THE UNIVERSITY OF CHICAGO

PROLYL OLIGOPEPTIDASE-BASED ARTIFICIAL METALLOENZYMES 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

KENNETH J. ELLIS-GUARDIOLA

CHICAGO, ILLINOIS

JUNE 2017

Para mi madre,

por su apoyo sin empuje,

y por el gran mundo que me ha dado

If your ball is too big for your mouth, it's not yours.

–Dog of Wisdom

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

LIST OF FIGURES

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

APPENDIX ONE

LIST OF SCHEMES

CHAPTER ONE

CHAPTER TWO

CHAPTER THREE

CHAPTER FOUR

ACKNOWLEDGEMENTS

The past five years at the University of Chicago have been a period of immense growth and self-discovery for me, both personally and as a scientist. On the scientific end, I foremost thank my advisor, Jared Lewis, for his guidance and consistent pressure to achieve the ambitious scientific goals laid out in this project. My skills as an experimentalist and have improved dramatically due to his close mentorship. Though this relationship has been at times colored by tension, it has forged me into a more critical and resilient scientist, and for this I am thankful.

I also wish to thank the many members of the Lewis group with whom I have collaborated on these projects. Dr. Hao Yang, Dr. Chen Zhang, Dr. Poonam Srivastava, Dr. Cathy Poor, Dr. Hyun June Park, Dr. Ketaki Belsare, Yifan Gu, Alan Swartz, Brian Koronkiewicz, David Upp, and Jacqueto Zephyr. The work described in this dissertation could not have been realized without the help of these talented scientists, and I have been privileged to have met and worked with each one. I also thank the rest of my labmates, past and present, for helpful discussion and support, particularly Dr. Landon Durak, Dr. James Payne, Mary Andorfer and Joe Gair. I thank my undergraduate mentees, Ryan Beckner and Marissa Parker, whose excitement and passion for this research has been inspiring to witness. I wish to thank my external collaborators Dr. Narayanasami Sukumar (Argonne National Laboratory), Dr. Sunhwan Jo (Roux group) and Gihoon Lee (Moellering group). Special thanks to Dr. Chang Qin, Dr. Antoni Jurkiewicz, Dr. Elena Solomaha, and Dr. Justin Jureller for their exceptional support with instrumentation.

I also thank the members of my committee, Prof. Chuan He and Prof. Raymond Moellering, for their input and support. Special thanks to Prof. Joseph Piccirilli for his support through the CBI Trainee program.

Graduate school is a period of intense pressure, one that I could not have survived without my close-knit community of friends. To my labmates Dr. Landon Durak, Dr. James Payne and Mary Andorfer: you guys have been the best with jokes, the best with commiseration, and three of the best people I have ever met. We would not have gotten through the early years without each other and I'm privileged to be your friend. I am also indebted to my friends outside the Lewis group: Jeff Montgomery, Dennis Tam, Nathan La Porte, Judith Kamm, Mike Boles, EW Malachosky, Alex Linkin, and the many that I am omitting from this list for brevity. My time spent with these people will undoubtedly be the most lasting and memory of this period of my life.

I am also indebted to my family and friends outside the University. I thank my mother, whose consistent wisdom, experience, and love have given me strength through this period. Gracias por todo, Mamá. I also thank my siblings, Inés, John, and Miriam, for their loving support. My family has kept me grounded and reminded me of who I truly am, and I owe them a debt of gratitude.

Thanks to Newspaper the Cat for being pretty cute and meowing a lot. I like that.

Finally, thanks to my girlfriend, Claire O'Hanlon. I am hugely indebted to her patient love and constant support, even through some of the most turbulent and stressful periods of graduate school. Our daily chats mingled with occasional temporary cohabitation have been the morsels of sweet routine that have sustained me throughout. Our love has shown resilience in the face of long distance and trying circumstances. I am forever grateful to have met you on that Friday evening in 2012.

ABSTRACT

This dissertation describes the preparation, characterization, and application of artificial metalloenzymes for the selective functionalization of organic molecules. These catalysts combine a catalytic metal complex and a protein scaffold, allowing one to exploit the reactivity of the transition metal catalyst while taking advantage of the selectivity and evolvability of proteins. The work described herein is specifically focused on the application of *Pyrococcus furiosus* prolyl oligopeptidase as an enabling protein scaffold for the generation of exceptionally robust and selective artificial metalloenzymes.

Chapter One introduces the concept of catalyst control of selectivity and the challenges surrounding it, particularly in the context of C-H functionalization. Site-selective C-H-functionalization, the process whereby a specific C-H bond is converted to a desired functional group, remains a challenging problem for small molecule catalysis. An overview of methods to achieve site-selective catalysis is described, encompassing strategies employing small-molecule transition metal catalysis and enzyme catalysis. A third class of catalyst, artificial metalloenzymes, is introduced as potentially combining the advantages of small molecule catalysts (reactivity) with enzymes (evolvability/selectivity). The many methods that have been employed to generate artificial metalloenzymes is then discussed in detail, with a particular focus on challenges in the characterization of these catalysts.

Chapter Two describes the formation of a wide variety of artificial metalloenzymes formed using serine hydrolase scaffolds. Electrophilic phosphonate-tethered metal catalysts have been covalently bioconjugated to serine hydrolases, exploiting the native mechanism of this class of enzymes. The synthetic procedures to generate such metal cofactors are described. A panel of serine hydrolase scaffolds is explored using these catalysts. A number of the resulting artificial

metalloenzymes have been applied successfully for C-H oxygenation, epoxidation, and cyclopropanation, albeit with limited selectivity. Strategies to enhance these selectivities are also elaborated in this chapter.

Chapter Three introduces the application of *Pyrococcus furiosus* prolyl oligopeptidase as an ideal protein scaffold for artificial metalloenzyme formation. Methodology for the expression of this scaffold has been optimized, and the stability of the scaffold has been investigated. Methods for the routine characterization of artificial metalloenzymes using this scaffold are described, including native activity assays, mass spectrometry, and LC/MS/MS. Finally, the *Pyrococcus furiosus* prolyl oligopeptidase crystal structure has been solved and a detailed description of its structural features is presented. A number of discoveries pertinent to the native activity of the enzyme and its application as an artificial metalloenzyme scaffold are elaborated.

Chapter Four describes the formation of dirhodium artificial metalloenzymes using *Pyrococcus furiosus* prolyl oligopeptidase. Enantioselective cyclopropanation has been catalyzed with these systems and catalyst performance has been optimized through rational mutagenesis and directed evolution. Biophysical studies into the origin of the observed selectivity were conducted, with a hypothetical histidine-rhodium interactions at the center of investigation. Finally, a detailed digestion LC/MS/MS study into artificial metalloenzyme self-modification with diazo- substrates is described. This revealed that self-modification occurs during catalysis, potentially impacting the selectivity of the catalyzed reaction. The methods described lay the groundwork for future detailed studies.

PREFACE

Each chapter of this dissertation is numbered independently. A given compound or reference 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. All NMR spectra for select compounds in Chapter Two are included in Appendix One.

CHAPTER ONE

AN INTRODUCTION TO SELECTIVE CATALYSIS AND ARTIFICIAL METALLOENZYMES

1.1 THE IMPORTANCE OF SELECTIVITY IN CATALYSIS

Chemical catalysis is a fundamentally important branch of chemistry, providing some of the most beautiful examples of creativity and complexity in chemical synthesis and nature. At its essence, catalysis is nothing more than the minimization of kinetic barriers between reactants and products in a chemical reaction. The mechanism of catalysis can therefore be attributed to the reduction of activation energy of a given transformation. By definition, all catalysts reduce free energy of activation by stabilization of the transition state of a chemical transformation through entropic and enthalpic means (Figure 1.1).[1] The chemical community has been tremendously successful at discovering new ways to catalyze the transformation of reactive handles—generally termed *functional groups*—to other desired chemical moieties. Reliable *reactivity* has been established for a huge number of catalysts. In contrast, stereo-, regio-, site-, and chemo-*selectivity* remain a particularly vexing challenge. *Selectivity* in catalysis entails the energetic differentiation of different functional groups on a substrate, or, in the case of asymmetric catalysis, the differentiation of the different faces of the same functional group. The rewards for unlocking catalyst selectivity are immense. Commonly, a substrate containing multiple functional groups (or sites) that exhibit reactivity toward a given catalyst means that, in the presence of a non-selective catalyst, one will ultimately obtain a complex mixture of products with differential extent of

conversion at the different available sites. Aside from the inherent loss of yield to side-products, the scientist is then tasked with the tedious separation and isolation of chemically-similar compounds, a source of major inefficiency in the process of chemical synthesis. The development of catalysts that could cleanly and efficiently convert the substrate to a given product of interest would obviate the need for difficult purification and would result in inherently improved yields of the desired product. It is therefore of great interest to scientists to develop methods to readily generate efficient catalysts with high selectivities.

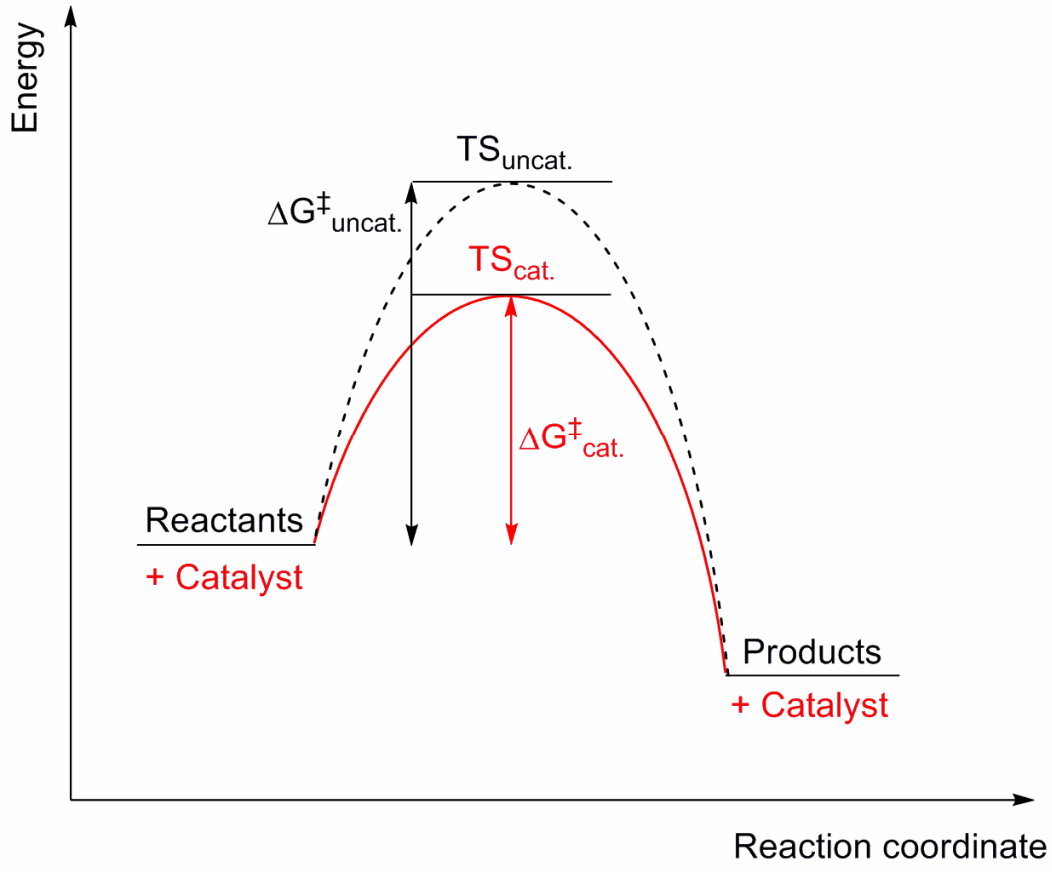


Figure 1.1 A general free energy diagram illustrating the effect of a catalyst in reducing the free energy of activation of a given transformation.

1.2 C-H FUNCTIONALIZATION CATALYSIS

The fundamental goal undertaken in this laboratory is the creation of selective catalysts that improve the efficiency of chemical synthesis. As discussed above, selectivity in catalysis enhances the efficiency of synthesis by enabling clean conversion to desired products, obviating the tedium of purification of intermediates along a synthetic route. However, this does not address another basic source of inefficiency in chemical synthesis: the need for substrate pre-functionalization –that is, the prerequisite that reactive handles be present on a given substrate. This has dominated the concept of retrosynthesis –bond disconnections are made based on the reactive functional groups available on a molecule.[2] Rarely is the ubiquitous C-H bond thought of as a viable reactive handle, as these are relatively inert compared to the polarized bonds that typify the handles used in retrosynthesis. However, the past 50 years has brought to light a number of catalysts capable of activating and functionalizing C-H bonds.[3] The ability to control the selectivity of these catalysts would hypothetically enable the selective targeting of specific C-H bonds on a hydrocarbon skeleton, obviating the necessity for substrate pre-functionalization (with polar carbon-halogen, carbon-chalcogen, or carbon-main group bonds), thereby minimizing the number of steps and the amount of waste in a given chemical synthesis. The combination of C-H functionalization catalysis with high stereo/regio/site/chemo-selectivity could provide some of the most concise tools for the precise conversion of simple building blocks to complex products.[4]

1.2.1 SELECTIVE CATALYSIS BY SYNTHETIC TRANSITION METAL CATALYSTS

Catalysis mediated by transition metals has been a dominant force in the development of important transformations in synthetic chemistry, including cross-couplings[5]. olefin polymerizations[6], olefin metathesis reactions[7], and C-H bond functionalizations[8]. An advantage of transition metal catalysis arises from the fact that catalytic activity (selectivity/reactivity) depends on the structural parameters of the metal complex itself. This enables ready perturbation of catalytic activity through the modulation of catalyst structure, determining its orbital/electronic structure as well as steric environment, both of which influence catalyst reactivity and selectivity. By changing ligands directly coordinated to the metal (the primary coordination sphere) and ligand functionality distal to the metal (the secondary coordination sphere), scientists can optimize the complex for desired selectivity and reactivity.

As noted above, a multitude of synthetic homogeneous transition metal catalysts have been developed for the activation and functionalization of C-H bonds. In chapter two, the C-H oxidation chemistry of first-row metal-polydentate nitrogen complexes will be described. In chapter four, C-H carbene insertion catalysts based on the dirhodium tetracarboxylate framework will be detailed. Small molecule C-H functionalization catalysts employ a variety of mechanisms to achieve activation and functionalization of inert C-H bonds. For the C-H oxidation chemistry explored with first-row metal-polydentate nitrogen complexes, this entails abstraction of a hydrogen atom by a highly oxidative metal-oxo species, followed by rebound of the oxygen fragment to generate a C-O bond.[9] In the case of dirhodium catalysis, the catalyst is capable of stabilizing high-

energy electrophilic carbenes, which can then insert into a variety of C-H bonds.[10] The diversity of metal-binding ligands that have been applied toward the optimization of both of these catalyst classes cannot be understated.[11-13] Homogeneous transition metal catalysts serve as excellent platforms for "coarse-tuning" of reactivity and chemoselectivity, as the structures of these compounds are often predictable and well-defined by a framework whose structural parameters are set by relatively stable molecular forces (covalent/dative bonds). Trends in reactivity can often be evaluated through modulation of this framework, and rational, predictable changes can be made to these relatively simple structures. As such, systematic optimization of complexes has become a norm in the organometallic community for development of new catalysts with novel reactivity and chemoselectivity.

Achieving catalyst-defined stereo-, regio-, and site-selectivity has been a major driving force in small-molecule catalysis. Unsurprisingly, small molecule catalysts have proven to be amenable for enabling asymmetric transformations, as simple stereochemical information can often be relayed readily from an asymmetric small molecule catalyst to its substrate.[14] Enantio-differentiation of substrate functional groups is well suited to small molecule catalysts, as the substrate/catalyst interactions mediating enantioselectivity occur on compatible length scales. A number of asymmetric metal catalysts have been hugely successful in their application toward organic synthesis, notably those introduced by Knowles,[15] Noyori,[16] Sharpless,[17] and Jacobsen.[18] In contrast, achieving site/regio selectivity with a small molecule catalyst is significantly more challenging, as selection of one specific site among functional groups with similar reactivity requires more complex intermolecular interactions (steric and electronic) in

order to energetically differentiate these positions on a substrate. In some cases, such as iridium-catalyzed aryl C-H borylation and silylation, site selectivity can be promoted almost entirely by steric repulsion between the catalyst and the substituents on the arene substrate, ultimately targeting the most accessible arene C-H bond available.[19,20] This is one of the simplest examples of catalyst-controlled site selectivity in C-H functionalization. Other catalysts, like Davies' recently reported site-selective dirhodium C-H insertion employ deep ligand-defined cavities that can sterically differentiate alkyl C-H bonds to enable site-selective reaction (Figure 1.2).[21] Beyond steric control, Crabtree was able to promote differentiation of the two benzylic sites on ibuprofen using hydrogen bonding provided by a ligand-bound pendant carboxylate.[22] These are examples of the sort of higher-order catalyst-substrate interaction, termed "molecular recognition," that can be introduced to enable catalyst-controlled site-selectivity.[23] However, these and many other instances also serve to highlight the fact that introduction of these properties can be synthetically challenging.[24] Therefore, the development of more general approaches toward the optimization of catalyst selectivity stands as an appealing goal in the field**.**

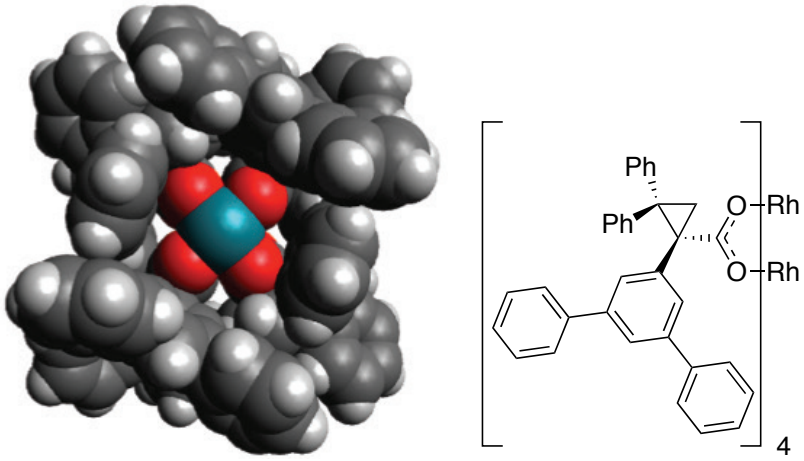


Figure 1.2 Figure borrowed from ref. 18. A "small" molecule synthetic catalyst capable of defining a chiral pocket for site- and enantioselective C-H carbene insertion

1.2.2 SELECTIVE CATALYSIS BY ENZYMES

Enzymes have proven to be incredibly effective catalysts for the conversion of their native substrates, often doing so with impressive rate enhancements and excellent selectivity. Their power arises from their ability to three-dimensionally organize substrate-catalyst interactions in order to reduce the activation energy of the catalyzed transformation.[25,26] This capacity for molecular recognition is a product of the complex three-dimensional structure of the enzyme, a parameter defined by its primary amino acid sequence as well as its constituent cofactors. The three-dimensional structures of enzymes are governed by weak intramolecular forces (hydrogen-bonding, hydrophobic interactions, electrostatics) that ultimately govern the tertiary structure of the molecule. Because this shape is predicated on the primary sequence of the polypeptide chain, alterations to the primary sequence translate to alterations in the three-dimensional structure, which ultimately describes the chemical properties of the molecule. These chemical properties (functional groups, electrostatics, steric environment, etc.) are

provided by the subset of chemical entities available for enzyme biosynthesis (20 canonical amino acids, biogenic cofactors, bioavailable metals). When assessed in a combinatorial fashion, the number of theoretical catalysts accessible through protein biosynthesis is astronomical.[27]

Among the vast array of natural enzymes, there are several notable C-H functionalization catalysts that have garnered great attention from the chemical community.[28] Of particular relevance to this work are those that employ metallocofactors as reactive centers. Cytochrome p450s exemplify the exquisite efficiency and selectivity of natural metalloenzyme catalysis, enabling the selective hydroxylation of aliphatic and aromatic C-H bonds on a plethora of substrates.[29] Similarly, a number of metallo-halogenases pose great promise in their capability to selectively halogenate aliphatic and aromatic C-H bonds, a transformation of great synthetic utility.[30] In both of these cases, a variety of proteinogenic functional groups in the secondary coordination sphere around the metal enable molecular recognition of the substrate, enabling the catalyst-controlled differentiation of energetically similar C-H bonds.[28]

A critical difference between small molecule catalysts and their enzymatic counterparts is that the synthesis of an enzyme is carried out through the templated action of ribosomal machinery. That is, structural information is encoded at the genetic level into DNA and flows according to the central dogma of biology: DNA is transcribed to RNA, which is then translated into a protein.[31] A scientist need only exploit the transcriptional and translation translational machinery of a host organism (*Escherichia coli* and *Saccharomyces cerevisiae* being a prime workhorses in protein engineering) in order to synthesize (express) a protein of interest. The technology for the introduction and

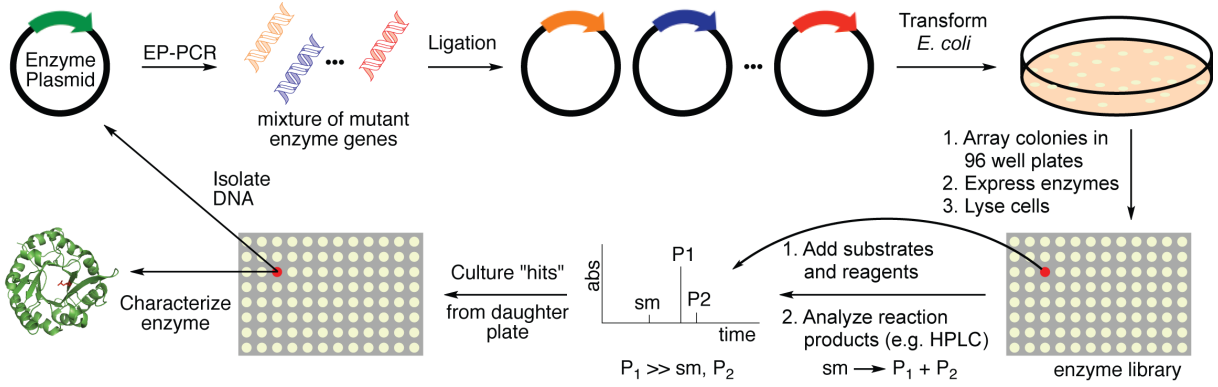
expression of heterologous genes in host organisms has reached a point that standard protocols can enable the ready biosynthesis and extraction of a wide variety of proteins in *E. coli* hosts (with the caveat that expression levels, solubility of protein product, and purification can be variable depending on the identity of the gene of interest.)

Perturbation of the protein structure is also facile since it entails the introduction of mutations to the gene, a task that, with modern cloning techniques, has become operationally simple to perform.[32] Given the relatively simple task of synthesis and structural diversification of proteins, one can appreciate a critical advantage in the optimization of enzyme catalysts over their small-molecule counterparts: the ready generation of libraries of similar yet structurally diverse catalysts. The parallel synthesis of panels of subtly different proteins enables the rapid evaluation of these variants for desired characteristics, like selectivity. In contrast, an analogous strategy for small molecule catalysts can be quite challenging for a synthetic chemist, given the non-generality of chemical synthesis, which holds true even for the introduction of subtle molecular differences. Examples wherein "mutants" of small molecule catalysts have been synthesized and evaluated in parallel generally rely on modular synthetic strategies (i.e. peptide synthesis) echoing the modular action of the ribosome.[33,34]

Importantly, while "rational" mutagenesis can be performed by deliberate introduction of specific mutations into the amino acid sequence of a protein, many methods enable semi-rational[35] or random mutagenesis[36] as well. Evaluation of libraries of enzyme variants created through random or semi-random mutagenesis forms the basis of "laboratory evolution" or "directed evolution."[37] (Scheme 1.1) This method will be explored in greater depth in chapter four. For now, it suffices to note that random

mutagenesis creates a diverse panel of structurally-related catalysts that can then be evaluated for desired function (increase in selectivity, for example). Variants showing improvements in desired function or "fitness" serve as the basis, or "parent," for a further round of mutagenesis and screening. This iterative approach enables the ready optimization of desired catalyst traits *without the need* for mechanistic/structural insight or guidance.[37]

Scheme 1.1 The general procedure for directed evolution as applied in this laboratory toward the optimization of selectivity



The directed evolution methodology described echoes the mechanisms at play in the natural evolution of enzymes. However, as noted, the "survival" of a gene is determined by the improvement observed by screening of function, which is often the bottleneck in the protocol. In natural evolution, the criterion for selection of the "fittest" variant is predicated on the survival of the organism under the selective pressures applied. Laboratory evolution methods have been developed that link desired function to gene survival, thereby obviating the need for tedious screening protocols.[38] These present

some of the most efficient biomolecular optimization protocols available to scientists, although more work must be done to enhance the generality of these methods in their application toward selective catalysis.

Enzymes are a powerful class of catalyst due to their exceptional selectivity and turnover as well as their versatility toward optimization. Indeed, the efficiency of these catalysts and the evolutionary algorithms that govern their adaptation for new function make them enviable systems for synthetic chemists. Natural enzymes should not be viewed as a panacea for all challenges in catalysis, however. They are limited by the pool of chemical entities available for their biosynthesis, for example, which excludes molecular features (metals/ligands/functional groups) present in many successful synthetic catalysts. Furthermore, artifacts of their biological origins (such as the need for aqueous or membrane-like chemical environments, poor adaptation to very high temperatures, intolerance of highly reactive abiotic species) can pose challenges for their practical implementation in a laboratory context. However, when enzymes are engineered for the goals of synthetic chemists, their versatility can be profoundly enabling.

1.3 THE HYBRID ARTIFICIAL METALLOENZYME

1.3.1 INTRODUCTION TO ARTIFICIAL METALLOENZYMES

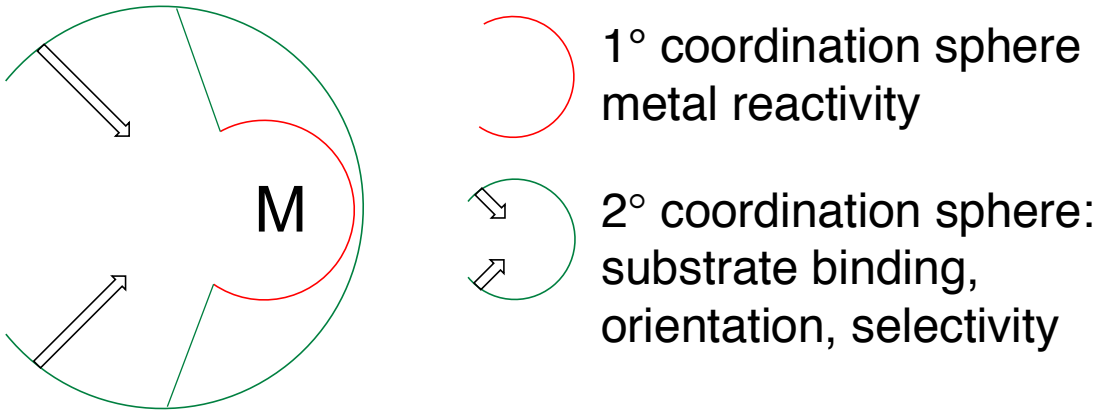


Figure 1.3 A general overview of an ArM catalyst. The primary coordination sphere is defined by the metal ligands. The green segment represents the protein scaffold, which controls the secondary coordination sphere of metal M.

Clearly, both synthetic transition metal complexes and enzymes carry a number of advantages and disadvantages in their implementation toward achieving novel selectivity and reactivity in catalysis, particularly in the context of C-H functionalization. The focus of this work centers on the bridge between these two powerful catalytic platforms: artificial metalloenzymes, or ArMs. ArMs, by definition, are comprised of two components: 1) a catalytic metal center (M), which serves as the locus of reactivity in these hybrid catalysts, and 2) a protein component (protein scaffold), which defines the parameters around this metal (Figure 1.3).[39] The creation of artificial metalloenzymes therefore relies on methodology to bring the catalytic metal center and the protein scaffold together. Three general methods are used to achieve this end: 1) the dative

approach, whereby a metal ion is bound directly through the coordination of scaffold residues, 2) the supramolecular approach, in which a metal complex (cofactor) is bound through non-covalent interactions, and 3) the covalent approach, in which the metal cofactor is covalently grafted to the protein scaffold. Often, combinations of these general methods can be used to enforce the junction of the metal fragment and the protein scaffold. These three methods carry certain advantages and disadvantages in their implementation, which will be discussed in the following sections.

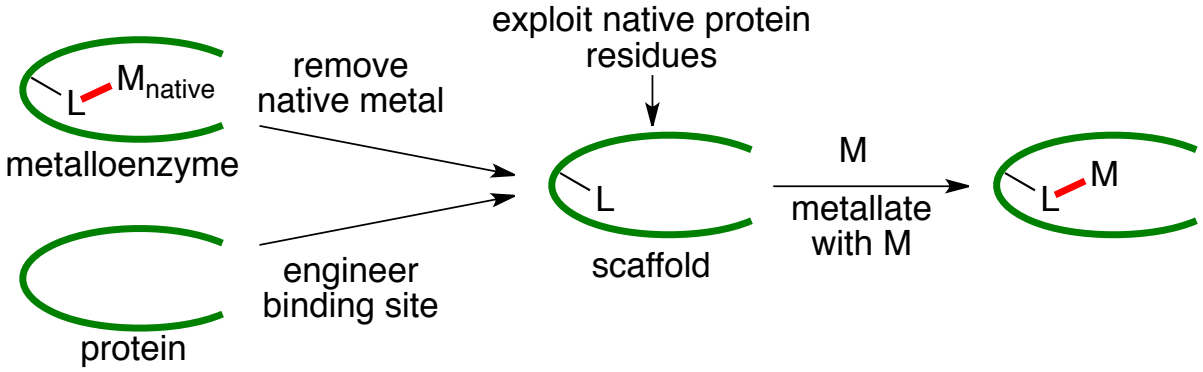
1.3.2 THE SYNTHESIS AND APPLICATION OF ARTIFICIAL METALLOENZYMES

1.3.2.1 DATIVE ARTIFICIAL METALLOENZYMES

An intuitive approach to the generation of artificial metalloenzymes mirrors the common preparation of small molecule transition metal catalysts, wherein a metal ion (M) is combined with an appropriate stoichiometry of small molecule ligands (L).[40] Given that proteins themselves contain a diverse set of metal binding functionality (N, O, or S donors), one may simply swap the small molecule ligands with a protein and, provided that the proper three-dimensional coordination environment is present, the protein scaffold will bind the metal. A variety of natural metalloenzymes exhibit this form of direct metal binding, inspiring chemists to mimic the strategy to generate new artificial metalloenzymes. The three principal methods employed for direct dative binding of metals with protein scaffolds include: 1) Repurposing of natural metalloproteins for binding of non-native metal ions, 2) Exploiting serendipitous metal binding by non-metalloproteins, and 3) Designing metal binding sites into an protein scaffold (Figure 1.4) Herein notable examples of each of these modes will be discussed.

The advantages of this method generally arise from the operational simplicity of the method, providing a rapid way of introducing the metal into the complex primary and secondary coordination environment of the protein scaffold. However, because the binding of the metal is predicated on direct coordination of proteinogenic ligands, some limitations arise. The first limitation stems from the fact that the metal-binding functionality must come from amino acids that can be ribosomally incorporated into a polypeptide chain. This is comprised of the 20 canonical amino acids and a subset unnatural amino acids (UAAs). Despite steadily growing in number, UAAs are limited in terms of practical incorporation by ribosomal machinery (see below).[41] The second limitation stems from the fact that, because coordination of the metal necessitates the scaffold to provide a binding context, mutagenesis strategies could destabilize these interactions, limiting the flexibility of the dative platform for optimization. In spite of this, a number of excellent examples of ArMs have been developed with this mode formation.

Figure 1.4. Approaches to generate ArMs via metal binding.



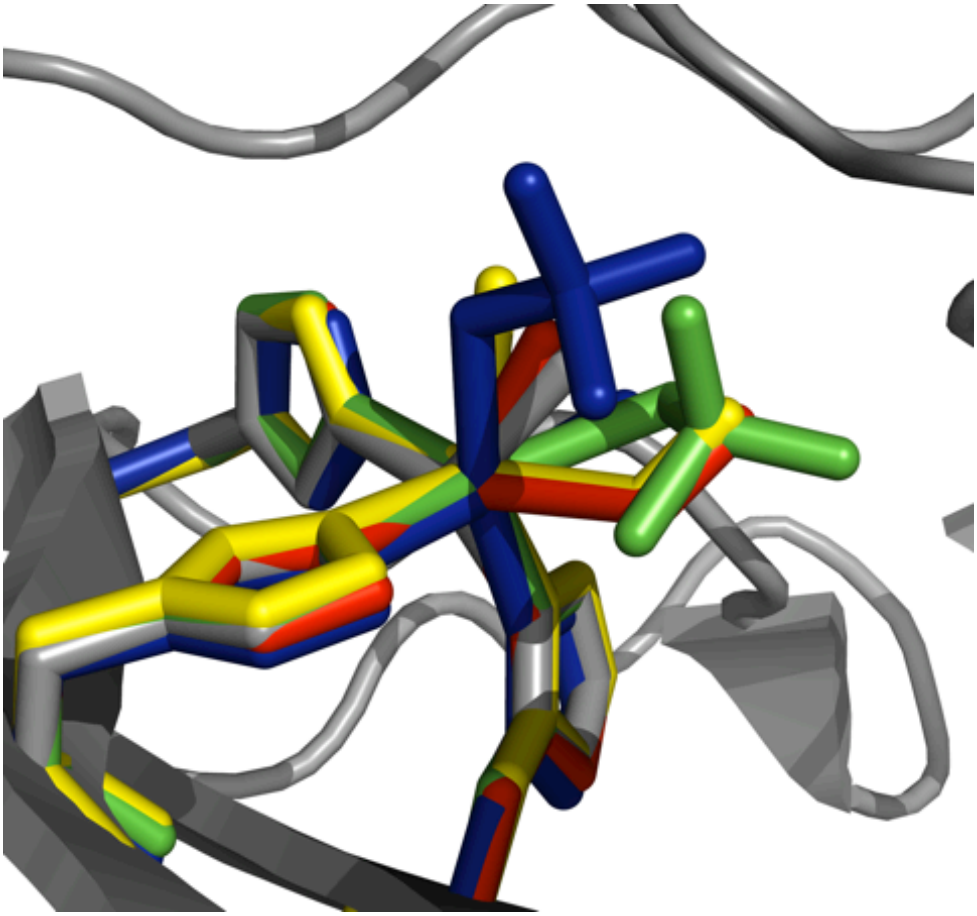
1.3.2.1.1 REPURPOSING NATURAL METALLOPROTEINS

Natural metalloproteins, by virtue of their evolved capacity for metal binding, already contain the requisite coordination pocket to accommodate metal ions. Therefore, to generate artificial metalloenzymes utilizing natural metalloprotein scaffolds, one need only develop reliable methods for generating the "apo" (non-metalated) form of the protein scaffold followed by its reconstitution with non-native metal ions. This has been accomplished for a number of metalloproteins.

One of the first examples of this methodology was applied to carboxypeptidase A, a zinc(II) metalloenzyme with a His/His/Glu binding site.[42] Dialysis against the chelator 1,10-phenanthroline was used to remove the native metal, and subsequent metalation with copper(II) led to a redox-active artificial metalloenzyme that catalyzed oxidation of ascorbic acid by O_2.[42] Another set of zinc(II) metalloenzymes, the ubiquitous carbonic anhydrases, have also been used as ArM scaffolds. The His/His/His binding motif has been reconstituted with manganese(II) to generate redox-active ArMs capable of catalyzing asymmetric alkene epoxidations.[43,44] In addition, rhodium(I) has been introduced into carbonic anhydrase to generate competent ArMs for the selective hydrogenation of alkenes as well as the hydroformylation of styrene.[39,40] The reconstitution of carbonic anhydrase has also been demonstrated with cobalt(II), copper(II), and nickel(II), as well, though with no catalysis reported (Figure 1.5).[45,46] The di-zinc(II) metallo-*β*-lactamase from *Stenotrophomonas maltophilia* was also demonstrated to be competent for the binding of two copper(II) ions after mutagenesis of Asp to His, which furnished two His/His/His binding sites, mimicking the active site architecture of type III copper proteins.[47] Unlike carbonic anhydrase and

carboxypeptidase A, metallo-β-lactamase can take up the non-native metal ion simply by expression of the protein with an excess of $Cu(SO_4)$. The resulting ArM proved to be a competent catechol oxidase.[47]

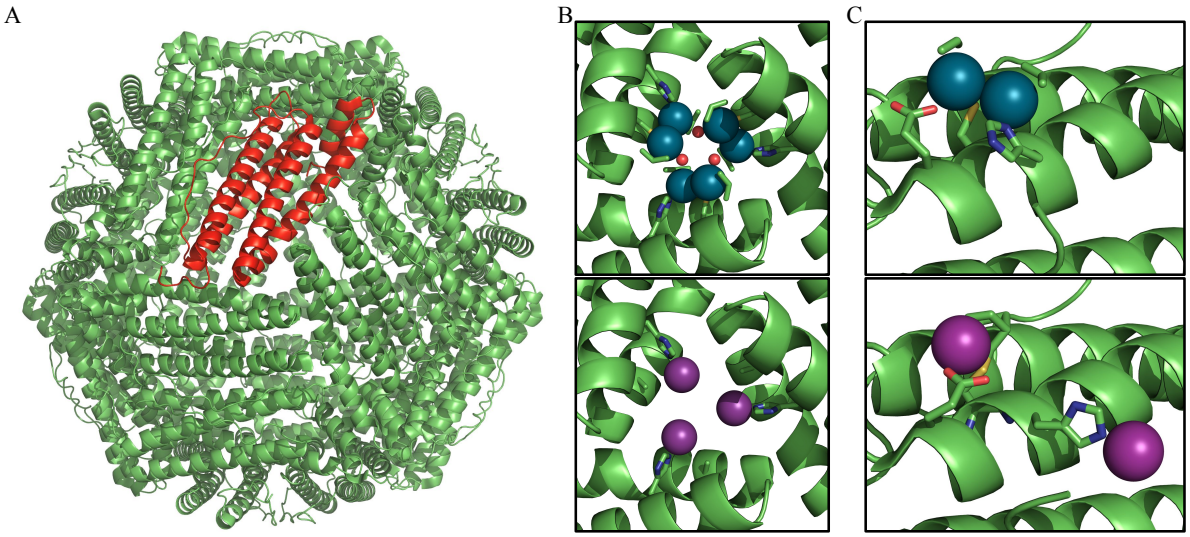
Figure 1.5. Overlay of His/His/His metal binding site in hCAII structures containing Zn(II) (gray), Co(II) (red), Cu(II) (yellow), Ni(II) (blue), and Mn(II) (green) bearing H_2O/O_2 (Zn, Cu, and Co) and sulfate (Ni and Mn) ligands



Among the most impressive examples of natural metalloprotein repurposing are those provided by ferritin. Ferritin is a massive 450 kDa cage-like spherical iron storage protein built up from 24 individual subunits. “Demineralization” of ferritin is carried out

by dialysis against thioglycolic acid.[48] The Watanabe group demonstrated that palladium(II) can be soaked into the apo-ferritin scaffold, generating a competent Suzukiase for the cross-coupling of phenylboronic acid with 4-iodoaniline (Figure 1.6).[49] Attempts were made at altering the observed coordination sites (confirmed by X-ray crystallography) through mutagenesis, though this had little impact on the observed selectivity.[50,51] Later, the same group used this approach to introduce rhodium(I) into the scaffold, which enabled the polymerization of phenylacetylene by the ArM (Figure 1.6).[52] This is a remarkable example of selectivity in ArM catalysis, as the polyphenylacetylene generated by this system showed lower average molecular weight and narrower polydispersity relative to the small-molecule $[Rh(nbd)Cl]_2$ catalyzed reaction.[52] Iridium(III) has also been introduced into the apo-ferritin scaffold, though no catalysis has yet been demonstrated.[53]

Figure 1.6. A) Apo-recombinant horse liver ferritin with highlighted subunit in red. B) Ferritin three-fold axis binding site occupied with Pd(II) (top, blue spheres) and Rh(I) (bottom, purple spheres). C) Ferritin accumulation binding site with Pd(II) (top, blue spheres) and Rh(I) (bottom, purple spheres)



1.3.2.1.2 SERENDIPITOUS METAL BINDING BY NON-METALLOPROTEINS

Exploring the metal binding capacity of non-metalloproteins expands the pool of viable protein scaffolds for dative ArM preparation. From an operational standpoint, this is an appealing strategy because any soluble and isolable protein can be screened for its capacity to bind metals and catalyze reactions of interest. Expectations for selectivity in catalysis using this method must be tempered, however, since fortuitous interactions govern not only the binding of the metal to the protein, but also the interface of the metal with its substrate. In addition, non-selective metal binding can lead to the formation of multiple active sites with different catalytic properties.

Serum albumins have attracted a great deal of attention as ArM scaffolds given their relative stability and commercial availability. These proteins qualify as non-metalloproteins, though it should be noted that their physiological role is in the transport of a number of species in blood, including metal ions like zinc and copper.[54] Serum albumins have seen use in the binding of rhodium(I) for application in the hydroformylation of olefins and the hydrogenation of α,β-unsaturated aldehydes and ketones.[55] However, no enantioselectivity could be observed using prochiral substrates with these systems, a likely symptom of multiple bound rhodium centers per albumin monomer.[56,57]

Albumins were also determined to be reasonable hosts for high-valent transition metal oxo-complexes like OsO_4. The resultant hybrids were employed for the dihydroxylation of alkenes, but again no enantioselectivity was observed.[58] Serendipitious binding of OsO_4 was also observed with the streptavidin protein scaffold. This protein has seen great utility in the generation of supramolecular non-covalent

anchoring ArMs (see below), but it has been proven to be competent at binding metal oxo complexes at sites proximal to the native biotin-binding site.[59] The streptavidin-OsO_4 hybrid catalyzed moderately selective dihyrdoxylation of alkenes.[59] A similar study employed $[VO]^{2+}$ for the generation of an oxidase for the enantioselective sulfoxidation of aryl thioethers. Importantly, the scaffold imparted a rate enhancement on the reaction relative to the free vanadium catalyst, showing a secondary benefit of this ArM platform.[60]

The phosphatase phytase has yielded a series of fascinating ArMs for the enantioselective sulfoxidation of sulfides. Structural alignment of phytase and vanadium chloroperoxidase revealed that both of these enzymes share remarkable similarities in structure. By treating phytase with vanadate VO_4^{3-}, Sheldon was able to demonstrate that the anionic oxo-complex not only inhibited native phosphatase activity, but also introduced chloroperoxidase activity into the phytase scaffold.[61] It was hypothesized that vanadate binds the oxoanion site of the phosphatase, as VO_4^{3-} and PO_4^{2-} are chemically similar. The same group showed that vanadate could bind numerous protein scaffolds to generate enantioselective sulfoxidases, albeit with lower selectivity and efficiency relative to the phytase/ VO_4^{3-} ArM.[62,63] This work demonstrates the utility of leveraging latent metal binding based on isostructural native substrate binding, as this can impart a well-defined metal active site for selective catalysis.

Finally, an exciting new area of ArM catalysis is exploring heterogeneous catalysis in the context of protein crystals. Ueno demonstrated that soaking hen egg white lysozyme crystals with ruthenium(II) complexes enabled the binding of the metal to discrete positions on the protein surface (determined by X-ray crystallography).[64] Cross-

linking the protein crystals with glutaraldehyde generated highly stable ArMs for the reduction of a number of acetophenone derivatives. Moderate enantioselectivity was observed with the system, and because the crystals were cross-linked, they were stable enough to be recycled for 10 rounds of catalysis.[61]

1.3.2.1.3 DESIGNING METAL BINDING SITES IN PROTEIN SCAFFOLDS

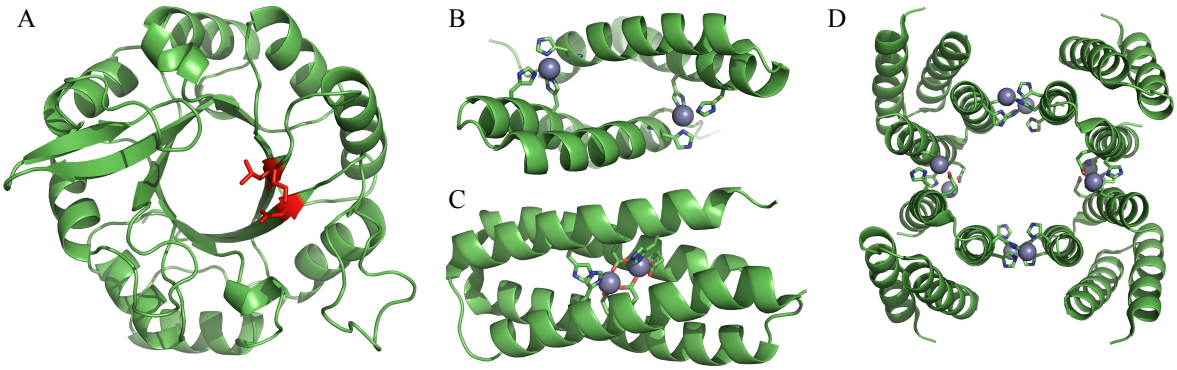
The previous two sections detail methods that utilize pre-existing, if latent, binding sites for the generation of ArMs. Structural modeling and *de novo* protein design, however, can be used to deliberately engineer metal binding sites into protein scaffolds. This expands the range of potential dative scaffolds and allows the scientist to rationally define the coordination chemistry and geometry of the site. Of course, this method requires the existence of structural information to guide the design strategy. Furthermore, metal binding is predicated on the ability of the protein scaffold to accommodate the necessary mutations, which may not always be the case. Despite these limitations, the design strategy has been used for the successful creation of a number of ArMs. In addition, this comprises a fundamentally distinct strategy from the previous two described, since the chemist is afforded the flexibility to not only choose the identity of the protein scaffold, but also to determine the specific location of metal binding in the context of the scaffold. This enables the judicious selection of binding site to leverage attractive structural features such as "vacant space" for substrate binding.[65]

A straightforward example of this method and the advantages afforded by vacant space was described by Reetz, who, by inspection of the structure of the thermostable TIM-barrel protein tHisF, was able to introduce a Cu(II) binding site comprised of a trigonally-arranged His/His/Asp triad (Figure 1.7 A). Enantioselective Diels-Alder

cycloaddition between aza-chalcones and cyclopentadiene was achieved with this ArM. It was hypothesized that the barrel shape of the scaffold furnished a well-defined active site in which substrates could bind.[66]

Another approach has been directed toward the design of metal binding sites at the interfaces of proteins. The Kuhlman group introduced His/His/His Zn(II) binding sites into the Rab4 binding domain of rabenosyn to enable the Zn(II)-directed assembly of homodimers. This generated an ArM esterase for the hydrolysis of *p*-nitrophenyl acetate (Figure 1.7 B).[67] A similar Zn(II)-templating approach was used by Tezcan to carry out the Zn(II)-dependent self-assembly of tetramers using cytochrome cb_{562} as a monomeric building block.[68] A second non-structural, catalytic Zn(II) binding site was introduced, which exhibited esterase activity. The resulting system was then optimized for β-lactamase-like hydrolysis of ampicillin *in vivo* through saturation mutagenesis with a survival-based selection (Figure 1.7 D).[69] This is one of the few examples in which catalytic activity of an ArM has been optimized through selection-based directed evolution.

Figure 1.7. Locations of metal binding sites introduced into scaffold proteins. A) tHisF scaffold with mutation sites in red. B) Rab4 Zn-directed homodimer crystal structure with Zn(II) represented with gray spheres. C) NMR structure of 3His-G4DFsc bound to two Zn(II) ions (gray spheres). D) Crystal structure of Zn_8:$^{A104/G57}$AB34 with structural Zn(II) sites on the vertical axis and catalytic Zn(II) sites on the horizontal axis.



In a departure from the trend of utilizing non-metalloenzyme scaffolds as hosts for designed metal binding sites, Yi Lu demonstrated that, through the design of a copper-binding site proximal to the heme-binding site of myoglobin, heme-copper oxidase-like activity could be introduced into the enzyme.[70,71] Indeed, the resulting enzyme exhibited heme copper oxidase-like O_2-reduction activity. By introducing a tyrosine residue proximal to the active site (mirroring the active site architecture of heme-copper oxidase), the efficiency of the catalyzed O_2 reduction to H_2O was significantly improved, demonstrating the power of tuning the metal coordination spheres for enhancement of ArM activity.[72] In addition, this demonstrates another exciting prospect: the ability to introduce natural metalloenzyme activity into scaffolds with desirable characteristics, like high expression efficiency or improved stability.

In the examples described, the rationale behind the chosen designs of binding sites has relied on the close inspection of near-atomic resolution crystal- or NMR-structures. However, a variety of computational methods have been developed to further facilitate the task. RosettaMatch[73] and RosettaDesign[74] have been successfully applied toward achieving this end. In addition, tools like STAMPS,[75,76] SUNS,[77] and Urantein[78] have been developed for the search and identification of viable three-dimensional motifs, which can be then used as supports to enforce metal binding. It should be noted that these methods cannot be used predictively in the context of catalysis. They do, however, serve as a basis for the identification of candidate scaffolds with optimal properties for metal binding. Ward has recently produced an example wherein the STAMPS algorithm was applied to the identification of facial triad motifs on a variety of scaffolds.[79] The computationally-generated list of candidates was then screened with a variety of metal salts, ultimately yielding a $CuSO_4$/6-phosphogluconolactonase combination that proved a competent ArM peroxidase. This demonstrates that these tools are useful not only for the engineering of new metal binding sites, but for enhancing the likelihood of "serendipitous" metal binding by protein scaffolds.[80]

De novo design of metal-binding proteins has seen great progress, particularly in systems comprised of α-helical bundles for the creation of metal binding sites.[81] These have traditionally been non-catalytic and comprised synthetic peptides, falling outside of the paradigm of ArMs. However, recent examples of *de novo* design have been applied to the creation of catalytic ArMs. Degrado was able to design and express a *due ferri N*-oxidase (DFsc), which was applied for the two-electron oxidation of *p*-aminophenol.[82] With further engineering, researchers were able to furnish catalysis of a relatively rare

transformation in Nature, *N*-oxidation.[83] Using the *p*-aminobenzoate *N*-oxidase AurF as inspiration, researchers in the group performed *in silico* design to mimic the di-iron binding site of this enzyme. The final engineered enzyme proved a competent *N*-oxidase for the hydroxylation of *p*-aminoanisole (Figure 1.7 C).[83] Pecoraro recently reported a *de novo* designed three-helix bundle with a designed Zn(II)-binding site, which exhibited carbonic anhydrase activity (hydration of CO_2 to carbonic acid). Remarkably, the catalytic efficiency of this enzyme was only 1-3 orders of magnitude lower than carbonic anhydrases I-III.[84] These examples provide evidence that *de novo* and *in silico* design can be tremendously successful in enabling the engineering of ArMs.

As mentioned earlier, one of the fundamental limitations of formation of ArMs through dative binding stems from the restricted pool of amino acid functionality that can be introduced into the main-chain of a protein polypeptide. The incorporation of unnatural amino acids serves to address this limitation, as these can be synthetically tailored with a variety of metal-binding functionalities and architectures. Early studies furnished unnatural amino acids through semisynthetic methodology, which, while incredibly important for establishing the viability of this method, is not practical for optimization via genetic engineering as discussed. The development of codon suppression methods for genetically encoding unnatural amino acids into proteins rendered the incorporation process compatible with standard protein expression technology, enabling ribosomal synthesis of UAA-containing proteins.[41,85] It should be noted, however, that engineering the tRNA/tRNA synthetase (aaRS) pairs required by these methods remains a challenging task.

Schultz provided the first example of a metal-binding amino acid to be incorporated into a protein through amber suppression.[86] A tRNA/aaRS pair was engineered to recognize and introduce bipyridyl alanine (BpyAla) into proteins, furnishing the sidechain bipyridine, a ligand capable of binding a wide range of metals (Figure 1.8 A).[87] BpyAla was then successfully introduced near the DNA-binding site of catabolite activator protein. Upon metalation with Cu(II) and Fe(II), this ArM could catalyze the site-specific cleavage of DNA in the presence of air and a mild reducing agent.[87] Genetic incorporation of BpyAla at the dimer interface of LmrR followed by metalation with Cu(II) produced an ArM for catalyzing enantioselective Friedel-Crafts alkylation.[88]

Besides BpyAla, a number of tRNA/aaRS pairs were developed for the incorporation of unnatural tyrosine derivatives for the optimization of the ArM heme-copper oygenases (HCOs) described earlier. Chemical variation of these tyrosine derivatives (Figure 1.8 B-D) allowed tailoring of their redox properties to enable higher efficiency in the catalysis of O_2 reduction.[89-91] Such an effort serves to demonstrate that unnatural amino acids can be introduced to tune stereoelectronic properties in the primary and secondary metal coordination spheres, introducing another layer of control for the chemist.

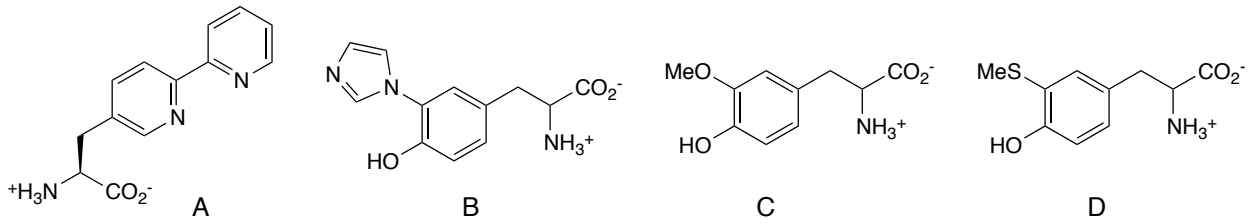
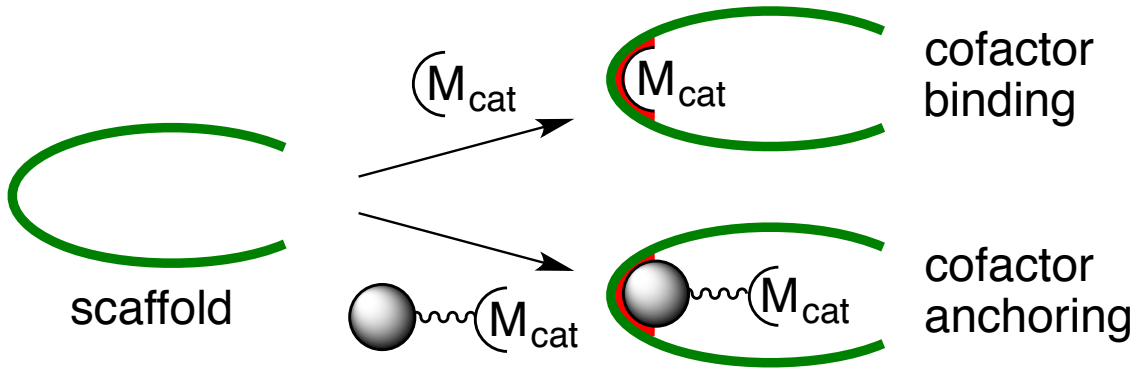


Figure 1.8 A set of metal-binding unnatural amino acids that have been ribosomally incorporated into proteins for ArM catalysis. A) BpyAla, B-D) tyrosine derivatives for HCO catalysis.

1.3.2.2 SUPRAMOLECULAR ARTIFICIAL METALLOENZYMES

The examples above rely on the protein scaffold to define (at least in part) the primary coordination sphere of the catalytic metal center in ArMs. Such a constraint limits the flexibility of this approach. Another class of ArMs utilizes supramolecular interactions to bind metal cofactors without requisite dative coordination of the scaffold to the metal (Figure 1.9). This provides chemists greater freedom to define the primary ligand sphere about the metal rather than relying on proteinogenic functionality to provide the main link between metal and scaffold. The supramolecular approach represents a step toward the decoupling of catalytic roles between the metal center and the protein scaffold, since the primary coordination sphere of the metal (defining reactivity) can be synthetically tuned independent of the secondary coordination sphere (defining selectivity) provided by the scaffold.

Figure 1.9. ArM formation through supramolecular binding directly to cofactor (sections 1.3.2.2.1-1.3.2.2.3) and by non-covalent anchoring (1.3.2.2.4).



The first three portions of this section cover supramolecular cofactor binding that relies primarily on the direct interaction of metal cofactors with the protein scaffold. This approach is appealing in its simplicity, as often one interrogates ArM formation by combination of a pre-formed complex and the target scaffold. Of course, the reality of non-covalent supramolecular interactions is that they can be quite complex, so small perturbations to the metal complex can profoundly impact its ability to be bound by the protein. In addition, because the reactivity of the metal complex pre-defined by the identity of its synthetic ligands, the free complex is a competent catalyst independent of its host scaffold. Selective ArM catalysis therefore necessitates that the protein scaffold possess high affinity for the metal cofactor. Otherwise, free cofactor may catalyze non-selective background reactions, impacting the overall selectivity of the reaction.

The fourth portion of this section describes a strategy to circumvent low affinity scaffold/cofactor interactions by appending a high affinity scaffold-anchoring group to

the metal. For most cases this enables reliable attachment of the metal complex to its target scaffold, but it obviously carries with it drawbacks in terms of the synthetic complexity of anchoring group installation as well as the narrowed scope of potential protein scaffolds.

1.3.2.2.1 EXPLOITING NATIVE COFACTOR BINDING

One class of supramolecular non-covalently bound complexes in ArMs exploits native cofactor binding functionality in a manner analogous to that observed with the repurposing of natural metalloenzymes for dative ArM formation. Heme proteins are the obvious first choice as potential scaffolds using this method, as the binding of heme is largely supramolecular and these proteins are ubiquitous in Nature, providing a variety of starting scaffolds to choose from.[92] Again, methodology must exist to replace their native cofactor with a synthetic construct. Besides some cases of direct expression of apo-forms of hemoproteins, heme-extraction and apo-protein reconstitution methods have long been known for various heme proteins.[93,94]

Myoglobin, one of the most studied proteins in biology, is particularly amenable to extraction of its native Fe(PPIX) cofactor (hemin) and reconstitution with non-native cofactors.[93]

The Watanabe group introduced a number of non-native Fe-porphyrins bearing artificial substituents that were able to garner peroxidase activity with changes in substrate selectivity.[95] Myoglobin has also been shown to be capable of hosting protoporphyrin IX complexes containing metals besides iron. Hartwig recently demonstrated that myoglobin could be successfully reconstituted with a whole panel of M(PPIX) complexes. Substitution with iridium enabled ArM catalyzed C-H carbene insertion and

cyclopropanation.[96] Mutagenesis of this system was pursued to alter the enantioselectivity of the cyclopropanation reaction.[96] Myoglobin has also been demonstrated to be a good host for planar aromatic heme-like complexes, like those based on the salen/salophen architecture. Substitution of myoglobin with Mn(III) and Cr(III) salen/salophen complexes yielded competent sulfoxidases for mildly enantioselective sulfoxidation. Alterations to the substitution pattern on these cofactors led to differential enantioselectivity, in one case inverting the selectivity of the myoglobin ArM.[97]

Cytochrome p450, a heme-dependent monooxygenase that has been extensively explored by protein engineers, has recently been exploited as a scaffold for non-native metal-substituted protoporphyrin IX.[98] Previous work had shown that this enzyme could be expressed containing non-native cofactors when co-expressed in the presence of the heme-loading enzyme ChuA.[99] In a recent study, Hartwig introduced Ir(III)-substituted protoporphyrin IX into the thermostable p450 CYP119 for the intra- and inter-molecular insertion of carbenes into C-H bonds. This ArM was optimized through site-directed mutagenesis to yield a highly efficient and enantioselective hybrid catalyst.[98]

Heme oxygenase has also been explored as an ArM scaffold. This is not explicitly a metalloenzyme, as hemin serves as its substrate for oxidation. However, it has been shown to bind Fe(III)-salophen derivatives. The resulting ArMs were able to interface with native electron transfer partners to perform O_2 reduction to superoxide, suggesting the exciting potential to employ native redox partners to enable non-native reactivity.[100]

1.3.2.2.2 SERENDIPITOUS COFACTOR BINDING

Similar to serendipitous metal binding, serendipitous cofactor binding of non-metalloenzyme scaffolds can also be explored. Again, this expands the scope of viable scaffolds beyond the realm of cofactor-dependent natural metalloenzymes. Imperfect cofactor binding with these systems can be an impediment, leading to low selectivity from poorly-bound free cofactor, multiple cofactor binding, and low cofactor rigidity within the protein scaffold. Despite these issues, the added scaffold scope allows for the evaluation of panels of different candidates to observe desired selectivity. Indeed, multiple selective ArMs have been discovered in this way.

Xylanases are competent scaffolds for the incorporation of different iron porphyrin derivatives.[101] These are appealing ArM scaffolds due to their deep substrate-binding clefts and their wide availability from thermophilic genomes, enhancing their utility as stable, selective environments for metal cofactors. Indeed, xylanase-Fe(porphyrin) ArMs have been applied as peroxidases for the oxidation of electron-rich arenes like guaiacol and *o*-dianisidine.[102] Importantly, the lifetime of the porphyrin catalyst was extended relative to that of the free cofactor, demonstrating another potential advantage of ArMs: catalyst protection. The sequestration of catalysts inside protein scaffolds could reduce the likelihood of auto-inactivation through cofactor damage, a known effect in oxidation catalysis.[103] Fe(porphyrin)-xylanase ArMs have also been applied thioaniosole sulfoxidation with some enantioselectivity.[102] Finally, enantioselective epoxidation has been achieved using Mn-porphyrin incorporated into xylanase as well.[104,105]

Unsurprisingly, serum albumins, by virtue of their capacity for binding a variety of substrates, have been demonstrated to be adequate hosts for various metalated porphyrin derivatives.[106] Gray reported a detailed study into the binding of Ga(III) and Mn(III) corrole cofactors to human serum albumin (HSA), determining the approximate location of the corrole within the scaffold.[107] In a follow-up study Fe- and Mn-substituted corroles were found to be competent catalysts for the enantioselective sulfoxidation of a variety of prochiral thioethers in the presence of H_2O_2 or iodosylbenzene.[108] Particularly encouraging about these systems was the solution of a hemin-bound HSA crystal structure, which enabled rational mutagenesis to alter the identity of the axial ligand, echoing the strategies applied for the design of metal-binding sites.[109] Substitution of the HSA mutant with Fe(PPIX) generated an ArM peroxidase for the one-electron oxidation of phenols,[110] while Mn(PPIX) substitution produced an ArM superoxide dismutase.[111] Further exploration with heme-like Mn(salen) enabled sulfide oxidation catalysis, though no enantioselectivity was observed with this system.[112] Switching to Co(salen) enabled moderately enantioselective sulfoxidation.[113] Finally, in a departure from redox catalysis, Reetz introduced a Cu(II)-phthalocyanine complex into a variety of serum albumins for the enantioselective Diels-Alder cycloaddition of azachalcones and cyclopentadiene.[114]

The hydrophobic interfacial cavity of the LmrR dimer has been explored by Roelfes group as a viable supramolecular binding site for planar aromatic cofactors, particularly those bearing bidentate aromatic nitrogen ligands.[115] A Cu(II)-phenanthroline complex was successfully introduced by this method and the subsequent hybrid could catalyze the Friedel-Crafts alkylation of numerous indoles with high enantioselectivity.

Importantly, perturbation of the binding site by removal of a critical tryptophan residue resulted in the loss of enantioselectivity, suggesting the importance of this residue for the binding of the cofactor.[115]

NikA, a periplasmic nickel-binding protein from *E. coli*, has been explored as a scaffold for supramolecular binding of metal cofactors because of its known propensity to bind anion-bearing metal complexes like Fe(EDTA) through salt-bridging interactions.[116] A number of Fe(II)-complexes bearing EDTA-like tetradentate cofactors have been explored for binding to NikA. Importantly, these have successfully promoted the NaOCl-driven sulfoxidation of a panel of aryl thioethers.[117] While no enantioselectivity was observed with this system, the ArM was able to promote substrate selectivity toward shape-matched thioethers, an effect that was predicted through computational substrate design.[117]

1.3.2.2.3 ENGINEERING COFACTOR BINDING

The cofactor binding strategies detailed so far have relied on the action of existing natural cofactor-binding sites or chance affinity of scaffold proteins for metal cofactors. However, it is possible to introduce cofactor-binding affinity into protein scaffolds through engineering of the scaffold. Theoretically, it should be possible to accomplish this through computational rational design, though to date, no such effort has been accomplished. Rather, metal cofactor binding has been introduced in the context of antibody engineering, which has been greatly successful in promoting high affinity binding of a variety of substrates.

Antibodies have been successfully raised for binding catalytic metal complexes, thus creating antibody ArMs.[118] In 1989, a site-specific ArM peptidase was produced by

raising antibodies against a trien-Co(III)-peptide hapten, illustrating the remarkable power of this engineering method.[119] Schultz demonstrated the generation of an antibody ArM for the binding of methylmesoporphyrin IX, which could catalyze the metalation of Fe-protoporphyrin IV. The antibody was also found to bind Fe-mesoporphyrin to enable ArM peroxidase activity. Further studies enabled the optimization of antibody-based ArM peroxidases with a phage display affinity-based selection strategy, comprising one of the few selection-based mutagenesis optimizations of an ArM.[120] Finally, antibodies have been successfully raised against a number of other porphyrin derivatives, all generating various ArM peroxidases.[118,121,122]

1.3.2.2.4 NON-COVALENT COFACTOR ANCHORING

The examples of cofactor binding shown above comprise instances in which the cofactor ligand(s) (i.e. protoporphyrin IX, salen, salophen, trien, EDTA, etc.) is directly bound through non-covalent interactions to the protein scaffold. Therefore, the affinity of the protein scaffold for the cofactor depends on the identity of the ligand(s). This poses a constraint: changes to the ligand(s) can ablate affinity of the protein scaffold for the cofactor. Therefore, scientists have sought to engender cofactor binding independent of the ligands in the primary coordination sphere of the metal. This has led to the development of non-covalent cofactor "anchoring" methods, wherein the metal cofactor is linked to a chemical motif with high affinity for the target scaffold. Such a strategy grants flexibility in the determination of the metal ligand sphere, but constrains the number of scaffolds that can be explored, as high-affinity for the anchoring group is a prerequisite with this ArM formation strategy. Of course, in theory one may interrogate cofactor binding to any scaffold, but in practice, chemists use this approach with a

specific target scaffold in mind, given the synthetic complexity of introducing said anchoring groups.

Whitesides reported one of the first instances of an ArM, formed using a biotinylated Rh-bisphosphine complex and the protein avidin to form an enantioselective olefin hydrogenase.[123] Avidin displays extremely tight affinity for biotin, with a $K_d \sim 10^{-12}$-10^{-15}. The binding event is rapid and effectively irreversible. Furthermore, the biotin-binding site in avidin leads to a defined chiral pocket in the structure. It was found that the incorporation of biotin-linked chiral hydrogenation catalyst Rh-Pyrphos into avidin could override the small molecule complex's sense of enantioinduction, instead leading to the opposite enantiomer of hydrogenation product.[124] Ward showed that streptavidin, a related biotin binding protein from *Streptomyces avidinii*, can be similarly employed as an ArM scaffold with the advantage of greater expression efficiency.[125] Research in the Ward laboratory has proven the versatility of the biotin-linkage strategy, with Rh-bisphosphine ArMs for hydrogenation[126], Ru-pianostool ArMs for transfer hydrogenation[127], Ru-Hoveyda-Grubbs ArMs for olefin metathesis[128], and Pd-*N*-heterocyclic carbene/Pd-phosphine ArMs for cross coupling (Figure 1.10).[129] These systems have consistently yielded ArMs with high selectivity, and the reliability of the streptavidin system has enabled systematic investigations on the mechanistic role of the scaffold and metal ligands during catalysis.[126] In addition, scaffold-accelerated reactivity has been observed with these systems, an often-hypothesized effect of incorporating transition metal catalysts in the context of protein scaffolds.[130] Beyond this, Ward has recently demonstrated the genetic optimization of a Ru-streptavidin-based olefin metathetase *in vivo* based on a fluorescence-based screening assay. This, again, presents

one of the few ArM systems to be optimized by directed evolution.[131] Of particular relevance to this work are the numerous detailed crystallographic and computational design studies that have been performed using the biotin/streptavidin system.[132,133] All of these examples serve as proof that the cofactor anchoring strategy can be immensely successful and provides a general platform to explore a number of different catalytic metal complexes in the context of an ArM.

