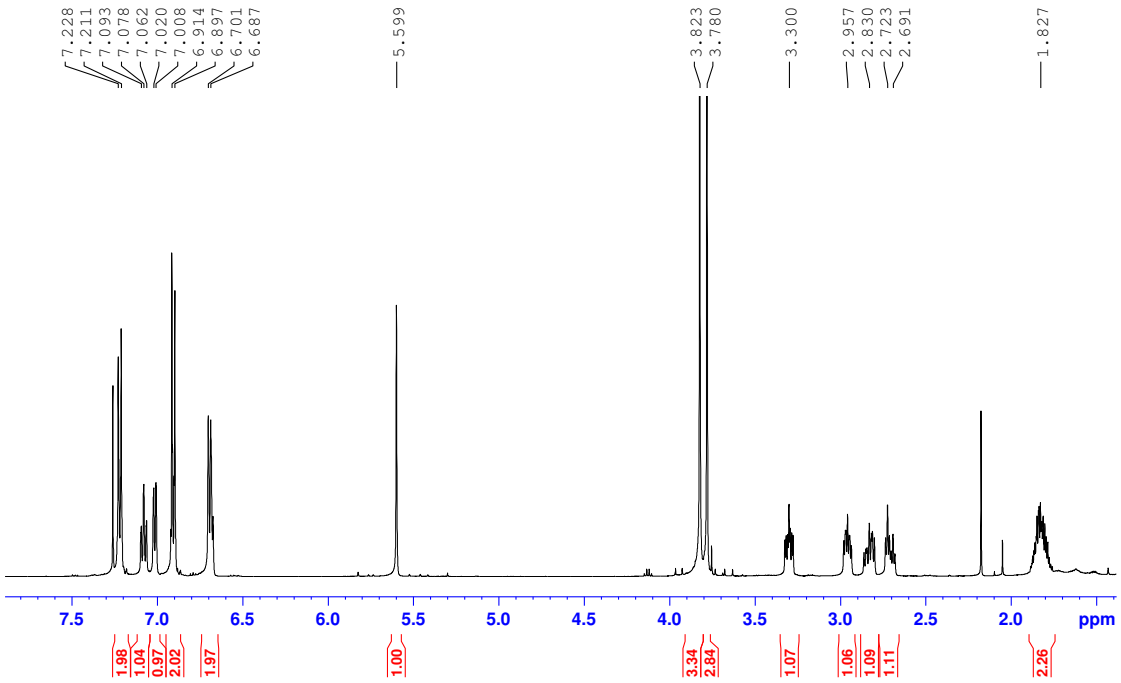
General procedure using benzenethiol as the substrate resulted in a product of white solids. Spectra match previous reports.[5]

- 1(2H)- Quinolineacetic acid, 3,4-dihydro-α-(4-methoxyphenyl)-, methyl ester

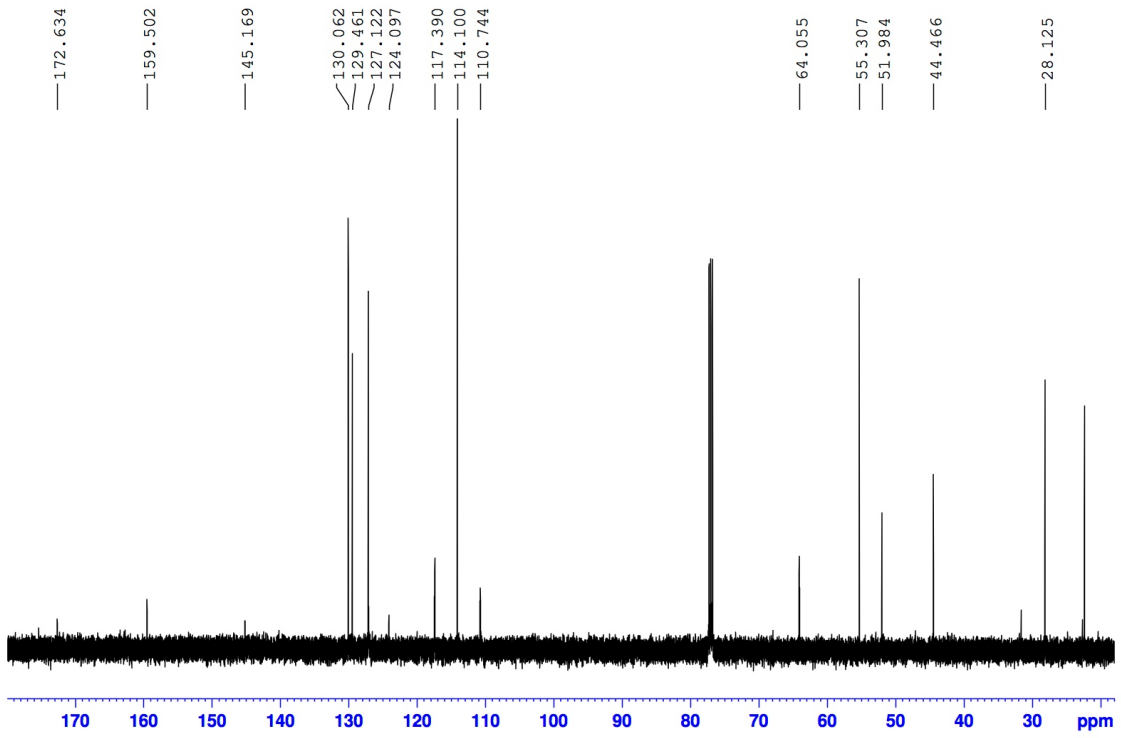


General procedure using 1,2,3,4-tetrahydroquinoline as the substrate resulted in a colorless oil product. ^{1}H NMR (500 MHz, CDCl3) ∂ 7.22 (d, J = 8.7 Hz, 2H), 7.08 (t, J = 7.8 Hz, 1H), 7.01 (d, J = 7.0 Hz, 1H), 6.91 (d, J = 8.7 Hz, 2H), 6.72-6.66 (m, 2H), 5.60 (s, 1H), 3.82 (s, 3H), 3.78 (s, 3H), 3.32-3.28 (m, 1H), 2.98-2.93 (m, 1H), 2.86-2.81 (m, 1H), 2.73-2.68 (m, 1H), 1.86-1.78 (m, 2H); 13C NMR (125 MHz, CDCl3) ∂ 172.6, 145.2, 130.1, 129.5, 127.1, 124.1, 117.4, 114.1, 110.7, 64.1, 55.3, 52.0, 44.5, 28.1;HRMS (ESI-MS) calculated for $C_{19}H_{21}NO_3$ ($M+H^+$) 312.1500, found 312.1496

NMR spectrum:



^{1}H NMR (500 MHz, $CDCl_3$)



^{13}C NMR (500 MHz, $CDCl_3$)

Scaffold Expression

The putative hits identified from the library screening were grown, expressed, lysed, and purified according to a previous report.[3] A colony for the selected mutant was inoculated in 5 mL 2YT medium with 50 μg/mL kanamycin and 20 μg/mL chloramphenicol. The culture was incubated overnight at 37 ºC with constant shaking at 250 rpm. On the following day, 5 mL of the overnight cultures was used to inoculate 500 mL of fresh 2YT media having the same antibiotics, in a 2.8L Fernbach flask. The culture was incubated at 37 ºC, 250 rpm, and protein expression was induced by adding isopropyl β-D-1-thiogalactopyranoside (1mM final concentration), 4-azido-L-phenylalanine (1mM final concentration) and 1% (w/v) L-arabinose when OD_{600}=1.0. The induced culture was allowed to grow for 12 hours, and then the cells were harvested by centrifugation at 4 ºC, 3000 x g for 20 minutes. Cell pellets were re-suspended in 30 mL PBS (pH 7.5) and sonicated (40 amplitude, 30 second burst, 10 minute total process). Lysed culture was clarified by centrifugation at 16000 x g, 4 ºC for 30 minutes and supernatant thus obtained was purified by Ni-NTA resin using the manufacturer's instructions. Purified protein was buffer exchanged to 50 mM Tris-HCl buffer (pH 7.4) and measured by Pierce® BCA Protein Assay Kit as recommended.

Bioconjugation

To set up bioconjugation, a solution of the POP mutant (480 μL, 75 μM in 50 mM Tris-HCl buffer, pH 7.4) and a solution of cofactor (120 μL, 0.75 mM in ACN, 0.594 mg/mL) were added to a 1.5 mL microcentrifuge tube and shaken at 750 rpm at 4 °C overnight. The final concentrations were: 60 μM POP, 150 μM, 20 vol% acetonitrile/Tris-HCl buffer. The resulting solution was treated with 100 μL azide agarose resin, and rotated on the Labquake™ Tube Shaker/Rotator in a 4 °C cold

cabinet for 24 hours to remove excess cofactor. The suspension was then centrifuged at 5000 rpm for 3 minutes and the supernatant was transferred to a new microcentrifuge tube. The resin was washed twice with 600 µL of 50 mM Tris-HCl buffer (pH 7.4) and centrifuged at 5000 rpm for 3 minutes. These supernatants were combined with the first supernatant and buffer exchanged to proper buffers for use in biocatalysis or characterization. ESI-MS was used to characterize the bioconjugates. The total protein concentration was calculated based on its absorbance at 280 nm (A_{280}) and the calculated extinction coefficient for the protein (109,210 $M^{-1}cm^{-1}$ from ExPASy), which is consistent with concentrations measured by Pierce® BCA Protein Assay Kit; the cofactor absorbance at 280 nm is negligible relative to POP in aqueous solution under the concentrations used. The efficiency of dirhodium cofactor incorporation was calculated based on the ratio of the high resolution ESI-MS peak intensity of the ArM and scaffold ($I_{ArM}/(I_{ArM}+I_{scaffold})$); the effective ArM concentration was calculated by multiplying the total protein concentration by the efficiency of dirhodium incorporation ($[\mathrm{ARM}] = [\mathrm{Total\ protein}] * I_{\mathrm{ARM}}/(I_{\mathrm{ARM}} + I_{\mathrm{scaffold}})$). The effective ArM loading was adjusted to 1 mol% with respect to the dirhodium cofactor in bioconversions. Example ESI-MS and deconvolution are shown in Supplementary Figure 4. Additionally, excess cofactor scavenging by azide agarose resin was verified by low resolution LC-MS performed on an Agilent Technologies 6130 LC-MS using a Grace Vydac 218TP C18 5µ column (4.6 mm i.d. x 250 mm) with a flow rate of 1.0 ml/min and a detection wavelength set at 280 nm. The following gradient was used: 20 % to 64% B from 0-10 min, 64 % to 64% B from 10-15 min, 64 % to 80 % B from 15-17 min, 80 % to 80 % B from 17-20 min, 80 % to 20 % B from 20-21 min, 5 min post-run (solvent A: water with 0.1% TFA; solvent B: acetonitrile with 0.1% TFA).

Biocatalysis Reactions

To set up biocatalysis, solutions of aryldiazoacetate (25 μL, 96 mM, in THF), X-H substrate (25 μL, 485 mM, in THF), and POP-ArM solution (500 μL, the effective ArM concentration adjusted to 48 μM with respect to the dirhodium cofactor according to the aforementioned method) were added to a 1.5 mL microcentrifuge tube. The final concentrations of the reagents were: 22 mM olefin, 4.4 mM aryldiazoacetate, 44 μM POP-ArM. The resulting mixture was left shaking at 750 rpm at 4 °C overnight (or for various time points, *vide infra*). The reaction was quenched by adding 20 μL 1,2,4-trimethoxybenzene solution or 1,3-benzodioxole solution (30 mM, in 10% isopropanol/hexanes) and 600 μL ethyl acetate. The mixture was vortexed and centrifuged (15,000 x g, 3 minutes). The top organic layer was collected and the bottom aqueous layer was extracted with 600 μL ethyl acetate twice. The organic extracts were combined, evaporated and re-dissolved in 200 μL 10% isopropanol/hexanes, and analyzed on NP-HPLC to determine conversions and enantioselectivity. The conversions and enantioselectivity were reported as the average of three trials from the same batch of ArM set up in parallel. The NP-HPLC was performed on an Agilent 1100 Series HPLC system using a Phenomenex Lux® 3u Cellulose-1 column (1000 Å, 3 μm, 4.6 mm i.d. x 250 mm), with a flow rate of 1.0 mL/min and detection wavelength set at 230 nm. The following gradient was used: 10 % to 70 % B from 0-10 min, 70 % B from 10-15 min, 70 % to 100 % B from 15-18 min, 100 % B from 18-22 min, 4 min post-run (solvent A: 2-propanol; solvent B: hexanes).

Ni-NTA Library Methodology

Library expression and lysis

Library colonies were picked using an automated colony picker (Norgren Systems) and arrayed in 1-mL 96-well plates containing 300 µL LB media with 50 µg/mL kanamycin and 20 µg/mL chloramphenicol. In each plate, 8 wells were saved for picking parent colonies as positive controls. Cells were grown overnight for 14-16 hours at 37 ºC, 250 rpm. 60 µL of the overnight culture was used to inoculate 6 mL 2YT media (with 50 µg/mL kanamycin and 20 µg/mL chloramphenicol) in 6-mL 24-well plates. In each 24 well plate, 2 wells were inoculated with parent as positive controls. Following growth at 37 ºC, 200 rpm for about 6 hours, to an OD_{600} = ~1.2, enzyme expression was induced by adding Isopropyl β-D-1-thiogalactopyranoside (final concentration 1mM), 4-azido-L-phenylalanine (final concentration 1mM) and 1% (w/v) L-arabinose. After growth at 37 ºC, 200 rpm for about 16 hours, to an OD_{600}= 3.4~3.6, the cultures were harvested by centrifugation at 3600 rpm, 4 °C, for 20 minutes and the supernatants were discarded. The cell pellets were washed by adding 4 mL Tris-HCl buffer (50 mM, pH 7.4) to each well and incubating the plate at 37 ºC, 200 rpm for 10 minutes to re-suspend the cells. The suspended cells were centrifuged at 3600 rpm, 4°C for 10 minutes and the supernatant was discarded.

The washed cell pellets were suspended in 630 µL Tris-HCl buffer (50 mM, pH 7.4) containing 1 mg/mL lysozyme and incubated at 37 ºC, 250 rpm for 60 minutes. The plate was then flash frozen in liquid nitrogen for 10 minutes and then thawed in a 37 ºC water bath. 70 µL Tris-HCl buffer (50 mM, pH 7.4) containing 1.0 mg/mL DNase was added and the plate was incubated at 37 ºC, 250 rpm for 30 minutes. The lysed cells were submitted to a heat treatment for 15 min at 75 ºC in a water bath, after which the plate was centrifuged at 3600 rpm, 4 °C for 30 minutes. 500 µL of the supernatant was transferred to a 1.5 mL 96-well filter plate (polypropylene, 0.45 µM).

Library immobilization and purification

100 μL of a 50% Ni-NTA resin suspension was added to each well while mixing the resin to ensure consistency. The 96-well plate was then shaken at 600 rpm at 4°C for 1 hour, at which point the plate was centrifuged at 3000 rpm for 5 minutes and the flow-through discarded. 500 μL of Ni-NTA wash buffer (25 mM imidazole, 300 mM NaCl, 25mM dibasic sodium phosphate, pH 7.4) was added to each well and the plate was centrifuged at 3000 rpm for 5 minutes, discarding the flow-through. 480 μL of 50 mM Tris-HCl buffer (pH 7.4) was added to each well, followed by 120 μL of a 0.013 μM solution of cofactor **1**. The plate was shaken at 600 rpm for 1 hour. The plate was centrifuged for 5 minutes at 3000 rpm and the flow through was discarded. To wash the resin-bound artificial metalloenzyme suspension, 500 μL of acetonitrile followed by 500 μL of reaction buffer (50 mM PIPES, 1.75 M NaCl) was added to each well. The plate was centrifuged at 3000 rpm for 5 minutes. This wash procedure was done twice more for a total of three washes, discarding the flow through after each one. The plate was centrifuged for 1 minute following the final wash to remove any excess liquid.

Library screening

450 μL of reaction buffer was then added to each well followed by 10 μL of a substrate solution in THF (0.048 M methyl 4-methoxyphenyldiazoacetate, 0.24 M 4-methoxystyrene). The plate was mixed at 600 rpm at 4°C for 4 hours. The reaction plate was centrifuged at 3000 rpm for 5 minutes using a new 1.5mL 96-well plate to collect the flow-through. To each well, 200 μL ethyl acetate was added and the plate was centrifuged at 3000 rpm for 5 minutes into the same collection plate. This organic wash was repeated twice more, for a total of three 200 μL ethyl acetate washes. The

collection plate (containing aqueous reaction buffer and 600 μL ethyl acetate) was sealed and shaken at 600 rpm for 10 minutes. This plate was centrifuged at 2000 rpm for 2 minutes to separate the aqueous and organic layers. Using a multichannel pipette, 300 μL ethyl acetate was removed from the top layer of each well and placed into a new 96-well plate which was evaporated under vacuum. 200 μL of 1:1 isopropanol/hexanes was added to each well and the plate was analyzed using SFC. The SFC analysis was performed on an Agilent 1260 Infinity Analytical SFC System using a Daicel Corporation Chiralpak IC-3 column (3 μm, 4.6 mm i.d. x 150 mm), with a flow rate of 3.0 ml/min and detection wavelengths set at 210, 230, 254, and 280 nm. An isocratic elution was used with 10% B from 0-3.5 min (solvent A: supercritical CO_2; solvent B: methanol).

2.5 Acknowledgements

I would like to thank Hao Yang and Alan Swartz for teaching me how to prepare dirhodium ArMs, use them to catalyze carbene insertion reactions, and synthesize many of the compounds explored in this work. Additionally, I would like to thank Chen Zhang for guidance and assistance in early N-H functionalization studies and Hyun June Park for assistance in learning how to use the Ni-NTA screening method.

2.6 References

[1] H. Yang, P. Srivastava, C. Zhang, J. C. Lewis, *Chembiochem* 2014, *15*, 223–227.

[2] P. Srivastava, H. Yang, K. Ellis-Guardiola, J. C. Lewis, *Nat Commun* 2015, *6*, 7789.

[3] J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *J Am Chem Soc* 2002, *124*, 9026–9027.

[4] H. Yang, A. M. Swartz, H. J. Park, P. Srivastava, K. Ellis-Guardiola, D. M. Upp, G. Lee, K. Belsare, Y. Gu, C. Zhang, R. E. Moellering, J. C. Lewis, *Nat Chem* 2018, *10*, 318–324.

[5] J. Hansen, J. Autschbach, H. M. L. Davies, *J Org Chem* 2009, *74*, 6555–6563.

[6] X.-C. Wang, X.-S. Song, L.-P. Guo, D. Qu, Z.-Z. Xie, F. Verpoort, J. Cao, *Organometallics* 2014, *33*, 4042–4050.

[7] Z.-Z. Xie, W.-J. Liao, J. Cao, L.-P. Guo, F. Verpoort, W. Fang, *Organometallics* 2014, *33*, 2448–2456.

[8] Y.-Y. Ren, S.-F. Zhu, Q.-L. Zhou, *Org Biomol Chem* 2018, *16*, 3087–3094.

[9] B. Xu, S. Zhu, X. Xie, J. Shen, Q. Zhou, *Angewandte Chemie Int Ed* 2011, *50*, 11483–11486.

[10] X. Xu, C. Li, Z. Tao, Y. Pan, *Adv Synth Catal* 2015, *357*, 3341–3345.

[11] H. Saito, T. Uchiyama, M. Miyake, M. Anada, S. Hashimoto, T. Takabatake, S. Miyairi, *Heterocycles* 2010, *81*, 1149.

[12] J. Wu, X. Li, X. Qi, X. Duan, W. L. Cracraft, I. A. Guzei, P. Liu, W. Tang, *J Am Chem Soc* 2019, *141*, 19902–19910.

[13] T. C. Nugent, M. El-Shazly, *Adv Synth Catal* 2010, *352*, 753–819.

[14] I. Slabu, J. L. Galman, R. C. Lloyd, N. J. Turner, *Acs Catal* 2017, *7*, 8263–8284.

[15] B. Xu, S. Zhu, X. Zuo, Z. Zhang, Q. Zhou, *Angewandte Chemie Int Ed* 2014, *53*, 3913–3916.

[16] X. Zhang, M. Ma, J. Wang, *Arkivoc* 2003, *2003*, 84–91.

[17] B. Xu, S.-F. Zhu, Z.-C. Zhang, Z.-X. Yu, Y. Ma, Q.-L. Zhou, *Chem Sci* 2014, *5*, 1442.

[18] V. Tyagi, R. B. Bonn, R. Fasan, *Chem Sci* 2015, *6*, 2488–2494.

[19] H. M. L. Davies, J. R. Manning, *Nature* 2008, *451*, 417–24.

[20] H. M. L. Davies, D. Morton, *Chem Soc Rev* 2011, *40*, 1857.

[21] K. Liao, S. Negretti, D. G. Musaev, J. Bacsa, H. M. L. Davies, *Nature* 2016, *533*, 230–234.

[22] W. Liu, Z. Ren, A. T. Bosse, K. Liao, E. L. Goldstein, J. Bacsa, D. G. Musaev, B. M. Stoltz, H. M. L. Davies, *J Am Chem Soc* 2018, *140*.

[23] J. C. Lewis, P. S. Coelho, F. H. Arnold, *Chem Soc Rev* 2010, *40*, 2003–2021.

[24] H. M. L. Davies, K. Liao, *Nat Rev Chem* 2019, *3*, 347–360.

Chapter 3: Engineering Dirhodium Artificial Metalloenzymes for Diazo Coupling Cascade Reactions

3.1 Introduction

3.1.1 Chemo- and Site-Selectivity in Catalysis

Transition metals catalysts are uniquely valuable in synthetic chemistry because they can bind and react with a wide range of chemical species and functional groups. The source of this value also presents an inherent challenge: selectively reacting with the desired functional group on the desired substrate over other reactive species in the reaction mixture. In the laboratory, this challenge is typically avoided through careful design of the synthetic sequence and reaction conditions so that the catalyst does not have to differentiate similar chemical species, frequently by masking problematic functionality on substrates using protecting groups. Without the concern of these alternate side reactions, the metal catalyst can be tuned to achieve the desired characteristics towards the primary reaction, whether that be selectivity or increased yield.

The standard strategy for the optimization of these small molecule metal catalysts is through ligand design since modifying the primary coordination sphere through steric and electronic perturbations can have significant effects on catalysis.[1] While less common, there has been significant effort devoted to using the secondary coordination sphere interactions, such as attractive substrate-catalyst interactions distal to a metal center, to modulate selectivity.[2,3] Thorough control over the primary and secondary coordination spheres of a catalytic metal results in highly selective catalysis.[4,5] Some of the best demonstrations of this exist in nature,[6,7] however, and many natural metalloenzymes demonstrate exquisite selectivity and reactivity through both primary and secondary sphere effects.[8] In fact, these examples have inspired many efforts to recapitulate enzyme-like secondary sphere effects in small molecule complexes.[9]

Enantio- and regioselectivity are impressive demonstrations of the importance of both primary and secondary sphere effects, but that these reactions are performed chemoselectively within the complex cellular milieu is equally compelling. Control over substrate positioning is required for enantioselectivity and is demonstrated by a plethora of synthetic small molecule transition metal catalysts. The capability of metalloenzymes to achieve this along with near-absolute chemoselectivity suggests that the molecular recognition imparted by second sphere interactions in enzymes enables far greater control over transition metal reactivity than can currently be achieved with small molecule ligands.[10] Similar control over synthetic metal complexes could enable reactions in complex environments, including enzymatic and chemoenzymatic cascades containing multiple catalysts, reagents, and intermediates.[11,12]

3.1.2 Artificial Metalloenzymes in Biosynthetic Cascade Reactions

As discussed in length in Chapter 1, ArMs have been used to merge the reactivity of synthetic catalysts with the selectivity and evolvability of protein scaffolds.[13] Thus far, the development of ArMs has primarily focused on their synthetic utility, where features such as high enantioselectivity and TON (turnover number) are desirable. These goals are frequently achievable by small molecule catalysts, so the potential for the protein scaffold to control the primary and secondary coordination spheres are not being fully harnessed. Towards this end, developing ArMs that can perform chemoselective chemistry in a complex reaction mixture containing reactive chemical species would necessitate control over the metal that is uncommon outside of natural metalloenzymes.

There are numerous examples of ArM-catalysis in complex reaction conditions. Indeed, streptavidin-,[14,15] LmrR-,[16] and albumin-based[17] ArMs have been used for *in vivo* catalysis. In

some of these cases, localization of the ArM in specific cellular compartments was required to avoid poisoning of the metal center. Additionally, streptavidin-,[18] FhuA-,[19] and P450-based[20] ArMs have been used for *in vitro* cascade reactions. In each of these examples, however, inherent cofactor reactivity alleviates the need for scaffold-controlled chemoselectivity. The ArMs produce the same products as the metal cofactors alone, albeit with impressive rate acceleration and enantioselectivity. To truly demonstrate scaffold-controlled chemoselectivity, reactive species must be actively selected against by the scaffold.

3.1.3 Dirhodium ArM-Catalyzed Diazo Coupling

Dirhodium complexes react with donor-acceptor diazo compounds to generate carbene complexes that react with water, thiols, amines, olefins, silanes, other diazo compounds, and even sp^3 C-H bonds,[21,22] so we anticipated that this extensive reactivity would enable studies on how protein scaffolds might modulate transition metal chemoselectivity through secondary sphere control, rather than through reaction condition optimization or ligand design. Previous studies, outlined in Chapter 2, demonstrated that ArM chemoselectivity can be evolved to favor cyclopropanation over formal carbene insertion into water O-H bonds,[23] an example of scaffold dependent chemoselectivity. Based on those results, we hypothesized that this level of control over dirhodium reactivity could enable cascade reactions involving a variety of additional species in solution.

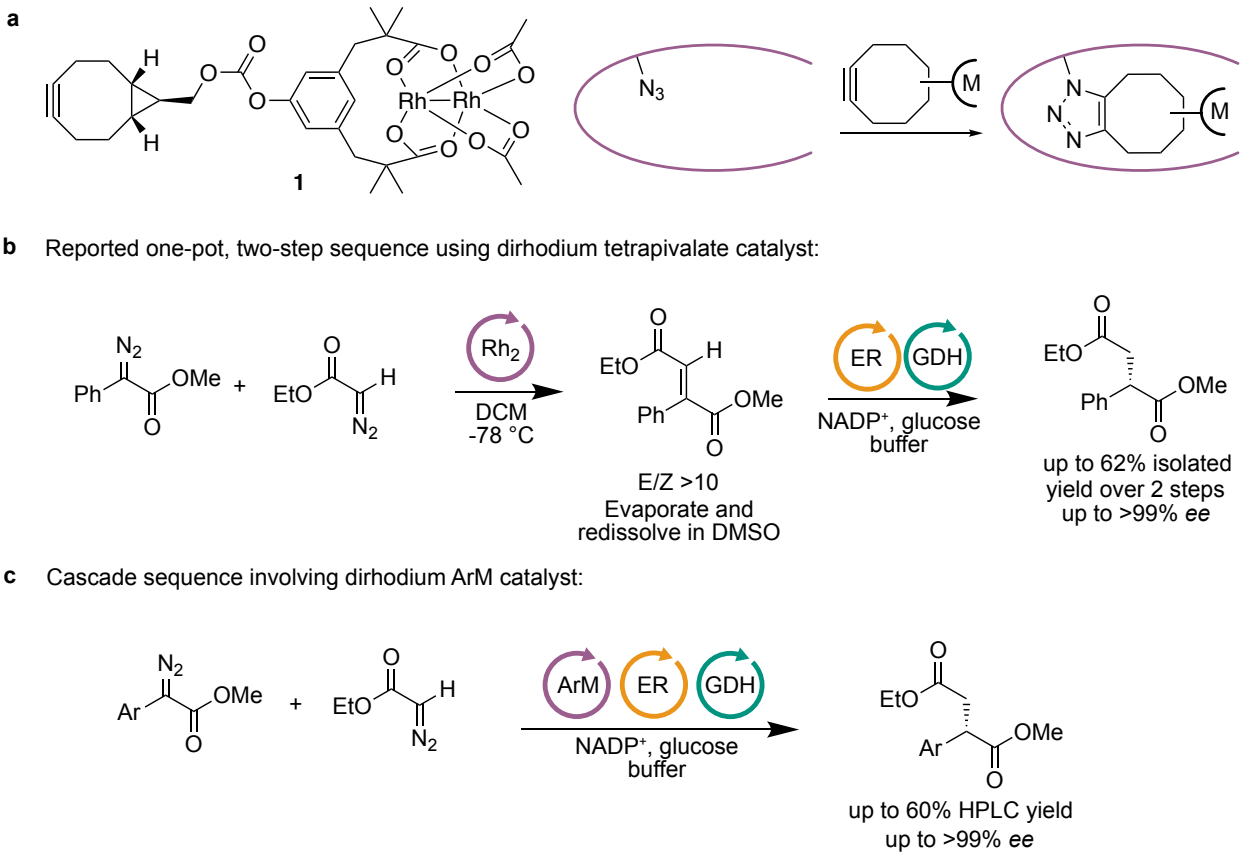


Figure 3.1 Dirhodium ArM formation and catalysis. A) Cofactor **1** and covalent bioconjugation via azidophenylalanine. B) Synthesis of chiral aryl succinates through a one-pot, two-step reaction.[24] C) POP-**1**/ER cascade catalysis.

For example, dirhodium-catalyzed diazo cross-coupling has been used to generate fumaric acid esters that are converted by alkene reductases to 2-substituted succinate derivatives (Scheme 1B).[24] In this study, and in the first report of the reaction by Davies[25], dirhodium catalysis was conducted in organic solvent at cryogenic temperatures. As the enzymatic reaction requires aqueous buffer, the solvent was evaporated and the residue was redissolved in DMSO for reduction. In addition to the incompatibility of the conditions for the two reaction steps, dirhodium is known to react with some of the necessary additives in the reduction reaction, such as glucose[26]

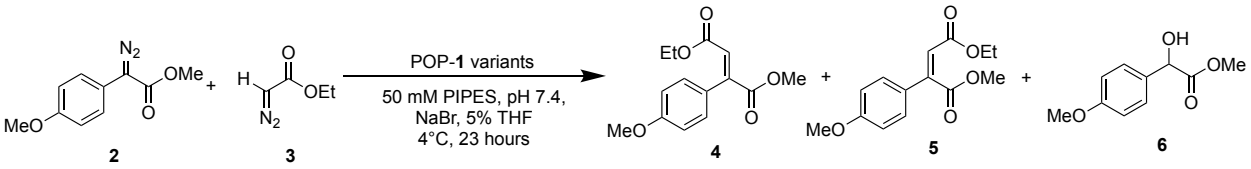
and proteins[27], making this an excellent system to demonstrate the capacity of the ArM scaffold for chemoselective catalysis in complex reaction mixtures.

3.2 Results and Discussion

3.2.1 Initial Investigation and Optimization of ArM-Catalyzed Diazo Coupling

To determine if dirhodium-ArMs successfully catalyze diazo coupling, previously-developed scaffolds were evaluated using a model diazo coupling reaction (Entries 1-3, Table 3.1). As the ene reductases investigated in this work only accept *E*-alkenes,[24] it was important to establish whether ArMs could provide this isomer in good yield under conditions suitable for biocatalysis. The initial ArM variant 0-ZA_4, containing only 4 alanine mutations and azidophenylalanine at position 477 catalyzed diazo coupling with 34% yield and a 2.9:1 *E*/*Z* ratio (**4**/**5**). Not surprisingly given the trends discussed in Chapter 2, the final variant in the cyclopropanation lineage, 3-VRVH,[28] was improved, yielding 41% of diazo coupled alkene with a 3.4:1 *E*/*Z* ratio. Variant 1-SGH,[28] which contains only the three mutations in 3-VRVH that are necessary for its cyclopropanation enantioselectivity, provided a similar yield and *E*/*Z* ratio as 3-VRVH. As scaffolds with fewer mutations from wild-type are generally more stable and thus more amenable for further evolution, 1-SGH was selected as the parent for further optimization. It was found that ArM loading could be reduced to 0.1 mol% with minimal change in yield, but excess **3** was required (Tables 3.1 and 3.2).

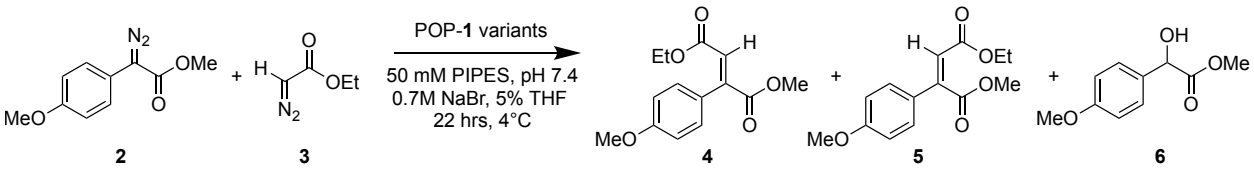
Table 3.1 Scaffold selection and reaction optimization.



Entry[a]	Variant	[ArM] (uM)	[**3**]	[NaBr]	% Yield[b] **4**	**5**	**6**	E/Z
1	0-ZA4	50	25	0.7	34	12	48	2.9
2	3-VRVH	50	25	0.7	41	12	35	3.4
3	1-SGH	50	25	0.7	43	13	39	3.3
4	1-SGH	5	25	0.7	44	12	31	3.7
5	1-SGH	5	5	0.7	39	13	37	3.1
6	1-SGH	5	25	0.1	34	19	13	1.8
7	1-SGH	5	25	1.75	46	8	33	5.7

[a]Standard reaction conditions: 5 mM **2**, 25 mM **3**, 50 mM PIPES pH 7.4, 5% cosolvent, 22 hours at 4°C with shaking. [b]Determined by SFC analysis using 1,3,5-trimethoxybenzene as internal standard. Reported yields and E/Z values are the average of triplicate reactions.

Table 3.2 Optimization of ArM concentration for diazo coupling.



Entry[a]	Variant	[ArM] (uM)	Time (hrs)	% Yield[b] 4	5	6	E/Z	TTN to 4
1	1-SGH	50	22	43	13	39	4.4	43
2	1-SGH	5	22	44	12	31	3.7	445
3	1-SGH	2	22	44	11	37	4.3	1103
4	1-SGH	0.5	22	40	11	38	3.6	4078
5	1-SGH	0.05	22	14	6	25	2.6	14339
6	1-SGH	0.05	96	27	17	36	1.6	26824
7	5-G	50	22	60	5	13	11.2	60
8	5-G	5	22	76	5	16	14.9	761
9	5-G	2	22	73	5	13	13.4	1826
10	5-G	0.5	22	68	6	14	11.2	6755
11	5-G	0.05	22	32	4	9	8.3	31928
12	5-G	0.05	96	45	16	11	2.8	44612

[a]Standard reaction conditions: 5 mM **2**, 25 mM **3**, 50 mM PIPES pH 7.4, 5% THF, 22 hours at 4°C with shaking. [b] Determined by SFC analysis using 1,3,5-trimethoxybenzene as internal standard. Reported yields and E/Z values are the average of triplicate reactions.

We previously established that high salt concentrations are required for high ArM activity and selectivity,[23] but activity of the alkene reductases were found to be negatively affected by the 1.75 M NaBr we had previously used. While this was the best concentration for ArM activity and selectivity, 0.7 M NaBr was found to be sufficient for both the ArM and alkene reductases (Table 3.1, Entries 4, 6, and 7). It is worth noting that in all cases, the ArM E/Z selectivity is lower than the analogous reaction using bulky dirhodium tetraacetate catalysts in organic solvent under cryogenic conditions (typically >10:1)[24], so improving this was a key goal as we turned to directed evolution to improve the scaffold for diazo coupling. In addition to improving the E/Z

ratio, we also sought to find variants that minimized the formation of the O-H insertion side product that was being formed (39% yield).

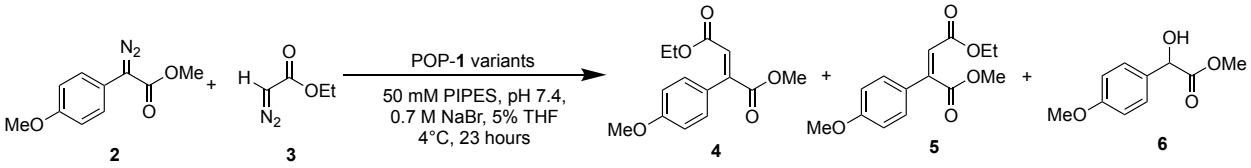
3.2.2 Directed Evolution of ArM Scaffold for Diazo Coupling

As significant effort had gone into optimization of the Ni-NTA high-throughput methodology for C-H functionalization (described in Section 2.2.5), the method was found to be quite robust when used with diazo coupling. As such, scaffold libraries containing the Z-477 mutation were expressed in 24-well plates, transferred to a 96-well filter plate and bioconjugated in lysate using cofactor **1**. The resulting ArMs were immobilized on Ni-NTA resin and 50% acetonitrile/50 mM Tris buffer was used to wash the resin in order to remove excess cofactor. The diazo coupling reactions were then conducted in the wells using the immobilized ArM in a heterogenous mixture. Following reaction, the reaction mixture was removed by centrifugal filtration, extracted with ethyl acetate, and the organic products were analyzed by SFC.

As discussed in Section 2.1.1, previous engineering efforts, both using error-prone PCR and CCM, revealed that mutations in a β-strand (residues 98-101) across from a putative Rh-binding histidine residue significantly affected ArM-catalyzed cyclopropanation activity and selectivity. Our initial evolution approach was therefore to target these positions to find the optimal residue at each of these positions. Site-saturation libraries of 1-SGH were constructed at each position in this β-strand using degenerate NNK codons. This effort revealed that Q98P improved the yield of **4**, increased the *E*/*Z* ratio, and decreased the yield of **6** (variant 2-P, Table 3.3). Site-saturation mutagenesis of S99 of 2-P provided variant 3-H (Table 3.3), which yielded a similar amount of product as 2-P but had an increased *E*/*Z* selectivity. Similar mutagenesis of residues F100 and T101 did not lead to further improvements, so we turned to an alternate mutagenic

approach, combinatorial codon mutagenesis (CCM)[29]. This strategy had been used in the evolution of the 0-413Z variant 1'-RFY by targeting 25 active site residues projecting into the interior of the β-propellor domain of the scaffold using degenerate NDT codons. Based on this success, the same residue pool was used in the mutagenesis of 3-H. The mutation V71G (4-G, Table 3.3) was identified using this approach, but a subsequent CCM library did not yield a positive variant.

Table 3.3 Directed evolution of ArM for diazo coupling.



Entry[a]	Variant	Mutagenesis Method	Mutations from previous generation	ArM mol%	%Yield[c] **4**	**5**	**6**	**4/5**	TTN to **4**	TTN **4**+**5**+**6**
1	1-SGH	Parent	G99S/S301G/Y326H	0.1	44	12	31	3.7	445	876
2[b]	1-SGH	Parent	G99S/S301G/Y326H	0.001	27	17	36	1.6	26824	79696
3	2-P	Q98NNK	Q98P	0.1	55	11	33	4.8	545	984
4	3-H	S99NNK	F99H	0.1	51	7	24	7.3	511	821
5	4-G	Active site CCM	V71G	0.1	72	8	21	8.5	717	1006
6	5-G	E283NNK	E283G	0.1	76	5	16	14.9	761	970
7[b]	5-G	E283NNK	E283G	0.001	45	16	11	2.8	44612	72196

[a] Standard reaction conditions: 0.1 mol% ArM, 5 mM **2**, 25 mM **3**, 50 mM PIPES pH 7.4, 5% THF, 22 hours at 4°C with shaking. [b] Standard conditions using 0.001 mol% (50 nM) ArM and 96 hour reaction time. [c] Determined by SFC analysis using 1,3,5-trimethoxybenzene as internal standard. Reported yields and E/Z values are the average of triplicate reactions.

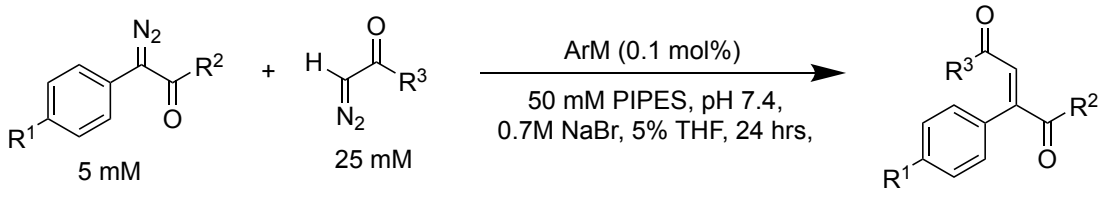
At the same time this evolution was being conducted, MD simulations (discussed in depth in Section 3.2.3) were used to identify CCM library residues proximal to the cofactor. To better explore the sequence space at these positions, site saturation libraries using NNK codons rather

than the NDT codons previously used, were used to screen six of these residues. This screen resulted in variant 5-G (E283G, Table 3.3), which was further improved for both in diazo coupling yield and selectivity. Remarkably, we found that decreasing the ArM loading 100-fold with respect to **2** to 0.001 mol% substantially increased the TTN, with 5-G catalyzing 44,612 turnovers to **4**. This is among the highest TTN reported for an ArM[30], but if turnovers associated with formation of **5** and **6** are included, a remarkable 72,196 TTN is observed, highlighting the high activity of the dirhodium cofactor within 5-G (Table 3.3). While the TTN is drastically improved in these conditions, there is a significant decrease in E/Z selectivity. Previous work had characterized an active site modification via carbene insertion during catalysis,[28] which could be responsible for the changing selectivity over thousands of turnovers.

3.2.3 Diazo Coupling Substrate Scope

The substrate scope of the parent 1-SGH and final variant 5-G were then examined. In all cases, 5-G yielded more of the desired *E*-alkene compared to 1-GSH (Table 3.4), indicating that mutations accumulated during directed evolution are generally improving the scaffold and not resulting in substantial substrate specificity. When the sterics and electronics of the aryl diazo acetate substrate are changed by varying the para substituent (R1) and ester (R2), yields are not drastically affected. Interestingly, changes to the acceptor-only diazo (R3) had a larger effect on alkene yield, with both diethyldiazoacetamide and benzyl diazoacetate resulting in lower yields of product. With the exception of the previously unreported amide substrate, these substrates are in line with the known scope for dirhodium-catalyzed diazo coupling,[24,31] indicating that the ArM enables the desired dirhodium activity while significantly reducing undesired side reactions such as water O-H insertion.

Table 3.4 Substrate scope of ArM-catalyzed diazo coupling.



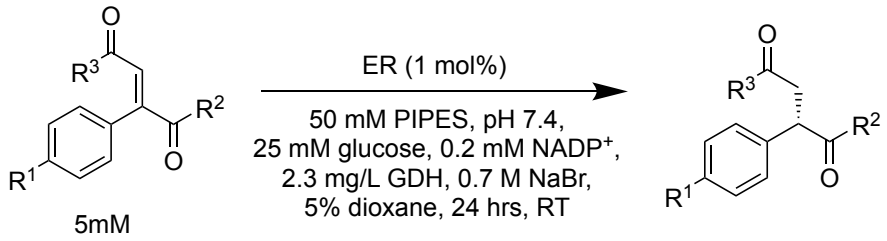
				Yield (%)[b]	
Entry[a]	R1	R2	R3	1-GSH	5-G
1	OMe	OMe	OEt	44	76
2	H	OMe	OEt	38	75
3	Cl	OMe	OEt	23	68
4	Br	OMe	OEt	38	64
5	OMe	OMe	NEt_2	33	58
6	OMe	OMe	OBn	18	37
7	Cl	Me	OEt	30	66

[a] 0.1 mol% ArM, 5 mM donor-acceptor diazo, 25 mM acceptor-only diazo, 50 mM PIPES pH 7.4, 0.7 M NaBr, 5% THF, 22 hrs at 4°C with shaking. [b] Determined by HPLC analysis using 1,3,5-trimethoxybenzene as internal standard. Reported yields are the average of triplicate reactions.

3.2.4 Cascade Reaction and Chemoselectivity of Dirhodium ArMs

Having engineered an ArM variant capable of catalyzing the desired *E*-alkene in good yields, we next turned to optimize a biocatalytic cascade reaction using an alkene reductase to reduce the alkene and form enantioenriched succinate derivatives in a single pot. The activities of several ERs were examined on the 2-aryl fumaric acid derivatives produced via ArM catalysis to select the optimal ER for each substrate. Three ERs were examined,[32,33] including alkene reductase from *Yersinia bercovieri* (YersER), enoate reductase-1 from *Kluyveromyces lactis* (KYE1), and 1,2-oxophytodienoate reductase from *Lycopersicum esculentum* (OPR1), and in each case except 4-chlorophenyldiazoacetone, there was an ER capable of forming the desired product with excellent enantioselectivity (>99%) (Table 3.5).

Table 3.5 Ene-reductase substrate selection.



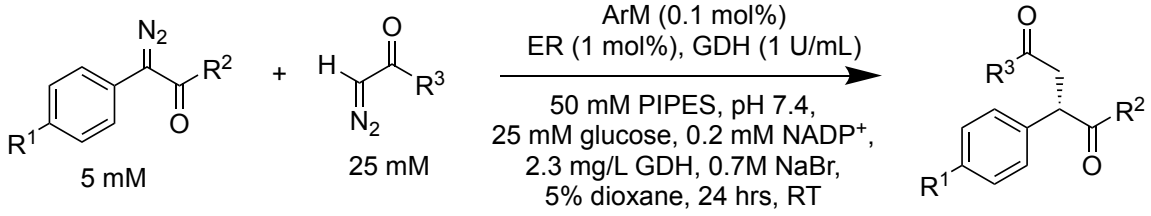
				KYE1		YersER		OPR1	
Substrate	R1	R2	R3	%Yield	% e.e.	%Yield	% e.e.	%Yield	% e.e.
1	OMe	OMe	OEt	87	>99	80	>99	63	>99
2	H	OMe	OEt	78	>99	87	>99	69	>99
3	Cl	OMe	OEt	50	>99	61	>99	41	>99
4	Br	OMe	OEt	73	>99	86	>99	74	>99
5	OMe	OMe	NEt2	60	>99	63	>99	96	>99
6	OMe	OMe	OBn	36	>99	35	>99	54	>99
7	Cl	Me	OEt	83	29	94	78	82	32

[a]Standard reaction conditions: 5 mM E-alkene, 50 μM ER, 25 mM glucose, 0.2 mM $NADP^+$, 2.3 mg/mL GDH, 50 mM PIPES pH 7.4, 0.7M NaBr, 5% dioxane, 24 hours shaking at room temperature. Yields determined by HPLC analysis using 1,3,5-trimethoxybenzene as internal standard. Reported yields are the average of triplicate reactions. Enantioselectivity was determined by normal phase HPLC after semipreparative purification from a single reaction.

The ArM/ER cascade requires that the ArM, the ER, and a glucose dehydrogenase (GDH, to supply the ER with reduced cofactor) all tolerate one another. The ArM must also not react with glucose or NADP(H) which are required for the regeneration of $FMNH_2$ in the ER. This challenge is particularly significant given that dirhodium carbenoid species react readily with water, proteins,[34] and a range of small molecule nucleophiles, including carbohydrates.[21] Remarkably, however, 1-GSH and 5-G catalyzed diazo coupling with only slightly reduced yields in the presence of all cascade components and the ERs successfully converted the majority of the fumaric

ester intermediates to the reduced products in good yields while maintaining excellent enantioselectivities (Table 3.6). In some cases, the cascade reaction scope was limited by the ERs examined. While some bulky R2 groups were coupled in good yields by the ArM, they were not reduced by any of the ERs examined. For example, 1-SGH coupled isopropyl (4-methoxyphenyl)diazoacetate with ethyl diazoacetate in 39% yield, but the bulky ester group prevented the reduction of the alkene by any of the ERs investigated. Within the known ER substrate scope, however, 5-G catalyzed the desired alkene formation with up to 723 turnovers in the cascade conditions, again highlighting the capacity of the ArM scaffold to protect the dirhodium center from deactivation and side reactions. This tolerance to additives appears to be present from the outset of the evolution as a feature inherent to the POP scaffold, as the yield of the final succinic acid derivatives tracked with the ArM alkene yields for both 1-SGH and 5-G.

Table 3.6 Substrate scope of ArM/ER cascade reactions.



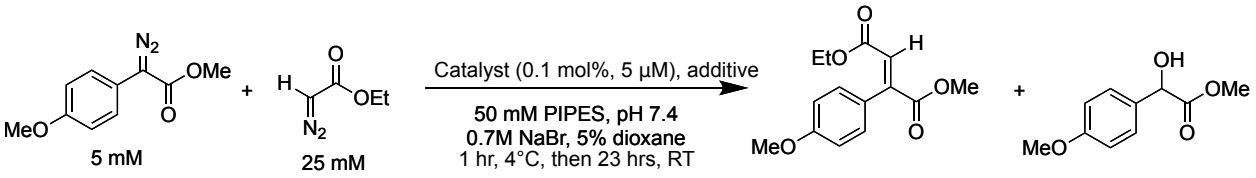
				1-SGH		5-G	
Entry[a]	R1	R2	R3	% Yield	% e.e.	% Yield	% e.e.
1	OMe	OMe	OEt	25	>99	61	>99
2	H	OMe	OEt	35	>99	56	>99
3	Cl	OMe	OEt	18	>99	47	>99
4	Br	OMe	OEt	32	>99	60	>99
5	OMe	OMe	NEt2	22	>99	40	>99
6	OMe	OMe	OBn	9	>99	12	>99
7	OMe	Me	OEt	34	79	52	78

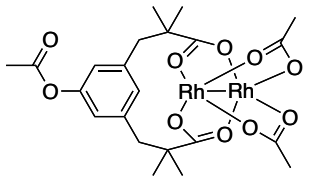
[a] 0.1 mol% ArM, 5 mM donor-acceptor diazo, 25 mM acceptor-only diazo, 50 mM PIPES pH 7.4, 0.7M NaBr, 5% dioxane, 1 hr at 4°C with shaking followed by 23 hrs shaking at 23 °C. Yields determined by HPLC analysis using 1,3,5-trimethoxybenzene as internal standard. Reported yields are the average of triplicate reactions. Enantioselectivity determined by chiral HPLC analysis.

The scaffold-based control over chemoselectivity is apparent when compared to the reaction of the dirhodium catalyst alone in water. When diazo coupling reactions catalyzed by an acetyl-substituted cofactor in aqueous buffer are performed, the OH insertion product **6** is formed almost exclusively and only trace **4** or **5** are observed (Table 3.7). Further, in the presence of glucose, formal OH insertion involving both water and glucose[35] was observed by mass spectrometry, but the latter is completely absent in the ArM catalyzed reaction (Figure 3.2). Finally, while dirhodium catalysts are capable of modifying surface-exposed protein residues,[34] no such modifications were observed by mass spectrometry for the ERs or GDHs used in the cascade reactions (Figure 3.2). Together, these results show that the scaffold modifies nucleophile

access to the dirhodium carbenoid, providing a hydrophobic environment that excludes polar nucleophiles like bulk water and glucose to enable selective diazo cross-coupling.[36]

Table 3.7 Modification of additives by dirhodium catalysts in aqueous buffer.



