Development of Cobalt Porphyrin Biocatalysis

by

Lydia Joanna Perkins

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Chemistry)

at the
UNIVERSITY OF WISCONSIN-MADISON
2023

Date of final oral examination: 07/17/2022

The dissertation is approved by the following members of the Final Oral Committee:

Andrew R. Buller, Assistant Professor, Chemistry – Chemical Biology
Judith N. Burstyn, Professor, Chemistry – Chemical Biology
Zachary K. Wickens, Assistant Professor, Chemistry – Organic
Jeffrey D. Martell, Assistant Professor, Chemistry – Chemical Biology

Acknowledgements

Both scientific pursuits and life are group efforts, and I have many people to thank for their help in the completion of this dissertation. Firstly, I would like to thank my advisor, Prof. Andrew R. Buller. Through your guidance I have learned not only how to ask questions and design experiments, but also how to be wrong—gracefully, how to be both a human and a scientist simultaneously, how to be make myself be rigorous, and how to explain my ideas to others clearly and concisely. Thank you for your patience with me for those first few years, and the freedom you gave me to explore in the later years.

Thank you also to the team that started the Buller Lab: Prasanth, Allwin, and Jon. I am thankful for our ability to learn from one another and work together in building the lab's body of research and the group's culture. Thank you to the rest of the Buller Buddies. You all made it a joy to come to work every day, and I learned a lot about both science and myself from working with you. I am extremely proud of the group we have all built together, and am excited to see it continue to grow.

Thank you also to Brian and Judith for their collaboration throughout the years. I have learned a tremendous amount from both of you, and owe virtually 100% of early successes in the lab to Brian's incredible insight and experience.

Thank you to my friends in Madison for being my friends, my support system, and my path to a good time. I am grateful to have met you all and to have had 5 (+) years to get to know you. Thank you for bearing with me.

Lastly, thank you to my family for their support throughout the years, for valuing my education, and for instilling in me a love for exploring the world around us.

Table of Contents

Abstract

Biocatalytic reactions are prized for their selectivity, mild reaction conditions, and high efficiency. However, the scope of biocatalytic reactions readily available to the synthetic chemist is limited by the reactivity of the naturally occurring enzyme cofactors. In this thesis, I explore and develop a method to produce new-to-nature cobalt porphyrin cofactor (CoPPIX) in a common laboratory strain of *E. coli*. This method dramatically improves access to CoPPIX-bearing enzymes and enables the development of CoPPIX-dependent biocatalytic reactions. To this end, we have discovered that the CoPPIX-bearing P450 enzyme CoCYP119 is able to perform metal-hydrogen atom transfer-like reactivity, a reaction which, to our knowledge, is not found in nature. We leverage de novo biosynthesis and incorporation of the required cofactor to engineer CoCYP119 for *O*-deallylation of nitrophenol. We demonstrate that these enzymes may also be engineered for other metal-hydrogen atom transfer reactions.

Chapter 1

Introduction

Chapter 1: Introduction

1. 1. Enzymes as catalysts for organic synthesis

The ability to synthesize small molecules selectively and efficiently is vital to the modern scientific enterprise. Synthetic small molecules aid in the study of biological systems, serve as life-saving therapeutics, enable the construction of multifunctional materials, and demonstrate and enlighten our understanding of the chemical world. Since the first chemical synthesis of Urea in 1828[1], the field of synthetic chemistry has grown enormously. The modern chemist can choose from a plethora of well-understood chemical transformations to assemble an array of strikingly complex small molecules, often on massive scales (Figure 1.1). Nevertheless, significant challenges in synthetic chemistry remain. Many common chemical reactions suffer from subpar selectivity, poor atom economy, low energy efficiency, and can require environmentally toxic reagents. Some bonds and chemical motifs are challenging to make by any means at all.[2] To address these challenges, researchers must continue to develop novel chemical transformations. Among these, catalytic reactions are particularly desirable because they often exhibit high atom economy, excellent selectivity, and can be used to promote challenging reactions under mild conditions.[3] Biocatalytic reactions, in particular, are prized for their high selectivity and efficiency, while operating under mild, aqueous conditions.[4]

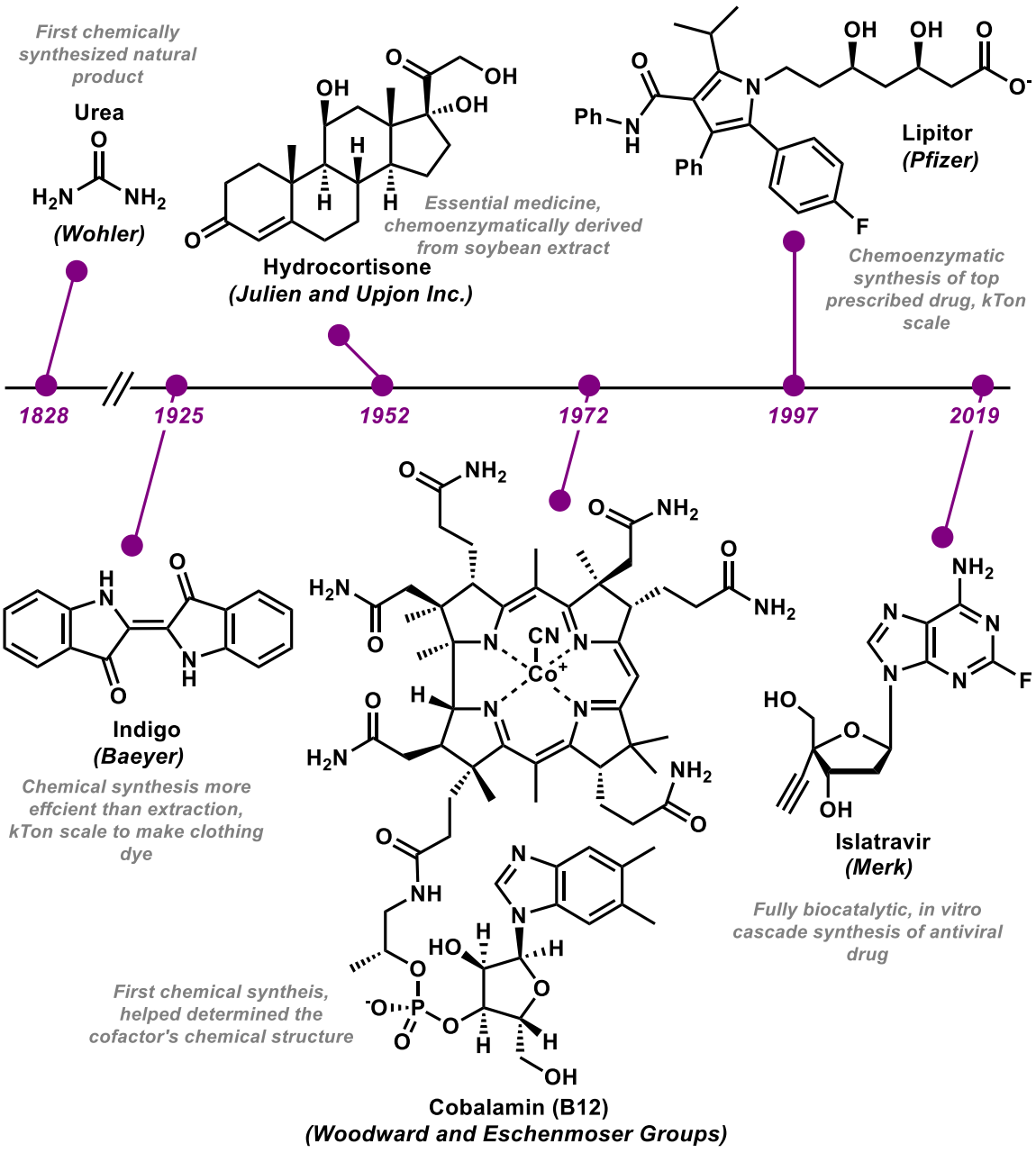


Figure 1.1. Timeline of notable small molecule syntheses over the past two centuries.

Catalysis, specifically *bio*catalysis is essential to life itself. For instance, the conversion of glucose and O_2 to CO_2 and water is the function of central carbon metabolism and is a catalyzed process. The overall transformation is highly exergonic and could be achieved by burning sugar (a non-catalyzed, chemical process). Instead, cells carefully control which reactions happen and in what order, funneling constituent atoms and the energy of each bond into useful biological

building blocks and enabling higher order cellular functions. The catalytic processes of turning food into energy and building blocks (catabolism) and assembling small molecules and biopolymers (anabolism) are ubiquitous across cell types, cellular organelles, and the domains of life. Regardless of location and host organism, enzymes serve as Nature's premiere catalysts for virtually all chemical transformations (with the notable exception of the catalyst for making proteins themselves). Enzymes have several unique features as catalysts, which are highlighted below and in Figure 1.2. These features originate from the role of enzymes in Nature but have numerous implications for the way that enzymes are used for synthesis.

I. **Assembled from biologically-available pieces.** Apart from RNA-based catalysts, all enzymes are sequence-defined protein polymers composed of a relatively restricted set of 20 homochiral α-amino acids. These proteins can be augmented by post-translational modifications (phosphorylation, oxidation, and glycosylation, among others) and cofactors (metals or organic small molecules). Notably, the building blocks for cofactors, proteins, and the other biopolymers must either be harvested from a cell's natural environment, as in the case of metal cofactors and vitamins, or biosynthesized enzymatically.

II. **Evolvable.** The amino acid sequence of every enzyme is encoded by DNA. Thus, the structure of each catalyst is both heritable and mutable. Small changes to catalytic activity, which arise through mutation, can be tested by natural selection. Beneficial mutations, which contribute to the fitness of the host organism, are passed to the next generation. This process of evolution results in catalysts that are highly optimized for their biological context.

III. **Selective.** An enzyme's native environment is an aqueous mixture of small molecules, ions, and the various biopolymers. Life depends on a complex, interconnect cascade of multiple ongoing biosynthetic and catabolic pathways, and enzymes need to function selectively within this context. This complex reaction

environment has multiple implications. Firstly, enzymes can be highly substrate specific such that, within the cellular milieu, they only react with a single substrate. Secondly, as many pathways and enzymes share substrates, each enzyme's activity is optimized such that one enzyme does not entirely consume a substrate which is required for a different pathway. Lastly, the substrates for enzymatic reactions tend to be relatively stable and unreactive with water. Many enzymes kinetically protect highly reactive intermediates from reacting with solvent.

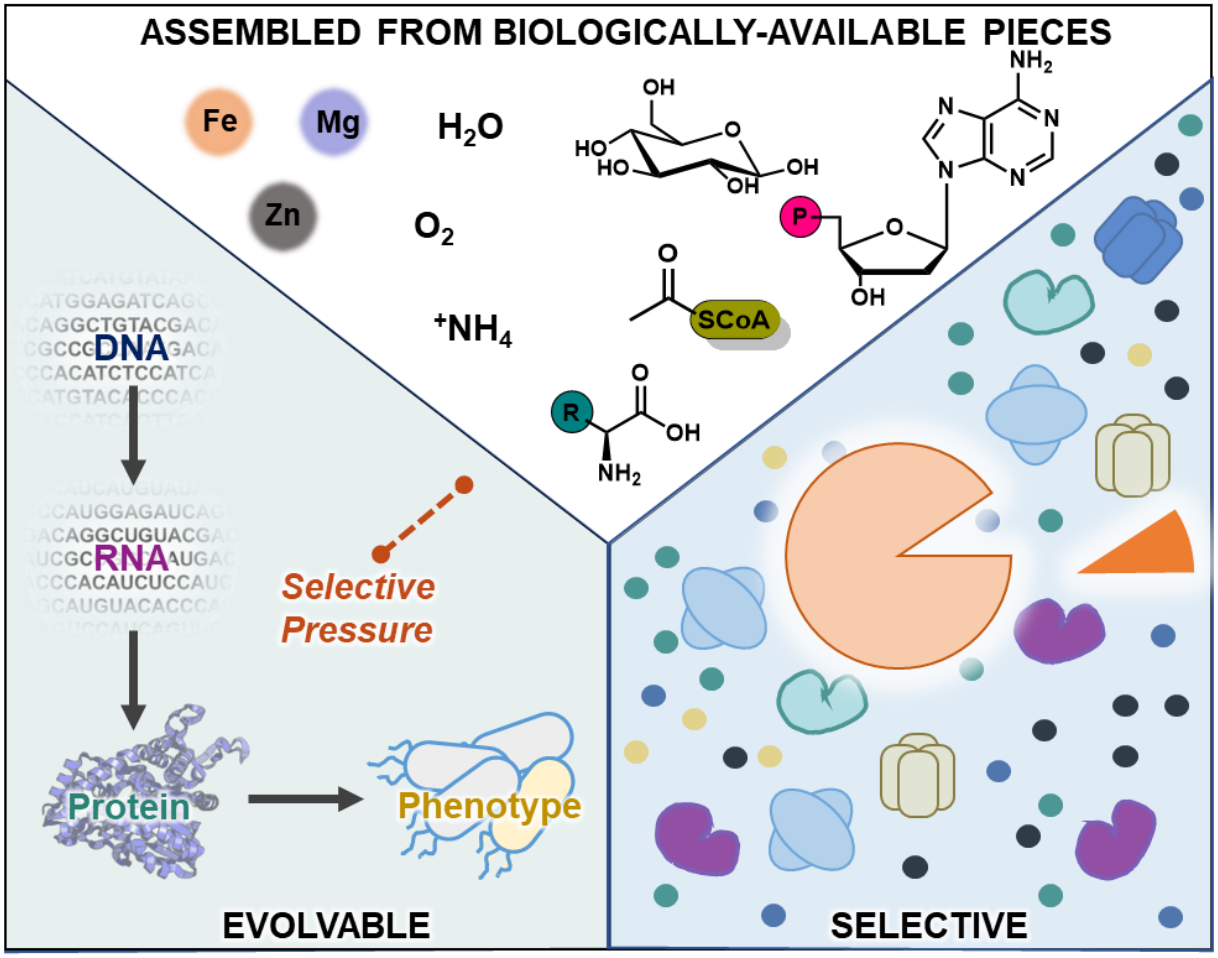


Figure 1.2. Common features shared by biocatalysts.

Biocatalysis technology has been regularly used for centuries. Processes such as the fermentation of beer and wine or the coagulation of milk *en route* to cheese are, at their core, biocatalytic reactions. Today, enzymes are used as catalysts in a variety of contexts, from the

multi-ton conversion of D-glucose to D-fructose by fructose isomerase, to the cascade synthesis of bespoke pharmaceuticals such as Islatravir[5] (Figure 1.1). Biocatalytic reactions are desirable because the catalysts can be obtained by fermentation with inexpensive materials, and the biocatalytic reactions themselves are highly selective, efficient, and environmentally benign. The central features shared by all enzymes, as described above, can both strengthen and stymie the widespread application of biocatalysts to solve synthetic challenges.

1.2. Protein engineering harnesses the evolvability of proteins to improve enzyme function for synthetic purposes.

While the collection of enzymes found in nature can carry out a dizzying array of transformations with high efficiency, there are many desirable chemical transformations for which existing biocatalysts are too inefficient to be synthetically useful. Enzymes have evolved to behave optimally in their natural context,[6] which often differs significantly from conditions that are most useful for synthetic applications. Many enzymes have diminished activity with even close analogs of the native substrate, others are less efficient at high substrate loadings, and many are intolerant of organic co-solvent. In these cases, protein engineering can be used to improve enzyme function in the desired, synthetically relevant context.

At its core, the protein engineering process harnesses the *evolvability* of enzymes to improve their activity.[7–9] The first step of protein engineering is to make changes to the template enzyme, or "parent" (Figure 1.3). This is achieved by making mutations to the gene encoding the parent, enabled by myriad modern molecular biology techniques.[10] Protein engineering strategies can be categorized as either rational, in which specific mutations are chosen, or agnostic, in which mutations are random. Single point mutants can be generated by site directed mutagenesis, typifying the former approach. Error prone PCR[11] and the use of mutagenic plasmids[12] are examples of the later, fully agnostic approach. Most modern protein engineering campaigns are a hybrid of rational and agnostic approaches. For instance, site saturation mutagenesis involves mutating a single amino acid residue that is suspected to affect activity to every other amino acid.[8,13] Combinatorial libraries are generated by mutating one or more residues of interest to a subset of supposed activating residues.[14]

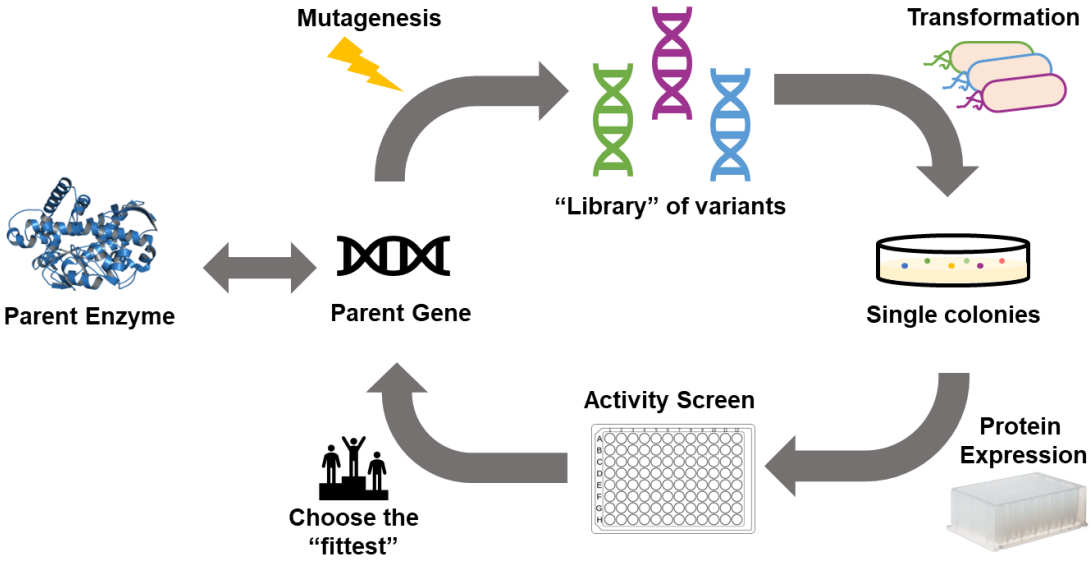


Figure 1.3. The directed evolution cycle.

After designing and generating a library of mutants, the members of the library are transformed into a host, such as *E. coli*, where corresponding variant enzymes can be expressed. The activity of each enzyme is tested, and improved catalysts can be selected. Selection can be engineer-determined (any assay in which the amount of product formed is assessed) or directed autonomously (such as in FACS[15], life-or-death screens, and PACE[16]). Once an improved variant is identified, this variant can be subsequently mutated and screened iteratively, beginning a process called directed evolution (Figure 1.3). [17]

Since its seminal demonstration in 1994,[18] directed evolution has been used to great success to improve enzymes for a variety of transformations. In most cases, these enzymes are engineered to have improved reactivity with non-native substrates (substrate analogs), or to operate under non-native conditions (high temp, organic cosolvent, or high substrate loading).

1.3. Catalytic promiscuity of existing enzyme scaffolds enables new-to-nature enzyme function.

On rare occasions, enzymes can be coopted to perform reactions that are wholly unrelated to their native reaction, termed “new-to-nature” reactivity. Despite Nature’s prolific catalytic repertoire, there still exist many transformations in synthetic organic chemistry for which there is no enzymatic counterpart. In an effort to expand the biocatalytic toolbox and address more synthetic challenges with biocatalysis, researchers have looked for ways to access non-natural reaction pathways and intermediates using enzymatic scaffolds.[19]

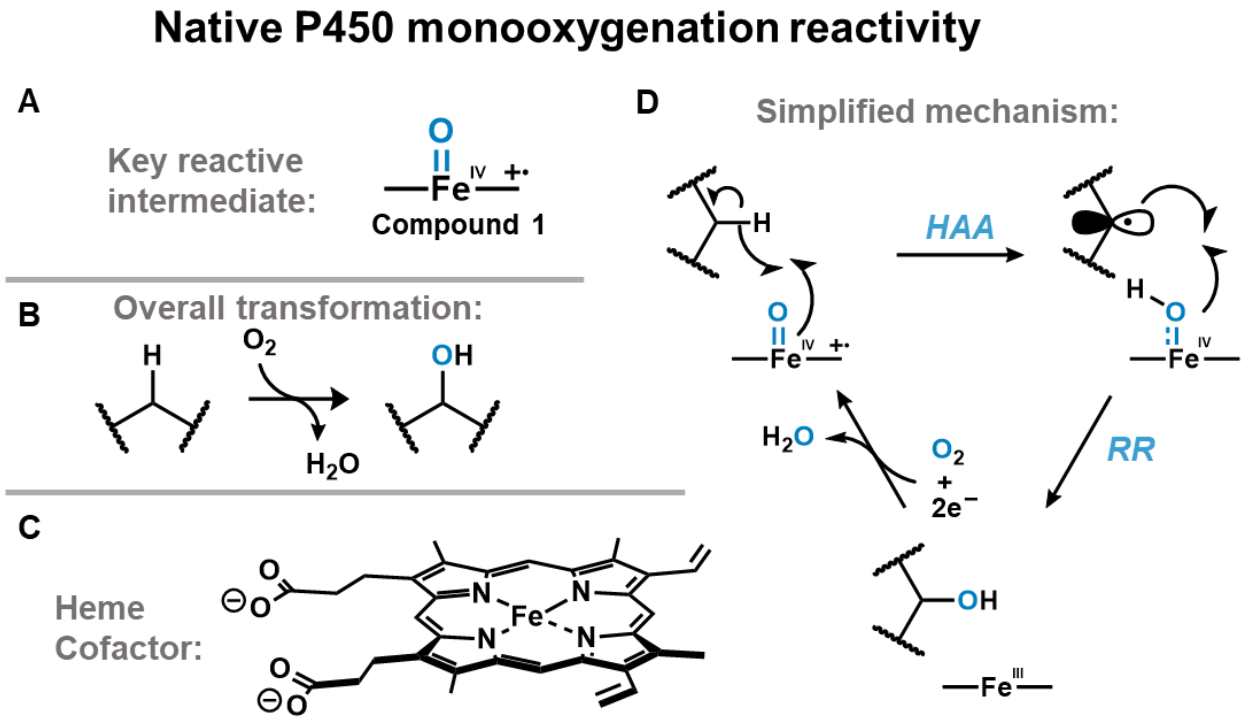


Figure 1.4. The native monooxygenase activity of heme dependent P450 enzymes. A) The key reactive intermediate for P450 monooxygenation is the iron-oxo radical cation intermediate, commonly referred to as Compound I. B) General scheme for the overall transformation catalyzed by P450s. C) The heme b cofactor. D) Simplified catalytic cycle for P450 monooxygenation reactions.

One particularly significant example of this approach is the development of P450 monooxygenase-derived carbene and nitrene transferases.[20] In nature, heme monooxygenases activate molecular oxygen, generating the canonical Compound I intermediate (Fig 1.4A). This intermediate elicits selective hydrogen atom abstraction and radical rebound on a C-H bond of

substrate, generating alcohols (Figure 1.4D). In the 1980s, Gellman, Breslow, and others demonstrated that a non-enzymatic heme analog, Iron-tetraphenylporphyrinate (FeTPP) could catalyze the transfer of nitrene groups into C-H bonds, generating amines.[21,22] The same catalyst was also shown to catalyze a related carbene transfer into alkenes, resulting in cyclopropane products.[23] P450 enzymes, which coordinate the heme cofactor through a cysteine axial ligand, were later shown to catalyze both nitrene[24] (Figure 1.5) and carbene transfer[25] reactions (Figure 1.6). The Arnold group has spearheaded substantial directed evolution of these enzymes, generating variants which can catalyze thousands of turnovers with high rates and selectivity for a variety of substrate classes. Compared to synthetic carbene and nitrene transfer catalysts, these enzymes can provoke otherwise inaccessible regioselectivity,[26,27] reactivity,[26,28] and overall reaction pathway outcomes (see Figures 1.5 and 1.6 for examples).[28,29]

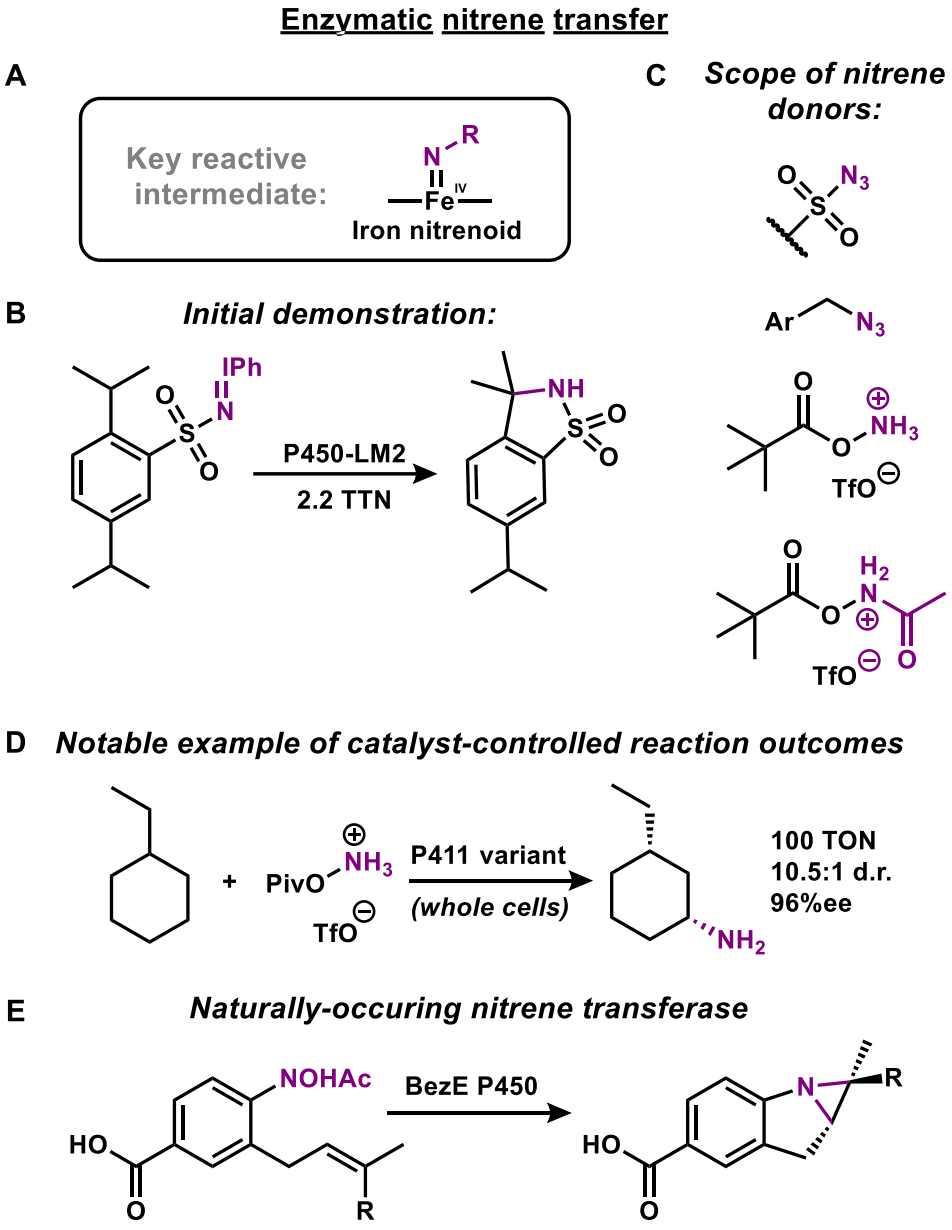


Figure 1.5. Enzymatic nitrene transferase reactivity and selected examples. A) The key intermediate for nitrene transfer catalyzed by P450-derived enzymes is an Iron(IV)nitrenoid. B) Initial demonstration of intramolecular nitrene transfer activity by a P450 enzyme derived from mouse liver.[24] C) Scope of nitrene donors that have been used for enzymatic nitrene transfer reactions to date. D) Example of site and stereoselective nitrene transfer into unactivated C-H bonds by a P411 variant.[26] This transformation is extremely challenging for state of the art synthetic catalysts. E) The only known example of P450-catalyzed nitrene transfer in a naturally occurring biosynthetic pathway. The enzyme BezE catalyzes a key aziridination step in the biosynthesis of benzastatin.[30]

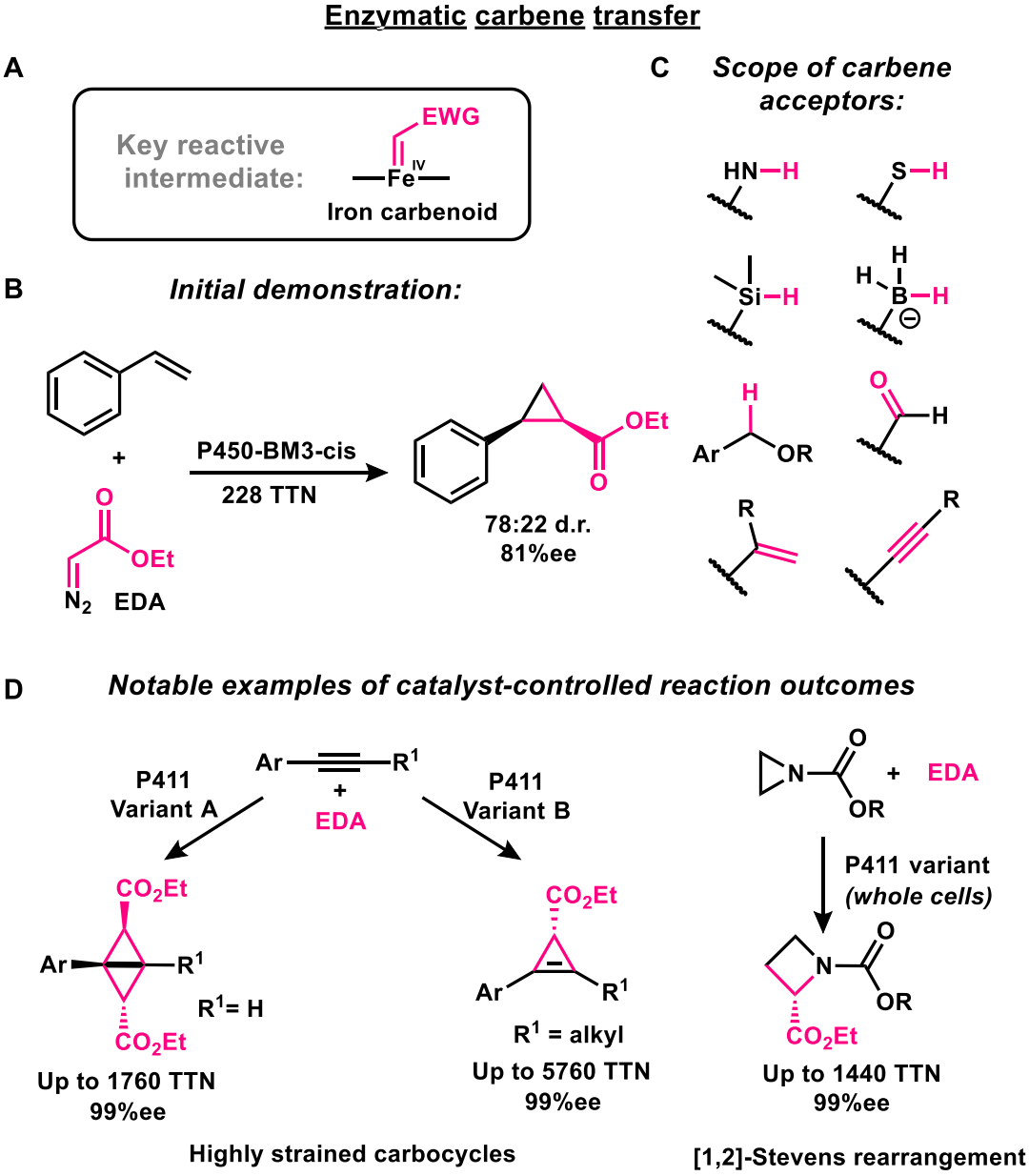


Figure 1.6. Enzymatic carbene transferase reactivity and selected examples. A) The key intermediate for nitrene transfer catalyzed by P450-derived enzymes is an Iron(IV)carbenoid. B) Initial demonstration of carbene transfer reactivity. P450-BM3-cis, originally engineered for improved monooxygenation activity on selected substrates, catalyzed the enantioselective cyclopropanation of styrene with ethyl diazoacetate (EDA) as the carbene donor. C) Scope of bond types into which carbenes have been inserted enzymatically. D) Two notable examples of carbene transferase activity. Two different P411 variants selectively cyclopropanate an activated alkyne either once (variant B) to make a highly strained cyclopropane, or twice (variant A) to make highly strained bicyclobutanes.[28] In another example, a P411 variant is able to elicit a [1,2]-Stevens rearrangement and ring expansion following carbene transfer to the nitrogen atom of an aziridine, furnishing an azetidine .[29] This reaction outcompetes the otherwise highly favorable cheletropic extrusion of ethylene to make a linear imine.

Another demonstration of new-to-nature enzymatic reactivity utilizes nicotinamide (NADH)-dependent reductases. Ketoreductases (KREDs) and imine reductases (IREDs) are widely used in industrial settings for the hydride-mediated reduction of carbonyls and imminiums, to chiral alcohols and amines, respectively, using NADH or NADPH cofactors (Figure 1.7A). These enzyme scaffolds have been demonstrated to have activity on a broad range of substrates and are highly engineerable.[31] In direct analogy to the development of carbene transferases, the new-to-nature mode of reactivity of these enzymes was inspired by the studies of free cofactor models. In 1983 Tanaka and coworkers showed that an analog of NADH, 1-benzyl-1,4-dihydronicotinamide, when irradiated with 350 nm light, reductively dehalogenated benzylic bromides and chlorides into the corresponding alkanes.[32] The mechanism of this reaction was hypothesized to involve photoexcitation of the nicotinamide co-substrate, followed by single electron transfer to the benzyl halide and mesolytic cleavage. The Hyster lab has exploited this mode of reactivity within the active site of an enzyme (Figure 1.7B). [33] In this system, binding of halogenated substrate in the enzyme active site, near the nicotinamide cofactor, forms a charge transfer complex. Irradiation of this complex with 390 nm light elicits single electron transfer from the cofactor to the substrate, mesolytic cleavage of the C-X bond, and subsequent quenching of the resulting radical by stereoselective hydrogen atom transfer from the nicotinamide cofactor. Investigations of Nature's other organic cofactor redox catalysts, the flavin-dependent ene-reductases (EREDs), have resulted in enzymes which can catalyze new-to-nature radical cyclization[34] and cross-coupling[35] reactions in addition to dehalogenation, all with good enantioselectivity. Irraditation of the reduced form of the cofactor (flavin hydroquinone) generates a potenent, excited state reductant (Figure 1.7C). This excited state reduces the substrate alkene and generates a flavin semiquinone intermediate. The fate of the radical (cyclization versus cross-coupling versus rebound to the cofactor) is controlled by the enzyme scaffold. Hydrogen atom transfer from the flavin semiquinone to the substrate generate the oxidized form of flavin, which

must subsequently be reduced by NADH to regenerate the flavin hydroquinone, and restart the catalytic cycle.

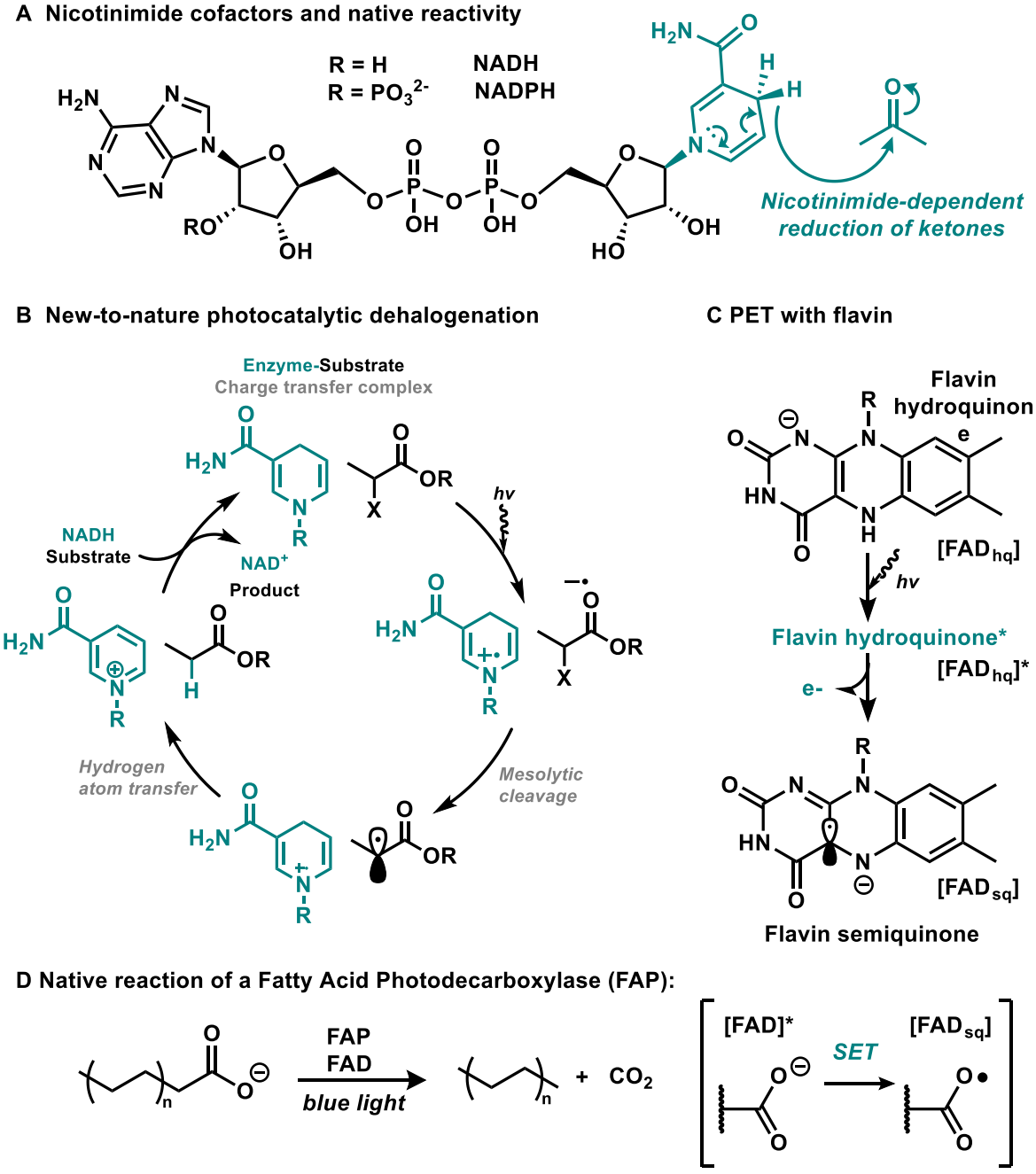


Figure 1.7. Photobiocatalytic reaction with reductase (nicotinamide and flavin dependent) enzymes. A) Native reactivity of nicotinamide cofactors. B) Mechanistic scheme for the new-to-nature photocatalytic dehalogenation of α-chloroesters with NAD-dependent reductases. C) Photoinduced electron transfer (PET) of the flavin cofactor. Irradiation with light accesses the potently reducing excited state species $[FAD_{hq}]^*$. D) Native reaction of a Fatty acid photodecarboxylase (FAP) enzyme. The key step involves photoexcitation of the FAD cofactor to access a highly oxidizing excited state flavin species, followed by single electron transfer.

Each of the above examples of new-to-nature enzymatic reactions have been improved immensely by protein engineering. The evolutionary trajectory of both carbene/nitrene transferases and photoreductases have been shaped by key mutations which substantially improved non-native reactivity, often by non-obvious mechanisms. For instance, the reactivity of carbene transferases was revolutionized by the introduction of an axial serine ligand in place of the canonical cysteine. This single mutation increased the redox potential of the iron center by 140 mV, and shut down the native monooxygenation activity of the catalysts, thereafter referred to as “P411” due to a characteristic shift in their CO-bound Soret absorbance. Similarly, the reactivity of photocatalytic EREDs was improved vastly by mutating an active site tyrosine to phenylalanine.[36] This single, subtle mutation decreased the rate of back-electron transfer relative to forward electron transfer and increased the quantum yield for the flavin cofactor.

These examples demonstrate two important principles regarding the development of enzymes for new-to-nature reactions. Firstly, new-to-nature reactivity can be enabled by accessing chemistry inherent to the bare cofactor. Cofactors can be catalytically promiscuous, and under the correct set of conditions, can be coopted to perform non-traditional reaction mechanisms. This idea may call into question whether the above examples truly are *new* to nature, since the naturally occurring cofactors innately possess this reactivity. Indeed, the existence of a naturally-occurring heme-dependent nitrene transferase enzyme (BezE) (Figure 1.5E),[30] and the photocatalytic reactivity of flavin-dependent Fatty Acid Photodecarboxylases (FAPs) (Figure 1.7D),[37] demonstrate that Nature can and does access these non-traditional mechanisms. Secondly, enzyme scaffolds are often have innate catalytic promiscuity. While enzymes have evolved to be highly selective within their natural context, there is virtually no evolutionary pressure to select against substrates and reactive intermediates that are not commonly encountered in nature. [38,39] Directed evolution enables repurposing of the enzyme environment to stabilize new reaction intermediates and is often necessary to achieve efficiencies similar to those of enzymes for their native reactions. Thus, new-to-nature reactivity is enabled by

the innate chemistry of the cofactor, and can be promoted and tuned by a suitable, engineerable enzyme scaffold.

1.4. Introducing new cofactors into enzymes imparts new reactivity.

The reactivity of cofactor-dependent enzymes is primarily derived from the cofactor itself, and most enzymes are cofactor-dependent.[40] Typically, the chemical role of cofactors is to stabilize otherwise highly reactive species. Pyridoxal-5'-phosphate (PLP), for instance, stabilizes anionic character at the alpha carbon of amino acids—a key intermediate for many chemical transformations of amino acids.[41] Thiamine Diphosphate (TDP) stabilizes acyl anion intermediates and enabled umpolung carbonyl reactivity.[42] Heme, which plays many roles in the cell, most notably activates molecular oxygen to form an electrophilically polarized oxygen species, as discussed earlier for P450 monooxygenases.[43,44]

Like the homochirality of life, the structures of 20 canonical amino acids and 4 nucleobases, and the universality of the central dogma, cofactors are remarkably well-conserved. Many cofactors, such as PLP, TDP, and heme, are found across all domains of life, indicating that the appearance of some cofactors may predate the last universal common ancestor (LUCA).[45] Some cofactors have been coopted by nature to carry out multiple different roles,[46] such as SAM, which carries out both two electron methylation and single electron radical rearrangements, among a host of other transformations.[47] Other cofactors are more specialized, such as the F_{430} cofactor which is found only in methanogenic archaea and is responsible for the organism's eponymous chemical reaction.[48] FeMoCo, the cofactor responsible for biologically-derived fixed nitrogen, is found only in certain cyano and root bacteria.[49] Despite these unique examples, a relatively limited set of cofactors seems to be responsible for majority of the chemistry of life. As a corollary, the chemistry of life is limited by the breadth of cofactors available to it, although that breadth is still quite considerable when the protein scaffold is used to tune reactivity.

In attempt to expand the breadth of chemical transformations available to enzymes and, correspondingly, biocatalytic reactions, researchers have invested tremendous effort into design and implementation of enzymes bearing new-to-nature cofactors. The first examples of artificial cofactors installed into enzymes sought to mimic the reactivity of the native cofactor-bearing

enzymes in unrelated or non-enzymatic model protein scaffolds. For instance, in the 1970s, several flavin derivatives were synthesized and introduced into a cysteine peptidase, resulting in a peptidase-based flavoprotein capable of oxidizing a nicotinamide cofactor with comparable efficiency to native flavoenzymes.[50] In another example, a disulfide containing mimic of PLP was appended to an active site cysteine in adipocyte lipid binding protein. This artificial PLP enzyme transaminated of a variety of α-keto-acids with good enantioselectivity, albeit non-catalytically.[51]

Inspired by the breadth and utility of organometallic chemistry in traditional synthetic chemistry settings, the incorporation of artificial metal centers into enzymes has been another major target. Perhaps the most famous example of this approach is the incorporation of Rhodium-complexes into a streptavidin or avidin protein, using a biotin tether, pioneered by Whitesides in the 1970s [52] The Ward group and others have expanded this concept for the successful catalysis of at least 12 distinct and new-to-nature transformations, including hydrogenation and metathesis, as well as new modes of ketone and imine reduction.[53] The directed evolution of the metallo-biotin tethered streptavidin system was heavily explored by both the Ward and Reetz groups.[54,55] Through rational design and iterative site saturation mutagenesis strategies, these enzymes have seen massive improvements in both rate, turnover number, and enantioselectivity, highlighting the ability of directed evolution tune the activity of cofactors for which the enzyme scaffold did not originally evolve.

Another key area of research involves substitution of the native iron center of hemoproteins for a different metal, rendering artificial hemoproteins. Inspired by the initial work of Arnold and coworkers, both the Hartwig and Fasan labs have explored the reactivity of metal substituted P450s towards carbene transfer reactions.[56–59] The introduction of Iridium-methyl-protoporphyrin IX (Ir-Me-PPIX) into P450s by the Hartwig group has received particular attention. IrMePPIX-P450 variants have been shown to catalyze the cyclopropanation of challenging, unactivated alkenes and C-H functionalization at high rates and with good enantioselectivity.[60–62] Unfortunately, many of these systems only marginally outcompete the reactivity of the free

cofactor (albeit, imparting significant improvements in enantioselectivity), and similarly challenging transformations can be achieved with highly engineered iron-containing enzymes.[20] Nevertheless, the biocompatibility of these artificial systems is exciting, demonstrated by the production of cyclopropane-containing L-carvone derivatives fully *in vivo*.[63] The incorporation of artificial cofactors into biosynthetic pathways has the potential to broaden the scope chemical transformations available to the scientific community through environmentally-friendly, fermentative processes.

1.5. A long standing challenge in the field: Easy access to biocatalysts bearing artificial cofactors

In each of the examples of artificial cofactors described in the previous section, the cofactor itself was synthesized exogenously, and then later appended to the enzyme scaffold. This method serviceably furnishes diverse artificial cofactor/enzyme pairs and has enabled the discovery of a host of new biocatalytic transformations. These enzymes, however, have not seen widespread adoption. A major hurdle in the field of artificial cofactor biocatalysis is access to the artificial cofactor bearing enzymes. While naturally-occurring cofactor-dependent enzymes can be obtained in the *holo* form in high titers through straightforward fermentative processes, routes to artificial cofactor-bearing enzymes are more challenging, since these cofactors typically cannot be biosynthesized *in vivo*.[64]

The many processes developed to procure metal-substituted hemoproteins illustrate the challenges associated with artificial cofactor substitution. Hemoproteins are a common target for artificial cofactor incorporation, as substitution of the metal center can impart substantial changes to the enzymes reactivity and spectroscopic properties, while the protoporphyrin IX (PPIX) prosthetic ligand maintains key interactions within the enzyme active site. The iron center in heme is tightly bound and cannot be exchanged through simple dialysis, so other approaches to metal substitution are needed. The most straightforward of these approaches is to express and isolate apo protein (lacking cofactor) and then add exogenous metalloporphyrin (Figure 1.8A). The holo protein must then be re-purified to remove any remaining free cofactor. One of the earliest examples of this approach was substitution of the heme in Sperm Whale Myoglobin for CoPPIX, which enables a more detailed study of oxygen binding and other spectroscopy [65]. Wanatabe and coworkers simplified this method by reconstituting the protein in cell lysate.[66] While this strategy is limited to enzymes that can be produced in their apo form, it is nonetheless powerful and capable of enabling the discovery of new biocatalytic functions. Hartwig and coworkers substituted apo-tolerant myoglobin and P450s with a variety of metal substituted PPIX in their

original demonstration of IrMe-P450 mediated carbene transfer [67]. In parallel with these *in vitro* approaches, researchers have actively sought methods to enable production of artificial metalloenzyme production *in vivo*.

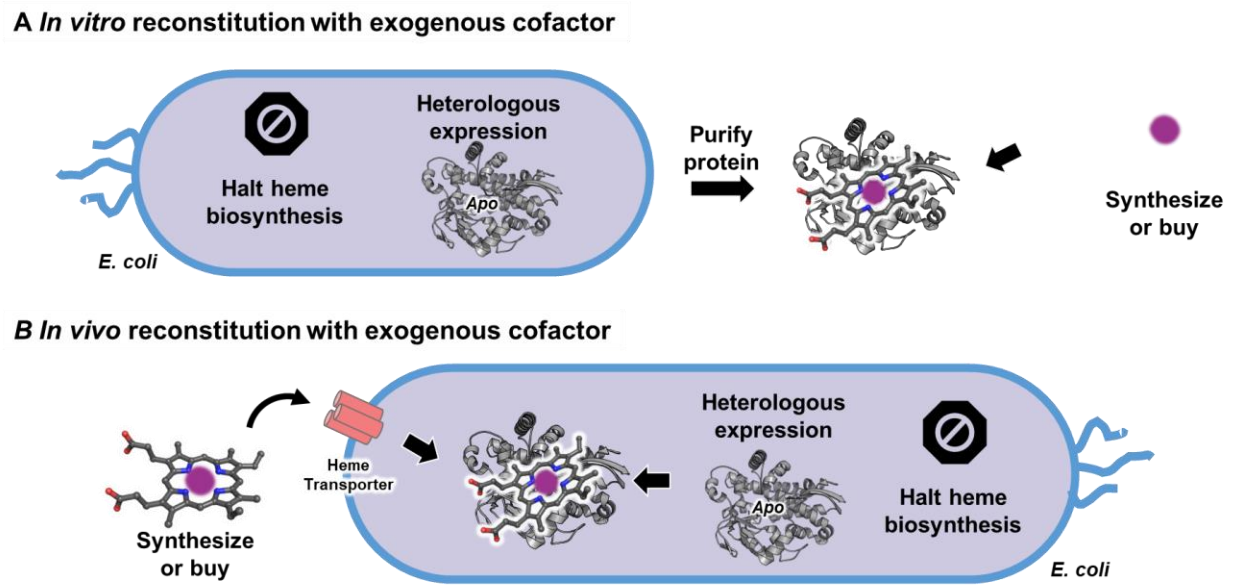


Figure 1.8. Routes to hemoproteins in which the native heme cofactor has been substituted for an alternatively metalated porphyrin cofactor. A) In vitro approaches to artificial hemoprotein production, requiring expression of the apo form of the hemoprotein. B) In vivo approaches to artificial hemoprotein production, usually involving a promiscuous heme transporter.

Several specialty strains of *E. coli* have been engineered to incorporate exogenously added non-native cofactors (Figure 1.8B) The strain RP523 has the native E. coli heme biosynthesis pathway disrupted and also carries an additional, uncharacterized mutation that makes the *E. coli* permeable to heme analogs, which can then be added to the media [68]. This method has been shown to accommodate two different proteins, two metals, and a variety of porphyrin synthetic analogs. However, the RP523 strain must be grown in anaerobic conditions unless media is supplemented with heme. Alternatively, the Fasan and Hartwig groups have overexpressed a heme transporter to facilitate incorporation of exogenous heme analogs [57]. This method has been used to incorporate diverse metal-PPIX into myoglobin and cytochrome P450 119 [58,69–71]. While these methods have enabled study of artificial hemoproteins, the efficiency of

these methods pale in comparison to that of producing native hemoproteins. Notably, each of these methods delegate the biosynthesis of the artificial cofactor outside the cell.

Nature's naturally occurring cofactors, as with every other chemical of life, must be either harvested or biosynthesized from earth-abundant materials. The acquisition of cofactors (vitamins and metal ions) and cofactor building blocks is dependent on the cell's natural environment, in addition to cellular homeostasis mechanisms. Of the metallocofactors, the most widely used cofactors, (Zn^{2+} and Mg^{2+}) are also the most bioavailable forms of highly earth-abundant metals.[72] The abundance of Fe^{2+} dependent proteins is perhaps the exception which proves this rule, as most of earth's iron is sequestered as non-bioavailable (and non-water-soluble) iron(III)oxide. It is hypothesized that Nature's dependance on iron arose prior to The Great Oxidation Event,(GOE) when soluble ferrous iron was plentiful.[73] Since GOE, living systems have developed complex, genetically-encoded systems to uptake iron from their environment and maintain its ferrous state. These genetically-encoded systems can be challenging to override.

A recent survey of the known organic cofactors highlighted that the majority of cofactors are biosynthesized from one or two amino acid or nucleotide building blocks.[45] Over millions of years of evolution, nature has orchestrated and genetically encoded the instructions for precise construction of the naturally-occurring cofactors we know today. In principle, both natural evolution and researcher-directed engineering of cofactor biosynthetic pathways have the potential to invent new, fermentatively-accessible cofactors.

1.6. Case Study: Biocatalytic cascade for the production of a tryptamine analog.

This section reflects contributions by L.J.P. to the following manuscript:

McDonald, A. D., Perkins, L. J.,Buller, A. R., ChemBioChem. **20**, 1939–1944 (2019).

The following technical example serves to illustrate the attractiveness of a typical biocatalytic transformation to the uninitiated synthetic chemist. These reactions are exceptionally cost effective, environmentally friendly, and safe. I will describe the process of obtaining catalyst, (called heterologous protein expression) and the execution of a one-pot, two step biocatalytic reaction for the synthesis of a tryptamine analog. The enzymes used for this reaction are cofactor (PLP) dependent enzymes. This cofactor is naturally produced by the host cell, and can also be supplemented in the reaction solution, as commercial supply is inexpesive and abundant, and the background reactivity of the bare cofactor is low.

Genetic construction of RgnTDC expression plasmid

The DNA sequence used to produce *Rgn*TDC is given in Table 1.1. Of special note are the elements of the sequence which have been colored. The green “ATG” represents the codon for methionine and is the universal start codon for all bacterial genes. TGA (red) encodes the stop codon and releases the protein from the ribosome after translation. The blue repeating CACCAT encodes for a 6-histidine extension to the C-terminus of the protein. This sequence binds metal ions tightly, and aids in the purification of proteins from the rest of the cellular material.

Table 1.1. Gene sequence for RgnTDC, from start codon (green) to stop codon(red).

Enzyme	Gene Sequence
*Rgn*TDC	**ATG**TCGCAGGTCATTAAGAAAAAACGCAATACGTTTATGATTGGAACGGAGTACATCCTTAATTC GACACAGTTAGAGGAGGCAATTAAGTCTTTCGTGCACGATTTTTGTGCGGAAAAACATGAGATC CATGATCAGCCCGTCGTTGTTGAAGCCAAGGAGCACCAGGAAGATAAAATTAAGCAGATCAAGA TCCCTGAAAAAGGACGCCCAGTAAATGAGGTCGTGAGTGAGATGATGAATGAAGTTTACCGCTA TCGCGGAGATGCGAACCACCCCCGTTTCTTCTCCTTCGTTCCGGGTCCAGCTTCGAGCGTCTC CTGGCTTGGAGACATCATGACGAGTGCATATAATATCCATGCCGGAGGCAGTAAATTGGCTCCC ATGGTAAACTGTATTGAGCAAGAAGTGCTGAAGTGGTTGGCAAAGCAAGTGGGATTTACTGAAA ATCCCGGCGGGGTGTTCGTCTCAGGTGGCTCGATGGCGAACATCACGGCGTTAACAGCAGCC CGTGACAATAAACTTACTGACATTAATTTGCATTTAGGAACGGCGTATATCAGCGACCAAACACA CTCCAGTGTAGCCAAGGGGTTACGTATTATTGGCATCACCGACAGCCGTATTCGCCGTATTCCC ACTAATTCGCACTTCCAAATGGATACGACCAAGTTGGAGGAGGCCATTGAAACCGATAAAAAGA GTGGCTATATCCCGTTTGTAGTGATCGGAACCGCTGGCACGACTAATACAGGATCCATTGACCC ATTAACGGAAATTTCTGCATTATGTAAAAAGCACGATATGTGGTTCCACATCGACGGTGCGTATG GTGCCTCCGTATTGCTTAGTCCAAAATATAAGTCCCTTTTGACAGGAACAGGATTAGCAGATAGT ATTTCTTGGGATGCTCACAAATGGTTATTCCAGACGTATGGGTGCGCCATGGTATTGGTGAAGG ACATCCGCAACCTGTTCCATTCCTTTCACGTTAACCCCGAATATCTGAAAGACCTTGAGAATGAC ATTGATAATGTCAATACGTGGGATATTGGGATGGAGTTAACACGTCCGGCACGTGGGCTTAAAC TTTGGCTGACCTTGCAGGTTCTTGGGTCCGACCTTATTGGGTCTGCAATTGAGCACGGTTTCCA ATTAGCGGTATGGGCGGAAGAAGCGCTGAATCCCAAAAAAGATTGGGAAATTGTTAGCCCTGC CCAGATGGCGATGATTAATTTTCGCTACGCGCCTAAGGATTTAACCAAAGAGGAGCAGGACATC CTTAATGAAAAGATTTCGCATCGCATCTTGGAATCAGGCTATGCCGCTATTTTTACTACTGTGCT GAATGGTAAGACAGTGTTACGCATTTGCGCGATTCACCCTGAGGCTACTCAAGAGGATATGCAG CACACCATTGATCTGTTGGACCAATACGGTCGCGAGATCTATACTGAAATGAAAAAGGCTCTCG AGCAC**CATCACCATCACCATTGA**

The gene sequence for *Rgn*TDC was inserted into a plasmid using the Gibson assembly method, which is described elsewhere.[74] In Nature, bacteria use plasmids to share genetic information with neighboring, competent bacteria on a short-term basis. Researchers have exploited the molecular biology of plasmids for decades to introduce genetic information (for example, the gene encoding a protein of interest) into *E. coli* and other bacteria without needing to alter the bacterium's native genome. Most plasmids for protein expression include (1) a selection marker gene (2) short a promoter sequence (Figure 1.9) For the *Rgn*TDC plasmid, the selection marker gene encodes a beta-lactamase enzyme for ampicillin. This gene renders bacteria bearing the plasmid as ampicillin-resistant. Growth in media supplemented with ampicillin ensures that only bacteria bearing the plasmid, and capable of expressing the gene of interest, proliferate. The short promoter sequence used to express *Rgn*TDC is the T7 promoter. This sequence acts as an on/off switch for protein expression. Addition of a small molecule "inducer", Isopropyl β-D-thiogalactoside (IPTG) (Figure 1.9), causes a change in regulatory pathways such

that a T7 RNA polymerase (encoded elsewhere on the bacterial genome) is produced, and binds to the T7 promoter on the plasmid. T7 polymerase transcribes the gene of interest (the *Rgn*TDC gene) into *m*RNA, which is subsequently translated into protein (*Rgn*TDC).

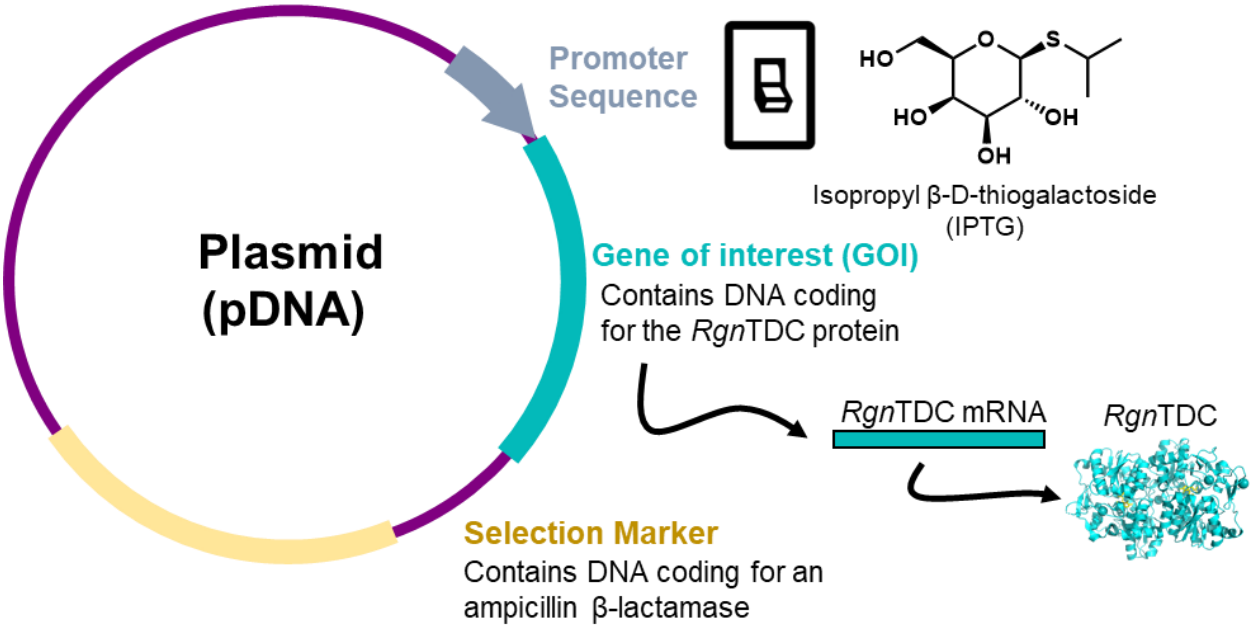


Figure 1.9 Simplified schematic of plasmid structure, depicting the central elements for heterologous protein expression.

Once constructed the plasmid DNA must be inserted into a cell line of *E. coli* BL21(DE3). BL21(DE3) is the most commonly used laboratory strain of *E.* coli for the production of proteins, and is fast-growing, expresses heterologous proteins at high titers, and is non-pathogenic.[75] The process of inserting DNA into bacteria is called “transformation” and, in this case, was accomplished via electroporation.[76]

The design and construction of the system used to produce the biocatalyst of interest is by far the most complex step of biocatalysis. In practice, most expression constructs can be obtained from a cell repository such as Addgene, or from a collaborator or generous author. Self-regeneration is one of Life’s defining features, and, consequently, requesting a cell stock from an academic researcher is not a zero-sum game. Indeed, giving away a small portion of an infinitely reproducible cell stock is a far cry from donating a few grams of precious organometallic catalyst.

Production of the biocatalyst (*Rgn*TDC)

A cell stock of *Rgn*TDC was used to inoculate 5 mL of Terrific Broth (TB), supplemented with ampicillin. These cultures were grown overnight at 37 °C, with shaking. Expression cultures, typically 1 L of TB with ampicillin were inoculated from these starter cultures and shaken at 37 °C. After 3 hours, the expression cultures were chilled on ice. After 90 min on ice, expression was induced with 1 mM IPTG, and the cultures were supplemented with 0.5 mM indole. Addition of central metabolites, such as indole, to growth media is a relatively common practice. In this case, indole is added because *Rgn*TDC is known to deplete intracellular levels of tryptophan, and indole upregulated the production of tryptophan. Cultures were incubated overnight at 20 °C with shaking. Cells were then harvested by centrifugation. Cell pellets were used fresh or frozen at -20 °C until lysis.

Like many biocatalysts, *Rgn*TDC can be prepared in a variety of forms including whole cells (the procedure outlined up until this point), cellular lysate, or purified protein. Each form contains *Rgn*TDC protein that has been overexpressed and purified from cell culture to varying degrees.

To generate *Rgn*TDC, cellular lysate, cell pellets were thawed on ice and then resuspended in lysis buffer (50 mM potassium phosphate buffer (pH = 8.0), 1 mg/ml Hen Egg White Lysozyme (commercially available reagent enzyme), 0.2 mg/ml DNaseI (commercially available reagent enzyme), 1 mM $MgCl_2$, and 150 µM pyridoxal 5′-phosphate (PLP, the cofactor for *Rgn*TDC)). A volume of 4 mL of lysis buffer per gram of wet cell pellet was used. After 30 min of shaking at 37 °C, the lysis suspension was disrupted using sonication. The resulting lysate was then centrifuged at 75,000 x*g* to pellet cell debris, yielding cellular lysate of *Rgn*TDC

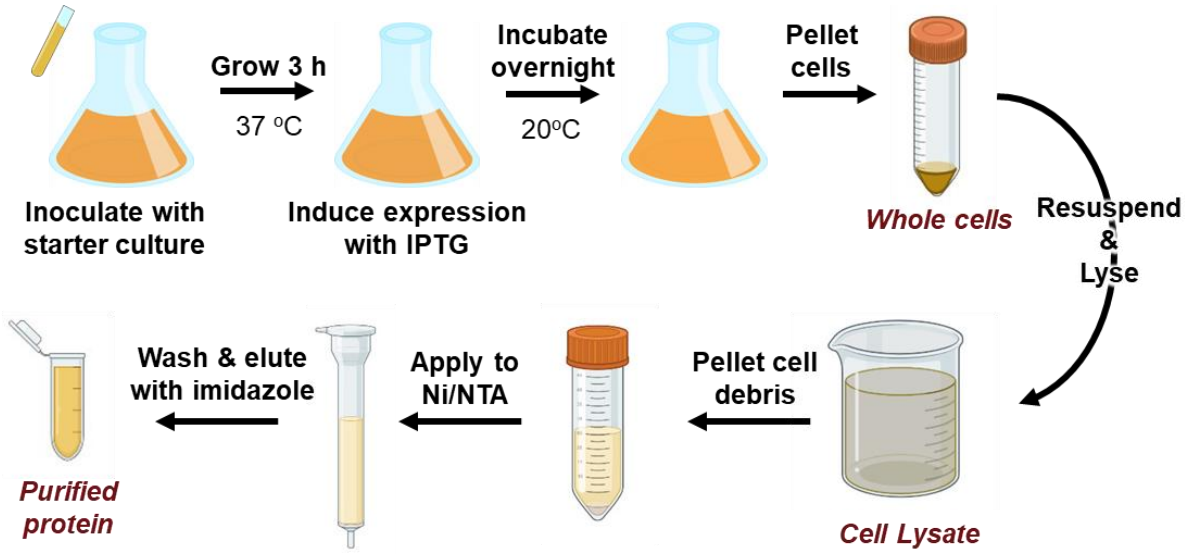


Figure 1.10. Basic workflow for the purification of heterologously expressed protein from E. coli cell culture. Key formulations for biocatalysts are highlighted in red.

To further purify *Rgn*TDC, metal affinity chromatography was used. Commercially available Ni/NTA resin, which comprises Ni^{2+} appended to a solid phase sepharose bead, was added to the *Rgn*TDC lysate and the mixture incubated on ice for 40 min. The mixture was then loaded into a gravity column, and the solid phase was washed with 3 column volumes of a buffer containing 20 mM imidazole, 100 mM NaCl, 50 mM potassium phosphate buffer (pH = 8.0). Protein was eluted with from the column with 250 mM imidazole, 100 mM NaCl, 50 mM potassium phosphate buffer (pH = 8.0). Elution of the desired protein product was monitored by the disappearance of its bright yellow color (resulting from the PLP cofactor) from the column. The protein product was dialyzed to < 1 μM imidazole, dripped into liquid nitrogen to flash freeze, and stored at -80 °C for no more than 1 month before use, as activity was found to decrease following long periods at -80 °C. The concentration of protein was determined by Bradford assay.[77] Generally, this procedure yielded 150 – 200 mg of purified protein per liter of cell culture.

Enzymatic synthesis of 7-iodotryptamine

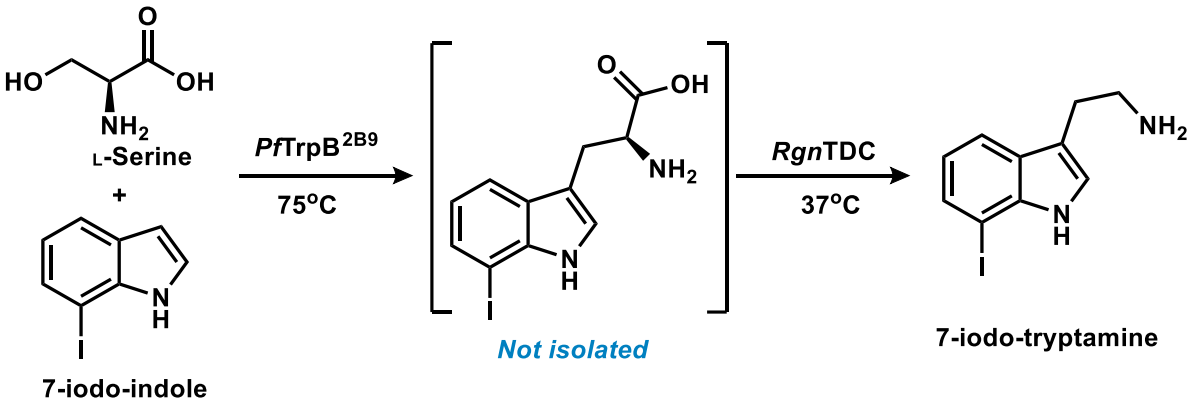


Figure 1.11. Reaction scheme for the one-pot, two enzyme reaction to make 7-iodo-tryptamine

The chemical transformation via the one-pot, two enzyme sequential reaction is exceedingly straightforward. 7-Iodo-Indole (1.0 mmol, 243 mg) and L-serine (3 mmol, 105 mg) were dissolved in 5 mL of methanol, in a 150 mL glass screw top vial. The reaction mixture was diluted to 100 mL with 50 mM potassium phosphate buffer (pH = 8.0). *Pf*TrpB^{2B9} solution was added to bring the total enzyme concentration to 2.0 μM (0.02 mol% catalyst, 5000 max TTN). *Pf*TrpB^{2B9} is a thermostable enzyme whose activity is greatest at high temperatures, and so the reaction was heated to 75 °C in a water bath overnight. The next day the vial was cooled to room temperature on the benchtop. Once cooled, 20 μM PLP is added, as *Rgn*TDC has been shown to operate more efficiently with excess PLP. Finally, *Rgn*TDC (2.0 μM, 5000 max TTN) is added to the reaction mixture, and the sealed vial is incubated at 37 °C. Once the reaction had reached an apparent 100% conversion, or after 12 hours, the reaction mixture was acidified, and the tryptamine was obtained as a pure material after sequential acid wash and basic extraction into ethyl acetate in 77% yield (206 mg). This reaction can also be conducted in a simultaneous cascade fashion, but the activity of *Pf*Trp^{2B9} at 37 °C is slow for the synthesis of Trp analogs, primarily due to the poor solubility of these analogs at room temperature (Figure 1.11)

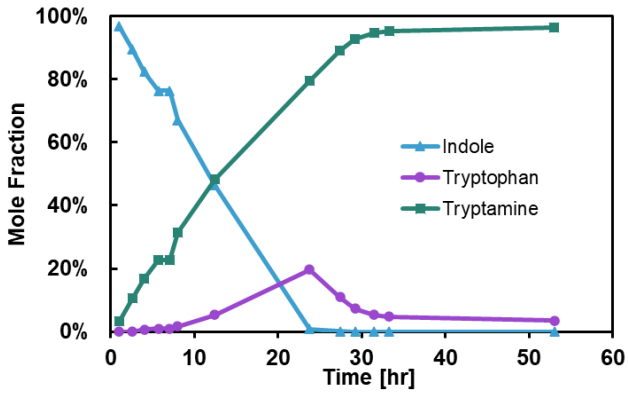


Figure 1.12. Progress curve for the coupled enzyme reaction producing tryptamine from indole (15 mM) and L-Serine (45 mM) in 1 ml 50 mM potassium phosphate buffer (pH=8.0) with 5% DMSO as cosolvent. Reaction utilized 3 μM (5000 Max TTN, 0.02% catalyst) each of Pyrococcus furiosus tryptophan synthase (PfTrpB) and Ruminococcus gnavus tryptophan decarboxylase (RgnTDC). At each time point, 50 μL of the reaction mixture was quenched with 50 μL of quenching solution (equal volume solution of acetonitrile and 1 M HCl). Quenched reactions were centrifuged for 1 min at 15,000 x g and then analyzed via UPLC-MS. Relative amounts of substrate, intermediate, and product were quantified by absorbance at 280 nm.

Concluding remarks

The key innovation of this thesis is the development and mechanistic understanding of methods for producing enzymes bearing unnatural cofactors. The above example of a biocatalytic reaction with PLP-dependent enzymes serves as a gold standard for ease of access – PLP enzymes are widely used for synthetic and industrial applications. Note the following:

1. The cofactor-loaded catalyst is produced without significant researcher intervention. The addition of various inexpensive additives is tolerable, but more expensive additives significantly affect the cost of catalyst production.
2. The holoenzyme can be obtained in a variety of forms, from whole cells to purified protein. This flexibility enables directed evolution (activity screens are often conducted in cell lysates or whole cells) and the development of fully *in vivo* biocatalytic cascades.
3. The synthetic reaction itself is extremely easy, efficient (5000 turnovers for each enzyme), and environmentally-friendly. Biocatalytic routes to small molecules are often preferable to synthetic routes, when available.

1. 7. Preface to remaining chapters

While enzymes bearing artificial cofactors have the potential to dramatically expand the repertoire of available biocatalytic transformations, accessing these enzymes remains a challenge to their discovery, development, and widespread use. In my doctoral work, I have explored both the biosynthesis and implementation of the new-to-nature cofactor, cobalt protoporphyrin IX (CoPPIX). In chapters 2 and 3 I discuss the discovery and mechanism of CoPPIX production in *E coli* BL21(DE3). Our explorations of this process have enabled the development of a method for artificial metalloenzyme production that is as straightforward as methods for the naturally occurring hemoprotein, paving the way for substantial engineering of this enzyme. In Chapter 3, a new-to-nature reaction, metal-hydrogen atom transfer (MHAT)-dependent O-deallylation, is presented. This reaction is enabled by the innate chemistry of a CoPPIX cofactor incorporated into cytochrome P450 119 (CoCYP119). We used directed evolution to tune the reactivity of the catalyst and study the reaction's mechanism. Lastly, Chapter 5 explores other new-to-nature MHAT reactions catalyzed by CoCYP119, and outlines our current efforts to discover synthetically useful enzymatic MHAT reactions.

1. 8. References

1. Wöhler, F. Ueber künstliche Bildung des Harnstoffs. *Ann. Phys.* **88**, 253–256 (1828).

2. Brown, D. G. Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? *J. Med. Chem.* 4443–4458 (2016) doi:10.1021/acs.jmedchem.5b01409.

3. Anastas, P. T., Kirchhoff, M. M. & Williamson, T. C. Catalysis as a foundational pillar of green chemistry. *Appl. Catal. A Gen.* **221**, 3–13 (2001).

4. Devine, P. N. *et al.* Extending the application of biocatalysis to meet the challenges of drug development. *Nat. Rev. Chem.* **2**, 409–421 (2018).

5. Huffman, M. A. *et al.* Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science (80-.).* **366**, 1255–1259 (2019).

6. Bar-Even, A. *et al.* The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402–4410 (2011).

7. Arnold, F. H. Design by Directed Evolution. *Acc. Chem. Res.* **31**, 125–131 (1998).

8. Carballeira, D. & Reetz, M. T. Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2**, 891–903 (2007).

9. Zeymer, C. & Hilvert, D. Directed Evolution of Protein Catalysts. *Annu. Rev. Biochem.* **87**, (2018).

10. Wang, Y. *et al.* Directed Evolution : Methodologies and Applications. *Chem. Rev.* 12384–12444 (2021) doi:10.1021/acs.chemrev.1c00260.

11. Wilson, D. & Keefe, A. D. Random Mutagenesis by PCR. *Curr. Protoc. Mol. Biol.* 8.3.1-8.3.9 (2000) doi:10.1002/0471142727.mb0803s51.

12. Morrison, M. S., Podracky, C. J. & Liu, D. R. The developing toolkit of continuous directed evolution. *Nat. Chem. Biol.* **16**, 610–619 (2020).

13. Kille, S. *et al.* Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. (2012) doi:10.1021/sb300037w.

14. Reetz, M. T. *et al.* Expanding the Substrate Scope of Enzymes : Combining Mutations Obtained by CASTing. *Chem. - A Eur. J.* **12**, 5895–6179 (2006).

15. Yang, G. & Withers, S. G. Ultrahigh-Throughput FACS-Based Screening for Directed Enzyme Evolution. *ChemBioChem* **10**, 2704–2715 (2009).

16. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nat. 2011 4727344* **472**, 499–503 (2011).

17. Cobb, R. E., Chao, R. & Zhao, H. Directed evolution: Past, present, and future. *AIChE J.* **59**, 1432–1440 (2013).

18. Chen, K. & Arnold, F. H. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc. Natl. Acad. Sci.* **90**, 5618–5622 (1993).

19. Chen, K. & Arnold, F. H. Engineering new catalytic activities in enzymes. *Nat. Catal.* **3**, 203–213 (2020).

20. Liu, Z. & Arnold, F. H. New-to-nature chemistry from old protein machinery: carbene and nitrene transferases. *Curr. Opin. Biotechnol.* **69**, 43–51 (2021).

21. Breslow, R. & Gellman, S. H. Tosylamidation of cyclohexane by a cytochrome P-450 model. *J. Chem. Soc. Chem. Commun.* **1**, 1400–1401 (1982).

22. Breslow, R. & Gellman, S. H. Intramolecular Nitrene C-H Insertions Mediated by Transition-

Metal Complexes as Nitrogen Analogues of Cytochrome P-450 Reactions. *J. Am. Chem. Soc.* **105**, 6728–6729 (1983).

23. Wolf, J. R., Djukic, J.-P., Woo, L. K., Hamaker, C. G. & Kodadek, T. Shape and stereoselective cyclopropanation of alkenes catalyzed by iron porphyrins. *J. Am. Chem. Soc.* **117**, 9194–9199 (2005).

24. Svastits, E. W., Dawson, J. H., Breslow, R. & Gellman. Functionalized Nitrogen Atom Transfer Catalyzed by Cytochrome P-4501. *J. Am. Chem. Soc* **107**, 6427–6428 (1985).

25. Coelho, P. S., Brustad, E. M., Kannan, A. & Arnold, F. H. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science (80-.).* **339**, 307–310 (2013).

26. Athavale, S. V. *et al.* Enzymatic Nitrogen Insertion into Unactivated C-H Bonds. *J. Am. Chem. Soc.* **144**, 19097–19105 (2022).

27. Zhang, R. K. *et al.* Enzymatic assembly of carbon–carbon bonds via iron-catalysed sp3 C–H functionalization. *Nature* **565**, 67–72 (2019).

28. Chen, K., Huang, X., Jennifer Kan, S. B., Zhang, R. K. & Arnold, F. H. Enzymatic construction of highly strained carbocycles. *Science (80-.).* **360**, 71–75 (2018).

29. Miller, D. C., Lal, R. G., Marchetti, L. A. & Arnold, F. H. Biocatalytic One-Carbon Ring Expansion of Aziridines to Azetidines via a Highly Enantioselective [1,2]-Stevens Rearrangement. *J. Am. Chem. Soc.* **144**, 4739–4745 (2022).

30. Tsutsumi, H. *et al.* Unprecedented Cyclization Catalyzed by a Cytochrome P450 in Benzastatin Biosynthesis. *J. Am. Chem. Soc.* jacs.8b02769 (2018) doi:10.1021/jacs.8b02769.

31. Nazor, J., Liu, J. & Huisman, G. Enzyme evolution for industrial biocatalytic cascades. *Curr.*

Opin. Biotechnol. **69**, 182–190 (2021).

32. Fukuzumi, S., Hironaka, K. & Tanaka, T. Photoreduction of Alkyl Halides by an NADH Model Compound. An Electron-Transfer Chain Mechanism. *J. Am. Chem. Soc.* **105**, 4722–4727 (1983).

33. Emmanuel, M. A., Greenberg, N. R., Oblinsky, D. G. & Hyster, T. K. Accessing non-natural reactivity by irradiating nicotinamide-dependent enzymes with light. *Nature* **540**, 414–417 (2016).

34. Clayman, P. D. & Hyster, T. K. Photoenzymatic Generation of Unstabilized Alkyl Radicals: An Asymmetric Reductive Cyclization. *J. Am. Chem. Soc.* **142**, 15673–15677 (2020).

35. Huang, X. *et al.* Photoenzymatic enantioselective intermolecular radical hydroalkylation. *Nature* **584**, 69–74 (2020).

36. Nicholls, B. T. *et al.* Engineering a Non-Natural Photoenzyme for Improved Photon Efficiency**. *Angew. Chemie* **134**, (2022).

37. Heyes, D. J. *et al.* Photochemical Mechanism of Light-Driven Fatty Acid Photodecarboxylase. *ACS Catal.* **10**, 6691–6696 (2020).

38. Hammer, S. C., Knight, A. M. & Arnold, F. H. Design and evolution of enzymes for non-natural chemistry. *Curr. Opin. Green Sustain. Chem.* **7**, 23–30 (2017).

39. Tawfik, D. S. Messy biology and the origins of evolutionary innovations. *Nat. Chem. Biol.* **6**, 692–696 (2010).

40. Christen, P. & Mehta, P. K. From cofactor to enzymes. The molecular evolution of pyridoxal-s'-pliosphate-dependent enzymes. *Chem. Rec.* **1**, 436–447 (2001).

41. Du, Y. L. & Ryan, K. S. Pyridoxal phosphate-dependent reactions in the biosynthesis of

natural products. *Nat. Prod. Rep.* **36**, 430–457 (2019).

42. Kluger, R. & Tittmann, K. Thiamin diphosphate catalysis: Enzymic and nonenzymic covalent intermediates. *Chem. Rev.* **108**, 1797–1833 (2008).

43. Ortiz de Montellano, P. R. Hydrocarbon hydroxylation by cytochrome P450 enzymes. *Chem. Rev.* **110**, 932–948 (2010).

44. Huang, X. & Groves, J. T. Oxygen Activation and Radical Transformations in Heme Proteins and Metalloporphyrins. *Chemical Reviews* vol. 118 2491–2553 (2018).

45. Fischer, J. D., Holliday, G. L., Rahman, S. A. & Thornton, J. M. The structures and physicochemical properties of organic cofactors in biocatalysis. *J. Mol. Biol.* **403**, 803–824 (2010).

46. Richter, M. Functional diversity of organic molecule enzyme cofactors. *Nat. Prod. Rep.* **30**, 1324–1345 (2013).

47. Sun, Q., Huang, M. & Wei, Y. Diversity of the reaction mechanisms of SAM-dependent enzymes. *Acta Pharm. Sin. B* **11**, 632–650 (2021).

48. Diekert, G., Konheiser, U., Piechulla, K. & Thauer, R. K. Nickel requirement and factor F430 content of methanogenic bacteria. *J. Bacteriol.* **148**, 459–464 (1981).

49. Einsle, O. & Rees, D. C. Structural Enzymology of Nitrogenase Enzymes. *Chem. Rev.* **120**, 4969–5004 (2020).

50. Levine, H. L., Nakagawa, Y. & Kaiser, E. T. Flavopapain: Synthesis and properties of semi-synthetic enzymes. *Biochem. Biophys. Res. Commun.* **76**, 64–70 (1977).

51. Kuang, H., Brown, M. L., Davies, R. R., Young, E. C. & Distefano, M. D. Enantioselective reductive amination of α-keto acids to α-amino acids by a pyridoxamine cofactor in a protein

cavity. *J. Am. Chem. Soc.* **118**, 10702–10706 (1996).

52. 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).

53. Ward, T. R. Artificial metalloenzymes based on the biotin - Avidin technology: Enantioselective catalysis and beyond. *Acc. Chem. Res.* **44**, 47–57 (2011).

54. Hyster, T. K., Knörr, L., Ward, T. R. & Rovis, T. Biotinylated Rh(III) complexes in engineered streptavidin for accelerated asymmetric C-H activation. *Science (80-.).* **338**, 500–503 (2012).

55. 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 (2006) doi:10.1039/b610461d.

56. Natoli, S. N. & Hartwig, J. F. Noble-Metal Substitution in Hemoproteins: An Emerging Strategy for Abiological Catalysis. *Acc. Chem. Res.* (2019) doi:10.1021/acs.accounts.8b00586.

57. Bordeaux, M., Singh, R. & Fasan, R. Intramolecular C(sp3)H amination of arylsulfonyl azides with engineered and artificial myoglobin-based catalysts. *Bioorganic Med. Chem.* **22**, 5697–5704 (2014).

58. Sreenilayam, G., Moore, E. J., Steck, V. & Fasan, R. Metal Substitution Modulates the Reactivity and Extends the Reaction Scope of Myoglobin Carbene Transfer Catalysts. *Adv. Synth. Catal.* **359**, 2076–2089 (2017).

59. Dydio, P. *et al.* An artificial metalloenzyme with the kinetics of native enzymes. *Science (80-.).* **354**, 102–106 (2016).

60. Dydio, P. *et al.* An artificial metalloenzyme with the kinetics of native enzymes. *Science* **354**, 102–106 (2016).

61. Gu, Y. *et al.* Directed Evolution of Artificial Metalloenzymes in Whole Cells. *Angew. Chemie - Int. Ed.* **61**, (2022).

62. Liu, Z. *et al.* Assembly and Evolution of Artificial Metalloenzymes within E. coli Nissle 1917 for Enantioselective and Site-Selective Functionalization of C─H and C═C Bonds. *J. Am. Chem. Soc.* **144**, 883–890 (2022).

63. Huang, J. *et al.* Unnatural biosynthesis by an engineered microorganism with heterologously expressed natural enzymes and an artificial metalloenzyme. *Nat. Chem.* **13**, 1186–1191 (2021).

64. Bloomer, B. J., Clark, D. S. & Hartwig, J. F. Progress, Challenges, and Opportunities with Artificial Metalloenzymes in Biosynthesis. *Biochemistry* **62**, 221–228 (2023).

65. Hoffman, B. M. & Petering, D. H. Coboglobins: Oxygen-Carrying Cobalt-Reconstituted Hemoglobin and Myoglobin*. *Proc. Natl. Acad. Sci.* **67**, 637–643 (1970).

66. 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).

67. 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).

68. Woodward, J. J., Martin, N. I. & Marletta, M. A. An Escherichia coli expression-based method for heme substitution. *Nat. Methods* **4**, 43–45 (2007).

69. Wei, Y., Tinoco, A., Steck, V., Fasan, R. & Zhang, Y. Cyclopropanations via Heme Carbenes: Basic Mechanism and Effects of Carbene Substituent, Protein Axial Ligand, and Porphyrin Substitution. *J. Am. Chem. Soc.* **140**, 1649–1662 (2018).

70. Moore, E. J., Steck, V., Bajaj, P. & Fasan, R. Chemoselective Cyclopropanation over Carbene Y−H Insertion Catalyzed by an Engineered Carbene Transferase. (2018) doi:10.1021/acs.joc.8b00946.

71. Liu, Z. *et al.* Assembly and Evolution of Artificial Metalloenzymes within E. coli Nissle 1917 for Enantioselective and Site-Selective Functionalization of C─H and C═C Bonds. *J. Am. Chem. Soc.* **144**, 883–890 (2022).

72. Waldron, K. J. & Robinson, N. J. How do bacterial cells ensure that metalloproteins get the correct metal? *Nat. Rev. Microbiol.* **7**, 25–35 (2009).

73. Ilbert, M. & Bonnefoy, V. Insight into the evolution of the iron oxidation pathways. *Biochim. Biophys. Acta - Bioenerg.* **1827**, 161–175 (2013).

74. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

75. Rosano, G. L., Morales, E. S. & Ceccarelli, E. A. New tools for recombinant protein production in Escherichia coli: A 5-year update. *Protein Sci.* **28**, 1412–1422 (2019).

76. Chassy, B. M., Mercenier, A. & Flickinger, J. Transformation of bacteria by electroporation. *Trends Biotechnol.* **6**, 303–309 (1988).