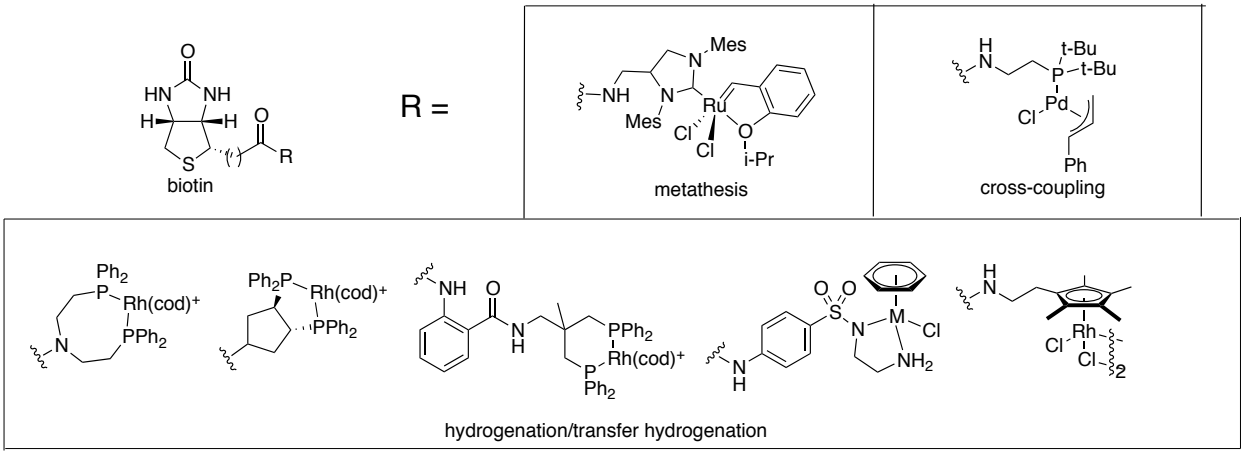


Figure 1.10 Different metal catalysts that have been appended to the biotin anchor.

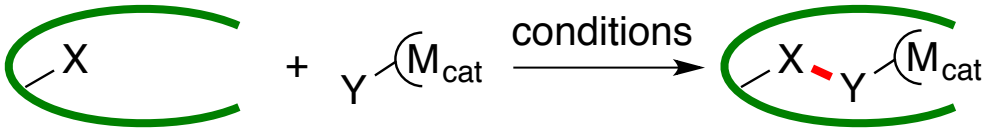
Another anchoring system has stemmed from the Ward laboratory using carbonic anhydrase as a protein scaffold. The Zn(II)-center in carbonic anhydrase displays remarkable affinity for aryl-sulfonamide moieties, which have been demonstrated to be quite efficacious inhibitors of these enzymes. Tethering Ir-pianostool and Ru-pianostool complexes to arylsulfonamides provides a facile method to introduce these complexes into the asymmetric environment of carbonic anhydrase, which enables enantioselective transfer hydrogenation.[134,135] A recent report detailed the computational design of this ArM to effect stronger cofactor affinity and impart enhanced enantioselectivity to the catalyzed transfer hydrogenation reaction.[136] In addition, a Grubbs-Hoveyda-type catalyst

could be appended to an aryl sulfonamide to achieve olefin metathesis with this system.[137]

A number of other anchoring strategies have been explored to a lesser extent. These all employ the same principle: 1) identify an effective anchoring group for the target scaffold, and 2) attach the metal cofactor/ligand through an interstitial linker. Given the high affinity of antibodies for their respective antigen, antibodies have been explored for non-covalent cofactor anchoring. Estradiol-binding antibody 7A3 was used as a scaffold for a variety of estradiol-linked Fe- and Mn- containing cofactors, which generated ArM peroxidases and sulfoxidases.[138,139] Further work using a neocarzinostatin scaffold engineered for the binding of testosterone has shown that other Fe(III), Zn(II), and Cu(II) cofactors could be linked to testosterone to generate ArMs based on this scaffold.[140,141] Ibuprofen, a known substrate for human serum albumin, was appended to an Fe(II)-BPMEN-like catalyst to generate an HSA ArM for the sulfoxidation of aryl thioesters.[142] An interesting example employed a bipyridine-modified hemin cofactor for anchoring to myoglobin. The resulting ArM was used for Diels-Alder cycloaddition.[143] Finally, a fatty acylated 2,2-dipyridylamine-Rh transfer hydrogenation cofactor was introduced into β-lactoglobulin, generating an ArM transfer hydrogenase.[144] A hallmark of these systems is the low selectivity observed during catalysis. In many of these cases, the non-covalent binding affinity is not very strong (micromolar K_d), increasing the likelihood of background non-selective catalysis. In addition, despite the deliberate anchor/scaffold pairing, many of these systems operate with only the assumption that the cofactor is located in the predicted position, since no further characterization is provided. Finally, it should be noted that many of these ArM cofactors employ relatively long

linkers. Excessive linker length enables cofactor movement, which can have a profound negative impact on selectivity, particularly when the scaffold has no necessary affinity for the metal complex itself, as is the case when utilizing non-covalent cofactor anchoring for ArM formation.[133]

1.3.2.3 COVALENT ARTIFICIAL METALLOENZYMES



Scheme 1.2 A general scheme of the covalent formation of artificial metalloenzymes

The covalent grafting of metal cofactors presents a logical step in the progression of ArM-formation strategies. For this approach, metal complexes must be linked to a reactive group to promote formation of a covalent bond between the protein scaffold and the cofactor. The reaction to generate this covalent linkage is termed "bioconugation." This method bears similarities to the non-covalent cofactor anchoring approach, since the main link between the metal and scaffold is predicated not on the identity of the metal complex, but rather on the anchoring group present. Therefore, this enables flexibility as to the choice of catalytic metal complex. In addition, this method surpasses the scope of supramolecular anchoring strategies with respect to scaffold choice. While some bioconjugation methods do depend on scaffold-specific chemical properties, a number of highly general bioconjugation methods have been developed, expanding the range of potential scaffolds to virtually any protein that can be solubly expressed and isolated.[145] As previously noted, non-covalent linkage of metal cofactors suffers from the potential for cofactor dissociation from its protein scaffold, increasing the potential for background

non-selective catalysis by the free complex. With covalent methods, the metal cofactor is irreversibly linked to the protein scaffold, effectively eliminating the possibility for such a background to occur. Of course, while the versatility of covalent bioconjugation strategies is exceptional, it should be noted that cofactor synthesis can present a major challenge in its implementation, as reactive metal complexes must be compatible with reactive anchoring groups and the anchoring groups themselves must be stable in the presence of water and polar functionality. Despite this constraint, covalent bioconjugation has been quite successful for ArM formation. Herein will be introduced examples of the most common methods for covalent bioconjugation of metal cofactors to protein scaffolds, ranging from narrow scaffold-specific native mechanism-based methods to highly general protein engineering-based approaches.

1.3.2.3.1 HYDROLASE MECHANISM-BASED COVALENT BIOCONJUGATION

Covalent bioconjugation methods have been developed to exploit the known mechanisms of cysteine and serine hydrolases. These enzymes feature nucleophilic cysteine/serine residues at the core of their active sites. Nucleophilicity in these sidechains is promoted by networks of activating residues (Ser-His-Asp/Glu for serine hydrolases, Cys-His-(Asp) for cysteine hydrolases), thereby enforcing a well-defined active site architecture.[146,147] These nucleophilic residues are known to form covalent adducts with a variety of electrophilic irreversible inhibitors.[148] One of the key advantages of this approach stems from the fact that native enzymes can be used directly for bioconjugation, obviating the need for protein engineering, a hallmark of other covalent methods that will be described. The second advantage of this approach arises from the fact that the nucleophilic residue sits within the active site of the enzyme, which

is often somewhat buried in the protein scaffold. Judicious selection of active sites with deep cavities allows for the thorough encapsulation of the metal cofactor, a desirable feature to achieve control of the metal primary/secondary coordination spheres.[149] Finally, since these electrophilic compounds inhibit native enzyme activity, the ablation of hydrolase activity can be used as a proxy for evaluating the success of the bioconjugation reaction.

One of the first ArM scaffolds to be explored for covalent linkage of metal cofactors was the enzyme papain.[150] This enzyme reacts with α-haloketone electrophiles, maleimides, electrophilic phosphonates, and epoxides to generate irreversible covalent adducts.[148] An α-bromoketone-linked phosphite cofactor was successfully incorporated into the papain scaffold enabling metalation with $Rh[COD]^+$ for the hydrogenation of methyl 2-acetamidoacrylate.[151] Despite the large cavity of papain, low enantioselectivity was observed with this ArM, with size mismatch of the scaffold being cast as the potential culprit.[151] A maleimide-substituted manganese(III) salen complex bioconjugated to papain was demonstrated to be a competent ArM for oxidation chemistry, albeit with low selectivity.[152] Maleimide bipyridine complexes were also successfully introduced into the papain scaffold, yielding low-efficiency and low-selectivity ArMs.[153] More efficient ArMs were generated for Diels-Alder cycloaddition using an α-chloroketone-linked Ru(III)-phenanthroline complex in conjunction with papain.[154] Papain hydrogenases based on a number of different Ru-pianostool and Ru-Cp* complexes tethered to maleimide fragments enabled the catalytic reduction of NAD^+ and $NADP^+$.[155] Further efforts extended the reactivity of these papain hybrids to transfer hydrogenation, albeit with no enantioselectivity.[156,157] Only by furnishing a peptide tail on an epoxide-

linked Rh-Cp*/Ru-pianostool complex was enantioselectivity observed with these papain ArM hydrogenases. The peptide fragment was introduced as a recognition element to enforce cofactor orientation in the papain active site.[158] This illustrates the necessity to introduce rigidity in ArMs in order to adequately observe selectivity.

In a similar vein, serine hydrolases have been demonstrated as competent ArM scaffolds. ArMs based on serine-hydrolase mechanism-based bioconjugation will be the focus of chapter two. These enzymes are covalently inhibited by electrophilic phosphonates and appropriately substituted α-haloketones. One of the notable advantages of the use of phosphonate inhibitors as covalent anchors in the bioconjugation to serine hydrolases arises from the fact that these can leverage the oxyanion hole, a general feature of this class of enzymes.[146] This motif provides a set of hydrogen bonds that stabilize the tetrahedral intermediate generated during native ester/amide hydrolysis. In the context of serine hydrolase ArMs, the oxyanion hole can stabilize the covalent phosphonate adduct and rigidify the link between scaffold and cofactor. Cofactor rigidity has been demonstrated to play a major role in promoting selectivity in ArM catalysis, so this strategy offers a relatively simple framework to achieve this end.[133,159]

Van Koten demonstrated the viability of employing serine hydrolases like lipase and cutinase for the anchoring of a variety of palladium(II) and platinum(II) pincer complexes (Figure 1.11 A).[160,161] However, these artificial metalloproteins were not demonstrated to be catalytic. More recently, a number of phosphonate-linked Ru-Hoveyda-Grubbs metathesis complexes was appended to cutinase for ring-closing metathesis (Figure 1.11 C).[162] Similarly, a phosphonate-linked Rh-*N*-heterocyclic carbene catalyst was successfully introduced into lipases cutinase and CalB (Figure 1.11 B). This

enabled the enantioselective hydrogenation of acetamidoacrylate and acetophenone, with observed substrate selectivity in competition reactions.[163] Chymotrypsin, which has been demonstrated to be an effective platform for covalent inhibition with electrophilic phosphates and phosphonates (as will be discussed in chapter two), has also been shown to be alkylated by α-haloketones bearing recognition peptides for promoting specificity toward the chymotrypsin S1 binding site. This strategy has been successfully applied in covalently linking a Ru-Grubbs-Hoveyda catalyst to chymotrypsin.[164]

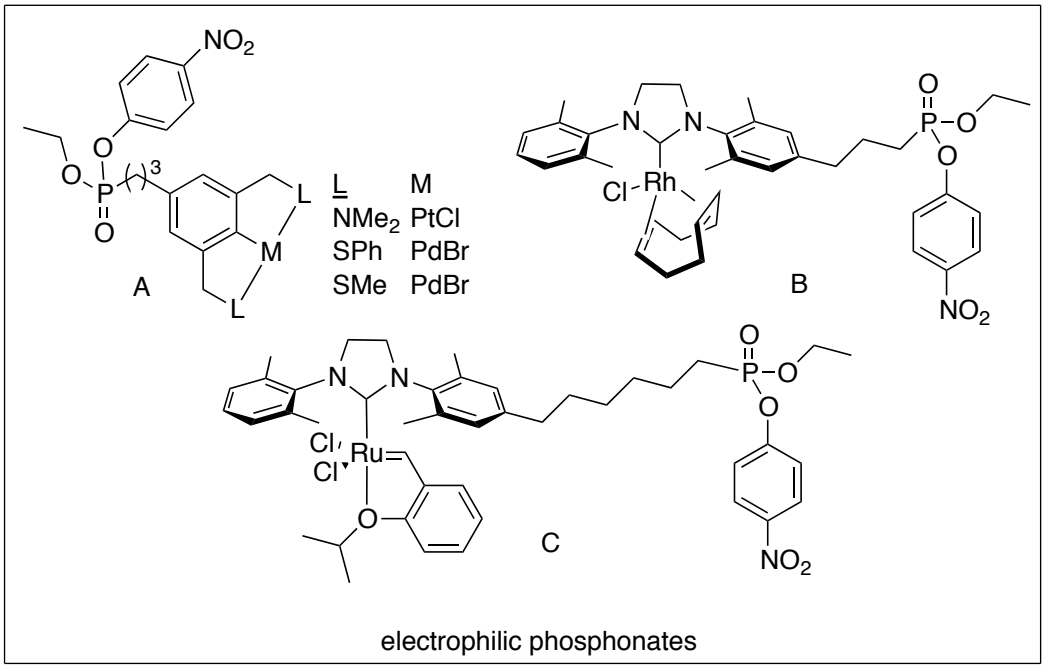


Figure 1.11 Several electrophilic phosphonate-linked complexes synthesized for bioconjugation to serine hydrolases. A) Pt/Pd pincer complexes B) NHC Rh hydrogenation catalyst C) Hoveyda-Grubbs catalyst

1.3.2.3.2 CYSTEINE-BASED COVALENT BIOCONJUGATION

The sulfhydryl moiety of cysteine has proven to be an attractive functionality for covalent bioconjugation due to its inherent nucleophilicity at physiological pH.[165] Cysteines are known to react readily with a number of electrophilic moieties such as α-halocarbonyl compounds, thioesters, and maleimides. In addition, they can form covalent adducts through disulfide exchange.[165] Because cysteines can be introduced through mutagenesis, one need only introduce a cysteine residue at the desired position on a protein scaffold (or utilize a pre-existing cysteine) to enable bioconjugation at that site. Of course, this comes with the prerequisite of removing competing nucleophiles (cysteine or otherwise) from the scaffold, an engineering task that might be detrimental to the stability of the protein. In spite of this drawback, cysteine bioconjugation has proven to be a reliable, operationally simple method of covalently grafting metal cofactors to protein scaffolds.

Some excellent examples of the power of the method and of the ability for ArMs to promote exceptional selectivity are seen in the cysteine-ArM nucleases. Fe(EDTA) and metal-phenanthroline complexes linked to α-haloketone anchors have been successfully appended to DNA-binding proteins, enabling site-selective DNA cleavage in the presence of an oxidant like H_2O_2 or reductant/O_2 mixtures. These operate through the generation of free hydroxyl radicals proximal to the site of the metal, promoting DNA cleavage in the general region near this position. Notably, these methods are reliable enough to form a part of the commercial biochemical toolbox.[166,167] More specific oxidative cleavage of DNA has been enabled through the use of maleimide-linked Mn-tetramethylpyridinium

porphyrin complexes attached to the T4 DNA clamp protein.[168] This remarkable system was able to achieve high site selectivity for AAA motifs on dual-stranded DNA.

ArMs synthesized by cysteine modification have also been successfully applied for enantioselective catalysis. Distefano produced an early example wherein an iodoacetamide-linked phenanthroline ligand was introduced site specifically to the interior of the adipocyte lipid binding protein (ALBP). This was subsequently metalated with copper(II), generating an ArM that could successfully catalyze the enantioselective hydrolysis of esters and amides.[169,170]

Maleimide-substituted cofactors have been particularly useful for covalent linkage to cysteines, since their relatively diminished reactivity makes them more compatible with complex synthetic methodology. In Chapter 2, an example of a maleimide-linked manganese(II) terpyridine cofactor will be discussed in its application to the bioconjugation of nitrobindin and subsequent oxidation chemistry (Figure 1.12 A).[171] The oxidative manifold has been explored with a number of other maleimide-linked polydentate nitrogen cofactors in conjunction with a variety of protein scaffolds (Figure 1.12 B and C).[144] Reetz showed the versatility of the thermostable TIM-barrel protein tHisF as a scaffold for the generation of ArMs with a variety of maleimide-linked metal complexes.[172] Hoveyda-Grubbs catalysts bearing maleimide functionality have also been introduced into nitrobindin[173] and FhuA,[171] signaling a trend toward the exploration of barrel-shaped proteins as ArM scaffolds due to their "vacant space," the unique chemical environment that enables control of secondary coordination sphere around the metal. In the case of FhuA, the metathease was applied toward ring-opening metathesis polymerization, garnering a slight preference for cis-alkene content relative to the free

small-molecule catalyst.[171] More explorations into "vacant space" have employed a Rh-Cp maleimide cofactor within nitrobindin for alkyne polymerization.[174] In an increase toward meso-scale catalysis, chaperonin thermosome protein cage was linked to TEDETA-maleimide ligand, which was metalated with $CuBr_2$ to generate an effective atom-transfer radical polymerization ArM. Again, confining a polymerization catalyst within a protein cage generated narrower molecular weight.[174]

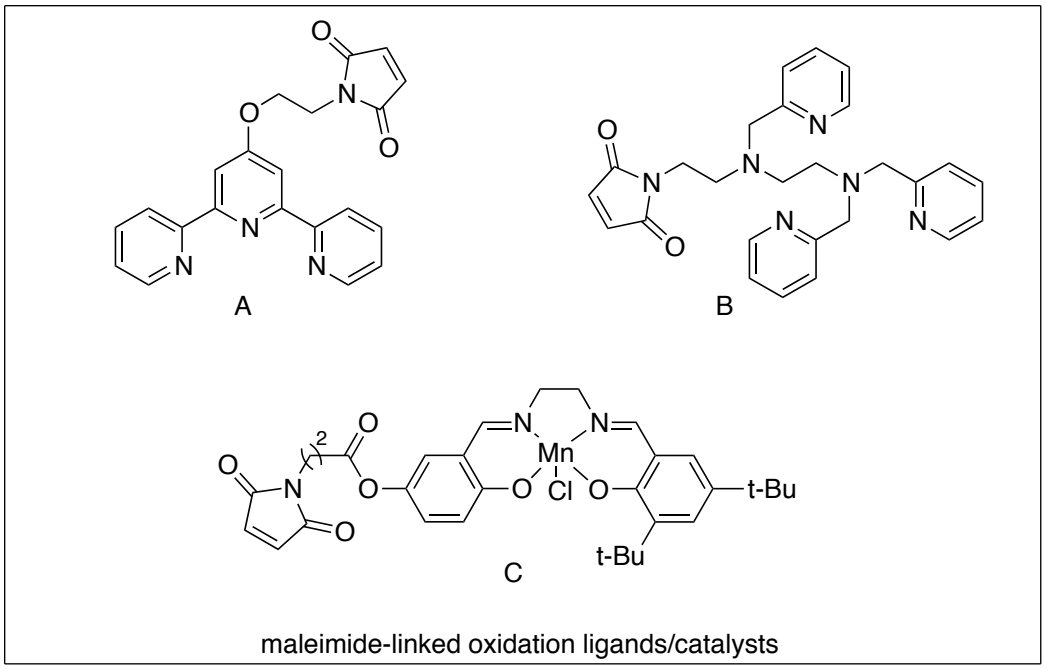


Figure 1.12 Maleimide-linked ligands/catalysts for oxidation catalysis. A) Terpyridine maleimide[171] B) maleimide tetradentate ligand[144] and C) malemide salen[152]

The versatility of the cysteine modification method has enabled the exploration of trends in ArM design. It should be noted that, in spite of the impressive variety of ArMs that has been achieved using cysteine modification, there are few examples that have truly demonstrated the power of ArMs to promote selectivity. This may be due to a

number of issues. Problems with multiple-addition of cofactor have been known to plague cysteine/electrophile bioconjugation, since other nucleophilic residues can compete for the electrophilic anchor of the cofactor, as mentioned earlier.[39,151] Another issue that has been shown to adversely affect covalent ArM selectivity is the improper positioning of the cofactor at sites that do not enable proper metal encapsulation or sufficient substrate/scaffold interactions to promote selectivity.[145] The Roelfes group presented a solution to poor cofactor encapsulation by anchoring an iodoacetamide-linked Cu(II)-phenanthroline complex at the interface of the LmrR dimer. The resultant ArM proved to be exceptionally efficient in catalyzing the Diels-Alder cyclization between azachalcones and cyclopentadiene and the hydration of azachalcones with high enantioselectivity.

In other cases, low selectivity may result form poor cofactor localization within the protein scaffold, which can arise from excessive linker flexibility or from poor conformational restriction imposed by the protein scaffold. The latter highlights one of the main drawbacks of covalent linkage for ArM formation –given that the only requisite interaction for creation of these ArMs is the covalent bond formed between scaffold and cofactor, secondary interactions that conformationally restrict the metal in supramolecular/dative ArMs are not necessarily present with this method. To address this issue, Lu adopted a dual-anchoring strategy to restrict the motion of a bis-methane-thiosulfonate-subsituted Mn-salen complex.[159] Using two appropriately positioned cysteine residues inside the apo-myoglobin scaffold, bioconjugation of the cofactor was enabled by disulfide exchange at both cysteines, pinning the cofactor in a restricted

position and enabling higher enantioselectivity in thioanisole sulfoxidation than that observed for singly-anchored counterparts.[159]

1.3.2.3.3 STRAIN-PROMOTED AZIDE/ALKYNE CYCLOADDITION BIOCONJUGATION

Cysteine bioconjugation has enabled great diversity in the elaboration of ArMs. The ability to control the site of cofactor anchoring through genetic incorporation of cysteines has proven to be a valuable and flexible tool for chemists. On the other hand, cross-reactivity of electrophilic cofactors with off-target nucleophiles necessitates the genetic removal of these nucleophiles and precludes the use of cysteine bioconjugation in complex media like cellular lysate. Both of these issues diminish the viability of cysteine bioconjugation as a general platform for library-based evolutionary approaches.

The discovery of bioorthogonal reactions has enabled the potential to leverage these reactions for the covalent bioconjugation of chemical species to biomolecules without the sort of cross-reactivity described above.[175] Development of amber-codon suppression for ribosomal incorporation of unnatural amino acids into proteins has enabled the genetic introduction of *p*-azidophenylalanine, which bears an azide functional handle.[86] This azide functionality can be employed for the strain-promoted azide/alkyne cycloaddition (SPAAC) reaction, a bioorthogonal reaction. With this in mind, coworkers in this laboratory were able to genetically incorporate an azidophenylalanine residue into the tHisF scaffold, to which a variety of strained bicyclononyne (BCN)-substituted cofactors could be readily bioconjugated. BCN-linked manganese(II) and copper(II) terpyridine complexes were introduced into the scaffold, along with a BCN-linked dirhodium tetracarboxylate catalyst (Figure 1.13). The latter was demonstrated to

successfully catalyze carbene insertion into Si-H bonds as well as cyclopropanation. This early system did not demonstrate any selectivity in catalysis, an issue thought to arise from incomplete encapsulation of the dirhodium cofactor within the tHisF scaffold. In chapter four, a selective dirhodium BCN ArM will be described based on the *Pfu* prolyl oligopeptidase protein scaffold, which exhibits more thorough encapsulation of the metal to achieve control of ArM selectivity.[176]

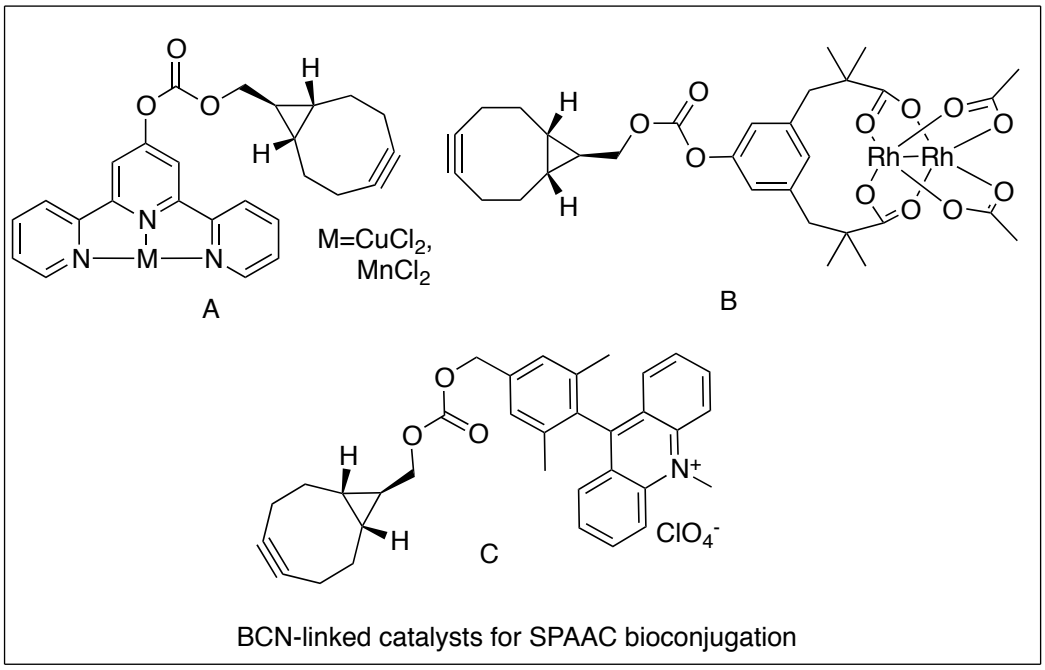


Figure 1.13 BCN-linked cofactors for SPAAC bioconjugation A) Mn/Cu BCN-terpyridine[145], B) Dirhodium BCN[145], and C) acridinium BCN[177]

1.3.3 THE CHARACTERIZATION OF ARTIFICIAL METALLOENZYMES

An important aspect in the development of ArMs is the proper characterization of these hybrid catalysts. Often, ArM engineers fall into the trap of utilizing low-resolution biochemical methodology to characterize these hybrid biocatalysts. However, due to the

intricacies of ArM catalysis, proper ArM characterization requires high-resolution, information-rich methods.

The examples of ArMs described above employ a number of different characterization methods to confirm the formation of the metal/protein hybrid and, in some cases, the fate of the catalyst during and after catalysis. For routine characterization of covalent artificial metalloenzymes, mass spectrometric techniques have proven indispensable. Intact protein MALDI and high-resolution ESI-MS enable the rapid evaluation of predicted bioconjugation and extent of bioconjugation, parameters that are critical for confirmation of the identity of the formed catalyst.[151,176,178] Protein digestion followed by LC/MS/MS enables further confirmation of the site of covalent attachment of a metal cofactor as well as sites of protein modification that may occur during the course of catalysis, as will be described in chapter four.[171,176] Furthermore, ICP-MS/OES is often employed to assess the stoichiometry of metal relative to protein scaffold.[51] UV/Vis, EPR, and Raman spectroscopies are regularly employed to probe the coordination environment of the metal in ArMs, much as is seen in the characterization of natural metalloenzyme sand small molecule metal complexes.[88] NMR can also be applied for the characterization of ArM systems, though its utility may be diminished by the high molecular weight of ArMs and the potential paramagnetism of the catalytic metal. Efforts toward NMR characterization of ArMs will be described in chapter two. Routine characterization of the scaffold folds following ArM formation can be carried out by analyzing native tryptophan fluorescence and circular dichroism spectra. These have proven valuable for the assessment of compatibility of ArMs with conditions for formation and catalytic reaction conditions.[157] In addition, Cotton effects in circular

dichroism can be used to confirm the asymmetric environment of bound metal.[142] Finally, X-ray crystallography stands as one of the most convincing methods for the confirmation of ArM formation and for the confirmation of metal location and coordination environment. This, of course, bears the caveat that metal complexes may be conformationally flexible and may not be present at full occupancy, thereby complicating characterization of the ArM by this method. That being said, even just a high-resolution model of the scaffold protein itself can often provide an enormous degree of predictive power in ArM design. This will be explored further in chapters two and four.

1.4 CONCLUSIONS

Artificial metalloenzymes have the potential to serve as a powerful platform for the development of novel selective catalysts. However, their application must be logical—that is, artificial metalloenzymes cannot be made through uncharacterized random combinations of metal and protein. Chemical logic must be applied. Structural aspects of metal encapsulation, structural stability, and catalyst robustness are all necessary for the development of powerful artificial metalloenzyme platforms. In addition, at the heart of an effective artificial metalloenzyme should lie a good catalyst, with predictable properties and, above all else, interesting reactivity. Proper characterization is needed to understand the mechanisms that underlie the function of these complex hybrid catalysts. This dissertation attempts to introduce: 1) a powerful, robust scaffold for artificial metalloenzymes (chapter three), 2) applications of these artificial metalloenzymes toward the goal of site-selective catalysis (chapters two and four), and 3) reproducible and reliable methods for the characterization of these hybrid catalysts (chapters two, three, and four).

The innovative catalysts that chemists have achieved through synthetic molecule building can now be diversified, optimized, and expanded with help from the beautifully complex molecular machinery of biology. Through these methods, we hope to achieve some of the most concise and efficient molecular transformations yet, converting simple organic molecules into highly complex and useful compounds with maximal efficiency and minimal waste. Fundamentally, artificial metalloenzymes epitomize an incredible future. As traditional fields of chemistry become more integrated with biological methods, the creation of new arrangements of matter—novel molecules—is becoming more advanced, more efficient, and more powerful than ever before.

1.5 REFERENCES

(1) Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry*, illustrated edition. University Science, 2005.

(2) Corey, E. J. **1989**.

(3) Bergman, R. G. *Nature* **2007**, *446* (7134), 391–393.

(4) Gutekunst, W. R.; Baran, P. S. *Chemical Society Reviews* **2011**, *40* (4), 1976–1991.

(5) Wu, X. F.; Anbarasan, P.; Neumann, H.; Beller, M. *Angew. Chem. Int. Ed. Engl.* **2010**, *49* (48), 9047–9050.

(6) *Metal Catalysts in Olefin Polymerization*; Guan, Z., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; Vol. 26.

(7) Hoveyda, A. H.; Zhugralin, A. R. *Nature* **2007**, *450* (7167), 243–251.

(8) Davies, H. M. L.; Bois, Du, J.; Yu, J.-Q. *Chemical Society Reviews* **2011**, *40* (4), 1855–1856.

(9) Borovik, A. S. *Chemical Society Reviews* **2011**, *40* (4), 1870–1874.

(10) Doyle, M. P.; Duffy, R.; Ratnikov, M.; Zhou, L. *Chem. Rev.* **2010**, *110* (2), 704–724.

(11) Davies, H. M. L.; Parr, B. T. *Rhodium Carbenes*; John Wiley & Sons, Inc, 2013; pp 363–403.

(12) Que, L.; Tolman, W. B. *Nature* **2008**, *455* (7211), 333–340.

(13) Hansen, J.; Davies, H. M. L. *Coordination Chemistry Reviews* **2008**, *252* (5-7), 545–555.

(14) Walsh, P. J.; Kozlowski, M. C. **2009**.

(15) Knowles, W. S. *Angew. Chem. Int. Ed. Engl.* **2002**, *41* (12), 1998–2007.

(16) Noyori, R. *Angew. Chem. Int. Ed. Engl.* **2002**, *41* (12), 2008–2022.

(17) Sharpless, K. B. *Searching for new reactivity (Nobel lecture).*; 2002; Vol. 41, pp 2024–2032.

(18) Zhang, W.; Loebach, J. L.; Wilson, S. R.; Jacobsen, E. N. *J Am Chem Soc* **2002**, *112* (7), 2801–2803.

(19) Cheng, C.; Hartwig, J. F. *J Am Chem Soc* **2014**, *136* (34), 12064–12072.

(20) Mkhalid, I. A. I.; Barnard, J. H.; Marder, T. B.; Murphy, J. M.; Hartwig, J. F. *Chem. Rev.* **2009**, *110* (2), 890–931.

(21) Liao, K.; Negretti, S.; Musaev, D. G.; Bacsa, J.; Davies, H. M. L. *Nature* **2016**, *533* (7602), 230–234.

(22) Das, S.; Incarvito, C. D.; Crabtree, R. H.; Brudvig, G. W. *Science* **2006**, *312* (5782), 1941–1943.

(23) Breslow, R. *Acc. Chem. Res.* **1995**, *28* (3), 146–153.

(24) Davis, H. J.; Phipps, R. J. *Chem. Sci.* **2016**.

(25) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303* (5655), 186–195.

(26) Benkovic, S. J.; Hammes-Schiffer, S. *Science* **2003**, *301* (5637), 1196–1202.

(27) Huang, P.-S.; Boyken, S. E.; Baker, D. *Nature* **2016**, *537* (7620), 320–327.

(28) Lewis, J. C.; Coelho, P. S.; Arnold, F. H. *Chemical Society Reviews* **2011**, *40* (4), 2003–2021.

(29) Julsing, M. K.; Cornelissen, S.; Bühler, B.; Schmid, A. *Curr Opin Chem Biol*

2008, *12* (2), 177–186.

(30) Butler, A.; Sandy, M. *Nature* **2009**, *460* (7257), 848–854.

(31) CRICK, F. *Nature* **1970**, *227* (5258), 561–563.

(32) Sambrook, J.; Russell, D. W. **2001**, No. v. 1.

(33) Francis, M. B.; Jamison, T. F.; Jacobsen, E. N. *Curr Opin Chem Biol* **1998**, *2* (3), 422–428.

(34) Lichtor, P. A.; Miller, S. J. *Nat Chem* **2012**, *4* (12), 990–995.

(35) Chica, R. A.; Doucet, N.; Pelletier, J. N. *Current Opinion in Biotechnology* **2005**, *16* (4), 378–384.

(36) Labrou, N. E. *Curr. Protein Pept. Sci.* **2010**, *11* (1), 91–100.

(37) Romero, P. A.; Arnold, F. H. *Nat. Rev. Mol. Cell Biol.* **2009**, *10* (12), 866–876.

(38) Esvelt, K. M.; Carlson, J. C.; Liu, D. R. *Nature* **2011**, *472* (7344), 499–503.

(39) Lewis, J. C. *ACS Catal.* **2013**, *3* (12), 2954–2975.

(40) Yoon, T. P.; Jacobsen, E. N. *Science* **2003**, *299* (5613), 1691–1693.

(41) Lewis, J. C. *Curr Opin Chem Biol* **2015**, *25*, 27–35.

(42) Yamamura, K.; Kaiser, E. T. *J. Chem. Soc., Chem. Commun.* **1976**, No. 20, 830–831.

(43) Okrasa, K.; Kazlauskas, R. J. *Chem. Eur. J.* **2006**, *12* (6), 1587–1596.

(44) Fernández-Gacio, A.; Codina, A.; Fastrez, J.; Riant, O.; Soumillion, P. *Chembiochem* **2006**, *7* (7), 1013–1016.

(45) Håkansson, K.; Carlsson, M.; Svensson, L. A.; Liljas, A. *J. Mol. Biol.* **1992**, *227* (4), 1192–1204.

(46) Håkansson, K.; Wehnert, A.; Liljas, A. *Acta Crystallogr. D Biol. Crystallogr.* **1994**, *50* (Pt 1), 93–100.

(47) Fujieda, N.; Hasegawa, A.; Ishihama, K.-I.; Itoh, S. *Chemistry – An Asian Journal* **2012**, *7* (6), 1203–1207.

(48) Wong, K. K. W.; Mann, S. *Advanced Materials* **1996**, *8* (11), 928–932.

(49) Abe, S.; Niemeyer, J.; Abe, M.; Takezawa, Y.; Ueno, T.; Hikage, T.; Erker, G.; Watanabe, Y. *Journal of the …* **2008**, *130* (32), 10512–10514.

(50) Ueno, T.; Abe, M.; Hirata, K.; Abe, S.; Suzuki, M.; Shimizu, N.; Yamamoto, M.; Takata, M.; Watanabe, Y. *J Am Chem Soc* **2009**, *131* (14), 5094–5100.

(51) Abe, S.; Hikage, T.; Watanabe, Y.; Kitagawa, S.; Ueno, T. *Inorg Chem* **2010**, *49* (15), 6967–6973.

(52) Abe, S.; Hirata, K.; Ueno, T.; Morino, K.; Shimizu, N.; Yamamoto, M.; Takata, M.; Yashima, E.; Watanabe, Y. *J Am Chem Soc* **2009**, *131* (20), 6958–6960.

(53) Maity, B.; Fukumori, K.; Abe, S.; Ueno, T. *Chem. Commun.* **2016**, *52* (31), 5463–5466.

(54) He, X. M.; Carter, D. C. *Nature* **1992**, *358* (6383), 209–215.

(55) Bertucci, C.; Botteghi, C.; Giunta, D.; Marchetti, M.; Paganelli, S. *Advanced Synthesis & Catalysis* **2002**, *344* (5), 556–562.

(56) Marchetti, M.; Minello, F.; Paganelli, S.; Piccolo, O. *Applied Catalysis A: General* **2010**, *373* (1-2), 76–80.

(57) Crobu, S.; Marchetti, M.; Sanna, G. *Journal of Inorganic Biochemistry* **2006**, *100* (9), 1514–1520.

(58) Kokubo, T.; Sugimoto, T.; Uchida, T.; Tanimoto, S.; Okano, M. *J. Chem. Soc., Chem. Commun.* **1983**, *0* (14), 769–770.

(59) Köhler, V.; Mao, J.; Heinisch, T.; Pordea, A.; Sardo, A.; Wilson, Y. M.; Knörr, L.; Creus, M.; Prost, J.-C.; Schirmer, T.; Ward, T. R. *Angewandte Chemie International Edition* **2011**, *50* (46), 10863–10866.

(60) Pordea, A.; Creus, M.; Panek, J.; Duboc, C.; Mathis, D.; Novic, M.; Ward, T. R. *Journal of the …* **2008**, *130* (25), 8085–8088.

(61) van de Velde, F.; Arends, I. W. C. E.; Sheldon, R. A. *Topics in Catalysis* **2000**, *13* (3), 259–265.

(62) van de Velde, F.; Könemann, L. *Chem. Commun. (Camb.)* **1998**, No. 17, 1891–1892.

(63) van de Velde, F.; Könemann, L.; van Rantwijk, F.; Sheldon, R. A. *Biotechnology & Bioengineering* **2000**, *67* (1), 87–96.

(64) McNae, I. W.; Fishburne, K.; Habtemariam, A.; Hunter, T. M.; Melchart, M.; Wang, F.; Walkinshaw, M. D.; Sadler, P. J. *Chem. Commun. (Camb.)* **2004**, No. 16, 1786–1787.

(65) Ueno, T.; ABE, S.; YOKOI, N.; WATANABE, Y. *Coordination Chemistry Reviews* **2007**, *251* (21-24), 2717–2731.

(66) Podtetenieff, J.; Taglieber, A.; Bill, E.; Reijerse, E. J.; Reetz, M. T. *Angew. Chem. Int. Ed. Engl.* **2010**, *49* (30), 5151–5155.

(67) Der, B. S.; Edwards, D. R.; Kuhlman, B. *Biochemistry* **2012**, *51* (18), 3933–3940.

(68) Brodin, J. D.; Ambroggio, X. I.; Tang, C.; Parent, K. N.; Baker, T. S.; Tezcan, F. A. *Nat Chem* **2012**, *4* (5), 375–382.

(69) Song, W. J.; Tezcan, F. A. *Science* **2014**, *346* (6216), 1525–1528.

(70) Jeffrey A Sigman; Brian C Kwok, A.; Lu, Y. *From Myoglobin to Heme-Copper Oxidase: Design and Engineering of a CuB Center into Sperm Whale Myoglobin*; American Chemical Society, 2000; Vol. 122, pp 8192–8196.

(71) Wang, N.; Zhao, X.; Lu, Y. *J. Am. Chem. Soc.* **2005**, *127* (47), 16541–16547.

(72) Miner, K. D.; Mukherjee, A.; Gao, Y.-G.; Null, E. L.; Petrik, I. D.; Zhao, X.; Yeung, N.; Robinson, H.; Lu, Y. *Angew. Chem. Int. Ed. Engl.* **2012**, *51* (23), 5589–5592.

(73) Zanghellini, A.; Jiang, L.; Wollacott, A. M.; Cheng, G.; Meiler, J.; Althoff, E. A.; Röthlisberger, D.; Baker, D. *Protein Sci.* **2006**, *15* (12), 2785–2794.

(74) Kuhlman, B.; Baker, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (19), 10383–10388.

(75) Magis, C.; Gasparini, D.; Lecoq, A.; Le Du, M. H.; Stura, E.; Charbonnier, J. B.; Mourier, G.; Boulain, J.-C.; Pardo, L.; Caruana, A.; Joly, A.; Lefranc, M.; Masella, M.; Menez, A.; Cuniasse, P. *J. Am. Chem. Soc.* **2006**, *128* (50), 16190–16205.

(76) Debret, G.; Martel, A.; Cuniasse, P. *Nucl. Acids Res.* **2009**, *37* (Web Server issue), W459–W464.

(77) Gonzalez, G.; Hannigan, B.; DeGrado, W. F. *PLoS Comput Biol* **2014**, *10* (7), e1003750.

(78) Zhou, L.; Bosscher, M.; Zhang, C.; Özçubukçu, S.; Zhang, L.; Zhang, W.; Li, C. J.; Liu, J.; Jensen, M. P.; Lai, L.; He, C. *Nat Chem* **2014**, *6* (3), 236–241.

(79) Fujieda, N.; Schätti, J.; Stuttfeld, E.; Ohkubo, K.; Maier, T.; Fukuzumi, S.; Ward, T. R. *Chemical Science* **2015**, *6* (7), 4060–4065.

(80) Pordea, A. *Curr Opin Chem Biol* **2015**, *25*, 124–132.

(81) Plegaria, J. S.; Pecoraro, V. L. *Methods Mol. Biol.* **2016**, *1414* (Chapter 11), 187–196.

(82) Faiella, M.; Andreozzi, C.; de Rosales, R. T. M.; Pavone, V.; Maglio, O.; Nastri, F.; DeGrado, W. F.; Lombardi, A. *Nat. Chem. Biol.* **2009**, *5* (12), 882–884.

(83) Reig, A. J.; Pires, M. M.; Snyder, R. A.; Wu, Y.; Jo, H.; Kulp, D. W.; Butch, S. E.; Calhoun, J. R.; Szyperski, T. G.; Solomon, E. I.; DeGrado, W. F. *Nat Chem* **2012**, *4* (11), 900–906.

(84) Cangelosi, V. M.; Deb, A.; Penner Hahn, J. E.; Pecoraro, V. L. *Angew. Chem. Int. Ed. Engl.* **2014**, *53* (30), 7900–7903.

(85) Hu, C.; Chan, S. I.; Sawyer, E. B.; Yu, Y.; Wang, J. *Chemical Society Reviews* **2014**, *43* (18), 6498–6510.

(86) Xie, J.; Schultz, P. G. *Methods* **2005**, *36* (3), 227–238.

(87) Xie, J.; Liu, W.; Schultz, P. G. *Angew. Chem. Int. Ed. Engl.* **2007**, *46* (48), 9239–9242.

(88) Drienovská, I.; Rioz-Martínez, A.; Draksharapu, A.; Roelfes, G. *Chemical Science* **2015**, *6* (1), 770–776.

(89) Liu, X.; Yu, Y.; Hu, C.; Zhang, W.; Lu, Y.; Wang, J. *Angew. Chem. Int. Ed. Engl.* **2012**, *51* (18), 4312–4316.

(90) Yu, Y.; Zhou, Q.; Wang, L.; Liu, X.; Zhang, W.; Hu, M.; Dong, J.; Li, J.; Lv, X.; Ouyang, H.; Li, H.; Gao, F.; Gong, W.; Lu, Y.; Wang, J. *Chemical Science* **2015**, *6* (7), 3881–3885.

(91) Yu, Y.; Cui, C.; Liu, X.; Petrik, I. D.; Wang, J.; Lu, Y. *Journal of the American ...* **2015**, *137* (36), 11570–11573.

(92) Reedy, C. J.; Elvekrog, M. M.; Gibney, B. R. *Nucl. Acids Res.* **2008**, *36* (Database issue), D307–D313.

(93) Fronticelli, C.; Bucci, E. *Biochim. Biophys. Acta* **1963**, *78* (3), 530–531.

(94) Ascoli, F.; Fanelli, M. R.; Antonini, E. *Meth. Enzymol.* **1981**, *76*, 72–87.

(95) Fruk, L.; Kuo, C. H.; Torres, E.; Niemeyer, C. M. *Angewandte Chemie International Edition* **2009**, *48* (9), 1550–1574.

(96) Key, H. M.; Dydio, P.; Clark, D. S.; Hartwig, J. F. *Nature* **2016**, *534* (7608), 534–537.

(97) Takafumi Ueno; Tomomi Koshiyama; Masataka Ohashi; Kazuyoshi Kondo; Masaharu Kono; Atsuo Suzuki; Takashi Yamane, A.; Yoshihito Watanabe. *Coordinated Design of Cofactor and Active Site Structures in Development of New Protein Catalysts*; American Chemical Society, 2005; Vol. 127, pp 6556–6562.

(98) Dydio, P.; Key, H. M.; Nazarenko, A.; Rha, J. Y. E.; Seyedkazemi, V.; Clark, D. S.; Hartwig, J. F. *Science* **2016**, *354* (6308), 102–106.

(99) Lelyveld, V. S.; Brustad, E.; Arnold, F. H.; Jasanoff, A. *J Am Chem Soc* **2011**, *133* (4), 649–651.

(100) Ueno, T.; Yokoi, N.; Unno, M.; Matsui, T.; Tokita, Y.; Yamada, M.; Ikeda-Saito, M.; Nakajima, H.; Watanabe, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (25), 9416–9421.

(101) Komatsu, T.; Ishihara, S.; Tsuchida, E.; Nishide, H.; Morokuma, C.; Nakamura, S. *Biomacromolecules* **2005**, *6* (3), 1489–1494.

(102) Ricoux, R.; Dubuc, R.; Dupont, C.; Maréchal, J.-D.; Martin, A.; Martin, A.; Sellier, M.; Sellier, M.; Mahy, J.-P. *Bioconjug. Chem.* **2008**, *19* (4), 899–910.

(103) Marchi-Delapierre, C.; Rondot, L.; Cavazza, C.; Ménage, S. *Israel Journal of Chemistry* **2015**, *55* (1), 61–75.

(104) Allard, M.; Dupont, C.; Muñoz Robles, V.; Doucet, N.; Lledós, A.; Maréchal, J.-D.; Urvoas, A.; Mahy, J.-P.; Ricoux, R. *Chembiochem* **2012**, *13* (2), 240–251.

(105) Ricoux, R.; Allard, M.; Dubuc, R.; Dupont, C.; Maréchal, J.-D.; Mahy, J.-P. *Organic & Biomolecular Chemistry* **2009**, *7* (16), 3208–3211.

(106) Albanese, D. C. M.; Gaggero, N. *RSC Adv.* **2015**, *5* (14), 10588–10598.

(107) Mahammed, A.; Gray, H. B.; Weaver, J. J.; Sorasaenee, K.; Gross, Z. *Bioconjug. Chem.* **2004**, *15* (4), 738–746.

(108) and, A. M.; Gross, Z. *J Am Chem Soc* **2005**, *127* (9), 2883–2887.

(109) Zunszain, P. A.; Ghuman, J.; Komatsu, T.; Tsuchida, E.; Curry, S. *BMC Structural Biology 2003 3:1* **2003**, *3* (1), 1.

(110) Watanabe, K.; Ishikawa, N.; Komatsu, T. *Chemistry – An Asian Journal* **2012**, *7* (11), 2534–2537.

(111) Kato, R.; Kobayashi, Y.; Akiyama, M.; Komatsu, T. *Dalton Trans* **2013**, *42* (45), 15889–15892.

(112) Rousselot-Pailley, P.; Bochot, C.; Marchi-Delapierre, C.; Jorge-Robin, A.; Martin, L.; Fontecilla-Camps, J. C.; Cavazza, C.; Ménage, S. *Chembiochem* **2009**, *10* (3), 545–552.

(113) Tang, J.; Huang, F.; Wei, Y.; Bian, H.; Zhang, W.; Liang, H. *Dalton Trans* **2016**, *45* (19), 8061–8072.

(114) Reetz, M. T.; Jiao, N. *Angew. Chem. Int. Ed. Engl.* **2006**, *45* (15), 2416–2419.

(115) Bos, J.; Browne, W. R.; Driessen, A. J. M.; Roelfes, G. *J Am Chem Soc* **2015**, *137* (31), 9796–9799.

(116) Mickaël V Cherrier; Lydie Martin; Christine Cavazza; Lilian Jacquamet; David Lemaire; Jacques Gaillard, A.; Juan C Fontecilla-Camps. *Crystallographic and Spectroscopic Evidence for High Affinity Binding of FeEDTA(H2O)- to the Periplasmic Nickel Transporter NikA*; American Chemical Society, 2005; Vol. 127, pp 10075–10082.

(117) Esmieu, C.; Cherrier, M. V.; Amara, P.; Girgenti, E.; Marchi-Delapierre, C.; Oddon, F.; Iannello, M.; Jorge-Robin, A.; Cavazza, C.; Ménage, S. *Angew. Chem. Int. Ed. Engl.* **2013**, *52* (14), 3922–3925.

(118) Mahy, J. P.; Maréchal, J. D.; Ricoux, R. *Chem. Commun. (Camb.)* **2015**, *51* (13), 2476–2494.

(119) Iverson, B. L.; Lerner, R. A. *Science* **1989**, *243* (4895), 1184–1188.

(120) Jun Yin; Jeremy H Mills, A.; Schultz, P. G. *A Catalysis-Based Selection for Peroxidase Antibodies with Increased Activity*; American Chemical Society, 2004; Vol. 126, pp 3006–3007.

(121) Ricoux, R.; Lukowska, E.; Pezzotti, F.; Mahy, J.-P. *The FEBS Journal* **2004**, *271* (7), 1277–1283.

(122) Quilez, R.; de Lauzon, S.; Desfosses, B.; Mansuy, D.; Mahy, J.-P. *FEBS Letters* **1996**, *395* (1), 73–76.

(123) Wilson, M. E.; Whitesides, G. M. *J. Am. Chem. Soc.* **1978**, *100* (1), 306–307.

(124) Lin, C.-C.; Lin, C.-W.; Chan, A. S. C. *Tetrahedron: Asymmetry* **1999**, *10* (10), 1887–1893.

(125) Zocchi, A.; Humbert, N.; Berta, T.; Ward, T. R. *CHIMIA International Journal for Chemistry* **2003**, *57* (10), 589–592.

(126) Jérôme Collot; Julieta Gradinaru; Nicolas Humbert; Myriem Skander; Andrea Zocchi, A.; Ward, T. R. *Artificial Metalloenzymes for Enantioselective Catalysis Based on Biotin−Avidin*; American Chemical Society, 2003; Vol. 125, pp 9030–9031.

(127) Pordea, A.; Creus, M.; Letondor, C.; Ivanova, A.; Ward, T. R. *Inorganica Chimica Acta* **2010**, *363* (3), 601–604.

(128) Kajetanowicz, A.; Chatterjee, A.; Reuter, R.; Ward, T. R. *Catal Lett* **2014**, *144* (3), 373–379.

(129) Chatterjee, A.; Mallin, H.; Klehr, J.; Vallapurackal, J.; Finke, A. D.; Vera, L.; Marsh, M.; Ward, T. R. *Chemical Science* **2016**, *7* (1), 673–677.

(130) Heinisch, T.; Ward, T. R. *European Journal of Inorganic …* **2015**.

(131) Jeschek, M.; Reuter, R.; Heinisch, T.; Trindler, C.; Klehr, J.; Panke, S.; Ward, T. R. *Nature* **2016**, *537* (7622), 661–665.

(132) Panek, J. J.; Ward, T. R.; Jezierska-Mazzarello, A. *… -aided molecular design* **2010**.

(133) Zimbron, J. M.; Heinisch, T.; Schmid, M.; Hamels, D.; Nogueira, E. S.; Schirmer, T.; Ward, T. R. *J Am Chem Soc* **2013**, *135* (14), 5384–5388.

(134) Monnard, F. W.; Heinisch, T.; Nogueira, E. S.; Schirmer, T.; Ward, T. R. *Chemical Communications* **2011**, *47* (29), 8238–8240.

(135) Monnard, F. W.; Nogueira, E. S.; Heinisch, T.; Schirmer, T.; Ward, T. R. *Chemical Science* **2013**, *4* (8), 3269–3274.

(136) Heinisch, T.; Pellizzoni, M.; Dürrenberger, M.; Tinberg, C. E.; Köhler, V.; Klehr, J.; Häussinger, D.; Baker, D.; Ward, T. R. *Journal of the …* **2015**, *137* (32), 10414–10419.

(137) Zhao, J.; Kajetanowicz, A.; Ward, T. R. *Organic & Biomolecular Chemistry* **2015**, *13* (20), 5652–5655.

(138) Raffy, Q.; Ricoux, R.; Sansiaume, E.; Pethe, S.; Mahy, J.-P. *Journal of Molecular Catalysis A: Chemical* **2010**, *317* (1-2), 19–26.

(139) Sansiaume, E.; Ricoux, R.; Gori, D.; Mahy, J.-P. *Tetrahedron: Asymmetry* **2010**, *21* (11-12), 1593–1600.

(140) Ghattas, W.; Cotchico Alonso, L.; Maréchal, J.-D.; Urvoas, A.; Rousseau, M.; Mahy, J.-P.; Ricoux, R. *Chembiochem* **2016**, *17* (5), 433–440.

(141) Sansiaume-Dagousset, E.; Urvoas, A.; Chelly, K.; Ghattas, W.; Maréchal, J.-D.; Mahy, J.-P.; Ricoux, R. *Dalton Transactions* **2014**, *43* (22), 8344–8354.

(142) Rondot, L.; Girgenti, E.; Oddon, F.; Marchi-Delapierre, C.; Jorge-Robin, A.; Ménage, S. *Journal of Molecular Catalysis A: Chemical* **2016**, *416*, 20–28.

(143) Kuo, C. H.; M Niemeyer, C.; Fruk, L. *Croatica Chemica Acta* **2011**, *84* (2), 269–275.

(144) Buron, C.; Sénéchal David, K.; Ricoux, R.; Le Caër, J. P.; Guérineau, V.; Méjanelle, P.; Guillot, R.; Herrero, C.; Mahy, J.-P.; Banse, F. *Chem. Eur. J.* **2015**, *21* (34), 12188–12193.

(145) Yang, H.; Srivastava, P.; Zhang, C.; Lewis, J. C. *Chembiochem* **2014**, *15* (2), 223–227.

(146) Kidd, D.; Liu, Y.; Cravatt, B. F. *Biochemistry* **2001**, *40* (13), 4005–4015.

(147) Buller, A. R.; Townsend, C. A. *Proceedings of the National Academy of Sciences* **2013**, *110* (8), E653–E661.

(148) Powers, J. C.; Asgian, J. L.; Ekici, O. D.; James, K. E. *Chem. Rev.* **2002**, *102* (12), 4639–4750.

(149) Hyster, T. K.; Ward, T. R. *Angew. Chem. Int. Ed. Engl.* **2016**, *55* (26), 7344–7357.

(150) Levine, H. L.; Nakagawa, Y.; Kaiser, E. T. *Biochemical and Biophysical Research Communications* **1977**, *76* (1), 64–70.

(151) Panella, L.; Broos, J.; Jin, J.; Fraaije, M. W.; Janssen, D. B.; Jeronimus-Stratingh, M.; Ben L Feringa; Minnaard, A. J.; de Vries, J. G. *Chemical Communications* **2005**, *0* (45), 5656–5658.

(152) Reetz, M. T.; Rentzsch, M.; Pletsch, A.; Maywald, M. *CHIMIA International ...* **2002**, *56* (12), 721–723.

(153) Reetz, M. T. *Tetrahedron* **2002**, *58* (32), 6595–6602.

(154) Talbi, B.; Haquette, P.; Martel, A.; de Montigny, F.; Fosse, C.; Cordier, S.; Roisnel, T.; Jaouen, G.; Salmain, M. *Dalton Trans* **2010**, *39* (24), 5605–5607.

(155) Haquette, P.; Talbi, B.; Barilleau, L.; Madern, N.; Fosse, C.; Salmain, M. *Organic & Biomolecular Chemistry* **2011**, *9* (16), 5720–5727.

(156) Madern, N.; Talbi, B.; Salmain, M. *Applied Organometallic Chemistry* **2012**, *27* (1), 6–12.

(157) Madern, N.; Queyriaux, N.; Chevalley, A.; Ghasemi, M.; Nicolotti, O.; Ciofini, I.; Mangiatordi, G. F.; Salmain, M. *Journal of Molecular Catalysis B: Enzymatic* **2015**, *122*, 314–322.

(158) Reiner, T.; Jantke, D.; Marziale, A. N.; Raba, A.; Eppinger, J. *ChemistryOpen* **2013**, *2* (2), 50–54.

(159) James R Carey; Steven K Ma; Thomas D Pfister; Dewain K Garner; Hyeon K Kim; Joseph A Abramite; Zhilin Wang; Zijian Guo, A.; Yi Lu. *A Site-Selective Dual Anchoring Strategy for Artificial Metalloprotein Design*; American Chemical Society, 2004; Vol. 126, pp 10812–10813.

(160) Kruithof, C. A.; Casado, M. A.; Guillena, G.; Egmond, M. R.; van der Kerk-van Hoof, A.; Heck, A. J. R.; Klein Gebbink, R. J. M.; van Koten, G. *Chem. Eur. J.* **2005**, *11* (23), 6869–6877.

(161) Kruithof, C. A.; Dijkstra, H. P.; Lutz, M.; Spek, A. L.; Egmond, M. R.; Klein Gebbink, R. J. M.; van Koten, G. *Eur. J. Inorg. Chem.* **2008**, *2008* (28), 4425–4432.

(162) Basauri Molina, M.; Verhoeven, D. G. A.; van Schaik, A. J.; Kleijn, H.; Klein Gebbink, R. J. M. *Chem. Eur. J.* **2015**, *21* (44), 15676–15685.

(163) Basauri-Molina, M.; Riemersma, C. F.; Würdemann, M. A.; Kleijn, H.; Gebbink, R. J. M. K. *Chemical Communications* **2015**, *51* (31), 6792–6795.

(164) Matsuo, T.; Imai, C.; Yoshida, T.; Saito, T.; Hayashi, T.; Hirota, S. *Chemical Communications* **2012**, *48* (11), 1662–1664.

(165) In *Bioconjugate Techniques (Third edition)*; Academic Press: Boston, 2013; pp i–iii.

(166) Dongfeng Qi; Cheng-Min Tann; Dietmar Haring, A.; Distefano, M. D. *Generation of New Enzymes via Covalent Modification of Existing Proteins*;

American Chemical Society, 2001; Vol. 101, pp 3081–3112.

(167) Ishihama, A. *Chemical Communications* **2000**, *0* (13), 1091–1094.

(168) van Dongen, S. F. M.; Clerx, J.; Nørgaard, K.; Bloemberg, T. G.; Cornelissen, J. J. L. M.; Trakselis, M. A.; Nelson, S. W.; Benkovic, S. J.; Rowan, A. E.; Nolte, R. J. M. *Nat Chem* **2013**, *5* (11), 945–951.

(169) and, R. R. D.; Distefano, M. D. *A Semisynthetic Metalloenzyme Based on a Protein Cavity That Catalyzes the Enantioselective Hydrolysis of Ester and Amide Substrates*; American Chemical Society, 1997; Vol. 119, pp 11643–11652.

(170) Davies, R. R.; Kuang, H.; Qi, D.; Mazhary, A.; Mayaan, E.; Distefano, M. D. *Bioorg. Med. Chem. Lett.* **1999**, *9* (1), 79–84.

(171) Zhang, C.; Srivastava, P.; Ellis-Guardiola, K.; Lewis, J. C. *Tetrahedron* **2014**, *70* (27-28), 4245–4249.

(172) Reetz, M. T.; Rentzsch, M.; Pletsch, A.; Taglieber, A.; Hollmann, F.; Mondière, R. J. G.; Dickmann, N.; Höcker, B.; Cerrone, S.; Haeger, M. C.; Sterner, R. *Chembiochem* **2008**, *9* (4), 552–564.

(173) Sauer, D. F.; Himiyama, T.; Tachikawa, K.; Fukumoto, K.; Onoda, A.; Mizohata, E.; Inoue, T.; Bocola, M.; Schwaneberg, U.; Hayashi, T.; Okuda, J. *ACS Catal.* **2015**, *5* (12), 7519–7522.

(174) Onoda, A.; Fukumoto, K.; Arlt, M.; Bocola, M.; Schwaneberg, U.; Hayashi, T. *Chemical Communications* **2012**, *48* (78), 9756–9758.

(175) Sletten, E. M.; Bertozzi, C. R. *Angew. Chem. Int. Ed. Engl.* **2009**, *48* (38), 6974–6998.

(176) Srivastava, P.; Yang, H.; Ellis-Guardiola, K.; Lewis, J. C. *Nat Commun* **2015**, *6*, 7789.

(177) Gu, Y.; Ellis-Guardiola, K.; Srivastava, P.; Lewis, J. C. *Chembiochem* **2015**, *16* (13), 1880–1883.

(178) Renggli, K.; Nussbaumer, M. G.; Urbani, R.; Pfohl, T.; Bruns, N. *Angewandte Chemie International Edition* **2014**, *53* (5), 1443–1447.

CHAPTER TWO

SERINE HYDROLASE MECHANISM-BASED ARTIFICIAL METALLOENZYMES

2.1 INTRODUCTION

2.1.1 SERINE HYDROLASES

As discussed in section 1.3.2.3.1, serine hydrolases constitute a viable class of scaffold for the covalent bioconjugation of metal cofactors linked to electrophilic anchoring groups.[1,2] Serine hydrolases comprise ~1% of the human proteome, consisting of approximately 200 individual protein species.[3,4] The structural diversity of this protein superfamily is sufficiently large to enable an enormous number of artificial metalloenzyme (ArM) permutations. We sought to employ the unique reactivity of the serine hydrolases to enable the rapid preparation and evaluation of a variety of novel ArMs.

The unique reactivity of the serine hydrolases arises from their "catalytic triad," a network of serine-histidine-(aspartate/glutamate) that works through a proton relay mechanism, nucleophilically activating this residue (Figure 2.1).[5]

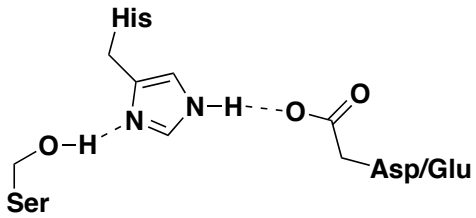


Figure 2.1 The catalytic triad of serine hydrolases. The "proton relay" is responsible for partially deprotonating serine, rendering it nucleophilic.

The nucleophilic serine is necessary to carry out native amide/ester hydrolysis catalyzed by these enzymes.[5] This nucleophilic mechanism can be exploited by electrophilic phosphates and phosphonates, which act as effective irreversible covalent inhibitors due to their mimicry of the tetrahedral intermediate generated during amide/ester hydrolysis (Figure 2.2).[6]

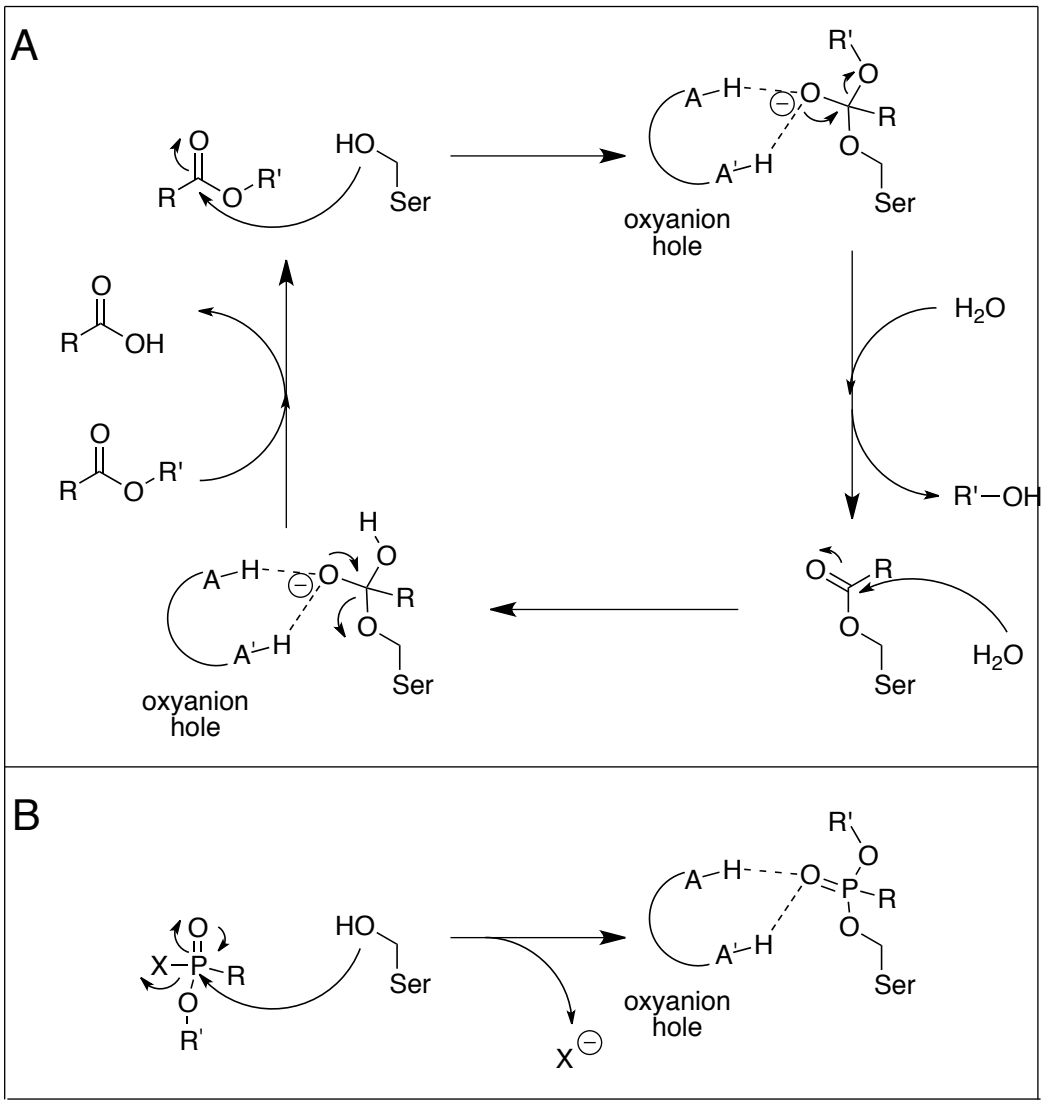


Figure 2.2 A) The catalytic mechanism of serine hydrolases in ester hydrolysis. The tetrahedral intermediate formed after nucleophilic attack of serine is stabilized by the

oxyanion hole. B) The inhibition mechanism of serine hydrolases. Phosphates and phosphonates mimic the tetrahedral intermediate formed during catalysis.

The mechanism of this inhibition reaction involves the nucleophilic substitution of the phosphate/phosphonate ester to generate the serine-substituted covalent adduct shown. This species is complementary to the site occupied by the tetrahedral intermediate during native catalysis. The "oxyanion hole" of serine hydrolases provides a set of hydrogen bonds for the stabilization of this covalent phosphate/phosphonate adduct.[6]

2.1.2 PHOSPHATE AND PHOSPHONATE INHIBITORS

Owing to the utility of electrophilic phosphates/phosphonates as covalent inhibitors, we pursued these molecules to create a set of covalent ArMs with serine hydrolase target scaffolds. This strategy affords a number of advantages over other bioconjugation methodologies. First, because the bioconjugation exploits the native mechanism of enzymes, no protein engineering is required for the covalent linkage of metal cofactors to these scaffolds. This enables the rapid evaluation of new scaffolds, as one need only obtain the native form of the enzyme to achieve bioconjugation. Furthermore, the anchoring of these cofactors proceeds selectively at the active site serine, reducing complications arising from multiple addition, a known problem for other bioconjugation methods.[7] This selectivity for the active site serine also confers the advantage of guiding the metal cofactor into the active site, often characterized by a cleft or cavity. Such a topology is desirable for the design of ArMs, as this alters the secondary coordination sphere about the metal. Finally, as a minor point, the complementarity of

phosphate/phosphonate cofactors for the catalytic site of serine hydrolases could potentially rigidify the cofactor anchor, a desirable feature in the design of ArMs.[8]

2.2 RESULTS AND DISCUSSION

2.2.1 ARTIFICIAL METALLOENZYME OXYGENATION CATALYSIS

C-H oxygenation is a highly important reaction in C-H functionalization.[9] We chose this as a desirable starting point in the design of new ArMs due to the strong precedent for enzymatic C-H oxygenation catalysis provided by cytochromes p450,[10] providing a clear framework for the compatibility of this reactivity with protein catalysts. In addition, as will be discussed in further detail, a number of groups have generated ArMs for oxygenation of alkenes (epoxidation) and sulfides (sulfoxidation), both of which can proceed through similar mechanisms to those driving C-H oxygenation catalysis.[11,12]

2.2.1.1 TERPYRIDINE-BASED ARTIFICIAL METALLOENZYMES

One of the first types of ArM explored in the Lewis group centered around the oxidative chemistry of first-row metal terpyridine complexes, with particular focus on the reactivity of the first-row metals manganese and iron. Several systems have employed the terpyridine ligand framework to enable oxidative chemistry, including water oxidation,[13] epoxidation,[14] and benzylic C-H oxygenation.[15] Crabtree and Brudvig were able append molecular recognition elements into small-molecule Mn(μ-O)terpyridine to effect the selective oxidation of alkenes and benzylic C-H bonds (Figure 2.3).[16,17] This work served as a guiding model for the explorations into manganese-terpyridine oxidations described below since it confirmed the potential for secondary coordination sphere effects to promote selective catalysis with these systems.