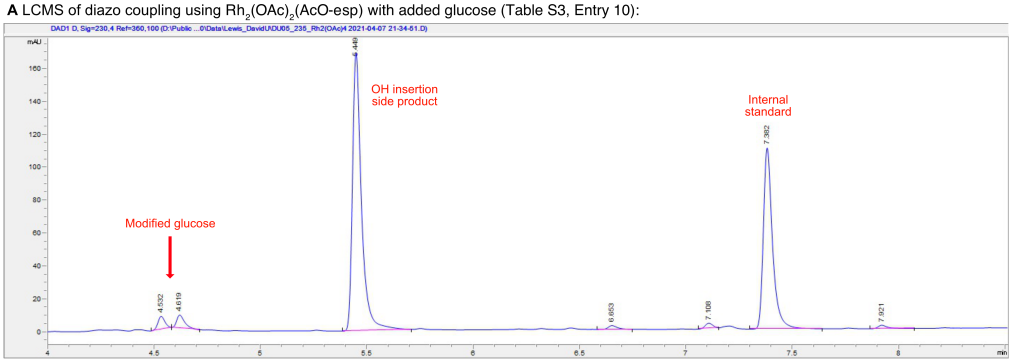


$Rh_2(OAc)_2(AcO\text{-}esp)$

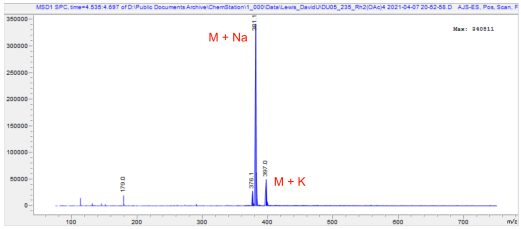
Entry	Catalyst	Additive	Modification results
1	Rh2(OAc)2(AcO-esp)	YersER (12 uM)	None detected
2	1-SGH	YersER (12 uM)	None detected
3	5-G	YersER (12 uM)	None detected
4	Rh2(OAc)2(AcO-esp)	KYE1 (12 uM)	None detected
5	1-SGH	KYE1 (12 uM)	None detected
6	5-G	KYE1 (12 uM)	None detected
7	Rh2(OAc)2(AcO-esp)	OPR1 (12 uM)	None detected
8	1-SGH	OPR1 (12 uM)	None detected
9	5-G	OPR1 (12 uM)	None detected
10	Rh2(OAc)2(AcO-esp)	Glucose (25 mM)	Modification observed
11	1-SGH	Glucose (25 mM)	None detected
12	5-G	Glucose (25 mM)	None detected

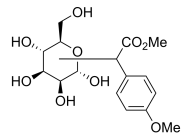
[a]Yields of small molecule products were not determined to allow for analysis of reaction components.

A LCMS of diazo coupling using $Rh_2(OAc)_2$(AcO-esp) with added glucose (Table S3, Entry 10):



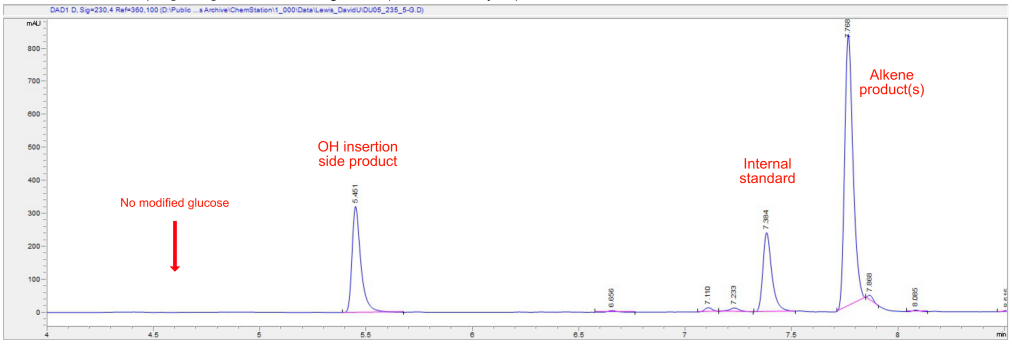
Mass spec of peaks at 4.53+4.62:





Chemical Formula: $C_{16}H_{22}O_9$
Exact Mass: 358.13
M+Na: 381.12
M+K: 397.09

B LCMS of diazo coupling using 5-G with added glucose (Table S3, Entry 12):



C Overlay of deconvoluted intact protein mass spectra of KYE1 before (orange) and after diazo coupling reaction with 5-G (Table S3, Entry 6, blue):

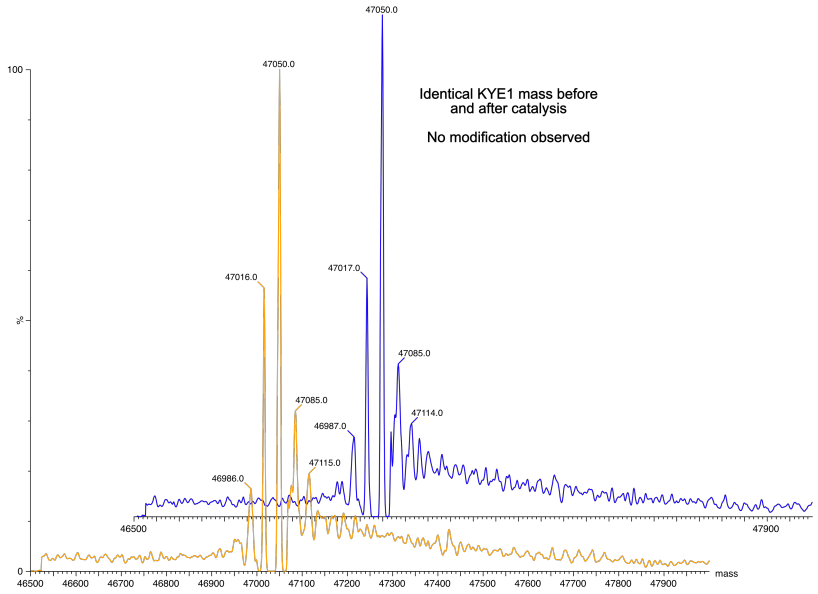


Figure 3.2 Mass spectra analyzing the modification of additives by dirhodium catalysts.

3.2.5 Molecular Dynamics of Dirhodium ArMs

To gain insight into how the POP scaffold might accomplish this level of substrate specificity, MD simulations were conducted on models of 5-G. As we have no crystal structure of the ArM, the cofactor orientation in the cofactor within the scaffold is unknown. Four cofactor geometries were prepared within the 5-G scaffold to contain a Rh-H326 bond. Short (20 ns) MD simulations were run, with 3 out of the 4 resulting in cofactors with unusual bond angles and orientations with respect to histidine, suggesting that these orientations may not form stable Rh-His interactions. The one cofactor orientation that did maintain a reasonable Rh-His bond was used for the following simulations. Three models of 5-G were prepared, one with no cofactor, a second containing the cofactor with a parameterized Rh-His interaction and a final identical model with no Rh-His bond. MD Previous modelling in the group had shown that apo-POP undergoes inter-domain opening and closing to form a solvent-exposed cleft,[37] and molecular mechanics modelling of apo-5-G found similar behavior over the 960 ns simulation (Figure 3.3A, B). We speculated that analogous dynamics in POP ArMs would facilitate cofactor bioconjugation in the open state and provide a more compact, hydrophobic environment for chemoselective catalysis in the closed state.[10] Decreasing the amount of time spent in an open conformation with a solvent exposed cofactor should result in less O-H insertion side product and more control by the scaffold over geometric selectivity.

A common mutation found to significantly improve the selectivity of POP ArM scaffolds for both cyclopropanation and diazo coupling is the incorporation of histidine in the β-propellor domain, observed in both the rationally designed H328 variant and as well as H326 in 3-VRVH found using random mutagenesis. As dirhodium tetracarboxylate complexes are known to bond to Lewis basic residues,[38] and histidine in particular,[39] we hypothesized that a cofactor-histidine

crosslink could enforce a closed conformation that could contribute to the observed improvements in ArM activity and selectivity. Supporting this notion, MD simulations of 5-G with the Rh-His bond intact showed that POP opening/closing was greatly reduced (Figure 3E). Despite its flexible linker, cofactor **1** holds 5-G closed when the Rh-His bond is present. Interestingly, a simulation starting from a state lacking a Rh-His bond was able to access an open structure for much of the simulation (676/960 ns, Figure 3D), though this system did not open to the same extent as apo 5-G (Figure 3C). This constraint appears to result from a persistent hydrophobic interaction between **1** and a number of residues on the interior surface of the 5-G scaffold. This finding could explain the improved selectivity of POP-based ArMs lacking an interior His residue,[23,28] but further free energy calculations and experimental validation will be required to establish this possibility.

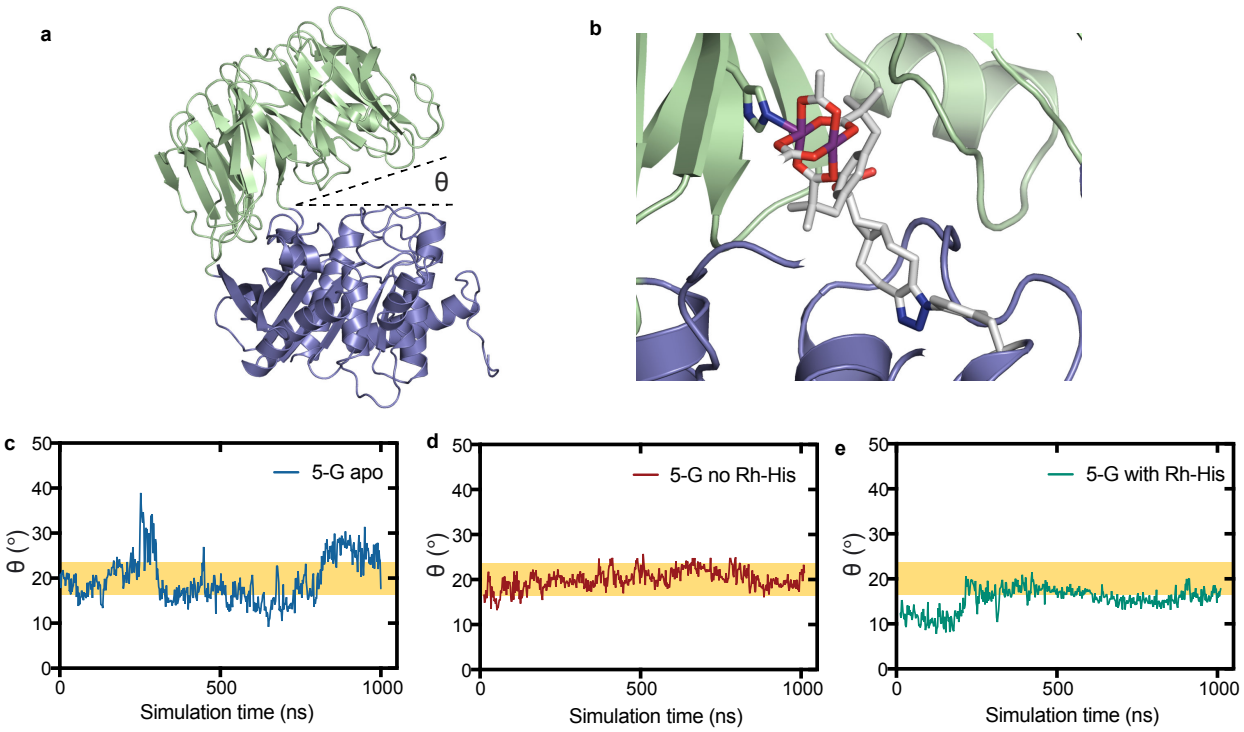


Figure 3.3 Domain dynamics of apo-POP and POP-**1** ArMs. a) Open state of apo-5-G showing the interdomain angle, θ. b) Representative trajectory showing the rhodium-histidine interaction in 5-G with the mutations found in this study highlighted in yellow. c) The interdomain angle of apo-5-G. The yellow bar (17-23°) indicates the open/closed transition. d) The interdomain angle of 5-G-**1** without a Rh-His bond. e) The interdomain angle of 5-G-**1** with a Rh-His bond.

3.3 Conclusion

Controlling the reactivity and selectivity of a transition metal catalyst requires precise tuning of its primary and secondary coordination spheres. Our previous efforts on a variety of carbene addition and insertion reactions (Chapter 2) have shown that the ArM scaffold can achieve good enantioselectivity and chemoselectivity for the desired reaction over water,[23,28] but achieving selectivity in cases where there are numerous potentially reactive species requires

greater control over the coordination environment than had been shown. In this study, we demonstrated that evolved dirhodium ArMs can achieve this level of control over the first and second sphere interactions to catalyze diazo cross-coupling reactions in complex cascade reaction mixtures, presumably by limiting substrate access to the rhodium carbenoid intermediate. Others have established that ArM scaffolds can protect catalysts from poisons such as glutathione, which reversibly bind to and deactivate metal centers,[14,17] but the current study highlights rare examples of chemoselectivity for a reaction with multiple possible substrates.

While this ArM-controlled chemoselectivity was present from the outset, directed evolution was used to increase the yield of the desired E-alkene over the Z-alkene and OH insertion side products of the initial parent 1-SGH. These improvements carried over to the cascade reaction with an ER, and the final variant, 5-G, provided high yields of the desired product under the complex reaction conditions. The four mutations leading to this variant are mostly likely involved in outer-sphere control and substrate positioning, but the H326 in 1-SGH and H99 in 3-H can bind the rhodium cofactor, potentially altering both the primary coordination sphere and interdomain dynamics. Indeed, the rhodium-His to H326 was found to alter POP dynamics in MD simulations, helping it to maintain a closed state. These models suggest a number of mechanisms by which the cofactor can affect the structural dynamics of the ArM scaffold, adopting dual catalytic and structural roles just as natural metalloenzyme cofactors do.[40] A variant from this lineage, 3-H, was found to achieve excellent selectivity for Si-H functionalization (discussed Chapter 5) suggesting that the alterations to the coordination sphere can also be impactful in other contexts. Further studies on this system in N-H and especially Si-H functionalization will help us understand the different ways that cofactor-scaffold interactions can give rise to these improvements.

3.4 Experimental

3.4.1 General materials and methods

Materials

Ethyl diazo acetate (87% in DCM), *p*-ABSA and KOD Hot Start DNA Polymerase were purchased from Sigma Aldrich. $Rh_2(OAc)_4$ was purchased from Pressure Chemicals. Other chemicals and solvents were purchased from Fisher Scientific. Vivaspin 20 centrifugation filters (20 mL volume, 30 kDa cutoff) from Sartorius were used to concentrate or wash protein solutions. Restriction enzymes were purchased from New England Biolabs. The DNA polymerase PrimeSTAR was purchased from Takara. The QIAprep Spin Miniprep Kit and QIAquick Gel Extraction Kit were purchased from Qiagen. Glucose dehydrogenase GDH-105 was purchased from Codexis. Restriction enzymes were purchased from NEB. The DNA polymerase PrimeSTAR was purchased from Takara. Zeba desalting columns (2 mL, 30 kDa cutoff) were purchased from Fisher Scientific. 2-mL 96-well filter plates (0.45 μm, PP membrane) were purchased from Agilent Technologies.

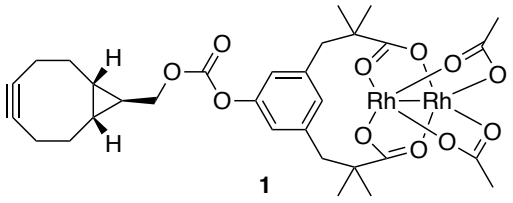
Methods

Unless otherwise specified, all reactions were prepared in flame or oven-dried glassware under an inert N_2 atmosphere using either syringe or cannula techniques. TLC plates were visualized using 254 nm ultraviolet light. Flash column chromatography was carried out using Silicycle 230-400 mesh silica gel. All ^{13}C NMR spectra were taken on a Varian 500 MHz spectrometer and ^{1}H NMR spectra were taken on either a Varian 500 MHz or Inova 400 MHz spectrometer and chemical shifts are reported relative to residual solvent peaks. Chemical shifts are reported in ppm and coupling constants are reported in Hz. Unless otherwise specified, ^{13}C NMR spectral shifts are reported as single carbons. Yields were determined by SFC and UHPLC with 1,3,5-

trimethoxybenzene as the internal standard internal standard and reported as the average of three trials from the same batch of ArM set up in parallel. SFC analysis was performed on an Agilent 1260 Infinity SFC. UHPLC analysis was performed on an Agilent 1290 Infinity UHPLC. Enantioselectivity of all chiral products were determined using an Agilent 1200 series HPLC in normal phase. Low resolution ESI mass spectra were obtained using an Agilent Technologies 1290/6135B quadrupole LC-MS. Intact protein mass spectrometry was performed using a Waters Synapt G2S HDMS using a C18 column. Protein samples were desalted using manufacturer specifications before MS analysis. Protein concentrations were measured using the Pierce® BCA Protein Assay Kit and protein stocks were then stored at -80 °C until use. Standard molecular cloning procedures were followed.

3.4.2 Synthetic Procedures

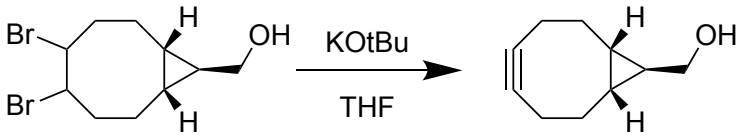
Synthesis of Cofactor 1



1

Improvements to the previously published[1] synthesis of cofactor **1** were used. Minor changes to the double elimination reaction forming BCN(exo)-OH (**2**) were found to increase yield and reproducibility and an alternate route to the esp-OH (**5**) fragment was used.

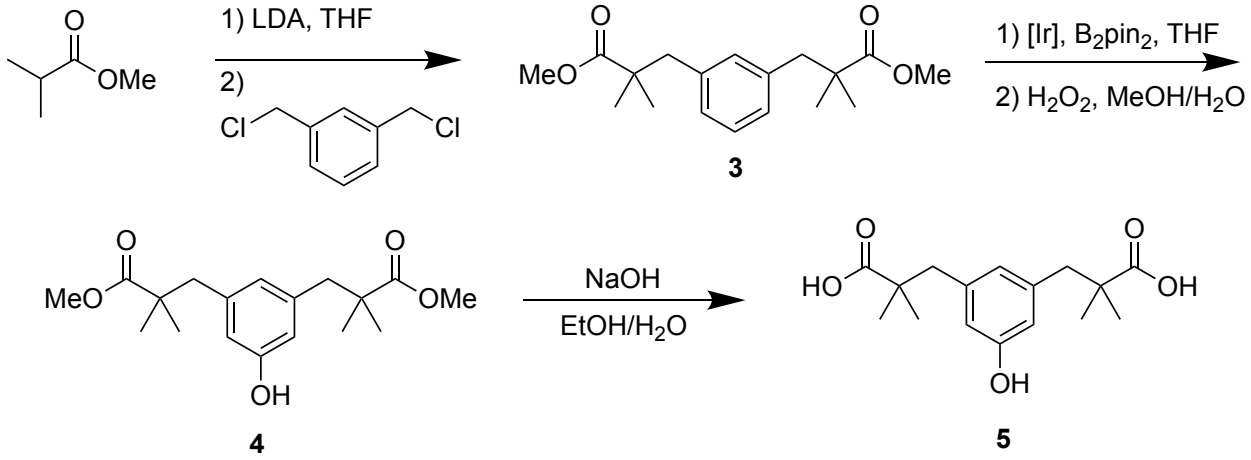
Synthesis of (1R,8S,9r)-Bicyclo[6.1.0]non-4-yn-9-ylmethanol (**2**):



2

The dibromide was prepared according to previous literature.[2] This material, a colorless oil, was found to contain trace water that was not removed under high vacuum. To remove this water, the oil was dissolved in anhydrous heptane and concentrated *in vacuo* three times, resulting in a colorless amorphous solid that was used in the double elimination reaction. An oven dried 100 mL round bottom flask containing a stir bar was evacuated and refilled with N_2 three times. The dried dibromide (700 mg, 2.24 mmol, 1 equiv) was taken up in anhydrous THF (25 mL, final concentration of 0.06 M) and added to the reaction flask under N_2. The flask was placed in an ice bath and cooled to 0 °C. Freshly sublimed potassium *tert*-butoxide (828 mg, 7.40 mmol, 3.3 equiv) was dissolved in 9 mL anhydrous THF and added dropwise over 10 minutes to the dibromide, resulting in an orange suspension. After 5 minutes of stirring at 0 °C, the reaction was placed in an oil bath and refluxed at 75 °C for two hours. The reaction was quenched with 20 mL of sat. NH_4Cl and the volatile solvent was removed using rotary evaporation. The suspension was extracted three times with 30 mL DCM and the combined organics were dried with brine, then Na_2SO_4. The solvent was removed using rotary evaporation and the crude product was purified using flash chromatography (1:1 EA/Hex) to yield the desired product **2** as a light yellow solid (198 mg, 59%). ^1H NMR spectra match previous reports and this material was used to form the carbonate BCN(exo)-PNP following previous literature.[2]

Alternate route to 5-hydroxy-α^1,α^1,α^3,α^3-tetramethyl-1,3-benzenedipropanoic acid (**5**):



Synthesis of diester **3**:

An oven dried 250 mL three-neck flask was evacuated and refilled with N_2 three times. Freshly distilled methylisobutyrate (8.2 mL, 71.5 mmol, 2.5 equiv) was added followed by 40 mL THF under a N_2 atomosphere. The solution was cooled to -78 °C and 34 mL LDA (2M in THF/hexane, 68 mmol, 2.4 equiv) was slowly added. The brown mixture was stirred at -78 °C for 10 minutes. In an oven dried pear-shaped flask, 5.0 g dichloroxylene (28.6 mmol, 1 equiv) was added and the atomosphere was purged with N_2. 10 mL THF was used to dissolve the dichloroxylene and the resulting solution was slowly cannula transferred to the reaction flask over 5 minutes. The reaction flask was stirred at -78 °C for 45 minutes then warmed to room temperature and stirred for 14 hours. The reaction was cooled in an ice bath before being carefully quenched with 150 mL of sat. NH_4Cl. The resulting mixture was extracted twice with 200 mL Et_2O. The combined organics were washed with brine and dried with Na_2SO_4 and the solvent was removed using rotary evaporation. The crude reaction mixture was purified using flash chromatography (10:1 Hexanes/EA) to yield a light yellow oil (7.12 g, 23.2 mmol, 81%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.12 (t, J = 7.6 Hz,

1H), 6.92 (d, J = 7.6 Hz, 2H), 6.83 (s, 1H), 3.64 (s, 6H), 2.80 (s, 4H), 1.14 (s, 12H).

Synthesis of phenol **4**:

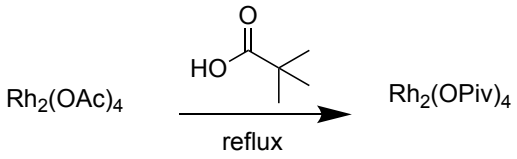
In a glovebox under N_2 atomosphere, 7.12 g of diester **3** (23.2 mmol, 1 equiv) was taken up in 20 mL anhydrous THF and added to an oven dried 250 mL Schlenk flask equipped with a stir bar. $[Ir(COD)(OMe)]_2$ (46 mg, 0.07 mmol, 0.003 equiv) was added in 2 mL THF and B_2pin_2 (4.11 g, 16.2 mmol, 0.7 equiv) was added as a solid, using 8 mL anhydrous THF to ensure complete transfer. 4,4'-Di-tert-butylbipyridine (37.6 mg, 0.14 mmol, 0.006 equiv) was taken up in 5 mL THF and added to the reaction last. The reaction vessel was sealed and heated to 80 °C for 48 hours. After completion of the reaction, the brown mixture was filtered through a silica plug, which was washed with 50 mL THF, before the solvent was removed by rotary evaporation. The crude product, a brown amorphous solid, was taken up in 250 mL MeOH and 50 mL of 30% H_2O_2 in water was added. The solution was stirred at room temperature for 2.5 hours, at which point the volatile solvent was evaporated and the remaining aqueous mixture was extracted 3 times with dichloromethane. The solvent was evaporated and the crude product was purified by flash chromatography (20:1 DCM:MeOH) to yield the desired product as a colorless oil (4.23 g, 13.1 mmol, 56% over two steps). ^{1}H NMR (400 MHz, $CDCl_3$) δ 6.42 (s, 2H), 6.39 (s, 1H), 3.64 (s, 6H), 3.47 (s, 6H), 2.74 (s, 4H), 1.14 (s, 12H).

Synthesis of diacid **5**:

4.23 g of phenol **4** (13.1 mmol, 1 equiv) was added to a 250 mL round bottom flask containing a solution of NaOH (3.14 g, 78.6 mmol, 6 equiv) in 65 mL 50% EtOH/H_2O. The flask was placed in an oil bath and the reaction was refluxed at 85 °C for 12 hours. The EtOH was removed by

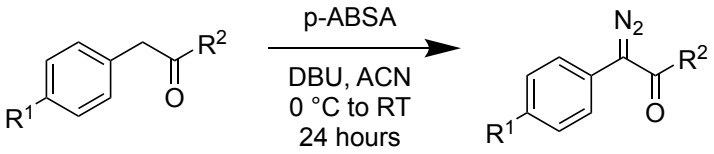
rotary evaporation and the remaining basic aqueous mixture was washed twice with DCM. The aqueous phase was acidifed to pH ~2 using 6M HCl, then extracted three times with DCM. The combined organics were washed with brine and dried with Na_2SO_4. Rotary evaporation yielded the product **5** as colorless crystals (3.55 g, 12.1 mmol, 92%). ^{1}H NMR analytic spectra match previous reports.[1]

Synthesis of Dihodium(II) tetrapivalate



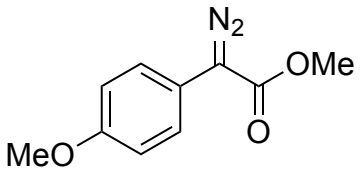
An oven dried flask containing a stir bar was evacatuate and refilled with N_2 three times. $Rh_2(OAc)_4$ (707 mg, 1.60 mmol, 1 equiv) and pivalic acid (13.6 g, 80 equiv) were added as solids and the mixture was heated to 80 °C for 3.5 hours. The pivalic acid melted rapidly to form a green solution and the reaction progress was monitored by thin layer chromatography until no rhodium tetraacetate remained. The pivalic acid was removed using rotary evaporation at 14 mmHg with heating at 50 °C until dry, leaving a green powder (694 mg, 1.14 mmol, 71%). ^{1}H NMR and ^{13}C NMR analytic spectra match previous reports.[3]

Synthesis of Donor-Acceptor Diazo Compounds

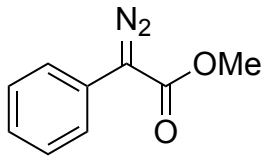


General Method: An oven-dried round bottom flask containing a 4-acetamidobenzenesulfonyl

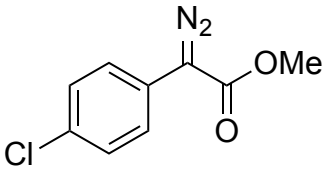
azide (1.5 equiv) was evacuated and refilled with N_2 three times. A solution of the ester substrate (1 equiv) in acetonitrile (final concentration of 0.1 M) was added via syringe under N_2. The reaction flask was placed into an ice bath and cooled to 0 °C. 1,8-diazabicylco[5.4.0]undec-7-ene (3 equiv) was added dropwise to the reaction mixture over thirty minutes. The reaction mixture was stirred at 0 °C for an additional 30 minutes then removed from the ice bath and stirred at room temperature for 24 hours. Over the course of the overnight reaction, a color shift to orange/yellow was observed. The solvent was evaporated using a rotary evaporator to yield a sticky orange oil. The crude product was purified using normal phase column chromatography (5-20% ethyl acetate/hexanes).



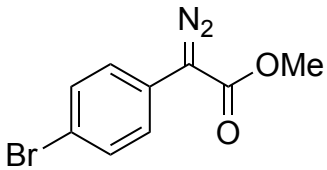
6



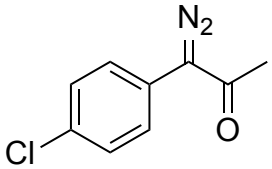
7



8



9



10

6, Methyl 2-diazo-2-(4-methoxyphenyl)acetate (2.219 mmol scale, yield: 178 mg, 39%) as orange solids; purified using 10% ethyl acetate/hexanes with silica gel. ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.38 (m, 2H), 6.94 (m, 2H), 6.84 (d, 1H), 3.85 (s, 3H), 3.81 (s, 3H). Characterization data matches those of previous reports.[4]

7, Methyl 2-diazo-2-phenylacetate (2.91 mmol scale, yield: 313 mg, 69%) as a yellow oil; purified using 10% ethyl acetate in hexanes with silica gel. ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.47 (d, J = 7.3

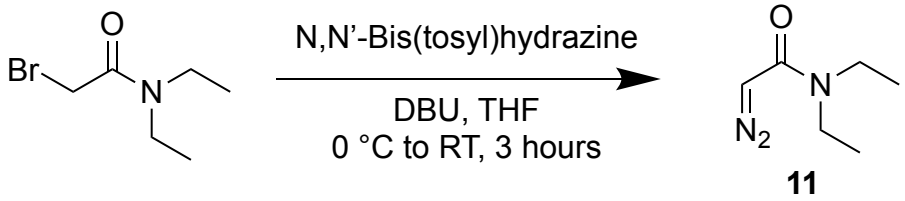
Hz, 2H), 7.39 (m, J = 7.4 Hz, 2H), 7.18 (t, J = 7.4 Hz, 1H), 3.85 (s, 3H). Characterization data matches those of previous reports.[5]

8, Methyl 2-(4-chlorophenyl)-2-diazoacetate (2.19 mmol scale, yield: 355 mg, 77%) as orange solids; purified by NH_4Cl and diethyl ether extraction followed by a brine wash; then a 10% ethyl acetate/hexanes mobile phase with silica gel column was run. ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.40 (d, J = 6.6 Hz, 2H), 7.33 (d, J = 6.6 Hz, 2H), 3.85 (s, 3H). Characterization data matches those of previous reports.[4]

9, Methyl 2-(4-bromophenyl)-2-diaozacetate (1.95 mmol scale, yield: 303 mg, 61%) as orange solids; purified using 10% ethyl acetate/hexanes with silica gel. ^{1}H NMR ($CDCl_3$, 400 MHz) δ 7.49 (d, J = 8.9 Hz, 2H), 7.37 (d, J = 8.9 Hz, 2H), 3.86 (s, 3H). Characterization data matches those of previous reports.[4]

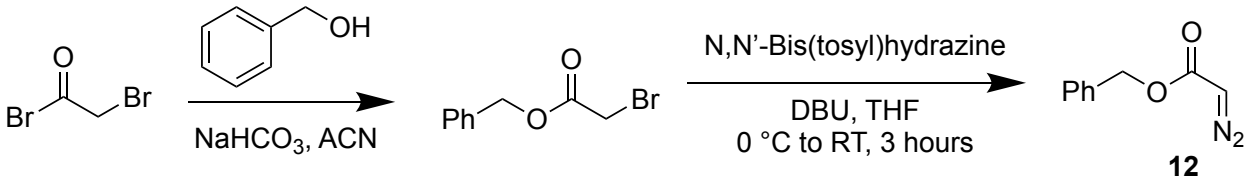
10, 1-(4-chlorophenyl)-1-diazopropan-2-one (2.38 mmol scale, yield: 148 mg, 39%) as yellow solids; purified using 15% ethyl acetate/hexanes with silica gel. Characterization data matches those of previous reports.[6]

Synthesis of Acceptor-Only Diazo Compounds



11

Diethyl diazoacetamide (11): An oven-dried 25 mL round bottom flask containing a stir bar was evacuated and refilled with N_2 three times. A solution of diethyl bromoacetamide (300 mg, 1.56 mmol, 1 equiv) and N,N'-Bis(p-toluenesulfonyl)hydrazine (1.064 g, 3.12 mmol, 2 equiv) in 7.8 mL of THF (0.2 M) was added to the flask under N_2. The flask was placed into an ice bath, and 1,8-diazabicylco[5.4.0]undec-7-ene (465 µL, 3.12 mmol, 2 equiv) was added dropwise over 30 minutes. The reaction was warmed to room temperature and stirred for three hours. Reaction progress was monitored by thin layer chromatography until all starting material was consumed. The solvent was removed by rotary evaporation and the residue extracted by diethyl ether/saturated sodium bicarbonate four times. The combined organic extracts were dried using $MgSO_4$ and filtered through cotton. The organic phase was evaporated to yield the diazo **11** as an orange oil (79 mg, 0.56 mmol, 36% yield). ^{1}H NMR (400 MHz, $CDCl_3$): δ 4.90 (s, 1H), 3.25 (m, 4H), 1.13 (t, 6H). Characterization data matches those of previous reports.[7]



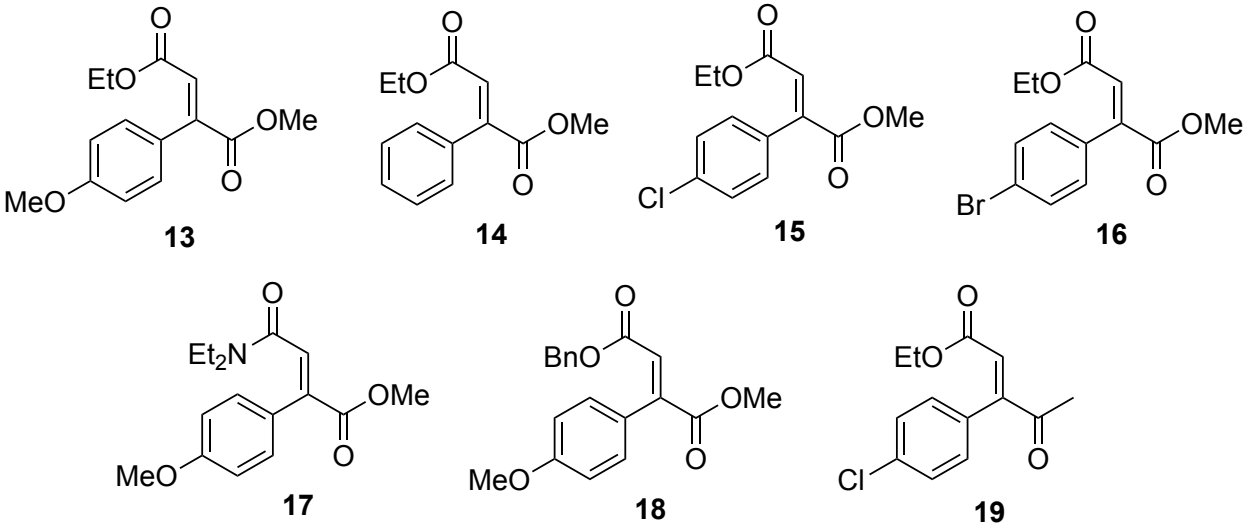
12

Benzyl bromoacetate: To an oven-dried 50 mL round bottom flask containing a stir bar was added sodium bicarbonate (892 mg, 10.59 mmol, 3 equiv), and the flask was evacuated and refilled with N_2 twice. A solution of benzyl alcohol (368 μL, 3.53 mmol, 1 equiv) in acetonitrile (final concentration of 0.2 M) was added, and the reaction mixture was placed into an ice bath. Bromo acetylbromide (1.00 g, 4.96 mmol, 1.4 equiv) was added to the reaction mixture dropwise over 15 minutes. The reaction was allowed to proceed at room temperature until thin layer chromatography showed no remaining benzyl alcohol (about 45 minutes). Water was added to the crude product, and the reaction mixture was extracted twice with dichloromethane. The combined organics were dried with Na_2SO_4 and filtered. The solution was concentrated using a rotary evaporator then purified using normal phase column chromatography to yield a colorless oil (674 mg, 2.95 mmol, 60% yield).

Benzyl diazoacetate (12): An oven-dried round bottom flask was evacuated and refilled with N_2 three times, then N,N'-Bis(toluenesulfonyl)hydrazine (2 equiv, 894 mg, 2.63 mmol) and benzyl bromoacetate (1 equiv, 300 mg, 1.315 mmol) were added in tetrahydrofuran (final concentration of 0.2 M). The flask was placed into an ice bath and 1,8-diazabicylco[5.4.0]undec-7-ene (2 equiv, 392 μL, 2.63 mmol) was added dropwise over 30 minutes. The reaction was monitored by thin layer chromatography until no bromoacetate remained, approximately 1 hour. Saturated sodium bicarbonate was used to quench the reaction, which was then extracted with diethyl ether (3x). The organic phases were combined and washed with brine then dried with Na_2SO_4, filtered, then

evaporated. The crude product was purified using normal phase column chromatography to give benzyl diazoacetate **8** (72 mg, 0.408 mmol, 31% yield). ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.35(m, 5H), 5.19 (s, 2H), 4.83 (s, 1H). Characterization data matches those of previous reports.[8]

Synthesis of Authentic Alkenes



General Method: An oven-dried round bottom flask containing a magnetic stir bar was evacuated and refilled with N_2 three times. Dried dichloromethane (80% of total amount, final substrate concentration of 0.2 M) was used to dissolve the dirhodium catalyst (1 mol%) then added to the reaction flask under N_2 via syringe. The remaining solvent was used to dissolve an equimolar mixture of the two diazo compounds (1 equiv each). The reaction mixture was cooled to -78 °C and the solution of diazo compounds was added dropwise over an hour. The reaction was stirred for an hour at -78°C before being allowed to warm to room temperature and stirred for 30 minutes. The reaction was concentrated using a rotary evaporator and purified using normal phase column chromatography.

13-E, 4-ethyl 1-methyl 2-(4-methoxyphenyl)fumarate (4.14 mmol scale, yield: 317 mg, 29%); prepared using $Rh_2(OAc)_4$, purified using 10% ethyl acetate/hexanes and silica gel. Characterization data matches those of previous reports.[9]

13-Z, 4-ethyl 1-methyl 2-(4-methoxyphenyl)maleate (4.14 mmol scale, yield: 42 mg, 4%); prepared using $Rh_2(OAc)_4$, purified using 10% ethyl acetate/hexanes and silica gel; Characterization data matches those of previous reports.[9]

14, 4-ethyl 1-methyl 2-phenylfumarate (1.05 mmol scale, yield: 100 mg, 40%); purified using 10% ethyl acetate in n-hexanes and silica gel; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.35 (t, J = 4.2 Hz, 3H), 7.22 (m, 2H), 7.00 (s, 1H), 4.02 (q, J = 7.1 Hz, 2H), 3.79 (s, 3H), 1.04 (t, J = 7.1 Hz, 3H); Characterization data matches those of previous reports.[9]

15, 4-ethyl 1-methyl 2-(4-chlorophenyl)fumarate (0.95 mmol scale, yield: 87 mg, 34%); purified using 5% ethyl acetate in n-hexanes with silica gel; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.32 (d, J = 8.5 Hz, 2H), 7.16 (d, J = 8.5 Hz, 2H), 4.05 (q, J = 7.1 Hz, 2H), 7.01 (s, 1H), 3.78 (s, 3H), 1.09 (t, J = 7.1 Hz, 3H); Characterization data matches those of previous reports.[10]

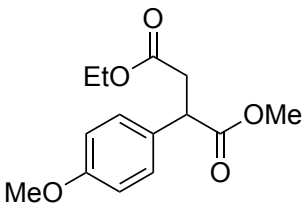
16, 4-ethyl 1-methyl 2-(4-bromophenyl)fumarate (1.18 mmol scale, yield: 165 mg, 44%); purified using 10% ethyl acetate in n-hexanes with silica gel; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.48 (d, J = 8.5 Hz, 2H), 7.10 (d, J = 8.5 Hz, 2H), 7.02 (s, 1H), 4.05 (q, J = 7.1 Hz, 2H), 3.79 (s, 3H), 1.09 (t, J = 7.1 Hz, 3H). Characterization data matches those of previous reports.[9]

17, Methyl (E)-4-(diethylamino)-2-(4-methoxyphenyl)-4-oxobut-2-enoate (0.44 mmol scale, yield: 73 mg, 56%); purified using 25% ethyl acetate in n-hexanes with silica gel; ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.27 (d, *J* = 8.8 Hz, 2H), 7.14 (s, 1H), 6.84 (d, *J* = 8.8 Hz, 2H), 3.79 (s, 3H), 3.78 (s, 3H), 3.29 (q, *J* = 7.1 Hz, 2H), 3.07 (q, *J* = 7.2 Hz, 2H), 0.93 (t, *J* = 7.1 Hz, 3H), 0.87 (t, *J* = 7.2 Hz, 3H); Characterization data matches those of previous reports.[10]

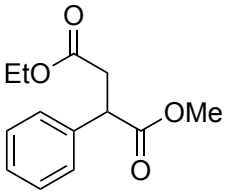
18, 4-benzyl 1-methyl 2-(4-methoxyphenyl)fumarate (0.40 mmol scale, yield: 78 mg, 59%); purified using 10% ethyl acetate in n-hexanes with silica gel; ^{1}H NMR (400 MHz, CD_3CN) δ 7.35 – 7.29 (m, 3H), 7.16 (m, 4H), 6.92 (s, 1H), 6.86 (d, *J* = 8.7 Hz, 2H), 5.04 (s, 2H), 3.80 (s, 3H), 3.75 (s, 3H); ^{13}C NMR in CD_3OD δ 167.08, 165.69, 160.13, 143.00, 135.34, 130.18 (2C), 127.99 (2C), 127.95 (2C), 127.81, 127.70, 112.90 (2C), 66.28, 54.31, 51.85.

19, Ethyl (E)-3-(4-chloroyphenyl)-4-oxopent-2-enoate (0.77 mmol scale, yield: 77 mg, 37%); purified using 10% ethyl acetate in n-hexanes with silica gel; ^{1}H NMR (400 MHz, $CDCl_3$) δ δ 7.34 (d, *J* = 8.1 Hz, 2H), 7.09 (d, *J* = 8.0 Hz, 2H), 6.75 (s, 1H), 4.05 (q, *J* = 7.1 Hz, 2H), 2.32 (s, 3H), 1.09 (t, *J* = 7.1 Hz, 3H); ^{13}C NMR in $CDCl_3$ δ 198.33, 165.21, 156.83, 136.91, 131.46, 130.05, 128.18, 116.1, 61.3, 30.36, 14.02. Characterization data matches those of previous reports.[11]

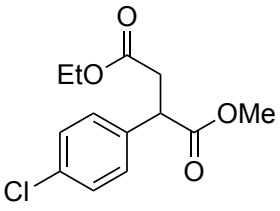
Synthesis of Authentic Alkanes



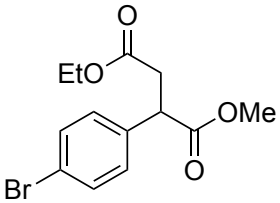
20



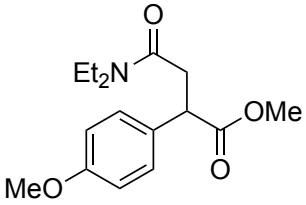
21



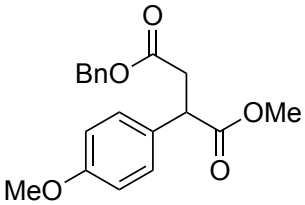
22



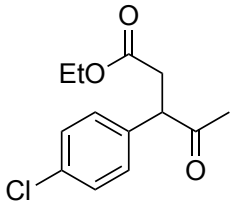
23



24



25



26

General Reduction of Alkenes: To an oven-dried round-bottom flask with magnetic stir bar was added palladium on activated carbon (0.33 equiv). The flask was then evacuated and refilled with N_2 three times. The alkene (1 equiv) was dissolved in 70% EtOH (0.5 M) and added via syringe. H_2 was added at atmospheric pressure using a balloon. The reaction was let stir overnight. Conversion was confirmed by thin layer chromatography.

Synthesis of 25: A separate method was necessary for this preparation as reduction with H_2/Pd would deprotect the benzylic ester in addition to reducing the alkene.

An oven dried three-necked round bottom flask containing a stir bar was purged and refilled with N_2 three times. The flask was then cooled to -78 °C, and a solution of methyl (4-methoxy)phenylacetate (59 mg, 0.327 mmol, 1.5 equiv) in anhydrous THF (0.1 M) was added. Lithium diisopropylamide (2M in THF/hexane, 0.109 mL, 0.218 mmol, 1 equiv) was then added, and the reaction mixture was stirred at -78 °C for 15 minutes. Benzyl diazoacetate (**12**) (50 mg, 0.218 mmol, 1 equiv) was dissolved in 100 µL anhydrous THF and added to the reaction. After 5 minutes, the flask was allowed warm to room temperature with stirring overnight. The reaction

mixture was quenched with MeOH then saturated NH_4Cl, and the mixture was extracted with dichloromethane twice. The combined organics were evaporated then purified using normal phase column chromatography to yield a colorless oil.

20, 4-ethyl 1-methyl 2-(4-methoxyphenyl)succinate (0.05 mmol, yield: 8.7 mg, 65%); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.13 (d, *J* = 8.7 Hz, 2H), 6.78 (d, *J* = 8.7 Hz, 2H), 4.05 (q, *J* = 7.1 Hz, 2H), 3.96 (dd, *J* = 9.9, 5.5 Hz, 1H), 3.09 (dd, *J* = 16.8, 10.0 Hz, 1H), 2.57 (dd, *J* = 16.8, 5.5 Hz, 1H), 1.15 (t, *J* = 7.1 Hz, 2H); ^{13}C NMR in $CDCl_3$ δ 173.1, 171.52, 159.04, 129.75, 128.80 (2C), 114.23 (2C), 60.71, 55.26, 52.25, 46.28, 37.99, 14.13.

21, 4-ethyl 1-methyl 2-phenylsuccinate (0.06 mmol scale, yield: 8.1 mg, 57%); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.31-7.22 (m, 5H), 4.10 (q, *J* = 7.2 Hz, 2H), 4.06 (dd, *J* = 10.3, 5.6 Hz, 1H), 3.65 (s, 3H), 3.17 (dd, *J* = 16.9, 10.1 Hz, 1H), 2.64 (dd, *J* = 16.9, 5.4 Hz, 1H), 1.20 (t, *J* = 7.1 Hz, 3H). Characterization data matches those of previous reports.[11]

22, 4-ethyl 1-methyl 2-(4-chlorophenyl)succinate (0.20 mmol scale, yield: 38 mg, 68%); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.35 – 7.27 (m, 5H), 4.17 – 4.05 (m, 4H), 3.68 (s, 3H), 3.19 (dd, *J* = 16.8, 10.2 Hz, 2H), 2.66 (dd, *J* = 16.8, 5.3 Hz, 1H), 1.22 (t, *J* = 7.1 Hz, 3H). ^{13}C NMR in CD_3OD δ 173.69, 171.67, 137.80, 128.48(2C), 127.42(2C), 127.29, 60.39, 51.30, 47.07, 37.41, 13.05. Characterization data matches those of previous reports.[11]

23, 4-ethyl 1-methyl 2-(4-bromophenyl)succinate (0.21 mmol scale, yield: 33 mg, 50%); ^{1}H NMR in $CDCl_3$ δ 7.35 – 7.25 (m, 4H), 4.16 – 4.03 (m, 4H), 3.66 (s, 3H), 3.17 (dd, *J* = 16.9, 10.1 Hz,

1H), 2.64 (dd, *J* = 16.9, 5.3 Hz, 1H), 1.21 (t, *J* = 7.1 Hz, 3H); ^{13}C NMR in $CDCl_3$ 173.47, 171.50, 137.71, 128.88(2C), 127.79(2C), 127.67, 60.84, 52.44, 47.18, 37.97, 14.18.

24, methyl 4-(diethylamino)-2-(4-methoxyphenyl)-4-oxobutanoate (0.05 mmol scale, yield: 10 mg, 68%); ^{1}H (400 MHz, $CDCl_3$) $CDCl_3$ δ 7.23 (d, *J* = 8.7 Hz, 2H), 6.85 (d, *J* = 8.7 Hz, 2H), 4.17 (dd, *J* = 10.6, 4.4 Hz, 1H), 3.78 (s, 3H), 3.67 (s, 3H), 3.40 (q, *J* = 7.0 Hz, 1H), 3.35-3.22 (m, 3H), 3.18 (dd, *J* = 16.1, 10.5 Hz, 1H), 2.53 (dd, *J* = 16.1, 4.3 Hz, 1H), 1.16 (t, *J* = 7.1 Hz, 3H), 1.08 (t, *J* = 7.1 Hz, 3H); ^{13}C NMR in $CDCl_3$ δ 174.49, 169.68, 158.91, 130.69, 128.85(2C), 114.19(2C), 55.27, 52.16, 46.60, 41.81, 40.27, 37.61, 14.13, 13.05.

25, 4-benzyl 1-methyl 2-(4-methoxyphenyl)succinate was prepared using the alternate synthetic route (0.218 mmol scale, yield: 26 mg, 37%); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.41 – 7.28 (m, 5H), 7.21 (d, *J* = 8.7 Hz, 2H), 6.86 (d, *J* = 8.7 Hz, 2H), 5.13 (q, *J* = 12.3 Hz, 2H), 4.07 (dd, *J* = 9.8, 5.7 Hz, 1H), 3.81 (s, 3H), 3.66 (s, 3H), 3.24 (dd, *J* = 16.9, 9.8 Hz, 1H), 2.73 (dd, *J* = 16.8, 5.7 Hz, 1H); ^{13}C NMR (126 MHz, $CDCl_3$) δ 173.60, 171.34, 159.09, 135.72, 129.62, 128.82, 128.54, 128.24, 128.20, 114.27, 66.53, 55.26, 52.28, 46.31, 38.01.

26, ethyl 3-(4-chlorophenyl)-4-oxopentanoate (0.067 mmol scale, yield: 8.3 mg, 49%); ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.45 – 6.92 (m, 4H), 4.18 (dd, *J* = 9.8, 5.1 Hz, 1H), 4.10 (dq, *J* = 7.1, 3.6 Hz, 2H), 3.20 (dd, *J* = 16.9, 9.7 Hz, 1H), 2.52 (dd, *J* = 16.9, 5.1 Hz, 1H), 2.11 (s, 3H), 1.22 (t, *J* = 7.1 Hz, 3H); ^{13}C NMR (126 MHz, $CDCl_3$) δ 206.83, 172.09, 137.47, 129.15, 128.25, 127.73, 60.66, 54.89, 37.05, 28.90, 14.13.

3.4.3 Overexpression and purification of proteins

Overexpression and purification of *Pfu*POP scaffolds

BL21 cells harboring pET28a-POP variants were induced and overexpressed in 1L of 2x YT medium containing AzF as previously described.[12] The cell pellet was collected by centrifugation at 3,600 rpm for 10 min, resuspended in equilibration buffer consisting of 20 mM phosphate buffer (pH 7.4), 300 mM NaCl, and 10 mM imidazole, then lyzed using sonication. The lysate was treated at 75 °C for 15 min to precipitate *E. coli* proteins. After centrifugation at 15,000 rpm for 30 min, the supernatant was bound to 2 mL of Ni-NTA resin. The column was washed with 10-column volumes (i.e., 20 mL) of wash buffer consisting of 50 mM phosphate buffer (pH 7.4), 300 mM NaCl, and 25 mM imidazole. Six-column volumes (i.e., 12 mL) of elution buffer containing 250 mM imidazole was used to elute the protein. The protein concentration was measured using Bradford assay. To standardize concentrations during the buffer exchange process, the protein concentration was diluted to 12.3 μM using 50 mM Tris-HCl (pH 7.4) and 15 mL of the diluted protein solution was added to a 20 mL centrifugation filter and concentrated to 0.5 mL by centrifugation at 3,600 rpm for ~20 min. The flow through was discarded, then 20 mL of Tris-HCl buffer was added to the spin filter, and the protein solution was concentrated to ~0.5 mL by centrifuge. Three rounds of buffer exchange (~60,000-fold total exchange) were performed by this method, after which the protein sample was transfered into a 1.5 mL microcentrifuge tube. Any insoluble scaffold present after buffer exchange was precipitated by centrifugation at 15,000 rpm for 2 mins. The supernatant was decanted into a fresh 1.5 mL microcentrigue tube and the protein concentration was measured using UV absorption at 280 nm. Samples were diluted to 75 μM with 50 mM Tris•HCl (pH 7.4) and 480 μL aliquots were prepared for bioconjugation (details in section S5). Any unused aliquots were snap frozen in LN_2 and stored at -80 °C.