77. Marion M. Bradford. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).

Chapter 2

De novo biosynthesis of a non-natural Co-porphyrin cofactor in *E. coli* and incorporation into hemoproteins

Adapted with permission from Brian Weaver, Judith Burstyn, and Andrew Buller. "*De novo* biosynthesis of a non-natural Co-porphyrin cofactor in *E. coli* and incorporation into hemoproteins". L. J. Perkins, B. R. Weaver, A. R. Buller, J. N. Burstyn, *Proc. Natl. Acad. Sci.* **118**, 2021 (2021).

B.R.W. collected and analyzed all electron paramagnetic resonance (EPR) and electronic absorption (UV-vis) spectroscopic data presented in this chapter. Serial passaging of BL21(DE3) through Co-supplemented media, Protein expression, and quantification of porphyrin loading was conducted by both L.J.P and B.R.W.. L.J.P collected and analyzed all *in vivo* growth data.

Chapter 2: De novo biosynthesis of a non-natural Co-porphyrin cofactor in E. coli and incorporation into hemoproteins

2. 1. Abstract

Enzymes that bear a non-native or artificially introduced metal center, can engender novel reactivity and enable new spectroscopic and structural studies. In the case of metal-organic cofactors, such as metalloporphyrins, no general methods exist to build and incorporate new-to-nature cofactor analogs in vivo. We report here that a common laboratory strain, *E. coli* BL21(DE3), biosynthesizes cobalt protoporphyrin IX (CoPPIX) under iron-limited, cobalt-rich growth conditions. In supplemented minimal media containing $CoCl_2$, the metabolically-produced CoPPIX is directly incorporated into multiple hemoproteins in place of native heme b (FePPIX). Five cobalt-substituted proteins were successfully expressed with this new-to-nature cobalt porphyrin cofactor: myoglobin H64V V68A, dye decolorizing peroxidase, aldoxime dehydratase, cytochrome P450 119, and catalase. We show conclusively that these proteins incorporate CoPPIX, with the CoPPIX making up at least 95% of the total porphyrin content. In cases where the native metal ligand is a sulfur or nitrogen, spectroscopic parameters are consistent with retention of native metal ligands. This method is an improvement on previous approaches with respect to both yield and ease-of-implementation. Significantly, this method overcomes a long-standing challenge to incorporate non-natural cofactors through *de novo* biosynthesis. By utilizing a ubiquitous laboratory strain, this process will facilitate spectroscopic studies and the development of enzymes for new, CoPPIX-mediated biocatalysis.

2. 2. Introduction

The identity of a metal center often dictates enzymatic activity and swapping the native metal for an alternative one or introducing a new metal center has profound effects. More generally, the chemical utility of natural cofactors has inspired decades of study into synthetic analogs with distinct properties and researchers have subsequently sought straightforward ways

to put these novel cofactors back into proteins .[1] Substituted metalloenzymes constitute one of the simplest cases. Changing the identity of the metal ion in metalloproteins has enabled powerful spectroscopic and functional studies of these proteins [2–10] in addition to new biocatalytic activities.[11–20] However, most methods for producing such proteins with new-to-nature cofactors are limited by the inability to produce the novel protein-cofactor complex *in vivo*.

Hemoproteins, in particular, have been studied through metal substitution because of their important biological functions and utility as biocatalysts. Heme is a ubiquitous and versatile cofactor in biology and heme-dependent proteins serve essential gas sensing functions,[21] metabolize an array of xenobiotic molecules,[22] and perform synthetically useful oxygen activation and radical-based chemistry.[23] Metal-substituted hemoproteins have enabled key spectroscopic studies of hemoprotein function and the development of biocatalysts with novel reactivity (Figure 2.1). For example, electron paramagnetic resonance (EPR) studies on cobalt-substituted sperm whale myoglobin (CoMb) enabled detailed characterization of the paramagnetic $CoMbO_2$ complex [3,4,24,25]. In analogous oxygen-binding studies in CoMb and Co-hemoglobin [5,6,26], resonance Raman was used to identify the O-O stretching mode because cobalt-substituted proteins exhibit enhancement of this vibrational mode compared to the native iron proteins.

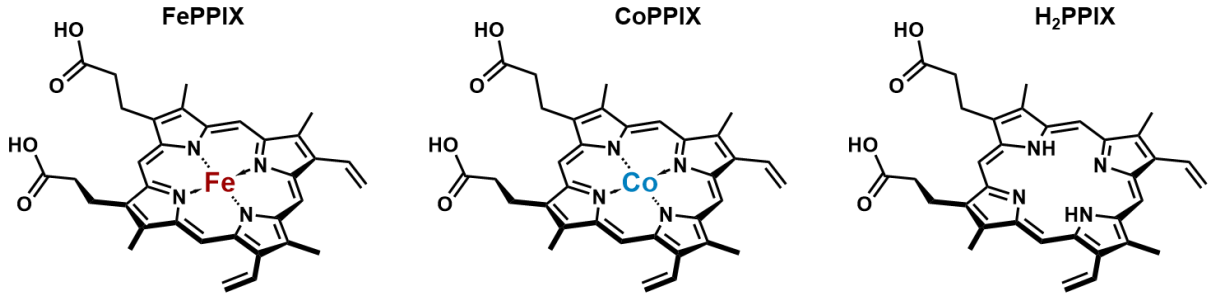


Figure 2.0.1. Chemical structures of iron protoporphyrin IX (FePPIX or heme b), cobalt protoporphyrin IX (CoPPIX) and free base protoporphyrin IX (H_2PPIX).

Metal substitution has a profound effect on catalytic activity of hemoproteins, enabling diverse new activities. Substitution of the native iron for cobalt in several hemoproteins, including a thermostable cytochrome c variant, enabled the reduction of water to H_2 under aerobic, aqueous

conditions. [27–29] Reconstitution of apoprotein with selected metalloporphyrins has been used to generate metal-substituted myoglobin and cytochrome P450s variants. These enzymes were effective as biocatalysts for C-H activation and carbene insertion reactions. [11–14] In a tour de force of directed evolution, which required purification and cofactor reconstitution of each individual variant, Hartwig and coworkers generated a cytochrome P450 variant that utilizes a non-native Ir(Me)mesoporphyrin cofactor to perform desirable C-H activation chemistry.[14] These activities may not be unique to the Ir-substituted protein, as synthetic Co-porphyrin complexes have been shown to mediate a variety of Co(III)-aminyl and -alkyl radical transformations, including C-H activation.[30–32] Indeed, a number of Co-porphyrin carbene complexes display significant carbon centered radical character, [33–35] whereas the corresponding Fe-porphyrin complexes are closed shell species, [36,37] indicating that Co-porphyrins may possess distinct, complementary modes of reactivity. [38–40]

Inspired by these applications, researchers have sought strategies for generating metal-substituted hemoproteins. For many metalloproteins, metal substitution is carried out by removal of the native metal with a chelator and replacement with an alternate metal of similar coordination preference. This method is inapplicable to hemoproteins, as porphyrins do not readily exchange metal ions. Consequently, diverse methods have been employed to make metal-substituted hemoproteins.[41–46] Early on, copper, cobalt, nickel and manganese-substituted horseradish Peroxidase (HRP) were prepared by a multistep process that subjected protein to strong acid and organic solvents.[41,42] Variations of this method have been used repeatedly. [24,43,47–49] However, this method is applicable only to a narrow range of hemoproteins that tolerate the harsh treatment. With the advent of overexpression methods, significant improvement of metalloporphyrin-substituted protein yield was achieved by direct expression of the apoprotein and reconstitution with the desired metalloporphyrin in lysate prior to purification.[50] Although this approach has many virtues, direct expression of apoprotein is ineffective for many hemoproteins, again limiting utility.

As an alternative to the above *in vitro* approaches, researchers have pursued systems for direct *in vivo* expression of metal substituted hemoproteins. Two specialty strains of *E. coli* were engineered to incorporate metalloporphyrin analogs from the growth medium into hemoproteins during protein expression. The engineered RP523 strain cannot biosynthesize heme and bears an uncharacterized heme permeability phenotype.[51] Together, these two features enable this strain to assimilate and incorporate various metalloporphyrins into overexpressed hemoproteins with no background heme incorporation.[44,52–54] However, heme auxotrophy makes RP523 cells exceedingly sensitive to O_2 and, in many situations, RP523 cultures must be grown anaerobically. An alternative BL21(DE3)-based engineered strain harbors a plasmid bearing the heme transporter ChuA, which facilitates import of exogenous heme analogs.[45] Production of metalloporphyrin-substituted protein with this ChuA-containing strain relies on growth in iron-limited minimal media, thereby diminishing heme biosynthesis. This method was used successfully to express metal-substituted versions of the heme domain of cytochrome P450 BM3 [45] and several myoglobin variants.[11,12] Because these cells biosynthesize a small quantity of their own heme, they are more robust than the RP523 cells. This advantage does come with a small penalty, increased heme contamination in the product protein (2–5%).[45]

A set of intriguing papers reported the production of cobalt-substituted human cystathionine β-synthase (CoCBS) that relies on the *de novo* biosynthesis of CoPPIX from $CoCl_2$ and δ-aminolevulinic acid (δALA), a biosynthetic precursor to heme.[46,55] This method yielded significant amounts of CoCBS—albeit with modest heme contamination (7.4%)—sufficient for spectroscopic and functional characterization of the CoPPIX-substituted protein.[8,46] As cobalt is known to be toxic to *E. coli*, the researchers passaged the CBS expression strain through cobalt-containing minimal media for 12 days with the hypothesis that this procedure enabled the cells to adapt to high concentrations of cobalt prior to protein expression. It is plausible that this serial passaging metabolically alters the *E. coli* cells, enabling the biosynthesis of CoPPIX and *in vivo* production of metal-substituted protein. However, the adaptation process is slow (>10 days) and

it is unknown how genomic instability under these mutagenic conditions affects the reproducibility of this passaging approach.

The possibility of facile CoPPIX production is particularly attractive for future biocatalysis efforts. As described above, synthetic Co-porphyrins have been shown to perform a range of radical-mediated reactions. The ability to produce a CoPPIX center in vivo may enable engineering these unusual reactivities via directed evolution, in addition to spectroscopic applications. We therefore set out to explore the unusual phenotype of CoPPIX production by *E. coli* and to ascertain whether it was possible to efficiently biosynthesize cobalt-containing hemoproteins *in vivo* from a single 'generalist' cell line. Our goal was to achieve an efficient method of cobalt-substituted hemoprotein production with minimal contamination of the native cofactor. Herein, we report the surprising discovery that native *E. coli* BL21(DE3) can biosynthesize a new-to-nature CoPPIX cofactor. We use this insight to produce cobalt-substituted hemoproteins *in vivo* without requirement for complex expression methods or specialized strains.

2. 3. Results and Discussion

2. 3. 1. Cobalt-tolerant *E. coli* are not optimal for CoPPIX-substituted hemoprotein expression

We envisioned a strategy for producing CoPPIX-substituted hemoproteins that would utilize a simple cobalt salt as the metal source. Our initial approach was inspired by the procedure reported by Majtan *et al*. [46], in which a CBS-expressing strain was serially-passaged until direct expression of CoCBS was achieved. We hypothesized that this serial passaging method might select for mutant strains with a cobalt tolerance phenotype, enabling CoPPIX production. Therefore, we attempted to evolve a cobalt-tolerant strain that could be used as a 'generalist' to express any CoPPIX-substituted hemoprotein. To this end, we passaged *E. coli* BL21(DE3) —a commonly used laboratory expression strain—through Co-supplemented media to select for cobalt-tolerant phenotypes. Specifically, we subjected *E. coli* to 15 rounds of passaging over 12 days (Figure 2.2). To initiate each passage, the previous culture was diluted into fresh $CoCl_2$-containing, augmented minimal media (M9*) that was supplemented with low concentrations of δALA to facilitate porphyrin production. The concentration of $CoCl_2$ in the media was increased in a stepwise fashion after the cells achieved a steady, high growth rate and a stable cell density at stationary phase. The resultant strains were able to grow in much higher concentrations of cobalt than the parent strain, with an IC_{50} (the concentration at which growth is inhibited by 50%) that was 10-fold higher than that of the parent strain of *E. coli* (Table 2.1, Figure 2.3). Further experimental details may be found in the Material and Methods section (2.5).

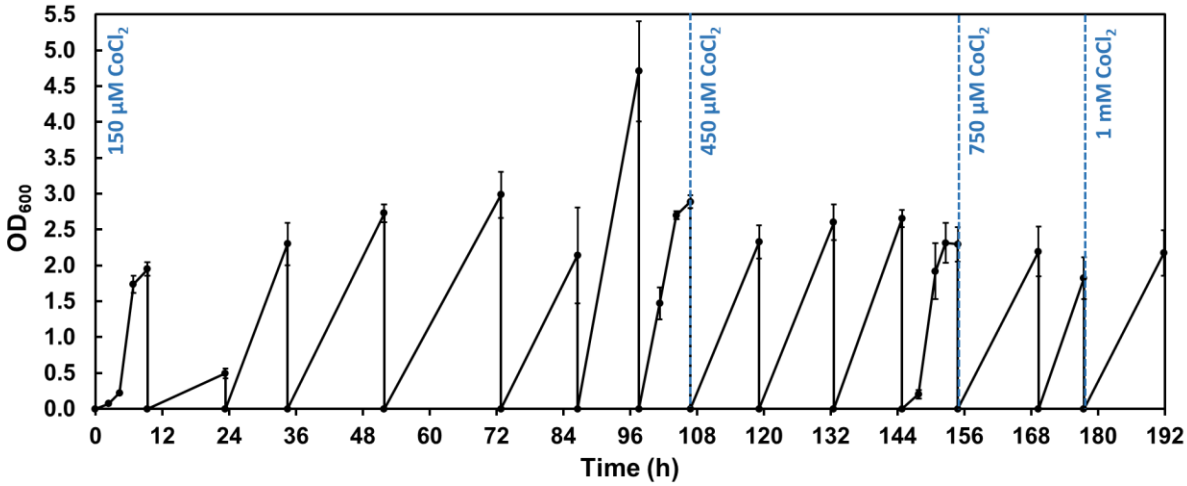


Figure 2.0.2. Growth of BL21(DE3) during passaging in $CoCl_2$-supplemented M9* media. The plotted OD_{600} values are the average of three biological replicates, which were passaged as independent lineages, A, S, and T (see the passaging procedure outlined in experimental methods for further description). Calculated standard deviations are shown by the error bars. The $CoCl_2$ concentration in the growth medium was increased during passaging, and these changes are demarked with blue dashed lines and the concentrations are noted in blue. We hypothesize that residual iron from the culture inoculum is responsible for the robust growth during the first passage.

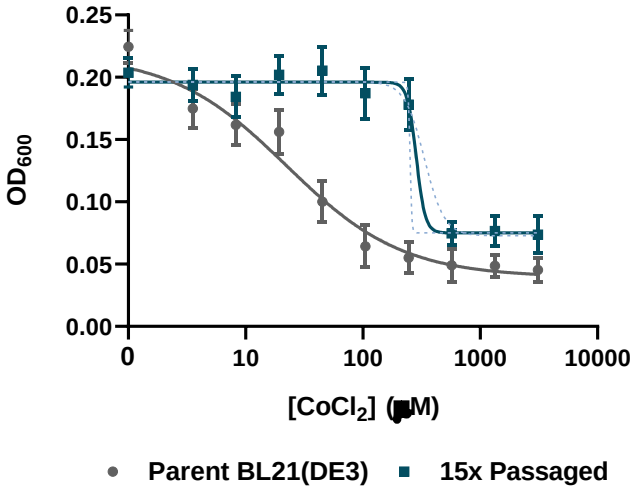


Figure 2.0.3. Optical densities of E. coli BL21(DE3) cultures, after 9 hours growth at variable concentrations of $CoCl_2$. In all graphs. Direct comparison of parent BL21(DE3) and data from all three passaged growths in variable concentration $CoCl_2$-containing media. Data points are the average of either nine (for the parent) or 27 (for the 15x passaged strain) individual data points, and error bars represent the standard deviations from the mean. The solid line represents the best fit to the data using the Hill equation (fit parameters given in Table 2.1). Blue dotted lines represent best fit curves when the Hill coefficient is set to specific values. More details of this analysis can be found in the IC_{50} assay description section of the experimental methods.

Table 2.0.1. Fit parameters for IC_{50} curves of parent BL21(DE3 and passaged strains as an aggregate. Because the passaged strain IC_{50} is not well defined, the fit parameters the result from constraining the Hill coefficient to the 95% confidence lower bound and an arbitrarily high value are also shown. The resulting IC_{50} value is 10-fold higher than that of the parent strain in any case.

Strain	Fit parameters [95% CI]				
	IC_{50}	Top	Bottom	Hill coefficient	R-squared
Parent (BL21(DE3))	20 [17, 29]	0.22 [0.206, 0.230]	0.040 [0.030, 0.047]	-1.1 [-1.24, -0.973]	0.93
15x Passaged (A15, S15, T15 analyzed together)	280 [???, ???]	0.20 [0.193, 0.199]	0.080 [0.070, ???]	-12 [???, -5.12]	0.91
	340 [314, 360]	0.20 [0.194, 0.200]	0.070 [0.069, 0.077]	-5.1[a]	0.91
	250 [251, 255]	0.20 [0.193, 0.199]	0.080 [0.071, 0.080]	-50[a]	0.91

[a]Hill coefficient is constrained to this value.

To determine whether this improved cobalt tolerance enabled production of a CoPPIX-substituted hemoprotein, we tested expression of a sperm whale Mb variant H64V V68A bearing a 6-His tag on the *C*-terminus (Mb*). This Mb variant, in which oxygen binding is eliminated by substitution of His64, was engineered for cyclopropanation of styrene [56] and recently was used by Sreenilayam *et al*. to carry out a variety of carbene transfer reactions using CoPPIX and other non-iron metalloporphyrins.[11] Importantly, electronic absorption spectra of the CoPPIX substituted Mb* were reported and the cobalt-substituted native Mb protein has been well-characterized [3–5,57,58]. Mb* was overexpressed in parent BL21(DE3) cells, and cells from one passaged strain (strain A15, see SI), in augmented minimal media (M9*) supplemented with 500 μM $CoCl_2$ and 250 μM δALA. The resulting Mb* proteins were purified by metal-affinity chromatography. To compare the effect of passaging on the metalloporphyrin incorporation, the two purified Mb* samples were analyzed for the presence of CoPPIX and FePPIX by basic extraction of the pyridine (py)-coordinated metalloporphyrin complexes.[59] The Co(II)PPIX(py) and Fe(II)PPIX$(py)_2$ complexes exhibit distinct visible spectra, which can be used for metalloporphyrin identification.[46]

Expression in parent and passaged strains resulted in an appreciable amount of CoPPIX-substituted Mb* (Figure 2.4), with protein yields of 3–5 mg L^{-1} culture.

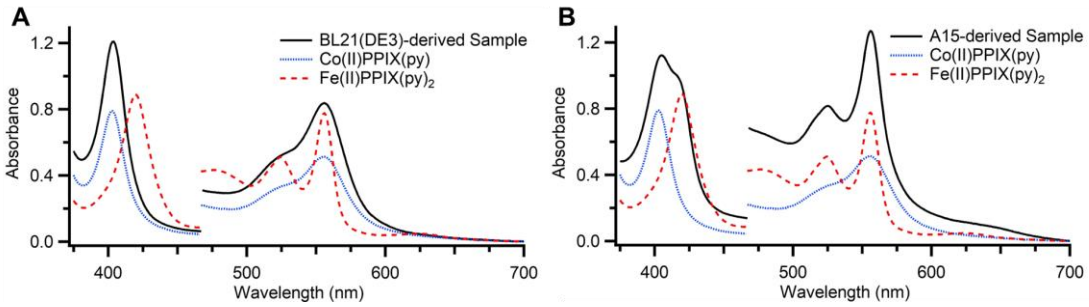


Figure 2.0.4. Comparison of electronic absorption spectra of M(II)PPIX(py)$_X$ complexes derived from CoPPIX-substituted Mb* expressed in (A) E. coli BL21(DE3) and (B) $CoCl_2$-passaged E. coli strain A15. Standard spectra of M(II)PPIX(py)$_x$ complexes (7.2 µM), prepared from Co(III)PPIX(Cl)and Fe(III)PPIX(Cl) standards, are shown for comparison. The standard spectrum of the hemochrome, Fe(II)PPIX(py)$_2$ is a red dashed line and that of Co(II)PPIX(py) is a blue dotted line. All pyridine complexes were generated using the modified pyridine hemochrome assay procedure outlined in experimental methods. Spectra above 466 nm are multiplied by five for clarity

To our surprise, and initial disappointment, the Mb* purified from the cobalt-resistant strain contained significant amounts of FePPIX. In contrast with the results from Majtan *et al*. [46], where the fraction of native heme in the isolated CoCBS was less than 10%, in Mb* the visible spectrum of the pyridine-extracted cofactor suggests that close to half of the incorporated metalloporphyrin cofactor was heme (Figure S2). This observation was unexpected given the iron limited growth conditions. Inductively coupled plasma mass spectrometry (ICP-MS) analysis of the iron content in the M9* expression medium revealed less than 1 ppb Fe. We hypothesize that the cobalt-tolerant strain has an increased capacity to assimilate iron and to discriminate iron from cobalt through unknown mechanisms. While passaging the BL21(DE3) cells, we deliberately increased the cobalt concentration as the cells adapted, with the result that the passaged cells grew significantly better in high cobalt (1 mM $CoCl_2$) media. This cobalt concentration was significantly higher than that employed by Majtan *et al*. (150 µM).[46,55] It is of note that in the prior CoCBS expression, the passaged cells contained a plasmid encoding either CBS (a heme-containing protein) or an FeS cluster containing protein. In our experiments, no plasmid was present during

the passaging process. Whether these differences altered the outcome of the passaging process, and its previously reported ability to cleanly produce CoPPIX-containing protein, is unknown. Remarkably, the Mb* sample expressed in the parent BL21(DE3), which was intended as a negative control, contained high levels of CoPPIX with no apparent FePPIX contamination. The visible spectrum of the pyridine-extracted cofactor from isolated Mb* is identical to that of the CoPPIX standard, with no evidence of heme (Figure 2.4). The efficient production of CoMb* was completely unexpected, especially since the cells were grown at a toxic cobalt concentration (500 μM). The measured IC_{50} for $CoCl_2$ in *E. coli* BL21(DE3) was 22 μM, (Figure 2.3), yet cells expressing Mb* grew effectively and produced good yields of CoMb*, motivating further investigation of these observations.

2. 3. 2. *E. coli* BL21(DE3) can biosynthesize CoPPIX *de novo* in cobalt-rich, iron-poor conditions

To characterize the effects of cobalt on *E. coli* growth and porphyrin production, we measured the ability of *E. coli* BL21(DE3) to grow in the presence and absence of high levels of iron and cobalt (500 μM), while supplementing with exogenous δALA to drive heme biosynthesis. Growth curves comparing the effects of these additives to M9* media on *E. coli* growth rate are shown in Figure 2.5A. Cells were harvested from each culture when growth reached stationary phase, and porphyrins were extracted with acidified pyridine and analyzed via ultra-performance liquid chromatography mass spectrometry (UPLC-MS) (Figure 2.5B). *E. coli* BL21(DE3) cells grew well in M9* broth and produced relatively little porphyrin. Cell cultures supplemented with δALA and no extra metal grew more slowly and produced a large quantity of free base protoporphyrin IX (H_2PPIX). Cell cultures co-supplemented with δALA and $FeCl_3$ showed a dramatic increase in growth rate to significantly higher cell density. These cells produced substantial FePPIX, with a small amount of residual H_2PPIX. Cultures co-supplemented with

δALA and $CoCl_2$ showed a dramatic lag in growth of almost 24 h and grew to a lower cell density in stationary phase. This result is likely a consequence of cobalt-induced disruption of iron metabolism, which is known to cause defects in iron-sulfur cluster biosynthesis and induce oxidative stress.[60,61] Cell pellets isolated from these cobalt-supplemented cultures were distinctly pink, suggesting the presence of a new porphyrin species (Figure 2.5C). Indeed, extraction and analysis revealed significant CoPPIX content in these cells. Free base H_2PPIX was below the limit of detection and the amount of FePPIX produced was comparable to that of the δALA-supplemented sample that lacked additional $FeCl_3$. These data clearly show that, in the presence of near-toxic levels of $CoCl_2$ and without any passaging in Co-rich media, *E. coli* BL21(DE3) cells produce significant amounts of CoPPIX.

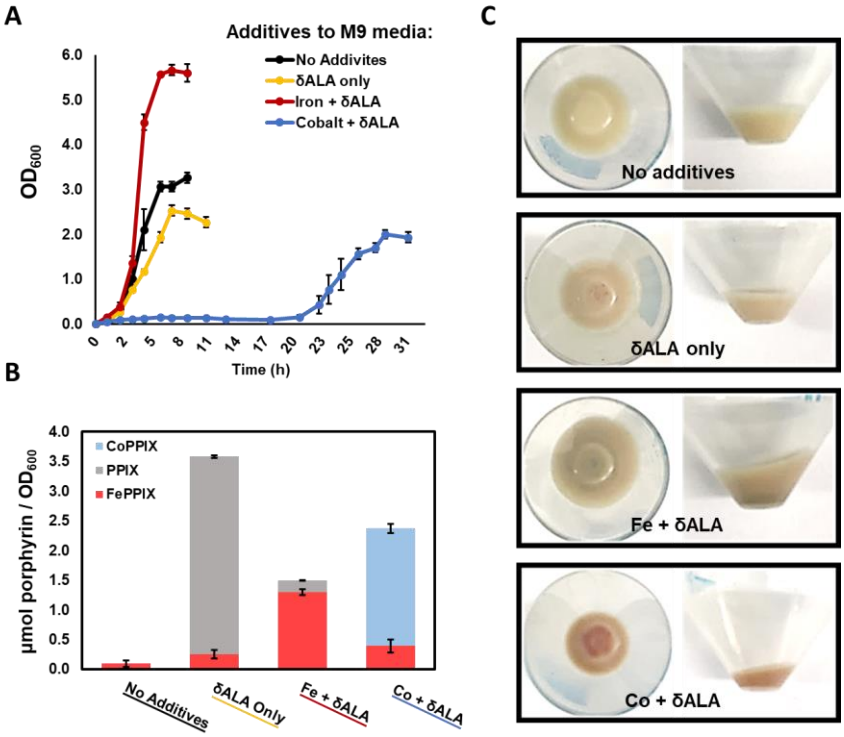


Figure 2.0.5. The effect of iron, cobalt and δALA on E. coli BL21(DE3) growth and porphyrin production. A) OD_{600} values monitoring cell growth over time in M9* media without additives (black), and supplemented with 250 mM δALA (yellow), 500 μM $FeCl_3$ and 250 mM δALA (red), and 500 mM $CoCl_2$ and 250 μM δALA (blue). Error bars are the standard deviation among OD_{600} values from three separate cultures. B) Porphyrin content in cells harvested in stationary phase from (A) as measured by LC-MS. Amount of porphyrin calculated is shown relative to the OD_{600} of cells when collected. See SI Experimental Methods for details..C) Photographic images of cell pellets of BL21(DE3) E. coli grown in M9* media with various supplements.

This observation raises questions as to the origin of this non-native metalloporphyrin cofactor. The final step of heme biosynthesis is divalent iron insertion into H_2PPIX, catalyzed by the enzyme ferrochelatase. Notably, the ferrochelatase homologs from *H. sapiens*, *M. musculus*, *S. cerevisiae* and *B. subtilis* have exhibited promiscuous activity *in vitro* with the alternative divalent metal ions cobalt and zinc.[48–51] Although the metal specificity of *E. coli* ferrochelatase has not been determined (and is the focus of Chapter 3), the data reported here suggest that this enzyme may possess promiscuous activity with cobalt, resulting in the biosynthesis of CoPPIX when *E. coli* cells are grown in the presence of high concentrations of cobalt and minimal iron. Given the ability of *E. coli* to biosynthesize CoPPIX, we sought to explore the properties of BL21(DE3) as an expression strain for cobalt-substituted hemoproteins.

2. 3. 3. Overexpression of heme-containing protein enhances cobalt tolerance of *E. coli* BL21(DE3)

To understand the interplay of hemoprotein overexpression and cobalt tolerance and inform development of an expression protocol, we tested the effect of heme and non-heme protein overexpression on *E. coli* cobalt tolerance. We hypothesized, based on our observations and the prior results of Majtan et al., [55] that overexpression of a hemoprotein in cobalt-supplemented broth may help to alleviate cobalt toxicity by enabling the cells to sequester cobalt within the CoPPIX-substituted hemoprotein. As our non-heme, cofactor-containing control that is expressed at high levels, we chose the PLP-containing β-subunit of tryptophan synthase from *Pyrococcus furiosus* (*Pf*TrpB) [62]. We tested whether expression of heme-containing protein Mb* conferred greater tolerance to $CoCl_2$ than expression of *Pf*TrpB. To measure the IC_{50} of $CoCl_2$ on *E. coli* growth, cells containing isopropyl β-D-1-thiogalactopyranoside (IPTG)-inducible, protein-encoding plasmids were grown in 96 well plates with variable concentrations of IPTG and $CoCl_2$. Growth was measured via OD_{600} using a plate reader (Figure 2.12). Overexpression of *Pf*TrpB in

unsupplemented M9* broth results in a small decrease in cell density relative to non-overexpressing (no IPTG added) controls, consistent with diversion of resources to protein production (Figure 2.6A). Nevertheless, the IC_{50} for $CoCl_2$ was not significantly altered upon overexpression of the *Pf*TrpB protein. As was observed upon expression of *Pf*TrpB, expression of Mb* in unsupplemented M9* broth results in a small decrease in cell density. In contrast to the *Pf*TrpB strain, overexpression of Mb* also causes a significant increase in the IC_{50} for $CoCl_2$ (Figure 2.6B), revealing that hemoprotein overexpression indeed helps to alleviate cobalt toxicity. Therefore, we adopted an approach to expressing CoPPIX-substituted hemoproteins utilizing relatively high concentrations of $CoCl_2$ (500 μM), where hemoprotein overexpression exerts a selective advantage on cell growth.

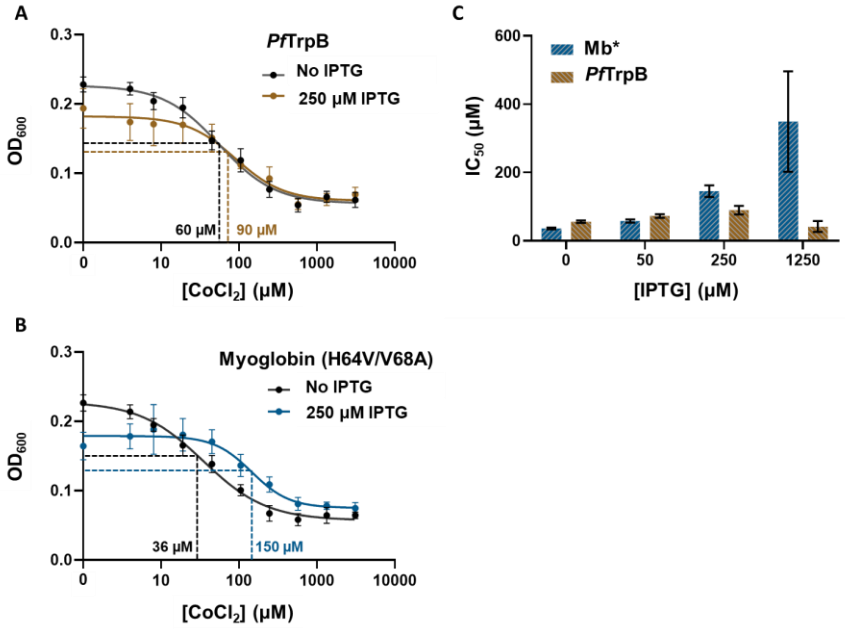


Figure 2.0.6. Effect of over-expressing a heme protein or a non-heme protein on cobalt tolerance of E. coli BL21(DE3). OD_{600} measured after 9 hours of growth in variable amounts of $CoCl_2$. A) Effect of inducing expression of PfTrpB, a non-heme cofactor-containing protein. B) Effect of inducing expression of P. catodon Mb*, a hemoprotein. In each graph, the grey line represents the IPTG-free, uninduced control. Error bars represent the standard deviation of 12 measurements, as described in SI Experimental Methods. C) Effect of protein expression on cobalt chloride IC50. IC50 values are calculated from the line of best fit for each strain at a given IPTG concentration. Error bars represent the standard error of the fit for the IC50 parameter, which are also listed in Table 2.2.

Table 2.2. Fit parameters for IC_{50} assay: Effect of expressing a hemoprotein of cobalt toxicity

Growth conditions		Fit parameters [95% CI]				
	[IPTG] (μM	IC_{50}	Top	Bottom	Hill coefficient	R-squared
BL21(DE3) bearing a Mb* plasmid	0	40 [31, 41]	0.23 [0.220, 0.232]	0.057 [0.052, 0.062]	-1.1 [-1.24, -0.973]	0.97
	50	60 [49, 69]	0.22 [0.210, 0.225]	0.062 [0.054, 0.069]	-1.9 [-2.41, -1.46]	0.93
	250	150 [115, 187]	0.18 [0.173, 0.185]	0.075 [0.066, 0.083]	-1.8 [-2.50, -1.33]	0.84
	1250	350 [165][a]	0.14 [0.131, 0.151]	0.062 [0.082][b]	-1.3 [-4.33, -0.322]	0.49
BL21(DE3) bearing a *Pfu*TrpB plasmid	0	60 [49, 64]	0.23 [0.220, 0.232]	0.056 [0.050, 0.062]	-1.2 [-1.42, -1.06]	0.96
	50	70 [63, 84]	0.20 [0.199, 0.210]	0.057 [0.051, 0.063]	-1.6 [-1.90, -1.30]	0.96
	250	90 [68, 118]	0.18 [0.174, 0.192]	0.060 [0.049, 0.069]	-1.4 [-1.98, -0.946]	0.85
	1250	40 [22, 124]	0.21 [0.193, 0.221]	0.051 [0.013, 0.069]	-0.58 [-0.819, -0.392]	0.78

[a]Value is a maximum, minimum value is ill-defined
[b]Value is a minimum, maximum value is ill-defined

2. 3. 4. *E. coli* BL21(DE3) is an effective host for *in vivo* production of CoMb*

We next established a method to produce CoMb* on a preparative scale using a two-step induction procedure. *E. coli* BL21(DE3) cells were grown in M9* broth (1.0 L in a 2.8 L Fernbach flask) to an intermediate OD_{600} of 0.2–0.3, whereupon δALA (250 μM) and $CoCl_2$ (500 μM) were added to stimulate formation of CoPPIX. Cells were grown in this supplemented medium to an OD_{600} of 0.6-0.8, after which Mb* expression was induced by addition of IPTG (250 μM). The overexpressed Mb* protein was purified using Ni-affinity chromatography and stored at -80 °C.

The logic of this expression protocol, where the cells were grown in unsupplemented M9* to a moderate cell density and then with δALA and $CoCl_2$, prior to induction of Mb* expression, is as follows. We observed that cells inoculated at low density in cobalt-containing media exhibit a substantial growth lag, which may be offset by allowing the cells to grow in cobalt-free media. Further, we were aware that the medium contained trace iron. Given the high efficiency with which iron is incorporated into heme, the pre-incubation period was designed to allow full incorporation of all available heme into native hemoproteins. In the next phase, addition of δALA and $CoCl_2$, CoPPIX is produced. Under these conditions, we anticipated that CoPPIX was present and available for incorporation into Mb* at the time of induction.

While this method provides a pool of CoPPIX, residual FePPIX or H_2PPIX might also be available for incorporation into hemoproteins. We therefore assessed the identity of the porphyrin cofactor bound within the isolated Mb* using three independent methods: pyridine extraction assay, ICP-MS, and UPLC-MS. First, we extracted the porphyrin cofactors from purified Mb* and characterized the metalloporphyrin pyridine complexes using visible spectroscopy. The spectral peak positions and band shapes for the extracted porphyrins closely matched that of a Co(II)PPIX(py) standard (Figure 2.7). The absence of any shoulder at 424 nm, the Soret peak position of Fe(II)PPIX(py)$_2$, suggested that there was minimal FePPIX in the Mb* protein. To quantify the metal content of Mb* more rigorously, we employed ICP-MS. From ICP-MS analysis, we determined that the vast majority of the metal content (>99.5%) was cobalt, and less than 0.5% was iron (Table 2.3). We further considered the possibility that the total metalloporphyrin content and metal content may not fully reflect the entire porphyrin content in the protein sample, as some porphyrin may be unmetallated. Therefore, we assessed the *relative* amounts of H_2PPIX, FePPIX and CoPPIX by UPLC-MS. UPLC-MS analysis of isolated Mb* confirmed that the vast majority of the porphyrin present in the protein was CoPPIX, representing greater than 99% of the total porphyrin (Table 2.3). FePPIX and H_2PPIX were present at less than 1% total porphyrin. These three experiments together revealed that the isolated Mb* contained CoPPIX essentially to the exclusion of other porphyrin cofactors.

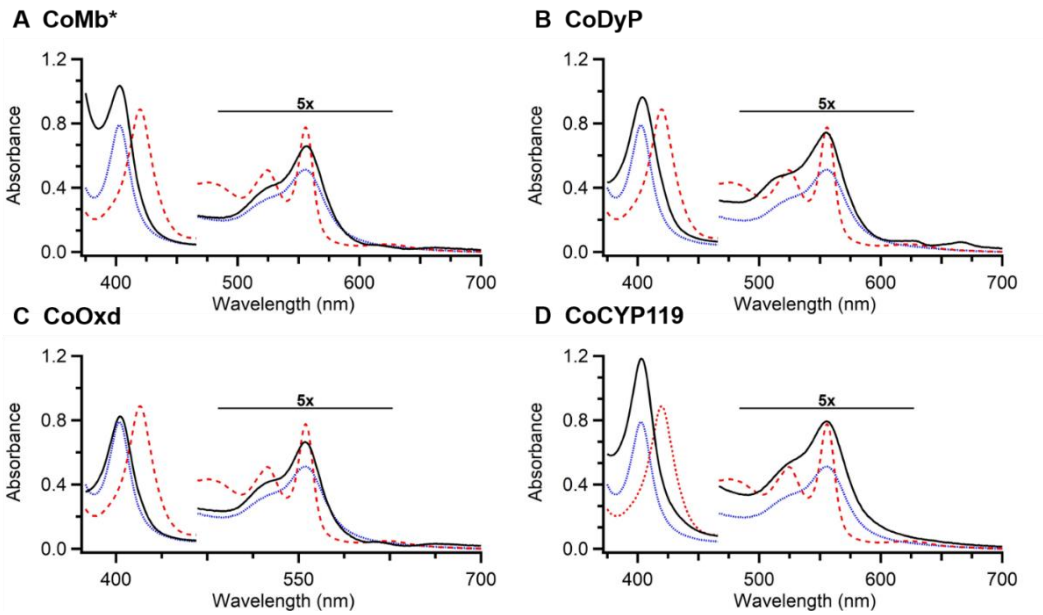


Figure 2.0.7. Comparison between the visible spectra of the Co(II)PPIX(py) complexes derived from Co(III)PPIX-substituted hemoproteins A) Mb*, B) DyP, C) CoOxd, D) CYP119 and those of reference standards. The black traces are the electronic absorption spectra of the M(II)PPIX(py)$_x$ complexes derived from CoPPIX-substituted hemoproteins with NaOH and pyridine. Spectra of 7.2 μM Fe(II)PPIX(py)$_2$ (red dashed) and Co(II)PPIX(py) (blue dotted) are shown for reference. Absorbance above 466 nm are multiplied by five for clarity.

Table 2. 3. Comparison of the relative porphyrin content, relative metal content, cofactor loading, and protein yield for CoPPIX-substituted hemoproteins produced by overexpression in $CoCl_2$-supplemented media.

Protein	Relative Porphyrin Content			Relative Metal Content		% holoprotein[a]	Yield (mg protein /L culture)[b]
	CoPPIX	FePPIX	H_2PPIX	Co	Fe		
CoMb*	> 99%	0.1 ± 0.1%	0.1 ± 0.1%	99.8 ± 0.3%	0.2 ± 0.3%	29 ± 2%	4.7
CoCYP119	99 ± 1%	< 1%	< 1 %	> 99.4%	< 0.6%	101 ± 2%	3.6
CoOxd	>99%	0.3 ± 0.7%	< 3%	95.9 ± 0.9%	4.1 ± 0.9%	63 ± 1%	3.7
CoCat	99 ± 1%	0.8 ± 0.1%	<0.1%	>99%	<1%	54 ± 1%	2.2
CoDyP	74 ± 1%	1.1 ± 0.4%	25 ± 1%	99.4 ± 0.4%	0.6 ± 0.4%	49 ± 0.2%	6.5
FeDyP	< 5%	70 ± 2%	25 ± 1%	NA[c]	NA	22 ± 3[d]	NA

[a]Percentage of [Co] (measured by ICP-MS) versus [protein] (measured by BCA assay)
[b]Calculated as mg of holoprotein per L culture. Yields were reproducible within a factor of 2 between batches.
[c]NA, not assessed
[d]Percentage of [heme] (measured by pyridine hemochrome assay) versus [protein] (measured by BCA assay)

Because heterologous overexpression may result in significant apo-protein, we assessed the fraction of isolated Mb* that was loaded with CoPPIX. The total cobalt concentration was measured by ICP-MS and compared to the protein concentration, as measured by bicinchoninic acid (BCA) assay. According to the measured CoPPIX to total Mb* ratio (Table 2.3), the isolated Mb* protein is only 30% loaded with CoPPIX. This observation suggests two non-exclusive possibilities. First, Mb* production outpaces CoPPIX production under the expression conditions. Second, CoPPIX may not bind efficiently to Mb*. Sub-stoichiometric cofactor loading is frequently observed for hemoproteins, and has previously been ascribed to incommensurate rates of apoprotein production and heme biosynthesis under overexpression conditions. [63,64]

We confirmed that native histidine coordination (Figure 2.8A) was maintained in CoMb* using electronic absorption and electron paramagnetic resonance (EPR) spectroscopies. Electronic absorption spectra of the as-isolated Co(III)Mb* and dithionite-reduced Co(II)Mb* (Figure 2.8B) are similar to those of CoMb* and CoMb previously produced by two different routes: *in vivo* expression of Mb* in engineered *E. coli* cells capable of importing CoPPIX from the media [11], and by reconstitution of apoMb with CoPPIX [24] (Table 2.4). In the Co(III)Mb* spectrum, we observe a major set of peaks, listed in Table 2.4, and a minor set of peaks, the latter appearing at 404, 520, and 561 nm. The presence of two sets of peaks clearly implicates the presence of two differently coordinated species. The major species exhibits peak maxima comparable to those of Co(III)Mb, where cobalt is six-coordinate bound by His94 and a water molecule (Table 2.4). The identity of the minor species is less clear. The distal pocket of this Mb* variant is hydrophobic and lacks the usual hydrogen-bond donating His64. We suggest the minor species is a five-coordinate complex lacking the sixth water ligand based on analogy to studies of other cobalt porphyrin complexes [65,66]. The relatively slow reduction of Co(III)Mb*, which requires 30 minutes to complete, is consistent with previous observations of the behavior of CoMb.[57]

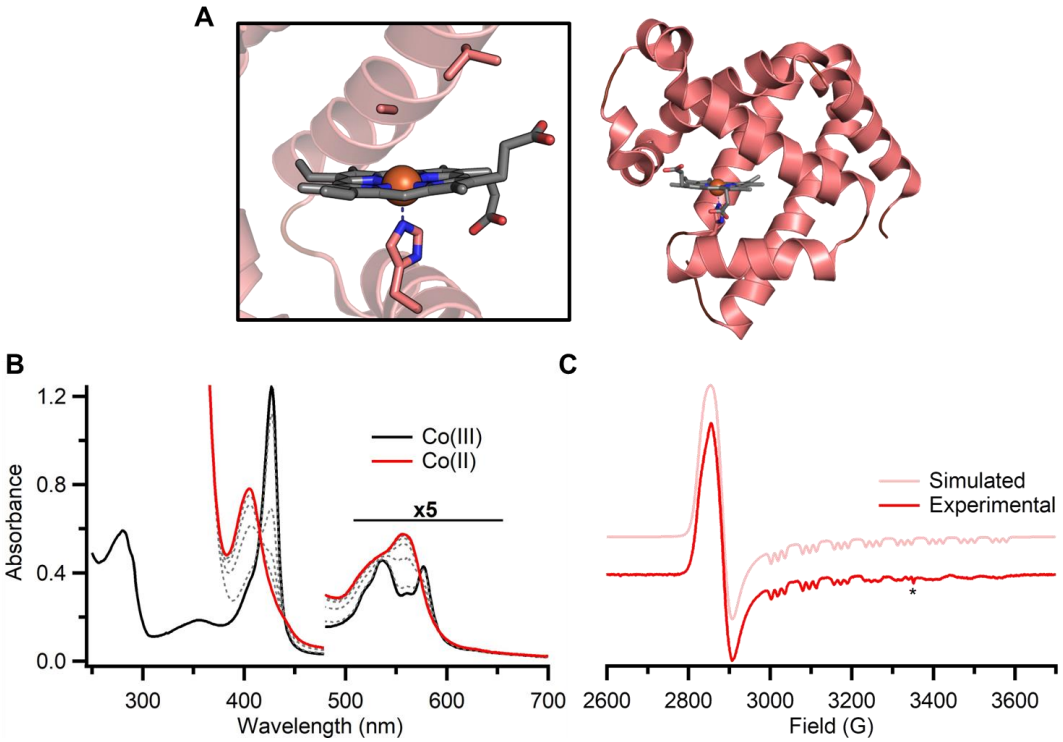


Figure 2.0.8. Spectroscopic characterization of CoPPIX-substituted Mb* produced by overexpression in $CoCl_2$-supplemented media. (A) Structure of sperm whale myoglobin H64V/V68A (Mb*) (PDB: 6G5T) shown as cartoon highlighting helical secondary structure. Structure of Mb* surrounding heme highlighting iron axial ligation by His93 and residues altered in the distal pocket. (B) Electronic absorption spectra of Co(III)Mb* (black) and Co(II)Mb* (red) at 35°C in Ar-sparged 50 mM potassium phosphate buffer, pH 7. Reduction was initiated by addition of $Na_2S_2O_4$ to a final concentration of 4 mM. Spectra tracking the progress of reduction are shown as gray dotted lines. (C) Experimental (red) and simulated (pink) EPR spectra of Co(II)Mb*. Parameters used for spectral simulation are provided in Table 2. Acquisition parameters are provided in SI Experimental Methods. Spectroscopic data collected and analyzed by B.R.W.

To conclusively establish the presence of the native histidine ligand in CoMb*, we turned to EPR spectroscopy. The EPR spectrum of CoMb* (Figure 2.8C) matches those previously reported for CoMb prepared by other methods [4]. The presence of triplet super-hyperfine coupling unequivocally confirms that cobalt is coordinated by a nitrogen donor ligand. As previously observed for CoMb, the Co(II)Mb* EPR spectrum is axially symmetric, consistent with a low-spin Co(II) complex with an unpaired electron in dz^2. The octaplet hyperfine and triplet super-hyperfine structures are indicative of coupling to S = 7/2 ^{59}Co and S = 1 ^{14}N nuclei, respectively (Figure 3.8C). Spin Hamiltonian parameters derived from simulation (Table 2.5) match well with those previously reported for wild-type sperm whale deoxy-Co(II)Mb,[4] in which the nitrogen super-

hyperfine coupling is attributed to axial nitrogen coordination by His93. These spectroscopic data are wholly consistent with a natively bound CoPPIX cofactor in Mb*.

Table 2.4. Comparison of electronic absorption maxima of CoPPIX-substituted hemoproteins produced in vivo with analogous proteins produced through in vitro reconstitution with CoPPIX.

Protein	Co(III) λ_{max} (nm)				Co(II) λ_{max} (nm)			ref.
	Ligands	Soret (γ)	β	α	Ligand	Soret (γ)	α/β	
Mb	His/H_2O	424	535	572	His	406	558	32
Mb*	His/H_2O	428	540	572	His	406	555	10
Mb*	His/H_2O	427	536	577	His	406	557	This work
$P450_{cam}$	Cys/H_2O	422	538	570	Cys	404	556	34
CYP119	Cys/H_2O	420	539	566	Cys	410	558	This work
Oxd	His/H_2O	429	542	574	His	402	559	This work
HRP(B,C)	His/H_2O	421	533	565	His	401	553	32
DyP	His/H_2O	422	534	567	His	403	564	This work
Cat	N/?	427	537	570	N	403	561	This work

Table 2.5. Comparison of EPR parameters for Co(II)PPIX-substituted hemoproteins produced using in vivo expression, with those of analogous proteins produced using in vitro reconstitution with CoPPIX.

Protein	Ligand	Effective *g* value		$A^{Co}_{\perp}$	$A^{Co}_{\parallel}$	$A^{N}_{\perp}$	$A^{N}_{\parallel}$	Ref
		$g_{\perp}$	$g_{\parallel}$					
Mb	His	2.32	2.03	6	79	NR[a]	17	3
Mb*	His	2.32	2.04	2.4	77.6	2.2	16.9	This work
$P450_{cam}$	Cys	2.32	2.03	2	71			34
CYP119	Cys	2.27	2.03	1.9	67.8			This work
Oxd	His	2.31	2.02	7.2	83.4	2.3	16.7	This work
HRP (A)	His	2.34	2.03	11.5	73.9	NR[a]	17.3	32
DyP	His	2.30	2.04	10.4	73.1	6.6	16.1	This work

[a]NR, not reported

Encouraged by our success with expression of CoMb*, we considered to what extent other CoPPIX-substituted hemoproteins might be produced by this method. We chose myoglobin as a starting point because it is an excellent model protein on which to test CoPPIX incorporation: apoMb is relatively stable and easy to express, heme has previously been introduced by multiple routes, and spectroscopic data on CoMb are available for comparison. While the stability and versatile cofactor scavenging properties of apoMb* are beneficial for incorporating non-native porphyrins, its stability may be responsible for the large proportion of apoMb* that we observed in isolated CoMb*. This stability of the apo form of the protein, paired with the solvent-accessible heme-binding pocket may facilitate i*n vitro* reconstitution. In contrast, *in vivo* heme insertion into myoglobin is hypothesized to be co-translational,[67,68] and may also be aided by chaperones.[69,70] A more stringent challenge of our method of direct *in vivo* expression is production of diverse hemoproteins, including those for which *in vitro* reconstitution is inefficient or untested.

2. 3. 5. Direct *in vivo* expression is a generalizable method for production of CoPPIX-substituted hemoproteins

We assembled a panel of hemoproteins from diverse fold families and with different size, axial ligation, and biological function (Figure 2.9) and tested their expression in cobalt-supplemented media using the approach that was successful for CoMb*. A practical expression method should give good yields of CoPPIX-loaded proteins, with minimal FePPIX background, for virtually any *E. coli*-expressible hemoprotein. The proteins we targeted belong to classes of hemoproteins with established or potential biotechnological applications, such as biocatalytic carbene and nitrene transfer (Cytochrome P450),[71] production of nitriles from aldoximes (aldoxime dehydratase),[72] and oxidative lignin depolymerization (dye-decolorizing peroxidase) [73]. Furthermore, all of these proteins were previously expressed in *E. coli* with native FePPIX bound.[74–76] The expression method described above for Mb* afforded the CoPPIX-substituted version of each hemoprotein with good yields (3–7 mg protein/ L culture), reasonable cofactor loading (at least 40% holoprotein in all cases) and minimal FePPIX as measured by UPLC-MS (Table 2.3). No method optimization was attempted, but we anticipate that expression efficiency could be tuned individually for any target protein. Notably, expression of CoCYP119 was particularly efficient: this protein was isolated in good yield, fully cofactor loaded and with >99% CoPPIX bound in the active site. CoOxd was isolated in similarly good yield, with the majority of protein isolated as cofactor-containing holoprotein bearing CoPPIX. Within the holoprotein fraction, the vast majority (>95%) bore CoPPIX with minimal FePPIX or H_2PPIX content. In contrast, DyP is produced in exceptionally high yield and we observed a high proportion of apoDyP (51%) and H_2PPIX-bound DyP (25%) (Table 2.3). Isolation of a significant fraction of H_2PPIX-bound DyP upon overexpression in *E. coli* was noted previously,[77] and we were curious whether our cobalt-supplemented expression method resulted in an increased fraction of either apoDyP or H_2PPIX-bound DyP compared to typical expression procedures of this hemoprotein.

To this end, we compared the cofactor loading resulting from our cobalt-supplemented expression to that of a typical expression in rich media and found that the cofactor loading profiles for both expression methods were similar (Table 2.3). As in the case of Mb*, the DyP apoprotein appears to be quite stable, and we surmise from the high protein yield that heme or CoPPIX production lags behind protein production, leading to significant amounts of apoprotein. These observations establish that the efficiency of CoPPIX incorporation into DyP closely parallels that of FePPIX.

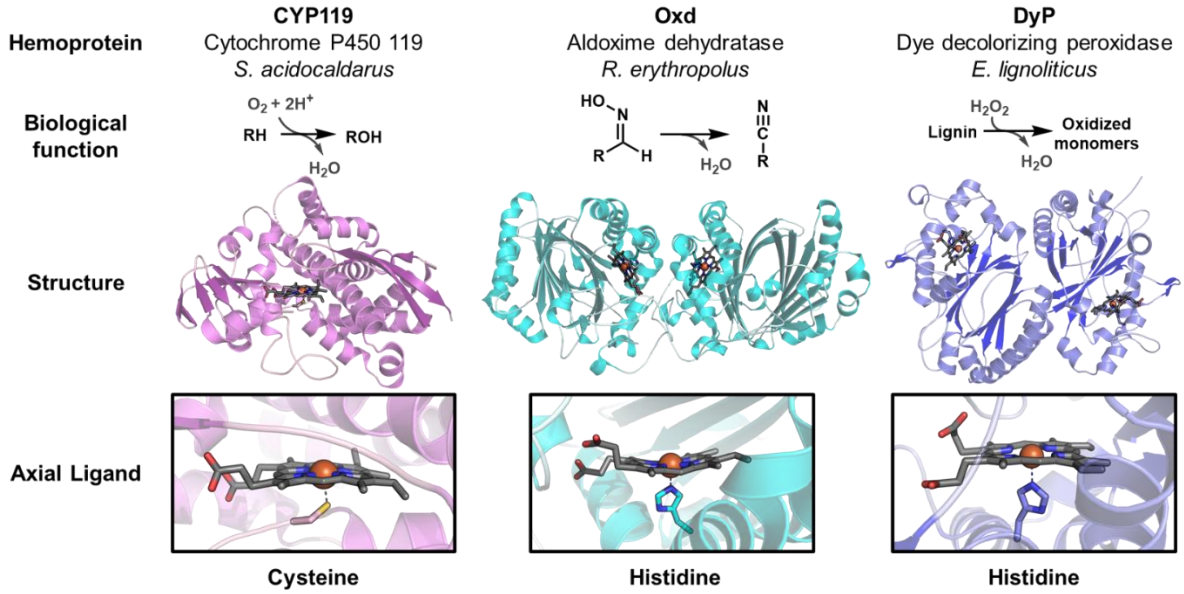


Figure 2.0.9. Characteristics of hemoproteins chosen for CoPPIX-substituted hemoprotein production. Each protein is listed with name, abbreviation, native organism, and biological function. Structures of CYP119 (PDB: 1F4U), Oxd (PDB: 3A15), and DyP (PDB: 5VJ0) are shown as cartoons highlighting secondary structure. Structures of heme b in each protein highlighting axial ligation are shown. In all structures, iron is shown as an orange sphere, and protoporphyrin IX and axial ligands are shown as sticks.

We characterized the environment of CoPPIX in these varied protein contexts using electronic absorption spectroscopy. Absorption spectra collected of the as-isolated and dithionite-reduced proteins were consistent with CoPPIX incorporation, and with retention of native axial coordination to the cobalt. The visible spectrum of Co(III)CYP119 exhibited peak maxima comparable to those of CoP450cam (Figure 2.10, Table 2.5). Reduction of Co(III)CYP119 was rapid, occurring in less than 2 minutes. Interestingly, the visible spectrum and facile reduction

behavior of Co(III)CYP119 were distinct from those of CoP450cam, which had been prepared through reconstitution of purified apo-protein with CoPPIX.[49] In the absence of substrate, reduction of Co(III)P450cam required >60 min, but in the presence of substrate, complete reduction occurred within 10 minutes. Although CoPPIX-substituted Oxd has not been prepared previously, the native coordination is analogous to that of Mb, with a single His ligand bound to the heme iron. The absorption maxima of Co(III)Oxd are comparable to that of Co(III)Mb and Co(III)Mb*, suggesting that the native histidine is coordinated to cobalt (Figure 2.10B). Reduction of Co(III)Oxd was slow and incomplete. A peak attributable to Co(III)Oxd remained at 429 nm after incubation with 4 mM sodium dithionite for 90 minutes at 35°C. The absorption maxima for Co(II)Oxd were comparable to those of Co(II)Mb and Co(II)Mb*. The visible spectra of CoDyP are more complex due to the presence of significant amounts of H_2PPIX-bound DyP (Figure 2.10C). The Co(III)DyP sample exhibited absorption maxima at 618 nm and 664 nm, which are similar to Q bands previously observed for H_2PPIX-bound *Tfu*DyP at 622 nm and 663 nm.[77] However, the loss in intensity of these bands upon reduction suggests they might arise from charge transfer rather than from H_2PPIX. Absorption maxima attributable to Co(III)DyP at 422 nm, 534 nm and 567 nm are at comparable positions to those of Co(III) horseradish peroxidase (HRP), again suggesting retention of the native histidine ligand. Co(III)DyP reduced sluggishly to Co(II)DyP, at a rate similar to the reduction of CoMb* and requiring 45 min to reach completion. The resulting absorption maxima attributable to Co(II)DyP are similar to those of Co(II)Mb and Co(II)HRP.

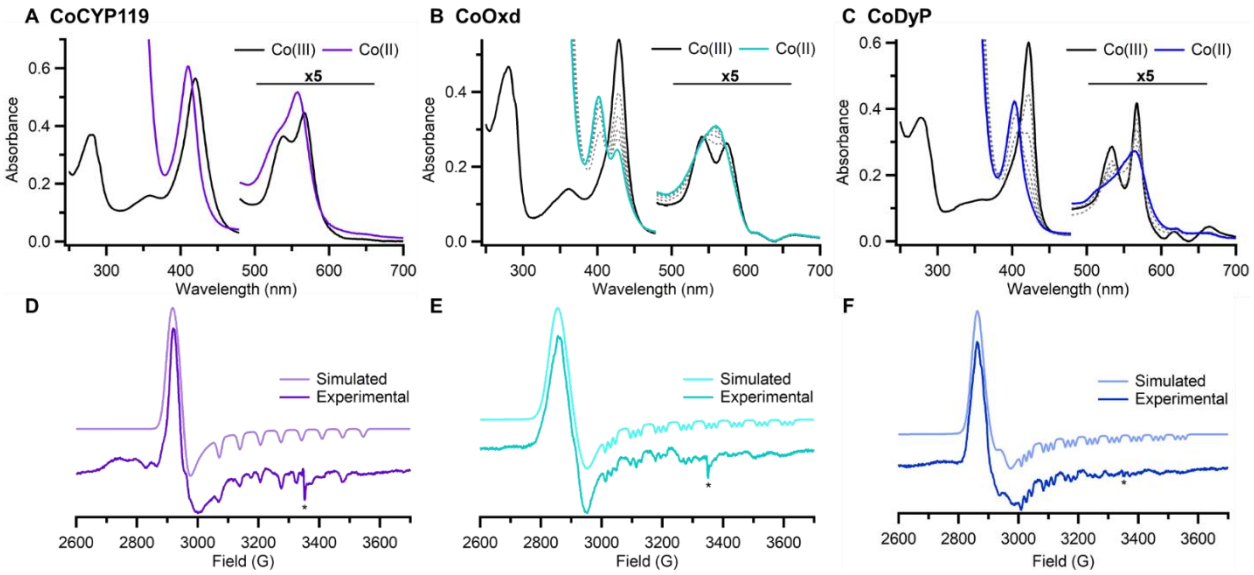


Figure 2.0.10. Spectroscopic characterization of other CoPPIX-substituted hemoproteins produced by overexpression in $CoCl_2$-supplemented media. (A-C) Electronic absorption spectra of CoCYP119, CoOxd, and CoDyP at 35°C in Ar-sparged 50 mM potassium phosphate buffer, pH 7. Co(III) spectra are shown in black. Co(II) spectra are shown in color. Reduction was initiated by addition of $Na_2S_2O_4$ to a final concentration of 3–4 mM. Spectra tracking the progress of reduction are shown as gray dotted lines. No intermediate spectra are shown for CoCYP119 as complete reduction occurred prior to the first spectral scan after addition of reductant. (D-F) EPR spectra of Co(II) hemoproteins. Darker colors are experimental spectra. Lighter colors are simulated spectra. Asterisks denote a consistent signal derived from sample cavity. Parameters from spectral simulations are provided in Table 2. Acquisition parameters are provided in SI Experimental Methods. Spectroscopic data was collected and analyzed by B.R.W.

We confirmed that native coordination to Co(II)PPIX was maintained in these proteins using EPR spectroscopy. Co(II)CYP119, Co(II)Oxd and Co(II)DyP exhibit EPR spectra consistent with low-spin, 5-coordinate, Co(II)porphyrin complexes, as indicated by the axial symmetry and eight hyperfine features due to coupling to the S = 7/2 ^{59}Co nucleus (Figure 2.10D)). Spin Hamiltonian parameters derived from simulations of these spectra are provided in Table 2.6. The EPR spectrum of Co(II)CYP119 showed an overall spectral shape that was comparable to that observed in Co(II)P450cam. This spectrum is notably devoid of the triplet superhyperfine features that are associated with nitrogen coordination to cobalt and is therefore indicative of coordination by an I = 0 ligand such as sulfur. Notable are two distinct 5-coordinate Co(II) signals: a major signal consistent with a stronger donor as the fifth ligand and a minor signal consistent with a weaker donor as the fifth ligand. These signals are analogous to those observed by Wagner *et*

al. [49] in Co(II)P450cam, who attributed the major signal with a strong donor to axial thiolate ligation and the minor signal with a weak donor to thiol or water ligation. In contrast, CoDyP (Figure 2.10F) and CoOxd (Figure 2.10E) exhibited triplet superhyperfine structure, conclusively demonstrating coordination by a nitrogen-bearing donor ligand. The presence of nitrogen superhyperfine coupling is consistent with the expected axial coordination by the native histidine residues in CoOxd and CoDyP. These EPR data are consistent with retention of the native coordination environment for Co(II)PPIX in the contexts of CoCYP119, CoOxd and CoDyP.

In contrast to thiolate-coordinated CoCYP119 and histidine-coordinated CoOxd and CoDyP, expression of tyrosine-coordinated catalase failed to produce natively-coordinated cofactor. Using the same expression conditions as previously, we successfully produced a CoPPIX-substituted catalase 3 from *N. crassa* with reasonable yields and 54% of the protein loaded with cofactor (Table 2.3). Of that cofactor, 99% was CoPPIX. However, spectroscopic characterization of the expressed protein revealed that the cobalt ion was not bound to the native tyrosinate ligand. Visible spectra of CoCat are shown in Figure 2.11A, and peak maxima are provided in Table 2.5. No spectra of tyrosinate-ligated CoPPIX are available for comparison, but the peak maxima of Co(III)Cat are comparable to those of CoMb, suggestive of histidine coordination to CoPPIX. The EPR spectrum of Co(II)Cat is highly informative and exhibits two distinct signals (Figure 2.11B). The first signal is axially symmetric, consistent with a 5-coordinate cobalt complex, and exhibits poorly resolved triplet superhyperfine structure, indicative of a nitrogen-containing ligand bound to Co(II). The second signal, the significant feature centered around g = 2 that exhibits cobalt hyperfine coupling, is suggestive of a superoxide complex of Co(III), which may have formed upon reaction with oxygen during sample transfer. The shapes and g values for this latter signal are reminiscent of those observed for $CoMbO_2$.[4] Importantly, neither of these signals are what would be expected for tyrosinate coordination. Interrogation of the crystal structure of *N. crassa* catalase 3 (PDB 5WHQ) reveals that His-102 resides near the heme iron on the distal side opposite the native tyrosine ligand. It is therefore plausible that in the

CoPPIX-substituted protein, this histidine coordinates to cobalt in place of the native tyrosinate. Failure to bind the native phenolate ligand is unsurprising given the poor oxophilicity of cobalt. Thus, ligation of cobalt-substituted hemoproteins produced via this method is expected to follow the chemical propensity of the cobalt center.

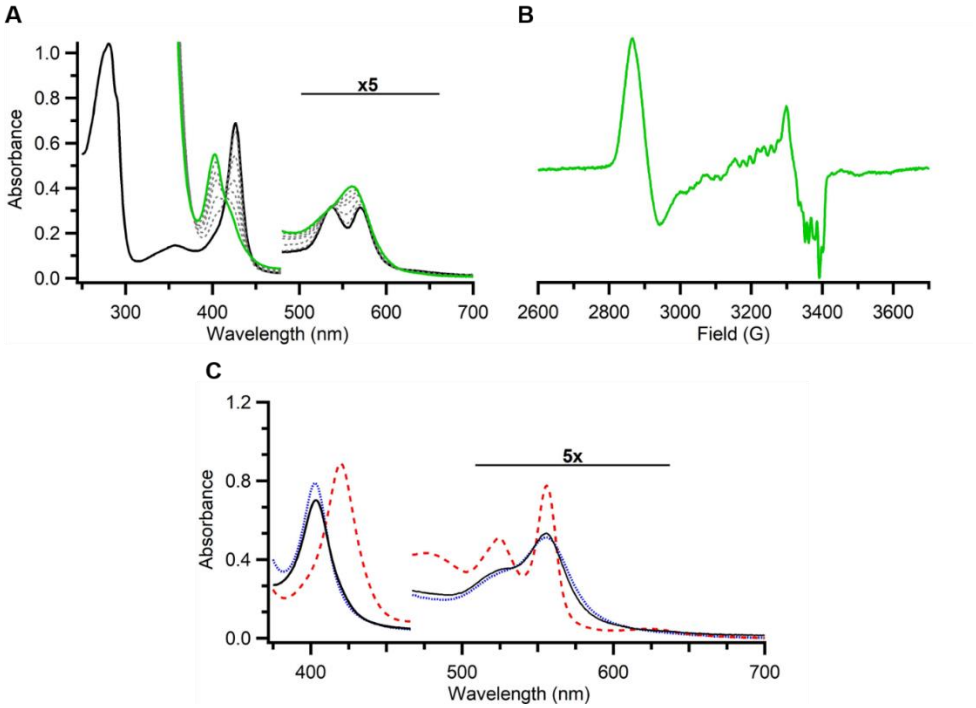


Figure 2.0.11. Spectroscopic analysis of CoPPIX-substituted catalase (Cat) shows that cobalt is not coordinated by the native tyrosinate ligand. A) Electronic absorption spectra of Co(III)Cat (black) and Co(II)Cat (green) at 35°C in 50 mM potassium phosphate buffer, pH 7.0. Intermediate spectra tracking the reduction progress after addition of sodium dithionite are shown as gray dotted lines. B) EPR spectrum of Co(II)Cat. The spectrum contains signals derived from a low spin, 5-coordinate Co(II) species and a $Co(II)O_2$ complex. Superhyperfine features observed in the low spin, 5-coordinate Co(II) spectrum between 3000 and 3200 G indicate the presence of a nitrogen-coordinating, non-native, axial ligand. C) Electronic absorption spectra of the $M(II)PPIX(py)_x$ complexes derived from CoCat (black) via the modified pyridine hemochrome assay procedure. The standard spectrum of the iron hemochrome, $Fe(II)PPIX(py)_2$ (red dashed line) and that of the cobaltochrome, Co(II)PPIX(py), (blue dotted line) are shown for reference.

2. 4. Conclusions

We serendipitously discovered that common *E. coli* BL21(DE3) is able to produce a new-to-nature CoPPIX cofactor without genetic engineering, evolutionary adaptation or auxiliary plasmids, and can efficiently incorporate this cofactor into hemoproteins. The toxicity of free cobalt, which is present in high concentrations in the growth medium, is partly offset by overexpression of the hemoprotein. We surmise that this enhanced cobalt tolerance is due to sequestration of toxic cobalt within the hemoprotein in the form of CoPPIX. To our surprise, cells rendered tolerant to cobalt by serial passaging were less efficient at incorporating CoPPIX and, in fact, selection for cobalt tolerance appeared to enhance incorporation of the native FePPIX. When plasmid-bearing *E. coli* BL21(DE3) cells are grown in augmented minimal media supplemented with 500 μM $CoCl_2$, CoPPIX is readily biosynthesized and incorporated into a variety of heterologously expressed hemoproteins with an efficiency that is comparable to their native FePPIX cofactor. Our analysis demonstrates that CoPPIX is usually incorporated into the native binding site in the expressed protein, with only trace amounts of FePPIX observed.

While the method here is only demonstrated with cobalt, future studies may explore the introduction of alternative metals. Regardless, the utility of cobalt substituted hemoproteins is broad. CoPPIX is an excellent spectroscopic analog of heme that has been used extensively in EPR and resonance Raman studies to characterize hemoproteins. Notably, the study of synthetic cobalt porphyrin complexes suggests there is a wealth of chemistry that remains to be explored and expanded upon in a biocatalytic setting.[30–32] Cobalt itself is attractive as it is abundant and inexpensive compared to many transition metals. This route to CoPPIX synthesis and bioincorporation leverages a common laboratory strain and requires no auxiliary plasmids. Protein expression does not require specialized equipment, anaerobic culture conditions, or supplementation of a pre-synthesized cofactor. Consequently, this method can be applied using common resources by any research group equipped for protein expression in *E. coli*. Further, the

straightforward nature of this approach lends itself to directed evolution, which we speculate may unlock new modes of biocatalytic CoPPIX reactivity.

2. 5. Materials and Methods

2. 5. 1. General experimental methods

All chemicals and chemical standards were purchased from commercial suppliers (Sigma-Aldrich, VWR, Goldbio, Frontier biosciences, Fluka), and used without further purification. Casamino acids were obtained from Biolabs or Research Products International. Unless otherwise noted, all media and solutions were prepared using ultrapure water (≥18 MΩ, from Thermo Scientific Barnstead Nanopure water purification system). Immobilized metal affinity chromatography resin or columns were purchased from GE Healthcare.

Equipment and instrumentation

New Brunswick I26R, 120 V/60 Hz shakers (Eppendorf) were used for cell growth. When needed, the humidity was controlled with a HumidiKit™ Auto Humidity System for Incubators purchased from IncubatorWarehouse.com. Cell disruption was accomplished via sonication with a Sonic Dismembrator 550 (Fischer Scientific) sonicator. Electroporation (for transformation and cloning) was achieved using an Eppendorf E-porator set to 2500V. Optical density measurements of liquid cultures were recorded on a Ultrospec 10 cell density meter (Amersham Biosciences). Fast protein liquid chromatography was carried out using an ÄKTA Prime Plus (GE Healthcare). An Envision® 2105 multimode plate reader (Perkin Elmer) was used to measure optical density in 96-well plates. UPLC/MS data were collected on an Acquity UHPLC with an Acquity QDA MS detector (Waters). Electronic absorption data were collected on a UV-2600 Shimadzu or a Varian Cary 4 Bio spectrophotometer, set to a spectral bandwidth of 0.5 nm. EPR spectra were collected with a Bruker ELEXSYS E500 spectrometer equipped with an Oxford ESR 900 continuous flow liquid helium cryostat and an Oxford ITC4 temperature controller. The microwave frequency was monitored using an EIP model 625A continuous-wave microwave frequency counter. ICPMS data

were collected with a Shimadzu ICP-MS 2030 equipped with an AS-10 autosampler. Curve fitting for IC_{50} assays was performed using Graphpad PRISM™ Software.

Media preparation and cell growth

Cells were grown in either Luria-Bertani (LB) broth, using the conventional recipe, or in supplemented M9 broth (M9*) using an adaptation of the recipe from Maijtan et al.[46] A stock solution of 10x M9 salts was prepared (75.2 g/L $Na_2H_2PO_4$, 30 g/L KH_2PO_4, NaCl 5 g/L, NH_4Cl 5 g/L), pH was adjusted to 8.0 with HCl or NaOH, and the solution was autoclaved. M9* media was assembled from 10x salts, 100x amino acid supplement stock (0.4 g/mL casamino acids, 1 mg/mL thiamine, sterile filtered), 100x glucose solution (20% glucose w/v, sterile filtered), 100 µM $CaCl_2$ (autoclaved as a 1 M aqueous solution), and 200 mM $MgSO_4$ (autoclaved as a 1 M aqueous solution). The media was diluted to the appropriate volume and stored at room temperature until use. All cultures were inoculated from single colonies and incubated at 37°C and 225 rpm unless otherwise stated. All growths of plasmid-containing bacteria were supplemented with 100 mg/L of ampicillin (Amp).

General cloning procedures

Hemoprotein genes appended with C-terminal 6-His tags (listed in Table SI3) and flanked by regions complementary to a pET22b vector were purchased as codon optimized gBlocks from Integrated DNA Technologies (see Table S7 for gene sequences). These DNA fragments were inserted into a pET22b vector by the Gibson Assembly method [78]. The Gibson products were subsequently transformed into electrocompetent *E. coli* BL21(DE3) cells via electroporation. After 30–45 minutes of recovery in LB broth at 37 °C and 225 rpm, cells were plated onto LB plates with 100 mg/L Amp and incubated at 37 °C overnight. Single colonies were used to inoculate 5 mL of LB broth and grown overnight. Plasmid DNA was extracted and purified using a Zyppy™ Plasmid MiniPrep Kit (Zymogen Research) and gene sequences were confirmed via Sanger sequencing (Functional Biosciences). Purified plasmids were transformed into electrocompetent *E. coli* BL21(DE3) (New England Biosciences) using the above procedure, unless otherwise noted.

2. 5. 2. Assessing effect of $CoCl_2$ supplementation on *E. coli* growth profile

We followed the growth of *E. coli* BL21(DE3) in M9* under various conditions. Single colonies of vector-less BL21(DE3) were used to inoculate 5 mL starter cultures of antibiotic-free M9* broth in triplicate and were grown overnight at 37 °C and 225 rpm. These starter cultures (500 µL of each) were used to inoculate growth cultures (50 mL each in 250 mL baffled Erlenmeyer flasks), prepared from 50 mL of M9* media supplemented with 250 µM δ-aminolevulinic acid (δALA) (for δALA only, Fe + δALA, and Co + δALA growths) and 500 µM metal salt ($FeCl_3$ for Fe + δALA growth, $CoCl_2$ for Co + δALA growth). These cultures were grown at 37 ºC and 225 rpm, measuring OD_{600} every 1–3 hours. Growth was stopped and cells were harvested once the OD_{600} did not significantly increase from one time point to the next. Of note, severely diminished growth was observed under all growth conditions when non-baffled flasks were used. Even after 3 days of incubation in non-baffled flasks (37 °C, 225 rpm), no growth was detected in the Co + δALA cultures (data not shown).

2. 5. 3. Passaging *E. coli* through $CoCl_2$ supplemented media

We generated three separate lineages of passaged *E. coli*, which we nicknamed **A**lvin, **S**imon, and **T**heodore and which are denoted A, S and T subsequently. A glycerol stock of BL21(DE3) *E. coli* (New England Biosciences) was streaked onto an antibiotic-free LB agar plate and three single colonies were used to inoculate 5 mL starter cultures of M9*. The cultures were shaken at 225 rpm at 37 ºC for 16 h. Each starter culture (50 µL) was used to inoculate 5 mL of M9*, supplemented with 150 µM $CoCl_2$. To passage, these cultures were grown for 8–15 hr, after which 50 µL of each was used to inoculate 5 mL of fresh M9*, supplemented with 150 µM $CoCl_2$. This passaging was repeated every 8–15 h for a total of 15 passages over 2 weeks, with three changes in $CoCl_2$ concentration. After the 8^{th} passage, the concentration of $CoCl_2$ was increased

to 450 μM, after the 12th passage the concentration was increased to 750 μM, and after the 14th passage the concentration was increased to 1 mM $CoCl_2$. OD_{600} values, taken at the end of each growth period, are shown in Figure S1. The 15x passaged cultures (A15, S15, and T15 populations) were used to make 25% glycerol stocks and were stored at -80 ºC until further use.

2. 5. 4. Quantitation of *E. coli* cobalt tolerance with IC_{50} assays

Effect of passaging on E. coli cobalt tolerance (IC_{50})

We tested the tolerance of parent and passaged *E. coli* BL21 (DE3) cells to $CoCl_2$ by measuring the IC_{50} for $CoCl_2$ in various cell lines. Glycerol stocks of un-passaged BL21(DE3) *E. coli* (New England Biosciences) and passaged populations (A15, S15, and T15) were streaked onto antibiotic-free LB agar plates to obtain single clones. Three single colonies from each plate were used to inoculate 5 mL of antibiotic-free M9*, and the cultures were grown at 37 °C for 16 h. These cultures (5 μL of each) were used to inoculate 195 μL of M9* supplemented with variable amounts of $CoCl_2$ (0–3 mM) in triplicate in 96-well plates, with plate layout as illustrated in Figure S12 (Eppendorf). Following inoculation, the 96-well plates were incubated at 37 °C with a shake speed of 180 rpm. To prevent excessive evaporation, plates were stacked with a water-filled plate above and below, and the incubator's humidity was maintained between 60-80% relative humidity with a humidifier. After 9 h, the optical density at 600 nm of each well was measured with a plate reader. To eliminate any contribution from the absorbance of or scattering from the $CoCl_2$ supplemented M9*, the OD_{600} value for each growth well was corrected by subtracting the average optical density of 6 wells with the same $CoCl_2$ concentration in an identical plate that was inoculated with 5 μL sterile M9* and incubated with the rest of the plates. The OD_{600} versus $CoCl_2$ concentration data were fit to the Hill equation (below):

$$\mathbf{OD_{600}} = \mathbf{Bottom} + \frac{\mathbf{Top} - \mathbf{Bottom}}{\mathbf{1} + \left(\frac{[\boldsymbol{CoCl_2}]}{\boldsymbol{IC_{50}}}\right)^{\boldsymbol{Hill\ coefficient}}}$$

Curve fits are shown in Figure S5 and fit parameters are listed in Table S1. Best fit curves for single clones of the passaged strains were superimposable, so the individual data points for passaged strains were aggregated and analyzed together. In the case of the passaged strains, the steepness of curves (plausibly limiting the applicability of the Hill equation) and a lack of datapoints with intermediate OD_{600} (due to limits in experimental design), made fit values for IC_{50} and Hill coefficient difficult to bound with high confidence. Thus, Figure S5 shows the best-fit curves that result from allowing the Hill coefficient to vary (dark blue), and from constraining the Hill coefficient to the lower 95% confidence bound (-5.12) and an arbitrarily large value (-50) (light blue dashed lines). Regardless of the treatment of Hill coefficient, the best-fit IC_{50} value for the 15x passaged strain is 10-fold higher than that of the parent BL21(DE3) (250 μM $CoCl_2$ (Hill coefficient = -50) versus 20 μM $CoCl_2$ for the un-passaged strain).

Effects of hemoprotein and non-heme-protein overexpression on E. coli cobalt tolerance (IC_{50})

We tested whether expression of a heme-containing protein (Mb*) conferred greater tolerance to $CoCl_2$ than expression of a non-heme, cofactor-containing protein. Single-colony derived overnight cultures (5 μL of each) of BL21(DE3) *E. coli* bearing a pET22b plasmid encoding either Mb* or *Pfu*TrpB (the PLP-dependent β subunit of tryptophan synthase from *P. furiosus*) were used to inoculate 195 μL of M9* with 100 mg/L Amp and variable amounts of $CoCl_2$ (0–3 mM) and IPTG (0–1000 mM) in a 96-well plate (Figure 2.12). Plates were incubated at 37 °C and 180 rpm for 9 h. Humidity was controlled to eliminate evaporation during incubation, as described above, after which OD_{600} values were measured. The measured OD_{600} values were corrected for absorbance of or scattering from $CoCl_2$-containing media as described above. Each starter culture inoculated three rows in the 96-well plate, yielding experimental triplicate measurement of OD_{600}. Data was processed as described above; fits for individual data sets are in Figure 2.13.

2. 5. 5. Comparing the Mb* expression efficiency between *E. coli* BL21(DE3) and passaged strains in $CoCl_2$ supplemented media

We tested the passaged, cobalt-tolerant strains for their ability to produce CoMb* and compared that ability to that of unpassaged BL21(DE3). Passaged population glycerol stocks (A15 and S15) and control BL21(DE3) cells were streaked onto antibiotic-free LB agar plates and grown at 37 °C. Single colonies from each plate were grown in LB broth, and the cells were rendered electrocompetent by washing multiple times with 10% glycerol in deionized water. Purified pET22b plasmid bearing a C-terminal 6-His-tagged Mb* gene (See Table S7) was transformed into these cells via electroporation according to the general cloning procedure given above. Starter cultures (a single colony into 5 mL M9*) were used to inoculate 50 mL of M9* broth in 250 mL Erlenmeyer flasks, which were incubated in a shaker at 37 ºC and 225 rpm. When culture OD_{600} values reached approximately 0.2, CoPPIX production was induced by addition of δALA and $CoCl_2$, to final concentrations of 200 μM and 500 μM, respectively. Incubation continued until the culture OD_{600} values reached 0.6–0.8, at which time cultures were removed and placed on ice for 40 minutes. Mb* production was induced by addition of IPTG (to final concentration of 1 mM), and the cultures were incubated at 23 ºC and 140 rpm for 16 h. Cells were harvested by centrifugation (4,300x*g*, 1 h, 4 °C) and stored at -20 °C for 24 h before lysis and purification, as described below. Metalloporphyrin content of the purified Mb* stocks were assessed using the modified pyridine hemochrome assay described below.

2. 5. 6. Hemoprotein overexpression and purification

Expression of cobalt-substituted hemoproteins

CoPPIX-substituted hemoproteins were overexpressed in the following manner. M9* overnight starter cultures of *E. coli* bearing hemoprotein expression vectors described earlier were used to inoculate 1 L of M9* media containing 100 mg/L Amp. These cultures were incubated at

37 °C and 225 rpm. After the OD_{600} reached 0.2–0.3, both $CoCl_2$ and δALA were added from sterile filtered 1.0 M stocks to final concentrations of 500 mM and 250 mM, respectively. The expression flasks were returned to the incubator at 37 °C and 180 rpm. Once the OD_{600} had reached 0.6–0.7, the expression cultures were chilled on ice for 20–30 minutes. Expression was induced with 250 mM IPTG and the cultures were incubated overnight at 23 °C and 140 rpm. Cells were harvested by centrifugation at 4300x*g* and 4 ºC. Cells were stored at -20 °C until lysis (at least 12 hours and no more than 2 weeks).

General method for protein purification

Cell pellets of derived from expression procedure described above were lysed and proteins were purified according to the following procedure. Frozen cell pellets were resuspended in lysis buffer (50 mM potassium phosphate buffer, 250 mM NaCl, 10 mM imidazole, 1 mg/mL hen egg white lysozyme, 0.2 mg/mL DNase I, 2 mM $MgCl_2$, pH 8.0). A volume of 4 mL of lysis buffer per gram of wet cell mass was used. After incubation for 30 minutes at 37 °C and 140 rpm, the cells were disrupted via sonication for 20 min on ice (0.8 second pulses followed by 0.2 second pause, at a power setting of 5). The resulting lysate was clarified of cellular debris via centrifugation at 75,000 x *g* for 30 min. Supernatant was applied to a Ni-NTA gravity column or a 5 mL HisTrap column equilibrated with binding buffer (50 mM potassium phosphate, 250 mM sodium chloride, 10 mM imidazole, pH 8.0). The column was washed with approximately 3 column volumes 15:85, 25:75 and 30:70 (v/v) mixtures of elution buffer (50 mM potassium phosphate, 250 mM sodium chloride, 200 mM imidazole, pH 8.0) and binding buffer. Proteins were eluted with 100% elution buffer. Eluent was pooled and dialyzed into storage buffer (50 mM potassium phosphate, pH 8.0) overnight at 4°C. Dialyzed protein solutions were concentrated to <500 µL using Amicon® Ultra 15 mL Centrifugal Filters (Millipore) at a molecular weight cutoff of 15 kDa

for Mb* and 30 kDa for Oxd, DyP and CYP119. Final protein solutions were flash frozen in liquid nitrogen and stored at -80°C.

Expression of FeDyP in rich media

DyP was expressed under iron-containing, rich media conditions for comparison of the heme (FePPIX) and H_2PPIX loading to that of the CoPPIX-substituted variant. Overnight starter cultures (5 mL) of *E. coli* bearing a pET22b with the DyP gene in LB broth were used to inoculate 1 L of Terrific Broth (TB) + 100 mg/L Amp. These cultures were incubated at 37 °C and 225 rpm. Once the OD_{600} reached 0.2–0.3, cultures were supplemented with 250 mM δALA, and the expression flasks were returned to the incubator at 37 °C and 180 rpm. When the OD_{600} had reached 0.6–0.7, the expression cultures were chilled on ice for 20–30 minutes. Expression was induced with 250 μM IPTG and the cultures were incubated overnight at 23 ºC and140 rpm. Cells were harvested by centrifugation at 4300x*g* at 4 ºC. Cell pellets were stored at -20 ºC prior to protein purification, as described above.

2. 5. 7. Spectroscopic characterization of purified CoPPIX-substituted hemoproteins

Electronic absorption spectroscopy of CoPPIX-substituted hemoproteins

We collected electronic absorption spectra to characterize the purified CoPPIX-substituted hemoproteins to compare their spectra with previous literature examples. Purified protein was mixed with argon-sparged 50 mM potassium phosphate buffer, pH 7.0 in septum-sealed cuvettes to ensure anaerobic conditions. Protein solutions were placed in the spectrometer at 35°C using a Cary temperature control module for 5 minutes prior to collecting as-isolated spectra. Reduction was initiated via syringe addition of an anaerobic stock solution of sodium dithionite in buffer, to achieve a final concentration of 3–4 mM. Spectra were recorded every five minutes until absorbance changes were no longer observed.

EPR spectroscopy of Co(II)PPIX-substituted hemoproteins

EPR spectroscopy was used to characterize the coordination environment of cobalt in each Co(II)PPIX-substituted hemoprotein. Protein samples were prepared in Ar-sparged buffer (50 mM potassium phosphate, pH 7.0) in septum-sealed cuvettes. Sodium dithionite was added as an anaerobic stock solution in the same buffer via syringe, to a final concentration of 4 mM. Samples were incubated for 2 hours at 35°C; complete reduction was confirmed via electronic absorption spectroscopy. Samples were brought into an anaerobic chamber and concentrated to 200 μL using 0.5 mL centrifugal filters (Amicon, Millipore-Sigma). Protein solutions were transferred to quartz EPR tubes via syringe, removed from the anaerobic chamber and immediately frozen in liquid nitrogen.

The CoPPIX concentration in the EPR samples varied from 100 to 270 μM. Spectra were recorded at 0.05024 mW microwave power, 40.96 s time constant, 40.96 s conversion time, 100 kHz modulation frequency, and 5 G modulation amplitude. These parameters were chosen to optimize the signal on the basis of a power saturation experiment, as shown in Figure 2.15.

2. 5. 8. Quantification of Porphyrin Content

UPLC-MS analysis of porphyrin content for cell pellets and purified protein

Cell pellet sample preparation:

We analyzed the porphyrin content of cell pellets grown under a variety of conditions by UPLC-MS. Stationary phase cell cultures were obtained from the growth curve analysis outlined on page S27. Cells were harvested by centrifugation (4,300*xg*, 5 min, 4 ºC), and cell pellets were immediately lysed via addition of 2 mL of an 8:2 pyridine:1.2 M HCl mixture and vortexed for 3 minutes. Saturated $MgSO_4$ (2 mL) was added, and lysates were centrifuged (4,300 x *g*, 5 min, 4 ºC) to yield defined aqueous and organic layers. The organic supernatant was removed and clarified by further centrifugation (14,000 rpm, 3 min). The clarified, organic-soluble portion of the

cell lysate was then injected onto UPLC-MS. Reported porphyrin concentrations are the average of biological triplicate measurements, and the standard error for these measurements was propagated from the standard error of the calibration curve (Figure S13).