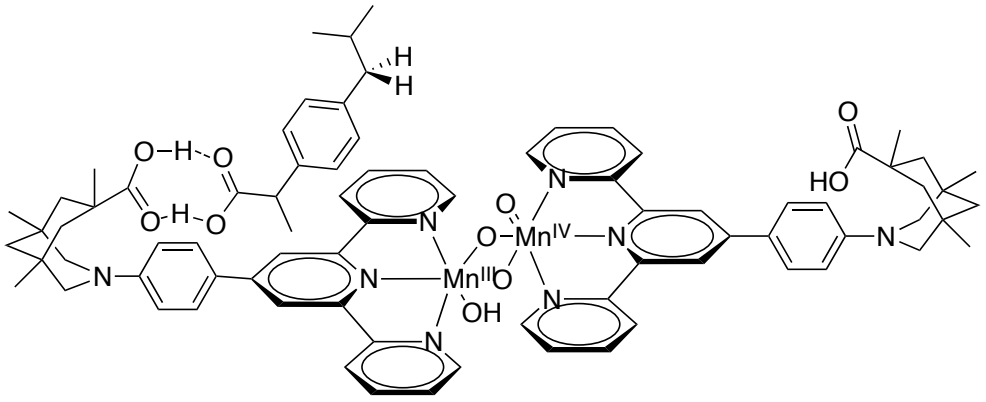


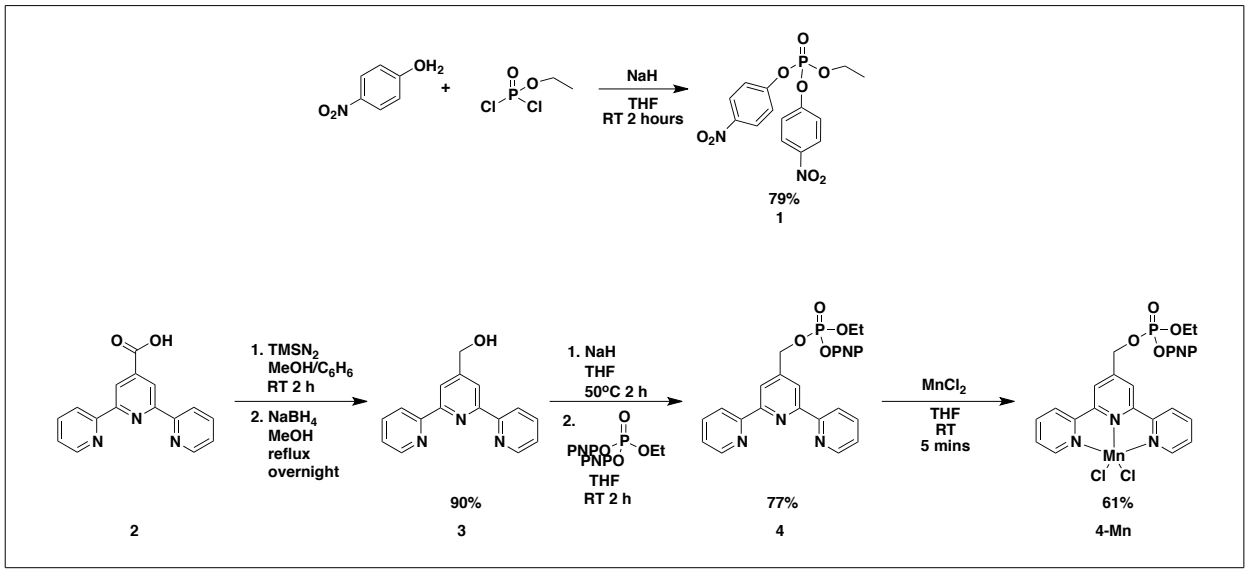
Figure 2.3 The active catalyst in ref. 16 with the putative hydrogen bond-mediated molecular recognition of substrate ibuprofen.

2.2.1.1.1 TERPYRIDINE PHOSPHATE ARTIFICIAL METALLOENZYMES

The first set of Mn-terpyridine-based ArMs explored were accessed on the basis of their practicality of synthesis and their ready introduction into serine hydrolase scaffolds, a number of which could be purchased in highly pure form from commercial sources. These employed phosphate-conjugated terpyridine ligand **4**. Their relatively simple preparation enabled the rapid evaluation of conditions to promote ArM catalysis, such as aqueous buffer choice, temperature, substrate/catalyst loading, and choice of oxidant.

A general route for accessing electrophilic phosphate functionality was developed through the use of electrophilic phosphate precursor **1**, which is accessed through the nucleophilic attack of *p*-nitrophenol on dichloroethylphosphate (Scheme 2.1). The benzylic alcohol **3** was readily generated through the two-step alkylation and reduction of terpyridine carboxylate precursor **2**. Nucleophilic attack of the benzylic alcohol **3** on **1** enabled the generation of phosphate cofactor **4**, which could then be metalated with a Mn(II) and Fe(II) salts to generate the corresponding metal complexes **4-Mn** and **4-Fe**.

One of the key advantages of using a *p*-nitrophenol leaving group with these electrophilic phosphate cofactors stems from the bright yellow color of its conjugate base, *p*-nitrophenolate. At basic pH, the extent of bioconjugation of the cofactor can be followed by monitoring the increase in UV/Vis absorbance at 400 nm.



Scheme 2.1 Synthesis of **4** and metalation to yield **4-Mn**

Using this method, cofactor **4** and its metalated derivatives **4-Mn** and **4-Fe** were determined to be competent covalent inhibitors of serine hydrolases. The bulk of early ArM work using this system was carried out with bovine α-chymotrypsin, a commercially-available serine protease that is primarily used in the digestion of proteins for proteomics applications. The UV/Vis trace of bioconjugation of chymotrypsin with **4** is shown in Figure 2.5. The conditions for bioconjugation with chymotrypsin (5% MeCN/100 mM Tris HCl pH 8.0 with 5:1 cofactor:protein) proved to be quite general for the bioconjugation of a number of other protein scaffolds (see below). In addition, removal of excess **4** was simple, requiring one step of gel filtration followed by

centrifugal membrane buffer exchange. The resulting chymotrypsin hybrids were simple to characterize by high-resolution ESI-MS, as shown in Figure 2.4.

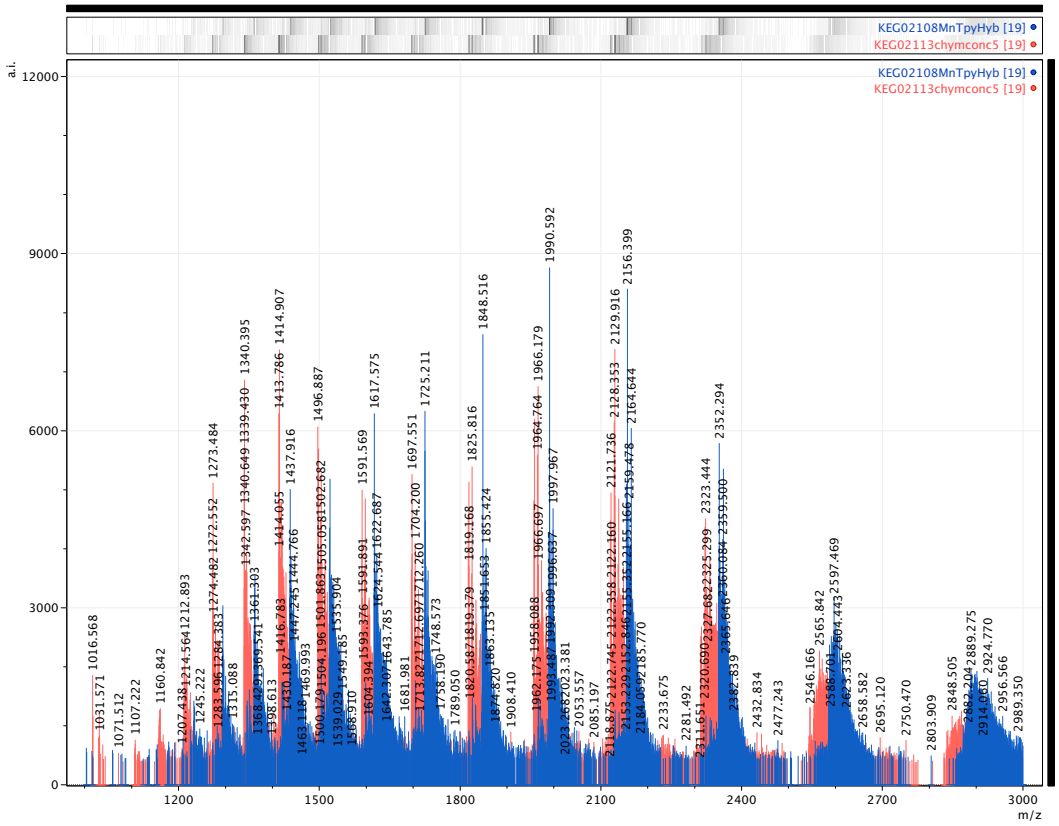


Figure 2.4 ESI-MS raw spectrum with the charge envelope of chymotrypsin (red) and its corresponding **4** hybrid (blue).

ESI-MS enabled the ready confirmation of extent of bioconjugation in the product hybrids. The extent of metalation could not be evaluated, however, as the metal was never observed in the mass adducts, in contrast to recent reports.[18] Therefore, alternative characterization approaches were explored to confirm the metalation of chymotrypsin hybrids. Initial attempts at characterization of the bound paramagnetic Mn(II) center by electron paramagnetic resonance (EPR) spectroscopy failed due to operational

inexperience with the method. A simpler approach was taken by employing UV/Vis spectroscopy, which enabled the observation of new absorbance features upon *in situ* metalation of chymotrypsin-**4** hybrids with iron(II) or manganese(II) (Figure 2.6). Indeed, *in situ* metalation of the chymotrypsin-**4** hybrid could be monitored by UV/Vis (Figure 2.6), and titration of metal salts into solutions of this hybrid enabled quantitative confirmation that the terpyridine ligand was being saturated by stoichiometric or low excesses of $MnCl_2$ and $FeCl_2$.

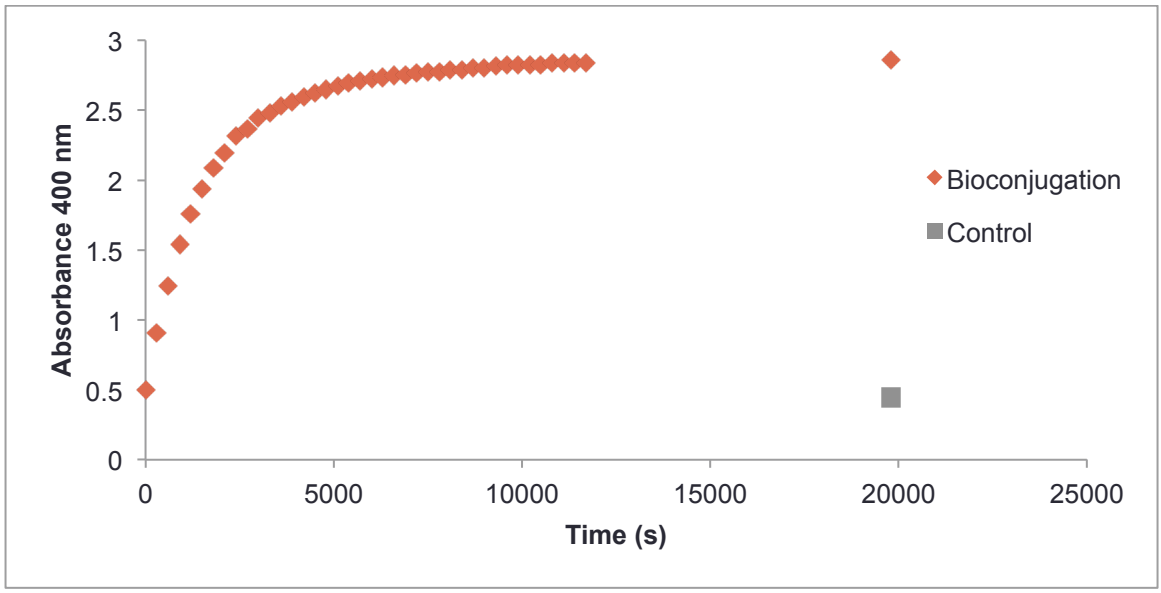


Figure 2.5 Monitoring the bioconjugation of chymotrypsin with **4** by release of *p*-nitrophenol. 5:1 **4**:chymotrypsin in 5% MeCN/100 mM Tris pH 8.0.

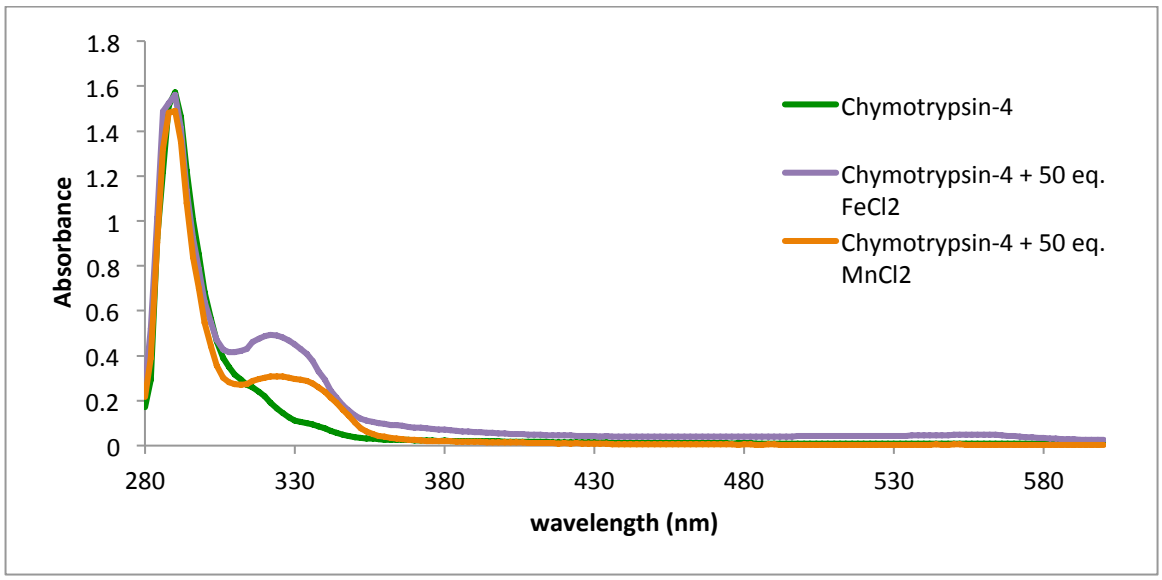


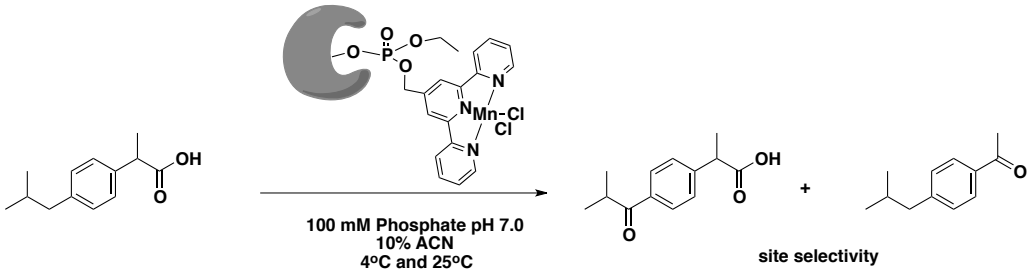
Figure 2.6 Observing metalation with chymotrypsin-4 by UV/Vis spectroscopy.

As a commercial lyophilized powder, chymotrypsin proved to be an operationally convenient, robust, and inexpensive scaffold for the development of oxidative reactivity in these ArM systems. Indeed, a number of optimizations were carried out with chymotrypsin-**4** metalated hybrids to enable ArM turnover in oxygenation catalysis. Initial investigations of reactivity identified the peroxyacid Oxone ($KHSO_5$) as a viable terminal oxidant to enable Mn-terpyridine-catalyzed benzylic oxidation and epoxidation, as was shown previously in the literature.[17,19] Large degrees of uncatalyzed background reaction were observed using Oxone, however, particularly in epoxidation. In addition, this oxidant led to rapid precipitation of protein due to its high acidity, necessitating relatively strong buffers to maintain pH for chymotrypsin stability. Shifting to peracetic acid enabled benzylic oxidation and epoxidation with minimal background and with a far less pronounced drop in medium pH. Phosphate buffer was found to promote higher yields of oxidation products. With these optimized conditions, we used two main test

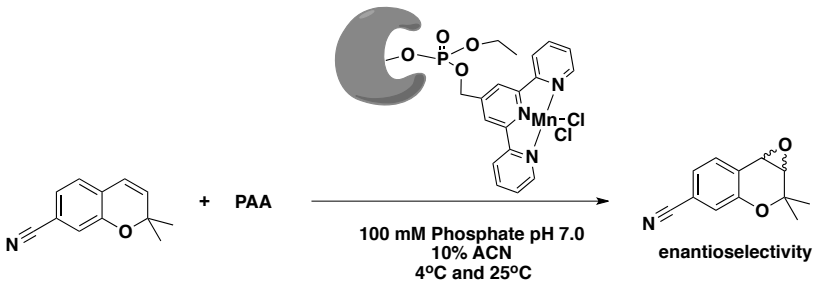
reactions to assess for selectivity: 1) C-H benzylic oxidation of ibuprofen, which bears two benzylic sites for oxidation to assess site selectivity (Scheme 2.2 A), and 2) epoxidation of 2,2-dimethyl-7-cyanobenzopyran, which is relatively bulky, potentially facilitating relay of stereochemical information from the scaffold and affording enantio-enriched epoxide products (Scheme 2.2 B).

Scheme 2.2 The two test reactions used to assess selectivity in terpyridine ArM-catalyzed oxidations

A) Benzylic oxidation of ibuprofen



B) Epoxidation of 2,2-dimethyl-7-cyanobenzopyran



With these optimized conditions and reactions in hand, the selectivity of oxidation was assessed. Unfortunately, chymotrypsin-**5** hybrids did not afford any notable selectivity in the epoxidation or in benzylic C-H oxidation. Reducing the reaction temperature to 4°C had little effect on improving the selectivity of the reactions.

Furthermore, altering the terminal oxidant also had little effect on the observed enantioselectivity or site selectivity, although yields were impacted significantly.

The lack of selectivity observed in the chymotrypsin-**4** based systems described proved to be puzzling. This was initially thought to be a symptom of poor cofactor encapsulation within the chymotrypsin protein scaffold. Indeed, the active site of chymotrypsin is relatively shallow and solvent exposed, as is common among proteases.[20] With this in mind, we explored a number of other viable serine hydrolase scaffolds for hosting the cofactor. Given the similar practicality of trypsin as a ubiquitous commercially available enzyme, bioconjugations of this enzyme were carried out, although ESI-MS characterization of the resulting hybrid was complicated by the heterogeneity of trypsin sources. In addition, trypsin suffers from a similarly shallow active site, decreasing its viability as an ArM scaffold. Esterases from pig liver and rabbit liver exhibited desirable deep pockets and were commercially available. Both enzymes were bioconjugated with **4-Mn**, but neither ArM exhibited selectivity in epoxidation. Finally, acetylcholinesterase from *Electrophorus electricus*, which has a very deep pocket and is commercially available, was explored as a potential scaffold ArM scaffold for **4-Mn**. Bioconjugation was observed by UV/Vis, but confirmation of bioconjugation by ESI-MS gave no signal. This is likely due to the high molecular weight of the enzyme, which forms a complex tetramer comprised of two disulfide-linked dimers.[21] Regardless, no selectivity in benzylic C-H oxidation of ibuprofen or epoxidation of multiple alkene derivatives was observed using this scaffold.

In a shift away from commercial scaffolds, we also looked to the PDB to identify serine hydrolases with desirable structural features. Using a list of serine hydrolases

provided from an activity-based protein profiling study by Cravatt,[5] a number of other potential scaffolds were identified. The criteria applied for the identification of these alternative scaffolds were driven by the availability of structural information to assess the depth of the active site serine in the protein scaffold as well as a practical consideration—had these scaffolds been expressed heterologously in *E. coli*? This led to the identification of four candidate serine hydrolase scaffolds: *Escherichia coli* esterase BioH,[22] *Rhodococcus* sp. strain MB1 cocaine esterase CocE,[23] *Rhodococcus* sp. strain H1 heroin esterase HerE,[24] and *Pyrococcus furiosus* prolyl oligopeptidase.[25] BioH was successfully expressed, purified, and bioconjugated with **4-Mn** (Figure 2.7). When exposed to standard oxidative reaction conditions, however, rapid precipitation of the BioH ArM was observed. *Pyrococcus furiosus* prolyl oligopeptidase (*Pfu* POP), in contrast, was readily expressed in *E. coli*, purified, bioconjugated with **4-Mn** and successfully applied for benzylic C-H oxidation, epoxidation, and oxidative dimethylation of a variety of substrates. Despite the lack of a crystal structure for this enzyme, a homology model existed, which granted assurance that the active site serine of *Pfu* POP sits at the base of a very deep binding pocket to guarantee cofactor encapsulation.[25] In spite of these promising scaffold properties, oxidation catalysis with the *Pfu* POP-**5a** hybrid failed to demonstrate any significant induction of selectivity.

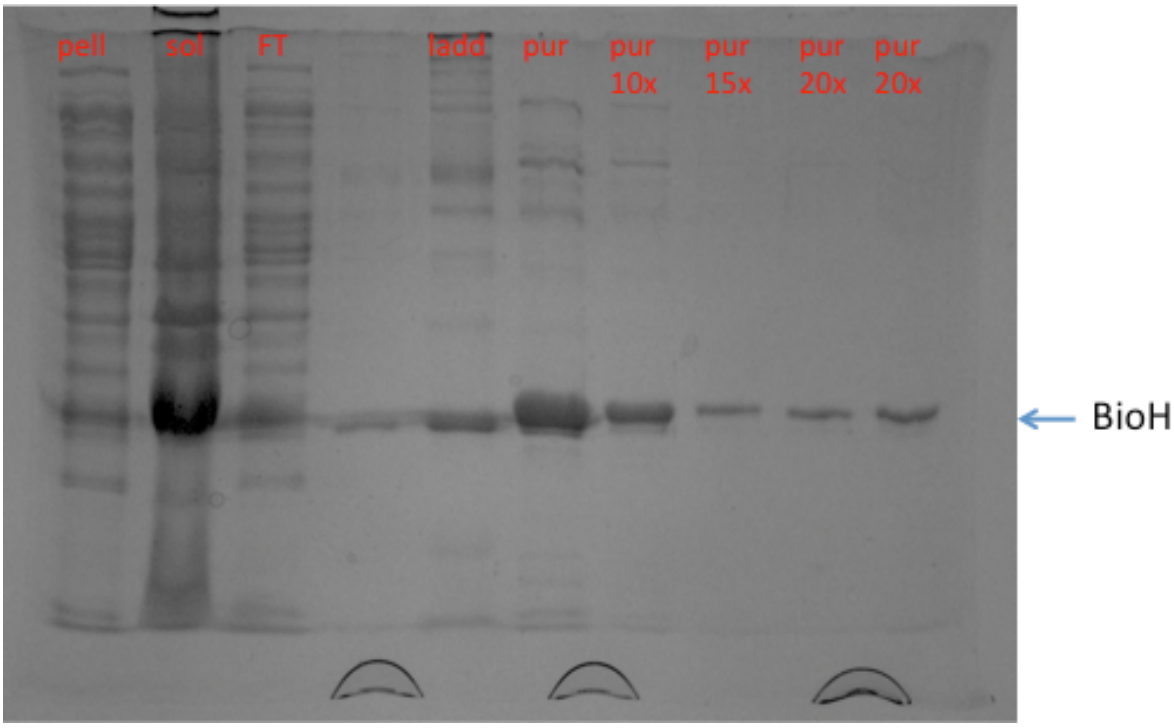
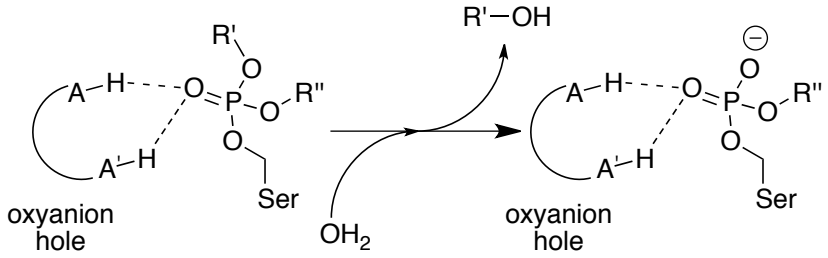


Figure 2.7 BioH expression and purification. pell = pellet, sol = soluble, FT = flowthrough,ladd = ladder, pur = pure NiNTA fraction and dilutions

A key deficiency in the terpyridine-phosphate system was identified and explored as a potential culprit for this lack of selectivity. Gradual hydrolysis of the phosphate moiety to release the ligand/metal fragment could promote low selectivity in catalysis, as this free fragment could catalyze non-selective background oxidation reactions. Such a flaw, known as "aging," is known to occur with organophosphate inhibitors of serine hydrolases (Scheme 2.3).[26]



Scheme 2.3 The process of "aging," wherein one of the alkoxy groups on the phosphate adduct is hydrolased from the enzyme-bound phosphate.

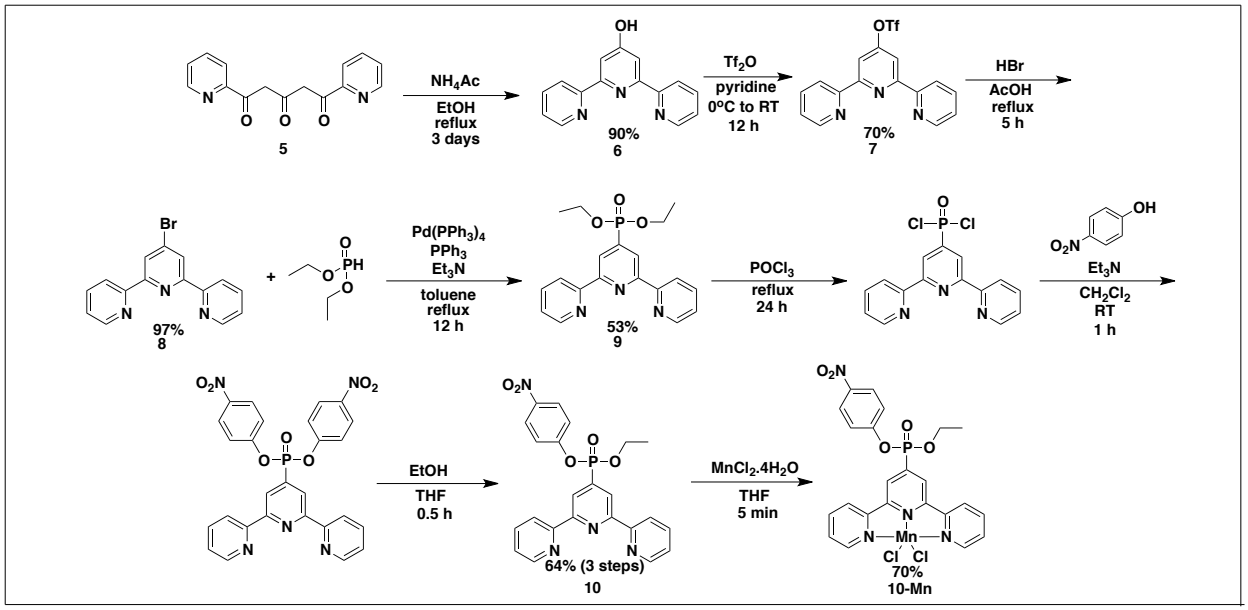
This mechanism was demonstrated to be potentially operative in enabling background reaction. The supernatant of an organic solvent-precipitated solution of *Pfu* POP-**4-Mn** was assessed for epoxidation activity under standard conditions to test for the reactivity of free cofactor fragment **3-Mn.** This showed some reactivity in epoxidation, though it was significantly lower than the ArM control. Between this result and prior reports of phosphate adduct lability, the terpyridine phosphate linkage was abandoned for more robust cofactor/anchor linking strategies.

2.2.1.1.2 TERPYRIDINE PHOSPHONATE ARTIFICIAL METALLOENZYMES

In light of the lability issues observed with phosphate-linked terpyridine cofactors, we sought to develop a cofactor with a more robust linkage between the electrophilic phosphorus moiety and the ligand/metal complex. For this, a C-P bond linkage was pursued, as has been demonstrated in the work of van Koten[27] and Gebbink.[28] van Koten confirmed the robustness of his C-P phosphonate lipase/pincer hybrids over more than two weeks of incubation at room temperature.[29] These observations were later recapitulated in this laboratory (Figure 2.11, see below). The first effort focused on the generation of aryl phosphonate **10**. This cofactor was an appealing

choice to address the lack of selectivity observed with phosphate cofactor **4-Mn**, since, aside from introducing a non-labile C-P bond, it shortened the link between the phosphonate center and the terpyridine moiety by two rotatable bonds. The hypothesis was that this might constrain the cofactor into a rigid position within the host protein scaffold, enabling enhanced selectivity.[30]

First, the hydroxyl terpyridine **6** was formed by the Hantzch-like pyridine condensation of triketone **5**. **6** was converted to terpyridine triflate **7** through treatment with triflic anhydride in pyridine solvent. Reflux of triflate **7** in HBr/AcOH yielded the aryl bromide **8,** which bears a heteroaryl bromide handle for Pd-catalyzed Hirao coupling to install the critical C-P bond in **9**. Chlorination of the diethylphosphonate **9** proceeded via treatment with $POCl_3$ to yield a dichlorinated intermediate, which was isolated as a crude mixture. Base-mediated nucleophilic attack of the dichlorophosphonate with *p*-nitrophenol yielded a highly unstable di-*p*-nitrophenolate phosphonate ester, which was isolated as a crude mixture. Removal of base *in vacuo* and addition of dry EtOH in THF gave partial solvolysis to yield the desired mono-ethyl mono-*p*-nitrophenyl phosphonate ester **10**. This product was used directly for bioconjugation/*in situ* metalation in *Pfu* POP or could be carried on for metalation with $MnCl_2$ to make catalytic cofactor **10-Mn**. (Scheme 2.4)



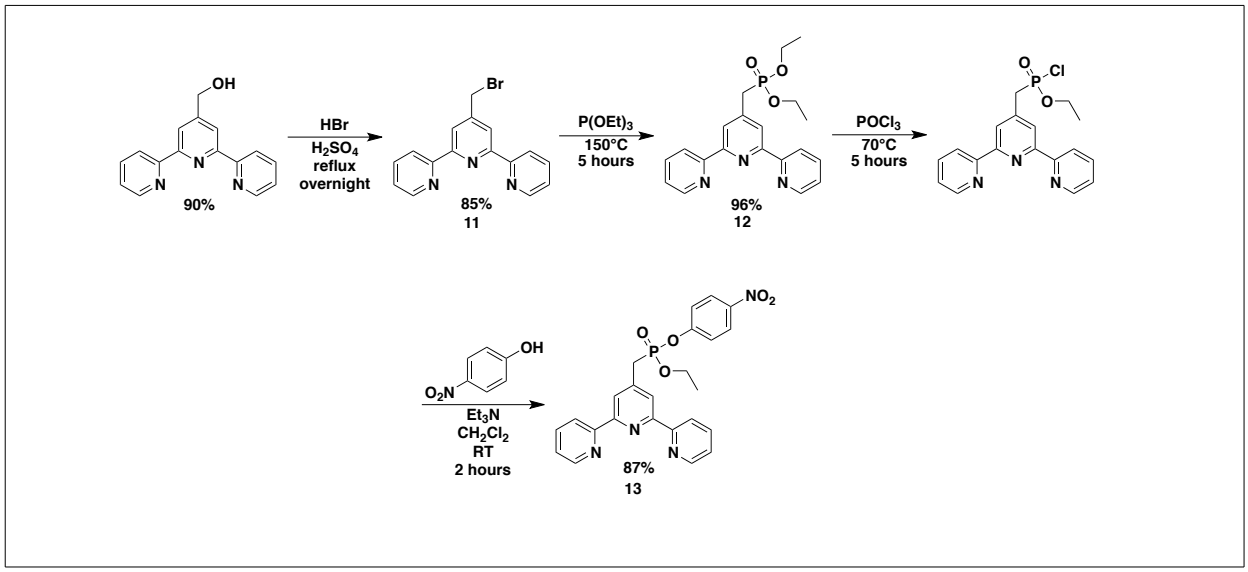
Scheme 2.4 The synthesis of cofactor **10** and metalated derivative **10-Mn**

Bioconjugation of chymotrypsin with **10** required a larger excess (10 equivalents) relative to that necessitated by **4** and its metalated derivatives (5 eq). This was thought to arise due to the greater electrophilicity of this phosphorus center relative to that of cofactor **4**, given the direct link of the phosphorus atom to an electron-poor heterocycle and the absence of the resonance-donating oxygen present in the corresponding phosphate.[31] This enhanced electrophilicity leads to competing hydrolysis of the cofactor by the aqueous buffer medium. In spite of this limitation, full bioconjugation of chymotrypsin with **10** was observed after 5 hours under standard conditions. Attempts at chymotrypsin bioconjugation with metalated cofactor **10-Mn** under identical conditions showed that only ~40% bioconjugation could be achieved after 5 hours of bioconjugation (ESI-MS), with all excess metalated cofactor undergoing hydrolysis in that time (based on UV/Vis traces monitoring release of *p*-nitrophenol). The coordinated metal is thought to enhance the electrophilicity of the phosphorus, leading to enhanced sensitivity toward hydrolysis.

The bioconjugation of *Pfu* POP with cofactors **10** and **10-Mn** is significantly more complicated due to the activity of this enzyme. *Pfu* POP, a hyperthermophilic enzyme with an activity optimum temperature of 85°C, reacts significantly more slowly than chymotrypsin at room temperature.[32] After 5 hours of bioconjugation with 10 equivalents of **10** at room temperature, only 20% conversion to hybrid could be observed by ESI-MS. Increasing the temperature of bioconjugation to 75°C led to 40% bioconjugation in 5 hours, with clear buffer hydrolysis of all remaining cofactor. At similarly elevated temperatures, **10-Mn** yields <5% bioconjugation, indicating the enhanced sensitivity of the metalated cofactor, likely due to induced electrophilicity by the coordinated metal. This removed the possibility of using pre-metalated **10-Mn** for bioconjugation with *Pfu* POP. Only by increasing the stoichiometry of the bioconjugation to 50 equivalents of **10** and using elevated temperatures was full conversion to the corresponding *Pfu* POP hybrid observed. This could then be used for *in situ* metalation to generate *Pfu* POP-**10-Mn**.

To overcome the challenges posed by the hydrolytic propensity of cofactors **10** and **10-Mn**, we sought to extend the terpyridine-phosphonate linkage by one methylene, which, while introducing a rotatable bond, enabled a more electron-rich phosphonate center and a slightly simplified synthesis. Starting from the hydroxymethyl terpyridine derivative **3**, the alcohol could be converted to the corresponding bromomethyl moiety by reflux in H_2SO_4/HBr to generate bromide **11**. Arbuzov reaction of **11** with neat triethylphosphite enabled conversion to diethyl phosphonate **12**. Reaction of **12** with $POCl_3$ at 70°C yielded a monochlorinated intermediate, which was isolated crude.

Nucleophilic attack of *p*-nitrophenol in the presence of Et_3N yielded the mono-ethyl/mono-*p*-nitrophenyl phosphonate cofactor **13**. (Scheme 2.5)



Scheme 2.5 Synthesis of cofactor **13**.

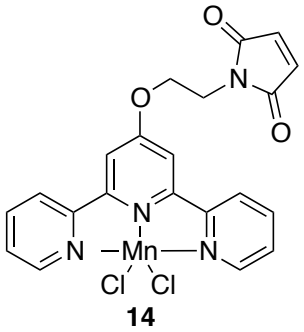
Indeed, once an interstitial methylene was introduced, terpyridine cofactor **13** showed significantly improved bioconjugation behavior, enabling near-complete bioconjugation of *Pfu* POP with a 5:1 excess after overnight incubation at room temperature or after only two hours at 75°C (confirmed by ESI-MS). This hybrid could then be metalated *in situ* for the generation of *Pfu* POP-**13-Mn** or *Pfu* POP-**13-Fe**, exhibiting similar absorbance behavior to that observed during *in situ* metalation with *Pfu* POP-**4** and *Pfu* POP-**10**.

With C-P terpyridine cofactors **10** and **13** in hand, catalysis of their corresponding metalated hybrids could be explored with the assurance that cofactor hydrolysis was not occurring during the course of oxidation catalysis. To date, catalysis has only been explored with hybrids derived from heteroaryl phosphonate cofactor **10**. Hybrids chymotrypsin-**10-Mn** and *Pfu* POP-**10-Mn** were evaluated using the standard conditions

developed for the epoxidation of benzopyran. Disappointingly, no enantioselectivity was observed with either one of these systems, indicating that the selectivity challenges within the terpyridine-manganese manifold arise from a more fundamental problem in catalysis. It is likely that freely-diffusing reactive oxygen species (ROS) or other unconstrained oxidizing species are generated during the course of reaction, allowing for non-selective oxidation reactivity.

2.2.1.1.3 TERPYRIDINE MALEIMIDE ARTIFICIAL METALLOENZYMES

During the time at which the various C-P cofactors were developed, coworkers were able to adapt the many advances in catalysis that were discovered with the chymotrypsin-**4** hybrids to another system employing a maleimide-equipped Mn-terpyridine derivative **14** (Figure 2.8).



14

Figure 2.8 Maleimide conjugated Mn-terpyridine cofactor **14**.

This system leveraged the cysteine-maleimide covalent bioconjugation strategy as described in section 1.3.2.3.2. Using engineered variants of the protein scaffold apo-nitrobindin from *Arabidopsis thaliana* and tHisF from *Thermotoga maritima*, we introduced **14** at specific sites within the scaffolds with good efficiency. The sites of bioconjugation were chosen for their ability to encapsulate the metal cofactor, utilizing

the natural heme-binding pocket of nitrobindin (Figure 2.9) and the interior of the tHisF β-barrel to encapsulate **14** to promote selective oxidation reactivity.

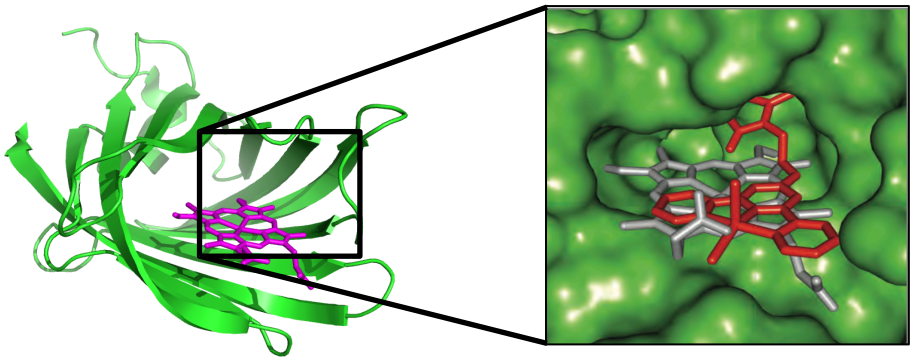


Figure 2.9 The structure of nitrobindin with a model of **14** bound in the heme-binding pocket of the protein

This system vindicated the generality of the catalysis conditions developed using the aforementioned chymotrypsin hybrids in that the various nitrobindin-**14** ArMs successfully catalyzed C-H benzylic oxidation, epoxidation, and oxidative demethylation under identical conditions to those discussed earlier. Even with rational positioning of **14** to enhance cofactor enclosure, however, no significant selectivity in catalysis was observed. This further corroborated the poor selectivity results obtained with terpyridine phosphate/phosphonate cofactors.[33]

2.2.1.1.4 TERPYRIDINE PHOSPHONATE ARMS: WAYS FORWARD

Despite the wealth of chemistry that has been developed with the ArMs described above, we have as of yet been unable to achieve the goal of selectivity in catalysis. A number of new directions are being pursued to leverage this highly versatile platform and enable the desired selectivity. These revolve around two strategies: 1) examining other

metals in the terpyridine ligand, and 2) supporting the terpyridine complexes with proteinogenic ligands.

The first strategy seeks to address the limited breadth of the periodic table that has been explored with this system. By focusing attention on manganese- and iron-catalyzed oxidations, we have omitted other metals that may react through more predictable oxidative mechanisms. Indeed, a number of catalytic terpyridine complexes have been discovered with a multitude of transition metals. These are currently being evaluated through metal screening efforts. It should be noted, however, that screening alternate metals also necessitates the evaluation of alternative oxidants, buffers, and additives, adding extra dimensions to the screening effort.

The second strategy seeks to change the primary coordination environment of the metal to achieve desired selectivity/reactivity in catalysis. We have taken a mutagenesis approach to this problem to leverage one of the key advantages of ArMs: the ability to easily generate mutant libraries of different scaffolds. The *Pfu* POP scaffold was chosen as a platform for this approach, since it exhibits high mutational tolerance due to its stability and it has desirable metal encapsulation properties, enabling proper control of the metal primary- and secondary-coordination spheres. The libraries being explored were generated through a semi-rational mutagenesis strategy developed by coworkers in the Lewis laboratory.[34] This approach, termed "combinatorial codon mutagenesis" (CCM), bears some resemblance to site-directed mutagenesis in that specific codons can be targeted for mutation. Multiple codons can be targeted for mutation simultaneously, with random combinations of these mutations arising per product protein variant. We selected a set of sites (D119, W142, F156, R158, F194, G213, W214, F399, Y401, F404,

I406, M411, R476, Y501, I514, V517, W518, R562, V563, and R600) that sit in or adjacent to the active site of *Pfu* POP (and therefore proximal to the location of the metal cofactor) for the introduction of metal-coordinating amino acids (Figure 2.10). The degenerate codon SAN (IUPAC naming) was identified using CASTer[35] for the introduction of metal-coordinating glutamic acid, aspartic acid, histidine, and asparagine at these sites.

The CCM method is amenable to tailoring the number of mutations expected per product protein polymer, as well. We introduced an average of one mutation/protein to minimize the occurrence of competing ligation by the presence of multiple metal-coordinating residues. This library has been successfully transformed into *E. coli* and will be screened for the identification of metalated derivatives *Pfu* POP-**13** that exhibit selective catalysis.

Figure 2.10 The *Pfu* POP residue positions pursued for CCM using the SAN library method. SAN encodes glutamate, aspartate, histidine, and glutamine, the sidechains of which bear O- and N-donors for metal coordination

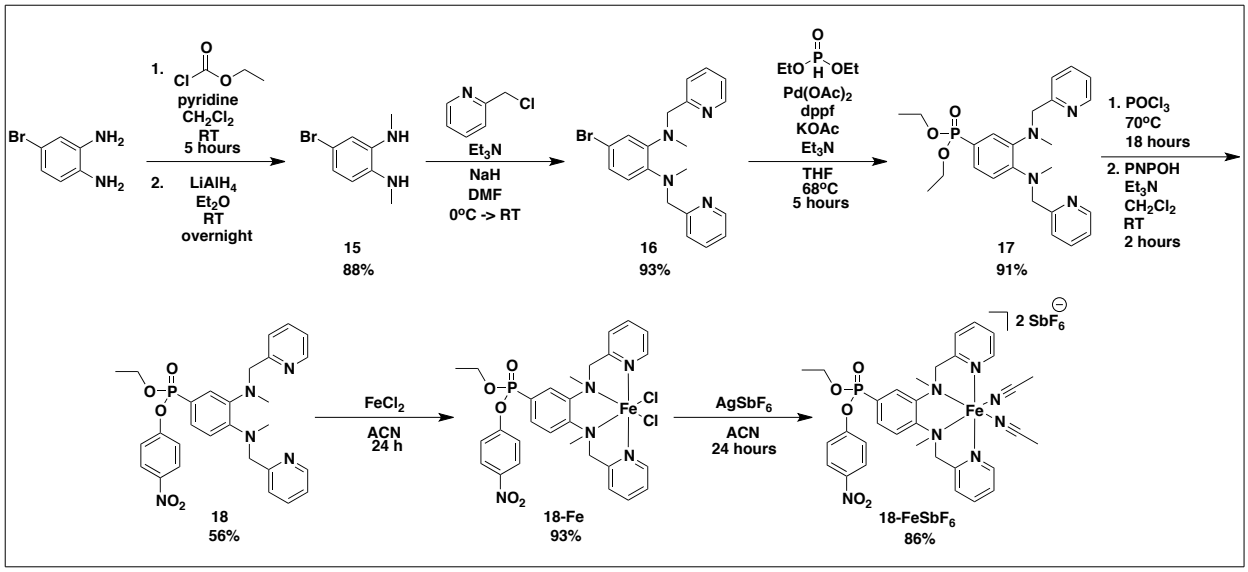
These two screening approaches exemplify the method coined by Ward as "chemo-genetic optimization," which is a strategy uniquely enabled in ArM catalysis.[36] That is, optimization is carried out through traditional chemical screening of metals/conditions as well as through the genetic diversification of protein scaffolds.

2.2.1.2 TETRADENTATE PHOSPHONATE ARTIFICIAL METALLONZYMES

Based on the failure of terpyridine-based ArMs to yield any appreciable selectivity in oxygenation catalysis, we also sought to extract selectivity by varying the

chemical ligand framework around the metal. Tetradentate nitrogen complexes of iron and manganese have been employed extensively in the field of small-molecule C-H oxidation catalysis for the purpose of recapitulating the oxygenation activity of non-heme oxygenases.[37] The Que[38] and White[39] groups have introduced some highly active complexes that enable the oxidation of unactivated aliphatic C-H bonds. Importantly, some of these catalysts have been reported to exhibit altered selectivity upon modification of the secondary coordination sphere[40], posing an appealing target for their introduction into protein scaffolds.

The first tetradentate phosphonate cofactor explored utilized a phenylenediamine-backbone. Fe(II)-complexes with this architecture have been demonstrated as effective aliphatic C-H oxidation catalysts, enabling the oxidation of cyclohexane to cyclohexanone in the presence of hydrogen peroxide.[41] In the context of generating phosphonate ArMs, these were appealing catalysts because of the relative ease with which they could be synthesized. 4-bromophenylenediamine was alkylated by ethylchloroformate followed by reduction with lithium-aluminum hydride to yield the *N,N*-dimethyl product **15**. This compound was then dialkylated with 2-picolyl chloride in the presence of sodium hydride to generate tetradentate bromide **16**. After this, Hirao coupling was performed with diethylphosphite to furnish the diethylphosphonate **17**, which could subsequently be monochlorinated with $POCl_3$ to generate an intermediate mono-phosphorochloridate. This was then treated with *p*-nitrophenol in the presence of base to generate desired cofactor **18**. This compound could be carried on to metalation with $FeCl_2$ to generate intermediate **18-Fe**, followed by treatment with $AgSbF_6$ for salt metathesis to form active catalyst **18-FeSbF$_6$**. (Scheme 2.6)



Scheme 2.6 Synthesis of the tetradentate cofactor **18** and its metalated derivative **18-FeSbF$_6$.**

Cofactors **18** and **18-FeSbF$_6$** were both applied in the bioconjugation to scaffold *Pfu* POP. These exhibited excellent behavior in bioconjugation, readily yielding hybrids *Pfu* POP-**18** and *Pfu* POP-**18-FeSbF$_6$** with high extent of bioconjugation despite the low excess of cofactor employed (5:1 ratio). Indeed, these cofactors serve as evidence that the electronics of the pendant arene in arylphosphonate cofactors determine the electrophilicity of the phosphorus center and, in turn, their stability toward hydrolysis by aqueous buffer.

In light of this enhanced stability, this cofactor was also employed to test the stability of phosphonate hybrids in general. During a period of 12 days, a *Pfu* POP-**18** hybrid was incubated at room temperature in buffers at pH 6.0, pH 7.0, and pH 8.0. The stability of this hybrid was assessed by LC/ESI-MS, revealing that after 12 days, no noticeable change could be observed in the extent of hybrid bioconjugation regardless of

pH (Figure 2.11). This stands to highlight the stability of the C-P linkage between the cofactor and its scaffold, corroborating the same result reported by van Koten.[29]

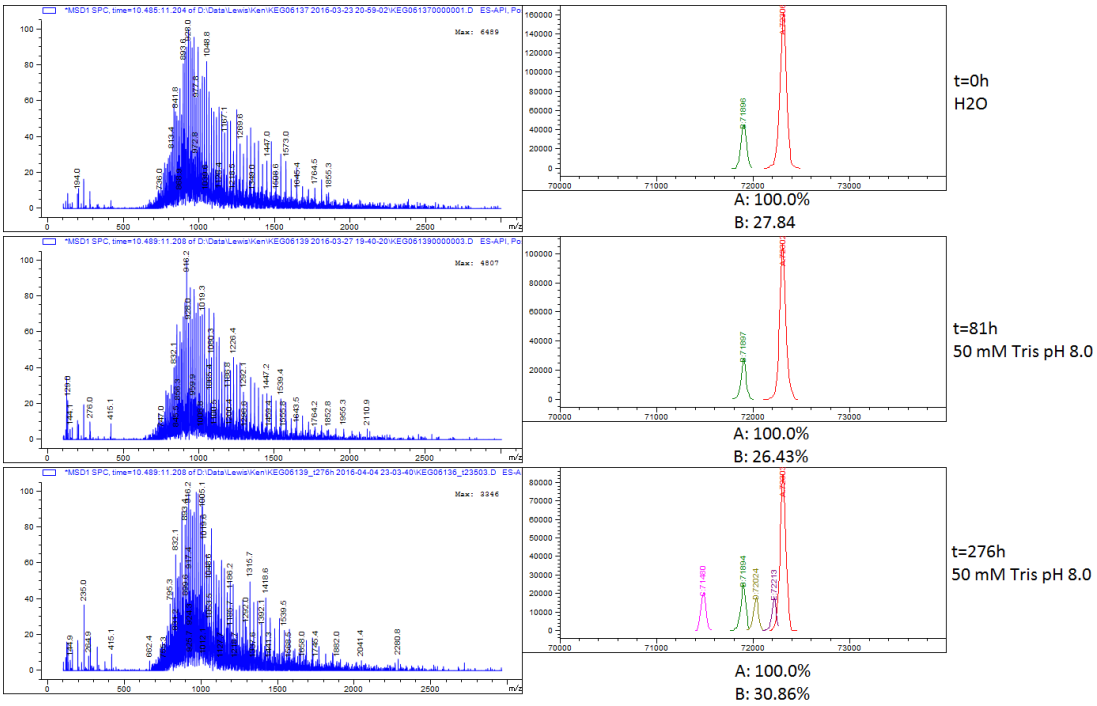
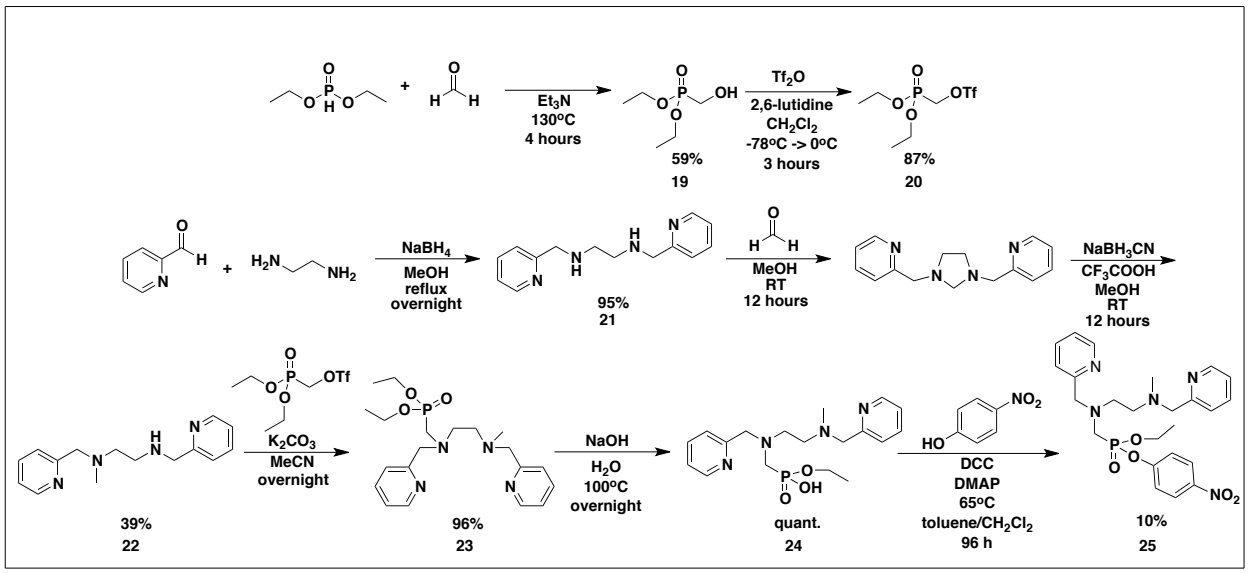


Figure 2.11 Monitoring the degradation of hybrid *Pfu* POP-**18** by ESI-MS (red peak=hybrid, green peak=unreacted scaffold) over 12 days shows minimal degradation of the hybrid mass at pH 8.0, as was seen for pH 6.0 and pH 7.0.

Despite the exciting prospect of catalysis with **18** and its corresponding metalated derivatives, we have not yet observed turnover of this catalyst in limited trials of ArM catalysis. Investigations into this catalyst are ongoing.

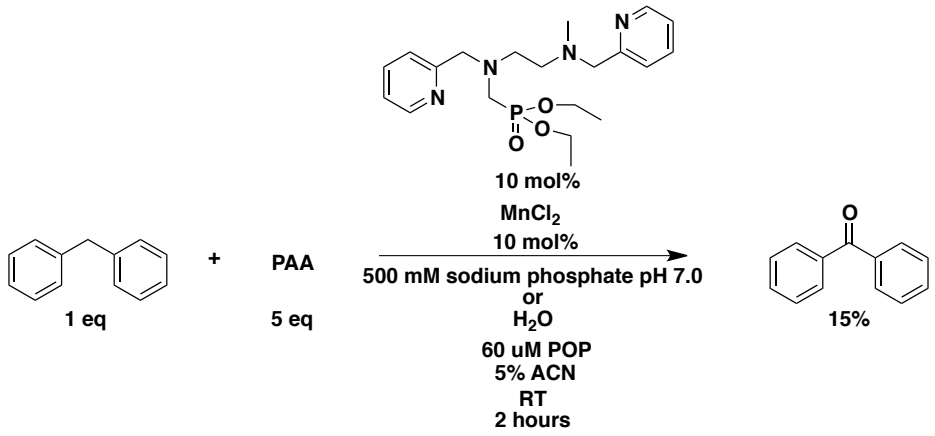
Another tetradentate nitrogen phosphonate ligand was designed based on results that were obtained from the Ménage group. Using a similar framework as that explored with NikA[42] and human serum albumin scaffolds[43] by this group, a new tetradentate catalyst was pursued. First, a general phosphonate alkylation reagent **20** was pursued.

The reaction of diethylphosphite with formaldehyde in the presence of Et_3N yielded hydroxymethylphosphonate **19**. Treatment with triflic anhydride generated triflate **20**, which can be used as a general electrophile for the generation of a number of different phosphonates. To generate the ligand fragment, reductive amination with ethylenediamine and 2-picolylaldehyde yielded the tetradentate ligand **21**. Monomethylation was achieved by cyclization of **21** with formaldehyde to form an imidazoline intermediate which was reduced with $NaBH_3CN$ in the presence of trifluoroacetic acid to make monomethyl ligand **22**. This was used for base-promoted nucleophilic attack on triflate **20** to alkylate the secondary amine and generate diethylphosphonate **23**. This product proved sensitive toward activation through chlorination procedures. Therefore, hydrolysis of **23** was carried out using NaOH to form the phosphonic acid intermediate **24**. High-temperature Steglich esterification of this product with *p*-nitrophenol enabled formation of the product cofactor **25** (Scheme 2.7). This product was exceptionally challenging to purify leading to great losses in yield.



Scheme 2.7 The synthesis of cofactor **25**.

To evaluate the oxidation reactivity of these cofactors under conditions amenable to ArM catalysis, we screened a number of conditions with ligand **23** to narrow down the efficacy of this catalyst for C-H oxidation. Ultimately, we found a set of conditions that enabled the ligand-dependent turnover of *in situ* generated **23-$MnCl_2$** for benzylic oxidation of diphenylmethane to benzophenone in the presence of the *Pfu* POP scaffold, albeit in relatively low yield (Scheme 2.8).



Scheme 2.8 Reactivity of **23-$MnCl_2$** for catalysis of benzylic oxidation in the presence of *Pfu* POP

These results suggest an exciting prospect for selective C-H oxidation catalysis enabled by the tetradentate ligand framework provided by **34**. Bioconjugation to *Pfu* POP is expected to be facile with **36**, as has been demonstrated for the aforementioned systems. Work to evaluate metalated ArMs containing cofactors **26** and **36** is underway and will hopefully yield interesting selectivity in oxidation catalysis.

2.2.1.3 ARTIFICIAL METALLOENZYME OXYGENATION CATALYSIS: CONCLUSIONS

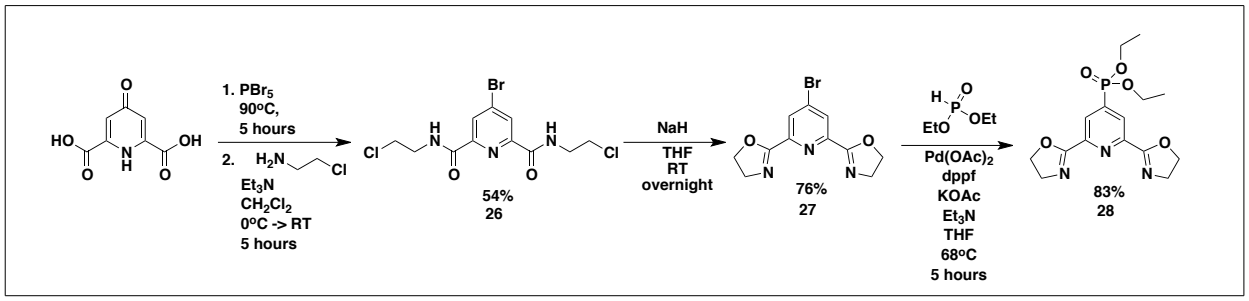
Much work remains to be done to extract the sort of selectivity that is expected from ArMs in the oxygenation manifold. These face competition from p450 catalysts,

which, as discussed earlier, present a number of beautifully compelling examples of site-selective C-H oxygenation catalysis.[10] It should be noted, however, that creating ArMs to achieve this end has the potential to expand on the scope of selectivity and reactivity that can be achieved by current biocatalysts. The flexibility to choose new scaffolds while retaining a common catalytic core should be appreciated as a major boon with the ArM platform. This could enable the rapid selective transformation of different substrates based on the natural substrate-binding propensity of the scaffold alone, as has been demonstrated with ArM nucleases.[44]

2.2.2 ARTIFICIAL METALLOENZYME AZIDATION CATALYSIS

Finally, it should be noted that complexes exhibiting alternative reactivity have been explored with the phosphonate anchoring method. An appealing target complex was identified when Hartwig presented an example of aliphatic C-H azidation using the simple pyridine bis-oxazoline ligand framework in conjunction with $Fe(OAc)_2$.[45] This system exhibits ligand-dependent reactivity with hypervalent azidoiodinane to generate an oxidizing species for the conversion of secondary and tertiary C-H bonds to C-N_3 bonds. The prospects of Fe-pybox are particularly exciting because of the small molecule's inherent site-selectivity, its ability to differentiate C-H bonds with similar reactivity.[45] Therefore, we embarked on an effort into isolating a phosphonate cofactor bearing a pybox ligand. Chelidamic acid was brominated by heating this solid in the presence of PBr_5, which was followed by attack with 2-chloroethanamine to make the diamide intermediate **26**. Cyclization of this molecule was promoted by deprotonation of the amide functionalities with NaH, generating *p*-bromo pybox **27**. This species proved amenable to Hirao coupling for the generation of diethylphosphonate **28**. Attempts to

chlorinate this system to furnish the activated *p*-nitrophenyl phosphonate have been unsuccessful to date owing to the sensitivity of the oxazoline motif. Recent developments in the methodology for high-temperature Steglich esterification (as seen in the conversion of **23** to **25**) are being explored to generate the desired cofactor **28**. (Scheme 2.9)



Scheme 2.9 Synthesis of phosphonate pybox **28**

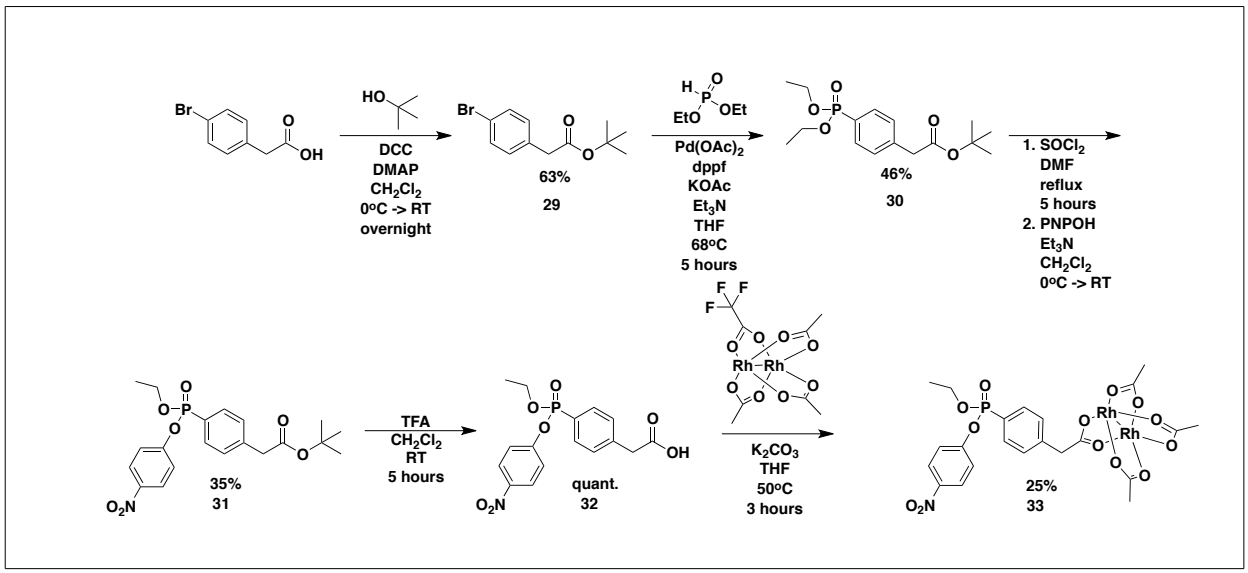
2.2.3 DIRHODIUM PHOSPHONATE ARTIFICIAL METALLOENZYMES

Beyond the oxidative manifold, we have also been greatly interested in the carbene insertion chemistry of dirhodium tetracarboxylate complexes. These have seen tremendous success in our laboratory in their application with *Pfu* POP, which has been shown to be an amenable scaffold for enabling enantioselective carbene insertion into alkenes, S-H, Si-H, N-H, and to a small extent, C-H bonds. These *Pfu* POP-based systems rely on an alternative covalent bioconjugation strategy, utilizing a genetically-encoded *p*-azidophenylalanine residue for the introduction of an azide bioorthogonal "click" handle, which is exploited for bioconjugation with a strained bis-cyclononyne-linked dirhodium cofactor (Figure 2.13).[46] Details of this system will be described in greater detail in chapter four.

In contrast to the above method, which requires protein engineering and synthetic biology tools to achieve the requisite link between cofactor and scaffold, covalent anchoring of phosphonate-based cofactors proceeds without such intervening

complexities. As seen, the serine hydrolase *Pfu* POP can be expressed and covalently modified in its native form using electrophilic phosphonate-based anchors. Therefore, we sought to develop various phosphonate-linked dirhodium cofactors to achieve this end.

The first dirhodium phosphonate system explored employed a relatively simple "monodentate" linkage to the dirhodium complex. That is, one equatorial carboxylate moiety was responsible for connecting the phosphonate anchor to the dirhodium lantern complex. This system was developed because of its relatively simple synthesis. Starting from 2-(4-bromophenyl)acetic acid, the *tert*-butyl ester **29** was generated. Protection of the carboxylate moiety is necessary, as Hirao coupling is incompatible with carboxylic acids. Hirao coupling using diethyl phosphite furnished the diethylphoshonate **30**. Chlorination in thionyl chloride followed by attack with *p*-nitrophenol in the presence of Et_3N generated activated cofactor **31**. Deprotection with trifluoroacetic acid yielded the free carboxylic acid **32**. This was followed by metalation with $Rh(O_2CCH_3)_3(O_2CCF_3)$ in the presence of K_2CO_3 to generate cofactor **33**. (Scheme 2.10)



Scheme 2.10 The synthesis of cofactor **33**.

Bioconjugations of **33** and *Pfu* POP have proven to be somewhat problematic, however. On its surface, the behavior of this cofactor echoes the behavior of the various polydentate nitrogen phosphonate cofactors, yielding enzyme-promoted release of *p*-nitrophenolate, which can be monitored by UV/Vis. Attempts to characterize the resulting hybrids by ESI-MS have been consistently hampered by poor ionization, even after LC chromatographic separation to remove interfering small-molecule components. This has proven detrimental to the validation of the identity of these ArMs. The presence of the dirhodium fragment seems to be the main culprit for the poor MS results observed, as bioconjugation with the non-metalated cofactor **32** yields strong ionization of the hybrid in close agreement with its theoretical mass. Further investigations seemed to reveal that higher temperatures in bioconjugation led to qualitatively inferior ESI-MS ionization. In addition, the use of 50 mM MES at pH 6.0 was found to give consistently superior ESI-MS data relative to buffers at pH 7.0 and 8.0. In addition, lower excesses of dirhodium cofactor were found to yield better hybrid product ions. With this in mind, bioconjugations using cofactor **33** and subsequent dirhodium phosphonate cofactors with *Pfu* POP were all performed at room temperature with low excesses (5:1) of cofactor:hybrid and at pH 6.0, requiring longer incubation times but leading to somewhat more consistent bioconjugation data. Some trials revealed significant extent of demetalation of the cofactor under room-temperature bioconjugation conditions, leading to the hypothesis that the lability of equatorial carboxylate ligands on these dirhodium complexes might be to blame for the heterogeneity of results. Indeed, lability of

dirhodium carboxylates has been noted by Ball, particularly in the context of complex media like the intracellular environment.[47]

In light of the observed characterization issues with the *Pfu* POP-**33** ArMs, we also sought to uncover the role of another interfering effect, that of non-specific binding of excess **33** to the POP scaffold after bioconjugation. This insidious effect was noted as a potential source of ionization issues and was corroborated by the persistent struggle to purify excess cofactor away from product hybrid after bioconjugation, as seen in Figure 2.12.

