THE UNIVERSITY OF CHICAGO

ENGINEERING ARTIFICIAL METALLOENZYME FOR SELECTIVE CATALYSIS

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

HAO YANG

CHICAGO, ILLINOIS

MARCH 2016

To my parents Shuangquan Yang and Guilan Wu,

and my sister Zenmin Yang for their enduring love and support

TABLE OF CONTENTS

CHAPTER II

CHAPTER III

LIST OF TABLES

CHAPTER I

CHAPTER II

CHAPTER III

LIST OF FIGURES

CHAPTER I

CHAPTER II

CHAPTER III

LIST OF SCHEMES

CHAPTER I

ACKNOWLEDGEMENTS

I would like to thank first and foremost my advisor, Jared C. Lewis, an enthusiastic, talented, and dedicated scientist. He has provided an unparalleled environment and opportunity for me to explore the exciting and challenging field of enzyme catalysis. I also thank the members of my thesis committee, Professor Bryan Dickinson and Professor Joseph A. Piccirilli for their discussions and oversight.

This work would not have been possible without the contributions of my fellow graduate students and postdocs I have had the privilege of working with during the last five years. I am indebted to Dr. Poonam Srivastava, Dr. Zhihui Zhang, Dr. Duo-Sheng Wang, Dr. Catherine Poor, Dr. Hyun June Park, Dr. Landon Durak, Dr. Chen Zhang, Dr. James Payne, Mary Andorfer, Yifan Gu, Ken Ellis-Guardiola, Joe Gair and so on, whose expertise and assistance has made much of the work presented here possible.

The research described in the dissertation has relied heavily on the support of departmental facilities and instruments. I thank Dr. Jin Qin and Dr. Antoni Jurkiewicz for their respective assistance with mass spectrometry and NMR spectroscopy. I extend my gratitude to Dr. Elena Solomaha, director of Biophysics Core Facility at University of Chicago for assistance in protein characterization.

Finally, I thank my parents and my sister for their unconditional love and support throughout my extended education period.

ABSTRACT

Chapter I. The concept and development in the field of artificial metalloenzymes (ArMs) are introduced. A general method of bioconjugation to construct ArMs via strain-promoted azide-alkyne cycloaddition (SPAAC) is described. Scaffold proteins containing a genetically encoded p-L-azidophenylalanine and catalytically active bicyclononyne-substituted metal complexes were covalently linked through SPAAC. The bioorthogonality of SPAAC allows for bioconjugation in the presence of cysteine residues in the scaffold, so no additional scaffold modification is necessary for ArM formation. The broad scope of this method with respect to both the scaffold and cofactor components was demonstrated. Catalytic study showed that a dirhodium ArM formed with this method catalyzed decomposition of diazo compounds and both SiH and olefin insertion reactions involving these compounds, but no selectivity was observed. The simplicity and modularity of the SPAAC approach should facilitate rapid optimization of the ArMs for selective catalysis.

Chapter II. Rational engineering of ArMs toward selective catalysis is described. An alkyne-substituted dirhodium catalyst was linked to a prolyl oligopeptidase (POP) containing a genetically encoded p-L-azidophenylalanine residue to create an ArM that catalyzes olefin cyclopropanation. Scaffold mutagenesis based on a reported homology mode was then used to improve the enantioselectivity of this reaction, and cyclopropanation of a range of styrenes and donor–acceptor carbene precursors was accepted. Of all the mutations introduced, a histidine residue in the POP active site led to the largest improvements in both selectivity, conversion and activity, probably due to its capability to coordinate rhodium. The formed dirhodium-POP ArM also improved substrate specificity by reduced the formation of byproducts, including those

resulting from the reaction of dirhodium–carbene intermediates with water. This indicated control of other water-sensitive organometallics could be possible by using solvent-sequestered POP active site.

Chapter III Directed evolution efforts to improve ArM selectivity is described. A streamlined, high-throughput screening protocol for dirhodium-POP ArMs was developed. The essential step in the protocol was to scavenge excess metal cofactor without causing significant enzyme loss. Using this protocol, A POP parent mutant was submitted to iterative random mutagenesis to improve enatioselectivity in olefin cyclopropanation. Library hits giving up to 94% ee were discovered from three rounds of directed evolution. Key mutations both proximal and distal to the active site were found, which demonstrated the importance of random mutagenesis in ArM evolution. In addition, immobilization of ArMs was explored and integrated into the library screening protocol, providing an effective method to evolve ArMs for expanded scope and novel reactivity.

PREFACE

Each chapter of this dissertation is numbered independently. A given compound may have a different number in different chapters. All experimental details, references, and notes for individual chapters are included at the end of each chapter.

CHAPTER I

GENERATION OF ARTIFICIAL METALLOENZYMES THROUGH STRAIN-PROMOTED AZIDE-ALKYNE CYCLOADDITION

Most of the work described in this chapter is published (Yang et. al., *ChemBioChem*. **2014**, *15*, 223-227). I performed chemical synthesis of small molecules (including cofactors, substrates, and products) and characterization of bioconjugation and biocatalysis reactions. Dr. Poonam Srivastava contributed to scaffold engineering, protein expression, and protein purification. Dr. Chen Zhang helped with the synthesis and characterization of some cofactors.

ABSTRACT

Transition metal catalysts and enzymes possess unique and often complementary properties that have made them important tools for chemical synthesis. The potential practical benefits of catalysts that combine the best properties of both have driven the development of artificial metalloenzymes (ArMs). Strain-promoted azide–alkyne cycloaddition (SPAAC) can be used to generate artificial metalloenzymes (ArMs) from scaffold proteins containing a *p*-azido-L-phenylalanine (Az) residue and catalytically active bicyclononyne-substituted metal complexes. The high efficiency and bioorthogonality of this reaction allows rapid ArM formation when using Az residues within the scaffold protein in the presence of cysteine residues or various reactive components of cellular lysate. In general, cofactor-based ArM formation allows the use of any desired metal complex to build unique inorganic protein materials. SPAAC covalent linkage further decouples the native function of the scaffold from the installation process because it is not affected by native amino acid residues; as long as an Az residue can be incorporated, an

ArM can be generated. In this chapter, we have demonstrated the scope of this method with respect to both the scaffold and cofactor components and established that the dirhodium ArMs generated can catalyze the decomposition of diazo compounds and both SiH and olefin insertion reactions involving these carbene precursors.

INTRODUCTION

Development of efficient catalysts has been a vibrant research field of great interest to both academic and industrial scientists. Catalysis stands as essential technology for chemical and material manufacturing, pharmaceutical and food production, energy-related applications[1]. The resulting products have a direct impact on the sustainable growth of society, environment and global economy[2]. Due to this scope and importance, a variety of catalytic systems, including heterogeneous solids, organocatalysts, homogeneous metal complexes, and enzymes, have been developed for chemical synthesis[3]. Each type of these catalysts possesses unique characteristics that differentiate their potential utility for particular chemical applications. For the sake of this dissertation, two types of catalysts are briefly described and compared.

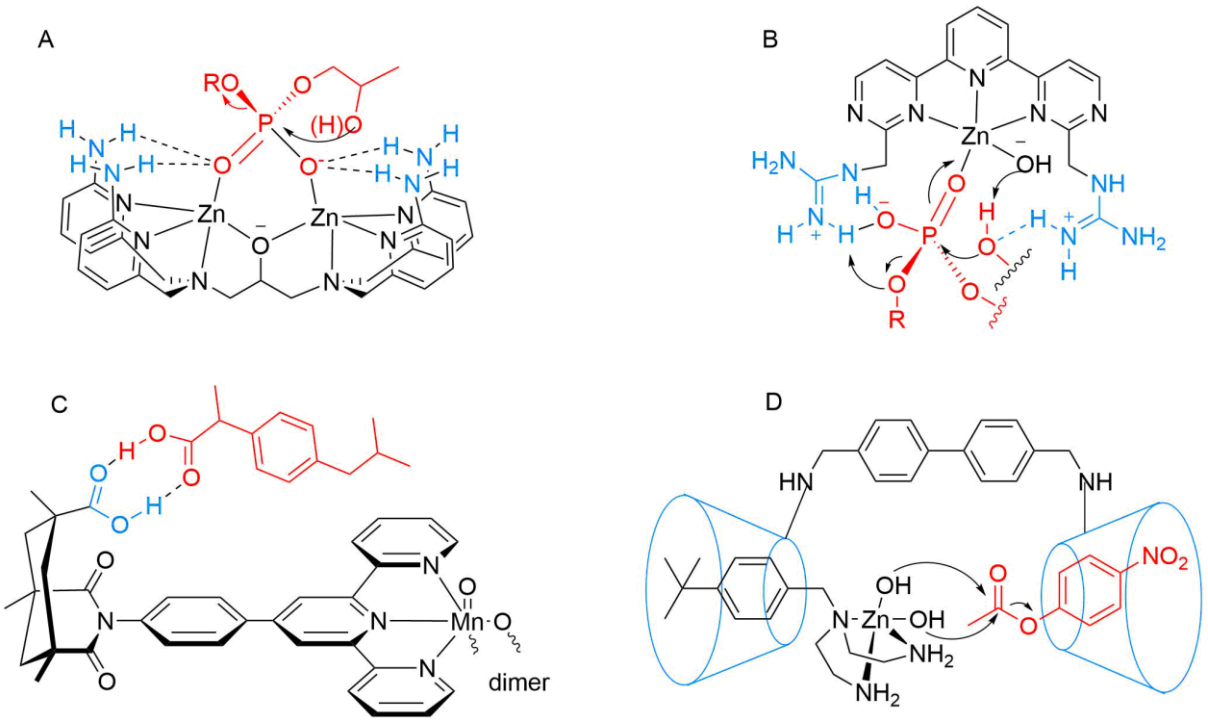
Transition metal catalysis

Homogeneous transition metal catalysts makes possible a broad range of important chemical transformations, including cross-coupling reactions[4], C−H bond functionalization[5], and olefin polymerization[6] and metathesis[7], many of which have no counterparts in nature's repertoire of chemical synthesis. Typically, a transition metal catalyst is comprised of one or more metal ion(s) bound to some number of ligands that constitute the primary coordination sphere of the metal[8]. Control of the reactivity and activity of metal catalysts can be readily achieved by rationally modifying ligand structures, and high levels of selectivity (stereo-, chemo- or regio-) have been frequently observed in organic synthesis. The engineering of such catalysts are often focused on

ligand substituents proximal to the metal center, those having direct interactions with metal center and substrates. These small molecule catalysts are generally considered to have a broad substrate scope. Despite so, such catalysts may fail to provide satisfactory solutions to some particular applications, in which for example, a novel selectivity overriding the selectivity defined by the substrate structure is pursued, or differentiating the same type of reactive sites with similar chemical environments is required. This challenge leads to an increased research interest[9-12] in modifying ligand substituents distal to metal centers, which comprise the secondary coordination sphere of the metal. Noncovalent interactions such as hydrogen bonding, hydrophobic, electrostatic, and steric interactions have been invoked to confer great impacts on catalysis outcomes[13-15]. These interactions tend to be weaker, less distance dependent and directional, and more affected by entropy. By operating in concert, multiple weak interactions derived from secondary coordination sphere may render high selectivity, reminiscent of enzyme-type control. To realize such cooperative weak interactions, functional groups including amine, guanidines and carboxyl groups have been employed to simulate the roles of secondary coordination sphere[13] (Scheme 1.1). Design and synthesis of these types of catalysts is already a daunting task, not to mention altering such structures to fine tune catalyst−substrate interactions. In this respect, both small molecule ligands and existing supramolecular hosts offer limited flexibility and control over the orientations of distal substituents in both a static and dynamic sense. Given such limitations, the need persists for transition metal catalysts with well-defined secondary coordination environments that can be readily and extensively fine-tuned for particular applications.

Scheme 1.1. Representative examples of engineering secondary coordination sphere of transition metal complexes. A) Phosphate hydrolysis by a dinuclear zinc(II) complexes with amino groups activating substrates[13c]; B) Accelerated cleavage of an RNA dinucleotide adenylyl phosphoadenine by a zinc(II) complex with guanidinium groups interacting with substrates[13d]; C) Selective oxidation of saturated hydrocarbons by a dimanganese complex with carboxylic groups

as substrate anchoring sites[13e]; D) Accelerated ester hydrolysis by a zinc(II) complexes encapsulated by two β-cyclodextrin molecules[13f]



Enzyme catalysis

Enzymes are nature's powerful catalyst to conduct all the biochemical transformations essential to life. Many natural enzymes[16] are well known for the extreme levels of rate acceleration and exquisite selectivity on their native substrates (although this is not universal).[17] Unlike small molecule catalysts, enzymes exercise selectivity/activity control through their huge three dimensional framework around the active site which is defined yet dynamically fluxional[18]. Natural evolutionary processes[19] have given rise to such an exquisite control that amino acid residues, cofactors and substrates are precisely oriented by the entire enzyme structure to modulate reaction energy surfaces[20] through synergistic operation of multiple weak interactions. To accelerate discovery of new enzymes, researchers have mimicked the natural evolution in the laboratory using iterative rounds of catalyst diversification and functional screening or selection, namely directed evolution, to engineer enzymes with improved efficiency.[21] However, whereas

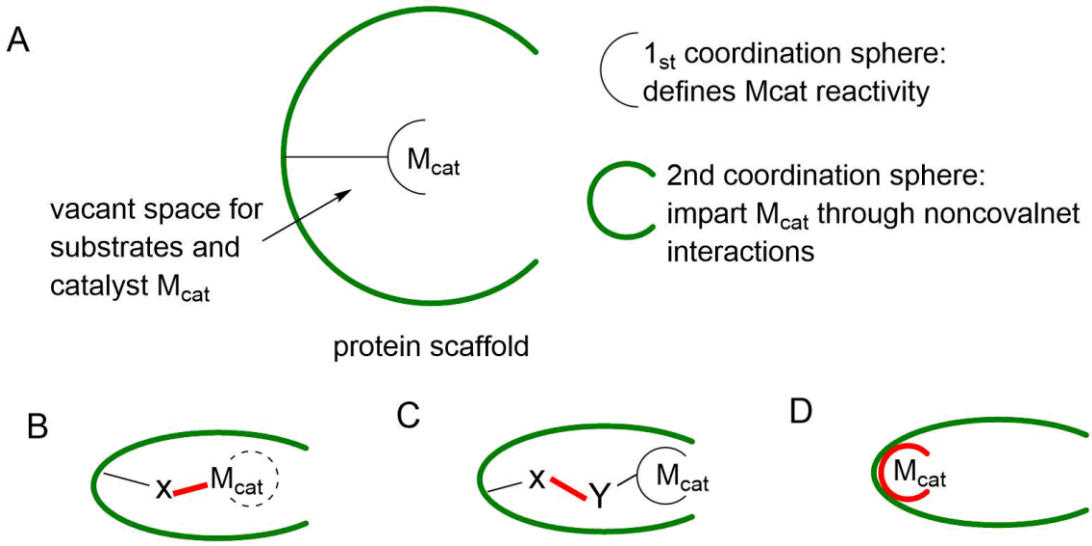
laboratory evolution has been used to create enzymes from noncatalytic scaffolds,[22] the most effective applications typically involves activities that are already present.[21] Chemists are also interested in many reactions with no biocatalytic counterparts in nature, particularly those catalyzed by abiotic metals. The lack of such chemical transformations in nature is probably because nature does not has a need for them or these reactions require reagents hardly available in nature[23]. Thus enabling reactions catalyzed by abiotic metals is likely not possible without introduction of corresponding metals.

Hybrid catalyst: metalloenzyme

Metalloenzyme, a hybrid catalyst comprised of a protein scaffold and metal cofactors, is not a new concept to nature. Actually, to diversify the possible reactivity scope, one-quarter to one-third (estimated) of all proteins require metals to carry out their functions[24]. These naturally-occurring metalloenzymes are responsible for catalyzing many important biological processes, such as photosynthesis, respiration, water oxidation, molecular oxygen reduction and nitrogen fixation[24]. In hopes of combining the selectivity control and evolvability of enzymes with the reactivity of metal catalysts[25], so that novel reactivity/selectivity not possible with either metal complex or enzyme alone may be discovered, researchers have invested intensive efforts to incorporate non-natural metal cofactors into protein scaffolds to create artificial metalloenzymes (ArM). We expect the large sizes, defined 3-D shapes, and dynamic properties of these systems allow them to manipulate molecules and reactivity in ways that conventional catalysts cannot. Because of this, they are well-suited to tackle complex selectivity problems, ranging from in vitro functionalization of biologically active molecules, to in vivo catalytic manipulation of cellular function and metabolic engineering.

The potential practical benefits of combing metal and protein catalysts and a desire to understand the structure-reactivity relationship within hybrid catalysts have driven researchers to create ArMs since the 1970s (Scheme 1.2)[24-26]. Central to these efforts are robust bioconjugation methods that incorporate transition metals into protein scaffolds. In terms of the nature of the scaffold-metal interaction, bioconjugation methods include dative binding (coordination of metal atoms by scaffold residues), covalent scaffold modification using functionalized catalysts, or supramolecular binding (catalyst binding through specific protein-ligand interactions) (Scheme 1.2, B−D)[26]. Each method has its advantages and disadvantages that make it more or less suitable for particular applications. Using one or combination of these methods, a diverse collection of ArM systems have been developed, in which secondary coordination sphere effects[27] impart selectivity to metal catalysts[28], accelerate chemical reactions,[29] and are systematically optimized via directed evolution.[30] Despite much work will be required to make such systems practically useful in organic synthesis, these capabilities have the potential to impact chemical synthesis in ways not readily achieved using small molecule catalysts.

Scheme 1.2. A) General ArM structure and bioconjugation through dative (B), covalent (C) and supramolecular (D) interactions



RESULTS

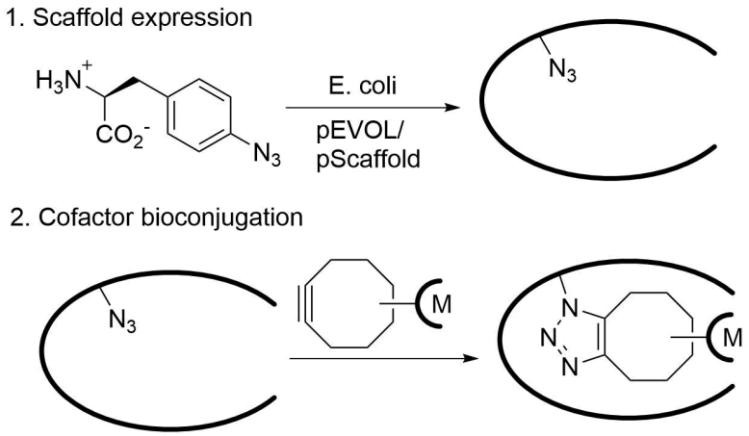
As described above, dative[31], supramolecular[32], or covalent[33] approaches have been developed to link metal cofactor and protein scaffold, each with its own advantage and limitation. Several factors came into consideration when thinking of a good ArM anchoring approach, such as the availability of protein scaffolds[28], the efficiency and selectivity of bioconjugation methods[33], and the ease of synthesizing or modifying metal cofactors.[34] The catalytic activity of an ArM is often independent of the original function of the scaffold protein, leading to research focused on other properties that improve ArM utility, including expression level, thermostability, and organic solvent tolerance[35]. Considering the need to screen different scaffolds or cofactors to find a good starting point for metalloenzyme development[28], and to conduct protein engineering for high selectivity[36], a general ArM construction platform that in principle works regardless what scaffold or metal complex is chosen would be ideal.

Among these methods, dative linkage or supramolecular binding requires strong and specific molecular recognitions which are present in very few proteins[31, 32]. Also, coordinating metals with proteins is typically limited to ligand sets composed of the 20 natural amino acids. Encoding unnatural amino acids into proteins can somehow expand the ligand scope, but this process itself requires extensive engineering for each desired amino acid and still does not work for metal complexes with non-natural ligands (for example, phosphine or carbene ligand). Comparatively, bioconjugation through covalent linkage places the fewest limitations on the nature of the metal catalyst and scaffold protein used for ArM formation[33]. The most common covalent linkage involves nucleophilic attack of a unique cysteine residue present in the protein scaffold on cofactors with an electrophilic linker (for example, maleimide or iodoacetamide). The

major drawback is that the reactive residue used for covalent linkage should be uniquely existent in the scaffold to achieve site-selective modification[33], which requires extra engineering efforts in some proteins. A covalent linkage orthogonal to all the present residues in protein scaffold will be ideal.

Such design requirements led us to consider the toolbox of bioorthogonal chemistry[37]. By introducing complementary functional handles inert to biological moieties, bioorthogonal chemistry enables probing biomolecules (glycans, lipids, and metabolites) *in vivo* through highly selective reactions, without interaction or interference with biological systems. Among different reactions developed, the strain-promoted azide-alkyne cycloaddition (SPAAC) developed by Bertozzi and coworkers[37] is of great interest to serve as a platform for metalloenzyme formation (Scheme 1.3).

Scheme 1.3. ArM preparation through SPAAC



SPAAC-based protein modification involves expressing a target protein with a genetically encoded azide- or alkyne-bearing amino acid[38] and reacting this protein with an alkyne or azide-bearing reagent[39]. Due to the high reaction rate, little interference with biomolecules, and site-specific modification, SPAAC has been widely applied in chemical biology[40]. Yet no example of labeling a protein with a metal catalyst through SPAAC reaction has been shown. In general, we

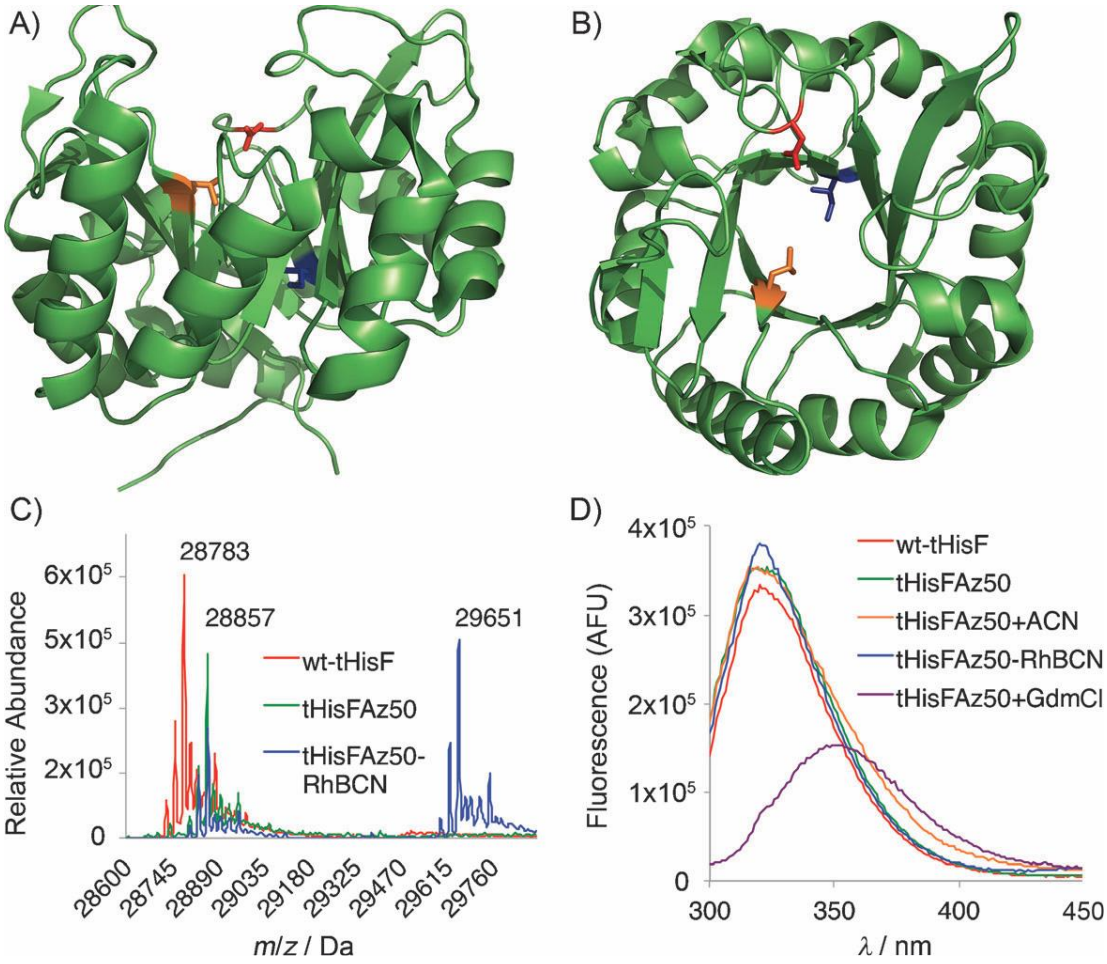
desired this SPAAC reaction to occur on the interior of the protein (Scheme 1.3) rather than its surface[41], so that the surrounding protein environment may have significant impacts on the metal center[42]. This is not a typical requirement for the chemical biology applications employing SPAAC (usually surface modifications are used) and demands robust bioconjugation methodology. We envisioned that the central pore of an α, β-barrel protein could provide a suitable environment for azide or alkyne incorporation[43]. Because SPAAC cofactors would be unreactive toward any native amino acid residue, any desired scaffold could in principle be exploited for ArM formation.

The choice of protein scaffold is essential to success of metalloenzyme development. Reetz and coworkers showed that a thermostable α, β-barrel protein tHisF from *Thermotoga maritimacan* can be a robust host for a variety of chelating ligands commonly used in transition metal catalysis and even a Pd-complex, though no examples of ArM catalysis was described[44]. This monomeric protein with a molecular mass of 27.7 KDa, constitutes the synthase subunit of the glutaminase–synthase bi-enzyme complex which catalyzes the formation of imidazole glycerol phosphate in histidine biosynthesis[45a, b]; its $(\beta\alpha)_8$-barrel structure is the most common enzyme fold found in nature[45c, d]. In *E. coli.*, tHisF is expressed in large amounts; its remarkable thermostability allows using heat treatment to parallelize purification in libraries, a valuable feature for further protein engineering[44].

Because the bottom of this protein is blocked by a salt bridge, we hoped that azide or alkyne mutations introduced deep within the 25Å-long central pore of the protein should project SPAAC cofactors up into the pore and place the metal complexes within reach of several loops near the pore opening (Figure 1.1, A and B).[45a, c] Extensive biochemical characterization of this protein has revealed that its fold is highly stable and that it possesses a tryptophan and four tyrosine residues

that enable spectroscopic characterization of its folding; both features make it an ideal test substrate for bioconjugation method development[46].

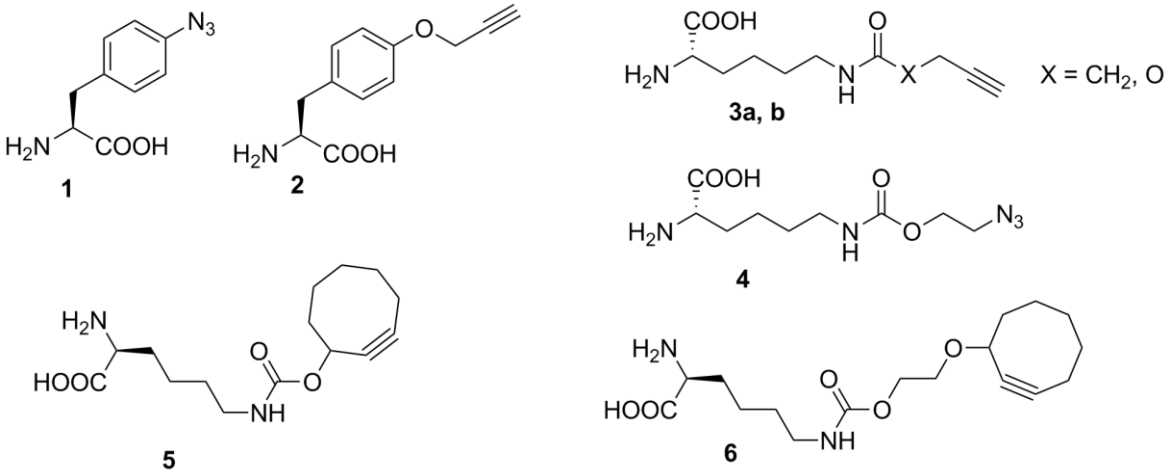
Figure 1.1. A), B) Structure of wt-tHisF (PDB ID: 1THF); colored residues are in positions 199 (blue), 50 (orange), and 176 (red). C) HR-ESI-MS of wt-tHisF, tHisF-Az50, and tHisFAz50-RhBCN. D) Fluorescence spectra (290 nm) of wt-tHisF, tHisF-Az50 (in buffer, 60% CH3CN, 6m guanidinium chloride [GdmCl]), and tHisF-Az50-RhBCN (CAN, CH3CN, GdmCl).



To install an unnatural amino acid as a bioorthogonal reaction handle for SPAAC chemistry, we implemented amber stop codon suppression[38] developed by Schultz and his coworkers. This method enables genetic encoding of unnatural amino acids with diverse physical, chemical or biological properties into proteins with high fidelity and efficiency, by using an orthogonal biosynthetic machinery composed of a unique codon (for example, the amber codon TAG) and

corresponding tRNA-synthetase pair. To date, diverse alkyne- or azide-bearing amino acids have been successfully encoded into proteins within *E. coli*, yeast and mammalian cells[39] (Figure 1.2). Considering that incorporation of an azide-containing amino acid (for example, compound **1** in Figure 1.2) is more efficient than incorporation of an cyclooctyne-containing amino acid (for example, compound **5** or **6** in Figure 1.2) and may cause probably smaller perturbation to scaffold structure, we decided to incorporate the azide into a protein scaffold and install cyclooctyne into a metal cofactor.

Figure 1.2. A list of azide- or alkyne-containing amino acids that have been encoded into proteins through the stop codon suppression method



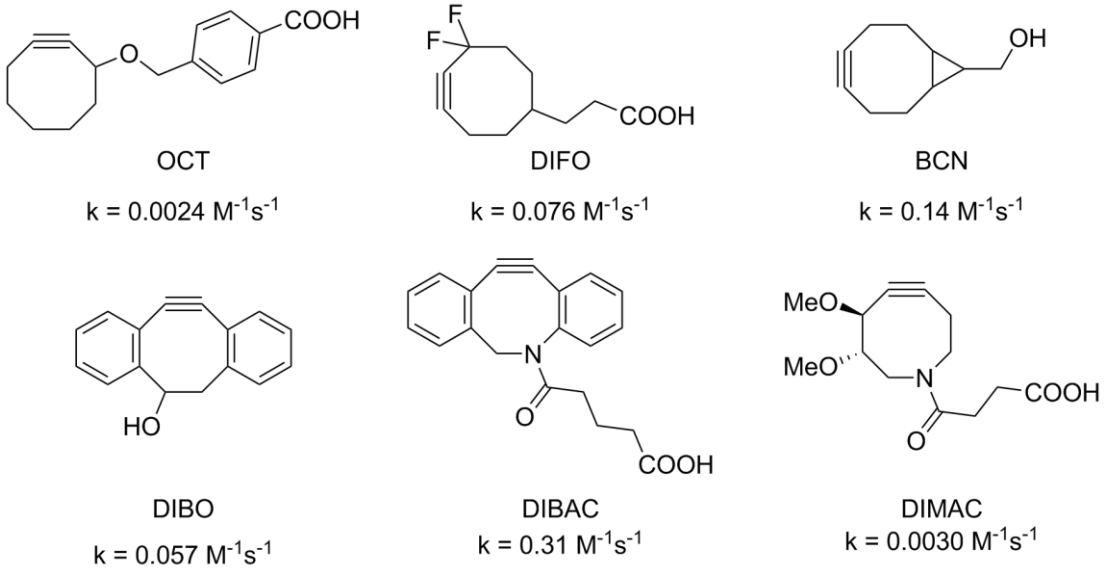
Incorporation of p-azido-l-phenylalanine (Az) was conducted by Dr. Poonam Srivastava at representative positions throughout the central pore of tHisF, including top (residue Ala 176), middle (residue Ala 50), and bottom (residue Ala 199)[45]. We observed high levels of scaffold expression and unnatural amino acid incorporation with no apparent azide photolysis based on high resolution ESI mass spectrometry (Figure 1.1, C), despite Ala 50 and Ala 199 being located on the protein interior[47]. The His6-tagged scaffold proteins were purified by Ni-affinity chromatography following an initial heat treatment[44].

In order to study potential structural perturbation in azide-containing tHisF due to UAA incorporation, fluorescence and CD spectroscopy were conducted on the azide mutants. Because tHisF contains a single tryptophan residue (W156) located in α-helix 5 and four tyrosine residues (Y39, Y143, Y182, and Y240) distributed across the protein, the fluorescence signal of these aromatic residues have been used as an indicator for global protein folding (tertiary structure)[46]. And because tHisF has a $(\beta\alpha)_8$-barrel structure, circular dichroism signal can be used to monitor the loss of secondary structures. Fluorescence spectra for tHisFAz50 mutant and wild-type tHisF in buffer were measured and compared (Figure 1.1, D), and no change was observed. Furthermore, the organic solvent tolerance was studied by measuring the fluorescence signal of tHisFAz50 mutant sample which were submitted to incubation in 9:1 (v/v) organic solvent/buffer mixture for 1 hour. While the presence of methanol or DMF significantly reduces the fluorescence signal of tHisFAz50, acetonitrile shows a milder influence (Figure S1.6). In 60% acetonitrile, no change in the fluorescence spectrum was observed (Figure 1.1, D), which highlights the organic solvent tolerance of this scaffold protein[35]. CD spectrum for tHisFAz50 and tHisF were measured (Figure S1.5), and the high similarity between the two proteins indicates UAA incorporation has no impact on secondary structure of tHisF.

A similar approach was used by Dr. Poonam Srivastava to express variants of a thermostable phytase from *Bacillus amyloliquefaciens*[48] with Az incorporated at residue 104. This enzyme has an overall cylindrical shape, built from six sheets of four to five anti-parallel-β-strands arranged around a central pore. The position of the Az residue was approximately 20 Å into this pore, so point mutations N99A, N100A, and D102A were introduced to facilitate access of bicyclo[6.1.0]nonyne (BCN) to the Az residue.

We next developed a modular approach to synthesize alkyne-substituted cofactors. A number of structurally varied cyclooctyne derivatives (e.g. DIFO, BCN, DIBAC, DIBO, ADIBO) have been developed that strongly differ in terms of reaction kinetics and hydrophilicity (Figure 1.3)[37]. Among them, we chose bicyclononyne (BCN) developed by van Delft and coworkers[49] as the linker between metal cofactor and protein scaffold, considering its symmetric and small structure, fast reaction rates, and relative ease of synthesis[37]. In van Delft's work, a para-nitrophenyl-carbonate(PNP)-substituted BCN[49] (carbonate **1**, see Scheme 1.4) was used as an electrophilic intermediate to link molecules of interest, such as biotin or fluorescent probes. We envisioned that this intermediate can be recapitulated to anchor metal complexes with appropriate nucleophile groups.

Figure 1.3. Cyclooctynes for Cu-free click chemistry in living systems and second order rate constants for their reaction with benzyl azide in acetonitrile or methanol[37]



Considering ArM catalysis will be mainly conducted in aqueous buffered conditions, only metal complexes that facilitate catalytic reactions with high efficiency in the presence of water and air can be good candidates for ArM incorporation. To date, existing ArM systems have focused on reactions including transfer hydrogenation, Diels-Alder reaction, epoxidation and so on[32b, 50]. We

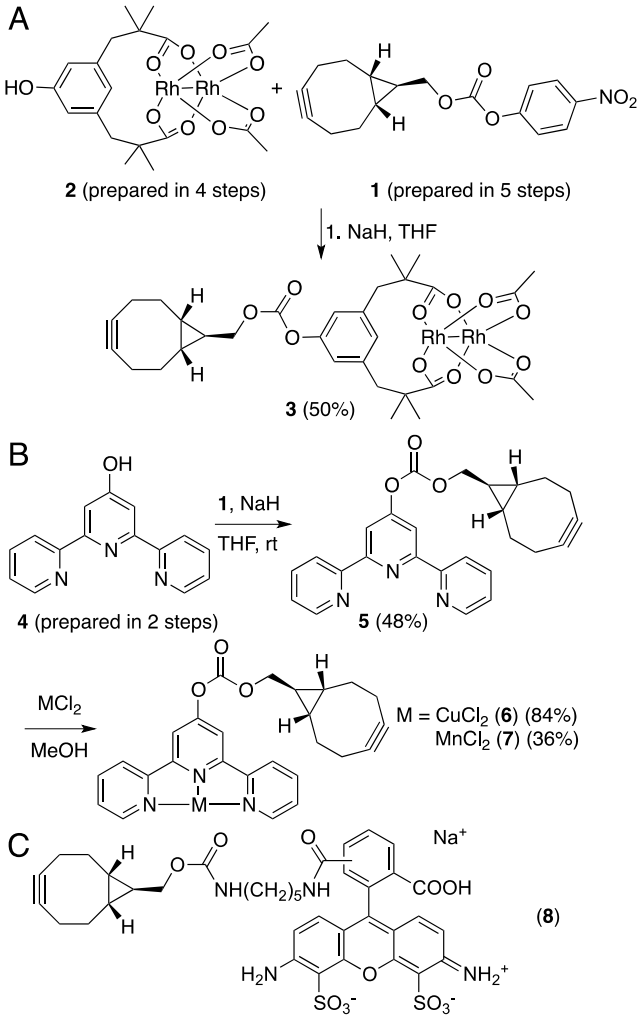
initially targeted dirhodium tetracarboxylate cofactors, due to the high activity of these complexes toward a range of carbene insertion reactions[51] and their stability in both air and water[52].

In general, rhodium-mediated carbene/nitrene insertion reactions (including cyclopropanation, C-H insertion, N-H insertion, etc.) employing diazo substrates are performed in anhydrous environment due to the competing O-H insertion reactions. However, several examples of dirhodium-catalyzed carbene insertion in aqueous environments indicate the potential of dirhodium complexes in metalloenzyme catalysis. Charette and coworkers[52b] demonstrated a cyclopropanation reaction involving ethyl diazoacetate and olefins proceeds with high efficiency in aqueous media using dirhodium(II) carboxylates, and a combination of hydrophobic catalyst and hydrophobic substrate provide the best efficiency, presumably by forming small catalyst/substrate beads or micelles. Afonso and coworkers[52e] reported preferential Rh(II) carbenoid intramolecular C-H versus O-H insertion derived from alpha-diazo-acetamides can be achieved in water by using an appropriate combination of the catalyst and amide groups, which forms a larger hydrophobic environment around the reactive carbenoid center; also, the regioselectivity of the C-H insertion depends on the structure of the catalyst and the hydrophobic nature of the amide substituents. The intermolecular version of C-H insertion in water was realized by Francis and coworkers[52c] in their study of chemoselective tryptophan labeling of peptides/proteins with rhodium carbenoids. However, the labeling method requires 1 equivalent of $Rh_2(OAc)_4$ catalyst and 100 equivalents of vinyl diazo substrate under optimized conditions, which limits its application in practical synthesis and implies the O-H insertion is a strong competitive side reaction. These examples show that dirhodium catalysis in aqueous media is possible. We also hypothesized that creating a hydrophobic microenvironment around the metal-

carbenoid center within the scaffold protein, which mimics the micelle effects in Charette's and Afonso's work, could suppress the competing O-H insertion reaction.

Inspired by the improvements in dirhodium catalysis shown by Du Bois and co-workers using tetramethyl *m*-benzenedipropionic acid ligands (esp)[53], we prepared hydroxy-esp derivative **2** and reacted this compound with $Rh_2(TFA)_2(OAc)_2$[54] to form the mixed esp/diacetate complex **3**[55] (Scheme 1.4). To demonstrate the versatility of SPAAC bioconjugation in terms of cofactor choice, two additional BCN cofactors, **6** and **7**, containing Cu[56] and Mn[57] terpyridine complexes, were prepared by metallating BCN-terpyridine **5**, which was formed from phenol **4** and carbonate **1** (Scheme 1.4, B). Similar metal–terpyridine complexes are known to catalyze a range of CH insertion reactions[58]. This metallation approach complements the convergent approach used to prepare **3** and provides additional flexibility for BCN cofactor formation to accommodate the unique reactivity of different metal complexes. Finally, fluorescent probe **8** was prepared in analogy to the approach developed by van Delft[49] (Scheme 1.4, C). The carbonate linkage in all of these cofactors was not hydrolyzed, even after extended room temperature incubation in various aqueous buffers (e.g., CH_3CN/Tris or THF/KP_i, pH 7.5), based on HPLC analysis.