Purified protein sample preparation:

The relative porphyrin content of purified protein samples was analyzed by UPLC-MS. Purified protein samples, obtained from the protein purification procedure outlined on page S31, were thawed on the benchtop. An aliquot of 2–30 µL purified protein was added to 200 µL 8:2 pyridine:1.2 M HCl mixture and the resulting mixture was vortexed for 1 min. Saturated $MgSO_4$ (200 µL) was added and the solution was vortexed vigorously once more. The emulsion was centrifuged (4,300 x *g*, 5 min, 4 ºC) to yield defined aqueous and organic layers. The organic supernatant was removed and clarified by further centrifugation (14,000 rpm, 3 min). The clarified, organic-soluble portion of the cell lysate was injected onto UPLC-MS. Reported porphyrin concentrations are the average of triplicate extractions from a single protein sample, and the standard error for these measurements was propagated from the standard error of the calibration curve (Figure S13).

UPLC analysis:

The UPLC chromatographic method consisted of an isocratic mobile phase (25% water in acetonitrile with 0.1% formic acid), flowing at a rate of 1 mL/min and 40 ºC, over an ACQUITY UPLC-CSH Phenyl-Hexyl 1.7µm column (Waters). A 3.0 µL injection volume was used. An LC trace for a purified protein sample is given in Figure SI7. Clear porphyrin signals were observed via absorbance at the porphyrin Soret maximum (398 nm for FePPIX, 400 nm for H_2PPIX, and 423 for CoPPIX), and positive single ion reads (SIR) (616.3 m/z for FePPIX, 563.3 m/z for H_2PPIX, and 619.3 m/z for CoPPIX). Porphyrin content was quantified by comparison to standard curves (see Figure S13). Integrated UV absorbance was used to quantify H_2PPIX and FePPIX. For

CoPPIX, interfering absorption at 423 nm and the CoPPIX retention time rendered absorption analysis less useful for cell lysates. Therefore, CoPPIX concentration was assessed using integrated SIR for cell pellet analyses. The calculated concentration of porphyrin in the analyte solution and the volume of organic supernatant obtained by centrifugation were used to calculate the concentration of porphyrin in the original sample.

ICP-MS analysis for total metal content of purified proteins

Metal content of each protein was determined by inductively coupled plasma-mass spectrometry. Metal samples were prepared by addition of trace analysis grade 70% nitric acid (350 μL) to protein stocks (<30 μL). Samples were incubated at 80°C for 4 hours to ensure complete digestion and subsequently diluted with ultrapure water to give a final concentration of 5% nitric acid. Cobalt and iron standard solutions were prepared in 5% nitric acid solutions from 1000 ppm atomic absorption standards. ICP-MS conditions and parameters are shown in Table S5. Metal content was quantified by comparison to a standard curve (Figure S11). To calculate CoPPIX-loading, cobalt concentrations were compared to protein concentrations measured using the Pierce™ bicinchoninic acid assay (BCA) assay kit (ThermoFisher Scientific) according to the manufacturer's instructions.

Modified pyridine hemochrome assay

A modified version of the pyridine hemochrome assay[59] was used to qualitatively assess metalloporphyrin content in purified proteins. Protein solution (50 μL) was added to 350 μL of pyridine hemochrome reagent (40 mL pyridine, 36 mL H_2O, and 4 mL 1 M NaOH) in a quartz cuvette. The sample was capped, mixed thoroughly, and heated to 70 ºC to fully denature the protein and form the pyridine metalloporphyrin complex. A few crystals of sodium dithionite were added and the sample was cooled to 30 ºC. Spectra were recorded at 30 ºC. Comparison of these

spectra to reduced pyridine adducts of commercially available Co(III)PPIX(Cl) and Fe(III)PPIX(Cl) yields qualitative assessment of metalloporphyrin content.

2. 6. Supplemental Figures and Tables

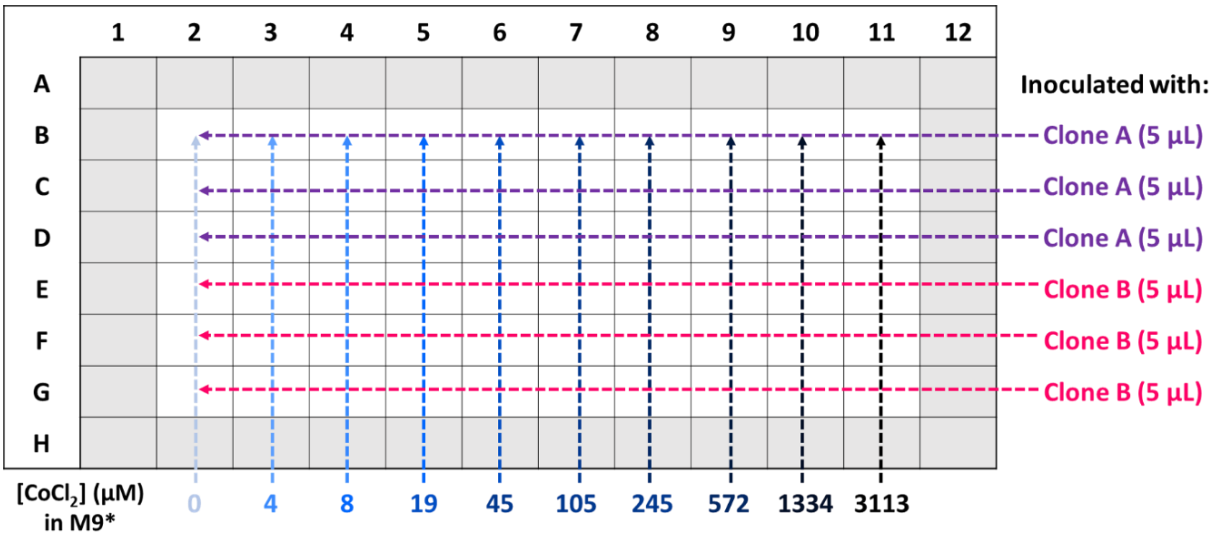


Figure 2.0.12. Representative layout of a 96-well plate for the IC_{50} assay to measure cobalt tolerance. Each plate was sufficient for triplicate measurements of two clones under one set of conditions. Outer wells were filled with 200 µL of water to prevent excessive evaporation in the wells on the edges of the plate.

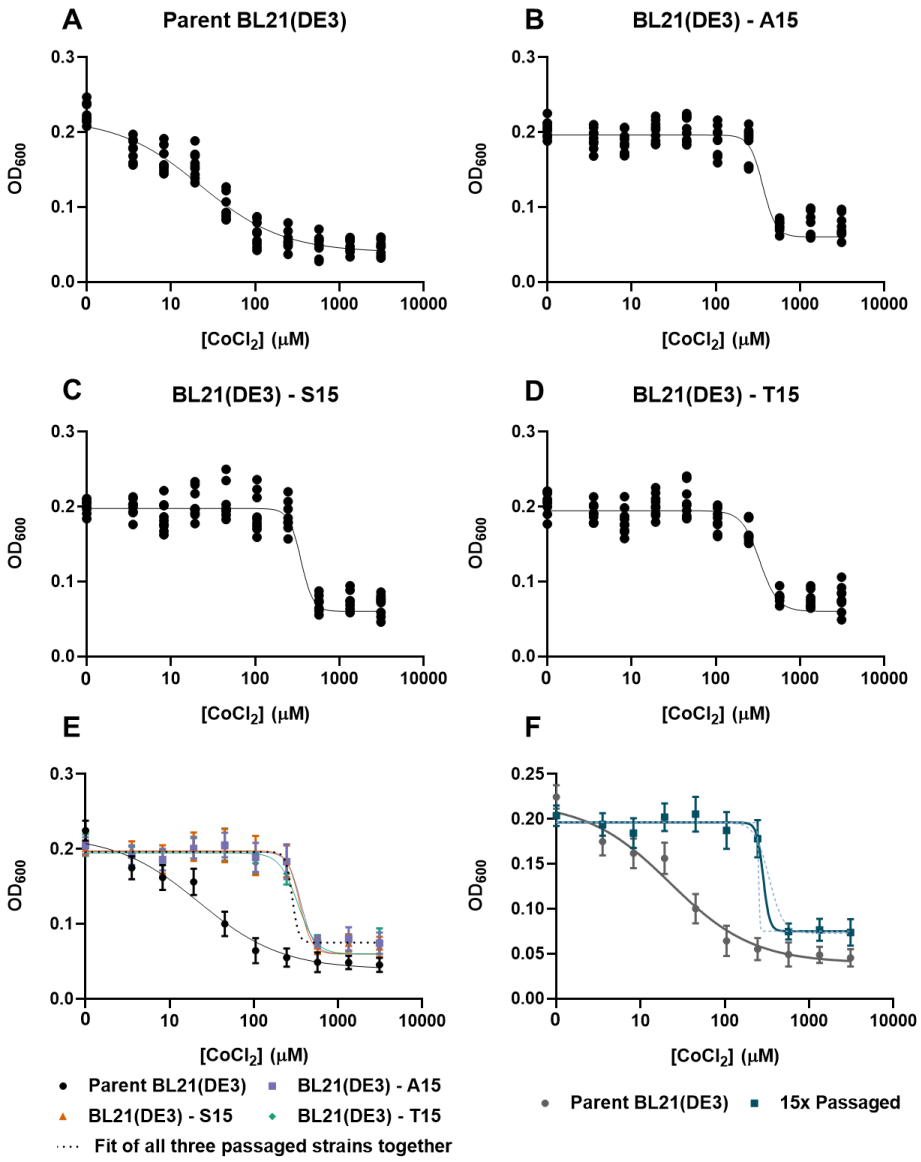


Figure 2.0.13. Optical densities of E. coli BL21(DE3) cultures, after 9 hours growth at variable concentrations of $CoCl_2$. In all graphs, the line represents the best fit to the data using the Hill equation (fit parameters given in Table S1). (A) Growth of parent BL21(DE3) E. coli. Individual data points, consisting of biological triplicate (three single colonies) and experimental triplicate (three wells inoculated from each starter culture), are shown. (B-D) Growth of passaged BL21(DE3) E. coli. Individual data points, consisting of biological triplicate (three single colonies from each passaged cell-derived population, A15, S15, and T15, were used to inoculate starter cultures) and experimental triplicate (three wells inoculated from each starter culture), are shown. (E) Superposition of the growth curves from all three passaged strains (A15, S15, and T15), with the best fit line to data from all three passaged strains (dotted line). (F) Direct comparison of parent BL21(DE3) and data from all three passaged growths in variable concentration $CoCl_2$-containing media. Data points are the average of either nine (for the parent) or 27 (for the 15x passaged strain) individual data points, and error bars represent the standard deviations from the mean. Blue dotted lines represent best fit curves when the Hill coefficient is set to specific values. More details of this analysis can be found in the IC_{50} assay description section of the experimental methods.

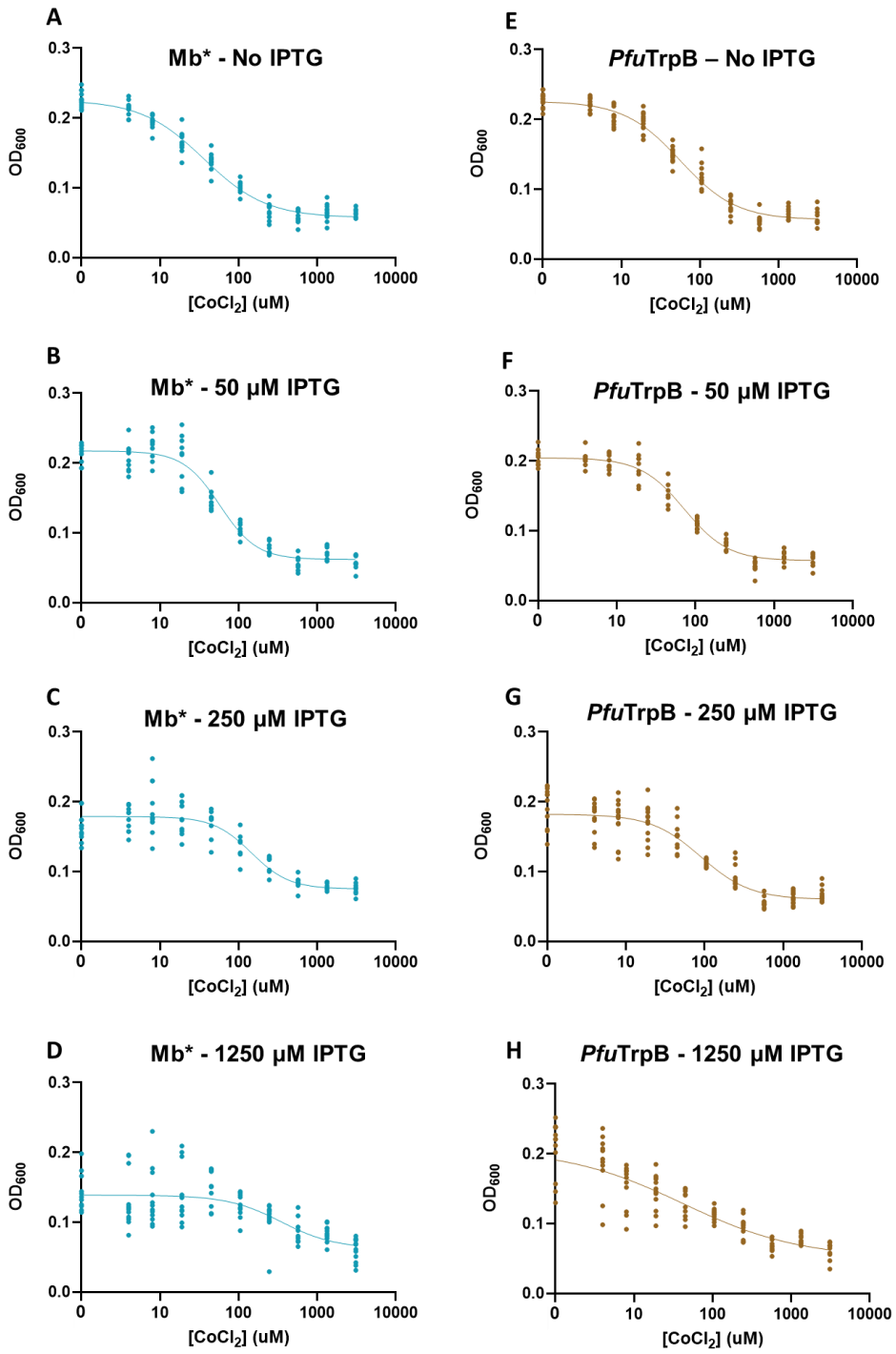


Figure 2.0.14. Optical densities for growth of Mb* expressing Bl21(DE3) *E. coli* (A-D) and *Pf*TrpB expressing Bl21(DE3) *E. coli (E-H)* in variable concentrations of $CoCl_2$. Individual data points, consisting of biological quadruplicate (four monoclonal single colonies were used to inoculate starter cultures) and experimental triplicate (three wells inoculated from each starter culture), are shown. In all graphs, the line represents the best fit to the data using the Hill equation, with fit parameters given in Table S3. Direct comparisons between these curves are shown in 2.15.

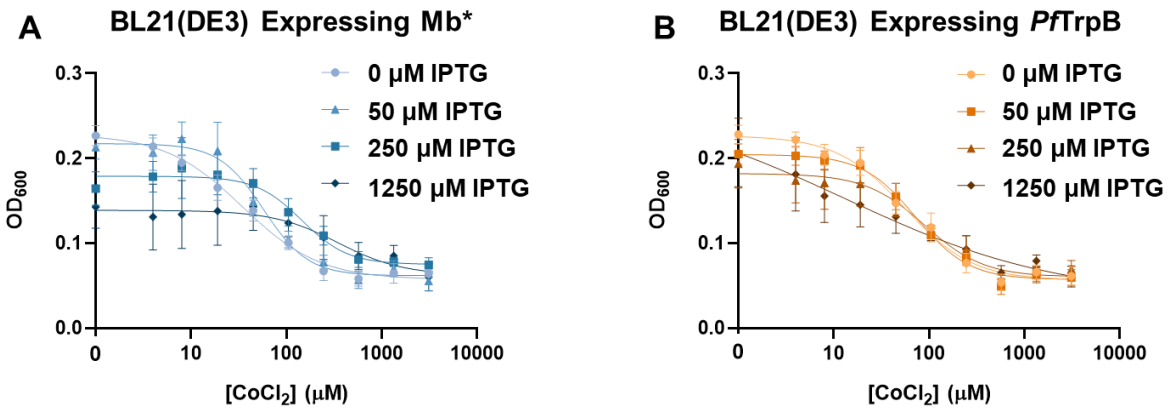


Figure 2.0.15. Effect of varied IPTG concentration on optical densities of BL21(DE3) cells expressing Mb* (A) or PfuTrpB (B). Individual data points represent the average of twelve total replicates, and error bars represent the standard deviation from the mean. Hill equation fit curves to each data set are the same as in Figure S6, and fit values for these curves can be found in Table S3. (C) IC_{50} values (concentration of $CoCl_2$ at which growth is inhibited by 50%) calculated from the line of best fit for each strain at a given IPTG concentration. Error bars represent the standard error of the fit for the IC_{50} parameter. These IC_{50} values are listed in Table S3. Figure C is also included in the main text.

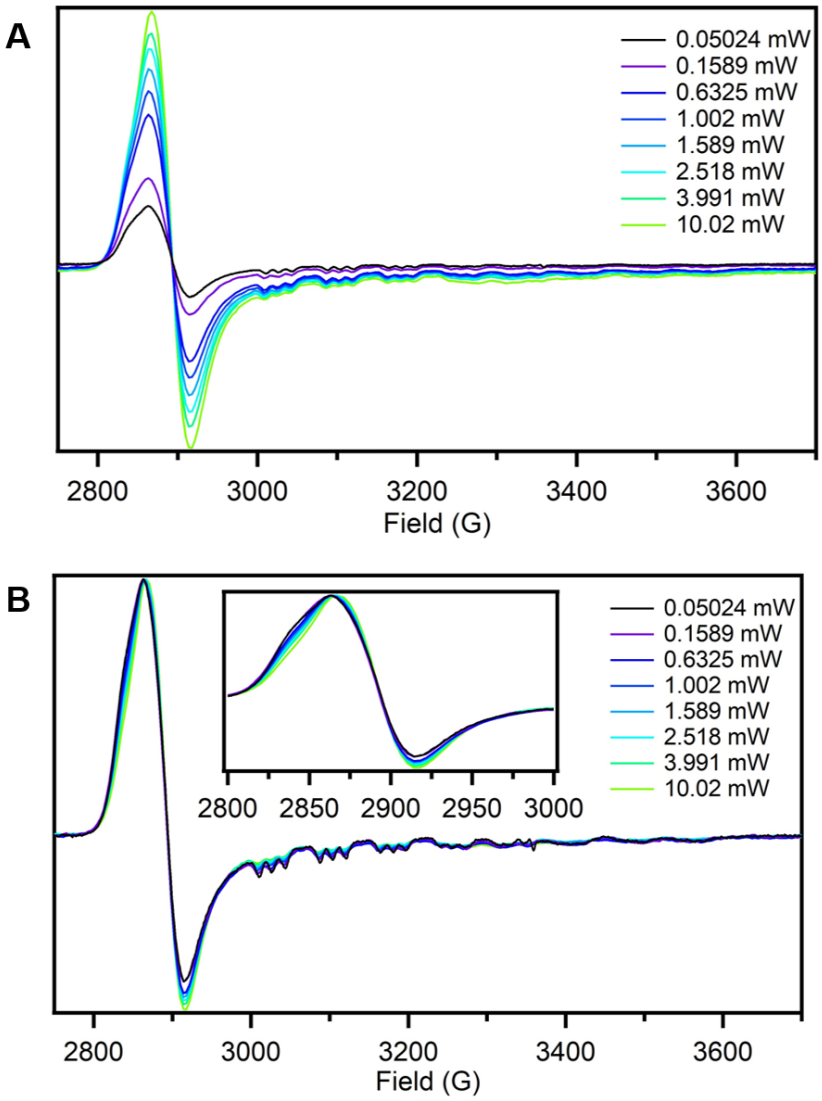


Figure 2.0.16. Power saturation experiment for the Co(II)Mb* EPR signal at 20 K to determine optimal EPR acquisition parameters. (A) Raw EPR spectra collected at the microwave powers indicated on the graph. (B) EPR spectra identical to those in (A) normalized in intensity at the top of the axial feature highlighting the loss in relative intensity of the hyperfine and superhyperfine features. Inset in (B) highlights the broadening of the axial feature at low power. EPR spectra of Co(II)PPIX-substituted proteins were collected at microwave power of 0.05024 mW (black), as this value gave the best spectral resolution.

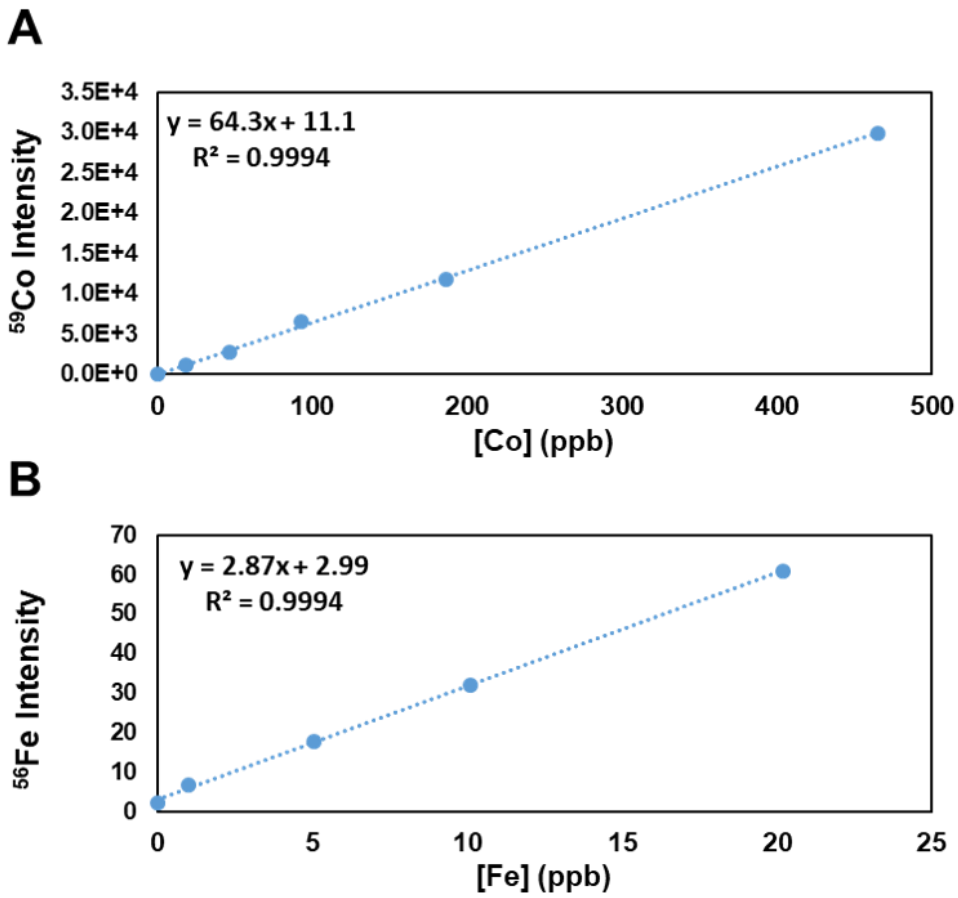


Figure 2.0.17. Calibration curves for metal content analysis by ICP-MS. (A) Cobalt-59 intensity of Co^{2+} calibration solutions (0.0 ppb, 18.6, 46.5, 93, 186, 465) measured by ICP-MS. (B) Iron-56 intensity of Fe^{3+} calibration solutions (0 ppb, 1.01, 5.05, 10.10, 20.20) measured by ICP-MS. Calibration solutions were prepared from 1000 ppb atomic absorption standards in trace analysis grade nitric acid (5%). Experimental details on data collection are provided in the experimental methods section, quantitation of porphyrin content

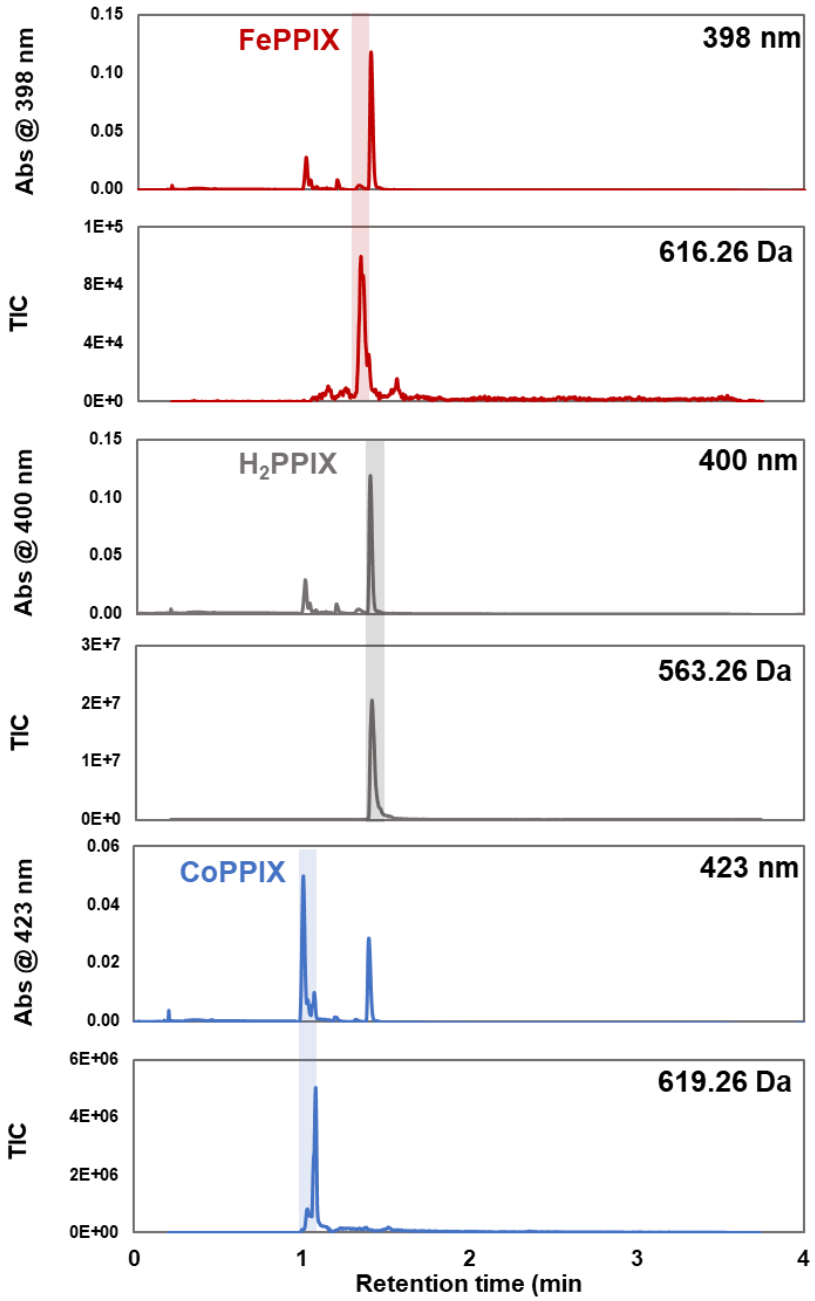


Figure 2.0.18. Example of a UPLC-MS trace for a purified protein sample. The trace shown is for a single replicate of a purified sample of CoPPIX-substituted dye-decolorizing peroxidase (CoDyP), which has a particularly high H_2PPIX loading. Total ion counts (TIC) at the mass of the m+1 ions align well with the Soret absorbance maxima for each porphyrin. Note that H_2PPIX and FePPIX have similar Soret absorbance maxima. Mass detection confirms peak identity of the FePPIX at a slightly shorter retention time than H_2PPIX

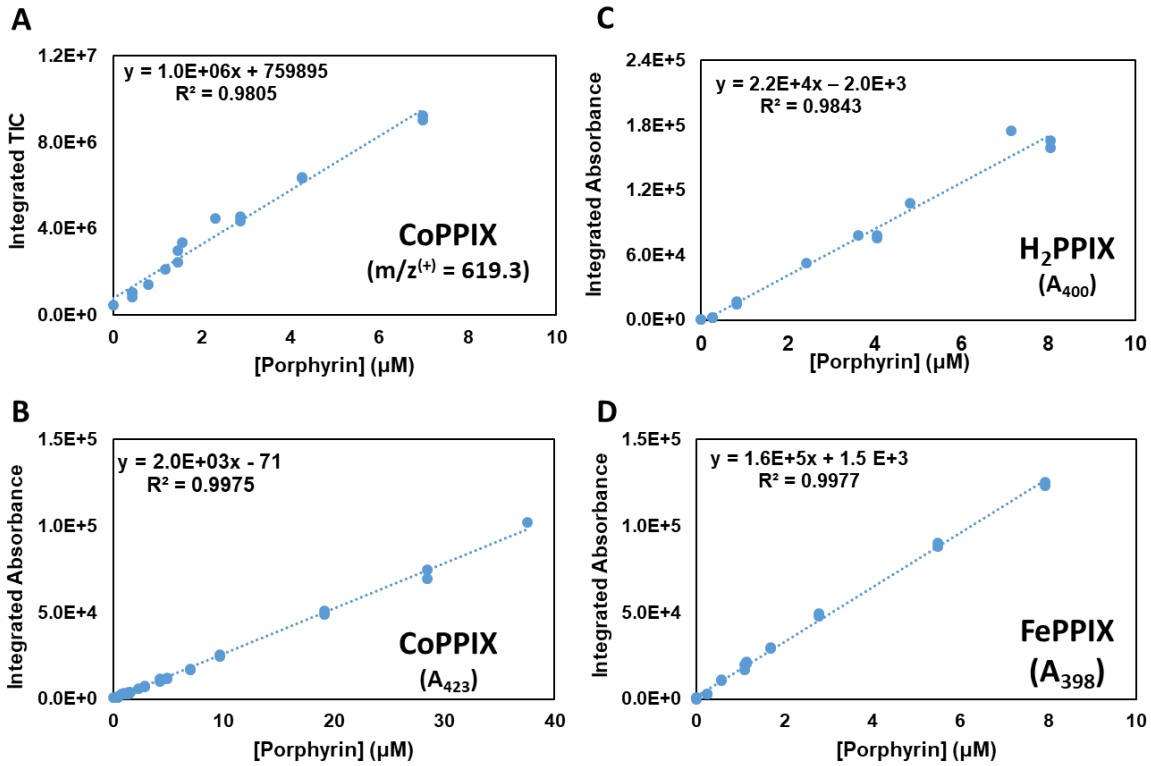


Figure 2.0.19. Calibration curves for UPLC-MS quantitation of porphyrin content. Standard solutions were made from commercially available CoPPIXCl, Na_2PPIX, and FePPIXCl. A) Calibration curve for the quantitation of CoPPIX by integrated total ion count (TIC). The m+1 mass (619.3 m/z) was used to determine the concentration of CoPPIX in cell pellet samples. B) Calibration curve for the quantitation of CoPPIX by UV absorbance. C-D) Calibration curve for the quantitation of FePPIX and H_2PPIX by UV absorbance, respectively. Further details are outlined in the experimental methods, porphyrin quantitation section.

Table 2.6. Calculated porphyrin concentrations in cell pellets grown under various conditions.

	Concentration of porphyrin per L cell culture (μM)			
Porphyrin	**No additives**	**δALA added**	**Fe + δALA added**	**Co + δALA added**
FePPIX	6 ± 0.6	14 ± 4	150 ± 6	14 ± 4
H_2PPIX	< 0.5	186 ± 1	4 ± 1	< 2
CoPPIX	< 6	< 6	< 6	71 ± 2
	Amount of porphyrin per OD_{600} (μmol)			
	No additives	**δALA added**	**Fe + δALA added**	**Co + δALA added**
FePPIX	0.09 ± 0.06	0.25 ± 0.07	1.29 ± 0.05	0.39 ± 0.11
H_2PPIX	< 0.01	3.33 ± 0.03	0.21 ± 0.01	< 0.01
CoPPIX	< 0.09	< 0.09	< 0.1	1.98 ± 0.08

Table 2.7. ICP-MS parameters and accessories used for data collection.

Parameters and Accessories	**Value**
Radio frequency power generator	1.2 kW
Gas type	Ar
Plasma gas flow rate	8.0 L min^{-1}
Auxiliary gas flow rate	1.1 L min^{-1}
Nebulization gas flow rate	0.6 L min^{-1}
Torch	Mini-torch (quartz)
Nebulizer	coaxial
Spray chamber temperature	5 °C
Drain	Gravity fed
Sampling Depth	5.0 mm
Collision cell gas (He	6.0 mL min^{-1}
Cell Voltage	-21 V
Energy Filter	7.0 V
Number of replicates	3

Table 2.8. Codon optimized gene sequences

***Physeter catadon* myoglobin H64V/V68A** (Mb*)
ATGGTCCTGTCAGAGGGGGAGTGGCAATTAGTGCTGCACGTCTGGGCTAAGGTAGAAGCTGACG TGGCCGGGCACGGGCAAGATATTTTAATCCGCTTATTTAAGTCCCACCCCGAGACTTTAGAAAAAT TCGACCGTTTTAAACATCTTAAAACGGAAGCCGAGATGAAAGCCTCAGAGGACTTAAAGAAGGTA GGCGTAACTGCTTTGACCGCACTGGGAGCCATCCTTAAGAAAAAAGGACACCATGAAGCCGAATT GAAGCCTCTTGCACAGAGTCACGCAACCAAACACAAGATCCCCATCAAATACTTGGAGTTTATCAG CGAAGCGATCATTCATGTGCTTCATTCACGCCACCCTGGGGATTTCGGGGCTGACGCCCAAGGC GCCATGAATAAAGCCCTGGAGTTATTCCGTAAGGACATCGCCGCTAAATATAAGGAGCTTGGGTA TCAGGGGCTCGAGCACCATCACCATCACCAT**TGA**
Sulfolobus acidocaldarius cytochrome P450 119 *(CYP119)*
ATGTACGATTGGTTTTCGGAGATGCGTAAAAAAGATCCTGTTTACTACGATGGTAATATTTGGCAA GTATTTTCATATCGCTACACCAAGGAGGTATTAAACAATTTTTCAAAATTTTCTTCCGATTTAACAGG GTACCACGAGCGCCTTGAGGACTTGCGCAATGGAAAGATCCGCTTTGACATCCCAACCCGCTATA CCATGCTTACGTCGGATCCGCCGCTGCATGACGAATTGCGTAGTATGAGTGCCGATATCTTTTCTC CGCAAAAGTTGCAAACCCTGGAAACTTTTATTCGCGAAACGACCCGTAGTTTATTGGACTCGATTG ACCCTCGCGAGGACGATATTGTTAAGAAGCTGGCCGTGCCCCTTCCAATTATCGTGATTTCCAAAA TCCTGGGTCTGCCAATCGAGGACAAAGAGAAGTTCAAGGAATGGAGCGACCTTGTAGCGTTTCGT CTGGGCAAACCTGGCGAGATCTTCGAATTAGGAAAGAAGTATCTGGAATTGATCGGATATGTGAA AGACCATCTTAATTCAGGTACGGAAGTGGTCAGCCGCGTAGTTAATTCGAATCTTTCTGACATTGA AAAATTAGGATATATCATTTTATTATTGATTGCCGGGAACGAGACGACTACTAATCTGATTAGCAAT TCAGTTATTGACTTCACGCGCTTCAACTTGTGGCAACGCATTCGCGAAGAAAACCTTTACCTGAAA GCTATTGAAGAGGCTCTTCGCTATTCACCCCCGGTTATGCGCACAGTTCGTAAAACCAAGGAGCG TGTAAAGTTGGGCGACCAGACGATTGAGGAAGGAGAATACGTGCGCGTCTGGATTGCATCAGCG AACCGCGACGAAGAGGTCTTTCACGACGGAGAAAAGTTCATTCCAGATCGTAACCCGAATCCTCA TCTGAGTTTCGGGAGTGGGATCCACCTGTGCCTGGGGGCACCTTTGGCACGTTTGGAAGCCCGT ATTGCCATCGAGGAGTTCTCGAAGCGTTTTCGTCACATCGAAATTCTTGACACAGAAAAGGTGCCA AACGAAGTGTTGAACGGGTATAAGCGTTTGGTAGTTCGTCTGAAGAGCAATGAGCTCGAGCACCA TCACCATCACCAT**TGA**
***Enterobacter lignolyticus* dye decolorizing peroxidase** (DyP)
ATGTCACAAGTGCAGTCTGGCATTCTTCCAGAGCACTGTCGTGCAGCCATCTGGATCGAGGCAAA CGTGAAAGGGGATGTCAATGCACTGCGCGAGTGCTCAAAAGTCTTTGCAGATAAATTAGCTGGTT TCGAGGCACAGTTTCCAGACGCACATCTTGGAGCGGTCGTTGCCTTTGGACATGACACATGGCGT GCTTTGTCGGGGGGGGTAGGGGCCGAGGAATTGAAGGATTTCACGCCTTATGGCAAGGGTTTGG CTCCAGCCACTCAGTACGATGTCCTTATCCATATTTTATCTCTGCGTCACGACGTAAATTTTTCCGT CGCCCAAGCGGCCATGGCCGCTTTTGGGGATGCTGTGGAGGTGAAAGAAGAGATTCACGGCTTT CGCTGGGTCGAAGAACGTGACTTATCTGGGTTCGTAGATGGGACAGAAAATCCAGCGGGGGAAG AAACACGCCGTGAAGTCGCCGTTATTAAAGATGGGGTTGACGCCGGAGGCAGTTACGTTTTTGTG CAACGCTGGGAGCATAACTTAAAACAACTTAACCGTATGTCAGTTCACGATCAGGAAATGATGATC GGTCGTACAAAAGTTGCAAACGAAGAAATTGATGGCGATGAGCGCCCAGAGACAAGCCACTTAAC GCGTGTCGACCTTAAAGAGAACGGAAAAGGCTTAAAAATTGTTCGTCAGAGCTTGCCGTACGGTA CTGCCAGCGGCACGCATGGATTGTATTTCTGCGCGTATTGCGCGCGCTTGTATAATATCGAGCAG CAGTTGCTGTCCATGTTCGGTGACACGGACGGAAAACGTGATGCAATGCTTCGCTTCACCAAGCC TGTTACAGGGGGTTACTACTTTGCGCCGTCGTTAGATAAATTACTTGCCTTGCTCGAGCACCATCA CCATCACCAT**TGA**
***Rhodococcus erythropolis* aldoxime dehydratase** (Oxd)
ATGGAATCAGCGATTGGCGAGCACCTTCAGTGCCCACGTACACTTACGCGTCGTGTGCCCGACA CGTACACGCCCCCTTTCCCGATGTGGGTCGGCCGCGCGGACGATGCTTTACAGCAAGTTGTAAT GGGTTATTTAGGCGTTCAGTTTCGTGATGAAGACCAACGCCCGGCGGCCCTTCAAGCAATGCGTG ACATTGTCGCAGGGTTTGATTTACCTGACGGTCCCGCCCACCATGACTTAACCCACCACATTGATA ATCAGGGATATGAAAATCTGATTGTAGTCGGATATTGGAAGGACGTTAGTAGTCAGCATCGCTGGT CAACTTCGACTCCCATCGCAAGTTGGTGGGAATCGGAGGATCGTTTATCAGACGGTCTTGGATTC

TTTCGTGAGATCGTAGCTCCACGCGCTGAACAGTTTGAGACCTTATATGCGTTTCAAGAAGACCTT
CCGGGAGTGGGTGCAGTGATGGACGGCATTTCCGGGGAAATTAATGAACACGGATACTGGGGGA
GCATGCGCGAGCGTTTTCCGATTAGCCAAACGGACTGGATGCAGGCCAGTGGTGAGCTGCGCGT
GATTGCGGGCGACCCGGCCGTAGGTGGCCGCGTTGTGGTACGTGGGCACGACAATATTGCGCT
GATTCGTTCCGGACAGGACTGGGCCGATGCCGAGGCAGATGAGCGCTCACTGTATTTGGATGAG
ATCCTTCCTACGCTTCAAAGTGGGATGGACTTCCTGCGCGATAATGGCCCGGCAGTAGGCTGTTA
CTCCAATCGTTTCGTGCGTAACATCGACATCGACGGTAACTTCTTGGATTTAAGCTATAATATCGG
GCACTGGGCGAGTTTGGATCAATTGGAGCGCTGGTCCGAGTCTCACCCGACGCACTTGCGCATC
TTTACGACTTTCTTCCGTGTCGCCGCTGGGTTAAGCAAACTGCGCTTGTACCATGAGGTCTCGGTA
TTCGACGCCGCTGACCAACTGTACGAATACATCAATTGCCATCCTGGGACAGGGATGCTGCGTGA
TGCCGTAACAATCGCCGAACACCTCGAGCACCATCACCATCACCAT**TGA**

***Neurospora crassa* catalase-3** (Cat)

ATGGGTAGCGCCTGCGAATTGTCCGAGTGTCCGGTACGTAAGTCGAATGTAGGTGGCGGTGGCA
CACGTAATCATGACTGGTGGCCCGCCCAGCTTCGCTTGAACATTTTACGTCAACATACACCAGTCA
GCAATCCTCTGGACAAAGACTTTGACTACGCTGCAGCTTTCAAATCATTAGACTACGAAGGTTTGA
AAAAGGATCTGACAAAACTGATGACAGACAGTCAGGACTGGTGGCCAGCGGACTTCGGCCATTAT
GGTGGCTTGTTCATTCGCATGGCATGGCACTCGGCAGGCACATACCGCGTTACTGACGGTCGTG
GTGGCGGAGGCGAGGGGCAACAGCGCTTTGCCCCGTTAAACAGCTGGCCGGATAACGTGTCGTT
GGACAAAGCGCGCCGTCTTCTTTGGCCTATTAAGCAGAAATATGGGAATAAAATTAGTTGGAGTGA
TCTTTTACTTCTTACAGGAAACGTCGCTTTAGAGTCTATGGGATTCAAGACCTTTGGTTTTGCTGGC
GGCCGCCCCGACACATGGGAGGCTGATGAGAGCGTCTATTGGGGTGCTGAGACCACATGGTTAG
GAAATGAAGATCGCTACTCAGAAGGACAAGAAGGACACGAAGGGCACGGGGTAGTGCAAGGCGA
CGAGTCTAAGAAGCAGCATACCGATATTCATAACCGTGACCTGCAATCCCCATTAGCAAGTTCCCA
TATGGGGCTGATCTACGTGAACCCTGAGGGCCCGGACGGTATCCCTGACCCAGTAGCTTCAGCA
AAGGACATTCGTGTCACGTTCGGCCGTATGGCCATGAACGACGAGGAAACTGTAGCGCTGATCG
CAGGGGGACATTCCTTCGGTAAGACTCACGGGGCCGGACCCACCCATCACGTGGGGAAAGAACC
CGAGGCAGCCCCCATCGAGCATCAGGGCCTTGGTTGGGCGAATTCATTTGGACAGGGTAAAGGC
CCAGACACTATCACCTCAGGTTTGGAGGTTACATGGACACCGACCCCAACCAAATGGGGAATGGG
ATATCTTGAATACTTGTACAAGTTTGATTGGGAGCCCACGAAGTCCCCGGCAGGTGCGAACCAGT
GGGTCGCTAAAAACGCCGAGCCAACCATCCCAGATGCTTACGACCCAAATAAGAAAAAATTGCCC
ACAATGCTTACAACCGACATCGCATTACGTATGGATCCAGCCTATGATAAAATCTGTCGTGATTATT
TAGCAAACCCTGATAAATTCGCAGACGCCTTCGCACGTGCGTGGTTCAAACTTCTTCACCGCGAC
ATGGGACCGCGCACTCGTTGGATCGGTCCCGAGGTCCCTTCGGAGATCTTGCCATGGGAGGATT
ACATCCCCCCTGTGGATTATCAGATCATCGACGACAATGATATCGCGGCTCTTAAAAAGGAGATCT
TGGCGACCGGAGTTGCGCCTAAAAAACTTATCTTCGTTGCTTGGTCCTCGGCGTCGAGTTTTCGC
GGGTCCGACAAACGTGGAGGGGCAAATGGGGCACGTATTCGTTTAGCACCGCAGAACGAGTGGA
AGGTGAATGATCCGTCCACATTGCGTGAAGTGTTGGCAGCGTTAGAGTCGGTTCAACAGAAATTT
AATGATTCGTCTTCGGGAAAGAAAGTAAGTCTTGCCGATCTTATCGTACTTGGAGGCGTGGCAGC
CTTAGAACAGGCTAGCGGGCTTGTGGTTCCGTTTACACCTGGACGTAACGACGCGACACAAGAGC
ATACTGATGTGCATTCCTTTACTCACTTAGAACCTCATGCAGACGGTTTTCGTTCCTATGGGAAAG
GAACCAAACGCGTTCGCACTGAACAATTTCTTATCGACCGTGCTTCTTTGCTTACTTTGAGTGCTC
CCGAACTGACAGCGCTTATTGGTGGGTTGCGTGTTCTGGAAGCTAATTACGATGGCTCCTCATAC
GGAGTTCTTACAAAAACTCCGGGCAAGTTAACGAATGATTACTTTGTTAACTTGTTGGATACGAAC
ACCGCATGGAAGGCTGCAGATAACGAAGGGGAAGTATTCATCGGCTATGATCGTAAAACACACGA
TAAGAAATGGACAGCCACCCGTGCCGATCTGATTTTTGGCGCGCATGCAGAGTTACGTGCGCTTG
CAGAAGTTTATGCTGCGGTGGATGGCGAAGAGAAATTCAAGCGTGACTTTGTGGCGGCATGGCAC
AAAGTTATGAACCTTGACCGTTTTGACCTTAAGCAAGAAGGACGCGGGCAAAATGCTCCGAAACT
GCTCGAGCACCATCACCATCACCAT**TGA**

***Pyrococcus furiosus* tryptophan synthase β subunit** (*Pfu*TrpB)

ATGTGGTTCGGTGAATTTGGTGGTCAGTACGTGCCAGAAACGCTGATTGAACCGCTGAAAGAGCT
GGAAAAAGCTTACAAACGTTTCAAAGATGACGAAGAATTCAATCGTCAACTGAATTACTACCTGAA
AACCTGGGCAGGTCGTCCAACCCCACTGTACTACGCAAAACGCCTGACTGAAAAAATCGGTGGTG
CTAAAATCTACCTGAAACGTGAAGACCTGGTTCACGGTGGTGCACACAAGACCAACAACGCCATC
GGTCAGGCACTGCTGGCAAAGTTCATGGGTAAAACTCGTCTGATCGCTGAGACCGGTGCTGGTCA

GCACGGCGTAGCGACTGCAATGGCTGGTGCACTGCTGGGCATGAAAGTGGACATTTACATGGGT
GCTGAGGACGTAGAACGTCAGAAAATGAACGTATTCCGTATGAAGCTGCTGGGTGCAAACGTAAT
TCCAGTTAACTCCGGTTCTCGCACCCTGAAAGACGCAATCAACGAGGCTCTGCGTGATTGGGTGG
CTACTTTTGAATACACCCACTACCTGATCGGTTCCGTGGTCGGTCCACATCCGTATCCGACCATCG
TTCGTGATTTTCAGTCTGTTATCGGTCGTGAGGCTAAAGCGCAGATCCTGGAGGCTGAAGGTCAG
CTGCCAGATGTAATCGTTGCTTGTGTTGGTGGTGGCTCTAACGCGATGGGTATCTTTTACCCGTTC
GTGAACGACAAAAAAGTTAAGCTGGTTGGCGTTGAGGCTGGTGGTAAAGGCCTGGAATCTGGTAA
GCATTCCGCTAGCCTGAACGCAGGTCAGGTTGGTGTGTTCCATGGCATGCTGTCCTACTTTCTGC
AGGACGAAGAAGGTCAGATCAAACCAACTCACTCCATCGCACCAGGTCTGGATTATCCAGGTGTT
GGTCCAGAACACGCTTACCTGAAAAAAATTCAGCGTGCTGAATACGTGACTGTAACCGATGAAGA
AGCACTGAAAGCGTTCCATGAACTGAGCCGTACCGAAGGTATCATCCCAGCTCTGGAATCTGCGC
ATGCTGTGGCTTACGCTATGAAACTGGCTAAGGAAATGTCTCGTGATGAGATCATCATCGTAAACC
TGTCTGGTCGTGGTGACAAAGACCTGGATATTGTTCTGAAAGTGTCTGGCAACGTGCTCGAGCAC
CATCACCATCACCAT**TGA**

2. 5. References

1. Prier, C. K. & Arnold, F. H. Chemomimetic Biocatalysis: Exploiting the Synthetic Potential of Cofactor-Dependent Enzymes to Create New Catalysts. *J. Am. Chem. Soc.* **137**, 13992–14006 (2015).

2. Maret, W. & Vallee, B. L. Cobalt as Probe and Label of Proteins. *Methods Enzymol.* **226**, 52–71 (1993).

3. Hoffman, B. M. & Petering, D. H. Coboglobins: Oxygen-Carrying Cobalt-Reconstituted Hemoglobin and Myoglobin*. *Proc. Natl. Acad. Sci.* **67**, 637–643 (1970).

4. Yonetani, T., Yamamoto, H. & Iizuka, T. Electron Paramagnetic Resonance Studies of Reversible Oxygenation of Cobalt Myoglobins and Hemoglobins. *J. Biol. Chem.* **249**, 2168–2174 (1974).

5. Bruha, A. & Kincaid, J. R. Resonance Raman Studies of Dioxygen Adducts of Co-Substituted Heme Proteins and Model Compounds. Vibrationally Coupled Dioxygen and the Issue of Multiple Structures and Distal Side Hydrogen Bonding. *J. Am. Chem. Soc.* **110**, 6006–60014 (1988).

6. Proniewicz, L. M. & Kincaid, J. R. A Quantitative Clarification of Vibrationally Coupled Dioxygen in the Resonance Raman Spectra of Cobalt-Substituted Heme Proteins and Model Compounds. *J. Am. Chem. Soc.* **112**, 675–681 (1990).

7. Manesis, A. C. *et al.* A Biochemical Nickel(I) State Supports Nucleophilic Alkyl Addition: A Roadmap for Methyl Reactivity in Acetyl Coenzyme A Synthase. *Inorg. Chem.* **58**, 8969–8982 (2019).

8. Smith, A. T. *et al.* Cobalt cystathionine β-synthase: A cobalt-substituted heme protein with

a unique thiolate ligation motif. *Inorg. Chem.* **50**, 4417–4427 (2011).

9. Bonander, N. *et al.* The metal site of Pseudomonas aeruginosa azurin, revealed by a crystal structure determination of the Co(II) derivative and co-EPR spectroscopy. *Proteins Struct. Funct. Genet.* **27**, 385–394 (1997).

10. Di Bilio, A. J. *et al.* Electronic absorption spectra of M(II)(Met121X) azurins (MCo, Ni, Cu; XLeu, Gly, Asp, Glu): charge-transfer energies and reduction potentials. *Inorganica Chim. Acta* **198**–**200**, 145–148 (1992).

11. Sreenilayam, G., Moore, E. J., Steck, V. & Fasan, R. Metal Substitution Modulates the Reactivity and Extends the Reaction Scope of Myoglobin Carbene Transfer Catalysts. *Adv. Synth. Catal.* **359**, 2076–2089 (2017).

12. Moore, E. J., Steck, V., Bajaj, P. & Fasan, R. Chemoselective Cyclopropanation over Carbene Y−H Insertion Catalyzed by an Engineered Carbene Transferase. (2018) doi:10.1021/acs.joc.8b00946.

13. 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).

14. Dydio, P. *et al.* An artificial metalloenzyme with the kinetics of native enzymes. *Science (80-.).* **354**, 102–106 (2016).

15. Yu, Y. *et al.* A Designed Metalloenzyme Achieving the Catalytic Rate of a Native Enzyme. *J. Am. Chem. Soc.* **137**, 11570–11573 (2015).

16. Miner, K. D. *et al.* A designed functional metalloenzyme that reduces O 2 to H 2O with over one thousand turnovers. *Angew. Chemie - Int. Ed.* **51**, 5589–5592 (2012).

17. Mirts, E. N., Petrik, I. D., Hosseinzadeh, P., Nilges, M. J. & Lu, Y. A designed heme-[4Fe-4S] metalloenzyme catalyzes sulfite reduction like the native enzyme. *Science* **361**, 1098–1101 (2018).

18. Olshansky, L. *et al.* Artificial Metalloproteins Containing Co4O4 Cubane Active Sites. *J. Am. Chem. Soc.* **140**, 2739–2742 (2018).

19. Ohashi, M. *et al.* Preparation of artificial metalloenzymes by insertion of chromium (III) Schiff base complexes into apomyoglobin mutants. *Chemtracts* **17**, 359–365 (2004).

20. Oohora, K., Kihira, Y., Mizohata, E., Inoue, T. & Hayashi, T. C(sp 3)−H Bond Hydroxylation Catalyzed by Myoglobin Reconstituted with Manganese Porphycene. *J. Am. Chem. Soc* **135**, 17 (2013).

21. Shimizu, T. *et al.* Gaseous O2, NO, and CO in Signal Transduction: Structure and Function Relationships of Heme-Based Gas Sensors and Heme-Redox Sensors. *Chemical Reviews* vol. 115 6491–6533 (2015).

22. Ortiz de Montellano, P. R. Hydrocarbon hydroxylation by cytochrome P450 enzymes. *Chem. Rev.* **110**, 932–948 (2010).

23. Huang, X. & Groves, J. T. Oxygen Activation and Radical Transformations in Heme Proteins and Metalloporphyrins. *Chemical Reviews* vol. 118 2491–2553 (2018).

24. Yonetani, T., Yamamoto, H., And, $ & Woodrow, G. V. *Studies on Cobalt Myoglobins and Hemoglobins I. PREPARATION AND OPTICAL PROPERTIES OF MYOGLOBINS AND HEMOGLOBINS CONTAINING COBALT PROTO-, MESO-, AND DEUTEROPORPHYRINS AND THERMODYNAMIC CHARACTERIZA-TION OF THEIR REVERSIBLE OXYGENATION**. *THE JOURNAL OF BIOLOGICAL CHEMISTRY* vol. 249

(1974).

25. Chien, J. C. & Dickinson, L. C. Electron paramagnetic resonance of single crystal oxycobaltmyoglobin and deoxycobaltmyoglobin. *Proc. Natl. Acad. Sci. U. S. A.* **69**, 2783–2787 (1972).

26. Mackin, H. C., Tsubaki, M. & Yu, N. T. Resonance Raman Studies of Co—O2 and O—O Stretching Vibrations in Oxy-Cobalt Hemes. *Biophys. J.* **41**, 349–357 (1983).

27. Kleingardner, J. G., Kandemir, B. & Bren, K. L. Hydrogen Evolution from Neutral Water under Aerobic Conditions Catalyzed by Cobalt Microperoxidase-11. *J. Am. Chem. Soc.* **136**, 4–7 (2014).

28. Sommer, D. J. *et al.* Reengineering cyt b_{562} for hydrogen production: A facile route to artificial hydrogenases. *Biochim. Biophys. Acta - Bioenerg.* **1857**, 598–603 (2016).

29. Kandemir, B., Chakraborty, S., Guo, Y. & Bren, K. L. Semisynthetic and Biomolecular Hydrogen Evolution Catalysts. *Inorg. Chem.* **55**, 467–477 (2016).

30. Wen, X., Wang, Y. & Zhang, X. P. Enantioselective radical process for synthesis of chiral indolines by metalloradical alkylation of diverse C(sp3)-H bonds. *Chem. Sci.* **9**, 5082–5086 (2018).

31. Li, C. *et al.* Catalytic Radical Process for Enantioselective Amination of C(sp3)−H Bonds. *Angew. Chemie - Int. Ed.* **57**, 16837–16841 (2018).

32. Wang, Y., Wen, X., Cui, X. & Zhang, X. P. Enantioselective Radical Cyclization for Construction of 5-Membered Ring Structures by Metalloradical C-H Alkylation. *J. Am. Chem. Soc.* **140**, 4792–4796 (2018).

33. Dzik, W. I., Xu, X., Zhang, X. P., Reek, J. N. H. & De Bruin, B. Carbene Radicals in CoII(por)-catalyzed olefin cyclopropanation. *J. Am. Chem. Soc.* **132**, 10891–10902 (2010).

34. Lu, H. *et al.* Experimental evidence for cobalt(III)-carbene radicals: Key intermediates in cobalt(II)-based metalloradical cyclopropanation. *J. Am. Chem. Soc.* **133**, 8518–8521 (2011).

35. Ikeno, T. *et al.* Cobalt-Carbene Complex with Single-Bond Character: Intermediate for the Cobalt Complex-Catalyzed Cyclopropanation. *J. AM. CHEM. SOC* **124**, 15152–15153 (2002).

36. Kan, S. B. J. *et al.* Catalytic iron-carbene intermediate revealed in a cytochrome c carbene transferase . *Proc. Natl. Acad. Sci.* **115**, 7308–7313 (2018).

37. Hayashi, T. *et al.* Capture and characterization of a reactive haem–carbenoid complex in an artificial metalloenzyme. *Nat. Catal.* **1**, 578–584 (2018).

38. Jensen, K. P. & Ryde, U. Comparison of the chemical properties of iron and cobalt porphyrins and corrins. *ChemBioChem* **4**, 413–424 (2003).

39. Liu, Y. *et al.* Iron- And cobalt-catalyzed C(sp3)-H bond functionalization reactions and their application in organic synthesis. *Chemical Society Reviews* vol. 49 5310–5358 (2020).

40. Chen, Y. & Zhang, X. P. Asymmetric cyclopropanation of styrenes catalyzed by metal complexes of D2-symmetrical chiral porphyrin: Superiority of cobalt over iron. *J. Org. Chem.* **72**, 5931–5934 (2007).

41. Theorell, H. Can the Iron in Peroxidase be Replaced by Manganese? *Nature* vol. 156 474 (1945).

42. Gjessing, E. C. & Sumner, J. B. No Title. *Arch. Biochem. Biophys.* **1**, 1 (1942).

43. Teale, F. W. J. Cleavage of the haem-protein link by acid methylethylketone. *BBA - Biochim. Biophys. Acta* **35**, 543 (1959).

44. Woodward, J. J., Martin, N. I. & Marletta, M. A. An Escherichia coli expression-based method for heme substitution. *Nat. Methods* **4**, 43–45 (2007).

45. 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 (2011).

46. Majtan, T., Freeman, K. M., Smith, A. T., Burstyn, J. N. & Kraus, J. P. Purification and characterization of cystathionine β-synthase bearing a cobalt protoporphyrin. *Arch. Biochem. Biophys.* **508**, 25–30 (2011).

47. Wang, M.-Y. R., Hoffman, B. M. & Hollenberg, P. F. Cobalt-substituted Horseradish Peroxidase*. *J. Biol. Chem.* **252**, 6268–6275 (1977).

48. Choi, S. *et al. Structural Correlations and Vinyl Influences in Resonance Raman Spectra of Protoheme Complexes and Proteins*. *C. R. Hebd. Seances Acad. Sci., Ser. C* vol. 104 (1982).

49. Wagner, G. C., Gunsalus, I. C., Wang, M.-Y. R. & Hoffman, B. M. Cobalt-substituted Cytochrome P-450$_{cam}$. *J. Biol. Chem.* **256**, 6266–6273 (1981).

50. 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).

51. Woodward, J. J., Martin, N. I. & Marletta, M. A. An Escherichia coli expression-based

method for heme substitution. *Nat. Methods* **4**, 43–45 (2007).

52. Winter, M. B., Klemm, P. J., Phillips-Piro, C. M., Raymond, K. N. & Marletta, M. A. Porphyrin-substituted H-NOX proteins as high-relaxivity MRI contrast agents. *Inorg. Chem.* **52**, 2277–2279 (2013).

53. Winter, M. B. *et al.* Ru-porphyrin protein scaffolds for sensing O2. *J. Am. Chem. Soc.* **132**, 5582–5583 (2010).

54. Nierth, A. & Marletta, M. A. Direct meso-alkynylation of metalloporphyrins through gold catalysis for hemoprotein engineering. *Angew. Chemie - Int. Ed.* **53**, 2611–2614 (2014).

55. Majtan, T., Frerman, F. E. & Kraus, J. P. Effect of cobalt on *Escherichia coli* metabolism and metalloporphyrin formation. *BioMetals* **24**, 335–347 (2011).

56. Bordeaux, M., Tyagi, V. & Fasan, R. Highly Diastereoselective and Enantioselective Olefin Cyclopropanation Using Engineered Myoglobin-Based Catalysts. *Angew. Chemie Int. Ed.* **54**, 1744–1748 (2015).

57. Hambright, P. *et al.* A Dissociative Mechanism for the Dithionite Reduction of Cobalt(III) Myoglobin. *Inorganica Chim. Acta* **92**, 167–172 (1984).

58. Li, C. zhong, Nishiyama, K. & Taniguchi, I. Electrochemical and spectroelectrochemical studies on cobalt myoglobin. *Electrochim. Acta* **45**, 2883–2888 (2000).

59. Berry, E. A. & Trumpower, B. L. Simultaneous determination of hemes a, b, and c from pyridine hemochrome spectra. *Anal. Biochem.* **161**, 1–15 (1987).

60. Ranquet, C., Ollagnier-de-Choudens, S., Loiseau, L., Barras, F. & Fontecave, M. Cobalt Stress in *Escherichia coli*: The Effect on the Iron-Sulfur Proteins. *J. Biol. Chem.* **282**,

30442–30451 (2007).

61. Fantino, J. R., Py, B., Fontecave, M. & Barras, F. A genetic analysis of the response of Escherichia coli to cobalt stress. *Env. Microbiol* **12**, 2846–2857 (2010).

62. Buller, A. R. *et al.* Directed evolution of the tryptophan synthase β-subunit for stand-alone function recapitulates allosteric activation. *Proc. Natl. Acad. Sci.* **112**, 14599–14604 (2015).

63. Sudhamsu, J. *et al.* Co-expression of ferrochelatase allows for complete heme incorporation into recombinant proteins produced in E. coli. *Protein Expr Purif* **73**, 78–82 (2010).

64. Honda, Y., Nanasawa, K. & Fujii, H. Coexpression of 5-Aminolevulinic Acid Synthase Gene Facilitates Heterologous Production of Thermostable Cytochrome P450, CYP119, in Holo Form in Escherichia coli. *ChemBioChem* **19**, 2156–2159 (2018).

65. Datta-Gupta, N. Preparation and Spectra of Some Cobaltic Complexes of meso-Tetra-p-Tolylporphyrin. *J. Inorg. Nucl. Chem.* **33**, 4219–4225 (1971).

66. Stynes, D. V., Stynes, H. C., Brian, R. & Ibers, J. A. Thermodynamics of Ligand and Oxygen Binding to Cobalt Protoporphyrin IX Dimethyl Ester in Toluene Solution. *J. Am. Chem. Soc.* **95**, 1796–1801 (1973).

67. Komar, A. A. Unraveling co-translational protein folding: Concepts and methods. *Methods* **137**, 71–81 (2018).

68. Komar, A. A., Kommer, A., Krasheninnikov, I. A. & Spirin, A. S. Cotranslational Folding of Globin. *J. Biol. Chem.* **272**, 10646–10651 (1997).

69. Haskamp, V. *et al.* The radical SAM protein HemW is a heme chaperone. *J. Biol. Chem.*

293, 2558–2572 (2017).

70. Sweeny, E. A. *et al.* Glyceraldehyde-3-phosphate dehydrogenase is a chaperone that allocates labile heme in cells. *J. Biol. Chem.* **293**, 14557–14568 (2018).

71. Brandenberg, O. F., Fasan, R. & Arnold, F. H. Exploiting and engineering hemoproteins for abiological carbene and nitrene transfer reactions. *Curr Opin Biotechnol* **47**, 102–111 (2017).

72. Bhalla, T. C., Kumar, V., Kumar, V., Thakur, N. & Savitri. Nitrile Metabolizing Enzymes in Biocatalysis and Biotransformation. *Applied Biochemistry and Biotechnology* vol. 185 925–946 (2018).

73. Bugg, T. D. H., Williamson, J. J. & Rashid, G. M. M. Bacterial enzymes for lignin depolymerisation: new biocatalysts for generation of renewable chemicals from biomass. *Current Opinion in Chemical Biology* vol. 55 26–33 (2020).

74. Park, S. *et al.* T hermophilic cytochrome P450 (CYP119) from Sulfolobus solfataricus : high resolution structure and functional properties. *J. Inorg. Biochem.* **91**, 491–501 (2002).

75. Sawai, H. *et al.* X-ray Crystal Structure of Michaelis Complex of Aldoxime Dehydratase. *J. Biol. Chem.* **284**, 32089–32096 (2009).

76. Shrestha, R., Huang, G., Meekins, D. A., Geisbrecht, B. V & Li, P. Mechanistic Insights into Dye-Decolorizing Peroxidase Revealed by Solvent Isotope and Viscosity Effects. *ACS Catal.* **7**, 6352–6364 (2017).

77. Colpa, D. I. & Fraaije, M. W. High overexpression of dye decolorizing peroxidase TfuDyP leads to the incorporation of heme precursor protoporphyrin IX. *J. Mol. Catal. B Enzym.* **134**, 372–377 (2016).

78. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

Chapter 3

Mechanism-guided improvement of Co-substituted hemoprotein production in *E. coli*

\

B.R.W. collected kinetic data for *Ec*HemH, generated HemH homolog LogoPlots, and quantified metal concentrations by ICP-MS. Protein engineering, protein over-expression and purification, and *in vitro* and *in vivo* metal specificity experiments were conducted by L.J.P.

Chapter 2 Chapter 3: Mechanism-guided improvement of Co-substituted hemoprotein production in *E. coli*

Chapter 3 3. 1. Abstract

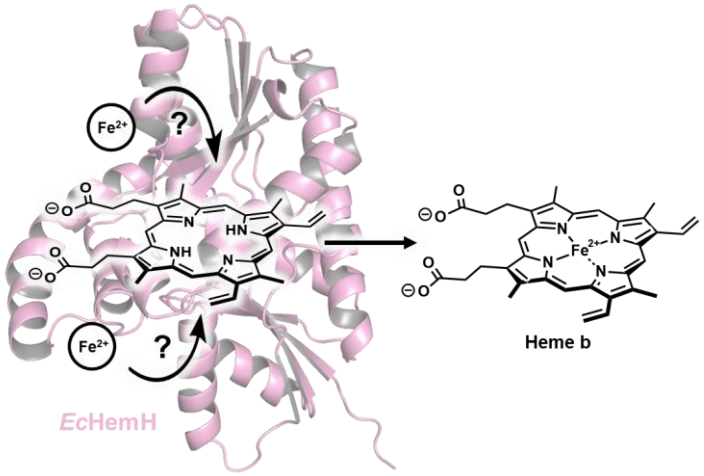
Exchanging the native iron of heme for other metals yields artificial metalloproteins with new properties for spectroscopic investigations and biocatalysis. Recently, we reported a method for biosynthesis and incorporation of a non-natural metallocofactor, cobalt protoporphyrin IX (CoPPIX), into hemoproteins using the common laboratory strain *E. coli* BL21(DE3). This discovery inspired us to explore the specificity determinants for metallocofactor biosynthesis in *E. coli*. Here, we report detailed kinetic analysis of the ferrochelatase responsible for metal insertion, *Ec*HemH. This enzyme exhibits less than a two-fold preference for Fe^{2+} over the non-native Co^{2+} substrate *in vitro*. To test how *Ec*HemH impacts bioincorporation, we used a surrogate metal specificity screen to identify variants with altered metal insertion preferences. This engineering led to a variant with a ~30-fold shift in specificity towards Co^{2+}. When assayed *in vivo*, however, the impact of this mutation is small compared to the effects of altering the external metal concentrations. These data suggest that Co-incorporation into protoporphyrin IX in BL21 is principally controlled by this strain's impaired ability to maintain transition metal homeostasis. With this knowledge, we generated a method for CoPPIX production and incorporation in rich media, which yields the Co-substituted hemoproteins with >95% cofactor purity and yields comparable to standard expression protocols for the analogous native hemoproteins.

Chapter 4 3. 2. Introduction

Proteins bearing unnatural metallocofactors are useful tools to study metalloprotein function, can expand the reactivity of natural proteins, and catalyze abiological reactions.[1,2] In particular, metal-substituted hemoproteins have been extensively explored because they have unique reactivity,[3–5] can serve as spectroscopic probes,[6–9]and have unique imaging properties.[8] While there are diverse methods to generate proteins loaded with unnatural metalloporphyrin cofactors both *in vivo* and *in vitro*, [3,10–12] many of these strategies require exogenous synthesis of the unnatural metalloporphyrin and inefficient methods for inserting the cofactor into the protein. We recently reported a method for the *de novo* biosynthesis of the non-native heme analog, cobalt protoporphyrin IX (CoPPIX), in *E. coli* and its incorporation into hemoproteins.[13] This fully biosynthetic method streamlined the production of CoPPIX artificial metalloenzymes and inspired this inquiry into the determinants of metal specificity in the metalloporphyrin assembly in *E. coli* BL21 (DE3).

Nature typically assembles metalloproteins and metallocofactors with impressive fidelity. Nevertheless, there are reported examples of improperly metalated heme cofactors occurring in living systems, each of which is hypothesized to be a marker of perturbed iron homeostasis. In humans, trace amounts of zinc protoporphyrin IX (ZnPPIX) in bone marrow is a biomarker for iron deficiency.[14] Majtan et al. found that *E. coli* BL21, when passaged for several days through iron-poor, cobalt-rich media, adventitiously biosynthesized and incorporated CoPPIX into both endogenous[15] and heterologously expressed[16] hemoproteins. We reported an expansion of this finding by showing that the standard laboratory strain *E. coli* BL21(DE3) innately possesses the ability to biosynthesize and incorporate CoPPIX, without passaging. Addition of an inexpensive cobalt salt to iron-deficient media was sufficient to produce cobalt substituted hemoproteins from diverse fold families with >95% cobalt loadings. Others similarly found that ZnPPIX can be over-produced by an engineered B-derived strain of *E. coli* under iron-poor, zinc-rich fermentation conditions.[17] These reports indicate that heme biosynthesis in *E. coli* may be manipulated toward

production of non-natural metalloporphyrin cofactors. However, these cases of alternative heme metalation in *E. coli* depend on iron-deficient growth conditions to limit production of the native heme cofactor. Such conditions severely limit cell growth and, correspondingly, the titers of expressed Co-substituted hemoproteins are low when compared to standard expression of hemoproteins. We envisioned that by understanding determinants of cofactor metalation specificity, we might engineer more efficient production and incorporation of the CoPPIX cofactor.



F**igure 3.1.** Scheme for the reaction of *Ec*HemH with ferrous iron and protoporphyrin IX (PPIX) to make heme b. The direction of metal insertion and residues responsible for orienting the metal prior to insertion are debated. The pink cartoon is an AlphaFold model of *Ec*HemH.

Metalation of tetrapyrrole-derived cofactors (heme, siroheme, cobalamin, and others) is a physiologically irreversible process catalyzed by a diverse class of enzymes called chelatases. In *E. coli*, the final step in heme b biosynthesis is incorporation of ferrous iron into protoporphyrin IX (PPIX), catalyzed by PPIX ferrochelatase (*Ec*HemH) (Figure 1A). Eukaryotic HemH homologs, denoted FECH, especially those from *S. cerevisiae* and *H. sapiens*, are well-studied.[18–20] *Ec*HemH is a membrane-associated enzyme,previously expressed and purified,[21,22] but no kinetic characterization of *Ec*HemH, or HemH from any other gram-negative bacterium, has been reported,. The location of the productive metal binding site and the face of the porphyrin from which metal is inserted are also debated in the literature,[20,23] reflecting the challenges associated

with study of ferrochelatases. Nevertheless, *in vitro* incorporation of non-native metals by several ferrochelatases (HemH homologs) is well-known.[20,24–27] To date, it is unclear how chelatases tune their activity for different metals, although distortion of the porphyrin, [28–30]specificity-determining active site residues,[31,32] and exogenous metal delivery[25] are suggested to contribute. Here, we explore *E. coli* BL21(DE3)'s unusual ability to produce CoPPIX *in vivo*. To better understand the native metal specificity of *Ec*HemH, we measured the enzyme's catalytic activity with Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, and Zn^{2+} *in vitro*. We next asked whether a more selective cobalt chelatase might enable efficient production and incorporation of CoPPIX, even in iron-rich growth media, and we identified a single point mutant (L13R) which is ~30 times more selective for Co^{2+} over Fe^{2+} than the parent *Ec*HemH. We analyzed the ability of *Ec*HemH variants to discriminately produce CoPPIX over heme in rich growth media and found that, while variants may produce altered specificity profiles *in vivo*, other metabolic factors play an outsized role in determining CoPPIX incorporation into co-expressed hemoproteins. Under sufficiently high cobalt concentrations in rich media and in the absence of a specifically engineered chelatase, BL21(DE3) can produce upwards of 95% cobalt loaded hemoprotein with titers comparable to typical hemoprotein expressions. Our results provide insight into the mechanism and metal selectivity of *Ec*HemH, inform future engineering efforts for metallopophryin biosynthesis, and demonstrate a straightforward and scalable route to cobalt-substituted hemoprotein production in rich media.

Chapter 5 3. 3. Results

3. 3. 1. *In vitro* activity of *E. coli* ferrochelatase

We began our investigation by probing the kinetic mechanism and metal specificity of *Ec*HemH (UniProt accession: A0A140NEM8). This protein was expressed with a C-terminal His-tag, over-expressed in *E. coli* BL21(DE3), and purified by nickel affinity chromatography yielding approximately 10 mg *Ec*HemH per L culture. While poly-His tags can interfere with the metal binding properties of some enzymes, such constructs for chelatase homologs have been used previously, without complication.[18,33–35] Sodium cholate was added to the lysis, purification, and storage buffers, as the enzyme has been shown to require detergents for solubility.[22] *E*cHemH is predicted to be membrane associated due to the presence of a 12-residue segment implicated in membrane association.[36]

Using a simple spectroscopic assay, we investigated the ability of *E. coli* ferrochelatase to catalyze insertion of Fe^{2+}, Co^{2+}, Ni^{2+}, Zn^{2+}, Cu^{2+}, Mn^{2+}, and Mg^{2+} into PPIX. The catalytic activity of ferrochelatase is well-suited to kinetic analysis by UV-visible spectroscopy, as the metalation of PPIX imparts unique spectral shifts to the Soret (~400 nm) and Q bands (~500-600 nm). We detected product formation for all metals tested except for Mn^{2+} and Mg^{2+} (Figure 3.2A and C, Figure 3.3). This qualitative reactivity profile is in good agreement with reported activity of previously characterized ferrochelatase enzymes.[20] To quantify the native iron chelatase activity of *Ec*HemH, we measured initial rates of heme formation spectrophotometrically with variable PPIX and Fe^{2+} concentrations at room temperature (Figure 3.2B). The resulting data fit well to an ordered sequential mechanism where PPIX binds first, followed by Fe^{2+} binding, as described by Scheme 3.1. The resulting fit yielded a k_{cat} of 30 min^{-1}, K_M^{Fe} of 0.48 μM Fe^{2+}, and K_M^{PPIX} of 1.1 μM PPIX. (Table 3.1).[27]

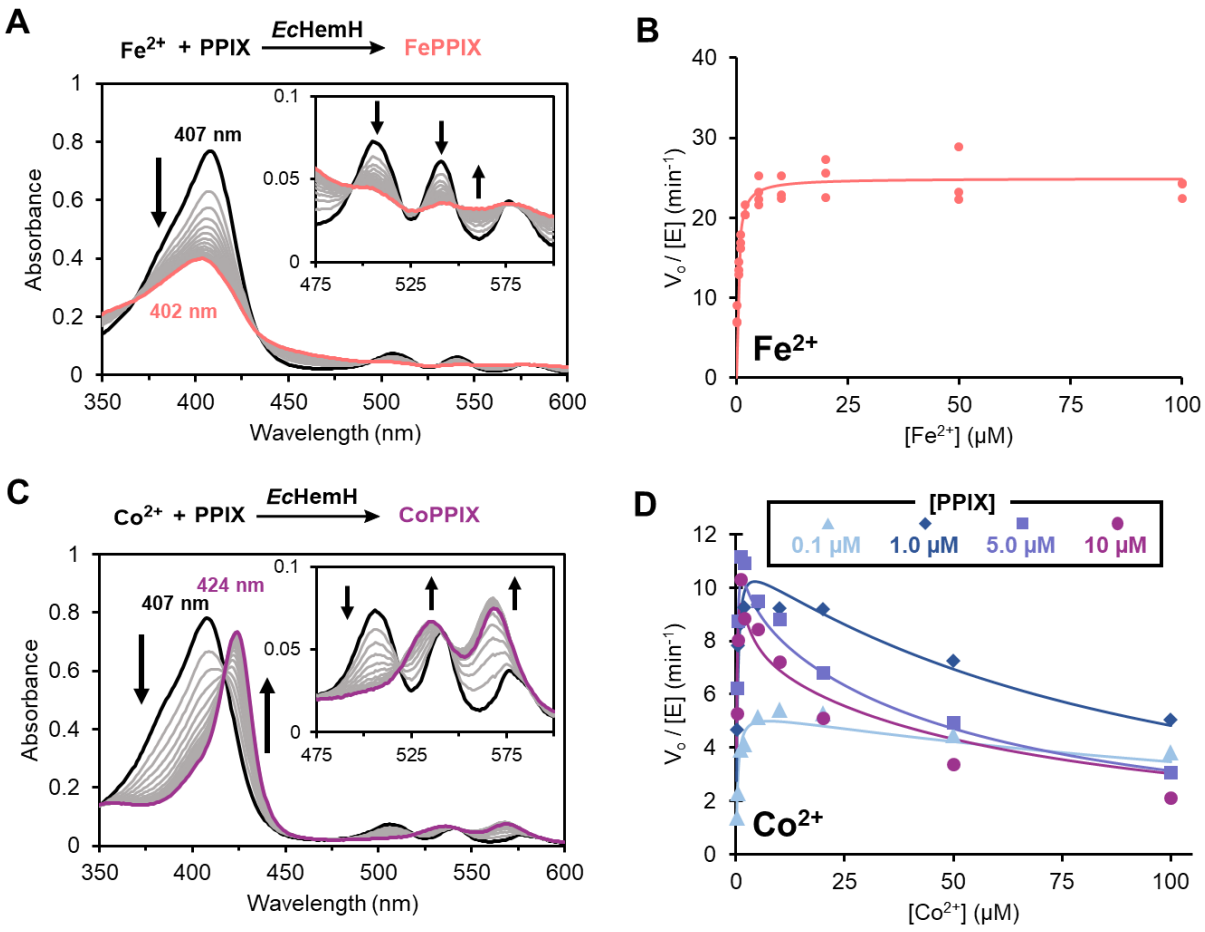


Figure 3.2. *E. coli* ferrochelatase (*Ec*HemH)-catalyzed insertion of cobalt and iron into protoporphyrin IX (PPIX). A) Progress spectra of Fe^{2+} insertion into PPIX by *Ec*HemH. The black trace represents the UV-visible spectrum of 7.5 μM protoporphyrin IX in reaction buffer (100 mM MOPS, 400 mM NaCl, 0.2% w/v Tween 80 at pH 7.0) with 100 nM *Ec*HemH. The metalation reaction was initiated at 25 °C with the addition of 10 μM of Fe^{2+}. Grey lines indicate absorption spectra taken during the reaction progress, and the pink line represents the absorption spectrum of heme b (FePPIX) at the end of the reaction. B) Initial rates of *Ec*HemH-catalyzed heme production plotted as a function of the Fe^{2+} concentration. The concentration of PPIX was 5 μM. These reactions were conducted at 25 °C. The solid line represents the best fit of the data by DynaFit 4 to the scheme shown in Figure 3.4A and corresponds to the kinetic parameters shown in Table 1. C) Progress spectra of Co^{2+} insertion into PPIX by EcHemH. The black trace represents the UV-visible spectrum of 7.5 μM protoporphyrin IX in reaction buffer with 100 nM EcHemH. The metalation reaction was initiated at 25 °C with the addition of 10 μM of Co^{2+}. Grey lines indicate absorption spectra taken during the reaction progress, and the purple line represents the absorption

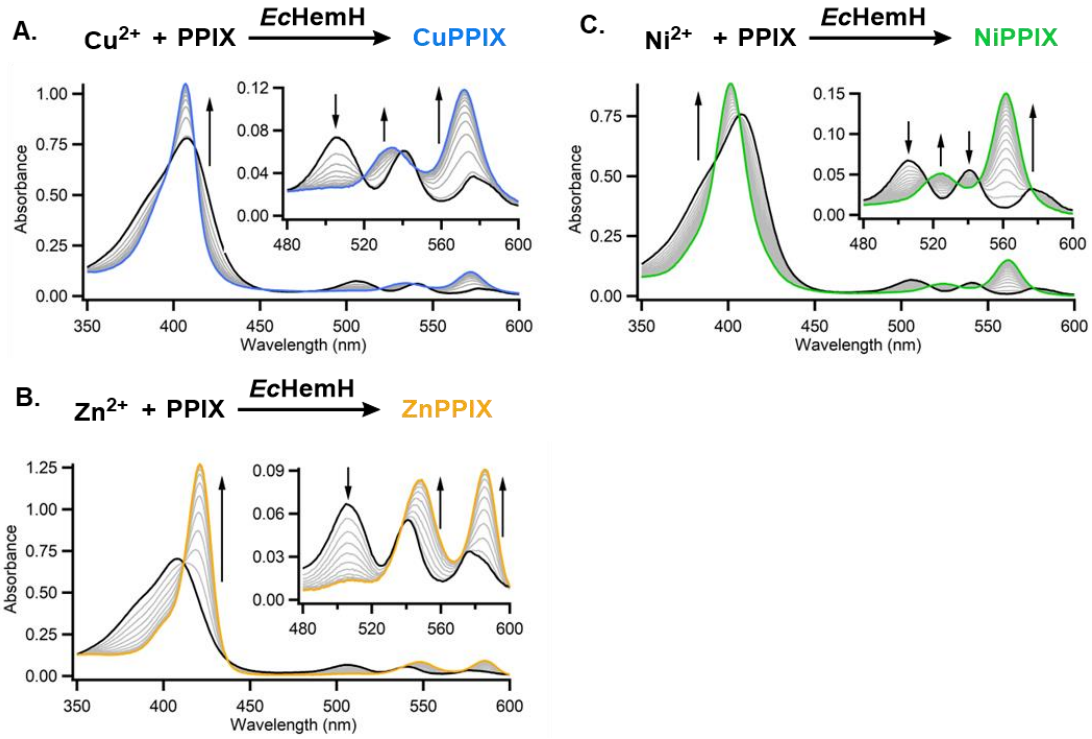


Figure 3.3. Spectra collected of ferrochelatase-calalyzed metal insertion reactions with various transition metals. Black traces indicate the absorption spectra of 7.5 µM protoporphyrin IX in 100 mM MOPS, 400 mM NaCl, 0.2% (w/v) Tween 80, pH 7.0 with 100 nM EcHemH. Reactions were initiated by addition of the indicated divalent metal chloride salt to a final concentration of 10 µM. Gray lines indicate the absorption spectra collected during reaction progress. Colored spectra in each graph reflects the absorbance of the final metalloporphyrin products. Arrows indicate the direction of absorbance changes during reaction progress. Manganese chloride and magnesium chloride yielded no change in absorption spectra upon addition to protoporphyrin IX and EcHemH reaction mixtures.

We next sought to test the kinetics of CoPPIX formation by *Ec*HemH. In an identical manner as with Fe^{2+}, we measured initial rates of CoPPIX production at 5 µM PPIX while varying the Co^{2+} concentration. In contrast to the kinetic behavior observed with Fe^{2+}, we observed a decrease in the rate of CoPPIX production with high concentrations of Co^{2+} (Figure 3.2D). This substrate inhibition has been observed previously with yeast[36] and mouse[31] FECH proteins, as well as with coproporphyrin (III) ferrochelatase (CpfC) of other species.[37,38] To characterize this inhibition further, we measured initial rates of CoPPIX production while varying the concentrations of both Co^{2+} and PPIX (Figure 3.2D). The resulting data show that increasing concentrations of PPIX exacerbate the Co^{2+} inhibitory effect. These data mirror observations by Davidson, et al. in studies of human HemH with a PPIX analog, mesoporphyrin IX, and Zn^{2+}.[25]

To better understand the nature of Co^{2+} inhibition, we turned to global analysis of the Co^{2+} initial rate data to identify the simplest kinetic mechanism that recapitulates this complex kinetic behavior. A complete accounting of each kinetic model that we tested is provided in the Supplemental Figures section 3.7.1(Figure 3.23-27). Notably, none of the existing kinetic models for chelatases in the literature fit the data. [25,27] On the basis of the simplest model that fit a global kinetic analysis, we propose an *Ec*HemH kinetic mechanism that includes two kinetically distinct metal binding modes, only one of which is kinetically productive (Figure 3.4A). A kinetically productive pathway is enabled when PPIX binds first, then the Co^{2+} substrate binds in metal-insertion-active mode, such that the cobalt ion is inserted into the porphyrin. A kinetically unproductive pathway is enabled when the substrate Co^{2+} binds in an alternative, noncompetitive mode, either prior or subsequent to PPIX binding. When PPIX is bound, the affinity of the Co^{2+} ion for the noncompetitive inhibitory mode is enhanced. Because of this PPIX-mediated change in affinity for the noncompetitive metal binding mode, the extent to which increasing concentrations of Co^{2+} inhibit the enzyme varies with the PPIX concentration. At low Co^{2+} concentrations, binding in the catalytically productive metal mode dominates, and the initial velocities increase as the concentration of PPIX increases. At high Co^{2+} concentrations, the noncompetitive binding mode is increasingly populated as the PPIX concentration increases, resulting in increasing PPIX-dependent inhibition with increasing Co^{2+} concentration. This model effectively reproduces the observed behavior, as seen in the best fit lines in Figure 3.3D.

