Olefin Cyclopropanation and Carbon-Hydrogen Amination via Carbene and Nitrene Transfers Catalyzed by Engineered Cytochrome P450 Enzymes

Thesis by

Pedro de Souza Leão Coelho

In Partial Fulfillment of the Requirements for the Degree

of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2013

(Defended February 27, 2013)

ACKNOWLEDGEMENTS

To my mentor and role model, Frances Arnold, whose career is proof that novel and useful research can have an immense impact in this world.

I am indebted to the postdocs and fellow graduate students who much contributed to my learning and development at Caltech, especially Jared Lewis, Eric Brustad, Mike Chen, Jane Wang, and John McIntosh. I am grateful for the contributions from the exceptional undergraduates, Arvind Kannan, Tristan Brown, and Eugenia Zah, who worked during the riskiest and most uncertain stages of this research project.

To my parents and my brothers for their unconditional support.

To Renata for her understanding in sharing me with the lab and for her always joyful company that made these four years in Los Angeles unforgettable.

Salomon saith: There is no new thing upon the earth.

So that as Plato had an imagination, that all knowledge was but remembrance;

so Salomon giveth his sentence, that all novelty is but oblivion.

Francis Bacon, *Essays*, LVIII

The certainty that everything has already been written annuls us,

or renders us phantasmal.

Jorge Luis Borges, *The Library of Babel*

SEPTIMUS: We shed as we pick up, like travellers who must carry everything in their arms, and what we let fall will be picked up by those behind. The procession is very long and life is very short. We die on the march. But there is nothing outside the march so nothing can be lost to it. The missing plays of Sophocles will turn up piece by piece, or be written again in another language. Ancient cures for disease will reveal themselves once more. Mathematical discoveries glimpsed and lost to view will have their time again. You do not suppose, my lady, that if all of Archimedes had been hiding in the great library of Alexandria, we would be at a loss for a corkscrew?

Tom Stoppard, *Arcadia*

ABSTRACT

Synthetic biology promises to transform organic synthesis by enabling artificial catalysis in living cells. I start by reviewing the state of the art in this young field and recognizing that new approaches are required for designing enzymes that catalyze nonnatural reactions, in order to expand the scope of biocatalytic transformations. Carbene and nitrene transfers to C=C and C-H bonds are reactions of tremendous synthetic utility that lack biological counterparts. I show that various heme proteins, including cytochrome $P450_{BM3}$, will catalyze promiscuous levels of olefin cyclopropanation when provided with the appropriate synthetic reagents (e.g., diazoesters and styrene). Only a few amino acid substitutions are required to install synthetically useful levels of stereoselective cyclopropanation activity in $P450_{BM3}$. Understanding that the ferrous-heme is the active species for catalysis and that the artificial reagents are unable to induce a spin-shift-dependent increase in the redox potential of the ferric P450, I design a high-potential serine-heme ligated P450 (P411) that can efficiently catalyze cyclopropanation using NAD(P)H. Intact *E. coli* whole-cells expressing P411 are highly efficient asymmetric catalysts for olefin cyclopropanation. I also show that engineered P450s can catalyze intramolecular amination of benzylic C-H bonds from arylsulfonyl azides. Finally, I review other examples of where synthetic reagents have been used to drive the evolution of novel enzymatic activity in the environment and in the laboratory. I invoke preadaptation to explain these observations and propose that other man-invented reactions may also be transferrable to natural enzymes by using a mechanism-based approach for choosing the enzymes and the reagents. Overall, this work shows that existing enzymes can be readily adapted for catalysis of synthetically important reactions not previously observed in nature.

CONTENTS

NOMENCLATURE

Bioinformatics: the study of methods for storing, retrieving, and analyzing biological data such as DNA, RNA, and protein sequence, structure, function, pathways, and genetic interactions

Directed evolution: a systematic algorithm of mutation and selection for improvements in the properties of a biological entity, such as an enzyme, a metabolic pathway, or a whole organism

Heterologous protein expression: expression of a gene coding for a protein not natively produced by the production host

Metabolic engineering: genetic alteration of metabolism to produce molecules of interest

Metagenomics: the study of genetic material recovered directly from environmental samples

Operon: a functioning unit of genomic DNA containing a cluster of genes under the control of a single promoter

Recombinant (metabolic) pathway: an artificial (metabolic) pathway created by combining one or more enzymes foreign to the host organism

Synthetic biology: forward engineering of genetic material (genes, promoters, operons, genomes) encoding an artificial function

ABBREVIATIONS

BM3	Cytochrome P450 BM3 from *Bacillus megaterium* (CYP102A1)
DMSO	Dimethyl sulfoxide
K_M	Michaelis-Menten constant
k_{cat}	Catalytic rate constant
LB	Luria-Bertani medium
M9Y	M9 minimal medium with 1.5% yeast extract
NADH	Reduced nicotinamide adenine dinucleotide
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
P411	Serine-heme ligated cytochrome P450s
P450	Cysteine-heme ligated cytochrome P450s
TB	Terrific-broth medium
TTN	Total turnover number

ADDITIONAL PUBLICATIONS

I coauthored the following two publications that were not included in this thesis in the interest of restraining the thematic focus to the catalysis of carbene and nitrene transfers by cytochrome P450s.

J. C. Lewis, P. S. Coelho, F. H. Arnold, Enzymatic functionalization of carbon-hydrogen bonds. *Chemical Society Reviews* **40**, 2003–2021 (2011).

Abstract. The development of new catalytic methods to functionalize carbon-hydrogen (C−H) bonds continues to progress at a rapid pace due to the significant economic and environmental benefits of these transformations over traditional synthetic methods. In nature, enzymes catalyze regio- and stereoselective C−H bond functionalization using transformations ranging from hydroxylation to hydroalkylation under ambient reaction conditions. The efficiency of these enzymes relative to analogous chemical processes has led to their increased use as biocatalysts in preparative and industrial applications. Furthermore, unlike small molecule catalysts, enzymes can be systematically optimized via directed evolution for a particular application and can be expressed *in vivo* to augment the biosynthetic capability of living organisms. While a variety of technical challenges must still be overcome for practical application of many enzymes for C−H bond functionalization, continued research on natural enzymes and on novel artificial metalloenzymes will lead to improved synthetic processes for efficient synthesis of

complex molecules. In this critical review, we discuss the most prevalent mechanistic strategies used by enzymes to functionalize nonacidic C−H bonds, the application and evolution of these enzymes for chemical synthesis, and a number of potential biosynthetic capabilities uniquely enabled by these powerful catalysts.

M. M. Chen, P. S. Coelho, F. H. Arnold, Utilizing terminal oxidants to achieve P450-catalyzed oxidation of methane. *Advanced Synthesis & Catalysis* **354**, 964–968 (2012).

Abstract. Terminal oxidant-supported P450 reactions alleviate the need for substrate binding to initiate catalysis by chemically generating "compound I." This allows investigation of the innate substrate range of the enzyme active site. Using iodosylbenzene as the oxidant, CYP153A6, a medium-chain terminal alkane hydroxylase, exhibits methanol formation in the presence of methane demonstrating that P450-mediated methane hydroxylation is possible.

Comment. Hydroxylations using iodosylbene are isoelectronic to C−H aminations using sulfonylazides and C=C cyclopropanations using diazoesters. An important difference is that iodozylbenzene can react directly with the resting state ferric form of cytochrome P450, whereas the N and C transfer reagents only react with the ferrous form of the enzyme (as shown in chapters 2, 3 and 4). This difference in reactivity requires that P450 aminations and cyclopropanations be conducted in the presence of a reductant [e.g., dithionite or NAD(P)H].

Chapter 1

SYNTHETIC BIOLOGY APPROACHES FOR ORGANIC SYNTHESIS

"Organic chemistry just now is enough to drive one mad.
It gives me the impression of a primeval forest
full of the most remarkable things,
a monstrous and boundless thicket,
with no way of escape,
into which one may well dread to enter."
– Friedrich Wöhler (1835)

Synthetic biology is to the 21st century chemist
what organic chemistry was to the 19th century inorganic chemist

This chapter is published as P. S. Coelho, J. C. Lewis, F. H. Arnold "Synthetic biology approaches for organic synthesis" in *Comprehensive Organic Synthesis II*, G. Molander, P. Knochel, eds. (Oxford, UK: Elsevier Ltd, 2014)

Key Words. Synthetic Biology | Enzyme Engineering | Directed Evolution | Metabolic Engineering | Microbial Cell Factories | Metagenomics | Biocatalysis

Abstract. Advances in DNA technologies, metagenomics, and bioinformatics have enabled the use of biological systems (i.e., enzymes, metabolic pathways, and cells) for chemical synthesis and production of chemicals from renewable resources such as plant sugars. We review these basic technologies, illustrate their connection to synthetic chemistry, and provide examples of enzyme and metabolic engineering for synthesis of organic molecules with high efficiency and selectivity. Finally, we anticipate the potential for increased integration of engineered enzymes in metabolic pathways as well as the creation of enzymes with completely novel activities to expand our biosynthetic capabilities.

1. Introduction

As the volumes of this collection illustrate, synthetic chemists have devised a powerful arsenal of chemical reactions—one that is often, if colloquially, claimed to enable the synthesis of any desired molecule given enough time. This factor, time, along with measures of material and energy economy, largely determines the efficiency, and thus utility, of chemical syntheses. Developing new catalysts, reactions, and functional molecules remains a time-intensive endeavor, and the costs of energy and material inputs along with waste from syntheses continue to increase as the global population swells and resources become more limited. Barring the removal of these pressures, advances in our ability to assemble molecules are essential.

Systematic and highly efficient approaches to the construction of molecular complexity have evolved in response to analogous supply and demand challenges over four billions years of natural selection. The emergence of myriad biocatalysts ultimately enabled the generation of the seemingly infinite array of natural products that inspire chemists (and provide immense human benefits). The efficiency of these evolved catalytic systems is typically far greater than any synthetic systems we have generated in the short 200 years of modern chemistry. This has driven extensive efforts to mimic these systems for synthetic applications—but stripping systems down to their simplest forms rarely replicates the levels of efficiency and selectivity that are hallmarks of biology.

More recently, increased understanding of biology has allowed researchers to begin to harness the capabilities of biological catalysts, including enormous catalyst diversity,

molecular recognition, selection, and evolution, for chemical synthesis. Furthermore, we are no longer limited to whole-cell or enzyme catalysts isolated from nature or cultured in the laboratory, but can readily modify these systems to suit our purposes. Efforts toward this end have taken many names over the course of their development, but are now encompassed by the umbrella term "synthetic biology". This review summarizes some key technologies and examples from this field, with the goal of illustrating how synthetic biology is already a powerful tool that can be used alongside and, increasingly, in place of conventional synthetic methods.

2. Enabling Technologies

2.1 DNA Sequencing and Synthesis

Since chemical function in biocatalysts is encoded and engineered at the DNA level, our ability to build systems for chemical synthesis is greatly influenced by technologies that enable efficient manipulation and analysis of genetic information. Reductions in the cost of DNA sequencing have outpaced Moore's law (a doubling in microprocessor computing power every two years), especially following the transition from Sanger-based (dideoxy chain termination sequencing) capillary electrophoresis sequencing to second-generation ("massively parallel") technologies in the mid 2000s.[1] Several competing massively parallel sequencing technologies have been brought to market, including but not limited to pyrosequencing (Roche/454, Basel, Switzerland), sequencing-by-ligation (Life Technologies, Carlsbad, USA), sequencing-by-synthesis (Illumina, San Diego, USA), optical single-molecule sequencing (Pacific Biosciences, Menlo Park, USA) and nanopore technology (Oxford Nanopore, Oxford, UK). All the

massively parallel technologies differ significantly from Sanger capillary sequencing in that they occur in a nucleotide-by-nucleotide stepwise fashion, eliminating the need for discrete separation and detection of Sanger sequencing reaction products on a capillary instrument. An entire human genome can now be sequenced in a matter of hours for less than US$1,000 (a dramatic reduction from US$70,000,000 in 2002).[2] This revolution in genomics has now made whole-genome sequencing experiments routine and has enabled initiatives to sequence entire ecosystems. [3]

The ability to chemically synthesize sequences of DNA greatly facilitates production and optimization of genetically encoded products. DNA synthesis capabilities have also improved at an exponential pace.[4] Synthetic oligonucleotides have been available for 20 years, and synthetic genes built commercially from those "oligos" for the last 10. The ability to synthesize increasingly longer stretches of DNA is improving rapidly as evidenced from the following milestone "firsts": a synthetic gene in 1979 (207 base pairs, bp), a synthetic plasmid in 1990 (2,100 bp), and a synthetic genome in 2008 (583,000 bp).[5] Because total chemical synthesis remains limited to relatively short DNA molecules (<200 nucleotides), novel enzymatic methods have been developed to assemble synthetic oligos into the desired DNA duplex length.[6] These methods can be broadly categorized as restriction enzyme-dependent or sequence-independent overlap techniques, either of which may be employed depending on the gene of interest.[7] *De novo* synthesis of entire genomes requires assembly at the chromosomal scale, which can be achieved by *in vivo* recombination in yeast or *Bacillus* by making use of native recombination mechanisms.[8] For example, in the 580 kb synthesis for the genome of

Mycoplasma genitalium, in vitro recombination was used to assemble 101 overlapping cassettes each 5–7 kb in length into 4 DNA fragments of 144 kb, which were subsequently assembled into the complete genome by *in vivo* recombination in yeast.[5c] *De novo* DNA synthesis is a key enabling technology for synthetic biology because it permits access to arbitrary genes and even whole biosynthetic pathways directly from digital information (DNA sequences from databases), thus bypassing the need to clone genetic material from biological samples.

2.2. Metagenomics and Bioinformatics

Nature presents an immense collection of genes encoding biocatalysts that can be directly applied in chemical syntheses. The most diverse source of biocatalysts in nature is arguably the prokaryotes (archaea and bacteria), comprising millions of species that display an extraordinary diversity of niche, habitat range, and metabolisms.[9] However, genomic information for only around 6,000 prokaryotic species is available in commercial culture collections, and it is estimated that 99% of all microorganisms may be recalcitrant to cultivation. Although it is not possible to characterize the enzymes and small molecules produced by uncultured bacteria using traditional microbiological methods, it is possible to extract microbial DNA directly from an environmental sample and clone this DNA into easily cultured bacteria. The utility of this approach for the discovery of novel biocatalysts and small molecules has been extensively reviewed.[10] In short, these “metagenomic” libraries can be analyzed by two general strategies: (i) functional screening, which requires the heterologous expression of environmental DNA

in a model cultured host to yield a phenotype of interest, and (ii) homology screening, which relies on DNA sequence similarity to identify genes of interest.

The first approach typically uses one of three classes of functional screens: (i) direct readout assays, (ii) reporter genes, and (iii) complementation. Direct readout assays are applicable when the enzyme reaction product can be coupled to a spectroscopic signal or to the formation of a halo around the clone encoding the gene for the desired enzyme. For example, halo formation resulting from tributyrin ester hydrolysis in agar plates was used to screen a metagenomic library from the Antarctic desert soil, which led to the discovery of a cold-active alkaliphilic esterase only distantly related to reported lipases.[11] Reporter genes may be used in special cases where the enzymatic product is known to elicit a response from a transcription factor. For example, a benzoate-sensitive transcription factor controlling green fluorescent protein expression was used to identify 11 amidases distantly related to known amidases from a wastewater sludge metagenomic library based on their activity on benzamide.[12] Genetic complementation assays may also be used, with very high throughput. Genencor-DuPont screened metagenomic libraries for a novel glycerol dehydratase that catalyzes the bottleneck reaction in the heterologous fermentation pathway to make 1,3-propanediol from sugars (see section 3.6).[13] The complementation screen was based on an *E. coli* strain that contained and expressed all genes necessary for anaerobic glycerol breakdown except the genes coding for glycerol dehydratase. The only way for such a strain to survive under anaerobic conditions with glycerol as the sole carbon and energy source is complementation by a gene that confers

the glycerol or diol dehydratase activity. Two out of 560,000 *E. coli* clones tested contained active dehydratases with superior properties.[13]

Novel enzymes can also be identified from their sequence similarity to enzymes that possess the desired activity. This homology-based screening can be carried out *in vitro*, using polymerase chain reaction (PCR) and colony hybridization, or *in silico*, using bioinformatics. This approach is particularly suited for studying the production of secondary metabolites in prokaryotes, since the genes involved are typically clustered on the chromosomes. For example, homology screening was used to identify gene clusters associated with the biosynthesis of natural products derived from bacteria that exist in symbiosis with animals, including medicinally important pederins and patellamides.[14] Ultimately, heterologous expression of the gene clusters provides a renewable source of these compounds, which could not otherwise be isolated in sufficient quantities to permit extensive biological testing.

"Synthetic metagenomics" is another form of sequence-based screening that builds on advances in gene synthesis and bioinformatics. Several free web-based bioinformatics tools are available for aiding experimental design (table 1). For example, the NCBI sequence database, which comprises both cultured and uncultured genomes, can be used to identify genes that are similar to a given enzyme gene. These genes can be optimized for expression in a particular heterologous host, chemically synthesized, cloned, and expressed in the selected host. This approach was used for the discovery and characterization of methyl halide transferases, which catalyze the transfer of a methyl group from the ubiquitous metabolite S-adenosyl methionine (SAM) to a halide ion.[15] A

BLAST search of the NCBI database using known methyl chloride transferases yielded 89 putative methyl halide transferases from sources spanning the biological kingdoms, with a remarkable degree of sequence diversity (average of 28% amino acid identity between sequences). These genes were codon-optimized, expressed in *E. coli*, and analyzed for methyl halide synthesis. More than half were active on halides (X^- = Cl^-, Br^- and I^-), and subsequent engineering of the most active methyl halide transferase in *S. cerevisiae* resulted in fermentative production of methyl halides from glucose with a volumetric productivity of 190 mg L^{-1} h^{-1}.[15]

Table 1. A compilation of generally useful publicly available bioinformatic tools and resources

Database	URL	Description
Genbank—National Center for Biotechnology Information (NCBI)	http://www.ncbi.nlm.nih.gov/genbank/	Annotated collection of all publicly available DNA sequences
BLAST (Basic Local Alignment Search Tool)—NCBI	http://blast.ncbi.nlm.nih.gov/Blast.cgi	Compares a query sequence with a database of nucleotide or amino acid sequences
BRENDA Enzyme Database	http://www.brenda-enzymes.info/	Kinetic, thermodynamic, structural and reaction information for enzymes
KEGG (Kyoto Encyclopedia of Genes and Genomes)	http://www.genome.jp/kegg/	Searchable metabolic pathway maps for various organisms
DR Nelson Cytochrome P450	http://drnelson.uthsc.edu/CytochromeP450.html	Reference for over 11,500 CYPs
eQuilibrator	http://equilibrator.weizmann.ac.il/	Computation of thermodynamic parameters for enzymatic reactions and metabolic pathways
Genome News Network (GNN)	http://www.genomenewsnetwork.org/	Updates on developments in genomics research
American Type Culture Collection (ATCC)	http://www.atcc.org/	Source of microbial strains, vectors, genomic and cDNA libraries
FoldX	http://foldx.crg.es/	Estimates the contributions of mutations/interactions to the stabilities of proteins and protein complexes

EcoCyc (Encyclopedia of *E. coli* K-12 genes and metabolism)	http://ecocyc.org/	Transcriptional regulators, transporters, metabolic pathways; flux balance metabolic model
YeastCyc	http://pathway.yeastgenome.org/	152 metabolic pathways from *S. cerevisiae*
MetaCyc	http://biocyc.org/metacyc/index.shtml	1,877 metabolic pathways from 2,263 organisms
GLAMM	http://glamm.lbl.gov/	Regulation of metabolic pathways under a range of environmental conditions
Gene Designer (DNA 2.0)	https://www.dna20.com/genedesigner2/	Codon optimization of genes for heterologous expression

2.3 Protein Expression

Once a gene encoding an enzyme of interest is identified, it must be expressed by some organism in a functional form. Expression level targets differ depending on whether the enzyme is used to catalyze a reaction in isolation or in the context of a well-balanced metabolic pathway. For example, metabolic engineering requires precise relative control of the expression of several enzymes (as opposed to the overexpression of a single enzyme) to prevent the accumulation of potentially toxic intracellular intermediates and to prevent the diversion of carbon and energy to other products. Expression levels must be carefully tuned rather than simply maximized to ensure high flux through the target pathway. This has been accomplished by varying the gene copy number, altering promoter sequences, and utilizing designer transcription factors.[16]

On the other hand, production of isolated enzymes for single reactions is largely an exercise in maximizing expression of functional enzyme. This is often accomplished by optimizing culture conditions (e.g., time, temperature, media, aeration, etc.), but more recently gene modification has also been used to improve enzyme expression levels.[17] Modified leader sequences and optimization of codon choices have improved transcription and translation, respectively, but solubility (if dealing with a cytosolic enzyme) cannot generally be improved using these methods. This can be addressed in some cases by lowering the expression temperature, fusing a solubility-inducing protein to the target enzyme, or even by expressing truncated soluble versions of the target enzyme.[18] Coexpressing target enzymes with chaperones has also been effective.[19] These are largely empirical techniques, and their mode of action remains the subject of active investigation, but they can be useful tools in the preparation of soluble enzymes.

The most important factor influencing protein expression is the host organism used. *E. coli* is often chosen due to the expansive set of genetic tools and relatively simple culturing requirements for this organism. However, *E. coli* is by no means appropriate for all enzymes and biosynthetic pathways, and one may choose from a wide range of organisms for a given application.[20] In general, a host organism must possess the ability not only to synthesize the gene product, which requires appropriate codon usage and GC content, but also to incorporate cofactors, form disulfide bonds, and perform any required posttranslational modifications. Many industrial enzymes are expressed in hosts with good secretion capabilities (e.g., filamentous fungi and *Bacillus sp*.), so that the

enzyme can be precipitated or filtered directly from the supernatant, or even used without purification.

3. Enzyme Engineering

Enzymes produced using the methods outlined above are valuable catalysts for a range of chemical reactions. Extensive reviews on the utility of such catalysts[21] and the engineering of substrates[22] to suit those catalysts have been published. In many cases, however, enzymes themselves are not ideally suited to the task for which they are charged, and properties such as thermal stability, organic cosolvent and pH tolerance, substrate scope, or selectivity may require improvement. Although some of these factors may be addressed through the use of specialized bioreactors, enzyme immobilization, or other process optimization,[23] modification of the enzyme itself has proven to be a powerful means of improving biocatalytic processes.[24] Indeed, the ability to engineer enzymes gives them a significant advantage over small molecule catalysts. For example, enzymes may be tuned for specific applications using a systematic process of mutation and selection known as directed evolution.[25]

We provide some examples of directed evolution of synthetically relevant enzymes,[24] to illustrate the types of improvements that have been made. We specifically focus on the use of directed evolution to alter substrate scope or reaction selectivity, rather than process-related properties such as thermal stability, solvent tolerance, or pH.[26] The choice of sequence modification approach depends on the problem; for example, how much is known about that enzyme? Is a crystal structure available or can a homology model be generated? Practical constraints also exist. Can libraries of mutant

enzymes be generated readily, and is a high throughput assay possible? Thus, the examples presented illustrate the results of various approaches and are grouped based on the type of reaction catalyzed rather than the methods employed.[24] We avoid discussion of specific technologies and strategies for introducing sequence modifications (EP-PCR, DNA shuffling, site-directed recombination, ISM, circular permutation, etc.), as they have been thoroughly discussed elsewhere.[27]

3.1. Oxygenation

Given the importance of oxygenation in synthetic organic chemistry, it is not surprising that several different classes of oxygenases have been engineered.[28] Of these, the cytochromes P450 (P450s) have been the subject of a particularly large volume of work due to their ability to hydroxylate unactivated C-H bonds and epoxidize olefins (figure 1).[29] Several recent examples highlight the power of engineered P450s to hydroxylate a wide range of substrates whose structures differ significantly from the native substrates of these enzymes. The applications of engineered P450s in biocatalysis have been reviewed.[29-30]

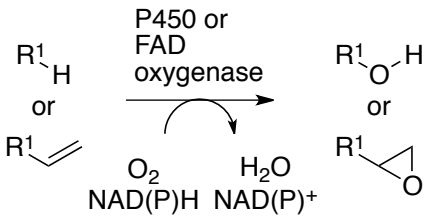


Figure 1. General scheme for oxygenase-catalyzed C-H hydroxylation and olefin epoxidation.

Cytochrome $P450_{BM3}$ from *Bacillus megaterium* (BM3) has been the subject of a large number of engineering efforts.[31] In this enzyme, the heme and diflavin reductase domains (FMN and FAD) are fused in a single polypeptide chain, rendering BM3 self-sufficient in generating the active oxidant from dioxygen and NAD(P)H. BM3 is soluble, readily overexpressed in a variety of heterologous hosts, and requires only atmospheric oxygen and a supply of nicotinamide adenine dinucleotide phosphate (NADPH) for monooxygenase activity. Arnold and Lewis have used directed evolution to alter the substrate scope of BM3, from long chain fatty acids to gaseous alkanes, including propane and ethane.[32] Some of the intermediate mutants along the evolutionary lineage of the so-called BM3 propane monooxygenase (PMO) were found to have activity on a wide range of small organic molecules. These enzymes commonly catalyze both olefin epoxidation and C-H hydroxylation, and hydroxylation of secondary C-H bonds or those alpha to heteroatom substitution is particularly efficient. This activity has been enhanced to provide catalysts suitable for preparative scale hydroxylation of drugs, protected sugars, steroids, alkaloids, and other small molecules (figure 2). Other research groups, including those of Reetz, Schmid, and Fasan, have described the directed evolution of BM3 for hydroxylation of substrates ranging from methane to aromatic heterocycles to steroids and terpenes.[29]

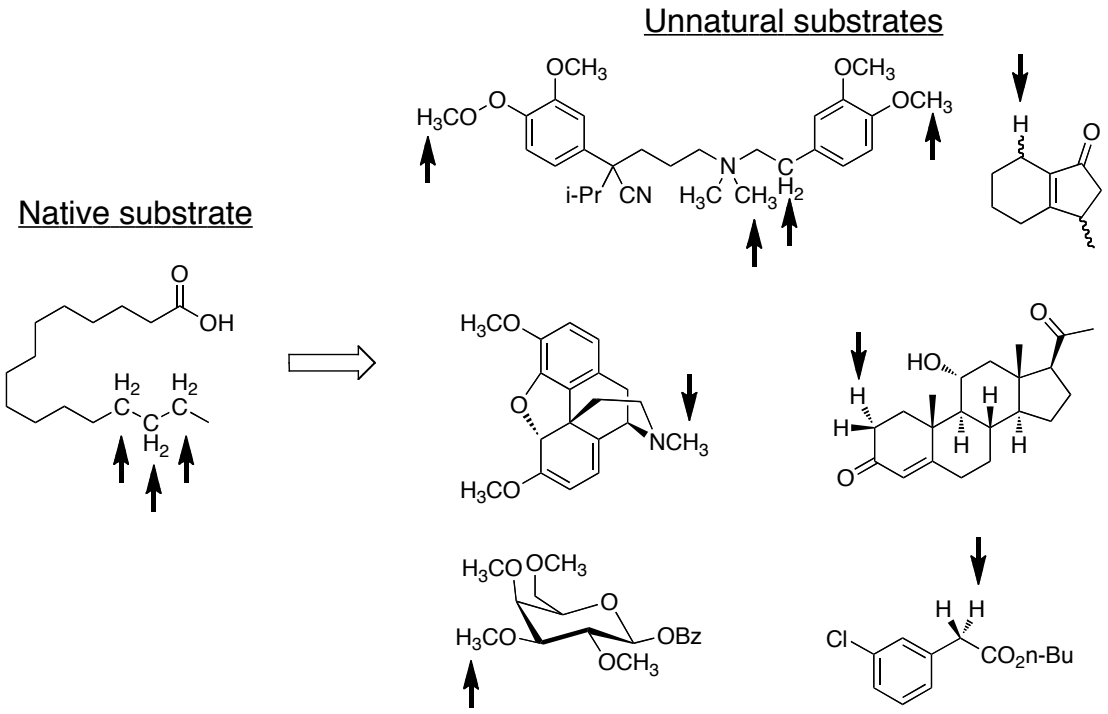


Figure 2. A native substrate for cytochrome $P450_{BM3}$ and representative substrates accepted by BM3 variants.[32] Arrows denote sites of hydroxylation. Block arrow denotes engineered activity on unnatural substrates.

Cytochrome P450 CAM from *Pseudomonas putida* has also been engineered as an oxygenation catalyst. Hydroxylations of substrates similar to those outlined above for BM3, including terpenes, arenes, polychlorinated arenes, and even ethane, have been accomplished using CAM variants.[29] Both cytochrome P450 CAM and BM3 accept small molecules (camphor and dodecanoic acid) as substrates, but significant sequence engineering effort is typically required to achieve high activity on larger substrates. In contrast, PikC, a P450 from *Streptomyces venezuelae* that catalyzes hydroxylation of an intermediate in the biosynthesis of the macrolactone YC-17 (22 carbon atoms),[33] can also hydroxylate a range of other macrocyclic compounds. Efforts are underway to expand this reactivity and enable hydroxylation of large substrates.

FAD-dependent oxygenases use an FAD hydroperoxide oxidant to catalyze oxygenation of activated substrates. Directed evolution of 2-hydroxybiphenyl 3-monooxygenase led to variants with improved activity on several substrates, including 2-sec-butylphenol, 2-methoxyphenol, and 2-tert-butylphenol (figure 3A).[34] Similar efforts improved the hydroxylase activity of vanillyl-alcohol hydroxylase on cresol[35] and epoxidation activity of styrene monoxygenase from *Pseudomonas putida* CA-3 on styrene, indene, and indole[36] (figure 3B, C).

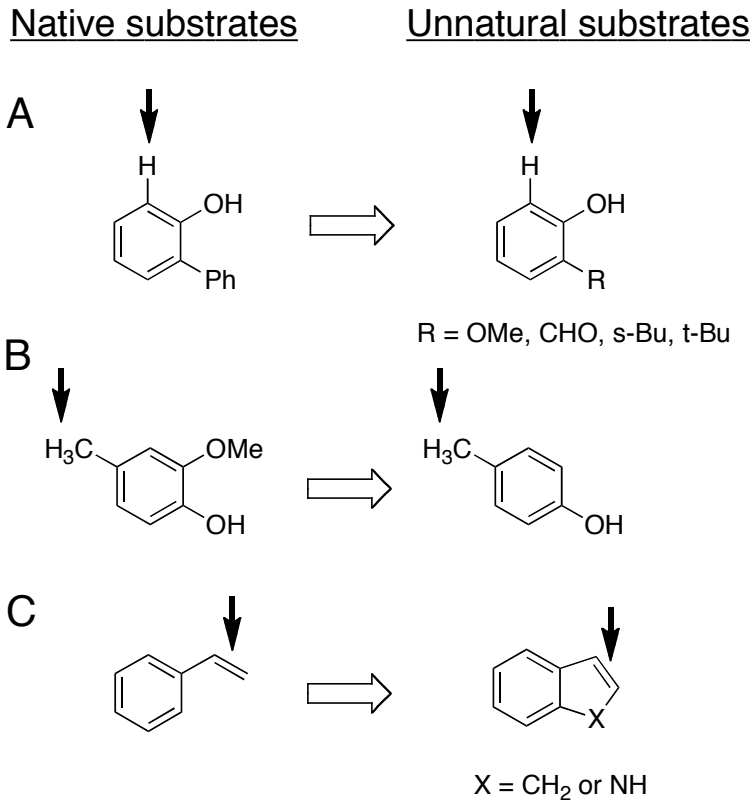


Figure 3. Native substrates and representative substrates accepted by variants of (A) 2-hydroxybiphenyl 3-monooxygenase,[34] (B) vanillyl-alcohol hydroxylase,[35] and (C) styrene monoxygenase from *Pseudomonas putida* CA-3.[36]

Extensive evolution of FAD-dependent Baeyer-Villiger monooxygenases,[37] which catalyze the eponymous chemical reaction on a range of native carbonyl and thiocarbonyl containing substrates (figure 4A), has been reported. This work focused on improving activity and enantioselectivity on unnatural substrates, typically involving desymmetrization of prochiral ketones or kinetic resolution of racemic ketones.[38] For example, cyclohexanone monooxygenase (CHMO) was engineered to oxidize 4-hydroxycyclohexanone to either enantiomer of the corresponding lactone (figure 4B), and related variants have been found to accept a wide range of substrates (figure 4C). Many of these enzymes have been employed on large scales (e.g., 200 L), including the reactions of rac-bicyclo-[3.2.0]-hept-2-en-6-one and rac-2-phenylcyclohexanone, which provide near-quantitative conversion and high enantioselectivities (E-values >100). Similarly high enantioselectivities are possible for the resolution of a range of substituted cyclohexanones using variants of phenylacetone monooxygenase (PAMO) (figure 4D).

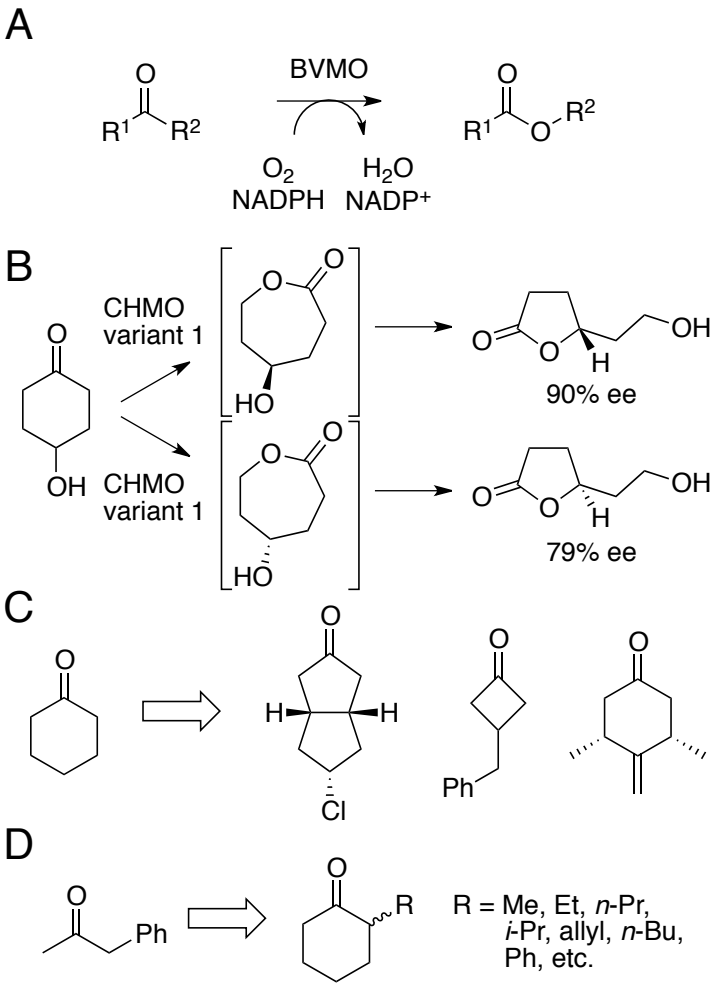


Figure 4. (A) General scheme for Baeyer-Villiger monooxygenase-catalyzed oxidation of carbonyl compounds. (B) Enantioselective oxidation of a *meso* ketone. (C) Native substrate and representative substrates accepted by variants of cyclohexanone monooxygenase. (D) Native substrate and representative substrates accepted by variants of phenylacetone monooxygenase.[37-38]

Finally, even multicomponent, nonheme iron oxygenases have been engineered by directed evolution. Mutations in the oxygenase subunit of these enzymes alter both their site selectivity and substrate scope. Various arene dioxygenases have been engineered to accept unnatural substrates in order to make useful diene diols and phenols (figure 5A).[39] For example, biphenyl dioxygenase variants were used to make the

dioxygenated (and in some cases monooxygenated) products of a wide range of heterocycles and arenes (figure 5B).[40]

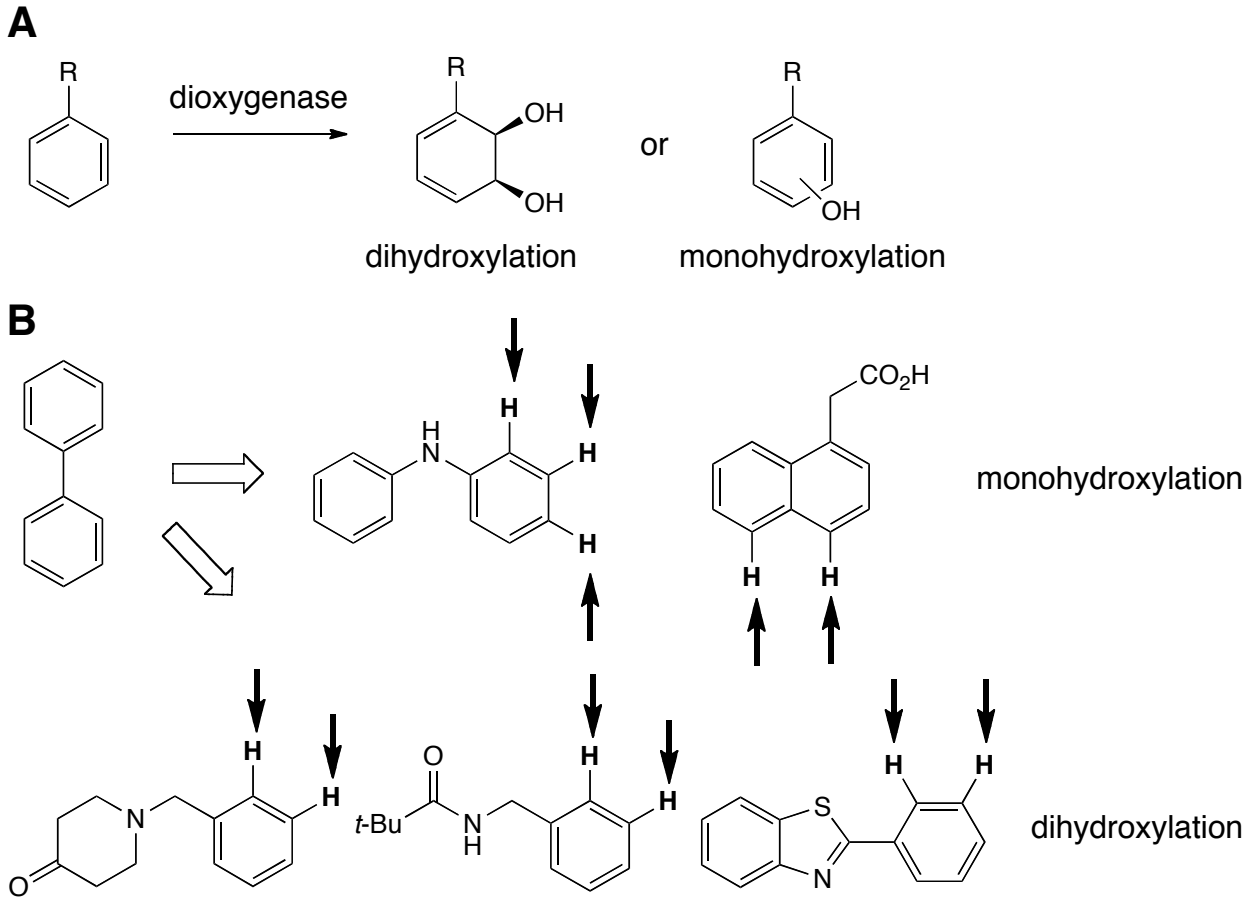


Figure 5. (A) General scheme for dioxygenase catalyzed arene oxidation.[39] (B) Native substrate and representative substrates accepted by variants of biphenyl dioxygenase. [40]

Wood and others have made extensive efforts to expand the scope of toluene and xylene monooxygenases. Variants with improved activity and altered regioselectivity on indole[41] and naphthalene[42] have been reported (figure 6A). Furthermore, variants of toluene p-monooxygenase from *R. pickettii* PKO1 are capable of hydroxylating toluene at the ortho, meta, and para positions (figure 6B).[43]

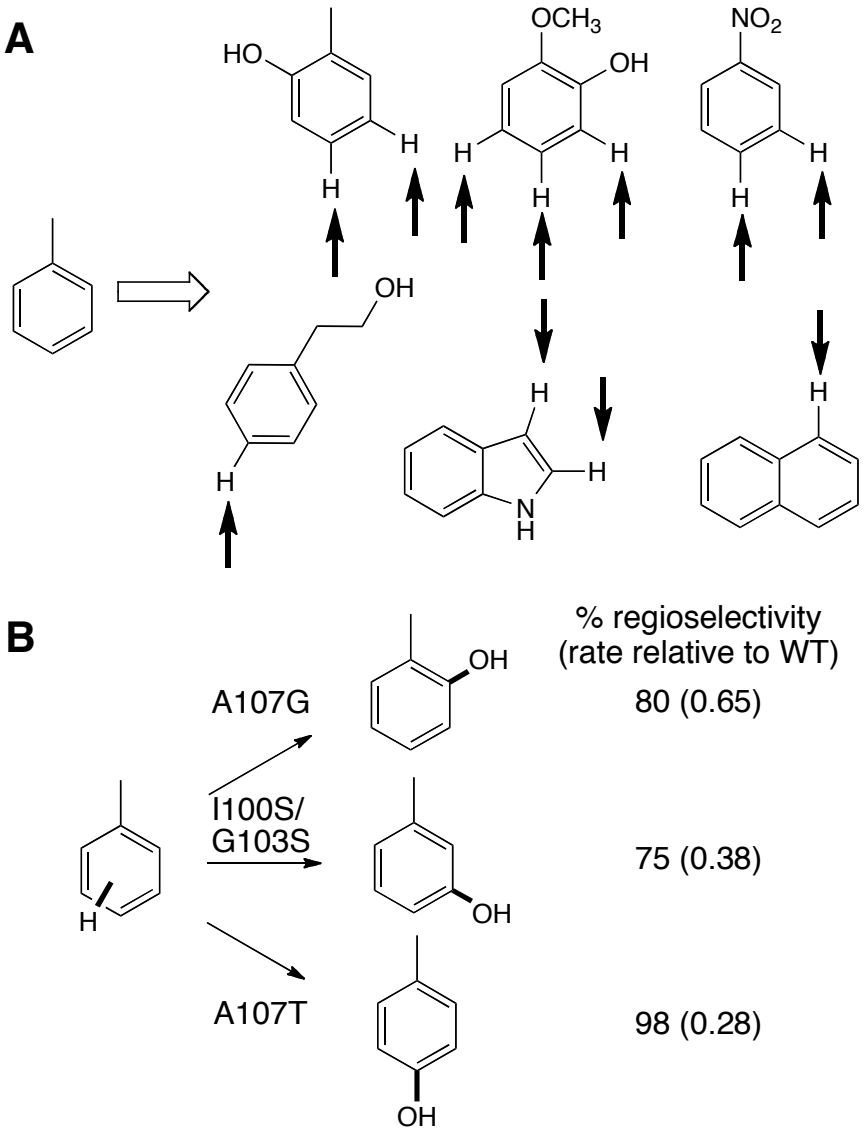


Figure 6. (A) Native substrate and representative substrates accepted by variants of toluene and xylene monooxygenases.[41-42] Sites of hydroxylation are denoted with arrows. (B) Hydroxylation of ortho, meta, and para positions of toluene using variants of a toluene monooxygenase.[43]

3.2. Halogenation

Enzymatic halogenations typically involve oxidation of halide anion to an X^+ (X = halogen) equivalent that can react with electron-rich substrates. Directed evolution of heme-[44] and vanadium-dependent[45] haloperoxidases has improved the scope of biocatalytic halogenation. The site selectivity of halogenation reactions catalyzed by haloperoxidases has been questioned due to their release of free HOCl into solution.[46] On the other hand, FAD-dependent halogenases are known to catalyze site-specific halogenation of their native substrates. Thus, Van Pee and co-workers utilized site directed mutagenesis to alter the site selectivity of tryptophan chlorination and bromination by the halogenase PrnA.[47] More recently, O'Connor and co-workers engineered the halogenase RebH to act on tryptamine, rather than its native substrate, tryptophan.[48]

3.3. Dehydrogenation

Dehydrogenases and a variety of oxidases catalyze the oxidation of organic substrates, including amines, alcohols, and sugars, via dehydrogenation (figure 7A). The activities of these enzymes can also be reversed under suitable conditions to enable hydrogenation of unsaturated compounds (vide infra). The utility of the products of both modes of operation has motivated many successful examples of engineering these enzymes for synthetic applications.

Amino acid dehydrogenases have been engineered to convert unnatural amino acids to the corresponding imines, which subsequently undergo hydrolysis to generate

alpha-keto acids (figure 7B).[49] Although not directly relevant to organic synthesis, much work has also focused on improving alcohol, sugar, and other dehydrogenases for producing reducing equivalents needed by other redox enzymes.[50] Several oxidases also catalyze dehydrogenation of C-O and C-N bonds. Amino acids are also substrates for these enzymes, and directed evolution has been used to expand the range of amino acids that these enzymes accept.[51] Monoamine oxidase from *Aspergillus niger* has been studied as a catalyst for chemo-enzymatic deracemization of amines. In this process, the oxidase converts a single enantiomer of the racemic amine to an imine, which is subsequently reduced by a nonspecific reducing agent (figure 7C).[52] Directed evolution has been used to expand the scope of this enzyme from several simple primary amines to a very large range of secondary amines (figure 7D).[52] In a similar fashion, galactose oxidase, which naturally oxidizes the primary alcohol of galactose to the corresponding aldehyde has been used to catalyze oxidation of secondary alcohols to the corresponding ketones, enabling kinetic resolution of the alcohol enantiomers (figure 7E).[53] Directed evolution of several other pyranose oxidase enzymes has also been described.[54]

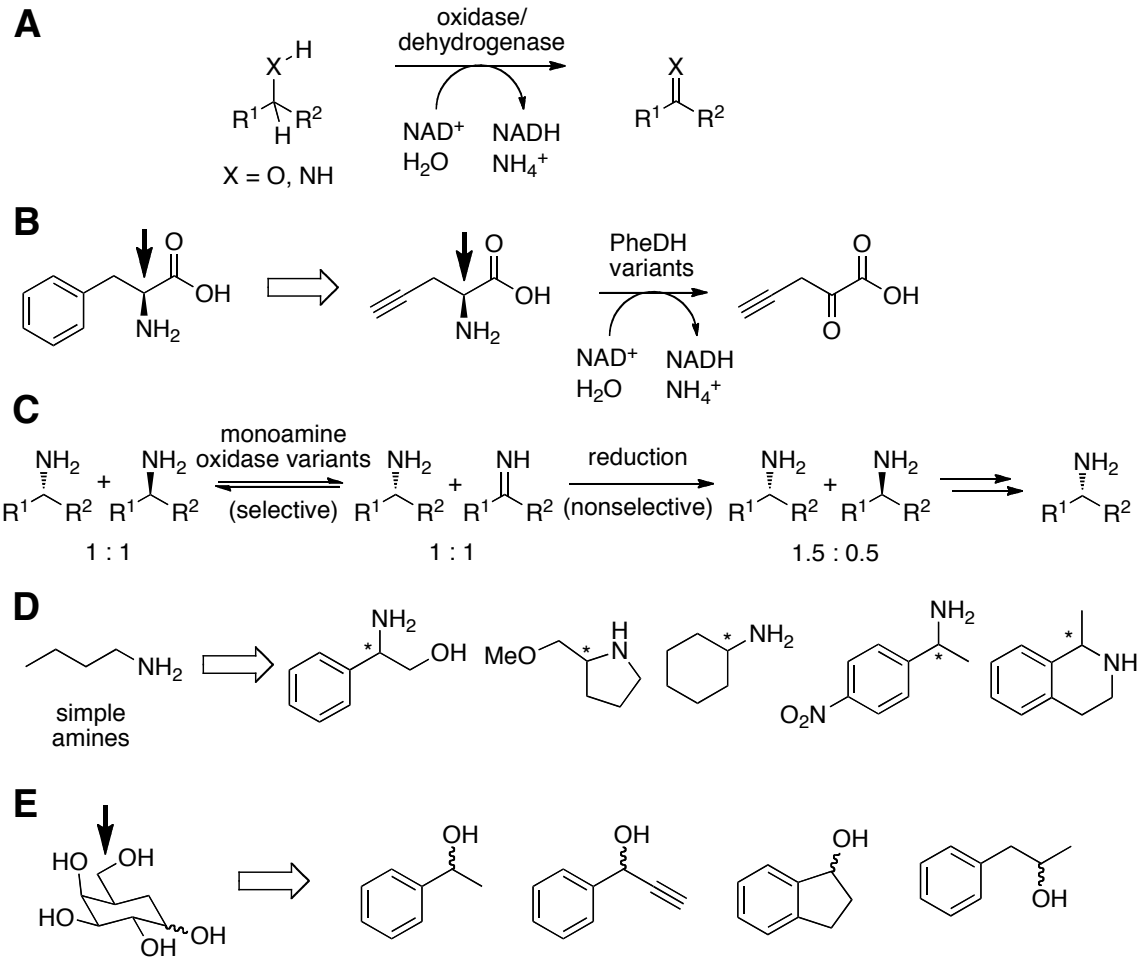


Figure 7. (A) General scheme for oxidase and dehydrogenase catalyzed reactions. (B) Native substrate and representative substrate accepted by a variant of phenylalanine dehydrogenase. The site of dehydrogenation is denoted by an arrow.[49] (C) General scheme for oxidase catalyzed deracemization of amines.[52] (D) Native substrate and representative substrates accepted by variants of a monoamine oxidase from *Aspergillus niger*. (E) Native substrate and representative substrates accepted by variants of a galactose oxidase. [53]

3.4. Hydrogenation

As noted above, dehydrogenases can operate in "reverse" to catalyze hydrogenation of unsaturated carbon-heteroatom bonds (i.e., the reverse of figure 7A). This direction has significant importance due to the formation, rather than destruction or resolution, of a stereogenic center. Furthermore, these enzymes possess a remarkably

broad substrate scope. For example, the S-selective dehydrogenase from *Rhodococcus erythropolis* and the R-selective dehydrogenase from *Lactobacillus kefir* have been used extensively as ketoreductases.[55] These enzymes accept a wide range of simple aldehydes and ketones as substrates for reduction and the corresponding alcohols as substrates for oxidation. In most cases a specific cognate substrate for the wild-type enzyme is not known. Despite the utility of these enzymes, the desired selectivity is not always achieved, and directed evolution has been used to achieve highly selective reduction of a wide range of ketones, including aryl ketones, α-amino ketones, cyclic ketones, and β-dicarbonyl compounds, many on a commercial scale (figure 8).[55] Importantly, both product enantiomers typically can be obtained using variants of one of the two aforementioned ketoreductases. Enolate reductases have also attracted much attention due to their ability to effect stereoselective hydrogenation of β-substituted α,β-unsaturated ketones.[56]

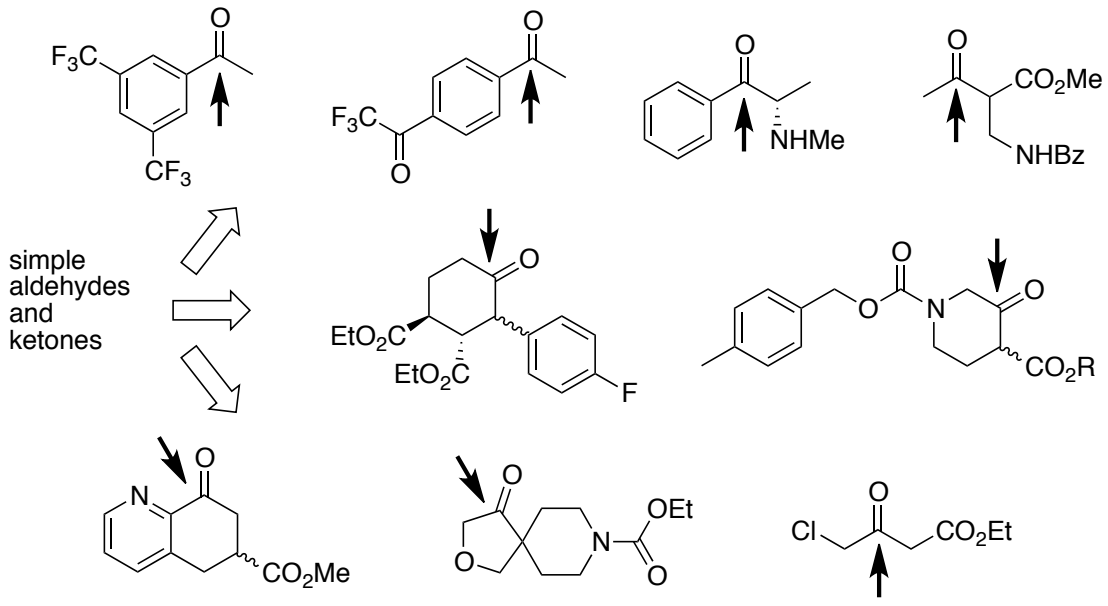


Figure 8. Representative substrates accepted by ketoreductase variants. The site of reduction is denoted by an arrow.[55]

3.5. Group Transfers

Several different enzyme classes transfer functional groups and even entire molecular fragments from donor substrates to target substrates. Particularly notable in this regard are the pyridoxyl phosphate dependent amino transferases, also known as transaminases, which enable synthesis of chiral amines by exchanging amine and ketone functional groups on donor (typically amino acids or simple short chain alkyl amines) and acceptor substrates, respectively (figure 9A). The ω-transaminases have been the subject of the most engineering efforts due to the fact that they can accept a wide range of amine (rather than amino acid) donor substrates. These have been engineered to utilize simple amines as donor substrates with high efficiency. The reaction scope of the ketone acceptor substrate has been greatly expanded to enable the amination of a wide range of acyclic and cyclic ketones bearing a variety of functional groups (figure 9B).[57] Variants capable of providing either enantiomer of the amine product have also been generated. Perhaps most impressively, as highlighted in another section of this volume, researchers at Merck and Codexis engineered a transaminase capable of installing a chiral amine in the final step of the synthesis of the drug Sitagliptin.[58]

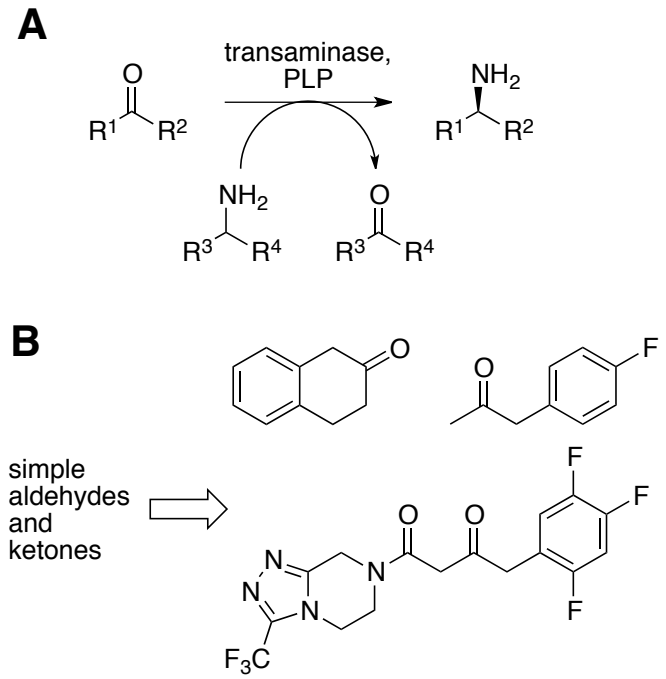


Figure 9. (A) General scheme for transaminase-catalyzed reductive amination. (B) Representative substrates accepted by transaminase variants.[57-58]

Glycotransferases have also benefited from directed evolution.[59] These enzymes transfer single sugar moieties to a range of aglycone (i.e., small molecule) substrates (figure 10 A), and directed evolution has been used to expand the scope of aglycones that may be utilized (figure 10B/C). Glycosidases have also been used as glycotransferase catalysts. These enzymes possess transglycosidase activity in addition to hydrolytic activity, and the former has been improved through rational design of variants lacking a nucleophilic active site residue required for hydrolytic activity.[60] The resulting artificial transglycosidases enable glycosylation of a range of substrates, including sugars (to generate oligosaccharides) as well as aglycones (figure 10D).

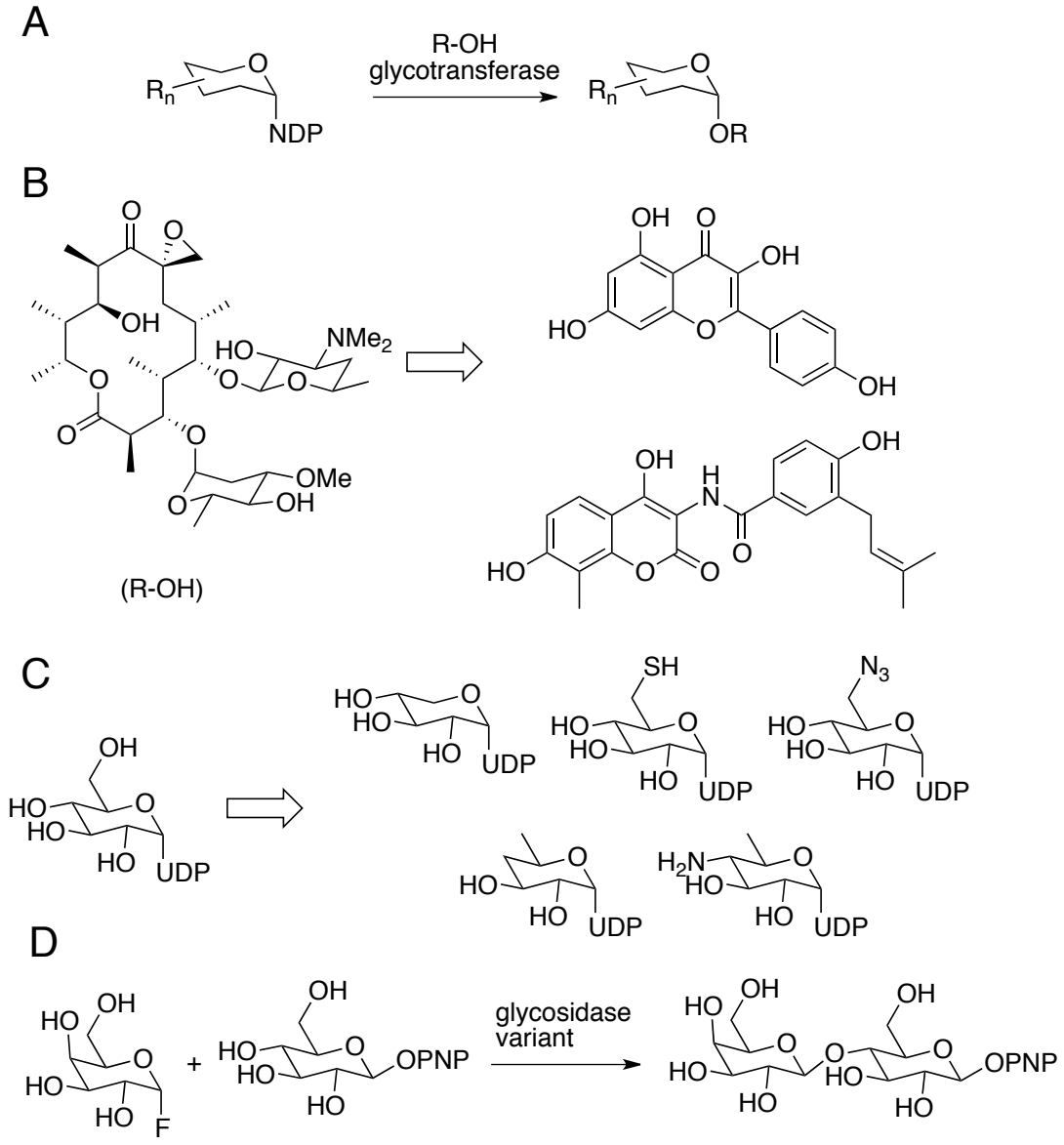


Figure 10. (A) General scheme for glycotransferase coupling of a nucleotide diphosphate (NDP) glycosyl donor and a glycosyl acceptor (R-OH). (B) Native glycosyl acceptor and acceptors accepted by glycotransferase variants.[59] (C) Native glycosyl donor and glycosyl donors accepted by glycotransferase variants.[59] (D) Representative glycosylation reaction catalyzed by an artificial transglycosidase.[60]

Directed evolution of sulfuryl-,[61] phosphoryl-,[62] glutathione,[63] and many other transferases has also been described. These enable highly selective tailoring of substrates

with biologically important functional groups that are highly challenging to install using conventional synthetic methods.

3.6. Hydrolysis and Condensation

Hydrolases remain among the most commonly used enzymes in organic synthesis. In the forward direction, these enzymes hydrolyze a range of functional groups, including esters, amides, and nitriles, and many of these enzymes can be conveniently operated in the reverse direction to enable formation of these same functional groups. Directed evolution has been applied extensively to optimize hydrolase activity in both directions. While not all of these can be covered here, the examples presented should provide readers with an idea of how these enzymes are typically used for chemical synthesis.

Lipases, which catalyze ester hydrolysis or formation (figure 11A), were among the first enzymes to be evolved for improved organic solvent tolerance and enantioselectivity, and have received extensive attention since these early efforts.[27] Engineered lipases enable efficient kinetic resolution of racemic esters and desymmetrization of meso diesters that are dramatically different, in many cases, from those accepted by the wild type enzymes. Carboxyl esterases possess very similar hydrolytic activities[64] and have been engineered to improve enantioselectivity toward a range of substrates. Hydrolysis of carboxylic acid esters, including ibuprofen esters, 2-phenylalkanoic acid esters, and allene esters, has been achieved, and various alcohol leaving groups ranging from phenols to primary alcohols to tertiary alcohols can be used (figure 11B). Among the most commonly used enzymes are those from *Pseudomonas*

aeruginosa (PAL), *Bacillus subtilis* (BSLA), C*andida antarctica* (CALA/CALB),[65] and *Burkholderia cepacia* (BCL).