Scheme 1.4. Syntheses of cofactor **3**, **6**, **7**; structure of probe **8** (the intermediate **5** was synthesized by Dr. Chen Zhang).



Bioconjugation between the metal cofactor and tHisF azide mutant was studied next. To find optimized bioconjugation conditions, a range of reaction parameters (temperature, CH_3CN%, and cofactor amounts) were explored. It was observed that 5 equivalents of cofactor, 20% (v/v) CH_3CN/Tris buffer and conducting the reaction at 4 ^oC provided the optimal ArM conversion. 20% CH_3CN was the highest co-solvent ratio that could be used without causing protein degradation. Although lower temperatures decreased the bioconjugation rate, the overall conversion was higher due to decreased ArM denaturation and precipitation over the course of the reaction, so these conditions were utilized for a preparative scale bioconjugation reaction for ArM isolation and

characterization. ArM formation was monitored by MALDI mass spectrometry, and cofactor consumption was followed by protein HPLC (due to the high level of structural similarity, the scaffold and ArM could not be resolved on HPLC). This analysis revealed a depth-dependent rate of bioconjugation and final conversions ranging from 50 % for Az199 (bottom) to 80 % for Az176 (top; Table 1.1).

Table 1.1. Mass spectrometry and conversion data for ArMs

Scaffold (MW)[a]	Cofactor (MW)	MW_{ArM}[a]	MW_{Obs}[b]	Conv.(%)[b]
tHisF-Az50 (28859)	**3** (792)	29651	29614	70
tHisF-Az176 (28857)	**3** (792)	29649	29630	80
tHisF-Az199 (28857)	**3** (792)	29649	29620	50
tHisF-Az176 (28857)	**6** (560)	29417	29435	90
tHisF-Az176 (28857)	**7** (550)	29407	29392	80
phytase-Az104 (40040)	**3** (792)	40832	40846	90

[a]Calculated protein MW using tools at www.expasy.org. [b]Observed MW and approximate conversion from MALDI mass spectrometry.

After maximum conversion was observed by MALDI mass spectrometry, preparative HPLC was used to remove all traces of unreacted cofactor from the inseparable ArM/scaffold mixture, the purity of which was validated by analytic HPLC. Extensive efforts to purify ArM/scaffold mixture through a range of protein purification techniques (size exclusion chromatography, ion exchange chromatography, Ni-NTA affinity chromatography, etc al.) failed to isolate protein from unreacted cofactor, probably due to the lipophilicity of the dirhodium cofactor and corresponding strong non-specific, hydrophobic association with protein scaffold. It is worth noting that the HPLC method we use includes a linear gradient from 20% CH_3CN to 80%

CH3CN, during which the purified ArM/scaffold mixture elutes at around 60% CH_3CN, with around 25% mass recovery in average. To test if the HPLC condition affects protein folding, fluorescence and CD spectroscopy of the eluted protein sample (Figure 1.1 D and S1.5) after buffer exchange to Tris buffer were measured. No noticeable difference from a sample of pure scaffold was observed, which indicated that the ArM maintained its secondary and tertiary folding or refolded as the solvent environment was restored to aqueous conditions. While the ArM could not be separated from the scaffold, this has no impact on catalysis, as only the ArM contains the cofactor required for catalytic activity and the scaffold showed zero catalytic activity in control experiments. Last, the isolated ArM/scaffold mixture was further characterized by ESI mass spectrometry to confirm the composition of the ArM (Figure 1.1, C).

To demonstrate the generality of our SPAAC ArM approach, we also covalently linked cofactor **3** to sites 176 and 199 of tHisF (Figure 1.1) and to Az104 in the central pore of the engineered phytase scaffold. Cofactors **6** and **7** were then linked to tHisF-Az176. In all cases, formation of the desired ArM was confirmed by MALDI mass spectrometry following purification (Table 1.1). Together, these examples are the first in which the SPAAC reaction has been used to link metal catalysts to proteins. The mild conditions required for this reaction are ideally suited for transition metal cofactors, which might react under other bioconjugation conditions. Also, unlike most SPAAC reactions[40], which are typically conducted on the surface of proteins, these reactions are conducted within the barrel of the protein, illustrating the high efficiency of this reaction, even when the azide is not exposed to bulk solvent.

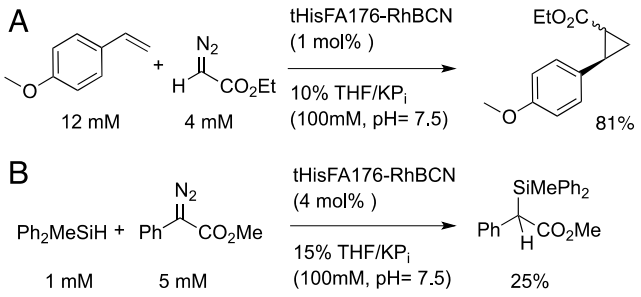
We next evaluated the catalytic activity of purified tHisF-RhBCN ArMs toward a number of dirhodium-catalyzed chemical reactions[51]. Small-molecule dirhodium carboxylates have been widely used in intra- and intermolecular X-H insertion reactions (X = C, N, O, Si, S, etc.), and

moderate to high stereoselectivity are observed. We are most interested in examining the application of tHisF-RhBCN ArMs in insertion into C-H bonds due to the ubiquitous existence of C-H bonds in organic molecules and the potential to streamline chemical synthesis with selective C-H functionalization methodology. However, initial substrate screening with both cofactor **2** (from Scheme 1.4) and tHisF-RhBCN revealed typical substrates used in carbenoid-mediated C-H insertion did not produce any observable formation of desired insertion products, and the decomposition of diazo precursor led to mainly hydrolysis products. While further protein engineering might enable C-H insertion, we decided to try other reactions with reasonable conversions as a starting point.

Davies and coworkers have shown that aryldiazoacetates provide remarkable chemoselectivity toward different types of substrates[51]. In their competition experiment of various substrates reacting with methyl phenyldiazoacetate in the presence of Rh_2(S-DOSP)$_4$, insertion into diallylic C-H bond, Si-H insertion and cyclopropanation of styrene are much more favorable than reaction with other substrates[51]. Hence, we tested these three reaction types with our ArM catalysts. Surprisingly, while Si-H insertion and cyclopropanation gave reasonable conversions in the presence of ArMs, no insertion product was observed when reacting the diallylic substrate with aryldiazoacetates (the diallylic substrate was not consumed). Using tHisF-Az50, -Az176, and -Az199 as catalysts in both intermolecular cyclopropanation and Si-H insertion (Scheme 1.5), a correlation between decreasing conversion and lower cofactor linkage site was observed (Table S1.3). This could be rationalized by the assumption that deeper location of the cofactor within scaffold hinders substrate access. Unfortunately, the ArMs provided lower conversions than cofactor alone (**2**-OAc provided 99 % and 80 % yields for cyclopropanation and Si-H insertion,

respectively), negligible enantioselectivity, and still significant amounts of diazo insertion into the water OH bond[52].

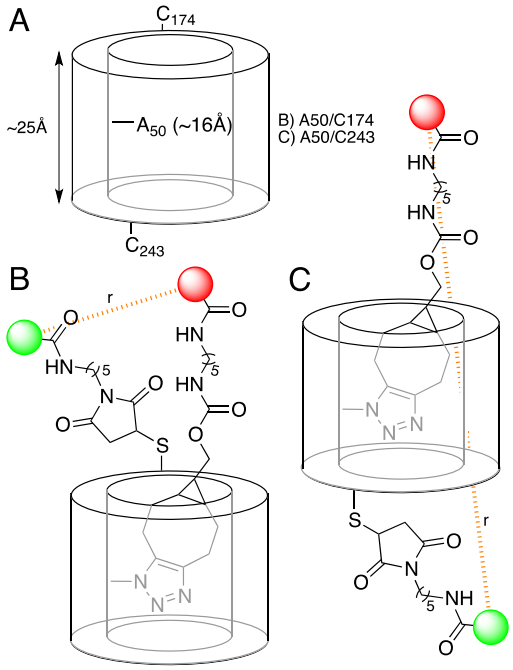
Scheme 1.5. tHisF-RhBCN-catalyzed A) cyclopropanation and B) Si-H insertion



The lack of enantioselectivity in our ArM-catalyzed reactions in not surprising. Although a number of selective ArM-catalyzed reactions have been reported, achieving such selectivity still remains challenging[36], and improving metal–protein interactions through rational design or directed evolution[28,59] is generally be required. Without detailed structure information (such as X-ray crystallography), the lack of selectivity could be ascribed to many possible reasons. We were particularly concerned about the possibility that the rhodium cofactor might assume a wrong orientation and project outside from the protein cavity[42]. As previously mentioned in the chapter, we assumed that the bottom of tHisF protein is blocked by a salt bridge, so that the azide amino acid and the corresponding dirhodium cofactor introduced deep within the 25 Å-long central pore of the protein should project SPAAC cofactors up into the pore and place the metal complexes within reach of amino acid residues around. If our assumption proved wrong, the rhodium cofactor might point into the solution and lead to non-selective reactions. Establishing cofactor orientation is difficult as the site of bioconjugation is distal to the metal catalyst, but could provide insight into the poor selectivity observed for the RhBCN ArMs. Although X-ray crystallography of the formed ArM is still underway, we reasoned that surrogate fluorescent probe **8** (Scheme 1.4) could also provide quantitative information about cofactor orientation.

Specifically, to establish whether cofactors linked to the central pore of tHisF protrude up through the central pore or down through the bottom of the protein (Figure 1.4), we used a dual-label FRET approach with surrogate fluorescent probe **8**.[60] Cysteine point mutations were introduced by Dr. Poonam Srivastava at the top (D174C, Figure 1.4, B) and bottom (D243C, Figure 1.4, C) of the tHisF-Az50 scaffold exterior. These proteins were reacted with a commercially available Alexa Fluor 594 maleimide probe. A BCN-conjugated Alexa Fluor 488 probe **8** was prepared[49] and reacted with the tHisF-Alexa 488 conjugates. Energy transfer from the Alexa Fluor 488 donor, excited at 495 nm, to the Alexa 594 acceptor was then measured by using both fluorescence intensity and lifetime methods and used to calculate approximate distances between the pore-linked (Az50) donor dye and the exterior-linked (Cys 174 and Cys 243) acceptor dyes (Table S1.2).[61] Both intensity and lifetime measurements provided similar results, consistent with the relative positions and linker lengths used, and both indicated that that the pore-linked dye resided substantially closer to the top of the tHisF than to the bottom[60].

Figure 1.4. A) Cartoon schematic of tHisF. B) "Top" (A50/C174) and C) "bottom" (A50/C343) dual-labeled constructs for FRET analysis



Given the identical BCN linkage in probe **8** and cofactors **3**, **6**, and **7**, these data provide good evidence for BCN cofactor projection up through the top of the scaffold as intended. We believe the effective length of **3** upon bioconjugation (ca. 20 Å from the α-carbon of the Az residue to either Rh atom), places the metal complex near the mouth of the α,β-barrel (see Figure S1.10). The lower yields of reactions catalyzed by the dirhodium ArMs, relative to those catalyzed by **2**-OAc, suggest that repulsive ArM–substrate interactions may occur during catalysis; but such interactions as well as a potentially correct cofactor projection indicated by FRET experiment are insufficient to impart selectivity in the reactions studied. Modified cofactor designs, use of alternate scaffolds, and protein engineering may provide a solution to this problem.

CONCLUSION

We have established that the SPAAC reaction[37] can be used to generate ArMs from scaffold proteins containing a genetically encoded Az residue and catalytically active BCN-linked cofactors. The high efficiency of this reaction allows for rapid ArM formation even when the Az residue is located within, rather than on the surface, of the scaffold protein, which enables the possibility of engineering the scaffold to tune the secondary coordination sphere[42] of the metal catalyst. The bioorthogonality of SPAAC allows for bioconjugation in the presence of cysteine residues[62] in the scaffold, so no additional scaffold modification is necessary for ArM formation. Although these properties are widely exploited for various applications in chemical biology, they have not yet been employed for ArM formation and thus provide a number of opportunities for catalysis. We have demonstrated the scope of this method with respect to both the scaffold (tHisF[45] and phytase[48]) and cofactor (Rh_2-tetraacetate[51] and Mn- and Cu-terpyridine[58] complexes) components. We also established that dirhodium ArMs can catalyze the decomposition of diazo compounds and both SiH and olefin insertion reactions involving these compounds[51]. The

simplicity and modularity of our SPAAC approach should facilitate rapid optimization of the ArMs reported herein for selective catalysis, work that is currently underway in our laboratory.

ACKNOWLEDGEMENTS

We thank Prof. Peter G. Schultz for providing pEVOL-pAzF and Prof. Dr. Reinhard Sterner for providing pET11c tHisF. We also thank Dr. Justin Jureller at the UC Institute for Biophysical Dynamics NanoBiology Facility, Dr. Jin Qin at the UC Department of Chemistry Mass Spectroscopy Facilities, and Dr. Elena Solomaha at the UC Biophysics Core Facility for assistance.

EXPERIMENTAL

Materials

Unless otherwise noted, all reagents were obtained from commercial suppliers and used without further purification. Benzene, dimethylformamide (DMF), acetonitrile (ACN), tetrahydrofuran (THF) and methylene chloride (CH_2Cl_2), were obtained from a PureSolv MD solvent purification system by Innovative Technology (solvent deoxygenated by N_2 sparge and dried over alumina). "Extra Dry" grade methanol purchased from Acros was utilized. Deuterated solvents were obtained from Cambridge Isotope labs. Two fluorescent probes (Alexa Fluor® 594 C_5 maleimide and Alexa Fluor® 488 cadaverine, sodium salt) were purchased from Life Technologies™. Silicycle silica gel plates (250 mm, 60 F254) were used for analytical TLC, and preparative chromatography was performed using SiliCycle SiliaFlash silica gel (230-400 mesh). Known compounds including α,α,α',α'-tetramethyl-1,3-benzenedipropionitrile[53], cis-

$Rh_2(OAc)_2(OCOCF_3)_2$[54], carbonate **1**[49], 2,6-Bis(2'-pyridyl)-4-pyridone[63], were prepared as previously reported.

Plasmid pET11c-tHisF was provided by the Sterner group of University of Regensburg, Germany.[44] Plasmid pEVOL-pAzF was provided by the Schultz group of the Scripps Research Institute, CA.[64] *E. coli* DH5α and BL21 (DE3) cells were purchased from Invitrogen (Carlsbad, CA). Bacillus amyloliquefaciens strain was purchased from American Type Culture Collection (Mannasas, VA). Nde I, Xho I restriction enzyme, T4 DNA ligase, Taq DNA polymerase and Phusion HF polymerase (Cat# 530S) were purchased from New England Biolabs (Ipswitch, MA). Luria broth (LB) and rich medium (2YT) and Agar media were purchased from Research Products International (Mt. Prospect, IL). Qiagen DNA extraction kit (Cat# 28706) and plasmid isolation kit (Cat# 27106) were purchased from QIAGEN Inc. (Valencia, CA) and used according to the manufacturer's instructions. DNA purification kit (Zymo, Cat# D4004) was purchased from Zymo research (Irvine, CA) and used as recommended. 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. Ni-nitrilotriacetic acid (Ni-NTA) resin and Pierce® BCA Protein Assay Kits (Cat# 23225) were purchased from Fisher Scientific International, Inc. (Hampton, NH), and the manufacturer's instructions were followed when using both products (for Ni-NTA resin, 8 mL resin were used, with buffers delivered by a peristaltic pump at a rate of 1 mL/min, in a 4 °C cold cabinet). Amicon® 10 kD spin filters for centrifugal concentration were purchased from EMD Millipore (Billerica, MA) and used at 4,000 g at 4 °C. PD-10 desalting columns (Cat# 17-0851-01) and Hitrap desalting columns (Cat# 11-0003-29) were purchased from GE Healthcare (Pittsburg, PA).

General procedures

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 125 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 determined by HPLC were calculated from internal standards (anisole for cyclopropanation and 1,2,4-trimethoxybenzene for silane insertion) and reported as the average of two trials set up in parallel. High resolution ESI mass spectra were obtained from the University of Chicago mass spectrometry facility using an Agilent Technologies 6224 TOF LC/MS. MALDI-MS spectra were recorded on AB SCIEX Voyager-DE PRO MALDI-Tof system. Amicon® 15 mL 10 kD cutoff centrifugal filter was 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 -20 °C until use.

Cloning, expression and protein purification

Standard cloning procedures and site directed mutagenesis:

A gene encoding the cyclase subunit (tHisF) of the imidazole glycerol phosphate synthase enzyme complex from Thermotoga maritima was amplified from pET11c-tHisF20 by PCR using gene specific primers containing NdeI (forward) and XhoI (reverse) restriction sites. The gene was cloned into the NdeI and XhoI sites of pET22b so that scaffolds would be expressed with a C-

terminal hexa-histidine tag for Ni-NTA affinity chromatography. Amber mutations were introduced into the tHisF gene at positions L50, G176 and I199 by site directed overlap extension (SOE) PCR[65]. For each mutation, two separate polymerase chain reactions were performed, each using a perfectly complementary flanking primer at the 5' and 3' end of the sequence and a mutagenic primer. The PCR conditions were as follows: Phusion HF buffer 1x, 0.2 mM dNTPs each, 0.5 μM forward primer, 0.5 μM reverse primer, 0.02 U/μL Phusion polymerase and 1 ng/mL template plasmid. The resulting two overlapping fragments that contained the base pair substitution were then assembled in a second PCR using the flanking primers resulting in the full-length mutated gene. For phytase gene cloning, genomic DNA was isolated from *Bacillus amyloliquefaciens* (ATCC#23350) and genomic PCR was done using gene specific primers having above restriction sites on the flanking region. Genomic PCR was performed in one step using the same conditions as above, except template (genomic DNA) concentration was increased to 500 ng/mL. An amber stop codon (Y104Az) was introduced for ArM formation, and three alanine mutations (N99A, N100A, D101A) were introduced to improve access to the site of Az mutation within the scaffold. Nucleotide sequences for the all the primers are summarized in Table S1.1.

Table S1.1. Nucleotide sequences for the primers

#	Primer name	Primer sequence
1	T7 for	5'-GCG AAA TTA ATA CGA CTC ACT ATA-3'
2	T7 rev	5'-TTA TGC TAG TTA TTG CTC AGC GG-3'
3	L50Az for	5'-GAA CTC GTT TTT TAG GAT ATC ACC GCG-3'
4	L50Az rev	5'-CGC GGT GAT ATC CTA AAA AAC GAG TTC-3'
5	G176Az for	5'-AGT ATC GAC AGA TAG GGC ACA AAA TCG-3'
6	G176Az rev	5'-CGA TTT TGT GCC CTA TCT GTC GAT ACT-3'
7	I199Az for	5'-ACA CTT CCC ATC TAG GCT TCC GGT GGT-3'
8	I199Az rev	5'-ACC ACC GGA AGC CTA GAT GGG AAG TGT-3'
9	C9ala for	5'-AGA ATA ATC GCG GCG CTC GAT GTG AAA-3'
10	C9ala rev	5'-TTT CAC ATC GAG CGC CGC GAT TAT TCT-3'
11	K243C for	5'-GAG TAC CTC AAA TGC CAC GGA GTG AAC-3'
12	K243C rev	5'-GTT CAC TCC GTG GCA TTT GAG GTA CTC-3'

Table S1.1. Nucleotide sequences for the primers, continued

#	Primer name	Primer sequence
13	D174C for	5'-C TCACC AGT ATC TGC AGA GAC GGC-3'
14	D174C rev	5'-GCC GTC TCT GCA GAT ACT GGTGA G-3'
15	phyA104 for	5'-CCTGCGATTTAGCTGGACCCCAAG-3'
16	phyA104 rev	5'-CTTGGGGTCCAGCTAAATCGCAGG-3'
17	Phy for	5'-GCAACATATGTCTGATCCTTATCATTTTACCG-3'
18	Phy rev	5'-AGCACTCGAGTTATTTTCCGCTTCTGTCAGTCA-3'

PCR amplified fragments and plasmid vector pET22b were restriction digest with Nde I and Xho I enzymes in recommended buffer at 37 °C for 2 hours. Digested DNA was cleaned by agarose gel extraction using commercial kit before ligation. Ligation was setup with a molar ratio of 1:3 (plasmid: insert) in 10 µL reaction mix. Typically a ligase reaction mix had 1 ng/L digested plasmid vector, 9 ng/mL of the insert, 1 µL 10X ligase buffer and 1 U/mL ligase. Reaction mix was incubated at 16 °C overnight, cleaned using DNA purification kits and transformed into E. coli DH5 cells. Cells were spread on LB ampicillin plates (6.25 g LB powder mix, 4 g agar, 250 mL DDI water, 0.1 mg/mL ampicillin) before recovering in SOC medium for 1 hour at 37 °C. Plates were incubated at 37 °C overnight; individual colonies that appeared next day were tested for gene fragments by colony PCR. Clones that showed amplification for desired fragments were inoculated on LB broth having 0.1 mg/mL ampicillin and grown overnight at 37 °C, 250 rpm. Recombinant plasmid from these overnight grown cultures were isolated using kit and given for sequencing. Plasmid sequencing was done by the U Chicago sequencing facility staff and T7 for and T7 rev primers were used for sequencing reactions. For FRET experiments, double mutants were created by replacing L50 with Az (L50Az) and C9 with A (C9A), and this construct was used as template to introduce cysteine residues at different positions (D174C and K243C).

Expression and purification protocol:

pET22b-tHisF (or mutants thereof) and pEVOL-pAzF were co-transformed into electrocompetent *E. coli* BL21 (DE3)[66], these cells were allowed to recover in SOC medium (37 °C, 50 min), then plated onto LB amp+Cm agar plates (6.25 g LB powder mix, 4 g agar, 250 mL DDI water, 0.1 mg/mL ampicillin, 0.05 mg/mL chloramphenicol), and incubated at 37 °C for 16 h. Several colonies appeared on overnight-incubated plates; a single colony from this plate was inoculated in 5 mL 2YT medium having antibiotics with the same concentrations as above. The culture was incubated overnight at 30 °C with constant shaking at 250 rpm. On the following day, 3 mL of the overnight cultures was used to inoculate 300 mL of fresh 2YT media having the same antibiotics, in 1 L Erlenmeyer flask. The culture was incubated at 30 °C, 250 rpm, and protein expression was induced by adding 1mM IPTG, 2mM 4-Azido-phenyl alanine and 1% (w/v) L-arabinose when OD_{600} reached 1. The induced culture was allowed to grow for 12 hours, and then the cells were harvested by centrifugation at 4 °C, 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 then clarified at 16000 x g, 4 °C for 30 minutes and supernatant thus obtained was purified by Ni-NTA resin using manufacturer's instructions. Purified protein was buffer exchanged to 10 mM Tris (pH 7.5) and measured by Pierce® BCA Protein Assay Kit as recommended.

Cofactor and Probe Synthesis

Compound **2**: To a 50 mL glass bomb were added α, α, α', α'-tetramethyl-1,3-benzenedipropionitrile[53] (3.365 g, 14 mmol), $[Ir(COD)(OMe)]_2$ (30.5 mg, 0.046 mmol, 0.003 equiv), 4,4'-di-tert-butylbipyridine (24.7 mg, 0.092 mmol, 0.006 equiv), and B_2pin_2 (2.489 g, 9.8 mmol, 0.7 equiv)[67]. The bomb was evacuated and refilled with N_2 three times. Under a positive

flow of N_2, THF (22.0 mL) was added. The bomb was then sealed and heated in an 80 °C oil bath for 48 h. The solvent was removed under reduced pressure, and the resulting residue was purified by flash chromatography (silica gel, 85:15 hexanes/EtOAc) to yield a crude product. The crude product was dissolved in 100 mL MeOH, and hydrogen peroxide solution (30 % (w/w) in H_2O, 19 mL, 167.6 mmol, 12.0 equiv) was added. The resulting solution was stirred at room temperature for 2 h, and the solvent was evaporated. Purification by flash chromatography (silica gel, 3:1 hexanes/EtOAc) afforded phenol 2a as a white solid (2.401 g, 68%). ^{1}H NMR (500 MHz, $CDCl_3$) δ 6.72 (d, *J* = 1.0 Hz, 2H), 6.71 (d, *J* = 1.0 Hz, 1H), 2.74 (s, 4H), 1.35 (s, 12H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 156.0, 137.5, 124.9, 124.5, 116.3, 46.5, 33.6, 26.7. HRMS-ESI (m/z): calcd for $C_{16}H_{19}N_2O$ $[M-H]^-$: 255.1492, found: 255.1504.

Phenol **2a**: (2.401 g, 9.37 mmol) and KOH (3.152 g, 56.2 mmol, 6.0 equiv) were dissolved in 16 mL ethylene glycol[53]. The resulting solution was heated at 180 °C for 6 h. After cooling the reaction to room temperature, the contents were partitioned between 24 mL of $CHCl_3$ and 24 mL of H_2O. The aqueous layer was collected, acidified with 6 M aqueous HCl to pH 1, and extracted with EtOAc (80 mL x 3). The organic layer was washed successively with 50 mL of H_2O and 50 mL of brine, dried over Na_2SO_4, and concentrated under reduced pressure. Purification by flash chromatography (silica gel, 19:1 CH_2Cl_2/MeOH) afforded 2b as white solid (2.185 g, 79%). ^{1}H NMR (500 MHz, CD_3OD) δ 6.47 (m, 3H), 2.74 (s, 4H), 1.13 (s, 12H). ^{13}C NMR (125 MHz, CD_3OD) δ 181.7, 157.8, 140.4, 125.0, 116.4, 47.2, 44.5, 25.8. HRMS-ESI (m/z): calcd for $C_{16}H_{22}O_5Na$ $[M+Na]^+$: 317.1365, found: 317.1357.

Phenol diacid **2b**: (48 mg, 0.16 mmol), cis-$Rh_2(OAc)_2(OCOCF_3)_2$[54] (100 mg, 0.16 mmol, 1.0 equiv) and potassium carbonate (47 mg, 0.32 mmol, 2.0 equiv) were added into a 25 mL round-bottom flask. 8 mL of THF was added, and the resulting suspension was heated at 50 °C for 3 h. The solvent was evaporated, and purification by flash chromatography (silica gel, 4:1 benzene/acetonitrile) yielded 2 as purple solid (77 mg, 80%, bis-acetonitrile adduct)[55]. ^{1}H NMR (500 MHz, CD_3CN) δ 6.37 (s, 1H), 6.33 (s, 2H), 2.52 (s, 4H), 2.00 (s, 6H), 1.81 (s, 6H), 0.96 (s, 12H). ^{13}C NMR (125 MHz, CD_3CN) δ 197.6, 191.7, 156.4, 140.5, 123.9, 115.6, 47.5, 46.4, 26.0, 23.5. HRMS-ESI (m/z): calcd for $C_{20}H_{26}ClO_9Rh_2$ $[M+Cl]^-$: 650.9376, found: 650.9395.

Cofactor **3**: Complex 2 (54 mg, 0.088 mmol) and sodium hydride (60 % dispersion in mineral oil, 3.2 mg, 0.08 mmol, 0.9 equiv) were added into a 10 mL round-bottom flask. 3 mL THF was added and, the resulting suspension was heated at 50 °C for 1 h and cooled to room temperature. A solution of carbonate **1**[49] (31 mg, 0.098 mmol, 1.1 equiv) in 1 mL THF was added. The reaction was stirred at room temperature for 1 h, and the solvent was evaporated. Purification by flash chromatography (silica gel, 4:1 benzene/acetonitrile) yielded cofactor **3** as a dark green solid (50mg, 72%). ^{1}H NMR (500 MHz, CD_3CN) δ 6.73 (d, J = 1.5 Hz, 1H), 6.68 (d, J = 1.5 Hz, 2H), 4.14 (d, J = 6.9 Hz, 2H), 2.58 (s, 4H), 2.44 – 2.22 (m, 4H), 2.20 (m, 2H), 1.49 (m, 2H), 0.92 (s, 6H), 0.84 (m, 3H). ^{13}C NMR (125 MHz, CD_3CN) δ 197.6, 192.0, 154.7, 150.9, 140.7, 129.6, 121.4, 99.0, 73.6, 47.1, 46.6, 33.5, 25.9, 23.9, 23.7, 23.5, 21.7. HRMS-ESI (m/z): calcd for $C_{31}H_{42}NO_{11}Rh_2$ $[M+NH_4]^+$: 810.0868, found: 810.0831.

Compound **5**: 2,6-bis(2'-pyridyl)-4-pyridone[63] (39 mg, 0.16 mmol) and 1 (48 mg, 0.15 mmol, 0.95 equiv) were dissolved in 12 mL THF. Cesium carbonate (73 mg, 0.22 mmol, 1.5 equiv) was

added, and the suspension was stirred at room temperature overnight. The mixture was filtered through celite and the filtrate was concentrated down to yellow oil. The crude was purified by flash chromatography (silica gel, 49:1 CH_2Cl_2/MeOH)) to afford compound 5 (32 mg, 48 %) as a white solid. ^{1}H NMR (500 MHz, $CDCl_3$) δ 8.80 (d, *J* = 4.3 Hz, 2H), 8.64 (d, *J* = 8.0 Hz, 2H), 8.34 (s, 2H), 7.95 (t, *J* = 7.7 Hz, 2H), 7.47 – 7.40 (m, 2H), 4.24 (d, *J* = 6.5 Hz, 2H), 2.47 (m, 2H), 2.32 (m, 2H), 2.18 (m, 2H), 1.42 (m, 2H), 0.87 (m, 3H). ^{13}C NMR (125 MHz, $CDCl_3$) δ 160.2, 157.9, 155.4, 152.7, 149.4, 137.0, 124.3, 121.4, 113.6, 98.9, 73.9, 33.3, 31.1, 23.4, 23.2, 21.4. HRMS-ESI (m/z): calcd for $C_{26}H_{24}N_3O_3$ $[M+H]^+$: 426.1818, found: 426.1812.

Cofactor **6**: 1 (29 mg, 0.068 mmol) was dissolved in 1 mL dichloromethane. A solution of $CuCl_2$ (9.2 mg, 0.068 mmol, 1.0 equiv) in 1 mL of DCM/methanol 1:1 was added dropwise. The resulting solution was stirred at room temperature for 4 h, and the solvent was evaporated. The residue was washed with cold dichloromethane and cold methanol and dried to yield **6** (32 mg, 84%) as a paramagnetic, blue solid[56]. HRMS-ESI (m/z): calcd for $C_{26}H_{23}ClCuN_3O_3$ $[M-Cl]^+$: 523.0724, found: 523.0722.

Cofactor **7**: 5 (21.2 mg, 0.05 mmol) was dissolved in 1 mL THF. A solution of $MnCl_2$•$4H_2O$ (49.5 mg, 0.25 mmol, 5.0 equiv) in 5 mL THF was added in one portion. A yellow precipitate formed immediately after addition. The resulting suspension was stirred for 15 minutes and filtered. The solid was washed with copious THF and dried under vacuum to afford 7 (10 mg, 36%) as a paramagnetic, light yellow solid[57]. HRMS-ESI (m/z): calcd for $C_{26}H_{23}ClMnN_3O_3$ $[M-Cl]^+$: 515.0809, found: 515.0792.

Probe **8**: In a 1.5 mL microcentrifuge tube, Alexa Fluor® 488 cadaverine, sodium salt (0.5 mg, 0.78 µmol), 1 (0.50 mg, 1.56 µmol, 2 equiv), and N-ethyldiisopropylamine (0.7 µL, 4.02 µmol, 5.2 equiv) were dissolved in 1 mL DMF. The tube was sealed with aluminum foil and shaken at room temperature for 48 h, and the solvent was evaporated. The residue was dissolved in 9:1 water/acetonitrile, purified by reversed-phase HPLC using water and acetonitrile, and lyophilized to a fine powder. See supporting information for HPLC trace. HRMS-ESI (m/z): calcd for $C_{37}H_{37}N_4O_{12}S_2$ $[M]^+$: 793.1850, found: 793.1890.

*HPLC trace of probe **8***

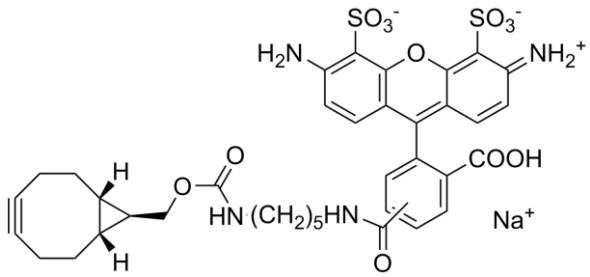
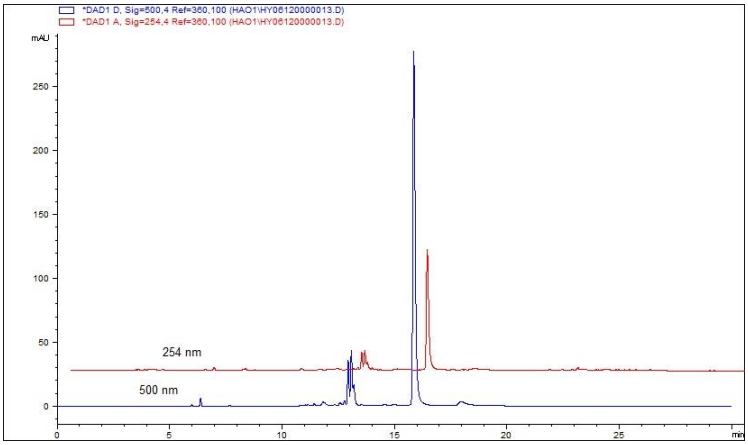


Figure S1.1. Analytical HPLC trace of purified **9**. Method: 0% to 50% B from 0-20 min, 50% to 100% B from 20-25 min. Solvent A: water + 0.1% TFA; solvent B: acetonitrile.



ArM formation

*Bioconjugation of **3**:*

A solution of tHisF (480 µL, 75 µM tHisF in 10 mM Tris, pH 7.5) and a solution of cofactor **3** (120 µL, 1.188 mg/mL in acetonitrile, 5.0 equiv) were added to an microcentrifuge tube and shaken

in the dark at 4 °C for 48 h. The final concentrations were: 60 μM tHisF-Az50, 300 μM **3**, 20 vol% acetonitrile. Due to the high hydrophobicity of cofactor **3**, coelution of **3** with tHisF in size exclusion chromatography or anion exchange chromatography was observed (other cofactors **6**, **7** were removed by simple gel filtration due to their relatively good hydrophilicity). Hence, the reaction was purified by reversed-phase preparative HPLC and exchanged to 10 mM Tris (pH 7.5) with centrifugal filters. The removal of excess cofactor **3** was confirmed by analytical HPLC analysis, the concentration of product was determined with Pierce® BCA Protein Assay Kit, and the conversion was estimated by MALDI-MS analysis. Because apparent in-situ bioconjugation under MALDI-MS conditions was observed, a modified sinapinic acid matrix was used to quench any unreacted azide (10 mg/mL of sinapinic acid and 3 mg/mL of cyclooctaalkyne alcohol in 50:50 water/acetonitrile, 0.1 % TFA final conc.).

*Bioconjugation of **6** or **7***:

General procedure for the conjugation of cofactor **6** or **7**: A tHisF solution in 10 mM Tris (pH 7.5) and a solution of cofactor in methanol were mixed and incubated in the dark at 4 °C for 24 h. The final concentrations were: 60 μM tHisF, 300 μM, 15 vol% methanol. The reaction was desalted by gel-filtration with PD-10 desalting columns to remove excess cofactors. The purity of metalloenzyme was determined by analytic HPLC analysis. Protein concentrations were measured using the Pierce® BCA Protein Assay Kit. The conversion was estimated by MALDI-MS analysis with a modified sinapinic acid matrix described above.

HPLC monitoring and analysis of RhBCN bioconjugation:

Analytic HPLC runs were performed on an Agilent 1100 Series HPLC system using Vydac 218TP54 column (C18, 300 Å, 5 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: 20 % to 64 % B from 0-15 min, 64 % from 15-20 min, 64 % to 80 % from 20-22 min, 80 % from 22-25 min, 80 % to 20 % from 25-28 min (solvent A: water containing 0.1% TFA; solvent B: CH_3CN).

Figure S1.2. Comparison of analytical HPLC runs of cofactor **3**, tHisF protein, crude bioconjugation reaction, and purified tHisF_RhBCN. The residual amount of **3** in tHisF_RhBCN is not detectable by HPLC (Slight variations in retention time of protein or cofactors were observed in HPLC).