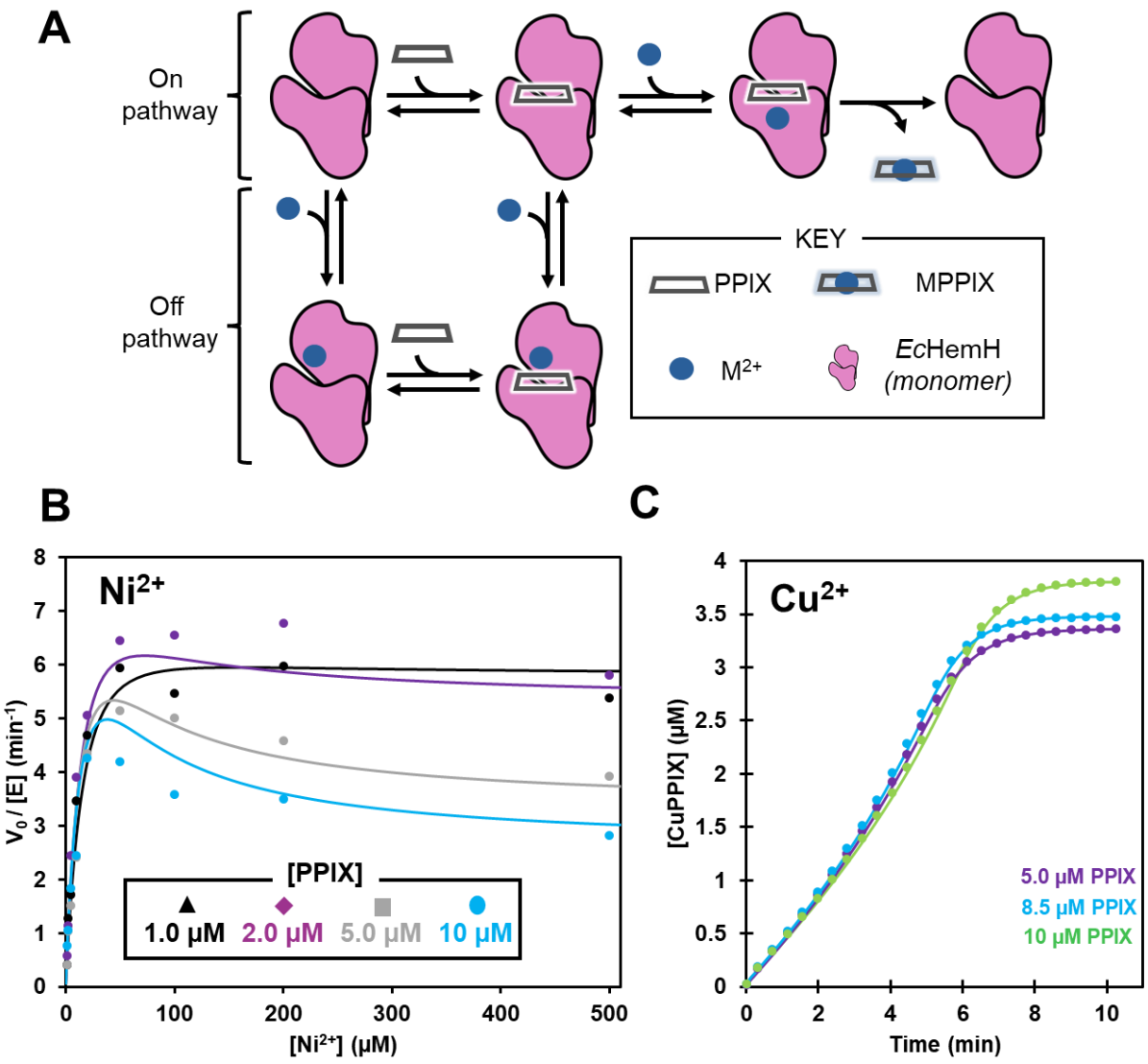


Figure 3.4. Updated kinetic model for E. coli ferrochelatase and additional supporting data. A) Cartoon representation of proposed kinetic scheme (See also: Scheme S1). The pink blob represents EcHemH, the blue circle represents divalent metal cation (M^{2+}), and the black rectangle represents protoporphyrin IX (PPIX). M^{2+} binds in both productive "on-pathway" and unproductive "off-pathway" modes, in the presence and absence of PPIX, respectively. The binding of PPIX stabilizes the off-pathway M^{2+} binding mode. Microscopic kinetic parameters for this model for each metal can be found in Table 3.1 B) Initial velocities of EcHemH-catalyzed NiPPIX formation as a function of Ni^{2+} concentration. Reactions were conducted at 25 °C with 1-10 μM PPIX and 100 nM EcHemH. The reaction was initiated by the addition of Ni^{2+}. The solid lines represent the best global fit of the data by DynaFit 4 to Scheme 3.S1 and correspond to the kinetic parameters shown in Table 3.1. C) Progress curves tracking CuPPIX production over time. Reactions were conducted at 25 °C with 5 μM PPIX and 100 nM EcHemH, and were initiated by the addition of 5 μM (black) or 10 μM (blue) Cu^{2+}. Solid lines represent the best global fit of the data by DynaFit 4 to Scheme S1 and correspond the kinetic parameters shown in Table 1. Points are representative experimental data points.

We next tested whether this kinetic model could be used to describe the catalytic insertion of other metals into PPIX. We measured initial rates of NiPPIX production at multiple concentrations of

PPIX and Ni^{2+}. These data also showed an increase in metal inhibition as a function of PPIX (Figure 3.4B) and the proposed mechanism fit these experimental data well. Lastly, we used our model to study Cu^{2+} incorporation into PPIX. Previous studies of Cu^{2+} with ferrochelatase homologs yielded distinctive 'S' shaped progress curves due to substantial substrate inhibition.[36] We tracked CuPPIX formation spectroscopically with three different PPIX and Cu^{2+} concentrations and observed the same S-shape in the progress curves. Satisfyingly, we found that the experimental data was well recapitulated by the model (Figure 3.4C), including the characteristic S-feature. Initial velocities of *Ec*HemH-catalyzed ZnPPIX production as a function of Zn^{2+} and PPIX concentration also fit well (Figure 3.5). Together, these data all support the same kinetic mechanism of ferrochelatase activity wherein metal-first binding leads to an unproductive inhibitor complex that is stabilized at higher concentrations of PPIX. On-path reactivity comes from PPIX binding first, followed by divalent metal at a kinetically distinct and catalytically competent site.

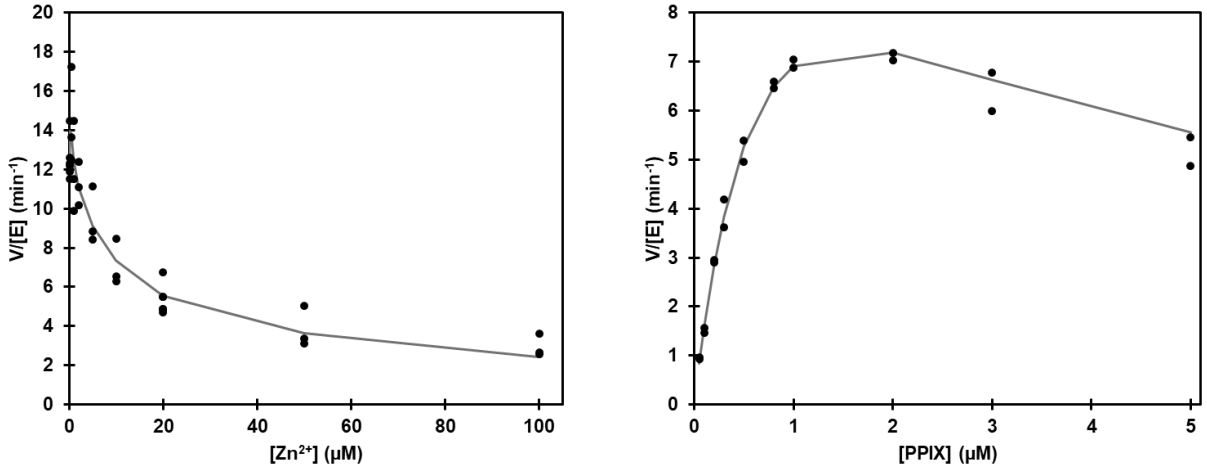


Figure 3.5. Initial velocities of EcHemH-catalyzed insertion of zinc into protoporphyrin IX. A) Initial velocity of EcHemH-catalyzed ZnPPIX production is plotted as a function of the Zn^{2+} concentration. In this experiment, the PPIX concentration was held at 5 μM. B) Initial velocity of EcHemH-catalyzed ZnPPIX production is plotted as a function of the protoporphyrin IX concentration. In this experiment, Zn^{2+} was held at 20 μM. These reactions were conducted in ferrochelatase reaction buffer, 100 mM MOPS, 400 mM NaCl, 0.2% (w/v) Tween 80 at 25 °C. The solid black lines represent the best fit of the data by DynaFit 4 to the uncompetitive substrate inhibition model shown in Figure 3.3A and correspond to the kinetic parameters are shown in Table 3.1.

Table 3.1. Apparent *Ec*HemH kinetic parameters. Values are derived from fitting initial rate data to the uncompetitive inhibition model shown in Figure 3.4A. These parameters are calculated from the microscopic rate constants as described in the Supplemental Information.

metal	k_{cat} (min^{-1})	K_M (μM)	k_{cat}/K_M (/μM/min)
Fe	30	0.48	63
Co	54	1.1	48
Ni	42	79	0.53
Zn	550	0.23	2437
Cu	34	0.94	36

With the new kinetic model in hand, we quantitatively compared the catalytic properties of *Ec*HemH with different transition metals. According to these data, ferrochelatase exhibits the highest catalytic efficiency (k_{cat}/K_M) with Zn^{2+} followed by Fe^{2+}, Co^{2+}, Cu^{2+}, and Ni^{2+} (Table 3.1). Notably, this analysis shows that *Ec*HemH's catalytic efficiency with Co^{2+} is only slightly less than that of the native metal substrate, Fe^{2+}.

3. 3. 2. *Ec*HemH structural analysis highlights promising residues for engineering metal specificity.

Given the modest Fe^{2+} versus Co^{2+} selectivity of *Ec*HemH in vitro, we hypothesized that the promiscuous *Ec*HemH activity with Co^{2+} may be the dominant factor in BL21(DE3)'s ability to produce and incorporate CoPPIX. We further hypothesized that altering the specificity of the native chelatase via protein engineering to further favor Co^{2+} incorporation might allow selective CoPPIX production in iron-rich media. The structural underpinnings of metal specificity in other ferrochelatase homologs are not well understood, and there are only a few examples of variants with altered specificity.[26]

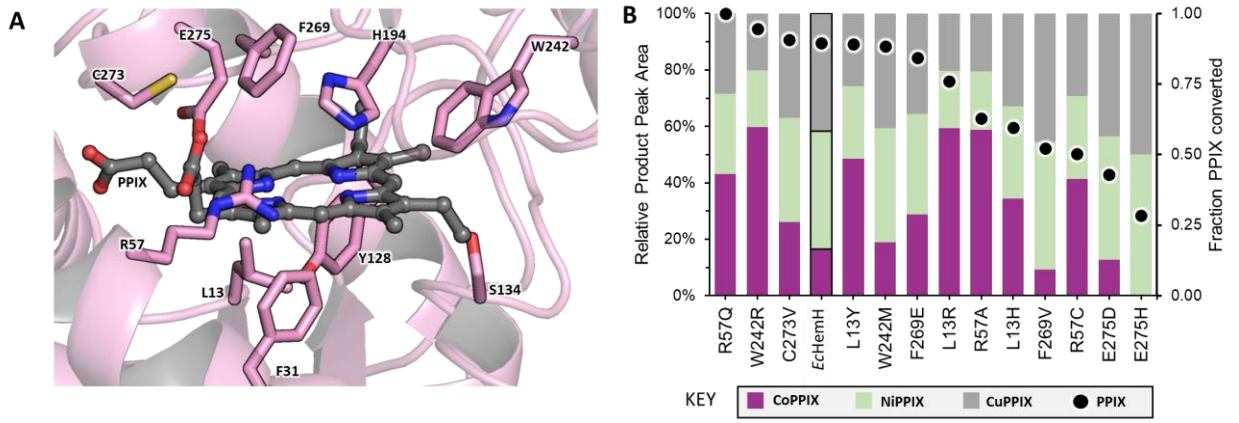


Figure 3.6. Structural model of E. coli ferrochelatase (EcHemH) and selection of residues for site saturation mutagenesis. A) AlphaFold active site model of EcHemH (pink) superimposed with PPIX (grey sticks) from PDB 2HRE. B) Results from re-screen of "hits" from site-saturation mutagenesis of EcHemH. Each variant was screened in cell lysate in biological quadruplicate with 10 μM PPIX and a total of 1 mM divalent metal substrate. The resulting porphyrin distribution was analyzed by UPLC. Bars represent the average relative product peak areas corresponding to CoPPIX (purple), NiPPIX (green), and CuPPIX (grey). Black dots represent the relative amount of PPIX substrate remaining, a measure of each variant's total activity. Parent EcHemH is indicated by black borders.

We used a structural model generated using AlphaFold to compare how structural features inferred from HemH homologs might map onto *Ec*HemH.[39,40]. Alignment of *Ec*HemH with the PPIX-bound *Hs*FECH structure (PDB ID: 2HRE) provided useful clues into how *Ec*HemH might interact with the porphyrin substrate. The alignment suggests that the human and *E. coli* homologs are structurally similar with an alpha carbon root mean square deviation of 2.7 Å (Figure 3.6A), despite sharing just 27% sequence identity. The main structural difference between *Hs*FECH and *Ec*HemH is the presence of an iron-sulfur cluster at the C-terminus of *Hs*FECH, the function of which is not known. The identity and position of residues in the active sites of the two structures are nearly identical. Distinctive residues in *Ec*HemH are L13, F31, and C273, which correspond to M76, L92, and H341 in *Hs*FECH, respectively (Figure 3.6A). The differences at L13 (*E. coli*) and M76 (human) are noteworthy because these residues lie directly below the porphyrin substrate.

To gain further insight into the active site, we generated a representative sequence alignment of 5,026 HemH homolog protein sequences. Using this sequence alignment, we generated logo plots for each predicted active site residue shown in Figures 3.7-17B.[37] The conservation patterns

depicted in these logo plots provides insight into the functional role of these residues within the ferrochelatase active site. For example, H194, W242, F269, and E275 (*E. coli* numbering) are >99% conserved, suggesting that these residues are especially important for ferrochelatase function. Additionally, the positions corresponding to L13 and F31 are almost exclusively occupied by hydrophobic residues, suggesting that these residues may interact directly with PPIX and potentially influence the porphyrin conformation. Taken as a whole, these plots provide insight into the possible roles of active sight residues and serve as a launch point for choosing which residues might be most amendable to mutation. We chose to screen site saturation mutagenesis (SSM) libraries at 10 active site residues (Figure 3.6A, 3.7-17A). Based on our structural model, we hypothesized these residues might impart changes in metal specificity either by altering metal chelation or by altering the extent and nature of the porphyrin distortion.

3. 3. 3. Metal multiplexed screening identifies EcHemH residues that impact substrate specificity

Substrate specificity can be a particularly challenging feature to optimize through engineering, since traditional engineering approaches typically monitor activity on a single, model substrate as a proxy for overall enzyme performance.[26,38–40] Instead, we developed a substrate-multiplexed screen (SUMS) with a mixture of metal ions, which provided information on enzyme activity with several substrates in a single experiment. The SUMS approach, as applied to protein engineering, has been shown to facilitate discovery of enzyme variants with altered specificity and identification of distal residues that impact substrate specificity.[41–43]

While relative activity of Fe^{2+} versus Co^{2+} was our principal interest, we found that the specificity between these two metals was exceedingly difficult to monitor through a single time point measurement in cell lysates. At least two factors confounded our analysis. First, Fe^{2+} quickly oxidizes in reaction buffer yielding Fe^{3+}, which is not a substrate for *Ec*HemH. Second, heme b is catabolized in cell lysate, leading to inconsistent ratios of CoPPIX to heme b. To work around

these challenges, we hypothesized that screening on a mixture of non-native metals would provide insight into metal specificity 'hot spots', or residues that provide the most significant changes to metal specificity relative to the parent. These hot spots could subsequently be analyzed to gain insight into their Fe^{2+} specificity using an alternative screening method.

Cell lysates expressing a *Ec*HemH variant were added to a mixture of Co^{2+}, Ni^{2+}, and Cu^{2+} at relatively high concentration (1 mM total M^{2+}) in the presence of 10 μM PPIX, giving pseudo-first order conditions with respect to the metal. The relative distribution of the porphyrin products was assessed using ultra-pressure liquid chromatography (UPLC) and the concentrations of each metal was adjusted such that the parent *Ec*HemH produced approximately equal signal for CoPPIX, NiPPIX, and CuPPIX (See materials and methods section). The resulting data for all 10 SSM libraries are compiled in Figures 3.7-16A).

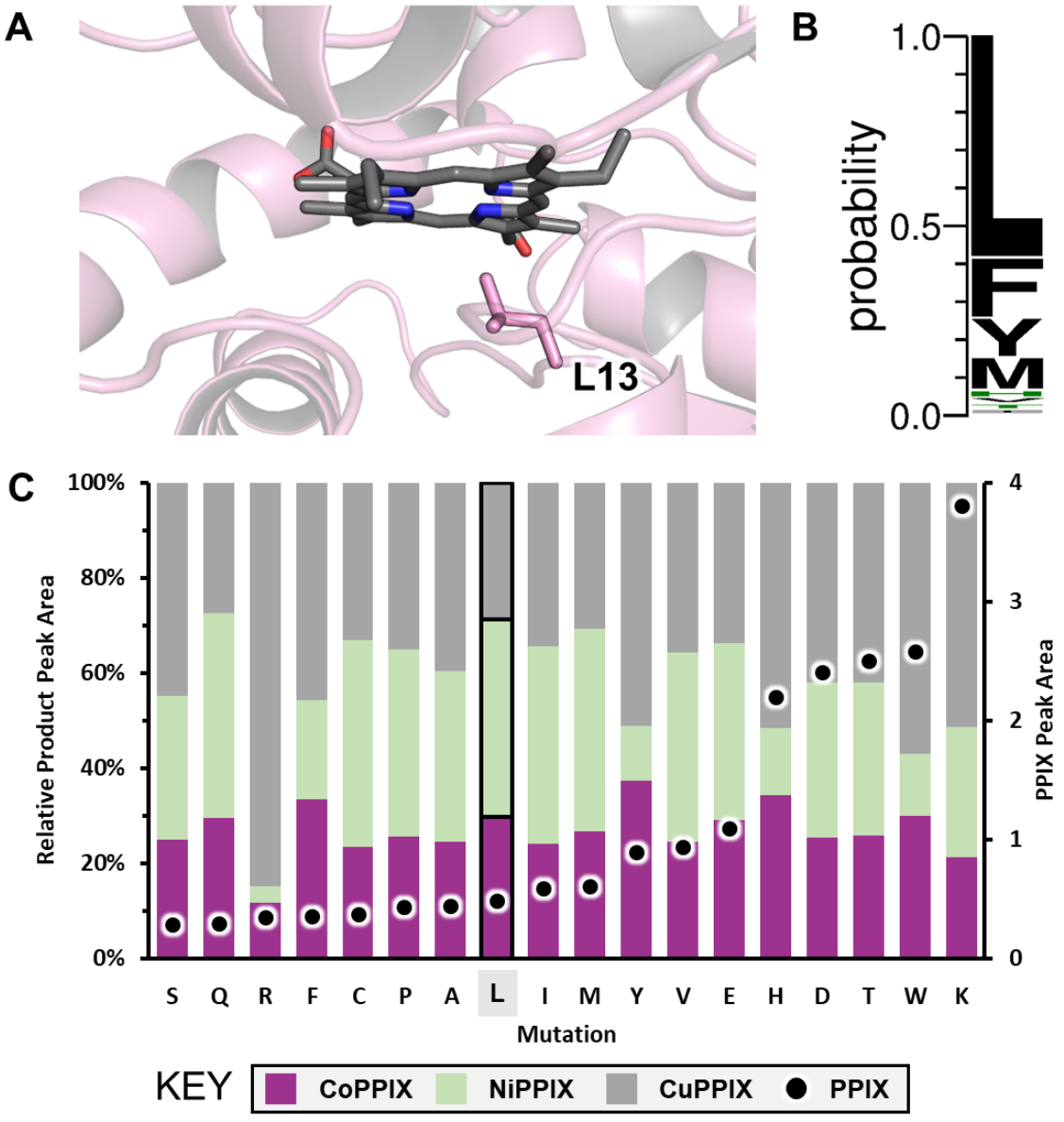


Figure 3.7.. Site saturation mutagenesis library at L13 of *Ec*HemH. A) AlphaFold model of the active site of *Ec*HemH. Protoporphyrin IX (PPIX) is shown in dark grey and is modeled from the PPIX-bound Human ferrochelatase structure (2HRE). B) Logoplot depicting conservation at L13 across homolog ferrochelatases. More information can be found in the supplemental methods section "Construction of HemH sequence alignment and LogoPlots". C) Screening results from metal SUMS of site saturation mutagenesis at L13. Reactions were conducted with cell lysate and 10 µM PPIX, 900 µM $NiCl_2$, 80 µM $CuCl_2$, and 20 µM $CoCl_2$ in 100 mM Tris-HCl, 400 mM NaCl buffer pH 7.0 at room temperature for 1 hour. Reactions were quenched with 3.0 M HCl and extracted into a 2:1 ethyl acetate/pyridine mixture before injection onto UPLC. Colored cars represent the relative peak area (left axis) for CoPPIX (purple), NiPPIX (green), and CuPPIX (grey) as measured at the appropriate λ_{max} and retention time, relative to an internal standard. Black dots represent the peak area of remaining PPIX substrate (right axis), relative to the internal standard. Sequencing information was obtained for each well of the plate, and wells with the same sequence were averaged in this plot. Data from wells with the same residue identity as the parent at position 13 (L13), are highlighted with black border.

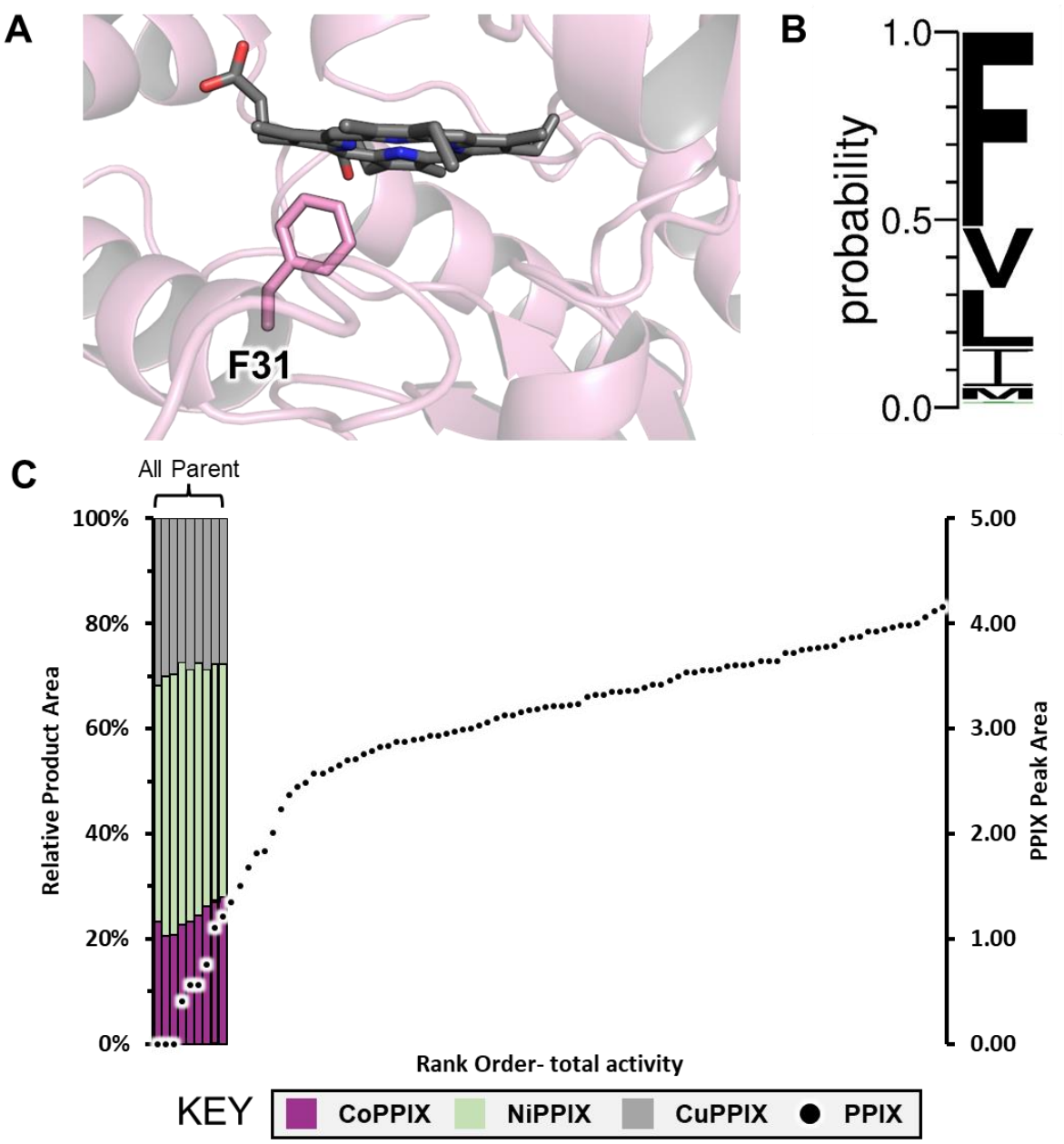


Figure 3.8. Site saturation mutagenesis library at F31 of EcHemH. A) AlphaFold model of the active site of EcHemH. Protoporphyrin IX (PPIX) is shown in dark grey and is modeled from the PPIX-bound Human ferrochelatase structure (2HRE). B) Logoplot depicting conservation at F31 across homolog ferrochelatases. More information can be found in the supplemental methods section "Construction of HemH sequence alignment and LogoPlots". C) Screening results from metal SUMS of site saturation mutagenesis at F31. Reactions were conducted with cell lysate and 10 µM PPIX, 900 µM $NiCl_2$, 80 µM $CuCl_2$, and 20 µM $CoCl_2$ in 100 mM Tris-HCl, 400 mM NaCl buffer pH 7.0 at room temperature for 1 hour. Reactions were quenched with 3.0 M HCl and extracted into a 2:1 ethyl acetate/pyridine mixture before injection onto UPLC. Colored cars represent the relative peak area (left axis) for CoPPIX (purple), NiPPIX (green), and CuPPIX (grey) as measured at the appropriate λ_{max} and retention time, relative to an internal standard. Black dots represent the peak area of remaining PPIX substrate (right axis), relative to the internal standard. Data from wells with the same residue identity as the parent at position 31 (F31), are highlighted with black border.

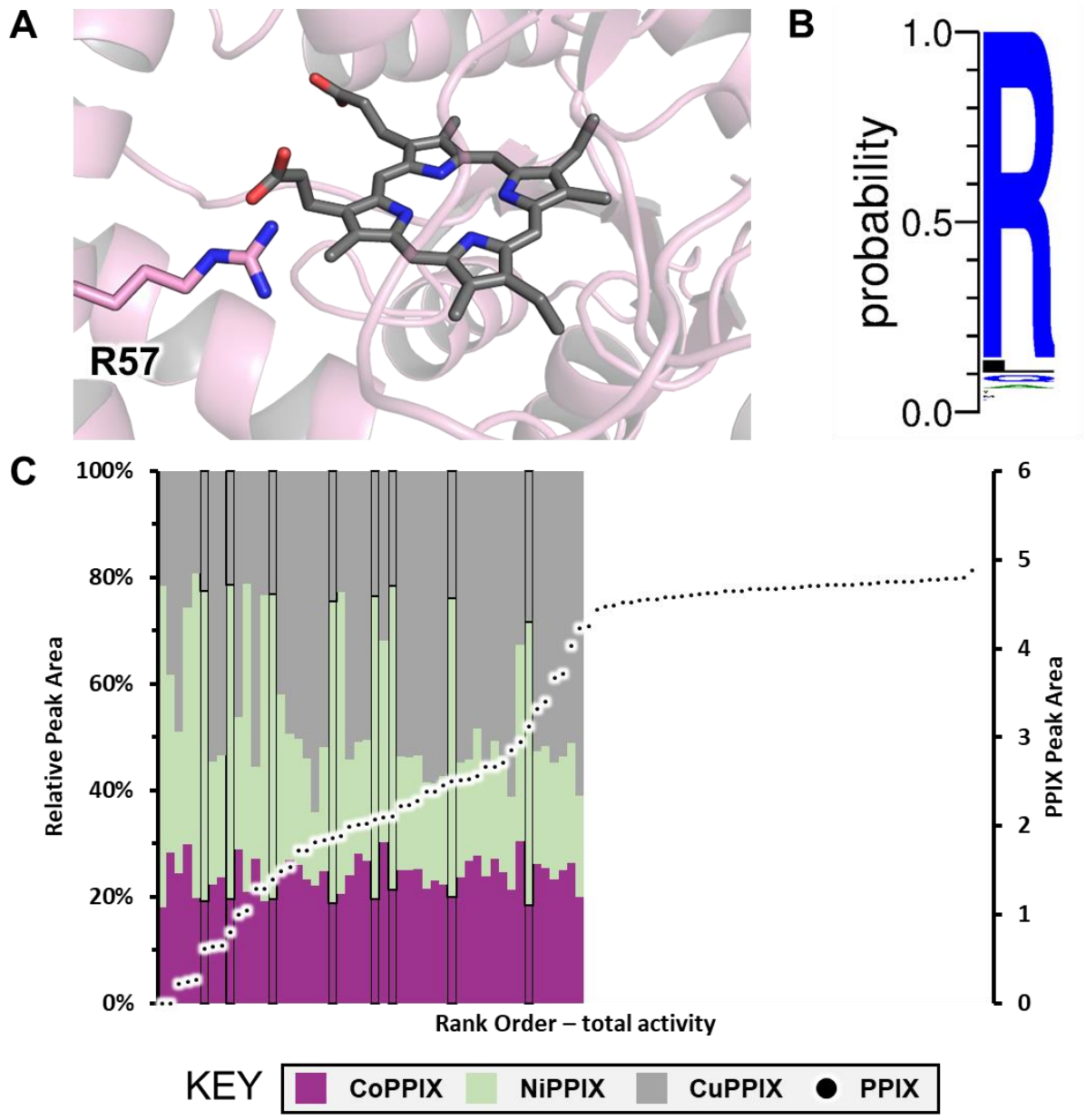


Figure 3.9. Site saturation mutagenesis library at R57 of EcHemH. A) AlphaFold model of the active site of EcHemH. Protoporphyrin IX (PPIX) is shown in dark grey and is modeled from the PPIX-bound Human ferrochelatase structure (2HRE). B) Logoplot depicting conservation at R57 across homolog ferrochelatases. More information can be found in the supplemental methods section "Construction of HemH sequence alignment and LogoPlots". C) Screening results from metal SUMS of site saturation mutagenesis at R57. Reactions were conducted with cell lysate and 10 µM PPIX, 900 µM $NiCl_2$, 80 µM $CuCl_2$, and 20 µM $CoCl_2$ in 100 mM Tris-HCl, 400 mM NaCl buffer pH 7.0 at room temperature for 1 hour. Reactions were quenched with 3.0 M HCl and extracted into a 2:1 ethyl acetate/pyridine mixture before injection onto UPLC. Colored cars represent the relative peak area (left axis) for CoPPIX (purple), NiPPIX (green), and CuPPIX (grey) as measured at the appropriate λ_{max} and retention time, relative to an internal standard. Black dots represent the peak area of remaining PPIX substrate (right axis), relative to the internal standard. Data from wells with the same residue identity as the parent at position 57 (R57), are highlighted with black border.

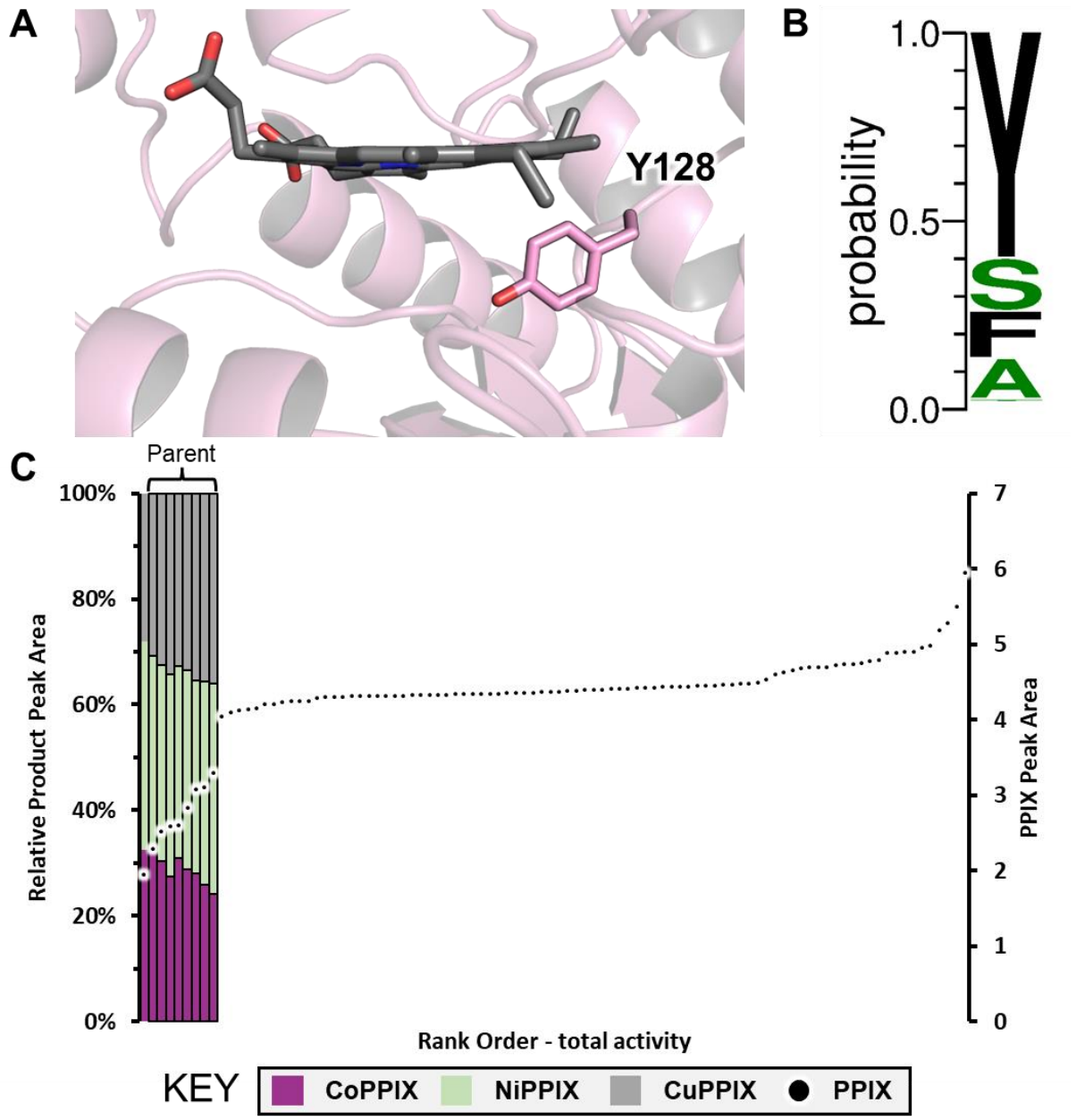


Figure 3.10. Site saturation mutagenesis library at Y128 of EcHemH. A) AlphaFold model of the active site of EcHemH. Protoporphyrin IX (PPIX) is shown in dark grey and is modeled from the PPIX-bound Human ferrochelatase structure (2HRE). B) Logoplot depicting conservation at F31 across homolog ferrochelatases. More information can be found in the supplemental methods section "Construction of HemH sequence alignment and LogoPlots". C) Screening results from metal SUMS of site saturation mutagenesis at Y128. Reactions were conducted with cell lysate and 10 μM PPIX, 900 μM $NiCl_2$, 80 μM $CuCl_2$, and 20 μM $CoCl_2$ in 100 mM Tris-HCl, 400 mM NaCl buffer pH 7.0 at room temperature for 1 hour. Reactions were quenched with 3.0 M HCl and extracted into a 2:1 ethyl acetate/pyridine mixture before injection onto UPLC. Colored cars represent the relative peak area (left axis) for CoPPIX (purple), NiPPIX (green), and CuPPIX (grey) as measured at the appropriate λ_{max} and retention time, relative to an internal standard. Black dots represent the peak area of remaining PPIX substrate (right axis), relative to the internal standard. Data from wells with the same residue identity as the parent at position 128 (Y128), are highlighted with black border.

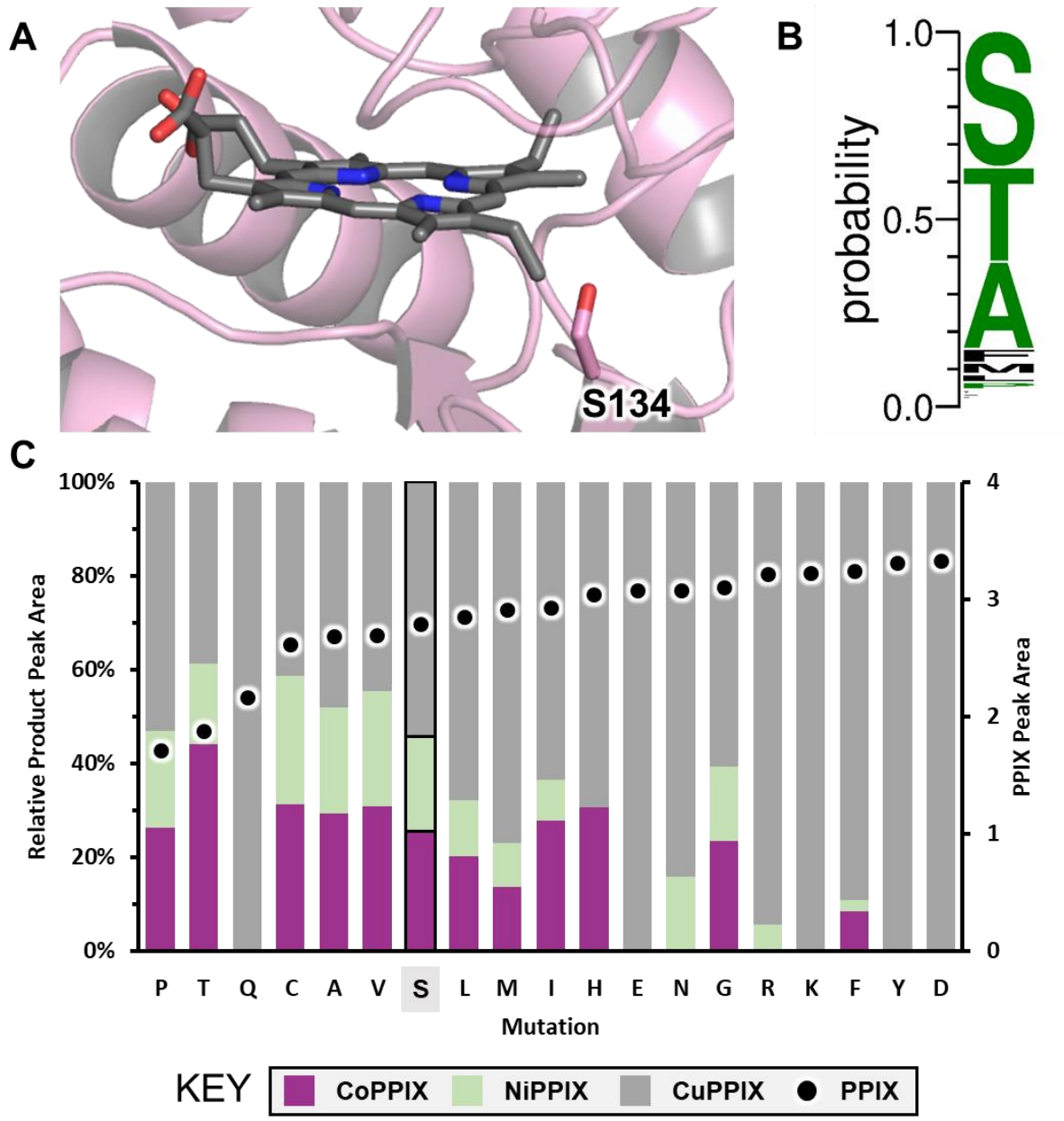


Figure 3.11. Site saturation mutagenesis library at S134 of EcHemH. A) AlphaFold model of the active site of EcHemH. Protoporphyrin IX (PPIX) is shown in dark grey and is modeled from the PPIX-bound Human ferrochelatase structure (2HRE). B) Logoplot depicting conservation at S134 across homolog ferrochelatases. More information can be found in the supplemental methods section "Construction of HemH sequence alignment and LogoPlots". C) Screening results from metal SUMS of site saturation mutagenesis at S134. Reactions were conducted with cell lysate and 10 µM PPIX, 900 µM $NiCl_2$, 80 µM $CuCl_2$, and 20 µM $CoCl_2$ in 100 mM Tris-HCl, 400 mM NaCl buffer pH 7.0 at room temperature for 1 hour. Reactions were quenched with 3.0 M HCl and extracted into a 2:1 ethyl acetate/pyridine mixture before injection onto UPLC. Colored cars represent the relative peak area (left axis) for CoPPIX (purple), NiPPIX (green), and CuPPIX (grey) as measured at the appropriate λ_{max} and retention time, relative to an internal standard. Black dots represent the peak area of remaining PPIX substrate (right axis), relative to the internal standard. Sequencing information was obtained for each well of the plate, and wells with the same sequence were averaged in this plot. Data from wells with the same residue identity as the parent at position 134 (S134), are highlighted with black border.

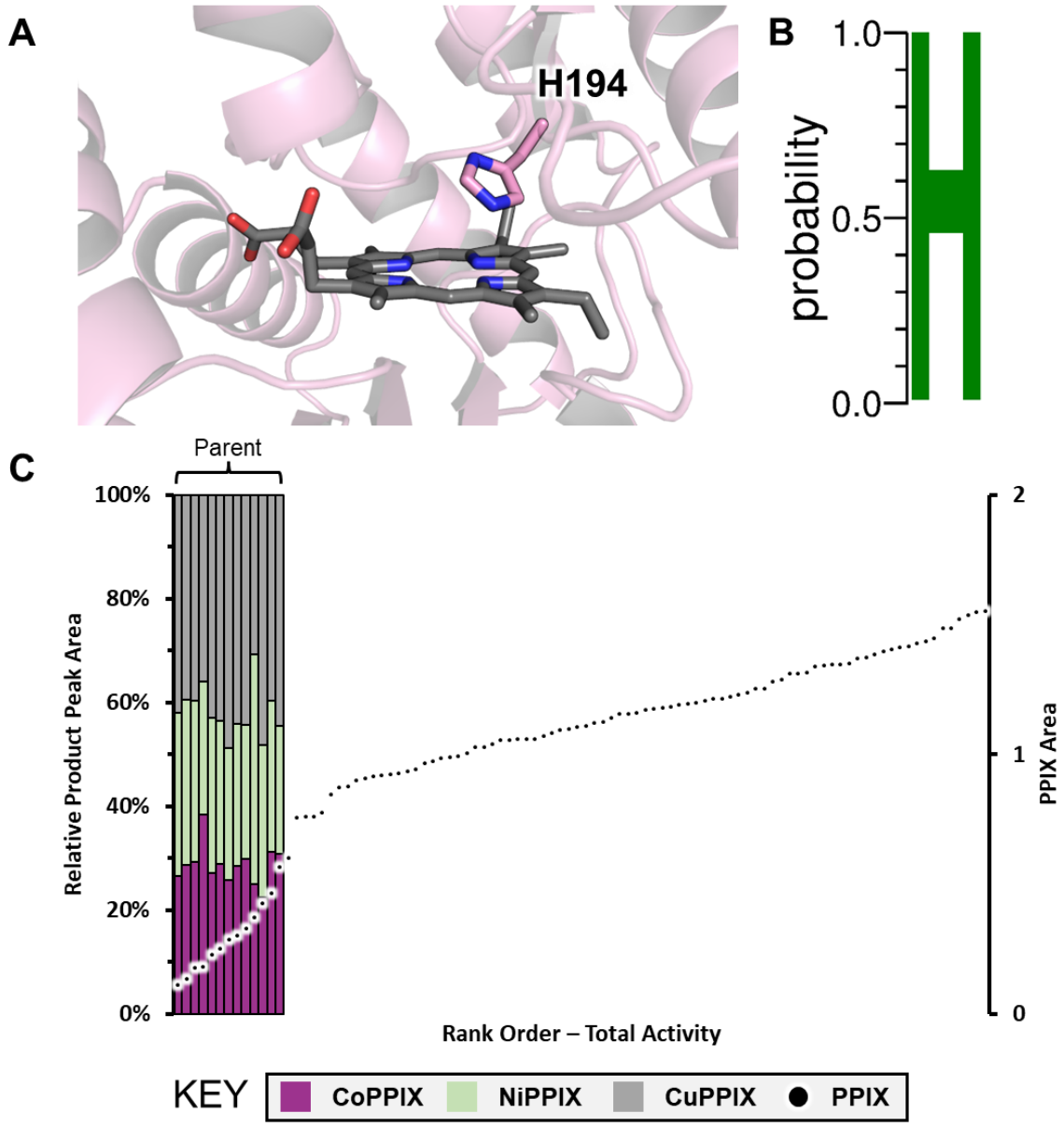


Figure 3.12. Site saturation mutagenesis library at H194 of EcHemH. A) AlphaFold model of the active site of EcHemH. Protoporphyrin IX (PPIX) is shown in dark grey and is modeled from the PPIX-bound Human ferrochelatase structure (2HRE). B) Logoplot depicting conservation at H194 across homolog ferrochelatases. More information can be found in the supplemental methods section "Construction of HemH sequence alignment and LogoPlots". C) Screening results from metal SUMS of site saturation mutagenesis at H194. Reactions were conducted with cell lysate and 10 µM PPIX, 900 µM $NiCl_2$, 80 µM $CuCl_2$, and 20 µM $CoCl_2$ in 100 mM Tris-HCl, 400 mM NaCl buffer pH 7.0 at room temperature for 1 hour. Reactions were quenched with 3.0 M HCl and extracted into a 2:1 ethyl acetate/pyridine mixture before injection onto UPLC. Colored cars represent the relative peak area (left axis) for CoPPIX (purple), NiPPIX (green), and CuPPIX (grey) as measured at the appropriate λ_{max} and retention time, relative to an internal standard. Black dots represent the peak area of remaining PPIX substrate (right axis), relative to the internal standard. Data from wells with the same residue identity as the parent at position 194 (H194), are highlighted with black border.

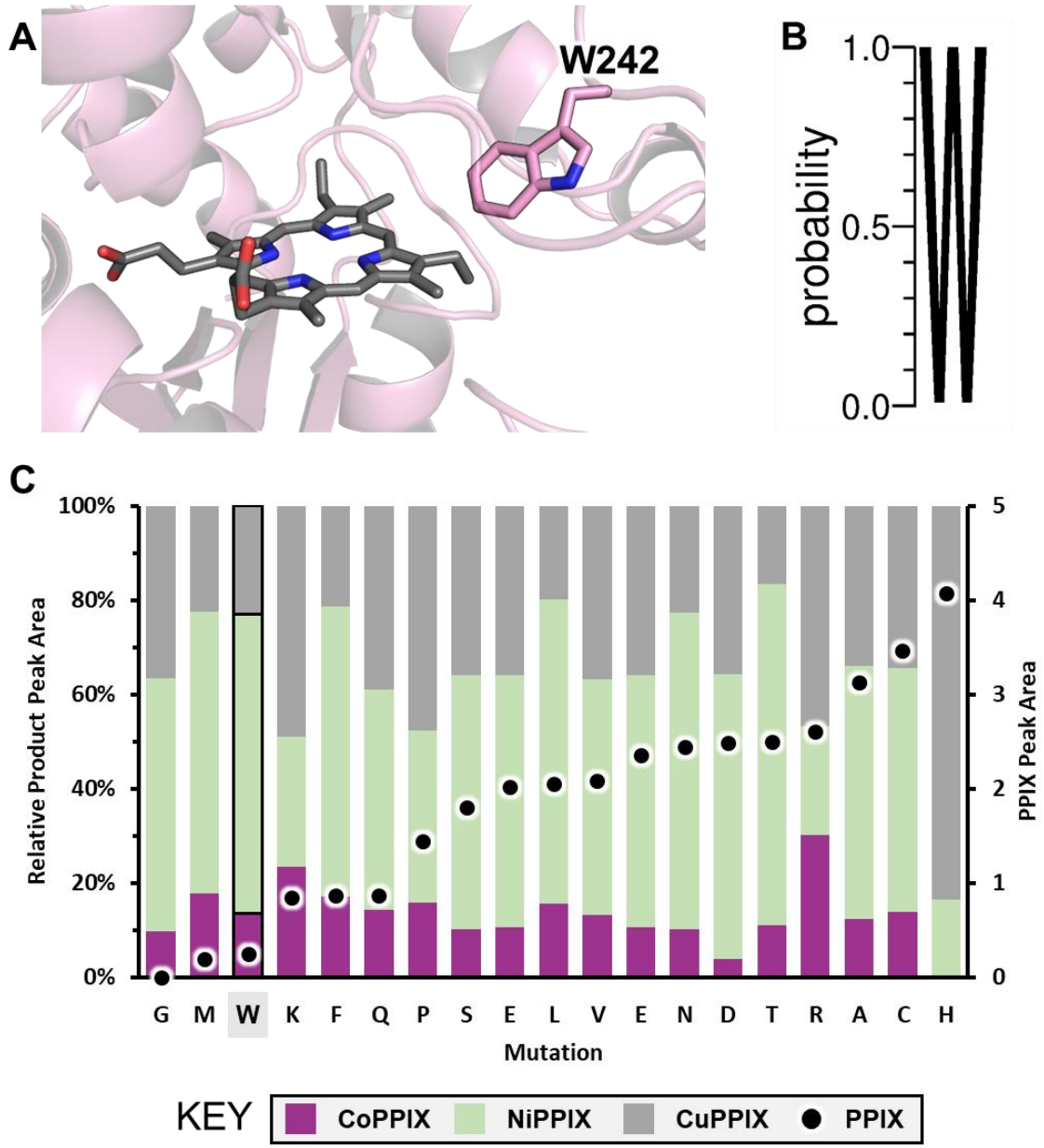


Figure 3.13. Site saturation mutagenesis library at W242 of *Ec*HemH. A) AlphaFold model of the active site of *Ec*HemH. Protoporphyrin IX (PPIX) is shown in dark grey and is modeled from the PPIX-bound Human ferrochelatase structure (2HRE). B) Logoplot depicting conservation at W242 across homolog ferrochelatases. More information can be found in the supplemental methods section "Construction of HemH sequence alignment and LogoPlots". C) Screening results from metal SUMS of site saturation mutagenesis at W242. Reactions were conducted with cell lysate and 10 µM PPIX, 900 µM $NiCl_2$, 80 µM $CuCl_2$, and 20 µM $CoCl_2$ in 100 mM Tris-HCl, 400 mM NaCl buffer pH 7.0 at room temperature for 1 hour. Reactions were quenched with 3.0 M HCl and extracted into a 2:1 ethyl acetate/pyridine mixture before injection onto UPLC. Colored cars represent the relative peak area (left axis) for CoPPIX (purple), NiPPIX (green), and CuPPIX (grey) as measured at the appropriate λ_{max} and retention time, relative to an internal standard. Black dots represent the peak area of remaining PPIX substrate (right axis), relative to the internal standard. Sequencing information was obtained for each well of the plate, and wells with the same sequence were averaged in this plot. Data from wells with the same residue identity as the parent at position 242 (W242), are highlighted with black border.

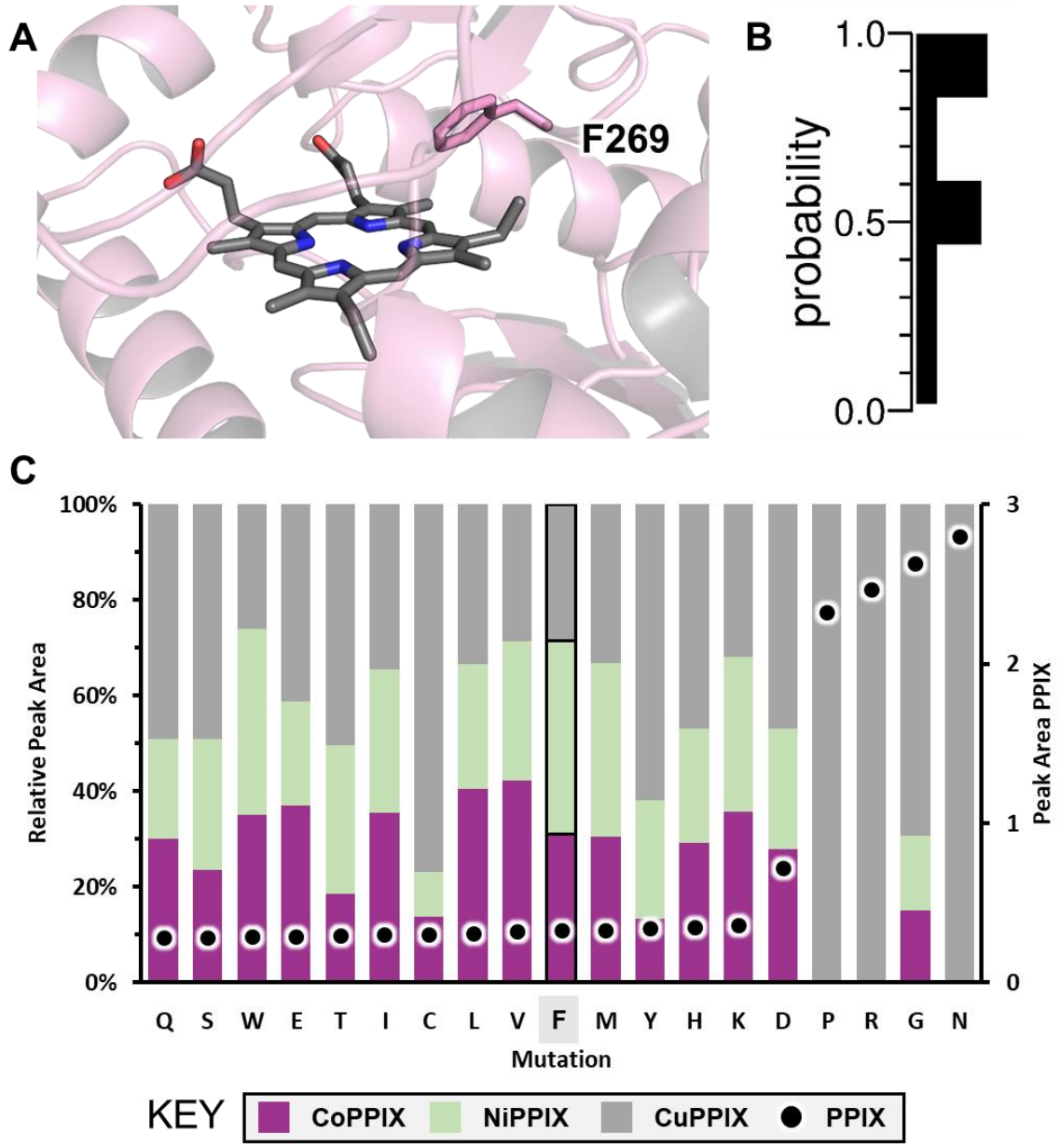


Figure 3.14. Site saturation mutagenesis library at F269 of EcHemH. A) AlphaFold model of the active site of EcHemH. Protoporphyrin IX (PPIX) is shown in dark grey and is modeled from the PPIX-bound Human ferrochelatase structure (2HRE). B) Logoplot depicting conservation at F269 across homolog ferrochelatases. More information can be found in the supplemental methods section "Construction of HemH sequence alignment and LogoPlots". C) Screening results from metal SUMS of site saturation mutagenesis at F269. Reactions were conducted with cell lysate and 10 µM PPIX, 900 µM $NiCl_2$, 80 µM $CuCl_2$, and 20 µM $CoCl_2$ in 100 mM Tris-HCl, 400 mM NaCl buffer pH 7.0 at room temperature for 1 hour. Reactions were quenched with 3.0 M HCl and extracted into a 2:1 ethyl acetate/pyridine mixture before injection onto UPLC. Colored cars represent the relative peak area (left axis) for CoPPIX (purple), NiPPIX (green), and CuPPIX (grey) as measured at the appropriate λ_{max} and retention time, relative to an internal standard. Black dots represent the peak area of remaining PPIX substrate (right axis), relative to the internal standard. Sequencing information was obtained for each well of the plate, and wells with the same sequence were averaged in this plot. Data from wells with the same residue identity as the parent at position 269 (F269), are highlighted with black border.

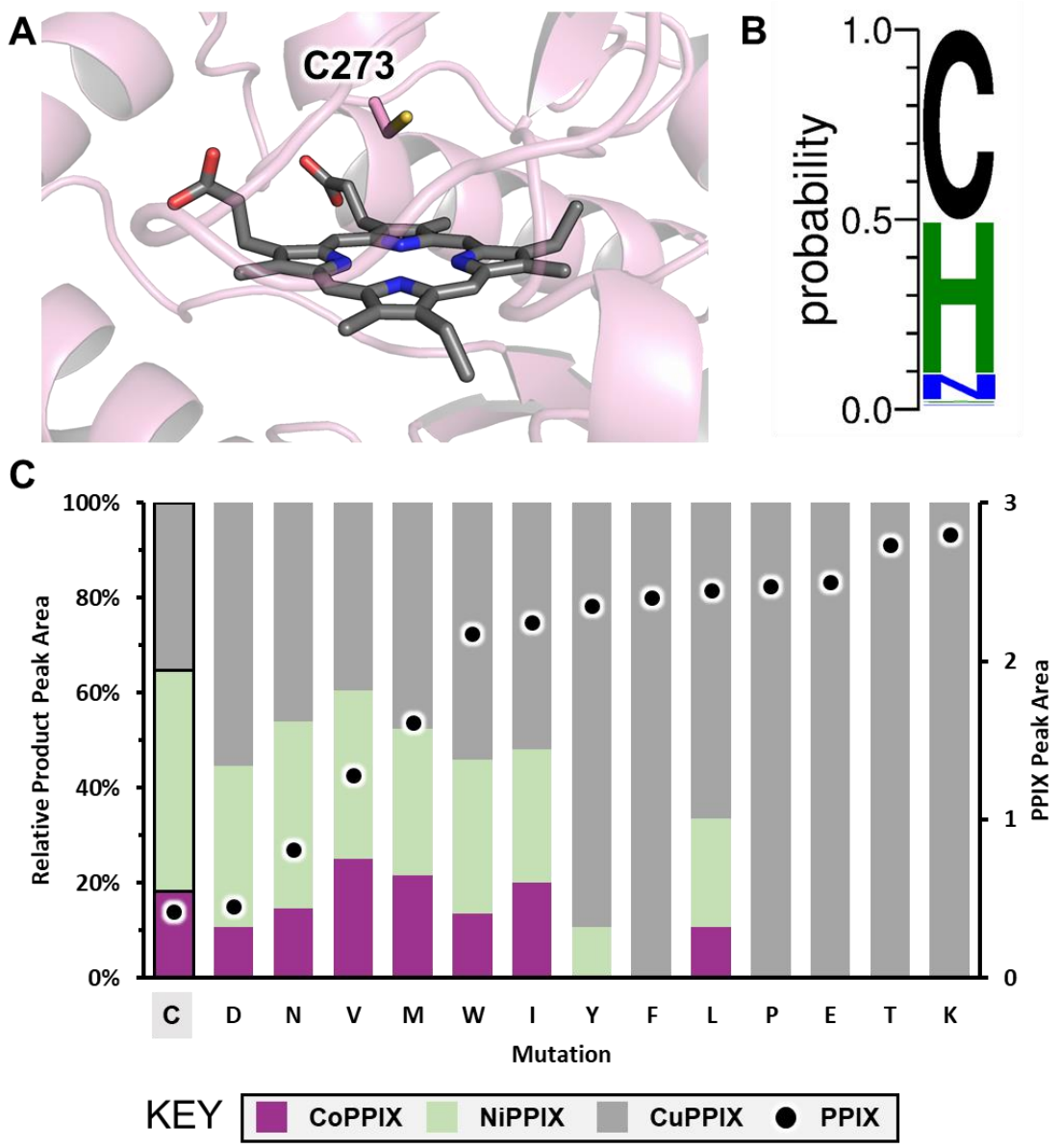


Figure 3.15. Site saturation mutagenesis library at C273 of EcHemH. A) AlphaFold model of the active site of EcHemH. Protoporphyrin IX (PPIX) is shown in dark grey and is modeled from the PPIX-bound Human ferrochelatase structure (2HRE). B) Logoplot depicting conservation at C273 across homolog ferrochelatases. More information can be found in the supplemental methods section "Construction of HemH sequence alignment and LogoPlots". C) Screening results from metal SUMS of site saturation mutagenesis at C273. Reactions were conducted with cell lysate and 10 µM PPIX, 900 µM $NiCl_2$, 80 µM $CuCl_2$, and 20 µM $CoCl_2$ in 100 mM Tris-HCl, 400 mM NaCl buffer pH 7.0 at room temperature for 1 hour. Reactions were quenched with 3.0 M HCl and extracted into a 2:1 ethyl acetate/pyridine mixture before injection onto UPLC. Colored cars represent the relative peak area (left axis) for CoPPIX (purple), NiPPIX (green), and CuPPIX (grey) as measured at the appropriate λ_{max} and retention time, relative to an internal standard. Black dots represent the peak area of remaining PPIX substrate (right axis), relative to the internal standard. Sequencing information was obtained for each well of the plate, and wells with the same sequence were averaged in this plot. Data from wells with the same residue identity as the parent at position 273 (C273), are highlighted with black border.

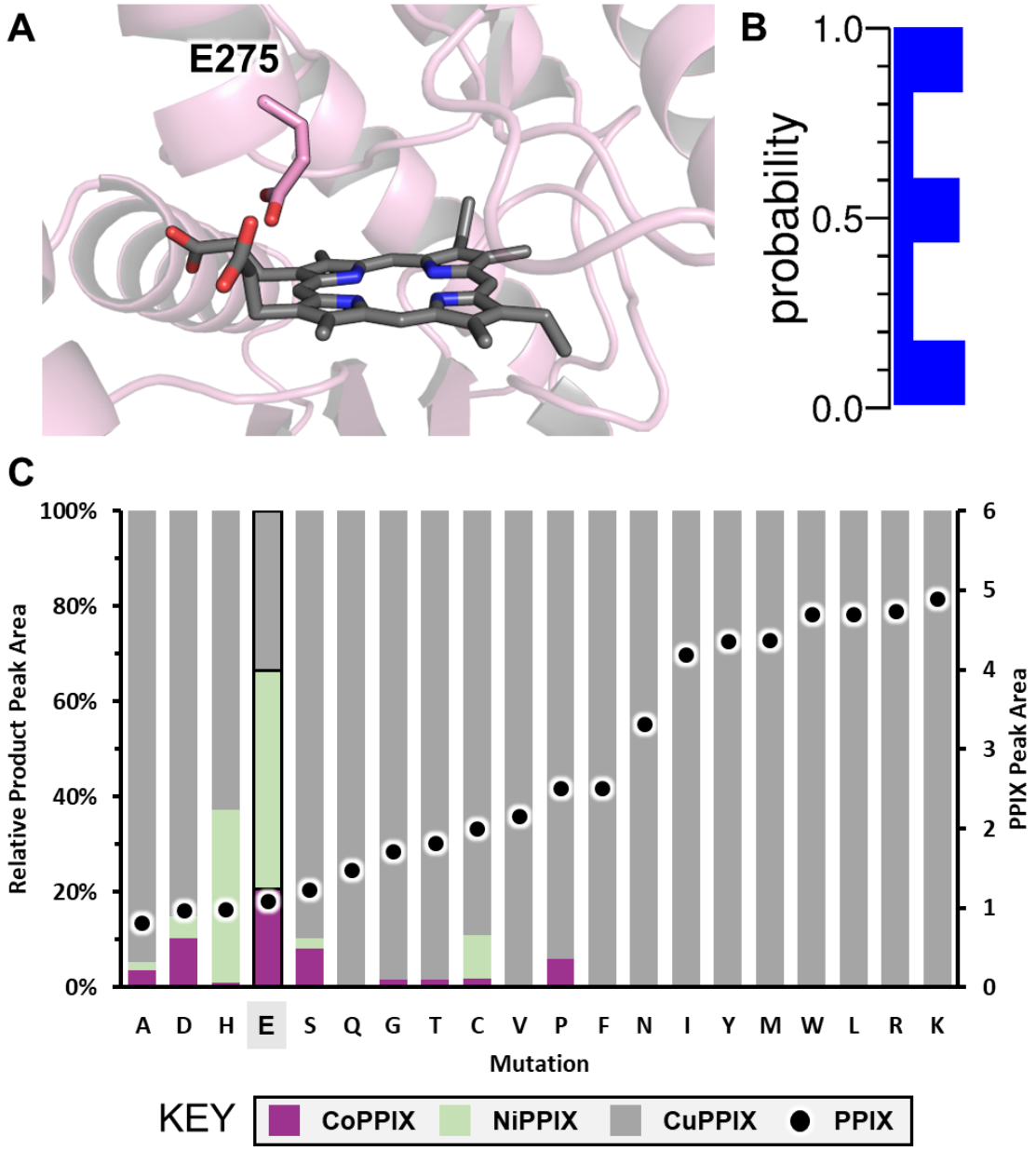


Figure 3.16. Site saturation mutagenesis library at E275 of *Ec*HemH. A) AlphaFold model of the active site of *Ec*HemH. Protoporphyrin IX (PPIX) is shown in dark grey and is modeled from the PPIX-bound Human ferrochelatase structure (2HRE). B) Logoplot depicting conservation at E275 across homolog ferrochelatases. More information can be found in the supplemental methods section "Construction of HemH sequence alignment and LogoPlots". C) Screening results from metal SUMS of site saturation mutagenesis at E275. Reactions were conducted with cell lysate and 10 µM PPIX, 900 µM $NiCl_2$, 80 µM $CuCl_2$, and 20 µM $CoCl_2$ in 100 mM Tris-HCl, 400 mM NaCl buffer pH 7.0 at room temperature for 1 hour. Reactions were quenched with 3.0 M HCl and extracted into a 2:1 ethyl acetate/pyridine mixture before injection onto UPLC. Colored cars represent the relative peak area (left axis) for CoPPIX (purple), NiPPIX (green), and CuPPIX (grey) as measured at the appropriate λ_{max} and retention time, relative to an internal standard. Black dots represent the peak area of remaining PPIX substrate (right axis), relative to the internal standard. Sequencing information was obtained for each well of the plate, and wells with the same sequence were averaged in this plot. Data from wells with the same residue identity as the parent at position 275 (E275), are highlighted with black border.

None of the variants we generated have been studied in *Ec*HemH previously, but many corresponding variants have been studied in HemH homologs, such as the murine, yeast, and human FECH enzymes. Given the indirect nature of the SUMS screen, comparing these screening data to published studies of corresponding variants in the FECH proteins provided an important means to benchmark the results. For example, the *Ec*HemH H194X library yielded no active variants (Figure 3.14). Studies of corresponding variants of FECH (H263C and H263A human FECH, H235L yeast FECH) indicated these variants were inactive or had only trace activity, demonstrating that our multiplexed screen results are consistent with literature precedent.[19,44,45] Indeed, H194 is nearly 100% conserved among HemH/FECH homologs (Figure 3.14B).

To directly compare specificity shifts among variants across different libraries, we re-screened a subset of variants showing the largest shifts in specificity in the initial screen. Specificity shifts were generally reproducible, but several variants had markedly different specificities in this follow up assessment when compared to the initial screens. We attribute these differences to small changes in pH and buffer concentrations between the two rounds of screening. From the re-screen data, we identified several variants with significantly altered metal specificity compared to the parent *Ec*HemH (Figure 4B).

3. 3. 4. Effect of EcHemH variant co-expression on heme and CoPPIX production in E. coli

We next sought to investigate how variants with *in vitro* changes in metal specificity altered metalloporphyrin formation *in vivo.* We wanted to monitor changes in activity with the native Fe^{2+} substrate, and to this end, we devised an *in vivo* screen to assess chelatase selectivity, comparing Fe^{2+} versus Co^{2+} insertion into PPIX. Cells were grown in rich media supplemented with a 4:1 ratio of exogenous Co^{2+} to Fe^{2+}, such that the bioavailable metal pool was consistent and well-defined. To circumvent uncertainties caused by heme b catabolism, we co-expressed *Ec*HemH

with the hemoprotein, dye-decolorizing peroxidase (DyP), which binds free PPIX, CoPPIX, and heme b promiscuously and tightly. Following expression, DyP was purified by nickel-affinity chromatography and the relative Co^{2+} and Fe^{2+} content of the protein samples was measured using inductively coupled plasma mass spectrometry (ICP-MS). A mixture of Co^{2+} and Fe^{2+} was observed in each DyP sample, enabling comparison of how amino acid identity in *Ec*HemH impacted cofactor distribution (Figure 3.17). We identified several variants that incorporated more CoPPIX relative to the parent *Ec*HemH, including L13R, L13H, R57Q, and E275D.

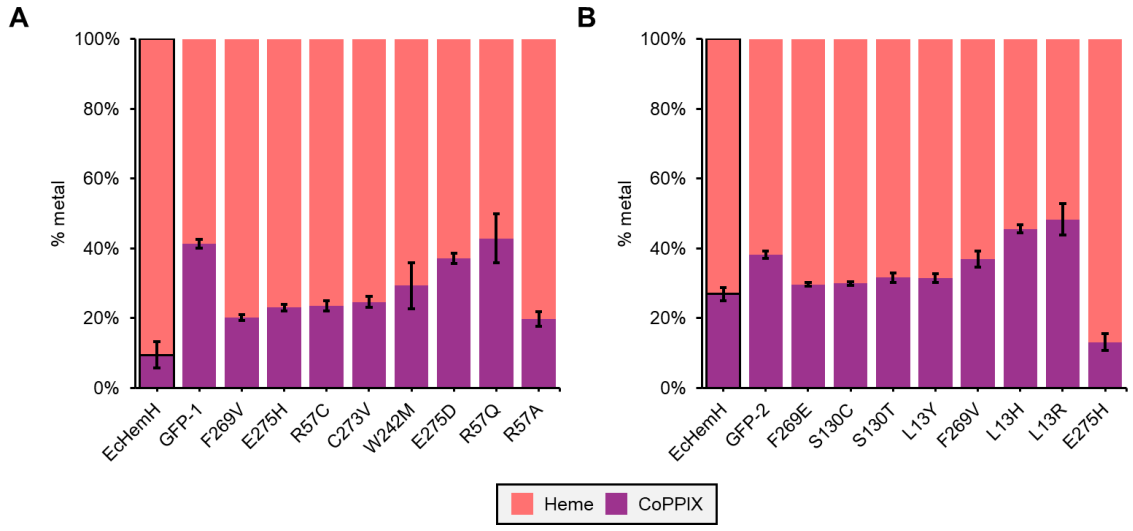


Figure 3.17. In vivo screen of CoPPIX and heme production by *Ec*HemH and *Ec*HemH variants as assessed by ICP-MS analysis of acid-digested DyP. Panels A and B reflect two independent expression experiments with several selected *Ec*HemH variants. The black bars represent the standard deviation of three independent ICP-MS measurements of a single sample of acid-digested DyP. Further experimental details can be found in the methods section "In vivo screen of Co^{2+} and Fe^{2+} activity of *Ec*HemH variants".

To validate that differences metal in incorporation were due to improved Co^{2+} versus Fe^{2+} selectivity, we expressed and purified one such variant, L13R, and assessed its selectivity *in vitro* by measuring the incorporation of Co^{2+} and Fe^{2+} into PPIX under direct competition. Excess ascorbate was added to the reaction solution to maintain the ferrous oxidation state of the iron substrate, and after 20 minutes, the porphyrin content of each reaction was extracted and analyzed by UPLC. For the parent *Ec*HemH enzyme, an apparent 3:1 ratio of MPPIX species from this direct competition experiment was in reasonable agreement with the 2:1 ratio inferred from the ratio of k_{cat}/K_M for the enzyme operating with each substrate (Table 3.1). Crucially, this

direct selectivity measurement showed that the L13R had increased activity with Co^{2+}, indicating a ~30-fold shift in selectivity relative to parent under these conditions (Figure 3.18).

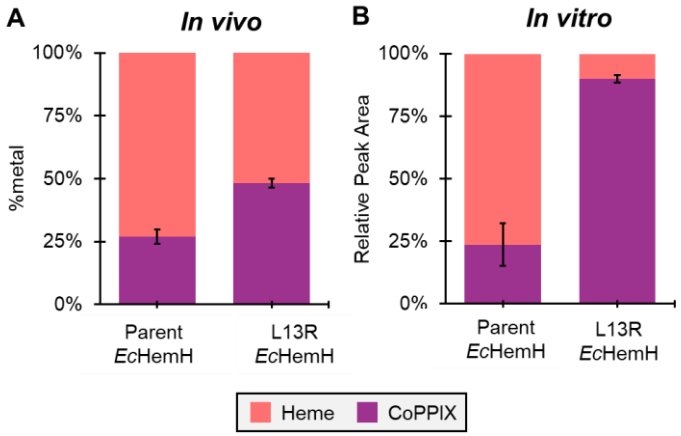


Figure 3.18. Competition experiments for Co^{2+} and Fe^{2+} incorporation into protoporphyrin IX (PPIX) by parent EcHemH and L13R EcHemH variant. A) Distribution of MPPIX products obtained from co-expression of parent and L13R EcHemH with dye-decolorizing peroxidase (DyP) with a defined mixture of 50 μM Fe^{2+} and 100 μM Co^{2+} added to the growth media. DyP was purified by Ni-affinity chromatography and digested in nitric acid prior to analysis by ICP-MS. Data are average of triplicate ICP-MS measurements. B) In vitro measurement of Co^{2+} versus Fe^{2+} incorporation into PPIX by purified parent and L13R EcHemH enzymes. Reactions were conducted in 100 mM Tris-HCl buffer with 400 mM NaCl and 0.2% Tween 80. The concentration of enzyme added was 250 nM and the concentration of PPIX was 5.0 μM. Reactions were initiated with the addition of 50 μM each of Fe^{2+} and Co^{2+} in 0.5 mM ascorbate. After 20 minutes, reactions were quenched with acid and extracted into organic solvent. The relative concentrations of porphyrin products were analyzed by UPLC using the absorbance at 400 nm (FePPIX and PPIX) and 423 nm (CoPPIX). Error bars represent the standard deviation of quadruplicate UPLC measurements.

In an effort to control for the effects of the native *Ec*HemH being overexpressed in *E. coli* BL21, we also expressed DyP without an exogenous copy of *Ec*HemH. Instead, cells were grown under an analogous double-antibiotic pressure with a plasmid expressing super folder green fluorescent protein (sfGFP). As this experiment was intended to be a negative control, we were surprised to observe that the CoPPIX content of DyP co-expressed with GFP was as good or higher than that of the *Ec*HemH co-expressed sample (Figure 3.17). We hypothesize that the differences in metalloporphyrin content from overexpression of *Ec*HemH may be due to differences in other aspects of heme metabolism, which is typically under tight regulation in *E. coli*.[46–48]

We next asked whether co-expression of the cobalt selective *Ec*HemH variant L13R could enable useful levels of CoPPIX versus heme incorporation (> 90% cofactor purity) in rich media. We therefore tested co-expression of parent *Ec*HemH versus the L13R variant with DyP at

substantially higher concentrations of cobalt (500 µM) in rich media without added iron. We compared these expressions to the expression of DyP in the absence of exogenous HemH expression. To our surprise, the resulting DyP hemoprotein samples were greater than 95% CoPPIX-loaded in all cases, even without co-expression of a ferrochelatase (Figure 3.19A). Therefore, while the L13R variant did succeed in yielding a robust method for producing CoPPIX, it was not substantially different than the effects of adding cobalt alone. These results refuted our initial hypothesis that alterations to chelatase specificity would lead to more efficient CoPPIX production in vivo, as the promiscuity of the native system is sufficient. Instead, we re-investigated other factors related to CoPPIX production.

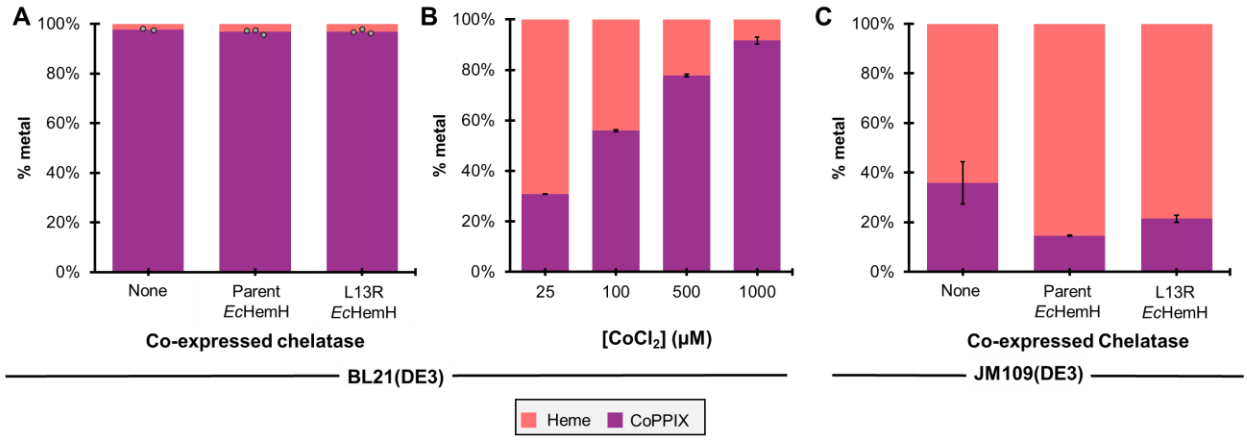


Figure 3.19. Effect of HemH expression, cobalt concentration, and E. coli expression strain on the distribution of CoPPIX- and heme-loaded DyP A) ICPMS measurements of DyP expressed in BL21(DE3) with parent and L13R EcHemH co-expression, and in the absence of co-expression. 1 mM $CoCl_2$ and 250 µM δ-aminolaevulinic acid was added at the time of induction. Dots represent average triplicate ICP-MS measurements from three separate expressions. B) ICPMS measurements of DyP expressed in BL21(DE3) with various amounts of cobalt. Cobalt was added as an aqueous solution of $CoCl_2$ as expression cultures when OD_{600} reached 0.2-0.3, prior to induction. Error bars represent standard deviation of triplicate ICP-MS measurements. C) ICPMS measurements of DyP expressed in JM109(DE3) with parent and L13R EcHemH co-expression, and in the absence of chelatase co-expression. Expression conditions were identical to that of Figure 6A. Error bars represent standard deviation of triplicate ICP-MS measurements.

3. 3. 5. Effect of cobalt concentration, E. coli strain, and metal identity on MPPIX production

To begin, we studied the effect of Co concentration in TB media on the metalloporphyrin content of expressed hemoprotein. We were delighted to see that addition of 1 mM of Co^{2+} yielded

>90% cobalt incorporation in rich media (Figure 3.19B). While the wet cell mass derived from these expressions decreased as more cobalt was added (Table 3.2), the holoprotein titers per liter cell culture were only slightly diminished as cobalt concentration was increased. Based on our earlier observations in minimal media,[13] we hypothesize that the cytotoxic effect of high intracellular cobalt concentrations may be mitigated by the expression of a hemoprotein, which serves as a sink for excess, toxic Co^{2+} and CoPPIX.

Table 3.2. Titers from expression of DyP in BL21(DE3) with addition of $CoCl_2$ at inoculation, early log phase, and at induction.

Condition	Holoprotein Titer* (mg/L culture)	% Co[§] (versus Fe)	Wet cell mass (g)	Final OD_{600}
1 mM $CoCl_2$	38	92 ± 1.4	4.1	6.4
500 µM $CoCl_2$	33	78 ± 0.49	5.4	8.8
100 µM $CoCl_2$	55	56 ± 0.35	5.2	9.6
25 µM $CoCl_2$	80	31 ± 0.10	6.1	11.2

* Measured by pyridine hemochromagen assay, accounting for the ratio of Co versus Fe as determined by ICP-MS. Error represents triplicate ICP-MS measurements.
[§]Measured by ICP-MS

We next tested whether CoPPIX production ability was unique to BL21(DE3). We carried out expression experiments with JM109(DE3), a K12-derived strain whose metal homeostasis is well-studied.[49] Overall, CoPPIX production was markedly less efficient in this strain, yielding less than 40% Co incorporation under the same expression conditions (Figure 3.20C). Overexpression of *Ec*HemH made CoPPIX versus heme production worse still, suggesting that the dynamics of porphyrin metabolism may be affected by HemH overexpression. Satisfyingly, L13R HemH co-expression did yield improved CoPPIX relative to the parent *Ec*HemH, although heme b was still the major cofactor bound.

The surprisingly efficient incorporation of cobalt into PPIX by BL21(DE3) described here inspired us to test the same method for the incorporation of other metals: copper, nickel, and zinc. Analogous to expression with cobalt, 1 mM of $CuCl_2$, $NiCl_2$, and $ZnCl_2$ were added to 1 L of culture expressing DyP in TB media at the time of induction. The metalloporphyrin content of resulting

purified DyP was interrogated by UPLC (Figure 3.20B). Zinc- and copper-supplemented cultures appeared to contain exclusively the native heme cofactor and unmetallated PPIX. The DyP resulting from expression with added nickel contained mostly heme and unmetallated PPIX, but UPLC analysis indicated that a small amount of NiPPIX was also incorporated. While we chose not to pursue the production of NiPPIX substituted hemoproteins in this study, we hypothesize that the methods described here could be optimized for this purpose.

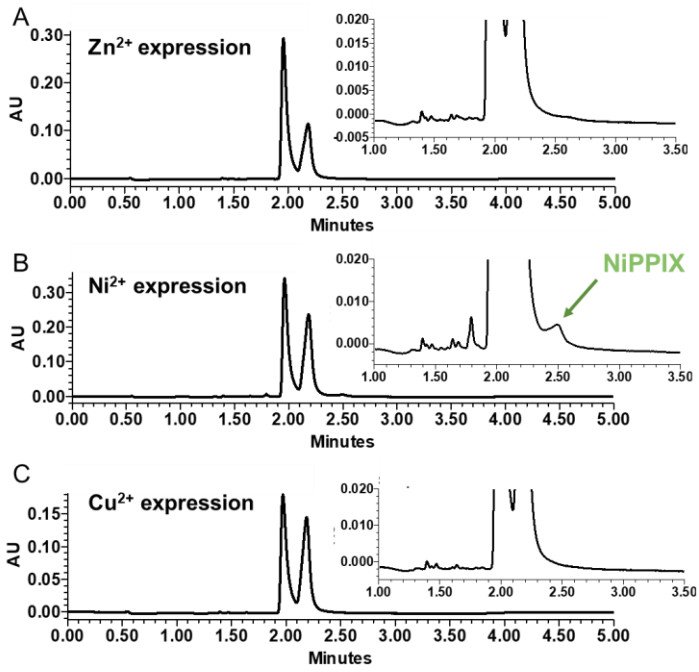


Figure 3.20. UPLC analysis of DyP porphyrin extracts resulting from expression of DyP in the presence of (A) 1 mM $ZnCl_2$, (B) 1 mM $NiCl_2$, and (C) 1 mM $CuCl_2$. Black traces represent 400 nm chromatographic trace. FePPIX has a retention time of 2.0 min (peak 1) and PPIX has a retention time of 2.2 min (peak 2). NiPPIX is indicated by a small peak with a retention time of 2.5 min, and a λ_{max} of 395 nm.

3. 3. 6. Validation of a method for cobalt-substituted hemoproteins in rich media

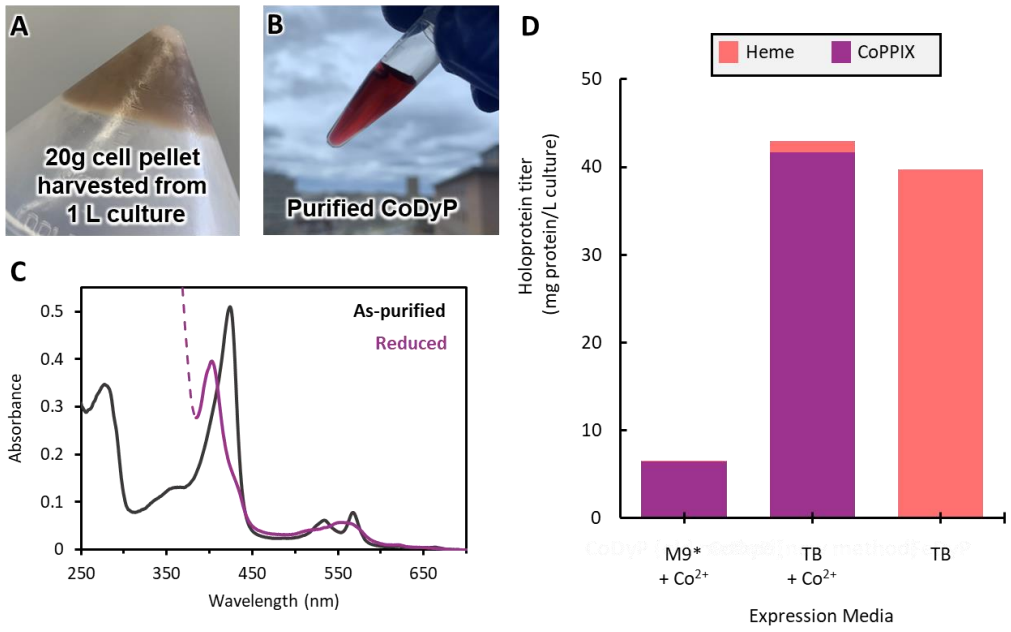


Figure 3.21. Heterologous expression and purification of cobalt substituted DyP from E. lignoliticus in rich media. A) Cell pellet derived from a 1 L culture expression of DyP supplemented with 1 mM $CoCl_2$. B) Sample of purified CoDyP. C) UV-visible spectra of as-purified CoDyP and after reduction with sodium dithionite (magenta). D) Expression titers for Co and native DyP in different growth media.

With this new perspective on the determinants of CoPPIX production, we set out to generate a robust method for production of cobalt substituted hemoprotein in rich media. As chelatase overexpression and selectivity has a surprisingly small and sometimes deleterious effects on CoPPIX production, we chose to forgo co-expresssion of a chelatase and use BL21(DE3) as the expression host. We tested the effect of adding 1 mM $CoCl_2$ at different times throughout the expression protocol, and found that there were only minor differences in yield and cofactor purity (Table 3.3). Adding metal at the time of induction is convenient and effective. With this method we were able to obtain CoDyP at titer of 43 mg protein per L cell culture, and the resulting proteins were >95% cobalt loaded relative to iron. This yield represents a nearly 10-fold improvement in protein titer over our previous method in minimal media,[13] gives similar titers to expression of the native heme-loaded protein (Figure 3.21).

Table 3.3. Titers from expression of DyP in BL21(DE3) with addition of $CoCl_2$ at inoculation, early log phase, and at induction.

$CoCl_2$ addition timing	**Holoprotein Titer* (mg/L culture)**	**% Co[§] (versus Fe)**	**Wet cell mass (g)**	**Final OD_{600}**	**Time to Induction**
At inoculation	30	93 ± 0.45	3.48	6.1	2.5 h
At OD = 0.2	24	81 ± 2.3	3.45	6.1	2.5 h
At induction	43	97 ± 0.06	3.85	8.2	2 h

* Measured by pyridine hemochromagen assay, accounting for the ratio of Co versus Fe as determined by ICP-MS. Error represents triplicate ICP-MS measurements.
[§]Measured by ICP-MS

Lastly, we tested the efficiency of cobalt-substituted hemoprotein production with a different protein scaffold, the P450 enzyme CYP119 from *S. acidcaldarous* under pET22b (IPTG-inducible) and pBAD (arabinose-inducible) expression systems. From 1 liter of cell culture in rich media we were able to obtain 2 mg and 43 mg of cobalt-substituted hemoprotein from the Pet22b and pBAD systems, respectively, albeit with 5% heme contamination. This expression efficiency compares favorably with that of the analogous heme proteins, which, in our hands, can be obtained at 3 mg and 40 mg per liter culture. Furthermore, this method requires virtually no additional steps or genetic manipulations relative to that of the heme protein, aside from addition of an inexpensive cobalt salt ($CoCl_2$) at the time of induction.