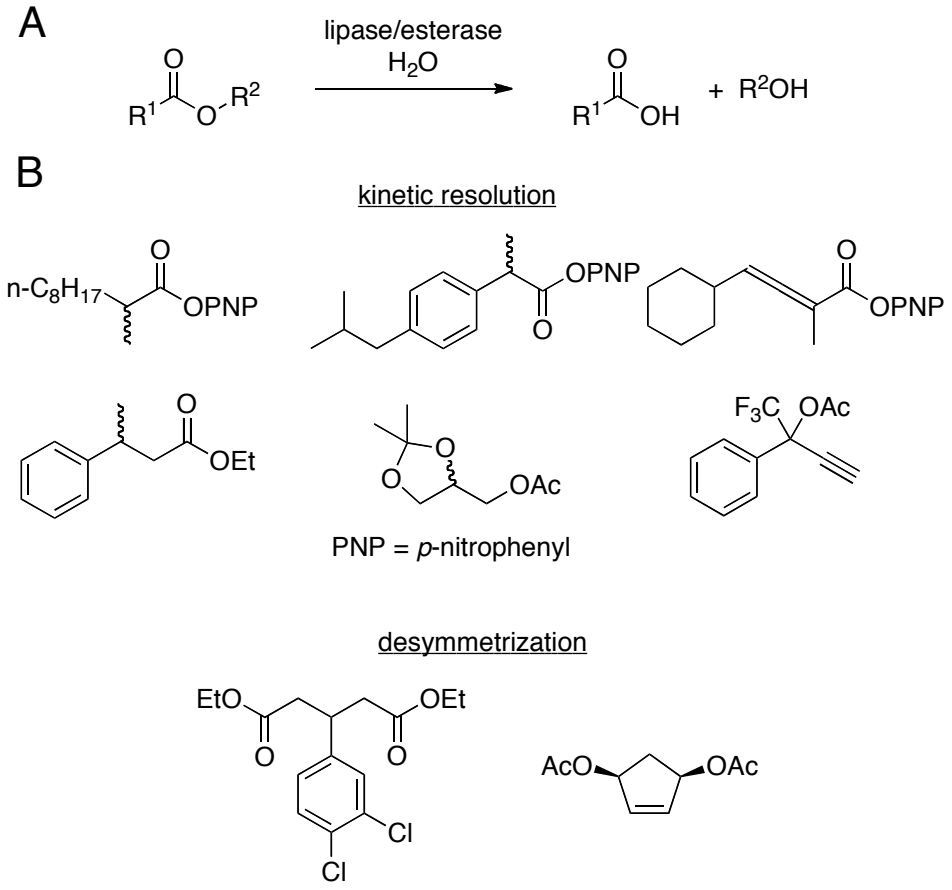


Figure 11. (A) General scheme for lipase- and esterase-catalyzed hydrolysis reactions. (B) Representative substrates accepted by lipase and esterase variants for either kinetic resolution or desymmetrization.[64-65]

Epoxides have long been recognized as valuable synthetic intermediates, and epoxide hydrolases convert these compounds to enantio-enriched diols (figure 12A). These enzymes are typically used as catalysts for kinetic resolution of racemic epoxides, and the epoxide hydrolase from *Aspergillus niger* has been a frequent target of

engineering efforts (figure 12B).[66] Halohydrin dehalogenases can also catalyze epoxide hydrolysis; however, these enzymes possess the added ability to resolve racemic halohydrins via epoxide formation prior to catalyzing epoxide opening with water (hydrolysis) and a range of additional nucleophiles, including azide, cyanide, and halide (figure 12C). In a particularly notable example, directed evolution improved a halohydrin dehalogenase for production of the side chain of the commercial drug atorvastatin (figure 12C, Nu = CN, R = CH_2CO_2Et).[67] Several examples of similar efforts have been described, including engineering the halohydrin dehalogenase from *Agrobacterium radiobacter* for greater enantioselectivity on 2-chloro-1-phenethyl alcohol.[24]

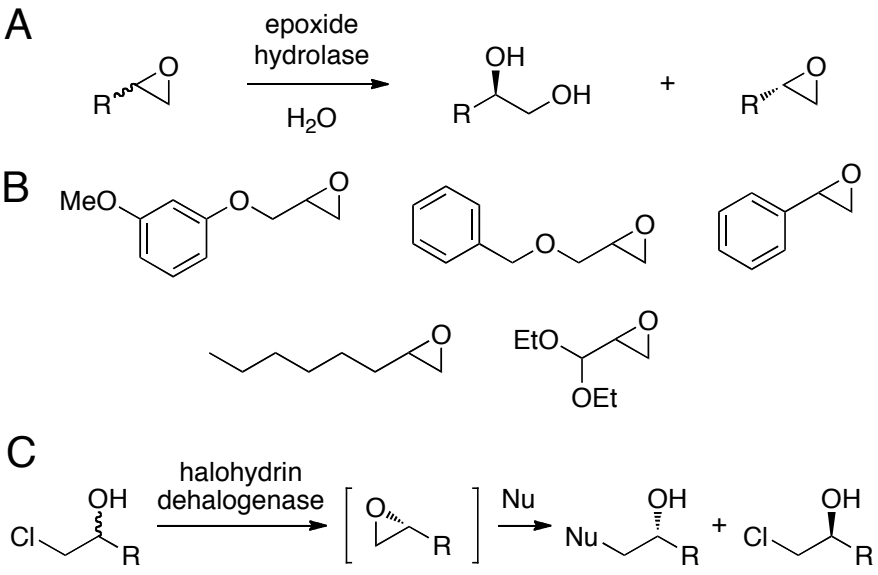


Figure 12. (A) General scheme for epoxide hydrolase catalyzed reaction. (B) Representative substrates accepted by epoxide hydrolase variants.[66] (C) General scheme for halohydrin dehalogenase-catalyzed addition of nucleophiles to 2-haloalcohols.[68]

3.7. Nonoxidative, Nonhydrolytic Bond Formation

Lyases catalyze nonoxidative, nonhydrolytic bond cleavage; when catalyzing the reverse direction (bond formation), these enzymes are called synthases. Perhaps among the most recognizable members of this class of enzymes to the synthetic chemist are the aldolases. Given the importance of the aldol reaction in synthetic organic chemistry, it is hardly surprising that enzymes capable of catalyzing this reaction on highly functionalized, unprotected substrates have received considerable attention. These enzymes utilize either an active site lysine (Class I aldolases) or zinc ion (Class II aldolases) to catalyze the cross aldol reaction between two carbonyl compounds (figure 13A). Directed evolution has improved the utility of these enzymes, which typically have stricter requirements on at least one of the aldol partner substrates than many of the enzymes outlined above.[69]

For example, many aldolases require phosphorylated substrates, which are not typically desired by synthetic chemists. Reymond and co-workers removed this requirement by evolving a L-ramulose-1-phosphate (RhaD) aldolase to accept dihydroxyacetone as a substrate, which greatly improves the enzyme's utility (figure 13B).[70] Similarly, 2-deoxyribose-5-phosphate aldolase (DERA), which naturally catalyzes the aldol reaction of acetaldehyde with D-glyceraldehyde 3-phosphate, was evolved to accept nonphosphorylated aldol acceptors.[71] Because this enzyme catalyzes the aldol reaction of two aldehydes, the product can in turn function as a substrate for a second aldol reaction (figure 13C). Of course, not all aldolases require phosphorylated substrates, but expanding the substrate scope of these enzymes to include unnatural

substrates remains highly desirable. For example, *N*-acetyl neuraminic acid lyase (NAL) catalyzes the aldol reaction between pyruvate and *N*-acetyl mannosamine and a variety of additional 5- and 6-carbon sugars. This scope has been extended by directed evolution to generate biologically relevant sialic acid derivatives (figure 13D).[72]

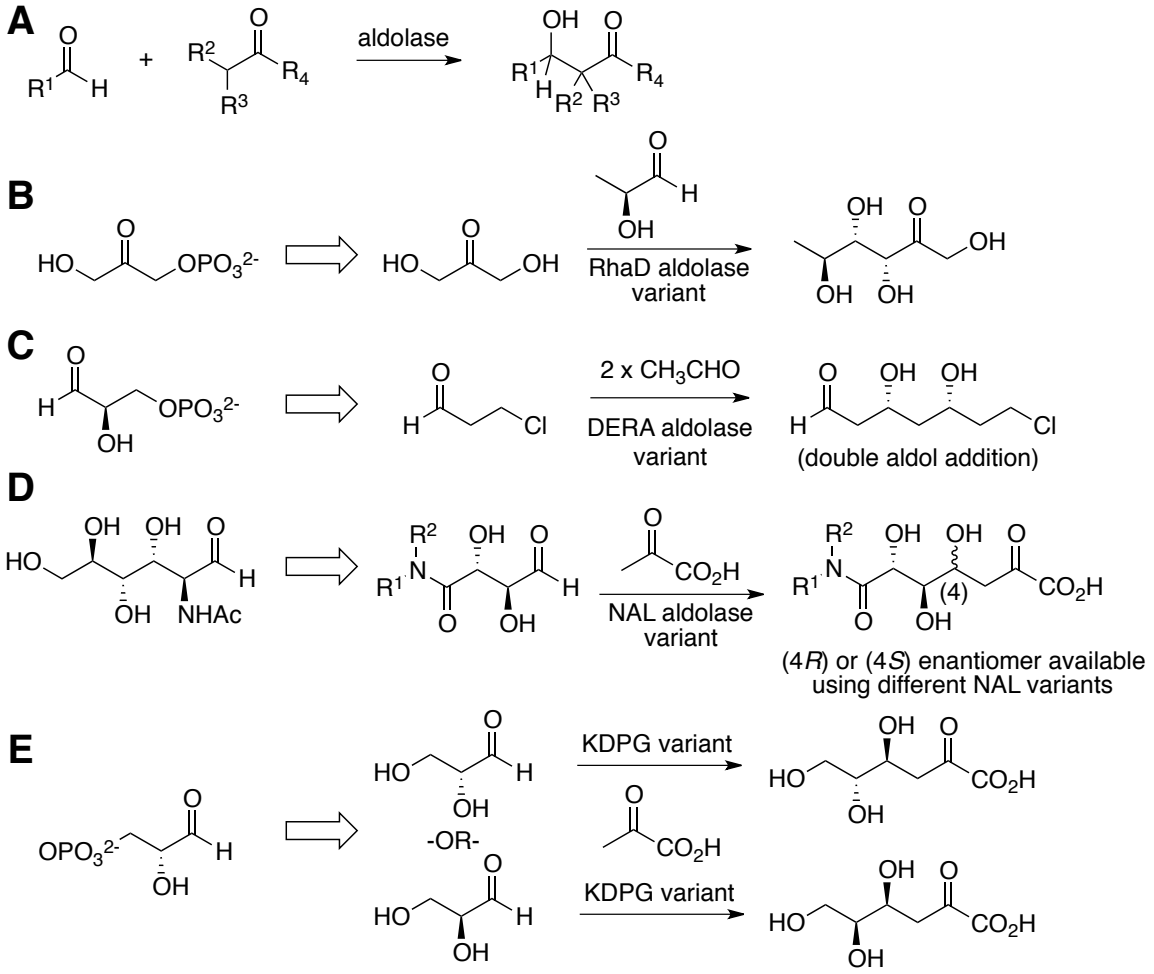


Figure 13. (A) General scheme for aldolase-catalyzed aldol reactions. (B) Native phosphorylated substrates and nonphosphorylated substrates accepted by DHAP and DERA aldolase variants.[70] (C) Native substrate and representative substrate accepted by an NAL aldolase variant. Different variants can be used to generate either the 4*R* or 4*S* product enantiomer.[72-73] (D) Native substrate and representative substrates accepted by KDPG aldolase variants to generate diastereomeric products.[74]

As these examples highlight, aldolases can generate products containing a high degree of stereochemical information. Control of both the enantioselectivity and diastereoselectivity of these reactions is therefore crucial. For example, building upon the NAL-catalyzed reaction outlined above, researchers used directed evolution to generate variants capable of providing either the (4*S*)- or (4*R*)-enantiomer of the aldol product (figure 13D).[73] Control over the diastereoselectivity of the aldol reaction requires that either enantiomer of the aldol acceptor function as a substrate for the aldolase. Indeed, D-2-keto-3-deoxy-6-phosphogluconate aldolase was evolved not only to accept non-phosphorylated substrates, including glyceraldehyde, but also to enable the reaction of either enantiomer of glyceraldehyde with pyruvate to provide diastereomeric products (figure 13E).[74]

Hydroxynitrile lyases catalyze the addition of cyanide to aldehydes to generate synthetically valuable cyanohydrins. These enzymes have been the targets of several engineering efforts, including that of Glieder and co-workers, who were able to significantly improve the activity and enantioselectivity of a hydroxynitrile lyase from *Prunus amygdalus* to (*R*)-pantolactone (figure 14A).[75] Amino acid lyases catalyze nonoxidative, reversible deamination of amino acids and hydroamination of a,b-unsaturated acids in the reverse direction. Altering the substrate scope of these enzymes has enabled the synthesis of unnatural amino acids (figure 14B).[76] Directed evolution has been used to expand the scope of a number of naturally occurring synthase complexes, including those responsible for the biosynthesis of terpenes,[77] polyketides,[78] and nonribosomal peptides.[79]

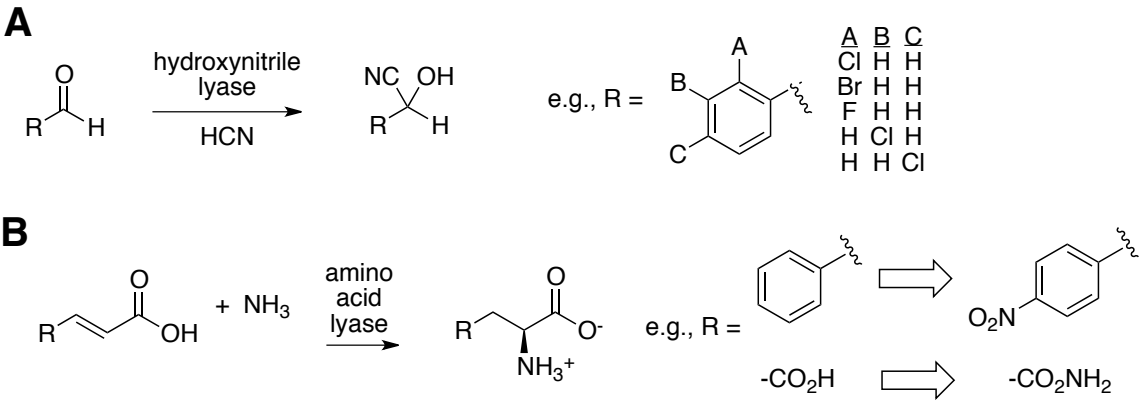


Figure 14. (A) General scheme for hydrocyanation of aldehydes using a hydroxynitrile lyase and representative substrates accepted by hydroxynitrile lyase variants.[75] (B) General scheme for hydroamination of a,b-unsaturated acids using an amino acid lyase and representative substrates accepted by amino acid lyase variants.[76]

3.8. Summary

Given that efforts to alter the properties of enzymes using directed evolution have only been underway for approximately 20 years,[80] and that the total range of enzymatic activities is far greater than those targeted for improvement thus far, much more progress can be expected in the tailoring of enzymes for synthetic applications. However, several general trends that have emerged are worth emphasizing, particularly to synthetic chemists.[81] The first is that many important transformations, including C-H bond functionalization (figure 2),[82] hydroamination (figure 14B),[76] and a wide range of carbon-carbon bond forming reactions, may be carried out in a preparative fashion using enzymes. Many chemists are comfortable with enzymatic hydrolysis reactions, but far greater synthetic power awaits early adopters of systems such as those described above. Second, many of these enzymes exhibit much broader substrate scope than is typically

appreciated.[32] This renders enzymes similar to the "general" catalysts typically employed by chemists. And, unlike those small molecule catalysts, which may be more or less applicable to a given target, enzymes can be systematically improved using directed evolution.

The examples outlined above show that changes in enzyme activity achieved using directed evolution can range from relatively minor, for example, increasing activity toward a known substrate, to surprisingly large, as exemplified by the commercial synthesis of Sitagliptin using an enzyme evolved from a completely inactive parent enzyme.[58] While the extent to which any given enzyme can be tuned appears to vary on a case-to-case basis, it is clear that for many enzymes, very large improvements can be realized. It should also be appreciated that in the course of evolving an enzyme for a given target, many additional enzymes with improved fitness on substrates not directly selected for are also obtained, which greatly expands the pool of catalysts available for synthesis.[32] This also provides starting points for rapid evolution of enzymes for new targets. Finally, while most work in this area has focused on improving enzymes for single reactions either *in vivo* or *in vitro*, the optimization of enzymes of inclusion in metabolic pathways, either natural or synthetic, holds great promise for the synthesis of a wide range of compounds. This will ultimately enable completely novel metabolic pathways to be developed for chemical production inside a cell.

4. Metabolic Engineering

Living organisms create molecular complexity under stringent limitations where cofactors are recycled and multiple catalysts coexist in the same medium without compromising efficiency and selectivity. This ability to catalyze multiple steps *in vivo* without the need to isolate intermediates and change conditions (solvent, temperature, pH, etc.)[83] underscores the tremendous potential of using metabolically engineered cells for efficient chemical synthesis. Natural selection has rarely acted to channel valuable energy and carbon towards production of a single compound. Thus most organisms are suboptimally designed to convert a feedstock (i.e., glucose) to a desired product. Engineering organisms to become high performance living factories is the ultimate goal of metabolic engineering.

Contemporary metabolic engineering utilizes various technologies (enzyme engineering, modeling, network analysis, manipulation of regulation, DNA synthesis, and careful analysis of genes, transcripts, and metabolites) to engineer organisms to synthesize chemical compounds of both native and heterologous origin.[84] This field has attracted intense industrial interest over the last decade, particularly to provide alternate, microbial sources of scarce compounds that are extracted from biological materials and cannot be synthesized economically by purely chemical means. Additionally, there is strong interest in fermentative processes to replace petrochemicals as a strategy to mitigate volatility in the price and supply of petroleum. We illustrate the state of the art by summarizing selected efforts directed at biomanufacturing commodities and specialty

chemicals. The examples were selected based on diversity of methodology and are organized below according to natural product class.

4.1 Isoprenoids

Isoprenoids make up the most numerous and structurally diverse family of natural products, with more than 50,000 different compounds identified to date from all forms of life. Commonly isolated from plants (in which case they are also referred to as terpenoids), microbes, and marine organisms, these valuable compounds are used as flavors, fragrances, nutraceuticals, biodegradable pesticides, as well as potent anticancer and antimalarial drugs. Representative examples include limonene (responsible for the strong smell in oranges and used in many fragrances and artificial flavors), the anticancer drug taxol, the antimalarial agent artemisinin, and lycopene (the red pigment in tomatoes that is used as a nutraceutical). Due to their structural complexity, isoprenoids are difficult, expensive, and inefficient to synthesize chemically and have to be sourced from biological materials. However, these compounds are produced only in small quantities, and extraction and purification is costly as well. The biosynthesis of these complex targets directly from inexpensive carbon sources such as glucose by metabolic engineering in microbial hosts has provided an elegant alternative to chemical synthesis and extraction from natural sources.

All isoprenoids are biosynthesized from two precursors, isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), which undergo isomeric conversion catalyzed by the enzyme IPP isomerase. Elimination of

pyrophosphate from DMAPP catalyzed by isoprene synthase forms isoprene, the simplest terpene (figure 15). On the other hand, the nucleophilic methylene group of IPP can attack the electrophilic allylic CH_2 group of DMAPP to condense in a head-to-tail fashion to form geranyl pyrophosphate (GPP), the common precursor to all monoterpenes ($C_{10}H_{16}$, e.g., limonene). Chain elongation of GPP by addition of another IPP unit forms farnesyl pyrophosphate (FPP), the common precursors to all sesquiterpenes ($C_{15}H_{24}$, e.g., amorphadiene). Addition of yet another IPP unit to FPP gives geranylgeranyl pyrophosphate (GGPP), from which all diterpenes are made ($C_{20}H_{32}$, e.g., taxadiene). The head-to-head condensation of two FPP molecules makes pre-squalene ($C_{30}H_{48}$) which is reduced to give squalene, the precursor to steroids (section 4.2). The tail-to-tail linkage of two GGPP molecules makes the tetraterpene phytoene ($C_{40}H_{64}$, i.e., lycopene), which serves as the precursor to carotenoids. This collection of precursors provides starting points for a wide range of molecules of interest to chemists, some of which are shown in figure 15.

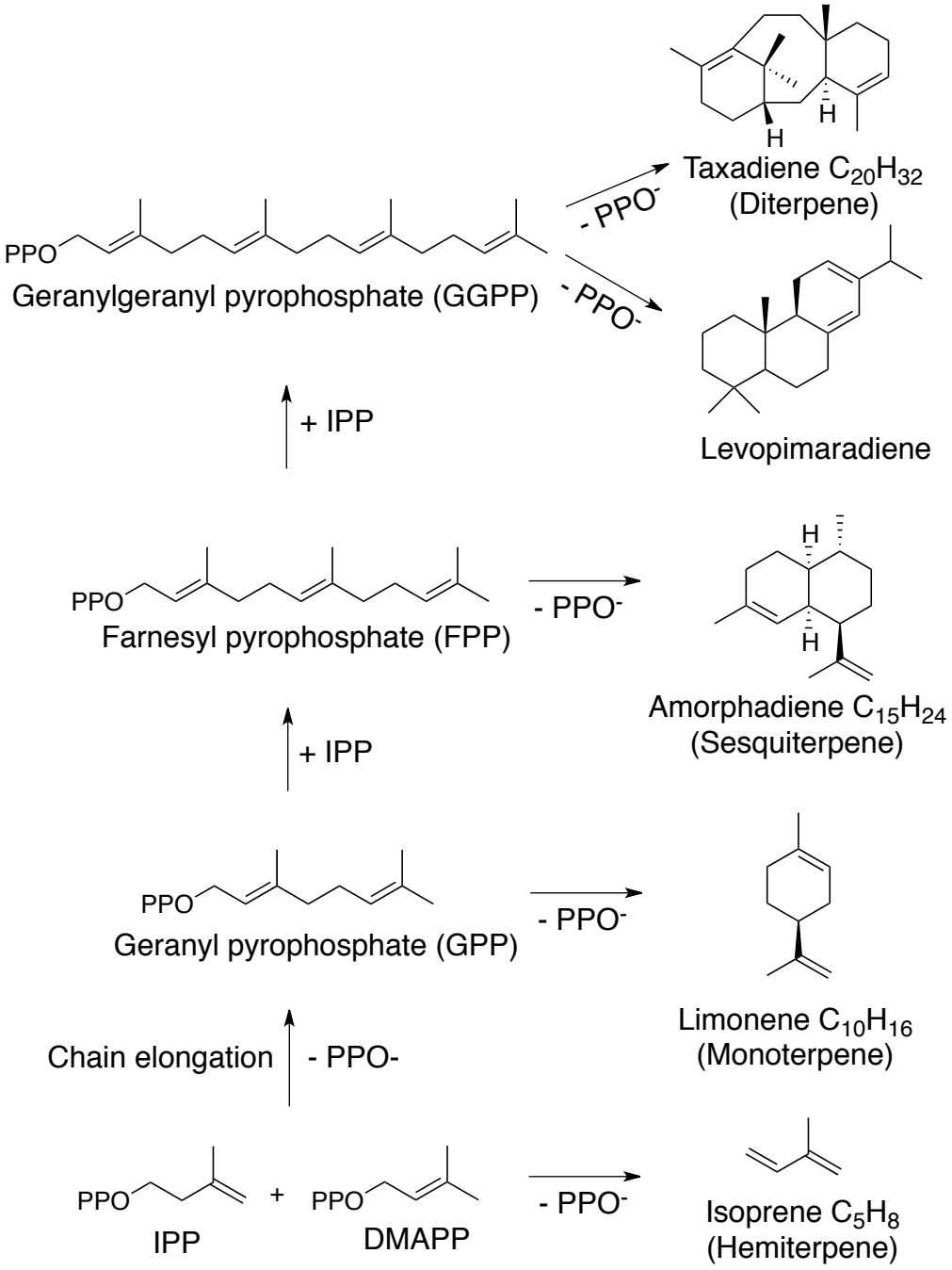


Figure 15. Terpene synthases make isoprenoids from pyrophosphate precursors derived from IPP and DMAPP.[85]

Given the importance of IPP and DMAPP in terpene biosynthesis, understanding the biosynthesis of these precursors is essential for any terpenoid engineering effort. In fact, IPP and DMAPP are produced by two distinct pathways that differ in their cofactor requirements and theoretical yields from glucose. The mevalonate (MVA) pathway converts the ubiquitous intermediate acetyl-coenzyme A (A-CoA) to IPP, whereas the 1-deoxy-D-xylulose-5-phosphate/2-*C*-methyl-D-erythritol-4-phosphate (DXP/MEP) pathway produces IPP and DMAPP from pyruvate and glyceraldehyde-3-phosphate derived from glycolysis. The MVA pathway offers a lower theoretical yield of conversion of glucose to isoprenoids (0.252 g g^{-1}) compared to the DXP/MEP pathway (0.302 g g^{-1}) due to its cofactor imbalance.[86] The choice of which pathway to engineer for maximizing flux toward IPP and DMAPP depends on the regulatory elements present in the host organism. It is important to note that accumulation of prenyl pyrophosphates has been found to be toxic to the host,[87] such that the metabolic engineering experiment has to go one step further and include the terpene synthase that will produce the hydrocarbon product.

4.1.1 Isoprene

Isoprene (2-methyl-1,3-butadiene) is an important commodity chemical used in a wide range of industrial applications, from production of synthetic rubber for tires and coatings to use in adhesives and specialty elastomers. Global industrial production of synthetic isoprene from petrochemical feedstocks is approximately 1 million tons per year.[86] Isoprene is naturally produced by bacteria, animals, and plants, which are

estimated to collectively release approximately 600 million tons per year into the atmosphere. Despite isoprene's pervasiveness in the environment, genetic approaches to identifying the genes encoding isoprene synthesis in prokaryotes have been pursued with limited success.[88] Thus far, only plant-derived isoprene synthases have been well characterized, and sequence data for isoprene synthases exist for only two plant families. Because the catalytic efficiencies of these enzymes are suboptimal for industrial applications (K_M ~ 1–10 mM and k_{cat} ~ 1 s^{-1}),[86, 89] protein engineering for superior kinetic parameters has been pursued by industrial groups.[89-90] As a proof of concept, recombinant production of isoprene in microbial hosts has been achieved by expressing plant isoprene synthases in *E. coli*, *S. cerevisiae*, and photosynthetic bacteria.[91]

Genencor-DuPont and The Goodyear Tire & Rubber Co. are developing a high efficiency fermentative route for polymer-grade isoprene via heterologous expression of the MVA pathway and plant derived isoprene synthases in *E. coli*.[86] The mevalonate pathway was chosen over the DXP/MEP pathway because it was better characterized and has been exploited industrially for production of isoprenoids in both yeast and bacteria.[87,92] The volatile nature of isoprene (b.p. 34 °C) allows gas phase recovery, simplifying purification and eliminating potential feedback inhibition by virtue of product accumulation. This process is reported to have an isoprene yield of 0.11 g g^{-1}, volumetric productivity of 2.0 g L^{-1} h^{-1}, and a titer of 60 g L^{-1}.[86]

4.1.2 Artemisinin

Artemisinin, an endoperoxide sesquiterpene lactone, is an antimalarial drug produced in the leaves of *Artemisia annua*. Its biosynthesis comprises at least four enzymatic steps starting from FPP, the common precursor to sesquiterpenes. The first committed step is the cyclization of FPP to amorpha-4,11-diene, catalyzed by amorphadiene synthase (ADS). The cyctochrome P450 CYP71AV1 then oxidizes amorphadiene to artemisinic aldehyde and also makes artemisinic acid as a by-product. Reduction to dihydroartemisinic aldehyde is catalyzed by artemisinic aldehyde Δ11 reductase, BDR2, and a specific aldehyde dehydrogenase is likely involved in the oxidation step to form hydroartemisinic acid, but the remaining steps in the pathway may be nonenzymatic.[93]

The semisynthesis of artemisinin via metabolic engineering is based on the *in vivo* overproduction of amorphadiene or artemisinic acid and their subsequent conversion to artemisinin through chemical reactions (figure 16). Initial experiments aimed at improving the isoprenoid flux in *E. coli* focused on introducing the heterologous MVA pathway from *S. cerevisiae* to circumvent the regulatory elements found in *E. coli*'s native MEP pathway.[87] The recombinant eight-gene biosynthetic pathway was divided into two operons: a "top" operon (MevT) that encodes the three enzymes required to convert the ubiquitous precursor acetyl-CoA to (*R*)-mevalonate and a "bottom" operon (MBIS) that converts (*R*)-mevalonate to FPP (figure 16). Coexpression of these operons with a synthetic ADS gene, optimized for expression in *E. coli*, resulted in the production

of amorphadiene from glucose and glycerol with a final concentration of 0.5 g L^{-1} following optimization of process conditions.[87, 94]

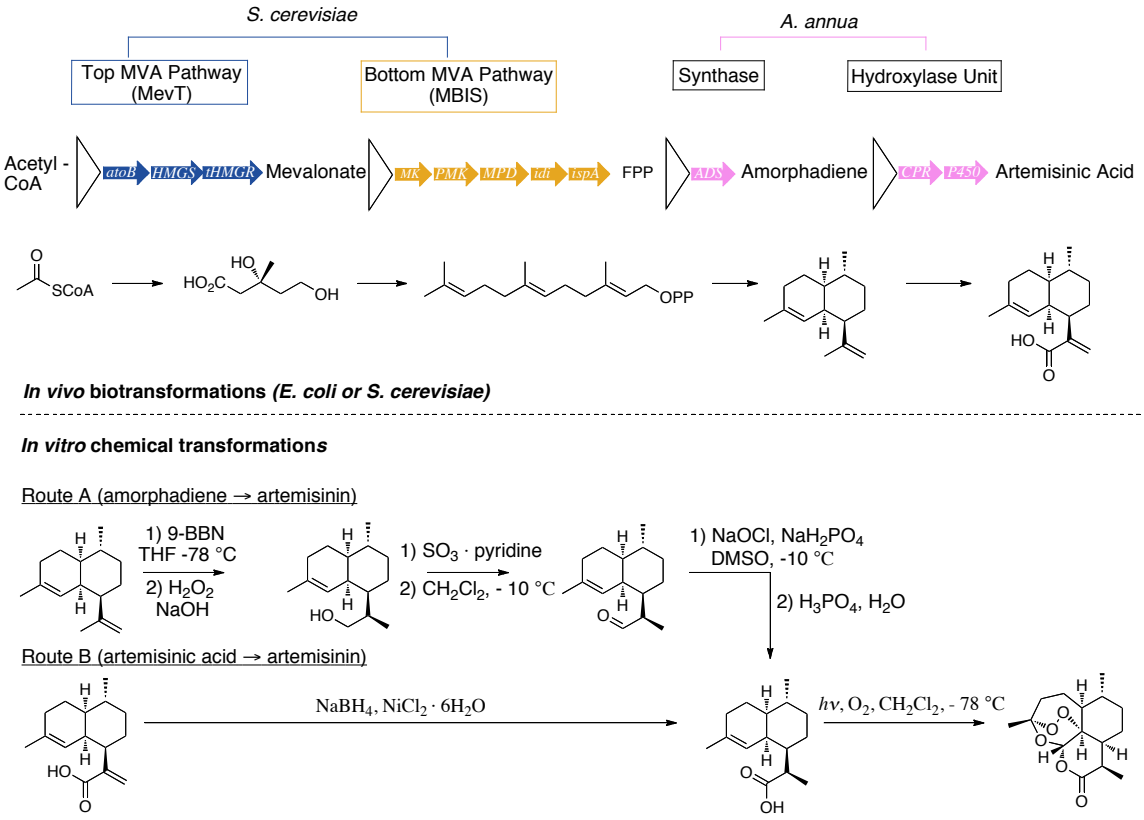


Figure 16. Semisynthetic scheme for artemisinin production. Enzymes encoded by the genes shown are: *atoB*, acetoacetyl-CoA thiolase; *HMGS*, HMG-CoA synthase; *tHMGR*, truncated HMG-CoA reductase (expressing only the C-terminal catalytic domain); *MK*, mevalonate kinase; *PMK*, phosphomevalonate kinase; *MPD*, mevalonate diphosphate decarboxylase; *idi*, IPP isomerase; *ispA*, farnesyl diphosphate synthase; *ADS*, amorphadiene synthase; *CPR*, cytochrome P450 reductase; *P450*, CYP71AV1.[95]

Mevalonate supplementation studies suggested that the production of amorphadiene in this system was limited by the conversion of acetyl-CoA to mevalonate.[96] However, increased expression of the three enzymes in the top operon (MevT) — by increasing plasmid copy number and by replacing the promoter — resulted in cell growth inhibition. Gene titration studies and metabolite profiling using LCMS

linked growth inhibition with the intracellular accumulation of the heterologous metabolite 3-hydroxy-3-methyl-glutaryl-coenzyme A (HMG-CoA), suggesting that low activity of a single enzyme, HMG-CoA reductase (*tHMGR*), encoded in MevT, was the major bottleneck. This flux imbalance was alleviated by coexpressing the MevT operon with an additional medium-copy-number plasmid carrying an extra copy of tHMGR, effectively increasing the concentration of the catalyst in the bottleneck step and thereby achieving a 3-fold increase in mevalonate production. Further optimizations included replacing the yeast genes for HMG-CoA synthase and HMG-CoA reductase with equivalent genes from *Staphylococcus aureus*, and improving the fermentation process under the restriction of carbon and nitrogen, so as to maximize flux toward the desired metabolic pathway, yielding commercially relevant titers of 27.4 g L^{-1} amorphadiene.[97]

The membrane-associated cytochrome P450 monooxygenase from *A. annua* (CYP71AV1) catalyzes the three-step oxidation of amorphadiene to artemisinic acid, producing the alcohol and aldehyde congeners of artemisinic acid as intermediates. To achieve the *in vivo* oxidation of amorphadiene to artemisinic acid, CYP71AV1 and its redox partner cytochrome P450 reductase (CPR), were codon-optimized for *E. coli* and coexpressed in the aforementioned strains. By engineering the *N*-terminal transmembrane sequence and using an expression plasmid suitable for P450 expression, artemisinic acid titers of 105 mg L^{-1} were achieved, when conducting the fermentation in the absence of a *n*-dodecane overlay.[98] Conversion of artemisinic acid to the drug aremisinin can be accomplished in just two *in vitro* chemical reactions, namely reduction to hydroartemisinic acid followed by a photooxidation cyclization reaction (figure 16).[99]

Biosynthesis of artemisinic acid has also been achieved in the yeast *S. cerevisiae,* by up-regulating the expression of several genes from the native MVA pathway and by heterologous expression, through chromosomal integration, of amorphadiene synthase (ADS), amorphadiene monooxygenase (CYP71AV1) and the reductase (CPR).[100] Since steroid biosynthesis in yeast competes with ADS for FPP availability, it was necessary to down-regulate expression of squalene synthase, which catalyzes the first step downstream of FPP in the sterol pathway. Just as for *E. coli*, integration of an additional copy of tHMGR into the chromosome increased amorphadiene production, altogether producing >100 mg L^{-1} artemisinic acid. Interestingly, artemisinic acid was found to accumulate on the outer side of the cell membrane, allowing for a one-step purification following its "total" synthesis in yeast. Artemisinic acid removed from the cell pellet by washing with alkaline buffer was extracted into ether and purified by silica gel chromatography to achieve more than 95% purity.[100] Further strain engineering could not improve titers beyond 0.2 g L^{-1} for glucose-fed fermentations.[95b] However, strains that had been cloned with the pathway only up to amorphadiene achieved titers of more than 1.2 g L^{-1}. Subsequent improvements in the fermentation process for amorphadiene were able to significantly increase product titers up to the commercially relevant target of 37 g L^{-1}.[95b] Key developments included restricting phosphate in fed-batch fermentations—this is thought to limit growth and channel greater carbon flux to product—and replacing the restricted glucose feed with a restricted ethanol feed, which is thought to increase intracellular acetyl-CoA levels for the MVA pathway. Deletion of *GAL1 GAL10 GAL7* and Gal80p, a

gene cluster required for the utilization of galactose, altogether eliminated the need to add expensive galactose for induction of amorphadiene production. The *S. cerevisiae*–based process has considerable advantages over the aforementioned *E. coli* process, most notably the absence of an expensive inducer such as IPTG and the lower oxygen uptake rate required for the restricted ethanol process, which reduces costs associated with aerating large fermentors.

The more facile biosynthesis of amorphadiene compared to artemisinic acid led to an alternative semisynthetic strategy where yeast fermentation converts sugars to amorphadiene and *in vitro* chemical reactions subsequently convert amorphadiene to dihydroartemisinic acid en route to artemisinin (figure 16).[95b] Industrial production of semisynthetic artemisinin is being pursued by Amyris Biotechnologies in collaboration with Sanofi-Aventis.

4.1.3 Taxol

Taxol (paclitaxel) and its structural analogs are among the most potent and commercially successful anticancer drugs. Taxol comprises a highly oxygenated 6-8-6 tricyclic ring system, featuring a total of 11 stereogenic centers, an ester side chain, and a distinctive oxetane ring. Six independent total syntheses have been reported involving elaborate routes that require 37 to 51 steps, with yields too low for practical synthesis (highest yield of 0.4%).[83] Taxol was first isolated from the bark of *Taxus brevifolia* (Pacific Yew tree), and its biosynthetic pathway consists of approximately 20 enzymatic steps from the primary intermediate geranylgeranyl diphosphate (GGPP).[93] Taxadiene

synthase catalyzes the cyclization of GGPP to taxa-4(5),11(12)-diene (figure 15), the first committed step of the pathway, with remarkable fidelity generating the 6-8-6 tricycle with the correct chirality for the first three stereogenic centers. The diterpenoid core is further decorated by 8 cytochrome P450-mediated oxygenations, four coenzyme A-dependent acylations, and a C13 side chain assembly and attachment en route to taxol.

Metabolic engineering has been applied successfully to overproduce taxol precursors in *E. coli*.[101] Since the MEP pathway for making IPP and DMAPP (figure 17, upstream module) is endogenous to *E. coli*, heterologous expression of only GGPP synthase and taxadiene synthase (figure 17, downstream module) allows for the production of taxa-4(5), 11(12)-diene. The pathway was partitioned into two modules separated at IPP, the key intermediate in terpenoid biosynthesis. A total of 32 strains were constructed wherein 8 native enzymes identified as bottlenecks in the upstream MEP pathway were expressed in one module and the two heterologous downstream enzymes were expressed in the other module. By systematically varying expression of each module, a nonlinear effect on taxadiene titers was observed for the relative dosage of the two pathways. Further optimizations such as chromosomal integration of the upstream operon, minimizing pathway inhibition by indole (a by-product), and a two-phase fermentation to avoid air-stripping of taxadiene resulted in titers as high as 1.02 g L^{-1}.[101] Finally, the plant-derived cytochrome P450, taxadiene 5α hydroxylase, an unusual monooxygenase that catalyzes hydroxylation along with double bond migration, was included in the pathway to generate taxadien-5α-ol (though this resulted in a much lower product titer of 50 mg L^{-1}).

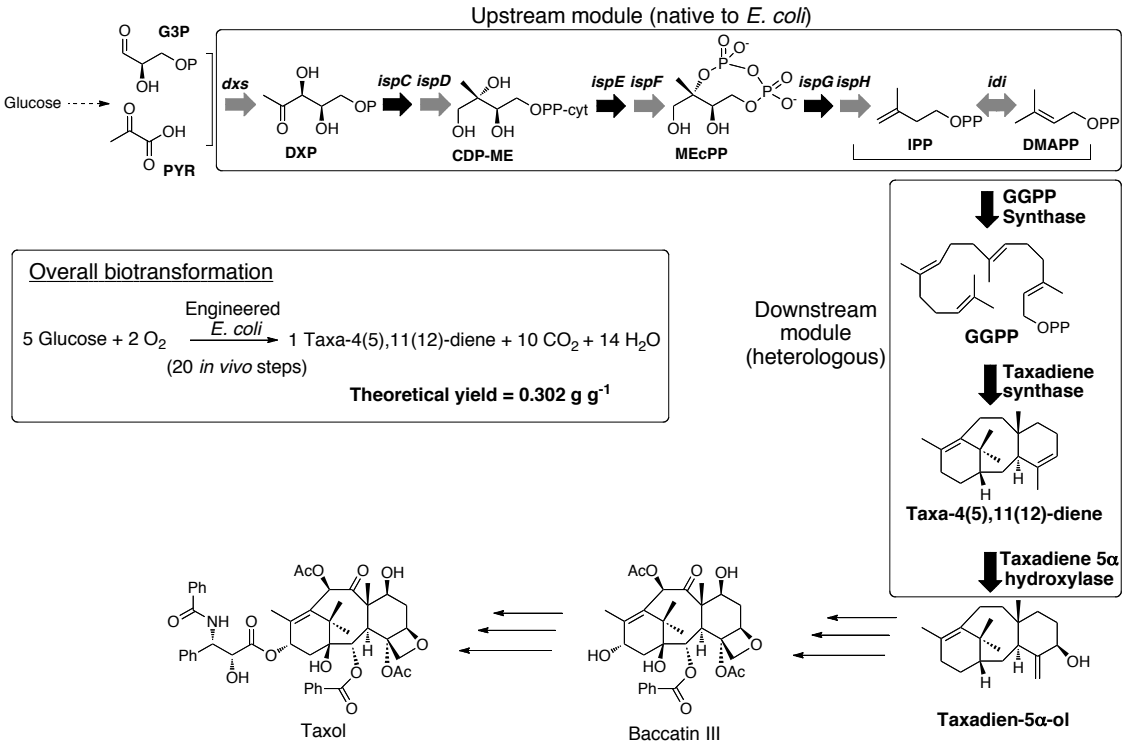


Figure 17. Biosynthesis of taxol precursors in *E. coli* via the DXP/MEP pathway (upstream module). Enzymes encoded by the genes shown are *dxs*, 1-deoxy-D-xylulose-5-phosphate synthase; *ispC*, 1-deoxy-D-xylulose-5-phosphate reductoisomerase; *ispD*, 4-diphosphocytidyl-2-*C*-methyl-D-erythritol-2,4-cyclodiphosphate synthase; *ispE*, 4-diphosphocytidyl-2-*C*-methyl- D-erythritol kinase; *ispF*, 2-*C*-Methyl-D-erythritol-2,4-cyclodiphosphate synthase; *ispG*, 1-hydroxy-2-methyl-2-(*E*)-butenyl-4- diphosphate synthase; *ispH*, 4-hydroxy-3-methyl-2-(*E*)-butenyl-4-diphosphate reductase; *idi*, IPP isomerase.[101]

4.1.4 Levopimaradiene

Levopimaradiene is the diterpenoid gateway precursor of the pharmaceutically important plant-derived ginkgolides. Heterologous production has been achieved in *E. coli* with a maximum titer of 700 mg L^{-1} using a combination of metabolic and protein engineering methods.[102] Introduction of plant-derived geranylgeranyl diphosphate synthase (GGPPS) and levopimaradiene synthase (LPS) codon-optimized for *E. coli* led to only marginal production of levopimaradiene (0.15 mg L^{-1}). Overexpression of the

upstream MEP/DXP pathway for increasing the intracellular concentration of the isoprenoid precursors IPP and DMAPP led to a titer of 92 mg L^{-1}. However, further product amplification could not be achieved by continuing to increase the concentration of the bottleneck enzymes in the MEP/DXP pathway. This suggested that the inherently low enzyme activity and specificity of the downstream steps had become rate limiting under high isoprenoid precursor flux. Thus, enzyme engineering at the key metabolic nodes was pursued in order to achieve substantial overproduction of the target diterpene.

LPS from *Ginkgo biloba* converts GGPP to levopimaradiene with about 87% selectivity, while also forming the isomers abietadiene (11%), sandaracopimaradiene (2%), and neoabietadiene (trace) (figure 18).[102] This product mixture results from initial isomerization of GGPP in the first active site of LPS to form the intermediate copalyl diphosphate, which subsequently undergoes a poorly selective diphosphate-ionization cyclization in the second active site of LPS. This example illustrates how the poor selectivity of some terpene synthases can result in unwanted product formation, which leads to low metabolic fluxes and large product losses.

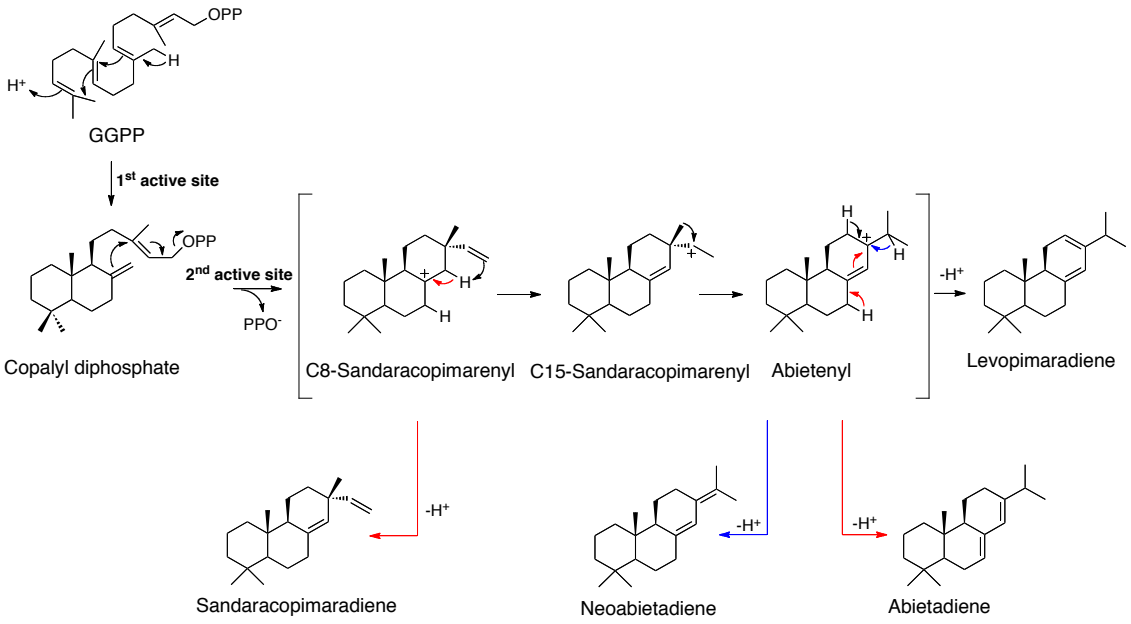


Figure 18. Levopimaradiene synthase (LPS) catalyzes the cyclization of geranylgeranyl diphosphate (GGPP) to levopimaradiene in two distinct active sites. In the first active site, deprotonation drives the formation of (+)-copalyl diphosphate (CPP). Reaction in a second active site leads to the formation of multiple products from GGPP with poor selectivity for levopimaradiene. CPP undergoes diphosphate-ionization cyclization to form the C8-sandaracopimarenyl cation, which can be deprotonated to release sandaracopimaradiene or undergo intramolecular proton transfer and 1,2-methyl migration to yield the abietenyl cation. Subsequent deprotonation of abietenyl cation at four possible sites produces abietadiene, levopimaradiene, neoabietadiene, and palustradiene (not shown).[102]

E. coli overexpressing the genes for the MEP/DXP pathway and expressing wild-type GGPPS was used to screen a library of LPS variants with mutations focused in the troublesome second active site. Mutations were selected to increase diterpene productivity and retain or improve the selectivity for levopimaradiene. Saturation mutagenesis at key residues and recombination of the top mutations improved productivity by 10-fold without significant changes in product distribution. Finally, the activity of GGPPS was improved 1.7-fold via random mutagenesis and an *in vivo* screen that utilized a lycopene biosynthetic pathway as a colorimetric reporter.[102] This example

illustrates the utility of protein engineering in manipulating secondary metabolic pathways with poor selectivity to optimize microbial overproduction of desired chemicals.

4.2 Steroids

Sterols are essential structural and regulatory components of eukaryotic cell membranes. Thus far, only a couple of members of this important class of natural products have been produced by engineered microbial hosts, most notably ergosterol and hydrocortisone. Ergosterol is the main sterol in yeast and is responsible for structural membrane features such as fluidity and permeability, similar to cholesterol in mammalian cells. Additionally, ergosterol is a precursor to vitamin D_2 and cortisone. Overproduction of ergosterol has been achieved in *S. cerevisiae* by overexpressing endogenous genes that are bottlenecks in the native ergosterol biosynthetic pathway (figure 19).[103] Transformation of squalene, the polyisoprene precursor, to ergosterol requires the action of 12 different enzymes. Deregulated expression of HMG-CoA reductase leads to accumulation of squalene (due to higher isoprenoid flux via the MVA pathway) but not of ergosterol, suggesting that the steps downstream of squalene become rate limiting. Simultaneous overexpression of two genes along the *ERG* pathway led to a significant decrease of the squalene pool and concomitant increase in the amount of ergosterol.[103] With optimization in fermentation conditions the final ergosterol titer has been reported to reach 1.98 g L^{-1}.[104]

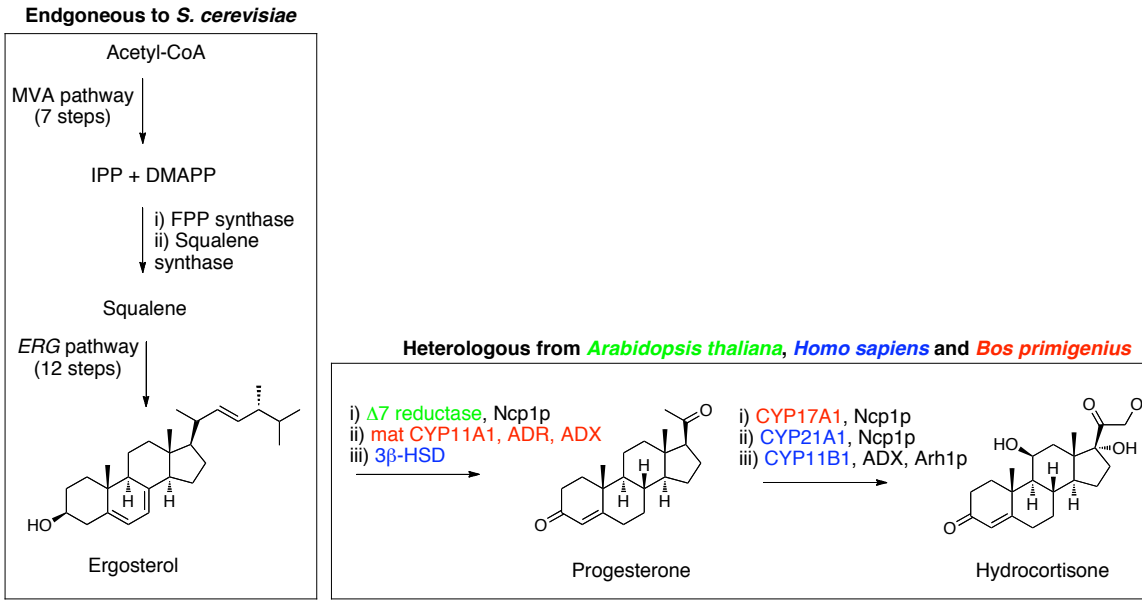


Figure 19. Endogenous biosynthesis of ergosterol in yeast and recombinant production of hydrocortisone involving mammalian and plant enzymes. Enzymes: matCYP11A1, mature form of CYP11A1 (P450 side chain cleaving); 3β-HSD, 3β-hydroxysteroid dehydrogenase-isomerase; ADX, bovine adrenodoxin (electron carrier); ADR, bovine adrenodoxin reductase; Δ7 reductase, Δ7 sterol reductase; Ncp1p, NADPH P450 reductase; Arh1p, adrenodoxin reductase homolog.[105]

Total biosynthesis of heterologous steroids in genetically programmed microorganisms has been pursued with limited success, with only modest titers reported. Notable advances include the overproduction of progesterone and hydrocortisone in yeast.[105-106] Hydrocortisone is the major steroid in mammals and an important intermediate of steroidal drug synthesis. Total biosynthesis of hydrocortisone in yeast involves 25 steps from acetyl-CoA (figure 19), of which the first 19 are endogenous to *S. cerevisiae*. Conversion of ergosterol to hydrocortisone requires the introduction of one plant enzyme and five additional enzymatic steps catalyzed by eight mammalian proteins (figure 19).[105] The heterologous genes were introduced by a combination of plasmid and genomic integration techniques. Elimination of competing pathways and balanced

expression of the key enzymes resulted in the production of hydrocortisone as the major sterol (70% selectivity), albeit at low titers (<20 mg L^{-1}), reflecting the challenges in transferring a highly complex and mainly membrane-bound mammalian pathway into a microbial host.

4.3 Amino Acids

Amino acids are major industrial products of microbial fermentations, with a global market of over three million tons per year and applications ranging from animal feed to food and pharmaceutical products. The microbial hosts used in industry typically do not require foreign DNA for overproduction, and engineering efforts are instead directed toward bypassing native regulation mechanisms (i.e., feedback inhibition) that limit titer and productivity. Metabolic engineering applied to amino acid fermentation has been reviewed elsewhere.[107] Here we limit discussion to a summary of efforts to produce L-lysine and L-alanine in bacteria to illustrate approaches of genomic reconstruction and metabolic evolution.

More than 1,000,000 tons of L-lysine per year are produced from the Gram-positive bacterium *Corynebacterium glutamicum.*[108] Industrial strains of *C. glutamicum* have been developed by classical methods of whole-cell mutagenesis and selection. This technique generated many industrial fermentation strains, but has the serious disadvantage of accumulating uncharacterized secondary mutations that are detrimental to important properties such as growth, sugar consumption, and stress tolerance. For instance, *C. glutamicum* B-6 was constructed from the wild-type strain (ATCC 13032) by

iterations of random mutation and selection, and is able to produce L-lysine to high titers (95 g L^{-1}) but with suboptimal rate (1.9 g L^{-1} h^{-1}), due to poor growth and sugar consumption.[109] It had been known for decades that the wild-type strain cannot accumulate L-lysine unless it is deficient in L-homoserine dehydrogenase, since this enzyme will divert flux toward the production of L-threonine, L-methionine, and L-isoleucine.[110] Comparative genomics of B-6 and the wild-type strain along the 16 proteins involved in the terminal pathway and the L-lysine efflux step revealed that only three enzymes had suffered amino acid mutations, with each differing from the wild-type copy by a single amino acid substitution.[109] Recombining the three evolved genes in the wild-type background leads to faster production of L-lysine (3.0 g L^{-1} h^{-1}) than in B-6, while retaining a high titer (80 g L^{-1}). This example of genome-based strain reconstruction shows how dissecting the genetic modifications responsible for phenotypic selection at the macroscopic level can lead to an industrially useful, genetically defined mutant without any foreign DNA.

In another effort to understand microbial metabolism at a systems biology level, in-depth profiling of intracellular metabolite concentrations, metabolic fluxes, and gene expression of lysine-producing *C. glutamicum* revealed that the phase shift from growth to lysine production is accompanied by a decrease in glucose uptake flux and the redirection of flux from the tricarboxylic acid (TCA) cycle toward lysine biosynthesis.[111] Additionally, metabolic engineering created an L-lysine hyper-producing *C. glutamicum* with impressive fermentation metrics: 120 g L^{-1} titer, 4.0 g L^{-1} h^{-1} volumetric productivity and 0.55 g g^{-1} yield on glucose.[112] The engineered strain contained only 12

mutations, which were integrated into the genome without introducing foreign DNA and without requiring selection markers.

Production of L-alanine has been achieved in *E. coli* with industrially relevant parameters.[113] An *E. coli* strain that makes L-alanine as the major product was created by replacing the genomic lactic acid dehydrogenase gene in a strain already optimized for lactic acid fermentation with a heterologous alanine dehydrogenase (figure 20). Alanine dehydrogenase produces alanine from pyruvate and ammonia using one equivalent of NADH (i.e., the biological equivalent of reductive amination, see section 3.4). The engineered pathway is redox-balanced such that NADH oxidation (required for alanine biosynthesis) is coupled to ATP production and cell growth (figure 20). This linkage provides a basis for metabolic evolution (i.e., serial or continuous cultivation) where selection for improvements in growth coselect for increased glycolytic flux and alanine production. Cells with increased growth rate due to spontaneous mutations will outcompete their parent strains and have increased alanine productivity. Deletion of other genes that divert metabolic flux away from the target pathway, optimizing fermentation conditions, and successive rounds of metabolic evolution achieved a final product titer of 114 g L^{-1}, 96% mass yield (with respect to glucose), 2.4 g L^{-1} h^{-1} average volumetric productivity, and product purity greater than 99.5%. In this example, chromosomal integration of a single heterologous gene provided sufficient levels of expression to support high levels of growth and product flux without the need for multicopy plasmids and selection agents for plasmid retention (thus obviating the need for expensive inducers and antibiotics). Impressively, the dried broth from batch fermentation in glucose mineral

salts medium afforded, after cell removal, approximately 95% (w/w) L-alanine without further purification.

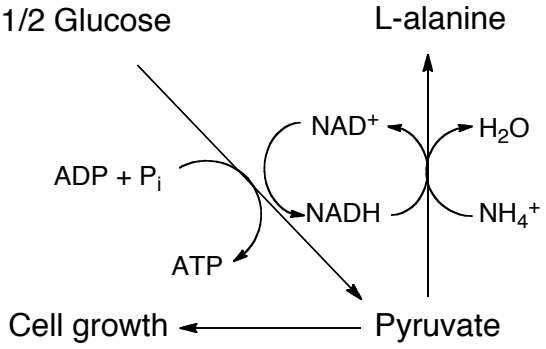


Figure 20. Fermentation of glucose to L-alanine generates a surplus of ATP allowing the pathway to be optimized by metabolic evolution.[113]

4.4 Diamines

Diamines have many industrial applications as components of polyamides, pharmaceuticals, agrochemicals, surfactants, and other materials. For example, putrescine (1,4-diaminobutane) is currently used to synthesize nylon-4,6 by polycondensation with adipic acid. Putrescine is naturally biosynthesized in *E. coli* and is thought to play a role in cell proliferation and normal cell growth; however, its intracellular concentration in *E. coli* is reported to be only ~30 mM, as it is elaborately regulated by biosynthesis, degradation, uptake and excretion.[114] Accordingly, overproduction of putrescine in *E. coli* requires the concerted overexpression of bottleneck enzymes, deletion of competing degradation pathways, enhancing excretion, and hindering uptake back into cells. Amplification of putrescine biosynthesis in *E. coli* is based on seven enzymatic steps from the TCA cycle intermediate α-ketoglutarate (figure 21).[115] The key step is decarboxylation of ornithine, a common precursor in arginine biosynthesis, to directly

form putrescine. The putrescine importer (*puuP*) was deleted to prevent the cells from importing extracellular putrescine at the end of batch fermentations under conditions of carbon starvation. Arginine biosynthesis from ornithine requires the action of one of two ornithine carbamoyltransferases (*argI* and *argF*). The *argI argF* double knock-out is auxotrophic with respect to arginine, but arginine supplementation is undesirable since the pathway is feedback-inhibited by arginine at the level of *N*-acetylglutamate synthase. Deletion of just the *argI* gene, encoding the major carbamoyltransferase, results in increased production of putrescine and eliminates the requirement for arginine supplementation. Constitutive overexpression of the antiporter *potE*, responsible for putrescine excretion, was achieved by replacing its native promoter by a stronger promoter. Finally, high cell density fed-batch fermentations in glucose mineral salts medium allowed for a putrescine accumulation of 24.2 g L^{-1} at 0.75 g L^{-1} h^{-1}, with 0.168 g g^{-1} yield on glucose.[115]

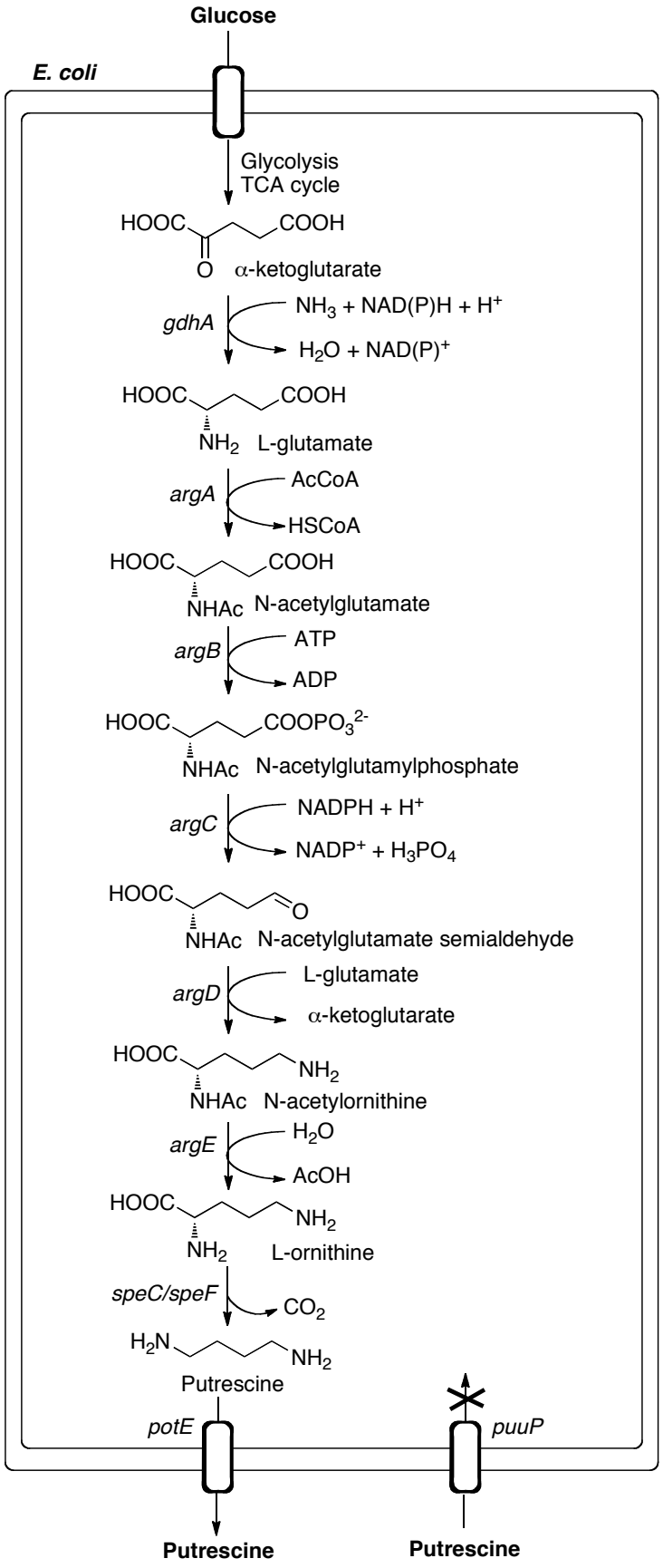

Glucose
E. coli
Glycolysis
TCA cycle
HOOC
COOH
O
α-ketoglutarate
gdhA
NH3 + NAD(P)H + H+
H2O + NAD(P)+
NH2
L-glutamate
argA
AcCoA
HSCoA
NHAc
N-acetylglutamate
argB
ATP
ADP
COOPO3 2-
N-acetylglutamylphosphate
argC
NADPH + H+
NADP+ + H3PO4
N-acetylglutamate semialdehyde
argD
L-glutamate
α-ketoglutarate
N-acetylornithine
argE
H2O
AcOH
L-ornithine
speC/speF
CO2
H2N
Putrescine
potE
puuP
Putrescine
Putrescine

Figure 21. Putrescine biosynthesis in *E. coli*. Enzymes encoded by the genes shown are *gdhA*, glutamate dehydrogenase; *argA*, N-acetylglutamate synthase; *argB*, N-acetylglutamate kinase; *argC*, *N*-acetylglutamyl phosphate reductase; *argD*, acetylornithine transaminase; *argE*, acetylornithine deacetylase; *speC/speF*, biosynthetic/degradative ornithine decarboxylase; *potE* putrescine/ornithine antiporter; *puuP*, putrescine importer. TCA cycle, tricarboxylic acid cycle. Overall biotransformation: 1.03 glucose + 2 NH_3 → putrescine + 2.18 CO_2 + NADH.[115]

Unlike putrescine, cadaverine is a minor polyamine found only in a few microorganisms.[116] In *E. coli*, cadaverine (1,5-pentanediamine) is biosynthesized from lysine in a single enzymatic step by the action of lysine decarboxylase. Accordingly, previously derived lysine-producing strains of *C. glutamicum* offer an excellent starting point for overproduction of this diamine. Toray Industries demonstrated that cadaverine accumulation can be achieved by chromosomal substitution of the native homoserine dehydrogenase gene (*hom*) in *C. glutamicum* for the lysine decarboxylase from *E. coli* (*cadA*).[117] Homoserine dehydrogenase deficiency is required for lysine accumulation (section 4.3) and thus is also important for improving flux toward cadaverine. Simultaneous *hom* removal and *cadA* expression resulted in a titer of 2.6 g L^{-1} cadaverine, with 0.052 g g^{-1} yield from glucose. Efficient secretion of cadaverine is important as its *in vivo* accumulation can inhibit lysine decarboxylase. Most cadaverine formed inside *C. glutamicum* is ionized (p*K*a = 9.13) such that its secretion requires active transport across the cell membrane. Genome-wide transcription analysis of cadaverine-producing *C. glutamicum* was used to identify the key exporter, a permease with an undefined physiological role, which when overexpressed resulted in improved yields of cadaverine (0.136 g g^{-1} on glucose).[118] *E. coli* has also been engineered to overproduce cadaverine by overexpressing *cadA* and dihydrodipicolinate synthetase.[119] Deletion of genes

involved in the utilization, degradation, and import of both putrascine and cadaverine was also required. Altogether, the engineered *E. coli* fermented sugars in mineral salts to a final cadaverine concentration of 9.61 g L^{-1}, at a volumetric productivity of 0.32 g L^{-1} h^{-1}, and a yield of 0.12 g g^{-1} on glucose.[119]

4.5 Alcohols

Alcoholic fermentation in *S. cerevisiae* is the largest scale biotechnology process in operation (>70 million tons per year). Strains of *S. cerevisiae* have been developed over millennia of artificial selection for ethanol production, which is the end product of anaerobic fermentation in yeast. Intense academic and industrial research has been directed toward novel fermentation routes for higher chain alcohols with superior fuel properties and more versatile chemical applications. Isobutanol (2-methylpropan-1-ol), for example, is a precursor to C4 petrochemical building blocks as well as an excellent gasoline blend. Glucose to isobutanol conversion has been demonstrated in *E. coli* via a modified amino acid pathway that requires only two nonnative steps (KDC and ADH) and uses keto-isovalerate as a precursor (figure 22).[120] Isobutanol titers of >20 g L^{-1} have been reached by using engineered strains with deletions in pathways competing for carbon and using cofactors that are available aerobically or microaerobically.[120] Anaerobic conditions are preferred, however, for lower operating costs, higher yields, and compatibility with existing ethanol plants. An anaerobic pathway that operates at nearly theoretical yield was achieved by engineering the KARI to switch cofactor

preference (i.e., to use NADH instead of NADPH) and thereby consume NADH produced during glycolysis (figure 22).[121]

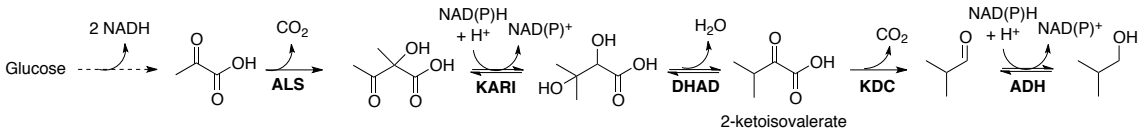


Figure 22. Glucose to isobutanol conversion via 2-ketoisovalerate. Enzymes: ALS, acetolactate synthase; KARI, ketol-acid reducto-isomerase; DHAD, dihydroxy-acid dehydratase; KDC, 2-ketoacid decarboxylase; ADH, alcohol dehydrogenase.[120]

4.6 Diols

Propane-1,3-diol (PDO) is used in fiber polymers prepared from terephthalic acid and PDO that are produced in large volumes (e.g., for apparel and carpeting). PDO is naturally produced by a large number of bacteria, such as *Citrobacter*, *Clostridium*, *Enterobacter*, *Klebsiella*, and *Lactobacillus*, as the result of anaerobic growth on glycerol.[122] In these systems, glycerol is converted to PDO in just two steps, namely dehydration to 3-hydroxypropionaldehyde (3-HPA) followed by NAD(P)H-dependent reduction to PDO (figure 23). The first reaction, which is rate limiting, proceeds via a free radical H-abstraction/rearrangement mechanism followed by spontaneous dehydration and is catalyzed by the coenzyme B_{12}-dependent glycerol dehydratase. This enzyme is eventually inactivated by a suicide reaction with glycerol, such that an ATP-dependent auxilliary reactivase (a chaperone-like protein) is required to reestablish dehydratase activity. The high cost of supplementing vitamin B_{12} to the growth medium is an additional drawback of relying on a B_{12}-dependent enzyme in a metabolic pathway to PDO. Dehydratases with higher resistance to inactivation by glycerol and inhibition by PDO were identified in a metagenomics study, as described in section 2.2.[13]

Alternatively, B_{12}-independent glycerol dehydratases have been described and may find application in PDO bioproduction.[123]

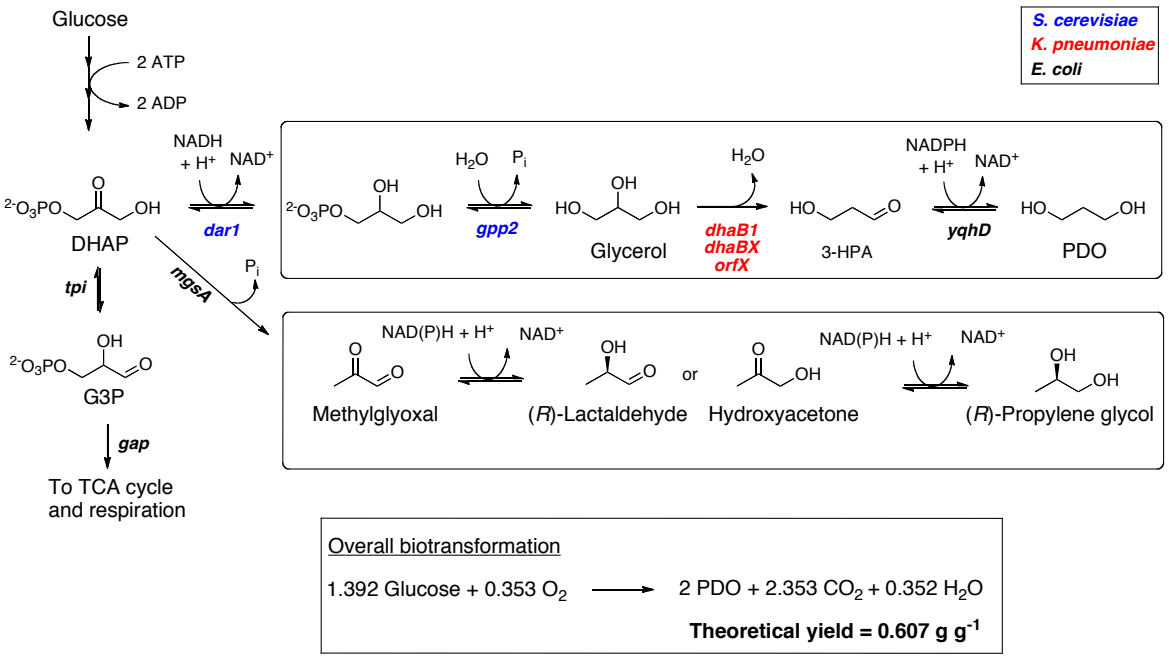


Figure 23. Recombinant pathways for the production of propanediols in *E. coli*. Enzymes encoded by the genes shown are *dar1*, glycerol 3-phosphate dehydrogenase (*S. cerevisiae*); *gpp2*, glycerol 3-phosphate phosphatase (*S. cerevisiae*); *dhaB1*, glycerol dehydratase (*K. pneumoniae*); *dhaBX*, *orfX*, reactivating factors for glycerol dehydratase (*K. pneumoniae*); *yqhD*, oxidoreductase (*E. coli*); *mgsA*, methylglyoxal synthase (*E. coli*); *tpi*, triosephosphate isomerase (*E. coli*); *gap*, glyceraldehyde 3-phosphate dehydrogenase (*E. coli*). Calculations assumed that cellular respiration of 1 mol of glucose yields 12 mol NAD(P)H equivalents and 34 mol ATP.[124]

No naturally occurring microbe has been reported to directly ferment glucose to PDO. In a landmark success of industrial metabolic engineering, Genencor-DuPont engineered *E. coli* to directly produce high levels of PDO from glucose.[122] The heterologous pathway imports two genes from *S. cerevisiae* (DAR1 and GPP2, figure 23) to convert the glycolytic intermediate dihydroxyacetone phosphate (DHAP) to glycerol and the *K. pneumoniae* glycerol dehydratase (*dhaB1*) and its reactivating factors (*dhaBX*,

orfX) that subsequently convert glycerol to 3-HPA. Reduction of 3-HPA to PDO is catalyzed by an oxidoreductase (*yqhD*) that is endogenous to *E. coli*. Since biosynthesis of the highly reduced product PDO from glucose requires additional inputs of NAD(P)H and ATP (figure 23), some glucose substrate needs to be oxidized through glycolysis and the TCA cycle to provide the necessary reducing equivalents and energy (figure 23). This creates a branch point at triose phosphate isomerase (*tpi*), which interconverts DHAP and glyceraldehyde-3-phosphate (G3P). G3P is fed into the TCA cycle to produce reducing equivalents and energy required for the PDO pathway and cell growth. Down-regulation of glyceraldehyde 3-phosphate dehydrogenase (*gap*) provides an improved flux control point while maximizing PDO yields. With additional modifications in the *E. coli* strain, the metabolically engineered organism provides PDO at a rate of 3.5 g L^{-1} h^{-1}, a titer of 135 g L^{-1}, and a yield of 0.51 g g^{-1} in glucose fed-batch aerobic fermentations.[122]

Fermentation of glucose to 1,2-propanediol (propylene glycol) in both homologous and heterologous hosts has also been investigated by various groups.[124-125] Major uses of propylene glycol are in unsaturated polyester resins, liquid laundry detergents, pharmaceuticals, cosmetics, antifreeze, and deicing formulations. Some microbes such as *Clostridium sphenoides* and *Thermoanaerobacterium thermosaccharolyticum* naturally ferment sugars to propylene glycol.[124] The pathways employed by these two organisms are very similar and have identical ATP and NAD(P)H cofactor requirements to the aforementioned PDO pathway (figure 23). DHAP is first dephosphorylated to methylglyoxal by the action of methylglyoxal synthase (an enzyme which is also native to *E. coli*). In *C. sphenoides*, the methylglyoxal is reduced to (*R*)-

lactaldehyde, whereas in *T. thermosaccharolyticum* the methylglyoxal is reduced to hydroxyacetone. Both of these intermediates are then reduced to (*R*)-1,2-propanediol. Propylene glycol production in these hosts is limited by low yields, a lack of genetic tools for rational biocatalyst improvement and, in the case of *C. sphenoides*, by the requirement for phosphate-free media (it is both difficult and expensive to obtain commercial-grade medium components that are free of phosphate). Since *E. coli* naturally produces methylglyoxal, recombinant production of propylene glycol can be readily achieved by providing the appropriate reductase activities.[124]

1,4-Butanediol (BDO) is an important commodity chemical used to manufacture more than 2.5 million tons annually of valuable polymers. Current production is based on feedstocks of fossil origin such as acetylene, propane, butane, and butadiene. Because BDO is not produced naturally in any known organism, its production in *E. coli* required the establishment of a novel heterologous biochemical pathway.[126] Computational design was used to screen reaction classes (rather than actual enzyme-catalyzed reactions) that could intercept common metabolites in *E. coli* and convert those to BDO. This broader approach identified more than 10,000 putative pathways capable of converting common metabolites (i.e., acetyl-CoA, succinyl-CoA, α-ketoglutarate, and glutamate) to BDO in just 4–6 steps. Computational methods were used to rank the pathways according to maximum theoretical BDO yield, pathway length, number of nonnative steps, number of novel steps, and thermodynamic feasibility. Constraint-based modeling was then used to eliminate pathways that displayed unfavorable thermodynamics or low theoretical yield. The remaining 1,000 pathways were ranked again according to the number of steps

catalyzed by uncharacterized enzymes (based on KEGG and EcoCyc databases, see section 2), the number of nonnative steps, and the total number of steps from central metabolism.