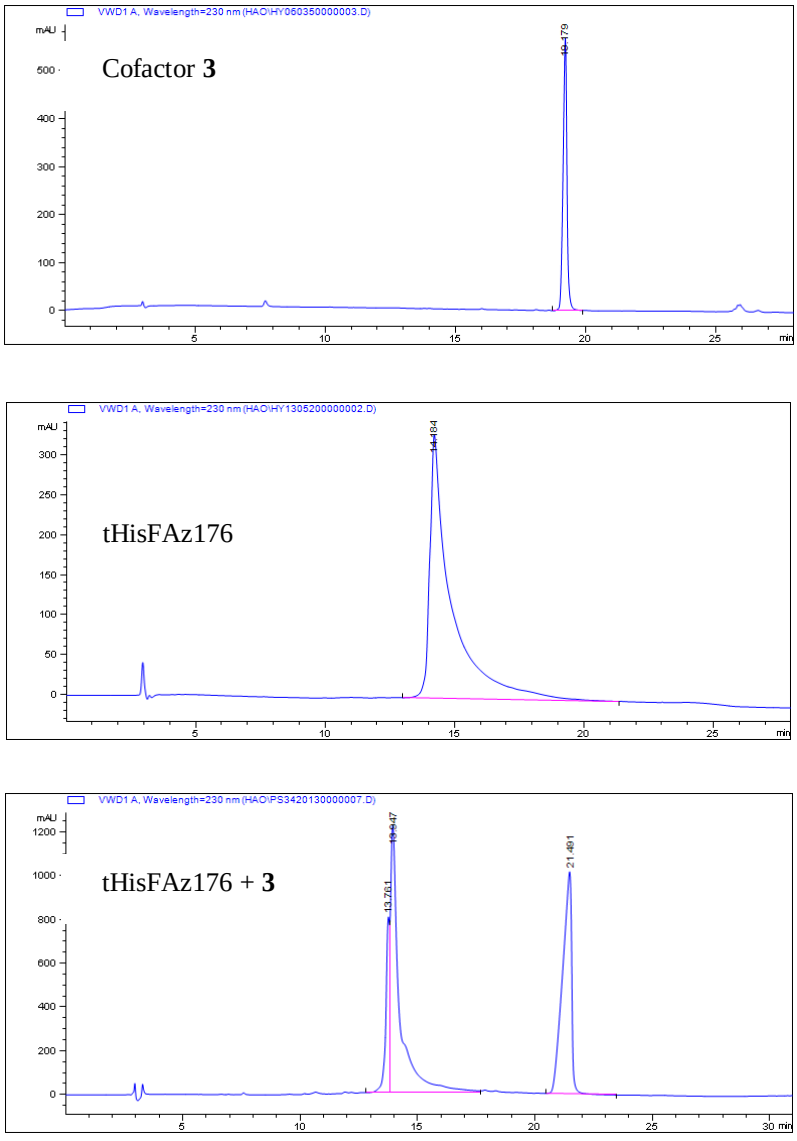
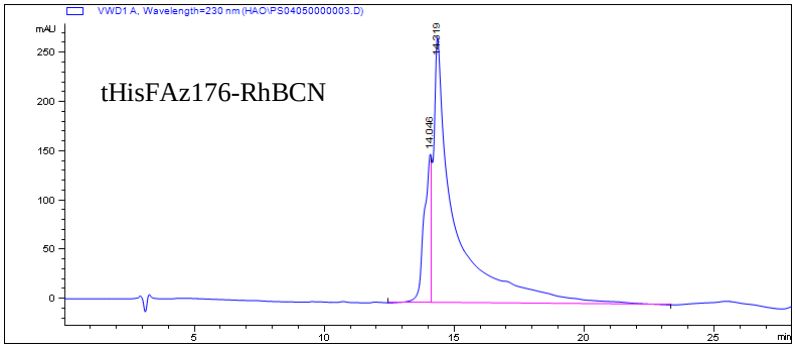
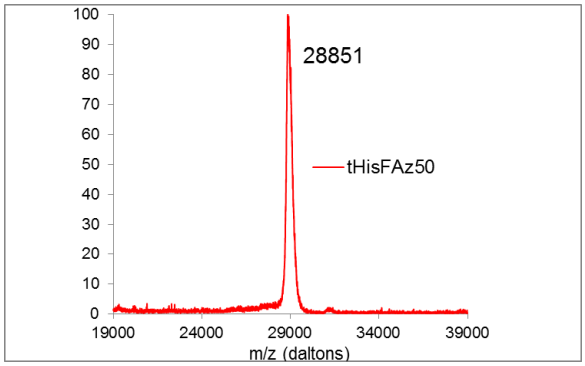


Figure S1.2. Comparison of analytical HPLC runs of cofactor **3**, tHisF protein, crude bioconjugation reaction, and purified tHisF_RhBCN, continued



Representative MALDI-MS spectra for ArM:

Figure S1.3. MALDI-MS spectra of representative bioconjugations



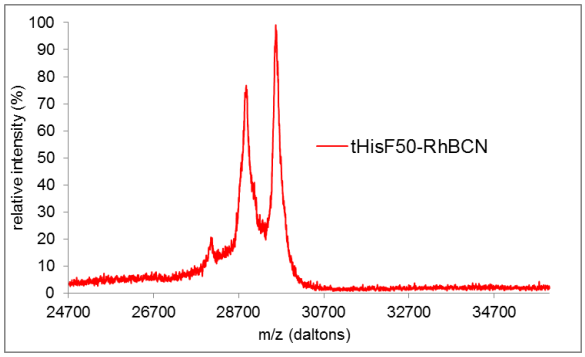
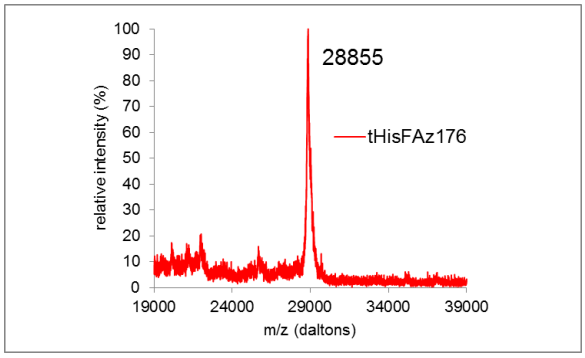
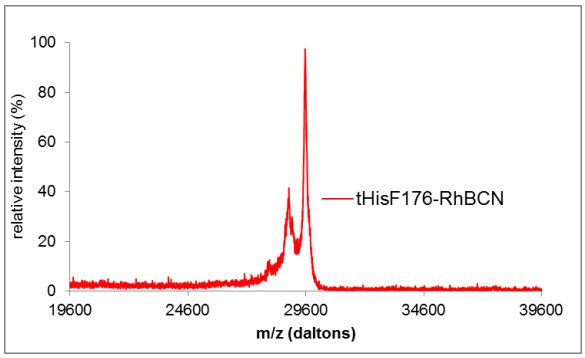


Figure S1.3. MALDI-MS spectra of representative bioconjugations, continued





Representative ESI-MS spectra for ArM (or scaffold):

In ESI-TOF MS analysis, a sample of protein was desalted with centrifugal filters to a mixture of water: acetonitrile: glacial acetic acid (49.5: 49.5: 1, v/v). The final protein concentration was around 50 μM.

Figure S1.4. ESI-MS spectra for tHisF-wt, tHisFAz50 and tHisF50-RhBCN (the corresponding deconvoluted spectra are shown in the article)

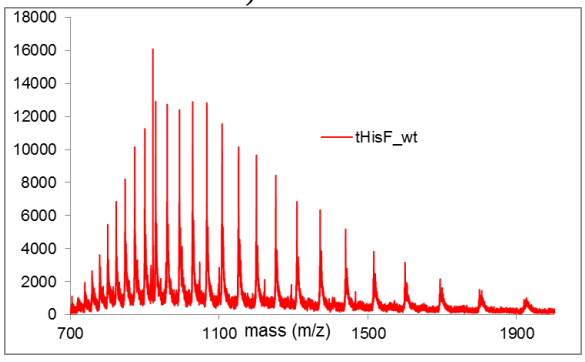
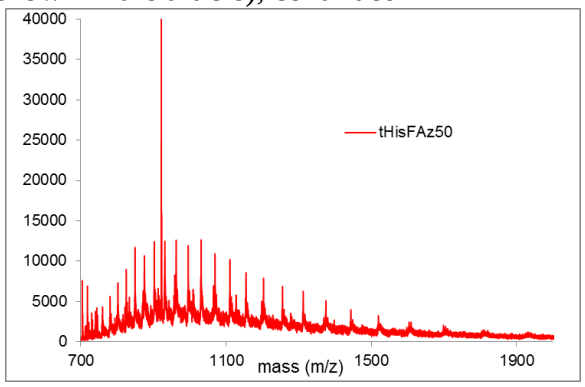
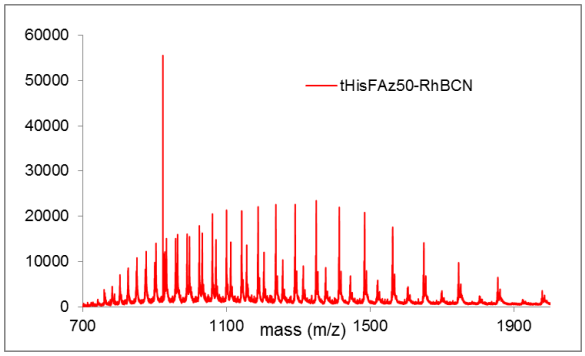


Figure S1.4. ESI-MS spectra for tHisF-wt, tHisFAz50 and tHisF50-RhBCN (the corresponding deconvoluted spectra are shown in the article), continued





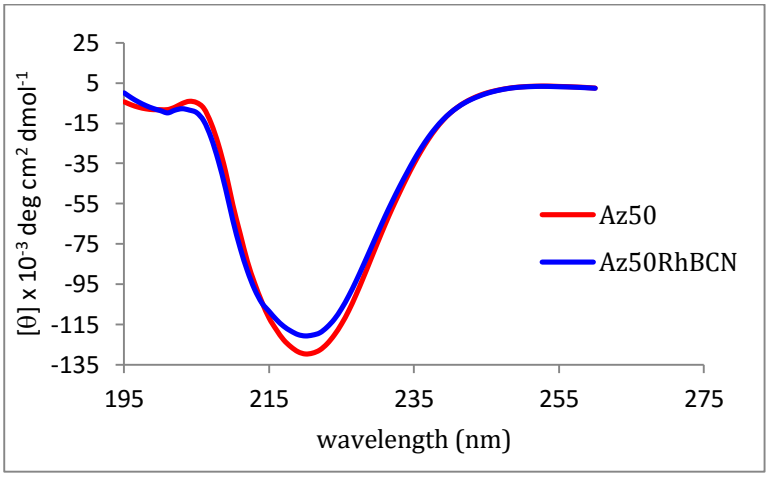
Circular Dichroism and Fluorescence Analysis

CD analysis of scaffold proteins and ArMs was conducted by loading the protein solutions into a 0.1 mm quartz cuvette. CD spectra were obtained on AVIV-202 CD spectrophotometer (AVIV Biomedical, Inc.) between 280 and 200 nm with a 1 nm increment at room temperature. Fluorescence measurements of scaffold proteins and ArMs were acquired at 290 nm using Fluorolog-3 spectrofluorometer (Horiba Jobin Yvon, Inc.) in phosphate buffer (pH 8.0). Proteins were treated with 60% acetonitrile and 6 M guanidine chloride (GdmCl) separately and fluorescence was measured at 290 nm along with non-treated protein to check for any structural perturbation.

Circular dichroism of tHisF and tHisFAz50:

Purified wild type tHisF and tHisFAz50 mutant were exchanged into 100 mM sodium phosphate buffer (pH 7.5, 150 mM NaCl), and concentrated to 10 mg/mL to provide 10X stocks for CD experiments. The protein was diluted to a final concentration of 2 mg/mL into buffer and 200 µL of each protein solution was loaded into a 2 mm quartz cuvette (Hellma QC) and CD spectra were obtained on an AVIV 202SF CD spectropolarimeter at 24 °C. For each protein, CD spectra were taken in triplicate between 260 nm and 195 nm in 1 nm increments with a 2 second integration time. Data were averaged and background spectra from free buffer solutions were subtracted.

Figure S1.5. Comparison of CD spectra of tHisFAz50 and tHisFAz50-RhBCN

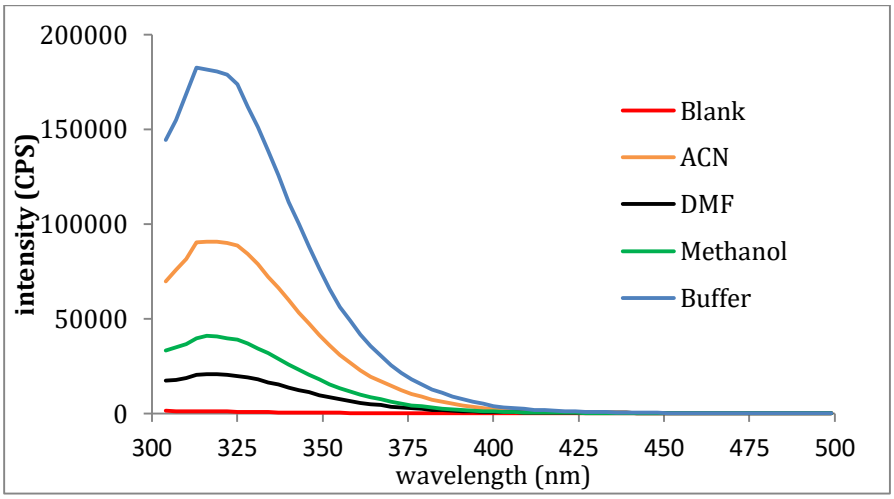


Fluorescence Analysis:

For organic solvent tolerance measurements, tHisFAz50 was first concentrated down to 100 µM in 10mM tris buffer (pH 7.5). 10 µL of this concentrated protein was then added to 90 µL Acetonitrile, 90 µL DMF and 90 µL methanol in separate 1.5 mL microcentrifuge tubes. Thus final protein concentration becomes 10 µM and buffer to organic solvent ratio becomes 10:90 (v/v) for each sample. A control sample was also prepared where 10 µL of concentrated protein was added to 90 µL buffer (10 mM Tris, pH 7.5). All the samples were incubated at room temperature for an hour and fluorescence emission spectra was measured in a Tecan infinite M200pro plate

reader. At 60% acetonitrile concentration no change was observed as described in the article. Typically, 20% acetonitrile concentration was used in bioconjugation reactions.

Figure S1.6. Fluorescence emission spectra of tHisFAz50 in different organic solvents



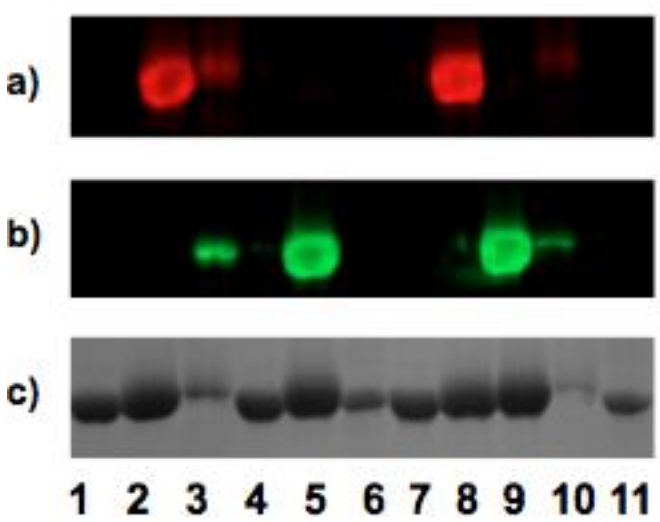
Dual-Labelling FRET Measurements[61]

Double labeling of Azide(1)/Cys mutants

Purified tHisFAz50/Cys double mutants were exchanged into labeling buffer (10 mM Tris pH 7.5), by washing 3X 100 fold dilutions using a 10 kDa cutoff 15 mL Amicon® Ultra Centrifugal Filtration Device. 10 equivalents of Alexa594-maleimide were then added to label the free cysteine thiol of 100 µM protein. Reaction was done at room temperature for 16 hours in the dark with constant shaking. Excess unreacted dye was removed by desalting columns followed by buffer exchange 3 times and concentrated to 80 µL. The dye to protein ratio (Alexa594-maleimide:tHisF) was estimated by measuring the absorbance at 280 and 595 nm for tHisF and Alexa594-maleimide, respectively. 10 equivalents of Alexa488-BCN (probe **8**) were then added and the reaction was allowed to proceed at room temperature for 16 h. Excess dye was removed by exhaustive dialysis against water. Protein was aliquoted and stored in 10 mM Tris buffer (pH 7.5) and 150 mM NaCl. The dye to protein ratio (Alexa488:tHisF) was determined by measuring the absorbance at 280

and 495 nm for tHisF and Alexa488, respectively. Extinction coefficients of tHisF (11460 $M^{-1}cm^{-1}$ at 280 nm)[68], Alexa488-BCN (70,000 $M^{-1}cm^{-1}$ at 495 nm, Provided by Invitrogen), Alexa594-maleimide (80000 $M^{-1}cm^{-1}$ at 595 nm)[69] were used to calculate molar concentration of the protein and the dyes. Reactions performed at 100 μM protein concentration typically gave > 80 % labeling after 16 h reaction at 24 °C. Labeling reactions were tried at different conditions and low concentrations of proteins (less than 100 μM) showed slow reaction and a significant decrease in yield. Also negligible labeling was observed when 50mM phosphate buffer (pH 7) was used. Fluorescence Scanning was done by running the labeled proteins (both singly and doubly) on 10% SDS PAGE. The protein gel was scanned for Alexa488-BCN and Alexa594-maleimide using a fluorescence scanner (Bio-Rad FX pro plus, see Figure S1.7).

Figure S1.7. Fluorescence scanned SDS-PAGE picture of labeled proteins



(Lane 1, 4 and 7: control protein (tHisF-C9A); lane 2, 8: tHisF-C9A-L50Az-D174C-Alexa488 and tHisF-C9A-L50Az-K243C-Alexa488 respectively; lane 3 and 10: doubly labeled, tHisF-C9A-L50Az-D174C-Alexa488-Alexa594 and tHisF-C9A-L50Az-K243C-Alexa488-Alex594 respectively; lane 5 and 9: tHisF-C9A-L50Az-D174C-Alexa594 and tHisF-C9A-L50Az-K243C-Alex594 respectively; lane 6 and 11: protein markers.)

Steady state FRET measurement

Steady-state fluorescence measurements were performed at 4 °C on a Fluoromax-3 spectrofluorometer (Horiba Jobin Yvon Inc.). Fluorescence was measured for tHisFAlexa488-BCN (donor) and doubly labeled mutant where tHisFAlexa594-maleimide was the acceptor. The reduction of donor fluorescence emission by the acceptor was recorded between 500 and 700 nm and corrected for the buffer blank. The protein concentrations of tHisFAlexa488-BCN and tHisFAlexa594-maleimide were kept at 2 μM. The experiments were conducted in 10 mM Tris-HCl buffer (pH 7.5). The distance between residues of tHisFAlexa488-BCN and tHisFAlexa594-maleimide was estimated spectroscopically by FRET. The distance is given by: $R = R_0(E^{-1}-1)^{1/6}$,[61] where R is calculated in Å, R_0 is the Foster critical distance, and E is the FRET efficiency given by $E=(1-F_{da}/F_d) \times 1/F_a$, where F_{da}, F_d and F_a are fluorescence intensities in both, donor, and acceptor channels. R_0 is given by: $R_0 = 9.79 \times 10^3 (\kappa^2 J \Phi_D \eta^{-4})^{1/6}$, where κ2 is the orientation factor, η is the refractive index of the buffer, Φ_D is the quantum yield of the donor, and J is the overlap integral in cm^3 /M given by, $J=\int F_D(\lambda)\varepsilon(\lambda)\lambda_4 d\lambda / F_D(\lambda) d\lambda$, where λ is the wavelength in cm, $F_D(\lambda)$ is the corrected fluorescence of the donor, and $\varepsilon(\lambda)$ is the acceptor molar absorption coefficient in M^{-1} cm^{-1}. J was obtained by numerical integration of normalized spectra. InstruView v-0.5 software (Columbia University) was used to calculate the R_0 values using the defined parameters.

Time-domain lifetimes were measured on a ChronosBH lifetime fluorometer (ISS, Inc.) using Time-Correlated Single Photon Counting (TCSPC) methods. The fluorometer contained Becker-Hickl SPC-130 detection electronics and a HPM-100-40 Hybrid PMT detector. Tunable picosecond pulsed excitation was provided by a Fianium SC400 supercontiuum laser source and integrated AOTF. Emission wavelengths were selected with bandpass filters (Semrock and Chroma). The Instrument Response Function (IRF) was measured to be approximately 120 ps

FWHM in a 1% scattering solution of Ludox LS colloidal silica. Lifetimes were fit via a forward convolution method in the Vinci control and analysis software.

Table S1.2. R_0 values calculated from steady state FRET measurement

Serial	Double mutant	Energy transfer $E=(1-F_{da}/F_d) \times 1/F_a$		Distance (R) in A^0 $E=1/1+(R/R_0)^6$	
		Steady state	Lifetime	Steady state	Lifetime
1	Az50, D174C (top)	0.54	0.63	58	55
2	Az50, K243C (bottom)	0.11	0.15	88	80

((F_{da}: Fluorescence of doubly labeled protein; F_d: Fluorescence of donor protein; F_a: Fluorescence of acceptor protein)

ArM Catalysis

Preparation of standard products for cyclopropanation

In a 25 mL round-bottom flask, 4-methoxystyrene (482 µL, 3.6 mmol) and rhodium acetate dimer (10.6 mg, 0.024 mmol, 0.0067 equiv) were dissolved in 7.5 mL ether. A solution of ethyl diazoacetate (351 µL, 3.0 mmol) in 5 mL ether was added dropwisely over 30 minutes. The reaction was stirred at room temperature for 8 h and the solvent was evaporated. Purification by flash chromatography (silica gel, 19:1 hexanes/EtOAc) afforded both *cis*-isomer (60 mg, 9 %) and *trans*-isomer (146 mg, 22 %). The characterization of products is consistent with literature reports[70,71].

Preparation of standard products for silane insertion

In a 25 mL round-bottom flask, methyldiphenylsilane (67 µL, 0.44 mmol, 1.1 equiv) and rhodium acetate dimer (1.8 mg, 0.004 mmol, 0.01 equiv) were dissolved in 4 mL hexane. A solution of methyl phenyldiazoacetate (59 µL, 0.4 mmol) in 4 mL hexane was added dropwisely over 1 h. The reaction was stirred at room temperature overnight and the solvent was evaporated. Purification by flash chromatography (silica gel, 19:1 hexanes/EtOAc) afforded the product (54

mg, 39 %), which is consistent with literature report[72]. The O-H insertion product methyl DL-mandelate was prepared as previously reported[73].

Catalytic cyclopropanation

In a 1.5 mL microcentrifuge tube, tHisFA176-RhBCN solution (80 µL, 100 µM), 90 µl phosphate buffer (0.1 M, pH 7.5), and 26 µL THF were added. A mixed solution of styrene and diazoacetate in THF (4 µL, styrene 600 mM, diazoacetate 200 mM) was added. The resulting solution was left shaking at room temperature overnight. The final concentrations of the reagents were: 12 mM styrene, 4 mM diazoacetate, 40 µM tHisF-Az176-RhBCN solution. The reaction was quenched by adding 800 µL chloroform to the closed vials and immediately vortexing the mixture. The vial was then opened and 20 µL internal standard (18.4 mM anisole in acetonitrile) was added. The mixture was vortexted and centrifuged (16,000xg, 1 min). The bottom organic layer was evaporated and re-dissolved in 200 µL acetonitrile and analyzed by HPLC.

Catalytic silane insertion

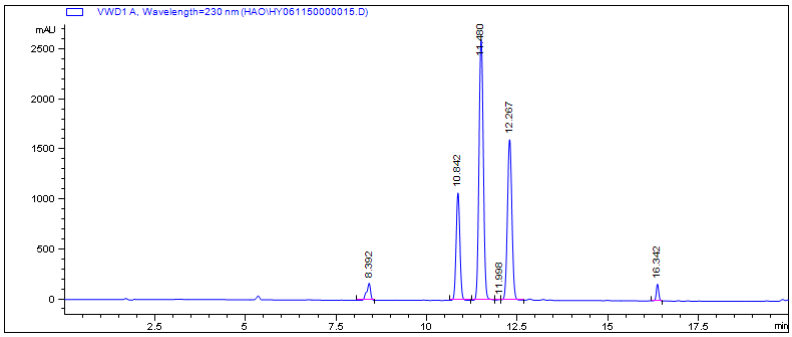
In a 1.5 mL microcentrifuge tube, tHisF-Az176-RhBCN solution (75 µL, 130 µM), 95 µL phosphate buffer (0.1 M, pH 7.5) and 22 µL THF were added. A mixed solution of silane and diazoacetate in THF (8 µL, styrene 25 mM, diazoacetate 125 mM) was added. The resulting solution was left stirring at room temperature overnight. The final concentrations of the reagents were: 1 mM silane, 5 mM diazoacetate, 40 µM tHisFA176-RhBCN solution. The reaction was quenched by adding 800 µL chloroform to the closed vials and immediately vortexing the mixture. The vial was then opened and 8 µL internal standard (26.8 mM 1,2,4-trimethoxybenzene in

acetonitrile) was added. The mixture was vortexted and centrifuged (16,000 x g, 1 min). The bottom organic layer was evaporated and re-dissolved in 200 µL hexane and analyzed by HPLC.

HPLC analysis of cyclopropanation:

The analytic HPLC run for cyclopropanation was performed on an Agilent 1100 Series HPLC system using an Agilent Eclipse Plus C18 column (95 Å, 3.5 µM, 4.6 mm i.d. x 150 mm), with with a flow rate of 1.0 mL/min and detection wavelength set at 230 nm. The following gradient was used: 20 % to 50 % B from 0-5 min, 50 % from 5-10 min, 50 % to 80 % from 10-15 min, 80 % from 15-18 min, 80 % to 20 % from 18-20 min (solvent A: water containing 0.1% TFA; solvent B: CH_3CN).

Figure S1.8. HPLC traces of cyclopropanation catalyzed by tHisF176-RhBCN (1: anisole internal standard; 2: cis-product isomer; 3: 4-methoxystyrene starting material; 4: trans-product isomer)



HPLC analysis of silane insertion:

The analytic HPLC run for cyclopropanation was performed on an Agilent 1200 UHPLC system using a Lux® 3u Cellulose-1 column (1000 Å, 3.0 µM, 4.6 mm i.d. x 250 mm), with with a flow rate of 1.0 mL/min and detection wavelength set at 230 nm. The following gradient was used: 97 % B from 0-15 min, 97-90 % from 15-17 min, 90 % from 17-26 min, 90-97 % from 26-28 min, 97 % from 28-30 min (solvent A: isopropanol; solvent B: hexane).

Figure S1.9. HPLC trace of silane insertion catalyzed by tHisF176-RhBCN (1: phenyldimethysilane; 2: one product enantiomer; 3: methyl phenyldiazoacetate; 4: the other enantiomer; 5: 1,2,4-trimethoxybenzene internal standard)

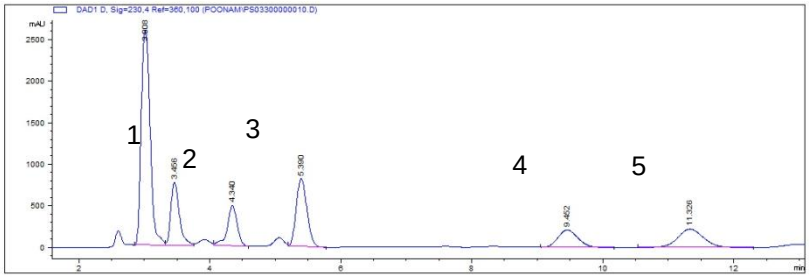


Table S1.3 Summary of yields for silane insertion and cyclopropanation by ArMand small molecule dirhodium catalyst

catalyst	silane insertion (relative to silane)[a]		cyclopropanation%
	Si-H insertion(%)	diazo hydrolysis(%)	
tHisFAz176-RhBCN (top)	28	322	81 (cis : trans = 1: 1.8)
tHisFAz50-RhBCN (middle)	5	327	69 (cis : trans = 1: 1.8)
tHisFAz199-RhBCN (bottom)	6	98	60 (cis : trans = 1: 1.8)
2-OAc	80	418	99 (cis : trans = 1: 1.8)

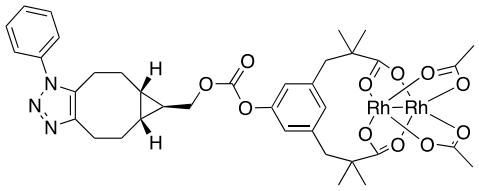
[a] All the yields were calculated relative to limiting reagent (silane for Si-H insertion and olefin for cyclopropanation) in the reactions by analysis of HPLC traces for crude reaction mixtures.

DFT model of ArM

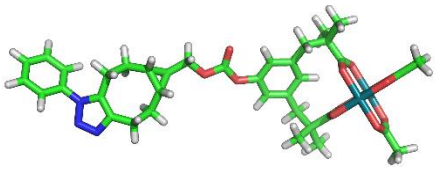
The phenyl azide adduct of cofactor **3** (Figure S1.10a) was optimized using density functional theory calculation (DFT, B3LYP, LANL2DZ) using Gaussian09 (Figure S1.10b). The mutation wizard in Pymol was used to convert residue Ile199 of tHisF (PDB#1THF) to Ala. The Ala methyl and terminal phenyl of the cofactor **3** adduct were fused using the "fuse" command in Pymol. Bond angles were manually adjusted to provide a rough model of the ArM (Figure S1.10c). While crude, this model provides some idea of the relative scale of scaffold and cofactor.

Figure S1.10. a) Structure of phenyl azide-**3** adduct; b) DFT-optimized structure of adduct; c) tHisF-Az199-RhBCN (note that higher mutants, 50 and 176, should project cofactor further into solution)

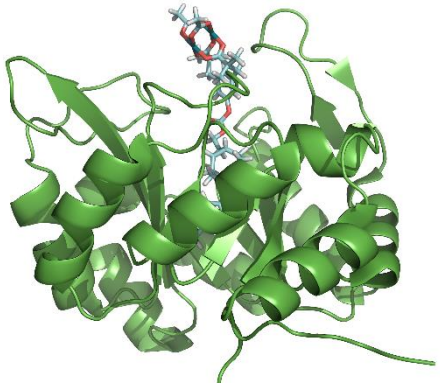
a



b



c



REFERENCES

1. National Research Council. *Catalysis for Energy*; National Academies Press: Washington, DC, 2009.

2. Umile, Thomas P. *Catalysis for Sustainability: Goals, Challenges, and Impacts*; CRC Press: Boca Raton, FL, 2016.

3. Swiegers, G., Ed.; *Mechanical Catalysis*; John Wiley & Sons: New York, 2008.

4. de Meijere, A., Brase, S., Oestreich, M., Eds.;¨Metal Catalyzed Cross-Coupling Reactions and More; Wiley-VCH: New York, 2013.

5. Yu, J.-Q., Shi, Z., Eds.; Topics in Current Chemistry, C−H Activation; Springer: Berlin, New York, 2010; Vol 292.

6. Guan, Z., Ed.; Topics in Current Chemistry, Metal Catalysts in Olefin Polymerization; Springer: Berlin, New York, 2009; Vol 26.

7. Cossy, J., Arseniyadis, S., Meyer, C., Eds.; Metathesis in Natural Product Synthesis; John Wiley & Sons: New York, 2010.

8. Hartwig, J. F. Organotransition Metal Chemistry: from Bonding to Catalysis; University Science Books: Sausalito, CA, 2010.

9. Balcells, D.; Moles, P.; Blakemore, J.; Raynaud, C.; Brudvig, G.; Crabtree, R.; Eisenstein, O., Molecular recognition in Mn-catalyzed C-H oxidation. Reaction mechanism and origin of selectivity from a DFT perspective. Dalton Transactions **2009**, (30), 5989-6000.

10. Yang, J.; Gabriele, B.; Belvedere, S.; Huang, Y.; Breslow, R., Catalytic oxidations of steroid substrates by artificial cytochrome P-450 enzymes. *Journal of Organic Chemistry* **2002,** *67* (15), 5057-5067.

11. Shaw, W.; Helm, M.; DuBois, D., A modular, energy-based approach to the development of nickel containing molecular electrocatalysts for hydrogen production and oxidation. *Biochimica Et Biophysica Acta-Bioenergetics* **2013,** *1827* (8-9), 1123-1139.

12. Hapiot, F.; Bricout, H.; Tilloy, S.; Monflier, E., Functionalized Cyclodextrins as First and Second Coordination Sphere Ligands for Aqueous Organometallic Catalysis. *European Journal of Inorganic Chemistry* **2012,** (10), 1571-1578.

13. a) Knowles, R.; Jacobsen, E., Attractive noncovalent interactions in asymmetric catalysis: Links between enzymes and small molecule catalysts. *Proceedings of the National Academy of Sciences of the United States of America* **2010,** *107* (48), 20678-20685. b) Zhao, M.; Wang, H.; Ji, L.; Mao, Z., Insights into metalloenzyme microenvironments: biomimetic metal complexes with a functional second coordination sphere. *Chemical Society Reviews* **2013,** *42* (21), 8360-8375. c) Linjalahti, H.; Feng, G.; Mareque-Rivas, J.; Mikkola, S.; Williams, N., Cleavage and isomerization of UpU promoted by dinuclear metal ion complexes. *Journal of the American Chemical Society* **2008,** *130* (13), 4232-4233. d) Zhang, T.; Edwards, N.; Bonizzoni, M.; Anslyn, E., the Use of Differential Receptors to Pattern Peptide Phosphorylation. *Journal of the American Chemical Society* **2009,** *131* (33), 11976-11984. e) Das, S.; Incarvito, C.; Crabtree, R.; Brudvig, G., Molecular recognition in the selective oxygenation of saturated C-H bonds by a dimanganese catalyst. *Science* **2006,** *312* (5782), 1941-1943. f) Zhou, Y.; Zhao, M.; Li, J.; Mao, Z.; Ji, L., Carboxylic ester hydrolysis catalyzed by a host-guest system constructed by cyclodextrin dimer and zinc complex. *Journal of Molecular Catalysis a-Chemical* **2008,** *293* (1-2), 59-64.

14. Bastero, A.; Gottker-Schnetmann, I.; Rohr, C.; Mecking, S., Polymer microstructure control in catalytic polymerization exclusively by electronic effects of remote substituents. *Advanced Synthesis & Catalysis* **2007,** *349* (14-15), 2307-2316.

15. Weberski, M.; Chen, C.; Delferro, M.; Zuccaccia, C.; Macchioni, A.; Marks, T., Suppression of beta-Hydride Chain Transfer in Nickel(II)-Catalyzed Ethylene

Polymerization via Weak Fluorocarbon Ligand-Product Interactions. *Organometallics* **2012,** *31* (9), 3773-3789.

16. Bar-Even, A.; Noor, E.; Savir, Y.; Liebermeister, W.; Davidi, D.; Tawfik, D.; Milo, R., The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* **2011,** *50* (21), 4402-4410.

17. Radzicka, A.; Wolfenden, R., A Proficient Enzyme. *Science* **1995,** *267* (5194), 90-93.

18. Henzler-Wildman, K.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M.; Petsko, G.; Karplus, M.; Hubner, C.; Kern, D., Intrinsic motions along an enzymatic reaction trajectory. *Nature* **2007,** *450* (7171), 838-844.

19. Wagner, A., Opinion Neutralism and selectionism: a network-based reconciliation. *Nature Reviews Genetics* **2008,** *9* (12), 965-974.

20. a) Benkovic, S.; Hammes-Schiffer, S., A perspective on enzyme catalysis. *Science* **2003,** *301* (5637), 1196-1202. b) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D., How enzymes work: Analysis by modern rate theory and computer simulations. *Science* **2004,** *303* (5655), 186-195.

21. a) Tracewell, C.; Arnold, F., Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Current Opinion in Chemical Biology* **2009,** *13* (1), 3-9. b) Bershtein, S.; Tawfik, D., Advances in laboratory evolution of enzymes. *Current Opinion in Chemical Biology* **2008,** *12* (2), 151-158.

22. a) Seelig, B.; Szostak, J., Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* **2007,** *448* (7155), 828-831. b) Moffet, D.; Hecht, M., De novo proteins from combinatorial libraries. *Chemical Reviews* **2001,** *101* (10), 3191-3203.

23. Renata, H.; Wang, Z.; Arnold, F., Expanding the Enzyme Universe: Accessing Non-Natural Reactions by Mechanism-Guided Directed Evolution. *Angewandte Chemie-International Edition* **2015**, *54* (11), 3351-3367.

24. a) Waldron, K.; Robinson, N., How do bacterial cells ensure that metalloproteins get the correct metal? *Nature Reviews Microbiology* **2009,** *7* (1), 25-35. b) Lu, Y.; Yeung, N.; Sieracki, N.; Marshall, N., Design of functional metalloproteins. *Nature* **2009,** *460* (7257), 855-862.

25. Thomas, C.; Ward, T., Artificial metalloenzymes: proteins as hosts for enantioselective catalysis. *Chemical Society Reviews* **2005,** *34* (4), 337-346.

26. a) Rosati, F.; Roelfes, G., Artificial Metalloenzymes. *Chemcatchem* **2010,** *2* (8), 916-927. b) Heinisch, T.; Ward, T., Design strategies for the creation of artificial metalloenzymes. *Current Opinion in Chemical Biology* **2010,** *14* (2), 184-199.

27. Davies, C.; Dux, E.; Duhme-Klair, A., Supramolecular interactions between functional metal complexes and proteins. *Dalton Transactions* **2009,** (46), 10141-10154.

28. a) Reetz, M. T.*Top. Organomet. Chem.***2009**, *25*, 63−92. b) Letondor, C.; Ward, T., Artificial metalloenzymes for enantioselective catalysis: Recent advances. *Chembiochem* **2006,** *7* (12), 1845-1852.

29. Collot, J.; Humbert, N.; Skander, M.; Klein, G.; Ward, T.R., *J. Organomet. Chem.* **2004**, *689*, 4868–4871.

30. Ilie, A.; Reetz, M., Directed Evolution of Artificial Metalloenzymes. *Israel Journal of Chemistry* **2015,** *55* (1), 51-60.

31. a) Jing, Q.; Kazlauskas, R., Regioselective Hydroformylation of Styrene Using Rhodium-Substituted Carbonic Anhydrase. *Chemcatchem* **2010,** *2* (8), 953-957. b) Popp, B.; Ball, Z., Structure-Selective Modification of Aromatic Side Chains with Dirhodium Metallopeptide Catalysts. *Journal of the American Chemical Society* **2010,** *132* (19), 6660-6662. c) Liu, X.; Yu, Y.; Hu, C.; Zhang, W.; Lu, Y.; Wang, J., Significant Increase of Oxidase Activity through the Genetic Incorporation of a Tyrosine-Histidine Cross-Link in a Myoglobin Model of Heme-Copper Oxidase. *Angewandte Chemie-International Edition* **2012,** *51* (18), 4312-4316.

32. a) Wilson, M.; Whitesides, G., Conversion Of a Protein To a Homogeneous Asymmetric Hydrogenation Catalyst by Site-Specific Modification with a Diphosphinerhodium(I) Moiety. *Journal of the American Chemical Society* **1978,** *100* (1), 306-307. b) Mao, J.; Ward, T., Artificial Metalloenzymes for Enantioselective Catalysis Based on the Biotin-Avidin Technology. *Chimia* **2008,** *62* (12), 956-961. c) Allard, M.; Dupont, C.; Robles, V.; Doucet, N.; Lledos, A.; Marechal, J.; Urvoas, A.; Mahy, J.; Ricoux, R., Incorporation of Manganese Complexes into Xylanase: New Artificial Metalloenzymes for Enantioselective Epoxidation. *Chembiochem* **2012,** *13* (2), 240-251.

33. a) Kaiser, E.; Lawrence, D., Chemical Mutation of Enzyme Active-Sites. *Science* **1984,** *226* (4674), 505-511. b) Reetz, M.; Peyralans, J.; Maichele, A.; Fu, Y.; Maywald, M., Directed Evolution of Hybrid Enzymes: Evolving Enantioselectivity of an Achiral Rh-complex Anchored to a Protein. *Chemical Communications* **2006,** *41*, 4318-4320. c) Deuss, P.; Popa, G.; Botting, C.; Laan, W.; Kamer, P., Highly Efficient and Site-Selective Phosphane Modification of Proteins through Hydrazone Linkage: Development of Artificial Metalloenzymes. *Angewandte Chemie-International Edition* **2010,** *49* (31), 5315-5317. d) den Heeten, R.; Munoz, B.; Popa, G.; Laan, W.; Kamer, P., Synthesis of Hybrid Transition-metalloproteins via Thiol-selective Covalent Anchoring of Rh-phosphine and Ru-phenanthroline Complexes. *Dalton Transactions* **2010,** *39* (36), 8477-8483.

34. Pordea, A.; Ward, T., Chemogenetic Protein Engineering: an Efficient Tool for the Optimization of Artificial Metalloenzymes. *Chemical Communications* **2008,** *36*, 4239-4249.

35. a) Groger, H.; Asano, Y.; Bornscheuer, U.; Ogawa, J., Development of Biocatalytic Processes in Japan and Germany: From Research Synergies to Industrial Applications. *Chemistry-an Asian Journal* **2012,** *7* (6), 1138-1153. b) Huisman, G.; Collier, S., On the Development of New Biocatalytic Processes for Practical Pharmaceutical Synthesis. *Current Opinion in Chemical Biology* **2013,** *17* (2), 284-292.

36. Reetz, M., Directed Evolution of Selective Enzymes and Hybrid Catalysts. *Tetrahedron* **2002,** *58* (32), 6595-6602.

37. a) Jewett, J. C.; Bertozzi, C. R., Cu-free Click Cycloaddition Reactions in Chemical Biology. *Chemical Society Reviews* **2010,** *39* (4), 1272-1279. b) Debets, M.; Van Berkel, S.; Dommerholt, J.; Dirks, A.; Rutjes, F.; Van Delft, F., Bioconjugation with Strained Alkenes and Alkynes. *Accounts of Chemical Research* **2011,** *44* (9), 805-815.

38. a) Xie, J.; Schultz, P., An Expanding Genetic Code. *Methods* **2005,** *36* (3), 227-238. b) Nguyen, D.; Lusic, H.; Neumann, H.; Kapadnis, P.; Deiters, A.; Chin, J., Genetic Encoding and Labeling of Aliphatic Azides and Alkynes in Recombinant Proteins via a Pyrrolysyl-tRNA Synthetase/tRNA(CUA) Pair and Click Chemistry. *Journal of the American Chemical Society* **2009,** *131* (25), 8720-8721. c) Wang, Q.; Parrish, A.; Wang, L., Expanding the Genetic Code for Biological Studies. *Chemistry & Biology* **2009,** *16* (3), 323-336.

39. Plass, T.; Milles, S.; Koehler, C.; Schultz, C.; Lemke, E., Genetically Encoded Copper-Free Click Chemistry. *Angewandte Chemie-International Edition* **2011,** *50* (17), 3878-3881.