Chapter 6 3. 4. Discussion

The production of artificial metalloenzymes, and of substituted hemoproteins in particular, is of interest to biocatalysis and spectroscopy communities. Metal substitution of hemoproteins is a relatively common practice, but methods for substitution rely on supplementing either cell culture or purified protein with exogenously synthesized metalloporphyrin. Developing fully biosynthetic approaches to producing these artificial metalloenzymes lowers barriers to studying these proteins and their development as biocatalysts. While we have focused on the production and incorporation of one metalloporphyrin, CoPPIX, other metal substitutions, such as nickel, zinc, and copper, may also be worthwhile. Our findings in this study may guide the future development of such systems.

Our initial hypothesis was that a low-level of promiscuity in the chelatase was the main enabler of adventitious CoPPIX production in BL21(DE3), warranting thorough kinetic characterization of *Ec*HemH. We found that that this enzyme exhibits a catalytic efficiency with Co^{2+} that is only 2-fold lower than that of the native metal substrate, Fe^{2+}, indicating that *Ec*HemH is an effective cobalt chelatase. Furthermore, in-depth kinetic characterization of *Ec*HemH with multiple metal substrates, along with modeling and site saturation mutagenesis of key active site residues, has provided new mechanistic understanding for this ubiquitous class of enzymes. Kinetic data support the presence of an unproductive metal binding mode, to which inhibitory metal binding is exacerbated by high concentrations of PPIX. One molecular interpretation consistent with this phenomenon is metal ion binding at two different sites to the protein, which has been proposed by other groups for other ferrocheletases.[20] The first binding mode, which is populated when no PPIX is present in the enzyme active site, is kinetically off-pathway and inhibitory. Structural analysis of the *S. cerevisiae* (PDB 1L8X) and *B. subtilis* (PDB 3M4Z) chelatases found consistent metal ion binding to E275 and H194 (*E. coli* numbering) in the absence of porphyrin. We hypothesize that these two residues may comprise the inhibitory binding site, which sits ‘on top’ of the porphyrin when the hydrophobic binding interface is oriented ‘below’ (Figure 3.22). This

combination of kinetic and structural evidence suggests that metalation occurs from the bottom face of the porphyrin, near L13, where we found that alteration of this residue strongly influenced metal specificity. In this scenario, the metal is dewetted as it binds to the PPIX pyrrole nitrogens in the hydrophobic enzyme environment, whereupon the metal begins to insert into the PPIX cofactor. A base is required to deprotonate the pyrrole rings and steric considerations demand that these protons are removed from the opposite face of from which metal insertion occurs. In *Ec*HemH, H194 is positioned to play the role of general base. This hypothesis is corroborated by the absence of any chelatase activity in the H194X site saturation mutagenesis library. Were the role of H194 simply to pre-coordinate metal prior to insertion from the top face, one might expect that other metal coordinating residues maintain activity with at least some metals. How the second, final proton is removed during the metal insertion process is unclear, but these data provide a revised molecular model that can be interrogated in future studies.

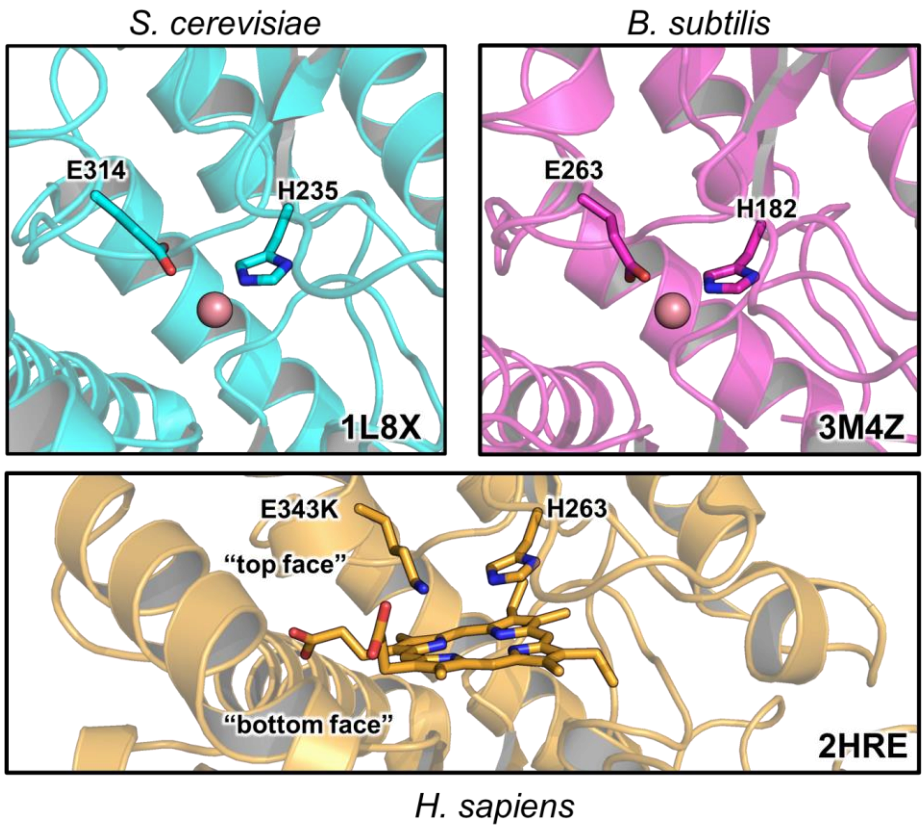


Figure 3.22. Structural analysis of crystal structures of HemH homologs with substrates bound. S. cervisiae HemH (top left – blue) is a protoporphyrin IX ferrochelatase and is crystalized with Co^{2+} bound to E213 and H235 active site residues in the absence of PPIX. B. subtilis HemH (top right – pink) is a coproporphyrin ferrochelatase and is crystalized with Co^{2+} bound to E263 and H182 active site residues. H. sapiens HemH (bottom – yellow) is a protoporphyrin IX ferrochelatase. The E343K variant of this enzyme is crystalized with PPIX bound in the active site. Each enzyme is depicted in analogous orientation, such that the Glu and His residues are positioned on the "top face" of the PPIX substrate. In the analogous EcHemH enzyme, we hypothesize that metalation occurs form the "bottom face" of the PPIX substrate, concomitant with deprotonation of PPIX by the His residue positioned on the "top face."

To test the effect of chelatase metal specificity on the production of CoPPIX versus heme in BL21(DE3), we undertook an engineering campaign to improve the Co selectivity of the native BL21(DE3) ferrochelatase. The enzyme was surprisingly tolerant to mutation. Site saturation mutagenesis at 7 of the 10 active site residues we tested yielded active chelatase variants. By screening on a mixture of metal substrates, we were able to probe the influence of residues on metal selectivity in a fashion that is decoupled from overall protein activity. That is, a single substrate screen cannot distinguish between mutations that decrease activity with a specific metal or if the mutation alters properties like catalyst concentration. Mutations throughout the active site

have marked effects on metal specificity indicating that complex and synergistic interactions with PPIX, metal, and active site residues all influence metal insertion. This assessment aligns with previous studies that link porphyrin distortions in the active site to metal specificity.[28] Through several iterations of screening both *in vivo* and *in vitro*, we were able to identify a mutation, L13R, that causes a 30-fold shift in cobalt versus iron specificity. The discovery of this variant enabled us to explore how expression of this cobalt-chelatase affected the distribution of MPPIX products generated *in vivo*.

Notably—and contrary to our initial hypothesis—we found that the specificity of an over-expressed chelatase plays only a minor role in determining the distribution of metalated porphyrins *in vivo*. Instead, the native promiscuity of *Ec*HemH, which is 2-fold selective for Fe^{2+} over Co^{2+}, is already sufficient for efficient CoPPIX production *in vivo*. Even large changes to Co^{2+} versus Fe^{2+} specificity, such as those exhibited by the variant L13R, do not outweigh the myriad other metabolic factors governing PPIX metalation.

Metal homeostasis is a key factor in determining the distribution of metalated porphyrin products. One recent study of metal homeostasis in *E. coli* JM109(DE3) highlights that *E. coli* BL21(DE3) lacks the entirety of the *rcn* operon,[49,50] which is responsible for Co^{2+} and Ni^{2+} efflux.[51,52] Indeed, when we tested the ability of JM109(DE3) *E. coli*, which contains a functional *rcn* operon, to biosynthesize and incorporate CoPPIX, the JM109(DE3) strain was far less effective than BL21(DE3). This observation suggests that the success of our cobalt-substituted hemoprotein expression method may be due to the inability of *E. coli* BL21(DE3) to efflux cobalt. This hypothesis is corroborated by our finding that BL21(DE3) can produce trace amounts of NiPPIX, despite the relatively low efficiency of Ni^{2+} insertion by *Ec*HemH (185-fold less efficient with Ni relative to Fe, Table 3.1). Uniquely in BL21(DE3), a high concentration of cobalt in the growth media leads to high cytosolic cobalt concentrations and, as a result, efficient production of CoPPIX by a sufficiently promiscuous chelatase.

While not explicitly addressed in this study, differences in the catabolism of heme and CoPPIX might also contribute to the distribution of these metalloporphyrins under our expression conditions. In pathogenic *E. coli*, heme is degraded by the heme oxygenase, ChuS (UniProt accession: A0A271QSA5).[53,54] These enzymes catalyze oxidative ring opening of the iron porphyrin cofactor, yielding the linear tetrapyrrole, biliverdin IXα, via an iron-oxo intermediate and are inhibited by non-native metalloporphyrins such as CoPPIX.[55,56] Heme catabolism in non-pathogenic *E. coli* such as BL21 is not as well understood. The peroxidase enzyme YfeX from BL21 has been hypothesized to be a heme degrading enzyme,[57] but these findings have been refuted by others.[58] It is also possible that *E. coli* BL21(DE3) harbors a yet uncharacterized heme-degrading enzyme or that our and others' observations of heme loss in cell lysates are due to an alternative, non-enzymatic process. Regardless, native heme catabolic pathways may be specific to the iron-chelated metalloporphyrin, allowing CoPPIX to accumulate under expression conditions while excess heme is quickly degraded.

Chapter 7 3. 5. Conclusion

Hemoproteins comprise a diverse family of widely studied and used biocatalysts. Often, BL21(DE3) *E. coli* is the organism used to produce these hemoproteins—and the cofactors they bear—at high titers. By studying the metalation and incorporation of a non-native heme analog, CoPPIX, we provide insight into how biological metalation is controlled in this ubiquitous model organism, and guide future campaigns to engineer metallocofactor biosynthesis. We found that metalloporphyrin distribution is controlled by many factors, including chelatase specificity, environmental metal concentrations, metal homeostasis machinery, and cofactor catabolism. In nature, these factors work in concert to ensure proper metalation of heme. In contrast, by studying the disruption of these factors, we developed a method for efficient biosynthesis of CoPPIX, which we hypothesize functions according to the following paradigm: In BL21(DE3) *E. coli*, the dysregulation of Co^{2+} and Ni^{2+} efflux, when exposed to a high concentration of cobalt in the media, leads to an excess of intracellular Co^{2+}. The native chelatase, *Ec*HemH, is sufficiently effective as a cobalt chelatase to produce CoPPIX when intracellular Co^{2+} is abundant. In the presence of hemoprotein overproduction, the excess metallocofactor is loaded into these hemoproteins, lessening the toxic effects of both the free cofactor and Co^{2+}. Any metallocofactor not bound by heterologously expressed hemoprotein is prone to degradation by heme catabolic pathways, to which the iron-bound cofactor is most susceptible. The net result is efficient heterologous expression of hemoproteins that are highly substituted with the CoPPIX cofactor.

The biosynthesis and incorporation of artificial cofactors is a long-standing and alluring challenge in the field of biocatalysis. New-to-nature cofactors imbue enzymes with alternate modes of reactivity, but because these enzymes are difficult and tedious to produce, widespread applications are rare. The enantioselective carbene and nitrene transfer reactivity of metal-substituted hemoproteins in particular has led to the development of several strategies for the incorporation of pre-fabricated metalloporphyrins into hemoproteins,[59] but the production of these artificial metalloenzymes remains significantly more difficult than that of the native hemoprotein.

Here, we report a fully-biosynthetic and highly-efficient method for the production and incorporation of the artificial cofactor CoPPIX into hemoproteins. This method is similar in efficiency to the production of native hemoproteins, and thus enables their widespread use as biocatalysts and development through directed evolution and high-throughput protein engineering.

Chapter 8 3. 6. Materials and Methods

3. 6. 1. General Experimental methods

All chemicals and chemical standards were purchased from commercial suppliers (Sigma-Aldrich, VWR, Goldbio, Thermo Fisher, Frontier biosciences, Fluka), and used without further purification. Unless otherwise noted, all media and solutions were prepared using ultrapure water (≥18 MΩ, from Thermo Scientific Barnstead Nanopure water purification system). Immobilized metal affinity chromatography resin was purchased from Thermo Fisher.

3. 6. 2. Equipment and instrumentation

New Brunswick I26R, 120 V/60 Hz shakers (Eppendorf) were used for cell growth. Cell disruption was accomplished via sonication with a Sonic Dismembrator 550 (Fischer Scientific). Electroporation (for transformation and cloning) was achieved using a BioRad Eporator. Optical density measurements of liquid cultures were recorded on an Ultrospec 10 cell density meter (Amersham Biosciences). Ultra pressure liquid chromatography – mass spectrometry (UPLC-MS) data were collected on an Acquity UHPLC with an Acquity QDA MS detector (Waters). Absorbance measurements used for kinetic studies were collected on a UV-2600 Shimadzu or a Varian Cary 4 Bio spectrophotometer, set to a spectral bandwidth of 0.5 nm. ICP-MS data were collected with an Agilent 8900 triple quadrupole ICP-MS. Parent glycerol stocks for all growth strains were obtained from New England Biosciences (BL21(DE3) *E. coli*) or Invitrogen (JM109(DE3) *E. coli*).

3. 6. 3. Cloning and expression of *Ec*HemH

The gene encoding ferrochelatase was obtained directly from genomic DNA isolated from *E. coli* BL21(DE3) using a Monarch® Genomic DNA purification kit (New England Biolabs). The

gene was amplified via PCR with the primers from table 1, inserted into pET22b(+) via Gibson assembly as previously described, and transformed into *E. coli* BL21(DE3) via electroporation. For overexpression of *Ec*HemH, 10 mL of overnight starter culture of *E. coli* BL21(DE3) harboring the HemH expression plasmid were added to 1 L of TB media supplemented with 100 mg/L ampicillin in a 2.8 L Fernbach flask. This expression culture was incubated at 37 C and 225 rpm until it reached an OD_{600} ~1.5. At this point, IPTG was added to 1 mM and the culture was returned to 30°C and 225 rpm to incubate for 16 to 20 hours. Cells were harvested via centrifugation at 4,000 xg and 4 °C for 15 minutes and stored at -20 until purification.

3. 6. 4. *Ec*HemH purification

Expression pellets harboring *Ec*HemH were resuspended in lysis buffer [50 mM tris(hydroxymethyl)aminomethane (Tris), 100 mM NaCl, 2 mM $MgCl_2$, 1.0% (w/v) sodium cholate, 1 mg/mL lysozyme, 0.2 mg/mL DNase, pH 8.0] using 4 mL per gram of cell pellet mass. This mixture was incubated at room temperature for 30 minutes. The lysate was then transferred to ice and sonicated for 10 minutes (1 s pulse, 1 s pause, 30% amplitude). Lysate was cleared of cellular debris by centrifugation at 20,000 x*g* and 4°C for 30 min. Supernatant was pooled and applied to Ni-NTA resin equilibrated with binding buffer (50 mM Tris, 100 mM NaCl, 1.0% (w/v) sodium cholate, pH 8.0) for 1 h. The resin was washed with 4 column volumes of binding buffer supplemented with 45 mM and 75 mM imidazole. Purified HemH was eluted using binding buffer supplemented with 300 mM imidazole and concentrated to < 5 mL with centrifugal protein concentrators (Amicon, 30 kDa molecular weight cutoff). The protein was then buffer exchanged into storage buffer (50 mM tris, 100 mM NaCl, 1.0% (w/v) sodium cholate, pH 7.0, 20% (v/v) glycerol) via desalting column, flash frozen with liquid N_2 and stored at -80°C.

3. 6. 5. Preparation of protoporphyrin IX and metal stock solutions

All procedures related to preparation of metal and protoporphyrin IX stock solutions are previously described by Hunter, et al.[23] Briefly, protoporphyrin IX stocks were generated by vortexing 1 to 3 mg of protoporphyrin IX disodium salt with 300 µL of 2 M NaOH. The substrate was completely solubilized by sequential addition of 1 mL 10% (w/v) tween 80, 500 µL of 1 M MOPS or Tris (to match reaction buffer), pH 7.0, and 200 µL of 3 M HCl with vortexing after each addition. The final solution was diluted to 5 mL with >18.2 MΩ water, and the PPIX concentration was measured spectrophotometrically via dilution of the stock into 3 M HCl using an extinction coefficient of 297,000 $mM^{-1}cm^{-1}$ at 408 nm. Metal stocks were prepared as 0.1 M divalent metal chloride salts in 1.2 M HCl. Metal stocks were diluted with >18.2 MΩ water to prepare 1000, 100, and 10 µM working solutions for assays.

3. 6. 5. Measurement of initial rates and progress curves

Initial rates and progress curves were collected by tracking absorbance of the *Ec*HemH reaction mixture over time. For reactions including Fe^{2+} and Co^{2+}, absorbance was tracked at 407 nm. For reactions including Ni^{2+}, Cu^{2+}, and Zn^{2+}, absorbance at 417 nm was tracked. Reactions were conducted in Tris-HCl reaction buffer (100 mM buffer, 400 mM NaCl, 0.2% (w/v) tween 80, pH 7.0). For each measurement, *Ec*HemH and PPIX were mixed with reaction buffer in a quartz cuvette. Each reaction was initiated by addition of metal solution and placed into the spectrometer. The delay time between addition of metal and data collection was approximately 1 s. Initial rate experiments were performed with 10 to 50 nM EcHemH, and all data was collected at 25°C. Initial velocities, V, were calculated using the slope of the first 20 to 50 s of absorbance data, calculated by linear least squares fit, and equation 1, normalizing for protein concentration:

$$V = \frac{-Slope}{(\varepsilon_{PPIX} - \varepsilon_{MPPIX}) * [HemH]} \quad (1)$$

where ε_{PPIX} and ε_{MPPIX} represent the extinction coefficients of protoporphyrin IX (PPIX) and the metalated PPIX product, respectively. Extinction coefficients for PPIX under reaction conditions are 1.07x10^5 M^{-1} cm^{-1} at 407 nm and 8.04x10^4 M^{-1} cm^{-1} at 417 nm, as measured by Hunter et al.[60] Apparent extinction coefficients for each metalloporphyrin species under reaction conditions were calculated by end point analysis in which 0.1 μM *Ec*HemH and excess metal ions were used to convert known concentrations of PPIX into corresponding metalloporphyrins under reaction conditions: FePPIX, 4.15x10^4 M^{-1} cm^{-1} (407 nm); CoPPIX, 4.55x10^4 M^{-1} cm^{-1} (407 nm); NiPPIX, 3.88x10^4 M^{-1} cm^{-1} (417 nm); CuPPIX, 4.62x10^4 M^{-1} cm^{-1} (417 nm); ZnPPIX, 1.58x10^5 M^{-1} cm^{-1}.

3. 6. 6. Mechanism modeling, global fitting of initial rates and progress curves

To determine the kinetic parameters for Fe^{2+} insertion into PPIX by *Ec*HemH, individual rates at various concentration of PPIX and Fe^{2+}. These rates were fit globally to the integrated rate law for sequential ordered binding. The kinetic model for the reaction with iron is shown in scheme 3.1.

$$\mathbf{E + PPIX \rightleftharpoons E{:}PPIX + M^{2+} \rightleftharpoons E{:}PPIX{:}M^{2+} \longrightarrow E + MPPIX}$$

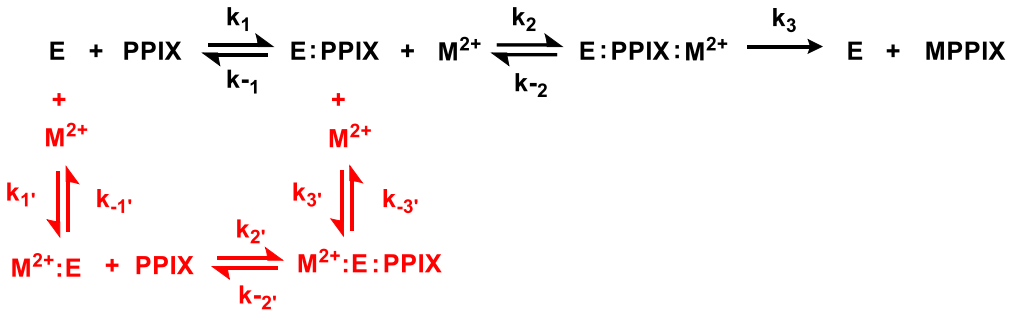
Scheme 3.1

The rate data was fit usinga least squares fit using the following integrated rate law:

$$v_0 = \frac{k_{cat}[PPIX][Fe]}{K_{iPPIX}\,K_M^{PPIX} + K_M^{Fe}[PPIX] + K_M^{PPIX}[Fe] + [Fe][PPIX]} \qquad (1)$$

To model possible mechanisms, initial rates of CoPPIX formation were modeled to different mechanisms as shown in Figure 3.24 using DynaFit 4.[61] This program identifies the best fit of the experimental data by numerical integration of a system of differential equations as defined by each unique kinetic mechanism. In each calculation, DynaFit was set to the following parameters: task, fit; data, rates; approximation, none; delay = 1 s. All forward and reverse rate constants were allowed to vary. We tested several kinetic mechanisms, as shown in Figures 3.23-

27. A discussion of each model and its ability to qualitatively recapitulate the experimental data are discussed as well. The non-competitive substrate inhibition kinetic model which we determined to be the best fit to the Co^{2+} and, and which was subsequently = used to fit the data for Ni^{2+}, Cu^{2+}, and Zn^{2+} is given in Scheme 3.2.



Scheme 3.2

To calculate kinetic parameters from the microscopic rate constants provided by non-linear fits of experimental data to Scheme 3.1, we used the following equations:

$$K_M^{Metal} = \frac{k_3 + k_{-2}}{k_2} \quad (2)$$

$$k_{cat} = k_3 \quad (3)$$

where the microscopic rate constants are defined as shown Scheme 3.2. The macroscopic rate constants are given in Table 3.1, and the calculated microscopic kinetic parameters are given in Table 3.7. The calculated %CV for many of these microscopic rate constants are exceedingly large, due to the complexity of the model and the codependency of these variables. However, we found that the macroscopic ratios (K_M and catalytic efficiency) of these values were well-determined.[62] Furthermore, the overall shape of the fit was relatively consistent.

3. 6. 7. Construction of HemH sequence alignment and LogoPlots

To formulate a representative sequence alignment of all known HemH homologs, we first queried InterPro for HemH protein sequences as defined the ferrochelatase family, IPR001015.[63]

This provided a dataset of 13,471 protein sequences. To eliminate sequence redundancy and reduce the size of this dataset, we clustered these sequences into 75% sequence identical bins using CD-HIT.[64,65] This reduced dataset, consisting of 6,539 sequences, was aligned using Muscle 5.1.[66] Outlier sequences that contained unique, >25 residue inserts were removed to clear any non-homologous proteins and mistranslated sequences from the dataset. Using the final sequence alignment of 5,026 proteins, we generated sequence logos for each active site residue using the WebLogo 3 server (https://weblogo.threeplusone.com/).[37,67]

3. 6. 8. EcHemH Library construction and expression

Based on homology models, bioinformatic information, and previously studied variants of interest, 10 sites were selected for site saturation mutagenesis. These sites were: L13, F31, R57, Y128, S134, H194, W242, C273, E275, and F269. Primers were ordered for these sites from Integrated DNA Technologies (IDT), (Table 3.4). For each mutation site, three primers for the forward direction, NDT, VHG, TGG, were mixed in a 12:9:1 ratio, respectively, to produce 22 unique codons accounting for 20 unique protein sequences.[68] Single overlap extension PCR was conducted with the HemH:Pet22b(+) construct and the template and forward and reverse primers to amplify the entire vector. The linearized vector was then re-circularized by Gibson Assembly (New England Biolabs) and then transformed into Electrocompetent BL21(DE3). After 45 minutes of recovery in TB broth, the outgrowth was plated onto LB+Amp media and incubated overnight at 37°C.

Single colonies were picked and used to inoculate 500 µL of TB + Amp broth in each well of a 96-well deep well plate. Two columns of the 96-well plate were used as internal controls, including 8 wells that were inoculated with colonies expressing parent *Ec*HemH, 4 wells inoculated with non-HemH pet22b construct negative controls, and 4 sterile control wells. The plate was grown overnight at 37°C with shaking at 220 rpm.

The following day, expression plates and glycerol plates were generated. For glycerol plates, 50 µL of each well of the overnight culture plate was mixed with 50 µL of a 50% glycerol aqueous solution. Glycerol plates were stored at -80°C until needed to obtain genetic material for Sanger sequencing (Functional Biosciences). For expression plates, 50 µL of each well of the overnight culture plate was used to inoculate 600 µL of fresh TB + Amp media in a 96-well deep well plate. These plates were grown for 5 hours at 37°C with shaking at 225 rpm, and then placed on ice-water for 20 minutes. Expression was induced by the addition of 100 µL of TB + Amp with 7.5 mM IPTG, bringing the total expression volume to 750 µL with 1 mM IPTG. The plates were then incubated for 16 hours at 30°C at 180 rpm. Cell pellets were harvested by centrifuging each plate at 4000 x g for 15 minutes and discarding spent media. Plates were stored at -20°C until lysis and screening.

3. 6. 9. *In vitro* screen of HemH Site Saturation Libraries

Expression plates were lysed using 300 µL of lysis buffer in each well, consisting of the following: 1 mg/mL Hen Egg white lysozyme, 0.2 mg/mL DNaseI from Bovine Pancreas, 2 mM $MgCl_2$, 0.5 % w/v Sodium Cholate, 100 mM NaCl, 50 mM Tris-HCl buffered to pH = 7.0. The plates were resuspended in this solution and shaken at 37°C. After 1 hour, lysate was clarified by centrifugation at 4000 x g for 20 minutes.

Plate screen reactions were conducted in 96-well deep-well plate using the following procedure. The reaction buffer was assembled as follows: 100 mM Tris buffer with 400 mM NaCl and 0.2% (w/v) Tween80. A stock solution of PPIX was made freshly each day of library screening, as the solution is unstable and formed soluble aggregates over time, as evidenced by characteristic changes to the UV-visible spectrum.[69] First, 2-3 mg of Protoporphyrin IX disodium salt were vortexed in 1.0 M NaOH until no large solid particles remained. Then, 1.0 mL of 10% Tween80 was added, and the solution was vortexed again until the solution became transparent

and red. Next, 0.5 mL of 1.0 M Tris-HCl was added, and the solution was diluted to approximately 5 mL with water. Lastly, 150 µL of 3.0 M HCl was added, and the concentration of the stock solution was calculated by taking a UV-Vis spectrum of the solution diluted in reaction buffer (see above). An extinction coefficient of 1.07 x 10^5 M^{-1} at 407 nm was used to calculate the concentration of the stock solution. Metal salts ($CoCl_2$, $NiCl_2$, and $CuCl_2$) were made as 100 mM stock solutions in water. To assemble the reaction master mix, PPIX (12 µM) was added to 60 mL of reaction buffer. $NiCl_2$ (1.05 mM), $CuCl_2$ (93 µM), and $CoCl_2$ (23 µM) were added, and the pH was adjusted to 7.0 using pH paper and 1.0 M HCl. To each well of a 96-well deep-well plate was added 600 µL of reaction master mix, and 100 µL of Lysate (above) was added to initiate the reaction. The concentration of components in the reaction plate was as follows: 10 µM PPIX, 900 µM $NiCl_2$, 80 µM $CuCl_2$, 20 µM $CoCl_2$. After 1 hour at room temperature, the reaction was quenched with the addition of 100 µL of 3.0 M HCl. Then, 0.4 mL of a 2:1 Ethyl Acetate-pyridine mixture was added, with 50 µM anthracene added as an internal injection standard. Each well was pipetted up and down thoroughly to mix the immiscible layers. The aqueous and organic layers were separated by centrifugation at 4000 x g for 5 minutes, and the upper organic layer was transferred to a separate 96-well plate for analysis by UPLC.

The UPLC chromatographic method consisted of an isocratic mobile phase (20% water in acetonitrile with 0.1% formic acid), flowing at a rate of 0.6 mL/min and 30 ºC, over an ACQUITY UPLC-CSH Phenyl-Hexyl 1.7µm column (Waters). A 3.0 µL injection volume was used. Clear porphyrin signals were observed at distinct retention times via absorbance at the porphyrin Soret maximum (CoPPIX: 1.4 min at 423 nm, PPIX: 2.0 min at 400 nm, NiPPIX: 2.3 min at 395 nm, CuPPIX: 2.4 min at 401 nm), as assessed by injection of authentic standards. Integrated absorbance values were measured relative to the anthracene internal standard (1.8 min at 365 nm). These relative integration values were used directly to assess shifts in porphyrin product content between each well.

3. 6. 10. *In vitro* screen of Co^{2+} and Fe^{2+} activity of EcHemH and L13R variant

In duplicate, purified HemH enzyme (5 µL of a 50 µM solution in 50 mM Tris-HCl, 100 mM NaCl, 1% v/v cholate, pH = 7.0) was added to 1 mL of reaction buffer (100 mM Tris-HCl, 400 mM NaCl, 0.2% Tween 80, pH = 7.0) in a 1.7 mL centrifuge tube, bringing to final concentration of enzyme to 250 nm. To this solution was added 10 µL of 0.25 mM PPIX solution, prepared as described previously, such that the concentration of PPIX was 5.0 µM. $Fe(NH_4)_2(SO_4)_2$ and $CoCl_2$ stock solutions (100 mM) were prepared freshly in 1.0 M sodium ascorbate, pH = 7.0. A metal master mix with equimolar amounts of each metal was prepared from these solutions (10 mM $Fe(NH_4)_2(SO_4)_2$, 10 mM $CoCl_2$, 200 mM ascorbate). This solution (5 µL) was added to the reaction mixture to initiate the reaction, such that the final concentration of each metal 100 µM and the final concentration of ascorbate was 2 mM. After 20 minutes at room temperature, 100 µL of 1.0 M HCl was added to halt reaction progress. A solution of 2:1 ethyl acetate/pyridine (300 µL) was added, and the reactions were vortexed briefly before centrifugation at 40,000 x g for 1 minute. The organic layer was isolated and analyzed by UPLC as described above. FePPIX eluted at 1.9 min and was measured via absorbance at 400 nm.

3. 6. 11. *In vivo* screen of Co^{2+} and Fe^{2+} activity of EcHemH

E. coli BL21(DE3) harboring expression vectors for DyP (pBAD) and *Ec*HemH (Pet22b) were grown overnight in TB broth containing 50 mg/L Kan, 100 mg/L Amp. These starter cultures were subcultured (2 mL) into a 1 L baffled Erlenmeyer flask containing 200 mL of TB media supplemented with the appropriate antibiotics, 100 µM $CoCl_2$, 50 µM $Fe(NH_4)_2(SO_4)_2$, 250 µM δ-aminolevulinic acid, and 100 µM IPTG. These cultures were grown at 37 °C and 225 rpm until they reached an optical density (OD_{600}) of 0.5 to 0.6. The amount of time from subculturing to induction varied from 4 to 8 h. Cultures were placed on ice for 30 minutes. At this point, L-arabinose was added to a final concentration of 0.2% (w/v), and the cultures were then grown at

25 °C and 225 rpm overnight. Cells were harvested by centrifugation and pellets were frozen at -20 °C until purification. Cells were lysed and purified according to a previously reported procedure.[13] After purification, proteins were buffer exchanged into 100 mM Tris buffer, pH 7.0 made with >18.2 MΩ water.

For experimental ease, the variant co-expression described above was repeated in two separate batches, with controls (Parent *Ec*HemH and GFP co-expression) repeated for each batch. The purified protein from each of these expressions were stored at -80°C until digestion and analysis by ICP-MS, described below.

3. 6. 12. ICP-MS Analysis of protein samples

ICP-MS analysis was conducted based a previous procedure.[13] Hemoprotein samples were digested by addition of 145 µL of trace analysis grade 70% nitric acid to protein stocks (200 to 400 µL). Samples were incubated at 90°C for 1 hour to ensure complete digestion and subsequently diluted with >18.2 MΩ water to give a final concentration of 1% nitric acid. Cobalt and iron standard solutions were prepared in 1% nitric acid solutions from 1000 ppm atomic absorption standards. Cobalt and iron content of protein and standard solutions was quantified by tandem mass spectrometry via detection of ions with *m/z* of 59→59 ion and 56→56 ion while using oxygen as the reaction gas.

3. 6. 13. Analysis of DyP protein expressed with different metals by UPLC

Expression of DyP protein with different metals.

DyP protein was expressed and purified according to the optimized method for expression of cobalt-substituted DyP in rich media, given below, except that 1.0 mM of $NiCl_2$, $CuCl_2$, or $ZnCl_2$ was added at the time of induction, rather than $CoCl_2$. Purification was analogous to preparation of the Co^{2+}-expressed protein.

UPLC Analysis of purified protein samples to assess metal content

Following expression and purification, 500 µL of purified protein sample was acidified with the addition of 200 µL of 3.0 M HCl in a 1.7 mL centrifuge tube. Next, 400 µL of a 2:1 mixture of ethyl acetate/pyridine were added to the acidified solution and vortexed briefly. The samples were centrifuged at 40,000 xg for 5 minutes, and the upper organic layer was analyzed by UPLC using the chromatographic method described previously for library screening.

3. 6. 14. Optimized method for expression of cobalt-substituted hemoproteins

Expression cultures of BL21(DE3) or JM109(DE3) cells bearing hemoprotein (DyP on a pBAD vector under arabainose inducible promoter, DyP_pBAD; or CYP119 on a Pet22b vector under an IPTG-inducible promoter, CYP119_Pet22b or pBAD vector, CYP119_pBAD) were inoculated by addition of 10 mL of overnight culture to 1.0 L of TB with the appropriate antibiotic (50 mg/L Kan for DyP_pBAD and CYP119_pBAD or 100 mg/L Amp for CYP119_Pet22b). Cultures were grown for 3 – 4 hours, until the OD_{600} reached ~ 1.0. At this point, cells were placed on ice for 20 minutes, and then 1.0 mM $CoCl_2$, 0.25 mM δALA, and either 0.2% w/v arabinose (DyP_pBAD) or 0.5 mM IPTG (CYP119_Pet22b) were added. Cells were incubated overnight at 25 °C for 16 to 20 h at 200 rpm. Cells were harvested by centrifugation at 4,000 xg and stored at -20 °C until lysis and purification.

DyP cell pellets were resuspended in 4 mL of lysis buffer per gram of wet cell pellet (100 mM KPi pH = 8.0, 100 mM NaCl with 1.0 mg/mL Hen Egg White Lysozyme and 0.2 mg/mL DNAse I, and 200 µM $MgCl_2$). Suspensions were incubated at room temperature for 1 hour, and then transferred to ice and sonicated for 10 minutes (1 s pulse, 1 s pause, 30% amplitude). Lysate was cleared of cellular debris by centrifugation at 20,000 x*g* and 4°C for 20 min. Supernatant was pooled and applied to Ni-NTA resin.

CYP119 cell pellets were resuspended 4 mL lysis buffer per gram of wet cell pellet (100 mM KPi pH = 6.0, 100 mM NaCl with 1.0 mg/mL Hen Egg White Lysozyme and 0.2 mg/mL DNAse I, and 200 µM $MgCl_2$). Suspensions were incubated at room temperature for 1 hour and then 75 °C for 45 minutes to lyse the cells. Heat treated lysates were clarified by centrifugation at 20,000 xg for 15 minutes. The pH of the supernatant was adjusted to 8.0 and then applied to Ni-NTA beads for purification.

It was noted that prolonged incubation of lysate supernatant with Ni-NTA resin resulted in a color change in the resin from blue to purple, indicating that the Ni^{2+} in the resin was gradually replaced with free Co^{2+} from expression supernatant. When the supernatant was *not* preincubated with Ni-NTA resin, the Ni-NTA beads remained blue, and protein yields were slightly higher. After applying supernatant to the Ni-NTA resin, the column was washed with 6 column volumes of wash buffer (100 mM KPi, 100 mM NaCl, 50 mM Imidazole, pH = 8.0) and then eluted with elution buffer (100 mM KPi, 100 mM NaCl, 250 mM Imidazole, pH = 8.0). The resulting eluent, which was orange-pink in color, was dialyzed into 100 mM KPi pH = 6.0 and flash frozen in liquid nitrogen prior to storage at -80 °C. Holoprotein concentration was determined by the modified pyridine hemachrome assay, as described previously[13].

Chapter 9 3. 7. Supplemental Materials

Chapter 10 3. 7. 1. Supplemental Figures

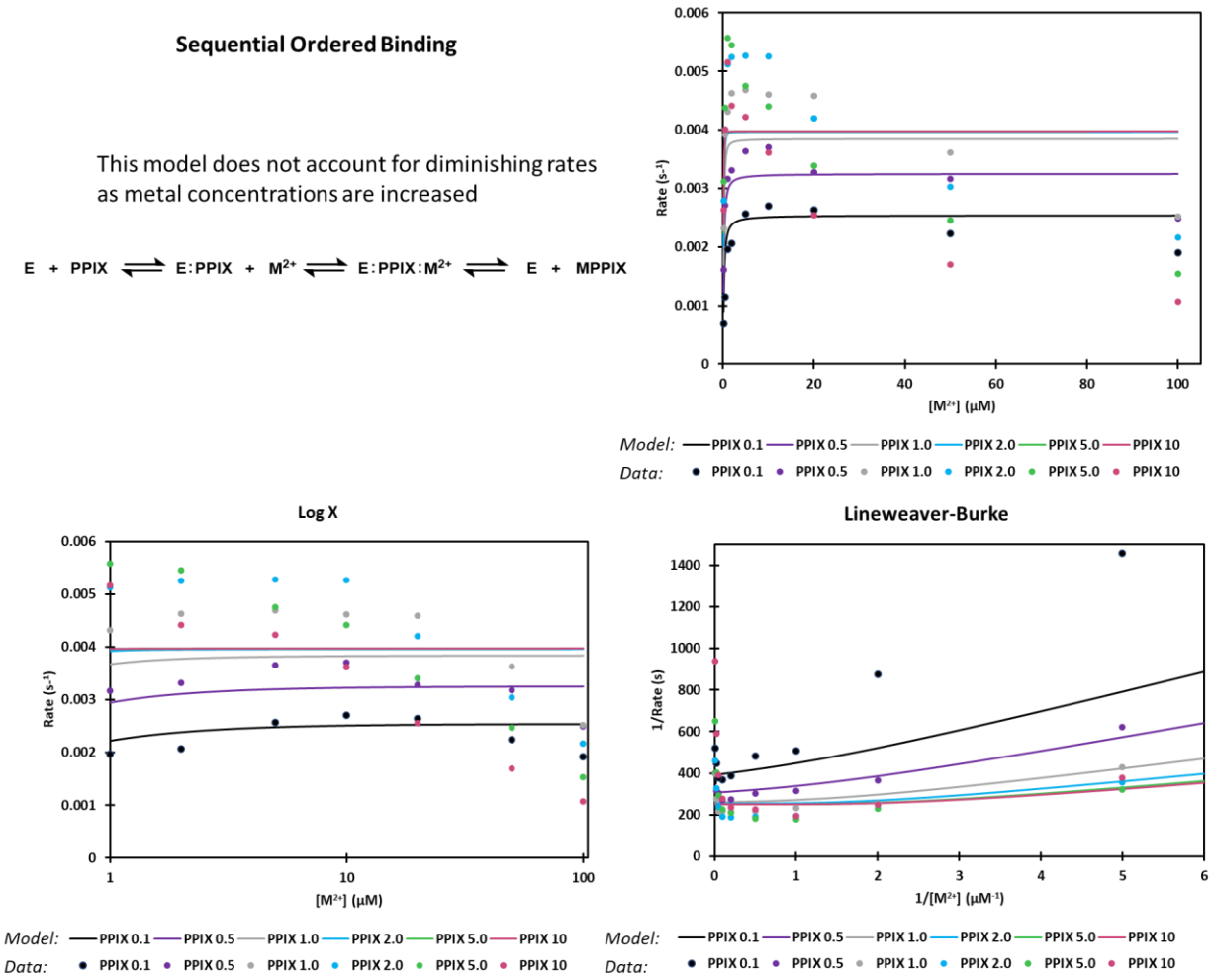


Figure 3.23. Fitting the initial rates for Co^{2+} insertion to a sequential ordered binding mechanism. Fits were generated using DynaFit 4 as described in the methods section,[61] and plotted in a variety of ways to assess fit.

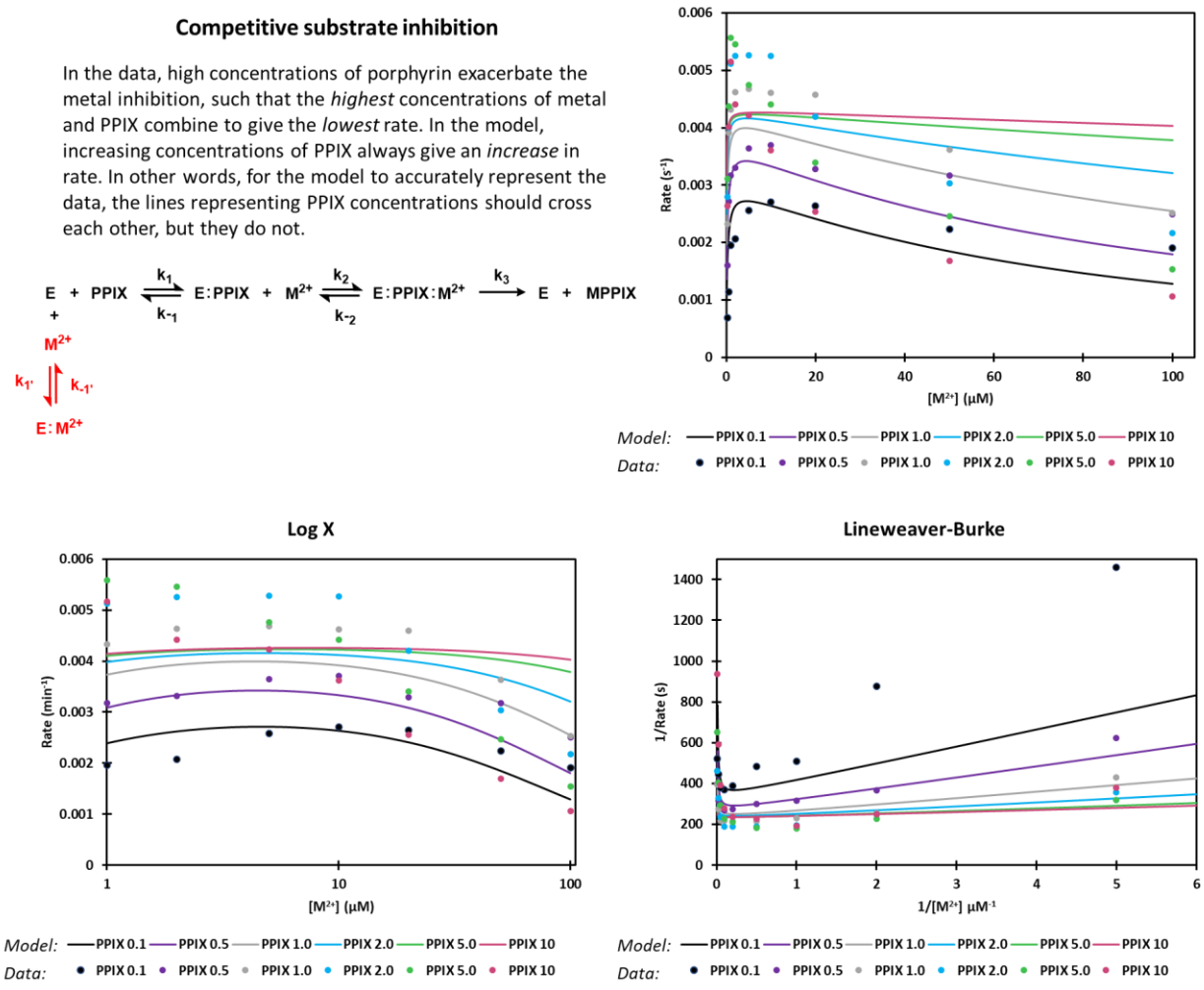


Figure 3.24. Fitting the initial rates for Co^{2+} insertion to a competitive substrate inhibition mechanism. Fits were generated using DynaFit 4 as described in the methods section,[61] and plotted in a variety of ways to assess fit.

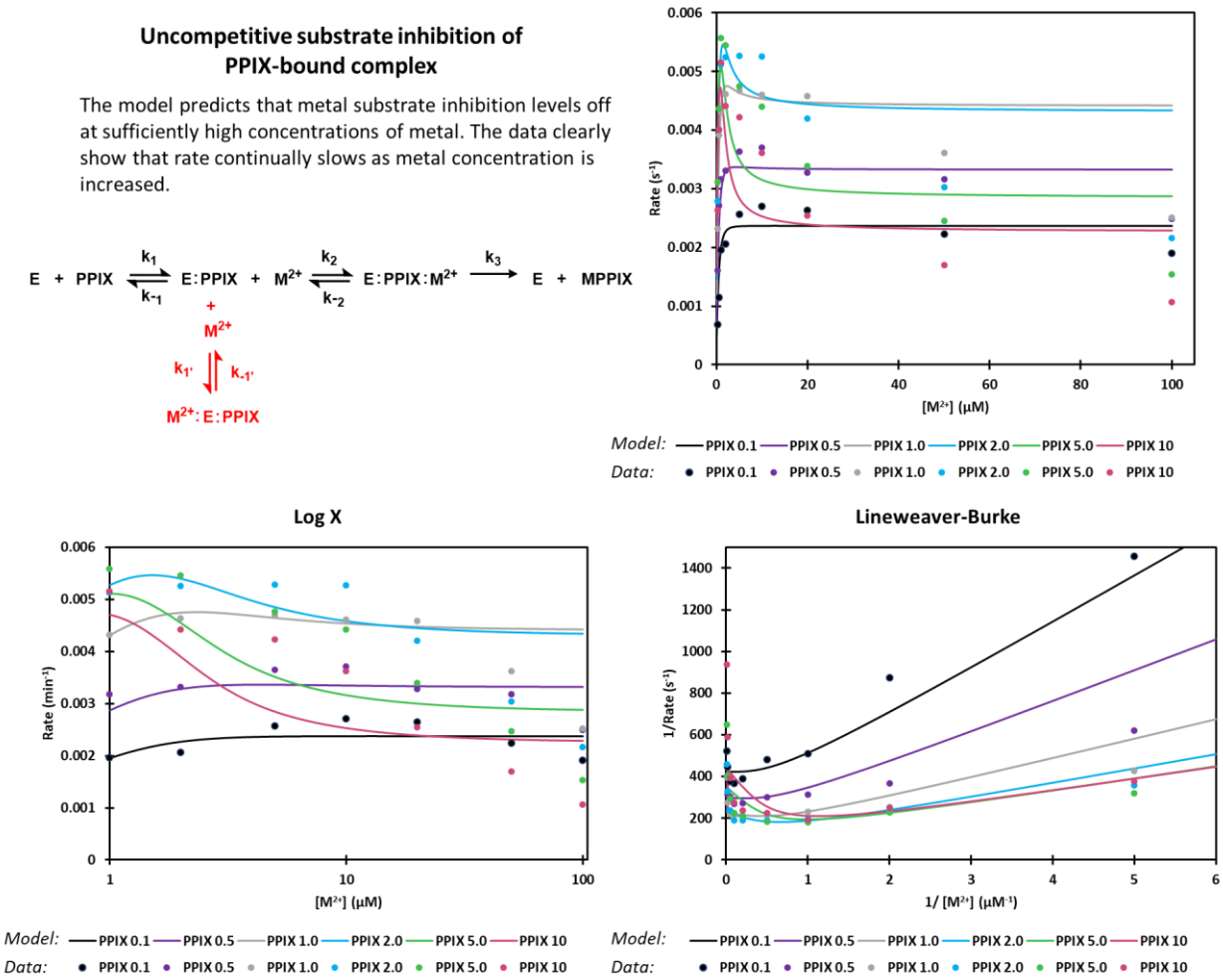


Figure 3.25. Fitting the initial rates for Co^{2+} insertion to an uncompetitive substrate inhibition of the PPIX-bound enzyme complex mechanism. Fits were generated using DynaFit 4 as described in the methods section,[61] and plotted in a variety of ways to assess fit.

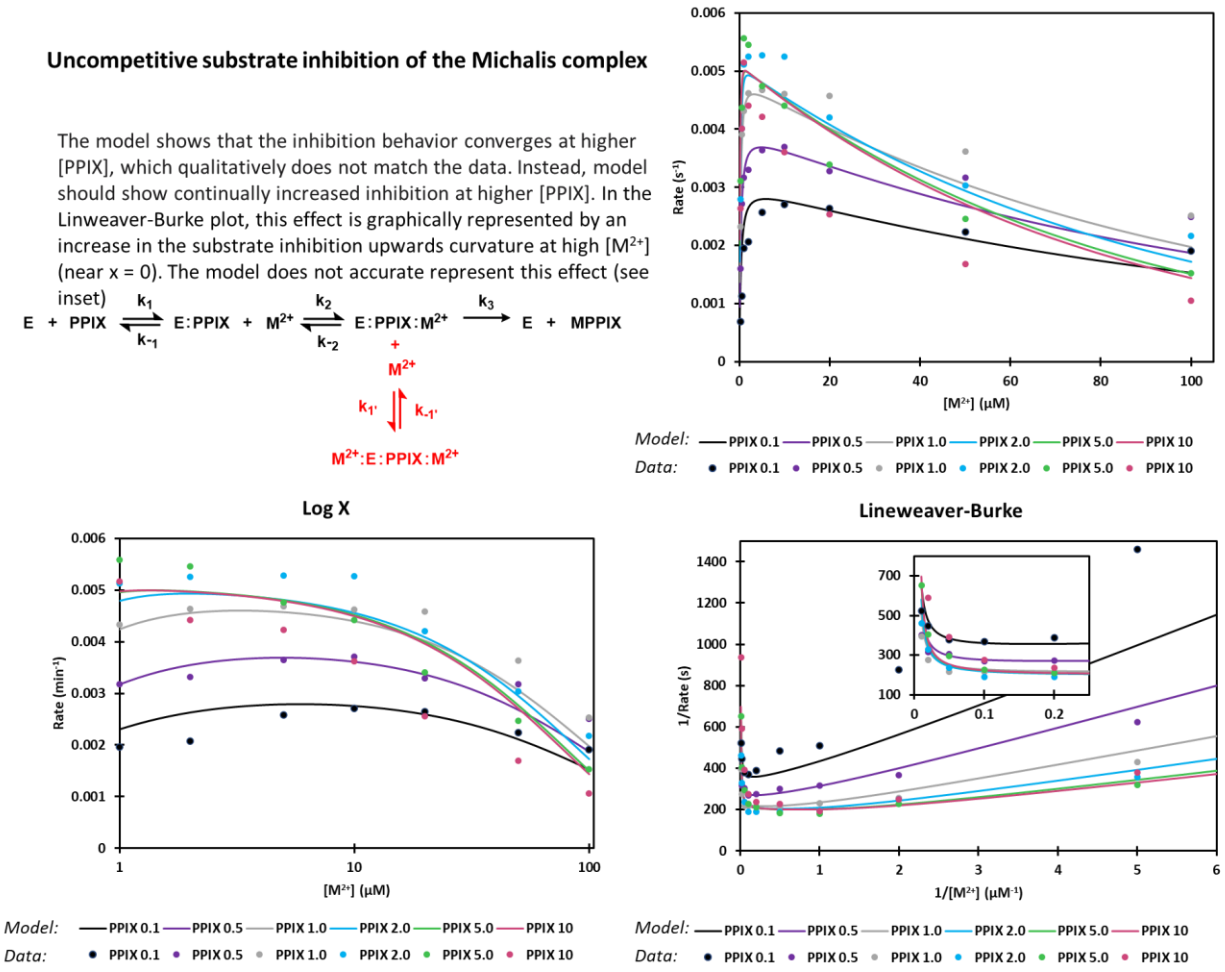


Figure 3.26. Fitting the initial rates for Co^{2+} insertion to an uncompetitive substrate inhibition of the Michalis-complex mechanism. Fits were generated using DynaFit 4 as described in the methods section,[61] and plotted in a variety of ways to assess fit.

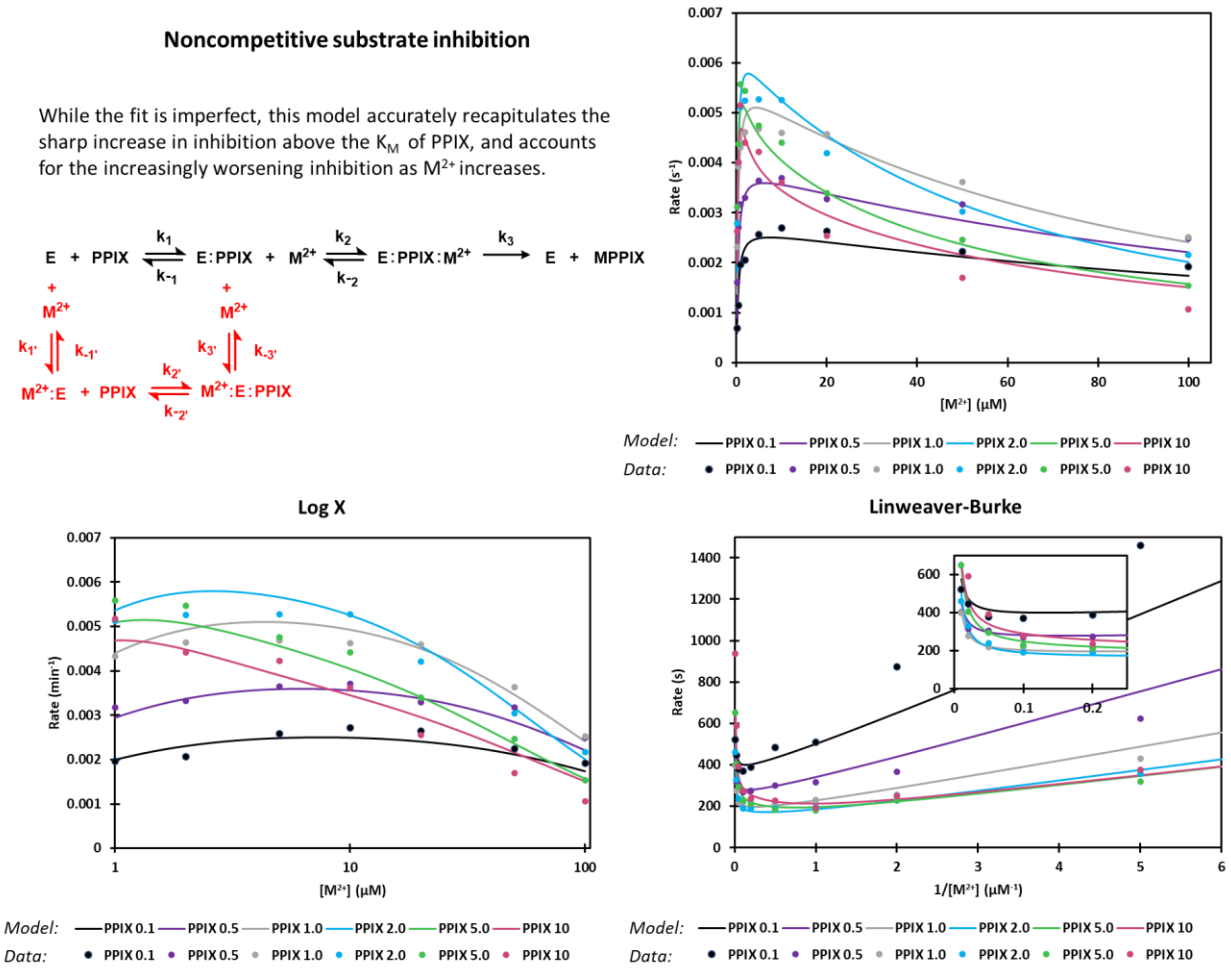


Figure 3.27. Fitting the initial rates for Co^{2+} insertion to a noncompetitive substrate inhibition of the Michalis-complex mechanism. Fits were generated using DynaFit 4 as described in the methods section,[61] and plotted in a variety of ways to assess fit.

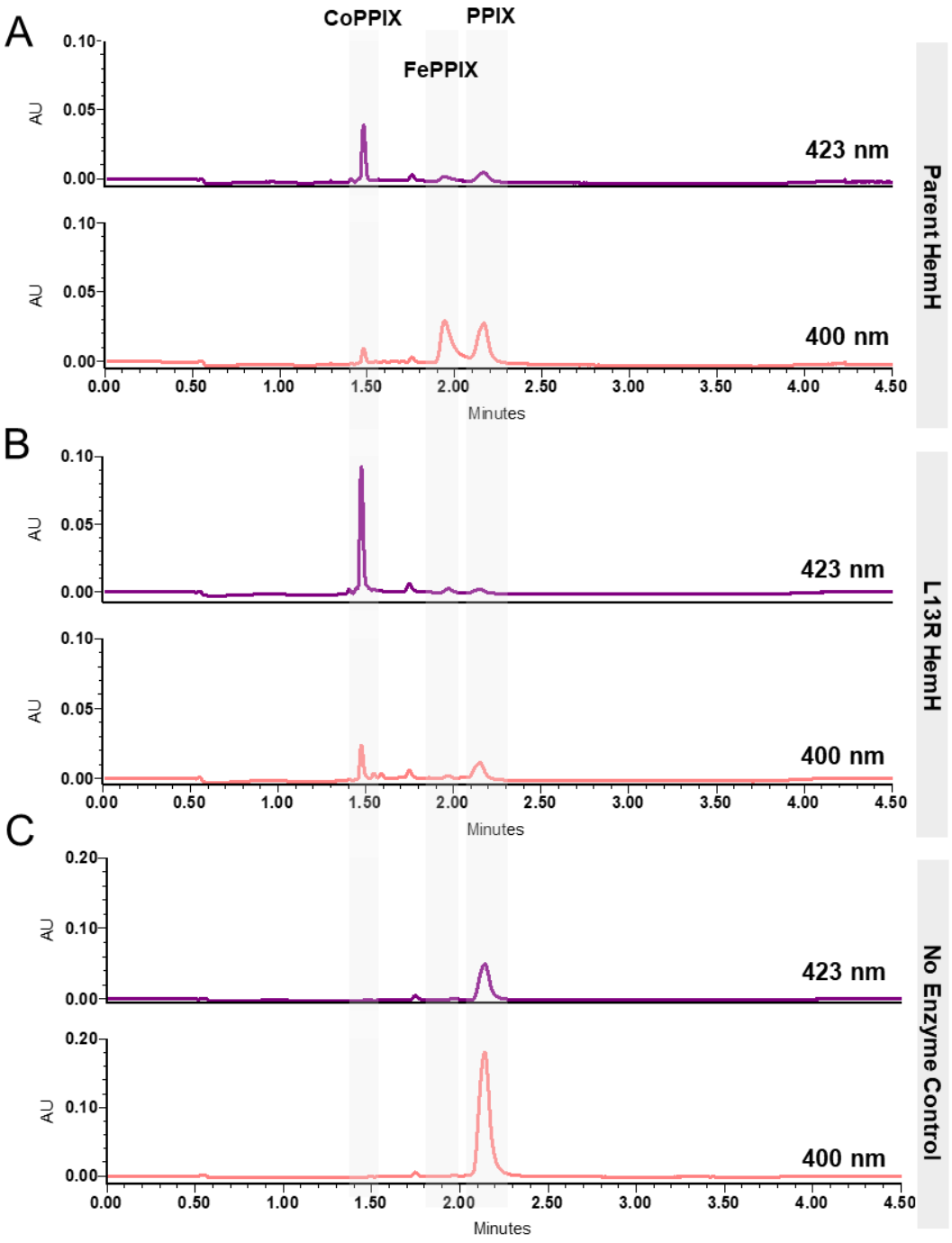


Figure 3.28. Distribution of porphyrin products from in vitro assay of EcHemH Co^{2+} versus Fe^{2+} specificity. Example UPLC traces resulting from the acid extracted porphyrin fraction of a reaction of Parent EcHemH (A) the L13R variant (B) and no enzyme (C) with Co^{2+}, Fe^{2+}, and PPIX. The purple lines represent chromatographic traces at 423nm and the pink lines represent chromatographic traces at 400 nm. The grey boxes represent the retention time of CoPPIX, FePPIX, and PPIX standards. Further details can be found in the methods section "In vitro screen of Co^{2+} and Fe^{2+} activity of EcHemH and L13R variant."

Chapter 11 3. 7. 2. Supplemental Tables

Table 3.4. Primers for amplification of EcHemH and generation of site-saturation libraries. ### denotes a 12:9:1 mixture of primers bearing NDT, VHG, TTG degenerate codons, respectively, for a total of 22-codons represented in the resulting libraries.[71]

Name	Sequence
HemH fw	GTTTAACTTTAAGAAGGAGATATACATATGCGTCAGACTAAAACCGGTATC
HemH rv	CCGGATCTCAATGGTGATGGTGATGGTGGCGATACGCGGCAACAAGATTAG
L13X fw	GTATCCTGCTGGCAAAC###GGTACGCCCGATGC
L13X rv	GTTTGCCAGCAGGATACCGGTTTTAGTCTGAC
F31X fw	GGTAAAACGCTATCTGAAACAA###TTAAGCGACAGACGC
F31X rv	TTGTTTCAGATAGCGTTTTACCGCTTCAGGTGTG
R57X fw	CGTGATTTTGCCGCTG###TCGCCGCGTGTGGC
R57X rv	GCGGCAAAATCACGCCGCGCAGCAATGGCC
Y128X fw	GTGGTGCTGCCGCTT###CCGCAATACTCC
Y128X rv	GCGGCAGCACCACAATATGATCTACATGC
S134X fw	GTGCTGCCGCTTTATCCGCAATACTCCTGT###ACGGTCGGTGC
S134X rv	GGATAAAGCGGCAGCACCACAATATGATCTACATGCTCTGCCAG
H194X fw	CTGCTGCTGCTCTCT###TGGGGCATTCCCCAGCG
H194X rv	TAAGAGAGCAGCAGCAGATCCGGTTCG
W242X fw	CGCTTTGGTCGGGAACCC###CTGATGCCTTATACCG
W242X rv	GGTTCCCGACCAAAGCGCGACTGAAAGGTCATCATCAC
F269X fw	GGTGATGTGCCCGGGC###GCTGCGGATTG
F269X rv	GCCCGGGCACATCACCTGTATATGAC
C273X fw	GCTTTGCTGCGGAT###CTGGAGACGCTG
C273X rv	ATCCGCAGCAAAGCCCGGGCAC
E275X fw	GCTGCGGATTGTCTG###ACGCTGGAAGAGATTGC
E275X rv	CAGACAATCCGCAGCAAAGCCCGG

Table 3.5. Gene sequences used. Start/Stop codons denoted in **red**

Name	Sequence
***Ec*HemH**	**ATG**CGTCAGACTAAAACCGGTATCCTGCTGGCAAACCTGGGTACGCCCGATGCCCCCACAC CTGAAGCGGTAAAACGCTATCTGAAACAATTTTTAAGCGACAGACGCGTGGTTGATACCTCA CGGTTGTTATGGTGGCCATTGCTGCGCGGCGTGATTTTGCCGCTGCGCTCGCCGCGTGTG GCGAAGCTGTATGCCTCTGTCTGGATGGAAGGTGGCTCGCCGCTGATGGTTTACAGCCGTC AGCAACAGCAGGCGCTGGCACAACGTTTACCGGAGACGCCCGTAGCGCTGGGAATGAGCT ACGGCTCGCCATCACTGGAAAGCGCCGTAGATGAACTCCTGGCAGAGCATGTAGATCATAT TGTGGTGCTGCCGCTTTATCCGCAATACTCCTGTTCAACGGTCGGTGCGGTATGGGATGAA CTGGCACGCATTCTGGCGCGCAAACGTAGCATTCCGGGGATATCGTTTATTCGTGATTACG CTGATAACCACGATTACATTAATGCACTGGCGAACAGCGTACGCGCTTCTTTTGCCAAACAT GGCGAACCGGATCTGCTGCTGCTCTCTTATCATGGCATTCCCCAGCGTTATGCAGATGAAG GCGATGATTACCCGCAACGTTGCCGCACAACGACTCGCGAACTGGCTTCCGCACTGGGGAT GGCACCGGAAAAAGTGATGATGACCTTTCAGTCGCGCTTTGGTCGGGAACCCTGGCTGATG CCTTATACCGACGAAACGCTGAAAATGCTCGGAGAAAAAGGCGTAGGTCATATACAGGTGA TGTGCCCGGGCTTTGCTGCGGATTGTCTGGAGACGCTGGAAGAGATTGCCGAGCAAAACC GTGAGGTCTTCCTCGGTGCCGGCGGGAAAAAATATGAATATATTCCAGCGCTTAATGCCAC GCCGGAACATATTGAAATGATGGCTAATCTTGTTGCCGCGTATCGCCACCATCACCATCACC AT**TGA**
DyP	**ATG**TCACAAGTGCAGTCTGGCATTCTTCCAGAGCACTGTCGTGCAGCCATCTGGATCGAGG CAAACGTGAAAGGGGATGTCAATGCACTGCGCGAGTGCTCAAAAGTCTTTGCAGATAAATTA GCTGGTTTCGAGGCACAGTTTCCAGACGCACATCTTGGAGCGGTCGTTGCCTTTGGACATG ACACATGGCGTGCTTTGTCGGGGGGGGTAGGGGCCGAGGAATTGAAGGATTTCACGCCTT ATGGCAAGGGTTTGGCTCCAGCCACTCAGTACGATGTCCTTATCCATATTTTATCTCTGCGT CACGACGTAAATTTTTCCGTCGCCCAAGCGGCCATGGCCGCTTTTGGGGATGCTGTGGAGG TGAAAGAAGAGATTCACGGCTTTCGCTGGGTCGAAGAACGTGACTTATCTGGGTTCGTAGAT GGGACAGAAAATCCAGCGGGGGAAGAAACACGCCGTGAAGTCGCCGTTATTAAAGATGGG GTTGACGCCGGAGGCAGTTACGTTTTTGTGCAACGCTGGGAGCATAACTTAAAACAACTTAA CCGTATGTCAGTTCACGATCAGGAAATGATGATCGGTCGTACAAAAGTTGCAAACGAAGAAA TTGATGGCGATGAGCGCCCAGAGACAAGCCACTTAACGCGTGTCGACCTTAAAGAGAACGG AAAAGGCTTAAAAATTGTTCGTCAGAGCTTGCCGTACGGTACTGCCAGCGGCACGCATGGA TTGTATTTCTGCGCGTATTGCGCGCGCTTGTATAATATCGAGCAGCAGTTGCTGTCCATGTT CGGTGACACGGACGGAAAACGTGATGCAATGCTTCGCTTCACCAAGCCTGTTACAGGGGGT TACTACTTTGCGCCGTCGTTAGATAAATTACTTGCCTTGCTCGAGCACCATCACCATCACCAT **TGA**
CYP119	**ATG**TACGATTGGTTTTCGGAGATGCGTAAAAAAGATCCTGTTTACTACGATGGTAATATTTGG CAAGTATTTTCATATCGCTACACCAAGGAGGTATTAAACAATTTTTCAAAATTTTCTTCCGATT TAACAGGGTACCACGAGCGCCTTGAGGACTTGCGCAATGGAAAGATCCGCTTTGACATCCC AACCCGCTATACCATGCTTACGTCGGATCCGCCGCTGCATGACGAATTGCGTAGTATGAGT GCCGATATCTTTTCTCCGCAAAAGTTGCAAACCCTGGAAACTTTTATTCGCGAAACGACCCG TAGTTTATTGGACTCGATTGACCCTCGCGAGGACGATATTGTTAAGAAGCTGGCCGTGCCC CTTCCAATTATCGTGATTTCCAAAATCCTGGGTCTGCCAATCGAGGACAAAGAGAAGTTCAA GGAATGGAGCGACCTTGTAGCGTTTCGTCTGGGCAAACCTGGCGAGATCTTCGAATTAGGA AAGAAGTATCTGGAATTGATCGGATATGTGAAAGACCATCTTAATTCAGGTACGGAAGTGGT CAGCCGCGTAGTTAATTCGAATCTTTCTGACATTGAAAAATTAGGATATATCATTTTATTATTG ATTGCCGGGAACGAGACGACTACTAATCTGATTAGCAATTCAGTTATTGACTTCACGCGCTT CAACTTGTGGCAACGCATTCGCGAAGAAAACCTTTACCTGAAAGCTATTGAAGAGGCTCTTC GCTATTCACCCCCGGTTATGCGCACAGTTCGTAAAACCAAGGAGCGTGTAAAGTTGGGCGA CCAGACGATTGAGGAAGGAGAATACGTGCGCGTCTGGATTGCATCAGCGAACCGCGACGA AGAGGTCTTTCACGACGGAGAAAAGTTCATTCCAGATCGTAACCCGAATCCTCATCTGAGTT TCGGGAGTGGGATCCACCTGTGCCTGGGGGCACCTTTGGCACGTTTGGAAGCCCGTATTG CCATCGAGGAGTTCTCGAAGCGTTTTCGTCACATCGAAATTCTTGACACAGAAAAGGTGCCA AACGAAGTGTTGAACGGGTATAAGCGTTTGGTAGTTCGTCTGAAGAGCAATGAGCTCGAGC ACCATCACCATCACCAT**TGA**

Table 3.6. Kinetic parameters of protoporphyrin IX ferrochelatases.

Organism	K_M^{PPIX} (μM)	K_M^{Fe} (μM)	k_{cat} (min^{-1})	reference
E. coli	1.1 ± 0.1	0.38 ± 0.03	25 ± 1	This study
M. xanthus	-	6.5 ± 2	2 ± 0.4	[72]
P. putida	-	20 ± 4		[72]
C. crescentus	-	11 ± 2		[72]
B. bacteriovorus	-	28 ± 3		[72]
S. cerevisiae	-	0.5 ± 0.1	45 ± 1	[60]
M. musculus	-	0.79 ± 0.2	6.6 ± 0.3	[60]
	1.4 ± 0.2	1.9 ± 0.3	4.1 ± 0.3	[73]
*H. sapiens**	9	9.3	6.6	[18]
B. taurus[§]	11	80	4.5[†]	[74]

* Error in reported parameters is less than or equal to 10%
[§]Error in kinetic parameters is not reported by the authors
[†]Reported value (105 nM heme mg^{-1} HemH min^{-1}) was converted using the reported molecular weight of active *B. taurus* HemH (43,133 g/mol).

Table 3.7 . Microscopic rate constants for EcHemH kinetic mechanism. Rate constants correspond to the model shown in Scheme 3.2.

	Co		Ni		Zn		Cu	
Parameter	Value	%CV	Value	%CV	Value	%CV	Value	%CV
k_1	1.0	3.2E+08	1.5	2.9E+08	0.84	207	1.00	3.4E+09
k_{-1}	2.3	2.0E+02	1.0E-07	4.4E+15	28	4153	0.96	1.7E+01
k_2	142	2.5E+02	0.20	2.2E+09	94	5866	28	1.2E+03
k_{-2}	158	2.3E+02	15	2.9E+07	12	9668	26	1.2E+03
k_3	0.9	1.8E+01	0.71	6.2E+08	9.2	2244	0.56	5.1E+01
$k_{1'}$	5.1	2.2E+02	1.0E-09	4.4E+17	0.24	770	12	3.7E+01
$k_{-1'}$	1,935	2.2E+02	1.0	4.2E+08	93	5443	1.8	1.4E+01
$k_{2'}$	68.4	1.0E+02	1.0	4.4E+08	27	4751	13	5.9E+02
$k_{-2'}$	1.7E-05	7.7E+03	0.95	9.8E+09	1.1	303	6.0	5.9E+02
$k_{3'}$	21	1.6E+01	0.044	9.1E+01	70	3517	2.8E-04	1.2E+03
$k_{-3'}$	0.0100	7.7E+03	0.0010	1.6E+05	0.091	1726	1.2E-05	1.2E+03

3. 8. References

1. Schwizer, F. *et al.* Artificial Metalloenzymes: Reaction Scope and Optimization Strategies. *Chemical Reviews* vol. 118 142–231 (2018).

2. Hartwig, J. F. & Ward, T. R. New “Cats” in the House: Chemistry Meets Biology in Artificial Metalloenzymes and Repurposed Metalloenzymes. *Acc. Chem. Res.* (2019) doi:10.1021/ACS.ACCOUNTS.9B00154.

3. Sreenilayam, G., Moore, E. J., Steck, V. & Fasan, R. Metal Substitution Modulates the Reactivity and Extends the Reaction Scope of Myoglobin Carbene Transfer Catalysts. *Adv Synth Catal* **359**, 2076–2089 (2017).

4. 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).

5. Brandenberg, O. F., Fasan, R. & Arnold, F. H. Exploiting and engineering hemoproteins for abiological carbene and nitrene transfer reactions. *Curr Opin Biotechnol* **47**, 102–111 (2017).

6. Hoffman, B. M. & Petering, D. H. Coboglobins: Oxygen-Carrying Cobalt-Reconstituted Hemoglobin and Myoglobin. *Proc. Natl. Acad. Sci.* **67**, 637–643 (1970).

7. Wagner, G. C., Gunsalus, I. C., Wang, M.-Y. R. & Hoffman, B. M. Cobalt-substituted Cytochrome P-450$_{cam}$. *J. Biol. Chem.* **256**, 6266–6273 (1981).

8. Wang, M.-Y. R., Hoffman, B. M. & Hollenberg, P. F. Cobalt-substituted Horseradish Peroxidase*. *J. Biol. Chem.* **252**, 6268–6275 (1977).

9. Bruha, A. & Kincaid, J. R. Resonance Raman Studies of Dioxygen Adducts of Co-Substituted Heme Proteins and Model Compounds. Vibrationally Coupled Dioxygen and the Issue of Multiple Structures and Distal Side Hydrogen Bonding. *J. Am. Chem. Soc.* **110**, 6006–60014 (1988).

10. Woodward, J. J., Martin, N. I. & Marletta, M. A. An Escherichia coli expression-based method for heme substitution. *Nat. Methods* **4**, 43–45 (2007).

11. Bordeaux, M., Singh, R. & Fasan, R. Intramolecular C(sp3)H amination of arylsulfonyl azides with engineered and artificial myoglobin-based catalysts. *Bioorganic Med. Chem.* **22**, 5697–5704 (2014).

12. Liu, Z. *et al.* Assembly and Evolution of Artificial Metalloenzymes within E. coli Nissle 1917 for Enantioselective and Site-Selective Functionalization of C─H and C═C Bonds. *J. Am. Chem. Soc.* **144**, 883–890 (2022).

13. Perkins, L. J., Weaver, B. R., Buller, A. R. & Burstyn, J. N. De novo biosynthesis of a nonnatural cobalt porphyrin cofactor in E. coli and incorporation into hemoproteins. *Proc. Natl. Acad. Sci.* **118**, 2021 (2021).

14. Labbé, R. F., Vreman, H. J. & Stevenson, D. K. Zinc protoporphyrin: A metabolite with a mission. *Clin. Chem.* **45**, 2060–2072 (1999).

15. Majtan, T., Frerman, F. E. & Kraus, J. P. Effect of cobalt on Escherichia coli metabolism and metalloporphyrin formation. *BioMetals* **24**, 335–347 (2011).

16. Majtan, T., Freeman, K. M., Smith, A. T., Burstyn, J. N. & Kraus, J. P. Purification and characterization of cystathionine β-synthase bearing a cobalt protoporphyrin. *Arch. Biochem. Biophys.* **508**, 25–30 (2011).

17. Choi, K. R., Yu, H. E. & Lee, S. Y. Production of zinc protoporphyrin IX by metabolically engineered Escherichia coli. *Biotechnol. Bioeng.* **119**, 3319–3325 (2022).

18. Medlock, A., Swartz, L., Dailey, T. A., Dailey, H. A. & Lanzilotta, W. N. Substrate interactions with human ferrochelatase. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 1789–1793 (2007).

19. Gora, M., Grzybowska, E., Rytka, J. & Labbe-Bois, R. Probing the Active-site Residues in Saccharomyces cerevisiae Ferrochelatase by Directed Mutagenesis. *J. Biol. Chem.* **271**, 11810–11816 (1996).

20. Dailey, H. A. *et al.* Ferrochelatase at the millennium: Structures, mechanisms and [2Fe-2S] clusters. *Cell. Mol. Life Sci.* **57**, 1909–1926 (2000).

21. Sudhamsu, J. *et al.* Co-expression of ferrochelatase allows for complete heme incorporation into recombinant proteins produced in E. coli. *Protein Expr Purif* **73**, 78–82 (2010).

22. Miyamoto, K., Kanaya, S., Morikawa, K. & Inokuchi, H. Overproduction, Purification, and Characterization of Ferrochelatase from Escherichia coli1. *J. Biochem.* **115**, 545–551 (1994).

23. Hunter, G. A. & Ferreira, G. C. Metal ion coordination sites in ferrochelatase. *Coord. Chem. Rev.* **460**, 214464 (2022).

24. Kohno, H., Okuda, M., Furukawa, T., Tokunaga, R. & Taketani, S. Site-directed mutagenesis of human ferrochelatase: Identification of histidine-263 as a binding site for metal ions. *Biochim. Biophys. Acta (BBA)/Protein Struct. Mol.* **1209**, 95–100 (1994).

25. Davidson, R. E., Chesters, C. J. & Reid, J. D. Metal Ion Selectivity and Substrate Inhibition in the Metal Ion Chelation Catalyzed by Human Ferrochelatase *. *J. Biol. Chem.* **284**, 33795–33799 (2009).

26. McIntyre, N. R., Franco, R., Shelnutt, J. A. & Ferreira, G. C. Nickel(II) chelatase variants directly evolved from murine ferrochelatase: Porphyrin distortion and kinetic mechanism. *Biochemistry* **50**, 1535–1544 (2011).

27. Hunter, G. A., Sampson, M. P. & Ferreira, G. C. Metal ion substrate inhibition of ferrochelatase. *J. Biol. Chem.* **283**, 23685–23691 (2008).

28. Al-Karadaghi, S. *et al.* Chelatases: distort to select? *Trends Biochem. Sci.* **31**, 135–142 (2006).

29. Blackwood, M. E., Rush, T. S., Romesberg, F., Schultz, P. G. & Spiro, T. G. Alternative

Modes of Substrate Distortion in Enzyme and Antibody Catalyzed Ferrochelation Reactions. *Biochemistry* **37**, 779–782 (1998).

30. Franco, R., Ma, J.-G., Lu, Y., Ferreira, G. C. & Shelnutt, J. A. Porphyrin Interactions with Wild-type and Mutant Mouse Ferrochelatase. *Biochemistry* **39**, 2517–2529 (2000).

31. Medlock, A. E. *et al.* Insight into the function of active site residues in the catalytic mechanism of human ferrochelatase. (2021) doi:10.1042/BCJ20210460.

32. Gillam, M. E., Hunter, G. A. & Ferreira, G. C. Ferrochelatase π-helix: Implications from examining the role of the conserved π-helix glutamates in porphyrin metalation and product release. *Arch. Biochem. Biophys.* **644**, 37–46 (2018).

33. Medlock, A. E. *et al.* Insight into the function of active site residues in the catalytic mechanism of human ferrochelatase. *Biochem. J.* **478**, 3239–3252 (2021).

34. Sellers, V. M., Wu, C. K., Dailey, T. A. & Dailey, H. A. Human ferrochelatase: Characterization of substrate - Iron binding and proton-abstracting residues. *Biochemistry* **40**, 9821–9827 (2001).

35. Brindley, A. A., Raux, E., Leech, H. K., Schubert, H. L. & Warren, M. J. A story of chelatase evolution: Identification and characterization of a small 13-15-kDa 'ancestral' cobaltochelatase (CbiXs) in the archaea. *J. Biol. Chem.* **278**, 22388–22395 (2003).

36. Dailey, H. A. *et al.* Ferrochelatase at the millennium: structures, mechanisms and [2Fe-2S] clusters. *Cell. Mol. Life Sci.* **57**, 1909–1926 (2000).

37. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: A Sequence Logo Generator: Figure 1. *Genome Res.* **14**, 1188–1190 (2004).

38. Arnold, F. H. Design by Directed Evolution. *Acc. Chem. Res.* **31**, 125–131 (1998).

39. Herger, M. *et al.* Synthesis of β-Branched Tryptophan Analogues Using an Engineered Subunit of Tryptophan Synthase. *J. Am. Chem. Soc.* **138**, 8388–8391 (2016).

40. Ellis, J. M. *et al.* Biocatalytic synthesis of non-standard amino acids by a decarboxylative aldol reaction. *Nat. Catal.* **5**, 136–143 (2022).

41. Weeks, A. M. & Wells, J. A. Engineering peptide ligase specificity by proteomic identification of ligation sites. *Nat. Chem. Biol.* **14**, 50–57 (2018).

42. Mcdonald, A. D., Higgins, P. M. & Buller, A. R. Substrate multiplexed protein engineering facilitates promiscuous biocatalytic synthesis. (2021) doi:10.26434/CHEMRXIV-2021-XL252.

43. McDonald, A. D., Bruffy, S. K., Kasat, A. T. & Buller, A. R. Engineering Enzyme Substrate Scope Complementarity for Promiscuous Cascade Synthesis of 1,2-Amino Alcohols. *Angew. Chemie Int. Ed.* **61**, e202212637 (2022).

44. Sellers, V. M., Wu, C.-K., Dailey, T. A. & Dailey, H. A. Human Ferrochelatase: Characterization of Substrate−Iron Binding and Proton-Abstracting Residues. *Biochemistry*

40, 9821–9827 (2001).

45. Medlock, A. E., Dailey, T. A., Ross, T. A., Dailey, H. A. & Lanzilotta, W. N. A π-Helix Switch Selective for Porphyrin Deprotonation and Product Release in Human Ferrochelatase. *J. Mol. Biol.* **373**, 1006–1016 (2007).

46. Hunter, G. A. & Ferreira, G. C. Molecular enzymology of 5-aminolevulinate synthase, the gatekeeper of heme biosynthesis. *Biochim. Biophys. Acta* **1814**, 1467–73 (2011).

47. Woodard, S. I. & Dailey, H. A. Regulation of heme biosynthesis in Escherichia coli. *Arch. Biochem. Biophys.* **316**, 110–115 (1995).

48. McNicholas, P. M., Javor, G., Darie, S. & Gunsalus, R. P. Expression of the heme biosynthetic pathway genes hemCD, hemH, hemM and hemA of Escherichia coli. *FEMS Microbiol. Lett.* **146**, 143–148 (2006).

49. Foster, A. W. *et al.* Metalation calculators for E. coli strain JM109 (DE3): aerobic, anaerobic, and hydrogen peroxide exposed cells cultured in LB media. *Metallomics* **14**, (2022).

50. Studier, F. W., Daegelen, P., Lenski, R. E., Maslov, S. & Kim, J. F. Understanding the Differences between Genome Sequences of Escherichia coli B Strains REL606 and BL21(DE3) and Comparison of the E. coli B and K-12 Genomes. *J. Mol. Biol.* **394**, 653–680 (2009).

51. Rodrigue, A., Effantin, G. & Mandrand-Berthelot, M.-A. Identification of rcnA (yohM), A Nickel and Cobalt Resistance Gene in Escherichia coli. *J. Bacteriol.* **187**, 2912–2916 (2005).

52. Blériot, C., Effantin, G., Lagarde, F., Mandrand-Berthelot, M.-A. & Rodrigue, A. RcnB Is a Periplasmic Protein Essential for Maintaining Intracellular Ni and Co Concentrations in Escherichia coli. *J. Bacteriol.* **193**, 3785–3793 (2011).

53. Suits, M. D. L., Jaffer, N. & Jia, Z. Structure of the Escherichia coli O157:H7 Heme Oxygenase ChuS in Complex with Heme and Enzymatic Inactivation by Mutation of the Heme Coordinating Residue His-193. *J. Biol. Chem.* **281**, 36776–36782 (2006).

54. Suits, M. D. L. *et al.* Identification of an Escherichia coli O157:H7 heme oxygenase with tandem functional repeats. *Proc. Natl. Acad. Sci.* **102**, 16955–16960 (2005).

55. Yoshinaga, T., Sassa, S. & Kappas, A. Purification and properties of bovine spleen heme oxygenase. Amino acid composition and sites of action of inhibitors of heme oxidation. *J. Biol. Chem.* **257**, 7778–7785 (1982).

56. Lee, W. C., Reniere, M. L., Skaar, E. P. & Murphy, M. E. P. Ruffling of Metalloporphyrins Bound to IsdG and IsdI, Two Heme-degrading Enzymes in Staphylococcus aureus. *J. Biol. Chem.* **283**, 30957–30963 (2008).

57. Létoffé, S., Heuck, G., Delepelaire, P., Lange, N. & Wandersman, C. Bacteria capture iron from heme by keeping tetrapyrrol skeleton intact. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 11719–11724 (2009).

58. Dailey, H. A. *et al.* The Escherichia coli Protein YfeX Functions as a Porphyrinogen Oxidase, Not a Heme Dechelatase. *MBio* **2**, (2011).

59. Bloomer, B. J., Clark, D. S. & Hartwig, J. F. Progress, Challenges, and Opportunities with Artificial Metalloenzymes in Biosynthesis. *Biochemistry* **62**, 221–228 (2023).

60. Hunter, G. A., Sampson, M. P. & Ferreira, G. C. Metal ion substrate inhibition of ferrochelatase. *J. Biol. Chem.* **283**, 23685–23691 (2008).

61. Kuzmič, P. Program DYNAFIT for the Analysis of Enzyme Kinetic Data: Application to HIV Proteinase. *Anal. Biochem.* **237**, 260–273 (1996).

62. Johnson, K. A. *Kinetic Analysis for the New Enzymology*. (Kintek Corporation, 2019).

63. Blum, M. *et al.* The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **49**, D344–D354 (2021).

64. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

65. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152 (2012).

66. Edgar, R. C. MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping. *bioRxiv* 2021.06.20.449169 (2021) doi:10.1101/2021.06.20.449169.

67. Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100 (1990).

68. Kille, S. *et al.* Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. *ACS Synth. Biol.* **2**, 83–92 (2013).

69. Scolaro, L. M. *et al.* Aggregation Behavior of Protoporphyrin IX in Aqueous Solutions: Clear Evidence of Vesicle Formation. *J. Phys. Chem. B* **106**, 2453–2459 (2002).

70. Davidson, R. E., Chesters, C. J. & Reid, J. D. Metal ion selectivity and substrate inhibition in the metal ion chelation catalyzed by human ferrochelatase. *J. Biol. Chem.* **284**, 33795–33799 (2009).

71. Kille, S. *et al.* Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis. (2012) doi:10.1021/sb300037w.

72. Shepherd, M., Dailey, T. A. & Dailey, H. A. A new class of [2Fe-2S]-cluster-containing protoporphyrin (IX) ferrochelatases. *Biochem. J.* **397**, 47–52 (2006).

73. Shi, Z. & Ferreira, G. C. Probing the Active Site Loop Motif of Murine Ferrochelatase by Random Mutagenesis. *J. Biol. Chem.* **279**, 19977–19986 (2004).

74. Dailey, H. A. & Fleming, J. E. Bovine ferrochelatase. Kinetic analysis of inhibition by N-methylprotoporphyrin, manganese, and heme. *J. Biol. Chem.* **258**, 11453–11459 (1983).