Overexpression and purification of alkene reductases

Cells containing the target ERs were incubated in TB media with 50 μg/mL of the relevant antibiotics at 37 °C and 250 rpm. IPTG (final concentration of 1 mM) was added to induce expression at OD_{600} = ~0.8. After overnight induction at 37 °C with 250 rpm shaking, the cell pellet was collected by centrifugation at 3,600 rpm for 10 min, resuspended in equilibration buffer consisting of 50 mM phosphate buffer (pH 7.4), 300 mM NaCl and 10 mM imidazole, then lysed by sonication on ice. The lysated was pelleted by centrifugation at 10,000 rpm for 30 minutes then the supernatant was added to Ni-NTA resin and the column was washed with 10-column volumes of 25 mM imidazole. Elution with 5-column volumes of 250 mM imidazole provided the purified ER. To maximize the flavin cofactor occupancy in the ER, a final concentration of 0.1 mM FMN was added the eluted protein sample. After incubation at room temperature for 5 min, the protein sample was buffer exchanged (>10,000-fold total exchange) into 50 mM Tris-HCl, pH 7.4. The protein concentration was measured using the standard Bradford assay protocol. The ER stock solution was snap frozen in LN_2 and stored at -80° C for future use if necessary.

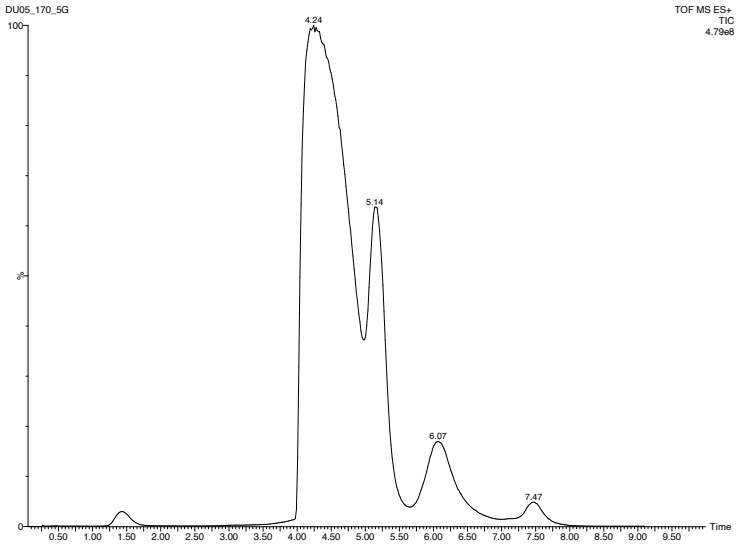
3.4.2 ArM Bioconjugation

General method: A method with minor improvements to previously reported[12] dirhodium bioconjugation was used. The changes include a shorter cofactor incubation time (1 hour instead of overnight) and a standardized buffer exchange protocol.

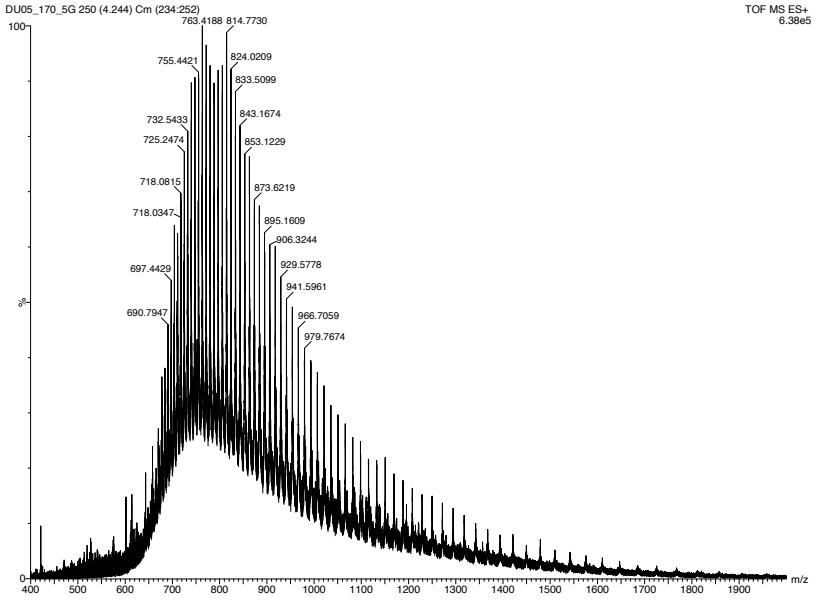
The bioconjugation of **1** to POP was performed by adding 120 μL of the cofactor **1** stock solution (830 μM (0.655 mg/mL) in ACN) to the scaffold solution (480 μL of 75 μM protein in 50 mM Tris•HCl, pH 7.4) while shaking at 4 °C, 750 rpm. The mixture was incubated at 4 °C, 750 rpm for 1 hour, at which point the shaking was stopped and 100 μL of a 50% suspension of N_3-

sepharose resin solution was added to scavenge the free cofactor. The mixture was rotated at 4°C overnight (12-16 hours). After the resin purification was complete, the resin was pelleted by centrifugation at 1,000 rpm for 2 minutes and the supernatant was buffer exchanged into 50 mM PIPES, pH 7.4 with 0.7 M NaBr using the same buffer exchange method detailed in section S4. The protein concentration of the ArM sample was measured by using UV analysis at 280 nm. 50 uL of 5 uM ArM was desalted using Zeba desalting columns and used for mass spectrometric analysis while the remaining ArM was used for catalysis. Intact protein mass spectrometry was used to analyze the extent of bioconjugation in order to determine the concentration of ArM for catalysis. A 10 minute LC method (A: H_2O with 0.1% formic acid, B = Acetonitrile with 0.1% formic acid) with a linear gradient from 95% A to 1% A over 6 minutes followed by a 4 minute flush at 95% A was used with mass spec recording between 400-2000 Da, providing an LC trace like the one shown below (ArM retention time = 4.24 mins). Deconvolution of the mass spectrum was performed using the 800-1300 Da mass window with a deconvoluted mass range of 71-74 kDa.

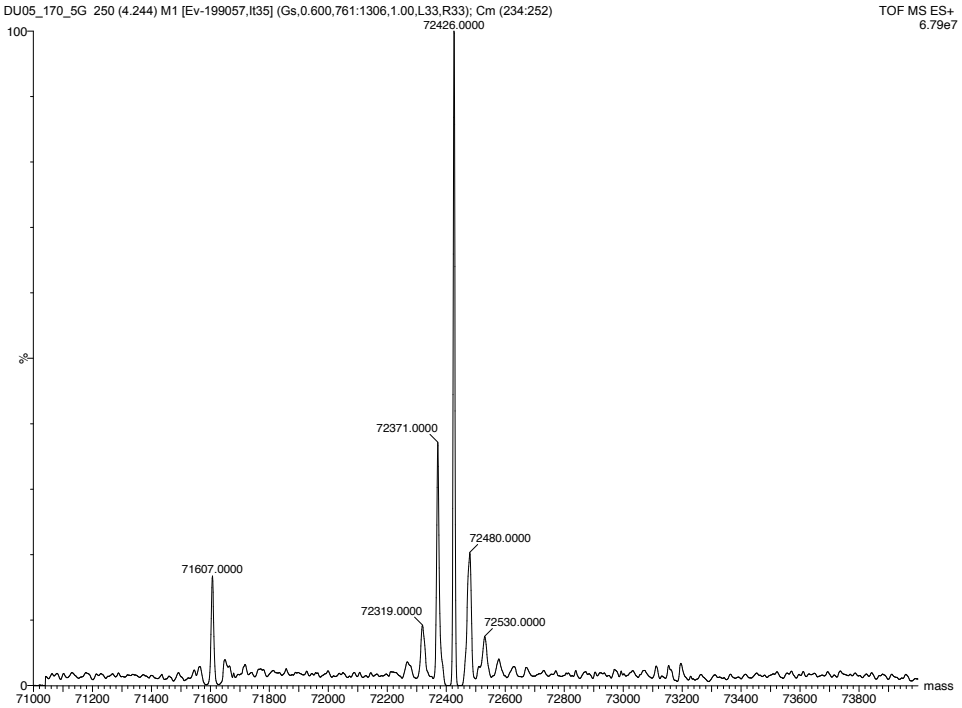
Representative LC trace of 5-G after bioconjugation:



Representative Mass Spectrum of 5-G after bioconjugation:

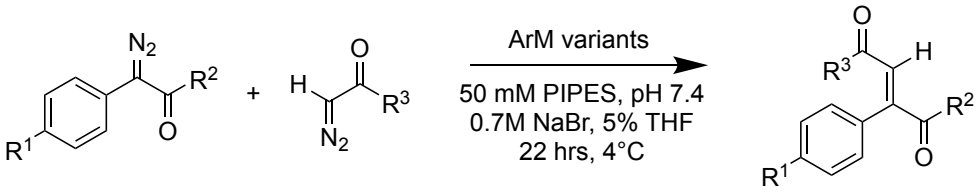


Representative deconvoluted mass of 5-G (>86% bioconjugation):



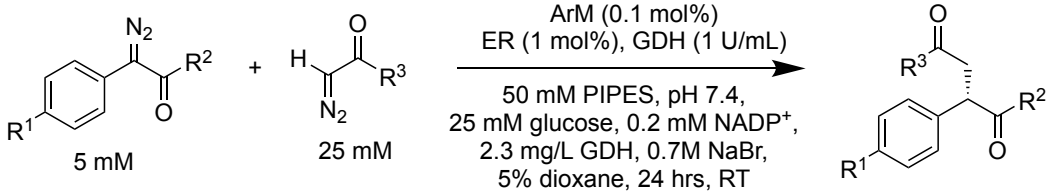
3.4.5 ArM Biocatalysis Method

General ArM-catalyzed diazo coupling method



A 285 µL solution of 5 µM ArM (corrected by the known bioconjugation efficiency, typically 90%) in 50 mM PIPES, pH 7.4 with 0.7 M NaBr was prepared in a 1.5 mL microcentrifuge tube. While shaking at 750 rpm at 4°C, 15 µL of the diazo stock solution (0.1 M donor-acceptor diazo, 0.5 M acceptor-only diazo) in THF was added to obtain a final concentration of 5 mM donor-acceptor diazo and 25 mM acceptor-only diazo. The reactions were sealed and briefly inverted to assist with mixing of the substrate, then placed in a thermomixer and shaken at 750 rpm at 4°C for 22 hours. The reaction was quenched by the addition of 50 µL internal standard (30 mM 1,3,5-trimethoxybenzene in ethyl acetate) and 100 µL ethyl acetate. The biphasic mixture was vortexed thoroughly then centrifuged at 20,000 xg for 1 minute to separate the phases. 100 µL of the upper organic layer was collected and filtered for LC analysis.

General cascade reaction method



The 300 µL reactions were carried out in 50 mM PIPES, pH 7.4 with 0.7 M NaBr. The reagent mixture includes 7.5 µL of glucose (1M), 0.6 µL of $NADP^+$ (100 mM), 0.68 µL of GDH-105 (1 mg/mL). Buffer, ArM and ER were combined to reach 285 µL at final concentrations of 5 µM

ArM, 25 mM glucose, 0.2 mM $NADP^+$, 0.05 mM ER and 2.3 mg/L GDH-105. While shaking at 750 rpm at 4°C, 15 μL of a diazo stock solution (0.1 M donor-acceptor diazo, 0.5 M acceptor-only diazo) in dioxane was added to obtain a final concentration of 5 mM donor-acceptor diazo and 25 mM acceptor-only diazo. The reactions were sealed and briefly inverted to assist with mixing of the substrate, then placed in a thermomixer and shaken at 750 rpm at 4°C for 1 hour, at which point the reaction temperature was raised to room temperature for 23 hours. The reaction was quenched by the addition of 50 μL internal standard (30 mM 1,3,5-trimethoxybenzene in ethyl acetate) and 100 μL pure ethyl acetate. The biphasic mixture was vortexed thoroughly then centrifuged at 20,000 xg for 1 minute to separate the phases. 100 μL of the upper organic layer was collected and filtered for LC analysis.

3.4.6 Library Cloning and Screening Method

Primer sequences

#	Primer name	Primer sequence
1	CCM insert F	CGTGAATTTATTGGTGAACTGAGTGATAAACTGTTTCCGGAAGTGTGGGAAC
2	CCM insert R	GGTCATCTTTAAATTCGTACAGACGATACGGAATCGTAAACGAGGTGTAG
3	CCM vect F	GTTCCCACACTTCCGGAAACAGTTTATCACTCAGTTCACCAATAAATTCACG
4	CCM vect R	CTACACCTCGTTTACGATTCCGTATCGTCTGTACGAATTTAAAGATGACC
5	98 BsaI NNK F	CATCGGTCTCCATCACCAACTTCACGTTCCAGTTCTTTAC
6	98 BsaI NNK R	CATCGGTCTCGTGATGAAGTCCTGCTGNNKAGCTTTACCACGGACGCGGAAG
7	98-101 SOE_F	CAGGACTTCATCACCAACTTCACGTTCCAGTTCTTTAC
8	98P/99NNK SOE_R	GAAGTTGGTGATGAAGTCCTGCTGCCGNNKTTTACCACGGACGCGGAAGGTG
9	98P/99H/100NNK SOE R	GAAGTTGGTGATGAAGTCCTGCTGCCGCATNNKACCACGGACGCGGAAGGTG
10	101 NNK R	GAAGTTGGTGATGAAGTCCTGCTGCAGAGCTTTNNKACGGACGCGGAAGGTG
11	283NNK SOE F	GAATTTCCGCTGNNKTGGGCAGTCATT
12	283NNK SOE R	AATGACTGCCCAMNNCAGCGGAAATTC
13	142NNK SOE F	AAACCGTCCATTNNKAACATCACCGCC
14	142NNK SOE R	GGCGGTGATGTTMNNAATGGACGGTTT
15	143NNK SOE F	CCGTCCATTTGGNNKATCACCGCCCTG
16	143NNK SOE R	CAGGGCGGTGATMNNCCAAATGGACGG
17	209NNK SOE F	TTCGCAATCGTTNNKCTGACGTATGGT
18	209NNK SOE R	ACCATACGTCAGMNNAACGATTGCGAA
19	241NNK SOE F	TCTGTGCCGGTTNNKGCTATTGACGTC
20	241NNK SOE R	GACGTCAATAGCMNNAACCGGCACAGA
21	338NNK SOE F	CGTGTTCTGCTGNNKTACACCTCGTTT
22	338NNK SOE R	AAACGAGGTGTAMNNCAGCAGAACACG

Note: The forward and reverse NDT primer pool used for the CCM library of POP was constructed as described previously.[12]

Library construction

A site saturation mutagenesis library of residue 98 in *pfu*POP 1-SGH was constructed using Golden gate assembly. The 50 µL of PCR mixture was prepared as follows: PrimeSTAR Max Premix (2x) 25 µL; primer 98 *Bsa*I NNK F (100 µM) 0.15 µL; primer 98 *Bsa*I NNK R (100 µM) 0.15 µL; plasmid pET28a-POP 1-SGH (132 ng/µL) 0.15 µL; water to reaction volume of 50 µL. The PCR reaction was conducted as followed: 1 cycle of 98 °C, 2 min; 23 cycles of 98 °C, 10 s; 62 °C, 15 s; 72 °C 3 min 30s; 1 cycle of 72 °C 5 min, 4 °C 10 min. The PCR product was separated by 1% agarose gel electrophoresis. The agarose region containing the desired PCR product were excised and purified by using QIAquick Gel Extraction Kit (Qiagen). The PCR product was digested using *Bsa*I and *Dpn*I at 37 °C for 3 hours using the manufacturer's protocol. The digested product was cleaned up using QIAquick Gel Extraction Kit (Qiagen) and ligated at room temperature for 10 min by using the manufacturer's protocol of Quick Ligation™ Kit (New England Biolabs). 20 µL of ligation product was added to 200 µL of chemical BL21(DE3)-pEVOL competent cell for transformation. The mixture was incubated on ice for 30 minutes, heat shocked at 42 °C for 25 seconds, then placed on ice for 3 minutes. 1 mL 2x YT medium was added to the vial and the cells were incubated at 37 °C for 1 hour. Cells were spread on 1.5% 2x YT agar plate with antibiotics (50 µg/mL kanamycin and 50 µg/mL chloramphenicol) and incubated at 37 °C for 12-16 hours.

The site saturation mutant libraries of residues 99, 100, 101, 142, 143, 209, 241, 283, and 338 were constructed using splicing by overhang extension (SOE) PCR[13] to avoid DNA fragment insertion or deletion observed when using Golden gate assembly-based techniques towards the relevant libraries. In the site saturation mutant library of residue 101, a 50 µL of PCR mixture was prepared for Fragment 1 as follows: PrimeSTAR Max Premix (2x) 25 µL; primer CCM insert F

(100 μM) 0.3 μL; primer 98-101 SOE_F (100 μM) 0.3 μL; plasmid pET28a-POP 1-SGH (~100 ng/μL) 0.3 μL; water to reaction volume of 50 μL. A 50 μL of PCR mixture was prepared for Fragment 2 as follows: PrimeSTAR Max Premix (2x) 25 μL; primer CCM insert R (100 μM) 0.3 μL; primer 101 NNK R (100 μM) 0.3 μL; plasmid of pET28a-POP 1-SGH (~100 ng/μL) 0.3 μL; water to reaction volume of 50 μL. The PCR reactions for fragments 1 and 2 were conducted as followed: 1 cycle of 98 °C, 2 min; 25 cycles of 98 °C, 10 s; 62 °C, 15 s; 72 °C 15 s; 1 cycle of 72 °C 5 min, 4 °C 10 min. The PCR product was gel purified as described above. The purified DNA fragments were mixed in a 50 μL of PCR mixture for gene assembly: PrimeSTAR Max Premix (2x) 25 μL; primer CCM insert F (100 μM) 0.3 μL; primer CCM insert R (100 μM) 0.3 μL; Fragment 1 30 ng; Fragment 2 70 ng; water to reaction volume of 50 μL. The assembly PCR reaction was conducted as follows: 1 cycle of 98 °C, 2 min; 25 cycles of 98 °C, 10 s; 62 °C, 15 s; 72 °C 15 s; 1 cycle of 72 °C 5 min, 4 °C 10 min. Gel purification was used to obtain the purified DNA for sub-cloning.

The site saturation mutant libraries of residues 99, 100, 142, 143, 209, 241,283, and 338 were constructed using SOE-PCR using the same protocol with the following primer and plasmid changes. The template plasmid used in the site saturation library of residue 99 was pET28a-POP 2-P with the primer 98P/99NNK_SOE_R used for Fragment 1 amplification. The template plasmid used in the site saturation library of residue 100 was pET28a-POP 3-H with the primer 98P/99H/100NNK SOE R used for Fragment 1 amplification. The template plasmid used in the site saturation library of residue 283 was pET28a-POP 4-G with the primer 283NNK SOE F used for Fragment 1 amplification and the primer 283NNK SOE R used for fragment 2 amplification. The template plasmid used in the site saturation library of residues 142, 143, 209, 241, and 338 was pET28a-POP 4-G with the relevant NNK SOE F primer used for Fragment 1 amplification

and the relevant NNK SOE R primer used for fragment 2 amplification.

A combinatorial codon mutagenesis (CCM) library of 3-H was constructed following the previously published protocol.[14] A 50 μL PCR mixture was prepared for inserted gene fragement as follows: PrimeSTAR Max Premix (2x) 25 μL; primer CCM insert F (100 μM) 0.3 μL; primer CCM insert R (100 μM) 0.3 μL; plasmid pET28a-POP 3-H (~100 ng/μL) 0.3 μL; water to reaction volume of 50 μL. The PCR reaction was conducted as follows: 1 cycle of 98 °C, 2 min; 25 cycles of 98 °C, 10 s; 62 °C, 15 s; 72 °C 15 s; 1 cycle of 72 °C 5 min, 4 °C 10 min. The PCR product was gel purified used in the next step. A forward and a reverse NDT primer pool of POP gene were constructed as described previously.[14] A 30 μL of PCR mixture was prepared for the forward fragmentation: KOD hot start premix (2x) 15 μL; forward mutagenic primer pool (4.5 μM) 2 μL; primer CCM insert R (4.5 μM) 2 μL; PCR product POP 3-H (3 ng/μL) 4 μL; water to reaction volume of 30 μL. The forward fragementation PCR reaction was conducted using the following method: 1 cycle of 95 °C, 2 min; 14 cycles of 95 °C, 20 s; 70 °C, 1 s; 50 °C for 30 seconds (cooling to 50 °C at 0.5 °C per second); 70 °C 40 s; 1 cycle of 4 °C 10 min. The PCR reaction for the reverse fragmentation was completed in the same manner, using the reverse primer pool and CCM insert R. A 30 μL of PCR mixture was then prepared for the joining reaction: KOD hot start premix (2x) 15 μL; forward fragment 1 μL; reverse fragment 1 μL; primer CCM insert F (4.5 μM) 2 μL; primer CCM insert R (4.5 μM) 2 μL; water to reaction volume of 30 μL. The PCR reaction of joining reaction was conducted: 1 cycle of 95 °C, 2 min; 20 cycles of 95 °C, 20 s; 70 °C, 1 s; 50 °C for 30 seconds (cooling to 50 °C at 0.5 °C per second); 70 °C 40 s; 1 cycle of 4 °C 10 min.

Prolonged overlap extension-PCR (POE-PCR)[15,16] was used to sub-clone the mutant library into the vector harboring the remaining POP gene. A 50 μL of PCR mixture was prepared for the vector as follows: PrimeSTAR Max Premix (2x) 25 μL; primer CCM vect F (100 μM) 0.3

μL; primer CCM vect R (100 μM) 0.3 μL; plasmid pET28a-POP 1-SGH (~100 ng/μL) 0.3 μL; water to reaction volume of 50 μL. The PCR reaction was conducted as follows: 1 cycle of 98 °C, 2 min; 25 cycles of 98 °C, 10 s; 62 °C, 15 s; 72 °C 3 min 30 s; 1 cycle of 72 °C 5 min, 4 °C 10 min. The PCR product was gel purified as described above.

To assemble the insertion fragment and vector fragment, a 50 μL of PCR mixture was prepared as follows: PrimeSTAR Max Premix (2x) 25 μL; insertion fragment 250 ng; vector fragment 149 ng (~1/10 equal mole as compared to insertion); water to reaction volume of 50 μL. The PCR reaction was conducted as follows: 1 cycle of 98 °C, 2 min; 20 cycles of 98 °C, 10 s; 60 °C, 15 s; 72 °C 4 min, 15 cycles of 98 °C, 10 s; 60 °C, 15 s; 72 °C 8 min; 1 cycle of 72 °C 10 min. To obtain the monomeric linearized plasmid containing a single insertion and vector fragment, the PCR product was digested using *Bgl*II at 37 °C for 16-20 hours in the following reaction: multimeric DNA 45 μL; 10x NEB 3.1 buffer 25 μL, water 225 μL, *Bgl*II (10,000 U/mL) 5 μL. The digested DNA fragment was cleaned up using QIAquick Gel Extraction Kit (Qiagen) and cyclized using Quick Ligation™ Kit (New England Biolabs). The ligation product was then transformed into BL21 (DE3) competent cells containing pEVOL- pAzF as described above to obtain the mutant library for screening.

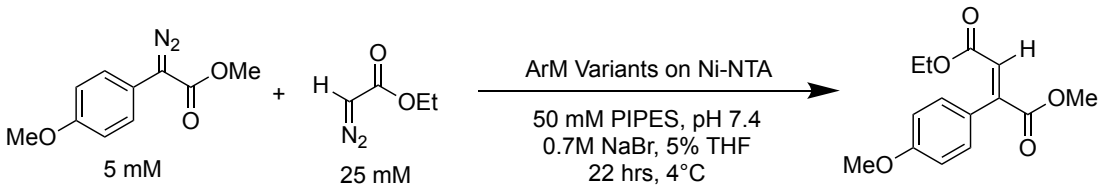
ArM library screening

The cell cultures were incubated, overexpressed and lyzed as described previously, using 5 mL overexpression cultures in 24-well plates.[12] The lysate was pelleted by centrifugation at 3,600 rpm and 480 μL of the clarified lysate from each well was transferred to a 96-deep well filter plate. 120 μL of cofactor **1** stock solution in ACN (12.8 μM, 0.010 mg/mL) was added to each well while shaking at 650 rpm at 4 °C. The mixture was incubated at 4 °C, 750 rpm for one hour, at which point 100 μL of 50% Ni-NTA resin was added to each well. ~1 cm of the multichannel

pipette tips were cut off using scissors to allow easier transfer of the resin. The plate was shaken at 4° C for 5 minutes to allow protein binding, then centrifuged at 2500 rpm for 5-10 minutes until the resin in each well was dry. The resin was washed to remove excess cofactor by the addition of 500 µL of 50% acetonitrile/50 mM PIPES buffer to each well. The plate was filtered by centrifugation at 3000 rpm for 3 minutes to remove the buffer. Three washes were done in total (1.5 mL total wash buffer) with a final centrifugation time of 5 minutes to ensure each well was dry.

285 µL reaction buffer (50 mM PIPES, pH 7.4 with 0.7 M NaBr) was added to each well and the plate was shaken at 4 °C, 750 rpm for 10 minutes. After this cooling period, 15 µL of substrate stock solution was added (final concentration of 5 mM methyl (4-methoxyphenyl)diazoacetate and 25 mM ethyldiazoacetate) at 4°C with 750 rpm shaking. The screening plate was sealed with a 96-well rubber mat to prevent evaporation and shaken at these conditions for 16 hours. The reaction was quenched by addition of 50 µL of 30 mM 1,3,5-trimethoxybenzene in ethyl acetate as internal standard. The plate filtered by centrifugation at 3,000 rpm for 5 min. To remove all organics from the resin, two additional filtrations of 75 µL ethyl acetate were applied and the biphasic mixture was clarified via centrifugation at 3,000 rpm for 5 min. 100 µL of the upper organic portion from each well were transferred to a 96-well microtiter plate for SFC analysis. Mutants displaying a higher product/internal ratio or improved E/Z isomer ratio than the parent were selected for DNA sequencing and validation.

Example plate data for 1-SGH-Q98NNK library:



Product (E-alkene)/Internal standard (hits highlighted in green, parent highlighted in blue):

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.08	0.16	0.11	0.10	0.06	0.10	0.56	0.14	0.18	0.12	0.22	0.49
B	0.16	0.18	0.08	0.10	0.21	0.06	0.25	0.31	0.09	0.14	0.05	0.62
C	0.11	0.52	0.27	0.37	0.12	0.43	0.15	0.06	0.08	0.08	0.08	0.22
D	0.11	0.56	0.09	0.12	0.50	0.10	0.11	0.03	0.53	0.29	0.09	0.05
E	0.12	0.16	0.11	0.10	0.08	0.05	0.11	0.09	0.14	0.11	0.13	0.13
F	0.21	0.21	0.18	0.12	0.23	0.10	0.17	0.58	0.14	0.58	0.13	0.10
G	0.08	0.22	0.14	0.25	0.23	0.57	0.35	0.23	0.19	0.49	0.11	0.05
H	0.10	0.09	0.11	0.08	0.07	0.09	0.08	0.20	0.06	0.17	0.23	0.09

E/Z selectivity (hits highlighted in green, parent highlighted in blue):

	1	2	3	4	5	6	7	8	9	10	11	12
A	1.9	3.5	2.1	1.8	1.5	2.1	6.8	2.2	2.7	2.1	3.1	6.6
B	2.1	3.8	1.7	1.8	4.1	1.6	3.2	3.4	2.1	2.3	1.4	8.5
C	2.3	6.6	2.8	4.9	2.0	6.1	2.6	0.0	1.7	1.6	1.8	2.5
D	2.0	7.1	1.7	1.8	6.2	1.8	1.9	0.9	6.8	4.6	1.9	1.4
E	2.1	1.8	2.2	1.7	1.8	1.0	2.1	1.7	2.2	2.1	2.3	2.4
F	2.9	2.9	2.6	2.1	2.5	2.0	2.5	7.4	2.4	7.3	2.2	2.0
G	1.9	2.2	1.8	2.2	2.9	6.9	4.8	2.2	2.7	6.4	2.0	1.5
H	2.0	1.8	1.9	1.6	1.5	2.1	1.8	2.3	1.4	1.4	2.2	1.5

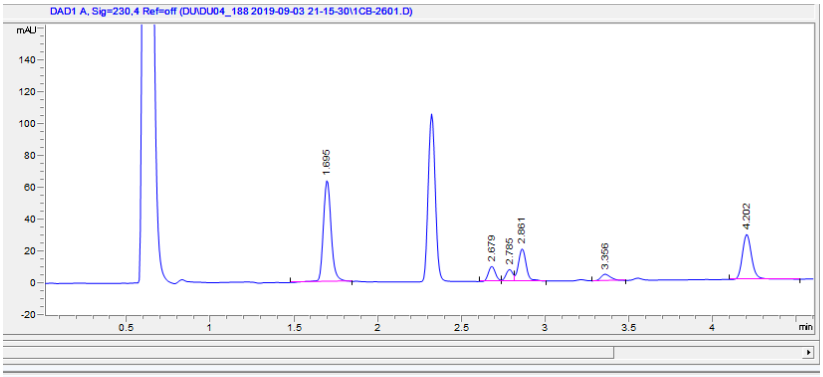
Example chromatogram for the hit C2 which was found to be Q98P (2-P):

E-alkene product: 4.2 mins

Z-alkene product: 3.4 mins

Internal standard: 1.7 mins

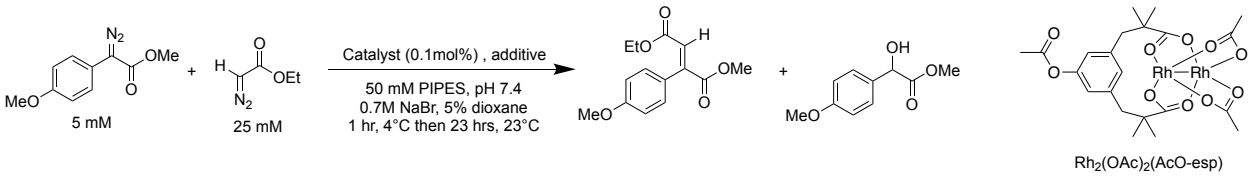
(Method and standard chromatograms can be found in Section S9)



#	Time	Area	Height	Width	Area%	Symmetry
1	1.695	209.5	63.8	0.0523	46.527	0.87
2	2.679	28.4	9.5	0.0471	6.315	0.901
3	2.785	21.5	7.4	0.0457	4.770	1.117
4	2.861	65.4	20.3	0.0496	14.527	0.857
5	3.356	16.6	4	0.0619	3.679	0.683
6	4.202	108.9	28.3	0.0589	24.182	0.823

3.4.7 ER and Glucose Modfication

General reaction setup



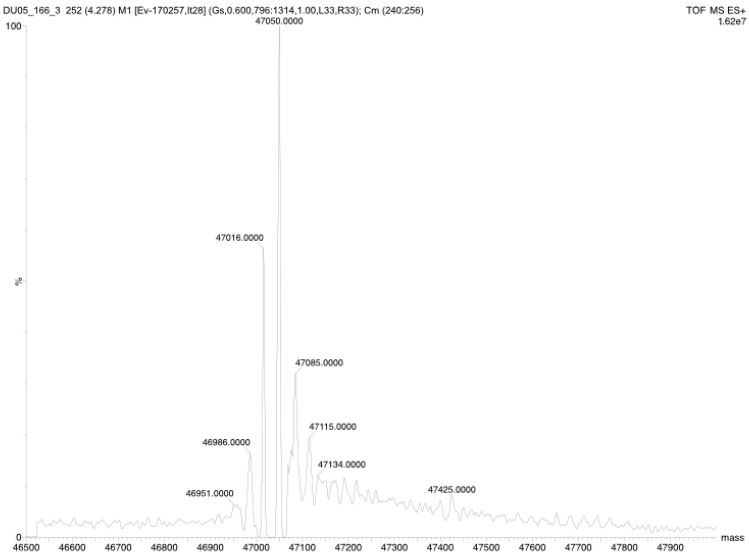
$Rh_2(OAc)_2(AcO\text{-}esp)$

A 285 µL solution of 5 µM catalyst (ArM-catalyzed reactions were corrected by the known bioconjugation efficiency, typically 90%) and additive in 50 mM PIPES, pH 7.4 with 0.7 M NaBr was prepared in a 1.5 mL microcentrifuge tube. While shaking at 750 rpm at 4 °C, 15 µL of the

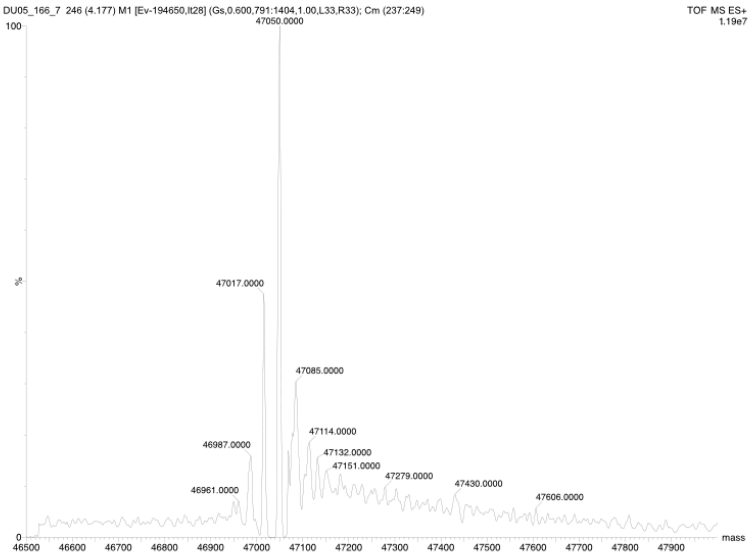
diazo stock solution in dioxane was added (final concentration of 5 mM methyl (4-methoxyphenyl)diazoacetate, 25 mM EDA) to obtain a final concentration of 5 mM donor-acceptor diazo and 25 mM acceptor-only diazo. The reactions were sealed and briefly inverted to assist with mixing of the substrate, then placed in a thermomixer and shaken at 750 rpm at 4 °C for 1 hour, at which point the reaction temperature was raised to room temperature for 23 hours. For the protein modification experiments, the reaction was desalted and analyzed by intact protein LCMS using the standard method described in the bioconjugation section (Section S5). The glucose modification experiments were first quenched with 300 µL MeOH with 15 mM 1,3,5-trimethoxybenzene (internal standard) and filtered before LCMS analysis.

Example intact protein mass spectrum of ene reductase before and after diazo coupling:

KYE1 control:

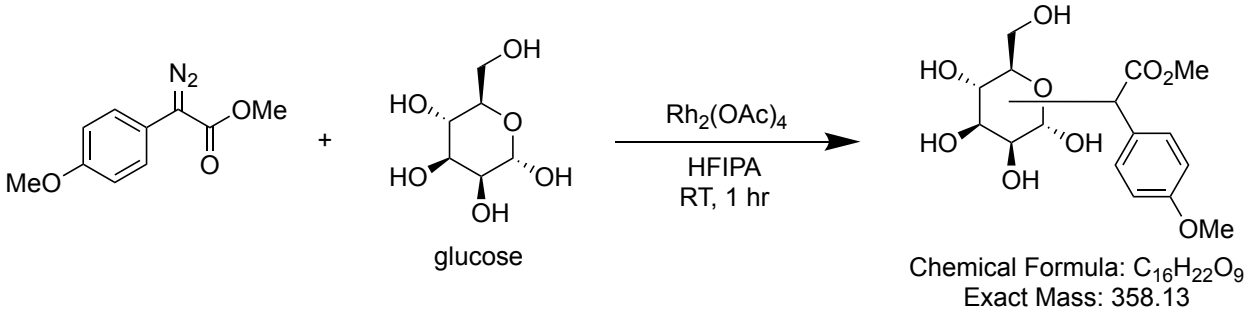


KYE1 after diazo coupling catalyzed by 5-G showing no observable modification:



HFIPA glucose modification

To verify that the product observed by LCMS is indeed the arylacetate-modified glucose, a comparable reaction was run in hexafluoroisopropanol to form the expected product.



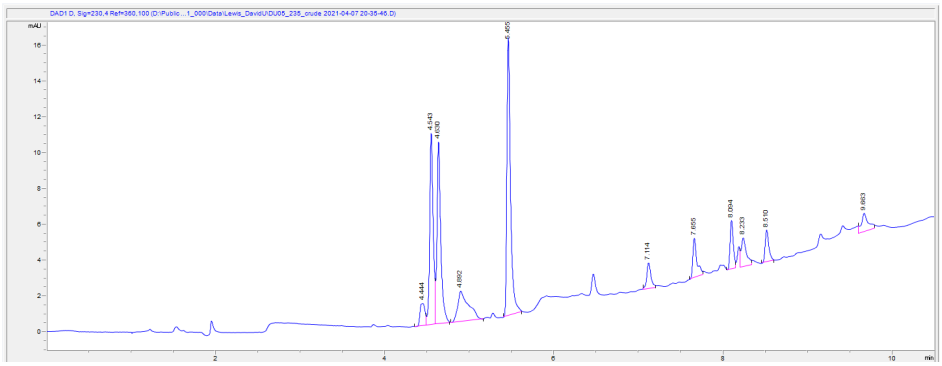
In an oven-dried 1.5 mL conical reaction vessel under N_2, 2.2 mg $Rh_2(OAc)_4$ (1 mol%) and glucose (90.1 mg, 0.5 mmol) were combined and 0.9 mL HFIPA (anhydrous) was added. To this heterogeneous mixture, 21.1 mg (0.1 mmol) diazo was added in 0.1 mL HFIPA dropwise over 5 minutes. The reaction was stirred vigorously for 1 hour under a N_2 atmosphere at which point no

orange color was observed. The solvent was evaporated and the crude reaction mixture was taken up in MeOH for LCMS analysis.

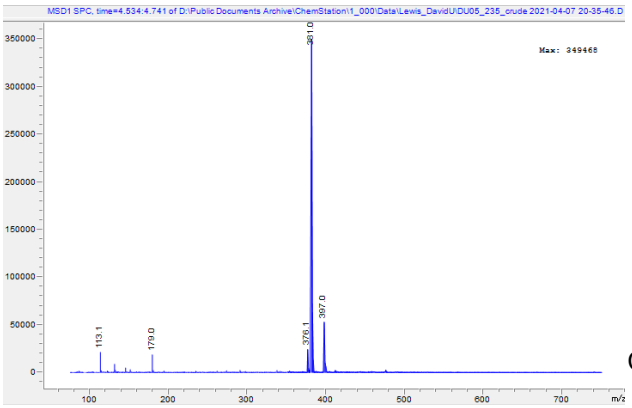
Small molecule LCMS showing glucose modification

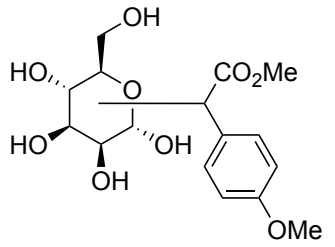
The modification of glucose was observed by LCMS. A 15cm Eclipse Plus C18 column was used. Mobile phase used was A: H_2O + 0.1% FA, B: ACN + 0.1% FA. Method: 10% B for 0.5 mins, 10% B to 90% B over 8 minutes, 2 minutes at 90% B.

Crude reaction mixture:



The peaks at 4.54 and 4.63 contain the mass spec shown below:





Chemical Formula: $C_{16}H_{22}O_9$
Exact Mass: 358.13

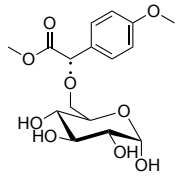
M+Na = 381.12

M+K = 397.09

Fragmentation peak at 179.0 corresponds to the breaking of the glucose-benzyl bond (exact

isomer(s) are unknown):

Chemical Formula: $C_{10}H_{11}O_3^{\bullet}$
Exact Mass: 179.07
Molecular Weight: 179.20



Chemical Formula: $C_6H_{11}O_6$
Exact Mass: 179.06
Molecular Weight: 179.15

This control reaction is consistent with the modification of glucose by $Rh_2(OAc)_2$(AcO-esp) in water.

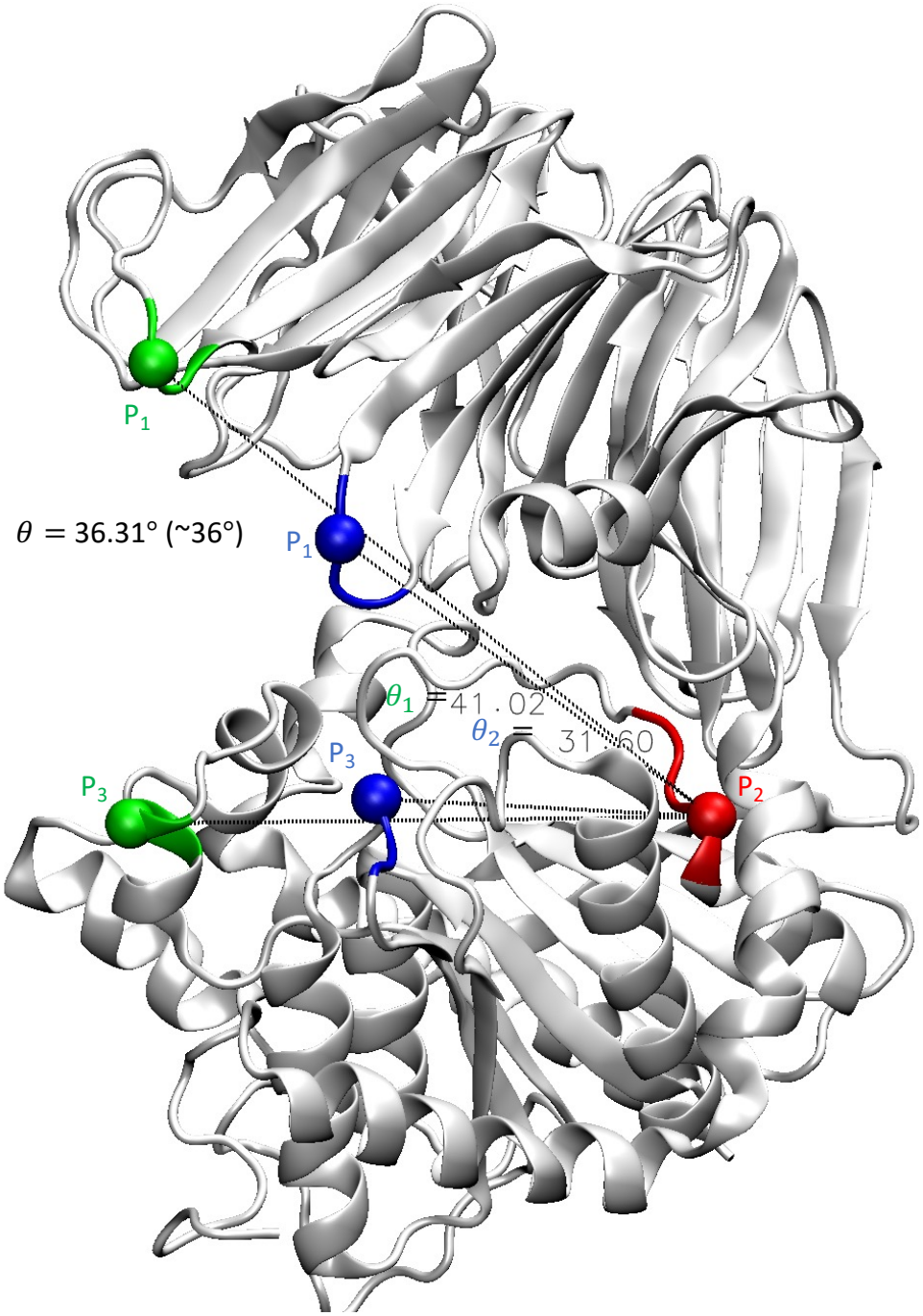
3.4.8 Computational Procedures

To gain insight into how the POP scaffold might give rise to the observed ArM selectivity, molecular dynamics simulations were conducted on models of 5-G that involved different starting coordination states of dirhodium cofactor **1**. Three starting coordination states (5-G apo, 5-G with parameterized Rh-His bond, 5-G without Rh-His bond) were studied. The initial structure of 5-G apo was constructed from the crystal structure of POP, with the following amino acids mutations made using VMD[17]: S477F/E104A/F146A/K199A/D202A/S301G/G99S/Y326H/Q98P/S99H/V71G/E283G. The 5-G with Rh-His and 5-G without Rh-His bond models were constructed by mutating S477Z (Z is azidophenylalanine, where dirhodium cofactor **1** covalently links to the protein), which is the only other mutation from 5-G apo. No other changes where made to the model without a Rh-His bond. The Rh-His model contains a parameterized H326 bond to one of the dirhodium atoms, as shown in Figure 3.3.

MD simulations were performed for the 5-G apo, with Rh-His and no Rh-His POP enzymes. All three structures were bathed in a 0.15M KCl solution using the *Solution Builder* Module in CHARMM-GUI.[18] These systems were roughly 100 × 100 × 100 $Å^3$ in dimension and contained ~ 110,000 atoms. The periodic boundary conditions were counted using the particle-mesh Ewald method with an automatic generated grid size.

Once the simulation systems were generated, they were subjected to equilibration at 358.15 K. The system was first equilibrated in an NVT ensemble for 10 ns. The equilibration simulations were performed using NAMD2.14 GPU acceleration version package[19] on Nvidia's P100 GPUs. After equilibration, the systems were simulated for 1000 ns each in an NPT ensemble with temperature set to 358.15 K and the isotropic pressure set to 1 atm. Langevin thermostats with a damping coefficient of 1 ps^{-1} were used to keep the temperature constant. The cutoff of the van der Waals interactions and short-range electrostatic interactions were set to ~25 Å as suggested by the guesser script. The additive C36 force field was used in all the simulations performed here.[20,21,22,23] The force field parameters for the covalently linked dirhodium cofactor **1** were generated using the GAAMP server.[24]

Example model showing θ calculation:



Theta, θ, is an average of two angles (θ_1 and θ_2), each one defined by three groups of Cα atoms (colored selections). For θ_1, the three groups of Cα atoms are from residues R158-E160, F412-Q415, and H510-L512. Among them, the first group (G_1) lies on the loop region in the propeller domain facing the peptidase domain, the second (G_2) is in the propeller-peptidase hinge region, and the third (G_3) is in the loop region of the peptidase domain facing the propeller blade 3. The centroid of each group is computed and named as P_1, P_2, and P_3 (spheres) according to their group names. θ_1 is computed as the angle between vectors $\mathbf{P_2P_1}$ and $\mathbf{P_2P_3}$ (dotted lines). θ_2 is defined using the same method with three groups of residues: G117-D119, F412-Q415, and D560-R562. G_1, P_1, and G_3, P_3 for θ_1 and θ_2 are indicated in green and blue respectively. The shared vertex G_2, P_2 is indicated in red. Positions used to determine θ for analogous POPs were identified via sequence alignment.

3.4.10 Experimental References

1. H. Yang, P. Srivastava, C. Zhang, J. C. Lewis, *ChemBioChem* **2014**, *15*, 223–227.
2. J. Dommerholt, S. Schmidt, R. Temming, L. J. A. Hendriks, F. P. J. T. Rutjes, J. C. M. Van Hest, D. J. Lefeber, P. Friedl, F. L. Van Delft, *Angew. Chem., Int. Ed.* **2010**, *49*, 9422–9425.
3. S. W. Kwok, L. Zhang, N. P. Grimster, V. V. Fokin, *Angew. Chem., Int. Ed.* **2014**, *53*, 3452-3456.
4. H. M. L. Davies, T. Hansen, M. Rowen Churchill, *J. Am. Chem. Soc.* **2000**, *122*, 3063–3070.
5. W.-Y. Yu, Y.-T. Tsoi, Z. Zhou, A. S. C. Chan, *Org. Lett.* **2008**, *11*, 469–472.
6. Y. Wei, S. Liu, M.-M. Li, Y. Li, Y. Lan, L.-Q. Lu, W.-J. Xiao, *J. Am. Chem. Soc.* **2018**, *141*, 133–137.
7. D. P. Hari, J. Waser, *J. Am. Chem. Soc.* **2016**, *138*, 7, 2190–2193
8. T. Toma, J. Shimokawa, T. Fukuyama, *Org. Lett.* **2007**, *9*, 3195–3197.
9. J. H. Hansen, B. T. Parr, P. Pelphrey, Q. Jin, J. Autschbach, H. M. L. Davies, *Angew. Chem., Int. Ed.* **2011**, *50*, 2544–2548.
10. T. Xiao, M. Mei, Y. He, L. Zhou, *Chem. Commun.* **2018**, *54*, 8865-8868.
11. Y. Wang, M. J. Bartlett, C. A. Denard, J. F. Hartwig, H. Zhao, *ACS Catal*. **2017**, 7, 2548–2552.
12. H. Yang, A. M. Swartz, H. J. Park, P. Srivastava, K. Ellis-Guardiola, D. M. Upp, G. Lee, K. Belsare, Y. Gu, C. Zhang, R. E. Moellering, J. C. Lewis, *Nat. Chem.* **2018**, *10*, 318–324.
13. K. L. Heckman, L. R. Pease, *Nat. Protoc.* **2007**, *2*, 924–932.

14. K. D. Belsare, M. C. Andorfer, F. Cardenas, J. R. Chael, H. J. Park, J. C. Lewis, *ACS Synth. Biol.* **2017**, 6, 416–420.
15. X. Z. Zhang, Y. H. P. Zhang, *Microb. Biotechnol.* **2011**, *4*, 98–105.
16. C. Zhong, C. You, P. Wei, Y.-H. P. Zhang, in *Synth. DNA Methods Protoc.* (Ed.: R.A. Hughes), Springer New York, New York, NY, **2017**, pp. 49–61.
17. Humphrey, W., Dalke, A., Schulten, K., VMD - Visual Molecular Dynamics, *J. Molec. Graphics* **1996**, *14*, 33-38.
18. Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: a Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865.
19. J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Henin, W. Jiang, R. McGreevy, M. C. R. Melo, B. K. Radak, R. D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. V. Kale, K. Schulten, C. Chipot, E. Tajkhorshid. *J. Chem. Phys.*, **2020**, *153* 044130.
20. Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
21. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
22. Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D., Jr.; Pastor, R. W. *J. Phys. Chem. B* **2010**, *114*, 7830–7843.
23. MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
24. Huang, L.; Roux, B. *J. Chem. Theory Comput.* **2013**, *9*, 3543–3556.

3.5 Acknowledgements

First and foremost, I would like to thank Rui Huang for his work in many of the parts of this project. His proficiency in cloning led to all of the libraries screened in this chapter and his passion for reaction optimization was critical in developing conditions that were suitable for cascade reactions. I would also like to thank Max Bultmann for his work on substrate synthesis throughout this project. His growth into a proficient synthetic chemist over the course of this project was impressive to witness. Additional thanks go out to Ying Li who developed and ran the simulation which contributed to our understanding of the dynamics of the ArM. Finally, I would like to thank Hyun June Park for work on the Ni-NTA screening method that was used here.

3.6 References

[1] J. Mahatthananchai, A. M. Dumas, J. W. Bode, *Angewandte Chemie International Edition in English* 2012, *51*, 10954–10990.

[2] R. R. Knowles, E. N. Jacobsen, *Proceedings Of The National Academy Of Sciences Of The United States Of America* 2010, *107*, 20678–20685.

[3] F. D. Toste, M. S. Sigman, S. J. Miller, *Accounts Of Chemical Research* 2017, *50*, 609–615.

[4] J. Trouvé, R. Gramage-Doria, *Chem Soc Rev* 2021, *50*, 3565–3584.

[5] L. V. A. Hale, N. K. Szymczak, *Acs Catal* 2018, *8*, 6446–6461.