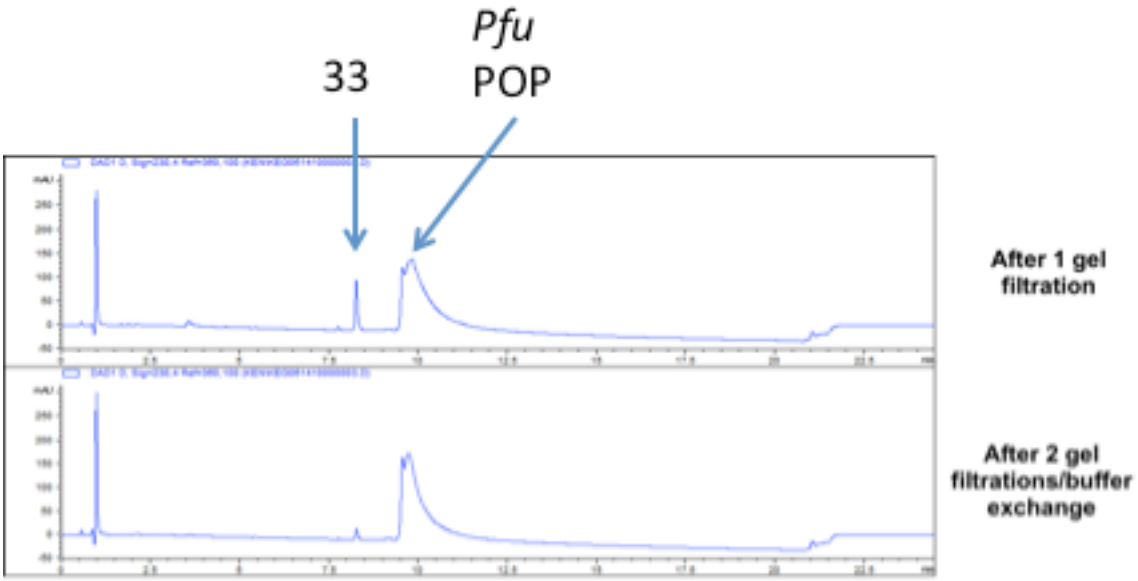


Figure 2.12 HPLC traces revealing the persistence of cofactor **46** with hybrid *Pfu* POP-**46** after multiple rounds of purification.

Size-exclusion chromatography and membrane buffer exchange failed to effect the full removal of excess cofactor, a major impediment to catalysis with this system, as non-specifically bound cofactor can lead to non-selective reactions with this catalyst. Indeed, non-specific binding of the cofactor was corroborated by observation of the

HPLC chromatograms of cofactor **33** added to a solution of *Pfu* POP S477A, a scaffold mutant lacking the requisite serine for bioconjugation. "Consumption" of the cofactor was observed immediately upon combination of the protein and the cofactor (as compared to a buffer-only control). This indicates that cofactor **33** was either non-specifically bound to the protein, or that carboxylate ligand exchange was occurring between the cofactor and protein-centered carboxylate sidechains.

To address the issues with *Pfu* POP-**33** LC/ESI-MS characterization and purification of these hybrids, we then pursued phosphonate cofactors with structural similarity to the bis-cyclononyne-esp-based dirhodium cofactor **(**Figure 2.13), which consistently produces excellent hybrid ESI-MS signal.

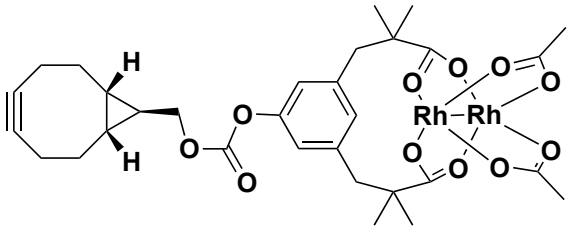
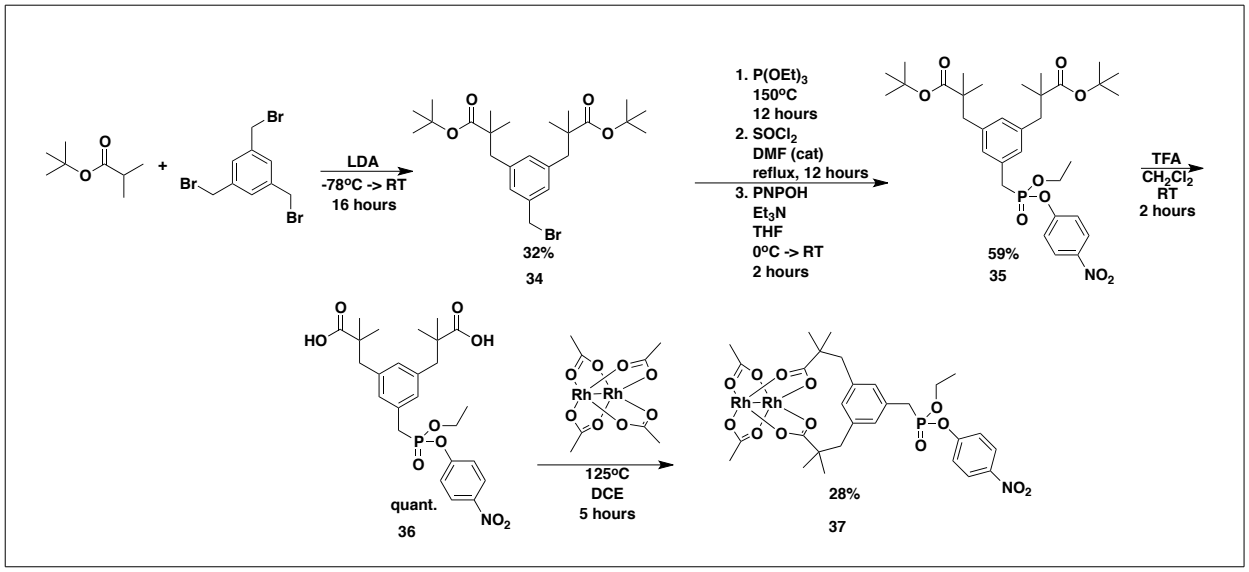


Figure 2.13 RhBCN, a click cofactor with an esp-chelating ligand architecture for dirhodium binding

We hypothesized that the esp ligand framework contributes greater stability against carboxylate ligand exchange, as Du Bois had hypothesized in the development of the esp ligand.[48] Two esp dirhodium phosphonate cofactors were pursued.

The first of these complexes was generated through a novel esp synthetic pathway. *tert*-butyl isobutyrate was deprotonated with LDA and added slowly to a solution of 1,3,5-tris-(bromomethyl)benzene to yield the dialklyated product **34**. This was converted to diethylphosphonate via Arbuzov reaction by refluxing in neat triethyl

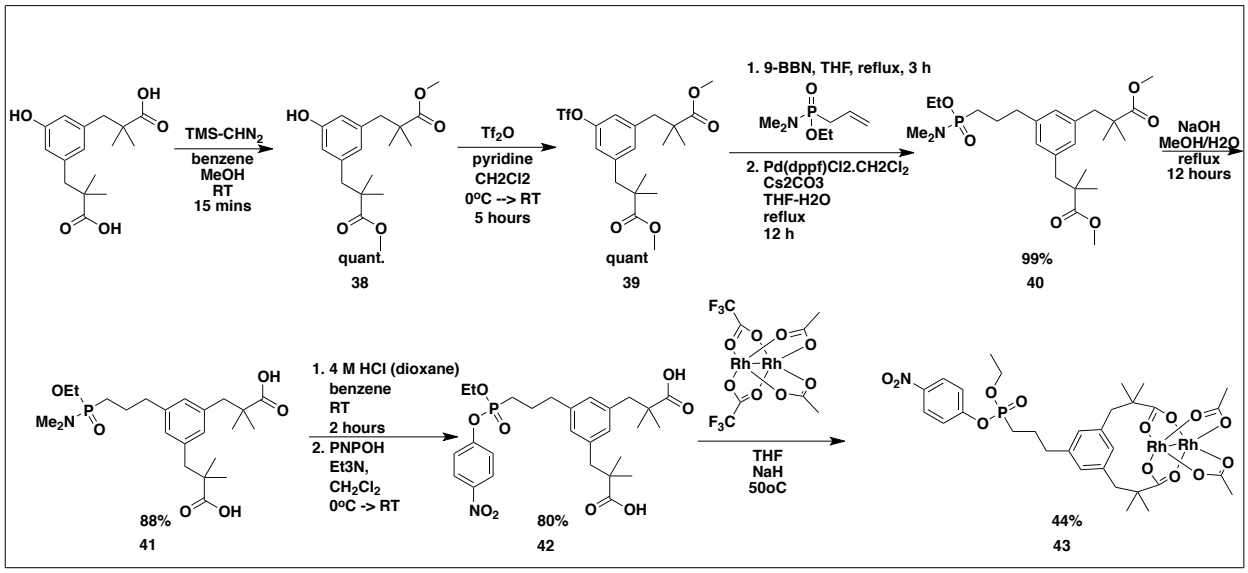
phosphite. The crude phosphonate was monochlorinated with $SOCl_2$, and this was substituted by nucleophilic addition of *p*-nitrophenol in the presence of Et_3N, yielding cofactor **35**. Deprotection with trifluoroacetic acid yielded esp product **36**. Metalation could be carried out directly by heating the ligand in the presence of $Rh_2(O_2CCH_3)_4$, generating the desired product dirhodium cofactor **37**. (Scheme 2.11)



Scheme 2.11 The synthesis of esp-Rh_2 cofactor **37**.

A more complex synthetic sequence was carried out to generate cofactor **43**, which is structurally similar to **37** but contains a longer propyl linker. Starting from a phenol esp ligand[49], the carboxylate arms were esterified by the action of TMS-CHN_2 to generate diester **38**. This compound was then carried on to the triflate **39** by reaction with triflic anhydride in pyridine. Suzuki coupling with the *in situ* generated boronate ester of allyl ethyl *N,N*-dimethylaminophosphinate[27] generated propyl-linked phosphoramidate **40** in excellent yield. Hydrolysis of **40** with NaOH generated dicarboxylate **41**. In contrast to the previously described methods, the intermediate phosphorochloridate was produced by treating the phosphoramidate with dry HCl, leading to protonation and substitution with

chloride. *p*-nitrophenol was added through base-promoted nucleophilic attack to generate the activated phosphonate cofactor **42**. Deprotonation of the carboxylate arms of **42** with NaH followed by addition of $Rh(O_2CCH_3)_3(O_2CCF_3)$ yielded dirhodium cofactor **4**. (Scheme 2.12)



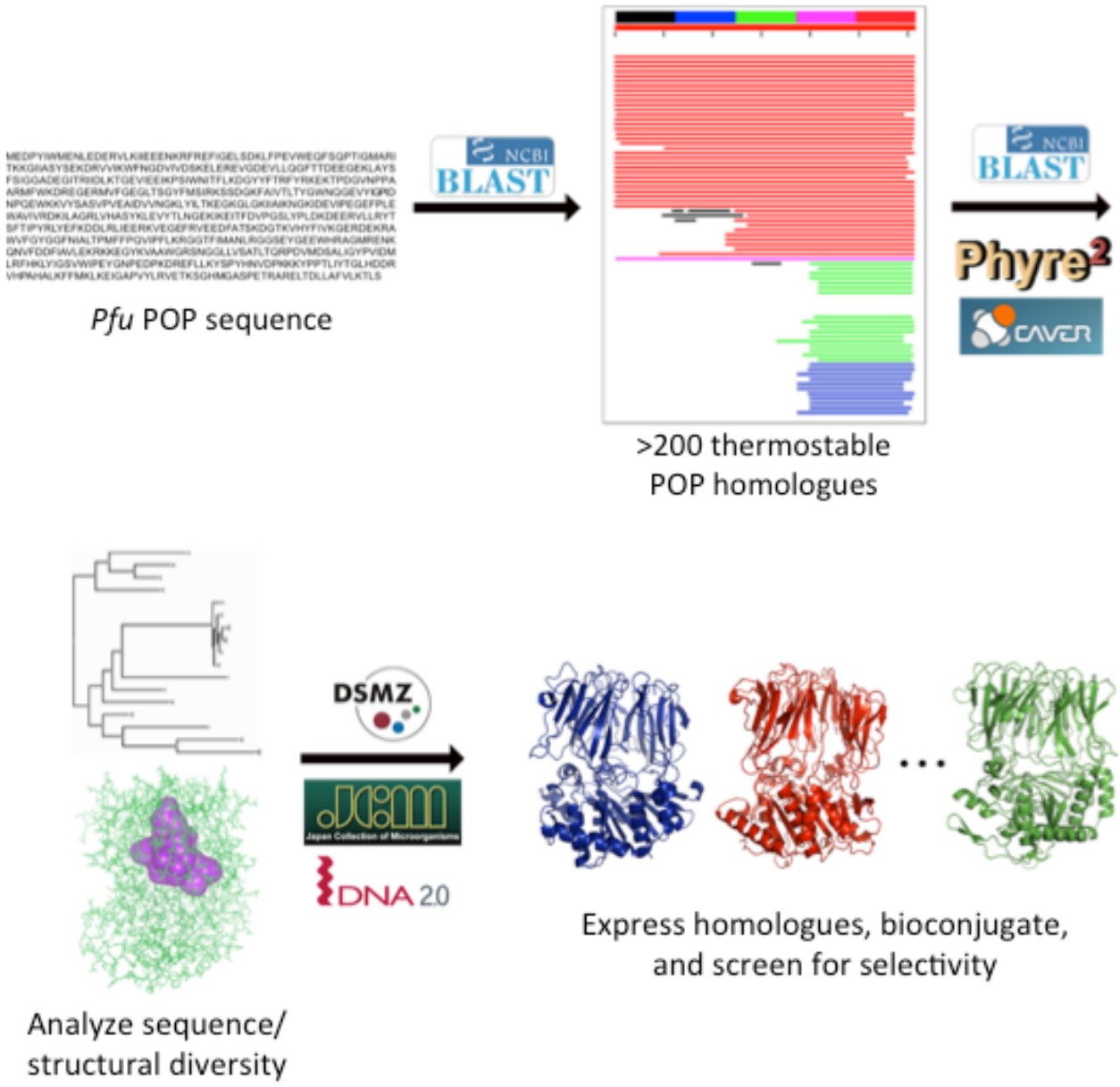
Scheme 2.12 The synthesis of extended-linker Rh_2-esp cofactor **43**

In spite of the exciting prospects of these cofactors, we encountered similarly vexing issues in the characterization of the corresponding *Pfu* POP hybrids, mainly with regard to removal of excess cofactor. Again, non-specific binding was suspected to be the source of these problems. Despite purification issues, *Pfu* POP-**43** was successfully applied toward the cyclopropanation of styrene and methyl 2-diazo-2-(4-methoxyphenyl)acetate. This hybrid induced a small degree of enantioselectivity in cyclopropanation (16% e.e.) in spite of the presence of some free/non-specifically bound cofactor, signaling promising potential once an adequate solution to these purification issues is achieved.

2.2.4 ALTERNATE ARCHAEAL PROLYL OLIGOPEPTIDASE FAMILY SCAFFOLDS

A critical advantage of the mechanism-based covalent inhibition of serine hydrolases arises from their ability to be utilized as scaffolds without the prerequisite of protein engineering. This can be highly enabling for the evaluation of new scaffolds for ArMs, as one need only obtain a wild-type serine hydrolase (commercially or through expression/isolation) to evaluate its utility as a potential scaffold. Indeed, this fact enabled the rapid evaluation of a panel of viable serine hydrolase scaffolds described above. In addition, since the nucleophilic serine often sits in a well-defined, deep active site, this approach effectively guides metal cofactors into environments likely to influence the secondary coordination sphere of the metal. *Pfu* POP, a demonstrably effective hydrolase ArM scaffold, was chosen for this reason.

Drawing from the success of *Pyrococcus furiosus* prolyl oligopeptidase as a scaffold for ArMs, we sought to explore the structural diversity provided by related homologues of this enzyme. Given that all homologues of this enzyme are serine hydrolases, they represent viable candidates as scaffolds for ArMs generated with any of the cofactors noted above. This structural diversity could be leveraged to rapidly evaluate panels of scaffolds that bore the same general architecture of *Pfu* POP (guaranteeing metal encapsulation) but sufficient differences to induce novel selectivity. The strategy described is summarized in Scheme 2.13.



Scheme 2.13 A summary of the methods used to generate the library of thermophilic POP homologues explored

An initial protein BLAST[50] search of the *Pfu* POP gene was conducted within the phylum euryarchaeota. This was chosen to ensure that the homologues were drawn from extremophilic source organisms, as enzymes from these organisms tend to exhibit high

stability. Structural stability is a desirable feature in ArM scaffolds, as they need to be robust toward non-native catalytic conditions and must be tolerant of mutational loads to facilitate genetic optimization.[51] From this initial BLAST search was discovered 254 candidate scaffold sequences. Methods to cull this pool of homologous scaffolds were explored, as 254 species presented an intractably large set of sequences for cloning and expression. To select scaffolds with the highest potential for successful induction of selectivity, we looked to categorize them by the size of their internal cavity. Constriction of the internal cavity of the enzymes is desirable, as this could lead to the conformational restriction of the metal cofactor within the scaffold context. Of course, measuring the internal cavity of the enzymes necessitated homology modeling, as most of these constructs have not yet been structurally characterized. Using the homology-modeling server Phyre2[52], a set of homology models was generated based on the primary sequences of the identified enzymes. The internal cavity volume of the homology models was measured using CaverAnalyst,[53] and the resulting volumes were tabulated and binned (Figure 2.14). With this list, a subset of the genes with the lowest identified volumes was cloned from the genomic DNA of the source organisms (ordered from Japan Collection of Microorganisms).

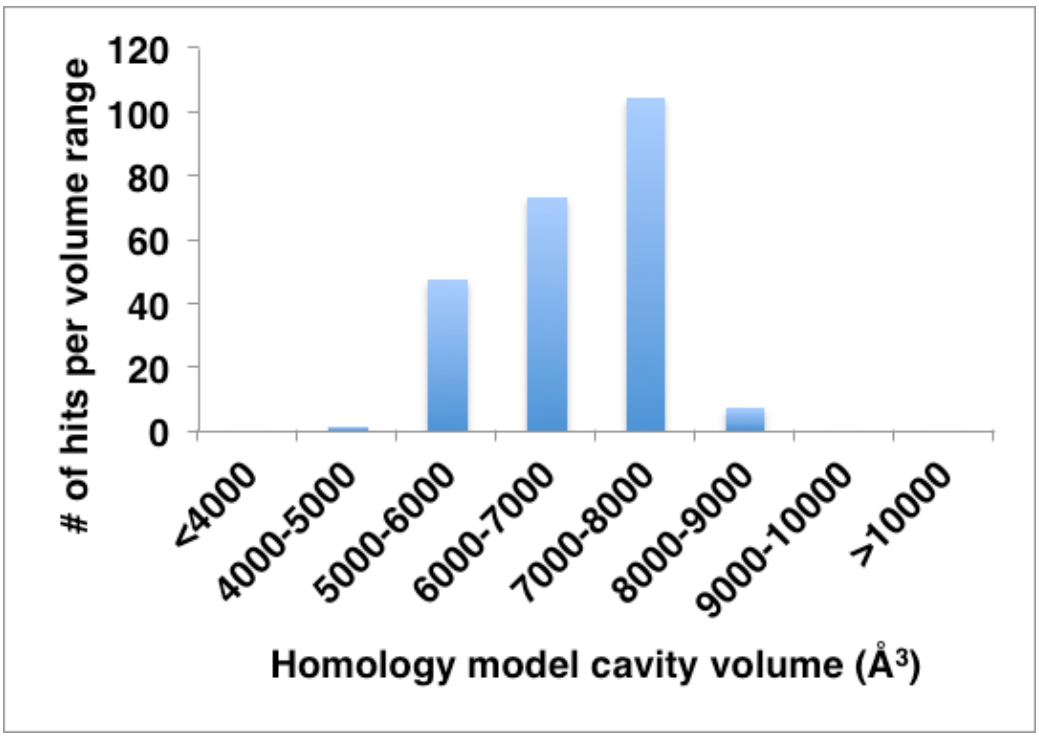


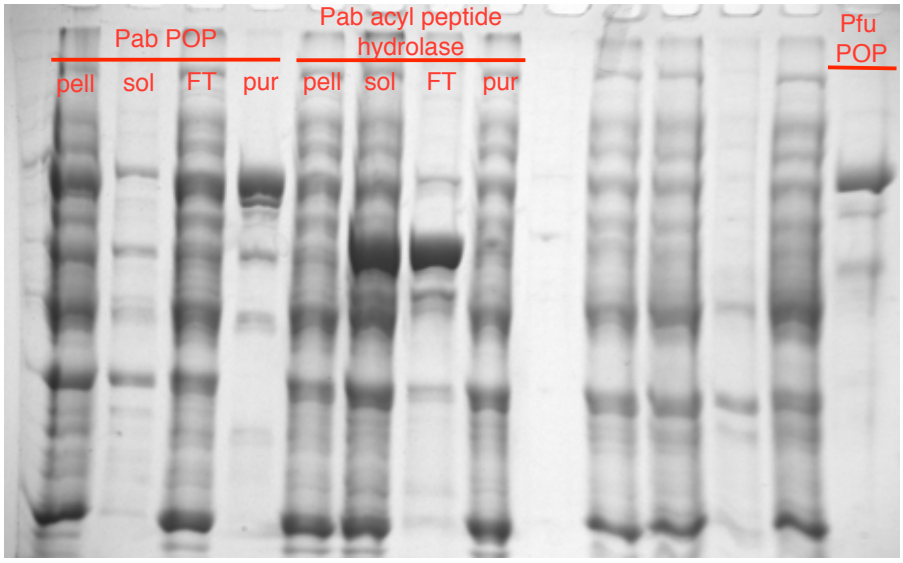
Figure 2.14 The cavity volumes obtained through Phyre2 homology modeling.

Most of the cloning performed on this gene set was performed by coworkers and will not be discussed in further detail. Three POP-homologues from the genome of archaeon *Pyrococcus abyssii* were successfully cloned in this work. This set includes *Pab* prolyl oligopeptidase (PAB0762), *Pab* acyl-peptide hydrolase (PAB1300), and *Pab* peptidase S9 (PAB1418). These genes were cloned from genomic DNA and introduced into pET28a. Due to the high rare codon content of these genes, Rosetta2(DE3) *E. coli* cells were used as host organisms, as these are supplemented with 7 tRNA synthetases for more efficient expression of rare-codon-rich genes. In expression tests, only *Pab* POP (Figure 2.15 A) and *Pab* peptidase S9 (Figure B) were successfully expressed and purified as full soluble constructs. Some optimization of inducer concentration was required to achieve adequate expression of *Pab* peptidase S9. *Pab* acyl-peptide hydrolase

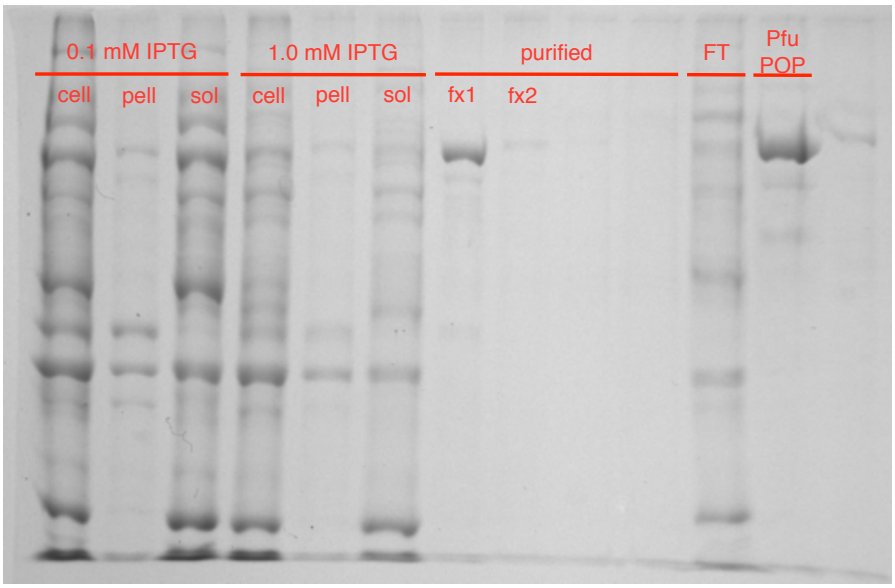
was expressed as a truncated insoluble construct likely due to expression initiated at an alternative downstream Met start codon. The well-behaved *Pab* POP and *Pab* peptidase S9 were both purified by NiNTA and isolated as soluble enzymes, albeit with somewhat low yields.

Figure 2.15 Expression SDS-PAGE for alternate POP-family thermophilic scaffolds

A) *Pab* POP and *Pab* acyl-peptide hydrolase expression



B) *Pab* POP S9 peptidase expression tests with variable inducer loading



Bioconjugation was successfully carried out with these scaffolds and cofactor **46**, as confirmed by UV/Vis monitoring of release of *p*-nitrophenolate. Comparison of the selectivity observed in the cyclopropanation reactions catalyzed by these ArMs was performed. Minimal selectivity was observed with hybrids of **46** and the *Pab* variants, but this could be ascribed to the difficulty of removing the excess contaminating cofactor. Selectivity notwithstanding, the successful expression and bioconjugation of these alternative scaffolds in their native form attests to the generality of hybrid formation with phosphonate-type inhibitors without the prerequisite for complex protein engineering.

2.3 CONCLUSIONS

The large variety of electrophilic phosphonate cofactors described herein speaks to the generality of this anchoring chemistry for the bioconjugation of a number of different metal catalysts. Of course, the synthesis of such cofactors has proven to be exceptionally challenging in some cases, which can detract from the overall simplicity of the method. With well-developed syntheses, however, this method can afford highly efficient routes to ArM formation.

In addition, these phosphonate cofactors deliver on the promise to enable bioconjugations to native serine hydrolases without the prerequisite of protein engineering, rapidly affording diverse sets of ArMs that would otherwise be difficult to achieve using alternative engineering-based covalent approaches. Bioconjugation may be complicated, of course, by sensitivity of the anchor to the conditions applied. Care must be made to optimize bioconjugation conditions to minimize anchor hydrolysis, which is an inevitable concern with this method. Furthermore, compatibility of the phosphonate with its linked metal complex should be ensured, as this can lead to complications in

bioconjugation. Pathologies like excessive electrophilicity of the phosphonate (as seen with terpyridyl phosphonate **10**) as well as challenging removal of non-specifically bound cofactor (as seen with the dirhodium phosphonate cofactors **25**, **37**, and **43**) can be a major source of problems with phosphonate cofactors. In spite of these issues, a number of excellent examples of ArM formation with phosphonate cofactors have been demonstrated.

Finally, the multitude of scaffolds that have been explored with these phosphonate illustrates the utility of exploiting this bioconjugation methodology to rapidly screen a great number of different ArM protein scaffolds. The number of readily achievable scaffold/metal permutations poses great promise in the exploration of novel ArM selectivity.

2.4 EXPERIMENTAL

Materials

Unless otherwise noted, all reagents were obtained from commercial suppliers and used without further purification. Benzene, dimethylformamide (DMF), acetonitrile (CH_3CN), tetrahydrofuran (THF), and methylene chloride (CH_2Cl_2) were obtained from a PureSolv MD solvent purification system by Innovative Technology (solvent deoxygenated by N2 sparge and dried over alumina). Acetonitrile (CH_3CN) 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). *E. coli* Rosetta2(DE3) cells were purchased from

Novagen. Nco I and Not I restriction enzymes, T4 DNA ligase, Taq DNA polymerase and Phusion HF polymerase (Cat# 530S) were purchased from New England Biolabs (Ipswitch, MA). Luria broth (LB) 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. 5 mL HisTrap NiNTA columns and 5 mL HiTrap Desalting columns were purchased from GE Healthcare Life Sciences (Pittsburgh, PA). 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 Methods

Unless otherwise specified, all reactions were prepared in flame or oven-dried glassware under an inert N2 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. Reverse phase chromatography was carried out using a Biotage Isolera One. Protein chromatography was performed using a GE Healthcare Life Sciences ÄKTA FPLC system. 1H and 31P NMR spectra were recorded at 500 MHz and 202 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. 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.

Experimental Procedures

Ethyl bis(4-nitrophenyl) phosphate (1): A 250 mL round bottom flask with a stir bar and septum was charged with 4-nitrophenol (1.948 g, 14.0 mmol) and NaH (0.980 g, 24.5 mmol) (60% suspension in mineral oil), evacuated and purged 3 times with N_2. The flask was placed in an ice bath and 50 mL dry THF was added to dissolve the solids. The reaction mixture was warmed to room temperature and stirred for 1 hour. Ethyl dichlorophosphate (0.831 mL, 7.0 mmol) was added dropwise. Upon full addition, the reaction mixture was stirred for 2 hours at room temperature. Filtered reaction mixture through celite and concentrated the filtrate to obtain a crystalline solid. Recrystallized from 50% EtOAc/Hexanes to obtain white crystals. 79% yield. (2.01 g, 5.46 mmol)
^{1}H NMR (500 MHz, $CDC1_3$) δ 8.28 (dt, J = 9.0, 0.7 Hz, 4H), 7.40 (dt, J = 9.0 Hz, 1.0 Hz, 4H), 4.10 (dq, J = 7.0 Hz, 9.0 Hz, 4H), 1.41 (dt, J = 7.0, 1.0 Hz, 4H).
^{31}P NMR ($CDCl_3$) δ 12.35

[2,2':6',2''-terpyridin]-4'-ylmethanol (3): A 50 mL round bottom flask with a stir bar and rubber septum was charged with 500 mg (1.803 mmol) 2,2':6',2''-terpyridine-4'-carboxylic acid **2**, evacuated and purged 3 times with N_2. The solid was dissolved in 14

mL dry benzene and 4 mL dry MeOH. A solution of (trimethylsilyl)diazomethane (2.0 M in hexanes) (1.082 mL, 2.164 mmol) was added dropwise to the reaction mixture. The reaction mixture was stirred at room temperature for 1 hour, after which volatiles were removed *in vacuo*. The crude 2,2':6',2"-terpyridine-4'-carboxylic acid methyl ester was redissolved in 10 mL dry MeOH and cooled to 0°C in an ice bath. Sodium borohydride (566.5 mg, 14.98 mmol) was carefully added to the stirring mixture. Upon full addition, the reaction flask was fitted with a reflux condenser and the reaction mixture was refluxed overnight. The reaction was cooled to room temperature and carefully quenched with 1 M HCl (10 mL). The reaction mixture was transferred to a separatory funnel and neutralized with saturated sodium bicarbonate. The aqueous phase was extracted with CH_2Cl_2 (3 x 50 mL) and ethyl acetate (3 x 50 mL). The organic layers were combined, dried with Na_2SO_4, and concentrated to afford **3** as a white/pink solid in 90% yield. (427 mg, 1.61 mmol). ESI-MS calculated for $C_{16}H_{14}N_3O$ $[M+H]^+$=264.11, found 264.1. ^{1}H NMR (500 MHz; CDCl3): δ 8.71 (d, *J* = 4.7 Hz, 2H), 8.63 (d, *J* = 7.9 Hz, 2H), 8.47 (s, 2H), 7.87 (t, *J* = 7.7 Hz, 2H), 7.35 (dd, *J* = 7.0, 4.6 Hz, 2H), 4.92 (d, *J* = 6.4 Hz, 2H).

[2,2':6',2''-terpyridin]-4'-ylmethyl ethyl (4-nitrophenyl) phosphate (4): A 25 mL round bottom flask with a stir bar and rubber septum was charged with 213.4 mg (0.811 mmol) [2,2':6',2"-terpyridin]-4'-ylmethanol **3**. The flask was evacuated and purged 3 times with N_2. The solid was dissolved in 8 mL dry THF. To this solution was added 245.6 mg (6.14 mmol) NaH (60% suspension in mineral oil). The reaction was heated to 50°C and stirred for 2 hours. In that time, a separate 100 mL round bottom flask containing a stirred solution of 896 mg (2.43 mmol) ethyl bis(4-nitrophenyl) phosphate **1**

in 40 mL dry THF was prepared and cooled to 0°C. The deprotonation mixture of **3** was cooled to room temperature and transferred dropwise to the cooled stirred solution of **1** via cannula. After full addition, the reaction mixture was warmed to room temperature and stirred for two additional hours. The mixture was concentrated, loaded onto silica, and the product **4** was purified by silica flash chromatography (4% MeOH/CH_2Cl_2). 77% yield. (278 mg, 0.624 mmol). ESI-MS calculated for $C_{24}H_{22}N_4O_6P$ $[M+H]^+$=493.13, found 493.1.

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.70 (s, 2H), 8.60 (d, *J* = 8.0 Hz, 2H), 8.41 (s, 2H), 8.14 (d, *J* = 8.0 Hz, 2H), 7.87 (t, *J* = 7.5 Hz, 2H), 7.36 (d, *J* = 9.0 Hz, 4H), 5.39-5.30 (m, 2H), 4.34 (dt, *J* = 14.5, 7.2 Hz, 2H), 1.41 (t, *J* = 6.9 Hz, 3H).

^{31}P-NMR (202 MHz; $CDCl_3$): δ -7.12.

Manganese(II) chloride [2,2':6',2''-terpyridin]-4'-ylmethyl ethyl (4-nitrophenyl) phosphate (4-Mn): Manganese chloride tetrahydrate (100.5 mg, 0.508 mmol) was dissolved in 10 mL dry THF after sonication. This solution was added to a stirring solution of **4** (50 mg, 0.102 mmol) dissolved in 2 mL dry THF in a 20 mL scintillation vial. A fine yellow precipitate formed immediately upon addition. The yellow solid was collected on a fine frit and washed with copious THF to afford 35.7 mg of the manganese complex **4-Mn** (0.062 mmol, 61% yield)

1,5-di(pyridin-2-yl)pentane-1,3,5-trione (5) was prepared according to literature procedure.[54]

[2,2':6',2''-terpyridin]-4'-ol (6): Trione **5** (958.7 mg, 3.57 mmol) and ammonium acetate (1.928 g, 25.0 mmol) were weighed into a 100 mL flask and dissolved in 25 mL EtOH. The reaction flask was fitted with a reflux condenser and refluxed for 72 hours. The reaction mixture was cooled to room temperature, concentrated to ~12.5 mL, cooled to 0°C and the white precipitate was collected by filtration. The solid was washed with Et_2O. 90% yield. (800 mg, 3.21 mmol). ESI-MS calculated for $C_{15}H_{12}N_3O$ $[M+H]^+$=250.10, found 250.1.

^1H NMR (500 MHz; $CDCl_3$): δ 8.72 (d, *J* = 4.0 Hz, 2H), 8.63 (d, *J* = 8.0 Hz, 2H), 8.40 (s, 2H), 7.90 (t, *J* = 8.1 Hz, 2H), 7.39 (t, *J* = 5.7 Hz, 2H).

[2,2':6',2''-terpyridin]-4'-yl trifluoromethanesulfonate (7): [2,2':6',2''-terpyridin]-4'-ol **6** (550 mg, 2.21 mmol) was weighed into a 25 mL round bottom flask equipped with a stir bar and evacuated and purged 3 times with N_2. The solid was dissolved in 2 mL pyridine and cooled to 0°C in an ice bath. Triflic anhydride (0.370 mL, 2.21 mmol) was added dropwise to the stirred mixture. After full addition, the reaction mixture was warmed to room temperature and stirred for 12 hours. Ice was added to the reaction mixture causing a white solid to precipitate out, which was filtered and washed with cold H_2O. 70% yield (590 mg, 1.55 mmol). ESI-MS calculated for $C_{16}H_{11}F_3N_3O_3S$ $[M+H]^+$=382.05, found 382.1.

^1H NMR (500 MHz; $CDCl_3$): δ 8.73 (d, *J* = 3.8 Hz, 2H), 8.63 (d, *J* = 7.9 Hz, 2H), 8.43 (s, 2H), 7.90 (t, *J* = 7.7 Hz, 2H), 7.40 (dd, *J* = 6.3, 4.8 Hz, 2H).

[2,2':6',2''-terpyridin]-4'-yl bromide (8): [2,2':6',2''-terpyridin]-4'-yl trifluoromethanesulfonate **7** (282.2 mg, 0.740 mmol) was weighed into a 25 mL round bottom flask equipped with a stir bar. The solid was dissolved in 5 mL aqueous HBr (48%) and 6.4 mL glacial acetic acid. The reaction flask was fitted with a reflux condenser and the reaction was refluxed for 5 hours. The reaction mixture was then cooled to room temperature and transferred to a separatory funnel. Neutralization with saturated sodium bicarbonate yielded a white precipitate. The aqueous suspension was extracted into CH_2Cl_2 (3 x 100 mL), the organic fractions were combined, dried over Na_2SO_4, and concentrated onto silica. The product was purified by silica flash chromatography (10% EtOAc/Hexanes) to afford **8** as a brown/white solid. 97% yield. (224.4 mg, 0.719 mmol) ESI-MS calculated for $C_{15}H_{11}BrN_3$ $[M+H]^+$=312.01, found 312.0.

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.72 (d, *J* = 4.3 Hz, 2H), 8.67 (s, 2H), 8.60 (d, *J* = 7.8 Hz, 2H), 7.89 (t, *J* = 7.6 Hz, 2H), 7.38 (t, *J* = 6.0 Hz, 2H).

Diethyl [2,2':6',2''-terpyridin]-4'-ylphosphonate (9): Into a dry 25 mL round bottom flask under positive flow of N_2 was weighed $Pd(PPh)_4$ (9.2 mg, 0.008 mmol), [2,2':6',2''-terpyridin]-4'-yl bromide **8** (50 mg, 0.160 mmol). The reaction components were dissolved in 0.5 mL dry toluene and 44.6 μL (0.320 mmol) triethylamine was added. Diethyl phosphite (41.2 μL, 0.320 mmol) was added and the mixture was heated to 110°C and stirred for 12 hours. The reaction mixture was cooled to room temperature and concentrated onto a celite pad. The crude mixture was then purified by reverse phase

chromatography using a Biotage C-18 12 g SNAP cartridge. 53% yield. (31 mg, 0.085 mmol) ESI-MS calculated for $C_{19}H_{21}N_3O_3P$ $[M+H]^+$=370.13, found 370.1.

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.85 (d, *J* = 13.9 Hz, 2H), 8.73 (d, *J* = 4.1 Hz, 2H), 8.61 (d, *J* = 7.9 Hz, 2H), 7.87 (t, *J* = 7.7 Hz, 2H), 7.36 (t, *J* = 6.0 Hz, 2H), 4.28-4.17 (m, 4H), 1.38 (t, *J* = 7.1 Hz, 6H).

^{31}P NMR (202 MHz; $CDCl_3$): δ 15.36.

Ethyl (4-nitrophenyl) [2,2':6',2''-terpyridin]-4'-ylphosphonate (10): Phosphonate **9** (68.3 mg, 0.185 mmol) was weighed into a 25 mL round bottom flask equipped with a stir bar, fitted with a reflux condenser and a rubber septum. The setup was evacuated and purged with 3 times with N_2. 2 mL $POCl_3$ was added to the reaction flask and this was refluxed for 24 hours. The reaction was cooled and excess $POCl_3$ was thoroughly removed *in vacuo* on a high vacuum line. 4-nitrophenol (257.2 mg, 1.849 mmol) was added to the residue under positive flow of N_2. 2 mL dry CH_2Cl_2 was added to the reaction mixture, which was subsequently cooled to 0°C in an ice bath and stirred for 5 minutes. Dry Et_3N (258 μL, 1.849 mmol) was added dropwise to the stirred reaction mixture. After full addition, the reaction mixture was stirred at room temperature for 1 hour. The reaction was then concentrated thoroughly to remove volatiles. The residue was redissolved in 6 mL 1:1 EtOH/THF and stirred for 30 minutes. The crude mixture was concentrated onto a celite pad and purified by reverse phase chromatography using a Biotage C-18 12 g SNAP cartridge. 64% yield. (54.5 mg, 0.118 mmol). ESI-MS calculated for $C_{23}H_{20}N_4O_5P$ $[M+H]^+$=463.12, found 463.1.

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.92 (d, *J* = 14.6 Hz, 2H), 8.74 (d, *J* = 4.4 Hz, 2H), 8.62 (d, *J* = 7.9 Hz, 2H), 8.21 (d, *J* = 9.1 Hz, 2H), 7.89 (td, *J* = 7.7, 1.6 Hz, 2H), 7.44 (d, *J* = 8.6 Hz, 2H), 7.39 (dd, *J* = 6.9, 5.5 Hz, 2H), 4.26 (ddd, *J* = 8.3, 7.2, 2.6 Hz, 2H), 1.43 (t, *J* = 7.1 Hz, 3H).

^{31}P NMR (202 MHz; CDCl3): δ 12.58.

Manganese(II) chloride ethyl (4-nitrophenyl) [2,2':6',2''-terpyridin]-4'-ylphosphonate (10-Mn): Manganese chloride tetrahydrate (87.2 mg, 0.441 mmol) was dissolved in 4 mL dry THF after sonication. This solution was added to a stirring solution of **10** (38.2 mg, 0.088 mmol) dissolved in 450 µL dry THF in a 20 mL scintillation vial. A fine yellow precipitate formed immediately upon addition. The reddish solid was collected on a fine frit and washed with copious THF to afford 35 mg of the manganese complex **10-Mn** (0.061 mmol, 70% yield). The product turns brown over time.

4'-(bromomethyl)-2,2':6',2''-terpyridine (11): [2,2':6',2"-terpyridin]-4'-ylmethanol **3** (531 mg, 2.01 mmol) was weighed into a 25 mL round bottom flask equipped with a stir bar and reflux condenser. The solid was dissolved in 20 mL HBr and 800 µL H_2SO_4. The reaction mixture was stirred at reflux overnight. The reaction mixture was then cooled to room temperature and transferred to a separatory funnel. The mixture was neutralized with saturated $NaHCO_3$ yielding a copious tan precipitate. The aqueous phase was extracted with CH_2Cl_2 (3 x 50 mL) and ethyl acetate (3 x 50 mL). The organic layers were combined, dried over Na_2SO_4, and concentrated to yield **11** as a pink/gray solid in

85% yield (556.9 mg, 1.71 mmol). ESI-MS calculated for $C_{16}H_{13}BrN_3$ $[M+H]^+$=326.03, found 326.0.

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.73 (d, *J* = 4.3 Hz, 2H), 8.63 (d, *J* = 7.9 Hz, 2H), 8.51 (s, 2H), 7.89 (t, *J* = 7.3 Hz, 2H), 7.37 (dd, *J* = 6.6, 5.2 Hz, 2H), 4.58 (s, 2H).

Diethyl ([2,2':6',2''-terpyridin]-4'-ylmethyl)phosphonate (12): 4'-(bromomethyl)-2,2':6',2''-terpyridine **11** (326.19 mg, 0.690 mmol) was weighed into a dry 25 mL round bottom flask equipped with a stir bar and a rubber septum. The setup was evacuated and purged with 3 times with N_2. The solid was dissolved in 7 mL triethyl phosphite (40.8 mmol) and heated to 150°C for 5 hours. The reaction mixture was subsequently cooled to room temperature and excess $P(OEt)_3$ was thoroughly removed *in vacuo* to yield product **12** as a pink/white solid. 96% yield (253.5 mg, 0.661 mmol). ESI-MS calculated for $C_{20}H_{23}N_3O_3P$ $[M+H]^+$=384.15, found 384.2.

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.70 (d, *J* = 4.5 Hz, 2H), 8.61 (d, *J* = 7.9 Hz, 2H), 8.43 (s, 2H), 7.85 (t, *J* = 7.7 Hz, 2H), 7.33 (dd, *J* = 6.4, 5.7 Hz, 2H), 4.10 (quintet, *J* = 7.1 Hz, 4H), 3.34 (d, *J* = 22.1 Hz, 2H), 1.29 (t, *J* = 7.0 Hz, 6H).

^{31}P NMR (202 MHz; $CDCl_3$): δ 24.50.

Ethyl (4-nitrophenyl) ([2,2':6',2''-terpyridin]-4'-ylmethyl)phosphonate (13): Diethyl ([2,2':6',2''-terpyridin]-4'-ylmethyl)phosphonate **12** (253.5 mg, 0.661 mmol) was weighed into a 25 mL round bottom flask equipped with a stir bar, reflux condenser and sealed with a rubber septum. The setup was evacuated and purged with 3 times with N_2. The solid was suspended in 6.6 mL $POCl_3$ and heated to 70°C for 5 hours. The reaction

mixture was cooled to room temperature and excess $POCl_3$ was removed *in vacuo*. 4-nitrophenol (460 mg, 3.31 mmol) was added to the residue under positive flow of N_2. The solids were resuspended in 6.6 mL dry CH_2Cl_2 and the reaction mixture was cooled to 0°C in an ice bath. Triethylamine (0.922 mL, 6.61 mmol) was added dropwise to the stirred reaction mixture. After full addition, the reaction was stirred at room temperature for 2 hours. The reaction mixture was concentrated to remove volatiles. The crude residue was redissolved in CH_2Cl_2 and concentrated onto a celite pad. This was purified by reverse phase chromatography using a Biotage C-18 12 g SNAP cartridge. 87% yield. (274 mg, 0.575 mmol) ESI-MS calculated for $C_{24}H_{21}N_4O_5P$ $[M+H]^+$=477.13, found 477.1.

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.69 (d, *J* = 4.6 Hz, 2H), 8.61 (d, *J* = 8.0 Hz, 2H), 8.43 (s, 2H), 8.15 (t, *J* = 8.8 Hz, 2H), 7.87 (t, *J* = 7.6 Hz, 2H), 7.36 (d, *J* = 8.2 Hz, 4H), 4.29-4.17 (m, 2H), 3.55 (d, *J* = 22.3 Hz, 2H), 1.30 (t, *J* = 7.0 Hz, 3H).

^{31}P-NMR (202 MHz; $CDCl_3$): δ 21.76.

Manganese (II) chloride 1-(2-([2,2':6',2''-terpyridin]-4'-yloxy)ethyl)-1H-pyrrole-2,5-dione (14) was synthesized according to the procedure indicated in the literature.[33]

4-bromo-*N1*,*N2*-dimethylbenzene-1,2-diamine (15): A dry 50 mL Schlenk flask equipped with a stir bar was charged with 4-bromobenzene-1,2-diamine (0.750 g, 4.009 mmol). The flask was evacuated and purged with 3 times with N_2. The solid was dissolved in 8 mL CH_2Cl_2 and 0.989 mL (12.029 mmol) pyridine was added to the solution. The reaction mixture was cooled in an ice bath and ethyl chloroformate was

added dropwise to the reaction mixture. The reaction was stirred for 5 hours at room temperature. The reaction was then washed with H_2O (50 mL) and brine (50 mL) and the organic layer was dried over $MgSO_4$. This was concentrated to afford a white solid. The residue was redissolved in 15 mL dry Et_2O and the solution transferred to a dry Schlenk flask containing a stir bar. $LiAlH_4$ (700 mg, 18.44 mmol) was added slowly to the reaction mixture and the reaction was stirred overnight at room temperature. 800 µL H_2O was slowly added to quench the reaction, followed by 800 µL 15% NaOH, followed by 2.4 mL H_2O. The mixture was stirred for one hour, producing a chunky white precipitate. The resultant mixture was filtered and the filtrate was concentrated to yield **15** as a brown oil in 88% yield. (758.8 mg, 3.53 mmol). ESI-MS calculated for $C_8H_{12}BrN_2$ $[M+H]^+$=215.02, found 215.0

^{1}H NMR (500 MHz; $CDCl_3$): δ 7.21 (d, *J* = 7.6 Hz, 1H), 7.02 (s, 1H), 6.98 (d, *J* = 7.6 Hz, 1H), 3.43 (s, 1H), 3.43 (s, 1H), 2.97 (s, 3H), 2.90 (s, 3H).

4-bromo-*N*^{1},*N*^{2}-dimethyl-*N*^{1},*N*^{2}-bis(pyridin-2-ylmethyl)benzene-1,2-diamine (16): 4-bromo-*N1*,*N2*-dimethylbenzene-1,2-diamine **15** (400 mg, 1.860 mmol) was weighed into a 50 mL Schlenk flask containing a stir bar and fitted with a rubber septum. The solid was dissolved in 4 mL DMF and the solution was cooled to 0°C in an ice bath. NaH (60% suspension in mineral oil) (446 mg, 11.17 mmol) was added slowly to the mixture. After full addition, the mixture was stirred at 0°C for 15 minutes. 2-(chloromethyl)pyridine (915 mg, 5.58 mmol) was added to the reaction mixture as a solution in 1:1 DMF:Et_3N. The reaction was gradually warmed to room temperature and stirred for 48 hours. The reaction mixture was then poured into 60 mL 1 M NaOH and the

aqueous phase was extracted with CH_2Cl_2 (3 x 50 mL). The organic layers were combined and dried over Na_2SO_4. The organic fractions were concentrated onto silica and purified by silica flash chromatography (1% Et_3N/EtOAc) yielded product **16** as a red oil in 93% yield. (688 mg, 1.73 mmol). ESI-MS calculated for $C_{20}H_{22}BrN_4$ $[M+H]^+$=397.10, found 397.1

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.56 (dd, *J* = 10.1, 4.6 Hz, 2H), 7.58 (dt, *J* = 15.9, 7.8 Hz, 2H), 7.19 (q, *J* = 6.3 Hz, 2H), 7.11 (dd, *J* = 11.4, 8.0 Hz, 2H), 7.06 (dd, *J* = 4.3, 2.3 Hz, 2H), 6.83 (d, *J* = 9.1 Hz, 1H), 4.66 (s, 2H), 4.59 (s, 2H), 2.82 (s, 3H), 2.80 (s, 3H).

Diethyl (3,4-bis(methyl(pyridin-2-ylmethyl)amino)phenyl)phosphonate (17): Into a dry 50 mL Schlenk flask under positive flow of N_2 was weighed $Pd(OAc)_2$ (2.6 mg, 0.014 mmol), 1,1'-bis(diphenylphosphino)ferrocene (12.7 mg, 0.023 mmol), **16** (200 mg, 0.504 mmol), and potassium acetate (4.5 mg, 0.046 mmol). The reaction components were dissolved in 5 mL dry THF and 77 µL (0.549 mmol) triethylamine was added. The mixture was heated to 68°C and stirred for 15 minutes. Diethyl phosphite (59 µL, 0.458 mmol) was added dropwise to the pre-equilibrated mixture and the reaction was stirred overnight. The reaction mixture was cooled to room temperature and concentrated onto a silica. The crude mixture was then purified by silica flash chromatography (1% MeOH/1% Et_3N/98% CH_2Cl_2 to afford **17** as a reddish/brown oil. 91% yield. (209.1 mg, 0.460 mmol). ESI-MS calculated for $C_{24}H_{32}N_4O_3P$ $[M+H]^+$=455.22, found 455.2.

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.52 (dd, *J* = 9.2, 4.6 Hz, 2H), 7.48 (dtd, *J* = 18.6, 7.7, 1.5 Hz, 2H), 7.38 (ddd, *J* = 12.7, 8.2, 1.5 Hz, 2H), 7.32 (dd, *J* = 14.2, 1.4 Hz, 1H), 7.12

(dt, J = 10.7, 5.6 Hz, 1H), 7.01 (dd, J = 16.9, 7.9 Hz, 2H), 6.97 (dd, J = 8.2, 4.8 Hz, 1H), 4.74 (s, 2H), 4.53 (s, 2H), 4.10-3.98 (m, 4H), 2.89 (s, 3H), 2.80 (s, 3H).

^{31}P-NMR (202 MHz; CDCl3): δ 20.45.

Ethyl (4-nitrophenyl) (3,4-bis(methyl(pyridin-2-ylmethyl)amino)phenyl)phosphonate (18): Diethyl (3,4-bis(methyl(pyridin-2-ylmethyl)amino)phenyl)phosphonate **17** (100 mg, 0.220 mmol) was weighed into a 25 mL round bottom flask equipped with a stir bar, reflux condenser and sealed with a rubber septum. The setup was evacuated and purged with 3 times with N_2. The solid was suspended in 4 mL (42.9 mmol) $POCl_3$ and heated to 70°C for 18 hours. The reaction mixture was cooled to room temperature and excess $POCl_3$ was removed *in vacuo*. 4-nitrophenol (76.5 mg, 0.550 mmol) was added to the residue under positive flow of N_2. The solids were resuspended in 8 mL dry CH_2Cl_2 and the reaction mixture was cooled to 0°C in an ice bath. Triethylamine (153 μL, 1.10 mmol) was added dropwise to the stirred reaction mixture. After full addition, the reaction was stirred at room temperature for 2 hours. 2 mL EtOH was added to the mixture and it was stirred for an additional 2 hours. The reaction mixture was concentrated to remove volatiles. The crude residue was redissolved in CH_2Cl_2 and concentrated onto silica. The product was purified on silica (100% CH_2Cl_2 → 2% MeOH/CH_2Cl_2 →10% MeOH/CH_2Cl_2) 56% yield. (65.4 mg, 0.119 mmol) ESI-MS calculated for $C_{28}H_{31}N_5O_5P$ $[M+H]^+$=548.21, found 548.2.

^{1}H NMR (500 MHz; CDCl3): δ 8.55 (d, J = 4.5 Hz, 1H), 8.51 (d, J = 4.5 Hz, 1H), 8.17 (s, 1H), 8.15 (s, 1H), 7.60 (d, J = 0.7 Hz, 1H), 7.52 (d, J = 0.5 Hz, 1H), 7.43 (dd, J = 12.6, 8.9 Hz, 1H), 7.36 (d, J = 14.8 Hz, 1H), 7.28 (d, J = 8.9 Hz, 2H), 7.20 (d, J = 23.0

Hz, 2H), 7.07 (d, *J* = 7.1 Hz, 1H), 7.02 (d, *J* = 7.2 Hz, 1H), 6.98 (dd, *J* = 8.2, 5.2 Hz, 1H), 4.86 (s, 2H), 4.58 (s, 2H), 4.27-4.19 (m, 2H), 2.93 (s, 3H), 2.81 (s, 3H), 1.36 (t, *J* = 7.1 Hz, 3H).

^{31}P NMR (202 MHz; $CDCl_3$): δ 17.58.

Iron (II) chloride ethyl (4-nitrophenyl) (3,4-bis(methyl(pyridin-2-ylmethyl)amino)phenyl)phosphonate (18-Fe): Under N_2 atmosphere, ethyl (4-nitrophenyl) (3,4-bis(methyl(pyridin-2-ylmethyl)amino)phenyl)phosphonate **18** (36.7 mg, 0.067 mmol) was weighed into a 5 mL conical vial containing a spin vane. The solid was dissolved in 0.395 mL dry CH_3CN and 8.5 mg (0.067 mmol) $FeCl_2$ was added to the reaction mixture. The mixture was stirred for 24 hours and 3 mL dry Et_2O was added, generating copious yellow precipitate. The suspension was filtered on a fine frit to yield a fine yellow powder. This solid was washed with copious Et_2O and collected. 93% yield. (42 mg 0.062 mmol).

Iron (II) hexafluoroantimonate ethyl (4-nitrophenyl) (3,4-bis(methyl(pyridin-2-ylmethyl)amino)phenyl)phosphonate (18-FeSbF$_6$): Complex **18-Fe** (42 mg, 0.062 mmol) was weighed into a foil-wrapped 5 mL conical vial containing a spin vane. The solid was suspended in 780 μL dry CH_3CN and $AgSbF_6$ (42.8 mg, 0.125 mmol) was added to the suspension yielding an immediate gray precipitate. The reaction mixture was stirred at room temperature for 24 hours and filtered through celite and washed with CH_3CN. The filtrate was blue/violet. This was concentrated and redissolved in CH_3CN. The solution was filtered through a 0.22 um PTFE syringe filter and concentrated again.

This was repeated twice more. **18-$FeSbF_6$** was obtained as a dark blue solid. 86% yield. (57.3 mg, 0.495 mmol)

Diethyl (hydroxymethyl)phosphonate (19): 1.50 g (50 mmol) paraformaldehyde was weighed into a 25 mL round bottom flask equipped with a stir bar and a reflux condenser. Diethyl phosphite (6.44 mL, 50 mmol) and triethylamine (0.697 mL, 5 mmol) was added to the mixture and the reaction was heated to 130°C for 4 hours. The reaction mixture was cooled to room temperature and excess Et_3N was removed *in vacuo*. The crude product was distilled by Kügelrohr at 0.05 mmHg, 125°C to yield product **19** as clear, colorless oil. 59% yield. (4.951 g, 29.4 mmol). ESI-MS calculated for $C_5H_{14}O_4P$ $[M+H]^+$=169.06, found 169.1.

^{1}H NMR (500 MHz; $CDCl_3$): δ 4.17 (dtt, *J* = 10.9, 7.2, 3.7 Hz, 4H), 3.90 (d, *J* = 6.1 Hz, 2H), 1.34 (t, *J* = 7.1 Hz, 6H).

^{31}P NMR (202 MHz; CDCl3): δ 24.10.

(Diethoxyphosphoryl)methyl trifluoromethanesulfonate (20): A dry 100 mL Schlenk flask equipped with a stir bar was charged with 2.00 g (11.896 mmol) diethyl (hydroxymethyl)phosphonate **19** and 1.663 mL (14.275 mmol) 2,6-lutidine. The reagents were dissolved in 20 mL dry CH_2Cl_2 and the reaction mixture was cooled to -78°C in a dry ice/acetone bath. Freshly distilled trifluoromethanesulfonic anhydride (2.302 mL, 13.680 mmol) was added to the reaction dropwise, resulting in an immediate red/orange color. The reaction was gradually warmed to 0°C over 3 hours. A brown precipitate was formed over that time. The reaction mixture was filtered and the solid was washed with

Et_2O. The filtrate was washed with 1 M HCl (3 x 20 mL), dried over $MgSO_4$, and concentrated to yield product **20** as an orange oil. 87% yield. (3.1197 g, 10.4 mmol). The product should be stored at low temperature to prevent decomposition. ESI-MS calculated for $C_6H_{13}F_3O_6PS$ $[M+H]^+$=301.01, found 301.0.

^{1}H NMR (500 MHz; $CDCl_3$): δ 4.61 (d, *J* = 8.8 Hz, 2H), 4.28-4.22 (m, 4H), 1.39 (t, *J* = 7.1 Hz, 6H).

^{31}P-NMR (202 MHz; CDCl3): δ 12.22.

***N*^{1},*N*^{2}-bis(pyridin-2-ylmethyl)ethane-1,2-diamine (21):** A 100 mL round bottom flask equipped with a stir bar and a reflux condenser was charged with 1,2-ethylenediamine (669 μL, 10.0 mmol) and 5 mL dry MeOH. 2-pyridinecarboxaldehyde (1.902 mL, 20.0 mmol) was added to the solution and the mixture was refluxed for 1 hour. The reaction was cooled to room temperature and placed in an ice bath. $NaBH_4$ (2.017 g, 53.3 mmol) was added slowly in small portions to control the resulting exotherm. The reaction mixture was then heated to reflux overnight. The reaction was cooled back to room temperature and 1M HCl was added in order to quench excess $NaBH_4$. The mixture was then concentrated to remove excess MeOH. The resulting aqueous solution was transferred to a separatory funnel and neutralized with saturated $NaHCO_3$. The aqueous layer was extracted with CH_2Cl_2 (3 x 100 mL), dried over Na_2SO_4, and concentrated to yield product **21** as a brown/orange oil. 95% yield. (2.3193 g, 9.54 mmol). ESI-MS calculated for $C_{14}H_{19}N_4$ $[M+H]^+$=243.16, found 243.2.

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.54 (d, *J* = 4.5 Hz, 2H), 7.63 (td, *J* = 7.6, 1.3 Hz, 2H), 7.32 (d, *J* = 7.7 Hz, 2H), 7.15 (t, *J* = 6.1 Hz, 2H), 3.93 (s, 4H), 2.84 (s, 4H).

N^1-methyl-N^1,N^2-bis(pyridin-2-ylmethyl)ethane-1,2-diamine (22): 2.00 g (8.254 mmol) **21** was weighed into a 50 mL round bottom flask equipped with a stir bar and a rubber septum. The oil was dissolved in 42 mL MeOH and formaldehyde (37% in H_2O) (16.304 mL, 201 mmol) was added to the solution. The mixture was stirred at room temperature for 12 hours. The reaction mixture was concentrated under high vacuum to remove excess formaldehyde. The crude imidazolidine intermediate was redissolved in 42 mL MeOH. 622 mg (9.904 mmol) $NaBH_3CN$ was added to the reaction mixture and CF_3COOH (1.263 mL, 16.507 mmol) was added dropwise, producing a mild exotherm. The reaction was stirred for 12 hours at room temperature, after which 70 mL 4 M NaOH was added to quench excess $NaBH_3CN$. The mixture was stirred for another hour at room temperature and then concentrated to remove volatiles. The resulting aqueous solution was extracted with CH_2Cl_2 (3 x 50 mL). The organic fractions were dried over Na_2SO_4 and concentrated onto silica. The crude product was purified by silica flash chromatography (2% MeOH/1% Et_3N/CH_2Cl_2). 39% yield. 834.3 mg, 3.26 mmol. ESI-MS calculated for $C_{15}H_{21}N_4$ $[M+H]^+$=257.18, found 257.2.

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.53 (t, *J* = 4.6 Hz, 2H), 7.64 (qd, *J* = 7.6, 1.3 Hz, 3H), 7.46 (d, *J* = 7.7 Hz, 1H), 7.40 (d, *J* = 7.8 Hz, 1H), 7.31 (d, *J* = 7.8 Hz, 1H), 7.15 (t, *J* = 6.0 Hz, 2H), 7.15 (t, *J* = 6.0 Hz, 2H), 3.90 (s, 2H), 3.68 (s, 2H), 3.66 (s, 2H), 2.64 (s, 2H), 2.27 (s, 3H).

Diethyl (((2-(methyl(pyridin-2-ylmethyl)amino)ethyl)(pyridin-2-ylmethyl)amino)methyl)phosphonate (23): A 50 mL Schlenk flask containing a stir bar

was charged with 400 mg (1.560 mmol) **22** and K_2CO_3 (862.6 mg, 6.241 mmol). After evacuation and purging 3 times with N_2, 7.5 mL dry CH_3CN was added to the reagents, followed by the dropwise addition of (diethoxyphosphoryl)methyl trifluoromethanesulfonate **20** (468.4 mg, 1.560 mmol) as a solution in 7.5 mL CH_3CN. The reaction mixture was stirred at room temperature overnight, after which it was concentrated, redissolved in 50 mL CH_2Cl_2, and washed with 1 M HCl (50 mL). The aqueous layer was neutralized with $NaHCO_3$ and extracted with CH_2Cl_2 (4 x 50 mL). The organic layers were combined, dried over Na_2SO_4, and concentrated. Product **23** was obtained as a brown/yellow oil. 96% yield. (608 mg, 1.50 mmol) ESI-MS calculated for $C_{20}H_{32}N_4O_3P$ $[M+H]^+$=407.22, found 407.2.

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.53 (t, *J* = 4.4 Hz, 2H), 7.64 (qd, *J* = 8.4, 1.5 Hz, 2H), 7.54 (d, *J* = 7.8 Hz, 1H), 7.41 (d, *J* = 7.8 Hz, 1H), 7.18-7.15 (m, 2H), 4.13 (quintet, *J* = 7.2 Hz, 4H), 3.98 (s, 2H), 3.67 (s, 2H), 3.08 (d, *J* = 10.2 Hz, 2H), 2.91 (t, *J* = 6.9 Hz, 2H), 2.65 (t, *J* = 6.9 Hz, 2H), 2.26 (s, 3H).

^{31}P NMR (202 MHz; CDCl3): δ 25.43.

Ethyl hydrogen (((2-(methyl(pyridin-2-ylmethyl)amino)ethyl)(pyridin-2-ylmethyl)amino)methyl)phosphonate (24): 23 (608 mg, 1.497 mmol) was weighed into a 25 mL round bottom flask equipped with a stir bar and a reflux condenser. 5.98 g (149.7 mmol) NaOH was added and the mixture was dissolved in 5 mL H_2O and heated to reflux overnight. The reaction was then concentrated to remove H_2O and the white solid was thoroughly extracted with CH_2Cl_2 and filtered. The filtrate was concentrated to

yield the sodium salt of **24** as a yellow waxy oil. 100% yield. 560 mg, 1.494 mmol, ESI-MS calculated for $C_{18}H_{28}N_4O_3P$ $[M+H]^+$=379.19, found 379.2.

Ethyl (4-nitrophenyl) (((2-(methyl(pyridin-2-ylmethyl)amino)ethyl)(pyridin-2-ylmethyl)amino)methyl)phosphonate (25): Phosphonate **24** (558 mg, 1.394 mmol) was weighed into a 25 mL round bottom flask equipped with a stir bar and reflux condenser. To this was added DMAP (85 mg, 0.697 mmol) and 4-nitrophenol (233 mg, 1.672 mmol). The reagents were dissolved in 6 mL dry toluene and 4 mL dry CH_2Cl_2 and the reaction mixture was heated to 65°C for 24 hours. After cooling to room temperature, the reaction was concentrated onto a celite pad and purified by reverse phase chromatography using a Biotage C-18 12 g SNAP cartridge. 10% yield. (70 mg, 0.139 mmol). ESI-MS calculated for $C_{24}H_{31}N_5O_5P$ $[M+H]^+$=500.21, found 500.2.

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.70 (d, *J* = 4.8 Hz, 1H), 8.54 (d, *J* = 4.3 Hz, 1H), 8.18 (d, *J* = 9.0 Hz, 2H), 7.93 (t, *J* = 7.3 Hz, 1H), 7.77 (td, *J* = 7.6, 1.2 Hz, 1H), 7.65 (d, *J* = 7.8 Hz, 1H), 7.48 (d, *J* = 7.8 Hz, 1H), 7.45 (t, *J* = 6.4 Hz, 1H), 7.33 (dd, *J* = 7.5, 5.3 Hz, 1H), 7.30 (d, *J* = 8.8 Hz, 2H), 4.51 (s, 2H), 4.21 (s, 2H), 4.21-4.13 (m, 2H), 3.54 (s, 2H), 3.37 (s, 2H), 2.90 (s, 3H), 1.25 (t, *J* = 7.0 Hz, 3H).

^{31}P NMR (202 MHz; CDCl3): δ 21.28.

4-bromo-N^2,N^6-bis(2-chloroethyl)pyridine-2,6-dicarboxamide (26): Br_2 (688 μL, 13.42 mmol) was added to an N_2-purged 100 mL round bottom flask equipped with a stir bar and a rubber septum. This was dissolved in 5 mL dry hexanes and stirred vigourously. PBr_3 (1.543 mL, 16.3 mmol) was added dropwise to the solution, resulting

in the rapid formation of a yellow solid. After full addition, the mixture was stirred for 1 hour at room temperature. The solid PBr_5 was then washed and decanted with copious hexanes to remove residual Br_2. The solid was then dried *in vacuo* for 1 hour on a high vacuum line. Chelidamic acid monohydrate (1.00 g, 4.972 mmol) was added to this and the solids were mixed to homogeneity. The solid mixture was then heated to 90°C, producing a melt, which was stirred for an additional 5 hours. The reaction mixture was cooled to room temperature and 3 mL dry $CHCl_3$ was added to the mixture. This was stirred for 30 minutes. The suspension was filtered and the remaining solid was washed with dry $CHCl_3$. The filtrate was concentrated leaving a crystalline acyl bromide derivative of chelidamic acid. In a separate 25 mL round bottom flask equipped with a stir bar and a rubber septum, 2-chloroethylamine hydrochloride (634.4 mg, 5.47 mmol) was dissolved in 12.5 mL $CHCl_3$ and 5 mL Et_3N. This solution was cooled to 0°C. The acyl bromide was dissolved in 12.5 mL $CHCl_3$ and the solution was added dropwise to the 2-chloroethylamine solution. The reaction mixture was then warmed to room temperature and stirred for 5 hours. After 5 hours, volatiles were removed *in vacuo* and the crude mixture was purified by silica flash chromatography (60% EtOAc/Hexanes) to afford **26** as a white solid. 54% yield. (990 mg, 2.68 mmol). ESI-MS calculated for $C_{11}H_{13}BrCl_2N_3O_2$ $[M+H]^+$=367.96, found 368.0.

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.55 (s, 2H), 8.08 (s, 2H), 3.90 (q, J = 5.7 Hz, 4H), 3.79 (t, J = 5.5 Hz, 4H).

2,2'-(4-bromopyridine-2,6-diyl)bis(4,5-dihydrooxazole) (27): 4-bromo-N^2,N^6-bis(2-chloroethyl)pyridine-2,6-dicarboxamide **26** (90 mg, 0.244 mmol) and sodium hydride

(60% suspension in mineral oil) (39 mg, 0.976 mmol) were added to a dry 50 mL round bottom flask equipped with a stir bar and rubber septum. The flask was evacuated and purged 3 times with N_2. The solids were suspended in 10 mL THF and stirred overnight. The mixture was then filtered on celite and the filtrate was concentrated, affording **27** as a pure white powder. 76% yield. (55 mg, 0.185 mmol). ESI-MS calculated for $C_{11}H_{11}BrN_3O_2$ $[M+H]^+$=296.00, found 296.0.

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.34 (s, 2H), 4.54 (t, *J* = 9.7 Hz, 4H), 4.13 (t, *J* = 9.7 Hz, 4H).