APPENDIX 1 : DynaFit4 Scripts

Determine the kinetic mechanism for the inhibition of HemH
from /Escherichia coli/ by cobalt.

```
;_____________________________________________________________________________
[task]
  task   = fit
  data   = rates
  approx = none
  model  = Co inihibition ?

[mechanism]


        E + PPIX <===> E.PPIX          :  k1 k-1
      E.PPIX + Co <===> E.PPIX.Co      :  k2 k-2
       E.PPIX.Co ----> E + CoPPIX      :  k3
      E.PPIX + Co <===> E.PPIX.Co'     :  k2' k-2'
         E + Co <===> E.Co'           :  k1' k-1'
     E.Co' + PPIX <===> E.PPIX.Co'     :  k3' k-3'



[constants]
k-1 = 100 ?
k1 = 1 ?
k2 = 1000 ?
k-2 = 10 ?
k3 = .01 ?
k3' = 10 ?
k-3' = 1 ?
k1' = 100 ?
k-1' = 1000 ?
k2' = 100 ?
k-2' = 1 ?

k1 = (k-1 k1' k2' k-3')/(k-1' k3' k-2')


[responses]
  CoPPIX = 1
[concentrations]
  E = 0.03
[data]
mesh logarithmic from 100 to 0.1 step 0.9
  delay = 0.1
  variable  Co

  directory C:\DynaFit_Data\Cobalt
  sheet     HemH_Co_initial_rates.csv
  column   2 | conc PPIX = 0.3 | label PPIX=0.3
```

```
   column   3 | conc PPIX = 0.5 | label PPIX=0.5
   column   4 | conc PPIX = 1 | label PPIX=1
   column   5 | conc PPIX = 2 | label PPIX=2
   column   6 | conc PPIX = 5 | label PPIX=5
   column   7 | conc PPIX = 10 | label PPIX=10



[output]

   directory C:\DynaFit_Data\Cobalt\publication_branch
[settings]

{Output}
   XAxisLabel = [Co2+] (uM)
   YAxisLabel = Rate (uM/s)
;_____________________________________________________________________________
;_____________________________________________________________________________
```

Determine the kinetic mechanism for the inhibition of HemH
from /Escherichia coli/ by copper.
;_____

```
[task]
  task   = fit
  data   = progress
  approx = none
  model  = Noncompetitive substrate inhibition ?

[mechanism]

              E + PPIX <===> E.PPIX          :  k1 k-1
            E.PPIX + Cu <===> E.PPIX.Cu      :  k2 k-2
             E.PPIX.Cu ----> E + CuPPIX      :  k3
              E + Cu  <===> E.Cu'            :  k1' k-1'
        E.Cu' + PPIX <===> E.PPIX.Cu'        :  k2' k-2'
           E.PPIX + Cu <===> E.PPIX.Cu'      :  k3' k-3'


[constants]
k-1 = 10 ?
k1 = 10 ?
k2 = 10 ?
k-2 = 10 ?
k3 = .01 ?
k1' = 10 ?
k-1' = 10 ?
k2' = 10 ?
k-2' = 10 ?
k3' = 10 ?
k-3' = 10 ?


k1 = (k-1 k1' k2' k-3')/(k-1' k3' k-2')



[responses]
  CuPPIX = 1
[concentrations]
   E = 0.7
[data]
directory C:\DynaFit_Data\Copper
sheet     Cu_progress_curves.txt
monitor   CuPPIX
column   2
offset = 0.8113 ?
conc PPIX = 10.01
conc Cu = 3.7 ?
label PPIX=10 uM
column   3
```

offset = 0.4381 ?
conc PPIX = 5.413
conc Cu = 3.7 ?
label PPIX=5 uM
column 4
offset = 0.6940 ?
conc PPIX = 8.46
conc Cu = 3.4 ?
label PPIX=8.5 uM
[output]
directory C:\DynaFit_Data\Copper\Cu_publication
[settings]
{Output}
XAxisLabel = time, sec
YAxisLabel = Absorbance

[end]

```
Determine the kinetic mechanism for the inhibition of HemH
from /Escherichia coli/ by Nickel.
;_____________________________________________________________________
[task]
  task   = fit
  data   = rates
  approx = none
  model  = Noncompetitive substrate inhibition ?
  algorithm = TR
[mechanism]
        E + PPIX <===> E.PPIX          :  k1 k-1
      E.PPIX + Ni <===> E.PPIX.Ni      :  k2 k-2
       E.PPIX.Ni ----> E + NiPPIX      :  k3
          E + Ni <===> E.Ni'           :  k1' k-1'
           E.Ni' <===> E.Ni.PPIX'      :  k2' k-2'
      E.PPIX + Ni <===> E.PPIX.Ni'     :  k3' k-3'


[constants]
k1 = 1 ?
k-1 = .1 ?
k2 = .1 ?
k-2 = 1 ?
k3 = 1 ?
k1' = .001 ?
k-1' = 1 ?
k2' = 1 ?
k-2' = 1 ?
k3' = .001 ?
k-3' = 1 ?



k2' = (k1 k3' k-2' k-1') / ( k-1 k1' k-3')


[responses]
  NiPPIX = 1
[concentrations]
  E = 0.05
[data]
mesh logarithmic from 600 to 0.1 step 0.9
  delay = 1
  variable  Ni
  directory C:\DynaFit_Data\Nickel
  sheet     Ni_PPIX_rates.txt
  column   2 | conc PPIX = 1.05 | label PPIX=1
  column   3 | conc PPIX = 2.16 | label PPIX=2
  column   4 | conc PPIX = 5.27 | label PPIX=5
  column   5 | conc PPIX = 10.06 | label PPIX=10
[output]
```

```
   directory C:\DynaFit_Data\Nickel\Ni_publication2
[settings]


{Output}
   XAxisLabel = [Ni2+] (uM)
   YAxisLabel = Rate (uM/s)
;_____
;_____
```

Determine the kinetic mechanism for the inhibition of HemH
from /Escherichia coli/ by Zinc
[task]
task = fit
data = rates
approx = none
model = uncompetitive substrate inhibition
[mechanism]

E + PPIX <===> E.PPIX : k1 k-1
E.PPIX + Zn <===> E.PPIX.Zn : k2 k-2
E.PPIX.Zn ----> E + ZnPPIX : k3
E + Zn <===> E.Zn' : k1' k-1'
E.Zn' + PPIX <===> E.PPIZ.Zn' : k2' k-2'
E.PPIX + Zn <===> E.PPIX.Zn' : k3' k-3'

[constants]

k-1 = 10 ?
k1 = 10 ?
k2 = 10 ?
k-2 = 10 ?
k3 = .01 ?
k1' = 10 ?
k-1' = 10 ?
k2' = 10 ?
k-2' = 10 ?
k3' = 10 ?
k-3' = 10 ?

[responses]
ZnPPIX = 1
[concentrations]
E = 0.03
[data]
delay = 1
variable Zn, PPIX
set data
[set:data]

Zn	PPIX	Rate
20	0.05	0.000467045
20	0.05	4.82E-04
20	0.1	0.000777406
20	0.1	0.000731378
20	0.2	0.001452966
20	0.2	0.001471632
20	0.3	0.002103889

20	0.3	0.001821095
20	0.5	0.002703953
20	0.5	0.002484367
20	0.8	0.003305859
20	0.8	0.003242277
20	1	0.003533282
20	1	0.003446924
20	2	0.003599312
20	2	0.00352358
20	3	0.003400631
20	3	0.003006768
20	5	0.002740046
20	5	0.002442869
0.1	5	0.005770774
0.1	5	0.005968224
0.1	5	0.006113869
0.2	5	0.006315869
0.2	5	0.007269114
0.2	5	0.006165693
0.5	5	0.006248175
0.5	5	0.008635706
0.5	5	0.006837577
1	5	0.005779358
1	5	0.007267302
1	5	0.004972117
2	5	0.005566715
2	5	0.006207255
2	5	0.005116248
5	5	0.004430949
5	5	0.005579616
5	5	0.00423638
10	5	0.003140861
10	5	0.004251516
10	5	0.003276263
20	5	0.002359124
20	5	0.003366296
20	5	0.002433781
50	5	0.001673679
50	5	0.002525246
50	5	0.00155362
100	5	0.001286194
100	5	0.001808478
100	5	0.001332258

[output]
directory C:\DynaFit_Data\Zinc\Zn_Publication
[settings]

{Output}
XAxisLabel = [PPIX] (uM)

YAxisLabel = Rate (uM/s)

Chapter 4

Engineering a cobalt-substituted P450 for metallohydrogen atom transfer (MHAT)

This research was conducted with experimental and intellectual contributions from Carly L. Masonheimer (C. L. M.)

Chapter 4: Engineering a cobalt-substituted P450 for metallohydrogen atom transfer

4. 1. Abstract

Metal hydrides are broadly useful intermediates in organic chemistry. In contrast, there are relatively few known examples of metal hydrides occurring naturally in the biological world. Those few metal hydrides that are known are quickly protonated to evolve hydrogen gas or react via two electron mechanisms for the reduction of ketones. Here, we develop the new-to-nature CoPPIX cofactor for metal hydrogen atom transfer (MHAT)-type reactivity in an enzyme. Mechanistic analysis supports the formation of bonafide metal-hydride intermediate which can react via hydrogen atom transfer. We evolve a cobalt substituted P450 (CoCYP119) for an MHAT-mediated *O*-deallylation reaction and discover several variants with improved reactivity. The electronic structures and catalytic behavior of these enzymes are distinct from those of the parent enzyme. These evolved catalysts open the door to exploring other MHAT-mediated enzymatic transformations.

4. 2. Introduction

Metal hydrides are important reactive intermediates in synthetic chemistry. Metal hydrides can exhibit protic, hydridic, or hydrogen atom-like character, depending on the oxidation state and identity of the metal, ligand environment, and solvent (Figure 4.1A).[1] Recently, metal-hydrogen atom transfer (MHAT) reactions have gained attention synthetic community, as these reactions enable a wide range of radical hydrofunctionalization and isomerization reactions with alkenes (Figure 4.1B).[2–7] Furthermore, these reactions tend to be highly functional group tolerant, and can be performed under mild conditions, adding to their attractiveness in a synthetic context.

Despite the ubiquity of metal hydrides in synthetic chemistry, there are relatively few examples of these intermediates in nature, and their reactivity is more limited. The hydrogenase enzyme is one notable example of a naturally occurring bimetallic hydride intermediate.[8] Additionally, the B_{12}-dependant photoreceptor CarH is proposed to access a hydridocobalamin intermediate following photolytic β-hydride-elimination of an adenosine cosubstrate.[9,10] It is hypothesized that the Co-H intermediate of CarH is rapidly protonated, or quenched with another equivalence of cobalt hydride, and spectroscopic evidence for this intermediate is sparse. The Bren lab has characterized several cytochrome c mimics that utilize a new-to-nature cobalt porphyrin cofactor (Figure 1.4C).[11–13] These catalysts have been demonstrated to perform hydrogen evolution in water, which is proposed to operate via protonation of a cobalt hydride. There also is limited evidence that the action of F_{430}-dependant Methyl-Coenzyme M reductase enzyme (MCR) involves nickel-hydride-like intermediates,[14] although this hypothesis has been refuted.[15] Recently, the Hartwig group and others have reported that a mononuclear zinc dependent enzymes can form a putative zinc-hydride intermediate (Figure 4.1C).[16,17] This non-native reactive intermediate subsequently reduces a range of ketone substrates enantioselectivity. There are notably no known examples of metal hydrides with hydrogen atom-like character found in Nature.

A biocatalytic MHAT reaction may have distinct advantages over traditional synthetic approaches. Biocatalysts are prized for their high activity and selectivity, and their ability to function under mild, aqueous conditions.[18] Many enzymes can be used in tandem, cascade settings, obviating the need for stepwise and inefficient intermediate purifications. Protein engineering and directed evolution enables the tuning of enzyme catalysts for specific reactions, and highly engineered enzymes have been shown to catalyze transformations that are not found in nature with exceedingly high rates and turnover numbers.[19,20] Lastly, the ability to make subtle and mechanism-agnostic changes to an enzyme's active site architecture via protein engineering fosters insight into new enzymatic reaction mechanisms.

However, formation of a metal hydride in the active site of an enzyme *en route* to MHAT presents unique challenges. The first of these is that there exists no known mononuclear metal cofactor for delivery of a hydrogen atom. The reactivity of metal hydrides (hydridic versus protic character) is heavily influenced by both the metal identity and its ligand environment. It is currently difficult to predict which environments best lend themselves to hydrogen atom transfer reactivity.[7] As described above, protonation of metal hydride intermediates in aqueous environments, is well known[11,21], though MHAT reactions are known to proceed cleanly in water.[22] Lastly, routes to access a metal hydride through a hydride donor, such as borohydride or reactive silane reagents, are also problematic under aqueous conditions as these substrates are subject to competing hydrolysis.[23]

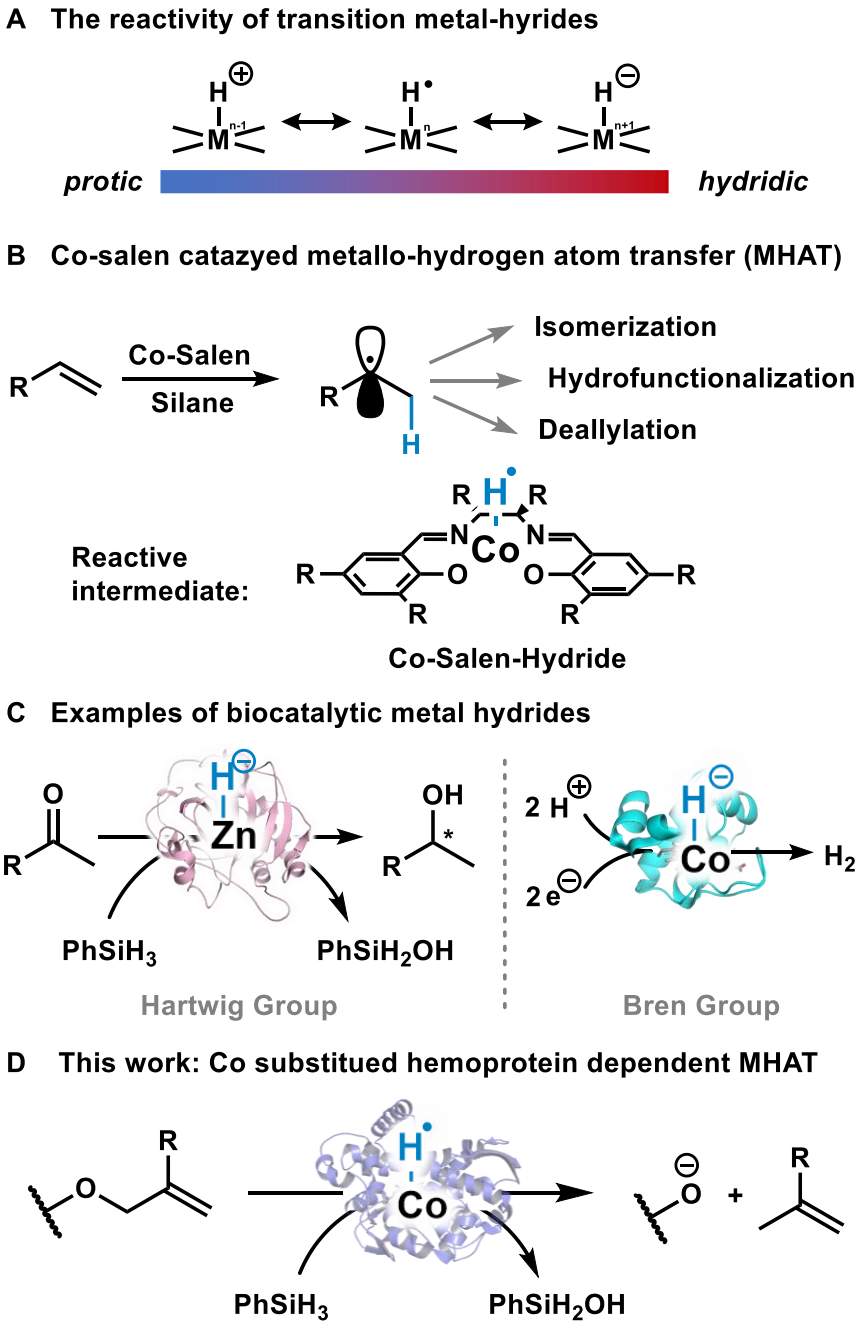


Figure 4.1. Examples of enzymatic and synthetic metal hydrides.

As no biological cofactor is known support MHAT in a natural enzyme, we envisioned that a new-to-nature cofactor might enable this chemistry. There are many examples of meta-stable metal hydride intermediates in synthetic chemistry from which to draw inspiration. The most commonly used MHAT catalysts are cobalt-salen complexes,[6,24,25] which feature a quasi-planar,

4-coordinate [L_2X_2]-type ligands. While salen complexes are not biologically derived, these ligand systems were originally inspired[26] by the tetrapyrrole class of cofactors, which are ubiquitous in nature. The most famous of these cofactors are heme and cobalamin (B_{12}). Intriguingly, cobalamin, which bears a [L_3X]-type corrin ligand structure, is known to form metal hydride intermediates that have protic behavior. Metal hydride complexes of heme, which features a [L_2X_2]-type porphyrin ligand) have not been studied, but the cobalt analog of heme (CoPPIX), has been shown to form metal hydrides with hydridic character, en route to hydrogen evolution.[12] We hypothesized that, in a suitable enzyme environment, the reactivity of CoPPIX could be tuned to react via hydrogen atom transfer. Indeed, the redox potential of cytochromes c and P450 hemoproteins has been shown to be highly tunable through mutation of the axial ligand and surrounding residues.[27–29] Importantly, our group's recent discovery of a method for producing cobalt-substituted hemoproteins *de novo* (see chapters 2 and 3) enables facile directed evolution of these artificial metalloproteins for MHAT reactivity. Development of these systems would broaden the scope of the biocatalytic toolbox to include a generalizable, new-to-nature, hydrogen atom transfer cofactor.

4. 3. Results

4. 3. 1. Design and "optimization" of a MHAT model reaction

We began by developing a probe for reactivity that would allow us to discover MHAT competent enzymes and study the reaction's mechanism. The MHAT reaction is typified by two key steps (Figure 4.2A). First, a metal hydride is formed either by hydride transfer to an oxidized metal species or protonation of a reduced metal species. Subsequently, hydrogen atom transfer of this metal hydride to an alkene substrate generates a carbon centered radical intermediate with Markovnikov selectivity. The various ways in which the carbon-centered radical can be trapped are diverse.[6]

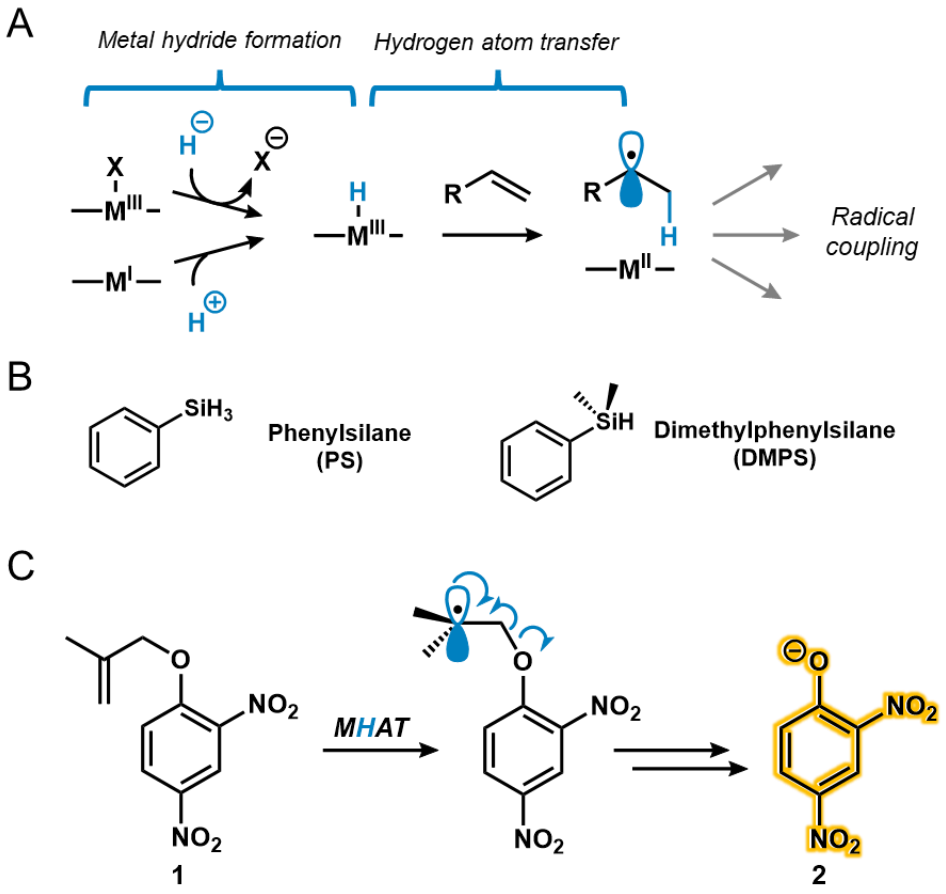


Figure 4.2. Design and implementation of a model MHAT reaction. A) Key steps of an MHAt reaction. B) Potential hydride donors for MHAT. C) Colorimetric detection of MHAT reactivity.

Formation of the metal hydride is the key step that would be shared among any potential MHAT transformation. Co-Salen rests at a Co(II) oxidation state, and metal hydride complexes of Co-Salen are routinely accessed by two different routes. Reduction of Co(II) to Co(I) yields a highly nucleophilic metal center, which can subsequently be protonated to form a cobalt hydride.[5,30–32] Conversely, Co(II) can be oxidized to Co(III), and a hydride donor, such as a silane reagent, can be used to generate the cobalt hydride.[2] Previous studies have shown that cobalt-substituted P450s rest at Co(III),[33] so we envisioned that addition of a hydride donor would be the most facile route to form biocatalytic cobalt hydride. Both dimethylphenylsilane (DMPS) and PS have been shown to be competent hydride donors for MHAT (Figure 4.2B).[6] PS is more hydridic but is also easily hydrolyzed in water.[23,34] DMPS hydrolyzes more slowly but its sterics render it a less reactive hydride donor. The Hartwig group previously used PS as a hydride donor for enzymatic reduction of ketones via a Zinc-hydride intermediate, but DMPS was not able to act as a hydride donor in this system.[16] The Shenvi group has also used phenylsilane in aqueous solvent as a hydride donor, although multiple equivalents of this reagent were required.[22]

Next, we considered the HAT step of the MHAT mechanism. Because we were primarily interested in hydrogen atom transfer (rather than hydride reduction), we chose an alkene substrate as a reactivity probe. HAT from the metal hydride to an alkene substrate would form a carbon centered radical on the substrate. To avoid complications in choosing a suitable coupling partner for this carbon radical, we chose to use a reactivity probe based on *O*-deallylation of a nitrophenol (Figure 4.2C). An MHAT-mediated deallylation mechanism was recently reported for the deprotection of aryl alcohols,[35] and the strong absorption of the product (a nitrophenolate) enables colorimetric detection. Use of the 2,4-dinitrophenol-derived substrate **1** enables detection at a wide range of pHs, as the deallylated product **2** has pKa of around 4.[36]

The P450 cytochrome 119 from *S. acidocaldarius* (CYP119) was chosen as an enzyme scaffold. This enzyme has been previously shown to be highly engineerable and thermostable, and can be used in a heat-treated lysate format.[37,38] Furthermore, we have previously shown that the cobalt-

substituted version of this enzyme (CoCYP119) can be easily obtained via heterologous expression in BL21(DE3) by supplementing the growth media with $CoCl_2$ (see Chapter 3).

We tested the ability of several catalysts for the deallylation of the allyl substrate (**2**), using DMPS as the hydride donor. The solubilities of both the allyl substrate and DMPS are poor in aqueous solvent. Addition of 20% DMSO solubilized both substrates (50 μM (**1**) and 200 μM DMPS) and allowed for spectral characterization of the reaction progress. The spectra of the substrate starting material (**1**) and the product, 2,4-dinitrophenolate **2**, are given in Figure 4.3. Transformation of **1** to **2** can be measured by tracking absorbance at 400nm (ε_{400}**(**1**)** = 9,600 M^{-1} cm^{-1}) We began by testing the reactivity of the bare cofactors, heme (FePPIX) and CoPPIX in aqueous buffer (KP_i, pH = 6.0) with 20% DMSO as a cosolvent. FePPIX showed no reactivity under these conditions, but we were delighted to see that reaction with CoPPIX yielded a modest 10 turnovers, and an initial rate of 0.14 min^{-1}. Next, we tested the activity of enzyme catalysts, both the native hemoprotein (FeCYP119) and the cobalt substituted version, CoCYP119. Again, the iron-dependent protein produced no nitrophenolate (**2**) under the conditions tested. CoCYP119, on the other hand, completed ~50 turnovers in 5 hours, with an initial rate approximately double of CoPPIX alone (0.27 min^{-1}), indicating that the enzyme environment alters the reactivity of a putative metal hydride intermediate.

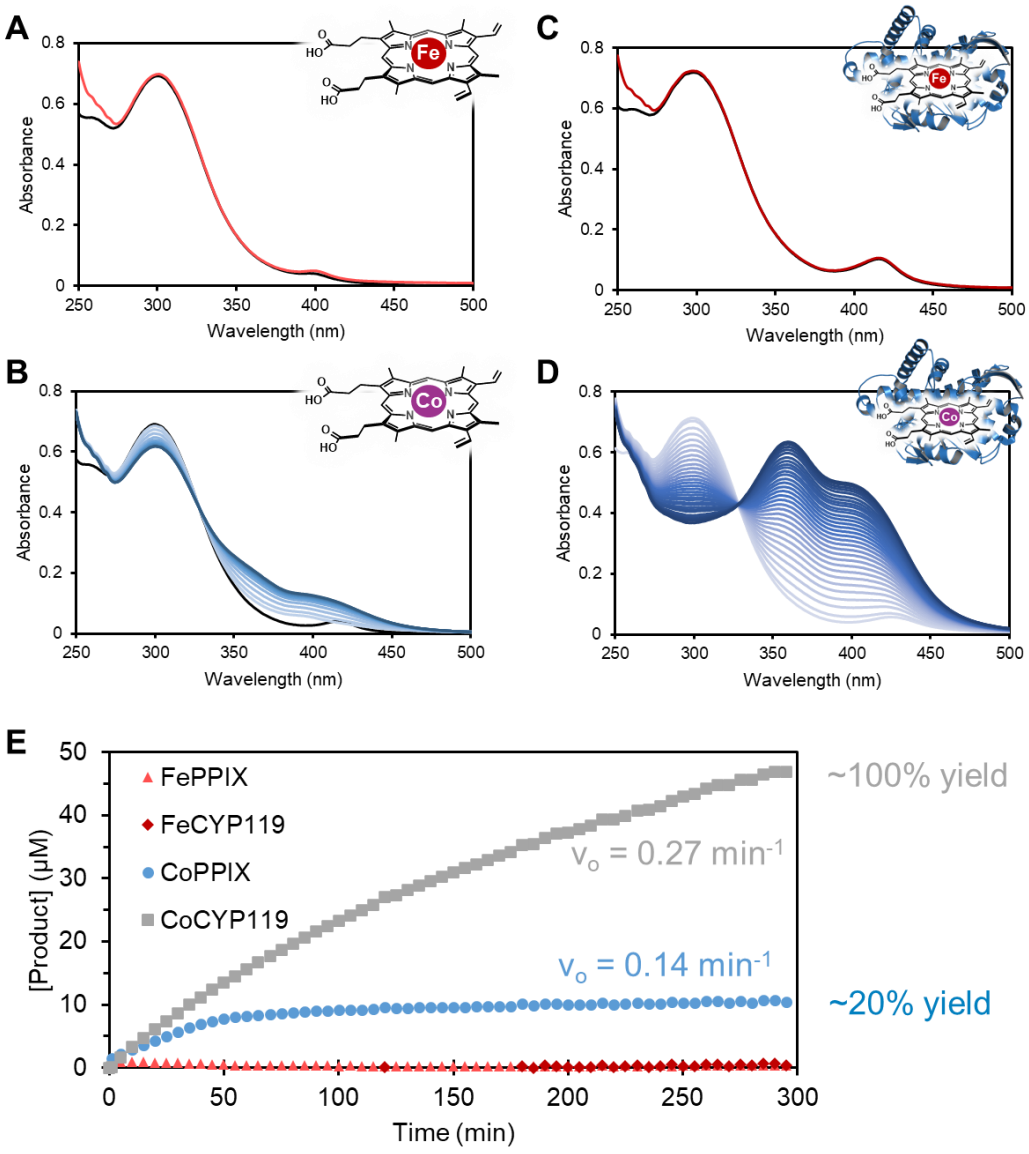


Figure 4.3. Spectroscopic measurements of deallylation of **1** by various catalytic systems. All reactions were conducted with 0.20 mM DMPS, 0.05 mM allyl substrate **1**, and 1.0 μM Catalyst in 100 mM KPi pH = 6.0 with 10% acetonitrile as the cosolvent. A) Reaction with FePPIX (heme). B) Reaction with FeCYP119. C) Reaction with CoPPIX. D) Reaction with CoCYP119. E) Reaction progress curves, measured by the absorbance at 400 nm.

We next tested the activity of CoCYP119 under a variety of reaction conditions (Table 4.1). Using PS or the hydride donor isopropoxyphenylsilane (IPS)[39] instead of DMPS dramatically improved the initial rate of the reaction, but reactions with PS and IPS went to overall lower yield (Table 4.1, entries 5, 7, and 9). Kinetic analysis of the PS reaction revealed that the lower yield

may be due to enzyme-independent depletion of PS, likely through hydrolysis (Figure 4.4). We tested the reaction with DMPS in several different buffers and found that the reaction proceeded similarly within the pH range 5.0-8.0, as measured by both initial rate and yield. Previous studies have shown pH = 6.0 is optimal for CYP119 epoxidation activity.[40] Analysis of the reaction progress curves in the first two hours shows that the initial velocity is maintained longer at higher pH, while the rate drops off quickly (within 30 min) under more acidic conditions.

Table 4.1. Reaction "Optimization" for deallylation of **1** by CoCYP119.

Entry	Buffer (pH)	Temp	Cosolvent	[Silane] (mM)	Silane	[1] (mM)	[Enzyme] (μM)	Yield (%)	v_0 (μM/min)
1	KPi (6.0)	25°C	20% DMSO	0.20	DMPS	0.05	1	100	0.27
2	KPi (6.0)	25°C	10% ACN	0.20	DMPS	0.05	1	100	0.16
3	Citrate (5.0)	25°C	10% ACN	0.20	DMPS	0.05	1	100	0.32
4	KPi (7.0)	25°C	10% ACN	0.20	DMPS	0.05	1	100	0.17
5	KPi (8.0)	25°C	10% ACN	0.20	DMPS	0.05	1	100	0.17
6	KPi (6.0)	25°C	20% DMSO	0.20	DMPS	0.10	2	80[a]	0.40
7	KPi (6.0)	50°C	20% DMSO	0.20	DMPS	0.10	2	73[a]	0.38
8	KPi (6.0)	25°C	20% DMSO	0.20	PS	0.10	2	36	5.3
9	KPi (6.0)	50°C	20% DMSO	0.20	PS	0.10	2	9	n.q.
10	KPi (6.0)	25°C	20% DMSO	0.20	IPS	0.10	2	29	4.3
11	KPi (6.0)	50°C	20% DMSO	0.20	IPS	0.10	2	19	3.5
12	KPi (6.0)	25°C	20% DMSO	0.20	PS	0.10	1	16	2.5

[a]Reaction had not reached completion, yield after 15 h.
n.q. = not quantified

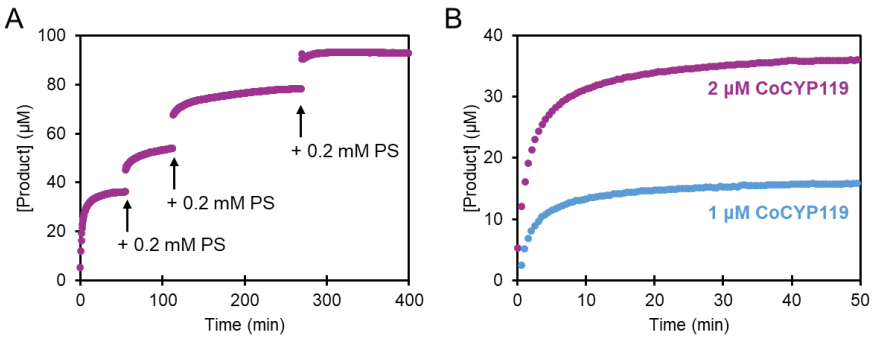


Figure 4.4. Evidence for non-enzymatic phenylsilane (PS) hydrolysis. A) Reaction yield is limited by PS. B) Doubling enzyme concentration exactly doubles the yield of the reaction with PS, indicating that the PS degradation is not enzyme dependent. Reaction conditions: 1 or 2 μM WT CoCYP119, 0.2 mM PS (added iteratively as shown), 0.1 mM allyl substrate **1**. Conditions for the purple traces reflect entry 8 in Table 4.1, conditions for the blue trace reflect entry 12.

Increasing the temperature to 50 °C had no effect on the initial rate of the deallylation reaction (Figure 4.5C-E, entries 6-11 in Table 4.1) but decreased the yield, presumably due to increased rate of background hydrolysis of the silane reagents. Lastly, both 10% acetonitrile and 20% DMSO were tolerated as cosolvents for the reaction. Acetonitrile more effectively solubilized the reagents, but the reaction was faster with 20% DMSO (compare entries 1 and 2 in Table 4.1).

We also tested the effect of proximal Co ligand on the reaction. The native proximal ligand for CoCYP119 (and all P450 enzymes) is Cys. We tested the C317S and C317A variants of CoCYP119, and a cobalt substituted dye decolorizing peroxidase (CoDyP) enzyme, whose native proximal ligand is His. None of the variants nor CoDyP had any deallylation activity, indicating that thiolate ligation is necessary for this reaction. Intriguingly, we found that CoCYP119 C317S purified in the Co(II) oxidation state, indicating that this variant may be useful for other new-to-nature reactions which require Co(II), such as carbene transfer.[41]

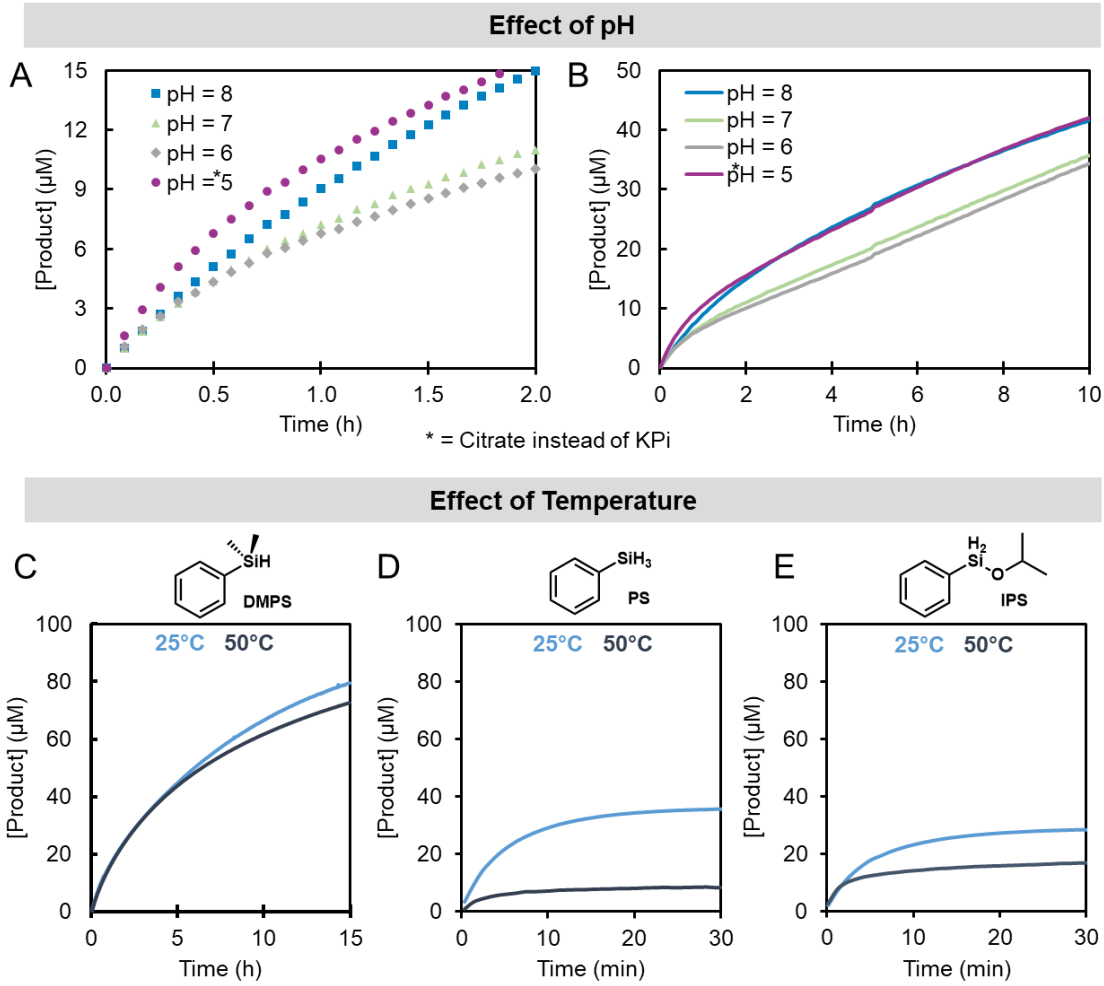


Figure 4.5. Effects of pH and Temperature on the deallylation of **1** by WT CoCYP119, as measured by absorbance at 400 nm as a function of time. A) Reaction progress for the first two hours of the reaction under different pH's. the Reaction at pH = 5.0 was conducted in citrate buffer, rather than KPi. Conditions for these reactions can be found in entries 2-4 in Table 4.1. B) 10 h progress curves for reaction in panel A. The yields recorded in Table 4.1 were taken after 24 h. C) Reaction progress curve for reaction with dimethylphenylsilane at 50°C and 25°C. Conditions can be found in entries 6 and 7 of Table 4.1. D) Reaction progress curve for reaction with phenylsilane at 50°C and 25°C. Conditions can be found in entries 8 and 9 of Table 4.1. E) Reaction progress curve for reaction with isopropoxyphenylsilane at 50°C and 25°C. Conditions can be found in entries 10 and 11 of Table 4.1

4. 3. 2. Spectroscopic evidence for a cobalt-hydride intermediate

Encouraged that the CoCYP119 scaffold had potential as an MHAT catalyst, we next turned our attention towards characterizing the putative Co-H intermediate. Weak metal hydrides, such as those of cobalt, are notoriously challenging to characterize, as they are only meta-stable.[7]

We were interested in understanding how the enzyme might be able to tune the reactivity of such an intermediate.

We began by using electronic absorption (UV-visible) spectroscopy. The UV-visible spectrum of CoCYP119 in the absence of a hydride donor is shown in Figure 4.6. The Soret maximal absorbance at 420 nm is typical of that for cobalt-substituted hemoproteins, and is indicative of a 6-coordinate, thiolate/hydroxide-ligated Co(III) center.[42] Upon reduction with dithionite, the Soret undergoes a characteristic blue shift, indicative of a Co(II) oxidation state (Figure 4.6A). This blue shift mirrors observations with the free cofactor (Figure 4.6B). These shifts to the CoPPIX absorbance are recapitulated by the addition of both PS and DMPS to the free cofactor, indicating rapid reduction of these species to Co(II). The spectral changes which occur upon addition of PS to Co(III)CYP119, however, are distinct from those of the putative Co(II) species (Figure 4.6C). Following addition of PS, the Soret absorbance is diminished, and two new peaks transiently appear at 385 and 430 nm. These two peaks are reminiscent of “hyperporphyrin” spectra, which result from interactions with hemoproteins (and their cobalt analogs) with strong distal donor ligands.[43,44] The Q-band features between 500-600 nm similarly lose intensity. Isosbestic points near 450 and 395 nm indicate clean interconversion between two distinct electronic states. The addition of DMPS (0.5 mM, the limit of solubility for this substrate) yields similar features, albeit at lower intensity (Figure 4.6E).

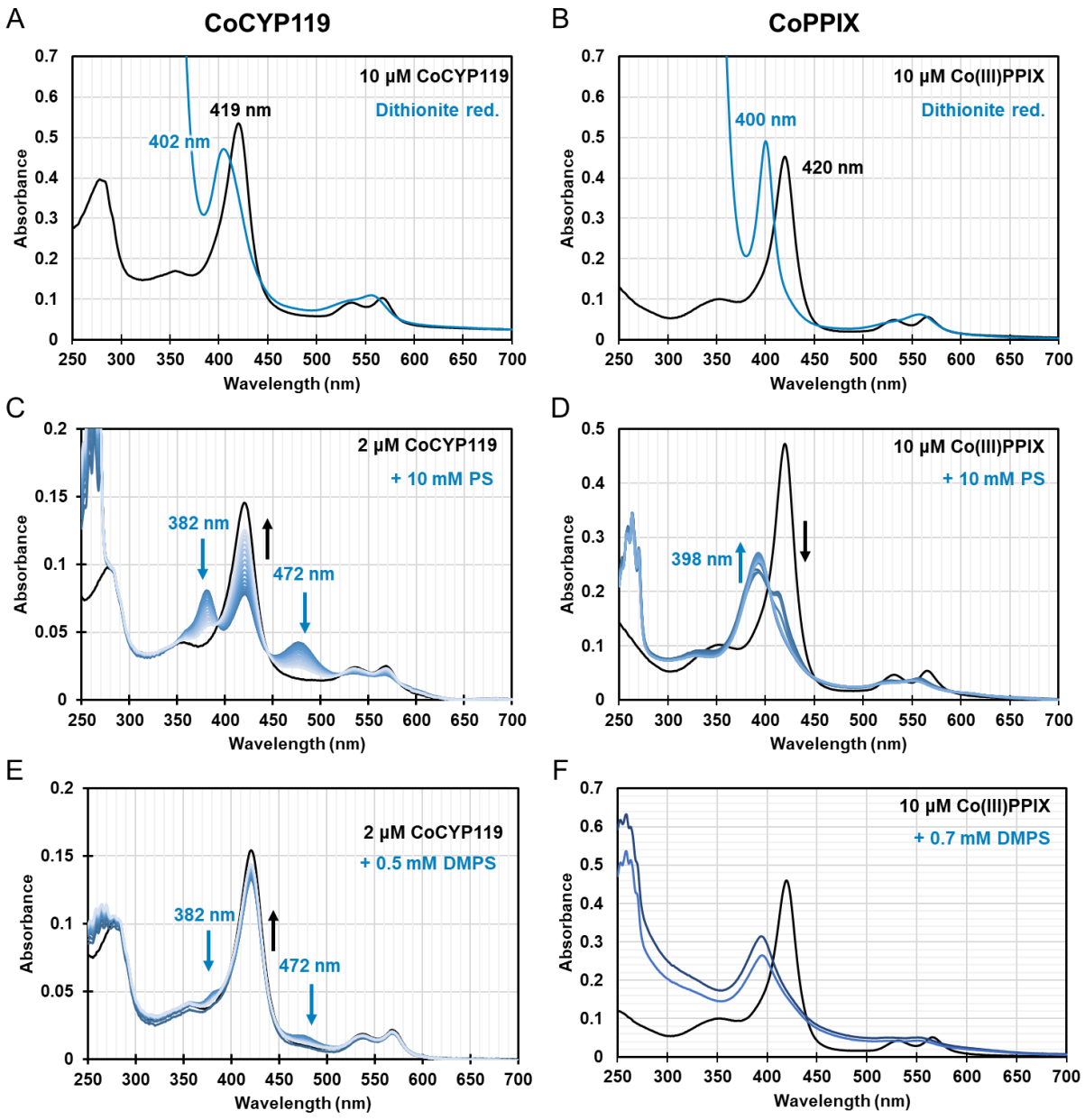


Figure 4.6. Changes to the electronic absorption spectra of CoCYP119 and CoPPIX upon addition of various reductants. The as-purified spectra are represented by the black trace, and changes over time upon addition of reductant are represented by the blue traces, from dark to light.

Under these conditions, these putative Co-H intermediates were transient with a lifetime of around 20 minutes, which was dependent on temperature, pH, and the concentration of CoCYP119 (Figure 4.7). The CoCYP119 dependence suggested that depletion of PS (and of the new Co intermediate) may be enzyme dependent. Puzzlingly, the decay of the putative Co-H intermediate features was faster at pH 8 than pH 6. This result may reflect the faster rate of non-

enzymatic hydrolysis of phenylsilane under more alkaline conditions.[23] The intermediate was also shorter lived at elevated temperatures, which matches previous observations of phenylsilane hydrolysis at 50°C (Figure 4.5D)

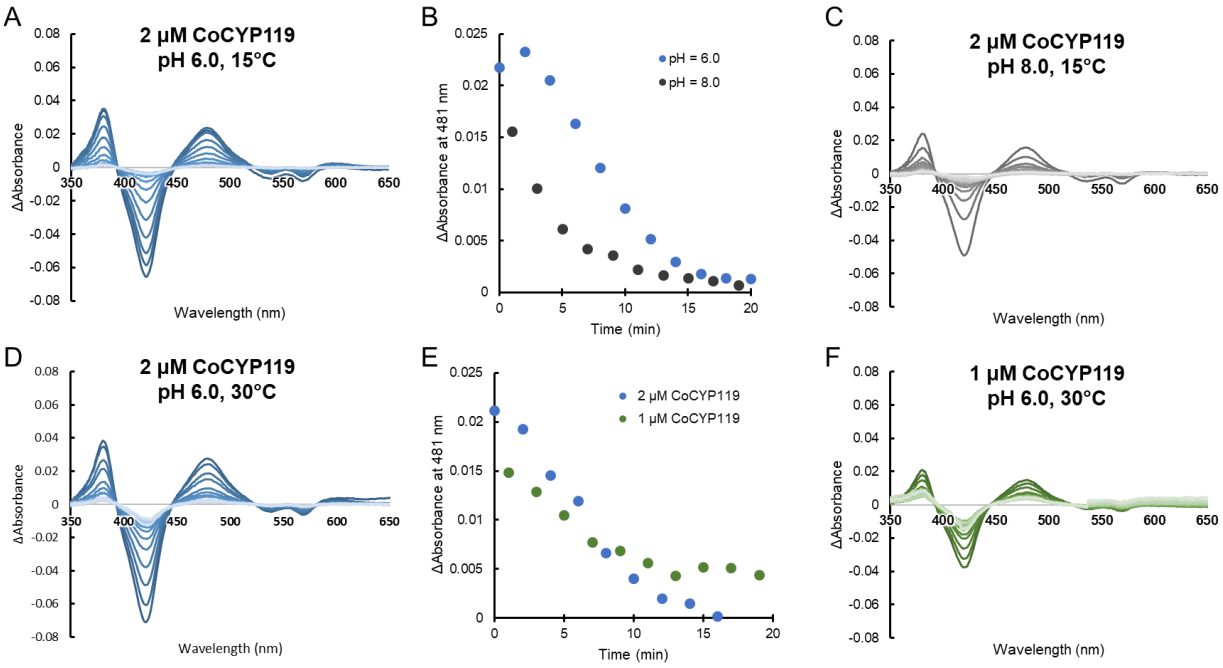


Figure 4.7. Effect of temperature, pH, and enzyme concentration on the behavior of a putative Co-hydride intermediate. A) Difference spectra for addition of 10 mM PS to 2 µM CoCYP119 at 15°C in KPi pH 6.0. B) Changes to the absorbance at 481 nm over time for conditions described in panels A and C. The intermediate appears to be shorter-lived at pH 8.0 than pH 6.0 C) Difference spectra for addition of 10 mM PS to 2 µM CoCYP119 at 15°C in KPi pH 8.0.D) Difference spectra for addition of 10 mM PS to 2 µM CoCYP119 at 30°C in KPi pH 6.0. E) Changes to the absorbance at 481 nm over time for conditions described in panels D and E. The intermediate has lower peak intensity, but decays more slowly with lower enzyme concentration. F) Difference spectra for addition of 10 mM PS to 1 µM CoCYP119 at 30°C in KPi pH 6.0.

Magnetic circular dichroism (MCD) spectroscopy enables additional interrogations the electronic environment of the cofactor. This work is ongoing and is in collaboration with Ryan Hall (R. H.) under the supervision of Prof. Thomas Brunold at UW-Madison. Spectral features analogous to those in in the UV-vis spectrum are observable in the MCD spectrum (Figure 4.8A). The 472 nm UV-vis feature appears near 2300 cm^{-1}, and the 381 nm feature appears around 2650 cm^{-1}. The intensity of these features do not change with respect to temperature, indicating

that the species is diamagnetic (Figure 4.8B). Unfortunately, this species is not highly abundant and the corresponding signals are dominated by the resting state Co(III) absorbance. Identification of conditions which promote a higher population of the putative cobalt hydride state would enable more in-depth characterization of these signals.

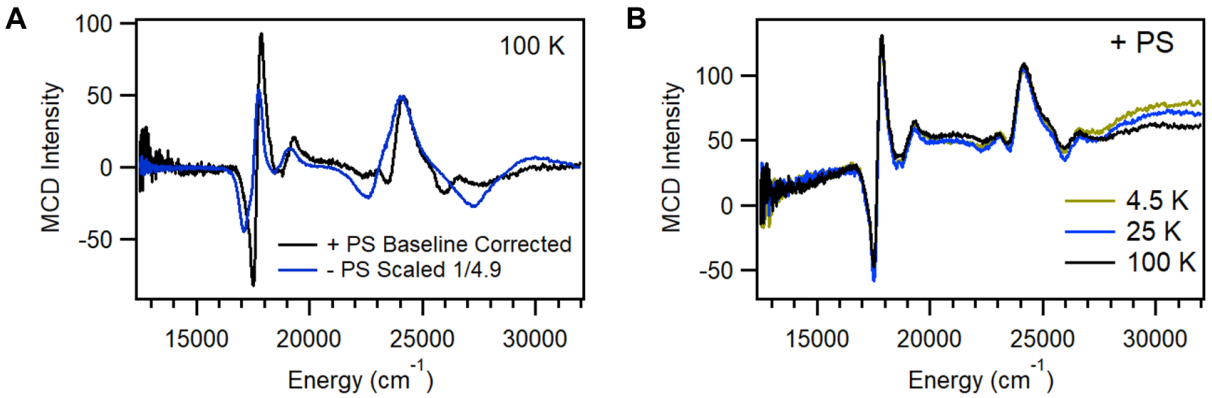


Figure 4.8. Magnetic circular dichroism (MCD) spectra of putative cobalt hydride intermediate. A)Comparison of CoCYP119 MCS spectra in the presence (black) and absence (blue) of 2 mM Phenylsilane (PS) in 100 mM KPi pH = 6.0 with 20% DMSO and 90% sucrose (w/v) as a glassing agent. Samples were flash frozen in liquid nitrogen to glass, immediately following assembly of the MCD cell. B) MCD spectra of CoCYP119 with 2 mM PS added. The satellite features at 2300 and 2600 cm^{-1} do not change with temperature, indicating that the putative cobalt hydride intermediate is diamagnetic.

DFT simulations of the putative cobalt hydride electronic absorption spectra are also ongoing (Figure 4.9). Initial findings suggest that the electronic structure of the intermediate is best described by a 5-coordinate, protonated Co(I) species, in which the proximal thiolate ligand is no longer coordinated to the cobalt center (Figure 4.9B, light blue). DFT simulated spectra of this complex yields low-absorbance features near 380 and 440 nm, closely matching the experimental spectra. Notably, the simulated 6-coordinate thiolate/hydride Co(III) complex gives only a single absorbance feature which is red-shifted from the resting state Soret. Investigations of the electronic absorption spectra in the IR region (beyond 800 nm) may help confirm or refute these hypotheses.

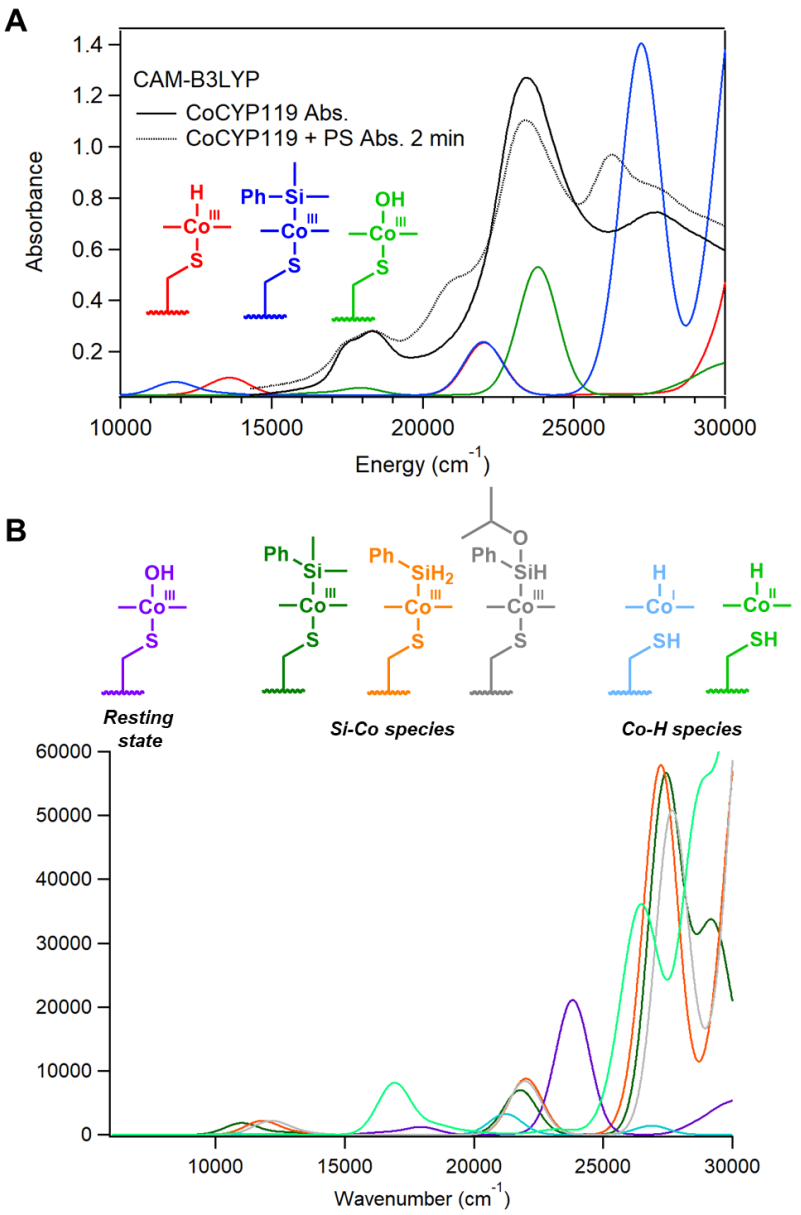


Figure 4.9. Experimental and Time-dependent density functional theory (TD-DFT) calculated electronic absorption spectra for enzyme resting state and potential intermediates which form upon addition of phenylsilane and substrate analogs. A) Experimental spectra (grey and black) superimposed with calculated electronic absorption spectra. The modeled structures are depicted in corresponding colors to the spectra. B) Additional modeled spectra. The Co(I)-proton calculated spectra (light blue) matches well with the observed spectra.

4. 3. 3. Inhibition of CoCYP119 deallylation activity by imidazole

Nitrogenous heterocycles, such as pyridine and imidazole, are known inhibitors of CYP119. These inhibitors act as ligands and occupy that distal ligand site of heme, preventing O_2 binding and catalysis. We measured the rate of deallylation by WT CoCYP119 in the presence of absence of imidazole, and found a significant decrease in rate (Figure 4.10A). This result indicates that catalysis is dependent on direct coordination to the cobalt center, and not through some other, secondary sphere interaction. Additionally, as imidazole is commonly used in the elution buffer for purifying His-tagged proteins via metal affinity chromatography, this study illustrates the importance of thorough dialysis of these proteins to remove excess imidazole. The apparent non-first order dependence on catalyst[45] depicted in Figure 4.10B can be attributed to residual imidazole in enzyme stock solution following two rounds of 1:100 dialysis. Four rounds of dialysis appears to sufficiently remove excess imidazole, as depicted by the return to first order catalyst dependence in Figure 4.10C.

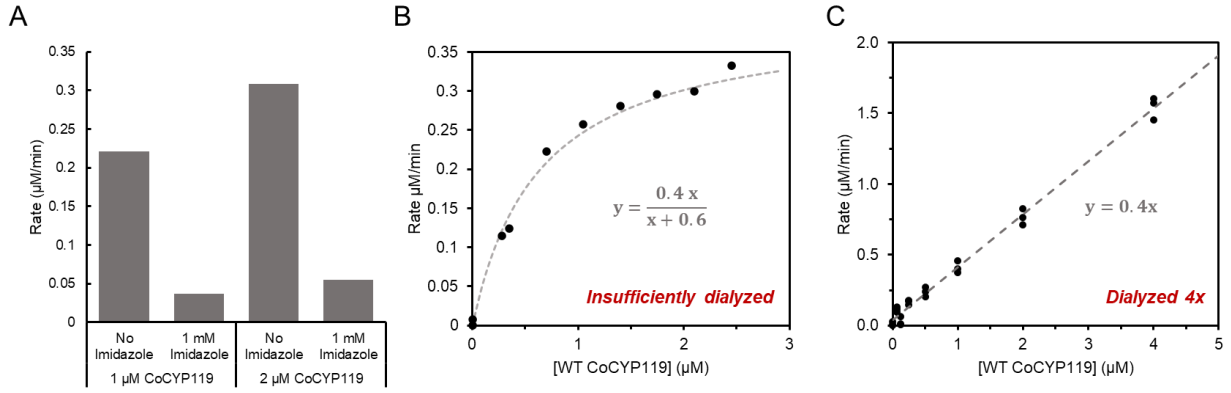


Figure 4.10. Effect of imidazole on the rate of deallylation of **1** by WT CoCYP119. A) Effect of adding 1 mM imidazole on the initial rate of the deallylation reaction. B) Initial rate of deallylation versus concentration of enzyme added. Not accounted for in this plot is the concentration of imidazole added proportionally to enzyme as a consequence of insufficiently dialyzed enzyme. Reaction conditions: 50 μM **1**, 200 μM DMPS in KPi pH 6.0 with 10% acetonitrile. The appropriate concentration of enzyme was added from a stock that was 70 μM to 1.0 mL total reaction volume. Black dots represent the experimental data points and the grey dotted line represents the best fit curve described by the equation in grey. C) Initial rate of deallylation versus concentration of enzyme added. Reaction conditions: 50 μM **1**, 200 μM DMPS in KPi pH 6.0 with 10% acetonitrile. The appropriate concentration of enzyme was added from a stock that was 100 μM to 1.0 mL total reaction volume. Black dots represent the experimental data points and the grey dotted line represents the best fit line described by the equation in grey.

4. 3. 4. Preliminary mechanistic investigations of MHAT deallylation

The proposed mechanism for the deallylation of **1** by CoCYP119 is given in Figure 4.11. Spectroscopic evidence in combination with DFT calculations suggest that the resting state of the catalyst is a 6-coordinate Co(III) species, with Cys317 thiolate as a proximal ligand and hydroxide bound in the distal site. A silane substrate enters the active site of the enzyme, and ligand exchange with the hydroxide distal ligand yields a cobalt hydride intermediate, as detected by UV-Visible and MCD spectroscopies.

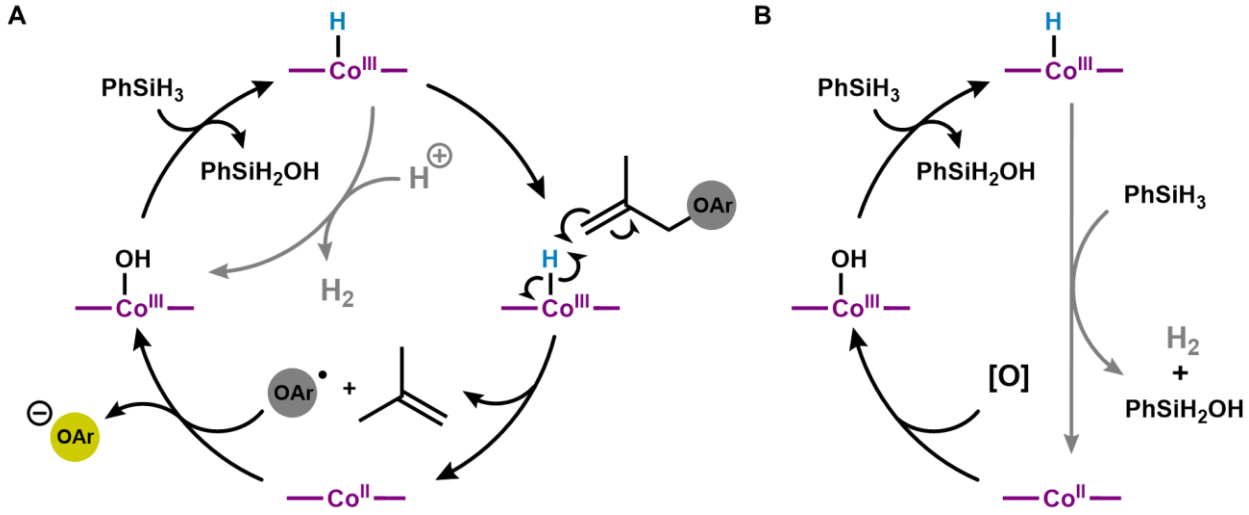


Figure 4.11. Proposed catalytic cycle for MHAT deallylation by CoCYP119. A) Main catalytic cycle with protonation shunt pathway. B) Alternative shunt pathway for H_2 production via HAT to PS.

This cobalt hydride intermediate can react via three possible pathways. In the first, a non-productive shunt pathway, the cobalt hydride is protonated, resulting in the formation of H_2, and returning the catalyst to the resting (Co(III)) state. Indeed, GC analysis of the headspace of these reactions revealed increased H_2 gas is produced in the presence of enzyme (See supplemental Figure 4.12). An alternative explanation for the formation of H_2 gas is the reaction of the cobalt hydride via hydrogen atom abstraction from an additional silane Si-H (Figure 4.11B). In the future, NMR analysis of the headspace of a reaction conducted in D_2O would clearly differentiate these two hypotheses. If H_2 is the predominant species detected in the headspace, HAT to silane is

more likely the operative pathway. Detection of significant HD would suggest that the cobalt hydride is protonated by water.

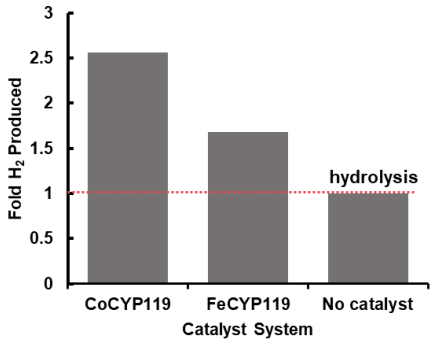


Figure 4.12. Relative amount of H_2 gas measured in the headspace of a reaction consisting of 5 μM catalyst, 1 mM DMPS, 20% DMSO in KPi pH=6.0 after 18 hours at 50°C. Significant H_2 gas was detected in the absence of catalyst, likely due to hydrolysis of DMPS at elevated temperatures.

In the third, productive pathway, allyl substrate binds in the active site, and HAT from the cobalt center to the alkene leads to a carbon centered radical with Markovnikov selectivity, and a Co(II) species. To probe this HAT step, we tested the reaction of CoCYP119 and DMPS with a variety of protected dinitrophenolate substrates and measured the initial rates (Figure 4.13). The initial rate with the monosubstituted alkene (**3**) was substantially slower than that with the disubstituted alkene (**1**). This result is consistent with a less stable radical intermediate species for the monosubstituted alkene (secondary versus tertiary radical for the disubstituted alkene substrate). The benzyl protected dinitrophenolate substrate (**4**) was unable to react with CoCYP119, indicating that the alkene is required as an HAT acceptor.

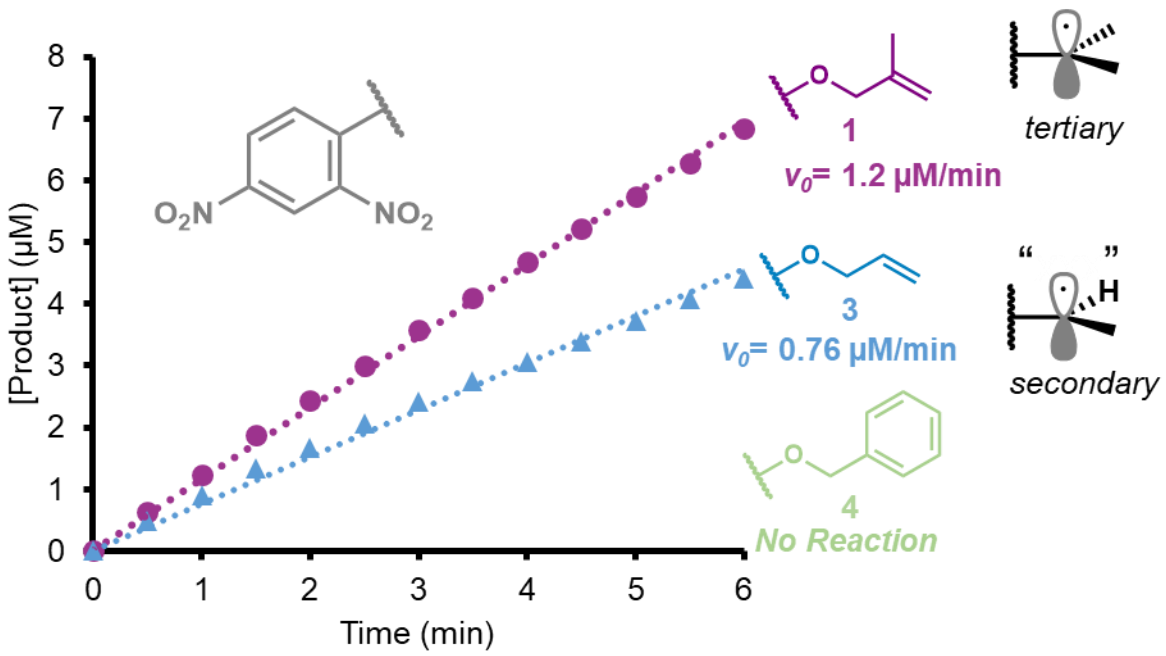


Figure 4.13. Relative rates for the deallylation of substrates **1**, **3** and **4** by CoCYP119. Reaction conditions: 0.1 mM allyl Substrate **1**, **3**, or **4**, 0.2 mM DMPS or 1 mM PS, 2 μM WT CoCYP119, in 100 mM KPi pH 6.0 and 20% DMSO.

Following HAT, decomposition of the radical yields isobutene (bp = -7° C), and dinitrophenoxy radical. GC analysis of the headspace of the reaction confirms the formation of isobutene (See Figure 55 in the supplemental figures section). Dinitrophenoxy radical is a potent oxidant (E_{red}~ 1.2 V vs Normal Hydrogen Electrode (NHE) for *p*-nitrophenolate,[46] E_{red} for the 2,4-dinitro compound is likely higher). The oxidation potential of Co(II) to Co(III) must be no less than 0.66 V vs. NHE, since the reduction of Co(III) to Co(II) can be accomplished by dithionite (E_{red} = -0.66 V vs. NHE for dithionite,[47] E_{red} (Co(III/II)) > 0.66 V vs. NHE). Thus, the phenoxide radical is sufficiently oxidizing to return the Co(II) species to the Co(III) catalyst resting state. In the absence of the phenoxide radical, molecular oxygen could also serve as an oxidant to return the metal to the Co(III) state.

4. 3. 5. Site saturation mutagenesis

Directed evolution has been used extensively to improve the activity of enzymes for non-native reactions.[20] We envisioned that directed evolution could likewise be harnessed to improve MHAT activity of CoCYP119. Importantly, directed evolution enables improvement of enzymatic activity without detailed prior knowledge about the enzyme's mechanism. Based on our prior mechanistic studies, however, there were clearly many features of the reaction that contributed to lower overall yields. Improved enzymes might improve the binding affinity for either substrate, improve the rate of formation of the Co-H intermediate, kinetically protect the fragile Co-H intermediate from protonation with water, or alter electronics of the cofactor to increase the hydrogen atom-like reactivity of the metal hydride intermediate. We hypothesized that by screening catalysts for improved deallylation activity, we might uncover CoCYP119 variants with improved activity.

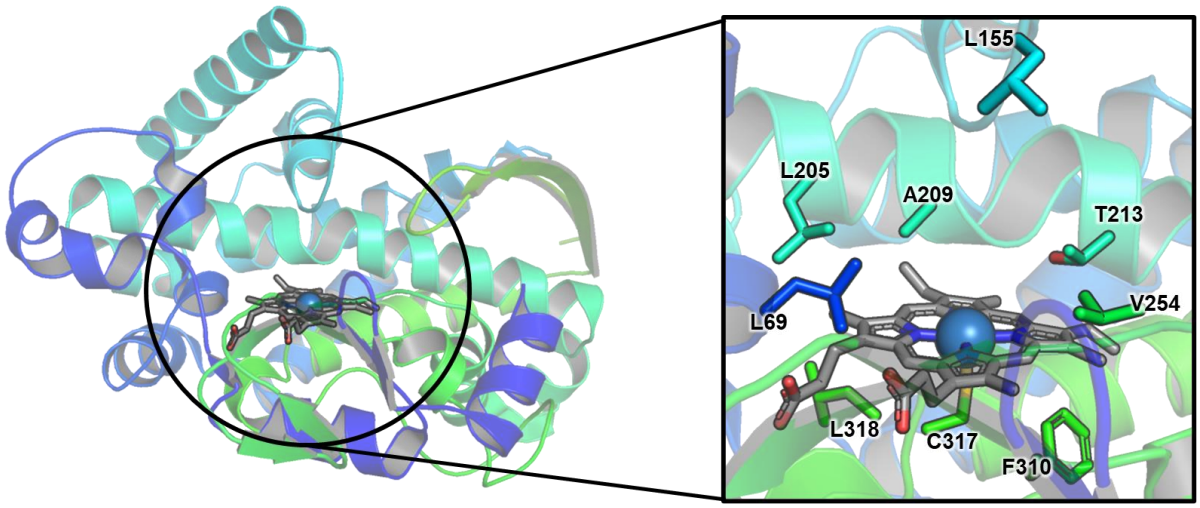


Figure 4.14. Active site of CYP119 with active site residues selected for site saturation mutagenesis. PDB 1IO7.

Previous engineering of CYP119 towards both improved hydroxylation (heme-dependent)[48] and non-native carbene transfer (Ir(Me)PPIX-dependent)[37,38] activities has revealed several residues in the active site that exert a profound influence on reactivity. We targeted ten of these active site residues for an initial round of site saturation mutagenesis (Figure 4.14). Each

of these libraries was expressed as the cobalt-substituted CYP119 (CoCYP119) construct in rich media (using the method optimized in chapters 2 and 3 of this thesis) and prepared as heat-treated lysates in a 96-well plate format. Libraries were screened at relatively high (and insoluble) substrate concentrations, as activity was below the limit of detection when lower concentrations of substrate were used. Even with these conditions, activity with DMPS was below the limit of detection under these conditions and we therefore elected to use PS instead. Due to the insolubility of the substrates at the beginning of the reaction, product formation was assessed by absorbance at 405 nm after the reaction solution had cleared, at 1 hour.

We found that mutation at several residues throughout the active site lead to improved deallylation activity. Retention of function curves for all site saturation mutagenesis libraries are given in Figures 4.15-4.23. Mutations to site L69, L205, A209, T213, and V254 were particularly fruitful. Notably, mutation to the proximal ligand to the cobalt (C317) yielded no active variants, in good agreement with earlier observations of the C213A and C317S variants as purified enzymes (Figure 4.17).

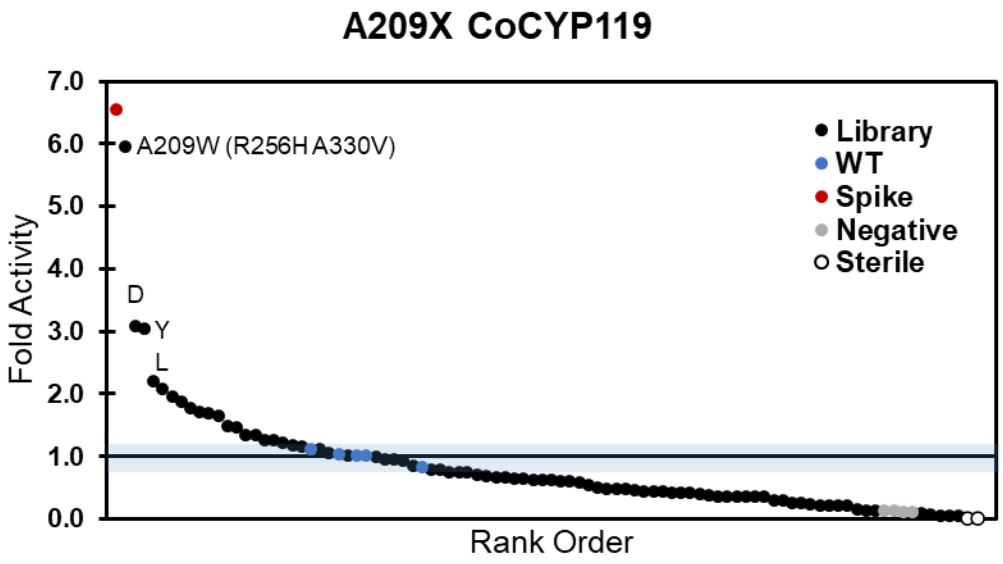


Figure 4.15. Retention of function curve for site saturation mutagenesis at A209. Screening conditions: 1 mM substrate **1**, 10 mM PS, 200 µL heat treated cell lysate in KPi 6.0 with 20% DMSO. Total reaction

volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the absorbance of the sterile control wells after 1 hour. The blue bar corresponds to the range of activity of WT enzyme.

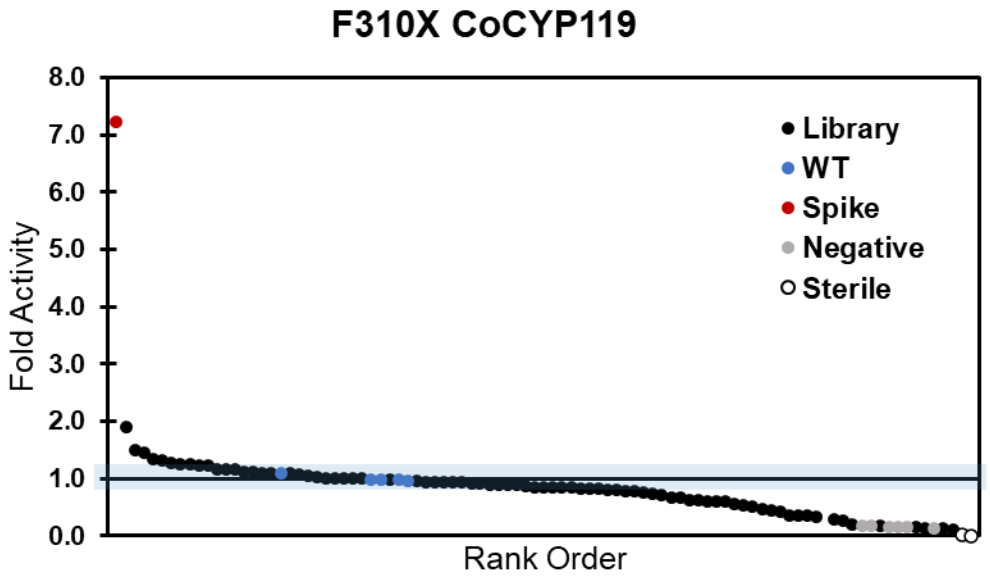


Figure 4.16. Retention of function curve for site saturation mutagenesis at F310. Screening conditions: 1 mM substrate **1**, 10 mM PS, 200 µL heat treated cell lysate in KPi 6.0 with 20% DMSO. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the absorbance of the sterile control wells after 1 hour. The blue bar corresponds to the range of activity of WT enzyme.

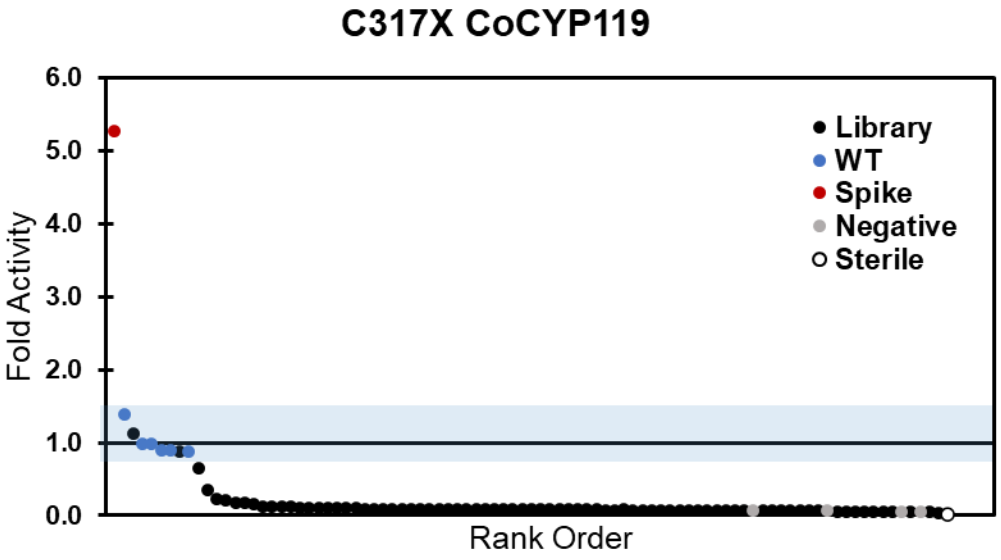


Figure 4.17. Retention of function curve for site saturation mutagenesis at C317. Screening conditions: 1 mM substrate **1**, 10 mM PS, 200 µL heat treated cell lysate in KPi 6.0 with 20% DMSO. Total reaction

volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the absorbance of the sterile control wells after 1 hour. The blue bar corresponds to the range of activity of WT enzyme.

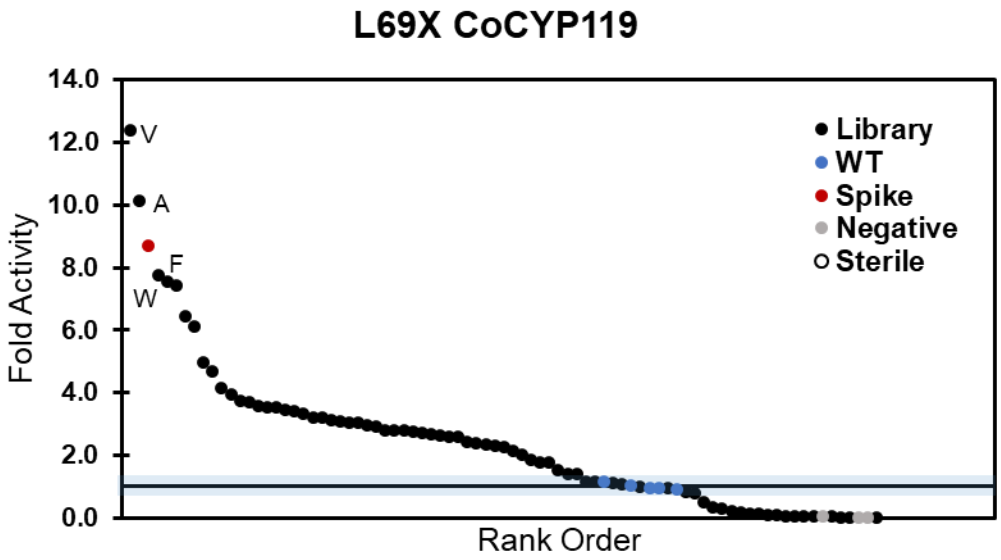


Figure 4.18. Retention of function curve for site saturation mutagenesis at L69. Screening conditions: 1 mM substrate **1**, 10 mM PS, 200 µL heat treated cell lysate in KPi 6.0 with 20% DMSO. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the absorbance of the sterile control wells after 1 hour. The blue bar corresponds to the range of activity of WT enzyme.

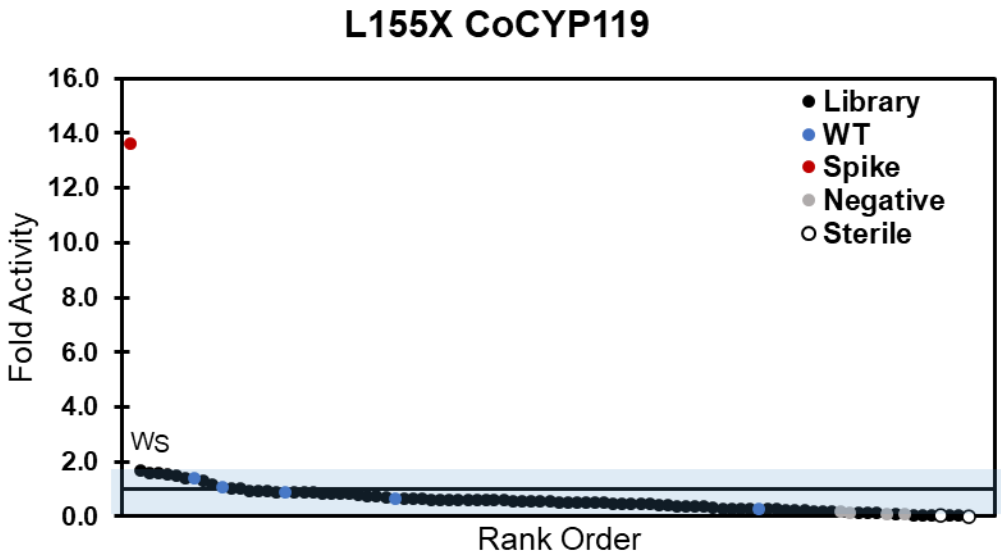


Figure 4.19. Retention of function curve for site saturation mutagenesis at L155. Screening conditions: 1 mM substrate **1**, 10 mM PS, 200 µL heat treated cell lysate in KPi 6.0 with 20% DMSO. Total reaction

volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the absorbance of the sterile control wells after 1 hour. The blue bar corresponds to the range of activity of WT enzyme.

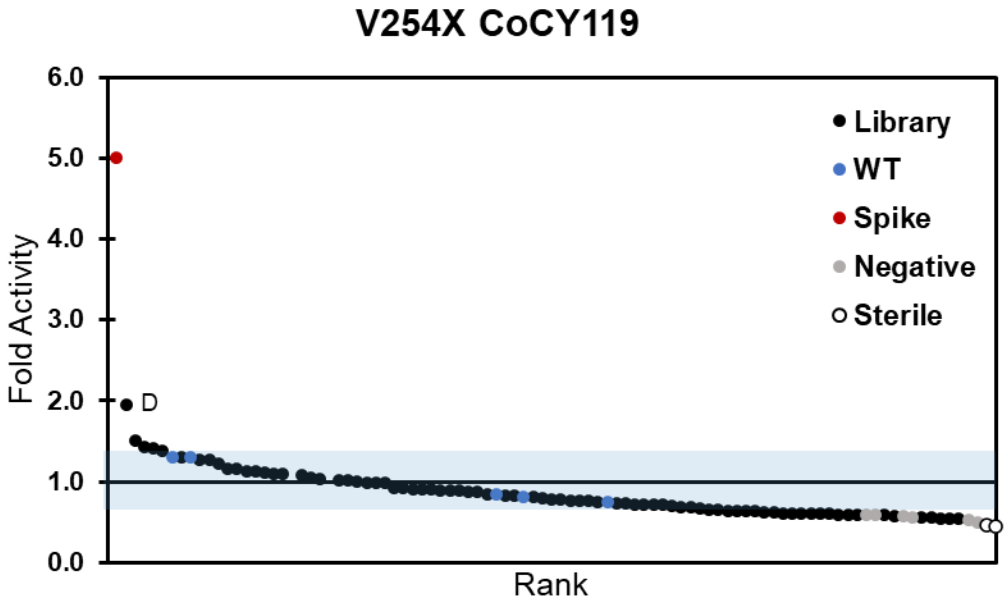


Figure 4.20. Retention of function curve for site saturation mutagenesis at V254. Screening conditions: 1 mM substrate **1**, 10 mM PS, 200 µL heat treated cell lysate in KPi 6.0 with 20% DMSO. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the absorbance of the sterile control wells after 1 hour. The blue bar corresponds to the range of activity of WT enzyme.

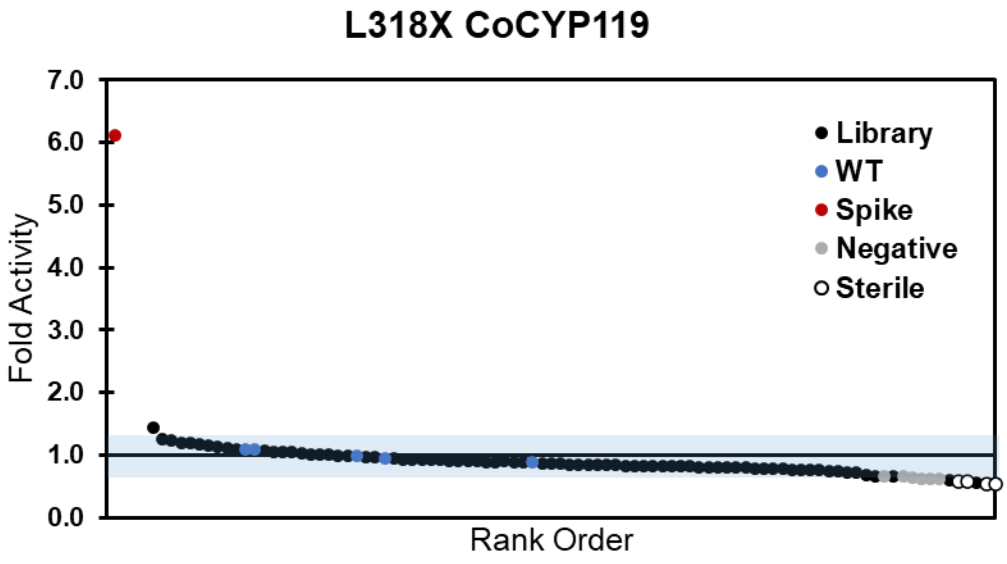


Figure 4.21. Retention of function curve for site saturation mutagenesis at L318. Screening conditions: 1 mM substrate **1**, 10 mM PS, 200 µL heat treated cell lysate in KPi 6.0 with 20% DMSO. Total reaction

volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the absorbance of the sterile control wells after 1 hour. The blue bar corresponds to the range of activity of WT enzyme.