[6] K. M. Lancaster, in *Molecular Electronic Structures of Transition Metal Complexes I. Structure and Bonding* (Eds.: M.D.D. P., D. J., D. P.), 2011.

[7] J. C. Lewis, P. S. Coelho, F. H. Arnold, *Chem Soc Rev* 2010, *40*, 2003–2021.

[8] S. W. Ragsdale, *Chemical Reviews* 2006, *106*, 3317–3337.

[9] R. Breslow, *Accounts Of Chemical Research* 1995, *28*, 146–153.

[10] J. C. Lewis, *Accounts Of Chemical Research* 2019, *52*, 576–584.

[11] V. Köhler, N. J. Turner, *Chemical Communications* 2014, *51*, 450–464.

[12] C. A. Denard, J. F. Hartwig, H. Zhao, *ACS Catalysis* 2013, *3*, 2856–2864.

[13] F. Schwizer, Y. Okamoto, T. Heinisch, Y. Gu, M. M. Pellizzoni, V. Lebrun, R. Reuter, V. Kohler, J. C. Lewis, T. R. Ward, *Chemical Reviews* 2017, *118*, 142–231.

[14] M. Jeschek, R. Reuter, T. Heinisch, C. Trindler, J. Klehr, S. Panke, T. R. Ward, *Nature* 2016, *537*, 661–665.

[15] A. D. Liang, J. Serrano-Plana, R. L. Peterson, T. R. Ward, *Accounts Of Chemical Research* 2019, *52*, 585–595.

[16] S. Chordia, S. Narasimhan, A. L. Paioni, M. Baldus, G. Roelfes, *Angewandte Chemie Int Ed* 2021, *60*, 5913–5920.

[17] S. Eda, I. Nasibullin, K. Vong, N. Kudo, M. Yoshida, A. Kurbangalieva, K. Tanaka, *Nat Catal* 2019, *2*, 780–792.

[18] V. Köhler, Y. M. Wilson, M. Dürrenberger, D. Ghislieri, E. Churakova, T. Quinto, L. Knörr, D. Häussinger, F. Hollmann, N. J. Turner, T. R. Ward, *Nature Chemistry* 2012, *5*, 93–99.

[19] M. A. S. Mertens, D. F. Sauer, U. Markel, J. Schiffels, J. Okuda, U. Schwaneberg, *Catal Sci Technol* 2019, *9*, 5572–5576.

[20] J. Huang, Z. Liu, B. Bloomer, D. Clark, A. Mukhopadhyay, J. Keasling, J. Hartwig, 2021, DOI 10.21203/rs.3.rs-130318/v1.

[21] D. Gillingham, N. Fei, *Chem Soc Rev* 2013, *42*, 4918–4931.

[22] H. M. L. Davies, D. Morton, *Chemical Society Reviews* 2011, *40*, 1857–1869.

[23] P. Srivastava, H. Yang, K. Ellis-Guardiola, J. C. Lewis, *Nature Communications* 2015, *6*, 7789.

[24] Y. Wang, M. J. Bartlett, C. A. Denard, J. F. Hartwig, H. Zhao, *ACS Catalysis* 2017, *7*, 2548–2552.

[25] J. H. Hansen, B. T. Parr, P. Pelphrey, Q. Jin, J. Autschbach, H. M. L. Davies, *Angewandte Chemie Int Ed* 2011, *50*, 2544–2548.

[26] J. Wu, X. Li, X. Qi, X. Duan, W. L. Cracraft, I. A. Guzei, P. Liu, W. Tang, *J Am Chem Soc* 2019, *141*, 19902–19910.

[27] J. Ohata, Z. T. Ball, *Dalton T* 2018, *47*, 14855–14860.

[28] H. Yang, A. M. Swartz, H.-J. Park, P. Srivastava, K. Ellis-Guardiola, D. M. Upp, G. Lee, K. Belsare, Y. Gu, C. Zhang, R. E. Moellering, J. C. Lewis, *Nature Chemistry* 2018, *10*, 318–324.

[29] K. D. Belsare, M. C. Andorfer, F. S. Cardenas, J. R. Chael, H.-J. Park, J. C. Lewis, *ACS Synthetic Biology* 2017, *6*, 416–420.

[30] P. Dydio, H. M. Key, A. Nazarenko, J. Y. E. Rha, V. Seyedkazemi, D. S. Clark, J. F. Hartwig, *Science* 2016, *354*, 102–106.

[31] J. H. Hansen, B. T. Parr, P. Pelphrey, Q. Jin, J. Autschbach, H. M. L. Davies, *Angewandte Chemie Int Ed* 2011, *50*, 2544–2548.

[32] Y. Yanto, C. K. Winkler, S. Lohr, M. Hall, K. Faber, A. S. Bommarius, *Org Lett* 2011, *13*, 2540–2543.

[33] C. Stueckler, M. Hall, H. Ehammer, E. Pointner, W. Kroutil, P. Macheroux, K. Faber, *Org Lett* 2007, *9*, 5409–5411.

[34] Z. Chen, F. Vohidov, J. M. Coughlin, L. J. Stagg, S. T. Arold, J. E. Ladbury, Z. T. Ball, *Journal of the American Chemical Society* 2012, *134*, 10138–10145.

[35] J. Wu, X. Li, X. Qi, X. Duan, W. L. Cracraft, I. A. Guzei, P. Liu, W. Tang, *J Am Chem Soc* 2019, *141*, 19902–19910.

[36] R. Breslow, S. Bandyopadhyay, M. Levine, W. Zhou, 2006, *7*, 1491–1496.

[37] K. Ellis-Guardiola, H. Rui, R. L. Beckner, P. Srivastava, N. Sukumar, B. Roux, J. C. Lewis, *Biochemistry* 2019, *58*, 1616–1626.

[38] Z. T. Ball, *Accounts Of Chemical Research* 2013, *46*, 560–570.

[39] Z. T. Ball, *Accounts Chem Res* 2013, *46*, 560–570.

[40] Y. Valasatava, A. Rosato, N. Furnham, J. M. Thornton, C. Andreini, *J Inorg Biochem* 2018, *179*, 40–53.

Chapter 4: Engineering Dirhodium ArMs for N-H Functionalization Reactions

4.1 Introduction

Alpha-chiral amines are a common motif in many biologically relevant compounds, including amino acids, natural products, and pharmaceuticals.[1,2] In fact, of the 200 top pharmaceuticals in 2018, approximately 35% contained a chiral amine group.[3] The pervasiveness of such compounds necessitates efficient methods to form C-N bonds that are both high yielding and highly enantioselective. A number of reactions have been developed with these goals in mind, such as enantiospecific enamine and imine hydrogenation,[1] additions to ketamines or nitriles, and Mannich reactions, amongst others. These diverse reaction types make up an excellent toolbox for the stereospecific formation of chiral amines, which has been supplemented by recent progress in biocatalysis. Taking advantage of an abundance of natural enzymes capable of forming similar compounds, recent engineering efforts have shown that enzymes, including transaminases[4,5] and monoamine oxidases[6], can catalyze the formation or enantioenrichment of alpha-chiral amines with yields and selectivities that rival state-of-the-art synthetic methods. A notable example of this is in the directed evolution of a transaminase for the synthesis of an alpha-chiral amine in the production of the drug sitagliptin.[7] The final transaminase resulting from this work was able to catalyze the formation of the desired chiral amine product with higher enantioselectivity (99.95% e.e.) than the previously developed rhodium-catalyzed enamine hydrogenation, and eliminate the use of a transition metal catalyst late in the synthesis.

A route that has seen increased interest in recent years is transition metal-catalyzed carbene insertion into N-H bonds. This is a particularly efficient method for forming chiral carbon–nitrogen bonds but, historically, the development of asymmetric versions of this chemistry had proven

challenging.[8] As dirhodium tetracarboxylate catalysts have been long recognized as proficient at carbene transfer chemistry, many early efforts centered on this class of catalysts.[9] In early studies with chiral dirhodium catalysts, for example, intermolecular N-H insertion reactions achieved only low to modest enantioselectivities (< 50 % ee).[10] As discussed in depth in Section 1.3.2, multiple theoretical investigations into the mechanism of dirhodium-catalyzed nucleophilic carbene insertion have been performed and, in all cases, the enantiodetermining proton transfer is found to be most energetically favorable when the planar intermediate has dissociated from the dirhodium catalyst.[11–13] A related example of this is shown in Figure 1.7, where the only energetically accessible pathway for proton transfer that was found in the reaction of water with a donor-acceptor carbenoid was after dissociation occurred.

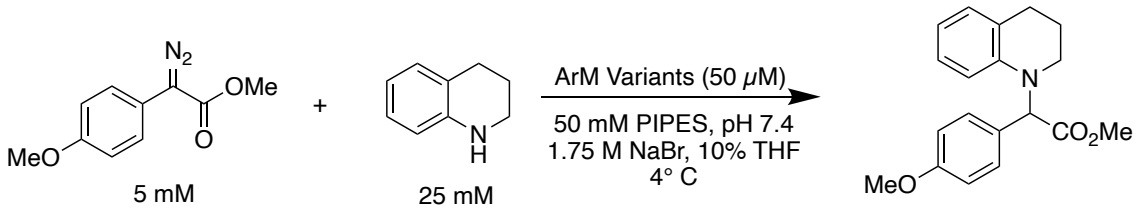
With a more thorough understanding of the stepwise nature of this reaction, alternate approaches have been pursued. As the dissociation of the prochiral intermediate is most pronounced for dirhodium catalysts, other achiral carbene transfer catalysts have been used with chiral proton-transfer cocatalysts.[8] A recent example using a Cu^{II} carbene transfer catalyst in combination with a chiral H-bonding catalyst enables highly selective alkyl N-H functionalization, highlighting how powerful this approach can be.[14] The necessity of controlling the proton transfer was also demonstrated in the directed evolution of a P450 for N-H functionalization of anilines.[15] In this case, substantial computation evidence suggested that the ylide intermediate dissociated from the iron center but was bound by the enzyme scaffold where a water-facilitated proton transfer occurred.

Controlling the protonation step has also been achieved using dirhodium catalysts. When used with chiral phosphoric acids or H-bonding catalysts, achiral dirhodium tetracarboxylate catalysts were found to enable selective aniline[16], amide, and carbamate[17] N-H functionalization.

An interesting absence in the literature is the asymmetric functionalization of alkyl N-H bonds using dirhodium catalysts. There are two reasons as to why this is more a challenging class of substrates: the basicity of amines is far higher than for anilines or amides, meaning they would be protonated when used with chiral phosphoric acids, and they are far better ligands at the axial position of the dirhodium catalyst, effectively preventing carbenoid formation.[14]

The N-H functionalization of anilines catalyzed by dirhodium ArMs has been shown using four of the variants developed in our work on cyclopropanation (Table 4.1).[18] As discussed in Section 2.2.1, we had previously found that 3-VRVH catalyzed the reaction of 1,2,3,4-tetrahydroquinoline with methyl (4-methoxyphenyl)diazoacetate, achieving high yields but modest enantioselectivity. The yields for this reaction were reasonably high for the engineered variants, suggesting that the aniline substrate is significantly out competing the O-H insertion side reaction. This may very well result from anilines higher nucleophilicity, with a Ritchie nucleophilicity parameter (N^+) of 13 in water compared to water itself with a N^+ value of 5.1.[19] The orders of magnitude difference in nucleophilicity would result in a significantly faster attack of the carbene by aniline. While the enantioselectivity was not impressive, these results suggest that the scaffold is operating in two ways: controlling the activity and chemoselectivity of the dirhodium cofactor as well as binding the planar intermediate to enable a chiral proton transfer to occur.[20]

Table 4.1 N-H functionalization catalyzed by the cyclopropanation lineage.



Entry	Variant	Lineage	Yield (%)	e.e. (%)
1	0-ZA4	Z477	55	8
2	1-NAGS	Z477	77	6
3	3-VRVH	Z477	73	40
4	1-RFY	Z413	87	-12

The latter role, binding a substate for selective proton transfer, is an uncommon one for an artificial metalloenzyme scaffold though there are some examples.[21] Of course, proton transfer is a reaction proficiently catalyzed by natural enzymes such as 4-oxalocrotonate tautomerase, where a terminal proline residue deprotonates the substrate alpha carbon and transfers the proton to the gamma position.[22] Additionally, despite lacking an ideal active site as is the case for tautomerases[15], we envision that the scaffold binds the intermediate in an orientation that enables an enantiospecific tautomerization, perhaps involving an intervening water molecule, as was recently proposed for N-H functionalization catalyzed by engineered P450s.[15] The compact active site of the P450s in that work provided a straightforward path to improvement, where the position of the ylide was predetermined by the location of carbene insertion. In the POP scaffold, this intermediate could reside throughout the large internal volume, making this step challenging to control. Engineering a dual-purpose ArM scaffold is rare in the ArM literature, but frequently observed in natural enzymes, so improving this function via directed evolution would further highlight the capacity of these hybrid catalysts to control the primary and secondary coordination spheres.[23]

4.2 Results and Discussion

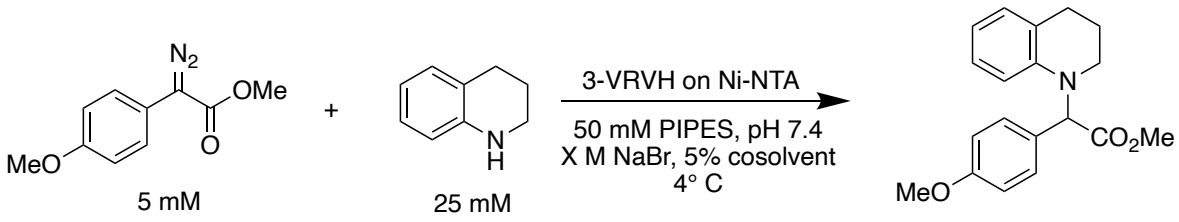
4.2.1 N-H Functionalization of Anilines

4.2.1.1 High-Throughput Method Optimization and Parent Selection

With a scaffold capable of achieving modest enantioselectivity, we needed to verify that our high throughput screening method was suitable for N-H functionalization before directed evolution could begin. The Ni-NTA-based screening method outlined in Section 3.2.2 was used again here.[18] In this method, bioconjugation of the scaffold is performed in lysate and the resulting ArM is immobilized on Ni-NTA in a filter plate. The resin is washed with 50% ACN/Tris buffer in order to remove excess cofactor, and the reaction buffer and substrates are then added. After the reaction is complete, the products are filtered and, after two washes with ethyl acetate, extracted for SFC analysis. After initial experiments indicated that this method was enabling ArM-catalyzed N-H functionalization as we had hoped, we needed to determine the best conditions for screening that resulted in high reproducibility. Previous data indicated that 3-VRVH was the most selective scaffold for N-H functionalization, so this was used as the scaffold for the optimization of reaction conditions (Table 4.2).

.

Table 4.2 Optimization of N-H functionalization screening conditions.



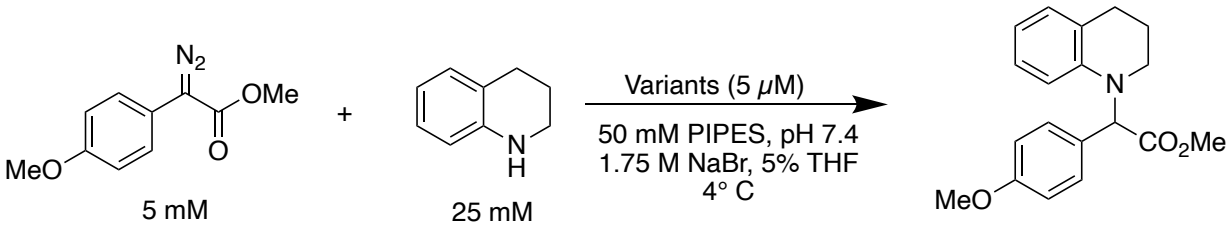
Entry	Variant	[NaBr]	Cosolvent	Yield (%)	e.e. (%)
1	3-VRVH	0	THF	15	12
2	3-VRVH	0.1	THF	20	23
3	3-VRVH	0.7	THF	23	30
4	3-VRVH	1.75	THF	36	34
5	3-VRVH	1.75	MeOH	41	25
6	3-VRVH	1.75	ACN	29	25
7	3-VRVH	1.75	DMSO	38	22
8	Blank	1.75	THF	5	-6

Similar to cyclopropanation[24] and diazo coupling[25], increased salt concentration was found to be very important for enantioselectivity, and the highest sodium bromide concentration tested, 1.75 M, resulted in the highest e.e. (Entry 4). As expected based on our previous experience with this method, there was only a slight drop in enantioselectivity (34%) compared to the reaction using homogenous, purified ArM (40%). Cosolvents were also screened as a part of the optimization, and it was found that THF was superior to the other polar solvents examined, just as in previous studies. Importantly, blank wells containing lysate with no ArM scaffold had very low background activity (Entry 8, the negative enantioselectivity is due to a small overlapping impurity on the minor enantiomer in the SFC chromatogram). With a method that could reliably detect ArM selectivity in high-throughput, we screened a panel of variants from previous directed evolution campaigns.

From this panel, 10 variants were selected for purified ArM catalysis (Table 4.3). Interestingly, the best variants for N-H functionalization were found to be 1-SGH and 3-VRVH,

which performed very similarly. Interestingly, the evolved variants for diazo coupling (Entries 4-8) were significantly less selective than the parent, indicating that the improvements required for diazo coupling were not impactful in this reaction. 1-SGH was selected as the parent for future evolution because it has fewer mutations from the wild-type enzyme, making it more stable and amenable to future mutations.[26]

Table 4.3 Parent screening for the N-H functionalization of THQ.



Entry	Variant	Lineage	Yield (%)	e.e. (%)
1	0-ZA4	Intial Parent	55	8
2	1-NAGS	Cyclopropanation	77	6
3	3-VRVH	Cyclopropanation	73	40
4	1-SGH	Cyclopropanation	81	40
5	2-P	Diazo coupling	80	37
6	3-H	Diazo coupling	80	37
7	4-G	Diazo coupling	83	36
8	5-G	Diazo coupling	79	33
9	2-P S99L	Diazo coupling	78	39
10	2-P S99Y	Diazo coupling	76	35

4.2.1.2 Directed Evolution

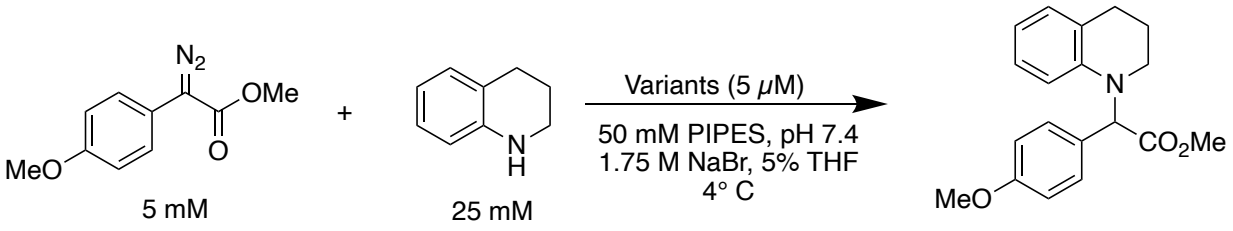
Having identified 1-SGH as a parent for ArM evolution, we next considered how best to generate variants of this scaffold. There are numerous ways to achieve mutagenesis that range from highly random, such as using error-prone PCR to introduce nearly random single base pair mutations in the gene of interest, to very targeted methods, such as site saturation mutagenesis to

optimize the amino acid at a single position. As our POP-ArM library screening methodology can evaluate only ~200 variants/day, selecting a mutagenesis strategy that maximized the likelihood of identifying mutants with impactful mutations was crucial. In the diazo coupling evolution detailed in Section 3.2.2, combinatorial codon mutagenesis (CCM)[27] on a pool of residues pointing towards the presumptive cofactor position in the interior cavity enabled rapid identification of improved variants,[25] so we decided to use the same technique and residue pool for library diversification for N-H functionalization.

A CCM library of 1-SGH was prepared, and the activity on 92 variants assayed. A hit, termed D7, was identified, and the purified ArM exhibited improved enantioselectivity (53%, Table 4, Entry 2). D7 contains 7 mutations, so individual reversions were made (Table 4, Entries 3-9). Interestingly, the H326Y mutation was found to have a negative impact on enantioselectivity, with the reversion catalyzing the N-H functionalization reaction of THQ with 58% e.e. The beneficial presence of two histidine residues that could bind the dirhodium cofactor from two positions in the same β-strand is noteworthy and unexpected. While it is possible that the cofactor remains bound to H326 and H328 acts in proton transfer or substrate orientation to improve selectivity, H328 was a rationally designed mutation (Section 1.3.3) that was found to improve selectivity through presumptive cofactor binding, so anchoring at this position is known to be possible. If this were the case, it appears that the minor shift in the location of the cofactor when bound to H326 or H328 does not have the negative effect on enantioselectivity that one might expect due to altered localization of the cofactor. Instead, this potentially negative effect is apparently overridden by the increased local concentration of ligating residues which should increase the fraction of time spent bound versus unbound. As we have previously shown in Section 3.2.5, when a Rh-His bond is present the scaffold maintains the closed structure to a greater extent

than when it is not bound.[25] Consequently, increasing the amount of time a Rh-His bond exists, as may have been done with the addition of a second histidine, should favor the closed conformation and decrease diffusion of the prochiral intermediate before proton transfer occurs in the chiral environment of the scaffold. The other mutations were either beneficial or neutral and were kept in the variant 2-HH (Table 4, Entry 3), which was taken forward as the parent for future evolution.

Table 4.4 Mutation deconvolution of D7 hit.



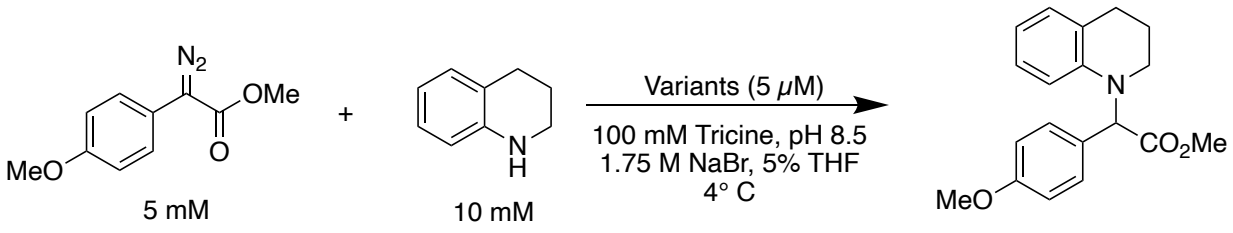
Entry	Variant	e.e. (%)
1	1-GSH	41
2	D7	53
3	D7+Y326H	58
4	D7+R98Q	53
5	D7+V209T	53
6	D7+F99S	47
7	D7+Y142W	46
8	D7+Y214W	47
9	D7+Y328L	38

The other 3 mutations that were found to have an impact on the enantioselectivity of 2-HH during deconvolution were F99S, Y142W, and Y214W. We had explored two variants with mutations at position 99 previously (Table 2, Entries 9 and 10) and found that they hadn't made a substantial impact on selectivity in that case, so we turned our focus to positions 142 and 214. As we had no reason to believe tyrosine was the best residue at either of these positions, we screened

site saturation libraries using NNK degenerate codons at both. Unfortunately, no improved variants were found in either case. We returned to another round of CCM mutagenesis which did result in hits, multiple of which had a mutation at position 194. To conclusively determine the best residue at this position, a site saturation library was screened, with the best variants validated by analysis of purified ArMs. Glutamate was found to be the best residue at this position of the scaffold, resulting in the variant 3-E. Optimization of the reaction pH and buffer was performed concurrently to this work (see Section 4.2.1.2 below), and these conditions were used in finding that 3-E maintained good yields (86%) with an improved e.e. of 66% (Table 4.5, Entry 3).

At this point in the directed evolution campaign, an investigation into the structural motifs that could be involved in binding of the intermediate (see Section 4.2.1.3 below) indicated that a pocket across from the presumptive cofactor location may be involved. A series of site saturation libraries at L97, D119, E120, and G121 were made to probe this pocket, and mutation at one of these positions, D119, was found to improve selectivity. An arginine at this position formed the improved variant 4-R (Table 4.5, Entry 4), which catalyzed the desired reaction with 68% e.e. and 93% yield and is the current parent for continued evolution efforts.

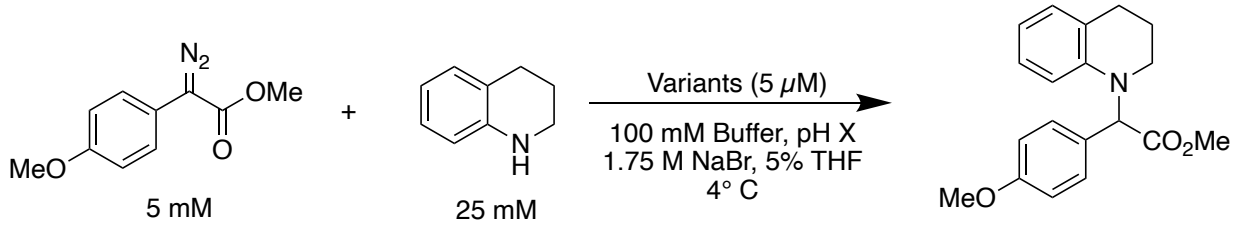
Table 4.5 Summary of the N-H Functionalization Lineage.



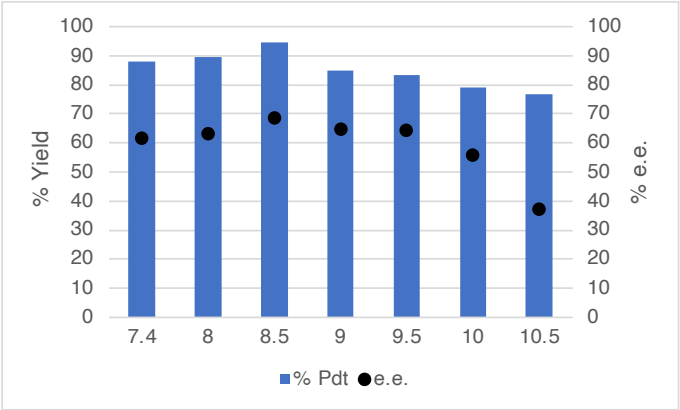
Entry	Variant	Yield (%)	e.e. (%)
1	1-SGH	89	43
2	2-HH	82	60
3	3-E	86	66
4	4-R	88	68

4.2.1.2 Reaction Optimization and pH Dependence of Enantioselectivity

While previous optimization was performed in order to improve the consistency and robustness of the screening on Ni-NTA, no changes from the purified ArM reaction conditions optimized for cyclopropanation had been explored. A thorough optimization of the reaction conditions was therefore performed using the variant 3-E. The reaction pH and buffer were found to be very impactful to the selectivity (Figure 4.1), with the best enantioselectivity found using Tricine buffer at pH 8.5. Substrate stoichiometry, buffer concentration, and sodium bromide concentration were also found to be impactful, with 2 equivalents of THQ and 100 mM Tricine buffer containing 1.75 M NaBr being ideal (Table 4.6, Entry 3).



a

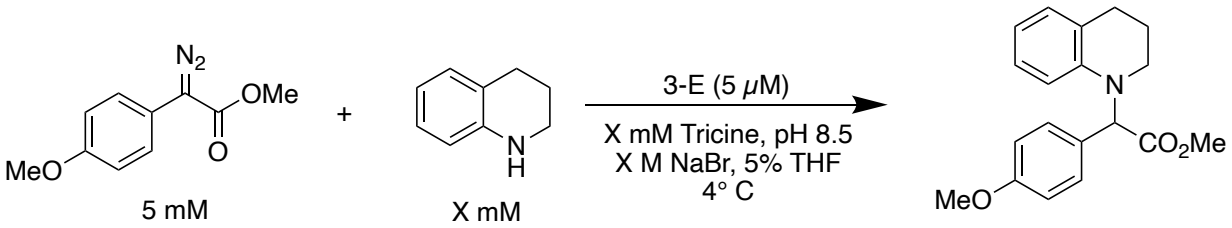


b

Entry	Variant	pH	e.e.
1	GSH	7.5	40%
2	GSH	8.5	43%
3	5-G	7.5	32%
4	5-G	8.5	34%
5	3-L	7.5	60%
6	3-L	8.5	66%

Figure 4.1 pH dependence of N-H functionalization.

Table 4.6 Optimization of reaction conditions.



Entry	[THQ] (mM)	[NaBr] (M)	[Tricine] (mM)	Yield (%)	e.e. (%)
1	25	1.75	100	81	65
2	25	1.75	50	71	62
3	10	1.75	100	86	66
4	10	1.75	50	68	63
5	25	0.5	100	86	39
6	25	1	100	84	49

The pH dependence on selectivity, with pH 8.5 resulting in better enantioselectivity, is unexpected, but has many potential explanations. The first is that the more basic reaction conditions change the rate of the tautomerization of the planar intermediate. Unfortunately, we were unable to find a study that analyzes the tautomerization kinetics of these enamine-enols across a variety of pH values, despite an extensive literature search. These molecules likely have a different rate of tautomerization than typical enols, which of course depend heavily on the pH and solvent. It is therefore challenging to predict how the increased pH will affect this step, though the intermediate would be less likely to diffuse from the chiral protein environment before tautomerization if tautomerization were faster. This is likely an important factor, as the three variants tested at both pH 7.4 and 8.5 had a minor improvement at the latter value (Figure 4.1b).

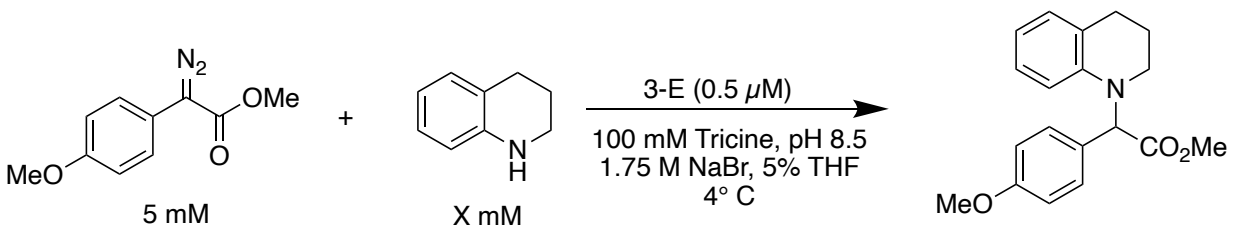
Alternate explanations for the observed increase in enantioselectivity require additional context. As the effect on pH was found for N-H functionalization, we also examined different pH values in the Si-H functionalization of dimethylphenylsilane (see Section 5.2.1.2 for more

information). Interestingly, there is a significant effect of pH on the enantioselectivity of this reaction with both 2-HH and 3-E, where the e.e. was found to increase with increasing pH to a maximum at 8.5, but this was not the case for 1-SGH or 3-H (from the diazo coupling lineage). As only 2-HH and 3-E saw significant pH dependence, the mutations in 2-HH would appear to be responsible for this effect. There are three mutations that were found to be impactful on selectivity (Table 4.4) and would be sensitive to a change in pH: W142Y, W214Y, and L328H. As histidine would have a pK_a near 6, well below the range we observe this effect, the Y142 and Y214 mutations with a phenol pK_a around 10.5 were of greater interest. As they are both protic residues, a possible source of their improved selectivity is simply facilitating the intermediate binding or proton transfer. Additionally, these are both found near the interdomain interface, so a sensible explanation for the observed improvements is that the tyrosine residues better enforce a closed conformation which keeps the prochiral intermediate contained.

In addition to potential dynamic effects, previous work on cyclopropanation had found that the residue at position 142 (tryptophan in 3-VRVH) was modified by the carbenoid during catalysis.[18] Mutation of tryptophan to the more nucleophilic tyrosine, especially when deprotonated at higher pH values, should increase the extent of modification. Therefore, a third hypothesis for the observed improvement is that this modification is somehow advantageous for enantioselective proton transfer. Despite significant effort to demonstrate this was occurring in the ArM scaffolds developed in this lineage, no covalent modification was observed using intact mass spectrometry under any of the reaction conditions examined (Table 4.7). Though modification was not found to be occurring to a significant extent, the mechanism by which selectivity might be altered by modification led us to investigate the ways that the prochiral intermediate may interact with the protein scaffold.

4.2.1.3 Structural Modeling and Mechanism

In order to understand how the ArM scaffold controls an enantioselective proton transfer, we first wanted to understand some of the individual steps of the reaction mechanism. An observation made early on in this project is that the N-H functionalization reaction rate is far lower than for any of the other reactions investigated. As dirhodium tetraacetate complexes are known to bind Lewis bases at the axial positions,[28] a potential explanation for this effect is that substrate binding to the dirhodium inhibits reaction with the diazo substrate. An examination of the initial rates supported this hypothesis, with a significant decrease observed at higher tetrahydroquinoline concentrations (Figure 4.2). An additional difference between N-H functionalization and the other ArM-catalyzed carbene transfer reactions investigated is that very low yields (<5%) of O-H insertion side product are observed for all variants examined (other than 0-ZA4). As discussed in Section 1.3.2, the rate limiting step for stepwise dirhodium catalyzed carbene insertion reactions is the loss of dinitrogen, the rest of the reaction occurs rapidly. The low O-H insertion side product observed suggests that the N-H substrate may be prepositioned in the active site before carbenoid formation occurs, a potential result of the substrate inhibition observed.



[THQ]	mM/min	TOF (s^{-1})
2.5	0.019	0.620
5	0.017	0.561
10	0.012	0.398
25	0.003	0.103

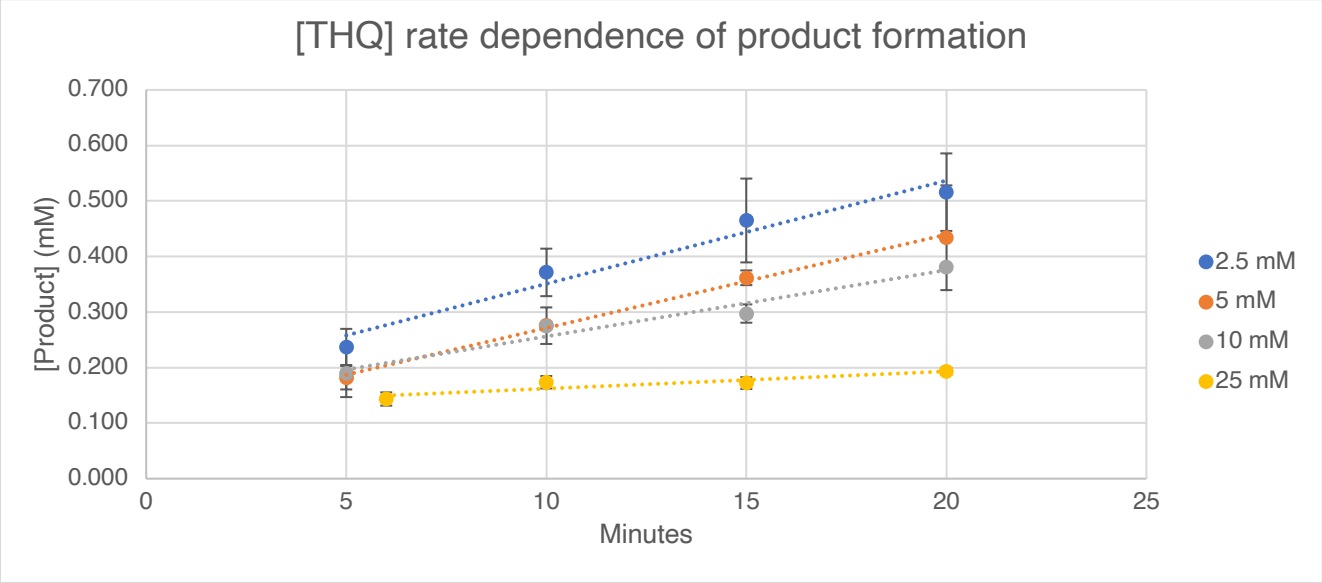
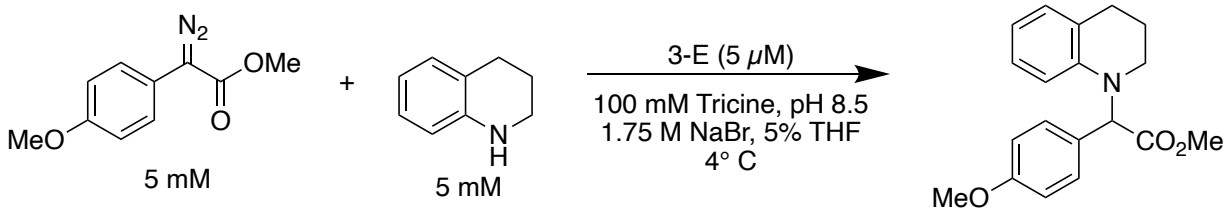


Figure 4.2 Substrate inhibition by tetrahydroquinoline.

Unfortunately, later steps in the reaction mechanism are more difficult to analyze. As discussed in Sections 4.1 and 1.3.2, there have been numerous theoretical investigations showing that the stepwise carbenoid insertion into N-H bonds proceeds through an achiral planar intermediate detached from the dirhodium catalyst when run in organic solvent.[11] In our ArM-catalyzed reactions run in water, however, the mechanism of proton transfer could be substantially altered through a water-facilitated proton transfer during any of the intermediate steps, either

bound or unbound to the dirhodium cofactor.[13,29] An investigation into the enantiospecific O-H functionalization of water facilitated by a chiral phosphoric acid[13,30] found that it was energetically unfavorable (4 kcal/mol) for the phosphate to bind to an enol-dirhodium complex compared to phosphate binding to the dissociated enol. The subsequent phosphoric acid-catalyzed proton transfer was found also found to be most favorable without the involvement of dirhodium or water.[30] Though our N-H functionalization reaction has significant differences from this reaction, we wondered if using phosphate as a buffer in our reaction might bind the dissociated enamine intermediate and alter the mechanism of proton transfer, as the limited theoretical evidence suggests it should not interact favorably with a dirhodium-bound enol complex. As shown in Table 4.7, the use of 50 mM phosphate buffer at pH 7.4 or 8.5 was found to lower the enantioselectivity modestly, which is consistent with a dissociated proton transfer mechanism.

Table 4.7 The effect of phosphate buffer on product enantioselectivity.



Entry	Buffer	pH	Substrate	Yield (%)	e.e. (%)
1	PIPES	7.4	THQ	85	60
2	Phosphate	7.4	THQ	74	49
3	Tricine	8.5	THQ	86	66
4	Phosphate	8.5	THQ	83	52

We sought to determine where potential scaffold-intermediate binding interactions could occur by using a structural modeling/substrate docking approach. As detailed in Section 3.2.5, a collaboration with the Roux group at the University of Chicago resulted in a ~1000 ns simulation

of the final variant in that work, 5-G, with the dirhodium cofactor bound to H326.[25] In this simulation, the two domains remain in the closed position for the entirety of the analysis, so we thought that docking of the intermediate to a frame from this model might illuminate positions that may be involved in intermediate binding. Frame 424 was selected as a representative snapshot, as it had a cofactor orientation that was frequently observed throughout the simulation and an average open/closed angle. Unfortunately, the docking software used in this effort, AutoDock Vina, could not tolerate the cofactor or azidophenylalanine, so these were removed and a mutation at position 477 to tyrosine was made using PyMol. In addition to the variant 5-G that the original simulations were performed on, we wanted to examine the variants obtained from the evolution efforts outlined above. To do this, the mutations corresponding to 1-SGH and 3-E were made in PyMol by selecting the side chain orientation with the lowest clashing. As these models still had significant side chain conflicts, each model (including the initial 5-G) was minimized using the ModRefiner server, a public server from the University of Michigan that was able to substantially reduce clashing, optimize side chain orientations, and improve backbone positioning.[31] Before docking, the THQ-enamine intermediate (Figure 4.3, left) was built and minimized using Spartan. An enamine containing an intramolecular O-H---N bond was found to be the lowest energy conformation in the gas phase. This molecular model was then used for docking in Chimera using AutoDock Vina. The entire scaffold, including both the interior and exterior of the protein, was set within the search radius. This analysis output 10 poses for each variant which were found to cluster into specific pockets around the protein. For clarity, representative poses for each cluster are shown in Figure 4.3 for 5-G (top) and 3-E (bottom).

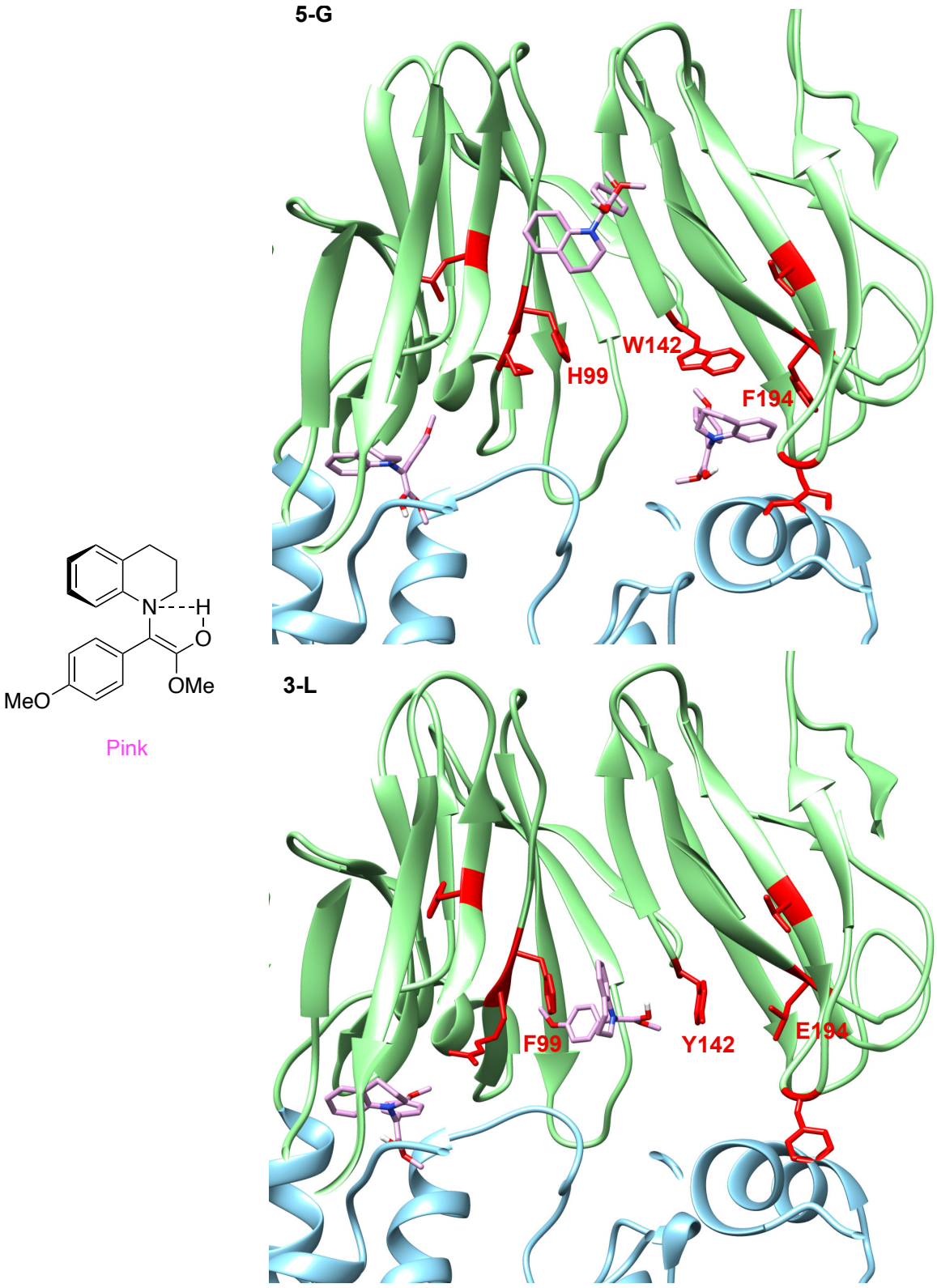


Figure 4.3. 5-G (top) and 3-E (bottom) docking of the THQ-enamine (pink sticks). Residues with different identities between the variants are shown in red. AAs at 99, 142 and 194 are labeled.

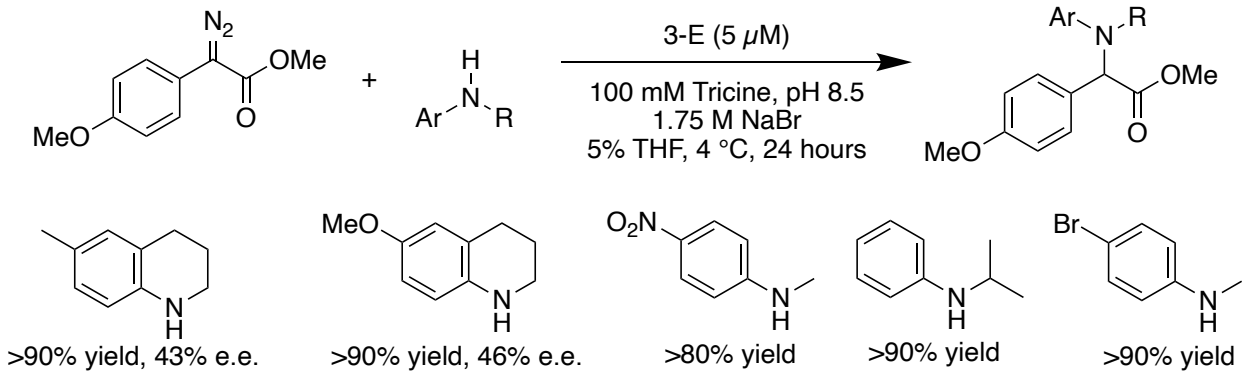
In each of the docking models, there is a cluster in the hinge region near V71, suggesting that the mutations between the variants do not alter this location significantly. The other major binding position near W142, on the other hand, has clear variant dependence. In 5-G the binding occurs between W142 and F194, whereas it is positioned to the other side of Y142, closer to F99, in 3-E. Because 3-E is a more selective catalyst than 5-G, we hypothesized that the binding location observed in 3-E may result in a more selective proton transfer. As there are no amino acid side chains capable of catalyzing the proton transfer pointed towards the enamine, the most likely mechanism is a water-facilitated process. Even still, we hypothesized that a pocket capable of better accommodating the enamine intermediate could orient this species and an active site water to enable selective tautomerization. Improved binding would also decrease the possibility of the enamine leaving the active site entirely and undergoing racemic tautomerization in solution. With this goal in mind, a number of site saturation libraries at were prepared at positions near the potential binding location, including at D119, E120, and G121. From these libraries, only D119 was found to be an impactful site, with mutation to arginine slightly improving enantioselectivity (4-R, Table 4.4, Entry 5). The docking analysis also suggested binding may occur in the hinge region and near the top of the β-propellor domain cavity, both of which are the targets of ongoing engineering efforts.

In addition to these approaches, we wondered if a complete redesign of the F99-Y142 pocket in 3-E would result in greater changes. To do this, we used the Rosetta package Coupled Moves,[32] a program that improves the binding of a ligand to a binding pocket *in silico* by changing the ligand orientation/conformation, making mutations, and repositioning the backbone. This software was used to construct an optimal binding pocket in 20 replicate models. The results of this computational approach indicated that several additional positions may be crucial to

improving binding, including residue 97 and the loop containing 141, 143 and 144. An initial SSM library of L97 did not result in any improved scaffolds, but a combinatorial mutagenesis library on the other 3 residues is currently underway. Additionally, the alternate binding location in the hinge region of the scaffold is being investigated for potential effects on enantioselectivity. The modelling efforts outlined in this section were used to create hypotheses about potential binding locations which will continue to guide library creation in addition to current combinatorial codon mutagenesis approaches.

4.2.1.4 Aniline Substrate Scope

In parallel with this computational analysis, the substrate scope of aniline N-H functionalization was investigated (Scheme 4.1). The yields for these substrates were determined by the relative amount of O-H insertion side product after consumption of the diazo. In all cases except N-methyl-4-nitroaniline, the yields are over 90%, retaining the excellent yields observed using tetrahydroquinoline. As was mentioned in Section 4.1, the nucleophilicity of aniline is considerably higher than that of water, which could explain the conserved chemoselectivity across anilines. Interestingly, the enantioselectivities for the substituted tetrahydroquinolines are lower than that observed for THQ. While not ideal for the present work, these results indicate that there may be substrate specificity involved in the enantiodetermining protonation step, supporting the hypothesis of a scaffold-enamine intermediate complex during protonation.



Scheme 4.1 Aniline substrate scope.

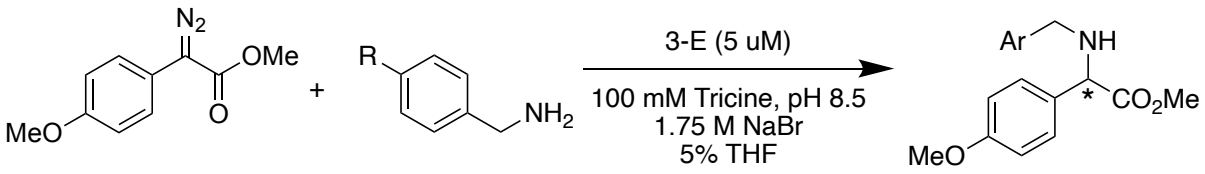
4.2.2 N-H Functionalization of Alkyl Amines

4.2.2.1 Initial Reaction Exploration and Optimization

While the studies outlined above focused on anilines, we also investigated the functionalization of alkyl amine N-H bonds. While the increased nucleophilicity of these compounds relative to anilines would benefit the desired nucleophilic attack on a dirhodium carbenoid intermediate, the increased basicity means that these compounds are protonated under the reaction conditions, ultimately preventing nucleophilic attack. As discussed in Section 4.1, this factor, along with the tighter binding of amines to the dirhodium cofactor, has prevented the use of the typical proton-transfer catalysts that had found success in the functionalization of anilines and amides with dirhodium tetracarboxylate catalysts.[14] There are no reports in the literature that use dirhodium catalysts to perform this type of reaction. Indeed, initial reaction with benzylamine using our dirhodium ArMs at pH 7.4 saw essentially no product formation, consistent with substrate protonation resulting in deactivation. By increasing the reaction pH to 8.5 we would expect less protonation of the amine which could enable the desired transformation to occur. As shown in Table 4.8, this was found to be the case, where the three benzylamines examined had yields that varied in relation to the pK_a of the corresponding ammonium. As there was a clear

improvement in yield with less basic amines, we thought that continuing to increase the pH of the reaction may enable reaction of the more basic benzylamines.

Table 4.8 Benzylamine substrate scope.



Entry	R	pK_a	Yield (%)	e.e. (%)
1	CF_3	8.6 (calc)	67	25
2	H	8.82	30	28
3	OMe	9.3 (calc)	8	8

The results shown in Figure 4.4 are consistent with this hypothesis. Higher reaction pH values were found to improve the yields for all substrates, with the best value trending with the pK_a of the ammonium. The substrate with the highest ammonium pK_a**,** (4-methoxy)benzylamine, performed best at the highest pH examined (pH = 10), while substrate with the highest ammonium pK_a, (4-trifluromethyl)benzylamine, performed best at pH = 9. The enantioselectivities for these substrates also change with pH, with the best pH for yield also typically providing the highest e.e. value. It is notable that the scaffold functions effectively up to pH 10 when (4-methoxy)benzylamine is used. These results show that the interplay between basicity and nucleophilicity is an important factor in the N-H functionalization of alkyl amines and is something that the reaction pH can be used to alter.