Diethyl (2,6-bis(4,5-dihydrooxazol-2-yl)pyridin-4-yl)phosphonate (28): A dry 50 mL Schlenk flask equipped with a stir bar under positive flow of N_2 was charged with pybox **27** (355 mg, 1.199 mmol), $Pd(OAc)_2$ (6.73 mg, 0.030 mmol), 1,1'-bis(diphenylphosphino)ferrocene (33.2 mg, 0.060 mmol), and potassium acetate (11.8 mg, 0.120 mmol). The reaction components were dissolved in 2.5 mL dry THF and 201 μL triethylamine was added. The mixture was heated to 68°C and stirred for 15 minutes. Diethyl phosphite (154.4 μL, 1.199 mmol) was added dropwise to the pre-equilibrated mixture and the reaction was stirred overnight. The reaction mixture was cooled to room temperature and concentrated onto silica. The crude mixture was purified by silica flash chromatography (0.5% Et_3N/2% MeOH/CH_2Cl_2) to afford **28** as a white solid. 83% yield. (350.6 mg, 0.992 mmol). ESI-MS calculated for $C_{15}H_{21}N_3O_5P$ $[M+H]^+$=354.12, found 354.1.

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.50 (d, *J* = 13.3 Hz, 2H), 4.55 (t, *J* = 9.7 Hz, 2H), 4.26-4.19 (m, 4H), 4.15 (t, *J* = 9.6 Hz, 4H), 1.35 (t, *J* = 7.6 Hz, 6H).

^{31}PNMR (202 MHz; $CDCl_3$): δ 12.56.

***tert*-butyl 2-(4-bromophenyl)acetate (29):** To a 50 mL Schlenk flask under positive flow of N_2 and equipped with a stir bar was added 4-bromophenylacetic acid (500 mg, 2.33 mmol), 4-dimethylaminopyridine (227 mg, 1.86 mmol) and *tert*-butanol (667 μL, 6.975 mmol). The reagents were dissolved in 2.33 mL dry CH_2Cl_2 and the mixture was cooled to 0°C in an ice bath. Dicyclohexylcarbodiimide (528.7 mg, 2.56 mmol) was added portionwise to the mixture and the reaction was stirred at room temperature overnight. The reaction mixture was concentrated and the crude product was purified by silica gel chromatography (5% EtOAc/Hexanes) to afford **29** as a clear colorless oil. 63% yield. (403.7 mg, 1.489 mmol). ESI-MS calculated for $C_{12}H_{15}BrO_2$ $[M+H]^+$=270.03, found 270.1.

^{1}H NMR (500 MHz; $CDCl_3$): δ 7.44 (d, J = 6.7 Hz, 2H), 7.14 (d, J = 6.9 Hz, 2H), 3.47 (s, 2H), 1.43 (s, 9H).

***tert*-butyl 2-(4-(diethoxyphosphoryl)phenyl)acetate (30):** *tert*-butyl 2-(4-bromophenyl)acetate **29** (403.7 mg, 1.49 mmol), $Pd(OAc)_2$ (7.6 mg, 0.034 mmol), 1,1'-bis(diphenylphosphino)ferrocene (37.5 mg, 0.068 mmol), and potassium acetate (13.2 mg, 0.135 mmol) were added to a dry 50 mL Schlenk flask equipped with a stir bar under positive flow of N_2. The reaction components were dissolved in 4 mL dry THF and 226.5 μL (1.624 mmol) triethylamine was added. The mixture was heated to 68°C and stirred for 15 minutes. Diethyl phosphite (174 μL, 1.353 mmol) was added dropwise to the pre-equilibrated mixture and the reaction was stirred overnight. The reaction mixture was

cooled to room temperature and concentrated onto silica. The crude mixture was purified by silica flash chromatography (50% EtOAc/Hexanes) to afford **30** as a white solid. 46% yield. (225 mg, 0.685 mmol). ESI-MS calculated for $C_{16}H_{26}O_5P$ $[M+H]^+$=329.15, found 329.2.

^{1}H NMR (500 MHz; $CDCl_3$): δ 7.77 (dd, *J* = 13.1, 7.8 Hz, 2H), 7.37 (dd, *J* = 7.6, 3.7 Hz, 2H), 4.20-4.03 (m, 4H), 3.58 (s, 2H), 1.44 (s, 9H), 1.32 (t, *J* = 7.0 Hz, 6H).

^{31}P-NMR (202 MHz; CDCl3): δ 18.89.

***tert*-butyl 2-(4-(ethoxy(4-nitrophenoxy)phosphoryl)phenyl)acetate (31):** *tert*-butyl 2-(4-(diethoxyphosphoryl)phenyl)acetate **30** (100 mg, 0.305 mmol) was added to a 50 mL round bottom flask equipped with a stir bar, a reflux condenser, and a rubber septum. The setup was evacuated and purged 3 times with N_2. The ester was dissolved in 6 mL $SOCl_2$ and a few drops of dry DMF were added to the solution. The reaction mixture was heated to reflux for 5 hours. The reaction was cooled to room temperature and excess $SOCl_2$ was removed *in vacuo* affording a crude solid. 4-nitrophenol (106 mg, 0.762 mmol) was added to the solid and the mixture was resuspended in 4 mL CH_2Cl_2. The reaction mixture was cooled to 0°C in an ice bath and dry triethylamine (212µL, 1.52 mmol) was added dropwise to the stirring suspension. The reaction was warmed to room temperature and stirred for 2 hours. Volatiles were removed *in vacuo* and the crude product was redissolved in CH_2Cl_2 and concentrated onto silica. Purification by silica flash chromatography (30% EtOAc/Hexanes) afforded product **31** as an off-white solid in 35% yield. (45.5 mg, 0.108 mmol). ESI-MS calculated for $C_{20}H_{25}NO_7P$ $[M+H]^+$=422.14, found 422.1.

^{1}H-NMR (500 MHz; $CDCl_3$): δ 8.20 (d, *J* = 9.1 Hz, 2H), 7.85 (dd, *J* = 13.7, 8.1 Hz, 2H), 7.44 (dd, *J* = 7.9, 4.2 Hz, 2H), 7.36 (d, *J* = 8.7 Hz, 2H), 4.38-4.25 (m, 2H), 3.62 (s, 2H), 1.57 (s, 9H), 1.41 (t, *J* = 7.1 Hz, 3H).

^{31}P-NMR (202 MHz; $CDCl_3$): δ 16.12 (s, 1P).

2-(4-(ethoxy(4-nitrophenoxy)phosphoryl)phenyl)acetic acid (32): *tert*-butyl 2-(4-(ethoxy(4-nitrophenoxy)phosphoryl)phenyl)acetate **31** (45.5 mg, 0.108 mmol) was weighed into a 5 mL conical vial equipped with a spin vane. The compound was dissolved in 2.5 mL dry CH_2Cl_2. Trifluoroacetic acid (377 μL, 4.859 mmol) was added to the mixture and the reaction was stirred at room temperature for 5 hours. Concentration to remove volatiles afforded product **32** as yellow solid in quantitative yield. (39.4 mg, 0.108 mmol). ESI-MS calculated for $C_{16}H_{17}NO_7P$ $[M+H]^+$=366.07, found 366.1.

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.21 (d, *J* = 9.0 Hz, 2H), 7.88 (dd, *J* = 13.6, 8.0 Hz, 2H), 7.48 (dd, *J* = 7.9, 4.2 Hz, 2H), 7.35 (d, *J* = 9.0 Hz, 2H), 4.37-4.28 (m, 2H), 3.76 (s, 2H), 1.41 (t, *J* = 7.1 Hz, 3H).

^{31}P NMR (202 MHz; $CDCl_3$): δ 15.70.

Dirhodium triacetate 2-(4-(ethoxy(4-nitrophenoxy)phosphoryl)phenyl)acetate (33): A 25 mL Schlenk flask equipped with a stir bar was charged with 17.44 mg (0.048 mmol) carboxylic acid **32**, 27.6 mg (0.048 mmol) $Rh_2(OCCH_3)_3(OCCF_3)$, and 13.2 mg (0.095 mmol) anhydrous K_2CO_3. The flask was evacuated and purged with N_2 3 times and the solids were suspended in 2.74 mL dry THF. The reaction mixture was then heated to 50°C for 3 hours. The mixture was then cooled to room temperature and concentrated

onto silica. Purification by silica flash chromatography (30% → 40% MeCN/CH_2Cl_2) afforded product **33** as a bright purple solid. 25% yield. (9.2 mg, 0.012 mmol). ESI-MS calculated for $C_{22}H_{27}NO_{14}PRh_2$ $[M+H_3O]^+$=765.93, found 766.1.

^{1}H NMR (500 MHz; CD_3CN): δ 8.22 (d, *J* = 9.1 Hz, 2H), 7.78 (dd, *J* = 13.5, 8.1 Hz, 2H), 7.41 (d, *J* = 9.1 Hz, 2H), 7.24 (dd, *J* = 8.2, 4.1 Hz, 2H), 4.26 (quintet, *J* = 7.4 Hz, 2H), 3.48 (s, 2H), 1.80 (s, 3H), 1.75 (s, 6H), 1.34 (t, *J* = 7.0 Hz, 3H).

^{31}P NMR (202 MHz; CD_3CN): δ 15.48.

di-*tert*-butyl 3,3'-(5-(bromomethyl)-1,3-phenylene)bis(2,2-dimethylpropanoate) (34): A solution of LDA (2.0 M in THF) (2.982 mL, 5.605 mmol) was added to a dry 50 mL Schlenk flask equipped with a stir bar and cooled to -78°C in a dry ice/acetone bath. *tert*-butyl isobutyrate (1.010 g, 7.007 mmol) was added dropwise as a solution in 3.5 mL THF. The mixture was gradually warmed to room temperature and stirred for 1.5 hours to allow full deprotonation. In a separate 50 mL Schlenk flask equipped with a stir bar, 1,3,5-tri(bromomethyl)benzene (1.00 g, 2.803 mmol) was dissolved in 5 mL dry THF. The deprotonation mixture was cannulated dropwise into the vigorously stirred 1,3,5-tri(bromomethyl)benzene solution. After full addition, the reaction mixture was stirred at room temperature for 16 hours. The reaction was quenched with H_2O and extracted into CH_2Cl_2 (3 x 50 mL). The organic fractions were combined, dried over Na_2SO_4, and concentrated onto silica. Purification by silica flash chromatography (2% EtOAc/Hexanes) afforded **34** as a pale yellow oil. 32% yield. (430 mg, 0.889 mmol). ESI-MS calculated for $C_{25}H_{40}BrO_4$ $[M+H]^+$=483.21, found 483.2.

^{1}H NMR (500 MHz; $CDCl_3$): δ 7.03 (s, 2H), 6.88 (s, 1H), 4.41 (s, 2H), 2.78 (s, 4H), 1.55 (s, 9H), 1.44 (s, 18H), 1.11 (s, 12H).

di-*tert*-butyl 3,3'-(5-((ethoxy(4-nitrophenoxy)phosphoryl)methyl)-1,3-phenylene)bis(2,2-dimethylpropanoate) (35): di-*tert*-butyl 3,3'-(5-(bromomethyl)-1,3-phenylene)bis(2,2-dimethylpropanoate) **34** (550 mg, 1.137 mmol) was weighed into a 25 mL round bottom flask equipped with a stir bar and a reflux condenser. Triethyl phosphite (5 mL) was added and the mixture was refluxed for 12 hours. The reaction was then cooled to room temperature and excess $P(OEt)_3$ was removed *in vacuo*. The residue was then redissolved in $SOCl_2$ (5 mL) along with a few drops of dry DMF. The mixture was refluxed overnight, after which the reaction was cooled and excess thionyl chloride was removed *in vacuo*. 4-nitrophenol (316 mg, 2.274 mmol) was added to the residue and the components were suspended in 5 mL THF (dry). The reaction mixture was cooled to 0°C and dry triethylamine (0.793 mL, 5.688 mmol) was added dropwise. The reaction was stirred at room temperature for 2 hours. After removal of volatiles *in vacuo*, the crude mixture was loaded onto silica and purified by silica flash chromatography (1:4 EtOAc/Hexanes → 1:1 EtOAc/Hexanes). 59% yield. (425 mg, 0.671 mmol). ESI-MS calculated for $C_{33}H_{49}NO_9P$ $[M+H]^+$=634.31, found 634.3.

^{1}H NMR (500 MHz; $CDCl_3$): δ 8.18 (d, J = 9.1 Hz, 2H), 7.21 (d, J = 8.7 Hz, 2H), 6.98 (s, 2H), 6.90 (s, 1H), 4.18-4.10 (m, 2H), 3.30 (d, J = 21.7 Hz, 2H), 2.78 (s, 4H), 1.46 (s, 18H), 1.27 (t, J = 7.1 Hz, 3H), 1.09 (s, 12H).

^{31}P NMR (202 MHz; $CDCl_3$): δ 24.13.

3,3'-(5-((ethoxy(4-nitrophenoxy)phosphoryl)methyl)-1,3-phenylene)bis(2,2-dimethylpropanoic acid) (36): di-*tert*-butyl 3,3'-(5-((ethoxy(4-nitrophenoxy)phosphoryl)methyl)-1,3-phenylene)bis(2,2-dimethylpropanoate) **35** (102.4 mg, 0.161 mmol) was weighed into a 20 mL scintillation vial loaded with a stir bar. The compound was dissolved in 4 mL dry CH_2Cl_2. Trifluoroacetic acid (0.556 mL, 7.259 mmol) was added to the solution and the reaction was stirred for 2 hours. 5 mL toluene was then added and the mixture was concentrated *in vacuo* to afford a waxy yellow solid in quantitative yield (83 mg, 0.161 mmol). ESI-MS calculated for $C_{25}H_{33}NO_9P$ $[M+H]^+$=522.19, found 522.2.

Dirhodium diacetate 3,3'-(5-((ethoxy(4-nitrophenoxy)phosphoryl)methyl)-1,3-phenylene)bis(2,2-dimethylpropanoate) (37): 3,3'-(5-((ethoxy(4-nitrophenoxy)phosphoryl)methyl)-1,3-phenylene)bis(2,2-dimethylpropanoic acid) **36** (75 mg, 0.144 mmol) and $Rh_2(OCCH_3)_4$ (12.7 mg, 0.029 mmol, 0.2 eq) were weighed into a 5 mL conical vial containing a spin vane. 0.5 mL dichloroethane was added to the mixture and the vial was tightly sealed. The reaction was heated to 125°C for 15 minutes, after which it was cooled to room temperature and an additional 12.7 mg portion of $Rh_2(OCCH_3)_4$ was added to the mixture. This was repeated 5 times until a full equivalent of $Rh_2(OCCH_3)_4$ was added. The reaction was stirred at 125°C for an additional 30 minutes, after which it was cooled to room temperature and concentrated onto a celite pad. The mixture was purified by reverse phase chromatography using a Biotage 12 g SNAP C-18 cartridge. The product **37** was isolated as a purple solid. 28% yield. (34 mg, 0.514 mmol). ESI-MS calculated for $C_{29}H_{39}NO_{14}PRh_2$ $[M+H_3O\ ]^+$=862.02, found 862.1.

^{1}H NMR (500 MHz; CD_3CN): δ 8.23 (d, *J* = 9.1 Hz, 2H), 7.29 (d, *J* = 8.8 Hz, 2H), 6.83 (s, 2H), 6.79 (s, 1H), 4.22-4.15 (m, 2H), 3.35 (d, *J* = 21.6 Hz, 2H), 2.58 (s, 4H), 1.80 (s, 6H), 1.28 (t, *J* = 7.1 Hz, 3H), 0.93 (s, 12H).

^{31}P NMR (202 MHz; CD_3CN): δ 24.50.

Dimethyl 3,3'-(5-hydroxy-1,3-phenylene)bis(2,2-dimethylpropanoate) (38): A 20 mL scintillation vial loaded with a stir bar was charged with 150 mg (0.510 mmol) 3,3'-(5-hydroxy-1,3-phenylene)bis(2,2-dimethylpropanoic acid)[49] and 2 mL 1:7 MeOH/benzene. To the stirred solution was added a 2.0 M solution of TMS-diazomethane in hexanes (1.02 mmol) and the reaction was stirred at room temperature for 15 minutes. The compound was concentrated and passed through a pad of silica eluting with 2:3 EtOAc:Hexanes. Diester **38** was obtained as a clear oil in quantitative yield (164 mg, 0.510 mmol). ESI-MS calculated for $C_{18}H_{27}O_5$ $[M+H]^+$=323.19, found 323.2.

Dimethyl 3,3'-(5-(((trifluoromethyl)sulfonyl)oxy)-1,3-phenylene)bis(2,2-dimethylpropanoate) (39): Phenol **38** (199 mg, 0.617 mmol) was weighed into a 25 mL round bottom flask equipped with a stir bar and a rubber septum. The flask was evacuated and purged 3 times with N_2. The phenol was dissolved in 3 mL dry CH_2Cl_2, after which dry pyridine (0.099 mL, 1.23 mmol) was added. The reaction mixture was cooled to 0°C and trifluoromethanesulfonic anhydride (0.155 mL, 0.926 mmol) was added dropwise. The reaction was warmed to room temperature and stirred for 5 hours. 10 mL H_2O was added to quench the reaction, and the aqueous layer was extracted with CH_2Cl_2 (3 x 30 mL). The organic layers were combined, dried over Na_2SO_4, and concentrated to yield **39**

as a white crystalline solid. Quantitative yield. (280 mg, 0.617 mmol). ESI-MS calculated for $C_{19}H_{26}F_3O_7S$ $[M+H]^+$=455.14, found 455.2.

Dimethyl 3,3'-(5-(3-((dimethylamino)(ethoxy)phosphoryl)propyl)-1,3-phenylene)bis(2,2-dimethylpropanoate) (40): Ethyl *P*-allyl-*N*,*N*-dimethylphosphonamidate[29] (78 mg, 0.440 mmol) was weighed into a 25 mL round bottom flask equipped with a stir bar, a reflux condenser, and a rubber septum. The flask was evacuated and purged 3 times with N_2. The compound was dissolved in 1 mL dry THF and a 0.5 M solution of 9-BBN (1.76 mL, 0.880 mmol) was added. The mixture was refluxed for 3 hours, after which it was cooled to room temperature and 2 mL dry THF was added, along with 200 mg (0.440 mmol) triflate **39**, 32.3 mg (0.040 mmol) $Pd(dppf)Cl_2.CH_2Cl_2$, and 328.82 mg (1.32 mmol) $CsCO_3$. This yielded a violet solution. 0.5 mL N_2-sparged H_2O was added to the mixture and it was heated to reflux for 12 hours. It was then cooled to room temperature and concentrated onto silica. The crude mixture was purified by silica flash chromatography (100% EtOAc → 100% CH_2Cl_2 → 2% MeOH/CH_2Cl_2). 218.2 mg **40** obtained as a red oil. 100% yield. ESI-MS calculated for $C_{25}H_{43}NO_6P$ $[M+H]^+$=484.28, found 484.3.

^{1}H NMR (500 MHz; $CDCl_3$): δ 6.75 (s, 2H), 6.67 (s, 1H), 4.02 (ddt, *J* = 14.3, 10.3, 7.1 Hz, 2H), 3.89-3.80 (m, 2H), 3.66 (s, 4H), 2.64 (d, *J* = 8.9 Hz, 6H), 2.63-2.54 (m, 2H), 1.27 (t, *J* = 7.0 Hz, 3H), 1.15 (s, 12H).

^{31}P NMR (202 MHz; $CDCl_3$): δ 36.61.

3,3'-(5-(3-((dimethylamino)(ethoxy)phosphoryl)propyl)-1,3-phenylene)bis(2,2-dimethylpropanoic acid) (41): Into a 50 mL round bottom flask equipped with a stir bar and reflux condenser was weighed 308.1 mg (0.637 mmol) diester **40**. This compound was dissolved in 10 mL MeOH and 7 mL 1 M NaOH. The solution was heated to reflux for 2 hours. After this it was cooled to room temperature, transferred to a separatory funnel and diluted with 50 mL H_2O. The aqueous layer was washed with CH_2Cl_2 (2 x 20 mL) and then acidified to pH 3.0, causing precipitation of product. This was extracted into CH_2Cl_2 (3 x 50 mL). The organic layers were combined, dried over Na_2SO_4, and concentrated to yield **41** as a red oil that solidified into a yellow solid. 88% yield. (230 mg, 0.561 mmol). ESI-MS calculated for $C_{23}H_{39}NO_6P$ $[M+H]^+$=456.25, found 456.3. ^{1}H NMR (500 MHz; $CDCl_3$): δ 6.92 (s, 2H), 6.80 (s, 1H), 4.06-4.00 (m, 1H), 3.95-3.88 (m, 1H), 2.79 (dq, *J* = 21.1, 10.1 Hz, 2H), 2.70-2.61 (m, 6H), 2.17 (s, 4H), 1.99 (dq, *J* = 15.1, 7.5 Hz, 2H), 1.65-1.50 (m, 2H), 1.30 (t, *J* = 7.1 Hz, 3H), 1.21 (d, *J* = 10.0 Hz, 12H). ^{31}P NMR (202 MHz; $CDCl_3$): δ 39.27.

3,3'-(5-(3-(ethoxy(4-nitrophenoxy)phosphoryl)propyl)-1,3-phenylene)bis(2,2-dimethylpropanoic acid) (42): Phosphoramidate **41** (110 mg, 0.241 mmol) was dissolved in 2 mL dry benzene in a 20 mL scintillation vial containing a small stir bar. To this solution was added 0.302 mL (1.21 mmol) 4.0 M HCl in dioxanes. The reaction mixture was stirred for 2 hours, producing a crystalline precipitate. This was filtered off through a glass wool plug and the filtrate was concentrated to remove benzene and excess HCl. 4-nitrophenol (33.5 mg, 0.241 mmol) was added to the residue and the reagents were suspended in 2 mL dry benzene and cooled to 0°C. 33.6 µL (0.241 mmol) dry Et_3N

was added to the reaction and this was stirred for 2 hours at room temperature. The reaction mixture was then concentrated, redissolved in CH_2Cl_2, and concentrated onto a celite pad. The crude mixture was purified by reverse phase chromatography using a Biotage 12 g SNAP C-18 cartridge to afford **42** in 80% yield. (105 mg, 0.193 mmol). ESI-MS calculated for $C_{27}H_{37}NO_9P$ $[M+H]^+$=550.22, found 550.2.

Dirhodium diacetate 3,3'-(5-(3-(ethoxy(4-nitrophenoxy)phosphoryl)propyl)-1,3-phenylene)bis(2,2-dimethylpropanoate) (43): Diacid **42** (20.2 mg, 0.037 mmol) and NaH (60% in mineral oil) (14.6 mg, 0.365 mmol) were weighed into a 5 mL conical vial equipped with a spin vane. The components were suspended in 2 mL dry THF and stirred at 50°C for 1 hour. *cis*-$Rh(OCCH_3)_2(OCCF_3)_2$[49](20.2 mg, 0.037 mmol) was added to the mixture generating a brilliant green solution. The reaction mixture was stirred at 50°C for an additional 10 hours, after which the reaction was cooled to room temperature and filtered. The filtrate was concentrated onto celite and the crude product was purified by reverse phase chromatography using a Biotage 12 g SNAP C-18 cartridge to afford **43** as a purple solid. 44% yield. (14.3 mg, 0.016 mmol). ESI-MS calculated for $C_{31}H_{43}NO_{14}PRh_2$ $[M+H_3O]^+$=890.05, found 890.1.

^{1}H NMR (500 MHz; CD_3CN): δ 8.30 (d, *J* = 9.0 Hz, 2H), 7.28 (d, *J* = 9.0 Hz, 2H), 7.14 (s, 1H), 6.78 (s, 1H), 6.73 (s, 1H), 3.54 (dt, *J* = 14.0, 7.0 Hz, 3H), 3.42-3.24 (m, 2H), 2.99-2.87 (m, 2H), 2.84-2.61 (m, 4H), 1.87-1.75 (m, 6H), 1.37-1.22 (m, 12H), 1.18-1.03 (m, 3H).

^{31}P NMR (202 MHz; CD_3CN): δ 53.14.

Expression of native *Pfu* POP

Chemically-competent Rosetta2(DE3) (Novagen) cells were transformed with plasmid pJS3 obtained from Prof. Harold Schreier (UMBC)[55]. Transformants were allowed to recover in SOC medium (37°C, 1 hour) and the mixture was spread onto agar plates containing ampicillin and chloramphenicol (6.25 g LB powder mix, 4 g agar, 250 ml DDI water, 100 ug/g ampicillin + 25 ug/g chloramphenicol) and the plates were incubated at 37°C for 16 hours. Colonies containing the pJS3 plasmid were obtained and inoculated into 5 mL primary cultures containing 100 ug/g ampicillin + 25 ug/g chloramphenicol in LB media. These were incubated overnight at 37°C with 250 rpm shaking. The following day, the primary cultures were used to inoculate 1 L of LB media containing 100 ug/g ampicillin + 25 ug/g chloramphenicol). These were incubated at 37°C with 250 rpm shaking for ~5 hours until reaching an OD_{600}≈1.0. Overexpression of *Pfu* POP was induced by addition of 1 mM isopropylthiogalactoside. The expression cultures were incubated at 37°C with 250 rpm shaking for an additional 18 hours. Cells were harvested by centrifugation at 3000 rpm for 10 minutes. They were resuspended in 100 mL 20 mM sodium phosphate buffer pH 6.5 and lysed by sonication. The cell lysate was clarified at 15000 rpm for 1 hour. Clarified lysate was applied to a 3 x 5 mL HiTrap QFF anion exchange columns (GE Healthcare Life Sciences) pre-equilibrated with 20 mM sodium phosphate buffer pH 6.5. The protein was purified using a gradient of 0-500 mM NaCl in 20 mM sodium phosphate buffer pH 6.5. Fraction purity was confirmed by SDS-PAGE. Pure fractions were combined and concentrated using Amicon® 30 kD spin filters. Pure *Pfu* POP was buffer exchanged into H_2O, snap frozen with liquid N_2, and stored at -20°C.

Standard bioconjugation protocol for serine hydrolases with phosphonate cofactors

585 μL of 60 μM serine hydrolase in either 100 mM MES pH 6.0, 100 mM NaPi pH 7.0, or 100 mM NaPi pH 8.0 was mixed with 15 μL 24 mM phosphonate cofactor solution in MeCN in a 1.5 mL microcentrifuge tube. Reactions were carried out at 25°C, 37°C, or 75°C in dry baths shaking at 600 rpm. Extent of bioconjugation was monitored by LC/ESI-MS (applicable for all conditions) or by release of *p*-nitrophenol (applicable only at pH >7.2). Upon completion of the bioconjugation reaction, hybrids were applied to a 5 mL HiTrap Desalting column pre-equilibrated with buffer. The hybrids were fractionated in 1 mL portions. Fractions were evaluated for protein content by Bradford assay. Hybrid fractions were combined and centrifuged with Amicon 10 kD or Amicon 30 kD spin membrane filters in order to concentrate the samples. Hybrids were buffer exchanged for at least three cycles into water and evaluated for adequate removal of cofactor. Then they were aliquotted, snap frozen, and stored at -20°C.

MS Characterization of POP and hybrids

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. For LC/ESI-Q-TOF and low-resolution LC/MS analysis, 10 μL samples of 60 μM protein in H_2O or 100 mM NaPi pH 7.0 were injected onto the Aeris WIDEPORE column using a standard gradient from 10% ACN/90% H_2O + 0.1% formic acid to 90% ACN/10% H_2O + 0.1% formic acid over 15

minutes. Raw ESI spectra were deconvoluted using the Agilent Chemstation LC/MSD data deconvolution module or Agilent MassHunter.

Standard bioconversion protocol for terpyridine variants

A solution of terpyridine hybrid (500 μL, 100 μM in H_2O), 500 μL NaPi buffer (200 mM, pH 7.0 with 10% MeCN), and 12.5 μL substrate (80 mM stock in MeCN) were added to a 1.5 mL microcentrifuge tube, and incubated at 4°C for 15 min. 5 μL of peracetic acid aqueous solution (1 M) was added to the mixture. The resulting solution was shaken at 4°C overnight. The final concentrations were: 1 mM substrate, 2.5 mM AcOOH, and 50 μM hybrid. The reaction was quenched by adding 350 mL CH_2Cl_2 and 25 μL 12 mM benzoic acid internal standard in MeCN. The pH was adjusted to pH 2 for carboxylic acid substrates. The closed vials were agitated using a vortexer for 10 seconds and centrifuged at 15,000 rpm for 3 min. The organic layer was isolated. Reverse phase HPLC analysis was conducted on 10 uL sample using an Agilent Eclipse C-18 column with a standard linear gradient method from 10% MeCN/90% H_2O(+ 0.1% TFA) to 90% MeCN/10% H_2O(+ 0.1% TFA) over 15 minutes. Chiral HPLC was conducted with a Lux Cellulose column using an isocratic method (10% iPrOH/Hexanes for 20 minutes).

Cloning and expression protocol for *Pab* mutants

Genomic DNA from *Pyrococcus abyssii* was obtained from Prof. Patrick Forterre at the Institut Pasteur, France and cloned into pET28a plasmid vector using NcoI and NotI restriction sites. The gene was cloned upstream of a C-terminal hexa-histidine tag for Ni-NTA affinity chromatography. 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 DNA. Nucleotide sequences for the primers are summarized:

PAB0762 prolyl endopeptidase forward	GTCATACCATGGATGGAAGACCCCTACATATGG
PAB0762 prolyl endopeptidase reverse	TAATATGCGGCCGCAAGGTGCAGGAGAACG
PAB1418 S9 peptidase forward	ATTAATCCATGGTTGATGCAATTAATATTTGGTGATACCGTTAT
PAB1418 peptidase reverse	TAAGATGCGGCCGCCTTGTTCTCACCCTTAAGTA
PAB1300 acyl-peptide hydrolase forward	ATTAATCCATGGGTGATGAAGAAGATTAGTGAAAAAGACATTGA
PAB1300 acyl-peptide hydrolase reverse	TATTAAGCGGCCGCTCTTCCCTTCAGCCA

Table 2.1 Primers used for cloning *Pab* genes.

The thermal cycler was programmed as follows:

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

Table 2.2 Program used for standard cloning of *Pab* genes from genomic DNA.

PCR amplified fragments and plasmid vector pET28a were restriction digested with NcoI and NotI enzymes in recommended buffer at 37 °C for 2 hours. Digested DNA was cleaned by agarose gel extraction using a Qiagen (Valencia, CA) gel extraction kit before ligation. Ligations were set-up with a molar ratio of 1:3 (plasmid: insert) in 10 µL

reaction mix. A typical ligase reaction mix had 3 ng/mL 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 Rosetta2(DE3) cells. Cells were spread on LB kanamycin/chloramphenicol plates (6.25 g LB powder mix, 4 g agar, 250 ml DDI water, 50 ug/g kanamycin + 25 ug/g chloramphenicol) 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/chloramphenicol and grown overnight at 37 °C, 250 rpm. Recombinant plasmids from these primary cultures were isolated using a MiniPrep kit from Qiagen (Valencia, CA) and given for sequencing. Plasmid sequencing was done at the UChicago sequencing facility and T7 forward and T7 reverse primers were used for sequencing reactions.

Colonies containing the *Pab*-pET28a genes were used to inoculate 5 mL primary cultures containing 50 ug/g kanamycin + 25 ug/g chloramphenicol in LB media were incubated overnight at 37°C with 250 rpm shaking. The following day, the primary cultures were used to inoculate 1 L of LB media containing 50 ug/g kanamycin + 25 ug/g chloramphenicol). These were incubated at 37°C with 250 rpm shaking for ~5 hours until reaching an OD_{600}≈1.0. Overexpression of *Pab* variants was induced by addition of 1 mM isopropylthiogalactoside. The expression cultures were incubated at 37°C with 250 rpm shaking for an additional 18 hours. Cells were harvested by centrifugation at 3000 rpm for 10 minutes. They were resuspended in 100 mL 20 mM sodium phosphate buffer

pH 7.4/50 mM imidazole/50 mM NaCl and lysed by sonication. The cell lysate was clarified at 15000 rpm for 1 hour. Clarified lysate was applied to a 3 x 5 mL HisTrap NiNTA columns (GE Healthcare Life Sciences) pre-equilibrated with 20 mM sodium phosphate buffer pH 6.5. The protein was purified using a gradient of 0-500 mM imidazole in 20 mM sodium phosphate buffer pH 6.5/50 mM NaCl. Fraction purity was confirmed by SDS-PAGE. Pure fractions were combined and concentrated using Amicon® 30 kD spin filters. Pure *Pab* variants were buffer exchanged into H_2O, snap frozen with liquid N_2, and stored at -20°C.

2.5 ACKNOWLEDGEMENTS

Plasmid containing the gene for *Pfu* POP was generously donated by Prof. Harold Schreier (UMBC). Plasmid containing the gene for BioH was generously donated by Prof. Alexei Savchenko (University of Toronto). Plasmid containing the gene for cocaine esterase and heroin esterase was donated by Prof. Neil Bruce (University of York). Dr. Ketaki Belsare generated SAN libraries by combinatorial codon mutagenesis.

2.6 REFERENCES

(1) Basauri Molina, M.; Verhoeven, D. G. A.; van Schaik, A. J.; Kleijn, H.; Klein Gebbink, R. J. M. *Chem. Eur. J.* **2015**, *21* (44), 15676–15685.

(2) Matsuo, T.; Imai, C.; Yoshida, T.; Saito, T.; Hayashi, T.; Hirota, S. *Chemical Communications* **2012**, *48* (11), 1662–1664.

(3) Liu, Y.; Patricelli, M. P.; Cravatt, B. F. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (26), 14694–14699.

(4) Simon, G. M.; Cravatt, B. F. *J. Biol. Chem.* **2010**, *285* (15), 11051–11055.

(5) Kidd, D.; Liu, Y.; Cravatt, B. F. *Biochemistry* **2001**, *40* (13), 4005–4015.

(6) Powers, J. C.; Asgian, J. L.; Ekici, O. D.; James, K. E. *Chem. Rev.* **2002**, *102* (12), 4639–4750.

(7) In *Bioconjugate Techniques (Third edition)*; Academic Press: Boston, 2013; pp i–iii.

(8) Lewis, J. C. *ACS Catal.* **2013**, *3* (12), 2954–2975.

(9) Doble, M. V.; Ward, A. C. C.; Deuss, P. J.; Jarvis, A. G.; Kamer, P. C. J. *Bioorganic & Medicinal Chemistry* **2014**, *22* (20), 5657–5677.

(10) Lewis, J. C.; Coelho, P. S.; Arnold, F. H. *Chemical Society Reviews* **2011**, *40* (4), 2003–2021.

(11) Marchi-Delapierre, C.; Rondot, L.; Cavazza, C.; Ménage, S. *Israel Journal of Chemistry* **2015**, *55* (1), 61–75.

(12) Pordea, A.; Creus, M.; Panek, J.; Duboc, C.; Mathis, D.; Novic, M.; Ward, T. R. *Journal of the …* **2008**, *130* (25), 8085–8088.

(13) Gao, Y.; Crabtree, R. H.; Brudvig, G. W. *Inorg Chem* **2012**, *51* (7), 4043–4050.

(14) Liu, P.; Wong, E. L.-M.; Yuen, A. W.-H.; Che, C.-M. *Org. Lett.* **2008**, *10* (15), 3275–3278.

(15) Kamijo, S.; Amaoka, Y.; Inoue, M. *Synthesis* **2010**, *2010* (14), 2475–2489.

(16) Das, S.; Incarvito, C. D.; Crabtree, R. H.; Brudvig, G. W. *Science* **2006**, *312* (5782), 1941–1943.

(17) Hull, J. F.; Sauer, E. L. O.; Incarvito, C. D.; Faller, J. W.; Brudvig, G. W.; Crabtree, R. H. *Inorg Chem* **2009**, *48* (2), 488–495.

(18) Himiyama, T.; Sauer, D. F.; Onoda, A.; Spaniol, T. P.; Okuda, J.; Hayashi, T. *Journal of Inorganic Biochemistry* **2016**, *158*, 55–61.

(19) Limburg, J.; Vrettos, J. S.; Chen, H.; de Paula, J. C.; Crabtree, R. H.; Brudvig, G. W. *J Am Chem Soc* **2001**, *123* (3), 423–430.

(20) Farady, C. J.; Craik, C. S. *Protease Inhibitors: Mechanisms*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2008.

(21) Bourne, Y. *Journal of Biological Chemistry* **1999**, *274* (43), 30370–30376.

(22) Sanishvili, R.; Yakunin, A. F.; Laskowski, R. A.; Skarina, T.; Evdokimova, E.;

Doherty-Kirby, A.; Lajoie, G. A.; Thornton, J. M.; Arrowsmith, C. H.; Savchenko, A.; Joachimiak, A.; Edwards, A. M. *J. Biol. Chem.* **2003**, *278* (28), 26039–26045.

(23) Larsen, N. A.; Turner, J. M.; Stevens, J.; Rosser, S. J.; Basran, A.; Lerner, R. A.; Bruce, N. C.; Wilson, I. A. *Nature Structural & Molecular Biology* **2002**, *9* (1), 17–21.

(24) Zhu, X. *Journal of Biological Chemistry* **2002**, *278* (3), 2008–2014.

(25) Harris, M. N. *Journal of Biological Chemistry* **2001**, *276* (22), 19310–19317.

(26) Li, H.; Schopfer, L. M.; Nachon, F.; Froment, M.-T.; Masson, P.; Lockridge, O. *Toxicol. Sci.* **2007**, *100* (1), 136–145.

(27) Kruithof, C. A.; Casado, M. A.; Guillena, G.; Egmond, M. R.; van der Kerk-van Hoof, A.; Heck, A. J. R.; Klein Gebbink, R. J. M.; van Koten, G. *Chem. Eur. J.* **2005**, *11* (23), 6869–6877.

(28) Basauri-Molina, M.; Riemersma, C. F.; Würdemann, M. A.; Kleijn, H.; Gebbink, R. J. M. K. *Chemical Communications* **2015**, *51* (31), 6792–6795.

(29) Kruithof, C. A.; Dijkstra, H. P.; Lutz, M.; Spek, A. L.; Egmond, M. R.; Klein Gebbink, R. J. M.; van Koten, G. *Eur. J. Inorg. Chem.* **2008**, *2008* (28), 4425–4432.

(30) James R Carey; Steven K Ma; Thomas D Pfister; Dewain K Garner; Hyeon K Kim; Joseph A Abramite; Zhilin Wang; Zijian Guo, A.; Yi Lu. *A Site-Selective Dual Anchoring Strategy for Artificial Metalloprotein Design*; American Chemical Society, 2004; Vol. 126, pp 10812–10813.

(31) *Chemistry and Application of H-Phosphonates*; Elsevier, 2006.

(32) Harwood, V. J.; Schreier, H. J. In *Methods in enzymology*; Methods in Enzymology; Elsevier, 2001; Vol. 330, pp 445–454.

(33) Zhang, C.; Srivastava, P.; Ellis-Guardiola, K.; Lewis, J. C. *Tetrahedron* **2014**, *70* (27-28), 4245–4249.

(34) Belsare, K. D.; Andorfer, M. C.; Cardenas, F. S.; Chael, J. R.; Park, H. J.; Lewis, J. C. *ACS Synth Biol* **2017**.

(35) Reetz, M. T.; Carballeira, J. D. *Nat Protoc* **2007**, *2* (4), 891–903.

(36) Thomas, C. M.; Ward, T. R. *Applied Organometallic Chemistry* **2005**, *19* (1), 35–39.

(37) Que, L.; Tolman, W. B. *Nature* **2008**, *455* (7211), 333–340.

(38) Wang, D.; Zhang, M.; Bühlmann, P.; Que, L. *J Am Chem Soc* **2010**, *132* (22), 7638–7644.

(39) Chen, M. S.; White, M. C. *Science* **2007**, *318* (5851), 783–787.

(40) Gormisky, P. E.; White, M. C. *J Am Chem Soc* **2013**, *135* (38), 14052–14055.

(41) England, J.; Davies, C. R.; Banaru, M.; White, A. J. P.; Britovsek, G. J. P. *Advanced Synthesis & Catalysis* **2008**, *350* (6), 883–897.

(42) Esmieu, C.; Cherrier, M. V.; Amara, P.; Girgenti, E.; Marchi-Delapierre, C.; Oddon, F.; Iannello, M.; Jorge-Robin, A.; Cavazza, C.; Ménage, S. *Angew. Chem. Int. Ed. Engl.* **2013**, *52* (14), 3922–3925.

(43) Rondot, L.; Girgenti, E.; Oddon, F.; Marchi-Delapierre, C.; Jorge-Robin, A.; Ménage, S. *Journal of Molecular Catalysis A: Chemical* **2016**, *416*, 20–28.

(44) Lee, H. S.; Schultz, P. G. *J Am Chem Soc* **2008**, *130* (40), 13194–13195.

(45) Sharma, A.; Hartwig, J. F. *Nature* **2015**, *517* (7536), 600–604.

(46) Srivastava, P.; Yang, H.; Ellis-Guardiola, K.; Lewis, J. C. *Nat Commun* **2015**, *6*, 7789.

(47) Minus, M. B.; Kang, M. K.; Knudsen, S. E.; Liu, W.; Krueger, M. J.; Smith, M. L.; Redell, M. S.; Ball, Z. T. *Chem. Commun.* **2016**, *52* (78), 11685–11688.

(48) Espino, C. G.; Fiori, K. W.; Kim, M.; Bois, Du, J. *J. Am. Chem. Soc.* **2004**, *126* (47), 15378–15379.

(49) Yang, H.; Srivastava, P.; Zhang, C.; Lewis, J. C. *Chembiochem* **2014**, *15* (2), 223–227.

(50) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215* (3), 403–410.

(51) Romero, P. A.; Arnold, F. H. *Nat. Rev. Mol. Cell Biol.* **2009**, *10* (12), 866–876.

(52) Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E. *Nat Protoc* **2015**, *10* (6), 845–858.

(53) Brezovsky, J.; Chovancova, E.; Gora, A.; Pavelka, A.; Biedermannova, L.; Damborsky, J. *Biotechnology Advances* **2013**, *31* (1), 38–49.

(54) Potts, K. T.; Konwar, D. *J. Org. Chem.* **1991**, *56* (15), 4815–4816.

(55) Harwood, V. J.; Denson, J. D.; Robinson-Bidle, K. A.; Schreier, H. J.

CHAPTER THREE

PYROCOCCUS FURIOSUS PROLYL OLIGOPEPTIDASE: METHODS AND STRUCTURE

3.1 INTRODUCTION

The encapsulation of metal cofactors within a protein scaffold is an important feature in the design of artificial metalloenzymes, as this defines the secondary coordination sphere about the metal, ideally enabling controlled alteration of this environment.[1,2] Such secondary coordination sphere control is desirable to impart selectivity on the metal catalyst at the core of the artificial metalloenzyme. As such, judicious selection of an ArM scaffold should be employed to engender encapsulation of the metal and, in turn, desired selectivity in ArM catalysis.

Prolyl oligopeptidase fits the description of an optimal ArM scaffold. Prolyl oligopeptidase (POP; EC 3.4.21.26) is a serine peptidase that catalyzes the hydrolysis of the C-terminal peptide bond of proline residues in short (<30 amino acid) peptides.[3,4] POP is found in all three domains of life and is widely distributed throughout a variety of tissue types, making its specific physiological role in different organisms a heavily investigated topic.[5] The architecture of POP makes it an appealing scaffold for the formation of ArMs. The two-domain protein is comprised of a canonical α/β-hydrolase domain capped by a β-propeller domain.[5] The active site serine of POP sits at the base of the central cavity enclosed by the β-propeller domain. In chapter two, we postulated that anchoring phosphonate cofactors to the active site serine could fully encapsulate the metal. This same logic was applied in chapter four, wherein a *p*-azidophenylalanine

residue was introduced at the location of the active site serine to enable the enclosure of a large bis-cyclononyne-linked (BCN) dirhodium catalyst.[6]

We obtained the gene for *Pyrococcus furiosus* POP (*Pfu* POP) to test it as a serine hydrolase scaffold for the generation of ArMs using phosphonate-based cofactors (chapter two). This particular homologue was chosen due to its reported robustness to high temperatures and its facile heterologous expression in *E. coli*.[6] In addition, despite the lack of a crystal structure for this enzyme at the time, researchers had built an adequate homology model, forming the basis for initial studies in rational design of *Pfu* POP ArMs (Chapter Four).[7]

3.2 RESULTS AND DISCUSSION

3.2.1 PYROCOCCUS FURIOSUS PROLYL OLIGOPEPTIDASE: AN ARM SCAFFOLD

3.2.1.1 EXPRESSION

As described in chapter two, selection criteria for selecting novel serine hydrolase scaffolds included not only a deep binding pocket to encapsulate the metal cofactor, but also efficiency of gene expression in *Escherichia coli*. Expression of this gene had already been reported in *E. coli* BL21(pLysS)DE3 cells. This gene[8] was obtained from Prof. Harold Schreier at UMBC as a bacterial stab in this bacterial host. The gene itself was contained in a pET11d plasmid, with inducible expression under the control of the *lac* operon. Initial attempts at expressing this construct in the original host BL21(pLysS)DE3 cells were successful, with IPTG-induced overexpression of the soluble construct of the expected molecular weight, as evidenced by SDS-PAGE.

Enzyme purification was accomplished by a combination of two rounds of heat treatment at 85 °C followed by anion exchange on mild anion exchange Q resin.[9]

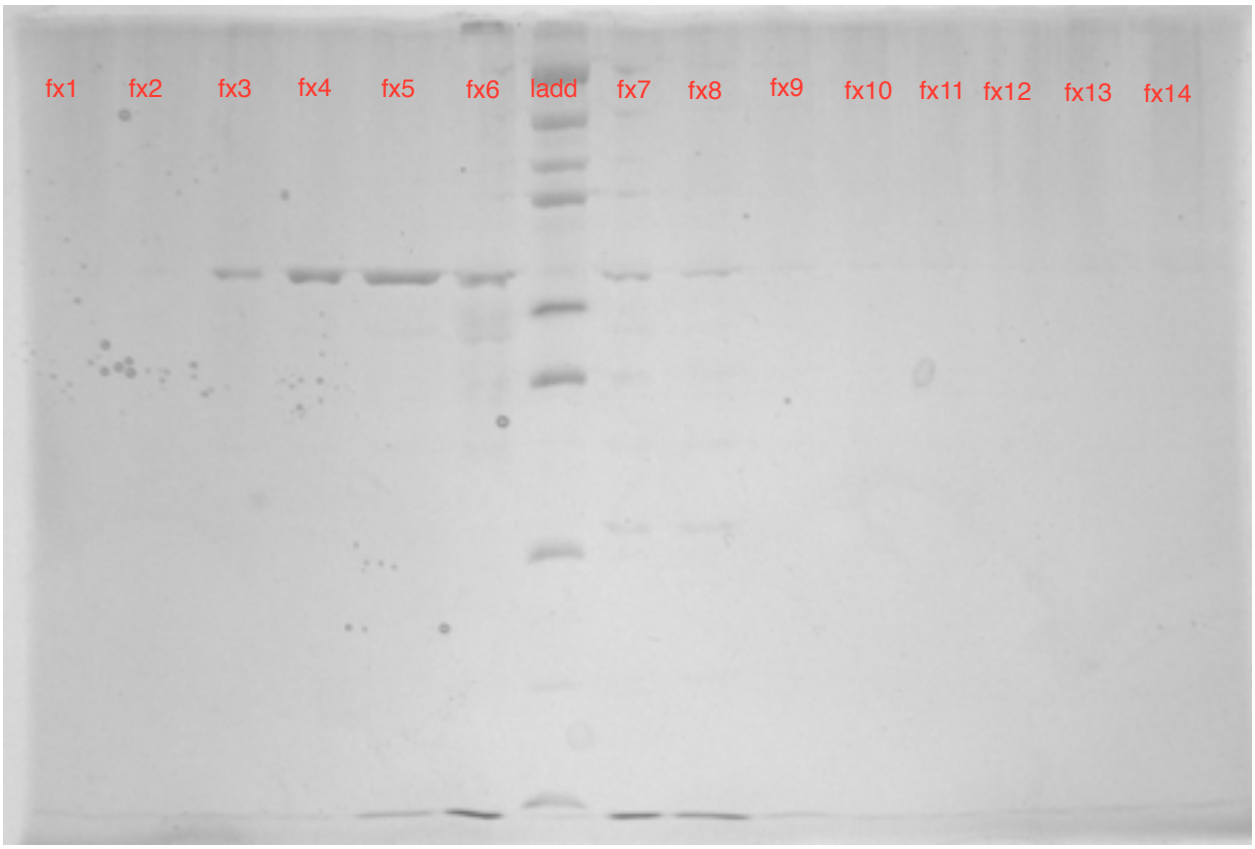


Figure 3.1 Anion exchange fractions from 1 L *Pfu* POP expression in BL21(pLysS)DE3.

Heat treatment effectively removed the vast majority of endogenous *E. coli* proteins, while anion exchange enabled isolation of pure protein fractions. Pure anion exchange fractions can be seen in Figure 3.1. High-speed centrifugal membrane concentration of these fractions proceeded without discernable precipitation of the protein, a testament to its high stability even at high, aggregation-prone concentrations. The ultimate yield of *Pfu* POP with this expression host proved to be disappointingly low, however, with ~10

mg/L isolated. Such low expression enabled preliminary bioconjugation/ArM bioconversion studies, but was unsustainable for long-term investigations or screening of catalysis conditions.

The low yield of *Pfu* POP could not be blamed on losses due to insoluble protein expression or problems in isolation, as SDS-PAGE revealed that most of the protein was expressed in soluble form and that little protein loss occurred during isolation/purification steps. Rather, the problem was deemed to be likely a symptom of rare codon abundance in the gene, a common problem in the expression of archaeal genes.[10] To address this, we transformed the *Pfu* POP plasmid into engineered *E. coli* strain BL21 Rosetta2(DE3) (Novagen), which encodes seven extra tRNAs for rare codon translation. With this new host, the yield of *Pfu* POP increased to ~30 mg/L, a 3-fold improvement. (Figure 3.2).

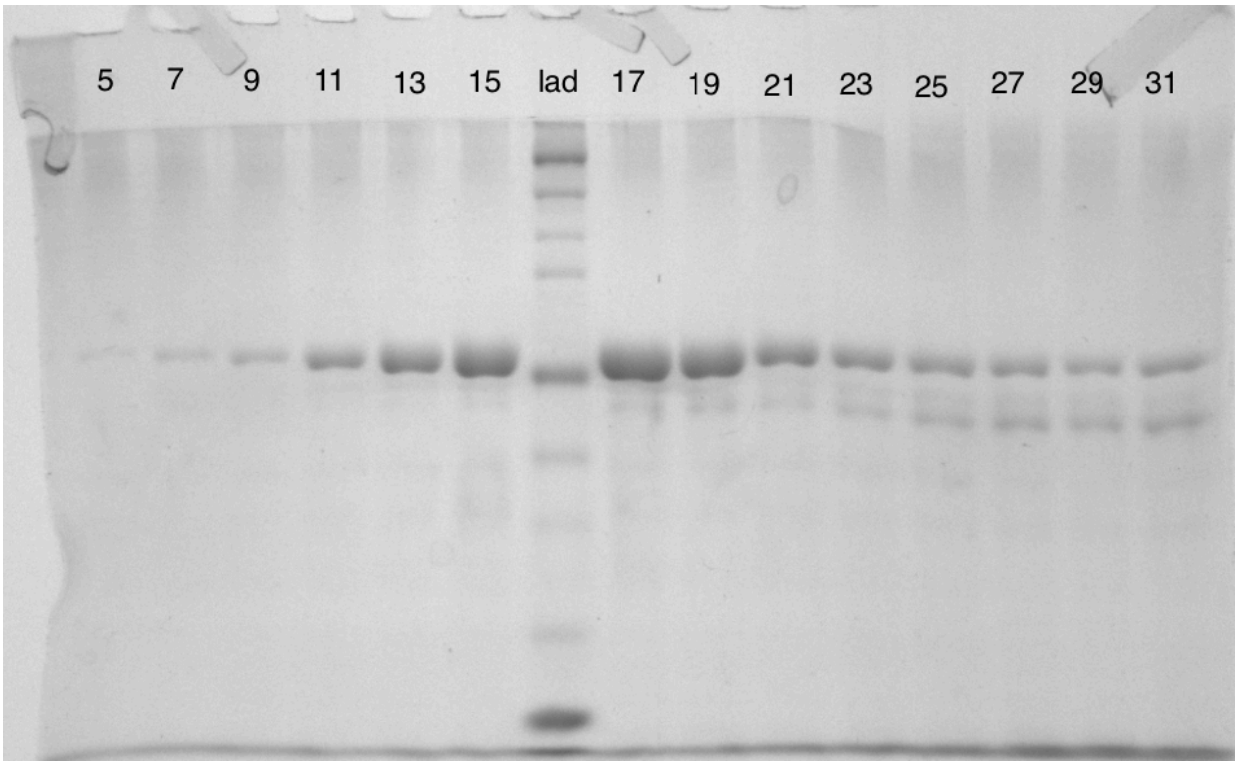


Figure 3.2 SDS-PAGE anion exchange fractions from 1 L expression of *Pfu* POP in Rosetta2(DE3) cells show improvement in overall expression yield.

This proved to be sufficient to sustainably generate protein for formation of ArMs with mechanism-based cofactors. Even greater expression yields were necessitated for the introduction of unnatural amino acids (UAAs) into the *Pfu* POP scaffold, a strategy that was used in the creation of SPAAC *Pfu* POP/dirhodium ArMs (chapter four). Since amber codon suppression relies on the co-expression of tRNA/tRNA synthetase pairs along with the introduction of exogenous amino acids, expression yields of proteins containing unnatural amino acids tend to be markedly lower than their native parents.[11] Therefore, we sought an alternative strategy to enhance expression yields.

Codon optimization is a strategy to convert rare codons to degenerate equivalents that are more optimal for expression in heterologous hosts. [12] The original *Pfu* POP gene was converted to a codon-optimized form by GenScript, after which it was cloned into a pET28a vector and transformed into BL21(DE3)Gold cells. This also introduced a C-terminal hexahistidine tag for NiNTA affinity chromatography, greatly facilitating the expression of the *Pfu* POP gene. With this improved gene construct, regular native *Pfu* POP expression yields of ~150 mg/L could be obtained, while expression yields of *Pfu* POP containing UAAs were ~100mg/L. In addition, purification was simplified to a single step of NiNTA affinity chromatography, enabling rapid processing and isolation of pure protein (Figure 3.3). This significantly enhanced the practicality of this scaffold in its use as a general scaffold for ArMs.

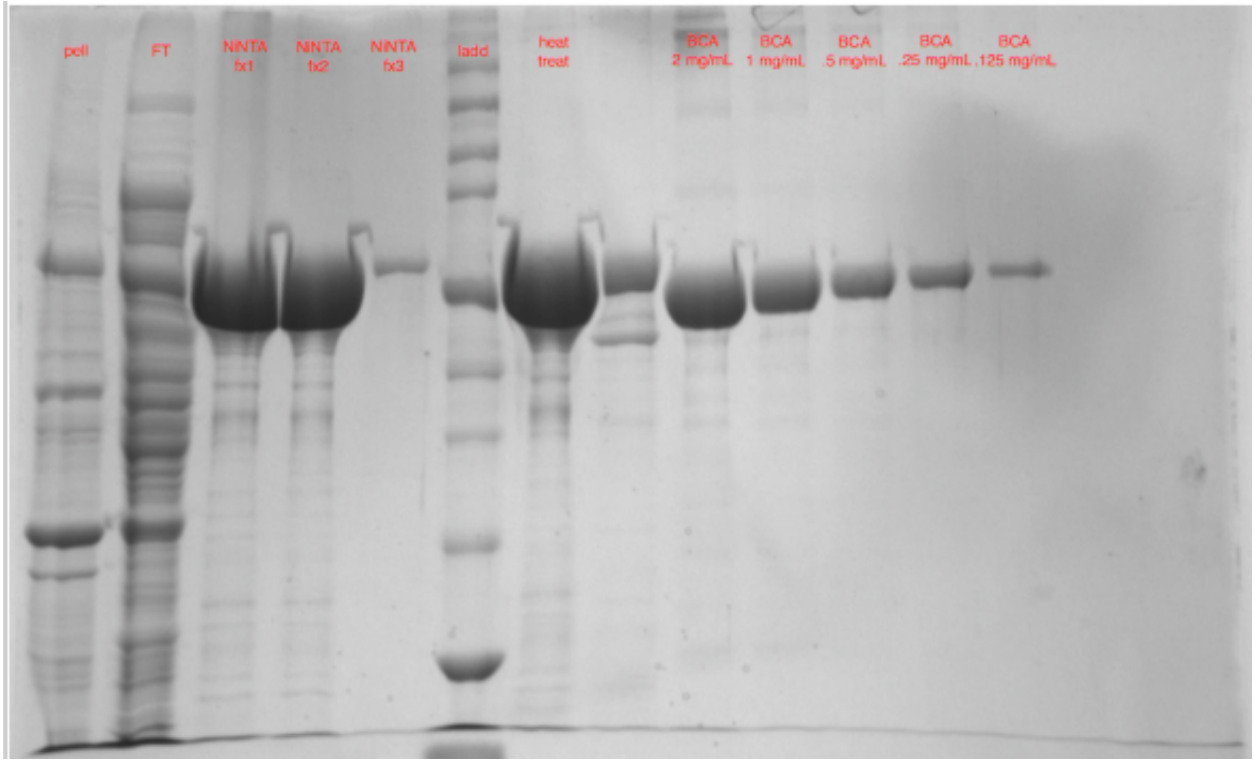


Figure 3.3 SDS-PAGE of *Pfu* POP expression of codon-optimized gene in BL21(DE3)Gold. Purification was carried out through NiNTA affinity chromatography to yield ~150 mg/L.

3.2.1.2 STABILITY

An important aspect of ArM design revolves around the practicality of the protein scaffold utilized in ArM formation. Given the ubiquity of prolyl oligopeptidases in every domain of life, the number of viable source organisms for this gene is massive.[13,14] The first crystal structure of POP was solved with crystals of heterologously-expressed POP from *Sus scrofa*, the common pig.[15] In spite of available structural data for this construct, we opted for POP from the hyperthermophilic archaeon *Pyrococcus furiosus*. Aside from practical concerns of expression, this POP variant had been proven to be exceptionally stable to high temperatures. No melting temperatures have been reported for this enzyme,

as *Pfu* POP can tolerate temperatures above 100°C. This exceptional robustness was an attractive feature of this scaffold for the creation of ArMs, as this confers a number of advantages. From a practical standpoint, scaffold stability enables a greater diversity of conditions to be explored for catalysis, enhancing the flexibility of this platform to accommodate chemical conditions that might not be endured by more sensitive protein scaffolds. Such parameters might include high temperatures, high organic co-solvent content, as well as resistance to proteolysis. Aside from these practical concerns, the high stability of the protein scaffold enables a greater degree of tolerance for mutational loads. This is important in the course of scaffold optimization by mutagenesis, since the protein must be folded and soluble in order to constitute an effective ArM scaffold.[16] Without this stability, scaffold mutagenesis becomes significantly more constrained.

The robustness of *Pfu* POP to high temperature has been reported previously. [17] The optimum temperature of this enzyme in native peptide hydrolysis is 85-95°C.[17] The temperature stability of this enzyme has been demonstrated with activity-based assays. Biophysical characterization of *Pfu* POP stability has not yet been demonstrated, however. In light of this, we sought to corroborate the reported stability studies by detailed measurement of the circular dichroism signal of this enzyme as a function of temperature. Even at 98°C (the maximum temperature allowed by the instrument), there was no change in the CD spectrum of *Pfu* POP, indicating its imperviousness to high temperatures. (Figure 3.4)

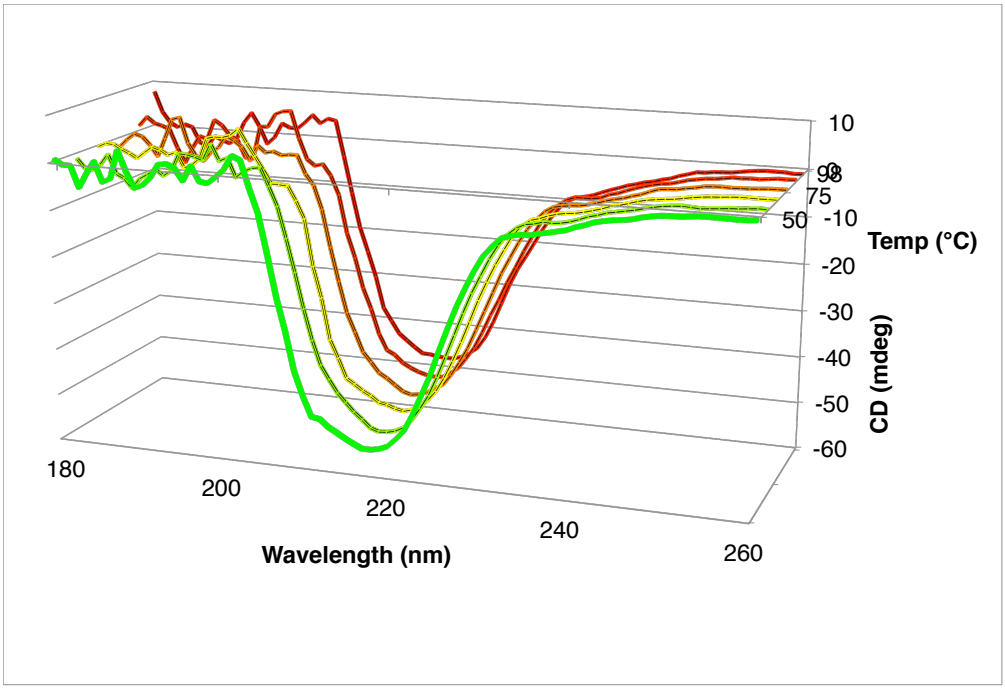


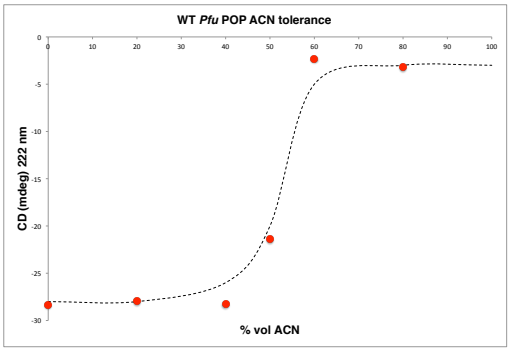
Figure 3.4 CD spectra of *Pfu* POP acquired at increasing temperatures from 50°C-98°C. Conditions: 10 uM *Pfu* POP in 100 mM NaPi pH 7.4.

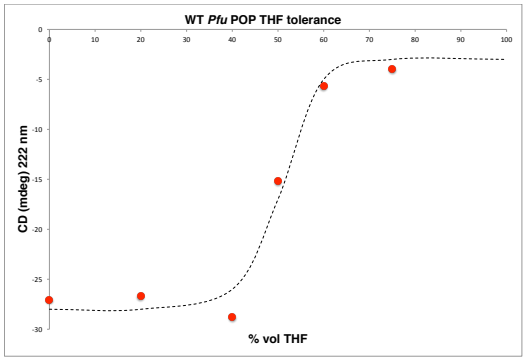
A similar CD study was conducted to determine the stability of *Pfu* POP toward increasing concentrations of organic co-solvent. Samples of *Pfu* POP were incubated for 5 minutes in medium containing the appropriate loading of cosolvent, after which the CD signal was measured. Solvent tolerance was measured against two common organic co-solvents used in our laboratory: tetrahydrofuran and acetonitrile (Figure 3.5). Remarkably, the scaffold proved resistant to 40% loading of both of these co-solvents, with clear denaturation starting at 50% solvent loading, finally leading to full denaturation of the sample at 60% solvent loading. This affirms the remarkable

robustness of *Pfu* POP, which may be a boon to future investigations where substrate solubility or purification might necessitate more hydrophobic media.

Figure 3.5 CD titration curves with increasing concentrations of organic cosolvent. Conditions:

5 uM Pfu POP in (100-x)% 100 mM sodium phosphate buffer pH 7.4/x% organic solvent





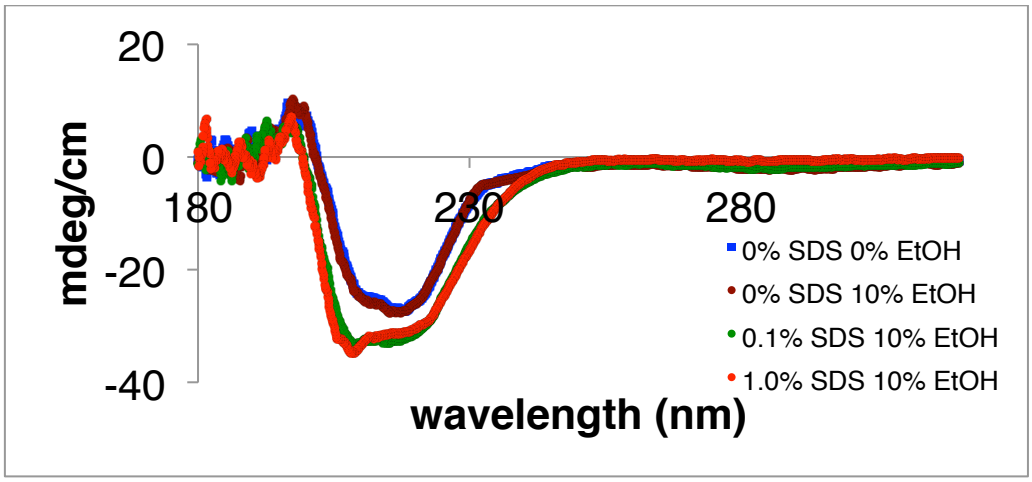
In a testament to the mutational tolerance of this enzyme, we investigated the temperature stability of a mutant *Pfu* POP-ZA$_4$-HFF, which contains eight mutations (G99H, E104A, F146A, K199A, D202A, L328H, S477Z, and G594F) relative to the wild-type parent (the rationale for these mutations is discussed in chapter four).

Remarkably, in spite of these mutations, the temperature robustness of this scaffold is indistinguishable from that observed with wild-type *Pfu* POP. This demonstrated that not only is high mutational load tolerated, but there is in fact room for even more mutagenic diversification using POP-ZA_4-HFF as a starting point.

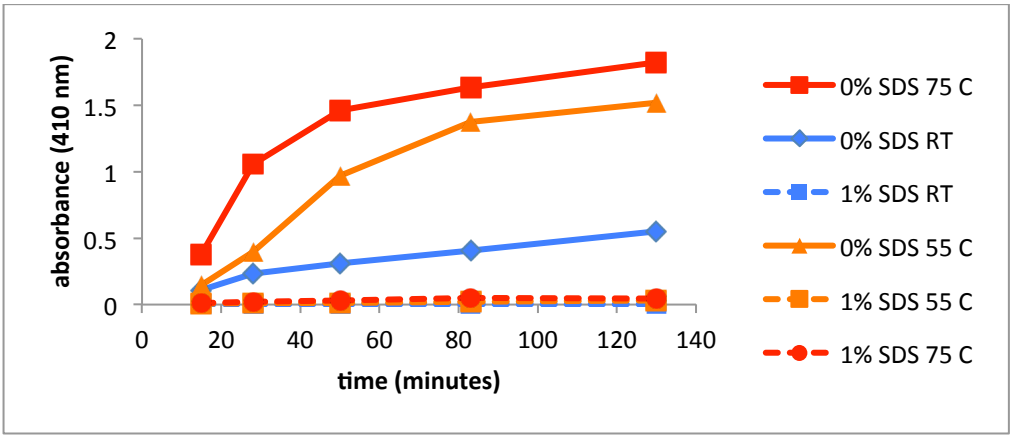
In spite of the importance of structural stability in ArMs, there are times when one must unfold the protein scaffold, either for characterization purposes or for enabling bioconjugations of large cofactors (this necessitates the ability to refold the protein and for bioconjugation to be dependent on inherent residue reactivity). Denaturation of *Pfu* POP can be carried out by precipitation with organic solvent (see above), low pH,[18] and through the introduction of even small quantities of denaturing detergents like sodium dodecylsulfate (SDS). This was apparent from CD spectra and ablated native activity of the enzyme upon exposure to SDS (Figure 3.6).

Figure 3.6 Characterization of effect of SDS on *Pfu* POP by CD and native activity

A) CD spectra of *Pfu* POP indicating a significant structural perturbation induced by exposure to SDS.



