Title
Synthesis, Characterization, and Stoichiometric and Catalytic Reactivity of Metalloproteins Binding Noble Metals in Place of Native Metals

Author
Key, Hanna Martel

Publication Date
2016

Peer reviewed|Thesis/dissertation

Synthesis, Characterization, and Stoichiometric and Catalytic Reactivity of Metalloproteins Binding Noble Metals in Place of Native Metals

By

Hanna Martel Key

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor John Hartwig, Chair

Professor Matthew Francis

Professor Douglas Clark

Fall 2016

Abstract

Synthesis, Characterization, and Stoichiometric and Catalytic Reactivity of Metalloproteins Binding Noble Metals in Place of Native Metals

By

Hanna Martel Key

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor John Hartwig, Chair

Catalysis by enzymes and transition metal complexes enables the modern synthetic strategies used to prepare bulk and fine chemicals. However, despite their exquisite selectivity and potential for evolution, metalloenzymes are rarely used in synthetic chemistry, in part due to the limited scope of reactions catalyzed by natural metalloenzymes. To expand the scope of substrates and the range of transformations with which metalloenzymes react, two distinct approaches have been followed. One approach designs promiscuous enzymes by directed evolution. An advantage of this method is that it exploits native, selective substrate binding sites. An alternative strategy creates artificial enzymes by affixing an abiological transition metal complex within a protein scaffold by cysteine bioconjugation or biotin-labeling. Although this method has been used successfully to impart abiological activity onto enzymes, a significant drawback is that the native binding pocket of the protein scaffold is occupied by the bulky metal complex that is introduced, excluding the substrate from its natural binding site.

At the intersection of these two approaches lies a third strategy that is rarely followed: the simple substitution of the native metal of a natural enzyme with an abiological metal. This strategy has the potential to create enzymes containing any metal of interest, while retaining the natural, evolvable substrate-binding site of the protein scaffold. We have sought new and superior approaches to the preparation, characterization, and application of diverse catalysts of this class, formed from both proteins that bind metals using organic cofactors and proteins that bind metals directly using amino acid side chains.

Despite challenges confronting their preparation, characterization, and application, our studies have revealed that diverse members of this class of artificial metalloenzymes can indeed be prepared efficiently and characterized as comprehensively as one would characterize natural enzymes and small molecule metal complexes. Moreover, the stoichiometric and catalytic reactivity of the noble metal sites within these proteins can be controlled by mutagenesis of the native substrate-binding site surrounding the abiological metal. With these studies as a foundation, we and others can now seek the creation of nearly limitless combinations of noble metals and protein scaffolds that can catalyze reactions with activities and selectivities not possible for either enzyme or transition metal catalysts alone.

Dedication

From afar, many goals seem impossibly out of reach. Yet, day-by-day, step-by-step, suddenly the once-impossible is on the doorstop.

To all my family, friends, and colleagues who have been a part of those innumerable small steps culminating in this thesis

Table of Contents

Acknowledgements

Thank you to those who prepared me to embark on a PhD in Chemistry. To my first chemistry teacher, Keith Heidmann. To the inspirational, dedicated faculty of Grinnell College who shaped me as a scholar, a scientist, an educator, and a person, including Stephen Sieck, Andy Mobley, James Lindberg, Leslie Lyons, and Mark Levandoski from the chemistry department, Kathy Jacobson, Vince Eckhart, Ben DeRidder, and Greg Ruthig from the biology department, Eric Carter from the anthropology department, and especially Jean Ketter and Paul Hutchison from the education department. To my intelligent, thoughtful, and motivated classmates who shared and shaped my time at Grinnell.

Thank you to those who were instrumental to the scientific achievements of my PhD studies. To my advisor Prof. John Hartwig, who graciously accepted me as a member of his research group and who guides his students' scientific curiosity, work ethic, perseverance, professionalism, and standard for excellence by example. To our collaborator, Prof. Douglas Clark, for his time, his advice, and his unfailingly positive outlook. To the scientists who lead the shared research facilities at Berkeley, especially Jeff Pelton from the 900 MHz NMR facility and Tony Iavarone from the mass spectrometry facility for sharing their time, their expertise, and their guidance, and to the staff of the DNA sequencing facility, for their quick service and dedication to serving the research groups on campus. And especially to Dr. Pawel Dydio, my collaborator for the final two years of the PhD studies. Working so closely with someone was an incredible research experience, and I could not have asked for research partner who was more dedicated, harder working, or easier to work with.

Thank you to the talented members of the Hartwig group among whom I was lucky enough to work for five years. To Andrey, Paramita, Mark, Adi, Julian, Pawel, Jing, Matthias, and Cynthia who had joined me in the "BioBay," thank you for your friendship and your (often-needed) senses of humor during our infancy in this research area. To Zhennan, Hiroki, and Rhett, I wish we had overlapped longer in the lab, and I wish you the best of luck in carrying on this research direction in our group. To Jenn, Andrew, and Vida, my undergraduate research students, who worked so hard to contribute to our studies, I also wish you the best as you figure out your steps. Finally, thank you to the "other" members of Hartwig group who were not luckly enough to work in the BioBay, but still managed to not think of us as total aliens, especially Ankit, Mike, Zach, Allie, Rashad, Jason, and Konstantin.

Finally, thank you to my family who has supported me my entire life. To Malka, Mom, Grandma, and Bobie, the strong, independent, women in my life whose perspectives, accomplishments, and successes have always inspired my own. To my Dad, who always believed his girls could do anything. And to Micah, who has endured the highs and lows, triumphs and tumbles, achievements and rejections of a PhD at my side, supporting me every step of the way. I am so proud of the work you do, and I admire your courage to leave everything you knew four years ago to join me in California.

Chapter I. Artificial Metalloproteins and Metalloenzymes

Overview and Motivation

From the simple proton to complex proteins, diverse molecules function as catalysts that enable modern approaches to synthetic organic chemistry.[1] Chemists have been long inspired by natural enzymes that effect transformations under mild conditions with rates and selectivities that are rarely mimicked by synthetic catalysts (Fig. 1).[2] Nonetheless, in certain respects, catalysts created by synthetic chemists have surpassed those created by nature. By employing molecules, particularly transition metal complexes not found in living systems, catalysts for transformations unknown in nature have been discovered and developed for synthetic processes.[1]

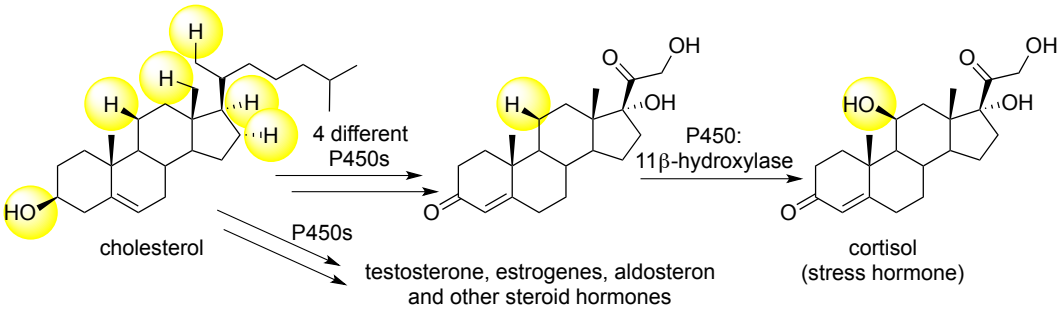


Figure 1. Biosynthesis of steroids, such as cortisol, from cholesterol by sequential oxidation reactions catalyzed by discrete cytochrome P450 enzymes.

The transition metals are a powerful platform from which to create catalysts for otherwise unachievable reactions, and both synthetic chemists and nature have incorporated transition metals into catalysts.[1,3] The reactivity and activity of natural metalloenzymes and synthetic metal catalysts are modulated by the electronic properties of the donor atom(s) binding the metal and the particular steric environment that the ligand or protein creates around the metal. Consequently, iterative modification of the molecular structure of the catalyst can lead to catalysts with improved functions (Figure 1).[4] In nature, evolution is the natural process by which this occurs over time. However, evolution of enzymes can also be accomplished in the laboratory to create variants of natural enzymes with new properties.[5] In specialized cases, the property of interest (eg. antibiotic resistance) can be evolved using life- and death-based selection, but more commonly, laboratory evolution requires a separate assay to determine whether the targeted properties (such as activity or selectivity) have been achieved, from which the most promising variants are manually identified and subjected to further variation.[4,6-8] Thus, the overall workflow to optimize an enzyme by laboratory evolution is not unlike the process typically used to optimize a synthetic catalyst, with cycles of structural modification ultimately giving rise to the most suitable catalyst for a transformation of interest.[9]

Despite these similarities, both the process to optimize an enzyme and nature of the enzyme structure offer considerable advantages for catalysis. Recent advances in molecular biology have streamlined the process to generate enzyme variants by which the genes encoding hundreds, thousands, or even millions of enzyme variants can be generated rapidly, and the library of unique proteins encoded by those genes are synthesized by bacterial hosts in parallel and in a

single step.[10] In contrast, the rapid preparation of libraries of organic ligands is difficult. The synthesis of ligands with varied donor atoms and with varied steric and electronic properties can rarely be accomplished by following a standard synthetic and purification protocol for all members of the library. Moreover, the synthesis of a complex chiral ligand is usually several steps, in sharp contrast to the recombinant expression of enzymes, which creates complex catalysts in a single step process.

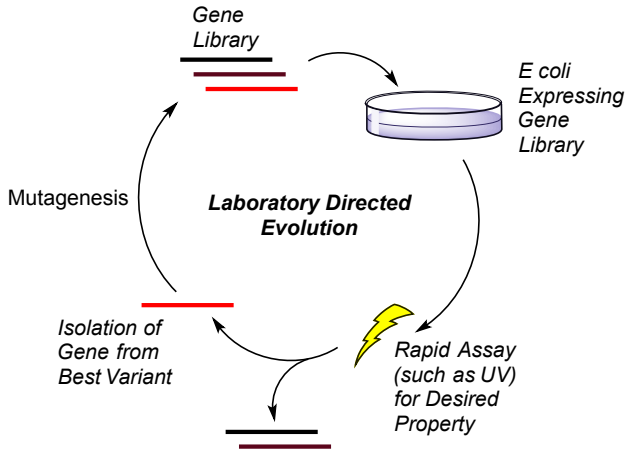


Figure 1. Graphical depiction of the strategy to create enzymes with new properties by laboratory evolution.

In addition to the ease with which many enzymes are prepared, the structurally complexity of enzymes is unparalleled in synthetic catalysts. The active site of an enzyme can surround the substrate completely, creating strong interactions with the entirety of the substrate molecule that enable favorable preorganization of the substrate and stabilization of the transition state.[11] Consequently, enzymes catalyze many processes with rates and activities that have never been accomplished with any other class of catalysts.[12] While the potential exists to create organic ligands of increasing complexity in order to enable comparable reaction outcomes in transition metal catalysis, this potential is inherently limited by the aforementioned challenges to synthesize libraries of such compounds.

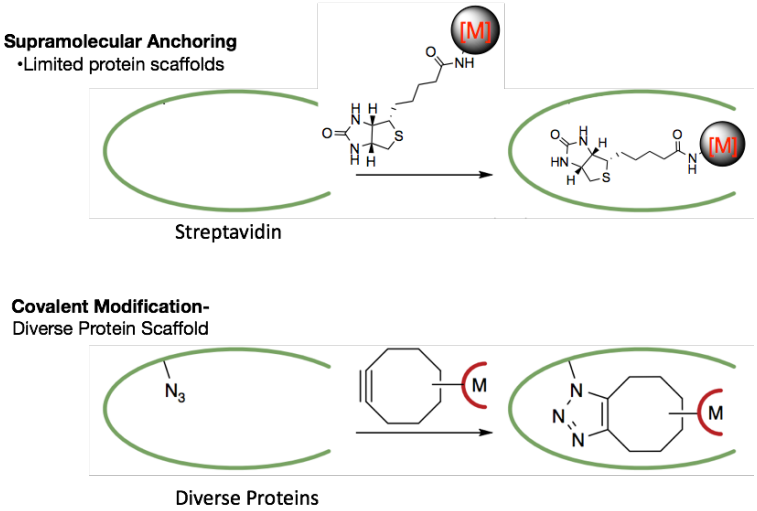


Figure 3. Graphical depiction of two strategies to create artificial metalloenzymes by anchoring pre-formed transition metal complexes within the pockets of proteins that do not typically contain metals.

To combine the activity, selectivity, and evolutionary potential of enzymes with the diverse reactivity of classic transition metal complexes, researchers have sought to create hybrid catalysts in which an abiological metal or metal complex is affixed within a protein scaffold.[13,14] Most commonly, such "artificial metalloenzymes" have been created by anchoring transition metal complexes within binding sites of natural proteins by covalent bioconjugation (e.g. cysteine alkylation) or dative affinity-labeling strategies (e.g. biotinylation) (Fig. 3). While in certain cases this strategy has been used to create enantioselective catalysts for abiological transformations (such as Rh-catalyzed annulation and cyclopropanation)[15,16] at the outset of our studies, there were no examples of artificial metalloenzymes reacting with all three crucial characteristics of natural enzymes: fast kinetics, high selectivity, and facile evolution. [17,18] Moreover, no artificial metalloenzyme had been engineered to catalyze any transformation that cannot already be accomplished with either a natural enzyme or a classic transition metal catalyst alone.

To create artificial metalloenzymes with strong potential to address these limitations, we hypothesized that suitable enzymes could be formed by replacing only the metal of a native metal protein or enzyme for a new metal with a distinct reactivity profile (Fig. 4).[19,20] This strategy confers several advantages. Most importantly, while incorporation of unnatural metal complexes within the cavities of proteins creates enzymes with no defined site to bind the substrate adjacent to the catalytic metal, metalloproteins reconstituted with abiological metal centers retain their native, evolvable binding site for a substrate. Therefore, these enyzmes should retain the potential to be evolved to generate variants displaying the desired levels of selectivity and activity for particular substrates of interest.

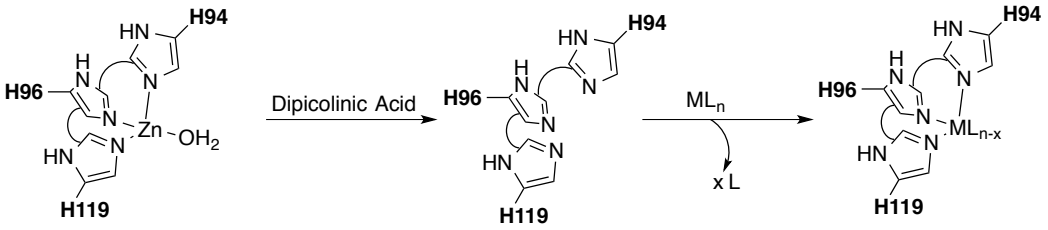


Figure 4. Strategy to create artificial metalloenzymes by replacing the native metal of a natural metalloenzyme with a different metal. This figure depicts the replacement of the zinc-site in carbonic anhydrase with a new metal by first removing the native Zn followed by addition of a new metal.

We have sought new and superior approaches to the preparation, characterization, and application of diverse catalysts of this class, formed from both proteins that bind metals using organic cofactors and proteins that bind metals directly using amino acid side chains. Though each class of protein scaffold presents unique challenges and opportunities, our research has revealed the possibility to exploit either in forming enzymes containing noble metals whose reactivity is controlled by the protein scaffold.

1.1 The Evolution of Biocatalysis in Synthetic Chemistry

The advantages that enzymes offer in synthetic chemistry are substantial.[18] Enzymes catalyze organic transformations with exquisite levels of selectivity, including chemo-, regio-, and stereoselectivity (both enantio- and diastereo-), as well as site-selectivity.[21] In addition, enzymes are often specific for one substrate among a mixture of potential reactants.[21] Finally, enzymes accomplish challenging transformations with high activity and under mild conditions, making enzymes attractive for both industrial scale processes, where mild conditions create cost savings,[22] and late-stage functionalization of complex molecules,[23] which cannot withstand harsh reaction conditions. However, the immense potential of enzymes in synthesis is tempered by the limited substrate scope with which natural enzymes react and the limited set of transformations that can be catalyzed by natural enzymes.

Due to these limitations of natural enzymes, the application of enzymes in synthesis was originally confined to 1) enzymes whose biological functions are of synthetic value or 2) enzymes that are naturally promiscuous – that is those that react with a wide range of substrates without the need for protein engineering.[10] Even with these limitations, a range of enzymes have been exploited in industrial chemistry. The formation and cleavage of ester and amide bonds, the resolution of alcohols and epoxides, and the reductions of carbonyls and alkenes catalyzed by natural enzymes have been routinely applied to the preparation of bulk chemicals and detergents for decades, and in more limited cases, these classes of enzyme-catalyzed transformations were also applied to the synthesis of fine chemicals and pharmaceuticals.[10]

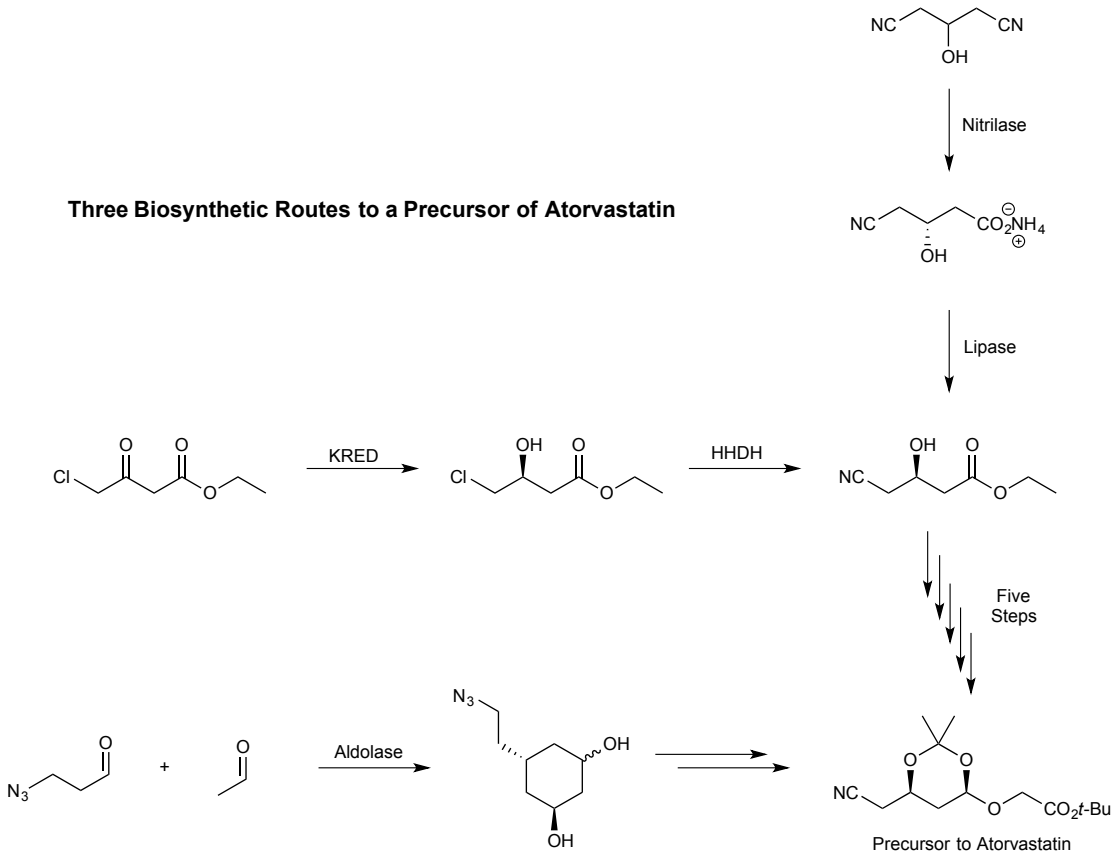


Figure 5. Three of several known biosynthetic routes used in the synthesis of atorvastatin, the active ingredient in Lipitor, a multibillion dollar drug. The various biosynthetic approaches differ in the class of enzyme used, the cost of the starting materials, and the number of stereocenters introduced.

Starting in the 1980's, the emergence of structure-based protein engineering strategies brought what is considered to be a new (second) wave of biocatalysis to synthetic chemistry.[10] New tools meant that the structures of enzymes could be modified to create variants that would react effectively with abiological substrates that reacted poorly with wild-type enzymes. Synthetic routes to drugs (such as diltiazem and statins, Fig. 5),[24] agrochemicals,[25] cosmetics (such as myristyl myristate),[26] and polymers (such as poly acrylamide)[27] were devised to include enzyme catalysis. However, the application of enzymes was still limited by the low efficiency of methods to design rationally, generate, and evaluate enzyme variants.

Most recently, the combination of newer molecular biology tools with superior approaches in bioinformatics and computer modeling has defined a third wave of biocatalysis in synthetic chemistry.[10] Frances Arnold and Pim Stemmer, among others, advanced the strategy of laboratory "directed evolution" of enzymes toward non-native properties.[28,29] Originally, their approach relied on cycles of iterative, random changes to the structure of enzymes (accomplished by various means), followed by screening of the resulting, often vast, libraries for variants with improved properties.[30] The continual emergence of superior structural data and methods to model enzymes has allowed the design of smaller, "smarter" libraries that net larger improvements in properties per mutant evaluated.[31] More efficient methods to evolve properties of interest into enzymes led to the creation of enzymes with higher suitability for organic

synthesis, including enzymes that are thermally stable (Fig. 6),[32] that operate in the presence of high concentrations of organic solvent (Fig. 7),[33] and that react with substrates for which the wild-type enzyme is completely inert.[8]

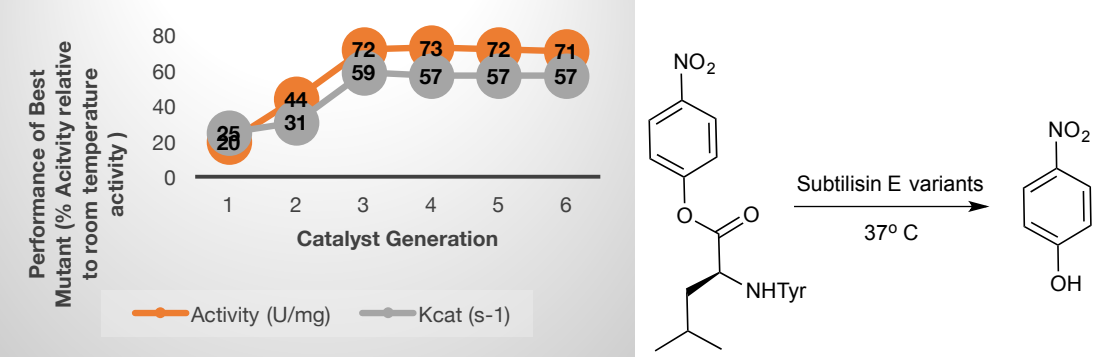


Figure 6. Optimization of the activity of subtilisin E at elevated temperature by a series of generations of directed evolution. The activity was optimized using a model reaction producing para-nitrophenol, which can be detected easily by UV.

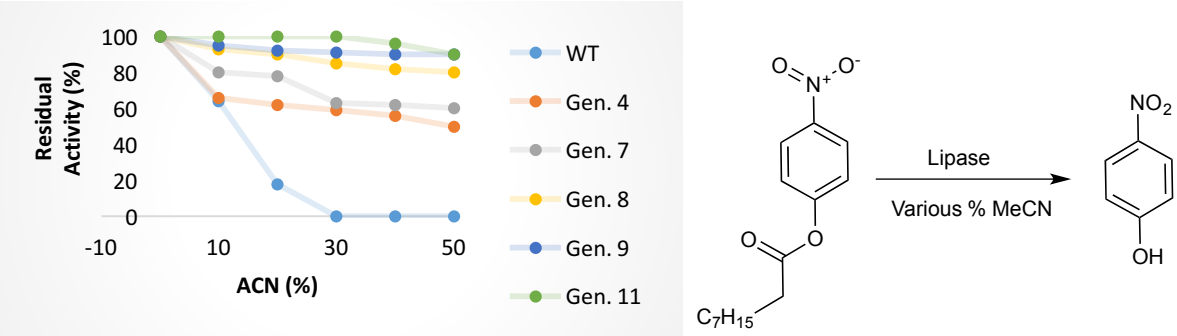


Figure 7. Optimization of the activity of a lipase in the presence of acetonitrile (CAN). The activity was optimized using a reaction producing para-nitrophenol.

In what could be considered a new, yet emerging era of biocatalysis, researchers seek to address the lingering limitation of the scope of reactions that enzymes catalyze. In some cases, directed evolution has been applied to seek enzymes that catalyze classes of reactions that are not catalyzed by the wild type enzyme.[34] In a few cases, enzymes for specific functions have been created by de novo design.[35] A third, more distinct approach, aims to create enzymes with new functions by incorporating abiological metal cofactors within proteins to create what are referred to as artificial metalloenzymes.[13,14] Though nature uses only a limited set of metals within natural metalloenzymes, chemists have created a vast number of reactions catalyzed by complexes containing metals that are not known to be used by living systems.[1,22] Compared to creating new enzymatic functions by evolution alone, artificial metalloenzymes offer an alternative strategy to access enzymes with predictable functions distinct from those created by nature.

1.2 Engineering Reaction Promiscuity into Natural Metalloenzymes

Metalloenzymes- enzymes binding a metal, a metal cluster, or a metal organic cofactor- are prevalent in nature.[3] Metalloenzymes are uniquely suited for promiscuity, because transition

metals have been found to display a range of catalytic activities.[1] In fact, nature itself exploits this phenomenon. By divergent evolution, among other evolutionary mechanisms, nature recycles metal sites to create enzymes with diverse functions.[36] For example, Fe-porphyrin (Fe-PIX) binding enzymes catalyze a broad range of reactions, including C-H oxidation,[36] C-H halogenation,[37] alcohol oxidation,[38] sulfide oxidation,[39] and ether[40] and amine[41] dealkylation, with the discrete protein structure controlling which substrates are accepted and which reaction pathways are followed (Fig. 8). Although evolution is typically a slow process, strong evolutionary pressure also has created metalloenzymes with new functions on a much shorter time scale. Less than forty years after the initial applications of the herbicide atrazine, researchers identified a new enzyme – atrazine chlorohydrolase- that cleaves C-Cl bonds to begin the catabolism of the herbicide.[42] The enzyme differs by eight amino acids from the previously characterized melamine hydrolase that cleaves C-N bonds.

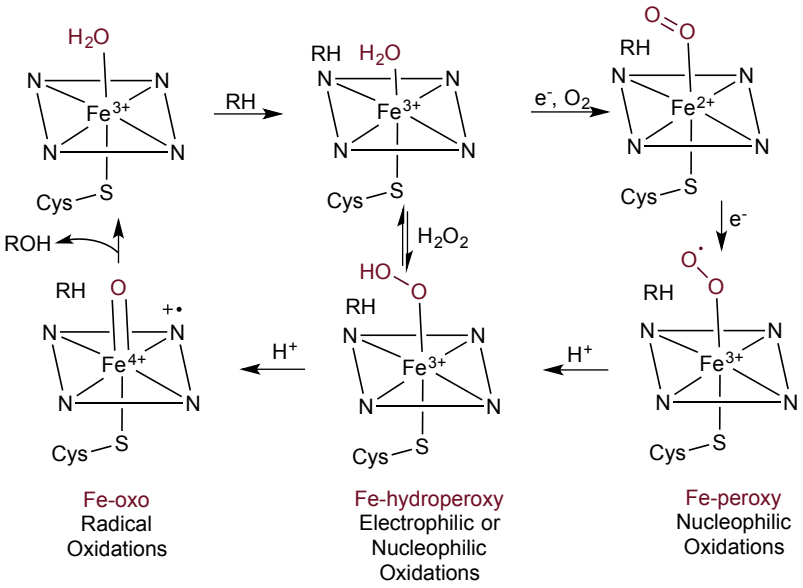


Figure 8. Reactivity of oxygen-species generated at the active sites of Fe-porphyrin containing enzymes. Interception of the various reactive oxygen species by a substrate leads to different reaction outcomes.

In other cases, metal sites within enzymes are sufficiently promiscuous that new reactions can be achieved without any evolution of the protein scaffold.[43] For example, while the zinc-site of carbonic anhydrase (CA) catalyzes the interconversion of CO_2 and bicarbonate,[44] this enzyme also catalyzes other reactions, such as the hydrolysis of phenyl acetates, that can be monitored conveniently by UV-Vis (Fig. 9).[44] Therefore, biochemical studies of CA are often accomplished using this promiscuous reactivity rather than the native activity.[45]

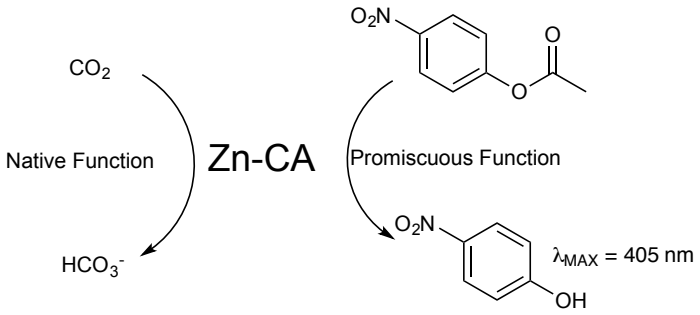


Figure 9. The native and promiscuous functions of the Zn-site of carbonic anhydrase.

These and other examples of reaction promiscuity that occur or evolve naturally have motivated research aimed at creating metalloenzymes with new functions by laboratory evolution. The most notable examples of this research direction have created heme (iron porphyrin, Fe-PIX) proteins with reaction promiscuity that is broader than that in known natural enzymes.[34] Fe-PIX proteins catalyze oxidations reactions that proceed through an Fe-oxo intermediate, which subsequently reacts with C-H bonds to form alcohols or adds to alkenes to form epoxides.[46] Pioneering work by Frances Arnold and coworkers showed that P450-BM3 (a widely studied P450) can activate ethyl diazo acetate to form an Fe-carbene intermediate that is structurally analogous to the native Fe-oxo intermediate. [47] They showed that this Fe-carbene can react with styrenes to form cyclopropanes.[47,48] Directed evolution produced mutants of P450-BM3 that catalyze this abiological transformation with high enantio- and diastereoselectivity (Fig. 10). Crucially, they found that modification of the amino acid that coordinates the Fe-center in the axial positions attenuated significantly to the activity of the enzyme.[34,48,49] From this initial success, Arnold and others have shown that heme proteins, including both P450-BM3 and myoglobin, can be used as catalysts for other transformations, including the insertion of carbenes into N-H[50,51] and S-H bonds,[52] the intramolecular insertion of nitrenes into C-H bonds,[53,54] and the addition of nitrenes to styrenes[55] and sulfides (Fig. 11-12).[56]

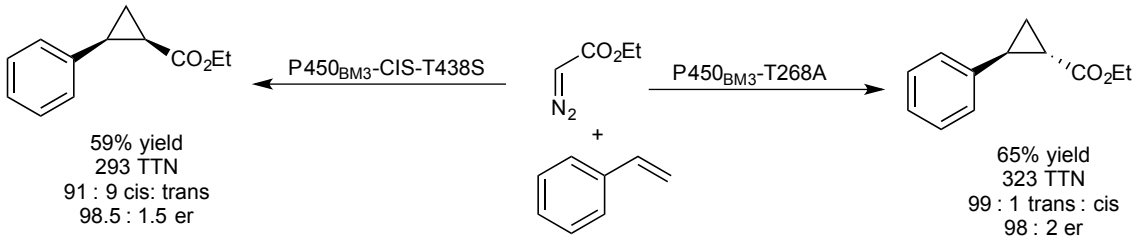


Figure 10. Evolved variants of P450-BM3 have been created to catalyze the addition of carbenes to styrenes to form cyclopropanes, in analogy to the native function of P450s that add oxygen to alkenes, forming epoxides.

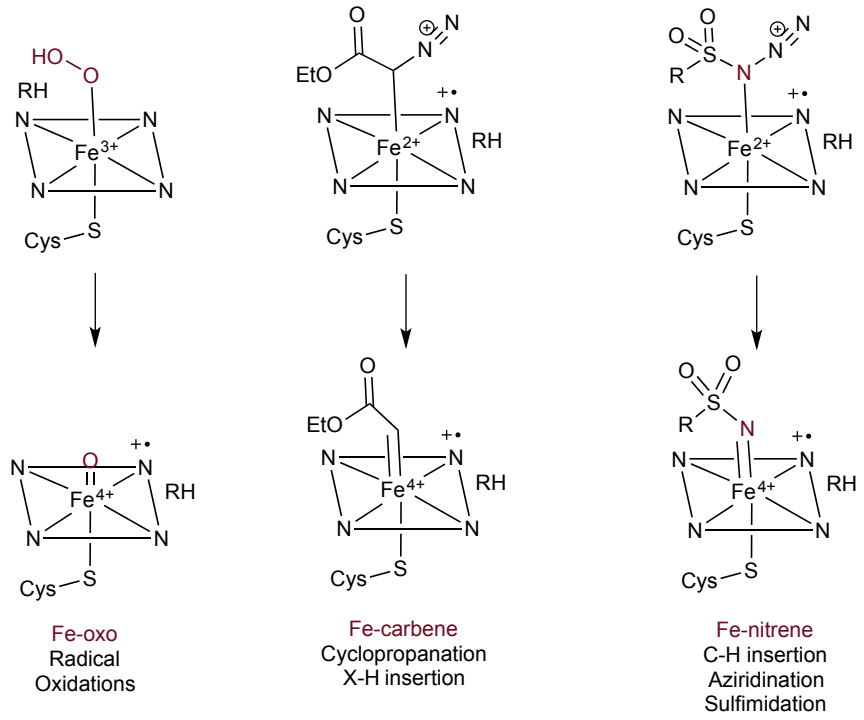
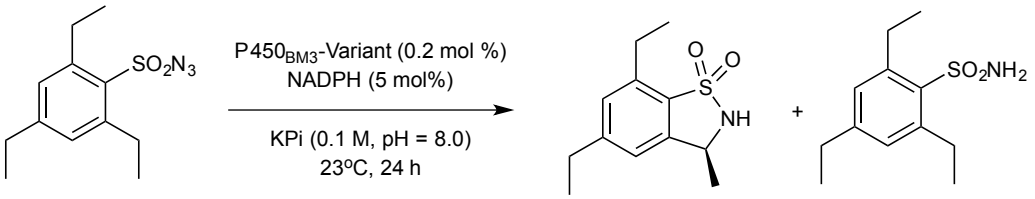


Figure 11. Depiction of the mechanism by which the Fe-site of Fe-porphyrin containing enzymes can activate molecular oxygen, diazo compounds, or azides to form isoelectronic reactive intermediates.



entry	variant	TTN	er
1	WT	2.1	n.d
2	$P411_{BM3}$	32	60 : 40
3	T268A	15	68 : 32
4	$P411_{BM3}$- T268A	120	79 : 21
5	$P411_{BM3}$- CIS-T438A	383	96.5 : 3.5

Figure 12. Evolved variants of P450-BM3 have been created to catalyze the insertion of sulfonyl-nitrenes into C-H bonds in an intramolecular fashion, in analogy to the native function of P450s that add oxygen to C-H bonds, forming alcohols.

Despite these remarkable accomplishments, the scope of transformations and range of substrates with which Fe-porphyrin containing enzymes react is limited. Cyclopropanation reactions occur only in the presence of terminal, activated alkenes.[34] Carbene insertion reactions proceed into weak S-H and N-H bonds, but not stronger O-H or C-H bonds.[34] Finally, nitrene insertion reactions occur only in an intramolecular fashion with a limited set of sulfonyl azide precursors, and the yields of the reactions are significantly limited by the poor chemoselectivity of the Fe-porphyrin, which is an equally or more efficient catalyst for the reduction of sulfonylazides to sulfonamides.[34]

These limitations encountered in the evolution of P450s toward new functionality exemplify the broader challenges and limitation of relying on only laboratory evolution to create new functions in existing enzymes. First, nature is inherently handicapped for catalysis, in that it does not incorporate many of the late second and third row metals (such as Rh, Ir, Ru, Pd, Pt, Au, and Ag) that offer modes of reactivity that are not accessible with other metals or by organocatalytic approaches.[1,3] While nature has indeed created enzymes to catalyze exceptionally challenging transformations,[12] to create by laboratory evolution an enzyme with a function for which no natural enzyme has a related function is exceptionally challenging because 1) it not obvious from what enzyme such an evolution should begin and 2) the strategy of directed evolution is predicated on achieving a measureable level of success from the start. Therefore, unless the initial enzyme(s) to be evolved or the first round of variants created from those enzymes display any activity, the system lacks a basis for directing further evolution.[5]

1.3 Synthesis and Application of Artificial Metalloenzymes Created by Dative or Covalent Anchoring Strategies

To create enzymes containing abiological metals, one of two related strategies have been followed most often (Fig. 13).[13] In both cases, an ancillary ligand of a typical organometallic complex is modified with functionally that will allow it to be anchored within the cavity of a

protein. By one strategy, a dative affinity label (such as biotin) is used to modify the ancillary ligand, enabling a supramolecular interaction between the metal complex and a complementary protein (such as streptavidin).[57] Alternatively, the ancillary ligand can be modified with a reactive group (such as a maleimide, an alpha-halo amide, or a strained alkyne[58]), such that the metal complex can be covalently attached to a genetically encoded natural or unnatural amino acid (such as cysteine or an amino acid containing an azide) by a classic organic reaction (S_N2, conjugate addition, or cycloaddition).[13]

These strategies are advantageous because they offer the potential to place nearly any type of transition metal complex within a protein. In addition, nearly any protein with a suitable pocket to contain a metal complex can serve as a host. A significant disadvantage of this strategy is that while the site at which the ancillary ligand is attached to the protein is well defined, the site at which the metal lies relative to the protein may be unknown, hindering the potential to evolve the resulting enzyme. Furthermore, if the cavity of the protein it too large, the metal freely rotate among many conformations within the protein, resulting in an amorphous active site. On the other hand, if the cavity it too small or too close to the surface, the metal site may not lie within the sphere of influence of the protein at all. Finally, even if the metal site is fixed rigidly at one site, an appropriate site adjacent to the metal to bind the substrate may not be present. Therefore, despite the broad potential of this strategy, judicious selection of metal complexes and protein hosts is still required in order to develop an effective catalyst. The research efforts to create artificial metalloenzymes that overcome these challenges are summarized in the following sections.

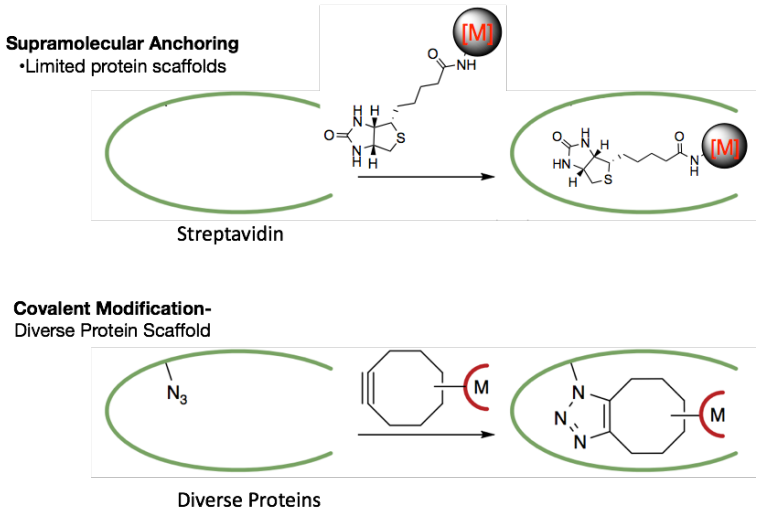


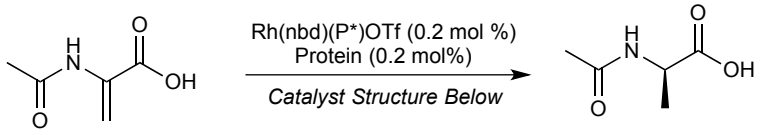
Figure 13. Graphical depiction of two strategies to create artificial metalloenzymes by anchoring pre-formed transition metal complexes within the pockets of proteins that do not typically contain metals.

1.3.1 Dative Anchoring Approach to Artificial Metalloenzymes

Whitesides and coworkers created the first artificial metalloenzymes by incorporating biotinylated rhodium complexes into streptavidin to create catalysts that hydrogenate activated alkenes (Fig. 14).[59] Although the reactions they reported occurred with modest enantioselectivity (~44%), this initial study nonetheless demonstrated the potential to create

active and selective enzymes by this approach. After a period of dormancy, this area of research resurged in the past twenty years during which time a wide variety of metal complexes have been introduced into a range of proteins.

Inspired by the early studies of Whitesides, Thomas Ward and coworkers contributed significantly to the development of artificial metalloenzymes based on streptavidin.[57,60] Their studies have aimed to create both artificial streptavidins that catalyze a range of reactions and to develop generalizable strategies that enhance the potential of streptavidin to function as an evolvable host protein. Recently, they have also expanded the dative anchoring strategy to include other proteins that have strong, specific dative interactions with particular molecules (such as carbonic anhydrase and aryl sulfonamides).[61] Together, contributions of Ward and others include streptavidins containing Rh, Ir, Pd, and Cu complexes that catalyze ester hydrolysis, dihydroxylation, epoxidation, sulfoxidation, hydrogenation, transfer hydrogenation, NAD+ reduction, and Diels Alder reactions.[13,62,63] Many of these reactions proceed with low to moderate selectivity and low activity when compared to either enzymes or synthetic metal complexes that function as catalysts for the same transformations.



entry	Protein	TON	ee
1	None	475	<2%
2	Lysozyme	450	<1%
3	BSA	150	<5%
4	Carbonic Anhydrase	50	<10%
5	Avidin (excess)	$>_{500}$	44%

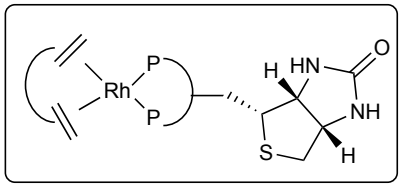
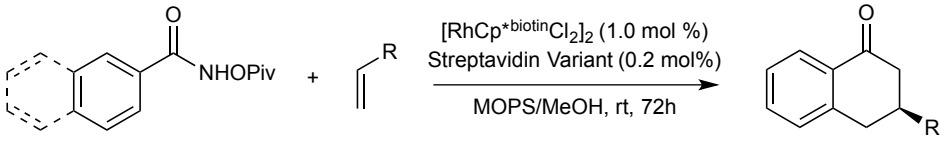


Figure 14. Initial work by Whitesides to show that incorporation of a metal complex within a protein by dative anchoring can create an enantioselective transition metal catalyst. In this example, an enantioselective catalyst for hydrogenation of alkenes was created.

In fewer cases, artificial streptavidins have been created to catalyze reactions for which natural enzymes are not known (abiological transformations). These examples, which are particularly notable due to the limited possibility to evolve a natural enzyme for these functions, include strepatividins for cross coupling,[64] metathesis,[65] allylic substitution,[66] and annulation reactions.[16] A landmark report from Thomas Ward described the enantioselective annulation of benzamides with acrylates using streptavidin containing a Rh(III) site as the catalyst (Fig. 15).[16]

This reaction occurred with up to 82% ee and 95% yield, with an observed rate enhancement of Rh-enzyme catalysts in comparison to the Rh-cofactor in the absence of the protein.



Substrate	Yield	ee
R = COEt	95%	12%
R = COOBn	61%	48%
R = COOMe	64%	76% ee
R = COOMe, Napthalene	80%	78% ee

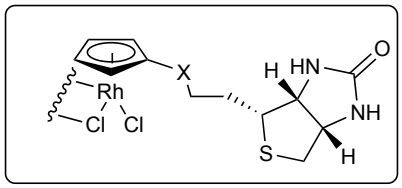
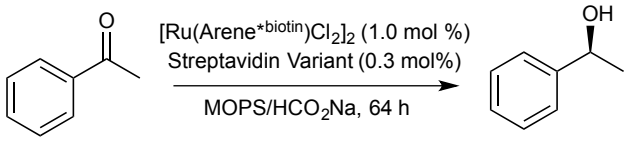


Figure 15. Enantioselective, Rh-catalyzed annulation reactions catalyzed within streptavidin.

Ward also reported enantioselective allylic alkylation reactions using streptavidin containing a Rh-phosphine site.[66] Though enantioselectivities of up to 90% were achieved in this case, the allylic acetate substrates were also consumed by hydrolysis, and the ratios between the alkylated and hydrolyzed products were not disclosed in the report. Other examples of abiological reactions catalyzed by artificial streptavidins proceeded with low to moderate enantioselectivities.[67]

To improve the approach to designing artificial metalloenzymes created by dative anchoring, Ward has investigated several strategies to optimize artificial metalloenzymes using the transfer hydrogenation of ketones and imines as model systems. In one approach, through what Ward coined "chemogenetic engineering,"[68] his group optimized the outcome of a catalytic reaction through combinatorial variation of the protein structure and metal cofactor (Fig. 16). In particular, they focused on creating mutations at two sites in the protein scaffold, and combined mutants of those site with metal complexes in which the linker between the biotin group and the metal center differed in length. These studies divulged the important interplay between the structure of the protein and the structure of the metal complex.

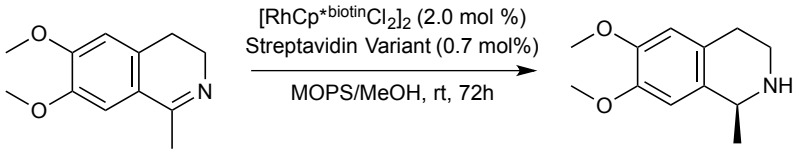


entry	Protein	Arene of Cofactor	Conv (%)	ee
1	S112Y Sav	*p*-cymene	95%	90% (*R*)
2	S112D Sav	*p*-cymene	0	---
3	S112K Sav	*p*-cymene	50%	20% (*S*)
4	S112K Sav	benzene	37%	55% (*S*)

Figure 16. Subset of results of chemogenetic engineering of Ru-streptavidins for transfer hydrogenation reactions. These results show the effect of both the structure of the protein and the structure of the cofactor on the outcome of the catalytic reaction.

Another advantageous strategy disclosed by Ward was to place a genetically encoded histidine residue at an appropriate position to coordinate directly to the metal, rigidifying the position of the metal site.[69] Reactions catalyzed by mutants bearing this histidine residue occurred with both higher activity and higher enantioselectivity compared to the same reactions catalyzed by proteins lacking this histidine residue (Fig. 17).

Computational design has been applied to the optimization of artificial metalloenzymes.[70,71] In collaboration with Ward, David Baker and coworkers used computational modeling to improve the affinity of carbonic anhydrase for metal complexes modified to contain sulfonamide affinity tags.[70] Their computational models identified mutants of carbonic anhydrase that bind Ir-piano stool complexes with 50-fold higher affinity compared to the wild type protein. Moreover, the mutants optimized by computation reacted with a four-fold improvement in TON and a 20% improvement in enantioselectivity compared to the wild type protein. The results obtained from these several strategies to design more effective artificial metalloenzymes suggest there is some potential to improve this class of enzymes as catalysts for a wider range of abiological transformations.



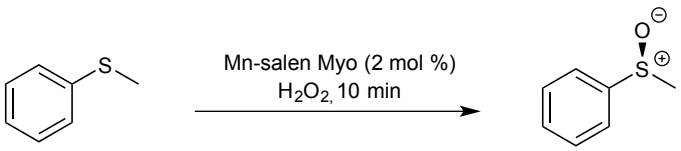
entry	Protein	Conv (%)	ee
1	WT	18%	racemic
2	S112H	100%	55% ee (*S*)
3	K121H	100%	79% (*R*)
4	S112H, K121H	40%	9% (*R*)

Figure 17. A dual-anchoring approach reported by Ward to improve the enantioselectivity of hydrogenation reactions catalyzed by Rh-streptavidins. In this case, a genetically encoded histidine residue, which can coordinate directed to the Rh-site, led to a catalyst with substantially improved enantioselectivity.

1.3.2 Covalent Anchoring Approach to Artificial Metalloenzymes

In addition to those created by dative anchoring, artificial metalloenzymes have been created by the covalent attachment of the metal complex to the protein. Most examples of enzymes created by this method react with low to moderate selectivities, likely because these enzymes tend to be compromised by poor rigidity of the metal complex within the protein scaffold.[13] To reduce the flexibility of this class of enzymes, the aforementioned strategy of dual anchoring has been used with some success. In one case, Yi Lu and coworkers showed that the enantioselectivity of sulfoxidation reactions catalyzed by myoglobins (Myo) containing manganese salens could be improved by bioconjugation of the salen complex to the protein at two genetically encoded cysteine residues, rather than attaching the salen at a single site (Fig. 18).[72] In this case, the rigidity of the metal center may also be enhanced though the coordination of the axial histidine residue of the myoglobin protein directly to the manganese center. Substitution reactions catalyzed by Mn-salen-Myo with a single anchoring point proceeded with up to 12% ee, while those catalyzed by a Mn-salen-Myo with two anchors proceeded with 55% ee. Subsequently, they showed that the enantioselectivity of the catalyst could be further tuned by changing the positions of the pair of anchoring groups.[73] Ultimately, the best catalyst they produced reacted with 61% ee and 16% yield.

In addition, a similar dual anchoring strategy to that initially described by Ward has also been applied to enzymes created by covalent anchoring. Recently, Jared Lewis and coworkers applied a bioconjugation strategy based on click chemistry developed in their group, to create oligiopeptidases containing Rh(II) sites that catalyze the cyclopropanation of styrenes with donor-acceptor carbenes (Fig. 19).[15] In this system, a genetically encoded histidine residue binds to the metal center, fixing the position of the metal at a particular site relative to the protein. By varying the site of this histidine, the authors were able to improve the enantioselectivity of the transformation, reaching selectivities of up to 92% ee without further evolution or optimization of the protein scaffold.[15]



entry	variant	rate (min^{-1})	ee
1	Y103C	0.05	10%
2	Y103C, S108C	0.09	-15%
3	Y103C, L72C	0.39	51%
4	T39C, S108C	0.18	1%
5	T39C, L72C	0.68	66%

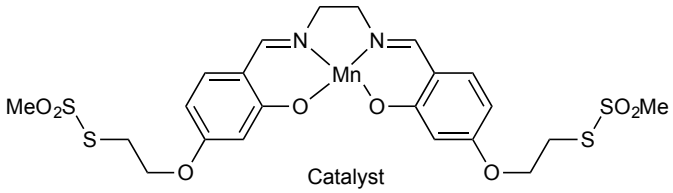
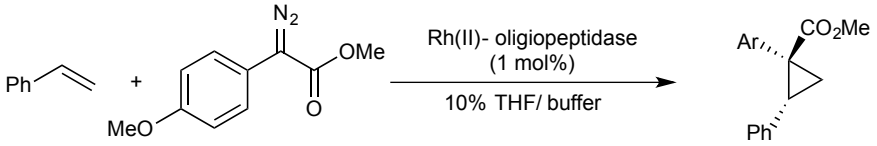


Figure 18. A sulfoxidation reaction catalyzed by myoglobin containing a salen ligand anchored to the porphyrin binding site of myoglobin at two sites.



entry	variant	Yield	ee
1	WT	25%	38%
2	L328F	14%	23%
3	L328C	21%	47%
4	L328H	61%	85%
5	L328H, F99, F594	74%	92%

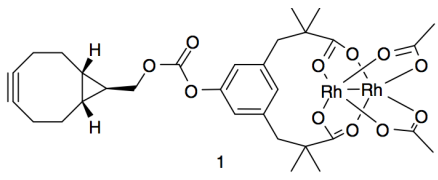


Figure 19. Cyclopropanation of styrenes with donor-acceptor carbenes catalyzed by a Rh(II) site anchored within a oliopeptidase by a click reaction at a genetically encoded azide residue

1.3.3 Outlook

To create highly active and selective artificial metalloenzymes by anchoring abiological metal complexes within proteins remains a developing area of research. While certain enantioselective catalysts for abiological reactions have been created, the levels of selectivity that have been achieved still fall short of the exquisite selectivities observed in a catalysis by natural enzymes. The cumulative studies of this class of artificial metalloenzymes have clearly shown the advantage conferred by creating additional interactions directly between the catalytic metal site and the protein. In the examples in which the highest enantioselectivity is achieved, the second anchor is an amino acid from the protein coordinating directly to the abiological metal site. This second anchor fixes and defines the site of the catalytic metal, allowing a discrete reaction environment to be achieved and mutagenesis to be directed in the vicinity of the metal site. The future will show whether this approach, likely followed in combination with computational modeling, will lead to examples of artificial metalloenzymes that react in abiological transformations with levels of selectivity more analogous to those of natural enzymes.

An additional, perhaps greater, challenge will be to evolve such artificial metalloenzymes to react with levels of activity resembling those of natural enzymes. While select examples of enantioselective artificial metalloenzymes exist, no artificial metalloenzymes display kinetic parameters resembling those of natural enzymes.[17] The absence of a natural substrate binding site adjacent to the metal in this class of artificial metalloenzymes compromises their potential to bind and react with substrates. Although the potential exists to evolve these types of substrate interactions into artificial metalloenzymes, multi-site directed evolution has rarely been applied to artificial metalloenzymes,[74,75] and the efficiencies of selective artificial metalloenzymes (as measured by K_M/k_{cat}) are more than 10,000X lower than the median efficiencies of natural enzymes involved in the synthesis of synthetic intermediate and secondary metabolites.[17]

Challenges that may have precluded the optimization of reported systems by directed evolution include the aforementioned amorphous structure of the active sites of these enzymes and the inefficiency of the methods used to prepare particular artificial metalloenzymes.[60] Therefore, to achieve an abiological reaction using an artificial metalloenzymes that is readily evolved to react with selectivities and activities comparable to natural enzymes would mark a genuine advancement in this field.

1.4 Synthesis and Applications of Artificial Metalloenzymes Created by Replacement of Native Metals with New Metals

Many of the challenges associated with creating highly selective and active artificial metalloenzymes by dative or covalent anchoring originate from the fundamental strategy by which they are created. Incorporation of unnatural metal complexes within cavities of proteins creates enzymes that lack a natural site adjacent to the metal to bind tightly a substrate. Without such a site, these artificial enzymes have diminished potential to bind and interact with substrates in a comparable manner to the interactions observed between natural enzymes and their native substrates,[11] and they may lack a suitable number of residues close to the metal site to optimize the active site by directed evolution.[31]

An alternative strategy to create artificial metalloenzymes that do contain natural sites to bind substrates adjacent to a metal site would be to reconstitute native metalloenzymes with abiological metals in place of the native metal. In this way, new reactivity can still be achieved due to the introduction of an abiological metal, but the natural affinity of the enzyme structure to bind a substrate can be retained. Consequently, such a system should have similar potential to that of the natural enzymes to be optimized by directed evolution.

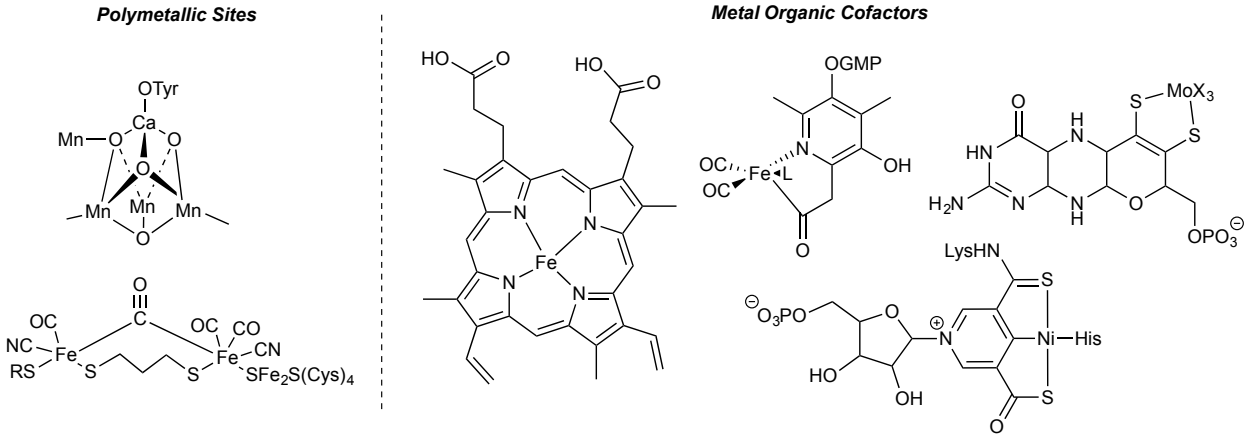
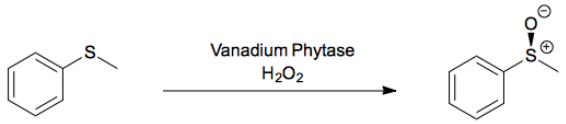


Figure 20. Examples of the structures of polymetallic sites and metal organic cofactors used in natural metalloenzymes.

Natural metalloproteins incorporate metals in one of three main modes: 1) as single metal sites coordinated to two or more amino acid side chains, 2) as polymetallic sites and metal clusters, or 3) within metal-organic cofactors (such as porphyrins) that are incorporated inside a protein (Fig. 20).[3] Of these three classes, the first and third, describing enzymes that contain single metal sites, should be most suitable to create selective artificial metalloenzymes because the presence of multiple metal sites will make it challenging to achieve a reaction in a single, discrete environment.

One of the earliest reports using this strategy converted a Zn-peptidase into Cu-sulfoxidase with low activity.[43] Subsequent work replaced zinc ions in thermolysin with W, Mo, and Se sites to create sulfoxidases that were slightly more active.[43] Finally, vanadium oxide sites were incorporated within a series of metalloproteins creating sulfoxidases that were also the first enantioselective artificial metalloenzymes created by metal reconstitution. Phytases containing

vanadium catalyzed the oxidation of arylsulfides to sulfoxides with up to 66% ee (Fig. 21).[19] These early studies did not use mutagenesis or directed evolution to seek more active or selective variants.



entry	mg Phytase	VO_4^{3-} (uM)	Temp (°C)	conversion	ee
1	0 mg	10	25	10%	0%
2	12 mg	10	25	100%	56%
3	12 mg	10	4	100%	66%
4	0 mg	0	25	10%	0%
5	1.7 mg	0	25	15%	25%

Figure 21. Enantioselective sulfoxidation reactions catalyzed by phytase substituted with a vanadium oxide site.

In most of these early examples, the native metals of metalloenzymes were replaced by alternative (typically first row metals) metals that are incorporated in other natural metalloenzymes. Reconstitution using first row metals is advantageous because first row metals form weaker bonds than second and third row metals, allowing the ions of these metals to equilibrate rapidly among many potential binding sites until the ion reaches the site that is most thermodynamically preferred.[1] Late second and third row transition metals, which typically form stronger bonds, are more likely to become kinetically trapped at sites other than the native metal binding site.[1] However, the creation of artificial metalloenzymes by the substitution of one first row metal for another has a limited potential to create functions that are not already possible with natural enzymes. For example, in the case of the aforementioned sulfoxidases, the activities of the artificial enzymes were 10^3-10^4 times lower than those of natural haloperoxidases that catalyze the same oxidation reactions.[18]

In fewer cases, the apo forms of natural metalloproteins have been reconstituted with metals that are not present in natural enzymes. Watanabe and coworkers have published a series of reports on the coordination of noble metals to ferritin, an Fe-transport protein cluster capable of binding dozens of Fe ions to each cluster.[76-79] In one case, apo-ferritin has been shown to bind approximately 58 Rh fragments and the resulting protein to catalyze size-selective polymerization of phenyl acetylene.[77] In these studies, X-ray crystallographic data support the incorporation of the noble metals at the same sites at which Fe coordinates in the native protein cluster. In another case, apo-carbonic anhydrase II (CA), a zinc-binding protein, was combined with rhodium precursors to generate Rh-CA species that catalyze hydrogenation and regioselective hydroformylation reactions (Fig. 22).[20,80] Compared to ferritin, apo-carbonic anhydrase is a simpler ligand because the natural protein binds a single metal center; however, the binding location and coordination environment of the Rh-center in the species generated by combining XX with apo-CA was not directed characterized in these studies. In fact, the complete, solution-state characterization of the metal center of a discrete, artificial organometallic protein had never been reported prior to the work described in Chapter 2 of this thesis.

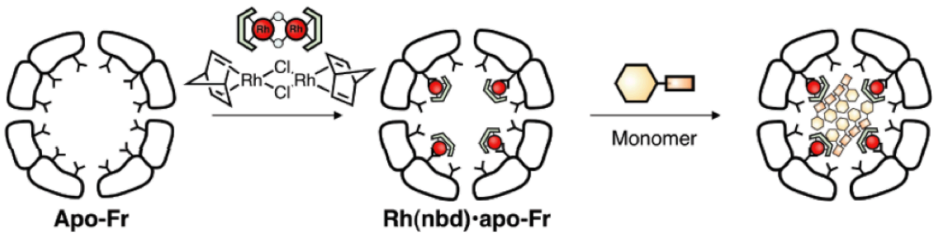


Figure 22. Reconstitution of apo-ferritin with Rh(nbd)-sites to create a catalyst for the polymerization of phenyl acetylene. Figure adapted from the original report of Watanabe.[80]

In a special case, one can circumvent many challenges associated with the reconstitution of apo-metalloproteins by choosing metalloproteins that bind metals as part of metal-organic cofactors. If only the metal of a biological metal-organic cofactor is changed, the host protein should retain a high affinity for the cofactor at its native binding site. This high affinity should guide the unnatural metal to the same site within the protein that is occupied by the natural metal. Although the number of known biological metal-organic cofactors is limited, those that are known, nonetheless, provide access to diverse and useful catalytic activity. For example, metal-porphyrins (M-PIX) are both biological cofactors and prevalent synthetic metal complexes.[81] Porphyrins containing abiological metals (Ru, Rh, Ir) catalyze a broader range of reactions than those containing Fe, including the formation of C-X bonds (via X-H insertion of carbenes or C-H insertion of nitrenes) and C-C bonds (via C-H insertion of carbenes), as well as cyclopropanes, epoxides, and aziridines (via addition of nitrenes and carbenes to C=X bonds), among other transformations that proceed through related mechanisms (Fig. 23).[82-85] Therefore, it one could incorporate porphyrins containing these abiological metals into heme proteins, enzymes with new functions could be created from heme proteins without disturbing the natural substrate binding sites present in those proteins.

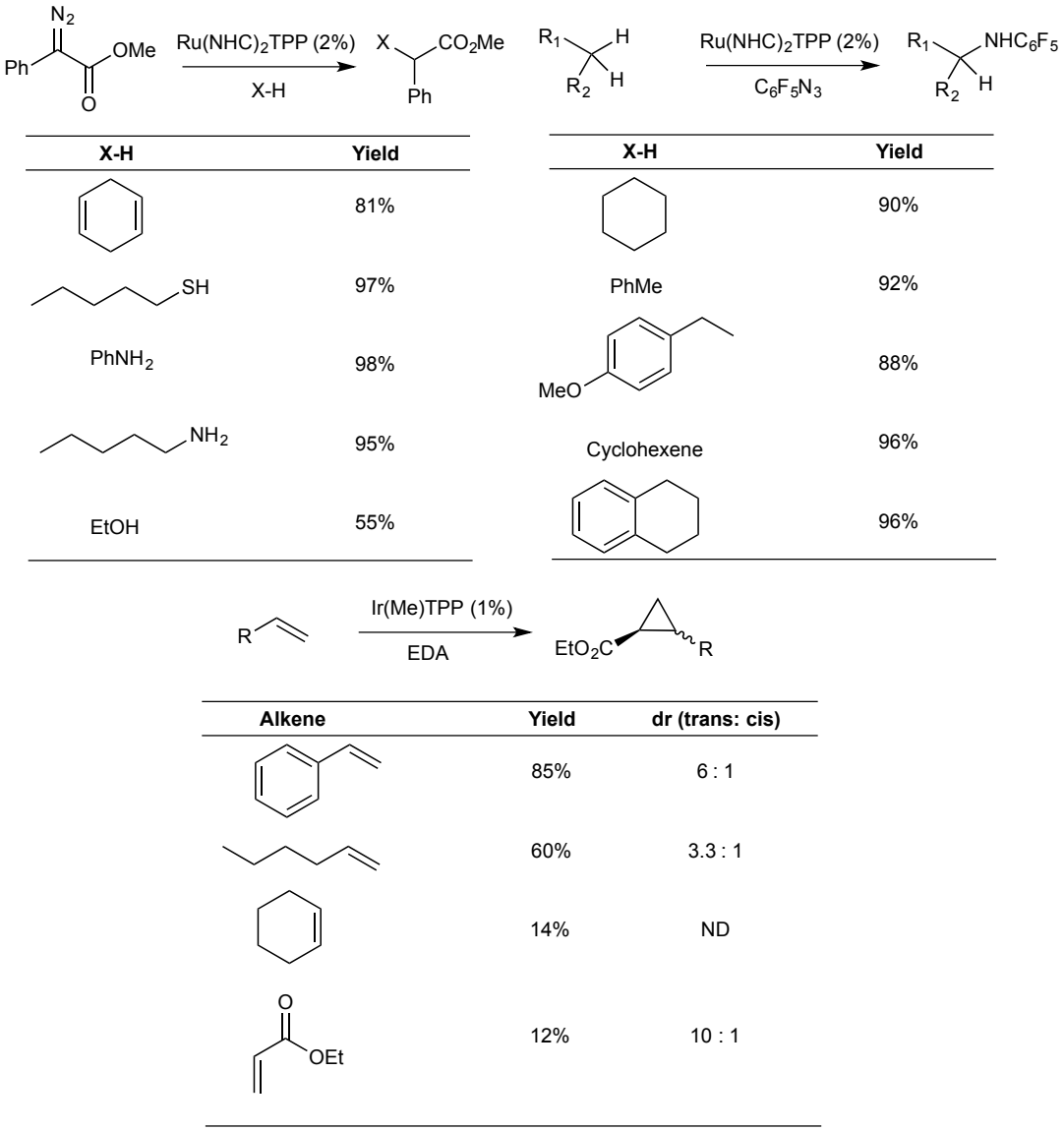
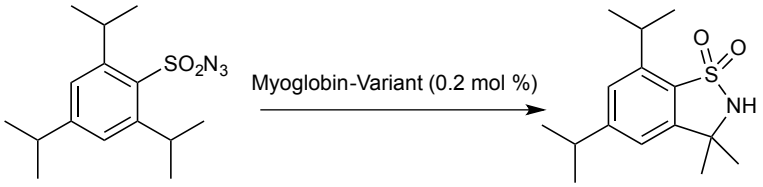


Figure 23. Examples of reactions catalyzed by noble metal porphyrins for which natural enzymes or their genetic variants are not known to be active catalysts.

Since the initial discovery of a method to generate the apo form of heme proteins from the Fe-form (by denaturing, acidic extraction of the heme cofactor),[86] numerous porphyrins containing metals other than Fe have been incorporated in heme-protein scaffold, including those of myoglobin, cytochrome P450-BM3, and cytochrome P450-CAM.[87-91] However, only a few recent studies have applied artificially metallated heme proteins as catalysts in synthetic organic chemistry. In these studies, the artificial heme proteins (containing Mn and Co) were only evaluated in reactions that can also be accomplished with native Fe-PIX proteins, and the variants containing Mn and Co in place of Fe were found to be less active than their Fe analogs (Fig. 24).[92,93] In addition, no enantioselectivity was observed in the transformations. Thus, artificially metallated heme proteins have never been used as catalysts for transformations beyond those catalyzed by Fe-PIX proteins.



entry	myo variant	metal	TTN
1	WT	Fe	175
2	H64V	Fe	197
3	V68A	Fe	185
4	N/A	Mn	145
5	N/A	Co	62

Figure 24. The single example from the literature to use a heme protein binding a porphyrin containing a metal other than Fe as a catalyst for an organic transformation. In this case, Mn and Co myoglobins were found to be less active catalysts for the insertion of nitrenes into C-H bonds than analogous (natural) Fe-containing myoglobins were found to be for the same reactions.

1.5 Challenges in Applying Reconstituted Metalloproteins in Catalysis

The synthesis of artificial metalloenzymes by formal reconstitution of the apo-forms of natural metalloenzymes is a promising, but rarely studied approach. The existing literature is sporadic; few systems have been studied in depth, improved upon over time, or grounded strongly in prior work by others. Despite the advantages of this class of artificial metalloenzymes previously discussed, the preparation and use of such enzymes confronts a series of challenges that may have impeded prior studies. Most notably, the thermodynamic driving force to bind the new metal selectively at same site as the native metal may be insufficient to preclude non-specific binding between the metal and other sites on the protein scaffold. Moreover, general, efficient methods to determine the site selectivity with which an abiological metal binds to an apo-metalloprotein have not been established. Finally, the fundamental reactivity of a metal complex directly ligating a protein may not be analogous to that of any small metal complex studied previously.

To overcome these challenges and to create the foundation for further development of such systems, it is essential to synthesize discrete organometallic proteins and to fully characterize the number and binding site of metal fragments coordinated to the protein and the number and identity of the amino acids bound to the metal. If well-characterized complexes can be generated, the elementary and catalytic reactivity of protein-ligated metal centers can then be observed, understood, and modulated to create the type of fundamental information that has fueled the development of small-molecule organometallic catalysts. The development of practical and efficient ways to synthesize these types of enzymes is also essential for efficient evolution of these enzymes; effective laboratory evolution requires the preparation of, at minimum, hundreds of variants of the wildtype enzyme.[5]

Additional challenges confronting catalysis using artificial metalloenzymes are those common to the broader field of biocatalysis. The range of useable reaction conditions is much more limited in biocatalysis than in standard transition metal catalysis. Reactions must be compatible with aqueous reaction conditions, and the reactions must occur at moderate temperatures and within a narrow pH range. Moreover, all reagents must be compatible with the wide range of reactive functionality (unprotected alcohols, amines, and phenols, sulfides, electron rich heterocycles, and others) present in the protein scaffold. As one example, the use of strong oxidants is challenging in the presence of proteins, due to both the consumption of the oxidant by reactions with the amino acid side chains and the precipitation of the protein after its structure reaches a heavily oxidized state. Beyond these considerable restrictions, substrates must be sufficiently soluble in water to interact with the enzyme, but when dissolved, they should not disrupt the the protein fold. Finally, reactions should occur under dilute reaction conditions and with >100 TON (which is more TON than are reported for many emerging transition metal catalysts), because proteins are generally unstable at high concentration or in the presence of high concentrations of organic substrates. If strong analogy to natural enzyme behavior is achieved, the hydrophobic effect can draw substrates present at low concentrations into the enzyme active site, at which point the structure of the active site can provide rate acceleration to enable challenging reactions under mild conditions.[11] However, initial systems are unlikely to display these favorable properties.

1.6 Conclusions and Objectives

The field of artificial metalloenzymes promises to combine the selectivity and evolutionary potential of enzymes with the diverse modes of reactivity possible using synthetic transition metal catalysts. Since its inception in the 1970s, research efforts in this area have increased, and the possibility to conduct abiological reactions at unnatural metal sites within enzymes has been demonstrated. However, presently these catalysts are used and behave with stronger analogy to synthetic transition metal catalysts than to enzymes. In comparison to natural enzymes, artificial metalloenzymes react with slow rates and modest selectivities.[13,17] Moreover, multi-site directed evolution has rarely been applied to artificial metalloenzymes due to challenges to create artificial metalloenzymes that are both sufficiently understood to apply directed evolution and that are sufficiently efficient to prepare.[75] Owing to these challenges, an artificial metalloenzymes has yet to be created to catalyze a reaction with a level of selectivity that cannot be accomplished with either a natural enzymes or a synthetic transition metal catalyst alone.

Based on the current literature in this field, we hypothesized that the alternative, rarely followed approach to create artificial metalloenzymes- by reconstitution of natural metalloenzymes with abiological metals- could create artificial enzymes with stronger structural analogy to natural enzymes. With stronger structural analogy should come reactivity profiles that are also more analogous to those of natural enzymes, including high activity and selectivity.

With limited examples of strategies to synthesize, characterize, and evolve artificial metalloenzymes created by metal replacement, our initial objectives were to develop suitable and generalizable methods to create, characterize, and evaluate the stoichiometric and catalytic activity of such artificial metalloenzymes. In particular, we aimed to prepare these enzymes by a strategy that is sufficiently sufficient for evolution.

In one case, using carbonic anhydrase as a model system,[94] we sought new strategies to introduce second- and third-row late transitions metals into proteins that bind metals directly

using amino acid side chains. Upon achieving a successful synthesis of carbonic anhydrases metallated with Rh- and Ir-sites, we sought to demonstrate that it is possible to characterize the structure of a metal that is ligated to a protein in the solution state in a fashion analogous to the way typical synthetic catalysts are characterized and understood. Finally, with suitable characterization methods in hand, we aimed to evaluate the fundamental, stoichiometric reactivity of the metal sites bound within the carbonic anhydrase matrix and to learn how this reactivity can be controlled through modification of the protein scaffold. These studies form a foundation from which to develop artificial metalloenzymes based on the same types of mechanistic data that have fueled the development of typical, synthetic metal catalysts.

In a second case, with myoglobin as a model system,[95] we aimed to develop an efficient process to form PIX-proteins containing metal porphyrins binding metals other than Fe. In particular, we targeted a process that was mild enough to preserve the structure of the protein and rapid enough to enable laboratory evolution of the resulting proteins. After developing a suitable process, we applied it to create a combinatorial array of artificial metalloenzymes combining different mutants of myoglobin with porphyrins containing different metals. This array was applied to search for artificial myoglobins with catalytic activity that is not known for any natural enzymes. Following the discovery of iridium-containing myoglobins that catalyze C-H insertion reactions, we applied laboratory evolution to develop catalysts for enantioselective variants of this and other transformations.

Subsequently, we applied the same methodology to create cytochrome P450s containing abiological metals. Because P450s naturally function as catalysts for organic transformations, we hypothesized that P450s containing abiological metal porphyrins could function as artificial metalloenzymes with high structural and functional analogy to natural enzymes. Elucidation of the kinetic parameters of these enzymes confirmed that their behavior strongly mirrored that of natural enzymes. This allowed us to exploit these enzymes in challenging transformations, such as intermolecular insertions of carbenes into a C-H bond and addition of carbenes to unactivated and internal alkenes that we could not accomplish using enzymes formed from myoglobin. Finally, we aimed to demonstrate the potential to control a range of selectivities using artificial metalloenzymes, such as diastereoselectivity, chemoselectivity, site-selectivity, and substrates selectivity, the latter three of which have not been shown previously with artificial metalloenzymes.

Despite challenges confronting preparation, characterization, and application, our studies have revealed that diverse artificial metalloenzymes created by the strategy of metal replacement can indeed be prepared efficiently and characterized as comprehensively as one would characterize natural enzymes and small molecule metal complexes. Moreover, the stoichiometric and catalytic reactivity of the noble metal sites within these proteins can be controlled by mutagenesis of the native substrate-binding site surrounding the abiological metal. With these studies as a foundation, we and others now can seek the creation of nearly limitless combinations of noble metals and protein scaffolds that can catalyze reactions with activities and selectivities not possible for either enzyme or transition metal catalysts alone.

1.7 References

1. Hartwig, J. F. *Organotransition Metal Chemistry*. (Univ Science Books, 2010).
2. Whitesides, G. M. Bioinspiration: something for everyone. *Interface Focus* **5,** 20150031–

20150031 (2015).
3. Andreini, C., Bertini, I., Cavallaro, G., Holliday, G. L. & Thornton, J. M. Metal ions in biological catalysis: from enzyme databases to general principles. *J Biol Inorg Chem* **13,** 1205–1218 (2008).
4. Reetz, M. T., Prasad, S., Carballeira, J. D., Gumulya, Y. & Bocola, M. Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods. *J. Am. Chem. Soc.* **132,** 9144–9152 (2010).
5. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16,** 379–394 (2015).
6. Roiban, G.-D. & Reetz, M. T. Expanding the toolbox of organic chemists: directed evolution of P450 monooxygenases as catalysts in regio- and stereoselective oxidative hydroxylation. *Chem. Commun.* **51,** 2208–2224 (2015).
7. Chen, M. M. Y., Snow, C. D., Vizcarra, C. L., Mayo, S. L. & Arnold, F. H. Comparison of random mutagenesis and semi-rational designed libraries for improved cytochrome P450 BM3-catalyzed hydroxylation of small alkanes. *Protein Eng. Des. Sel.* **25,** 171–178 (2012).
8. Peters, M. W., Meinhold, P., Glieder, A. & Arnold, F. H. Regio- and Enantioselective Alkane Hydroxylation with Engineered Cytochromes P450 BM-3. *J. Am. Chem. Soc.* **125,** 13442–13450 (2003).
9. Han, F.-S. Transition-metal-catalyzed Suzuki-Miyaura cross-coupling reactions: a remarkable advance from palladium to nickel catalysts. *Chem. Soc. Rev.* **42,** 5270–5298 (2013).
10. Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485,** 185–194 (2012).
11. Ringe, D. & Petsko, G. A. Biochemistry. How enzymes work. *Science* **320,** 1428–1429 (2008).
12. Bevers, L. E., Pinkse, M. W. H., Verhaert, P. D. E. M. & Hagen, W. R. Oleate hydratase catalyzes the hydration of a nonactivated carbon-carbon bond. *J. Bacteriol.* **191,** 5010–5012 (2009).
13. Lewis, J. C. Artificial Metalloenzymes and Metallopeptide Catalysts for Organic Synthesis. *ACS Catal.* **3,** 2954–2975 (2013).
14. Hyster, T. K. & Ward, T. R. Genetic Optimization of Metalloenzymes: Enhancing Enzymes for Non-Natural Reactions. *Angew. Chem. Int. Ed.* n/a–n/a (2016). doi:10.1002/anie.201508816
15. Srivastava, P., Yang, H., Ellis-Guardiola, K. & Lewis, J. C. Engineering a dirhodium artificial metalloenzyme for selective olefin cyclopropanation. *Nat Commun* **6,** 7789 (2015).
16. Hyster, T. K., Knörr, L., Ward, T. R. & Rovis, T. Biotinylated Rh(III) complexes in engineered streptavidin for accelerated asymmetric C-H activation. *Science* **338,** 500–503 (2012).
17. Bar-Even, A. *et al.* The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50,** 4402–4410 (2011).
18. Koeller, K. M. & Wong, C.-H. Enzymes for chemical synthesis. *Nature* **409,** 232–240 (2001).
19. van de Velde, F., Arends, I. W. C. E. & Sheldon, R. A. Vanadium-catalysed enantioselective sulfoxidations: rational design of biocatalytic and biomimetic systems.

Topics in Catalysis **13,** 259–265 (2000).
20. Jing, Q. & Kazlauskas, R. J. Regioselective Hydroformylation of Styrene Using Rhodium-Substituted Carbonic Anhydrase. *ChemCatChem* **2,** 953–957 (2010).
21. *The Ubiquitous Roles of Cytochrome P450 Proteins*. (John Wiley & Sons, Ltd, 2007). doi:10.1002/9780470028155
22. *Applied Homogeneous Catalysis with Organometallic Compounds*. (Wiley-VCH Verlag GmbH, 2002). doi:10.1002/9783527618231
23. Godula, K. & Sames, D. C-H bond functionalization in complex organic synthesis. *Science* **312,** 67–72 (2006).
24. Schulze, B. & Wubbolts, M. G. Biocatalysis for industrial production of fine chemicals. *Curr. Opin. Biotechnol.* **10,** 609–615 (1999).
25. Aleu, J., Bustillo, A., Hernandez-Galan, R. & Collado, I. Biocatalysis Applied to the Synthesis of Agrochemicals. *Current Organic Chemistry* **10,** 2037–2054 (2006).
26. Garcia, T., Martinez, M. & Aracil, J. Enzymatic Synthesis of Myristyl Myristate. Estimation of Parameters and Optimization Of the Process. *Biocatalysis and Biotransformation* **14,** 67–85 (2009).
27. Akin, C. Biocatalysis with immobilized cells. *Biotechnology and Genetic Engineering Reviews* (1987).
28. Kolkman, J. A.Willem P.C. Stemmer. Directed evolution of proteins by exon shuffling. *Nature Biotechnology* **19,** 423–428 (2001).
29. Romero, P. A. & Arnold, F. H. Exploring protein fitness landscapes by directed evolution. *Nature Reviews Molecular Cell Biology* **10,** 866–876 (2009).
30. Arnold, F. H. Combinatorial and computational challenges for biocatalyst design. *Nature* **409,** 253–257 (2001).
31. Sen, S., Dasu, V. V. & Mandal, B. Developments in Directed Evolution for Improving Enzyme Functions. *Appl Biochem Biotechnol* **143,** 212–223 (2007).
32. Lehmann, M. & Wyss, M. Engineering proteins for thermostability: the use of sequence alignments versus rational design and directed evolution. *Curr. Opin. Biotechnol.* **12,** 371–375 (2001).
33. Reetz, M. T., Soni, P., Fernández, L., Gumulya, Y. & Carballeira, J. D. Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method. *Chem. Comm.* **46,** 8657–8658 (2010).
34. Hyster, T. K. & Arnold, F. H. P450BM3-Axial Mutations: A Gateway to Non-Natural Reactivity. *Isr. J. Chem.* **55,** 14–20 (2015).
35. Richter, F., Leaver-Fay, A., Khare, S. D., Bjelic, S. & Baker, D. De Novo Enzyme Design Using Rosetta3. *PLOS ONE* **6,** e19230 (2011).
36. Sono, M., Roach, M. P., Coulter, E. D. & Dawson, J. H. Heme-containing oxygenases. *Chem. Rev.* **96,** 2841–2887 (1996).
37. Vaillancourt, F. H., Yeh, E., Vosburg, D. A., Garneau-Tsodikova, S. & Walsh, C. T. Nature's inventory of halogenation catalysts: oxidative strategies predominate. *Chem. Rev.* **106,** 3364–3378 (2006).
38. Le Bouquin, R. *et al.* CYP94A5, a new cytochrome P450 from Nicotiana tabacum is able to catalyze the oxidation of fatty acids to the ω-alcohol and to the corresponding diacid. *Eur. j. Biochem.* **268,** 3083–3090 (2001).
39. de Montellano, P. O. & Voss, J. Substrate oxidation by cytochrome P450 enzymes.

Cytochrome P450 (2005).
40. Relling, M. V. *et al.* O-demethylation of epipodophyllotoxins is catalyzed by human cytochrome P450 3A4. *Mol Pharmacol* **45,** 352–358 (1994).
41. Christelle Iribarne *et al.* Involvement of Cytochrome P450 3A4 Enzyme in the N-Demethylation of Methadone in Human Liver Microsomes. *Chem. Res. Toxicol.* **9,** 365–373 (1996).
42. de Souza, M. L., Sadowsky, M. J. & Wackett, L. P. Atrazine chlorohydrolase from Pseudomonas sp. strain ADP: gene sequence, enzyme purification, and protein characterization. *J. Bacteriol.* **178,** 4894–4900 (1996).
43. Bornscheuer, U. T. & Kazlauskas, R. J. Catalytic Promiscuity in Biocatalysis: Using Old Enzymes to Form New Bonds and Follow New Pathways. *Angew. Chem. Int. Ed.* **43,** 6032–6040 (2004).
44. Krishnamurthy, V. M. *et al.* Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein-ligand binding. *Chem. Rev.* **108,** 946–1051 (2008).
45. Thorslund, A. & Lindskog, S. Studies of the Esterase Activity and the Anion Inhibition of Bovine Zinc and Cobalt Carbonic Anhydrases. *Eur. j. Biochem.* **3,** 117–123 (1967).
46. Meunier, B., De Visser, S. P. & Shaik, S. Mechanism of oxidation reactions catalyzed by cytochrome P450 enzymes. *Chem. Rev.* (2004).
47. Coelho, P. S., Brustad, E. M., Kannan, A. & Arnold, F. H. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **339,** 307–310 (2013).
48. Coelho, P. S. *et al.* A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo. *Nat. Chem. Biol.* **9,** 485–487 (2013).
49. Wang, Z. J. *et al.* Improved Cyclopropanation Activity of Histidine-Ligated Cytochrome P450 Enables the Enantioselective Formal Synthesis of Levomilnacipran. *Angew. Chem. Int. Ed.* **53,** 6810–6813 (2014).
50. Sreenilayam, G. & Fasan, R. Myoglobin-catalyzed intermolecular carbene N-H insertion with arylamine substrates. *Chem. Commun.* **51,** 1532–1534 (2015).
51. Wang, Z. J., Peck, N. E., Renata, H. & Arnold, F. H. Cytochrome P450-Catalyzed Insertion of Carbenoids into N-H Bonds. *Chem Sci* **5,** 598–601 (2014).
52. Tyagi, V., Bonn, R. B. & Fasan, R. Intermolecular carbene S-H insertion catalysed by engineered myoglobin-based catalysts†. *Chem Sci* **6,** 2488–2494 (2015).
53. Hyster, T. K., Farwell, C. C., Buller, A. R., McIntosh, J. A. & Arnold, F. H. Enzyme-Controlled Nitrogen-Atom Transfer Enables Regiodivergent C–H Amination. *J. Am. Chem. Soc.* **136,** 15505–15508 (2014).
54. Singh, R., Bordeaux, M. & Fasan, R. P450-catalyzed intramolecular sp(3) C-H amination with arylsulfonyl azide substrates. *ACS Catal.* **4,** 546–552 (2014).
55. Farwell, C. C., Zhang, R. K., McIntosh, J. A., Hyster, T. K. & Arnold, F. H. Enantioselective Enzyme-Catalyzed Aziridination Enabled by Active-Site Evolution of a Cytochrome P450. *ACS Cent. Sci.* **1,** 89–93 (2015).
56. Farwell, C. C., McIntosh, J. A., Hyster, T. K., Wang, Z. J. & Arnold, F. H. Enantioselective Imidation of Sulfides via Enzyme-Catalyzed Intermolecular Nitrogen-Atom Transfer. *J. Am. Chem. Soc.* **136,** 8766–8771 (2014).
57. Ward, T. R. Artificial Metalloenzymes Based on the Biotin−Avidin Technology: Enantioselective Catalysis and Beyond. *Acc. Chem. Res.* **44,** 47–57 (2010).
58. Srivastava, P., Yang, H., Ellis-Guardiola, K. & Lewis, J. C. Engineering a dirhodium

artificial metalloenzyme for selective olefin cyclopropanation. *Nat Commun* **6,** 7789 (2015).
59. Wilson, M. E. & Whitesides, G. M. Conversion of a protein to a homogeneous asymmetric hydrogenation catalyst by site-specific modification with a diphosphinerhodium(I) moiety. *J. Am. Chem. Soc.* **100,** 306–307 (1978).
60. Letondor, C. *et al.* Artificial transfer hydrogenases based on the biotin-(strept)avidin technology: fine tuning the selectivity by saturation mutagenesis of the host protein. *J. Am. Chem. Soc.* **128,** 8320–8328 (2006).
61. Monnard, F. W., Heinisch, T., Nogueira, E. S., Schirmer, T. & Ward, T. R. Human Carbonic Anhydrase II as a host for piano-stool complexes bearing a sulfonamide anchor. *Chem. Comm.* **47,** 8238–8240 (2011).
62. Rosati, F. & Roelfes, G. Artificial Metalloenzymes. *ChemCatChem* **2,** 916–927 (2010).
63. Ringenberg, M. R. & Ward, T. R. Merging the best of two worlds: artificial metalloenzymes for enantioselective catalysis. *Chem. Commun.* **47,** 8470–8476 (2011).
64. Chatterjee, A. *et al.* An enantioselective artificial Suzukiase based on the biotin–streptavidin technology. *Chemical Science* **7,** 673–677 (2016).
65. Lo, C., Ringenberg, M. R., Gnandt, D., Wilson, Y. & Ward, T. R. Artificial metalloenzymes for olefin metathesis based on the biotin-(strept)avidin technology. *Chem. Comm.* **47,** 12065–12067 (2011).
66. Pierron, J. *et al.* Artificial Metalloenzymes for Asymmetric Allylic Alkylation on the Basis of the Biotin–Avidin Technology. *Angew. Chem. Int. Ed.* **120,** 713–717 (2008).
67. Jérôme Collot *et al. Artificial Metalloenzymes for Enantioselective Catalysis Based on Biotin−Avidin*. *Journal of the American Chemical Society* **125,** 9030–9031 (American Chemical Society, 2003).
68. Pordea, A. & Ward, T. R. Chemogenetic protein engineering: an efficient tool for the optimization of artificial metalloenzymes. *Chem. Commun.* 4239–4249 (2008). doi:10.1039/b806652c
69. Zimbron, J. M. *et al.* A Dual Anchoring Strategy for the Localization and Activation of Artificial Metalloenzymes Based on the Biotin–Streptavidin Technology. *J. Am. Chem. Soc.* **135,** 5384–5388 (2013).
70. Heinisch, T. *et al.* Improving the Catalytic Performance of an Artificial Metalloenzyme by Computational Design. *J. Am. Chem. Soc.* **137,** 10414–10419 (2015).
71. Ward, T. R. Artificial Enzymes Made to Order: Combination of Computational Design and Directed Evolution. *Angew. Chem. Int. Ed.* **47,** 7802–7803 (2008).
72. Carey, J. R. *et al.* A site-selective dual anchoring strategy for artificial metalloprotein design. *J. Am. Chem. Soc.* **126,** 10812–10813 (2004).
73. Garner, D. K., Liang, L., Barrios, D. A., Zhang, J.-L. & Lu, Y. Covalent Anchor Positions Play an Important Role in Tuning Catalytic Properties of a Rationally Designed MnSalen-containing Metalloenzyme. *ACS Catal.* **1,** 1083–1089 (2011).
74. Reetz, M. T., Peyralans, J. J.-P., Maichele, A., Fu, Y. & Maywald, M. Directed evolution of hybrid enzymes: Evolving enantioselectivity of an achiral Rh-complex anchored to a protein. *Chem. Commun.* 4318–4320 (2006). doi:10.1039/b610461d
75. Ilie, A. & Reetz, M. T. Directed Evolution of Artificial Metalloenzymes. *Isr. J. Chem.* **55,** 51–60 (2015).
76. Abe, S., Hikage, T., Watanabe, Y., Kitagawa, S. & Ueno, T. Mechanism of Accumulation and Incorporation of Organometallic Pd Complexes into the Protein Nanocage of apo-

Ferritin. *Inorg. Chem.* **49,** 6967–6973 (2010).
77. Abe, S. *et al.* Polymerization of Phenylacetylene by Rhodium Complexes within a Discrete Space of apo-Ferritin. *J. Am. Chem. Soc.* **131,** 6958–6960 (2009).
78. Abe, S. *et al.* Control of the Coordination Structure of Organometallic Palladium Complexes in an apo-Ferritin Cage. *J. Am. Chem. Soc.* **130,** 10512–10514 (2008).
79. Ueno, T. *et al.* Size-Selective Olefin Hydrogenation by a Pd Nanocluster Provided in an Apo-Ferritin Cage. *Angew. Chem. Int. Ed.* **116,** 2581–2584 (2004).
80. Jing, Q., Okrasa, K. & Kazlauskas, R. J. Stereoselective Hydrogenation of Olefins Using Rhodium-Substituted Carbonic Anhydrase—A New Reductase. *Chem. Eur. J.* **15,** 1370–1376 (2009).
81. Lu, H. & Zhang, X. P. Catalytic C–H functionalization by metalloporphyrins : recent developments and future directions. *Chem. Soc. Rev.* **40,** 1899–1909 (2011).
82. Doyle, M. P. & Forbes, D. C. Recent advances in asymmetric catalytic metal carbene transformations. *Chem. Rev.* (1998).
83. Chan, K. H., Guan, X., Lo, V. K. Y. & Che, C.-M. Elevated Catalytic Activity of Ruthenium(II)–Porphyrin-Catalyzed Carbene/Nitrene Transfer and Insertion Reactions with N-Heterocyclic Carbene Ligands. *Angew. Chem. Int. Ed.* **53,** 2982–2987 (2014).
84. Anding, B. J., Brgoch, J., Miller, G. J. & Woo, L. K. C–H Insertion Catalyzed by Tetratolylporphyrinato Methyliridium via a Metal–Carbene Intermediate. *Organometallics* **31,** 5586–5590 (2012).
85. Doyle, M. P. Catalytic methods for metal carbene transformations. *Chem. Rev.* **86,** 919–939 (1986).
86. Teale, F. W. Cleavage of the haem-protein link by acid methylethylketone. *Biochim Biophys Acta* **35,** 543 (1959).
87. Brucker, E. A., Olson, J. S., Phillips, G. N., Dou, Y. & Ikeda-Saito, M. High resolution crystal structures of the deoxy, oxy, and aquomet forms of cobalt myoglobin. *J. Biol. Chem.* **271,** 25419–25422 (1996).
88. Ikeda-Saito, M., Yamamoto, H., Imai, K., Kayne, F. J. & Yonetani, T. Studies on cobalt myoglobins and hemoglobins. Preparation of isolated chains containing cobaltous protoporphyrin IX and characterization of their equilibrium and kinetic properties of oxygenation and EPR spectra. *J. Biol. Chem.* **252,** 620–624 (1977).
89. Ikeda-Saito, M., Hori, H., Inubushi, T. & Yonetani, T. Studies on cobalt myoglobins and hemoglobins. The interaction of molecular oxygen with leghemoglobin. *J. Biol. Chem.* **256,** 10267–10271 (1981).
90. Li, C.-Z., Taniguchi, I. & Mulchandani, A. Redox properties of engineered ruthenium myoglobin. *Bioelectrochemistry* **75,** 182–188 (2009).
91. Hoffman, B. M. *et al.* Manganese-Substituted Hemoglobin AND Myoglobin. *Annals of the New York Academy of Sciences* **244,** 174–186 (1975).
92. Oohora, K., Kihira, Y., Mizohata, E., Inoue, T. & Hayashi, T. C(sp3)–H Bond Hydroxylation Catalyzed by Myoglobin Reconstituted with Manganese Porphycene. *J. Am. Chem. Soc.* **135,** 17282–17285 (2013).
93. Bordeaux, M., Singh, R. & Fasan, R. Intramolecular C(sp(3))H amination of arylsulfonyl azides with engineered and artificial myoglobin-based catalysts. *Bioorg. Med. Chem.* **22,** 5697–5704 (2014).
94. Key, H. M., Clark, D. S. & Hartwig, J. F. Generation, Characterization, and Tunable Reactivity of Organometallic Fragments Bound to a Protein Ligand. *J. Am. Chem. Soc.*

137, 8261–8268 (2015).
95. Key, H. M., Dydio, P., Clark, D. S. & Hartwig, J. F. Abiological catalysis by artificial haem proteins containing noble metals in place of iron. *Nature* **534,** 534–537 (2016).

Chapter 2: Generation, Characterization, and Tunable Reactivity of Organometallic Fragments Bound to a Protein Ligand

This chapter is modified from the following publication with permission of the university and all coauthors: Key, H.M. ; Clark, D. S. ; Hartwig, J.F. *JACS* **2015,** *137***,** 8261-8268.

ABSTRACT: Organotransition metal complexes catalyze important synthetic transformations, and the development of these systems has rested on the detailed understanding of the structures and elementary reactions of discrete organometallic complexes bound to organic ligands. An alternative strategy for the creation of new organometallic systems is to exploit the intricate and highly structured ligands in natural metalloproteins. To understand the relationships between the structure and reactivity of a protein-ligated organometallic site and those of a conventional organometallic complex, discrete, well-characterized transition metal complexes containing protein ligands need to be generated and their reactivity observed directly. We report the preparation and characterization of discrete rhodium and iridium fragments bound site-specifically in a κ_2-fashion to carbonic anhydrase as a ligand. With methods devised to determine the extent, specificity, and mode of bonding of the organometallic fragment to the protein, we have observed stoichiometric reactions occurring at an artificial organometallic protein. The preparation and reactivity of these complexes demonstrate the potential to study a new genre of organometallic complexes for which the rates and outcomes of organometallic reactions can be controlled by genetic manipulations of the protein scaffold.

2.1 Introduction.

Transition metal complexes catalyze many important chemical reactions used to prepare molecules ranging from small commodity chemicals to complex, medicinally important natural products.[1] Fundamental studies on the synthesis and reactivity of organometallic complexes have created the framework on which new catalytic processes are developed. These studies have shown that the ligands in these catalysts modulate reactivity and selectivity via the electronic properties of the donor atom(s) and the particular steric environment that the ligand creates around the metal. Consequently, the expanding utility of this class of catalyst is inseparable from the development and iterative modification of organic ligands. While this strategy has produced innumerable active and selective metal complexes for stoichiometric and catalytic transformations of interest, the synthesis of complex ligands is inefficient, and active and selective complexes undergoing desirable transformations and reacting with substrates of interest have yet to be identified.[1]

An alternative class of organometallic system to those containing ligands based on small molecules could exploit the intricate and highly structured structures of natural metalloproteins.[2] Unlike many organic ligands, diverse proteins that bind metals can be prepared conveniently in a single step via recombinant expression, consistently purified in a single step by affinity

chromatography, and rapidly derivatized by a variety of well-established mutagenesis techniques and molecular evolution.[3-5] Due to these properties, protein-derived ligands could enable a strategy to control organometallic and catalytic reactivity that complements that based on the more typical selection of small organic ligands, including the efficient evolution of ligands providing targeted activities and selectivities on a substrate specific basis.

Organometallic complexes containing abiotic metals directly coordinated to proteins are a limited part of the growing field of artificial metalloenzymes.[6-8] More frequently, artificial metalloenzymes have been prepared by a strategy in which a known transition metal complex is anchored within a protein scaffold in a covalent or supramolecular fashion.[9-11] This strategy is distinguished by the fact that complete complexes of known catalytic activity are incorporated into the system; however, in this case the native binding site of the protein must be sufficiently large to accommodate an entire organometallic complex, and although the binding or attachment site of anchoring group may be well-defined, the catalytic metal center does not necessarily fall within the native binding pocket of the protein host. Thus, the application of apo-metalloproteins as ancillary ligands for organometallic complexes is a complementary tactic in which the new metal is positioned directly in the native binding pocket of the protein. In this way, the secondary coordination sphere surrounding the new metal can be reasonably predicted provided that the structure of the natural metalloenzymes is known.

Although the number of examples in which metal complexes ligate a protein as the primary ancillary ligand and the resulting complex serves as a catalyst are limited, these reports nevertheless validate that protein-derived ligands have the potential to impose selectivity in catalytic reactions.[12-16] In one case, apo-ferritin has been shown to bind approximately 58 Rh fragments and the resulting protein to catalyze size-selective polymerization of phenyl acetylene.[14] In another case, apo-carbonic anhydrase II was reconstituted with manganese to create an enantioselective artificial epoxidase.[17] The same approach was then used to create carbonic anhydrase analogs containing rhodium in place of zinc. The resulting protein is reported to catalyze hydrogenation and regioselective hydroformylation reactions.[18,19] Compared to ferritin, apo-carbonic anhydrase is a simpler ligand because the natural protein binds a single metal center; however, the binding location and coordination environment of the Rh-center in the carbonic anhydrase analog was not directed characterized in these studies.[18,19] In fact, the complete, solution-state characterization of the metal center of a discrete, artificial organometallic protein was not reported prior to our publication of the research described here.

To create the foundation for further development of organometallic proteins formed from the direct ligation of abiological metals to the metal binding sites of apo-metalloproteins, it is essential to synthesize discrete organometallic proteins and to fully characterize the number of metal fragments bound to the protein, the site to which the protein is bound, and the number and identity of the amino acids bound to the metal. If well-characterized complexes can be generated, the elementary reactivity of protein-ligated metal centers can then be observed, understood, and modulated to create the type of fundamental information that has fueled the development of small-molecule organometallic catalysts. The synthesis and characterization of such an organometallic protein confronts a number of challenges not encountered in traditional organometallic synthesis. Preparing such complexes in high yield, determining which amino acid residues bind the metal, and monitoring the reactions of small molecules at the site created within a large protein requires the judicious combination of methods used to characterize both organometallic complexes and proteins. Indeed, stoichiometric reactions of an abiological organometallic center bound to a protein ligand have not been observed directly.

My PhD studies revealed the efficient synthesis and characterization of a series of discrete, well-defined organometallic complexes that ligate carbonic anhydrase through two of the three histidines of the native Zn-binding site. Generalizable methods to determine the extent and location of metal binding in a general fashion were developed, and, for the first time, the conversion of one organometallic protein to a second organometallic protein by the simultaneous cleavage and formation of metal-carbon bonds was observed directly using these methods. Furthermore, we show that the structure of the protein surrounding the organometallic center controls the rates and products of these reactions in a manner that would be challenging to achieve using small-molecule ligands.

2.2 Synthesis of Discrete Organometallic Proteins

To generate protein-derived ancillary ligands suitable for coordinating organometallic fragments in a defined manner, we followed a documented approach of removing native metals from natural metalloproteins also followed by others (Fig. 1).[19] The ligand substitution reaction of the resulting apo protein with an organometallic fragment presents a series of challenges that are not encountered when conducting analogous preparations of small organometallic complexes. In particular, the separation of apo and metal-bound proteins and the separation of proteins in which the metal is bound to different sites is impractical. Therefore, the generation of discrete artificial organometallic proteins requires high-yielding and site-specific metallations. To develop such a reaction, efficient methods are needed to identify the metal precursor, reaction conditions, and protein mutants that form single complexes in high yield. X-ray crystallography, non-native mass spectrometry, and metal analysis are the predominant methods used to characterize artificial metalloenzymes,[9,14] but these methods, respectively, are too slow to evaluate the yields for formation of potential organometallic proteins and reactions of them, release the metal due to denaturation, and are silent concerning the binding site. Furthermore, x-ray crystallography cannot be used to characterize metalloprotein species with that are unstable over time, such as the Rh-proteins described in this study. Therefore, we sought alternative, solution state methods that efficiently reveal the site-specificity of metal binding.

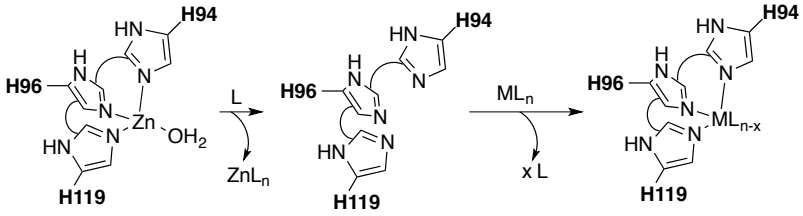


Figure 1. Standard method used to remove the native Zn fragment from CA.

Initially, we explored methods to generate and characterize protein-ligated transition metal complexes using a mutant of human carbonic anhydrase II. CA is an enzyme that binds zinc through three histidines,[20] and these strongly coordinating imidazole units could mimic classic polydentate ligands of organometallic complexes once Zn is removed to generate the apo-CA.[21] The mutant we studied (3*His-CA) lacks three surface histidine residues (Fig. 1, Table 1) that could interfere with binding of the organometallic fragment to the original Zn binding site.[18,19]

Mutant	Sequence (Mutations from WT sequence are highlighted)
3*His	SNAMARHWGYGKRNGPEFWHKDFPIAKGERQSPVDIDTHTAKYDPSLKPLSVSYDQATSLRILNNGH AFNVEFDDSQDKAVLKGGPLDGTYRLIQFHFHWGSLDGQGSEHTVDKKKYAAELHLVHWNTKYGDFG KAVQQPDGLAVLGIFLKVGSAKPGLQKVVDVLDSIKTKGKSADFTNFDPRGLLPESLDYWTYPGSLT TPPLLECVTWIVLKEPISVSSEQVLKFRKLNFNGEGEPEELMVDNWRPAQPLKNRQIKASFK
6*His	SNAMSRRWGYGKRNGPEFWFKDFPIAKGERQSPVDIDTRTAKYDPSLKPLSVSYDQATSLRILNNGH AFNVEFDDSQDKAVLKGGPLDGTYRLIQFHFHWGSLDGQGSEHTVDKKKYAAELHLVHWNTKYGDFG KAVQQPDGLAVLGIFLKVGSAKPGLQKVVDVLDSIKTKGKSADFTNFDPRGLLPESLDYWTYPGSLT TPPLLECVTWIVLKEPISVSSEQVLKFRKLNFNGEGEPEELMVDNWRPAQPLKNRQIKASFK
7*His	SNAMSRRWGYGKRNGPEFWFKDFPIAKGERQSPVDIDTRTAKYDPSLKPLSVSYDQATSLRILNNGA AFNVEFDDSQDKAVLKGGPLDGTYRLIQFHFHWGSLDGQGSEHTVDKKKYAAELHLVHWNTKYGDFG KAVQQPDGLAVLGIFLKVGSAKPGLQKVVDVLDSIKTKGKSADFTNFDPRGLLPESLDYWTYPGSLT TPPLLECVTWIVLKEPISVSSEQVLKFRKLNFNGEGEPEELMVDNWRPAQPLKNRQIKASFK

Table 1. Amino acid sequences of surface histidine mutants of carbonic anhydrase used in this study. Initial pre-methionine amino acids (red) are residual from TEV cleavage of N-terminal 6xHis Tag.

Recombinant carbonic anhydrase and its associated mutants were produced in E coli using rich, auto-induction media, and purified using Ni-NTA affinity chromatography. Reported methods to express carbonic anhydrase using IPTG induction and to purify the resulting protein using DEAE sephacel anion exchange chromatography resulted in lower expression yields and lower protein purity. The purified Zn-CA protein was characterized by SDS-PAGE and denaturing LC-MS (Figure 2). Following purification, the polyhistidine tag was cleaved using a genetically encoded TEV cleavage site. Following the affinity tag cleavage, Apo-CA was generated by exposing Zn-CA to dipicolinic acid, followed by extensive dialysis to separate the chelator from the apo-protein. A standard *p*-nitro phenylacetate assay was used to confirm the absence of coordinated zinc in the CA active site.

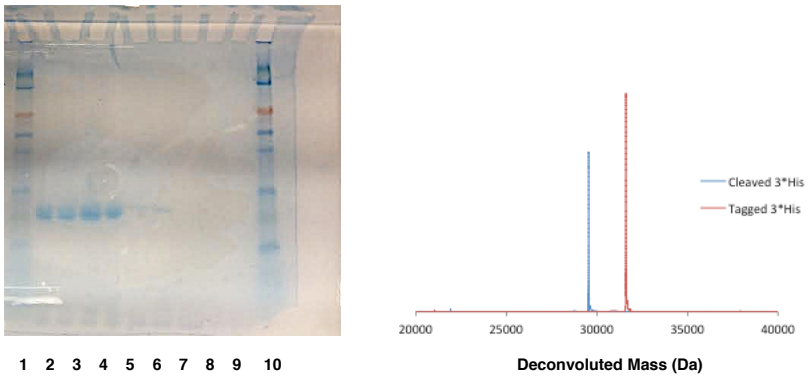


Figure 2. Protein purity achieved by Ni-NTA chromatography (left, SDS-PAGE gel) and conversion in TEV cleavage reaction (right, LC-TOF-MS) for 3*His-CA. Lanes 1, 10: MW ladder. Lanes 2-7: 3*His-CA (Ni-NTA Elution).