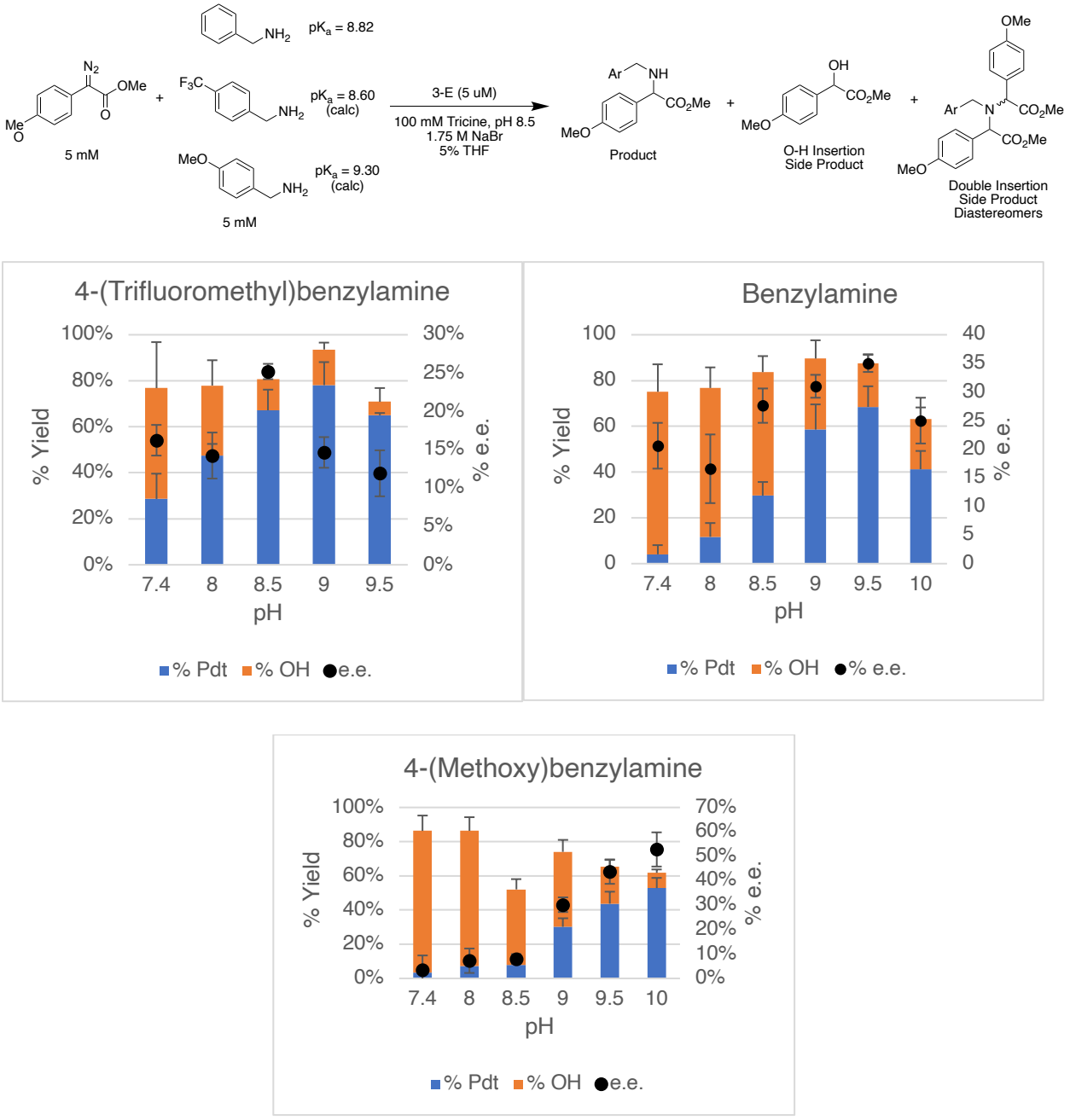


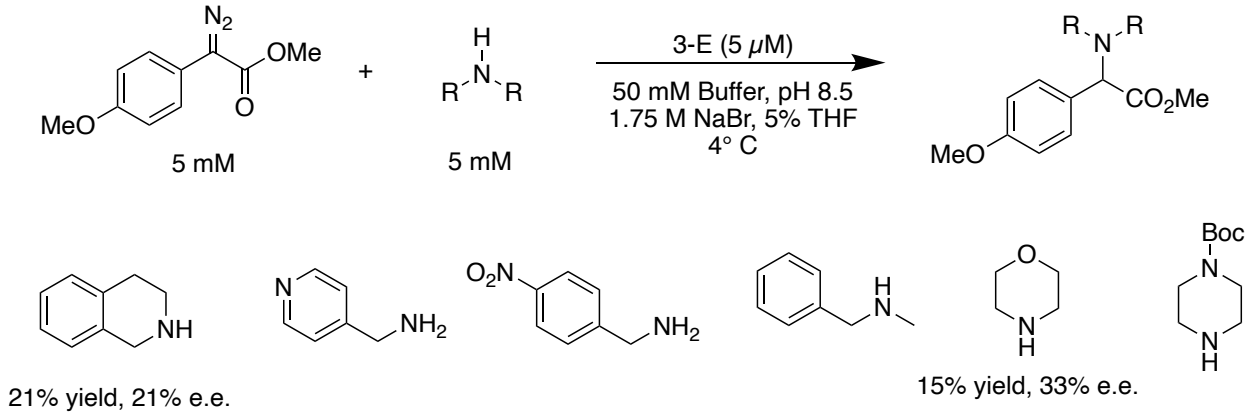
Figure 4.4 Effect of pH on benzylamine N-H functionalization.

The major N-H insertion product was the single N-H functionalized product in most cases, but we also observed significant amounts of the double N-H insertion product diastereomers. Interestingly, increasing the concentration of the amine starting material did not decrease the

relative amount of these over-functionalized products. This observation suggests that there may be binding of the product in the active site, which rapidly undergoes a second addition. These preliminary results warrant a more detailed investigation in the future. Fortunately, the use of secondary amines that contain only a single N-H bond is a straightforward way to circumvent this issue.

4.2.2.2 Amine Substrate Scope

With the clear impact of the amine protonation state established, we wanted to understand the scope of this reactivity (Scheme 4.2). Each of these substrates produce a significant amount of product by LCMS, but due to the lower yields and complex product mixtures, yields cannot be estimated. For the two products that have been confirmed by authentic standard synthesis, 1,2,3,4-tetrahydroisoquinoline and morpholine, low yields and enantioselectivities are observed. The ability of the ArM to control substrate positioning to allow enantioselective protonation of to occur, even if not to a significant extent, make this an excellent starting point for future directed evolution.



Scheme 4.2 Substrate Scope of Alkyl N-H Functionalization.

4.3 Conclusion

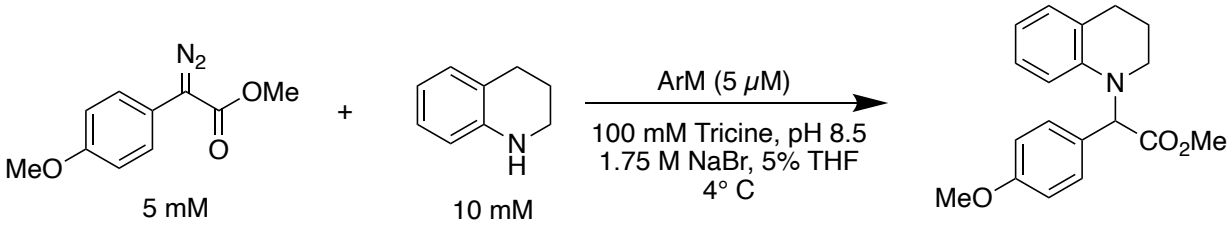
The results outlined above show that POP-based dirhodium ArMs can function as dual-role catalysts, controlling both the nucleophilic attack of the rhodium carbenoid by providing a hydrophobic environment for selective N-H insertion over O-H insertion as well as binding the dissociated prochiral intermediate to enable enantioselective tautomerization. The high nucleophilicity of anilines resulted in excellent chemoselectivity for the desired reaction over water O-H functionalization. This was the case with all variants after 0-ZA_4, where yields of >75% were observed. The enantioselectivity of these transformations was improved from 43% to 68% over 4 rounds of directed evolution. The ArM was also found to accept anilines with a variety of steric and electronic properties, though the enantioselectivity was lower in the cases examined. Further evolution of the scaffold for enantioselective N-H functionalization of THQ is ongoing.

We also probed the ArM-catalyzed reaction to begin to understand how the ArM controls the enantioselective proton transfer. Initial rates of N-H functionalization indicated that substrate inhibition by THQ is occurring, presumably through dirhodium binding. To explore the specific positions that intermediate binding could occur, the active sites of different variants were modeled, and enamine intermediates were docked using Auto Dock Vina. From this analysis, a potential binding site in the β-propellor domain was found. Interestingly, the specific location of the intermediate in the binding pocket was found to shift across the variants, moving closer to the 96-101 β-sheet in 3-E than in the least-enantioselective variant modeled, 5-G. The pocket in 3-E was further optimized for binding of the intermediate using the Rosetta package Coupled Moves. These analyses informed a series of libraries targeting positions that may impact binding at this site, resulting in the variant 4-R and informing future library design decisions. Additionally, evolved variants were shown to catalyze the N-H functionalization of alkyl amines with modest yields and

enantioselectivities. The selective functionalization of these molecules has not previously been reported using dirhodium catalysts, highlighting the utility of using these hybrid catalysts.

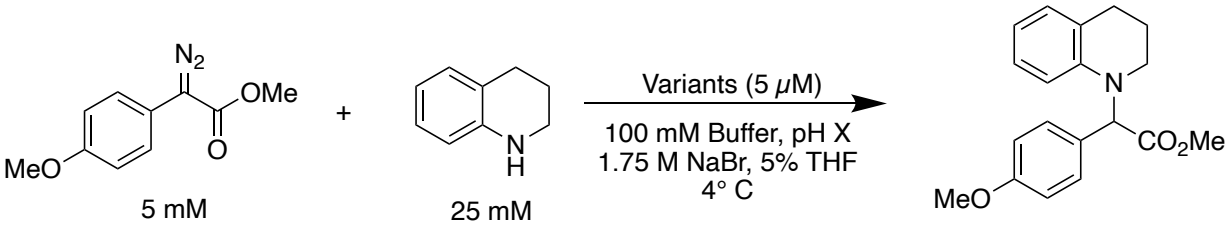
4.4 Experimental

General ArM-catalyzed N-H functionalization Method



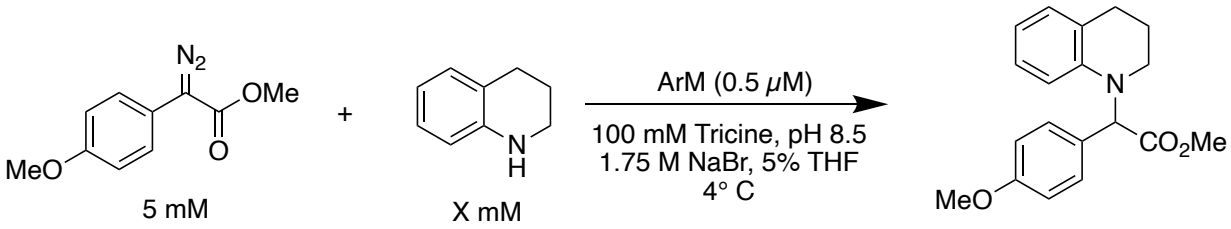
A 285 µL solution of 5 µM ArM (corrected by the known bioconjugation efficiency, typically 90%) in 100 mM Tricine, pH 8.5 with 1.75 M NaBr was prepared in a 1.5 mL microcentrifuge tube. While shaking at 750 rpm at 4°C, 15 µL of the diazo stock solution (0.1 M donor-acceptor diazo, 0.2 M THQ) in THF was added to obtain a final concentration of 5 mM donor-acceptor diazo and 10 mM THQ. The reactions were sealed and briefly inverted to assist with mixing of the substrate, then placed in a thermomixer and shaken at 750 rpm at 4°C for 22 hours. The reaction was quenched by the addition of 50 µL internal standard (30 mM 1,3,5-trimethoxybenzene in ethyl acetate) and 100 µL ethyl acetate. The biphasic mixture was vortexed thoroughly then centrifuged at 20,000 xg for 1 minute to separate the phases. 100 µL of the upper organic layer was collected and filtered for SFC analysis.

Exploration of pH effect on N-H functionalization



A 285 μL solution of 5 μM ArM (corrected by the known bioconjugation efficiency, typically 90%) in 100 mM buffer containing 1.75 M NaBr at varying pH values were prepared in 1.5 mL microcentrifuge tubes. While shaking at 750 rpm at 4°C, 15 μL of the diazo stock solution (0.1 M donor-acceptor diazo, 0.5 M THQ) in THF was added to obtain a final concentration of 5 mM donor-acceptor diazo and 25 mM THQ. The reactions were sealed and briefly inverted to assist with mixing of the substrate, then placed in a thermomixer and shaken at 750 rpm at 4°C for the necessary amount of time. The reaction was quenched by the addition of 50 μL internal standard (30 mM 1,3,5-trimethoxybenzene in ethyl acetate) and 100 μL ethyl acetate, then the tube was vortexed and spun down for phase separation. 100 μL of the upper organic layer was collected and filtered for SFC analysis.

Determination of Initial Rates of ArM-catalyzed N-H functionalization



A 285 μL solution of 0.5 μM ArM (corrected by the known bioconjugation efficiency, typically 90%) in 100 mM Tricine, pH 8.5 with 1.75 M NaBr was prepared in a 1.5 mL microcentrifuge tube. While shaking at 750 rpm at 4°C, 15 μL of the diazo stock solution (0.1 M

donor-acceptor diazo, X M THQ) in THF was added to obtain a final concentration of 5 mM donor-acceptor diazo and X mM THQ. The reactions were sealed and briefly inverted to assist with mixing of the substrate, then placed in a thermomixer and shaken at 750 rpm at 4°C for the necessary amount of time. The reaction was quenched by the addition of 50 μL internal standard (30 mM 1,3,5-trimethoxybenzene in ethyl acetate) and 100 μL ethyl acetate and vortexed and spun down for phase separation. No additional EA was added. 100 μL of the upper organic layer was collected and filtered for SFC analysis as quickly as possible.

ArM library screening

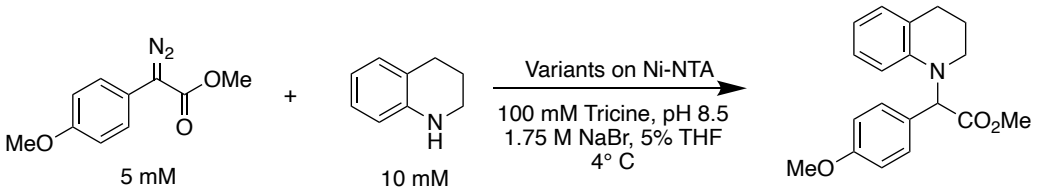
Libraries were constructed as described previously (Section 3.4.6). When CCM was used for diversification in the first round of directed evolution, an average of 2.1 residue mutations were made per gene. The variants were incubated, overexpressed and lyzed as described previously in Chapter 3, using 5 mL overexpression cultures in 24-well plates. The lysate was pelleted by centrifugation at 3,600 rpm and 480 μL of the clarified lysate from each well was transferred to a 96-deep well filter plate. 120 μL of cofactor **1** stock solution in ACN (12.8 μM, 0.010 mg/mL) was added to each well while shaking at 650 rpm at 4 °C. The mixture was incubated at 4 °C, 750 rpm for one hour, at which point 100 μL of 50% Ni-NTA resin was added to each well. ~1 cm of the multichannel pipette tips were cut off using scissors to allow easier transfer of the resin. The plate was shaken at 4° C for 5 minutes to allow protein binding, then centrifuged at 2500 rpm for 5-10 minutes until the resin in each well was dry. The resin was washed to remove excess cofactor by the addition of 500 μL of 50% acetonitrile/50 mM Tris buffer to each well. The plate was filtered by centrifugation at 3000 rpm for 3 minutes to remove the buffer. Three washes were done in total (1.5 mL total wash buffer) with a final centrifugation time of 5 minutes to ensure each well

was dry.

285 μL reaction buffer (100 mM Tricine, pH 8.5 with 1.75 M NaBr) was added to each well and the plate was shaken at 4 °C, 750 rpm for 10 minutes. After this cooling period, 15 μL of substrate stock solution was added (final concentration of 5 mM methyl (4-methoxyphenyl)diazoacetate and 10 mM THQ) at 4°C with 750 rpm shaking. The screening plate was sealed with a 96-well rubber mat to prevent evaporation and shaken at these conditions for 16 hours. The reaction was quenched by addition of 50 μL of 30 mM 1,3,5-trimethoxybenzene in ethyl acetate as internal standard. The plate filtered by centrifugation at 3,000 rpm for 5 min. To remove all organics from the resin, two additional filtrations of 75 μL ethyl acetate were applied and the biphasic mixture was clarified via centrifugation at 3,000 rpm for 5 min. 100 μL of the upper organic portion from each well were transferred to a 96-well microtiter plate for SFC analysis. Mutants displaying a higher product/internal ratio or improved E/Z isomer ratio than the parent were selected for DNA sequencing and validation. Each library contained four parent wells which were used to calculate an average and standard deviation, which are shown in the table below for the 3 libraries that generated positive variants.

Library	Parent e.e. w/ purified ArM	Parent e.e. in plate		
1-SGH CCM	43%	27%	+/-	0.5
2-HH 194NNK	60%	60%	+/-	1.1
3-E 119NNK	66%	62%	+/-	0.5

Example Data from 2-HH-194NNK Library



Product/IS:

	1	2	3	4	5	6	7	8	9	10	11	12
A	N/A	N/A	N/A	N/A	0.29	0.43	0.57	0.48	0.27	0.49	0.13	0.40
B	0.45	0.15	0.59	0.47	0.47	0.46	0.58	0.14	0.14	0.44	0.54	0.11
C	0.13	0.46	0.36	0.58	0.53	0.12	0.12	0.71	0.62	0.39	0.11	0.52
D	0.52	0.24	0.60	0.57	0.11	0.59	0.16	0.56	0.60	0.12	0.54	0.47
E	0.75	0.77	0.80	0.26	0.79	0.13	0.12	0.71	0.46	0.55	0.74	0.13
F	0.76	0.13	0.69	0.66	0.14	0.72	0.13	0.76	0.64	0.71	0.56	0.66
G	0.64	0.72	0.74	0.72	0.72	0.58	0.69	0.64	0.47	0.63	0.68	0.63
H	0.19	0.18	0.68	0.12	0.68	0.00	0.66	0.14	0.64	0.12	0.66	0.61

Enantioselectivity:

	1	2	3	4	5	6	7	8	9	10	11	12
A	N/A	N/A	N/A	29	43	53	59	57	33	58	-11	49
B	58	-10	60	53	54	57	60	-12	-12	52	59	4
C	-12	58	48	59	59	-14	-19	61	61	47	-15	58
D	54	35	60	55	-18	61	-16	52	57	-17	52	55
E	66	65	66	22	66	-19	-19	63	54	58	63	-17
F	65	-19	63	61	-19	63	-19	65	60	63	58	62
G	61	61	63	61	63	56	61	58	56	61	60	57
H	-18	-13	62	-19	60	0	60	-15	60	60	59	58

Docking analysis

The model of 5-G with Rh-His bond which was previously prepared for diazo coupling (Section 3.2.5) was used as the starting point for this analysis. Frame 424 was selected as representative, so this frame was saved as an individual structure for modification. Using PyMol, the cofactor was deleted and the mutation Z477Y was made using the Mutate function. The side

chain orientation with the lowest clashing was selected. To prepare the 1-SGH and 3-E models, the necessary mutations were made using a similar method. Each of these models were minimized using the ModRefiner server (https://zhanggroup.org/ModRefiner/). The minimized structures were then prepared for docking in Chimera using the Dock Prep function. The THQ-enamine and benzylamine-enamines optimized using Spartan DFT in gas phase..mol2 files of the were placed into the interior of the scaffold and docking was performed using the default settings with AutoDock Vina. The search volume was configured to contain the entire protein and 10 poses were found for each model. From the clusters of poses that were found in this analysis, a single representative pose was selected for the simplified figures.

4.5 Acknowledgements

Once again, Rui Huang played a significant role in the progress that was made in this chapter. His efforts in cloning and library design were critical in pushing this project forward, for which I am extremely grateful. I would like to thank Yasmine Zubi for considerable assistance with Coupled Moves. Her knowledge (and patience) was critical in my ability to perform those simulations. Additionally, I am grateful to the work that Lucía Cores Sarría put into writing scripts to aid in the analysis of the Coupled Moves results. Many of the trends we found were thanks to her work. Finally, I would like to thank Bingqing Liu and Prabir Saha for their assistence in the synthesis of some of the compounds used throughout this chapter.

4.6 References

[1] J.-H. Xie, S.-F. Zhu, Q.-L. Zhou, *Chem Rev* 2011, *111*, 1713–1760.

[2] T. C. Nugent, M. El-Shazly, *Adv Synth Catal* 2010, *352*, 753–819.

[3] Q. Yin, Y. Shi, J. Wang, X. Zhang, *Chem Soc Rev* 2020, *49*, 6141–6153.

[4] I. Slabu, J. L. Galman, R. C. Lloyd, N. J. Turner, *Acs Catal* 2017, *7*, 8263–8284.

[5] H. Kohls, F. Steffen-Munsberg, M. Höhne, *Curr Opin Chem Biol* 2014, *19*, 180–192.

[6] V. F. Batista, J. L. Galman, D. C. G. A. Pinto, A. M. S. Silva, N. J. Turner, *Acs Catal* 2018, *8*, 11889–11907.

[7] C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, P. N. Devine, G. W. Huisman, G. J. Hughes, *Science* 2010, *329*, 305–309.

[8] D. Gillingham, N. Fei, *Chem Soc Rev* 2013, *42*, 4918–4931.

[9] P. Bulugahapitiya, Y. Landais, L. Parra-Rapado, D. Planchenault, V. Weber, *J Org Chem* 1997, *62*, 1630–1641.

[10] R. T. Buck, C. J. Moody*, A. G. Pepper, *ARKIVOC* 2002, 16–33.

[11] X.-C. Wang, X.-S. Song, L.-P. Guo, D. Qu, Z.-Z. Xie, F. Verpoort, J. Cao, *Organometallics* 2014, *33*, 4042–4050.

[12] Y. Liang, H. Zhou, Z.-X. Yu, *J Am Chem Soc* 2009, *131*, 17783–17785.

[13] Z.-Z. Xie, W.-J. Liao, J. Cao, L.-P. Guo, F. Verpoort, W. Fang, *Organometallics* 2014, *33*, 2448–2456.

[14] M.-L. Li, J.-H. Yu, Y.-H. Li, S.-F. Zhu, Q.-L. Zhou, *Science* 2019, *366*, 990–994.

[15] Z. Liu, C. Calvó-Tusell, A. Z. Zhou, K. Chen, M. Garcia-Borràs, F. H. Arnold, *Nat Chem* 2021, 1–7.

[16] H. Saito, T. Uchiyama, M. Miyake, M. Anada, S. Hashimoto, T. Takabatake, S. Miyairi, *Heterocycles* 2010, *81*, 1149.

[17] B. Xu, S. Zhu, X. Xie, J. Shen, Q. Zhou, *Angewandte Chemie Int Ed* 2011, *50*, 11483–11486.

[18] H. Yang, A. M. Swartz, H. J. Park, P. Srivastava, K. Ellis-Guardiola, D. M. Upp, G. Lee, K. Belsare, Y. Gu, C. Zhang, R. E. Moellering, J. C. Lewis, *Nat Chem* 2018, *10*, 318–324.

[19] F. Brotzel, Y. C. Chu, H. Mayr, *J Org Chem* 2007, *72*, 3679–3688.

[20] J. C. Lewis, *Accounts Chem Res* 2019, *52*, 576–584.

[21] L. Villarino, S. Chordia, L. Alonso-Cotchico, E. Reddem, Z. Zhou, A. M. W. H. Thunnissen, J.-D. Maréchal, G. Roelfes, *Acs Catal* 2020, *10*, 11783–11790.

[22] A. B. Taylor, R. M. Czerwinski, W. H. Johnson, C. P. Whitman, M. L. Hackert, *Biochemistry-us* 1998, *37*, 14692–14700.

[23] F. Schwizer, Y. Okamoto, T. Heinisch, Y. Gu, M. M. Pellizzoni, V. Lebrun, R. Reuter, V. Köhler, J. C. Lewis, T. R. Ward, *Chem Rev* 2018, *118*, 142–231.

[24] P. Srivastava, H. Yang, K. Ellis-Guardiola, J. C. Lewis, *Nat Commun* 2015, *6*, 7789.

[25] D. M. Upp, R. Huang, Y. Li, M. J. Bultman, B. Roux, J. C. Lewis, *Angewandte Chemie Int Ed* 2021, *60*, 23672–23677.

[26] J. D. Bloom, F. H. Arnold, *Proc National Acad Sci* 2009, *106*, 9995–10000.

[27] K. D. Belsare, M. C. Andorfer, F. S. Cardenas, J. R. Chael, H. J. Park, J. C. Lewis, *Acs Synth Biol* 2017, *6*, 416–420.

[28] J. F. Berry, *Dalton T* 2011, *41*, 700–713.

[29] J. Wu, X. Li, X. Qi, X. Duan, W. L. Cracraft, I. A. Guzei, P. Liu, W. Tang, *J Am Chem Soc* 2019, *141*, 19902–19910.

[30] Y. Zhang, Y. Yao, L. He, Y. Liu, L. Shi, *Adv Synthesis Amp Catal* 2017, *359*, 2754–2761.

[31] D. Xu, Y. Zhang, *Biophys J* 2011, *101*, 2525–2534.

[32] N. Ollikainen, R. M. de Jong, T. Kortemme, *Plos Comput Biol* 2015, *11*, e1004335.

Chapter 5: Engineering Dirhodium ArMs for Concerted Carbene Insertion Reactions

5.1 Introduction

Carbon-hydrogen bonds are ubiquitous in organic molecules, making the selective functionalization of these bonds a goal that would greatly simplify the synthesis of organic molecules.[1–3] In place of complex functional group manipulation sequences that are currently necessary in standard organic synthesis, C-H bonds could be used directly as handles to forge carbon-carbon and carbon-heteroatom bonds. Eliminating the need for these steps would signficantly improve atom economy and enable late-stage diversification of complex molecules.[4] While the universality of C-H bonds gives rise to these possibilities, it also necessitates the development of catalysts capable of functionalizing specific C-H bonds among the many are present on complex substrates. The majority of recent progress toward this end has focused on metal catalysts that coordinate to a functional group on the substrate in order to direct the metal towards reactivity at a single position.[5,6]

Despite the prevalence of directed C-H functionalization in synthetic methodology, there are notable examples of non-directed C-H functionalization[7] using catalysts that can discriminate subtle steric,[8] electronic,[9] or stereoelectronic[10,11] properties of different C-H bonds. Dirhodium tetracarboxylate catalysts have proven particularly useful for non-directed carbene insertion into C-H bonds, as the high reactivity of the dirhodium carbenoid intermediate involved in these transformations enables insertion into primary, secondary, and tertiary sp^3 sites of diverse molecules.[11] As shown in Figure 1.6, dirhodium carbenoid insertion into a non-polar bond proceeds through an asymmetric, concerted type mechanism.[12,13] Controlling the specificity of this reaction therefore requires precise positioning of the substrate with respect to the carbenoid to

permit reaction at only a single C-H bond. Remarkable examples from the Davies group[14–16] demonstrate that catalyst-controlled site- and sterero-specificity can be achieved on simple alkane substrates when sufficiently bulky ligands are used. In these cases, extensive ligand design was necessary to optimize the catalyst structure for the steric and electronic differentiation of C-H bonds on the substrate.[17] Moreover, the selectivity demonstrated in these systems results from repulsive substrate/catalyst interactions that prevent reaction with undesired C-H bonds, rather than facilitating reaction at desired sites. This has led to dramatically lower rates (60-fold lower TON/hr) for highly selective catalysts.[14]

Despite these impressive advances in catalytic C-H bond functionalization, controlling and, just as importantly, tuning the selectivity of a catalyst to functionalize different C-H bonds remains challenging. The synthetic power of C-H functionalization thus remains best illustrated in the biosynthesis of natural products, which frequently involve enzyme-catalyzed functionalization of unactivated C-H bonds on both simple and complex molecular frameworks.[18,19] For example, the biosynthesis of the diterpenoid taxol involves more than 6 site and stereoselective C-H oxidation reactions after the carbon skeleton is assembled.[20–22] While natural enzymes are proficient in the context of their natural biosynthesis, the use of these catalysts for chemical synthesis requires an expanded substrate scope and selectivity. The use of repurposed and artificial metalloenzymes for C-H functionalization is covered in a perspective by the author.[23]

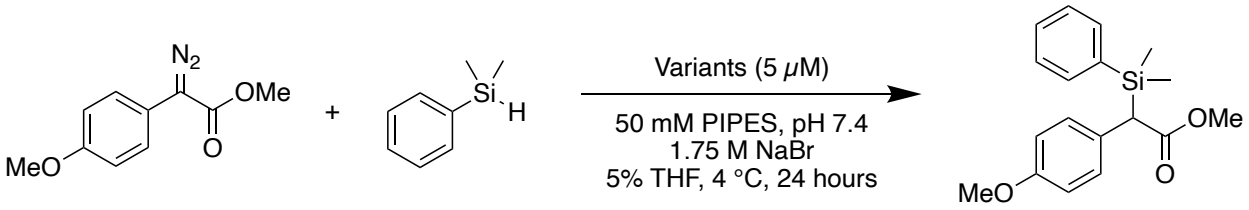
Directed evolution can, of course, be used to achieve selectivity within the reactivity of natural enzymes.[24–26] To expand outside of natural reactivity, artificial metalloenzymes have been designed to make use of scaffold-enabled molecular recognition for non-native chemistry.[27,28] The dirhodium-POP ArMs discussed throughout this dissertation exhibit this type of control, enabling enantio- and chemo-selectivity for a range of reactions that can be improved via directed

evolution.[29,30] Unfortunately, progress towards selective C-H functionalization using dirhodium ArMs has not been as straightforward. As discussed in depth in Section 2.2.4, 3-VRVH catalyzed the C-H functionalization of the benzylether bonds of phthalan with low yields and poor selectivity when using methyl phenyldiazoacetates. Due to the increased reactivity of trifluoroethyl phenyldiazoacetates, the ArM was able to functionalize activated C-H bonds with modest yields but no enantioselectivity using these compounds. Attempts at that time to evolve the scaffold for improved yields using either of these diazos were unsuccessful due to high background in all the screening conditions examined (Section 2.2.5).

On the other hand, the concerted insertion into Si-H bonds, which proceeds through a similar mechanism to C-H functionalization,[31] was controlled by the ArM scaffold with modest selectivity. As detailed in Section 2.2.3, we had found that 3-VRVH catalyzed the Si-H functionalization of dimethylphenylsilane with 64% e.e., albeit with poor chemoselectivity over water insertion (Table 5.1). The challenge of chemoselectivity likely arises from the relative rates of the competing reactions, with nucleophilic attack by the solvent, water, outcompeting the slower Si-H functionalization reaction. As this is the same fundamental problem that C-H functionalization likely faces, we decided to use Si-H functionalization as a model reaction to improve the scaffold for concerted carbene insertion into non-polar R-H bonds, expecting that variants with improved yields and selectivities for Si-H functionalization would also be better for the functionalization of C-H bonds. The resulting ArM could then be directly evolved on substrates with C-H bonds in place of Si-H bonds. This approach is analogous to substrate walking, in which directed evolution is conducted using model substrates increasing similar to a target substrate that extant enzymes do not accept, except that the group undergoing reaction, rather than the substrate periphery, is being change during the evolution effort. This chapter describes the improvement of

the scaffold for Si-H functionalization as well as other approaches to achieve C-H functionalization through substrate design and reaction conditions.

Table 5.1 Si-H functionalization catalyzed by the cyclopropanation lineage.



Entry	Variant	Yield (%)	e.e. (%)
1	0-ZA_4	45	6
2	1-NAGS	43	14
3	3-VRVH	35	64
4	1-RFY	40	-30

5.2 Results and Discussion

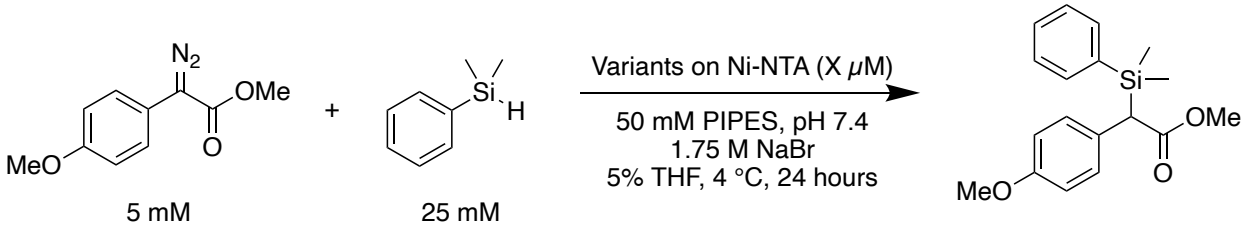
5.2.1 Engineering ArMs for Si-H Functionalization

5.2.1.1 High Throughput Screening Validation

Initial attempts to evolve ArMs for Si-H functionalization used the original Ni-NTA-based screening method developed for cyclopropanation. The best scaffolds at the time for this reaction, 3-VRVH and 1-SGH, resulted in low yields (<10%) and unexpectedly poor enantioselectivities (~30%) under the screening conditions. As these results were significantly worse than when the purified ArMs were used, we were concerned that the screening method would not be able to accurately identify improved variants. During other evolution efforts (Section 3.2.2), we developed scaffolds that were significantly improved for diazo coupling, so we wondered if these newer scaffolds might achieve better catalytic performance under screening conditions. The best variant from that lineage, 5-G, was compared to 1-SGH in an experiment mimicking the screening

conditions, where a controlled amount of purified ArM was loaded onto Ni-NTA resin rather than the unknown quantity of ArM bioconjugated in lysate that occurs during typical screening. In the final 300 μL reaction volumes, between 0.5-3 μM of POP scaffold is typically obtained from the 5 mL expression cultures used in screening, so the reaction was performed using 0.5 μM ArM. Encouragingly, 5-G performed far better than 1-SGH under these conditions (Table 5.2, Entries 3 and 4) and provided the desired product in 31% yield and 72% enantioselectivity with 0.5 μM ArM, results comparable to those obtained using 5 μM purified ArM. As expected based on previous work, 1-SGH was considerably worse than 5-G under these conditions, even though they performed similarly in homogeneous conditions (Table 5.4). While 0.1 μM ArM is lower than what we would expect the ArM concentration to be using this method, the smaller drop in enantioselectivity for 5-G at this concentration was a positive sign for the robustness of future screening efforts (Table 5.2, Entry 2).

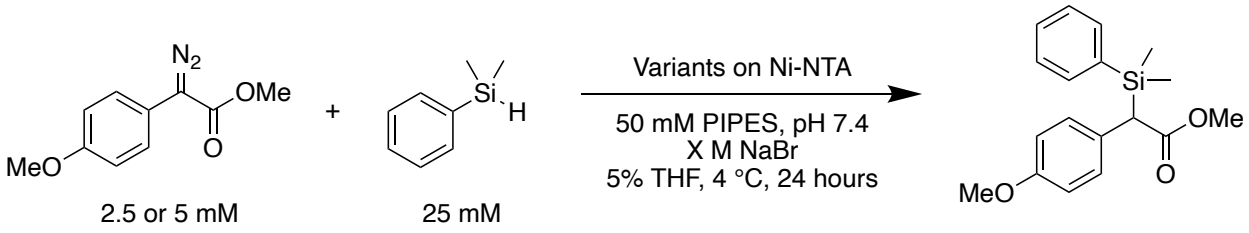
Table 5.2 Ni-NTA method with controlled ArM loading.



Entry	Variant	[ArM] (uM)	Yield (%)	e.e. (%)
1	1-SGH	0.1	3	26
2	5-G	0.1	8	60
3	1-SGH	0.5	11	46
4	5-G	0.5	31	72

These results were encouraging, but we needed to verify that this improvement would carry over to scaffold that was expressed under the screening conditions and bioconjugated in lysate. To this end, both 1-SGH and 5-G were expressed, bioconjugated, and bound to Ni-NTA using the standard screening methodology. Si-H functionalization reactions using either 1.75 M or 0.7 M sodium bromide and 2.5 or 5 mM methyl (4-methoxyphenyl)diazoacetate were then conducted in order to explore these key variables, which were previously found to affect cyclopropanation[29,30,32] and diazo cross-coupling[30] (Table 5.3).

Table 5.3 Validation and optimization of Ni-NTA screening methodology.



Entry	Variant	[N_2] (mM)	[NaBr] (M)	Yield %	Yield +/-	e.e. %	e.e. +/-
1	1-SGH	5	1.75	20	8	17	5
2	5-G	5	1.75	32	2	77	2
3	1-SGH	2.5	1.75	17	4	31	7
4	5-G	2.5	1.75	29	2	80	3
5	1-SGH	5	0.7	13	5	21	7
6	5-G	5	0.7	34	2	74	4
7	1-SGH	2.5	0.7	18	5	24	1
8	5-G	2.5	0.7	30	4	75	1

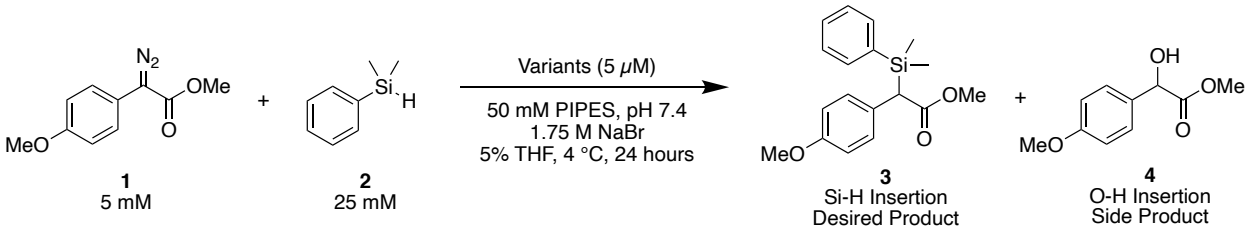
The superiority of 5-G over 1-SGH under screening conditions is even more apparent here, where both the yield and enantioselectivity of 5-G are comparable to the results using purified ArM under homogenous reaction conditions (Table 5.4). Additionally, the standard deviations of both the yields and enantioselectivities of 5-G were significantly lower than 1-SGH

under essentially all conditions tested. The low standard deviations means that this screen is reasonably robust, an important factor since validation of hits is very time consuming. This experiment also revealed that good activity could be observed across a range of reaction conditions, which enabled screening in the optimized conditions with 0.5 M NaBr as discussed below.

5.2.1.2 Parent Selection and Reaction Optimization

With a method that was sufficiently reliable for detecting the Si-H functionalization activity and selectivity of 5-G, we needed to select the best parent for evolution.The selectivity and yield of Si-H functionalization catalyzed by ArMs developed for the cyclopropanation,[33] diazo coupling,[30] and N-H functionalization lineages (Table 5.4, Entries 1-7) outlined above using the previously developed standard reaction conditions. Notably, the yield for 3-VRVH in this experiment (23%) is lower than previously reported (35%, Table 5.1, Entry 3), but it also achieves a higher enantioselectivity in this case, 70% compared to 64%. A likely cause of these differences are the many improvements to ArM preparation method that were made to increase the reproducibility and quality of dirhodium ArMs following our initial publications.[29,32] These modifications are documented in Chapter 3 of this thesis. From this panel of ArMs, variant 3-H from the diazo coupling lineage was the most selective, achieving 91% enantioselectivity with a yield of 24%. The majority of the diazo (67%) was converted to the O-H insertion side product **4**, resulting in a poor **3**/**4** ratio of 0.36 that would need to be significantly improved to demonstrate high scaffold-based control over the secondary coordination sphere.

Table 5.4 Exploration of a panel of ArMs for Si-H functionalization.

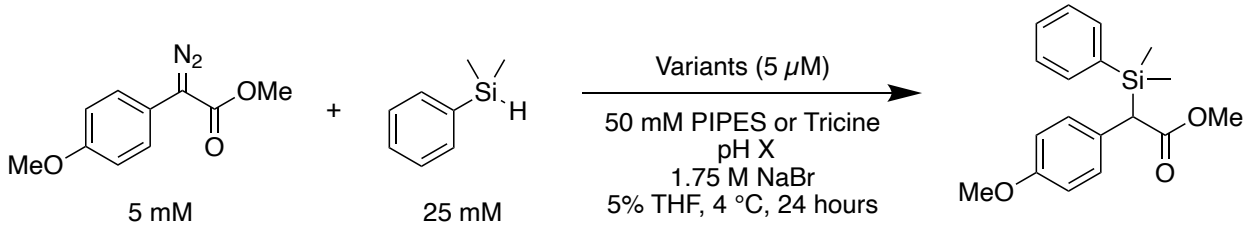


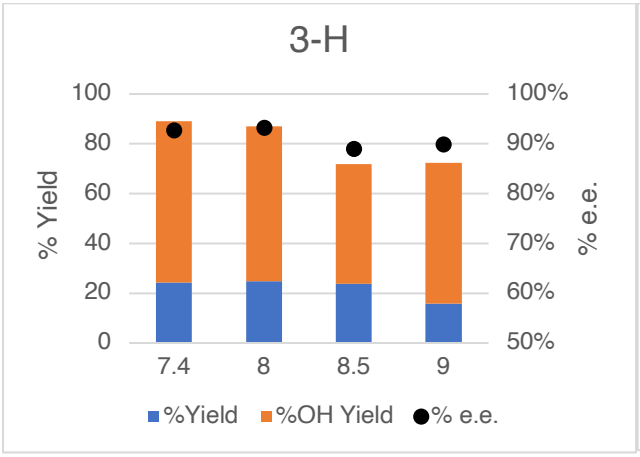
				3		4	
Entry	Variant	Lineage	pH	Yield (%)	e.e. (%)	Yield (%)	**3**/**4**
1	3-VRVH	Cyclopropanation	7.4	23	70	66	0.35
2	1-SGH	Cyclopropanation	7.4	24	72	67	0.36
3	3-H	Diazo coupling	7.4	24	91	66	0.36
4	4-G	Diazo coupling	7.4	25	87	64	0.39
5	5-G	Diazo coupling	7.4	28	80	58	0.48
6	2-HH	N-H Functionalization	7.4	26	74	69	0.38
7	3-E	N-H Functionalization	7.4	23	70	57	0.40
8	1-GSH	Cyclopropanation	8.5	22	79	48	0.46
9	5-G	Diazo coupling	8.5	23	80	47	0.49
10	3-E	N-H Functionalization	8.5	24	89	48	0.50

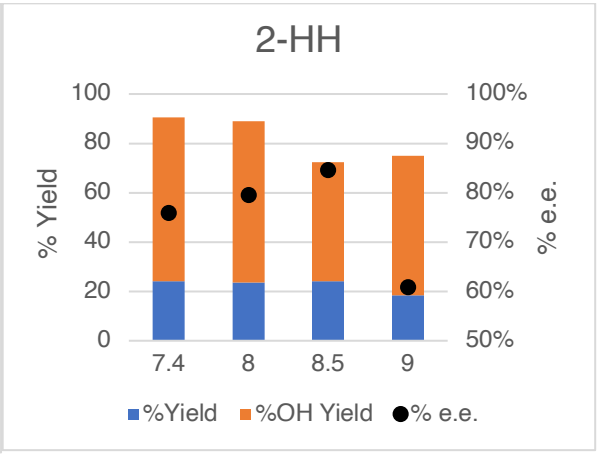
As 3-E was found to be more selective for N-H functionalization at pH 8.5 than 7.4 (Section 4.2.1.2), we examined this variant as well as 1-SGH and 5-G at this higher pH value (Table 5.4, Entries 8-10). Interestingly, the enantioselectivity of 3-E significantly increased at pH 8.5 while 1-SGH had only a minor improvement and 5-G had no change at all, suggesting that the mutations contained in the 3-E scaffold may be responsible for this pH-dependence. Unlike N-H functionalization, there is no proton transfer intermediate step that may be affected by the pH, so we wanted to probe this effect for Si-H functionalization further.

To do this, 3-H, 2-HH, and 3-E were used to catalyze Si-H functionalization reactions at 4 pH values ranging from 7.4 to 9 (Figure 5.1). Not surprisingly, only minor yield and selectivity differences were observed for 3-H, a variant evolved for diazo coupling at pH 7.4. For 2-HH and 3-E, on the other hand, there was a clear pH-dependent trend in enantioselectivity. In both cases,

the selectivity increased from pH 7.4 to 8.5, then dropped substantially at pH 9. Because this reaction proceeds through a single concerted step, the change in selectivity suggests that the higher pH causes a change in the secondary coordination sphere of 2-HH and 3-E (discussed in depth in Section 4.2.1.2). Despite this improved enantioselectivity at pH 8.5, neither 2-HH or 3-E achieved higher selectivity than 3-H at pH 7.4, so this variant was selected as the parent for further reaction optimization and evolution.







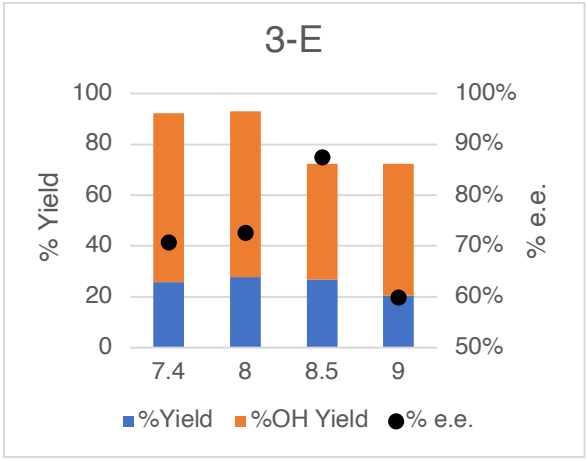
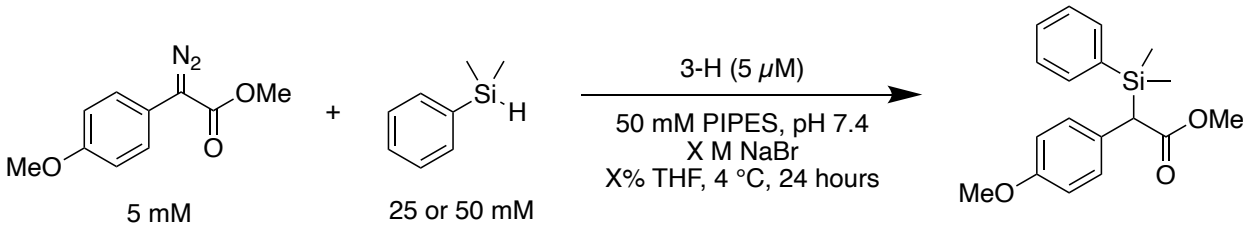


Figure 5.1 The effect of pH on yield and enantioselectivity of Si-H functionalization.

Optimization of other aspects of the reaction conditions was then performed for 3-H. The concentration of the PIPES buffer did not substantially alter selectivity between 10 and 200 mM, so the 50 mM concentration previously used for further study. Sodium bromide, cosolvent percentage, and silane stoichiometry did influence the reaction, however, so each of these was optimized (Table 5.5). Increasing the THF percentage slightly decreased the product yield, so we continued to use 5% THF (Table 5.5, Entries 1-3). Unexpectedly, lower sodium bromide concentrations were found to slightly increase yield without sacrificing enantioselectivity, with 0.5 M NaBr being optimal. This is an observation that had goes counter to our experience in cyclopropanation,[29,32] diazo coupling,[30] and N-H functionalization (Section 4.2.1.2), since these reactions were found to be most selective and highest-yielding with 1.75 M sodium bromide. It is not clear what the cause of this difference is, but it is notable that 3-H was evolved for improved diazo coupling using 0.7 M sodium bromide. Finally, the silane stoichiometry was increased to 10 equivalents (50 mM). This resulted in slightly higher yields at both 0.3 and 0.5 M NaBr, but the enantioselectivity was found to drop slightly, to 88% and 87%, respectively. With sufficiently optimized conditions and a highly selective scaffold in hand, we turned to evolution to improve the yield and chemoselectivity of the reaction.

Table 5.5 Further optimization of reaction conditions.



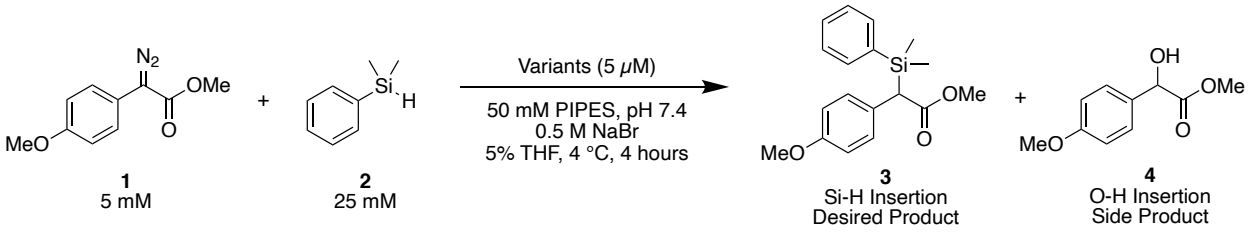
Entry	[NaBr]	% THF	[SiH]	Yield (%)	e.e. (%)
1	1.75	5	25	22	93
2	1.75	10	25	21	94
3	1.75	15	25	19	91
4	0.8	5	25	22	93
5	0.5	5	25	24	93
6	0.3	5	25	24	91
7	0.5	5	50	27	88
8	0.3	5	50	25	87

5.2.1.3 Directed Evolution of Si-H Chemoselectivity

The lower salt concentration of 0.5 M NaBr did not negatively impact the reliability of the Ni-NTA-based high throughput screening method, so only this change was made from the previously optimized protocol. As we had previously had success using combinatorial codon mutagenesis (CCM) of residues facing the interior of the β-propellor domain, this approach was once again used for scaffold diversification. This method proved efficient for generating improved variants, with multiple hits found from the 92 mutants screened. The best of these was determined to be a variant with E241S and L328H mutations (2-SH, Table 5.6, Entry 2). As was the case in N-H functionalization, an additional histidine positioned directed above H326 was found to improve the scaffold for carbene transfer reactions. The E241L mutation also increased the yield and **3**/**4** ratio in a deconvolution analysis, so a site saturation mutagenesis library using E241NNK degeneracy was constructed and screened. Mutation from serine to aspartate, forming the variant

3-D, slightly improved yield and enantioselectivity (Table 5.6, Entry 3). Another round of CCM was performed, and the variant 4-LIY was found to achieve improved yield (38%) and **3**/**4** (0.64, Table 5.6, Entry 4). While there is still a considerable amount of diazo starting material **2** converted to O-H insertion side product **4**, 4-LIY catalyzes 383 total turnovers to **3**, a high value in the context of artificial metalloenzyme catalysis. Further evolution of this variant is underway to continue improving the yield and chemoselectivity of the scaffold for Si-H functionalization as well as building a better starting point for future C-H functionalization efforts.

Table 5.6 Summary of directed evolution lineage.