This process led to the selection of two artificial routes for BDO production from succinyl-CoA and α-ketoglutarate proceeding via 4-hydroxybutyrate as a common intermediate (figure 24). The pathway was divided into an upstream module (succinyl-CoA / α-ketoglutarate to 4-hydroxybutyrate) and a downstream module (4-hydroxybutyrate to BDO) for experimental validation and metabolic engineering. Only two out of the seven enzymes are native to *E. coli*; the remaining enzymes have to be imported from other organisms, as shown in figure 24. All enzymes act on their native substrates, with the exception of the aldehyde and alcohol dehydrogenases. A computational algorithm was used to predict gene knock-outs that couple BDO production to growth while preserving the maximum BDO theoretical yield. Knocking-out those genes eliminated *E. coli*'s ability to produce the usual fermentation products (ethanol, formate, lactate, succinate) thereby forcing production toward BDO to achieve redox balance. This led to improved BDO titers, but other intermediates were still formed preferentially, such as 4-hydroxybutyrate, succinate, and notably butyrolactone (likely formed spontaneously from 4-hydroxybutyrate-CoA, figure 24). Finally, improvements in expression of TCA cycle enzymes under microaerobic conditions by lifting transcriptional repression and NADH inhibition of citrate synthase, as well as selecting an aldehyde dehydrogenase that acts on 4-hydroxybutyraldehyde but not on acetyl-CoA, led to formation of BDO as the main product with a final titer of 18 g L^{-1}.[126]

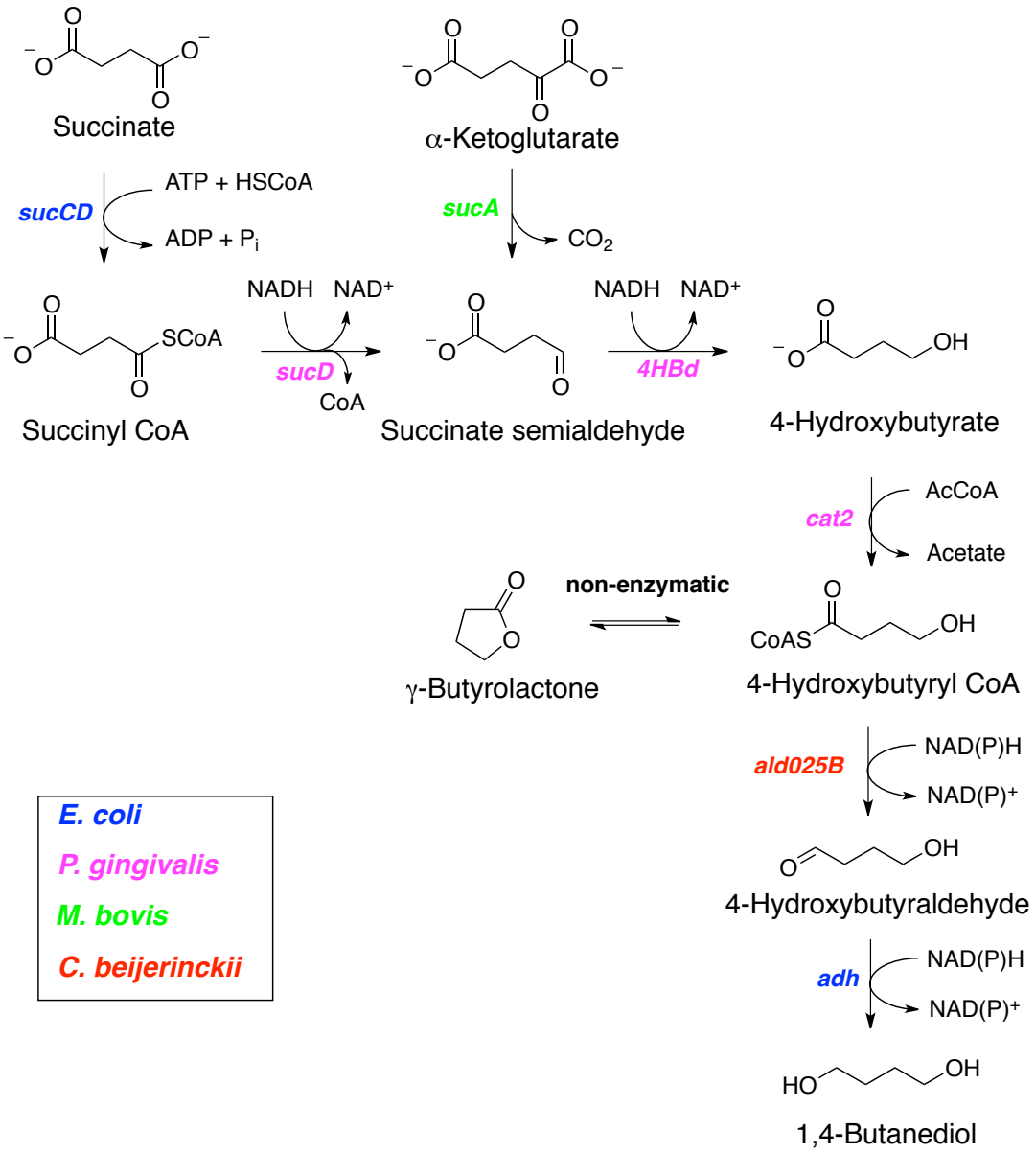


Figure 24. Computationally designed pathway for 1,4-butanediol production in *E. coli*. Enzymes encoded by the genes shown are: *sucCD*, succinyl-CoA synthetase (*E. coli*); *sucD*, semialdehyde dehydrogenase (*Porphyromonas gingivalis*); *4HBd*, 4-hydroxybutyrate dehydrogenase (*P. gingivalis*); *sucA*, alpha-keto acid decarboxylase (*Mycobacterium bovis*); *cat2*, 4-hydroxybutyryl-CoA transferase; *ald025B*, aldehyde dehydrogenase (*Clostridium beijerinckii*); *adh*, alcohol dehydrogenase (*E. coli*).[126]

4.7 Summary

Designing new synthetic pathways into microbial hosts demands careful consideration of redox balance, ATP balance, product toxicity, feedback inhibition, and host selection, the last factor accounting for growth rate and ability to grow on simple media. There is no universal approach for engineering microbial metabolism, and successful efforts necessarily employ a range of technologies as outlined above. It is noteworthy that most case studies thus far have not involved extensive use of enzyme engineering, but focused instead on optimizing and balancing expression of naturally occurring enzymes in a convenient host. Notable exceptions include the engineered levopimaradiene and isobutanol pathways listed above (sections 4.1.4 and 4.5). Fine-tuning of many recombinant pathways will likely involve more enzyme engineering to deliver enhanced kinetic parameters and substrate and product specificities at each catalytic node. However, the very possibility of such sensitive modulation should provide chemists with clear appreciation of the potential utility of biological catalysts relative to small molecule catalysts.

The three key metrics used to assess the utility of microbial processes—yield, volumetric productivity, and titer—are listed in table 2 for some of the pathways discussed in this section. These examples developed over the last decade affirm our ability to manipulate microbial metabolism for industrial production of chemicals. However, most of these processes do not yet match the thermodynamic and kinetic efficiencies of yeast ethanol fermentations, adapted and improved over millennia by artificial selection. Process robustness is also an important consideration for large-scale

implementation of these technologies. Ideally, the recombinant pathway should not require expensive inducers and selection markers (antibiotics), and the engineered strain should not require complex nutrients or ideally even sterile conditions for growth and should be tolerant to high product concentrations. This field is still in its infancy, and as our techniques grow in sophistication and our understanding of metabolic control deepens, we can look forward to enabling the vision of living factories for sustainable chemical production.

Table 2. Summary of key metrics for the pathways described in this section

Target product	Host	Theoretical yield (g_{pdt} $g_{Csource}^{-1}$)	Practical yield (g_{pdt} $g_{Csource}^{-1}$)	Titer (g L^{-1})	Volumetric productivity (g L^{-1} h^{-1})[c]
Isoprene[86]	*E. coli*	0.25[a]	0.11 (44%)	>60	2.0
Amorphadiene[95b]	*S. cerevisiae*	0.25	N/A	37	N/A
Taxadiene[101]	*E. coli*	0.30[b]	0.025 (7%)	1.00	N/A
L-Alanine[113]	*E. coli*	0.989	0.950 (96%)	114	2.37
L-Lysine[112]	*C. glutamicum*	0.75	0.55 (73%)	120	4.0
Putrescine[115]	*E. coli*	0.475	0.168 (35%)	24.2	0.75
Cadaverine[119]	*E. coli*	0.524	0.12 (23%)	9.61	0.32
Isobutanol	*S. cerevisiae*	0.41	0.41 (>95%)	N/A	2.0
1,3-Propanediol[122]	*E. coli*	0.607	0.51 (84%)	135	3.5
1,4-Butanediol [126]	*E. coli*	0.55	N/A	18.0	0.15
Ethanol	*S. cerevisiae*	0.51	~0.49 (96%)	~ 95	~ 8.0

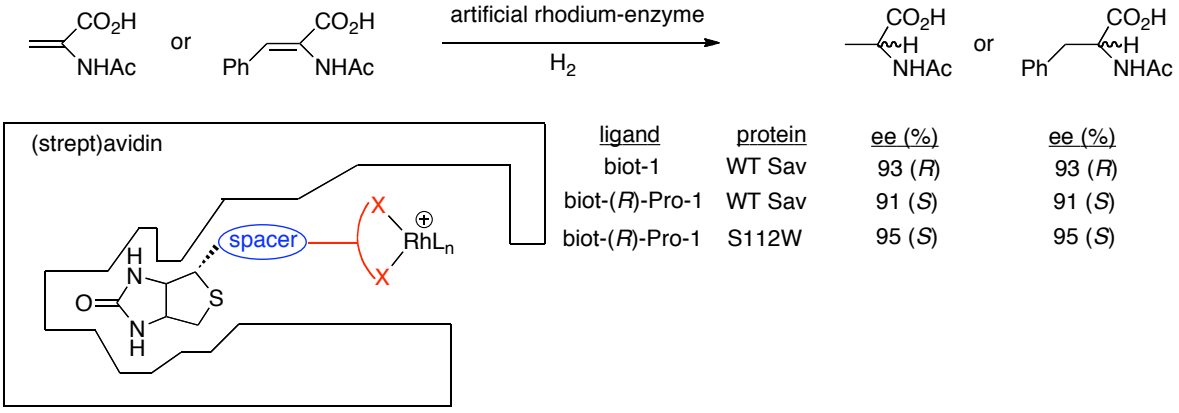
[a] Based on MVA pathway utilizing glucose as the carbon source. [b] Based on MEP pathway utilizing glycerol as the carbon source. [c] Average volumetric productivity was calculated based on total product produced per total fermentation time.

5. Beyond Nature: Design and Evolution of Artificial Function/Catalysis

Given the rate at which the biotechnologies outlined above are being applied to chemical synthesis, it is reasonable to consider more recent advances that might soon have a similar impact. While directed evolution of enzymes has clearly provided an extremely powerful tool to generate efficient catalysts, this process is only reliable once some initial activity is identified. Accordingly, researchers are starting to design enzymes *de novo* in order to reap the benefits of enzyme catalysis for reactions not found in nature; notable examples include designed enzymes for the Diels-Alder and retro-aldol reactions.[127] A range of nonnatural transition metal-catalyzed reactions have also been enabled by incorporating metal catalysts into protein scaffolds.[128] The several different systems that have been devised can be classified generally according to how the metal catalyst is anchored into the protein, namely covalent, supramolecular, or dative attachment. Among the most well-studied systems are the (strep)avidin-biotin metalloenzymes (supramolecular anchoring) devised by Whitesides and greatly improved by Ward, which have been used for a range of transition-metal-catalyzed reactions.[129]

Covalently linking a catalyst precursor (e.g., bisdiphenylphosphino rhodium) to a biotin anchor ensures that upon stoichiometric addition of (strept)avidin, the metal cofactor is quantitatively incorporated within the host protein. The resulting hybrid catalyst exhibits features of both homogenous and enzyme catalysis. For example, features reminiscent of homogeneous catalysis include straightforward access to both enantiomers of the product, broad substrate scope, organic solvent tolerance, whereas

acquired enzyme-like properties include access to genetic optimization, aqueous solubility and activity, and Michaelis-Menten behavior. This strategy has been used to create asymmetric systems for catalyzing rhodium-diphosphine hydrogenation of *N*-protected dehydroaminoacids (figure 25, up to 95% ee), palladium-diphosphine allylic alkylation (up to 95% ee), and vanadyl oxidation of sulfides (up to 93% ee).[129]



ligand	protein	ee (%)	ee (%)
biot-1	WT Sav	93 (*R*)	93 (*R*)
biot-(*R*)-Pro-1	WT Sav	91 (*S*)	91 (*S*)
biot-(*R*)-Pro-1	S112W	95 (*S*)	95 (*S*)

Figure 25. Reduction of *N*-acetamidoacrylic acid and *N*-acetamidocinnamic acid by artificial hydrogenases based on the biotin-avidin technology. The hybrid catalyst can be improved by chemogenetic optimization by varying the spacer (oval) or the chelating ligand, or by amino acid substitutions in the host protein [(strept)avidin].[129]

Nonnatural activities can also be installed into naturally occurring enzymes without the addition of synthetic cofactors. Taking advantage of the catalytic promiscuity of enzymes, panels of enzymes can be screened against synthetic reagents to discover abiological reactions. For example, Purkarthofer et al. screened a hydroxynitrile lyase for nitro-aldol (Henry reaction) activity on nitromethane and various aldehydes and found that the wild-type enzyme could catalyze the carboligation of nitromethane to benzaldehyde with high enantioselectivity (92% ee).[130] Similarly, a panel of engineered cytochrome P450s was screened against diazoester and azide reagents for C=C and C-H

functionalization by carbenoid and nitrenoid insertions.[131] P450 variants (with as few as one mutation from the wild-type) could catalyze the cyclopropanation of styrene from ethyl diazoacetate with high enantioselectivity and exquisite diastereocontrol (figure 26A) as well as the intramolecular C-H amination from arylsulfonyl azides (figure 26B).[131]

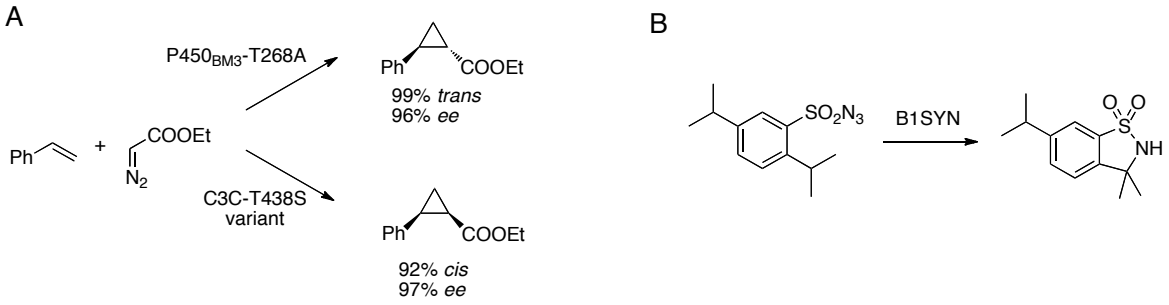


Figure 26. (A) $P450_{BM3}$-catalyzed cyclopropanation of styrene utilizing ethyl diazoacetate as a carbene precursor. Variant C3C-T438S has 14 mutations with respect to $P450_{BM3}$. (B) $P450_{BM3}$-catalyzed intramolecular C-H amination from an arylsulfonyl azide. Variant B1SYN has 23 mutations with respect to $P450_{BM3}$. [131]

Evolution of *de novo* designed enzymes and artificial metalloenzymes will be useful for their optimization for synthetic applications.[132] As illustrated in the previous section, most metabolic engineering projects have aimed at establishing heterologous production of molecules with well-understood natural pathways. As protein engineering advances in installing nonnatural activities into enzymes, we can expect that entirely artificial pathways will emerge for producing compounds of industrial relevance that are not synthesized naturally.

Of course both the catalysts and the format in which they are used can have a large impact on the utility and efficiency of a process. For example, some of the challenges associated with conducting reactions *in vivo* for metabolic engineering efforts

were outlined above. Due to these challenges, researchers have sought to design cells with minimal genomes that might simplify metabolic networks to minimize undesired reactivity, divert more energy and materials toward target products, and simplify downstream bioprocessing.[133] Further simplifications have also been realized using cell-free biosynthetic systems in which cells are grown to provide the necessary cofactors and catalysts for a target process, then lysed to release these materials in a reactor. Substrates, cosolvents, and other reagents that may be detrimental to cell survival can then be utilized, and the concentration of these species in the reaction medium can be precisely controlled.[134] The relative concentrations of enzymes can be controlled by combining isolated enzymes in a single reactor to enable selective sequential reactions on a substrate via *in vitro* cascades.[135] Thus, in all of these cases, chemists are given increasing control over the complex milieu in living cells. As these processes are improved, chemists will thus have an increasing palette of systems to draw on to carry out target reactions with increased efficiency.

References

1. Wetterstrand, K. A. DNA sequencing costs: Data from the NHGRI large-scale genome sequencing program. http://www.genome.gov/sequencingcosts.
2. http://www.lifetechnologies.com/us/en/home/about-us/news-gallery/press-releases/2012/life-techologies-itroduces-the-bechtop-io-proto.html.html.
3. (a) Venter, J. C.; Remington, K.; Heidelberg, J. F.; Halpern, A. L.; Rusch, D., et al., Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **2004,** *304* (5667), 66–74; (b) Vogel, T. M.; Simonet, P.; Jansson, J. K.; Hirsch, P. R.; Tiedje, J. M.; van Elsas, J. D.; Bailey, M. J.; Nalin, R.; Philippot, L., TerraGenome: a consortium for the sequencing of a soil metagenome. *Nature Reviews Microbiology* **2009,** *7* (4), 252.
4. Carlson, R., The changing economics of DNA synthesis. *Nature Biotechnology* **2009,** *27* (12), 1091–1094.
5. (a) Khorana, H. G., Total synthesis of a gene. *Science* **1979,** *203* (4381), 614–625; (b) Mandecki, W.; Hayden, M. A.; Shallcross, M. A.; Stotland, E. A totally synthetic plasmid for general cloning, gene expression and mutagenesis in Escherichia coli. *Gene* **1990,** *94* (1), 103–107; (c) Gibson, D. G.; Benders, G. A.; Andrews-Pfannkoch, C.; Denisova, E. A.; Baden-Tillson, *et al.*, Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **2008,** *319* (5867), 1215–1220.
6. Hughes, R. A.; Miklos, A. E.; Ellington, A. D., Gene synthesis: methods and applications. *Methods in Enzymology* **2011,** *498* (Synthetic Biology, Part B), 277–309.
7. Ellis, T.; Adie, T.; Baldwin, G. S., DNA assembly for synthetic biology: from parts to pathways and beyond. *Integrative Biology* **2011,** *3* (2), 109–118.
8. (a) Gibson, D. G.; Benders, G. A.; Axelrod, K. C.; Zaveri, J.; Algire, M. A.; Moodie, M.; Montague, M. G.; Venter, J. C.; Smith, H.; Hutchison, C. A., III, One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. *Proceedings of the National Academy of Sciences of the United States of America* **2008,** *105* (51), 20404–20409; (b) Itaya, M.; Fujita, K.; Kuroki, A.; Tsuge, K., Bottom-up genome assembly using the Bacillus subtilis genome vector. *Nature Methods* **2008,** *5* (1), 41–43.
9. Torsvik, V.; Ovrebs, L.; Thingstad, T. F., Prokaryotic diversity: Magnitude, dynamics, and controlling factors. *Science* **2002,** *296* (5570), 1064–1066.
10. (a) Lorenz, P.; Eck, J., Metagenomics and industrial applications. *Nature Reviews Microbiology* **2005,** *3* (6), 510–516; (b) Iqbal, H. A.; Feng, Z.; Brady, S. F., Biocatalysts and small molecule products from metagenomic studies. *Current Opinion in Chemical Biology* **2012,** *16* (1-2), 109–116.
11. Hu, X. P.; Heath, C.; Taylor, M. P.; Tuffin, M.; Cowan, D., A novel, extremely alkaliphilic and cold-active esterase from Antarctic desert soil. *Extremophiles* **2012,** *16* (1), 79–86.
12. Uchiyama, T.; Miyazaki, K., Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. *Applied and Environmental Microbiology* **2010,** *76* (21), 7029–7035.
13. Knietsch, A.; Bowien, S.; Whited, G.; Gottschalk, G.; Daniel, R., Identification and characterization of coenzyme B12-dependent glycerol dehydratase- and diol

dehydratase-encoding genes from metagenomic DNA libraries derived from enrichment cultures. *Applied and Environmental Microbiology* **2003,** *69* (6), 3048-3060.
14. (a) Piel, J., A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. *Proceedings of the National Academy of Sciences of the United States of America* **2002,** *99* (22), 14002-14007; (b) Schmidt, E. W.; Nelson, J. T.; Rasko, D. A.; Sudek, S.; Eisen, J. A.; Haygood, M. G.; Ravel, J., Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102* (20), 7315-7320.
15. Bayer, T. S.; Widmaier, D. M.; Temme, K.; Mirsky, E. A.; Santi, D. V.; Voigt, C. A., Synthesis of Methyl Halides from Biomass Using Engineered Microbes. *J. Am. Chem. Soc.* **2009,** *131* (18), 6508-6515.
16. Keasling, J. D., Synthetic Biology for Synthetic Chemistry. *ACS Chemical Biology* **2008,** *3* (1), 64-76.
17. Welch, M.; Villalobos, A.; Gustafsson, C.; Minshull, J., You're one in a googol: optimizing genes for protein expression. *Journal of the Royal Society, Interface* **2009,** *6* (Suppl. 4), S467-S476.
18. Esposito, D.; Chatterjee, D. K., Enhancement of soluble protein expression through the use of fusion tags. *Current Opinion in Biotechnology* **2006,** *17* (4), 353-358.
19. Thomas, J. G.; Ayling, A.; Baneyx, F., Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from E. coli: to fold or to refold. *Applied Biochemistry and Biotechnology* **1997,** *66* (3), 197-238.
20. Makino, T.; Skretas, G.; Georgiou, G., Strain engineering for improved expression of recombinant proteins in bacteria. *Microbial Cell Factories* **2011,** *10*, 32.
21. Clouthier, C. M.; Pelletier, J. N., Expanding the organic toolbox: a guide to integrating biocatalysis in synthesis. *Chemical Society Reviews* **2012,** *41* (4), 1585-1605.
22. Lairson, L. L.; Watts, A. G.; Wakarchuk, W. W.; Withers, S. G., Using substrate engineering to harness enzymatic promiscuity and expand biological catalysis. *Nature Chemical Biology* **2006,** *2* (12), 724-728.
23. Kim, P.-Y.; Pollard, D. J.; Woodley, J. M., Substrate Supply for Effective Biocatalysis. *Biotechnology Progress* **2007,** *23* (1), 74-82.
24. Strohmeier, G. A.; Pichler, H.; May, O.; Gruber-Khadjawi, M., Application of Designed Enzymes in Organic Synthesis. *Chemical Reviews* **2011,** *111* (7), 4141-4164.
25. Tracewell, C. A.; Arnold, F. H., Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Current Opinion in Chemical Biology* **2009,** *13* (1), 3-9.
26. Arnold, F. H., Design by directed evolution. *Accounts of Chemical Research* **1998,** *31* (3), 125-131.
27. (a) Turner, N. J., Directed evolution drives the next generation of biocatalysts. *Nature Chemical Biology* **2009,** *5* (8), 567-573; (b) Reetz, M. T., Laboratory Evolution of Stereoselective Enzymes: A Prolific Source of Catalysts for Asymmetric Reactions. *Angewandte Chemie, International Edition* **2011,** *50* (1), 138-174.
28. Tee, K. L.; Schwaneberg, U., Directed evolution of oxygenases: Screening systems, success stories and challenges. *Combinatorial Chemistry & High Throughput Screening* **2007,** *10* (3), 197-217.

29. Fasan, R., Tuning P450 Enzymes as Oxidation Catalysts. *ACS Catalysis* **2012,** *2* (4), 647-666.
30. Jung, S. T.; Lauchli, R.; Arnold, F. H., Cytochrome P 450: taming a wild-type enzyme. *Current Opinion in Biotechnology* **2011,** *22* (6), 809-817.
31. Warman, A. J.; Roitel, O.; Neeli, R.; Girvan, H. M.; Seward, H. E.; Murray, S. A.; McLean, K. J.; Joyce, M. G.; Toogood, H.; Holt, R. A.; Leys, D.; Scrutton, N. S.; Munro, A. W., Flavocytochrome P 450 BM3: an update on structure and mechanism of a biotechnologically important enzyme. *Biochemical Society Transactions* **2005,** *33* (4), 747-753.
32. Lewis, J. C.; Arnold, F. H., Catalysts on demand: selective oxidations by laboratory-evolved cytochrome P450 BM3. *Chimia* **2009,** *63* (6), 309-312.
33. Li, S.; Chaulagain, M. R.; Knauff, A. R.; Podust, L. M.; Montgomery, J.; Sherman, D. H., Selective oxidation of carbolide C-H bonds by an engineered macrolide P450 mono-oxygenase. *Proceedings of the National Academy of Sciences of the United States of America* **2009,** *106* (44), 18463-18468.
34. Meyer, A.; Wursten, M.; Schmid, A.; Kohler, H.-P. E.; Witholt, B., Hydroxylation of Indole by Laboratory-evolved 2-Hydroxybiphenyl 3-Monooxygenase. *Journal of Biological Chemistry* **2002,** *277* (37), 34161-34167.
35. van den Heuvel, R. H. H.; van den Berg, W. A. M.; Rovida, S.; van Berkel, W. J. H., Laboratory-evolved vanillyl-alcohol oxidase produces natural vanillin. *Journal of Biological Chemistry* **2004,** *279* (32), 33492-33500.
36. Gursky, L. J.; Nikodinovic-Runic, J.; Feenstra, K. A.; O'Connor, K. E., In vitro evolution of styrene monooxygenase from Pseudomonas putida CA-3 for improved epoxide synthesis. *Applied Microbiology and Biotechnology* **2010,** *85* (4), 995-1004.
37. Rehdorf, J.; Bornscheuer, U. T., Monooxygenases, Baeyer-Villiger applications in organic synthesis. *Encyclopedia of Industrial Biotechnology* **2010,** *5*, 3572-3607.
38. Zhang, Z.-G.; Parra, L. P.; Reetz, M. T., Protein Engineering of Stereoselective Baeyer-Villiger Monooxygenases. *Chemistry, A European Journal* **2012,** *18* (33), 10160-10172.
39. Boyd, D. R.; Bugg, T. D. H., Arene cis-dihydrodiol formation: from biology to application. *Organic & Biomolecular Chemistry* **2006,** *4* (2), 181-192.
40. Furukawa, K.; Suenaga, H.; Goto, M., Biphenyl dioxygenases: Functional versatilities and directed evolution. *Journal of Bacteriology* **2004,** *186* (16), 5189-5196.
41. Rui, L.; Reardon, K. F.; Wood, T. K., Protein engineering of toluene ortho-monooxygenase of Burkholderia cepacia G4 for regiospecific hydroxylation of indole to form various indigoid compounds. *Applied Microbiology and Biotechnology* **2005,** *66* (4), 422-429.
42. Vardar, G.; Ryu, K.; Wood, T. K., Protein engineering of toluene-o-xylene monooxygenase from Pseudomonas stutzeri OX1 for oxidizing nitrobenzene to 3-nitrocatechol, 4-nitrocatechol, and nitrohydroquinone. *Journal of Biotechnology* **2005,** *115* (2), 145-156.
43. Fishman, A.; Tao, Y.; Rui, L.; Wood, T. K., Controlling the Regiospecific Oxidation of Aromatics via Active Site Engineering of Toluene para-Monooxygenase of Ralstonia pickettii PKO1. *Journal of Biological Chemistry* **2005,** *280* (1), 506-514.

44. Rai, G. P.; Sakai, S.; Florez, A. M.; Mogollon, L.; Hager, L. P., Directed evolution of chloroperoxidase for improved epoxidation and chlorination catalysis. *Advanced Synthesis & Catalysis* **2001,** *343* (6+7), 638-645.
45. Hasan, Z.; Renirie, R.; Kerkman, R.; Ruijssenaars, H. J.; Hartog, A. F.; Wever, R., Laboratory-evolved Vanadium Chloroperoxidase Exhibits 100-Fold Higher Halogenating Activity at Alkaline pH: Catalytic effects from first and second coordination sphere mutations. *Journal of Biological Chemistry* **2006,** *281* (14), 9738-9744.
46. Butler, A.; Sandy, M., Mechanistic considerations of halogenating enzymes. *Nature* **2009,** *460* (7257), 848-854.
47. Lang, A.; Polnick, S.; Nicke, T.; William, P.; Patallo, E. P.; Naismith, J. H.; van Pee, K.-H., Changing the Regioselectivity of the Tryptophan 7-Halogenase PrnA by Site-Directed Mutagenesis. *Angewandte Chemie, International Edition* **2011,** *50* (13), 2951-2953.
48. Glenn, W. S.; Nims, E.; O'Connor, S. E., Reengineering a tryptophan halogenase to preferentially chlorinate a direct alkaloid precursor. *Journal of the American Chemical Society* **2011,** *133* (48), 19346-19349.
49. Chen, S.; Engel, P. C., Efficient screening for new amino acid dehydrogenase activity: Directed evolution of Bacillus sphaericus phenylalanine dehydrogenase towards activity with an unsaturated non-natural amino acid. *Journal of Biotechnology* **2009,** *142* (2), 127-134.
50. Wichmann, R.; Vasic-Racki, D., Cofactor regeneration at the lab scale. *Advanced Biochemical Engineering/Biotechnology* **2005,** *92* (Technology Transfer in Biotechnology), 225-260.
51. Sacchi, S.; Rosini, E.; Molla, G.; Pilone, M. S.; Pollegioni, L., Modulating D-amino acid oxidase substrate specificity: production of an enzyme for analytical determination of all D-amino acids by directed evolution. *Protein Engineering, Design & Selection* **2004,** *17* (6), 517-525.
52. Carr, R.; Alexeeva, M.; Enright, A.; Eve, T. S. C.; Dawson, M. J.; Turner, N. J., Directed evolution of an amine oxidase possessing both broad substrate specificity and high enantioselectivity. *Angewandte Chemie, International Edition* **2003,** *42* (39), 4807-4810.
53. Escalettes, F.; Turner, N. J., Directed evolution of galactose oxidase: generation of enantioselective secondary alcohol oxidases. *Chembiochem* **2008,** *9* (6), 857-860.
54. (a) Bastian, S.; Rekowski, M. J.; Witte, K.; Heckmann-Pohl, D. M.; Giffhorn, F., Engineering of pyranose 2-oxidase from Peniophora gigantea towards improved thermostability and catalytic efficiency. *Applied Microbiology and Biotechnology* **2005,** *67* (5), 654-663; (b) Sun, L.; Petrounia, I. P.; Yagasaki, M.; Bandara, G.; Arnold, F. H., Expression and stabilization of galactose oxidase in Escherichia coli by directed evolution. *Protein Engineering* **2001,** *14* (9), 699-704.
55. Moore, J. C.; Pollard, D. J.; Kosjek, B.; Devine, P. N., Advances in the Enzymatic Reduction of Ketones. *Accounts of Chemical Research* **2007,** *40* (12), 1412-1419.

56. Bougioukou, D. J.; Kille, S.; Taglieber, A.; Reetz, M. T., Directed Evolution of an Enantioselective Enoate-Reductase: Testing the Utility of Iterative Saturation Mutagenesis. *Advanced Synthesis & Catalysis* **2009,** *351* (18), 3287-3305.
57. Svedendahl, M.; Branneby, C.; Lindberg, L.; Berglund, P., Reversed Enantiopreference of an omega-Transaminase by a Single-Point Mutation. *ChemCatChem* **2010,** *2* (8), 976-980.
58. Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J., Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science* **2010,** *329* (5989), 305-309.
59. Williams, G. J.; Zhang, C.; Thorson, J. S., Expanding the promiscuity of a natural-product glycosyltransferase by directed evolution. *Nature Chemical Biology* **2007,** *3* (10), 657-662.
60. Shaikh, F. A.; Withers, S. G., Teaching old enzymes new tricks: engineering and evolution of glycosidases and glycosyl transferases for improved glycoside synthesis. *Biochemistry and Cell Biology* **2008,** *86* (2), 169-177.
61. Koryakina, I.; Neville, J.; Nonaka, K.; Van Lanen, S. G.; Williams, G. J., A high-throughput screen for directed evolution of the natural product sulfotransferase LipB. *Journal of Biomolecular Screening* **2011,** *16* (8), 845-851.
62. Mihara, Y.; Utagawa, T.; Yamada, H.; Asano, Y., Phosphorylation of nucleosides by the mutated acid phosphatase from Morganella morganii. *Applied and Environmental Microbiology* **2000,** *66* (7), 2811-2816.
63. Griswold, K. E.; Aiyappan, N. S.; Iverson, B. L.; Georgiou, G., The Evolution of Catalytic Efficiency and Substrate Promiscuity in Human Theta Class 1-1 Glutathione Transferase. *Journal of Molecular Biology* **2006,** *364* (3), 400-410.
64. Schmidt, M.; Boettcher, D.; Bornscheuer, U. T., Protein engineering of carboxyl esterases by rational design and directed evolution. *Protein & Peptide Letters* **2009,** *16* (10), 1162-1171.
65. Lutz, S., Engineering lipase B from Candida antarctica. *Tetrahedron: Asymmetry* **2004,** *15* (18), 2743-2748.
66. Reetz, M. T.; Bocola, M.; Wang, L.-W.; Sanchis, J.; Cronin, A.; Arand, M.; Zou, J.; Archelas, A.; Bottalla, A.-L.; Naworyta, A.; Mowbray, S. L., Directed Evolution of an Enantioselective Epoxide Hydrolase: Uncovering the Source of Enantioselectivity at Each Evolutionary Stage. *Journal of the American Chemical Society* **2009,** *131* (21), 7334-7343.
67. Ma, S. K.; Gruber, J.; Davis, C.; Newman, L.; Gray, D.; Wang, A.; Grate, J.; Huisman, G. W.; Sheldon, R. A., A green-by-design biocatalytic process for atorvastatin intermediate. *Green Chemistry* **2010,** *12* (1), 81-86.
68. Tang, L.; Zhu, X.; Zheng, H.; Jiang, R.; Elenkov, M. M., Key residues for controlling enantioselectivity of halohydrin dehalogenase from Arthrobacter sp. strain AD2, revealed by structure-guided directed evolution. *Applied and Environmental Microbiology* **2012,** *78* (8), 2631-2637.

69. Bolt, A.; Berry, A.; Nelson, A., Directed evolution of aldolases for exploitation in synthetic organic chemistry. *Archives of Biochemistry and Biophysics* **2008,** *474* (2), 318-330.
70. (a) González-García, E.; Helaine, V.; Klein, G.; Schuermann, M.; Sprenger, G. A.; Fessner, W. -D.; Reymond, J.-L., Fluorogenic stereochemical probes of transaldolases. *Chemistry, an European Journal* **2003**, *9,* 893–899 (b) Sugiyama, M.; Greenberg, W.; Wong, C.-H., Recent advances in aldolase-catalyzed synthesis of unnatural sugars and iminocyclitols. *Yuki Gosei Kagaku Kyokaishi* **2008,** *66* (6), 605-615.
71. Greenberg, W. A.; Varvak, A.; Hanson, S. R.; Wong, K.; Huang, H.; Chen, P.; Burk, M. J., Development of an efficient, scalable, aldolase-catalyzed process for enantioselective synthesis of statin intermediates. *Proceedings of the National Academy of Sciences of the United States of America* **2004,** *101* (16), 5788-5793.
72. Woodhall, T.; Williams, G.; Berry, A.; Nelson, A., Creation of a tailored aldolase for the parallel synthesis of sialic acid mimetics. *Angewandte Chemie, International Edition* **2005,** *44* (14), 2109-2112.
73. Williams, G. J.; Woodhall, T.; Farnsworth, L. M.; Nelson, A.; Berry, A., Creation of a Pair of Stereochemically Complementary Biocatalysts. *Journal of the American Chemical Society* **2006,** *128* (50), 16238-16247.
74. Fong, S.; Machajewski, T. D.; Mak, C. C.; Wong, C.-H., Directed evolution of D-2-keto-3-deoxy-6-phosphogluconate aldolase to new variants for the efficient synthesis of D- and L-sugars. *Chemistry & Biology* **2000,** *7* (11), 873-883.
75. Liu, Z.; Pscheidt, B.; Avi, M.; Gaisberger, R.; Hartner, F. S.; Schuster, C.; Skranc, W.; Gruber, K.; Glieder, A., Laboratory evolved biocatalysts for stereoselective syntheses of substituted benzaldehyde cyanohydrins. *Chembiochem* **2008,** *9* (1), 58-61.
76. (a) Asano, Y.; Kira, I.; Yokozeki, K., Alteration of substrate specificity of aspartase by directed evolution. *Biomolecular Engineering* **2005,** *22* (1-3), 95-101; (b) Bartsch, S.; Bornscheuer, U. T., Mutational analysis of phenylalanine ammonia lyase to improve reactions rates for various substrates. *Protein Engineering, Design & Selection* **2010,** *23* (12), 929-933.
77. Yoshikuni, Y.; Ferrin, T. E.; Keasling, J. D., Designed divergent evolution of enzyme function. *Nature* **2006,** *440* (7087), 1078-1082.
78. Lee, H. Y.; Harvey, C. J. B.; Cane, D. E.; Khosla, C., Improved precursor-directed biosynthesis in E. coli via directed evolution. *Journal of Antibiotics* **2011,** *64* (1), 59-64.
79. Evans, B. S.; Chen, Y.; Metcalf, W. W.; Zhao, H.; Kelleher, N. L., Directed evolution of the nonribosomal peptide synthetase AdmK generates new andrimid derivatives in vivo. *Chemistry & Biology* **2011,** *18* (5), 601-607.
80. (a) Chen, K.; Arnold, F. H., Tuning the activity of an enzyme for unusual environments: Sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proceedings of the National Academy of Sciences of the United States of America* **1993,** *90* (12), 5618-22; (b) Stemmer, W. P. C., Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **1994,** *370* (6488), 389-91.
81. Schoemaker, H. E.; Mink, D.; Wubbolts, M. G., Dispelling the Myths-Biocatalysis in Industrial Synthesis. *Science* **2003,** *299* (5613), 1694-1698.

82. Lewis, J. C.; Coelho, P. S.; Arnold, F. H., Enzymatic functionalization of carbon-hydrogen bonds. *Chemical Society Reviews* **2011,** *40* (4), 2003-2021.
83. Walji, A. M.; MacMillan, D. W. C., Strategies to bypass the taxol problem. Enantioselective cascade catalysis, a new approach for the efficient construction of molecular complexity. *Synlett* **2007,** (10), 1477-1489.
84. Lee, J. W.; Kim, T. Y.; Jang, Y.-S.; Choi, S.; Lee, S. Y., Systems metabolic engineering for chemicals and materials. *Trends in Biotechnology* **2011,** *29* (8), 370-378.
85. Christianson David, W., Chemistry. Roots of biosynthetic diversity. *Science* **2007,** *316* (5821), 60-1.
86. Whited, G. M.; Feher, F. J.; Benko, D. A.; Cervin, M. A.; Chotani, G. K.; McAuliffe, J. C.; La Duca, R. J.; Ben-Shoshan, E. A.; Sanford, K. J., Development of a gas-phase bioprocess for isoprene-monomer production using metabolic pathway engineering. *Industrial Biotechnology* **2010,** *6* (3), 152-163.
87. Martin, V. J. J.; Pitera, D. J.; Withers, S. T.; Newman, J. D.; Keasling, J. D., Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nature Biotechnology* **2003,** *21* (7), 796-802.
88. Julsing, M. K.; Rijpkema, M.; Woerdenbag, H. J.; Quax, W. J.; Kayser, O., Functional analysis of genes involved in the biosynthesis of isoprene in Bacillus subtilis. *Applied Microbiology and Biotechnology* **2007,** *75* (6), 1377-1384.
89. Cervin, M. A.; Whited, G. M.; Miasnikov, A.; Peres, C. M.; Weyler, W.; Wells, D. H.; Bott, R. R. Plant isoprene synthase variants for improved microbial production of isoprene. 2009-US415812009132220, 20090423., 2009.
90. Beck, Z. Q.; Estell, D. A.; Rife, C. L.; Wells, D. H.; Miller, J. V. Isoprene synthase variants for improved production of isoprene. 2011-US581882012058494, 20111027., 2012.
91. (a) Miller, B.; Oschinski, C.; Zimmer, W., First isolation of an isoprene synthase gene from poplar and successful expression of the gene in Escherichia coli. *Planta* **2001,** *213* (3), 483-487; (b) Lindberg, P.; Park, S.; Melis, A., Engineering a platform for photosynthetic isoprene production in cyanobacteria, using Synechocystis as the model organism. *Metabolic Engineering* **2010,** *12* (1), 70-79; (c) Hong, S.-Y.; Zurbriggen Andreas, S.; Melis, A., Isoprene hydrocarbons production upon heterologous transformation of Saccharomyces cerevisiae. *Journal of Applied Microbiology* **2012**, *113* (1), 52-65.
92. Anthony, J. R.; Anthony, L. C.; Nowroozi, F.; Kwon, G.; Newman, J. D.; Keasling, J. D., Optimization of the mevalonate-based isoprenoid biosynthetic pathway in Escherichia coli for production of the anti-malarial drug precursor amorpha-4,11-diene. *Metabolic Engineering* **2009,** *11* (1), 13-19.
93. Ye, V. M.; Bhatia, S. K., Metabolic engineering for the production of clinically important molecules: Omega-3 fatty acids, artemisinin, and taxol. *Biotechnology Journal* **2012,** *7* (1), 20-33.
94. Newman, J. D.; Marshall, J.; Chang, M.; Nowroozi, F.; Paradise, E.; Pitera, D.; Newman, K. L.; Keasling, J. D., High-level production of amorpha-4,11-diene in a two-phase partitioning bioreactor of metabolically engineered Escherichia coli. *Biotechnol. Bioeng.* **2006,** *95* (4), 684-691.

95. (a) Keasling, J. D., Synthetic biology and the development of tools for metabolic engineering. *Metabolic Engineering* **2012,** *14* (3), 189-195; (b) Westfall, P. J.; Pitera, D. J.; Lenihan, J. R.; Eng, D.; Woolard, *et al.*, Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proceedings of the National Academy of Sciences of the United States of America* **2012,** *109* (3), E111-E118.
96. Pitera, D. J.; Paddon, C. J.; Newman, J. D.; Keasling, J. D., Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli. *Metabolic Engineering* **2007,** *9* (2), 193-207.
97. Tsuruta, H.; Paddon, C. J.; Eng, D.; Lenihan, J. R.; Horning, T.; Anthony, L. C.; Regentin, R.; Keasling, J. D.; Renninger, N. S.; Newman, J. D., High-level production of amorpha-4,11-diene, a precursor of the antimalarial agent artemisinin, in Escherichia coli. *PLoS One* **2009,** *4* (2), No pp given.
98. Chang, M. C. Y.; Eachus, R. A.; Trieu, W.; Ro, D.-K.; Keasling, J. D., Engineering Escherichia coli for production of functionalized terpenoids using plant P450s. *Nature Chemical Biology* **2007,** *3* (5), 274-277.
99. Roth, R. J.; Acton, N., A simple conversion of artemisinic acid into artemisinin. *Journal of Natural Products* **1989,** *52* (5), 1183-1185.
100. Ro, D.-K.; Paradise Eric, M.; Ouellet, M.; Fisher Karl, J.; Newman Karyn, L.; Ndungu John, M.; Ho Kimberly, A.; Eachus Rachel, A.; Ham Timothy, S.; Kirby, J.; Chang Michelle, C. Y.; Withers Sydnor, T.; Shiba, Y.; Sarpong, R.; Keasling Jay, D., Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **2006,** *440* (7086), 940-943.
101. Ajikumar, P. K.; Xiao, W.-H.; Tyo, K. E. J.; Wang, Y.; Simeon, F.; Leonard, E.; Mucha, O.; Phon, T. H.; Pfeifer, B.; Stephanopoulos, G., Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in Escherichia coli. *Science* **2010,** *330* (6000), 70-74.
102. Leonard, E.; Ajikumar, P. K.; Thayer, K.; Xiao, W.-H.; Mo, J. D.; Tidor, B.; Stephanopoulos, G.; Prather, K. L. J., Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proceedings of the National Academy of Sciences of the United States of America* **2010,** *107* (31), 13654-13659.
103. Veen, M.; Stahl, U.; Lang, C., Combined overexpression of genes of the ergosterol biosynthetic pathway leads to accumulation of sterols in Saccharomyces cerevisiae. *FEMS Yeast Research* **2003,** *4* (1), 87-95.
104. Shang, F.; Wen, S.; Wang, X.; Tan, T., Effect of nitrogen limitation on the ergosterol production by fed-batch culture of Saccharomyces cerevisiae. *Journal of Biotechnology* **2006,** *122* (3), 285-292.
105. Szczebara, F. M.; Chandelier, C.; Villeret, C.; Masurel, A.; Bourot, S.; Duport, C.; Blanchard, S.; Groisillier, A.; Testet, E.; Costaglioli, P.; Cauet, G.; Degryse, E.; Balbuena, D.; Winter, J.; Achstetter, T.; Spagnoli, R.; Pompon, D.; Dumas, B., Total biosynthesis of hydrocortisone from a simple carbon source in yeast. *Nature Biotechnology* **2003,** *21* (2), 143-149.

106. Duport, C.; Spagnoli, R.; Degryse, E.; Pompon, D., Self-sufficient biosynthesis of pregnenolone and progesterone in engineered yeast. *Nature Biotechnology* **1998,** *16* (2), 186-189.
107. Park, J. H.; Lee, S. Y., Metabolic pathways and fermentative production of L-aspartate family amino acids. *Biotechnology Journal* **2010,** *5* (6), 560-577.
108. Kind, S.; Jeong, W. K.; Schroeder, H.; Wittmann, C., Systems-wide metabolic pathway engineering in Corynebacterium glutamicum for bio-based production of diaminopentane. *Metabolic Engineering* **2010,** *12* (4), 341-351.
109. Ohnishi, J.; Mitsuhashi, S.; Hayashi, M.; Ando, S.; Yokoi, H.; Ochiai, K.; Ikeda, M., A novel methodology employing Corynebacterium glutamicum genome information to generate a new L-lysine-producing mutant. *Applied Microbiology and Biotechnology* **2002,** *58* (2), 217-23.
110. Kinoshita, S.; Nakayama, K.; Kitada, S., L-Lysine production using microbial auxotroph, Preliminary report. *The Journal of General and Applied Microbiology* **1958,** *4* (2), 128-129.
111. Kromer Jens, O.; Sorgenfrei, O.; Klopprogge, K.; Heinzle, E.; Wittmann, C., In-depth profiling of lysine-producing Corynebacterium glutamicum by combined analysis of the transcriptome, metabolome, and fluxome. *Jorunal of Bacteriology* **2004,** *186* (6), 1769-84.
112. Becker, J.; Zelder, O.; Häfner, S.; Schröder, H.; Wittmann, C., From zero to hero, Design-based systems metabolic engineering of Corynebacterium glutamicum for l-lysine production. *Metabolic Engineering* **2011,** *13* (2), 159-168.
113. Zhang, X.; Jantama, K.; Moore, J. C.; Shanmugam, K. T.; Ingram, L. O., Production of L-alanine by metabolically engineered Escherichia coli. *Applied Microbiology and Biotechnology* **2007,** *77* (2), 355-366.
114. Igarashi, K.; Kashiwagi, K., Polyamines: Mysterious Modulators of Cellular Functions. *Biochemical and Biophysical Research Communications* **2000,** *271* (3), 559-564.
115. Qian, Z.-G.; Xia, X.-X.; Lee, S. Y., Metabolic engineering of Escherichia coli for the production of putrescine: A four carbon diamine. *Biotechnology and Bioengineering* **2009,** *104* (4), 651-662.
116. Tabor, C. W.; Tabor, H., Polyamines in microorganisms. *Microbiological Reviews* **1985,** *49* (1), 81-99.
117. Mimitsuka, T.; Sawai, H.; Hatsu, M.; Yamada, K., Metabolic engineering of Corynebacterium glutamicum for cadaverine fermentation. *Bioscience, Biotechnology, and Biochemistry* **2007,** *71* (9), 2130-2135.
118. Kind, S.; Kreye, S.; Wittmann, C., Metabolic engineering of cellular transport for overproduction of the platform chemical 1,5-diaminopentane in Corynebacterium glutamicum. *Metabolic Engineering* **2011,** *13* (5), 617-627.
119. Qian, Z.-G.; Xia, X.-X.; Lee, S. Y., Metabolic engineering of Escherichia coli for the production of cadaverine: A five carbon diamine. *Biotechnology and Bioengineering* **2010,** *108* (1), 93-103.
120. Atsumi, S.; Hanai, T.; Liao, J. C., Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **2008,** *451* (7174), 86-89.

121. Bastian, S.; Liu, X.; Meyerowitz, J. T.; Snow, C. D.; Chen, M. M. Y.; Arnold, F. H., Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in Escherichia coli. *Metabolic Engineering* **2011,** *13* (3), 345-352.
122. Nakamura, C. E.; Whited, G. M., Metabolic engineering for the microbial production of 1,3-propanediol. *Current Opinion in Biotechnology* **2003,** *14* (5), 454-459.
123. Raynaud, C.; Sarcabal, P.; Meynial-Salles, I.; Croux, C.; Soucaille, P., Molecular characterization of the 1,3-propanediol (1,3-PD) operon of Clostridium butyricum. *Proceedings of the National Academy of Sciences of the United States of America* **2003,** *100* (9), 5010-5015.
124. Cameron, D. C.; Altaras, N. E.; Hoffman, M. L.; Shaw, A. J., Metabolic engineering of propanediol pathways. *Biotechnology Progress* **1998,** *14* (1), 116-25.
125. (a) Altaras, N. E.; Cameron, D. C., Enhanced Production of (R)-1,2-Propanediol by Metabolically Engineered Escherichia coli. *Biotechnology Progress* **2000,** *16* (6), 940-946; (b) Berrios-Rivera, S. J.; San, K.-Y.; Bennett, G. N., The effect of carbon sources and lactate dehydrogenase deletion on 1,2-propanediol production in Escherichia coli. *Journal of Industrial Microbiology & Biotechnology* **2003,** *30* (1), 34-40.
126. Yim, H.; Haselbeck, R.; Niu, W.; Pujol-Baxley, C.; Burgard, A.; Boldt, J.; Khandurina, J.; Trawick, J. D.; Osterhout, R. E.; Stephen, R.; Estadilla, J.; Teisan, S.; Schreyer, H. B.; Andrae, S.; Yang, T. H.; Lee, S. Y.; Burk, M. J.; Van Dien, S., Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. *Nature Chemical Biology* **2011,** *7* (7), 445-452.
127. Richter, F.; Leaver-Fay, A.; Khare, S. D.; Bjelic, S.; Baker, D., De novo enzyme design using Rosetta3. *PLoS One* **2011,** *6* (5), e19230.
128. Kohler, V.; Wilson Yvonne, M.; Lo, C.; Sardo, A.; Ward Thomas, R., Protein-based hybrid catalysts--design and evolution. *Current Opinion in Biotechnology* **2010,** *21* (6), 744-52.
129. Ward, T. R., Artificial Metalloenzymes Based on the Biotin-Avidin Technology: Enantioselective Catalysis and Beyond. *Accounts of Chemical Research* **2011,** *44* (1), 47-57.
130. Purkarthofer, T.; Gruber, K.; Gruber-Khadjawi, M.; Waich, K.; Skranc, W.; Mink, D.; Griengl, H., A biocatalytic Henry reaction - the hydroxynitrile lyase from Hevea brasiliensis also catalyzes nitroaldol reactions. *Angewandte Chemie, International Edition* **2006,** *45* (21), 3454-3456.
131. Coelho, P. S.; Brustad, E. M.; Kannan, A.; McIntosh, J. A.; Brown, T. R.; Lewis, J. C.; Arnold Frances, H., C=C and C-H Functionalization by Enzyme-Catalyzed Carbenoid and Nitrenoid Insertion. **2012,** *Manuscript submitted*.
132. Brustad, E. M.; Arnold, F. H., Optimizing non-natural protein function with directed evolution. *Current Opinion in Chemical Biology* **2011,** *15* (2), 201-210.
133. Chiarabelli, C.; Stano, P.; Anella, F.; Carrara, P.; Luisi, P. L., Approaches to chemical synthetic biology. *FEBS Letters* **2012,** *586* (15), 2138-2145.
134. Hodgman, C. E.; Jewett, M. C., Cell-free synthetic biology: Thinking outside the cell. *Metabolic Engineering* **2012,** *14* (3), 261-269.

135. Lopez-Gallego, F.; Schmidt-Dannert, C., Multi-enzymatic synthesis. *Current Opinion in Chemical Biology* **2010,** *14* (2), 174-183.

Chapter 2

OLEFIN CYCLOPROPANATION VIA CARBENE TRANSFER CATALYZED BY ENGINEERED CYTOCHROME P450 ENZYMES

This chapter is published as P. S. Coelho, E. M. Brustad, A. Kannan, F. H. Arnold, *Science* **339**, 307 (2013).

Abstract

Transition metal-catalyzed transfers of carbenes, nitrenes and oxenes are powerful methods for functionalizing C=C and C–H bonds. Nature has evolved a diverse toolbox for oxene transfers, as exemplified by the myriad monooxygenation reactions catalyzed by cytochrome P450 enzymes. The isoelectronic carbene transfer to olefins, a widely used C–C bond forming reaction in organic synthesis, has no biological counterpart. Here, we report engineered variants of cytochrome $P450_{BM3}$ that catalyze highly diastereo- and enantioselective cyclopropanation of styrenes from diazoester reagents via putative carbene transfer. This work highlights the capacity to adapt existing enzymes for the catalysis of synthetically important reactions not previously observed in nature.

One sentence summary. A bacterial cytochrome P450 has been readily engineered to catalyze highly stereoselective carbene transfers to aryl-substituted olefins, enabling cyclopropanation through a route not observed in Nature.

The many strategies for functionalizing C=C and C–H bonds that have evolved in nature have captivated the imaginations of chemists and form the foundation of biomimetic chemistry (*1, 2*). The reverse of this, using inspiration from synthetic chemistry to discover and develop new biocatalysts, is a nascent frontier in molecular engineering, whose recent highlights include C–H activation by artificial rhodium enzymes (*3*) and the *de novo* design of Diels-Alderases (*3, 4*). Synthetic chemists have developed powerful methods for direct C=C and C–H functionalization based on transition metal-catalyzed carbenoid and nitrenoid transfers, reactions that are widely used to synthesize natural product intermediates and pharmaceuticals (*5*). The asymmetric cyclopropanation of olefins with high-energy carbene precursors (e.g., acceptor-substituted diazo reagents) is a hallmark reaction that generates up to three stereogenic centers in a single step to make the important cyclopropane motif, featured in many natural products and therapeutic agents (*6*). Limited to using physiologically accessible reagents, nature catalyzes intermolecular cyclopropane formation through wholly different strategies, typically involving olefin addition to the methyl cation of *S*-adenosyl methionine or through cyclization of dimethylallyl pyrophosphate-derived allylic carbenium ions (*7*). As a result, the diverse cyclopropanation products that can be formed by metallocarbene chemistry cannot be readily accessed by engineering natural cyclopropanation enzymes. We hypothesized that a natural metalloenzyme, the iron-heme-containing cytochrome P450, could be engineered to catalyze formal carbenoid transfers, thereby combining the high levels of regio- and stereoselectivity of enzymes with the synthetic versatility of carbene-based strategies.

Members of the cytochrome P450 enzyme family catalyze myriad oxidative transformations, including hydroxylation, epoxidation, oxidative ring coupling, heteroatom release, and heteroatom oxygenation (*8*). Most transformations encompassed by this broad catalytic scope manifest the reactivity of the same high-valent iron-oxene intermediate, compound I (figure 1). Inspired by the impressive chemo-, regio- and stereoselectivities with which cytochrome P450s can insert O atoms into C–H and C=C bonds, we investigated whether these enzymes could be engineered to mimic this chemistry for isoelectronic carbene transfer reactions via a high-valent iron-carbenoid species (figure 1). Here we report that variants of the cytochrome P450 from *Bacillus megaterium* (CYP102A1, or $P450_{BM3}$) are efficient catalysts for the asymmetric metallocarbene-mediated cyclopropanation of styrenes.

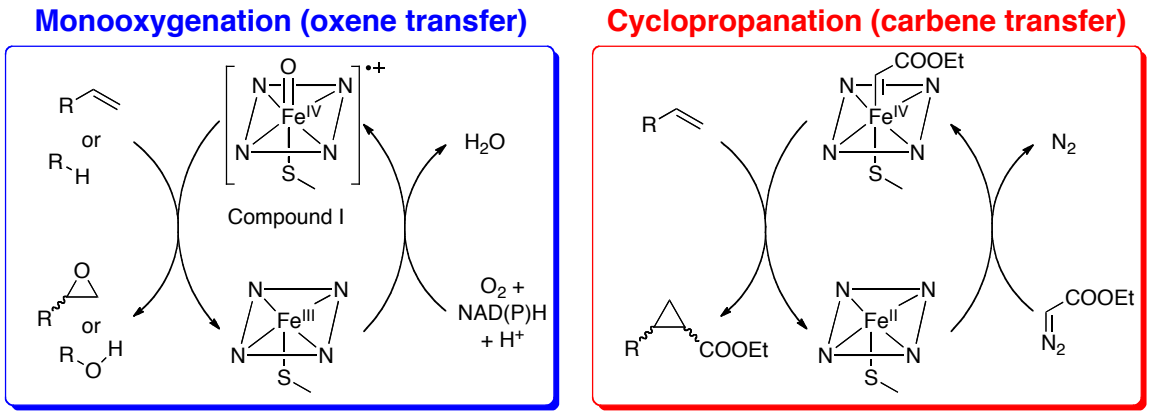


Figure 1. Canonical mode of reactivity of cytochrome P450s (left): monooxygenation of olefins and C-H bonds to epoxides and alcohols catalyzed by the ferryl porphyrin radical intermediate (Compound I). Artificial mode of formal carbene transfer activity of cytochrome P450s utilizing diazoester reagents as carbene precursors (right).

Because iron porphyrins catalyze carbene-based cyclopropanations (*9, 10*), we first probed whether some common heme proteins display measurable levels of cyclopropanation activity in aqueous media (phosphate buffer, with 5% methanol co-solvent). We chose the reaction between styrene and ethyl diazoacetate (EDA), (figure 2), a well-recognized model system for validating new cyclopropanation catalysts. Initial experiments showed that optimal formation of the desired cyclopropanation products occurred in the presence of a reducing agent (e.g., sodium dithionite, $Na_2S_2O_4$) under anaerobic conditions (tables S1-S4). Horseradish peroxidase (HRP), cytochrome *c* (cyt *c*), myoglobin (Mb), and $P450_{BM3}$ all displayed multiple turnovers toward the cyclopropane products, with HRP, cyt *c*, and Mb showing negligible enantioinduction, and formed the *trans* cyclopropane with over 90% diastereoselectivity, which is comparable to the diastereoselectivity induced by free hemin (table S1). $P450_{BM3}$, despite forming the cyclopropane products in low yield, catalyzed the reaction with different diasteroselectivity (*cis*:*trans* 37:63) and slight enantioinduction (table 1), showing that carbene transfer and selectivity are dictated by the heme cofactor bound in the enzyme active site.

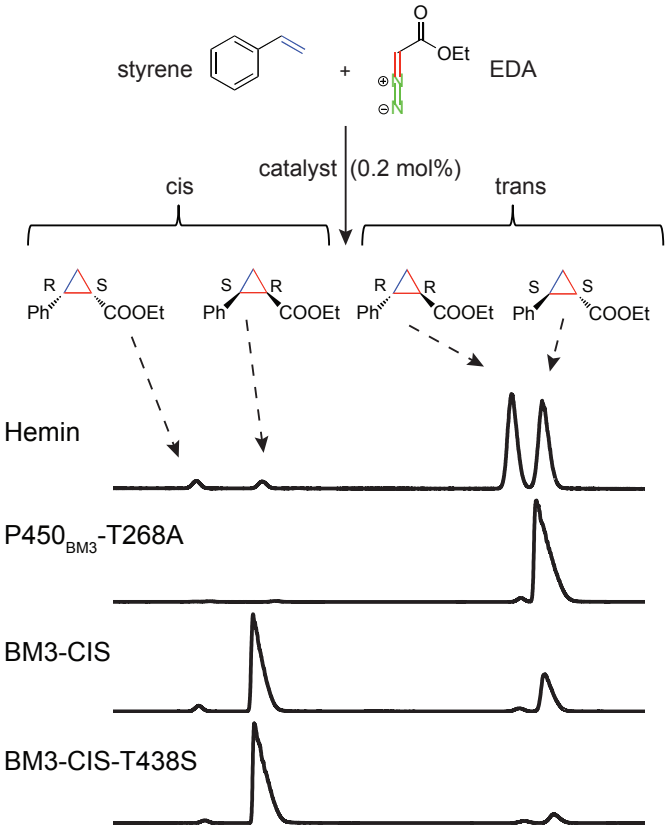


Figure 2. Absolute stereoselectivity of select $P450_{BM3}$ cyclopropanation catalysts. Reaction conditions: 20 μM catalyst, 30 mM styrene, 10 mM EDA, 10 mM $Na_2S_2O_4$, under argon in aqueous potassium phosphate buffer (pH 8.0) and 5% MeOH cosolvent for 2 hours at 298 K. Enzyme loading is 0.2 mol% with respect to EDA. The structures of each product stereoisomer are shown above the reaction gas chromatograms.