Although the combination of Rh-precursors and apo-carbonic anhydrases has previously reported,[18,19] the metallation protocol that was employed is not suitable for efficient evaluation of

a wide range of metal precursors. Previously, apo-carbonic anhydrase was dialyzed against a buffer solution that contained orders of magnitude excess rhodium compound. The elimination of excess Rh-metal from the buffer and the de-coordination of Rh bound outside the Zn-site were achieved by subsequent dialysis against Rh-free buffer. Beyond the impracticality of employing such a large excess of precious metal complexes, this method is limited to metal precursors with high water solubility and to those for which for site-selective de-coordination of ligated metal fragments is possible.

To create an alternative, more practical and general method to form organometallic proteins rapidly, we evaluated direct metalations involving the addition of nearly stoichiometric amounts of an organic stock solution of the appropriate rhodium precursor to a buffered solution of Apo-CA. These conditions should both serve as a practical strategy for the synthesis of a range of artificial organometallic proteins and reduce the potential to coordinate more than one metal fragment to each protein. Nonetheless, the success of this strategy was predicated on the possibility to determine easily identify the site at which the metal fragment coordinates to the protein.

To determine the selectivity of the noble metal fragments for coordination to the native Zn-binding site of CA, rather than the protein surface, we exploited the known hydrolysis of 4-nitrophenyl acetate catalyzed by cobalt-CA to produce 4-nitrophenol (λ_{max} = 405 nm) (Fig. 3).[22,23] Following the addition of a noble metal complex to the apo form of 3*His-CA (apo-3*His-CA), $CoCl_2$ was added to convert the remaining Apo-CA to Co-CA. The amount of protein with noble metal coordinated at the Zn-site would then correspond to the difference between the amount of apo-CA at the start of the reaction and the amount of Co-CA determined by the hydrolysis assay. This assay can be conducted with only 10 ug of protein in a 96-well format in five minutes, allowing the rapid assessment of different metal precursors, reaction stoichiometries, organic so-solvents, buffers, and other parameters. This method for assessing metal binding should be applicable to many other metalloprotein scaffolds for which activity or metal binding can be evaluated with a colorimetric assay.[24-27]

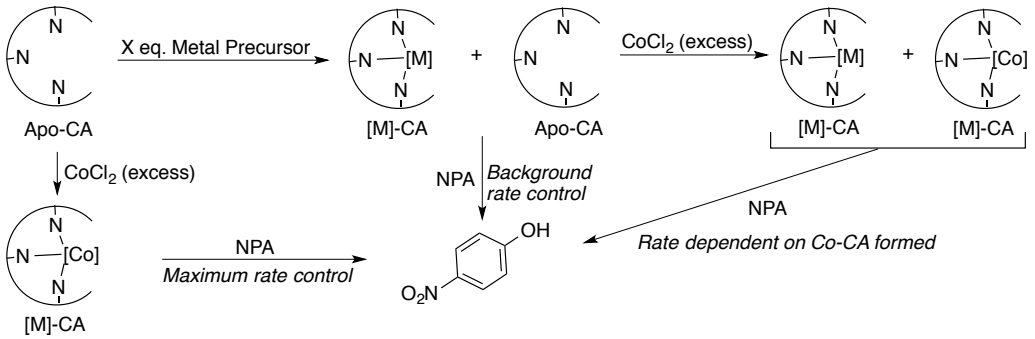


Figure 3. Activity assay to assess the selectivity of metal-substitution at the CA active site. The reaction rate for the hydrolysis of 4-nitro-phenylacetate (NPA), catalyzed by Co-CA, was used to compare the amount of Apo-CA (reaction starting material) before and after the reaction. The conversion (%) of the starting material was used to estimate the extent of metallation.

With this assay, we evaluated the binding of a series of 16-electron Rh(I) precursors for which small-molecule organometallic chemistry has been established and ligand substitution is

often facile.[28] The addition of 1 equivalent of $[Rh(cod)_2]BF_4$ (cod=cyclooctadiene) to apo-3*His-CA led to 45% Rh-occupancy of the native Zn-binding site (Fig. 4a, S1, S2).

Inductively coupled plasma optical emission spectroscopy (ICP-OES) revealed that an average of 0.94 Rh were coordinated per CA; thus, about half of this rhodium complex binds to the Zn-site and half binds to other sites on the protein. The addition of super-stoichiometric $[Rh(cod)_2]BF_4$ led to higher Rh-occupancy of the native zinc binding site, but resulted in concomitant binding of the excess Rh elsewhere on the protein (Fig 4a, S1-S2).[29] Together, these methods revealed that $[Rh(COD)_2]BF_4$ undergoes ligand substitution with the Apo-CA ligand, but that this reaction does not proceed with sufficient site specificity to form a discrete organometallic protein. Therefore, we sought alternative conditions to achieve more selective ligand substitution reactions that form Rh(bis-olefin)-CA complexes in high yield.

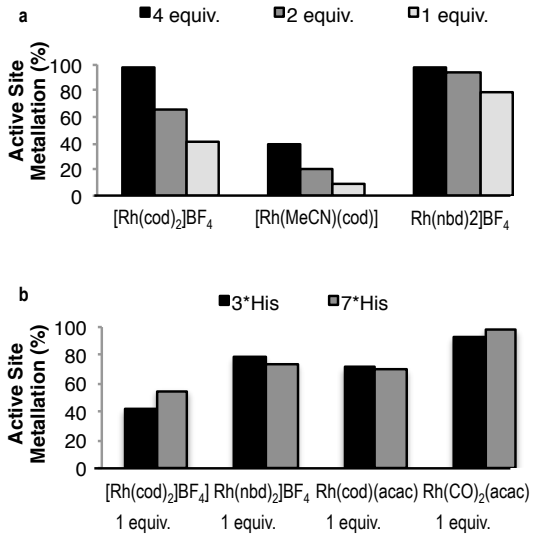


Figure 4. Preparation of organometallic proteins from carbonic anhydrase a. Comparison of the specificity of active site metallation of Rh(I) precursors upon the addition of 1-4 equiv of Rh to Apo-3*His-CA (67 uM) in 50 mM MES buffer, pH = 5.3. Metallation of the active site was determined using the activity assay described in the text (Fig. 3). b. Comparison of the selectivity of metallation for the addition of 1 equiv. cationic and neutral rhodium source

Given this initial result, we analyzed the effect of the buffer and the surface amino acid composition had a minor effect on the selectivity of this reaction, the ligands on rhodium did affect significantly the ligand substitution reaction that binds the metal precursor to the Zn-site of the CA (Fig. 4). The reaction of $[Rh(cod)(MeCN)_2]BF_4$ with apo-3*His-CA resulted in low occupancy of the Zn-site, likely because this cationic precursor with a labile nitrile binds predominantly to the protein surface. However, the reaction of 1 equivalent of $[Rh(nbd)_2]BF_4$ with the same protein resulted in 79% occupation of the Zn-site. In contrast to the low selectivity of metalation with the previously reported $[Rh(cod)_2]BF_4$ precursor, ligand substitution at the Zn-site with Rh(cod)(acac) occurs with high selectivity. Metallation of apo-3*His-CA with 1 equiv Rh(cod)(acac) (acac = acetylacetonate) 71% yield of CA containing Rh at the native Zn-site. Given this promising result using a neutral Rh complex with dative ligands that do not readily dissociate, we employed $Rh(CO)_2(acac)$ as a Rh-precursor and achieved >90% yield of protein containing rhodium in the original Zn-site (Fig. 2). ICP-OES revealed that addition of 1 equiv of either $[Rh(nbd)_2]BF_4$ or $Rh(CO)_2(acac)$ to Apo-CA resulted in the formation of a Rh-CA containing a Rh : protein ratio of approximately 1:1 (Fig. 5). Thus, a vast majority of these two rhodium complexes bind to the original Zn site, with little bound elsewhere on the protein.

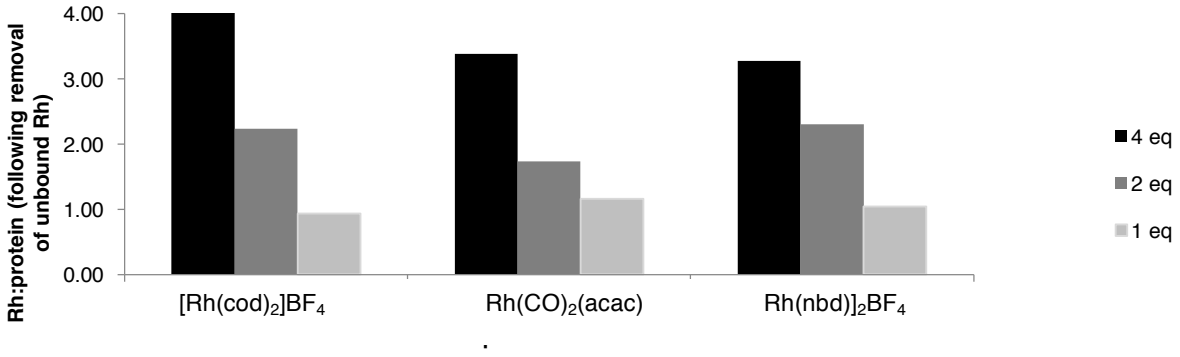


Figure 5. Determination of the Rh : protein ratio of Rh-CA samples by ICP-OES following the removal of unbound Rh. Bars are the average of three separate experiments. In all cases, the standard deviation of the measured [Rh] was less than 15% of the average value.

Following this comparison of the selectivity of Rh-precursors for coordination to the Zn-site, the resulting Rh-CA complexes were compared by UV-vis spectroscopy. The UV-vis spectrum of the Rh-CA prepared from $[Rh(nbd)_2]BF_4$ contained an absorption band at 410 nm, which is much different from the absorption of free $[Rh(nbd)_2]BF_4$ and distinguishable from the absorption of the Rh-CA generated from $[Rh(cod)_2]BF_4$ (Fig. 6).

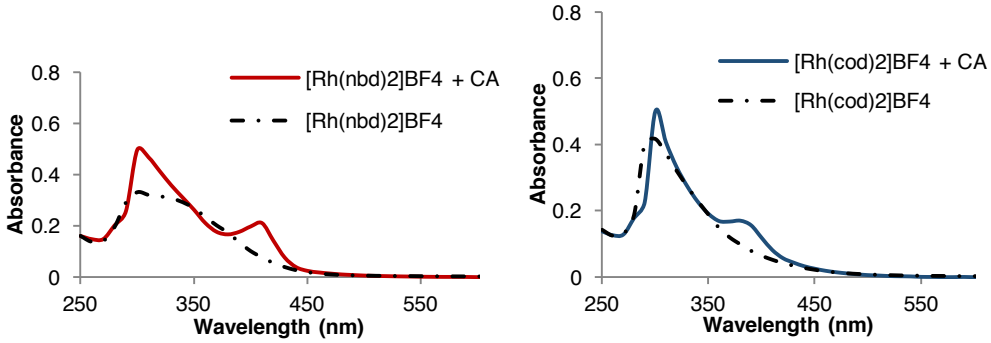


Figure 6. UV-vis spectra of Rh-substituted CA's prepared from the addition of Rh-precursors to Apo-3*His-CA. For comparison, the UV-vis spectra of the free metal complexes are also shown. A. Rh(nbd)-CA from $[Rh(nbd)_2]BF_4$. B. Rh(cod)-CA from $[Rh(nbd)_2]BF_4$.

The inequivalent spectra for the proteins generated from $[Rh(nbd)_2]BF_4$ and $[Rh(cod)_2]BF_4$ are likely due to slightly different diene ligands at the Rh-center. The UV-vis spectrum of the Rh-CA protein generated from Rh(cod)(acac) matched that of the Rh-CA generated from $[Rh(cod)_2]BF_4$, consistent with generation of the same Rh(COD) fragment at the Zn-site with these two precursors (Fig. 7). The UV-vis spectrum of the Rh-CA resulting from addition of $Rh(CO)_2(acac)$ is distinct from that of the other complexes (Fig. 7).

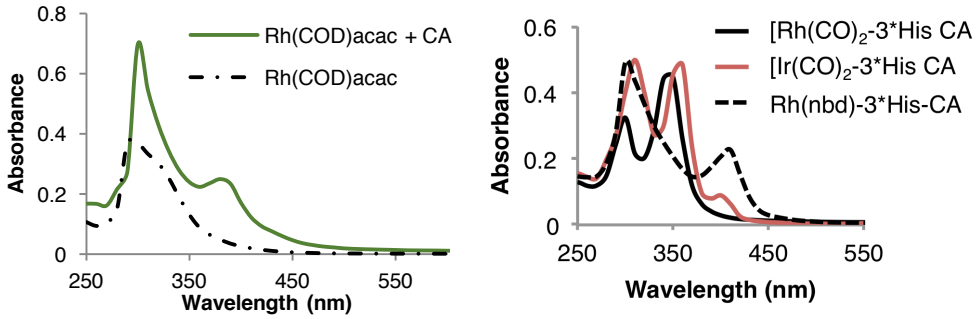


Figure 7. UV-vis spectra of Rh-substituted CA's prepared from the addition of Rh-precursors to Apo-3*His-CA. For comparison, the UV-vis spectra of the free metal complexes are also shown. Left: Rh(cod)-CA from [Rh(cod)acac]. Right: Comparison of UV spectra of Rh- and Ir-CA complexes that were produced with the highest site selectivity.

Together, these operationally simple, and broadly applicable methods to conduct and characterize ligand substitution reactions between apo-metalloproteins and noble metal complexes enabled us to identify Rh-precursors that react with apo-CA site specifically in a high yield.

2.3 Characterization of the Protein-Metal Coordination Sphere.

To characterize the coordination sphere of these protein-ligated organometallic complexes fully and to enable monitoring of the reactivity of these complexes with organic substrates, we sought additional, direct, solution-state characterization of the protein-ligated metal center. Because the metal is likely to be bound by histidines, we used ^{15}N NMR spectroscopy to evaluate the identity and number of these amino acids bound to rhodium, much like one uses ^{31}P NMR spectroscopy to characterize phosphine-ligated metal complexes. Prior work has shown that ^{15}N-^{1}H HSQC spectroscopy can distinguish Zn-CA from Apo-CA.[30]

A combination of [^{1}H]-[^{15}N]-HSQC NMR experiments revealed the coordination site of the rhodium bound to the CA mutants. 1-Bond HSQC spectra of ^{15}N labeled CA required short experiment times with only 0.2 mM protein. However, this method detects only nitrogen atoms for which N-H proton exchanges are slower than the NMR time scale, and it does not detect the histidine nitrogens bound to rhodium (because of the absence of a proton bound to this nitrogen). 2-Bond HSQC, which requires higher concentrations and longer experiment times, detects all histidine nitrogen atoms with or without a bound proton. Thus, we used the 1-bond HSQC experiment to monitor reactions, and the 2-bond HSQC experiment to fully characterize the products.

The site at which the rhodium fragment binds was revealed by the differences in chemical shifts between the ^{15}N-^{1}H HSQC spectra of 3*His-Apo-CA and the analogous spectra of the product from addition of the Rh-precursors to this protein. The 1-bond HSQC spectrum of Apo-3*His-CA contains resonances for four of the nine histidine residues (Fig. 8).

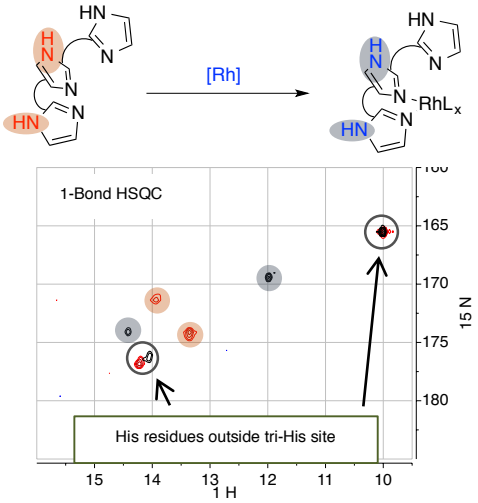


Figure 8. Comparison of the 1-Bond [^{15}N]-[^{1}H] HSQC spectra of Apo-3*His-CA (red) and the Rh-CA formed from the addition of [Rh(nbd)$_2$]BF$_4$ to 3*His-Apo-CA (blue). Only N-H nitrogens that do not rapidly exchange (highlighted) are observed in this experiment.

By comparing this spectrum to that of 3*His-Zn-CA, we determined that two of the four observed signals correspond to histidines in the tri-His site that binds zinc, and two correspond to histidines outside the tri-His site. Thus, if rhodium binds to the tri-His site, the chemical shift of the two tri-His signals in Apo-3*His-CA should change upon addition of the rhodium complex, while those of the two signals from histidines outside the tri-His site should not change. Indeed, addition of 1 equiv of [Rh(nbd)$_2$]BF$_4$ to Apo-3*His-CA caused only the chemical shifts of the signals corresponding to histidines in the tri-His site to change. A control experiment revealed that the addition of 1 equiv. of [Rh(nbd)$_2$]BF$_4$ to Zn-CA did not affect the chemical shifts corresponding to any of these histidine residues (Fig. 9) These results provide strong evidence that the Rh fragment binds to the tri-His site of CA.

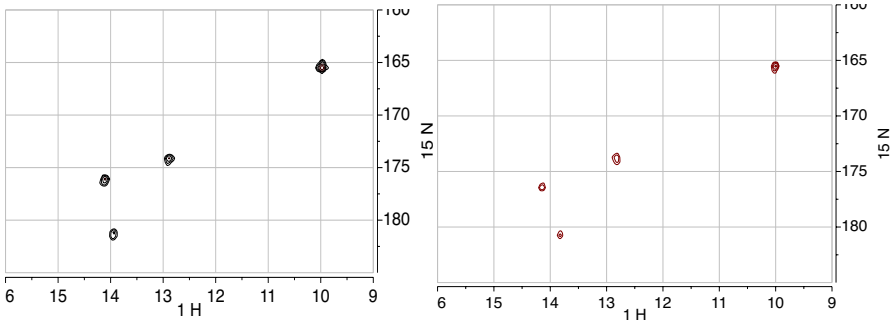


Figure 9. Comparison of the [^{15}N]-[^{1}H]-HSQC spectra of Zn-CA (left) and Zn-CA + [Rh(COD)$_2$]BF$_4$ (right).

The two Rh-CA proteins generated by addition of [Rh(nbd)$_2$]BF$_4$ and Rh(CO)$_2$(acac) to Apo-3*His-CA were also distinguishable by 1-bond HSQC spectroscopy (Fig. 8). The addition of 1 equiv of Rh(CO)$_2$(acac) to Apo-3*His-CA generated a Rh-3*His-CA protein with an ^{15}N-^{1}H HSQC spectrum that is distinct from that of the Rh -3*His-CA generated from [Rh(nbd)$_2$]BF$_4$, as well as that of Apo-3*His-CA (Fig. 10). These data corroborate our conclusion from UV-vis spectroscopy that these two Rh-precursors generate different Rh-CA proteins. Furthermore, the ability to distinguish among different Rh-CA's rapidly on solution-state samples using a protein concentration of 0.2 mM makes this method directly applicable to the monitoring of catalytic reactions in situ because catalysts are commonly used at this concentration or higher.

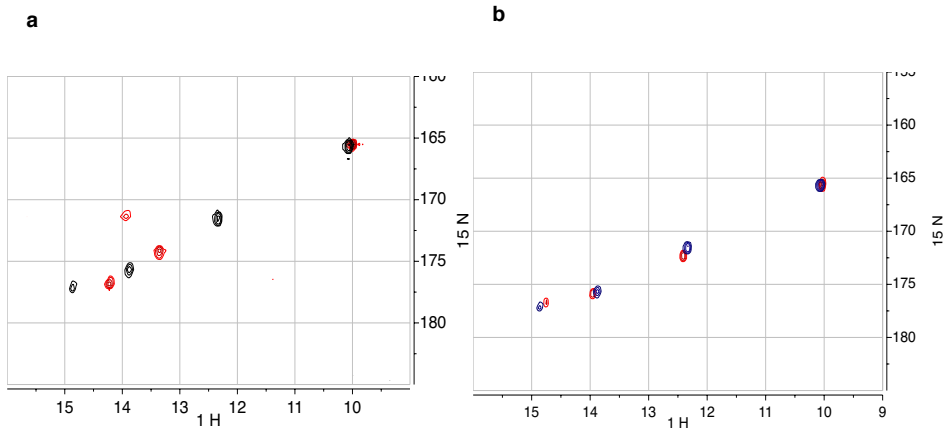


Figure 10. Characterization of the product of the reaction of Apo-3*His-CA with $Rh(CO)_2(acac)$ by 1-bond [^{15}N]-[^{1}H] HSQC a. The 1-Bond [^{15}N]-[^{1}H] HSQC spectrum of Apo-3*His-CA (red) and the spectra of the Rh-CA formed from the addition of $Rh(CO)_2(acac)$ to 3*His-Apo-CA (blue). b. The 1-Bond [^{15}N]-[^{1}H] HSQC spectra of $Rh(CO)_2$-3*His-CA (red) and $Rh(CO)_2$-H94A-3*His-CA (blue).

The identity of the Rh-bound nitrogens and the coordination number of the Rh in the organometallic protein was revealed by 2-bond [^{15}N]-[^{1}H] HSQC spectroscopy. To reduce the total number of histidine signals in the 2-bond HSQC spectrum, we acquired HSQC data with the mutant 6*His-CA (that lacks six surface histidine residues) instead of with 3*His-CA that lacks three (3*His-CA, Table 1). The 2-bond [^{15}N]-[^{1}H] HSQC spectrum of the product from addition of 1 equiv of $[Rh(nbd)_2]BF_4$ to Apo-6*His-CA contains ^{15}N NMR signals corresponding to the four nitrogens in the two histidines bound to rhodium (Fig. 11). The ^{15}N NMR chemical shifts corresponding to two of the nitrogens lie near the region in which Zn-bound nitrogens resonate in the ^{15}N spectrum of Zn-CA, and the signals corresponding to the other two nitrogen atoms lie in the region in which unligated histidines resonate. The spectrum of the product from addition of $Rh(CO)_2(acac)$ to Apo-6*His also contained signals corresponding to two metallated nitrogens (Fig. S3). From these data, we conclude that in both Rh-CAs the rhodium is bound by two histidine CA ligand.

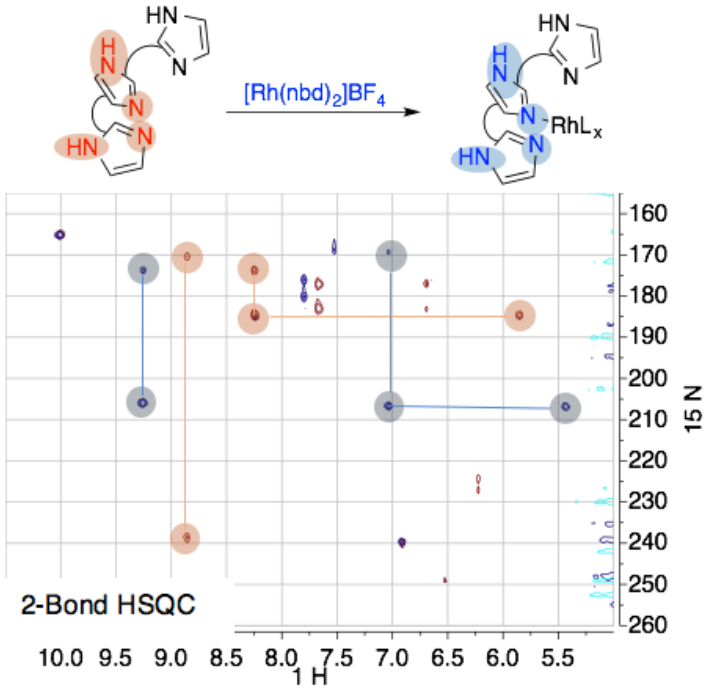


Figure 11. Comparison of the 2-Bond [^{15}N]-[^{1}H] HSQC spectra of Apo-6*His-CA (red) and the Rh-CA formed from the addition of $[Rh(nbd)_2]BF_4$ to 6*His-Apo-CA (blue). As a result of metallation, the signals from the histidine residues of the tri-His site in Apo-6*His-CA are different from those in the Rh-CA (highlighted).

To determine which two His residues ligate Rh, we prepared Rh-CA complexes from the three mutants 3*His-H94A-CA, 3*His-H96A-CA, and 3*His-H119A-CA, each of which lacks one His-residue in the metal-binding site. The UV-vis spectra of the proteins formed from adding $[Rh(nbd)_2]BF_4$ or $Rh(CO)_2(acac)$ to the three mutants are similar to each other and to those obtained from the Apo-3*His protein (Fig. 12).

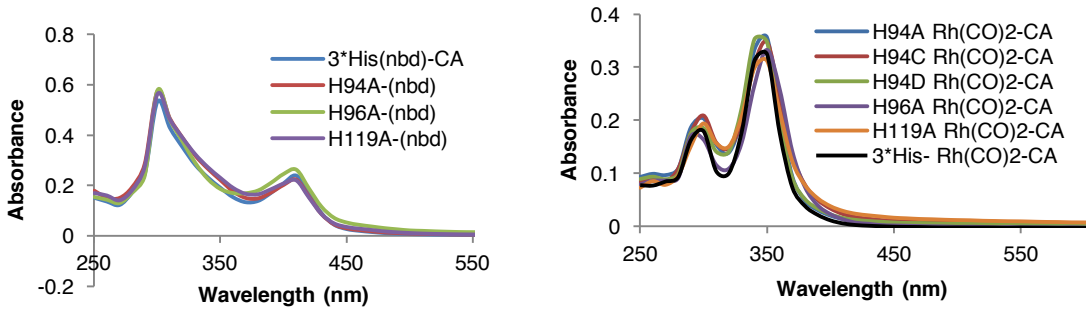


Figure 12. UV-vis spectra of the products from addition of Rh-complexes to apo-3*His-CA and the products from addition of the same Rh complexes to the three mutants with one histidine in the active site of 3*His-CA changed to alanine. Left: $Rh(nbd)_2BF_4$. Right: $Rh(CO)_2(acac)$.

Thus, any pair of histidine residues from the original Zn-binding site is capable of

CA ligand. However, the resonances in the [^{15}N]-[^{1}H]-HSQC spectra of the two proteins Rh-3*His-H94A and Rh-3*His-CA formed from addition of $Rh(CO)_2(acac)$ are identical to each other, while the resonances of the analogous proteins Rh-3*His-H96A-CA and Rh-3*His-H119A-CA are different from those of Rh-3*His-CA (Fig. 13). These data show that H96 and H119 are the two histidine residues that coordinate Rh within the native tri-histidine site.

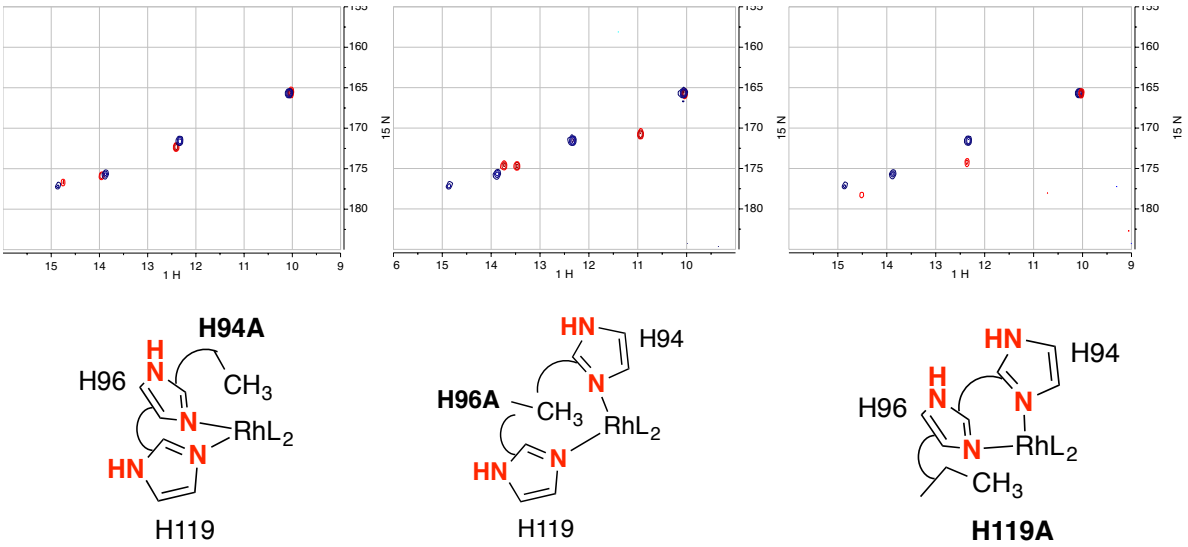


Figure 13. Comparison of the products from addition of $Rh(CO)_2(acac)$ to three mutants with one histidine in the active site of 3*His-CA changed to alanine in each case (red). In each case, the product from the alanine mutants is shown in comparison to the product formed from the addition of $Rh(CO)_2(acac)$ to 3*His-CA (blue).

To reveal the small-molecule ligands bound to rhodium in the Rh-3*His-CA proteins we used native nanoESI-MS. In the mass spectrum of the Rh-CA generated from $[Rh(nbd)_2]BF_4$, a mass (Fig. 14). In the mass spectrum of the Rh-CA generated from $Rh(CO)_2(acac)$, masses corresponding to both Rh(CO)-CA and $Rh(CO)_2$-CA were observed. These observed ions could result from a mixture of products or the dissociation of one CO ligand during the mass spectrometry (Fig. 14).

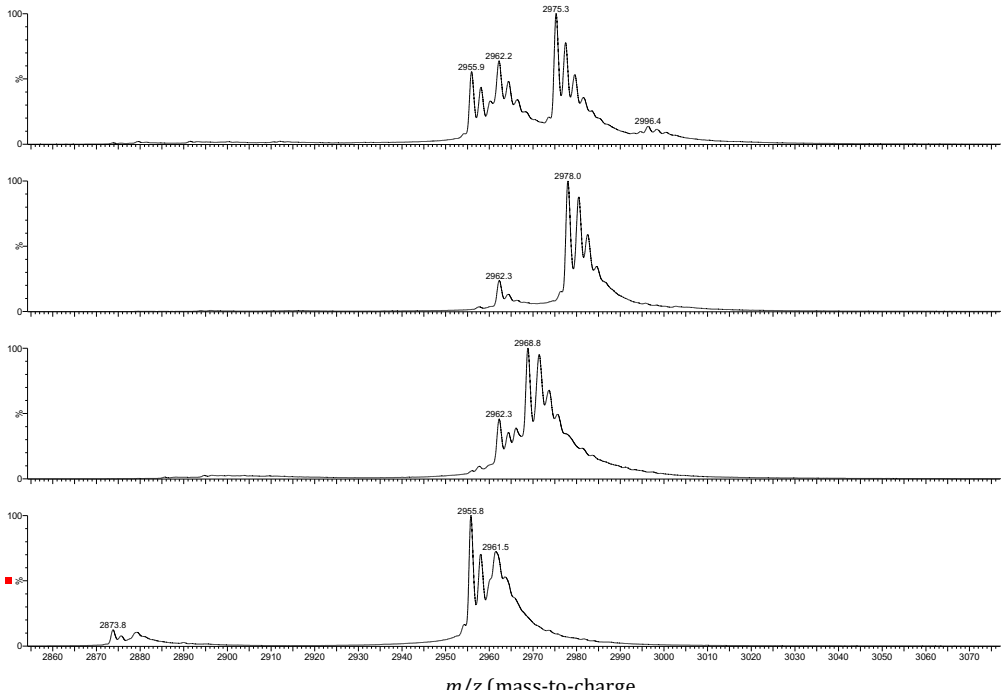


Figure 14. NS-ESI-MS of products from addition of organometallic fragments to CA. Top to Bottom: 1. Rh(nbd)-CA. 2. Rh4CO)$_2$-CA. 3. Ir(CO)$_2$-CA 4. Apo-CA.

To clarify the number of CO ligands in the Rh(CO)$_x$-CA protein, IR, EXAFS, and ^{13}C-NMR spectroscopy were performed (Fig. 15). The IR spectrum contains two CO stretching frequencies, consistent with the symmetric and anti-symmetric stretches of two CO ligands in equivalent or nearly equivalent environments (Fig. 13a), and the ^{13}C NMR spectrum contains two ^{13}CO signals. (Fig. 15b). The EXAFS spectrum of the sample matched the predicted spectrum of Rh(histidine)$_2$(CO)$_2$ and poorly matched the predicted spectrum of Rh(histidine)$_3$(CO) (Fig. 15c, Table S1). Together, these data indicate that the structure of the product of the reaction of Apo-CA CA), in which the two CO ligands lie in similar, but chemically inequivalent environments.

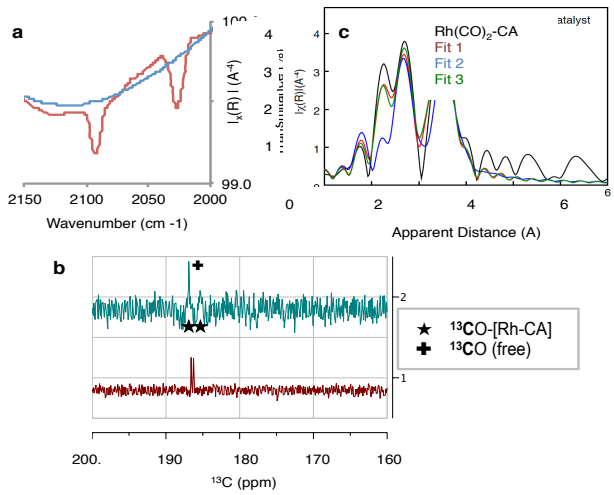


Figure 15. Characterization of $Rh(CO)_2$-CA. **a**. Comparison of the IR spectra of $Rh(CO)_2$-CA (red) and Apo-CA (blue) **b**. [^{13}C] NMR spectra of $Rh(^{13}CO)_2$-CA (top) and $Rh(^{13}CO)_2(acac)$ (bottom). **c**. Comparison of the EXAFS of $Rh(CO)_2$-CA to that of the three model complexes $Rh(His)_2(CO)_2$ (Fit 1), $Rh(His)_3(CO)_1$ (Fit 2), and $Rh(His)_3(CO)_2$ (Fit 3)

To assess the generality of these synthetic and analytical methods, we studied the reactions of apo-CA with $[Ir(cod)_2]BF_4$ and $Ir(CO)_2(acac)$. As observed for the [Rh]-analog, the reaction of $[Ir(cod)_2]BF_4$ with apo-CA led to low occupancy of the metal-binding site, whereas the reaction of apo-CA with $Ir(CO)_2(acac)$ led to >95% yield of the protein containing [Ir] at the original Zn-site (Figure 16).

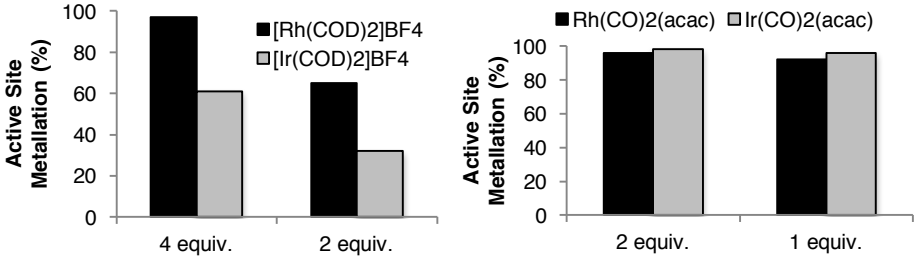


Figure 16. Comparison of the specificity for metallation of the active site of apo-3*His-CA by [Ir] and [Rh] precursors.

UV-vis spectroscopy (Fig. 7), [^{15}N]-[^{1}H]-HSQC spectroscopy (Fig. S4), and NS-ESI-MS (Fig 14) were fully consistent with a protein containing an $Ir(CA)(CO)_2$ site. Thus, this set of characterizations to determine the coordination mode of an organometallic fragment bound to a protein ligand can be broadly applied to protein scaffolds expected to coordinate metals using at least one histidine residue.

2.4 Characterization of Organometallic Reactions:

The methods used to characterize the structures of the artificial organometallic proteins in the solution phase were selected to be equally suited to monitor the stoichiometric reactions of the protein-ligated organometallic complexes. Specifically, we investigated ligand substitutions of Rh(nbd)-CA that would cleave and form metal-carbon bonds in the same reaction (Figure 17).

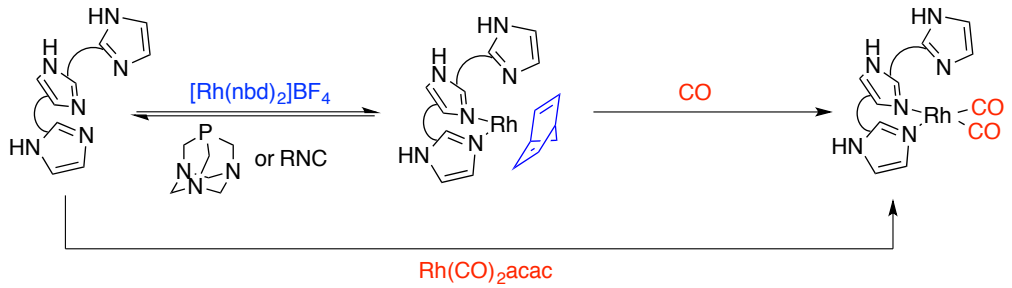


Figure 17. Summary of ligand substitution reactions achieved with Rh(nbd)-CA. Depiction of imidazole groups represents their relative orientation as histidine residues within CA.

The reaction of Rh(nbd)-CA with CO (1 atm) resulted in significant conversion of the starting material, as observed by UV-vis spectroscopy (Fig. 18). The decay of the absorption corresponding to Rh(nbd)-3*His-CA was accompanied by the appearance of an absorption at 350 nm, which matches the λ_{max} of $Rh(CO)_2$-CA. Based on the A_{350} value, $Rh(CO)_2$-CA formed in 53% yield, while 15% of the Rh(nbd)-CA starting material remained.

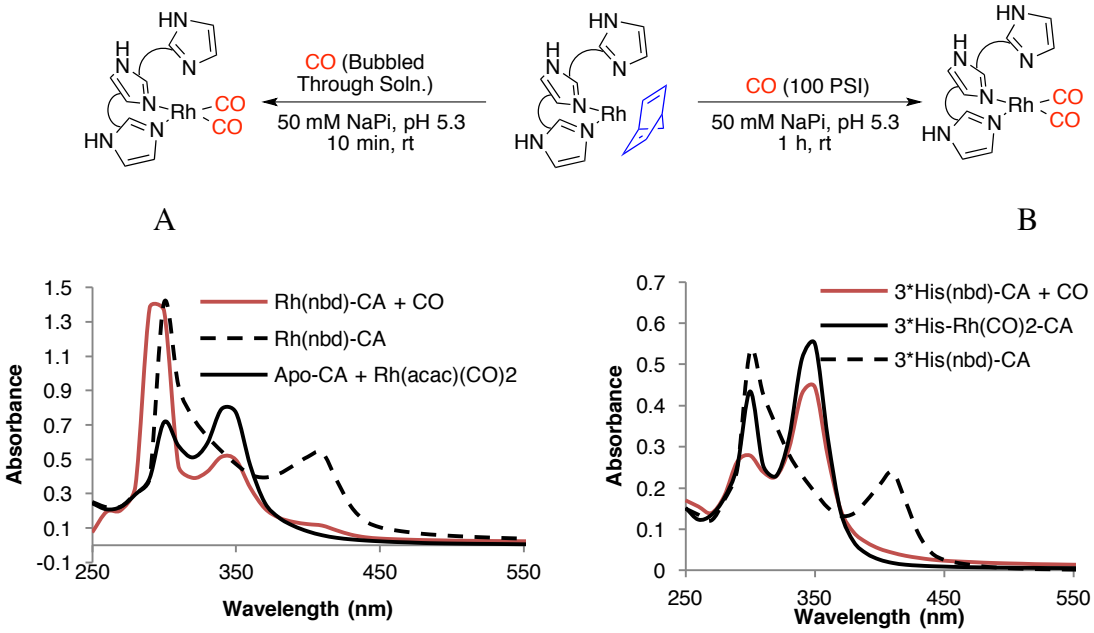


Figure 18. Characterization of the reaction between Rh(nbd)-CA and CO by UV-vis spectroscopy. A. Reaction conditions: CO was bubbled through a solution of Rh(nbd)-CA (0.5 mM) for 10 minutes. B. Reaction conditions: Rh(nbd)-CA (0.25 mM) was stirred in an autoclave under 100 PSI CO for 1 hour.

Further analysis of the reaction composition by [^{15}N]-[^{1}H]-HSQC spectroscopy also was consistent with the presence of $Rh(CO)_2$-CA and Rh(nbd)-CA; the remaining protein in solution CA ligand (Fig. 19).

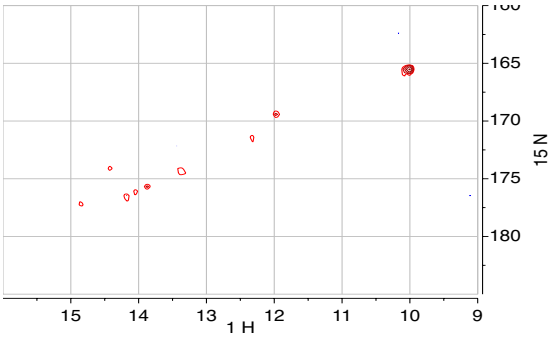


Figure 19. 1-Bond ^{15}N-^{1}H HSQC spectrum of the reaction mixture formed by adding CO to a solution of Rh(nbd)-CATo ensure full conversion, the reaction was conducted with a higher pressure of CO; indeed, the reaction of Rh(nbd)-3*His-CA with 100 psi of CO formed $Rh(CO)_2$-CA in 80% yield, as determined by UV-vis spectroscopy. (Fig. 16). This reaction is the first documented stoichiometric reaction to cleave and form metal-carbon bonds within an artificial protein.

2.5 Protein Controlled Reactivity

With our established methods to observe directly ligand substitution reactions occurring at the metal center, we investigated the potential of the protein scaffold to affect the rates and selectivities of ligand substitutions at the Rh-CA sites. The reactions of phosphines and isocyanides with Rh(nbd)-3*His-CA displace the CA ligand from the Rh-center (Fig. 17), and the protein scaffold can be modified to modulate both the rate and the selectivity of these reactions occurring at the protein-ligated Rh. The extent of these reactions was monitored by UV-Vis spectroscopy and the NPA hydrolysis activity assay (vide supra). The substitution of the CA ligand from the metal center by isocyanides to form Apo-CA depended on the steric properties of the added ligand. The reactions of Rh(nbd)-3*His-CA with the unencumbered *n*BuNC ligand progressed to the highest conversion, while the reactions with the most hindered *t*BuNC proceeded to the lowest conversion (Fig. 20).

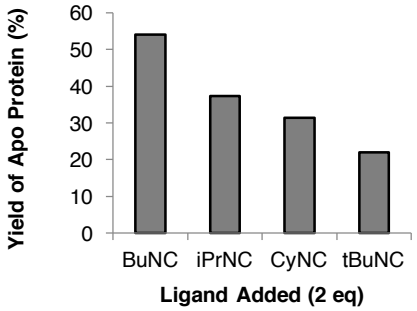


Figure 20. Analysis of the size selectivity of the substitution of the protein ligand from Rh(nbd)-CA with isocyanides.

Addition of the water soluble phosphine 1,3,5-Triaza-7-phosphatricyclo[3.3.1.13.7]decane (PTA) to 3*His-CA also resulted in complete conversion of the starting material. This reaction yielded [Rh(nbd)(PTA)2]+ and Apo-CA, as determined by UV-vis (Fig. 21), 31P NMR (Fig. 22), and [15N]-[1H] HSQC (Fig. 23) spectroscopy. Addition of excess ZnSO4 to the resulting Apo-CA generated Zn-3*His-CA, indicating that the protein structure remained intact throughout the introduction and elimination of rhodium (Fig. 23).

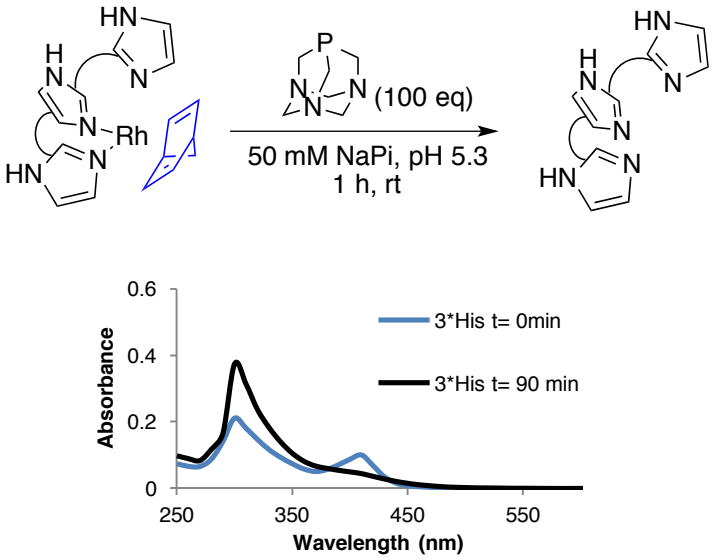


Figure 21. Analysis of the reaction of Rh(nbd)-CA with excess TPA by UV-vis spectroscopy. Blue = Rh(nbd)-3*His-CA before TPA addition. Black = Rh(nbd)-3*His-CA + TPA following 90 minutes of reaction time

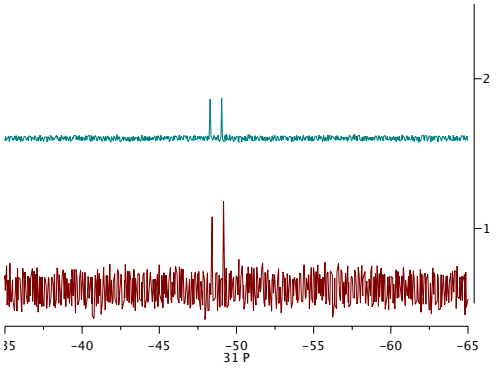


Figure 22. Analysis of the reaction of Rh(nbd)-CA with excess TPA by ^{31}P NMR spectroscopy. Top: Reaction between Rh(nbd)-CA and PTA. Bottom: Reaction between $Rh(nbd)_2BF_4$ and PTA.

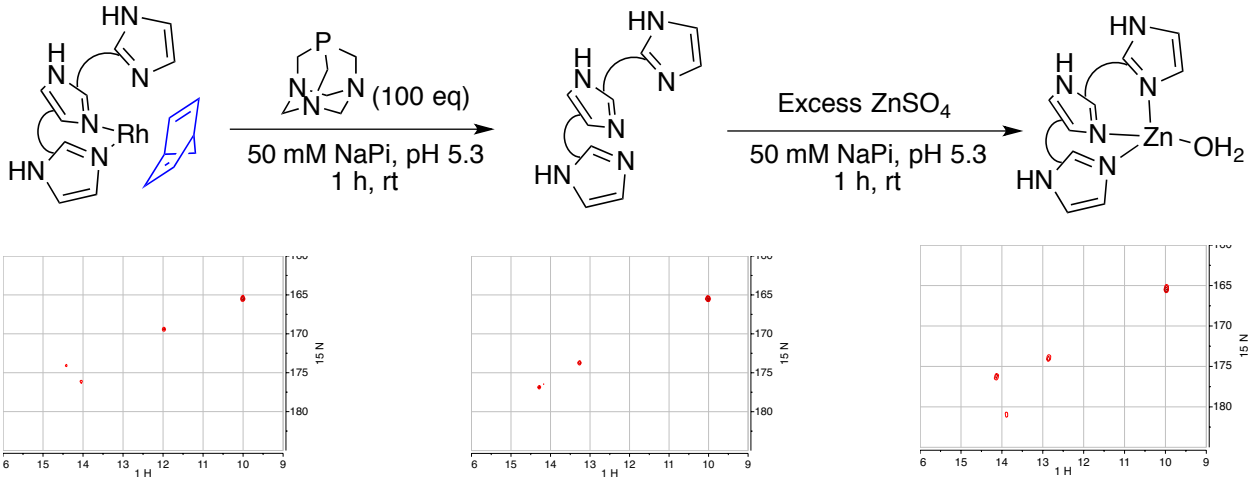


Figure 23. Characterization of the reaction of Rh(nbd)-CA with excess PTA by 1-Bond ^{15}N-^{1}H HSQC spectroscopy. Left: Rh(nbd)-CA SM. Center: Product following addition of PTA. Right: Subsequent product following addition of Zn-CA.

Reactions of PTA with Rh(nbd)-3*His-CA containing different mutations near the metal center indicated that the progress of this substitution reaction can be controlled by the protein structure. The proteins Rh-3*His-H94A and Rh-3*His-H94D were less reactive toward this ligand substitution than were Rh-3*His and Rh-3*His-H94C, while the Rh-3*His-H96A mutant was more reactive (Fig. 24, S5). The dependence of the reactivity of Rh(nbd)-3*His-CA on the size of the reagent and the identity of the mutant illustrate the potential to assess directly the effect of mutations and protein structure on an elementary reaction of an organometallic protein.

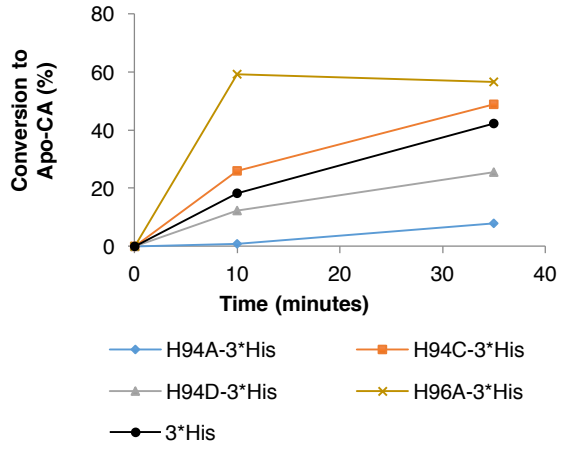


Figure 24. Time course for the formation of Apo-CA in the ligand substitution reaction of the phosphine PTA with several mutants of Rh(nbd)-CA.

Previous publications state that Rh-CA's prepared from $[Rh(COD)_2]BF_4$ and $Rh(CO)_2$acac are active catalysts for hydrogenation and hydroformylation of stillbenes and styrenes.[11] However, we find that the discrete organometallic protein complexes Rh(cod)-CA, Rh(nbd)-CA, and $Rh(CO)_2$-CA do not catalyze the hydrogenation or hydroformylation of a range of potential substrates for these reactions, including terminal, 1,1-disubstituted, and internal alkenes and vinylarenes. These findings suggest that the active catalyst of the previously reported systems was not a Rh-center

ligated at the native Zn-site of carbonic anhydrase. Instead, it is more likely that these reactions are catalyzed by a dissociated Rh-fragment or a fragment associated with a different site on the protein. Most important for the current work, our methods to synthesize and characterize discrete organometallic proteins by methods described here permit critical assessment of catalysis by specific complexes, much like one tests the competence of a small organometallic complex to be an intermediate in a catalytic process.

2.6 Conclusion

Although the catalytic reactivity of artificial organometallic proteins has been previously documented, the potential to understand and improve this class of catalysts has been limited by the absence of methods to synthesize and characterize a range of discrete catalyst systems and to evaluate their fundamental reactivity. The synthesis, characterization, and reactivity of the artificial organometallic proteins documented in the current work demonstrate the potential to gain information on the elementary reactions of discrete organometallic-protein complexes.

The expedient, solution state characterization methods presented in this work are crucial for determining the site and extent of metal binding and for evaluating the organometallic reactivity of the Rh-CA's. The efficient use of colorimetric activity assays to indirectly quantify the specificity of metal binding should be broadly applicable for the rapid assessment of a wide range of combinations of protein scaffold libraries and transition metal precursors. After identifying a selective coordination process, the binding mode of the protein ligand can be fully characterized by heteronuclear NMR spectroscopy, and the small molecule ligands at the metal site can be identified by the combination of ESI-MS, IR, and NMR spectroscopy. Even the elementary reactions of these site can be studied rapidly, setting the stage to reveal the ability of these macromolecular ligands to control the reactivity of the organometallic site in these artificial proteins and to create the potential to achieve selectivities beyond those observed with traditional organometallic complexes containing small organic ligands.

2.7 Experimental Information

2.7.1 General Methods

Unless otherwise noted, the chemicals, salts, and solvents used were reagent grade and used as received from commercial suppliers without further purification. All expression media and buffers were prepared using ddH_2O (MilliQ A10 Advantage purification system, Millipore). All expression media were sterilized using either an autoclave (45 min, 121^{o}C) or a sterile syringe filter (0.22 um). To maintain sterile conditions, sterile materials and *E. coli* cells were manipulated near a lit Bunsen burner. TEV protease was obtained from the UC Berkeley Macrolab.

2.7.2 Instrumentation:

a. Gel Electrophoresis

Protein purity was analyzed by sodium dodecyl sulfate-polyacrylamide (SDS-PAGE) gel electrophoresis using precast gels (polyacrylamide, 10-20% linear gradient, Biorad) according to standard protocols

b. Mass Spectrometry

Apo carbonic anhydrases were analyzed with an Agilent 1200 series liquid chromatograph connected in-line with an Agilent 6224 Time-of-Flight (TOF) LC/MS system using a Turbospray ion source. Rh and Ir carbonic anhydrases were analyzed by native nanoelectrospray ionization mass spectrometry (nanoESI-MS) using a Waters Q-Tof Premier quadrupole time-of-flight mass spectrometer equipped with a nanoESI source (Milford, MA). Mass spectra were acquired in the positive ion mode and processed using MassLynx software (version 4.1, Waters). The instrument is located in the QB3/Chemistry Mass Spectrometry Facility at UC Berkeley.

c. NMR Spectrometry

[^{15}N]-[^{1}H]-HSQC experiments were conducted using a Bruker Advance II 900 MHz NMR spectrometer and a TXI cryoprobe, which is part of the QB3 Central California 900MHz NMR facility at UC Berkeley. The general pulse program that was used has been previously described.[31] In the case of 1-bond [^{15}N]- [^{1}H] HSQC (J = 90 Hz, 256 points), the ^{15}N carrier frequency was 150 ppm (sw = 100 ppm, 200 points), the ^{1}H carrier frequency was centered on H_2O (~4.76 ppm, sw =30 ppm). In the case of 2-bond [^{15}N]- [^{1}H] HSQC (J = 30 Hz), the ^{15}N carrier frequency was 205 ppm (sw = 180 ppm), and the ^{1}H carrier frequency was centered on H_2O (~4.76 ppm, sw =30 ppm).

d. IR Spectroscopy

Apo and Rh-carbonic anhydrases were analyzed by IR spectroscopy using a Nicolet iS5 IR spectrometer (Thermoscientific) equipped with a iD1 transmission unit. For water compatibility, samples were analyzed on CaF_2 plates (33 x 3 mm).

e. X-ray Absorption Spectroscopy (XAS)

Data collection: Rh measurements were performed at the Stanford Synchrotron Radiation Laboratory (SSRL) on Beamline 7-3 at an electron energy of 3.0 GeV and an average current of 500 mA. The intensity of the incident X-ray beam was monitored using a N_2-filled ion chamber (I_0) in front of the sample. A Si(220) double-crystal monochromator was used. The data was collected as fluorescence excitation spectra with a Ge 30-element detector (Canberra). For Rh XAS, the monochromator energy was calibrated using the rising edge position of Rh foil (23220.00 eV) determined by the 1st derivative spectrum. XAS solution samples were prepared by loading the solution into a 0.8mm thick Lucite frame backed with Mylar tape. Samples were kept under liq. N_2 prior to the beamtime. Data were collected at 10K in a liquid helium flow cryostat (Oxford Inc.) under the exchanging helium gas atmosphere. To ensure that no X-ray induced radiation damage occurred, the XANES spectra over several scans were closely monitored for any reduction.

Data reduction and analysis: Data reduction was performed using SamView (SixPACK software, Dr. Samuel M. Webb, SSRL). The pre-edge and post-edge backgrounds were subtracted from the XAS spectra using Athena IFEFFIT software)[32,33], and the resulting spectra were normalized with respect to the edge height. Curve fitting was performed with Artemis and IFEFFIT software using *ab initio*-calculated phases and amplitudes from the program FEFF 8.2. These *ab initio* phases and amplitudes were used in the EXAFS equation: EXAFS curve fitting was carried out using the

crystal structural distances and the number of interactions of $Rh(CO)_2(acac)$ as starting parameters. The goodness of the fit was evaluated by the EXAFS R-factor that represents the absolute misfit (least-square fit) between theory and data (Table S2).

f. UV-Vis Spectroscopy

Protein concentration was determined using a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific). UV-Vis spectra of Apo and Rh-carbonic anhydrases were measured using a SpectraMax microplate reader (Molecular Devices) configured in the 96-well plate format. The hydrolysis of 4-nitrophenyl acetate to 4-nitrophenol was also measured using this instrument.

2.7.3 Plasmid Preparation

The wildtype carbonic anhydrase gene was obtained as a gift from the Fierke lab (University of Michigan) and cloned in the expression vector 2B-T (UC Berkeley/QB3 Macrolab, AddGene plasmid 29666), which contains a C-terminal six-histidine tag, a TEV cleavage site, and an ampicillin resistance gene. Site-directed mutagenesis was used to prepare all carbonic anhydrase variants. Reagents for PCR reactions were purchased either as a Quickchange II Kit (Agilent) or as individual components with no adverse effect. In a 200 uL PCR tube, 5 uL Pfu polymerase reaction buffer (Agilent) was added to 36.5 uL ddH_2O, followed by the addition of 50 ng dsDNA template, 130 ng (1.3 uL) forward primer, 130 ng (1.3 uL) reverse primer, 5 uL dNTP (Promega, 0.2 mM), and 1 uL Pfu polymerase. The reaction was spun in a microfuge (1 min, 5000 rpm) and subjected to the following thermocycler program: Segment 1 (1 cycle): 95° C for 30 s; Segment 2 (25 cycles): 95° C for 30 s, 55 °C for 1 min, 68 °C for 5 min; Segment 3 (1 cycle): 10 °C for 10 min. Directly to the reaction, 1 uL (20U) DPN 1 (New England BioLabs) was added, and the mixture was incubated at 37° C for 3 hours followed by transformation into chemically competent XL-Blue *E. coli* cells (UC Berkeley/QB3 Macrolab). Freshly transformed cells were plated on LB/ampicillin media and grown overnight at 37° C. Single colonies were used to inoculate 4 mL LB/amp cultures, which were shaken for 12 h at 250 rpm. Plasmid DNA was isolated using a Qiaprep Spin Miniprep Kit (Qiagen) and sequenced to verify mutation incorporation (UC Berkeley Sequencing Facility). DNA concentration was determined using a Nanodrop. All plasmids were stored at -20° C and used without further purification.

2.7.4 Protein Expression and Purification

a. Unlabeled Carbonic Anhydrase

Carbonic anhydrase was over-expressed in chemically competent BL21 DE3-RIL *E. coli.* cells (UC Berkeley Macro Lab) using a variation of the Studier autoinduction method. Freshly transformed cells were plated on LB/ampicillin media and grown overnight at 37° C. Single colonies were used to inoculate 3 mL LB/amp cultures, which were shaken at 37° C at 250 rpm for 6 h. Each 3 mL culture was used to inoculate 1 L of Studier rich autoinduction media,[34] which was supplemented with $ZnSO_4$ and vitamins[5]. Expressions were typically carried out in 4x1L batches in baffled flasks. Flasks were shaken at 37°C/ 250 rpm for six hours, after which the temperature was lowered to 28°C. Expression was allowed to proceed for 12 additional hours, after which cells were recovered by centrifugation (5000 rpm, 15 minutes, 4°C). Cell pellets were

frozen at -80°C, thawed, resuspended in Ni-NTA lysis buffer (50 mM NaPi, 300 mM NaCl, 10 mM imidazole, pH 8.0), and flash frozen until purification.

b. ^{15}N-Labeled Carbonic Anhydrase

Carbonic anhydrase was over-expressed in chemically competent BL21 DE3-RIL *E. coli.* cells (UC Berkeley Macro Lab) using the Studier autoinduction method for the preparation of uniformly labeled ^{15}N proteins with some modifications.[34] Freshly transformed cells were plated on P-0.5G media plates and grown 22h at 37° C. Single colonies were used to inoculate 4 mL P-0.5G/amp cultures, which were shaken at 37° C/ 250 rpm for 14 h. Each 4 mL culture was used to inoculate 50 mL, P-0.5G cultures, which was further shaken 10 h (37° C/ 250 rpm). Each 50 mL culture was used to inoculate 1 L of minimal auto-inducing media. Large cultures were shaken at 37°C/ 250 rpm for fourteen hours, after which the temperature was lowered to 30°C. Expression was allowed to proceed for 24 additional hours, after which cells were recovered by centrifugation (5000 rpm, 15 minutes, 4°C). Cell pellets were frozen at -80°C, thawed, resuspended in Ni-NTA lysis buffer (50 mM NaPi, 300 mM NaCl, 10 mM imidazole, pH 8.0), and flash frozen until purification.

c. Protein Purification

Frozen cells pellets were thawed and resuspended in 30 mL lysis buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH = 8.0). The cells were lysed in a cell homogenizer (5 passes, 10,000 PSI), and the cell debris were removed by centrifugation (90 min, 10,300 rpm) followed by syringe filtration of the supernatant (0.45 um filter). Ni-NTA (Qiagen) was added to the filtrate (8 mL resin per 40 mL supernatant), and the suspension was mixed on an end-over-end shaker (1 h, 4° C) and then poured into a glass frit. The resin was washed with wash buffer (50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, pH = 8.0) until the flow through contained no protein (Bradford assay). His-tagged carbonic anhydrase was eluded with 50 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH = 8.0. Protein purity was determined by SDS-PAGE electrophoresis (10%-20% polyacrylamide precast gels, Biorad). Pure fractions were pooled and dialyzed overnight to TEV cleavage buffer (50mM sodium phosphate, 50 mM NaCl, 50 mM Tris, 0.5mM EDTA). TEV protease (1: 50, w:w, TEV: CA) was added, and the reaction was incubated without mixing (20 h, 4° C). The extent of cleavage was monitored by LC-MS. If necessary, additional TEV protease was added to give give full conversion. The crude cleaved protein was dialyzed back to Ni-NTA lysis buffer (overnight, 4° C), and TEV protease and residual tagged carbonic anhydrase were removed by the addition 0.25 mL Ni-NTA resin to the crude cleaved protein solution, equilibration on an end-over-end shaker (1 h, 4° C), and filtration to remove the Ni-NTA resin. To generate apo protein, a solution of dipicolinic acid (250 mM in 50mM phosphate buffer, pH = 7.3) was added to the filtrate (1:4, v:v) and the resulting solution was incubated overnight without stirring (4° C). Dipicolinic acid was removed by sequential overnight dialysis against (1) 50 mM sodium phosphate, 50 mM NaCl, pH = 7.0, (2) 50 mM sodium phosphate, 50 mM NaCl, pH = 6.3, and (3) 50 mM sodium phosphate, 50 mM NaCl, pH = 5.3. The resulting apo protein was concentrated in Amicon spin concentrator (Corning, 10k Da MWCO) to the desired concentration and passed through a Sephadex G25 desalting column equilibrated with 50 mM sodium phosphate, pH = 5.3. Purified apo protein was stored at 4° C up to three weeks or flash frozen with the addition of 25% glycerol and stored at -80° C.

d. Determination of Protein Purity

SDS-PAGE gel electrophoresis was used to determine the purity of protein eluded from purification by the Ni-NTA chromatography. Protein identity and purity and ^{15}N incorporation (if applicable) was further confirmed by LC-TOF-MS (Fig. S6).

2.7.5 Metal Substitution Reactions

a. General Method

To a room temperature solution of apo-carbonic anhydrase in 50 mM phosphate buffer, pH = 5.3 (unless otherwise specified), a stock solution containing the desired equivalency of metal in DMSO (cationic complexes) or dioxane (neutral complexes) was added. Stock solutions were prepared at the appropriate concentration such that the final organic solvent concentration of the reaction was 2.5%. The reaction was mixed briefly by pipet and incubated without mixing at room temperature for 30 min followed by storage at 4° C. All analytical experiments presented were conducted with freshly prepared Ir or Rh-carbonic anhydrases, which were prepared from fresh stocks of metal complexes. Metal complexes were stored for the long term under inert atmosphere, however stock solutions were prepared under air without the use of degassed or rigorously dried solvents. The use of inert atmosphere conditions was not found to either improve or alter the results of metallation reactions.

b. Assay for Active Site Specificity of Metal Binding

Metal-carbonic anhydrases were prepared as described in the general method using 69 uM apo-CA. A 5 uL aliquot of each crude reaction mixture was added to 195 uL assay buffer (100 mM BES, 3 mM $CoCl_2$, pH = 7.2) in a separate well of a UV-Vis compatible 96-well plate. Background activity was determined using Co-free buffer. Just before analysis by UV-vis spectroscopy, a stock solution of 4-nitrophenylacetate (200 mM in acetonitrile) was diluted in assay buffer (1:25, v:v), and 50 uL of this diluted stock was added to each well of the 96-well plate. Using a microplate reader, ΔA_{405} of each reaction was determined over the course of 5 minutes (15 s intervals). The rate of hydrolysis of each metal-carbonic anhydrase was compared to the rate of hydrolysis in this assay in the absence of added [Co] and for Apo-CA at the same protein concentration (Fig. S7).

2.7.6 Characterization of Artificial Carbonic Anhydrases

a. UV-Vis Spectroscopy:

UV-Vis spectra were acquired using samples prepared by the standard metallation reaction conditions using 0.2 mM CA in NaPi buffer (50 mM, pH 5.3). In the case of metallated proteins, the spectrum of an equal-concentration apo-protein sample was recorded as a background that was subtracted from the presented spectra.

b. ^{15}N-^{1}H HSQC NMR Spectroscopy

NMR samples were prepared using 450 uL of either 0.166 mM (1-bond analysis) or 1.0 mM (2-bond analysis) CA in 50 mM phosphate buffer, pH 5.3. To this, 50 uL D_2O was added, and the

sample was mixed by pipet. Metal stock solutions were added in DMSO-d_6 (cationic complexes) or dioxane-H_{12} (neutral complexes), and samples analysis by NMR began within 1 hour of sample preparation. Additional spectra to those described in the main text can be found in supplemental supporting figures. (Fig. S8-S13).

c. NS-ESI-MS Sample Preparation

Samples for NS-ESI-MS were prepared via the standard metallation reaction using 0.25 mL of 0.135 mM Apo-3*His-CA. The entire reaction was loaded onto a NAP-5 column that was pre-equilibrated with NH_4OAc buffer (10 mL, 250 mM, pH=6) and the initial flow-through was discarded. Following sample loading, 250 uL equilibration buffer was added to the column, and the resulting flow-through was discarded, after which 700 uL additional equilibration buffer was added to elude the CA from the column. The NAP-5 column was re-equilibrated with 10 mL equilibration buffer, after which 500 uL of the previously eluded CA was added to the column, and the resulting flow-though was discarded. The CA was then eluded using 1 mL equilibration buffer. This sample was stored at 4^{o}C or on ice until analysis by ESI-MS.

d. EXAFS

Samples for EXAFS were prepared using 0.5 mL 1 mM Apo-3*His- CA using the standard protein metallation reaction conditions, except that the metal stock solution was added in two equal portions with pipet mixing between additions. The entire sample was loaded onto NAP-5 column that had been pre-equilibrated with NaPi buffer (50 mM, pH 5.3, 40% glycerol) with glycerol added as a cryoprotectant. The sample was flash-frozen on a house-made EXAFS analysis slide and stored until analysis.

e. IR

Samples for IR were prepared using 0.5 mL 0.5 mM or 1 mM Apo-3*His- CA using the standard protein metallation reaction conditions, except that the metal stock solution was added in two equal portions with pipet mixing between additions in the case of 1 mM Apo-CA. The samples were analyzed by IR spectroscopy as aqueous solutions (10 uL) between two CaF_2 plates. The background was collected in the same way using the sample buffer (50 mM NaPi buffer, pH = 5.3), and the software automatically subtracted this background. 500 scans were collected for both the background and the samples. The 0.5 mL 1 mM sample was then loaded on to a NAP-5 column that had been pre-equilibrated with the same sample buffer. Elusion with 1 mL of the same buffer resulted in a 0.5 mM samples. This sample was also analyzed by IR spectroscopy in the same manner. Auto baseline correction was used to analyze the data.

f. ICP-OES:

Apo-carbonic anhydrase (0.25 mL, 0.12 mM, 50 mM NaPi, pH = 5.3) was metallated with Rh-precursors using the general method previously described. Control experiments were performed by adding Rh to the same buffer containing no CA. Each 0.25 mL sample of crude Rh-CA was diluted to 1 mL, and 0.5 mL of this sample was then purified using a NAP-5 column to remove unbound Rh, resulting in a 2-fold dilution of the sample. The remaining, unpurified sample was diluted 2-fold and used in a control experiment. Control reactions containing no protein subjected to the same experiments. An 0.333 mL portion of each 1 mL sample was diluted with nitric acid (0.667 mL, 50% in H_2O, trace metals grade), and the samples were incubated first overnight at room temperature and then at 80^{o}C for 1 hour. An 0.5 mL portion of each digested sample was

diluted with nitric acid (4.5 mL, 2.5%, trace metals grade). These samples were analyzed with a Perkin Elmer Optima 7000 DV ICP-OES calibrated with commercial Rh standards (0 ppb, 25 ppb, 50 ppb, 100 ppb, 250 ppb). Analyses were conducted in triplicate and reported data are the average of these replicates (Fig. 3).

2.7.7 Ligand Substitution Reactions of Rh(nbd)-3*His-CA

a. Substitution with CO

Ligand substitution reactions were conducted with 0.25 mM Rh(nbd)-3*His-CA in 50 mM NaPi buffer, pH = 5.3, prepared as described in the general method. The Rh-CA starting material was characterized by UV-Vis spectroscopy and [^{1}H]-[^{15}N] HSQC in order to directly compare the starting material and reaction product. Two methods (A or B) were used to introduce carbon monoxide gas. Following the reactions, the ratio of Rh-CA (Rh(nbd)-CA plus $Rh(CO)_2$-CA) to Apo-3*His-CA (product) was determined by the hydrolysis assay previously described for the determination of metal binding specificity.

Method A: CO was slowly bubbled through a solution of Rh(nbd)- 3*His-CA for 10 minutes at room temperature.

Method B: The solution of Rh(nbd)- 3*His-CA was placed in an autoclave, the autoclave was pressurized with 100 PSI of CO, and the ensuing reaction was allowed to proceed for 1 hour at room temperature.

b. Substitution with Isocyanides

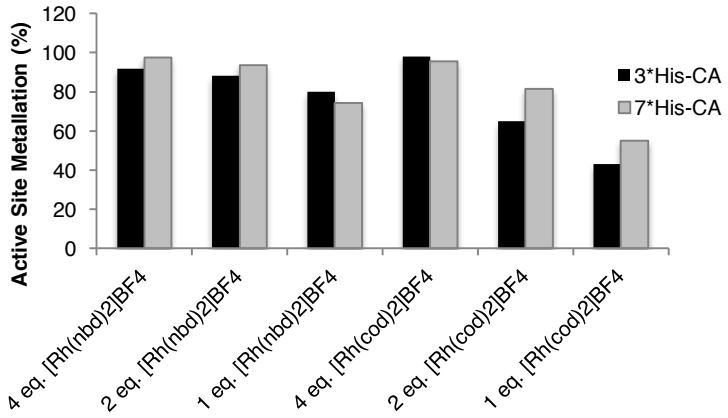
Stock solutions of isocyanides were prepared in dioxane in a glovebox under an atmosphere of nitrogen. Solutions were then removed from the glovebox, but kept closed until use. Rh(nbd)-3*His-CA (250 uL, 67 uM in 50 mM NaPi buffer, pH = 5.3) was prepared as described in the general method. Stock solutions of isocyanides (10 uL, 2 eq) were added to the Rh(nbd)-3*His-CA solution, and the reaction was allowed to proceed at room temperature for 10 minutes. The ratio of Rh-CA (SM) to Apo-3*His-CA (product) was determined by the hydrolysis assay previously described for the determination of metal binding specificity (Figure S14).

c. Substitution with PTA

Rh(nbd)-3*His-CA (250 uL, 67 uM in 50 mM NaPi buffer, pH = 5.3) was prepared as described in the general method. A solution of 1,3,5-Triaza-7-phosphatricyclo[3.3.1.13.7]decane (PTA) in water (10 uL, 100 eq.), freshly prepared in order to avoid slow oxidation to the phosphine oxide, was added to the protein solution, and the reaction was allowed to proceed. The amount of starting material remaining was monitored over time by UV-vis spectroscopy (A_{410}).

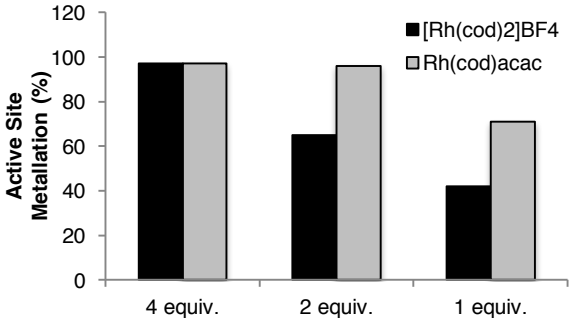
2.7.8 Supplementary Figures and Tables

Supplementary Figure 1



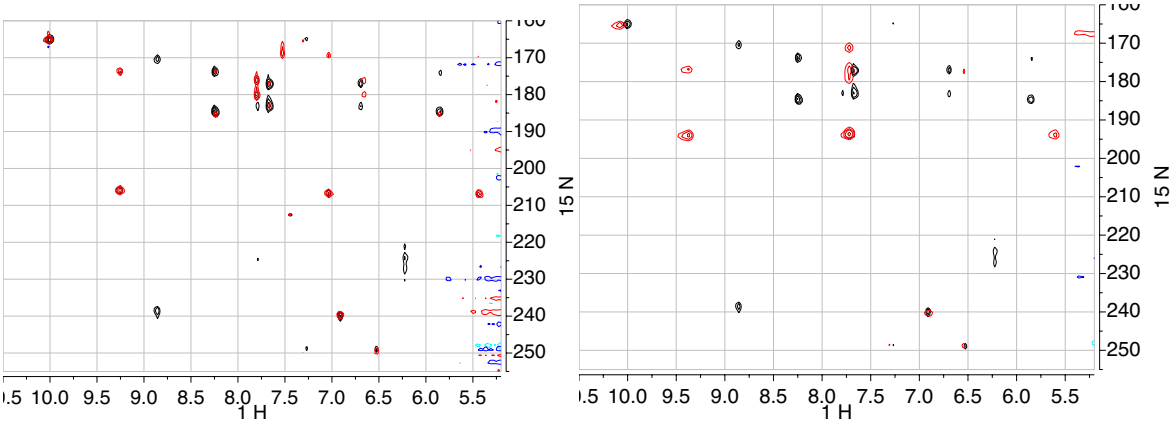
Comparison of the specificity for metallation of the active site in two mutants of CA. 3*His-CA refers to a mutant in which three of the nine histidines outside the active site have been changed to Arg or Phe (Table S1). Four of the remaining histidine residues are solvent exposed. 7*His-CA refers to a mutant in which seven of the nine histidines outside the active site have been changed to Arg, Phe, or Ala (Table S1). The remaining two histidine residues are buried and not expected to bind metals.

Supplementary Figure 2



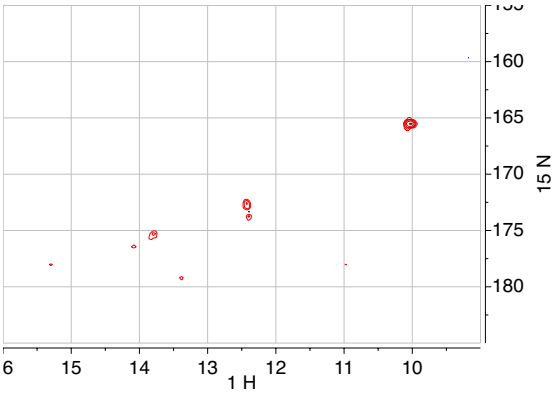
Comparison of the specificity for metallation of the active site of apo-3*His-CA with $[Rh(cod)_2]BF_4$ and Rh(cod)(acac).

Supplementary Figure 3



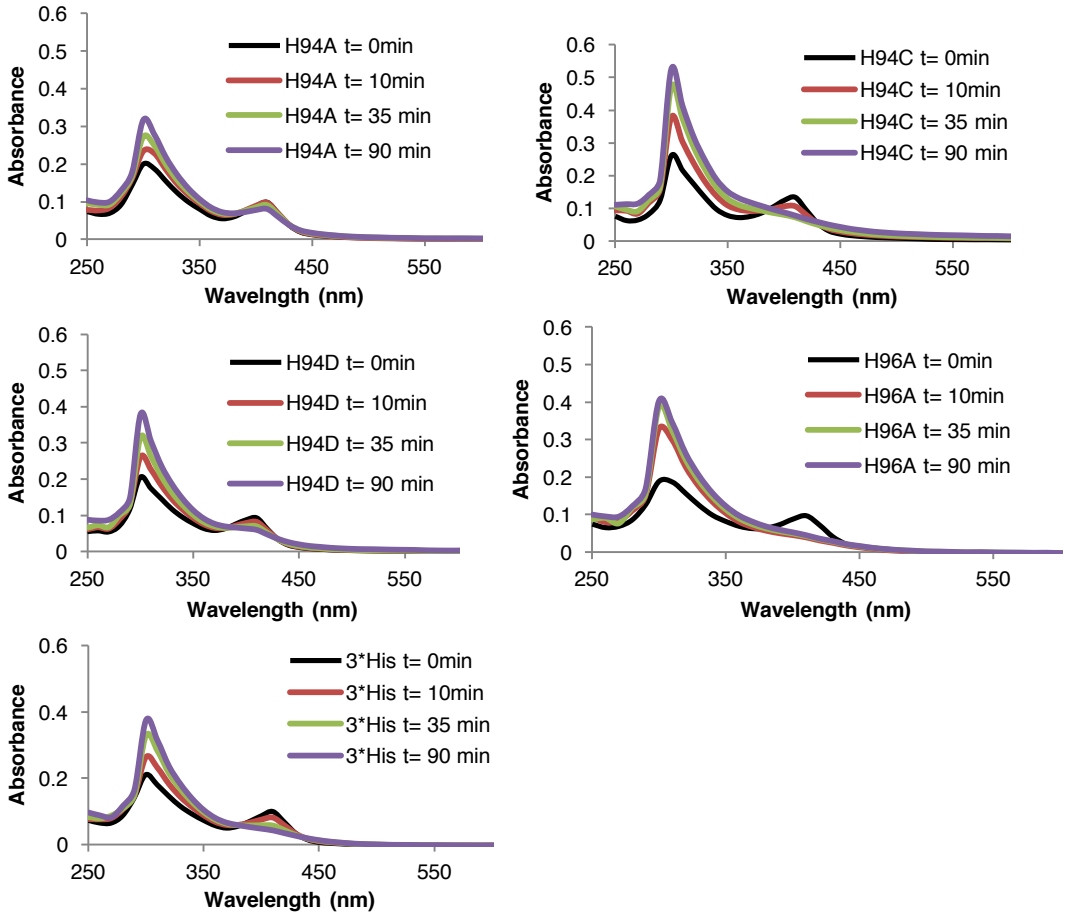
2-Bond HSQC spectra of Rh(nbd)-CA (top) and $Rh(CO)_2$-CA (bottom). In each case, the Rh-CA spectrum (red) is overlaid with the Apo-CA spectrum (gray).

Supplementary Figure 4



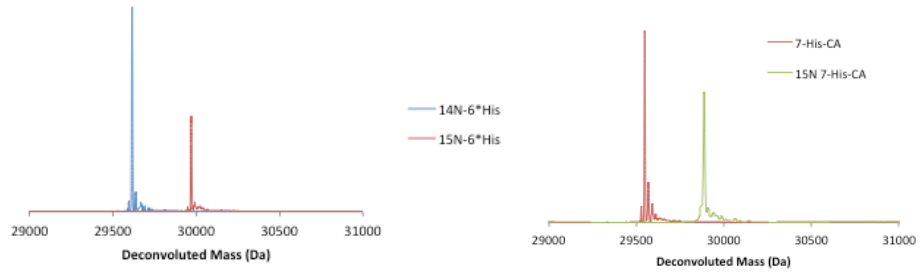
1-Bond ^{15}N-^{1}H-HSQC spectra of $Ir(CO)_2$-3*His-CA. Residual Zn-CA from the starting material can also be observed.

Supplementary Figure 5



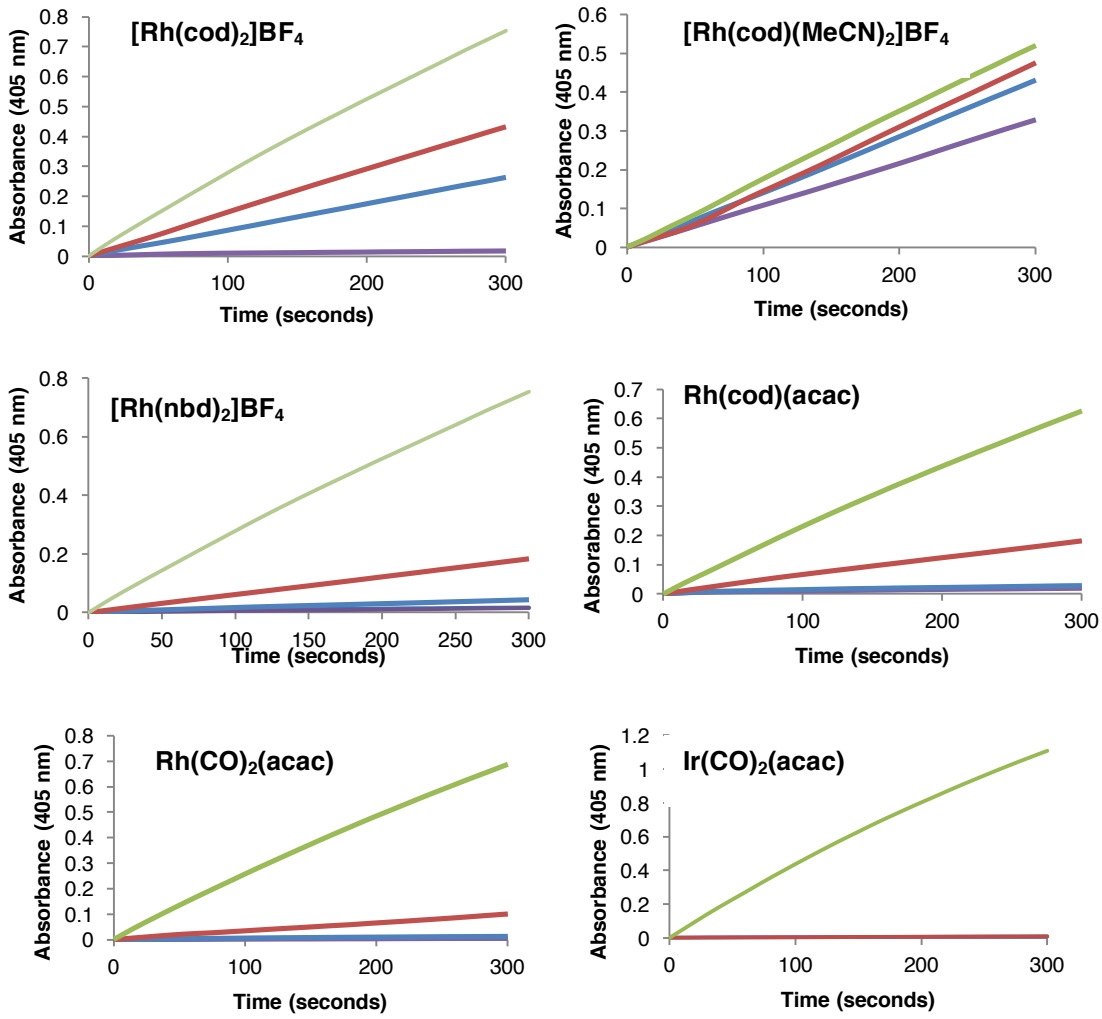
Comparison of the reaction of several Rh(nbd)-3*His-CA's with PTA. In each case, one residue near the ligated Rh-center has been mutated in order to perturb reactivity. Progress of reaction is evidenced by the loss of signal from the starting material at 410 nm.

Supplementary Figure 6



Comparison of unlabeled and ^{15}N labeled carbonic anhydrase (LC-TOF-MS). Left: 6*His-CA, Right: 7*His-CA.

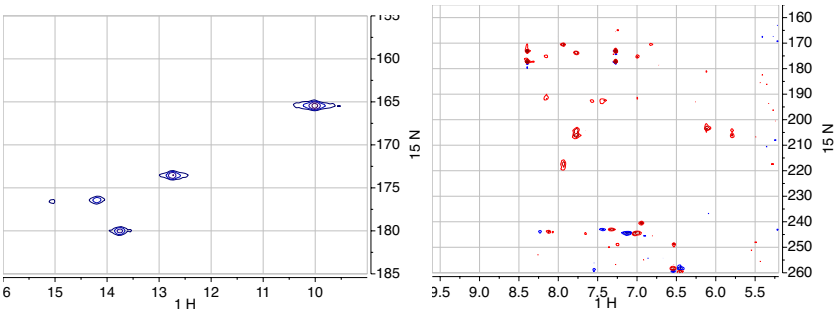
Supplementary Figure 7



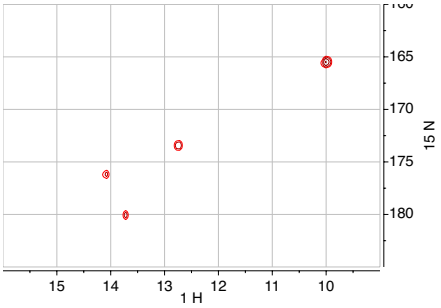
UV-Vis A_{405} data used to determine the active site metallation specificity of various metal precursors. Increasing A_{405} over time is due to the hydrolysis of 4-nitrophenyl acetate to 4-nitrophenol, catalyzed by Co-CA. Red = 1 equiv. added metal; Blue = 2 equiv. added metal; Purple = 4 equiv. added metal; Green = 0 equiv. added metal (control).

Supplementary Figure 8

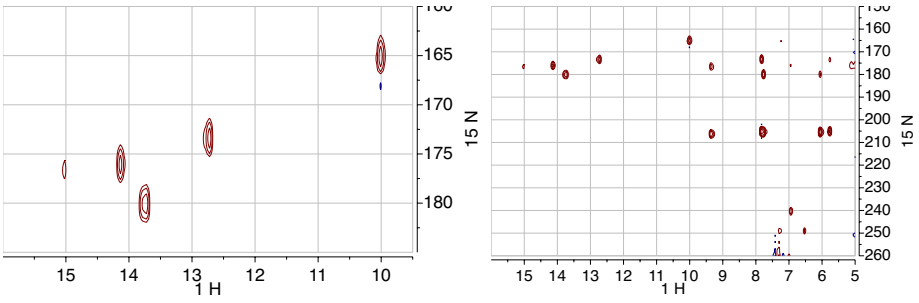
3*His:



6*His:



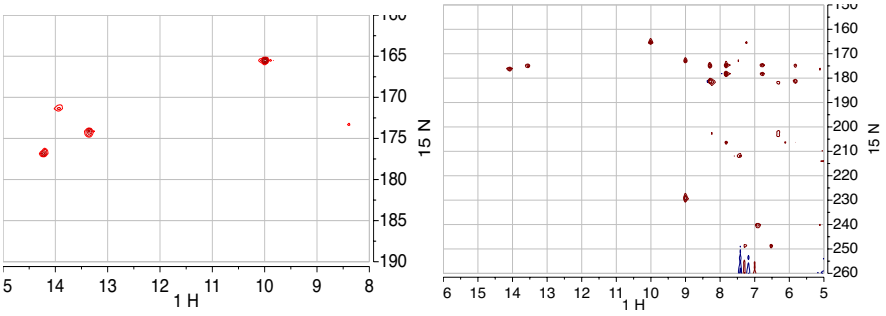
7*His:



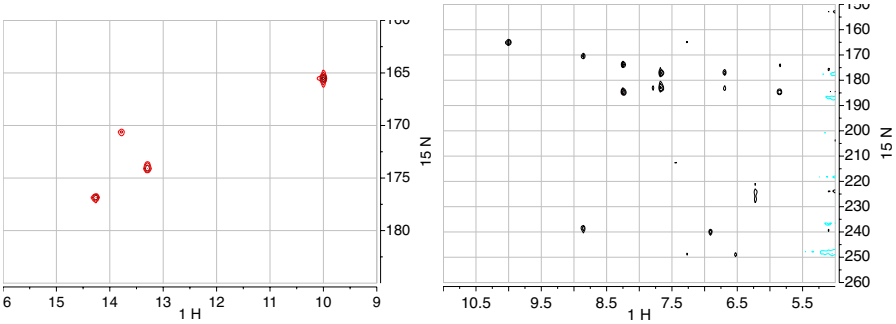
The [^{15}N]-[^{1}H] HSQC spectra of Zn-CA mutants (Left = 1-Bond; Right = 2-Bond (if applicable)).

Supplementary Figure 9

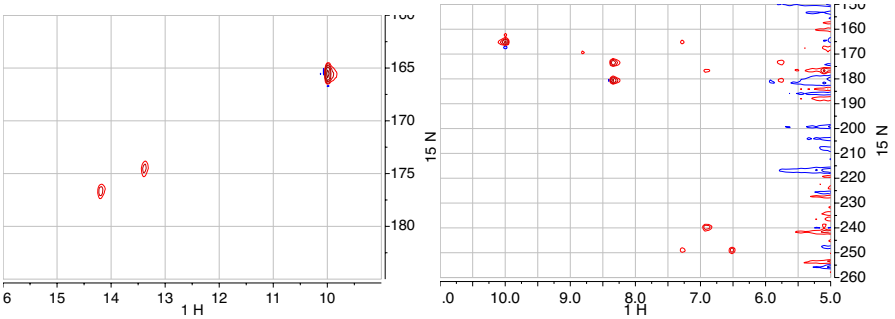
3*His:



6*His:

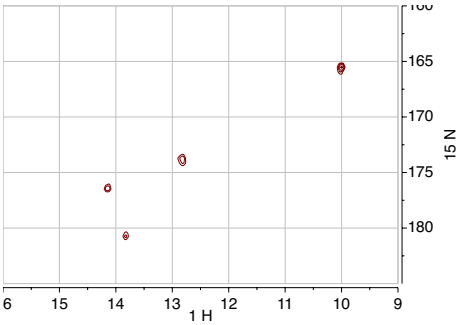


7*His:



The [^{15}N]-[^{1}H] HSQC spectra of Apo-CA mutants (Left = 1-Bond; Right = 2-Bond)

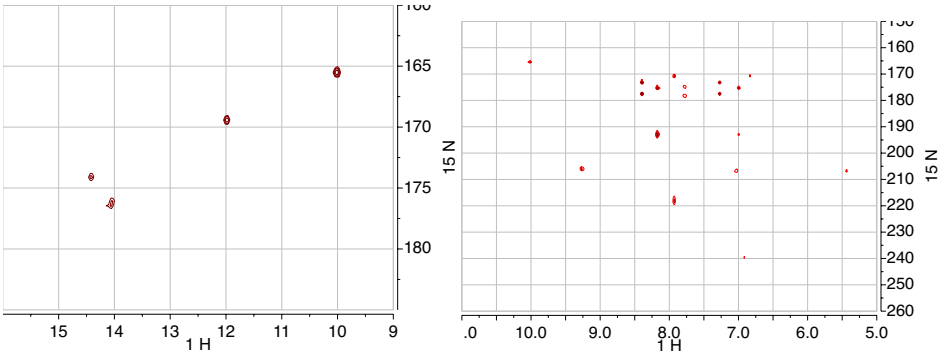
Supplementary Figure 10



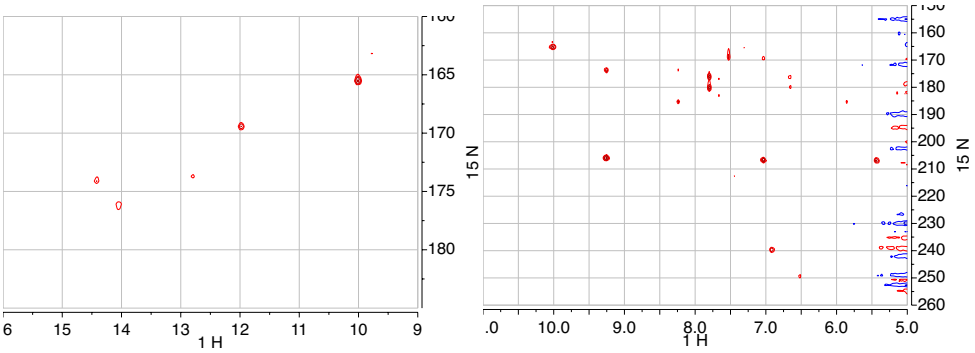
The [^{15}N]-[^{1}H] HSQC spectrum of 3*His-Zn-CA + $[Rh(cod)_2]BF_4$

Supplementary Figure 11

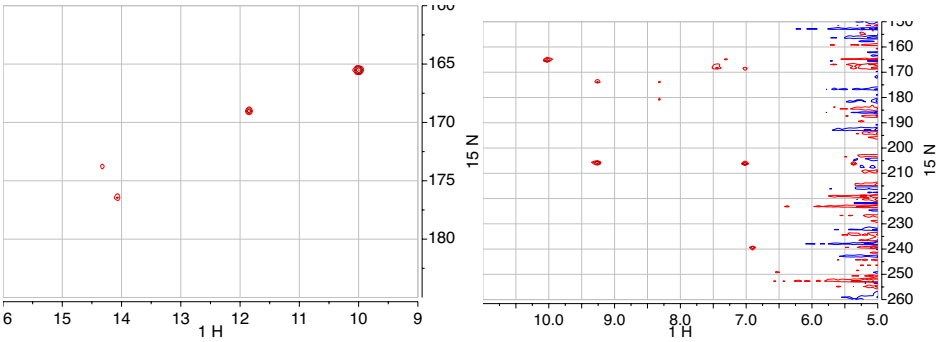
3*His:



6*His:

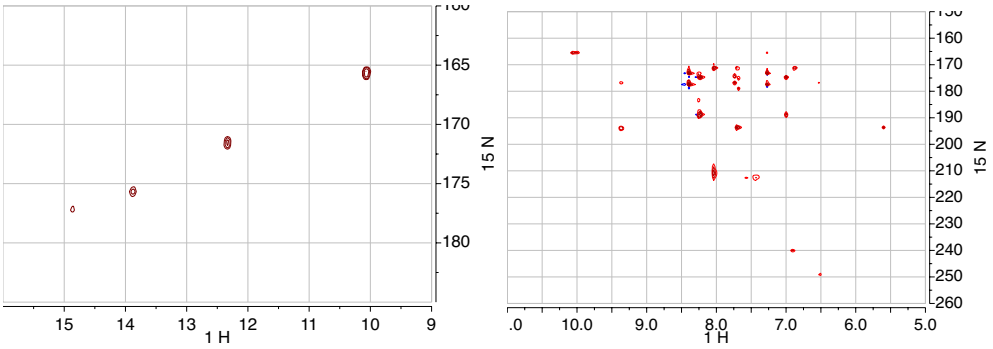


7*His:

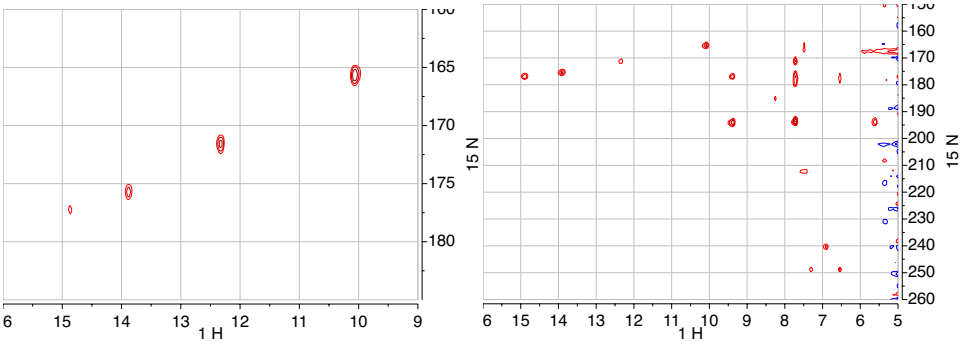


The [^{15}N]-[^{1}H] HSQC of Rh(nbd)-CA mutants (Left = 1-Bond; Right = 2-Bond)

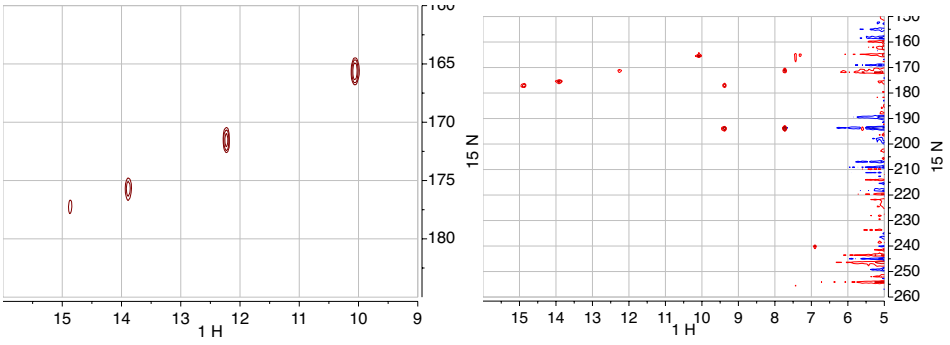
Supplementary Figure 12
3*His:



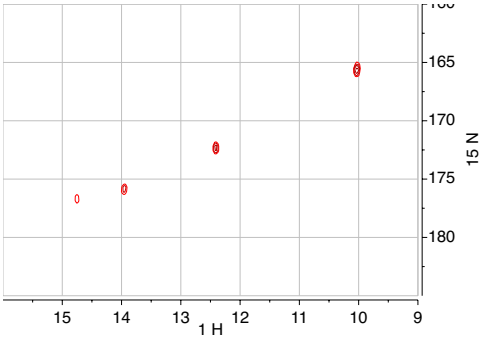
6*His:



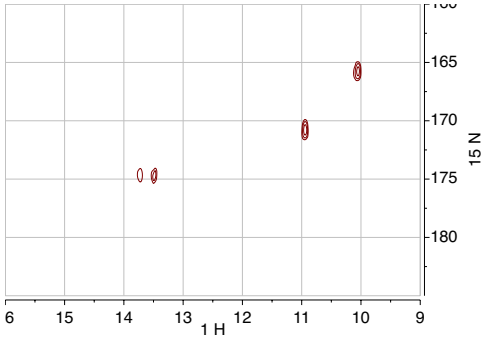
7*His:



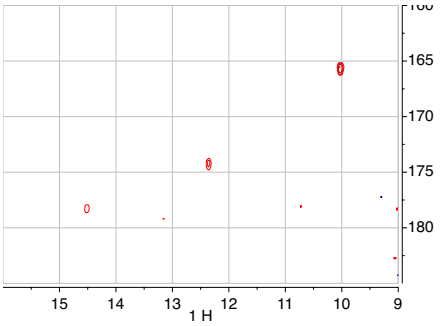
3*His- H94A



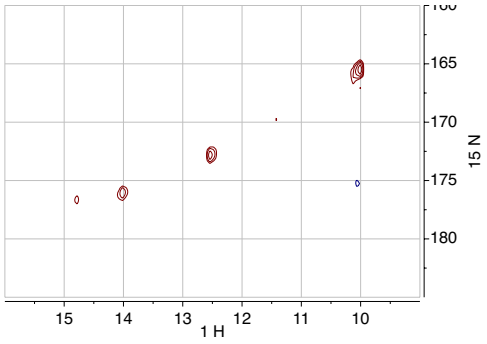
3*His- H96A



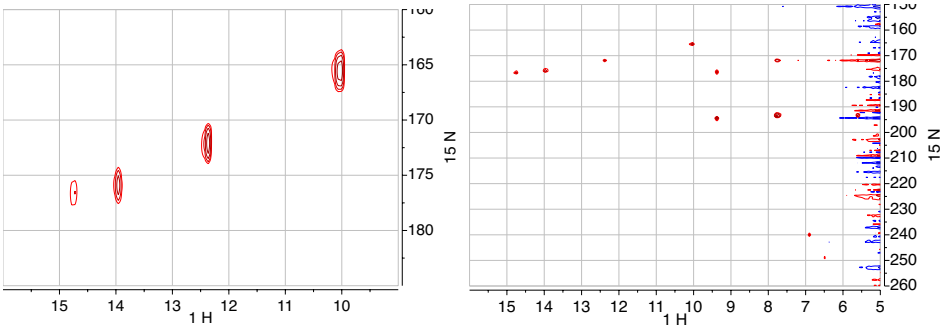
3*His- H119A



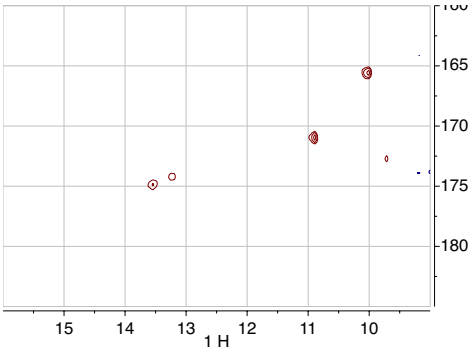
7*His- H94C



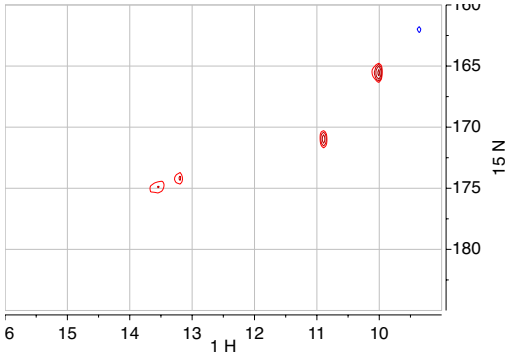
7*His- H94A



7*His- H96A



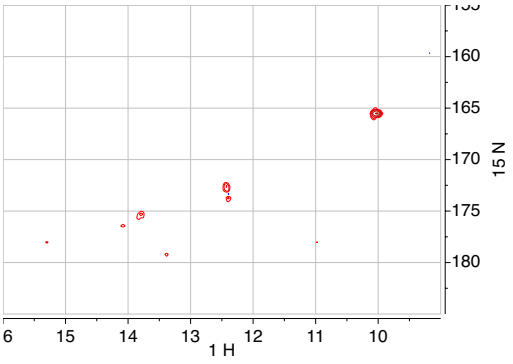
7*His- H96C



The [^{15}N]-[^{1}H] HSQC spectra of $Rh(CO)_2$-CA (Left = 1-Bond; Right = 2-Bond, if applicable).

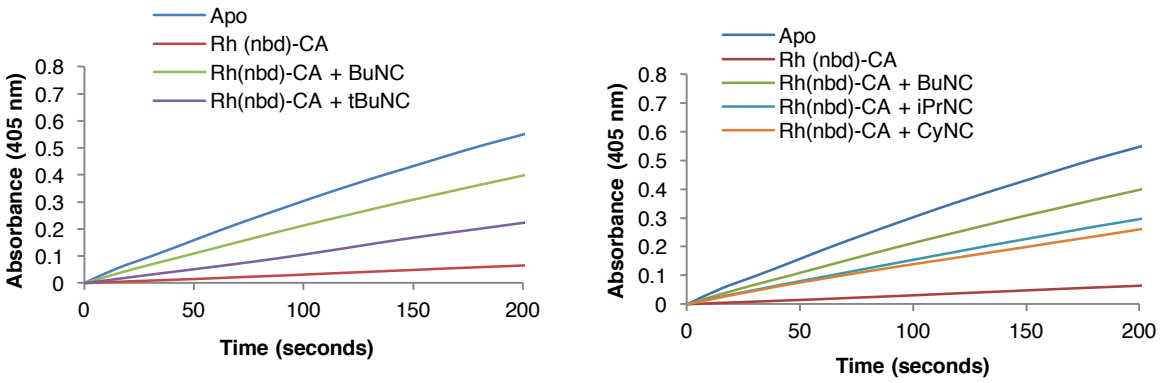
Supplementary Figure 13

7*His:



The [^{15}N]-[^{1}H] HSQC spectrum of Ir(CO)$_2$-CA. Small residual Zn-CA is present in sample.

Supplementary Figure 14



Activity Assay A_{405} traces used to quantify the amount of Apo-CA formed due to the ligand substitution reaction between Rh(nbd)-CA and isocyanides of various steric properties.

Supporting Table 1

EXAFS curve fitting parameters

Sample/Fit #		Path	R (Å)		N	σ^2 (Å^2)	R (%)
			XRD	EXAFS			
Rh reference		RhC	1.83	1.83 (0.03)	**2**	0.002 (0.001)	15.0
		RhO	2.04	2.03 (0.03)	**2**	0.004 (0.005)	ΔE=5.8
		RhCO(MS)	2.98 – 2.99	3.04 (0.04)	**6**	0.004 (0.003)	
		RhRh	3.25	3.16 (0.03)	**2**	0.003 (0.002)	
Rh catalyst	1	RhC		1.84 (0.02)	**2**	0.002 (0.002)	9.7
		RhN		2.11 (0.04)	**2**	0.006 (0.004)	ΔE=5.3
		RhCO(MS)		3.03 (0.06)	**6**	0.005 (0.002)	
	2	RhC		1.85 (0.03)	**1**	0.002 (0.001)	16.3
		RhN		2.13 (0.05)	**3**	0.008 (0.005)	ΔE=7.7
		RhCO(MS)		3.04 (0.03)	**1**	0.002 (0.002)	
	3	RhC		1.84 (0.02)	**2**	0.003 (0.002)	10.4
		RhN		2.11 (0.04)	**3**	0.009 (0.005)	ΔE=5.3
		RhCO(MS)		3.03 (0.03)	**2**	0.005 (0.002)	

S_0 was fixed to 1.0 based on the best fit parameter of Rh reference. N is the coordination number and σ^2 is the Debye-Waller factor. ΔE is the EXAFS threshold energy. R factor (%) indicates the goodness of the fit. Bold letters are the fixed parameters. MS shows multi-scattering path.