40. a) Laughlin, S.; Baskin, J.; Amacher, S.; Bertozzi, C., In Vivo Imaging of Membrane-associated Glycans in Developing Zebrafish. *Science* **2008,** *320* (5876), 664-667. b) Ning, X.; Guo, J.; Wolfert, M.; Boons, G., Visualizing Metabolically Labeled Glycoconjugates of Living Cells by Copper-free and Fast Huisgen Cycloadditions. *Angewandte Chemie-International Edition* **2008,** *47* (12), 2253-2255. c) Thomas, J.; Cui, H.; North, P.; Hofer, T.; Rader, C.; Burke, T., Application of Strain-Promoted Azide-Alkyne Cycloaddition and Tetrazine Ligation to Targeted Fc-Drug Conjugates. *Bioconjugate Chemistry* **2012,** *23* (10), 2007-2013.

41. Bos, J.; Fusetti, F.; Driessen, A.; Roelfes, G., Enantioselective Artificial Metalloenzymes by Creation of a Novel Active Site at the Protein Dimer Interface. *Angewandte Chemie-International Edition* **2012,** *51* (30), 7472-7475.

42. a) Davies, C.; Dux, E.; Duhme-Klair, A., Supramolecular Interactions between Functional Metal Complexes and Proteins. *Dalton Transactions* **2009,** *46*, 10141-10154. b) Hadt, R.; Sun, N.; Marshall, N.; Hodgson, K.; Hedman, B.; Lu, Y.; Solomon, E., Spectroscopic and DFT Studies of Second-Sphere Variants of the Type 1 Copper Site in Azurin: Covalent and Nonlocal Electrostatic Contributions to Reduction Potentials. *Journal of the American Chemical Society* **2012,** *134* (40), 16701-16716.

43. Wierenga, R., The TIM-barrel fold: A Versatile Framework for Efficient Enzymes. *Febs Letters* **2001,** *492* (3), 193-198.

44. Reetz, M. T.; Rentzsch, M.; Pletsch, A.; Taglieber, A.; Hollmann, F.; Mondiere, R. J. G.; Dickmann, N.; Hocker, B.; Cerrone, S.; Haeger, M. C.; Sterner, R., A Robust Protein Host for Anchoring Chelating Ligands and Organocatalysts. *Chembiochem* **2008,** *9* (4), 552-564.

45. a) Lang, D.; Thoma, R.; Henn-Sax, M.; Sterner, R.; Wilmanns, M., Structural Evidence for Evolution of the β/α Barrel Scaffold by Gene Duplication and Fusion. *Science* **2000,** *289* (5484), 1546-1550. b) Douangamath, A.; Walker, M.; Beismann-Driemeyer, S.; Vega-Fernandez, M.; Sterner, R.; Wilmanns, M., Structural Evidence for Ammonia Tunneling across the $(\beta\alpha)_8$ Barrel of the Imidazole Glycerol Phosphate Synthase Bienzyme Complex. *Structure* **2002,** *10* (2), 185-193. c) Beismann-Driemeyer, S.; Sterner, R., Imidazole Glycerol Phosphate Synthase from Thermotoga maritima - Quaternary Structure, Steady-state Kinetics, and Reaction Mechanism of the Bienzyme Complex. *Journal of Biological Chemistry* **2001,** *276* (23), 20387-20396. d) Hocker, B.; Jurgens, C.; Wilmanns, M.; Sterner, R., Stability, Catalytic Versatility and Evolution of the $(\beta\alpha)_8$ Barrel fold. *Current Opinion in Biotechnology* **2001,** *12* (4), 376-381.

46. Carstensen, L.; Zoldak, G.; Schmid, F.; Sterner, R., Folding Mechanism of an Extremely Thermostable (beta alpha)(8)-Barrel Enzyme: A High Kinetic Barrier Protects the Protein from Denaturation. *Biochemistry* **2012,** *51* (16), 3420-3432.
47. Krueger, A.; Imperiali, B., Fluorescent Amino Acids: Modular Building Blocks for the Assembly of New Tools for Chemical Biology. *Chembiochem* **2013,** *14* (7), 788-799.

48. Ha, N.; Oh, B.; Shin, S.; Kim, H.; Oh, T.; Kim, Y.; Choi, K.; Oh, B., Crystal Structures of a Novel, Thermostable Phytase in Partially and Fully Calcium-loaded States. *Nature Structural Biology* **2000,** *7* (2), 147-153.

49. Dommerholt, J.; Schmidt, S.; Temming, R.; Hendriks, L.; Rutjes, F.; van Hest, J.; Lefeber, D.; Friedl, P.; van Delft, F., Readily Accessible Bicyclononynes for Bioorthogonal Labeling and Three-Dimensional Imaging of Living Cells. *Angewandte Chemie-International Edition* **2010,** *49* (49), 9422-9425.

50. a) Pierron, J.; Malan, C.; Creus, M.; Gradinaru, J.; Hafner, I.; Ivanova, A.; Sardo, A.; Ward, T., Artificial Metalloenzymes for Asymmetric Allylic Alkylation on the Basis of the Biotin-avidin Technology. *Angewandte Chemie-International Edition* **2008,** *47* (4), 701-705. b) Carey, J.; Ma, S.; Pfister, T.; Garner, D.; Kim, H.; Abramite, J.; Wang, Z.; Guo, Z.; Lu, Y., A Site-selective Dual Anchoring Strategy for Artificial Metalloprotein Design. *Journal of the American Chemical Society* **2004,** *126* (35), 10812-10813. c) Lo, C.; Ringenberg, M.; Gnandt, D.; Wilson, Y.; Ward, T., Artificial Metalloenzymes for Olefin Metathesis Based on the Biotin-(strept)avidin Technology. *Chemical Communications* **2011,** *47* (44), 12065-12067. d) Abe, S.; Hirata, K.; Ueno, T.; Morino, K.; Shimizu, N.; Yamamoto, M.; Takata, M.; Yashima, E.; Watanabe, Y., Polymerization of Phenylacetylene by Rhodium Complexes within a Discrete Space of apo-Ferritin. *Journal*

of the American Chemical Society **2009,** *131* (20), 6958-6960.

51. Davies, H.; Beckwith, R., Catalytic Enantioselective C-H Activation by Means of Metal-carbenoid-induced C-H Insertion. *Chemical Reviews* **2003,** *103* (8), 2861-2903.

52. a) Nicolas, I.; Le Maux, P.; Simonneaux, G., Asymmetric Catalytic Cyclopropanation Reactions in Water. *Coordination Chemistry Reviews* **2008,** *252* (5-7), 727-735. b) Wurz, R.; Charette, A., Transition Metal-catalyzed Cyclopropanation of Alkenes in Water: Catalyst Efficiency and in Situ Generation of the Diazo Reagent. *Organic Letters* **2002,** *4* (25), 4531-4533. c) Antos, J.; McFarland, J.; Iavarone, A.; Francis, M., Chemoselective Tryptophan Labeling with Rhodium Carbenoids at Mild pH. *Journal of the American Chemical Society* **2009,** *131* (17), 6301-6308. d) Tishinov, K.; Schmidt, K.; Haussinger, D.; Gillingham, D., Structure-Selective Catalytic Alkylation of DNA and RNA. *Angewandte Chemie-International Edition* **2012,** *51* (48), 12000-12004. e) Candeias, N.; Gois, P.; Afonso, C., Rh(II)-catalyzed Intramolecular C-H Insertion of Diazo Substrates in Water: Scope and Limitations. *Journal of Organic Chemistry* **2006,** *71* (15), 5489-5497.

53. Espino, C.; Fiori, K.; Kim, M.; Du Bois, J., Expanding the Scope of C-H Amination through Catalyst Design. *Journal of the American Chemical Society* **2004,** *126* (47), 15378-15379.

54. Lou, Y.; Remarchuk, T.; Corey, E., Catalysis of Enantioselective [2+1]-cycloaddition Reactions of Ethyl Diazoacetate and Terminal Acetylenes using Mixed-ligand Complexes of the Series $Rh_2(RCO_2)(n)$ (L-4-n(star)). Stereochemical Heuristics for Ligand Exchange and Catalyst Synthesis. *Journal of the American Chemical Society* **2005,** *127* (41), 14223-14230.

55. Sambasivan, R.; Ball, Z., Metallopeptides for Asymmetric Dirhodium Catalysis. *Journal of the American Chemical Society* **2010,** *132* (27), 9289-9291.

56. Suntharalingam, K.; White, A.; Vilar, R., Two Metals Are Better than One: Investigations on the Interactions between Dinuclear Metal Complexes and Quadruplex DNA. *Inorganic Chemistry* **2010,** *49* (18), 8371-8380.

57. Chen, H.; Tagore, R.; Das, S.; Incarvito, C.; Faller, J.; Crabtree, R.; Brudvig, G., General Synthesis of Di-mu-oxo Dimanganese Complexes as Functional Models for the Oxygen Evolving Complex of Photosystem II. *Inorganic Chemistry* **2005,** *44* (21), 7661-7670.

58. a) Das, S.; Incarvito, C.; Crabtree, R.; Brudvig, G., Molecular Recognition in the Selective Oxygenation of Saturated C-H Bonds by a Dimanganese Catalyst. *Science* **2006,** *312* (5782), 1941-1943. b) Kwong, H.; Lee, W., First Enantioselective Catalyst Based on a Chiral Terpyridine: Synthesis of New C-2-symmetric 2,2 ': 6 ',2 "-terpyridine Ligands and Copper-catalyzed Enantioselective Cyclopropanation of Alkenes. *Tetrahedron-Asymmetry* **2000,** *11* (11), 2299-2308.

59. a) Tracewell, C.; Arnold, F., Directed Enzyme Evolution: Climbing Fitness Peaks One

Amino Acid at a Time. *Current Opinion in Chemical Biology* **2009,** *13* (1), 3-9. b) Reetz, M.; Rentzsch, M.; Pletsch, A.; Maywald, M.; Maiwald, P.; Peyralans, J.; Maichele, A.; Fu, Y.; Jiao, N.; Hollmann, F.; Mondiere, R.; Taglieber, A., Directed Evolution of Enantioselective Hybrid Catalysts: a Novel Concept in Asymmetric Catalysis. *Tetrahedron* **2007,** *63* (28), 6404-6414.

60. Roy, R.; Hohng, S.; Ha, T., A Practical Guide to Single-molecule FRET. *Nature Methods* **2008,** *5* (6), 507-516.

61. Lakowicz, J. R., in Principles of Fluorescence Spectroscopy, 3rd ed., Springer, New York, 2006, pp. 443 –472.

62. Stephanopoulos, N.; Francis, M., Choosing an Effective Protein Bioconjugation Strategy. *Nature Chemical Biology* **2011,** *7* (12), 876-884.

63. Constable, E.; Ward, M., Synthesis And Coordination Behavior of 6',6''-Bis(2-Pyridyl)-2,2'-4,4''-2'',2'''-Quaterpyridine-Back-To-Back 2,2'-6',2''-Terpyridine. *Journal of the Chemical Society-Dalton Transactions* **1990,** (4), 1405-1409.

64. Young, T.; Ahmad, I.; Yin, J.; Schultz, P., An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. *Journal of Molecular Biology* **2010,** *395* (2), 361-374.

65. Heckman, K.; Pease, L., Gene Splicing and Mutagenesis by PCR-driven Overlap Extension. *Nature Protocols* **2007,** *2* (4), 924-932.

66. Sambrook, J.; Fritsch, E. F.; Maniatis, T., Molecular cloning: A laboratory manual, Cold Spring Harbor Laboratory Press: 4th edition, **2012**, Cold Spring Harbor, N.Y.

67. Murphy, J.; Liao, X.; Hartwig, J., Meta Halogenation of 1,3-disubstituted Arenes via Iridium-catalyzed Arene Borylation. *Journal of the American Chemical Society* **2007,** *129* (50), 15434-15435.

68. The extinction coefficient of tHisF was determined according to the Lambert-Beer law by using the concentration obtained from a Pierce® BCA Protein Assay Kit.

69. Guo, Q.; Jureller, J.; Warren, J.; Solomaha, E.; Florian, J.; Tang, W., Protein-protein Docking and Analysis Reveal that Two Homologous Bacterial Adenylyl Cyclase Toxins Interact with Calmodulin Differently. *Journal of Biological Chemistry* **2008,** *283* (35), 23836-23845.

70. Rosenberg, M.; Krivokapic, A.; Tilset, M., Highly cis-Selective Cyclopropanations with Ethyl Diazoacetate Using a Novel Rh(I) Catalyst with a Chelating N-Heterocyclic Iminocarbene Ligand. *Organic Letters* **2009,** *11* (3), 547-550.

71. Chen, Y.; Zhang, X., Asymmetric Cyclopropanation of Styrenes Catalyzed by Metal Complexes of D-2-Symmetrical Chiral Porphyrin: Superiority of Cobalt over Iron. *Journal*

of Organic Chemistry **2007,** *72* (15), 5931-5934.

72. Dakin, L.; Ong, P.; Panek, J.; Staples, R.; Stavropoulos, P., Speciation and Mechanistic Studies of Chiral Copper(I) Schiff Base Precursors Mediating Asymmetric Carbenoid Insertion Reactions of Diazoacetates into the Si-H Bond of Silanes. *Organometallics* **2000,** *19* (15), 2896-2908.

73. Jones, J.; Kresge, A., Methylthio Group Migration in the Acid-Catalyzed Hydrolysis of S-Methyl Phenyldiazothioacetate - Kinetics and Mechanism of the Reaction. *Journal of Organic Chemistry* **1993,** *58* (10), 2658-2662.

CHAPTER II

STRUCTURE-GUIDED ENGINEERING OF AN ARTIFICIAL METALLOENZYME FOR ENANTIOSELECTIVE CYCLOPROPANATION

Most of the work described in this chapter is published (Srivastava et al., *Nat. Commun.* **2015**, *6*, 7789-7798). I conducted all chemical synthesis, biocatalysis, and bioconjugation optimization, and the final biocatalysis experiments. Dr. Srivastava conducted all cloning and initial protein expression and biocatalysis experiments. Guardiola KE identified the POP scaffold and carried out all spectroscopic characterizations.

ABSTRACT

In Chapter I, we developed a SPAAC-based bioconjugation method for rapid ArM formation and demonstrated the scope of this method with respect to both the scaffold and the cofactor. However, no selectivity was observed in reactions catalyzed by these ArM systems. In this work, we covalently link an alkyne-substituted dirhodium catalyst to a prolyl oligopeptidase containing a genetically encoded L-4-azidophenylalanine residue to create an ArM that catalyzes olefin cyclopropanation. Scaffold mutagenesis is then used to improve the enantioselectivity of this reaction, and cyclopropanation of a range of styrenes and donor–acceptor carbene precursors is accepted. The ArM reduces the formation of byproducts, including those resulting from the reaction of dirhodium–carbene intermediates with water. This shows that an ArM can improve the substrate specificity of a catalyst and, for the first time, the water tolerance of a metal-catalyzed

reaction. Given the diversity of reactions catalyzed by dirhodium complexes, we anticipate that dirhodium ArMs will provide many unique opportunities for selective catalysis.

INTRODUCTION

New approaches to control the selectivity and specificity of catalysts remain the subject of intense academic and industrial research because of the importance of selective catalysis for the synthesis of chemicals ranging from fuels to pharmaceuticals[1]. Weak interactions between substrates and catalysts imparted by functional groups distal to catalyst-active sites[2, 3] and supramolecular catalyst scaffolds[4, 5] are increasingly used to improve catalyst selectivity. Of course, such features are ubiquitous in enzymes[6] and contribute to their often stunning activities and selectivities. To exploit the substrate-binding and activation capabilities of enzymes for reactions and catalysts not found in nature, researchers have developed a range of methods to link synthetic catalysts and protein scaffolds to create artificial metalloenzymes (ArMs)[1, 7]. These efforts have culminated in ArMs for enantioselective, regioselective and chemoselective reactions, but, despite several notable examples[8–13], engineering scaffolds to further improve these parameters remains challenging[14]. The majority of successful optimization efforts exploit the binding of biotinylated metal cofactors to (strept)avidin[10]; therefore, the development of new scaffolds capable of imparting high levels of selectivity metal catalysts could significantly expand the scope of ArM catalysis[9]. Furthermore, the ArM-catalysed reactions explored to date rarely involve catalytic intermediates that can react irreversibly with water in a deleterious manner, and no examples have been reported in which an ArM can mitigate this undesired reactivity[7].

Chapter I outlined a new method for ArM formation via strain-promoted azide–alkyne cycloaddition (SPAAC) of bicyclo[6.1.0]nonyne (BCN)-substituted cofactors and scaffold

proteins containing a genetically encoded L-4-azidophenylalanine (Z) residue (Fig. 2.1a)[15]. Unlike non-covalent methods for ArM formation, this approach allows the use of any desired protein as a scaffold, and, unlike most covalent methods, the bioorthogonality of SPAAC eliminates the need to remove residues (for example, cysteine) in the scaffold that might react with electrophiles used in conventional bioconjugation methods (for example, maleimides)[7]. ArM formation from various cofactors, including the Esp-based[16] dirhodium cofactor **1** (Fig. 2.1b), was demonstrated with a range of protein scaffolds, but no selectivity was observed in reactions catalyzed by these systems. We attributed this lack of selectivity to the inability of the protein scaffolds selected for bioconjugation method development to fully encapsulate the cofactors selected for catalysis. Given the broad range of reactions catalyzed by dirhodium complexes (Fig. 2.1c), including cyclopropanation and X–H insertion (X = C, N, O, and so on)[17], and the selectivity challenges that persist for many of these reactions[18], we sought to identify a scaffold protein that could impart selectivity to **1**. This would validate our hypothesis regarding the poor selectivity of our initial ArM designs, illustrate the importance of scaffold selection in ArM design, and provide a platform for the development of future ArMs using different metal cofactors.

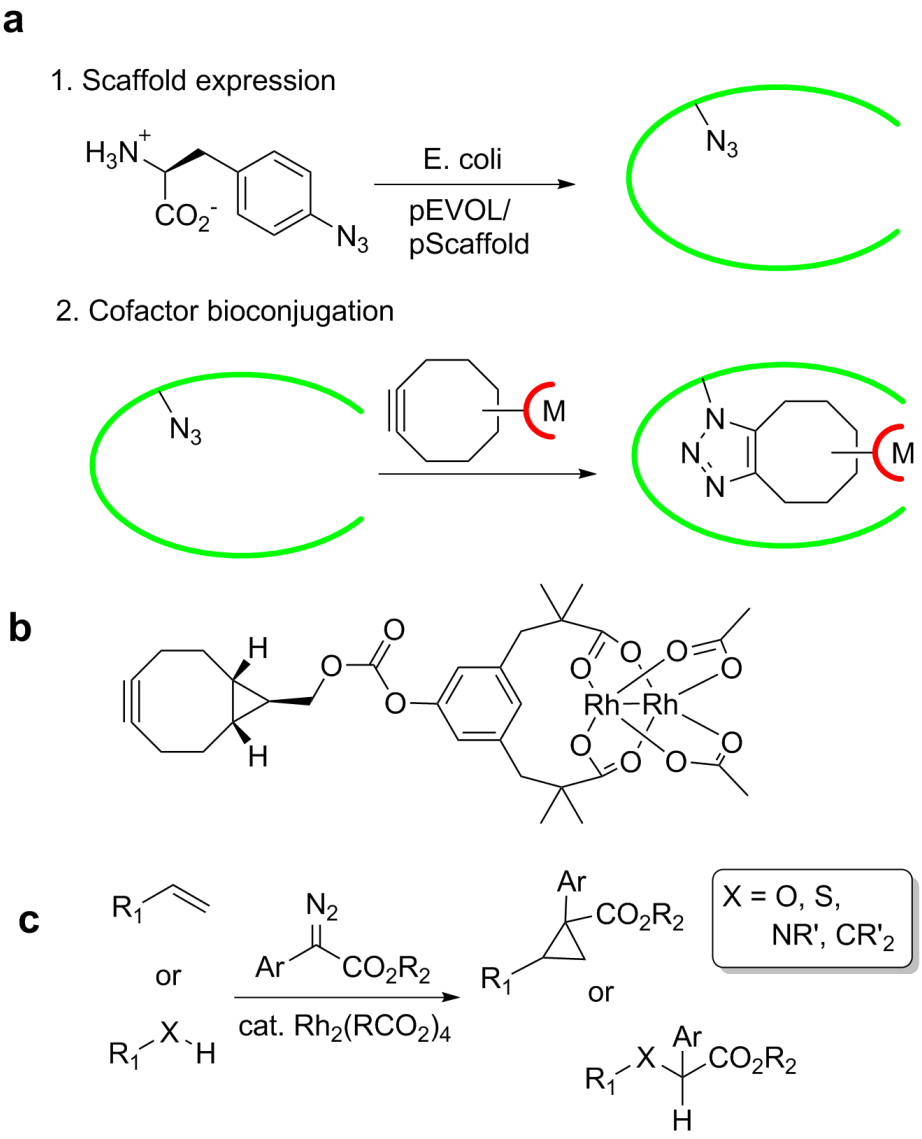
Here we show that a prolyl oligopeptidase (POP) scaffold can be used to generate dirhodium ArMs that catalyse asymmetric cyclopropanation. Genetic optimization of these ArMs led to high levels of enantioselectivity and reduced levels of byproducts resulting from the reaction of catalytic intermediates with water.

RESULTS

Scaffold selection and bioconjugation

An extensive search of different protein X-ray structures in the protein data bank (PDB) led to the identification of several members of the prolyl oligopeptidase family as potential ArM scaffolds. Similar to tHisF protein discussed in Chapter I, these proteins possess roughly cylindrical shapes

Figure 2.1. ArM formation and reactivity.(a) ArM formation using the SPAAC reaction. (b) Structure of cofactor1.(c) Representative reactions catalyzed by dirhodium complexes.



(30 X 60 Å) and large internal volumes (5–8 X 10^3 $Å^3$) for cofactor enclosure[19]. This family includes POPs, dipeptidyl peptidases IV, oligopeptidases B and acylaminoacyl peptidases. All of

these enzymes share a common fold comprising an α/β hydrolase domain, which contains a Ser-Asp-His triad for amide bond hydrolysis, capped by a β-barrel domain. We initially selected a POP from *Pyrococcus furiosus* (*Pfu*) as a scaffold for ArM formation because of its exceptional thermal stability[20]. It is worth noting that a BLAST analysis of the *Pfu* POP protein sequence revealed >8000 POP family enzymes, >200 of which come from halo/thermophilic organisms that could also serve as ArM scaffolds. Despite the abundance of POP structures in the PDB, however, the structure of *Pfu* POP was not solved yet when our study initiated[1]; therefore, a previously reported homology model[21] of this enzyme was used for initial engineering efforts (Fig. 2.2). An amber codon was introduced into the POP gene to replace the catalytically active serine (S477) with a Z residue (Z477), abolish the native proteolytic activity of the enzyme and position the cofactor centrally within the active site. A POP gene whose codon usage was optimized for expression in *E. coli* was used as a template for genetic manipulation, and the resulting scaffold, POP-Z, was expressed in high yield (>100 versus ~10 mg/L before codon optimization) with essentially quantitative Z incorporation. Unfortunately, however, no reaction occurred between POP-Z and **1**. POP variants in which other active site residues had been replaced with Z proved similarly unreactive towards **1**, but rapid reaction of surface-exposed Z residues was observed[22].

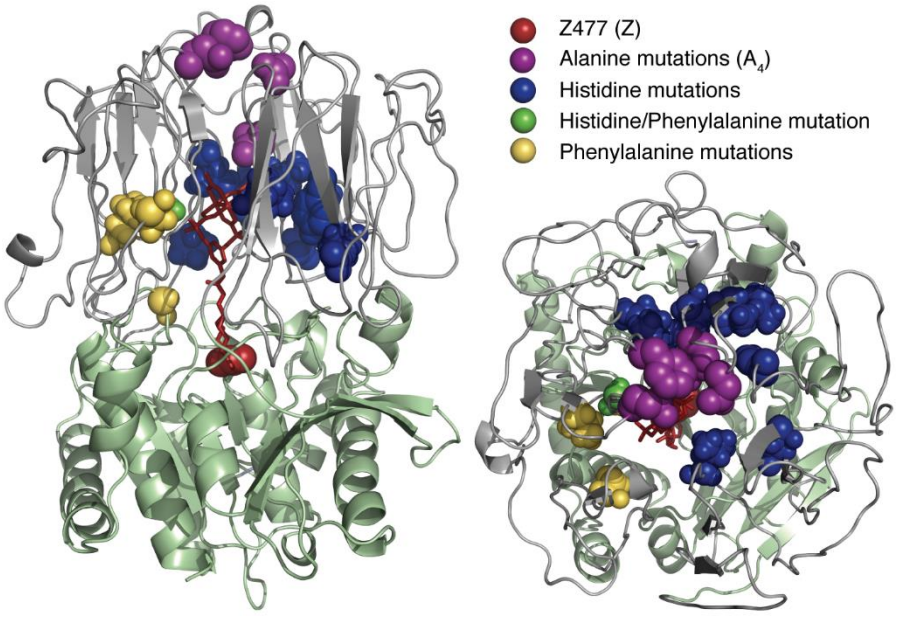
POP family enzymes have been crystallized in open and closed conformations[23, 24] and are proposed to sample both conformations during catalysis[25]. Active site residues, including Z477, should be accessible for bioconjugation in the open conformation. We hypothesized that the lack of POP-Z bioconjugation resulted from the enzyme existing predominantly in the closed conformation under the reaction conditions explored[26] and that **1** is too large to enter the POP-active site in this conformation. Because the closed conformation of POP possesses the cylindrical

[1] The crystal structures of *Pfu* POP wide-type mutant and a cysteine mutant S477C have been recently solved by K. E. Guardiola in the group, who is currently working on the structure of *Pfu* POP azide mutant.

shape and solvent-sequestered active site that we hoped to exploit for ArM catalysis, this indicated that POP-Z modification would be required for bioconjugation.

Early proposals for the substrate specificity of POP, which acts only on short peptides (<30 residues), invoked the entry of these substrates through a small pore at the end of the b-barrel domain where the β-sheets comprising this domain converge.[27] More recent studies have concluded that substrates do not enter via this pore and that it does not appear to be relevant to POP protease activity,[24, 28] but we envisioned that this pore could be coopted for ArM formation. Examining the pore structure of *Pfu* POP in the aforementioned homology model[21] suggested that four residues (E104, F146, K199, and D202) could block access to the active site (Fig. 2.2). We mutated these residues in POP-Z to alanine, and the resulting protein, POP-ZA_4, underwent rapid bioconjugation in the presence of cofactor **1** at 4 °C to form POP-ZA_4-**1**. The simplest explanation for this result is that the A_4 mutations expand the pore to enable cofactor access to the POP active site. It may also be that these mutations facilitate conformational changes that enable domain opening[24], and subsequent experiments will be required to differentiate these mechanisms. Interestingly, the crystal structure of wild-type POP recently resolved by Guardiola KE in our lab suggests that the homology model is significantly different from the real structure and only two of the A_4 mutations (K199 and D202) are in positions that may affect cofactor entry to the active site. These two mutations were then introduced by Guardiola KE to make a POP-ZA_2 mutant that showed rapid bioconjugation as well. Regardless of the mechanism by which the A_4 (or A_2) mutations enable bioconjugation of POP-Z, the success of this strategy highlights the potential for mutagenesis to allow the use of otherwise unreactive proteins as ArM scaffolds.

Figure 2.2. Homology model[21] of *Pfu* POP and mutations introduced to improve ArMs

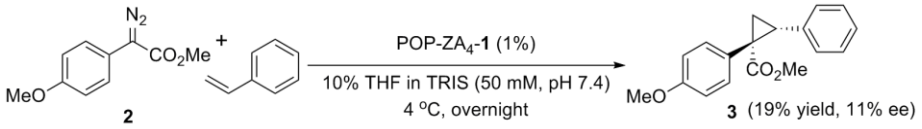


(The hydrolase domain is shown in green, the propeller domain is shown in grey and cofactor **1** linked at Z477 is shown in red. Sites of different mutations introduced into *Pfu* POP are shown as colored spheres)

ArM catalysis and optimization of reaction conditions

Due to variations in the extent of bioconjugation observed for different POP variants[15], ArM concentration was determined by multiplying the total protein concentration in purified ArM/scaffold mixtures by the ratio of the high resolution ESI-MS peak intensities of the ArM and scaffold in these mixtures. In this way, consistent dirhodium loadings were used regardless of the extent of bioconjugation. After successful construction of POP-based dirhodium ArM through protein engineering, our focus became investigation of catalytic functions of the formed ArM. The cyclopropanation of styrene with donor–acceptor diazo **2** as a model reaction (Fig. 1.3), and cyclopropane **3** was formed as a single diastereomer in 19% yield and 11% ee. This enantioselectivity, while low, showed that the POP scaffold could impart selectivity to cofactor **1**[16], unlike previously described scaffolds[15].

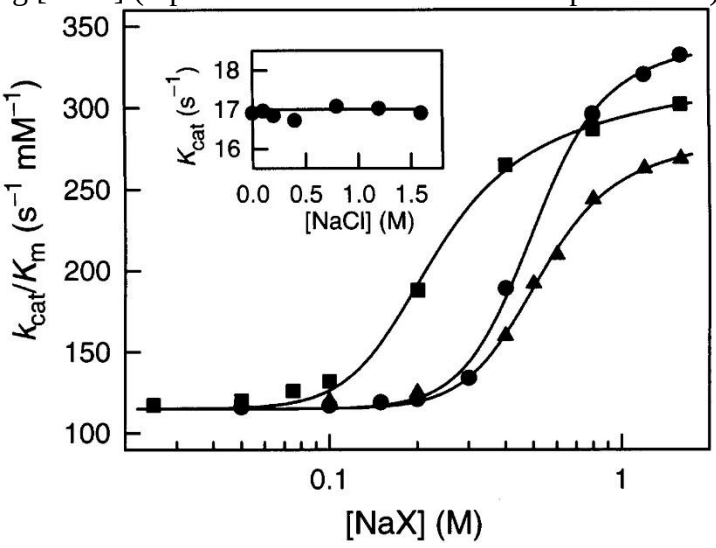
Figure 2.3. Initial result of cyclopropanation catalyzed by POP-Rh ArM



To identify other potential linkage sites within the active site in addition to Az477, A number of sites (214, 251, 326, and 401) were mutated to Z residue and examined in bioconjugation and biocatalysis. While POP-A_4-Z251 and POP-A_4-Z326 were unreactive toward cofactor **1**, POP-A_4-Z214 and POP-A_4-Z401 formed ArM successfully and provided 8% and 4% e.e. respectively under the same reaction conditions. Since none of these mutants provided a better selectivity, we used Z477 as the linkage site in further optimization.

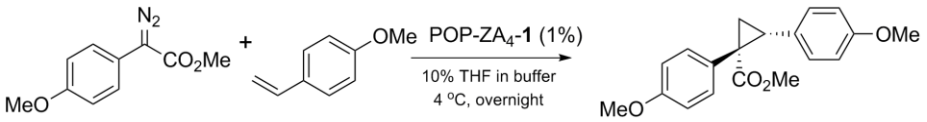
A range of reaction parameters were systematically explored to improve the enantioselectivity. According to a detailed kinetic and mechanistic study of *Pfu* POP conducted by Harwood et al., the kinetic parameters (K_{cat} and K_{cat}/K_m) for its native activity are dependent on choice of substrate, pH, reaction temperature, and halide binding[21]. The effect of ionic strength on catalytic rate for *Pfu* POP is particularly interesting. The hydrolytic rate of a model substrate catalyzed by *Pfu* POP is activated by halide salts such as NaF, NaCl, and NaBr; the plot of K_{cat}/K_m versus salt concentration displays a sigmoidal pattern for the binding of NaF, NaCl and NaBr (Fig. 1.4), which suggests that there are multiple binding sites for the halide ions, reminiscent of oxygen binding in hemoglobin. By fitting experimental data into kinetic equations, it is found that there are five "non-productive" binding sites (that have no effect on enzyme activity) for Cl^-, and two for F^- and Br^-. An activation mechanism with more than one halide binding site was proposed to explain the halide activation, and the sequential binding observed for halide ions suggests that a conformational change might occur at the hinge region between the two domains, allowing substrate access to the active site.

Figure 2.4. Plot of k_{cat}/K_m versus [NaF] (▪), [NaCl] (●), and [NaBr] (▲). The fittings were obtained by non-linear regression analysis. The inset is a plot of k_{cat} versus [NaCl], showing that k_{cat} remains unchanged with increasing [NaCl] (reprinted from reference 21 with permission).



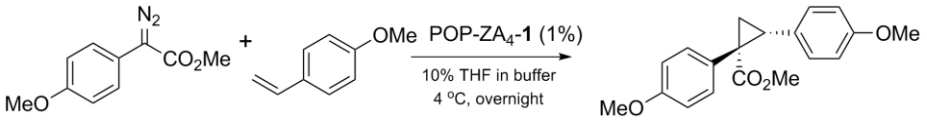
The above findings prompted us to study the effect of halide salts on POP-ZA_4-**1** using cyclopropanation of 4-methoxy phenyldiazoacetate and 4-methoxystyrene as a model reaction[29]. Initially, different concentrations of NaCl or NaBr were added to the reaction buffer (Table 2.1). It was observed that higher salt concentration led to significantly increased enantioselectivity, and at the same concentration NaBr produced a slightly larger improvement. Thus, a broader range of salt conditions involving NaF and NaI were examined (Table 2.2). Unlike NaBr/NaCl, NaF had a negligible effect on enantioselectivity, and highly concentrated NaI provided a decreased selectivity. A scrutiny of the influence of NaBr at a wider concentration range revealed that overly concentrated NaBr did not produce a further improvement and the optimal concentration was around 1.75 M. The mechanism by which these halide salts have an impact on POP-ZA_4-**1** catalyzed cyclopropanation is not clear, and further studies are undergoing in the lab to rationalize these improvements.

Table 2.1. Optimization of reaction conditions (NaBr, NaCl)



entry	buffer	halide salt	e.e. (%)
1	50 mM Tris (pH 7.4)	-	16.4
2	50 mM Tris (pH 7.4)	0.1 M NaCl	17.3
3	50 mM Tris (pH 7.4)	0.5 M NaCl	18.2
4	50 mM Tris (pH 7.4)	1.0 M NaCl	26.2
5	50 mM Tris (pH 7.4)	0.1 M NaBr	20.3
6	50 mM Tris (pH 7.4)	0.5 M NaBr	30.8
7	50 mM Tris (pH 7.4)	1.0 M NaBr	37.1

Table 2.2. Optimization of reaction conditions (NaX)

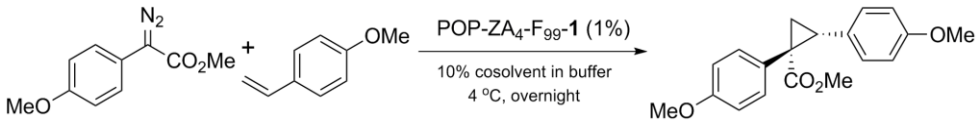


entry	buffer	halide salt	e.e. (%)
1	50 mM Tris (pH 7.4)	-	16.4
2	50 mM Tris (pH 7.4)	0.1 M NaBr	18.8
3	50 mM Tris (pH 7.4)	0.5 M NaBr	26.2
4	50 mM Tris (pH 7.4)	1.0 M NaBr	28.9
5	50 mM Tris (pH 7.4)	1.5 M NaBr	30.5
6	50 mM Tris (pH 7.4)	1.75 M NaBr	38.0
7	50 mM Tris (pH 7.4)	2.0 M NaBr	34.7
8	50 mM Tris (pH 7.4)	0.1 M NaF	14.2
9	50 mM Tris (pH 7.4)	0.5 M NaF	16.5
10	50 mM Tris (pH 7.4)	1.0 M NaF	15.4
11	50 mM Tris (pH 7.4)	0.1 M NaI	14.1
12	50 mM Tris (pH 7.4)	0.5 M NaI	14.5
13	50 mM Tris (pH 7.4)	1.0 M NaI	8.2

Organic solvent is often required in ArM catalysis and may influence catalysis result by changing enzyme stability and conformation. A set of common water-miscible solvents were examined in ArM-catalyzed cyclopropanation (Table 2.3), and it was observed that catalysis

conducted with THF provided the highest enantioselectivity, and DMSO conformed the largest decrease in selectivity.

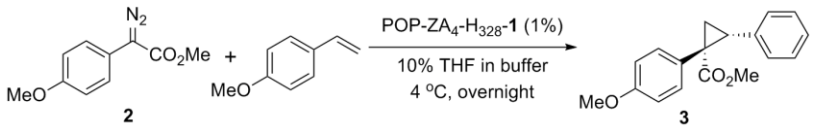
Table 2.3. Optimization of reaction conditions (cosolvent)



entry	cosolvent	buffer	e.e. (%)
1	THF	50 mM Tris (pH 7.4), 1.0 M NaBr	53.0
2	CH_3CN	50 mM Tris (pH 7.4), 1.0 M NaBr	30.5
3	DMSO	50 mM Tris (pH 7.4), 1.0 M NaBr	4.0
4	2,2,2-trifluoroethanol	50 mM Tris (pH 7.4), 1.0 M NaBr	31.6
5	isopropanol	50 mM Tris (pH 7.4), 1.0 M NaBr	29.8

Further optimization of reaction parameters was continued on a different mutant POP-ZA_4-H328-**1**, which will be discussed later. The original *Pfu* POP has an interesting double-sigmoidal pH-rate profile in its native function[21], which suggests pH may also affect the cyclopropanation. Considering one of the substrate diazoacetate is sensitive to acidic conditions, only pH conditions higher than 7 were explored (Table 2.4, entries 1~4). It was observed that higher pH caused a decrease in enantioselectivity, and the optimal pH was around 7.4. A few buffers were tested as well in cyclopropanation (Table 2.4, entries 5~7), and PIPES buffer provided a slight increase in enantioselectivity compared to other tested buffers.

Table 2.4. Optimization of reaction conditions (pH and buffer)



entry	pH	buffer	e.e. (%)
1	7.4	50 mM Tris, 1.0 M NaBr	71.7
2	8.0	50 mM Tris, 1.0 M NaBr	65.1
3	8.5	50 mM Tris, 1.0 M NaBr	56.2

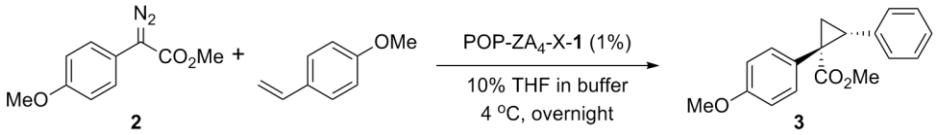
Table 2.4. Optimization of reaction conditions (pH and buffer), continued

entry	pH	buffer	e.e. (%)
4	9.0	50 mM Tris, 1.0 M NaBr	44.4
5	7.4	50 mM NaH_2PO_4, 1.0 M NaBr	75.4
6	7.4	50 mM BIS-Tris, 1.0 M NaBr	72.7
7	7.4	50 mM PIPES, 1.0 M NaBr	76.2

ArM catalysis and structural design

After development of optimal reaction conditions for cyclopropanation, we focused on structure design strategies to further improve enantioselectivity. Our first strategy involved strategic introduction of steric bulky substituents around the metal center to improve selectivity by disfavoring one possible transition-state conformation. Based on this concept, we looked for sites within the active-site cavity of *Pfu* POP which may project toward the rhodium center and interact with substrates and mutated them to sterically bulky residues such as phenylalanine or tryptophan. Four mutations (G99F, Y251W, A594W, and A594F) were separately introduced to investigate the effect of bulky residues (Table 2.5), and all the corresponding single mutants led to marked increase in enantioselectivity. Next, these beneficial mutations were combined into doubly mutants, which however did not provide a further boost in selectivity (data not shown).