We then explored whether the activity and selectivity of heme-catalyzed cyclopropanation could be enhanced by engineering the protein sequence. $P450_{BM3}$ is a well-studied, soluble, self-sufficient (heme and diflavin reductase domains are fused in a single polypeptide, ~120 kD), long-chain fatty acid monooxygenase. More than a decade of protein engineering attests to the functional plasticity of this biocatalyst (*11*). From our work using directed evolution to engineer cytochrome $P450_{BM3}$ for synthetic applications

we have accumulated thousands of variants that exhibit monooxygenase activity on a wide range of substrates (*12*). We tested some of these variants for altered cyclopropanation diastero- and enantioselectivity by analysis of product distributions using gas chromatography (GC) with a chiral stationary phase. A panel of 92 $P450_{BM3}$ variants, chosen for diversity of activity and protein sequence, was screened in *Escherichia coli* lysate for the reaction of styrene and EDA under aerobic conditions in the presence of $Na_2S_2O_4$ (tables S5 and S6). The 10 most promising hits were selected for purification and characterization under standardized anaerobic reaction conditions (tables 1 and S7).

Five of the 10 selected P450s showed improvements in activity as compared to the wild type [total turnover numbers (TTNs) > 100], a comprehensive range of diastereoselectivities with *cis*:*trans* ratios varying from 9:91 to 60:40, and up to 95% enantioselectivities (table S7). For example, variant H2-5-F10, which contains 16 amino acid substitutions, catalyzes 294 total turnovers, equivalent to ~58% yield under these conditions (0.2% enzyme loading with respect to EDA). This represents a 50-fold improvement over wild-type $P450_{BM3}$. Furthermore, mutations affect both the diastereo- and enantioselectivity of cyclopropanation: H2-5-F10 favors the *trans* cyclopropanation product (*cis*:*trans* 16:84) with 63% enantiomeric excess (ee_{trans}), whereas H2A10, with a TTN of 167, shows reversed diastereoselectivity (*cis*:*trans* 60:40) with high enantioselectivity (95% ee_{cis}).

We used H2A10 to verify the role of the enzyme in catalysis and identify optimal conditions (table S8, figures S1 and S2). Heat inactivation produced diastereo- and

enantioselectivities similar to those obtained with free hemin, consistent with protein denaturation and release of the cofactor. Complete inhibition was achieved by preincubating the reaction mixture with carbon monoxide, which irreversibly binds the reduced P450 heme, confirming that catalysis occurs at the active site. Air inhibited the cyclopropanation reaction by about 50%, showing that dioxygen and EDA compete for reduced Fe^{II}. Cyclopropanation was also achieved with NADPH as the reductant, confirming that the activity can also be driven by the endogenous electron transport machinery of the diflavin-containing reductase domain. The presence of a reducing agent in substoichiometric amounts proved essential for cyclopropanation (table S9), implying that the active species is Fe^{II} rather than the resting state Fe^{III}.

Highly active $P450_{BM3}$ variants H2A10, H2-5-F10, and H2-4-D4 have three to five active-site alanine substitutions with respect to 9-10A-TS-F87V (12 mutations from $P450_{BM3}$, supplementary material text), which itself shows negligible cyclopropanation activity. These variants exhibit a range of TTNs, diastereoselectivity, and enantioselectivity (table 1). To better understand how protein sequence controls P450-mediated cyclopropanation, we constructed 12 variants to assess the contributions of individual alanines to catalysis and stability [table S10, (*13*)]. T268A is key for achieving high cyclopropanation activity, and this mutation alone converts inactive 9-10A-TS-F87V into an active cyclopropanation catalyst. Variant 9-10A-TS-F87V-T268A (here called BM3-CIS) is a competent cyclopropanation catalyst (199 TTNs), displays strong preference for the *cis* product (*cis*:*trans* 71:29), forms both diastereomers with over 90% *ee,* and is as stable as wild-type $P450_{BM3}$. Other active site alanine mutations tune the

product distribution. The addition of I263A to BM3-CIS reverses diastereoselectivity (*cis*:*trans* 19:81). We also investigated the effects of similar mutations introduced in the poorly active wild type $P450_{BM3}$ (table S11). $P450_{BM3}$-T268A, with a single mutation, is an active cyclopropanation catalyst (323 TTNs, table 1) with exquisite *trans*-selectivity (*cis*:*trans* 1:99) and high enantioselectivity for the major diastereomer (−96% ee_{trans}, figure 1). Whereas BM3-CIS is a *cis*-selective cyclopropanation catalyst, identical active site mutations in wild-type $P450_{BM3}$ result in a *trans*-selective enzyme (table S11), demonstrating that mutations outside of the active site also influence the stereochemical outcome.

Table 1. Stereoselective $P450_{BM3}$ cyclopropanation catalysts. Reactions were run in aqueous phosphate buffer (pH 8.0) and 5% MeOH cosolvent at room temperature under argon with 30 mM styrene, 10 mM EDA, 0.2 mol% catalyst (with respect to EDA), and 10 mM $Na_2S_2O_4$. Yields, diastereomeric ratios, and enantiomeric excess were determined by GC analysis. Yields based on EDA. TTN = total turnover number. *(*R*,*S*) – (*S*,*R*). †(*R*,*R*) – (*S*,*S*). See supplementary text for protein sequences indicating mutations from wild type $P450_{BM3}$.

Catalyst	% yield	TTN	*cis*:*trans*	%ee_{cis}*	%ee_{trans}†
Hemin	15	73	6:94	−1	0
$P450_{BM3}$	1	5	37:63	−27	−2
$P450_{BM3}$ -T268A	65	323	1:99	−15	−96
9-10A-TS-F87V	1	7	35:65	−41	−8
H2-5-F10	59	294	16:84	−41	−63
H2A10	33	167	60:40	−95	−78
H2-4-D4	41	206	53:47	−79	−33
BM3-CIS	40	199	71:29	−94	−91
BM3-CIS-I263A	38	190	19:81	−62	−91
BM3-CIS-A328G	37	186	83:17	52	−45
BM3-CIS-T438S	59	293	92:8	−97	−66

Because the design of *cis*-selective small-molecule catalysts for diazocarbonyl-mediated cyclopropanations has proven more challenging than that of their *trans* counterparts (*14*), we investigated whether active site engineering of $P450_{BM3}$ could provide robust *cis*-selective water-compatible catalysts to complement existing organometallic systems (*15*). We chose five active site residues (L181, I263, A328, L437, and T438) for individual site-saturation mutagenesis (*13*). The A328G, T438A, T438S and T438P variants exhibited enhanced *cis*-selectivity (table S12). A328G also reversed the enantioselectivity for the *cis*-diastereomer (table 1). BM3-CIS-T438S displayed the

highest diastereo- and enantioselectivities (*cis*:*trans* 92:8 and −97% ee_{cis}) and maintained TTNs comparable to those of BM3-CIS (table 1).

BM3-CIS exhibits Michaelis-Menten kinetics (figure S3 and table S13) with relatively high K_M values for the olefin (~1.5 mM) and the diazoester (~5 mM), reflecting the lack of evolutionary pressure for this enzyme to bind these substrates. BM3-CIS exhibits a notable k_{cat} for cyclopropanation of 100 min^{-1}, comparable to the k_{cat} of many native P450s for hydroxylation, but about 50 times less than $P450_{BM3}$-catalyzed fatty acid hydroxylation (table S14). Free hemin does not exhibit saturation kinetics and displays slower initial rates than BM3-CIS (only 30 min^{-1} at 10 mM styrene and 15 mM EDA), indicating that the protein scaffold enhances k_{cat} as compared to the free cofactor in solution. When used at 0.2 mole % (mol %) equivalent, BM3-CIS catalyzed cyclopropanations reached completion after 30 minutes. Adding more EDA enhanced turnovers for cyclopropanes and preserved BM3-CIS stereoselectivity (table S15), confirming catalyst integrity and implying that the reaction stops because of EDA depletion rather than inactivation.

To assess the substrate scope of $P450_{BM3}$-catalyzed cyclopropanation, we investigated the activities of six variants against a panel of olefins and diazo compounds (table 2 and tables S16 to S20). P450 cyclopropanation is robust to both electron-donating (*p*-vinylanisole and *p*-vinyltoluene) and electron-withdrawing (*p*-trifluoromethylstyrene) substitutions on styrene, and variant 7-11D showed consistent *cis*-selectivity for these substrates. The P450s were also active on 1,1-disubstituted olefins (i.e., α-methyl styrene), with chimeric P450 C2G9R1 forming cyclopropanes in

77% yield (with respect to EDA). The P450s were only moderately active with *t*-butyl diazoacetate as substrate (<30% yield), forming the *trans* product with >87% selectivity and offering no advantage over free hemin (table S20). For reactions involving EDA and aryl-substituted olefins, however, the P450s consistently outperformed the free cofactor in both activity and stereoselectivity.

Table 2. Scope of P450 catalyzed cyclopropanation of styrenyl substrates. Ar = *p*-X-C_6H_4. Reaction conditions: 20 μM catalyst, 30 mM olefin, 10 mM diazoester, 10 mM $Na_2S_2O_4$, under argon in aqueous potassium phosphate buffer (pH 8.0) and 5% MeOH cosolvent for 2 hours at 298 K. Enzyme loading is 0.2 mol % with respect to diazoester. N/A = not available when enantiomers did not separate to baseline resolution.

Reagents	P450 catalyst	TTN	*Z* : *E*	%*ee*$_Z$	%*ee*$_E$
R_1 = H, X = Me, R_2 = Et	BM3-CIS	228	78 : 22	-81	N/A
R_1 = H, X = OMe, R_2 = Et	H2-5-F10	364	11 : 89	38	N/A
R_1 = H, X = CF_3, R_2 = Et	7-11D	120	76 : 24	31	59
R_1 = Me, X = H, R_2 = Et	7-11D	157	41 : 49	42	N/A
R_1 = H, X = H, R_2 = *t*-Bu	H2A10	120	3 : 97	N/A	N/A

Screening natural enzymes against synthetic reagents chosen based on chemical intuition offers a simple strategy for identifying enzymes with basal levels of nonnative activity. As we have shown, a single mutation can be enough to promote such activity and achieve synthetically useful stereoselectivities. The accumulation of beneficial mutations by directed evolution or other protein engineering strategies can generate a spectrum of highly active catalysts for desired substrate and product specificities. The established reaction promiscuity of natural enzymes (*16, 17*) and the surprising ease with which cyclopropanation activity could be installed into $P450_{BM3}$ suggest that this approach will be useful for other synthetically important transformations for which biological counterparts do not yet exist.

* * *

Supplementary Materials for

Olefin Cyclopropanation via Carbene Insertion Catalyzed by Engineered Cytochrome P450 Enzymes

I. Materials and Methods

Unless otherwise noted, all chemicals and reagents for chemical reactions were obtained from commercial suppliers (Sigma-Aldrich, Acros) and used without further purification. The following heme proteins were all purchased from Sigma-Aldrich: myoglobin (from equine heart), peroxidase II (from horseradish), cytochrome c (from bovine heart), catalase (from *Corynebacterium glutamicum*), and chloroperoxidase (from *Caldariomyces fumago*). Silica gel chromatography purifications were carried out using AMD Silica Gel 60, 230-400 mesh. ^{1}H and ^{13}C NMR spectra were recorded on either a Varian Mercury 300 spectrometer (300 and 75 MHz, respectively), or a Varian Inova 500 MHz (500 and 125 MHz, respectively), and are internally referenced to residual solvent peak. Data for ^{1}H NMR are reported in the conventional form: chemical shift (δ ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad), coupling constant (Hz), integration. Data for ^{13}C are reported in terms of chemical shift (δ ppm) and multiplicity. High-resolution mass spectra were obtained with a JEOL JMS-600H High Resolution Mass Spectrometer at the California Institute of Technology Mass Spectral Facility. Reactions were monitored using thin layer chromatography (Merck 60 silica gel plates) using an UV-lamp for visualization. Optical rotation was measured using a JASCO P-2000 Polarimeter.

Gas chromatography (GC) analyses were carried out using a Shimadzu GC-17A gas chromatograph, a FID detector, and J&W scientific cyclosil-B columns (30 m × 0.32 mm, 0.25 μm film and 30 m × 0.25 mm, 0.25 μm film). High-performance liquid chromatography (HPLC) was carried out using an Agilent 1200 series, an UV detector,

and an Agilent XDB-C18 column (4.6 x 150 mm, 5 μm). Cyclopropane product standards for the reaction of ethyl diazoacetate (EDA) with styrene (ethyl 2-phenylcyclopropane-1-carboxylate) and α-methylstyrene (ethyl 2-methyl-2-phenylcyclopropane-1-carboxylate) were prepared as reported (*18*). These standards and enzyme-prepared cyclopropanes demonstrated identical retention times in gas chromatograms when coinjected, confirming product identity. Absolute stereoconfiguration of cyclopropane enantiomers was determined by measuring optical rotation of purified cyclopropane products from preparative bioconversion reactions using enantioselective P450-BM3 variants and referenced to values taken from reference (*19*). Authentic P450-catalyzed cyclopropane samples were also prepared as described in section VIII and were characterized by NMR (^{1}H and ^{13}C) and mass spectrometry.

Plasmids pCWori[BM3] and pET22 were used as cloning vectors. Site-directed mutagenesis was accomplished by standard overlap mutagenesis using primers bearing desired mutations (IDT, San Diego, CA). Primer sequences are available upon request. Electrocompetent *Escherichia coli* cells were prepared following the protocol of Sambrook *et al.* (*20*). Restriction enzymes BamHI, EcoRI, XhoI, Phusion polymerase, and T4 ligase were purchased from New England Biolabs (NEB, Ipswich, MA). Alkaline phosphatase was obtained from Roche (Nutley, NJ). The 1,000x trace metal mix used in expression cultures contained 50 mM $FeCl_3$, 20 mM $CaCl_2$, 10 mM $MnSO_4$, 10 mM $ZnSO_4$, 2 mM $CoSO_4$, 2 mM $CuCl_2$, 2 mM $NiCl_2$, 2 mM Na_2MoO_4, and 2 mM H_3BO_3.

II. General Procedures

Enzyme library screening. Libraries maintained in our laboratory are stored at -78 °C as glycerol stocks (Luria-Bertani medium (LB_{amp}), 150 µL, 25% v/v glycerol with 0.1 mg mL^{-1} ampicillin) in 96-well plates. These stocks were used to inoculate 96-shallow-well plates containing 300 µL LB_{amp} medium using a 96-pin stamp. Single colonies from site-saturation libraries were picked with toothpicks and used to inoculate 300 µL of LB_{amp}. The cells were incubated at 37 °C, 250 rpm shaking, and 80% relative humidity overnight. After 16 h, 50 µL aliquots of these overnight cultures were transferred into 2 mL, deep-well plates containing terrific broth (TB_{amp}) (800 µL containing 0.1 mg mL^{-1} ampicillin, 1 µL mL^{-1} trace metal mix and 20 mg L^{-1} aminolevulinic acid) using a Multimek 96-channel pipetting robot (Beckman Coulter, Fullerton, CA). The cultures were incubated at 37 °C for 3 h and 30 min, and 30 min after reducing the incubation temperature to 25 °C (250 rpm, 80% relative humidity), 50 µL isopropyl β-D-1-thiogalactopyranoside (IPTG, 4.5 mM in TB_{amp}) was added, and the cultures were allowed to continue for another 24 h at 25 °C (250 rpm, 80% relative humidity). Cells were then pelleted (3,000 × g, 15 min, 4 °C) and stored at -20 °C until further use, but at least for 2 h. For cell lysis, plates were allowed to thaw for 30 min at room temperature and then cell pellets were resuspended in 275 µL phosphate buffer (0.1 M, pH = 8.0, 0.65 mg mL^{-1} lysozyme, 10 mM magnesium chloride and 40 U mL^{-1} DNAse I). The lysing cells were incubated at 37 °C for 1 h. Cell debris was separated by centrifugation at 5,000 × g and 4 °C for 15 min. The resulting crude lysates were then transferred to 96-well microtiter plates for CO assays and to 2 mL deep-well plates for bioconversions.

CO binding assay. $Na_2S_2O_4$ (160 μL, 0.1 M in phosphate buffer, 0.1 M, pH = 8.0) was added to $P450_{BM3}$ variants in cell lysate (40 μL). The absorbance at 450 and 490 nm was recorded using a Tecan M1000 UV/Vis plate reader, and the microtiter plates were placed in a vacuum chamber. The chamber was sealed, evacuated to approximately - 15 in Hg, purged with CO gas, and incubated for 30 min. The plates were then removed and the absorbance at 450 and 490 nm was again recorded using a plate reader. The difference spectra could then be used to determine the P450 concentration in each well as previously described (*21*).

P450 expression and purification. For the enzymatic transformations, $P450_{BM3}$ variants were used in purified form. Enzyme batches were prepared as follows. One liter TB_{amp} was inoculated with an overnight culture (100 mL, LB_{amp}) of recombinant *E. coli* DH5α cells harboring a pCWori plasmid encoding the P450 variant under the control of the *tac* promoter. After 3.5 h of incubation at 37 °C and 250 rpm shaking (OD_{600} ca. 1.8), the incubation temperature was reduced to 25 °C (30 min), and the cultures were induced by adding IPTG to a final concentration of 0.5 mM. The cultures were allowed to continue for another 24 hours at this temperature. After harvesting the cells by centrifugation (4 °C, 15 min, 3,000 × g), the cell pellet was stored at -20 °C until further use but at least for 2 h. The cell pellet was resuspended in 25 mM Tris.HCl buffer (pH 7.5 at 25 °C) and cells were lysed by sonication (2x1 min, output control 5, 50% duty cycle; Sonicator, Heat Systems—Ultrasonic, Inc.). Cell debris was removed by centrifugation for 20 min at 4 °C and 27,000 × g and the supernatant was subjected to anion exchange

chromatography on a Q Sepharose column (HiTrapTM Q HP, GE Healthcare, Piscataway, NJ) using an AKTAxpress purifier FPLC system (GE healthcare). The P450 was eluted from the Q column by running a gradient from 0 to 0.5 M NaCl over 10 column volumes (P450 elutes at 0.35 M NaCl). The P450 fractions were collected and concentrated using a 30 kDa molecular weight cutoff centrifugal filter and buffer-exchanged with 0.1 M phosphate buffer (pH = 8.0). The purified protein was flash-frozen on dry ice and stored at -20 °C. P450 concentration was determined in triplicate using the CO binding assay described above (10 μL P450 and 190 μL 0.1 M phosphate buffer, pH 8.0, per well).

Thermostability measurements. Duplicate measurements were taken for all values reported in tables S10 and S11. Purified P450 solutions (4 μM, 200 μL) were heated in a thermocycler (Eppendorf) over a range of temperatures (38–65 °C) for 10 min followed by rapid cooling to 4 °C for 1 min. The precipitate was removed by centrifugation. The concentration of folded P450 remaining in the supernatant was measured by CO-difference spectroscopy (as described above). The temperature at which half of the protein was denatured (T_{50}) was determined by fitting the data to the equation: $f(T) = 100 / (1 + \exp(a*(T - T_{50})))$.

Typical procedure for small-scale cyclopropanation bioconversions under anaerobic conditions. Small-scale reactions (400 μL) were conducted in 2 mL crimp vials (Agilent Technologies, San Diego, CA). P450 solution (80 μL, 100 μM) was added to the vial

with a small stir bar before crimp sealing with a silicone septum. Phosphate buffer (260 μL, 0.1 M, pH = 8.0) and 40 μL of a solution of the reductant (100 mM $Na_2S_2O_4$, or 20 mM NADPH) were combined in a larger crimp-sealed vial and degassed by bubbling argon through the solution for at least 5 min (figure S1). In the meantime, the headspace of the 2 mL reaction vial with the P450 solution was made anaerobic by flushing argon over the protein solution (with no bubbling). When multiple reactions were conducted in parallel, up to 8 reaction vials were degassed in series via cannulae. The buffer/reductant solution (300 μL) was syringed into the reaction vial, while under argon. The gas lines were disconnected from the reaction vial before placing the vials on a plate stirrer. A 40x styrene solution in MeOH (10 μL, typically 1.2 M) was added to the reaction vial via a glass syringe, and left to stir for about 30 s. A 40x EDA solution in MeOH was then added (10 μL, typically 400 mM) and the reaction was left stirring for the appropriate time. The final concentrations of the reagents were typically: 30 mM styrene, 10 mM EDA, 10 mM $Na_2S_2O_4$, 20 μM P450.

The reaction was quenched by adding 30 μL HCl (3M) via syringe to the sealed reaction vial. The vials were opened and 20 μL internal standard (20 mM 2-phenylethanol in MeOH) was added followed by 1 mL ethyl acetate. This mixture was transferred to a 1.8 mL eppendorf tube which was vortexed and centrifuged (16,000xg, 1 min). The top organic layer was dried over an anhydrous sodium sulfate plug and analyzed by chiral phase GC.

A slightly modified workup was implemented for kinetic experiments. The reactions were quenched after the set time by syringing 1 mL EtOAc to the closed vials

and immediately vortexing the mixture. The vials were then opened and 20 μL internal standard was added. The mixture was transferred to a 1.8 mL eppendorf tube, vortexed and centrifuged (16,000 × g, 1 min). The top organic layer was dried over an anhydrous sodium sulfate plug and analyzed by GC.

Typical procedure for preparative-scale cyclopropanation bioconversions under anaerobic conditions. The P450 solution was added to a Schlenk flask with a stir bar. With the flask kept on ice, the headspace was evacuated and backfilled with argon (4×) with care not to foam the protein solution. Phosphate buffer and reductant were premixed and degassed together in a separate round-bottom flask by bubbling argon through the solution for 20 min. The buffer/reductant solution was transferred to the Schlenk flask via syringe. Styrene was added under argon and left to mix for 1 min. EDA was added dropwise under argon. The solution was left to stir under argon until reaction completion. The reaction was quenched under argon by adding hydrochloric acid (3 M) to adjust the pH to 4, before opening the Schlenk flask. The reaction mixture was stirred with sodium chloride and dichloromethane (CH_2Cl_2). The combined emulsion layers were then filtered through Celite to break the emulsion, and the Celite pad was rinsed with 3 × 20 mL CH_2Cl_2. The resulting biphasic mixture was transferred to a separating funnel and the organic phase was removed. The remaining aqueous phase was reextracted with 3 × 40 mL CH_2Cl_2. The combined organic extracts were dried with sodium sulfate, filtered, and concentrated. The resulting residue was purified by SiO_2 chromatography.

III. Preliminary Experiments with Heme Proteins

The following six heme proteins were initially screened for cyclopropanation activity: catalase, chloroperoxidase (CPO), horseradish peroxidase (HRP), cytochrome C (cyt *c*), myoglobin (Mb) and $P450_{BM3}$. Small-scale (400 µL) reactions were conducted as described in section II and were analyzed by GC (cyclosil-B 30 m x 0.25 mm x 0.25 µm): oven temperature = 130 °C.

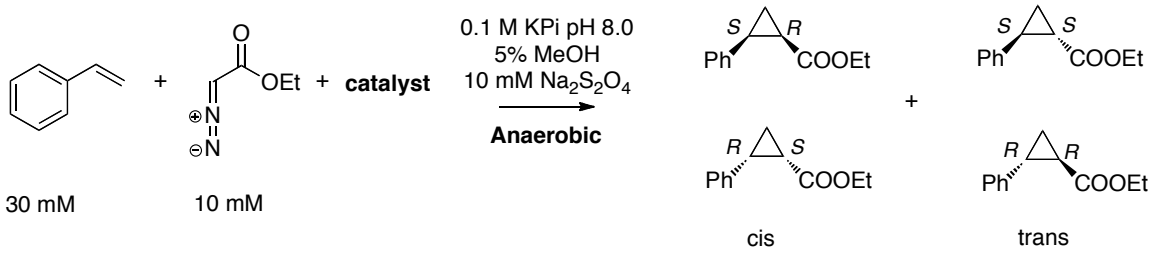


Table S1. Heme catalysts under **anaerobic** conditions **with sodium dithionite** ($Na_2S_2O_4$)

catalyst	axial ligand	cat. loading (% mol eq)	TTN	*cis*:*trans**	%ee *cis*[†]	%ee *trans*[‡]
catalase	O-Tyr	0.16	0	-	-	-
CPO[§]	S-Cys	0.40	0	-	-	-
HRP	N-His	1.00	9	7:93	8	−7
cyt *c*	N-His, S-Met	1.00	19	6:94	0	12
Mb	N-His	1.00	43	6:94	−1	2
$P450_{BM3}$	S-Cys	0.20	5	37:63	−27	−2
hemin	-	0.20	73	6:94	−1	0

* Diastereomeric ratios and enantiomeric excess were determined by GC analysis. [†] (*R*,*S*) – (*S*,*R*). [‡] (*R*,*R*) – (*S*,*S*). [§] Bioconversion conducted at 0.1 M citrate buffer pH = 4.0.

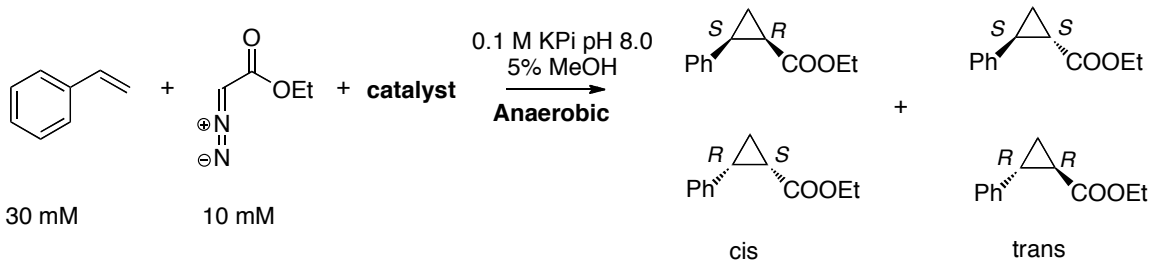


Table S2. Heme catalysts under **anaerobic** conditions **without $Na_2S_2O_4$**

catalyst	axial ligand	cat. loading (% mol eq)	TTN	*cis*:*trans**	%ee *cis*[†]	%ee *trans*[‡]
catalase	O-Tyr	0.16	0	-	-	-
CPO[§]	S-Cys	0.40	0	-	-	-
HRP	N-His	1.00	0	-	-	-
cyt *c*	N-His, S-Met	1.00	12	8:92	8	−3
Mb	N-His	1.00	0.8	11:89	−2	8
$P450_{BM3}$	S-Cys	0.20	0	0	-	-
hemin	-	0.20	4	11:89	−1	3

* Diastereomeric ratios and enantiomeric excess were determined by GC analysis. [†] (*R*,*S*) – (*S*,*R*). [‡] (*R*,*R*) – (*S*,*S*). [§] Bioconversion conducted at 0.1 M citrate buffer pH = 4.0.

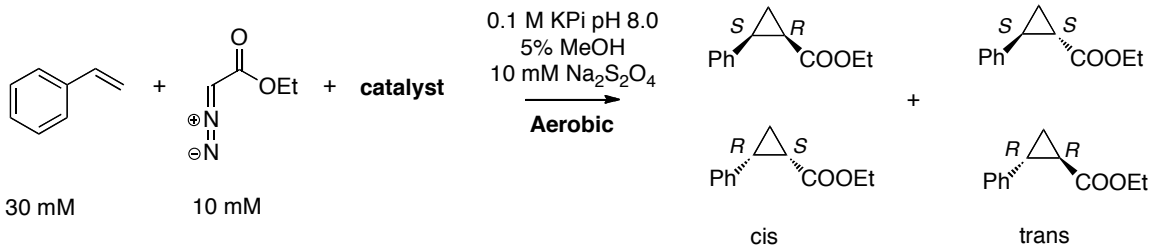


Table S3. Heme catalysts under **aerobic** conditions **with $Na_2S_2O_4$**

catalyst	axial ligand	cat. loading (% mol eq)	TTN	*cis*:*trans**	%ee *cis*[†]	%ee *trans*[‡]
catalase	O-Tyr	0.16	0	-	-	-
CPO[§]	S-Cys	0.40	0	-	-	-
HRP	N-His	1.00	1	12:88	−3	−7
cyt *c*	N-His, S-Met	1.00	3	9:91	−6	16
Mb	N-His	1.00	6	7:93	−13	12
$P450_{BM3}$	S-Cys	0.20	1	13:87	−38	−8
hemin	-	0.20	6	8:92	−5	1

* Diastereomeric ratios and enantiomeric excess were determined by GC analysis. [†] (*R*,*S*) – (*S*,*R*). [‡] (*R*,*R*) – (*S*,*S*). [§] Bioconversion conducted at 0.1 M citrate buffer pH = 4.0.

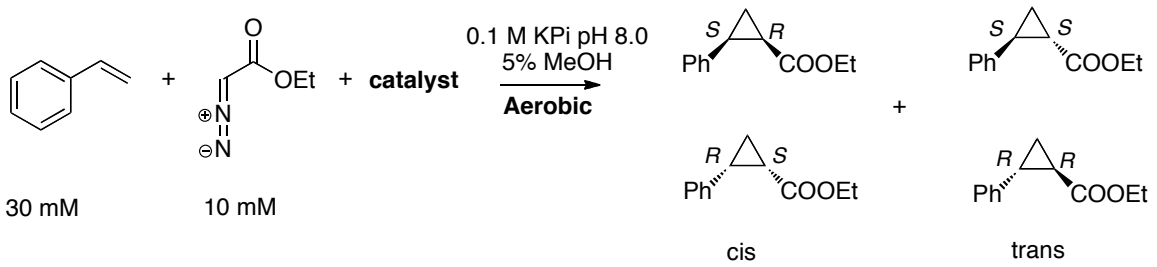


Table S4. Heme catalysts under **aerobic** conditions **without $Na_2S_2O_4$**

catalyst	axial ligand	cat. loading (% mol eq)	TTN	*cis*:*trans**	%ee *cis*[†]	%ee *trans*[‡]
catalase	O-Tyr	0.16	0	-	-	-
CPO[§]	S-Cys	0.40	0	-	-	-
HRP	N-His	1.00	0	-	-	-
cyt *c*	N-His, S-Met	1.00	0	-	-	-
Mb	N-His	1.00	0	-	-	-
$P450_{BM3}$	S-Cys	0.20	0.4	46:54	−46	36
hemin	-	0.20	0	-	-	-

* Diastereomeric ratios and enantiomeric excess were determined by GC analysis. [†] (*R*,*S*) – (*S*,*R*). [‡] (*R*,*R*) – (*S*,*S*). [§] Bioconversion conducted at 0.1 M citrate buffer pH = 4.0.

IV. Screening $P450_{BM3}$ Variants for Cyclopropanation Activity

Lysate screening under aerobic conditions. The 92 variants in our compilation plate (table S5, following pages) represent a diverse selection of $P450_{BM3}$ variants that have previously been engineered for monooxygenase activity on a variety of substrates, including but not limited to short alkane hydroxylation, demethylation of protected monosaccharides, and oxidation of lead drug compounds. These $P450_{BM3}$ variants carry various mutations accumulated along sequential rounds of engineering efforts for activity towards the target substrates (table S5) or were generated by recombination with homologous enzymes (table S6). The compilation plate was expressed and lysed as described in section II (**enzyme library screening**). Lysate (150 μL) was transferred (Multimek 96-channel pipetting robot, Beckman Coulter, Fullerton, CA) to a 2 mL deep-well plate, with 50 μL of 120 mM $Na_2S_2O_4$ in 0.1 M KPi (pH = 8.0). 100 μL of a 30 mM styrene, 60 mM EDA mixed solution in 15% MeOH in 0.1 M KPi (pH = 8.0) was added to the plate to initiate the reaction. The plate was sealed and was left shaking (300 rpm) for four hours. The plastic seal was removed and 30 μL HCl (3 M) was added to quench the reaction followed by 20 μL of an internal standard solution (20 mM α-methylstyrene in methanol). The reactions were extracted by adding 500 μL EtOAc and carefully vortexing the plate. The plate was centrifuged (1,700 × g) to separate the biphasic mixture. The top organic layer was transferred (2 × 150 μL) to a separate deep-well plate. The extracts for each of the 92 reactions were dried through 92 separate anhydrous sodium sulfate plugs. The dried extracts were analyzed by GC (cyclosil-B 30 m × 0.32 mm × 0.25 μm): oven temperature = 60 °C 3 min, 7.5 °C / min to 160 °C,

20 °C / min to 250 °C, 250 °C 2 min, *cis*-cyclopropanes (20.3 min and 20.45 min), *trans*-cyclopropanes (21.8 min). The top 10 protein variants of importance with respect to this report are highlighted in tables S5 and S6.

Table S5. Raw data from $P450_{BM3}$ compilation plate screen

Diastereo- and enantioselectivity were determined by gas chromatography using a chiral β-CDX column as the stationary phase.

$P450_{BM3}$ variant	**mutations with respect to wild-type $P450_{BM3}$**	**absolute activity***	**de**[†]	**ee (*cis*)** [‡]
CYP102A3	N/A	0.004053	-74	-8
CYP102A2	N/A	0.002963	-76	-36
CYP102A1 ($P450_{BM3}$)	None	0.002240	-81	7
WT F87A	F87A	0.001704	-28	57
WT T88L	T88L	0.004522	-78	23
WT A328V	A328V	0.000830	-100	N/A
J (*22*)	V78A, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, L353V	0.001334	-100	N/A
139-3 (*23*)	V78A, H138Y, T175I, V178I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, L353V	0.001386	-86	0
9-10A (*22*)	R47C, V78A, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, L353V	0.004292	-74	-20
9-10A L75W (*24*)	**9-10A** L75W	0.005191	-83	-8
9-10A L75I (*24*)	**9-10A** L75I	0.002267	-85	-3
9-10A A78F (*24*)	**9-10A** L78F	0.002008	-82	-35
9-10A A78S (*24*)	**9-10A** A78S	0.005098	-81	-6
9-10A A82G (*24*)	**9-10A** A82G	0.002245	-76	-7
9-10A A82F (*24*)	**9-10A** A82F	#VALUE!	N/A	N/A
9-10A A82C (*24*)	**9-10A** A82C	0.002487	-74	16
9-10A A82I (*24*)	**9-10A** A82I	0.001031	-100	N/A
9-10A A82S (*24*)	**9-10A** A82S	0.001483	-82	14
9-10A A82L (*22*)	**9-10A** A82L	0.000591	-100	N/A
9-10A F87A	**9-10A** F87A	0.001701	-61	-10
9-10A F87V (*24*)	**9-10A** F87V	0.000000	N/A	N/A
9-10A F87I (*24*)	**9-10A** F87I	0.000983	-100	N/A
9-10A F87L (*24*)	**9-10A** F87L	0.000710	-100	N/A
9-10A T88C (*24*)	**9-10A** T88C	0.002516	-77	3
9-10A T260S	**9-10A** T260S	0.004259	-82	-6

(*24*)				
9-10A T260N (*24*)	**9-10A** T260N	0.003882	-77	15
9-10A T260L (*24*)	**9-10A** T260L	0.006173	-77	-2
9-10A A328V (*22*)	**9-10A** A328V	0.006471	-68	-8
9-10A A328M (*24*)	**9-10A** A328M	0.005180	-82	6
9-10A A328F (*24*)	**9-10A** A328F	0.002009	-63	-32
49-1A	R47C, V78T, A82G, F87V, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, A328L, L353V	0.001874	-75	-32
35-7F	R47C, V78F, A82S, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, A328L, L353V	0.004514	-73	-52
53-5H (*24*)	**9-10A** A78F, A82S, A328F	0.002840	-80	2
7-11D	**9-10A** A82F, A328V	0.036840	-24	-28
49-9B	R47C, V78A, A82G, F87V, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, A328L, L353V	0.000000	N/A	N/A
41-5B	R47C, V78F, A82G, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, A328V, L353V	0.008391	-77	-17
13-7C (*24*)	**9-10A** A78T, A328L	0.005493	-73	-43
12-10C	R47C, V78A, A82G, F87V, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, A328V, L353V	0.004566	-73	-21
77-9H (*24*)	**9-10A** A78T, A82G, A328L	0.003053	-73	-34
11-8E	R47C, V78A, F87V, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, A328L, L353V	0.001453	-77	15
1-12G (*22*)	**9-10A** A82L, A328V	0.003884	-70	-19
29-3E	R47C, V78A, A82F, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, A328F, L353V	0.003425	-80	15
29-10E	R47C, V78F, A82G, K94I,	0.001935	-70	16

	P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, A328F, L353V			
68-8F (*24*)	**9-10A** A78F, A82G, A328L	0.004127	-72	-32
35E11 (*25*)	R47C, V78F, A82S, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, A328F, L353V, E464G, I710T	0.003600	-71	-14
19A12 (*25*)	**35E11** L52I, L188P, I366V	0.006909	-70	-27
ETS8 (*25*)	**35E11** L52I, I366V	0.003966	-79	-19
(11-3) (*25*)	**35E11** L52I, A74S, L188P, I366V	0.005633	-76	-39
(7-7) (*25*)	**35E11** L52I, A74E, S82G, A184V, L188P, I366V	0.010499	-77	-9
H2A10	**9-10A TS** F87V, L75A, L181A, T268A	0.066422	-8	-94
SL2-6F8	R47C, L52I, V78F, A82S, K94I, P142S,T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, A328L, K349N, L353V, I366V, E464G, I710T	0.000778	-100	N/A
A12SL-17-4	R47C, L52I, A74E, V78F, A82S, K94I, P142S,T175I, A184V, L188P, F205C, S226R, H236Q, E252G, R255S, A290V, A328F, L353V, I366V, E464G, I710T	0.010935	-80	6
H2-2-A1 (*26*)	**9-10A TS** F87V, L75A, L181A, L437A	0.003042	-75	-11
A12RM-2-8	R47C, L52I, A74E, V78F, A82S, K94I, P142S,T175I, A184S, L188P, F205C, S226R, H236Q, E252G, R255S, A290V, A328F, L353V, I366V, E464G, I710T	0.007705	-77	-13
H2-5-F10	**9-10A TS** F87V, L75A, I263A, T268A, L437A	0.141237	-46	-56
13C9R1	L52I, I58V, L75R, F87A, H100R, S106R, F107L, A135S, A184V, N239H, S274T, L324I, V340M, I366V, K434E, E442K, V446I	0.001980	-100	N/A
22A3	13C9R1 F162I E434K K442E I446V	0.004053	-70	4
2C6 (*27*)	**9-10A** A78L, F87A, V184T, G315S, A330V	0.004257	-78	-15
9C7 (*27*)	**9-10A** C47R, A78L, F87G, I94K, A180V, V184T, G315S, A330V, Y345C	0.007258	-79	-5

B1 (*27*)	**9-10A** C47R, A78L, F87A, I94K, V184T, I263M, G315S, A330V	0.002246	-61	-14
B1SYN (*27*)	**9-10A** C47S, N70Y, A78L, F87A, I174N, I94K, V184T, I263M, G315S, A330V	0.002705	-76	-23
H2-4-D4	**9-10A TS** F87V, L75A, M177A, L181A, T268A, L437A	0.052439	57	-84
E12 A87V (*27*)	**9-10A** C47R, A78L, F87V, I94K, A111V, V141I, A180V, V184T, G315S, A330V	0.001990	-65	-52
GlcA4 T180A	**9-10A** C47R, F81W, A82S, F87A, I94K	0.004925	-78	12
H2-8-C7 (*26*)	**9-10A TS** F87V, L75A, L181A	0.000808	-100	N/A
CH-F8	**9-10A** L51A, C47A, F87V, I94K, L181A, C205F, S254R, I366V, L437A, E442K	0.001126	-100	N/A
H2-4-H5 (*26*)	**9-10A TS** F87V, L75A, M177A, L181A	0.001229	-100	N/A
SA9	**9-10A** C47R, F81W, A82I, F87A, I94K, A180T, A197V	0.004170	-81	11
ManA10	**9-10A** C47R, F81S, A82V, F87A, I94K, A180T, A197V	0.006340	-82	14
Man1	**9-10A** C47R, F81L, A82T, F87A, I94K	0.003053	-73	21
MB2	**9-10A** C47R, F81W, A82I, F87A, I94K	0.003282	-77	10
HA62	**9-10A** C47R, F81A, A82L, F87A, I94K	0.003375	-81	-5
9-10A TS	V78A, P142S, T175I, A184V, S226R, H236Q, E252G, A290V, L353V, I366V, E442K	0.001920	-75	-54
9-10A TS F87A	**9-10A TS** F87A	0.001546	-60	5
25F7	**9-10A** C47R, A74F, A78S, F87A, I282K, C205F, S255R	0.001829	-81	43
24C4	**9-10A** C47R, A74I, A78L, F87A, I94K, C205F, S255R	0.000783	-100	N/A
5A1	**9-10A** M30T, C47R, A74F, A78S, I94K, C205F, S255R, Q310L, I366V, E442K	0.002471	-80	15
8B3	**9-10A** M30T, C47R, A74F, A78S, I94K, C205F, C255R, L310Q, Q323L, I366V, N381K, R398H, E441K	0.001315	-100	N/A

* Reported as the sum of the area of the cyclopropane peaks over the area of the internal standard. † Diastereomeric excess = ([cis]-[trans])/([cis]+[trans]). ‡ (*R*,*S*) – (*S*,*R*).

Table S6. Raw GC screening data for the chimeric P450s in the compilation plate

P450	chimeric P450s (heme domain block sequence)	absolute activity*	de†	ee (*cis*) ‡
CYP102A1 ($P450_{BM3}$) F87A (*28*)	11111111	0.001704	-28	56
CYP102A2 F88A (*28*)	22222222	N/A	N/A	N/A
CYP102A3 F88A (*28*)	33333333	N/A	N/A	N/A
5R1 (*29*)	32312231	0.008625	58	19
9R1 (*29*)	12112333	0.0042707	58	24
12R1 (*29*)	12112333	0.0701514	32	-49
C1D11R1 (*29*)	21113312	0.007138	51	9
C2B12R1 (*29*)	32313233	0.005914	38	-5
C2C12R1 (*29*)	21313111	0.006226	28	9
C2E6R1 (*29*)	11113311	0.008731	25	6
C2G9R1 (*29*)	22213132	0.007975	15	31
C3D10R1 (*29*)	22132231	0.004898	-16	-2
C3E4R1 (*29*)	21313311	0.007893	14	17
F3H12R1 (*29*)	21333233	0.005586	-56	-17
F6D8R1 (*29*)	22313233	0.008088	-76	-6
C3B5R1 (*29*)	23132233	0.014722	-81	4
X7R1 (*29*)	22312333	0.017305	-4	-34

* Reported as the sum of the area of the cyclopropane peaks over the area of the internal standard. † Diastereomeric excess = ([cis]-[trans])/([cis]+[trans]). ‡ (*R*,*S*) – (*S*,*R*).

Determining the cyclopropanation activity of the top 10 hits (highlighted in yellow on tables S5-6) under anaerobic conditions. Small-scale reactions (400 μL total volume) were conducted as described in section II and were analyzed by GC (cyclosil-B 30 m x 0.32 mm x 0.25 μm): oven temperature = 100 °C 5 min, 1 °C / min to 135 °C, 135 °C 10 min, 10 °C / min to 200 °C, 200 °C 5 min, *cis*-cyclopropanes (39.40 min and 40.20 min), *trans*-cyclopropanes (44.69 min and 45.00 min).

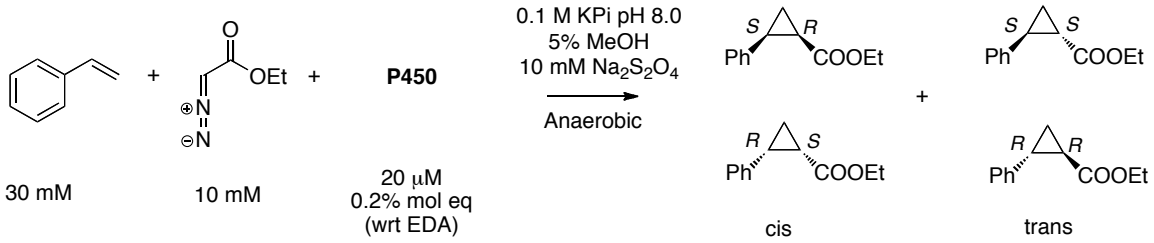


Table S7. Stereoselective $P450_{BM3}$-based cyclopropanation catalysts

P450	% yield*	TTN	*cis*:*trans*†	%ee *cis*‡	%ee *trans*§
WT	1	5	37:63	−10	−9
WTF87A	1.2	6	37:63	26	−6
H2A10	33.4	167	60:40	−95	−78
H2-4-D4	41.2	206	53:47	−79	−33
H2-5-F10	58.8	294	16:84	−41	−63
C2C12R1	1.6	8	36:64	45	1
C3E4R1	1.6	8	43:57	51	−7
X7R1	2.4	12	33:67	23	−4
12 R1	6.2	31	17:83	9	−2
C2E6 R1	4.6	23	27:73	25	−6
C2G9 R1	48	240	9:91	10	−2
7-11D	32	160	35:65	−22	−18

* based on EDA. † Diastereomeric ratios and enantiomeric excess were determined by GC analysis. ‡ (*R*,*S*) – (*S*,*R*). § (*R*,*R*) – (*S*,*S*).

V. Experimental Characterization of $P450_{BM3}$ Cyclopropanation Catalysts

V.I. Controls to confirm the enzymatic cyclopropanation activity of variant H2A10. Small-scale reactions (400 μL total volume) were set up and worked up as described in section II. For the carbon monoxide (CO) inhibition experiment, the reaction vial and the buffer/reductant vial were purged with CO after having been purged with argon. For the boiled P450 experiment, a 100 μM solution of variant H2A10 was heated at 60 °C for 10 min. For the hemin experiment, hemin (80 μL) was added from a 1 mM solution in 50% DMSO-H2O, such that its final concentration in the reaction was 200 μM. Complete System = 10 mM styrene, 20 mM EDA, 20 mM $Na_2S_2O_4$, 20 μM P450 (H2A10) under anaerobic conditions. The dried ethyl acetate extracts were analyzed by chiral phase GC, using 2-phenylethanol as an internal standard (injector temperature = 300 °C, oven temperature = 100 °C for 5 min, 1 °C / min ramp up to 135 °C, 135 °C for 10 min, 10 °C / min ramp up to 200 °C, 200 °C for 5 min). Elution time: *cis*-cyclopropanes (39.40 min and 40.20 min), *trans*-cyclopropanes (44.69 min and 45.00 min).

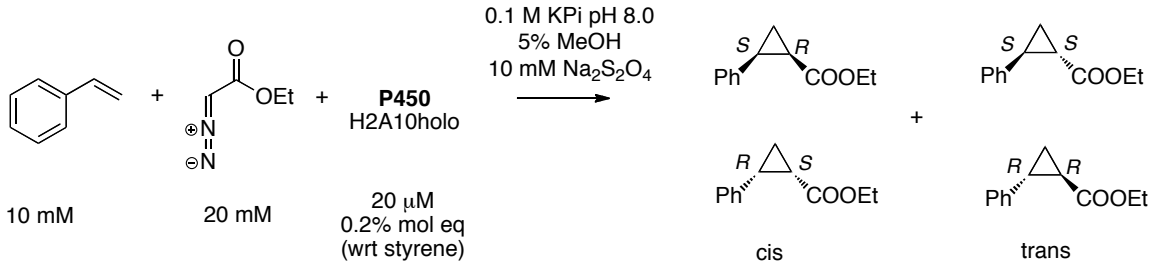


Table S8. Controls for P450 based cyclopropanation using variant H2A10

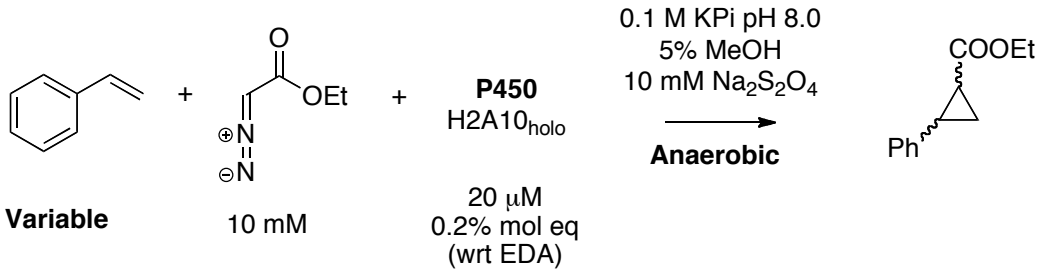
conditions	TTN	% inhibition	*cis*:*trans**	% ee *cis*[†]	% ee *trans*[‡]
complete system (CS)	101	-	70:30	−95	−78
CS-$Na_2S_2O_4$+NADPH	45	−55	61:39	−87	−31
CS-$Na_2S_2O_4$+NADH	38	−62	53:47	−76	−19
CS-$Na_2S_2O_4$	0	−100	-	-	-
CS-P450	0	−100	-	-	-
CS+CO	0	−100	-	-	-
boiled P450	146	+45	16:84	2	−2
$H2A10_{heme}$	85	−16	67:33	−92	−67
CS-P450+hemin	16	−84	15:85	−1	−2
CS (aerobic)	43	−57	67:33	−94	−76

* Diastereomeric ratios and enantiomeric excess were determined by GC analysis. [†] (*R*,*S*) – (*S*,*R*). [‡] (*R*,*R*) – (*S*,*S*).

V.II. Optimizing cyclopropanation reaction conditions for variant H2A10. Small-scale reactions (400 μL final volume) were set up and worked up as described in section II. The dried ethyl acetate extracts were analyzed by chiral phase GC, using 2-phenylethanol as an internal standard (injector temperature = 300 °C, oven temperature = 100 °C for 5 min, 5 °C / min ramp up to 200 °C, 20 °C / min ramp up to 250 °C, 250 °C for 5 min). Elution time: *cis*-cyclopropanes (19.20 min and 19.33 min), *trans*-cyclopropanes (20.44 min). The reaction conditions that gave optimal yields of

cyclopropanes (with respect to EDA) were: 30 mM styrene, 10 mM EDA and 20 μM P450 and were used in subsequent experiments.

V.II.I. Styrene concentration



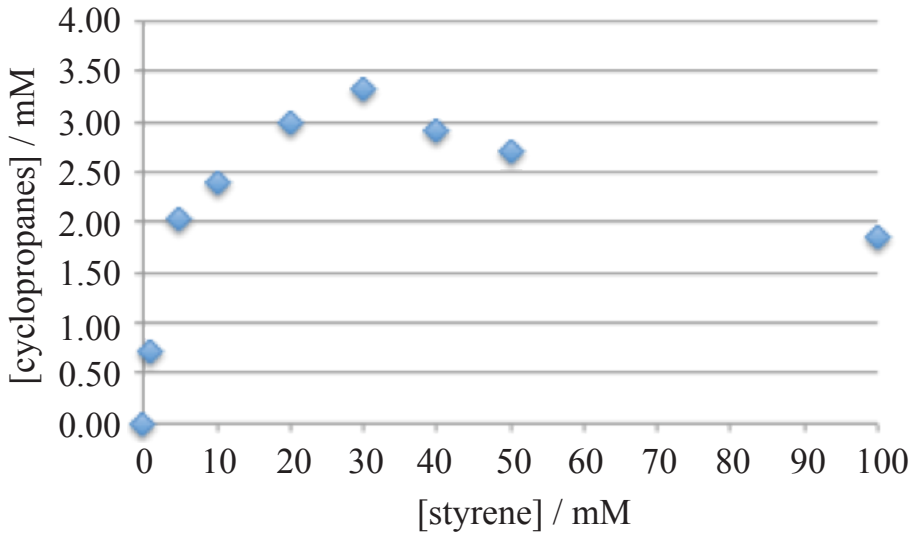
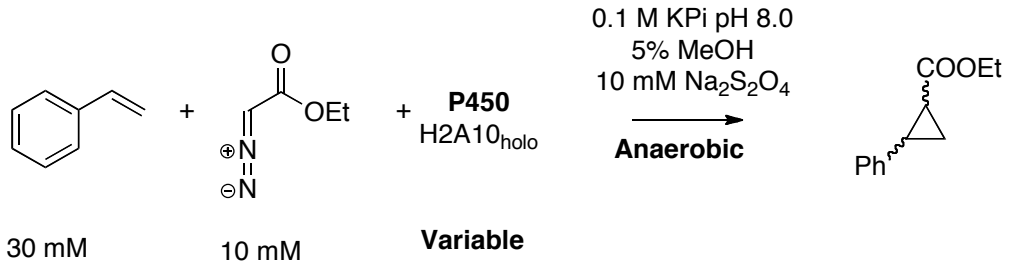


Figure S1. Effect of styrene concentration on cyclopropane yield.

V.II.II. P450 concentration



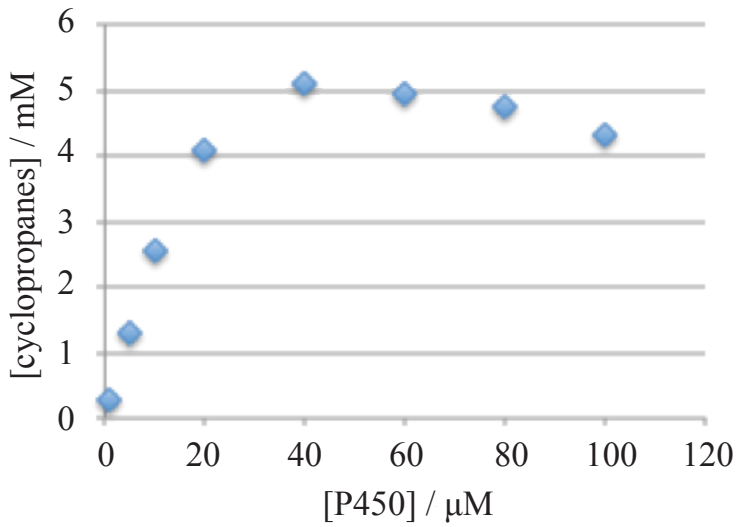


Figure S2. Effect of P450 (H2A10) concentration on cyclopropane yield

V.II.III. Dithionite concentration

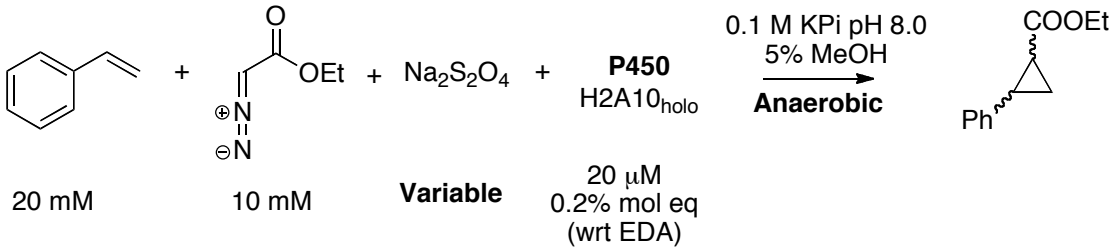


Table S9. Effect of concentration of $Na_2S_2O_4$ on cyclopropane yield

$[Na_2S_2O_4]$ / mM	[cyclopropanes] / mM	TTN
0	0	0
1	2.59	129
5	2.72	136
10	3.34	167
20	3.13	156
50	2.79	140
100	2.71	136

V.III. Mutational analysis of active site alanine substitutions in 9-10A TS F87V

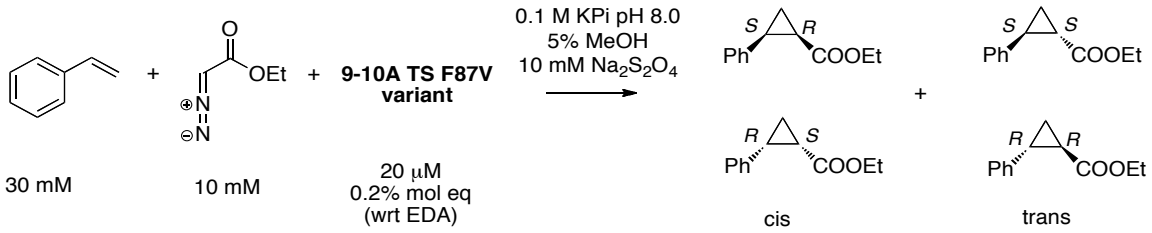


Table S10. Mutational analysis of alanine substitutions on 9-10A TS F87V

Mutations relative to 9-10A TS F87V

P450 (Holo)	L75A	M177A	L181A	I263A	T268A	L437A	TTN	*cis*:*trans**	%ee *cis*†	%ee *trans*‡	T_{50} (°C)
9-10A TS F87V	No	No	No	No	No	No	7	35:65	−41	−8	59.5
9-10A TS F87V L75A	Yes	No	No	No	No	No	5	42:58	−59	−11	52.3
9-10ATS F87V L181A	No	No	Yes	No	No	No	5	41:59	−27	−7	53.3
9-10A TS F87V I263A	No	No	No	Yes	No	No	8	29:71	−31	−39	55.4
9-10A TS F87V T268A (BM3-CIS)	No	No	No	No	Yes	No	199	71:29	−94	−91	55.2
BM3-CIS I263A	No	No	No	Yes	Yes	No	190	19:81	−62	−91	54.0
BM3-CIS L181A	No	No	Yes	No	Yes	No	159	56:44	−92	−94	50.8
H2A10	Yes	No	Yes	No	Yes	No	167	60:40	−95	−78	48.9
BM3-CIS L181A I263A	No	No	Yes	Yes	Yes	No	203	14:86	−46	−95	50.9
BM3-CIS L181A L437A	No	No	Yes	No	Yes	Yes	180	27:73	−74	−98	48.4
BM3-CIS L181A I263A L437A	No	No	Yes	Yes	Yes	Yes	218	9:91	−55	−96	48.2
4H5	Yes	Yes	Yes	No	No	No	7	32:68	−9	0	49.4
BM3-CIS I263A L437A	No	No	No	Yes	Yes	Yes	267	16:84	−59	−89	50.4
H2-5-F10	Yes	No	No	Yes	Yes	Yes	294	16:84	−41	−63	47.5
H2-4-D4	Yes	Yes	Yes	No	Yes	Yes	206	53:47	−79	−33	46.4

* Diastereomeric ratios and enantiomeric excess were determined by GC analysis. † (*R*,*S*) – (*S*,*R*). ‡ (*R*,*R*) – (*S*,*S*).

V. IV. Sequential introduction of BM3-CIS active site mutations in wild-type $P450_{BM3}$

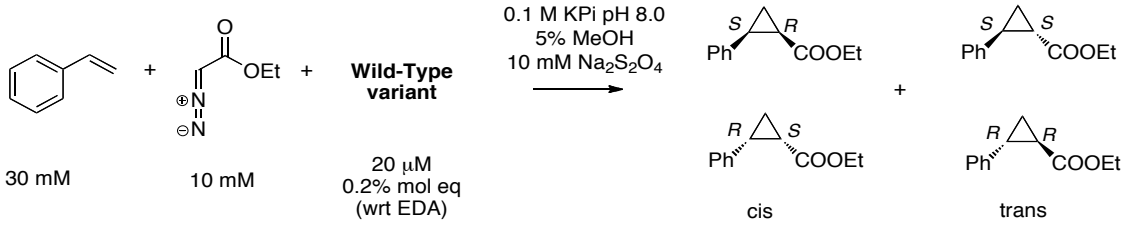


Table S11. Introducing BM3-CIS related active site mutations in wild-type $P450_{BM3}$

	Mutations relative to wild-type $P450_{BM3}$								
P450 (Holo)	**V78**	**F87**	**T268**	**I263**	**TTN**	***cis*:*trans****	**%ee *cis*†**	**%ee *trans*‡**	**T_{50} (°C)**
WT	-	-	-	-	5	37:63	−10	−9	56.0
WT-F87A	-	A	-	-	6	38:62	26	−6	53.0
WT-F87V	-	V	-	-	9	30:70	−33	−26	52.9
WT-T268A	-	-	A	-	323	1:99	−15	−96	53.6
WT-F87V/T268A	-	V	A	-	274	32:68	−77	−99	52.0
WT-V78A/F87V/T268A	A	V	A	-	190	32:68	−70	−20	50.8
WT-F87V/I263A/T268A	-	V	A	A	246	7:93	8	−94	50.0

* Diastereomeric ratios and enantiomeric excess were determined by GC analysis. † (*R*,*S*) – (*S*,*R*). ‡ (*R*,*R*) – (*S*,*S*).

VI. Active Site Saturation Mutagenesis of BM3-CIS$_{heme}$

Library construction. To simplify library construction and screening, only the BM3-CIS heme domain, which comprises residues 1-462 was used. This truncated enzymes lacks the P450 native reductase and exhibits similar activity and stereochemical control to the full length enzyme using $Na_2S_2O_4$ as a reductant, but not NADPH (data not shown). P450 site-directed mutagenesis and site-saturation libraries were assembled from PCR fragments generated from oligonucleotides containing the desired codon mutation or a degenerate NNK (or for reverse primers, the reverse complement MNN; where N = A,T,G,C, K = G,T and M = A,C) codon, which codes for all 20 amino acids and the TAG stop codon. PCR fragments were assembled using either standard overlap extension PCR or through restriction cloning using the Type IIS restriction enzyme, BsaI, depending on convenience.

Lysate screening under aerobic conditions. The compilation plate was expressed and lysed as described in section II (**enzyme library screening**). 150 μL lysate was transferred (Multimek 96-channel pipetting robot, Beckman Coulter, Fullerton, CA) to a 2 mL deep-well plate, with 50 μL of 120 mM $Na_2S_2O_4$ in 0.1 M KPi (pH = 8.0). 100 μL of a 90 mM styrene, 30 mM EDA mixed solution in 15% MeOH in 0.1 M KPi (pH = 8.0) was added to the plate to initiate the reaction. The plate was sealed and was left shaking (300 rpm) for four hours. The plastic seal was removed and 30 μL HCl (3 M) was added to quench the reaction followed by 20 μL of an internal standard solution (20 mM 2-phenylethanol in methanol). Acetonitrile (400 μL) was added before carefully vortexing

the plate. The plate was centrifuged (1,700 × g), the supernatant was filtered (1 μm glass, 96 well filter plate, Pall) and transferred (150 μL) to a 96-well microtiter plate (Agilent). Reactions were analyzed by reverse-phase HPLC (210 nm): 50% acetonitrile-water, 1.0 mL min^{-1}, *cis*-cyclopropanes (7.6 min), *trans*-cyclopropanes (9.7 min). Hits were selected based on enhancement of *cis*-selectivity over parent BM3-CIS.

Determining the cyclopropanation activity of hits from the site-saturation libraries under anaerobic conditions. Small-scale reactions (400 μL total volume) were conducted as described in section II and were analyzed by GC (cyclosil-B 30 m x 0.25 mm x 0.25 μm): oven temperature = 130 °C, 175 kPa, *cis*-cyclopropanes (39.40 min and 40.20 min), *trans*-cyclopropanes (44.69 min and 45.00 min).

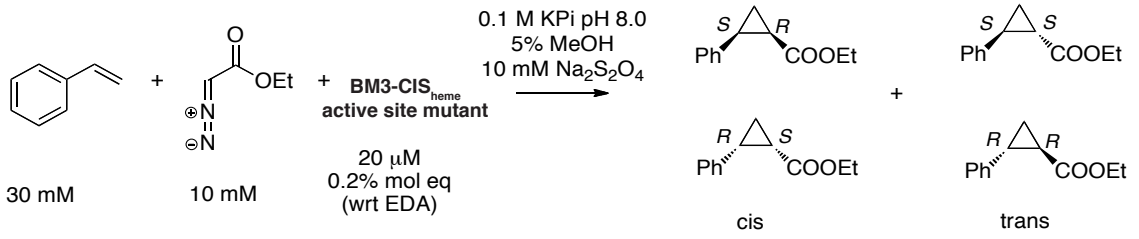


Table S12. Cyclopropanation activity of selected BM3-CIS$_{heme}$ active site variants

P450$_{heme}$	yield (%)*	TTN	*cis*:*trans*[†]	%ee *cis*[‡]	%ee *trans*[§]
BM3-CIS	57	286	71:29	−92	−88
BM3-CIS-L181G	47	234	59:41	−89	−90
BM3-CIS-A328G	37	186	83:17	52	−45
BM3-CIS-L437F	53	265	53:47	−82	−85
BM3-CIS-L437Q	30	148	53:47	−73	−87
BM3-CIS-L437G	58	290	54:46	−88	−91
BM3-CIS-L437A	39	194	38:62	−84	−11
BM3-CIS-T438A	54	273	91:9	−92	−75
BM3-CIS-T438G	15	78	73:27	−87	−59
BM3-CIS-T438S	59	293	92:8	−97	−66
BM3-CIS-T438Q	41	206	38:62	67	70
BM3-CIS-T438P	32	161	90:10	−91	−50

* based on EDA. [†] Diastereomeric ratios and enantiomeric excess were determined by GC analysis. [‡] (*R*,*S*) – (*S*,*R*). [§] (*R*,*R*) – (*S*,*S*).

VII. Kinetic Characterization of BM3-CIS

Determination of initial rates. Both styrene and EDA concentrations were varied in the presence of the P450s expressed as the heme-domain (0.5 or 1.0 μM BM3-CIS$_{heme}$). Reactions were set up in phosphate buffer (pH = 8.0) with $Na_2S_2O_4$ as the reductant at 298 K, and were worked up as described in section II. Three time points were taken and used to determine the rate of product formation by GC (cyclosil-B 30 m x 0.32 mm x 0.25 μm): oven temperature = 100 °C 5 min, 5 °C / min to 200 °C, 20 °C / min to 250 °C, 250 °C for 5 min. Elution time: *cis*-cyclopropanes (19.20 min and 19.33 min), *trans*-cyclopropanes (20.44 min). Kinetic parameters were determined by fitting the data to the standard Michaelis-Menten model.