2.8 Acknowledgements

This work was supported by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02- 05CH11231 and by the NSF (graduate research fellowship). We thank the QB3 MacroLab facility (sub-cloning, shared equipment), Dr. Jeffery Pelton and the QB3 Central California 900 MHz NMR facility (useful discussion and instrumentation, supported by NIH grant gm 68933), Dr. Tony Iavarone and the QB3 Mass Spectrometry Facility (NS-ESI data collection, supported by NIH grant 1S10RR022393-01), Dr. Junko Yano for EXAFS data collection and analysis, and Prof. Carol Fierke (U. Michigan) for the wild type carbonic anhydrase gene.

3.9 References

1. Hartwig, J. F. *Organotransition Metal Chemistry*. (Univ Science Books, 2010).
2. Andreini, C., Bertini, I., Cavallaro, G., Holliday, G. L. & Thornton, J. M. Metal ions in biological catalysis: from enzyme databases to general principles. *J Biol Inorg Chem* **13,** 1205–1218 (2008).
3. Roiban, G.-D. & Reetz, M. T. Expanding the toolbox of organic chemists: directed evolution of P450 monooxygenases as catalysts in regio- and stereoselective oxidative hydroxylation. *Chem. Commun.* **51,** 2208–2224 (2015).
4. Acevedo-Rocha, C. G., Hoebenreich, S. & Reetz, M. T. Iterative saturation mutagenesis: a powerful approach to engineer proteins by systematically simulating Darwinian evolution. *Methods Mol. Biol.* **1179,** 103–128 (2014).
5. Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485,** 185–194 (2012).
6. Wilson, M. E. & Whitesides, G. M. Conversion of a protein to a homogeneous asymmetric hydrogenation catalyst by site-specific modification with a diphosphinerhodium(I) moiety. *J. Am. Chem. Soc.* **100,** 306–307 (1978).
7. Rosati, F. & Roelfes, G. Artificial Metalloenzymes. *ChemCatChem* **2,** 916–927 (2010).
8. Lewis, J. C. Artificial Metalloenzymes and Metallopeptide Catalysts for Organic Synthesis. *ACS Catal.* **3,** 2954–2975 (2013).
9. Carey, J. R. *et al.* A site-selective dual anchoring strategy for artificial metalloprotein design. *J. Am. Chem. Soc.* **126,** 10812–10813 (2004).
10. Ohashi, M. *et al.* Preparation of artificial metalloenzymes by insertion of chromium(III) Schiff base complexes into apomyoglobin mutants. *Angew. Chem. Int. Ed.* **42,** 1005–1008 (2003).
11. Ward, T. R. Artificial Metalloenzymes Based on the Biotin−Avidin Technology: Enantioselective Catalysis and Beyond. *Acc. Chem. Res.* **44,** 47–57 (2010).
12. Abe, S. *et al.* Control of the Coordination Structure of Organometallic Palladium Complexes in an apo-Ferritin Cage. *J. Am. Chem. Soc.* **130,** 10512–10514 (2008).
13. Abe, S., Hikage, T., Watanabe, Y., Kitagawa, S. & Ueno, T. Mechanism of Accumulation and Incorporation of Organometallic Pd Complexes into the Protein Nanocage of apo-Ferritin. *Inorg. Chem.* **49,** 6967–6973 (2010).
14. Abe, S. *et al.* Polymerization of Phenylacetylene by Rhodium Complexes within a Discrete Space of apo-Ferritin. *J. Am. Chem. Soc.* **131,** 6958–6960 (2009).
15. Podtetenieff, J., Taglieber, A., Bill, E., Reijerse, E. J. & Reetz, M. T. An Artificial Metalloenzyme: Creation of a Designed Copper Binding Site in a Thermostable Protein. *Angew. Chem. Int. Ed.* **49,** 5151–5155 (2010).
16. Ueno, T. *et al.* Size-Selective Olefin Hydrogenation by a Pd Nanocluster Provided in an Apo-Ferritin Cage. *Angew. Chem. Int. Ed.* **116,** 2581–2584 (2004).
17. Okrasa, K. & Kazlauskas, R. J. Manganese-Substituted Carbonic Anhydrase as a New Peroxidase. *Chem. Eur. J.* **12,** 1587–1596 (2006).
18. Jing, Q. & Kazlauskas, R. J. Regioselective Hydroformylation of Styrene Using Rhodium-Substituted Carbonic Anhydrase. *ChemCatChem* **2,** 953–957 (2010).
19. Jing, Q., Okrasa, K. & Kazlauskas, R. J. Stereoselective Hydrogenation of Olefins Using Rhodium-Substituted Carbonic Anhydrase—A New Reductase. *Chem. Eur. J.* **15,** 1370–1376 (2009).

20. Krishnamurthy, V. M. *et al.* Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein-ligand binding. *Chem. Rev.* **108,** 946–1051 (2008).
21. Togni, A. & Venanzi, L. M. Nitrogen Donors in Organometallic Chemistry and Homogeneous Catalysis. *Angew. Chem. Int. Ed.* **33,** 497–526 (1994).
22. Thorslund, A. & Lindskog, S. Studies of the Esterase Activity and the Anion Inhibition of Bovine Zinc and Cobalt Carbonic Anhydrases. *Eur. j. Biochem.* **3,** 117–123 (1967).
23. Hakansson, K., Wehnert, A. & Liljas, A. X-ray analysis of metal-substituted human carbonic anhydrase II derivatives. *Acta Crystallogr Sect D Biol Crystallogr* **50,** 93–100 (1994).
24. Winn-Deen, E. S., David, H., Sigler, G. & Chavez, R. Development of a direct assay for alpha-amylase. *Clinical Chemistry* **34,** 2005–2008 (1988).
25. Plocke, D. J., Levinthal, C. & Vallee, B. L. Alkaline Phosphatase of Escherichia coli: A Zinc Metalloenzyme *. *Biochemistry* **1,** 373–378 (1962).
26. Jieyi Wang *et al.* Physiologically Relevant Metal Cofactor for Methionine Aminopeptidase-2 Is Manganese. *Biochemistry* **42,** 5035–5042 (2003).
27. A rapid colorimetric method for the quantitative determination of copper oxidase activity (ceruloplasmin). **4,** 519–523 (1958).
28. in **1,** 531–616 (Thieme, Stuttgurt).
29. Although previous work indicates that surface Rh(COD) fragments can be removed from the CA scaffold by dialysis{Jing:2009gk}, we were not able to reproduce this selective removal of Rh bound outside the Zn-site. Likewise, the insolubility of Rh(CO)2(acac) prevented us from preparing the dialysis solution for introduction of this precursor as has been previously reported.
30. Shimahara, H. *et al.* Tautomerism of histidine 64 associated with proton transfer in catalysis of carbonic anhydrase. *J. Biol. Chem.* **282,** 9646–9656 (2007).
31. Mori, S., Abeygunawardana, C., Johnson, M. O. & van Zijl, P. C. Improved sensitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation. *J Magn Reson B* **108,** 94–98 (1995).
32. Ravel, B., Newville, M.IUCr. ATHENA, ARTEMIS, HEPHAESTUS: data analysis for X-ray absorption spectroscopy using IFEFFIT. *J Synchrotron Radiat* **12,** 537–541 (2005).
33. Newville, M.IUCr. IFEFFIT : interactive XAFS analysis and FEFF fitting. *J Synchrotron Radiat* **8,** 322–324 (2001).
34. Tyler, R. C. *et al.* Auto-induction medium for the production of [U-15N]- and [U-13C, U-15N]-labeled proteins for NMR screening and structure determination. *Protein Expr. Purif.* **40,** 268–278 (2005).

Chapter 3: Abiological Catalysis by Artificial Myoglobins Containing Noble Metals in Place of Iron

This chapter is modified from the following publication with permission of the university and all coauthors: Key, H.M.*; Dydio, P.* Clark, D. S. ; Hartwig, J.F. *Nature* **2016,** *534*, 534-537.

*Authors contributed equally

ABSTRACT: Catalysis enables modern synthetic chemistry. Transition-metal catalysts and enzymes are used widely in the production of bulk chemicals and the fabrication of fine chemicals and biologically active compounds. However, despite their exquisite selectivity and potential for evolution, metalloenzymes are rarely used, in part due to the limited scope of reactions catalyzed by natural enzymes. Thus, researchers have been seeking to expand the scope of substrates and the range of transformations with which metalloenzymes react. Approaches toward this goal have included the design of enzymes with substrate promiscuity by directed evolution and the constructions of proteins with new catalytic functions by incorporating transition metal complexes into protein scaffolds. At the intersection of these two approaches lies a third, more powerful strategy that we report in the submitted manuscript combining the attributes of each: the replacement of the native metal of a natural metalloenzyme with abiological metals to bestow new reactivity on an otherwise intact, selective and evolvable enzyme. We disclose artificial heme-protein analogs containing abiological metals in place of iron that catalyze reactions that the iron analogs do not. These discoveries rest on a method to generate and assess the catalytic activity of a diverse array of [M]-PIX-proteins rapidly. In particular, we show that myoglobin containing Ir(Me)-PIX catalyzes the enantioselective formation of carbon–carbon bonds by insertion of prochiral carbenes into C-H bonds. The same enzymes also catalyze the cyclopropanation of β-substituted styrenes and unactivated α-olefins. *No native or artificial enzymes are known to catalyze the C-H insertion process or the cyclopropanation of these olefins*. Moreover, in analogy to the native enzymes, these iridium myoglobins are suitable for directed evolution. Several rounds of mutagenesis of Ir(Me)-PIX-Myo generated enzymes that form either enantiomer of the C-H insertion products for several substrates and that catalyze enantio- and diastereoselective cyclopropanation of 1-octene. Given the rich chemistry of free metalloporphyrins and the ease of preparation and evolution of heme proteins containing diverse metals by the methods we report, we expect that many artificial metalloenzymes will be accessible for reactions inaccessible with any natural enzymes.

3.1 Introduction

Metalloenzymes combine the reactivity of transition metal catalysts with the potential for laboratory evolution.[1] Recently, by exploiting substrate promiscuity and site-directed mutagenesis,

the scope of reactions catalyzed by native metalloenzymes has been expanded to include abiological transformations.[2,3] However, this strategy is limited by the inherent reactivity of metal centers in native metalloenzymes. To further expand the catalytic potential of metalloproteins, artificial metalloproteins have been created by incorporating complete, noble metal complexes within proteins lacking native metal sites.[1,4,5] However, the interactions of the substrate with the protein in these systems are distinct from those of the native protein because the metal complex occupies the substrate binding site. At the intersection of these approaches lies a third strategy, by which the native metal of a metalloenzyme is replaced with an abiological metal that possesses distinct reactivity from that of the metal in a native protein.[6-8] This strategy could create artificial enzymes that catalyze abiological reactions within the natural substrate binding site of an enzyme and that can be subjected to directed evolution.

Here, we report the formal replacement of iron in Fe-porphyrin IX (Fe-PIX) proteins with abiological, noble metals to create enzymes that catalyze reactions not catalyzed by native Fe-enzymes or other metalloenzymes.[9,10] Variants of myoglobin containing an Ir(Me) site catalyze the functionalization of C-H bonds to form C-C bonds by carbene insertion and add carbenes to both b-substituted vinylarenes and unactivated aliphatic a-olefins. Directed evolution of the Ir(Me)-myoglobin generates mutants that form either enantiomer of the products of C-H insertion and catalyze the enantio- and diastereoselective cyclopropanantion of unactivated olefins. These discoveries were enabled by establishing an efficient method to prepare heme-proteins containing abiological metal-porphyrins, and they set the stage to generate artificial enzymes from innumerable combinations of PIX-protein scaffolds and unnatural metal cofactors to catalyze a wide range of abiological transformations.

To create artificial metalloenzymes formed by combining abiological metals and natural metalloprotein scaffolds, we focused on heme proteins containing Fe-porphyrin IX (Fe-PIX) as a metal cofactor. Native heme enzymes catalyze reactions including C-H oxidation and halogenation,[11] and they have been successfully evolved to oxidize abiological substrates.[12,13] Fe-PIX proteins have also been shown to catalyze abiological reactions involving the addition and insertion of carbenes and nitrenes to olefins and X-H bonds.[2,3,9,14] However, the reactivity of the Fe-center in heme proteins limits the scope of these transformations. For example, Fe-PIX proteins catalyze the cyclopropanation of activated terminal vinylarenes,[9,10] but they do not catalyze reactions with internal vinylarenes or unactivated alkenes. Likewise, they catalyze insertions of carbenes into reactive N-H and S-H bonds, but do not catalyze the insertion into less reactive C-H bonds.[3,14]

Because the repertoire of reactions catalyzed by free Ru,[15] Rh,[16] and Ir[17]-porphyrin complexes is much greater than that of the free Fe-analogs, we hypothesized that their incorporation into PIX-proteins could create new functions for PIX-proteins. Artificial PIX-proteins containing Mn, Cr, and Co cofactors have been prepared to mimic the intrinsic chemistry of the native heme proteins,[18-21] but the reactivities and selectivities of these processes are lower than those achieved in the same reactions catalyzed by native Fe-PIX enzymes. Thus, artificially-metallated PIX-proteins that catalyze reactions that are not catalyzed by native Fe-PIX proteins are unknown, and the current, inefficient methods to prepare PIX proteins containing non-native metals have hindered the potential for directed evolution of the resulting enzymes.[22-25]

3.2: Synthesis of Artificially Metallated PIX Proteins

To evaluate rapidly the potential of artificial [M]-PIX enzymes, we envisioned creating an array of catalysts formed by pairing numerous mutants of apo-PIX proteins and [M]-cofactors in a combinatorial fashion. Previously, apo-PIX proteins have been prepared from native Fe-PIX enzymes by acidic, denaturing extraction of the Fe-cofactor, followed by extensive dialysis to refold the protein.[22] This multistep process is too lengthy for directed evolution, and the harsh, acidic conditions are known to result in proteins that are heterogeneous in structure, which would be detrimental for selective catalysis.[26] Alternatively, Ru, Mn and Co-PIX proteins have been expressed directly,[23-25] but these methods are not general, require a gross excess of metal cofactor, and would require a time-consuming purification of each combination of metal and protein. To avoid the aforementioned liabilities of these reported methods in the creation of the proposed catalyst library, we sought to express directly and purify apo-PIX proteins lacking the entire heme unit and to reconstitute them with metal-cofactors containing metals other than iron in a stoichiometric fashion (Fig. 1A).

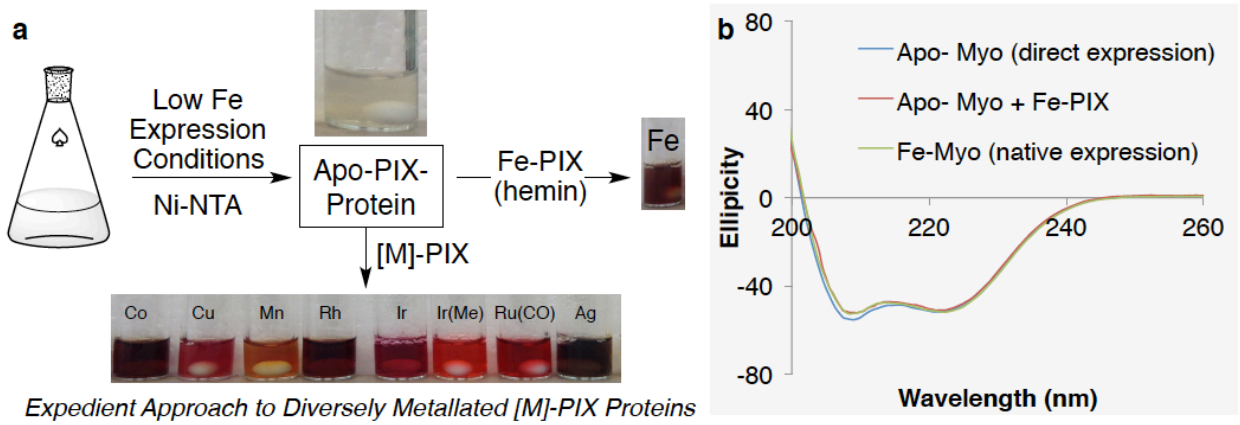


Fig. 1. Strategy for expedient preparation of [M]-PIX-proteins. a. Direct expression, purification, and diverse metallation of apo-PIX proteins. b. Comparison of the CD spectra obtained from directly expressed apo-Myo, the same protein reconstituted with Fe-PIX (hemin), and the same mutant expressed as a native Fe-PIX protein.

Evaluation of a series of expression conditions revealed those suitable for recombinant expression of the apo-form of heme proteins in *E. coli* (Table 1, 2).

	Induction Method	OD at Induction	Induction Time	Induction Temp	Media pH	Isolated Yield of Protein / L
A	IPTG	0.61	16 h	30 °C	7.2	5 mg
B	Auto	---	24 h	30 °C	7.2	5 mg
C	IPTG	0.82	16 h	30 °C	7.2	3 mg
D	IPTG	0.98	16 h	30 °C	7.2	3 mg
E	IPTG	0.98	16 h	30 °C	8.2	20 mg
F	IPTG	1.4	16 h	15 °C	8.2	16 mg
G	IPTG	1.4	16 h	20 °C	8.2	40 mg
H	IPTG	1.4	16 h	25 °C	8.2	15 mg
I	IPTG	1.4	16 h	30 °C	8.2	18 mg

Table 1. Selected expression conditions evaluated in the optimization of the direct expression of apo-Myo (103C/108C)

Protein	Organism	Vector/ Construction	Sequence (Wild Type Protein)
Myo-globin	*Physeter* macroceph-alus	2BT-6xHis-TEV-Myo	EGDIHMKSSHHHHHHENLYFQSNAVLSEGEWQLVLHVWAKVE ADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASED LKKHGVTVLTALGAILKKKGHHEAELKPLAQSHATKHKIPIK YLEFISEAIIHVLHSRHPGDFGADAQGAMNKALELFRKDIAA KYKELGYQG
mOCR-Myo-globin	*Physeter* macroceph-alus	2BT-6xHis-TEV-mOCR-Myo	EGDIHMKSSHHHHHHENLYFQSNMSNMTYNNVFDHAYEMLKE NIRYDDIRDTDDLHDAIHMAADNAVPHYYADIRSVMASEGID LEFEDSGLMPDTKDDIRILQARIYEQLTIDLWEDAEDLLNEY LEEVEEYEEDEEGTGSETPGTSESGVLSEGEWQLVLHVWAKV EADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASE DLKKHGVTVLTALGAILKKKGHHEAELKPLAQSHATKHKIPI KYLEFISEAIIHVLHSRHPGDFGADAQGAMNKALELFRKDIA AKYKELGYQG
P450 BM3	*Bacillus megaterium*	pcWori P450-BM3-6xHis	MTIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADELGEIFK FEAPGRVTRYLSSQRLIKEACDESRFDKNLSQALKFARDFAG DGLVTSWTHEKNWKKAHNILLPSFSQQAMKGYHAMMVDIAVQ LVQKWERLNADEHIEVSEDMTRLTLDTIGLCGFNYRFNSFYR DQPHPFIISMVRALDEVMNKLQRANPDDPAYDENKRQFQEDI KVMNDLVDKIIADRKARGEQSDDLLTQMLNGKDPETGEPLDD GNIRYQIITFLIAGHEATSGLLSFALYFLVKNPHVLQKVAEE AARVLVDPVPSYKQVKQLKYVGMVLNEALRLWPTAPAFSLYA KEDTVLGGEYPLEKGDEVMVLIPQLHRDKTVWGDDVEEFRPE RFENPSAIPQHAFKPFGNGQRASIGQQFALHEATLVLGMMLK HFDFEDHTNYELDIKETLTLKPKGFVVKAKSKKIPLGGIPSP STHHHHHH

Table 2. Sequences and vector constructions of recombinant proteins expressed in this study

Under the optimized conditions, using minimal media lacking Fe to hinder the biosynthesis of hemin and low temperature to mitigate the instability of the apo-form, we expressed successfully the protein containing less than 5% of the Fe-PIX cofactor (Fig 2.), as determined by ICP-OES

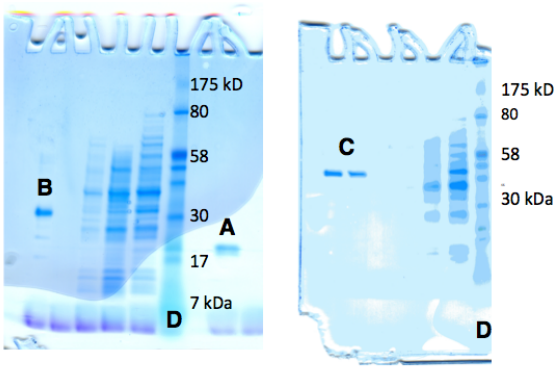


Fig. 2 Characterization of purified apo myoglobin (A), mOCR-myoglobin (B), and P411-CIS (C) proteins by SDS-PAGE gel electrophoresis. Protein samples are shown in comparison to a standard protein ladder (D)

In particular, mutants of *Physeter macrocephalus* myoglobin (Myo) and *Bacillus megaterium* cytochrome P450 BM3h (P450) with and without an mOCR stability tag were overexpressed in high yields and purified (up to 70 mg/L of protein; Fig. 1A, Table S1, S2).[9,10,27] CD spectroscopy revealed that these apo-proteins retain the fold of their native Fe-PIX analogs (Fig. 1B, 3).

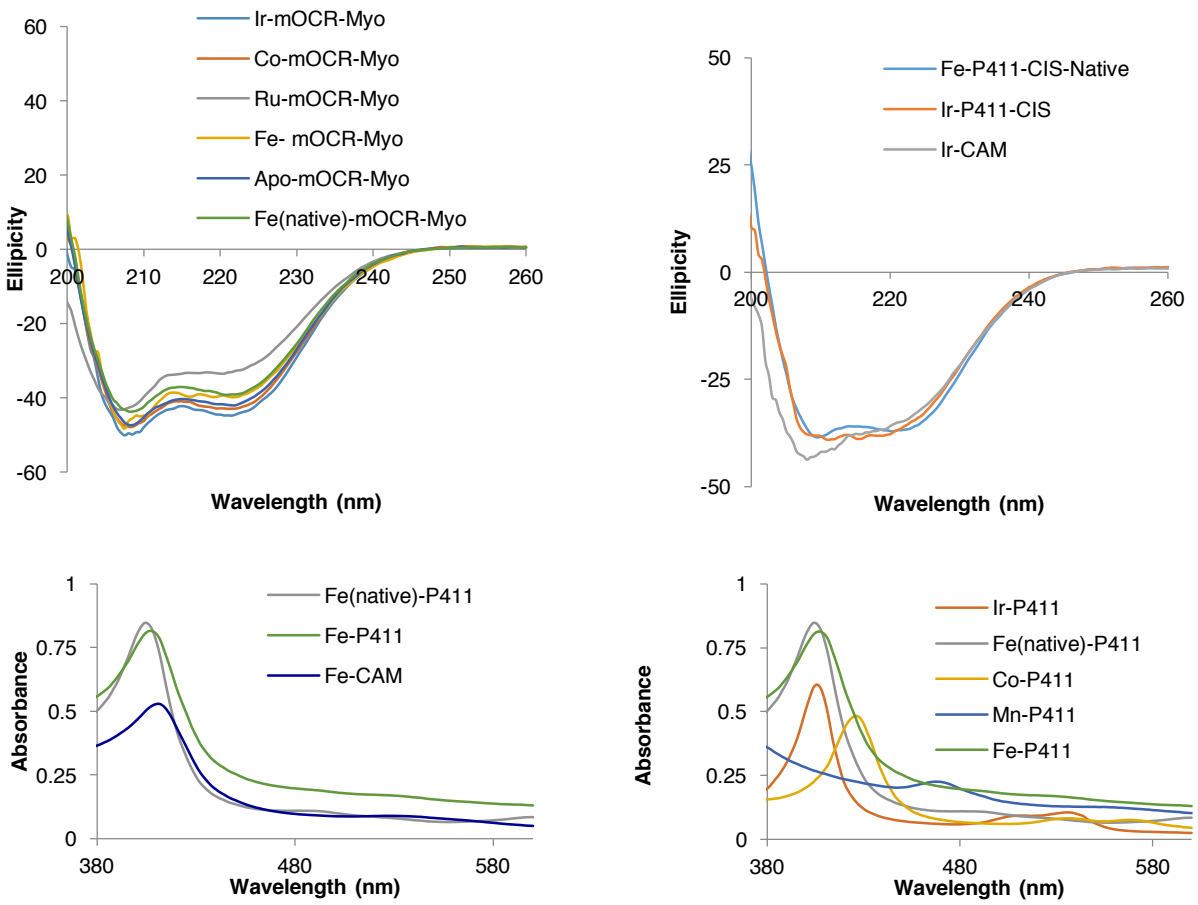


Figure 3. CD (above) and UV spectra (below) of native and artificially metallated myoglobins and P450s. "Fe" proteins contain the native hemin cofactor, while "Ir" protein contain the cofactor Ir(Me)-mesoporphyrin-IX. "Native" Fe proteins were expressed in rich media, such that the E coli

incorporated Fe into the protein during expression. Other Fe proteins were expressed in their apo form followed by reconstitution with hemin post-purification of the apo-protein.

The apo-proteins were reconstituted quantitatively upon addition of stoichiometric amounts of various [M]-PIX cofactors, as determined by native nanoESI-MS (Fig. 4).

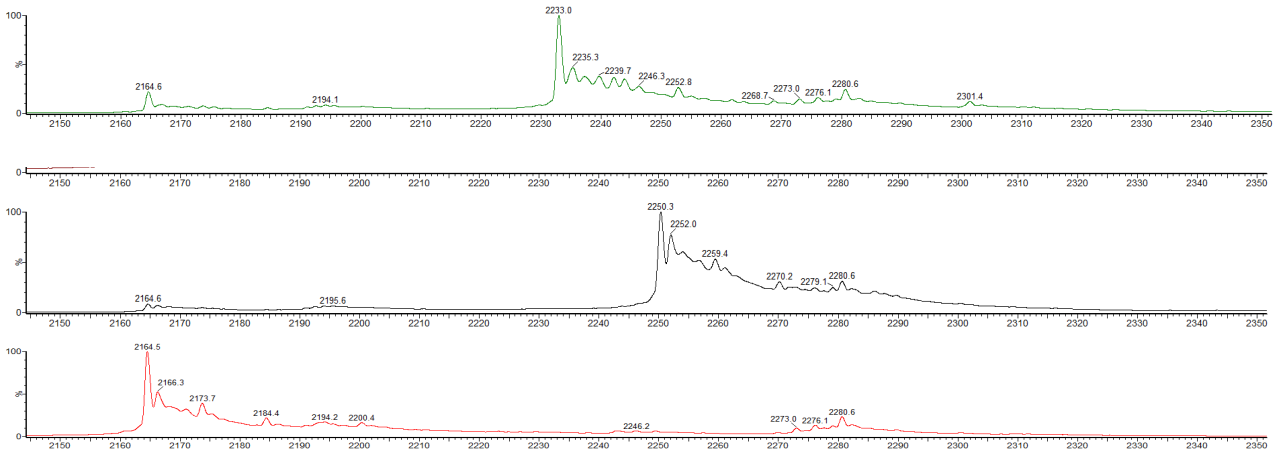


Figure 4. Native NS-ESI-MS showing the binding of hemin (top) and Ir(Me)-PIX (middle) to apo-Myo (bottom) in a 1:1 stoichiometry to the apo protein.

Moreover, reactions catalyzed by reconstituted Fe-myoglobin and Fe-P450 occurred with the same enantioselectivities as those catalyzed by native Fe-proteins (Fig. 5),[9,10] providing strong evidence that this method indeed generates [M]-PIX-proteins with the intact active site and with the cofactors bound at the native PIX-binding site.

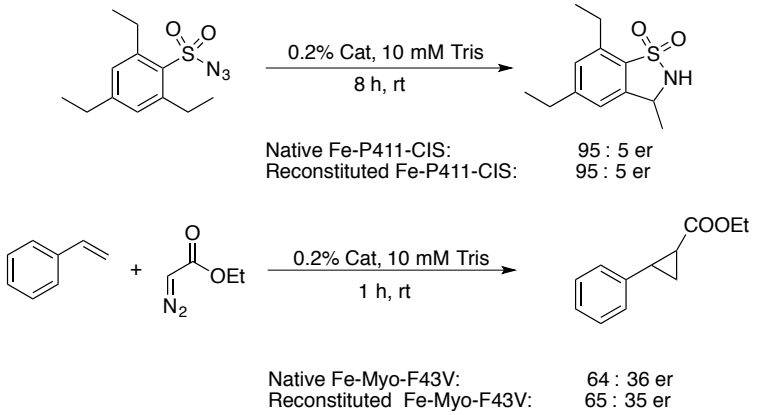
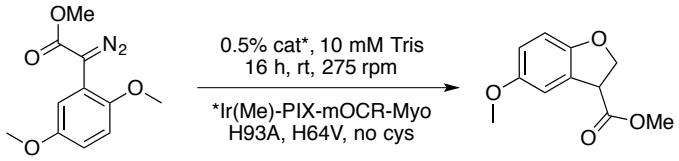


Figure 5. Selectivities of reactions of Fe P450s and Fe-myoglobins in which the protein was prepared by either 1) directly expression the protein in its Fe-form, or 2) expressing the protein in its apo form and reconstituting it with hemin following the purification of the apo-protein.

Further studies revealed that reconstituted mOCR-myoglobins are stable upon storage; reactions catalyzed by freshly prepared, frozen, and lyophilized enzymes proceeded with comparable enantioselectivity (vide infra, Figure 6).



Catalyst Preparation	
Freshly prepared	85:15 er
Frozen (12.5% glycerol), thawed	
Glycerol removed (NAP column)	85:15 er
Glycerol not removed	83:17 er
Lyophilized and redissolved	83:17 er

Fig 6. Comparison of enantioselectivies obtained from reactions catalyzed by Ir(Me)-PIX-Myo prepared and then stored in various ways.

Following this method, we directly expressed eight variants of apo-mOCR-Myo-H93X, each carrying a different mutation to the axial ligand position. Upon reconstitution of each variant with nine different porphyrin cofactors (containing Fe(Cl)-, Co(Cl)-, Cu-, Mn(Cl)-, Rh-, Ir(Cl)-, Ir(Me)-, Ru(CO)- and Ag-sites, Table S2), we rapidly accessed 72 potential catalysts whose activity profiles are distinct from those of wild type myoglobin due to the identity of the metal center and the amino acid residue serving as the axial ligand (Fig. 7).[3]

MUTANT OF MYOGLOBIN

METAL COFACTOR	C-H Insertion	93H	93C	93D	93E	93M	93S	93A	93G
	Fe(Cl)-PIX								
	Co(Cl)-PIX								
	Cu-PIX								
	Mn(Cl)-PIX								
	Rh-PIX								
	Ir(Cl)-PIX								
	Ir(Me)-PIX								
	Ru(CO)-PIX								
	Ag-PIX								

Fig 7. Array of catalysts obtained by combinatorial pairing of nine [M]-PIX cofactors with 8 mutants of myoglobin each containing a different amino acid as the position of the axial ligand.

Cofactor Abbreviation	Metal	Axial Ligand	Porphyrin Ligand	Source
Fe(Cl)-PIX	Iron	Chloride	Protoporphyrin IX	Commercial (Aldrich)
Co(Cl)-PIX	Cobalt	Chloride	Protoporphyrin IX	Commercial (Frontier Scientific)
Cu-PIX	Copper	None	Protoporphyrin IX	Commercial (Frontier Scientific)
Mn(Cl)-PIX	Manganese	Chloride	Protoporphyrin IX	Commercial (Frontier Scientific)
Rh-PIX	Rhodium	None	Protoporphyrin IX	Commercial (Frontier Scientific)
Ir(Cl)-PIX	Iridium	Chloride	Mesoporphyrin IX	Synthetic (Compound)
Ir(Me)-PIX	Iridium	Methyl	Mesoporphyrin IX	Synthetic (Compound)
Ru(CO)-PIX	Ruthenium	Carbon monoxide	Protoporphyrin IX	Commercial (Frontier Scientific)
Ag-PIX	Silver	None	Protoporphyrin IX	Commercial (Frontier Scientific)

Table 3. Structural information for metal cofactors used in this study

3.3 Multi-dimentional Evaluation of the Catalytic Activity of Artificially Metallated Heme Proteins

Natural heme proteins functionalize C-H bonds to form C-O bonds,[11] but no heme protein has functionalized a C-H bond to form a C-C bond. To identify an enzyme for the insertion of a carbene into a C-H bond, the array of artificial [M]-mOCR-myoglobins containing various metals and axial ligands was evaluated for the reaction of diazoester **1** to form chiral dihydrobenzofuran **2** (Fig. 8).

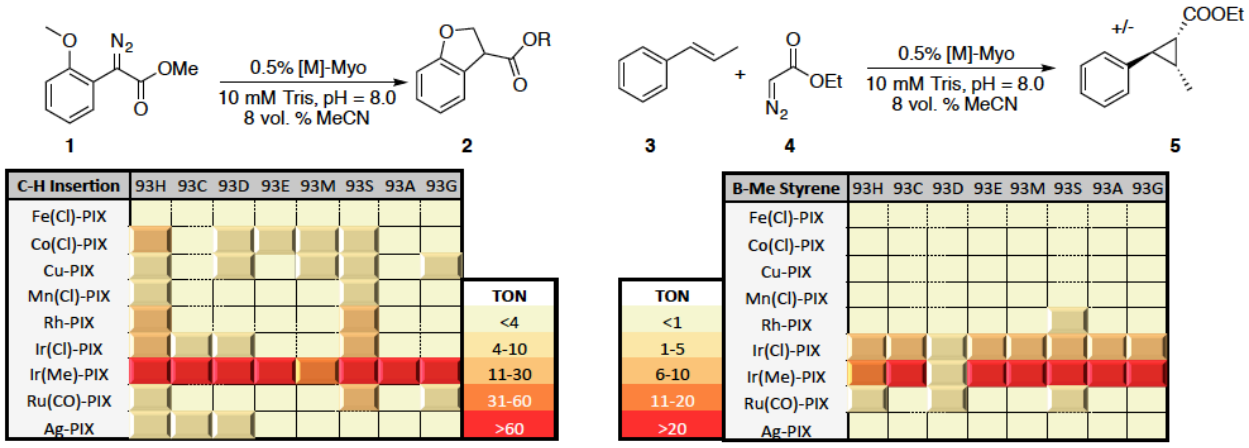


Fig. 8. Evaluation of artificial [M]-PIX mOCR-myoglobins as catalyts for the insertion of carbenes into C-H bonds (left) and the addition of carbenes to an internal olefin (right). TON: turnover number. Reaction conditions for C-H insertion reaction: 10 mM substrate and 0.5% catalyst in 250 uL buffer (10 mM Tris, pH 8.0 containing 8 vol% MeCN). Reaction conditions for

cyclopropanation reaction: 10 mM olefin, 30 mM EDA, and 0.5% catalyst in 250 uL buffer (10 mM Tris, pH 8.0 containing 8 vol% MeCN).

All myoglobins formed from the native Fe-PIX cofactor were inactive, regardless of the axial ligand. In contrast, non-native metals formed active catalysts when paired with an appropriate axial ligand. The most active catalysts, those containing Ir(Me)-PIX, were formed by incorporating both an abiological metal (Ir) and an abiological axial ligand (-CH_3) that cannot be incorporated though standard mutagenesis techniques. The eight myoglobins containing Ir(Me)-PIX formed enantioenriched dibenzohydrofuran **2** in up to 50% yield before any further mutagenesis (*vide infra*). Moreover, this artificial enzyme tolerated modifications to all portions of the substrate; diazoesters **6-11**, containing varied ester, arene, and alkoxy functionalities, (Fig. 9) also underwent C-H insertion in the presence of Ir(Me)-PIX-Myo. Together, these results show that the multi-dimensional evaluation of reconstituted PIX-enzymes can identify new artificial metalloenzymes that catalyze reactions that biological Fe-PIX-proteins do not catalyze.

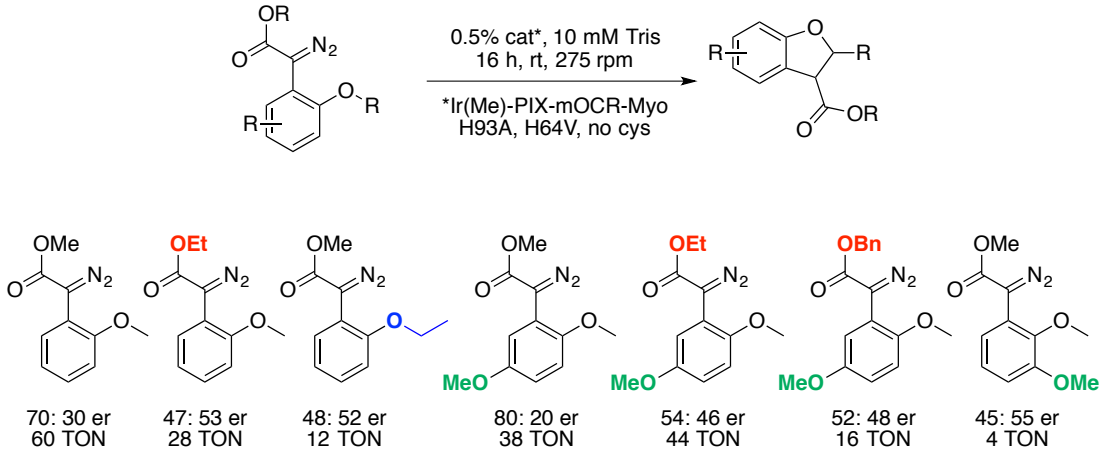


Fig. 9. Comparison of activities and selectivities for C-H insertion reactions catalyzed by the same Ir(Me)-PIX-mOCR-Myo mutant (H93A/H64V) for substrates varied at the arene, ester, and alkoxy-functionalities. TON determined by GC.

3.4: Directed Evolution of Ir(Me)-PIX Myoglobins for Enantioselective Insertions of Carbenes into C-H Bonds

A significant benefit of using enzymes for synthetic applications is the potential to use directed evolution to obtain a catalyst with desired properties.[13,28] To develop an enantioselective catalyst for each of the seven substrates undergoing C-H insertion, we followed a hybrid strategy based on stepwise optimization of small sets of amino acids progressively more distal from the reaction site (Fig. 10). In the first phase, the axial ligand (H93A,G) and the residue directly above the metal center (H64A,V,L,I) were modified (Fig. 4) to give an initial set of eight mutants. In the second phase, these initial eight mutants were modified at positions F43 and V68, which are located in the binding site (Fig. 4), to generate 225 prospective enzymes. To retain the hydrophobicity of the site that binds the porphyrin and substrate, only hydrophobic and uncharged residues (V, A, G, F, Y, S, T) were introduced at positions F43 and V68. From this

second phase of the evolution, discrete mutants were identified to provide either enantiomer of the product arising from each substrate with moderate to high enantioselectivity (Table 4).

Substrate	Mutant Providing (-) enantiomer					Mutant Providing (+) enantiomer				
	H93X	H64X	F43X	V68X	er (TON)	H93X	H64X	F43X	V68X	er (TON)
1	A	V	--	F	79:21 (42)	G	L	--	A	23:77 (50)
6	G	L	L	A	80:20 (68)	G	L	V	G	31:69 (46)
7	G	I	V	--	84:16 (62)	A	L	Y	G	31:69 (36)
8	A	V	--	A	90:10 (98)[a]	A	L	W	T	23:77 (164)
9	G	L	Y	--	84:16 (100)	A	V	I	T	25:75 (138)
10	A	V	H	S	77:23 (44)	G	L	L	G	33:67 (42)
11	A	A	W	G	90:10 (52)	G	L	--	A	20:80 (10)

Table 4. Most selective mutants obtained for substrates for C-H insertion in the second phase of directed evolution. [a]Contains 103C, 107C as additional mutations

Of these 225 mutants, 22 that were among the most selective for one or more of the substrates in Figure 3 were subjected to a further round of evolution during which the residues at four additional positions (L32, F33, H97, and I99) were modified to generate 217 more mutants (Fig. 4), from which mutants reacting with higher enantioselectivity than were previously achieved were identified (Table 5).

Substrate	Mutations Providing (-) enantiomer				Mutations Providing (+) enantiomer			
	H93/H64	F43/V68	Additional Mutations	er (TON)	H93/H64	F43/V68	Additional Mutations	er (TON)
1	93A, 64L	43L	33V	85:15 (134)[a]	93G, 64L	68A	32F, 99V	16:84 (62)
6	93A, 64V	43Y, 68A	97W	81:19 (128)	93G, 64L	68A	99V	20:80 (26)[a]
7	93G, 64L	43L	99F	92:8 (92)	93A, 64L	68A, 43W	97Y	25:75 (68)
8	93A, 64V	68A	103C, 108C	90:10 (98)	93A, 64L	43W, 68T	--	23:77 (164)
9	93A, 64L	43W, 68A	33I	85:15 (86)	93A, 64V	43I, 68T	--	25:75 (138)
10	93A, 64V	43H, 68S	--	77:23 (44)	93A, 64V	68A	33V, 97Y	17:83 (42)
11	93A, 64A	43W, 68G	--	90:10 (52) 90:10 (160)[b] 84:16 (7260)[c]	93G, 64L	68A	32F, 97Y	13:87 (144)

Table 5. Most selective mutants obtained for substrates for C-H insertion in the directed evolution of Ir(Me)-PIX-mOCR-myo as a catalyst. [a] Reaction performed under modified conditions: 10 mM NaPi, pH = 7.5, no inhibitor. [b] Reaction on a synthetic scale (Section V.g): 0.12 mmol of substrate,

no inhibitor, TON based on the isolated yield of the product (80%). [c] Reaction performed under modified conditions: [substrate] = 0.2 M, [catalyst] = 0.005 mM, 72 h, no inhibitor

The complete results of the carbene insertion reaction with these mutants are provided in 3.8 and are summarized Table 5 and Figure 10-12. The directed evolution of Ir(Me)-myoglobins uncovered distinct enzymes catalyzing the C-H functionalization to form either enantiomer of the products containing a new C-C bond formed from substrates **1** and **6-11** (Figure 10). The reactions occurred with selectivities up to 92:8 er and with yields up to 97% with enzymes that were evolved from those giving nearly racemic product (Figure 11). The Ir-myoglobins are suitable catalysts for synthetic-scale reactions; the carbene insertion of substrate **11** formed the product containing a new C–C bond in 80% isolated yield from a reaction of 28 mg of **11** with nearly the same enantioselectivity as observed on smaller scale (Fig. 3). A reaction conducted with a 40,000:1 ratio of substrate to Ir(Me)-mOCR-myo occurred with 7,200 turnovers (Fig. 3).

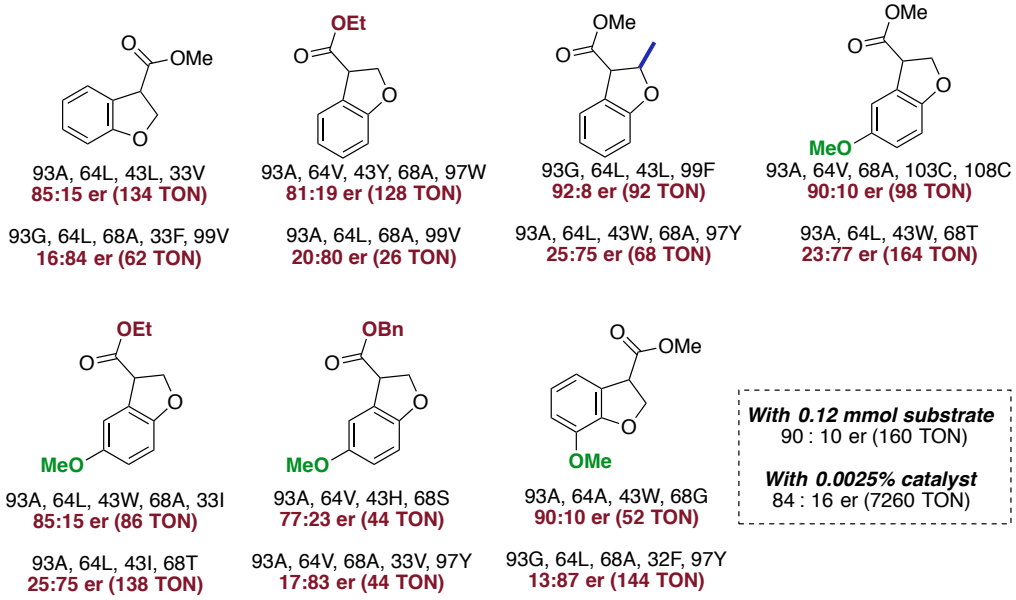
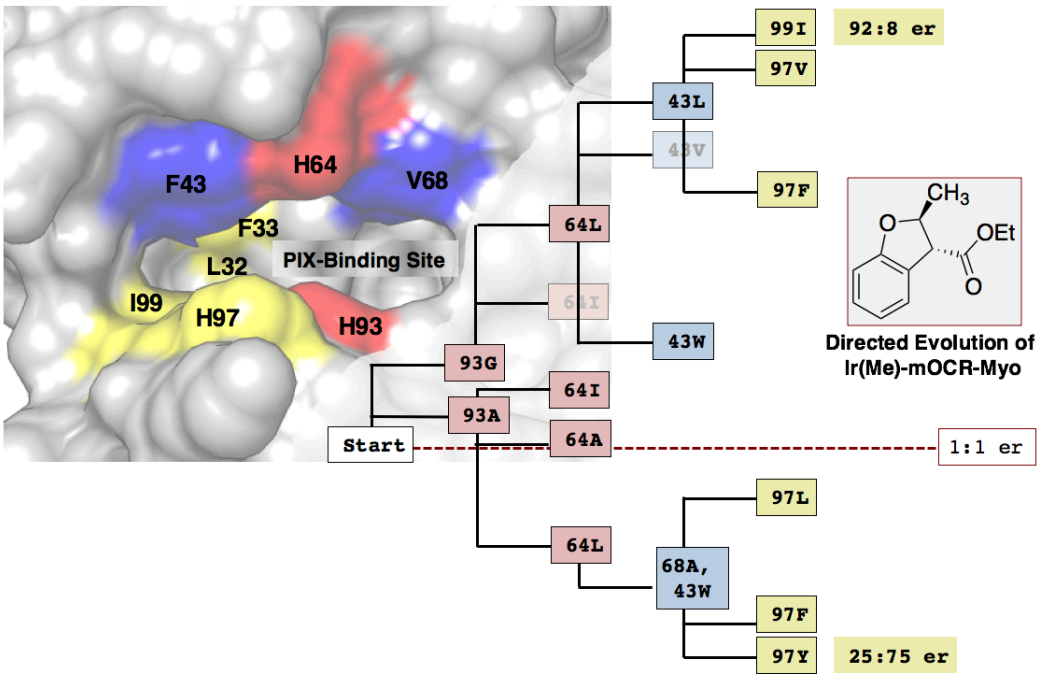
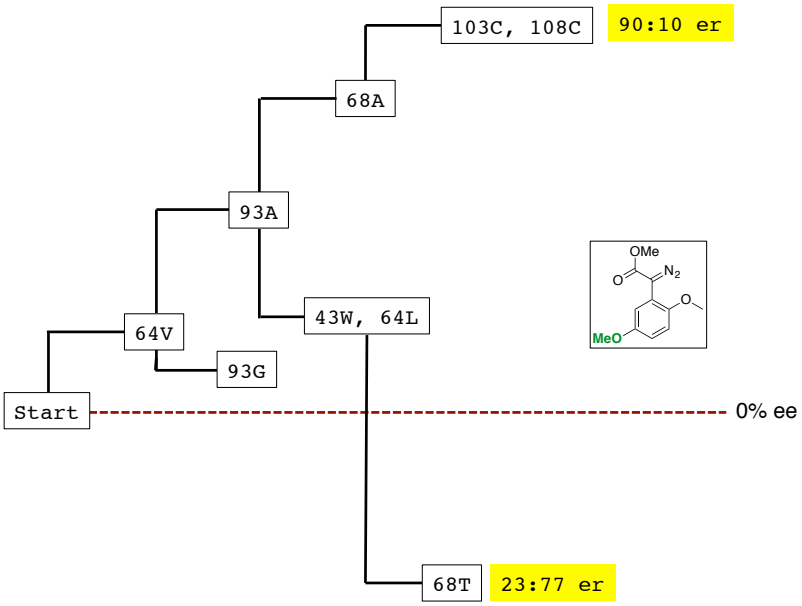
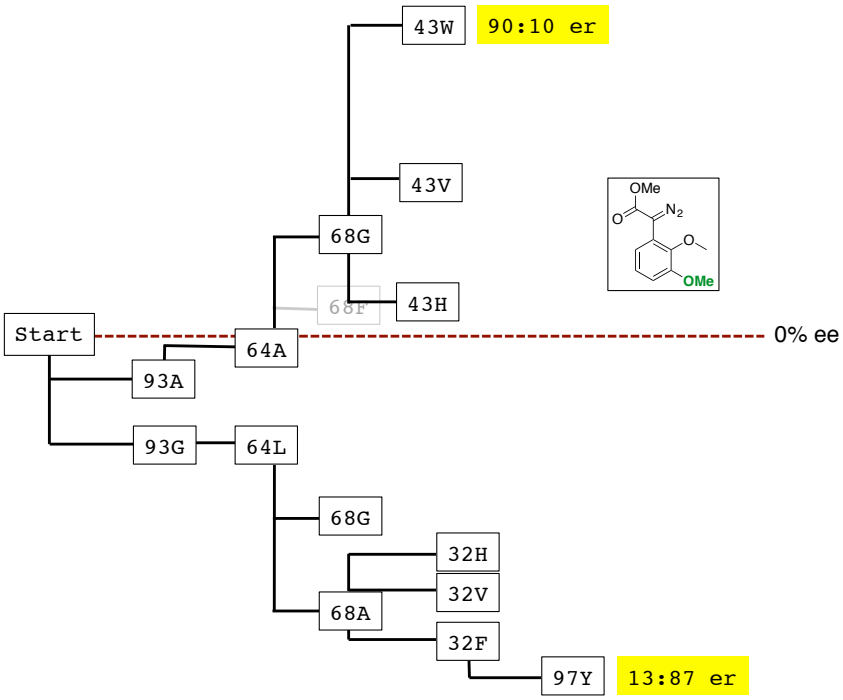


Fig. 10: Most selective mutants identified to provide either enantiomer of the product from a series of additional substrates.





64X 93X 68X, 43X Further Positions


43W
90:10 er
43V
68G
68F
43H
Start
64A
0% ee
93A
93G
64L
68G
32H
32V
68A
32F
97Y
13:87 er
OMe
OMe
93X
64X
68X, 43X
32/33/97/99X

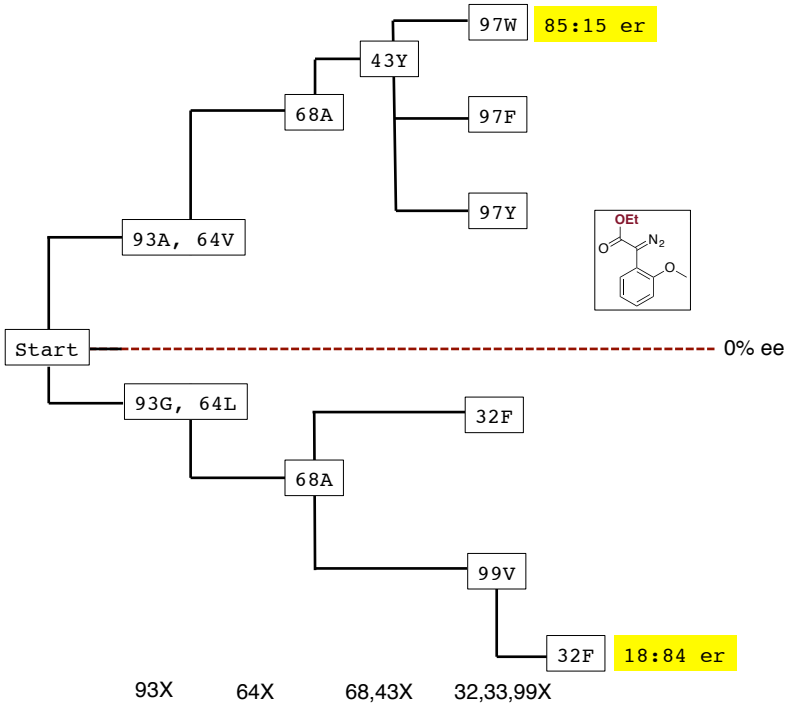

97W
85:15 er
43Y
68A
97F
97Y
93A, 64V
Start
0% ee
93G, 64L
32F
68A
99V
32F
18:84 er
OEt
93X
64X
68,43X
32,33,99X

Fig. 11. Directed evolution strategy used to obtain Ir(Me)-mOCR-Myo mutants capable of producing either enantiomer of the products of C-H insertion reactions of varied substrates. Top: Inner sphere (red), middle sphere (blue), and outer sphere (yellow) residues are highlighted in the depiction of the active site and in the evolutionary tree (Image produced in Chimera from PDB: 1MBN[29]). Bottom: Analogous evolutionary trajectories used to optimize the selectivities of reactions of other substrates.

In contrast to the few directed evolutions of artificial enzymes reported previously,[28] our method to prepare variants of the Ir(Me)-PIX-enzyme enabled us to pursue an individual, eight-site evolutionary trajectory for the reaction of each substrate that identified catalysts selectively forming either enantiomer of all targeted products. These results demonstrate that Ir(Me)-PIX-myoglobins are highly evolvable for different substrates containing varied structural modifications. These results, along with the high isolated yield and the observation of high turnovers, demonstrate that the direct expression of apo-Myo, the insertion of diverse [M]-PIX cofactors, and the subsequent directed evolution of the most active enzymes identified is a robust strategy that can be applied in a general way to create stereoselective catalysts for reactions that cannot be accomplished by any natural enzymes.

3.6 Directed Evolution of Ir(Me)-PIX Myoglobins for Enantio- and Diastereoselective Additions of Carbenes to Internal and Aliphatic Alkenes

To assess the generality of this approach further, we sought catalysts for the cyclopropanation of internal alkenes and alpha-olefins that, like carbene insertion into C-H bonds, have not been accomplished with natural or artificial enzymes. As a starting point, we evaluated the 2D array of [M]-PIX catalysts shown in Figure 7 for the cyclopropanation of β-methylstyrene **3** with ethyl diazoacetate **4** (EDA). In agreement with literature reports,[10] Fe-PIX enzymes did not catalyze this reaction (Fig. 8). In contrast, Rh-, Ru-, and Ir-PIX enzymes furnished the cyclopropane product **5**. The enzyme containing Ir(Me)-PIX was the most active (Fig. 8). Although further work is needed to obtain full conversion and high ee, the reaction of EDA with β-methylstyrene catalyzed by the Ir(Me)-mOCR-Myo mutant H93A, H64V, F43Y, V68A, H97F formed the cyclopropane **5** with 40 turnovers with 40% ee, and with a high >33:1 ratio of diastereomers, favoring the trans isomer (Fig. 12).

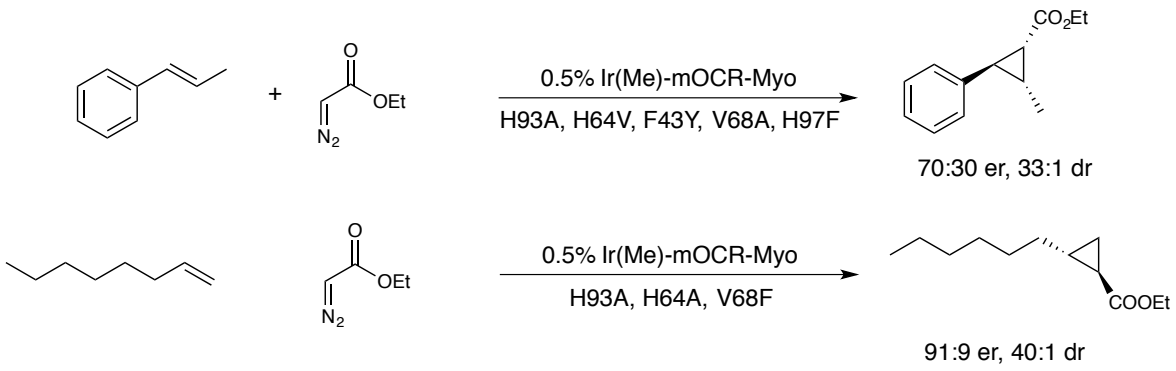


Fig. 12. Enantio- and diastereoselective cyclopropanation of alkenes catalyzed by variants of Ir(Me)-PIX-mOCR-Myo. Reactions were conducted using 6 equiv. of EDA added over 1 hour using a syringe pump.

Having observed the expanded scope of enzyme-catalyzed cyclopropanation, we assessed the ability of Ir(Me)-PIX enzymes to catalyze the cyclopropanation of 1-octene, an unactivated, aliphatic olefin. The series of Ir(Me)-PIX-Myo enzymes assessed for C-H insertion reactions were tested as catalysts for the reaction of EDA with 1-octene. Although unselective for C-H insertion, the mutant H93A, H64A, V68F, formed the cyclopropane product in 91:9 er and 40:1 trans:cis ratio (Fig. 12). Cyclopropanations of aliphatic alkenes are typically conducted with small-molecule catalysts with excess of the alkene.[30] In contrast, the Ir(Me)-PIX-Myo mutant catalyzes the reaction with excess of EDA (42 TON with 10:1 ratio of EDA : 1-octene), suggesting that reactions can be developed with valuable alkenes as limiting reagent. The reactions with fewer equivalents of EDA occur with fewer TON due to consumption of EDA by dimerization or O-H insertion of water. These cyclopropanations of unactivated alkenes show the broad potential to evolve artificial myoglobins containing abiological active sites for reactions that are not catalyzed by enzymes containing native metals.

3.7 Conclusions and Outlook

The work presented here demonstrates that unknown enzymatic reactivity can be achieved by incorporating just a metal ion with an accompanying small ligand into a well-known metalloprotein, while retaining the native structure of the active site. Selectivity for specific substrates, then, can be achieved readily by directed evolution. Considering the rich chemistry of free metalloporphyrins and the ease of preparation and evolution of heme proteins containing diverse metals by the methods just described, this methodology should seed the creation of many new artificial metalloenzymes with diverse, unnatural reactivity. Moreover, the facile, direct expression of apo-heme proteins can be used in tandem with strategies to incorporate highly active noble metal complexes of ligands beyond porphyrins. Access to such a range of artificial heme proteins provides a nearly limitless opportunity to achieve catalytic reactions with selectivity derived from the interaction of the substrate with a natural, evolvable binding site.

3.8 Supporting Information:

3.8.1 Protein Expression, Purification, and Characterization

a. General Methods

Unless otherwise noted, the chemicals, salts, and solvents used were reagent grade and used as received from commercial suppliers without further purification. All expression media and buffers were prepared using ddH_2O (MilliQ A10 Advantage purification system, Millipore). All expression media were sterilized using either an autoclave (45 min, 121^oC) or a sterile syringe filter (0.22 um). To maintain sterile conditions, sterile materials and *E. coli* cells were manipulated near a lit Bunsen burner. TEV protease was obtained from the UC Berkeley Macrolab.

b. Cloning

A double mutant of the gene encoding sperm whale myoglobin (Y103C, S108C) was purchased from GenScript with codon optimization for *E. coli* (Table S2). The gene was cloned into the expression vectors 2-BT (6xHis-TEV-ORF; AddGene #29666) and 2OT (6xHis-TEV-mOCR-ORF; AddGene 29710) at the QB3 Macrolab at UC Berkeley.

c. Media Preparation

Preparation of optimized minimal expression media: Salts (15 g Na_2HPO_4, 7.5 g K_2HPO_4, 0.3 g NaH_2PO_4, 0.3 g KH_2PO_4, 1.5 g NaCl, 5 g NH_4Cl) were dissolved in 2 L ddH_2O and autoclaved to give a media with pH ~8.0 - 8.2. Solutions of glucose (20%), casamino acids (BD Company, low Fe, 20%), and $MgSO_4$ (1 M), were autoclaved separately. Solutions of ampicillin (100 mg/ mL) and $CaCl_2$ (1 M) were sterilized by syringe filter. The following amounts of the listed solutions were added per 2 L of sterile salt solution: 40 mL glucose, 20 mL casamino acids, 4 mL $MgSO_4$, 100 uL $CaCl_2$, 2 mL ampicillin. Stock solutions were stored for several weeks; prepared media was stored for less than 1 day. Minimal media plates were prepared from the same media with the addition of 17 g agar/L media. In this case, agar was autoclaved in 1 L ddH_2O, and salts were autoclaved separately as a 20X solution, after which they were added to the agar solution.

Preparation of rich expression media: LB media (10 g tryptone, 5 g yeast, 10 g NaCl, 1 L ddH_2O) was autoclaved and supplemented with $MgSO_4$ (2 mL, 1 M), glucose (20 mL, 20%), ampicillin (1 mL, 100 mg/mL), and M9 salts (50 mL, 20X). Minimal media plates were prepared from the same media with the addition of 17 g agar/L media.

d. Mutagenesis

Site-directed mutagenesis was performed using the QuickChange Lightening mutagenesis kit (Agilent); requisite double stranded DNA primers were designed according to the Agilent Primer Design Program and purchased from Integrated DNA Technology. PCR reactions were performed according to the manufacturer's directions. PCR reactions contained: 5 uL reaction buffer, 34 uL

ddH_2O, 1.5 uL QuickSolution, 1 uL plamids (50 ng/uL), 1.25 uL sense primer (100 ng/uL), 1.25 uL antisense primer (100 ng/ uL), 5 uL dNTPs (2 mM/base), and 1 uL polymerase.

PCR Program: Phase 1 (1 cycle): 95 °C, 1.5 min; Phase 2 (18 cycles): 95 °C, 20 sec, 60 °C, 10 sec, 68 °C, 3 min; Phase 3 (1 cycle): 68 °C, 3 min; Phase 4 (storage): 4 °C

Plasmid Preparation for Phase 1 of Directed Evolution: The general PCR method was used with degenerate primers encoding H93A, G and H64A, V, L, I.

Preparation of the Combinatorial Library of Mutants of Positions 43/68 (Phase 2): The general PCR method was used with degenerate primers encoding F43Y, W, L, I, T, H, V and V68A, G, F, Y, S, T.

Preparation of the Combinatorial Library of Mutants of Positions 32/33/97/99 (Phase 3): The general PCR method was used with degenerate primers encoding L32V, I, F, H, W, Y, F33V, L, I, H, W, Y, H97V, L, I, F, W, Y, and I99V, L, F, H, W, Y.

DNA Isolation and Storage: Following the completion of the above set of PCR procedures, 1.5 uL DPN 1 was added to each reaction, and the reactions were further incubated (3 h, 37°C). The crude PCR mixture was used to transform XL-10 Gold Ultracompetent cells (45 L cells, 2 L PCR reactions). The mixture was incubated on ice (30 min), heat shocked (30 s, 42 °C), recovered with SOC media (1 h, 37 °C, 275 rpm), and plated on LB plates. Plates were grown (18 h, 37 °C), and individual colonies were used to inoculate 1 mL rich media cultures, which were grown in 96-well plates (13 h, 37 °C, 300 rpm). DNA was isolated from the 96-well cultures using magnetic bead technology at the UC Berkeley DNA Sequencing Facility. Alternatively, individual colonies were used to inoculate 4 mL rich media cultures and grown overnight (13 h, 37 °C, 300 rpm), and the plasmids were purified using a Qiagen DNA Miniprep kit according to the manufacturer's instructions.

e. Protein Expression

Optimized Expression of Apo Myoglobin: BL21 Star competent *E.* coli cells (50 L, QB3 Macrolab, UC Berkeley) were thawed on ice, transferred to 14 mL Falcon tubes, and transformed with the desired plasmid solution (2 L, 50-250 ng/ L). The cells were incubated on ice (30 min), heat shocked (20 sec, 42°C), re-cooled on ice (2 min), and recovered with SOC media (37 °C, 1 h, 250 rpm). Aliquots of the cultures were diluted (0.02X), plated on minimal media plates (expression media supplemented with 17 g agar/L), and incubated (20 h, 37 °C) to produce approximately 10-100 colonies per plate. Single colonies were used to inoculate starter cultures (3 mL, expression media), which were grown (6-8 hours, 37°C, 275 rpm) and used to inoculate 100 mL overnight cultures (minimal media, 37° C, 275 rpm). Each overnight culture was used to inoculate 750 mL of minimal media, which was further grown (9 h, 37°C, 275 rpm). Expression

was induced with IPTG (800 uL, 1M), and the cultures were further grown (15 h, 20° C, 275 rpm). Cells were harvested by centrifugation (5000 rpm, 15 min, 4° C), and the pellets were resuspended in 20 mL Ni-NTA lysis buffer (50 mM NaPi, 250 mM NaCl, 10 mM Imidazole, pH = 8.0) and stored at -80° C until purification.

Expression of Fe Myoglobin: Fe-myo was expressed following a procedure modified from the literature[10]. BL21 Star competent *E.* coli cells (50 uL, QB3 Macrolab, UC Berkeley) were thawed on ice, transferred to 14 mL Falcon tubes, and transformed with the desired plasmid solution (2 uL, 50-250 ng/uL). The cells were incubated on ice (30 min), heat shocked (20 sec, 42°C), re-cooled on ice (2 min), and recovered with SOC media (37°C, 1 h, 250 rpm). Aliquots of the cultures were diluted (0.02X), plated on minimal media (expression media supplemented with 17g agar/L), and incubated (20 h, 37°C) to produce approximately 10-100 colonies per plate. Single colonies were used to inoculate starter cultures (3 mL, expression media), which were grown (6-8 hours, 37°C, 275 rpm) and used to inoculate 100 mL overnight cultures (minimal media, 37° C, 275 rpm). A 5 mL portion of the overnight culture was used to inoculate 750 mL of LB media supplemented with 1X M9 salts, $MgSO_4$ (1.5 mL, 1M) and glucose (15 mL, 20% glucose), which was further grown (9 h, 37°C, 275 rpm). Expression was induced with IPTG (800 uL, 1M) and ALA (0.8 mL, 0.3 M) was added, and the cultures were further grown (15 h, 25° C, 250 rpm). Cells were harvested by centrifugation (5000 rpm, 15 min, 4° C), and the pellets were resuspended in 20 mL Ni-NTA lysis buffer (50 mM NaPi, 250 mM NaCl, 10 mM Imidazole, pH = 8.0) and stored at -80° C until purification.

Protein Purification: Cell suspensions were thawed in a room temperature ice bath, decanted to 50 mL glass beakers, and lysed on ice by sonication (3x30 sec on, 2x2 min off, 60% power). Crude lysates were transferred to 50 mL Falcon tubes, treated with triton X (100 uL, 2% in H_2O), and incubated on an end-over-end shaker (30 min, rt, 15 rpm). Cell debris was removed by centrifugation (10 000 rpm, 60 min, 4° C), and Ni-NTA (5 mL, 50% suspension per 850 mL cell culture) was added. The lysates were briefly incubated with Ni-NTA (30 min, 4° C, 20 rpm) and poured into glass frits (coarse, 50 mL). The resin was washed with Ni-NTA lysis buffer (3 x 35 mL), and the wash fractions were monitored using Bradford assay dye. The desired protein was eluded with 18 mL Ni-NTA elusion buffer (50 mM NaPi, 250 mM NaCl, 250 mM Imidazole, pH = 8.0), dialyzed against Tris buffer (10 mM, pH = 8.0, 12 h, 4° C), concentrated to the desired concentration using a spin concentrator, and metallated within several hours. Apo protein was not stored for more than 8 hours.

Metallation of the Apo-Protein (General Method): Stock solutions of metal cofactors in DMF were added to solutions of apo protein (0.3-0.6 mM) at room temperature in the desired stoichiometry with a final DMF concentration of 2%. The proteins were briefly incubated at room temperature (5 minutes), and DMF was removed by using a NAP column, according to manufactures instructions.

Protein Storage: Glycerol was added to a solution of the protein (3:1 v:v protein solution: 50% glycerol), the solution was divided into 1.5 mL eppendorf tubes (0.5 mL per aliquot), and the tubes

were flash frozen in liquid nitrogen and stored at -80 ° C until further use. Frozen aliquots of the protein were thawed in a room-temperature water bath.

f. Protein Characterization

Gel Electrophoresis: Protein purity was analyzed by sodium dodecyl sulfate-polyacrylamide (SDS-PAGE) gel electrophoresis using precast gels (polyacrylamide, 10-20% linear gradient, Biorad).

Mass Spectrometry: Apo-proteins were analyzed with an Agilent 1200 series liquid chromatograph connected in-line with an Agilent 6224 time-of-flight (TOF) LC/MS system using a Turbospray ion source. Metallated proteins were analyzed by native nanoelectrospray ionization mass spectrometry (nanoESI-MS) using a Waters Q-Tof Premier quadrupole time-of-flight mass spectrometer equipped with a nanoESI source (Milford, MA). Mass spectra were acquired in the positive ion mode and processed using MassLynx software (version 4.1, Waters). The instrument is located in the QB3/Chemistry Mass Spectrometry Facility at UC Berkeley.

UV-Vis Spectroscopy: Protein concentration was determined using a NanoDrop 2000 UV-Vis Spectrophotometer (Thermo Scientific). UV-Vis spectra of proteins were measured using a SpectraMax microplate reader (Molecular Devices) configured in the 96-well plate format.

CD Spectroscopy: CD spectra were acquired on a JASCO J-815 instrument in 3 mL quartz cuvettes using 2 mL of the appropriate protein solution (0.03 mg/ mL).

3.8.2 Organic Synthesis and Characterization

a. General methods and materials

Unless stated otherwise, all reactions and manipulations were conducted on the laboratory bench in air with reagent grade solvents. Reactions under inert gas atmosphere were carried out in the oven dried glassware in a nitrogen-filled glovebox or by standard Schlenk techniques under nitrogen.

NMR spectra were acquired on 400 MHz, 500 MHz, 600 MHz, or 900 MHz Bruker instruments at the University of California, Berkeley. NMR spectra were processed with MestReNova 9.0 (Mestrelab Research SL). Chemical shifts are reported in ppm and referenced to residual solvent peaks[31]. Coupling constants are reported in hertz. GC analyses were obtained on an Agilent 6890 GC equipped with either, an HP-5 column (25 m x 0.20 mm ID x 0.33 m film) for achiral analysis or Cyclosil-B column (30m x 0.25mm x 0.25 um film) for chiral analysis, and an FID detector. GC yields were calculated using dodecane as the internal standard and not corrected for response

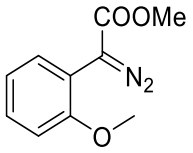
factors of minor isomers. High-resolution mass spectra and elemental analysis were obtained via the Micro-Mass/Analytical Facility operated by the College of Chemistry, University of California, Berkeley.

Unless noted otherwise, all reagents and solvents were purchased from commercial suppliers and used without further purification. If required, dichloromethane (DCM) and tetrahydrofuran (THF) were degassed by purging with argon for 15 minutes and dried with a solvent purification system containing a one-meter column of activated alumina; dried and degassed acetonitrile, 1,2-xylene, toluene, N,N-dimethylformamide (DMF), ethanol and methanol were purchased form commercial suppliers and used as received.

b. Substrates

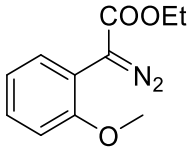
The synthesis of the following substrates was conducted by Pawel Dydio.

Methyl (2-methoxyphenyl)diazoacetate (1):



To a stirred solution of methyl (2-methoxyphenyl)acetate (3.2 ml, 20 mmol) and 4-acetamidobenzenesulfonyl azide (p-ABSA, 7.2 g, 30 mmol) in acetonitrile (40 ml) at 0 °C, 1,8-diazabicycloundec-7-ene (DBU, 4.8 ml, 32 mmol) was added dropwise. The cooling bath was removed, and the reaction was allowed to continue stirring overnight. The reaction mixture was diluted with dichloromethane (~60 ml), washed with water (2 x ~50 ml), dried over $MgSO_4$. After filtration, the volatile material from the filtrate was evaporated under reduced pressure.The crude product was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 $\rightarrow$ 95:5 gradient) as the eluent. Fractions of the pure product were combined, and the solvent evaporated, yielding 3.9 g (95%) of product. The NMR data match those of the reported molecule [32].

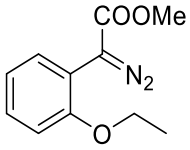
Ethyl (2-methoxyphenyl)diazoacetate (6):



In a closed vial, a solution of (2-methoxyphenyl)acetic acid (3.3 g, 20 mmol) in ethanol (20 ml), containing several drops of sulfuric acid, was stirred overnight at 80 °C. The volatile materials were evaporated under vacuum. The residue was dissolved in ethyl acetate (~40 ml), washed with $NaHCO_3$ sat. (40 ml) and water (40 ml), and dried over $MgSO_4$. After filtration, the volatile material from the filtrate was evaporated under reduced pressure. The resulting crude product was used in the next step without further purification.

To a stirred solution of ethyl (2-methoxyphenyl)acetate and 4-acetamidobenzenesulfonyl azide (p-ABSA, 7.2 g, 30 mmol) in acetonitrile (40 ml) at 0 °C, 1,8-diazabicycloundec-7-ene (DBU, 4.8 ml, 32 mmol) was added dropwise. The cooling bath was removed, and the reaction was allowed to continue stirring overnight. The reaction mixture was diluted with dichloromethane (~60 ml), washed with water (2 x ~50 ml), and dried over $MgSO_4$. After filtration, the volatile material from the filtrate was evaporated under reduced pressure. The crude product was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 95:5 gradient) as the eluent. Fractions of the pure product were combined, and the solvent evaporated, yielding 2.15 g (49%) of product. The NMR data match those of the reported molecule [33].

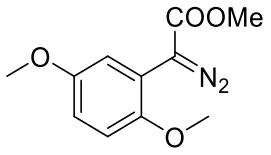
Methyl (2-ethoxyphenyl)diazoacetate (7):



To a solution of (2-ethoxyphenyl)diazoacetate acid (3.8 g, 20 mmol) in toluene (40 ml) and methanol (20 ml), a solution of trimethysilyldiazomethane in diethyl ether (15 ml, 2 M, 30 mmol) was added dropwise while stirring, and stirring was continued for 2 h. Upon evaporation of the volatile materials under vacuum, the product was obtained in quantitative yield without the need for further purification.

To a stirred solution of methyl (2-ethoxyphenyl)acetate (20 mmol) and 4-acetamidobenzenesulfonyl azide (p-ABSA, 7.2 g, 30 mmol) in acetonitrile (40 ml) at 0 °C, 1,8-diazabicycloundec-7-ene (DBU, 4.8 ml, 32 mmol) was added dropwise. The cooling bath was removed, and the reaction was allowed to continue stirring overnight. The reaction mixture was diluted with dichloromethane (~60 ml), washed with water (2 x ~50 ml), dried over $MgSO_4$ and evaporated. The crude product was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 95:5 gradient) as the eluent. Fractions of the pure product were combined, and the solvent evaporated, yielding 2.6 g (59%) of product. The NMR data match those of the reported molecule [33].

Methyl (2,5-dimethoxyphenyl)diazoacetate (8):



In a closed vial, a solution of (2,5-dimethoxyphenyl)acetic acid (3.8 g, 20 mmol) in methanol (20 ml) containing several drops of sulfuric acid, was stirred overnight at 80 °C. The volatile materials were evaporated under vacuum. The residue was dissolved in ethyl acetate (~40 ml), washed with

$NaHCO_3$ sat. (40 ml) and water (40 ml), dried over $MgSO_4$ and evaporated. The crude product was used in the next step without further purification.

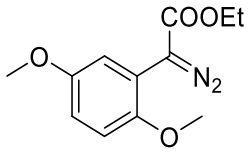
To a stirred solution of methyl (2,5-dimethoxyphenyl)acetate and 4-acetamidobenzenesulfonyl azide (p-ABSA, 7.2 g, 30 mmol) in acetonitrile (40 ml) at 0 °C, 1,8-diazabicycloundec-7-ene (DBU, 4.8 ml, 32 mmol) was added dropwise. The cooling bath was removed, and the reaction was allowed to continue stirring for 48 h (the reaction progress was followed by TLC). The reaction mixture was diluted with dichloromethane (~60 ml), washed with water (2 x ~50 ml), and dried over $MgSO_4$. After filtration, the volatile material from the filtrate was evaporated under reduced pressure. The crude product was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 95:5 gradient) as the eluent. Fractions of the pure product were combined, and the solvent evaporated, yielding 4.1 g (87%) of product.

^{1}H NMR (900 MHz, $CDCl_3$): d = 7.14 (d, J = 3.0 Hz, 1H), 6.78 (d, J = 8.9 Hz, 1H), 6.75 (dd, J = 8.9 Hz, J = 3.0 Hz, 1H), 3.80 (s, 3H), 3.76 (s, 3H), 3.74 (s, 3H);

^{13}C NMR (225 MHz, $CDCl_3$): d = 166.7, 154.1, 149.8, 115.2, 114.7, 114.0, 112.3, 56.4, 56.0, 52.2 ($C=N_2$ signal missing, as observed before for related molecules (*4*));

HR MS (EI): calcd. for $C_{11}H_{12}N_2O_4$ $[M]^+$: 236.0797, found: 236.0801.

Ethyl (2,5-dimethoxyphenyl)diazoacetate (9):



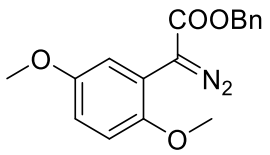
To a stirred solution of ethyl (2,5-dimethoxyphenyl)acetate (4.3 g, 19.2 mmol) and 4-acetamidobenzenesulfonyl azide (p-ABSA, 7.2 g, 30 mmol) in acetonitrile (40 ml) at 0 °C, 1,8-diazabicycloundec-7-ene (DBU, 4.8 ml, 32 mmol) was added dropwise. The cooling bath was removed, and the reaction was allowed to continue stirring for 48 h (the reaction progress was followed by TLC). The reaction mixture was diluted with dichloromethane (~60 ml), washed with water (2 x ~50 ml), and dried over $MgSO_4$. After filtration, the volatile material from the filtrate was evaporated under reduced pressure. The crude product was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 95:5 gradient) as the eluent. Fractions of the pure product were combined, and the solvent evaporated, yielding 3.9 g (81%) of product.

^{1}H NMR (900 MHz, $CDCl_3$): d = 7.16 (d, J = 3.0 Hz, 1H), 6.78 (d, J = 9.0 Hz, 1H), 6.74 (dd, J = 9.0 Hz, J = 3.0 Hz, 1H), 4.27 (q, J = 7.2 Hz, 2H, OCH_2), 3.77 (s, 3H), 3.74 (s, 3H), 1.28 (t, J = 7.2 Hz, 3H, $CH_2C\underline{H}_3$);

^{13}C NMR (225 MHz, $CDCl_3$): d = 166.3, 154.1, 149.8, 115.0, 114.9, 114.0, 112.3, 61.1, 56.4, 56.0, 14.8 ($C=N_2$ signal missing, as observed before for related molecules (*4*));

HR MS (EI): calcd. for $C_{12}H_{14}N_2O_4$ $[M]^+$: 250.0954, found: 250.0957.

Benzyl (2,5-dimethoxyphenyl)diazoacetate (10):



To a solution of (2,5-dimethoxyphenyl)acetic acid (0.95 g, 5 mmol) in dichloromethane (20 ml) thionyl chloride (2 ml) was added dropwise and the reaction mixture was stirred under reflux for 1 h. The volatile materials were evaporated under vacuum. The residue was dissolved in dichloromethane (40 ml), benzyl alcohol (1 ml) was added, followed by slow addition of trimethylamine (1 ml), and the reaction mixture was stirred for 48 h (the reaction progress was followed by TLC). The reaction mixture was washed with HCl (0.5 M, 40 ml) and water (40 ml), dried over $MgSO_4$ and evaporated. The crude product was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 95:5 gradient) as the eluent. Fractions of the product were combined, and the solvent evaporated, yielding 1.43 g (quantitative) of benzyl (2,5-dimethoxyphenyl)acetate.

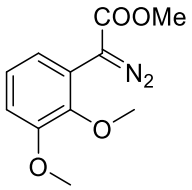
To a stirred solution of benzyl (2,5-dimethoxyphenyl)acetate (1.43 g, 5 mmol) and 4-acetamidobenzenesulfonyl azide (p-ABSA, 1.8 g, 7.5 mmol) in acetonitrile (20 ml) at 0 °C, 1,8-diazabicycloundec-7-ene (DBU, 1.2 ml, 8 mmol) was added dropwise. The cooling bath was removed, and the reaction was allowed to continue stirring overnight. The reaction mixture was diluted with dichloromethane (~50 ml), washed with water (2 x ~50 ml), dried over $MgSO_4$ and evaporated. The crude product was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 80:20 gradient) as the eluent. Fractions of the pure product were combined, and the solvent evaporated, yielding 0.94 g (60%) of product.

^{1}H NMR (900 MHz, $CDCl_3$): d = 7.37–7.32 (m, 4H), 7.31–7.28 (m, 1H), 7.16 (bs, 1H), 6.78 (d, J = 9.0 Hz, 1H), 6.75 (dd, J = 9.0 Hz, J = 3.0 Hz, 1H), 5.26 (s, 2H), 3.76 (s, 3H), 3.71 (s, 3H);

^{13}C NMR (225 MHz, $CDCl_3$): d = 166.1, 154.1, 149.8, 136.3, 128.8, 128.4, 115.0, 114.6, 114.2, 112.3, 66.6, 56.4, 56.0 (C=N_2 signal missing, as observed before for related molecules (*4*));

HR MS (EI): calcd. for $C_{17}H_{16}N_2O_4$ $[M]^+$: 312.1110, found: 312.1111.

Methyl (2,3-dimethoxyphenyl)diazoacetate (11):



In a closed vial, a solution of (2,3-dimethoxyphenyl)acetic acid (1.5 g, 7.9 mmol) in methanol (20 ml) containing several drops of sulfuric acid, was stirred overnight at 80 °C. The volatile materials

were evaporated under vacuum. The residue was dissolved in ethyl acetate (~40 ml), washed with $NaHCO_3$ sat. (40 ml) and water (40 ml), dried over $MgSO_4$ and evaporated. The crude product was used in the next step without further purification.

To a stirred solution of methyl (2,3-dimethoxyphenyl)acetate and 4-acetamidobenzenesulfonyl azide (p-ABSA, 2.9 g, 12 mmol) in acetonitrile (30 ml) at 0 °C, 1,8-diazabicycloundec-7-ene (DBU, 1.9 ml, 13 mmol) was added dropwise. The cooling bath was removed, and the reaction was allowed to continue stirring for 48 h (the reaction progress was followed by TLC). The reaction mixture was diluted with dichloromethane (~50 ml), washed with water (2 x ~50 ml), and dried over $MgSO_4$. After filtration, the volatile material from the filtrate was evaporated under reduced pressure. The crude product was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 80:20 gradient) as the eluent. Fractions of the pure product were combined, and the solvent evaporated, yielding 1.55 g (83%) of product.

^{1}H NMR (900 MHz, $CDCl_3$): d = 7.19 (d, J = 7.9 Hz, 1H), 7.05 (dd, J = 8.1 Hz, J = 8.1 Hz, 1H), 6.79 (d, J = 8.2 Hz, 1H), 3.83 (s, 3H), 3.81 (s, 3H), 3.79 (s, 3H);

^{13}C NMR (225 MHz, $CDCl_3$): d = 166.6, 152.9, 145.3, 124.5, 121.3, 120.0, 111.4, 60.8, 56.0, 52.2 ($C=N_2$ signal missing, as observed before for related molecules (*4*));

HR MS (EI): calcd. for $C_{11}H_{12}N_2O_4$ $[M]^+$: 236.0797, found: 236.0800.

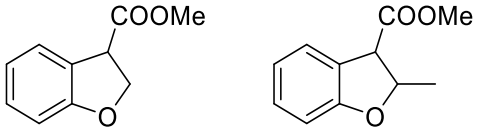
c. Synthesis of Authentic Products

The synthesis of the following compounds was conducted by Pawel Dydio.

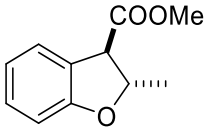
General procedure for synthesis of dihydrobenzofurans

To a solution of a derivative of methyl (2-methoxyphenyl)diazoacetate (~50 mM) in toluene a solution of Ir(Me)-PIX (8 mM, 0.2-2 mol%) in DMF was added, and the reaction mixture was vigorously stirred. The reaction progress was monitored by TLC. Upon completion, the volatile materials were removed, and the residue was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 80:20 gradient) as eluent. Fractions of the pure product were combined, and the solvent evaporated, yielding 20-90% of desired product.

Methyl 2,3-dihydrobenzofuran-3-carboxylate (2) and methyl 2-methyl-2,3-dihydrobenzofuran-3-carboxylate (**P7**) were characterized before [32,34].



Trans-methyl 2-methyl-2,3-dihydrobenzofuran-3-carboxylate (P7):

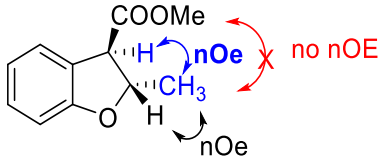


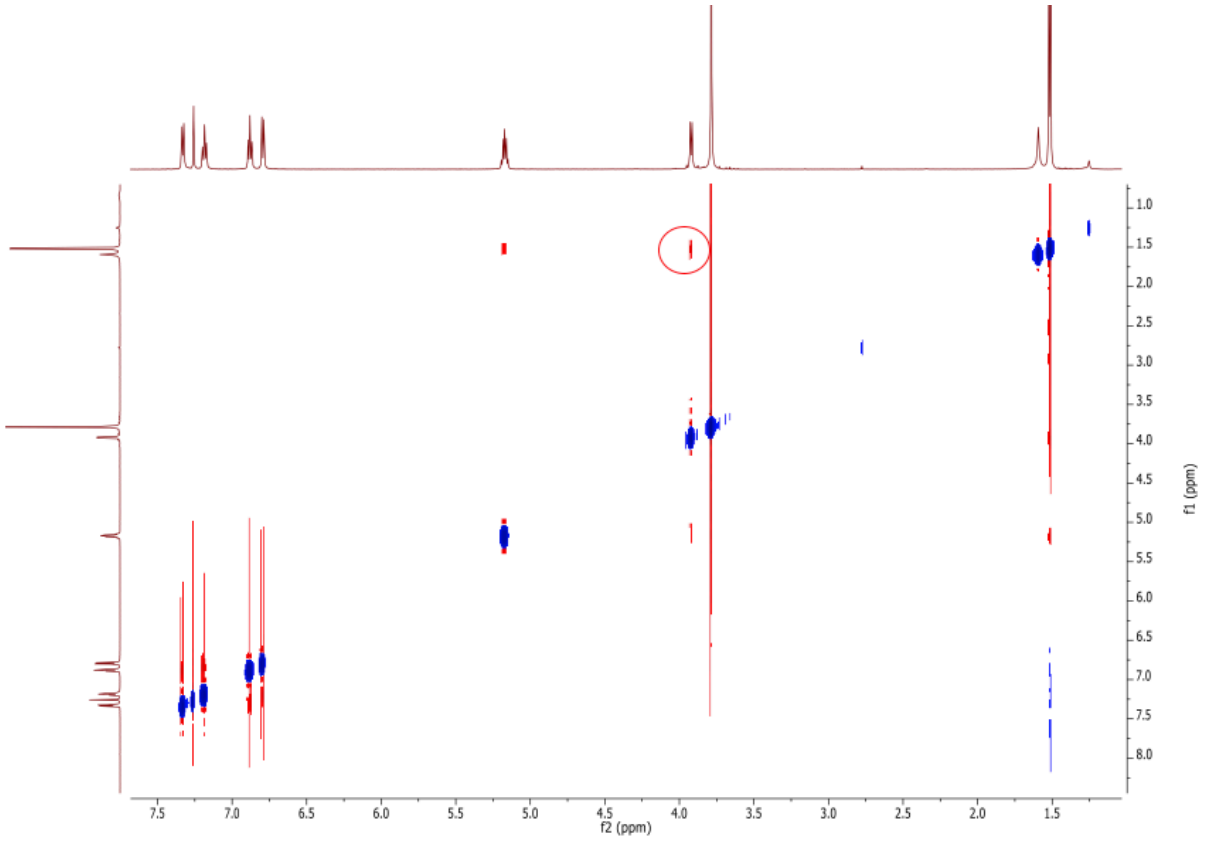
^{1}H NMR (900 MHz, $CDCl_3$): d = 7.29 (d, J = 7.7 Hz, 1H), 7.15 (dd, J = 7.8 Hz, J = 7.6 Hz, 1H), 6.84 (dd, J = 7.6 Hz, J = 7.6 Hz, 1H), 6.76 (d, J = 7.9 Hz, 1H), 5.13 (dq, J = 7.9 Hz, J = 6.5 Hz, 1H), 3.75 (s, 3H), 1.48 (d, J = 6.5 Hz, 3H);

^{13}C NMR (225 MHz, $CDCl_3$): d = 171.7, 159.4, 129.7, 125.6, 124.4, 120.7, 110.1, 81.5, 54.5, 52.7, 21.4;

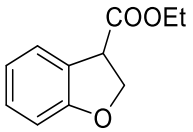
HR MS (EI): calcd. for $C_{11}H_{12}O_3$ $[M]^+$: 192.0786, found: 192.0784.

^{1}H-^{1}H NOESY NMR:





Ethyl 2,3-dihydrobenzofuran-3-carboxylate (P6):

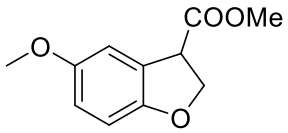


^{1}H NMR (900 MHz, $CDCl_3$): d = 7.34 (d, J = 7.5 Hz, 1H), 7.14 (dd, J = 7.9 Hz, J = 7.9 Hz, 1H), 6.85 (dd, J = 7.5 Hz, J = 7.5 Hz, 1H), 6.78 (d, J = 8.0 Hz, 1H), 4.90 (dd, J = 9.0 Hz, J = 6.8 Hz, 1H), 4.63 (dd, J = 9.4 Hz, J = 9.4 Hz, 1H), 4.29 (dd, J = 9.7 Hz, J = 6.8 Hz, 1H), 4.19 (dq, J = 13.3 Hz, J = 7.1 Hz, 2H), 1.27 (t, J = 7.1 Hz, 3H);

^{13}C NMR (225 MHz, $CDCl_3$): d = 173.9, 162.6, 132.2, 128.1, 127.1, 123.4, 112.7, 75.2, 64.3, 49.9, 17.0;

HR MS (EI): calcd. for $C_{11}H_{12}O_3$ $[M]^+$: 192.0786, found: 192.0782.

Methyl 5-methoxy-2,3-dihydrobenzofuran-3-carboxylate (P8):

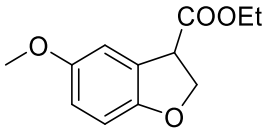


^{1}H NMR (900 MHz, $CDCl_3$): d = 6.91 (s, 1H), 6.71-6.68 (m, 2H), 4.86 (dd, J = 9.1 Hz, J = 6.7 Hz, 1H), 4.61 (dd, J = 9.5 Hz, J = 9.3 Hz, 1H), 4.27 (dd, J = 9.5 Hz, J = 6.7 Hz, 1H), 3.74 (s, 3H), 3.73 (s, 3H);

^{13}C NMR (225 MHz, $CDCl_3$): d = 171.7, 154.4, 154.1, 125.2, 114.9, 111.5, 110.1, 72.9, 56.3, 52.8, 47.8;

HR MS (EI): calcd. for $C_{11}H_{12}O_4$ $[M]^+$: 208.0736, found: 208.0740.

Ethyl 5-methoxy-2,3-dihydrobenzofuran-3-carboxylate (P9):

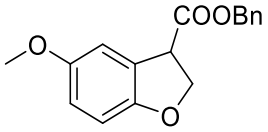


^{1}H NMR (900 MHz, $CDCl_3$): d = 6.92 (s, 1H), 6.70-6.68 (m, 2H), 4.86 (dd, J = 9.0 Hz, J = 6.9 Hz, 1H), 4.61 (dd, J = 9.6 Hz, J = 9.1 Hz, 1H), 4.26 (dd, J = 9.5 Hz, J = 7.0 Hz, 1H), 4.19 (dq, J = 19.2 Hz, J = 7.1 Hz, 2H), 3.72 (s, 3H), 1.27 (t, J = 7.1 Hz, 3H);

^{13}C NMR (225 MHz, $CDCl_3$): d = 171.2, 154.3, 154.2, 125.3, 114.9, 111.4, 110.1, 61.7, 56.3, 47.8, 14.5;

HR MS (EI): calcd. for $C_{12}H_{14}O_4$ [M] : 222.0892, found: 222.0893.

Benzyl 5-methoxy-2,3-dihydrobenzofuran-3-carboxylate (P10):

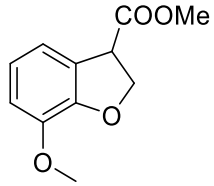


^{1}H NMR (900 MHz, $CDCl_3$): d = 7.35-7.29 (m, 5H), 6.86 (s, 1H), 6.70-6.68 (m, 2H), 5.17 (dd, J = 43.3 Hz, J = 12.2 Hz, 2H), 4.88 (dd, J = 9.1 Hz, J = 6.8 Hz, 1H), 4.62 (dd, J = 9.5 Hz, J = 9.2 Hz, 1H), 4.31 (dd, J = 9.6 Hz, J = 7.0 Hz, 1H), 3.66 (s, 3H);

^{13}C NMR (225 MHz, $CDCl_3$): d = 171.0, 154.3, 154.1, 135.7, 128.9, 128.7, 128.6, 125.0, 115.4, 111.1, 110.2, 72.9, 67.5, 56.2, 47.8;

HR MS (EI): calcd. for $C_{17}H_{16}O_4$ $[M]^+$: 284.1049, found: 284.1053.

Methyl 7-methoxy-2,3-dihydrobenzofuran-3-carboxylate (P11):



^{1}H NMR (900 MHz, $CDCl_3$): d = 6.95 (d, J = 7.7 Hz, 1H), 6.81 (dd, J = 7.9 Hz, J = 7.7 Hz, 1H), 6.76 (d, J = 8.2 Hz, 1H), 4.95 (dd, J = 9.2 Hz, J = 6.8 Hz, 1H), 4.69 (dd, J = 9.7 Hz, J = 9.3 Hz, 1H), 4.33 (dd, J = 9.7 Hz, J = 6.9 Hz, 1H), 3.83 (s, 3H), 3.73 (s, 3H);

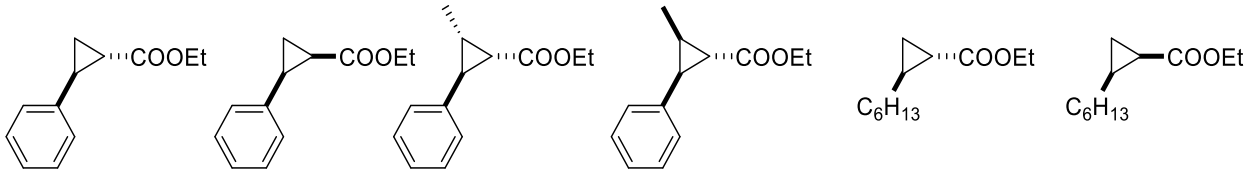
^{13}C NMR (225 MHz, $CDCl_3$): d = 171.7, 148.6, 145.1, 125.2, 121.5, 117.6, 112.6, 73.4, 56.2, 52.8, 47.9;

HR MS (EI): calcd. for $C_{11}H_{12}O_4$ $[M]^+$: 208.0738, found: 208.0740.

General procedure for synthesis of cyclopropanes

To a solution of an alkene (~0.1-0.5 M, 5-10 equiv.) in toluene, a solution of Ir(Me)-PIX (8 mM, 0.2 mol%) in DMF was added, followed by slow addition of a solution of ethyl diazoacetate (1 eq.) in toluene, while the reaction mixture was vigorously stirred. Upon completion, the volatile materials were removed and the residue was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 95:5 gradient) as the eluent. Fractions of the pure product(s) were combined, and the solvent evaporated, yielding desired products.

Cis- and trans- ethyl 2-phenylcyclopropane-1-carboxylates (P12), ethyl syn-2-methyl-anti-3-phenylcyclopropane-1-carboxylate (5), ethyl syn-2-methyl-syn-3-phenylcyclopropane-1-carboxylate and cis- and trans-ethyl 2-hexylcyclopropane-1-carboxylates (**P13**) have been reported previously [35].

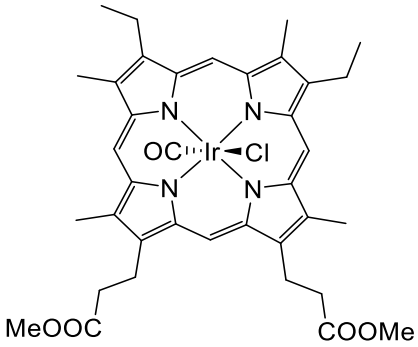


The absolute configurations of cis- and trans- ethyl 2-phenylcyclopropane-1-carboxylates, and cis- and trans-ethyl 2-hexylcyclopropane-1-carboxylate were assigned on the basis of previous reports [9,36].

d. Metal cofactors

The synthesis of the Ir-porphyrin complexes was conducted by Pawel Dydio. Characterization of the complexes was accomplished by Pawel Dydio and Hanna. Key.

Ir(CO)Cl-Mesoporphyrin IX dimethyl ester:



Under a nitrogen atmosphere, a suspension of mesoporphyrin IX dimethyl ester (250 mg, 0.42 mmol) and $[Ir(COD)Cl]_2$ (400 mg, 0.6 mmol) in dry and degassed 1,2-xylene (20 ml) was stirred at 150 °C for 7 days. The crude reaction mixture was loaded onto a plug of silica gel (SiO_2). The organic material was first eluted with a mixture of hexanes and ethyl acetate (100:0 → 75:25 gradient) to remove the remaining starting material and a side product. Then it was eluted with a mixture of hexanes and ethyl acetate (75:25 → 50:50 gradient) to collect the product. Fractions containing the pure product were combined, and the solvent evaporated, yielding 85 mg (24%) of a deep red solid.

^{1}H NMR (500 MHz, $CDCl_3$): d = 10.33–10.28 (m, 4H), 4.53–4.41 (m, 4H), 4.20–4.01 (m, 4H), 3.75–3.71 (m, 6H), 3.71–3.67 (m, 9H), 3.64 (s, 3HH), 3.38 (t, J = 7.8 Hz, 4H), 1.95 (t, J = 7.6 HZ, 6H);

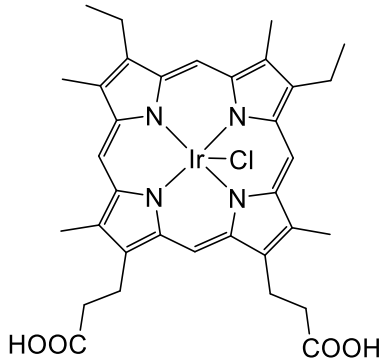
^{13}C NMR (225 MHz, $CDCl_3$): d = 173.83, 173.82, 143.2, 143.1, 140.0, 139.8, 139.4, 139.23, 139.20, 139.19, 138.9, 138.5, 138.45, 138.4, 137.38, 137.35, 136.4, 136.3, 135.0, 99.2, 99.1, 99.0, 98.9, 51.99, 51.96, 37.05, 37.04, 22.07, 22.01, 20.10, 20.09, 17.85, 17.83, 12.02, 11.98, 11.85, 11.81;

HR MS (ESI): calcd. for $C_{36}H_{40}IrN_4O_4$ $[M\text{-}Cl\text{-}CO]^+$: 785.2673, found: 785.2715;

IR (neat): 2031, 1735 cm^{-1};

UV/Vis (DMF, C=5·10^{-6} M): $_{max}$ (log) 327 (4.23), 393 (5.29), 510 (4.02), 540 nm (4.27).

Ir(Cl)-Mesoporphyrin IX:



A solution of Ir(CO)Cl-mesoporphyrin IX dimethyl ester (25 mg, 0.03 mmol) and LiOH (100 mg) in THF (3 ml), methanol (1 ml) and water (1 ml) was stirred at rt for 2 h. The reaction mixture was concentrated (~1 ml) under vacuum, diluted with sodium phosphate buffer (4 ml, 0.1 M, pH = 5) and slowly acidified with HCl (0.5 M) to ~ pH = 5. The red precipitate was separated from the liquid by centrifugation (2000 rpm, 15 min, 4 °C) and subsequent decanting of the liquid. The red solid residue was suspended in water (15 ml), the mixture was centrifuged, and the liquid was decanted. The resulting solid was dried for overnight under high vacuum at rt, yielding 21 mg (88%) of dark red powder.

(Alternatively, the precipitate can be isolated by filtration, followed by washing with water and drying).

^{1}H NMR (600 MHz, D_2O+0.5% NaOD): d = 10.22 (bs, 1H), 9.98 (bs, 3H), 4.35 (bs, 4H), 4.08 (bs, 4H), 3.67 (bs, 12H), 3.12 (bs, 4H), 1.84 (bs, 6H);

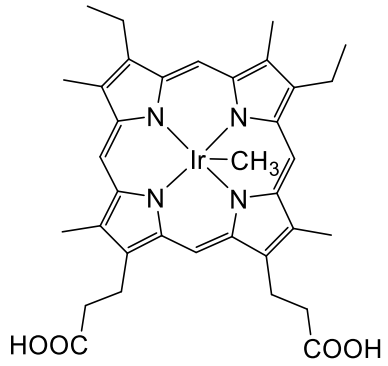
^{13}C NMR (225 MHz, D_2O+0.5% NaOD): δ = 185.7, 145.8, 145.6, 145.16, 145.10, 145.07, 145.0, 144.09, 144.08, 144.05, 143.9, 142.53, 142.51, 139.74, 139.71, 138.7, 138.5, 102.37, 102.14, 101.99, 101.96, 43.7, 25.4, 21.5, 19.95, 19.94, 13.12, 13.10, 12.88;

HR MS (ESI): calcd. for $C_{34}H_{36}IrN_4O_4$ $[M\text{-}Cl]^+$: 757.2360, found: 757.2393;

IR (neat): 1706 cm^{-1};

UV/Vis (DMF, C=5·10^{-6} M): $_{max}$ (log) 326 (4.40), 393 (5.26), 510 (4.15), 540 nm (4.37).

Ir(Me)-Mesoporphyrin IX:



Under a nitrogen atmosphere, to a suspension of Ir(CO)Cl-mesoporphyrin IX dimethyl ester (110 mg, 0.13 mmol) in degassed ethanol (10 ml), a degassed solution of $NaBH_4$ (25 mg) and NaOH (1 M) in water (3 ml) was added. The reaction mixture was stirred at 50 °C for 1 h in the dark, cooled to rt, and followed by addition of methyl iodide (10 ml). The reaction mixture was stirred overnight at rt. The reaction mixture was concentrated (~3 ml) under vacuum (NOTE: Evaporation of unreacted methyl iodide is important), followed by addition of LiOH (200 mg). The reaction mixture was stirred at rt for 1h, diluted with sodium phosphate buffer (8 ml, 0.1 M, pH = 5) and slowly acidified with HCl (0.5 M) to ~ pH = 5. The red precipitate was separated from the liquid by centrifugation (2000 rpm, 15 min, 4 °C) and subsequent decanting of the liquid. The red solid residue was suspended in water (15 ml), the mixture was centrifuged, and the liquid was decanted. The resulting solid was dried for overnight under high vacuum at rt, yielding product quantitatively; dark red powder.

(Alternatively, the precipitate can be isolated by filtration, followed by washing with water and drying).

^{1}H NMR (900 MHz, DMF-d_7): d = 12.56 (bs, 2H), 9.93 (s, 1H), 9.78 (s, 3H), 4.38–4.33 (m, 2H), 4.29–4.24 (m, 2H), 4.03–3.98 (m, 4H), 3.61 (s, 6H), 3.58 (s, 3H), 3.57 (s, 3H), 3.28 (q, J = 7.5 Hz, 6H), 1.83 (t, J = 7.5 Hz, 6H), –7.61 (s, 3H);

^{13}C NMR (225 MHz, DMF-d_7): d = 175.40, 175.39, 143.5, 143.2, 143.0, 142.91, 142.88, 142.79, 142.4, 142.0, 141.83, 141.82, 139.82, 139.76, 136.95, 136.90, 135.9, 135.7, 101.0, 100.9, 100.7, 100.5, 38.34, 38.32, 22.50, 22.49, 20.09, 20.06, 18.59, 18.55, 11.63, 11.61, 11.48, 11.44, -45.0;

HR MS (ESI): calcd. for $C_{35}H_{40}IrN_4O_4$ $[M+H]^+$: 773.2673, found: 773.2708;

IR (neat): 1706 cm^{-1};

UV/Vis (DMF, C=5·10^{-6} M): $_{max}$ (log) 341 (4.39), 392 (5.07), 530 (4.24).

3.8.3 Preparation of the Multi-Dimensional Catalyst Array

In a nitrogen-atmosphere glove box, stock solutions of 9 metal cofactors (8.3 uL per well, 6 mM, DMF) were distributed down the columns of a 96-well plate containing 1.2 mL glass vials (Table S2). Eight mutants (Table S2) of apo-mOCR-myo were expressed and purified according to the procedure in Section I.e. Each mutant was concentrated to 0.6 mM and degassed on a Schlenk line (3 cycles vacuum/refill), and the mutants were distributed across the rows of the same 96-well plate (166 uL protein per vial) to generate 72 unique catalysts formed from different combinations of cofactors and mutants. The prepared portion of each catalyst was evenly divided among 4 separate 96-well reaction plates of the same type (42 uL per well), and each well was diluted to 250 uL with 10 mM tris buffer, pH = 8.0.

3.8.4 Catalytic Experiments

a. *General Methods*:
Unless otherwise noted, catalytic reactions were performed in 4 mL individually-capped vials or in 1.2 mL vials as part of a 96-well array fitted with a screw on cover. Reactions were either (1) assembled in a nitrogen atmosphere glove box or (2) assembled on the bench. In the latter case, the headspace of the vial purged with nitrogen through a septum cap. Solutions of Ir(Me)-PIX-mOCR-Myo were gently degassed on a Schlenk line (3 cycles vacuum/refill) before being pumped into a glove box in sealed vials. Organic reagents were added as stock solutions in MeCN, such that the final amount of MeCN in the reaction was approximately 8% by volume. Protein catalysts were diluted to reaction concentration in Tris buffer (10 mM, pH = 8.0) before being added to reaction vials. Unless otherwise noted, all reactions were performed with catalysts generated from a 1:2 ratio of [M]-cofactor : apo protein, with 0.5% catalyst loading with respect to [M] cofactor and limiting reagent, if not noted otherwise. All reactions were conducted in a shaking incubator (20 °C, 16 h, 275 rpm).

b. Reaction Setup and Composition for Each Class of Transformation

Cyclopropanation: Catalyst solution (240 uL, 0.1 mM protein, 0.24 mol) was added to a vial. A stock solution of the appropriate olefin (2.5 umol in 10 uL MeCN) was added, followed by a stock solution of the appropriate diazo compound (15 umol in 10 uL MeCN).