B) *Pfu* POP native peptidase activity in the presence and absence of SDS. Rate of Z-gly-pro-*p*-nitroanilide amide hydrolysis is monitored by release of *p*-nitroaniline. Measurements were taken at three temperatures

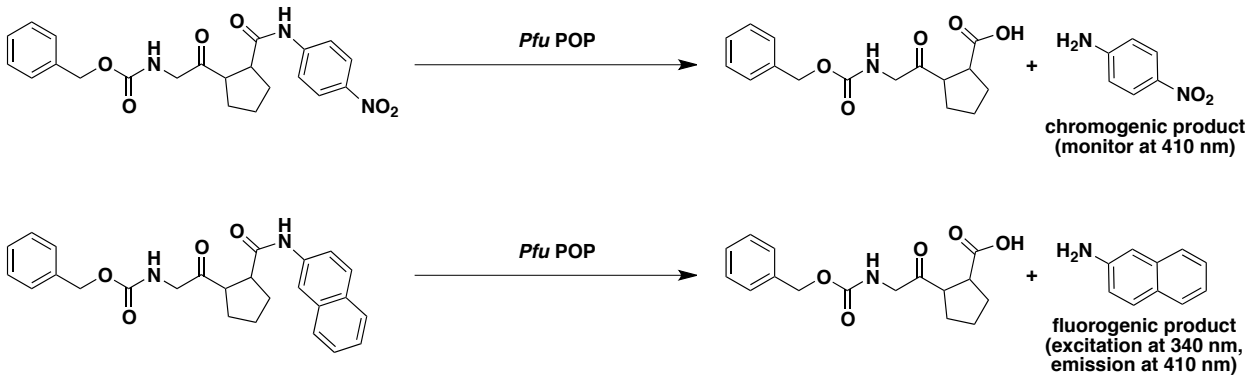


3.2.1.3 ROUTINE CHARACTERIZATION OF PFU POP ARTIFICIAL METALLOENZYMES

A key concern in the development of artificial metalloenzymes is their adequate characterization, which is often a deficiency in ArM literature. *Pfu* POP has proven to be a robust platform for detailed characterization of these highly complex hybrid catalysts.

3.2.1.3.1 NATIVE ACTIVITY

As shown above, a proxy for proper folding of *Pfu* POP is its ability to catalyze its native peptide hydrolysis reactions. Native activity can also be assayed for the evaluation of bioconjugation with mechanism-based inhibitor cofactors, like the phosphonates discussed in Chapter Two. Reliable chromogenic and fluorogenic assays have been developed for this purpose using the short substrate peptide Z-Gly-Pro-*p*-nitroaniline and Z-Gly-Pro-2-naphthylamine (Scheme 3.1).[17,19] These methods have been employed by a number of coworkers in the search for alternative mechanism-based cofactors. The preference in this work has been for robust physical characterization by mass spectrometry.



Scheme 3.1 Monitoring of *Pfu* POP native activity can be readily performed with chromogenic or fluorogenic substrates Z-gly-pro-*p*-nitroaniline (top) and Z-gly-pro-2-naphthylamine (bottom)

3.2.1.3.2 INTACT ELECTROSRAY IONIZATION MASS SPECTROMETRY

Mass spectrometry is one of the most important tools for the routine characterization of covalent ArMs, as this enables rapid confirmation of extent of bioconjugation and other modifications based on the observed shift in mass of the hybrid relative to its parent scaffold. With advances in modern mass spectrometry, it is possible to acquire high-resolution mass data for the characterization of proteins and covalent modifications thereof. While MALDI has seen extensive application for small biomolecule characterization (up to ~25 kDa), its utility for species of higher molecular weight is somewhat limited, as the resolution of determined MALDI masses tends to deteriorate in higher mass ranges.[20,21] In contrast, electrospray ionization mass spectrometry (ESI-MS) is extremely powerful for the observation of covalent modifications, as its resolution is maintained in high mass ranges.[22]

After significant optimization efforts, we have developed methods for the routine characterization of *Pfu* POP by ESI-MS, enabling high-resolution characterization of perturbations in the scaffold mass due to mutation and covalent modification. Reliable characterization of *Pfu* POP ArMs can be carried out via direct injection of these species into ESI-MS instruments, which requires a significant effort to remove contaminating components from the ArM sample, as these can contribute to poor ionization of the analyte. A more general method was developed employing liquid chromatography/ESI-MS (LC/ESI-MS) to enable rapid analysis of ArM mixtures without the need for purification of the samples. Indeed, this has become the standard to confirm the identity of ArMs derived from *Pfu* POP. A critical advantage of this method arises from the ability to simultaneously assess both the extent of ArM bioconjugation and the purity of the sample (importantly, to identify residual cofactor that may interfere with downstream catalysis).

This methodology has enabled the identification of unexpected behavior of the scaffold itself. A notable degradation of mass spectral quality was observed when *Pfu* POP and variants thereof were retained in solution for extended periods of time, either at 4°C or room temperature. This was perplexing, as native activity was retained with this enzyme indefinitely, indicating that the mass spectral issues were not a result of denaturation or aggregation. A long-term investigation of the fate of a wild-type *Pfu* POP sample incubated at room temperature for three weeks showed the gradual drop in protein mass to a species ~410 Da lower than the calculated native mass (Figure 3.7). In addition, this gradual drop proceeded in increments of ~-140 Da. This was determined to arise due gradual deterioration of the C-terminal hexahistidine tag, which was further corroborated

by the analysis of aged samples by protein digestion (see below). Of course, such degradation could be prevented through the proper frozen storage of the protein scaffold.

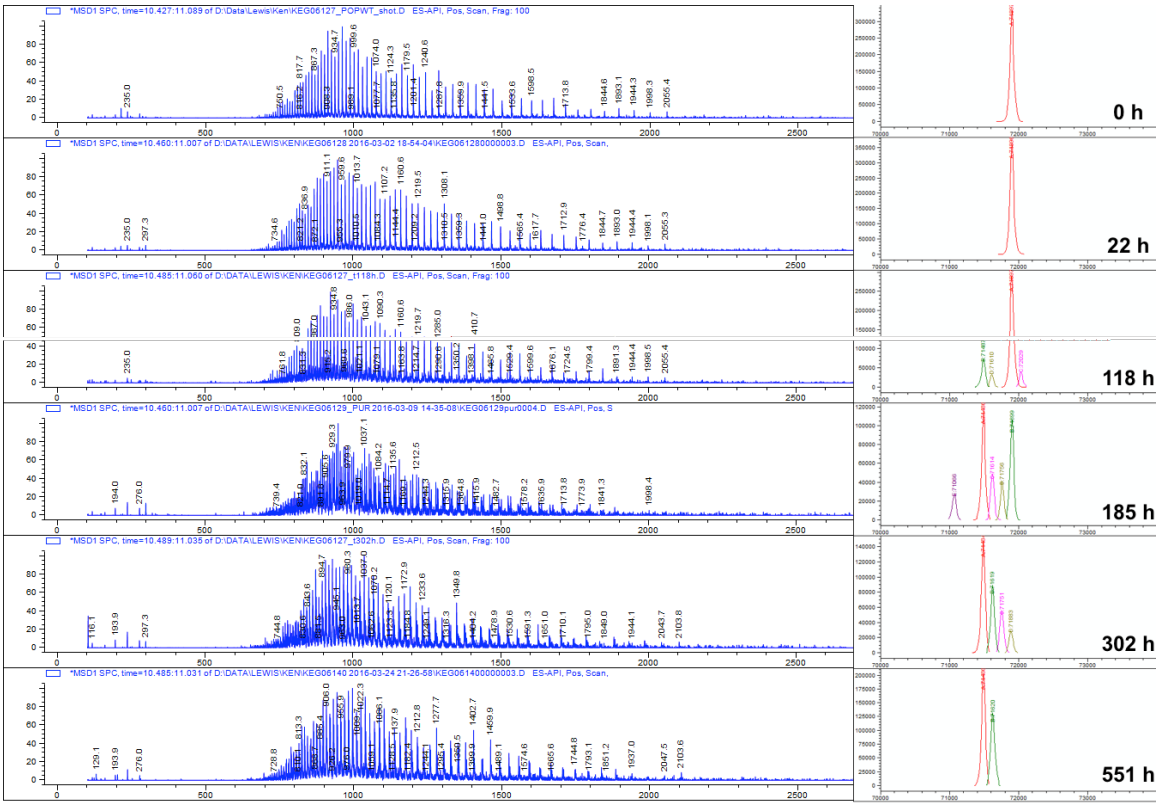


Figure 3.7 Gradual decomposition of the hexahistidine tag of *Pfu* POP incubated at 60 uM concentration at 25°C over three weeks. Raw ESI-MS spectra are on the left. Deconvoluted spectra are on the right, indicating conversion to a product ~410 Da shorter than the initial scaffold.

The careful and routine analysis of *Pfu* POP and its corresponding ArMs by intact electrospray ionization mass spectrometry have thus been proven to enable the understanding of subtle chemical effects that would otherwise be overlooked by traditional biochemical methods. Therefore, the protocols developed herein provides a robust framework for the routine yet detailed characterization of *Pfu* POP ArMs.

3.2.1.3.3 PROTEIN DIGESTION AND LC/MS/MS

While the mass spectrometric characterization of intact protein samples enables the evaluation of changes to the protein scaffold in a global sense, it does not provide information about the localization of changes to the protein scaffold. In order to locate changes at the primary sequence level, alternate methods must be explored. A ubiquitous methodology in the field of proteomics relies on enzymatic or chemical digestion of proteins followed by their analysis by LC/MS/MS.[23] This can provide a wealth of primary sequence level information about the analyzed proteins. Of greatest interest in its application to ArMs is the ability of digest-LC/MS/MS methodology to identify the residue-specific location of covalent modifications arising either from bioconjugation or ArM self-modification during catalysis. Drawing from the extensive studies into protein post-translational modification (PTM), we sought to leverage digestion/LC/MS/MS methodology to extract sequence-level information of covalent modification of ArMs. The methods described herein form the basis for a detailed study of scaffold modification during ArM catalysis, which will be described in chapter four.

Owing to the exceptional stability of *Pfu* POP, the development of reliable methods to digest this enzyme was quite challenging. Initial attempts at traditional protein digestion using trypsin all failed, even under the denaturing conditions tolerated by trypsin.[24] In light of this, we sought alternative proteases for digestion of *Pfu* POP. High-temperature digestion with thermostable protease thermolysin failed to yield appreciable digestion (Figure 3.8).

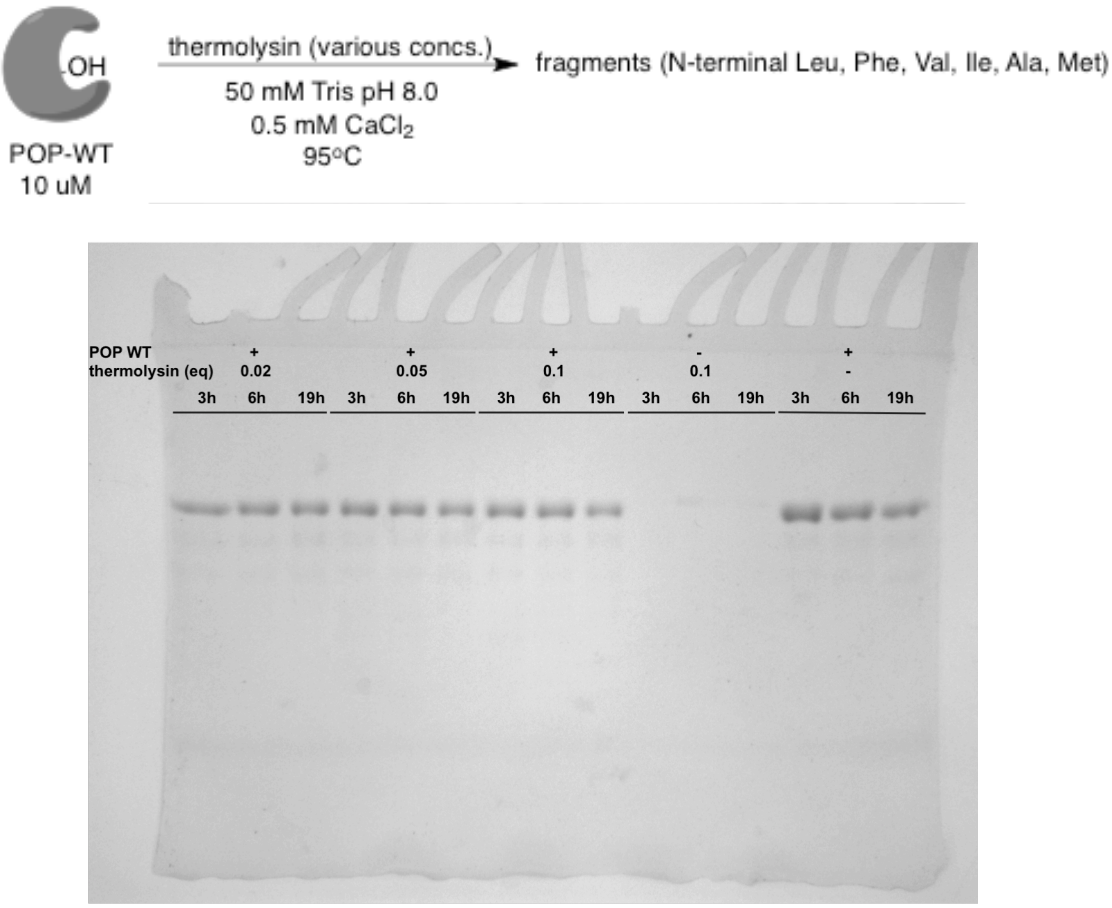


Figure 3.8 Attempted proteolysis of *Pfu* POP with thermolysin.

Pfu POP digestion under highly acidic conditions (pH 1.0) with the acid-tolerant protease pepsin yielded ready digestion of *Pfu* POP, owing to the acid-promoted denaturation of this enzyme (Figure 3.9)

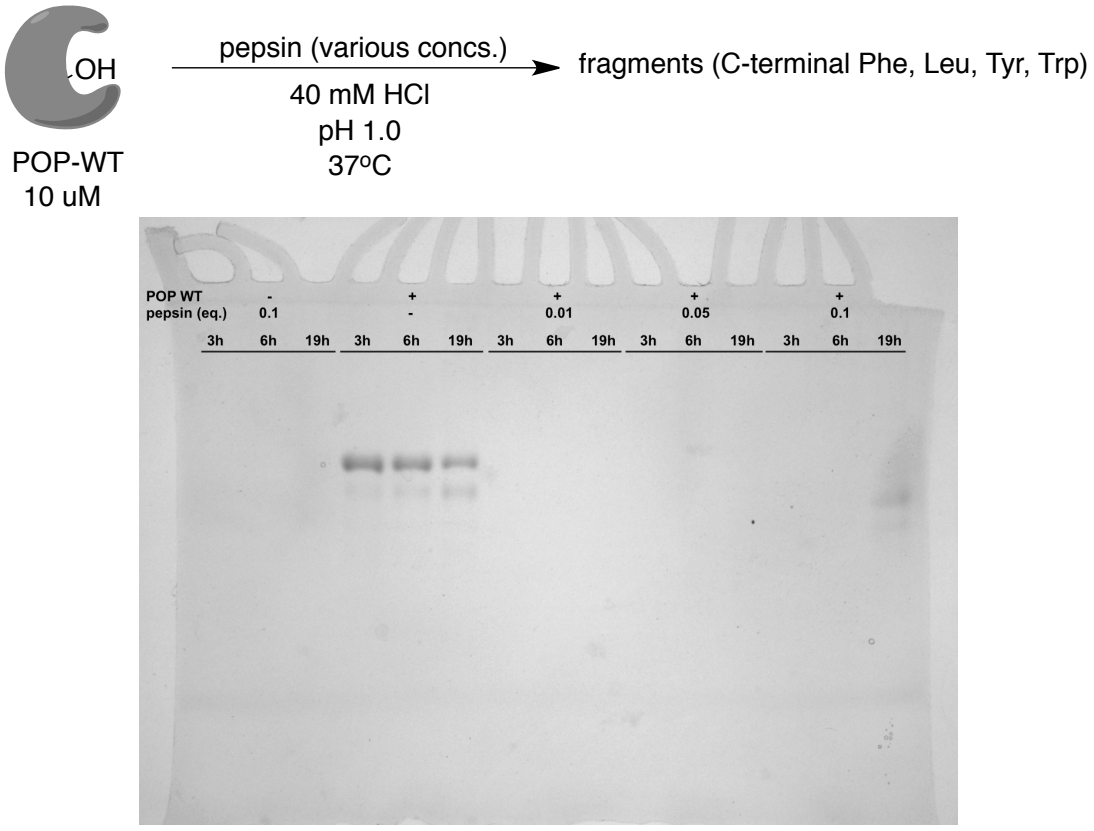
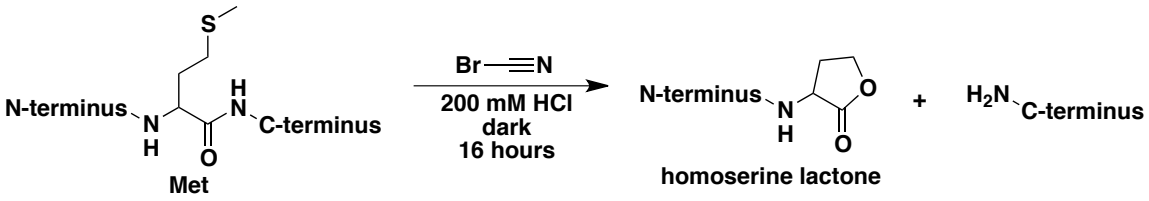


Figure 3.9 Proteolysis of *Pfu* POP with pepsin proceeds efficiently at al concentrations tested.

Unfortunately, pepsin suffers from low specificity, targeting different cleavage sites with even subtle changes in the pH of the medium. This was evidenced by poorly reproducible LC/MS chromatograms of the product peptides. In the search for alternative digestion methods reliant on low pH (to ensure denaturation of *Pfu* POP), we identified cyanogen bromide (CNBr), which is a chemical protein digestion reagent with high specificity for methionine residues. The digestion reaction proceeds as shown in Scheme 3.2. CNBr provided highly reproducible digestion of *Pfu* POP and mutants thereof, as evidenced by LC/MS chromatograms of product peptide mixtures (Figure 3.10)



Scheme 3.2 CNBr reaction with proteins at methionines yielding homoserine lactone on N-terminal segment and free amine on C-terminal segment.

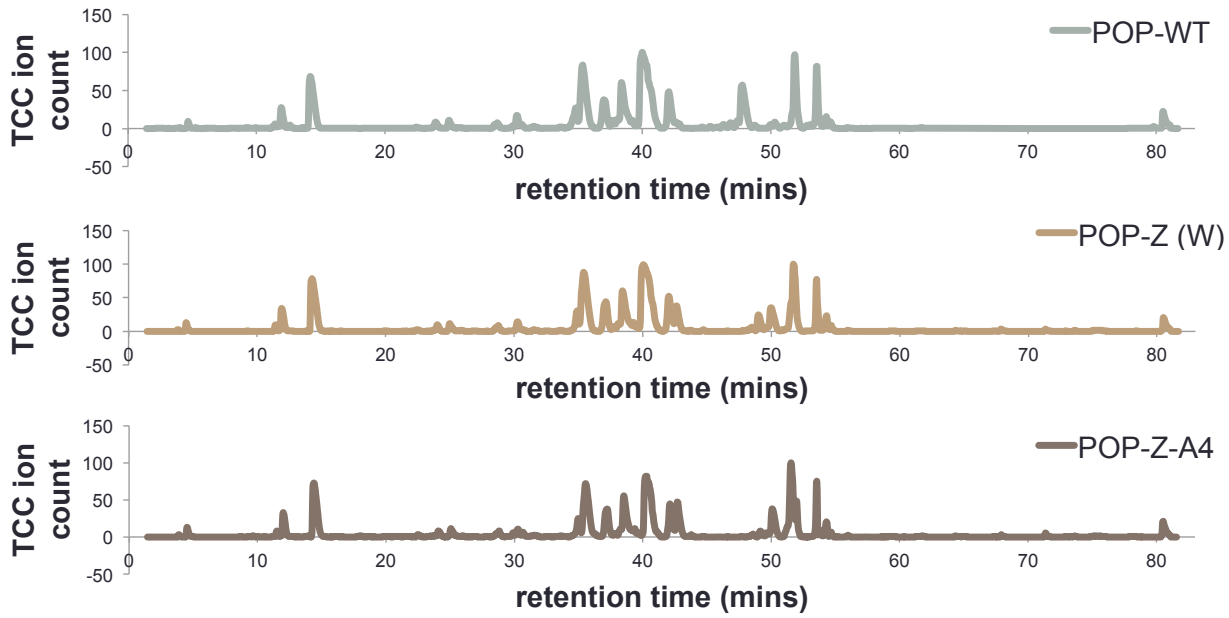


Figure 3.10 Reproducibility of CNBr digestion as demonstrated by chromatograms of three different *Pfu* POP mutants *Pfu* POP, *Pfu* POP S477Z, and *Pfu* POP-A_4-S477Z

To demonstrate the utility of this methodology in the identification of sequence-local changes to the *Pfu* POP scaffold, we compared the digest chromatograms of wild-type *Pfu* POP and *Pfu* POP S477A, an inactive variant lacking the active site S477. A

major change in the retention time was observed for the peptide fragment containing position 477, a remarkable shift given the subtlety of the mutation, which only consists of the deletion of a hydroxyl group. (Figure 3.11). The ability to separate species with such chemical similarity presents an alternative method for the evaluation of extent of bioconjugation, wherein species are only differentiated by the presence of a covalent modification. Comparison of the peak areas corresponding to modified and unmodified fragments could yield higher-resolution information about the extent of bioconjugation.

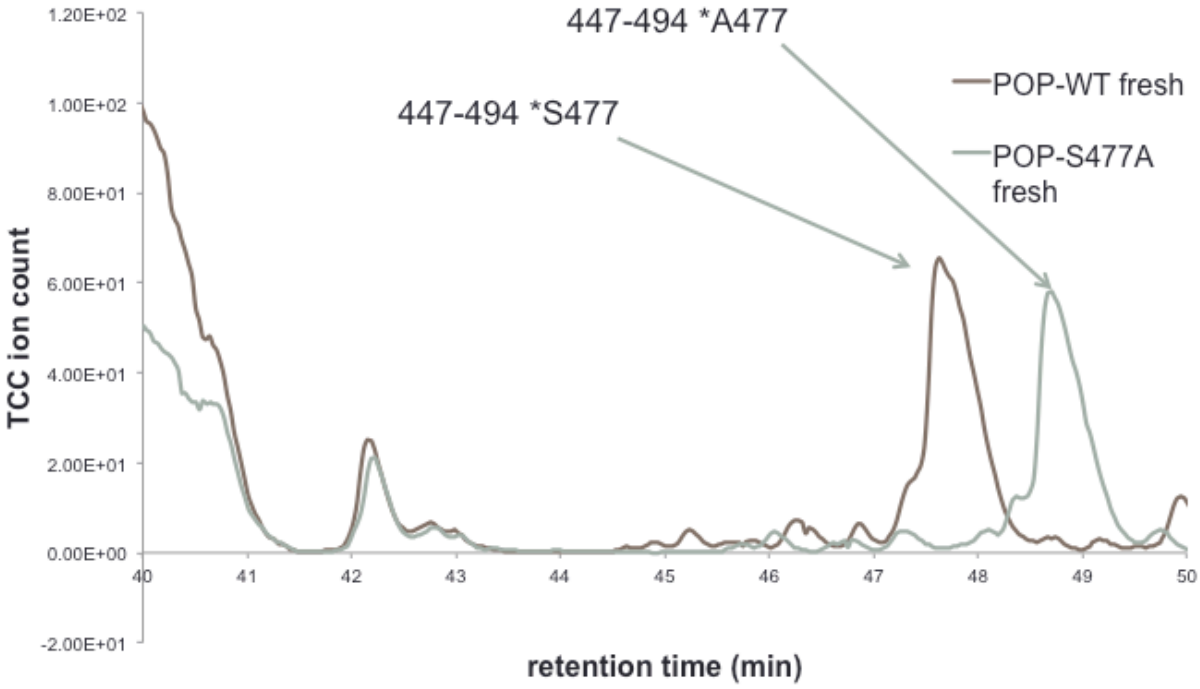


Figure 3.11 LC/Q-TOF-MS chromatogram indicting the ability to resolve two 47-residue-long peptides differing by one hydroxyl group. Peaks correspond to residues 447-494 of *Pfu* POP (brown) and residues 447-494 of *Pfu* POP S477A (gray)

Using this CNBr digestion protocol, more sequence-specific information was obtained concerning the aforementioned hexahistidine tag degradation observed with *Pfu* POP. By comparing the digest chromatograms of freshly-prepared samples of wild-type *Pfu* POP and *Pfu* POP S477A with 10 day-aged samples, we were able to identify several products with masses corresponding to C-terminal peptides with the loss of one, two, or three histidine residues. (Figure 3.12)

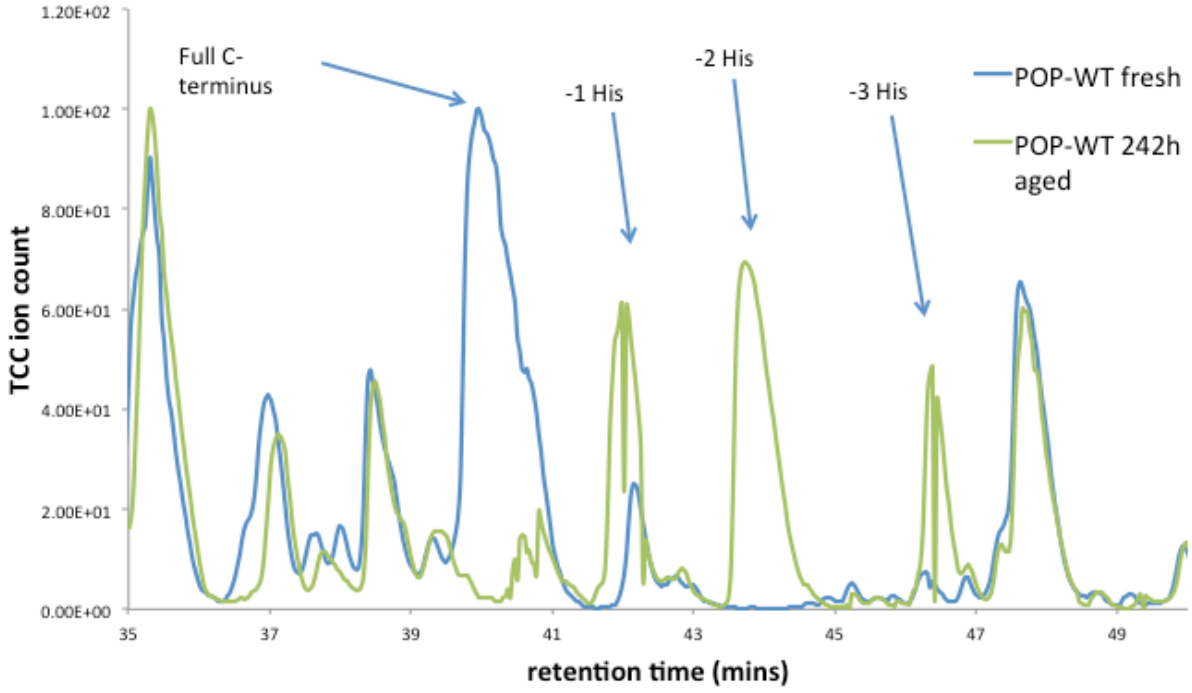


Figure 3.12 LC/Q-TOF-MS chromatogram of CNBr digests of freshly prepared *Pfu* POP versus *Pfu* POP aged for 10 days at room temperature. This corroborated the gradual degradation of the *Pfu* POP C-terminal hexahistidine tag by incremental loss of histidine segments.

The high-accuracy of the observed masses proved unequivocally that degradation of the histidine tag was occurring. In addition, given that the same degradation products

were observed with both *Pfu* POP and *Pfu* POP S477A, we concluded that the degradation was due to chemical hydrolysis, not hydrolysis owing to off-target hydrolase activity of the enzyme. In spite of the excellent reproducibility observed with CNBr digestion of *Pfu* POP and its mutants, it should be noted that the product peptides formed are relatively large, as is highlighted in Figure 3.13. For example, a 215 residue-long fragment, long enough to be considered a protein of its own, is generated by CNBr digestion of *Pfu* POP.

```
10          20          30          40          50          60
MEDPYIWMEN LEDERVLKII EEENKRFREF IGELSDKLFP EVWEQFSQPT IGMARITKKG

         70          80          90         100         110         120
IIASYSEKDR VVIKWFNGDV IVDSKELERE VGDEVLLQGF TTDEEGEKLA YSFSIGGADE

        130         140         150         160         170         180
GITRIIDLKT GEVIEEIKPS IWNITFLKDG YYFTRFYRKE KTPDGVNPPA ARMFWKDREG

        190         200         210         220         230         240
ERMVFGEGLT SGYFMSIRKS SDGKFAIVTL TYGWNQGEVY IGPIDNPQEW KKVYSASVPV

        250         260         270         280         290         300
EAIDVVNGKL YILTKEGKGL GKIIAIKNGK IDEVIPEGEF PLEWAVIVRD KILAGRLVHA

        310         320         330         340         350         360
SYKLEVYTLN GEKIKEITFD VPGSLYPLDK DEERVLLRYT SFTIPYRLYE FKDDLRLIEE

        370         380         390         400         410         420
RKVEGEFRVE EDFATSKDGT KVHYFIVKGE RDEKRAWVFG YGGFNIALTP MFFPQVIPFL

        430         440         450         460         470         480
KRGGTFIMAN LRGGSEYGEE WHRAGMRENK QNVFDDFIAV LEKLKKEGYK VAAWGRSNGG

        490         500         510         520         530         540
LLVSATLTQR PDVMDSALIG YPVIDMLRFH KLYIGSVWIP EYGNPEDPKD REFLLKYSPY

        550         560         570         580         590         600
HNVDPKKKYP PTLIYTGLHD DRVHPAHALK FFMKLKEIGA PVYLRVETKS GHMGASPETR

        610         620
ARELTDLLAF VLKTLSLEHH HHHH
```

Figure 3.13 The primary sequence of *Pfu* POP with methionine positions highlighted. This residue is relatively rare, making CNBr digestion alone a suboptimal digestion method for detailed MS/MS studies.

Large fragment sizes impede efforts to obtained residue-level MS/MS information on POP modification. MS/MS methodology relies on the bombardment of relatively short (>2000 Da) peptides to acquire residue-level fragmentation data.[25] Longer fragments are not amenable to this methodology. For this reason, we sought to generate shorter digest peptide fragments. Luckily, tryptic digestion of *Pfu* POP CNBr peptide digest mixtures generated adequately short peptides with high reproducibility (Figure 3.14). In chapter four, the application of this method to acquire high-resolution information about scaffold modification will be discussed.

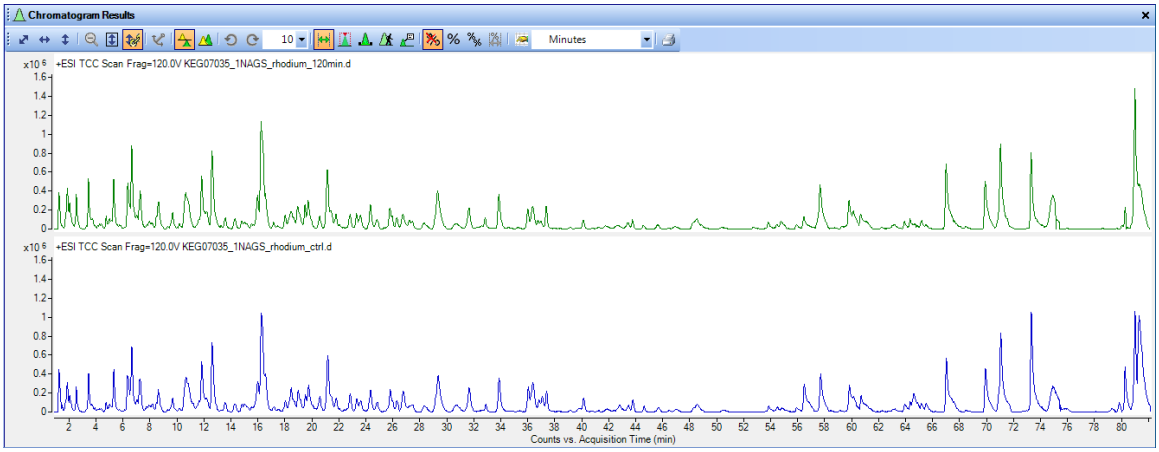


Figure 3.14 LC/MS chromatograms of two separate *Pfu* POP samples indicating the high reproducibility of CNBr/trypsin double digestion.

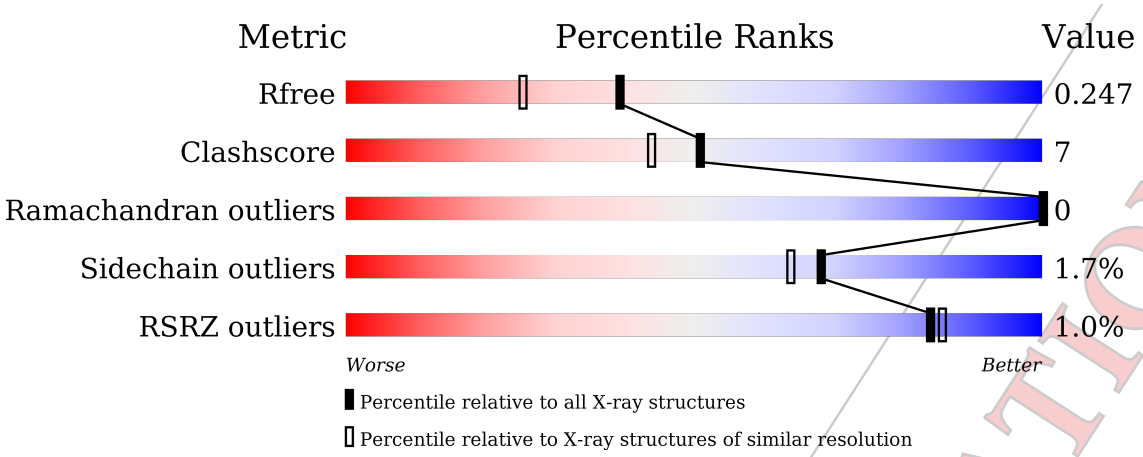
3.2.2 PYROCOCCUS FURIOSUS PROLYL OLIGOPEPTIDASE

3.2.2.1 X-RAY CRYSTALLOGRAPHY

One of the main deficiencies of the initial selection of *Pfu* POP as an ArM scaffold was the lack of a crystal structure of this enzyme. This was prohibitive for making well-informed rational mutations and prevented the rationalization of mutations

that arose through random mutagenesis (as discussed in chapter four). For this reason, once *Pfu* POP became a standardized ArM scaffold in this laboratory, its structural characterization became a top priority. After a comprehensive screening effort, the native *Pfu* POP (i.e. not containing a C-terminal hexahistine tag) of *Pfu* POP and its mutant *Pfu* POP S477C were crystallized under similar conditions, yielding high quality X-ray diffraction with resolutions of 1.9 Å and 2.2 Å, respectively.

Solving these structures posed a major challenge due to the poor quality of the molecular replacement solutions obtained with a variety of homologous search models (PDB: 1QFS, 1E8M, 2YRT). As expected, molecular replacement succeeded for the adequate placement of the hydrolase domain of the enzyme, which is the most conserved segment of POP-family enzymes. Poor placement of the β domain was observed, however. Our collaborator Dr. Narayanasami Sukumar determined that the initial struggles with model building and refinement were due to crystallographic polymorphism. Reprocessing the crystallographic data in an alternate P1 spacegroup and application of non-crystallographic symmetry-averaged maps yielded higher electron density corresponding to the β domain, enabling rebuilding of this segment of the model. Iterative rounds of model building and refinement were applied, while iterative build OMIT maps[26] were used for the removal of model bias from the ED map. The final refinement statistics for the native *Pfu* POP crystal structure are shown in Table 3.1.



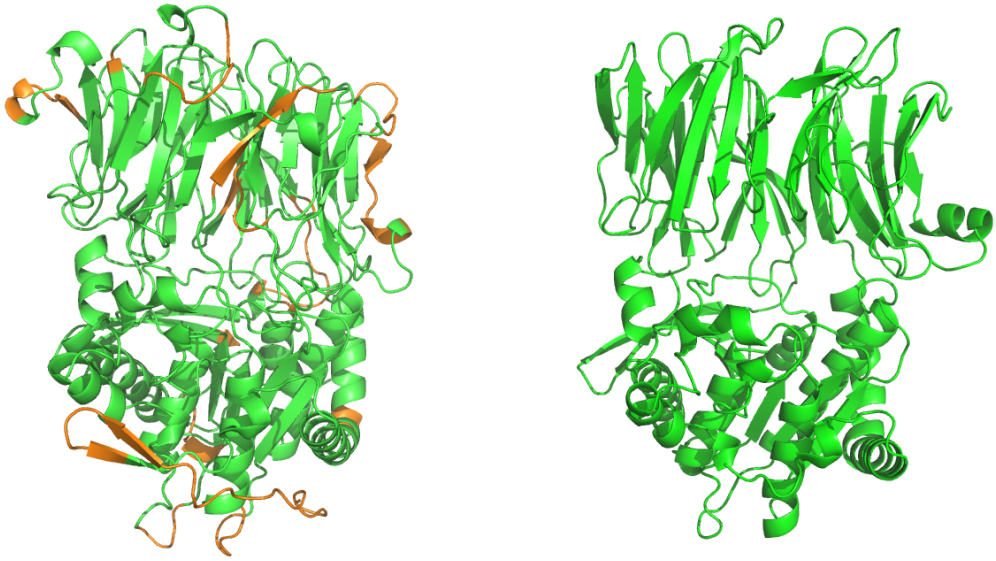
Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	55.53Å 176.76Å 57.90Å 90.00° 106.03° 90.00°	Depositor
Resolution (Å)	47.09 – 1.90 47.09 – 1.90	Depositor EDS
% Data completeness (in resolution range)	99.5 (47.09-1.90) 93.1 (47.09-1.90)	Depositor EDS
R_{merge}	0.17	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.03 (at 1.90Å)	Xtriage
Refinement program	PHENIX (1.10_2155: ???)	Depositor
R, R_{free}	0.193 , 0.247 0.193 , 0.247	Depositor DCC
R_{free} test set	2002 reflections (2.40%)	DCC
Wilson B-factor (Å^2)	22.9	Xtriage
Anisotropy	0.299	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.40 , 50.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	0.086 for l,-k,h	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	21236	wwPDB-VP
Average B, all atoms (Å^2)	34.0	wwPDB-VP

Table 3.1 Crystal structure collection and refinement statistics.

3.2.2.1.1 OVERALL STRUCTURE

Like other POPs, *Pfu* POP features a two-domain architecture, with an α/β-hydrolase domain capped by a 7-blade β-propeller domain. The N-terminus of the enzyme (residues 1-47) consists of an alpha-helical segment that is wrapped around the large C-terminal hydrolase domain (residues 367-616). The intermediate segment (residues 48-366) comprises the β β -barrel domain. Like its mesophilic homologues, the first and seventh blade of the propeller domain are joined by hydrophobic interactions between the blade surfaces. This stands in contrast to the "Velcro" of hydrogen bonds or disulfides implicated in the stabilization of many β -barrels proteins. The hydrolase and β -barrel domains are joined covalently through a hinge region comprised of residues 47-50 and residues 361-367.

Pfu POP is significantly more compact than its mesophilic homologues, at 71,889 Da versus the average of ~80,000 Da observed for mesophiles. Compactness in structure has often been implicated as a mechanism for thermostabilization in hyperthermophilic enzymes.[27] BLAST alignment of *Pfu* POP with its porcine homologue reveals that it achieves its relative compactness by abridging its N-terminal region and by systematically shortening and omitting loops and β -strands relative to its porcine homologue. Comparing the two enzymes reveals that *Pfu* POP is significantly smaller in both volume and cross-section than its porcine homologue.[28] (Figure 3.15) The compact internal volume of *Pfu* POP is an appealing feature in its application as an ArM scaffold, as this may help in conformational restriction of enclosed cofactors.



Porcine POP *Pfu* POP

Figure 3.15 The side-by-side comparison of structural deletions (in orange) of porcine POP (1QFS) versus *Pfu* POP. These deletions are mostly clustered in loop regions.

3.2.2.1.2 OPEN STRUCTURE

Mammalian POPs (*Sus scrofa* and *Homo sapiens*) have been crystallized exclusively in a closed conformation,[28,29] while various bacterial POPs (*Myxococcus xanthus, Sphingomonas capsulata*, and *Aeromonas punctata*) have been captured in both closed and open conformations.[30,31] The archaeal *Pyrococcus furiosus* POP reveals an open structure, with the two individual chains in the unit cell exhibiting slightly different interdomain angles. Interestingly, the interdomain angle of this species is lower than that observed in its bacterial homologues. Small differences in sidechain torsion angles and loop positions can be observed between the two protein chains as well. Li and coworkers have previously crystallized *Aeromonas punctata* POP in varying degrees of openness,[31]indicating that this is not a unique phenomenon with these enzymes. This

open/closed conformational heterogeneity gives a glimpse into the span of conformations available to *Pfu* POP. Concurrent changes to the active site on the enzyme opening and closing can be extrapolated based on the cystallized homologues of *Pfu* POP. Altogether, however, the implications of interdomain opening and closing are not entirely positive in the context of ArM formations with the *Pfu* POP scaffold, as this suggests that cofactor enclosure might not be as simple as expected.

3.2.2.1.3 ACTIVE SITE

Much of the active site of *Pfu* POP retains features common to prolyl oligopeptidases. The hydrophobic proline-binding site, comprising Trp518, Phe404, and Tyr522, is effectively superimposable with that of other POPs. Likewise, the active-site serine Ser477 is predictably positioned on the nucleophile elbow, a common feature among serine hydrolases. Loop-(560-564), containing the catalytic Asp560, shows a greater degree of conformational freedom based on its B-factors. Crucially, this loop also contains Arg562, which is conserved across all POPs plays an important role in substrate binding. In closed structures, Arg562 forms a salt bridge with the β-domain-localized Asp119, stabilizing it in an optimal position to hydrogen bond with substrate.[32] Upon opening, Arg562 has been demonstrated to rotate to ion pair with Asp560, occupying the vacant position of the highly mobile catalytic His592.[31,32]. In agreement with this model, Arg562 shows high side-chain B-factors and is held in two distinct rotamer positions on either chain. Asp119, which resides on the flexible loop-(115-121), is also characterized by high temperature factors and holds two distinct positions on either chain.

As with all previously characterized POPs, the catalytic His592 resides on a highly flexible loop (residues 590-596), as is determined by its relatively high B-factor.

In existing POP family crystal structures, the histidine loop tends to be ordered and well-resolved in the electron density when the enzyme is in its closed conformation.[28,30,32,33] In contrast, this loop has been consistently omitted in open structures due to its high disorder.[30,33] The structure of *Pfu* POP offers a rare view of the position of the catalytic histidine, which, despite its high temperature factor, can be seen retracted away from its presumed position in the closed, catalytically competent conformation (Figure 3.16). Coupling of interdomain angle with His-loop position has been postulated previously.[34]

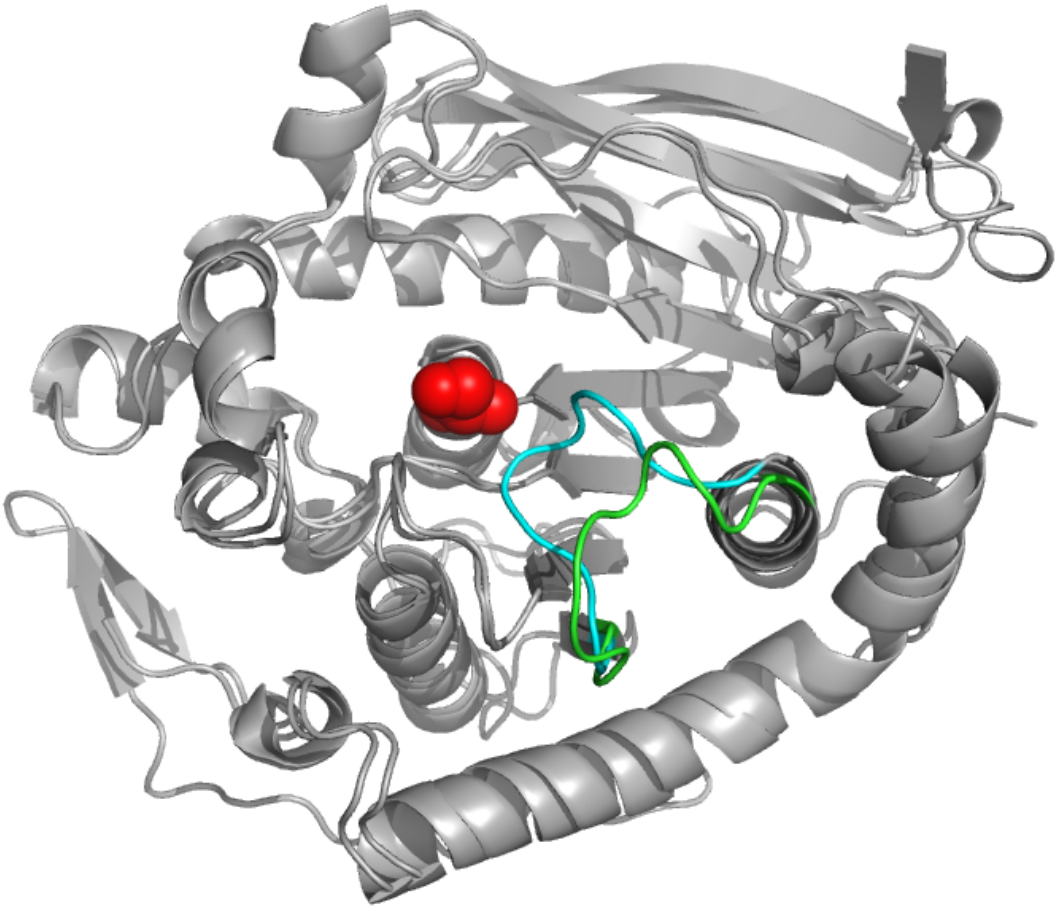


Figure 3.16 Alignment of the hydrolase domains of *Pfu* POP and porcine POP show the His-loop of *Pfu* POP (green) significantly retracted away from S477 (red spheres) as compared to the His-loop of porcine POP (cyan).

An interesting structural feature of *Pfu* POP is the loop connecting the β-strands of blade 3. This loop corresponds to the position of "loop A", or the hydrophilic loop, as it has been referred to in the porcine model (Figure 3.17).[34,35] This loop is significantly shorter in *Pfu* POP (residues 157-170) than in previously characterized homologues. It also is decidedly more rigid, with three prolines (P163, P168, P169) restricting its conformational flexibility (as determined by its low observed B-factor). Two sequential prolines (P168 and P169) lock the loop in a position projected away from the hydrolase domain and folded up against the β-domain. This position is further reinforced by a stabilizing salt bridge between the tip of the loop (Asp164) and the β-domain (Arg172). In contrast, other homologues show this loop closely latched to the hydrolase domain (in closed structures) or highly disordered in bulk solvent (open structures).[28,29,31] Based on the temperature factors observed on both chains of the enzyme, it is unlikely that this loop undergoes major conformational changes during interdomain opening or closing, suggesting a different role for this loop relative to that of its mesophilic cousins.

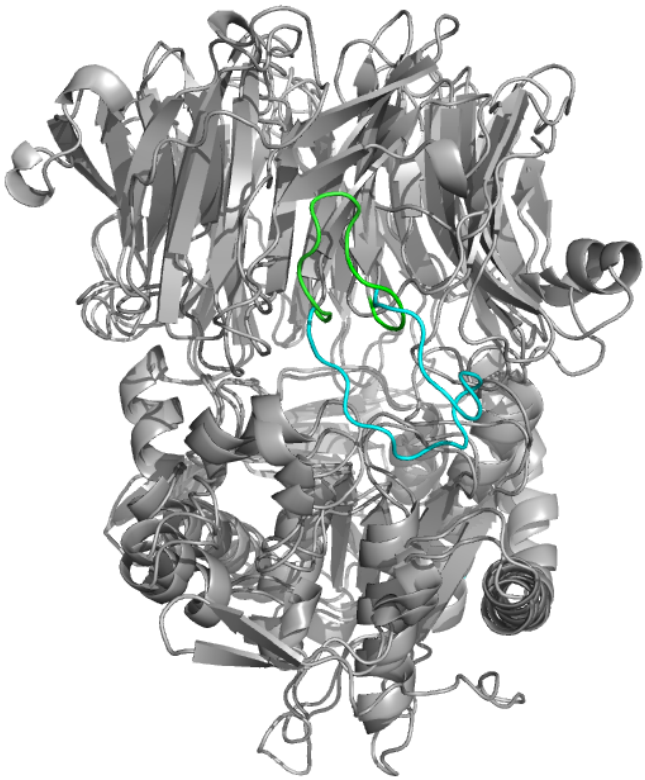


Figure 3.17 Alignment of *Pfu* POP with porcine POP (1QFS) reveals major differences in the position of loop A. *Pfu* POP loop A (residues 157-170, green) is much shorter and remains wrapped onto the β-domain. Porcine POP loop A (cyan) is significantly longer and comes into contact with the hydrolase domain.

Aligning the β-domain of *Pfu* POP with *Ap* or porcine POP reveals that this loop projects the side chain of Arg158 into the position typically occupied by *Ap* POP K173 or porcine K172. In these structures, the cationic side chain provides a salt bridge to the conserved Asp561. Asp561 sits between two functionally important residues (Asp560 and Arg562, discussed above) and is likely an important component of the extensive hydrogen bond network that promotes optimal active site geometry (Figure 3.18).[19] In contrast to the proline-loop position of Arg158, *Ap* and porcine Lys172(173) resides on a

connecting loop between blades 2 and 3, a segment that is not conserved in *Pfu* POP. Spatial conservation of charged residues has been reported previously.[36]

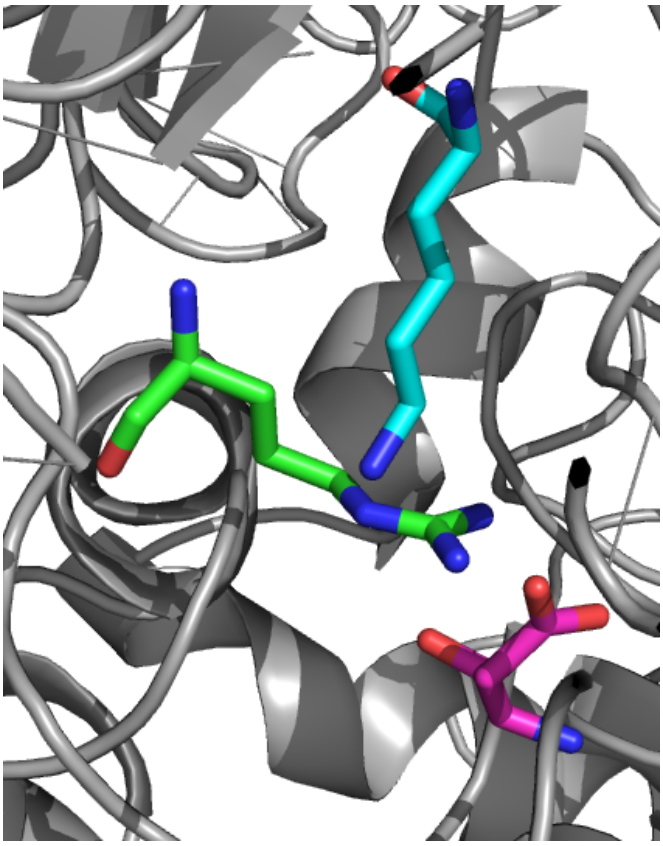


Figure 3.18 An overlay of Arg158 of *Pfu* POP (green) with Lys172 of porcine POP (cyan) as it relates to the salt bridge formed with Asp561 (pink).

While knowledge of residue positions with POP will greatly aid in mutagenesis efforts, it uncertain how the fine details of the native *Pfu* POP active site will impact ArM design. It should be noted, however, that POP family enzymes that have been crystallized with mechanism-based inhibitors have generally been found to be closed in structure, with the active site histidine forming a hydrogen bond with the serine/phosphonate. This might signal an encouraging effect—the use of phosphonate-linked cofactors may in fact force the enzyme to adopt a closed conformation, driving encapsulation of the metal.

Efforts to solve the structure of ArMs containing such cofactors are underway, and the apo structure described here will greatly facilitate solution of those structures.

3.2.2.1.4 HALIDE BINDING

Prolyl oligopeptidase has long been known to exhibit marked sensitivity to ionic strength.[37] Increasing salt concentrations up to 2.0 M has been shown to increase the specificity constant of the enzyme. Harris et al. demonstrated that *Pfu* POP shows a similar ionic strength dependence to that observed in the porcine ortholog, with the hypothesis that this arose through explicit halide binding events proximal to the enzyme active site.[17] In our labotory, high concentrations of NaBr were also shown to encourage greater enantioselectivity in cyclopropanations catalyzed by *Pfu* POP dirhodium ArMs, as discussed in chapter four.

Corroborating the observed effects of halide and putative halide binding, the active site of *Pfu* POP contains two clearly bound chloride ions in this crystal structure (Figure 3.19) It is unclear what functional role these chlorides play in the active site, though from their relative B-factors they appear to be well-bound, forming part of an extended hydrogen bond network in the active site. Arg476 and Arg600 appear to both be engaged in salt bridging of these chlorides. Such a structured active site might help explain why NaBr enhances selectivity of the *Pfu* POP-dirhodium catalyzed cyclopropanation, as cofactor rigidifaction may be occurring due to engagement of the adjacent Arg476 in a stable salt bridge.

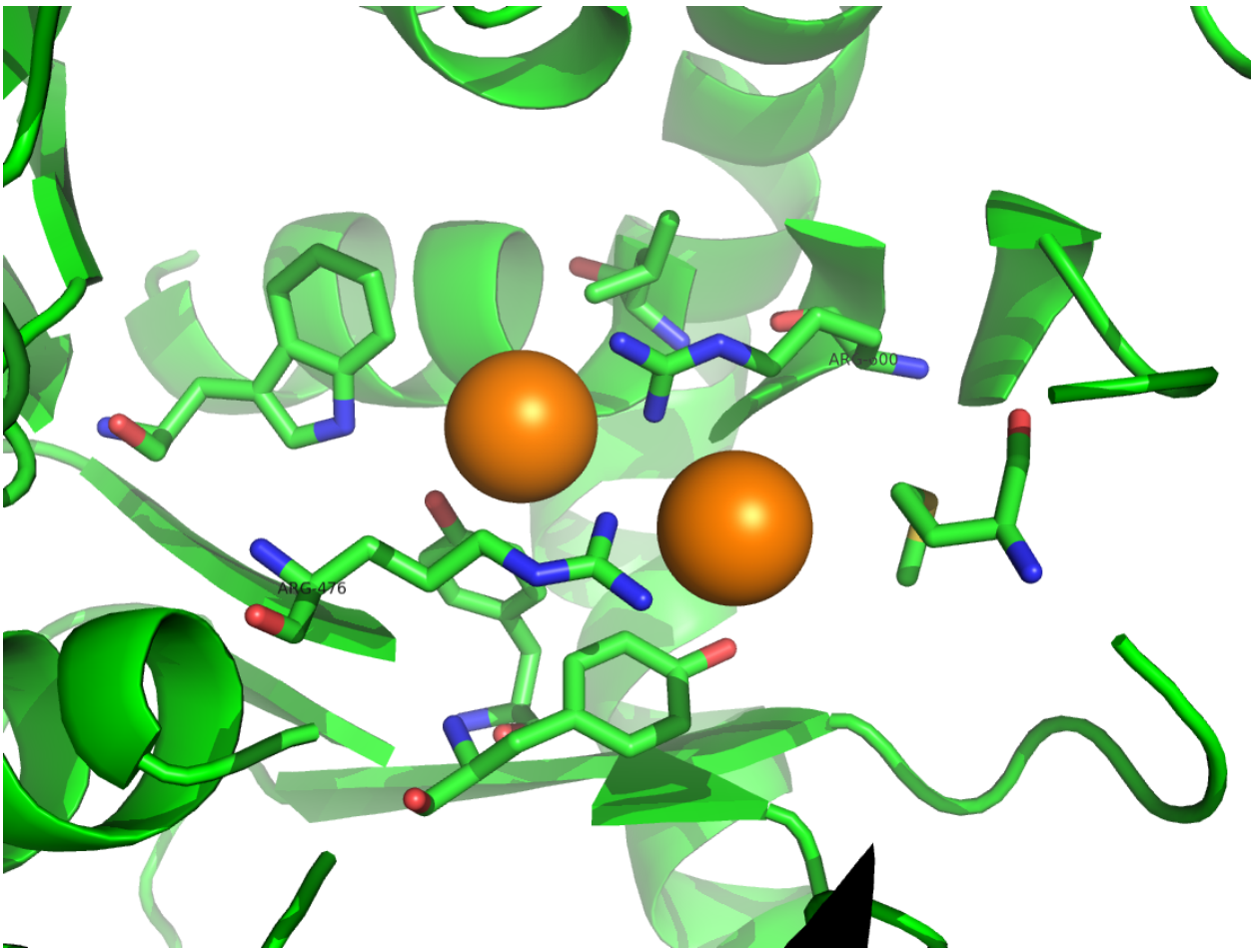


Figure 3.19 Chlorides (orange) bound near the active site of *Pfu* POP

3.2.2.2 AUTOPROTEOLYTIC ACTIVITY

It is important to note that the crystal structure of *Pfu* POP contains an indeterminate proline-containing peptide loosely bound to the active site (Figure 3.20). We are not certain of the identity of this peptide, though it is not unreasonable to expect that some degree of autoproteolysis/chemical hydrolysis could occur under the crystallization conditions. The current assignment of a free proline is based on the best fit for the density observed, though further studies will need to be carried out to definitively establish the identity of this substrate.

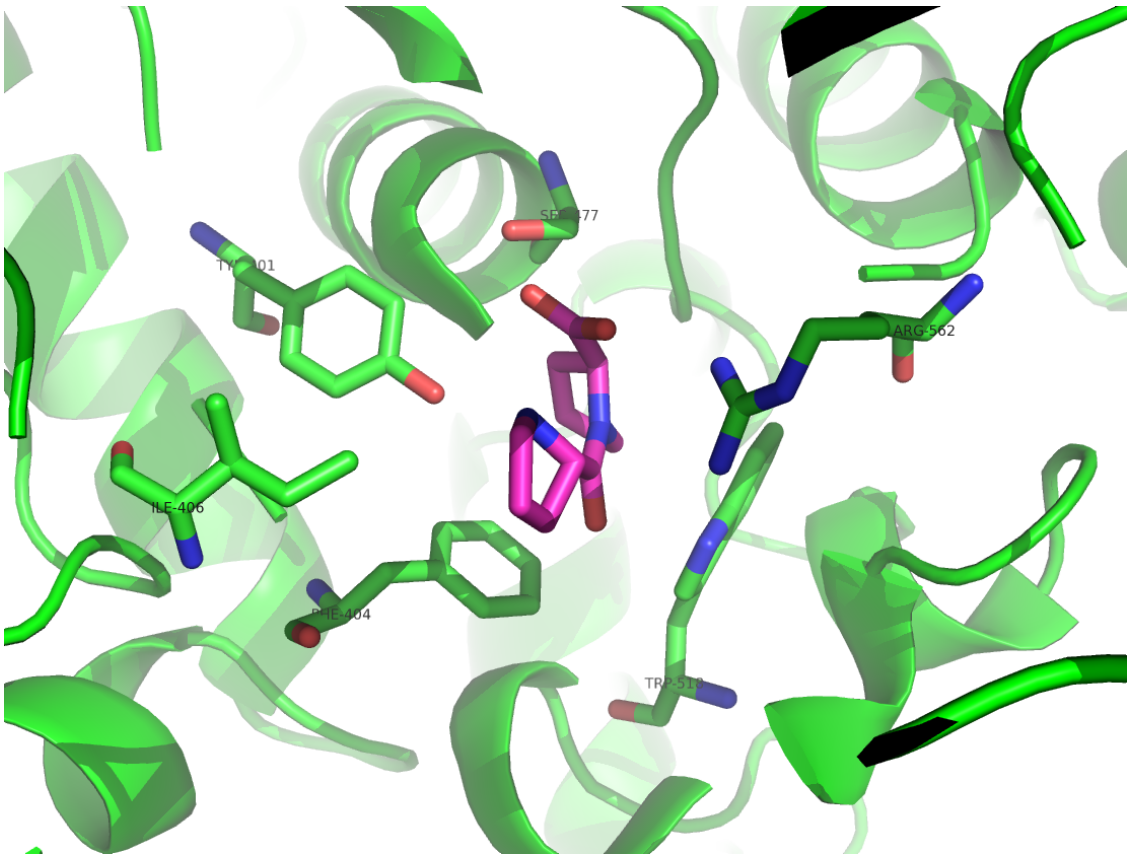


Figure 3.20 Bound Pro-Pro peptide (magenta) in the active site of *Pfu* POP.

Harwood and coworkers first hypothesized the autoproteolysis of *Pfu* POP in 1997.[38] To confirm this effect, we investigated the autoproteolytic activity of *Pfu* POP under incubation at room temperature and 95°C. With overnight incubation of the enzyme at pH 9.0 significant conversion to a slightly protein species could be observed by SDS PAGE (Figure 3.21). HR-ESI-MS analysis indicated that protein cleavage occurred predominantly at three sites: Asn77, Asn524, and Pro502. The two asparagine sites could be ascribed to thermal hydrolysis, given that these occurred with both wild type *Pfu* POP and the catalytically inactivated S477A mutant. Thermal hydrolysis of polypeptides at the C-terminus of asparagine residues is well documented.[27] Given the

observed specificity of the thermal hydrolysis, which occurs at a temperature within the physiological norm of *Pyrococcus furiosus*, it is tempting to ascribe a functional role to these "weak points" in the polypeptide chain.[39,40] This question remains to be investigated. More importantly, peptide hydrolysis can be observed at the C-terminus of P502, the products of which are only detected in the presence of the catalytically competent wild type POP. This provides evidence that autoproteolysis of *Pfu* POP indeed does occur. This is highly surprising, particularly given the relatively buried position of P502. It is unclear whether enzymatic cleavage at this site is an inter- or intra-molecular event.

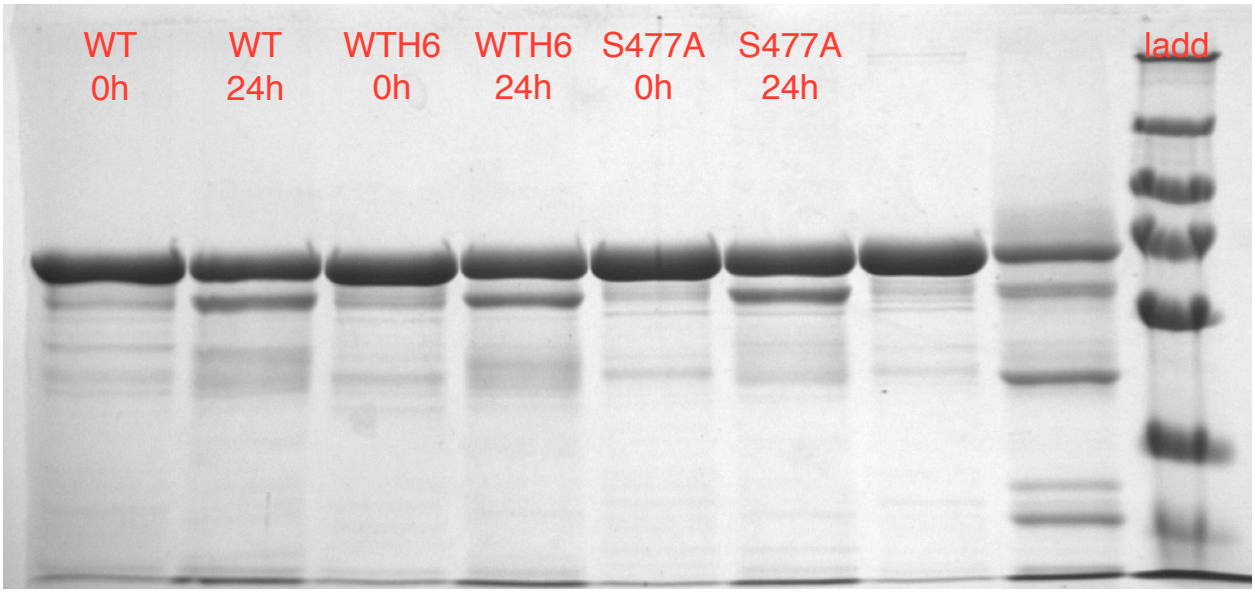


Figure 3.21 SDS-PAGE of different POP variants before and after incubation at 95°C for 24 hours. Clear development of a shorter protein product is observed.

Prolonged room temperature incubation (18 weeks) of wild type and S477A *Pfu* POP also results in protein cleavage, though the products are different from those observed at high temperature. Chemical lysis of the C-terminal hexahistidine tag is a

prominent feature, with dominant cleavage at His621 and His622. Non-enzymatic cleavage at His541 can be observed as well, resulting in a fragment that, in the presence of catalytically-competent wild type POP, is further cleaved on the C-terminus of Val543. This off-target enzymatic hydrolysis is unexpected but not unprecedented for POPs.

The observations made here suggest numerous pathways for main-chain cleavage of *Pfu* POP, a fact that may influence the dynamic and structural nature of the enzyme, impacting its properties as an ArM scaffold. It should be noted that these cleavage events apply only under relatively forcing conditions (prolonged room temperature and high temperature incubation) and only to a small extent. The majority of the protein species remain intact, speaking to the remarkable robustness of *Pfu* POP.

3.2.2.3 MOLECULAR DYNAMICS: MECHANISM OF THERMOSTABILIZATION

A primary mechanism for thermostabilization by hyperthermophilic enzymes is through ion pairing. Typically, the contribution of ion pairs to the thermodynamic stability of and enzyme is marginally, since the energetic gain of the electrostatic ionic interaction is readily offset by the penalty of desolvation of charged residues.[41] This tradeoff holds true at ambient temperatures. At higher temperatures, however, desolvation is less energetically costly.[39] Thermostable enzymes tend to feature large networks of ion pairs, as ion pair networking lowers the overall entropic cost of constraining residues to ion pairing.[42] In light of this, we pursued a molecular dynamics-based study of the dynamic networking of ion pairs in *Pfu* POP as it compares to its mesophilic porcine homologue. The crystal structures of *Pfu* POP and porcine POP

(PDB: 1QFS) were simulated for 100 ns at 298 K. Using the methods described by Papaleo[43] , we conducted a thorough analysis of the dynamic ion pair networks maintained by *Pfu* POP and porcine POP (Figure 3.22). By this metric, it was clear that the degree of ion pairing and the extent of ion pair networking was significantly greater in the smaller *Pfu* POP. *Pfu* POP retained 0.146 ion pairs/residue in contrast to porcine POP, which retained only 0.086 ion pairs/residue. In addition, the largest maintained ion pair network was comprised of 7 ion pairs in *Pfu* POP in contrast to 4 ion pairs in porcine POP (Table 3.2). These results corroborated the findings of similar studies comparing thermophilic and mesophilic enzymes.[41]

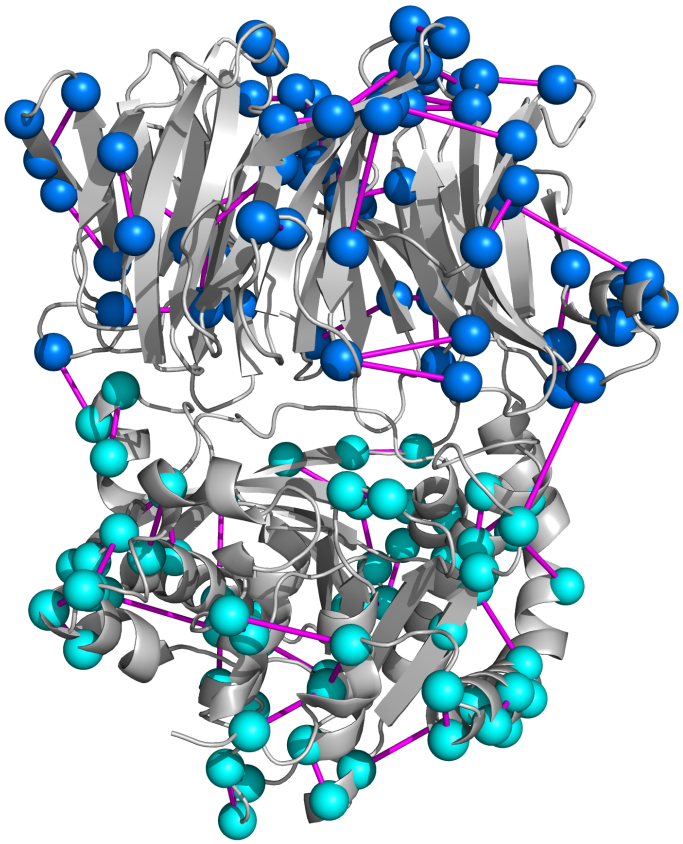


Figure 3.22 Ion pairs maintained over 50% of the 100 ns MD simulation period.