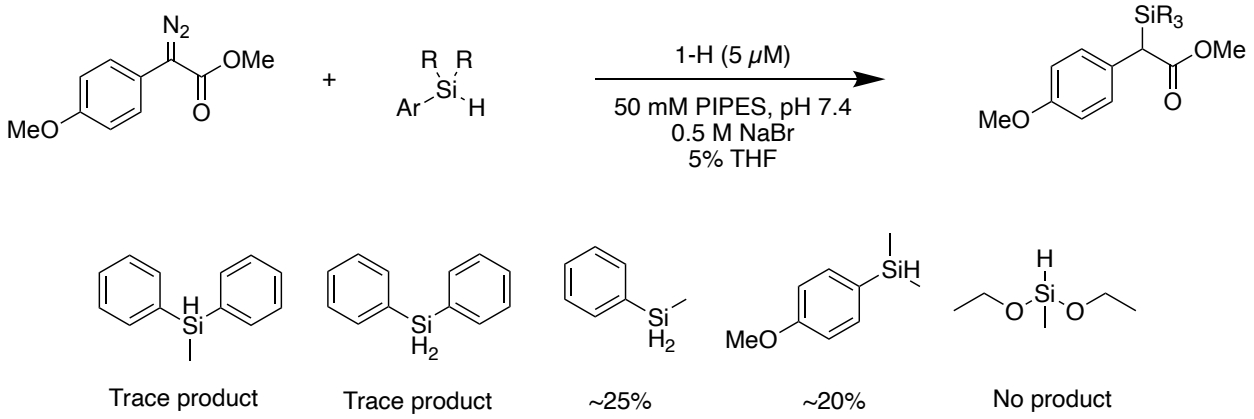


			3		**4**		
Entry	Variant	Mutations	Yield (%)	e.e. (%)	Yield (%)	**3**/**4**	TTN to **3**
1	3-H	Parent	24	93	71	0.34	241
2	2-SH	E241S/L328H	30	90	68	0.44	298
3	3-D	S241D	32	92	64	0.50	324
4	4-LIY	F194L/T211I/D241Y	38	93	59	0.64	383

5.2.1.4 Substrate Scope

As the 3-H scaffold catalyzes the Si-H functionalization of dimethylphenylsilane with excellent enantioselectivity but poor yield, we wondered if other substrates may be higher yielding. A collection of silanes with different steric and electronic properties were examined using mass spectrometry to determine if product was formed (Scheme 5.1). Diphenylsilanes were not tolerated, presumably due to the additional steric bulk of a second phenyl ring and decreased

solubility in the aqueous buffer. As expected, the less bulky phenylmethylsilane was successfully converted to the desired product with an approximate yield of 25%, as was the electron-rich (4-methoxy)phenyldimethylsilane. Approximate yields of these reactions were determined by the relative amount of O-H insertion side product after consumption of the diazo starting material. The final silane examined, diethoxymethylsilane, produced no desired product. While none of these substrates achieved higher yields as we had hoped, these data will greatly assist in the exploration of the substrate scope in the future.



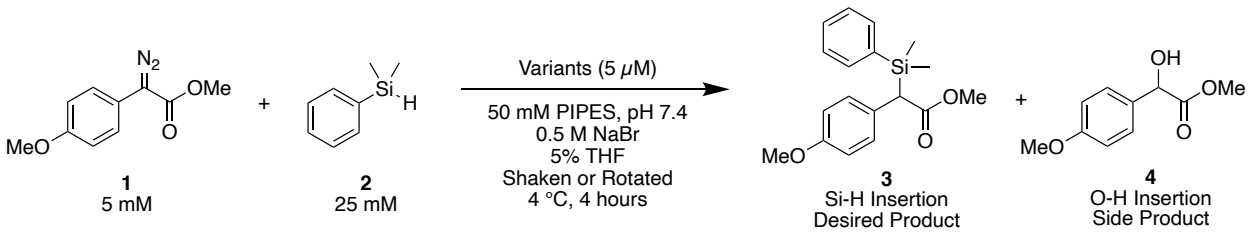
Scheme 5.1 Silane substrate scope.

5.2.1.5 Alternate Approaches to Improving Si-H Functionalization Yield

In addition to directed evolution, we explored a number of alternate approaches to improve the yield and **3**/**4** ratio of the ArM-catalyzed Si-H functionalization reaction. The optimized reaction conditions contain 5% THF that remains as an organic layer on the top of the reaction mixture, even with rapid shaking. We were concerned that, due to its high hydrophobicity, the substrate was not diffusing into the aqueous buffer at a concentration that was sufficient to saturate the scaffold active site before carbene formation. Though neither higher THF percentages nor

increased silane concentration resulted in improved yields during the initial reaction optimization (Table 5.5), we explored other options to overcome this potential issue. As the reactions are set up and shaken in microcentrifuge tubes, we wondered if rotating the reactions upside down throughout the reaction would better mix the biphasic mixture. As shown in Table 5.7, this was not the case. The yield for both 2-SH and 3-D was essentially the same when the reactions were rotated instead of shaken. The rotated reactions also gave slightly higher enantioselectivity and lower **3**/**4** ratios, but none of these differences were significant enough to pursue further.

Table 5.7 Rotation compared to shaking in Si-H functionalization.

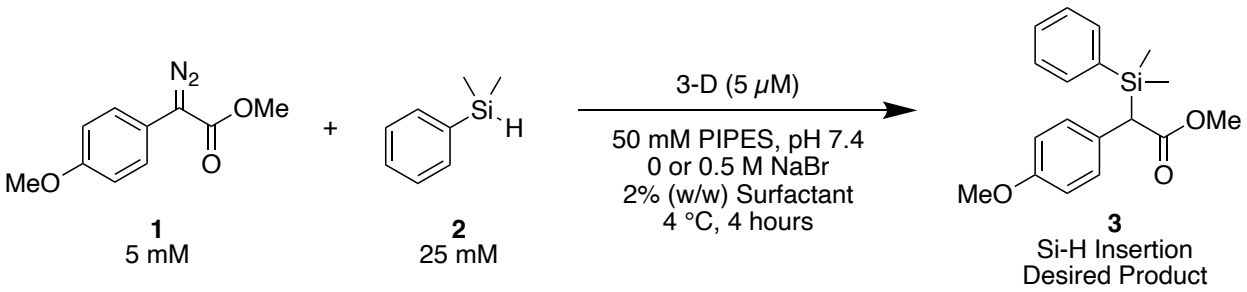


			3		
Entry	Variant	Mixing	Yield (%)	e.e. (%)	**3**/**4**
1	2-SH	Shake	28	91	0.41
2	2-SH	Rotate	27	94	0.35
3	3-D	Shake	31	92	0.46
4	3-D	Rotate	27	95	0.37

There have been numerous reports of surfactants improving biocatalysis on hydrophobic substrates.[34,35] These surfactants form micelles throughout the reaction mixture that hydrophobic substrates can more easily diffuse into and out of, so we thought that this strategy may improve our solubility issue. ArM reactions were prepared without organic cosolvent to facilitate micelle formation, but none of the three surfactants examined had a positive effect on ArM catalysis (Table

5.8). In each case, both the yield and enantioselectivity was considerably lower than under standard reaction conditions.

Table 5.8 Use of surfactants in ArM-catalyzed Si-H functionalization.

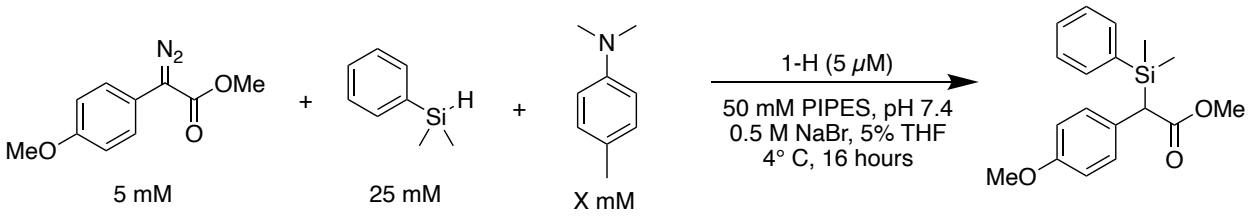


		[NaBr]	**3**	
Entry	Surfactant	(M)	Yield (%)	e.e. (%)
1	Tween 20	0	12	51
2	Tween 20	0.5	7	69
3	PS-750-M	0	14	58
4	PS-750-M	0.5	11	71
5	TPGS-750-M	0	9	62
6	TPGS-750-M	0.5	7	73

As changing substrate solubility did not improve the yield as we had hoped, we considered ways that we might alter the active site environment. Inspired by an observation in our N-H functionalization work, that binding of an amine to the axial position of dirhodium inhibits reaction (Section 4.2.1.3), we wondered if this might have a favorable effect on Si-H functionalization. To test this, we added *p*-methyldimethylaniline (pMDMA, Table 8.9) to the Si-H functionalization reaction mixture. Instead of the typical consumption of diazo in less than an hour, the orange color indicative of remaining diazo was observed even after 8 hours in reactions with 25 mM pMDMA (Table 5.9, Entry 4), so all reactions were allowed to continue for 16 total hours, at which point the diazo was completely consumed. Though inhibition had occurred as expected, the presence of

DMA resulted in lower yield of **3** and **3**/**4** ratio, as shown in Table 5.9. Interestingly, the enantioselectivity substantially improved at higher DMA concentrations to >98% with 25 mM DMA. The origin of this effect was not probed in depth, but one hypothesis is that the binding and release of aniline fills the already restricted active site space, essentially blocking off approach from the disfavored face. A direct analogue of the model substrate in N-H functionalization, *N*-methyltetrahydroquinoline, along with other molecules containing functionality capable of ligating the dirhodium cofactor, including trialkylamines, pyridines and polar triarylphosphines, were also examined. While a lower rate of reaction was observed in nearly every case, none of these inhibitors resulted in a higher yield of the desired Si-H functionalized product.

Table 5.9 Effect of aniline inhibitors on Si-H functionalization.



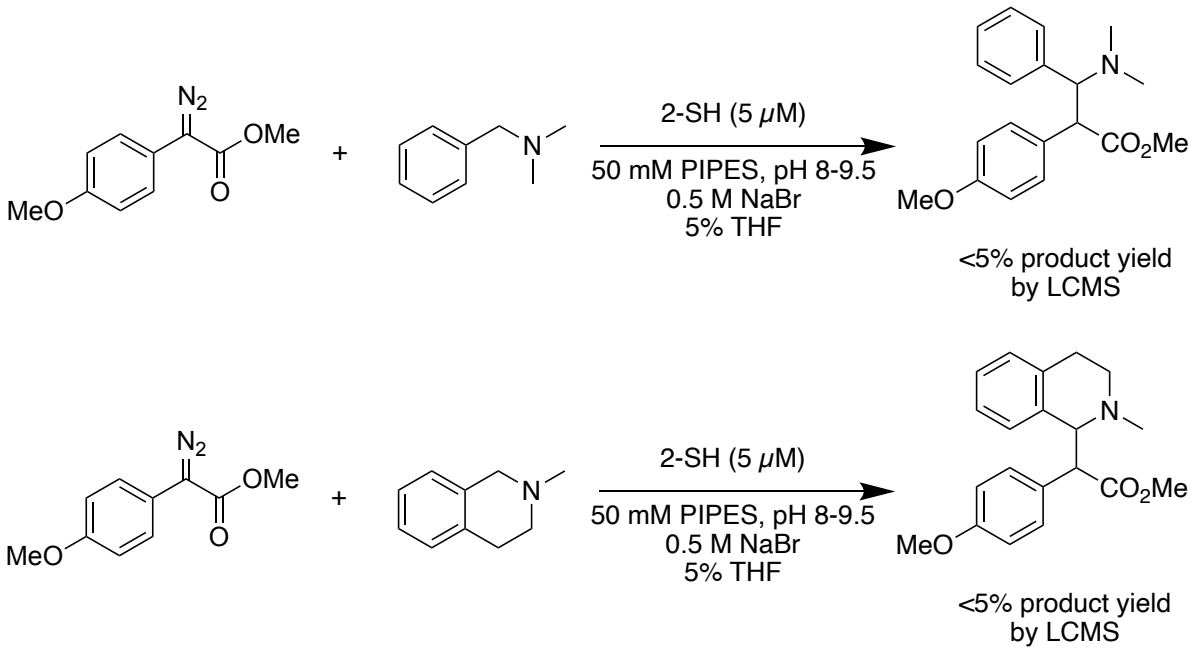
	[DMA]	**3**		
Entry	(mM)	Yield (%)	e.e. (%)	**3**/**4**
1	0	24	93	0.34
2	1	23	93	0.33
3	5	21	96	0.31
4	25	12	>98	0.21

As was discussed at length Section 1.2, a reliable way to improve ArM catalysis is through scaffold-based control over the primary and secondary coordination spheres. This section has described a number of alternate routes towards the goal of increased Si-H functionalization yield

but the only method that has succeeded thus far is the use of directed evolution to improve the scaffold's control over chemoselectivity.

5.2.2 Alternate Approaches Towards C-H Functionalization

During the process of evolving improved scaffolds for Si-H functionalization, we considered other options to improve yields and selectivities for C-H functionalization. Inspired by the pMDMA results noted above, we postulated that substrates containing a tertiary amine and an activated C-H bond, such as *N*,*N*'-dimethylbenzylamine or *N*-methyltetrahydroisoquinoline (Scheme 5.2), could bind to the cofactor and be well-positioned for C-H functionalization after dissociation and carbenoid formation. Unfortunately, this approach did not lead to higher C-H functionalization yields. At the 4 pH values examined (pH 8, 8.5, 9, 9.5), yields lower than 5% were observed for both substrates, determined relative to the yield of the O-H insertion side product. While this could be an interesting route to pursue in the future, it may take considerable substrate design and protein engineering to achieve high levels of C-H functionalization desired.



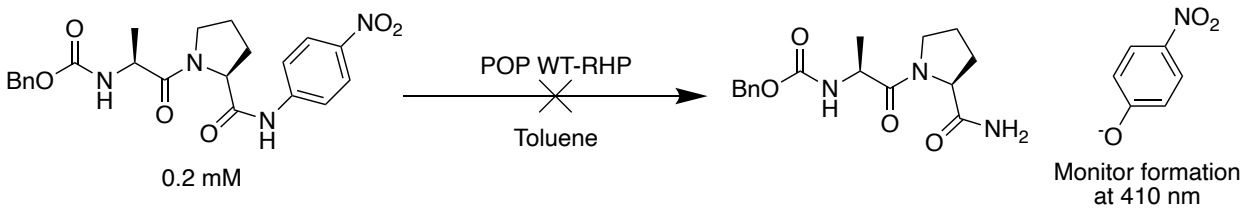
Scheme 5.2 C-H functionalization of rhodium-binding substrates.

For both Si-H and C-H functionalization reactions catalyzed by our POP-ArMs, the major product is the undesired O-H insertion of water. This suggests that the substrate positioning necessary to achieve carbene insertion into non-polar bonds is apparently slower or less favorable than the reaction with water in the active site. We therefore speculated that anhydrous reaction conditions would eliminate the possibility of this more rapid side reaction and enable higher yields of the desired reaction. As soluble enzymes have evolved to be stable in aqueous conditions, running reactions in other solvents typically results in denaturation and significant, if not complete, loss of activity. There are some examples of lipases that can be used in organic solvents,[36] but these enzymes evolved to function on fatty acids where tolerance to hydrophobicity is necessary. To enable general biocatalysis in non-aqueous solvents, a variety of approaches have been developed.

Recent reports have used random heteropolymers (RHPs) to stabilize protein structure and prevent denaturation in organic solvents,[37,38] which appeared to be a promising path towards anhydrous ArM chemistry. In these investigations, the authors showed that polymers containing a mixture of randomly distributed hydrophobic and hydrophilic monomers could encapsulate stabilize proteins when the mixture was dissolved and stored in organic solvents for less than 1 day. Though no catalysis was performed in organic solvent, the enzymes maintained activity after evaporation and redissolution in aqueous buffer. In a collaboration with the Xu group at UC Berkeley, which developed the RHPs noted above, the activity of robust, model POP-RHP systems were analyzed in organic solvents before more challenging reactions were attempted.

The POP WT-RHP hydrolysis of Z-Gly-Pro, a substrate analogue, was examined using 3 RHPs provided by the Xu group. Each of these RHPs has a different ratio of hydrophilic to hydrophobic monomers, so we thought that there may be one more suited to maintain the structure and activity of POP WT in toluene. The POP-WT-RHP mixture was prepared according to the optimized method, where a 5:1 RHP:enzyme (w/w) mixture was prepared in water, then lyophilized. The resulting solids were resuspended in toluene, where the activity was measured. As shown in Table 5.10, no product formation (monitored by Abs 410 nm) was observed using high (2.8 μM) or low (0.54 μM) POP WT concentration. The enzyme is known to rapid hydrolyze the substrate at <100 nM concentrations in aqueous buffer.[39] Indeed, when the toluene was evaporated from these reactions and the residue was redissolved in aqueous buffer, product formation occurred quite rapidly, indicating that the enzyme structure was maintained in organic solvent due to RHP stabilization, but no reaction occurred.

Table 5.10 POP WT-RHP activity in toluene.

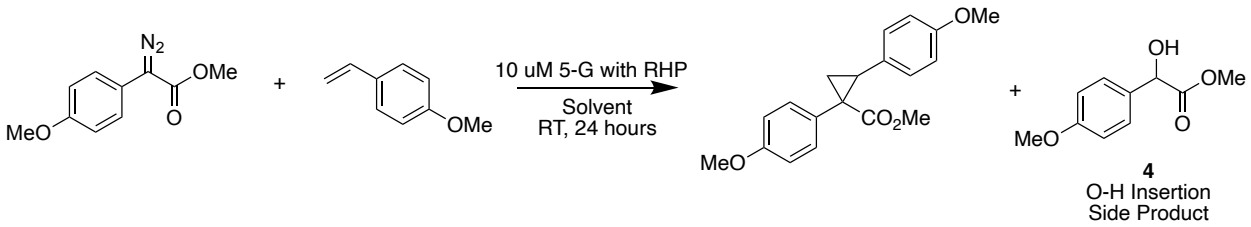


Entry	RHP	[POP WT] (μM)	Activity
1	K	2.78	No product formation
2	ZR16B	2.78	No product formation
3	ZR16E	2.78	No product formation
4	K	0.54	No product formation
5	ZR16B	0.54	No product formation
6	ZR16E	0.54	No product formation

Though these results were negative, potential causes, such as poor substrate solubility and a lack of domain dynamics when encapsulated in RHPs, would not necessarily by prohibitive in the context of ArM reactions. We therefore used cyclopropanation, catalyzed with high yields and selectivities in aqueous buffer, as a model reaction to understand how the ArMs may function in organic solvent. After demonstrating that POP-ArMs can tolerate lyophilization, a potentially useful fact for future efforts in water-free systems, 5-G-RHP aggregates were prepared according to the published methods. The RHP/ArM ratio was varied from 10:1 to 0.6:1, with no conversion observed using >3:1. As shown in Table 5.11, poor activity was observed for the desired cyclopropanation reaction in all cases. Even in cases with a detectable yield, with product was formed with no enantioselectivity, suggesting that the ArMs are not functioning as they do in aqueous buffer. Interestingly, even after lyophilization, there is a considerable amount of O-H insertion side product observed, indicating that persistent water remains embedded in the protein-RHP mixture. These results imply that C-H functionalization would likely not be controlled by the

scaffold under these conditions either, so an alternate approach towards anhydrous reaction conditions was examined.

Table 5.11 Cyclopropanation catalyzed ArM-RHP mixtures in organic solvents.

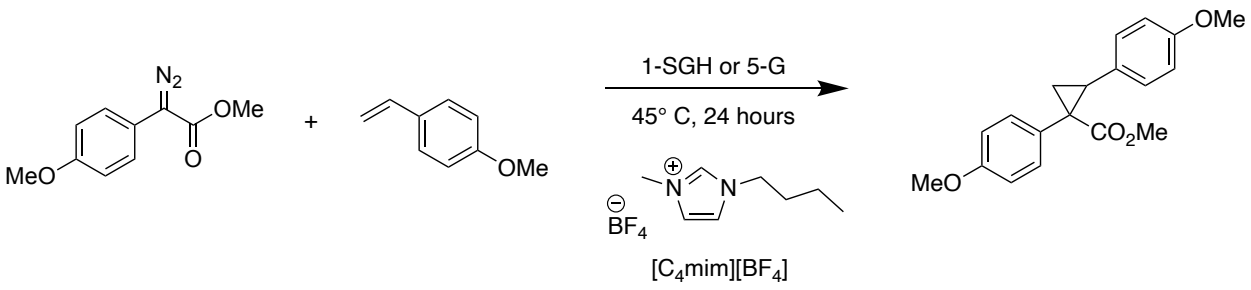


Entry	Cosolvent	RHP/ArM (w/w)	Cyclopropane	**4**
1	Toluene	2.8	No product formation	~5%
2	Toluene	0.6	No product formation	~5%
3	THF	2.8	No product formation	~5%
4	THF	0.6	No product formation	~5%
5	DCM	2.8	<5%	~20%
6	DCM	0.6	<5%	~20%
7	EA	2.8	No product formation	~5%
8	EA	0.6	No product formation	~5%

Ionic liquids (ILs) have found use as a solvent for biocatalytic reactions, with numerous enzymes maintaining activity without denaturation.[40] The ionic nature of these solvents has been suggested to stabilize the hydrophilic exterior of soluble proteins, so we thought that these liquids could enable scaffold-controlled carbene transfer catalysis in anhydrous conditions. A variety of ILs have been used for this purpose, with methylimidazoliums (mim) frequently the best at maintaining enzyme activity compared to aqueous conditions.[40] Three mim-based ILs containing different counter ions ([C_4mim][PF_6], [C_4mim][BF_4], [C_6mim][Cl]) were therefore screened for POP WT hydrolase activity at 2 μM enzyme loading. No activity was observed in any of these cases, but as was the case with RHPs, the ArM may still be functional. A single IL was selected for ArM-catalyzed cyclopropanation, [C_4mim][BF_4], as it was the least viscous and best

solubilized the diazo substrate. Many biocatalytic reactions involving ILs are performed in IL/buffer mixtures, so both 100% IL and 1:1 IL:buffer reaction conditions were analyzed using 50 mM PIPES without salt as the buffer. Analysis of the reactions was performed using an ethyl acetate extraction, but significant IL in the organic extracts resulted in poor chromatographic results. The results, as shown in Table 5.12, are clear regardless. Under anhydrous conditions (Entries 1-4), there is essentially no conversion. With 50% buffer, however, some product was observed with modest enantioselectivity. As the purpose of using ILs was to achieve relevant ArM catalysis under anhydrous conditions, the lack of activity with 100% IL meant this method was not investigated further.

Table 5.12 ArM-catalyzed cyclopropanation in an ionic liquid.



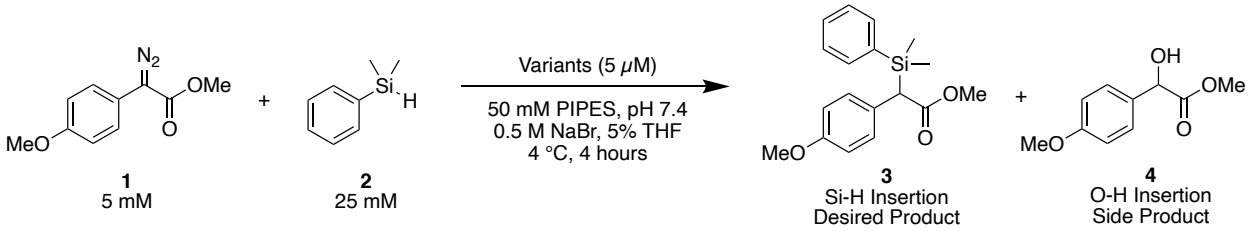
Entry	Variant	[ArM] (μM)	% IL	Result	Approx. e.e. (%)
1	1-SGH	2	100	No product formation	N/A
2	5-G	2	100	No product formation	N/A
3	1-SGH	10	100	No product formation	N/A
4	5-G	10	100	No product formation	N/A
5	1-SGH	2	50	Low product formation	14
6	5-G	2	50	Low product formation	56
7	1-SGH	10	50	Low product formation	17
8	5-G	10	50	Low product formation	31

5.3 Conclusion

Though performing ArM catalysis in anhydrous conditions was not found to preserve ArM reactivity using the methods examined, our initial approach towards improved concerted carbene transfer has proven fruitful. After thorough optimization of the reaction conditions and identification of the variant 3-H, ArM-catalyzed Si-H functionalization was achieved with excellent enantioselectivity (93%) but poor chemoselectivity, with a **3**/**4** ratio of only 0.34. Directed evolution of the scaffold was used to improve this characteristic, with the current variant, 4-FIY, catalyzing Si-H functionalization with the same enantioselectivity but a **3**/**4** ratio of 0.64, nearly double that of the parent. Continued evolution of the scaffold-controlled chemoselectivity is underway, with the variants developed for this reaction to be eventually used for C-H functionalization. Unfortunately, other attempts at improving the chemoselectivity of the desired transformation have not proven effective. While still in an early stage, mono-aryl silanes were acceptable substrates for ArM-catalyzed Si-H functionalization, which achieved similar yields to the model substrate. Other efforts using tertiary aniline inhibitors resulted in increased selectivity but not the improved yield we were pursuing. Scaffold-based control over the enantio- and chemoselectivity of carbene insertion into non-polar bonds has proven challenging, but the results laid out in this chapter show that improvement of the scaffold using directed evolution has, and will continue to, tune the primary and secondary spheres for improved ArM-catalysis.

5.4 Experimental

General ArM-catalyzed Si-H functionalization Method



A 285 µL solution of 5 µM ArM (corrected by the known bioconjugation efficiency, typically 90%) in 50 mM PIPES, pH 7.4 with 0.5 M NaBr was prepared in a 1.5 mL microcentrifuge tube. While shaking at 750 rpm at 4 °C, 15 µL of the diazo stock solution (0.1 M donor-acceptor diazo, 0.5 M dimethylphenylsilane) in THF was added to obtain a final concentration of 5 mM donor-acceptor diazo and 25 mM dimethylphenylsilane. The reactions were sealed and briefly inverted to assist with mixing of the substrate, then placed in a thermomixer and shaken at 750 rpm at 4 °C for 4 hours. The reaction was quenched by the addition of 50 µL internal standard (30 mM 1,3,5-trimethoxybenzene in ethyl acetate) and 100 µL ethyl acetate. The biphasic mixture was vortexed thoroughly then centrifuged at 20,000 xg for 1 minute to separate the phases. 100 µL of the upper organic layer was collected and filtered for SFC analysis.

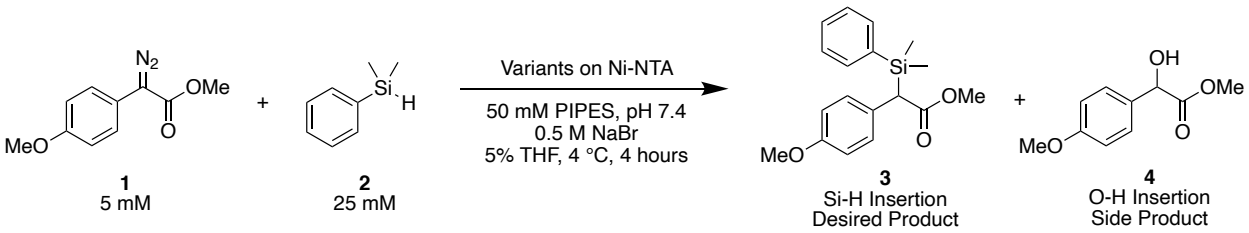
ArM library screening

Libraries were constructed as described previously (Section 3.4.6). When CCM was used for diversification in the first round of directed evolution (3-H CCM), an average of 2.0 residue mutations were made per gene. The variants were incubated, overexpressed and lyzed as described previously in Chapter 3, using 5 mL overexpression cultures in 24-well plates. The lysate was pelleted by centrifugation at 3,600 rpm and 480 µL of the clarified lysate from each well was transferred to a 96-deep well filter plate. 120 µL of cofactor **1** stock solution in ACN (12.8 µM,

0.010 mg/mL) was added to each well while shaking at 650 rpm at 4 °C. The mixture was incubated at 4 °C, 750 rpm for one hour, at which point 100 μL of 50% Ni-NTA resin was added to each well. ~1 cm of the multichannel pipette tips were cut off using scissors to allow easier transfer of the resin. The plate was shaken at 4° C for 5 minutes to allow protein binding, then centrifuged at 2500 rpm for 5-10 minutes until the resin in each well was dry. The resin was washed to remove excess cofactor by the addition of 500 μL of 50% acetonitrile/50 mM Tris buffer to each well. The plate was filtered by centrifugation at 3000 rpm for 3 minutes to remove the buffer. Three washes were done in total (1.5 mL total wash buffer) with a final centrifugation time of 5 minutes to ensure each well was dry.

285 μL reaction buffer (50 mM PIPES, pH 7.4 with 0.5 M NaBr) was added to each well and the plate was shaken at 4 °C, 750 rpm for 10 minutes. After this cooling period, 15 μL of substrate stock solution was added (final concentration of 5 mM methyl (4-methoxyphenyl)diazoacetate and 25 mM dimethylphenylsilane) at 4°C with 750 rpm shaking. The screening plate was sealed with a 96-well rubber mat to prevent evaporation and shaken at these conditions for 16 hours. The reaction was quenched by addition of 50 μL of 30 mM 1,3,5-trimethoxybenzene in ethyl acetate as internal standard. The plate filtered by centrifugation at 3,000 rpm for 5 min. To remove all organics from the resin, two additional filtrations of 75 μL ethyl acetate were applied and the biphasic mixture was clarified via centrifugation at 3,000 rpm for 5 min. 100 μL of the upper organic portion from each well were transferred to a 96-well microtiter plate for SFC analysis. Mutants displaying a higher product/internal ratio or improved E/Z isomer ratio than the parent were selected for DNA sequencing and validation.

Example Library Data



After extraction, the reactions were analyzed using SFC chromatography. A ChiralPak OJ-3 column was used with a flow rate of 3 mL/min of A: CO_2, B: MeOH. The method was as follows:

0 mins: 1%B

0.5 mins: 1%B

4 mins: 20%B

Example Chromatogram of Library Screening (showing hit 4-LIY (well B1)):

Product/IS ratio:

	1	2	3	4	5	6	7	8	9	10	11	12
A	N/A	N/A	N/A	0.47	0.20	0.16	0.46	0.03	0.51	0.22	0.04	0.47
B	0.50	0.02	0.45	0.46	0.19	0.43	0.35	0.05	0.31	0.09	0.48	0.31
C	0.40	0.38	0.02	0.40	0.04	0.19	0.48	0.30	0.47	0.41	0.05	0.03
D	0.15	0.42	0.05	0.41	0.03	0.04	0.34	0.34	0.05	0.07	0.38	0.15
E	0.06	0.05	0.39	0.03	0.06	0.15	N/A	0.37	N/A	0.41	0.37	0.40
F	0.23	0.21	0.45	0.45	0.09	0.18	0.13	0.58	0.37	0.30	0.27	0.45
G	0.12	0.41	0.36	0.13	0.05	0.05	0.47	0.33	0.27	0.47	0.33	0.48
H	0.13	0.08	0.18	0.42	N/A	0.00	0.44	N/A	0.20	0.06	0.50	0.49

Enantioselectivity:

	1	2	3	4	5	6	7	8	9	10	11	12
A	N/A	N/A	70.5	70.3	1.6	24.3	88.8	N/A	87.0	37.2	N/A	89.0
B	80.8	N/A	90.2	90.6	40.7	85.4	76.0	N/A	71.5	N/A	90.3	64.6
C	85.4	81.5	N/A	76.8	N/A	43.6	88.9	80.1	87.0	88.8	N/A	N/A
D	16.3	71.5	15.9	81.6	N/A	N/A	64.9	82.0	-12.3	-4.9	87.0	19.7
E	N/A	N/A	90.2	N/A	N/A	-15.7	N/A	66.0	N/A	90.3	85.0	85.7
F	14.2	21.6	92.3	89.0	N/A	32.0	2.7	90.1	85.2	80.9	40.9	88.0
G	-1.8	87.1	88.3	7.9	N/A	N/A	85.2	64.8	68.5	90.3	62.3	88.4
H	N/A	9.8	11.6	92.6	N/A	0.0	79.6	N/A	20.1	20.1	86.5	85.5

Blue wells are parent, green well (B1) was found to be the hit 4-LIY.

General Reaction Method for C-H Functionalization Reactions

Reactions were prepared according to the general Si-H functionalization reaction procedure using the desired C-H functionalization substrate at 25 mM unless otherwise specified.

Protocol for the lyophilization of dirhodium ArMs

After bioconjugation and resin purification, the ArM was buffer exchanged into 50 mM Tris•HCl (no salt). The resulting ArM stocks were diluted to 200 µM in the same buffer. 7.5 µL of this ArM stock was added to a microcentrifuge, followed by 42.5 µL of pure water. This mixture was snap frozen in liquid nitrogen and lyophilized for >16 hours. After lyophilization, the solids were redissolved in the desired buffer and used without additional treatment.

Protocols for Random Heteropolymer Reactions

Preparation of Protein-RHP materials

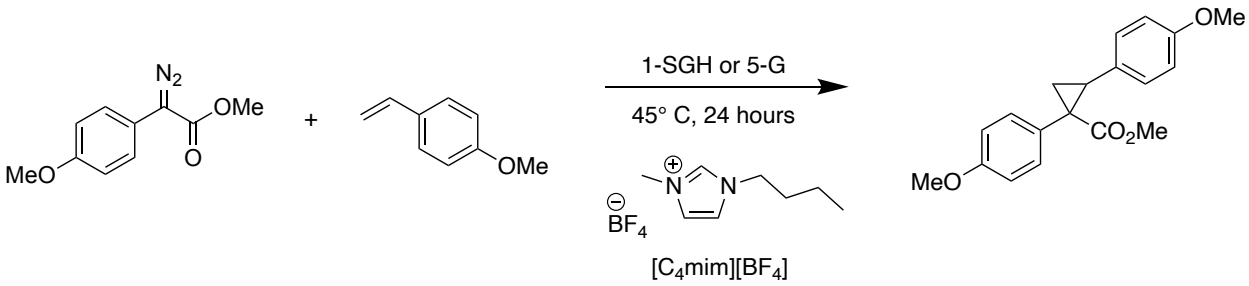
A modification of the protocol from ref. 37 was used. To prepare a 5:1 (m:m) mixture of RHP to protein (2.8 µM final protein concentration), the following method was used: POP WT in

pure water (1025 μM) was diluted to 139 μM (10 mg/mL). 2 μL of this mixture was diluted in 48 μL pure water (0.4 mg/mL) in a microcentrifuge tube, to which 50 μL of 2 mg/mL resin (in pure water) was added. The mixture was vortexed and lightly spun down, then incubated at RT for 10-30 mins. The mixture was snap frozen in liquid nitrogen and lyophilized for >16 hours. Small amounts of colorless solids were observed. Both POP-WT and RHP-ArM mixtures were prepared the same way, with the ArM buffer exchanged into water after bioconjugation.

Reactions run using Protein-RHP

The lyophilized Protein-RHP solids were redissolved in the desired solvent and sonicated for 10 minutes to ensure complete the dissolution of the material. The diazo and olefin substrates were prepared in stock solutions in the necessary solvent and added directly to the dissolved protein-RHP mixtures. After the reactions were complete, 50 μL of internal standard in ethyl acetate was added to the reactions, which were filtered for SFC analysis.

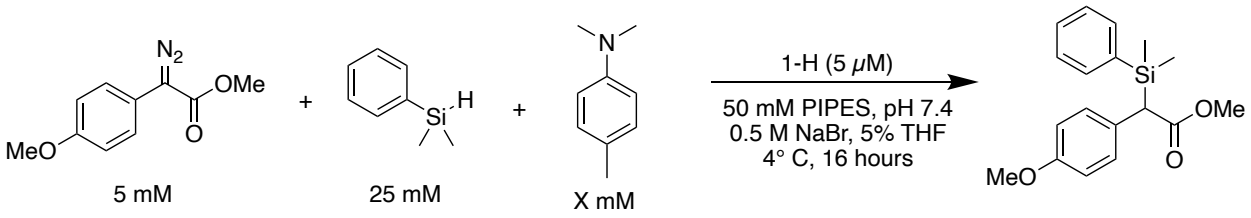
Protocols for Ionic Liquids Reactions



The desired concentration of ArM for a 300 μL reaction was prepared in pure water in a microcentrifuge tube. This was lyophilized according to the general procedure, then brought into a dry glovebox. As the ionic liquids were highly viscous, approximately 1 cm was cut off the end of 1 mL pipette tips to enable transfer. A stock solution containing the diazo and silane were

prepared in the appropriate ionic liquid in the glovebox. To each microcentrifuge tube containing lyophilized ArM, 285 µL of ionic liquid (stored under N_2 in a glovebox) was added and all mixtures were brought out of the box. To reactions containing 50% PIPES buffer, 142.5 µL IL was added and, after the tubes were brough out of the box, an additional 142.5 µL of aqueous buffer was added. 15 µL of the substrate stock solution was added to each microcentrifuge tube and the reactions were shaken vigorously by hand to ensure the solution was mixed. The reactions were shaken at 750 rpm at room temperature or 45° C for 24 hours. Analysis was performed by removing a 50 µL aliquout and diluting with 50 µL internal standard and 200 µL ethyl acetate. The resulting solutions were filtered before SFC analysis.

General Reaction Method for Aniline Si-H Functionalization Reactions



The reactions were run using the general method with the addition of X mM N,N'-dimethylanilide and 16 hour reaction time rather than the typical 4 hours.

5.5 Acknowledgements

I would like to thank Rui Huang for his efforts in cloning and library creation throughout this project as well as for the fruitful discussions that led to many of the ideas tested in this chapter. Additionally, Prabir Saha provided assistance in synthesis and library screening.

5.6 References

[1] D. Y. K. Chen, S. W. Youn, *Chem - European J* 2012, *18*, 9452–9474.

[2] M. C. White, *Science* 2012, *335*, 807–809.

[3] R. G. Bergman, *Nature* 2007, *446*, 391–393.

[4] T. Cernak, K. D. Dykstra, S. Tyagarajan, P. Vachal, S. W. Krska, *Chem Soc Rev* 2016, *45*, 546–576.

[5] K. M. Engle, T. S. Mei, M. Wasa, J. Q. Yu, *Accounts Chem Res* 2012, *45*, 788–802.

[6] S. R. Neufeldt, M. S. Sanford, *Accounts Chem Res* 2012, *45*, 936–946.

[7] J. F. Hartwig, M. A. Larsen, *Acs Central Sci* 2016, *2*, 281–292.

[8] J. F. Hartwig, *Accounts Chem Res* 2012, *45*, 864–873.

[9] F. O'Hara, D. G. Blackmond, P. S. Baran, *J Am Chem Soc* 2013, *135*, 12122–12134.

[10] P. E. Gormisky, M. C. White, *J Am Chem Soc* 2013, *135*, 14052–14055.

[11] H. M. L. Davies, D. Morton, *Chem Soc Rev* 2011, *40*, 1857.

[12] J. Hansen, J. Autschbach, H. M. L. Davies, *J Org Chem* 2009, *74*, 6555–6563.

[13] E. Nakamura, N. Yoshikai, M. Yamanaka, *J Am Chem Soc* 2002, *124*, 7181–7192.

[14] K. Liao, S. Negretti, D. G. Musaev, J. Bacsa, H. M. L. Davies, *Nature* 2016, *533*, 230–234.

[15] J. He, L. G. Hamann, H. M. L. Davies, R. E. J. Beckwith, *Nat Commun* 2015, *6*, 5943.

[16] W. Liu, Z. Ren, A. T. Bosse, K. Liao, E. L. Goldstein, J. Bacsa, D. G. Musaev, B. M. Stoltz, H. M. L. Davies, *J Am Chem Soc* 2018, *140*, jacs.8b07534.

[17] H. M. L. Davies, D. Morton, *J Org Chem* 2016, *81*, 343–350.

[18] P. S. Coelho, J. C. Lewis, F. H. Arnold, in *Comprehensive Organic Synthesis* (Eds.: ["Knochel, Paul and Molander, Gary A"]), Elsevier Ltd, n.d., pp. 390–420.

[19] J. C. Lewis, P. S. Coelho, F. H. Arnold, *Chem Soc Rev* 2010, *40*, 2003–2021.

[20] R. Kaspera, R. Croteau, *Phytochem Rev* 2006, *5*, 433–444.

[21] A. M. Walji, D. W. C. MacMillan, *Synlett* 2007, *12*, 1477–1489.

[22] Y. Ishihara, P. Baran, *Synlett* 2010, *2010*, 1733–1745.

[23] D. M. Upp, J. C. Lewis, *Curr Opin Chem Biol* 2017, *37*, 48–55.

[24] P. A. Romero, F. H. Arnold, *Nat Rev Mol Cell Bio* 2009, *10*, 866–76.

[25] E. L. Bell, W. Finnigan, S. P. France, A. P. Green, M. A. Hayes, L. J. Hepworth, S. L. Lovelock, H. Niikura, S. Osuna, E. Romero, K. S. Ryan, N. J. Turner, S. L. Flitsch, *Nat Rev Methods Primers* 2021, *1*, 46.

[26] T. K. Hyster, T. R. Ward, *Angewandte Chemie Int Ed* 2016, *55*, 7344–7357.

[27] F. Schwizer, Y. Okamoto, T. Heinisch, Y. Gu, M. M. Pellizzoni, V. Lebrun, R. Reuter, V. Köhler, J. C. Lewis, T. R. Ward, *Chem Rev* 2018, *118*, 142–231.

[28] J. C. Lewis, *Accounts Chem Res* 2019, *52*, 576–584.

[29] H. Yang, A. M. Swartz, H. J. Park, P. Srivastava, K. Ellis-Guardiola, D. M. Upp, G. Lee, K. Belsare, Y. Gu, C. Zhang, R. E. Moellering, J. C. Lewis, *Nat Chem* 2018, *10*, 318–324.

[30] D. M. Upp, R. Huang, Y. Li, M. J. Bultman, B. Roux, J. C. Lewis, *Angewandte Chemie Int Ed* 2021, *60*, 23672–23677.

[31] R. T. Buck, M. P. Doyle, M. J. Drysdale, L. Ferris, D. C. Forbes, D. Haigh, C. J. Moody, N. D. Pearson, Q.-L. Zhou, *Tetrahedron Lett* 1996, *37*, 7631–7634.

[32] P. Srivastava, H. Yang, K. Ellis-Guardiola, J. C. Lewis, *Nat Commun* 2015, *6*, 7789.

[33] H. Yang, A. M. Swartz, H. J. Park, P. Srivastava, K. Ellis-Guardiola, D. M. Upp, G. Lee, K. Belsare, Y. Gu, C. Zhang, R. E. Moellering, J. C. Lewis, *Nat Chem* 2018, *10*, 318–324.

[34] M. Cortes-Clerget, N. Akporji, J. Zhou, F. Gao, P. Guo, M. Parmentier, F. Gallou, J.-Y. Berthon, B. H. Lipshutz, *Nat Commun* 2019, *10*, 2169.

[35] D. K. Romney, F. H. Arnold, B. H. Lipshutz, C.-J. Li, *J Org Chem* 2018, *83*, 7319–7322.

[36] A. Kumar, K. Dhar, S. S. Kanwar, P. K. Arora, *Biol Proced Online* 2016, *18*, 2.

[37] B. Panganiban, B. Qiao, T. Jiang, C. DelRe, M. M. Obadia, T. D. Nguyen, A. A. A. Smith, A. Hall, I. Sit, M. G. Crosby, P. B. Dennis, E. Drockenmuller, M. O. de la Cruz, T. Xu, *Science* 2018, *359*, 1239–1243.

[38] C. DelRe, Y. Jiang, P. Kang, J. Kwon, A. Hall, I. Jayapurna, Z. Ruan, L. Ma, K. Zolkin, T. Li, C. D. Scown, R. O. Ritchie, T. P. Russell, T. Xu, *Nature* 2021, *592*, 558–563.

[39] M. N. Harris, J. D. Madura, L.-J. Ming, V. J. Harwood, *J Biol Chem* 2001, *276*, 19310–19317.

[40] T. Itoh, *Chem Rev* 2017, *117*, 10567–10607.

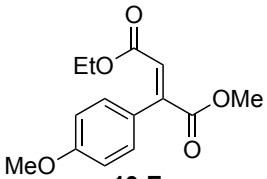
Appendix – Supplemental Spectra

Representative LC Traces

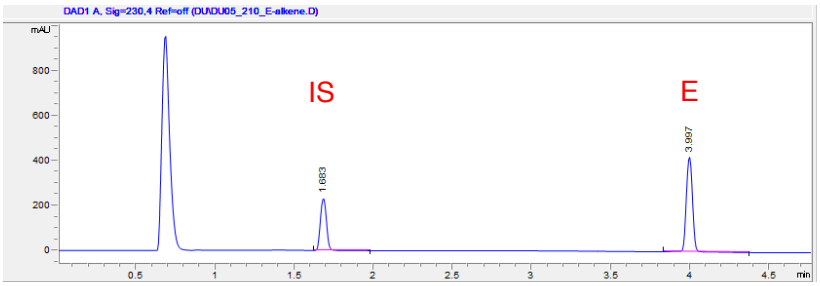
SFC traces

ChiralTek OJ-3 SFC column was used. Mobile phase: A: CO_2, B: MeOH. Method: 1% B for 0.5 mins, 1% B to 25% B over 3.5 minutes, 1 minute at 25%B.