Intramolecular C-H Insertion: A stock solution of the appropriate diazo compound (2.5 umol in 20 uL MeCN) was added to a vial followed by the catalyst solution (240 uL, 0.1 mM protein, 0.24 umol).

c. Evaluation of the Activity of Catalysts in the Array:

Plates containing the catalyst array were prepared as described in Section III. Reactions were assembled in a nitrogen atmosphere glove box according to the general methods described above. Plates were sealed before being removed from the glove box.

d. Mutant Screening (directed evolution)

Mutants were evaluated for C-H insertion reactions in 96 well plates with the reaction conditions described in the general method (Section IVa,b). Aliquots of the twelve Ir(Me)-Myo mutants were added to each well of the 12 columns of a 96 well reaction plate, followed by addition of 4-(pyridin-4-yl)morpholine (1 mM) – the size selective inhibitor, while aliquots of 7 substrates for C-H insertion were added to each of the rows (A-G) to generate 84 unique reactions. The most selective mutants were subsequently evaluated without the inhibitor (4-(pyridin-4-yl)morpholine).

e. Procedure for typical catalytic experiments

Catalyst solution (240 uL, 0.1 mM protein) was added to a vial, followed by addition of stock solutions of substrates: the appropriate diazo compound (2.5 umol in 20 uL MeCN) – in case of the intramolecular C-H insertion reactions, or the appropriate olefin (2.5 umol in 10 uL MeCN) and EDA (15 umol in 10 uL MeCN) – in case of the cyclopropanation reaction. The vial was sealed with a cap, removed from a glovebox and shaken (20 °C, 275 rpm). Upon completion, the reaction mixture was analyzed in the general method described in section VI.

f. Experiments with slow addition of EDA

Catalyst solution (960 uL, 0.1 mM protein) was added to a vial equipped with a magnetic stirring bar. A stock solution of the appropriate olefin (10 umol in 40 uL MeCN) was added. The vial was sealed with a septum-cap and removed from a glovebox. The stock solution of EDA (60 umol in 40 uL MeCN) was slowly added from a gas-tight syringe (a needle extending into the reaction mixture) over ~8 h, using a syringe pump, with slow stirring (100 rpm). Upon completion, the reaction mixture was incubated for additional 1 h and analyzed in the general method described in section VI.

g. Reaction on a synthetic scale

Catalyst solution (12 mL, 0.1 mM protein (mOCR-myo-93A,64A,43W,68A)) was added to a Schlenk flask and gently degassed on a Schlenk line (3 cycles vacuum/refill). A solution of substrate **11** (28.3 mg, 0.12 mmol in 0.96 mL ACN) was added. The flask was sealed and gently shaken (120 rpm) at 20 °C for 72h. The reaction was diluted with brine (30 ml) and extracted with ethyl acetate (3·50 ml). If required, the phase separation was achieved by centrifuging (2000 rpm, 3 min) the mixture. The combined organic fractions were washed with brine (30 ml), dried over $MgSO_4$. After filtration, the volatile material from the filtrate was evaporated under reduced pressure. The residue was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 90:10 gradient) as the eluent. Fractions of the pure product were combined, and the solvent evaporated, yielding 20.1 mg (80%) of product **P11**. Er: 90:10, $[\alpha]_D^{20}$ = 52° (c = 0.8, $CHCl_3$].

3.8.5 Analysis of Yield and Enantiomeric Ratio (er)

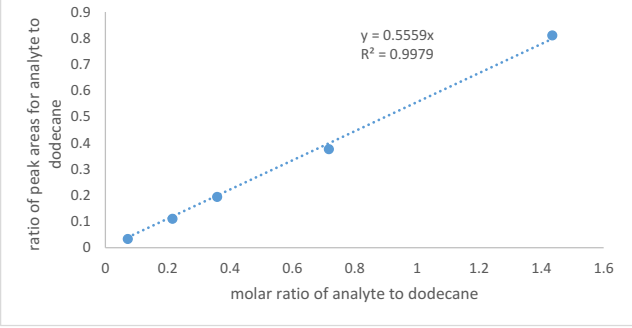
Yields were determined by (a) chiral GC using dodecane as an internal standard. Enantiomeric ratios were determined either by chiral GC or chiral SFC (Figures S21-S29). Achiral GC was used to determine the yields for reactions analyzed by SFC, while the yields for reactions analyzed by chiral GC were determined concurrently with that analysis. The methods used to determine the er of each product are summarized subsequently. Representative traces can be found in Figures S21-29. Samples for analysis were prepared as follows, depending on the analysis method:

SFC/achiral GC: Saturated NaCl (200 uL) was added to each reaction vial, followed by a solution of dodecane (500 ul, 1 ul/ml) in EtOAc. The contents of the vial were mixed by pipet, and the phases were allowed to separate. A portion (250 uL) of the organic layer was removed from the top of the vial by pipet, transferred to a new vial, evaporated and redissoved in MeOH for analysis by SFC. An additional portion of EtOAc (250 uL) was added to the original reaction vial, and the reaction was quenched by the addition of HBr (40 uL). After separation of the layers, approximately 400 uL of the aqueous phase was removed from the bottom of the vial by pipet. The remaining contents of the vial were neutralized by the addition of sat. $NaHCO_3$ (200 uL), and the organic layer was further diluted with EtOAc (500 uL). The organic layer was then transferred to a separate vial for GC analysis. In the case of experiments performed in a 96-well array, all manipulations were performed using a multichannel pipet.

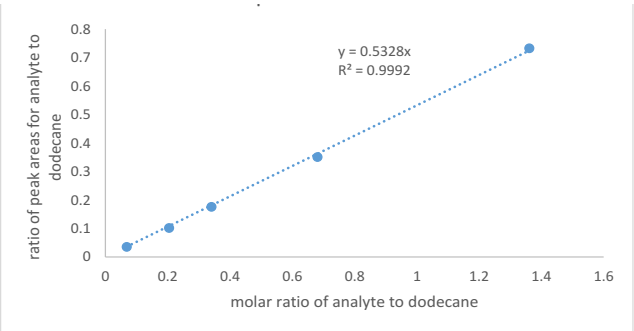
Chiral GC: Saturated NaCl (200 uL) was added to each reaction vial, followed by a solution of dodecane (500 ul, 1 ul/ml) in EtOAc. The contents of the vial were mixed by pipet and the phases were allowed to separate. The reaction was then quenched by the addition HBr (40 uL). After separation of the layers, approximately 400 uL of the aqueous phase was removed from the bottom of the vial by pipet. The remaining contents of the vial were neutralized by the addition of sat. $NaHCO_3$ (200 uL), and the organic layer was further diluted with EtOAc (500 uL). The organic layer was then transferred to a separate vial for GC analysis. In the case of experiments performed in a 96-well array, all manipulations were performed using a multichannel pipet.

3.8.6 Calibration Curves Used to Determine Yields from GC

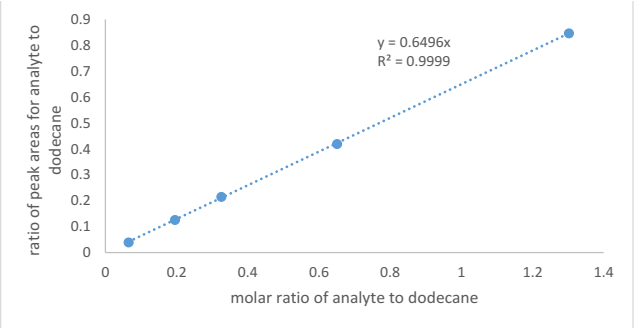
a. *Product **2***



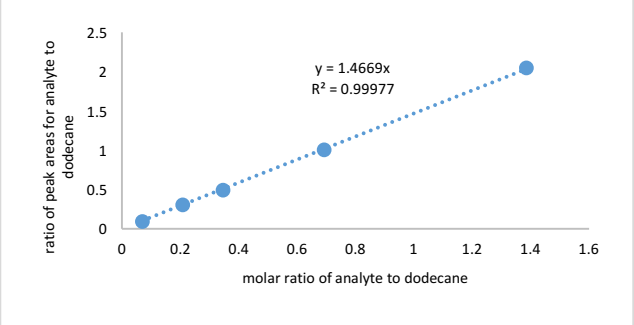
b. *Product **6**.*



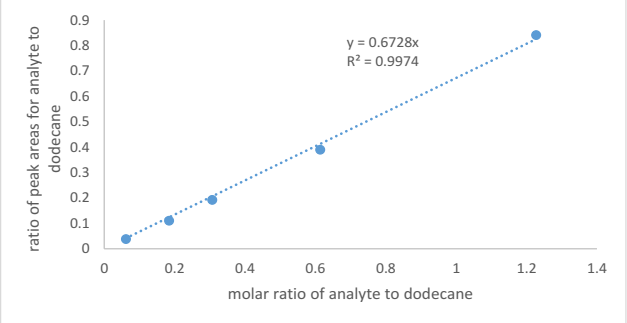
c. *Product **7***



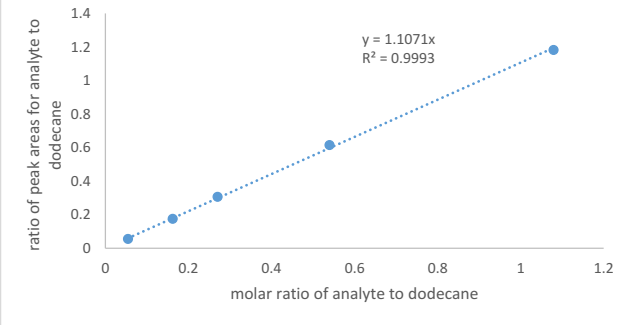
d. *Product **8***



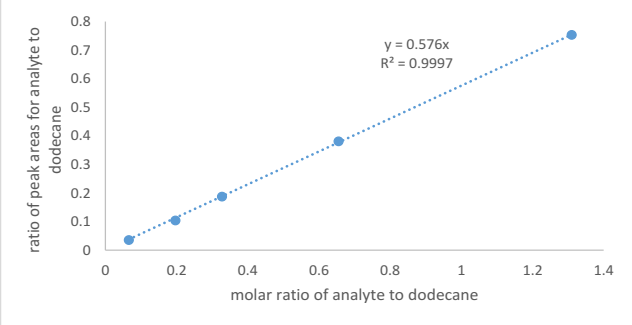
e. *Product **9***



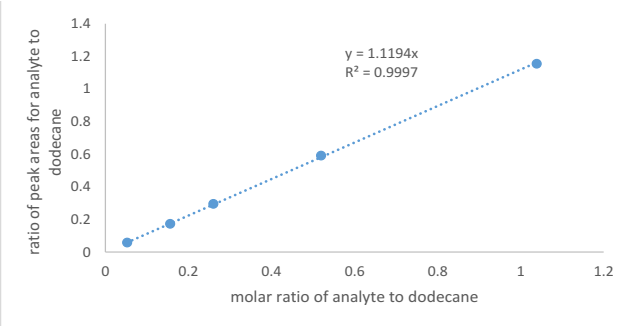
f. *Product **10***



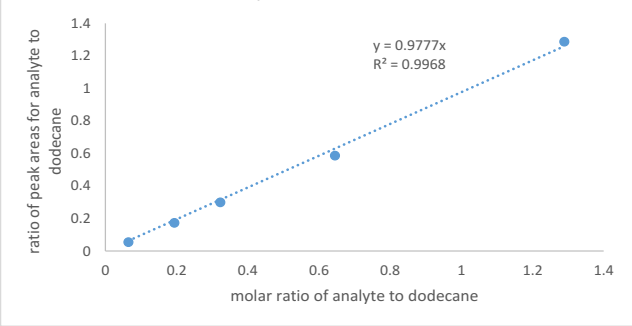
g. *Product **11***



h. *Product trans-**13***

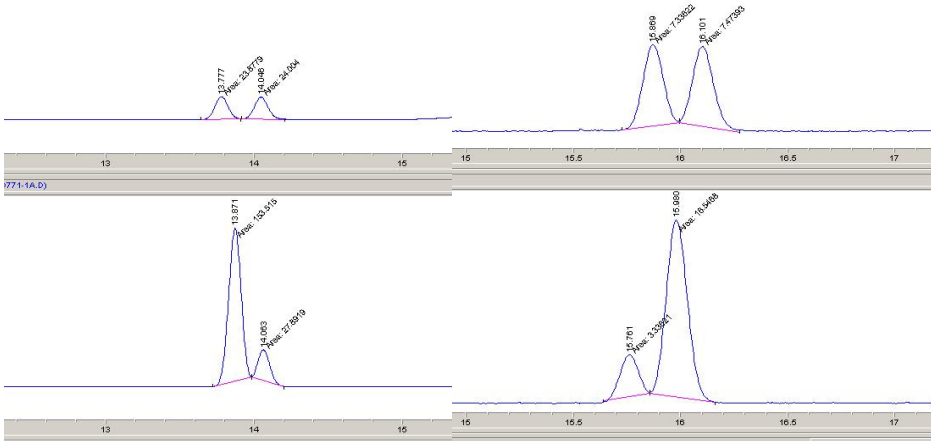


i. *Product cis-**13***

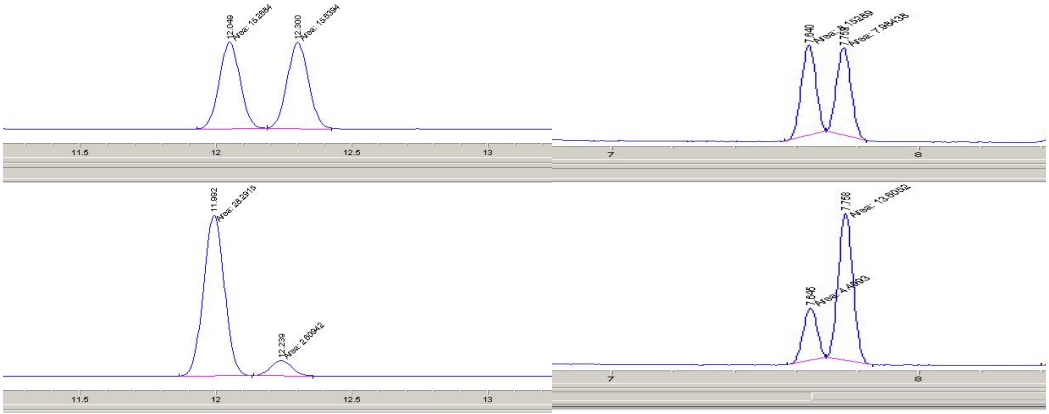


3.8.7 Representative Traces from Chiral GC and Chiral SFC

a. Chiral GC traces for product **2**; racemate (above) and enantioenriched samples (below).

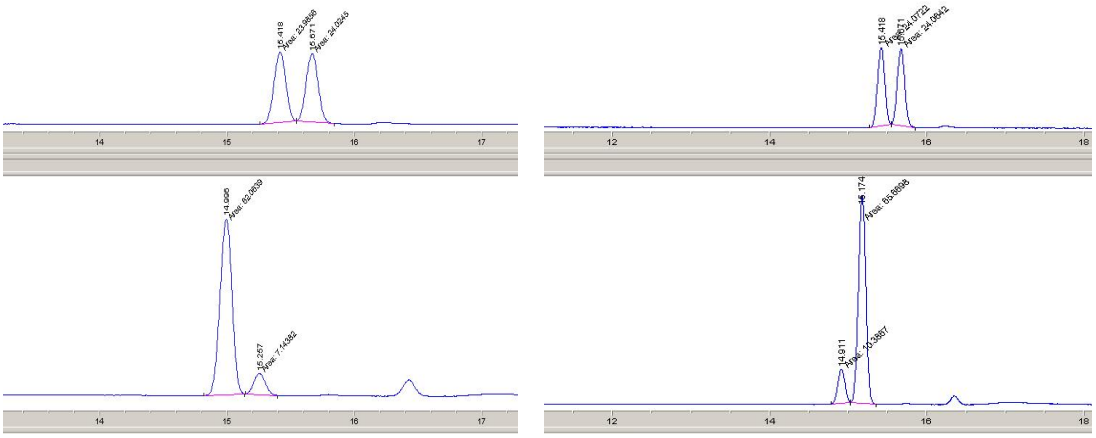


b. Chiral GC traces for product **P7**; racemate (above) and enantioenriched samples (below).

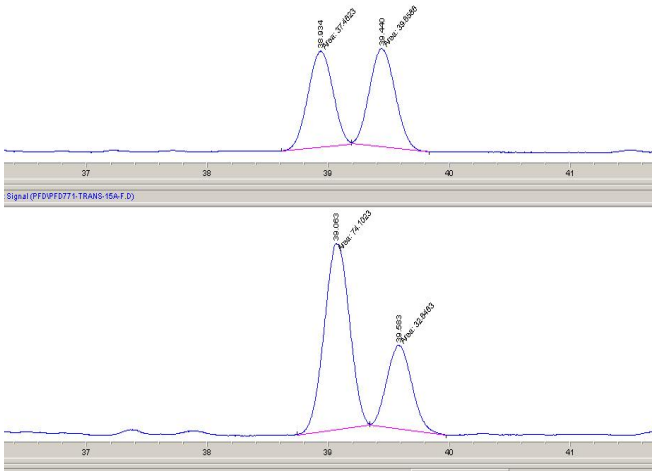


c. Chiral GC traces for product **P8**; racemate (above) and enantioenriched samples (below).

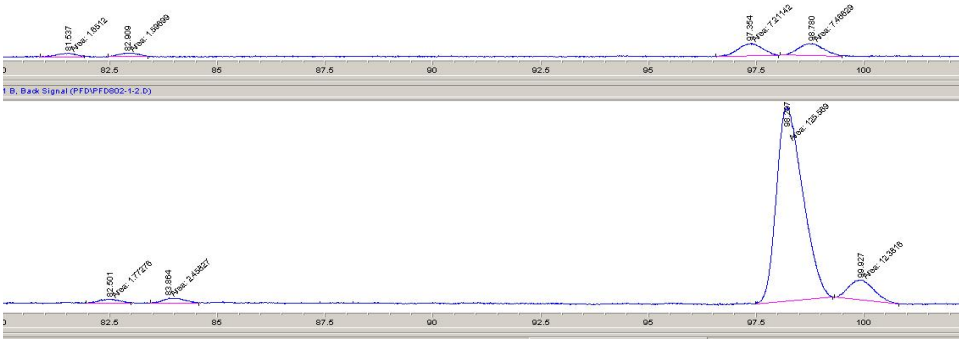
d. Chiral GC traces for product **P11**; racemate (above) and enantioenriched samples (below).



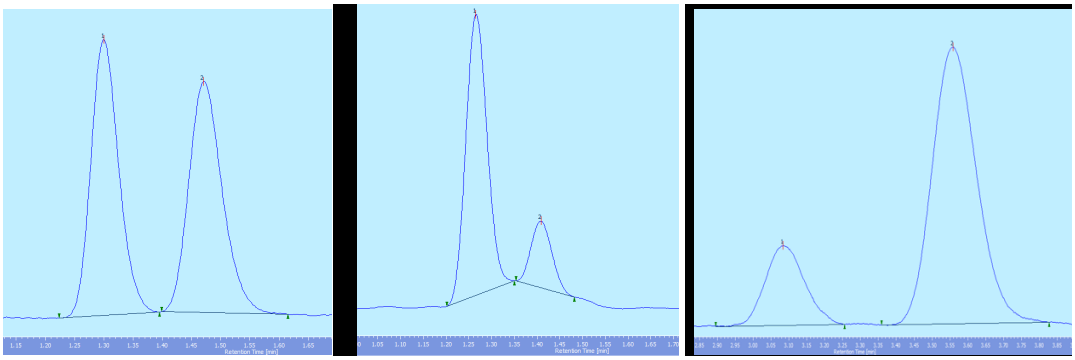
e. Chiral GC traces for ethyl syn-2-methyl-anti-3-phenylcyclopropane-1-carboxylate (**P13**) (product of cyclopropanation of trans-β-methylstyrene with EDA); racemate (above) and enantioenriched sample (below).



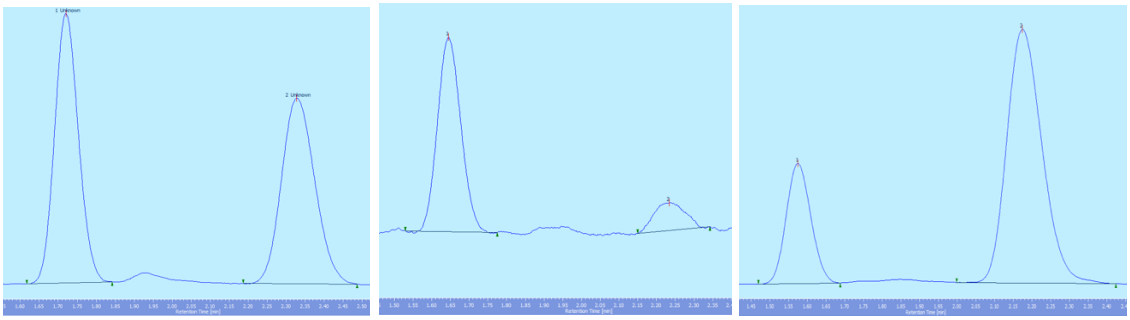
e. Chiral GC traces for cis- and trans-ethyl 2-hexylcyclopropane-1-carboxylates (**P13**) (products of cyclopropanation of 1-octene with EDA); racemate (above) and enantioenriched sample (below).



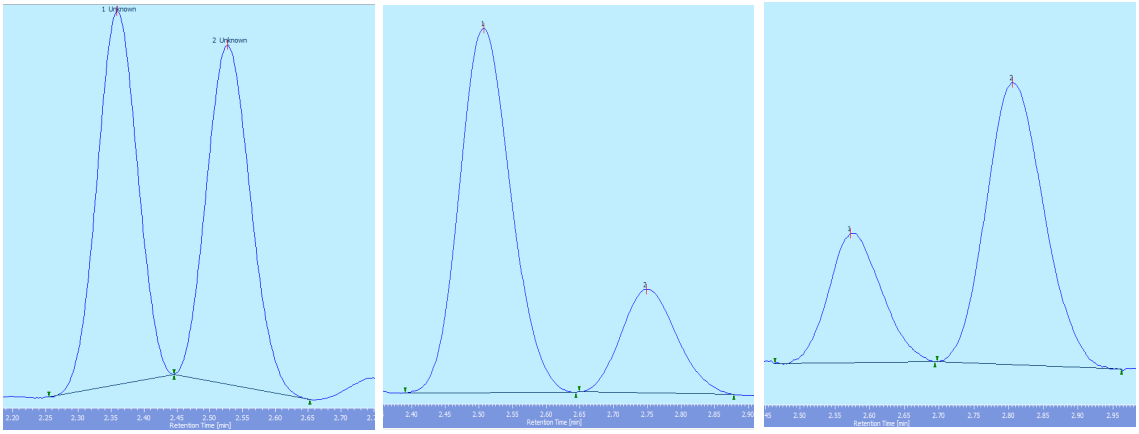
f. Chiral SFC traces for product **P6**; racemate (left) and enantioenriched samples (center, right).



g. Chiral SFC traces for product **P9**; racemate (left) and enantioenriched samples (center, right).



h. Chiral SFC traces for product **P10**; racemate (left) and enantioenriched samples (center, right).



3.8.8. Methods used to separate enantiomers of the generated reaction products

Product	Instrument	Column	Method	Retention Times
2[a]	GC	CYCLOSIL-B (30m x 0.25mm x 0.25u)	Isothermal: 150 °C	t1(R)(-)=13.8 min t2(S)(+)=14.0 min
P7	GC	CYCLOSIL-B (30m x 0.25mm x 0.25u)	Isothermal: 145 °C	t1=12.0 min t2=12.3 min
P8	GC	CYCLOSIL-B (30m x 0.25mm x 0.25u)	Isothermal: 145 °C	t1=47.1 min t2=47.9 min
P11	GC	CYCLOSIL-B (30m x 0.25mm x 0.25u)	Isothermal: 170 °C	t1=15.0 min t2=15.3 min
5 (trans-β-methylstyrene + EDA)	GC	CYCLOSIL-B (30m x 0.25mm x 0.25u)	Isothermal: 130 °C	t1=39.1 min t2=39.6 min
P13[a] (1-octene+EDA)	GC	CYCLOSIL-B (30m x 0.25mm x 0.25u)	Isothermal: 90 °C	Cis: (1S,2R)(+) t1=81.5 min (1R,2S)(-) t2=82.9 min Trans: (1R,2R)(-) t1=97.4 min (1S,2S)(+) t2=98.8 min
P12[a] (styrene+EDA)	GC	CYCLOSIL-B (30m x 0.25mm x 0.25u)	Isothermal: 120 °C	Cis: (1S,2R)(+) t1=37.2 min (1R,2S)(-) t2=39.0 min Trans: (1R,2R)(-) t1=48.3 min (1S,2S)(+) t2=49.2 min
P6	SFC	Chiracel OD-H (Diacel)	Isocratic: 1% MeOH, 4 ml/min flow	t1=1.3 min t2=1.5 min
P9	SFC	Chiracel AD-H (Diacel)	Isocratic: 1% MeOH, 4 ml/min flow	t1=1.6 min t2=2.25 min
P10	SFC	Chiracel AD-H (Diacel)	Isocratic: 4% MeOH, 4 ml/min flow	t1=2.5 min t2=2.75 min

a – absolute geometry assigned by comparison with literature [9,36].

3.8.9. Complete, Tabulated Results of Directed Evolution

Complete, tabulated results for the directed evolution of Ir(Me)-PIX-mOCR-Myo for the C-H insertion of substrate 1 under the standard, screening reactions conditions. Mutants are designated by the substitutions listed in the table header plus additional mutations indicated in the row and column headers of each table. Data are presented as: enantiomeric ratio (GC yield).

mOCR-Myo. H93G, H64A- 2-OMe, COOMe							
6A	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	35 : 65 (12)	32 : 68 (8)	35 : 65 (20)	39 : 61 (16)		39 : 61 (21)	41 : 59 (14)
F43Y	49 : 51 (14)	61 : 39 (22)	59 : 41 (41)			50 : 50 (17)	
F43W	49 : 51 (12)	55 : 45 (24)	62 : 38 (38)	52 : 48 (19)			
F43L	47 : 53 (14)	67 : 33 (20)	50 : 50 (41)	50 : 50 (23)	51 : 49 (24)	58 : 42 (18)	
F43I	48 : 52 (13)	59 : 41 (28)	49 : 51 (57)	49 : 51 (19)			
F43T					46 : 54 (24)		
F43H	49 : 51 (13)	51 : 49 (12)		58 : 42 (15)	49 : 51 (13)	48 : 52 (14)	
F43V			49 : 51 (21)	44 : 56 (24)		50 : 50 (13)	

mOCR-Myo. H93G, H64V- 2-OMe, COOMe							
6B	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	30 : 70 (18)	45 : 55 NA	45 : 55 NA	42 : 58 (15)	47 : 53 NA	45 : 55 NA	44 : 56 NA
F43Y	49 : 51 NA						
F43W	49 : 51 NA	55 : 45 (29)					
F43L	47 : 53 NA	55 : 45 (21)					
F43I	50 : 50 NA	52 : 48 (43)					
F43T	51 : 49 NA		51 : 49 (54)				
F43H		57 : 43 (35)					
F43V							

mOCR-Myo. H93G, H64L- 2-OMe, COOMe							
6C	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	26 : 74 (37)	23 : 77 (25)	29 : 71 NA	46 : 54 (12)	77 : 23 (20)		26 : 74 (22)
F43Y		40 : 60 (24)		38 : 62 NA			
F43W	43 : 57 (16)	52 : 48 (23)	51 : 49 NA		48 : 52 (14)	47 : 53 (22)	46 : 54 (15)
F43L	32 : 68 (31)	82 : 18 (55)	48 : 52 (34)	63 : 37 (15)	72 : 28 (46)		
F43I	50 : 50 (29)	42 : 58 NA		52 : 48 (13)			
F43T		54 : 46 NA		51 : 49 (15)			
F43H	63 : 37 (40)	50 : 50 NA	53 : 47 (18)	50 : 50 (15)			
F43V	77 : 23 (20)		24 : 76 (30)	48 : 52 (15)			

mOCR-Myo. H93G, H64I- 2-OMe, COOMe							
6D	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	36 : 64 (17)	37 : 63 NA	34 : 66 (21)	47 : 53 NA			
F43Y		56 : 44 (49)					
F43W	46 : 54 (16)	56 : 44 NA					
F43L	53 : 47 (11)		65 : 35 NA				
F43I	46 : 54 (11)	55 : 45 (25)			49 : 51 (13)		
F43T		56 : 44 NA		48 : 52 (8)			
F43H	51 : 49 (14)	56 : 44 NA					
F43V	41 : 59 (24)		48 : 52 (18)				

mOCR-Myo. H93A, H64A- 2-OMe, COOMe							
6E	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	64 : 36 (17)	44 : 56 (19)	48 : 52 (34)	62 : 38 (7)	58 : 42 (24)	55 : 45 (31)	61 : 39 (32)
F43Y	52 : 48 (12)	54 : 46 (28)		43 : 57 (15)	50 : 50 (20)	56 : 44 (24)	36 : 64 (36)
F43W			59 : 41 (51)	55 : 45 (12)	51 : 49 (24)		54 : 46 (25)
F43L							
F43I						43 : 57 (19)	
F43T	53 : 47 (11)	56 : 44 (23)					
F43H	58 : 42 (9)	60 : 40 (14)	63 : 37 (41)			58 : 42 (20)	
F43V	60 : 40 (18)	56 : 44 (42)	61 : 39 (24)		44 : 56 (31)		

mOCR-Myo. H93A, H64V- 2-OMe, COOMe							
6F	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	70 : 30 (30)	76 : 24 (37)		79 : 21 (21)	76 : 24 (63)	69 : 31 (28)	
F43Y	52 : 48 (6)	76 : 24 (33)	53 : 47 (19)		56 : 44 (11)	65 : 35 (26)	50 : 50 (8)
F43W	56 : 44 (9)			52 : 49 (18)	51 : 50 (10)	66 : 34 (18)	54 : 46 (11)
F43L	53 : 48 (10)	72 : 28 (27)	59 : 41 (32)	58 : 43 (4)			
F43I	52 : 48 (9)	62 : 38 (22)	53 : 47 (28)	51 : 49 (5)			43 : 57 (11)
F43T	51 : 49 (9)					52 : 48 (9)	
F43H	61 39 (5)		68 : 32 (26)			66 : 34 (21)	50 : 50 (12)
F43V							

mOCR-Myo. H93A, H64L- 2-OMe, COOMe							
6G	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	77 : 23 (50)	78 : 22 (27)	64 : 36 NA	76 : 24 (29)	65 : 35 (17)	55 : 45 (7)	36 : 64 (17)
F43Y	66 : 34 (26)	75 : 25 (52)	71 : 29 (43)		54 : 46 (10)	66 : 34 (32)	
F43W	57 : 43 (17)	71 : 29 NA	67 : 33 (25)	62 : 38 (18)	51 : 49 (14)	64 : 36 (20)	44 : 56 (17)
F43L		78 : 22 (32)	68 : 32 (46)	68 : 32 (42)			
F43I	49 : 51 (26)	76 : 24 (16)	55 : 45 (12)	69 : 31 (19)	60 : 40 (15)	49 : 51 (20)	
F43T	51 : 49 (21)	72 : 29 NA	50 : 50 (16)	50 : 50 (8)	50 : 50 (15)		
F43H	68 : 32 (17)		66 : 34 (27)	70 : 30 (17)	52 : 48 (12)		
F43V	57 : 43 (6)	65 : 35 (39)		55 : 45 (25)	64 : 37 (34)	63 : 37 (25)	

mOCR-Myo. H93G, H64A- 2-OMe, COOMe							
6H	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	73 : 27 (40)	76 : 24 (49)	68 : 32 (48)	67 : 33 (19)	69 : 31 (24)		
F43Y	55 : 45 NA	34 : 66 (28)	65 : 35 (34)				
F43W	57 : 43 (11)						
F43L	47 : 53 (13)		60 : 40 (30)				
F43I	55 : 45 (7)	61 : 39 (26)			50 : 50 (11)		
F43T	44 : 56 (12)	78 : 22 (51)	59 : 41 NA	48 : 52 NA			
F43H	56 : 44 (8)	39 : 61 (42)	62 : 38 (30)	55 : 45 NA			
F43V	60 : 40 (8)	50 : 50 (25)	62 : 38 (30)	46 : 54 (10)			

mOCR-Myo 93A, 64L, 43Y, 68A				
6I	**L32X**	**F33X**	**H97X**	**I99X**
V		68:32 (23)		
L			59:41 (21)	
I				
F		70:30 (22)	68:32 (27)	
H				
W			72:28 (42)	
Y			73:27 (59)	

mOCR-Myo 93A, 64L, 43L, 68A				
6J	**L32X**	**F33X**	**H97X**	**I99X**
V		82:18 (30)	68:32 (43)	
L		82:18 (36)		
I	73:27 (42)	78:22 (25)	68:32 (30)	
F	74:26 (36)			
H	54:46 (14)			
W	56:44 (18)	69:31 (21)	77:23 (60)	
Y	65:35 (43)	76:24 (33)	78:22 (66)	

mOCR-Myo 93G, 64L, 68A				
6K	**L32X**	**F33X**	**H97X**	**I99X**
V	25:75 (30)	31:69 (11)		21:79 (14)
L		28:72 (29)	30:70 (24)	54:46 (17)
I		33:67 (15)	50:50 (38)	
F	26:74 (31)			46:54 (7)
H	41:59 (9)			46:54 (11)
W		27:73 (21)		
Y		26:74 (26)	46:54 0	50:50 (11)

mOCR-Myo 93G, 64I, 68A				
6L	**L32X**	**F33X**	**H97X**	**I99X**
V	40:60 (27)	42:58 (13)		
L		41:59 (24)		
I	46:54 (58)			
F	49:51 (46)			
H				
W				
Y	47:53 (18)	44:56 (14)		

mOCR-Myo 93A, 64L, 43W, 68A				
6M	**L32X**	**F33X**	**H97X**	**I99X**
V		67:33 (25)		
L			64:36 (54)	
I		68:32 (33)		
F			65:35 (54)	
H		67:33 (25)		
W		65:35 (29)	76:24 (97)	
Y		67:33 (22)	73:27 (64)	

mOCR-Myo 93A, 64V, 68A. 103C, 108C				
6N	**L32X**	**F33X**	**H97X**	**I99X**
V			67:33 (41)	67:33 (32)
L		67:33 (22)	61:39 (47)	
I	59:41 (35)		62:38 (42)	
F	58:42 (32)			
H		61:39 (9)		49:51 (10)
W	53:47 (12)	63:37 (25)	70:30 (72)	67:33 (11)
Y	72:28 (62)	65:35 (19)	73:27 (69)	51:49 (11)

mOCR-Myo 93G, 64L, 43V				
6P	**L32X**	**F33X**	**H97X**	**I99X**
V		46:54 (19)		28:72 (46)
L				
I				
F				
H				
W			43:57 (41)	
Y		42:58 (26)	38:62 (57)	

mOCR-Myo 93G, 64L, 43L				
6Q	**L32X**	**F33X**	**H97X**	**I99X**
V	39:61 (17)		45:55 (45)	19:81 (61)
L			37:63 (40)	42:58 (30)
I	41:59 (27)			
F	44:56 (14)		31:69 (37)	22:78 (32)
H	44:56 (12)			
W	49:51 (12)		36:64 (29)	
Y	49:51 (18)		29:71 (34)	

mOCR-Myo 93A, 64I, 43H, 68A				
6R	**L32X**	**F33X**	**H97X**	**I99X**
V		64:36 (33)	64:36 (18)	72:28 (37)
L			60:40 (17)	
I		70:30 (15)		
F	55:45 (103)		65:35 (33)	75:25 (107)
H				
W			78:22 (58)	
Y	68:32 (21)	73:27 (35)	76:24 (41)	

mOCR-Myo 93A, 64L, 68A				
6S	**L32X**	**F33X**	**H97X**	**I99X**
V		79:21 (38)	76:24 (98)	75:25 (74)
L			54:46 (92)	
I	72:28 (69)		74:26 (93)	
F			53:47 (26)	
H				
W			66:34 (91)	
Y		69:31 (87)		

mOCR-Myo 93A, 64V, 43Y, 68A				
6T	**L32X**	**F33X**	**H97X**	**I99X**
V	71:29 (52)	60:40 (89)	56:44 (45)	75:25 (56)
L		78:22 (80)	62:38 (72)	58:42 (11)
I	75:25 (68)	69:31 (28)	59:41 (67)	76:24 (58)
F	70:30 (57)		64:36 (62)	
H		70:30 (38)		
W	65:35 (21)	68:32 (48)	78:22 (92)	51:49 (6)
Y		79:21 (77)	61:39 (81)	

mOCR-Myo 93A, 64L, 43L, 33V				
6U	**L32X**	**F33X**	**H97X**	**I99X**
V			73:27 (45)	75:25 (37)
L				
I			66:34 (56)	
F			80:20 (76)	58:42 (8)
H				
W			79:21 (67)	
Y			80:20 (76)	

mOCR-Myo 93A, 64V, 68S				
6V	**L32X**	**F33X**	**H97X**	**I99X**
V	61:39 (72)			
L			73:27 (96)	
I			66:34 (103)	
F			58:42 (89)	
H				
W			59:41 (101)	
Y	56:44 (20)	72:28 (39)		

mOCR-Myo 93A, 64L, 43W, 68A. 97Y				
6W	**L32X**	**F33X**	**H97X**	**I99X**
V	72:28 (67)	72:28 (66)		
L		74:26 (71)		
I	71:29 (66)	75:25 (55)		
F	71:29 (70)			
H	72:28 (79)	68:32 (56)		
W	68:32 (59)			
Y	67:33 (74)	68:32 (66)		

mOCR-Myo 93A, 64V, 68A				
6X	**L32X**	**F33X**	**H97X**	**I99X**
V				63:37 (71)
L		60:40 (49)		
I				
F				
H				
W				
Y		70:30 (57)		

mOCR-Myo 93G, 64L, 68A. 32F				
6Y	**L32X**	**F33X**	**H97X**	**I99X**
V			51:49 (68)	16:84 (31)
L			34:66 (69)	56:44 (14)
I			51:49 (68)	
F				33:67 (30)
H				46:54 (11)
W			39:61 (91)	
Y			25:75 (99)	54:46 (58)

mOCR-Myo 93G, 64L, 43L. 99V				
6Z	L32X	F33X	H97X	I99X
V	35:65 (30)	38:62 (22)		
L		27:73 (39)		
I	34:66 (50)	29:71 (36)		
F	40:60 (30)			
H	35:65 (33)			
W	44:56 (30)	33:67 (36)		
Y	45:55 (42)	27:73 (38)		

mOCR-Myo 93G, 64L, 43L. 99F				
6AA	L32X	F33X	H97X	I99X
V	45:55 (25)	37:63 (10)		
L				
I	31:69 (27)			
F	27:73 (26)			
H	49:51 (17)	47:53 (11)		
W	49:51 (24)	36:64 (26)		
Y	48:52 (22)	33:67 (26)		

mOCR-Myo 93A, 64V, 43Y. 68A, L32I				
6BB	L32X	F33X	H97X	I99X
V				70:30 (57)
L				58:42 (43)
I				
F				29:71 (45)
H				
W				
Y				50:50 (15)

Complete, tabulated results for the directed evolution of Ir(Me)-PIX-mOCR-Myo for the C-H insertion of substrate **6** under the standard, screening reactions conditions. Mutants are designated by the substitutions listed in the table header plus additional mutations indicated in the row and column headers of each table. Data are presented as: enantiomeric ratio (GC yield).

mOCR-Myo. H93G, H64A- 2-OMe, COOMe							
7A	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	54 : 46 (8)	59 : 41 (9)	57 : 43 (10)	56 : 44 (6)			
F43Y	52 : 48 (8)	43 : 57 (10)	46 : 54 (24)				
F43W	51 : 49 (9)		47 : 53 (22)	47 : 53 (7)			
F43L	52 : 48 (10)	39 : 61 (47)	51 : 49 (24)	53 : 47 (9)			
F43I	53 : 47 (9)			51 : 50 (9)			
F43T							
F43H	52 : 48 (9)	50 : 50 (6)		55 : 45 (7)			
F43V							

mOCR-Myo. H93G, H64V- 2-OMe, COOMe							
7B	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	69 : 31 (9)	53 : 47 NA	53 : 47 NA	48 : 52 (15)	48 : 52 NA	48 : 52 NA	55 : 45 (1)
F43Y	52 : 48 NA						
F43W	51 : 49 NA	48 : 52 (20)					
F43L	57 : 43 (8)	54 : 46 (13)					
F43I	54 : 46 NA	53 : 47 (23)					
F43T	54 : 46 NA		51 : 49 (23)				
F43H		44 : 56 (23)					
F43V							

mOCR-Myo. H93G, H64L- 2-OMe, COOMe							
7C	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	78 : 22 (26)	69 : 31 (10)	56 : 44 NA	51 : 49 NA	38 : 63 (10)		66 : 34 (11)
F43Y		56 : 44 (12)		55 : 45 (22)			
F43W	59 : 41 (10)	51 : 49 (10)	47 : 53 NA		52 : 48 (5)		
F43L	90 : 10 (40)	40 : 60 (23)	48 : 52 (19)	52 : 48 (8)	35 : 65 (26)	54 : 46 (8)	56 : 44 (6)
F43I	56 : 44 (19)	58 : 42 NA		46 : 54 (3)			
F43T		52 : 48 NA		46 : 54 (9)			
F43H	50 : 50 (24)	55 : 45 NA	47 : 53 (9)	53 : 47 (15)			
F43V	83 : 17 (60)		80 : 20 (16)	46 : 54 (8)			

mOCR-Myo. H93G, H64I- 2-OMe, COOMe							
7D	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	60 : 40 (17)	58 : 42 (10)	57 : 43 (15)	50 : 50 NA			
F43Y		53 : 47 (23)					
F43W	61 : 39 (9)	47 : 53 NA					
F43L	50 : 50 (8)		45 : 55 NA				
F43I	66 : 34 (10)	53 : 47 (16)			49 : 51 (8)		
F43T		48 : 52 NA		49 : 51 (5)			
F43H	57 : 43 (11)	52 : 48 NA					
F43V	84 : 17 (31)		54 : 47 (8)				

mOCR-Myo. H93A, H64A- 2-OMe, COOMe							
7E	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	53 : 47 (10)		49 : 51 (18)	57 : 44 (2)	62 : 38 (58)		
F43Y		35 : 65 (67)		46 : 54 (15)	51 : 49 (17)	42 : 58 (47)	48 : 52 (46)
F43W			38 : 62 (78)	40 : 60 (15)	48 : 52 (23)		46 : 54 (29)
F43L	49 : 51 (8)						
F43I						55 : 45 (36)	
F43T		48 : 52 (70)					
F43H		51 : 49 (27)	39 : 61 (68)			48 : 52 (37)	
F43V					49 : 51 (27)		

mOCR-Myo. H93A, H64V- 2-OMe, COOMe							
7F	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	48 : 52 (6)	39 : 61 (19)		48 : 52 (5)		52 : 48 (14)	
F43Y	51 : 49 (4)	39 : 61 (19)	46 : 54 (11)		49 : 51 (7)	43 : 57 (29)	47 : 53 (6)
F43W	53 : 47 (4)			40 : 60 (3)	55 : 45 (6)	54 : 46 (13)	55 : 45 (6)
F43L	51 : 49 (6)	43 : 57 (13)	37 : 63 (18)	46 : 54 (2)			
F43I	49 : 51 (5)	45 : 55 (12)	40 : 60 (15)	46 : 54 (2)			53 : 47 (6)
F43T	51 : 49 (5)					53 : 47 (6)	
F43H	52 48 (5)		39 : 61 (13)			47 : 53 (13)	53 : 47 (7)
F43V							

mOCR-Myo. H93A, H64L- 2-OMe, COOMe							
7G	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	39 : 61 (14)	41 : 59 (12)	44 : 56 NA	38 : 62 NA	49 : 51 (12)	41 : 59 (8)	48 : 52 (5)
F43Y	50 : 50 (8)	35 : 65 (21)	31 : 69 (18)		48 : 52 (10)	44 : 56 (16)	
F43W	49 : 51 (10)	33 : 68 NA	31 : 69 (13)	35 : 65 (8)	49 : 51 (8)	36 : 64 (5)	54 : 46 (5)
F43L		40 : 60 (19)	35 : 65 (21)	35 : 65 (18)			
F43I	52 : 48 NA	36 : 64 NA		31 : 69 (10)	37 : 63 (8)		
F43T	59 : 41 NA	38 : 62 NA	51 : 49 (8)		46 : 54 (8)		
F43H	41 : 59 (21)		37 : 63 (11)	35 : 65 (11)	49 : 51 (8)		
F43V	51 : 49 (3)	36 : 64 (24)		39 61 (14)	27 : 73 (19)	52 : 48 (15)	45 : 55 (10)

mOCR-Myo. H93G, H64A- 2-OMe, COOMe							
7H	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	58 : 42 (19)	44 : 56 (28)		50 : 50 (9)	54 : 47 (11)		
F43Y	51 : 49 NA	65 : 35 (15)	39 : 61 (18)				
F43W	52 : 48 (7)						
F43L	56 : 44 (8)	52 : 48 NA	42 : 58 (17)				
F43I	54 : 46 (6)	44 : 56 (13)			50 : 50 (8)		
F43T	47 : 53 (7)	35 : 65 (26)	44 : 56 NA	49 : 51 NA			
F43H	51 : 49 (7)	53 : 47 (22)	42 : 58 (22)	48 : 52 NA			
F43V	51 : 49 (5)	45 : 55 (12)	39 : 61 (14)	40 : 60 (4)			

mOCR-Myo 93A, 64L, 43Y, 68A				
7I	**L32X**	**F33X**	**H97X**	**I99X**
V		38:62 (11)		
L			61:39 (15)	
I				
F			44:56 (23	
H				
W			38:62 (29)	
Y			36:64 (30)	

mOCR-Myo 93A, 64L, 43L, 68A				
7J	**L32X**	**F33X**	**H97X**	**I99X**
V		39:61 (18)	69:31 (24)	
L		35:65 (21)		
I	38:62 (25)	39:61 (15)	68:32 (23)	
F	37:63 (22)			
H	48:52 (9)			
W	45:55 (13)	45:55 (18)	42:58 (30)	
Y	48:52 (32)	43:57 (17)	41:59 (33)	

mOCR-Myo 93G, 64L, 68A				
7K	**L32X**	**F33X**	**H97X**	**I99X**
V	64:36 (17)	61:39 (5)		71:29 (6)
L		66:34 (13)	69:31 (12)	62:38 (10)
I		63:37 (7)	71:29 (24)	
F	67:33 (20)			64:36 (4)
H	59:41 (7)			73:27 (9)
W	56:44 (10)	62:38 (8)		
Y		61:39 (15)	54:46 (3)	70:30 (8)

mOCR-Myo 93G, 64I, 68A				
7L	**L32X**	**F33X**	**H97X**	**I99X**
V	53:47 (15)	61:39 (9)		
L		59:41 (12)		
I	58:42 (33)			
F	55:45 (21)			
H				
W				
Y	57:43 (10)	53:47 (10)		

mOCR-Myo 93A, 64L, 43W, 68A				
7M	**L32X**	**F33X**	**H97X**	**I99X**
V		38:62 (13)	71:29 (38)	
L			41:59 (32)	
I		35:65 (16)		
F			31:69 (29)	
H		43:57 (16)		
W		40:60 (14)	27:73 (51)	
Y		42:58 (12)	25:75 (34)	

mOCR-Myo 93A, 64V, 68A. 103C, 108C				
7N	**L32X**	**F33X**	**H97X**	**I99X**
V			51:49 (28)	55:45 (12)
L		53:47 (9)	56:44 (22)	
I	57:43 (12)		56:44 (25)	
F	54:46 (12)			
H		53:47 (5)		62:38 (10)
W	54:46 (8)	57:43 (22)	40:60 (30)	45:55 (6)
Y	55:45 (36)	58:42 (22)	46:54 (46)	53:47 (8)

mOCR-Myo 93G, 64L, 43V				
7P	**L32X**	**F33X**	**H97X**	**I99X**
V		71:29 (11)		87:13 (58)
L				
I				
F				
H				
W			72:28 (51)	
Y		77:23 (30)	80:20 (51)	

mOCR-Myo 93G, 64L, 43L				
7Q	**L32X**	**F33X**	**H97X**	**I99X**
V	79:21 (20)		89:11 (58)	90:10 (64)
L			89:11 (64)	83:17 (39)
I	82:18 (37)			
F	76:24 (11)		86:14 (40)	92:8 (46)
H	83:17 (18)			
W	58:42 (5)		80:20 (27)	
Y	72:28 (15)		86:14 (38)	

mOCR-Myo 93A, 64I, 43H, 68A				
7R	**L32X**	**F33X**	**H97X**	**I99X**
V		44:56 (17)	52:48 (12)	41:59 (20)
L			52:48 (11)	
I		40:60 (9)		
F	45:55 (71)		47:53 (14)	40:60 (34)
H				
W			37:63 (24)	
Y	38:62 (15)	42:58 (28)	38:62 (21)	

mOCR-Myo 93A, 64L, 68A				
7S	**L32X**	**F33X**	**H97X**	**I99X**
V		39:61 (16)	46:54 (53)	44:56 (58)
L			51:49 (68)	
I	42:58 (37)		51:49 (54)	
F			51:49 (16)	
H				
W			45:55 (55)	
Y		46:54 (34)		

mOCR-Myo 93A, 64V, 43Y, 68A				
7T	**L32X**	**F33X**	**H97X**	**I99X**
V	44:56 (33)	56:44 (66)	59:41 (56)	39:61 (28)
L		36:64 (65)	50:50 (43)	52:48 (5)
I	38:62 (43)	44:56 (17)	56:44 (44)	40:60 (31)
F	33:67 (33)		46:54 (41)	
H		41:59 (23)		
W	39:61 (16)	44:56 (34)	33:67 (61)	55:45 (6)
Y		35:65 (57)	57:43 (52)	

mOCR-Myo 93A, 64L, 43L, 33V				
7U	**L32X**	**F33X**	**H97X**	**I99X**
V			60:40 (29)	41:59 (20)
L				
I			60:40 (36)	
F			42:58 (38)	53:47 (6)
H				
W			39:61 (24)	
Y			38:62 (38)	

mOCR-Myo 93A, 64V, 68S				
7V	**L32X**	**F33X**	**H97X**	**I99X**
V	55:45 (48)			
L			40:60 (58)	
I			58:42 (69)	
F			55:45 (69)	
H				
W			47:53 (61)	
Y	53:47 (13)	51:49 (28)		

mOCR-Myo 93A, 64L, 43W, 68A. 97Y				
7W	**L32X**	**F33X**	**H97X**	**I99X**
V	32:68 (42)	33:67 (43)		
L		29:71 (43)		
I	29:71 (44)	31:69 (63)		
F	29:71 (47)			
H	30:70 (55)	38:62 (67)		
W	42:58 (65)	80:20 (56)		
Y	41:59 (43)	36:64 (36)		

mOCR-Myo 93A, 64V, 68A				
7X	**L32X**	**F33X**	**H97X**	**I99X**
V				53:47 (33)
L		55:45 (28)		
I				
F				
H				
W				
Y		49:51 (37)		

mOCR-Myo 93G, 64L, 68A. 32F				
7Y	**L32X**	**F33X**	**H97X**	**I99X**
V			59:41 (36)	63:37 (35)
L			57:43 (59)	48:52 (16)
I			59:41 (49)	
F				79:21 (20)
H				63:37 (11)
W			56:44 (63)	
Y			61:39 (54)	59:41 (35)

mOCR-Myo 93G, 64L, 43L. 99V				
7Z	**L32X**	**F33X**	**H97X**	**I99X**
V	80:20 (35)	78:22 (37)		
L		88:12 (50)		
I	84:16 (58)	86:14 (43)		
F	78:22 (27)			
H	79:21 (22)			
W	62:38 (14)	85:15 (44)		
Y	72:28 (36)	82:18 (59)		

mOCR-Myo 93G, 64L, 43L. 99F				
7AA	**L32X**	**F33X**	**H97X**	**I99X**
V	74:26 (17)	65:35 (21)		
L				
I	86:14 (31)			
F	88:12 (25)			
H	58:42 (13)	57:43 (12)		
W	53:47 (22)	84:16 (28)		
Y	61:39 (14)	81:19 (31)		

mOCR-Myo 93A, 64V, 43Y. 68A, L32I				
7BB	**L32X**	**F33X**	**H97X**	**I99X**
V				39:61 (57)
L				46:54 (20)
I				
F				79:21 (41)
H				
W				
Y				62:38 (11)

Complete, tabulated results for the directed evolution of Ir(Me)-PIX-mOCR-Myo for the C-H insertion of substrate **7** under the standard, screening reactions conditions. Mutants are designated by the substitutions listed in the table header plus additional mutations indicated in the row and column headers of each table. Data are presented as: enantiomeric ratio (GC yield).

mOCR-Myo. H93G, H64A- 2-OMe, COOMe							
8A	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	58 : 42 (40)	69 : 31 (5)	60 : 40 (32)	49 : 51 (26)		63 : 37 (21)	55 : 45 (18)
F43Y	54 : 46 (19)	80 : 20 (38)	68 : 32 (61)			49 : 51 (25)	
F43W	52 : 48 (20)	70 : 30 (32)	67 : 33 (53)	47 : 53 (7)			
F43L	53 : 47 (19)	69 : 31 (24)	51 : 49 (55)	55 : 45 (37)	52 : 48 (42)	71 : 29 (27)	
F43I	54 : 46 (20)	71 : 29 (33)	49 : 51 (88)	48 : 52 (27)			
F43T					45 : 55 (38)		
F43H	53 : 47 (19)	55 : 45 (37)		51 : 49 (19)	47 : 53 (21)	55 : 45 (18)	
F43V			55 : 45 (25)	46 : 54 (43)		52 : 48 (13)	

mOCR-Myo. H93G, H64V- 2-OMe, COOMe							
8B	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	60 : 40 (15)	54 : 46 NA	49 : 51 (51)	65 : 35 (16)	50 : 50 0	53 : 47 NA	51 : 49 NA
F43Y	54 : 46 NA						
F43W	52 : 48 NA	59 : 41 (33)					
F43L	53 : 47 (19)	60 : 40 (29)					
F43I	55 : 45 NA	60 : 40 (40)					
F43T	53 : 47 NA		54 : 46 (45)				
F43H		67 : 33 (41)					
F43V							

mOCR-Myo. H93G, H64L- 2-OMe, COOMe							
8C	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	59 : 41 (13)	55 : 45 (21)	51 : 49 NA	56 : 44 NA	65 : 36 (32)		51 : 50 (21)
F43Y		67 : 33 (27)		50 : 50 0			
F43W	52 : 48 (21)	64 : 36 (29)	55 : 45 NA		54 : 46 (23)	58 : 42 (22)	46 : 54 (19)
F43L	54 : 46 (36)	90 : 10 (54)	57 : 43 (42)	64 : 36 (19)	78 : 22 (53)		
F43I	51 : 49 (42)	61 : 39 NA		50 : 50 0			
F43T		63 : 37 NA		55 : 45 (21)			
F43H	64 : 36 (52)	66 : 34 NA	53 : 47 (28)	47 : 53 (39)			
F43V	61 : 39 (25)		60 : 40 (26)	62 : 38 (17)			

mOCR-Myo. H93G, H64I- 2-OMe, COOMe							
8D	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	69 : 31 (46)	64 : 36 (17)	56 : 44 (45)	54 : 46 (34)			
F43Y		70 : 30 (43)					
F43W	57 : 43 (19)	69 : 31 (29)					
F43L	58 : 42 (15)		71 : 29 NA				
F43I	61 : 39 (14)	65 : 35 (28)			49 : 51 (21)		
F43T		71 : 29 NA		54 : 46 (14)			
F43H	58 : 42 (18)	70 : 30 NA					
F43V	60 : 40 (25)		55 : 45 (26)				

mOCR-Myo. H93A, H64A- 2-OMe, COOMe							
8E	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	69 : 31 (26)	48 : 52 (31)	73 : 27 (38)	45 : 55 (9)	56 : 44 (65)	71 : 29 (34)	64 : 36 (48)
F43Y	63 : 37 (20)	83 : 18 (61)		51 : 49 (15)	56 : 44 (21)		
F43W			71 : 29 (55)	53 : 47 (14)	54 : 46 (24)		55 : 45 (31)
F43L							
F43I						63 : 38 (30)	
F43T	65 : 35 (17)	74 : 26 (56)					
F43H	44 : 56 (23)	78 : 22 (25)	78 : 22 (55)				
F43V	68 : 32 (26)	72 : 28 (64)	72 : 28 (41)				

mOCR-Myo. H93A, H64V- 2-OMe, COOMe							
8F	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	80 : 20 (19)	86 : 14 (36)		63 : 37 (28)	66 : 34 (71)	88 : 12 (20)	
F43Y	54 : 46 (12)	89 : 11 (50)	61 : 39 (23)		55 : 45 (21)	80 : 20 (51)	49 : 51 (14)
F43W	65 : 35 (12)			61 : 39 (10)	52 : 48 (26)	75 : 25 (34)	
F43L	54 : 46 (18)	82 : 18 (37)	65 : 35 (46)	52 : 48 (6)			
F43I	55 : 45 (14)	76 : 24 (31)	60 : 40 (42)	36 : 64 (7)			31 : 69 (22)
F43T	52 : 48 (12)					62 : 38 (21)	
F43H	66 34 (12)		80 : 20 (41)			81 : 19 (50)	38 : 62 (33)
F43V							

mOCR-Myo. H93A, H64L- 2-OMe, COOMe							
8G	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	83 : 17 (74)	84 : 16 (36)		79 : 21 NA	63 : 37 (43)	64 : 36 (13)	57 : 43 (28)
F43Y	76 : 24 (33)	85 : 15 (66)	81 : 19 (55)		57 : 43 (27)	83 : 17 (40)	
F43W	60 : 40 (17)	78 : 22 NA	76 : 24 (39)	54 : 46 (26)	53 : 47 (16)	65 : 35 (29)	23 : 77 (82)
F43L		83 : 17 (35)	76 : 24 (55)	76 : 24 (73)			
F43I	51 : 49 (23)	73 : 27 NA	60 : 40 (22)	58 : 42 (31)	56 : 44 (8)	33 : 67 (52)	
F43T	52 : 48 (30)	75 : 25 NA	51 : 49 (19)	51 : 49 (8)	56 : 44 NA		
F43H	70 : 30 (49)		73 : 27 (39)	60 : 40 (24)	50 : 50 (13)		
F43V	67 : 33 (12)	76 : 24 (62)		45 : 55 (49)	45 : 55 NA	64 : 36 (30)	79 : 21 (26)

mOCR-Myo. H93G, H64A- 2-OMe, COOMe							
8H	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	68 : 32 (51)	84 : 16 (56)	84 : 16 (76)	57 : 43 (27)	61 : 40 (48)		
F43Y	50 : 50 (46)		72 : 28 (56)				
F43W	68 : 32 (15)	60 : 40 (29)					
F43L	54 : 46 (18)	54 : 46 NA	67 : 33 (42)				
F43I	67 : 33 (13)	67 : 33 (51)			48 : 52 (16)		
F43T	35 : 65 (23)	74 : 26 (53)	63 : 37 NA	50 : 50 (46)			
F43H	65 : 35 (14)	80 : 20 (82)	73 : 27 (36)	54 : 46 (34)			
F43V		53 : 47 (33)	70 : 30 (38)	51 : 49 (13)			

mOCR-Myo 93A, 64L, 43Y, 68A				
8I	**L32X**	**F33X**	**H97X**	**I99X**
V		74:26 (39)		
L			66:34 (32)	
I				
F		77:23 (26)	76:24 (38)	
H				
W			73:27 (58)	
Y			85:15 (76)	

mOCR-Myo 93A, 64L, 43L, 68A				
8J	**L32X**	**F33X**	**H97X**	**I99X**
V		80:20 (33)	65:35 (37)	
L		82:18 (34)		
I	76:24 (44)	74:26 (34)	64:36 (24)	
F	77:23 (40)			
H	64:36 (23)			
W	65:35 (28)	72:28 (32)	81:19 (61)	
Y	72:28 (63)	79:21 (32)	81:19 (66)	

mOCR-Myo 93G, 64L, 68A				
8K	**L32X**	**F33X**	**H97X**	**I99X**
V	58:42 (30)	59:41 (12)		41:59 (10)
L		58:42 (28)	61:39 (23)	61:39 (11)
I		63:37 (14)	65:35 (46)	
F	57:43 (42)			53:47 (6)
H	58:42 (11)			46:54 (12)
W	51:49 (20)	54:46 (19)		
Y		54:46 (31)	52:48 (5)	57:43 (13)

mOCR-Myo 93G, 64I, 68A				
8L	**L32X**	**F33X**	**H97X**	**I99X**
V	61:39 (43)	63:37 (17)		
L		60:40 (42)		
I	67:33 (71)			
F	66:34 (64)			
H				
W				
Y	64:36 (26)	58:42 (32)		

mOCR-Myo 93A, 64L, 43W, 68A				
8M	**L32X**	**F33X**	**H97X**	**I99X**
V		72:28 (43)		
L			67:33 (90)	
I		71:29 (53)		
F			76:24 (57)	
H		71:29 (58)		
W		69:31 (61)	79:21 (92)	
Y		73:27 (53)	78:22 (57)	

mOCR-Myo 93A, 64V, 68A. 103C, 108C				
8N	**L32X**	**F33X**	**H97X**	**I99X**
V			69:31 (61)	81:19 (52)
L		77:23 (31)	62:38 (71)	
I	73:27 (37)		65:35 (74)	
F	73:27 (27)			
H		73:27 (23)		52:48 (17)
W	53:47 (17)	78:22 (38)	76:24 (75)	65:35)16)
Y	76:24 (69)	78:22 (30)	81:19 (71)	52:48 (12)

mOCR-Myo 93G, 64L, 43V				
8P	**L32X**	**F33X**	**H97X**	**I99X**
V		56:44 (13)		55:45 (38)
L				
I				
F				
H				
W			60:40 (61)	
Y		57:43 (33)	60:40 (63)	

mOCR-Myo 93G, 64L, 43L				
8Q	**L32X**	**F33X**	**H97X**	**I99X**
V	57:43 (21)		56:44 (56)	47:53 (28)
L			55:45 (44)	52:48 (28)
I	58:42 (37)			
F	61:39 (13)		58:42 (47)	45:55 (15)
H	61:39 (16)			
W	52:48 (14)		56:44 (35)	
Y	67:33 (21)		55:45 (37)	

mOCR-Myo 93A, 64I, 43H, 68A				
8R	**L32X**	**F33X**	**H97X**	**I99X**
V		75:25 (43)	72:28 (43)	85:15 (57)
L			68:32 (40)	
I		81:19 (28)		
F	76:24 (96)		84:16 (48)	87:13 (95)
H				
W			88:12 (82)	
Y	74:26 (43)	78:22 (71)	89:11 (58)	

mOCR-Myo 93A, 64L, 68A				
8S	**L32X**	**F33X**	**H97X**	**I99X**
V		83:17 (46)	79:21 (80)	84:16 (66)
L			77:23 (55)	
I	81:19 (62)		80:20 (72)	
F			55:45 (34)	
H				
W			83:17 (94)	
Y		86:14 (67)		

mOCR-Myo 93A, 64V, 43Y, 68A				
8T	**L32X**	**F33X**	**H97X**	**I99X**
V	78:22 (57)	65:35 (89)	61:39 (53)	81:19 (72)
L			66:34 (76)	73:27 (15)
I	76:24 (81)	80:20 (44)	64:36 (83)	84:16 (72)
F	83:17 (63)		80:20 (47)	
H		83:17 (53)		
W	77:23 (40)	81:19 (40)	84:16 (60)	58:42 (15)
Y		35:65 (57)	64:36 (50)	

mOCR-Myo 93A, 64L, 43L, 33V				
8U	**L32X**	**F33X**	**H97X**	**I99X**
V			70:30 (60)	78:22 (51)
L				
I			67:33 (58)	
F			81:19 (76)	55:45 (17)
H				
W			82:18 (76)	
Y			80:20 (95)	

mOCR-Myo 93A, 64V, 68S				
8V	**L32X**	**F33X**	**H97X**	**I99X**
V	83:17 (75)			
L			85:15 (55)	
I			77:23 (103)	
F			81:19 (107)	
H				
W			76:24 (105)	
Y	62:38 (27)	85:15 (63)		

mOCR-Myo 93A, 64L, 43W, 68A. 97Y				
8W	**L32X**	**F33X**	**H97X**	**I99X**
V	75:25 (84)	78:22 (86)		
L		79:21 (99)		
I	73:27 (85)	80:20 (58)		
F	70:30 (86)			
H	78:22 (104)	82:18 (54)		
W	62:38 (104)			
Y	53:47 (89)	3.5:21.5 (77)		

mOCR-Myo 93A, 64V, 68A				
8X	**L32X**	**F33X**	**H97X**	**I99X**
V				53:47 (33)
L		55:45 (28)		
I				
F				
H				
W				
Y		49:51 (37)		

mOCR-Myo 93G, 64L, 68A. 32F				
8Y	**L32X**	**F33X**	**H97X**	**I99X**
V			59:41 (83)	62:38 (58)
L			54:46 (72)	61:39 (20)
I			59:41 (81)	
F				53:47 (27)
H				50:50 (21)
W			59:41 (80)	
Y			52:48 (88)	61:39 (63)

mOCR-Myo 93G, 64L, 43L. 99V				
8Z	**L32X**	**F33X**	**H97X**	**I99X**
V	54:46 (37)	47:53 (22)		
L		49:51 (26)		
I	53:47 (48)	52:48 (28)		
F	57:43 (27)			
H	54:46 (24)			
W	51:49 (55)	49:51 (30)		
Y	68:32 (48)	51:49 (52)		

mOCR-Myo 93G, 64L, 43L. 99F				
8AA	**L32X**	**F33X**	**H97X**	**I99X**
V	49:51 (21)	49:51 (23)		
L				
I	47:53 (21)			
F	46:54 (21)			
H	49:51 (34)	50:50 (11)		
W	51:49 (24)	48:52 (24)		
Y	52:48 (39)	49:51 (19)		

mOCR-Myo 93A, 64V, 43Y. 68A, L32I				
8BB	**L32X**	**F33X**	**H97X**	**I99X**
V				79:21 58)
L				68:32 (45)
I				
F				58:42 (50)
H				
W				
Y				55:45 (33)

Table S8A-BB. Complete, tabulated results for the directed evolution of Ir(Me)-PIX-mOCR-Myo for the C-H insertion of substrate **8** under the standard, screening reactions conditions. Mutants are designated by the substitutions listed in the table header plus additional mutations indicated in the row and column headers of each table. Data are presented as: enantiomeric ratio (GC yield).

mOCR-Myo. H93G, H64A- 2-OMe, COOMe							
9A	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	36 : 64 (7)	21 : 79 (5)	34 : 66 (6)	43 : 57 (2)		22 : 78 (9)	41 : 59 (4)
F43Y		23 : 77 (11)	29 : 71 (13)			47 : 53 (3)	
F43W		27 : 73 (8)	64 : 36 (19)	45 : 55 (3)			
F43L		51 : 49 (25)	30 : 70 (16)	47 : 53 (3)	51 : 49 (4)	47 : 53 (13)	
F43I		28 : 72 (9)	38 : 62 (20)	47 : 53 (3)			
F43T	47 : 53 (2)				50 : 50 (3)		
F43H		50 : 50 0		45 : 55 (3)	48 : 52 (2)	37 : 63 (2)	
F43V			41 : 59 (7)	42 : 58 (5)		42 : 58 (1)	

mOCR-Myo. H93G, H64V- 2-OMe, COOMe							
9B	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	34 : 66 (3)			36 : 64 (4)			
F43Y							
F43W		44 : 56 NA					
F43L		32 : 68 NA					
F43I		33 : 68 NA					
F43T			35 : 65 NA				
F43H		41 : 60 NA					
F43V							

mOCR-Myo. H93G, H64L- 2-OMe, COOMe							
9C	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	35 : 65 (3)	20 : 80 (5)	30 : 70 (15)	48 : 52 (7)	53 : 47 (6)		33 : 67 (5)
F43Y	71 : 29 (51)	26 : 74 (11)		48 : 52 NA			
F43W		38 : 62 (16)		25 : 75 (5)	43 : 57 (23)	41 : 59 (5)	44 : 56 (3)
F43L	31 : 69 (5)	37 : 63 (15)	40 : 60 (25)	45 : 55 (4)	68 : 32 (32)		
F43I	49 : 51 NA	31 : 69 (8)		42 : 58 (1)			
F43T		39 : 61 (4)		49 : 51 (4)			
F43H	50 : 50 NA	32 : 68 (7)	38 : 62 (19)	49 : 51 (6)			
F43V	32 : 68 (4)		35 : 65 (15)	57 : 43 (4)			

mOCR-Myo. H93G, H64I- 2-OMe, COOMe							
9D	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	34 : 66 (7)	20 : 80 (14)	29 : 71 (22)				
F43Y		27 : 73 NA					
F43W	43 : 57 (5)	41 : 59 NA					
F43L							
F43I	38 : 62 (2)	26 : 74 NA					
F43T							
F43H	44 : 56 (2)						
F43V	29 : 71 (15)		27 : 73 (10)				

mOCR-Myo. H93A, H64A- 2-OMe, COOMe							
9E	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	47 : 53 (4)	49 : 51 (5)	60 : 40 (8)	53 : 47 (3)	38 : 62 (6)	63 : 37 (12)	42 : 58 (6)
F43Y	56 : 44 0	79 : 21 (29)		55 : 45 (4)	49 : 51 (3)		
F43W			93 : 7 (71)	72 : 28 (7)	48 : 52 (4)		55 : 45 (8)
F43L							
F43I							
F43T	57 : 43 (3)	67 : 33 (18)					
F43H	55 : 45 (3)	22 : 78 (7)	53 : 47 (14)				
F43V	52 : 48 (6)	63 : 37 (32)	66 : 34 (8)				

mOCR-Myo. H93A, H64V- 2-OMe, COOMe							
9F	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	45 : 55 (2)	49 : 51 (13)		57 : 43 (2)	47 : 53 (14)	43 : 57 (4)	
F43Y		70 : 30 (29)	41 : 59 (3)		61 : 39 (3)	67 : 33 (14)	52 : 49 (1)
F43W	57 : 43 (3)			80 : 20 (7)	61 : 39 NA	81 : 19 (22)	52 : 49 (3)
F43L	61 : 39 (17)		55 : 45 (24)	53 : 47 (1)			
F43I	59 : 41 (14)	55 : 45 (22)		53 : 47 (1)			46 : 54 (1)
F43T							
F43H	52 : 48 (1)		53 : 47 (31)				45 : 55 (1)
F43V							

mOCR-Myo. H93A, H64L- 2-OMe, COOMe							
9G	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	39 : 61 (19)	49 : 51 (8)	46 : 54 (7)	53 : 47 (9)	49 : 51 (17)	42 : 58 (2)	37 : 63 (3)
F43Y	47 : 53 (9)	61 : 39 (33)	65 : 35 (17)		47 : 53 (4)	56 : 44 (14)	
F43W	52 : 48 NA	77 : 23 (26)	76 : 24 (29)	60 : 40 (12)	47 : 53 (2)	69 : 31 (27)	39 : 61 (7)
F43L		54 : 46 (11)	32 : 68 (14)	54 : 46 (23)			
F43I	52 : 48 NA	46 : 54 (4)	53 : 47 (6)	53 : 47 (6)	52 : 48 (4)	41 : 59 (5)	
F43T	52 : 48 NA	52 : 48 (34)	49 : 51 (2)	54 : 46 (1)	52 : 48 (4)		
F43H	47 : 53 (10)		56 : 45 (12)	57 : 43 (5)			
F43V	52 : 48 (2)	62 : 38 (23)		53 : 47 (4)	55 : 45 (8)	50 : 50 (3)	54 : 46 (6)

mOCR-Myo. H93G, H64A- 2-OMe, COOMe							
9H	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	41 : 59 (8)	52 : 48 (25)	55 : 45 (21)				
F43Y		46 : 54 (6)					
F43W	54 : 46 (13)						
F43L							
F43I		57 : 43 NA					
F43T	48 : 52 (2)	55 : 45 (17)					
F43H	49 : 51 (11)	34 : 66 (5)					
F43V	49 : 51 (14)	41 : 59 (10)	55 : 45 (18)				

mOCR-Myo 93A, 64L, 43Y, 68A				
9I	**L32X**	**F33X**	**H97X**	**I99X**
V		57:43 (6)		
L			47:53 (4)	
I				
F			59:41 (12)	
H				
W			74:26 (18)	
Y			68:32 (13)	

mOCR-Myo 93A, 64L, 43L, 68A				
9J	**L32X**	**F33X**	**H97X**	**I99X**
V		57:43 (3)	46:54 (10)	
L		55:45 (10)		
I	70:30 (26)	54:46 (11)	48:52 (6)	
F	53:47 NA			
H	49:51 (6)			
W	45:55 (3)	60:40 (18)	58:42 (35)	
Y	52:48 (10)	55:45 (13)	54:46 (49)	

mOCR-Myo 93G, 64L, 68A				
9K	**L32X**	**F33X**	**H97X**	**I99X**
V	21:79 (12)	25:75 (3)		28:72 (2)
L		24:76 (8)	26:74 (4)	37:63 (6)
I		24:76 (4)	43:57 (12)	
F	15:85 (16)			41:59 (2)
H	31:69 (2)			45:55 (1)
W	24:76 (5)	25:75 (3)		
Y		30:70 (7)	43:57 (3)	43:57 (3)

mOCR-Myo 93G, 64I, 68A				
9L	**L32X**	**F33X**	**H97X**	**I99X**
V	22:78 (18)	23:77 (5)		
L		23:77 (13)		
I	30:70 (61)			
F	18:82 (44)			
H				
W				
Y	22:78 (11)	35:65 (6)		

mOCR-Myo 93A, 64L, 43W, 68A				
9M	**L32X**	**F33X**	**H97X**	**I99X**
V				
L			67:33 (23)	
I		65:35 (17)		
F			75:25 (54)	
H		58:42 (10)		
W		61:39 (15)	82:18 (100)	
Y		57:43 (6)	79:21 (56)	

mOCR-Myo 93A, 64V, 68A. 103C, 108C				
9N	**L32X**	**F33X**	**H97X**	**I99X**
V			51:49 (19)	46:54 (12)
L		57:43 (10)	46:54 (28)	
I	49:51 (11)		48:52 (15)	
F	46:54 (8)			
H		44:56 (8)		52:48 (4)
W	51:49 (6)	49:51 (12)	56:44 (59)	53:47 (3)
Y	51:49 (24)	49:51 (6)	53:47 (86)	49:51 (2)

mOCR-Myo 93G, 64L, 43V				
9P	**L32X**	**F33X**	**H97X**	**I99X**
V		48:52 (2)		40:60 (11)
L				
I				
F				
H				
W			37:63 (16)	
Y		42:58 (6)	32:68 (14)	

mOCR-Myo 93G, 64L, 43L				
9Q	**L32X**	**F33X**	**H97X**	**I99X**
V	46:54 (8)		44:56 (17)	35:65 (8)
L			43:57 (11)	44:56 (13)
I	44:56 (18)			
F	47:53 (20)		41:59 (17)	41:59 (4)
H	47:53 (9)			
W	49:51 (1)		45:55 (19)	
Y	47:53 (13)		42:58 (14)	

mOCR-Myo 93A, 64I, 43H, 68A				
9R	**L32X**	**F33X**	**H97X**	**I99X**
V		50:50 (23)	48:52 (9)	47:53 (19)
L			61:39 (6)	
I		53:47 (10)		
F	51:49 (57)		40:60 (14)	40:60 (41)
H				
W			44:56 (33)	
Y	53:47 (6)	54:46 (27)	52:48 (26)	

mOCR-Myo 93A, 64L, 68A				
9S	**L32X**	**F33X**	**H97X**	**I99X**
V		47:53 (11)	44:56 -12	41:59 (21)
L			51:49 (23)	
I	40:60 (29)		43:57 (33)	
F			49:51 (12)	
H				
W			37:63 (39)	
Y		28:72 (17)		

mOCR-Myo 93A, 64V, 43Y, 68A				
9T	**L32X**	**F33X**	**H97X**	**I99X**
V	71:29 (57)	62:38 (81)	59:41 (28)	64:36 (35)
L		76:24 (93)	60:40 (39)	53:47 (3)
I	83:17 (81)	65:35 (16)	60:40 (28)	72:28 (49)
F	73:27 (63)		58:42 (41)	
H		56:44 (18)		
W	62:38 (6)	53:47 (46)	73:27 (54)	51:49 (1)
Y		71:29 (89)	59:41 (49)	

mOCR-Myo 93A, 64L, 43L, 33V				
9U	**L32X**	**F33X**	**H97X**	**I99X**
V			50:50 (18)	42:58 (11)
L				
I			51:49 (11)	
F			56:44 (44)	49:51 (18)
H				
W			59:41 (37)	
Y			57:43 (68)	

mOCR-Myo 93A, 64V, 68S				
9V	**L32X**	**F33X**	**H97X**	**I99X**
V	39:61 (9)			
L			36:64 (32)	
I			38:62 (35)	
F			48:52 (53)	
H				
W			51:49 (57)	
Y	47:53 (4)	46:54 (18)		

mOCR-Myo 93A, 64L, 43W, 68A. 97Y				
9W	**L32X**	**F33X**	**H97X**	**I99X**
V				40:60 (33)
L		50:50 (19)		
I				
F				
H				
W				
Y		48:52 (30)		

mOCR-Myo 93A, 64V, 68A				
9X	**L32X**	**F33X**	**H97X**	**I99X**
V				53:47 (33)
L		55:45 (28)		
I				
F				
H				
W				
Y		49:51 (37)		

mOCR-Myo 93G, 64L, 68A. 32F				
9Y	**L32X**	**F33X**	**H97X**	**I99X**
V			36:64 (15)	13:87 (13)
L			19:81 (20)	45:55 (30)
I			40:60 (20)	
F				47:53 (24)
H				49:51 (17)
W			16:84 (62)	
Y			13:87 (63)	46:54 (28)

mOCR-Myo 93G, 64L, 43L. 99V				
9Z	L32X	F33X	H97X	I99X
V	31:69 (6)	46:54 (1)		
L		32:68 (5)		
I	27:73 (11)	33:67 (4)		
F	35:65 (4)			
H	47:53 (28)			
W	47:53 (20)	33:67 (5)		
Y	42:58 (11)	29:71 (4)		

mOCR-Myo 93G, 64L, 43L. 99F				
9AA	L32X	F33X	H97X	I99X
V	41:59 (4)	49:51 (1)		
L				
I	39:61 (4)			
F	39:61 (4)			
H	49:51 (4)	50:50 (2)		
W	50:50 (27)	48:52 (4)		
Y	51:49 (34)	49:51 (12)		

mOCR-Myo 93A, 64V, 43Y. 68A, L32I				
9BB	L32X	F33X	H97X	I99X
V				82:18 (74)
L				59:41 (15)
I				
F				40:60 (19)
H				
W				
Y				49:51 (16)

Complete, tabulated results for the directed evolution of Ir(Me)-PIX-mOCR-Myo for the C-H insertion of substrate **9** under the standard, screening reactions conditions. Mutants are designated by the substitutions listed in the table header plus additional mutations indicated in the row and column headers of each table. Data are presented as: enantiomeric ratio (GC yield).

mOCR-Myo. H93G, H64A- 2-OMe, COOMe							
10A	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	49 : 51 (20)	48 : 52 (18)	48 : 52 (18)	43 : 57 (19)		46 : 54 (NA)	46 : 54 (NA)
F43Y	49 : 51 (15)	41 : 59 (27)	70 : 30 (56)			46 : 54 (NA)	
F43W	50 : 50 (15)	63 : 37 (NA)	62 : 38 (48)	49 : 52 (23)			
F43L	50 : 50 (15)	61 : 39 (NA)	67 : 33 (62)	49 : 51 (41)	49 : 52 (NA)	58 : 42 (NA)	
F43I	51 : 49 (15)	57 : 43 (NA)	65 : 35 (NA)	44 : 56 (39)			
F43T		60 40			44 : 57 (NA)		
F43H	50 : 50 (16)	51 : 49 (16)		47 : 53 (22)	50 : 51 (NA)	53 : 47 (NA)	
F43V		52 48	71 : 29 (NA)	38 : 62 (NA)		51 : 50 (NA)	

mOCR-Myo. H93G, H64V- 2-OMe, COOMe							
10B	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	47 : 53 (14)	47 : 53 NA	47 : 53 NA	NA (9)	48 : 52 NA	46 : 54 NA	48 : 52 NA
F43Y	51 : 49 NA						
F43W	48 : 52 NA	NA (33)					
F43L	51 : 49 NA	NA (25)					
F43I	50 : 50 NA	56 : 44 (40)					
F43T	50 : 50 NA		NA (43)				
F43H		NA (33)					
F43V							

mOCR-Myo. H93G, H64L- 2-OMe, COOMe							
10C	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	47 : 53 (28)	31 : 69 (21)	46 : 54 (31)	49 : 51 (7)	50 : 50 (25)		34 : 66 (20)
F43Y	77 : 23 (38)			47 : 53 NA	45 : 55 (18)		
F43W		52 : 48 (27)	51 : 49 NA		48 : 52 (21)	49 : 51 (24)	48 : 52 (18)
F43L	42 : 58 (31)	80 : 20 (34)	54 : 46 (28)	61 : 39 (4)	72 : 28 (43)		
F43I	NA (24)	53 : 47 (26)		54 : 46 (1)			
F43T		59 : 41 (8)		50 : 50 (4)			
F43H	NA (45)	57 : 43 (21)	57 : 43 (28)	NA (6)			49 : 51 (16)
F43V	50 : 50 (22)		31 : 69 (23)	51 : 49 (4)			

mOCR-Myo. H93G, H64I- 2-OMe, COOMe							
10D	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	48 : 52 (36)	46 : 54 (18)	NA (26)	48 : 52 (30)			
F43Y		55 : 45 NA					
F43W	49 : 51 (14)	55 : 45 (24)					
F43L	NA (10)		65 : 35 NA				
F43I	49 : 51 (11)	53 : 47 NA			55 : 45 (18)		
F43T		58 : 42 NA		50 : 50 (14)			
F43H	51 : 49 (15)	56 : 44 NA					
F43V	51 : 49 (25)		52 : 48 (25)				

mOCR-Myo. H93A, H64A- 2-OMe, COOMe							
10E	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	52 : 48 (14)	48 : 52 (NA)	49 : 51 (26)	48 : 52 (7)		47 : 53 (NA)	54 : 46 (NA)
F43Y	52 : 48 (NA)	60 : 40 (NA)				59 : 41 (NA)	39 : 61 (NA)
F43W	52 : 48 (NA)		67 : 33 (NA)				
F43L	55 : 45 (13)						
F43I							
F43T	53 : 47 (NA)						
F43H	57 : 43 (NA)		57 : 43 (NA)			61 : 40 (NA)	
F43V	51 : 49 (NA)	54 : 46 (NA)			37 : 63 (NA)		

mOCR-Myo. H93A, H64V- 2-OMe, COOMe							
10F	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	53 : 47 (14)	71 : 29 (39)		73 : 27 (14)	57 : 43 (36)	69 : 31 (26)	
F43Y	55 : 45 (8)	76 : 24 (29)	57 : 43 (29)		55 : 45 (17)	70 : 30 (21)	54 : 46 (8)
F43W	49 : 51 (12)			56 : 44 (5)	48 : 52 (10)	65 : 35 (13)	52 : 48 (11)
F43L	48 : 52 (10)	64 : 36 (36)	55 : 45 (43)	54 : 46 (4)			
F43I	56 : 44 (11)	60 : 40 (37)	53 : 47 (41)	50 : 50 (7)			45 : 55 (11)
F43T	49 : 51 (10)					48 : 52 (11)	
F43H			68 : 32 (34)			69 : 31 (23)	54 : 46 (12)
F43V							

mOCR-Myo. H93A, H64L- 2-OMe, COOMe							
10G	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	68 : 32 (52)	71 : 29 (26)	62 : 38 (44)	64 : 36 (29)	27 : 73 (17)		42 : 58 (20)
F43Y	61 : 39 (27)	70 : 30 (37)	67 : 33 (32)		51 : 49 (27)	65 : 35 (28)	
F43W	53 : 47 (24)	63 : 37 (20)	64 : 36 (39)	48 : 52 (18)	47 : 53 (11)	51 : 49 (20)	44 : 57 (21)
F43L		69 : 31 (27)	62 : 38 (29)	63 : 37 (47)			
F43I	NA (20)	73 : 27 (16)	55 : 45 (29)	61 : 39 (23)	62 : 38 (15)	48 : 52 (30)	
F43T	NA (18)	66 : 34 (41)	56 : 44 (25)	47 : 53 (6)	52 : 48 (27)		
F43H	59 : 41 (37)		63 : 37 (32)	61 : 39 (24)	51 : 49 (9)		
F43V	53 : 47 (10)	66 : 34 (44)		51 : 49 (38)	56 : 44 (39)	59 : 41 (26)	63 : 37 (20)

mOCR-Myo. H93G, H64A- 2-OMe, COOMe							
10H	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	65 : 35 (24)	65 : 35 (47)	65 : 35 (33)	48 : 52 (16)	49 : 52 (19)		
F43Y	51 : 49 (30)	41 : 59 (20)	60 : 40 (49)				
F43W	NA (11)						
F43L	51 : 49 (14)	46 : 54 NA	52 : 48 (33)				
F43I	54 : 46 (8)	58 : 42 (27)			52 : 48 (14)		
F43T	47 : 53 (15)	68 : 32 (50)		49 : 51 (29)			
F43H	56 : 44 (9)	33 : 67 (41)	51 : 49 (24)	52 : 48 (32)			
F43V	NA (10)	61 : 39 (40)	59 : 41 (43)	48 : 52 (11)			

mOCR-Myo 93A, 64L, 43Y, 68A				
10I	**L32X**	**F33X**	**H97X**	**I99X**
V		75:25 NA		
L			58:42 NA	
I				
F		67:33 NA	66:34 NA	
H				
W			70:30 NA	
Y			76:24 NA	

mOCR-Myo 93A, 64L, 43L, 68A				
10J	**L32X**	**F33X**	**H97X**	**I99X**
V		76:24 (36)	55:45 (42)	
L		76:24 (34)		
I	62:38 (42)	72:28 NA	62:38 (37)	
F	71:29 (41)			
H	58:42 NA			
W	66:34 NA	66:34 NA	60:40 (52)	
Y	50:50 0	69:31 (35)	68:32 (54)	

mOCR-Myo 93G, 64L, 68A				
10K	**L32X**	**F33X**	**H97X**	**I99X**
V	34:66 (29)	41:59 (10)		26:74 (11)
L		39:61 (29)	35:65 (19)	62:38 (16)
I		41:59 (13)	43:57 (43)	
F	42:58 (35)			54:46 (6)
H	46:54 (9)			53:47 (6)
W		65:35 (15)		
Y		35:65 (25)	43:57 NA	54:46 (10)

mOCR-Myo 93G, 64I, 68A				
10L	**L32X**	**F33X**	**H97X**	**I99X**
V	46:54 NA	50:50 (12)		
L		51:49 NA		
I	54:46 NA			
F	59:41 NA			
H				
W				
Y	64:36 NA	51:49 NA		

mOCR-Myo 93A, 64L, 43W, 68A				
10M	**L32X**	**F33X**	**H97X**	**I99X**
V		67:33 NA		
L			56:44 NA	
I		69:31 NA		
F			55:45 (53)	
H		67:33 NA		
W		66:34 NA	59:41 (56)	
Y		70:30 NA	60:40 (54)	

mOCR-Myo 93A, 64V, 68A. 103C, 108C				
10N	**L32X**	**F33X**	**H97X**	**I99X**
V			59:41 NA	65:35 NA
L		63:37 NA	58:42 NA	
I	62:38 NA		60:40 NA	
F	61:39 NA			
H		57:43 NA		52:48 (11)
W	55:45 (25)	60:40 NA	65:35 NA	64:36 (13)
Y	64:36 (57)	63:37 (19)	66:34 NA	52:48 NA

mOCR-Myo 93G, 64L, 43V				
10P	**L32X**	**F33X**	**H97X**	**I99X**
V		48:52 (9)		43:57 (43)
L				
I				
F				
H				
W			55:45 (31)	
Y		55:45 (21)	47:53 (38)	

mOCR-Myo 93G, 64L, 43L				
10Q	**L32X**	**F33X**	**H97X**	**I99X**
V	45:55 (12)		60:40 (29)	27:73 (23)
L			52:48 (23)	62:38 (14)
I	48:52 (19)			
F	52:48 (19)		43:57 (28)	33:67 (11)
H	50:50 (10)			
W	49:51 (6)		45:55 (23)	
Y	61:39 (12)		37:63 (21)	

mOCR-Myo 93A, 64I, 43H, 68A				
10R	**L32X**	**F33X**	**H97X**	**I99X**
V		70:30 (25)	62:38 (29)	73:27 (44)
L			53:47 (31)	
I		72:28 (20)		
F	53:47 (23)		69:31 (39)	71:29 (98)
H				
W			74:26 (61)	
Y	80:20 (33)	71:29 (45)	74:26 (50)	

mOCR-Myo 93A, 64L, 68A				
10S	**L32X**	**F33X**	**H97X**	**I99X**
V		70:30 (27)	71:29 (93)	76:24 (65)
L			56:44 (100)	
I	66:34 (45)		68:32 (78)	
F			59:41 (11)	
H				
W			63:37 (105)	
Y		73:27 (71)		

mOCR-Myo 93A, 64V, 43Y, 68A				
10T	**L32X**	**F33X**	**H97X**	**I99X**
V	73:27 (36)	58:42 (75)	55:45 (63)	79:21 (38)
L		77:23 (59)	56:44 (57)	61:39 (8)
I	67:33 (68)	71:29 (25)	56:44 (56)	77:23 (41)
F	79:21 (43)		72:28 (40)	
H		71:29 (29)		
W	74:26 (32)	69:31 (36)	81:19 (64)	55:45 (5)
Y		77:23 (66)	60:40 (54)	

mOCR-Myo 93A, 64L, 43L, 33V				
10U	**L32X**	**F33X**	**H97X**	**I99X**
V			64:36 (71)	..5:28.5 (41)
L				
I			62:38 (61)	
F			67:33 (81)	61:39 (6)
H				
W			NA (87)	
Y			66:34 (102)	

mOCR-Myo 93A, 64V, 68S				
10V	**L32X**	**F33X**	**H97X**	**I99X**
V	64:36 (93)			
L			71:29 (59)	
I			61:39 (58)	
F			61:39 (59)	
H				
W			62:38 (65)	
Y	63:37 (37)	65:35 (60)		

mOCR-Myo 93A, 64L, 43W, 68A. 97Y				
10W	**L32X**	**F33X**	**H97X**	**I99X**
V	51:49 (91)	68:32 (89)		
L		63:37 (80)		
I	49:51 (91)	67:33 (NA)		
F	63:37 (95)			
H	66:34 (61)	64:36 (NA)		
W	64:36 (70)	64:36		
Y	72:28 (86)	66:34 (81)		

mOCR-Myo 93A, 64V, 68A				
10X	**L32X**	**F33X**	**H97X**	**I99X**
V				65:35 (84)
L		53:47 (72)		
I				
F				
H				
W				
Y		64:36 (73)		

mOCR-Myo 93G, 64L, 68A. 32F				
10Y	**L32X**	**F33X**	**H97X**	**I99X**
V			51:49 (62)	39:61 (30)
L			42:58 (67)	55:45 (22)
I			52:48 (67)	
F				49:51 (17)
H				46:54 (8)
W			55:45 (91)	
Y			43:57 (73)	53:47 (41)

mOCR-Myo 93G, 64L, 43L. 99V				
10Z	**L32X**	**F33X**	**H97X**	**I99X**
V	NA (45)	47:53 (11)		
L		NA (39)		
I	NA (57)	NA (31)		
F	48:52 (32)			
H	53:47 (19)			
W	46:54 (24)	NA (43)		
Y	56:44 (49)	42:58 (20)		

mOCR-Myo 93G, 64L, 43L. 99F				
10AA	**L32X**	**F33X**	**H97X**	**I99X**
V	NA (22)	51:49 (7)		
L				
I	NA (4)			
F	NA (4)			
H	48:52 (4)	51:49 (44)		
W	50:50 (27)	NA (35)		
Y	51:49 (34)	49:51 (14)		

mOCR-Myo 93A, 64V, 43Y. 68A, L32I				
10BB	**L32X**	**F33X**	**H97X**	**I99X**
V				58:42 (74)
L				54:46 (15)
I				
F				42:58 (19)
H				
W				50:50 0
Y				55:45 (14)

Table S10A-BB. Complete, tabulated results for the directed evolution of Ir(Me)-PIX-mOCR-Myo for the C-H insertion of substrate **10** under the standard, screening reactions conditions. Mutants are designated by the substitutions listed in the table header plus additional mutations indicated in the row and column headers of each table. Data are presented as: enantiomeric ratio (GC yield).

mOCR-Myo. H93G, H64A- 2-OMe, COOMe							
11A	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	48 : 52 (32)	60 : 40 (7)	59 : 41 (30)	43 : 57 (21)		63 : 37 (16)	58 : 42 (13)
F43Y	53 : 47 (22)	74 : 26 (35)	72 : 28 (65)			48 : 52 NA	
F43W	56 : 44 NA	65 : 35 (31)	65 : 35 (55)	48 : 52 (29)			
F43L	55 : 45 (1)	62 : 38 (25)	48 : 52 (57)	48 : 52 (43)	51 : 49 (41)	64 : 36 (23)	
F43I	56 : 44 (0)	67 : 33 (38)	48 : 52 (NA)	42 : 58 (44)			
F43T	61 40 NA	60 : 40			45 : 55 (31)		
F43H	56 : 44 NA	55 : 45 (15)		46 : 54 (27)	54 : 46 (15)	60 : 40 (14)	
F43V			54 : 46 (34)	45 : 55 (31)		61 : 39 (14)	

mOCR-Myo. H93G, H64V- 2-OMe, COOMe							
11B	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	47 : 53 (12)	51 : 49 NA	52 : 48 NA	52 : 48 (7)	52 : 48 NA	51 : 49 NA	47 : 53 NA
F43Y	47 : 53 NA						
F43W	51 : 49 NA	55 : 45 (44)					
F43L	53 : 47 NA	60 : 40 (38)					
F43I	53 : 47 NA	59 : 41 (52)					
F43T	51 : 49 NA		55 : 45 (49)				
F43H		61 : 39 (52)					
F43V							

mOCR-Myo. H93G, H64L- 2-OMe, COOMe							
11C	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	50 : 50 (38)	50 : 50 (20)	52 : 48 (18)	52 : 48 (12)	49 : 51 (33)		47 : 53 (23)
F43Y	84 : 16 (50)			52 : 48 (50)	50 : 50 (20)		
F43W		56 : 44 (32)	49 : 51 NA		49 : 51 (27)	54 : 46 (28)	46 : 54 (21)
F43L	54 : 46 (36)	82 : 18 (25)	50 : 51 (42)	51 : 49 (21)	77 : 23 (51)		
F43I	57 : 43 (37)	54 : 46 (20)		57 : 43 (15)			
F43T		60 : 40 (5)		71 : 29 (6)			
F43H	60 : 40 (65)	60 : 40 (17)	52 : 48 (39)	41 : 59 (42)			53 : 47 (0)
F43V	57 : 43 (32)		52 : 48 (28)	50 : 50 (20)			

mOCR-Myo. H93G, H64I- 2-OMe, COOMe							
11D	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	49 : 51 (44)	NA (18)	54 : 46 (34)	49 : 51 (43)			
F43Y		61 : 39 (43)					
F43W	55 : 45 (20)	58 : 42 (45)					
F43L	57 : 43 (18)		60 : 41 NA				
F43I	59 : 41 (18)	63 : 37 (32)			52 : 48 (26)		
F43T		62 : 38 NA		59 : 41 NA			
F43H	57 : 43 (20)	60 : 40 NA					
F43V	60 : 40 (44)		48 : 52 (33)				

mOCR-Myo. H93A, H64A- 2-OMe, COOMe							
11E	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	50 : 50 (20)	47 : 53 (26)	64 : 36 (18)	41 : 59 (6)	49 : 51 (43)	63 : 38 (19)	55 : 45 (30)
F43Y	64 : 36 (18)	75 : 25 (42)		53 : 47 (14)	51 : 49 (27)	68 : 32 (27)	52 : 48 (43)
F43W	60 : 40 (30)		63 : 37 (33)	51 : 49 (3)	52 : 48 (25)		52 : 48 (40)
F43L	55 : 45 (18)						
F43I							
F43T	66 : 34 (NA)	64 : 36 (48)					
F43H	65 : 35 (13)	74 : 26 (33)	74 : 26 (55)			72 : 28 (29)	
F43V	57 : 43 (27)	61 : 39 (NA)	49 : 51 (25)				

mOCR-Myo. H93A, H64V- 2-OMe, COOMe							
11F	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	54 : 46 (22)	72 : 28 (30)		52 : 48 (18)	NA (42)	78 : 22 (25)	
F43Y	53 : 47 (16)	80 : 20 (31)	53 : 47 (25)		47 : 53 (19)	59 : 41 (37)	55 : 45 (9)
F43W	53 : 47 (12)			51 : 49 (6)	48 : 52 (16)	67 : 33 (20)	51 : 49 (16)
F43L	52 : 48 (13)	67 : 33 (38)	58 : 42 (49)	NA (6)			
F43I	51 : 49 (20)	64 : 36 (31)	52 : 48 (38)	39 : 61 (9)			33 : 67 (31)
F43T	51 : 49 (16)					55 : 45 (20)	
F43H	59 : 41 (11)		69 : 31 (38)			78 : 22 (37)	49 : 51 (26)
F43V							

mOCR-Myo. H93A, H64L- 2-OMe, COOMe							
11G	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	69 : 31 (58)	72 : 28 (28)	62 : 38 (13)	62 : 38 (24)	54 : 46 (23)		51 : 50 (25)
F43Y	63 : 37 (30)	75 : 25 (39)	70 : 30 (57)		51 : 49 (32)	64 : 36 (33)	
F43W	52 : 48 (30)	66 : 34 (22)	66 : 34 (43)	41 : 59 (27)	48 : 52 (11)	47 : 53 (21)	32 : 68 (38)
F43L		65 : 35 (30)	60 : 40 (46)	59 : 41 (51)			
F43I	51 : 49 (22)	68 : 32 (14)	52 : 48 36)	45 : 55 (33)	43 : 57 (20)	37 : 63 (45)	
F43T	58 : 42 (24)	62 : 38 (28)	50 : 50 (31)	63 : 37 (6)	40 : 60 (35)		
F43H	61 : 39 (44)		63 : 37 (35)	NA (28)	NA (0)		
F43V	53 : 47 (14)	71 : 29 (46)		38 : 62 (26)	39 : 61 (47)	56 : 44 (31)	74 : 26 (26)

mOCR-Myo. H93G, H64A- 2-OMe, COOMe							
11H	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	59 : 41 (41)	71 : 29 (41)	55 : 45 (41)	52 : 48 (23)	53 : 47 (33)		
F43Y	51 : 49 (56)	62 : 38 (23)	66 : 34 (44)				
F43W	55 : 45 (12)						
F43L		47 : 53 NA	59 : 41 (44)				
F43I	72 : 28 (17)	58 : 42 (41)			49 : 51 (25)		
F43T	42 : 58 (22)	62 : 38 (48)	54 : 46 NA	49 : 51 (46)			
F43H	56 : 44 (13)	25 : 75 (69)	63 : 37 (34)	52 : 48 (51)			
F43V	57 : 43 (19)	52 : 48 (46)	58 : 42 (15)	47 : 53 (16)			

mOCR-Myo 93A, 64L, 43Y, 68A				
11I	**L32X**	**F33X**	**H97X**	**I99X**
V		72:28 (45)		
L			56:44 (33)	
I				
F		69:31 (20)	68:32 (27)	
H				
W			71:29 (52)	
Y			75:25 (79)	

mOCR-Myo 93A, 64L, 43L, 68A				
11J	**L32X**	**F33X**	**H97X**	**I99X**
V		64:36 (42)	54:46 (53)	
L		69:31 (38)		
I	62:38 (80)	65:35 (34)	53:47 (44)	
F	68:32 (44)			
H	60:40 (30)			
W	63:37 (38)	59:41 (32)	65:35 (68)	
Y	73:27 (50)	66:34 (44)	63:37 (55)	

mOCR-Myo 93G, 64L, 68A				
11K	**L32X**	**F33X**	**H97X**	**I99X**
V	53:47 (31)	55:45 (13)		42:58 (11)
L		54:46 (34)	56:44 -23	51:49 (19)
I		57:43 (17)	41:59 -43	
F	61:39 (49)			52:48 (8)
H	59:41 (14)			52:48 (10)
W		52:48 (13)		
Y		52:48 (27)	44:56 -6	55:45 (12)

mOCR-Myo 93G, 64I, 68A				
11L	**L32X**	**F33X**	**H97X**	**I99X**
V	68:32 (24)	60:40 (18)		
L		61:39 (24)		
I	63:37 (51)			
F	51:49 (47)			
H				
W				
Y	69:31 (33)	57:43 (24)		

mOCR-Myo 93A, 64L, 43W, 68A				
11M	**L32X**	**F33X**	**H97X**	**I99X**
V		67:33 (41)		
L			59:41 (80)	
I		85:15 (43)		
F			62:38 (64)	
H		71:29 (48)		
W		70:30 (38)	69:31 (76)	
Y		73:27 (33)	62:38 (63)	

mOCR-Myo 93A, 64V, 68A. 103C, 108C				
11N	**L32X**	**F33X**	**H97X**	**I99X**
V			61:39 (69)	69:31 (38)
L		67:33 (31)	59:41 (76)	
I	63:37 (42)		61:39 (76)	
F	67:33 (36)			
H		65:35 (22)		48:52 (16)
W	51:49 (16)	71:29 (29)	64:36 (64)	59:41 (18)
Y	64:36 (66)	61:39 (25)	71:29 (88)	52:48 (16)

mOCR-Myo 93G, 64L, 43V				
11P	**L32X**	**F33X**	**H97X**	**I99X**
V		52:48 (12)		66:34 (20)
L				
I				
F				
H				
W			65:35 (52)	
Y		59:41 (26)	61:39 (59)	

mOCR-Myo 93G, 64L, 43L				
11Q	**L32X**	**F33X**	**H97X**	**I99X**
V	56:44 (73)		53:47 (55)	
L			53:47 (64)	53:47 (18)
I	56:44 (64)			
F	59:41 (58)		54:46 (56)	49:51 (14)
H	65:35 (59)			
W	55:45 (51)		56:44 (64)	
Y	73:27 (74)		55:45 (58)	

mOCR-Myo 93A, 64I, 43H, 68A				
11R	**L32X**	**F33X**	**H97X**	**I99X**
V		74:26 (35)	66:34 (4)	76:24 (4)
L			67:33 (5)	
I		75:25 (3)		
F	71:29 (102)		78:22 (52)	81:19 (70)
H				
W			84:16 (77)	
Y	79:21 (4)	76:24 (53)	77:23 (59)	

mOCR-Myo 93A, 64L, 68A				
11S	**L32X**	**F33X**	**H97X**	**I99X**
V		70:30 (28)	76:24 (68)	82:18 (65)
L			71:29 (87)	
I	70:30 (54)		72:28 (64)	
F			65:35 (27)	
H				
W			72:28 (83)	
Y		81:19 (68)		

mOCR-Myo 93A, 64V, 43Y, 68A				
11T	**L32X**	**F33X**	**H97X**	**I99X**
V	73:27 (54)	59:41 (66)	59:41 (97)	61:39 (61)
L		78:22 (59)	58:42 (99)	54:46 (16)
I	66:34 (78)	68:32 (47)	57:43 (95)	74:26 (68)
F	79:21 (74)		78:22 (60)	
H		71:29 (65)		
W	72:28 (47)	78:22 (62)	79:21 (99)	52:48 (21)
Y		75:25 (104)	60:40 (91)	

mOCR-Myo 93A, 64L, 43L, 33V				
11U	**L32X**	**F33X**	**H97X**	**I99X**
V			60:40 (64)	72:28 (42)
L				
I			65:35 (66)	
F			69:31 (83)	NA (6)
H				
W			71:29 (83)	
Y			66:34 (94)	

mOCR-Myo 93A, 64V, 68S				
11V	**L32X**	**F33X**	**H97X**	**I99X**
V	79:21 (53)			
L			78:22 (53)	
I			67:33 (92)	
F			74:26 (94)	
H				
W			70:30 (91)	
Y	67:33 (15)	82:18 (32)		

mOCR-Myo 93A, 64L, 43W, 68A. 97Y				
11W	**L32X**	**F33X**	**H97X**	**I99X**
V	58:42 (8)	67:33 (8)		
L		64:36 (8)		
I	57:43 (8)	72:28 (53)		
F	62:38 (8)			
H	69:31 (79)	73:27 (76)		
W	61:39 (67)	73:27		
Y	60:40 (9)	72:28 (46)		

mOCR-Myo 93A, 64V, 68A				
11X	**L32X**	**F33X**	**H97X**	**I99X**
V				70:30 (63)
L		71:29 (56)		
I				
F				
H				
W			61:39 (59)	
Y		74:26 (56)		

mOCR-Myo 93G, 64L, 68A. 32F				
11Y	**L32X**	**F33X**	**H97X**	**I99X**
V			56:44 (67)	73:27 (39)
L			54:46 (66)	56:44 (26)
I			56:44 (69)	
F				61:39 (16)
H				46:54 (8)
W			65:35 (77)	
Y			60:40 (71)	57:43 (58)

mOCR-Myo 93G, 64L, 43L. 99V				
11Z	**L32X**	**F33X**	**H97X**	**I99X**
V	56:44 (39)	50:50 (16)		
L		53:47 (28)		
I	56:44 (54)	53:47 (34)		
F	58:42 (34)			
H	59:41 (26)			
W	53:47 (32)	5:38.5 (36)		
Y	75:25 (58)	64:36 (25)		

mOCR-Myo 93G, 64L, 43L. 99F				
11AA	**L32X**	**F33X**	**H97X**	**I99X**
V	52:48 (24)	51:49 (14)		
L				
I	54:46 (22)			
F	52:48 (23)			
H	53:47 (17)	55:45 (15)		
W	49:51 (18)	51:49 (40)		
Y	58:42 (35)	51:49 (18)		

mOCR-Myo 93A, 64V, 43Y. 68A, L32I				
11BB	**L32X**	**F33X**	**H97X**	**I99X**
V				67:33 (73)
L				61:39 (36)
I				
F				61:39 (52)
H				
W				
Y				57:43 (16)

Table S11A-BB. Complete, tabulated results for the directed evolution of Ir(Me)-PIX-mOCR-Myo for the C-H insertion of substrate **11** under the standard, screening reactions conditions. Mutants are designated by the substitutions listed in the table header plus additional mutations indicated in the row and column headers of each table. Data are presented as: enantiomeric ratio (GC yield).

mOCR-Myo. H93G, H64A- 2-OMe, COOMe							
12A	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	49 : 51 (14)	33 : 67 (5)	44 : 56 (18)	53 : 47 (5)		49 : 51 (8)	52 : 48 (5)
F43Y		33 : 67 (10)	76 : 24 (15)			53 : 47 (8)	
F43W	53 : 47 (22)	68 : 32 (16)	54 : 46 (21)	56 : 44 (10)			
F43L	52 : 48 (23)	51 : 49 (8)	43 : 57 (24)	56 : 44 (15)	50 : 50 (12)	58 : 42 (10)	
F43I	53 : 47 (22)	52 : 48 (12)	47 : 53 (14)	57 : 43 (15)			
F43T		57 43			50 : 50 (13)		
F43H	53 : 47 (27)	64 : 36 (12)		63 : 37 (10)	51 : 49 (6)	53 : 47 (6)	
F43V		65 35	51 : 49 (0)	60 : 40 (30)		51 : 49 (0)	

mOCR-Myo. H93G, H64V- 2-OMe, COOMe							
12B	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	48 : 52 (14)	50 : 50 NA	50 : 50 NA	51 : 49 (3)	47 : 53 NA	47 : 53 NA	47 : 53 NA
F43Y	48 : 53 NA						
F43W	49 : 51 NA	49 : 51 (17)					
F43L		46 : 54 (17)					
F43I	49 : 51 NA	48 : 52 (!4)					
F43T	50 : 50 NA		47 : 53 (20)				
F43H		53 : 47 (21)					
F43V							

mOCR-Myo. H93G, H64L- 2-OMe, COOMe							
12C	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	49 : 51 (14)	NA (20)	44 : 56 (18)	53 : 47 (7)	47 : 53 (13)		46 : 54 NA
F43Y				51 : 49 (26)	51 : 49 (10)		
F43W	51 : 49 (12)	45 : 55 (17)	42 : 58 NA		54 : 46 NA	45 : 55 (15)	49 : 51 (12)
F43L	53 : 47 (17)	36 : 64 (15)	33 : 67 (21)	46 : 54 (10)	76 : 24 (28)		
F43I	49 : 51 (13)	39 : 61 (12)		45 : 55 (0)			
F43T		45 : 55 (6)		53 : 47 (12)			
F43H	51 : 49 (22)	53 : 47 (10)	44 : 56 (12)	51 : 49 (0)			52 : 48 (25)
F43V	60 : 40 (11)		49 : 51 (6)	45 : 55 (11)			

mOCR-Myo. H93G, H64I- 2-OMe, COOMe							
12D	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	49 : 51 (19)	42 : 58 (18)	47 : 53 (15)	50 : 50 (23)			
F43Y		51 : 49 (19)					
F43W	53 : 47 (6)	48 : 52 (17)	49 : 51 (13)				
F43L			47 : 53 NA				
F43I	66 : 34 (9)	48 : 52 (16)			56 : 44 NA		
F43T		53 : 47 NA		57 : 43 (15)			
F43H	53 : 47 (9)	54 : 46 NA					
F43V	57 : 43 (15)		33 : 67 (13)				

mOCR-Myo. H93A, H64A- 2-OMe, COOMe							
12E	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	49 : 51 (11)	20 : 80 (10)	36 : 64 (6)	50 : 50 (6)	53 : 47 (6)	48 : 52 (9)	51 : 49 (9)
F43Y	51 : 49 (0)	59 : 41 (!4)		59 : 41 (5)	56 : 44 (9)	65 : 35 (9)	51 : 49 (10)
F43W			42 : 58 (13)	55 : 45 (4)	52 : 48 (13)		53 : 47 (11)
F43L							
F43I							
F43T	51 : 49 (4)	50 : 50 (18)					
F43H	57 : 43 (5)	65 : 35 (!4)	52 : 48 (14)			70 : 30 (11)	
F43V	52 : 48 (12)	52 : 48 (NA)	51 : 49 (10)		47 : 53 (13)		

mOCR-Myo. H93A, H64V- 2-OMe, COOMe							
12F	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	51 : 49 (8)	34 : 66 14)		50 : 50 (2)	49 : 51 (14)	46 : 54 (10)	
F43Y	50 : 50 (5)	69 : 31 (13)	44 : 56 (27)		57 : 43 (5)	66 : 34 (22)	47 : 53 (4)
F43W	51 : 49 (4)			56 : 44 (6)	53 : 47 (5)	68 : 32 (8)	49 : 51 (6)
F43L	51 : 49 (7)	69 : 31 (13)	44 : 56 (13)	50 : 50 (1)			
F43I	51 : 49 (9)	54 : 46 (24)	47 : 53 (17)	50 : 50 (3)			50 : 50 (11)
F43T	50 : 50 (8)					63 : 37 (16)	
F43H			61 : 39 (15)			77 : 23 (22)	55 : 45 (11)
F43V							

mOCR-Myo. H93A, H64L- 2-OMe, COOMe							
12G	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	58 : 42 (19)	44 : 56 (15)	52 : 48 (8)	38 : 62 (15)	53 : 47 (12)		49 : 51 (8)
F43Y	54 : 46 (15)	55 : 45 (17)	57 : 43 (22)		52 : 48 (16)	57 : 43 (13)	
F43W		46 : 54 (11)	69 : 31 (15)	59 : 41 (7)	53 : 47 (0)	57 : 43 (8)	39 : 61 (20)
F43L		43 : 57 (12)	39 : 61 (18)	41 : 59 (21)			
F43I	51 : 49 (13)	60 : 40 (5)	41 : 59 (18)	45 : 55 (13)	52 : 48 (5)	49 : 51 (16)	
F43T	49 : 51 (22)	44 : 56 (18)	50 : 50 (8)	52 : 48 (5)	49 : 51 (8)		
F43H	41 : 59 (31)		56 : 44 (16)	60 : 40 (11)	52 : 48 (0)		
F43V	64 : 36 (7)	58 : 42 (36)		49 : 51 (15)	47 : 53 (25)	61 : 39 (19)	63 : 37 (19)

mOCR-Myo. H93G, H64A- 2-OMe, COOMe							
12H	**68V**	**V68A**	**V68G**	**V68F**	**V68Y**	**V68S**	**V68T**
43F	52 : 48 (12)	40 : 60 (21)	52 : 48 (16)	52 : 48 (10)	51 : 49 (8)		
F43Y	50 : 50 (22)	60 : 40 (17)	54 : 46 (26)				
F43W	41 : 59 (4)						
F43L		54 : 46 NA	51 : 49 (13)				
F43I	51 : 49 (13)	51 : 49 (15)			50 : 50 (10)		
F43T	39 : 61 (6)	69 : 31 (26)	43 : 57 NA	52 : 48 (16)			
F43H	51 : 49 (4)	44 : 56 (27)	43 : 57 (16)	53 : 47 (22)			
F43V	56 : 44 (7)	53 : 47 (11)	47 : 53 (17)	58 : 42 (4)			

mOCR-Myo 93A, 64L, 43Y, 68A				
12I	**L32X**	**F33X**	**H97X**	**I99X**
V		48:52 (16)		
L			41:59 (10)	
I				
F		51:49 (12)	57:43 (18)	
H				
W			62:38 (24)	
Y			74:26 (22)	

mOCR-Myo 93A, 64L, 43L, 68A				
12J	**L32X**	**F33X**	**H97X**	**I99X**
V		31:69 (16)	37:63 (18)	
L		41:59 (15)		
I	32:68 (15)	29:71 (18)	35:65 (16)	
F	39:61 (15)			
H	55:45 (13)			
W	56:44 (14)	34:66 (14)	59:41 (26)	
Y	35:65 (15)	39:61 (17)	55:45 (20)	

mOCR-Myo 93G, 64L, 68A				
12K	**L32X**	**F33X**	**H97X**	**I99X**
V	40:60 (21)	32:68 (8)		45:55 (4)
L		41:59 (22)	35:65 -8	47:53 (10)
I		37:63 (9)	45:55 -16	
F	34:66 (25)			47:53 (3)
H	42:58 (7)			45:55 (3)
W		42:58 (7)		
Y		44:56 (17)	50:50 -3	43:57 (7)

mOCR-Myo 93G, 64I, 68A				
12L	**L32X**	**F33X**	**H97X**	**I99X**
V	40:60 (11)	41:59 (8)		
L		45:55 (14)		
I	35:65 (15)			
F	36:64 (14)			
H				
W				
Y	32:68 (13)	46:54 (11)		

mOCR-Myo 93A, 64L, 43W, 68A				
12M	**L32X**	**F33X**	**H97X**	**I99X**
V		46:54 (16)		
L			43:57 (18)	
I		51:49 (12)		
F			55:45 (26)	
H		44:56 (18)		
W		46:54 (16)	63:37 (41)	
Y		46:54 (5)	59:41 (38)	

mOCR-Myo 93A, 64V, 68A. 103C, 108C				
12N	**L32X**	**F33X**	**H97X**	**I99X**
V			41:59 (32)	38:62 (21)
L		60:40 (23)	41:59 (37)	
I	31:69 (28)		42:58 (33)	
F	30:70 (24)			
H		55:45 (17)		48:52 (8)
W	46:54 (16)	44:56 (22)	37:63 (42)	51:49 (8)
Y	42:58 (24)	52:48 (9)	29:71 (52)	47:53 (7)

mOCR-Myo 93G, 64L, 43V				
12P	**L32X**	**F33X**	**H97X**	**I99X**
V		53:47 (8)		61:39 (58)
L				
I				
F				
H				
W			62:38 (22)	
Y		59:41 (9)	66:34 (29)	

mOCR-Myo 93G, 64L, 43L				
12Q	**L32X**	**F33X**	**H97X**	**I99X**
V	55:45 (73)		52:48 (12)	66:34 (11)
L			46:54 (10)	43:57 (10)
I	55:45 (64)			
F	48:52 (58)		57:43 (41)	54:46 (6)
H	.5:42.5 (59)			
W	48:52 (51)		53:47 (40)	
Y	38:62 (74)		55:45 (34)	

mOCR-Myo 93A, 64I, 43H, 68A				
12R	**L32X**	**F33X**	**H97X**	**I99X**
V		55:45 (15)	65:35 (15)	61:39 (16)
L			51:49 (16)	
I		62:38 (7)		
F	42:58 (NA)		65:35 (23)	
H				
W			74:26 (28)	
Y	47:53 (14)	60:40 (21)	77:23 (20)	

mOCR-Myo 93A, 64L, 68A				
12S	**L32X**	**F33X**	**H97X**	**I99X**
V		24:76 (15)	32:68 (NA)	35:65 (NA)
L			57:43 (NA)	
I	21:79 (22)		29:71 (NA)	
F			45:55 (NA)	
H				
W			47:53 (NA)	
Y		37:63 (NA)		

mOCR-Myo 93A, 64V, 43Y, 68A				
12T	**L32X**	**F33X**	**H97X**	**I99X**
V	53:47 (11)		46:54 (43)	63:37 (15)
L		69:31 (48)	52:48 (22)	50:50 (11)
I	40:60 (16)	58:42 (16)	54:46 (25)	64:36 (18)
F	55:45 (15)		69:31 (27)	
H		53:47 (24)		
W	52:48 (13)	72:28 (20)	73:27 (38)	48:52 (5)
Y		71:29 (61)	55:45 (38)	

mOCR-Myo 93A, 64L, 43L, 33V				
12U	**L32X**	**F33X**	**H97X**	**I99X**
V			26:74 (30)	37:63 (17)
L				
I			27:73 (20)	
F			39:61 (35)	52:48 (7)
H				
W			56:44 (42)	
Y			38:62 (46)	

mOCR-Myo 93A, 64V, 68S				
12V	**L32X**	**F33X**	**H97X**	**I99X**
V	33:67 (8)			
L			47:53 (17)	
I			41:59 (15)	
F			40:60 (20)	
H				
W			54:46 (28)	
Y	43:57 (5)	61:39 (9)		

mOCR-Myo 93A, 64L, 43W, 68A. 97Y				
12W	**L32X**	**F33X**	**H97X**	**I99X**
V	58:42 (23)	57:43 (18)		
L		58:42 (26)		
I	45:55 (27)	64:36 (41)		
F	51:49 (34)			
H	69:31 (27)	54:46 (33)		
W	59:41 (30)	55:45 (43)		
Y	48:52 (42)	52:48 (29)		

mOCR-Myo 93A, 64V, 68A				
12X	**L32X**	**F33X**	**H97X**	**I99X**
V				33:67 (33)
L		29:71 (31)		
I				
F				
H				
W			59:41 (23)	
Y		35:65 (4%)		

mOCR-Myo 93G, 64L, 68A. 32F				
12Y	**L32X**	**F33X**	**H97X**	**I99X**
V			33:67 (6)	39:61 (7)
L			22:78 (9)	38:62 (17)
I			34:66 (5)	
F				49:51 (21)
H				46:54 (8)
W			21:79 (15)	
Y			21:79 (7)	45:55 (23)

mOCR-Myo 93G, 64L, 43L. 99V				
12Z	**L32X**	**F33X**	**H97X**	**I99X**
V	54:46 (13)	51:49 (4)		
L		53:47 (7)		
I	53:47 (18)	54:46 (10)		
F	55:45 (11)			
H	49:51 (13)			
W	52:48 (12)	54:46 (9)		
Y	52:48 (10)	60:40 (3)		

mOCR-Myo 93G, 64L, 43L. 99F				
12AA	**L32X**	**F33X**	**H97X**	**I99X**
V	49:51 (8)	52:48 (1)		
L				
I	53:47 (10)			
F	54:46 (9)			
H	53:47 (9)	48:52 (4)		
W	55:45 (12)	50:50 (7)		
Y	57:43 (16)	53:47 (2)		

mOCR-Myo 93A, 64V, 43Y. 68A, L32I				
12BB	**L32X**	**F33X**	**H97X**	**I99X**
V				27:73 (7)
L				31:69 (34)
I				
F				63:37 (14)
H				
W				50:50 0
Y				41:59 (8)

3.8.10. NMR Spectra of New Compounds

3.10 Acknowledgements:

This work was supported by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02- 05CH11231, by the NSF (graduate research fellowship to HMK), and the NWO Netherlands Organization for Scientific Research (the Rubicon postdoctoral fellowship, No. 680-50-1306 to PD). We thank the QB3 MacroLab facility (sub-cloning), the UC Berkeley DNA Sequencing Facility (plasmid sequencing), Dr. Tony Iavarone and the QB3 Mass Spectrometry Facility (NS-ESI data collection, supported by NIH grant 1S10RR022393-01), for nativeNS-ESI-MS data and analysis, and Prof. Huimin Zhao (University of Illinois-Champaign Urbana) for the P411-CIS gene.