	S. scrofa	***P. furiosus***
# of residues	710	616
# of ion pairs	61	90
IPs per residue	0.086	0.146
Largest network	4 residues	7 residues

Table 3.2 A comparison of ion pairing parameters determined by analysis of protein structure networks[44] and the number of ion pairs maintained over 50% of the time during a 100 ns MD simulation.

Other mechanisms of thermostabilization have also been investigated. The proline content of *Pfu* POP is also significantly higher than that of its porcine homologue, another feature of hyperthermostabilization.[39] Furthermore, as mentioned, we have also noted increased compactness and lower flexible loop content of *Pfu* POP, which seems to be a common feature among hyperthermostable enzymes.[39] Finally, we are looking to quantitate the extent of hydrophobic networking in *Pfu* POP, as large hydrophobic networks are often implicated in the thermostabilization of enzymes.[45]

3.3 CONCLUSIONS

Pyrococcus furiosus prolyl oligopeptidase has been demonstrated to be an exceptionally convenient host scaffold for the formation of artificial metalloenzymes

through a variety of methods. Its ability to enclose large metal cofactors, its high stability, and its ease-of-use in the laboratory make it an ideal platform for the study of artificial metalloenzyme catalysis. The work described in this chapter serves as a primer to illustrate the many aspects of its utility as a protein scaffold, as well as a guide for its standard use in the development of artificial metalloenzymes. This scaffold is amenable to a number of methods for routine characterization of ArMs

In addition, *Pfu* POP has finally been structurally characterized in its native form at 1.9 Å resolution. Aside from its fundamental utility in elucidating the enzymology of POPs, this crystal structure serves as a map to guide future rational design efforts as well as a reference for the rationalization of new observed effects in *Pfu* POP ArM catalysis.

3.4 EXPERIMENTAL

Materials

Unless otherwise noted, all reagents were obtained from commercial suppliers and used without further purification. Luria broth (LB; Cat# L24040), rich medium (2YT; Cat# X15600) and Agar (Cat# A20020) were purchased from Research Products International, Corp (Mt. Prospect, IL). Amicon® Ultra-15 Centrifugal Filter Units with Ultracel-30 membrane (Cat# UFC903024) were purchased from EMD Millipore (Billerica, MA) and used according to manufacturer's recommendations. 24-well crystallization trays and coverslips for protein crystallization were purchased from Hampton Research (Aliso Viejo, CA). High resolution ESI mass spectra were obtained using an Agilent Technologies 6224 TOF LC/MS and an Agilent Technologies 6540 Q-TOF MS-MS. Low resolution ESI mass spectra were obtained using an Agilent Technologies 6130 LC-

MS. LC/MS experiments were performed using an Aeris 3.6 μm WIDEPORE C4 200 Å LC Column 150 x 4.6 mm. LC-MS/MS experiments were performed with an Easy-nLC 1000 ultra-high pressure LC system (ThermoFisher) using a PepMap RSLC C18 column (column: 75 μm x 15 cm; 3 μm, 100 Å) coupled to a Q Exactive HF Orbitrap and Easy-Spray Nanosource (ThermoFisher). Circular dichroism (CD) spectra were obtained on a JASCO J-1500 CD Spectrometer.

Experimental Procedures

Expression of POP

Anion exchange (native *Pfu* POP): Chemically-competent Rosetta2(DE3) (Novagen) cells were transformed with plasmid pJS3 obtained from Prof. Harold Schreier (UMBC)[9]. Transformants were allowed to recover in SOC medium (37°C, 1 hour) and the mixture was spread onto agar plates containing ampicillin and chloramphenicol (6.25 g LB powder mix, 4 g agar, 250 ml DDI water, 100 ug/g ampicillin + 25 ug/g chloramphenicol) and the plates were incubated at 37°C for 16 hours. Colonies containing the pJS3 plasmid were obtained and inoculated into 5 mL primary cultures containing 100 ug/g ampicillin + 25 ug/g chloramphenicol in LB media. These were incubated overnight at 37°C with 250 rpm shaking. The following day, the primary cultures were used to inoculate 1 L of LB media containing 100 ug/g ampicillin + 25 ug/g chloramphenicol). These were incubated at 37°C with 250 rpm shaking for ~5 hours until reaching an OD_{600}≈1.0. Overexpression of *Pfu* POP was induced by addition of 1 mM isopropylthiogalactoside. The expression cultures were incubated at 37°C with 250 rpm shaking for an additional 18 hours. Cells were harvested by centrifugation at 3000 rpm

for 10 minutes. They were resuspended in 100 mL 20 mM sodium phosphate buffer pH 6.5 and lysed by sonication. The cell lysate was clarified at 15000 rpm for 1 hour. Clarified lysate was applied to a 3 x 5 mL HiTrap QFF anion exchange columns pre-equilibrated with 20 mM sodium phosphate buffer pH 6.5. The protein was purified using a gradient of 0-500 mM NaCl in 20 mM sodium phosphate buffer pH 6.5. Fraction purity was confirmed by SDS-PAGE. Pure fractions were combined and concentrated using Amicon® 30 kD spin filters. Pure *Pfu* POP was buffer exchanged into H_2O at least three times, snap frozen with liquid N_2, and stored at -20°C.

NiNTA chromatography (C-terminal His-tag codon-optimized *Pfu* POP): Codon-optimized pET28a-POP was transformed into electrocompetent *E. coli* BL21 (DE3). Transformed cells were allowed to recover in SOC medium (37 °C, 1 hour), then plated onto LB kanamycin agar plates (6.25 g LB powder mix, 4 g agar, 250 mL DDI water, 0.05 mg/mL kanamycin) and incubated at 37 °C for 18 h. Several colonies appeared on overnight-incubated plates; a single colony from this plate was inoculated in 5 mL 2YT medium containing 0.05 mg/mL kanamycin. 5 mL primary cultures containing 50 ug/mL kanamycin in LB media were incubated overnight at 37°C with 250 rpm shaking. The following day, the primary cultures were used to inoculate 1 L of LB media containing 50 ug/mL kanamycin. These were incubated at 37°C with 250 rpm shaking for ~5 hours until reaching an $OD_{600}\approx1.0$. Overexpression of *Pab* variants was induced by addition of 1 mM isopropylthiogalactoside. The expression cultures were incubated at 37°C with 250 rpm shaking for an additional 18 hours. Cells were harvested by centrifugation at 3000 rpm for 10 minutes. They were resuspended in 100 mL 20 mM sodium phosphate buffer

pH 7.4/50 mM imidazole/50 mM NaCl and lysed by sonication. The cell lysate was clarified at 15000 rpm for 1 hour. Clarified lysate was applied to a 3 x 5 mL HisTrap NiNTA columns (GE Healthcare Life Sciences) pre-equilibrated with 20 mM sodium phosphate buffer pH 7.4/50 mM imidazole/50 mM NaCl. The protein was purified using step elution from 20 mM sodium phosphate buffer pH 7.4/500 mM imidazole/50 mM NaCl.. Fraction purity was confirmed by SDS-PAGE. Pure fractions were combined and concentrated using Amicon® 30 kD spin filters. Pure *Pfu* POP variants were buffer exchanged into H_2O at least three times, snap frozen with liquid N_2, and stored at -20°C.

Circular Dichroism (CD) stability profiles

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. Solvent/detergent stability assays were performed by pre-incubation of 10 uM protein samples prepared at 5 minutes under the trial conditions probed. Full CD sweep spectra were acquired

Standard native peptidase activity assays

Activity assays were adapted from protocols described by Harwood[17] and Polgár[46]. Reactions were conducted on 2000 uL scale. 1 uL of 20 uM *Pfu* POP in 20 mM NaPi pH

7.4 was added to 1799 uL buffer (20 mM NaPi pH 7.4 as a standard). The reactions were allowed to thermally equilibrate (if temperature dependence was being probed.) The reaction was initiated by the addition of 200 uL 5 mM Z-Gly-Pro-pNA or 5 mM Z-Gly-Pro-Nap. The reactions were monitored by UV/Vis at 410 nm (in the case of Z-Gly-Pro-pNA) or with fluorescence excitation at 340 nm and detection at 410 nm (in the case of Z-Gly-Pro-2-naphthylamine). The final concentrations were 10 nM *Pfu* POP, 500 uM substrate, 10% v/v MeOH.

MS Characterization of POP and hybrids

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. For LC/ESI-Q-TOF and low-resolution LC/MS analysis, 10 uL samples of 60 uM protein in H_2O or 100 mM NaPi pH 7.0 were injected onto the Aeris WIDEPORE column using a standard gradient from 10% ACN/90% H_2O + 0.1% formic acid to 90% ACN/10% H_2O + 0.1% formic acid over 15 minutes. Raw ESI spectra were deconvoluted using the Agilent Chemstation LC/MSD data deconvolution module or Agilent MassHunter.

Digestion protocols

CNBr digestions were performed according to a modified literature procedure[47]: 100 μL of 60 uM protein sample in 100 mM NaPi pH 7.0 was treated with 37.7 μL 1 N HCl and 4.8 μL 1 mM neurotensin standard. Then, the samples were wrapped in aluminum foil, at

which point 7.5 µL 5.0 M CNBr in acetonitrile was added to the mixture. These samples were incubated with 250 rpm shaking for 16 hours (fully wrapped in aluminum foil). Once the digestions were completed, they were concentrated on a Speedvac at 55°C to remove solvent and excess CNBr. The residues were then resuspended in 150 µL H_2O. These were analyzed by LC/Q-TOF MS using an Aeris WIDEPORE C8 peptide/protein column. The mobile phase consisted of: Solvent A: [H_2O + 0.1% formic acid] and Solvent B: [MeCN + 0.1% formic acid]. 20 uL of sample were analyzed using a linear gradient from 3% B to 70% B over 84 minutes.

Crystallization

Crystals of native (non-His-tagged) wild-type and S477C *Pyrococcus furiosus* prolyl oligopeptidase were grown at 20°C using the hanging-drop vapor-diffusion technique. A highly purified (>95%, determined by SDS-PAGE) protein solution of 8 mg/mL (determined by A_{280}) in H_2O was used. Crystallization drops contained 1 uL protein solution with 1 uL of well solution containing 30% PEG 8K and 100 mM Tris pH 9.0. Diffraction quality crystals were obtained after 4 months of incubation. Crystals were harvested and cryoprotected in a solution of 15% glycerol in well solution. These were then flash frozen in liquid nitrogen and mounted for data collection.

Crystal Data Collection and Refinement

Crystallographic data were collected on NE-CAT 24 ID-C and ID-E beamlines at the Advanced Photon Source at Argonne National Laboratory. The diffraction data were obtained at 100 K using a wavelength of 0.979 Å and an ADSC QUANTUM 315

detector. The data were indexed, integrated and scaled in HKL2000 to 2.20 Å resolution for the *Pfu* POP S477C mutant and to 1.90 Å resolution for wild-type *Pfu* POP. Crystal polymorphism was observed, requiring reprocessing in P1 space group with larger cell dimension in order to achieve adequate refinement. *Myxococcus xanthus* prolyl oligopeptidase[30] (PDB ID: 2BKL) provided a viable molecular replacement solution in PHASER-MR[48], yielding an initial set of phases. Phenix.refine[49] was employed for refinement. Strong NCS restraints were used initially in refinement, with gradual loosening of these restraints as R_{free} decreased. Rounds of refinement and model building in COOT[50] were pursued until R_{free} and R_{work} converged and Ramachandran/Clashscore parameters had acceptable scores. Refinement statistics for this solution can be found in Table 3.3.

Molecular Dynamics and Ion Pair Networks

Molecular dynamics simulations were performed using the *Pfu* POP S477C model determined by X-ray crystallography as well as the porcine POP model (PDB: 1QFS). NAMD simulation input files were generated using CHARMM-GUI. [51] The models were first solvated and equilibrated for 10,000 steps, followed by a 100 ns production simulation. The persistence of ion pairing interactions in both models was compared using methodology outlined by Papaleo et al.[44]

3.5 ACKNOWLEDGEMENTS

Special thanks to Dr. Narayanasami Sukumar for assistance with protein crystallography. Thanks to Dr. Sunhwan Jo (Roux lab) for molecular dynamics

simulations. Dr. Hyun June Park carried out comparative molecular dynamics and ion-pair network analysis.

3.6 REFERENCES

(1) Basauri Molina, M.; Verhoeven, D. G. A.; van Schaik, A. J.; Kleijn, H.; Klein Gebbink, R. J. M. *Chem. Eur. J.* **2015**, *21* (44), 15676–15685.

(2) Matsuo, T.; Imai, C.; Yoshida, T.; Saito, T.; Hayashi, T.; Hirota, S. *Chemical Communications* **2012**, *48* (11), 1662–1664.

(3) Liu, Y.; Patricelli, M. P.; Cravatt, B. F. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96* (26), 14694–14699.

(4) Simon, G. M.; Cravatt, B. F. *J. Biol. Chem.* **2010**, *285* (15), 11051–11055.

(5) Kidd, D.; Liu, Y.; Cravatt, B. F. *Biochemistry* **2001**, *40* (13), 4005–4015.

(6) Powers, J. C.; Asgian, J. L.; Ekici, O. D.; James, K. E. *Chem. Rev.* **2002**, *102* (12), 4639–4750.

(7) In *Bioconjugate Techniques (Third edition)*; Academic Press: Boston, 2013; pp i–iii.

(8) Lewis, J. C. *ACS Catal.* **2013**, *3* (12), 2954–2975.

(9) Harwood, V. J.; Denson, J. D.; Robinson-Bidle, K. A.; Schreier, H. J.

(10) Doble, M. V.; Ward, A. C. C.; Deuss, P. J.; Jarvis, A. G.; Kamer, P. C. J. *Bioorganic & Medicinal Chemistry* **2014**, *22* (20), 5657–5677.

(11) Lewis, J. C.; Coelho, P. S.; Arnold, F. H. *Chemical Society Reviews* **2011**, *40* (4), 2003–2021.

(12) Hershberg, R.; Petrov, D. A. *PLoS Genet.* **2009**, *5* (7), e1000556.

(13) Marchi-Delapierre, C.; Rondot, L.; Cavazza, C.; Ménage, S. *Israel Journal of Chemistry* **2015**, *55* (1), 61–75.

(14) Pordea, A.; Creus, M.; Panek, J.; Duboc, C.; Mathis, D.; Novic, M.; Ward, T. R. *Journal of the …* **2008**, *130* (25), 8085–8088.

(15) Gao, Y.; Crabtree, R. H.; Brudvig, G. W. *Inorg Chem* **2012**, *51* (7), 4043–4050.

(16) Liu, P.; Wong, E. L.-M.; Yuen, A. W.-H.; Che, C.-M. *Org. Lett.* **2008**, *10* (15), 3275–3278.

(17) Harris, M. N. *Journal of Biological Chemistry* **2001**, *276* (22), 19310–19317.

(18) Kamijo, S.; Amaoka, Y.; Inoue, M. *Synthesis* **2010**, *2010* (14), 2475–2489.

(19) Szeltner, Z.; Renner, V.; Polgar, L. *Journal of Biological Chemistry* **2000**.

(20) Das, S.; Incarvito, C. D.; Crabtree, R. H.; Brudvig, G. W. *Science* **2006**, *312* (5782), 1941–1943.

(21) Hull, J. F.; Sauer, E. L. O.; Incarvito, C. D.; Faller, J. W.; Brudvig, G. W.; Crabtree, R. H. *Inorg Chem* **2009**, *48* (2), 488–495.

(22) Himiyama, T.; Sauer, D. F.; Onoda, A.; Spaniol, T. P.; Okuda, J.; Hayashi, T. *Journal of Inorganic Biochemistry* **2016**, *158*, 55–61.

(23) Aebersold, R.; Mann, M. *Nature* **2003**, *422* (6928), 198–207.

(24) Vandermarliere, E.; Mueller, M.; Martens, L. *Mass Spectrom Rev* **2013**, *32* (6), 453–465.

(25) Cui, W.; Rohrs, H. W.; Gross, M. L. *Analyst* **2011**, *136* (19), 3854–3864.

(26) Terwilliger, T. C.; Grosse-Kunstleve, R. W.; Afonine, P. V.; Moriarty, N. W.; Adams, P. D.; Read, R. J.; Zwart, P. H.; Hung, L. W.; IUCr. *Acta Crystallogr. D Biol. Crystallogr.* **2008**, *64* (5), 515–524.

(27) Vieille, C.; Zeikus, G. J. *Microbiol. Mol. Biol. Rev.* **2001**, *65* (1), 1–43.

(28) Fülöp, V.; Böcskei, Z.; Polgár, L. *Cell* **1998**, *94* (2), 161–170.

(29) Haffner, C. D.; Diaz, C. J.; Miller, A. B.; Reid, R. A.; Madauss, K. P.; Hassell, A.; Hanlon, M. H.; Porter, D. J. T.; Becherer, J. D.; Carter, L. H. *Bioorg. Med. Chem. Lett.* **2008**, *18* (15), 4360–4363.

(30) Shan, L. *Proceedings of the National Academy of Sciences* **2005**, *102* (10), 3599–3604.

(31) Li, M.; Chen, C.; Davies, D. R.; Chiu, T. K. *J. Biol. Chem.* **2010**, *285* (28), 21487–21495.

(32) Canning, P.; Rea, D.; Morty, R. E.; Fülöp, V. *PLoS ONE* **2013**, *8* (11), e79349.

(33) Menyhárd, D. K.; Orgován, Z.; Szeltner, Z.; Szamosi, I.; Harmat, V. *Acta Crystallogr. D Biol. Crystallogr.* **2015**, *71* (Pt 3), 461–472.

(34) Szeltner, Z.; Juhász, T.; Szamosi, I.; Rea, D.; Fülöp, V.; Módos, K.; Juliano, L.; Polgár, L. *Biochim. Biophys. Acta* **2013**, *1834* (1), 98–111.

(35) Polgár, L.; Patthy, A. *Biochemistry* **1992**, *31* (44), 10769–10773.

(36) Akyuz, N.; Georgieva, E. R.; Zhou, Z.; Stolzenberg, S.; Cuendet, M. A.; Khelashvili, G.; Altman, R. B.; Terry, D. S.; Freed, J. H.; Weinstein, H.; Boudker, O.; Blanchard, S. C. *Nature* **2015**, *518* (7537), 68–73.

(37) Polgar, L. *Eur. J. Biochem.* **1991**, *197* (2), 441–447.

(38) Harwood, V. J.; Denson, J. D.; Robinson-Bidle, K. A.; Schreier, H. J. *J. Bacteriol.* **1997**, *179* (11), 3613–3618.

(39) Vieille, C.; Zeikus, G. J. *Microbiol. Mol. Biol. Rev.* **2001**, *65* (1), 1–43.

(40) Walden, H.; Taylor, G. L.; Lorentzen, E.; Pohl, E.; Lilie, H.; Schramm, A.; Knura, T.; Stubbe, K.; Tjaden, B.; Hensel, R. *J. Mol. Biol.* **2004**, *342* (3), 861–875.

(41) Yip, K. S.; Stillman, T. J.; Britton, K. L.; Artymiuk, P. J.; Baker, P. J.; Sedelnikova, S. E.; Engel, P. C.; Pasquo, A.; Chiaraluce, R.; Consalvi, V. *Structure* **1995**, *3* (11), 1147–1158.

(42) Sterpone, F.; Melchionna, S. *Chemical Society Reviews* **2012**, *41* (5), 1665–1676.

(43) Jónsdóttir, L. B.; Ellertsson, B. Ö.; Invernizzi, G.; Magnúsdóttir, M.; Thorbjarnardóttir, S. H.; Papaleo, E.; Kristjánsson, M. M. *Biochim. Biophys. Acta* **2014**, *1844* (12), 2174–2181.

(44) Papaleo, E.; Renzetti, G.; Tiberti, M. *PLoS ONE* **2012**, *7* (4), e35686.

(45) Tiberti, M.; Invernizzi, G.; Lambrughi, M.; Inbar, Y.; Schreiber, G.; Papaleo, E. *Journal of chemical …* **2014**, *54* (5), 1537–1551.

(46) Polgár, L. *Biochemistry* **1992**, *31* (33), 7729–7735.

(47) Andreev, Y. A.; Kozlov, S. A.; Vassilevski, A. A.; Grishin, E. V. *Anal. Biochem.* **2010**, *407* (1), 144–146.

(48) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. *Journal of Applied Crystallography* **2007**, *40* (Pt 4), 658–674.

(49) Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D. *Acta Crystallogr. D Biol. Crystallogr.* **2012**, *68* (Pt 4), 352–367.

(50) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (Pt 4), 486–501.

(51) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. *Journal of computational chemistry*. Wiley Subscription Services, Inc., A Wiley Company August 2008, pp 1859–1865.

CHAPTER FOUR

DIRHODIUM PYROCOCCUS FURIOSUS PROLYL OLIGOPEPTIDASE ARTIFICIAL METALLOENZYMES

4.1 INTRODUCTION

Members of the Lewis group recently developed a general method for the creation of artificial metalloenzymes using strain-promoted azide–alkyne cycloaddition (SPAAC) of bicyclo[6.1.0]nonyne (BCN)-substituted cofactors.[1] Through the genetic introduction of a *p*-azidophenylalanine (Z) using amber-codon suppression,[2] they were able site-specifically introduce this bioorthogonal handle. Owing to the bioorthogonality of this approach, biococonjugation could be achieved site-specifically without the need to remove competing functional groups from the protein scaffold, a key deficiency with approaches employing cysteines as nucleophilic reactive handles.[3]

Using this approach, they introduced the BCN-substituted esp-dirhodium catalyst RhBCN (Figure 4.1) into various Z mutants of tHisF. Dirhodium tertracarboxylate catalysts like RhBCN are attractive targets for enhancing selectivity, as they can catalyze alkene cyclopropanation and carbene insertion into C-H, N-H, O-H, Si-H, and S-H bonds. The use of tHisF as an ArM scaffold for this cofactor proved useful for the development of catalytic methodology for dirhodium ArM catalysis. No selectivity was observed with hybrids based on this tHisF scaffold, however, indicating a failure of the scaffold to induce selectivity in the catalyzed reactions (cyclopropanation and Si-H insertion). Poor encapsulation of the metal in the tHisF scaffold was hypothesized to be culprit for lack of selectivity.[1] *Pyrococcus furiosus* prolyl oligopeptidase posed a

promising alternative scaffold for the RhBCN cofactor, with the appropriate size and shape to potentially result in metal enclosure (chapter three). As such, we initiated an engineering effort to incorporate RhBCN into the *Pfu* POP scaffold. This yielded a number of interesting ArMs for asymmetric cyclopropanation.[4]

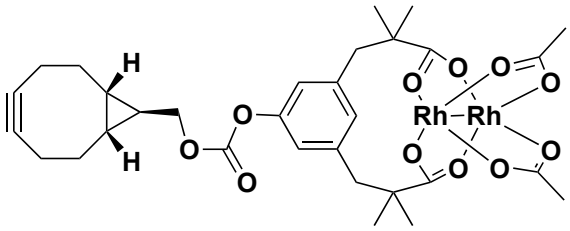


Figure 4.1 RhBCN cofactor used in ref. 1 and in the investigations described.

4.2 RESULTS AND DISCUSSION

4.2.1 ENGINEERING PFU POP Z FOR BIOCONJUGATION

The utility of *Pfu* POP as a potential scaffold to encapsulate the RhBCN cofactor was recognized soon after the first expression tests of this enzyme were conducted in the laboratory (chapter three). A model of this cofactor inside of the *Pfu* POP homology model suggested ideal placement of the dirhodium fragment in the central cavity provided by the β-domain (Figure 4.2).[5] The anchoring position of the cofactor was selected at the active site serine (S477) of *Pfu* POP. According to the model, this serine occupied a central position in the active site of the enzyme. Furthermore, S477 forms part of a conserved consensus motif in all POPs: GXSXG. [6] In the absence of experimental structural data, this motif allowed us to position the anchoring *p*-azidophenylalanine (Z) with confidence that it would lie at the bottom of the central active site cavity of the scaffold. (Figure 4.2)

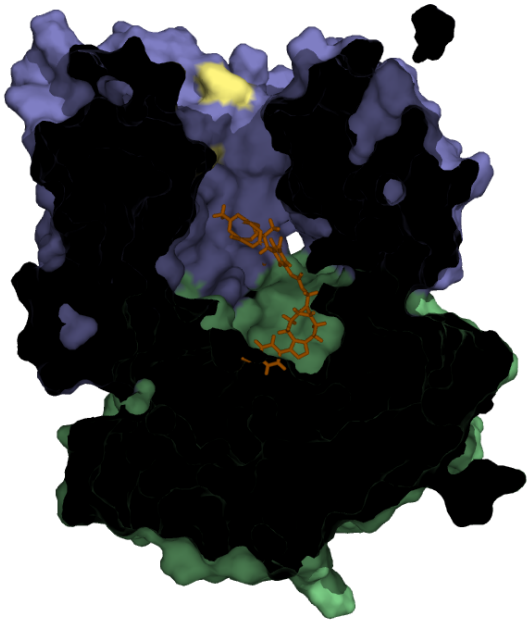


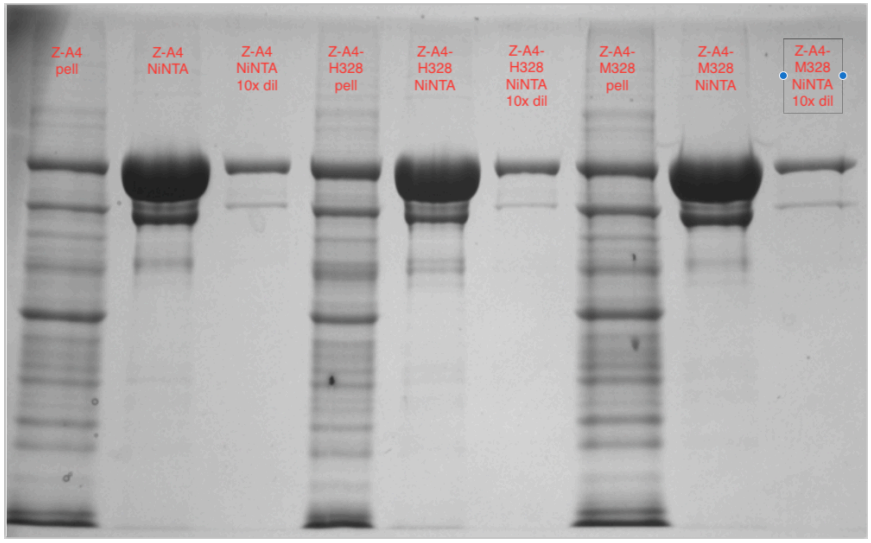
Figure 4.2 Cross-section of the *Pfu* POP homology model with RhBCN (orange) anchored at Z477 (hydrolase domain, green) and contained within the central cavity defined by the β-domain (purple)

With this in mind, an amber stop codon (UAG) was introduced at position 477 of the *Pfu* POP primary sequence. The protein was properly expressed with *p*-azidophenylalanine when co-expressed with pEVOL-pAzF (encoding the tRNA/tRNA synthetase pair for *p*-azidophenylalanine incorporation) (Figure 4.3 A).[7] This species is denoted *Pfu* POP-Z. In a curious side note, when expression of *Pfu* POP Z variants areinduced in the absence of pEVOL induction, truncated *Pfu* POP Δ(477-624) is produced in soluble form (Figure 4.3 B). Indeed, this truncation product is consistently observed with expression of Z mutants of *Pfu* POP, as the pEVOL-pAzF system is not perfectly efficient in amber codon suppression. In addition, the truncation product is

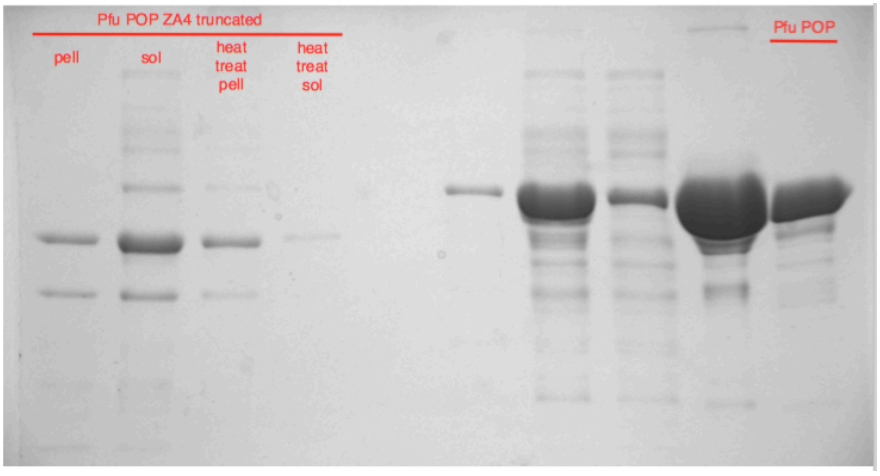
persistent despite heat treatment and NiNTA affinity chromatography, a remarkable fact given the lack of a C-terminal hexahistidine tag.

Figure 4.3 A) *Pfu* POP-Z mutant expression of different variants. A contaminant truncated fragment is persistent after NiNTA chromatography, however. **B)** *Pfu* POP Δ(477-624) expression. The truncated species is tolerant of heat treatment, particularly in the presence of full-length *Pfu* POP species, making it difficult to remove from samples of *Pfu* POP Z mutants.

A)



B)



Initial attempts at bioconjugation of *Pfu* POP-Z with RhBCN were found to be unsuccessful. It should be noted that at the time of these experiments, reliable ESI-MS characterization had not yet been developed for characterization of the product hybrid. In light of these bioconjugation issues, we inspected the homology model of *Pfu* POP and hypothesized that restriction of the central pore of the enzyme was preventing introduction of the RhBCN cofactor. To solve this, four alanine mutations (A_4: E104, F146, K199 and D202) were introduced, removing the bulky residues that were deemed to be blocking entry of the cofactor. Ready bioconjugation to the *Pfu* POP-ZA_4 scaffold was observed by MALDI, and, after more reliable methodology had been developed, ESI-MS. In later bioconjugation trials using ESI-MS for validation, it was found that *Pfu* POP-Z was readily bioconjugated with RhBCN, indicating that these alanine mutations were not necessary for bioconjugation. At the time, however, *Pfu* POP-ZA_4-RhBCN was carried on as the platform for investigations into its reactivity and selectivity.

As discussed in Chapter Two, the removal of excess cofactor from the bioconjugation mixtures was found to be operationally non-trivial. In order to effectively separate the hybrid from the cofactor, coworkers employed an azide-functionalized sepharose resin to scavenge excess cofactor after bioconjugation. In this way, a full ArM formation and purification strategy was developed for *Pfu* POP Z mutants.[8]

Several important details about SPAAC bioconjugation to *Pfu* POP Z mutants were discovered owing to the excellent reliability of low-resolution and high-resolution LC/ESI-MS characterization of these systems. Incomplete bioconjugation was consistently observed regardless of the identity of the Z scaffold or BCN cofactor. In addition, increasing cofactor stoichiometry had no effect in improving bioconjugation

efficiency. Through the systematic mass spectrometry effort described, it was found that reduction of the *p*-azidophenylalanine residue to *p*-aminophenylalanine (AmF) was the culprit in limiting bioconjugation efficiency. This was corroborated by high-resolution ESI-MS, which could resolve the *Pfu* POP-Z-A_4 scaffold mass from that of *Pfu* POP-AmF-A_4. CNBr digestion followed by LC/Q-TOF-MS of *Pfu* POP-Z-A_4 scaffold further corroborated this result, as the product AmF477-containing peptide could be chromatographically separated from the Z477-containing product peptide (Figure 4.4).

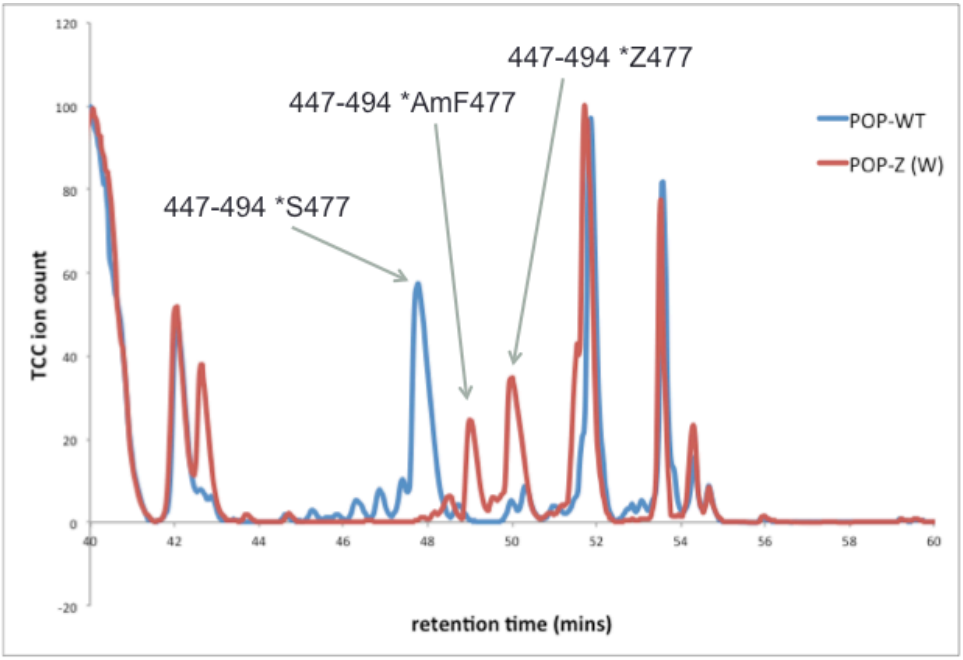
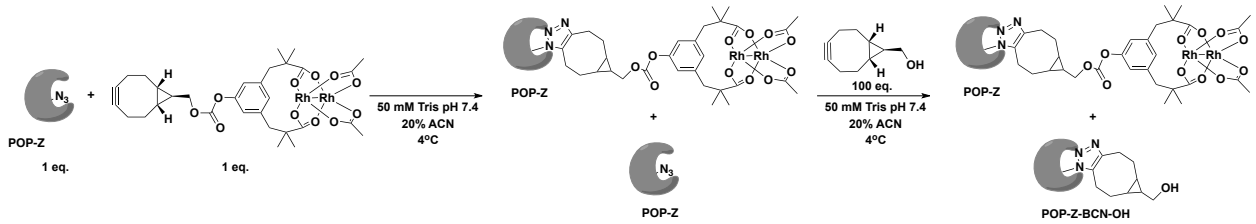


Figure 4.4 LC/Q-TOF-MS chromatogram indicating resolution of CNBr digest fragment 447-494 containing S477 (wt*Pfu* POP, blue), AmF477 (*Pfu* POP-Z, red) and Z477 (*Pfu* POP-Z, red)

The cause of Z reduction is still unknown, though photoreduction[9] and chemical reduction (primarily by sulfhydryl compounds)[10] have been reported to be quite facile for such aryl azides. To address this issue, coworkers in the laboratory are exploring *p*-

methylazidophenylalanine incorporation into *Pfu* POP, as tRNA/tRNA synthetase pairs have already been developed for this azide-containing residue, which promises to be more stable toward reduction.[11]

Bioconjugation reactions were also analyzed at different time points to shed light on the aforementioned bioconjugation problems and the role of the A_4 aperture mutations on the rate of the SPAAC reaction. A significant difference in the bioconjugation rates of *Pfu* POP-Z and *Pfu* POP-Z-A_4 was expected. Bioconjugation reactions were conducted using equimolar *Pfu* POP-Z-(A_4) and RhBCN and quenched with a large excess (100 eq.) of the truncated strained alkyne BCN-OH at designated time points (Table 4.1). The mixtures were analyzed by ESI-MS, and the ratio of RhBCN- and BCN-OH-conjugated POP was evaluated as a proxy for extent of bioconjugation. This experiment consistently showed that the SPAAC bioconjugation reaction was extremely rapid, regardless of the presence or absence of aperture mutations. After only 5 seconds, complete bioconjugation of azide-containing scaffold could be observed for both *Pfu* POP-Z and *Pfu* POP-Z-A_4. (Table 4.1)



time	*Pfu* POP-Z-RhBCN %bioconj	*Pfu* POP-Z-BCNOH %bioconj	*Pfu* POP-Z-A_4 – RhBCN %bioconj	*Pfu* POP-Z-A_4-BCNOH %bioconj
5 s	56%	1%	45%	0%
30 s	54%	0%	47%	0%
60 s	54%	0%	50%	0%
120 s	58%	0%	51%	0%

Table 4.1 LC/ESI-MS bioconjugation values obtained at different timepoints for the bioconjugation/quenching experiment shown. Bioconjugation extent was measured as a fraction of total protein (including AmF477 incompetent scaffold)

4.2.2 SELECTIVE CATALYSIS WITH PFU POP DIRHODIUM SPAAC ARMS

With the *Pfu* POP-Z-A_4-RhBCN hybrid successfully synthesized, we interrogated this system for catalytic activity and selectivity in the insertion of carbenes into alkenes (cyclopropanation). Using a naïve set of standard conditions, this system was found to give 11% e.e. and 19% yield in the cyclopropanation of styrene with methyl 2-diazo-2-(4-methoxyphenyl)acetate (Table 4.2). Optimizing reaction conditions, in part based on the known halide-binding effects on *Pfu* POP (chapter three), led to improved enantioselectivity and yield for this reaction (38% e.e. and 25% yield, respectively). This served as the starting point for two branches of optimization through mutagenesis:

rational design and directed evolution. These will be discussed in detail along with mechanistic studies that arose from interesting effects in the mutants obtained.

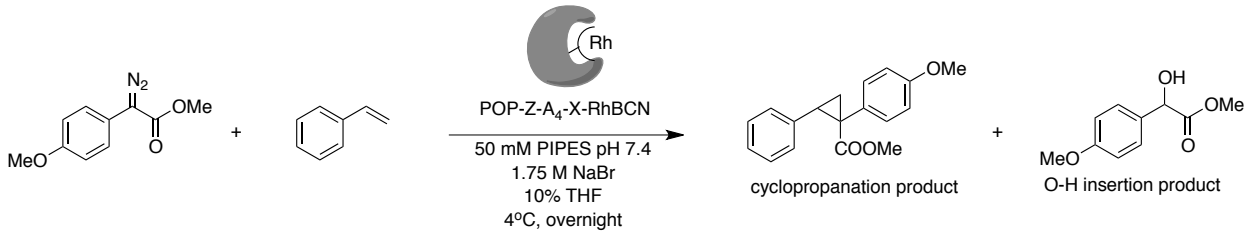
4.2.3 ARM OPTIMIZATION THROUGH RATIONAL DESIGN

As discussed in chapter one, an important feature of ArM catalysis is the capacity for catalyst design and optimization through mutagenesis. By applying a number of rational semi-combinatorial mutagenesis strategies to *Pfu* POP-Z-A_4-RhBCN, we were able to significantly enhance the selectivity of the catalyzed cyclopropanation reaction, a vindication for the *Pfu* POP scaffold and its application in ArM catalysis.[4]

Starting with the optimized conditions for *Pfu* POP-Z-A_4-RhBCN-catalyzed cyclopropanation, members of the Lewis group sought to strategically introduce mutations into the scaffold to enhance the observed selectivity. As has been mentioned multiple times in this dissertation and in ArM literature, restricting cofactor movement within protein scaffolds has been shown to afford enhanced selectivity in ArM-catalyzed reactions.[3,12] Cofactor flexibility was a reasonable pathology to address, as the RhBCN cofactor contains a relatively long linker with multiple rotatable bonds. To effect cofactor restraint, two strategies were carried out. The first relied on the known propensity of dirhodium catalysts to be axially coordinated by dative donors.[13] It was hypothesized that axial coordination by a histidine imidazole sidechain (a known ligand for dirhodium catalysts)[14] could restrict the movement of the dirhodium cofactor within the *Pfu* POP scaffold. Individual histidine mutations were systematically introduced in the vicinity of the putative position of the dirhodium center within the protein scaffold (as determined by inspection of the homology model). The resulting set of variants-*Pfu* POP-ZA_4-(G99H, P139H, I141H, I197H, T209H, E218H, V219H, Y251H, E283H and L328H) was

assessed for its ability to yield mutants with altered and enhanced enantioselectivity in cyclopropanation. Mutant *Pfu* POP-ZA$_4$-L328H yielded a substantial improvement in enantioselectivity to 85% e.e., with a concurrent increase in yield to 61%.

In order to qualitatively assess whether this improvement was due to axial coordination of the histidine sidechain, position 328 was mutated to cysteine and methionine, the sidechains of which are also metal coordinating ligands. These exhibited slightly lower yields and enantioselectivities relative to L328H, but they were both improved when compared to the parent *Pfu* POP-Z-A$_4$. This provided some evidence that axial coordination could play a role in ArM enantioselectivity. In addition, the corresponding phenylalanine mutant hybrid *Pfu* POP-Z-A4-L328F-RhBCN showed decreased selectivity and yield, indicating that coordinating residues are preferred at this position. Further investigations to physically characterize the axial coordination of histidine, methionine, and cysteine at position 328 in these hybrids are discussed below.



Entry	*Pfu* POP-Z-A_4 Mutant	Conditions	Conv. (%)	e.e. (%)	cyclopropanation/O-H insertion
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, RhEsp-OAc	PIPES, 1.75 M NaBr	12	0	0.4

Table 4.2 Results of rational design effort. All reactions conducted using 4 mM styrene and 20 mM diazo. Conversion and enantioselectivity determined by HPLC relative to internal standard.

With *Pfu* POP-ZA_4-L328H in hand, members of the Lewis group attempted to further restrain cofactor movement in the cavity of the scaffold by introducing a variety of phenylalanine mutations projecting towed the putative position of the cofactor (as predicted by the homology model of *Pfu* POP.) With this in mind, mutants *Pfu* POP-ZA_4-L328H-(S64F, L97F, G99F and G594F) were systematically assessed for enhancement of enantioselectivity and yield. The double phenylalanine mutant scaffold *Pfu* POP-ZA_4-L328H- G99F-G594F (HFF) was found to exhibit the strongest improvement in selectivity and yield, with 92% e.e. and 74% yield. Whether the enhanced selectivity arises from the hypothesized steric crowding of the cofactor within the *Pfu* POP scaffold cavity remains an open question. The selective ArM catalyst *Pfu* POP-ZA_4-HFF-RhBCN was shown to accept a variety of styrene and diazo substrates for cyclopropanation, all with moderate to good enantioselectivity. These substrates contain both electron donating and withdrawing groups, speaking to the generality of the catalyst reactivity.

Arguably the most compelling aspect of this rational ArM mutagenesis study arose from the fact that, during the lineage of mutagenesis, a trend of increasing cyclopropanation product yield was observed concurrently with improvements in enantioselectivity. This fact was met simultaneously with the observation that the efficiency of cyclopropanation relative to the competing carbene insertion reaction into the O-H bond of water also increased (Table 4.2) . These two trends suggest that over the course of the rational mutagenesis lineage, the complementarity between the substrates and the *Pfu* POP ArMs was enhanced. This was corroborated by an observed rate enhancement in cyclopropanation. The significance of the observed increase in substrate/catalyst complementarity cannot be understated: such an effect suggests that

reaction specificity of the formed dirhodium-carbene intermediate can be modulated to preferentially select styrene substrates over the plentiful water in the medium.

4.2.3.1 BIOPHYSICAL VALIDATION OF ARM STABILITY

As discussed in chapter three, one of the main advantages of the use of *Pfu* POP as a scaffold arises from the exceptional mutational tolerance that such a stable scaffold confers. The effect of the dirhodium cofactor RhBCN on the stability of the scaffold was also investigated. The thermal stability of hybrid *Pfu* POP-ZA_4-HFF-RhBCN was assessed by observation of the circular dichroism signal over increasing temperatures in a manner similar to that reported in section 3. (Figure 4.5) With this, it was evident that the presence of the covalently-appended cofactor had little effect on the scaffold thermostability, with no significant change in the temperature-dependent CD curve. This stands as further testament to the remarkable robustness of *Pfu* POP as an ArM scaffold.

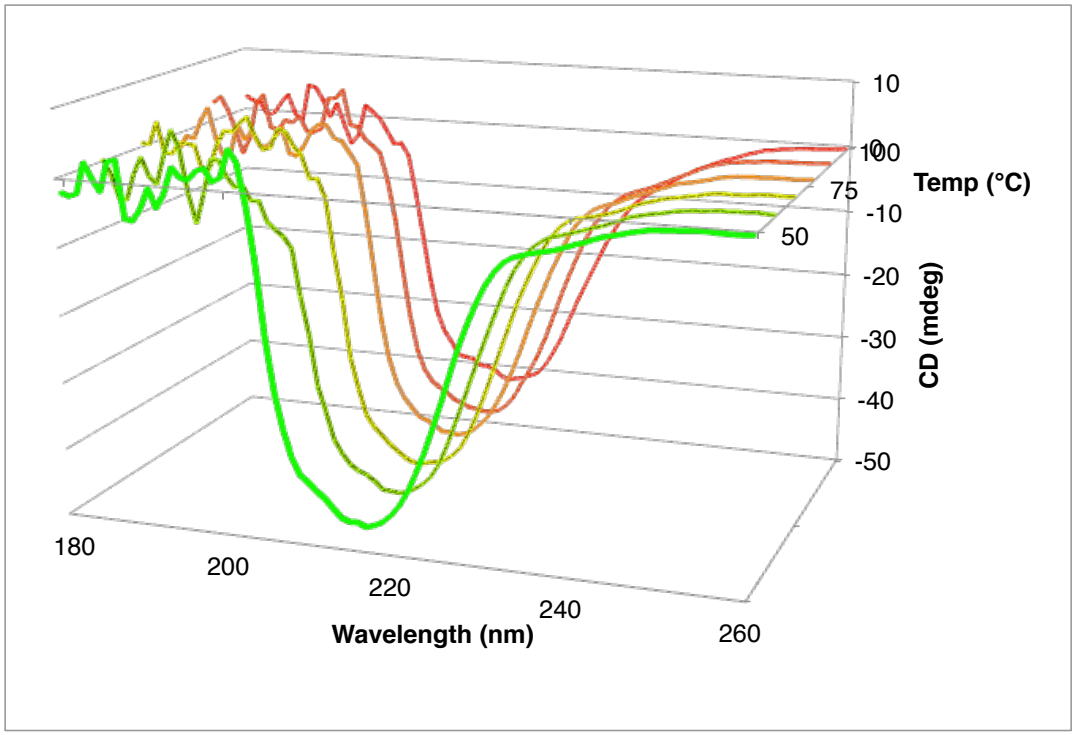


Figure 4.5 Temperature-dependent CD curves of *Pfu* POP-ZA_4-HFF-RhBCN indicating that scaffold stability is not impacted by these mutations nor the presence of bound cofactor

4.2.3.2 VALIDATION OF AXIAL LIGATION

After the identification of the beneficial mutation L328H, we sought to confirm the presence of axial coordination (Figure 4.6) by three methods of physical characterization: UV/Visible spectroscopy, resonance Raman spectroscopy, and protein NMR spectroscopy. These methods will be discussed below:

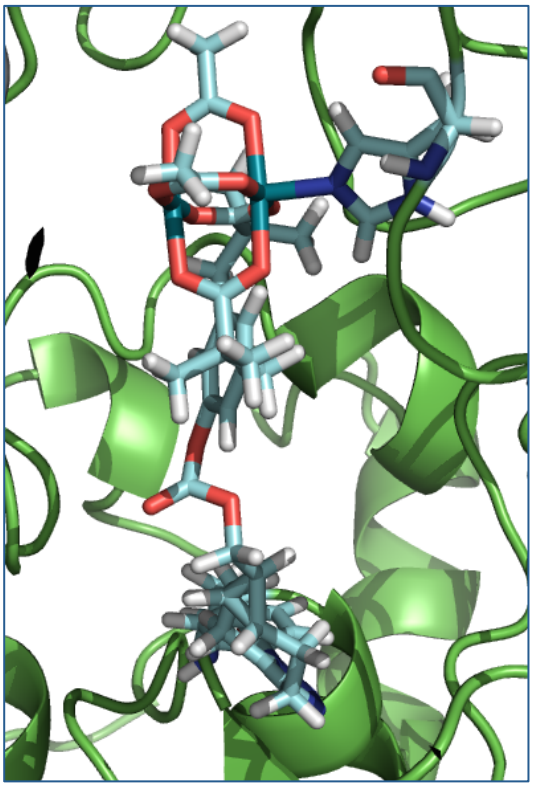


Figure 4.6 Putative axial coordination of histidine to bound RhBCN in *Pfu* POP-ZA_4-HFF-RhBCN built in Pymol.

4.2.3.2.1 UV/VISIBLE SPECTROSCOPY

UV/Visible spectroscopy has been extensively employed for the identification of axial ligation of dirhodium complexes, as absorbance profile of these molecules is quite sensitive to the identity of its axial ligands.[15] Axial coordination of dative oxygen donors like tetrahydrofuran gives the complexes brilliant blue colors, while dative nitrogen donors like acetonitrile tend to yield purple complexes.[16] The observed coordination-dependent absorbance shifts arise from the diagnostic Rh–Rh π^*–σ^* transition of dirhodium tetracarboxylate complexes.[17]

In order to assess the effect of axial coordination by histidine, cysteine, or methionine at position 328, a differential absorbance UV/Vis strategy was taken. This is

due to the fact that, while sufficient to give color to concentrated ArM samples, the molar absorptivity associated with the Rh–Rh π^*–σ^* is quite weak, and therefore this suffers from competitive absorbance of scaffold. Therefore, scaffold absorbance must be simultaneously subtracted from the corresponding dirhodium hybrid, a method that has been applied for assessing the binding of dirhodium tetracarboxylate complexes to serum albumins.[18] Applying this method with samples of *Pfu* POP-ZA_4-RhBCN, *Pfu* POP-ZA_4-H328-RhBCN, *Pfu* POP-ZA_4-L328C-RhBCN, and *Pfu* POP-ZA_4-L328M-RhBCN, we were able to successfully observe the chromophore due to the bound dirhodium cofactor (Figure 4.7). No significant shift in the absorbance spectrum of these species was observed, however. Though disappointing, this result did not conclusively rule out the expected axial ligation, as the change in absorbance might be subtle. Furthermore, it does not preclude the possibility that axial coordination occurs during the course of catalyst turnover.

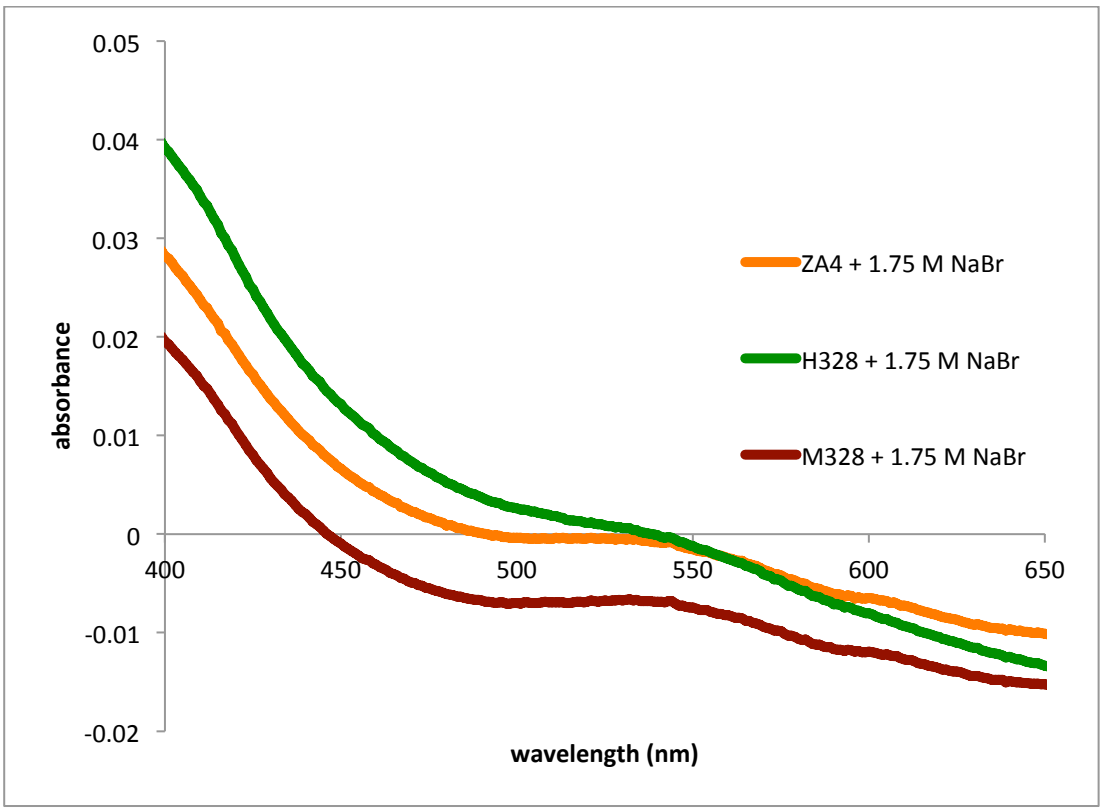


Figure 4.7 Difference UV/Vis spectra of 27 uM samples of hybrids *Pfu* POP-ZA_4-RhBCN, *Pfu* POP-ZA_4-H328-RhBCN, and *Pfu* POP-ZA_4-M328-RhBCN in 50 mM PIPES pH 7.4 + 1.75 M NaBr

4.2.3.2.2 RAMAN SPECTROSCOPY

Resonance Raman spectroscopy was also explored to probe dirhodium axial ligation in collaboration with the Berry group at UW-Madison. Resonance Raman has been successfully applied to probe the frequency of the Rh-Rh stretching mode in dirhodium tetracarboxylate complexes.[19] This vibrational mode is known to be sensitive to axial ligation, perhaps granting a diagnostic handle for characterization in the context of *Pfu* POP-RhBCN ArMs.[19] In addition, the resonance signal due to dirhodium would not suffer from the same scaffold background that plagued the UV/Vis measurements.

Initial attempts at the acquisition of adequate Raman spectra have been unsuccessful, potentially due to operational inexperience with aqueous protein samples.

4.2.3.2.3 PROTEIN NMR SPECTROSCOPY

Protein NMR was also attempted for the biophysical characterization of axial dirhodium histidine coordination in the *Pfu* POP-ZA_4-L328H-RhBCN hybrid. Protein NMR studies have been previously been applied to probe perturbations in the chemical environment of metal-binding histidine residues in natural metalloenzymes. Hartwig demonstrated that the replacement of the active site Zn(II) of human carbonic anhydrase II (hCAII) with Rh(I) led to a significant shift in the resonances associated with the aromatic histidine sidechains that form the His/His/His motif responsible for metal binding in hCAII[20], echoing a similar study observing the metalation of apo-hCAII with Zn(II).[21] In 2D-HSQC and 2D-HMBC experiments, histidine sidechains reside in the aromatic region separate from the peptide backbone resonances used for solving protein NMR structures. This greatly simplifies the spectrum and enables the identification of diagnostic shifts in histidine sidechain resonances.

The structure of the NMR experiment conducted on *Pfu* POP-ZA_4-L328H was planned as follows: double labeled ^{15}N-^{13}C-labelled *Pfu* POP-ZA_4 and its histidine mutant *Pfu* POP-ZA_4-L328H would first be analyzed by 2D-HSQC and 2D-HMQC in order to enable identification of the aromatic resonances arising from H328 by simple comparison of the two spectra. Once this residue was identified, the two scaffolds would be bioconjugated with RhBCN, and the resulting hybrids would be analyzed again by NMR. If axial ligation of the dirhodium complex were occurring, a shift in the resonance corresponding to H328 would be observed.

Optimization of expression conditions for double-labeled *Pfu* POP-ZA_4 and *Pfu* POP-ZA_4-H328 was conducted, as maximal yields were desired due to the high cost of ^{15}C-labeled glucose and ^{15}N-labeled ammonium chloride. (Figure 4.8) This proved to be fruitful, as relatively high yields of protein were observed despite the requisite co-induction of pEVOL, a competitor for the sparse nutrients in the minimal labeling media. Once adequate quantities of the two labeled variants were obtained, the protein was taken to the Center for Structural Biology NMR Facility at UIC with guidance from Dr. Ben Ramirez. Owing to the large size of *Pfu* POP, a 900 MHz NMR instrument was used to acquire 2D HSQC and 2D HMQC spectra. TROSY was also applied, which has precedent in NMR studies with *Homo sapiens* POP.[22] In a further step to sharpen the NMR spectra acquired with this method, we conducted experiments at 50°C to counteract slow tumbling rates with this system. Spectra acquired with these methods enabled the identification of some aromatic nitrogen resonances, but not all expected histidine resonances were observed. This posed a major problem, as it made it impossible to confidently assign H328 sidechain resonances, a crucial result for the success of this experiment.

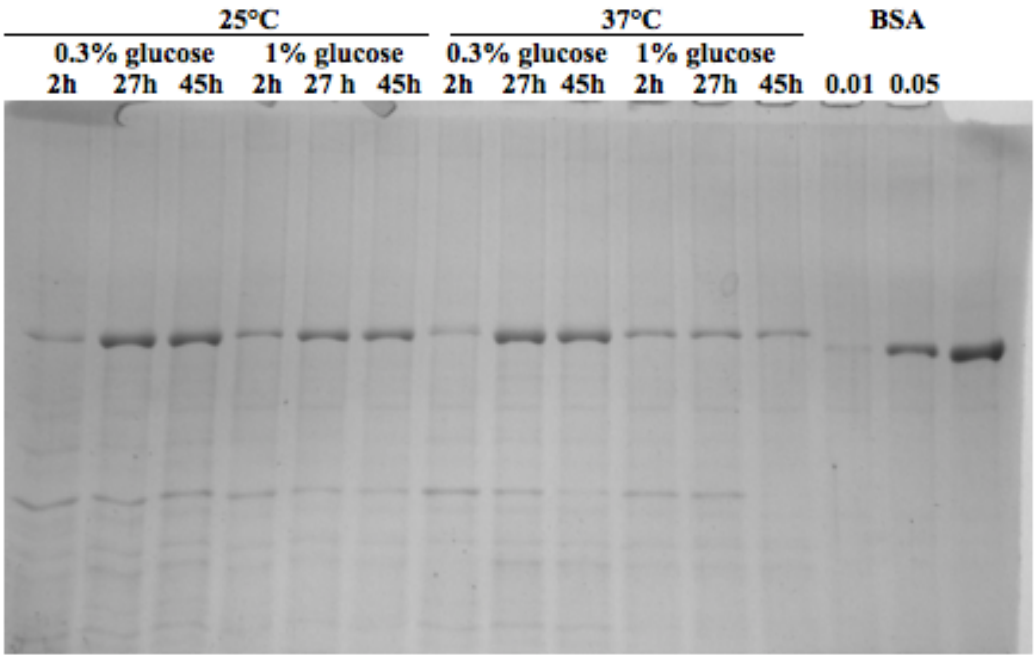


Figure 4.8 Optimization of minimal media expression conditions for double-labeling of *Pfu* POP Z variants for NMR

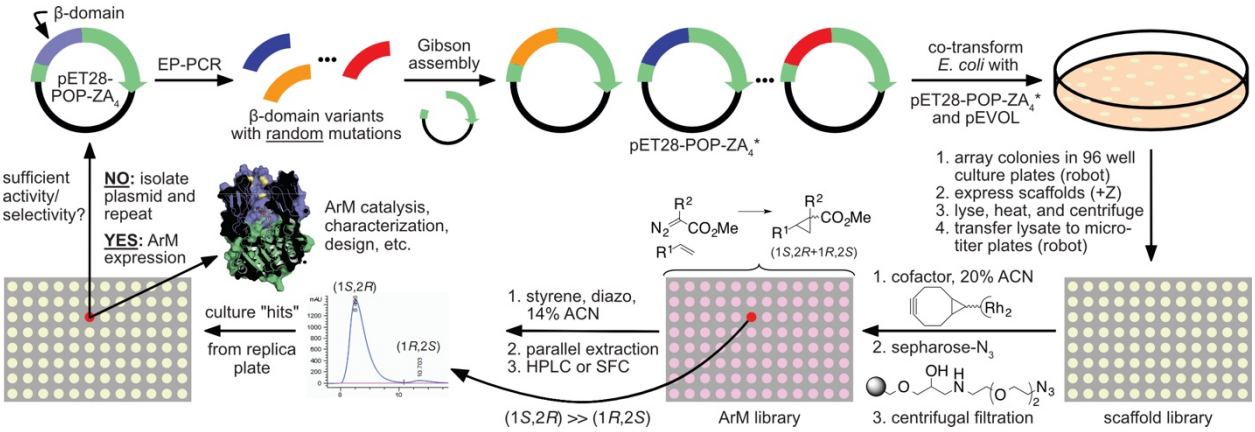
Further complicating the execution of the experiment was the noted inconsistency of the amide backbone region between *Pfu* POP-ZA_4 and *Pfu* POP-ZA_4-H328. While the cause of this inconsistency is unknown, there is reason to suspect interference from two contaminating species that may have been present in each of the samples to differing extent. The first of these is the truncated *Pfu* POP Δ(477-624), which, as noted, is a persistent contaminant in all Z mutants of *Pfu* POP. The second contaminant may arise from reduction of *p*-azidophenylalanine to *p*-aminophenylalanine. While scaffold reduction is not expected to yield greatly perturbed 2D NMR spectra, it does pose another complication for the second segment of this experiment. That is, bioconjugation of the double-labeled *Pfu* POP-ZA_4 and *Pfu* POP-ZA_4-H328 hybrids would inherently result a

heterogeneous mixture of hybrid and aniline-containing scaffold, thereby introducing uncertainty in the correct assignment of histidine resonances.

These issues have impeded progress toward the goal of this experiment. Currently, coworkers are generating a *Pfu* POP ΔHis scaffold to remove the signal interference due to histidines besides H328, the histidine of interest. In addition, as mentioned, we are developing standard methodology for the incorporation of *p*-methylazidophenylalanine, which poses an opportunity to eliminate complications arising from azide reduction. Purification methods to remove interference by truncated *Pfu* POP Δ(477-624) are also being pursued.

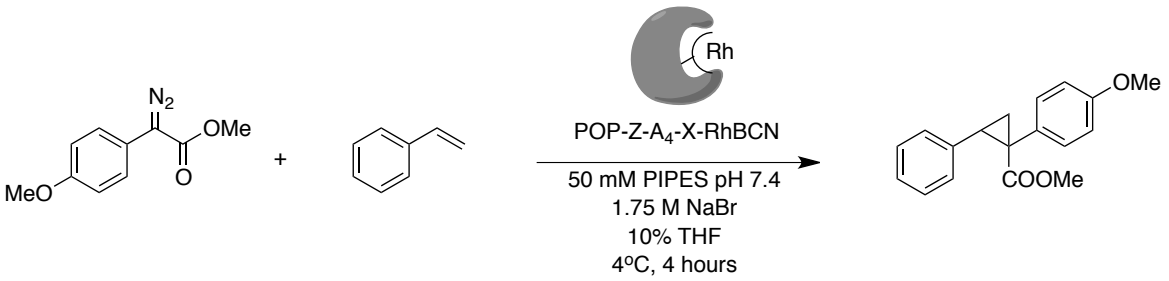
4.2.4 ARM OPTIMIZATION THROUGH DIRECTED EVOLUTION

In a separate effort from the rational design approach described, several members of the Lewis group recently demonstrated an ArM optimization strategy employing iterative rounds of random mutagenesis in the directed evolution of *Pfu* POP-Z-A_4-RhBCN. As discussed in chapter one, directed evolution presents another major advantage of ArM catalysts over traditional small molecule catalysts, as the selectivity of an ArM can be optimized iteratively and agnostically. That is, by applying semi-random or random mutagenesis to generate libraries of ArM scaffolds, one can obtain enhancements in desired properties without the prerequisite for structural or mechanistic knowledge. Directed evolution has seen more limited use in the context of ArM optimization than rational design approaches, but multiple notable examples exist.[23 24]



Scheme 4.1 The directed evolution strategy developed by coworkers for the optimization of *Pfu* POP RhBCN ArMs.

Members of the Lewis group developed an efficient protocol for the optimization of ArMs by directed evolution (Scheme 4.1). With this procedure, coworkers in the laboratory performed the iterative optimization of *Pfu* POP ArM cyclopropanases starting from the platform variant *Pfu* POP-Z-A_4-RhBCN. In just three rounds of random mutagenesis using 96-576 variants per library round, they identified a competent ArM cyclopropanase that exhibited 94% e.e. in cyclopropanation (Table 4.3).



Entry	Variant	Mutations[b]	ee **3** (%)[c]	Yield **3** (%)[c]
1	0-ZA_4	E104A/F146A/K199A/D202A	80:20	25
2	1-NAGS	K161N/V166A/S301G/T308S	84:16	33
3	2-NSIA	S84N/G99S/K330I/V335A	89:11	37
4	3-VRVH	I221V/Q228R/A265V/Y326H	97:3	47
5	4-GSH	S301G/G99S/Y326H	95:5	27

Table 4.3 Reactions conducted using 24 mM styrene, 4.8 mM diazo, and 1 mol% ArM [b] Mutations indicated in addition to those from previous generation. [c] ee and yield of cyclopropane determined by analysis of UPLC chromatograms for crude reaction mixtures relative to internal standards.

This impressive result arose from four mutations per round. The painstaking deconvolution of these mutations revealed a minimal set of beneficial mutations: S301G, G99S, and Y326H. Two of these mutations bear striking resemblance to the beneficial mutations obtained through rational design: G99F and L328H. Indeed, rationalization of these mutations seems feasible: position 99 seems to play an important role in directly imparting selectivity to the dirhodium catalyst. Y326H positions a histidine in a similar site to that identified by rational mutagenesis, H328, suggesting that a similar axial

coordination mechanism might be at play in imparting selectivity to the ArM. This newly identified histidine site is being investigated for its potential role as an axial ligand to the dirhodium catalyst, as was done for H328. In contrast to these mutations, the role of S301G is not quite as simple to rationalize, as this site is distal to the catalyst. As such, the identification of this mutation stands as testament to the power of random mutagenesis to uncover beneficial mutations in positions that would not be identified through rational design approaches.

4.2.4.1 DECREASING SELECTIVITY: PROBING SCAFFOLD MODIFICATION

In the rational design study discussed above, unintentional increases in yield, rate, and reaction specificity were uncovered along with the intended improvements in enantioselectivity. Similarly, along the lineage of directed evolution mutants, which were screened exclusively for increases in enantioselectivity, parallel improvements were uncovered. Significant deterioration of enantioselectivity was observed as the cyclopropanation reaction time was extended, particularly in the case of the parent variant *Pfu* POP-Z-A_4-RhBCN (termed 0-ZA_4 in the lineage nomenclature) (Figure 4.9).

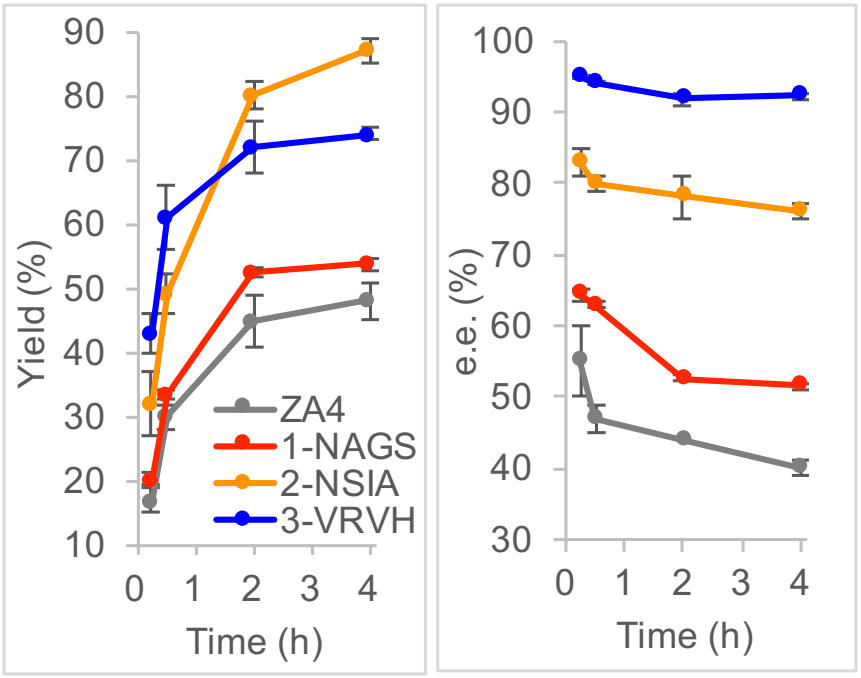


Figure 4.9 Yields of ArM-catalyzed cyclopropanation consistently increase during the course of the reaction, while enantioselectivity can be seen to deteriorate over time. The extent of loss of enantioselectivity is itself decreased along the evolved lineage.

These decreases in enantioselectivity themselves seemed to decrease along the evolved lineage. Given that the decreases in selectivity occur as a function of reaction time, the source of this effect was hypothesized to likely arise from changes to the catalyst as reaction time progresses. Scaffold modification by reactive carbene insertion was hypothesized as a likely cause of loss of catalyst selectivity. Preliminary mass spectrometry results to observe the ArM after catalysis revealed mass adducts corresponding to the predicted carbene insertion. A more systematic study was needed in order to confirm 1) the occurrence of scaffold modification, 2) the site(s) of scaffold modification, and 3) the effect that scaffold modification has on selectivity.