Table 2.5. Effect of bulky mutations on selectivity



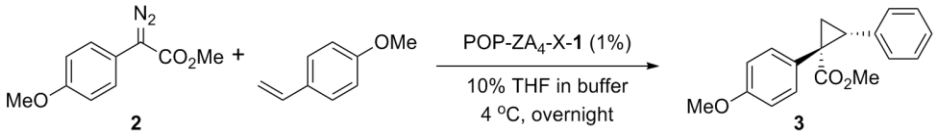
entry	POP mutant	buffer	e.e. (%)
1	F99	50 mM Tris (pH 7.4), 1.0 M NaBr	61.2
2	W594	50 mM Tris (pH 7.4), 1.0 M NaBr	58.2
3	F594	50 mM Tris (pH 7.4), 1.0 M NaBr	61.3
4	W251	50 mM Tris (pH 7.4), 1.0 M NaBr	63.2

Our second strategy toward improving enantioselectivity was inspired by precedent work in the field of metalloenzyme catalysis. Several researchers have proposed that free cofactor movement within protein scaffolds can reduce ArM selectivity. Various methods have been developed to circumvent this problem by restricting conformational freedom through additional binding forces. Lu and coworkers[30] reported in an artificial metalloenzyme composed of a manganese salen complex and an apo sperm whale myoglobin, a site-selective two-point covalent attachment affects catalytic sulfoxidation of thioanisole with much higher ee and rate than either noncovalent or single-point covalent attachment strategies. This method was recapitulated by Ward and coworkers[31] in their study of artificial transfer hydrogenase based on biotin-streptavidin technology. It was shown that coordination of the metal cofactor to a suitably positioned histidine residue has a significant impact on the catalyst's performance, both in terms of activity and of selectivity. A more relevant example from Ball and coworkers[32] focused on development of peptide-based dirhodium catalysts. They found that the introduction of an axial-binding histidine residue on the peptide ligand led to dramatically increased enantioselectivity in cyclopropanation with α-diazophenylacetate. In all the above examples, extra binding site to the metal complex were generated by modifying protein/peptide scaffold.

We pursued the same strategy as Ball's example to improve POP-ZA_4-**1**, given the established success of this method in peptide-based dirhodium catalysts[33]. Based on the homology model of *Pfu* POP, histidine mutations were individually introduced at several residues within POP-ZA_4 that projected towards the POP active site cavity (G99, P139, I141, I197, T209, E218, V219, Y251, E283 and L328, see Fig. 2.2), and the enantioselectivity of the resulting ArMs was examined (Table 2.6). Of these, P139H, I197H and G99H showed slight improvement in enantioselectivity in the corresponding ArM compared to POP-ZA_4-**1**; the largest increase was

observed in L328H. We hypothesize that histidine coordination to the proximal Rh of cofactor **1** projects the distal Rh towards a specific region of the POP-active site and that the improved enantioselectivity of POP-ZA$_4$-L328H-**1** results from the ability of residues near the distal rhodium atom to impart selectivity to cyclopropanation reactions occurring at this center.

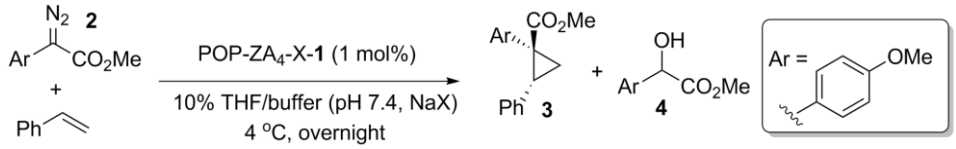
Table 2.6. Histidine mutants in ArM-catalyzed cyclopropanation



entry	POP mutant	buffer	e.e. (%)
1	I141H	50 mM Tris (pH 7.4), 1.0 M NaBr	30.1
2	E218H	50 mM Tris (pH 7.4), 1.0 M NaBr	33.1
3	V219H	50 mM Tris (pH 7.4), 1.0 M NaBr	13.2
4	P139H	50 mM Tris (pH 7.4), 1.0 M NaBr	43.1
5	T209H	50 mM Tris (pH 7.4), 1.0 M NaBr	30.5
6	I197H	50 mM Tris (pH 7.4), 1.0 M NaBr	40.2
7	G99H	50 mM Tris (pH 7.4), 1.0 M NaBr	42.1
8	E283H	50 mM Tris (pH 7.4), 1.0 M NaBr	28.0
9	L328H	50 mM Tris (pH 7.4), 1.0 M NaBr	56.6
10	Y251H	50 mM Tris (pH 7.4), 1.0 M NaBr	26.8

With the aim of further improving the selectivity of this ArM, we mutated to phenylalanine several residues (64, 97, 99 and 594, see Fig. 2.2) near and projecting towards the putative location of the distal Rh (Table 2.7, entries 9–14). We tried to combine the beneficial bulky mutations described above (and two new sites including 64 and 97) with L328H. While only F99 improved enantioselectivity significantly, the F99/F97 and F99/F594 double mutants provided modest further improvements. The synergistic combination ultimately led to cyclopropanation with 92% ee using POP-ZA$_4$-HFF-**1**(Table 2.7, entry 14).

Table 2.7. Summary of reaction condition optimization and active site mutations



Entry	POP Mutant (X)	Conditions	Yield (%)	e.e. (%)	**3**/**4**
1	L328 (WT)	TRIS, 0.1M NaCl	19	11	0.6
2	L328 (WT)	TRIS, 0.1M NaBr	23	18	0.6
3	L328 (WT)	TRIS, 1.75M NaBr	29	38	0.7
4	L328 (WT)	PIPES, 1.75M NaBr	25	38	0.6
5	F328	PIPES, 1.75M NaBr	14	23	0.5
6	C328	PIPES, 1.75M NaBr	24	47	0.5
7	M328	PIPES, 1.75M NaBr	33	68	0.7
8	H328	PIPES, 1.75 M NaBr	61	85	1.6
9	H328-F64	PIPES, 1.75 M NaBr	36	67	0.9
10	H328-F97	PIPES, 1.75 M NaBr	43	82	1.3
11	H328-F99	PIPES, 1.75 M NaBr	55	89	2.1
12	H328-F594	PIPES, 1.75 M NaBr	50	80	1.3
13	H328-F99-F97	PIPES, 1.75 M NaBr	73	91	2.3
14	H328-F99-F594	PIPES, 1.75 M NaBr	74	92	2.4
15	n/a, **5** (Fig. 2.3B)	PIPES, 1.75 M NaBr	12	0	0.4

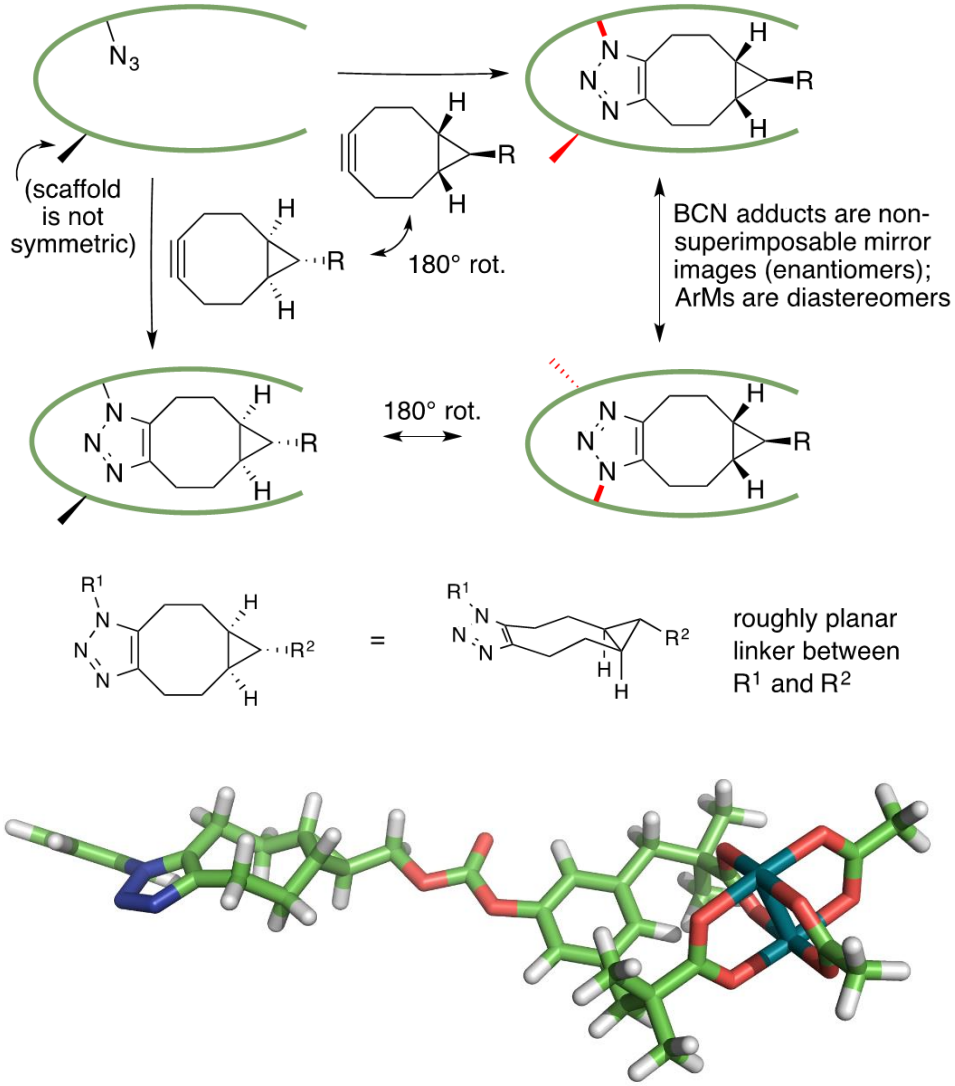
[a]All reactions conducted using 4 mM **2** and 20 mM styrene. Yield and enantioselectivity determined by HPLC relative to internal standard.

ArM selectivity and specificity

Several aspects of the activities exhibited by ArMs in the POP-ZA$_4$-HFF-**1** lineage deserve comment. First, the mechanism of asymmetric induction in our ArM system is worth discussion. The majority of ArMs capable of selective catalysis developed to date involve either chiral-at-metal complexes or complexes with flexible ligands. For example, Ward and his coworkers have focused on biotin-substituted cofactors derived from either fluxional, bidentate bisphosphine−Rh(I) complexes or racemic, readily racemized, chiral-at-metal d6 transition metal piano stool complexes[7]. In the former case, a relay of chirality from the scaffold to the bisphosphine ligand to generate a chiral Pd(II)−bisphosphine complex was used to explain the enantioselectivity. In the latter case, crystal structures for ArMs show only a single cofactor enantiomer, which suggests

that the scaffold assists even resolution of stereogenic metal centers. In these examples, a relay of chirality from a protein scaffold to metal complexes could contribute to enantioselectivity[10].

Figure 2.5. Explanation of bioconjugation stereochemistry and a geometry optimized (DFT, B3LYP, LANL2DZ) structure of a phenylazide-**1** adduct.



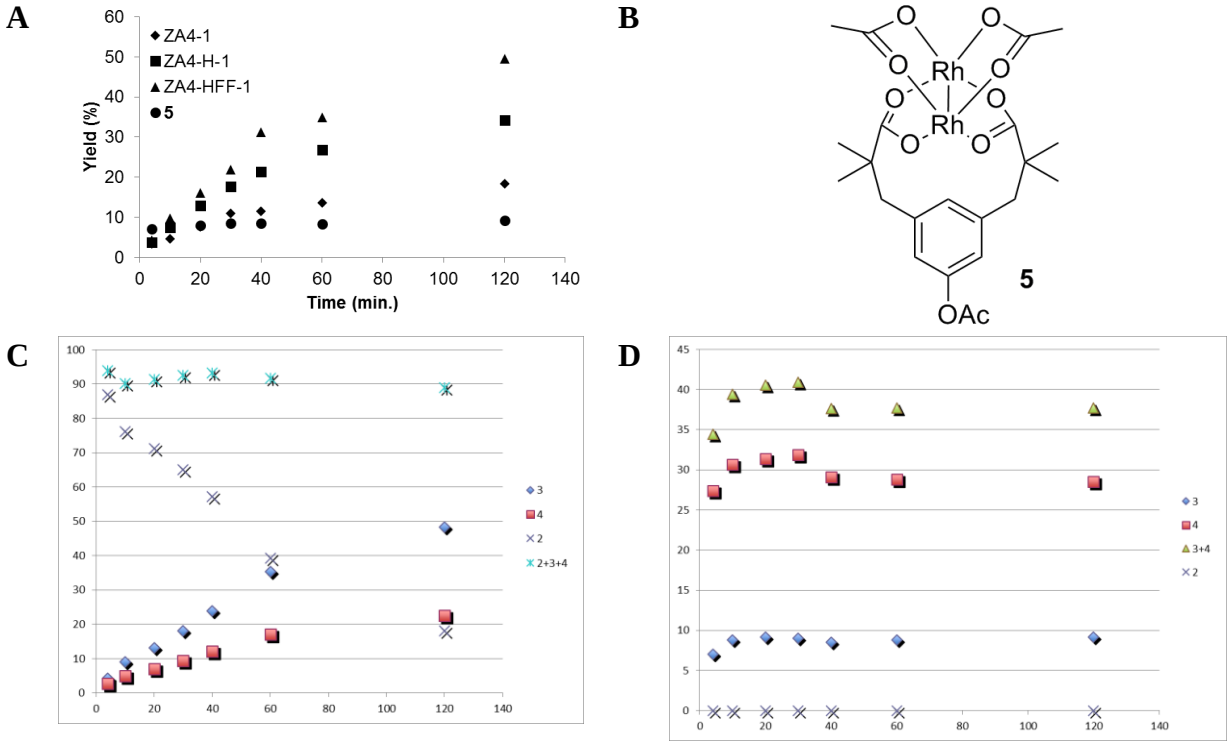
This mechanism of asymmetric induction seems unlikely for POP-ZA$_4$-HFF-**1**, given the rigidity of **1**. The observed selectivity is more consistent with direct interactions between active site residues, substrates and catalytic intermediates[9, 33], which suggests that the POP scaffold could be used to impart selectivity to a wide range of metal complexes. It is also interesting to note that

high selectivity is achieved despite the C_s-symmetry of the BCN moiety in **1**, which would lead to enantiomeric cycloadducts on reaction with the Z residue in POP-ZA_4-HFF-**1** (Fig. 2.5). The extended conformation of the exo BCN diastereomer and the distance between the BCN moiety and the dirhodium center in **1** could render structural differences between these enantiomers small enough that they have only a minor impact on cofactor position in the POP active site. On the other hand, POP could impart enantioselectivity to the stoichiometric cycloaddition, in which case mutations introduced to improve the enantioselectivity of ArM catalyzed cyclopropanation could have done so by improving cycloaddition enantioselectivity and thus ArM diastereopurity (Fig. 2.5). Structural studies of the POP ArMs described in this work are underway and could shed light on these possibilities.

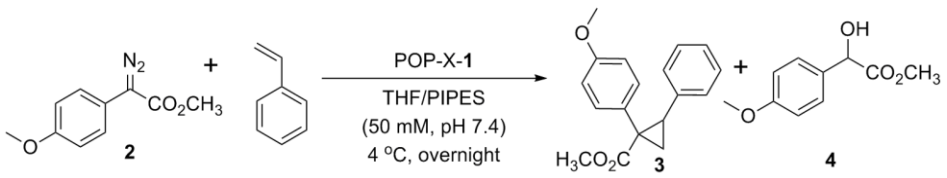
Second, the only screening criterion used in our engineering effort was improved enantioselectivity, but increased conversion was also observed (Table 2.7). This trend is particularly notable relative to soluble small molecule catalyst **5**, which gave lower conversion than any of the ArMs investigated (Table 2.7, Entry 15; Fig. 2.6B). While detailed kinetic analysis of these reactions was complicated by poor substrate solubility at elevated concentrations, monitoring the reaction of styrene with diazo **2** shows that ArMs that provide improved enantioselectivity also have increased cyclopropanation rates (Fig. 2.6). The rate of diazo consumption by these ArMs is well below that of **5**, which leads to nearly instantaneous consumption of **2** (Fig. 2.6C, D). Subsequent additions of **2** to reactions catalyzed by **5** lead to similarly rapid conversion of this species with only minor increases in cyclopropanation conversion, indicating that **5** remains active even after the first aliquot of **2** is consumed (Fig. 2.7). The discrepancy between diazo consumption and cyclopropanation catalyzed by **5** results from the poor substrate specificity of this catalyst in aqueous solution. Under these conditions, formal

carbene insertion into the O–H bond of water (rather than the olefin π bond) readily occurs to form α-hydroxyester **4**[17] a problem that has long-complicated aqueous dirhodium-catalysed carbene insertion reactions. Importantly, however, the cyclopropane/α-hydroxyester ratio (**3**/**4**) increases from 0.4 using 5 to 2.4 using POP-ZA$_4$-HFF-**1**. This six-fold increase occurred in increments that parallel increases in enantioselectivity (Table 2.7, Entries 4, 8, 11, 14). Together, these conversion, rate and selectivity data highlight the improved complementarity between POP and styrene in the engineered ArMs. The specificity of POP-ZA$_4$-HFF-**1** for styrene over water ultimately leads to increased cyclopropanation conversion even though **5** provides much faster conversion of **2** under the reaction conditions (Fig. 2.6C, D).

Figure 2.6. Kinetic analysis of cyclopropanation reactions. (a) Comparison of product yield versus time for cyclopropanation of styrene using **2** catalyzed by various ArMs or **5** (0.5 mol %). (b) Structure of **5**. (c) Conversion of **2**, **3**, **4** over time for POP- ZA$_4$-HFF-**1**. (d) Conversion of **2**, **3**, **4** over time for **5**. (e) Reaction scheme.



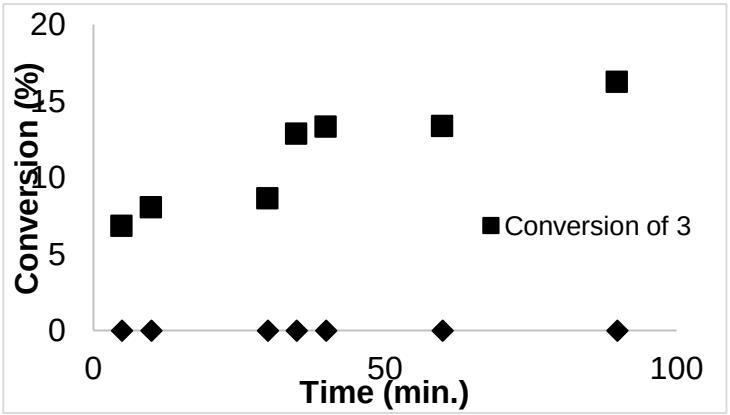
E



Third, despite the specificity of POP-ZA$_4$-HFF-**1** for styrene over water, this ArM also catalyzes enantioselective cyclopropanation of different styrenes using a variety of donor–acceptor diazo compounds (Table 2.8). Electron-withdrawing and electron-donating substituents were tolerated on the aromatic groups of both the styrene and diazo substrates. A styrenyl diazo substrate also reacted, albeit with significantly reduced selectivity relative to aryl diazoacetates. Ethyl diazoacetate, as mentioned earlier, an acceptor-only carbene precursor, was also a competent substrate, but provided negligible enantioselectivity.

Figure 2.7. (a) Plot of conversion of **2** and **3** catalyzed by **5** (equal amount of **2** was added at 0, 30, 60 min). Following each addition, **2** is immediately consumed and additional conversion to **3** is observed. (b) Table of conversion of **2** and **3** catalyzed by **5** (equal amount of **2** was added at 0, 30, 60 min).

A



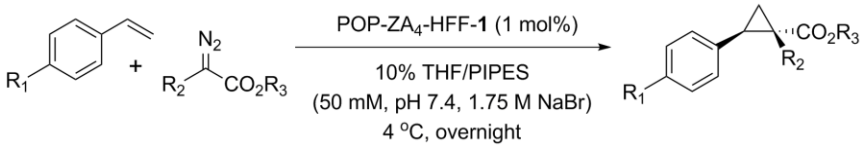
B

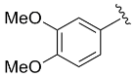
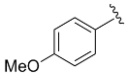
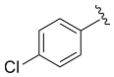
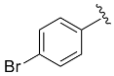
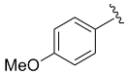
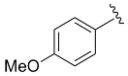
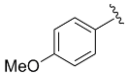
time (min)	5	10	30	35	40	60	90
2%	0	0	0	0	0	0	0
3%	6.8	8.0	8.6	12.8	13.3	13.3	16.2

The unique catalytic properties of POP-ZA$_4$-HFF-**1** result from extensive protein engineering on the wild-type POP protein. The introduction of eight mutations (four mutations to enable bioconjugation and four mutation to render improved selectivity) and dirhodium cofactor

1 into the interior of the POP scaffold, brings far more scaffold modifications than most ArM efforts[7]. Despite these perturbations, essentially identical circular dichroism (CD) spectra were obtained for several POP variants and POP-ZA$_4$-HFF-**1**, suggesting little difference in secondary structures of these proteins (Fig. 2.8A)[34]. Remarkably, the study of CD spectrum of POP-ZA$_4$-HFF-**1** at different temperatures reveals this ArM remains its secondary structure up to 100 ^{o}C (Fig. 2.8B), indicating that the stability of POP itself is also not reduced to a relevant extent. This stability clearly highlights the utility of protein scaffolds from hyperthermophilic organisms that can form robust ArMs even when extensive mutagenesis is required to achieve high selectivity and will greatly facilitate further efforts to evolve ArMs derived from the POP scaffold[35].

Table 2.8. Representative substrate scope of POP-ZA$_4$-HFF-**1** catalyzed cyclopropanation

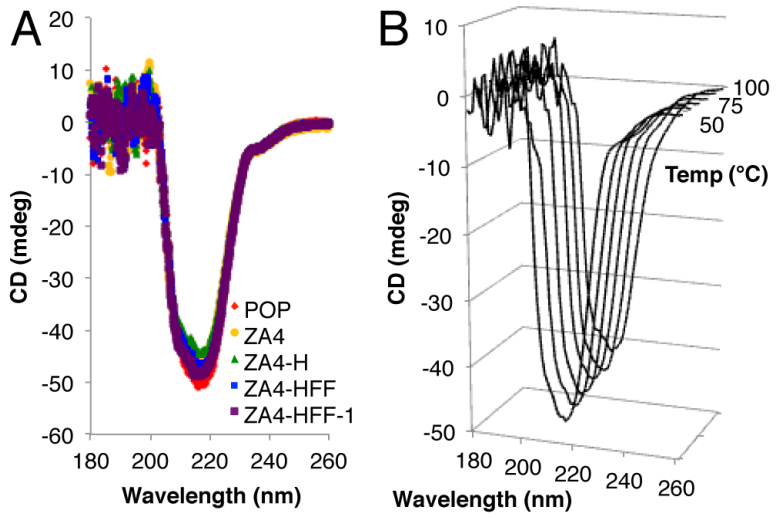


Entry	R^1	R^2	R^3	Yield (%)[a]	e.e. (%)
1	H		Me	73	74
2	H		Me	74	92
3	H	Ph	Me	14	74
4	H		Me	30	80
5	H		Me	43	80
6	H		Me	31	31
7	H		Et	40	90
8	OMe		Me	56	86
9	Cl		Me	37	80

[a] All reactions conducted using 20 mM olefin and 4 mM diazo. Conversion and enantioselectivity determined by HPLC relative to internal standard.

Of the mutations introduced, L328H led to the largest improvements in both selectivity, conversion and activity (Table 2.7, entry 8). As previously noted, this mutation was introduced based on the improved selectivity of peptide-based dirhodium catalysts containing a histidine residue capable of coordinating to Rh[32]. It is important to note, however, that axial coordination of ligands to dirhodium complexes in peptide and small molecule catalysts typically leads to decreased activity[32]. Given the difference in the effects of histidine incorporation into peptide catalysts and POP-ZA_4-HFF-**1**, several additional ArM variants were examined to clarify the role of H328 in POP-ZA_4-HFF-**1** (Table 2.7, entries 4–8). First, POP-ZA_4-L328F-**1** was prepared to examine the impact of a non-coordinating aromatic residue at position 328. The L328F variant possesses significantly lower selectivity than the L328H variant, suggesting that purely steric factors are not responsible for the improved selectivity of the latter. In addition, the L328M and L328C variants show that other residues capable of coordinating to Rh also improve ArM selectivity. The structural differences between histidine, methionine and cysteine suggest that their common metal-coordinating ability is responsible for the improved selectivity ArMs containing these residues, including POP-ZA_4-HFF-**1**. Initial attempts to characterize histidine coordination to 1in this ArM via NMR spectroscopy[36, 37] and ultraviolet–vis spectroscopy[32, 38, 39] have been complicated by the high molecular weight of POP (ca. 70 kDa) and the weak absorbance associated with the diagnostic Rh–Rh π^*–σ^* transition[40] in **1**, respectively. Further spectroscopic and crystallographic analysis of this ArM is underway to rigorously characterize the nature of cofactor binding within its active site and thus provide a mechanistic rationale for its high selectivity and improved specificity.

Figure 2.8. CD spectra for POP variants and ArMs. (a) Comparing different constructs (10 mM). (b) CD spectra of POP-ZA$_4$-HFF acquired at 10 ^{o}C intervals from 50 to 100 ^{o}C (see also Fig. S2.3)



DISCUSSION

The unique structure of *Pfu* POP has allowed us to engineer ArMs using this enzyme and **1** to catalyze enantioselective cyclopropanation. This effort required genetic incorporation of a 4-L-azidophenylalanine residue to covalently link **1**, four alanine mutations (A$_4$) to enable cofactor entry into the POP active site and three additional active site mutations (HFF) to improve the enatioselectivity and substrate specificity of the initial ArM construct. The use of SPAAC for cofactor bioconjugation provided the flexibility to choose POP as a scaffold because of its physical properties (shape, size and stability), rather than native cofactor-binding ability[10, 41], which, in turn, allowed the extensive mutagenesis required for ArM formation and selective catalysis. Despite this mutagenesis, the optimized ArM, POP-ZA$_4$-HFF-**1**, is extremely stable (Fig. 2.8), which will facilitate subsequent evolution[35] of ArMs with improved activity and selectivity for different substrates and reactions.

POP-ZA$_4$-HFF-**1** accepts a range of styrene and donor–acceptor carbene precursor[29] substrates. In the latter respect, it contrasts with recent reports from Arnold[42, 43] and Fasan[44] who have shown that naturally occurring haeme proteins catalyze olefin cyclopropanation using ethyl diazoacetate (an acceptor-only carbene precursor). Furthermore, while exciting developments, these systems exploit the native folds of enzymes and proteins that evolved to bind haeme in a manner appropriate for interacting with substrates in well-defined active sites. In contrast, selective ArM catalysis involves incorporating a synthetic metal complex into a protein scaffold and engineering an active site suitable for imparting selectivity to that complex. In the current case, this effort led to improved specificity of POP-ZA$_4$-HFF-**1** for styrene over water, which is remarkable, given the known reactivity of dirhodium donor–acceptor carbene intermediates[45] towards water[18] and suggests that similar control of other water-sensitive organometallics could be possible using the solvent-sequestered POP active site. This contrasts significantly with peptide scaffolds, which, while being capable of imparting high levels of selectivity to dirhodium catalysts, require the use of organic solvents or a large excess of diazo substrate[46] to compensate for reactivity of carbene intermediates with water. Given the wide range of reactions catalysed by dirhodium complexes[17] and the selectivity of POP-ZA$_4$-HFF-**1**, we anticipate that dirhodium ArMs will provide many unique opportunities for selective catalysis. Furthermore, the ability of POP to impart selectivity to the rigid dirhodium complex suggests that similar selectivity should be possible for a wide range of additional metal complexes regardless of their stereochemical properties[10]. POP will thus serve as a robust scaffold to explore this possibility and to study the effects of attractive interactions, molecular recognition and scaffold dynamics on transition metal catalysis.

ACKNOWLEDGMENTS

We thank Professor Jeffry Madura at Duquesne University for providing us with the POP homology model used in these studies, Professor Harold Schreier at the University of Maryland, Baltimore County for providing the *Pfu* POP gene, Dr. Landon Durak for reviewing the manuscript and CCHF members for many helpful discussions.

EXPERIMENTAL

Materials

Unless otherwise noted, all reagents were obtained from commercial suppliers and used without further purification. Benzene, dimethylformamide (DMF), acetonitrile (ACN), pentane, tetrahydrofuran (THF), and methylene chloride (CH_2Cl_2) were obtained from a PureSolv MD solvent purification system by Innovative Technology (solvent deoxygenated by N_2 sparge and dried over alumina). Acetonitrile (ACN) was purchased from Fisher Chemical, HPLC grade. Deuterated solvents were obtained from Cambridge Isotope labs. Silicycle silica gel plates (250 mm, 60 F254) were used for analytical TLC, and preparative chromatography was performed using SiliCycle SiliaFlash silica gel (230-400 mesh). Rh_2(R-DOSP)$_4$ was purchased from Strem Chemicals. Azide Agarose was purchased from Click Chemistry Tools LLC. Labquake™ Tube Shaker/Rotators was purchased from Thermo Scientific (Catalog# 4002110Q).

Plasmid pEVOL-pAzF was provided by the Schultz group of the Scripps Research Institute, CA[50]. *E. coli* DH5α and BL21 (DE3) cells were purchased from Invitrogen (Carlsbad, CA). Nco I, Xho I restriction enzyme, T4 DNA ligase, Taq DNA polymerase and Phusion HF polymerase (Cat#

530S) were purchased from New England Biolabs (Ipswitch, MA). Luria broth (LB), rich medium (2YT) and Agar media were purchased from Research Products International (Mt. Prospect, IL). Qiagen DNA extraction kit (Cat# 28706) and plasmid isolation kit (Cat# 27106) were purchased from QIAGEN Inc. (Valencia, CA) and used according to the manufacturer's instructions. DNA purification kit (Zymo, Cat# D4004) was purchased from Zymo research (Irvine, CA) and used as recommended. 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. Ni-nitrilotriacetic acid (Ni-NTA) resin and Pierce® BCA Protein Assay Kits (Cat# 23225) were purchased from Fisher Scientific International, Inc. (Hampton, NH), and the manufacturer's instructions were followed when using both products (for Ni-NTA resin, 8 mL resin was used with buffers delivered by a peristaltic pump at a rate of 1 mL/min, in a 4 °C cold cabinet). Amicon® 30 kD spin filters for centrifugal concentration were purchased from EMD Millipore (Billerica, MA) and used at 4,000 g at 4 °C.

General Procedures

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 as the internal standard and

reported as the average of two 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. Low resolution ESI mass spectra were obtained using Agilent 6130 LC-MS. Amicon® 50 mL 30 kD cutoff centrifugal filter was 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. Circular dichroism (CD) spectra were obtained on a JASCO J-1500 CD Spectrometer.

Cloning, expression and protein purification

Standard cloning procedures and site directed mutagenesis:

A codon optimized gene for Prolyl oligopeptidase (POP) was obtained from GenScript USA Inc (Piscataway, NJ) and cloned into pET28a plasmid vector using NcoI and XhoI restriction sites. The gene was cloned upstream of a C-terminal hexa-histidine tag for Ni-NTA affinity chromatography. Alanine mutations (at positions E104A, F146A, K199A and D202A), histidine mutations (at positions G99H, P139H, I141H, I197H, T209H, E218H, V219H, Y251H, E283H, and L328H), and phenylalanine mutations (at positions S64F, L97F, G99F, G594F) were introduced into the POP gene by site directed overlap extension PCR[51]. To introduce mutations, two separate polymerase chain reactions were performed, each using a perfectly complementary flanking primer at the 5' and 3' end of the sequence and a mutagenic primer. The PCR conditions were as follows: Phusion HF buffer 1x, 0.2 mM dNTPs each, 0.5 μM forward primer, 0.5 μM reverse primer, 0.02 U/μL Phusion polymerase and 0.5 ng/mL template plasmid.

Thermal cycler was programmed as:

1. 98 °C-60 seconds
2. 95 °C-20 seconds

3. 54 °C-45 seconds

4. 72 °C- 120 seconds

5. 72 °C-10 minutes

25 repeat cycles from #2 to #4

The resulting two overlapping fragments that contained the base pair substitution were then assembled in a second PCR using the flanking primers resulting in the full-length mutated gene. The same PCR program was used with a slightly altered annealing temperature of 52 °C. Nucleotide sequences for the all the primers are summarized in Table S2.1.

Table S2.1. Nucleotide sequences for the primers

#	Primer name	Primer sequence
1	T7 for	5'-GCG AAA TTA ATA CGA CTC ACT ATA-3'
2	T7 rev	5'-TTA TGC TAG TTA TTG CTC AGC GG-3'
3	E104A for	5'- ACC ACG GAC GCG GAA GGT GAA A -3'
4	E104A rev	5'- T TTC ACC TTC CGC GTC CGT GGT -3'
5	F146A for	5'- AAC ATC ACC GCC CTG AAA GAT G -3'
6	F146A rev	5'- C ATC TTT CAG GGC GGT GAT GTT-3'
7	K199D202A for	5'- G TCC ATT CGC GCA AGC TCT GCT GGT AAA TTC G-3'
8	K199D202A rev	5'- C GAA TTT ACC AGC AGA GCT TGC GCG AAT GGA C-3'
9	POPz477 for	5'-A GCT TGG GGT CGT TAG AAT GGC GGT CTG-3'
10	POPz477 rev	5'-CAG ACC GCC ATT CTA ACG ACC CCA AGC T- 3'
11	HisG99 for	5'- TC CTG CTG CAG CAC TTT ACC ACG G-3'
12	HisG99 rev	5'- C CGT GGT AAA GTG CTG CAG CAG GA-3'
13	HisP139 for	5'- GAA GAA ATC AAA CAC TCC ATT TGG AAC-3'
14	HisP139 rev	5'- GTT CCA AAT GGA GTG TTT GAT TTC TTC -3'
15	HisI141 for	5'- C AAA CCG TCC CAC TGG AAC ATC ACC -3'
16	HisI141 rev	5'- GGT GAT GTT CCA GTG GGA CGG TTT G -3'
17	HisI197 for	5'- AT TTC ATG TCC CAC CGC GCA AGC TC-3'
18	HisI197 rev	5'- GA GCT TGC GCG GTG GGA CAT GAA AT-3'
19	HisT209 for	5'- TTC GCA ATC GTT CAC CTG ACG TAT GGT -3'
20	HisT209 rev	5'- ACC ATA CGT CAG GTG AAC GAT TGC GAA - 3'

Table S2.1. Nucleotide sequences for the primers, continued

#	Primer name	Primer sequence
21	HisE218 for	5'- AAC CAG GGC CAC GTC TAC ATT GG -3'
22	HisE218 rev	5'- CC AAT GTA GAC GTG GCC CTG GTT -3'
23	HisV219 for	5'- CAG GGC GAA CAC TAC ATT GGT CC -3'
24	HisV219 rev	5'- GG ACC AAT GTA GTG TTC GCC CTG -3'
25	HisY251 for	5'- GGC AAA CTG CAC ATC CTG ACC -3'
26	HisY251 rev	5'- GGT CAG GAT GTG CAG TTT GCC- 3'
27	HisE283 for	5'- AA TTT CCG CTG CAC TGG GCA GTC ATT GT -3'
28	HisE283 rev	5'- AC AAT GAC TGC CCA GTG CAG CGG AAA TT -3'
29	HisW284 for	5'- A TTT CCG CTG GAA CAC GCA GTC ATT GT -3'
30	HisW284 rev	5'- AC AAT GAC TGC GTG TTC CAG CGG AAA T -3'
31	HisL328 for	5'- CA CTG TAT CCG CAC GAT AAA GAC GA -3'
32	HisL328 rev	5'- TC GTC TTT ATC GTG CGG ATA CAG TG -3'
33	G99F for	5'- C CTG CTG CAG TTC TTT ACC ACG GA -3'
34	G99F rev	5'-TC CGT GGT AAA GAA CTG CAG CAG G- 3'
35	G594F for	5'-CA GGT CAC ATG TTC GCG TCG CCG G- 3'
36	G594F rev	5'- C CGG CGA CGC GAA CAT GTG ACC TG- 3'
37	L97F for	5'- T GAA GTC CTG TTT CAG GGC TTT ACC- 3'
38	L97F rev	5'- GGT AAA GCC CTG AAA CAG GAC TTC A- 3'
39	S64F for	5'- GGT ATT ATC GCT TTT TAT TCC GAA AAA- 3'
40	S64F rev	5'- TTT TTC GGA ATA AAA AGC GAT AAT ACC-3'

PCR amplified fragments and plasmid vector pET28a were restriction digested with Nco I and Xho I enzymes in recommended buffer at 37 °C for 2 hours. Digested DNA was cleaned by agarose gel extraction using commercial kit before ligation. Ligation was set-up with a molar ratio of 1:3 (plasmid: insert) in 10 µL reaction mix. Typically a ligase reaction mix had 3 ng/L digested plasmid vector, 9 ng/mL of the insert, 1 µL 10X ligase buffer and 1 U/mL ligase. The reaction mixture was incubated at 16 °C overnight, cleaned using DNA purification kits and transformed into *E. coli* DH5 cells. Cells were spread on LB kanamycin plates (6.25 g LB powder mix, 4 g agar, 250 mL DDI water, 0.05 mg/mL kanamycin) before recovering in SOC medium for 1 hour at 37 °C. Plates

were incubated at 37 °C overnight; individual colonies that appeared next day were tested for gene fragments by colony PCR. Clones that showed amplification for desired fragments were inoculated on LB broth having 0.05 mg/mL kanamycin and grown overnight at 37 °C, 250 rpm. Recombinant plasmid from these overnight grown cultures were isolated using kit from Qiagen (Valencia, CA) and given for sequencing. Plasmid sequencing was done at the U Chicago sequencing facility and T7 for and T7 rev primers were used for sequencing reactions.

Standard expression and purification procedure:

pET28a-POP-ZA_4 and pEVOL-$pAzF^{50}$ were co-transformed into electrocompetent *E. coli* BL21 (DE3). Transformed cells were allowed to recover in SOC medium (37 °C, 50 min), then plated onto LB kan+Cm agar plates (6.25 g LB powder mix, 4 g agar, 250 mL DDI water, 0.05 mg/mL kanamycin, 0.05 mg/mL chloramphenicol), and incubated at 37 °C for 16 h. Several colonies appeared on overnight-incubated plates; a single colony from this plate was inoculated in 5 mL 2YT medium having antibiotics with the same concentrations as above. The culture was incubated overnight at 37 °C 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 5 L Erlenmeyer flask. The culture was incubated at 37 °C, 250 rpm, and protein expression was induced by adding 1mM IPTG, 2mM 4-Azido-phenyl alanine and 1% (w/v) L-arabinose when OD_{600} reached 1. The induced culture was allowed to grow for 12 hours, and then the cells were harvested by centrifugation at 4 °C, 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 °C for 30 minutes and supernatant thus obtained was purified by Ni-NTA resin using manufacturer's instructions. Purified protein was buffer

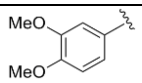
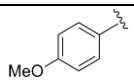
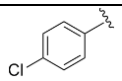
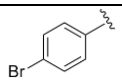
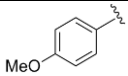
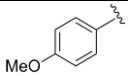
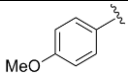
exchanged to 10 mM Tris (pH 7.4) and measured by Pierce® BCA Protein Assay Kit as recommended.

Cofactor and Standard Product Synthesis

Cofactor **1** was prepared with previously reported methods[49].

Synthesis of aryldiazoacetates and cyclopropanes in the table below:

Table S2.2. List of Substrates

Entry	R^1	R^2	R^3	Diazo	cyclopropane
1	H		Me	**2a**	**3a**
2	H		Me	**2b**	**3b**
3	H	Ph	Me	**2c**	**3c**
4	H		Me	**2d**	**3d**
5	H		Me	**2e**	**3e**
6	H		Me	**2f**	**3f**
7	H		Et	**2g**	**3g**
8	OMe		Me	**2b**	**3h**
9	Cl		Me	**2b**	**3i**

General procedure for the synthesis of aryldiazoacetates[S4]:

The arylacetate (3 mmol, 1 equiv), *p*-ABSA (1.3 equiv), 20 mL acetonitrile were added to a 100-mL round-bottom flask with a magnetic stir bar. The reaction mixture was cooled to 0 °C using an

ice bath under nitrogen. 1,8-Diazabicycloundec-7-ene (DBU, 1.4 equiv) was then added to the stirring mixture over the duration of 5 min. After the addition of the DBU, the reaction mixture was stirred at 0 °C for an additional 30 min. The ice bath was removed and the reaction mixture was stirred for 24 h at room temperature. The resulting solution was quenched with saturated NH_4Cl solution and the aqueous layer was extracted with diethyl ether (3 x 50 mL). The organic extracts were combined, washed with H_2O, and dried over $MgSO_4$. The organic layer was then concentrated under reduced pressure and purified using silica gel column chromatography (10:1 hexanes/EtOAc). Diazoacetates **2a-f**[S5, S6] were synthesized according to the general procedure and characterization match previous literature.