3.11 Author Contributions

HMK, PD, and JFH conceived the work, designed the initial experiments, and interpreted the data. All authors discussed the results to design subsequent experiments. HMK and PD performed all of the experiments. HMK, PD, and JFH wrote the manuscript and all authors reviewed the manuscript.

3.12 References

1. Lewis, J. C. Artificial Metalloenzymes and Metallopeptide Catalysts for Organic Synthesis. *ACS Catal.* **3,** 2954–2975 (2013).
2. Farwell, C. C., Zhang, R. K., McIntosh, J. A., Hyster, T. K. & Arnold, F. H. Enantioselective Enzyme-Catalyzed Aziridination Enabled by Active-Site Evolution of a Cytochrome P450. *ACS Cent. Sci.* **1,** 89–93 (2015).
3. Hyster, T. K. & Arnold, F. H. P450BM3-Axial Mutations: A Gateway to Non-Natural Reactivity. *Isr. J. Chem.* **55,** 14–20 (2015).
4. Ringenberg, M. R. & Ward, T. R. Merging the best of two worlds: artificial metalloenzymes for enantioselective catalysis. *Chem. Commun.* **47,** 8470–8476 (2011).
5. Ward, T. R. Artificial Metalloenzymes Based on the Biotin−Avidin Technology: Enantioselective Catalysis and Beyond. *Acc. Chem. Res.* **44,** 47–57 (2010).
6. Jing, Q. & Kazlauskas, R. J. Regioselective Hydroformylation of Styrene Using Rhodium-Substituted Carbonic Anhydrase. *ChemCatChem* **2,** 953–957 (2010).
7. Abe, S. *et al.* Polymerization of Phenylacetylene by Rhodium Complexes within a Discrete Space of apo-Ferritin. *J. Am. Chem. Soc.* **131,** 6958–6960 (2009).
8. Key, H. M., Clark, D. S. & Hartwig, J. F. Generation, Characterization, and Tunable Reactivity of Organometallic Fragments Bound to a Protein Ligand. *J. Am. Chem. Soc.* **137,** 8261–8268 (2015).
9. Coelho, P. S., Brustad, E. M., Kannan, A. & Arnold, F. H. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **339,** 307–310 (2013).
10. Bordeaux, M., Tyagi, V. & Fasan, R. Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts. *Angew. Chem. Int. Ed.* **54,** 1744–1748 (2015).
11. de Montellano, P. O. Hydrocarbon hydroxylation by cytochrome P450 enzymes. *Chem. Rev.* **110,** 932–948 (2010).

12. Peters, M. W., Meinhold, P., Glieder, A. & Arnold, F. H. Regio- and Enantioselective Alkane Hydroxylation with Engineered Cytochromes P450 BM-3. *J. Am. Chem. Soc.* **125,** 13442–13450 (2003).
13. Roiban, G.-D. & Reetz, M. T. Expanding the toolbox of organic chemists: directed evolution of P450 monooxygenases as catalysts in regio- and stereoselective oxidative hydroxylation. *Chem. Commun.* **51,** 2208–2224 (2015).
14. Tyagi, V., Bonn, R. B. & Fasan, R. Intermolecular carbene S-H insertion catalysed by engineered myoglobin-based catalysts†. *Chem Sci* **6,** 2488–2494 (2015).
15. Chan, K. H., Guan, X., Lo, V. K. Y. & Che, C.-M. Elevated Catalytic Activity of Ruthenium(II)–Porphyrin-Catalyzed Carbene/Nitrene Transfer and Insertion Reactions with N-Heterocyclic Carbene Ligands. *Angew. Chem. Int. Ed.* **53,** 2982–2987 (2014).
16. Maxwell, J. L., O'Malley, S., Brown, K. C. & Kodadek, T. Shape-selective and asymmetric cyclopropanation of alkenes catalyzed by rhodium porphyrins. *Organometallics* **11,** 645–652 (1992).
17. Anding, B. J., Ellern, A. & Woo, L. K. Olefin Cyclopropanation Catalyzed by Iridium(III) Porphyrin Complexes. *Organometallics* **31,** 3628–3635 (2012).
18. Carey, J. R. *et al.* A site-selective dual anchoring strategy for artificial metalloprotein design. *J. Am. Chem. Soc.* **126,** 10812–10813 (2004).
19. Ohashi, M. *et al.* Preparation of artificial metalloenzymes by insertion of chromium(III) Schiff base complexes into apomyoglobin mutants. *Angew. Chem. Int. Ed.* **42,** 1005–1008 (2003).
20. Oohora, K., Kihira, Y., Mizohata, E., Inoue, T. & Hayashi, T. C(sp3)–H Bond Hydroxylation Catalyzed by Myoglobin Reconstituted with Manganese Porphycene. *J. Am. Chem. Soc.* **135,** 17282–17285 (2013).
21. Bordeaux, M., Singh, R. & Fasan, R. Intramolecular C(sp(3))H amination of arylsulfonyl azides with engineered and artificial myoglobin-based catalysts. *Bioorg. Med. Chem.* **22,** 5697–5704 (2014).
22. Teale, F. W. Cleavage of the haem-protein link by acid methylethylketone. *Biochim Biophys Acta* **35,** 543 (1959).
23. Lelyveld, V. S., Brustad, E., Arnold, F. H. & Jasanoff, A. Metal-Substituted Protein MRI Contrast Agents Engineered for Enhanced Relaxivity and Ligand Sensitivity. *J. Am. Chem. Soc.* **133,** 649–651 (2010).
24. Kawakami, N., Shoji, O. & Watanabe, Y. Single-Step Reconstitution of Apo-Hemoproteins at the Disruption Stage of Escherichia coli Cells. *ChemBioChem* **13,** 2045–2047 (2012).
25. Woodward, J. J., Martin, N. I. & Marletta, M. A. An Escherichia coli expression-based method for heme substitution. *Nat. Methods* **4,** 43–45 (2007).
26. Paulson, D. R., Addison, A. W., Dolphin, D. & James, B. R. Preparation of ruthenium(II) and ruthenium(III) myoglobin and the reaction of dioxygen, and carbon monoxide, with ruthenium(II) myoglobin. *J. Biol. Chem.* **254,** 7002–7006 (1979).
27. DelProposto, J., Majmudar, C. Y., Smith, J. L. & Brown, W. C. Mocr: a novel fusion tag for enhancing solubility that is compatible with structural biology applications. *Protein Expr. Purif.* **63,** 40–49 (2009).
28. Ilie, A. & Reetz, M. T. Directed Evolution of Artificial Metalloenzymes. *Isr. J. Chem.* **55,** 51–60 (2015).
29. Watson, H. C. The stereochemistry of the protein myoglobin. *Protein Stereochem.* **4,** 299

(1969).
30. Suematsu, H., Kanchiku, S., Uchida, T. & Katsuki, T. Construction of aryliridium-salen complexes: enantio- and cis-selective cyclopropanation of conjugated and nonconjugated olefins. *J. Am. Chem. Soc.* **130,** 10327–10337 (2008).
31. Fulmer, G. R. *et al.* NMR Chemical Shifts of Trace Impurities: Common Laboratory Solvents, Organics, and Gases in Deuterated Solvents Relevant to the Organometallic Chemist. *Organometallics* **29,** 2176–2179 (2010).
32. Huw M L Davies, Mônica V A Grazini, A. & Aouad, E. Asymmetric Intramolecular C−H Insertions of Aryldiazoacetates. *Org. Lett.* **3,** 1475–1477 (2001).
33. Nicolle, S. M. & Moody, C. J. Potassium N-Iodo p-Toluenesulfonamide (TsNIK, Iodamine-T): A New Reagent for the Oxidation of Hydrazones to Diazo Compounds. *Chem. Eur. J.* **20,** 4420–4425 (2014).
34. Bongen, P., Pietruszka, J. & Simon, R. C. Dynamic kinetic resolution of 2,3-dihydrobenzo[b]furans: chemoenzymatic synthesis of analgesic agent BRL 37959. *Chem. Eur. J.* **18,** 11063–11070 (2012).
35. Kamata, K., Kimura, T. & Mizuno, N. Cyclopropanation of Olefins with Diazo Compounds Catalyzed by a Dicopper-substituted Silicotungstate [γ-H2SiW10O36Cu2(μ-1,1-N3)2]4−. *Chemistry Letters* **39,** 702–703 (2010).
36. Bonaccorsi, C. & Mezzetti, A. Optimization or Breakthrough? The First Highly cis- and Enantioselective Asymmetric Cyclopropanation of 1-Octene by 'Electronic and Counterion' Tuning of [RuCl(PNNP)]+ Catalysts. *Organometallics* **24,** 4953–4960 (2005).

Chapter 4: An Artificial Iridium-P450 with the Kinetics of Native Enzymes

This chapter is modified from the following manuscript submitted for publication with permission of the university and all coauthors

Pawel Dydio contributed equally to the work described in this chapter.

ABSTRACT: Natural enzymes contain highly evolved active sites that bind substrates with affinities and in conformations leading to fast rates and high selectivities for chemical reactions. While artificial metalloenzymes have been developed to catalyze abiological transformations with high stereoselectivity, the activities and productivities of these artificial systems are inferior to those of natural enzymes. Here, we report an artificial metalloenzyme that reacts with kinetic parameters similar to those displayed by natural enzymes created by reconstituting a mutant of a natural P450 enzyme with an abiological metalloporphyrin. In particular, CYP119 variants containing an Ir(Me)-porphyrin catalyze the insertions of carbenes into C-H bonds in high yields with up to 98% ee, 35,000 TON, and 2,550 h^{-1} TOF, and this activity leads to intramolecular insertions into unactivated C-H bonds and *inter*molecular insertions into C-H bonds. The catalyst efficiency derived from Michaelis-Menton kinetic parameters is more than 1000 times greater than that of any other selective artificial metalloenzyme.

4.1 Introduction

Native enzymes undergo a range of synthetically valuable reactions with high activity and high stereo-, regio- and site-selectivities enabled by the structure of their active sites. [1,2] Metalloenzymes form a distinct class of natural enzymes, which contain a metal in the active site; this metal is often contained in a cofactor embedded within this site. The catalytic activity of a metalloenzyme is determined by both the primary coordination sphere of the metal and the surrounding protein scaffold. In some cases, laboratory evolution has been used to develop variants of metalloenzymes for selective reactions of unnatural substrates. [3,4] Yet, with few exceptions,[5] the classes of reactions that such enzymes undergo are limited to those of biological transformations.

To combine the favorable qualities of enzymes with the diverse reactivity of synthetic transition metal catalysts, abiological transition-metal centers have been incorporated into native proteins. The resulting systems, called artificial metalloenzymes, catalyze classes of reactions for which there is no known enzyme (i.e. abiological transformations). [5,6] Despite this progress, current artificial metalloenzymes lack many of the fundamental characteristics of natural enzymes, such as high activity (turnover frequency, TOF) and high productivity (turnover number, TON). They also lack many of the practical characteristics of enzymes used in synthesis,

such as the ability to be evolved in the laboratory, suitability for preparative-scale reactions, and potential to be recovered and reused. [2]

In many cases, artificial metalloenzymes have been prepared by anchoring abiological metal-organic catalysts within natural proteins. [7] In some cases, these constructs functioned as catalysts for abiological transformations, and in some cases their selectivities have been increased through mutagenesis of the protein scaffold. Some of the most successful developments of artificial metalloenzymes include enantioselective Rh-catalyzed annulations [8] and cyclopropanations [9] catalyzed within streptavidin and prolyl oligopeptidase. Although these artificial enzymes react with high stereoselectivity, they do not react with rates that rival natural enzymes. [10] These reactions occur with low turnover frequencies (TOF < 120 h^{-1}) and turnover numbers (TON<100).

Natural enzymes generally bind their substrates with high affinity and in a conformation that leads to extremely fast rates and high selectivity. [11] One reason that artificial metalloenzymes react more slowly than native enzymes is the absence of a defined binding site for the substrate. If the artificial metalloenzyme is generated by incorporation of full metal-ligand complexes into the substrate binding site of a natural enzyme or protein, the space remaining to bind a reactant for a catalytic process is limited and the interactions by which the protein binds the reactant are compromised. [7]

The previous chapter described that formal replacement of a biological metal for an abiological metal in a metalloprotein creates new reactivity, while leaving the natural substrate-binding site of the metalloprotein intact. [12] By introducing an iridium center in place of the iron in myoglobin (myo), we created and evolved the first metalloprotein that catalyzes the insertion of carbenes into C-H bonds to form C-C bonds. Although these artificial metalloenzymes catalyze reactions that are not catalyzed by any known natural enzyme, [6] the reactivity for these processes remained far from that of a natural enzyme for its native substrate [10] and the productivities and enantioselectivities of these enzymes were too low for synthetic applications. [2] Reaction were only accomplished on 20 mg scale and with up to 84% ee and 50 TON h^{-1}.

In addition, the Ir-myoglobins were suitable catalysts for only a limited set of substrates and transformations. For example, while the C-H bonds of methyl ethers were transformed in high yield (up to 98%), the somewhat more hindered C-H bonds of ethyl ethers underwent C-H insertion with lower yields (up to 50%), and the most hindered C-H bonds of benzyl ethers did not react in the presence of Ir-Myo catalysts. Unactivated C-H bonds were also inert, and attempts to achieve intermolecular reactions between ethyl diazoacetate and substrates containing highly activated C-H bonds resulted in formation of products only from competative diazo coupling and EDA insertion into water, rather than C-H insertion. One reason for this limited reactivity is that in all cases, the presence of the myoglobin resulted in reactions with slower rates than the rates of the same reactions catalyzed by the Ir(Me)-PIX cofactor alone. Therefore, in contrast to natural enzymes that enable difficult transformation via rate accelerate, in this case the myoglobin likely blocked access to the catalytic metal.

To address these limitations, we considered that alternative heme-binding proteins could form artificial M-PIX enzymes whose properties more strongly resemble those of natural enzymes. Though myoglobin offered many suitable properties for our initial studies (such as high stability, high yield of expression, and reported suitability for mutagenesis of the active site), myoglobin evolved to bind O_2, rather than to function as a selective catalyst, and its

structure reflects this function. The heme bind closely to the surface of myoglobin, and the substrate binding site is small (to accommodate O_2); thus the potential for strong and stabilizing interactions between large organic molecules and the protein may be limited.[13] Other heme proteins, such as cytochrome P450s, have evolved naturally to bind and react with large organic molecules.[14] In the case of P450s, the heme is buried deeply within the protein structure, such that that the adjacent substrate binding site can surround completely a large substrate. Therefore, P450s should be especially suitable scaffolds for the formation of artificial M-PIX proteins, as replacing only the metal of these enzymes should leave intact the native substrate binding site of the natural enzyme.

In numerous cases, Fe-P450s have been evolved to oxidize with abiological substrates, including substrates that have no structural similarity to the native substrates. In most cases, a single P450 (P450-BM3) has been used for the basis of these studies, because P450-BM3 conveniently contains both a "heme domain" and a "redox domain" fused into a single protein.[15] Other P450s can also be used, however in those cases it can be challenging to identify either the natural redox partner or a suitable replacement to work efficiently in concert with the heme protein. In our studies to evaluate M-PIX proteins for reactions other than oxidations, it is not necessary to employ a redox partner in the reaction;[12] therefore any P450 could form a suitable scaffold if it can be prepared in suitable yield and metallated with an abiological cofactor. Therefore, the P450 offering advantageous properties (such as the size of the substrate binding site or the thermal stability of the protein scaffold) can be chosen.

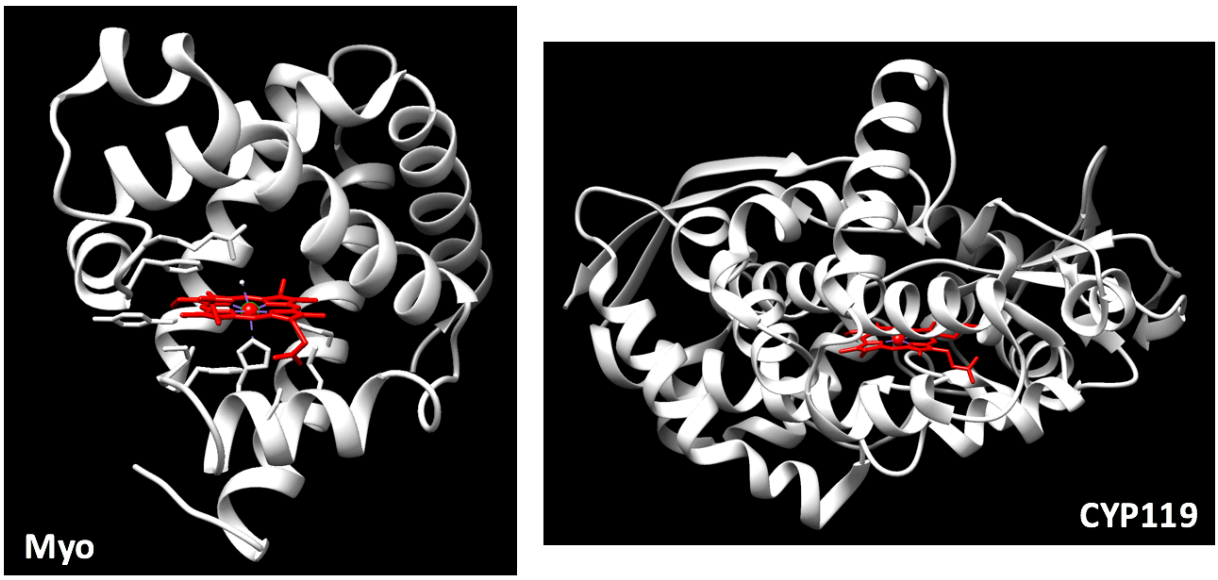


Figure 1. Comparison of the PIX binding locations in the cases of sperm whale myoglobin and CYP119 (a cyctochrome P450). The porphyrin cofactor is shown in red. Images produced in chimera using structural information from the PDB.

Here, we report that the artificial metalation of thermally stable cytochrome P450 enzymes (P450s) rather than myoglobin creates artificial metalloenzymes that can be evolved to catalyze abiological reactions with activities that are comparable to those of native enzymes [10]. In particular, we show that Ir(Me)-PIX enzymes derived from CYP119 catalyze insertions of carbenes into C-H bonds to form C-C bonds with up to 98% ee, 35,000 turnovers, and 2,550 h^{-1}

turnover frequency. The efficiency of these artificial metalloenzymes is more than 1000 times greater than that of any selective artificial metalloenzyme reported previously, [6] and Ir(Me)-PIX P450s are the only enantioselective artificial metalloenzymes that exhibit kinetic parameters that are similar to those displayed by natural enzymes. [10] This high reactivity enables the first enzymatic insertions into unactivated C-H bonds and the first intermolecular enzymatic insertions into any C-H bond.

4.2 Synthesis and Laboratory Evolution of CYP119 Enzymes Containing an Ir(Me)-PIX

Our studies to develop prospective artificial metalloenzymes with natural substrate binding sites began by using our method for artificial metalation of heme proteins to insert a methyliridium fragment into the iron site of a range of cytochrome P450 enzymes. P450s constitute a superfamily of heme-binding monooxygenases that are involved in various biosynthetic pathways, catalyzing reactions that encompass chemo-, regio-, stereo- and site-selective C-H hydroxylation reactions of complex natural products. [14] We hypothesized that artificial metalloenzymes that are created from P450s in thermophiles, such as CYP119 [16] in *Sulfolobus solfataricus*, could improve the thermal stability of the resulting artificial metalloenzyme and create the potential to conduct reactions at elevated temperatures. Thus, we conducted studies with CYP119, as well as the more commonly used P450-BM3 and P450-CAM.

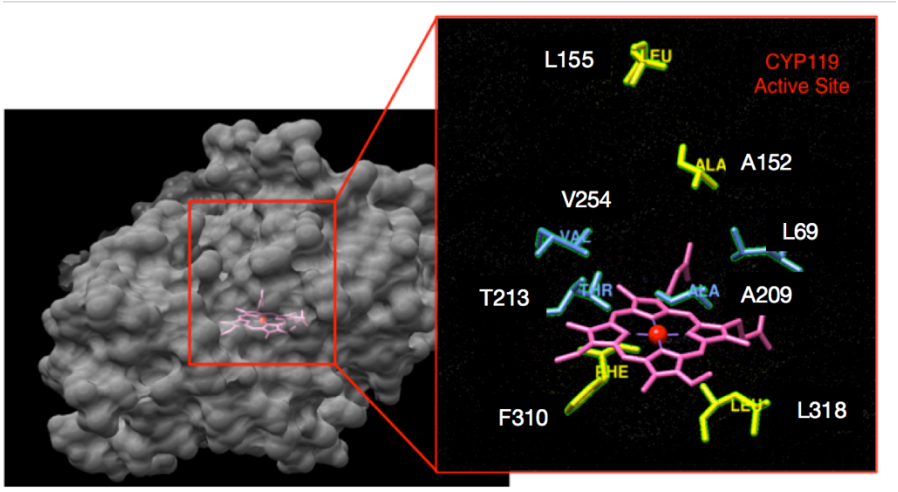


Figure 2. Structure of WT Fe-CYP119 (image prepared in Chimera from PDB 1IO7). Left: Complete structure of Fe-CYP119. Right: Active-site residues targeted in directed evolution of the protein scaffold to increase activity and selectivity in C-H insertion reactions.

Simple, general protocols to prepare prospective artificial enzymes are essential for both efficient directed evolution strategies and for synthetic applications. The apo forms of P450-BM3, CAM, and CYP119 were expressed in good yields (15-25 mg/L) under conditions similar to those used to express apo-myoglobin. [12] The apo forms of P450-BM3, CAM, and CYP119 were readily and quantitatively reconstituted with Ir(Me)-PIX to yield the corresponding Ir-proteins, with no need for further purification (Fig. S1). As hypothesized, the Ir(Me)-PIX protein

formed from CYP119 had a much higher Tm (69 °C) than those formed from P450-BM3 (45 °C) or P450-CAM (40 °C) (Fig. S2). This higher Tm suggested that enzymes created from the scaffold of CYP119 could be used at elevated temperatures. Thus, this protein was used for our studies on catalytic reactions.

With thermally stable Ir(Me)-PIX CYP119 enzymes in hand, we initially evaluated them as catalysts for similar transformations to those catalyzed by Ir(Me)-Myo in order to compare the potential of these two protein hosts to function as active and selective enzymes for the insertion of carbenes into C-H bonds. By studying the model reaction to convert diazoester **1** into dihydrobenzofuran **2**, we found that the activity and selectivity of Ir(Me)-PIX CYP119 enzymes are readily evolved through molecular evolution of the natural substrate binding site of the CYP119 scaffold (Fig. 2 and 3). The wild type (WT) Ir(Me)-CYP119 enzyme and its variant C317G (bearing a mutation that introduces space to accommodate the axial ligand of the Ir(Me)-PIX cofactor) catalyze the intramolecular C-H insertion reaction to form **2**, although with low rates (TOF = 8 and 7 h^{-1}), yields (21% and 18%), and enantioselectivities (ee = 0% and 14%) for reactions conducted on a 2.5 μmol scale with 0.17 mol% catalyst. In particular, these rates were slower than those achieved for a comparable mutant of Ir(Me)-PIX Myo (H93A, TOF = 25 h^{-1}), possible because the Ir-site is buried more deeply within the protein (Fig. 1), and the substrate cannot easily access the metal site.

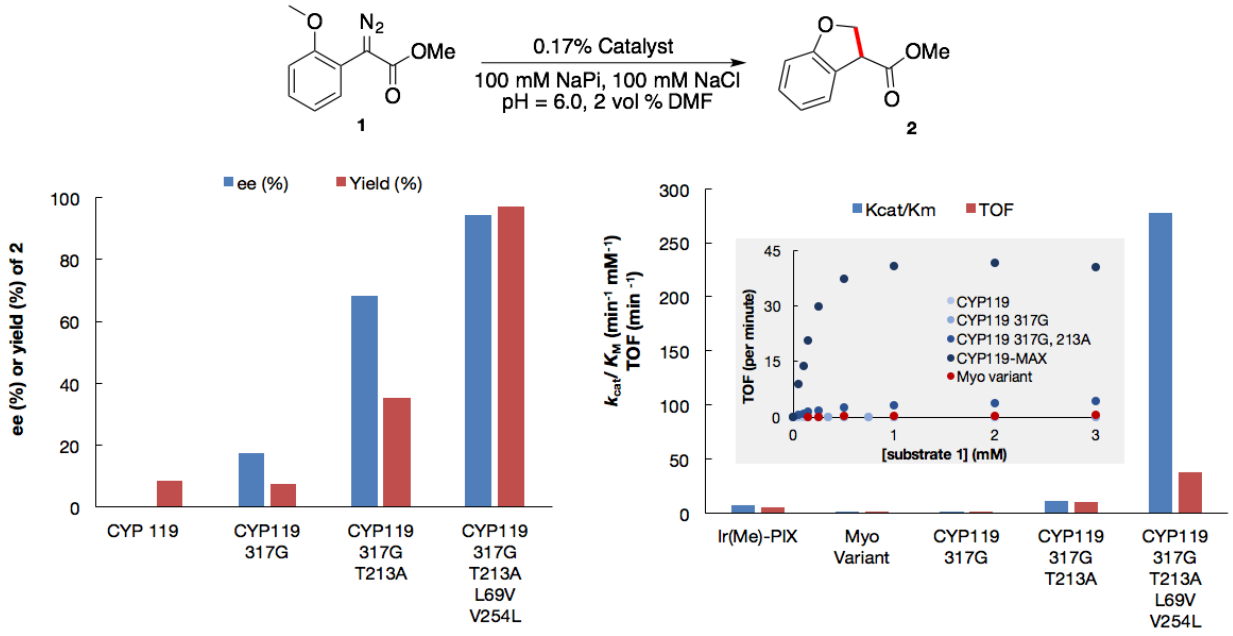


Figure 3. Directed evolution of Ir(Me)-PIX CYP119 for enantioselective insertions of carbenes into C-H bonds. Top: Model reaction converting diazoester **1** to dihydrobenzofuran **2**. Top Left: enantioselectivity and yields for the formation of **2** catalyzed by sequentially evolved variants of CYP119. Top Right: Kinetic parameters describing the formation of **2** by variants of CYP119.

To identify mutants that form **2** with higher rates and enantioselectivity, we incorporated mutations at a series of additional residues close to the site where the substrates should bind adjacent to the Ir-PIX cofactor (L69, A209, T213, and V254, Fig. 2). We found that by reducing the steric bulk at the residue T213 from that of threonine to that of glycine, the activity of the enzyme could be increased substantially. The double mutant C317G, T213G formed **2** with 68% ee and 80-fold higher activity (TOF = 556 h^{-1}) than the single mutant C317G (Fig. 3). These data support the hypothesis that the WT and single C317G mutant react with poor activity due to low accessibility of the substrate to the metal site. Therefore, we created further mutants containing the mutations C317G and T213G plus additional mutations to the sites L69, A209, and V254.

Two additional rounds of evolution identified the quadruple mutant C317G, T213G, L69V, V254L (CYP119-Max) that formed **2** with 94% enantioselectivity and with an initial TOF of 2550 h^{-1} (Fig. 3). This rate is more than 300-times faster than that of the WT variant. Such rates are unprecedented for abiological transformations catalyzed by artificial metalloenzymes with high enantioselectivity. In total, this highly active and selective mutant was identified from the evaluation of only 150 variants of CYP119, compared to 450 variants of myoglobin that were evaluated to be able to achieve the same reaction with only 68% ee and 50 TON h^{-1}. Together these results support the hypothesis that Ir(Me)-PIX enzymes created from P450s hold greater potential to be evolved towards high activity and selectivity than those derived from myoglobin.

To further validate this hypothesis, we applied directed evolution to identify suitable mutants that form products **3-5** with high enantioselectivity. We found that these reactions could be achieved with up to 98% ee, which is the highest level of enantioselectivity ever observed in a reaction catalyzed by an artificial metalloenzymes (Fig. 4). Moreover, for products **2- 4**, each enantiomer of the product could be formed with higher ee than was achieved using catalysts formed from myoglobin, as discussed in the previous chapter. The reactions to form products **3-5** show that the enzyme reacts as effectively with substrates containing substituents on the aryl ring as it does with the unsubstituted **2**. Such reactivity is relevant to contemporary synthetic challenges, as compound **(*S*)-5** is an intermediate in the synthesis of BRL 37959 – a potent analgesic – and was prepared previously by kinetic resolution. [17] A single enantiomer of this product could be formed directly by the strategy of asymmetric carbene insertion into a C-H bond catalyzed by the Ir(Me)-PIX CYP119 variant 69Y-152W-213G in 94% ee (Fig. 4).

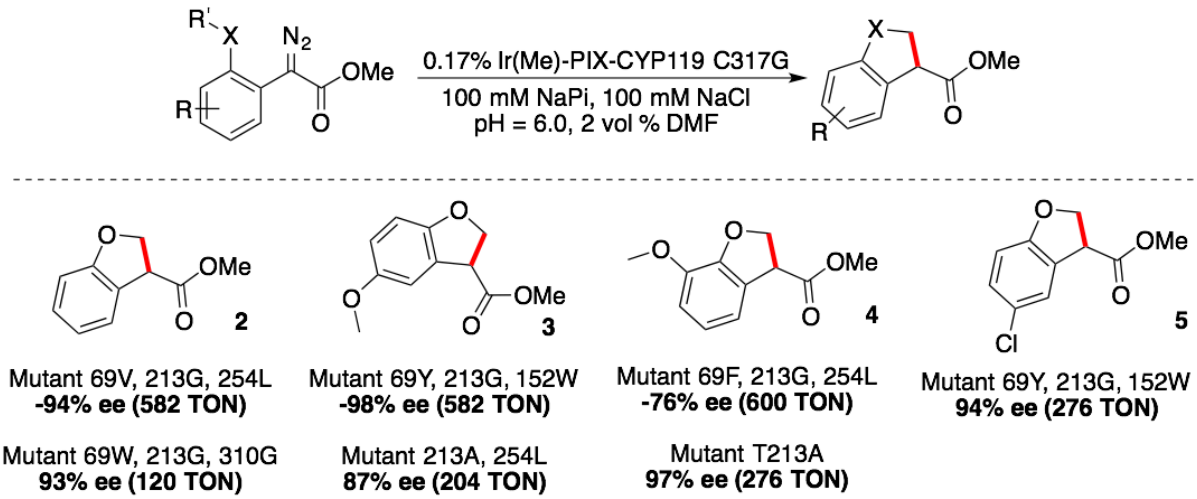


Figure 4. Most selective variants of Ir(Me)-PIX CYP119 identified to catalyze enantioselective intra- and intermolecular C-H insertion reactions of activated and unactivated C-H bonds.

4.3 Evaluation of the Kinetics of Artificial Ir(Me)-CYP119 Enzymes in Comparison to Natural Enzymes Involved in Biosynthesis

To gain insight into the origin of the differences in the enzymatic activity between the various mutants of Ir(Me)-CYP119, we evaluated the kinetic parameters of particular mutants from each phase of the directed evolution that led to the variant CYP119-Max. These studies found that the affinity of substrate **1** for the WT enzyme is weak, as revealed by a high K_m (> 5 mM). The single mutant C317G, which lacks a sidechain at this position that could act as an axial ligand, exhibits higher substrate affinity (K_m = 3.14 mM), catalytic activity (k_{cat} = 0.22 min^{-1}), and, therefore, overall efficiency (k_{cat}/K_m = 0.071 $min^{-1}mM^{-1}$) than the WT enzyme (Fig. 3). In fact, the efficiency of this single mutant of CYP119 is comparable to that of the previously identified most enantioselective variant of Ir(Me)-PIX-mOCR-Myo, which contains 5 mutations and formed **2** in 68% ee (Fig. 3). The double mutant T213G, C317G of Ir(Me)-CYP119 reacts with far more favorable Michaelis-Menten parameters (k_{cat} = 4.88 min^{-1} and K_m = 0.43 mM, k_{cat}/K_m = 11 $min^{-1}mM^{-1}$) than those of this single mutant C317G. The incorporation of two additional mutations (L69V, V254L) led to further improvements of both k_{cat} and K_m, creating an enzyme (CYP119-Max) with 4,000-times higher efficiency (k_{cat} = 48.0 min^{-1}, K_m = 0.17 mM, and k_{cat}/K_m = 278 $min^{-1}mM^{-1}$) than that of the WT system (Fig. 3). Furthermore, the rates of reactions catalyzed by this mutant at concentrations below saturation are almost an order of magnitude faster than those catalyzed by the free iridium-porphyrin (TOF = 37 min^{-1} versus TOF = 4.5 min^{-1} at 2 mmol of substrate **1**). These results show the value of conducting this iridium-catalyzed reaction within the enzyme active site to control selectivity and increase the reaction rate simultaneously.

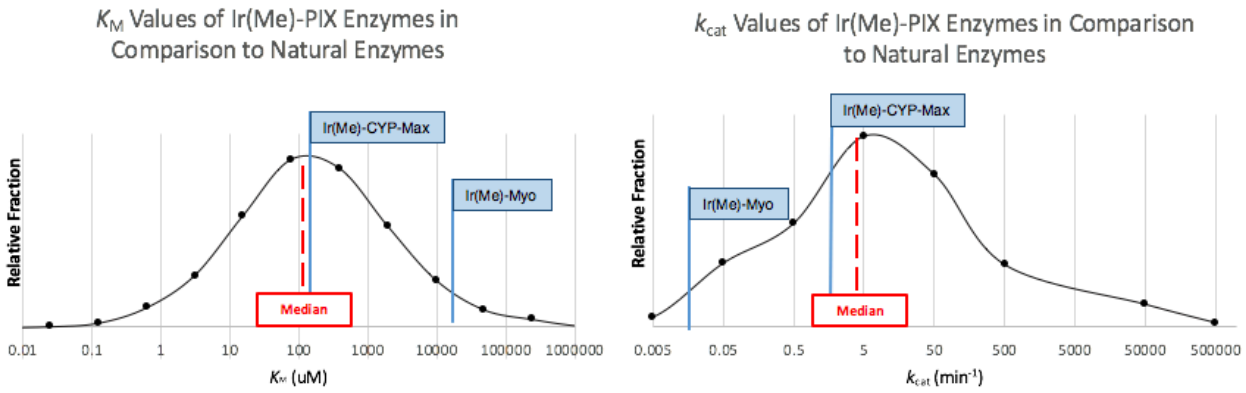


Figure 5: Comparison of K_M and k_{cat} values for CYP119-Max with those of natural enzymes involved in the metabolism of intermediate and secondary metabolites [10]. For comparison, the kinetic parameters for Ir(Me)-PIX mOCR-myoglobin (H93A, H64V) catalyzing the same transformation are shown.

A comparison of the kinetic parameters of reactions catalyzed by the Ir(Me)-PIX CYP119-Max enzyme to those of natural enzymes involved in intermediate and secondary metabolism is shown in Figure 5. This comparison indicates that Ir(Me)-PIX CYP119-Max reacts with kinetic parameters that are comparable to those of natural enzymes. [10] The binding affinity of Ir(Me)-CYP119-Max for the abiological substrate **1** is even higher than the affinity of P450s for their native substrates (compare K_m = 0.17 mM for CYP119-Max to K_m = 0.298 mM for the native substrate lauric acid of P450-BM3) [18] and similar to the median K_m value for natural enzymes (0.13 mM). [10] In addition, the k_{cat} of 48.0 min^{-1} for this enzyme is within an order of magnitude of the median k_{cat} of natural enzymes responsible for the production of biosynthetic intermediates (312 min^{-1}) and secondary metabolites (150 min^{-1}). [10] This analysis demonstrates that direct replacement of the metal found in a natural metalloenzyme by an abiological metal creates artificial metalloenzymes that can be evolved into catalysts that react with tight binding and favorable preorganization of unnatural substrates, and thus with high activity and selectivity.

4.4 Application of Ir(Me)-CYP119 Enzymes to Functionalize Sterically Hindered or Unactivated C-H Bonds

The potential to evolve proteins having advantageous enzyme-substrate interactions should also create the possibility to catalyze C-H insertion reactions involving less reactive substrates (Fig. 6-7). Product **6** results from carbene insertion into a sterically hindered C-H bond of a benzyl ether (Fig. 6). Directed evolution furnished a mutant capable of forming **6** with high enantio- and diastereoselectivity favoring the cis isomer (94% ee, 12 : 1 dr (cis : trans)). The Ir(Me)-PIX enzymes based on myoglobin reported previously produced this product in only trace amounts, and the free Ir(Me)-PIX cofactor formed predominantly the trans isomer (3:1 dr, trans : cis). This reversal of diastereoselectivity from that of the free Ir(Me)-PIX cofactor to that of the

artificial metalloenzyme highlights the ability of strong substrate-enzyme interactions to override the inherent selectivity of a metal cofactor or substrate.

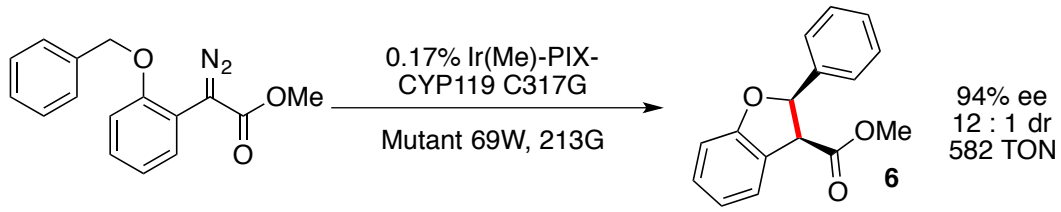
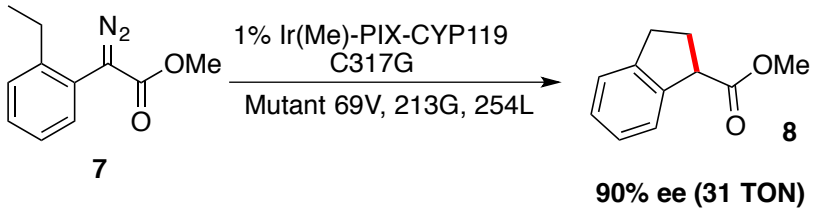


Figure 6. Enantio- and diastereoselective intramolecular C-H insertion reaction of a sterically hindered C-H bond.

Ir(Me)-CYP119-Max also catalyzes the insertion of carbenes into fully unactivated C-H bonds (Fig. 7). Although substrates **1** and **7** are structurally similar, the primary C-H bonds in **7** are stronger and less reactive than those in **1**, which are located alpha to an oxygen atom. [19] In fact, there are no metal catalysts of any type reported to form indanes by carbene insertion into an unactivated C-H bond with synthetically useful enantioselectivities. [20] In contrast, Ir(Me)-CYP119-Max catalyzed the formation of **8** in 90% ee. To observe substantial amounts of this product (TON = 31), the reaction was conducted at 40° C; the enantioselectivity of the product at this temperature was the same as that of the small quantity of product formed at room temperature. These data constitute the first enzyme-catalyzed carbene insertion into an unactivated C-H bond and illustrate the value of employing a thermally stable enzyme scaffold.



First enzymatic insertion of a carbene into an unactivated C-H bond

Highest enantioselectivity attained in the synthesis of an indane by metal catalyzed insertion of a carbene into an unactivated C-H bond

Figure 7. Intramolecular C-H insertion reaction of an unactivated double bond.

4.5 Application of Ir(Me)-CYP119 Enzymes to Intermolecular C-H Functionalization Reactions

In addition to catalyzing intramolecular reactions with unactivated C-H bonds, Ir(Me)-CYP119-Max catalyzes the first enzyme-catalyzed, intermolecular carbene insertion into a C-H bond (Fig. 8). Intermolecular insertions of carbenes into C-H bonds are challenging because the metal-carbene intermediate can undergo competitive diazo coupling or insert the carbene unit into the O-H bond of water. [9] In fact, the model reaction between phthalan and ethyl diazoacetate (EDA) forms alkene and alcohol as the dominant products when catalyzed by the free Ir(Me)-

PIX cofactor; only trace amounts of insertion product **11** were formed. In sharp contrast, the same reaction catalyzed by the mutant Ir(Me)-PIX CYP119-Max-A152F occurred to form **11** in 55% yield with 330 TON and 68% ee. Dimerization of the carbene when catalyzed by Ir(Me)-CYP119-Max is limited, presumably because of selective binding and preorganization of the substrate; the selectivity for formation of the C-H insertion product **11** over side product **12** was 70-fold higher when catalyzed by Ir(Me)-PIX CYP119-Max-A152F than when catalyzed by the free cofactor. This intermolecular, enzyme-catalyzed C-H insertion reaction opens new possibilities for selective C-H functionalizations.

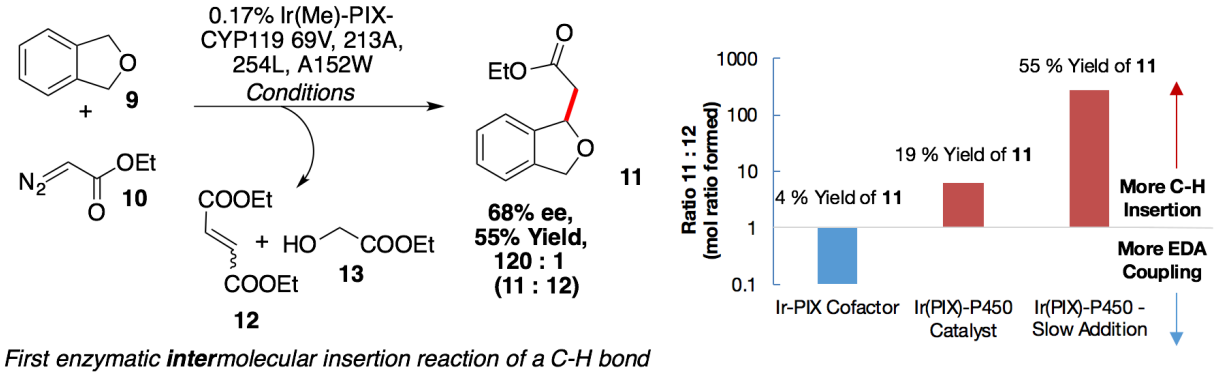
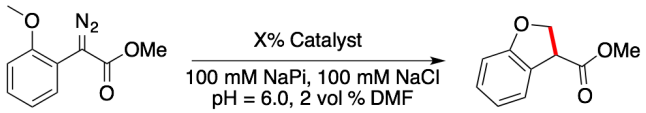


Figure 8. Intermolecular C-H insertion reaction of a carbene. Left: Reaction of 1 equiv. pthalan **9** and 3 equiv. EDA **10**, with addition of EDA over 1 hour. Right: Comparison of the chemoselectivity of the metal-carbene intermediate in the case of reactions catalyzed by Ir(Me)-PIX cofactor and those catalyzed by Ir(Me)-PIX

4.6 Initial Studies To Evaluate the Synthetic Potential of Ir(Me)-PIX CYP119 Enzymes

With the ultimate goal of applying artificial metalloenzymes to the synthesis of organic molecules for fine chemicals, the reactions conducted by such catalytic systems should be suitable for preparative scales, occur with high substrate concentrations, react with high TONs and be attached to a solid support for recycling. We found that a series of reactions containing between 40 mg and 1 g of substrate **1** catalyzed by Ir(Me)-CYP-Max occurred with yields and enantioselectivities that were similar to each other, showing that the outcome of the reaction is independent of the scale (Figure 9). Moreover, with 200 mM of substrate, reactions catalyzed by Ir(Me)-CYP-Max (0.0025 mM) formed product **2** with up to 35,000 TON (Figure 9). Thus, this artificial metalloenzyme operates with high productivity under conditions suitable for preparative scales. Finally, Ir(Me)-CYP-Max supported on CNBr-activated sepharose catalyzed the conversion of **1** to **2** by carbene insertion into a C-H bond in 52% yield and 83% ee. This supported catalyst was used, recovered, and recycled four times without loss of the enantioselectivity for formation of **2** (Figure S11).



Reactions Conducted at a Range of Scales

mmol/ mg Sub	% catalyst	TON	Yield
0.0025 mmol, 0.5 mg	0.17%	330	55% yield (GC)
0.2 mmol, 40 mg	0.25%	384	48% yield (Isolated)
0.2 mmol, 40 mg	0.05%	2120	53% yield (Isolated)
1.0 mmol, 206 mg	0.017%	3529	60% yield (Isolated)
5.0 mmol, 1 gram	0.017%	3235	55% yield (Isolated)

Reactions Conducted with a Range of Catalyst Loadings

[Sub], mg Sub	[catalyst]	TON	Yield
100 mM, 5 mg	0.025 mM	4791	95% yield
100 mM, 5 mg	0.01 mM	8326	83% yield
100 mM, 10 mg	0.005 mM	30293	76% yield
200 mM, 10 mg	0.0025 mM	35129	43% yield

Figure 9. Catalyst productivity in intramolecular C-H insertion reactions catalyzed by Ir(Me)-PIX CYP119-Max under synthetically relevant reaction conditions.

4.7 Conclusions and Future Directions

Enzymes containing artificial, transition-metal active sites that react with the kinetics, selectivity, and evolutionary potential of natural enzymes has been a major goal of catalyst design. Here we show that artificial metalloenzymes catalyzing abiological processes can possess the two fundamental characteristics of natural enzymes – fast kinetics and high selectivity – while being amenable to laboratory evolution to enhance their reactivity, selectivity, and stability. By exploiting these favorable properties, intramolecular functionalization of C-H bonds in substrates bearing varied steric properties, intramolecular functionalization of fully unactivated C-H bonds, and intermolecular reactions of C-H bonds can occur. Moreover, artificially metallated P450s can be used in a practical fashion on a preparative scale. Together, our results lift the restrictions on the kinetics of artificial metalloenzymes that has limited this direction for merging chemical and biocatalysis.

4.8 Experimental Details

4.8.1 Protein Expression, Purification, and Characterization

a. General Methods

Unless otherwise noted, the chemicals, salts, and solvents used were reagent grade and used as received from commercial suppliers without further purification. All expression media and buffers were prepared using ddH_2O (MilliQ A10 Advantage purification system, Millipore). All expression media were sterilized using either an autoclave (45 min, 121^{o}C) or a sterile syringe filter (0.22 um). To maintain sterile conditions, sterile materials and *E. coli* cells were manipulated near a lit Bunsen burner.

b. Genes and Cloning

The WT CYP119 gene cloned into the vector 2BT (6xHis-TEV-ORF; AddGene #29666) was purchased from GenScript with codon optimization for *E. coli* (Table S1). The genes for WT P450 BM3 and P450-CAM were obtained as a gift from Prof. Humin Zhao (University of Illinois) and cloned to the vector 2BT at the QB3 Macrolab at UC Berkeley.

c. Media Preparation

Preparation of optimized minimal expression media: Salts (15 g Na_2HPO_4, 7.5 g K_2HPO_4, 0.3 g NaH_2PO_4, 0.3 g KH_2PO_4, 1.5 g NaCl, 5.0 g NH_4Cl) were dissolved in 2 L ddH_2O and autoclaved to give a media with pH ~8.0 - 8.2. Solutions of glucose (20%), casamino acids (BD Company, low Fe, 20%), and $MgSO_4$ (1 M), were autoclaved separately. Solutions of ampicillin (100 mg/ mL) and $CaCl_2$ (1 M) were sterilized by syringe filter. The following amounts of the listed solutions were added per 2 L of sterile salt solution: 40 mL glucose, 20 mL casamino acids, 4 mL $MgSO_4$, 100 uL $CaCl_2$, 2 mL ampicillin. Stock solutions were stored for several weeks; prepared media was stored for less than 1 day. Minimal media plates were prepared from the same media with the addition of 17 g agar/L media. In this case, agar was autoclaved in 1 L ddH_2O, and salts were autoclaved separately as a 20X solution, after which they were added to the agar solution.

d. Mutagenesis

Site-directed mutagenesis was performed using the QuickChange Lightening mutagenesis kit (Agilent); requisite double stranded DNA primers were designed according to the Agilent Primer Design Program and purchased from Integrated DNA Technology. PCR reactions were performed according to the manufacturer's directions. PCR reactions contained 5 uL reaction buffer, 34 uL ddH_2O, 1.5 uL QuickSolution, 1 uL plamids (50 ng/uL), 1.25 uL sense primer (100 ng/uL), 1.25 uL antisense primer (100 ng/ uL), 5 uL dNTPs (2 mM/base), and 1 uL polymerase.

PCR Program: Phase 1 (1 cycle): 95 °C, 1.5 min; Phase 2 (18 cycles): 95 °C, 20 sec, 60 °C, 10 sec, 68 °C, 4.5 min; Phase 3 (1 cycle): 68 °C, 3 min; Phase 4 (storage): 4 °C

DNA Isolation and Storage: Following the completion of the above set of PCR procedures, 1.5 uL DPN 1 was added to each reaction, and the reactions were further incubated (3 h, 37 °C). The crude PCR mixture was used to transform XL-10 Gold Ultracompetent cells (45 L cells, 2 L PCR reactions). The mixture was incubated on ice (30 min), heat shocked (30 s, 42 °C), recovered with SOC media (1 h, 37 °C, 275 rpm), and plated on LB plates. Plates were grown (18 h, 37 °C), and individual colonies were used to inoculate 1 mL rich media cultures, which were grown in 96-well plates (13 h, 37 °C, 300 rpm). DNA was isolated from the 96-well cultures using magnetic bead technology at the UC Berkeley DNA Sequencing Facility. Alternatively, individual colonies were used to inoculate 4 mL rich media cultures and grown overnight (13 h, 37 °C, 300 rpm), and the plasmids were purified using a Qiagen DNA Miniprep kit according to the manufacturer's instructions.

e. Protein Expression

Optimized Expression of Apo CYP119: BL21 Star competent *E.* coli cells (50 L, QB3 Macrolab, UC Berkeley) were thawed on ice, transferred to 14 mL Falcon tubes, and transformed with the desired plasmid solution (2 L, 50-250 ng/ L). The cells were incubated on ice (30 min), heat shocked (20 sec, 42°C), re-cooled on ice (2 min), and recovered with SOC media (37 °C, 1 h, 250 rpm). Aliquots of the cultures (50uL) were plated on minimal media plates (expression media supplemented with 17 g agar/L) and incubated (20 h, 37 °C) to produce approximately 10-100 colonies per plate. Single colonies were used to inoculate starter cultures (3 mL, expression media), which were grown (4-8 hours, 37°C, 275 rpm) and used to inoculate 100 mL overnight cultures (minimal media, 37° C, 275 rpm). Each culture grown overnight was used to inoculate 750 mL of minimal media, which was then grown further (9 h, 37 °C, 275 rpm). Expression was induced with IPTG (800 uL, 1M), and the cultures were grown further (15 h, 30 °C, 275 rpm). Cells were harvested by centrifugation (5000 rpm, 15 min, 4° C), and the pellets were resuspended in 20 mL Ni-NTA lysis buffer (50 mM NaPi, 250 mM NaCl, 10 mM Imidazole, pH = 8.0) and stored at -80 °C until purification.

Protein Purification: Cell suspensions were thawed in a room-temperature water bath, decanted to 50 mL glass beakers, and lysed on ice by sonication (3x30 sec on, 2x2 min off, 65% power). Cell debris was removed by centrifugation (10 000 rpm, 30 min, 4 °C), and Ni-NTA (5 mL, 50% suspension per 850 mL cell culture) was added. The lysates were briefly incubated with Ni-NTA (30 min, rt, 20 rpm) and poured into glass frits (coarse, 50 mL). The resin was washed with Ni-NTA lysis buffer (3 x 35 mL), and the wash fractions were monitored using a Bradford assay dye. The desired protein was eluted with 18 mL Ni-NTA elution buffer (50 mM NaPi, 250 mM NaCl, 250 mM Imidazole, pH = 8.0), dialyzed twice against Tris buffer (10 mM, pH = 8.0, 1 h, rt), concentrated to the desired concentration using a spin concentrator, and metallated within several hours. Apo protein was not stored for more than 8 hours.

Metallation of the Apo-Protein (General Method): Stock solutions of metal cofactors in DMF were added to solutions of apo protein (0.12 mM) at room temperature in the desired stoichiometry with a final DMF concentration of 2%. The proteins were briefly incubated at room temperature (5 minutes), and DMF was removed by using a NAP column equilibrated with the reaction buffer (100 mM NaPi, 100 mM NaCl, pH = 6.0).

Protein Storage: Glycerol was added to a solution of the protein (3:1 v:v protein solution: 50% glycerol), the solution was divided into 1.5 mL Eppendorf tubes (0.5 mL per aliquot), and the tubes were flash frozen in liquid nitrogen and stored at -80 °C until further use. Frozen aliquots of the protein were thawed in a room-temperature water bath.

f. Protein Characterization

Gel Electrophoresis: Protein purity was analyzed by sodium dodecyl sulfate-polyacrylamide (SDS-PAGE) gel electrophoresis using precast gels (polyacrylamide, 10-20% linear gradient, Biorad).

Mass Spectrometry: Apo-proteins were analyzed with an Agilent 1200 series liquid chromatograph connected in-line with an Agilent 6224 time-of-flight (TOF) LC/MS system using a Turbospray ion source. Metallated proteins were analyzed by native nanoelectrospray

ionization mass spectrometry (nanoESI-MS) using a Waters Q-ToF Premier quadrupole time-of-flight mass spectrometer equipped with a nanoESI source (Milford, MA). Mass spectra were acquired in the positive ion mode and processed using MassLynx software (version 4.1, Waters). The instrument is located in the QB3/Chemistry Mass Spectrometry Facility at UC Berkeley.

Melting Temperature: The melting temperature of Ir(Me)PIX containing heme proteins were determined by differential scanning fluorimetry (DSF) using SYPRO Orange as the fluorescent reporter, according to an established protocol. DSF curves were obtained using a CFX96 Touch Real Time PCR Detection System (BioRad), and data were processed using the Solver extension of Microsoft Excel. Conditions and melting temperatures can be found in Table S7. DSF curves can be found in Figure S1.

4.8.2 Organic Synthesis and Characterization

a. General methods and materials

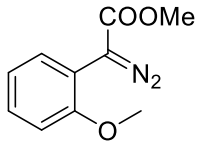
Unless stated otherwise, all reactions and manipulations were conducted on the laboratory bench in air with reagent grade solvents. Reactions under inert gas atmosphere were carried out in the oven dried glassware in a nitrogen-filled glovebox or by standard Schlenk techniques under nitrogen.

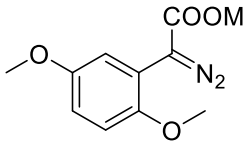
NMR spectra were acquired on 400 MHz, 500 MHz, 600 MHz, or 900 MHz Bruker instruments at the University of California, Berkeley. NMR spectra were processed with MestReNova 9.0 (Mestrelab Research SL). Chemical shifts are reported in ppm and referenced to residual solvent peaks[21]. Coupling constants are reported in hertz. Chiral SFC analysis was conducted on a JASCO SF-2000 integrated analytical SFC system. GC analyses were obtained on an Agilent 6890 GC equipped with either, an HP-5 column (25 m x 0.20 mm ID x 0.33 m film) for achiral analysis or Cyclosil-B column (30m x 0.25mm x 0.25 um film) for chiral analysis, and an FID detector. GC yields were calculated using dodecane as the internal. High-resolution mass spectra and elemental analysis were obtained via the Micro-Mass/Analytical Facility operated by the College of Chemistry, University of California, Berkeley.

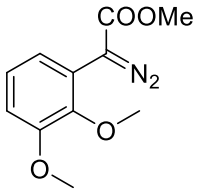
Unless noted otherwise, all reagents and solvents were purchased from commercial suppliers and used without further purification. If required, dichloromethane (DCM) and tetrahydrofuran (THF) were degassed by purging with argon for 15 minutes and dried with a solvent purification system containing a one-meter column of activated alumina; dried and degassed acetonitrile, 1,2-xylene, toluene, N,N-dimethylformamide (DMF), ethanol and methanol were purchased form commercial suppliers and used as received.

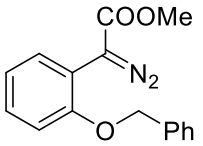
b. Substrates

The synthetic procedures and characterization of ethyl (2-methoxyphenyl)diazoacetate (**1**), methyl (2,5-dimethoxyphenyl)diazoacetate (**S3**), methyl (2,3-dimethoxyphenyl)diazoacetate (**S4**)[22] and methyl (2-benzyloxyphenyl)diazoacetate (**S6**)[23] were reported previously.

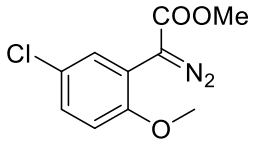








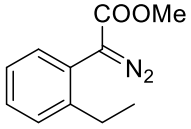
Methyl (5-chloro-2-methoxyphenyl)diazoacetate (S5):



In a closed vial, a solution of 5-chloro-2-methoxyphenylacetic acid (4.0 g, 20 mmol) in methanol (15 ml) containing several drops of sulfuric acid, was stirred overnight at 80 °C. The volatile materials were evaporated under vacuum. The residue was dissolved in ethyl acetate (~40 ml), washed with $NaHCO_3$ sat. (40 ml) and water (40 ml), dried over $MgSO_4$ and evaporated. The crude product was used in the next step without further purification.

To a stirred solution of methyl 5-chloro-2-methoxyphenylacetate (20 mmol) and 4-acetamidobenzenesulfonyl azide (p-ABSA, 7.2 g, 30 mmol) in acetonitrile (40 ml) at 0 °C, 1,8-diazabicycloundec-7-ene (DBU, 4.8 ml, 32 mmol) was added dropwise. The cooling bath was removed, and the reaction was allowed to continue stirring overnight. The reaction mixture was diluted with dichloromethane (~60 ml), washed with water (2 x ~50 ml), dried over $MgSO_4$ and evaporated. The crude product was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 90:10 gradient) as the eluent. Fractions of the pure product were combined, and the solvent evaporated, yielding 3.9 g (81%) of product.
^{1}H NMR (500 MHz, $CDCl_3$): d = 7.55 (d, J = 2.7 Hz, 1H), 7.15 (dd, J = 8.8, 2.6 Hz, 1H), 6.76 (d, J = 8.8 Hz, 1H), 3.80 (s, 3H), 3.80 (s, 3H);
^{13}C NMR (151 MHz, $CDCl_3$): d = 166.19, 153.85, 129.29, 127.97, 126.32, 115.57, 112.03, 112.02, 55.99, 52.20 ($C=N_2$ signal missing, as observed before for related molecules[24]);
HR MS (EI): calcd. for $C_{10}H_9N_2O_4Cl$ $[M]^+$: 240.0302, found: 240.0303.

Methyl (2-ethylphenyl)diazoacetate (7):



Methyl (2-ethylphenyl)acetate was prepared following a synthetic procedure developed for analogous compounds:[20] to a stirred solution of 2-bromobenzyl acetate (2 g, 8.7 mmol) in 30 mL of THF were added freshly prepared ethylzinc bromide (20 mL, ~0.6-0.7 M in THF)[25] and Pd(t-$Bu_3P)_2$ (90 mg, 0.175 mmol) sequentially at room temperature. The reaction mixture was stirred at room temperature for 8 h, then quenched with 1N HCl (30 mL). The product was extracted with ethyl acetate (3 x 50 mL), and the combined organic layers were washed with brine (30 mL), dried over $MgSO_4$ and evaporated. The crude product was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 90:10

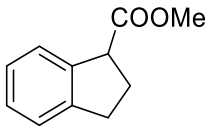
gradient) as the eluent. Fractions of the pure product were combined, and the solvent evaporated, yielding 1.29 g (83%) of product as colorless liquid.
^{1}H NMR (500 MHz, $CDCl_3$): 7.21 – 7.08 (m, 4H), 3.65 (s, 3H), 3.63 (s, 2H), 2.62 (q, J = 7.6 Hz, 2H), 1.17 (t, J = 7.6 Hz, 3H);
^{13}C NMR (151 MHz, $CDCl_3$): d = 172.36, 142.76, 132.15, 130.63, 128.74, 127.78, 126.22, 52.20, 38.62, 25.98, 15.02;
HR MS (EI): calcd. for $C_{11}H_{14}O_2$ $[M]^+$: 178.0994, found: 178.0997.

To a stirred solution of methyl (2-ethylphenyl)acetate (1.29 g, 7.1 mmol) and 4-acetamidobenzenesulfonyl azide (p-ABSA, 2.4 g, 10 mmol) in acetonitrile (40 ml) at 0 °C, 1,8-diazabicycloundec-7-ene (DBU, 1.8 ml, 12 mmol) was added dropwise. The cooling bath was removed, and the reaction was allowed to continue stirring for 48 h (the reaction progress was followed by TLC). The reaction mixture was diluted with dichloromethane (~50 ml), washed with water (2 x ~50 ml), and dried over $MgSO_4$. After filtration, the volatile material from the filtrate was evaporated under reduced pressure. The crude product was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 95:5 gradient) as the eluent. Fractions of the pure product were combined, and the solvent evaporated, yielding 0.72 g (50%) of product.
^{1}H NMR (400 MHz, $CDCl_3$): d = 7.37 – 7.17 (m, 4H), 3.79 (s, 3H), 2.61 (q, J = 7.6 Hz, 2H), 1.21 (t, J = 7.5 Hz, 3H);
^{13}C NMR (151 MHz, $CDCl_3$): d = 166.89, 144.17, 131.71, 129.58, 128.96, 126.47, 123.74, 52.25, 26.33, 14.55 ($C=N_2$ signal missing, as observed before for related molecules[24]);
HR MS (EI): calcd. for $C_{11}H_{12}N_2O_2$ $[M]^+$: 204.0899, found: 204.0899.

c. Authentic Products

General procedure for synthesis of dihydrobenzofurans
To 5 ml of a solution of a derivative of methyl (2-methoxyphenyl)diazoacetate (~50 mM) in toluene 60-300 ul of a solution of Ir(Me)-PIX (8 mM, 0.2-1 mol%) in DMF was added, and the reaction mixture was stirred vigorously. The reaction progress was monitored by TLC. Upon completion, the volatile materials were evaporated under reduced pressure, and the residue was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 80:20 gradient) as eluent. Fractions of the pure product were combined, and the solvent evaporated, yielding 20-90% of desired product. The NMR data match those of the reported molecules: methyl 2,3-dihydrobenzofuran-3-carboxylate (**2**),[26] methyl 5-methoxy-2,3-dihydrobenzofuran-3-carboxylate (**3**), [22] 7-methoxy-2,3-dihydrobenzofuran-3-carboxylate (**4**), [22] 5-chloro-2,3-dihydrobenzofuran-3-carboxylate (**5**),[17] methyl trans-2-phenyl-2,3-dihydrobenzofuran-3-carboxylate (trans-**6**),[23] and methyl cis-2-phenyl-2,3-dihydrobenzofuran-3-carboxylate (cis-**6**).[23]

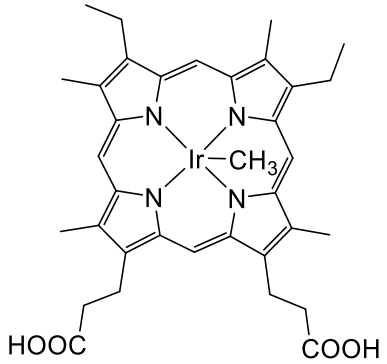
Methyl indane-1-carboxylate (11):



In a closed vial, a solution of methyl (2-ethylphenyl)diazoacetate (50 mg) and Ir(Me)-PIX (~1 mg) in toluene (10 ml) was stirred at 80 °C. The reaction progress was monitored by TLC. Upon

completion (~16 h), the volatile materials were evaporated under reduced pressure, and the residue was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 95:5 gradient) as eluent. Fractions of the pure product were combined, and the solvent evaporated, yielding title product. The NMR data match those of the reported molecule.[27]

d. Ir(Me)-Mesoporphyrin IX (Ir(Me)-PIX) cofactor used in the studies:



The synthetic procedures and characterization of Ir(Me)-PIX were reported previously. [22]

4.8.3 Catalytic Experiments

a. General Methods:

Unless otherwise noted, catalytic reactions were performed in 4 mL individually-capped vials or in 1.2 mL vials as part of a 96-well array fitted with a cover that is attached by screws. Reactions were either (1) assembled in a nitrogen atmosphere glove box or (2) assembled on the bench. In the latter case, the headspace of the vial was purged with nitrogen through a septum cap. Solutions of Ir(Me)-PIX-CYP119 were gently degassed on a Schlenk line (3 cycles vacuum/refill) before being pumped into a glove box in sealed vials. Organic reagents were added as stock solutions in DMF, such that the final amount of DMF in the reaction was approximately 2% by volume (unless noted otherwise). Protein catalysts were diluted to reaction concentrations in sodium phosphate buffer (100 mM, pH = 6.0) before being added to reaction vials. Unless otherwise noted, all reactions were performed with catalysts generated from a 1:2 ratio of Ir(Me)-cofactor : apo protein, with 0.17 mol % catalyst loading, based on the ratio limiting reagent to metal cofactor. Unless otherwise noted, all reactions were conducted in a shaking incubator (20 °C, 16 h, 275 rpm).

b. Procedure for typical catalytic experiments

The catalyst stock solution: The Ir(Me)-CYP119 catalyst was prepared by addition of a stock solution of Ir(Me)-PIX (3.1 mM in DMF) to a solution of the apo-protein (0.13mM in 10 mM Tris buffer, pH = 8), such that the resulting solution had a 1 : 2 ratio of Ir(Me)-PIX : CYP119. This ratio was used to ensure that all Ir-PIX was bound. The mixture was incubated for 5 min, and desalted with the NAP-10 desalting column equilibrated with reaction buffer (100 mM NaPi, 100 mM NaCl, pH = 6.0). The protein mixture was diluted to the required reaction concentration with the same reaction buffer.

Intramolecular C-H Insertion: 250 ul of the catalyst stock solution (0.0425 umol of [Ir]) was added to a vial, followed by addition of a stock solution of the appropriate diazo compound (2.5 umol in 5 uL DMF). For experiments with higher substrate concentration (> 50 mM), the substrate was added directly to a vial, followed by the appropriate organic solvent and 250 ul of the catalyst stock solution. The vial was sealed with a cap, removed from a glovebox and incubated in a shaker (20 °C, 275 rpm). For experiments at elevated temperatures, the vial was incubated in a metal heat block at the reported temperature. Upon completion, the reaction mixture was analyzed as described in section 4.8.6.

Intermolecular C-H Insertion: 1 mL of the catalyst stock solution (0.017 mM Ir(Me)-CYP119 in 100 mM NaPi, 100 mM NaCl, pH =6.0) was added to a 4 mL vial containing a micro stir bar. The vial was capped with a cap containing a septa, and the catalyst solution was degassed by three cycles of vacuum/nitrogen refill, after which it was tranfered to a nitrogen atmosphere wet box. In the box, 10 umol phthalan (1.1 uL) and 10 uL DMF were added to the vial, and a separate solution of EDA was prepared in DMF (50X relative to final targeted EDA concentration). The reaction vial and EDA solution were covered with septa caps and removed from the glovebox. A syringe pump outfitted with a series of gas tight syringes was used to add the solution of EDA (20 uL) to each reaction vial over 1 hour. After the conclusion of the reaction, the reaction mixture was quenched and analyzed as described in section 4.8.6.

Mutant Screening (directed evolution): Mutants were evaluated for C-H insertion reactions in 96-well plates with the reaction conditions described in the general method (Section Va,b). Twelve mutants were analyzed per each 96-well plate. Aliquots of each Ir(Me)-P450 mutant were distributed down the columns of a 96-well plate. Next, aliquots of substrates for intramolecular C-H insertion were added to each of the rows (A-E) to generate up to 60 unique reactions. The reaction with the most selective mutants were repeated. The same enantioselectivities and similar yields were measured in both sets of experiments.

4.8.4 Kinetic experiments

The initial rates of selected mutants for C-H insertion reactions were evaluated in 96-well plates with the catalyst stock solution prepared as described in the general method section. The stock solutions of the substrate in different concentrations (2 – 100 mM in DMF) were distributed in the first row of a 96-well plate. Aliquots of 250 ul of each solution of Ir(Me)-P450 mutant were distributed to remaining rows of a 96-well plate (rows B-H). Next, 12.5 ul aliquots of the substrate stock solutions from row A were added to each of the rows containing catalyst solutions (B-H) to initiate reactions. After suitable period of time (2-80 min, depending on the activity of the variant), the reactions were simultaneously quenched by the addition HBr (60 uL, 50% in water). All manipulations were performed using a multichannel pipet. The reaction mixtures were analyzed as described in section IV. (Similar enantioselectivities were measured as in the experiments conducted under standard conditions). The measured yields were used to calculate reaction rates. Repeated experiments, including experiments with different time points and different catalyst concentrations resulted in similar reaction parameters for the same variants of Ir(Me)-CYP119. The kinetic parameters of the enzymes were obtained by the data fitting with the standard Michaelis-Menten kinetic model, using the data fitting software package Origin 8.0.

4.8.5 Procedure for Reactions on Synthetic Scale

The catalyst stock solution – prepared as in described section Vb – was added to a Schlenk flask and gently degassed on a Schlenk line (3 cycles vacuum/refill). A solution of the appropriate diazo substrate in DMF was added (the concentration of the stock solution adjusted such that the final amount of DMF in the reaction was approximately 2% by volume). The flask was sealed and gently agitated by shaking (120 rpm) or end over end rotation (20 rpm) at 20 °C. The reaction was diluted with brine (30 ml) and extracted with ethyl acetate (3·50 ml). If required, the phase separation was achieved by centrifuging (2000 rpm, 3 min) the mixture. The combined organic fractions were washed with brine (30 ml) and dried over $MgSO_4$. After filtration, the volatile material from the filtrate was evaporated under reduced pressure. The residue was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 90:10 gradient) as the eluent. Fractions of the pure product were combined, and the solvent evaporated. The specific conditions (variants of the catalyst, loadings, and substrate concentrations), yields, enantioselectivites and specific rotations of products for reactions with different substrates are reported in Table S3.

4.8.6 Analysis of Yield and Enantiomeric Ratio (er)

Yields were determined by achiral or chiral GC using dodecane as an internal standard (Figure S2). Enantiomeric ratios were determined either by chiral GC or chiral SFC (Figures S2-S9). Achiral GC was used to determine the yields for reactions analyzed by SFC, while the yields for reactions analyzed by chiral GC were determined concurrently with that analysis. The methods used to determine the e.r. of each product are summarized in a subsequent section. Samples for analysis were prepared as follows, depending on the analysis method:

SFC/achiral GC: Saturated NaCl (200 uL) was added to each reaction vial, followed by a solution of dodecane (500 ul, 1 ul/ml) in EtOAc. The contents of the vial were mixed by pipet, and the phases were allowed to separate. A portion (250 uL) of the organic layer was removed from the top of the vial by pipet, transferred to a new vial, evaporated and redissoved in MeOH for analysis by SFC. An additional portion of EtOAc (250 uL) was added to the original reaction vial, and the reaction was quenched by the addition of HBr (60 uL, 50% in water). After separation of the layers, approximately 400 uL of the aqueous phase was removed from the bottom of the vial by pipet. The remaining contents of the vial were neutralized by the addition of sat. $NaHCO_3$ (200 uL), and the organic layer was further diluted with EtOAc (500 uL). The organic layer was then transferred to a separate vial for GC analysis. In the case of experiments performed in a 96-well array, all manipulations were performed using a multichannel pipet.

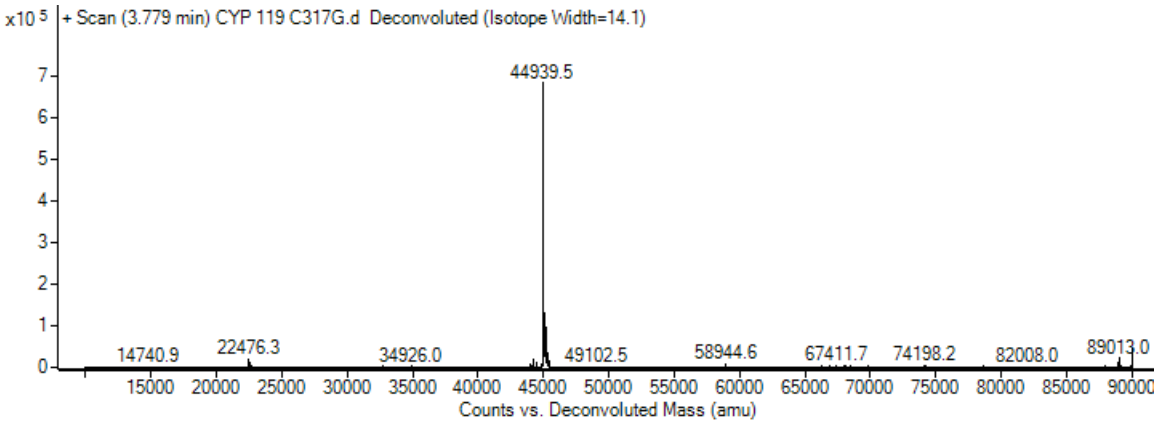
Chiral GC: Saturated NaCl (200 uL) was added to each reaction vial, followed by a solution of dodecane (500 ul, 1 ul/ml) in EtOAc. The contents of the vial were mixed by pipet, and the phases were allowed to separate. The reaction was then quenched by the addition HBr (40 uL, 50% in water). After separation of the layers, approximately 400 uL of the aqueous phase was removed from the bottom of the vial by pipet. The remaining contents of the vial were neutralized by the addition of sat. $NaHCO_3$ (200 uL), and the organic layer was further diluted with EtOAc

(500 uL). The organic layer was then transferred to a separate vial for GC analysis. In the case of experiments performed in a 96-well array, all manipulations were performed using a multichannel pipet.

4.9 Supporting Figures

Figure S1. Characterization of CYP119 protein by liquid chromatography/mass spectrometry (mass determination, above) and differential scanning fluorimetry (melting temperature, below

Deconvoluted mass spectrum of apo CYP119-C317G acquired by LC-MS.



Differential scanning fluorimetry curves used to determine the melting temperature of Ir(Me)-PIX containing heme proteins.

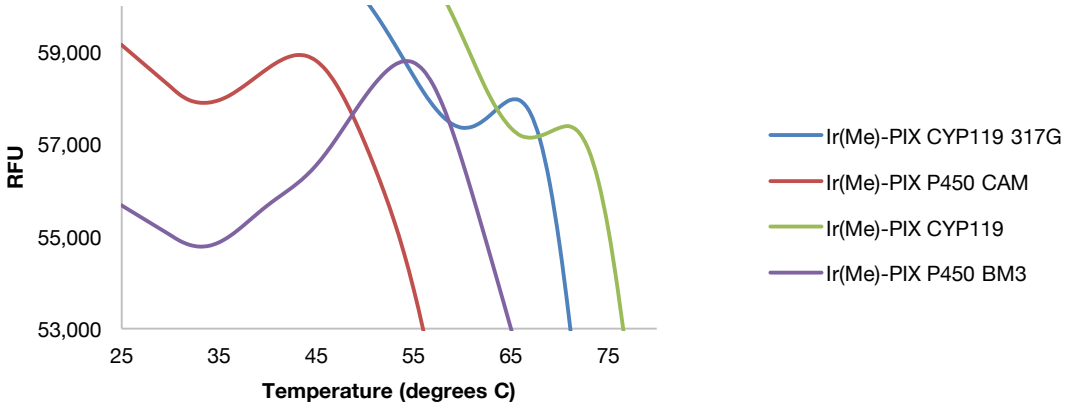


Figure S2. Chiral GC traces for product **2**; racemate (above) and enantioenriched sample (below).

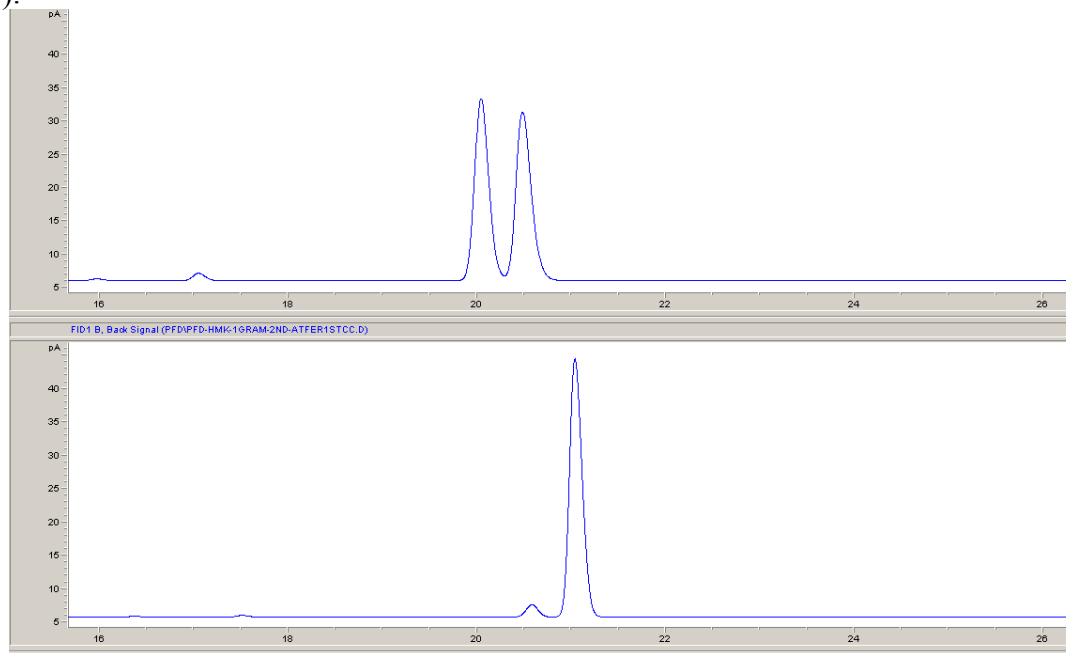


Figure S3. Chiral GC traces for product **3**; racemate (above) and enantioenriched sample (below).

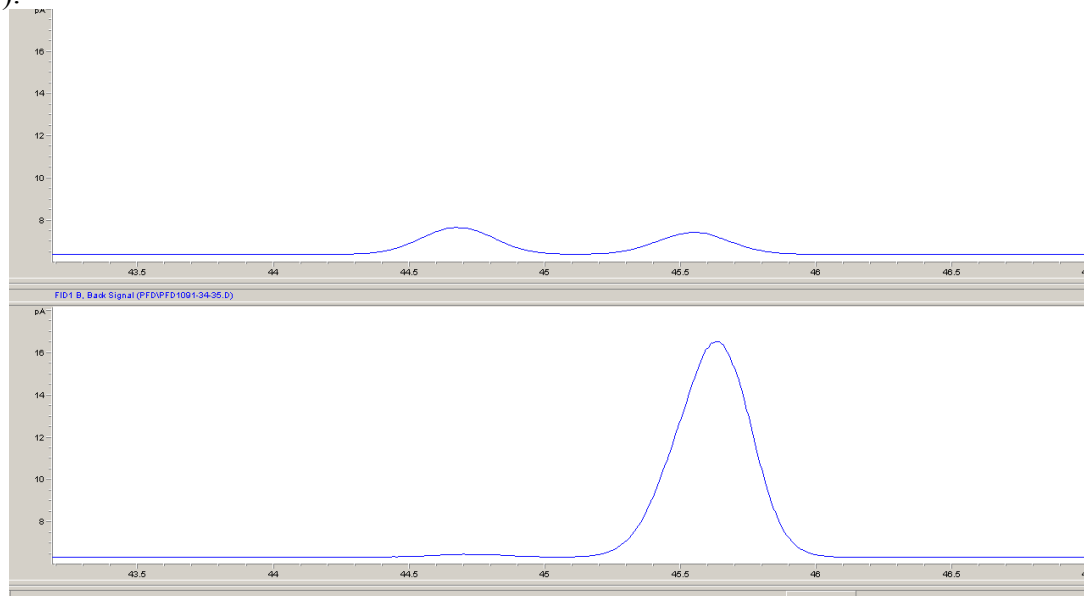


Figure S4. Chiral GC traces for product **4**; racemate (above) and enantioenriched sample (below).

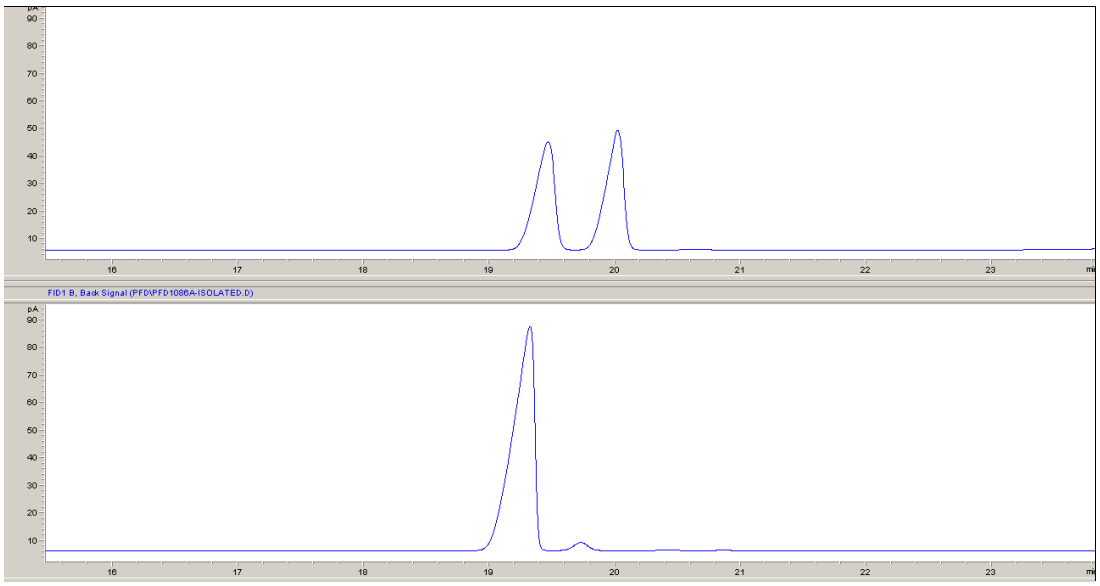


Figure S5. Chiral GC traces for product **5**; racemate (above) and enantioenriched sample (below).

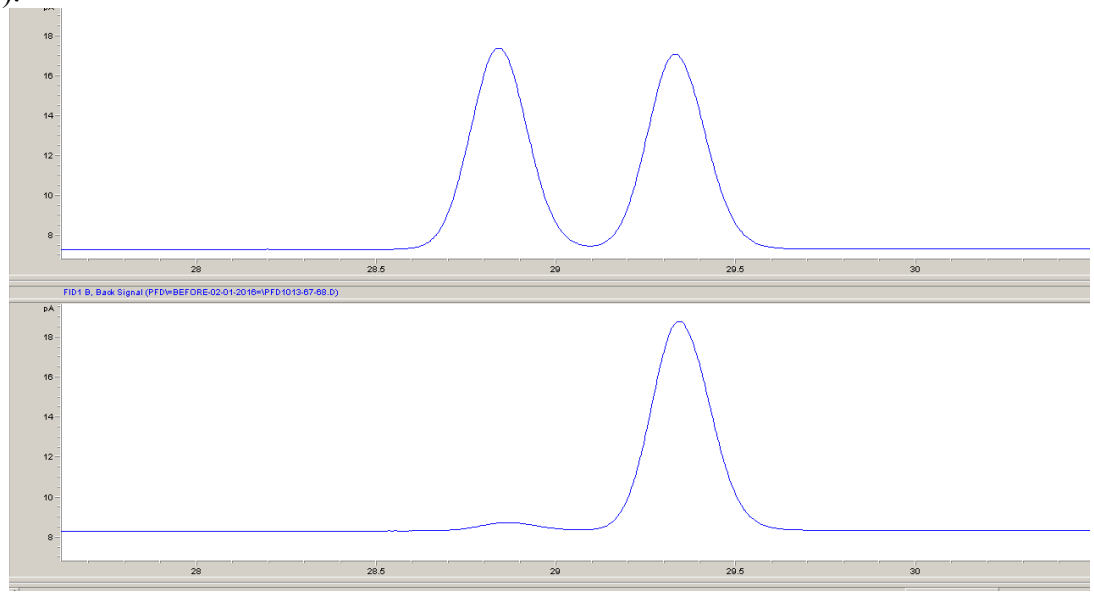


Figure S6. Chiral SFC traces for product **6**; racemate (above) and enantioenriched sample of cis-**6** (middle), and trans-**6** (below).

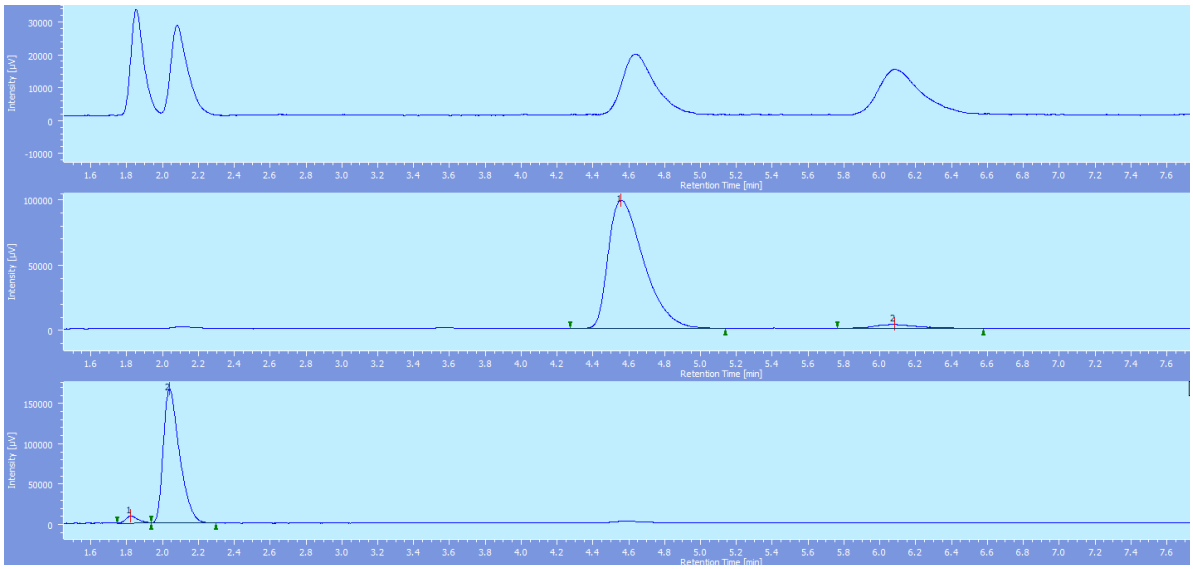


Figure S7. Chiral GC traces for product **8**; racemate (above) and enantioenriched sample (below).

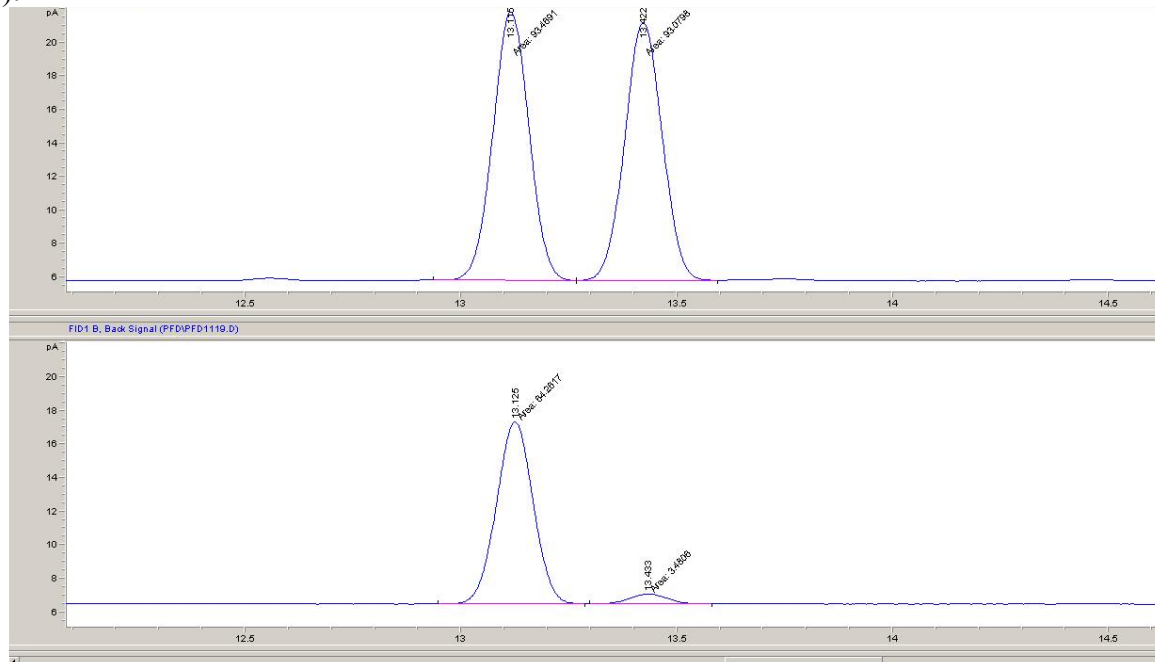


Figure S8. Chiral SFC traces for product **11**; racemate (above) and enantioenriched sample (below).