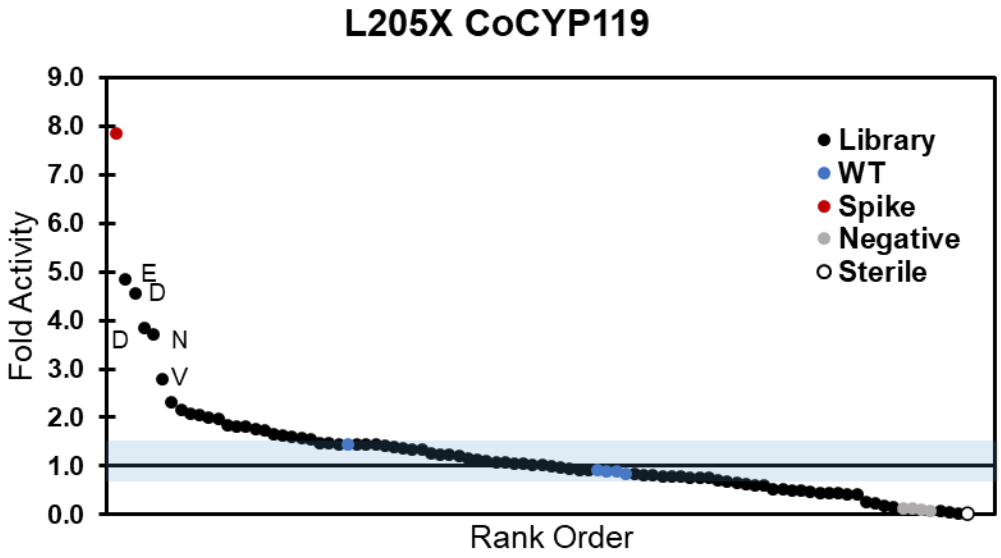


Figure 4.22. Retention of function curve for site saturation mutagenesis at L205. Screening conditions: 1 mM substrate **1**, 10 mM PS, 200 µL heat treated cell lysate in KPi 6.0 with 20% DMSO. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the absorbance of the sterile control wells after 1 hour. The blue bar corresponds to the range of activity of WT enzyme.

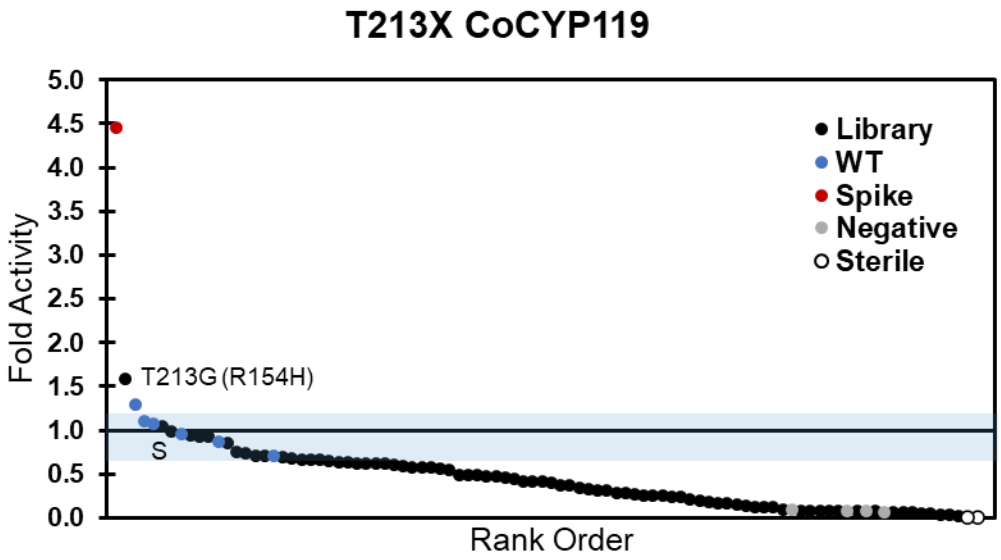


Figure 4.23. Retention of function curve for site saturation mutagenesis at T213. Screening conditions: 1 mM substrate **1**, 10 mM PS, 200 µL heat treated cell lysate in KPi 6.0 with 20% DMSO. Total reaction

volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the absorbance of the sterile control wells after 1 hour. The blue bar corresponds to the range of activity of WT enzyme.

We selected three variants, L69N, L69H, and V254D for validation. We began by measuring the initial rate and overall yield of reaction with the allyl substrate **1** (0.1 mM) and PS (0.2 mM) by monitoring absorbance at 400 nm. Under these conditions, all three variants outperformed the parent CoCYP119 in terms of initial rate and endpoint (1 h) product yield, reported as total turnover numbers (Figure 4.24A). L69H was the best of the variants tested, with an initial rate that was 9.8-fold that of the WT reaction, and a 5.4-fold improvement in yield when assayed with NN eq of PS. V245D only modestly outperformed WT under these conditions (3-fold rate increase, 2-fold yield increase), which recapitulates the screening data.

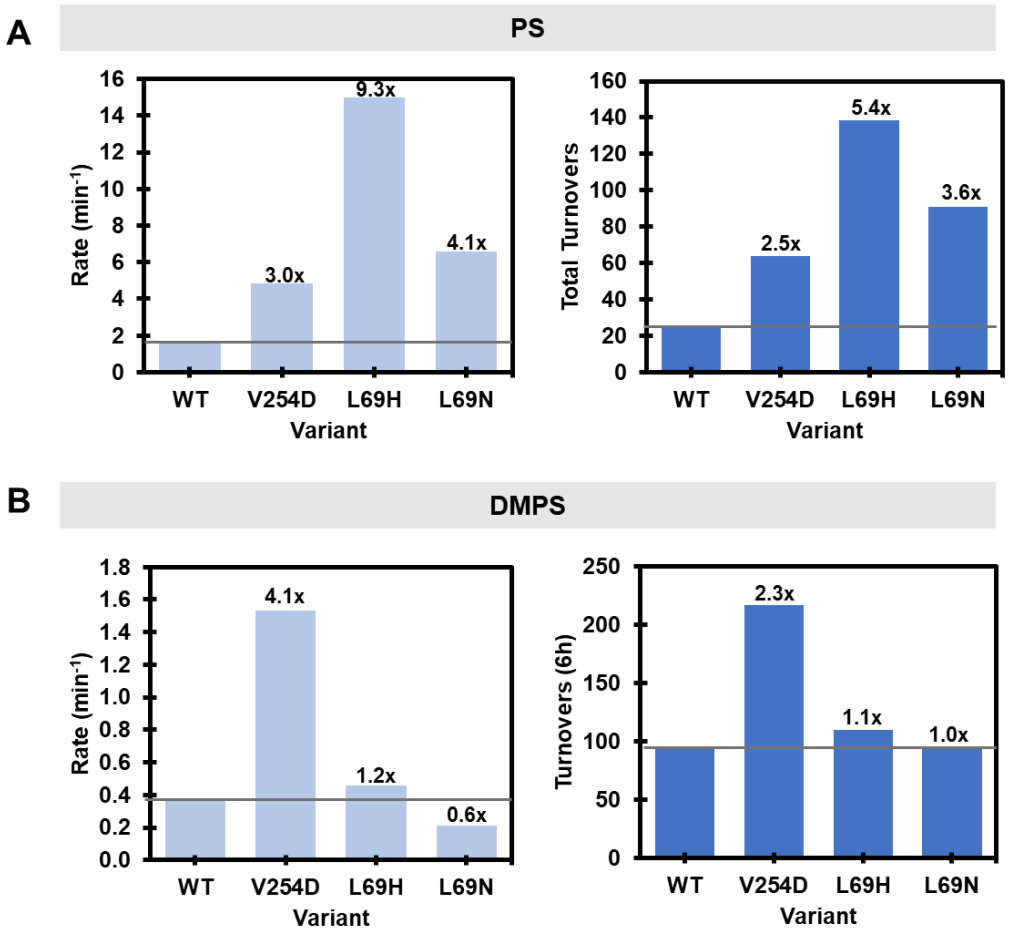


Figure 4.24. Validation of hits from the first round of site saturation mutagenesis. Screening conditions: KPi pH = 6.0 with 20% DMSO, 0.1 µM Enzyme (1000 max TON), 0.1 mM allyl substrate **1**, 0.2 mM PS (A) or DMPS (B). Light blue bars represent the initial rates of each reaction, and dark blue bars represent the total

turnover number (in A) or the number of turnovers completed in 6 h (B), as the reaction had not yet reached completion.

We previously found that PS quickly hydrolyzed in water. We were therefore motivated to explore the reactivity of improved variants with DMPS, which is more stable under aqueous conditions. Reactions with DMPS are sluggish with the WT enzyme, likely because the enzyme struggles to activate the sterically bulky and less hydridic DMPS. We tested the reaction of WT and the three variants with DMPS and measured initial rates and product formation after 6 h Figure 4.24B). Neither of the L69 variants showed improvement with DMPS relative to WT. However, the DMPS reaction with V254D exhibited a 2.3-fold improvement in yield and 4.1-fold improvement in rate relative to the WT enzyme.

4. 3. 6. Recombination of activating mutations

Of the ten active site residues initially targeted for mutagenesis, site saturation mutagenesis at five of these sites yielded variants with improved activity towards deallylation with PS. To probe cooperative effects in the active site and further improve CoCYP119 MHAT activity, we designed a focused recombination library that simultaneously mutagenized all five sites for a subsequent sequence space comprised of 3840 unique recombinants. The degenerate codons and possible mutations selected are given in Table 4.2. Of these 3840 variants, we screened fifteen 96-well plates, for a total of up to 1320 unique variants, or ~35% of the sequence space.

Because the variant V254D showed improved activity with DMPS, we chose to switch to this substrate for further engineering. WT CoCYP119 activity was undetectable under these conditions, and so V254D was chosen as the “parent” enzyme to which variants in the recombination library were directly compared. To improve solubility of substrates and select for variants with improved activity under more synthetically-useful conditions, we also switched from 20% DMSO to 10% Acetonitrile as the cosolvent. Retention of function curves for each plate screened from this recombination library can be found in Figures 4.25-4.39.

Table 4.2. Degenerate codons the amino acids they encode for the recombination library. The identity of the "parent" variant is denoted in blue. A 1:1 mixture WT and V254D plasmids was used as the template for cloning.

Residue:	L69	L205	A209	T213	V254
Deg. Codon:	HWC	VWK	DHC	RSC	1:1
Mutations	L	L	A	T	V
	N	N	N	G	D
	H	D	D	S	
	F	Q	I	A	
	Y	E	F		
	I	H	S		
		I	T		
		K	V		
		M			
		V			
# muts	6	10	8	4	2

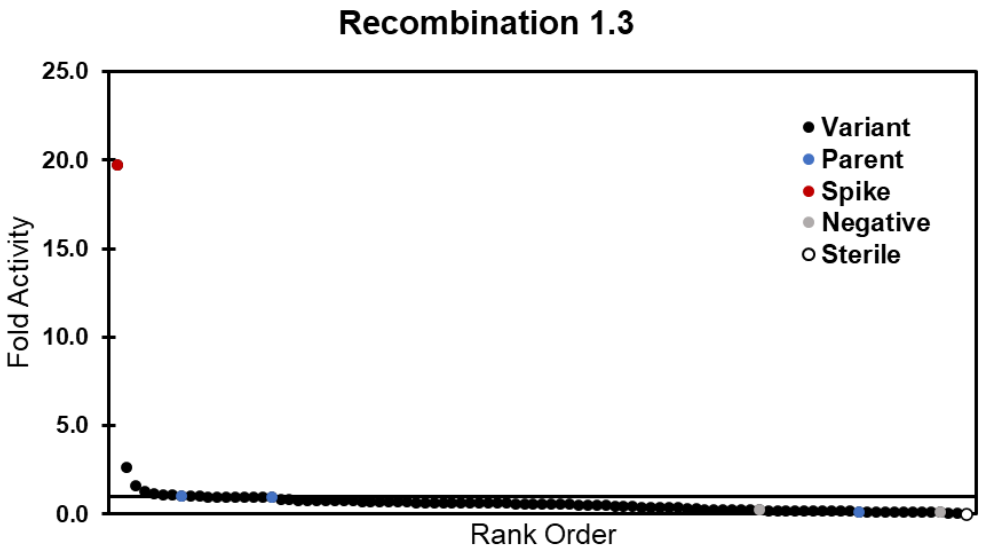


Figure 4.25. Retention of function curve for plate 1.3 of the recombination library. Screening conditions: 1 mM DMPS, 1 mM substrate **1**, 200 µL heat treated cell lysate, 10% acetonitrile in KPi pH 6.0. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the sterile control well. Fold activity is given relative to wells expressing the V245D parent.

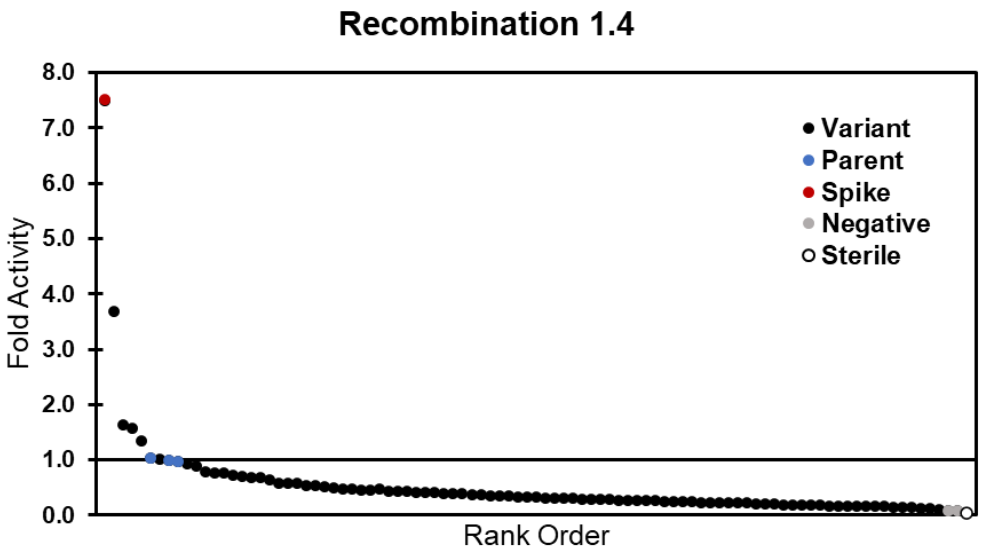


Figure 4.26. Retention of function curve for plate 1.4 of the recombination library. Screening conditions: 1 mM DMPS, 1 mM substrate **1**, 200 µL heat treated cell lysate, 10% acetonitrile in KPi pH 6.0. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the sterile control well. Fold activity is given relative to wells expressing the V245D parent.

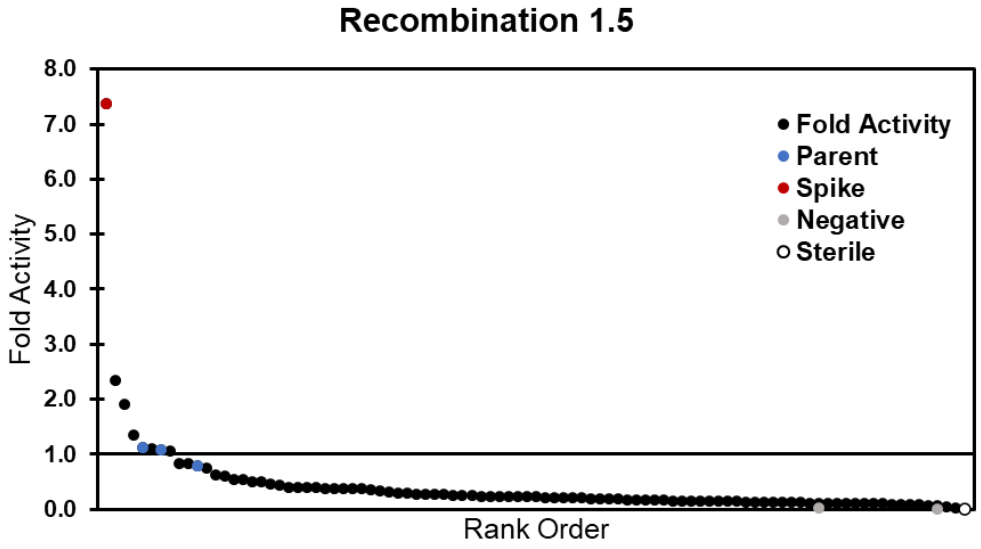


Figure 4.27. Retention of function curve for plate 1.5 of the recombination library. Screening conditions: 1 mM DMPS, 1 mM substrate **1**, 200 µL heat treated cell lysate, 10% acetonitrile in KPi pH 6.0. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the sterile control well. Fold activity is given relative to wells expressing the V245D parent.

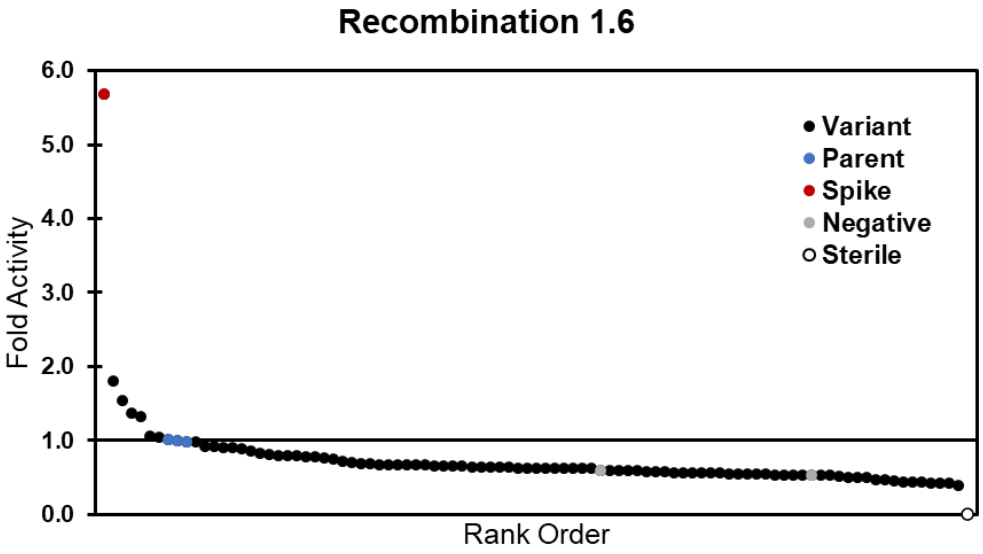


Figure 4.28. Retention of function curve for plate 1.6 of the recombination library. Screening conditions: 1 mM DMPS, 1 mM substrate **1**, 200 µL heat treated cell lysate, 10% acetonitrile in KPi pH 6.0. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the sterile control well. Fold activity is given relative to wells expressing the V245D parent.

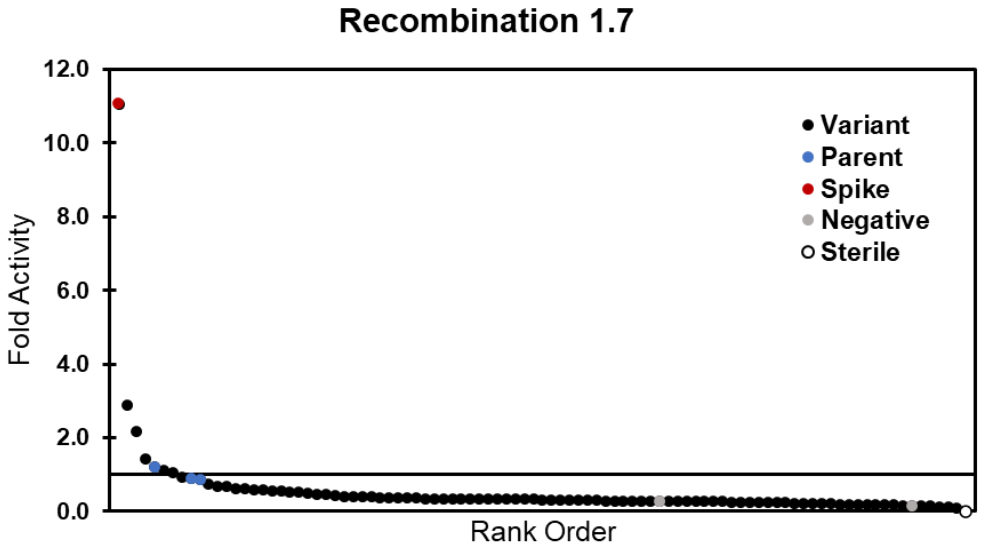


Figure 4.29. Retention of function curve for plate 1.7 of the recombination library. Screening conditions: 1 mM DMPS, 1 mM substrate **1**, 200 µL heat treated cell lysate, 10% acetonitrile in KPi pH 6.0. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the sterile control well. Fold activity is given relative to wells expressing the V245D parent.

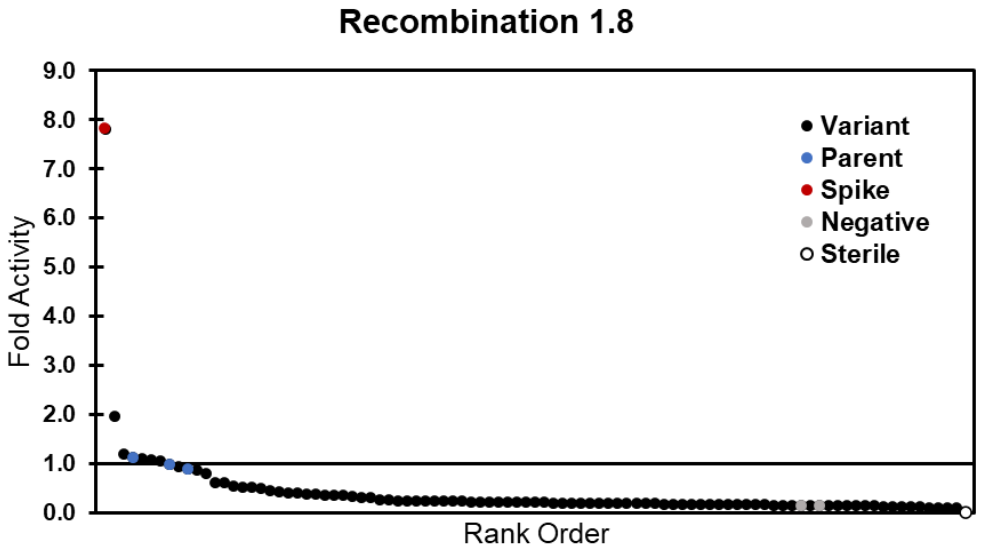


Figure 4.30. Retention of function curve for plate 1.8 of the recombination library. Screening conditions: 1 mM DMPS, 1 mM substrate **1**, 200 µL heat treated cell lysate, 10% acetonitrile in KPi pH 6.0. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the sterile control well. Fold activity is given relative to wells expressing the V245D parent.

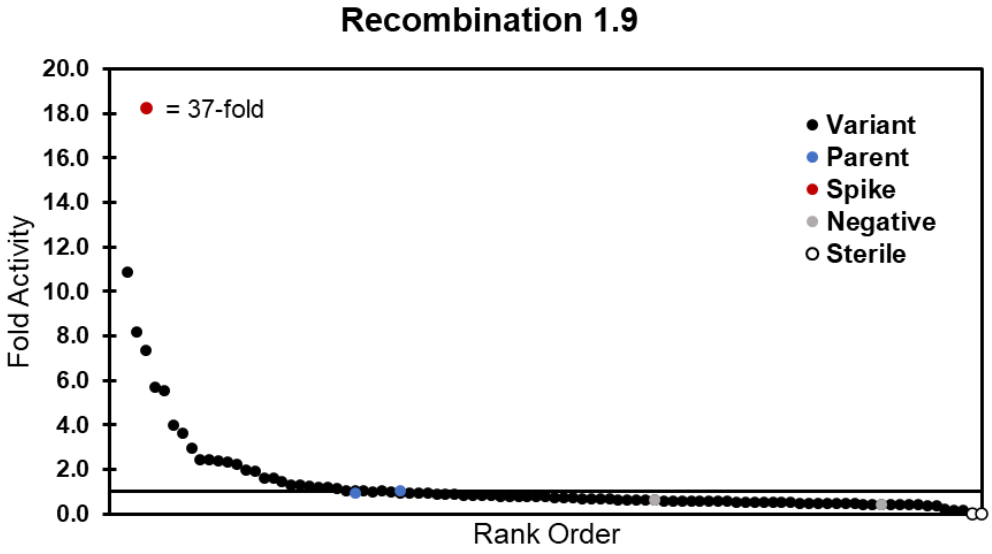


Figure 4.31. Retention of function curve for plate 1.9 of the recombination library. Screening conditions: 1 mM DMPS, 1 mM substrate **1**, 200 µL heat treated cell lysate, 10% acetonitrile in KPi pH 6.0. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the sterile control well. Fold activity is given relative to wells expressing the V245D parent.

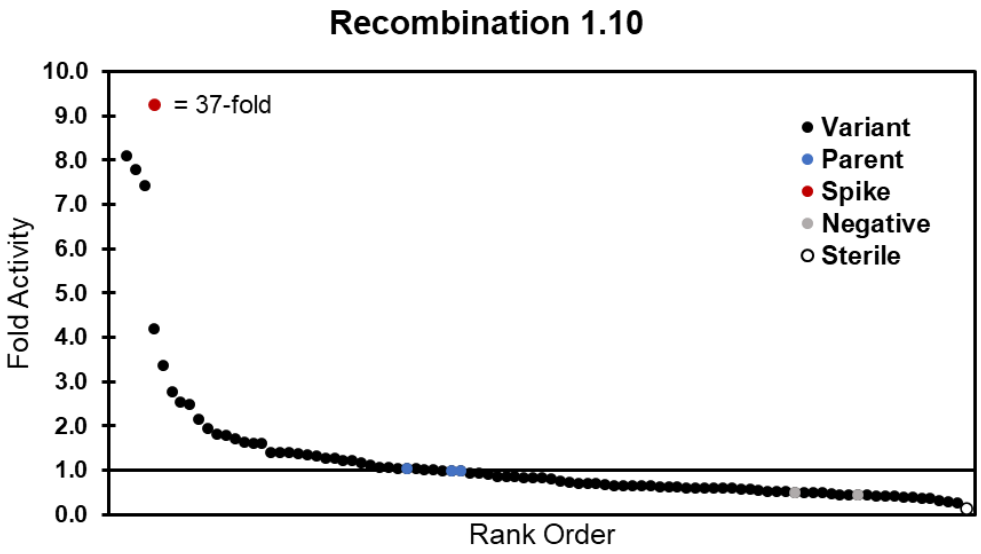


Figure 4.32. Retention of function curve for plate 1.10 of the recombination library. Screening conditions: 1 mM DMPS, 1 mM substrate **1**, 200 µL heat treated cell lysate, 10% acetonitrile in KPi pH 6.0. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the sterile control well. Fold activity is given relative to wells expressing the V245D parent.

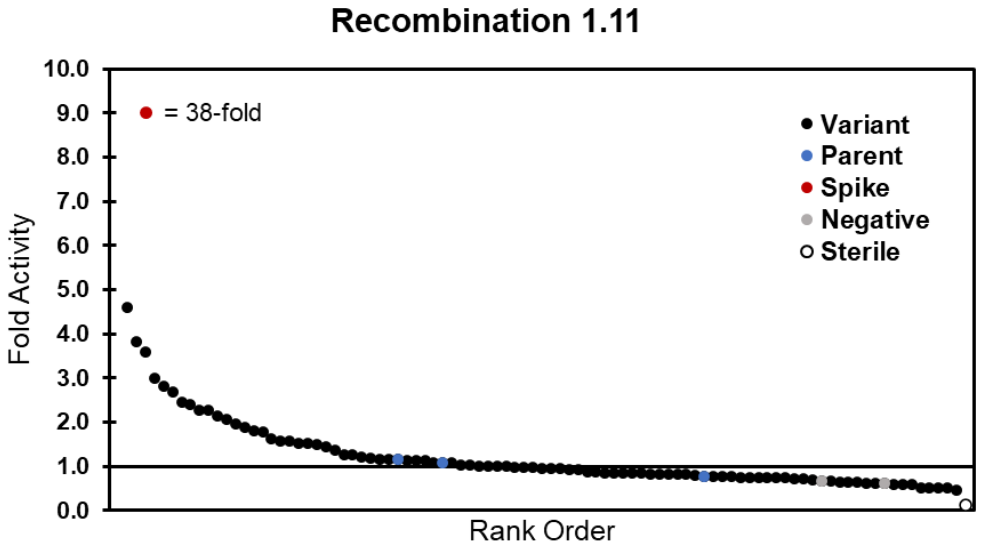


Figure 4.33. Retention of function curve for plate 1.11 of the recombination library. Screening conditions: 1 mM DMPS, 1 mM substrate **1**, 200 µL heat treated cell lysate, 10% acetonitrile in KPi pH 6.0. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the sterile control well. Fold activity is given relative to wells expressing the V245D parent.

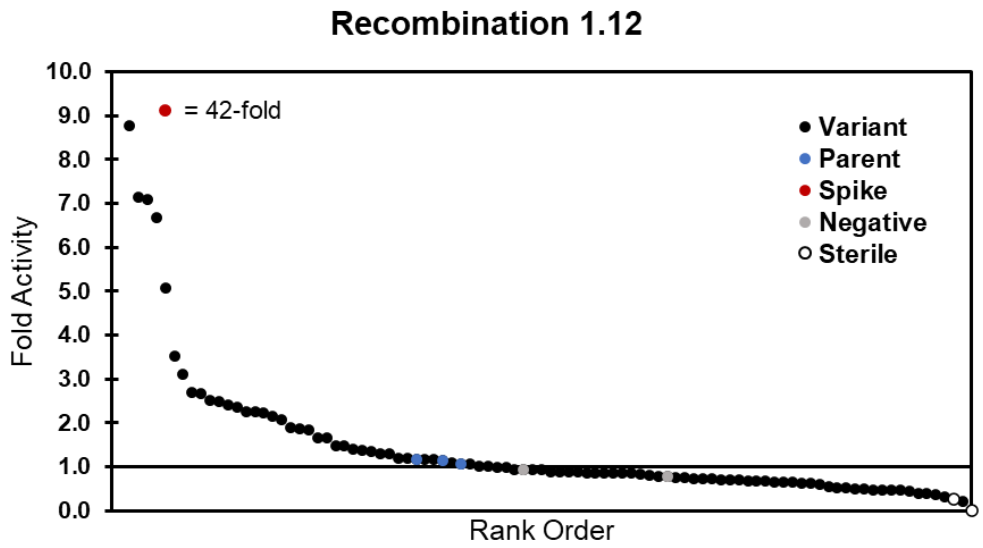


Figure 4.34. Retention of function curve for plate 1.12 of the recombination library. Screening conditions: 1 mM DMPS, 1 mM substrate **1**, 200 µL heat treated cell lysate, 10% acetonitrile in KPi pH 6.0. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the sterile control well. Fold activity is given relative to wells expressing the V245D parent. Parent wells expressed poorly due to a deleterious mutation in the parent plasmid.

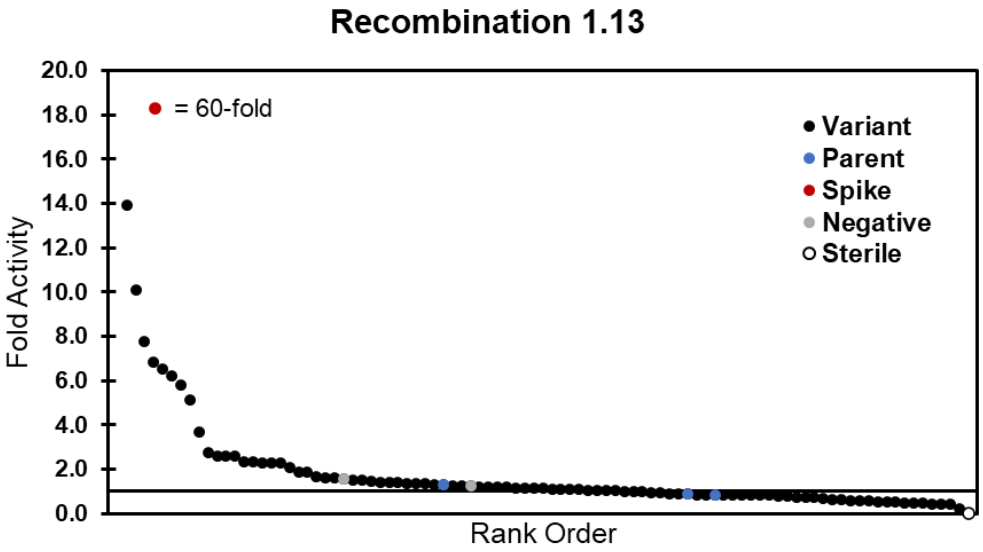


Figure 4.35. Retention of function curve for plate 1.13 of the recombination library. Screening conditions: 1 mM DMPS, 1 mM substrate **1**, 200 µL heat treated cell lysate, 10% acetonitrile in KPi pH 6.0. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the sterile control well. Fold activity is given relative to wells expressing the V245D parent. Parent wells expressed poorly due to a deleterious mutation in the parent plasmid.

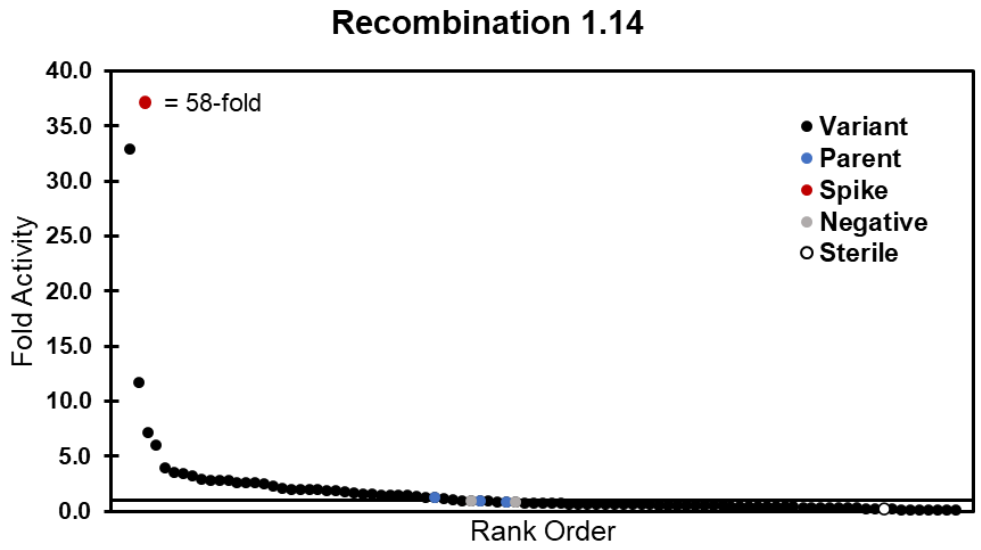


Figure 4.36. Retention of function curve for plate 1.14 of the recombination library. Screening conditions: 1 mM DMPS, 1 mM substrate **1**, 200 µL heat treated cell lysate, 10% acetonitrile in KPi pH 6.0. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the sterile control well. Fold activity is given relative to wells expressing the V245D parent. Parent wells expressed poorly due to a deleterious mutation in the parent plasmid.

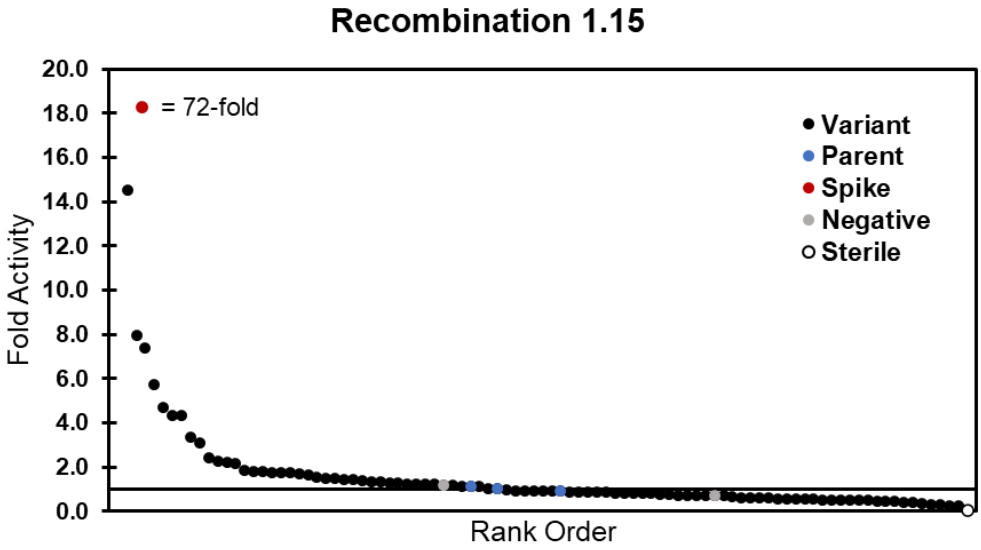


Figure 4.37. Retention of function curve for plate 1.15 of the recombination library. Screening conditions: 1 mM DMPS, 1 mM substrate **1**, 200 µL heat treated cell lysate, 10% acetonitrile in KPi pH 6.0. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the sterile control well. Fold activity is given relative to wells expressing the V245D parent. Parent wells expressed poorly due to a deleterious mutation in the parent plasmid.

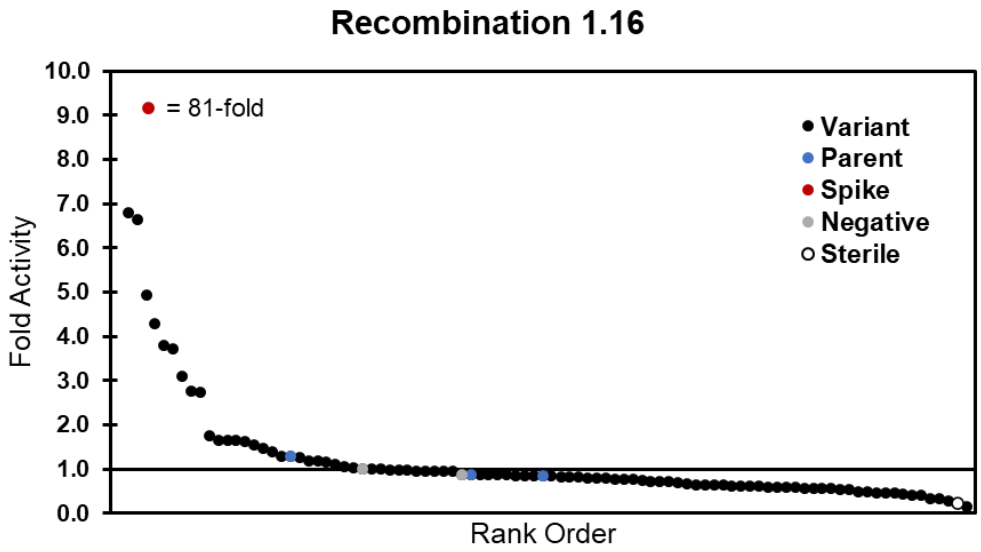


Figure 4.38. Retention of function curve for plate 1.16 of the recombination library. Screening conditions: 1 mM DMPS, 1 mM substrate **1**, 200 µL heat treated cell lysate, 10% acetonitrile in KPi pH 6.0. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the sterile control well. Fold activity is given relative to wells expressing the V245D parent.

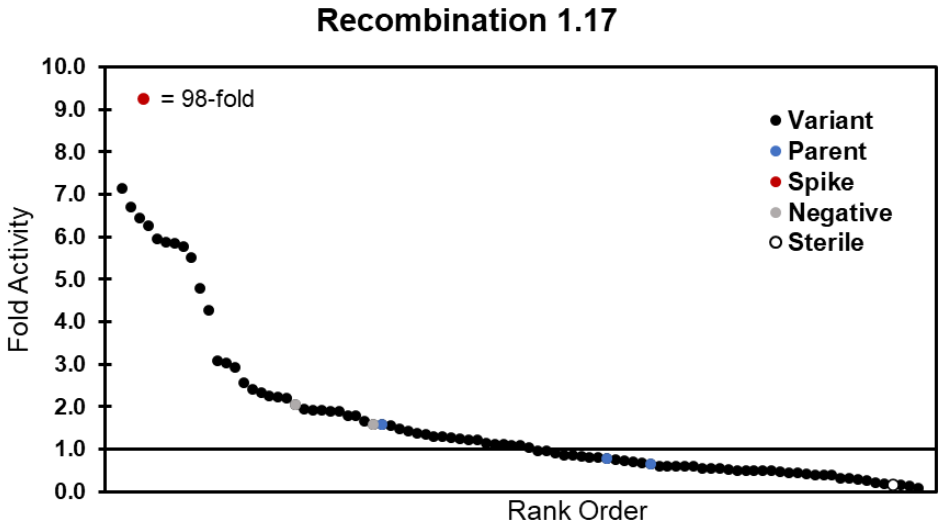


Figure 4.39. Retention of function curve for plate 1.17 of the recombination library. Screening conditions: 1 mM DMPS, 1 mM substrate **1**, 200 µL heat treated cell lysate, 10% acetonitrile in KPi pH 6.0. Total reaction volume was 250 µL. Product formation was measured by absorbance at 405 nm relative to the sterile control well. Fold activity is given relative to wells expressing the V245D parent.

After screening of the recombination library, the variants that were most strongly activated compared to V254D variant were sequenced and assembled into a master plate of 90 active variants. This master plate was rescreened under identical conditions to enable direct comparison of activated recombination variants. The retention of function curve for this master plate can be found in Figure 4.40. In this head-to-head assay, all but two of these variants reproduced as more active than V254D, attesting to the quality of the upstream screening of the wider recombination space.

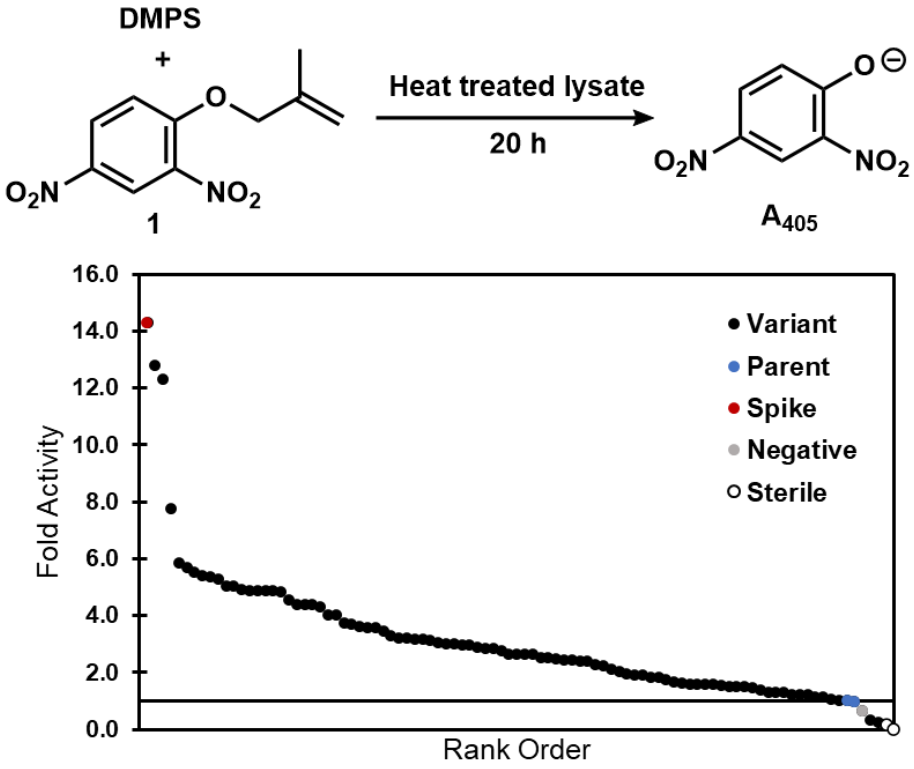


Figure 4.40. Retention of function curve for master plate screened with substrate **3**. Screening conditions: 200 µL cell lysate in KPi pH 6.0, 1 mM substrate **1**, 5 mM DMPS, 10% Acetonitrile. The total reaction volume was 250 µL. Reactions were run at room temperature for 20 hours at which point absorbance at 405 nm was measured using a plate reader. Data were background subtracted by the sterile control well and reported as fold activity relative to the parent (V254D).

Sequencing of this master plate revealed significant diversity at each of the mutated sites. Logoplot depictions of both the intended distribution of mutations from the recombination library

(Figure 4.41A) and the distribution of mutations found in activated subset of variants chosen for the master plate (Figure 4.41B) can be found below.

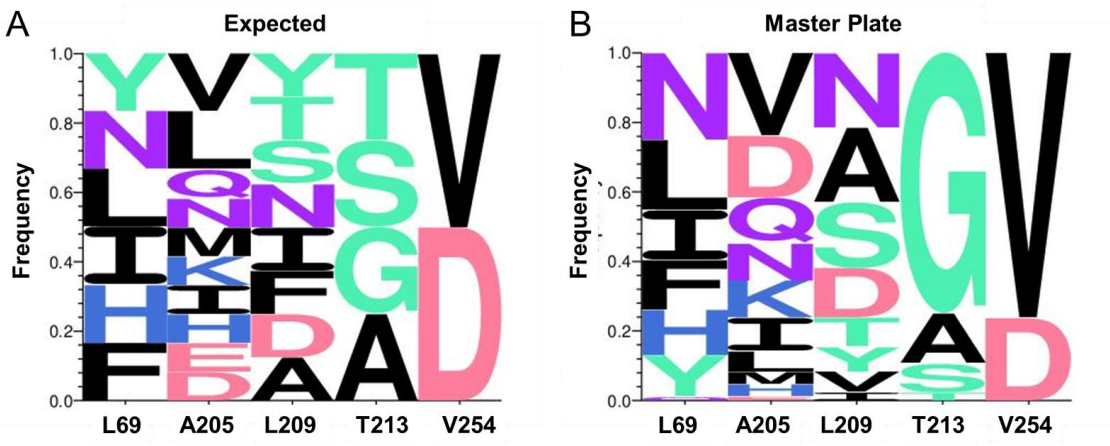


Figure 4.41. Logoplots depicting the frequency of mutations expected, based on the degenerate codons used for library construction (A), and the frequency of mutations measured in the master plate of activated variants (B). The size of each letter represents the relative frequency of that mutation, and the colors roughly correspond to the chemical features of each residue.

Despite the improved activity of the V254D mutation with DMPS, this mutation is less common in the master plate of activated variants. Indeed, both of the top two hits, 14A2 and 9B12, maintain the native Val at position 254. Mutation at position 213, however, was more common. T213G was the most common mutation at this position, and both the native Thr213 and the more conservative T213S mutations were de-enriched in the master plate. 9B12 features the T213G mutation, in addition to an A209S mutation. Notably, the A209S T213G combination, which is expected to occur in only 3.1% (1/8)*(1/4) = 0.031) of the plate based on the corresponding mutational frequencies, was present in about 15% of the variants screened in the master plate and is pair of mutations is found in four of the top seven variants. Intriguingly, these two residues

inhibit *i* and *i+4* positions on the I helix of CYP119, and lie above the plane of the porphyrin cofactor, opposite the site of the proximal thiolate ligand (Figure 4.42).

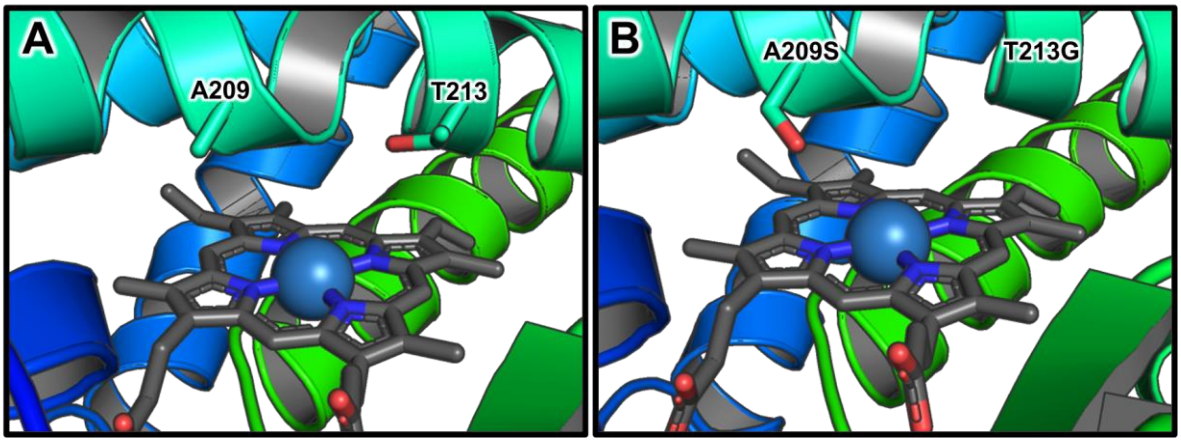


Figure 4.42. Model of A209S and T213G mutations in the active site of CoCYP119. PBD 1IO7. A) structure of the native CoCYP119 active site. B) model of active site with A209 and T213G mutations.

4. 3. 10. Discovery of a reductive dearomitization mode of reactivity

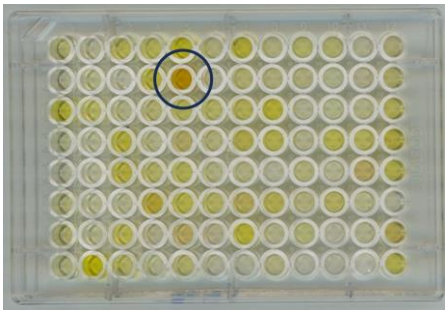


Figure 4.43. Example of an orange-colored well from a 96-well plate screen of activated variants.

Throughout screening, we noticed the appearance of a distinctly black-brown color at early time points in a handful of the reaction wells. These wells subsequently turned orange over the course of the reaction (Figure 4.49). This color was not correlated with increased absorbance at 405 nm, indicating that a different product may be accumulating in these wells. One early discovered, orange-color producing variant was the T213G point mutant. We purified this variant and assayed its activity by monitoring changes in the absorbance of the starting material over the course of the reaction. These spectra are shown in Figure 4.50. Noticeably, the isosbestic point

at 327 nm in the time course of the WT reaction is not present in the time course with the T213G variant, indicating that the T213G time course can no longer be described as a clean interconversion between two species. Two new spectral features appeared in the reaction with T213G. A broad feature between 550 and 600 nm grows in and then begins to diminish over time. An additional feature at ~450 nm is also present, and its growth is monotonic.

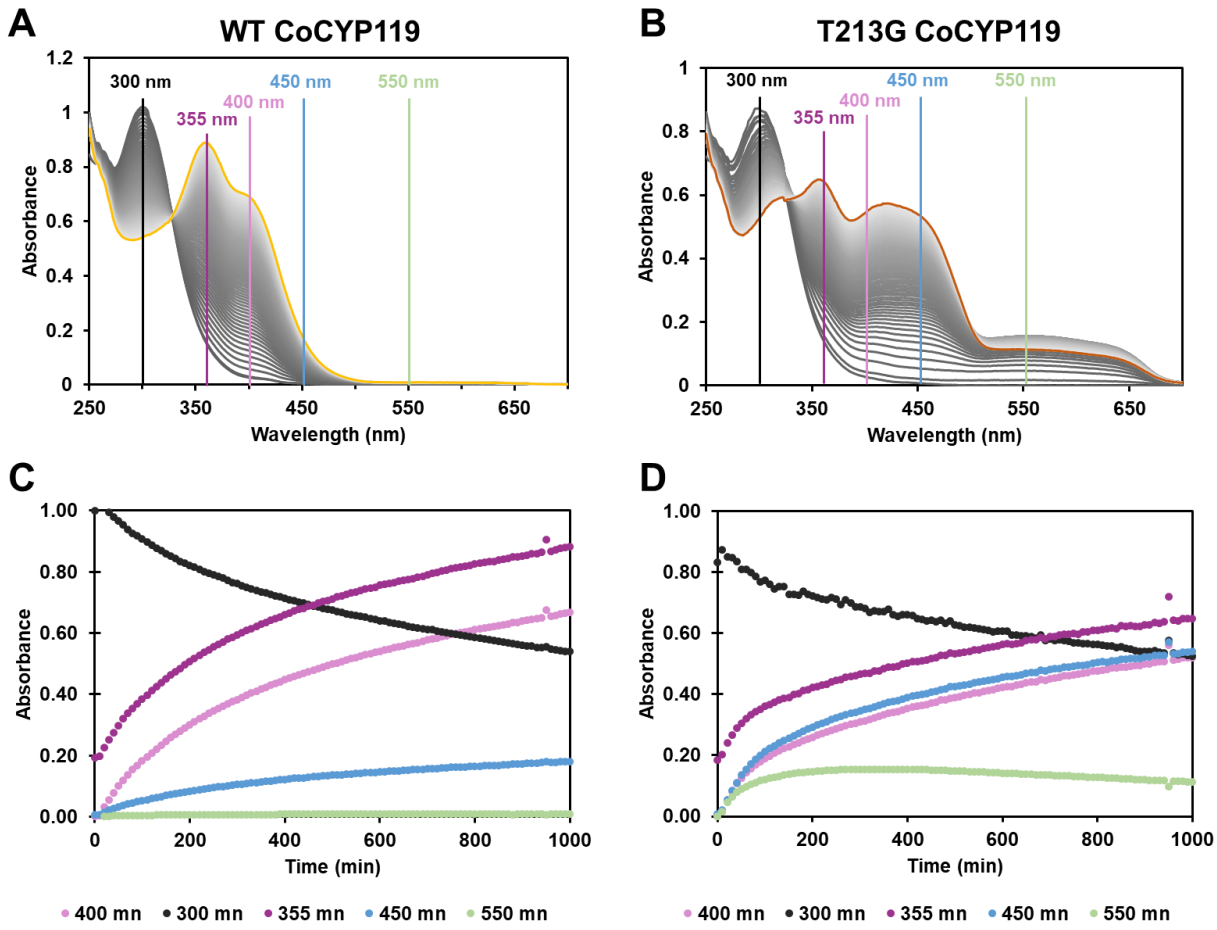


Figure 4.44. Characterization of a new product by UV-visible spectroscopy. A) Time course of deallylation of substrate **1** by WT CoCYP119. The spectral features at 355 nm and 400 nm are identical to those of the expected deallylated product, **3**. B) Time course for the reaction of T213G with substrate **1**. C) Change in absorbance over time of key wavelengths from the reaction of substrate **1** with WT CoCYP119. D) Change in absorbance over time of key wavelengths from the reaction of substrate **1** with T213G CoCYP119.

The reactivity of 9B12, and indeed many of the highly active variants from the master plate deallylation screen, showed similar modes of reactivity (Figure 4.45). To gain further insight into the structure of this species, we analyzed the product of the reaction with 9B12 by UPLC-MS.

Both starting material and the deallylated product were detected by UPLC. The starting material ionizes poorly, but a fragment ion consistent with loss of the allyl group (-m/z = 183, negative mode ionization) is observed for this peak. A new peak at 2.6 min is also observed in the 9B12 reaction, with a red-shifted absorbance (λ_{max} ~ 454 nm) and a -m/z = 185. The +2 mass relative to the starting material indicates that the product results from a net hydrogenation reaction. Indeed, when the reaction is conducted in D_2O, the product at 2.6 min has a -m/z = 186, which indicates that exactly one proton (or deuteron) is obtained from solvent. We subjected an alternative substrate, 2,4-dinitroanisole, to the reaction conditions and found that a similar brown to orange color transition was observed. UPLC analysis of this product again reveals a +2 mass with a different retention time (2.45 min), indicative of reduction of the arene ring. No orange color is seen when 2,4-dinitrophenol (**2**) is subjected to identical reaction conditions. We scaled up reactions with 2,4-dinitroanisole (50 mg) both is proto-buffer and in D_2O buffer, and purified their products for NMR analysis. NMR analysis suggests that the product of these reactions is a dearomatized 6-membered ring, in which both nitro groups are maintained. The product obtained from the deuterated buffer experiment reveals that a buffer-derived proton is appended to the carbon adjacent to the anisole group (Figure 4.46). Additionally, NMR analysis is consistent with de-alkylation of the anisole group. We hypothesize that this dealkylation (hydrolysis) occurs non-enzymatically, following enzymatic dearomitization.

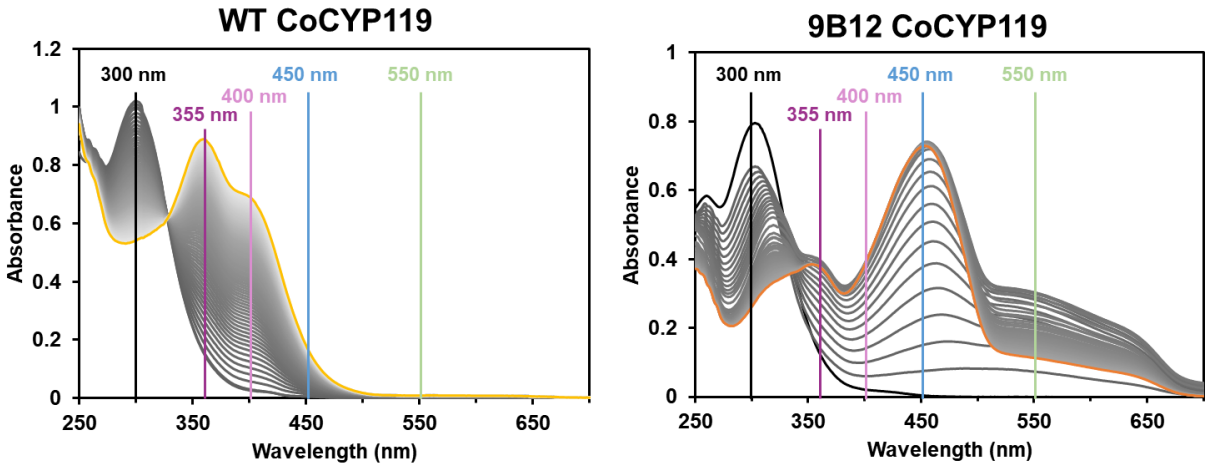


Figure 4.45. Progress spectra for the reaction of CoCYP119 and variants with model substrate **1**. Reaction conditions: 50 µM **1**, 200 µM DMPS, 1 µM (WT) or 0.2 µM (9B12) CoCYP119 in 100 mM KPi pH 6.0 at room temperature. The black trace represents the initial spectrum, grey lines represent the course of the reaction over time (every 3 min) and the yellow/orange lines represent the final trace of the product.

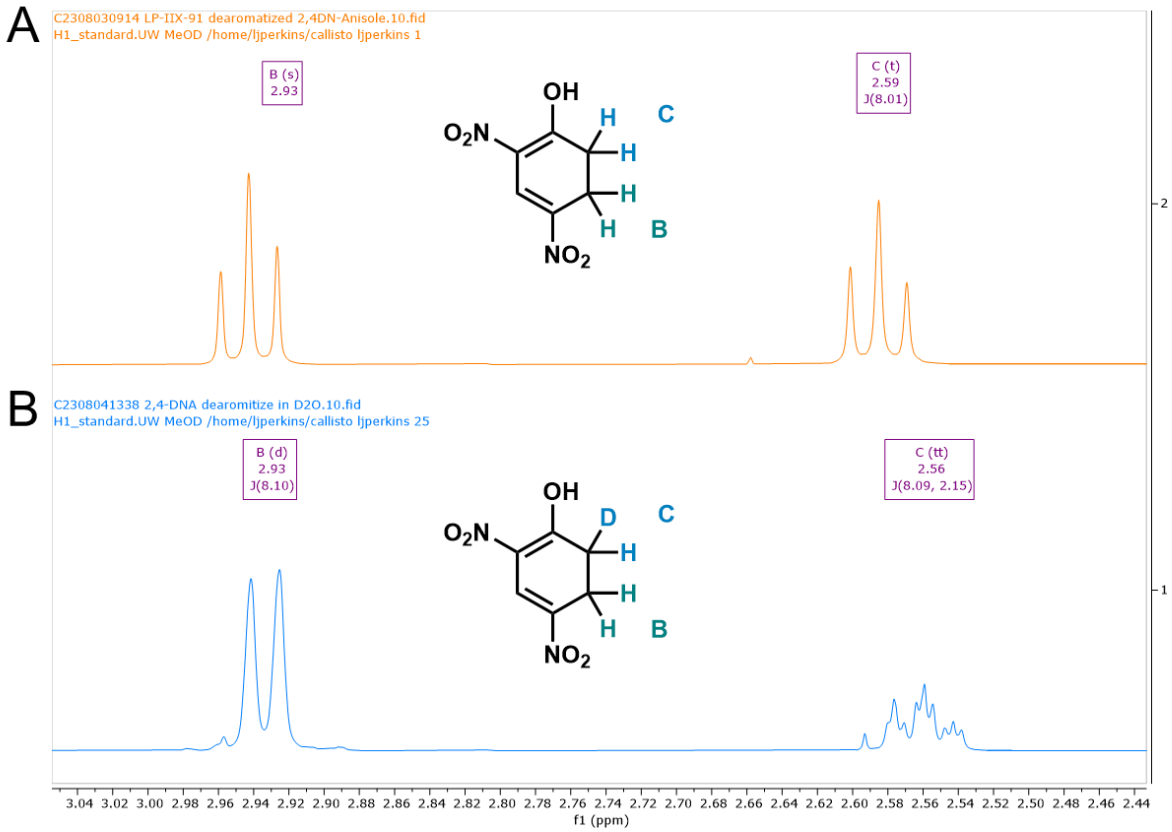


Figure 4.46. NMR Analysis of purified, dearomatized products of the reaction of 2,4-dinitroanisole with 9B12 CoCYP119 and DMPS in both proto (A) and deuterated (B) buffers. An additional singlet at ~8.67 ppm, corresponding to the vinylic proton, is also present in both spectra.

4. 3. 11. Spectroscopic investigations of improved variants

To better understand changes to the electronic structure of the cofactor and reactive intermediate in the improved variants, we once again turned to UV-visible spectroscopy. We collected spectra for WT, V254D and T213G point mutants, and the 14A2 and 9B12 variants at room temperature in KPi buffer at pH 6.0 with 10% acetonitrile. Spectra for the as-purified and dithionite-reduced CoCYP119 variants are given in Figure 4.51.

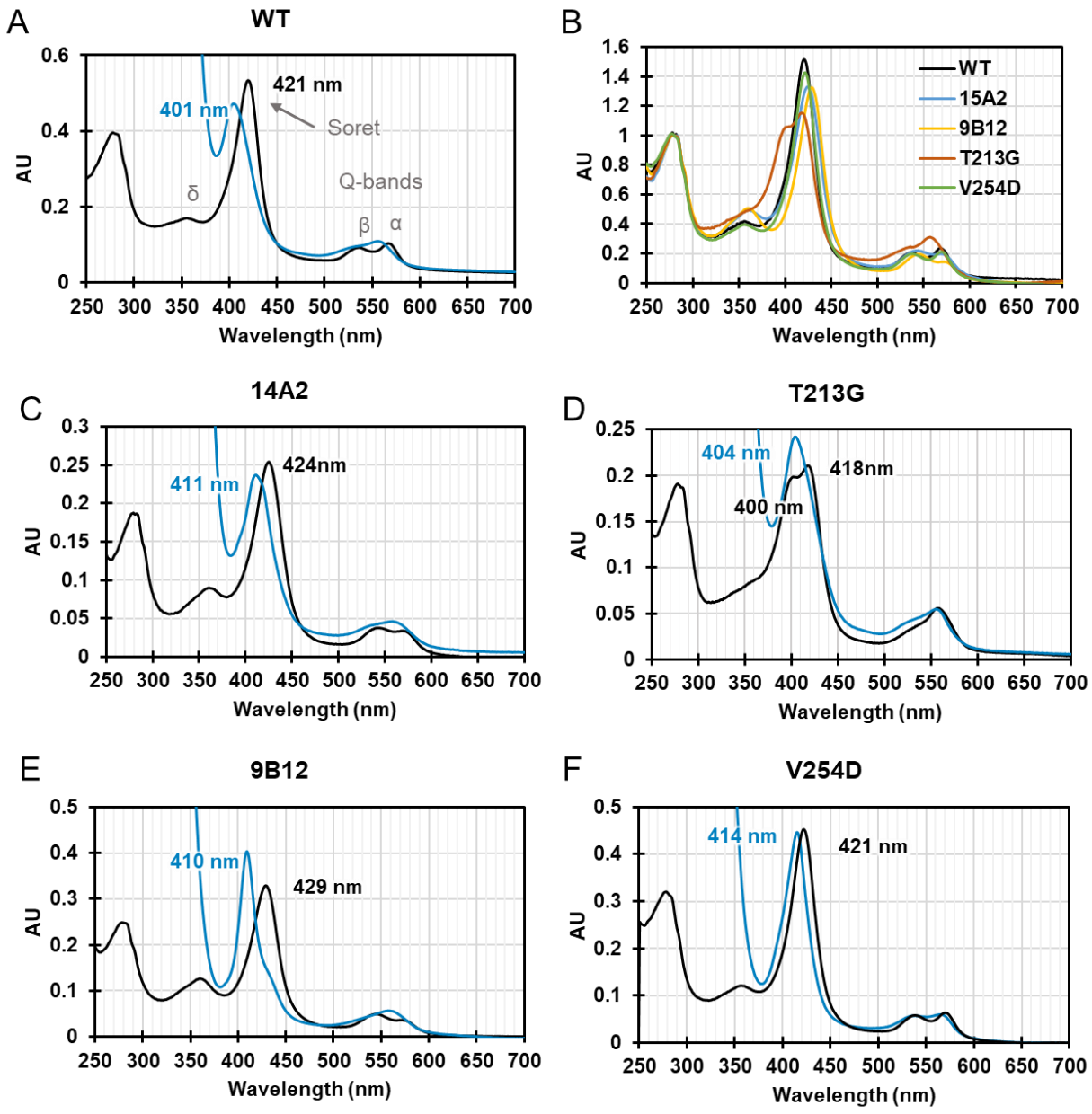


Figure 4.47. As-purified (black) and dithionite-reduced (blue) spectra of WT CoCYP119 and variants.

Figure 4.51B directly compares the UV-visible spectra of each variant as-purified, normalized by absorbance at 280 nm. The similarity between the WT and V254D spectra becomes apparent by this analysis, indicating that these enzymes may have similar electronic structures. For the other variants, unique spectral characteristics indicate electronic changes to

the metal center and porphyrin complex that have arisen as a result of mutation to the enzyme scaffold.

The maximal Co(III) Soret absorbance for each of the variants is substantially different from one another. 9B12 has the most red-shifted Soret compared to WT at 429 nm. T213G exhibits two peaks at the typical Soret wavelengths, at 418 and 400 nm. Pyridine hemachrome analysis confirmed that these two peaks were not the result of heme contamination in the protein sample, and this feature is consistent across multiple purifications of the T213G variant.

The shape of the Q bands, centered around 550 nm, is also significantly different in each of these samples. The relative intensity of the α (~570 nm) and β (~540 nm) bands is reversed in the 9B12 and 14A2 variants relative to the WT. Previous studies of the iron version of CYP119, have suggested that inversion of the α/β intensities and red-shifted Soret are diagnostic of transition to an inactive “P420” form of the enzyme, in which the proximal thiolate is protonated.[52] The cause of such features in the cobalt-substituted version of the protein has not been explored. T213G also exhibits a unique Q band spectral feature, in which the α-band is much higher in intensity than the broadened β-band shoulder.

Lastly, the δ-band feature, a defining characteristic of P450 enzymes which results from a ligand to charge transfer from the proximal thiolate residue,[53] differs between the variants. This feature is negligible in the T213G sample but has increased intensity in the 9B12 and 14A2 variants. Increased intensity of this feature may correlate with stronger orbital overlap between the thiolate and cobalt center, indicative of a stronger Co-S bond. The δ features of the 9B12 and 14A2 variants are also slightly redshifted relative to that of the WT enzyme.

Incubation of each variant with sodium dithionite (DTH) typically results in reduction of the cobalt center to Co(II). A blue-shift in the Soret absorbance is diagnostic of Co-substituted hemoproteins, and such a shift is clearly observed for all the variants tested, with the exception V254D. Even after heating with DTH, the V254D only exhibited a very small blue-shift in Soret absorbance and a slight broadening of the Q-band features. Further investigations are required in order to

determine whether the V254D variant is truly reduced to the Co(II) state by DTH. Reduction of the T213G cobalt causes the Soret to shift to a single, broad peak at 404 nm. It is not clear from electronic absorption spectroscopy if this broad peak represents two distinct Co(II) signals, or if the metal center is not fully reduced.

While it is clear that the electronic environment for each of these variants is distinct, the molecular determinants of these changes and their implications (e.g. change in redox potential) cannot be inferred from these spectra alone. To better understand these differences, other forms of spectroscopy, such as magnetic circular dichroism (MCD) or Electronic paramagnetic resonance spectroscopy (EPR) may be employed. Co(II) is EPR active, so EPR interrogations are the reduced form are possible, while interrogations of the Co(III) state are not. Metal substitution to the native iron version of these proteins could enable direct interrogation of the enzyme in a 3+ oxidation state. In addition, UV-visible spectroscopy of the heme versions of these proteins may be enlightening, as these spectra could be directly compared to known P450s and changes to electronic structure in variants could be inferred by analogy.

To collect information about the putative cobalt-hydride intermediate for each variant, Enzyme (2 μM) was incubated at 15°C in KPi pH 6.0 with 10% acetonitrile. PS (10 mM) or DMPS (0.5 mM) was added to each sample, and the spectra were taken every minute for a total of 20 minutes. The UV-visible spectra resulting from the addition of DMPS and PS to each variant are given in Figure 4.52, along with difference spectra.

These spectra are complex, and without definitive assignments of individual spectral features, it will be challenging to draw conclusions. Comparisons between variants, however, can help to form hypotheses to guide future studies. In the reaction with the WT CoCYP119 enzyme, the putative Co-H intermediate, as indicated by the hyperporphyrin-like absorption peaks at 381 and 470 nm, is short lived. After no more than an hour, the absorption spectrum of the WT enzyme is superimposable with the spectrum of the enzyme alone, indicating the enzyme has returned to its resting state. Similar behavior is observed for the V254D variant, although the intermediate

hyperporphyrin species is formed at a lower concentration, stymying efforts to quantify a lifetime. In contrast with this behavior, the 14A2 and 9B12 variant spectra before and after addition of silane are not superimposable. Both enzymes exhibit a decrease in their Soret absorbance, and an increase in absorbance near 400 nm. Such a shift may be indicative of reduction to a Co(II) state. Lastly, the T213G behavior is again distinct and will require additional future studies for authoritative interpret. The hyperpophryin species is relatively long lived, and as the hyperporphyrin species disappears, the two soret peaks reappear, although their relative intensities are changed from the resting state.

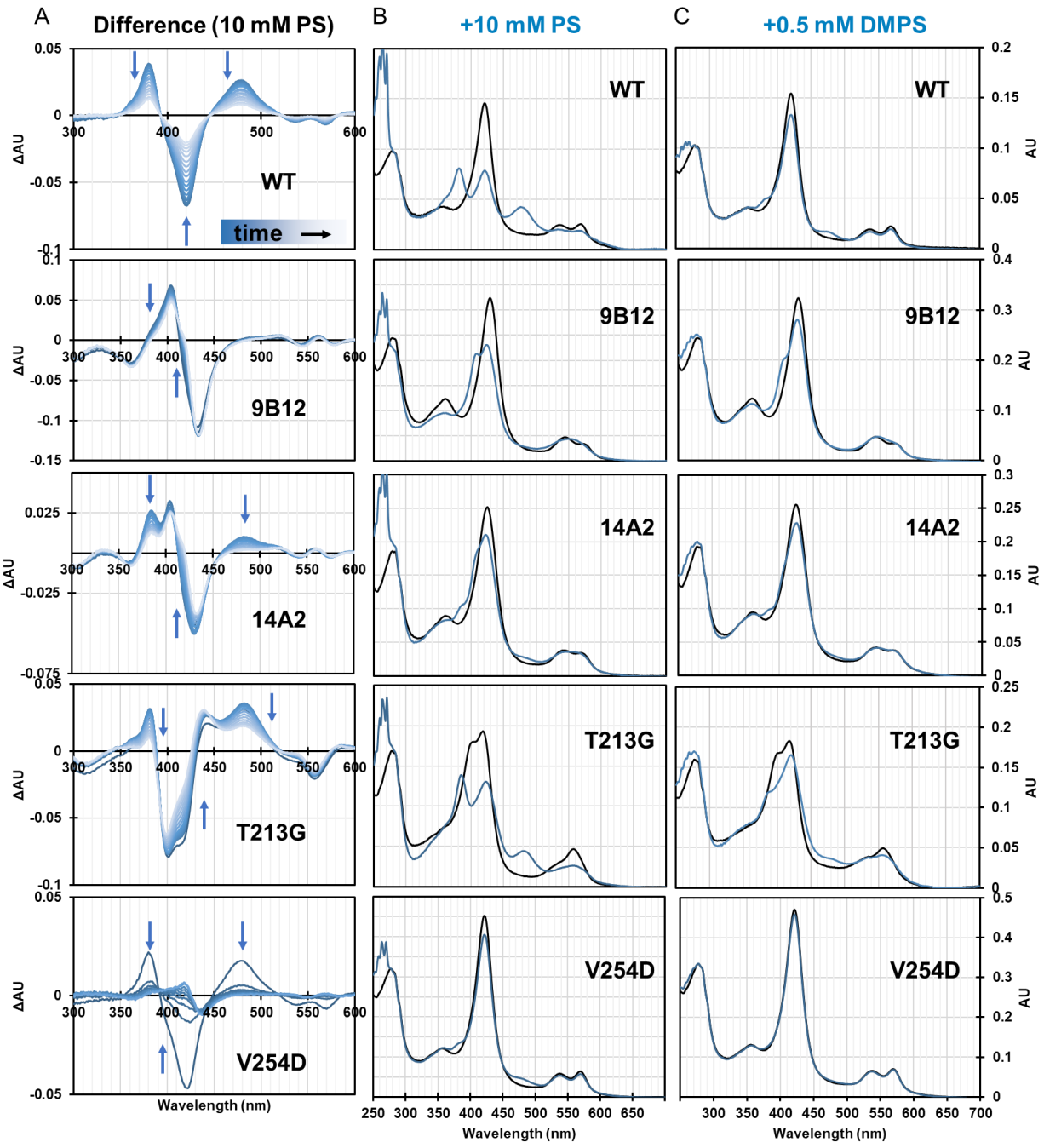


Figure 4.48. UV-visible spectra of CoCYP119 and variants with PS and DMPS added. Difference spectra upon addition of 10 mM PS are given in column A. The spectra with the greatest changes from the ground state spectra following addition of 10 mM PS are plotted with the as-purified (ground state) spectra in column B. Column C shows spectral changes upon addition of 0.5 mM DMPS.

4. 4. Discussion

MHAT is, to date, not known to exist in nature. Nevertheless, a biocatalytic version of MHAT may enable a range of stereo- and regio- selective radical coupling reactions. The challenges associated with implementing a complex MHAT reaction in an enzymatic system, when no known MHAT enzyme exists, prompted us to explore this mode of reactivity with a model deallylation reaction. Furthermore, directed evolution of these enzymes has provided us with several variants of CoCYP119 with improved activity and altered spectroscopic and catalytic properties. Our investigations of CoCYP119 catalyzed MHAT are ongoing, and this thesis chapter represents our progress to date towards understanding the mechanism of this reaction and how the enzyme scaffold tunes reactivity.

Spectroscopic and mechanistic evidence supports formation of a cobalt-hydride intermediate that resides in the enzyme active site and is responsible for the observed reactivity. The non-natural CoPPIX cofactor appears to be deterministic of deallylation of activity, as stand-alone CoPPIX is a competent, albeit poor, catalyst for this reaction. Neither the heme cofactor nor the heme-dependent protein are competent catalysts. The addition of a hydride donor to CoCYP119 imparts unique spectral shifts to the UV-visible spectrum of the enzyme, indicative of a strong donor ligand, such as hydride. The fact that these spectral features are identical, regardless of the hydride donor used, is further evidence of a cobalt hydride intermediate. Enzymatic CoPPIX-hydride intermediates have been inferred previously,[12] but to our knowledge, have not been spectroscopically characterized. Lastly, the formation of hydrogen gas in the absence of alkene substrate is additional evidence for a bona fide cobalt-hydride intermediate. Investigations into the electronic structure of this intermediate are still ongoing.

A hallmark of directed evolution is the ability to make changes to enzyme's activity with minimal a priori knowledge of the enzyme's mechanism. The MHAT reaction developed in this work is fundamentally different, both electronically and sterically, from the native monooxygenation reactivity of CYP119. The electronic properties of the new-to-to nature CoPPIX

cofactor likewise, are unique. A modest sampling of the mutational landscape (~1k variants) led to numerous improved variants, suggesting that there may be multiple mechanistic routes to improved activity for this reaction.

Our mechanistic understanding of how these variants have improved MHAT reactivity remains limited. Preliminary mechanistic investigations have identified three potential steps of MHAT deallylation in which variants may have improved behavior: Activation of the hydride donor to form a cobalt hydride, decreasing the rate of a competing hydrogen evolution pathway, and HAT to the alkene substrate. We have identified variants for which HAT is rate limiting (V254D and the WT enzyme). We have also identified at least one variant (9B12) with increased catalytic efficiency with DMPS.

Previous studies of cobalt catalysts for Mukaiyama hydration reactions (a subclass of MHAT reactions) have shown that a narrow window of redox potentials enables MHAT chemistry.[25] Indeed, a competent MHAT catalyst must be both electron poor enough to accept a strong hydride ligand, and sufficiently electron rich to enable homolysis of the cobalt hydride bond to yield Co(II), rather than heterolytic cleavage to yield Co(III). Thus, tuning the redox potential of the cobalt center is likely to have a large impact on the reactivity. While we have not explicitly measured the reduction potential of CoCYP119 or variants, the differences in the ground state and dithionite reduced UV-visible spectra of these proteins indicate that the electronic structure of the cofactor has changed significantly (Figure 4.51). Further spectroscopic investigations of these variants are warranted, to better understand the electronic structure of their reactive intermediates.

4. 5. Conclusion and future directions

We have demonstrated the directed evolution of a protein bearing a non-natural cofactor (CoPPIX) for a new-to-nature mode of reactivity (MHAT). This engineering campaign was enabled by the *de novo* biosynthesis of a CoPPIX cofactor and incorporation of this cofactor into hemoprotein *in vivo*, and the development of a high throughput colorimetric screen for MHAT based on *O*-deallylation. While the synthetic utility of this deallylation reaction is currently limited, we envision that variants which are activated for MHAT deallylation may have improved activity for other MHAT reactions as well. We will therefore continue to engineer CoCYP119 for MHT deallylation, investigate the molecular determinants of improved reactivity, and test other MHAT reactions with improved variants.

4. 6. Materials and Methods

4. 6. 1. General experimental methods

All chemicals and chemical standards were purchased from commercial suppliers (Sigma-Aldrich, VWR, Goldbio, Frontier biosciences, Fluka), and used without further purification. Casamino acids were obtained from Biolabs or Research Products International. Unless otherwise noted, all media and solutions were prepared using ultrapure water (≥18 MΩ, from Thermo Scientific Barnstead Nanopure water purification system). Immobilized metal affinity chromatography resin or columns were purchased from GE Healthcare.

Equipment and instrumentation

New Brunswick I26R, 120 V/60 Hz shakers (Eppendorf) were used for cell growth. When needed, the humidity was controlled with a HumidiKit™ Auto Humidity System for Incubators purchased from IncubatorWarehouse.com. Cell disruption was accomplished via sonication with a Sonic Dismembrator 550 (Fischer Scientific) sonicator. Electroporation (for transformation and cloning) was achieved using an Eppendorf E-porator set to 2500V. Optical density measurements of liquid cultures were recorded on a Ultrospec 10 cell density meter (Amersham Biosciences). An Envision® 2105 multimode plate reader (Perkin Elmer) was used to measure optical density in 96-well plates. UPLC/MS data were collected on an Acquity UHPLC with an Acquity QDA MS detector (Waters). Electronic absorption data were collected on a UV-2600 Shimadzu spectrophotometer, set to a spectral bandwidth of 0.5 nm.

4. 6. 2. Expression and purification of CoCYP119 and variants

BL21(DE3) electrocompetent cells were freshly transformed with the CYP119 gene on an IPTG-inducible Pet22b and recovered in Terrific Broth (TB) for 45 min at 37°C. Cells were then plated onto Luria-Burtani (LB) plates with 100 µg/mL ampicillin (AMP) and incubated overnight. Single colonies were used to inoculate 10 mL TB + 100 µg/mL AMP, which were grown overnight at 37 °C and 200 rpm. Expression cultures, typically 1 L of TB-AMP in a 2.8 L baffled Fernbach flask, were inoculated from these starter cultures and shaken at 37 °C and 220 rpm. After 3.5 hours, the expression cultures were chilled on ice. After 30 min. on ice, CYP119 expression was induced

with 0.5 mM IPTG and CoPPIX production was promoted with 0.25 mM δ-aminolevulinic acid (δALA) and 1 mM $CoCl_2$. The cultures were incubated overnight at 25 °C and 180 rpm. Cells were then harvested by centrifugation at 4,000xg for 15 min at 25 °C. Cell pellets were usually pinkish-brown in color. These pellets were frozen and stored at -20 °C until lysis.

To purify CoCYP119, cell pellets were thawed and resuspended in lysis buffer (100 mM potassium phosphate buffer (pH = 6.0), 100 mM NaCl, 1 mg/mL hen egg white lysozyme (GoldBio), 0.2 mg/mL DNase I (GoldBio), 2 mM MgCl2, ~14 μM CoPPIX in DMSO). A volume of 4 mL of lysis buffer per 1 gram of wet cell pellet was used. After 45 min. of shaking at 37 °C and 200 rpm, the cells heat treated at 65°C for 15-20 min. The resulting lysate was clarified of cellular debris by centrifugation at 30,000xg for 15 min, resulting in a red-orange supernatant. The pH of the supernatant was adjudted to 7.5 with 1.0 M NaOH, and the supernatant was applied to a gravity column of Ni/NTA beads (GoldBio) that were equilibrated with 100 mM potassium phosphate buffer (pH = 8.0). The column was washed with approximately 3 column volumes of 50 mM imidazole, 100 mM NaCl, 100 mM potassium phosphate buffer (pH = 8.0). CoCYP119 was eluted with 250 mM imidazole, 100 mM NaCl, 100 mM potassium phosphate buffer (pH = 8.0). Elution of the desired protein was monitored by the disappearance of the red color (resulting from the release of CoCYP119) from the column. The eluent was dialyzed into storage buffer (100 mM potassium phosphate buffer, pH = 8.0) overnight at 4 °C, at least 4 times to remove excess imidazole. Dialyzed protein solutions were concentrated to roughly 0.5 mM using Amicon® Ultra 15 mL Centrifugal Filters (Millipore) at a molecular weight cutoff of 10 kDa. Purified enzyme was flash frozen in pellet form by pipetting enzyme dropwise into a crystallization dish filled with liquid nitrogen. The enzyme pellets were transferred to a plastic conical and stored at -80 °C until further use. Generally, this procedure yielded about 20 mg per L culture.

4. 6. 3. General cloning method

PCR amplification was used to introduce the mutation at each sites using primers with degenerate codons. The resulting gene fragments were stitched together by PCA, purified by gel electrophoresis, and extracted using the Macherey-Nagel Gel DNA Recovery Kit. The resulting DNA fragments was inserted into the pET22b vector by the Gibson Assembly method and E. coli cells were transformed with the resulting cyclized DNA product by electroporation. After 45 min. of recovery in TB media at 37 °C and 200 rpm, cells were plated onto LB plates with 100 µg/mL AMP and incubated overnight. Single colonies were used to inoculate 6 mL TB + 100 µg/mL AMP, which were grown overnight at 37 °C and 200 rpm. Plasmid was extracted and purified from these starter cultures using the Macherey-Nagel Plasmid Miniprep Kit, and the gene inserts were sequenced by Functional Biosciences.

4. 6. 4. General procedure for library screening.

A 96-deep well plate containing TB + 100 ug/mL AMP (500 uL) was inoculated by a single colony in each well and starter plates were grown overnight at 37 °C and 200 rpm. Expression cultures were inoculated in a fresh 96-deep well plate containing TB + 100 ug/mL AMP (650 uL) using starter culture (50 uL) and grown at 37 °C and 200 rpm for 4.5 hours. Expression cultures were chilled on ice for 30 min. and then induced with 50 µL of a 3.75 mM δALA, 15 mM $CoCl_2$, 15 mM IPTG in H_2O stock (final concentrations of 0.25 mM, 1 mM, and 1 mM, respectively). Cultures were expressed overnight at 25 °C and 180 rpm. The plate was spun down at 4,000xg for 15 min. at 25 °C and the pellets were frozen at -20 °C until use.

Before screening, frozen cell pellets in a 96-deep well plate were thawed and resuspended in 400 uL 100 mM potassium phosphate buffer (pH = 6.0). After applying an Alumaseal sealing film (VWR) to the plate with a rubber roller (VWR), the plate was heat treated in a water bath at 65 °C

for 15 min. The resulting lysate was clarified of cellular debris by centrifugation at 4,000xg at 25 °C for 20 min.

First round of directed evolution- Screening with PS

An organic master mix was created by adding 25 mM PS and 2.5 mM (**1**) to DMSO (final volume of 6,250 uL). A flat bottom polystyrene 96-well microplate (Corning) was charged with 50 uL organic master mix in each well. After removing the sealing film from the 96-deep well plate, the reaction was initiated by the addition of 200 uL clarified lysate to each well (5 mM PS, 0.5 mM (2) for final concentrations, 20% DMSO). Reaction plate was sealed with aluminum sealing film and absorbance at 405 nm was recorded after 1 hours with a UV-plate reader.

Subsequent rounds of directed evolution- Screening with DMPS

Organic master mix was assembled consisting of 10 mM DMPS, 10 mM **1** in Acetonitrile. 25 µL of this mixture was added to each well of a 96-well plate that had been charged with 200 µL of heat treated CoCYP119 lysate (pH 6.0), plus 25 µL 100 mM KPi pH = 6.0. Reaction plate was sealed with aluminum sealing film and absorbance at 405 nm was recorded after 20 hours with a UV-plate reader.

4. 6. 5. Synthesis and purification of starting materials

*Preparation of 1-((2-methylallyl)oxy)-2,4-dinitrobenzene (**1**):*

A 100 mL round bottom flask was charged with 1-fluoro-2,4-dinitrobenzene (186 g, 1 mmol), 5 mL of dichloromethane, and a magetic sitr bar. To this solution, 2-methylprop-2-en-1-ol (244 µL, 2.2 mmol) and triethylamine (140 µL, 1.0 mmol) were added and the reaction was stirred at room temperature overnight. After 16 h., the solvent was evaporated and the resulting oil was resuspended in ethylacetate. The solution was washed thrice with brine. The organic layer was

dried with magnesium sulfate and the solvent was evaporated under vaccuum. The resulting residue was purified via silica column chromatography (Ethylacetate/hexane) using a biotage flash purification system, and **1** was obtained as a yellow oil which solidified to a waxy yellow solid over time, (190 mg, 80% yield). **^{1}H NMR** (400 MHz, $CDCl_3$) δ = 8.76 (d, J = 2.8 Hz, 1H), 8.41 (dd, J = 9.3, 2.8 Hz, 1H), 7.19 (d, J = 9.3 Hz, 1H), 5.17 (p, J = 1.2 Hz, 1H), 5.10 (p, J = 1.3 Hz, 1H), 4.70 (s, 2H), 2.04 (s, 1H), 1.86 (t, J = 1.1 Hz, 3H). Trace triethylamine observed as δ = 4.12 (q, J = 7.1 Hz), 1.26 (t, J = 7.1 Hz).

*Preparation of 1-(allyl)oxy)-2,4-dinitrobenzene (**3**):*

A 100 mL round bottom flask was charged with 1-fluoro-2,4-dinitrobenzene (930 g, 5 mmol), 5 mL of dichloromethane, and a magetic sitr bar. To this solution, prop-2-en-1-ol (1 mL, 14 mmoll) and triethylamine (900 µL, 6.0 mmol) were added and the reaction was stirred at room temperature overnight. After 16 h., the solvent was evaporated and the resulting oil was resuspended in ethylacetate. The solution was washed thrice with brine. The organic layer was dried with magnesium sulfate and the solvent was evaporated under vaccuum. The resulting residue was purified via silica column chromatography (Ethylacetate/hexane) using a biotage flash purification system, and **1** was obtained as a yellow oil which solidified to a waxy yellow solid over time, (913 mg, 83% yield). **^{1}H NMR** (500 MHz, $CDCl_3$) δ 8.78 (d, *J* = 2.8 Hz, 1H), 8.44 (dd, *J* = 9.3, 2.8 Hz, 1H), 7.22 (d, *J* = 9.2 Hz, 1H), 6.06 (ddt, *J* = 17.3, 10.4, 5.0 Hz, 1H), 5.55 (dq, *J* = 17.2, 1.6 Hz, 1H), 5.45 (dq, *J* = 10.7, 1.4 Hz, 1H), 4.85 (dt, *J* = 5.0, 1.6 Hz, 2H), 1.56 (s, 1H).