Ethyl 4-methoxyphenyldiazoacetate (2g): Title compound was prepared by the general procedure and obtained as an orange solid with 47% yield. ^{1}H NMR (500 MHz; $CDCl_3$) δ 7.38 (d, *J* = 9.0 Hz, 2H), 6.94 (d, *J* = 9.0 Hz, 2H), 4.32 (q, *J* = 7.1 Hz, 2H), 3.81 (s, 3H), 1.33 (t, *J* = 7.1 Hz, 3H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 165.9, 158.2, 126.1, 117.2, 114.7, 61.0, 55.5, 14.6 ($C=N_2$ signal missing); IR (KBr, cm^{-1}): 2086.2, 1700.4. HRMS (ESI-MS) calcd for $C_{11}H_{13}O_3$ $(M-N_2+H)^+$ 193.0865, found 193.0868.

General procedure for the synthesis of cyclopropanes with Rh_2(R-DOSP)$_4$[S4]:

Styrene (5 equiv) and Rh_2(R-DOSP)$_4$ (0.01 equiv) were added to a 25-mL round bottom flask (flask A) equipped with a magnetic stir bar and degassed using vacuum/nitrogen cycles (x3). 3 mL pentane was added to flask A under nitrogen. The aryldiazoacetate (0.5 mmol, 1 equiv) was added to a separate 25-mL round bottom flask (flask B) and degassed using vacuum/nitrogen cycles (x3). 5 mL pentane was added to flask B under nitrogen. The contents in flask B were then added to flask A using a syringe pump for the duration of 1 h. After the addition, the reaction mixture was

stirred for one additional hour. The reaction mixture was concentrated under reduced pressure and purified using silica gel column chromatography (increasing gradient starting at 10:1 hexanes/EtOAc). Cyclopropanes (**3a-f**, **3h-i**)[S4, S7] were synthesized according to the general procedure and characterization match previous literature.

(1*S*,2*R*)-ethyl 1-(4-methoxyphenyl)-2-phenylcyclopropanecarboxylate (3g): Title compound was prepared by the general procedure and obtained as a white solid with 58% yield. ^{1}H NMR (500 MHz; $CDCl_3$) δ 7.06-6.92 (m, 5H), 6.77 (m, 2H), 6.65 (d, *J* = 8.8 Hz, 2H), 4.12 (m, 2H), 3.72 (s, 3H), 3.05 (dd, *J* = 9.2 and 7.2 Hz, 1H), 2.11 (dd, *J* = 9.3 and 4.8 Hz, 1H), 1.81 (dd, *J* = 7.2 and 4.8 Hz, 1H), 1.18 (t, *J* = 7.2 Hz, 3H); ^{13}C NMR (125 MHz, $CDCl_3$) δ 174.1, 158.5, 136.8, 133.0, 128.2, 127.8, 127.1, 126.3, 113.2, 61.3, 55.2, 37.0, 33.0, 20.5, 14.3; HRMS (ESI-MS) calcd for $C_{19}H_{20}O_3$ $(M+H)^+$ 297.1491, found 297.1495.

Product **4** was prepared with previous reported methods[S8].

Synthesis of Artificial Enzyme and Characterization

Preparation of Metalloenzyme (bioconjugation):

A solution of the POP-Z mutant (480 μL, 75 μM in 50 mM Tris-HCl buffer, pH 7.4) and a solution of cofactor **1** (120 μL, 0.75 mM in ACN, 0.655 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 **1**, 20 vol% acetonitrile/Tris 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 h to remove excess cofactor. The suspension was then centrifuged at 5000 rpm for 3 min and the supernatant was transferred to a new microcentrifuge tube. The resin was rinsed twice with 600 μL 50 mM Tris-HCl buffer and centrifuged at 5000 rpm for 3 min. These supernatants were combined with

the first supernatant and buffer exchanged to proper buffers for use in biocatalysis or characterization. ESI-MS were used to characterize the bioconjugates. It is worth noting that the bioconjugation reaction often does not go to completion (40 % ~ 60 % incorporation of the dirhodium cofactor was typically observed), depending on specific mutations in the POP scaffold, based on high resolution ESI-MS. This results in part from reduction of azide to anline as indicated by HR ESI-MS, although we did not observe this process in our earlier work.S3 Because of this, the effective ArM concentration was determined according to the following method: the total protein concentration was calculated based on its absorbance at 280 nm (A280) and the calculated extinction coefficient for the protein (109,210 M-1cm-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 incorporation was calculated based on the ratio of the high resolution ESI-MS peak intensity of the ArM and scaffold (IArM/(IArM+Iscaffold)); the effective ArM concentration was calculated by multiplying the total protein concentration by the efficiency of dirhodium incorporation ([ArM]=[Total protein]*(IArM/(IArM+Iscaffold)). The effective ArM loading was adjusted to 1 mol% with respect to the dirhodium cofactor in bioconversions and 0.5 % with respect to the dirhodium cofactor in kinetic study.

MS Characterization of POP metalloenzyme:

For ESI-TOF MS analysis, a sample of protein was desalted with centrifugal filters to a mixture of water: acetonitrile: glacial acetic acid (49.5: 49.5: 1, v/v). The final protein concentration was 50 μM. Acquisition of the spectra was perfomed by flow injection analysis with fragmenter set at 100V-200V. Raw ESI spectra (shown in Fig. S2.1) were deconvoluted using the Agilent

Chemstation LC/MSD data deconvolution module. The deconvoluted massesare in good agreement with the predicted masses (Table S2.2). It should be noted that the mass for POP- ZA_4-HFF is typically observed 20-30 Da lower than its theoretical mass, consistent with putative loss of N_2 from the azide.

Figure S2.1. Raw ESI spectra of **A)** POP- ZA_4-HFF (green) and **B)** POP- ZA_4-HFF-**1** (blue)

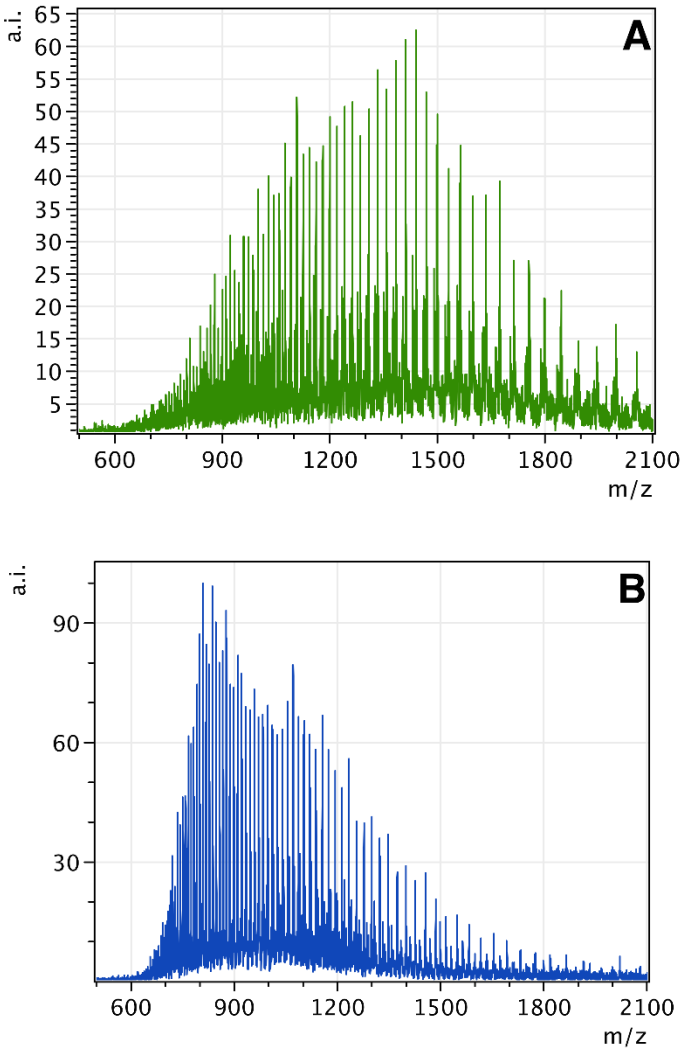


Table S2.3. Calculated masses versus observed deconvoluted masses.

Protein Species	Calculated Mass (Da)	Observed Mass (Da)	Δ Mass (Da)
POP- ZA_4-HFF	71959.4	71935.6	-23.6

Table S2.3. Calculated masses versus observed deconvoluted masses, continued

Protein Species	Calculated Mass (Da)	Observed Mass (Da)	Δ Mass (Da)
POP- ZA_4-HFF-**1**	72751.9	72745.3	-6.6

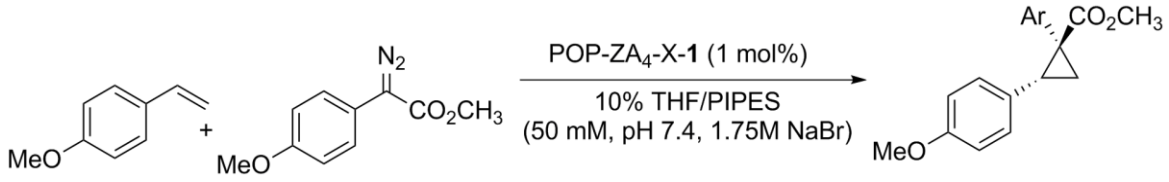
Bioconversion and Kinetics

Bioconversion:

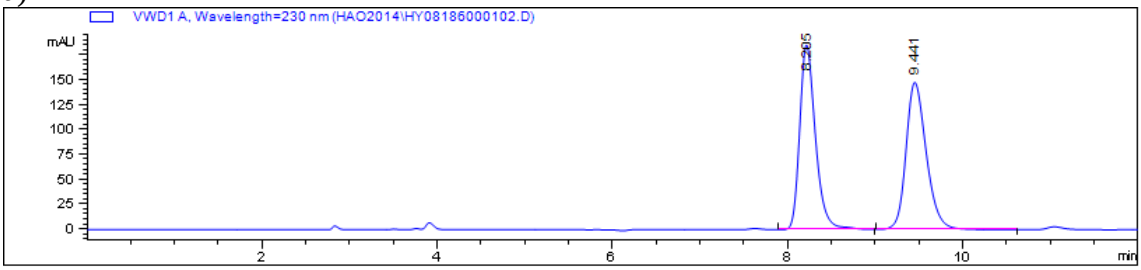
Solutions of aryldiazoacetate (25 µL, 96 mM, in THF), styrene (25 µL, 485 mM, in THF), and POP-ZA_4-X-**1** solution (500 µL, the effective ArM concentration adjusted to 48 µM with respect to the dirhodium cofactor according to the aforementioned method) were added to a 1.5 mL microcentrifuge tube. The final concentrations of the reagents were: 22 mM olefin, 4.4 mM aryldiazoacetate, 44 µM POP- ZA_4-X-**1**. The resulting mixture was left shaking at 750 rpm at 4 °C overnight. The reaction was quenched by adding 20 µL 1,2,4-trimethoxybenzene solution (30 mM, in THF) and 600 µL ethyl acetate. The mixture was vortexed and centrifuged (15,000 x g, 3 min). 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 THF. 4 µL THF solution of the crude product was analyzed on RP-HPLC to determine conversions; 50 µL THF solution of the crude product was purified on preparative-HPLC to isolate the cyclopropane product, which was analyzed on NP-HPLC to determine enantioselectivities. The conversions and enantioselectivities were reported as the average of two trials from the same batch of ArM set up in parallel. The RP-HPLC to determine conversions was performed on an Agilent 1100 Series HPLC system using an Agilent Eclipse Plus C18 column (95 Å, 3.5 µM, 4.6 mm i.d. x 150 mm), with 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: water containing 0.1% TFA; solvent B: CH_3CN). The preparative HPLC used the same method as above. The NP-

HPLC to determine enantioselectivities was performed on 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. Representative traces of chiral-HPLC for determining enantioselectivities in bioconversions are shown in Fig. S2.2. Because the dirhodium cofactor concentration in ArM was normalized as described above, the batch-to-batch variations in the catalytic experiments were minute (<1% for ee and <5% for conversion), regardless of bioconjugation conversion, as shown in Table S2.3.

Figure S2.2. Representative HPLC traces for a) a racemic mixture and b) enantiomeric mixture made by POP- ZA_4-HFF-**1**



a)



b)

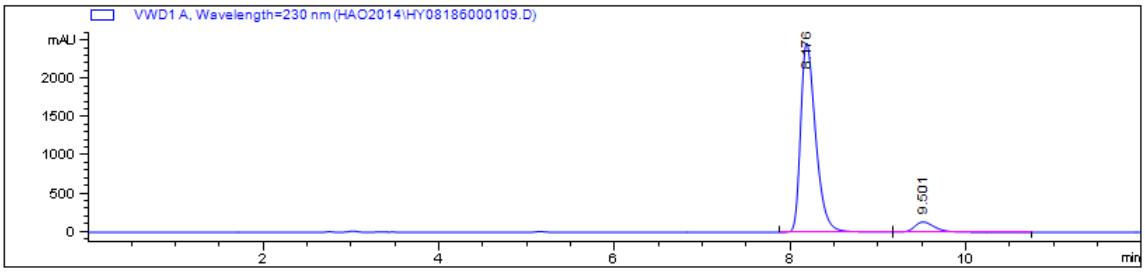


Table S2.4. Summary of bioconversion (for the reaction in Fig. S2.2) catalyzed by duplicates from three independent batches of POP- ZA_4-HFF-1.

	Batch 1		Batch 2		Batch 3	
	Trial 1	Trial 2	Trial 1	Trial 2	Trial 1	Trial 2
Yield (%)	70	74	71	74	75	73
e.e.(%)	92	91	91	92	92	92

Qualitative Kinetic Analysis:

The conditions used in bioconversions were slightly modified for kinetic experiment. In a 1.5 mL microcentrifuge tube, a solution of **2** (12.5 µL, 96 mM, in THF), a styrene solution (12.5 µL, 485 mM, in THF) and POP-ZA_4-X-**1** solution (250 µL, the effective ArM concentration adjusted to 24 µM with respect to the dirhodium cofactor according to the aforementioned method) were added. The resulting mixture was left shaking at 750 rpm at 4 °C. The final concentrations of the reagents were: 22 mM styrene, 4.4 mM **2**, 22 µM POP- ZA_4-X-**1**.

To determine conversion of product (**3**), the following workup was used: after the set time, the reaction was quenched by adding 30 µL 1,3-dimethoxybenzene solution (30 mM, in THF) and 1000 µL ethyl acetate. The mixture was vortexed and centrifuged (15,000 x g, 3 min). The top organic layer was analyzed by RP-HPLC. The conversions at all the time points were reported as the average of two trials from the same batch of ArM set up in parallel. The RP-HPLC to determine conversions was performed on an Agilent 1290 Series HPLC system using an Agilent Eclipse Plus C18 RRHD column (300 Å, 1.8 µM, 2.1 mm i.d. x 50 mm), with with a flow rate of 0.4 mL/min and detection wavelength set at 230 nm. The following gradient was used: 60 % B from 0-5 min, 60 % to 100 % from 5-7 min, 100 % from 7-8 min, 100 % from 8-8.5 min, 1.5 min post-run (solvent A: water containing 0.1% TFA; solvent B: CH_3CN). The data are shown in Table S2.4 and Fig. S2.3.

To determine conversions of **2, 3, 4**, a slightly different workup was used: after the set time, the reaction was quenched by adding 30 µL 1,3-dimethoxybenzene solution (30 mM, in THF) and 400 µL dichloromethane. The mixture was vortexed and centrifuged (15,000 x g, 3 min). The organic layer was collected and the aqueous layer was extracted twice with 400 µL dichloromethane. The

organic layer was combined and analyzed by RP-HPLC. The conversions at all the time points were reported as the average of two trials from the same batch of ArM set up in parallel. The RP-HPLC to determine conversions was performed on an Agilent 1290 Series HPLC system using an Agilent Eclipse Plus C18 column (95 Å, 3.5 µM, 4.6 mm i.d. x 150 mm), with with a flow rate of 1 mL/min and detection wavelength set at 230 nm. The following gradient was used: 10 % to 73 % from 0-7 min, 73 % from 7-10 min, 73 % to 100 % from 10-12 min, 4 min post-run (solvent A: water containing 0.1% TFA; solvent B: CH_3CN). The data are shown below in Table S2.5, 2.6 and Fig. S2.4, 2.5.

Table S2.5. Yield for biocatalysis catalyzed by **5** and selective ArM hybrids as a function of time.

	Average yield (%)			
Time (min)	ZA_4-1	ZA_4-H328-1	ZA_4-HFF-1	**5**
4	3.4	3.7	4.3	7.1
10	4.7	7.5	9.7	8.2
20	7.6	13.0	16.2	8.0
30	11.1	17.7	21.9	8.5
40	11.5	21.3	31.3	8.6
60	13.6	26.8	35.0	8.3
120	18.4	34.1	49.6	9.2

Table S2.6. Conversion of **2**, **3**, **4** over time for POP- ZA_4-HFF-**1**.

Time (min)	**2**%	**3**%	**4**%	(**2**+**3**+**4**)%
4	86.9	4.3	2.7	93.9
10	76.1	9.1	5.0	90.2
20	71.2	13.2	7.0	91.4
30	65.0	18.2	9.4	92.6
40	57.2	23.9	12.1	93.2
60	39.3	35.3	17.1	91.7
120	18.0	48.3	22.6	89.0

Table S2.7. Conversion of **2**, **3**, **4** over time for **5**.

Time (min)	**2**%	**3**%	**4**%	(**2**+**3**+**4**)%
4	0	7.0	27.4	34.4
10	0	8.8	30.6	39.4
20	0	9.2	31.4	40.5
30	0	9.0	31.9	40.9
40	0	8.5	29.1	37.7

Table S2.7. Conversion of **2**, **3**, **4** over time for **5**, continued

Time (min)	**2**%	**3**%	**4**%	(**2**+**3**+**4**)%
60	0	8.8	28.8	37.7
120	0	9.2	28.5	37.7

*Additional kinetic experiment for **5***:

Solutions of **2** (12.5 µL, 96 mM, in THF), styrene (12.5 µL, 485 mM, in THF), and POP- ZA_4-X-**1** (250 µL, 24 µM) were added to a 1.5 mL microcentrifuge tube. The resulting mixture was left shaking at 750 rpm at 4 °C. At 30-min and 60-min time points, solutions of additional **2** and styrene were added (**2:** 12.5 µL, 96 mM, in THF; styrene: 12.5 µL, 485 mM, in THF). The following workup was used: after the set time, the reaction was quenched by adding 30 µL 1,3-dimethoxybenzene solution (30 mM, in THF) and 1000 µL ethyl acetate. The mixture was vortexed and centrifuged (15,000 x g, 3 min). The top organic layer was analyzed by RP-HPLC. The conversions at all the time points were reported as the average of two trials from the same batch of ArM set up in parallel. The data are shown below in Table S2.7 and Supplementary Fig. S2.6.

Table S2.8. Conversion of **2** and **3** catalyzed by **5** (equal amount of **2** added at 0, 30, 60 min).

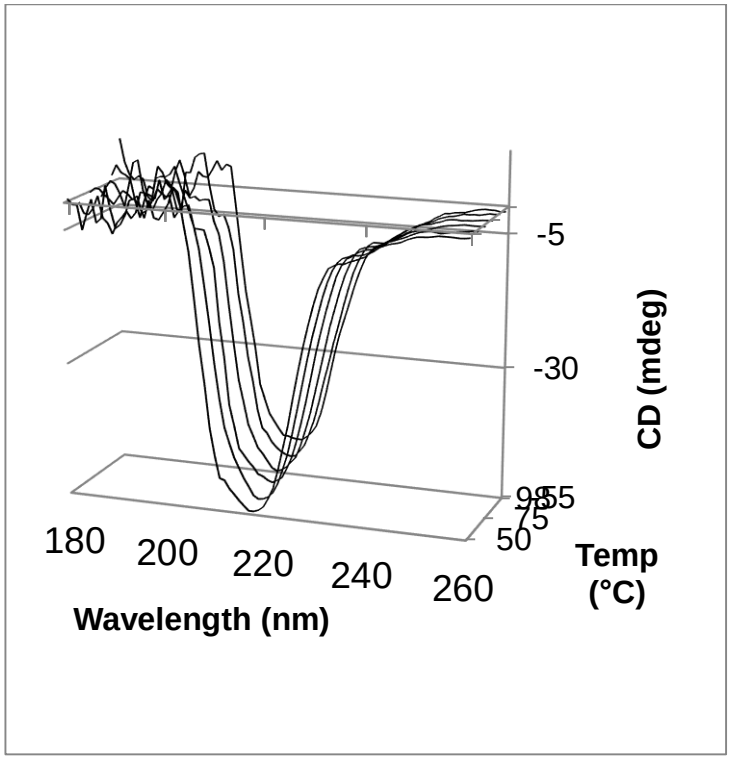
Time (min)	**2**%	**3**%
5	0	6.8
10	0	8.0
30	0	8.6
35	0	12.8
40	0	13.3
60	0	13.3
90	0	16.2

Circular Dichroism (CD) Analysis

CD spectra were acquired using a 10 mm pathlength quartz cuvette. All spectra were acquired at 25°C. Protein concentration was fixed at 10 uM (determined by A_{280}) in 100 mM sodium

phosphate buffer pH 7.0. Temperature stability profiles were acquired at 10 uM protein concentration in 100 mM sodium phosphate buffer pH 7.0. CD curves were acquired at 10°C intervals from 50°C to 100°C, with a heating gradient of 2°C/min. Acquisition was commenced after samples were equilibrated for 2 minutes at each temperature step. Shown in Supplementary Fig. S2.3 is the temperature stability profile for wild-type POP.

Figure S2.3. CD temperature stability profile for WT POP



REFERENCES

1. Knowles, R.; Jacobsen, E., Attractive noncovalent interactions in asymmetric catalysis: Links between enzymes and small molecule catalysts. *Proceedings of the National Academy of Sciences of the United States of America* **2010,** *107* (48), 20678-20685.

2. Lewis, C.; Gustafson, J.; Chiu, A.; Balsells, J.; Pollard, D.; Murry, J.; Reamer, R.; Hansen, K.; Miller, S., A Case of Remote Asymmetric Induction in the Peptide-Catalyzed Desymmetrization of a Bis(phenol). *Journal of the American Chemical Society* **2008,** *130*

(48), 16358-16365.

3. Weberski, M.; Chen, C.; Delferro, M.; Zuccaccia, C.; Macchioni, A.; Marks, T., Suppression of beta-Hydride Chain Transfer in Nickel(II)-Catalyzed Ethylene Polymerization via Weak Fluorocarbon Ligand-Product Interactions. *Organometallics* **2012,** *31* (9), 3773-3789.

4. Das, S.; Incarvito, C.; Crabtree, R.; Brudvig, G., Molecular recognition in the selective oxygenation of saturated C-H bonds by a dimanganese catalyst. *Science* **2006,** *312* (5782), 1941-1943.
5. Chen, Z.; Vohidov, F.; Coughlin, J.; Stagg, L.; Arold, S.; Ladbury, J.; Ball, Z., Catalytic Protein Modification with Dirhodium Metallopeptides: Specificity in Designed and Natural Systems. *Journal of the American Chemical Society* **2012,** *134* (24), 10138-10145.

6. Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D., How enzymes work: Analysis by modern rate theory and computer simulations. *Science* **2004,** *303* (5655), 186-195.

7. Lewis, J., Artificial Metalloenzymes and Metallopeptide Catalysts for Organic Synthesis. *Acs Catalysis* **2013,** *3* (12), 2954-2975.

8. Reetz, M.; Peyralans, J.; Maichele, A.; Fu, Y.; Maywald, M., Directed evolution of hybrid enzymes: Evolving enantioselectivity of an achiral Rh-complex anchored to a protein. *Chemical Communications* **2006,** (41), 4318-4320.

9. Bos, J.; Garcia-Herraiz, A.; Roelfes, G., An enantioselective artificial metallo-hydratase. *Chemical Science* **2013,** *4* (9), 3578-3582.

10. Ward, T., Artificial Metalloenzymes Based on the Biotin-Avidin Technology: Enantioselective Catalysis and Beyond. *Accounts of Chemical Research* **2011,** *44* (1), 47-57.

11. Schwizer, F.; Kohler, V.; Durrenberger, M.; Knorr, L.; Ward, T., Genetic Optimization of the Catalytic Efficiency of Artificial Imine Reductases Based on Biotin-Streptavidin Technology. *Acs Catalysis* **2013,** *3* (8), 1752-1755.

12. Hyster, T.; Knorr, L.; Ward, T.; Rovis, T., Biotinylated Rh(III) Complexes in Engineered Streptavidin for Accelerated Asymmetric C-H Activation. *Science* **2012,** *338* (6106), 500-503.

13. Song, W.; Tezcan, F., A designed supramolecular protein assembly with in vivo enzymatic activity. *Science* **2014,** *346* (6216), 1525-1528.

14. Kohler, V.; Wilson, Y.; Lo, C.; Sardo, A.; Ward, T., Protein-based hybrid catalysts-design and evolution. *Current Opinion in Biotechnology* **2010,** *21* (6), 744-752.

15. Yang, H.; Srivastava, P.; Zhang, C.; Lewis, J., A General Method for Artificial

Metalloenzyme Formation through Strain-Promoted Azide-Alkyne Cycloaddition. *Chembiochem* **2014,** *15* (2), 223-227.

16. Espino, C.; Fiori, K.; Kim, M.; Du Bois, J., Expanding the scope of C-H amination through catalyst design. *Journal of the American Chemical Society* **2004,** *126* (47), 15378-15379.

17. Timmons, D. J.; Doyle, M. P. in Multiple Bonds between Metal Atoms. (eds Cotton, A. F., Murillo, C. A. & Walton, R. A.) 591–632 (Springer, 2005)

18. Nicolas, I.; Le Maux, P.; Simonneaux, G., Asymmetric catalytic cyclopropanation reactions in water. *Coordination Chemistry Reviews* **2008,** *252* (5-7), 727-735.

19. Polgar, L., The prolyl oligopeptidase family. *Cellular and Molecular Life Sciences* **2002,** *59* (2), 349-362.

20. Harwood, V.; Denson, J.; RobinsonBidle, K.; Schreier, H., Overexpression and characterization of a prolyl endopeptidase from the hyperthermophilic archaeon Pyrococcus furiosus. *Journal of Bacteriology* **1997,** *179* (11), 3613-3618.

21. Harris, M.; Madura, J.; Ming, L.; Harwood, V., Kinetic and mechanistic studies of prolyl oligopeptidase from the hyperthermophile Pyrococcus furiosus. *Journal of Biological Chemistry* **2001,** *276* (22), 19310-19317.

22. Reddington, S.; Tippmann, E.; Jones, D., Residue choice defines efficiency and influence of bioorthogonal protein modification via genetically encoded strain promoted Click chemistry. *Chemical Communications* **2012,** *48* (67), 8419-8421.

23. Rea, D.; Fulop, V., Structure-function properties of prolyl oligopeptidase family enzymes. *Cell Biochemistry and Biophysics* **2006,** *44* (3), 349-365.

24. Kaszuba, K.; Rog, T.; Danne, R.; Canning, P.; Fulop, V.; Juhasz, T.; Szeltner, Z.; St Pierre, J.; Garcia-Horsman, A.; Mannisto, P.; Karttunen, M.; Hokkanen, J.; Bunker, A., Molecular dynamics, crystallography and mutagenesis studies on the substrate gating mechanism of prolyl oligopeptidase. *Biochimie* **2012,** *94* (6), 1398-1411.

25. Kichik, N.; Tarrago, T.; Claasen, B.; Gairi, M.; Millet, O.; Giralt, E., N-15 Relaxation NMR Studies of Prolyl Oligopeptidase, an 80 kDa Enzyme, Reveal a Pre-existing Equilibrium between Different Conformational States. *Chembiochem* **2011,** *12* (18), 2737-2739.

26. Juhasz, R.; Szeltner, Z.; Polgar, L., Truncated prolyl oligopeptidase from Pyrocaccus furiosus. *Proteins-Structure Function and Bioinformatics* **2007,** *69* (3), 633-643.

27. Fulop, V.; Bocskei, Z.; Polgar, L., Prolyl oligopeptidase: An unusual beta-propeller domain regulates proteolysis. *Cell* **1998,** *94* (2), 161-170.

28. Tarrago, T.; Martin-Benito, J.; Sabido, E.; Claasen, B.; Madurga, S.; Gairi, M.; Valpuesta, J.; Giralt, E., A new side opening on prolyl oligopeptidase revealed by electron microscopy. *Febs Letters* **2009,** *583* (20), 3344-3348.

29. Hansen, J.; Autschbach, J.; Davies, H., Computational Study on the Selectivity of Donor/Acceptor-Substituted Rhodium Carbenoids. *Journal of Organic Chemistry* **2009,** *74* (17), 6555-6563.

30. Carey, J.; Ma, S.; Pfister, T.; Garner, D.; Kim, H.; Abramite, J.; Wang, Z.; Guo, Z.; Lu, Y., A site-selective dual anchoring strategy for artificial metalloprotein design. *Journal of the American Chemical Society* **2004,** *126* (35), 10812-10813.

31. Zimbron, J.; Heinisch, T.; Schmid, M.; Hamels, D.; Nogueira, E.; Schirmer, T.; Ward, T., A Dual Anchoring Strategy for the Localization and Activation of Artificial Metalloenzymes Based on the Biotin-Streptavidin Technology. *Journal of the American Chemical Society* **2013,** *135* (14), 5384-5388.

32. Sambasivan, R.; Zheng, W.; Burya, S.; Popp, B.; Turro, C.; Clementi, C.; Ball, Z., A tripodal peptide ligand for asymmetric Rh(II) catalysis highlights unique features of on-bead catalyst development. *Chemical Science* **2014,** *5* (4), 1401-1407.

33. Durrenberger, M.; Heinisch, T.; Wilson, Y.; Rossel, T.; Nogueira, E.; Knorr, L.; Mutschler, A.; Kersten, K.; Zimbron, M.; Pierron, J.; Schirmer, T.; Ward, T., Artificial Transfer Hydrogenases for the Enantioselective Reduction of Cyclic Imines. *Angewandte Chemie-International Edition* **2011,** *50* (13), 3026-3029.

34. Kelly, S.; Jess, T.; Price, N., How to study proteins by circular dichroism. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2005,** *1751* (2), 119-139.

35. Bloom, J.; Labthavikul, S.; Otey, C.; Arnold, F., Protein stability promotes evolvability. *Proceedings of the National Academy of Sciences of the United States of America* **2006,** *103* (15), 5869-5874.

36. Miller, A.; Yikilmaz, E.; Vathyam, S., N-15-NMR characterization of His residues in and around the active site of FeSOD. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **2010,** *1804* (2), 275-284.

37. Legge, G.; Martinez-Yamout, M.; Hambly, D.; Trinh, T.; Lee, B.; Dyson, H.; Wright, P., ZZ domain of CBP: An unusual zinc finger fold in a protein interaction module. *Journal of Molecular Biology* **2004,** *343* (4), 1081-1093.

38. Trynda, L.; Pruchnik, F., Interaction Of Tetra-Mu-Acetatodirhodium(Ii) With Human Serum-Albumin. *Journal of Inorganic Biochemistry* **1995,** *58* (1), 69-77.

39. TryndaLemiesz, L.; Pruchnik, F., Studies on the interaction between human serum albumin and [Rh-2(OAc)(2)(bpy)(2)(H2O)(2)](OAc)(2). *Journal of Inorganic Biochemistry* **1997,**

66 (3), 187-192.

40. Norman, J.; Kolari, H., Strength And Trans Influence Of Rh-Rh Bond In Rhodium(Ii) Carboxylate Dimers. *Journal of the American Chemical Society* **1978,** *100* (3), 791-799.
41. Monnard, F.; Nogueira, E.; Heinisch, T.; Schirmer, T.; Ward, T., Human carbonic anhydrase II as host protein for the creation of artificial metalloenzymes: the asymmetric transfer hydrogenation of imines. *Chemical Science* **2013,** *4* (8), 3269-3274.

42. Coelho, P.; Brustad, E.; Kannan, A.; Arnold, F., Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science* **2013,** *339* (6117), 307-310.

43. Heel, T.; McIntosh, J.; Dodani, S.; Meyerowitz, J.; Arnold, F., Non-natural Olefin Cyclopropanation Catalyzed by Diverse Cytochrome P450s and Other Hemoproteins. *Chembiochem* **2014,** *15* (17), 2556-2562.

44. Bordeaux, M.; Tyagi, V.; Fasan, R., Highly Diastereoselective and Enantioselective Olefin Cyclopropanation Using Engineered Myoglobin-Based Catalysts. *Angewandte Chemie-International Edition* **2015,** *54* (6), 1744-1748.

45. Kornecki, K.; Briones, J.; Boyarskikh, V.; Fullilove, F.; Autschbach, J.; Schrote, K.; Lancaster, K.; Davies, H.; Berry, J., Direct Spectroscopic Characterization of a Transitory Dirhodium Donor-Acceptor Carbene Complex. *Science* **2013,** *342* (6156), 351-354.

46. Popp, B.; Ball, Z., Proximity-driven metallopeptide catalysis: Remarkable side-chain scope enables modification of the Fos bZip domain. *Chemical Science* **2011,** *2* (4), 690-695.

47. Chepiga, K.; Qin, C.; Alford, J.; Chennamadhavuni, S.; Gregg, T.; Olson, J.; Davies, H., Guide to enantioselective dirhodium(II)-catalyzed cyclopropanation with aryldiazoacetates. *Tetrahedron* **2013,** *69* (27-28), 5765-5771.

48. Davies, H. M. L. & Antoulinakis, E. G. Intermolecular metal-catalyzed carbenoid cyclopropanations.Org. React. **2013**, 57, 1–313.

49. Davies, H.; Hansen, T.; Churchill, M., Catalytic asymmetric C-H activation of alkanes and tetrahydrofuran. *Journal of the American Chemical Society* **2000,** *122* (13), 3063-3070.

50. Young, T.; Ahmad, I.; Yin, J.; Schultz, P., An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. *Journal of Molecular Biology* **2010,** *395* (2), 361-374.

51. Heckman, K.; Pease, L., Gene splicing and mutagenesis by PCR-driven overlap extension. *Nature Protocols* **2007,** *2* (4), 924-932.

CHAPTER III
DIRECTED EVOLUTION OF ARTIFICIAL METALLOENZYME FOR ENANTIOSELECTIVE CYCLOPROPANATION

The project was conducted in collaboration with Dr. Chen Zhang, Dr. Poonam Srivastava, Dr. Hyun June Park and Alan Swartz in the Lewis lab. Dr. Srivastava designed the initial version of library screening protocol (optimized by Dr. Zhang and Dr. Park) and screened the first-generation library; I designed another library screening protocol with protein immobilization (optimized by Dr. Park) and screened later generations of protein library; Alan Swartz performed deconvolution of library hits. For the sake of continuity, these results are presented herein and clearly designated.

ABSTRACT

Chapter II detailed how rational protein engineering could impart dramatic changes to the catalytic properties of artificial metalloenzymes. However, it is not clear how similar rational strategy could be used in general to engineer metalloenzymes for other particular applications given that a model to rationalize improvements observed in enantioselectivity is absent. Directed evolution would seem to offer the perfect solution to this problem. Despite great many examples in which natural enzymes have been evolved for synthetic applications, no examples of similar efforts (involving random mutagenesis throughout the entire scaffold gene) for ArMs have been reported. This chapter outlines the development of a streamlined, high-throughput protocol for design, expression,

purification and screening of POP-based ArM libraries for enantioselective cyclopropanation. Starting with POP-ZA_4-**1**, the β-propeller domain of *Pfu* POP has been subject to iterative random mutagenesis (through error-prone PCR) to generate ArM isoforms that were screened for enantioselective cyclopropanation. Through a few rounds of directed evolution, mutants with up to 94% were evolved, and key mutations both proximal and distal to the active site were discovered.

INTRODUCTION

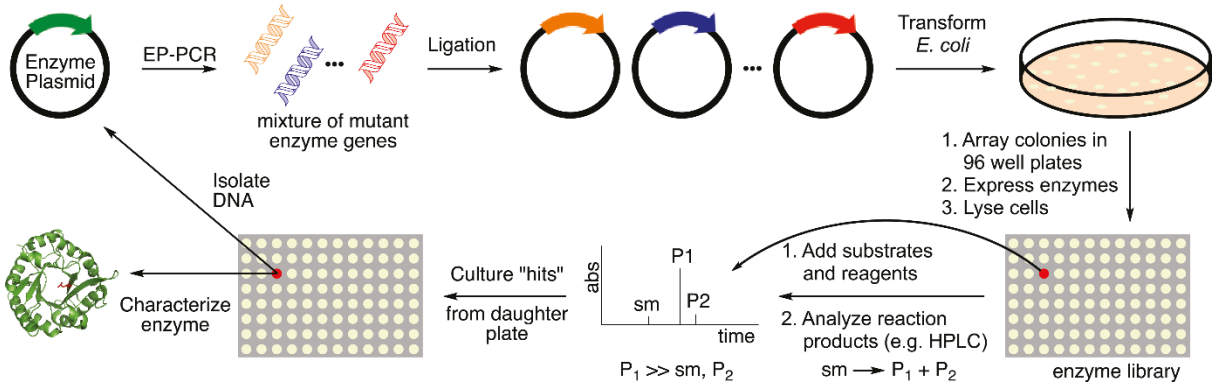
The potential utility of enzymes in synthetic chemistry, coupled with challenges associated with *de novo* design of such catalysts for particular applications, has driven intensive research efforts to improve existing enzymes through structure-guided rational modifications[1]. This protein engineering strategy is a powerful approach to optimize enzymes toward desirable functions, and has been used effectively many enzymes and artificial metalloenzymes. Chapter II detailed our efforts to engineer a dirhodium artificial metalloenzyme based on rational analysis of a homology model for a prolyl oligopeptidase (POP) and observed remarkable improvement in enantioselectivity and substrate specificity in biocatalysis[2].

However, the success of this method relies greatly on prior knowledge of the enzyme to be optimized, such as the crystal structure, the mechanism of enzyme activation/deactivation, enzyme dynamics, and so on[3]. Therefore, less-studied proteins may not benefit from the above approach. In fact, when we started to research the POP enzyme from *Pyrococcus furiosus*, a homology protein structure reported[4] was the only information we could derive from. Without a reliable model to rationalize improvements we observed in enantioselectivity, it is not clear how general the same strategy could be utilized to improve other functions of enzymes. Furthermore, whereas we were able to engineer POP-ZA_4-HFF-**1** to achieve high selectivity, our work entailed a large

amount of trial and error, and the highest selectivities and conversions for reactions catalyzed by this ArM were observed on the specific substrate used for screening purposes. Other substrates provided lower selectivity and conversion, and only one enantiomer of the product was obtained. It is not clear how POP scaffold could be further engineered to better suit other particular applications, such as different substrates, the other product enantiomer, or other dirhodium-catalyzed reactions (including C-H functionalization).

An alternative approach, namely directed evolution (Fig. 3.1), would seem to offer the best solution to this problem. Just as enzymes evolve in nature through many generations of mutations without prior knowledge of enzyme structure and function, directed evolution accumulate beneficial mutations for desired functions through iterative rounds of mutagenesis and screening[5]. This approach has been applied with great success to manipulate enzyme catalysts for essentially any catalytic properties such as thermostability, non-native reactivity, stereo-/regioselectivity, substrate scope[6].

Figure 3.1. The process of directed evolution of enzymes



Despite countless examples of improving enzyme functions through directed evolution in the literature, no examples in which directed evolution, involving random mutagenesis throughout the entire scaffold gene, have been reported to improve ArM function. Directed evolution of cytochrome P450 enzymes[7], nature's premier oxidation catalysts, is the most successful and

relevant example in this field. Nevertheless, ArMs formed via incorporation of unnatural metal-containing cofactors is very different from P450 enzyme, the latter being a mature catalyst evolved by nature with high efficiency and selectivity[8]. Reetz and Ward[9a-d] demonstrated seminal efforts to engineer streptavidin-based ArMs through scaffold mutagenesis on targeted sites proximal to the metal cofactor. Their screening systems, however, suffered from low-throughput library size due to poor expression efficiency of streptavidin and dependence on known crystal structure. Recently, Tezcan[9e] reported an artificial, supramolecular protein assembly with *in vivo* metallo-β-lactamase activity, which has allowed its functional screening and optimization via directed evolution. Contrasting with evolution efforts reported by Reetz and Ward (which was mainly toward improvements in enantioselectivity), the enzyme activity is directly linked to bacterial survival, thus a rapid *in vivo* selection was used to expedite the evolution process. A feature worth noting is that the screen was limited to targeted point mutations within a small active site, and higher screening efficiency might be achieved through simultaneous randomization of expanded enzyme domains.