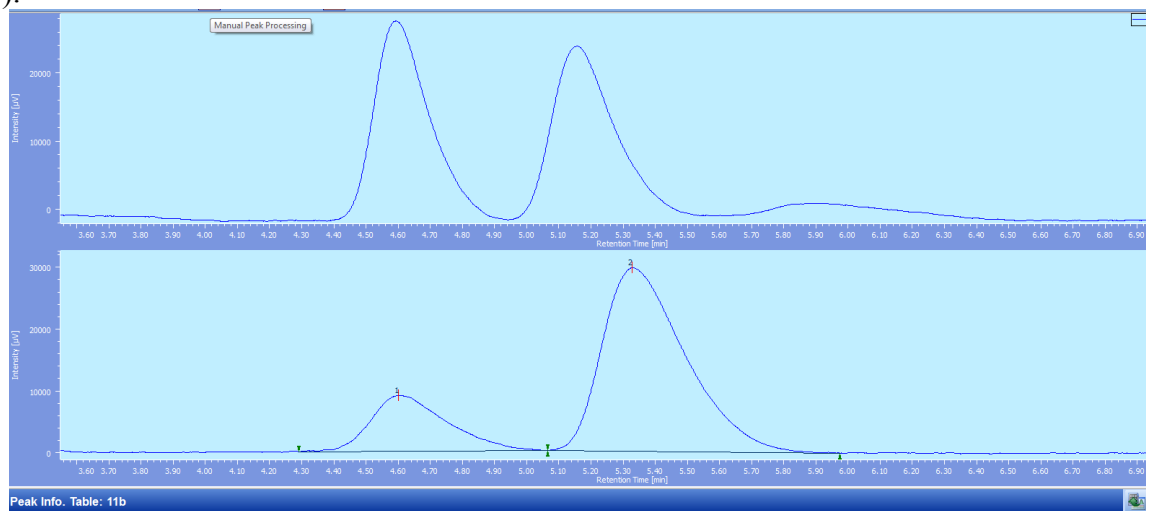
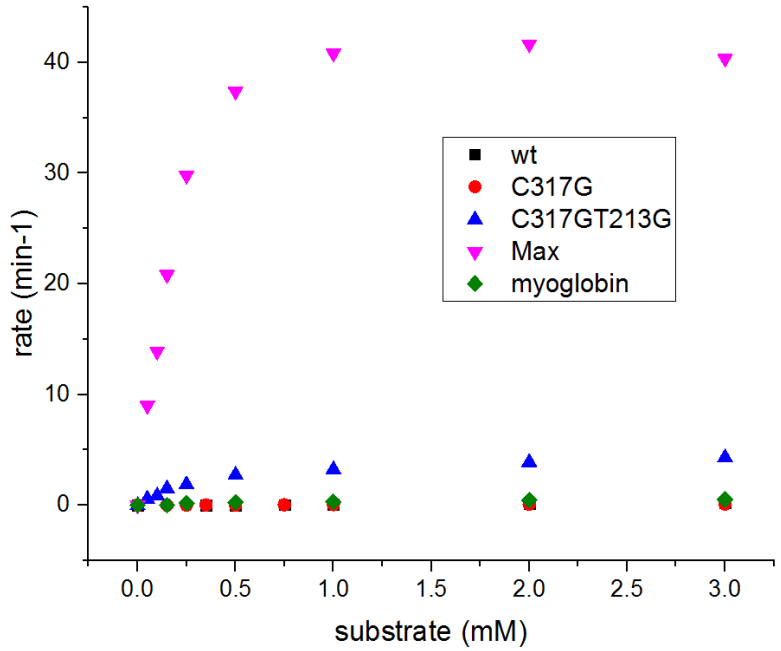
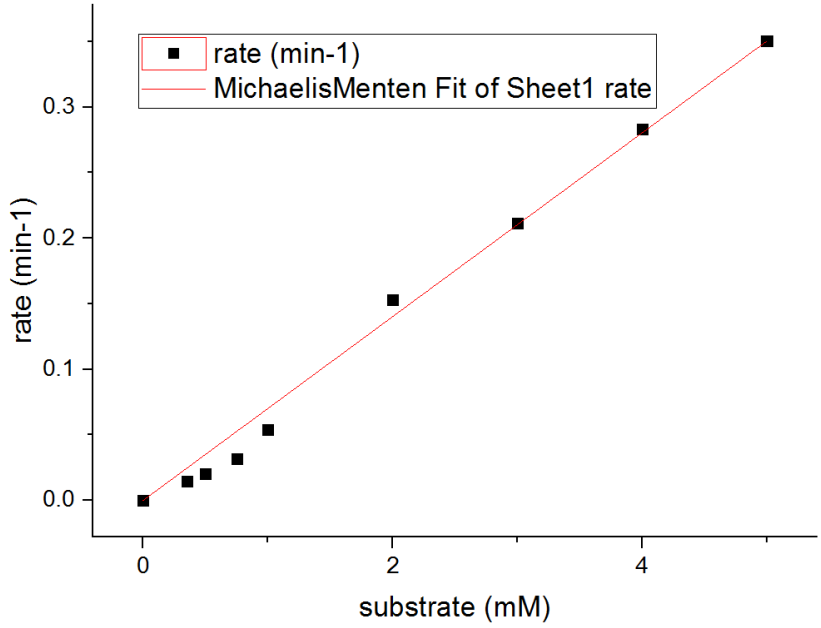


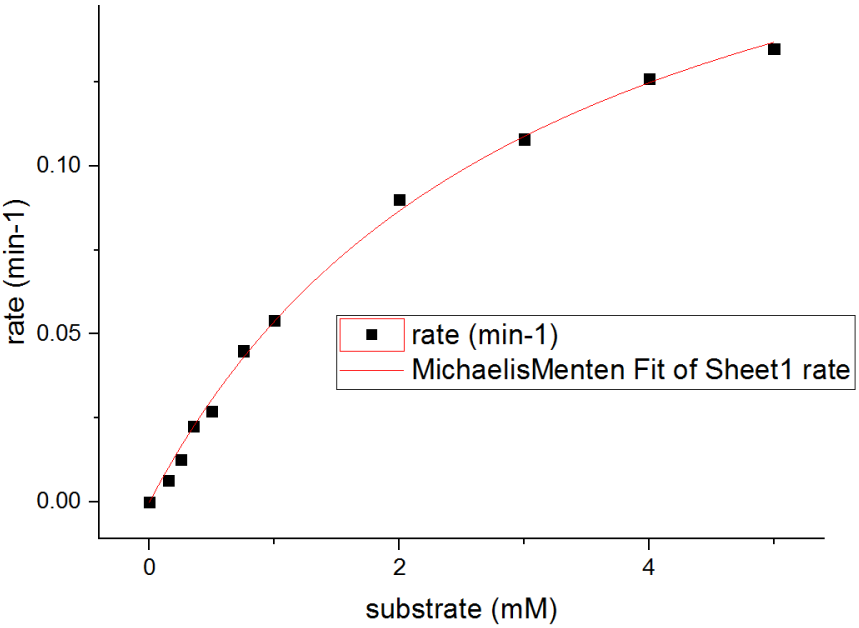
Figure S9. Plots of initial rates for reactions of substrate **1** to form **2** in the presence of different variants of Ir(Me)-CYP119 and selective variant of Ir(Me)-mOCR-myoglobin (H93A, H64V). Comparison of the activity of different enzymes (above) and data and the fit to the Michaelis-Menten model for each reaction (below).

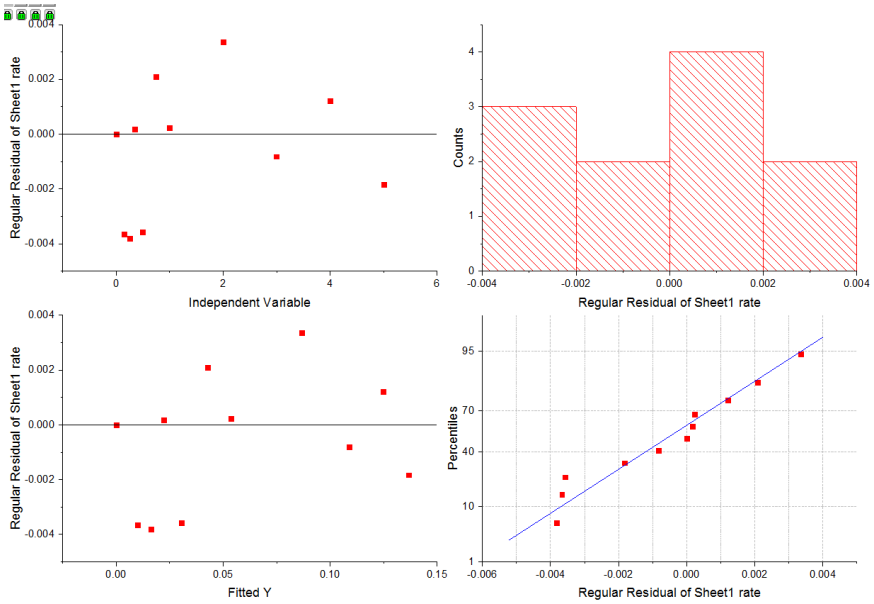


Ir(Me)-CYP119 WT – no satisfactory fit to the Michaelis-Menten kinetic model was obtained, presumably because the Km value is too high for this enzyme.

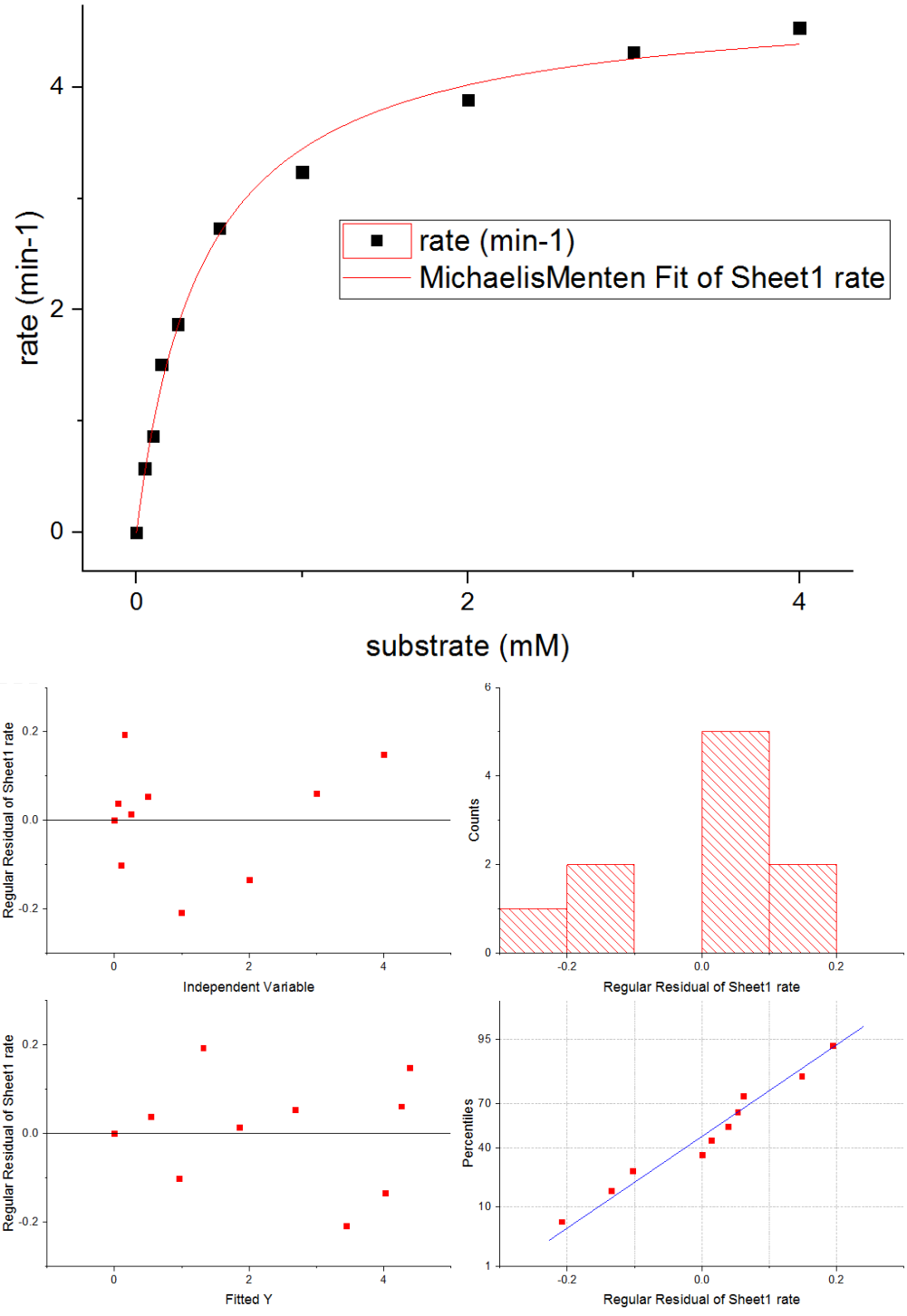


Ir(Me)-CYP119- C317G

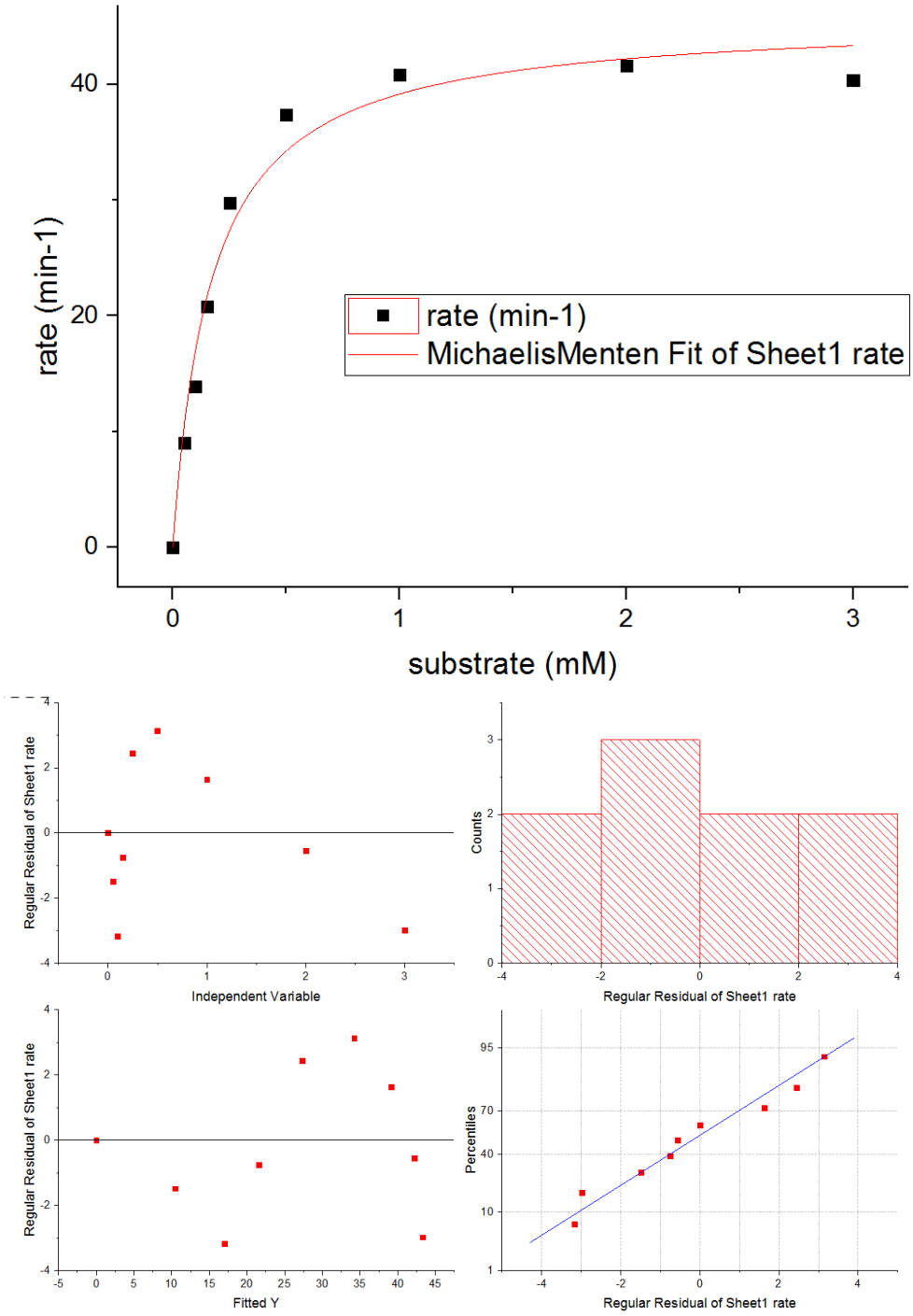




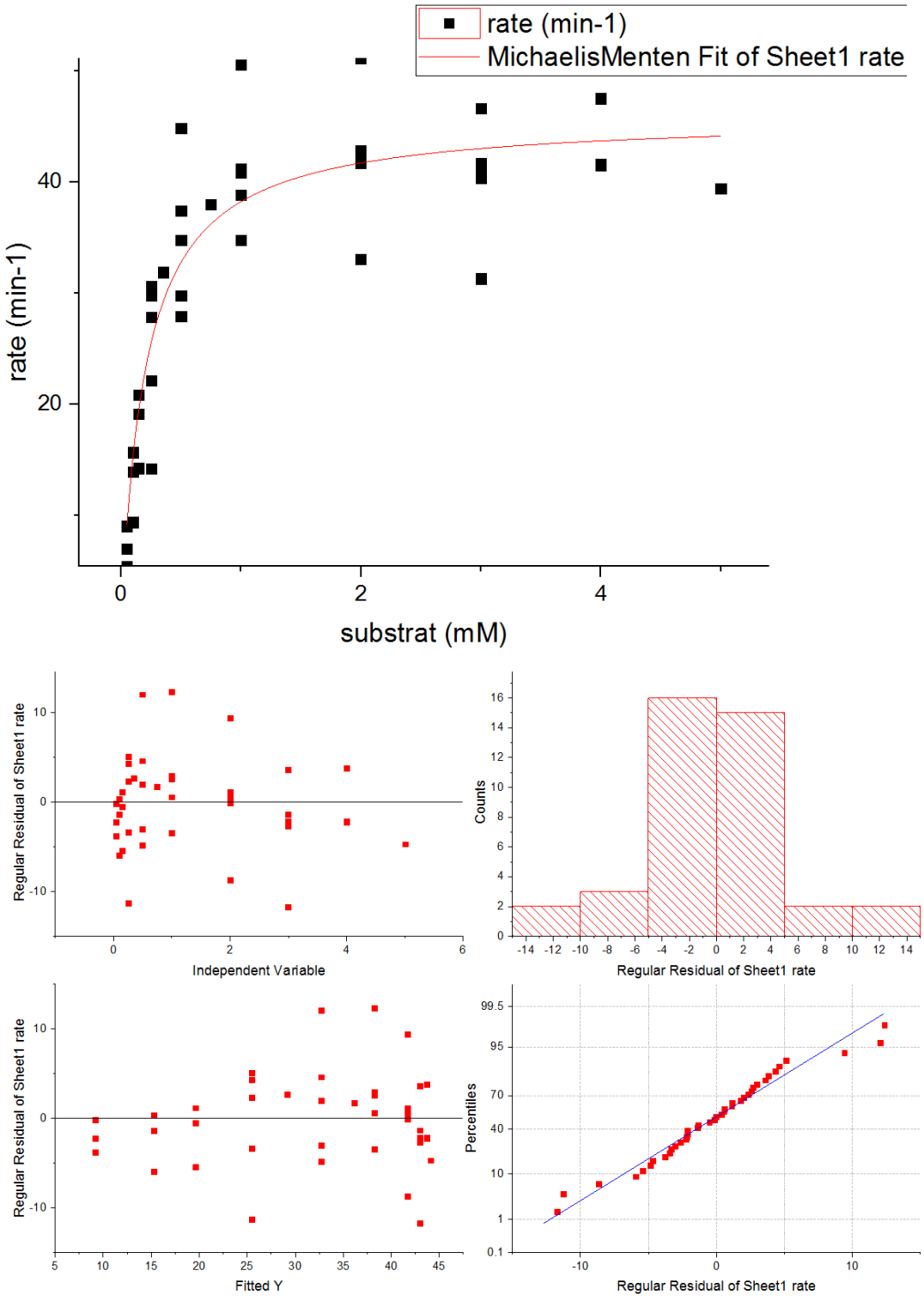
Ir(Me)-CYP119- C317G,T213A



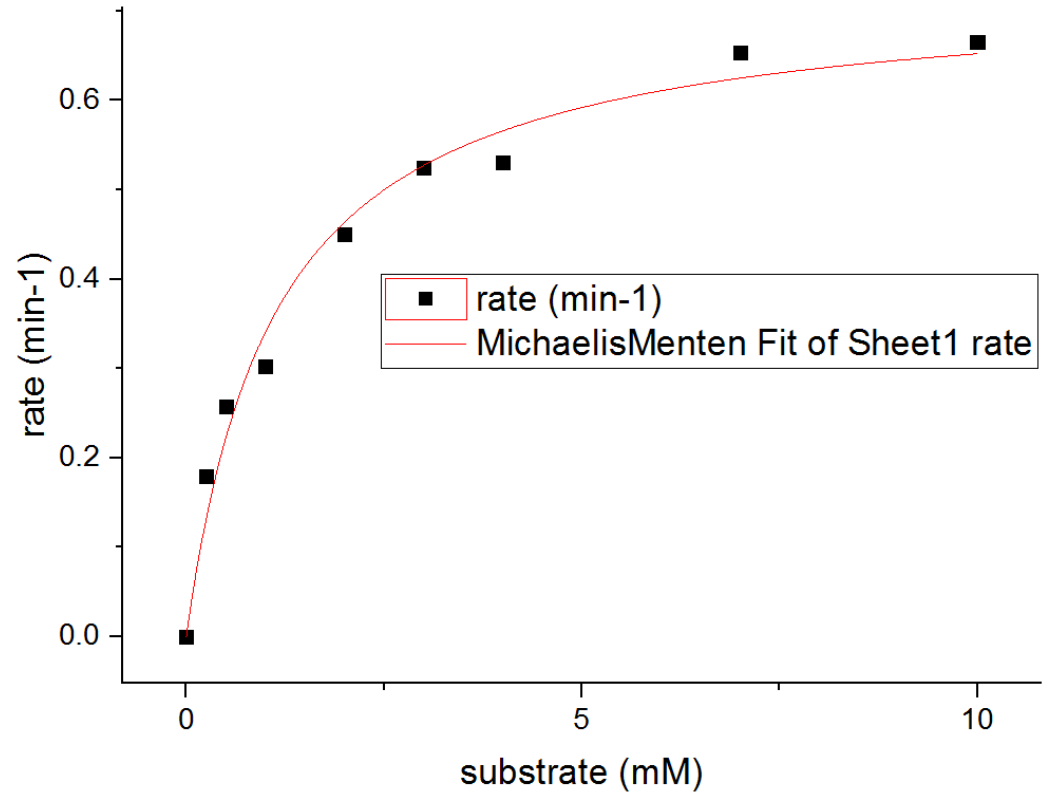
Ir(Me)-CYP119-Max (C317G,T213G,L69V,V254L) – data from one time point



Ir(Me)-CYP119-Max (C317G,T213G,L69V,V254L) - data from multiple experiments (multiple time point and catalyst concentrations):



Ir(Me)-myoglobin-H93A, H64V:



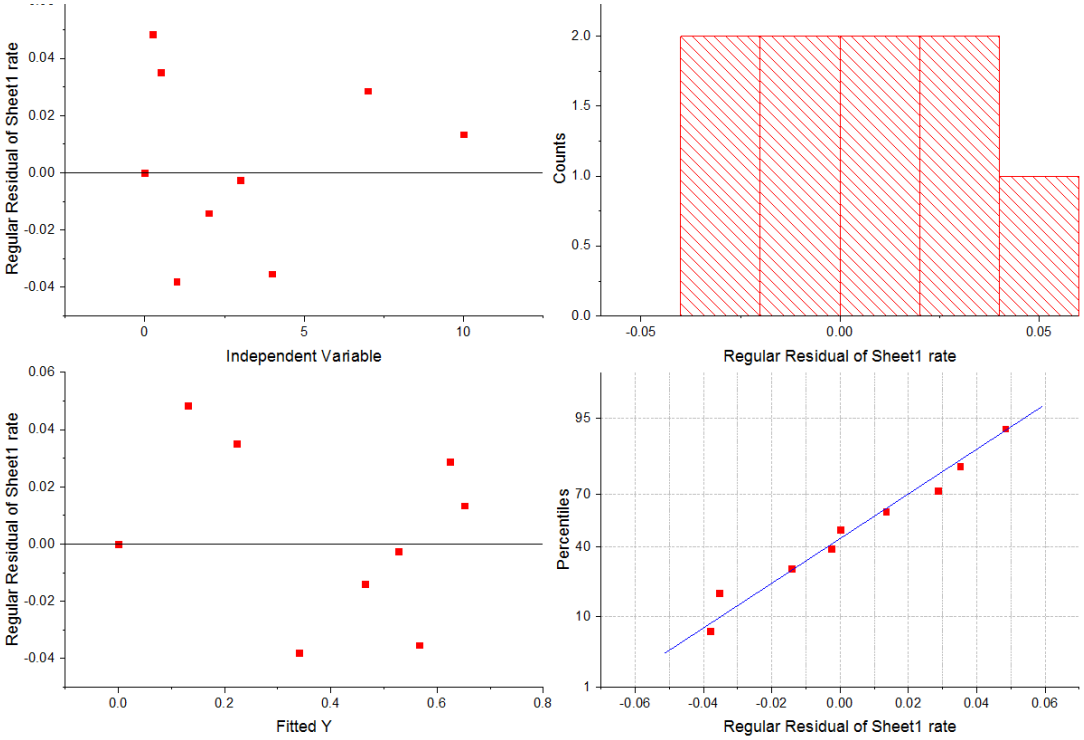
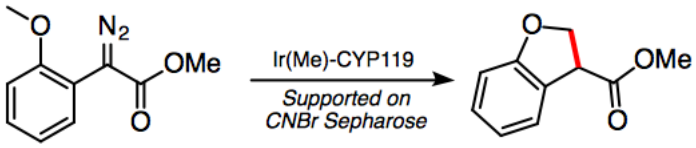


Figure S10. Enantioselectivities and TOFs of C-H insertion reactions catalyzed by Ir(Me)-PIX CYP119-MAX supported on sepharose activated with cyanogen bromide (CNBr). Catalyst preparation: 0.017 umol of Ir(Me)-CYP119-Max was supported on 0.1 mL of CNBr resin, according to the manufacturer's instructions. The resin was suspended in 0.25 mL NaPi, pH = 6.0 containing 100 mM NaCl, and 2.5 umol substrate was added as a solution in 5 uL of DMF. After 1 hour, the resin was collected by gravity filtration, washed with 10 column volumes of reaction buffer and re-subjected to a subsequent reaction under identical conditions. The flow through (containing the product) collected upon filtering the catalyst from the reaction was analyzed as described in section IV.



Initial use	-82% ee, 240 TOF (h^{-1})
Recycled 1 time	-81% ee, 244 TOF (h^{-1})
Recycled 4 times	-84% ee, 154 TOF (h^{-1})

4.10 Supporting Tables

Table S1. Sequence information for heme proteins used in this study

Amino acid sequences of proteins evaluated in this study

Protein	Organism	Construction	Vector	Sequence
mOCR-Myo-globin	*Physeter* macrocephalus	6xHis-TEV-mOCR-Myo	2BT	EGDIHMKSSHHHHHHENLYFQSNMSNMTYNNVFDHAYEML KENIRYDDIRDTDDLHDAIHMAADNAVPHYYADIRSVMAS EGIDLEFEDSGLMPDTKDDIRILQARIYEQLTIDLWEDAE DLLNEYLEEVEEYEEDEEGTGSETPGTSESGVLSEGEWQL VLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHL KTEAEMKASEDLKKHGVTVLTALGAILKKKGHHEAELKPL AQSHATKHKIPIKYLEFISEAIIHVLHSRHPGDFGADAQG AMNKALELFRKDIAAKYKELGYQG
P450 BM3	*Bacillus megaterium*	P450-BM3-6xHis	pcWori	MTIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADELGEI FKFEAPGRVTRYLSSQRLIKEACDESRFDKNLSQALKFAR DFAGDGLVTSWTHEKNWKKAHNILLPSFSQQAMKGYHAMM VDIAVQLVQKWERLNADEHIEVSEDMTRLTLDTIGLCGFN YRFNSFYRDQPHPFIISMVRALDEVMNKLQRANPDDPAYD ENKRQFQEDIKVMNDLVDKIIADRKARGEQSDDLLTQMLN GKDPETGEPLDDGNIRYQIITFLIAGHEATSGLLSFALYF LVKNPHVLQKVAEEAARVLVDPVPSYKQVKQLKYVGMVLN EALRLWPTAPAFSLYAKEDTVLGGEYPLEKGDEVMVLIPQ LHRDKTVWGDDVEEFRPERFENPSAIPQHAFKPFGNGQRA SIGQQFALHEATLVLGMMLKHFDFEDHTNYELDIKETLTL KPKGFVVKAKSKKIPLGGIPSPSTHHHHHH
P450 CAM	*Pseudomonas putida*	P450-CAM-6xHis	pcWori	MTTETIQSNANLAPLPPHVPEHLVFDFDMYNPSNLSAGVQ EAWAVLQESNVPDLVWTRCNGGHWIATRGQLIREAYEDYR HFSSECPFIPREAGEAYDFIPTSMDPPEQRQFRALANQVV GMPVVDKLENRIQELACSLIESLRPQGQCNFTEDYAEPFP IRIFMLLAGLPEEDIPHLKYLTDQMTRPDGSMTFAEAKEA

				LYDYLIPIIEQRRQKPGTDAISIVANGQVNGRPITSDEAK RMCGLLLVGGLDTVVNFLSFSMEFLAKSPEHRQELIQRPE RIPAACEELLRRFSLVADGRILTSDYEFHGVQLKKGDQIL L PQMLSGLDERENACPMHVDFSRQKVSHTTFGHGSHLCLGQ HLARRIIVTLKEWLTRIPDFSIAPGAQIQHKSGIVSGVQA LPLVWDPATTKAVHHHHHH
CYP 119	*Sulfolobus solfataricus*	6xHis-TEV-CYP119	2BT	EGDIHMKSSHHHHHHENLYFQSNAMYDWFSEMRKKDPVYY DGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLEDLRNG KIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLE TFIRETTRSLLDSIDPREDDIVKKLAVPLPIIVISKILGL PIEDKEKFKEWSDLVAFRLGKPGEIFELGKKYLELIGYVK DHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAGNETTTN LISNSVIDFTRFNLWQRIREENLYLKAIEEALRYSPPVMR TVRKTKERVKLGDQTIEEGEYVRVWIASANRDEEVFHDGE KFIPDRNPNPHLSFGSGIHLCLGAPLARLEARIAIEEFSK RFRHIEILDTEKVPNEVLNGYKRLVVRLKSNE

Table S2. Michaelis-Menten parameters for the carbene insertion of substrate **1** catalyzed by variants of Ir(Me)-CYP119 and Ir(Me)-myoglobin.

#	enzyme	k_{cat} (min^{-1})	K_m (mM)	k_{cat}/K_m ($min^{-1}mM^{-1}$)
1	Ir(Me)-CYP119 WT	-	> 5	-
2	Ir(Me)-CYP119-C317G	0.22 ± 0.01	3.1 ± 0.3	0.32
3	Ir(Me)-CYP119-C317G,T213A	4.8 ± 0.4	0.40 ± 0.03	12
4	Ir(Me)-CYP119-Max (C317G,T213G,L69V,V254L) - single experiment	45.8 ± 1.8	0.17 ± 0.03	269
5	Ir(Me)-CYP119-Max (C317G,T213G,L69V,V254L) - all data	45.9 ± 1.7	0.20 ± 0.03	230
6	Ir(Me)-myoglobin H93A, H64V	0.73 ± 0.03	1.1 ± 0.2	0.66

Table S3. Reactions of **2-6** on synthetic scale with variants of Ir(Me)-CYP119.

Subs.	Scale	Catalyst (variant)	Vol.	Time	Yield	ee	$[L]_D^{20}$
1	20.6 mg	0.17 mol% C317G, T213A, L155W	10 ml	24 h	5 mg of (R)-**2**	+81%	-22.0° (c=0.5 $CHCl_3$)
1	40 mg	0.05 mol% (317G, L69V, T213G, V254L)	10 ml	24 h	18 mg of (S)-**2**	-92%	
1	206 mg	0.017 mol% (317G, L69V, T213G, V254L)	50 ml	16 h	107 mg of (S)-**2**	-	
1	1000 mg	0.017 mol% (317G, L69V, T213G, V254L)	250 ml	16 h	480 mg of (S)-**2** + 72 mg rsm[b]	-93%	
S3	35.4 mg	0.17 mol% (317G, L69V, T213G, A152W)	15 ml	24 h	28.6 mg of **3**	-94%	+8.3° (c=0.6 $CHCl_3$)
S4	23.6 mg	0.17 mol% (317G, T213G, L155F)	10 ml	24h	10.3 mg of **4**	+95%	
S5	18.0 mg	0.17 mol% (317G, L69V, T213G, A152W)	7.5 ml	24 h	10.0 mg of (S)-**5**[a] (63%)	-85%	-25.3° (c=1.0 $CHCl_3$)
S6	20.0 mg	0.17 mol% (C317, V69W, T213G)	10 ml	24 h	10.0 mg of cis-	+94%	-56.8°

					(2S,3R)-**6**[a]		(c=0.5 $CHCl_3$)

[a] Absolute configuration assigned based on the specific rotation – product **2**[17], **5** and cis-**6**[17,23]
[b] rsm = recovered starting material

Table S4. Plasmid library obtained from site directed mutagenesis of the CYP119 gene within the 2B plasmid in the course of directed evolution.

	69X	209X	213X	254X
L		115-C10	116-G2	115-B12
V	115-D1	115-C2	116-F7	
A			116-H3	115-B6
G	115-G4	115-C8	115-F2	115-E8
Y	115-D3		116-D1	115-B8
W		115-C12	116-F1	115-B10
F	115-D10		115-H8	115-B11
T	115-D2			117-A1

T213G + …

	69X	209X	254X	310X	155X	152X
L			115-E3			
V	116-B1					
A			117-E4			
G	115-G2		117-E7			
Y	115-G5		116-E6			140-H7
W	116-A12		115-E2	140-B9		
F	115-H1		115-E7		141-C1	
T	115-G10	115-F7	115-E4		141-B1	

	69L	69V	69A	69G	69Y	69W	69F	69T
213L						116-B8		115-H9
213V		116-B10		116-A11	116-A10		116-B11	
213A		116-B3	116-B6			115-H11	116-B5	
213G								
213Y			116-B7		116-A9			115-H4
213W					116-B4			116-A5
213F			115-H5			115-H2	115-H7	
213T								

	254L	254V	254A	254G	254Y	254W	254F	254T
213L	116-D3					116-C10		116-C4
213V				116-C12				
213A	116-C3			116-E8			116-C6	
213G								
213Y			116-D7					
213W	116-E12				116-D12			116-E3
213F					116-E4	116-E10		
213T								

	209L	209V	209A	209G	209Y	209W	209F	209T
213L		116-F2				116-G6		
213V								
213A								
213G								
213Y								
213W							116-D2	
213F								
213T								

	209L	209V	209A	209G	209Y	209W	209F	209T
254L	117-C3					117-A10		
254V								
254A						117-A3		
254G								
254Y		117-A4				117-B3		
254W	117-C5	117-B11						
254F		117-C2		117-B7				
254T								

	69L	69V	69A	69G	69Y	69W	69F	69T
209L		117-F6		117-G8	117-F5	117-G6		117-H12
209V			117-F9			117-H3		117-H7
209A								
209G		117-H8	117-G4				117-G9	
209Y								
209W		117-H2		117-H6			117-F7	117-F11
209F								
209T			117-G5	117-H4	117-H5	117-H10		117-G1

213G, 254L (115-E3) + Additional Mutation

	69X	310X	318X
L		140-B10	
V	140-B6		
A	140-A3	140-D2	
G	140-B3	140-B8	
Y	140-A4	140-C5	
W	140-B2		
F	140-A8		
T		140-C1	

213G, 254L (115-E3) +

	152X	155X	
L	140-H1		
V	141-A5	141-C10	
A		141-B3	
G	141-A9	141-C3	
Y	140-H2	141-B11	
W	141-A11	141-B2	
F	140-H11	141-D3	
T	140-G9	141-G4	

213G, 69Y (115-G5) +

	69X	310X	318X
L	142-H4		
V			
A			
G		140-C11	
Y	142-H11		
W	142-G7	140-D9	
F			
T	142-H12		

213G, 69Y (115-G5) +

	152X	155X	
L			
V	140-G8		
A			
G			
Y	140-H10		
W	140-H6		
F			
T			

T213A (116-H3) +

	69X	254X	310X
L			166-A12
V			
A			
G		165-H1	
Y		165-H3	166-A6
W			
F		165-H6	166-A9
T		165-H7	166-A7

T213A (116-H3) +

	318X	152X	155X
L		166-C7	
V	166-B11		166-D2
A			
G			
Y	166-B1	166-C4	166-D7
W	166-B6		166-D5
F	166-B7	166-C3	166-D4
T	166-B3		166-D1

254X

	69F, 213G	69V, 213G	
L	141-F6	141-E9	
V			
A		141-E1	
G			
Y	141-D7	141-D9	
W	141-F1	141-E3	
F		141-E7	
T			

69Y, 254L, 213G +

	310X	155X	
L			
V			
A			
G			
Y			
W	140-B9	141-B12	
F			
T		141-B7	

69W, 213G (116-A12) +

	209X	254X	310X
L	165-A4	165-B6	
V	165-A5		165-C11
A		165-B2	165-C10
G	165-A11		165-C1
Y	165-A7	165-B8	
W		165-B3	165-C5
F	165-A2		
T		165-B7	165-C2

69W, 213G (116-A12) +

	318X	152X	155X
L		165-E8	165-F11
V	165-D9		
A			
G	165-D3		
Y	165-D7	165-E5	165-F5
W	165-D2	165-E7	165-F2
F		165-E1	165-F4
T	165-D1		165-F1

Table S5. Complete, tabulated results for the directed evolution of Ir(Me)-CYP119 for intramolecular C-H insertion reactions performed under the standard conditions for screening the enzymes. The identity of the mutants can be found in Table S5. Blank cells reflect that the mutant was not evaluated for the listed substrate.

Variant	Mutant Code	2-OMe		2,5-OMe		2,3-OMe		5-Cl		OBn		
		GC Yield (%)	ee (%)	GC Yield (%)	ee (%)	GC Yield (%)	ee (%)	GC Yield (%)	ee (%)	GC Yield (%)	ee (%)	dr (cis : trans)
1	116-B1	12	-28	17	-57	1	35			6	22	
2	116-C3	10	31	17	10	5	-17			10	-5	
3	116-F2	6	0	12	0	1	1			7	-2	
4	117-H5	6	-1	11	2	3	-61			8	7	
5	115-C8	10	24	18	17	3	29			12	0	
6	115-B6	8	2	17	1	4	-44			7	7	
7	115-F2	16	54	10	5	20	60					
8	115-E2	11	24	9	3	3	-41					
9	115-E4	14	29	12	4	4	22					
10	115-E7	7	9	8	4	4	-48					
11	116-B3	8	1	9	-8	2	11					
12	116-D1	23	67	46	69	1	20			7	-31	1.9
13	116-F7	5	14	17	11	3	21			2	10	0.8
14	116-G2	5	14	7	21	3	70			2	3	0.9
15	140-B9	21	-34	24	-14	16	-63			8	2	0.8
16	115-B11	8	23	8	40	3	5			2	-14	0.4
17	116-A12	64	-75	100	-97	87	59			33	90	4.4
18	116-H3	14	84	7	79	46	97			2	27	0.3
19	141-C1	43	49	80	43	92	83			16	82	2.3
20	116-E6	16	-19	50	23	65	31			22	-7	8.0
21	115-G5	42	-36	91	-93	81	87			34	79	4.1
22	141-B1	42	47	60	30	82	79			22	80	4.0
23	115-G4	9	7	8	12	1	26			3	8	0.5
24	115-H8	29	-16	80	-64	46	79			25	56	1.6
25	115-D1	14	21	17	20	12	67			4	52	1.2
26	115-H1	58	-45	75	-95	100	90			30	79	5.6
27	116-B11	10	4	12	-5	4	9			2	19	0.5
28	115-G2	15	-12	23	-47	6	-2			8	44	1.0
29	116-F1	25	-16	49	-22	65	36			21	-15	9.4
30	115-G10	15	-24	90	-83	10	69			15	66	4.3
31	115-E3	28	14	43	32	13	-44			10	-61	1.7
32	117-A1	28	6	36	3	7	15			10	10	0.8
33	115-F7	45	26	36	-26	50	65			22	76	7.0
34	117-E7	35	28	31	30	43	65			13	-1	1.2
35	115-D3	11	-2	13	8	1	34			3	26	0.4
36	116-B5	8	24	13	16	26	95			14	73	3.7
37	115-G11	8	12	11	3	10	89			4	34	0.8
38	115-E8	19	9	16	7	29	-47			8	-39	1.0
39	116-E8	36	56	27	47	69	48			6	-53	1.9
40	116-B11	15	6	25	3	3	25			3	25	0.7
41	115-D10	12	4	10	3	3	54			4	45	0.9
42	115-D2	12	4	9	14	1	2			2	2	0.5
43	115-B12	17	22	10	39	8	86			3	38	1.0
44	115-C12	13	7	9	4	11	27			2	6	0.5
45	115-C10	15	2	21	-63	3	61			9	40	0.8
46	142-H4	73	-82	56	-95	1	-67					
47	141-E1	60	-44	66	-55	73	27					
48	115-B8	6	67	25	54	5	-18					
49	141-E3	51	-60	28	-35	18	-31					
50	140-C11	48	1	72	-2	27	-1					
51	116-A10	3	0	1	0	0	ND					
52	140-D9	21	-10	31	-52	21	32					
53	142-H11	11	-30	14	-80	13	67					
54	117-G4	17	50	13	42	16	89					
55	142-G7	20	24	8	-25	7	7					
56	116-G6	9	3	8	19	2	11					
57	115-B10	29	2	9	7	4	-7					
58	140-H10	33	-6	85	-90	1	51					

Table S5 continued from previous page.

59	140-A4	36	21	68	-82	89	-64		
60	140-B2	2	-2	6	5	6	-12		
61	140-G8	51	-49	97	-95	83	65		
62	140-A3	97	-88	100	-88	97	-44		
63	140-B3	35	-59	51	-71	40	-32		
64	140-B6	77	-91	105	-92	107	-59		
65	141-F6	92	-82	95	-89	111	-55		
66	142-H12	26	-28	104	-81	35	39		
67	141-E7	24	-65	25	-61	21	0		
68	141-E9	97	-91	83	-93	101	-61		
69	141-D9	29	-32	14	2	16	-3		
70	116-C4	9	15	15	20	1	41	0	ND
71	116-D3	13	22	4	11	4	65	20	-6
72	116-A9	17	0	36	-2	3	29	4	-9
73	141-D7	26	-5	10	-59	26	61	10	-6
74	116-B4	0	ND	7	0	5	22	6	1
75	116-A11	11	9	17	2	2	9	11	15
76	141-F1	17	33	12	-20	8	17	8	-4
77	140-H6	0	ND	145	-97	15	67	54	-93
78	116-C12	23	0	4	-6	5	-47	7	-10
79	116-B8	22	1	26	3	2	16	2	3
80	116-C3	0	ND	34	87	70	92	14	12
81	140-A8	88	-85	105	-94	135	-76	23	-88
82	140-G9	66	-32	59	5	114	1	19	2
83	116-B10	6	12	12	12	12	10	48	0
84	140-H11	63	-49	67	-19	94	-27	21	2
85	141-A5	50	-46	47	4	57	-11		
86	140-A8	5	4	4	-12	6	-22		
87	140-C5	25	-6	51	-34	81	35		
88	141-B3	44	-51	55	-2	108	-14		
89	140-B10	23	-17	40	-16	50	19		
90	141-A11	71	-52	66	-39	82	-9		
91	141-B2	57	-53	64	8	83	2		
92	140-H1	68	-52	76	-15	109	-26		
93	141-C10	52	-34	30	-3	1	-1		
94	141-B11	79	-38	68	-18	129	1		
95	141-A9	56	-32	17	8	80	0		
96	141-D3	86	-35	62	-10	131	3		
97	115-C2	9	8	4	1				
98	140-D2	39	-13	24	-22				
99	117-H4	10	2	3	-1				
100	117-G6	6	2	7	-4				
101	117-H3	8	5	12	1				
102	117-H10	13	1	4	-5				
103	117-F9	6	7	4	-8				
104	117-H6	16	4	23	-3				
105	117-H2	3	-1	7	1	1	64	1	-1
106	117-G5	1	0	2	0	1	0	2	4
107	117-H7	1	3	2	1	1	-9	2	-7
108	117-H7	8	7	3	1	1	1	4	1
109	117-H8	11	63	19	65	39	96	6	3
110	117-F5	1	3	5	-1	2	47	1	1
111	117-G9	9	29	11	56	35	95		
112	117-F11	5	-4	7	1	2	11		
113	117-A10	16	31	7	30	6	50		
114	117-H6	5	72	3	7	20	95		
115	167-B4	8	41	15	72	13	39		
116	166-F2	36	51	69	65	83	62		
117	167-A6	9	24	15	22	1	30		
118	166-H9	6	3	15	4	4	7		
119	166-G3	12	3	24	0	5	3		
120	167-A3	3	26	8	17	10	48		

Table S5 continued from previous page.

121	166-H3	16	3	22	0	4	9					
122	166-G6	7	0	19	1	6	3					
123	167-A4	9	11	19	9	8	19					
124	166-D4	46	88	57	84	94	97					
125	166-A7	10	22	6	18	13	92					
126	166-B6	5	1	1	14	1	44					
127	166-C3	8	90	9	72	20	93					
128	166-A12	5	27	6	12	7	80					
129	166-A6	9	49	9	39	14	90					
130	166-C7	10	87	8	75	20	88					
131	166-B3	2	29	3	16	2	79					
132	166-D7	16	91	14	73	ND	ND					
133	165-C1	19	0	47	15	3	6					
134	166-D2	8	80	6	60	8	92					
135	166-B11	7	55	4	49	6	89					
136	166-D5	40	93	19	70	19	94					
137	166-A9	4	84	4	57	10	95					
138	166-B7	1	59	2	11	2	49					
139	167-B5	21	16	21	-42	1	22					
140	167-A2	13	18	29	10	20	51					
141	167-B2	20	14	36	1	9	15					
142	167-B1	16	21	23	13	15	43					
143	166-f8	55	22	54	42	69	-67					
144	167-B10	17	26	28	12	17	31					
145	166-F1	44	15	3	38	11	-12					
146	167-B9	51	6	7	4	11	21					
147	166-F5	24	2	13	5	8	-36					
148	166-F3	25	-4	23	20	8	19					
149	165-C11	22	2	10	-25	5	3					
150	165-D9	56	-52	98	-80	43	54					
151	165-C10	8	-25	11	-81	1	18	6	-66			
152	165-C2	15	-44	55	-89	14	36	9	-84			
153	165-D3	6	-25	8	-62	3	9	2	-72			
154	165-E8	79	-60	88	-97	64	52	21	-56			
155	165-D2	2	-32	0	ND	1	33	1	-83			
156	165-D1	8	-52	37	-94	9	42	4	-91			
157	165-E5	52	-67	96	-95	51	34	26	2			
158	165-D7	6	-25	20	-81	4	31	4	-82			
159	165-F5	80	-68	111	-95	65	42	14	-84			
160	165-F2	80	-66	124	-95	54	42	36	10			
161	165-F4	79	-66	102	-95	53	49	32	-94			
162	165-C11	0	ND	139	-96	66	25	71	-93			

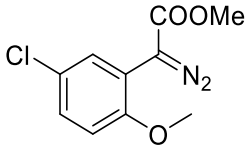
Table S6. Methods used to separate enantiomers of the reaction products.

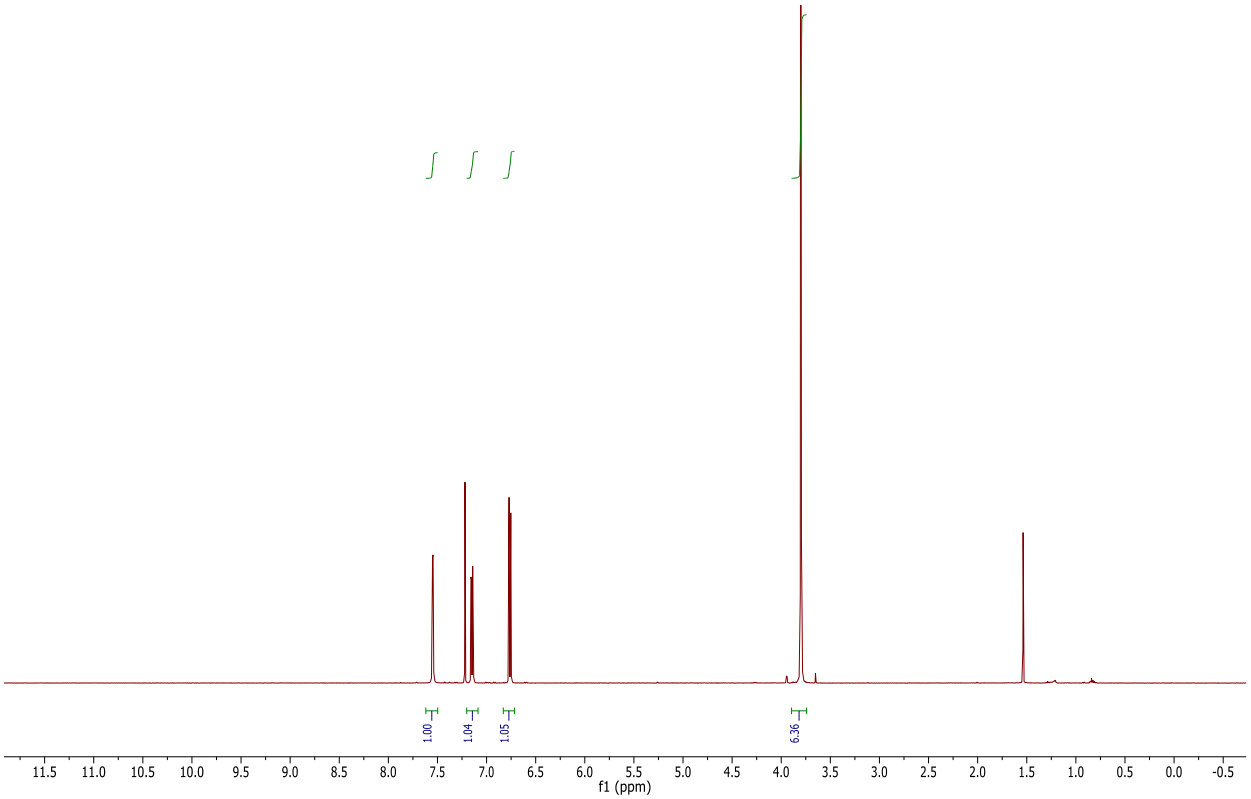
Product	Instrument	Column	Method	Retention Times
2	GC	CYCLOSIL-B (30m x 0.25mm x 0.25u)	Isothermal: 135 °C	t1(R)(-)=20.0 min t2(S)(+)=20.5 min
3	GC	CYCLOSIL-B (30m x 0.25mm x 0.25u)	Isothermal: 145 °C	t1(-)=44.7 min t2(+)=45.6 min
3	SFC	Chiracel OD-H (Diacel)	Isocratic: 1% MeOH, 4 ml/min flow	t1(-)=2.3 min t2(+)=3.4 min
4	GC	CYCLOSIL-B (30m x 0.25mm x 0.25u)	Isothermal: 165 °C	t1(-)=19.5 min t2(+)=20.0 min
5	GC	CYCLOSIL-B (30m x 0.25mm x 0.25u)	Isothermal: 150 °C	t1(R)(-)=28.8 min t2(S)(+)=29. min
6	SFC	Chiracel OJ-H (Diacel)	Isocratic: 1% MeOH, 4 ml/min flow	Trans: t1=1.85 min t2=2.1 min Cis: t1(2S,3R)(-)=4.6 min t2(2R,3S)(+)=6.1 min
8	GC	CYCLOSIL-B (30m x 0.25mm x 0.25u)	Isothermal: 140 °C	t1=13.1 min t2=13.4 min
11	SFC	Chiracel AZ-H (Diacel)	Isocratic: 2% MeOH, 4 ml/min flow	t1=4.6 min t2=5.3 min

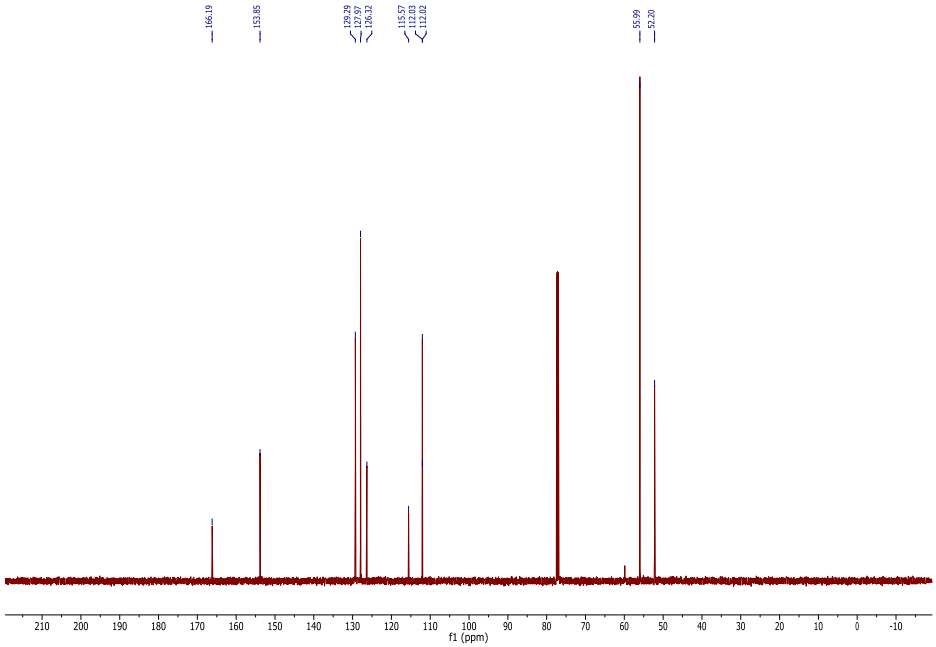
Table S7. Melting temperatures of various Ir(Me)-PIX heme proteins, as determined by DSF using SYPRO Orange as a fluorescent reporter.

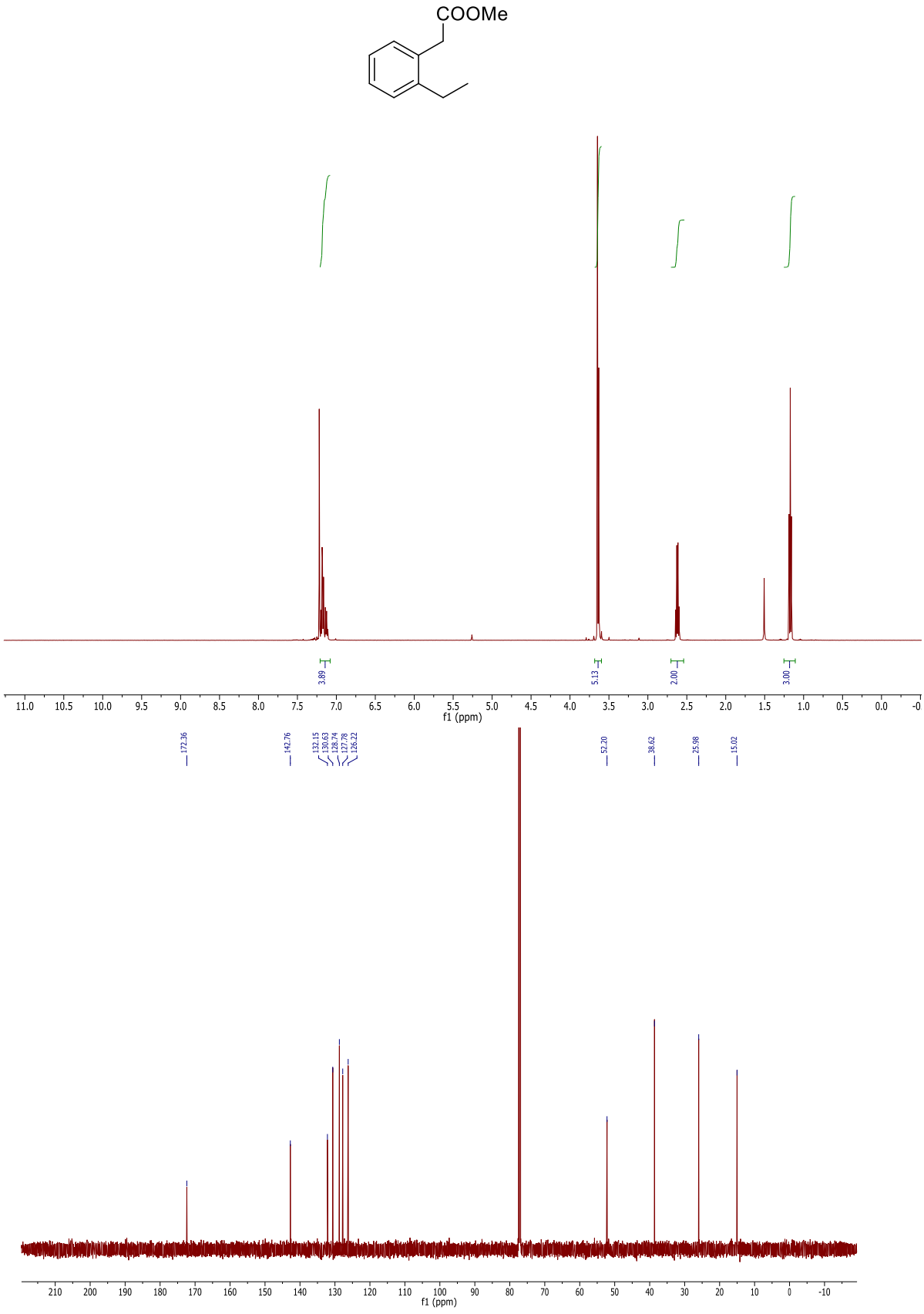
Protein	**[Protein] (mM)**	**PIX : Protein Stoichiometry**	**[SYPRO Orange]**	**Heating Rate**	**T_m (^{o} C)**
Ir(Me)-PIX CYP119	0.1 mM	1 : 1	20 X	1^o C/min	69^o C
Ir(Me)-PIX P450 BM3	0.1 mM	1 : 1	20 X	1^o C/min	45^o C
Ir(Me)-PIX P450 CAM	0.1 mM	1 : 1	20 X	1^o C/min	40^o C

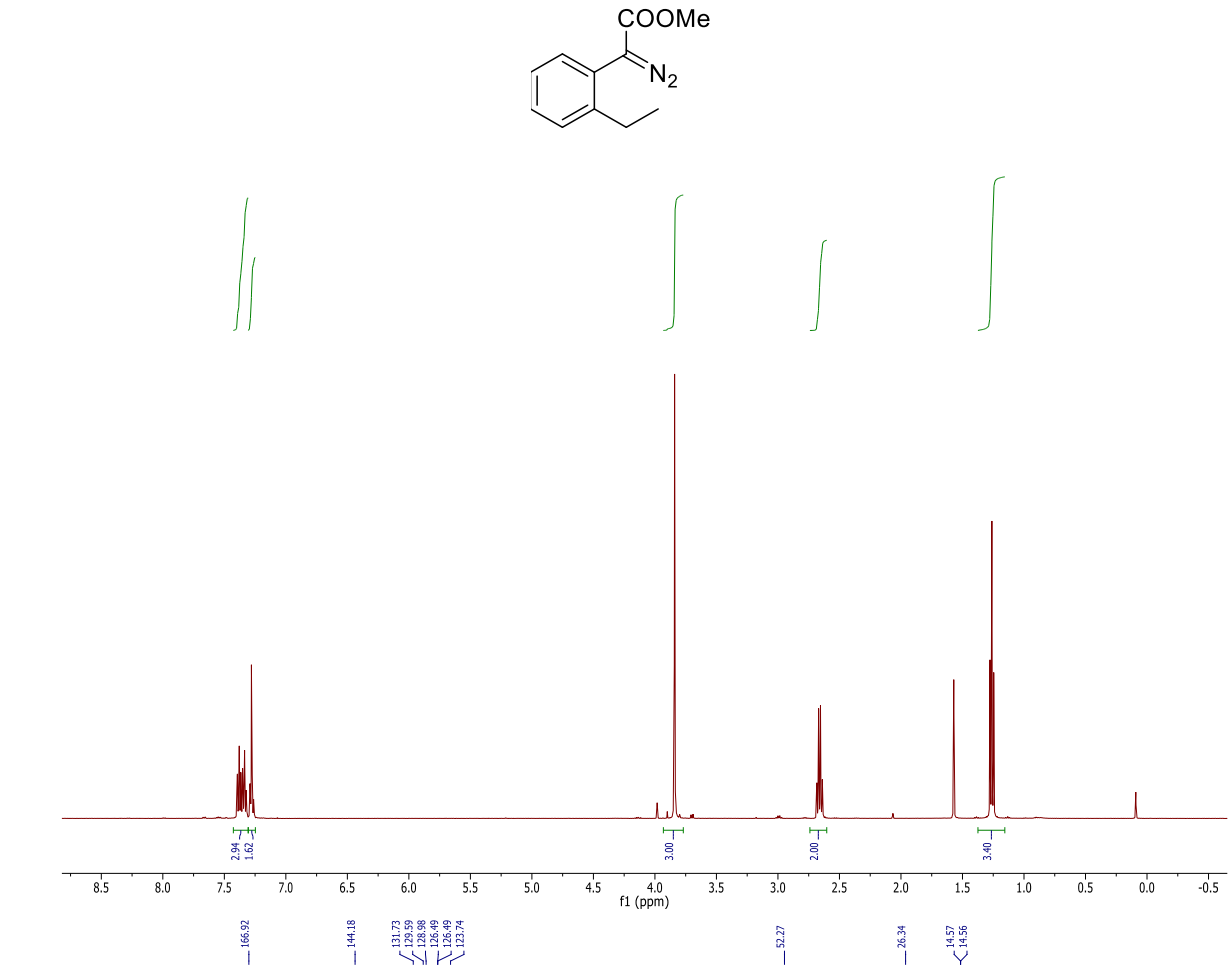
4.11 NMR Spectra

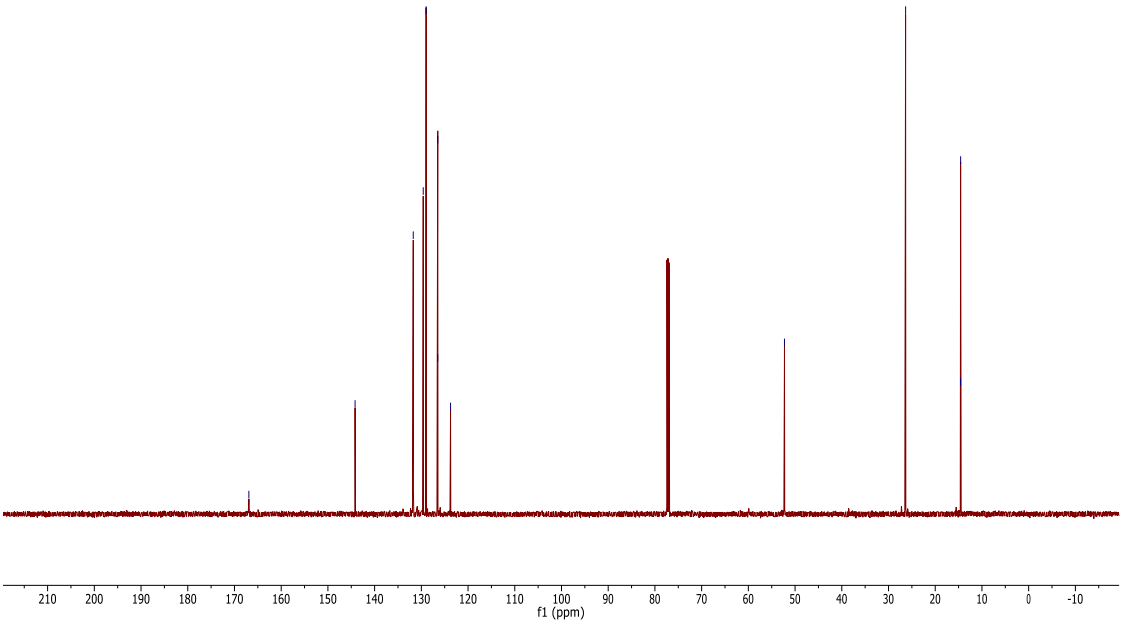







COOMe
3.89
5.13
2.00
3.00
f1 (ppm)
172.36
142.76
132.15
130.63
128.74
127.78
126.22
52.20
38.62
25.98
15.02
f1 (ppm)


COOMe
N2
2.94
1.62
3.00
2.00
3.40
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.5
f1 (ppm)
166.92
144.18
131.73
129.59
128.98
126.49
126.49
123.74
52.27
26.34
14.57
14.56


210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)

4.12 Acknowledgements

This work was supported by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02- 05CH11231, by the NSF (graduate research fellowship to HMK), and the NWO Netherlands Organization for Scientific Research (the Rubicon postdoctoral fellowship, No. 680-50-1306 to PD). We thank the QB3 MacroLab facility (competent cells), the UC Berkeley DNA Sequencing Facility (plasmid sequencing), and Dr. Tony Iavarone and the QB3 Mass Spectrometry Facility (NS-ESI data collection, supported by NIH grant 1S10OD020062-01), for nativeNS-ESI-MS data and analysis.

References and Notes:

1. Bornscheuer, U. T. *et al*. Engineering the third wave of biocatalysis. *Nature* **485,** 185–194 (2012).
2. Koeller, K. M. & Wong, C.-H. Enzymes for chemical synthesis. *Nature* **409,** 232–240 (2001).
3. Peters, M. W., Meinhold, P., Glieder, A. & Arnold, F. H. Regio- and Enantioselective Alkane Hydroxylation with Engineered Cytochromes P450 BM-3. *J. Am. Chem. Soc.* **125,** 13442–13450 (2003).
4. Agudo, R., Roiban, G.-D. & Reetz, M. T. Achieving regio- and enantioselectivity of P450-catalyzed oxidative CH activation of small functionalized molecules by structure-guided directed evolution. *ChemBioChem* **13,** 1465–1473 (2012).
5. Hyster, T. K. & Arnold, F. H. P450BM3-Axial Mutations: A Gateway to Non-Natural Reactivity. *Isr. J. Chem.* **55,** 14–20 (2015).
6. Lewis, J. C. Artificial Metalloenzymes and Metallopeptide Catalysts for Organic Synthesis. *ACS Catal.* **3,** 2954–2975 (2013).
7. Ringenberg, M. R. & Ward, T. R. Merging the best of two worlds: artificial metalloenzymes for enantioselective catalysis. *Chem. Commun.* **47,** 8470–8476 (2011).
8. Hyster, T. K., Knörr, L., Ward, T. R. & Rovis, T. Biotinylated Rh(III) complexes in engineered streptavidin for accelerated asymmetric C-H activation. *Science* **338,** 500–503 (2012).
9. Srivastava, P., Yang, H., Ellis-Guardiola, K. & Lewis, J. C. Engineering a dirhodium artificial metalloenzyme for selective olefin cyclopropanation. *Nat Commun* **6,** 7789 (2015).
10. Bar-Even, A. *et al*. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50,** 4402–4410 (2011).
11. Ringe, D. & Petsko, G. A. Biochemistry. How enzymes work. *Science* **320,** 1428–1429 (2008).
12. Key, H. M., Dydio, P., Clark, D. S. & Hartwig, J. F. Abiological catalysis by artificial haem proteins containing noble metals in place of iron. *Nature* **534,** 534–537 (2016).
13. Brucker, E. A., Olson, J. S., Phillips, G. N., Dou, Y. & Ikeda-Saito, M. High resolution crystal structures of the deoxy, oxy, and aquomet forms of cobalt myoglobin. *J. Biol. Chem.* **271,** 25419–25422 (1996).
14. *The Ubiquitous Roles of Cytochrome P450 Proteins*. (John Wiley & Sons, Ltd, 2007). doi:10.1002/9780470028155
15. Whitehouse, C. J. C., Bell, S. G. & Wong, L.-L. P450(BM3) (CYP102A1): connecting the

dots. *Chem. Soc. Rev.* **41,** 1218–1260 (2012).
16. Rabe, K. S., Kiko, K. & Niemeyer, C. M. Characterization of the Peroxidase Activity of CYP119, a Thermostable P450 FromSulfolobus acidocaldarius. *ChemBioChem* **9,** 420–425 (2008).
17. Bongen, P., Pietruszka, J. & Simon, R. C. Dynamic kinetic resolution of 2,3-dihydrobenzo[b]furans: chemoenzymatic synthesis of analgesic agent BRL 37959. *Chem. Eur. J.* **18,** 11063–11070 (2012).
18. Noble, M. A. *et al.* Roles of key active-site residues in flavocytochrome P450 BM3. *Biochemical Journal* **339 (Pt 2),** 371–379 (1999).
19. Paradine, S. M. & White, M. C. Iron-Catalyzed Intramolecular Allylic C–H Amination. *J. Am. Chem. Soc.* **134,** 2036–2039 (2012).
20. Hong, B. *et al.* Enantioselective Total Synthesis of (–)-Incarviatone A. *J. Am. Chem. Soc.* **137,** 11946–11949 (2015).
21. Fulmer, G. R. *et al.* NMR Chemical Shifts of Trace Impurities: Common Laboratory Solvents, Organics, and Gases in Deuterated Solvents Relevant to the Organometallic Chemist. *Organometallics* **29,** 2176–2179 (2010).
22. Ward, T. R. Artificial Metalloenzymes Based on the Biotin–Avidin Technology: Enantioselective Catalysis and Beyond. *Acc. Chem. Res.* **Online,** 47–57 (2010).
23. Hiroaki Saito *et al.* Enantio- and Diastereoselective Synthesis of cis-2-Aryl-3-methoxycarbonyl-2,3-dihydrobenzofurans via the Rh(II)-Catalyzed C–H Insertion Process. *Org. Lett.* **4,** 3887–3890 (2002).
24. Nicolle, S. M. & Moody, C. J. Potassium N-Iodo p-Toluenesulfonamide (TsNIK, Iodamine-T): A New Reagent for the Oxidation of Hydrazones to Diazo Compounds. *Chem. Eur. J.* **20,** 4420–4425 (2014).
25. Han, C. & Buchwald, S. L. Negishi Coupling of Secondary Alkylzinc Halides with Aryl Bromides and Chlorides. *J. Am. Chem. Soc.* **131,** 7532–7533 (2009).
26. Huw M L Davies, Mônica V A Grazini, A. & Aouad, E. Asymmetric Intramolecular C–H Insertions of Aryldiazoacetates. *Org. Lett.* **3,** 1475–1477 (2001).
27. Pietruszka, J. R., Simon, R. C., Kruska, F. & Braun, M. Dynamic Enzymatic Kinetic Resolution of Methyl 2,3-Dihydro-1 H-indene-1-carboxylate. *European Journal of Organic Chemistry* **2009,** 6217–6224 (2009).

Chapter 5: Iridium-Containing P450 Enzymes for Highly Selective Cyclopropanations of Structurally Diverse Alkenes

This chapter is modified from a manuscript that has been prepared for publication with permission of the university and all coauthors

Pawel Dydio contributed equally to the work described in this chapter.

ABSTRACT: Enzymes catalyze organic transformations with exquisite levels of selectivity, including chemo-, regio-, stereoselectivity, and substrate selectivity. Our prior work revealed that artificial P450s, containing an Ir(Me) site in place of Fe in the porphyrin cofactor, form catalysts for enantioselective insertions of carbenes into C-H insertion bonds. To expand upon the scope of transformation catalyzed by Ir(Me)-P450s and to more fully delineate the potential of these enzymes to catalyze reactions with a range of modes of selectivity, we applied variants of Ir(Me)-PIX CYP119 enzymes as catalysts for cyclopropanation reactions of alkenes. In particular, we sought catalysts to react with alkenes that are less reactive than simple vinylarenes, substrates for which Fe-P450s are active catalysts for cyclopropanation. Here we report that directed evolution of CYP119 containing Ir(Me)-PIX forms artificial enzymes that are highly active and highly stereoselective catalysts for the addition of carbenes to a variety of alkenes that are less reactive than terminal vinylarenes, including terminal and internal, activated and unactivated, electron-rich and electron-deficient alkenes. In particular, we show that Ir(Me)-PIX enzymes derived from CYP119 catalyze highly enantio-, diastereo-, and site-selective cyclopropanations of vinylarenes with up to 99% ee, 200 : 1 dr, 80% yield, and 10,000 turnovers (TON) and of internal and unactivated alkenes with up to 99% ee, 200 : 1 de, 76% yield, and 1300 TON. Finally, we show that certain evolved mutants of Ir(Me)-PIX P450s exhibit substrate selectivity reminiscent of natural enzymes, reacting preferentially with less reactive internal alkenes in the presence of more reactive terminal alkenes. In addition to expanding substantially the scope of substrates undergoing enzyme-catalyzed cyclopropanations, Ir(Me)-PIX P450s constitute a class of artificial metalloenzymes that possesses a range of selectivity modes including those previously attained only with natural enzymes.

5.1 Introduction

Enzymes catalyze organic transformations with exquisite levels of selectivity, including chemo-, regio-, and stereoselectivity (both enantio- and diastereo-), as well as substrate selectivity.[1] In addition, enzymes are often specific for one substrate among a mixture of potential reactants. In fact, many enzymatic reactions are accomplished by combining all of these forms of selectivity to effect a single transformation (Fig. 1).[2] The selective oxidation of cholesterol to various steroids is an important example of this capability of enzymes.

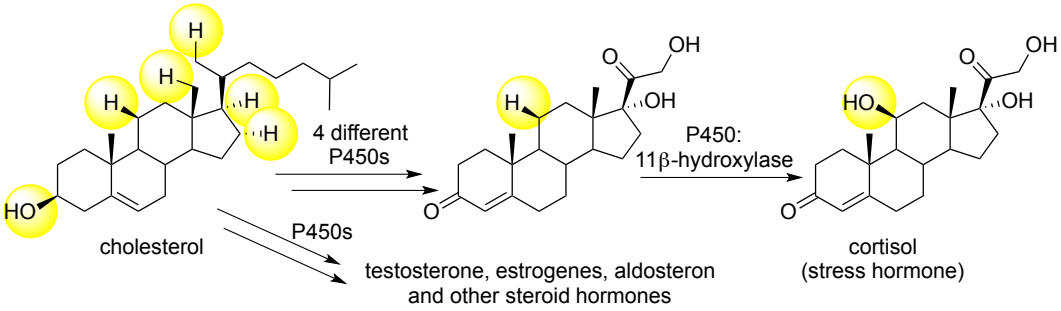


Figure 1. The biosynthesis of cortisol (and other steroid hormones) from cholesterol by discrete, sequential P450 oxidation.

In many cases, the scope of substrates with which natural enzymes react can be readily expanded by protein engineering and directed evolution, providing enzymes that perform reactions for abiological substrates with the mode and levels of selectivity typical of those achieved in biological reactions.[3] For instance, cytochrome P450 enzymes that catalyze highly selective C-H hydroxylation reactions of complex natural products have been evolved to react selectively with various unnatural substrates ranging from small hydrocarbons to large steroids.[4] In these cases, directed evolution furnishes not only enzymes that react with new substrates, but enzymes that functionalize individual C-H bonds in an enantio- and diastereo-selective fashion.

As described in the introduction and previous chapters, selective artificial metalloenzymes created by incorporating abiological transition-metal complexes into natural proteins have been achieved.[5] However artificial metalloenzymes that control selectivity beyond enantioselectivity are particularly rare. For example, site-selective reactions of substrates possessing multiple similarly reactive sites have not been achieved by artificial metalloenzymes. Likewise, the reaction of an inherently less reactive substrate in the presence of a more reactive analog has not been achieved. The strategy to prepare artificial metalloenzymes by anchoring transition metal complexes within binding sites of natural proteins imposes significant challenges when seeking to mimic the exquisite selectivities of natural enzymes. Incorporation of unnatural metal complexes within the binding sites of proteins changes significantly the potential interactions between the protein and the substrate.

As described in the prior chapters (Chapter 3-4), we disclosed that natural metalloproteins reconstituted with native cofactors carrying abiological metal centers display abiological catalytic activity, while retaining a native, evolvable binding site for the substrate. Our original system describing myoglobin containing an Ir(Me)-porphyrin (Ir(Me)-PIX) in place of the natural heme catalyzed enantioselective C-H insertion reactions with moderate enantioselectivities (50-84% ee).[6] Preliminary studies described in the same initial report found

that Ir(Me)-PIX myoglobins catalyzed the cyclopropanation of internal and terminal unactivated alkenes, but with low activity (20-40 TON) and modest selectivity (40-80% ee). Subsequently, we found that replacing the myoglobin scaffold with that of a thermally stable P450 created enzymes with higher suitability to mimic the properties of natural enzymes, such as high activity and high enantioselectivity in intramolecular C-H insertion reactions. Therefore, we considered that Ir(Me)-PIX-P450 enzymes could also function as highly selective and active catalysts for intermolecular cyclopropanation reactions.

In addition to the synthetic utility of cyclopropanation reactions, this reaction offers the chance to explore more widely the potential Ir(Me)-PIX P450s to be evolved to react with a range of substrates and with a range of modes of selectivity. The reactivity of alkenes toward cyclopropanation varies widely, from the most reactive simple styrenes to the less reactive internal, aliphatic or electron-deficient alkenes, and few catalysts react with this entire range of olefins.[7] Furthermore, the addition of a carbene to an alkene usually forms two or more stereocenters, which means that catalysts must react with both high enantioselectivity and high diastereoselectivity. Finally, poly-alkenes could be useful synthetic building blocks if methods were known for site-selective functionalization of alkenes. Therefore, using the cyclopropanation reaction as a test case, we can evaluate the possibility to use artificial metalloenzymes for functionalization reactions that are site- and substrate-selective.

Here we report that directed evolution of P450s containing Ir(Me)-PIX forms artificial enzymes that are highly active and highly stereoselective catalysts for the addition of carbenes to a variety of alkenes that are less reactive than terminal vinylarenes, including terminal and internal, activated and unactivated, electron-rich and electron-deficient alkenes. In particular, we show that Ir(Me)-PIX enzymes derived from CYP119 catalyze highly enantio-, diastereo-, and site-selective cyclopropanations of vinylarenes with up to 99% ee, 200 : 1 dr, 80% yield, and 10,000 turnovers (TON) and of internal and unactivated alkenes with up to 99% ee, 200 : 1 de, 76% yield, and 1300 TON. In the cases of enantioselective cyclopropanation reactions, discrete mutants are identified to form each enantiomer of the products. Finally, certain evolved mutants of Ir(Me)-PIX P450s react preferentially with less reactive internal alkenes in the presence of more reactive terminal alkenes, offering an unprecedented example of substrate selectivity by an artificial metalloenzyme. In addition to expanding substantially the scope of substrates undergoing enzyme-catalyzed cyclopropanations, Ir(Me)-PIX P450s constitute a class of artificial metalloenzymes that possesses a wide range of selectivity modes previously attained only with natural enzymes.

5.2 Cyclopropanation of vinylarenes

To generate a new class of cyclopropanation catalysts utilizing the substrate binding site of P450 enzymes, we used thermophilic cytochrome CYP119 from Sulfolobus solfataricus as a protein scaffold.[8] CYP119 containing Ir(Me)-PIX can be prepared easily by of direct expression of apo-protein followed by its direct reconstitution with Ir(Me)-PIX cofactor without any additional steps of purification (methodology described fully in Chapter 4).

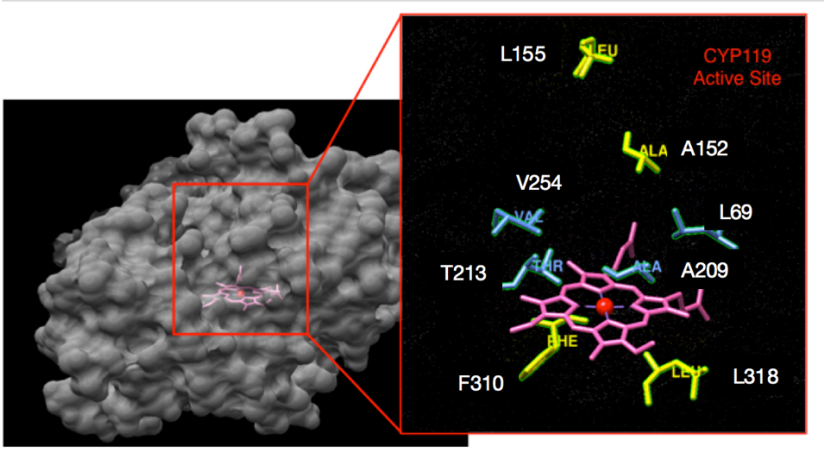


Figure 2. The residues surrounding the poprphyrin and active site of CYP119 that were modified in the molecular evolution of CYP119 containing an Ir(Me)-PIX cofactor.

To begin to evaluate the potential of Ir(Me)-PIX CYP119 enzymes to catalyze cyclopropanation (Fig. 2), we evaluated reactions of ethyl diazoacetate (EDA) with a series of vinylarenes **1-4** in the presence of Ir(Me)-CYP119 (Fig. 3). While highly reactive terminal alkenes **1** and **2** react with EDA in the presence of engineered P450s and myoglobins containing the native Fe-PIX cofactor,[9] no Fe-proteins have been found to catalyze a reaction of EDA with more sterically hindered internal alkenes **3** and **4**. Although the activity of WT Ir(Me)-PIX CYP119 for cyclopropanation of **1-4** was low, the incorporation of mutations at several sites within the active site of the enzymes generated mutants with activities and stereoselectivities that are higher than those of reactions catalyzed by the wild type protein (Fig. 3-4). Moreover, while the free cofactor produces the trans cyclopropanes in the reactions of **1**, **2**, and **4** with EDA, a pair of mutants was identified that provides each enantiomer of the cis cyclopropane from each of the initial substrates (**1-4**) with up to 40 : 1 dr and moderate enantioselectivity. Specifically, the double mutant C317G, V254A (CYP119(+)) formed the (1S,2R)-enantiomer of the cis cyclopropane from styrene **1** with 73% ee, while the triple mutant C317G, L69F, T213V (CYP119(-)) gave the opposite (1R,2S)-enantiomer of the same product with -74% ee. The mutant CYP119(+) formed one enantiomer of the cis cyclopropanes from **2-4** with (+) 73-86% ee, while the mutant CYP119(-) gave the opposite enantiomer of each of the cis cyclopropanes. These results suggest the potential to create stereodivergent catalysts for enantio- and diastereoselective formation of cis-cyclopropanes from diverse alkenes from just two mutants (Fig. 3).

Alkene	CYP119 (+)		CYP119 (-)		Free Ir(Me)-PIX Cofactor	
	ee	dr (cis : trans)	ee	dr (cis : trans)	ee	dr (cis : trans)
Styrene	73%	2: 1	-74%	5 : 1	0%	1 : 4
α-Me Styrene	82%	11 : 1	-86%	3 : 1	0%	1 : 2
B-Me Styrene (cis)	75%	27 : 1	-18%	40 : 1	0%	10 : 1
B-Me Styrene (trans)	86%	13 : 1	32%	4 : 1	0%	1 : >25
Hexen-5-one	79%	4 : 1	-88%	8 : 1	0%	1 : 4
Cyclopentene-3-COOMe	---	228 : 1 : t : 1	---	2 : 1 : t : t	---	3 : 1 : 1 : 3
1-Octene	66%	4 : 1	-78%	6 : 1	0%	1 : 4
Cis-2-octene	39%	8 : 1	-60%	1 : 1	0%	5 : 1

CYP119 (+): Ir(Me)-PIX CYP119 with the mutations C317G, V254A
CYP119 (-): Ir(Me)-PIX CYP119 with the mutations C317G, L69F, T213V
t = trace

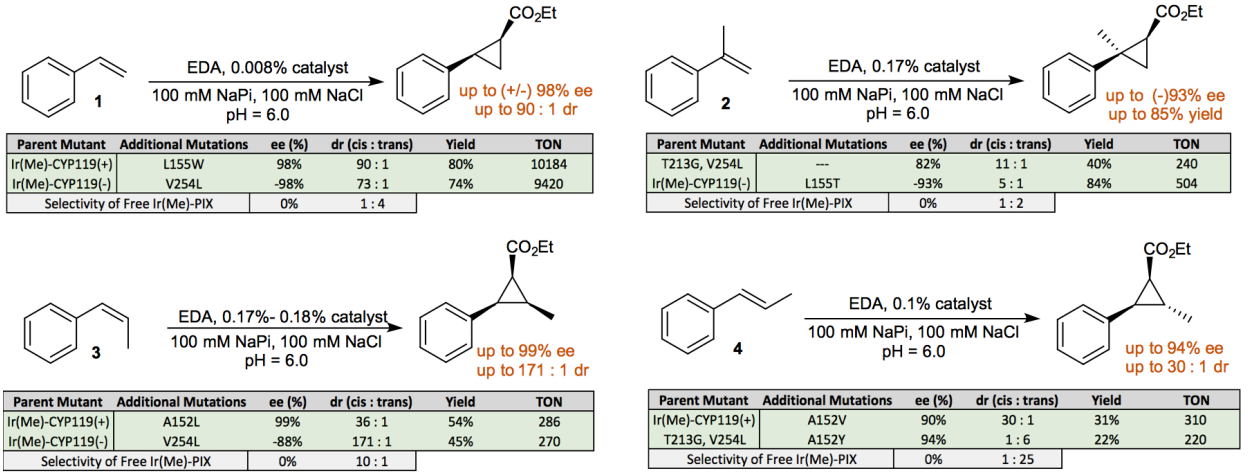
Figure 3. Cyclopropanation of alkenes of differing steric and electronic properties. The outcome of the reactions using two mutants of CYP119 identified in the initial phase of directed evolution are shown in comparison to the outcome of the same reactions catalyzed by the free Ir(Me)-PIX cofactor.

While the mutants CYP119(+) and CYP119(-) catalyzed enantioselective and diastereoselective cyclopropanation reactions of a range of substrates, we considered that additional rounds of evolution could create discrete mutants that react with higher selectivity for each individual substrate. To create mutants that were highly stereoselective for the cyclopropanation of **1-4**, we introduced additional mutations to CYP119(+) and CYP119(-) at positions A152, L155, F310, and L318 (Fig. 2). From these mutants, we identified distinct variants of Ir(Me)-CYP119 that function as catalysts for cyclopropanation of vinylarenes with enantioselectivity (up to 99% ee) and diastereoselectivity (up to 171 : 1) that are significantly higher than ?????. These are the highest levels of enantio- and diastereoselectivity ever achieved using artificial metalloenzymes catalysts (Fig 4).

More specifically, styrene **1** underwent cyclopropanation by EDA in high yield (up to 80%) and with high catalyst productivity (up to 10,000 TON). The mutant CYP119(+)-L155W formed the (1S,2R)-enantiomer of the cis cyclopropane in (+)98% ee, 90 : 1 dr, and 80 % yield, and the mutant CYP119(-)-V254L gave the opposite (1R,2S)-enantiomer with (-) 98% ee, 73 : 1 dr, and similar yield (75%). For this substrate, comparable levels of enantioselectivity have been achieved by Fe-containing P450s, but Fe-450s catalyze reactions forming the cis cyclopropane with ten times lower diastereoselectivity (9 : 1 cis : trans) than the diastereoselectivities that we achieve using variants of Ir(Me)-PIX CYP119 (90 : 1 cis : trans). In addition, while Fe-P450 enzymes have been identified to form one enantiomer of the cis cyclopropane from **1**, we have identified discrete mutants to form either enantiomer of the cis cyclopropane with 98% ee. In contrast, Ir(Me)-PIX CYP119 variants were identified that provide either enantiomers of the cis-cyclopropane product with high enantioselectivity.

Cis and trans β-Me styrenes (**3** and **4**), substrates that do not react with Fe-PIX proteins, also underwent highly selective cyclopropanation reactions in the presence of variants of Ir(Me)-CYP119(+) and (-). In the presence the mutant CYP119(+)-A152L, alkene **3** reacted with EDA to form the cis product in 45% yield, 99% ee, and 45 : 1 dr (cis : trans). The mutant CYP119(-) afforded the opposite enantiomer of the cyclopropane with 100 : 1 dr (cis : trans) and -73% ee (Fig. 4). In the presence of mutant CYP119(+)-A152V, substrate **4** reacted with EDA to form

the cis product in 55% yield, with 34 : 1 dr (cis : trans), and 92% ee (Fig. 4). These results are a substantial improvement over those we attained in our initial studies using artificial metalloenzymes created by combining Ir(Me)-PIX with apo myoglobin. In those cases, reactions between **3** and EDA gave only trace yields in the presence of various mutants of the Ir-myo catalyst, while reactions between **4** and EDA occurred with low TON and ee (up to 20 TON and 40% ee).



Parent Mutant	Additional Mutations	ee (%)	dr (cis : trans)	Yield	TON
Ir(Me)-CYP119(+)	L155W	98%	90 : 1	80%	10184
Ir(Me)-CYP119(-)	V254L	-98%	73 : 1	74%	9420
Selectivity of Free Ir(Me)-PIX		0%	1 : 4		

Parent Mutant	Additional Mutations	ee (%)	dr (cis : trans)	Yield	TON
T213G, V254L	---	82%	11 : 1	40%	240
Ir(Me)-CYP119(-)	L155T	-93%	5 : 1	84%	504
Selectivity of Free Ir(Me)-PIX		0%	1 : 2		

Parent Mutant	Additional Mutations	ee (%)	dr (cis : trans)	Yield	TON
Ir(Me)-CYP119(+)	A152L	99%	36 : 1	54%	286
Ir(Me)-CYP119(-)	V254L	-88%	171 : 1	45%	270
Selectivity of Free Ir(Me)-PIX		0%	10 : 1		

Parent Mutant	Additional Mutations	ee (%)	dr (cis : trans)	Yield	TON
Ir(Me)-CYP119(+)	A152V	90%	30 : 1	31%	310
T213G, V254L	A152Y	94%	1 : 6	22%	220
Selectivity of Free Ir(Me)-PIX		0%	1 : 25		

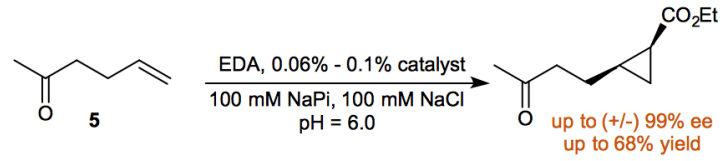
Figure 4. Addition of EDA to terminal and internal vinylarenes catalyzed by variants of Ir(Me)-PIX CYP119.

Although our studies focused on the preparation of cis-cyclopropanes, as this is generally the opposite diastereomer to that produced by the free Ir(Me)-PIX cofactor, we also found that different evolutionary trajectories can be followed to identify mutants to form trans cyclopropanes as the major diastereomer with high enantioselectivity (Fig. 4). For example, the variant of Ir(Me)-CYP119 containing the mutations C317G, L69V, T213G, V254L afforded the trans cyclopropane from the reaction of **4** with EDA in 52% yield, and with 96% ee and 6 : 1 dr (trans : cis). This demonstration of the enantio- and diastereo-divergent selectivity of Ir(Me)-CYP119 variants confirms the potential of the native active sites of artificially metallated P450s to control the outcome of the catalytic reactions in a comparable fashion to that observed for natural Fe-P450s reacting with their biological substrates.

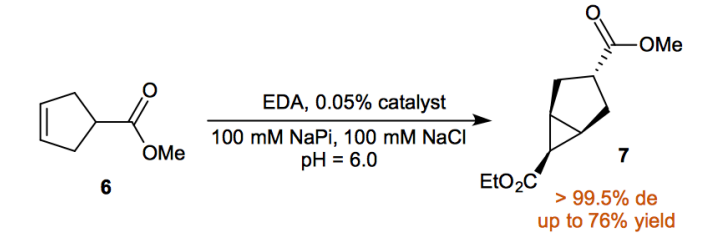
5.3 Cyclopropanation of unactivated and deactivated alkenes

Given these promising results with activated vinyl arenes, we considered the potential to functionalize less reactive aliphatic alkenes using Ir(Me)-CYP119 catalysts. Fe-PIX enzymes have not been reported to catalyze the cyclopropanation of unactivated alkenes previously. We selected the terminal aliphatic alkene hexen-5-one **5** and the internal alkene **6** as model substrates for these studies (Fig. 5) due to their polar functionality that will increase the solubility of the substrate in the aqueous reaction media. In contrast to Fe-containing P450s, we found that Ir(Me)-CYP119 variants catalyze the functionalization of both of these substrates with EDA. Moreover, the same mutants CYP119(+) and CYP119(-), which were identified to react in a stereodivergent fashion with vinyl arenes, were also enantio- and diastereoselective catalysts for the reactions with **5** and **6** (Fig 3) to form the opposite enantiomers of the products, allowing highly selective catalysts to be identified without extensive screening of additional mutants. In

the presence of CYP119(-)-V254L-L155T, the terminal and unactivated alkene **5** underwent cyclopropanation with EDA in 67% yield with 1006 TON, even when **5** was the limiting reagent, forming the cis-cyclopropane in (-)99% ee and 7 : 1 dr (cis : trans). The variant CYP119(+)-L155Y was equally enantioselective in forming the opposite enantiomer of the cis cyclopropane from **5** (42% yield, 650 TON, 99% ee, 25 : 1 dr). As was the case for the reactions of vinylarenes, the reaction of **5** catalyzed by the free cofactor produced the trans cyclopropane as the major product (Fig. 5). Other terminal, aliphatic alkenes were also transformed with high stereoselectivity; 1-octene was transformed with up to 90% ee, 15 : 1 dr, and 496 TON (Fig. 6).



Parent Mutant	Additional Mutations	ee (%)	dr (cis : trans)	Yield	TON
Ir(Me)-CYP119(+)	L155W	99%	19 : 1	44%	440
Ir(Me)-CYP119(-)	V254L, L155T	-99%	7 : 1	68%	1006
Selectivity of Free Ir(Me)-PIX		0%	1 : 4		



Parent Mutant	Additional Mutations	de (%)	dr (cis : trans)	Yield	TON
Ir(Me)-CYP119(-)	L155W	>99.5%	204 : 1 : t : t	76%	1300
Selectivity of Free Ir(Me)-PIX		0%	3 : 1: 1 : 3		

Figure 5. Addition of EDA to terminal and internal aliphatic alkenes catalyzed by variants of Ir(Me)-PIX CYP119.

In addition to terminal, aliphatic alkenes, both cyclic and acyclic, aliphatic internal alkenes underwent highly stereoselective cyclopropanation in the presence of the variants of Ir(Me)-PIX CYP119(+) and (-) (Fig 3) and their further variants (Fig 5). The reaction of cyclopentene **6** and EDA occurred in 76% yield with 1300 TON in the presence of CYP119(-)-F310W from a reaction conducted with alkene as the limiting reagent. Thus, Ir(Me)-PIX CYP119 variants can catalyze the cyclopropanation of unactivated and internal alkenes with nearly the same yields as those attained for the analogous reactions of styrene. The enzymatic reaction by this mutant produced **7** as nearly a single diastereomer (>99.5% diastereomeric excess). This selectivity contrasts sharply with that of the reaction in the presence of the free Ir(Me)-PIX cofactor, which formed a mixture of four diastereomers, none of which accounted for >45% of the product. The acyclic alkene cis-2-octene underwent cyclopropanation with EDA in up to -94% ee and 108 : 1 dr in the presence of Ir(Me)-CYP119(-)-V254L (Fig. 6). For this challenging class of substrate (acyclic, internal, aliphatic alkenes), no comparable stereoselectivities have been achieved using any small molecule or enzyme as catalyst.

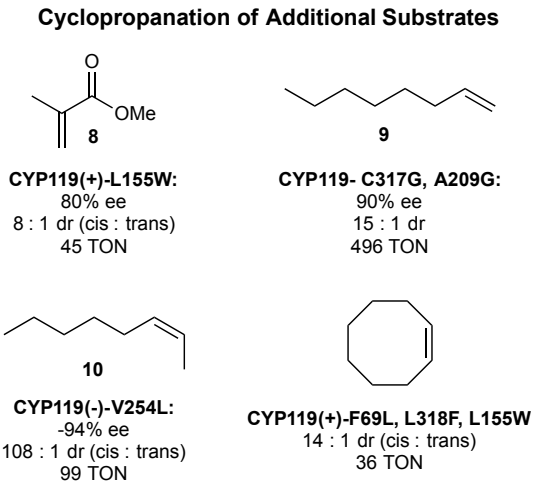


Figure 6. Stereoselective cyclopropanation of additional substrates with ethyl diazoacetates. The results were obtained from reactions containing 20 mM alkene, 0.01 mM catalyst, and 3 equiv. EDA added over 3 hours using a syringe pump.

Even substrates containing di-substituted, electron-deficient alkenes were found to undergo selective cyclopropanation in the presence of Ir(Me)-CYP119 enzymes. Although the yields and turnover numbers were modest, the variant CYP119(+)-155W catalyzed the reaction of methyl methacrylate **8** and EDA to form the cis product in 80% ee and 8 : 1 dr (cis : trans) with 47 TON. The variant CYP119(-)-L155V afford the opposite enantiomer with -88% ee and 145 TON.

5.4 Substrate Selective Cyclopropanation

Selective binding of a substrate within the active site of a natural enzyme allows reactions of less reactive substrates to occur in the presence more reactive substrates that are less suitable to bind to the active site. We evaluated whether the active site of Ir(Me)-CYP119 retains the potential for the substrate-selective catalysis of natural P450s. We evaluated a cyclopropanation reaction of EDA in the presence of an equal molar mixture of 1-octene **9** and cis-2-octene **10**, which are two substrates having the same size but different positions of the double bond. In the presence of EDA and equimolar **9** and **10**, the free co-factor Ir(Me)-PIX reacted with high substrate selectivity, giving a 12 : 1 ratio of products **11** and **12**, respectively (Fig. 7). In sharp contrast, the reaction of EDA with a mixture of **9** and **10** catalyzed by Ir(Me)-CYP119(-)-V254L occurred with the opposite substrate selectively; stereoisomers of **12** accounted for 63% of the overall product (Fig 7). This result is the first example of substrate-selective catalysis achieved with artificial metalloenzymes, and this example is particularly remarkable in that the enzyme can react selectively with the substrate that is much less reactive in the presence of the free cofactor alone.

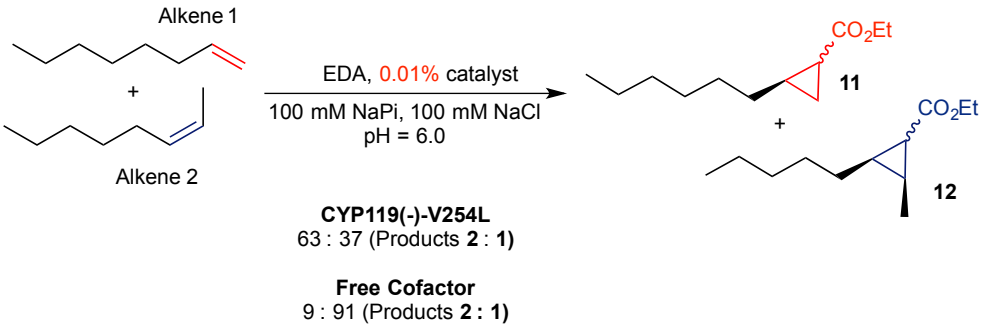


Figure 7. Substrate-selective cyclopropanation of a mixture of octene regioisomers by a variant of Ir(Me)-PIX CYP119 enzymes.

5.5 Conclusions and Future Directions

The results of the study just described demonstrate that artificially metallated P450 enzymes retain the properties of the active sites of native enzymes. The results obtained for substrates **1-10**, that is substrates of different size and shape, with different substitution patterns, and with varied electronic properties of the double bond, underscore the combined potential of the active site of CYP119 to be accommodate various substrates and the Ir(Me)-center to promote substantially broader reactivity than is possible with native, Fe-PIX enzymes. Such artificial metalloenzymes can perform catalytic transformations not only with exquisite stereoselectivity, but with access to other modes of selectivity of natural enzymes previously unrealized by artificial metalloenzymes. This approach creates avenues to truly combine the abiological reactivity of transition metals with exquisite selectivity of enzymes.

5.7 Experimental Information

5.7.1 Protein Expression, Purification, and Characterization

a. General Methods

Unless otherwise noted, the chemicals, salts, and solvents used were reagent grade and used as received from commercial suppliers without further purification. All expression media and buffers were prepared using ddH_2O (MilliQ A10 Advantage purification system, Millipore). All expression media were sterilized using either an autoclave (45 min, 121^oC) or a sterile syringe filter (0.22 um). To maintain sterile conditions, sterile materials and *E. coli* cells were manipulated near a lit Bunsen burner.

b. Genes and Cloning

The WT CYP119 gene cloned into the vector 2BT (6xHis-TEV-ORF; AddGene #29666) was purchased from GenScript with codon optimization for *E. coli* (Table S1).

c. Media Preparation

Preparation of optimized minimal expression media: Salts (15 g Na_2HPO_4, 7.5 g K_2HPO_4, 0.3 g NaH_2PO_4, 0.3 g KH_2PO_4, 1.5 g NaCl, 5 g NH_4Cl) were dissolved in 2 L ddH_2O and autoclaved to give a media with pH ~8.0 - 8.2. Solutions of glucose (20%), casamino acids (BD Company, low Fe, 20%), and $MgSO_4$ (1 M), were autoclaved separately. Solutions of ampicillin (100 mg/ mL) and $CaCl_2$ (1 M) were sterilized by syringe filter. The following amounts of the listed solutions were added per 2 L of sterile salt solution: 40 mL glucose, 20 mL casamino acids, 4 mL $MgSO_4$, 100 uL $CaCl_2$, 2 mL ampicillin. Stock solutions were stored for several weeks; prepared media was stored for less than 1 day. Minimal media plates were prepared from the same media with the addition of 17 g agar/L media. In this case, agar was autoclaved in 1 L ddH_2O, and salts were autoclaved separately as a 20X solution, after which they were added to the agar solution.

d. Mutagenesis

Site-directed mutagenesis was performed using the QuickChange Lightening mutagenesis kit (Agilent); requisite double stranded DNA primers were designed according to the Agilent Primer Design Program and purchased from Integrated DNA Technology. PCR reactions were performed according to the manufacturer's directions. PCR reactions contained: 5 uL reaction buffer, 34 uL ddH_2O, 1.5 uL QuickSolution, 1 uL plamids (50 ng/uL), 1.25 uL sense primer (100 ng/uL), 1.25 uL antisense primer (100 ng/ uL), 5 uL dNTPs (2 mM/base), and 1 uL polymerase.

PCR Program: Phase 1 (1 cycle): 95 °C, 1.5 min; Phase 2 (18 cycles): 95 °C, 20 sec, 60 °C, 10 sec, 68 °C, 4.5 min; Phase 3 (1 cycle): 68 °C, 3 min; Phase 4 (storage): 4 °C

DNA Isolation and Storage: Following the completion of the above set of PCR procedures, 1.5 uL DPN 1 was added to each reaction, and the reactions were further incubated (3 h, 37°C). The crude PCR mixture was used to transform XL-10 Gold Ultracompetent cells (45 L cells, 2 L

PCR reactions). The mixture was incubated on ice (30 min), heat shocked (30 s, 42 °C), recovered with SOC media (1 h, 37 °C, 275 rpm), and plated on LB plates. Plates were grown (18 h, 37 °C), and individual colonies were used to inoculate 1 mL rich media cultures, which were grown in 96-well plates (13 h, 37 °C, 300 rpm). DNA was isolated from the 96-well cultures using magnetic bead technology at the UC Berkeley DNA Sequencing Facility. Alternatively, individual colonies were used to inoculate 4 mL rich media cultures and grown overnight (13 h, 37 °C, 300 rpm), and the plasmids were purified using a Qiagen DNA Miniprep kit according to the manufacturer's instructions.

e. Protein Expression

Optimized Expression of Apo CYP119 for production of purified enzyme: BL21 Star chemically competent *E.* coli cells (45 L, QB3 Macrolab, UC Berkeley) were thawed on ice, transferred to 14 mL falcon tubes , and transformed with the desired plasmid solution (1.5 L, 50-250 ng/ L). The cells were incubated on ice (30 min), heat shocked (30 sec, 42°C), re-cooled on ice (2 min), and recovered by incubation with SOC media (500 uL, 37 °C, 1 h, 250 rpm). Aliquots of the cultures (50uL) were plated on plates containing minimal media agar (expression media supplemented with 17 g agar/L) and incubated (20 h, 37 °C) to produce approximately 10-100 colonies per plate. Single colonies were used to inoculate starter cultures (3 mL, expression media), which were grown in 14 mL round bottom tubes (4-8 hours, 37°C, 250 rpm) and used to inoculate 100 mL overnight cultures (minimal media, 37° C, 250 rpm). Each overnight culture was used to inoculate 750 mL of expression media, which was cultured further (9 h, 37°C, 275 rpm). Expression was induced with IPTG (800 uL, 1M), and expression was allowed to continue for 15 h (30° C, 250 rpm). Cells were harvested by centrifugation (5000 rpm, 15 min, 4° C), and the pellets were resuspended in 20 mL Ni-NTA lysis buffer (50 mM NaPi, 250 mM NaCl, 10 mM Imidazole, pH = 8.0) and stored at -80° C until purification.

Protein Purification: Cell suspensions were thawed in a room temperature ice bath, decanted to 50 mL glass beakers, and lysed on ice by sonication (4x30 sec on, 2x2 min off, 65% power). Cell debris was removed by centrifugation (10 000 rpm, 30 min, 4° C), and Ni-NTA (5 mL, 50% suspension per 850 mL cell culture) was added. The lysates were briefly incubated with Ni-NTA on an end-over-end mixer (30 min, rt, 20 rpm) and poured into glass frits (coarse, 50 mL). The resin was washed with Ni-NTA lysis buffer (3 x 35 mL), and the wash fractions were monitored using Bradford assay dye. The desired protein was eluded with 18 mL Ni-NTA elusion buffer (50 mM NaPi, 250 mM NaCl, 250 mM Imidazole, pH = 8.0), dialyzed against Tris buffer (10 mM, pH = 8.0, 12 h, 4° C), concentrated to the desired concentration using a spin concentrator, and metallated within several hours. Apo protein was not stored for more than 8 hours.

f. Protein Characterization

Gel Electrophoresis: Protein purity was analyzed by sodium dodecyl sulfate-polyacrylamide (SDS-PAGE) gel electrophoresis using precast gels (polyacrylamide, 10-20% linear gradient, Biorad).

Mass Spectrometry: Apo-proteins were analyzed with an Agilent 1200 series liquid chromatograph connected in-line with an Agilent 6224 time-of-flight (TOF) LC/MS system using a Turbospray ion source.

5.7.2 Organic Synthesis and Characterization

a. General methods and materials

Unless stated otherwise, all reactions and manipulations were conducted on the laboratory bench in air with reagent grade solvents. Reactions under inert gas atmosphere were carried out in the oven dried glassware in a nitrogen-filled glovebox or by standard Schlenk techniques under nitrogen.

NMR spectra were acquired on 400 MHz, 500 MHz, 600 MHz, or 900 MHz Bruker instruments at the University of California, Berkeley. NMR spectra were processed with MestReNova 9.0 (Mestrelab Research SL). Chemical shifts are reported in ppm and referenced to residual solvent peaks[10]. Coupling constants are reported in hertz. GC analyses were obtained on an Agilent 6890 GC equipped with either, an HP-5 column (25 m x 0.20 mm ID x 0.33 m film) for achiral analysis, Cyclosil-B column (30 m x 0.25 mm ID x 0.25 um film) or CP-Chirasil-Dex CB column (25 m x 0.25 mm ID x 0.25 um film) for chiral analysis, and an FID detector. GC yields were calculated using dodecane as the internal standard and not corrected for response factors of minor isomers. High-resolution mass spectra were obtained via the Micro-Mass/Analytical Facility operated by the College of Chemistry, University of California, Berkeley.