4. 6. 6. General procedure for analytical scale reactions and UV-vis detection of product

Organic solvent (acetonitrile or DMSO) and allyl substrate were added to a quartz UV cuvette. The appropriate volume of buffer and enzyme was added and the cuvette was inverted to mix. Silane, suspended in organic solvent, was added to initiate the reaction, and the cuvette was once

again inverted to mix. Product was detected via absorbance at 400 nm relative the absorbance at 700 nm, to adjust for any changes to the baseline measurement.

4. 6. 7. Supplemental Figures

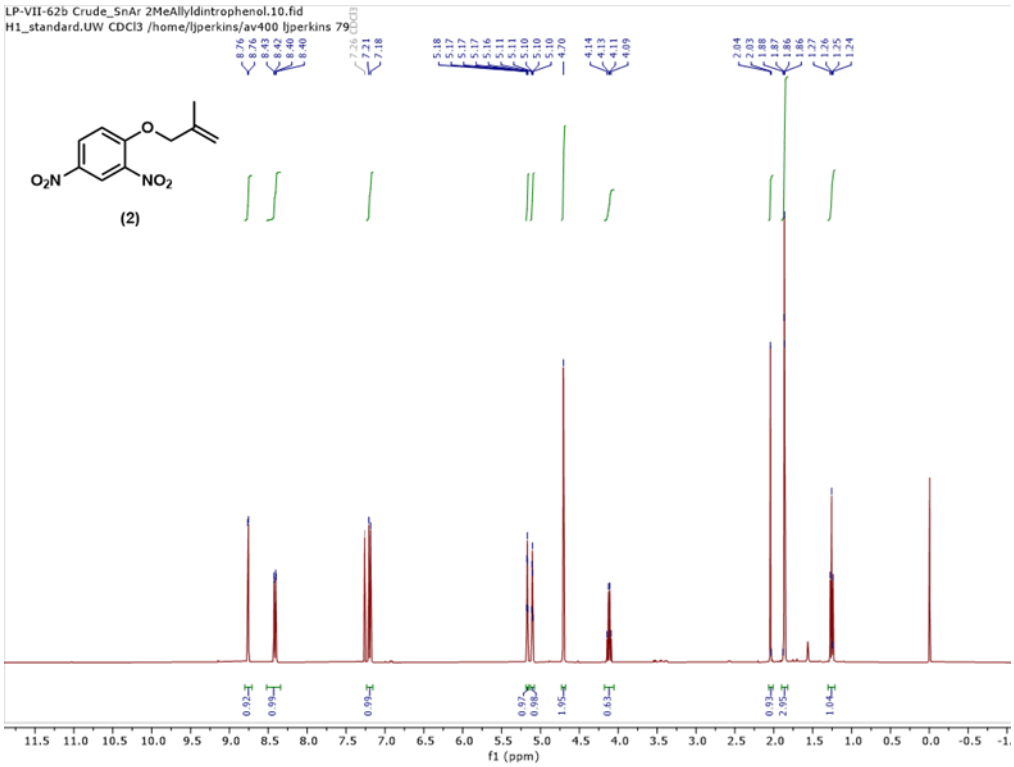


Figure 4.49. ^{1}H NMR of substrate **1**, 1-((2-methylallyl)oxy)-2,4-dinitrobenzene.

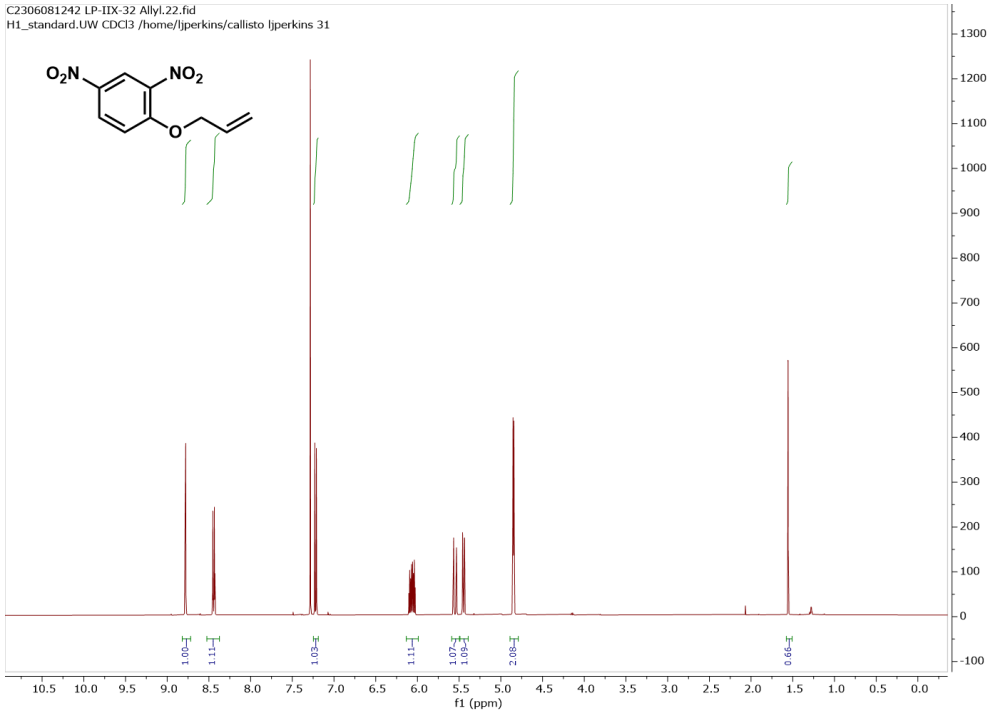


Figure 4.50. ^{1}H NMR of substrate **1**, 1-((allyl)oxy)-2,4-dinitrobenzene.

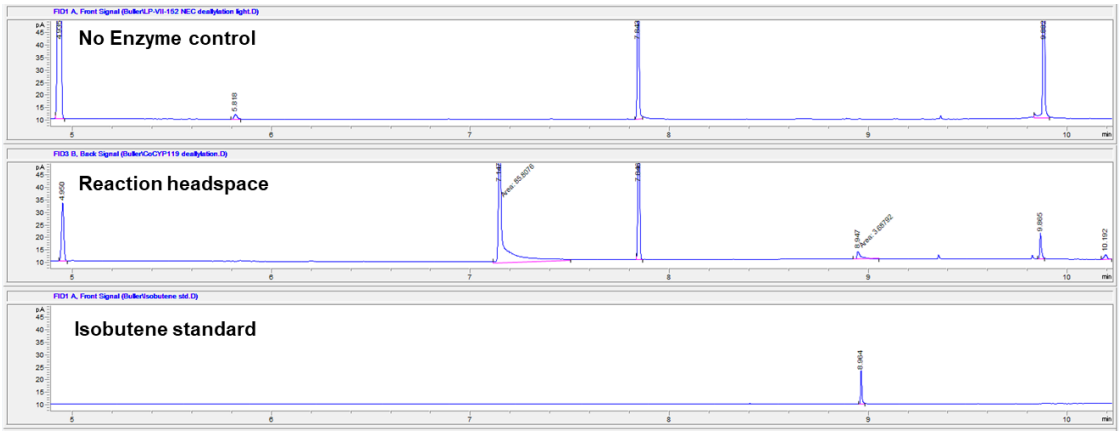


Figure 4.51. Detection of isobutene in the headspace of a reaction with **1** and CoCYP119. Reaction conditions: 2 mM DMPS, 1 mM substrate **1**, 10 µM WT CoCYP119 in KPi pH 6.0 with 20% DMSO. Reaction was run overnight at room temperature in a sealed 2 mL crimp vial. 500 µL of Ethyl acetate was added to the sealed vial to quench, and the organic supernatant was analyzed by GC.

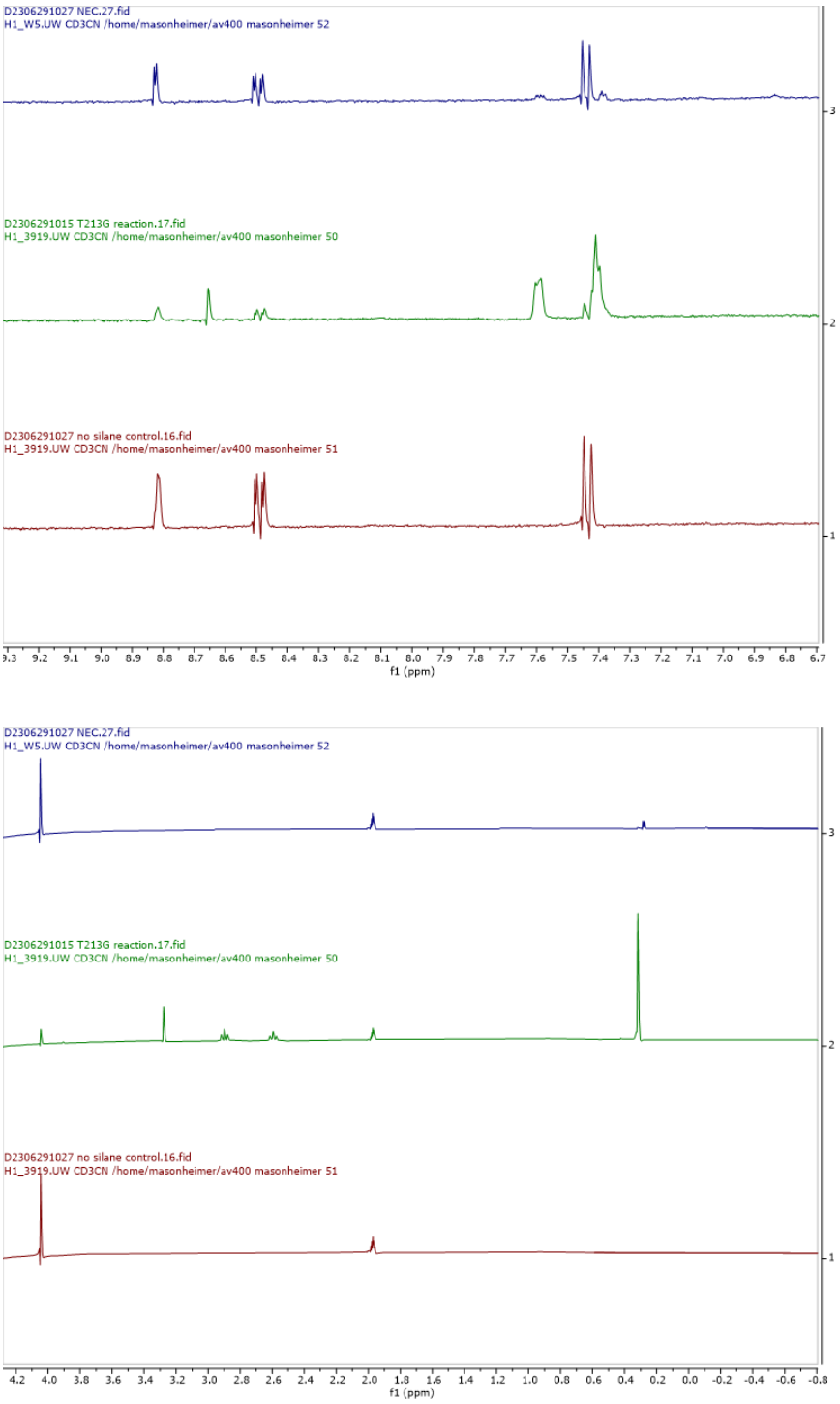


Figure 4.52. Crude NMR analysis of unknown product resulting from reaction of 2,4-dinitroanisole with the T213G CYP119 variant. Reaction conditions: 10 mM 2,4-dinitroanisole, 10 mM DMPS, 10% Acetonitrile-d_3 in 100 mM KPi pH 6.0. Reaction ran overnight at room temperature in an NMR tube, and NMR was taken with a 400 MHz spectrometer using a H1_3919 pulse program for H_2O solvent suppression. Blue trace = No enzyme control. Green trace = T213G CoCYP119 (50 µM). Red trace = No silane added control.

4. 7. References

1. Pearson, R. G. The Transition-Metal-Hydrogen Bond. *Chem. Rev.* 41–49 (1985).

2. Green, S. A. *et al.* The High Chemofidelity of Metal-Catalyzed Hydrogen Atom Transfer. *Acc. Chem. Res.* **51**, 2628–2640 (2018).

3. Li, G. *et al.* Radical Isomerization and Cycloisomerization Initiated by H• Transfer. *J. Am. Chem. Soc.* **138**, 7698–7704 (2016).

4. Ma, X. & Herzon, S. B. Intermolecular Hydropyridylation of Unactivated Alkenes. *J. Am. Chem. Soc.* **138**, 8718–8721 (2016).

5. Gnaim, S. *et al.* Cobalt-electrocatalytic HAT for functionalization of unsaturated C–C bonds. *Nat. 2022 6057911* **605**, 687–695 (2022).

6. Crossley, S. W. M., Obradors, C., Martinez, R. M. & Shenvi, R. A. Mn-, Fe-, and Co-Catalyzed Radical Hydrofunctionalizations of Olefins. *Chem. Rev.* **116**, 8912–9000 (2016).

7. Shevick, S. L. *et al.* Catalytic hydrogen atom transfer to alkenes: A roadmap for metal hydrides and radicals. *Chem. Sci.* **11**, 12401–12422 (2020).

8. Schilter, D., Camara, J. M., Huynh, M. T., Hammes-schi, S. & Rauchfuss, T. B. Hydrogenase Enzymes and Their Synthetic Models : The Role of Metal Hydrides. *Chem Rev* **116**, 8693–8749 (2016).

9. Jost, M., Simpson, J. H. & Drennan, C. L. The transcription factor CarH safeguards use of adenosylcobalamin as a light sensor by altering the photolysis products. *Biochemistry* **54**, 3231–3234 (2015).

10. Kutta, R. J. *et al.* The photochemical mechanism of a B 12-dependent photoreceptor protein. *Nat. Commun.* **6**, (2015).

11. Kleingardner, J. G., Kandemir, B. & Bren, K. L. Hydrogen Evolution from Neutral Water under Aerobic Conditions Catalyzed by Cobalt Microperoxidase-11. *J. Am. Chem. Soc.* **136**, 4–7 (2014).

12. Firpo, V., Le, J. M., Pavone, V., Lombardi, A. & Bren, K. L. Hydrogen evolution from water catalyzed by cobalt-mimochrome VI*a, a synthetic mini-protein. *Chem. Sci.* **9**, 8582–8589 (2018).

13. Kandemir, B., Chakraborty, S., Guo, Y. & Bren, K. L. Semisynthetic and Biomolecular Hydrogen Evolution Catalysts. *Inorg. Chem.* **55**, 467–477 (2016).

14. Harmer, J. *et al.* A nickel hydride complex in the active site of methyl-coenzyme M reductase: Implications for the catalytic cycle. *J. Am. Chem. Soc.* **130**, 10907–10920 (2008).

15. Wongnate, T. *et al.* The radical mechanism of biological methane synthesis by methylcoenzyme M reductase. *Science (80-.).* **352**, 953–958 (2016).

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

17. Hoffnagle, A. M. & Tezcan, F. A. Atomically Accurate Design of Metalloproteins with Predefined Coordination Geometries. (2023) doi:10.1021/jacs.3c04047.

18. Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485**, 185–194 (2012).

19. Yang, Y. & Arnold, F. H. Navigating the Unnatural Reaction Space: Directed Evolution of Heme Proteins for Selective Carbene and Nitrene Transfer. *Acc. Chem. Res.* **54**, 1209–1225 (2021).

20. Hammer, S. C., Knight, A. M. & Arnold, F. H. Design and evolution of enzymes for non-natural chemistry. *Curr. Opin. Green Sustain. Chem.* **7**, 23–30 (2017).

21. Tsay, C. & Yang, J. Y. Electrocatalytic Hydrogen Evolution under Acidic Aqueous Conditions and Mechanistic Studies of a Highly Stable Molecular Catalyst. *J. Am. Chem. Soc.* **138**, 14174–14177 (2016).

22. Matos, J. L. M. *et al.* Cycloisomerization of Olefins in Water. *Angew. Chemie - Int. Ed.* **59**, 12998–13003 (2020).

23. Nebergall, W. H. Some Reactions of Phenylsilane. *J. Am. Chem. Soc.* **72**, 4702–4704 (1950).

24. Lo, J. C., Yabe, Y. & Baran, P. S. A practical and catalytic reductive olefin coupling. *J. Am. Chem. Soc.* **136**, 1304–1307 (2014).

25. Mukaiyama, T. & Yamada, T. Recent Advances in Aerobic Oxygenation. *Bull. Chem. Soc. Jpn.* **68**, 17–35 (1995).

26. Tsumaki, T. Nebenvalenzringverbindungen. IV. Über einige innerkomplexe Kobaltsalze der Oxyaldimine. *Bull. Chem. Soc. Jpn.* **13**, 252–260 (1938).

27. Liu, J. *et al.* Metalloproteins Containing Cytochrome, Iron–Sulfur, or Copper Redox Centers. *Chem. Rev.* **114**, 4366–4469 (2014).

28. Coelho, P. S. *et al.* A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo. *Nat. Chem. Biol. 2013 98* **9**, 485–487 (2013).

29. Bren, K. L. Going with the Electron Flow: Heme Electronic Structure and Electron Transfer in Cytochrome c. *Isr. J. Chem.* **56**, 693–704 (2016).

30. Edwards, E. H., Jelušić, J., Chakraborty, S. & Bren, K. L. Photochemical hydrogen evolution from cobalt microperoxidase-11. *J. Inorg. Biochem.* **217**, (2021).

31. Giedyk, M., Goliszewska, K. & Gryko, D. Vitamin B12 catalysed reactions. *Chem. Soc. Rev.* **44**, 3391–404 (2015).

32. Wu, X. *et al.* Intercepting Hydrogen Evolution with Hydrogen-Atom Transfer: Electron-

Initiated Hydrofunctionalization of Alkenes. *J. Am. Chem. Soc* **144**, 41 (2022).

33. Wagner, G. C., Gunsalus, I. C., Wang, M.-Y. R. & Hoffman, B. M. Cobalt-substituted cytochrome P-450cam. *J. Biol. Chem.* **256**, 6266–6273 (1981).

34. Krajewski, A. E. & Lee, J. K. Nucleophilicity and Electrophilicity in the Gas Phase: Silane Hydricity. *J. Org. Chem.* **87**, 1840–1849 (2022).

35. Yin, Y.-N., Liu, H.-Y., Ouyang, D.-C., Zhang, Q. & Zhu, R. Efficient O-deallylation triggered by cobalt hydride-catalyzed oxidative hydrofunctionalization. *Green Synth. Catal.* **4**, 64–66 (2023).

36. Hossen, M. & Hossain, M. Spectrophotometric method for determination of dissociation constant of Spectrophotometric Method for Determination of Dissociation Constant of. (2016).

37. 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).

38. Gu, Y. *et al.* Directed Evolution of Artificial Metalloenzymes in Whole Cells. *Angew. Chemie - Int. Ed.* **61**, (2022).

39. Obradors, C., Martinez, R. M. & Shenvi, R. A. Ph(i-PrO)SiH2: An Exceptional Reductant for Metal-Catalyzed Hydrogen Atom Transfers. *J. Am. Chem. Soc.* **138**, 4962–4971 (2016).

40. Koo, L. S. *et al.* The active site of the thermophilic CYP119 from Sulfolobus solfataricus. *J. Biol. Chem.* **275**, 14112–14123 (2000).

41. Sreenilayam, G., Moore, E. J., Steck, V. & Fasan, R. Metal Substitution Modulates the Reactivity and Extends the Reaction Scope of Myoglobin Carbene Transfer Catalysts. *Adv Synth Catal* **359**, 2076–2089 (2017).

42. Wagner, G. C., Gunsalus, I. C., Wang, M.-Y. R. & Hoffman, B. M. Cobalt-substituted Cytochrome P-450$_{cam}$. *J. Biol. Chem.* **256**, 6266–6273 (1981).

43. Sono, M., Andersson, L. A. & Dawson, J. H. Sulfur donor ligand binding to ferric cytochrome P-450-CAM and myoglobin. Ultraviolet-visible absorption, magnetic circular dichroism, and electron paramagnetic resonance spectroscopic investigation of the complexes. *J. Biol. Chem.* **257**, 8308–8320 (1982).

44. Sakurai, H., Ishizu, K., Sugimoto, H. & Gunsalus, I. C. Model of the coordination site for cobalt-substituted cytochrome P450cam. *J. Inorg. Biochem.* **26**, 55–62 (1986).

45. Steinhoff, B. A., Guzei, I. A. & Stahl, S. S. Mechanistic characterization of aerobic alcohol oxidation catalyzed by Pd(OAc)2/pyridine including identification of the catalyst resting state and the origin of nonlinear [catalyst] dependence. *J. Am. Chem. Soc.* **126**, 11268–11278 (2004).

46. Li, C. & Hoffman, M. Z. One-Electron Redox Potentials of Phenols in Aqueous Solution. *J. Phys. Chem. B* **103**, 6653–6656 (1999).

47. MAYHEW, S. G. The Redox Potential of Dithionite and SO−2 from Equilibrium Reactions with Flavodoxins, Methyl Viologen and Hydrogen plus Hydrogenase. *Eur. J. Biochem.* **85**, 535–547 (1978).

48. Nishida, C. R. & Ortiz De Montellano, P. R. Thermophilic cytochrome P450 enzymes. *Biochem. Biophys. Res. Commun.* **338**, 437–445 (2005).

49. Wangler, A. *et al.* Co-solvent effects on reaction rate and reaction equilibrium of an enzymatic peptide hydrolysis. *Phys. Chem. Chem. Phys.* **20**, 11317–11326 (2018).

50. Jr, M. R. D. *et al.* solvent. **57**, 4263–4275 (2019).

51. Johnson, K. A. *Kinetic Analysis for the New Enzymology*. (Kintek Corporation, 2019).

52. Tschirret-Guth, R. A., Koo, L. S., Hui Bon Hoa, G. & Ortiz De Montellano, P. R. Reversible pressure deformation of a thermophilic cytochrome P450 enzyme (CYP119) and its active-site mutants. *J. Am. Chem. Soc.* **123**, 3412–3417 (2001).

53. Dawson, J. H. & Sono, M. Cytochrome P-450 and Chloroperoxidase: Thiolate-Ligated Heme Enzymes. Spectroscopic Determination of Their Active Site Structures and Mechanistic Implications of Thiolate Ligation. *Chem. Rev.* **87**, 1255–1276 (1987).

Chapter 5

Radical hydrofunctionalization reactions by CoCYP119

This research was conducted with experimental and intellectual contributions from Carly L. Masonheimer (C. L. M.)

Chapter 5: Radical hydrofunctionalization reactions by CoCYP119

5. 1. Abstract

Biocatalytic metal hydrogen atom transfer (MHAT) reactions have the potential to overcome many challenges in asymmetric catalysis and green chemistry. In this chapter, we explore the ability of CoCYP119 to carry out a number of classic MHAT hydrofunctionalization reactions, as well as carbonyl reduction. We show that the MHAT reactivity of CoCYP119 can be tuned through directed evolution, and lay the groundwork for the development of synthetically useful biocatalytic MHAT reactions.

5. 2. Introduction

Metal-hydrogen atom transfer (MHAT) reactions, have been leveraged for numerous hydrofunctionalizations, isomerization, and radical coupling reactions.[1–10] Select examples of these transformations are given in Figure 5.1. Despite the synthetic utility of this class of reactions, no biocatalytic MHAT reaction have yet been developed. A biocatalytic MHAT may have improved selectivity compared to synthetic approaches, and the tunability of the enzyme scaffold may enable improved catalyst-controlled selectivity. Our goal is to eventually leverage these reactions to achieve transformations that are not yet accessible to the synthetic MHAT toolbox.

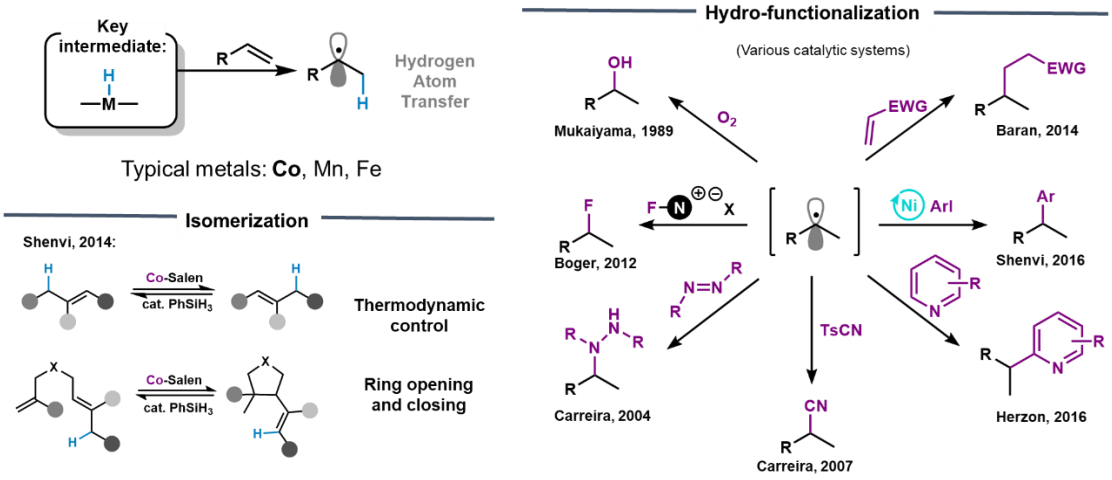


Figure 5.1. Examples of metal hydrogen atom transfer (MHAT) hydrofunctionalizations and isomerization reactions.

In the previous chapter, we developed a platform of enzymatic MHAT based on a new-to-nature CoPPIX cofactor, which we engineered for improved MHAT-deallylation activity. This deallylation reaction enabled us to study the mechanism and track improvements throughout evolution. In an effort to demonstrate the synthetic utility of biocatalytic MHAT, we have, in parallel, tested a variety of classic MHAT type reactions with the CoCYP119 system, including isomerization, cycloisomerization, and Mukaiyama hydration. Currently, most of these reactions are low yielding, and result in mixtures of products. Preliminary screens for these reactions with the master plate of variants developed in Chapter 4 demonstrate that we may be to tune the product profile of these reactions through directed evolution.

5. 3. CoCYP119 catalyzed reactions

5. 3. 1. Ketone reduction

Enzymatic metal hydrides have been used previously for the enantioselective reduction of ketones.[11,12] We were interested in exploring this reactivity in a CoCYP119 system to 1) further substantiate that CoCYP119 forms a metal hydride intermediate in the active site of the enzyme and 2) understand how the enzyme tunes hydrogen atom transfer versus hydride reduction activity.

We tested the ability of CoCYP119, FeCYP119 and their respective cofactors for the reduction of acetophenone (Figure 5.2B). We detected no activity for the iron version of the cofactor and enzyme. Instead, similar to our findings with deallylation, we observed that the CoCYP119 was more efficient than the free cofactor. We used chiral super-critical fluid chromatography (SFC) to analyze the products, which showed that the reduction occurs enantioselectivity. When the same reaction was probed with the T213A variant of CoCYP119, we found that the reaction occurs with opposite enantioselectivity to the WT enzyme. These results establish enzyme-mediated reduction within the active site of the enzyme, but do not themselves indicate the mechanism of this reduction.

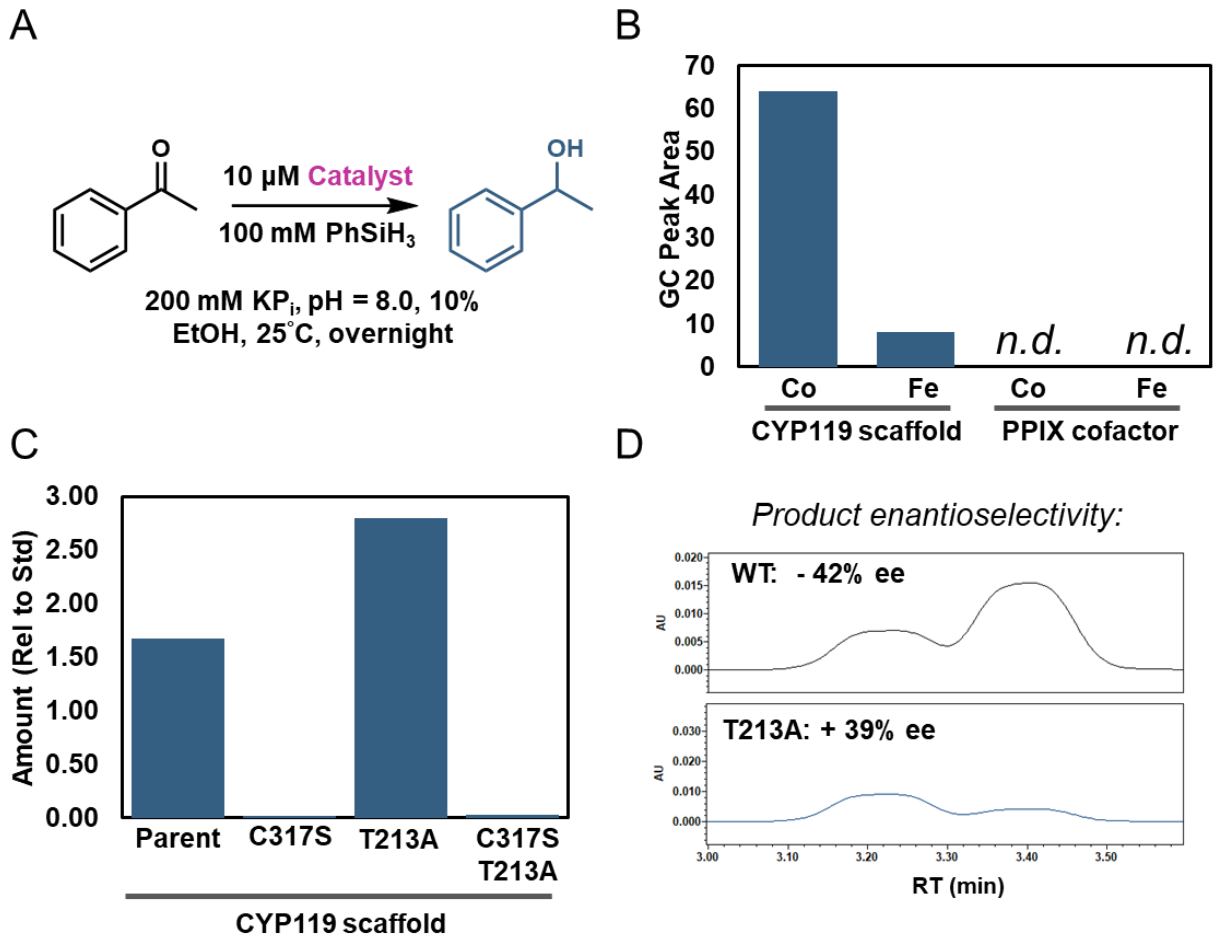


Figure 5.2. Ketone reduction reactivity for WT CoCYP119 and select variants. A) Overall reaction scheme and conditions for reaction. B) Product areas for ketone reduction reaction with various catalysts. Reactions were quenched with equal volume ethyl acetate, extracted, and analyzed by GC. C) Testing CoCYP119 variants for ketone reduction reactivity. Reactions were quenched with equal volume ethyl acetate, extracted, and analyzed by GC. Peak areas are given relative to a dodecane internal standard. D) Chiral SFC traces for the alcohol product of acetophenone reduction by WT CoCYP119 (Top) and the T213A variant (bottom). The absolute stereochemistry of this product is not known.

We hypothesized that the reduction of ketones could occur through two distinct mechanisms. Reduction could occur through a classic, two electron hydride reduction mechanism. Alternatively, the reduction could occur through a proton-coupled electron transfer-mechanism, resulting in a ketyl radical intermediate. To explore these two possibilities, we measured the relative reduction of ketone versus aldehyde, both in isolation and in direct competition. We hypothesized that, were the cobalt hydride to react via a hydrogen atom transfer-type reaction,

the resulting ketyl radical intermediate would be more stabilized for ketone substrates than for aldehyde substrates, leading to more facile reduction of ketones than aldehydes. Alternatively, aldehydes are far more electrophilic than ketones, so a two-electron hydride reduction pathway would favor aldehyde reduction over that of the ketone. We observed that both acetophenone and benzaldehyde are reduced to the corresponding alcohol in the presence of CoCYP119 and PS. Furthermore, in direct competition, the aldehyde is reduced more readily than the ketone, suggesting that direct hydride reduction is the more likely reaction pathway. This finding is further corroborated by the reactivity pattern observed with the reduction of a cyclopropyl-substituted ketone, (**X**). The product of the reaction of X with CoCYP119 and PS has an identical mass and retention time to the product of reduction of X with sodium borohydride, suggestive of a 2 electron hydride reduction pathway.

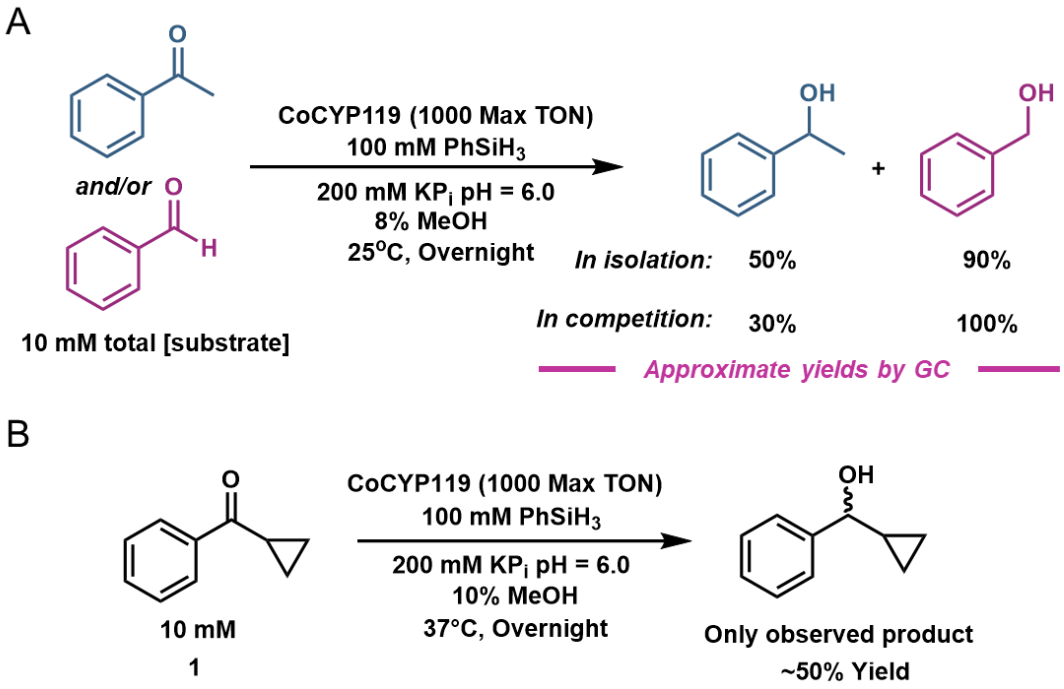


Figure 5.3. Mechanistic investigations for carbonyl reduction reactivity of WT CoCYP119. A) overall reaction scheme for competition reaction with acetophenone (blue) and benzaldehyde (pink). Reactions were quenched with equal volume ethyl acetate, extracted, and analyzed by GC using a standard curve. B) Overall reaction scheme and conditions for reduction of ketone **1** by WT CoCYP119. Reactions were quenched with equal volume ethyl acetate, extracted, and analyzed by GC. Yield was approximated based on relative peak areas of starting material and product, compared to the no enzyme control.

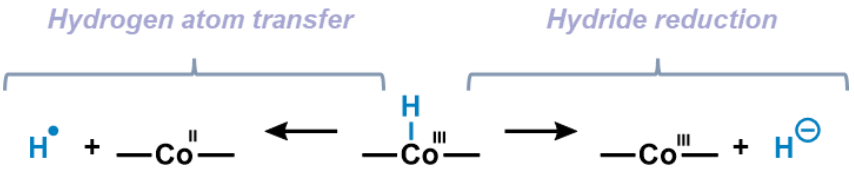


Figure 5.4. Two potential modes for metal hydride reactivity

According to the above results for reduction of ketones and the radical deallyllation activity we explored in chapter 4, we hypothesized that the putative cobalt hydride intermediate of CoCYP119 can operate via two modes of reactivity, hydride reduction and hydrogen atom transfer reactivity. These two modes have distinct requirements for metal oxidation state and suggest an intermediary reduction potential for the WT CoCYP119 cobalt center (Figure 5.4).

We hypothesized that the enzyme's predisposition for hydride reduction versus hydrogen atom transfer may be controlled by the enzymatic scaffold. We therefore screened the master plate (developed in chapter 4) for the reduction of ketone **1**, as this substrate retained well on UPLC. The master plate of variants was screened in heat-treated cell lysate format, with 3 equivalents (15 mM) of DMPS relative to ketone **1** (5 mM). Reactions were quenched after 5 hours and analyzed by UPLC-MS. The retention of function curve for this screen is given in Figure 5.5. The V254D parent enzyme had negligible activity for this reaction, and there were many variants with improved activity (Figure 5.5A). Intriguingly, none of the highly active variants in this screen were found to be highly activating for deallylation (Ch 4). Indeed, when we broadly compared the activity for deallylation versus ketone reduction for the variants, we were surprised to find nearly an anticorrelation for these two modes of reactivity. The best variants for deallylation performed very poorly for ketone reduction, and vice versa. This observation was surprising because both proposed reactions require the same reactive intermediate for activity. We surveyed the sequences for the top 10 variants for ketone reduction and found that V254D was the most common mutation at this site. In contrast, the native Val was preferred for this site in the deallylation reaction screen. The variant with the highest ketone

reduction activity in this screen, 7B6, is a quintuple variant, bearing the sequence L69N, L205Q, A209S, T213G, V254D.

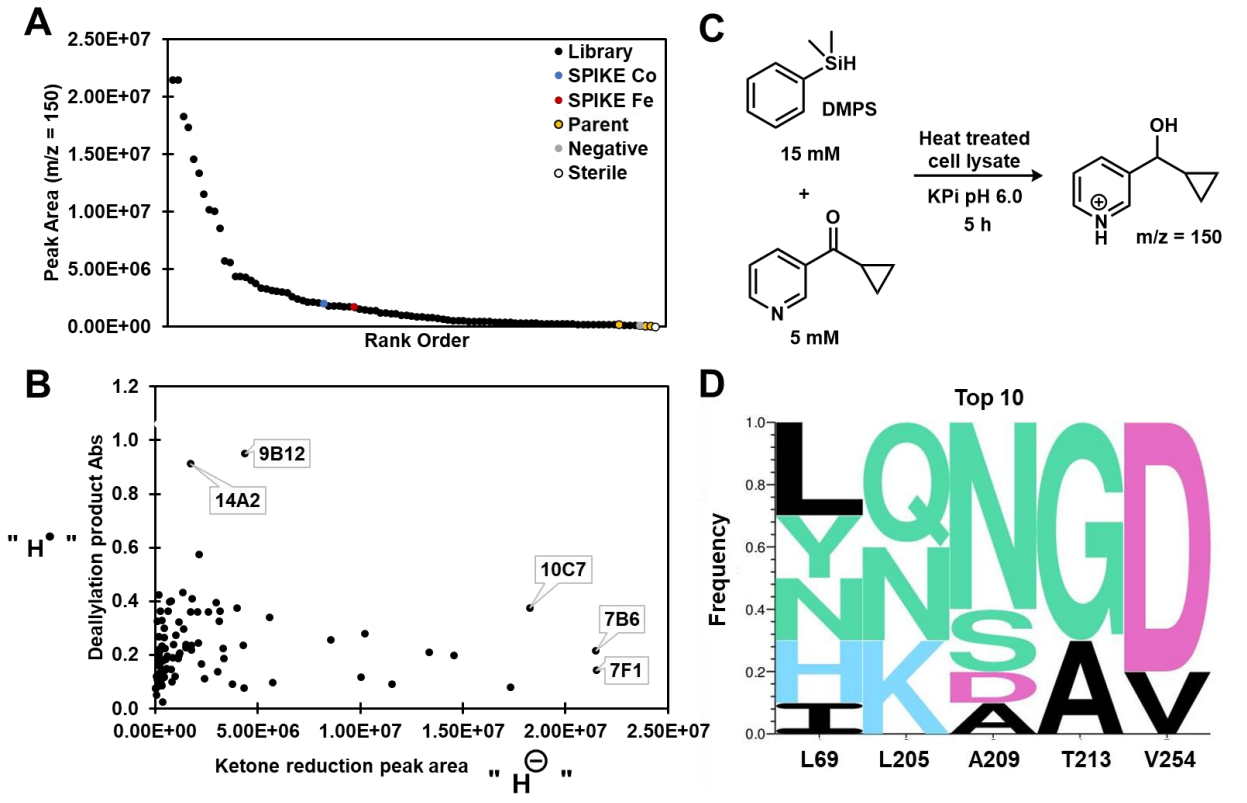


Figure 5.5. Screening the recombination master plate (see chapter 4) for ketone reduction reactivity. A) retention of function curve for reduction of ketone **1** to the corresponding alcohol. B) correlation of variant activity towards deallylation (y-axis) versus towards ketone reduction (x-axis). The identities of the best variants for each mode of reactivity are labeled. C) Conditions and overall reaction scheme for ketone reduction screen. D) Logo plot depicting the frequency of mutations in the top 10 ketone reduction variants.

5. 3. 2. Mukaiyama hydration

We next explored the ability of CoCYP119 and variants to carry out MHAT reactions. We envisioned that styrene analogs would serve as excellent HAT acceptors, as the resulting benzylic radical is highly stabilized and could be trapped by diverse coupling partners. In one classic MHAT reaction, the Mukaiyama hydration reaction, molecular oxygen serves as the radical coupling partner. We hypothesized that the natural evolution of P450s to accept molecular oxygen as a substrate would make this a good target for enzymatic MHAT. Indeed, in the presence of PS,

styrene, and aerobic atmosphere, we found that WT CoCYP119 generates two products, acetophenone and 1-phenylethanol (Figure 5.6). This mixture of products has previously been associated with Mukaiyama hydration reactions[1] and is thought to arise from β-scission pathway of a putative peroxy radical intermediate. In contrast to previous MHAT assays, reactions with the Fe form of the enzyme showed evidence of reactivity, but these data are not included in this thesis.

We measured the enantiomeric ratio of the chiral secondary alcohol resulting from the CoCYP119 reaction, and were pleased to see that this alcohol was enantioenriched—promising evidence that the O_2 radical trapping step occurred within the active site of the enzyme. The alcohol formed is identical to that generated from ketone reduction (above), affording the possibility that some alcohol is formed through reduction of the acetophenone that is formed in situ. However, we serendipitously observed that the stereochemistry of the alcohol formed via the putative Mukaiyama hydration was reversed relative to ketone reduction. This result indicates that the two alcohols are generated via distinct mechanisms, and that directed hydration was an operative pathway for the reaction.

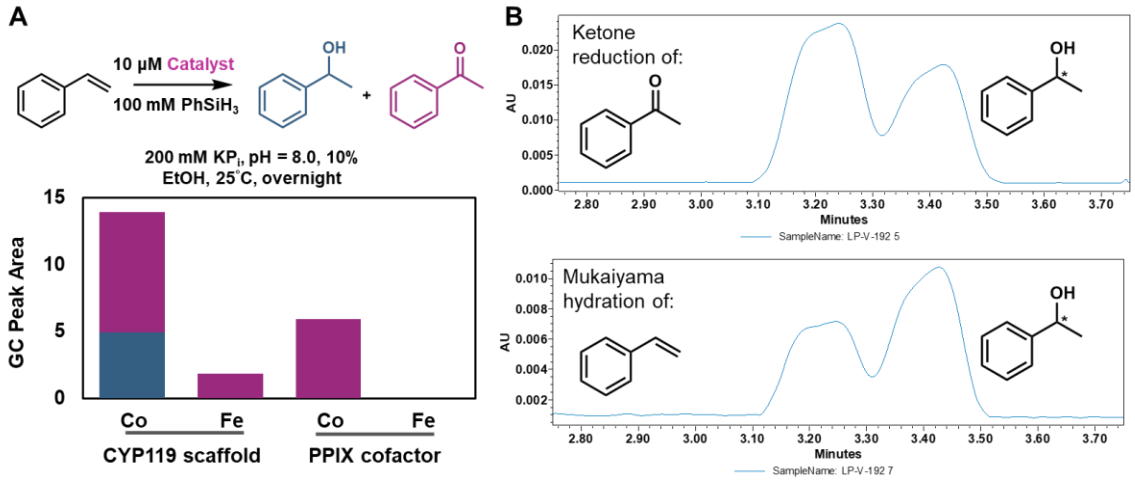


Figure 5.6. Products of Mukaiyama hydration of styrene with PS and various catalytic system. A) Overall reaction scheme, conditions, and product profiles for Mukaiyama hydration reaction with various catalyst systems. Blue represents the alcohol product and pink represents the acetophenone product. Reactions were quenched with equal volume ethyl acetate, extracted, and analyzed by GC. B) Chiral SFC trace depicting the enantiopurity of the benzylic alcohol resulting from reactions with either acetophenone (top) and styrene (bottom).

When α-methyl styrene is subjected to the reaction conditions, a related set of products is obtained (Figure 5.7). The acetophenone product may result from beta scission of a hydroperoxy radical intermediate.

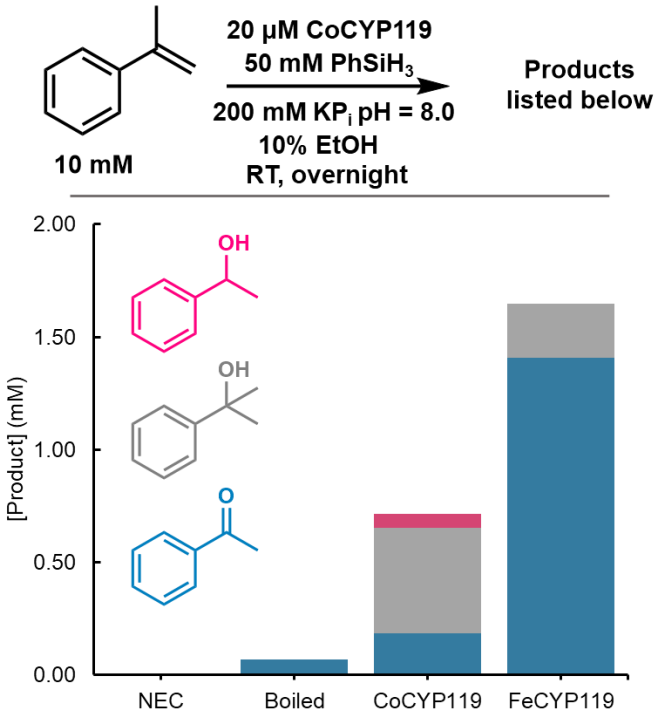


Figure 5.7. Product profile resulting from reaction of α-methylstyrene with PS under with various catalyst systems. “Boiled” refers to a sample of CoCYP119 that was heated to 90°C for 20 minutes prior to adding substrate, presumably denaturing any CoCYP119 enzyme. Reactions were quenched with equal volume ethyl acetate, extracted, and analyzed by GC using a standard curve.

In all reactions tested with styrene derivatives, we noticed that starting material was consistently fully depleted, even when the product was formed in low yield. A white, fluffy precipitate was present in many reaction solutions, including the no enzyme control reactions. We hypothesized that some background sytrene polymerization reaction may be responsible for both the loss of starting material and the white precipitant. We therefore chose to test reactivity with the more water-soluble and less-reactive vinylpyridine styrene analogs, shown in Figure 5.8.

Under an identical set of conditions, the putative Mukaiyama hydration of 3-vinylpyridine by CoCYP119 results in almost exclusively the alcohol product. We measured the relative

formation of this product in the presence and absence of oxygen, to measure whether molecular oxygen indeed served as the radical trap for this reaction (Figure 5.8). Indeed, product formation was severely diminished when the reaction was run anaerobically. This result indicates that oxygen likely serves as the radical trap for this reaction. We further speculate that the previously-observed ketone product is not observed because it is more electrophilic than acetophenone and that it is being fully reduced through a 2-electron hydride reduction. Further experiments will be needed to substantiate this reactivity.

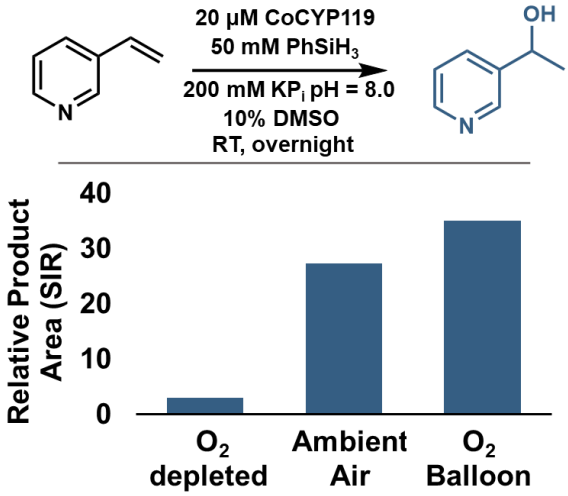


Figure 5.8. Evidence for O_2 as the radical trap in the mukaiyama hydration of 3-vinylpyridine. Reactions were quenched with equal volume methanol, centrifuged, and analyzed by UPLC-MS. m/z= 122 in positive mode was used to quantify the amount of the alcohol produced.

We were surprised to observe in our control reactions that native iron version of this enzyme, FeCYP119, also appears to catalyze a Mukaiyama hydration-like reaction. The Fe-catalyzed reaction varies in efficiency compared to the CoCYP119 catalyzed reaction, and in some cases (such as for α-methylstyrene), FeCYP119 reactivity is greater than that of CoCYP119. This result is puzzling, as we have no evidence for FeCYP119 reactivity in ketone reduction or deallylation-type reactions, and we observed no iron-hydride-like intermediate for FeCYP119 enzymes. We are still trying to understand the nature of the FeCYP119 reaction, and whether it occurs via the same mechanism as the CoCYP119 reaction.

We next tested whether a different radical acceptor could serve as a trap for the benzylic radical. When 50 mM TEMPO was added to a reaction with silane and α-methylvinylpyridine in the presence of a CoCYP119 variant, a mass corresponding to the TEMPO adduct product was detected by UPLC-MS (Figure 5.12). Whether this radical coupling occurs within the active site of the enzyme (and is therefore enantioselective) remains to be determined.

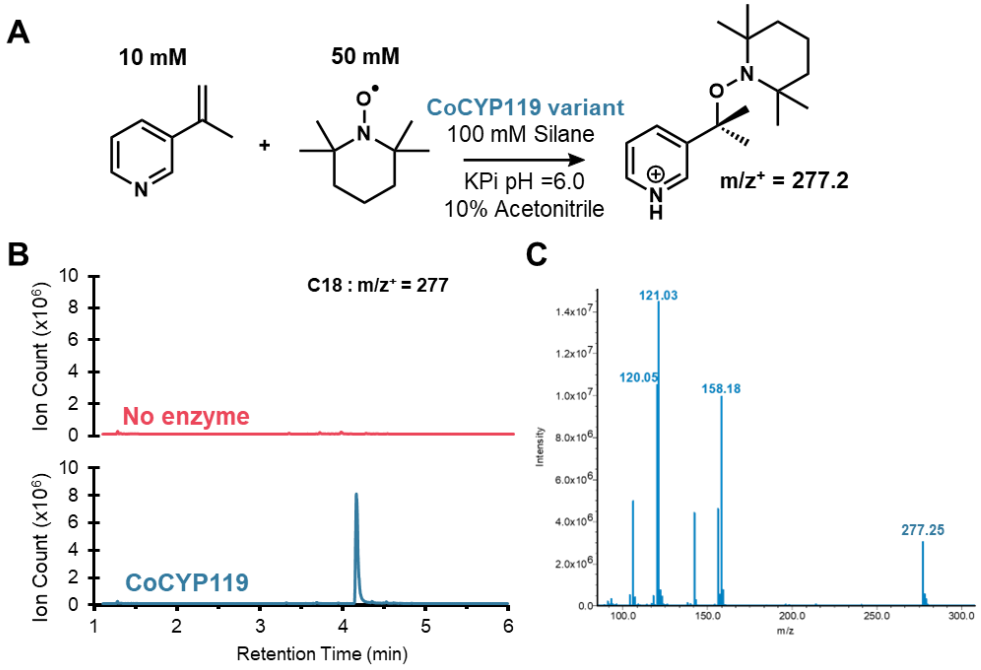


Figure 5.9. Trapping the putative benzylic radical from 3-(α-methyl)-vinyl pyridine with TEMPO. A) Overall reaction scheme and conditions. This reaction was conducted with variant 5B2 at approximately 100 max TON catalyst loading. B) Comparison of the XIC-UPLC trace for the reaction with CoCYP119 variant and the no enzyme control. C) Mass fragmentation of putative TEMPO adduct.

Mukaiyama hydration reactions with the α-methylvinylpyridine substrate yield a mixture of products, including the benzylic hydroperoxide, the alcohol, and the ketone. We were interested in understanding how the enzyme scaffold is able to control the product profile of this reaction. Thus, we screened the master plate library (see chapter 4) against the Mukaiyama hydration/hydroperoxidation of α-methylvinylpyridine in the presence of DMPS. The results of this screen can be seen in Figures 5.9-11. Retention of function curves for each product indicate that there are variants with improved activity relative to WT for each product (Figure 5.9). The relative

proportion of each product produced also varies throughout the library , indicating that the enzyme scaffold may divert reactive intermediates towards particular products of interest (Figure 5.10). However, none of the variants demonstrated exceptional selectivity for any one product.

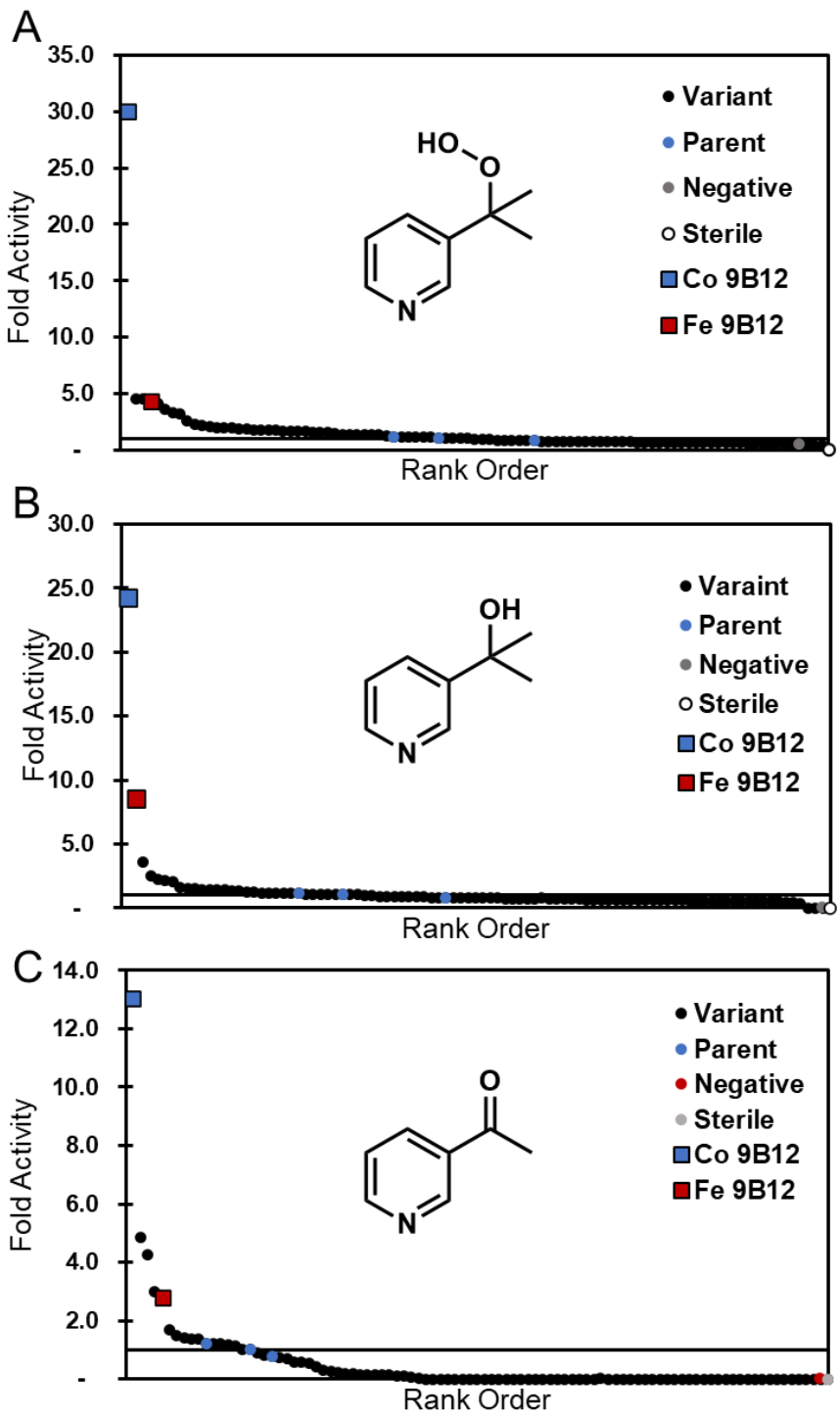


Figure 5.10. Retention of function curves for mukaiyama hydration with the recombination master plate developed in chapter 4. A) Fold improvement in benzylic hydroperoxide product formation. B) Fold improvement in formation of the benzylic tertiary alcohol product. C) Fold improvement in the production of the acetophenone product.

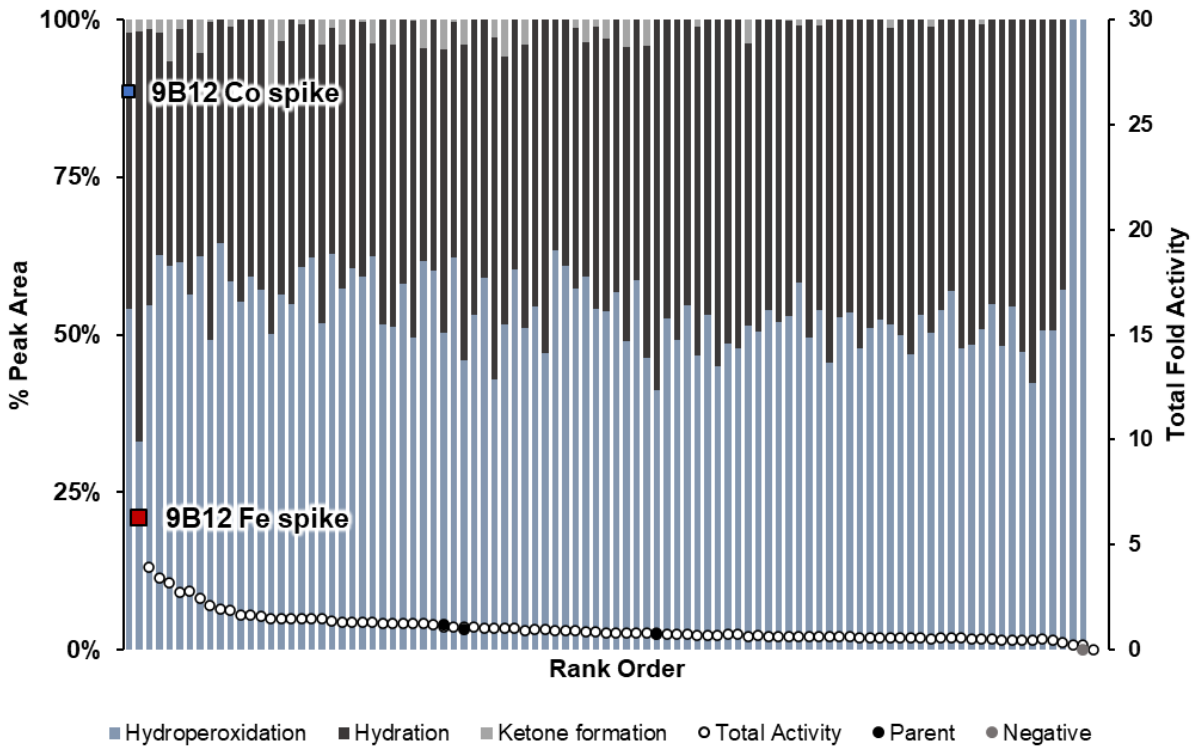


Figure 5.11. Product profile for each variant in the mukaiyama hydration screen of the recombination master plate. Grey bars indicate the acetophenone product, black bars represent the alcohol product, and blue bars represent the hydroperoxide product. The dots represent the total fold increase in peak areas, relative to the parent wells (indicated with black dots).

We also assessed whether Mukaiyama hydration activity was correlated with deallylation activity. Figure 5.11 plots the activity towards mukaiyama hydration and hydroperoxidation versus deallylation activity. We observed a weak (R^2 = 0.65) correlation for reactivity for both products. The top hydration catalyst, 8A9, is a quadruple variant, L69Y/L205K/T213G/P252S. The P252S mutation was serendipitous, and was not intended to be included in the original library. The best variants for deallylation, 9B12 and 14A2, have measurable Mukaiyama hydration activity as well, with 9B12 outcompeting 14A2.

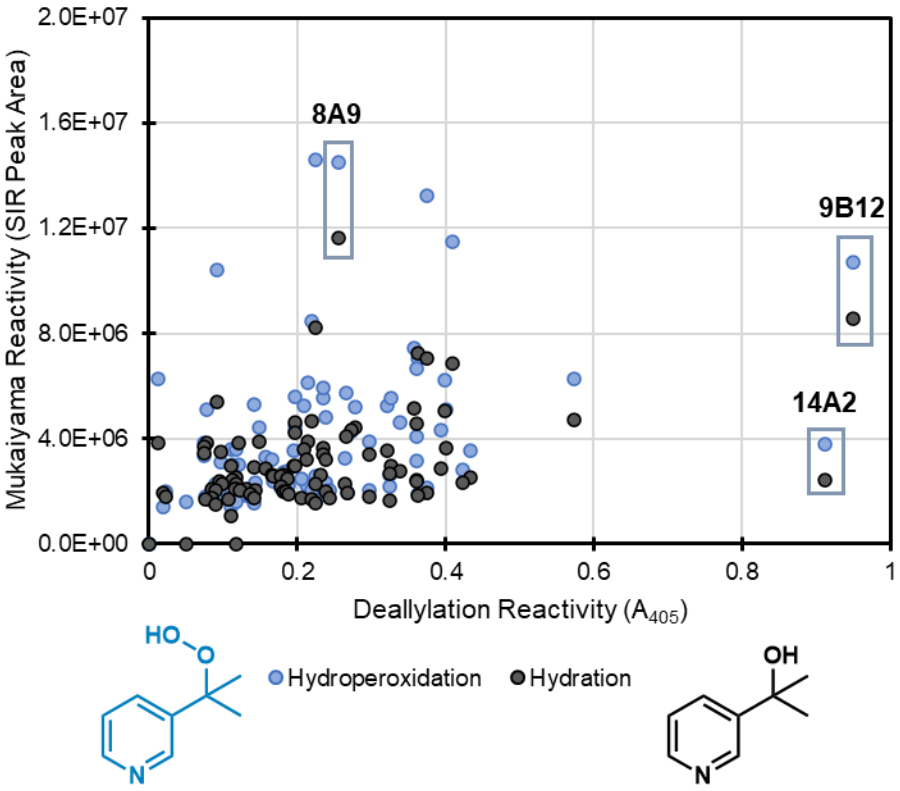


Figure 5.12. Correlation of deallylation activity (x-axis) to mukaiyama activity (y-axis). The peak area for the hydroperoxide product (blue) and alcohol (black) are plotted together. Top variants for each mode of reactivity are indicated on the plot.

5. 3. 3. Isomerization

We were surprised to discover that hydrogen atom transfer to olefins under aerobic conditions does not exclusively lead to hydration products via trapping with molecular oxygen. In certain cases, we also observed isomerization products. For instance, when prenol is subjected to conditions with silane and CoCYP119, isoprenol is observed in a 20:1 ratio of starting material to product. Assuming a 1.1 kcal/mol difference in energy between these two products, we estimate that the enzyme performed approximately 300 turnovers, scrambling the alkene substitution pattern. A proposed mechanism for this isomerization is given in Figure 5.13.

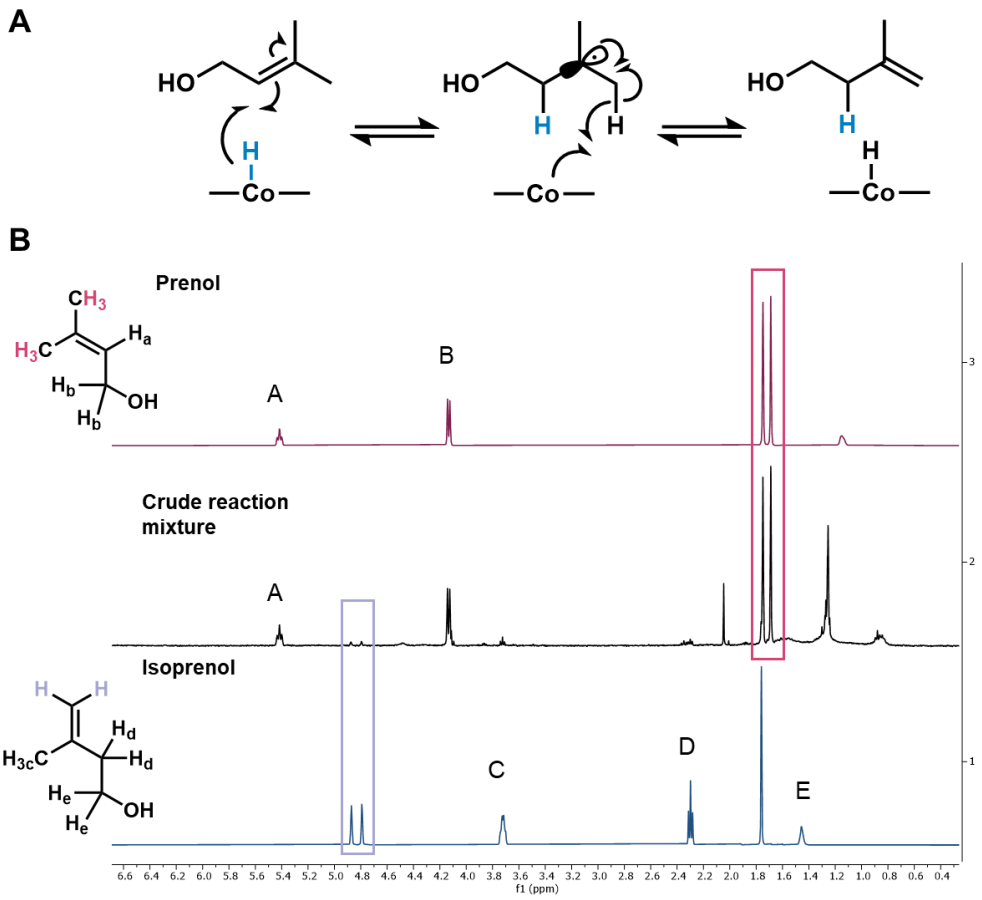


Figure 5.13. Isomerization of prenol by CoCYP119. A) proposed mechanism for isomerization. B) Crude NMR analysis of reaction products. Reaction conditions: 10 mM Prenol, 50 mM $PhSiH_3$, 10 µM WT CoCYP119 in 100 mM KPi pH = 8.0.

5. 3. 4. Radical Cyclization

Given the ability of CoCYP119 to enact isomerization reactions, we next asked if the catalyst could enable a radical cyclization reaction. To this end, we synthesized a diallylated amine substrate (structures I, II, and II in Figure 5.14B), and tested their ability to be cyclized with CoCYP119. Each reaction gave a mixture of products. By studying parent ion and fragmentation patterns, we were able to assign putative products from cycloisomerization followed by hydration and direct deallylation of the amine. Putative structure assignments are shown in the corresponding reaction product UPLC traces in Figure 5.14C.

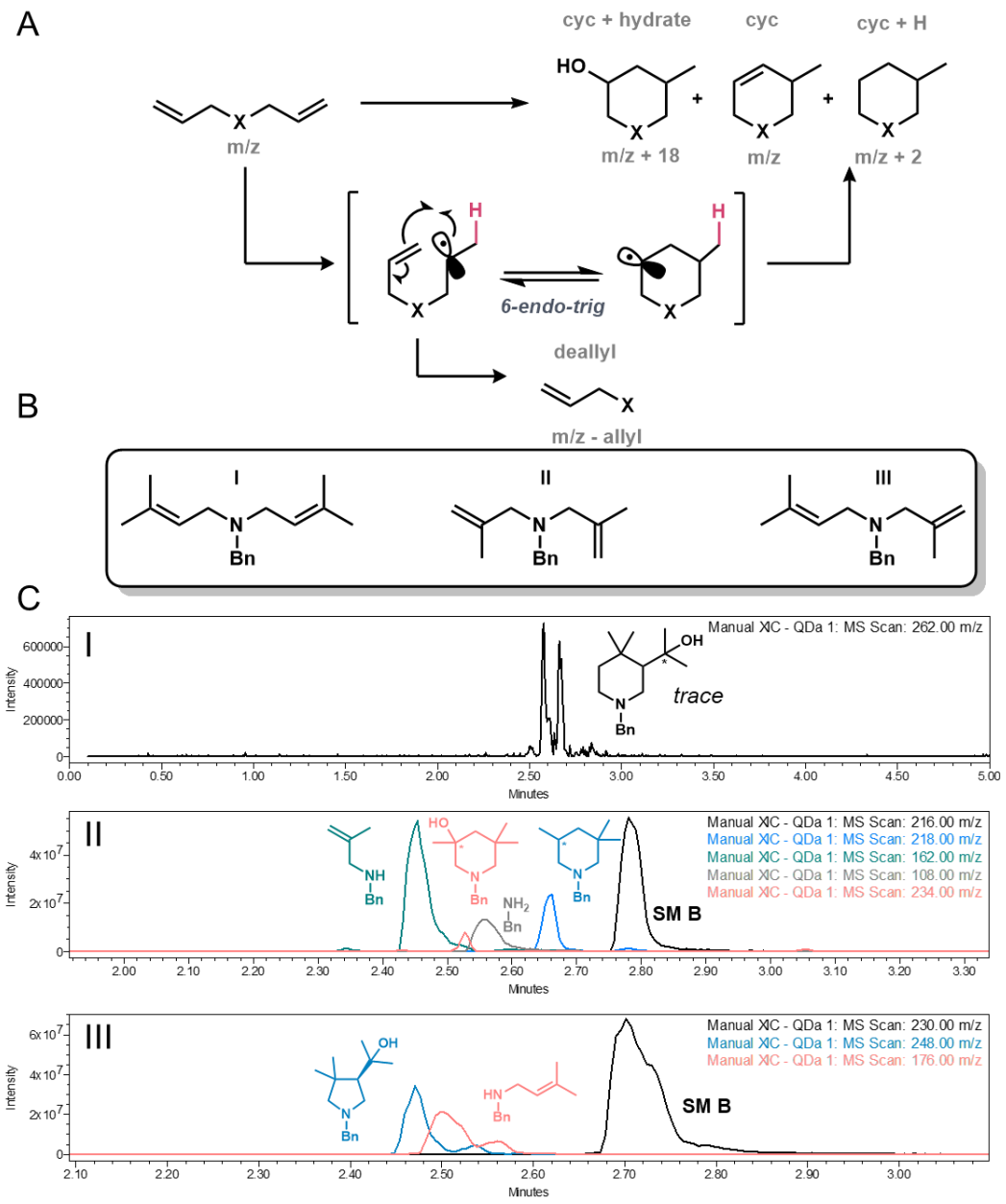


Figure 5.14. Radical cyclization catalyzed by CoCYP119. A) Putative cyclization mechanism and potential product profiles. B) Substrates tested for radical cyclization. C) UPLC-traces for the product of reaction of substrates I-III with WT CoCYP119. Reaction conditions: 10 mM Substrate I, II, or III, 100 mM $PhSiH_3$, 100 µM WT CoCYP119 in 200 mM KPi pH = 6.0. Reaction were run overnight at room temperature and then quenched with equal volume MeOH, centrifuged, and then analyzed by UPLC-MS.

To explore how the enzyme scaffold could tune reactivity for a radical cyclization product, we screened the master plate against substrate II (Figure 5.15). We were pleased to see that different variants gave different reaction outcomes. We are presently working to definitively assign structures to each of these peaks and validate changes to product profiles.

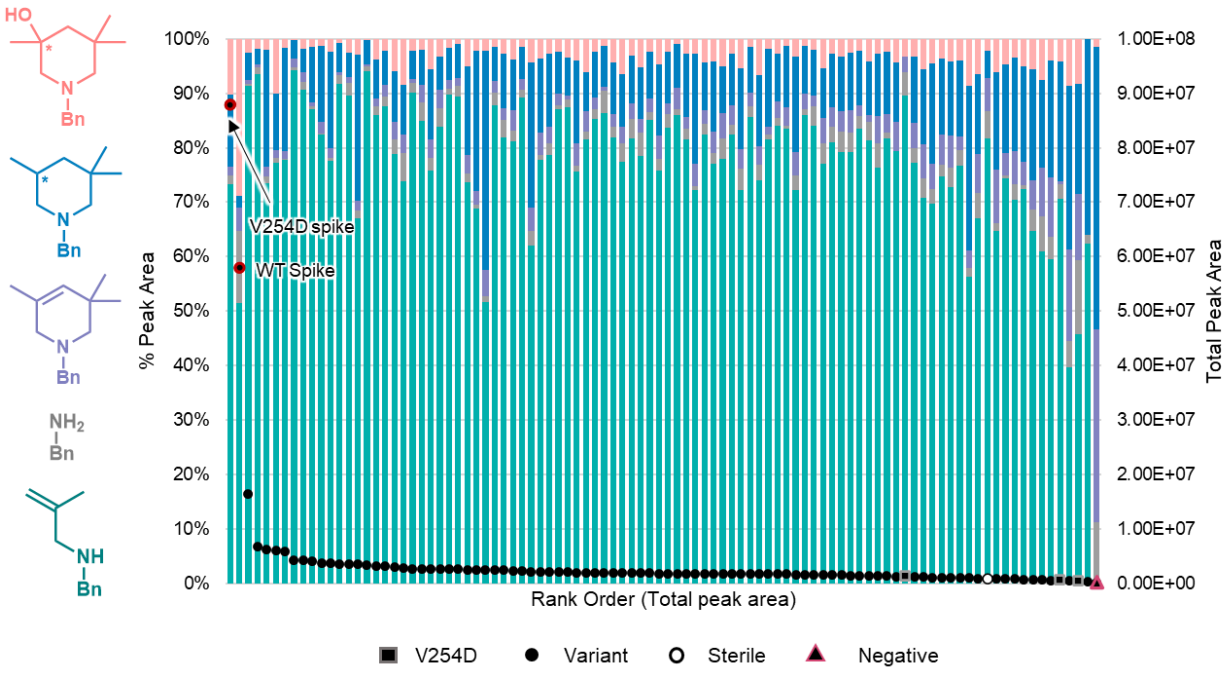


Figure 5.15. Retention of function curve for radical cyclization of substrate **II** by variants from the recombination master plate. Black dots represent the total peak area for each variant, and the colored bars represent the relative peak area for each potential product (putative structures drawn at right). Screening conditions: 5 mM substrate II, 15 mM DMPS, 180 µL heat treated cell lysate in 100 mM KPi pH 6.0 with 10% acetonitrile. Reactions were incubated at room temperature for 20 h, and then quenched with 200 µL methanol and analyzed by UPLC-MS.

5. 4. Discussion

The data presented in this chapter explore the potential for CoCYP119 as a synthetic catalyst. While the demonstration of diverse modes of reactivity is exciting, applications with wild-type enzyme are limited because the reactions suffer from forming mixtures of products. In the world of protein engineering, this is a good, and somewhat rare, problem to have. Selectivity is a hallmark of enzymatic reactions, and there are numerous examples of directed evolution campaigns to tune the reactivity of an enzyme such that a single product is produced out of myriad potential products.[13–16] We envision that similar directed campaigns for specific desired products will be useful for CoCYP119 engineering as well.

The reactivity of the native Fe version of CYP119 in MHAT reactions was unexpected. While there are many examples of Fe-dependent MHAT reactions in synthetic chemistry,[7,10] we observed many no MHAT reactivity for the FeCYP119 enzyme with our model MHAT deallylation reaction. It is possible that the electronic requirements of the deallylation reaction are sufficiently distinct from the other MHAT reactions presented in this chapter. Alternatively, it is possible that the FeCYP119 enzyme performs these reactions via a distinct mechanism. Our mechanistic investigation of these other MHAT reactions are minimal, and more experiments are required to confirm that MHAT is indeed the operative mechanism for both the Co and Fe-dependent reactions.

4. 6. Materials and Methods

4. 6. 1. General experimental methods

All chemicals and chemical standards were purchased from commercial suppliers (Sigma-Aldrich, VWR, Goldbio, Frontier biosciences, Fluka), and used without further purification. Casamino acids were obtained from Biolabs or Research Products International. Unless otherwise noted, all media and solutions were prepared using ultrapure water (≥18 MΩ, from Thermo Scientific Barnstead Nanopure water purification system). Immobilized metal affinity chromatography resin or columns were purchased from GE Healthcare.

Equipment and instrumentation

New Brunswick I26R, 120 V/60 Hz shakers (Eppendorf) were used for cell growth. When needed, the humidity was controlled with a HumidiKit™ Auto Humidity System for Incubators purchased from IncubatorWarehouse.com. Cell disruption was accomplished via sonication with a Sonic Dismembrator 550 (Fischer Scientific) sonicator. Electroporation (for transformation and cloning) was achieved using an Eppendorf E-porator set to 2500V. Optical density measurements of liquid cultures were recorded on a Ultrospec 10 cell density meter (Amersham Biosciences). An Envision® 2105 multimode plate reader (Perkin Elmer) was used to measure optical density in 96-well plates. UPLC/MS data were collected on an Acquity UHPLC with an Acquity QDA MS detector (Waters). Electronic absorption data were collected on a UV-2600 Shimadzu spectrophotometer, set to a spectral bandwidth of 0.5 nm.

5. 6. 2. Cloning and Purification of CoCYP119

Gene block fragments were purchased from Integrated DNA Technologies (IDT) and inserted into the pET22b vector by the Gibson Assembly method and E. coli cells were transformed with the resulting cyclized DNA product by electroporation. After 45 min. of recovery in TB media at 37 °C and 200 rpm, cells were plated onto LB plates with 100 µg/mL AMP and incubated overnight.

Single colonies were used to inoculate 6 mL TB + 100 µg/mL AMP, which were grown overnight at 37 °C and 200 rpm. Plasmid was extracted and purified from these starter cultures using the Macherey-Nagel Plasmid Miniprep Kit, and the gene inserts were sequenced by Functional Biosciences.

To obtain purified protein, BL21(DE3) electrocompetent cells were freshly transformed with the CYP119 gene on an IPTG-inducible Pet22b and recovered in Terrific Broth (TB) for 45 min at 37°C. Cells were then plated onto Luria-Burtani (LB) plates with 100 µg/mL ampicillin (AMP) and incubated overnight. Single colonies were used to inoculate 10 mL TB + 100 µg/mL AMP, which were grown overnight at 37 °C and 200 rpm. Expression cultures, typically 1 L of TB-AMP in a 2.8 L baffled Fernbach flask, were inoculated from these starter cultures and shaken at 37 °C and 220 rpm. After 3.5 hours, the expression cultures were chilled on ice. After 30 min. on ice, CYP119 expression was induced with 0.5 mM IPTG and CoPPIX production was promoted with 0.25 mM δ-aminolevulinic acid (δALA) and 1 mM $CoCl_2$. The cultures were incubated overnight at 25 °C and 180 rpm. Cells were then harvested by centrifugation at 4,000xg for 15 min at 25 °C. Cell pellets were usually pinkish-brown in color. These pellets were frozen and stored at -20 °C until lysis.

To purify CoCYP119, cell pellets were thawed and resuspended in lysis buffer (100 mM potassium phosphate buffer (pH = 6.0), 100 mM NaCl, 1 mg/mL hen egg white lysozyme (GoldBio), 0.2 mg/mL DNase I (GoldBio), 2 mM MgCl2, ~14 µM CoPPIX in DMSO). A volume of 4 mL of lysis buffer per 1 gram of wet cell pellet was used. After 45 min. of shaking at 37 °C and 200 rpm, the cells heat treated at 65°C for 15-20 min. The resulting lysate was clarified of cellular debris by centrifugation at 30,000xg for 15 min, resulting in a red-orange supernatant. The pH of the supernatant was adjudted to 7.5 with 1.0 M NaOH, and the supernatant was applied to a gravity column of Ni/NTA beads (GoldBio) that were equilibrated with 100 mM potassium phosphate buffer (pH = 8.0). The column was washed with approximately 3 column volumes of 50 mM imidazole, 100 mM NaCl, 100 mM potassium phosphate buffer (pH = 8.0). CoCYP119

was eluted with 250 mM imidazole, 100 mM NaCl, 100 mM potassium phosphate buffer (pH = 8.0). Elution of the desired protein was monitored by the disappearance of the red color (resulting from the release of CoCYP119) from the column. The eluent was dialyzed into storage buffer (100 mM potassium phosphate buffer, pH = 8.0) overnight at 4 °C, at least 4 times to remove excess imidazole. Dialyzed protein solutions were concentrated to roughly 0.5 mM using Amicon® Ultra 15 mL Centrifugal Filters (Millipore) at a molecular weight cutoff of 10 kDa. Purified enzyme was flash frozen in pellet form by pipetting enzyme dropwise into a crystallization dish filled with liquid nitrogen. The enzyme pellets were transferred to a plastic conical and stored at -80 °C until further use. Generally, this procedure yielded about 20 mg per L culture.

5. 6. 3. General Procedure for plate screening

A 96-deep well plate containing TB + 100 ug/mL AMP (500 uL) was inoculated by a single colony in each well and starter plates were grown overnight at 37 °C and 200 rpm. Expression cultures were inoculated in a fresh 96-deep well plate containing TB + 100 ug/mL AMP (650 uL) using starter culture (50 uL) and grown at 37 °C and 200 rpm for 4.5 hours. Expression cultures were chilled on ice for 30 min. and then induced with 50 µL of a 3.75 mM δALA, 15 mM $CoCl_2$, 15 mM IPTG in H_2O stock (final concentrations of 0.25 mM, 1 mM, and 1 mM, respectively). Cultures were expressed overnight at 25 °C and 180 rpm. The plate was spun down at 4,000xg for 15 min. at 25 °C and the pellets were frozen at -20 °C until use.

Before screening, frozen cell pellets in a 96-deep well plate were thawed and resuspended in 400 uL 100 mM potassium phosphate buffer (pH = 6.0). After applying an Alumaseal sealing film (VWR) to the plate with a rubber roller (VWR), the plate was heat treated in a water bath at 65 °C for 15 min. The resulting lysate was clarified of cellular debris by centrifugation at 4,000xg at 25 °C for 20 min.

To run reactions, 180 µL of heat treated cell lysate was added to a plastic 1 mL deep well plate. 20 µL of organic master mix, consisting of 150 mM DMPS and 50 mM of the appropriate substrate in acetonitrile were added. The reaction was mixed by pipetting up and down, and seal with an aluminum seal and incubated for the prescribed reaction time at room temperature. The seal was then removed and the reaction was quenched with 200 µL of Methanol. The plate was centrifuged at 4,300 xg for 20 minutes, and the supernatant was transferred to a filter plate (PAL). The filter plate was centrifuged at 1000 xg for 15 minutes, or until all of the liquid had passed through the filter plate. This filtrate was then analyzed by UPLC-MS using a 3 min acetonitrile/H_2O + 0.1% Formic acid gradient using a T3 column.

5. 7. References

1. Isayama, S. & Mukaiyama, T. A New Method for Preparation of Alcohols from Olefins with Molecular Oxygen and Phenylsilane by the Use of Bis(acetylacetonato)cobalt(II). *Chem. Lett.* **18**, 1071–1074 (1989).
2. Lo, J. C., Yabe, Y. & Baran, P. S. A practical and catalytic reductive olefin coupling. *J. Am. Chem. Soc.* **136**, 1304–1307 (2014).
3. Shevick, S. L., Obradors, C. & Shenvi, R. A. Mechanistic Interrogation of Co/Ni-Dual Catalyzed Hydroarylation. *J. Am. Chem. Soc.* **140**, 12056–12068 (2018).
4. Ma, X. & Herzon, S. B. Intermolecular Hydropyridylation of Unactivated Alkenes. *J. Am. Chem. Soc.* **138**, 8718–8721 (2016).
5. Gaspar, B. & Carreira, E. M. Mild cobalt-catalyzed hydrocyanation of olefins with tosyl cyanide. *Angew. Chemie - Int. Ed.* **46**, 4519–4522 (2007).
6. Rô Me Waser, J. & Carreira, E. M. Convenient Synthesis of Alkylhydrazides by the Cobalt-Catalyzed Hydrohydrazination Reaction of Olefins and Azodicarboxylates. *J. AM. CHEM. SOC* **126**, 5676–5677 (2004).
7. Barker, T. J. & Boger, D. L. Fe(III)/NaBH 4-mediated free radical hydrofluorination of

unactivated alkenes. *J. Am. Chem. Soc.* **134**, 13588–13591 (2012).

8. Crossley, S. W. M., Barabé, F. & Shenvi, R. A. Simple, chemoselective, catalytic olefin isomerization. *J. Am. Chem. Soc.* **136**, 16788–16791 (2014).
9. Green, S. A. *et al.* The High Chemofidelity of Metal-Catalyzed Hydrogen Atom Transfer. *Acc. Chem. Res.* **51**, 2628–2640 (2018).
10. Crossley, S. W. M., Obradors, C., Martinez, R. M. & Shenvi, R. A. Mn-, Fe-, and Co-Catalyzed Radical Hydrofunctionalizations of Olefins. *Chem. Rev.* **116**, 8912–9000 (2016).
11. Ji, P., Park, J., Gu, Y., Clark, D. S. & Hartwig, J. F. Abiotic reduction of ketones with silanes catalysed by carbonic anhydrase through an enzymatic zinc hydride. *Nat. Chem.* **13**, 312–318 (2021).
12. Hoffnagle, A. M. & Tezcan, F. A. Atomically Accurate Design of Metalloproteins with Predefined Coordination Geometries. (2023) doi:10.1021/jacs.3c04047.
13. Zetzsche, L. E. *et al.* Biocatalytic oxidative cross-coupling reactions for biaryl bond formation. *Nature* **603**, 79–85 (2022).
14. Miller, D. C., Lal, R. G., Marchetti, L. A. & Arnold, F. H. Biocatalytic One-Carbon Ring Expansion of Aziridines to Azetidines via a Highly Enantioselective [1,2]-Stevens Rearrangement. *J. Am. Chem. Soc.* **144**, 4739–4745 (2022).
15. Dodani, S. C. *et al.* Discovery of a regioselectivity switch in nitrating P450s guided by molecular dynamics simulations and Markov models. *Nat. Chem.* **8**, 419–425 (2016).
16. Hammer, S. C. *et al.* Anti-Markovnikov alkene oxidation by metal-oxo–mediated enzyme catalysis. *Science (80-.).* **358**, 215–218 (2017).