Careful evaluation of existing ArM platforms indicates a few prerequisites should be fulfilled for implementing directed evolution on ArMs. First, scaffolds should be amenable to high throughput expression, mutagenesis, and other manipulations involved in standard *in vitro* evolution protocols. Second, bioconjugation methods should allow high throughput removal of unbound cofactor, which would otherwise catalyze non-selective background reactions. ArMs derived from POP via SPAAC bioconjugation fulfill all of these criteria. First, as mentioned in Chapter II, our study indicates POP is an incredibly stable enzyme. The introduction of extensive mutations and a dirhodium cofactor did not cause noticeable perturbation in the secondary structure, and the resulting hybrid catalyst even remained its secondary structure up to 100 ^{o}C,

which enabled scaffold purification via high temperature precipitation. Our initial expression trials suggests that POP can be expressed in *E. coli* on 96-well microtiter plates with minimal variations in expression yield for identical mutants. Second, the SPAAC reaction eliminates potential background interference from cell lysates due to its bio-orthogonality. This facilitates both introducing cofactor to the scaffold and removing excess cofactor from solution following bioconjugation. All these features of our ArM system prompted us to develop a high-throughput protocol for non-targeted ArM evolution. In addition to a workflow for screening free ArMs in solution, we also explored the use of enzyme immobilization in ArM directed evolution. The first catalytic property we intend to evolve in directed evolution is enantioselectivity, due to our established knowledge in POP-based ArMs described in Chapter II.

POP ENGINEERING THROUGH DIRECTED EVOLUTION

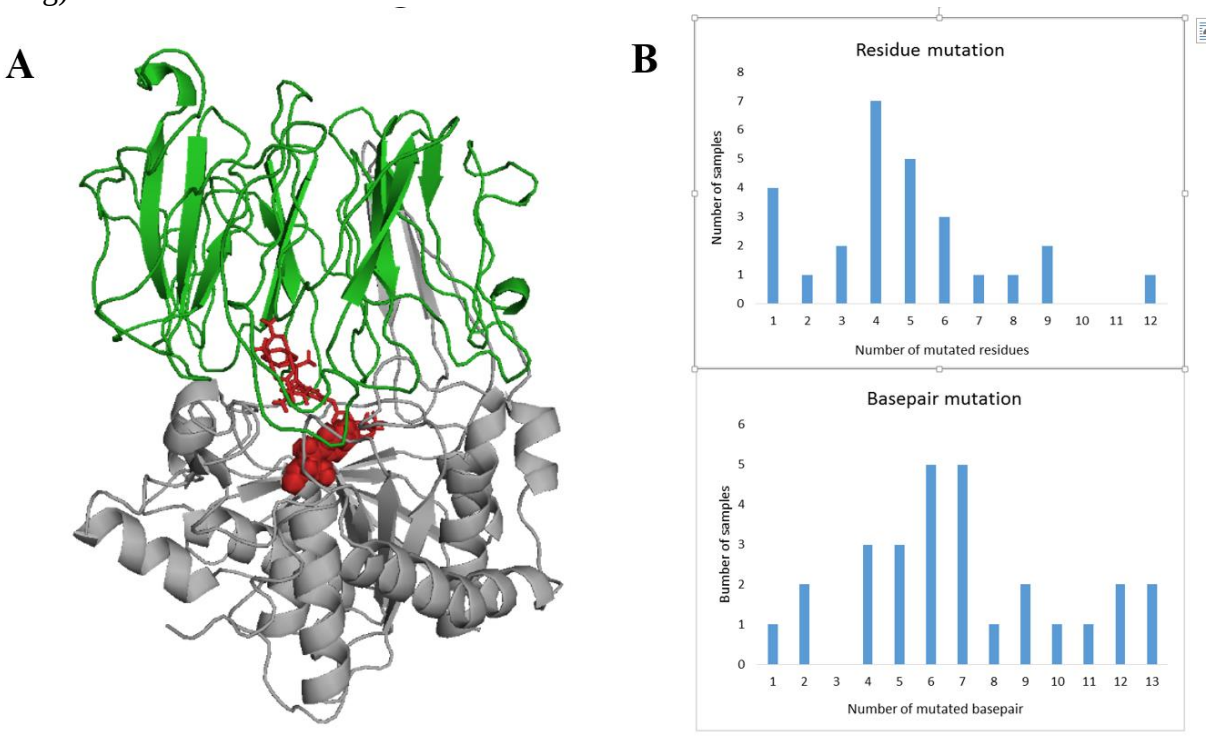
POP library design

Diverse methods have been employed to introduce genetic diversity for enzyme library creation, including error-prone PCR, saturation mutagenesis, and genetic recombination[10]. Since we aim to employ simple methods to look for effects of mutations not only in the active site but throughout a given scaffold protein, and since structural information required for planning a focused mutagenesis strategy do not exist for many potentially interesting scaffold proteins (e.g. thermostable variants of structurally characterized mesophiles), we used error-prone PCR to introduce diversity. This method has been widely used in similar situations due to high mutation rates, ease of use and fairly broad mutation spectrum[10]. To begin with, we sought to optimize the selectivity POP-ZA_4, an intermediate ArM scaffold identified in our rational engineering effort for cyclopropanation. This mutant not only provides a moderate selectivity, but serves as the minimal

ArM system required for evolution, as the presence of L-4-azidophenylalanine is needed for cofactor linkage and the "A_4" mutations were initially observed to facilitate bioconjugation.

First, we applied error-prone PCR to the POP β-domain[11] (Fig. 3.2A, green portion, from 48Q to 335V), which we believe encapsulates the dirhodium cofactor. While mutations could be made throughout the entire POP structure, we reasoned that those in the hydrolase domain (Fig. 3.2A, grey portion) would be less likely to impact cofactor selectivity. Addition of $MnCl_2$ to the PCR mixture allows for introducing errors, and a range of concentrations was examined. As expected, higher $MnCl_2$ concentration led to higher average mutation rates and lower DNA yields. A few initial library trials (data not shown) with an average of 1~2 amino acid mutations per mutant failed to provide hits with marked increase in selectivity. Considering the balance between high mutation rates and PCR efficiency, we used 300 μM $MnCl_2$ in the PCR and obtained libraries with an average of 4 amino acid mutations per variant (Fig. 3.2B). To assemble the mutated gene fragment with the complementary vector, we used Gibson assembly[12], a simple one-step ligation method to join the two DNA pieces. Compared with conventional restriction enzyme/ligation cloning, this method required fewer reagents, and more importantly saved much time for more challenging downstream operations in the screening protocol. The gene mixture was then used to co-transform *E. coli* along with pEVOL[13], which encodes the orthogonal tRNA and tRNA synthetase pair needed to genetically encode the L-4-azidophenylalanine residue for SPAAC.

Figure 3.2. (A) POP structure: β domain highlighted in green and hydrolase domain in grey; (B) residue/basepair mutations distribution using 300uM $MnCl_2$ (sequence data obtained by Dr. Chen Zhang).



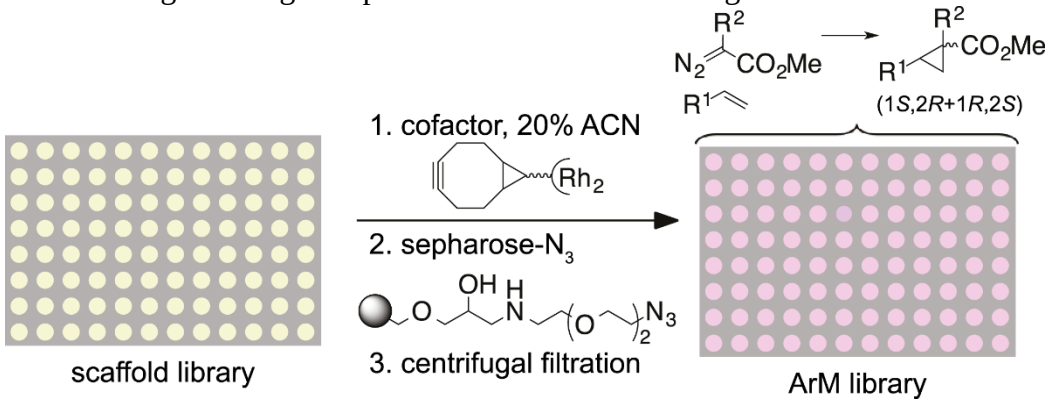
Library expression

The resulting library transformants were initially arrayed in the 96-well plates and expressed in the presence of L-4-azidophenylalanine to generate POP libraries containing members with the L-4-azidophenylalanine residue needed for metal incorporation and random mutations throughout their beta domains. Cell lysis, followed by heat treatment and centrifugation, provided cell lysate containing POP variants essentially free of other proteins. With the clarified cell lysates, we followed a screening workflow shown in Fig. 3.3A. A solution of dirhodium cofactor was added to initiate bioconjugation, which was followed by excess cofactor scavenging with an azide-substituted sepharose (explained later in the text). After centrifugation to remove sepharose, supernatants containing metalloenzymes were used in biocatalysis. We examined the biocatalysis of library mutants in cyclopropanation (Fig. 3.3B), from which very low ee (< 5%)

was observed. We reasoned the low selectivity in the library could result from either insufficient expression of POP scaffold in each well or addition of much excess cofactor that could lead to nonselective background catalysis. Dr. Park then analyzed POP concentration by gel electrophoresis throughout the screening process and found substantial protein loss occurred following cofactor scavenging (Fig. 3.3C).

Figure 3.3. (A). Workflow from bioconjugation to biocatalysis; (B). Reaction to be screened[a]; (C). SDS-PAGE monitoring of changes in protein concentrations through the workflow[b]

A



B

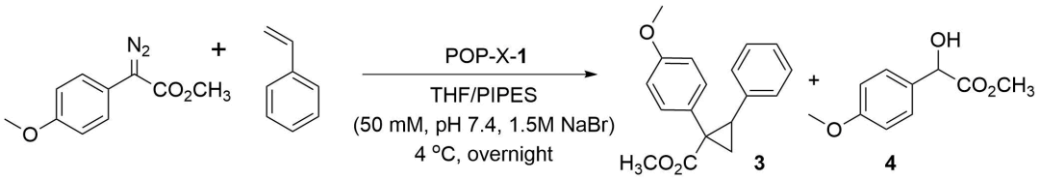
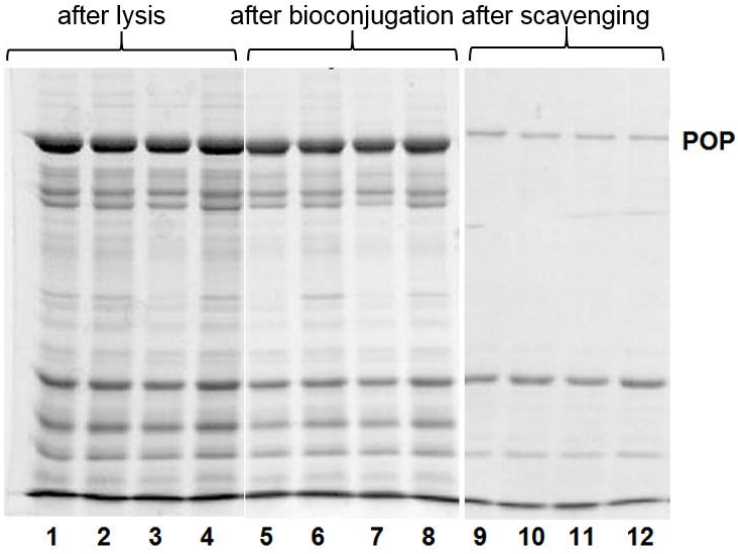


Figure 3.3. (A). Workflow from bioconjugation to biocatalysis; (B). Reaction to be screened[a]; (C). SDS-PAGE monitoring of changes in protein concentrations through the workflow[b], continued

C



[a] All reactions conducted using 4 mM diazoacetate and 20 mM styrene. Yield and enantioselectivity determined by HPLC relative to internal standard. [b] 4 repeat samples were assayed for three stages in the workflow respectively: after lysis (lanes 1-4), after bioconjugation (lanes 5-8), after cofactor scavenging (lanes 9-12).

To maximize scaffold expression in plates, a range of expression parameters (time, IPTG concentration, media, etc.) were systematically explored by Chen Zhang on both 96-well plates and 24-well plates. OD measurements revealed that total biomass in 24-well plates was 10-fold that obtained in 96-well plates (Table 3.1, entries 1 and 2). This is potentially due to better aeration in 24-well plate. Increasing volume to 6mL could further improved total biomass by 12-fold (Table 3.1, entry 3). Neither adding a pipette tip to improve aeration nor doubling broth nutrition were able to further improve expression levels (Table 3.1, entries 4 and 5).

Table 3.1. POP expression optimization

Entry	Plate type	Media broth	volume/mL	Shaking RPM	Final OD[a]
1	96-well	2YT	1	250	2.6

Table 3.1. POP expression optimization, continued

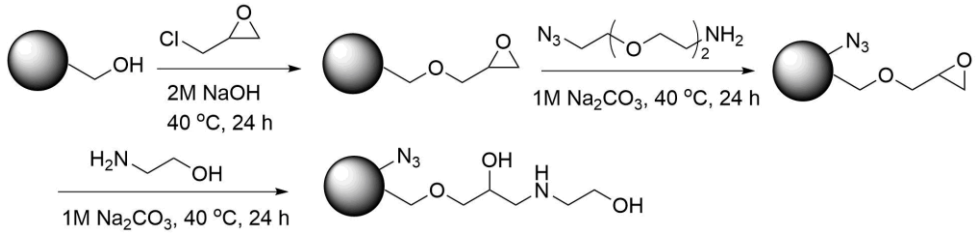
Entry	Plate type	Media broth	volume/mL	Shaking RPM	Final OD[a]
2	24-well	2YT	4	250	6.3
3	24-well	2YT	6	210	5.3
4[b]	24-well	2YT	6	210	5.0
5	24-well	double 2YT[c]	6	210	5.2

[a] OD detected by plastic cuvette under 600nm. [b]Autoclaved pipette tip added in each well. [c]doubled 2YT nutrition. The experiment was conducted by Chen Zhang.

Bioconjugation and cofactor scavenging

The high efficiency of SPAAC allowed for rapid bioconjugation of cofactor to these proteins in 96 well plates. Effective scavenging of residual cofactor is the essential step in the screening protocol (Fig. 3.3A). Crucially, we found that a commercially available azide-substituted sepharose resin[14] could be used to scavenge the excess cofactor with overnight incubation, and we established a facile synthesis of a comparable resin (Fig. 3.4). If this was not done, non-selective reaction catalyzed by unbound cofactor in solution competes with ArM catalysis, dramatically reducing the observed selectivity. The efficiency of cofactor scavenging was dependent on the amount of azide functional group loaded on the resin. To keep track of the sepharose resin we prepared, a coumarin-containing fluorescent bicyclononyne cofactor FBCN (Fig. 3.5A) was synthesized to monitor effective azide loading on resin. A fluorescence-based assay (excitation at 325 nm, emission at 500 nm) was developed by incubating FBCN with resin and measuring residual FBCN amount (Fig. 3.5B, 3.5C). According to this assay, the effective azide loading of the prepared resin was around 35~40 mM.

Figure 3.4. Synthesis of an azide-substituted sepharose resin



A potential problem associated with cofactor scavenging was that the use of sepharose resin could led to a significant decrease in the POP scaffold concentration (Fig. 3.3C), which caused low conversions in biocatalysis. To remedy the problem, Dr. Park optimized the amount of resin used in scavenging and monitored the change in protein concentrations and biocatalysis selectivity (Fig. 3.6). It was found that dropping the resin amount by 4-fold led to obtain a much higher ArM concentration following bioconversion and marked increase in cyclopropanation selectivity in ArM-catalyzed reactions as well.

Figure 3.5[a]. (A) FBCN structure; (B) calibration curve of FBCN, (C) schematic illustration of azide loading assay and (D) fluorescence intensity trace of azide loading assay

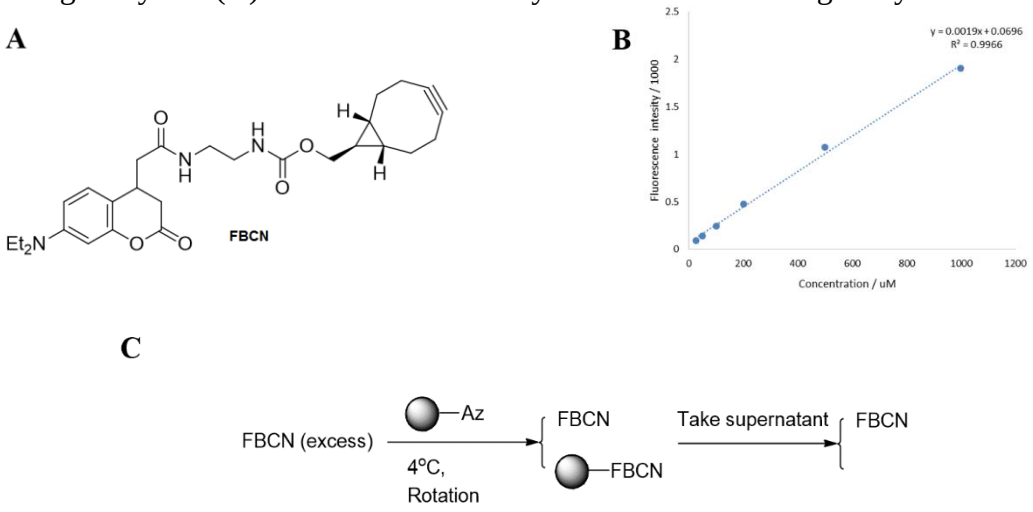
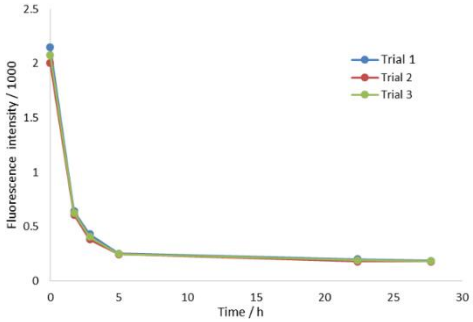


Figure 3.5[a]. (A) FBCN structure; (B) calibration curve of FBCN, (C) schematic illustration of azide loading assay and (D) fluorescence intensity trace of azide loading assay, continued

D



[a] The experiment was conducted by Chen Zhang.

Following removal of excess cofactor, the remaining portion of the workflow is essentially identical to that for natural enzymes. ArM catalyzed cyclopropanation was screened using HPLC or SFC[15], hits were identified based on desired criteria (high enantioselectivity, high conversion, etc.), and the genes encoding "hits" were used as parents for subsequent rounds of evolution. Combining all the modifications mentioned above to the general directed evolution workflow (Fig. 3.1), we developed a new workflow to enable directed evolution of POP ArMs (described in Fig. 3.7).

Figure 3.6[a]. (A). workflow from bioconjugation to biocatalysis; (B). SDS-PAGE monitoring of changes in protein concentrations through the workflow at different resin amounts; (C) summary of biocatalysis results in cyclopropanation obtained at different resin amounts

A

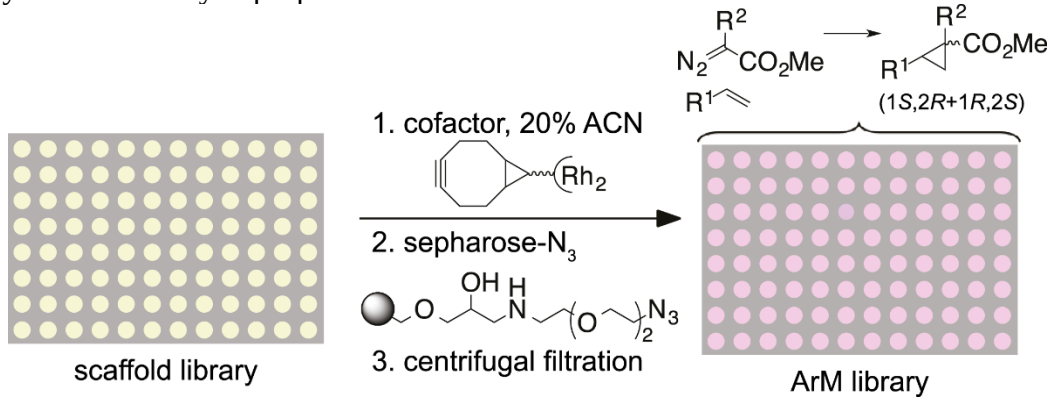
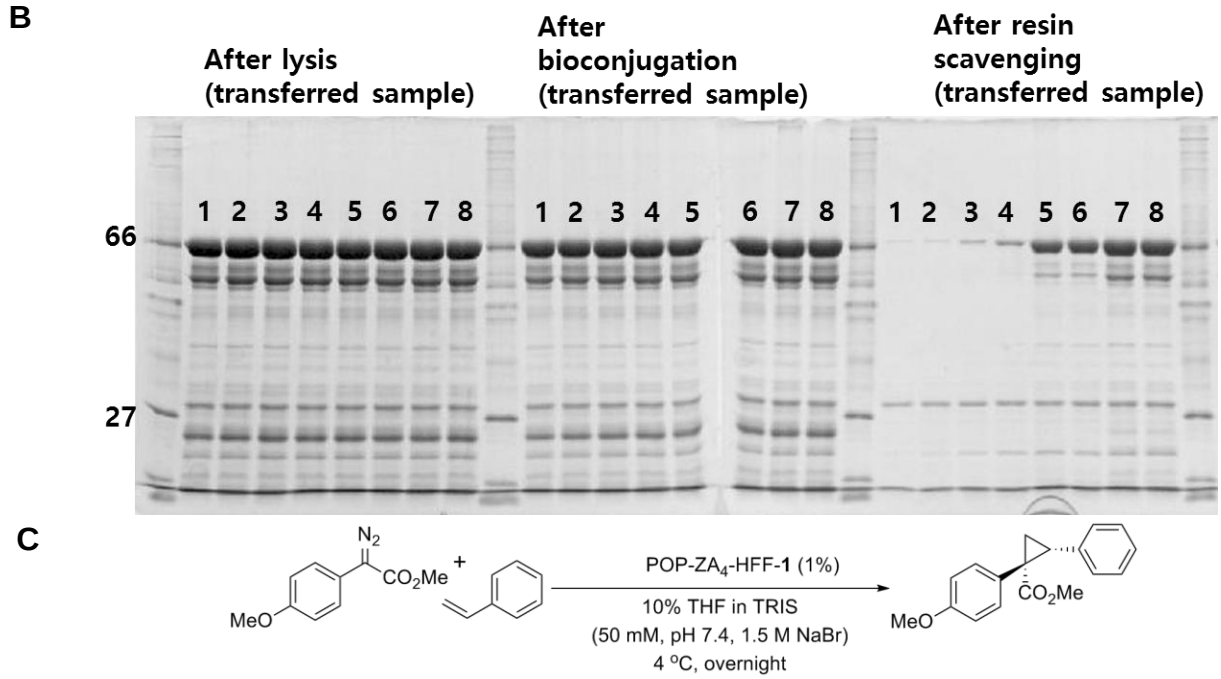


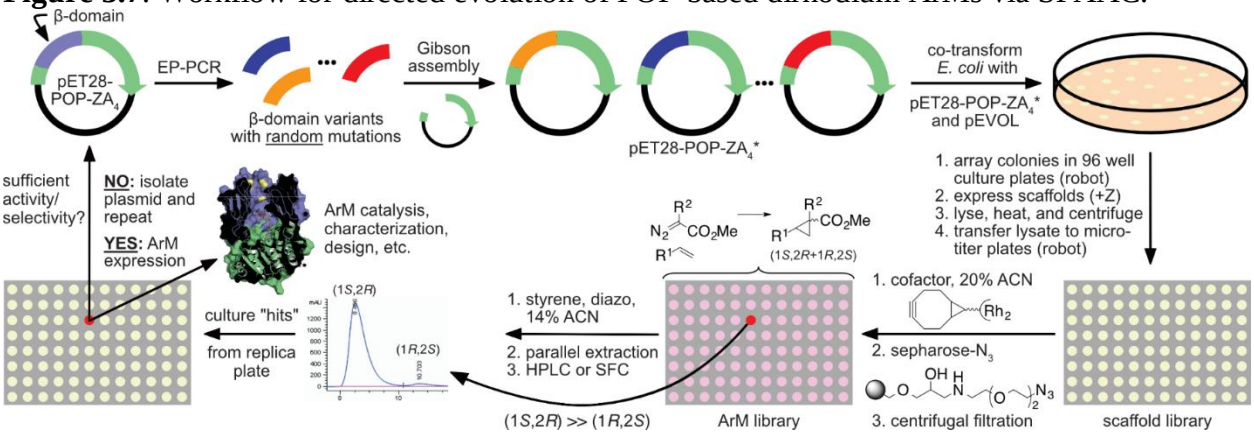
Figure 3.6[a]. (A). workflow from bioconjugation to biocatalysis; (B). SDS-PAGE monitoring of changes in protein concentrations through the workflow at different resin amounts; (C) summary of biocatalysis results in cyclopropanation obtained at different resin amounts, continued



entry[b]	1	2	3	4	5	6	7	8
resin/µl	200	200	100	200	50	50	25	25
ee%	31	32	38	48	52	55	40	47

[a] Dr. Park conducted the experiment and collected the data. [b] All reactions conducted using 4 mM diazoacetate and 20 mM styrene. Yield and enantioselectivity determined by HPLC relative to internal standard.

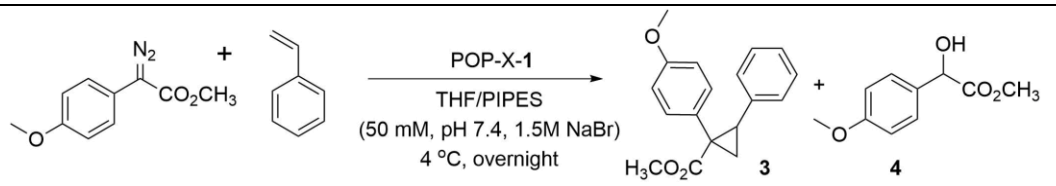
Figure 3.7. Workflow for directed evolution of POP-based dirhodium ArMs via SPAAC.



ArM evolution results

We employed cyclopropanation of aryldiazoacetates and olefins (Fig. 3.8A), the same reaction we discussed in Chapter II as a model reaction to test the utility of our library screening protocol to improve enantioselectivity (in Fig. 3.7). Using POP-ZA_4 as the starting parent, we used a high mutation level in our initial library (4 residue mutations/sequence) and screened approximately 96 variants against cyclopropanation. Remarkably, we identified two hits with significantly improved selectivities (Fig. 3.8). One variant 1-H possessing a single Y326H mutation in the active site provided 92% ee for styrene cyclopropanation. The mutation resembles the mutation L328H discussed in Chapter II, and conferred an even larger boost to selectivity. This indicates that directed evolution could be used to substitute the rational approach to engineer POP-based ArMs. The other mutant 1-NAGS gave 69% ee, and all the four mutations are outside of the POP active site, which makes their effects difficult to predict or rationalize and highlights the importance of random mutagenesis in ArM engineering. Deconvolution of the four mutations indicated one mutation S301G was particularly essential for ee enhancement (Table 3.2). An NNK library targeting the 301th position was generated and assayed (Table 3.3), and it was observed that quite a few amino acid substitutions could improve selectivity to the same extent.

Figure 3.8. Selectivities of dirhodium ArM variants obtained via directed evolution of POP-ZA_4



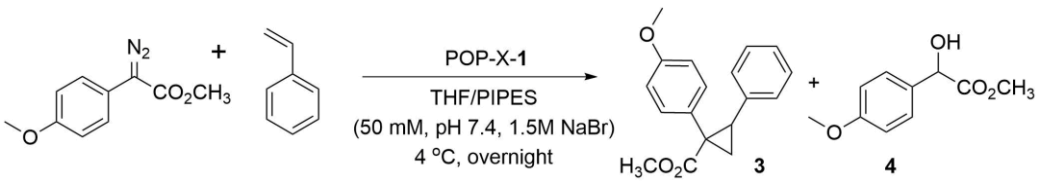
generation	enzyme	Mutation[a]	Method[b]	ee%	yield%	**3/4**
0	ZA_4	-	epPCR[c]	66	26	1.0
1	1-H	Y326H	epPCR	92	39	1.6
1	1-NAGS	K161N/V166A/S301G/T308S	epPCR	69	33	1.2
1	1-G	S301G	point mutation	75	40	1.3
2	2-NSIA	S84N/G99S/K330I/V335A	epPCR	77	37	1.3
3	3-VRVH	I221V/Q228R/A265V/Y326H	epPCR	94	47	1.8
3	3-AT	T211A/I221T	epPCR	80	38	1.5

Figure 3.8. Selectivities of dirhodium ArM variants obtained via directed evolution of POP-ZA$_4$, continued

[a] Mutations relative to the parent. [b] Method used to introduce mutations. [c]Error-prone PCR.

The second-round library was constructed as described above but with 1-NAGS as the parent, and around 96 variants were again screened. From this library, we identified a hit 2-NSIA (Fig. 3.8) that provided 77% ee in the cyclopropanation reaction. Afterwards, a third-round of evolution with 2-NSIA as the parent was conducted, and about 576 variants submitted to screening. Interestingly, despite starting from a different parent, another hit 3-VRVH (Fig. 3.8) that includes Y326H mutation was found in this round, giving >90% ee for cyclopropanation. Additionally, another hit 3-AT with two mutations was observed to provide slight improvement relative to its parent. All hits were verified by gene sequencing and large-scale reactions conducted with purified enzyme. Mutation deconvolution for these two hits is underway to elucidate the impacts of found mutations. Meanwhile, initial results on optimization of cyclopropanation reaction conditions revealed that the observed enantioselectivity could be tuned by varying reaction time, and systematic study is underway to understand such effects. In conclusion, our success in evolving ArMs suggests the potential to evolve ArM for particular applications in synthesis.

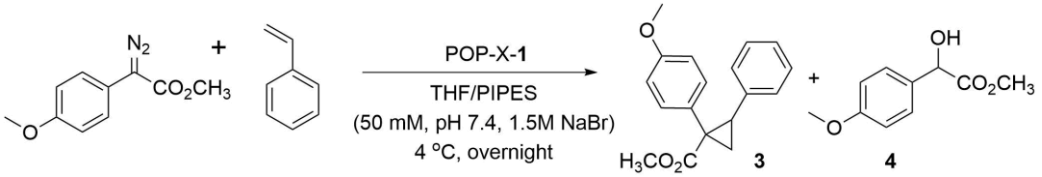
Table 3.2. Biocatalysis results for deconvolution of 1-NAGS



entry	mutation[a]	ee%[b]	conversion%	3/4
1	N161K	69	13	1.0
2	A166V	72	19	1.0
3	G301S	49	11	0.7
4	S308T	66	18	1.0
5	-	69	33	1.2

[a] Mutations relative to POP-ZA$_4$-NAGS. [b] The experiment was conducted by Alan Swartz.

Table 3.3. Biocatalysis results for hits in NNK library of POP-ZA$_4$-301X



entry	hit mutation[a]	ee%	conversion%	3/4
1	S301L	71	32	1.4
2	S301R	76	38	1.5
3	S301P	74	37	1.4
4	S301K	75	41	1.5
5	S301I	71	40	1.4
6	S301G	75	40	1.3

[a] Mutations relative to POP-ZA$_4$.

ENZYME IMMOBILIZATION AND ITS APPLICATION IN ARM SCREENING

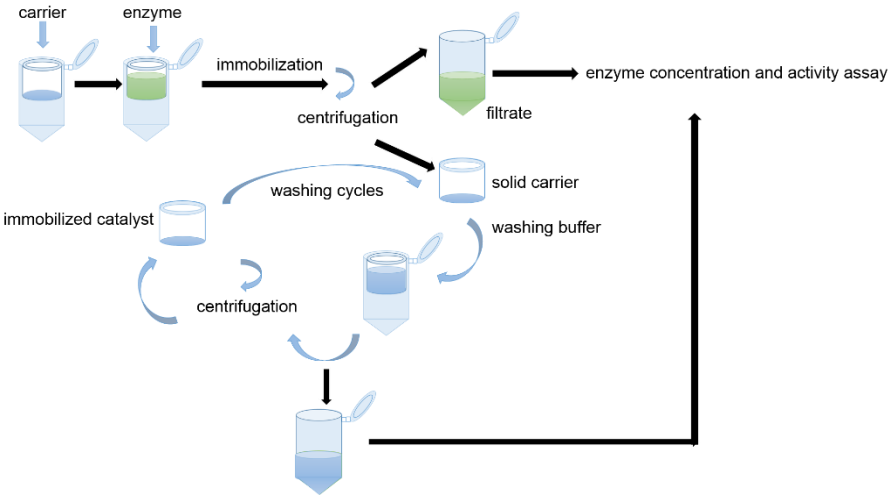
Whereas we were able to conduct ArM directed evolution with the above screening, there still existed a few potential problems. First, the intracellular azidophenylalanine released into the lysate mixture might react with metal cofactor to form adducts which cannot be scavenged by azide-substituted sepharose. This can cause nonselective background reactions and interfere with ArM catalysis. Second, the *E. coli* lysate contains a range of small metabolites[16] (e.g. glutathione), some of which have been reported to be detrimental to ArM catalysis[17]. It is difficult to determine which, if any of these may affect the biocatalysis outcome in our system, nor is the concentration of these species necessarily constant throughout the library, so that such effects could cancel. All these potential problems can be resolved if we achieve efficient separation of scaffold proteins from small-molecule lysate components. Also, from a practical perspective, the cofactor scavenging step involves tedious labor, time cost and use of expensive, disposable sepharose resin.

We reasoned that enzyme immobilization might provide a solution to all the mentioned problems[18]. Enzyme immobilization has been widely studied and utilized in industry-scale applications to improve enzyme operational stability, enable catalyst recovery and reuse, and modify enzyme reactivity/selectivity[19]. Specifically, for our application, a simple and efficient method for direct immobilization of POP scaffolds from cell lysates replete with azidophenylalanine and other metabolites is highly desirable. However, selection of a proper support and optimization of immobilization conditions are often an empirical process: different supports and experimental conditions were tested, until a satisfactory system has been developed[20]. To expedite the carrier selection process, we used a micro-scale procedure based on microcentrifuge filter tubes developed by Plou and his coworkers[21] (Fig. 3.9A) and screened a few commercially available solid supports. Briefly, 50-200 mg (or 50-200 µL for slurry) of the carrier is typically loaded on microcentrifuge filter tubes with cellulose acetate membrane (for aqueous applications). POP scaffold with the native Ser_{477} was added and the immobilization mixture was incubated for a certain time on a rotary mixer to facilitate the contact between carrier and enzyme. After centrifugation, the immobilization yield was determined by measuring the residual protein concentration of the filtrate via an activity assay based on hydrolysis of Z-Gly-Pro-PNP[22] (Fig. 3.9B). The retained solid was washed and its apparent activity could be determined by the same hydrolysis assay. Through screening of a few carrier supports, it was observed that Ni-NTA agarose resin gave the overall best performance in immobilization and the optimal resin amount required was determined (Fig. 3.10). Given the POP scaffold was the only protein with hexa-histidine tag in cell lysates, this support ensured selective immobilization of the POP enzyme in the presence of other background proteins (see Fig. 3.3). Several similar commercially available

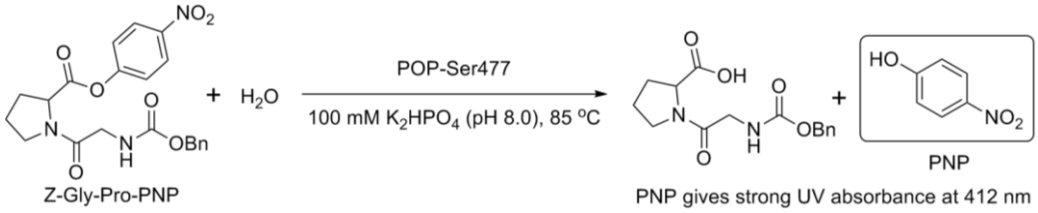
nitrilotriacetic acid-based agarose supports loaded with iron[23] were also tested to look for a better alternative to Ni-NTA agarose, but all showed much lower binding affinity.

Figure 3.9. (A). Micro-scale procedure to screen carriers and immobilization conditions[a]; (B) UV-absorbance assay based on POP-catalyzed hydrolysis of Z-Gly-Pro-PNP.

A



B



[a] The figure in 3.9 (A) was adapted from Ref. 21.

Figure 3.10. (A) Ni-NTA agarose: structure and interaction with proteins; (B) Calibration curve for PNP hydrolysis assay; (C) Determination of Ni-NTA agarose amount required for immobilization

A

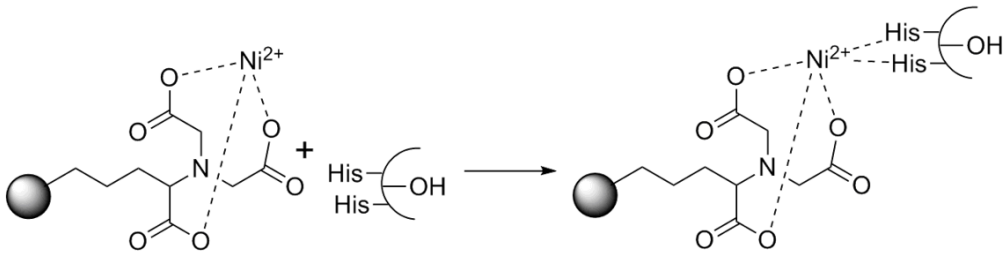
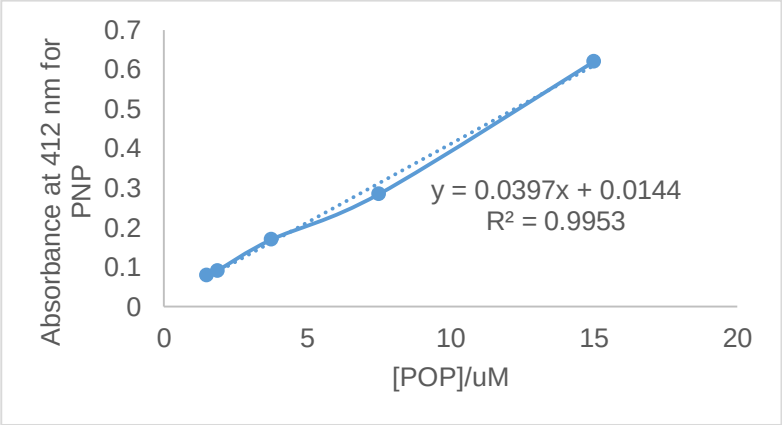


Figure 3.10. (A) Ni-NTA agarose: structure and interaction with proteins; (B) Calibration curve for PNP hydrolysis assay; (C) Determination of Ni-NTA agarose amount required for immobilization, continued

B



C

entry	resin amount/μL	A_{412}	POP% in filtrate	POP% immobilized
1	0	0.688	100	0
2	20	0.268	42.6	57.4
3	40	0.110	16.2	83.8
4	80	0.035	3.6	96.4
5	100	0.033	3.3	96.7

Afterwards, we made some modifications on the general screening procedure presented in Fig. 3.6 and studied the biocatalysis performance of immobilized ArMs (Fig. 3.11). The key challenge was to obtain effective cofactor scavenging by simply buffer washing so that nonselective background reaction can be avoided. Using POP-ZA_4-$F_{99}H_{328}$ (in purified form), a POP mutant we discovered in our rational engineering effort, we optimized a range of operational parameters in microcentrifuge tubes (Fig. 3.11A) to improve cyclopropanation selectivity (Fig. 3.12). It was observed that cofactor amount (Fig. 3.12B, entries 1-4), bioconjugation time (Fig. 3.12B, entries 8 and 9) and NaBr concentration in wash buffers (Fig. 3.12B, entries 7 and 8) all showed significant effects on cyclopropanation selectivity and conversions, while the influence of wash cycle numbers (Fig. 3.12B, entries 1, 5, 6) was minimal. Adjustment of cofactor amount (Fig.