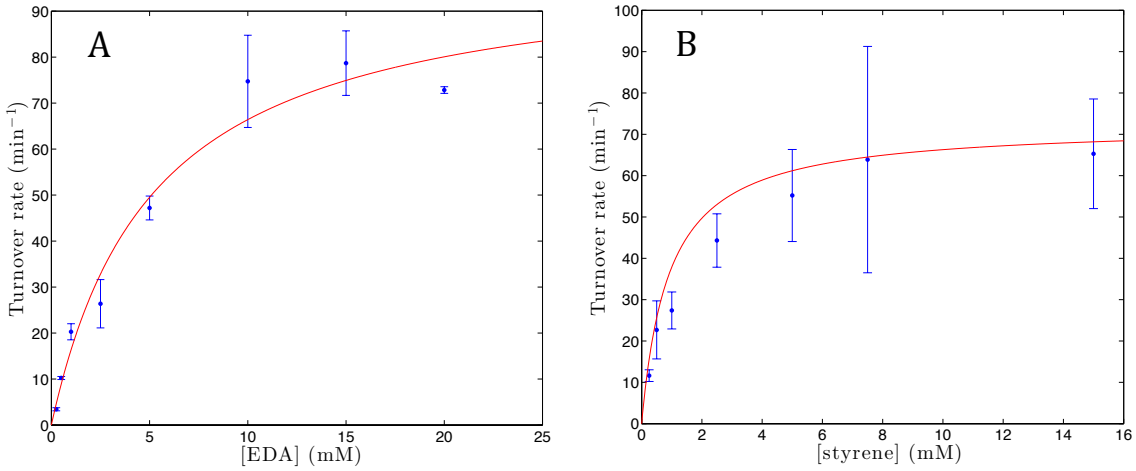


Figure S3. Initial velocities plot for BM3-CIS$_{heme}$. A) EDA concentration was varied at a saturating concentration of styrene (30 mM). B) Styrene concentration was varied at a fixed concentration of EDA (20 mM). Initial rates were computed as the slope of a zero-intercept linear fit of three different time points from independent reactions. Error bars correspond to 1-σ (68.3%) confidence intervals for the slope.

Table S13. Michaelis-Menten parameters for P450 cyclopropanation catalysts

catalyst	k_{cat} **(min^{-1})**	$K_{M\text{-}EDA}$ **(mM)**	$K_{M\text{-}styrene}$ **(mM)**	$k_{cat} / K_{M\text{-}EDA}$ **(s^{-1} M^{-1})**	$k_{cat} / K_{M\text{-}styrene}$ **(s^{-1} M^{-1})**	$k_{cat} / (K_{M\text{-}EDA} \times K_{M\text{-}styrene})$ **(s^{-1} M^{-1} M^{-1})**
BM3-CIS_{heme}	100 ± 24	5.2 ± 3.5	1.4 ± 0.5	320	1,100	2.1×10^5

Table S14. Kinetic parameters for wild-type cytochrome P450s acting on their native substrates and for an engineered variant of $P450_{BM3}$ (propane monooxygenase, PMO) acting on the nonnative substrate propane

P450	**substrate**	k_{cat} **(min^{-1})**	K_M **(mM)**	k_{cat} / K_M **(s^{-1} M^{-1})**
CYP153A6 (*30*)	octane	75	0.32	3,900
$P450_{BM3}$ (*31*)	lauric acid	5140	0.29	3.0×10^5
PMO (*32*)	propane	450	0.17	4.4×10^4

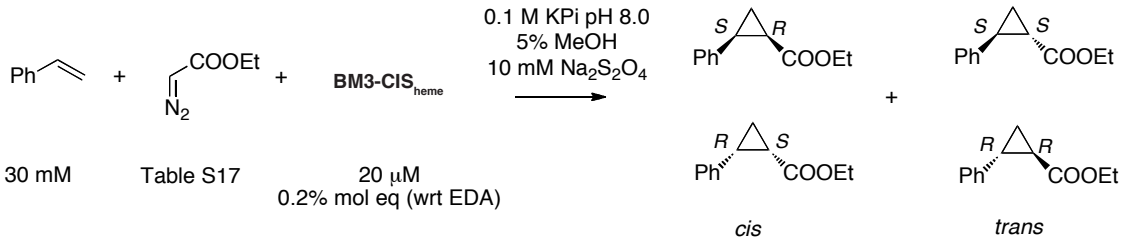


Table S15. Effect of EDA addition at t = 30 min on BM3-CIS-catalyzed cyclopropanations

conditions	TTN	*cis*:*trans**	%ee *cis* †	%ee *trans*‡
10 mM EDA added at t = 0	273 ± 2.5	72 : 28	−92	−90
10 mM EDA added at t = 0 + 10 mM EDA at t = 30 min	425 ± 17	73 : 27	−93	−89

TTN values are reported as the mean of triplicates ± standard deviation. * Diastereomeric ratios and enantiomeric excess were determined by GC analysis. † (*R*,*S*) – (*S*,*R*). ‡ (*R*,*R*) – (*S*,*S*)

VIII. Substrate Scope of P450 Cyclopropanation Catalysts

Small-scale reactions. Selected P450 catalysts were surveyed at a small-scale (400 μL total volume) for each combination of reagents (olefins and diazo esters). The small-scale anaerobic bioconversions were conducted as described in section II and were analyzed by GC.

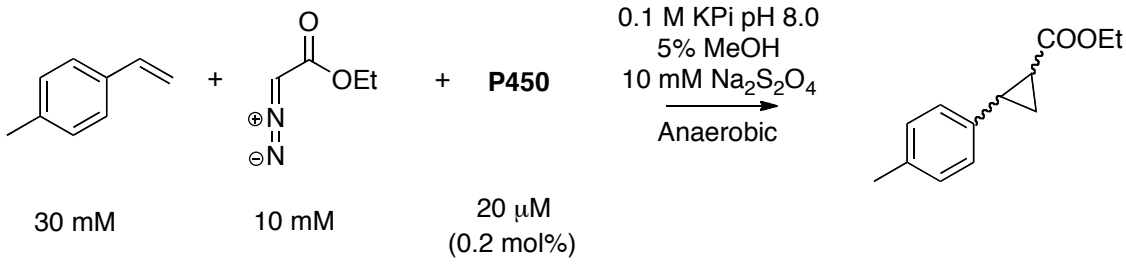


Table S16. Substrate scope of P450 cyclopropanation catalysts: *p*-methylstyrene + EDA

P450	% yield	TTN	*cis*:*trans*	%ee *cis*	%ee *trans**
7-11D	21	104	54:46	0.3	N/A
H2-5-F10	44	222	11:89	14.9	N/A
C2G9 R1	18	92	10:90	8.9	N/A
H2A10	10	50	43:57	−84.3	N/A
BM3-CIS	46	228	78:22	−81.4	N/A
Hemin	7	37	6:94	−1.6	N/A

GC (cyclosil-B column 30 m x 0.32 mm, 0.25 μm film): oven temperature = 100 °C for 5 min, 5 °C / min to 200 °C, 20 °C / min to 250 °C, 250 °C for 5 min. Elution times: *cis*-cyclopropanes (21.03 and 21.18 min), *trans*-cyclopropanes (22.71 min). * *trans*-enantiomers did not resolve.

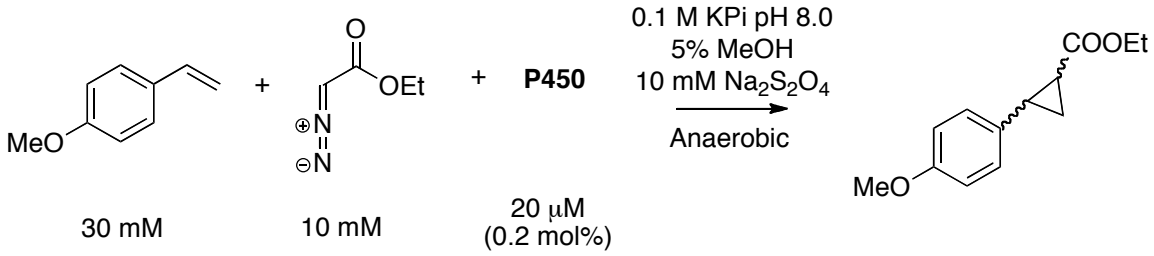


Table S17. Substrate scope of P450 cyclopropanation catalysts: *p*-vinylanisole + EDA

P450	% yield	TTN	*cis*:*trans*	%ee *cis*	%ee *trans**
7-11D	59	297	70:30	−27	N/A
H2-5-F10	73	364	11:89	38	N/A
C2G9 R1	39	196	10:90	−1	N/A
H2A10	16	80	40:60	−75	N/A
BM3-CIS	43	214	48:52	−44	N/A
hemin	19	96	7:93	0	N/A

GC oven temperature = 110 °C for 8 min, 2 °C / min to 180 °C then 180 °C for 30 min, 175 kPa. Cyclosil-B column (30 m x 0.25 mm, 0.25 μm film). Elution times: *cis*-cyclopropanes (38.74 and 39.52 min), *trans*-cyclopropanes (43.07 min). * Baseline resolution could not be achieved for the *trans*-enantiomers.

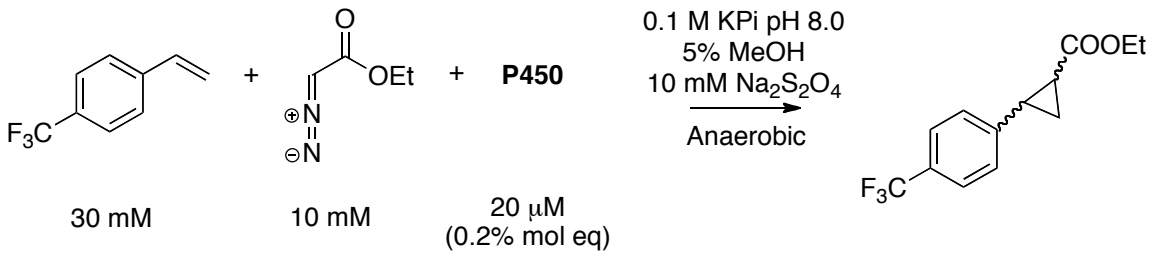


Table S18. Substrate scope of P450 cyclopropanation catalysts: *p*-(trifluoromethyl)styrene

P450	% yield*	TTN*	*cis*:*trans*	%ee *cis*	%ee *trans*
7-11D	24	120	76:24	31	59
H2-5-F10	40	198	26:74	72	−65
C2G9 R1	18	89	10:90	4	0
H2A10	9	47	26:74	−24	22
BM3-CIS	42	211	39:61	54	−93
hemin	2	9	11:89	1	1

* Assumed the same detector response factor as for ethyl 2-(4-methylphenyl)cyclopropane-1-carboxylate. GC (cyclosil-B column 30 m x 0.25 mm, 0.25 μm film): oven temperature = 110 °C for 8 min, 2 °C / min to 180 °C then 180 °C for 30 min, 175 kPa. Elution times: *cis*-cyclopropanes (27.26 and 28.11 min), *trans*-cyclopropanes (30.78 and 30.99 min).

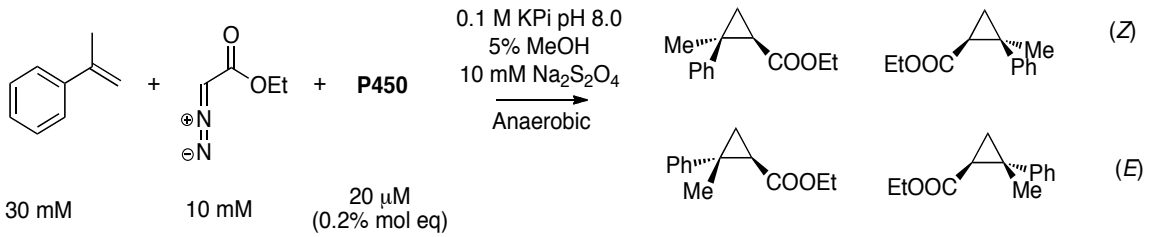


Table S19. Substrate scope of P450 cyclopropanation catalysts: α-methyl styrene

P450	% yield	TTN	*Z*:*E*	%ee (*Z*)	%ee (*E*) *
7-11D	31	157	41:49	42	N/A
H2-5-F10	66	329	21:79	−14	N/A
C2G9 R1	77	387	16:84	−4	N/A
H2A10	34	168	19:81	−31	N/A
BM3-CIS	26	127	16:84	−6	N/A
WT F87V T268A	62	312	7:93	3	N/A
hemin	15	77	24:76	0	N/A

GC oven temperature = 100 °C for 5 min, 1 °C / min up to 135 °C, 135 °C for 10 min, 10 °C / min up to 200 °C, 200 °C for 5 min. Cyclosil-B column (30 m x 0.32 mm, 0.25 μm film). Elution times: *Z*-cyclopropanes (34.96 and 35.33 min), *E*-cyclopropanes (39.34 and 39.61 min). * *trans*-enantiomers did not separate to baseline resolution.

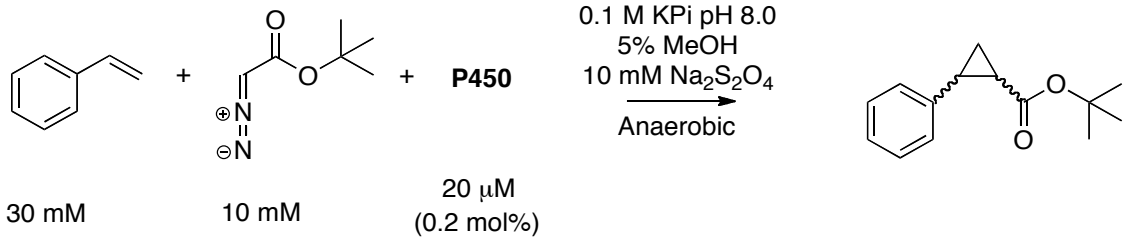


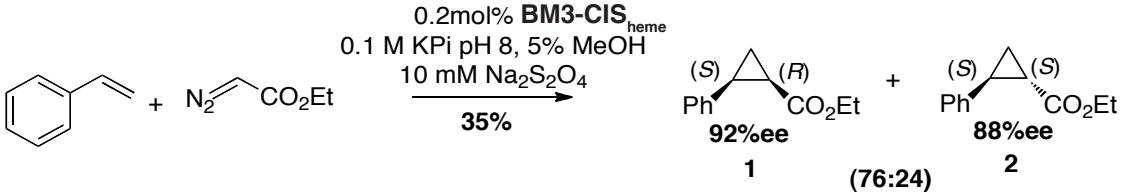
Table S20. Substrate scope of P450 cyclopropanation catalysts: *t*-butyl diazoacetate

P450	% yield*	TTN*	*cis*:*trans*
WT F87V T268A	1.4	7	4:96
7-11D	11	54	13:87
H2-5-F10	18	90	3:97
H2A10	24	120	3:97
BM3-CIS	0.3	2	3:97
hemin	20	100	4:96

* Assumed the same detector response factor as for ethyl 2-(4-methylphenyl)cyclopropane-1-carboxylate. GC (cyclosil-B column 30 m x 0.32 mm, 0.25 μm film): oven temperature = 100 °C for 5 min, 5 °C / min to 200 °C, 20 °C / min to 250 °C, 250 °C for 5 min. Elution times: *cis*-cyclopropanes (21.66 min), *trans*-cyclopropanes (23.31 min). *Cis*- and *trans*-enantiomers did not resolve.

Preparative-scale bioconversions. These reactions were conducted anaerobically as described in section II.

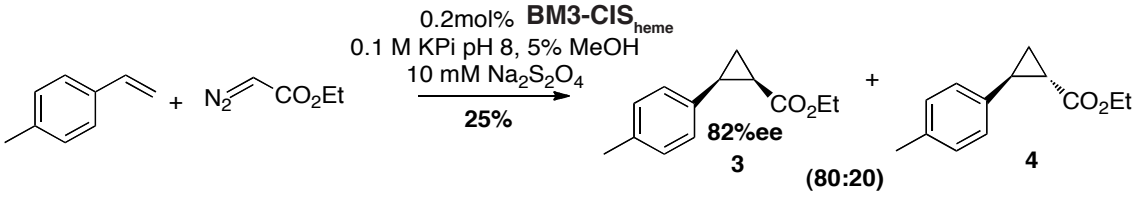
Cyclopropanation of styrene with EDA



Prepared using 1.5 mmol styrene (3 equiv), 0.5 mmol EDA (1 equiv) and 1 μmol BM3-CIS_{heme} (0.002 equiv). The product was purified by SiO_2 chromatography (9:1 hexanes-diethyl ether) to give 25 mg of the *cis*-cyclopropane (**1**) and 8 mg of a mixture of cyclopropanes with *trans* (**2**) in 5 : 1 excess over *cis* (*33-35*). Diagnostic data for the *cis*-cyclopropane **1**: ^{1}H NMR ($CDCl_3$, 500 MHz): δ 7.28 (m, 4H), 7.21 (m, 1H), 3.89 (q, J = 7.1 Hz, 2H), 2.60 (m, 1H), 2.10 (m, 1H), 1.73 (m, 1H), 1.35 (m, 1H), 0.99 (t, J = 7.1 Hz, 3H); ^{13}C NMR ($CDCl_3$, 125 MHz): δ 170.99, 136.56, 129.31, 127.88, 126.63, 60.18, 25.47, 21.80, 14.02, 11.12; $[\alpha]^{25}_D$ = −7.056° (*c* 0.83, $CHCl_3$). Diagnostic data for the *trans*-cyclopropane **2**: ^{1}H NMR ($CDCl_3$, 500 MHz): δ 7.20 (m, 3H), 7.03 (m, 2H), 4.10 (q, J = 7.1 Hz, 2H), 2.45 (m, 1H), 1.83 (m, 1H), 1.53 (m, 1H), 1.23 (m, 1H), 1.21 (t, J = 7.1 Hz, 3H); ^{13}C NMR ($CDCl_3$, 125 MHz): δ 173.43, 140.13, 128.46, 126.55, 126.16, 60.72, 26.18, 24.20, 17.09, 14.27; $[\alpha]^{25}_D$ = +199.2° (*c* 0.50, $CHCl_3$). MS (EI^+) *m/z*: 190 (M^+), 162 ($PhCH(CH_2)CHCO2^+$), 145 ($PhCH(CH_2)CHCO^+$). The absolute configuration of compounds **1** and **2** was determined by comparison of the sign of their optical rotations

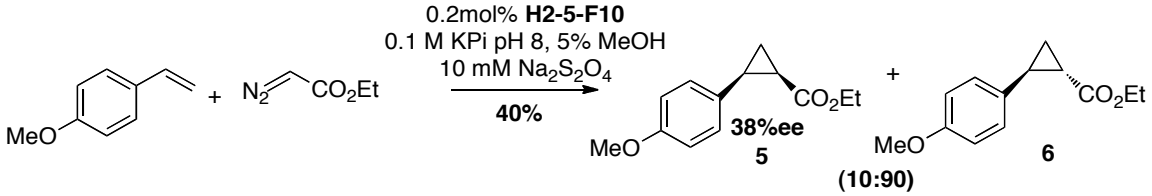
with that reported (*19*). The enantiomeric excess was determined to be 92% for the *cis*-cyclopropane and 88% for the *trans*-cyclopropane by GC.

Cyclopropanation of *p*-methylstyrene with EDA



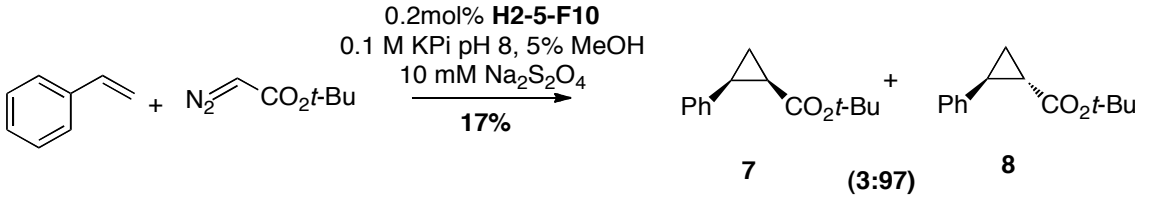
Prepared using 1.5 mmol styrene (3 equiv), 0.5 mmol EDA (1 equiv) and 1 μmol BM3-CIS_{heme} (0.002 equiv). The product was purified by SiO_2 chromatography (9:1 hexanes-diethyl ether) to give 10 mg of the *cis*-cyclopropane (**3**) and 16 mg of a mixture of cyclopropanes with *trans*(**4**) : *cis* / 2 : 1 (*35*). Diagnostic data for the *cis*-cyclopropane **3**: ^{1}H NMR ($CDCl_3$, 500 MHz): δ 7.17 (d, J = 8.0 Hz, 2H), 7.08 (d, J = 8.0 Hz, 2H), 3.91 (q, J = 7.1 Hz, 2H), 2.56 (m, 1H), 2.32 (s, 3H) 2.06 (m, 1H), 1.69 (m, 1H), 1.32 (m, 1H), 1.02 (t, J = 7.1 Hz, 3H); ^{13}C NMR ($CDCl_3$, 125 MHz): δ 171.12, 136.12, 133.42, 129.14, 128.60, 60.17, 25.23, 21.68, 21.10, 14.08, 11.21. Diagnostic data for the *trans*-cyclopropane **4**: ^{1}H NMR ($CDCl_3$, 500 MHz): δ 7.09 (d, J = 8.0 Hz, 2H), 7.01 (d, J = 8.0 Hz, 2H), 4.19 (q, J = 7.1 Hz, 2H), 2.50 (m, 1H), 2.33 (s, 3H), 1.88 (m, 1H), 1.59 (m, 1H), 1.33 (m, 1H), 1.29 (t, J = 7.1 Hz, 3H); ^{13}C NMR ($CDCl_3$, 125 MHz): δ 173.58, 137.04, 136.08, 129.12, 126.10, 60.66, 25.94, 24.06, 21.11, 16.96, 14.28. MS (EI^+) *m/z*: 204 (M^+), 175 ([M-Et]$^+$) 131 ([M-COOEt]$^+$). The enantiomeric excess was determined to be 82% for the *cis*-cyclopropane by GC. Baseline resolution of the *trans*-enantiomers could not be achieved.

Cyclopropanation of *p*-methoxystyrene with EDA



Prepared using 1.5 mmol styrene (3 equiv), 0.5 mmol EDA (1 equiv) and 1 μmol BM3-CIS_{heme} (0.002 equiv). The product was purified by SiO_2 chromatography (9:1 hexanes-diethyl ether) to give 16 mg of the *trans*-cyclopropane (**6**) and 3 mg of a mixture of cyclopropanes with *cis* : *trans* / 5 : 1 (*35*). Diagnostic data for the *trans*-cyclopropane **6**: 6.96 (m, 3H), 6.75 (m, 2H), 4.09 (q, J = 7.1 Hz, 2H), 3.72 (s, 3H), 2.41 (m, 1H), 1.75 (m, 1H), 1.48 (m, 1H), 1.21 (t, J = 7.1 Hz, 3H), 1.18 (m, 1H). MS (EI^+) *m/z*: 220 (M^+), 191 ($[M\text{-}Et]^+$), 175 ($[M\text{-}EtO]^+$), 147 ($[M\text{-}COOEt]^+$). The enantiomeric excess was determined to be 38% for the *cis*-cyclopropane by GC. The *trans*-enantiomers did not resolve to baseline resolution.

Cyclopropanation of styrene with *t*-butyl diazo acetate



Prepared using 0.75 mmol styrene (3 equiv), 0.24 mmol *t*-BuDA (1 equiv) and 0.5 μmol BM3-CIS_{heme} (0.002 equiv). The product was purified by SiO_2 chromatography (9:1 hexanes-diethyl ether) to give 9 mg of the *trans*-cyclopropane (**8**) (*33, 35*). Diagnostic data for the *trans*-cyclopropane **4**: ^1H NMR ($CDCl_3$, 500 MHz): δ 7.20 (m, 2H), 7.12 (m, 1H), 7.02 (m, 2H), 2.36 (m, 1H), 1.76 (m, 1H), 1.45 (m, 1H), 1.40 (s, 9H), 1.16 (m, 1H);

^{13}C NMR ($CDCl_3$, 125 MHz): δ 172.58, 140.52, 128.42, 126.32, 126.07, 80.57, 28.17, 25.75, 25.31, 17.08. MS (EI^+) *m/z*: 218 (M^+), 145 ($[M\text{-}OtBu]^+$).

IX. Summary of Mutations in $P450_{BM3}$ Variants

Mutations in variant P450 cyclopropanation catalysts are reported with respect to wild-type $P450_{BM3}$.

7-11D (*24*): R47C, V78A, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, L353V, A82F, A328V

9-10A TS (*26*): V78A, P142S, T175I, A184V, S226R, H236Q, E252G, A290V, L353V, I366V, E442K

H2A10: **9-10A TS +** F87V, L75A, L181A, T268A

H2-5-F10: **9-10A TS** + F87V, L75A, I263A, T268A, L437A

H2-4-D4: **9-10A TS** + F87V, L75A, M177A, L181A, T268A, L437A

BM3-CIS: **9-10A TS** + F87V, T268A

X. References

1. J. T. Groves, *Proc. Nat. Acad. Sci. U.S.A.* **100**, 3569 (2003).
2. R. Breslow, *J. Biol. Chem.* **284**, 1337 (2009).
3. T. K. Hyster, L. Knoerr, T. R. Ward, T. Rovis, *Science* **338**, 500 (2012).
4. J. B. Siegel *et al.*, *Science* **329**, 309 (2010).
5. H. M. L. Davies, J. R. Manning, *Nature* **451**, 417 (2008).
6. H. Lebel, J.-F. Marcoux, C. Molinaro, A. B. Charette, *Chem. Rev.* **103**, 977 (2003).
7. L. A. Wessjohann, W. Brandt, T. Thiemann, *Chem. Rev.* **103**, 1625 (2003).
8. E. M. Isin, F. P. Guengerich, *Biochim. Biophys. Acta, Gen. Subj.* **1770**, 314 (2007).
9. J. R. Wolf, C. G. Hamaker, J.-P. Djukic, T. Kodadek, L. K. Woo, *J. Am. Chem. Soc.* **117**, 9194 (1995).
10. B. Morandi, E. M. Carreira, *Science* **335**, 1471 (2012).
11. C. J. C. Whitehouse, S. G. Bell, L.-L. Wong, *Chem. Soc. Rev.* **41**, 1218 (2012).
12. J. C. Lewis, F. H. Arnold, *Chimia* **63**, 309 (2009).
13. See supplementary materials.
14. A. Caballero, A. Prieto, M. M. Diaz-Requejo, P. J. Perez, *Eur. J. Inorg. Chem.* 1137 (2009).
15. I. Nicolas, P. Le Maux, G. Simonneaux, *Coord. Chem. Rev.* **252**, 727 (2008).
16. U. T. Bornscheuer, R. J. Kazlauskas, *Angew. Chem. Int. Ed.* **43**, 6032 (2004).
17. O. Khersonsky, C. Roodveldt, D. S. Tawfik, *Curr. Opin. Chem. Biol.* **10**, 498 (2006).
18. A. Penoni *et al.*, *Eur. J. Inor. Chem.* **7**, 1452 (2003).
19. N. Watanabe, H. Matsuda, H. Kuribayashi, S.-I. Hashimoto, *Heterocycles* **42**, 537 (1996).
20. J. Sambrook, E. Frisch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Laboratory Press, New York, 1989), vol. 2.
21. C. R. Otey, in *Methods in Molecular Biology: Directed Enzyme Evolution,* F. H. Arnold, G. Georgiou, eds. (Humana Press, Totowa, NJ, 2003), vol. 230.
22. M. W. Peters, P. Meinhold, A. Glieder, F. H. Arnold, *J. Am. Chem. Soc.* **125**, 13442 (2003).
23. A. Glieder, E. T. Farinas, F. H. Arnold, *Nat. Biotechnol.* **20**, 1135 (2002).
24. P. Meinhold, M. W. Peters, A. Hartwick, A. R. Hernandez, F. H. Arnold, *Adv. Synth. Catal.* **348**, 763 (2006).
25. R. Fasan, M. M. Chen, N. C. Crook, F. H. Arnold, *Angew. Chem., Int. Ed.* **46**, 8414 (2007).
26. J. C. Lewis *et al.*, *Chembiochem* **11**, 2502 (2010).
27. J. C. Lewis *et al.*, *Proc. Nat. Acad. Sci. U.S.A* **106**, 16550 (2009).
28. C. R. Otey *et al.*, *PLoS Biol.* **4**, 789 (2006).
29. M. Landwehr, M. Carbone, C. R. Otey, Y. Li, F. H. Arnold, *Chem. Biol.* **14**, 269 (2007).

30. M. M. Chen, P. S. Coelho, F. H. Arnold, *Adv. Syn. Cat.* **354**, 964 (2012).
31. M. A. Noble *et al.*, *Biochem. J.* **339 (Pt 2)**, 371 (1999).
32. R. Fasan, Y. T. Meharenna, C. D. Snow, T. L. Poulos, F. H. Arnold, *J. Mol. Biol.* **383**, 1069 (2008).
33. C. J. Sanders, K. M. Gillespie, P. Scott, *Tetrahedron: Asymmetry* **12**, 1055 (2001).
34. M. Lenes Rosenberg, A. Krivokapic, M. Tilset, *Org. Lett.* **11**, 547 (2009).
35. Y. Chen, X. P. Zhang, *J. Org. Chem.* **72**, 5931 (2007).

Chapter 3

HIGHLY EFFICIENT CARBENE TRANSFER TO OLEFINS CATALYZED *IN VIVO*

Material from this chapter appears in: P. S. Coelho, Z. J. Wang, M. E. Ener, S. Baril, A. A. Kannan, F. H. Arnold,* E. M. Brustad* "Highly Efficient Carbene Transfer to Olefins Catalyzed *in vivo*" submitted. EMB and SB crystallized and solved the X-ray structures for the proteins described in this chapter. MEE was responsible for the redox titrations.

Abstract

The ability to genetically encode catalysts for nonnatural chemical reactions will open new routes to sustainable production of chemicals, discovery of biologically active agents, and selective labeling of biological targets. We describe the design of a unique serine-heme ligated cytochrome "P411" that catalyzes formal carbene transfers from diazoesters to olefins in intact *E. coli* cells with very high efficiency. Proximal cysteine to serine substitution in a bacterial cytochrome P450 gives a signature ferrous-CO Soret peak at 411 nm, raises the resting state Fe^{III}/Fe^{II} redox potential, abolishes monooxygenation activity and dramatically improves NAD(P)H-driven cyclopropanation. The serine-heme enzyme performs up to 60,000 turnovers *in vivo,* is highly stereoselective, and is more active than the purified protein at the same concentration *in vitro*. Reactions in intact *E. coli* cells require no organic cosolvents, and the cells can be stored as a dry powder without loss of activity. "ABC" catalysts based on cytochrome-P411 should be broadly useful for catalysis of nonnatural carbene transfer reactions.

One sentence summary. Engineering a cytochrome P450 with axial serine ligation creates a "P411" enzyme that catalyzes tens of thousands of turnovers for nonnatural olefin cyclopropanation in bacterial cells.

Genetically programmed whole-cell biocatalysts are readily produced in simple growth media, do not require further purification or isolation, and can be engineered with metabolic pathways for the elaboration of complex molecules (*1*-*3*). The range of accessible transformations, however, is currently limited by the chemical repertoire of natural enzymes. Designing enzymes for nonnatural reactions *in vivo* has been challenging due to the requirements for assembly of the functional catalyst, the compatibility of synthetic reagents in the cellular milieu, and cell permeability to allow substrate influx and product release. The catalysis of synthetic transformations inside cells will enable alternative metabolic routes to natural and artificial products, biobased production of chemicals currently made using synthetic reactions, and will enhance control over the reactivity of abiotic functional groups in "bioorthogonal" chemistry (*4*).

We recently showed that a few amino acid mutations in a bacterial cytochrome P450 monooxygenase can unlock significant cyclopropanation activity *in vitro.* Variants of P450 BM3 from *Bacillus megaterium* (BM3) catalyze hundreds of turnovers of formal carbene transfers from diazoesters (e.g., ethyl diazoacetate, EDA) to olefins (e.g., styrene) in the presence of a reductant, forming cyclopropane products with high levels of diastereo- and enantioselectivity (*5*). Olefin cyclopropanation is widely used in the synthesis of fine chemicals (*6*), and state of the art asymmetric organometallic catalysts are able to catalyze thousands to tens of thousands of turnovers (*7*-*9*). Because BM3 variants are readily expressed in functional form *in vivo* and can catalyze nonnatural carbene transfers without requiring artificial cofactors or posttranslational modifications, we hypothesized that this system may be suitable for catalysis *in vivo*. To initiate the catalytic cycle inside a cell, it is necessary to drive the reduction to the catalytically active

ferrous-P450 with an endogenous reducing agent such as NAD(P)H. Based on consideration of heme ligation control of the P450 Fe^{III}/Fe^{II} redox potential, we designed a genetically encoded "ABC" catalyst that catalyzes efficient and selective olefin cyclopropanation in intact cells.

Cytochrome P450-catalyzed cyclopropanations require substoichiometric (with respect to diazoester and olefin) reductant and proceed optimally under anaerobic conditions (*5*). This suggests that diazoester activation and carbene transfer involve a reduced P450-bound Fe^{II}-heme cofactor as opposed to the resting state Fe^{III}-heme (figure 1A). Active P450-derived cyclopropanation catalysts show marked preference for strong reducing agents such as sodium dithionite ($E°'$ = −660 mV, all potentials versus SHE) over native NAD(P)H ($E°'$ = −320 mV) (*5*). This suggests a limited substrate-induced low-spin ($E°'$ $Fe^{III/II}$ = −430 mV) to high-spin ($E°'$ $Fe^{III/II}$ = −290 mV) transition of the P450 heme-iron (*10*), which, while essential for monooxygenation, may not be achievable in this engineered system due to the poor affinity for the nonnatural substrates (K_M ~5 mM) (*5*). We hypothesized that raising the redox potential of the resting state enzyme to facilitate NAD(P)H reduction would be important for enhancing Fe^{II} catalysis *in vivo*. Because the redox potential of heme proteins can be tuned by axial ligand substitutions (*11, 12*), we reasoned that mutating the axial cysteine in BM3 to the weaker donor serine should raise the $Fe^{III/II}$ potential (*13*). Furthermore, axial cysteinate ligation is essential for dioxygen activation and stabilization of the active ferryl-porphyrin cation radical oxidant (compound I) during monooxygenation (*14*), and axial cysteine to serine substitutions have been reported to abolish monooxygenation activity in mammalian P450s (*15*). Because free hemin is also a (poor) cyclopropanation catalyst (*5*), we

anticipated that the axial cysteine to serine mutation (C400S in BM3) would support carbene transfer activity while abolishing monooxygenation activity.

We chose to introduce the C400S mutation into a *cis*-selective cyclopropanation catalyst from our previous work, BM3-CIS [13 mutations from BM3, (*5, 13*)]. BM3-CIS catalyzes hundreds of turnovers in the presence of dithionite *in vitro* and forms the ethyl 2-phenylcyclopropane-1-carboxylate product with 71% *cis*-selectivity and −94% enantioselectivity (ee_{cis}) (*5*). Heme-serine ligation in BM3-CIS-C400S (hereafter called ABC-CIS) was confirmed by determining the crystal structures of the BM3-CIS and ABC-CIS heme domains at 2.5 and 3.3 Å, respectively (figure S1 and table S1); the structures are superimposable (RMSD = 0.52 Å, figure S1). Despite the limited resolution of the ABC-CIS structure, simulated annealing omit maps generated in the absence of modeled heme and C400S show density consistent with heme coordination by a proximal amino acid side chain (figure 1B and figure S2). The UV-vis spectra for the green-brown ABC-CIS (figures S3-S5) provide further evidence for heme-serine ligation and are consistent with those reported for a Ser-ligated mammalian P450 (*15, 16*) (table S2), marked by a ferrous carbon monoxide-bound complex at 411 nm. Redox titrations using the truncated heme domains of wild-type BM3, the C400S variant (which we call ABC), BM3-CIS and ABC-CIS (figure 1C and figures S6 to S9) showed that the C400S mutation raises the reduction potential of the resting state enzyme by +120 mV (E°' $Fe^{III/II}_{Ser}$ = −295 mV and −265 mV for ABC and ABC-CIS, respectively). This shift is similar in magnitude to that which occurs in BM3 upon substrate binding (*10*), and should therefore allow ABC-CIS to be reduced by NAD(P)H even in the absence of substrate.

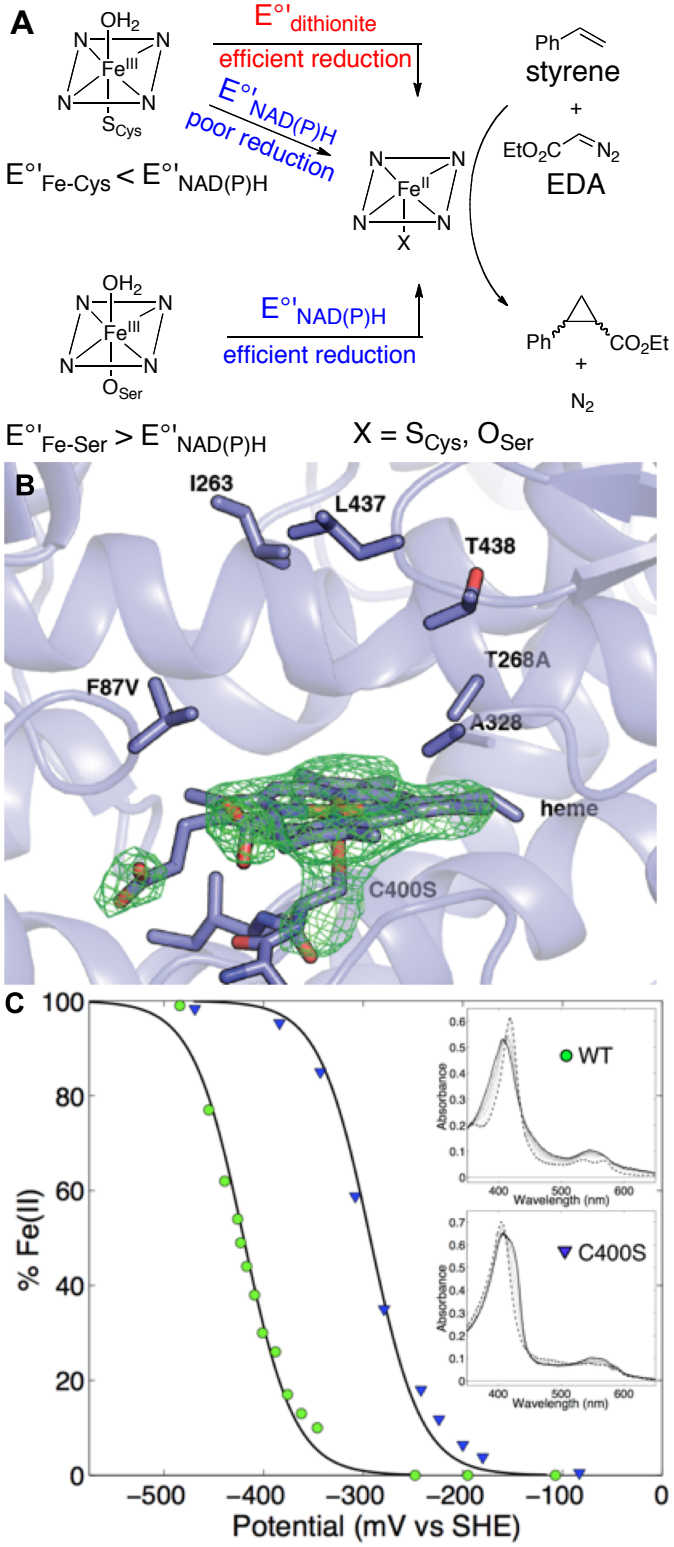

A
E°'dithionite
efficient reduction
E°'NAD(P)H
poor reduction
E°'Fe-Cys < E°'NAD(P)H
styrene
EDA
E°'NAD(P)H
efficient reduction
E°'Fe-Ser > E°'NAD(P)H
X = SCys, OSer
B
I263
L437
T438
T268A
F87V
A328
heme
C400S
C
% Fe(II)
Potential (mV vs SHE)
WT
C400S
Absorbance
Wavelength (nm)

Figure 1

Figure 1. (**A**) Cytochrome P450s inefficiently catalyze cyclopropanation using NAD(P)H as a reductant because the Fe^{III}/Fe^{II} redox potential for the low-spin resting state ($E°'_{Fe\text{-}cys}$ = −370 mV) is lower than that of $NAD(P)^+$/NAD(P)H ($E°'_1$ = −320 mV). We hypothesized that serine ligation should increase the resting state redox potential and allow reduction from NAD(P)H to the active Fe^{II} species *in vivo*. (**B**) Close-up of the ABC-CIS active site (pdb: 4H24) superimposed with an F_o - F_c simulated annealing omit map contoured at 3 σ showing electron density (green mesh) corresponding to the bound heme and C400S mutation. Interconnected density between C400S and the heme iron is consistent with proximal heme ligation by the C400S side chain hydroxyl. The heme, C400S and additional active site amino acid side chains are shown as sticks. (**C**) Potentiometric redox titrations for BM3 (green circles) and ABC (blue triangles) with overlaid one-electron Nernst curves. Insets show spectral changes between ferric (dashed line) and ferrous (solid line) enzymes. The changes in absorbance near 450-470 nm (BM3) and 420-440 nm (ABC) were used to determine the Fe^{II}/Fe^{III} ratio after reduction with dithionite. The reduction is reversible (with reoxidation by potassium ferricyanide), and little or no hysteresis was observed. The midpoint potential of the serine-ligated mutant (-295 mV) is shifted 125 mV positive compared to WT (-420 mV).

ABC-CIS is an active dithionite-driven cyclopropanation catalyst *in vitro*, with Michaelis-Menten parameters (k_{cat} = 82 min^{-1}, $K_{M\text{-}styrene}$ = 4.6 mM, $K_{M\text{-}EDA}$ = 5.7 mM, figure S11) comparable to those of BM3-CIS (table S3). ABC-CIS displays considerably improved diastereo- (*cis*:*trans* 93:7) and enantioselectivity (−99% ee_{cis}) compared to its cysteine homologue (figure S12), an unexpected result given the similar active site geometries of the two catalysts (figure S1). For a variety of styrenyl substrates, ABC-CIS showed superior *cis*-selectivity relative to BM3-CIS (table S4). ABC-CIS shows increased activity compared to BM3-CIS when NADPH is used as the reductant under anaerobic conditions (figure 2 and table S5). BM3-CIS only forms small amounts of cyclopropanes when NADPH is used, and forms styrene oxide, via monooxygenation, as the major product under aerobic conditions. In contrast, ABC-CIS produces negligible amounts of styrene oxide and is still able to form cyclopropanes under aerobic conditions, albeit with lower yields (43 TTN) due to oxygen inhibition (figure 2). Dioxygen

inhibition could be due to a two-electron oxidase activity as reported for CYP2B4-C436S (*15*). NADH drives ABC-CIS-mediated cyclopropanation as efficiently as NADPH (table S6), suggesting that ABC-CIS is well suited for *in vivo* catalysis under anaerobic conditions where NADPH biosynthesis in *E. coli* does not take place.

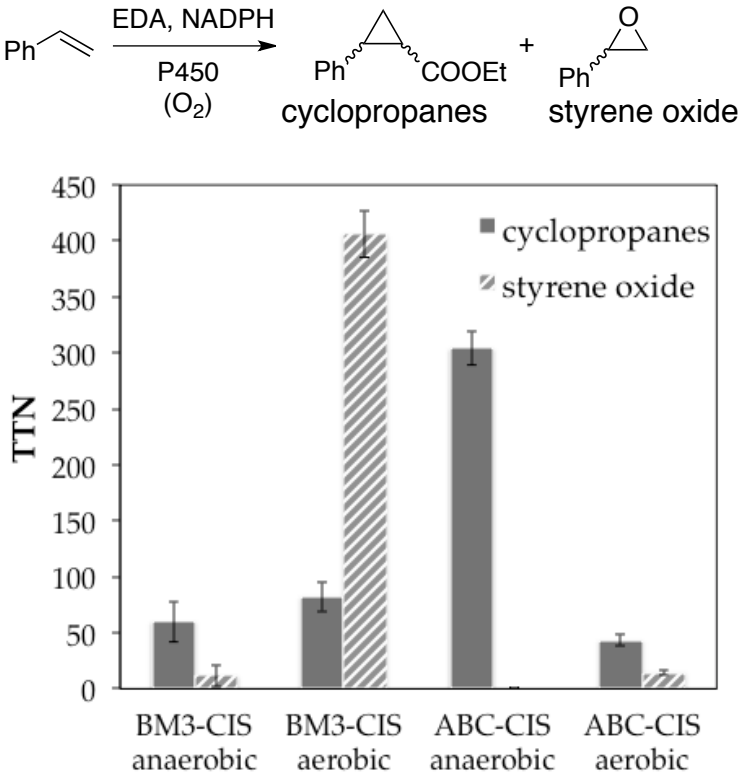


Figure 2. Cyclopropanation vs. epoxidation of styrene catalyzed by BM3-CIS and ABC-CIS under anaerobic and aerobic conditions. Reaction conditions were as follows: 30 mM styrene, 10 mM EDA, 0.5 mM NADPH, 25 mM glucose, 2 U ml^{-1} glucose dehydrogenase and 20 μM P450 catalyst in aqueous potassium phosphate buffer and 5% MeOH cosolvent for six hours at 298 K. Measurements were taken in triplicate and the error bars represent the standard deviation of the measurements.

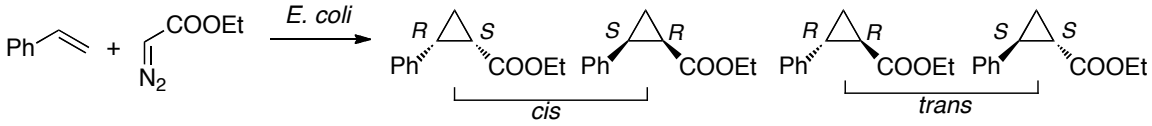
We next investigated the efficiency of cyclopropanation using resting *Escherichia coli* (BL21) cells grown in M9Y media [M9, 1.5% yeast extract, (*13*)] expressing BM3-CIS and ABC-CIS. Addition of glucose under anaerobic conditions significantly

increased product yield (figure S13), presumably due to enhanced intracellular production of NADH. ABC-CIS catalyzes thousands of turnovers *in vivo*, is about four times more active than BM3-CIS in whole-cells, and provides the cyclopropane products with enhanced *cis*- enantioselectivity (table 1, entries 1 and 2). The C400S mutation compromises protein expression such that ABC-CIS accounts for 2% of dry cell mass compared to 6% for BM3-CIS. The reduced expression is not due to decreased protein stability, as C400S contributes to increased thermostability in the purified ABC-CIS heme domain (figure S10). The holo enzyme, which contains both heme and diflavin reductase domains, is over two times more active on a molar basis than the heme domain alone (ABC-CIS$_{heme}$) (entry 3), confirming that reduction to the ferrous state *in vivo* is important, but also showing that the reducing intracellular environment achieves heme reduction even in the absence of the reductase domain. The C400S mutation (ABC) improves the *in vivo* cyclopropanation activity of BM3 by over two orders of magnitude (entries 4 and 5). Purified ABC is also an efficient NADH-driven cyclopropanation catalyst *in vitro*, whereas BM3 is barely active (table S7).

Both ABC-CIS and BM3-CIS whole cells are significantly inhibited by dioxygen (figure S14). Whole-cells containing the ABC-CIS gene but with no induction and whole cells devoid of the ABC-CIS gene are able to form small amounts of cyclopropanes, but do so with stereoselectivity similar to that of free hemin (figure S15). This is not surprising since free hemin and other heme proteins present in cells are also able to catalyze styrene cyclopropanation at low levels (*5*). Whole-cell ABC-CIS catalysts are as stereoselective as purified ABC-CIS *in vitro* at equivalent catalyst loading (*vide infra*), demonstrating that the overexpressed P411 enzyme favorably outcompetes background

catalysis. *In vivo* cyclopropanation is strongly inhibited by carbon monoxide (table 1, entry 6), which irreversibly binds ferrous heme, confirming that catalysis occurs in the enzyme active site. Yields could be increased to 80% by increasing the cell density up to OD_{600} = 50 (figure S16). Reaction yield was only slightly improved by using excess styrene (figure S17). Lysate of cells expressing ABC-CIS and with NADH added retain only about 30% of the activity of the intact whole-cells and are not active in the absence of exogenous reductant (table S8). Addition of dithionite inhibited ABC-CIS whole-cell reactions and was less efficient than NADH in driving the reaction in cell lysate (figure S15 and table S8)

Table 1. Cyclopropanation activities for intact *E. coli* cells expressing BM3, ABC, BM3-CIS, ABC-CIS and ABC-CIS_{heme}. Reaction conditions were as follows: 2 eq styrene, 1 eq EDA, 0.2 eq glucose, *E. coli* whole-cells in aqueous nitrogen-free M9 minimal medium and 5% MeOH cosolvent under anaerobic conditions for twelve hours at 298 K. Yields, diastereomeric ratios, and enantiomeric excess were determined by GC analysis. Yields based on EDA. TTN = total turnover number. *(2*R*,1*S*) – (2*S*,1*R*). †(2*R*,1*R*) – (2*S*,1*S*). ‡Neat reagents were used without addition of MeOH; reactions were left for 24 h. §Isolated yield (1.63 g cyclopropanes). The data represent the averages of triplicate experiments. Standard errors are within 20% of the reported average.



Entry	Catalyst	[EDA] (mM)	Cell Density (g_{cdw}/L)	[P450] (μM)	Yield (%)	TTN	*cis*:*trans*	ee_{cis} (%)*	ee_{trans} (%)†
1	BM3-CIS	8.5	7.7	3.7	42	950	22:78	–60	–22
2	ABC-CIS	8.5	7.7	1.3	55	3,700	76:24	–96	–25
3	ABC-CIS_{heme}	8.5	7.7	3.6	67	1,600	71:29	–95	–17
4	BM3	8.5	13.4	4.8	0.9	15	25:75	–24	–21
5	ABC	8.5	8.1	1.5	50	2,900	13:87	–12	–8
6	ABC-CIS + CO	8.5	8.4	1.8	0.6	30	20:80	–35	–20
7	ABC-CIS‡	170	8.4	1.8	72	67,800	90:10	–99	–43
8	ABC-CIS‡	200	20	3.2	78§	48,800	88:12	–99	–35

In order to provide a direct comparison of enzyme activity *in vivo* versus *in vitro*, we monitored both reactions at the same enzyme concentration over 8 hours (figure 3). On a molar basis, the *in vivo* catalyst showed almost 6 times higher TTN than the purified enzyme *in vitro* after 8 hours and retained the same stereoselectivity (75:25 cis:trans, -95% ee_{cis}). Both catalysts remained active over 6 hours, suggesting that the observed differences in yield and TTN are due to improved activity rather than enhanced catalyst stability *in vivo*. Gradual addition of EDA did not improve the reaction yield.

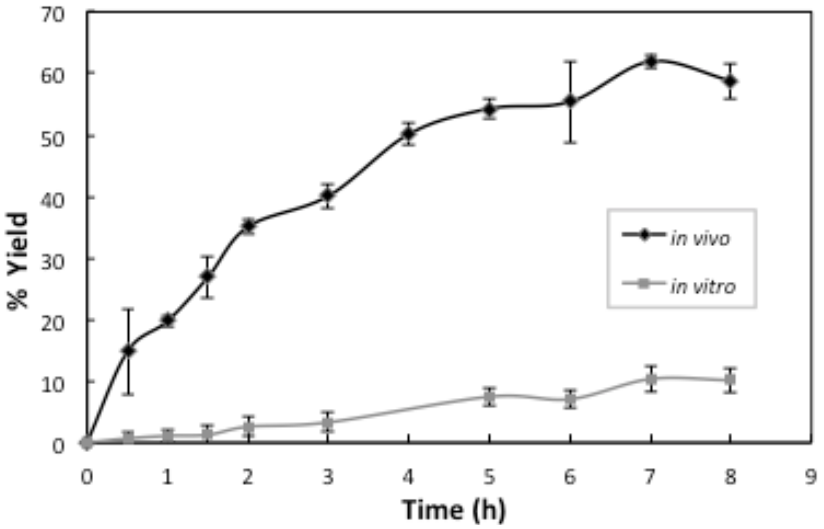


Figure 3. Time course for *in vivo* and *in vitro* ABC-CIS at a P411 loading of 1.6 μM [$\varepsilon_{411\text{-}490}$ = 103 mM^{-1} cm^{-1} for the ferrous-CO complex (*15*)]. Reaction conditions were as follows: 40 mM styrene, 20 mM EDA, 2 mM glucose, 10.2 g_{cdw} L^{-1} whole cell ABC-CIS (*in vivo*), 1.6 μM purified ABC-CIS, 1 mM NADH (*in vitro*) in aqueous nitrogen-free M9 minimal medium and 5% MeOH cosolvent under anaerobic conditions at 298 K. Yields at each time point are reported as averages of two independent reactions (*13*).

At high substrate loading (170 mM EDA, 400 mM styrene, added as neat reagents), more than 60,000 catalytic turnovers were observed in the *in vivo* reaction with

ABC-CIS (table 1, entry 6). ABC-CIS whole-cell reactions are readily scalable to make gram quantities of cyclopropanes with high stereoselectivity, product titer (27 g L^{-1}) and yield (78%, entry 8). No organic cosolvents are necessary, and the cyclopropane products can be readily obtained by extraction with organic solvent at the end of the reaction (*13*). Furthermore, the cells can be lyophilized with a cryoprotectant such as sucrose and stored as a powder for weeks at 4 °C without degradation of catalytic activity or regio- and enantioselectivity (table S9). The lyophilized cells can be readily packaged and distributed. These features render the whole-cell process extremely attractive for facile benchtop synthesis.

The ABC catalyst based on a Ser-ligated cytochrome-P411 is spectroscopically, electrochemically, and catalytically distinct from cytochrome P450s. Whole-cell ABC catalysts are easy to use and deliver high conversion, optical purity and yield for substrate input in the tens of grams per liter. The ability to catalyze this nonnatural C−C bond forming reaction *in vivo* will expand the scope of transformations accessible to microbial organic synthesis and should provide artificial metabolic pathways to complement nature's existing strategies for making cyclopropanes (*17*).

* * *

Supplementary Materials for

Highly Efficient Carbene Transfer to Olefins Catalyzed *In Vivo*

I. Materials and Methods

Unless otherwise noted, all chemicals and reagents for chemical reactions were obtained from commercial suppliers (Sigma-Aldrich, Acros) and used without further purification. Silica gel chromatography purifications were carried out using AMD Silica Gel 60, 230-400 mesh. ^{1}H and ^{13}C NMR spectra were recorded on either a Varian Mercury 300 spectrometer (300 and 75 MHz, respectively), or a Varian Inova 500 MHz (500 and 125 MHz, respectively), and are internally referenced to residual solvent peak. Data for ^{1}H NMR are reported in the conventional form: chemical shift (δ ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad), coupling constant (Hz), integration. Data for ^{13}C are reported in terms of chemical shift (δ ppm) and multiplicity. High-resolution mass spectra were obtained with a JEOL JMS-600H High Resolution Mass Spectrometer at the California Institute of Technology Mass Spectral Facility. Reactions were monitored using thin layer chromatography (Merck 60 silica gel plates) using a UV lamp for visualization.

Gas chromatography (GC) analyses were carried out using a Shimadzu GC-17A gas chromatograph, a FID detector, and an Agilent J&W cyclosil-B column (30 m x 0.25 mm, 0.25 μm film) and 2-phenylethanol as an internal standard. Injector temperature = 300 °C, oven temperature = 130 °C for 30 min, pressure = 175 kPa. Elution time: *cis*-cyclopropanes (19.7 and 21.0 min), *trans*-cyclopropanes (25.8 and 26.4 min). Cyclopropane product standards for the reaction of ethyl diazoacetate (EDA) with styrene (ethyl 2-phenylcyclopropane-1-carboxylate) and α-methylstyrene (ethyl 2-methyl-2-phenylcyclopropane-1-carboxylate) were prepared as reported (*18*). These standards and enzyme-prepared cyclopropanes demonstrated identical retention times in gas

chromatograms when coinjected, confirming product identity. Absolute stereoconfiguration of cyclopropane enantiomers was determined by measuring optical rotation of purified cyclopropane products from preparative bioconversion reactions using enantioselective P450-BM3 variants and referenced to values taken from reference (*19*). Authentic P450-catalyzed cyclopropane samples were also prepared as described in section II and were characterized by NMR (^{1}H and ^{13}C) and mass spectrometry.

Plasmids pCWori[BM3] and pET22 were used as cloning vectors. Site-directed mutagenesis was accomplished by standard overlap mutagenesis using primers bearing desired mutations (IDT, San Diego, CA). Primer sequences are available upon request. Electrocompetent *Escherichia coli* cells were prepared following the protocol of Sambrook *et al.* (*20*). Restriction enzymes BamHI, EcoRI, XhoI, Phusion polymerase, and T4 ligase were purchased from New England Biolabs (NEB, Ipswich, MA). Alkaline phosphatase was obtained from Roche (Nutley, NJ). The 1,000x trace metal mix used in expression cultures contained 50 mM $FeCl_3$, 20 mM $CaCl_2$, 10 mM $MnSO_4$, 10 mM $ZnSO_4$, 2 mM $CoSO_4$, 2 mM $CuCl_2$, 2 mM $NiCl_2$, 2 mM Na_2MoO_4, and 2 mM H_3BO_3.

II. General Procedures

CO binding assay. P450 concentration was determined from ferrous CO binding difference spectra using extinction coefficients of $\varepsilon_{450\text{-}490}$ = 91 mM^{-1} cm^{-1} for cysteine-ligated BM3 (*21*) and $\varepsilon_{411\text{-}490}$ = 103 mM^{-1} cm^{-1} for serine ligated ABC (*15*).

P450 expression and purification. For the *in vitro* transformations, $P450_{BM3}$ variants were used in purified form. Enzyme batches were prepared as follows. One liter TB_{amp} was inoculated with an overnight culture (100 mL, LB_{amp}) of recombinant *E. coli* BL21 cells harboring a pCWori plasmid encoding the P450 variant under the control of the *tac* promoter. After 3.5 h of incubation at 37 °C and 250 rpm shaking (OD_{600} ca. 1.8), the incubation temperature was reduced to 25 °C (30 min), and the cultures were induced by adding IPTG to a final concentration of 0.5 mM. The cultures were allowed to continue for another 24 hours at this temperature. After harvesting the cells by centrifugation (4 °C, 15 min, 3,000 × g), the cell pellet was stored at -20 °C until further use but at least for 2 h. The cell pellet was resuspended in 25 mM Tris.HCl buffer (pH 7.5 at 25 °C) and cells were lysed by sonication (2 × 1 min, output control 5, 50% duty cycle; Sonicator, Heat Systems—Ultrasonic, Inc.). Cell debris was removed by centrifugation for 20 min at 4 °C and 27,000xg and the supernatant was subjected to anion exchange chromatography on a Q Sepharose column (HiTrapTM Q HP, GE Healthcare, Piscataway, NJ) using an AKTAxpress purifier FPLC system (GE healthcare). The P450 (or P411) was eluted from the Q column by running a gradient from 0 to 0.5 M NaCl over 10 column volumes (P450 elutes at 0.35 M NaCl). The P450 (or P411) fractions were collected and concentrated using a 30 kDa molecular weight cutoff centrifugal filter and buffer-

exchanged with 0.1 M phosphate buffer (pH = 8.0). The purified protein was flash-frozen on dry ice and stored at -20 °C. P450 and P411 concentrations were determined in triplicate using the CO binding assay described above (10 μL P450 and 190 μL 0.1 M phosphate buffer, pH 8.0, per well).

For crystallization experiments, a two step purification was performed using the AKTAxpress purifier FPLC system. Frozen cell pellets containing expressed, 6XHis tagged heme domains were resuspended in Ni-NTA buffer A (25 mM Tris.HCl, 200 mM NaCl, 25 mM imidazole, pH 8.0, 0.5 mL/gcw) and lysed by sonication (2 × 1 min, output control 5, 50% duty cycle). The lysate was centrifuged at 27,000xg for 20 min at 4 °C to remove cell debris. The collected supernatant was first subjected to a Ni-NTA chromatography step using a Ni sepharose column (HisTrap-HP, GE healthcare, Piscataway, NJ). The P450 (or P411) was eluted from the Ni sepharose column using 25 mM Tris.HCl, 200 mM NaCl, 300 mM imidazole, pH 8.0. Ni-purified protein was buffer exchanged into 25 mM Tris.HCl pH 7.5 using a 30 kDa molecular weight cutoff centrifugal filter and subsequently loaded onto a Q sepharose column (HiTrapTM Q HP, GE healthcare, Piscataway, NJ) and purified to homogeneity by anion exchange. The P450 (or P411) was eluted from the Q column by running a gradient from 0 to 0.5 M NaCl over 10 column volumes. P450 (or P411) fractions were collected and buffer exchanged into 25 mM Tris.HCl pH 7.5, 25 mM NaCl. The purified protein was concentrated with a 30 kDa molecular weight cutoff centrifugal filter to approximately 10 mg/mL. 50 μL aliquots were flash frozen on dry ice and stored at −80 °C until needed.

Protein crystallography. BM3-CIS and ABC-CIS were crystallized by vapor diffusion. A 1:1 mixture of protein stock (10 mg/mL in 25 mM Tris.HCl pH 7.5, 25 mM NaCl) and mother liquor was combined in 24-well sitting drop plates (Hampton Research). Optimal crystallization conditions for BM3-CIS were found in 0.1 M sodium cacadolyte, pH 5.7, 0.14 $MgCl_2$ and 17 % PEG 3350. CM3-CIS crystals typically grew over a span of 7–14 days. ABC-CIS crystals optimally formed in 0.1 M Bis-Tris, pH 5.3, 0.2 M sodium formate and 18% PEG 3350. Initial ABC-CIS drops are marked with a dense layer of protein precipitate; however, after 36–48 hours, noticeable protein crystals were observed underneath the precipitate layer.

X-ray data collection and protein structure determination. X-ray diffraction data were collected at the General Medical Sciences and Cancer Institutes Structural Biology Facility (GM/CA) at the Advanced Photon Source (APS, Argonne National Laboratory) using beamline ID23-D and a MAR300 CCD detector. Data were collected at 100K and a wavelength of 1.033 Å. Data collections statistics are listed in table S1. Diffraction datasets were integrated with XDS (*22*) and scaled using SCALA (*23*). Initial phases were determined by molecular replacement against the closed form of wild type P450-BM3h structure taken from PDB 1JPZ (*24*), chain B using MOLREP software (*25*), a component of the CCP4 crystallography software suite (*26*). Refinement was accomplished by iterative cycles of manual model building within COOT (*27*) and automated refinement using REFMAC (*28*) within CCP4. Final cycles of REFMAC refinement included TLS parameters. Non-crystallographic symmetry constraints were not used during refinement. Model quality was assessed using the “complete validation”

tool inside of the PHENIX software suite (*29*). Ramachandran outliers generally lie in poorly structured loops connecting $P450_{BM3}$ F and G helices. These residues are often missing or marked by poor density in these and other $P450_{BM3}$ structures within the protein database. All protein structure figures and alignments were generated using PyMol software (The PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC.).

Thermostability measurements. Duplicate measurements were taken for all values reported in figure S10. Purified P450 (or P411) solutions (4 μM, 200 μL) were heated in a thermocycler (Eppendorf) over a range of temperatures (40–70 °C) for 10 min followed by rapid cooling to 4 °C for 1 min. The precipitate was removed by centrifugation. The concentration of folded P450 (or P411) remaining in the supernatant was measured by CO-difference spectroscopy (as described above). The temperature at which half of the protein was denatured (T_{50}) was determined by fitting the data to the equation:

$$f(T) = \frac{100}{1+e^{-a\left(\frac{1}{T}-\frac{1}{T_{50}}\right)}} .(30)$$

Typical procedure for *in vitro* small-scale cyclopropanation bioconversions under anaerobic conditions. Small-scale reactions (400 μL) were conducted in 2 mL crimp vials (Agilent Technologies, San Diego, CA). P450 solution (80 μL, 100 μM) was added to the vial with a small stir bar before crimp sealing with a silicone septum. Phosphate buffer (260 μL, 0.1 M, pH = 8.0) and 40 μL of a solution of the reductant (100 mM $Na_2S_2O_4$, or 20 mM NADPH) were combined in a larger crimp-sealed vial and degassed

by bubbling argon through the solution for at least 5 min (figure S1). In the meantime, the headspace of the 2 mL reaction vial with the P450 (or P411) solution was made anaerobic by flushing argon over the protein solution (with no bubbling). When multiple reactions were conducted in parallel, up to 8 reaction vials were degassed in series via cannulae. The buffer/reductant solution (300 μL) was syringed into the reaction vial, while under argon. The gas lines were disconnected from the reaction vial before placing the vials on a plate stirrer. A 40x styrene solution in MeOH (10 μL, typically 1.2 M) was added to the reaction vial via a glass syringe, and left to stir for about 30 s. A 40x EDA solution in MeOH was then added (10 μL, typically 400 mM) and the reaction was left stirring for the appropriate time. The final concentrations of the reagents were typically: 30 mM styrene, 10 mM EDA, 10 mM $Na_2S_2O_4$, 20 μM P450.

The reaction was quenched by adding 30 μL HCl (3M) via syringe to the sealed reaction vial. The vials were opened and 20 μL internal standard (20 mM 2-phenylethanol in MeOH) was added followed by 1 mL ethyl acetate. This mixture was transferred to a 1.8 mL eppendorf tube which was vortexed and centrifuged (16,000xg, 1 min). The top organic layer was dried over an anhydrous sodium sulfate plug and analyzed by chiral phase GC.

A slightly modified workup was implemented for kinetic experiments. The reactions were quenched after the set time by syringing 1 mL EtOAc to the closed vials and immediately vortexing the mixture. The vials were then opened and 20 μL internal standard was added. The mixture was transferred to a 1.8 mL eppendorf tube, vortexed and centrifuged (16,000xg, 1 min). The top organic layer was dried over an anhydrous sodium sulfate plug and analyzed by GC.

Media and cell cultures for *in vivo* cyclopropanation. *E. coli* (BL21) cells were grown from glycerol stock overnight (37 °C, 250 rpm) in 5 mL M9Y medium (1 L: 31 g Na_2HPO_4, 15 g KH_2PO_4, 2.5 g NaCl, 5.0 g NH_4Cl, 0.24 g $MgSO_4$, 0.01 g $CaCl_2$, 1.5% yeast extract, 1 mL micronutrients, 0.1 mg mL^{-1} ampicillin). The preculture was used to inoculate 45 mL of M9Y medium in a 125 mL Erlenmeyer flask and this culture was incubated at 37 °C, 250 rpm for 2 hours and 30 minutes. At OD_{600} = 1.2, the cultures were cooled to 25 °C and the shaking was reduced to 160 rpm before inducing with IPTG (0.25 mM) and δ-aminolevulinic acid (0.25 mM). Cultures were harvested after 20 hours and resuspended (OD_{600} = 30) in nitrogen-free M9 medium (1 L: 31 g Na_2HPO_4, 15 g KH_2PO_4, 2.5 g NaCl, 0.24 g $MgSO_4$, 0.01 g $CaCl_2$, 1 mL micronutrients). The micronutrient solution contains 0.15 mM $(NH_4)_6Mo_7O_{24}$, 20.0 mM H_3BO_3, 1.5 mM $CoCl_2$, 0.5 mM $CuSO_4$, 4.0 mM $MnCl_2$, and 0.5 mM $ZnSO_4$. Aliquots of the cell suspension were used for determination of the cell dry weight (cdw, 2 mL) and P450 (or P411) expression level (4 mL).