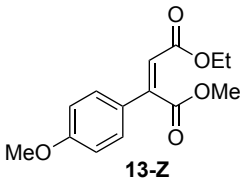
Pure E-alkene:



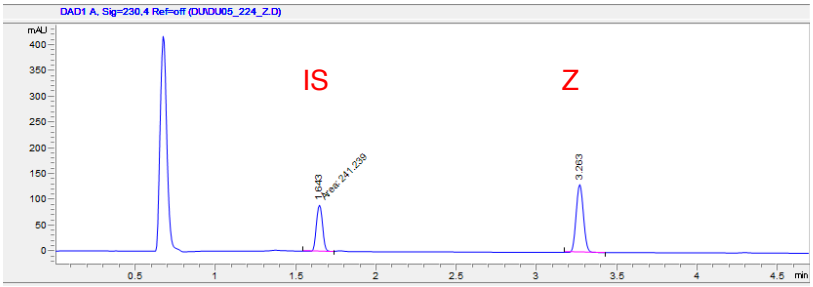
13-E
4.0 mins



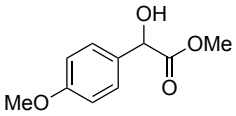
Pure Z-alkene:

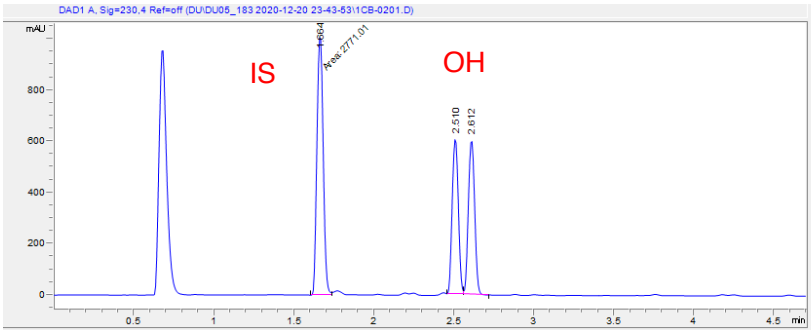


13-Z
3.2 mins

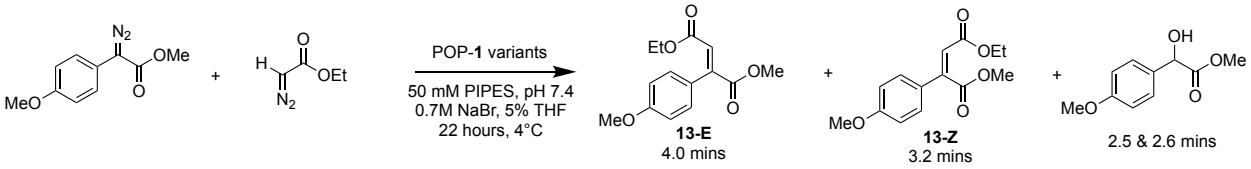


OH insertion side product:

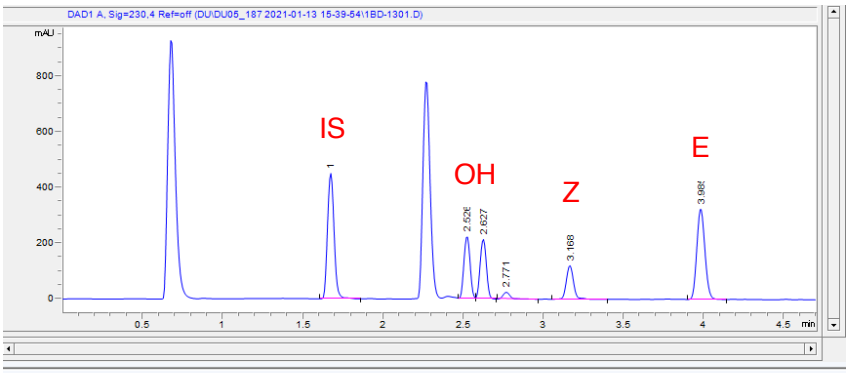




Reaction chromatograms using model substrates:

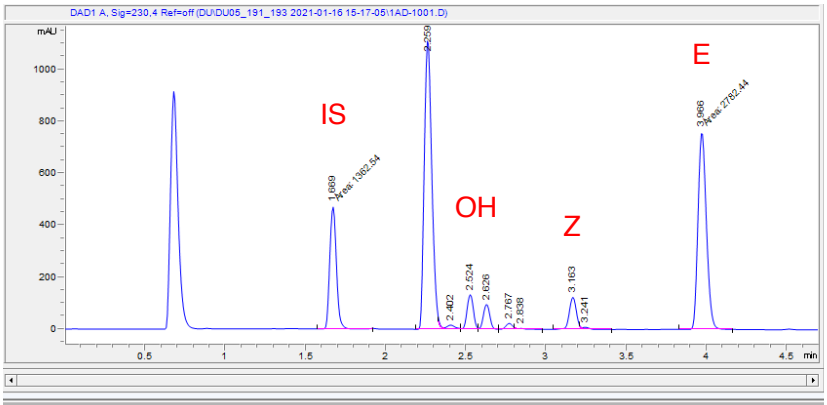


Diazo coupling catalyzed by 1-SGH (5 μM):



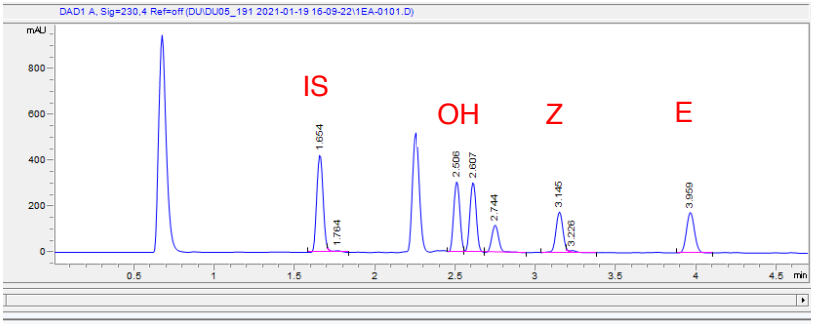
#	Time	Area	Height	Width	Area%	Symmetry
1	1.675	1284.6	449.9	0.0453	31.611	0.887
2	2.526	604.8	224.5	0.0434	14.882	0.975
3	2.627	583.3	214.3	0.0438	14.353	0.972
4	2.771	75.2	24.3	0.0481	1.851	0.8
5	3.168	376.4	120.7	0.0484	9.264	0.901
6	3.985	1139.4	325	0.0549	28.039	0.878

Diazo coupling catalyzed by 5-G (5 μM):



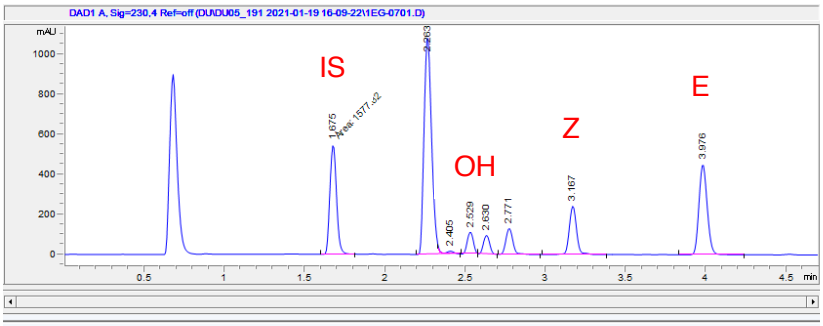
#	Time	Area	Height	Width	Area%	Symmetry
1	1.669	1362.5	473.2	0.048	15.210	0
2	2.259	3657.6	1117.9	0.0522	40.829	0.793
3	2.402	58.7	15.9	0.0551	0.655	0.826
4	2.524	360.6	133	0.0436	4.025	0.99
5	2.626	259.7	95.3	0.0438	2.900	0.965
6	2.767	66.2	22.1	0.0469	0.739	0.888
7	2.838	8.5	2.8	0.0459	0.095	0.477
8	3.163	381.8	123.7	0.05	4.262	0.96
9	3.241	20.2	7.1	0.0411	0.226	0.529
10	3.966	2782.4	764.3	0.0607	31.060	0.82

Diazo coupling catalyzed by 1-SGH (50 nM):



#	Time	Area	Height	Width	Area%	Symmetry
1	1.654	1180	424.3	0.0444	27.097	0.913
2	1.764	25.4	6.8	0.0574	0.584	1.097
3	2.506	806.8	303.8	0.045	18.526	0.964
4	2.607	821.5	302.8	0.0436	18.863	0.955
5	2.744	356.1	118.2	0.0471	8.177	0.866
6	3.145	536.4	176.4	0.0495	12.316	0.951
7	3.226	23	7.9	0.0439	0.527	0.58
8	3.959	605.7	175.9	0.0542	13.909	0.901

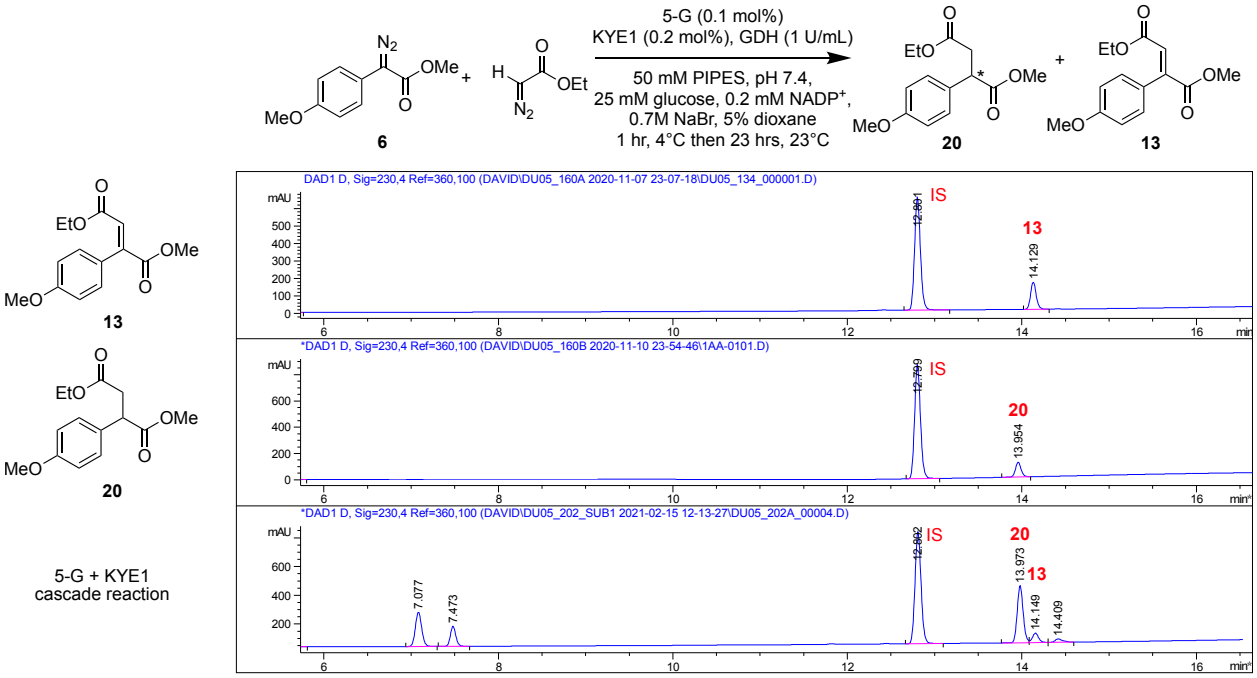
Diazo coupling catalyzed by 5-G (50 nM):



#	Time	Area	Height	Width	Area%	Symmetry
1	1.675	1577.8	550.9	0.0477	18.564	0.914
2	2.263	3571.9	1096.1	0.052	42.025	0.803
3	2.405	50.9	14.8	0.0521	0.598	1.077
4	2.529	293.8	110.4	0.043	3.457	0.965
5	2.63	255	94.4	0.0435	3.001	0.966
6	2.771	390.2	130.2	0.047	4.590	0.884
7	3.167	754.1	242.8	0.0502	8.872	0.901
8	3.976	1605.9	452.1	0.0554	18.894	0.855

Substrate Scope

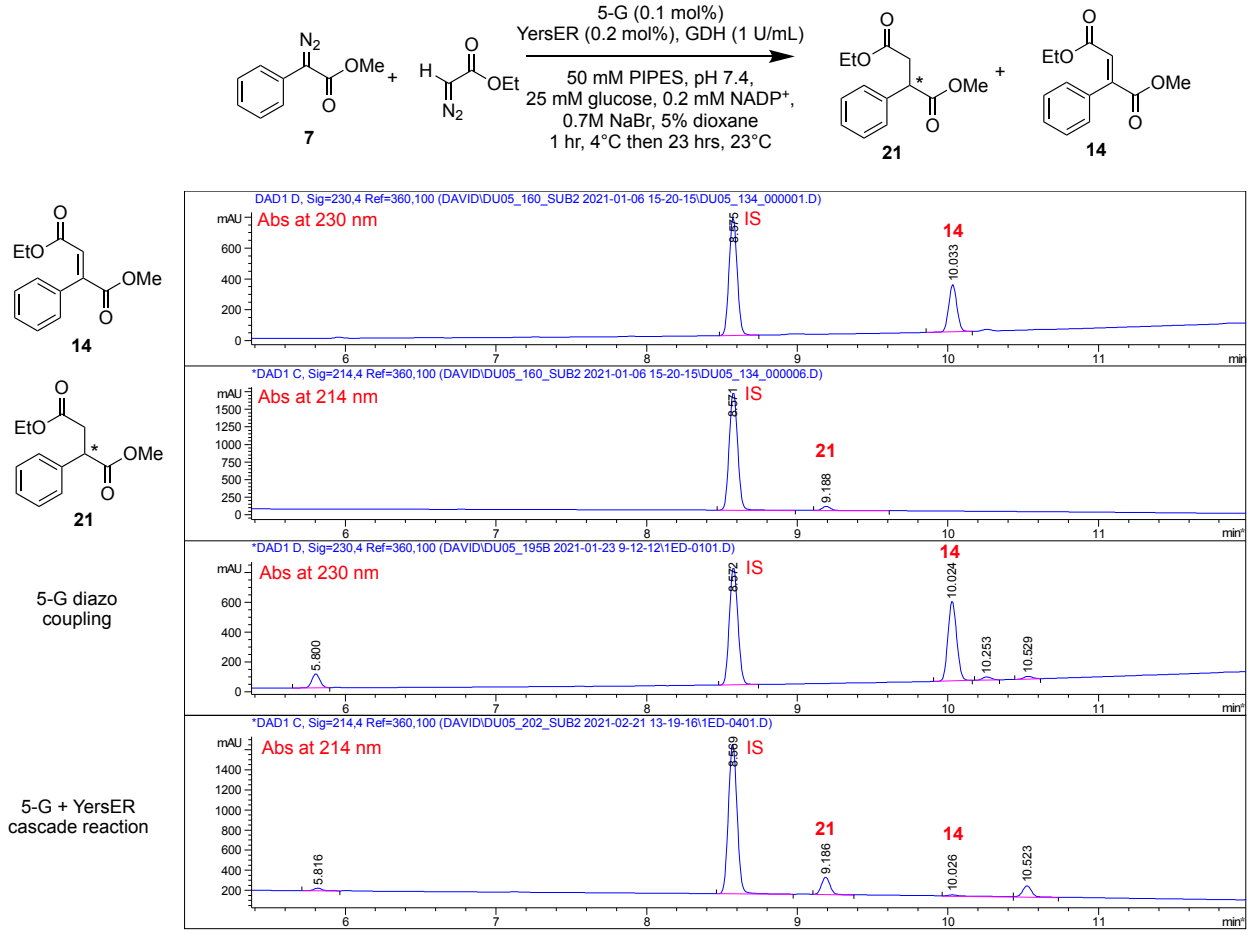
5-G + KYE1 cascade reaction with methyl (4-methoxyphenyl)diazoacetate (**6**) + ethydiazoacetate: LC method: 15cm Eclipse Plus C18 column was used. Mobile phase used was A: H_2O + 0.1% FA, B: ACN + 0.1% FA. Method: 10% B for 0.5 mins, 10% B to 90% B over 19.5 minutes, 3 minutes at 90% B.



5-G + YersER cascade reaction with methyl phenyldiazoacetate (**7**) + ethydiazoacetate:

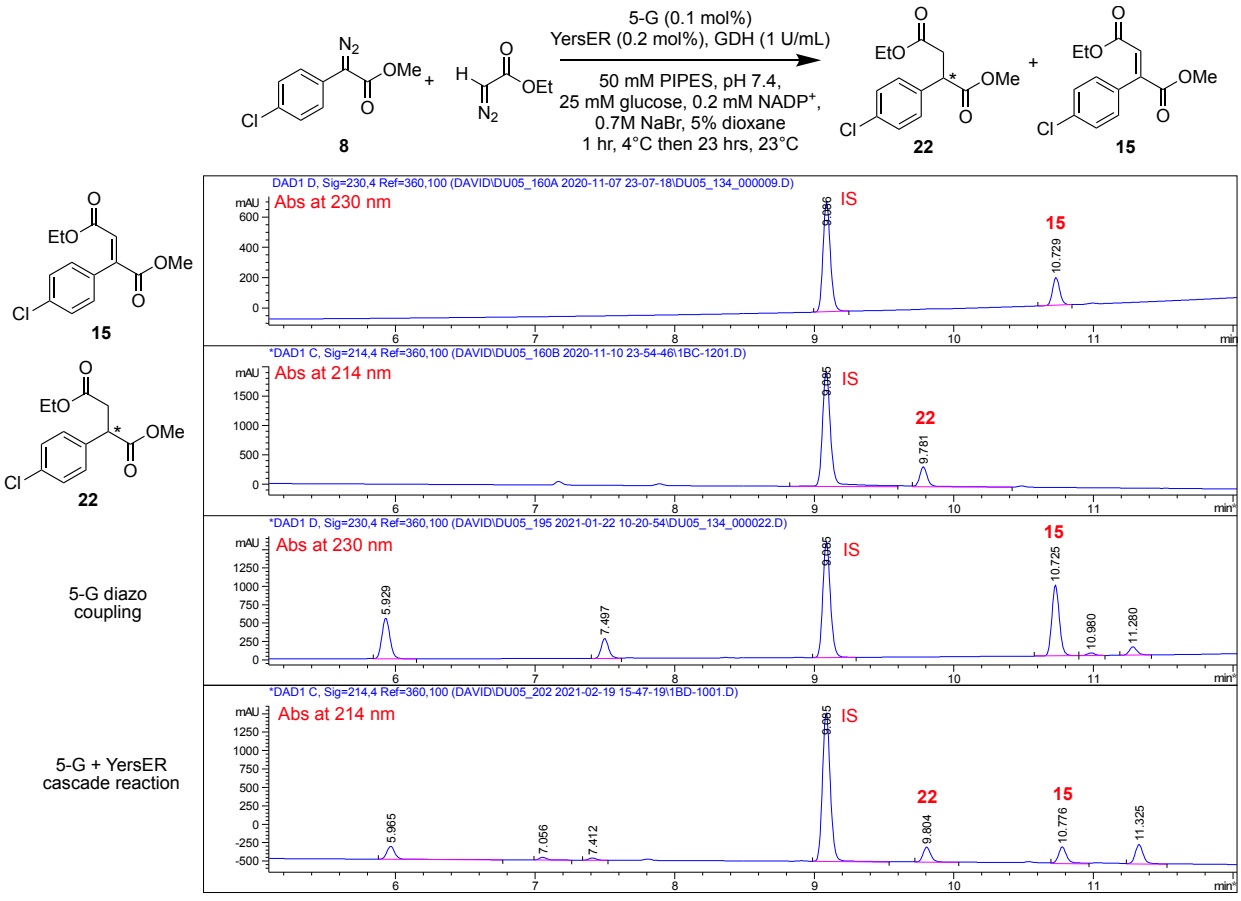
Note: absorbance of the alkane product **21** at 230 nm is too low for accurate peak integration. Integration at 214 nm was used for alkane **21** compared to the integration of the internal standard and alkene **14** at 230 nm.

LC method: 15cm Eclipse Plus C18 column was used. Mobile phase used was A: H_2O + 0.1% FA, B: ACN + 0.1% FA. Method: 10% B for 0.5 mins, 10% B to 80% B over 7 minutes, 3 minutes at 90% B.



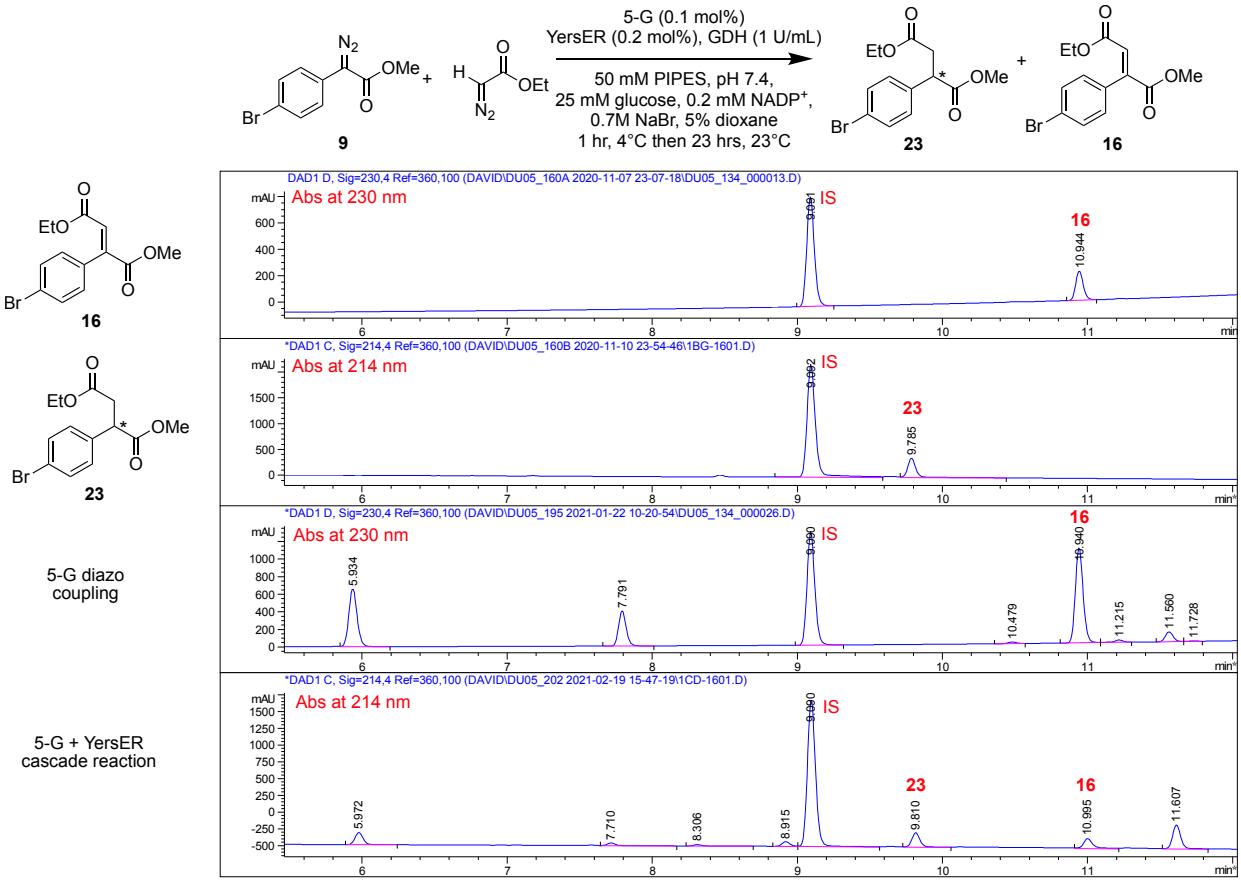
5-G + YersER cascade reaction with methyl (4-chlorophenyl)diazoacetate (**8**) + ethydiazoacetate: Note: absorbance of the alkane **22** at 230 nm is too low for accurate peak integration. Integration at 214 nm was used for alkane **22** compared to the integration of the internal standard and alkene **15** at 230 nm.

LC method: 15cm Eclipse Plus C18 column was used. Mobile phase used was A: H_2O + 0.1% FA, B: ACN + 0.1% FA. Method: 10% B for 0.5 mins, 10% B to 90% B over 8 minutes, 2 minutes at 90% B.



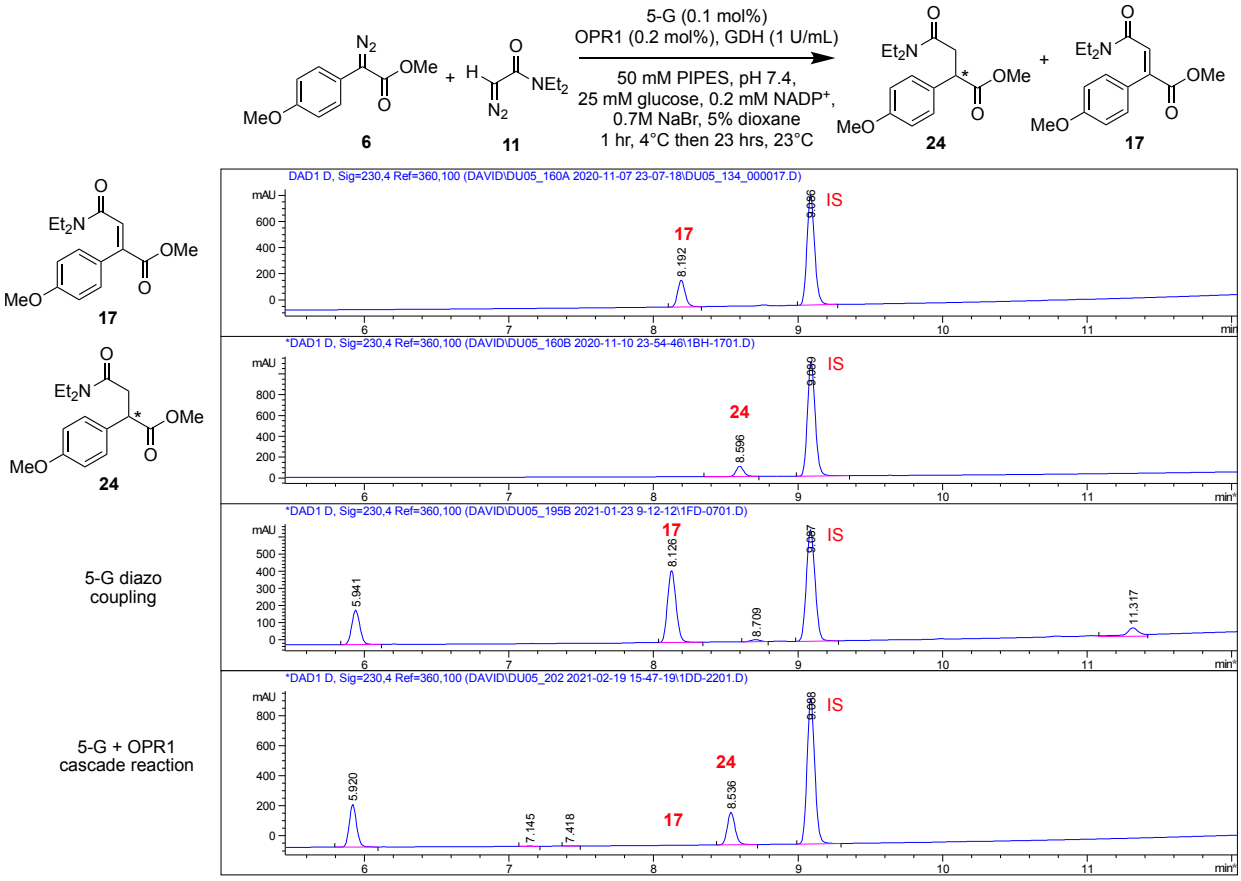
5-G + YersER cascade reaction with methyl (4-bromophenyl)diazoacetate (**9**) + ethydiazoacetate: Note: absorbance of the alkane **23** at 230 nm is too low for accurate peak integration. Integration at 214 nm was used for alkane **23** compared to the integration of the internal standard and alkene **16** at 230 nm.

LC method: 15cm Eclipse Plus C18 column was used. Mobile phase used was A: H_2O + 0.1% FA, B: ACN + 0.1% FA. Method: 10% B for 0.5 mins, 10% B to 90% B over 8 minutes, 2 minutes at 90% B.



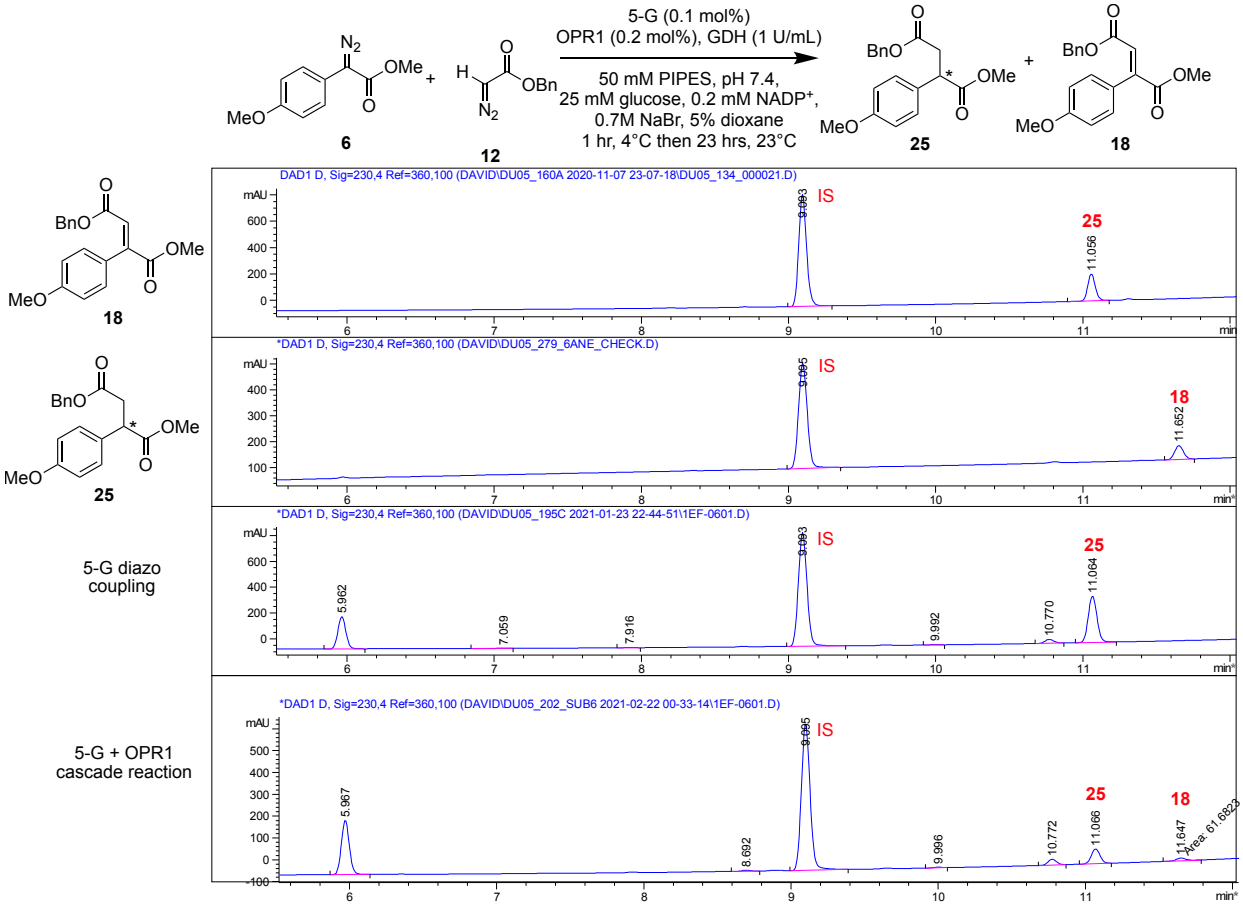
5-G + OPR1 cascade reaction with methyl (4-methoxyphenyl)diazoacetate (**6**) + diethydiazoaceamide (**11**):

LC method: 15cm Eclipse Plus C18 column was used. Mobile phase used was A: H_2O + 0.1% FA, B: ACN + 0.1% FA. Method: 10% B for 0.5 mins, 10% B to 90% B over 8 minutes, 2 minutes at 90% B.



5-G + OPR1 cascade reaction with methyl (4-methoxyphenyl)diazoacetate (**6**) + benzyl diazoacetate (**12**):

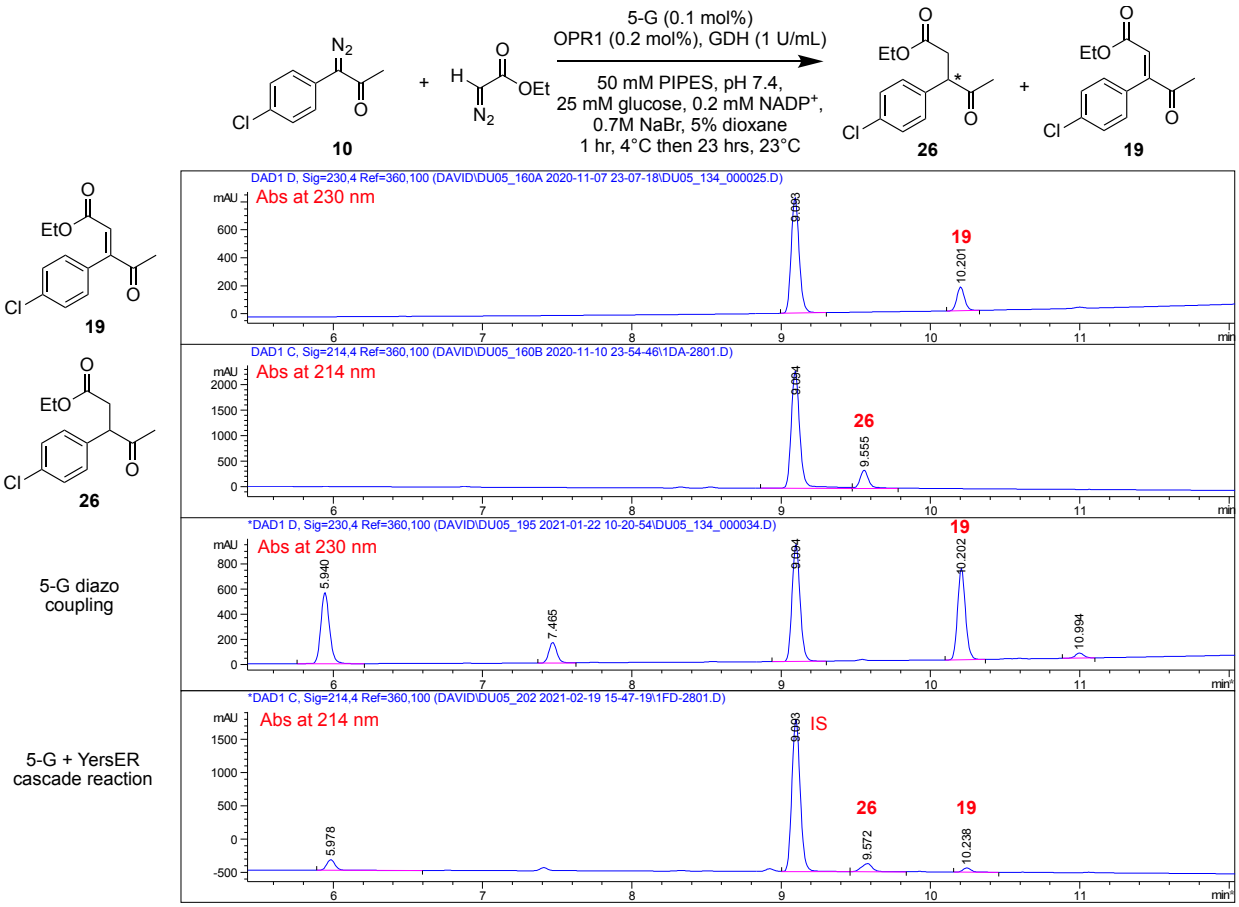
LC method: 15cm Eclipse Plus C18 column was used. Mobile phase used was A: H_2O + 0.1% FA, B: ACN + 0.1% FA. Method: 10% B for 0.5 mins, 10% B to 90% B over 8 minutes, 2 minutes at 90% B.



5-G + YersER cascade reaction with 1-(4-chlorophenyl)-1-diazopropan-2-one (**9**) + ethydiazoacetate:

Note: absorbance of the alkane **26** at 230 nm is too low for accurate peak integration. Integration at 214 nm was used for alkane **26** compared to the integration of the internal standard and alkene **19** at 230 nm.

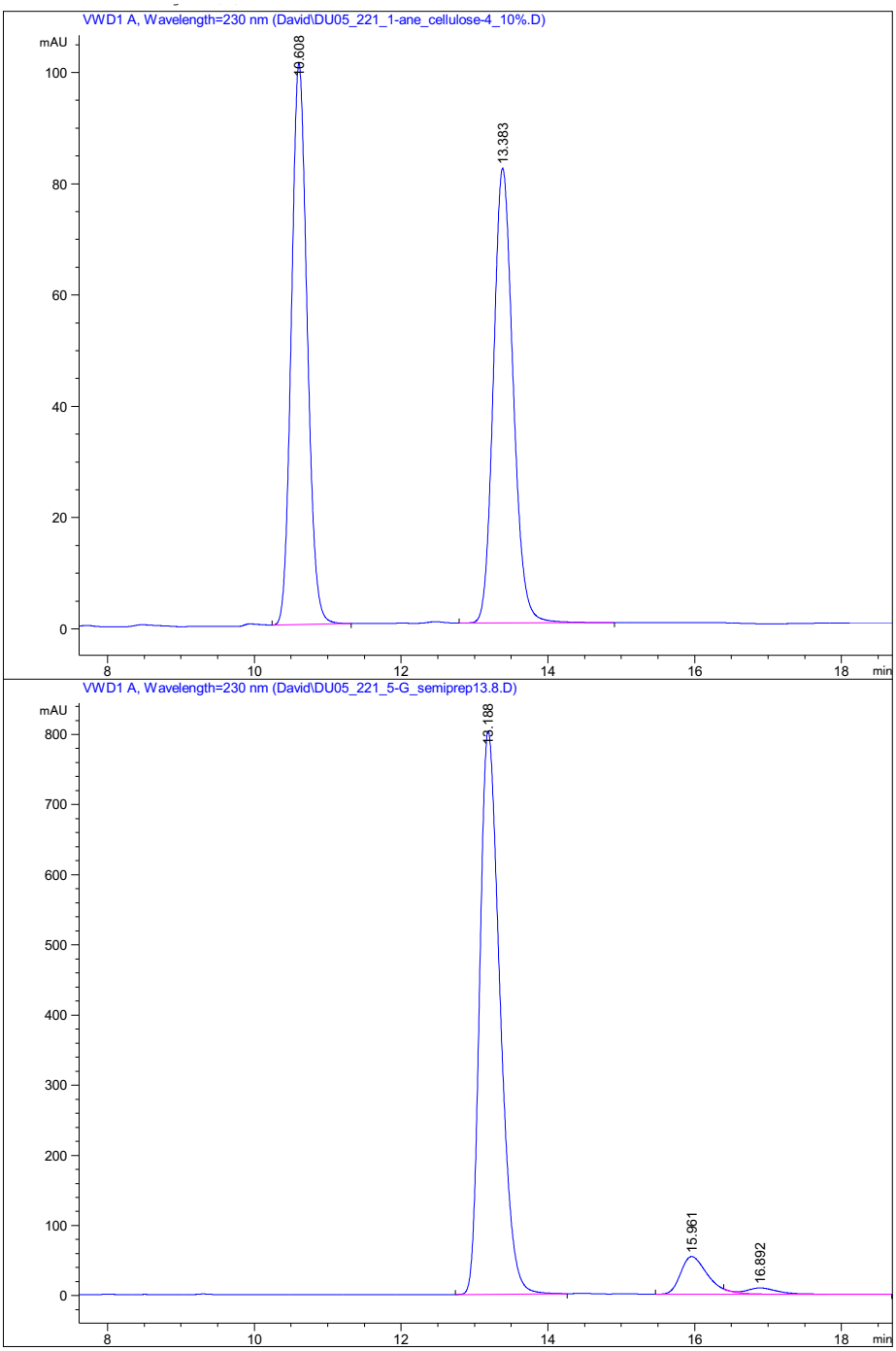
LC method: 15cm Eclipse Plus C18 column was used. Mobile phase used was A: H_2O + 0.1% FA, B: ACN + 0.1% FA. Method: 10% B for 0.5 mins, 10% B to 90% B over 8 minutes, 2 minutes at 90% B.



Enantioselectivity

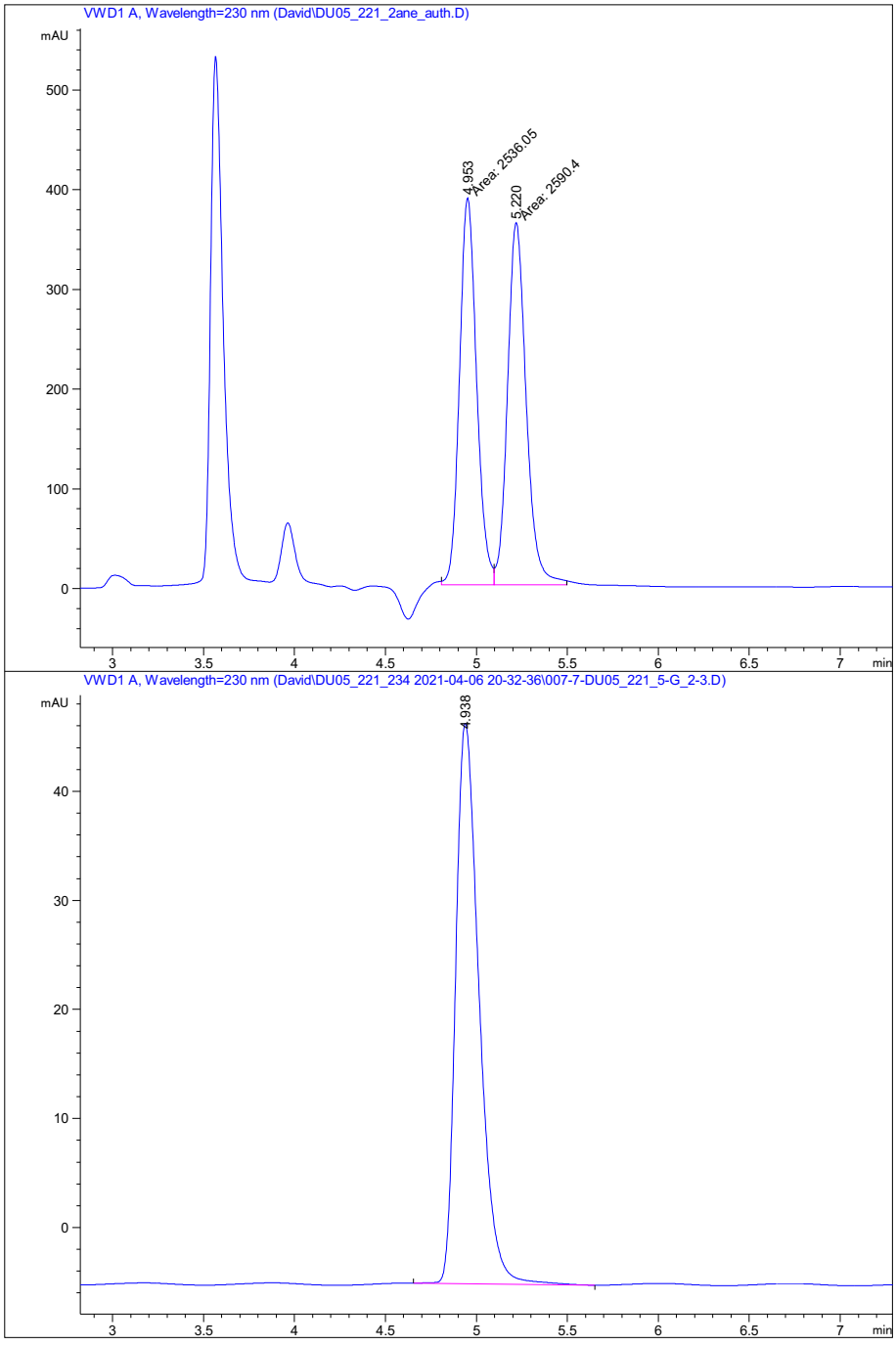
Alkane **20** authentic (top) and 5-G/KYE1 cascade reaction product (bottom):

LC method: 25cm Lux Cellulose-4 column was used. Mobile phase used was A: iPrOH, B: Hexanes. Method: 10% A for 25 mins, 10% A to 25% A over 5 minutes, 5 minutes at 25% A.



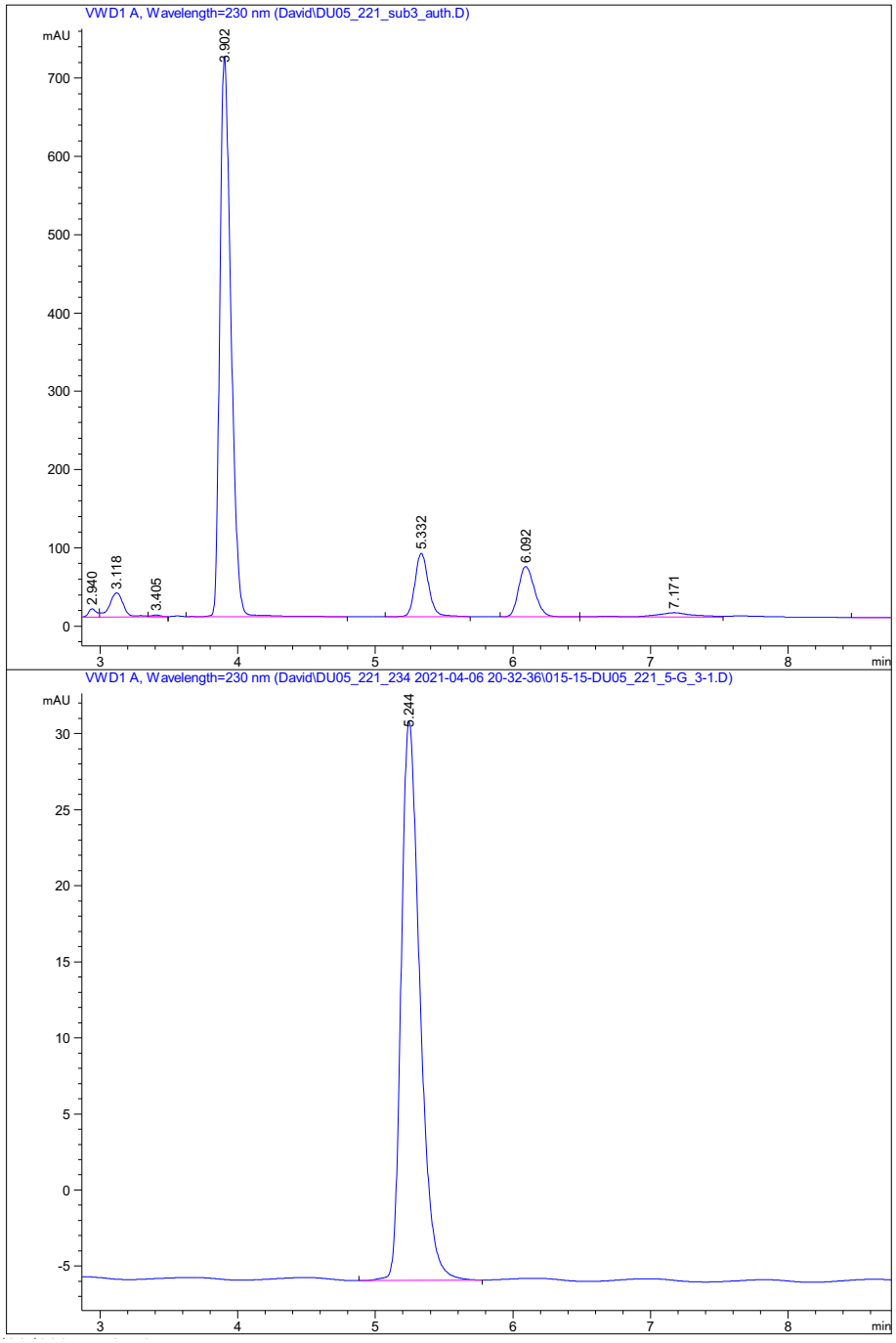
Alkane **21** authentic (top) and 5-G/YersER cascade reaction product (bottom):

LC method: 25cm Lux Cellulose-1 column was used. Mobile phase used was A: iPrOH, B: Hexanes. Method: 1% A for 25 mins, 1% A to 25% A over 5 minutes, 5 minutes at 25% A.



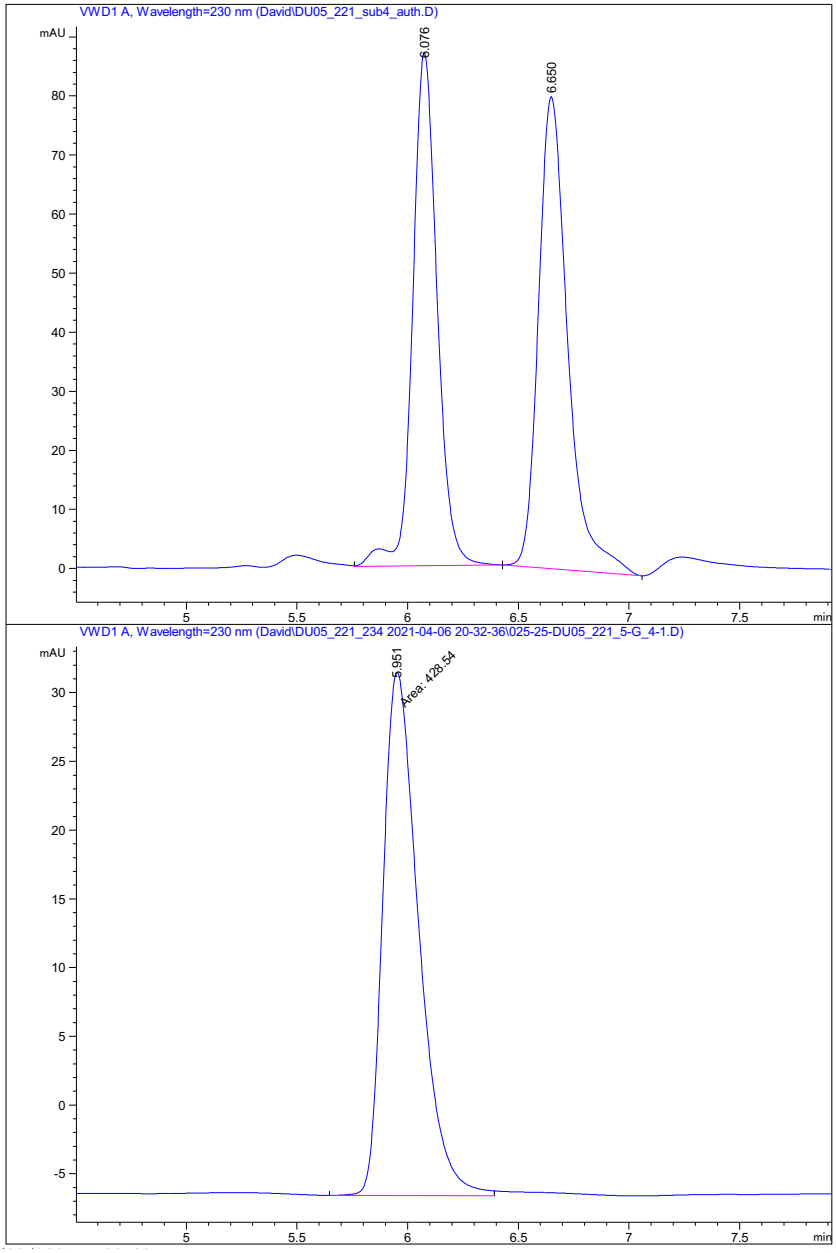
Alkane **22** authentic (top) and 5-G/YersER cascade reaction product (bottom):

LC method: 25cm Lux Cellulose-1 column was used. Mobile phase used was A: iPrOH, B: Hexanes. Method: 1.2% A for 25 mins, 1.2% A to 25% A over 5 minutes, 5 minutes at 25% A.



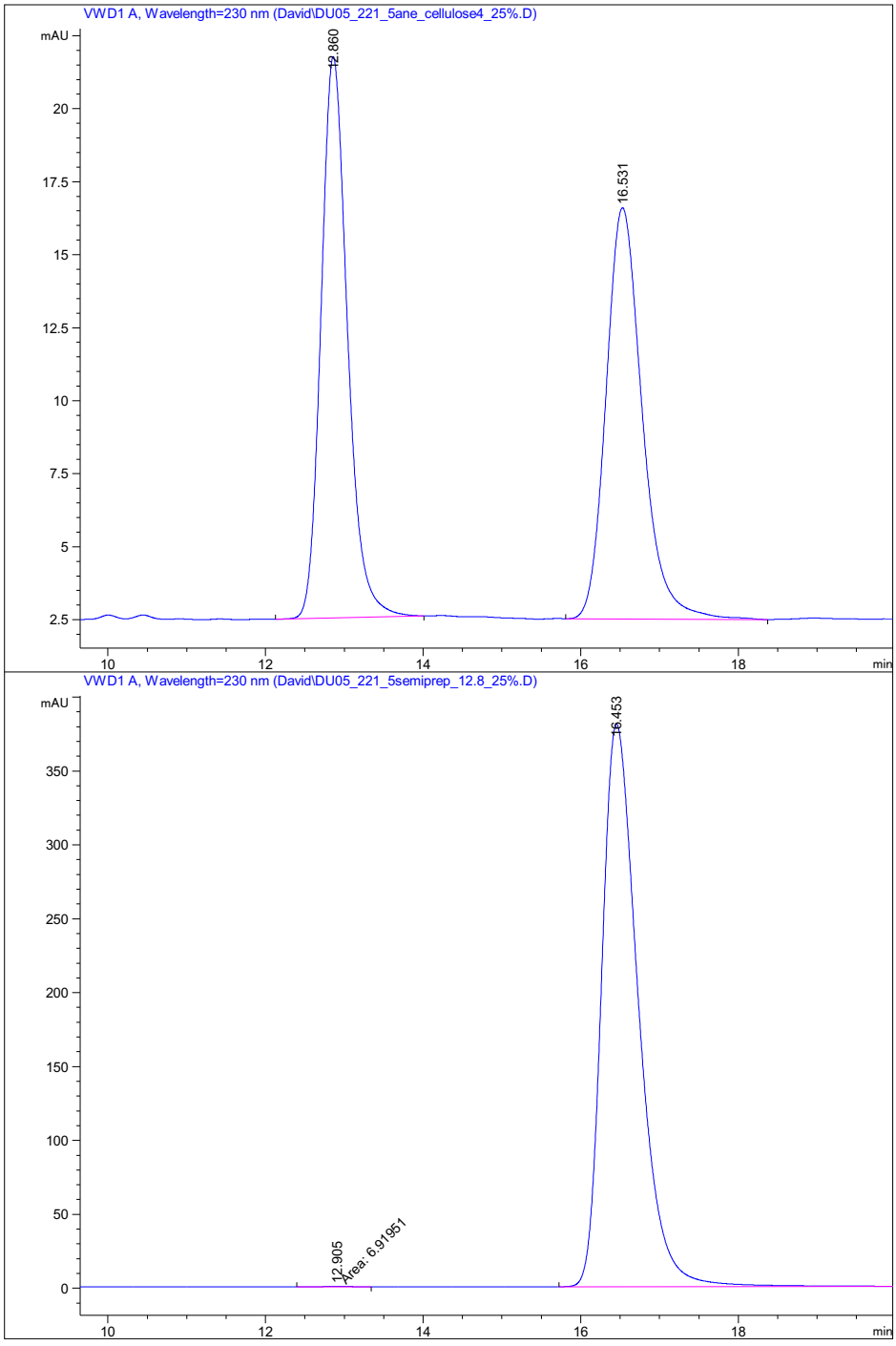
Alkane **23** authentic (top) and 5-G/YersER cascade reaction product (bottom):

LC method: 25cm Lux Cellulose-1 column was used. Mobile phase used was A: iPrOH, B: Hexanes. Method: 1.2% A for 25 mins, 1.2% A to 25% A over 5 minutes, 5 minutes at 25% A.



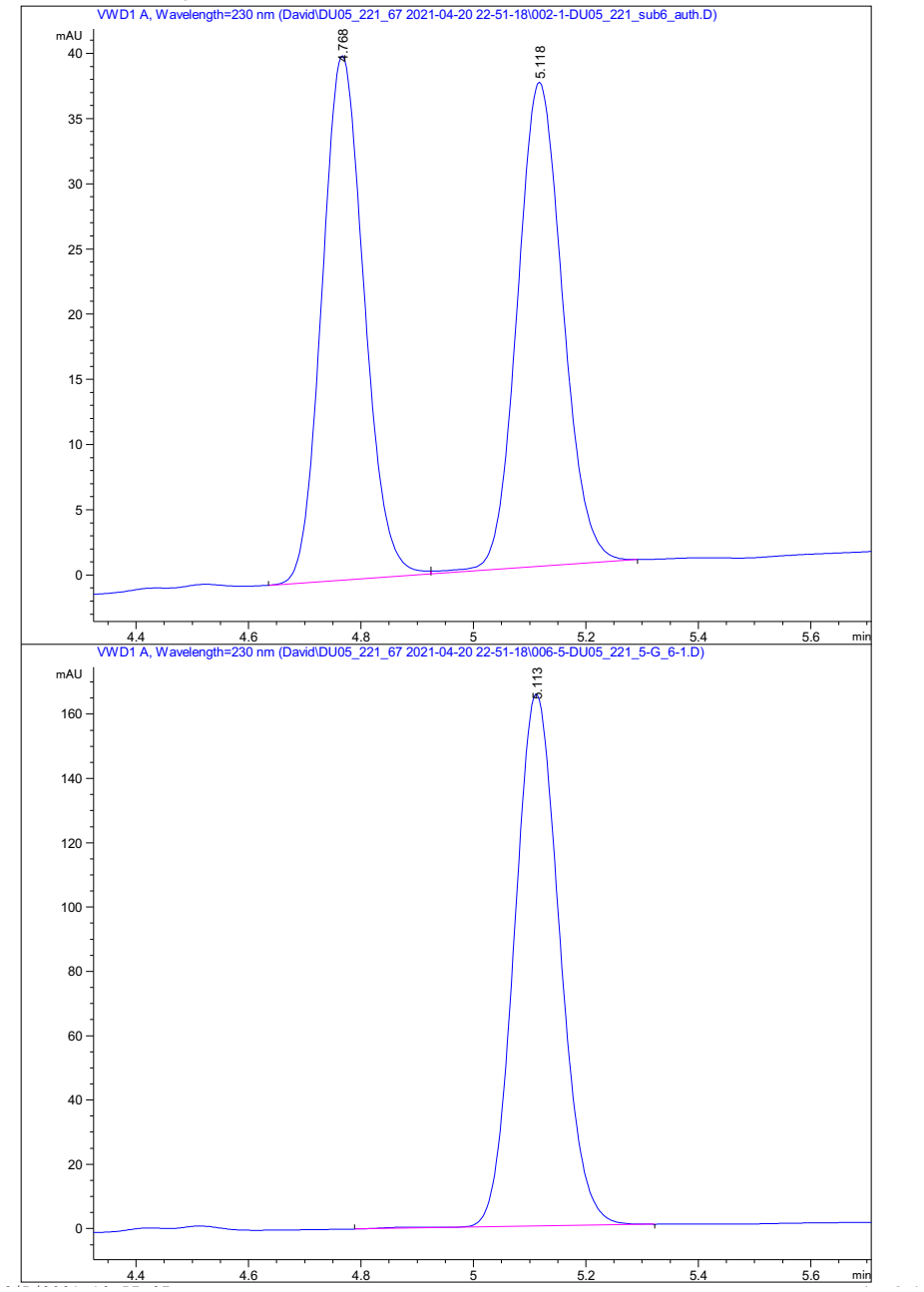
Alkane **24** authentic (top) and 5-G/OPR1 cascade reaction product (bottom):

LC method: 25cm Lux Cellulose-4 column was used. Mobile phase used was A: iPrOH, B: Hexanes. Method: 25% A for 25 mins, 25% A to 35% A over 5 minutes, 5 minutes at 35% A.