A thorough mass spectrometry effort was undertaken to confirm the occurrence of scaffold modification and to pinpoint its location on the *Pfu* POP scaffold. ArM-catalyzed cyclopropanations were prepared using three ArMs in the directed evolution lineage: 0-ZA_4 (parent), 1-NAGS, and 3-VRVH. In order to evaluate scaffold modification over time, high-resolution intact ESI-MS was employed. A method for the rapid "quenching" and purification of the ArM was needed to achieve accurate measurement of scaffold modification at specific timepoints. A rapid centrifugal gel-filtration method was developed toward this end. Intact HR-ESI-MS revealed a qualitatively similar rate and extent of scaffold carbene insertion for all ArM variants. Interestingly, controls containing the corresponding variant scaffolds (no dirhodium) showed some carbene modification as well, but to much lower extent.

The same timepoint purification protocol was employed for the evaluation of modified ArM samples by protein digestion. Purified timepoint samples of ArM were treated with CNBr (as described in chapter three) and the product peptide mixtures were analyzed by LC/Q-TOF-MS. We identified consistent time-dependent modification of the digest peptide corresponding to residues 174-183 of all three ArMs studied. Importantly, the rate and extent of modification appeared to be similar for all three variants. Again, minor modification could be observed when only the scaffolds themselves were employed. (Figure 4.10)

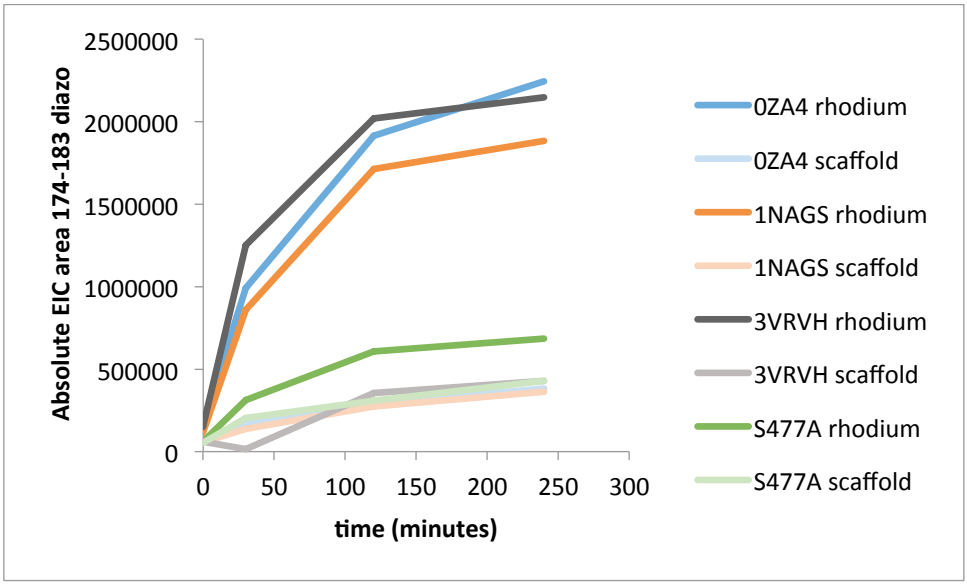


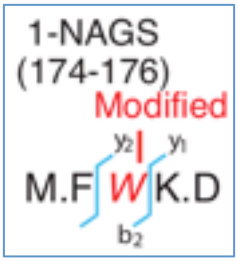
Figure 4.10 Increase in modified CNBr fragment 174-183 over the course of the cyclopropanation reaction. Scaffold-only controls exhibited much lower extents of modification than their respective ArMs. All variants exhibited similar extents and rates of modification. *Pfu* POP S477A was used in conjunction with free RhBCN to control for modification by free cofactor.

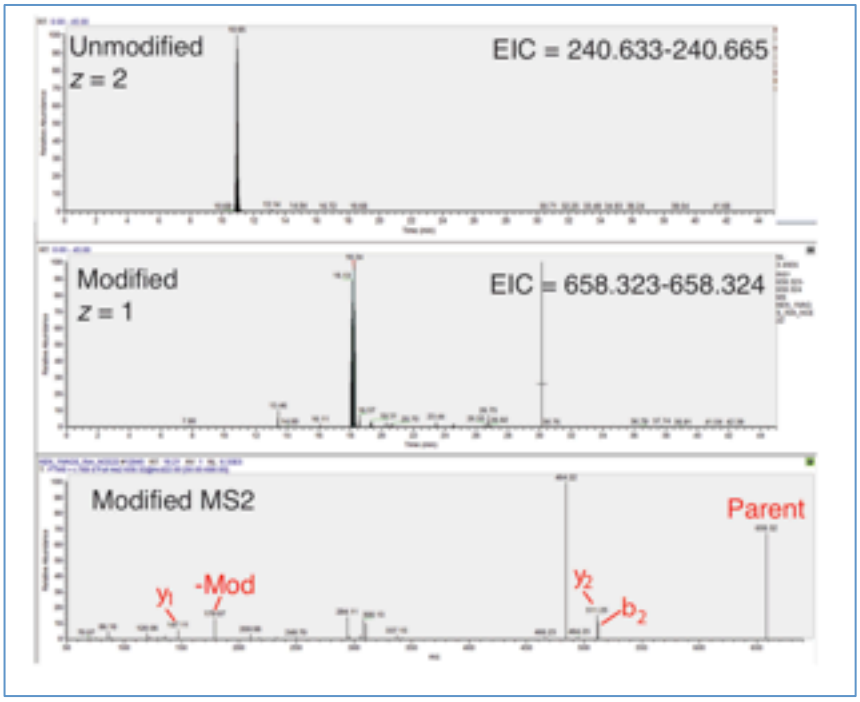
As discussed in chapter three, CNBr digestion of *Pfu* POP does not result in sufficiently small peptide fragments to enable confident coverage of the full protein sequence, nor does in enable MS/MS sequencing methodology to be applied for identifying the specific residues being modified during the course of cyclopropanation catalysis. For this reason, we applied a secondary tryptic digestion step to further cleave the CNBr peptide mixtures. With this approach, reproducible and complete digestion of *Pfu* POP ArM variants could be achieved. The product peptide mixtures obtained from two-step CNBr/trypsin digestion were analyzed by LC/MS/MS using an Q Exactive Orbitrap mass spectrometer courtesy of collaborators in the Moellering group at the

University of Chicago. This enabled sequencing of the 174-176 CNBr/tryptic fragment to confirm modification on W175 (Figure 4.11 A). In addition, we were able to identify a second major site of modification on W142, which had been missed in the analysis of CNBr digestion mixtures alone (Figure 4.11 B). The relative abundance of modified peptides was evaluated by comparison of the normalized abundance of parent unmodified peptides in cyclopropanation samples relative to their controls (Figure 4.11 C). This revealed significantly greater extent of modification of W142 versus W175. W142 sits in a position proximal to the putative location of the dirhodium cofactor, so modification at this site can be rationalized as an intramolecular effect driven by proximity to the catalyst. The rationale behind W175 modification is less clear. This site resides on the surface of *Pfu* POP, so its modification would seem to necessitate an intermolecular mechanism. This could explain the lower efficiency of modification at this site.

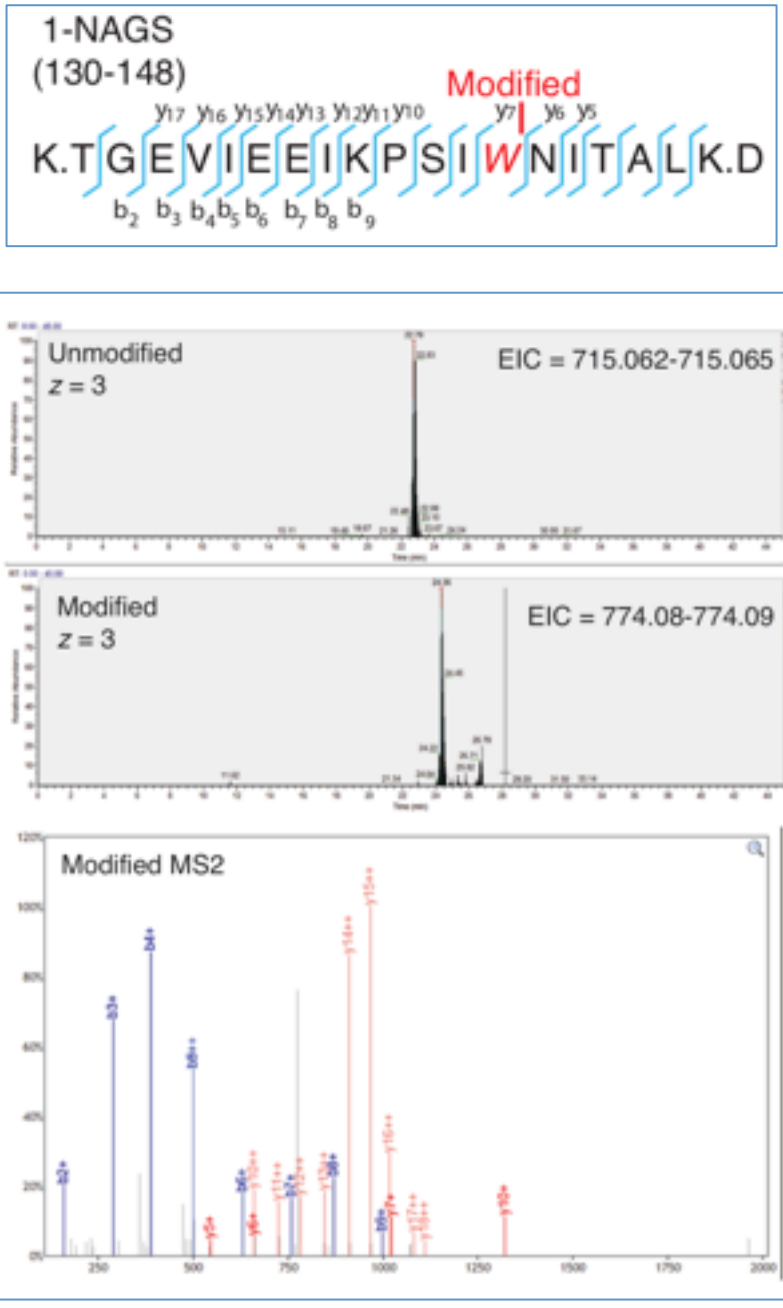
Figure 4.11 MS/MS results of CNBr/tryptic digestion of hybrid variants before (control) and after catalysis of cyclopropanation

A) LC/MS/MS chromatograms corresponding to the unmodified (top) and modified (middle) 174-176 (FWK) fragments of hybrid variant 1-NAGS. The MS2 spectrum (bottom) shows product –y and –b fragments confirming modification on W175

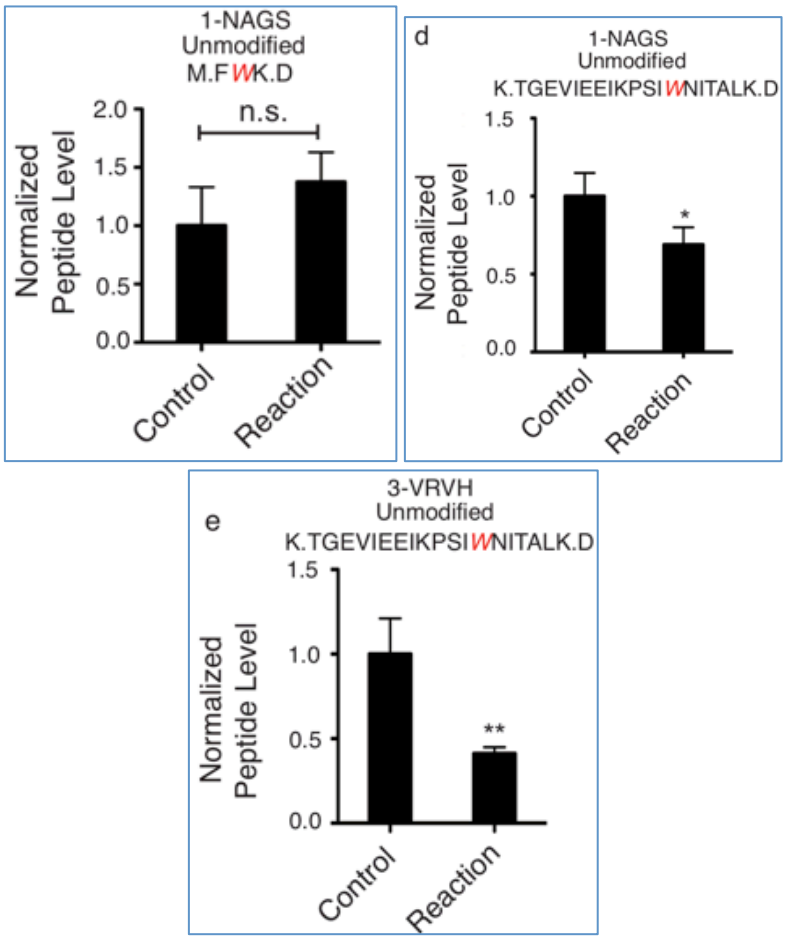




B) LC/MS/MS chromatograms corresponding to the unmodified (top) and modified (middle) 130-148 (KTGEVIEEIKPSIWNITALKD) fragments of hybrid variant 1-NAGS. The MS2 spectrum (bottom) shows product –y and –b fragments confirming modification on W142.



C) Relative abundances of unmodified fragments before (control) and after catalysis of cyclopropanation reaction (reaction).



The two identified sites of scaffold modification (W142 and W175) were mutated to alanine in the *Pfu* POP-Z-A_4 (0-ZA_4) scaffold to evaluate their effect on the drop in selectivity over time. Neither 0-ZA_4-W142A nor 0-ZA_4-W175A showed any change with

respect to the observed decrease in enantioselectivity. That is, modification at these sites did not seem to be the cause for the loss of catalyst selectivity.

In light of this result, we have not yet been able to identify the cause of drop in selectivity. One hypothesis implicates changes in the reaction medium as cyclopropanation proceeds over time. Heterogeneity of the reaction mixture increases as more of the cyclopropane product is generated, an unsurprising fact given the hydrophobicity of the product and the high ionic strength of the aqueous reaction medium. It has been noted that hydrophobic organic molecules can form colloidal aggregates under such conditions, which can lead to the partial or complete denaturation of proteins exposed to them.[25] Partial denaturation of the ArM scaffold could easily explain the observed drop in enantioselectivity, as the dirhodium catalyst can still react in a non-selective environment. Further investigations using non-denaturing detergents like Triton X-100 (a known solution to small-molecule aggregation) are being carried out to test this hypothesis.[26]

4.3 CONCLUSIONS

The work described in this chapter attests to the power of *Pfu* POP in providing a readily-optimized scaffold for selective ArM catalysis. While it is unclear whether *Pfu* POP encapsulates the dirhodium catalyst in the manner we have hypothesized, the scaffold has shown its utility in promoting selectivity on the otherwise symmetric dirhodium core of RhBCN. Furthermore, the selectivity of these ArMs can unequivocally be enhanced by mutagenesis. There is still much work to be done in order to fully uncover the mechanisms at play in granting selectivity to the ArMs optimized through

rational mutagenesis and directed evolution. The methods described herein serve as a basis to enable elucidation into the functioning of these highly complex catalysts.

4.4 EXPERIMENTAL

General Materials

Unless otherwise noted, all reagents were obtained from commercial suppliers and used without further purification. Deuterated compounds were obtained from Cambridge Isotope Labs. Labquake™ Tube Shaker/Rotators was purchased from Thermo Scientific (Catalog# 4002110Q). pEVOL-pAzF was a gift from Peter Schultz (Addgene plasmid # 31186) of the Scripps Research Institute, CAS1. E. coli DH5α and BL21 (DE3) cells were purchased from Invitrogen (Carlsbad, CA). Luria broth (LB; Cat# L24040), rich medium (2YT; Cat# X15600) and Agar (Cat# A20020) were purchased from Research Products International, Corp (Mt. Prospect, IL). Amicon® Ultra-15 Centrifugal Filter Units with Ultracel-30 membrane (Cat# UFC903024) were purchased from EMD Millipore (Billerica, MA) and used according to manufacturer's recommendations. 0.5 mL Zeba® centrifugal desalting columns were purchased from ThermoFisher Scientific. 5 mL HisTrap NiNTA columns were purchased from GE Healthcare Life Sciences (Pittsburgh, PA).

General Experimental

Unless otherwise noted, all reagents were obtained from commercial suppliers and used without further purification. Yields were determined by HPLC with 1,2,4-trimethoxybenzene or 1,3-benzodioxole as the internal standard and reported as the

average of three trials from the same batch of ArM set up in parallel. High resolution ESI mass spectra were obtained using an Agilent Technologies 6224 TOF LC/MS and an Agilent Technologies 6540 Q-TOF MS-MS. Low resolution ESI mass spectra were obtained using an Agilent Technologies 6130 LC-MS. LC-MS/MS experiments were performed with an Easy-nLC 1000 ultra-high pressure LC system (ThermoFisher) using a PepMap RSLC C18 column (column: 75 μm x 15 cm; 3 μm, 100 Å) coupled to a Q Exactive HF Orbitrap and Easy-Spray Nanosource (ThermoFisher). LC/MS experiments were performed using an Aeris 3.6 μm WIDEPORE C4 200 Å LC Column 150 x 4.6 mm. Amicon® Ultra-15 Centrifugal Filter Units with Ultracel-30 membrane (50 mL volume, 30 kDa cutoff) were used to concentrate or wash protein solutions. Protein concentrations were measured using the Pierce® BCA Protein Assay Kit and protein stocks were then stored at -80 °C until use. Circular dichroism (CD) spectra were obtained on a JASCO J-1500 CD Spectrometer. UV/Visible spectroscopy was conducted using a dual beam Cary 5000 UV/NIS/NIR Spectrophotometer.

Experimental Procedures

Standard *Pfu* POP Z-mutant expression protocol

The selected mutant was inoculated in 5 mL 2YT medium with 50 μg/mL kanamycin and 20 μg/mL chloramphenicol. The culture was incubated overnight at 37 ^{o}C with constant shaking at 250 rpm. On the following day, 5 mL of the overnight cultures was used to inoculate 1 L of fresh 2YT media 50 μg/mL kanamycin and 20 μg/mL chloramphenicol in a 2.8L Fernbach flask. The culture was incubated at 37 ^{o}C, 250 rpm, and protein expression was induced by adding 1mM IPTG, 2mM 4-Azido-l-phenylalanine and 1%

(w/v) l-arabinose when OD_{600} =1.0. The expression cultures were incubated at 37°C with 250 rpm shaking for an additional 18 hours. Cells were harvested by centrifugation at 3000 rpm for 10 minutes. They were resuspended in 100 mL 20 mM sodium phosphate buffer pH 7.4/50 mM imidazole/50 mM NaCl and lysed by sonication. The cell lysate was clarified at 15000 rpm for 1 hour. Clarified lysate was applied to a 3 x 5 mL HisTrap NiNTA columns (GE Healthcare Life Sciences) pre-equilibrated with 20 mM sodium phosphate buffer pH 6.5. The protein was purified using a gradient of 0-500 mM imidazole in 20 mM sodium phosphate buffer pH 6.5/50 mM NaCl. Fraction purity was confirmed by SDS-PAGE. Pure fractions were combined and concentrated using Amicon® 30 kD spin filters. Pure *Pfu* POP variants were buffer exchanged into H_2O, snap frozen with liquid N_2, and stored at -20°C.

Standard bioconjugation protocol

To set up bioconjugation, a solution of the POP mutant (480 μL, 75 μM in 50 mM Tris-Cl buffer, pH 7.4) and a solution of cofactor RhBCN (120 μL, 0.75 mM in ACN, 0.594 mg/mL) were added to a 1.5 mL microcentrifuge tube and shaken at 750 rpm at 4 °C overnight. The final concentrations were: 60 μM POP, 150 μM RhBCN, 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 hours to remove excess cofactor. The suspension was then centrifuged at 5000 rpm for 3 minutes and the supernatant was transferred to a new microcentrifuge tube. The resin was washed twice with 600 μL of 50 mM Tris-Cl 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 was 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-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.

Standard bioconversion protocol

To set up biocatalysis, solutions of aryldiazoacetate (25 μL, 96 mM, in THF), styrene (25 μL, 485 mM, in THF), and POP-ZA4-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 °C overnight (or for various time points, vide infra). The reaction was quenched by adding 20 μL 1,2,4-trimethoxybenzene solution or 1,3-benzodioxole solution (30 mM, in 90/10 hexanes/isopropanol) 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 redissolved in 200 μL 90/10 hexanes/isopropanol, and analyzed on NP-HPLC to determine conversions and enantioselectivities.

MS Characterization of POP and hybrids

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. For LC/ESI-Q-TOF and low-resolution LC/MS analysis, 10 uL samples of 60 uM protein in H_2O or 100 mM NaPi pH 7.0 were injected onto the Aeris WIDEPORE column using a standard gradient from 10% ACN/90% H_2O + 0.1% formic acid to 90% ACN/10% H_2O + 0.1% formic acid over 15 minutes. Raw ESI spectra were deconvoluted using the Agilent Chemstation LC/MSD data deconvolution module or Agilent MassHunter.

Circular Dichroism (CD) stability profiles

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.

UV/Vis protocol

Samples of POP-ZA_4, POP-ZA_4-H328, and POP-ZA_4-M328 were expressed according to the protocol described. The scaffolds were prepared for bioconjugation (75 μM in 50 mM Tris-HCl buffer, pH 7.4) and split into 2 x 480 uL portions. One portion was treated with RhBCN solution (120 μL, 0.75 mM in MeCN, 0.594 mg/mL) while the second was treated with pure MeCN. The reactions were purified according to the standard bioconjugation procedure. Each hybrid and its corresponding control scaffold was buffer exchanged into 50 mM PIPES pH 7.4 + 1.75 mM NaBr and adjusted to 27 uM concentration based on A_{280}. The samples were loaded into parallel cuvettes on a dual beam Cary 5000 UV/NIS/NIR Spectrophotometer and their UV/Vis spectra were recorded from 280 nm to 650 nm. The control scaffold sample absorbance was actively subtracted from that of the hybrid sample to yield simplified difference spectra.

Expression of ^{13}C-^{15}N-labeled POP and preparation of protein NMR samples

Inoculated 2.5 mL primary cultures of *Pfu* POP-Z variants in LB (0.50 mg/mL kanamycin, 0.20 mg/mL chloramphenicol). Incubated at 37°C overnight. Prepared M9 minimal media using the following recipe:

1. M9 salts (10x):
    a. 60 g Na_2HPO_4
    b. 30 g KH_2PO_4
    c. 5 g NaCl
    d. 1 L H_2O
    e. Adjust pH to 7.4 with NaOH
    f. Autoclave
2. $^{15}NH_4Cl$ stock (10x):
    a. 5 g $^{15}NH_4Cl$
    b. 500 mL H_2O
    c. Autoclave
3. Trace elements solution (100x):
    a. 5 g EDTA
    b. 800 mL H_2O
    c. Adjust to pH 7.5
    d. 0.83 g $FeCl_3 \bullet 6H_2O$
    e. 84 mg $ZnCl_2$
    f. 13 mg $CuCl_2 \bullet 2H_2O$
    g. 10 mg $CoCl_2 \bullet 6H_2O$
    h. 10 mg H_3BO_3
    i. 1.6 mg $MnCl_2 \bullet 4H_2O$
    j. Dilute to 1 L
    k. Autoclave
4. 20% U-^{13}C glucose solution:
    a. 3 g U-^{13}C glucose
    b. Dilute to 15 mL with H_2O
    c. Sterile filter through 0.22 micron filter
5. 1 M $MgSO_4$:
    a. 12.0366 g $MgSO_4$
    b. 100 mL H_2O
    c. Autoclave
6. 1 M $CaCl_2$:
    a. 11.098 g $CaCl_2$
    b. 100 mL H_2O
    c. Autoclave
7. 1 mg/mL biotin:
    a. 10 mg biotin
    b. 10 mL H_2O
    c. Sterile filter through 0.22 micron filter
8. 1 mg/mL thiamine:
    a. 10 mg thiamine
    b. 10 mL H_2O
    c. Sterile filter through 0.22 micron filter
    d. Wrap with foil (light sensitive)
9. 771.7 mL autoclaved H_2O
10. Assemble ingredients in biosafety cabinet:
    a. 100 mL M9 salts (10x)
    b. 100 mL $^{15}NH_4Cl$ stock (10x)
    c. 10 mL Trace elements solution (100x)
    d. 15 mL 20% U-^{13}C glucose solution
    e. 1 mL 1 M $MgSO_4$
    f. 0.3 mL 1 M $CaCl_2$ (Will precipitate. Swirl to return into solution)
    g. 1 mL biotin (1 mg/mL)
    h. 1 mL thiamine (1 mg/mL)
    i. 771.7 mL autoclaved H_2O

Table 4.4 Minimal media recipe for isotope labeling of *Pfu* POP.

Antibiotic was added to the expression media (0.50 mg/mL kanamycin, 0.20 mg/mL chloramphenicol) and the cultures were split into 4 x 250 mL portions and inoculated with primary cultures. The expression cultures were incubated at 37°C with 250 rpm continuous stirring. After 10 hours, the OD=1.0 and protein expression was induced by adding 1mM IPTG, 2mM 4-Azido-l-phenylalanine and 1% (w/v) l-arabinose. The cultures were incubated for 28 hours. Cells were harvested by centrifugation at 3000 rpm for 10 minutes. They were resuspended in 25 mL 20 mM sodium phosphate buffer pH 7.4/50 mM imidazole/50 mM NaCl and lysed by sonication. The cell lysate was clarified at 15000 rpm for 1 hour. Clarified lysate was applied to a 3 x 5 mL HisTrap NiNTA columns (GE Healthcare Life Sciences) pre-equilibrated with 20 mM sodium phosphate buffer pH 6.5. The protein was purified using a gradient of 0-500 mM imidazole in 20 mM sodium phosphate buffer pH 6.5/50 mM NaCl. Fraction purity was confirmed by SDS-PAGE. Pure fractions were combined and concentrated using Amicon® 30 kD spin filters. Pure ^{13}C-^{15}N *Pfu* POP-Z variants were buffer exchanged into 20 mM NaPi pH 6.0 and concentrated to >1 mM concentration and transferred to NMR tubes. These were taken to the Center for Structural Biology NMR Facility at UIC for 2D-NMR experiments.

Scaffold modification and LC-MS/MS characterization

Biocatalysis reactions were conducted as described above. Solutions of methyl 4-methoxyphenyldiazoacetate (25 μL, 96 mM in THF) and styrene (25 μL, 485 mM in THF) were added to 480 μL solutions of ArM adjusted to 48 μM loading of dirhodium-conjugated ArM and incubated at 4°C with 600 rpm stirring for 4 hours. Control reactions containing bare scaffold (no conjugated metal) were prepared in an identical fashion with equal total protein loading

relative to the corresponding metalloenzyme-catalyzed reactions. Controls were also prepared containing no diazo or styrene (THF). 150 μL aliquots of the reaction mixtures were pulled at designated timepoints and loaded onto 0.5 mL 40K Zeba Spin Desalting columns (Thermo Fisher) pre-equilibrated with 50 mM Tris buffer pH 7.4. These were spun at 1500 g for 2 minutes. The eluent was then collected. Intact protein ESI-MS was collected with this fraction.

CNBr digestions were performed according to a modified literature procedure[27]: 100 μL of 60 uM protein sample in 100 mM NaPi pH 7.0 was treated with 37.7 μL 1 N HCl and 4.8 μL 1 mM neurotensin standard. Then, the samples were wrapped in aluminum foil, at which point 7.5 μL 5.0 M CNBr in acetonitrile was added to the mixture. These samples were incubated with 250 rpm shaking for 16 hours (fully wrapped in aluminum foil). Once the digestions were completed, they were concentrated on a Speedvac at 55°C to remove solvent and excess CNBr. The residues were then resuspended in 150 μL H_2O. These were analyzed by LC/Q-TOF MS using an Aeris WIDEPORE C8 peptide/protein column. The mobile phase consisted of: Solvent A: [H_2O + 0.1% formic acid] and Solvent B: [MeCN + 0.1% formic acid]. 20 uL of sample were analyzed using a linear gradient from 3% B to 70% B over 84 minutes.

LC-MS/MS sample preparation, mass spectrometry and proteomic data analysis

Owing to the thermal stability of POP variants, a two-step digestion procedure was employed prior to LC-MS/MS analyses. First, POP protein was cleaved at methionine residues with cyanogen bromide treatment as described above. Each CNBr digest (18 μL) was diluted into trypsin digest buffer (1 M Urea, 50 mM NH_4CO_3H, 1 mM $CaCl_2$, pH 7.8) and digested with 1 μg modified bovine trypsin (1:100 enzyme/protein ratio; Promega) at 30°C overnight. It should be

noted that the POP enzyme variants used here do not contain cysteine, so reduction and alkylation steps were omitted. Digestion was stopped by acidification to pH 2-3 with formic acid, and peptides were then desalted on P10 ZipTips (Millipore), lyophilized and resuspended in 0.1% TFA-containing mass spectrometry grade water (Sigma).

LC-MS/MS experiments were performed with an Easy-nLC 1000 ultra-high pressure LC system (ThermoFisher) using a PepMap RSLC C18 column (column: 75 μm x 15 cm; 3 μm, 100 Å) coupled to a Q Exactive HF Orbitrap and Easy-Spray Nanosource (ThermoFisher). CNBr-trypsin digested peptides (500 ng) were injected onto the column in buffer A (0.1% TFA water) and separated using the following linear gradient of buffer B (80% acetonitrile with 0.1% TFA) at 300 nL/min: 0-40% buffer B over 30 minutes, 40-90% buffer B over 5 minutes, 90-90% buffer B over 10 minutes, and re-stabilized to 0% buffer B over 5minutes. MS/MS spectra were collected from 0 to 45 minutes using a data-dependent, top-10 ion setting. Data acquisition for differential modification searches were performed with the following settings: Full MS scans were acquired at a resolution of 120,000, scan range of 150-1600 m/z, maximum IT of 50 ms, AGC target of 1e6, and data type in profile mode. Sequencing was performed by HCD fragmentation with a resolution of 15,000, AGC target of 1e5, maximum IT of 30 ms, and data type in centroid mode. Isolation window for precursor ions was set to 1.5 m/*z* with an underfill ratio of 0.5%. Peptides with charge state >5 or undefined were excluded and dynamic exclusion for all others was set to 5.0 seconds. Furthermore, S-lens RF level was set to 60 with a spray voltage value of 2.60kV. Modified-peptide searches were enabled using a lower collision energy (NCE = 18), and runs used to quantify unmodified peptides used NCE = 26.

MS2 files were generated and searched using the ProLuCID algorithm in the Integrated Proteomics Pipeline (IP2) software platform. Custom search databases were created using each POP variant protein sequence, and a concatenated decoy database was included for false discovery rate estimations. Basic searches were performed with the following search parameters: HCD fragmentation method; monoisotopic precursor ions; precursor mass range 150-6000 and initial fragment tolerance at 600 p.p.m.; C-terminal enzyme cleavage specificity at lysine and arginine residues with 2 missed cleavage sites permitted; primary scoring type by XCorr and secondary by Zscore; minimum peptide length of 3 (as known peptides of this length were present after double digestion) with a candidate peptide threshold of 500. A minimum of one peptide per protein and half-tryptic peptide specificity were required. Starting statistics were performed with a Δmass cutoff = 15 p.p.m. with modstat, and trypstat settings. False-discovery rates were set initially to 1%, however actual FDRs across all runs and searches was 0, and Δmass of precursor ions was below 5 p.p.m.. Differential modification searches to identify carbene-modified residues were performed with the settings above and allowing for up to two total differential modification sites per peptide, including oxidized methionines (+15.9949), and methyl 2-(4-methoxyphenyl)acetate-modified residues (+178.063). Differential modification searches were performed by parallel searches of each residue individually (i.e. one search of W for 178.063 adducts, one search of F for 178.063 adducts etc.). Residue localization on larger peptides was accomplished by manual inspection of fragment spectra for characteristic fragment ions as well as overall spectral match statistics (Xcorr and DeltCN). The modified FWK peptide was first observed by Q-TOF analysis and was too short for detection by standard ProLuCID searches. Therefore, this peptide and its modified variant were validated by manual inspection and assignment of chromatograms and mass spectra. Relative modification levels were

calculated by comparing the extracted ion intensity area for the unmodified peptide of interest between control and reaction scaffolds; these ratios were further normalized for potential fluctuation between runs by normalizing to extracted ion intensity area ratios of high abundance peptides observed in all runs.

4.5 ACKNOWLEDGMENTS

Special thanks to Dr. Poonam Srivastava, Dr. Hyun June Park, Dr. Hao Yang, Dr. Chen Zhang, Alan Swartz, and Brian Koronkiewicz for the development of rational mutagenesis and directed evolution methodology for ArM optimization. Additional thanks to Gihoon Lee (Moellering lab) for assistance with LC/MS/MS.

4.6 REFERENCES

(1) Yang, H.; Srivastava, P.; Zhang, C.; Lewis, J. C. *Chembiochem* **2014**, *15* (2), 223–227.

(2) Ryu, Y.; Schultz, P. G. *Nature Methods* **2006**, *3* (4), 263–265.

(3) Lewis, J. C. *ACS Catal.* **2013**, *3* (12), 2954–2975.

(4) Srivastava, P.; Yang, H.; Ellis-Guardiola, K.; Lewis, J. C. *Nat Commun* **2015**, *6*, 7789.

(5) Harris, M. N. *Journal of Biological Chemistry* **2001**, *276* (22), 19310–19317.

(6) Szeltner, Z.; Polgár, L. *Curr. Protein Pept. Sci.* **2008**, *9* (1), 96–107.

(7) Xie, J.; Schultz, P. G. *Methods* **2005**, *36* (3), 227–238.

(8) Srivastava, P.; Yang, H.; Ellis-Guardiola, K.; Lewis, J. C. *Nat Commun* **2015**, *6*, 7789.

(9) Jason W Chin; Stephen W Santoro; Andrew B Martin; David S King; Lei Wang, A.; Peter G Schultz. *Addition of p-Azido-l-phenylalanine to the Genetic Code of Escherichia coli*; American Chemical Society, 2002; Vol. 124, pp 9026–9027.

(10) In *Bioconjugate Techniques (Third edition)*; Academic Press: Boston, 2013; pp i–iii.

(11) Tookmanian, E. M.; Fenlon, E. E.; Brewer, S. H. *RSC Adv.* **2015**, *5* (2), 1274–1281.

(12) James R Carey; Steven K Ma; Thomas D Pfister; Dewain K Garner; Hyeon K Kim; Joseph A Abramite; Zhilin Wang; Zijian Guo, A.; Yi Lu. *A Site-Selective Dual Anchoring Strategy for Artificial Metalloprotein Design*; American Chemical Society, 2004; Vol. 126, pp 10812–10813.

(13) Aquino, M. A. S.; Macartney, D. H. *Inorg Chem* **1987**, *26* (16), 2696–2699.

(14) Sambasivan, R.; Zheng, W.; Burya, S. J.; Popp, B. V.; Turro, C.; Clementi, C.; Ball, Z. T. *Chem. Sci.* **2014**.

(15) Chavan, M. Y.; Zhu, T. P.; Lin, X. Q.; Ahsan, M. Q.; Bear, J. L.; Kadish, K. M. *Inorg Chem* **1984**, *23* (26), 4538–4545.

(16) Boyar, E. B.; Robinson, S. D. *Coordination Chemistry Reviews* **1983**, *50* (1-2), 109–208.

(17) Norman, J. G.; Kolari, H. J. *J. Am. Chem. Soc.* **1978**, *100* (3), 791–799.

(18) Trynda, L.; Pruchnik, F. *Journal of Inorganic Biochemistry* **1995**, *58* (1), 69–77.

(19) Clark, R. J. H.; West, D. J.; Withnall, R. *Inorg Chem* **1992**, *31* (3), 456–459.

(20) Key, H. M.; Clark, D. S.; Hartwig, J. F. *J Am Chem Soc* **2015**, *137* (25), 8261–8268.

(21) Shimahara, H.; Yoshida, T.; Shibata, Y.; Shimizu, M.; Kyogoku, Y.; Sakiyama, F.; Nakazawa, T.; Tate, S.-I.; Ohki, S.-Y.; Kato, T.; Moriyama, H.; Kishida, K.-I.; Tano, Y.; Ohkubo, T.; Kobayashi, Y. *J. Biol. Chem.* **2007**, *282* (13), 9646–9656.

(22) Kichik, N.; Tarragó, T.; Claasen, B.; Gairí, M.; Millet, O.; Giralt, E. *Chembiochem* **2011**, *12* (18), 2737–2739.

(23) Song, W. J.; Tezcan, F. A. *Science* **2014**, *346* (6216), 1525–1528.

(24) Jeschek, M.; Reuter, R.; Heinisch, T.; Trindler, C.; Klehr, J.; Panke, S.; Ward, T. R. *Nature* **2016**, *537* (7622), 661–665.

(25) Irwin, J. J.; Duan, D.; Torosyan, H.; Doak, A. K.; Ziebart, K. T.; Sterling, T.; Tumanian, G.; Shoichet, B. K. *J. Med. Chem.* **2015**, *58* (17), 7076–7087.

(26) Feng, B. Y.; Shoichet, B. K. *Nat Protoc* **2006**, *1* (2), 550–553.

(27) Andreev, Y. A.; Kozlov, S. A.; Vassilevski, A. A.; Grishin, E. V. *Anal. Biochem.* **2010**, *407* (1), 144–146.

APPENDIX I

SELECT NMR SPECTRA FOR COMPOUNDS FROM CHAPTER TWO

Figure A1.1. ^{1}H NMR spectrum of 13.

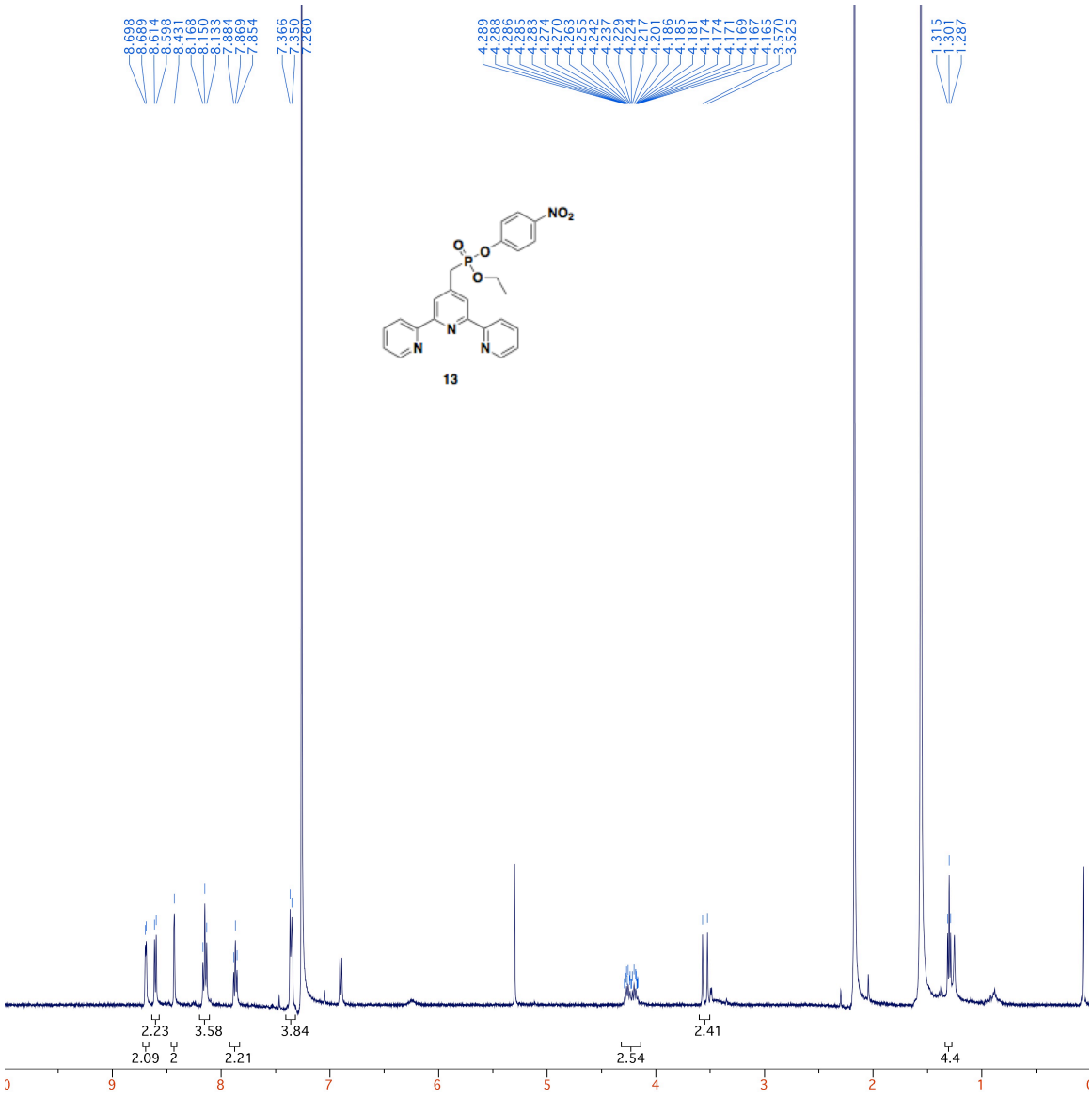


Figure A1.2. ^{31}P NMR spectrum of 13.

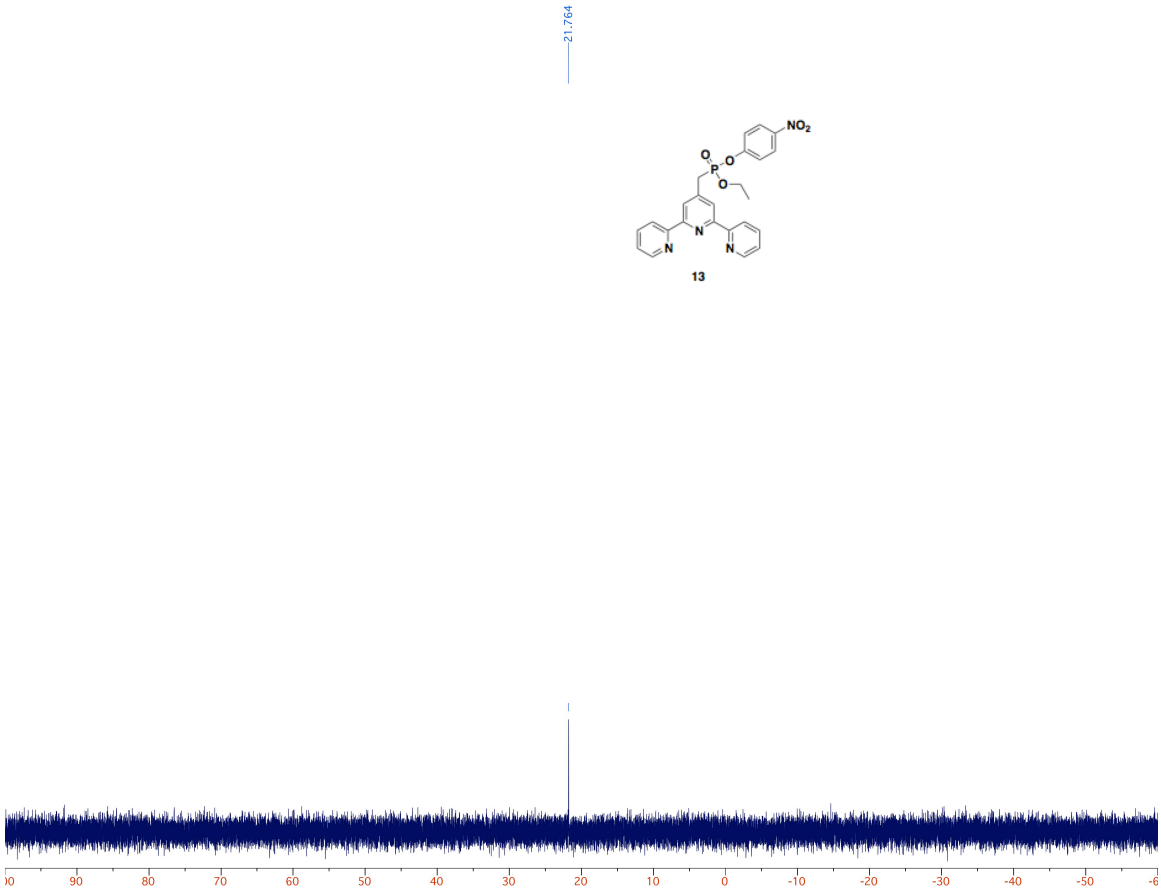


Figure A1.3. ^{1}H NMR spectrum of 18.

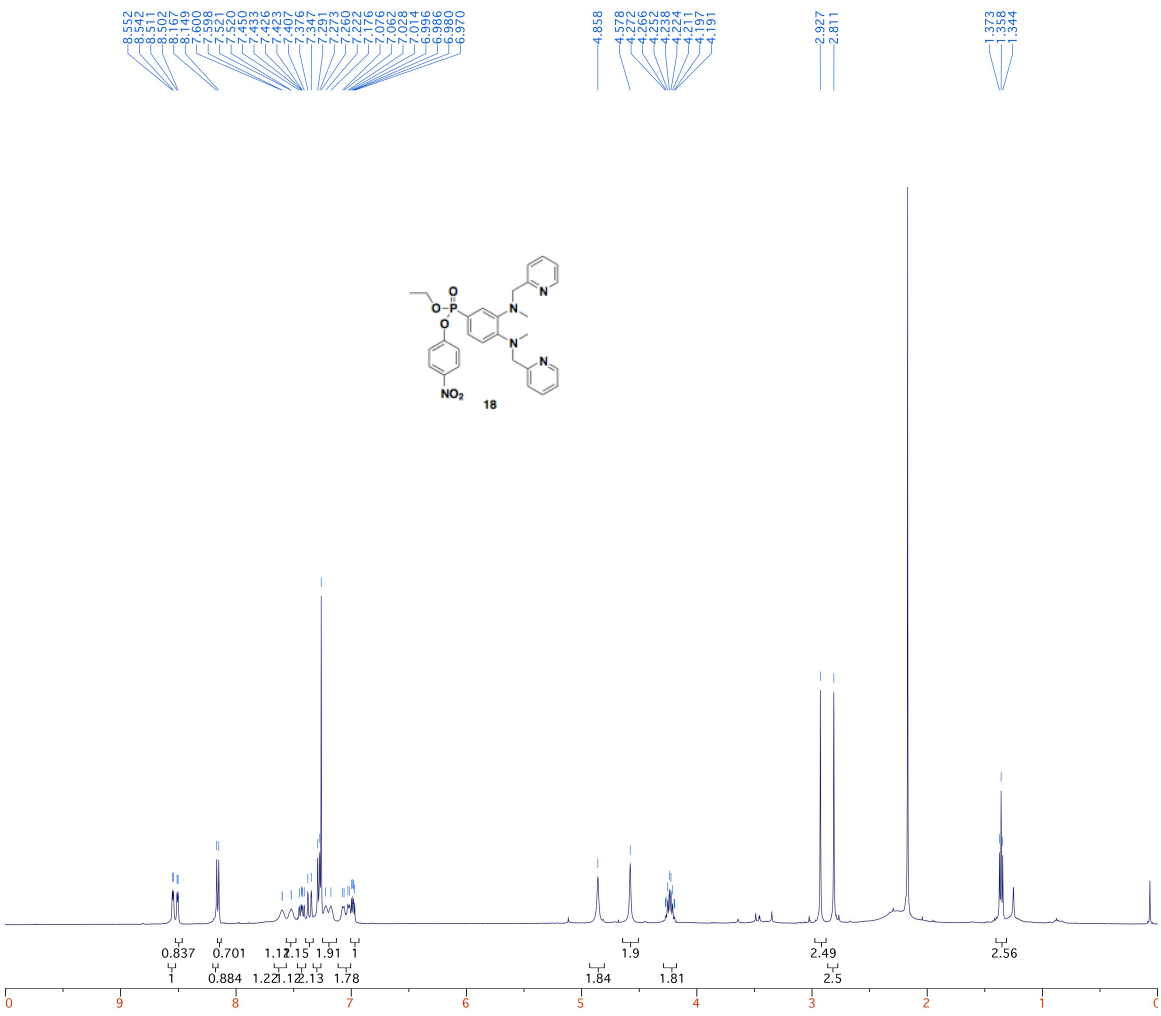


Figure A1.4. ^{31}P NMR spectrum of 18.

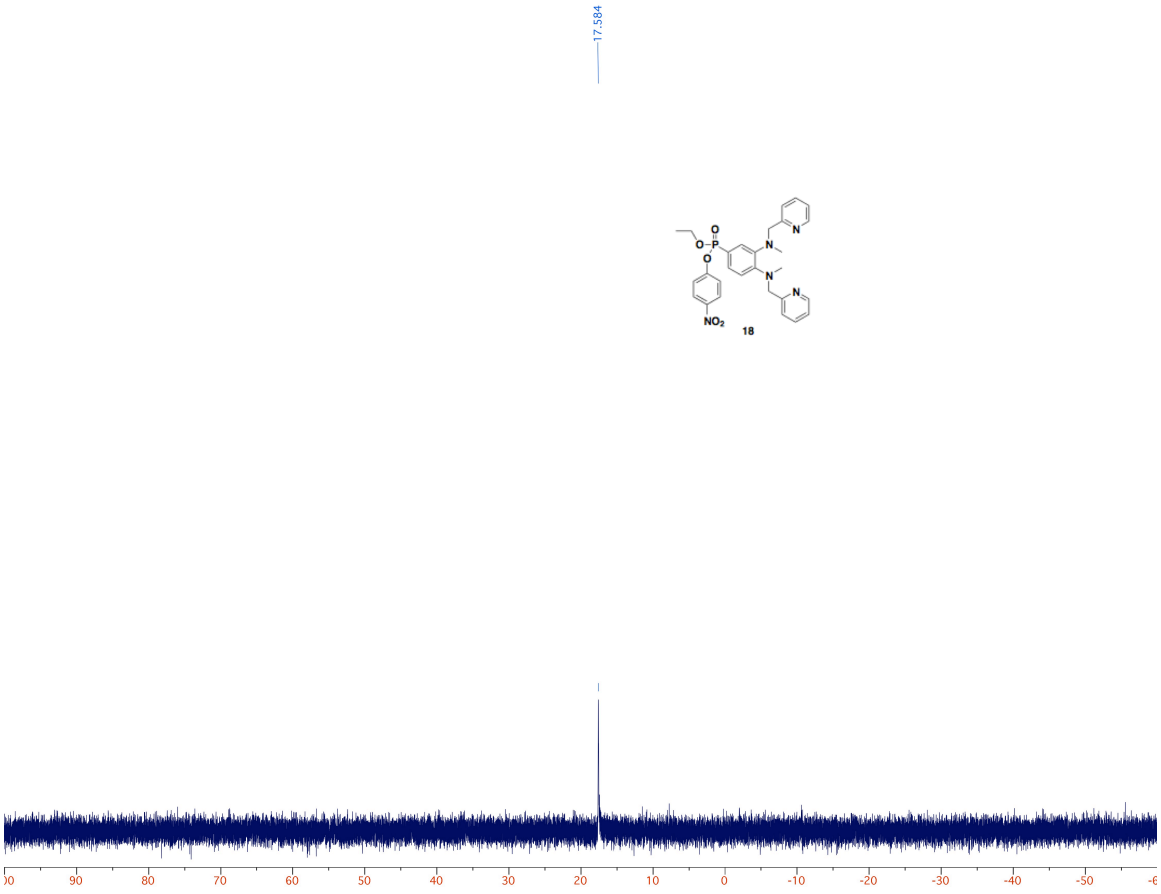


Figure A1.5. ^{1}H NMR spectrum of 25.

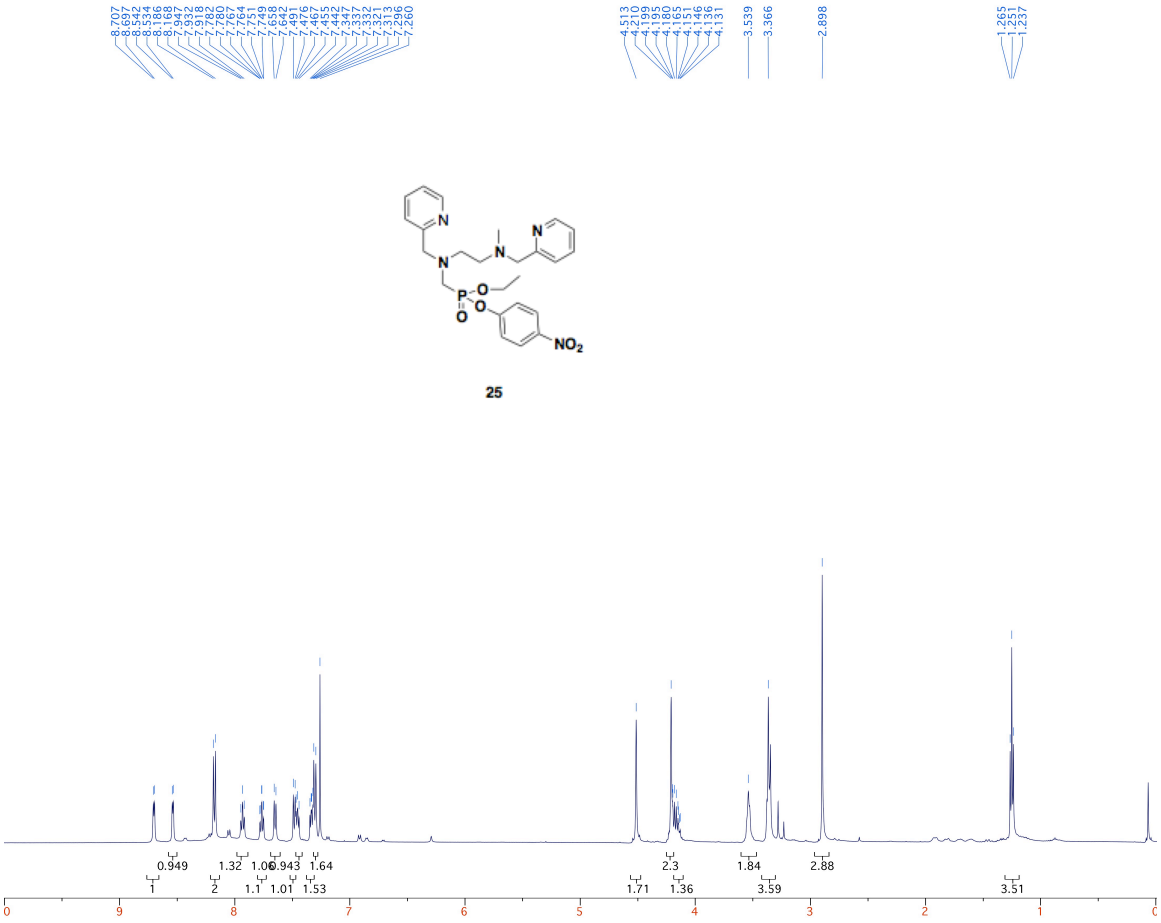


Figure A1.6. ^{31}P NMR spectrum of 25.

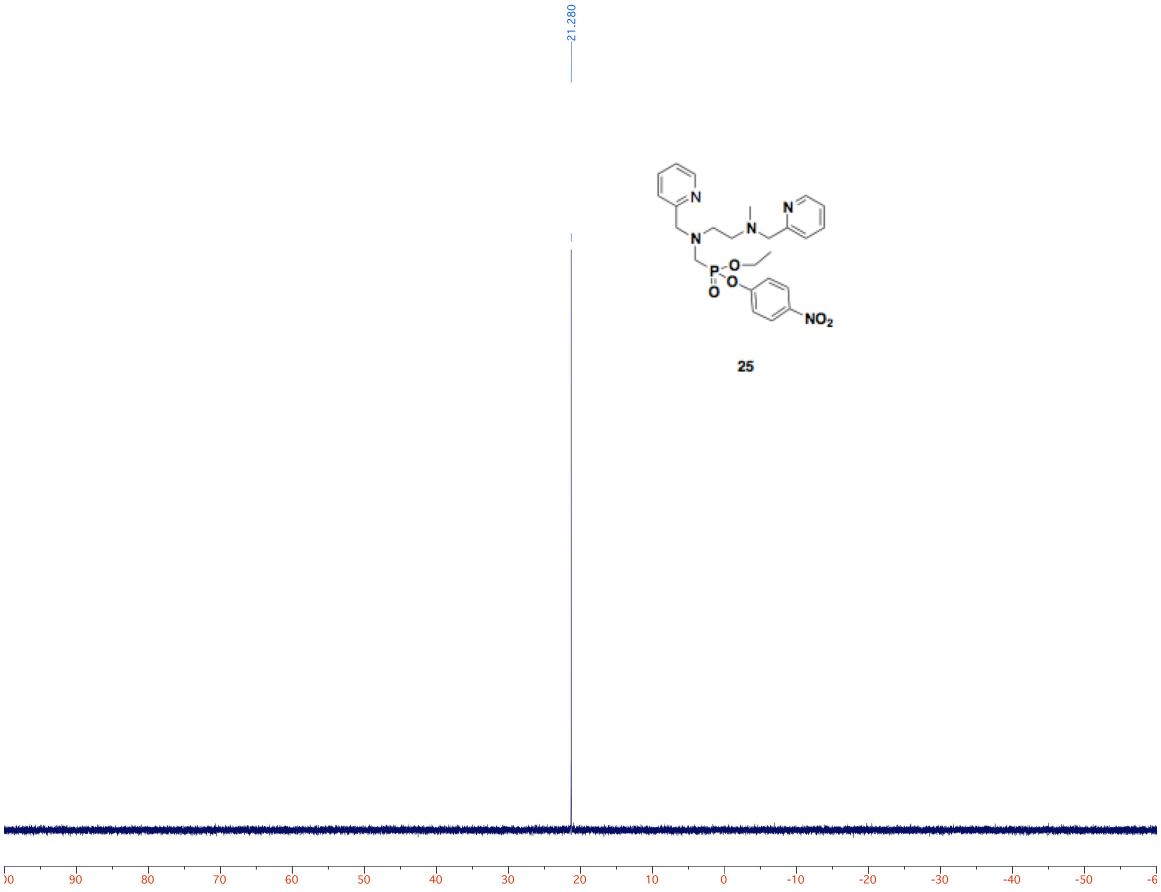


Figure A1.7. ^{1}H NMR spectrum of 33.

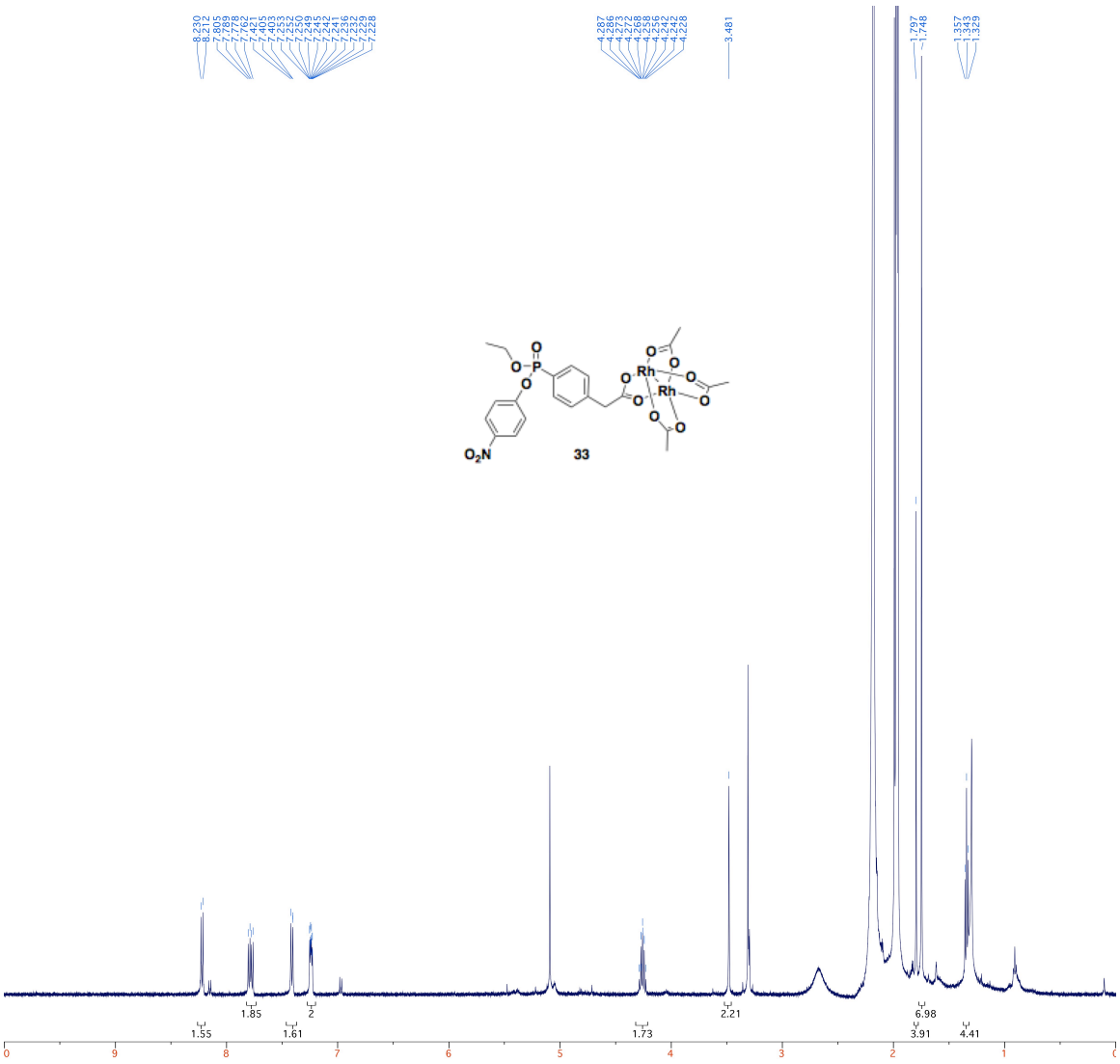


Figure A1.8. ^{31}P NMR spectrum of 33.

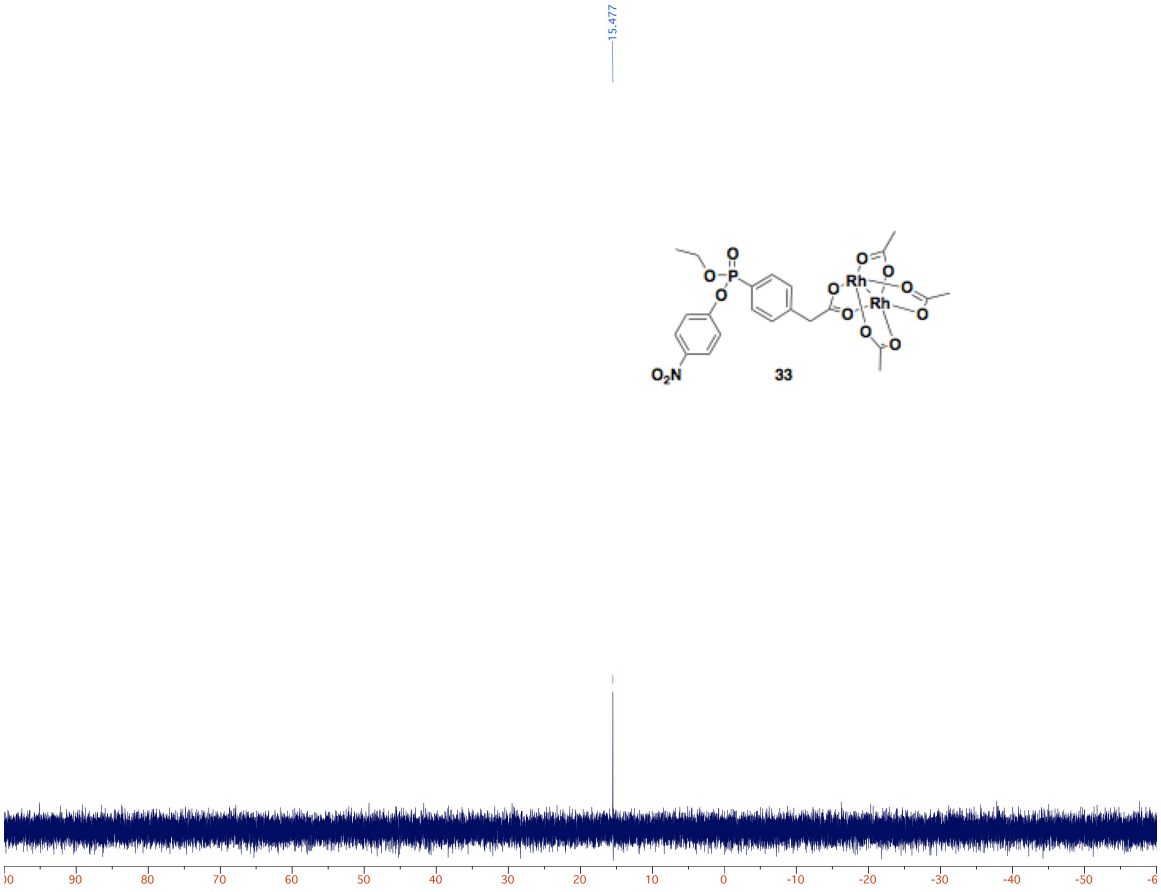


Figure A1.9. ^{1}H NMR spectrum of 37.

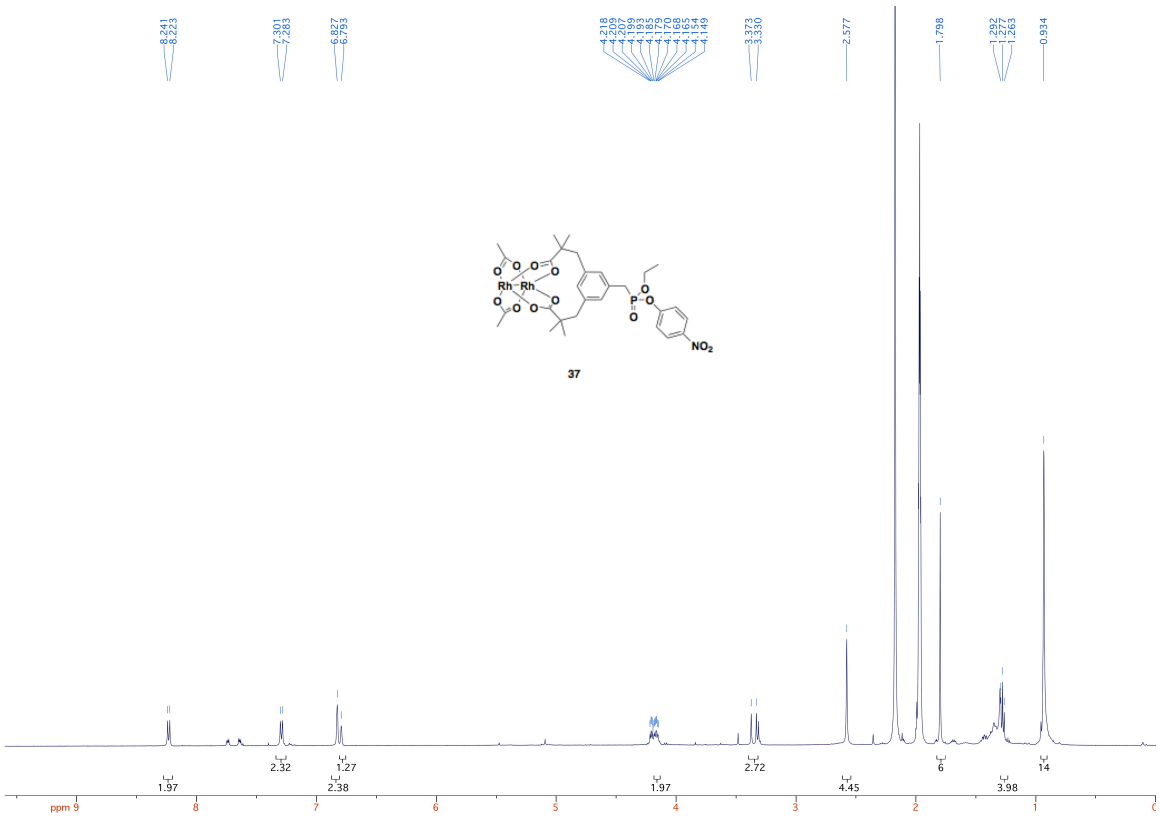


Figure A1.10. ^{31}P NMR spectrum of 37.