Small-scale whole-cell bioconversions. *E. coli* cells (OD_{600} = 30, 425 μL) were made anaerobic by bubbling argon through the cell suspension in a crimped 2 mL vial. A degassed solution of glucose (50 μL, 20 mM) was added to the cells before adding EDA (12.5 μL of a 400 mM solution in MeOH) and olefin (12.5 μL of a 1.2 M solution in MeOH). The reactions were stirred at room temperature for the appropriate and were worked up by adding 20 μL of the internal standard (20 mM 2-phenylethanol) and

extracting with 1 mL ethyl acetate. The organic layer was dried over Na_2SO_4 before analyzing the product mixture by chiral phase GC.

Preparative-scale whole-cell bioconversions. *E. coli* (BL21) cells were grown from glycerol stock overnight (37 °C, 250 rpm) in 50 ml M9Y medium. The preculture was used to inoculate 2-475 mL of M9Y medium in 2-1 L Erlenmeyer flask (using 25 mL each) and this culture was incubated at 37 °C, 250 rpm for 2 h and 30 min. At OD_{600} = 1.8, the cultures were cooled to 25 °C and the shaking was reduced to 150 rpm before inducing with IPTG (0.25 mM) and δ-aminolevulinic acid (0.25 mM). Cultures were harvested after 24 h and resuspended (OD_{600} = 75) in nitrogen-free M9 medium. Aliquots of the cell suspension were used for determination of the cell dry weight (cdw, 2 mL) and P450 (or P411) expression level (2 mL). *E. coli* cells (OD_{600} = 70, 53.6 mL) were made anaerobic by bubbling argon through the cell suspension in a 500 mL sealed round bottom flask. A degassed solution of glucose (1.4 mL, 500 mM) was added to the cells before adding EDA (1.36 mL, 85% EDA in DCM as packaged by Sigma Aldrich) and styrene (2.5 mL, neat). The reaction was stirred at room temperature under positive argon pressure for 24 h. The crude mixture was poured into 3-50 mL conical tubes, and the reaction was quenched by the addition of HCl (1 mL, 3 M) to each tube. The aqueous mixtures were extracted with 1:1 EtOAc: hexanes (20 mL each) and centrifuged (5,000 rpm, 5 min). The organics were collected and this extraction sequence was performed two more times. The organics were combined, dried over Na_2SO_4 then concentrated. Excess styrene was removed via azeotrope with H_2O/benzene and 1.85 g of crude product was isolated. *Cis*/*trans* selectivity of the reaction was determined via gas

chromatography of this crude mixture. Column chromatography of the crude product with 8% Et_2O/hexanes afforded the desired products as a mixture of cis and trans isomers (1.63 g combined, 78% yield). Based on comparison of crude and purified yields, the crude product was approximately 88% pure. NMR of the isolated products were identical to those reported in the literature (*5*).

Time course of *in vivo* and *in vitro* reactions. Following the procedure for small scale bioconversions, a series of *in vivo* and *in vitro* reactions were set up and EDA was added to each sample at time 0 hours. Time points were taken at 0.5, 1, 1.5, 2, 3, 4, 5, 6, 7, and 8 hours. Each reported yield reflects an average of two independent reactions that were allowed to stir for the indicated amount of time. The error bars shown reflect the two unaveraged data points. Yields of each reaction were determined by GC.

Potentiometric titrations. Enzyme samples were buffer-exchanged into 100 mM KPO4, 100 mM KCl, pH 7.4, and deoxygenated via 4x20 gentle pump-backfill cycles with argon, with care taken to avoid bubbling. Potentiometric redox titrations were performed in an anaerobic glove box, using a quartz spectroelectrochemical cell with path length of 1mm, platinum mesh working electrode, platinum wire counter electrode, and a Ag/AgCl electrode (Bioanalytical Systems, Inc.) Was used as the reference (Ag/AgCl vs NHE: +197 mV). Protein solutions consisted of approximately 600 µl of 50-100 µM protein with the following mediators added to ensure electrochemical communication between the protein and electrode: methyl viologen (5 µM), benzyl viologen (10µM) and 2-hydroxy-1,4-napthaquinone (20 µM). Enzyme samples were titrated using sodium

dithionite (reduction) and potassium ferricyanide (re-oxidation). The open circuit potential of the cell was monitored (WaveNow potentiostat, Pine Research Instrumentation) over a 10-minute equilibration period, and spectra were recorded using a Ocean Optics spectrometer (USB2000+).

III. Summary of P450-Derived Cyclopropanation Catalysts

Mutations in cyclopropanation catalysts are reported with respect to wild-type $P450_{BM3}$. The heme domain comprises the first 462 amino acids in the $P450_{BM3}$ sequence.

ABC: BM3 + C400S

BM3-CIS: V78A, F87V, P142S, T175I, A184V, S226R, H236Q, E252G, T268A A290V, L353V, I366V, E442K

ABC-CIS: BM3-CIS + C400S

IV. Rationale for the C400S Mutation

The very negative $Fe^{III/II}$ potential of cytochromes P450 relative to other heme proteins has been ascribed to the strong donating character of the axial cysteinate ligand. This effect has been modeled in cytochrome c: substitution of the native axial methionine for cysteine decreases the $Fe^{III/II}$ potential by an impressive 652 mV, from 262 vs SHE for the Met/His ligated variant to -390 mV vs SHE for the Cys/His variant (*31*). Even within cytochrome P450, the reduction potential can be further reduced by increasing the electron donating character of the cysteinate ligand. Removal of a single amide proton proposed to stabilize the cysteinate negative charge shifts the $Fe^{III/II}$ potential negative by 35-45 mV with respect to wild-type (*32*).

In order to facillitate cyclopropanation activity in vivo, it was necessary to shift the reduction potential sufficiently positive to allow reduction by NADPH. In cytochrome c, it was observed that axial ligation by a weakly donating water molecule raises the reduction potential of the His/H_2O ligated variant ($Fe^{III/II}$: -45 mV vs SHE) by 345 mV compared to the Cys/His variant (*31*). To that end, we hypothesized in P450, substitution of the cysteinate axial ligand with the weakly donating serine alcohol would be expected to shift the C400S reduction potential positive compared to wild-type. The pK_a of serine (~15) is approximately 7 pH units above that of cysteine (~8), and so while cysteine remains deprotonated as the cysteinate ligand in both ferric and ferrous states of the enzyme, we anticipated that serine would remain protonated in at least the ferrous form. The serine-ligated mammalian P450 mutant has been suggested to be serinate in the ferric form, and serine in the ferrous form based on analysis of absorption spectra and magnetic circular dichroism (*16*), and we anticipated that similar ligation in Ser-P450-BM3 would result in a more positive $Fe^{III/II}$ potential.

V. Physical Characterization of ABC-CIS

V.I. X-Ray Crystallography Statistics

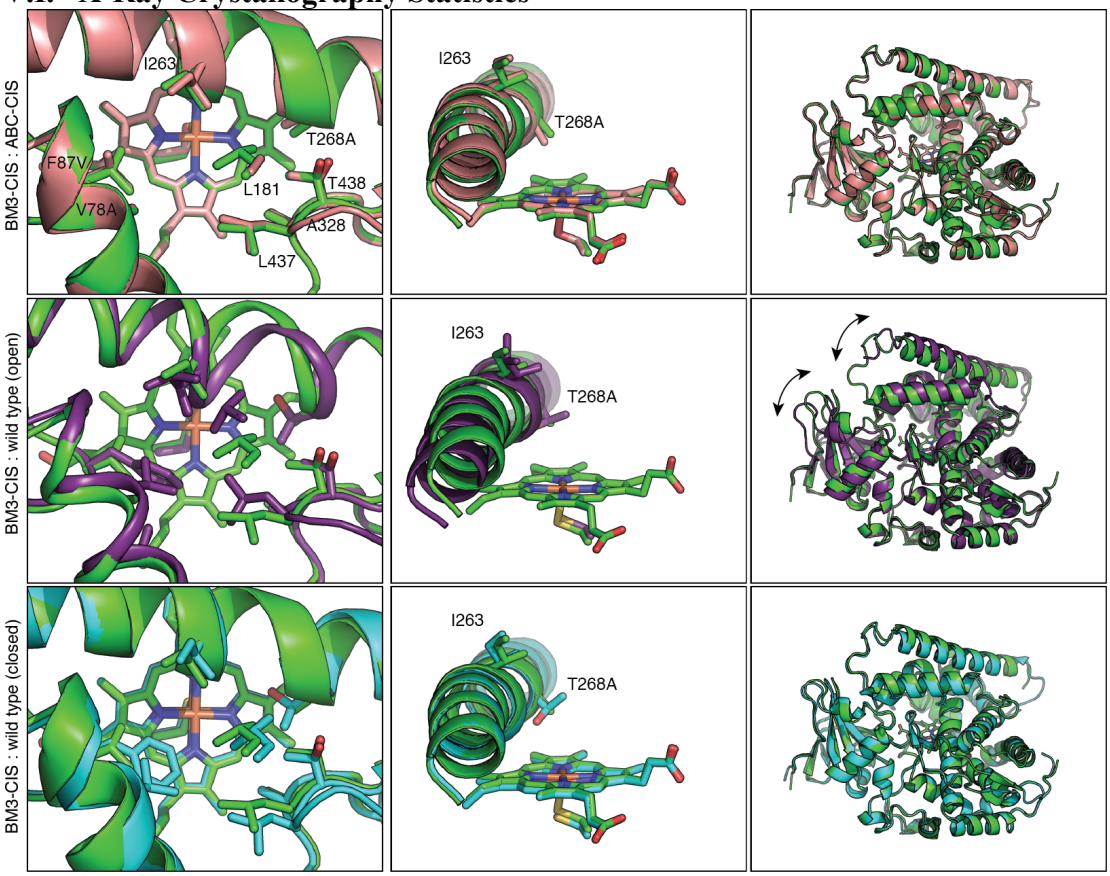


Figure S1. Active site and protein alignments of BM3-CIS with ABC-CIS and wild type $P450_{BM3}$.

To investigate the nature of enhanced stereoselectivity in ABC-CIS, crystal structures of both proteins were determined to assess any structural changes that may have occurred due to the axial Cys → Ser mutation. The top panels shows alignments of BM3-CIS (green) and ABC-CIS (peach) with left, middle, and right panels showing active site residues, the active site I-helix, and global protein fold, respectively. No significant structural changes were observed (RMSD 0.52 Å). Middle panels: Large variations are observed upon comparing BM3-CIS with the open (ligand-free) form of wild type BM3 (purple, taken from PDB# 2IJ2, RMSD 1.2 Å). Pronounced

rearrangements are observed in active site side chain residues (left) as well as rotations within the I-helix. Global movements are also observed in the N-terminal beta domain as well as F- and G-helices (right, marked by double-headed arrows). These movements are consistent with well-known transitions that occur upon substrate binding and are important for native monooxygenation catalysis. Bottom panels: Alignment of BM3-CIS with a ligand-bound BM3 structure (cyan, taken from PDB# 1JPZ, RMSD 0.52 Å) demonstrates that BM3-CIS and ABC-CIS mimic the closed protein conformation even in the absence of substrate.

Table S1. **Data collection and refinement statistics for $P450_{BM3}$ crystals**

	BM3-CIS	ABC-CIS
pdb accession #	**4H23**	**4H24**
Data collection*		
Space group	I 1 2 1	P 2 21 21
wavelength	1.033	1.033
Cell dimensions		
a, *b*, *c* (Å)	187.79, 62.74, 210.28	63.16, 124.46, 127.69
α, β, γ (°)	90.00, 115.75, 90.00	90.00, 90.00, 90.00
Resolution (Å)	48.6 - 2.5 (2.5 - 2.6) **	44.9 – 3.3 (3.3 – 3.5) **
R_{merge}	5.3(39.5)	17.6(51.4)
$I / \sigma I$	13.4(3.0)	11.8(5.7)
Completeness (%)	98.7(99.2)	99.9(99.9)
Redundancy	2.6(2.6)	5.3(5.4)
Refinement		
Resolution (Å)	48.6 - 2.5	44.9 – 3.3
No. reflections	72085	14884
R_{work} / R_{free}	0.19 / 0.25	0.18/0.26
No. atoms		
Protein	14401	6890
Ligand/ion	128	86
Water	196	24
B-factors		
Protein	33.9	25.4
Ligand/ion	25.4	19.9
Water	26.7	18.9
R.m.s. deviations		
Bond lengths (Å)	0.017	0.016
Bond angles (°)	1.70	1.65
Ramachandran outliers***	0.3%	0.7%

*All data sets were collected from single crystals.
**Highest-resolution shell is shown in parentheses.
*** Ramachandran outliers lie in regions of protein that are known to be flexible and show similar disorder among $P450_{BM3}$ structures in the literature.

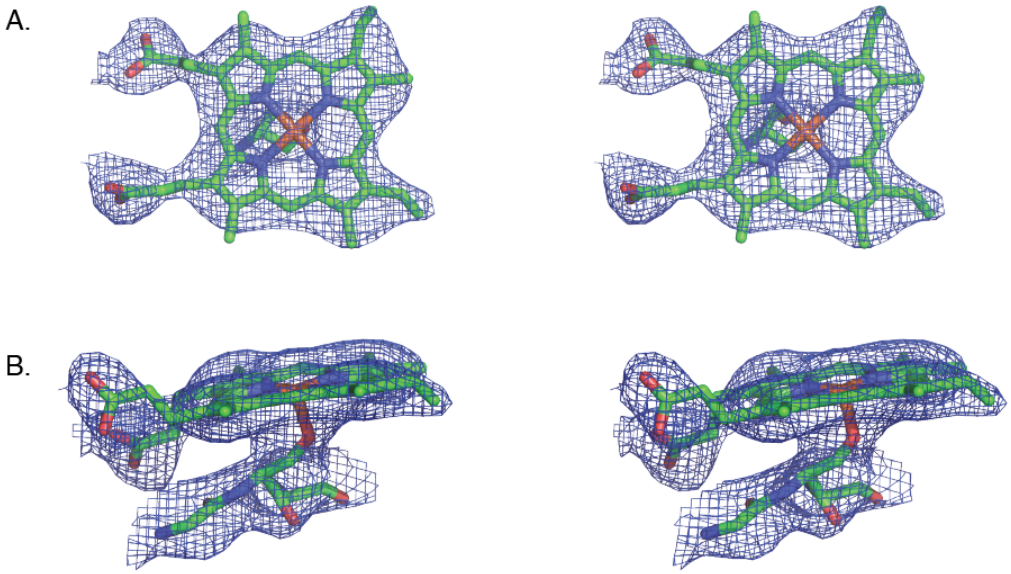


Fig. S2. Heme electron density of ABC-CIS: Maximum likelihood weighted electron density maps of serine-ligated heme in ABC-CIS. (A) Stereo image of heme bound to ABC-CIS viewed from the top of the heme in the active site. (B) Stereo image of heme bound to ABC-CIS rotated ~90° from panel A shows clear indication of heme-iron ligation by the side chain hydroxyl of Cys400Ser. All atoms are shown as sticks. All electron density maps were contoured at $\sigma = 1.0$. For perspective, in panel B, the main chain atoms of residues 399 and 401 are also shown as sticks.

V.II. UV-vis Spectroscopy

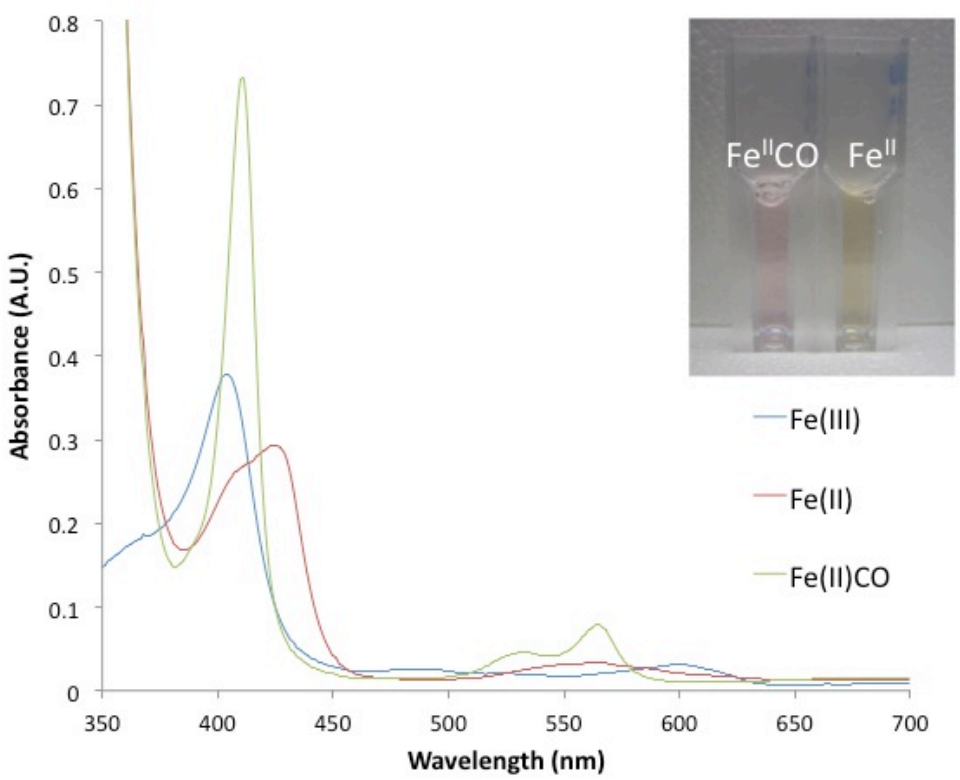


Figure S3. Absolute spectra for ferric (blue), dithionite reduced ferrous (red) and carbon monoxide bound ferrous (green) ABC-CIS_{heme}. Soret bands (nm): Fe^{III}, 404; Fe^{II}, 410, 425; Fe^{II}-CO, 411. Fe^{II}-CO displays α and β bands at 532 and 565 nm. Insert shows the carbon monoxide ferrous (pink) and the dithionite reduced ferrous (yellow) enzymes at 4.5 μM protein concentration.

The shoulder at 410 nm for the ferrous spectrum is due to incomplete reduction to FeII under the aerobic conditions in which these spectra were taken. Reduction of cytochrome P411 under strict anaerobic conditions, as is the case for the redox titrations (figures S7 and S9) gives a single peak at 421 nm.

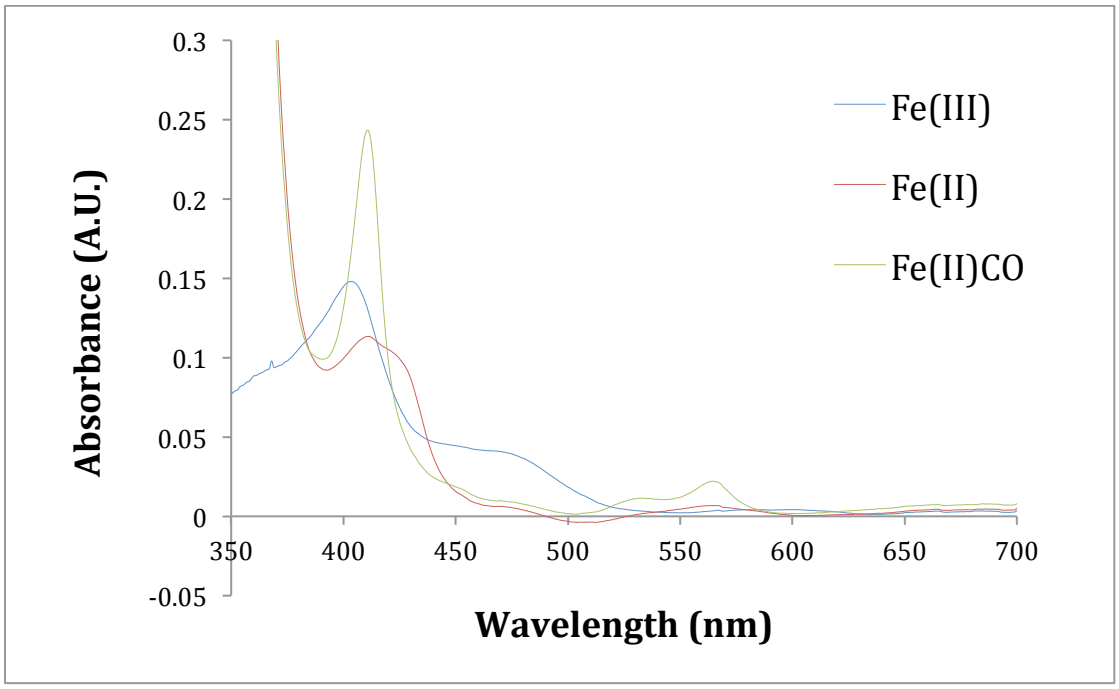


Figure S4. Absolute spectra for ferric (blue), dithionite reduced ferrous (red) and carbon monoxide bound ferrous (green) ABC-CIS$_{holo}$. Soret bands (nm): Fe^{III}, 404; Fe^{II}, 410, 422; Fe^{II}-CO, 411. Fe^{II}-CO displays α and β bands at 533 and 566 nm. Ferric spectrum displays a broad peak at 465 nm.

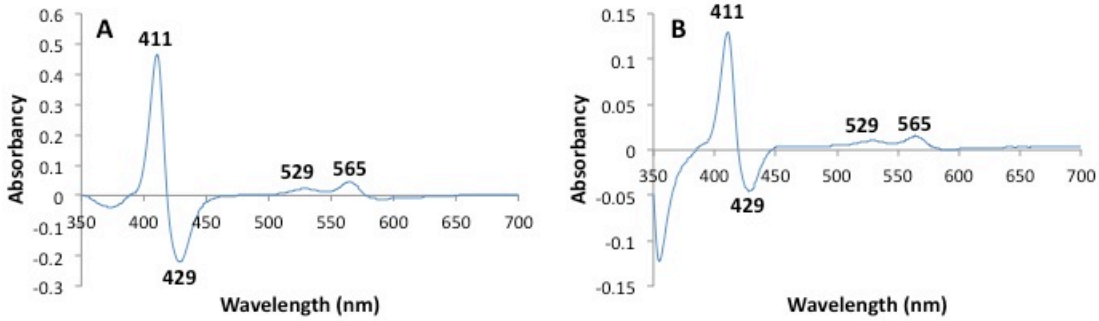


Figure S5. Difference spectra for ferrous carbonyl with respect to ferrous for (**A)** ABC-CIS$_{heme}$ and (**B**) ABC-CIS$_{holo}$.

Table S2: Comparison of λmax for ABC-CISheme and CYP2B4-C436S (*15*)

	ABC-CIS$_{heme}$ (nm)	**CYP2B4-C436S (nm)**
Ferric resting state	404	405
Ferrous	425	422
Ferrous-CO	411	413

V.III. Redox Titrations

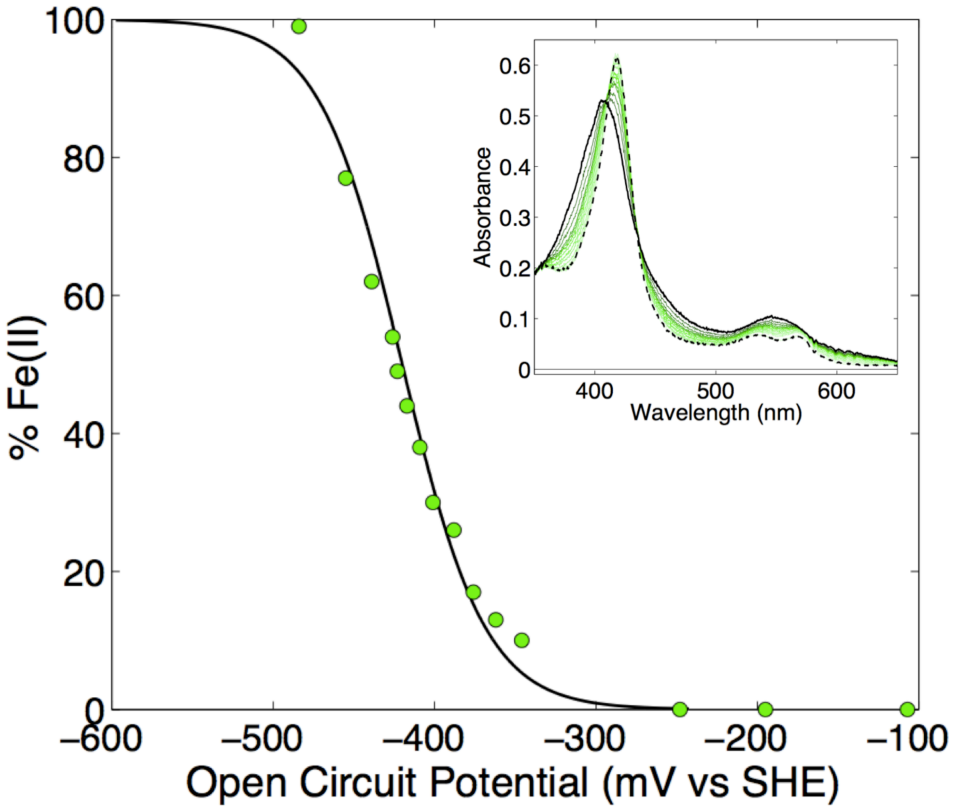


Figure S6. Potentiometric Titration for $BM3_{heme}$; $E°' = -420$ mV.

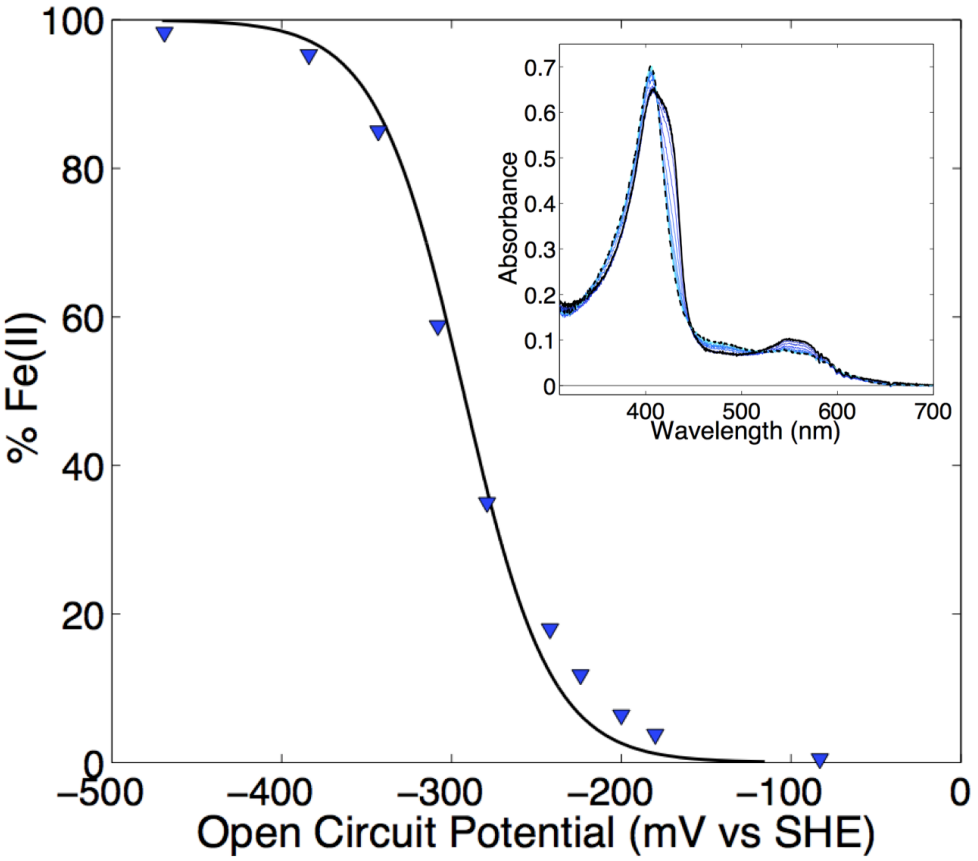


Figure S7. Potentiometric titration of ABC_{heme}; $E°' = -293$ mV.

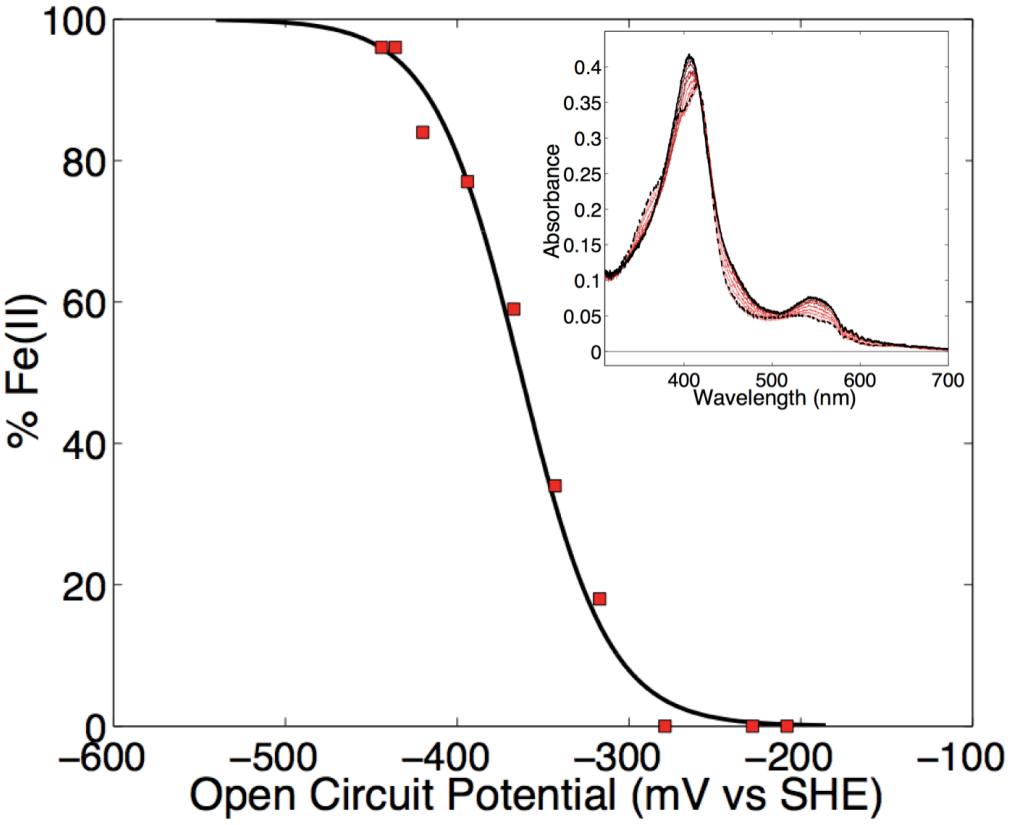


Figure S8. Potentiometric Titration of BM3-CIS$_{heme}$; $E°'$ = −360 mV.

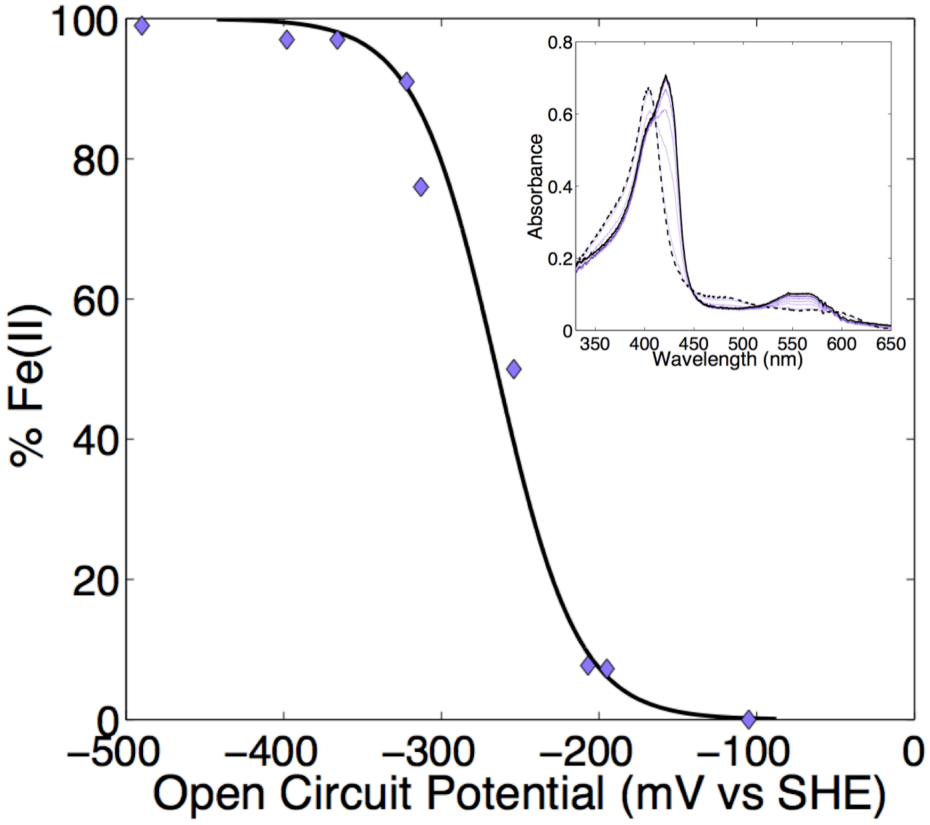


Figure S9. Potentiometric Titration of ABC-CIS$_{heme}$; $E°' = -265$ mV.

V.IV. Effect of C400S on Thermostability

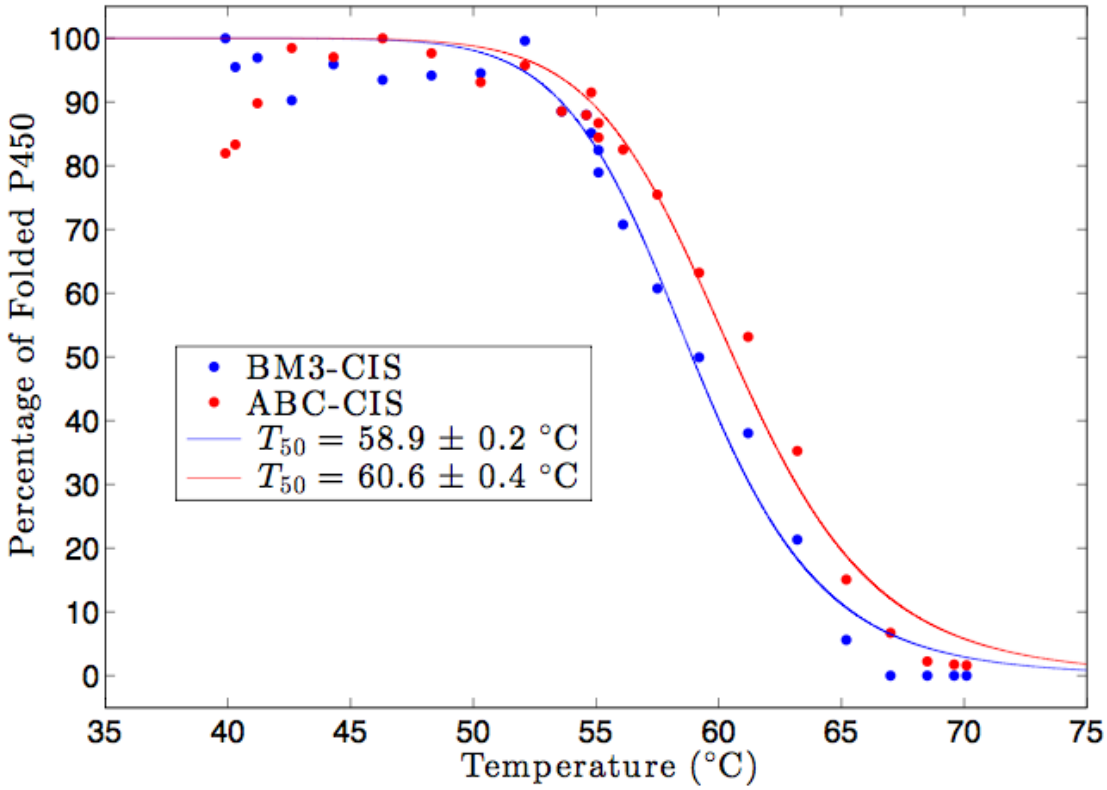


Figure S10. Thermostabilities of heme domains of BM3-CIS (blue) and ABC-CIS (red). The C400S mutation stabilizes the heme domain by +1.7 °C. The T_{50} is the temperature at which half of the enzyme population has unfolded. Error bars correspond to 1-σ (68.3%) confidence intervals for the T_{50}.

VI. *In vitro* Cyclopropanation Activities of ABC-CIS and ABC

VI.I. Michaelis-Menten Kinetics

Determination of initial rates. Both styrene and EDA concentrations were varied in the presence of the enzymes expressed as the heme-domain (0.5 or 1.0 μM BM3-CIS$_{heme}$). Reactions were set up in phosphate buffer (pH = 8.0) with $Na_2S_2O_4$ as the reductant at 298 K, and were worked up as described in section II. Three time points were taken and used to determine the rate of product formation by GC (cyclosil-B 30 m x 0.32 mm x

0.25 µm): oven temperature = 100 °C 5 min, 5 °C / min to 200 °C, 20 °C / min to 250 °C, 250 °C for 5 min. Elution time: *cis*-cyclopropanes (19.20 min and 19.33 min), *trans*-cyclopropanes (20.44 min). Apparent kinetic parameters were determined by fitting the data to the standard Michaelis-Menten model.

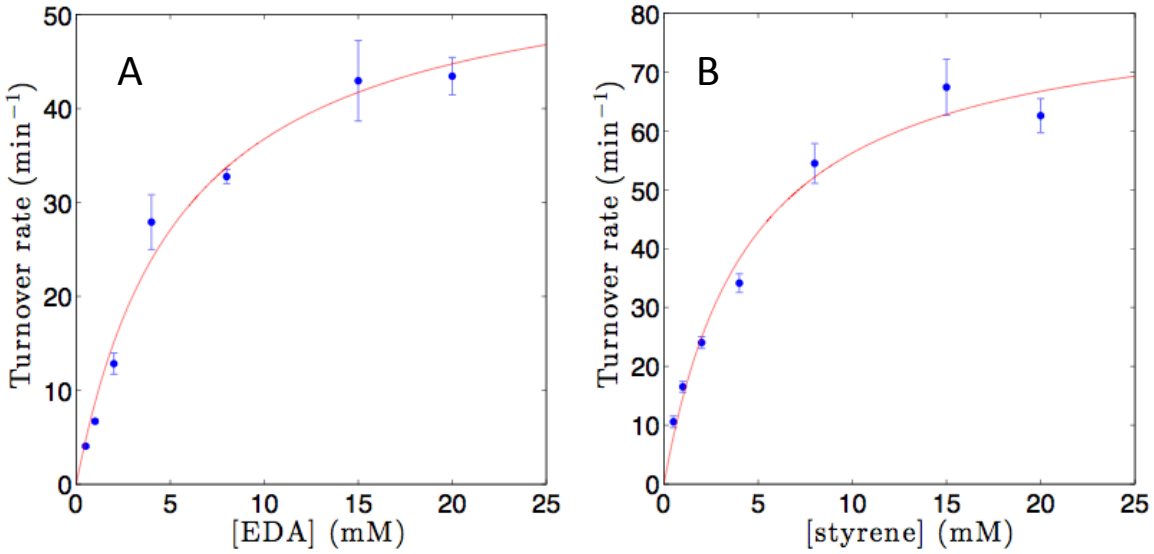


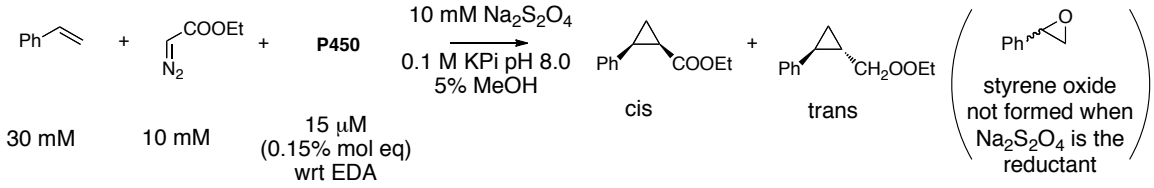
Figure S11. Initial velocities plot for ABC-CIS$_{heme}$. (**A**) EDA concentration was varied at a saturating concentration of styrene (30 mM). (**B**) Styrene concentration was varied at a fixed concentration of EDA (20 mM). Initial rates were computed as the slope of a zero-intercept linear fit of three different time points from independent reactions. Error bars correspond to 1-σ (68.3%) confidence intervals for the slope.

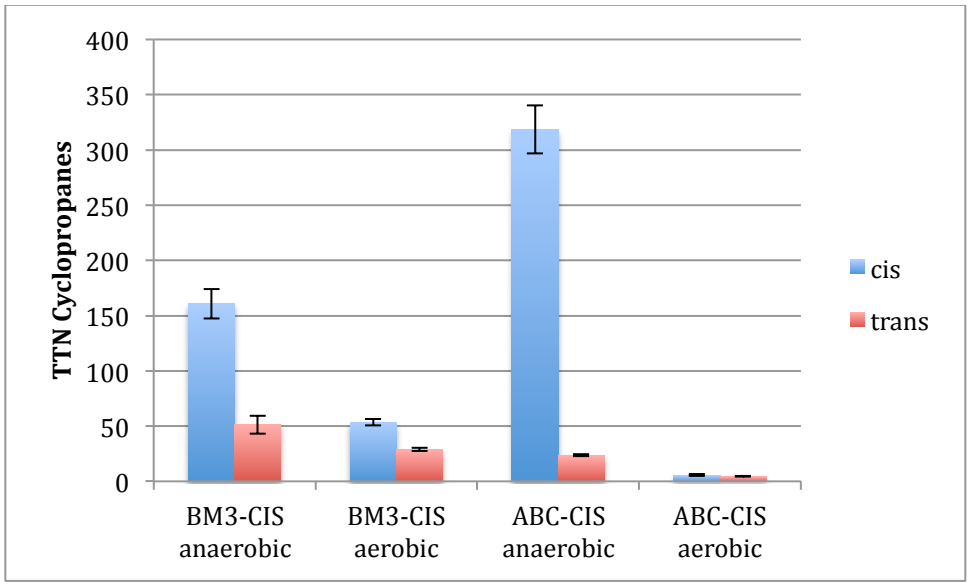
Table S3. Michaelis-Menten parameters for P450 cyclopropanation catalysts. Error bars correspond to 99% confidence intervals for the fitted parameters.

Catalyst	k_{cat} (min^{-1})	$K_{M\text{-}EDA}$ (mM)	$K_{M\text{-}styrene}$ (mM)	k_{cat} / $K_{M\text{-}EDA}$ (s^{-1} M^{-1})	k_{cat} / $K_{M\text{-}styrene}$ (s^{-1} M^{-1})	k_{cat} / ($K_{M\text{-}EDA}$ x $K_{M\text{-}styrene}$) (s^{-1} M^{-1} M^{-1})
BM3-CIS$_{heme}$ (*5*)	100 ± 24	5.2 ± 3.5	1.4 ± 0.5	320	1,100	2.1 x 10^{5}
ABC-CIS$_{heme}$	82 ± 15	5.7 ± 2.9	4.6 ± 2.4	240	300	5.5 x 10^{4}

VI.II. Enhanced *Cis* Selectivity and Substrate Scope

Activity under Anaerobic versus Aerobic Conditions with Sodium Dithionite as the Reductant





Catalyst	Conditions	Yield (%)*	TTN	O_2 inhibition (%)	*cis*:*trans*†	%ee *cis*‡	%ee *trans*§
BM3-CIS_{heme}	Anaerobic	32	212 ± 20	-	77:23	−94	−91
BM3-CIS_{heme}	Aerobic	12	83 ± 5	−61	65:35	−87	−86
ABC-CIS_{heme}	Anaerobic	51	342 ± 21	-	93:7	−99	−51
ABC-CIS_{heme}	Aerobic	2	10 ± 1	−97	45:55	−79	−31

* Based on EDA. † Diastereomeric ratios and enantiomeric excess were determined by GC analysis. ‡ (2*R*,1*S*) – (2*S*,1*R*). § (2*R*,1*R*) – (2*S*,1*S*).

Figure S12. Cyclopropanation activities of BM3-CIS and ABC-CIS driven by sodium dithionite under anaerobic and aerobic conditions. Measurements were taken in triplicate, and the error bars represent the standard deviation of the measurements.

Small-scale reactions. Selected P450 catalysts were surveyed at a small-scale reaction (400 μL total volume) for each combination of reagents (olefins and diazo esters). The small-scale anaerobic bioconversions were conducted as described in section II and were analyzed by GC. GC methods for these products are reported in reference (*5*).

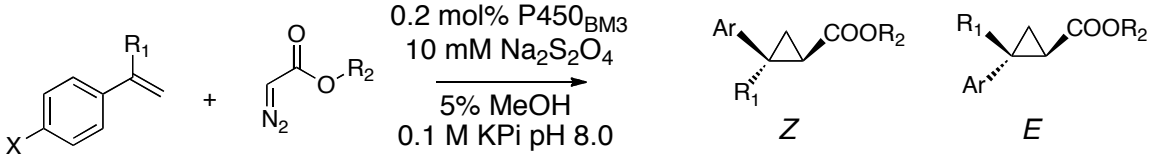


Table S4. Enhanced *Z* selectivity for ABC-CIS over BM3-CIS

Reagents	P450	% yield*	TTN	*Z*:*E*†	%ee *Z*‡	%ee *E*‡
R_1 = H, X = Me, R_2	BM3-CIS	46	228	78:22	-81.4	N/A
= Et	ABC-CIS	32	157	93:7	-87.1	N/A
R_1 = H, X = OMe,	BM3-CIS	43	214	48:52	-44	N/A
R_2 = Et	ABC-CIS	47	235	81:19	-61	N/A
R_1 = H, X = CF_3, R_2	BM3-CIS	42	211	39:61	54	-93
= Et	ABC-CIS	24	121	76:24	55	-75
R_1 = Me, X = H, R_2	BM3-CIS	26	127	16:84	-6	N/A
= Et	ABC-CIS	17	86	30:70	34	N/A
R_1 = H, X = H, R_2 =	BM3-CIS	0.3	2	3:97	N/A	N/A
t-Bu	ABC-CIS	15	76	8:92	N/A	N/A

* Based on EDA. † Diastereomeric ratios and enantiomeric excess were determined by GC analysis. ‡Enantiomeric excess is only reported when the enantiomers resolved to baseline resolution.

VI.III. Monooxygenation versus Cyclopropanation Activities for BM3-CIS and ABC-CIS

Activity under anaerobic versus aerobic conditions with NADPH as the reductant. Small-scale reactions (400 μL total volume) were conducted as described in section II with the following modifications: glucose dehydrogenase (GDH, 4 μL, 225 U mL^{-1}) was added to the reaction vial together with the P450 solution. Glucose (40 μL, 250 mM) and NADPH (40 μL, 5 mM) were degassed together with the buffer solution. Measurements were taken in triplicate and the error bars represent the standard deviation from the mean value.

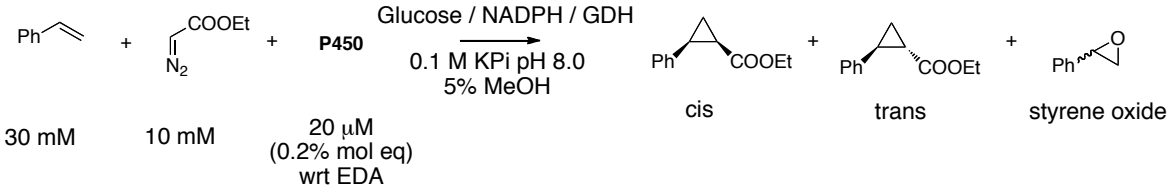


Table S5. *In vitro* activities for purified ABC-CIS vs BM3-CIS driven by NADPH

Cat.	Conditions	Cyclopropanes (TTN_{cyc})	Styrene oxide (TTN_{epo})	TTN_{cyc} / TTN_{epo}	O_2 inhibition (%)	*cis*:*trans*[*]	%ee *cis*[†]	%ee *trans*[‡]
BM3-CIS	Anaerobic	60 ± 18	12 ± 10	5	-	60:40	−89	−53
BM3-CIS	Aerobic	82 ± 13	406 ± 21	0.20	+36	56:44	−88	−58
ABC-CIS	Anaerobic	304 ± 15	0	-	-	72:28	−92	−19
ABC-CIS	Aerobic	43 ± 5	14 ± 2	3.1	-86	49:51	−74	−14

Diastereomeric ratios and enantiomeric excess were determined by GC analysis.
[†] (2*R*,1*S*) – (2*S*,1*R*). [‡] (2*R*,1*R*) – (2*S*,1*S*).

VI.IV. Choice of Reductant: NADPH vs NADH

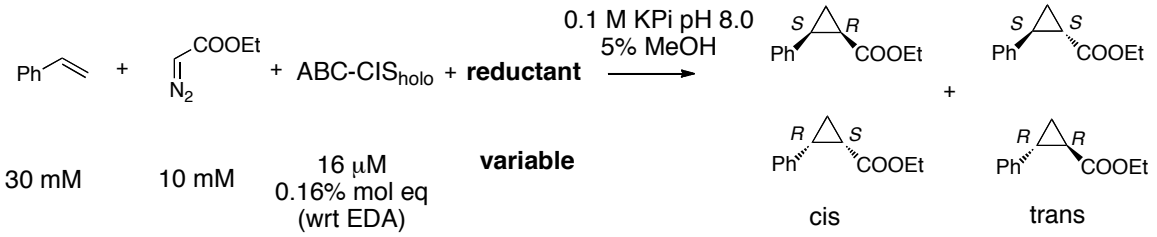


Table S6. *In vitro* ABC-CIS$_{holo}$ cyclopropanation driven by $Na_2S_2O_4$, NADPH and NADH

[NADPH] / mM	[NADH] / mM	[$Na_2S_2O_4$] / mM	Yield (%)*	TTN	*cis*:*trans*[†]	%ee *cis*[‡]	%ee *trans*[§]
0	0	0	1	3	54:46	-	-
0	0	10	35	218	94:6	−99	−43
1	0	0	47	294	93:7	−98	−40
5	0	0	58	364	90:10	−97	−28
10	0	0	49	305	91:9	−98	−26
0	1	0	70	437	93:7	−98	−38
0	5	0	68	428	93:7	−98	−34
0	10	0	47	295	93:7	−98	−34

* Based on EDA. [†] Diastereomeric ratios and enantiomeric excess were determined by GC analysis. [‡] (2*R*,1*S*) – (2*S*,1*R*). [§] (2*R*,1*R*) – (2*S*,1*S*).

VI.V. *In vitro* Cyclopropanation Activities of BM3 and ABC

Small-scale (500 μL) reactions were conducted as described in section II with purified BM3 and ABC catalysts.

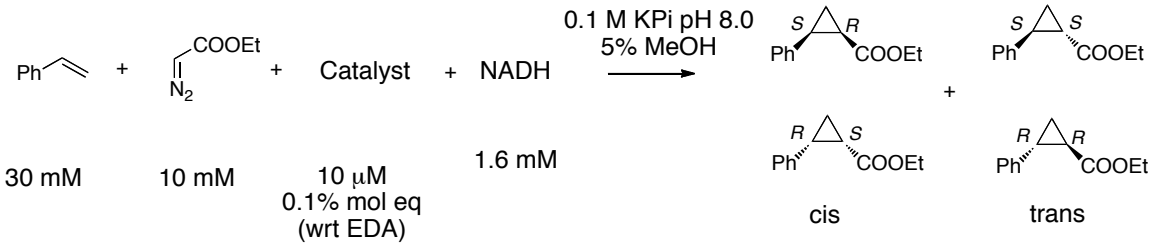


Table S7. In vitro cyclopropanation activities of BM3 and ABC driven by NADH

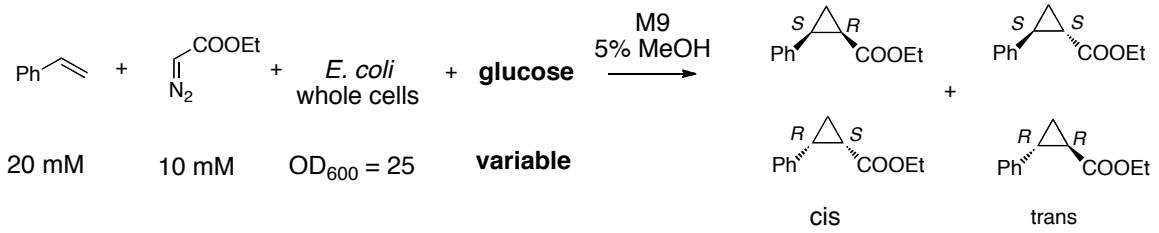
Catalyst	Yield (%)*	TTN	*cis*:*trans*[†]	%ee *cis*[‡]	%ee *trans*[§]
BM3	0.2	2 ± 0.5	17:83	−29	−21
ABC	36	364 ± 50	12:88	−5	−1

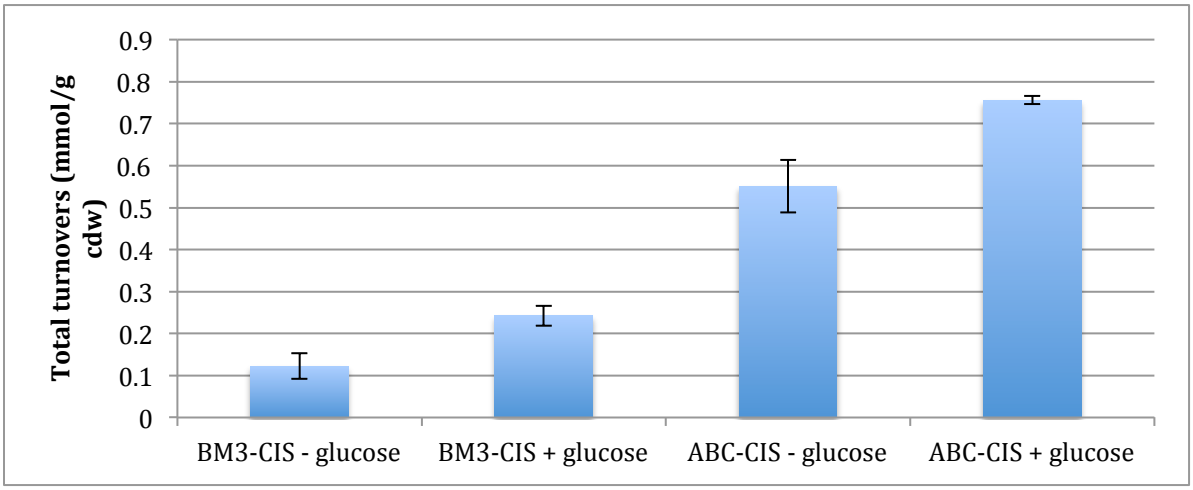
* Based on EDA. [†] Diastereomeric ratios and enantiomeric excess were determined by GC analysis. [‡] (2*R*,1*S*) – (2*S*,1*R*). [§] (2*R*,1*R*) – (2*S*,1*S*)

VII. Whole-Cell Cyclopropanation Catalysts

All experiments using whole-cells were done in triplicate; the error bars represent the standard deviation of the measurements. “Total turnovers” is defined here as the amount of cyclopropane product (mmol) formed per mass of catalyst (g_{cdw}).

VII.I. Effect of Glucose Addition



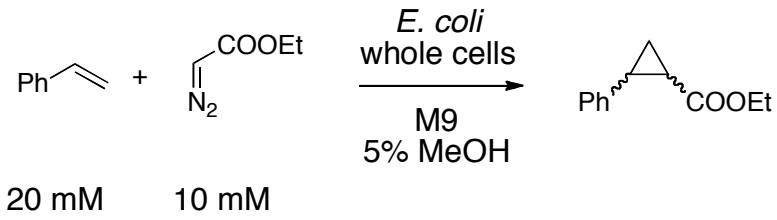


Whole-Cell Catalyst	[Glucose] (mM)	Yield (%)*	Total Turnover (mmol g_{cdw}^{-1})	Effect of glucose addition (%)	Cell density (g_{cdw} L^{-1})	*cis*:*trans*[†]	%ee *cis*[‡]	%ee *trans*[§]
BM3-CIS	0	7	0.120 ± 0.03	-	5.53	34:66	−73	−22
BM3-CIS	2	13	0.240 ± 0.02	+98	5.53	48:52	−86	−30
ABC-CIS	0	35	0.550 ± 0.06	-	6.38	70:30	−95	−11
ABC-CIS	2	48	0.760 ± 0.01	+37	6.38	76:24	−96	−14

* Based on EDA. [†] Diastereomeric ratios and enantiomeric excess were determined by GC analysis. [‡] (2*R*,1*S*) – (2*S*,1*R*). [§] (2*R*,1*R*) – (2*S*,1*S*).

Figure S13. Effect of adding exogenous glucose (2 mM) on olefin cyclopropanation catalyzed by *E. coli* whole cells.

VII.II. Anaerobic vs Aerobic Reaction Conditions



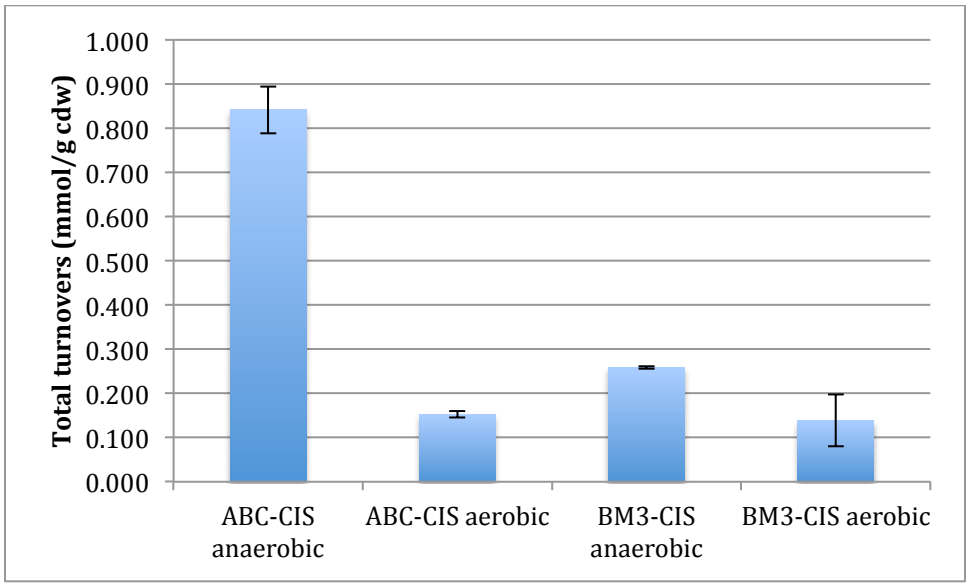
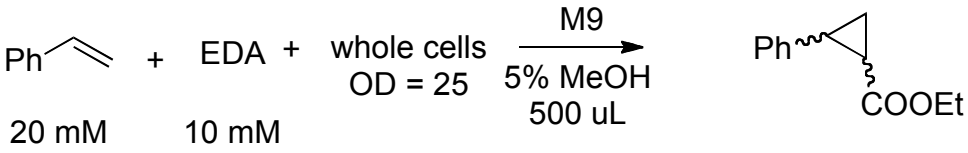
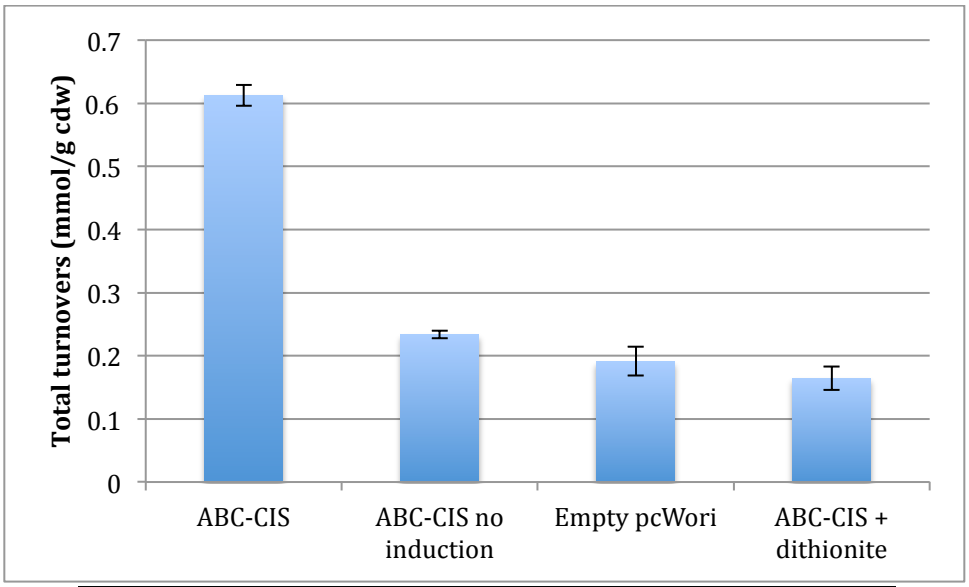


Figure S14. Effect of dioxygen exposure on whole-cell catalyzed cyclopropanation. ABC-CIS is strongly inhibited by dioxygen *in vivo*. All reactions had a cell density equivalent to OD_{600} = 25. Reactions were conducted in the absence of exogenous glucose.

VII.III. Empty Plasmid, No Induction Controls and Dithionite Addition to Whole-Cells



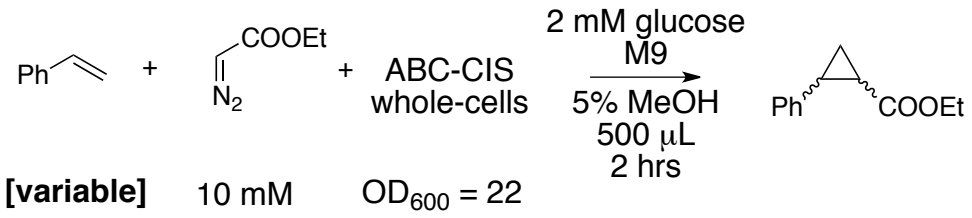


Catalyst	Yield (%)*	*cis*:*trans*[†]	%ee *cis*[‡]	%ee *trans*[§]
ABC-CIS	49	60:40	−93	−9
ABC-CIS no induction	16	17:83	−50	−9
Empty pcWori	15	11:89	−10	−12
ABC-CIS + dithionite	13	46:54	−87	−5

* Based on EDA. [†] Diastereomeric ratios and enantiomeric excess were determined by GC analysis. [‡] (2*R*,1*S*) – (2*S*,1*R*). [§] (2*R*,1*R*) – (2*S*,1*S*).

Figure S15. *E. coli* cells carrying the ABC-CIS gene but grown without the addition of IPTG **(**ABC-CIS no induction); *E. coli* cells carrying the pcWori plasmid but not the ABC-CIS gene (empty pcWori); ABC-CIS reaction with the addition of exogenous dithionite instead of glucose (ABC-CIS + dithionite). Reactions were left for two hours at 298 K.

VII.IV. Effect of Styrene Concentration



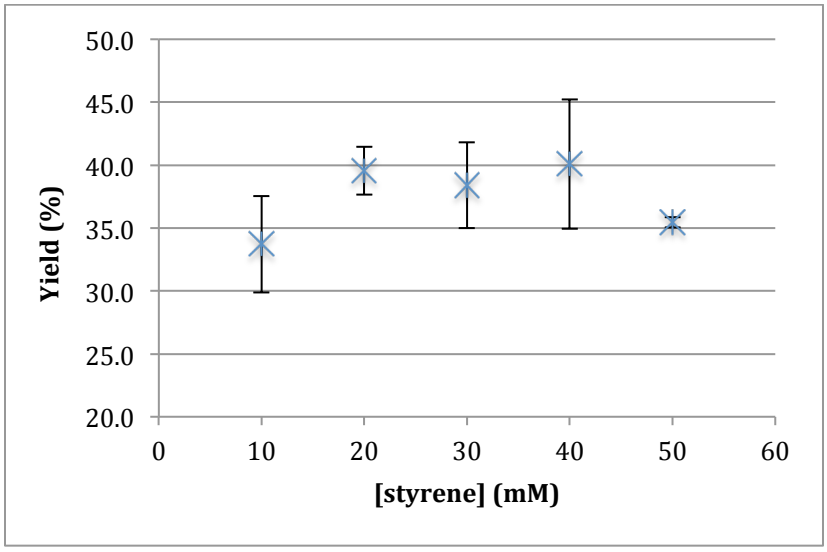
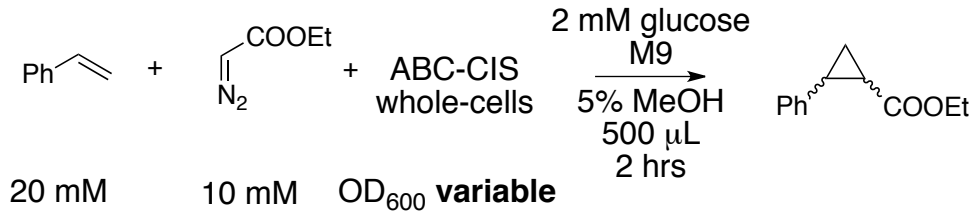
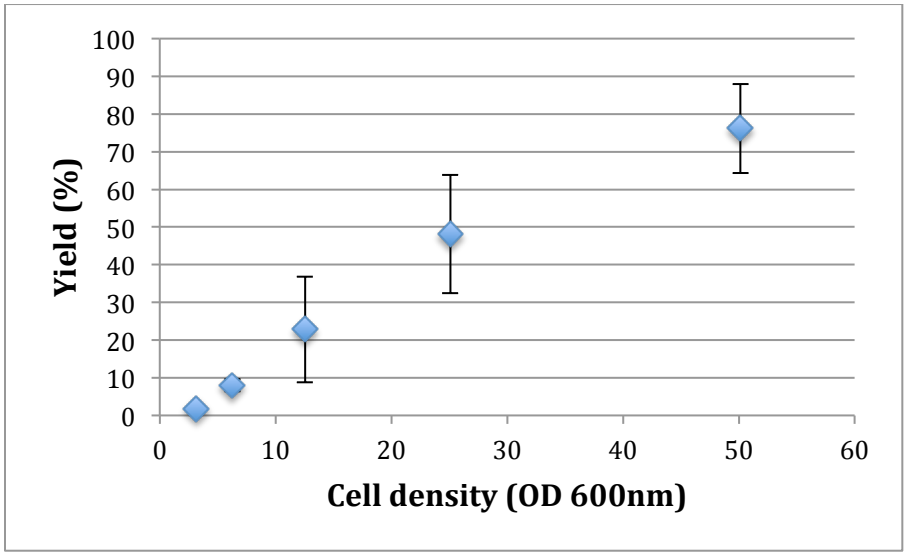


Figure S16. Effect of using 1, 2, 3, 4, and 5 equivalents of styrene on reaction yield. Excess styrene gives only small improvements in yield.