Unless noted otherwise, all reagents and solvents were purchased from commercial suppliers and used without further purification. If required, dichloromethane (DCM) and tetrahydrofuran (THF) were degassed by purging with argon for 15 minutes and dried with a solvent purification system containing a one-meter column of activated alumina; dried and degassed acetonitrile, 1,2-xylene, toluene, N,N-dimethylformamide (DMF), ethanol and methanol were purchased form commercial suppliers and used as received.

b. Substrates

All substrates were purchased from commercial suppliers and used as received.

c. Authentic Products

General procedures for synthesis of cyclopropanes:

Procedure A:

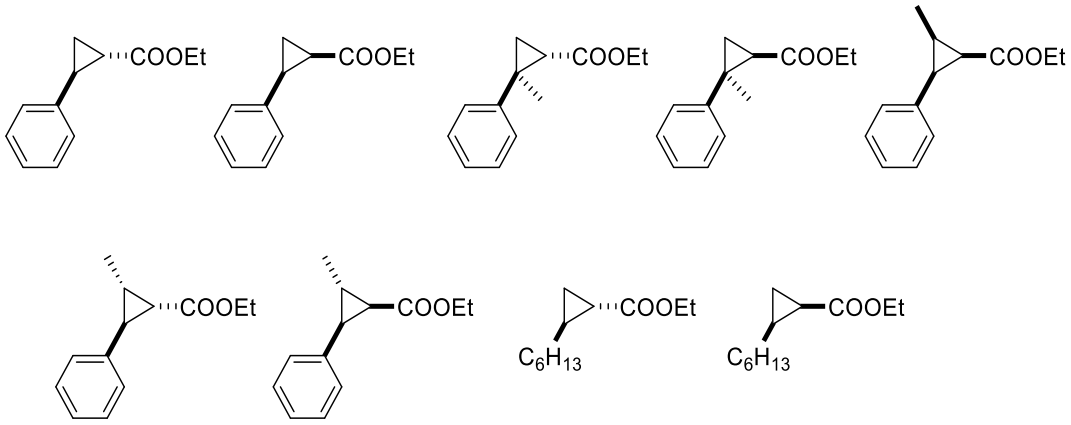
To a vial charged with the alkene (5-10 mmol, 5-10 equiv.) in toluene (10 ml), was added 100 ul of an 8 mM solution of Ir(Me)-PIX (0.0008 equiv.) in DMF. A solution of ethyl diazoacetate (1 mmol, 1 equiv.) in toluene (1 ml) was then added slowly while the reaction mixture was vigorously stirred. After complete addition of EDA, the reaction was stirred for 30-60 min, after which time the evolution of nitrogen stopped, indicating full consumption of EDA. Then, the volatile materials were removed, and the residue was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 85:15 gradient) as the eluent. Fractions of the pure product(s) were combined, and the solvent evaporated, yielding cyclopropanation products.

Procedure B:

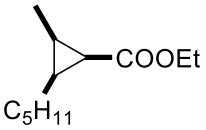
To a solution of alkene (~0.2 M) and Rh_2AcO_4 (~0.1-1 mol% in respect to EDA) in dry DCM, a solution of ethyl diazoacetate (~1 M) in dry DCM was added slowly while the reaction mixture was vigorously stirred. After complete addition of EDA, the reaction was stirred for 30-60 min, after which time the evolution of nitrogen stopped, indicating full consumption of EDA. Then, the volatile materials were removed, and the residue was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 85:15 gradient) as the eluent. Fractions of the pure product(s) were combined, and the solvent evaporated, yielding cyclopropanation products.

The NMR data of isolated products matched those for reported molecules: Cis- and trans- ethyl 2-phenylcyclopropane-1-carboxylates (**1**+eda, prepared following Procedure A for the trans-isomer and Procedure B for cis-isomer, both with styrene as alkene)[11], cis- and trans- ethyl 2-methyl-2-phenylcyclopropane-1-carboxylates (**2**+eda, both prepared following Procedure A, with alpha-methylstyrene as alkene)[12], ethyl syn-2-methyl-syn-3-phenylcyclopropane-1-carboxylate (**3**, prepared following Procedure A, with cis-betamethylstyrene as alkene),[11] ethyl syn-2-methyl-anti-3-phenylcyclopropane-1-carboxylate (**4**, prepared following Procedure B, with trans-betamethylstyrene as alkene),[13,14] ethyl syn-2-methyl-syn-3-phenylcyclopropane-1-carboxylate (**4**+eda, prepared following Procedure B, with trans-betamethylstyrene as alkene),[13] and cis- and trans-ethyl 2-hexylcyclopropane-1-carboxylates (**8** +EDA prepared following Procedure B for the cis-isomer and Procedure A for trans-isomer, with 1-octene as alkene).

The absolute configurations of cis- and trans- ethyl 2-phenylcyclopropane-1-carboxylates,[9] and cis- and trans-ethyl 2-hexylcyclopropane-1-carboxylate[15] were assigned on the basis of reported retention times,7,8 conducting analysis according to literature methods (Table 15).



Ethyl syn-2-methyl-syn-3-pentylcyclopropane-1-carboxylate



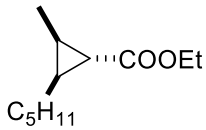
The product was isolated from a reaction of cis-2-octene (1 ml) with EDA (0.25 ml) conducted following Procedure B.

^{1}H NMR (900 MHz, $CDCl_3$): d = 4.09–4.05 (m, 2H), 1.60–1.57 (m, 3H), 1.38–1.36 (dq, J =6.5 Hz, J = 2.2 Hz, 1H), 1.31-1.22 (m, 11H), 1.18 (d, J = 6.2 Hz, 3H), 0.85 (m, J = 6.7 Hz, 3H).

^{13}C NMR (151 MHz, $CDCl_3$): d = 172.34, 59.76, 31.84, 29.57, 25.36, 22.82, 22.10, 20.76, 19.04, 14.55, 14.20, 7.53;

HR MS (EI): calcd. for $C_{12}H_{21}O_2$ $[M]^+$: 198.1620, found: 198.1622.

Ethyl anti-2-methyl-anti-3-pentylcyclopropane-1-carboxylate



The product was isolated from a reaction of cis-2-octene (1 ml) with EDA (0.25 ml) conducted following Procedure B.

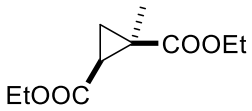
^{1}H NMR (900 MHz, $CDCl_3$): d = 4.08 (q, J= 7.2 Hz, 2H), 1.46 (m, 1H), 1.39-1.30 (m, 5H), 1.28-1.24 (m, 4H), 1.23 (t, J = 7.2 Hz, 3H), 1.08 (d, J = 6.3 Hz, 3H), 0.99 (t, J = 4.8 Hz, 1H), 0.87 (t, J = 7.4 Hz, 3H).

^{13}C NMR (151 MHz, $CDCl_3$): d = 174.85, 60.33, 31.77, 29.31, 28.15, 27.99, 27.23, 22.76, 21.92, 14.47, 14.19, 12.16;

HR MS (EI): calcd. for $C_{12}H_{21}O_2$ $[M]^+$: 198.1620, found: 198.1619.

Syn-2-ethyl 1-ethyl 1-methylcyclopropane-1,2-dicarboxylate

Stereochemistry assigned based on comparison of NMR data to analogous compound [16]



The product was isolated from a reaction of ethyl methacrylate (3 ml) with EDA (1 ml) conducted following Procedure B. The isolated product contains 5% impurity by GC.

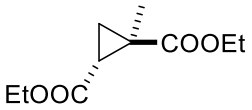
^{1}H NMR (500 MHz, $CDCl_3$): d = 4.10–4.20 (m, 4H), 1.78–1.84 (m, 2H), 1.41 (s, 3H), 1.24-1.28 (m, 6H), 1.04-1.10 (m, 1H);

^{13}C NMR (151 MHz, $CDCl_3$): d = 171.73, 170.55, 61.12, 60.93, 29.06, 28.97, 21.33, 19.64, 14.33, 14.24;

HR MS (EI): calcd. for $C_{10}H_{16}O_4$ $[M]^+$: 200.1049, found: 200.1051.

Anti-2-ethyl 1-ethyl 1-methylcyclopropane-1,2-dicarboxylate[16]

Relative stereochemistry was assigned based on comparison with the NMR data for the methyl ester analogue: anti-2-ethyl 1-methyl 1-methylcyclopropane-1,2-dicarboxylate



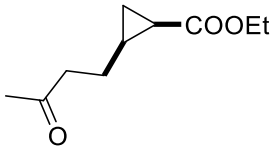
The product was isolated from a reaction of ethyl methacrylate (3 ml) with EDA (1 ml) conducted following Procedure B. The isolated product contains 5% impurity by GC.

^{1}H NMR (600 MHz, $CDCl_3$): d = 4.07-4.16 (m, 4H), 2.29 (dd, J = 8.7, 6.5 Hz, 1H), 1.53 (dd, J = 8.7, 4.2 Hz, 1H), 1.36 (s, 3H), 1.31 – 1.26 (m, 1H), 1.23 (dt, J = 10.7, 7.1 Hz, 6H);

^{13}C NMR (151 MHz, $CDCl_3$): d = 173.64, 170.63, 61.37, 61.04, 27.98, 27.10, 21.12, 14.45, 14.28, 14.26, 13.21;

HR MS (EI): calcd. for $C_{10}H_{16}O_4$ $[M]^+$: 200.1049, found: 200.1051.

Cis-ethyl 2-(3-oxobutyl)cyclopropane-1-carboxylate



The product was isolated from a reaction of hex-5-en-2-one (2 ml) with EDA (1 ml) conducted following Procedure B.

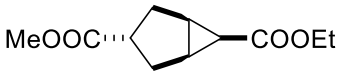
^{1}H NMR (600 MHz, $CDCl_3$): d = 4.13 – 4.04 (m, 2H), 2.41 (td, J = 7.3, 4.9 Hz, 2H), 2.08 (s, 2H), 1.83 (ddt, J = 14.4, 7.2, 7.2 Hz, 1H), 1.72 (ddt, J = 14.6, 7.4, 7.4 Hz, 1H), 1.63 (ddd, J = 8.3, 5.5 Hz, 1H), 1.17-1.26 (m, 3+1H), 0.96 (ddd, J = 8.2, 8.2, 4.5 Hz, 1H), 0.88 (ddd, J = 6.9, 5.0, 5.0 Hz, 1H);

^{13}C NMR (151 MHz, $CDCl_3$): d = 208.44, 172.89, 60.46, 43.50, 29.92, 21.67, 21.65, 21.09, 21.08, 18.37, 14.46, 13.52, 13.51;

HR MS (EI): calcd. for $C_{10}H_{16}O_3$ $[M]^+$: 184.1099, found: 184.1199.

6-ethyl 3-methyl (1R,3s,5S,6s)-bicyclo[3.1.0]hexane-3,6-dicarboxylate

Relative stereochemistry was assigned based on comparison with the NMR data for the analogue: methyl syn-3-carbomethoxybicyclo[3.1 .O]hexane-6-acetate[17]



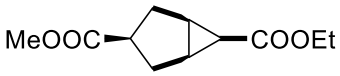
The product was isolated from a reaction of methyl cyclopent-3-ene-1-carboxylate (0.3 ml) with EDA (0.5 ml) conducted following Procedure B.

^{1}H NMR (900 MHz, $CDCl_3$): d = 4.11 (q, J = 7.1 Hz, 2H), 3.63 (s, 3H), 2.93 (p, J = 8.7 Hz, 1H), 2.22 (m, 4H), 1.82 (m, 2H), 1.67 (t, J = 8.3 Hz, 1H), 1.25 (t, J = 7.4 Hz, 3H).

^{13}C NMR (226 MHz, $CDCl_3$): d = 176.50, 171.61, 61.24, 51.75, 43.08, 29.80, 25.71, 23.60, 14.26.

HR MS (EI): calcd. for $C_{11}H_{16}O_4$ $[M]^+$: 212.1049, found: 212.1048.

6-ethyl 3-methyl (1R,3r,5S,6s)-bicyclo[3.1.0]hexane-3,6-dicarboxylate



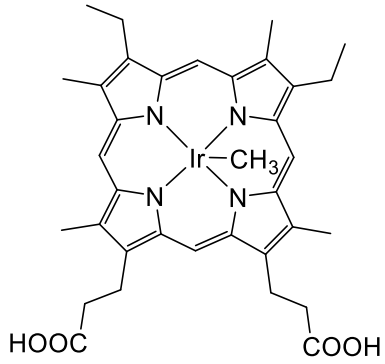
The product was isolated from a reaction of methyl cyclopent-3-ene-1-carboxylate (0.3 ml) with EDA (0.5 ml) conducted following Procedure B.

^{1}H NMR (900 MHz, $CDCl_3$): d = 4.09 (q, J= 7.1 Hz, 2H), 3.65 (s, 3H), 3.34 (p, J = 9.5 Hz, 1H), 2.26 (m, 4H), 1.88 (m, 3H), 1.66 (t, J = 8.5 Hz, 1H), 1.22 (t, J= 7.2 Hz, 3H);

^{13}C NMR (226 MHz, $CDCl_3$): d = 175.18, 171.25, 59.90, 51.87, 49.45, 27.67, 26.29, 14.33.

HR MS (EI): calcd. for $C_{11}H_{16}O_4$ $[M]^+$: 212.1049, found: 212.1050.

d. Ir(Me)-Mesoporphyrin IX (Ir(Me)-PIX) cofactor used in the studies:



The synthetic procedures and characterization of Ir(Me)-PIX were reported previously. [18]

5.7.3 Catalytic Experiments

a. General Methods

Unless otherwise noted, catalytic reactions were performed in 4 mL individually-capped vials or in 1.2 mL glass vials arranged in a 96-well metal block covered by lid attached by screws. Reactions on 96 well plates were assembled in a nitrogen atmosphere glove box. Reactions in individual vials were conducted under air after determining that comparable results are obtained under atmospheres of either nitrogen or air. Solutions of Ir(Me)-CYP119 variants were gently degassed on a Schlenk line (3 cycles vacuum/refill) before being brought into a glove box in sealed vials. Organic reagents were added as stock solutions in DMF, such that the final amount of DMF in the reaction was approximately 2% by volume (unless noted otherwise). If necessary, purified, degassed protein catalysts were diluted to the suitable reaction concentration using sodium phosphate buffer (100 mM, pH = 6.0, containing 100 mM NaCl) before being added to reaction vials. Unless otherwise noted, all reactions with purified enzymes were performed with catalysts generated from a 1:2 ratio of Ir(Me)-cofactor : apo protein, with 0.17 mol% catalyst loading with respect to metal-cofactor and limiting reagent. Unless otherwise noted, all reactions were conducted in a shaking incubator (20 °C, 16 h, 275 rpm).

b. Procedure for catalytic experiments conducted using a 96 well plate format.

The catalyst stock solution: To apo-CYP119 (0.14 mM protein in 10 mM tris buffer) was added Ir(Me)PIX as a 3.5 mM stock solution in DMF. The apo protein and Ir(Me)-PIX cofactor were combined in a 50 : 1 (v:v) ratio, producing a solution with a 2 : 1 mol ratio of protein : [Ir]. The buffer of the protein solution was subsequently exchanged to 100 mM NaPi, 100 mM NaCl, pH

= 6.0 using a NAP column. The protein solution was degassed and introduced to a nitrogen atmosphere glove box as described in the general methods.

Reaction Set Up: The appropriate catalyst stock solution (250 uL, 0.01 mM) was added to each well of 96 well array of glass vials. To each catalyst solution, the appropriate alkene was added (~3-5 uL, depending on the alkene) to give a final alkene concentration of 100 mM. After the addition of the alkene, ethyl diazoacetate (EDA) was added as a stock solution in DMF (5 uL of a 500 mM solution of EDA in DMF) to give a final EDA concentration of 10 mM in the reaction. The plate was sealed and removed from the glovebox, and the reactions were incubated overnight at rt in an orbital shaker (150 rpm).

c. Procedure for catalytic experiments conducted in 4 mL vials using a syringe pump.

The catalyst stock solution: The catalyst solution was prepared as described in section IIIb, except that the protein solution was not degassed and the reaction was not set up inside a glove box.

Reaction Set Up: Under air, 1 mL of catalyst stock solution (concentrations as listed in the tabulated results) was added to a 4 mL glass vial equipped with a micro stirbar. To the reaction vial, 10-20 umol of alkene were added as a stock solution in 10 uL DMF. The reaction vial was sealed using a cap containing a septa. The syringes of a syringe pump were loaded with a stock solution of EDA in DMF (20 uL of 0.5-5 M solution of EDA in DMF), and the EDA was added to the reactions over 1-6 hours, as specified in the tabulated results. The reaction was quenched upon the conclusion of the addition of EDA as described in section d.

d. Analysis of Yield and Enantiomeric Ratio (er)

Yields were determined by achiral or chiral GC using dodecane as an internal standard. Enantiomeric ratios were determined either by chiral GC or chiral SFC. Achiral GC was used to determine the yields for reactions analyzed by SFC, while the yields for reactions analyzed by chiral GC were determined concurrently with that analysis. The methods used to determine the er of each product are summarized in a subsequent section. Samples for analysis were prepared as follows, depending on the analysis method:

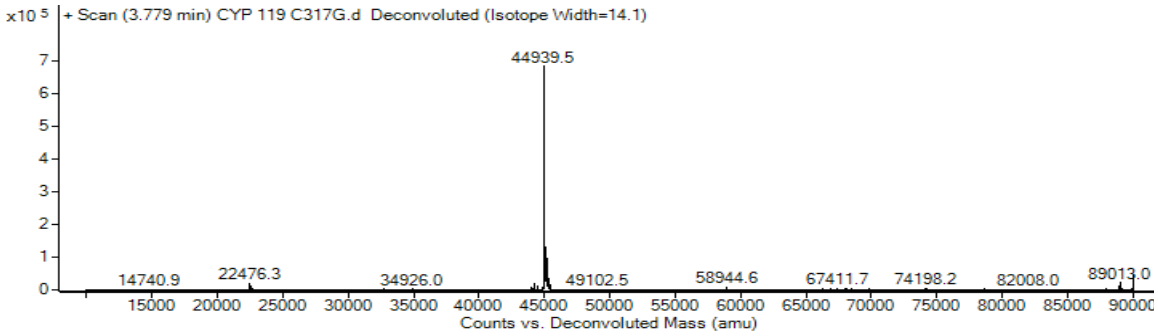
Analysis of Reaction from 96 well plates by chiral GC: The reaction was quenched by the addition of HBr (60 uL, 50% in water) to each reaction vial, followed with the addition of saturated NaCl (200 uL) and a solution of dodecane (500 ul, 1 ul/ml) in EtOAc. The contents of the vial were mixed by pipet, and the phases were allowed to separate. After separation of the layers, approximately 400 uL of the aqueous phase was removed from the bottom of the vial by pipet. The remaining contents of the vial were neutralized by the addition of sat. $NaHCO_3$ (200 uL), and the organic layer was further diluted with EtOAc (500 uL). The organic layer was then transferred to a separate vial for GC analysis. In the case of experiments performed in a 96-well array, all manipulations were performed using a multichannel pipet.

Analysis of Reaction from 4 mL vials by chiral GC: Saturated NaCl (500 uL) was added to each reaction vial, followed by a solution of dodecane (500 ul, 1 ul/ml) in EtOAc. The reaction was quenched by the addition of HBr (100 uL, 50% in water). The reaction mixture was mixed by pipet, and the layer were separated by centrifugation of the reaction vial (2 min, 2500 rpm). The aqueous phase was removed from the bottom of the vial by pipet, after which the remaining contents of the vial were neutralized by the addition of sat. $NaHCO_3$ (500 uL). The organic layer was then transferred to a separate vial for GC analysis.

5.7.4 Supporting Figures

Figure S1. Characterization of Ir(Me)-PIX CYP119 enzymes by mass spectrometry

Deconvoluted mass spectrum of CYP119-C317G acquired by LC-MS



Native Nanospray Ionization Mass Spectrum of Ir(Me)-CYP119 (12+ to 14+) charge states shown. Initial shoulder of each charge state corresponds to the apo-protein, while the most intense signal corresponds to the metallated protein (addition of Ir(Me)-PIX cofactor).

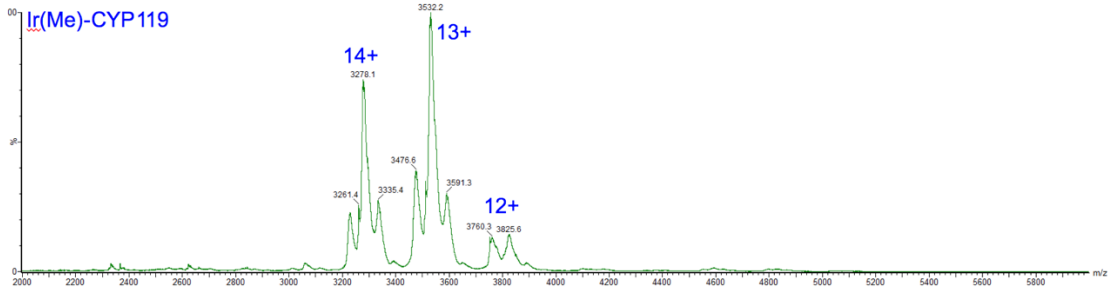
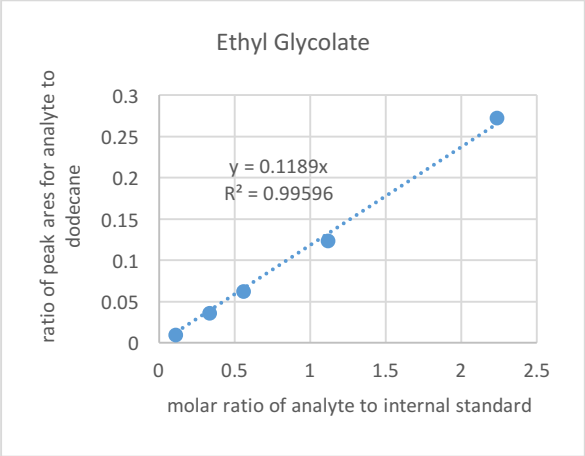
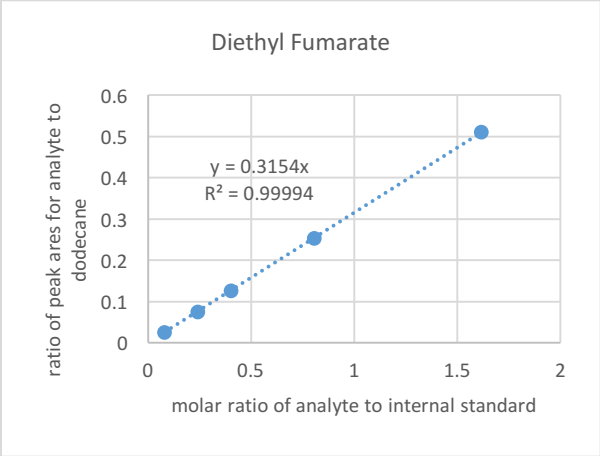
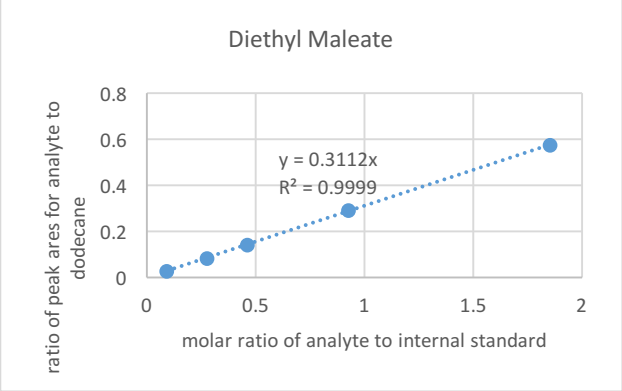
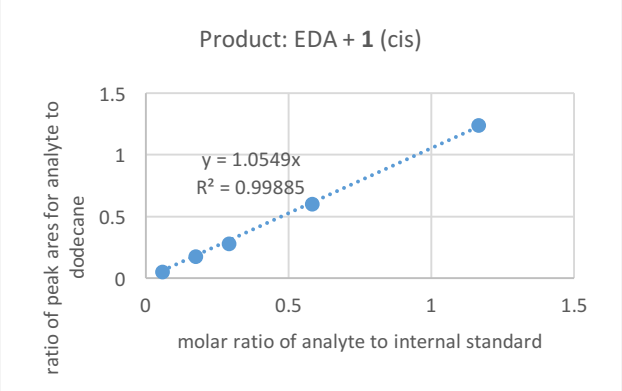


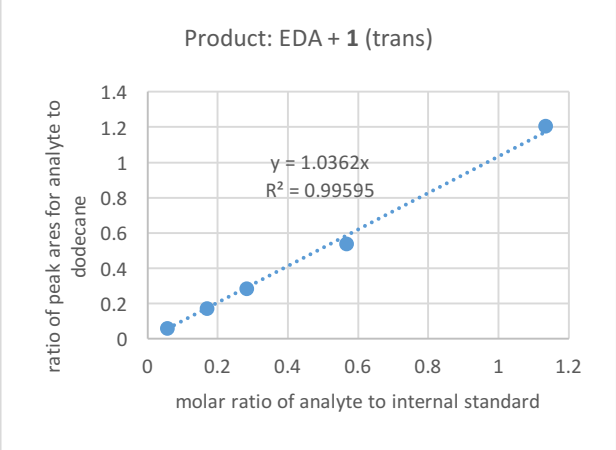
Figure S2. Calibration curves for products of cyclopropanation reactions.

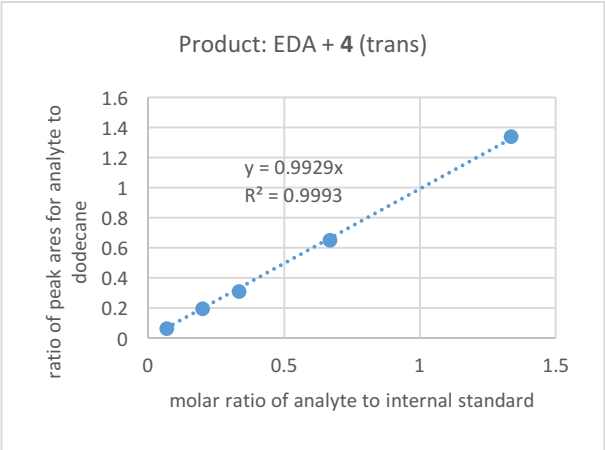

Ethyl Glycolate
ratio of peak ares for analyte to dodecane
molar ratio of analyte to internal standard
y = 0.1189x
R² = 0.99596
0
0.05
0.1
0.15
0.2
0.25
0.3
0
0.5
1
1.5
2
2.5

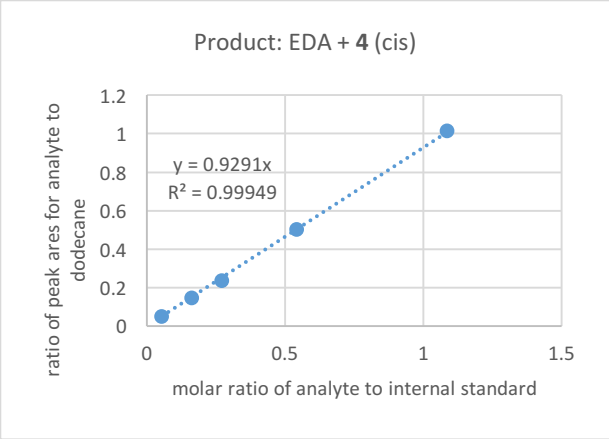

Diethyl Fumarate
ratio of peak ares for analyte to dodecane
molar ratio of analyte to internal standard
y = 0.3154x
R² = 0.99994
0
0.1
0.2
0.3
0.4
0.5
0.6
0
0.5
1
1.5
2

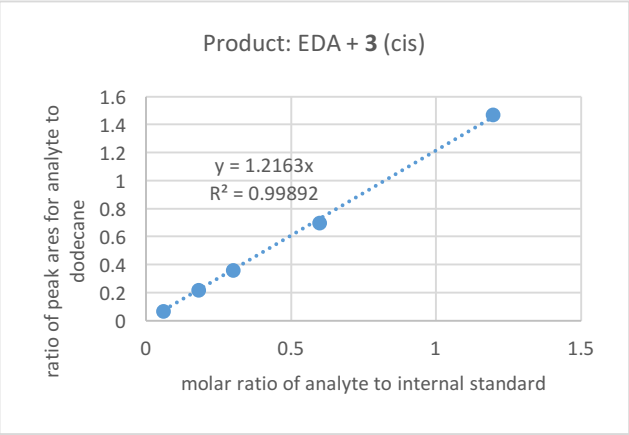

Diethyl Maleate
ratio of peak ares for analyte to dodecane
molar ratio of analyte to internal standard
y = 0.3112x
R² = 0.9999
0
0.2
0.4
0.6
0.8
0
0.5
1
1.5
2

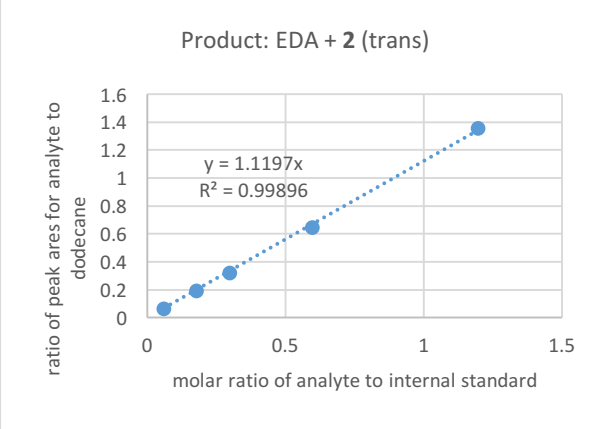

Product: EDA + 1 (cis)
ratio of peak ares for analyte to dodecane
molar ratio of analyte to internal standard
y = 1.0549x
R² = 0.99885
0
0.5
1
1.5
0
0.5
1
1.5

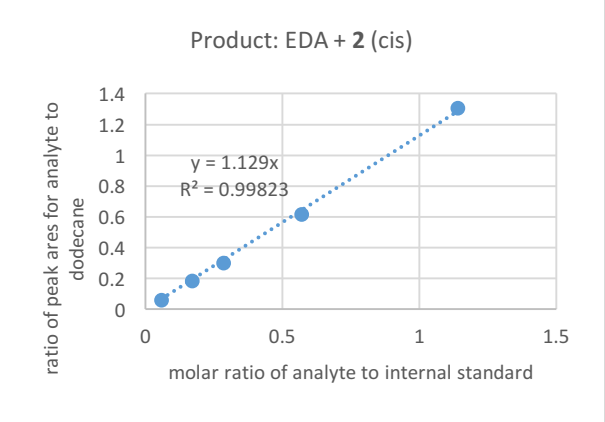

Product: EDA + 1 (trans)
ratio of peak ares for analyte to dodecane
molar ratio of analyte to internal standard
y = 1.0362x
R² = 0.99595
0
0.2
0.4
0.6
0.8
1
1.2
1.4
0
0.2
0.4
0.6
0.8
1
1.2

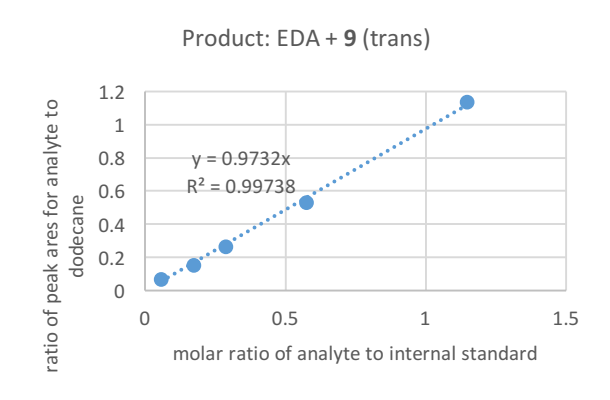

Product: EDA + 4 (trans)
ratio of peak ares for analyte to dodecane
molar ratio of analyte to internal standard
y = 0.9929x
R² = 0.9993
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
0
0.5
1
1.5

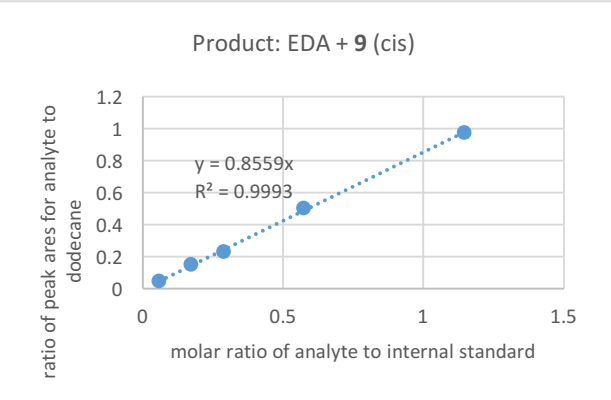

Product: EDA + 4 (cis)
ratio of peak ares for analyte to dodecane
molar ratio of analyte to internal standard
y = 0.9291x
R² = 0.99949
0
0.2
0.4
0.6
0.8
1
1.2
0
0.5
1
1.5

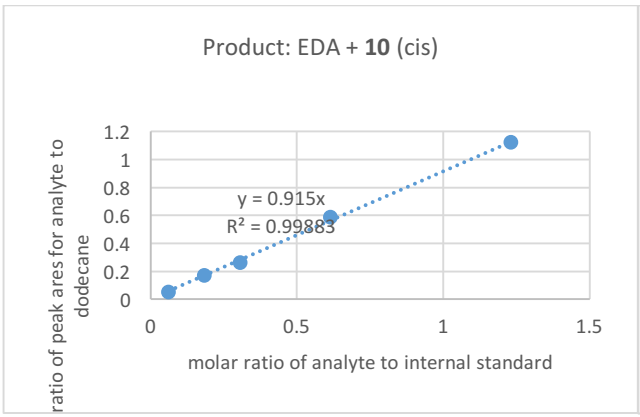

Product: EDA + 3 (cis)
ratio of peak ares for analyte to dodecane
molar ratio of analyte to internal standard
y = 1.2163x
R² = 0.99892
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
0
0.5
1
1.5

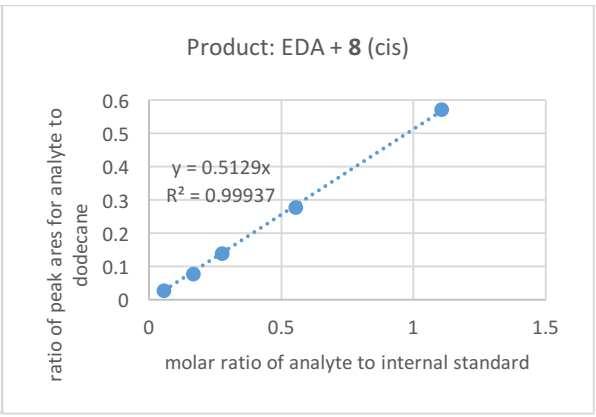

Product: EDA + 2 (trans)
ratio of peak ares for analyte to dodecane
molar ratio of analyte to internal standard
y = 1.1197x
R² = 0.99896
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
0
0.5
1
1.5

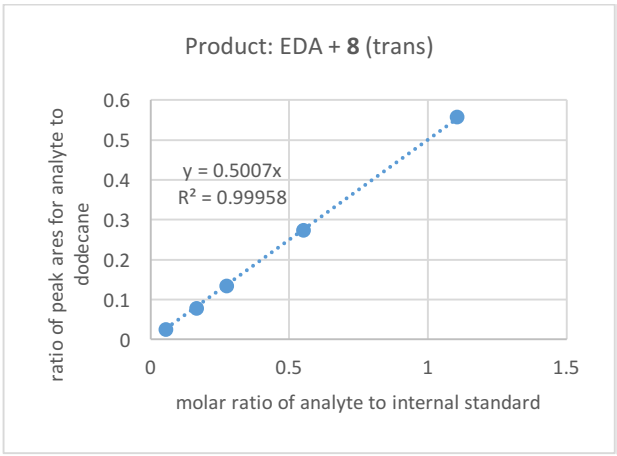

Product: EDA + 2 (cis)
ratio of peak ares for analyte to dodecane
molar ratio of analyte to internal standard
y = 1.129x
R² = 0.99823
0
0.2
0.4
0.6
0.8
1
1.2
1.4
0
0.5
1
1.5

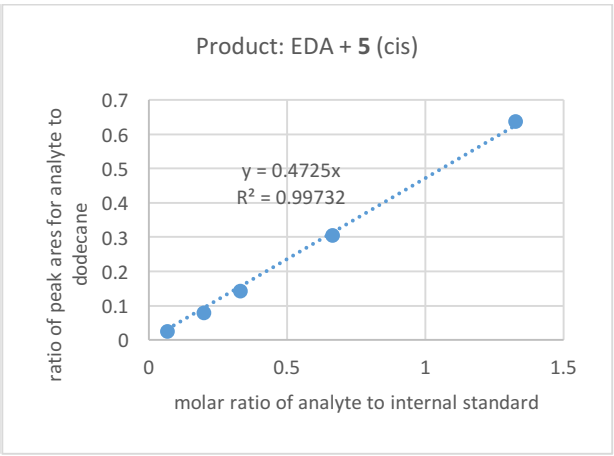

Product: EDA + 9 (trans)
ratio of peak ares for analyte to dodecane
molar ratio of analyte to internal standard
y = 0.9732x
R² = 0.99738
0
0.2
0.4
0.6
0.8
1
1.2
0
0.5
1
1.5


Product: EDA + 9 (cis)
ratio of peak ares for analyte to dodecane
molar ratio of analyte to internal standard
y = 0.8559x
R² = 0.9993
0
0.2
0.4
0.6
0.8
1
1.2
0
0.5
1
1.5


Product: EDA + **10** (cis)
ratio of peak ares for analyte to dodecane
molar ratio of analyte to internal standard
y = 0.915x
R² = 0.99883
0
0.2
0.4
0.6
0.8
1
1.2
0
0.5
1
1.5


Product: EDA + **8** (cis)
ratio of peak ares for analyte to dodecane
molar ratio of analyte to internal standard
y = 0.5129x
R² = 0.99937
0
0.1
0.2
0.3
0.4
0.5
0.6
0
0.5
1
1.5


Product: EDA + **8** (trans)
ratio of peak ares for analyte to dodecane
molar ratio of analyte to internal standard
y = 0.5007x
R² = 0.99958
0
0.1
0.2
0.3
0.4
0.5
0.6
0
0.5
1
1.5


Product: EDA + **5** (cis)
ratio of peak ares for analyte to dodecane
molar ratio of analyte to internal standard
y = 0.4725x
R² = 0.99732
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
0.5
1
1.5

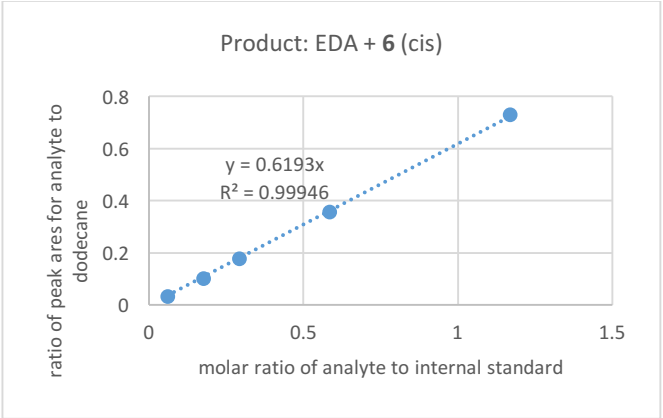
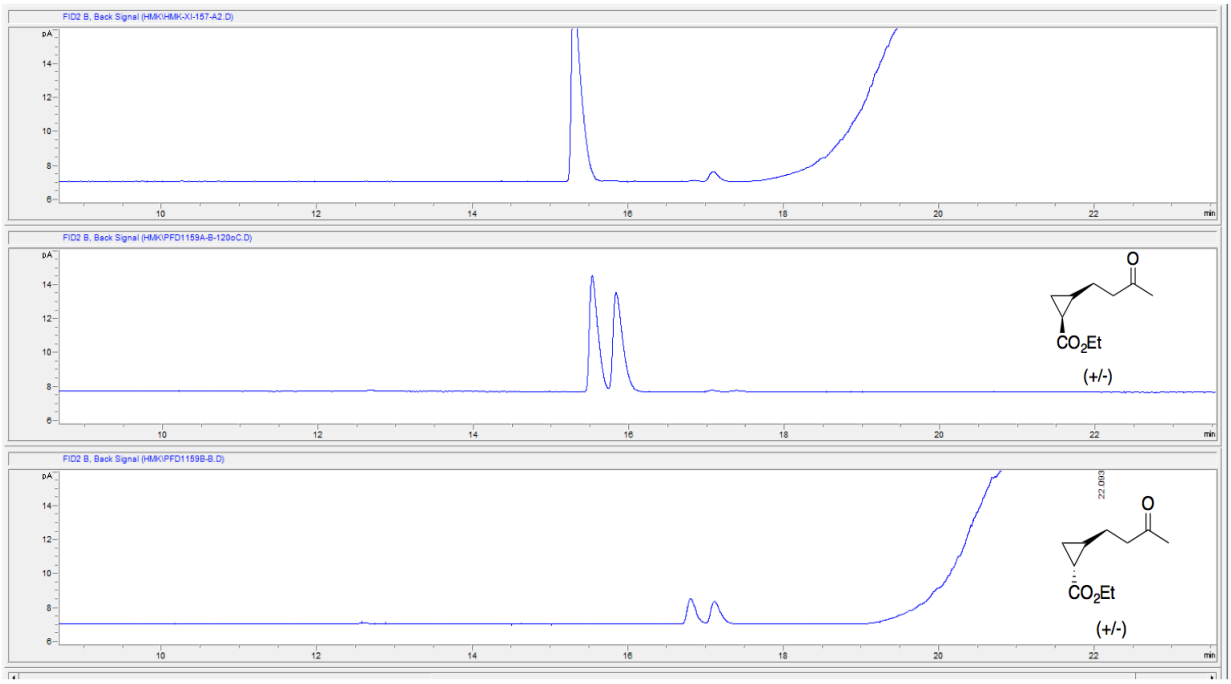
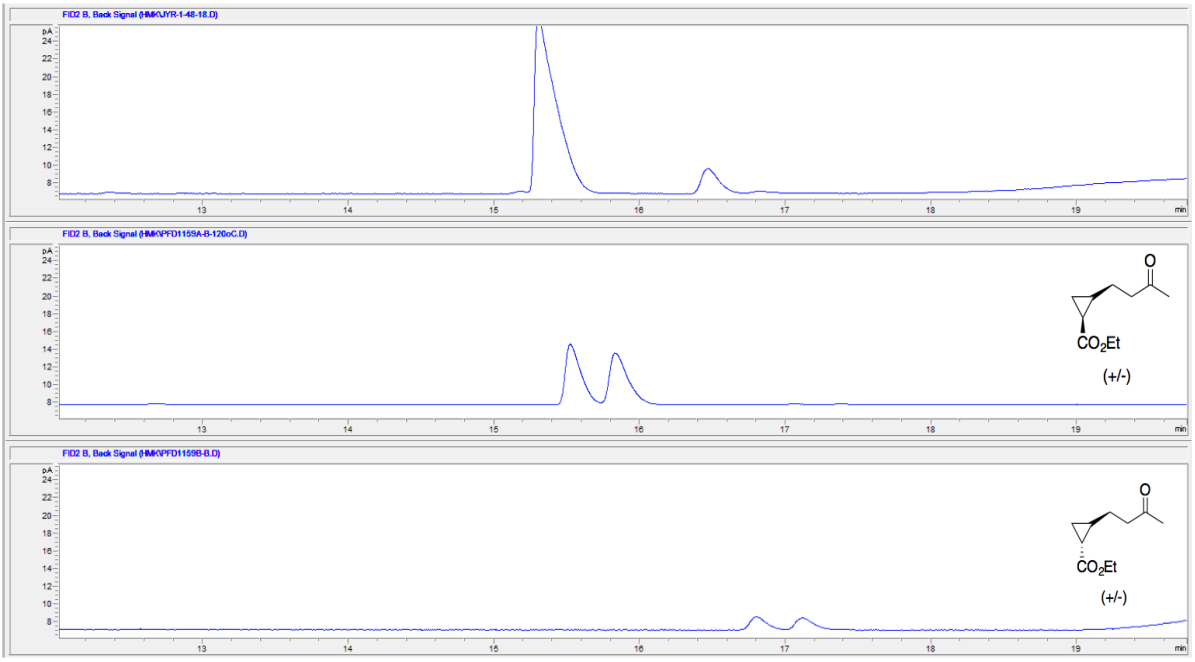
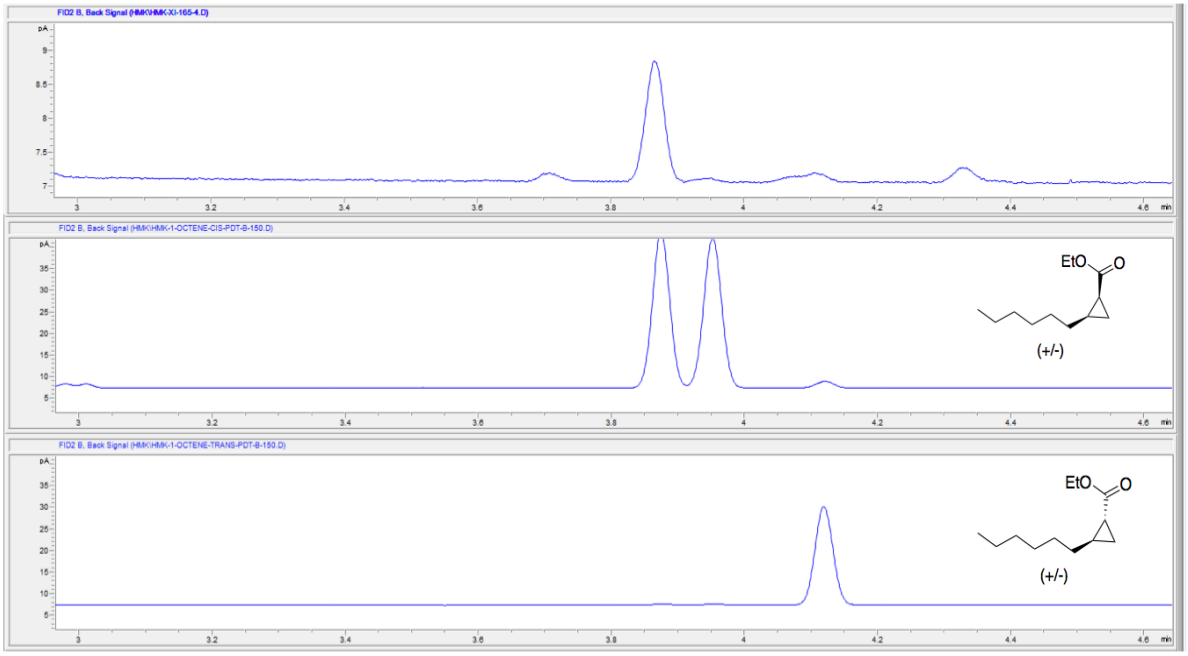
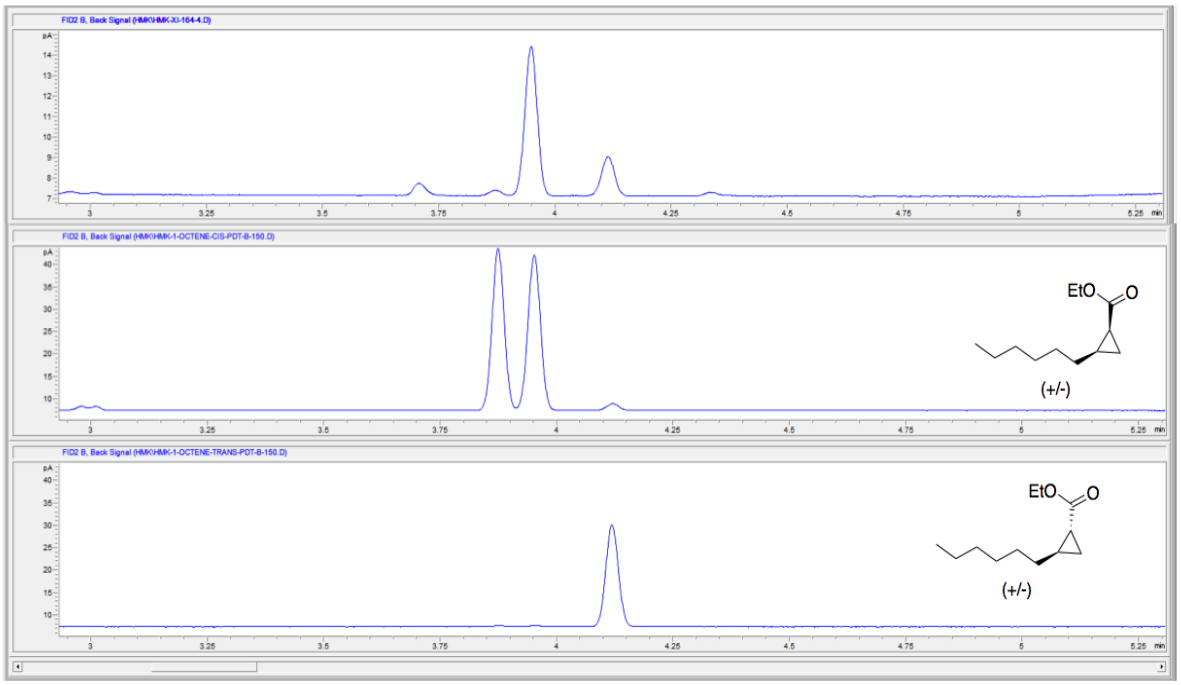

Product: EDA + **6** (cis)
ratio of peak ares for analyte to dodecane
molar ratio of analyte to internal standard
y = 0.6193x
R² = 0.99946
0
0.2
0.4
0.6
0.8
0
0.5
1
1.5

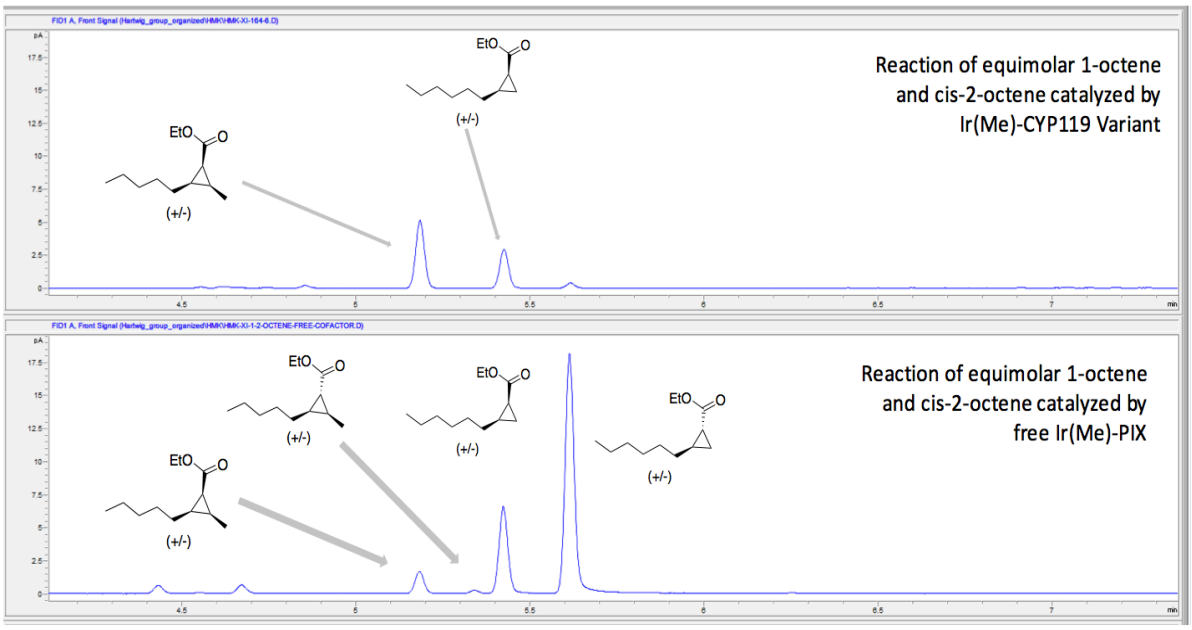
Figure S3. Representative chiral GC traces for stereoselective cyclopropanation reactions. The spectra of authentic compounds are shown below the spectra obtained in reactions with mutants of CYP119.

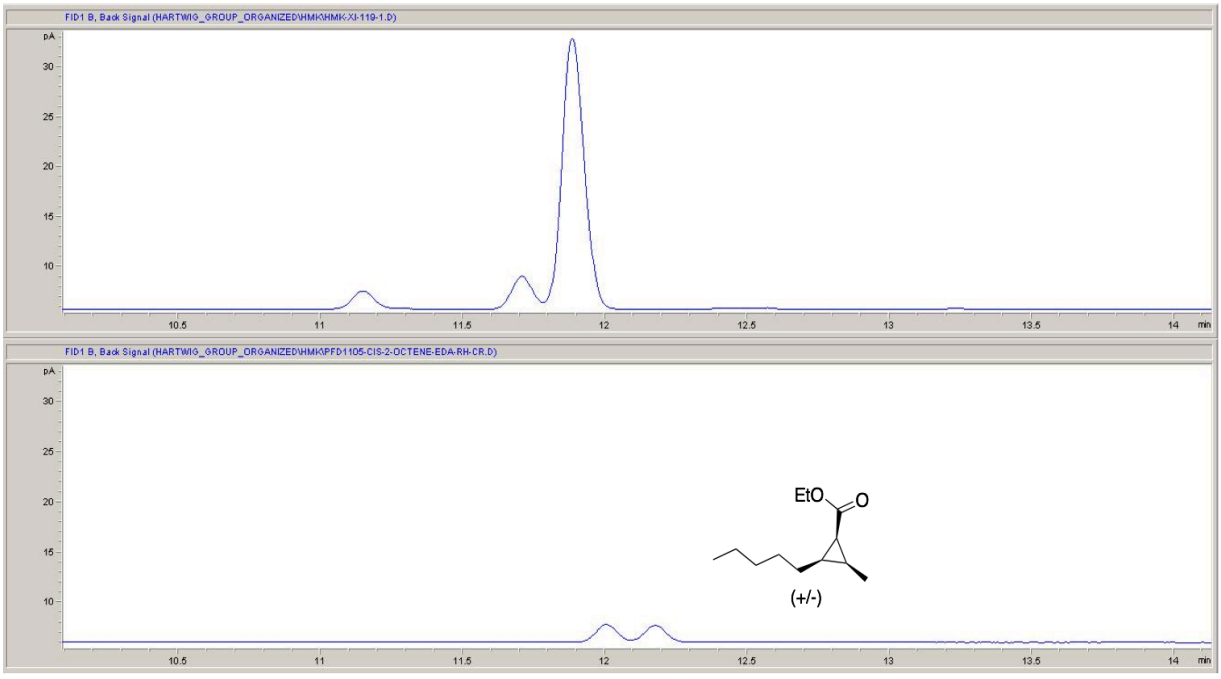


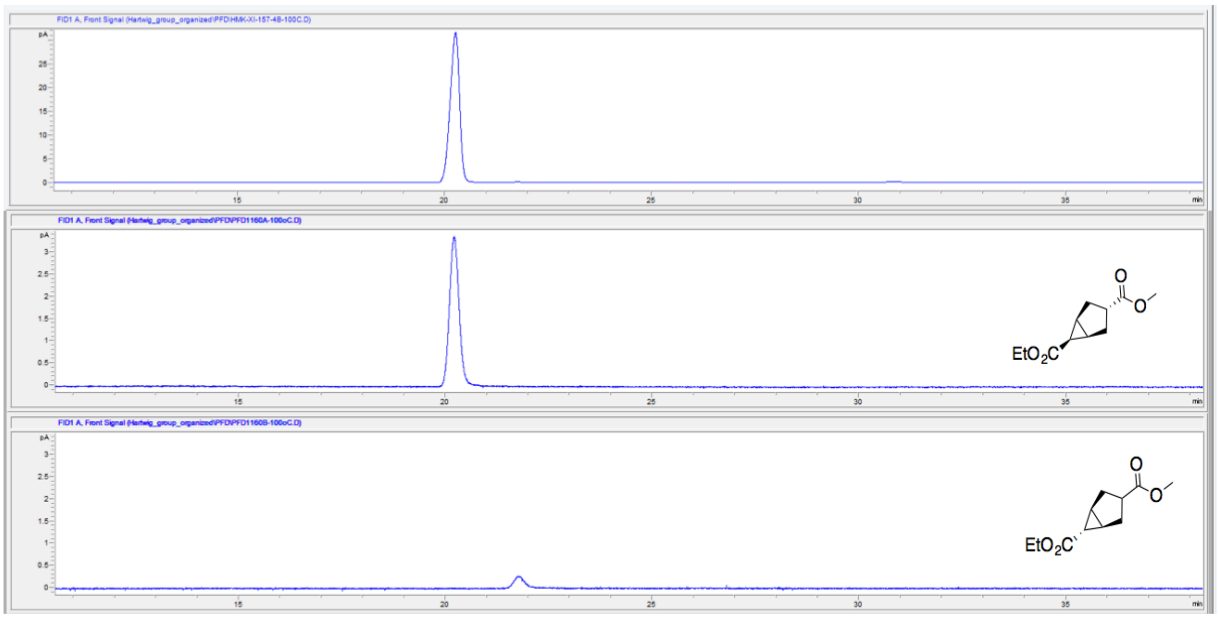

FID2 B, Back Signal (HMK\YR-1-48-18.D)
FID2 B, Back Signal (HMK\PFD1159A-B-120oC.D)
O
CO2Et
(+/-)
FID2 B, Back Signal (HMK\PFD1159B-B.D)
O
CO2Et
(+/-)

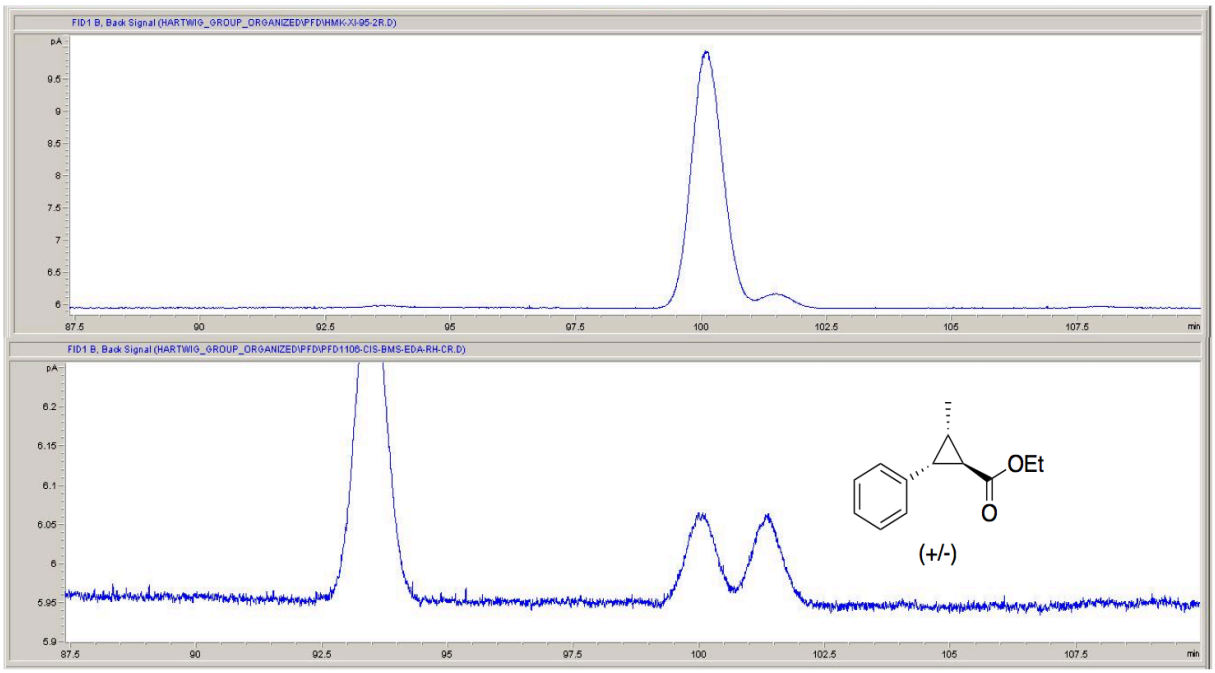

FID2 B, Back Signal (HMK\HMK-XI-165-4.D)
FID2 B, Back Signal (HMK\HMK-1-OCTENE-CIS-PDT-B-150.D)
EtO
O
(+/-)
FID2 B, Back Signal (HMK\HMK-1-OCTENE-TRANS-PDT-B-150.D)
EtO
O
(+/-)


EtO
O
(+/-)
EtO
O
(+/-)


EtO
O
(+/-)
EtO
O
(+/-)
Reaction of equimolar 1-octene and cis-2-octene catalyzed by Ir(Me)-CYP119 Variant
EtO
O
(+/-)
EtO
O
(+/-)
EtO
O
(+/-)
EtO
O
(+/-)
Reaction of equimolar 1-octene and cis-2-octene catalyzed by free Ir(Me)-PIX


FID1 B, Back Signal (HARTWIG_GROUP_ORGANIZED\HMK\HMK-XI-119-1.D)
pA
30
25
20
15
10
10.5
11
11.5
12
12.5
13
13.5
14
min
FID1 B, Back Signal (HARTWIG_GROUP_ORGANIZED\HMK\PFD1105-CIS-2-OCTENE-EDA-RH-CR.D)
pA
30
25
20
15
10
10.5
11
11.5
12
12.5
13
13.5
14
min
EtO
O
(+/-)


FID1 A, Front Signal (Hartwig_group_organized\PFD\HMK-XI-157-48-100C.D)
pA
25
20
15
10
5
0
15
20
25
30
35
min
FID1 A, Front Signal (Hartwig_group_organized\PFD\PFD1160A-100oC.D)
pA
3
2.5
2
1.5
1
0.5
0
15
20
25
30
35
min
O
O
EtO2C
FID1 A, Front Signal (Hartwig_group_organized\PFD\PFD1160B-100oC.D)
pA
3
2.5
2
1.5
1
0.5
0
15
20
25
30
35
min
O
O
EtO2C



5.7.5 Supporting Tables

Table S1. Sequence information for heme proteins used in this study

Protein	Organism	Construction	Vector	Sequence
CYP 119	*Sulfolobus solfataricus*	6xHis-TEV-CYP119	2BT	EGDIHMKSSHHHHHHENLYFQSNAMYDWFSEMRKKDPVY YDGNIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLEDLR NGKIRFDIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQ TLETFIRETTRSLLDSIDPREDDIVKKLAVPLPIIVISK ILGLPIEDKEKFKEWSDLVAFRLGKPGEIFELGKKYLEL IGYVKDHLNSGTEVVSRVVNSNLSDIEKLGYIILLLIAG NETTTNLISNSVIDFTRFNLWQRIREENLYLKAIEEALR YSPPVMRTVRKTKERVKLGDQTIEEGEYVRVWIASANRD EEVFHDGEKFIPDRNPNPHLSFGSGIHLCLGAPLARLEA RIAIEEFSKRFRHIEILDTEKVPNEVLNGYKRLVVRLKS NE

Table S2. Plasmid library obtained from site directed mutagenesis of the CYP119 gene within the 2B plasmid in the course of directed evolution.

	69X	209X	213X	254X
L		115-C10	116-G2	115-B12
V	115-D1	115-C2	116-F7	
A			116-H3	115-B6
G	115-G4	115-C8	115-F2	115-E8
Y	115-D3		116-D1	115-B8
W		115-C12	116-F1	115-B10
F	115-D10		115-H8	115-B11
T	115-D2			117-A1

T213G + …

	69X	209X	254X	310X	155X	152X
L			115-E3			
V	116-B1					
A			117-E4			
G	115-G2		117-E7			
Y	115-G5		116-E6			140-H7
W	116-A12		115-E2	140-B9		
F	115-H1		115-E7		141-C1	
T	115-G10	115-F7	115-E4		141-B1	

	69L	69V	69A	69G	69Y	69W	69F	69T
213L						116-B8		115-H9
213V		116-B10		116-A11	116-A10		116-B11	
213A		116-B3	116-B6			115-H11	116-B5	
213G								
213Y			116-B7		116-A9			115-H4
213W					116-B4			116-A5
213F			115-H5			115-H2	115-H7	
213T								

	254L	254V	254A	254G	254Y	254W	254F	254T
213L	116-D3					116-C10		116-C4
213V				116-C12				
213A	116-C3			116-E8			116-C6	
213G								
213Y			116-D7					
213W	116-E12				116-D12			116-E3
213F					116-E4	116-E10		
213T								

	209L	209V	209A	209G	209Y	209W	209F	209T
213L		116-F2				116-G6		
213V								
213A								
213G								
213Y								
213W							116-D2	
213F								
213T								

	209L	209V	209A	209G	209Y	209W	209F	209T
254L	117-C3					117-A10		
254V								
254A						117-A3		
254G								
254Y		117-A4				117-B3		
254W	117-C5	117-B11						
254F		117-C2		117-B7				
254T								

	69L	69V	69A	69G	69Y	69W	69F	69T
209L		117-F6		117-G8	117-F5	117-G6		117-H12
209V			117-F9			117-H3		117-H7
209A								
209G		117-H8	117-G4				117-G9	
209Y								
209W		117-H2		117-H6			117-F7	117-F11
209F								
209T			117-G5	117-H4	117-H5	117-H10		117-G1

213G, 254L (115-E3) + Additional Mutation

	69X	310X	318X
L		140-B10	
V	140-B6		
A	140-A3	140-D2	
G	140-B3	140-B8	
Y	140-A4	140-C5	
W	140-B2		
F	140-A8		
T		140-C1	

213G, 254L (115-E3) +

	152X	155X	
L	140-H1		
V	141-A5	141-C10	
A		141-B3	
G	141-A9	141-C3	
Y	140-H2	141-B11	
W	141-A11	141-B2	
F	140-H11	141-D3	
T	140-G9	141-G4	

213G, 69Y (115-G5) +

	69X	310X	318X
L	142-H4		
V			
A			
G		140-C11	
Y	142-H11		
W	142-G7	140-D9	
F			
T	142-H12		

213G, 69Y (115-G5) +

	152X	155X	
L			
V	140-G8		
A			
G			
Y	140-H10		
W	140-H6		
F			
T			

T213A (116-H3) +

	69X	254X	310X
L			166-A12
V			
A			
G		165-H1	
Y		165-H3	166-A6
W			
F		165-H6	166-A9
T		165-H7	166-A7

T213A (116-H3) +

	318X	152X	155X
L		166-C7	
V	166-B11		166-D2
A			
G			
Y	166-B1	166-C4	166-D7
W	166-B6		166-D5
F	166-B7	166-C3	166-D4
T	166-B3		166-D1

254X

	69F, 213G	69V, 213G	
L	141-F6	141-E9	
V			
A		141-E1	
G			
Y	141-D7	141-D9	
W	141-F1	141-E3	
F		141-E7	
T			

69Y, 254L, 213G +

	310X	155X	
L			
V			
A			
G			
Y			
W	140-B9	141-B12	
F			
T		141-B7	

69W, 213G (116-A12) +

	209X	254X	310X
L	165-A4	165-B6	
V	165-A5		165-C11
A		165-B2	165-C10
G	165-A11		165-C1
Y	165-A7	165-B8	
W		165-B3	165-C5
F	165-A2		
T		165-B7	165-C2

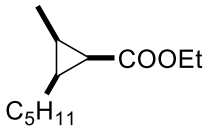
69W, 213G (116-A12) +

	318X	152X	155X
L		165-E8	165-F11
V	165-D9		
A			
G	165-D3		
Y	165-D7	165-E5	165-F5
W	165-D2	165-E7	165-F2
F		165-E1	165-F4
T	165-D1		165-F1

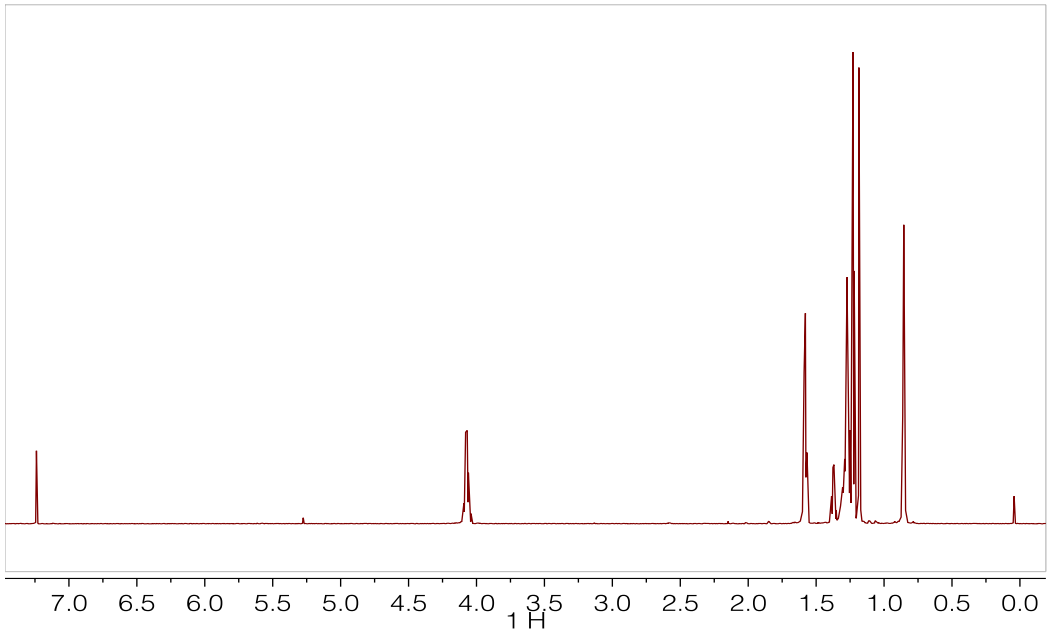
Table S3. Methods used to separate enantiomers of the generated reaction products.

Product from EDA + Alkene X	Instrument	Column	Method
Styrene	GC	CYCLOSIL-B (30m x 0.25mm x 0.25u)	Isothermal: 120 °C
a-Me Styrene	GC	CYCLOSIL-B (30m x 0.25mm x 0.25u)	Isothermal: 120 °C
Trans B-Me Styrene	GC	CYCLOSIL-B (30m x 0.25mm x 0.25u)	
Cis B-Me Styrene	GC	CYCLOSIL-B (30m x 0.25mm x 0.25u)	Isothermal: 110 °C
1-octene	GC	CP-Chirocel Dex-CB (30m x 0.25mm x 0.25u)	Isothermal: 150 °C
Cis-2-octene	GC	CP-Chirocel Dex-CB (30m x 0.25mm x 0.25u)	Isothermal: 120 °C
1-Hexen-5-one	GC	CP-Chirocel Dex-CB (30m x 0.25mm x 0.25u)	Isothermal: 120 °C
Cyclopentene-3-methyl-ester	GC (achiral)	HP-5	Isothermal: 100 °C

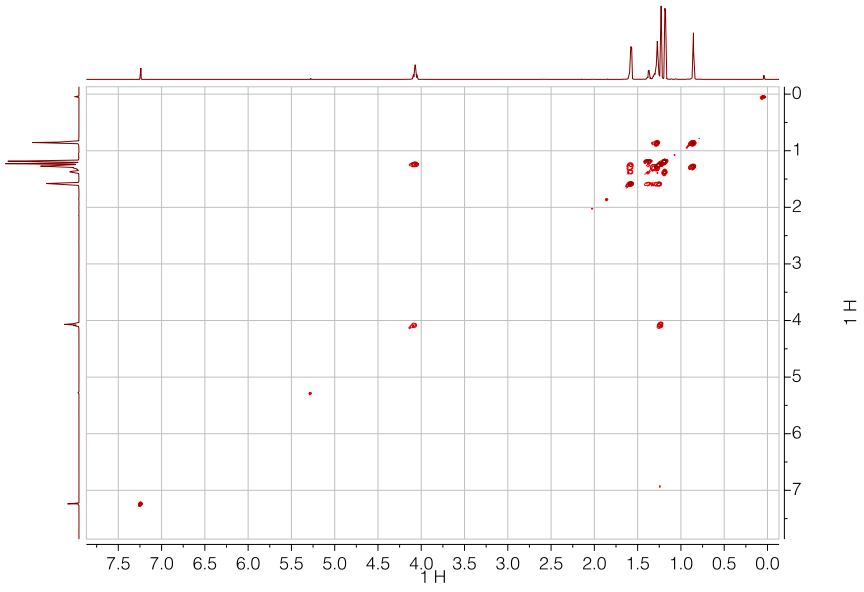
5.8 NMR Spectra of New Compounds



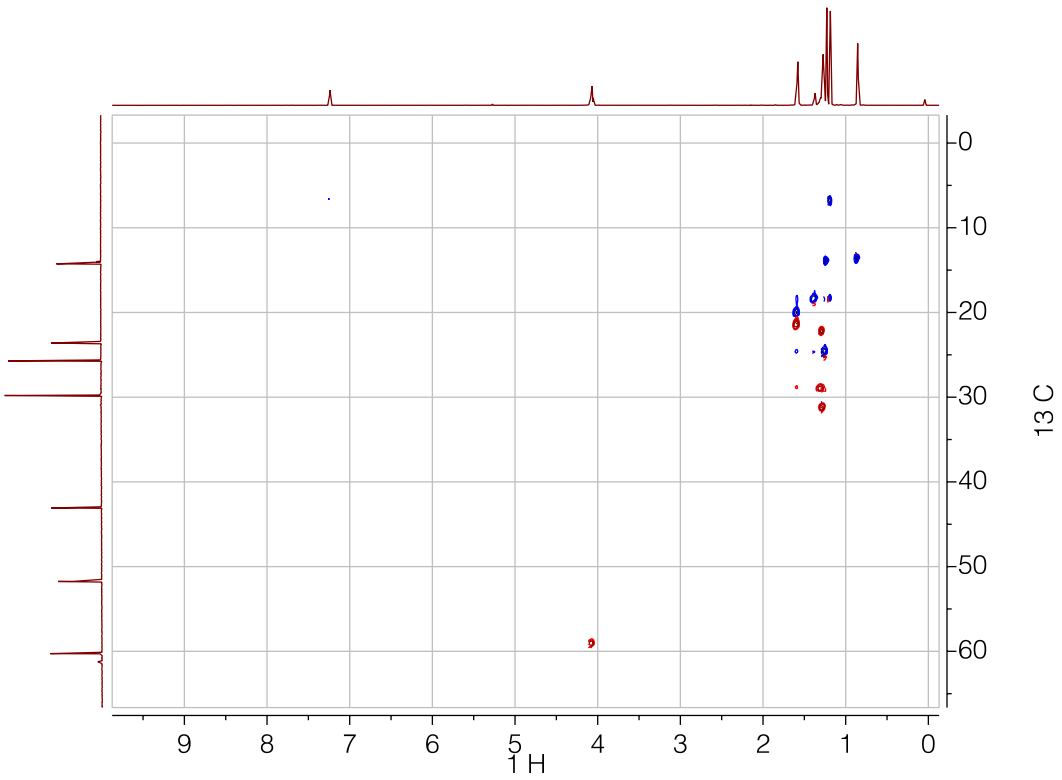
^{1}H NMR



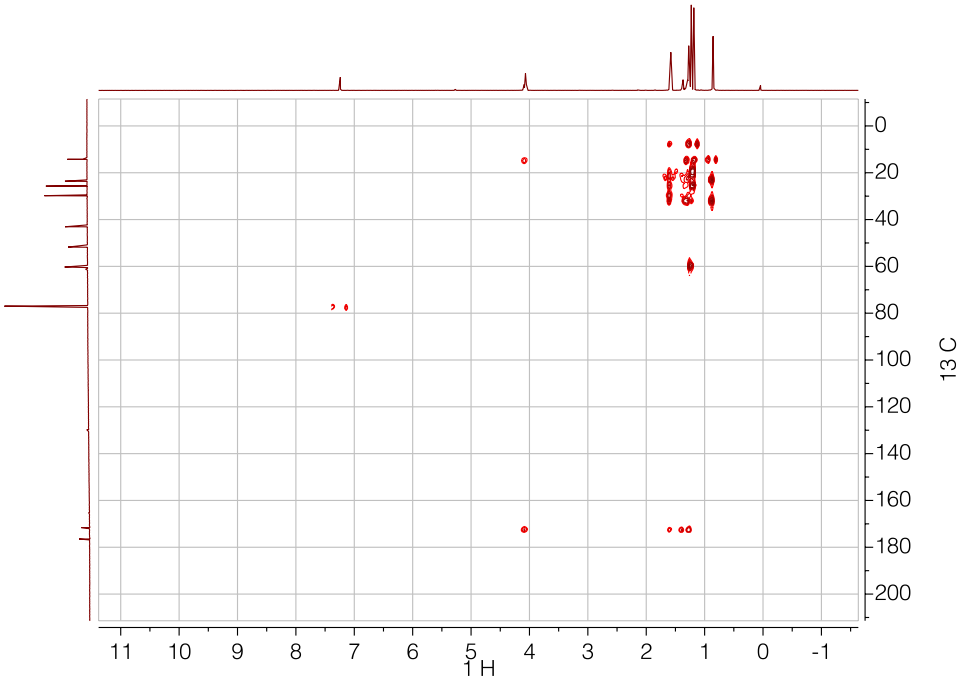
COSY



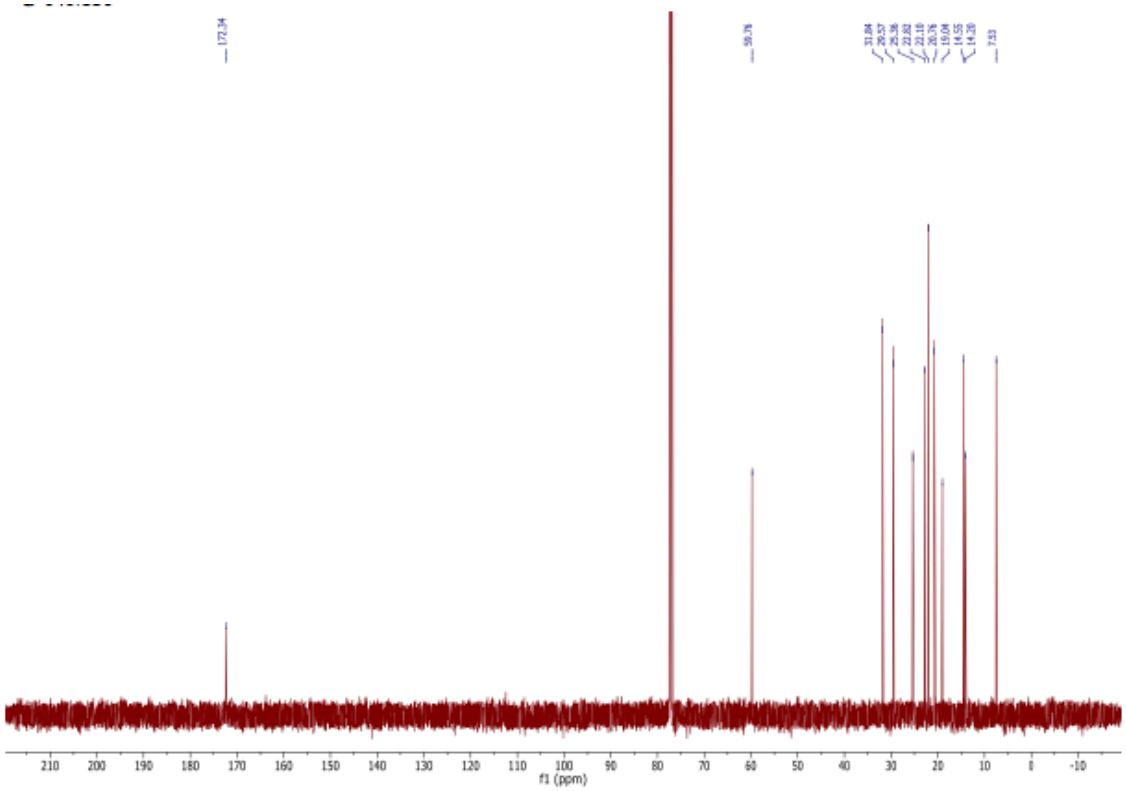
HSQC

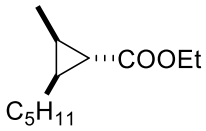

0
10
20
30
40
50
60
13 C
9
8
7
6
5
4
3
2
1
0
1 H

HMBC

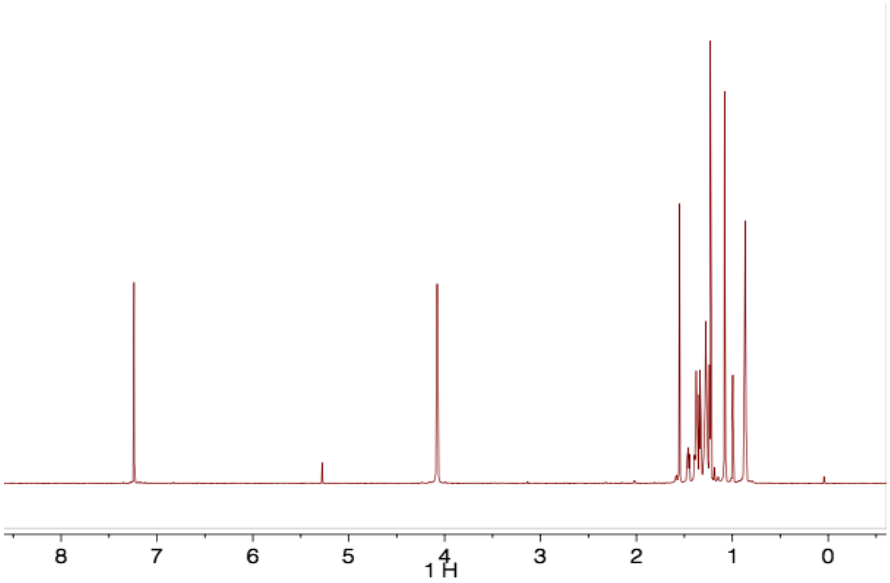

0
20
40
60
80
100
120
140
160
180
200
13 C
11
10
9
8
7
6
5
4
3
2
1
0
-1
1 H

^{13}C NMR

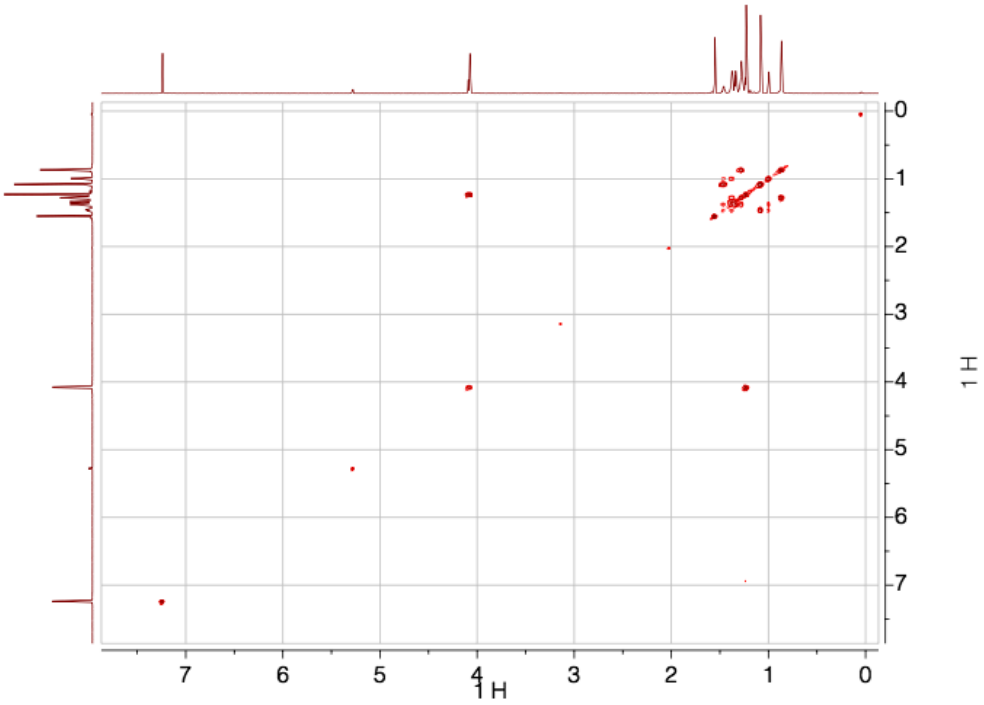

172.34
59.76
31.84
29.57
25.36
22.82
22.10
20.76
19.04
14.55
14.20
7.53
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)



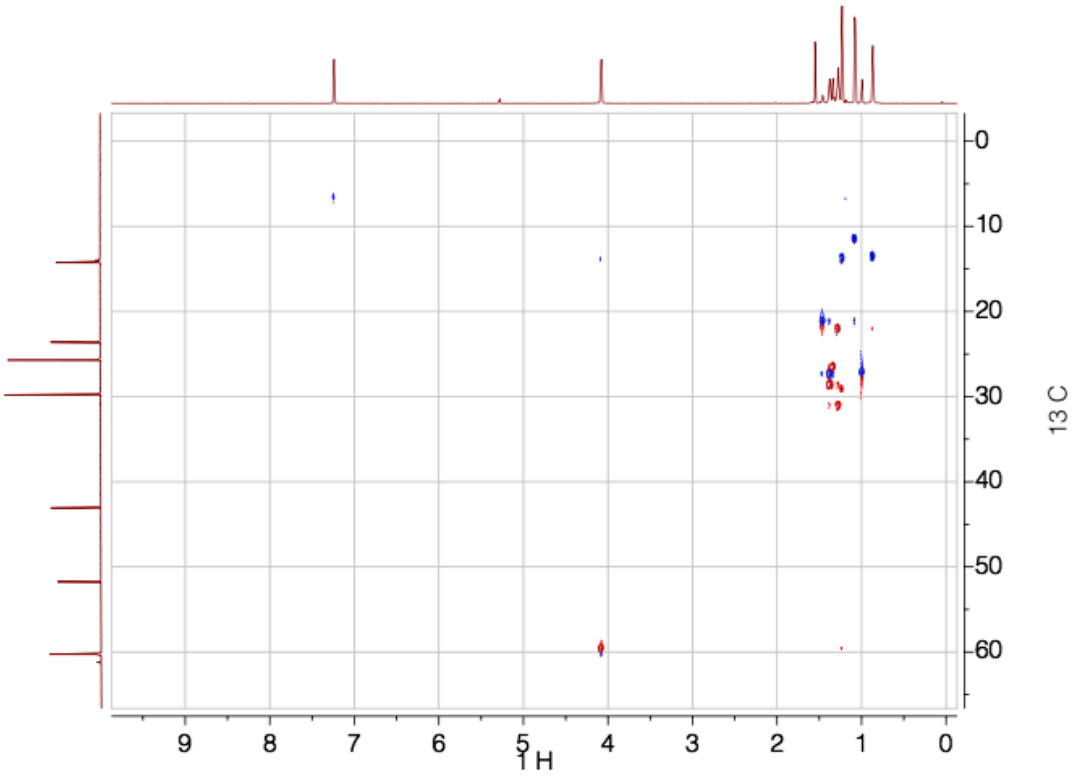
^{1}H NMR



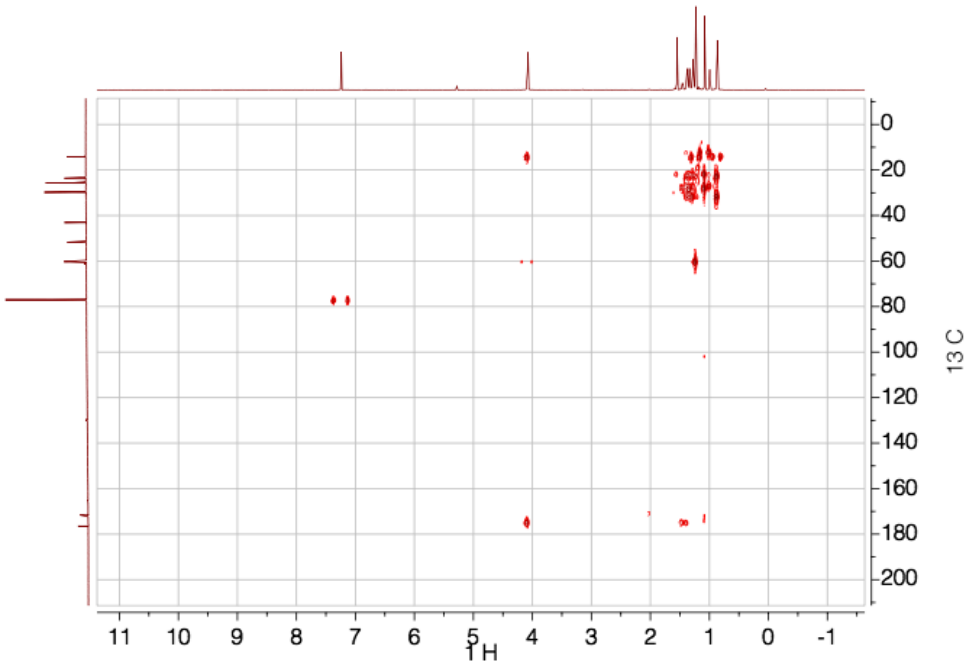
COSY



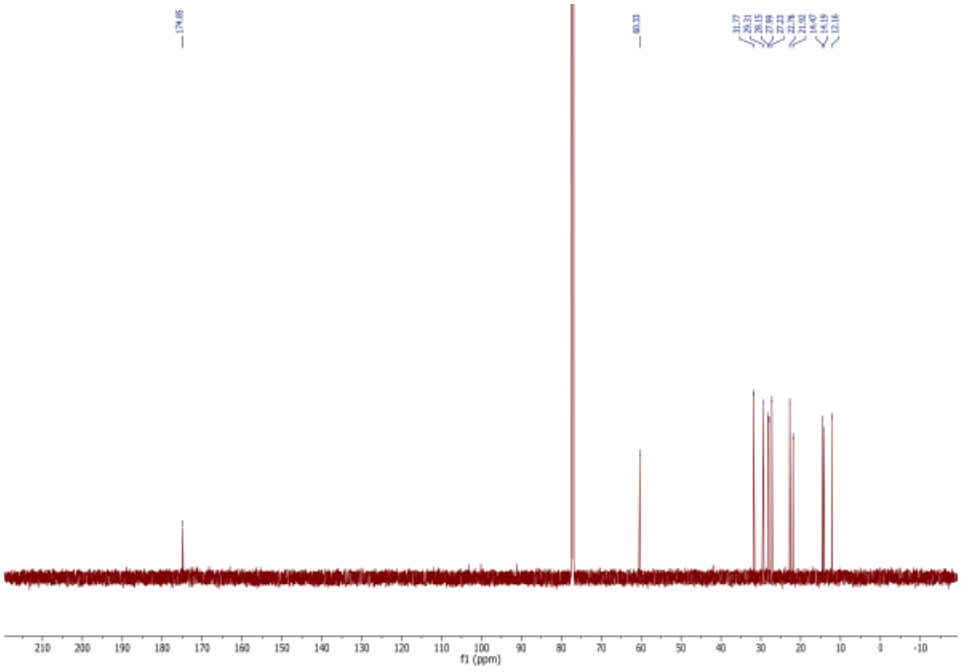
HSQC

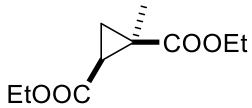

0
10
20
30
40
50
60
13 C
9
8
7
6
5
4
3
2
1
0
1 H

HMBC

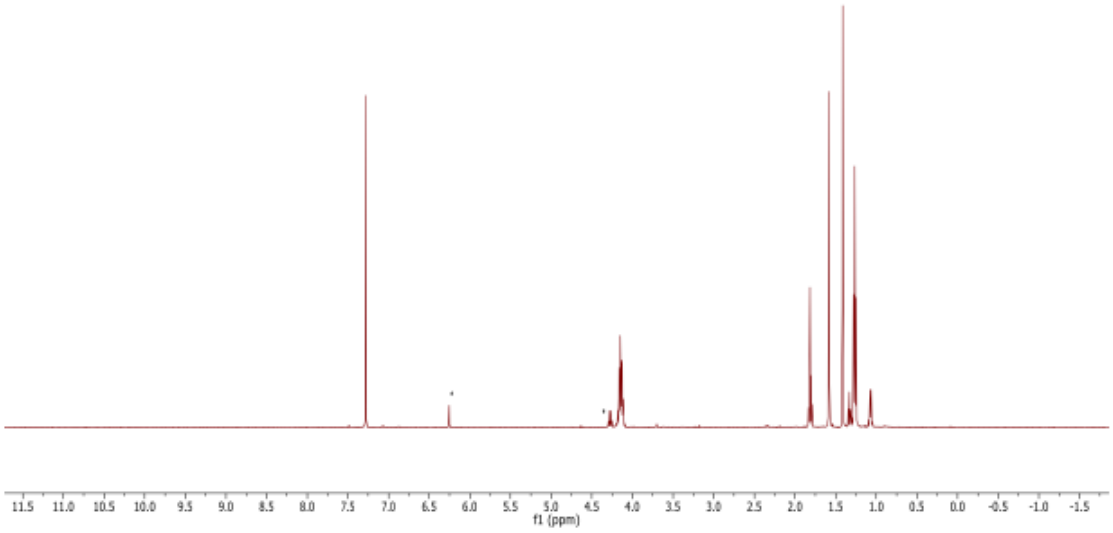

0
20
40
60
80
100
120
140
160
180
200
13 C
11
10
9
8
7
6
5
4
3
2
1
0
-1
1 H

^{13}C NMR

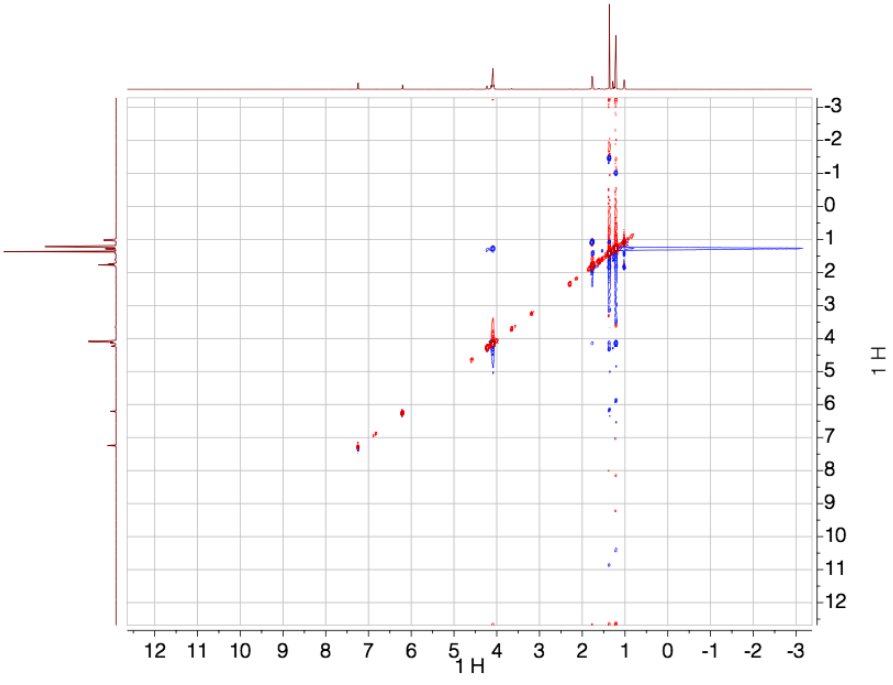

174.65
60.33
31.77
29.31
28.15
27.99
27.33
22.78
21.92
14.47
14.19
13.16
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)


COOEt
EtOOC

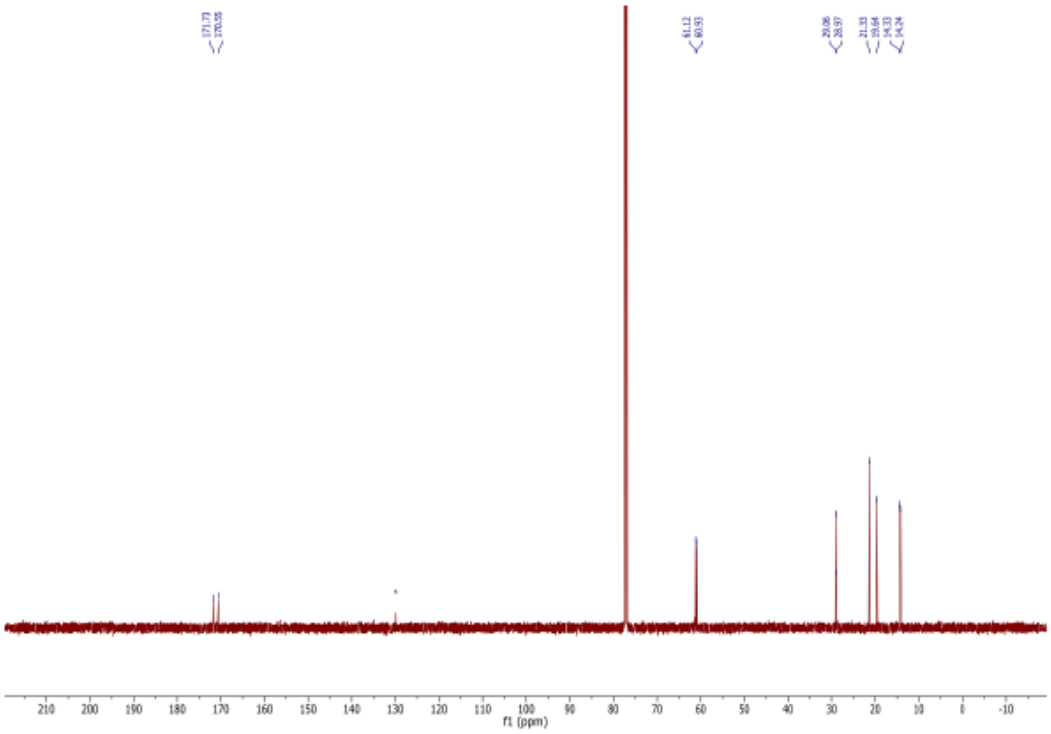
^{1}H NMR

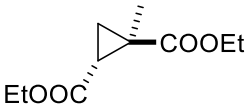

11.5 11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5 -1.0 -1.5
f1 (ppm)

NOESY

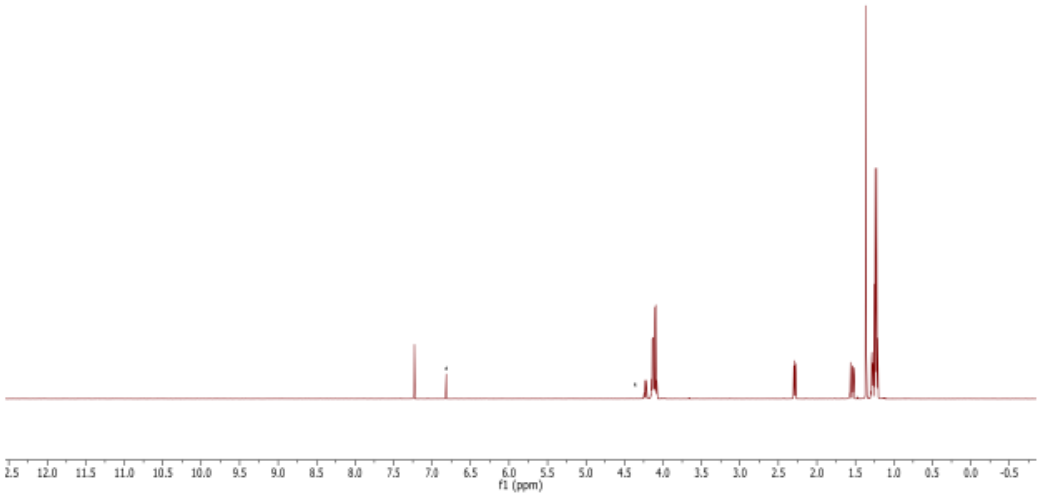

-3 -2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12
1 H
12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2 -3
1 H

^{13}C NMR

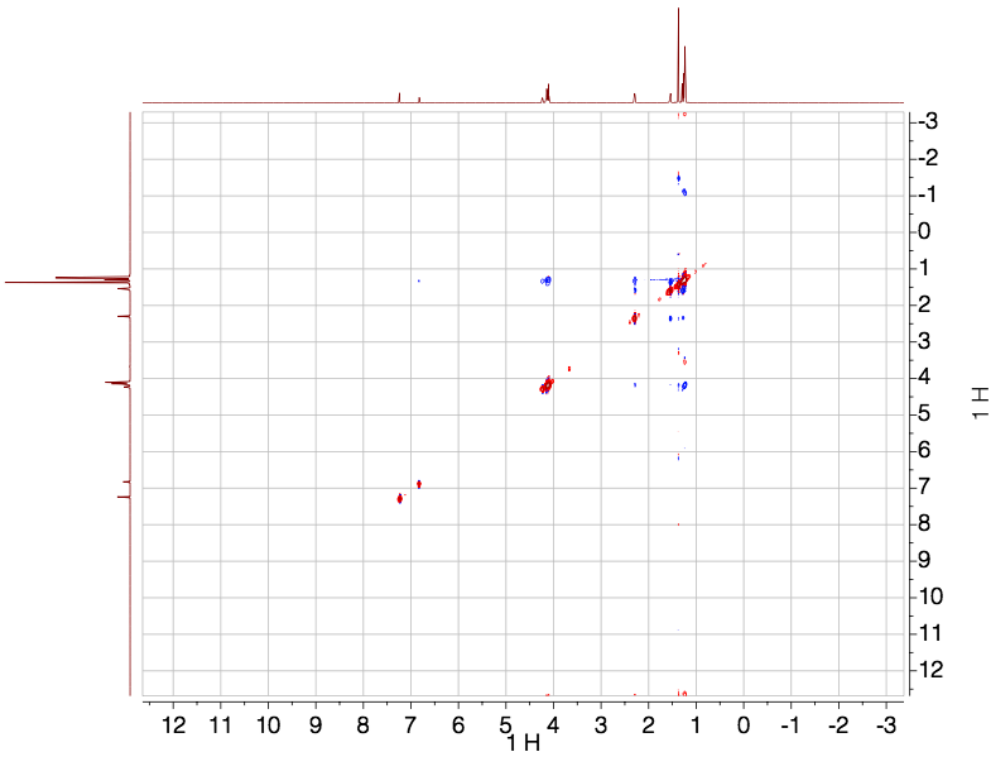

171.73
170.56
61.12
60.93
29.06
28.97
21.33
21.04
14.33
14.24
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)