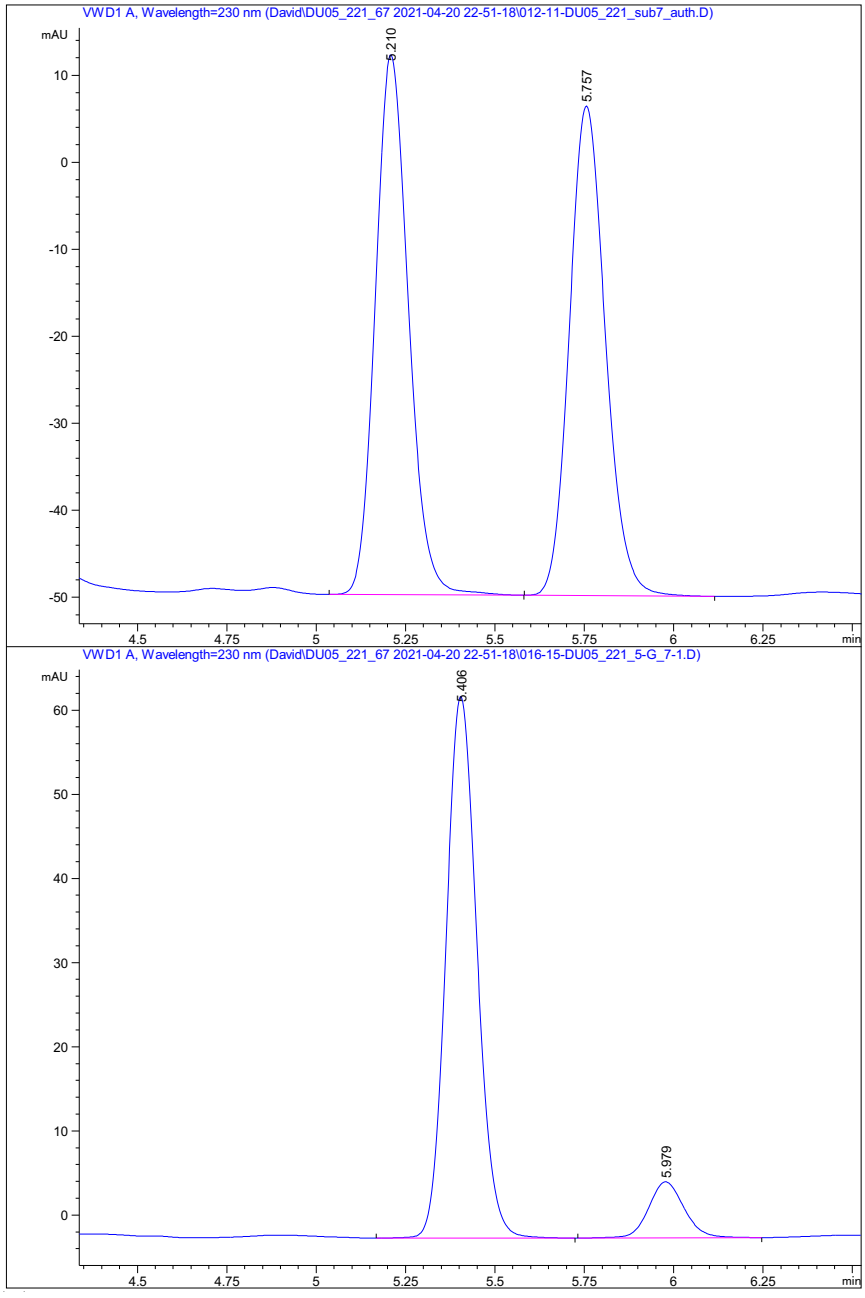
Alkane **25** authentic (top) and 5-G/OPR1 cascade reaction product (bottom):

LC method: 25cm Lux Cellulose-4 column was used. Mobile phase used was A: iPrOH, B: Hexanes. Method: 1.4% A for 25 mins, 1.4% A to 25% A over 5 minutes, 5 minutes at 25% A.



Alkane **26** authentic (top) and 5-G/YersER cascade reaction product (bottom):

LC method: 25cm Lux Cellulose-1 column was used. Mobile phase used was A: iPrOH, B: Hexanes. Method: 0.2% A for 25 mins, 0.2% A to 25% A over 5 minutes, 5 minutes at 25% A.



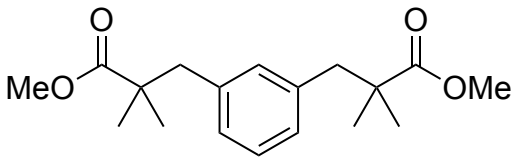
Signal 1: VWD1 A, Wavelength=230 nm

Peak #	RetTime [min]	Type	Width [min]	Area [mAU*s]	Height [mAU]	Area %
1	5.406	BB	0.0930	386.54861	64.38349	89.3696
2	5.979	BB	0.1066	45.97926	6.65840	10.6304

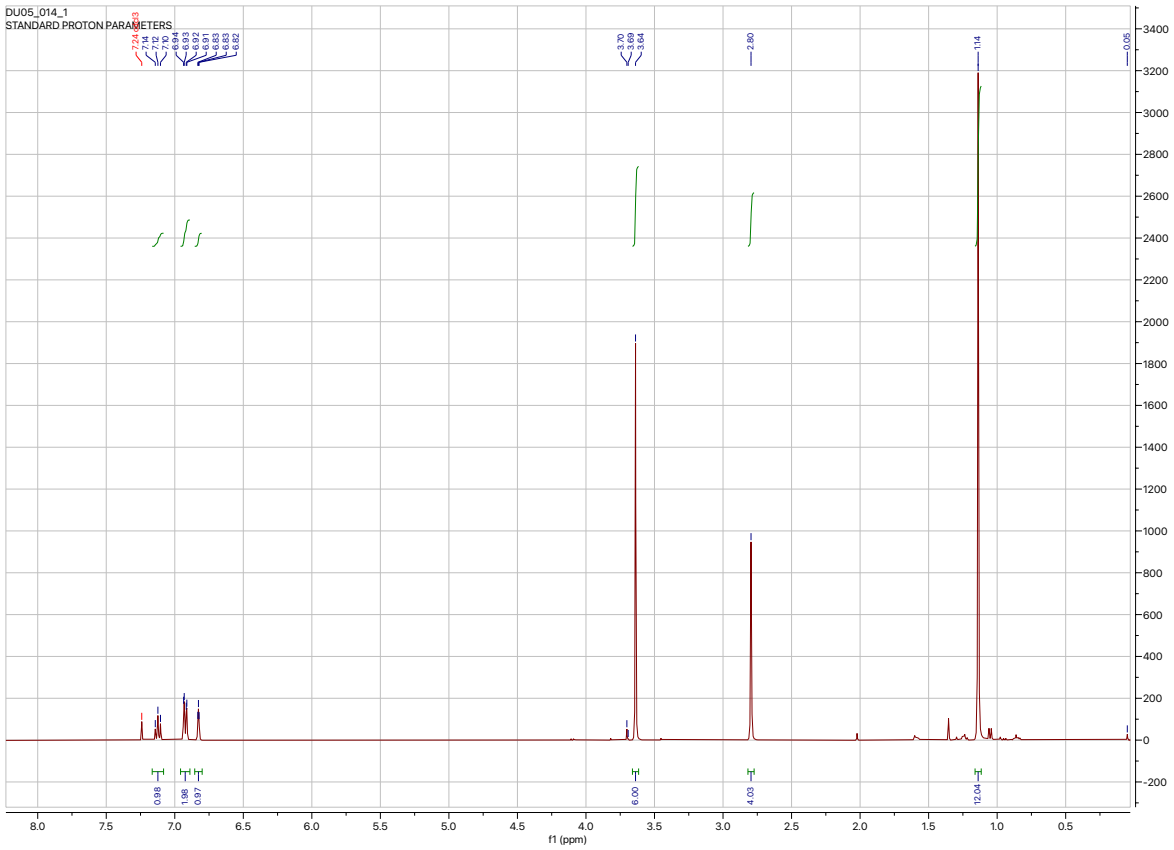
3.4.9 NMR Spectra of Novel Compounds

Cofactor precursors

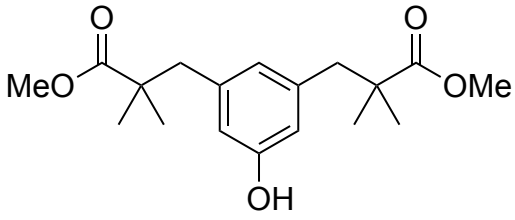
3, 1,3-dimethyl $\alpha^1,\alpha^1,\alpha^3,\alpha^3$-tetramethyl-1,3-benzenedipropanoate



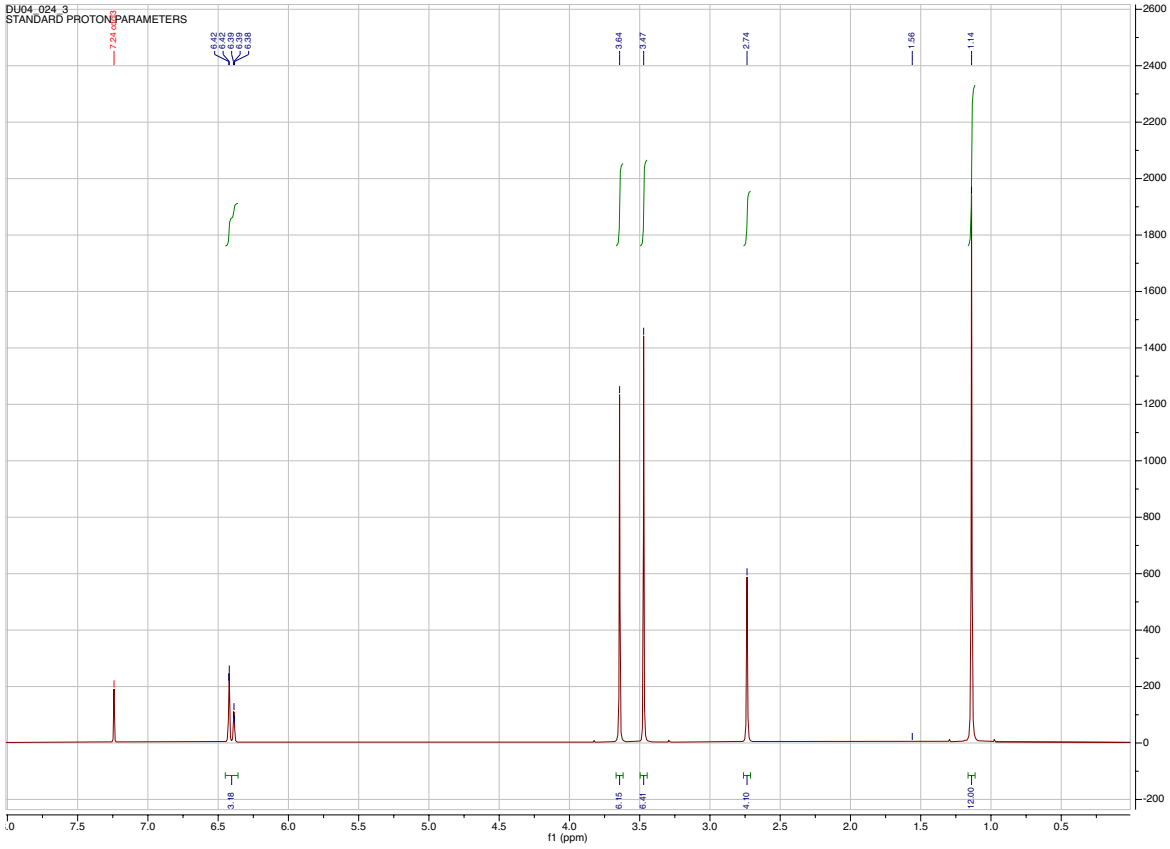
3



4, 1,3-dimethyl 5-hydroxy-α^1,α^1,α^3,α^3-tetramethyl-1,3-benzenedipropanoate

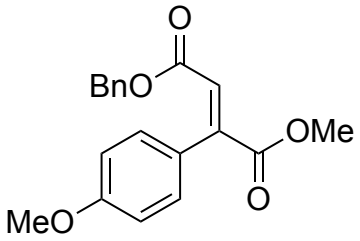


4

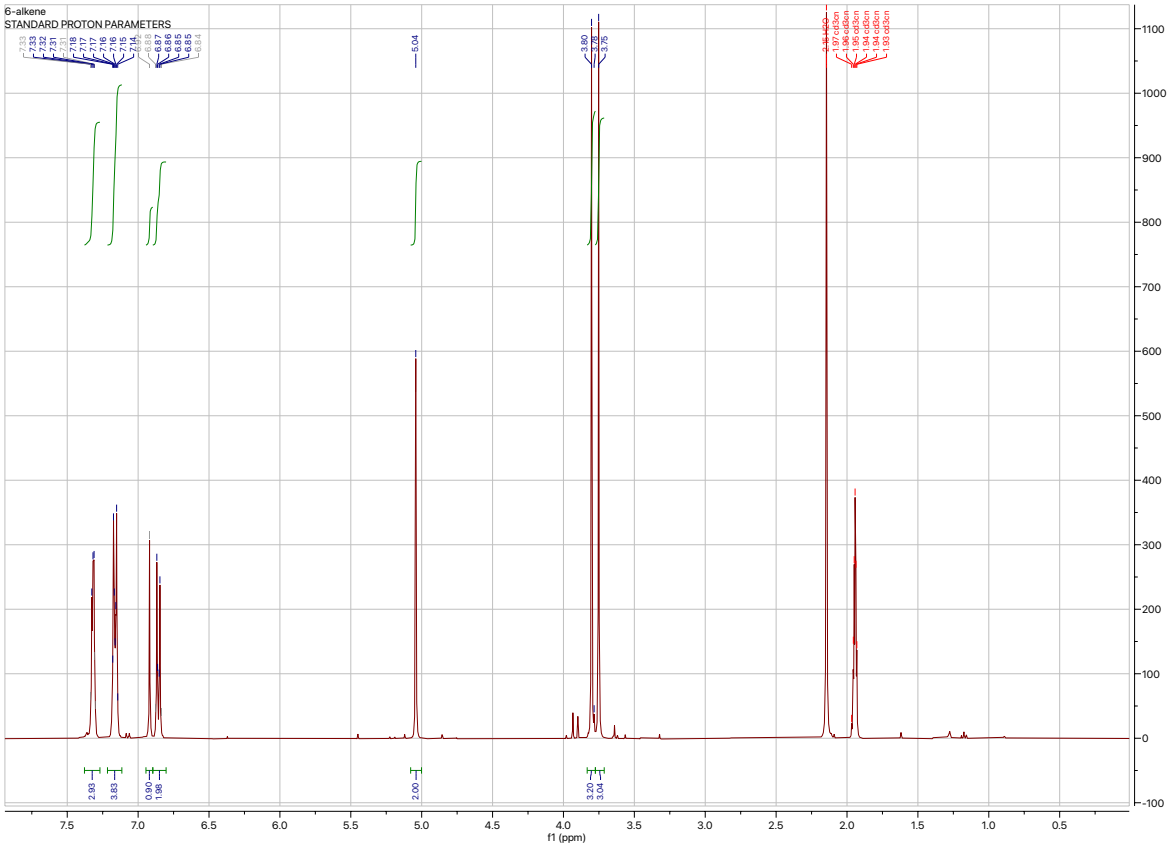


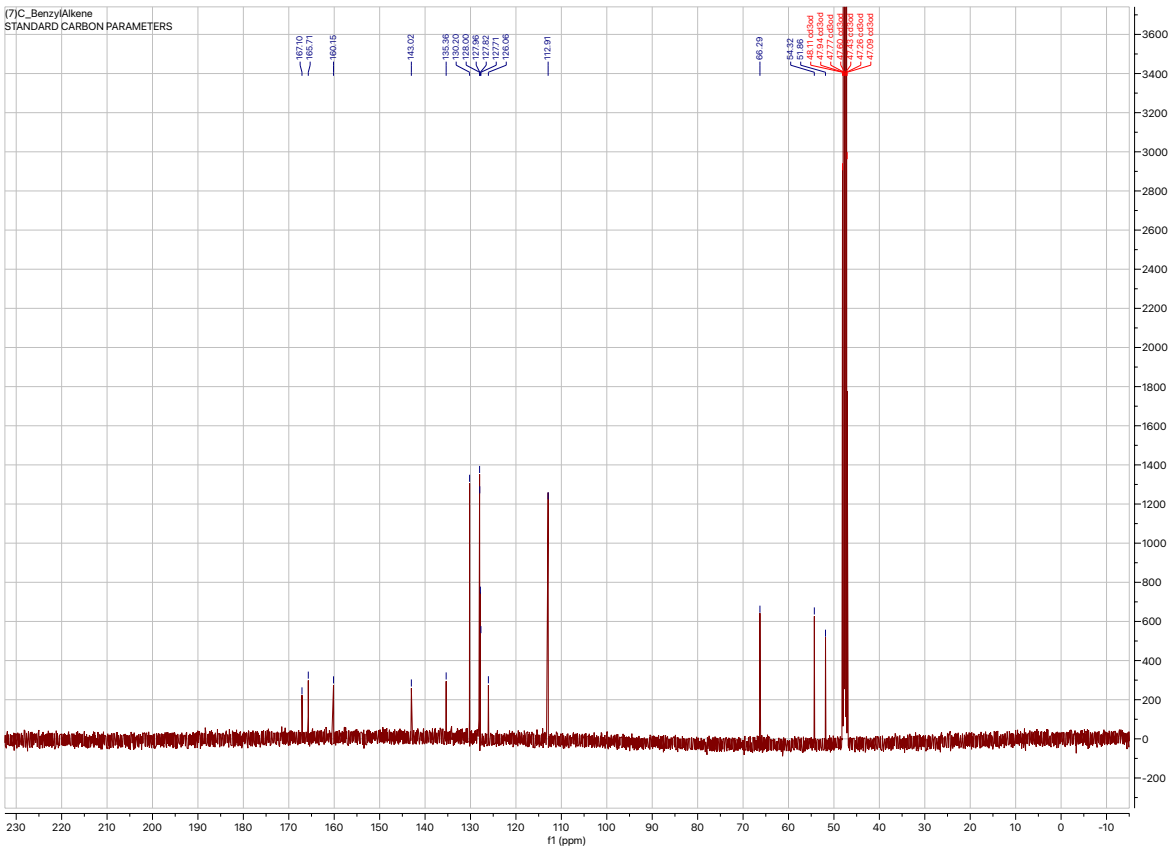
Alkenes

18, 4-benzyl 1-methyl 2-(4-methoxyphenyl)fumarate



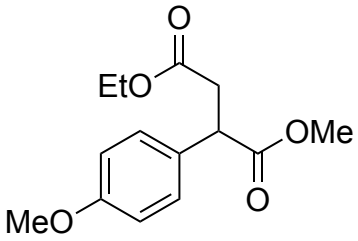
18



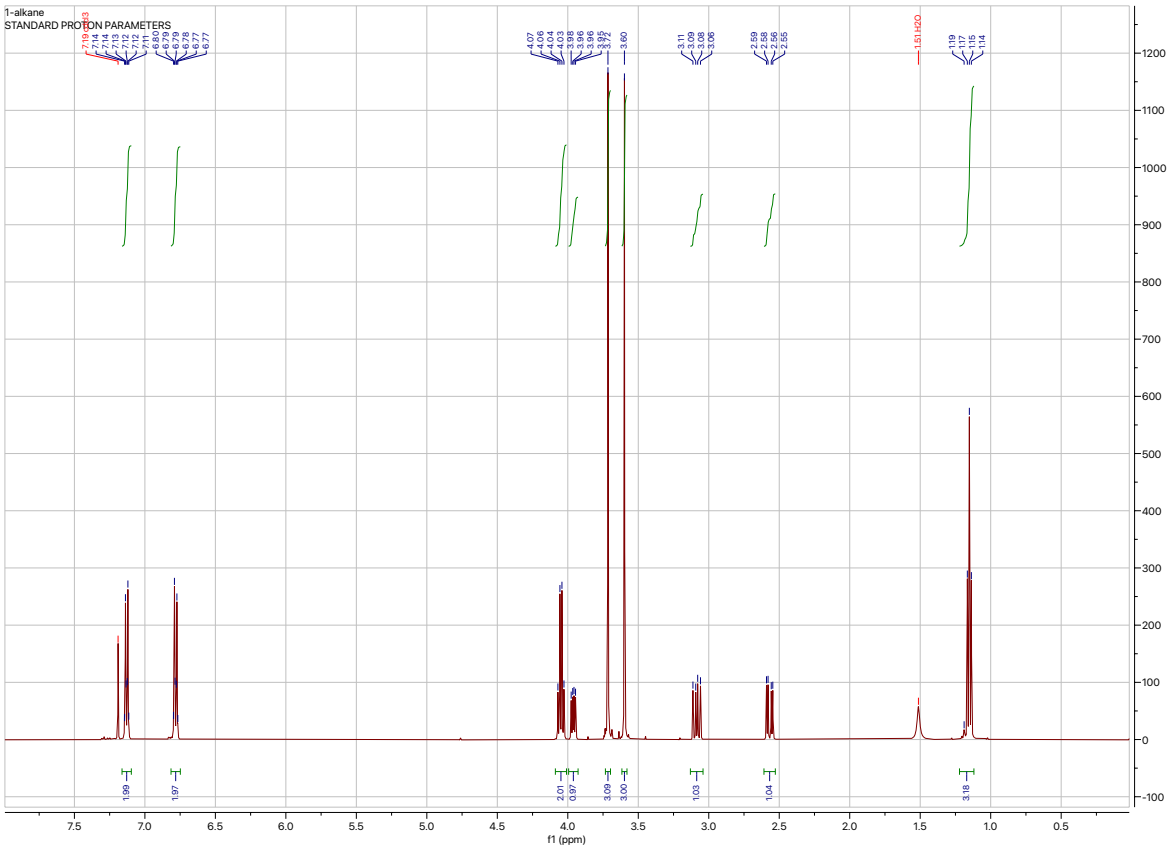

(7)C_BenzylAlkene
STANDARD CARBON PARAMETERS
167.10
165.71
160.15
143.02
135.36
130.20
129.00
127.96
127.82
127.71
126.04
112.91
66.29
54.32
53.95
48.11 cd3od
47.94 cd3od
47.77 cd3od
47.60 cd3od
47.43 cd3od
47.26 cd3od
47.09 cd3od
f1 (ppm)

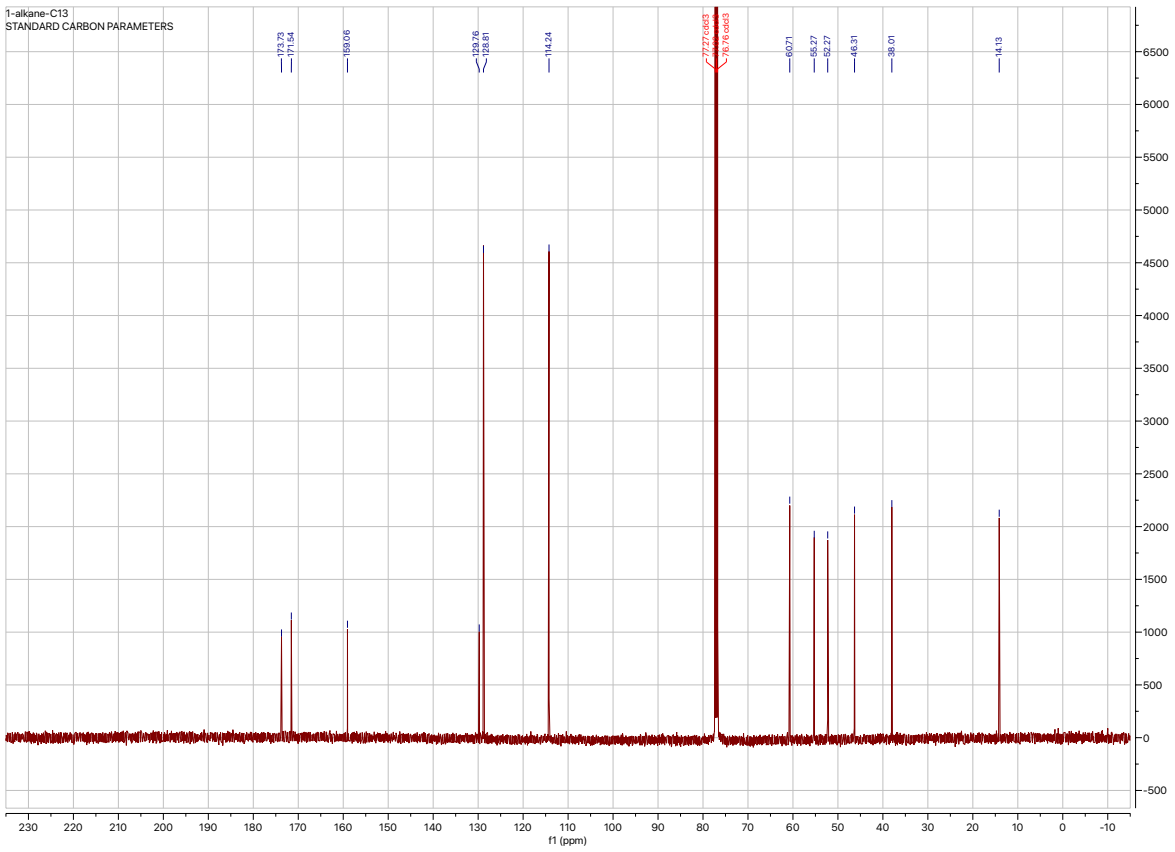
Alkanes

20, 4-ethyl 1-methyl 2-(4-methoxyphenyl)succinate

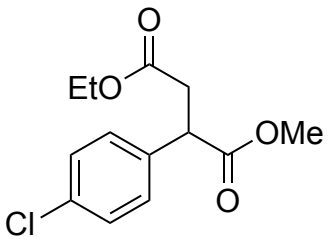


20

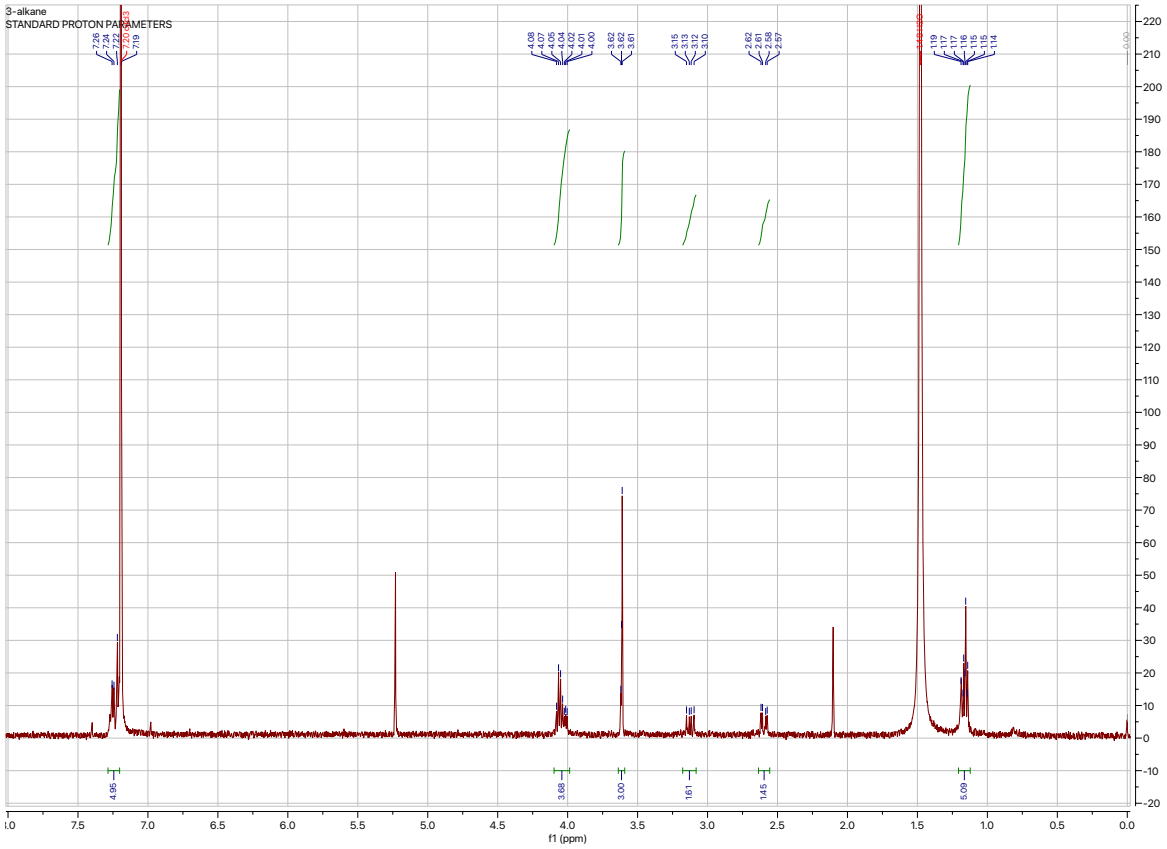


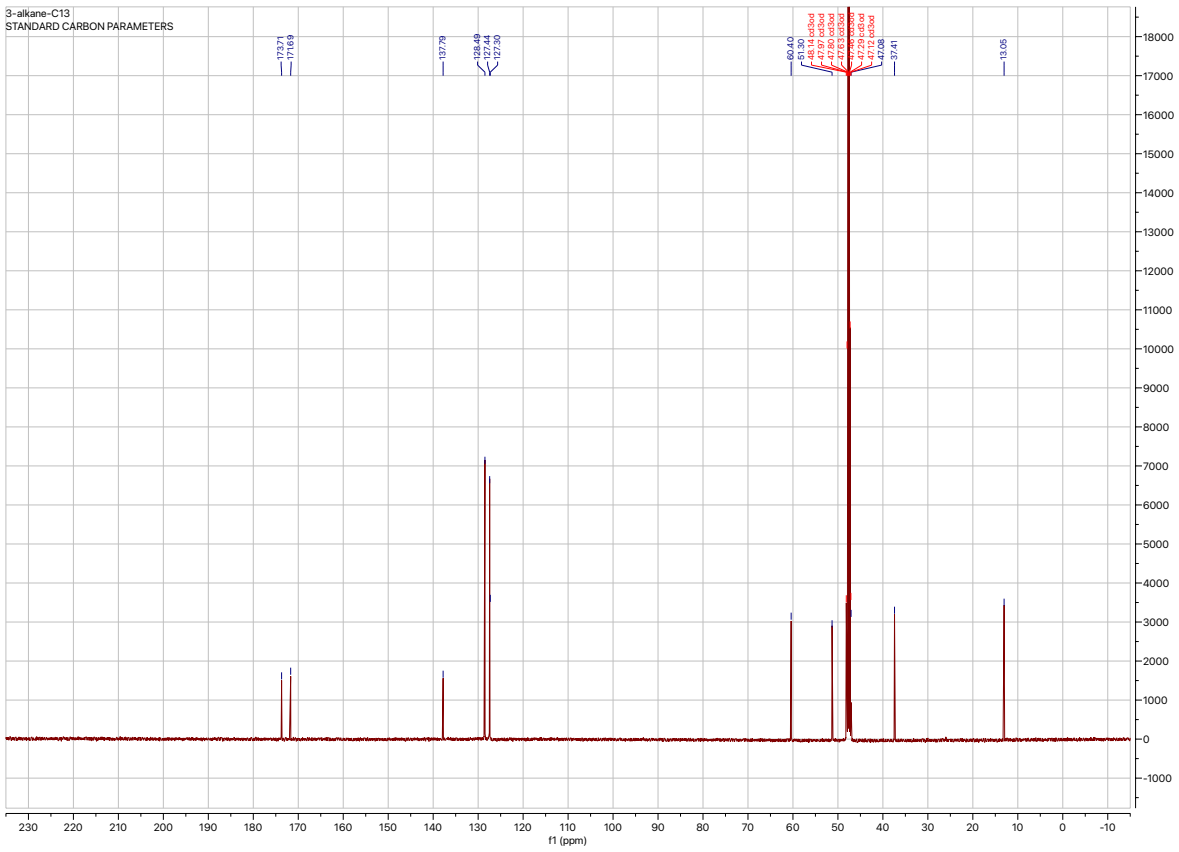

1-alkane-C13
STANDARD CARBON PARAMETERS
173.73
171.54
159.06
129.76
128.81
114.24
77.27 cdcl3
77.00 cdcl3
76.76 cdcl3
60.71
55.27
52.27
46.31
38.01
14.13
f1 (ppm)

22, 4-ethyl 1-methyl 2-(4-chlorophenyl)succinate

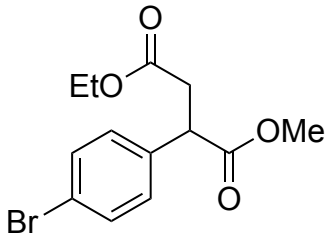


22

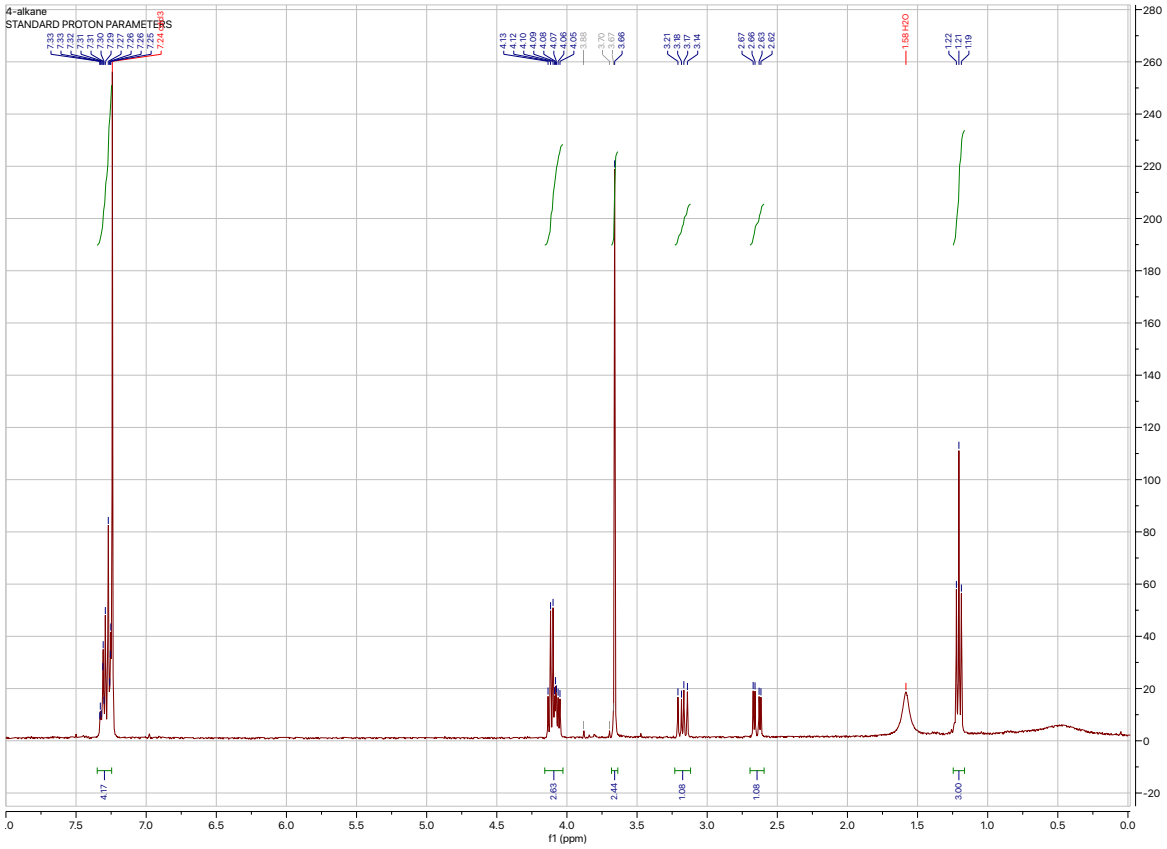


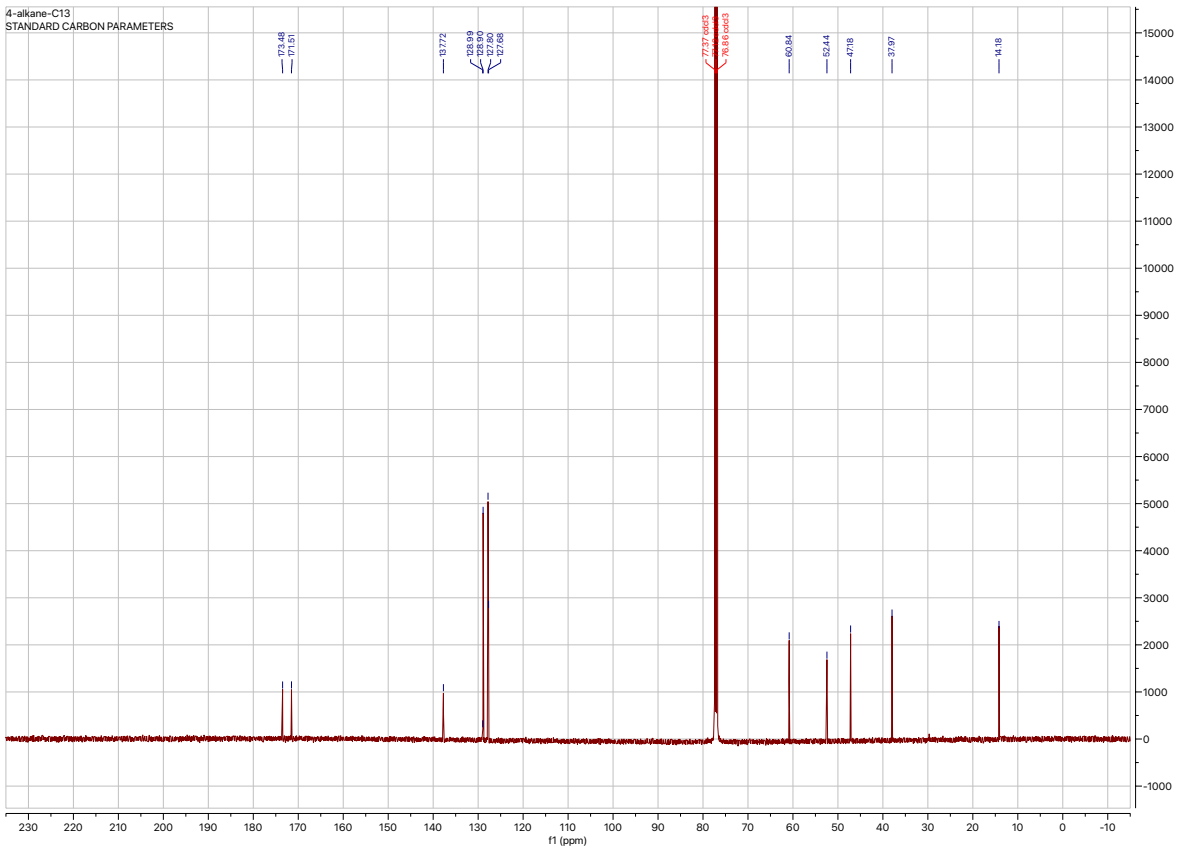

3-alkane-C13
STANDARD CARBON PARAMETERS
173.71
171.69
137.79
128.49
127.44
127.30
60.40
53.90
48.14 cd3od
47.97 cd3od
47.80 cd3od
47.63 cd3od
47.46 cd3od
47.29 cd3od
47.12 cd3od
47.08
37.41
13.05
f1 (ppm)

23, 4-ethyl 1-methyl 2-(4-bromophenyl)succinate

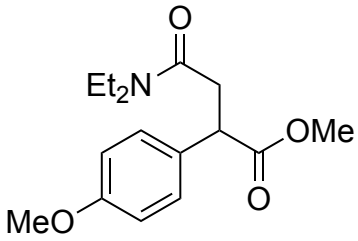


23

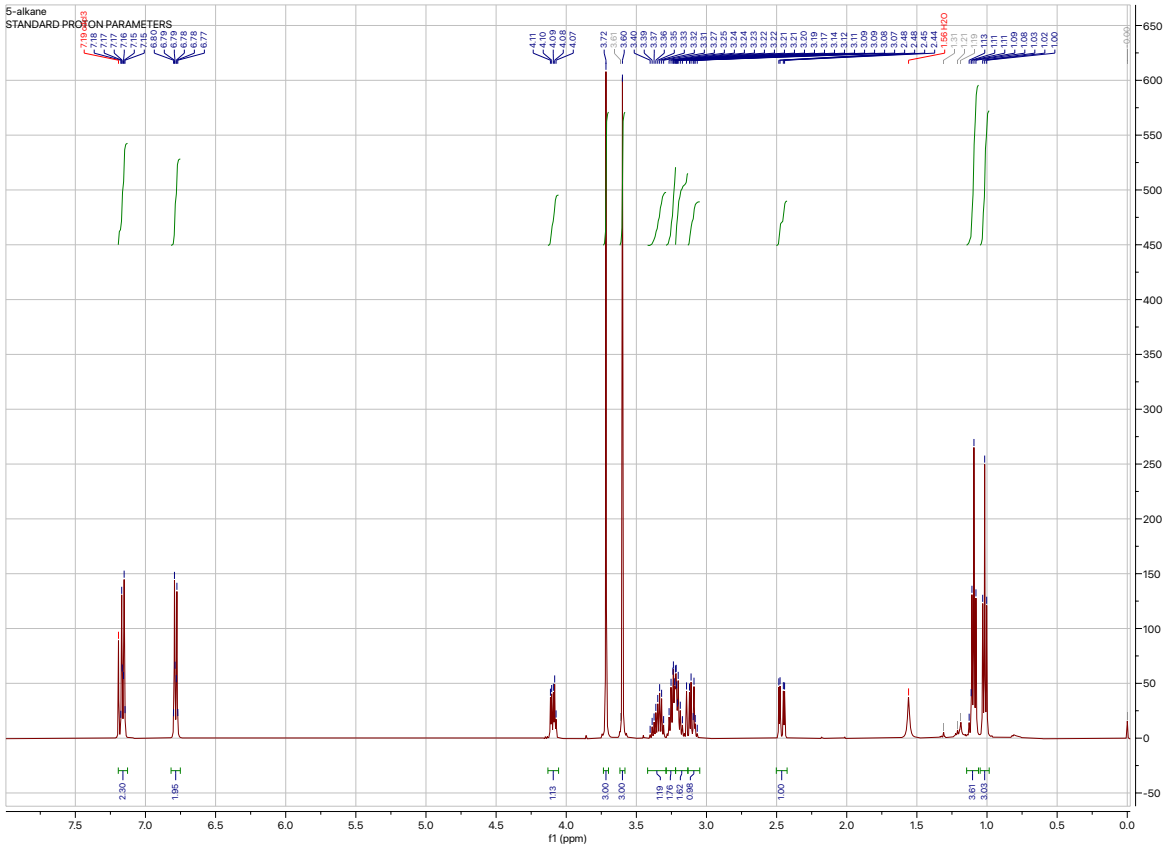



4-alkane-C13
STANDARD CARBON PARAMETERS
173.48
171.51
137.72
128.99
128.60
128.30
127.68
77.27 cdcl3
77.00 cdcl3
76.86 cdcl3
60.84
52.44
47.18
37.97
14.18
f1 (ppm)

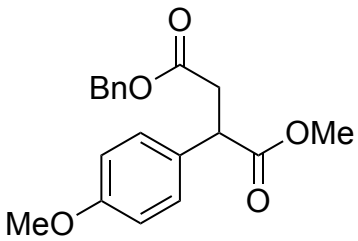
24, methyl 4-(diethylamino)-2-(4-methoxyphenyl)-4-oxobutanoate



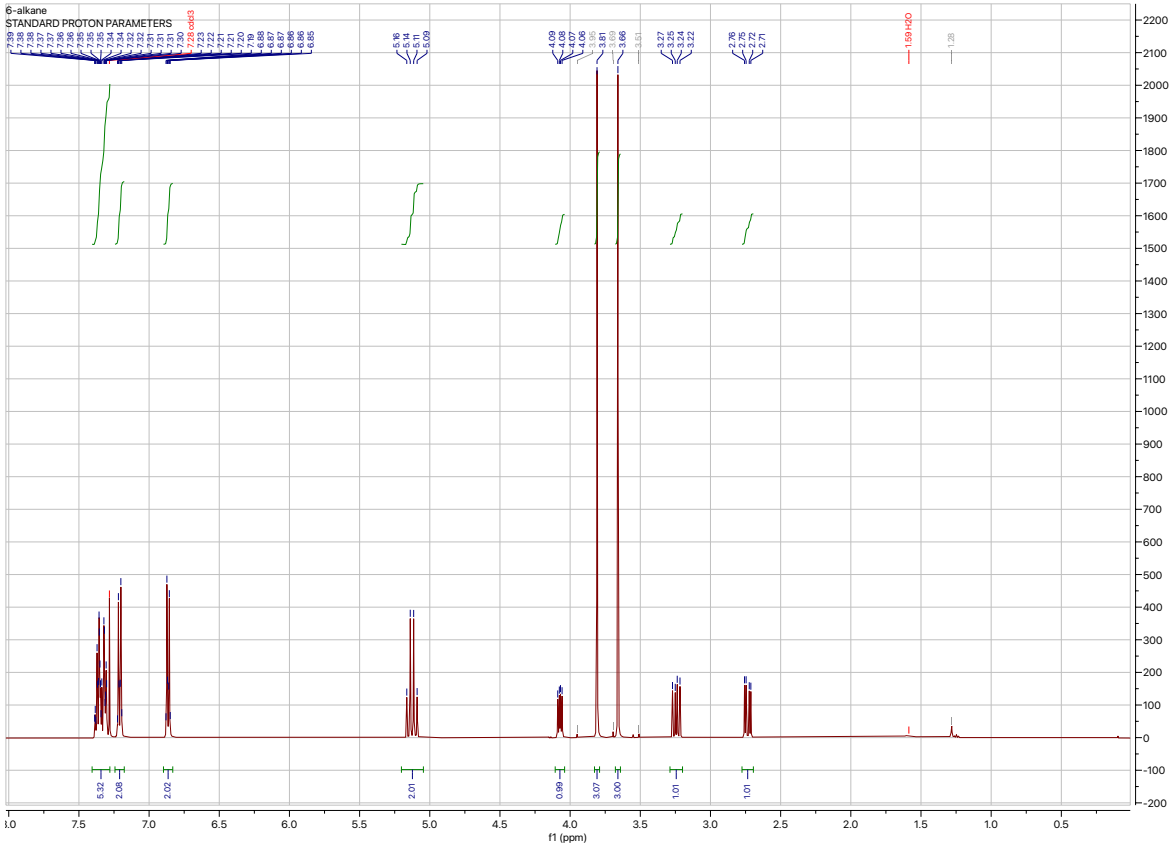
24

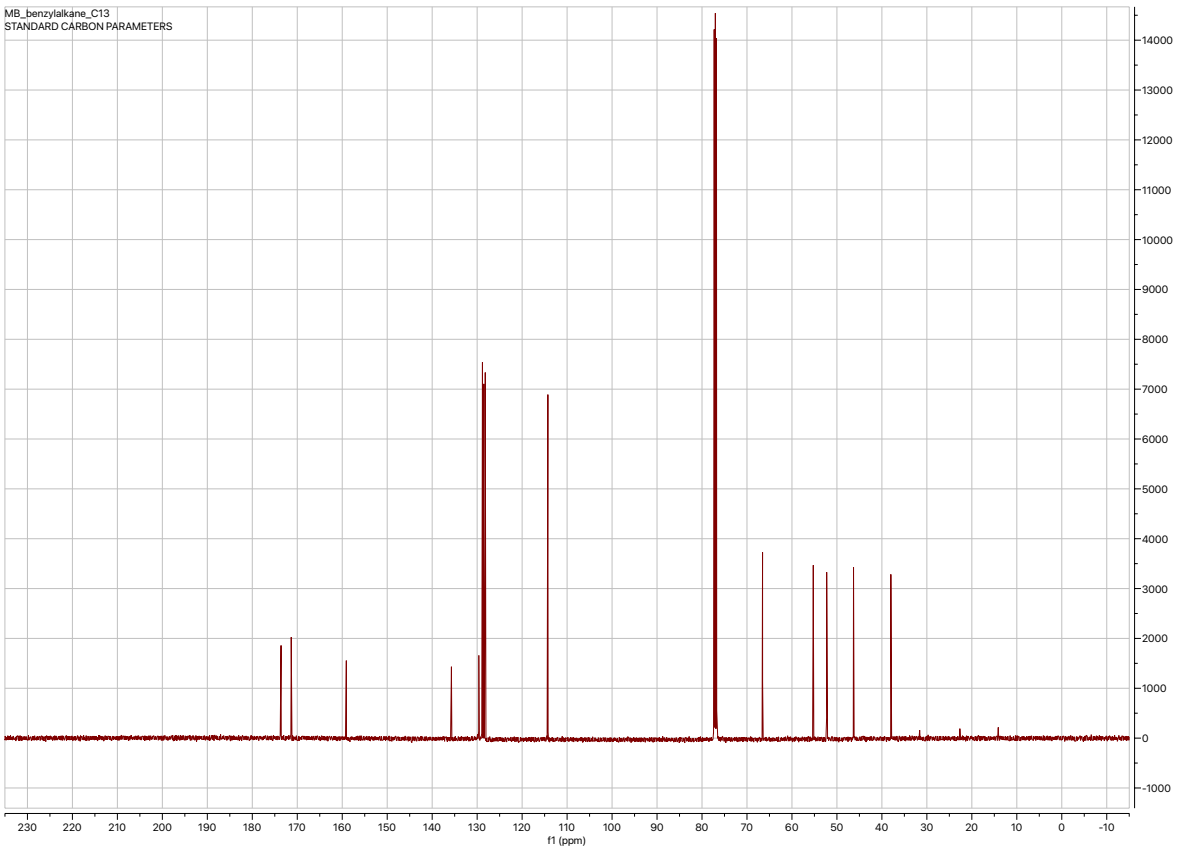


25, 4-benzyl 1-methyl 2-(4-methoxyphenyl)succinate



25




MB_benzylalkane_C13
STANDARD CARBON PARAMETERS
14000
13000
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
-1000
230 220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)