VII.V. Effect of Cell Density





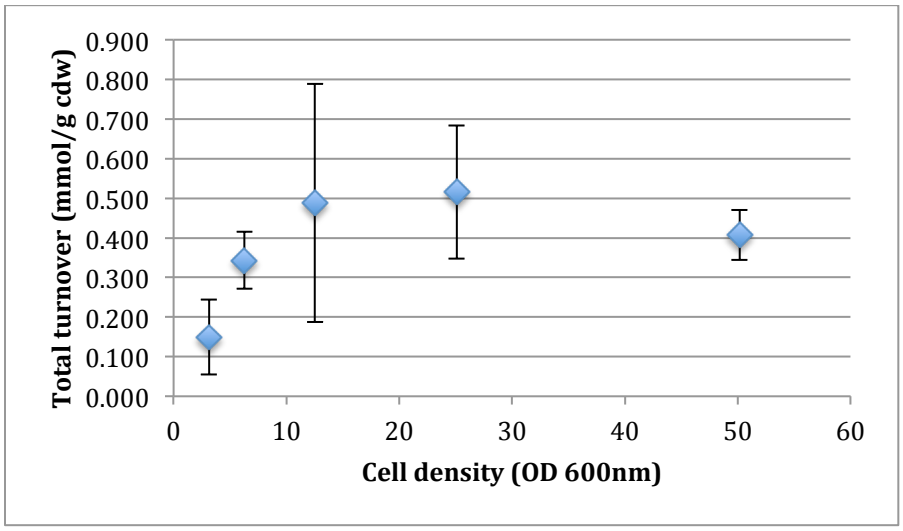


Figure S17. Increasing cell density increases cyclopropane yields up ~80%. Total turnovers do not increase for cell densities higher than OD_{600} = 20.

VII.VI. Lysate Compared to Intact Whole-Cells

Table S8. Lysate activity compared to *in vivo* activity

Catalyst	Conditions	[P411] (μM)	Yield (%)*	TTN	cis: trans†	%ee *cis*‡
ABC-CIS	*In vivo*	1.0	44	5120	80:20	−96
ABC-CIS	Lysate, no reductant	1.0	0.6	55	67:33	−92
ABC-CIS	Lysate + NADH	1.0	18	1780	80:20	−97
ABC-CIS	Lysate + dithionite	1.0	0.8	79	64:36	−86

* Based on EDA. † Diastereomeric ratios and enantiomeric excess were determined by GC analysis. ‡ (2*R*,1*S*) – (2*S*,1*R*).

VII.VII. Lyophilization of Whole-Cell Catalysts

Cells were lyophilized in 10% sucrose (m/V) and were stored at 4 °C for two weeks. An appropriate mass of the resulting powder was transferred to a 2 mL glass vial, which was crimp sealed and purged with argon. Degassed solutions of nitrogen-free M9 medium and glucose (20 mM) were added via syringe. Cells were resuspended to OD_{600} = 25 and 2 mM final concentration of glucose.

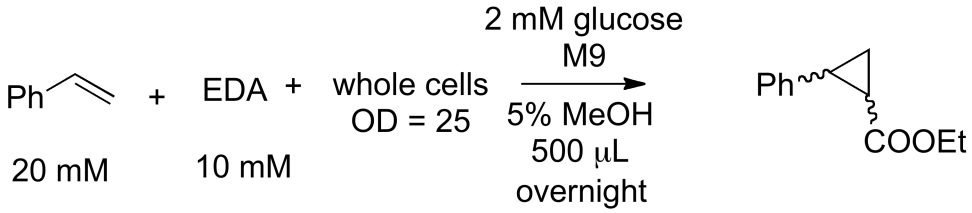


Table S9. Cyclopropanation activity of lyophilized ABC-CIS whole-cell catalysts.

Catalyst	[Cell density] (g_{cdw} L^{-1})	[P411] (μM)	Yield (%)*	Total Turnover (mmol g_{cdw}^{-1})	TTN	*cis*: *trans*[†]	%ee *cis*[‡]	%ee *trans*[§]
ABC-CIS	6.0	0.8	43	0.710 ± 0.08	5300 ± 600	67:33	−93	−25

* Based on EDA. [†] Diastereomeric ratios and enantiomeric excess were determined by GC analysis. [‡] (2*R*,1*S*) – (2*S*,1*R*). [§] (2*R*,1*R*) – (2*S*,1*S*).

VIII. References

1. P. K. Ajikumar *et al.*, Isoprenoid pathway optimization for taxol precursor overproduction in *Escherichia coli*. *Science* **330**, 70 (2010).
2. P. J. Westfall *et al.*, Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E111 (2012).
3. M. Kataoka *et al.*, Novel bioreduction system for the production of chiral alcohols. *Appl Microbiol Biotechnol* **62**, 437 (2003).
4. M. Boyce, C. R. Bertozzi, Bringing chemistry to life. *Nature Methods* **8**, 638 (2011).
5. P. S. Coelho, E. M. Brustad, A. Kannan, F. H. Arnold, Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **339**, 307 (2013).
6. H. Lebel, J.-F. Marcoux, C. Molinaro, A. B. Charette, Stereoselective Cyclopropanation Reactions. *Chemical Reviews* **103**, 977 (2003).
7. D. A. Evans, K. A. Woerpel, M. M. Hinman, M. M. Faul, Bis(oxazolines) as chiral ligands in metal-catalyzed asymmetric reactions. Catalytic, asymmetric cyclopropanation of olefins. *J. Am. Chem. Soc.* **113**, 726 (1991).
8. H. M. L. Davies, C. Venkataramani, Dirhodium Tetraprolinate-Catalyzed Asymmetric Cyclopropanations with High Turnover Numbers. *Organic Letters* **5**, 1403 (2003).
9. G. Maas, Ruthenium-catalyzed carbenoid cyclopropanation reactions with diazo compounds. *Chemical Society Reviews* **33**, 183 (2004).
10. T. W. B. Ost *et al.*, Phenylalanine 393 Exerts Thermodynamic Control over the Heme of Flavocytochrome P450 BM3. *Biochemistry* **40**, 13421 (2001).
11. D. S. Wuttke, H. B. Gray, Protein engineering as a tool for understanding electron transfer. *Curr. Opin. Struct. Biol.* **3**, 555 (1993).
12. C. J. Reedy, M. M. Elvekrog, B. R. Gibney, Development of a heme protein structure-electrochemical function database. *Nucleic Acids Research* **36**, D307 (2008).
13. See supplementary materials.
14. J. H. Dawson, Probing structure-function relations in heme-containing oxygenases and peroxidases. *Science* **240**, 433 (1988).
15. K. P. Vatsis, H.-M. Peng, M. J. Coon, Replacement of active-site cysteine-436 by serine converts cytochrome P450 2B4 into an NADPH oxidase with negligible monooxygenase activity. *Journal of Inorganic Biochemistry* **91**, 542 (2002).
16. R. Perera, M. Sono, H. L. Voegtle, J. H. Dawson, Molecular basis for the inability of an oxygen atom donor ligand to replace the natural sulfur donor heme axial ligand in cytochrome P450 catalysis. Spectroscopic characterization of the Cys436Ser CYP2B4 mutant. *Archives of Biochemistry and Biophysics* **507**, 119 (2011).
17. L. A. Wessjohann, W. Brandt, T. Thiemann, Biosynthesis and Metabolism of Cyclopropane Rings in Natural Compounds. *Chemical Reviews* **103**, 1625 (2003).

18. A. Penoni *et al.*, Cyclopropanation of olefins with diazoalkanes, catalyzed by CoII(porphyrin) complexes—a synthetic and mechanistic investigation and the molecular structure of CoIII(TPP)(CH2CO2Et) (TPP = dianion of meso-tetraphenylporphyrin). *European Journal of Inorganic Chemistry*, 1452 (2003).
19. N. Watanabe, H. Matsuda, H. Kuribayashi, S.-I. Hashimoto, Dirhodium(II) tetrakis[3(S)-phthalimido-2-piperidinonate]: a novel dirhodium(II) carboxamidate catalyst for asymmetric cyclopropanation. *Heterocycles* **42**, 537 (1996).
20. J. Sambrook, E. Frisch, T. Maniatis, *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor Laboratory Press, New York, 1989), vol. 2.
21. T. Omura, R. Sato, Carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. *Journal of Biological Chemistry* **239**, 2370 (1964).
22. W. Kabsch, Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica, Section D: Biological Crystallography* **D66**, 133 (2010).
23. P. Evans, Scaling and assessment of data quality. *Acta Crystallographica, Section D: Biological Crystallography* **D62**, 72 (2006).
24. D. C. Haines, D. R. Tomchick, M. Machius, J. A. Peterson, Pivotal Role of Water in the Mechanism of P450BM-3. *Biochemistry* **40**, 13456 (2001).
25. A. Vagin, A. Teplyakov, MOLREP: an automated program for molecular replacement. *Journal of Applied Crystallography* **30**, 1022 (1997).
26. S. Bailey, The CCP4 suite: programs for protein crystallography. *Acta Crystallographica, Section D: Biological Crystallography* **D50**, 760 (1994).
27. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallographica, Section D: Biological Crystallography* **D60**, 2126 (2004).
28. G. N. Murshudov, A. A. Vagin, E. J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallographica, Section D: Biological Crystallography* **D53**, 240 (1997).
29. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica, Section D: Biological Crystallography* **D66**, 213 (2010).
30. J. D. Bloom, S. T. Labthavikul, C. R. Otey, F. H. Arnold, Protein stability promotes evolvability. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 5869 (2006).
31. A. L. Raphael, H. B. Gray, Semisynthesis of axial-ligand (position 80) mutants of cytochrome c. *J. Am. Chem. Soc.* **113**, 1038 (1991).
32. S. Yoshioka *et al.*, Roles of the proximal hydrogen bonding network in cytochrome P450cam-catalyzed oxygenation. *J. Am. Chem. Soc.* **124**, 14571 (2002).

Chapter 4

INTRAMOLECULAR C-H AMINATION CATALYZED BY ENGINEERED CYTOCHROME P450 ENZYMES

Material from this chapter appears in J. A. McIntosh,* P. S. Coelho,* J. C. Lewis, C. Farwell, Z. J. Wang, T. R. Brown, and F. H. Arnold. "Enantioselective intramolecular C−H amination catalyzed by engineered cytochrome P450 enzymes *in vitro* and *in vivo*", submitted. *Authors contributed equally to this work. JAM conducted the 2D NMR experiments described in this chapter.

Abstract

Direct nitrene transfer to unactivated C-H bonds is a powerful method to introduce nitrogen atoms into organic scaffolds. Transition metal-catalyzed C-H amination proceeds through a nitrenoid intermediate with no mechanistic parallel in natural enzymes, but is isoelectronic with oxene transfers catalyzed by cytochromes P450. Engineered cytochrome $P450_{BM3}$ enzymes bind arylsulfonyl azides with micromolar affinity and show promiscuous levels of intramolecular benzylic C-H amination activity with these synthetic reagents. Formation of dimerized products and reduction of the sulfonyl azide to the sulfonamide are the major competing reactions. Sterically hindered multisubstituted arylsulfonyl azides favor C-H amination. These preliminary findings demonstrate the plasticity of cytochrome P450s to catalyze abiological bond disconnections.

Natural products and fine chemicals are often highly functionalized with amines and amides. The efficient installation of amines is thus of primary importance to organic synthesis. Methods for C-N bond formation often require functional group interconversion and the use of protecting groups, rendering such approaches redox and atom inefficient. The ability to insert nitrogen directly—via formal nitrene transfers—into unactivated C-H bonds allows for more convenient synthesis of otherwise difficult to make molecules (*1, 2*). Significant progress in this direction has been made by the development of organometallic catalysts that can transfer nitrene equivalents to C-H bonds with high levels of regio- and stereoselectivities (*3, 4*). These approaches, however, require organic solvents, expensive transition metals, elaborate ligand designs and often high catalyst loadings (>1% mol eq).

Nature makes amines using entirely different strategies (figure S1). For example, dehydrogenases and transaminases can convert a carbonyl to an amine (i.e., reductive amination), and aminomutases can catalyze the redox neutral radical rearrangement of α to β amino acids (*5*). Oxidative aminations, whereby a C-H bond is directly converted to an amine, are not known in nature. A chemomimetic approach to achieve enzymatic nitrene transfer can provide a more direct biosynthetic route to amines whilst complementing the existing synthetic methods. Since iron-porphyrin amination catalysts have been described (*6*), we reasoned that heme enzymes might also catalyze this reaction if provided with appropriate nitrene precursors. Furthermore, mammalian cytochrome P450s have been shown to catalyze trace levels of C-H amination from iminoiodane nitrene precursors (*7*), a transformation which is isoelectronic to the well established P450 catalyzed transfers of “oxenes” from iodosylbenzene.

We chose to use sulfonylazides instead of iminoiodanes as nitrene precursors due to their ease of synthesis, greater solubility in P450-compatible cosolvents, atom efficiency (since activation to the nitrene is accompanied by loss of only dinitrogen) and the demonstrated viability of these reagents for C-H amination (*8*). We focused our investigation on intramolecular C-H amination from arylsulfonyl azides so as to increase the likelihood of productive reaction (given the high local concentration of relatively weak benzylic C-H bonds in 2-substituted arylsulfonylazides). We chose to screen engineered variants of the bacterial cytochrome $P450_{BM3}$ (CYP102A1 or BM3) owing to its remarkable plasticity as well as the ready availability of many diverse and active variants in our laboratory (*5*).

We first screened a selection of 92 BM3 variants for binding to 2-isopropylbenzenesulfonyl azide (**1**), by monitoring changes in the ferric Soret peak in the UV-vis spectra, as a strategy to narrow our search for P450s with C-H amination activity. Many BM3 variants showed "type I" binding for azide **1** (figure S2), with spectral features characteristic for nondonor ligands (figure S3 to S6). We assessed several binders for reaction with azide **1** (2 mM) under anaerobic conditions in the presence of NADPH (0, 0.05, and 1 equivalent, with respect to azide), and 1 mole % (mol %) P450 in phosphate buffer with 2.5% DMSO cosolvent (figure 1). The desired benzosultam, however, was not formed in any of these reactions (figures 2-4). When NADPH is supplied in substoichiomteric amounts, some P450s are able to catalyze efficient formation of sulfonamide **2**. For example, variant 12-10C [16 amino acid mutations from BM3 (*5*)] makes **2** with ~60 TTN (TTN, total turnover number) when NADPH is supplied in 0.05 equivalents, and with ~20 TTN in the complete absence of exogenous

NADPH. NADPH-independent formation of sulfonamide **2** could be consistent with the hydrolysis of an "aza analogue" of compound I as proposed by White and McCarthy for the *p*-tosyl iminoiodane (PhINTs) dependent hydroxylation of cyclohexane (*9*) (figure S7). This mechanism would also explain the formation of the oxidative products [olefin **3** and alcohol **10** (*5*)] under anaerobic conditions that are formed in small amounts by 12-10C and other P450s (figures 3 to 5). Efficient reduction of the sulfonyl azide to the sulfonamide was readily achieved by various P450s when NADPH is supplied in one equivalent (figure S12). For example, wild-type BM3 (BM3) catalyzes formation of 2-isopropylbenzenesulfonamide (**2**) to 90% yield, corresponding to 90 TTNs (figure 4). Only variants H2A10 (15 mutations from BM3, SM) and H2-4-H5 (2 mutations from H2-A10, SM) did not form sulfonamide **2** as the major product with 1 eq. NADPH.

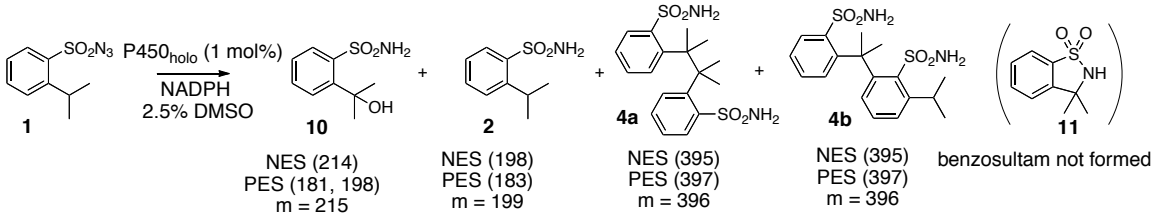


Figure 1. P450 bioconversions with 2-isopropylbenzenesulfonyl azide (**1**) under anaerobic conditions. NES = negative electrospray, PES = positive electrospray. Authentic standards for **7** and **9** were coinjected to assess formation of these compounds. Reaction conditions: 2 mM azide, 20 μM P450, oxygen depletion system (*5*) in 2.5% DMSO at 298 K for 12 hours. Reactions were analyzed by LC-MS.

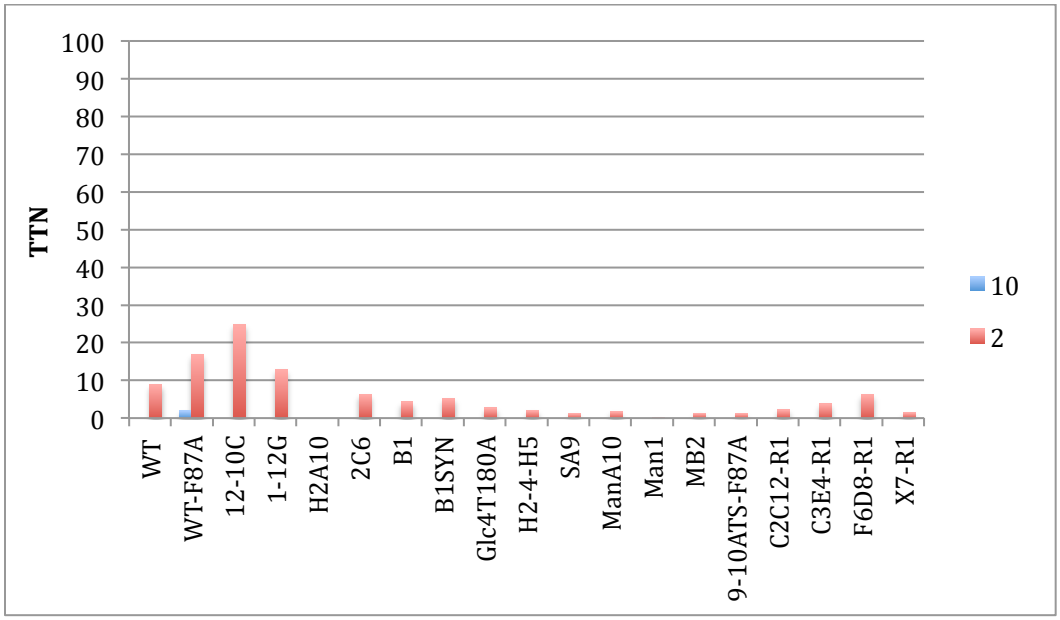


Figure 2. P450 reactions with azide **1** in the absence of NADPH. Alcohol **10** and arylsulfonamide **2** are defined in figure 1.

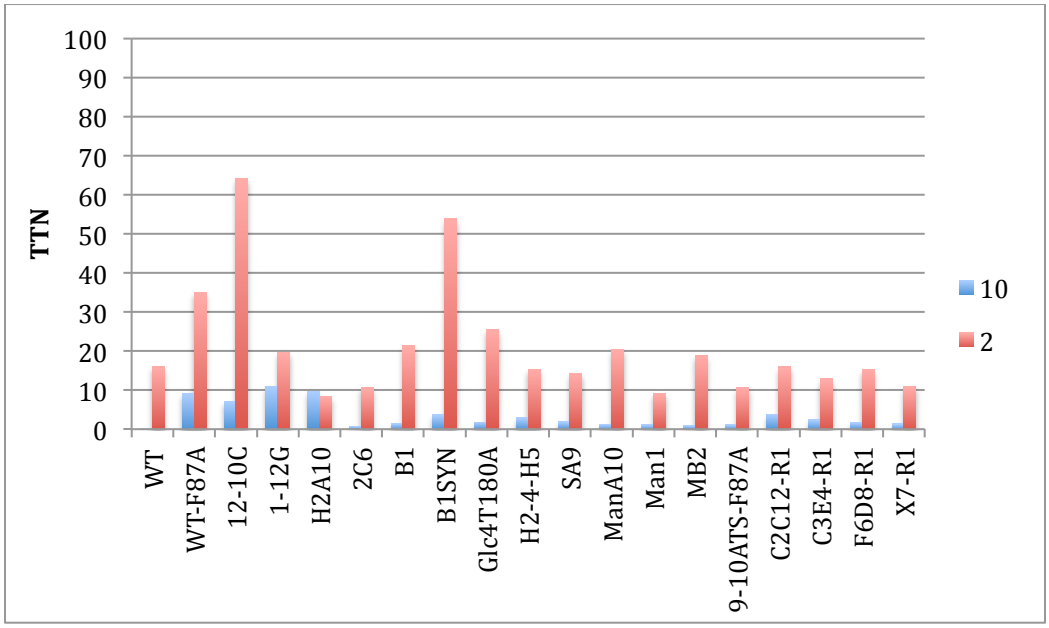


Figure 3. P450 reactions with azide **1** in presence of 0.1 mM NADPH (0.05 eq). Alcohol **10** and arylsulfonamide **2** are defined in figure 1.

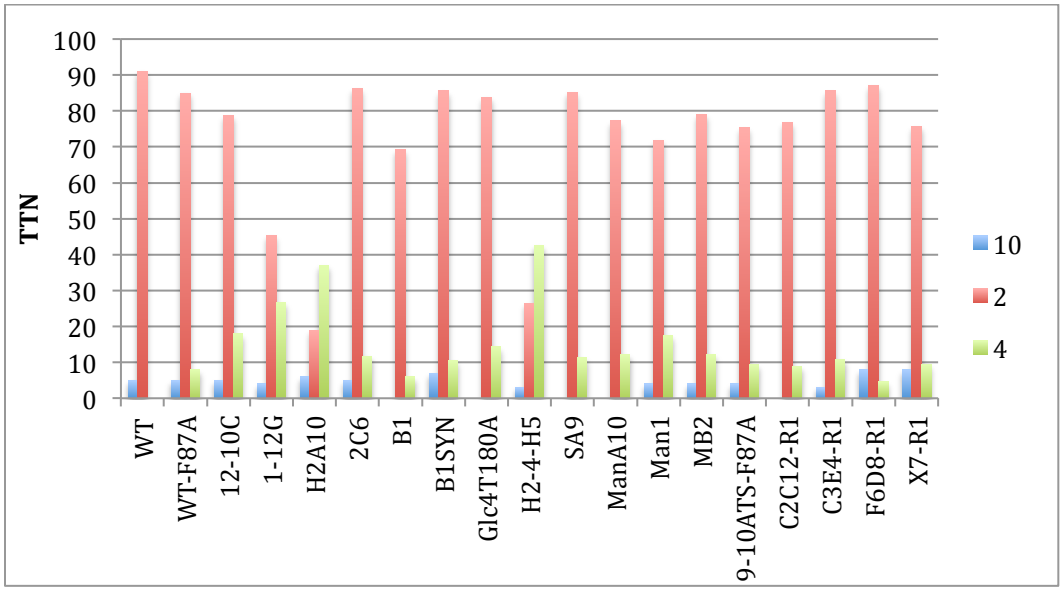


Figure 4. P450 reactions with azide **1** in presence of 2 mM NADPH (1 eq). Alcohol **10**, arylsulfonamide **2** and dimer **4** are defined in figure 1.

The effect of varying the concentration of NADPH on azide conversion and product distribution was further investigated for BM3, BM3-F87A, H2A10 and H2A10-A75L (figure S8). For BM3 and BM3-F87A, yields for sulfonamide **2** increased with increasing concentration of NADPH reaching close to 90% when stoichiometric amounts of NADPH are used. For H2A10 and H2A10-A75L, sulfonamide **2** was not the dominant product at any of the concentrations of NADPH investigated. H2A10-A75L formed alcohol **10** as the major product (20 TTN) when low concentrations of NADPH were used, but formed dimer **4a** (35 TTN) in the presence of a high concentration of NADPH. H2A10-A75L production of alcohol **10** is inhibited by dioxygen (table S1), confirming that **10** is not made via monooxygenation due to trace oxygen in the anaerobic reaction. Heat denaturation of H2A10-A75L significantly reduced production of alcohol **10** (table S1), confirming the involvement of the P450 in catalysis.

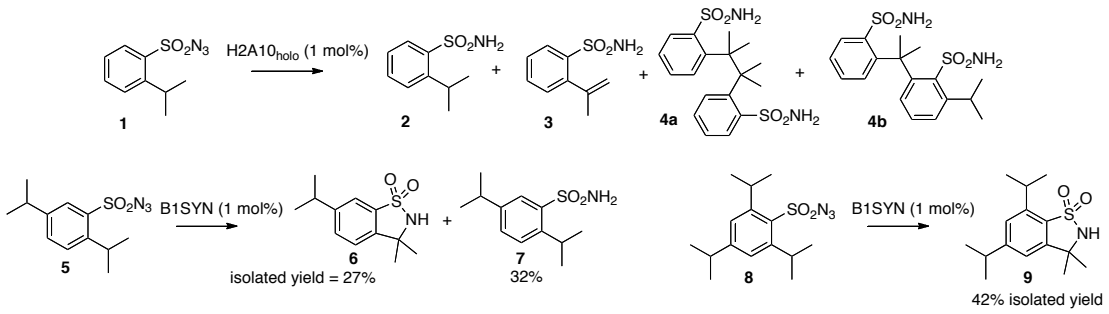


Figure 5. P450-catalyzed amination of benzylic C–H bonds from arylsulfonyl azides. Reaction conditions: 2 mM azide, 20 μM P450, 2 mM NADPH under argon in phosphate buffer (pH 8.0) for 4 hours at 298 K. Products isolated from small-scale (30 mg azide) bioconversions were analyzed by NMR and mass spectrometry.

We scaled up the reaction of azide **1** with H2A10, purified the products by reverse-phase HPLC, and characterized the by-products shown in figure 2 by NMR and mass spectrometry (*5*). Dehydrogenations to form alkenes, as observed for olefin **3**, were also reported by Breslow and Gellman for iron-porphyrins reacting with iminoiodanes (*6*). Formation of the dimerization products **4a** and **4b** require the nitrenoid intermediate (or an equivalent species at the same oxidation state) to react with a second molecule of sulfonamide **2** (as speculated in figure S7). That is the active site is likely able to simultaneously accommodate at least one molecule of **1** and one of **2**. The H2A10 active site has high affinity for binding azide **1** (K_d = 40 μM), and only weak affinity for sulfonamide **2** (K_d = 900 μM) (figures S9 and S10).

Due to the unexpected dimerizations (**4a** and **4b**), we reasoned that intramolecular C–H amination might be favored in more bulky multisubstituted arylsulfonyl azides. P450s chosen based on their ability to bind **1** could in fact catalyze benzosultam formation from 2,5-diisopropylbenzenesulfonyl azide (**5**) and 2,4,6-triisopropylbenzenesulfonyl azide (**8**) (figures 6 to 10). None of the P450s could form

benzosultam **6** as the major product, though B1SYN and chimera X7-R1 catalyzed over 30 TTNs of C-H amination. BM3, 2C6, B1SYN, Glc4T180A, ManA10, MB2 and 9-10A TS F87A all formed benzosultam **9** as the major product, with B1SYN catalyzing 45 TTN (45% yield). B1SYN (23 mutations from $P450_{BM3}$) binds azides **5** and **8** with micromolar affinity [K_d (**5**) = 1.5 μM, K_d (**8**) = 19 μM, figure S11]. B1SYN was used for preparative-scale amination reactions (figure 2), making benzosultams **6** and **9** with moderate isolated yields (27% for **6** and 42% for **9**).

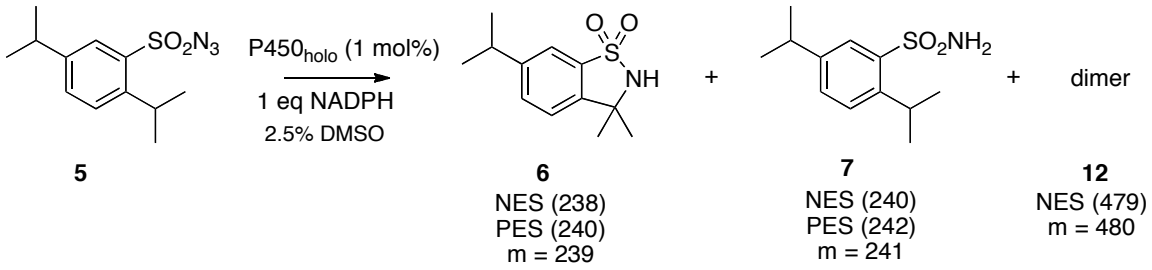


Figure 6. P450 bioconversions with 2,5-diisopropylbenzenesulfonyl azide (**5**) under anaerobic conditions. NES = negative electrospray, PES = positive electrospray. Reaction conditions: 2 mM azide, 20 μM P450, oxygen depletion system (*5*) in 2.5% DMSO at 298 K for 12 hours. Reactions were analyzed by LC-MS.

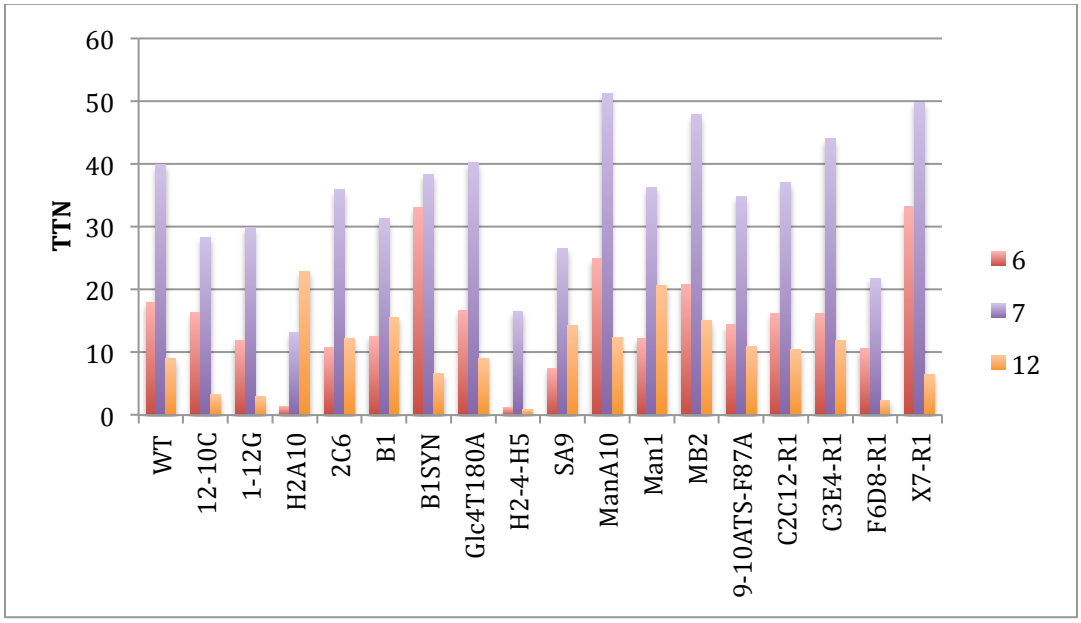


Figure 7. P450 reactions with azide **5** in presence of 2 mM NADPH (1 eq). Benzosultam **6**, arylsulfonamide **7**, and dimer **12** are defined in figure 6.

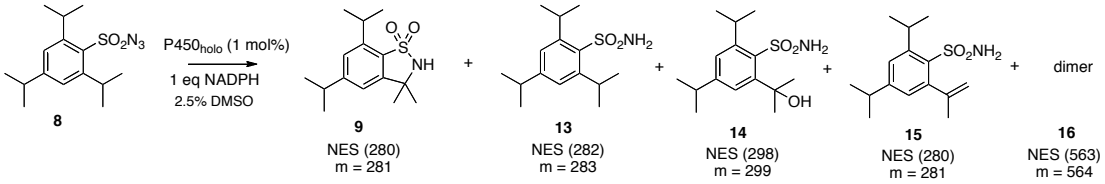


Figure 8. P450 bioconversions with 2,4,6-triisopropylbenzenesulfonyl azide (**8**) under anaerobic conditions. NES = negative electrospray. Reaction conditions: 2 mM azide, 20 µM P450, oxygen depletion system (*5*) in 2.5% DMSO at 298 K for 12 hours. Reactions were analyzed by LC-MS.

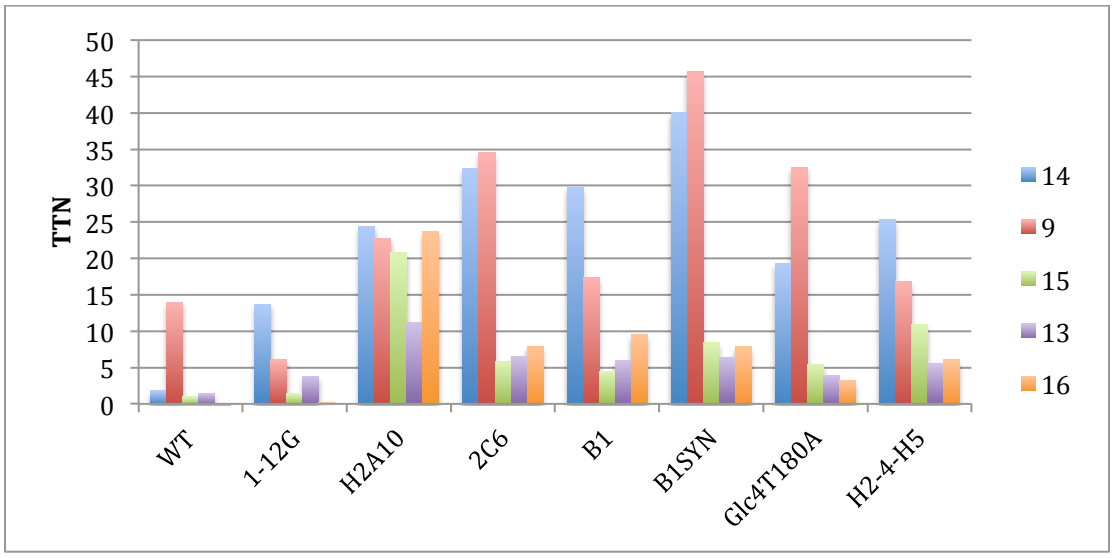


Figure 9. P450 reactions with azide **8** in presence of 2 mM NADPH (1 eq). Benzosultam **9**, arylsulfonamide **13**, alcohol **14**, alkene **15** and dimer **16** are defined in figure 8.

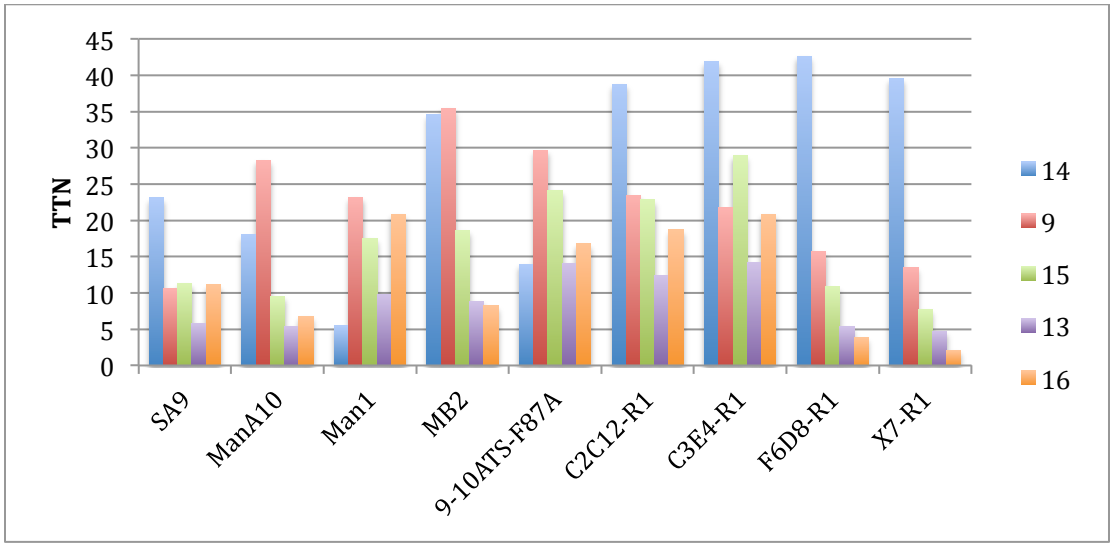


Figure 10. P450 reactions with azide **8** in presence of 2 mM NADPH (1 eq). Benzosultam **9**, arylsulfonamide **13**, alcohol **14**, alkene **15** and dimer **16** are defined in figure S15.

Free hemin was only an effective amination catalyst for the triisopropyl-substituted azide **8** and not the diisopropyl-substituted azide **5** (using dithionite as the reductant, tables S2-S3). Hemin was able to make both sulfonamides **7** and **13**. The fact that hemin can activate both azides but unlike P450s it can only form the benzosultam for the more hindered **8** (which also has two equivalent tertiary C-H bonds), suggests that the enzyme serves primarily to impose a conformation on the substrate that is favorable for nitrene C–H insertion.

These findings attest to the catalytic promiscuity of engineered $P450_{BM3}$ variants to catalyze nonnatural nitrene transfer reactions from synthetic azide reagents. Further engineering of the P450 active site will be required to improve C-H amination activity.

Supplementary Materials for

Intramolecular C-H Amination Catalyzed by Engineered Cytochrome P450 Enzymes

I. Materials and Methods

Unless otherwise noted, all chemicals and reagents for chemical reactions were obtained from commercial suppliers (Sigma-Aldrich, Acros) and used without further purification. Silica gel chromatography purifications were carried out using AMD Silica Gel 60, 230-400 mesh. ^{1}H and ^{13}C NMR spectra were recorded on either a Varian Mercury 300 spectrometer (300 and 75 MHz, respectively), or a Varian Inova 500 MHz (500 and 125 MHz, respectively), and are internally referenced to residual solvent peak. Data for ^{1}H NMR are reported in the conventional form: chemical shift (δ ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, br = broad), coupling constant (Hz), integration. Data for ^{13}C are reported in terms of chemical shift (δ ppm) and multiplicity. High-resolution mass spectra were obtained with a JEOL JMS-600H High Resolution Mass Spectrometer at the California Institute of Technology Mass Spectral Facility. Reactions were monitored using thin layer chromatography (Merck 60 silica gel plates) using an UV-lamp for visualization.

High-performance liquid chromatography (HPLC) was carried out using an Agilent 1200 series, an UV detector, and an Agilent XDB-C18 column (4.6 x 150 mm, 5 μm). Azides **5** and **8**, and benzosultam standards **6** and **9** were prepared as reported (*8*). Benzosultam **11** was purchased from Sigma. These standards were used in co-injection experiments to determine the authenticity of P450-catalyzed benzosultams. Authentic P450 catalyzed benzosultam samples were also prepared as described in section VI and were characterized by NMR (^{1}H and ^{13}C) and mass spectrometry.

II. General Procedures

Typical procedure for small-scale amination bioconversions under anaerobic conditions. Small-scale reactions (400 μL) were conducted in 2 mL crimp vials (Agilent Technologies, San Diego, CA) as for cyclopropanation reactions but an oxygen depletion system, composed of glucose oxidase and catalase, was also added to ensure strict anaerobic conditions. P450 solution (80 μL, 100 μM), glucose oxidase (20 μL, 2,000 U mL^{-1}) and catalase (20 μL, 28,000 U mL^{-1}) were added to the vial with a small stir bar before crimp sealing with a silicone septum. Phosphate buffer (190 μL, 0.1 M, pH = 8.0), glucose (40 μL, 250 mM) and NADPH (40 μL, 20 mM) were combined in a larger crimp sealed vial and degassed by bubbling argon through the solution for at least 5 min. In the meantime, the headspace of the 2 mL reaction vial with the P450 solution was made anaerobic by flushing argon over the protein solution (with no bubbling). The buffer/reductant/glucose solution (270 μL) was syringed into the reaction vial, while under argon. The gas lines were disconnected from the reaction vial before placing the vials on a plate stirrer. A 40x arylsulfonyl azide solution in DMSO (10 μL, typically 80 mM) was added to the reaction vial via a glass syringe, and the reaction was left stirring for the appropriate time. The final concentrations of the reagents were typically 2 mM arylsulfonyl azide, 2 mM NADPH, 25 mM glucose, 20 μM P450. Reactions were quenched by adding 30 μL 3 M HCl under argon. The vials were opened and 50 μL of internal standard (10 mM *ortho*-toluenesulfonamide) was added. The mixture was extracted twice with 500 μL ethyl acetate. The combined organic phase was dried under a light argon stream. The residue was dissolved in 100 μL of 50% water-acetonitrile for HPLC analysis.

Typical procedure for preparative-scale amination bioconversions under anaerobic conditions. Enzymes (P450, glucose oxidase and catalase) were added to a Schlenk flask with a stir bar. With the flask kept on ice, the headspace was evacuated and back-filled with argon (4×) with care not to foam the protein solution. Phosphate buffer, NADPH and glucose were premixed and degassed together in a separate round-bottom flask by bubbling argon through the solution for 20 min. The buffer/reductant solution was transferred to the Schlenk flask via syringe. The arylsulfonyl azide (80 mM in DMSO) was added under argon, and the solution was left to stir under argon until reaction completion. The reaction was quenched under argon by adding hydrochloric acid (3 M) to adjust the pH to 4, before opening the Schlenk flask. The reaction mixture was stirred with sodium chloride and dichloromethane (CH_2Cl_2). The combined emulsion layers were then filtered through Celite to break the emulsion and the Celite pad was rinsed with 3x20 mL CH_2Cl_2. The resulting biphasic mixture was transferred to a separating funnel and the organic phase was removed. The remaining aqueous phase was reextracted with 3x40 mL CH_2Cl_2. The combined organic extracts were dried with sodium sulfate, filtered, and concentrated. The resulting residue was purified by reverse-phase HPLC .

Small-scale amination reactions under anaerobic conditions for azide 1. Reaction conditions were as described in section II and were analyzed by reverse-phase LC-MS (Agilent 1100 series LC-MSD), acetonitrile-water, using a C18 column (Peeke Scientific, Kromasil 100 5 μm, 50 x 4.6 mm ID). Acetonitrile gradient for 2-isopropylbenzenesulfonyl azide reactions: 10%-22% (8 min), 22%-60% (10 min), 60% (2

min), at 1.5 mL min^{-1}. Retention times: alcohol **10** (6.3 min), sulfonamide **2** (10.6 min), dimer **4** (12.3 min), azide **1** (17.6 min). Acetonitrile gradient for 2,5-diisopropylbenzenesulfonyl azide and 2,4,6-triisopropylbenzenesulfonyl azide reactions: 30%-50% (10 min), 50%-90% (8 min), 90% (2 min), at 1.5 mL min^{-1}. Retention times: benzosultam **6** (5.4 min), sulfonamide **7** (7.2 min), dimer **12** (8.6 min), azide **5** (14.4 min); alcohol **14** (8.3 min), benzosultam **9** (9.8 min), olefin **15** (10.8 min), sulfonamide **13** (11.8 min), dimer **16** (15.4 min), azide **8** (16.5 min).

III. Summary of Mutations (with Respect to Wild-Type $P450_{BM3}$) for P450 Amination Catalysts

12-10C: R47C, V78A, A82G, F87V, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, A328V, L353V

9-10A TS: V78A, P142S, T175I, A184V, S226R, H236Q, E252G, A290V, L353V, I366V, E442K

H2A10: **9-10A TS** F87V, L75A, L181A, T268A

H2A10-A75L: 9-10A TS F87V, L181A, T268A

H2-4-H5: H2-A10 M177A A268T

H2-5-F10: **9-10A TS** F87V, L75A, I263A, T268A, L437A

H2-4-D4: **9-10A TS** F87V, L75A, M177A, L181A, T268A, L437A

B1SYN: R47C, V78A, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, L353V, C47S, N70Y, A78L, F87A, I174N, I94K, V184T, I263M, G315S, A330V

IV. Natural Approaches for Amine Biosynthesis

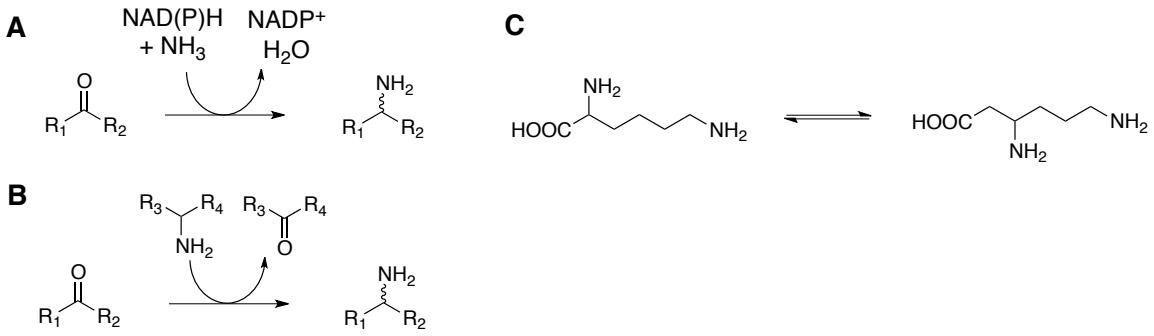


Fig S1. Natural approaches for amine biosynthesis. (**A**) Dehydrogenases convert carbonyl groups to amines using NAD(P)H and ammonia (e.g., glutamate dehydrogenase). (**B**) Transaminases convert carbonyls to amines using another amine (or amino acid) as the amine donor (e.g., acetylornithine transaminase, involved in putrescine biosynthesis). (**C**) Amino mutases catalyze the radical rearrangement of α to β amino acids.

V. P450 Amination Catalysts

Arylsulfonyl azide binding screen. Cell lysate of the compilation plate described in chapter 2 was scanned from 500 to 350 nm in a plate reader (Tecan M1000 UV/Vis) in the absence and presence of 100 μM 2-isopropylbenzenesulfonyl azide. Selected absorbance difference spectra that displayed type I binding to the azide (figure S2) are shown below (figures S3-6).

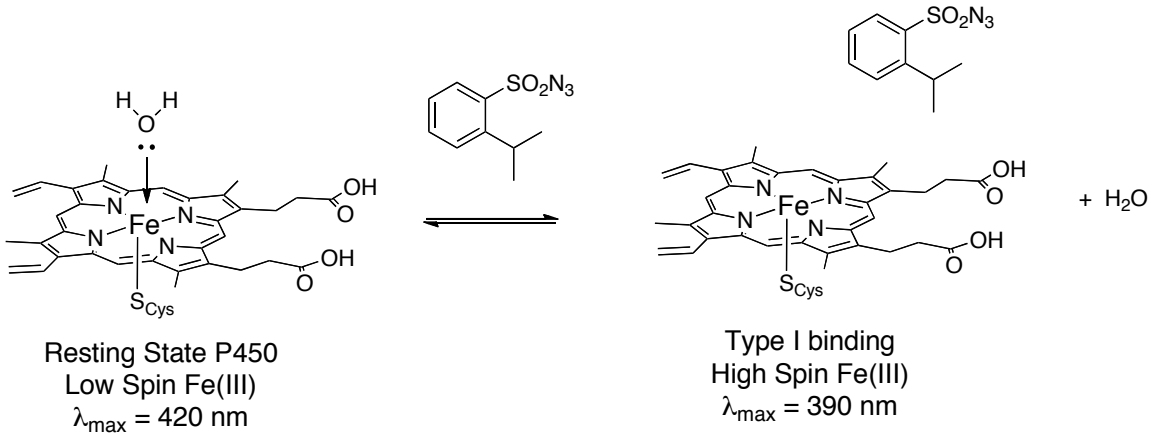


Figure S2. P450BM3 variants display type I binding to arylsulfonyl azides.

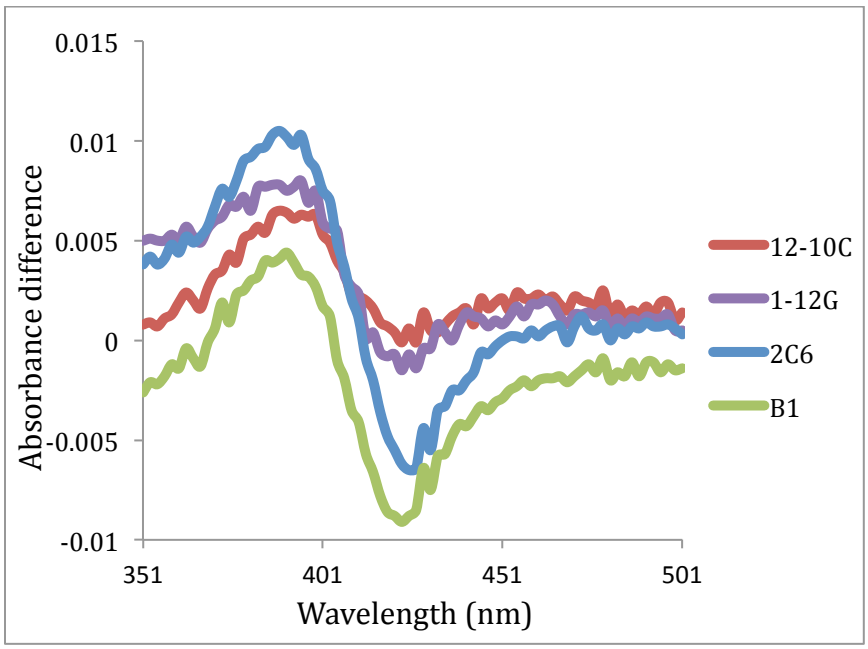


Figure S3. Absorbance difference spectra for $P450_{BM3}$ variants binding 2-isopropylbenzenesulfonyl azide. Sequence identities are shown on table S5 in chapter 2.

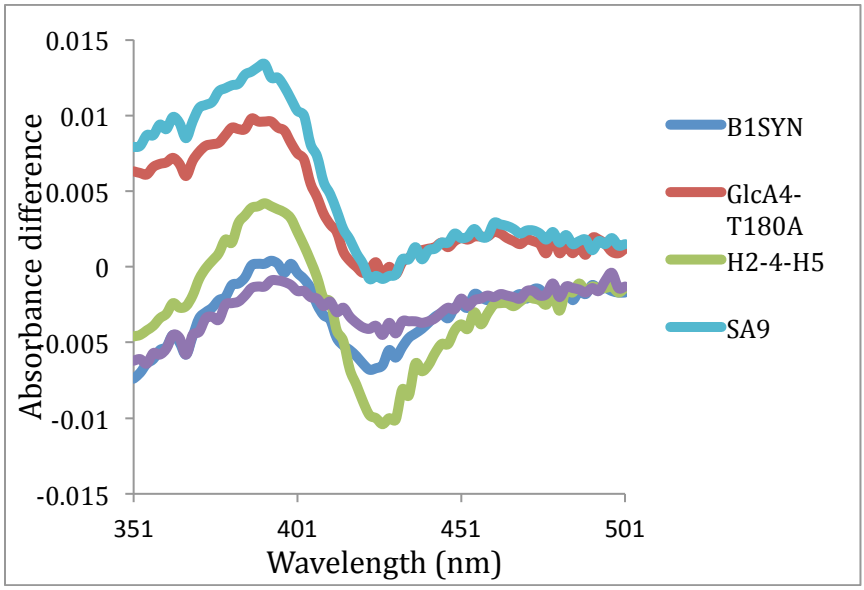


Figure S4. Absorbance difference spectra for $P450_{BM3}$ variants binding 2-isopropylbenzenesulfonyl azide. Sequence identities are shown on table S5 in chapter 2.

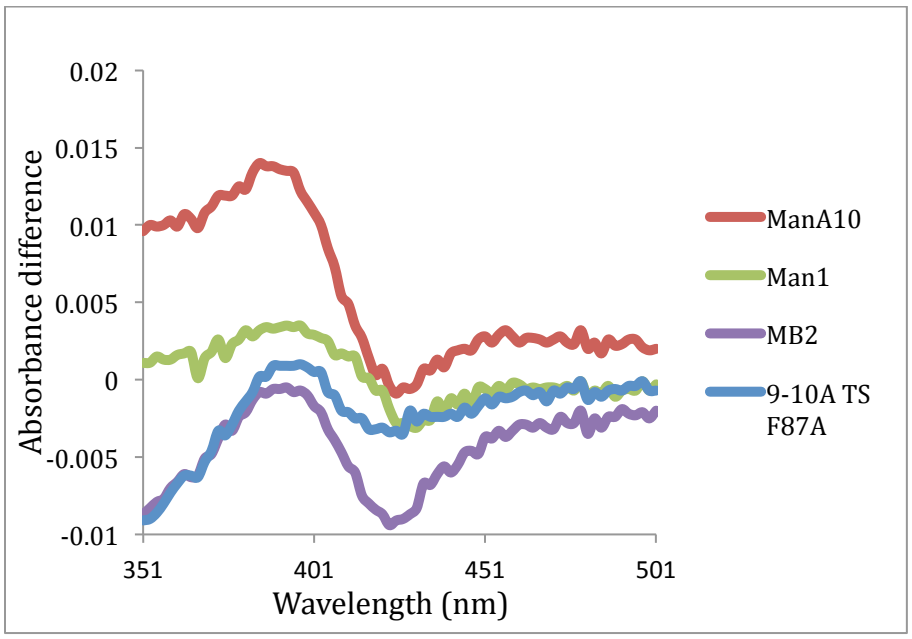


Figure S5. Absorbance difference spectra for $P450_{BM3}$ variants binding 2-isopropylbenzenesulfonyl azide. Sequence identities are shown on table S5 in chapter 2.

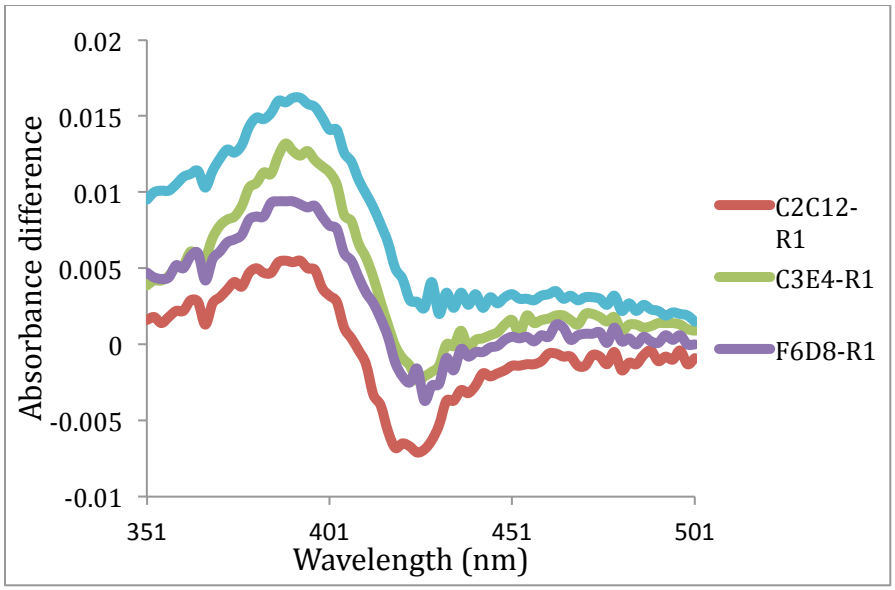


Figure S6. Absorbance difference spectra for $P450_{BM3}$ variants binding 2-isopropylbenzenesulfonyl azide. Sequence identities are shown on table S6 in chapter 2.

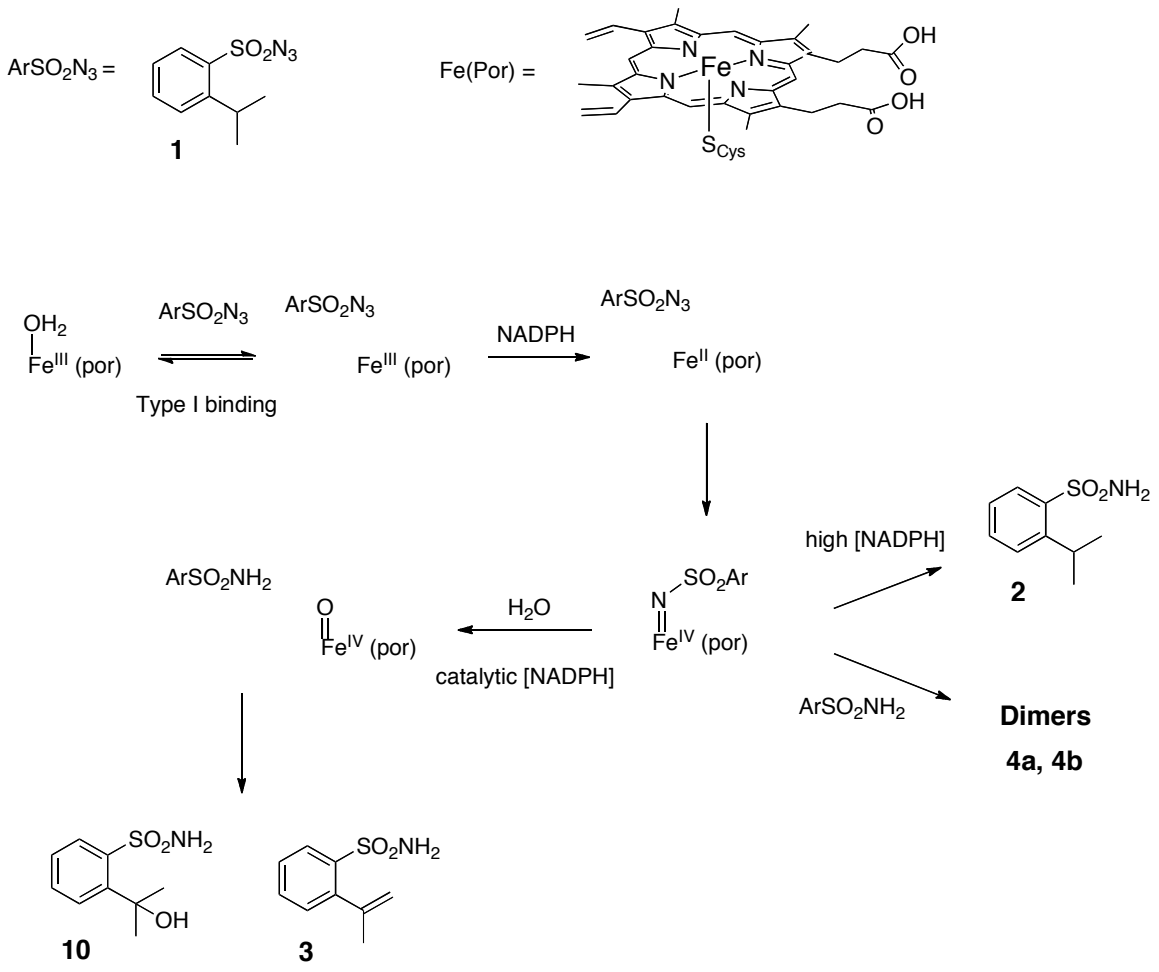


Figure S7. Putative mechanism for azide activation and decomposition catalyzed by cytochrome P450. Nitrenoid hydrolysis and subsequent to a ferryl species which catalyzes subsequent oxidation is based on the mechanism proposed by White and McCarthy (*9*).

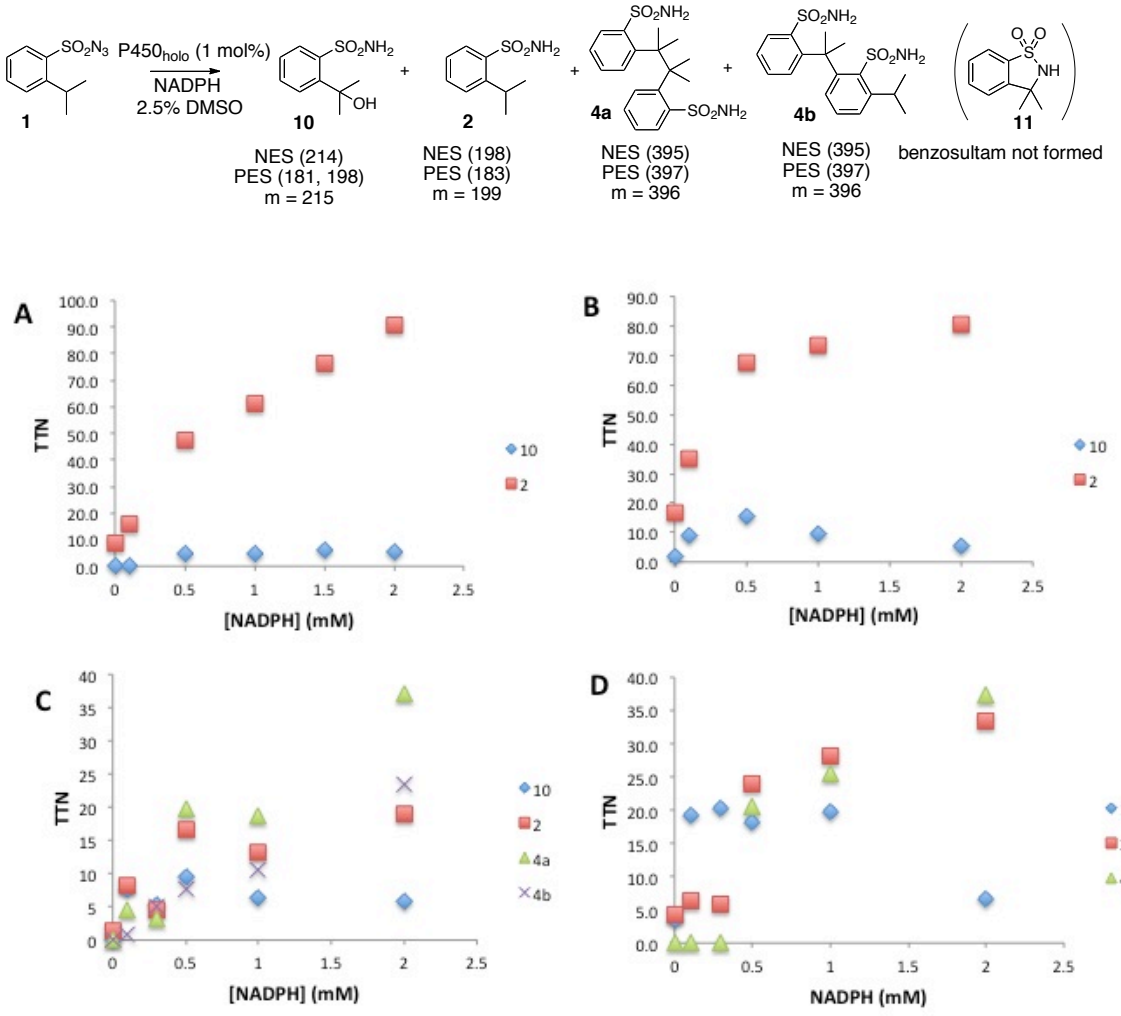
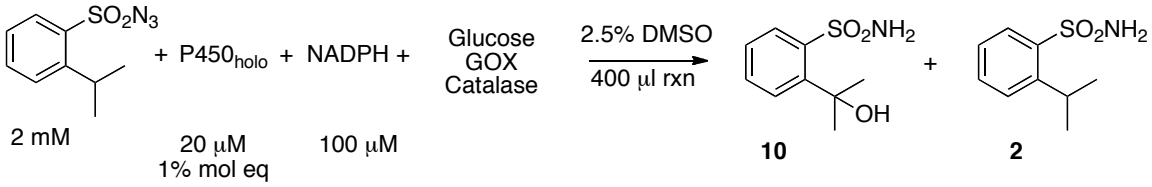


Figure S8. Effect of varying the concentration of NADPH on the reaction of azide **1** with: (**A**) BM3; (**B**) BM3-F87A; (**C**) H2A10; (**D**) H2A10-A75L. For plots A and B, [NADPH] = 0.1, 0.5, 1.0, 1.5, 2.0 mM. For plots C and D, [NADPH] = 0.1, 0.3, 0.5, 1.0, 2.0 mM NADPH.

Table S1. Controls for the reaction of H2A10-A75L with azide **1** with 0.05 eq. NADPH. Complete System (CS) = 2 mM azide, 20 μM P450, 100 μM NADPH in anaerobic conditions with an oxygen depletion system.



Conditions	**10** (TTN)	**2** (TTN)
CS	29	16
CS-NADPH	5	4
CS+O2	7	3.8
CS+O2-ODS	0	1.6
boiled P450	3.8	13.9

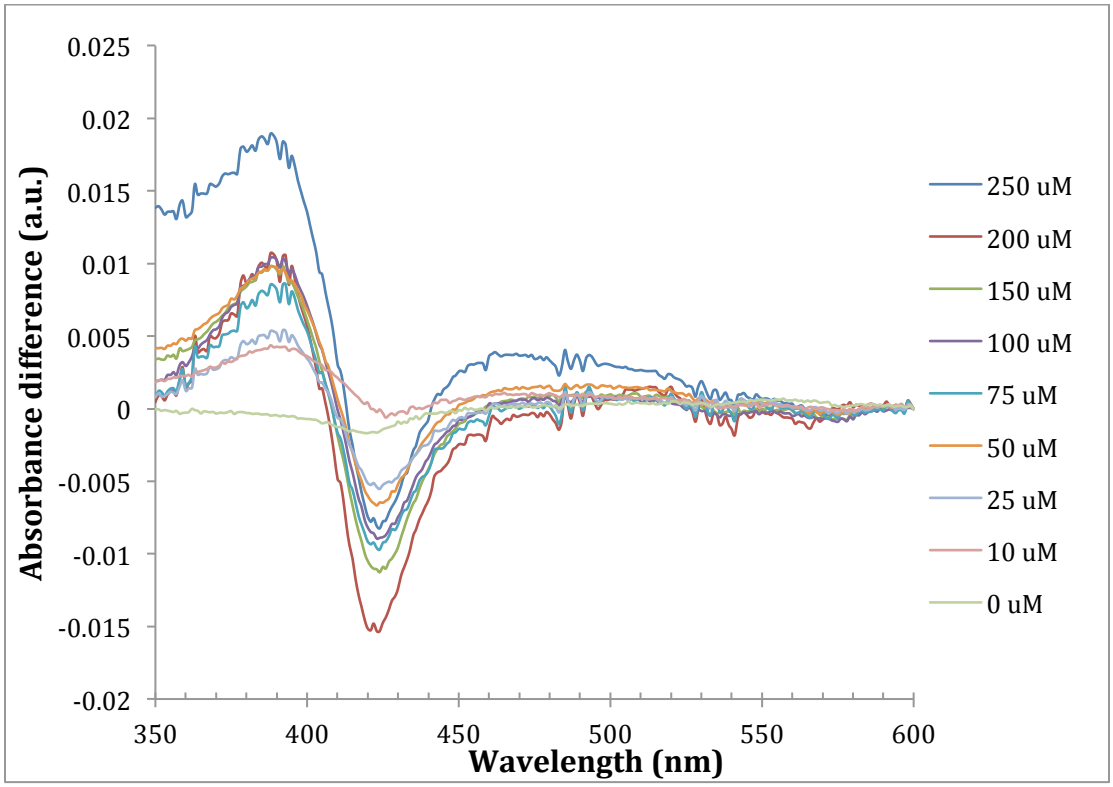


Figure S9. Absorbance difference spectrum for azide **1** (concentration shown in the legend) binding to H2A10. All spectra are referenced against a 1 μM H2A10 solution with 1% DMSO.