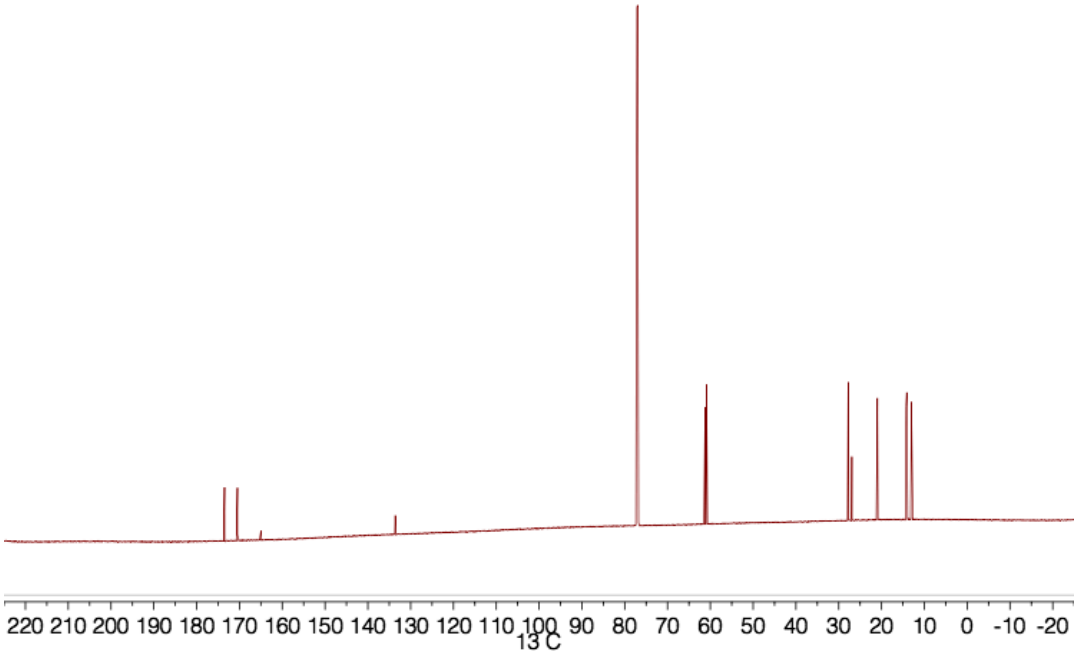
^{1}H NMR

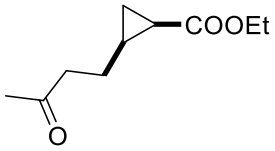


NOESY

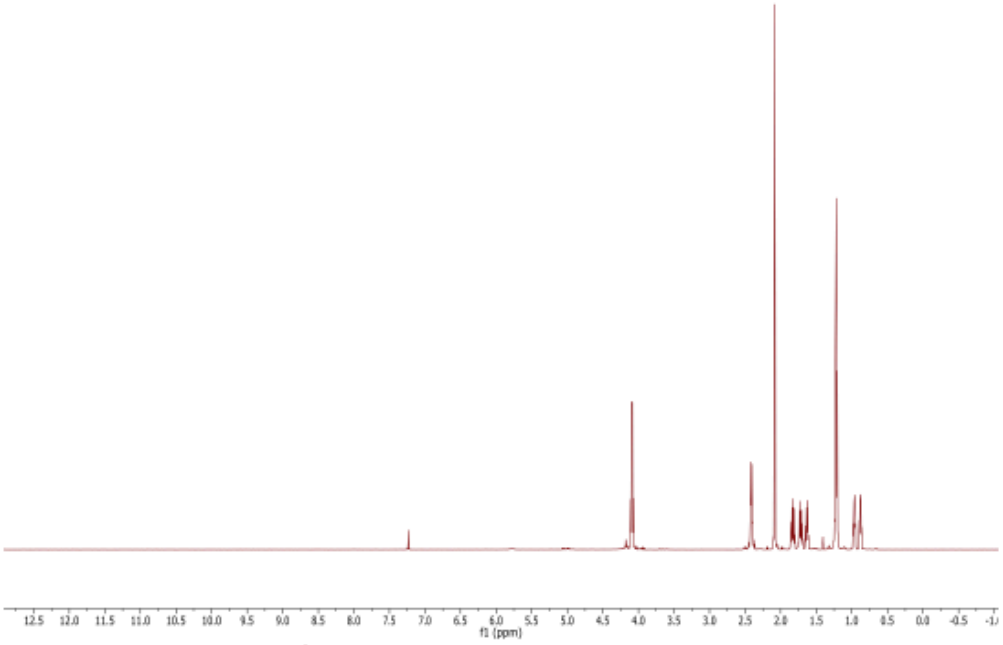


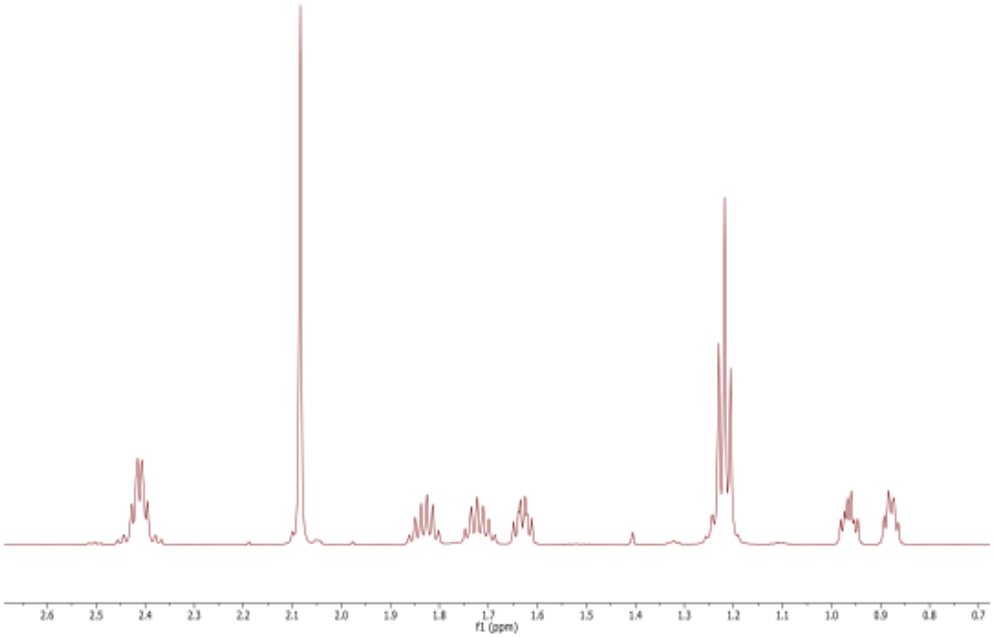
^{13}C NMR


220 210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10 -20
13 C

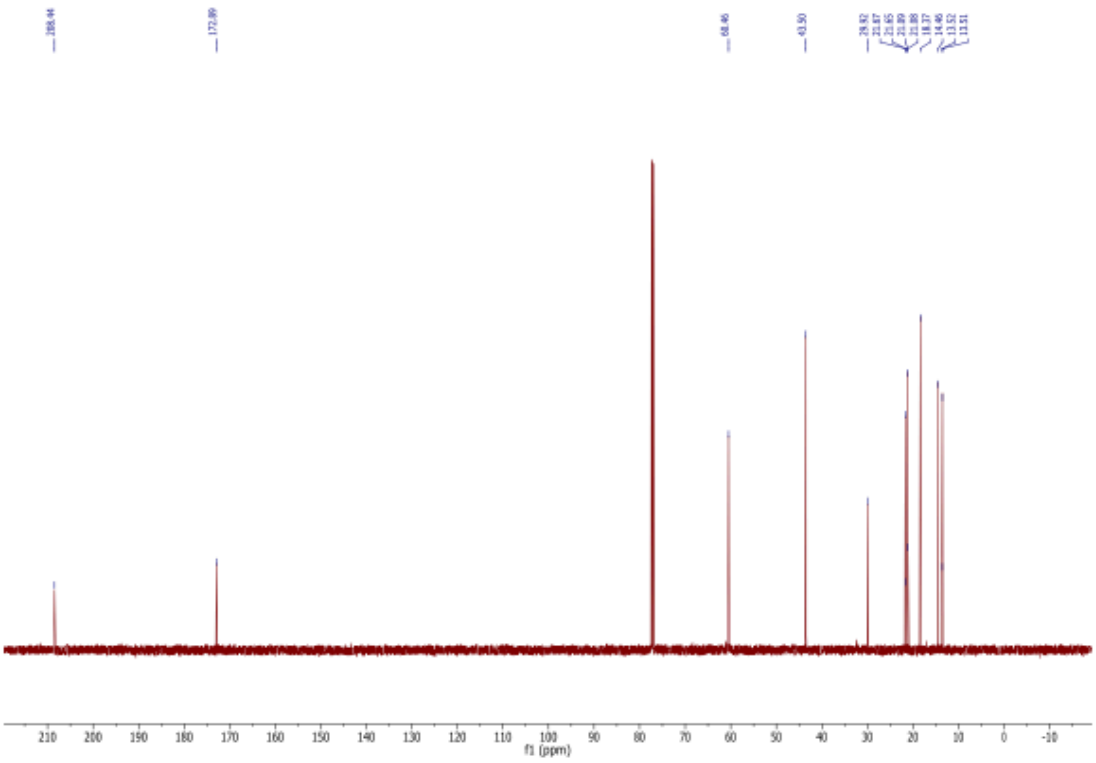


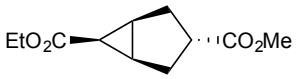
^{1}H NMR



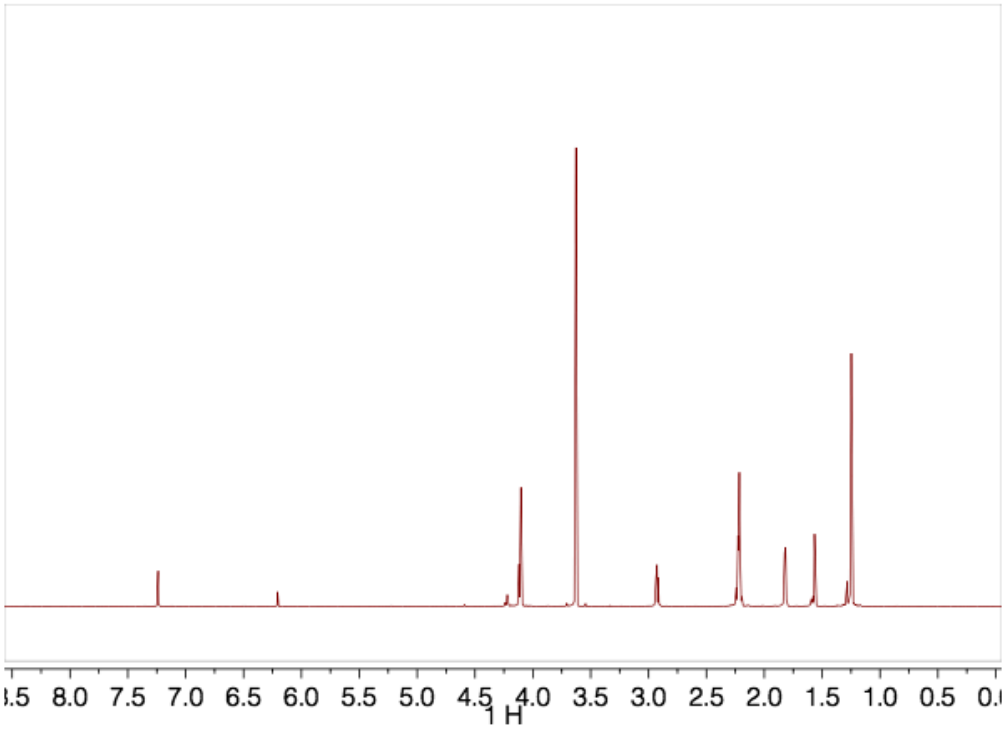


^{13}C NMR

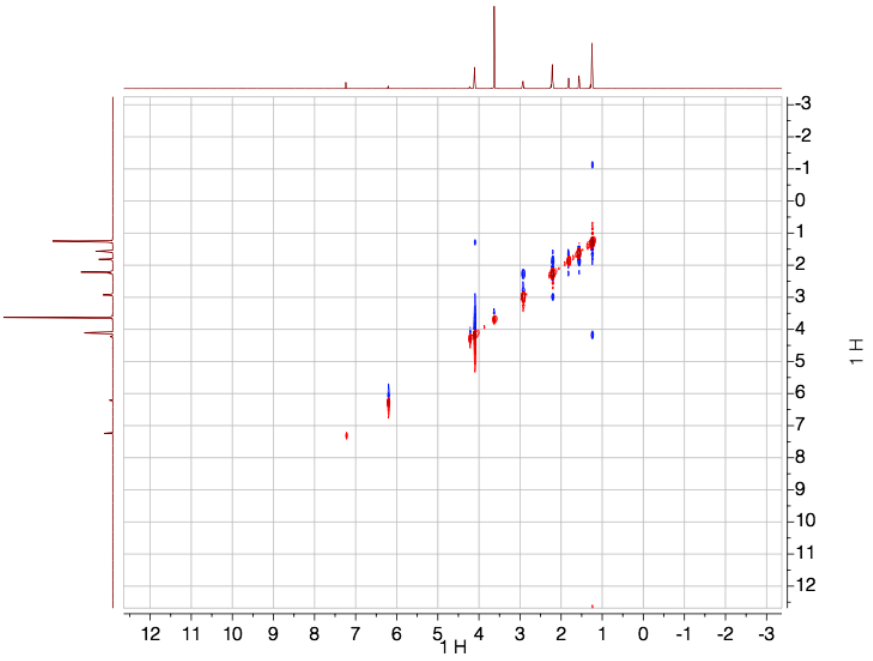

208.84
172.89
68.46
63.50
29.82
22.87
21.85
21.89
21.88
18.37
14.46
13.52
13.51
210
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
-10
f1 (ppm)


EtO2C
CO2Me

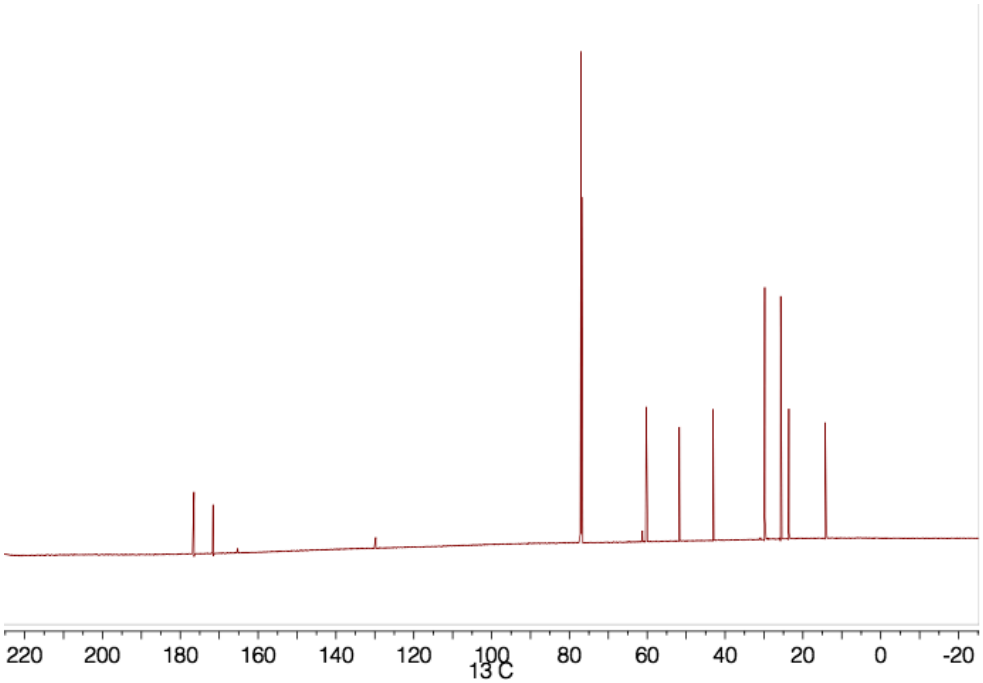
^{1}H NMR

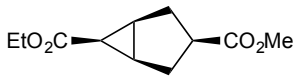

8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5
1 H

NOESY

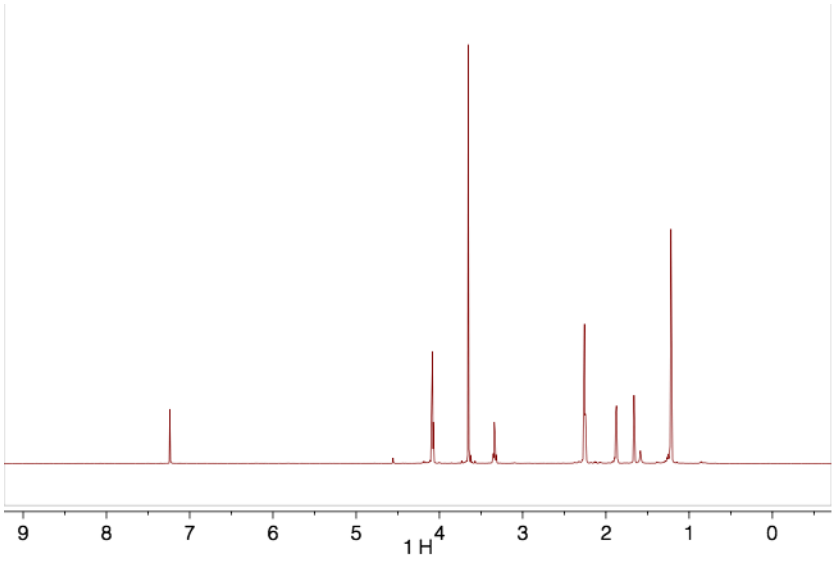

12 11 10 9 8 7 6 5 4 3 2 1 0 -1 -2 -3
1 H
-3 -2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12
1 H

^{13}C NMR

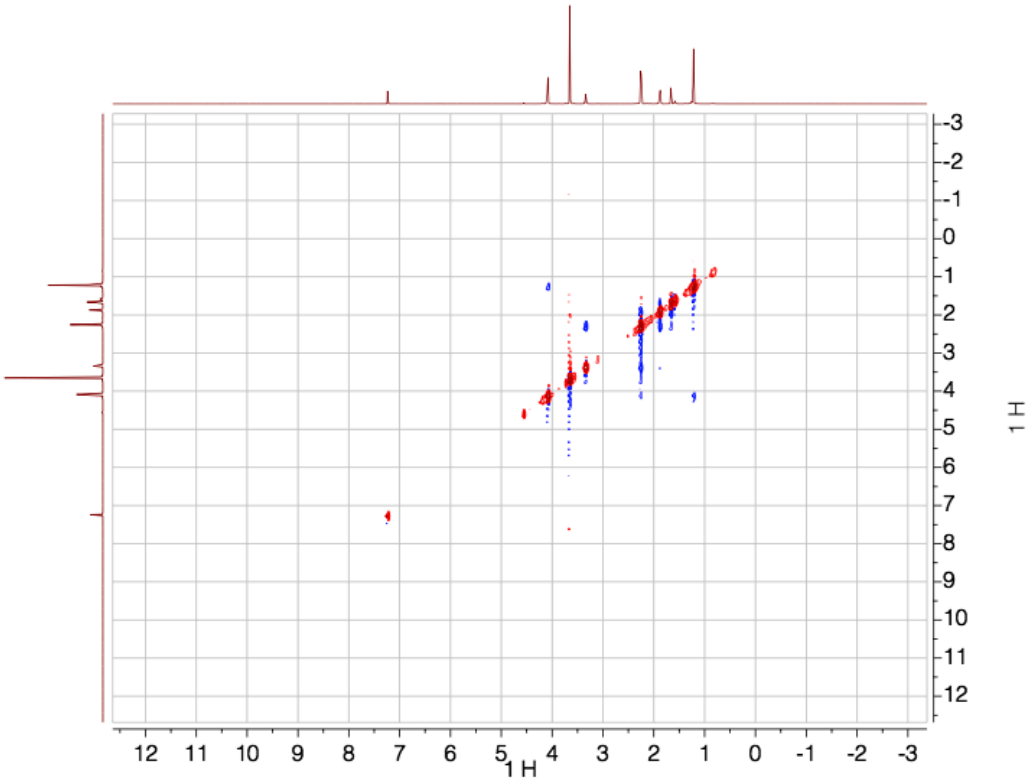




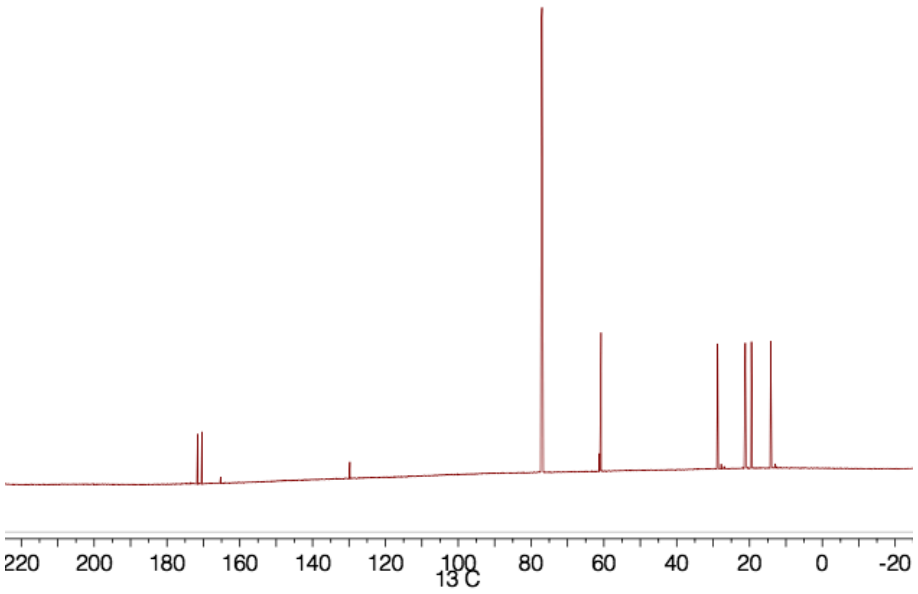
^{1}H NMR



NOESY



^{13}C NMR



5.9 Acknowledgements

This work was supported by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02- 05CH11231, by the NSF (graduate research fellowship to HMK), and the NWO Netherlands Organization for Scientific Research (the Rubicon postdoctoral fellowship, No. 680-50-1306 to PD). We thank the QB3 MacroLab facility (competent cells), the UC Berkeley DNA Sequencing Facility (plasmid sequencing), and Dr. Tony Iavarone and the QB3 Mass Spectrometry Facility (NS-ESI data collection, supported by NIH grant 1S10OD020062-01), for nativeNS-ESI-MS data and analysis.

5.10 References

1. Koeller, K. M. & Wong, C.-H. Enzymes for chemical synthesis. *Nature* **409,** 232–240 (2001).
2. *The Ubiquitous Roles of Cytochrome P450 Proteins*. (John Wiley & Sons, Ltd, 2007). doi:10.1002/9780470028155
3. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16,** 379–394 (2015).
4. Roiban, G.-D. & Reetz, M. T. Expanding the toolbox of organic chemists: directed evolution of P450 monooxygenases as catalysts in regio- and stereoselective oxidative hydroxylation. *Chem. Commun.* **51,** 2208–2224 (2015).
5. Lewis, J. C. Artificial Metalloenzymes and Metallopeptide Catalysts for Organic Synthesis. *ACS Catal.* **3,** 2954–2975 (2013).
6. Key, H. M., Dydio, P., Clark, D. S. & Hartwig, J. F. Abiological catalysis by artificial haem proteins containing noble metals in place of iron. *Nature* **534,** 534–537 (2016).
7. Chan, K. H., Guan, X., Lo, V. K. Y. & Che, C.-M. Elevated Catalytic Activity of Ruthenium(II)–Porphyrin-Catalyzed Carbene/Nitrene Transfer and Insertion Reactions with N-Heterocyclic Carbene Ligands. *Angew. Chem. Int. Ed.* **53,** 2982–2987 (2014).
8. Rabe, K. S., Kiko, K. & Niemeyer, C. M. Characterization of the Peroxidase Activity of CYP119, a Thermostable P450 FromSulfolobus acidocaldarius. *ChemBioChem* **9,** 420–425 (2008).
9. Coelho, P. S., Brustad, E. M., Kannan, A. & Arnold, F. H. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **339,** 307–310 (2013).
10. Fulmer, G. R. *et al.* NMR Chemical Shifts of Trace Impurities: Common Laboratory Solvents, Organics, and Gases in Deuterated Solvents Relevant to the Organometallic Chemist. *Organometallics* **29,** 2176–2179 (2010).
11. Kamata, K., Kimura, T. & Mizuno, N. Cyclopropanation of Olefins with Diazo Compounds Catalyzed by a Dicopper-substituted Silicotungstate [γ-H2SiW10O36Cu2(μ-1,1-N3)2]4−. *Chemistry Letters* **39,** 702–703 (2010).
12. Bordeaux, M., Tyagi, V. & Fasan, R. Highly diastereoselective and enantioselective olefin cyclopropanation using engineered myoglobin-based catalysts. *Angew. Chem. Int. Ed.* **54,**

1744–1748 (2015).
13. Feldman, K. S. & Simpson, R. E. The oxygenation of substituted vinylcyclopropanes: preparative and mechanistic studies. *J. Am. Chem. Soc.* **111,** 4878–4886 (1989).
14. Anding, B. J., Ellern, A. & Woo, L. K. Olefin Cyclopropanation Catalyzed by Iridium(III) Porphyrin Complexes. *Organometallics* **31,** 3628–3635 (2012).
15. Bonaccorsi, C. & Mezzetti, A. Optimization or Breakthrough? The First Highly cis- and Enantioselective Asymmetric Cyclopropanation of 1-Octene by 'Electronic and Counterion' Tuning of [RuCl(PNNP)]+ Catalysts. *Organometallics* **24,** 4953–4960 (2005).
16. Chen, Y., Ruppel, J. V. & Zhang, X. P. Cobalt-Catalyzed Asymmetric Cyclopropanation of Electron-Deficient Olefins. *J. Am. Chem. Soc.* **129,** 12074–12075 (2007).
17. Carfagna, C., Mariani, L., Musco, A., Sallese, G. & Santi, R. The regio- and stereoselectivities of the reaction of allyl acetates and silyl ketene acetals catalyzed by palladium(0) complexes: a new route to cyclopropane derivatives. *J. Org. Chem.* **56,** 3924–3927 (1991).
18. Ward, T. R. Artificial Metalloenzymes Based on the Biotin−Avidin Technology: Enantioselective Catalysis and Beyond. *Acc. Chem. Res.* **Online,** 47–57 (2010).

Chapter 6: Enzymatic, Chemoselective C-H Amination Catalyzed by CYP119 Containing an Ir(Me)-PIX Cofactor

This chapter is modified from a manuscript prepared for submission with permission of the university and all coauthors.

Pawel Dydio contributed equally to the work described in this chapter.

ABSTRACT: Cytochrome P450 enzymes, enzymes containing Fe-porphyrin (Fe-PIX) cofactors, catalyze reactions that oxidize C-H bonds with high selectivity. Recently, Fe-PIX enzymes have been engineered to catalyze analogous C-H amination reactions. However, the utility of these transformations are limited by poor chemoselectivity for the C-H amination pathway over competing reduction pathways. Recently, we reported that the introduction of abiological metals within PIX-proteins can create enzymes with reactivity profiles that are distinct from those of native Fe-PIX enzymes. Here we report that Ir(Me)-PIX containing CYP119 variants form active and selective catalysts for the insertion of nitrenes into C-H bonds with high chemoselectivity for C-H insertion over nitrene reduction. These C-H insertion reactions are accomplished with up to 81% yield, 93:7 er, 4 :1 site selectivity, and > 25 : 1 chemoselectivity. These results exemplify the merits of incorporating new metals into PIX enzymes in order to achieve reaction outcomes previously not achieve using natural enzymes and sets the stage to discover suitable enzymes to catalyze a wider range of C-H amination reactions than has been catalyzed by enzymes in previous work.

6.1 Introduction

Cytochrome P450 enzymes (P450s), metalloproteins containing Fe-porphyrin (Fe-PIX) cofactors, are important biocatalysts that perform oxidation reactions that are involved in various metabolic pathways. For example, Fe-P450s catalyze the oxidation of C-H bonds to alcohols with high stereo-, regio-, site-, and chemoselectivies to form exclusively single products.[1] This level of specificity in C-H oxidation reactions is unrivaled by any small molecule oxidation catalyst. In fact, the specificity and selectivity of P450s is not matched by any synthetic catalyst for any C-H functionalization reaction.[2] However, while enzymes catalyzing C-H oxidation reactions are commonplace, few natural enzymes are known for the analogous transformations of C-H bonds to C-N bonds.[3] Therefore, to exploit the selectivity of enzymes to formally accomplish C-H amination reaction, an alternative strategy is required.

One strategy is to use an enzyme cascade involving C-H oxidation and reductive amination.[4,5] While certain elegant examples of this type of tandem process have been reported,[4] this strategy is predicated on the challenging development of two or more selective enzymes to

produce a single product. As an alternative, more direct approach, recent research has sought to create enzymes for C-H amination reactions by directed evolution of Fe-PIX proteins.[6-8] Though Fe-PIX enzymes react naturally with molecular oxygen to form Fe-oxo compounds, porphyrins containing Fe and other metals (such as Mn, Co, Ru, and Rh) are known to activate azides to form isoelectronic metal-nitrene complexes (Fig. 1). Although Fe-porphyrins are not the most reactive metal complexes among synthetic porphyrin catalysts, distinct mutants of Fe-P450-BM3 have been found to activate sulfonyl azides to form [Fe]-nitrenes, which then insert the nitrene units into C-H bonds. The resulting intramolecular C-H amination reactions occur with up to 54% yield and 89 : 11 er, but the yields of the formation of C-H amination products are inherently limited by poor chemoselectivity of this transformation for C-H amination over nitrenes reduction (Fig. 2).[9-13] In addition to the targeted C-H insertion reactions, the [Fe]-nitrene intermediate can undergo reduction to a primary sulfonamide (Fig. 2). In all reports of P450 catalyzed C-H amination, the yield of this byproduct is comparable to or greater than that the product of the C-H nitrene insertion. Given this observation in the case of intramolecular C-H amination reactions, the propensity of Fe-nitrenes to undergo the competitive side-reactions stands as a major impedance to the development of further intramolecular or more challenging intermolecular transformations.

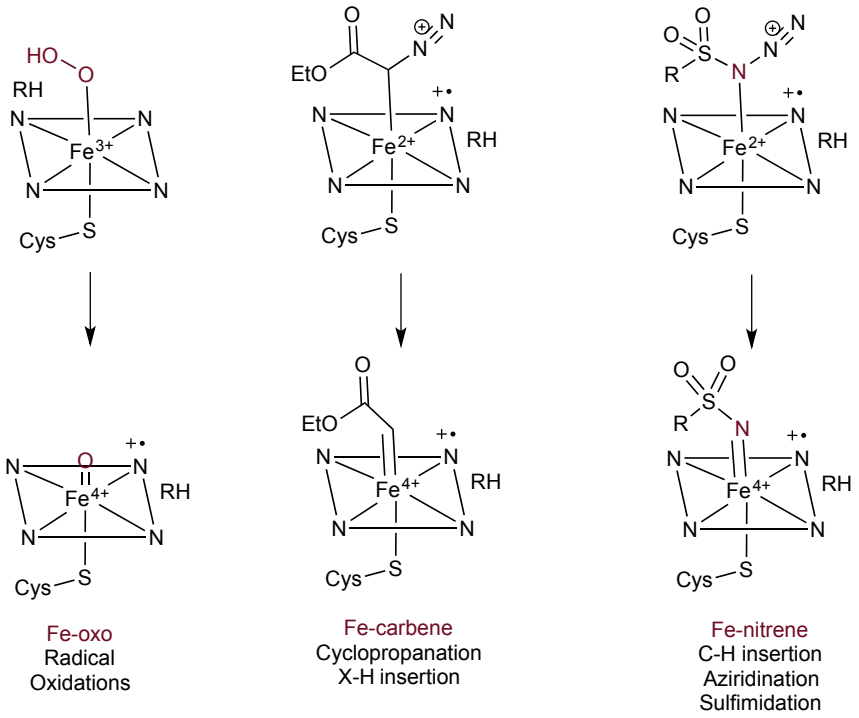


Figure 1. Activation of molecular oxygen, diazo esters, and sulfonyl azides by metal-porphyrins, forming Fe-oxo, Fe-carbene, and Fe-nitrene complexes, respectively.

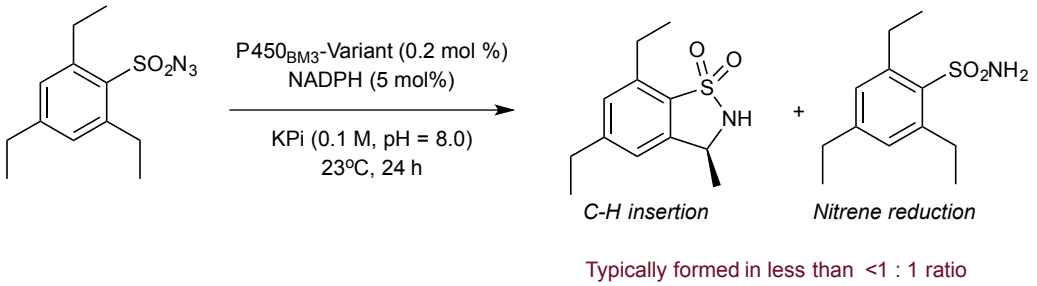


Figure 2. Typical chemoselectivity of C-H amination reactions catalyzed by Fe-PIX enzymes. The most chemoselective mutants form the product due to C-H insertion in approximately the same yield as the product of nitrene insertion.

In our previous studies, described in chapters 3-5, we reported that apo-(PIX)-proteins can be readily prepared and reconstituted with various metallo-mesoporphyrin-IX (M-PIX, M = Co, Cu, Rh, Ru, Ir etc.), and that the reactivity profiles of these artificial PIX proteins are distinct from those of the analogous Fe-containing heme proteins.[14] Given these results, we considered that the incorporation of new metals into heme proteins could create artificial enzymes that catalyze C-H amination reactions with chemoselectivity for the insertion of metal-nitrenes into C-H bonds over the competing pathway leading to the sulfonamide side product that is higher than the chemoselective observed in analogous reactions catalyzed by Fe-P450s . Many traditional metal porphyrin catalysts are known to catalyze C-H amination reactions of sulfone azide substrates, but the reactions usually occur at elevated temperature.[15-17] Therefore, we used the protein CYP119,[18] a thermally stable P450, for the construction of artificially metallated heme enzymes for C-H amination, which enables the use of such metalloproteins at elevated temperatures.

Here we report that Ir(Me)-PIX containing CYP119 variants form active and selective catalysts for the insertion of nitrenes into C-H bonds with > 25 : 1 chemoselectivity for C-H insertion over nitrene reduction. These C-H insertion reactions are accomplished with up to 67% yield and 240 TON. Prospective mutants for this transformation were evaluated in crude lysate by rapid, in situ formation of the Ir(Me)-PIX enzymes, allowing efficient evaluation of libraries of variants for desired modes of selectivity. We show that the evaluation of Ir(Me)-PIX CYP119 variants in this fashion led to mutants affording either enantiomer of the products of C-H insertion reactions with up to 93:7 er. Mutants for site selective C-H insertion to form 6-membered sultams over 5-membered sultams with up to 4:1 site selectivity were also identified. Together, these results exemplify the potential of artificially metallated heme proteins to form catalysts with distinct and favorable reactivity profiles for the development of enzymes for C-H amination reactions.

6.2 Comparison of Synthetic Metal Porphyrins as Catalysts for C-H Amination Reactions Under Aqueous Reaction Conditions

To create an artificial P450s for chemoselective C-H amination, we first assessed the reactivity of a set of free metallo-porphyrins IX (M-PIX, Fig. 3) for the model reactions to convert sulfonylazides **1** and **2** into sultams **3** and **4** (Fig. 3). These model substrates require

insertion of the nitrene into either tertiary or secondary C-H bonds respectively. The metal-PIX complexes containing Fe, Cu, and Mn either did not react with the sulfonyl azide or reacted with chemoselectivity strongly favoring the sulfonamide product over the sultam product. Reactions of sulfonyl azide **1** in the presence of the metal-PIX complexes containing Co, Ru, or Rh formed preferentially the sultam over the sulfonamide product with modest chemoselectivity. However, the reactions of sulfonyl azide **2**, containing stronger, secondary C-H bonds, catalyzed by the same metal porphyrins formed predominantly sulfonamide over sultam.

Although porphyrins containing iridium have not been reported for the insertion of nitrenes into C-H bonds,[19] Ir(Me)-PIX was found to be the most active and the most chemoselective catalyst from the series of tested M-PIXs for the formation of C-H insertion products under aqueous conditions, producing the sultam over the sulfonamide with > 10:1 chemoselectivity for both substrates (Fig. 3). In the presence of Ir(Me)-PIX, substrate **1** reacted to form sultam **3** in 92% yield, with 14:1 selectivity over the formation of the sulfonamide. Under otherwise identical conditions, substrate **2** containing less reactive secondary C-H bonds underwent the reaction to form **4** in 72% yield along with 6% of sulfonamide side product. Upon heating the reaction to 37 °C, the yield for the reaction of substrate **2** was increased to 89% of sultam **4**.

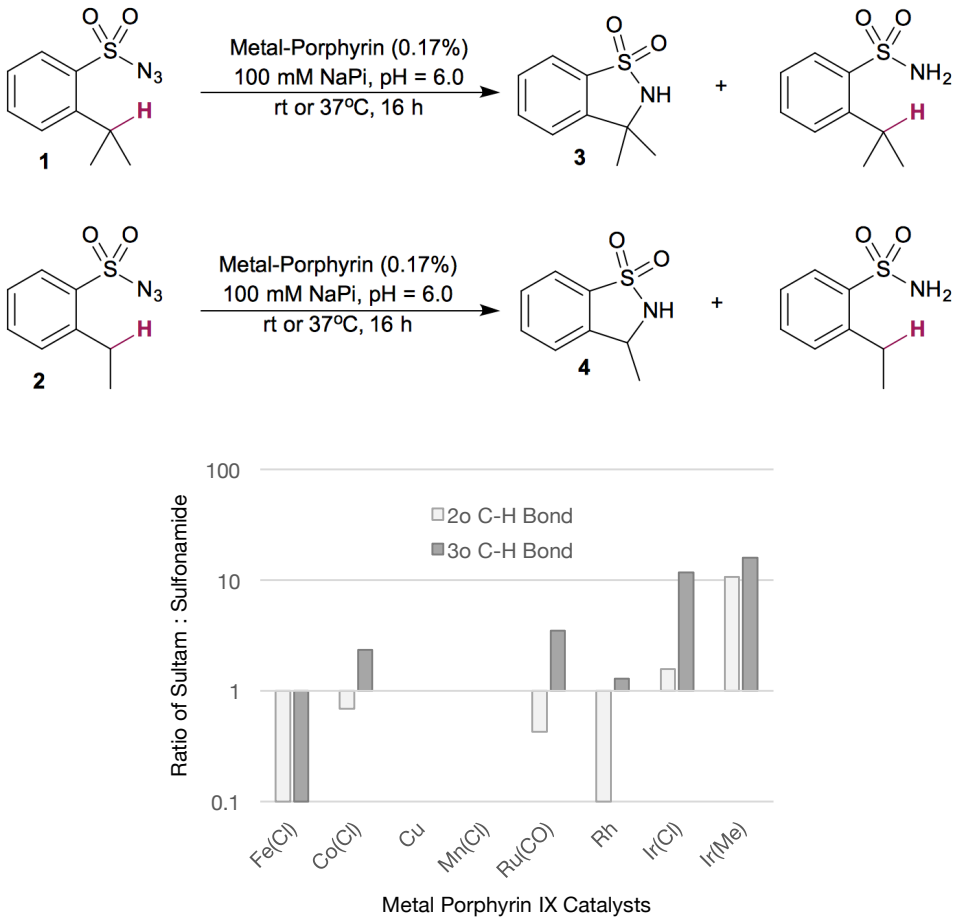


Figure 3. Evaluation of [M]-PIX-IX complexes as catalysts to form **3** and **4** via metal catalyzed C-H amination reactions of substrates **1** and **2**. Top: C-H amination reaction involving the

insertion of tertiary and secondary C-H bonds. The targeted products of the C-H insertion reactions (sultams **3** and **4)** are shown in addition to the side products (sulfonamides) formed from metal-nitrene reduction. Bottom: Outcome of C-H insertion reactions catalyzed by each metal complex. Bars reflect the mol ratio of two products (sultam and sulfonamide) formed, and a comparison of the outcomes in the case of insertions into secondary C-H bonds (substrate **1**, dark gray bars) and tertiary C-H bonds (substrate **2**, light gray bars).

6.3 Application of Ir(Me)-PIX CYP119 Enzymes as Catalysts for C-H Amination

In light of these results, we employed thermally stable P450s containing Ir(Me)-PIX cofactor, which can be used at elevated temperatures, to achieve enzymatic C-H amination. CYP119 from *Sulfolobus solfataricus* is among the most widely studied P450s isolated from a thermophilic organism.[18] To create Ir(Me)-PIX CYP119 enzymes, we followed a protocol analogous to that developed for the creation of Ir(Me)-PIX myoglobins.[14] The apo-form of WT CYP119 was expressed recombinantly in *E. coli* in high yield (10-20 mg/L cell culture) using minimal media without supplementation with any source of iron to limit heme biosynthesis. The apo protein was purified directly using Ni-NTA chromatography, after which the protein was reconstituted by the addition of a stoichiometric amount of Ir(Me)-PIX cofactor, without any need for subsequent purification.

Initially, the WT Ir(Me)-PIX enzyme and the enzyme containing the single mutation C317G to the axial ligand were evaluated as catalysts for a series of intramolecular C-H amination reactions. Reactions catalyzed by these Ir(Me)-CYP119 proteins occurred to form the sultam as the major products (up to 43% yield, 125 TON) with only trace amounts of sulfonamide byproducts, demonstrating the advantage of these artificial enzymes over the natural Fe-containing enzymes.

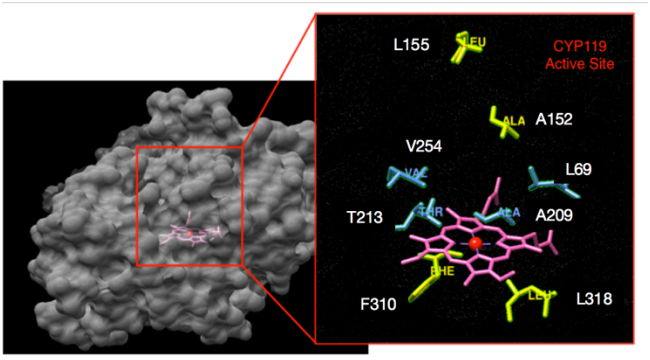


Figure 4. Structure of WT Fe-CYP119 (image prepared in Chimera from PDB 1IO7). Left: Complete structure of Fe-CYP119. Right: Active-site residues targeted in directed evolution of the protein scaffold to increase activity and selectivity in C-H insertion reactions.

To identify Ir(Me)-CYP119 enzymes with higher activity and selectivity in C-H amidation, we designed a library of variants carrying mutations at different positions within the active site (Fig. 4). Eight sites of the P450 scaffold were targeted for mutagenesis (Fig. 4), and a series of 144 mutants containing between 1 and 3 mutations at these sites were evaluated in the model reaction of **2** to form **4**. To enable rapid evaluation of Ir(Me)-PIX CYP119s, we established a screening protocol to evaluate variants directly in the cell lysate. Using this method,

we evaluated 144 mutants of Ir(Me)-CYP119 within three days. The reactions were conducted at 37 °C. Many of the mutants evaluated were inactive; however, the screening identified several variants that did form product **4** in an enantioselective fashion. The mutant T213G, V254L, F310L formed product **4** with 84:16 er, while the mutant T213A, A152L formed predominantly the opposite enantiomer of the product (26:74 er).

A limited set of mutants that formed **4** with the highest enantioselectivity were then evaluated as purified enzymes in the reaction to form **4** and in reactions of additional sulfonyl azides to form products **5-7**. These experiments revealed that selective mutants identified by screening using crude lysates formed product **4** with similar or higher selectivity when used as purified enzymes. In particular, the mutant C317G, T213G, V254L, F310G functioned as the most active and selective catalyst for the formation of **4-7**. Under optimized conditions, this mutant catalyzed the formation of various sultams with up to 93:7 er, 200 TON, 67% yield, and >25:1 chemoselectivity (Fig. 5).

The reaction of product **7** can occur by insertion of the nitrene into either alpha C-H bonds or beta C-H bonds forming either five- or six-membered sultam products (Fig. 5). The free Ir(Me)-PIX cofactor forms the five-membered product with a slight preference over the six-membered product (60 : 40 ratio). The mutant T213G, V254L, F310L catalyzes the reaction with the opposite site selectivity controlled by the protein scaffold, forming the six-membered product **7** with 4:1 site selectivity and with 83:17 er, while producing less than 1% yield of the free sulfonamide byproduct.

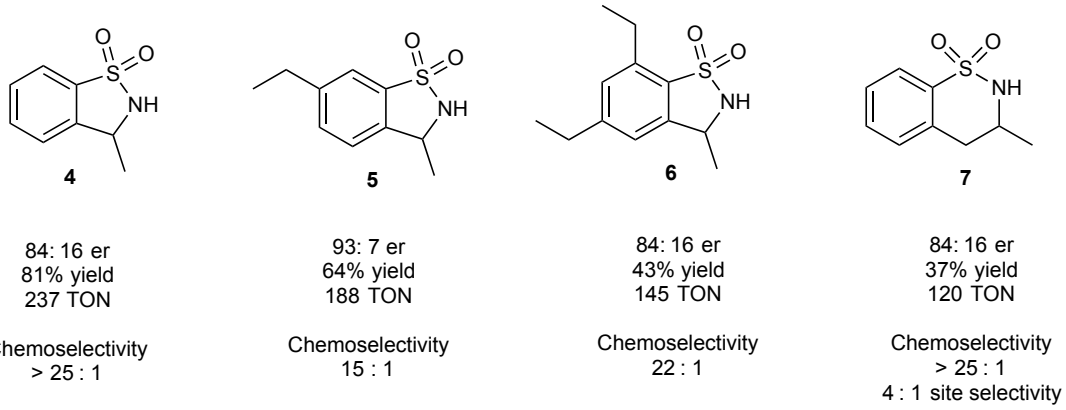


Figure 5. Outcome of C-H insertion reactions of nitrenes using various substrates (**4-7**). Conditions: 0.17% Ir(Me)-PIX CYP119 (mutations as shown), 10 mM substrate, 0.5 mL solvent (100 mM NaPi, 100 mM NaCl, pH = 6.0 containing 2 vol% DMF). Reaction time = 66 hours. Chemoselectivity refers to mol ratio of sultam formed in comparison to sulfonamide formed. Yield and TON refer to the formation of the sultam product

6.4 Conclusions and Outlook

Artificially metallated heme proteins offer the potential to function as evolvable catalysts with reactivity profiles distinct from those of natural enzymes. Here, we show that Ir(Me)-PIX CYP119 enzymes are the first enzymes known to catalyze C-H amination reactions with high chemoselectivity for insertion of nitrenes over reduction to the sulfonamide. Although Ir-containing porphyrins have not been reported to catalyze C-H amination

reactions,{Uchida:2014ja} Ir(Me)-PIX enzymes furnish sultams from sulfonyl azides in up to 82% yield, 93 : 7 er, and with 237 TON, while giving only traces of the sulfonamide byproducts typically observed in substantial amounts in the same reactions catalyzed by Fe-PIX enzymes. Variants displaying these favorable selectivities were identified rapidly by screening mutants as cell lysates, rather than as purified enzymes. These enzymes display rate enhancement for the insertion of nitrenes into secondary C-H bonds compared to the same reactions catalyzed by the free Ir(Me)-PIX cofactor. Finally, discrete variants of the enzyme differentiate among multiple, similarly reactive C-H bonds present in a substrate to achieve C-H amination reactions in a site-selective fashion. Together, these results exemplify the merits of incorporating new metals into PIX enzymes in order to achieve reaction outcomes previously not achieved using natural enzymes.

6.5 Experimental Details

6.6.1 Protein Expression, Purification, and Characterization

a. General Methods

Unless otherwise noted, the chemicals, salts, and solvents used were reagent grade and used as received from commercial suppliers without further purification. All expression media and buffers were prepared using ddH_2O (MilliQ A10 Advantage purification system, Millipore). All expression media were sterilized using either an autoclave (45 min, 121^oC) or a sterile syringe filter (0.22 um). To maintain sterile conditions, sterile materials and *E. coli* cells were manipulated near a lit Bunsen burner.

b. Genes and Cloning

The WT CYP119 gene cloned into the vector 2BT (6xHis-TEV-ORF; AddGene #29666) was purchased from GenScript with codon optimization for *E. coli* (Table S1). The genes for WT P450 BM3 and P450-CAM were obtained as a gift from Prof. Humin Zhao (University of Illinois) and cloned to the vector 2BT at the QB3 Macrolab at UC Berkeley.

c. Media Preparation

Preparation of optimized minimal expression media: Salts (15 g Na_2HPO_4, 7.5 g K_2HPO_4, 0.3 g NaH_2PO_4, 0.3 g KH_2PO_4, 1.5 g NaCl, 5 g NH_4Cl) were dissolved in 2 L ddH_2O and autoclaved to give a media with pH ~8.0 - 8.2. Solutions of glucose (20%), casamino acids (BD Company, low Fe, 20%), and $MgSO_4$ (1 M), were autoclaved separately. Solutions of ampicillin (100 mg/ mL) and $CaCl_2$ (1 M) were sterilized by syringe filter. The following amounts of the listed solutions were added per 2 L of sterile salt solution: 40 mL glucose, 20 mL casamino acids, 4 mL $MgSO_4$, 100 uL $CaCl_2$, 2 mL ampicillin. Stock solutions were stored for several weeks; prepared media was stored for less than 1 day. Minimal media plates were prepared from the same media with the addition of 17 g agar/L media. In this case, agar was autoclaved in 1 L

ddH_2O, and salts were autoclaved separately as a 20X solution, after which they were added to the agar solution.

d. Mutagenesis

Site-directed mutagenesis was performed using the QuickChange Lightening mutagenesis kit (Agilent); requisite double stranded DNA primers were designed according to the Agilent Primer Design Program and purchased from Integrated DNA Technology. PCR reactions were performed according to the manufacturer's directions. PCR reactions contained: 5 uL reaction buffer, 34 uL ddH_2O, 1.5 uL QuickSolution, 1 uL plamids (50 ng/uL), 1.25 uL sense primer (100 ng/uL), 1.25 uL antisense primer (100 ng/ uL), 5 uL dNTPs (2 mM/base), and 1 uL polymerase.

PCR Program: Phase 1 (1 cycle): 95 °C, 1.5 min; Phase 2 (18 cycles): 95 °C, 20 sec, 60 °C, 10 sec, 68 °C, 4.5 min; Phase 3 (1 cycle): 68 °C, 3 min; Phase 4 (storage): 4 °C

DNA Isolation and Storage: Following the completion of the above set of PCR procedures, 1.5 uL DPN 1 was added to each reaction, and the reactions were further incubated (3 h, 37°C). The crude PCR mixture was used to transform XL-10 Gold Ultracompetent cells (45 L cells, 2 L PCR reactions). The mixture was incubated on ice (30 min), heat shocked (30 s, 42 °C), recovered with SOC media (1 h, 37 °C, 275 rpm), and plated on LB plates. Plates were grown (18 h, 37 °C), and individual colonies were used to inoculate 1 mL rich media cultures, which were grown in 96-well plates (13 h, 37 °C, 300 rpm). DNA was isolated from the 96-well cultures using magnetic bead technology at the UC Berkeley DNA Sequencing Facility. Alternatively, individual colonies were used to inoculate 4 mL rich media cultures and grown overnight (13 h, 37 °C, 300 rpm), and the plasmids were purified using a Qiagen DNA Miniprep kit according to the manufacturer's instructions.

e. Protein Expression

Optimized Expression of Apo CYP119 for production of purified enzyme: BL21 Star competent *E.* coli cells (25 L, QB3 Macrolab, UC Berkeley) were thawed on ice, transferred to PCR tubes, and transformed with the desired plasmid solution (1 L, 50-250 ng/ L). The cells were incubated on ice (30 min), heat shocked (30 sec, 42°C), re-cooled on ice (2 min), and recovered by incubation with SOC media (100 uL, 37 °C, 1 h, 250 rpm). Aliquots of the cultures (20uL) were plated on minimal media agar (expression media supplemented with 17 g agar/L) poured in a standard 24-well array plate and incubated (20 h, 37 °C) to produce approximately 4-40 colonies per plate. Single colonies were used to inoculate starter cultures (1 mL, expression media), which were grown in 24 well plates (4-8 hours, 37°C, 250 rpm) and used to inoculate 100 mL overnight cultures (minimal media, 37° C, 250 rpm). Each overnight culture was used to inoculate 750 mL of expression media, which was further grown (9 h, 37°C, 275 rpm). Expression was induced with IPTG (800 uL, 1M), and expression was allowed to continue for 15 h (30° C, 250 rpm). Cells were harvested by centrifugation (5000 rpm, 15 min, 4° C), and the pellets were resuspended in 20 mL Ni-NTA lysis buffer (50 mM NaPi, 250 mM NaCl, 10 mM Imidazole, pH = 8.0) and stored at -80° C until purification.

f. Protein Purification

Cell suspensions were thawed in a room temperature ice bath, decanted to 50 mL glass beakers, and lysed on ice by sonication (4x30 sec on, 2x2 min off, 65% power). Cell debris was removed by centrifugation (10 000 rpm, 30 min, 4° C), and Ni-NTA (5 mL, 50% suspension per 850 mL cell culture) was added. The lysates were briefly incubated with Ni-NTA on an end-over-end mixer (30 min, rt, 20 rpm) and poured into glass frits (coarse, 50 mL). The resin was washed with Ni-NTA lysis buffer (3 x 35 mL), and the wash fractions were monitored using Bradford assay dye. The desired protein was eluded with 18 mL Ni-NTA elusion buffer (50 mM NaPi, 250 mM NaCl, 250 mM Imidazole, pH = 8.0), dialyzed against Tris buffer (10 mM, pH = 8.0, 12 h, 4° C), concentrated to the desired concentration using a spin concentrator, and metallated within several hours as described in section III. Apo protein was not stored for more than 8 hours.

g. Preparation of Catalysts for Mutant Screening in Cell Lysate

BL21 Star competent *E.* coli cells (25 L, QB3 Macrolab, UC Berkeley) were thawed on ice, transferred to standard PCR tubes fused in a 96-well array, and transformed with the desired plasmid solution (1 L, 50-250 ng/ L). The array of PCR tubes was incubated on ice (30 min), heat shocked (30 sec, 42°C), re-cooled on ice (2 min), and recovered by incubation addition of 100 uL SOC media. The cell cultures were transferred to a standard 96-well tissue culture plate and incubated at 37 °C for 1 h (250 rpm). Aliquots of the cultures (20uL) were plated on minimal media agar (expression media supplemented with 17 g agar/L) prepared in a standard 24-well array plate and incubated (20 h, 37 °C) to produce approximately 4-40 colonies per plate well. Single colonies were used to inoculate starter cultures (1 mL, expression media), which were grown in 24 well plates (3 hours, 37°C, 250 rpm) and used to inoculate 45 mL cultures (minimal media, 37° C, 250 rpm, 250 mL flask) that were further grown for 7 hours under the same conditions. Each culture was used induced with IPTG (50 uL, 1M), and expression was allowed to continue for 15 h (30° C, 250 rpm). Cells were harvested by centrifugation (10000 rpm, 10 min, 4° C), and the pellets were resuspended in 0.5 mL lysis buffer (100 mM NaPi, 250 mM NaCl, pH = 8.0) and stored at -80° C until purification.

Lysate Preparation: Cell suspensions were thawed in a room temperature ice bath, transferred to a deep well, 96-well plate, and lysed on ice by sonication (5x2 sec on, 4x2 sec off, 65% power). Cell debris was removed by centrifugation (4 000 rpm, 30 min, 4° C), and the crude lysate was transferred to the upper chamber of the Slide-A-Lyzer™ MINI Dialysis Devices (ThermoFisher, part # 88405). The lysate was dialyzed for a total of 4 hours against Tric buffer (10 mM, pH = 8), during which time the buffer in the lower chamber was replaced 4 times. During dialysis, the tubes were gently agitated on an orbital shaker operating at rt, 150 rpm. Ir(Me)-PIX CYP119 was generated in situ in the lysate as described in section III.

h. Protein Characterization

Gel Electrophoresis: Protein purity was analyzed by sodium dodecyl sulfate-polyacrylamide (SDS-PAGE) gel electrophoresis using precast gels (polyacrylamide, 10-20% linear gradient, Biorad).

Mass Spectrometry: Apo-proteins were analyzed with an Agilent 1200 series liquid chromatograph connected in-line with an Agilent 6224 time-of-flight (TOF) LC/MS system using a Turbospray ion source.

6.6.2 Organic Synthesis and Characterization

a. General methods and materials

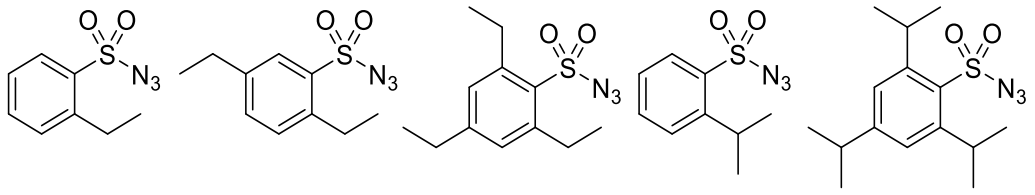
Unless stated otherwise, all reactions and manipulations were conducted on the laboratory bench in air with reagent grade solvents. Reactions under inert gas atmosphere were carried out in the oven dried glassware in a nitrogen-filled glovebox or by standard Schlenk techniques under nitrogen.

NMR spectra were acquired on 400 MHz, 500 MHz, 600 MHz, or 900 MHz Bruker instruments at the University of California, Berkeley. NMR spectra were processed with MestReNova 9.0 (Mestrelab Research SL). Chemical shifts are reported in ppm and referenced to residual solvent peaks[20]. Coupling constants are reported in hertz. Chiral SFC analysis was conducted on a JASCO SF-2000 integrated analytical SFC system. GC analyses were obtained on an Agilent 6890 GC equipped with an HP-5 column (25 m x 0.20 mm ID x 0.33 m film) and an FID detector. High-resolution mass spectra were obtained via the Micro-Mass/Analytical Facility operated by the College of Chemistry, University of California, Berkeley.

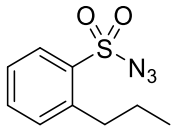
Unless noted otherwise, all reagents and solvents were purchased from commercial suppliers and used without further purification. If required, dichloromethane (DCM) and tetrahydrofuran (THF) were degassed by purging with argon for 15 minutes and dried with a solvent purification system containing a one-meter column of activated alumina; dried and degassed acetonitrile, 1,2-xylene, toluene, N,N-dimethylformamide (DMF), ethanol and methanol were purchased form commercial suppliers and used as received.

b. Synthesis and Characterization of Substrates

The synthetic procedures and characterization of 2-ethylbenzenesulfonyl azide (**3**),[21] 2,5-diethylbenzenesulfonyl azide (**8**),[22] 2,4,6-triethylbenzenesulfonyl azide (**9**)[22], and 2-isopropylbenzenesulfonyl azide (**1**)[13] were reported previously. 2,4,6-triisopropylbenzene sulfonyl azide (**10**) was purchased from a commercial supplier.



2-Propylbenzenesulfonyl azide (11):



2-Propylbenzenesulfonyl azide **11** was prepared following a synthetic procedure developed for analogous compounds:[13] to a stirred solution of 1-bromo-2-propylbenzene[23] (5 g, 25 mmol) in 50 mL of dry THF was added n-butyllithium (12 mL, 2.5 M in hexanes, 30 mmol) dropwise at 78° C, and the reaction mixture was stirred 1 h. Sulfuryl chloride (2.5 ml, 31 mmol) was added at 78° C, the cooling bath was removed and the reaction mixture was stirred overnight at room temperature. Then the reaction was quenched with water (30 mL). The product was extracted with diethyl ether (3 x 50 mL), and the combined organic layers were washed with brine (30 mL), dried over $MgSO_4$ and evaporated. The crude product was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 95:5 gradient) as the eluent. Fractions of the pure 2-propylbenzenesulfonyl chloride were combined, and the solvent evaporated, yielding 500 mg of product as colorless liquid, which was subjected to the next step without further purification.

To a solution of 2-propylbenzenesulfonyl chloride in an acetone : water mixture (40 ml, 1:1 v/v) was added sodium azide (500 mg, 7.7 mmol) at 0 C. The cooling bath was removed and the reaction mixture was let to stir for 24 h. Then the reaction mixture was concentrated to ca. 20 ml. The product was extracted with diethyl ether (3 x 30 mL), and the combined organic layers were washed with brine (30 mL), dried over $MgSO_4$ and evaporated. The crude product was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 95:5 gradient) as the eluent. Fractions of the pure product were combined, and the solvent evaporated, yielding 400 mg (7%, 2 steps) of product as colorless liquid.

^{1}H NMR (500 MHz, $CDCl_3$): d = 8.01 (d, J = 8.0 Hz, 1H), 7.59 (t, J = 7.5 Hz, 1H), 7.42 (d, J = 7.7 Hz, 1H), 7.36 (t, J = 7.7 Hz, 1H), 2.99 – 2.89 (m, 2H), 1.69 (dq, J = 15.0, 7.4 Hz, 2H), 1.00 (t, J = 7.3 Hz, 3H);

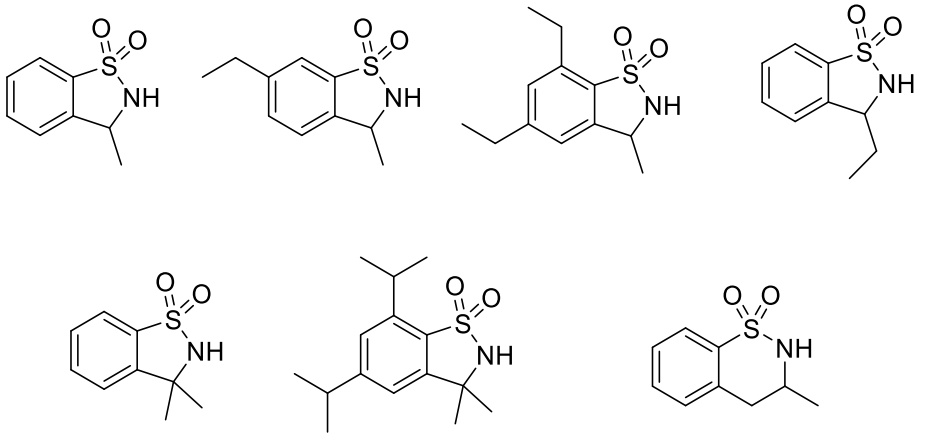
^{13}C NMR (151 MHz, $CDCl_3$): d = 143.38, 136.73, 134.74, 134.71, 132.20, 129.69, 126.56, 35.15, 24.65, 14.26;

HR MS (EI): calcd. for $C_{11}H_{14}O_2$ $[M]^+$: 178.0994, found: 178.0997.

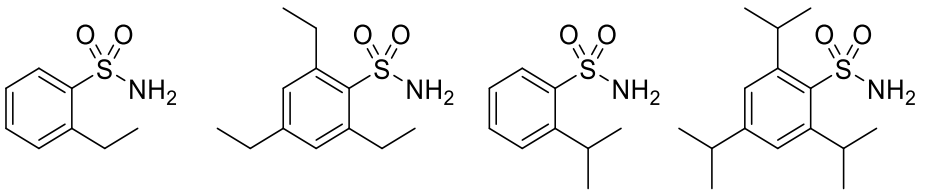
c. Synthesis and Characterization of Authentic Products

General procedure for synthesis of sultams: In a closed vial, a solution of benzenesulfonyl azide (100 mg) and Ir(Me)-PIX (~4 mg) in toluene (6 ml) was stirred at 80 °C. The reaction progress was monitored by TLC. Upon completion (~16 h), the volatile materials were evaporated under reduced pressure, and the residue was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 70:30 gradient) as eluent. Fractions of the pure product were combined, and the solvent evaporated, yielding sultam products. The NMR data match those of the reported molecules: 3-methyl-2,3-dihydrobenzo[d]isothiazole 1,1-dioxide (**4**),[21] 6-ethyl-3-methyl-2,3-dihydrobenzo[d]isothiazole 1,1-dioxide (**5**),[22] 5,7-diethyl-3-methyl-2,3-dihydrobenzo[d]isothiazole 1,1-dioxide (**6**),[22] 3,3-

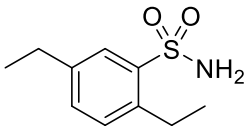
dimethyl-2,3-dihydrobenzo[d]isothiazole 1,1-dioxide (**2**),[13] 5,7-diisopropyl-3,3-dimethyl-2,3-dihydrobenzo[d]isothiazole 1,1-dioxide (**12**),[22] 3-ethyl-2,3-dihydrobenzo[d]isothiazole 1,1-dioxide (**13** – 5-memebred ring),[24] and 3-methyl-3,4-dihydro-2H-benzo[e][1,2]thiazine 1,1-dioxide (**7**– 6-member ring).[25]



General procedure for synthesis of benzenesulfonamide: To a solution of benzenesulfonyl azide (100 mg, ~0.5 mmol) in THF (5 ml) sodium borohydride (100 mg, 2.6 mmol) was added at room temperature. The reaction progress was monitored by TLC. Upon completion (~1 h), the reaction was quenched with water (10 ml). The reaction mixture was concentrated to ca. 10 ml. The product was extracted with ethyl acetate (3 x 30 mL), and the combined organic layers were washed with brine (30 mL), dried over $MgSO_4$ and evaporated. The crude product was purified by column chromatography on silica gel, with a mixture of hexanes and ethyl acetate (100:0 → 70:30 gradient) as the eluent. Fractions of the pure product were combined, and the solvent evaporated, yielding benzenesulfonamide products. The NMR data match those of the reported molecules: 2,4,6-triethylbenzenesulfonamide (**14**),[13] 2-ethylbenzenesulfonamide (**15**), [26]2,4,6-triisopropylbenzenesulfonamide (**16**),[13] and 2-isopropylbenzenesulfonamide (**17**).[27]



2,5-Diethylbenzenesulfonamide (18):



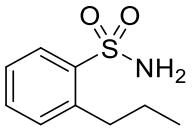
^{1}H NMR (600 MHz, $CDCl_3$): d = 1H NMR (400 MHz, Chloroform-d) δ 7.82 (s, 1H), 7.36 – 7.25 (m, 2H), 3.00 (q, J = 7.5 Hz, 2H), 2.64 (q, J = 7.6 Hz, 2H), 1.28 (t, J = 7.5 Hz, 3H), 1.21 (t, J = 7.6 Hz, 3H);

^{1}H NMR (600 MHz, DMSO-d6): d = 7.70 (s, 1H), 7.40-7.30 (m, 4H), 2.96 (q, J = 7.5 Hz, 2H), 2.63 (q, J = 7.6 Hz, 2H), 1.19 (q, J = 7.2 Hz, 6H);

^{13}C NMR (151 MHz, DMSO-d6): d = 141.70, 141.30, 139.11, 131.25, 130.55, 126.26, 27.62, 24.89, 15.48, 15.40;

HR MS (EI): calcd. for $C_9H_{13}NO_2S$ $[M]^+$: 199.0667, found: 199.0668.

2-Propylbenzenesulfonamide (19):

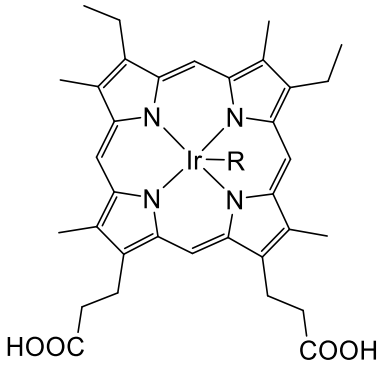


^{1}H NMR (600 MHz, $CDCl_3$): d = 7.96 (d, J = 8.0 Hz, 1H), 7.46 (t, J = 7.5 Hz, 1H), 7.34 (d, J = 7.6 Hz, 1H), 7.26 (t, J = 7.7 Hz, 1H), 4.94 (s, 2H), 2.98 – 2.92 (m, 2H), 1.70 (h, J = 7.4 Hz, 2H), 0.99 (t, J = 7.3 Hz, 3H);

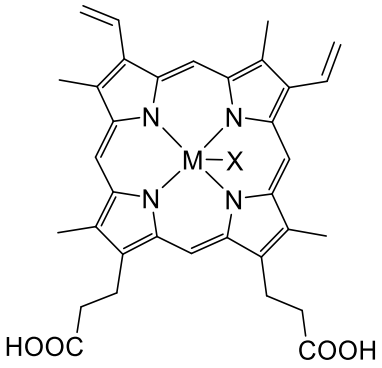
^{13}C NMR (151 MHz, $CDCl_3$): d = 141.68, 139.87, 132.89, 131.46, 128.40, 126.23, 35.15, 24.48, 14.410;

HR MS (EI): calcd. for $C_9H_{13}NO_2S$ $[M]^+$: 199.0667, found: 199.0668.

d. Structure, synthesis and/or commercial source of [M]-porphyrin-IX complexes used in this study



Ir(Me)-PIX: R=CH_3
Ir(Cl)-PIX: R=Cl



Fe(Cl)-PIX: M=Fe, X=Cl
Co(Cl)-PIX: M=Co, X=Cl
Cu-PIX: M=Cu, X= none
Mn(Cl)-PIX: M=Mn, X=Cl
Ru(CO)-PIX: M=Ru, X=CO
Rh-PIX: M=Rh, X= none

The synthetic procedures and characterization of Ir(Me)-PIX and Ir(Cl)-PIX were reported previously.[14] The other cofactors (Fe(Cl)-PIX, Co(Cl)-PIX, Cu-PIX, Mn(Cl)-PIX, Ru(CO)-PIX, Rh-PIX) were purchased from commercial suppliers (Aldrich and Frontier Scientific).

6.6.3 Catalytic Experiments

a. General Methods

Unless otherwise noted, catalytic reactions were performed in 4 mL individually-capped vials. Reactions were assembled in a nitrogen atmosphere glove box. Solutions of catalyst (either crude lysate supplemented with Ir(Me)-PIX-CYP119 or purified Ir(Me)-PIX-CYP119) were gently degassed on a Schlenk line (3 cycles vacuum/refill) before being brought into a glove box in sealed vials. Organic reagents were added as stock solutions in DMF, such that the final amount of DMF in the reaction was approximately 2% by volume. If necessary, protein catalysts were diluted to reaction concentrations in sodium phosphate buffer (100 mM, pH = 6.0) before being added to reaction vials. Unless otherwise noted, all reactions with purified enzymes were performed with catalysts generated from a 1:2 ratio of Ir(Me)-cofactor : apo protein, with 0.34 mol% catalyst loading with respect to metal-cofactor and limiting reagent. All reactions with lysate contained 5.8 uM catalyst. Unless otherwise noted, all reactions were conducted in a shaking incubator (37 °C, 275 rpm).

b. Preparation of Stock Solutions of Catalysts and Method to

The catalyst stock solution generated using purified protein

The Ir(Me)-CYP119 catalyst was prepared by addition of a stock solution of Ir(Me)-PIX (3.1 mM in DMF) to a solution of the apo-protein (0.13mM in 10 mM Tris buffer, pH = 8), such that the resulting solution had a 1 : 2 ratio of Ir(Me)-PIX : CYP119. This ratio was used to ensure that all Ir-PIX was bound. The mixture was incubated for 5 min, and desalted with the NAP-10 desalting column equilibrated with reaction buffer (100 mM NaPi, 100 mM NaCl, pH = 6.0). The protein mixture was diluted to the required reaction concentration with the same reaction buffer.

The catalyst stock solution using lysate

0.5 mL of cell lysate was prepared as described in Section Ie. Upon the last dialysis of the lysate using the Slide-A-Lyzer™ MINI Dialysis Devices (ThermoFisher, part # 88405) the bottom chamber was emptied, and the stock solution of Ir(Me)-PIX cofactor (0.29 mM Ir(Me)-PIX in DMF) was added to the cell lysate in the upper chamber. The mixture was incubated for 5 min. Then, the bottom chamber of the dialyzer was refilled with the reaction buffer (100 mM NaPi, 100 mM NaCl, pH = 6.0) and the tube was incubated on an orbital shaking platform (1 h, 150 rpm, room temp.). The catalyst solution from the upper chamber was transferred to a 4 mL glass vial, the catalyst solution was degassed and brought into a nitrogen atmosphere glovebox as described in Section IIIa.

Procedure for C-H Amination reactions:

500 ul of the catalyst stock solution was added to a vial, followed by addition of a stock solution of the appropriate sulfone azide substrate (2.5 umol in 10 uL DMF). The vial was sealed with a cap, removed from a glovebox and incubated in a shaker (37 °C, 275 rpm, 66h). Upon completion, the reaction mixture was analyzed as described in section IV.

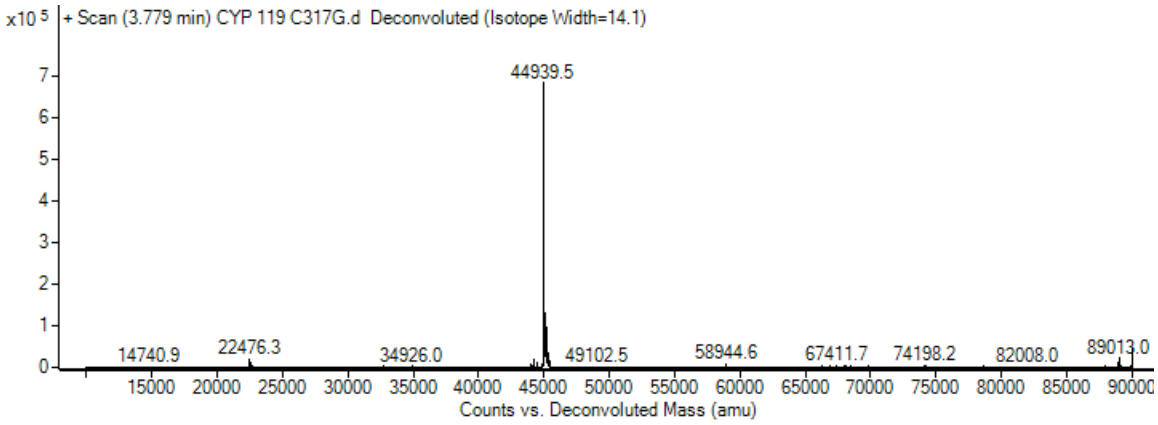
6.6.4 Analysis of Yield and Enantiomeric Ratio

Yields of products and side products, as well as, enantiomeric ratios were determined by chiral SFC using acetanilide as an internal standard. The methods used to analyze reaction mixtures for each substrate are summarized in Table S4. Samples for analysis were prepared as follows: Saturated NaCl (500 uL) was added to each reaction vial, followed by a solution of acetanilide (1 ml, 0.05 ug/ml) in EtOAc. Each vial was capped and the contents of the vial were mixed by shaking. The phases separation was achieved by centrifugation (2 500 rpm, 5 min, rt). The organic layer was removed from the top of the vial by pipet, transferred to a new vial. The aqueous phase was extracted twice with EtOAc (2 x 500 ul). The organic extract of each vial were combined, evaporated and re-dissoved in MeOH (200 ul) for analysis by SFC.

Control extraction studies: To 500 ul of either a catalyst stock solution of Ir(Me)-CYP119 – prepared as in section Vb – or phosphate buffer (100 mM, pH = 6), a stock solution of product **X** (2.5 umol in 7.5 uL MeOH) was added. Saturated NaCl (500 uL) was added to each vial, followed by a solution of acetanilide (1 ml, 0.05 ug/ml) in EtOAc. Each vial was capped and the contents of the vial were mixed by shaking. The separation of phases was achieved by centrifugation (2 500 rpm, 5 min, rt). The organic layer was removed from the top of the vial by pipet, transferred to a new vial. The aqueous phase was extracted twice with EtOAc (2 x 500 ul). The organic extract of each vial were combined, evaporated and redissoved in MeOH (200 ul) for analysis by SFC. These experiments showed that the product was extracted nearly quantitatively from the catalyst solution mixture – Table 16. These experiments confirmed the accuracy of the method.

6.6.5 Supporting Figures

Figure S1. Deconvoluted mass spectrum of CYP119-C317G acquired by LC-MS



6.6.6 Supporting Tables

Table S1. Sequence information for the heme protein used in this study

Amino acid sequence of protein used in this study

Protein	Organism	Construction	Vector	Sequence
CYP 119	*Sulfolobus solfataricus*	6xHis-TEV-CYP119	2BT	EGDIHMKSSHHHHHHENLYFQSNAMYDWFSEMRKKDPVYYDG NIWQVFSYRYTKEVLNNFSKFSSDLTGYHERLEDLRNGKIRF DIPTRYTMLTSDPPLHDELRSMSADIFSPQKLQTLETFIRET TRSLLDSIDPREDDIVKKLAVPLPIIVISKILGLPIEDKEKF KEWSDLVAFRLGKPGEIFELGKKYLELIGYVKDHLNSGTEVV SRVVNSNLSDIEKLGYIILLLIAGNETTTNLISNSVIDFTRF NLWQRIREENLYLKAIEEALRYSPPVMRTVRKTKERVKLGDQ TIEEGEYVRVWIASANRDEEVFHDGEKFIPDRNPNPHLSFGS GIHLCLGAPLARLEARIAIEEFSKRFRHIEILDTEKVPNEVL NGYKRLVVRLKSNE

Table S2. Plasmid library obtained from site directed mutagenesis of the CYP119 gene within the 2B plasmid in the course of directed evolution.

	69X	209X	213X	254X
L		115-C10	116-G2	115-B12
V	115-D1	115-C2	116-F7	
A			116-H3	115-B6
G	115-G4	115-C8	115-F2	115-E8
Y	115-D3		116-D1	115-B8
W		115-C12	116-F1	115-B10
F	115-D10		115-H8	115-B11
T	115-D2			117-A1

T213G + …

	69X	209X	254X	310X	155X	152X
L			115-E3			
V	116-B1					
A			117-E4			
G	115-G2		117-E7			
Y	115-G5		116-E6			140-H7
W	116-A12		115-E2	140-B9		
F	115-H1		115-E7		141-C1	
T	115-G10	115-F7	115-E4		141-B1	

	69L	69V	69A	69G	69Y	69W	69F	69T
213L						116-B8		115-H9
213V		116-B10		116-A11	116-A10		116-B11	
213A		116-B3	116-B6			115-H11	116-B5	
213G								
213Y			116-B7		116-A9			115-H4
213W					116-B4			116-A5
213F			115-H5			115-H2	115-H7	
213T								

	254L	254V	254A	254G	254Y	254W	254F	254T
213L	116-D3					116-C10		116-C4
213V				116-C12				
213A	116-C3			116-E8			116-C6	
213G								
213Y			116-D7					
213W	116-E12				116-D12			116-E3
213F					116-E4	116-E10		
213T								

	209L	209V	209A	209G	209Y	209W	209F	209T
213L		116-F2				116-G6		
213V								
213A								
213G								
213Y								
213W							116-D2	
213F								
213T								

	209L	209V	209A	209G	209Y	209W	209F	209T
254L	117-C3					117-A10		
254V								
254A						117-A3		
254G								
254Y		117-A4				117-B3		
254W	117-C5	117-B11						
254F		117-C2		117-B7				
254T								

	69L	69V	69A	69G	69Y	69W	69F	69T
209L		117-F6		117-G8	117-F5	117-G6		117-H12
209V			117-F9			117-H3		117-H7
209A								
209G		117-H8	117-G4				117-G9	
209Y								
209W		117-H2		117-H6			117-F7	117-F11
209F								
209T			117-G5	117-H4	117-H5	117-H10		117-G1

213G, 254L (115-E3) + Additional Mutation

	69X	310X	318X
L		140-B10	
V	140-B6		
A	140-A3	140-D2	
G	140-B3	140-B8	
Y	140-A4	140-C5	
W	140-B2		
F	140-A8		
T		140-C1	

213G, 254L (115-E3) +

	152X	155X	
L	140-H1		
V	141-A5	141-C10	
A		141-B3	
G	141-A9	141-C3	
Y	140-H2	141-B11	
W	141-A11	141-B2	
F	140-H11	141-D3	
T	140-G9	141-G4	

213G, 69Y (115-G5) +

	69X	310X	318X
L	142-H4		
V			
A			
G		140-C11	
Y	142-H11		
W	142-G7	140-D9	
F			
T	142-H12		

213G, 69Y (115-G5) +

	152X	155X	
L			
V	140-G8		
A			
G			
Y	140-H10		
W	140-H6		
F			
T			

T213A (116-H3) +

	69X	254X	310X
L			166-A12
V			
A			
G		165-H1	
Y		165-H3	166-A6
W			
F		165-H6	166-A9
T		165-H7	166-A7

T213A (116-H3) +

	318X	152X	155X
L		166-C7	
V	166-B11		166-D2
A			
G			
Y	166-B1	166-C4	166-D7
W	166-B6		166-D5
F	166-B7	166-C3	166-D4
T	166-B3		166-D1

254X

	69F, 213G	69V, 213G	
L	141-F6	141-E9	
V			
A		141-E1	
G			
Y	141-D7	141-D9	
W	141-F1	141-E3	
F		141-E7	
T			

69Y, 254L, 213G +

	310X	155X	
L			
V			
A			
G			
Y			
W	140-B9	141-B12	
F			
T		141-B7	

69W, 213G (116-A12) +

	209X	254X	310X
L	165-A4	165-B6	
V	165-A5		165-C11
A		165-B2	165-C10
G	165-A11		165-C1
Y	165-A7	165-B8	
W		165-B3	165-C5
F	165-A2		
T		165-B7	165-C2

69W, 213G (116-A12) +

	318X	152X	155X
L		165-E8	165-F11
V	165-D9		
A			
G	165-D3		
Y	165-D7	165-E5	165-F5
W	165-D2	165-E7	165-F2
F		165-E1	165-F4
T	165-D1		165-F1

Table S3. The SFC methods used to separate enantiomers of the sultam products and sulfonamides.

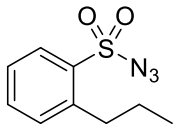
Substrate	Column	Method	Compound	Retention Times
1	Chiracel OD-H (Diacel)	Isocratic: 7% MeOH, 4 ml/min flow	AcNHPh Substrate Sultam Sulfonamide	t(ref)=1.9 min t(sub)=0.7 min t(prd+)=2.25 min t(prd-)=2.65 min t(by-prd)=3.8 min
1	Chiracel OZ-H (Diacel)	Isocratic: 7% MeOH, 4 ml/min flow	AcNHPh Substrate Sultam Sulfonamide	t(ref)=1.5 min t(sub)=0.8 min t(prd+)=2.6 min t(prd-)=2.9 min t(by-prd)=3.6 min
3	Chiracel OD-H (Diacel)	Isocratic: 4% MeOH, 4 ml/min flow	AcNHPh Substrate Sultam Sulfonamide	t(ref)=4.0 min t(sub)=0.8 min t(prd)=3.3 min t(by-prd)=5.2 min
3	Chiracel OZ-H (Diacel)	Isocratic: 7% MeOH, 4 ml/min flow	AcNHPh Substrate Sultam Sulfonamide	t(ref)=1.5 min t(sub)=0.8 min t(prd)=2.5 min t(by-prd)=3.3 min
8	Chiracel OD-H (Diacel)	Isocratic: 7% MeOH, 4 ml/min flow	AcNHPh Substrate Sultam Sulfonamide	t(ref)=1.9 min t(sub)=0.7 min t(prd+)=2.1 min t(prd-)=2.7 min t(by-prd)=2.9 min
8	Chiracel OZ-H (Diacel)	Isocratic: 7% MeOH, 4 ml/min flow	AcNHPh Substrate Sultam	t(ref)=1.5 min t(sub)=1.3 min t(prd+)=3.0 min

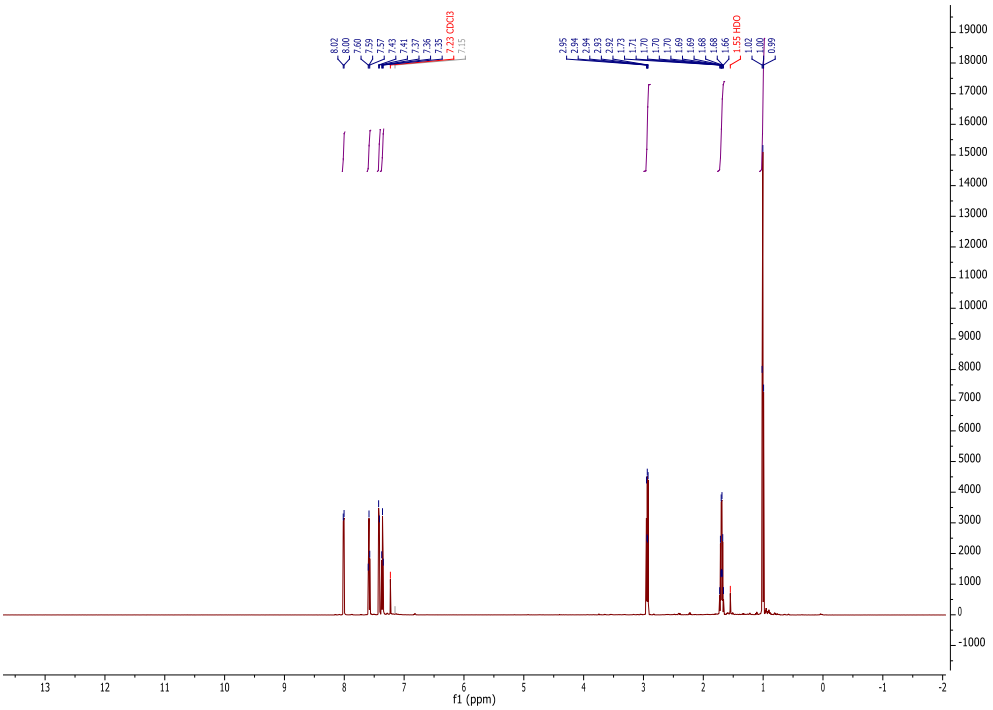
			Sulfonamide	t(prd-)=3.6 min t(by-prd)=3.8 min
9	Chiracel OD-H (Diacel)	Isocratic: 7% MeOH, 4 ml/min flow	AcNHPh Substrate Sultam Sulfonamide	t(ref)=1.9 min t(sub)=0.7 min t(prd+)=2.2 min t(prd-)=2.4 min t(by-prd)=4.6 min
9	Chiracel OJ-H (Diacel)	Isocratic: 4% MeOH, 4 ml/min flow	TsNH2 Substrate Sultam Sulfonamide	t(ref)=4.4 min t(sub)=0.55 min t(prd+)=1.8 min t(prd-)=2.25 min t(by-prd)=3.6 min
11	Chiracel AS-H (Diacel)	Isocratic: 5% MeOH, 2.5 ml/min flow	AcNHPh Substrate 5-Membered sultam (**A**) 6-Membered sultam (**B**) Sulfonamide	t(ref)=2.4 min t(sub)=0.55 min t(prd**A**+)=8.1 min t(prd**A**-)=12.5 min t(prd**B**+)=4.3 min t(prd**B**-)=5.8 min t(by-prd)=7.2 min
10	Chiracel OJ-H (Diacel)	Isocratic: 1.5% MeOH, 4 ml/min flow	AcNHPh Substrate Sultam Sulfonamide	t(ref)=5.6 min t(sub)=0.8 min t(prd)=1.6 min t(by-prd)=3.6 min

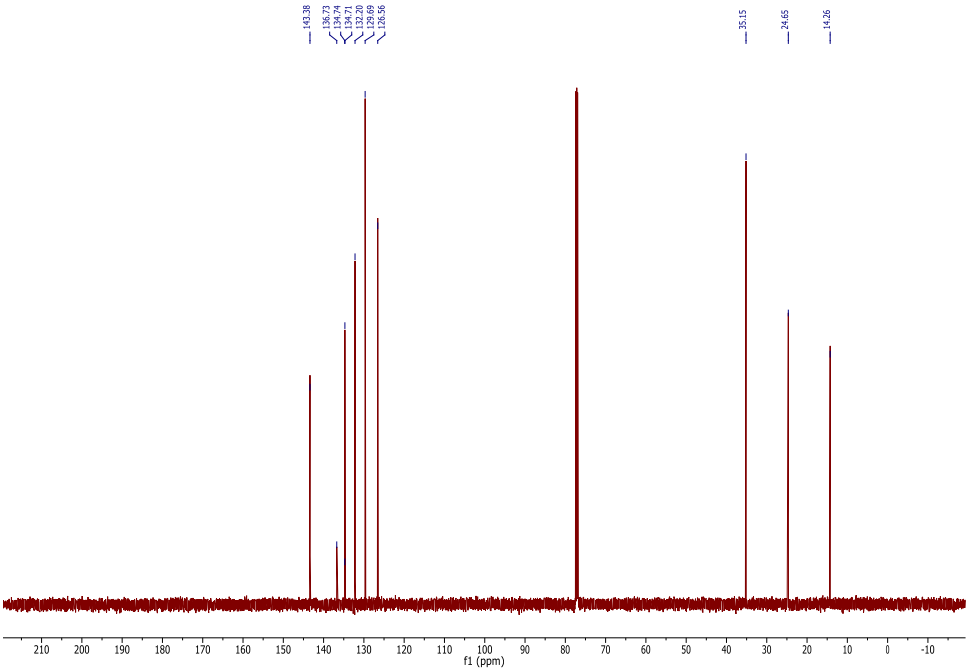
Table 5. Validation of SFC yields: Control studies of the extraction of the known amount of the product from the pseudo-reaction mixture.

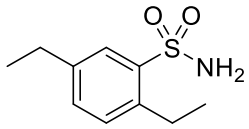
Experiment	**1**	**2**	**3**	**4**	**5**	**6**
Product Extracted	97.9%	96.3%	101.4%	107.7%	113.0%	101.8%

5.1.7 NMR spectra:

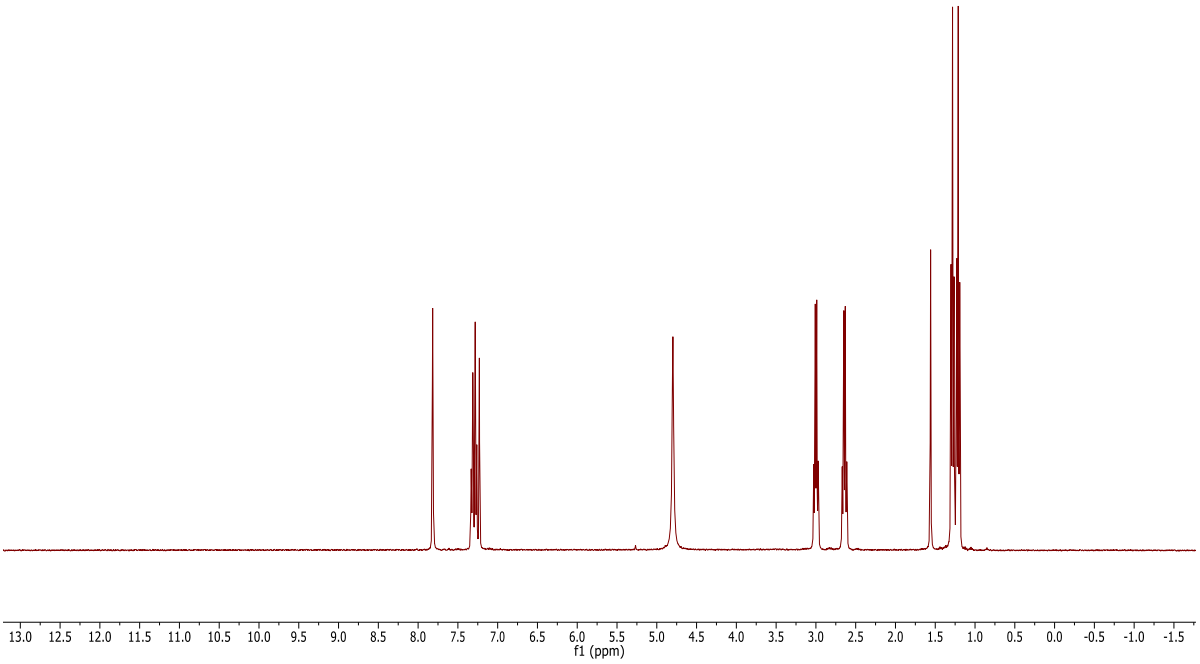




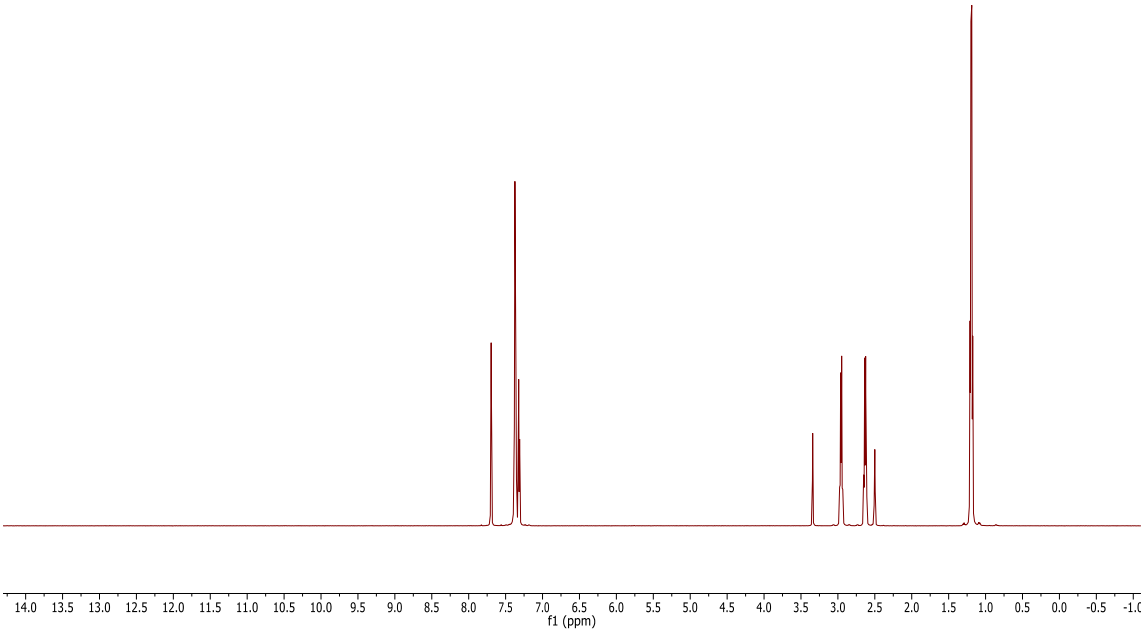


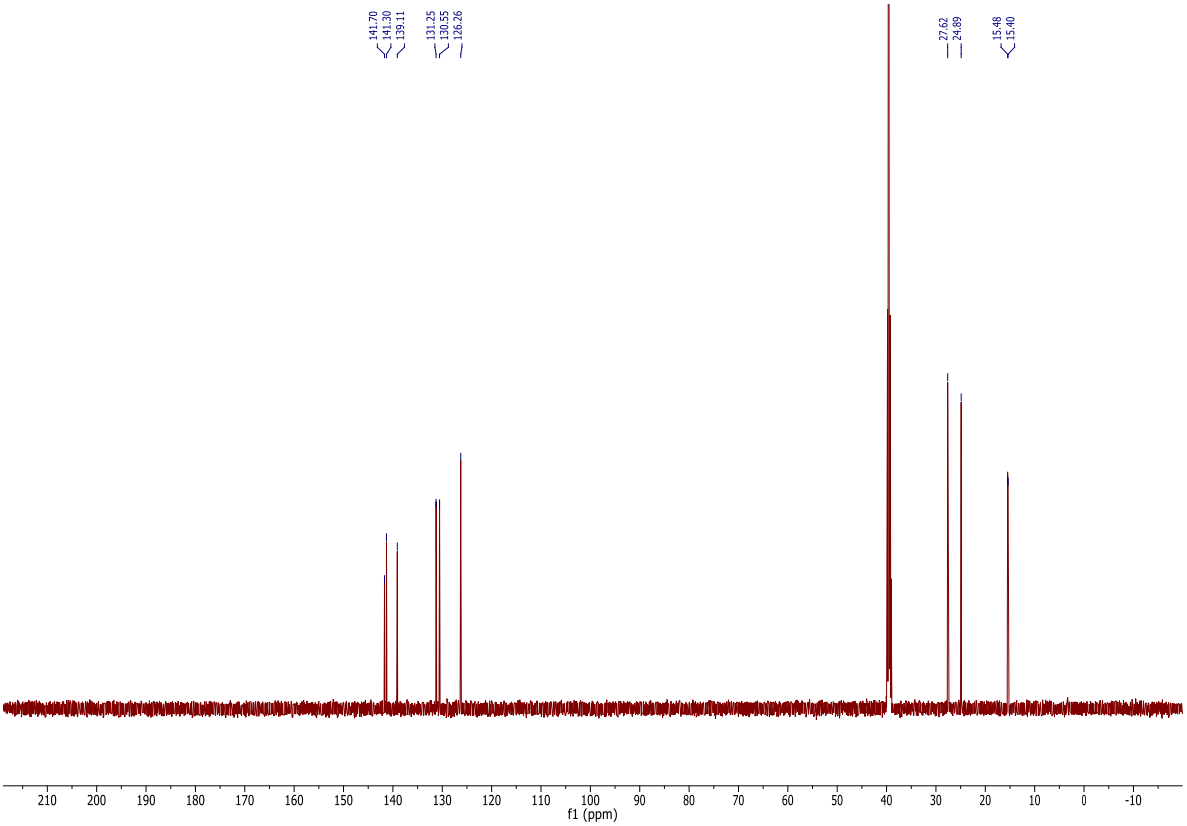


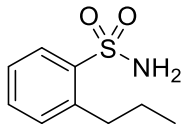
In $CDCl_3$:

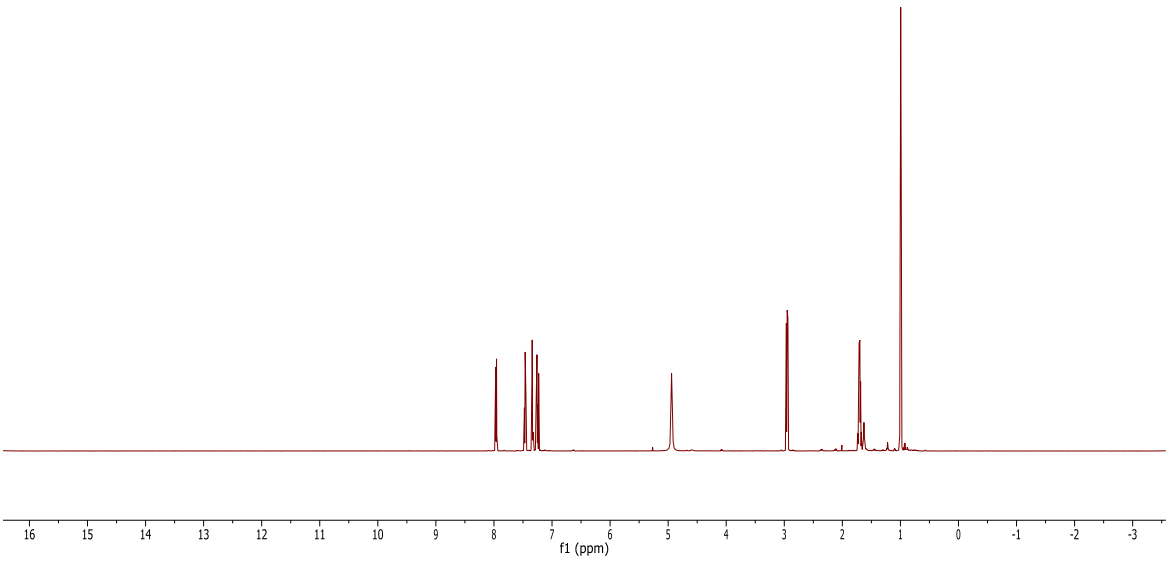


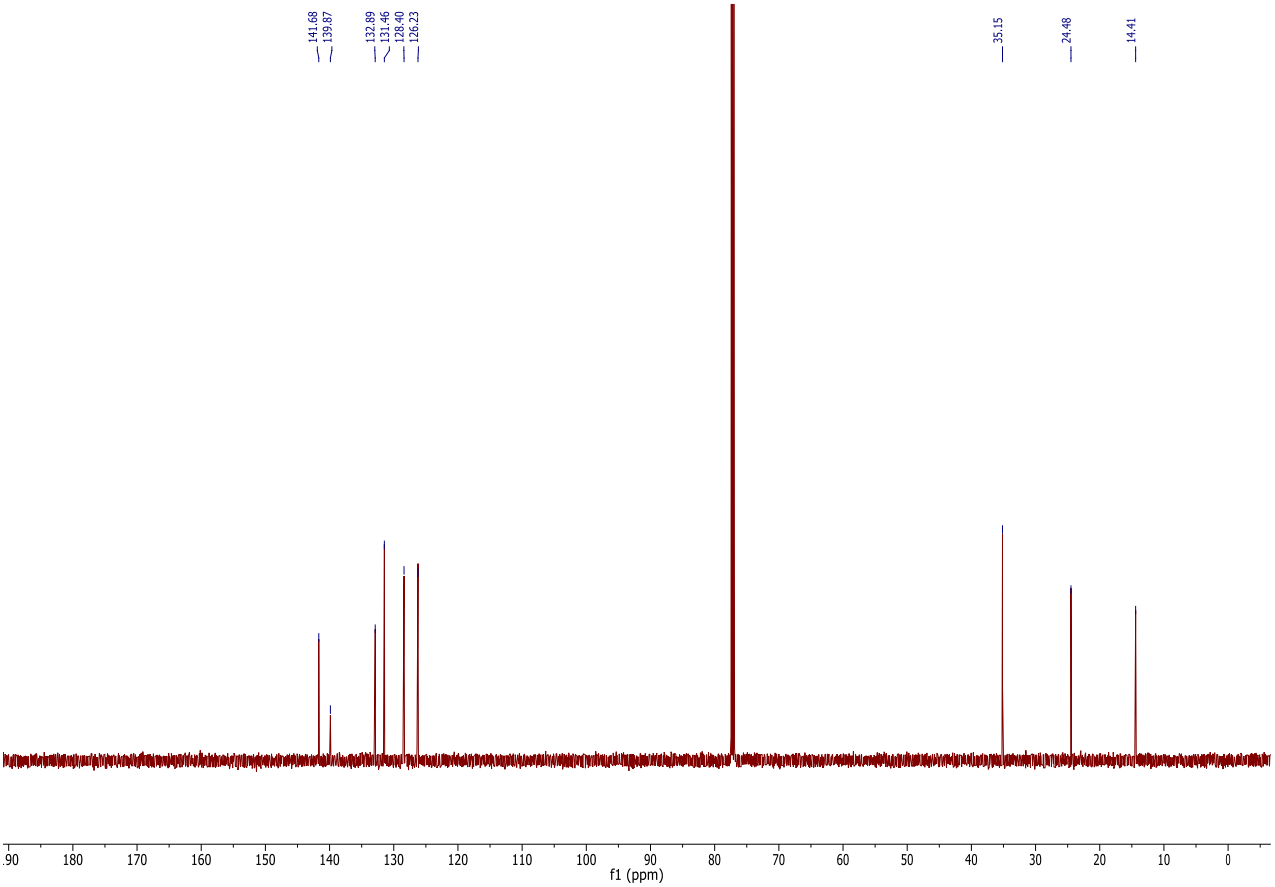
In DMSO-d6:




141.70
141.30
139.11
131.25
130.55
126.26
27.62
24.89
15.48
15.40
210 200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)


O
O
S
NH2


16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
-1
-2
-3
f1 (ppm)


141.68
139.87
132.89
131.46
128.40
126.23
35.15
24.48
14.41
f1 (ppm)

References:

1. *The Ubiquitous Roles of Cytochrome P450 Proteins*. (John Wiley & Sons, Ltd, 2007). doi:10.1002/9780470028155
2. Whitehouse, C. J. C., Bell, S. G. & Wong, L.-L. P450(BM3) (CYP102A1): connecting the dots. *Chem. Soc. Rev.* **41,** 1218–1260 (2012).
3. Barry, S. M. *et al*. Cytochrome P450–catalyzed L-tryptophan nitration in thaxtomin phytotoxin biosynthesis. *Nat. Chem. Biol.* **8,** 814–816 (2012).
4. Mutti, F. G., Knaus, T., Scrutton, N. S., Breuer, M. & Turner, N. J. Conversion of alcohols to enantiopure amines through dual-enzyme hydrogen-borrowing cascades. *Science* **349,** 1525–1529 (2015).
5. O'Reilly, E. & Turner, N. J. Enzymatic cascades for the regio- and stereoselective synthesis of chiral amines. *Perspectives in Science* **4,** 55–61 (2015).
6. Hyster, T. K. & Arnold, F. H. P450BM3-Axial Mutations: A Gateway to Non-Natural Reactivity. *Isr. J. Chem.* **55,** 14–20 (2015).
7. Hyster, T. K. & Ward, T. R. Genetic Optimization of Metalloenzymes: Enhancing Enzymes for Non-Natural Reactions. *Angew. Chem. Int. Ed.* n/a–n/a (2016). doi:10.1002/anie.201508816
8. Lewis, J. C. Artificial Metalloenzymes and Metallopeptide Catalysts for Organic Synthesis. *ACS Catal.* **3,** 2954–2975 (2013).
9. McIntosh, J. A. *et al*. Enantioselective intramolecular C-H amination catalyzed by engineered cytochrome P450 enzymes in vitro and in vivo. *Angew. Chem. Int. Ed.* **52,** 9309–9312 (2013).
10. Hyster, T. K., Farwell, C. C., Buller, A. R., McIntosh, J. A. & Arnold, F. H. Enzyme-Controlled Nitrogen-Atom Transfer Enables Regiodivergent C–H Amination. *J. Am. Chem. Soc.* **136,** 15505–15508 (2014).
11. Singh, R., Kolev, J. N., Sutera, P. A. & Fasan, R. Enzymatic C(sp(3))-H Amination: P450-Catalyzed Conversion of Carbonazidates into Oxazolidinones. *ACS Catal.* **5,** 1685–1691 (2015).
12. Bordeaux, M., Singh, R. & Fasan, R. Intramolecular C(sp(3))H amination of arylsulfonyl azides with engineered and artificial myoglobin-based catalysts. *Bioorg. Med. Chem.* **22,** 5697–5704 (2014).
13. Singh, R., Bordeaux, M. & Fasan, R. P450-catalyzed intramolecular sp(3) C-H amination with arylsulfonyl azide substrates. *ACS Catal.* **4,** 546–552 (2014).
14. Key, H. M., Dydio, P., Clark, D. S. & Hartwig, J. F. Abiological catalysis by artificial haem proteins containing noble metals in place of iron. *Nature* **534,** 534–537 (2016).
15. Uchida, T. & Katsuki, T. Asymmetric nitrene transfer reactions: sulfimidation, aziridination and C-H amination using azide compounds as nitrene precursors. *Chem Rec* **14,** 117–129 (2014).
16. Lu, H. & Zhang, X. P. Catalytic C–H functionalization by metalloporphyrins : recent developments and future directions. *Chem. Soc. Rev.* **40,** 1899–1909 (2011).
17. Chan, K. H., Guan, X., Lo, V. K. Y. & Che, C.-M. Elevated Catalytic Activity of Ruthenium(II)–Porphyrin-Catalyzed Carbene/Nitrene Transfer and Insertion Reactions with N-Heterocyclic Carbene Ligands. *Angew. Chem. Int. Ed.* **53,** 2982–2987 (2014).
18. Rabe, K. S., Kiko, K. & Niemeyer, C. M. Characterization of the Peroxidase Activity of CYP119, a Thermostable P450 FromSulfolobus acidocaldarius. *ChemBioChem* **9,** 420–

425 (2008).
19. Suematsu, H., Kanchiku, S., Uchida, T. & Katsuki, T. Construction of aryliridium-salen complexes: enantio- and cis-selective cyclopropanation of conjugated and nonconjugated olefins. *J. Am. Chem. Soc.* **130,** 10327–10337 (2008).
20. Fulmer, G. R. *et al*. NMR Chemical Shifts of Trace Impurities: Common Laboratory Solvents, Organics, and Gases in Deuterated Solvents Relevant to the Organometallic Chemist. *Organometallics* **29,** 2176–2179 (2010).
21. Ichinose, M. *et al*. Enantioselective Intramolecular Benzylic C H Bond Amination: Efficient Synthesis of Optically Active Benzosultams. *Angew. Chem. Int. Ed.* **50,** 9884–9887 (2011).
22. Ruppel, J. V., Kamble, R. M. & Zhang, X. P. Cobalt-catalyzed intramolecular C-H amination with arylsulfonyl azides. *Org. Lett.* **9,** 4889–4892 (2007).
23. Ruano, J., Aranda, M. T. & Puente, M. Remote stereocontrol by sulfinyl groups: asymmetric alkylation of chiral 2-p-tolylsulfinyl benzyl carbanions. *Tetrahedron* (2005).
24. Enders, D. & Seppelt, M. Catalytic Enantioselective Synthesis of 3-Substituted Benzosultams via Corey-Bakshi-Shibata Reduction of Cyclic N-Sulfonylimines. *Synlett* **2011,** 402–404 (2011).
25. Yu, C.-B., Gao, K., Wang, D.-S., Shi, L. & Zhou, Y.-G. Enantioselective Pd-catalyzed hydrogenation of enesulfonamides. *Chem. Comm.* **47,** 5052–5054 (2011).
26. Instructions for Authors. *Helvetica Chimica Acta* **87,** NA–NA (2004).
27. In vivo and in vitro carbene insertion and nitrene transfer reactions catalyzed by engineered and chimeric heme enzymes.