3.12B, entries 10 and 11) led to cyclopropanation with improved selectivity and conversion (88% ee, 68% conversion), which is identical to the corresponding free ArM. However, the optimal condition for purified scaffolds delivered a much lower selectivity when used in handling crude cell lysate samples on the 96-well microtiter plate format (Fig. 3.11B), which is not surprising given the effective POP concentration in the latter was significantly lower despite optimization efforts on expression (Table 3.1). Thus, fine-tuning of operational parameters (mainly cofactor amount and cosolvent) were conducted (Table S3.2) to improve catalysis selectivity with given POP concentration on plates. Compared with the original protocol (Fig. 3.7), the immobilization protocol avoids interference from intracellular metabolites and use of azide-substituted sepharose, provides more automated and efficient screening, and offer the chance to extend ArM directed evolution to broader applications (e.g., catalysis in organic solvents). Currently, attempts to evolve POP ArMs for site-selective C-H insertion reactions using the immobilization protocol is underway.

Figure 3.11. (A). procedure for applying Ni-NTA immobilization in microcentrifuge tubes; (B) procedure for applying Ni-NTA immobilization in 96-well plates.

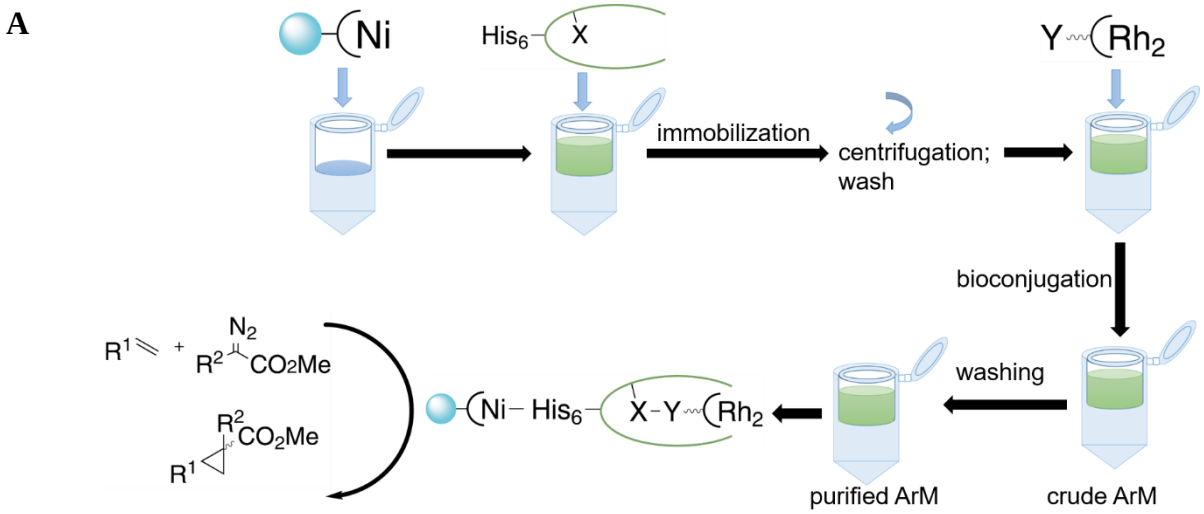


Figure 3.11. (A). procedure for applying Ni-NTA immobilization in microcentrifuge tubes; (B) procedure for applying Ni-NTA immobilization in 96-well plates, continued

B

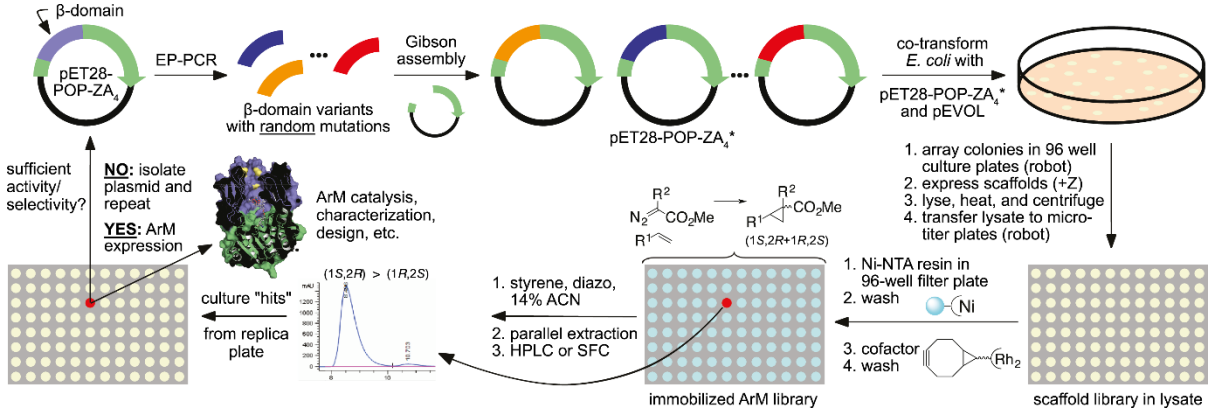
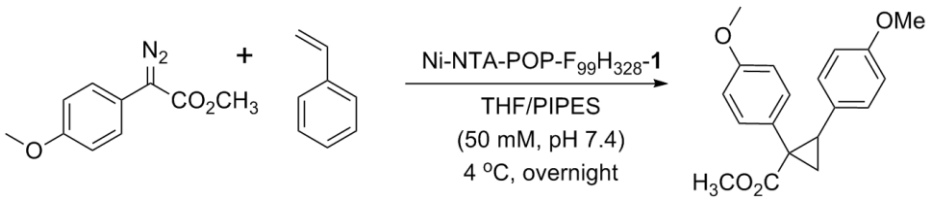


Figure 3.12. (A). the model reaction studied[a]; (B) optimization of operation parameters in the protocol

A



B

entry	RhBCN equiv	wash times	bioconjugation time	wash buffer additive	ee%	conversion%
1	2.5	10	5 h	-	15	17
2	1.25	10	5 h	-	18	14
3	0.625	10	5 h	-	21	10
4	0.375	10	5 h	-	27	7
5	2.5	5	5 h	-	14	14
6	2.5	3	5 h	-	16	11
7	1.25	5	5 h	1 M NaBr	20	22
8	1.25	5	5 h	1.75 M NaBr	24	23
9	1.25	5	24 h	1.75 M NaBr	51	36
10	0.625	5	24 h	1.75 M NaBr	63	46
11	0.3125	5	24 h	1.75 M NaBr	88	68

[a] All reactions conducted using 4 mM diazoacetate and 20 mM styrene. Yield and enantioselectivity determined by HPLC relative to internal standard.

DISCUSSION

The results outlined in this section are remarkable for several reasons. First, this marks the first example in which random mutagenesis was used to improve an ArM to our best knowledge. Second, given that one round of mutagenesis, screening, and hit validation currently takes about two weeks, the speed with which this can be accomplished is far greater than that associated with the trial and error (mostly error) of rational design. Third, the fact that our evolution efforts led to discovery of mutations both inside and outside of the active site clearly shows the importance of random mutagenesis for ArM engineering. There is currently no way to reliably predict how these distal mutations could have impacted enantioselectivity, and no other methods for ArM formation have proven compatible with the screening required to identify such mutations. Further computational and structural analysis is underway in our laboratory to rationalize these effects. Together, these remarkable results suggest that directed evolution will have an enormous impact on ArM engineering.

To date, we have demonstrated that directed evolution can be used to improve enantioselectivity, a catalytic function that can be achieved by small molecule transition metal catalysts as well. Given the complexity in design and preparation of ArM enzymes, and the fact that their performance is only comparable (or inferior in certain aspects) to small-molecule catalysts, our dirhodium enzyme will not be a first choice for synthetic chemists. The real challenge is to evolve ArMs that can be used in a practical manner in those applications in which conventional metal catalysts fail[24]. We are therefore continuing with these efforts, with a strong focus on evolving POP variants with expanded scope, the unknown selectivity, and novel reactivity.

ACKNOWLEDGEMENT

We thank Professor Jeffry Madura at Duquesne University for providing us with the POP homology model used in these studies, Professor Harold Schreier at the University of Maryland, Baltimore County for providing the *Pfu* POP gene, Prof. Peter G. Schultz for providing pEVOL-pAzF. We also thank Dr. Jin Qin at the UC Department of Chemistry Mass Spectroscopy Facilities, and Dr. Elena Solomaha at the UC Biophysics Core Facility for assistance.

EXPERIMENTAL

1. Materials

Unless otherwise noted, all reagents were obtained from commercial suppliers and used without further purification. Deuterated solvents were obtained from Cambridge Isotope labs. Silicycle silica gel plates (250 mm, 60 F254) were used for analytical TLC, and preparative chromatography was performed using SiliCycle SiliaFlash silica gel (230-400 mesh). Azide Agarose 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). Plasmid pEVOL-pAzF was provided by the Schultz group of the Scripps Research Institute, CA[S1]. *E. coli* DH5α and BL21 (DE3) cells were purchased from Invitrogen (Carlsbad, CA). Nco I, Xho I restriction enzyme, T4 DNA ligase, Taq DNA polymerase and Phusion HF polymerase (Cat# 530S) were purchased from New England Biolabs (Ipswitch, MA). Luria broth (LB), rich medium (2YT) and Agar media were purchased from Research Products International (Mt. Prospect, IL). Qiagen DNA extraction kit (Cat# 28706) and plasmid isolation kit (Cat# 27106) were purchased from QIAGEN Inc. (Valencia, CA) and used according to the manufacturer's instructions. DNA purification kit (Zymo, Cat# D4004) was purchased from

Zymo research (Irvine, CA) and used as recommended. Library colonies were picked using an automated colony picker (Norgren Systems). 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 Ec3. Ninitrilotriacetic acid (Ni-NTA) resin and Pierce® BCA Protein Assay Kits were purchased from Fisher Scientific International, Inc. (Hampton, NH), and the manufacturer's instructions were following when using both products (for Ni-NTA resin, 5 mL resin were used, with buffers delivered by a peristaltic pump at a rate of 1 mL/min, in a 4 °C cold cabinet). Amicon® 30 kD spin filters for centrifugal concentration were purchased from EMD Millipore (Billerica, MA) and used at 4,000 g at 4 °C. Biotage reverse phase columns (SNAPKP-C18-HS) were purchased from Biotage. 96-well filter microplates (Cat# 201009-100) were purchased from Seahorse Bioscience (North Billerica, MA). Microcentrifuge filter tubes were from the QIAquick® Gel extraction kit (Cat# 28706) from QIAGEN Inc. (Valencia, CA).

2. General procedures

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 as the internal standard and reported as the average of two 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. Low resolution ESI mass spectra were obtained using Agilent 6130 LC-MS. Amicon® 50 mL 30 kD cutoff centrifugal filter was 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. Circular dichroism (CD) spectra were obtained on a JASCO J-1500 CD Spectrometer. Standard molecular cloning procedures were followed. To introduce single mutations, the same procedure using site directed overlap extension PCR reported in Chapter II was followed.

3. Experimental procedure for ArM evolution using azide-substituted sepharose

Library construction

A codon optimized gene for Prolyl oligopeptidase (POP) was obtained from GenScript USA Inc (Piscataway, NJ) and cloned into pET28a plasmid vector using NcoI and XhoI restriction sites. The gene was cloned upstream of a C-terminal hexa-histidine tag for Ni-NTA affinity chromatography. Mutant libraries were constructed by conducting error-prone PCR for the β-domain and regular PCR for the vector from a template plasmid (the library parent). The error-prone PCR conditions for mutating the target domain were as follows: Taq polymerase buffer 10x, 0.2 mM dNTPs each, 0.5 μM forward primer (ep-forward), 0.5 μM reverse primer (ep-reverse), 0.3 mM $MnCl_2$, 0.05 U/μL Taq polymerase and 1.0 ng/mL template plasmid.

Thermal cycler was programmed as:

1. 98 °C-120 seconds
2. 95 °C-45 seconds
3. 52 °C-30 seconds
4. 68 °C- 90 seconds

5. Repeat cycles from #2 to #4 25 times

6. 68 ˚C- 10 mins

7. 4 ˚C- hold

The regular PCR conditions for amplifying the vector were as follows: Phusion polymerase buffer 1x, 0.2 mM dNTPs each, 0.5 µM forward primer (vec-forward), 0.5 µM reverse primer (vec-reverse), 0.02 U/µL Phusion HF polymerase and 1.0 ng/mL template plasmid.

Thermal cycler was programmed as:

1. 99 ˚C-60 seconds

2. 95 ˚C-30 seconds

3. 54 ˚C-45 seconds

4. 72 ˚C- 130 seconds

5. Repeat cycles from #2 to #4 25 times

6. 72 ˚C- 10 mins

7. 4 ˚C- hold

The resulting two overlapping fragments (the mutated domain and the vector) were then assembled using Gibson assembly kit resulting in the full-length mutated gene. The reaction conditions were as follows: in a 200 µL PCR tube, prepare a 5:1 (molar ratio) mixture of mutated domain and vector (total mass of the vector is ~50 ng) in 10 µL molecular grade water. Add 10 µL 2X Gibson Assembly Master Mix in a final volume of 20 µL and incubate the reaction at 50°C for 60 minutes. The reaction mixture was cleaned using DNA purification kits and transformed into *E. coli* BL21 cells with pEVOL-pAzF plasmid. Cells were spread on LB kanamycin plates (6.25 g LB powder mix, 4 g agar, 250 mL DDI water, 0.05 mg/mL kanamycin, 0.02 mg/mL chloramphenicol) before recovering in SOC medium for 1 hour at 37 ˚C. Plates were incubated at 37 ˚C overnight; typically,

greater than 100 colonies were observed when 1/5 of the outgrowth was used. To examine the mutant rates for a library, around 20 colonies were inoculated on LB media (with 0.05 mg/mL kanamycin, 0.02 mg/mL chloramphenicol) and grown overnight at 37 ˚C, 250 rpm. Mutant plasmid from these overnight grown cultures were isolated using kit from Qiagen (Valencia, CA) and given for sequencing. Plasmid sequencing was done at the U Chicago sequencing facility, and partB-for and partC-rev primers were used for sequencing reactions.

Nucleotide sequences for all the primers used above are summarized in Table S3.1.

Table S3.1. Nucleotide sequences for the primers.

Primer name	Primer Sequence
ep-forward	5'- CTG AGT GAT AAA CTG TTT CCG GAA GTG TG -3'
ep-reverse	5'- AGA CGA TAC GGA ATC GTA AAC GAG GTG T -3'
vec-forward	5'- GCG CTA CAC CTC GTT TAC GAT TCC GTA TC -3'
vec-reverse	5'- GGG AGA ACT GTT CCC ACA CTT CCG GAAA -3'
partB-for	5'- TCT GGA TGG AAA ACC TGG AA -3'
partC-rev	5'- TGC AAT GAA GTC ATC GAA CA -3'

Library expression, lysis and purification

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, 2~4 wells were saved for picking parent colonies as positive controls, and 2 wells were left blank as negative controls. Cells were grown overnight for 14~16 hours at 37 ºC, 250 rpm. 50 μL of the overnight culture was used to inoculate 6 mL 2YT media (with 50 μg/mL kanamycin and 20 μg/mL chloramphenicol) in 10-mL 24-well plates. 2~4 parent mutant were inoculated in each 24-well plates as positive controls. Following growth at 37 ºC, 200 rpm for about 6 hours, to an OD_{600} = 1.3~1.4, enzyme expression was induced by adding 1mM IPTG, 2mM 4-Azido-phenyl alanine 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 2000 rpm, 20 mins and the supernatants were discarded. The cell pellets were washed by adding 4 mL Tris-Cl buffer (50 mM, pH 7.4) to each well and incubating the plate at 37 ºC, 200 rpm (for 10 mins) until the pellets were loosen up, after which the suspension was submitted to centrifugation at 2000 rpm, 10 min. The supernatants were discarded and the above washing process was repeated again.

The washed cell pellets were suspended in 600 μL Tris-Cl buffer (50 mM, pH 7.4) containing 0.75 mg/mL lysozyme. After incubation plate at 37 ºC, 250 rpm for 60 mins, cells were flash frozen in liquid nitrogen for 10 mins and thawed at 37 ºC water bath. 60 μL Tris-Cl buffer (50 mM, pH 7.4) containing 1.0 mg/mL lysozyme was added and the plate was incubated at 37 ºC, 250 rpm for 30 mins. The lysed cells were submitted to a heat treatment for 15 mins at 75 ºC water bath, after which the plate was centrifuged at 3600 rpm, 20 mins.

Library bioconjugation and cofactor scavenging

480 μL of the lysate supernatant was transferred to a 2-mL 96-well plates for bioconjugation. Additionally, 10 μL of the lysate supernatants from randomly-picked wells were saved in a 1.5-mL microcentrifuge tube for electrophoresis analysis.

For bioconjugation, 120 μL of cofactor **1** in acetonitrile (93.75 μM) was added to lysate of each mutant and the reaction mixture was incubated at 4 ºC, 600 rpm overnight. 10 μL of the bioconjugation mixtures from wells picked above were saved in a 1.5-mL microcentrifuge tube for electrophoresis analysis. 50 μL azide agarose resin was then added. The plate was sealed tightly with a cap mat (Arctic white LLC, cat#AWSM-1003S), wrapped with aluminum foil, and rotated on the Labquake™ Tube Shaker/Rotator in a 4 °C cold cabinet for 24 h to scavenge excess cofactor. The resulting suspension was centrifuged at 3600 rpm, 10 mins. 300 μL of the supernatant was

carefully transferred to a new 2-mL 96-well plate with a Microlab NIMBUS liquid handling robot. Additionally, 10 μL of the supernatants from wells picked above were saved in a 1.5-mL microcentrifuge tube for electrophoresis analysis.

Library screening

150 μL Tris-Cl buffer (50 mM, pH 7.4, with 4.5 M NaBr) was added to each well and the plate was incubated at 4 °C, 20 mins. A combined stock solution containing 4-methoxystyrene (2.64 mM final concentration) and methyl 4-methoxyphenyldiazoacetate (0.53 mM final concentration) was added to initiate biocatalysis. The plate was sealed with a plastic lid and incubated overnight at 4 °C, 600 rpm. The reaction was quenched by adding 300 μL hexane and incubated at 600 rpm, 10 mins. The biphasic mixture was clarified by centrifugation at 2000 rpm, 10 mins, after which the plate was flash frozen with liquid nitrogen for 10 mins and kept at -20 °C for 10 mins. 300 μL from the top hexane layer was carefully transferred to a 200-μL 96-well polypropylene plate without interrupting the frozen aqueous layer. Hexane was evaporated in a vacuum desiccator at room temperature and 120 μL 10% isopropanol/hexane was added to redissolve the mixture. The reactions were filtered through a 96-well filter plate (PTFE, 0.45 μm) and analyzed for cyclopropanation on HPLC. The chiral-HPLC to determine enantioselectivity was performed on 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: 90% B from 0-10 min, 90 % B to 80% B from 10-15 min, 80 % to 75 % B from 15-18 min, 75 % B from 18-22 min, 75% to 90% from xx-xx min, 4 min post-run (solvent A: isopropanol; solvent B: hexane).

4. Experimental procedure for ArM evolution using Ni-NTA agarose

Library construction, expression, lysis and purification was conducted exactly in the same way as described above.

Library bioconjugation and cofactor scavenging

480 μL of the lysate supernatant was transferred to a 2-mL 96-well plates for bioconjugation. Additionally, 10 μL of the lysate supernatants from randomly-picked wells were saved in a 1.5-mL microcentrifuge tube for electrophoresis analysis. 100 μL Ni-NTA agarose slurry was added to each well and the plate was incubated at 4 ^{o}C, 600 rpm for 1 h. The plate was centrifuged at 3600 rpm, 10mins and resuspended in 480 μL Tris-HCl buffer (50 mM, pH 7.4). For bioconjugation, 120 μL of cofactor **1** in acetonitrile (46.875 μM) was added to lysate of each mutant and the reaction mixture was incubated at 4 ^{o}C, 600 rpm for 24 h. 10 μL of the bioconjugation mixtures from wells picked above were saved in a 1.5-mL microcentrifuge tube for electrophoresis analysis. The plate was then centrifuged at 3600 rpm, 10 mins. Into the slurry 1 mL of 20% acetonitrile in Tris-HCl buffer (50 mM, pH 7.4) was added and centrifuged again at 3600 rpm, 10 mins. The buffer wash was repeated three times, and the resulting slurry was resuspended in PIPES buffer (50 mM, pH 7.4, 1.75 M NaBr).

Library screening

A combined stock solution containing 4-methoxystyrene (2.64 mM final concentration) and methyl 4-methoxyphenyldiazoacetate (0.53 mM final concentration) was added to initiate biocatalysis. The plate was sealed with a plastic lid and incubated overnight at 4 °C, 600 rpm. The reaction was quenched by adding 300 μL hexane and incubate at 600 rpm, 10 mins. The plate was centrifuged at 3600 rpm, 10 mins and a 96-well microtitier plate was used as a receiver plate to collect the biphasic mixture. The receiver plate was flash frozen with liquid nitrogen for 10 mins and kept at -20 °C for 10 mins. 300 μL from the top hexane layer was carefully transferred to a

200-μL 96-well polypropylene plate without interrupting the frozen aqueous layer. Hexane was evaporated in a vacuum desiccator at room temperature and 120 μL 10% isopropanol/hexane was added to redissolve the mixture. The reactions were filtered through a 96-well filter plate (PTFE, 0.45 μm) and analyzed for cyclopropanation on HPLC.

5. Hits purification and verification

The putative hits identified from the library screening was grown, expressed, lysed, and purified according to a previous report[2]. A colony for the selected mutant was inoculated in 5 mL 2YT medium having antibiotics with the same concentrations as above. The culture was incubated overnight at 37 °C 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 5 L Erlenmeyer flask. The culture was incubated at 37 °C, 250 rpm, and protein expression was induced by adding 1mM IPTG, 2mM 4-Azido-phenyl alanine and 1% (w/v) L-arabinose when OD_{600} reached 1. The induced culture was allowed to grow for 12 hours, and then the cells were harvested by centrifugation at 4 °C, 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 °C for 30 minutes and supernatant thus obtained was purified by Ni-NTA resin using manufacturer's instructions. Purified protein was buffer exchanged to 10 mM Tris (pH 7.5) and measured by Pierce® BCA Protein Assay Kit as recommended.

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 **1** (120 μL, 0.75 mM in ACN, 0.655 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 **1**, 20 vol% acetonitrile/Tris 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 h to remove excess cofactor. The suspension was then centrifuged at 5000 rpm for 3 min and the supernatant was transferred to a new microcentrifuge tube. The resin was rinsed twice with 600 μL 50 mM Tris-HCl buffer and centrifuged at 5000 rpm for 3 min. These supernatants were combined with the first supernatant and buffer exchanged to proper buffers for use in biocatalysis or characterization. ESI-MS were used to characterize the bioconjugates. The total protein concentration was calculated based on its absorbance at 280 nm (A280) 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 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 ([ArM]=[Total protein]*($I_{ArM}/(I_{ArM}+I_{scaffold})$). The effective ArM loading was adjusted to 1 mol% with respect to the dirhodium cofactor in bioconversions.

To set up biocatalysis, solutions of aryldiazoacetate (25 μL, 96 mM, in THF), styrene (25 μL, 485 mM, in THF), and POP-ZA_4-X-**1** 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- ZA4-X-1. The resulting mixture was left shaking at 750 rpm at 4 ℃ overnight. The reaction was quenched by adding 20 μL 1,2,4-trimethoxybenzene solution (30 mM, in THF) and 600 μL ethyl acetate. The mixture was vortexed and centrifuged

(15,000 x g, 3 min). 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 THF. 4 µL THF solution of the crude product was analyzed on RP-HPLC to determine conversions; 50 µL THF solution of the crude product was purified on preparative-HPLC to isolate the cyclopropane product, which was analyzed on NP-HPLC to determine enantioselectivities. The conversions and enantioselectivities were reported as the average of two trials from the same batch of ArM set up in parallel. The RP-HPLC to determine conversions was performed on an Agilent 1100 Series HPLC system using an Agilent Eclipse Plus C18 column (95 Å, 3.5 µM, 4.6 mm i.d. x 150 mm), with 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: water containing 0.1% TFA; solvent B: CH_3CN). The preparative HPLC used the same method as above. The NP-HPLC to determine enantioselectivities was performed on 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.

6. Procedure for Ni-NTA Immobilization in Microcentrifuge Tubes

A general procedure for optimizing resin immobilization in microcentrifuge tubes is described as follows: a certain amount of carrier of interest was placed inside a microcentrifuge filter tube with 0.45 µm cellulose membrane and washed twice with 500 µL immobilization buffer to equilibrate the support. The washing step included the mixture of the carrier with the buffer by closing the microcentrifuge filter tube and inverting five times, followed by centrifuging at 1000 rpm, 1min to separate the washing buffer from the solid particles, which remained inside the filter holder.

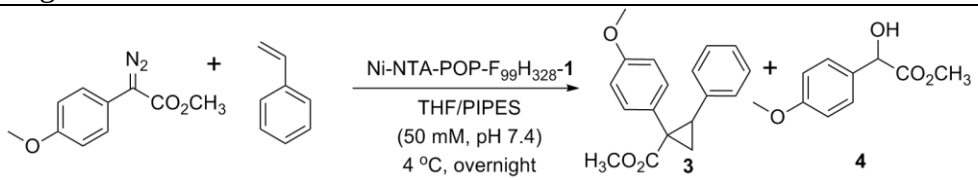
After washing, the carrier support was mixed (inside the filter holder) with 480 µL POP solution (purified enzyme or crude cell lysates) in an immobilization buffer (50 mM Tris-Cl buffer, pH 7.4). After incubation at room temperature for a certain period of time, the filter tube was centrifuged and the filtrate was collected. 30 µL of the filtrate was diluted to 900 µL with the immobilization buffer. A mixture of 940 µL reaction buffer (100 mM KH_2PO_4, pH 7.0) and 40 µL Z-Gly-Pro-PNP in methanol were mixed and heated at at 85 ^oC, 750 rpm, for 3 mins. 20 µL of the diluted enzyme was added to the preheated mixture and the reaction mixture was heated at 85 ^oC, 750 rpm, for 1 min. 900 µL of the reaction mixture was transferred to a cuvette for measurement of UV-vis absorbance at 412 nm on a TECAN Infinite® 200PRO microplate reader.

A general procedure to optimize Ni-NTA immobilization for ArM catalysis in microcentrifuge tubes is described as follows:

80 µL Ni-NTA agarose was placed inside a microcentrifuge filter tube with 0.45 µm cellulose membrane and washed twice with 500 µL immobilization buffer to equilibrate the support. The washing step included the mixture of the carrier with the buffer by closing the microcentrifuge filter tube and inverting five times, followed by centrifuging at 1000 rpm, 1min to separate the washing buffer from the solid particles, which remained inside the filter holder. After washing, the agarose was mixed (inside the filter holder) with 480 µL POP solution (purified enzyme or crude cell lysates) in an immobilization buffer (50 mM Tris-Cl buffer, pH 7.4). After incubation at room temperature for 1.5 h on a roller mixer, the filter tube was centrifuged and the filtrate was collected to test if any residual protein was left. The agarose was resuspended with 480 µL Tris-Cl buffer (50 mM, pH 7.4), 120 µL of RhBCN cofactor at a certain concentration was added, and the tube was incubated at 4 ^oC overnight on a roller mixer. Afterwards, the reaction mixture was centrifuged and washed with 600 µL of a wash buffer for a few times. The immobilized ArM was carefully

transferred to another 1.5 mL microcentrifuge tube in which 500 µL biocatalysis buffer (50 mM PIPES, pH 7.4, 1.75 M NaBr) was added. A solution of aryldiazoacetate (25 µL, 96 mM, in THF) and styrene (25 µL, 485 mM, in THF) was added to the ArM and the tube was incubated at 4 °C overnight on a roller mixer. The biocatalysis result was examined using HPLC according to the same procedure described above.

Table S3.2. Optimization of Ni-NTA immobilization procedure for crude cell lysates in microcentrifuge tubes



entry	mutant	RhBCN equiv	ee%[a]	conversion%	3/4
1	POP-ZA_4	0.3125	22	25	0.43
2	POP-ZA_4	0.15625	23	25	0.46
3	POP-ZA_4-H_{326}	0.3125	44	32	0.52
4	POP-ZA_4-H_{326}	0.15625	50	35	0.58
[a] ee%, conversion%, and 3/4 are based on average of two trials.					

7. Mutation deconvolution and NNK library

To create deconvolution mutations, the required single mutation was introduced using the same procedure for site-directed overlap extension PCR reported in Chapter III. The hits purification and verification procedure described in **5** was applied to each mutant to examine its performance in biocatalysis.

To construct NNK library for a specific site, the NNK primers were designed (see Table S3.3, entries 19-20). The same procedure for site-directed overlap extension PCR in Chapter III was followed to produce a plasmid mixture of NNK mutants. The plasmid mixture was cleaned using DNA purification kits and transformed into *E. coli* BL21 cells with pEVOL-pAzF plasmid. Cells were spread on LB kanamycin plates (6.25 g LB powder mix, 4 g agar, 250 mL DDI water, 0.05 mg/mL kanamycin, 0.02 mg/mL chloramphenicol) before recovering in SOC medium for 1 hour

at 37 °C. Plates were incubated at 37 °C overnight; typically, greater than 100 colonies were observed when 1/5 of the outgrowth was used. To examine the mutant rates for a library, around 20 colonies were inoculated on LB media (with 0.05 mg/mL kanamycin, 0.02 mg/mL chloramphenicol) and grown overnight at 37 °C, 250 rpm. Mutant plasmid from these overnight grown cultures were isolated using kit from Qiagen (Valencia, CA) and given for sequencing. Plasmid sequencing was done at the U Chicago sequencing facility, and partB-for and partC-rev primers were used for sequencing reactions. The library expression, lysis, purification, bioconjugation, cofactor scavenging and screening was done using the procedure described in **3**.

Table S3.3. Nucleotide sequences for the primers required

#	Primer name	Primer sequence
1	T7 for	5’-GCG AAA TTA ATA CGA CTC ACT ATA-3’
2	T7 rev	5’-TTA TGC TAG TTA TTG CTC AGC GG-3’
3	N84S for	5’-GTT ATC GTG GAT AGT AAA GAA CTG GAA CGT G-3’
4	N84S rev	5’-CAC GTT CCA GTT CTT TAC TAT CCA CGA TAA C-3’
5	S99G for	5’-GTC CTG CTG CAG GGC TTT ACC ACG GAC GAG-3’
6	S99G rev	5’-CTC GTC CGT GGT AAA GCC CTG CAG CAG GAC-3’
7	I330K for	5’-CCG CTG GAT AAA GAC GAA GAA CGT GCA CTG-3’
8	I330K rev	5’-CAG TGC ACG TTC TTC GTC TTT ATC CAG GCC-3’
9	A335V for	5’-GAC GAA GAA CGT GTT CTG CTG CGC TAC ACC-3’
10	A335V rev	5’-GGT GTA GCG CAG CAG AAC ACG TTC TTC GTC-3’
11	N161K for	5’-CCG CAA AGA AAA AAC GCC GGA TG-3’
12	N161K rev	5’-CAT CCG GCG TTT TTT CTT TGC GG-3’
13	A166V for	5’-GCC GGA TGG TGT CAA TCC GCC GG-3’
14	A166V rev	5’-CCG GCG GAT TGA CAC CAT CCG GC-3’
15	G301S for	5’-GGT CCA TGC CAG CTA TAA ACT GG-3’
16	G301S rev	5’-CCA GTT TAT AGC TGG CAT GGA CC-3’
17	S308T for	5’-GGA AGT GTA CAC CCT GAA CGG CG-3’
18	S308T rev	5’-CGC CGT TCA GGG TGT ACA CTT CC-3’
19	301NNK for	5'-CTG GTC CAT GCC NNK TAT AAA CTG GAA GTG-3'
20	301NNK rev	5'-CAC TTC CAG TTT ATA MNN GGC ATG GAC CAG-3'

REFERENCES

1. Benson, D.; Wisz, M.; Hellinga, H., Rational design of nascent metalloenzymes. Proceedings of the National Academy of Sciences of the United States of America **2000**, *97* (12), 6292-6297.

2. Srivastava, P.; Yang, H.; Ellis-Guardiola, K.; Lewis, J., Engineering a dirhodium artificial metalloenzyme for selective olefin cyclopropanation. Nature Communications **2015**, *6*, 7789-7798.

3. Petrik, I.; Liu, J.; Lu, Y., Metalloenzyme design and engineering through strategic modifications of native protein scaffolds. Current Opinion in Chemical Biology **2014**, *19*, 67-75.

4. Harris, M.; Madura, J.; Ming, L.; Harwood, V., Kinetic and mechanistic studies of prolyl oligopeptidase from the hyperthermophile Pyrococcus furiosus. Journal of Biological Chemistry **2001**, *276* (22), 19310-19317.

5. a) Arnold, F., Design by directed evolution. Accounts of Chemical Research 1998, 31 (3), 125-131. b) Renata, H.; Wang, Z.; Arnold, F., Expanding the Enzyme Universe: Accessing Non-Natural Reactions by Mechanism-Guided Directed Evolution. Angewandte Chemie-International Edition **2015**, *54* (11), 3351-3367.

6. a) Arnold, F., Directed evolution: Creating biocatalysts for the future. Chemical Engineering Science **1996**, *51* (23), 5091-5102. b) Moore, J.; Arnold, F., Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. Nature Biotechnology **1996**, *14* (4), 458-467. c) Reetz, M., Biocatalysis in Organic Chemistry and Biotechnology: Past, Present, and Future. Journal of the American Chemical Society **2013**, *135* (34), 12480-12496. d) Reetz, M.; Zonta, A.; Schimossek, K.; Liebeton, K.; Jaeger, K., Creation of enantioselective biocatalysts for organic chemistry by in vitro evolution. Angewandte Chemie-International Edition **1997**, *36* (24), 2830-2832. e) Shao, Z.; Arnold, F., Engineering new functions and altering existing functions. Current Opinion in Structural Biology **1996**, *6* (4), 513-518.

7. a) Fasan, R.; Meharenna, Y.; Snow, C.; Poulos, T.; Arnold, F., Evolutionary History of a Specialized P450 Propane Monooxygenase. Journal of Molecular Biology **2008**, *383* (5), 1069-1080. b) Lewis, J.; Bastian, S.; Bennett, C.; Fu, Y.; Mitsuda, Y.; Chen, M.; Greenberg, W.; Wong, C.; Arnold, F., Chemoenzymatic elaboration of monosaccharides using engineered cytochrome P450(BM3) demethylases. Proceedings of the National Academy of Sciences of the United States of America **2009**, *106* (39), 16550-16555. c) Jung, S.; Lauchli, R.; Arnold, F., Cytochrome P450: taming a wild type enzyme. Current Opinion in Biotechnology **2011**, *22* (6), 809-817. d) Arnold, F., The nature of chemical innovation: new enzymes by evolution. Quarterly Reviews of Biophysics **2015**, *48* (4), 404-410.

8. Lewis, J., Artificial Metalloenzymes and Metallopeptide Catalysts for Organic Synthesis. *Acs Catalysis* **2013,** *3* (12), 2954-2975.

9. a) Letondor, C.; Pordea, A.; Humbert, N.; Ivanova, A.; Mazurek, S.; Novic, M.; Ward, T., Artificial transfer hydrogenases based on the biotin-(strept)avidin technology: Fine tuning the selectivity by saturation mutagenesis of the host protein. *Journal of the American Chemical Society* **2006,** *128* (25), 8320-8328. b) Reetz, M.; Peyralans, J.; Maichele, A.; Fu, Y.; Maywald, M., Directed evolution of hybrid enzymes: Evolving enantioselectivity of an achiral Rh-complex anchored to a protein. *Chemical Communications* **2006,** (41), 4318-4320. c) Creus, M.; Pordea, A.; Rossel, T.; Sardo, A.; Letondor, C.; Ivanova, A.; Le Trong, I.; Stenkamp, R.; Ward, T., X-ray structure and designed evolution of an artificial transfer hydrogenase. *Angewandte Chemie-International Edition* **2008,** *47* (8), 1400-1404. d) Reetz, M.; Rentzsch, M.; Pletsch, A.; Taglieber, A.; Hollmann, F.; Mondiere, R.; Dickmann, N.; Hocker, B.; Cerrone, S.; Haeger, M.; Sterner, R., A robust protein host for anchoring chelating ligands and organocatalysts. *Chembiochem* **2008,** *9* (4), 552-564. e) Song, W.; Tezcan, F., A designed supramolecular protein assembly with in vivo enzymatic activity. *Science* **2014,** *346* (6216), 1525-1528.

10. Packer, M.; Liu, D., Methods for the directed evolution of proteins. *Nature Reviews Genetics* **2015,** *16* (7), 379-394.

11. Harris, M.; Madura, J.; Ming, L.; Harwood, V., Kinetic and mechanistic studies of prolyl oligopeptidase from the hyperthermophile Pyrococcus furiosus. *Journal of Biological Chemistry* **2001,** *276* (22), 19310-19317.

12. Gibson, D.; Young, L.; Chuang, R.; Venter, J.; Hutchison, C.; Smith, H., Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **2009,** *6* (5), 343-341.

13. Young, T.; Ahmad, I.; Yin, J.; Schultz, P., An Enhanced System for Unnatural Amino Acid Mutagenesis in E. coli. *Journal of Molecular Biology* **2010,** *395* (2), 361-374.

14. It is azide agarose (product # 1038) from Click Chemistry Tools, LLC.

15. Hamman, C.; Wong, M.; Hayes, M.; Gibbons, P., A high throughput approach to purifying chiral molecules using 3 µm analytical chiral stationary phases via supercritical fluid chromatography. *Journal of Chromatography a* **2011,** *1218* (22), 3529-3536.

16. Bennett, B.; Kimball, E.; Gao, M.; Osterhout, R.; Van Dien, S.; Rabinowitz, J., Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nature Chemical Biology* **2009,** *5* (8), 593-599.

17. Wilson, Y.; Durrenberger, M.; Nogueira, E.; Ward, T., Neutralizing the Detrimental Effect of Glutathione on Precious Metal Catalysts. *Journal of the American Chemical Society* **2014,** *136* (25), 8928-8932.

18. a) Filice, M.; Romero, O.; Gutierrez-Fernandez, J.; de las Rivas, B.; Hermoso, J.; Palomo, J., Synthesis of a heterogeneous artificial metallolipase with chimeric catalytic activity. *Chemical Communications* **2015,** *51* (45), 9324-9327. b) Vahidi, A.; Yang, Y.; Ngo, T.; Li, Z., Simple and Efficient Immobilization of Extracellular His-Tagged Enzyme Directly from Cell Culture Supernatant as Active and Recyclable Nanobiocatalyst: High-Performance Production of Biodiesel from Waste Grease. *Acs Catalysis* **2015,** *5* (6), 3157-3161.

19. Sheldon, R.; van Pelt, S., Enzyme immobilisation in biocatalysis: why, what and how. *Chemical Society Reviews* **2013,** *42* (15), 6223-6235.

20. Hanefeld, U.; Gardossi, L.; Magner, E., Understanding enzyme immobilisation. *Chemical Society Reviews* **2009,** *38* (2), 453-468.

21. Fernandez-Arrojo, L.; Santos-Moriano, P.; Rodriguez-Colinas, B.; Ballesteros, A.; Plou, F., Micro-scale procedure for enzyme immobilization screening and operational stability assays. *Biotechnology Letters* **2015,** *37* (8), 1593-1600.

22. Juhasz, R.; Szeltner, Z.; Polgar, L., Truncated prolyl oligopeptidase from Pyrococcus furiosus. *Proteins-Structure Function and Bioinformatics* **2007,** *69* (3), 633-643.

23. They include EziG-1-Fe, EziG-2-Fe, EziG-3-Fe from EnginZyme, Inc.

24. Ilie, A.; Reetz, M., Directed Evolution of Artificial Metalloenzymes. *Israel Journal of Chemistry* **2015,** *55* (1), 51-60.