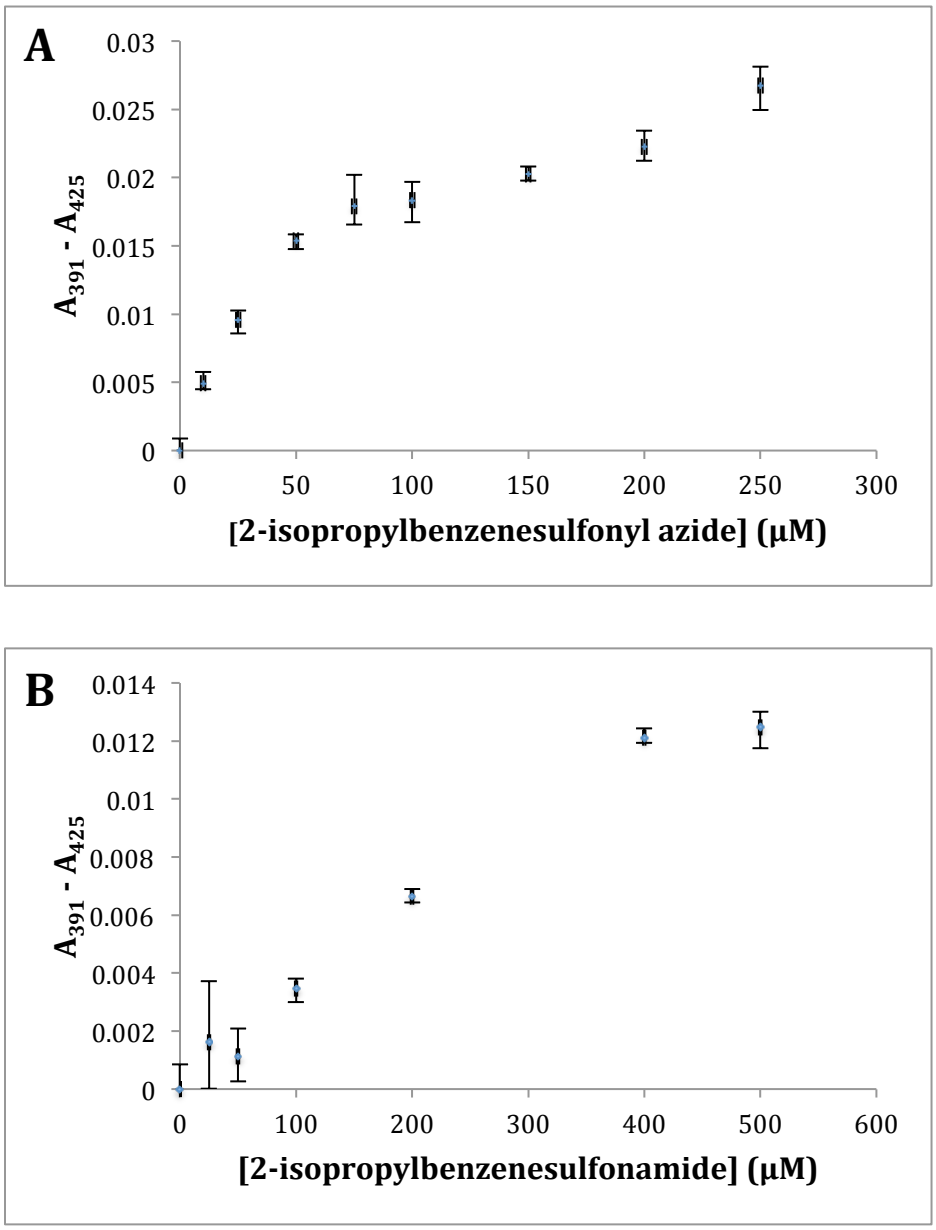


Figure S10. Biding curves for azide **1** (**A**) and sulfonamide **2** (**B**) to H2A10. The *y*-axes show the difference in absorbance at 391 and 425 nm with respect to 1 μM H2A10 solution with 1% DMSO and no ligand. K_d (**1**) = 40 μM. K_d (**2**) = 900 μM.

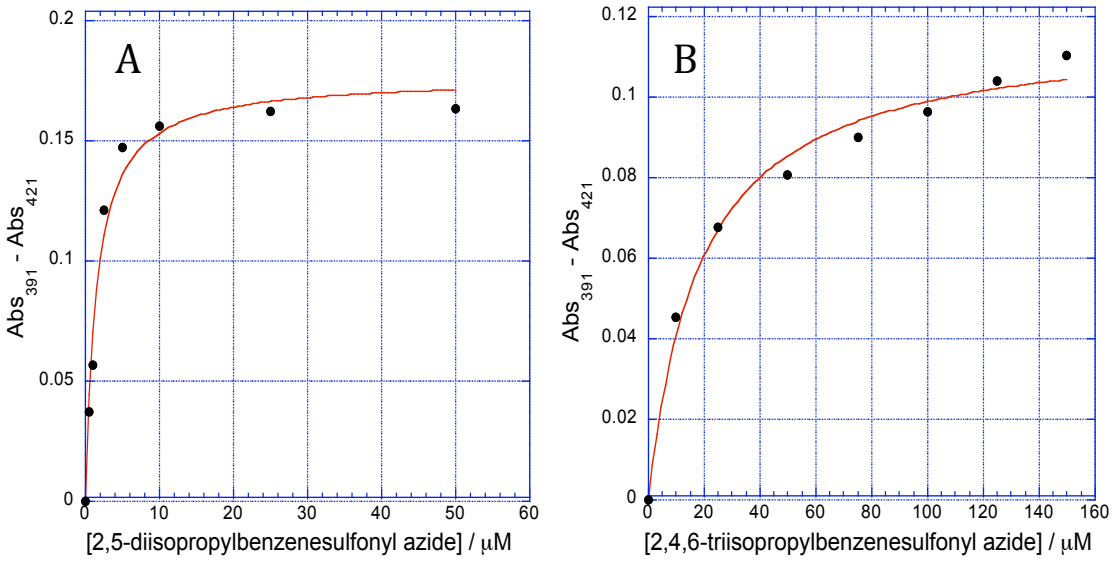
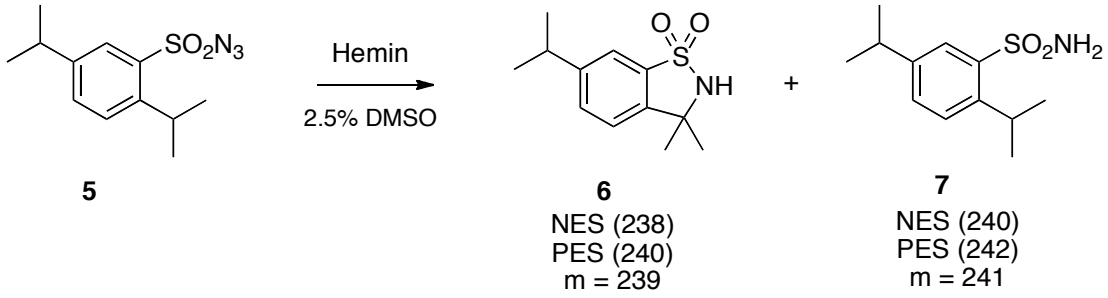


Figure S11. B1SYN type I binding curves for azides **5** (**A**) and **8** (**B**). K_d (**5**) = 1.5 μM, K_d (**8**) = 19 μM.



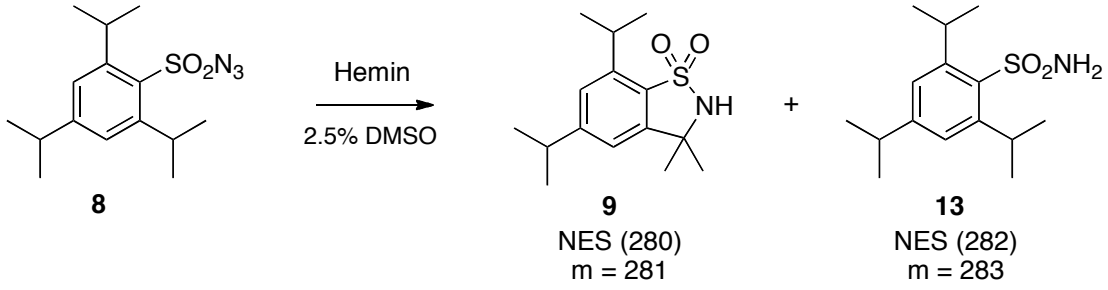
5

6
NES (238)
PES (240)
m = 239

7
NES (240)
PES (242)
m = 241

Table S2. Hemin reactions with azide **2a** under anaerobic conditions

Cat. loading (mol%)	**Reductant**	**TTN benzosultam 6**	**TTN arylsulfonamide 7 (% yield)**
1	None	0	0
1	2 mM $Na_2S_2O_4$	0	5 (5)
1	10 mM $Na_2S_2O_4$	0	58 (58)
1	2 mM NADPH	0	1 (1)
10	2 mM $Na_2S_2O_4$	0	1 (10)
10	10 mM $Na_2S_2O_4$	0	6 (60)



8

9
NES (280)
m = 281

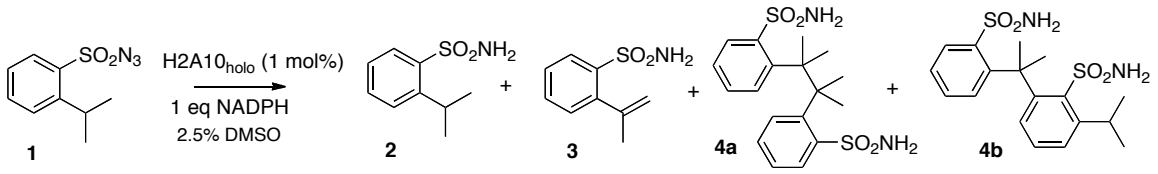
13
NES (282)
m = 283

Table S3. Hemin reactions with azide **2b** under anaerobic conditions

Cat. loading (mol%)	**Reductant**	**TTN benzosultam 9 (% yield)**	**TTN arylsulfonamide 13 (% yield)**
1	None	0	0
1	2 mM $Na_2S_2O_4$	4 (4)	1 (1)
1	10 mM $Na_2S_2O_4$	61 (61)	23 (23)
1	2 mM NADPH	0	1 (1)
10	2 mM $Na_2S_2O_4$	0	0
10	10 mM $Na_2S_2O_4$	4 (40)	3 (30)

VI. Preparative-Scale Bioconversions

These reactions were conducted anaerobically as described in section II.



H2A10 scale-up with 2-isopropylbenzenesulfonyl azide (1). Preparation used 48 mg of azide **1** and 2 μmol $H2A10_{holo}$ (0.01 equiv). The products were purified by reverse phase HPLC to give 6 mg of arylsulfonamide **2** (15%), 2 mg of olefin **3** (5%), 11 mg of dimer **4a** (25%) and 4 mg of dimer **4b** (5%).

2-isopropylbenzenesulfonamide (**2**). ^{1}H NMR (500 MHz, DMSO): d 7.82 (d, *J* = 8.14, 1H), 7.54 (m, 2H), 7.45 (br s, 2H), 7.31 (ddd, *J* = 2.16, 6.03, 8.16, 1H), 3.84 (sep, *J* = 6.71, 1H), 1.20 (d, *J* = 6.71, 6H). ^{13}C NMR (125 MHz, DMSO): d 147.04, 141.35, 132.11, 127.63, 126.76, 125.64, 28.54, 23.87. Expected *m/z* for $C9H13NO2NaS^{+}$ 222.0559. Observed *m/z* 222.0552.

2-(prop-1-en-2-yl)benzenesulfonamide (**3**). ^{1}H NMR (500 MHz, DMSO): d 7.89 (dd, *J* = 8.08, 1.23, 1H), 7.53 (m, 1H), 7.44 (m, 1H), 7.26 (dd, *J* = 7.60, 1.34, 1H), 7.22 (s, 2H), 5.20 (ap p, *J* = 1.60, 1H), 4.86 (m, 1H), 2.05 (br s, 3H). ^{13}C NMR (HMBC/HSQC 500 MHz, DMSO): d 141.83, 141.39, 131.27, 129.76, 126.89, 126.70, 115.79, 24.86. Expected *m/z* for $C9H12NO2S^{+}$ 198.0583. Observed *m/z* 198.19.

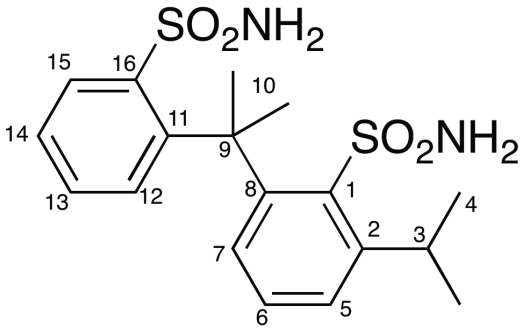
2,2'-(2,3-dimethylbutane-2,3-diyl)dibenzenesulfonamide (**4a**). ^{1}H NMR (500 MHz, DMSO): d 8.16 (dd, *J* = 8.22, 1.45, 1H), 7.53 (s, 2H), 7.33 (ddd, *J* = 8.18, 6.74, 1.26, 1H), 7.25 (m, 1H), 7.18 (d, *J* = 8.06, 1H), 1.59 (s, 6H). ^{13}C NMR (HMBC/HSQC 500 MHz,

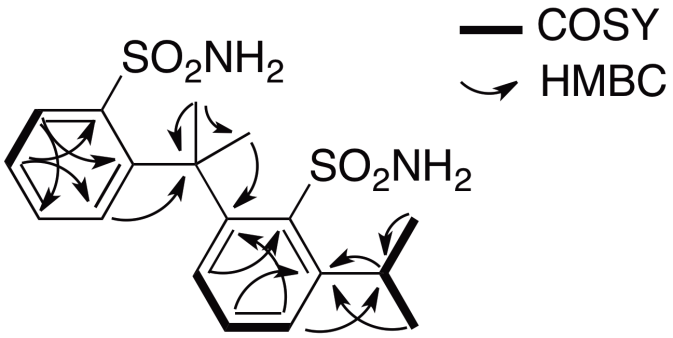
DMSO): d 145.11, 144.81, 134.17, 128.95, 128.26, 125.77, 48.79, 29.78. Expected *m/z* for $C18H25N2O4S2^{+}$ =397.1250. Observed *m/z* 397.0147.

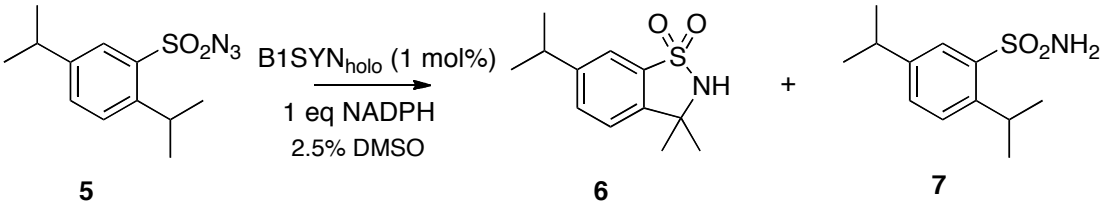
Summary of NMR data for dimer 4b.

position	δ_C, type	δ_H (*J* in Hz)	HMBC	COSY
1	140.45, C	-	-	-
2	143.51, C	-	-	-
3	27.97, CH	3.77, sep (6.77)	1, 2, 4, 5	4
4	23.72, CH3	1.17, d (6.72)	2, 3	3
5	126.72, CH	7.38, d (8.22)	1, 3, 8	6
6	129.67, CH	7.18, d (8.23)	2, 7, 9	5, 7
7	124.17, CH	7.63, m	1, 2, 6, 9	7
8	148.33, C	-	-	-
9	44.56, C	-	-	-
10	31.82, CH3	1.80, s	8, 11, 12	-
11	145.77, C	-	-	-
12	129.96, CH	7.57, m	9, 14, 16	-
13	131.47, CH	7.56, m	-	-
14	126.59, CH	7.46, m	12, 16	15
15	129.74, CH	8.08, d (7.97)	11, 13	14
16	143.15, C	-	-	-
NH		6.83, br s	-	-
NH		7.30, br s	-	-

Expected *m/z* for C18H25N2O4S2+ 397.1250. Observed *m/z* 397.1245.

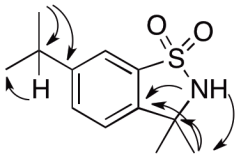






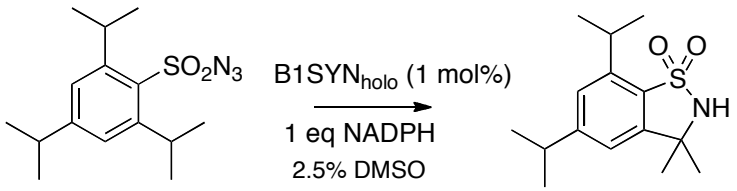
B1SYN scale-up with 2,5-diisopropylbenzenesulfonyl azide (5). Preparation used 24 mg of azide **5** and 0.9 μmol $B1SYN_{holo}$ (0.01 equiv). The products were purified by reverse phase HPLC to give 6 mg of benzosultam **6** (27%) and 7 mg of arylsulfonamide **7** (32%).

Enzymatically produced diisopropyl benzosultam (**6**). ^{1}H NMR (600 MHz, $CDCl_3$): d 7.59 (s, 1H), 7.48 (d, J = 8.03, 1H), 7.29 (d, J = 8.11, 1H), 4.50 (s, 1H), 3.01 (sep, J = 6.95, 1H), 1.64 (s, 6H), 1.28 (d, J = 6 .98, 6H). ^{13}C NMR (150 MHz, $CDCl_3$): d 150.83, 143.69, 135.34, 132.39, 122.74, 118.70, 60.78, 34.17, 29.92, 23.92. Expected *m/z* for $C12H18NO2S^{+}$ 240.1053. Observed *m/z* 240.1059. NMR spectra of enzymatically produced diisopropylbenzosultam were identical with those of a synthetic standard produced according to Ruppel et al (*8*). Moreover, the identity of the benzosultam could be further supported by the observation of HMBC correlations from the amide proton to the geminal dimethyl groups.



2,5-diisopropylbenzenesulfonamide (**7**). ^{1}H NMR (500 MHz, $CDCl_3$): d 7.86 (d, J = 1.7 Hz, 1H), 7.41 (m, 2H), 4.81 (2H, br s), 3.76 (sep, J = 6.80 Hz, 1H), 2.94 (sep, J = 6.97 Hz, 1H), 1.30 (J = 6.80 Hz, 6H), 1.26 (d, J = 6.96 Hz, 6H). ^{13}C NMR (125 MHz,

$CDCl_3$): d 146.93, 145.34, 138.67, 131.43, 128.03, 126.16, 33.85, 29.53, 24.21, 23.92. Expected *m/z* for C12H20NO2S+ 242.1209. Observed *m/z* 242.1210.

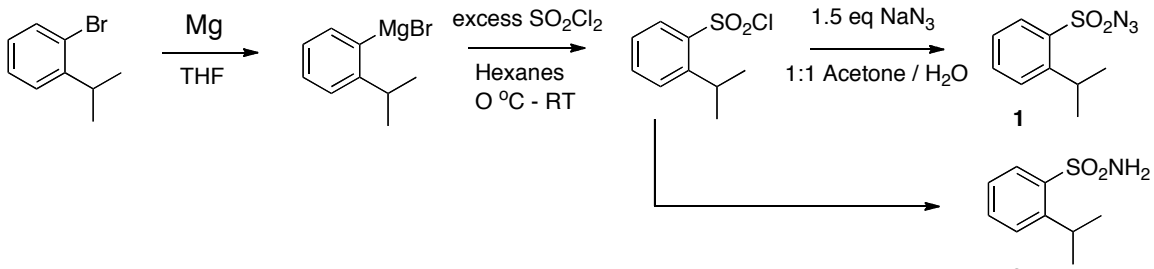


8 **9**

B1SYN scale-up with 2,5-diisopropylbenzenesulfonyl azide (8). Preparation used 13 mg of azide **8** and 0.4 μmol $B1SYN_{holo}$ (0.01 equiv). The product was purified by reverse phase HPLC to give 5 mg of benzosultam **9** (42%). Enzymatically produced triisopropyl benzosultam (**9**). ^{1}H NMR (500 MHz, $CDCl_3$): d 7.22 (d, *J* = 1.37, 1H), 6.98 (d, *J* = 1.38, 1H), 4.47 (s, 1H), 3.61 (sep, *J* = 6.80, 1H), 2.98 (sep, *J* = 7.01, 1H), 1.63 (s, 6H), 1.35 (d, 6.81, 6H), 1.27 (d, 6.92, 6H). ^{13}C NMR (125 MHz, $CDCl_3$): d 155.7, 146.8, 145.5, 131.0, 124.5, 117.9, 59.9, 34.8, 30.0, 29.6, 24.05, 23.72. Expected *m/z* for C15H24NO2S+ 282.1522. Observed *m/z* 282.1528.

VII. Synthesis of Arylsulfonyl Azides

Synthesis of 2-isopropylbenzenesulfonyl azide (1)



2-isopropylbenzenesulfonyl chloride. Freshly polished magnesium turnings (0.488 g, 20.1 mmol) were suspended in dry THF (16 mL) and stirred vigorously. An aliquot of a solution of 2-bromocumene (2.00 g, 10 mmol) in 8 mL dry THF was added and the reaction was initiated by heating to a brief boil. The remainder of the starting material was slowly added to maintain reaction. After three hours, the reaction was cooled with an ice bath, and the solution was transferred under nitrogen via Teflon tubing to a solution of SO_2Cl_2 (4.0 mL, 50 mmol) in dry hexanes (25 mL) also at 0 °C and left overnight. The reaction was slowly poured over ice cold water (50 mL) and extracted with DCM 4 times. The organic layer was dried (Na_2SO_4), filtered and concentrated *in vacuo*. The crude product was purified by chromatography (SiO_2, 10% ether/hexanes) to afford the sulfonyl chloride (1.302 g, 60%). ^{1}H NMR ($CDCl_3$, 300 MHz): δ = 8.00 (m, 1H), 7.73 – 7.55 (m, 2H), 7.40 – 7.31 (m, 1H), 4.22 – 4.08 (m, 1H), 1.36 (d, J = 6.8 Hz, 6H).

2-isopropylbenzenesulfonyl azide (1). The chloride (0.800 g, 3.7 mmol) was dissolved in acetone (9.5 ml) and cooled with an ice bath. A cold solution of sodium azide (0.358 g, 5.5 mmol) in water (9.5 ml) was added dropwise and left to react overnight. The reaction

mixture was extracted with DCM, dried, filtered and solvent was evaporated *in vacuo*. Flash chromatography (SiO_2, 10% ether/hexanes) gave the sulfonyl azide **1** (0.666 g, 80%). ^{1}H NMR ($CDCl_3$, 300 MHz): δ = 8.04 (dd, *J* = 8.1, 1.3 Hz, 1H), 7.70 – 7.56 (m, 2H), 7.38 (m, 1H), 3.82 – 3.67 (m, 1H), 1.36 – 1.27 (d, *J* = 6.8 Hz, 6H). HRMS (EI+): Calcd. for $C_9H_{11}SO_2N_3$ (M^+) m/z 225.0572; found 225.0581.

2-isopropylbenzenesulfonyl amide (2). The chloride (0.241 g, 1.11 mmol) was dissolved in chloroform (9 ml) and cooled with an ice bath. Ammonium hydroxide (30%, 0.35 mL, 5.6 mmol) was added dropwise and left to react overnight. The reaction mixture was extracted with DCM, dried, filtered and solvent was evaporated *in vacuo*. Flash chromatography (SiO_2, 10% ether/hexanes) gave the sulfonyl amide **7** (0.68 g, 62%). ^{1}H NMR ($CDCl_3$, 300 MHz): δ = 8.02 (d, *J* = 9.3 Hz, 1H), 7.60 – 7.47 (m, 2H), 7.35 – 7.27 (m, 1H), 4.78 (s, 2H), 3.87 – 3.74 (m, 1H), 1.32 (d, *J* = 6.8 Hz, 6H). HRMS (EI+): Calcd. for $C_9H_{13}SO_2N$ (M^+) m/z 199.0667; found 199.0627.

2,5-diisopropylbenzenesulfonyl amide (7). Same procedure as used above for **2**. Sulfonamide **7** was obtained in 75% yield (0.345 g). ^{1}H NMR ($CDCl_3$, 300 MHz): δ = 7.92 – 7.80 (m, 1H), 7.46 – 7.36 (m, 2H), 4.75 (s, 2H), 3.85 – 3.68 (m, 1H), 3.02 – 2.85 (m, 1H), 1.31 (d, *J* = 6.8 Hz, 6H), 1.26 (d, *J* = 6.9 Hz, 6H).

VIII. References

1. H. M. L. Davies, J. R. Manning, Catalytic C-H functionalization by metal carbenoid and nitrenoid insertion. *Nature* **451**, 417 (2008).
2. J. Du Bois, Rhodium-catalyzed C-H amination: versatile methodology for the selective preparation of amines and amine derivatives. *Chemtracts* **18**, 1 (2005).
3. D. N. Zalatan, J. Du Bois, Metal-catalyzed oxidations of C-H to C-N bonds. *Topics in Current Chemistry* **292**, 347 (2010).
4. P. Dauban, R. H. Dodd, Catalytic C-H amination with nitrenes. *Amino Group Chemistry*, 55 (2008).
5. See supplementary materials.
6. R. Breslow, S. H. Gellman, Intramolecular nitrene carbon-hydrogen insertions mediated by transition-metal complexes as nitrogen analogs of cytochrome P-450 reactions. *J. Am. Chem. Soc.* **105**, 6728 (1983).
7. E. W. Svastits, J. H. Dawson, R. Breslow, S. H. Gellman, Functionalized nitrogen atom transfer catalyzed by cytochrome P-450. *J. Am. Chem. Soc.* **107**, 6427 (1985).
8. J. V. Ruppel, R. M. Kamble, X. P. Zhang, Cobalt-catalyzed intramolecular C-H amination with arylsulfonyl azides. *Organic Letters* **9**, 4889 (2007).
9. R. E. White, M. B. McCarthy, Aliphatic hydroxylation by cytochrome P-450. Evidence for rapid hydrolysis of an intermediate iron-nitrene complex. *J. Am. Chem. Soc.* **106**, 4922 (1984).

Chapter 5

ON PREADAPTATION AND THE EVOLUTION OF ENZYMES WITH SYNTHETIC CATALYTIC FUNCTION

"A structure is said to be preadapted for a new function if its present form which enables it to discharge its original function also enables it to assume the new function whenever need for this function arises." — W. J. Bock (1959)

I was particularly influenced by the work of Seffernick et al. on the evolution of atrazine chlorohydrolase and by S. J. Gould's beautifully simple illustration of preadaptation in his book "Ever Since Darwin".

Abstract. The introduction of synthetic compounds into the environment since the industrial revolution has led, in some cases, to the rapid evolution of microbial catabolic enzymes with novel catalytic functions. A hallmark example is the evolution of atrazine chlorohydrolase, a unique metal-dependent hydrolase that cleaves C-Cl bonds in triazine containing herbicides. Synthetic reagents can also be used in the laboratory for screening existing enzymes to discover synthetic catalytic functions. Our own work shows that only a few mutations can unlock substantial C=C cyclopropanation and C-H amination in naturally occurring cytochrome P450 enzymes. We invoke preadaptation to explain how some structures are poised for novel catalytic functions. We propose that a mechanism-based approach, which requires only minimal knowledge of the mechanism of the target reaction and of structural features of the natural enzyme, can be used to engineer other abiological synthetically useful transformations into naturally occurring enzymes.

Using Evolutionary Design to Install Synthetic Activities in Existing Enzymes

It seems at first counterintuitive to use evolutionary approaches in attempting to create biocatalysts with wholly nonnatural functions. Since Darwinian evolution works by the incremental accumulation of beneficial mutations, how can this process be successful when the novel function has absolutely no natural precedence? Two requirements must be fulfilled for evolutionary methods to work: (i) the proposed novel reaction has to share mechanistic features with the existing wild-type function; and (ii) there must exist an evolutionary pathway for improving the novel activity (i.e., improvements can be reached by accumulating sequential amino acid mutations).

Over the last twenty years, evidence has mounted that naturally occurring enzymes often exhibit measurable levels of "promiscuous" catalytic activity for nonnatural reactions. Examples of catalytic promiscuity have been found in environmental microbiology and *in vitro* biocatalysis. Rather than being a special property of a few enzymes, catalytic promiscuity is now widely established, and many recent advances in biocatalysis are based on this phenomenon.[1] Searches for promiscuous enzymes are based on mechanistic hypotheses as to why the novel reaction may be catalyzed by a given enzyme.

Whilst many enzymes will exhibit basal levels of promiscuous activity on a given nonnatural target reaction, some enzymes may be more preadapted for undergoing optimization for the novel function. That is, synthetically useful levels of activity and selectivity can be reached by just a few amino acid mutations from the wild-type enzyme. We evaluate the effects that influence preadaptation and illustrate how to choose natural enzymes as candidates for catalyzing synthetic bond disconnections

The Introduction of Synthetic Compounds in the Environment Drives the Evolution of Enzymes with Novel Catabolic Activities

Melamine has been used since the early 20th century as a plastic component (Formica) and as a colorant. Melamine is a nonnatural compound and its persistency in the environment was described as non-biodegradable in the 1930s, slightly more biodegradable in the 1960s, and readily biodegradable in the 2000s.[2] This increasing biodegradability is due to the evolution of melamine deaminase (TriA), an enzyme originally isolated from *Pseudomonas* sp. strain NRRL B-12227.[3] TriA likely evolved from a pyrimidine deaminase,[2] an enzyme that belongs to the aminohydrolase superfamily. Aminohydrolases share a common $(\alpha\beta)_8$ barrel structure and the coordination sites for one or two divalent metals that serve to activate water for nucleophilic attack on the substrate.[4] Pyrimidine deaminases typically catalyze the hydrolysis reaction shown in figure 1A. Deamination of melamine involves hydrolysis of an amine from the *s*-triazine ring (figure 1B), which is more activated than the pyrimidine ring toward nucleophilic aromatic substitution. Dense and diverse microbial populations and relatively low levels of nutrients in the soil make for strong selective pressure, favoring bacteria that can rapidly adapt to using alternative nutrient sources. Since melamine can be used as a source of nitrogen,[3] positive selection can be expected for microbes that can catalyze the initial hydrolysis of melamine en route to cyanuric acid and ammonium. TriA catalyzes melamine deamination with a catalytic rate constant (k_{cat}) of 4.1 s^{-1} and a Michaelis-Menten constant (K_M) of 305 μM ($k_{cat}/K_M = 1.3 \times 10^4\ s^{-1}\ M^{-1}$).[5]

Atrazine was introduced in the late 1950s as a potent herbicide, and it is estimated that over 1 million tons have been applied globally. Until 1993 atrazine did not readily

degrade in the soil; however, rapid degradation has been reported from 1993 onwards.[2] Since 1993, various bacteria were ascertained to initiate atrazine catabolism via a chlorohydrolase (figure 1C). The genes encoding atrazine chlorohydrolase (AtzA, EC 3.8.1.8) from bacteria independently isolated from four continents have been shown to be essentially identical, suggesting recent evolution of AtzA (i.e., after the introduction of atrazine) and fast dissemination of this phenotype across the globe.[6] AtzA from *Pseudomonas* sp ADP shares 98% amino acid sequence identity to TriA, the two enzymes differing by just nine out of 475 amino acids.[7] The small number of amino acid substitutions, the lack of silent DNA mutations between the two genes, and the observation that AtzA and TriA can be found in the same bacterial species are all consistent with rapid evolution of AtzA from TriA due to an intense evolutionary pressure operating over a short time period.[2, 7] Both melamine and atrazine contain the *s*-triazine ring, but nucleophilic aromatic substitution for atrazine is expected to be more facile because chloride better stabilizes the anionic intermediate as well as being a better leaving group. TriA shows modest promiscuous atrazine chlorohydrolase activity (about two orders of magnitude lower than its deamination activity with comparable triazine substrates). Interestingly, AtzA shows no vestigial aminohydrolase activity on melamine (nor on an equivalent amine analogue of atrazine), and catalyzes atrazine dechlorination just as efficiently (k_{cat} = 2.2 s^{-1}, K_M = 153 μM, k_{cat}/K_M = 1.5 x 10^4 s^{-1} M^{-1}), thus showing a complete switch in reaction specificity.[5]

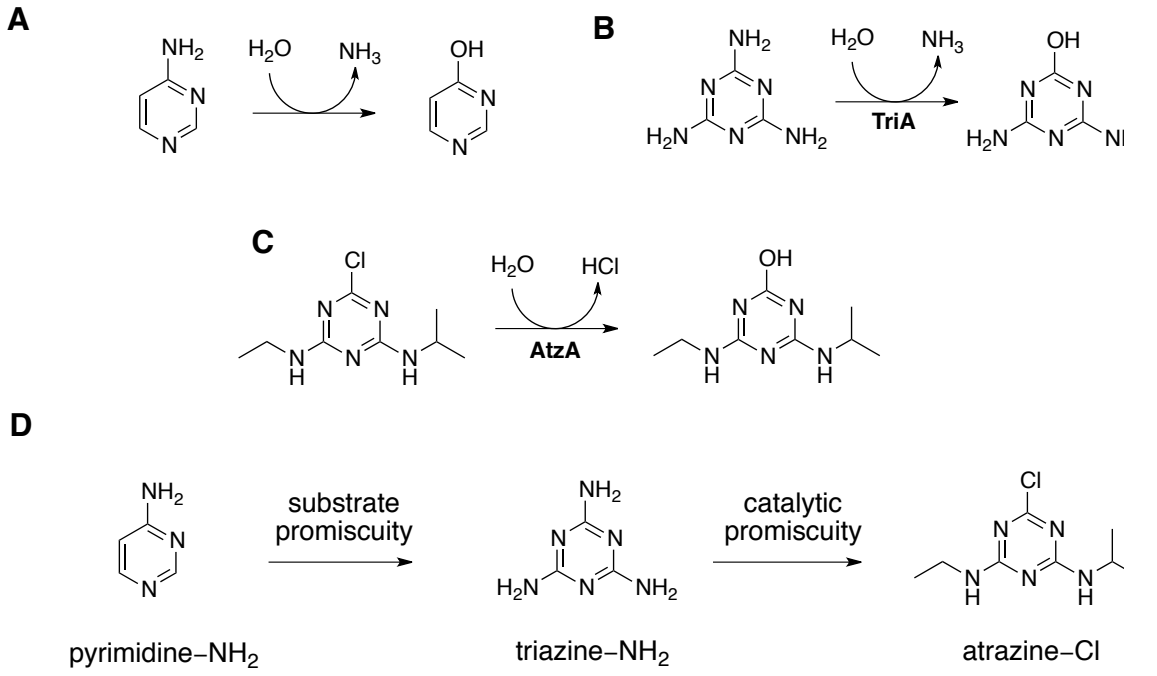


Figure 1: (**A**) The prototypical reaction for a pyrimidine aminohydrolase. (**B**) Melamine aminohydrolase (TriA) cleaves an *s*-triazine-NH_2 bond. (**C**) Atrazine chlorohydrolase (AtzA) cleaves an *s*-triazine-Cl bond. (**D**) Evolution of TriA is based on substrate promiscuity since the novel and native reactions involve the same bond disconnection, whereas evolution of AtzA is a case of catalytic promiscuity as the novel reaction requires a bond disconnection that is different from the native reaction.

The progression from pyrimidine to melamine to atrazine is an example of a substrate walk (figure 1D), whereby activity on a substrate of intermediate structural similarity (melamine) is first developed in order to create a catalyst that will show some promiscuous activity on the final target substrate (atrazine). Substrate walk approaches have been used with great success in directed evolution experiments in the laboratory.[8] The pyrimidine to atrazine walk is particularly interesting because it achieved not only a change of substrate (pyrimidine to triazine; i.e., substrate promiscuity) but also a change of reaction (cleavage of a C−N bond to cleavage of a C−Cl bond; i.e., catalytic promiscuity). Therefore a nonnatural reaction eventually emerged, whilst maintaining

similar mechanistic features to the native aminohydrolase activity. AtzA is a unique metal-dependent halohydrolase.[2] Other hydrolytic dehalogenases are known in nature; but these use a carboxylate nucleophile instead of water activated by a divalent metal, and do not belong to the amidohydrolase superfamily. The catalytic promiscuity of AtzA has been further developed by directed evolution, which has expanded the substrate range of AtzA to include C−S and C−O bond cleavage.[9] It can, therefore, be seen that very small changes in structure are sufficient to enable the emergence of novel catalytic function in homologous proteins with very high sequence similarity.

There are several other cases where the introduction of a synthetic compound into the environment drives the evolution of a novel bacterial catabolic activity. For example, the dry cleaning solvent tetrachloroethene, once thought to be nonmetabolizable, can be decomposed by reductive dehalogenation (figure 2A), thus serving as the final electron acceptor in microbial anaerobic respiration occurring in contaminated sediments.[10] Tetrachloroethene reductive dehalogenase (EC 1.97.1.8) contains two 4Fe-4S clusters and a cobalamin.[11] It is not clear if this dehalogenase evolved from an ancient reductive-type enzyme scaffold, and what its natural substrate may have been.[12] Anaerobic degradation of methyl aromatics involves benzylsuccinate synthase (EC 4.1.99.1), a glycyl radical carbon-carbon lyase, that formally catalyzes the hydroalkylation of fumarate with toluene, cresol, xylenes or methylnaphthalene (figure 2B).[13] The organophosphate insecticide parathion is degraded by phosphotriesterase, which cleaves the phosphorous-oxygen bond shown in figure 2C.[14] This enzyme belongs to the aminohydrolase superfamily together with TriA and AtzA; its natural substrate has not

been identified.[14] It is also possible that phosphotriesterase evolved under selective pressure since the introduction of phosphotriester insecticides some decades ago.

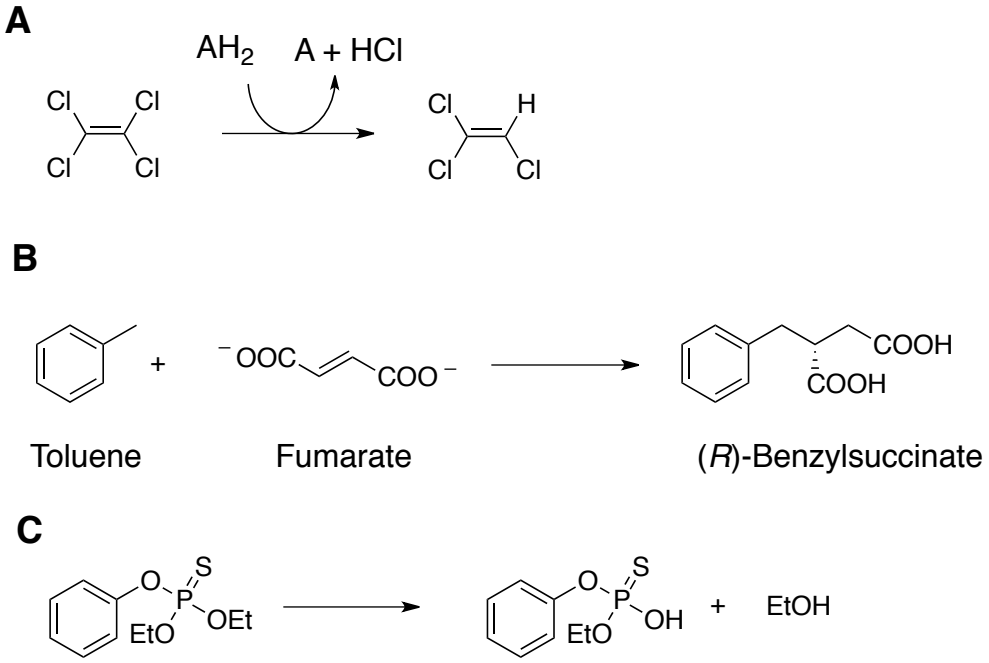


Figure 2: (**A**) Tetrachloroethene reductive dehalogenase catalyzes the reductive dechlorination of tetrachloroethene. AH_2 is an unknown electron donor. **(B)** Benzylsuccinate synthase catalyzes the first step in the anaerobic degradation of toluene, which involves the formal hydroalkylation of fumarate. (**C**) Phosphotriesterase, a member of the aminohydrolase superfamily, catalyzes the hydrolysis of the phosphotriester insecticide parathion.

Novel Enzymatic Activity can be Discovered by Screening Natural Enzymes Against Synthetic Reagents in the Laboratory

Reaction promiscuity in enzymes is often the starting point for development of unnatural bond-making or bond-breaking catalysts. In this context, reaction promiscuity refers to the ability of enzymes to utilize functional groups not found in nature or else different from those present in its natural substrate. The products from transformations with synthetic or unnatural reagents can be quite different from the enzyme's natural

product; however, both reactions often proceed via a common intermediate or share mechanistic similarities such as a deprotonation/nucleophilic addition sequence. Many examples of reaction promiscuity exist in nature,[1] such as the ability of some lipases to hydrolyze ester and amide bonds, but only bond connections that have no natural counterpart or involve unnatural functional groups will be discussed here.

Hydroxynitrile lyases (HNLs) constitute a family of versatile enzymes that catalyze the reversible cleavage of α-hydroxy nitriles and are utilized for the production of enantiopure cyanohydrins from aldehydes or ketones and HCN.[15] Based on the mechanism of this biotransformation,[16] Purkarthofer *et al.* recognized that HCN could be replaced by other nucleophiles of similar size and acidity to HCN (pK_a ~9).[17] The nitroaldol (or the Henry) reaction is a synthetically useful carboligation process involving the addition of nitroalkanes to carbonyl compounds to furnish vicinal nitroalcohols. Addition of nitromethane to benzaldehyde in the presence of the hydroxynitrile lyase from *Hevea brasiliensis* gave 2-nitro-1-phenylethanol in 63% yield with an enantiomeric excess of 92% (figure 3A).[17] This constituted the first example of a biocatalytic asymmetric nitroaldol reaction. Nitroalkanes are not common in nature. The enzyme also showed promiscuous activity for the addition of nitroethane to benzaldehyde, a reaction that generates two stereocenters simultaneously. The product mixture contained almost 90% of the main product (1*S*,2*R*)-2-nitro-1-phenylpropanol with a diastereoselectivity of 9:1 *anti*/*syn* ratio and an enantiomeric excess of 95% for the *anti* isomer.[17]

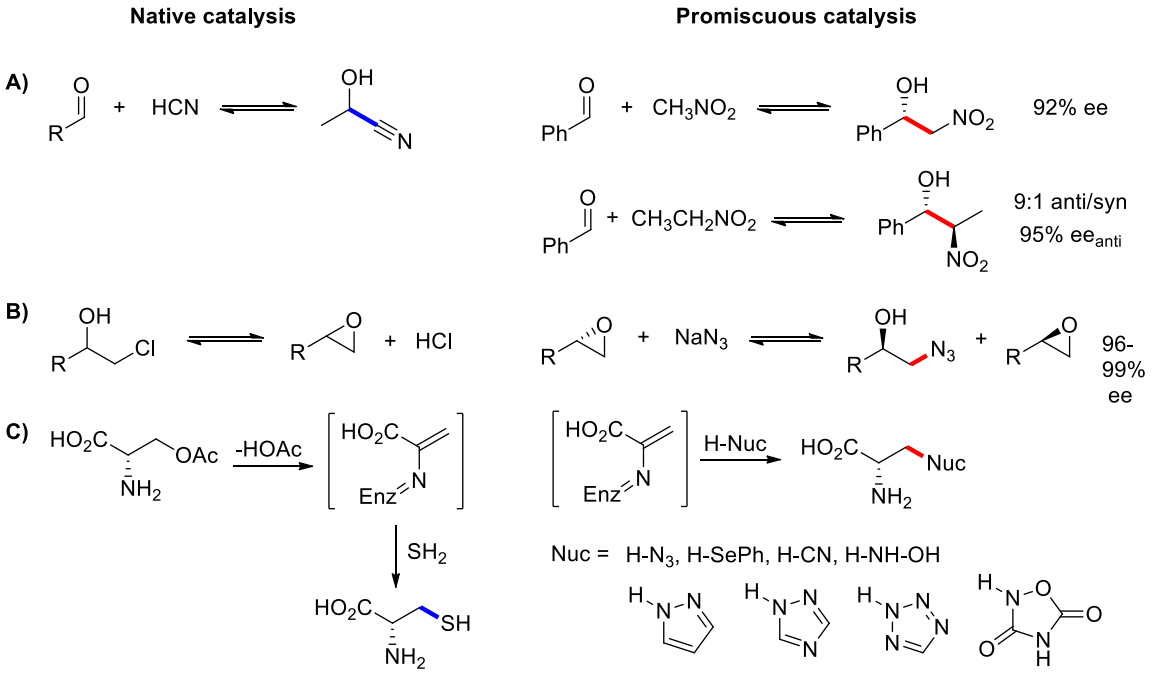


Figure 3: Native and promiscuous activities for (**A**) hydroxynitrile lyases; (**B**) halohydrin dehydrogenases; and (**C**) *O*-acetylserine sulfhydrylase.

In nature, halohydrin dehalogenases (HHDHs) catalyze both the decomposition of chloro- and bromohydrins into epoxides.[18] Due to the ability of *Agrobacterium radiobacter* HHDH strain AD1 to act on a variety of substrates and catalyze both epoxide formation and the microscopic reverse epoxide ring opening with nucleophile, Spelberg *et al.* posited that native enzyme may be used for styrene epoxide ring opening with azides.[19] In contrast to halohydrins, the azido-alcohol product that would be formed from epoxide ring opening with azide is thermodynamically favored (K_{eq} >33). Reaction of the racemic styrene epoxide with *Agrobacterium radiobacter* AD1 yielded both the (*S*)-styrene oxide and the (*R*)-azido alcohol in 95% combined yield and 96%-99% enantioselectivity (figure 3B). Furthermore, in 2007, Codexis reported an engineered

variant of *Agrobacterium radiobacter* AD1 for the enantioselective synthesis of ethyl (R)-4-cyano-3-hydroxybutyrate, the starting material for commercial synthesis of atorvastatin (Lipitor), from (*S*)-4-chloro-3-hydroxybutyrate and sodium cyanide.[20] The reaction proceeds through two HHDH-mediated steps: initial epoxide formation from the chlorohydrin, then subsequent epoxide opening by the cyanide nucleophile. After 18 rounds of recombination-driven directed evolution using protein sequence-activity relationships, mutants were identified that provided an approximately 4,000-fold improvement in volumetric productivity compared to the wild-type enzyme.

Another enzyme that has been found to react with a variety of synthetic nucleophiles is *O*-acetylserine sulfhydrylase, which catalyzes the biosynthesis of cysteine from serine and sulfide in nature (figure 3C).[21] The enzyme facilitates the elimination of acetate from L-*O*-acetylserine to form an aminoacrylate intermediate that was found to be reactive toward nucleophiles such selenol, azide, hydroxylamine, and a few N-containing heterocycles in addition sulfide. Discovery of these transformations has enabled *in vivo* synthesis of a variety of unnatural amino acids in 45%-91% yield.

Members of the cytochrome P450 enzyme family catalyze myriad oxidative transformations involving different bond disconnections, including hydroxylation, epoxidation, oxidative ring coupling, oxidative decarboxylation, heteroatom release, and heteroatom oxygenation.[22] Most transformations encompassed by this broad catalytic scope manifest the reactivity of the same high-valent iron-oxene intermediate, compound I (figure 4). We hypothesized that P450s could be engineered to catalyze the transfer of carbenes and nitrenes from diazoesters and azides, respectively, to C-H and C=C bonds, akin to the well-established isoelectronic oxene transfers from iodosylbenzene (figure

4A). Ethyl diazocarboxylate (EDA) is a commercially available and highly reactive carbene precursor that has been used to generate many transition metal carbenoids for cyclopropanation, and we hypothesized this synthetic starting material could be used to generate Fe-carbenoids for cyclopropanation. When we combined EDA with wild-type BM3 and styrene alone, no reaction was observed. However, addition of EDA to BM3 and styrene in the presence of a reductant led to trace amounts of cyclopropanes (~20 catalytic turnovers), suggesting that the reduced Fe(II)-heme could access a high oxidation state Fe-carbenoid intermediate. From this starting point, directed evolution and site saturation mutagenesis provided variants that "traded" native monoxygenation activity for cyclopropanation activity and produced styrenyl cyclopropanes in excellent selectivity and high turnover. Similarly, azides have been known to react with transition metal complexes to generate metal nitrenes that can then react with benzylic C-H bonds. When wild type BM3 was combined with a benzylsulfonyl azide and a reductant, the nitrene insertion product was observed, albeit with 2 total turnovers. Again, optimization of this promiscuous activity yielded variants that could catalyze C-H insertion in good yield and selectivity. By exploiting the ability of P450 to accommodate high oxidation state iron complexes and identifying analogous relationship of metal-carbenes, metal-nitrenes and metal-oxenes, we were able to adapt P450 for reaction with synthetic, bioorthogonal reagents. Studies to elucidate the reactive intermediates and catalyst resting state in these pathways are ongoing. While the formation of the cyclopropane and nitrene C-H insertion products do not preclude the possibility that the reactions proceed through mechanisms that do not involve Fe-carbenoid and Fe-nitrenoid complexes,

analysis of P450 reaction promiscuity nonetheless led to discovery of two novel enzyme-mediated bond connections.

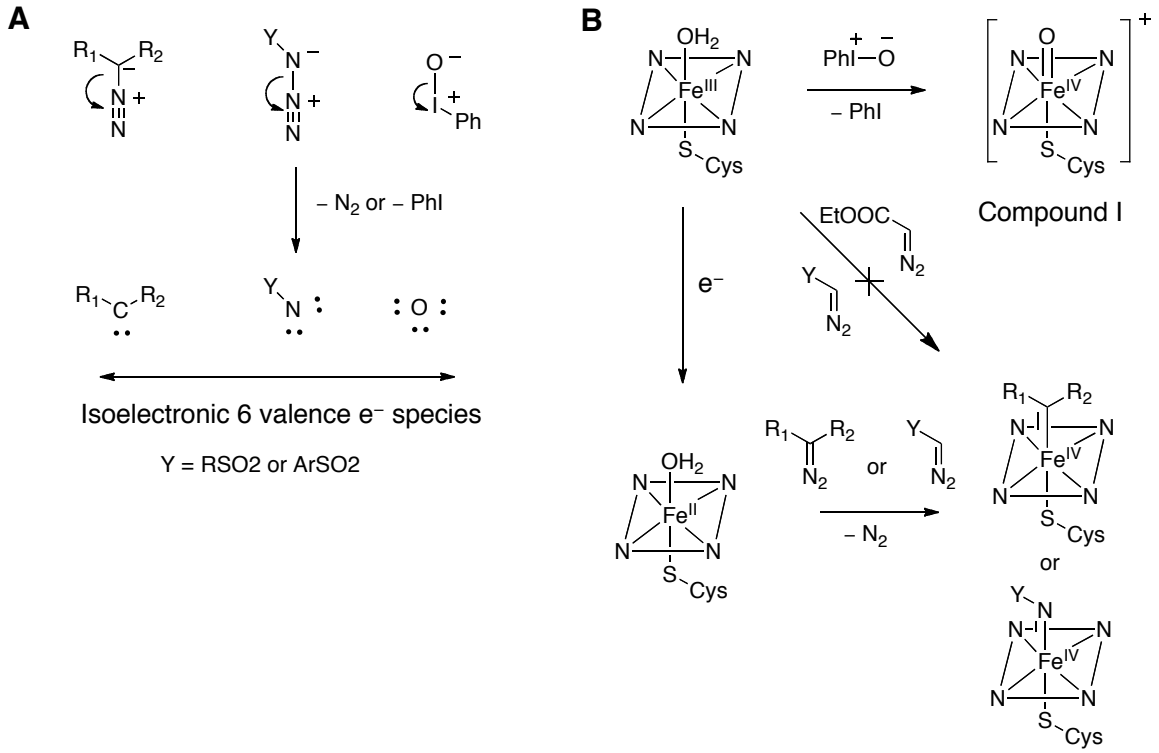


Figure 4: (**A**) Carbene, nitrene and oxene transfer reagents are isoelectronic. (**B**) Putative intermediates for carbene and nitrene transfers are not isoelctronic to compound I.

Preadaptation

Simply put, the principle of preadaptation asserts that a structure can change its function radically without altering its form as much.[23] Evolutionary biologists use preadaptation to address the question: how can selection act on a structure before that structure has appeared? That is, how can a structure be selected for before it has acquired selective value? For example, preadaptation is often invoked to explain the evolution of "perfect" organs by devising adaptive significance for the incipient stages in their

development. The question "what use would 5% of an eye be?" is avoided by arguing that the possessor of such an incipient structure did not use it for sight.

Biologists use preadaptation to rationalize events in historical time. For instance, given the present observation of complex organs, they try to recreate intermediate stages that use elements of such structure for different functions. We would like to use concepts from preadaptation to engineer synthetic catalytic functions in natural enzymes. Given our knowledge of the structure of existing enzymes, can we predict which structures would be more suitable or preadapted for the desired novel function? In Biology, structures are fortuitously preadapted for new functions because the organism is ignorant of future environmental changes. In the lab, however, we can equate environmental changes to the introduction of synthetic reagents, which we have complete control of. We define a novel function here as the catalysis of a nonnatural bond disconnection that results from changes in the amino acid sequence of the enzyme. We illustrate this approach using olefin cyclopropanation by cytochrome P450s as an example.

Cytochrome c (cyt c), myoglobin (Mb), horseradish peroxidase (HRP), and cytochrome $P450_{BM3}$ all show basal levels (<50 TTNs) of promiscuous catalytic activity for olefin cyclopropanation because they all share the heme cofactor, which is also a cyclopropanation catalyst on its own (note that other heme enzymes, such as catalase and chloroperoxidase, do not catalyze olefin cyclopropanation).[24] A single mutant of $P450_{BM3}$ (T268A) is able to make hundreds of turnovers of cyclopropananes with 96% enantioselectivity and 99% diastereoselectivity for the *trans* isomer. Are there single mutants of the other heme proteins (cyt *c*, Mb and HRP) that will also efficiently and selectively catalyze olefin cyclopropanation? That is, are all promiscuous enzymes

equally preadapted for the new function? We argue that the answer is no because the native function of those proteins never required them to evolve the ability to catalyze asymmetric chemical transformations.

Clearly, there are other preexisting elements in BM3-T268A, apart from Ala268, that enable the high performance of this catalyst. These preexisting elements evolved under selection for the native function, fatty acid monooxygenation, but also happen to be useful for olefin cyclopropanation. BM3 monooxygenations are known to be stereoselective, and so its uniquely buried active site is endowed with the capability to induce asymmetry for both oxene and carbene transfer reactions. Furthermore, axial cysteine to serine substitution (BM3-C400S, which we call ABC) enables thousands of NADH-driven cyclopropanation turnovers, whilst abolishing the native monooxygenation activity. This is similar to the evolution of TriA from AtzA, but in the P450 case a single mutation achieves specialization to the new function. An outside observer without knowledge of TriA or BM3 would not be able to recreate evolutionary trajectories for AtzA and ABC because the intermediates would be virtually nonfunctional for the ultimate chlorohydrolase or olefin cyclopropanation activities. The evolution of AtzA and ABC is possible because the intermediates were used for a different function (i.e., pyrimidine-NH_2 hydrolysis for AtzA and C-H monoooxygenation for ABC). Structural features that were selected for the old function were then fortuitously adapted for the new synthetic function.

Let us now reverse the order of this argument. Given TriA, BM3, atrazine, and diazoesters. could we have engineered AtzA and ABC? The reason why the answer to this question is yes is because the synthetic activities (chlorohydrolase and C=C

cyclopropanation) share mechanistic similarities with the native activities (*vide supra*). It is important to realize that a detailed knowledge of the structure of TriA, BM3 or of the intricate details of the mechanism of the target reactions are not necessary for choosing TriA and BM3 as starting points for directed evolution projects. All that is required is the formulation of a basic hypothesis based on a simplified view of the target mechanism, as shown in figure 5 for carbene transfers. We see that by following the mechanism-based approach, we would hypothesize that any enzyme that evolved for catalyzing oxene transfers are likely to be preadapted for carbene transfers. Shortlisted natural enzymes would include cytochrome P450, α-ketoglutarate dependent monooxygenase, diiron monooxygenases, and dicopper monooxygenases. Knowledge about the cofactor composition of the active site of these enzymes is not required to arrive at this short list. Variants of these enzymes can then be screened against diazoesters and olefins to find those suitable for the target reaction. The alternative structure-based approach (figure 5) is somewhat more limited because there is often only a small overlap between the structure of cofactors used in nature and those used in synthetic chemistry. We believe it is more fruitful to focus on the bond disconnections involved in the reaction and the necessary mechanistic features that the candidate enzyme must have to enable the new function.

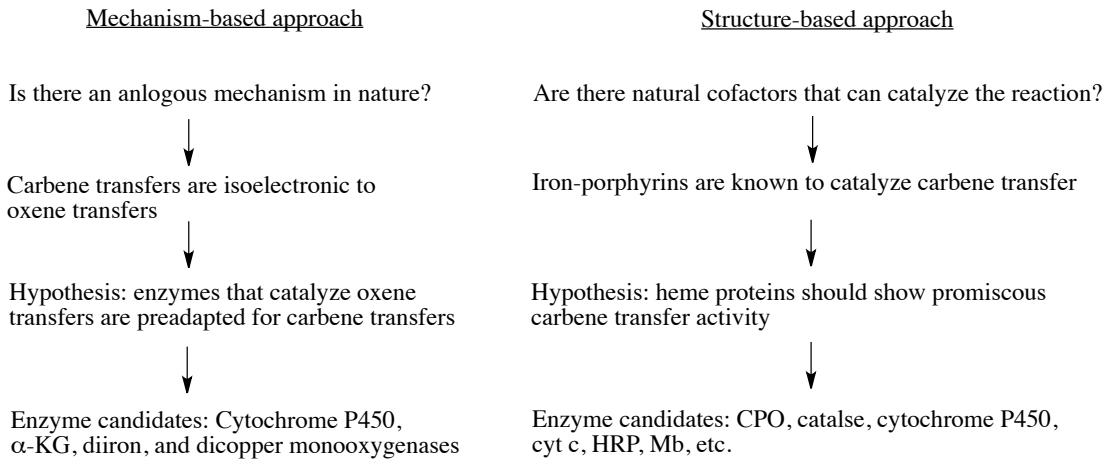


Figure 5: Mechanism and structure based approaches for finding natural enzymes that can be adapted for synthetic reactions. α-KG (α-ketoglutarate); CPO (chloroperoxidase).

The alternative strategies for creating enzymes with synthetic catalytic functions, namely computational design and protein-bound transition metal catalysts, were reviewed in the end of chapter 1. Computational design requires a detailed understanding of the mechanism of the target reaction and has thus far not been applied for transition-metal catalyzed transformations. In protein-bound transition metal catalysts, the new chemical capabilities are all enabled by the abiological organometallic cofactor. While the approach is highly modular, tethering of the cofactor has only been reported *in vitro*, and it is not clear if the platform could enable *in vivo* catalysis. The comparative advantages of the approach outlined here are that by using existing enzymes one retains several features that are desirable for both *in vitro* and *in vivo* catalysis such as thermostability and functional expression *in vivo* without the addition of exogenous cofactors. Accordingly, the ABC catalyst described in chapter 3 is the first report of efficient catalysis of a synthetic bond disconnection *in vivo*.

.

References

1. Bornscheuer, U. T.; Kazlauskas, R. J., Reaction specificity of enzymes: Catalytic promiscuity in biocatalysis: Using old enzymes to form new bonds and follow new pathways. *Angewandte Chemie, International Edition* **2004,** *43* (45), 6032-6040.
2. Seffernick, J. L.; Wackett, L. P., Rapid Evolution of Bacterial Catabolic Enzymes: A Case Study with Atrazine Chlorohydrolase. *Biochemistry* **2001,** *40* (43), 12747-12753.
3. Cook, A. M.; Huetter, R., s-Triazines as nitrogen sources for bacteria. *Journal of Agricultural and Food Chemistry* **1981,** *29* (6), 1135-43.
4. Holm, L.; Sander, C., An evolutionary treasure: unification of a broad set of amidohydrolases related to urease. *Proteins: Structure, Function, and Genetics* **1997,** *28* (1), 72-82.
5. Scott, C.; Jackson, C. J.; Coppin, C. W.; Mourant, R. G.; Hilton, M. E.; Sutherland, T. D.; Russell, R. J.; Oakeshott, J. G., Catalytic improvement and evolution of atrazine chlorohydrolase. *Applied and Environmental Microbiology* **2009,** *75* (7), 2184-2191.
6. De Souza, M. L.; Seffernick, J.; Martinez, B.; Sadowsky, M. J.; Wackett, L. P., The atrazine catabolism genes atzABC are widespread and highly conserved. *Journal of Bacteriology* **1998,** *180* (7), 1951-1954.
7. Seffernick, J. L.; De Souza, M. L.; Sadowsky, M. J.; Wackett, L. P., Melamine deaminase and atrazine chlorohydrolase: 98 percent identical but functionally different. *Journal of Bacteriology* **2001,** *183* (8), 2405-2410.
8. Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J., Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to Sitagliptin Manufacture. *Science* **2010,** *329* (5989), 305-309.
9. Raillard, S.; Krebber, A.; Chen, Y.; Ness, J. E.; Bermudez, E.; Trinidad, R.; Fullem, R.; Davis, C.; Welch, M.; Seffernick, J.; Wackett, L. P.; Stemmer, W. P. C.; Minshull, J., Novel enzyme activities and functional plasticity revealed by recombining highly homologous enzymes. *Chemistry & Biology* **2001,** *8* (9), 891-898.
10. Maymo-Gatell, X.; Chien, Y.-t.; Gossett, J. M.; Zinder, S. H., Isolation of a bacterium that reductively dechlorinates tetrachloroethene to ethene. *Science* **1997,** *276* (5318), 1568-1571.
11. Neumann, A.; Wohlfarth, G.; Kiekert, G., Purification and characterization of tetrachloroethene reductive dehalogenase from Dehalospirillum multivorans. *Journal of Biological Chemistry* **1996,** *271* (28), 16515-16519.
12. Magnuson, J. K.; Romine, M. F.; Burris, D. R.; Kingsley, M. T., Trichloroethene reductive dehalogenase from Dehalococcoides ethenogenes: sequence of tceA and substrate range characterization. *Applied and Environmental Microbiology* **2000,** *66* (12), 5141-5147.
13. Boll, M.; Fuchs, G.; Heider, J., Anaerobic oxidation of aromatic compounds and hydrocarbons. *Current Opinion in Chemical Biology* **2002,** *6* (5), 604-611.

14. Raushel, F. M.; Holden, H. M., Phosphotriesterase: an enzyme in search of its natural substrate. *Advances in Enzymology and Related Areas of Molecular Biology* **2000,** *74*, 51-93.
15. Johnson, D. V.; Zabelinskaja-Mackova, A. A.; Griengl, H., Oxynitrilases for asymmetric C-C bond formation. *Current Opinion in Chemical Biology* **2000,** *4* (1), 103-109.
16. Gruber, K.; Gartler, G.; Krammer, B.; Schwab, H.; Kratky, C., Reaction Mechanism of Hydroxynitrile Lyases of the α/β-Hydrolase Superfamily: The three-dimensional structure of the transient enzyme-substrate complex certifies the crucial role of LYS236. *Journal of Biological Chemistry* **2004,** *279* (19), 20501-20510.
17. Purkarthofer, T.; Gruber, K.; Gruber-Khadjawi, M.; Waich, K.; Skranc, W.; Mink, D.; Griengl, H., A biocatalytic Henry reaction—the hydroxynitrile lyase from Hevea brasiliensis also catalyzes nitroaldol reactions. *Angewandte Chemie, International Edition* **2006,** *45* (21), 3454-3456.
18. Van Hylckama Vlieg, J. E. T.; Tang, L.; Spelberg, J. H. L.; Smilda, T.; Poelarends, G. J.; Bosma, T.; Van Merode, A. E. J.; Fraaije, M. W.; Janssen, D. B., Halohydrin dehalogenases are structurally and mechanistically related to short-chain dehydrogenases/reductases. *Journal of Bacteriology* **2001,** *183* (17), 5058-5066.
19. Spelberg, J. H. L.; van Hylckama Vlieg, J. E. T.; Tang, L.; Janssen, D. B.; Kellogg, R. M., Highly enantioselective and regioselective biocatalytic azidolysis of aromatic epoxides. *Organic Letters* **2001,** *3* (1), 41-43.
20. Fox, R. J.; Davis, S. C.; Mundorff, E. C.; Newman, L. M.; Gavrilovic, V.; Ma, S. K.; Chung, L. M.; Ching, C.; Tam, S.; Muley, S.; Grate, J.; Gruber, J.; Whitman, J. C.; Sheldon, R. A.; Huisman, G. W., Improving catalytic function by ProSAR-driven enzyme evolution. *Nature Biotechnology* **2007,** *25* (3), 338-344.
21. Flint, D. H.; Tuminello, J. F.; Miller, T. J., Studies on the synthesis of the Fe-S cluster of dihydroxy-acid dehydratase in Escherichia coli crude extract. Isolation of O-acetylserine sulfhydrylases A and B and cystathionase based on their ability to mobilize sulfur from cysteine and to participate in Fe-S cluster synthesis. *Journal of Biological Chemistry* **1996,** *271* (27), 16053-16067.
22. Isin, E. M.; Guengerich, F. P., Complex reactions catalyzed by cytochrome P 450 enzymes. *Biochim. Biophys. Acta, Gen. Subj.* **2007,** *1770* (3), 314-329.
23. Gould, S. J., *Ever Since Darwin*. Norton: New York, 1979.
24. Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H., Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science* **2013,** *339* (6117), 307-310.