New-to-Nature Selective C–H Alkylation Using Engineered Carbene Transferases

Thesis by

Juner Zhang

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy in Chemistry

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2022

(Defended May 9, 2022)

Juner Zhang
ORCID: 0000-0001-8181-1187

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor and role model, Prof. Frances Arnold, for her great support and guidance throughout my PhD. Frances puts up really high standards about how we should do science: She has a skill for identifying some of the most important questions and is never afraid to push boundaries between different subjects to seek answers. I also really appreciate that she is the person who often asks the most challenging questions and prompts us to push back. Throughout these discussions, I have developed a more independent mind and taste to scrutinize and fearlessly pursue scientific questions about which I am passionate.

Thanks to all of my committee members, Prof. Sarah Reisman, Prof. Theo Agapie, and Prof. Lu Wei. Each of them has shared their own expertise and provided me tremendous support throughout each meeting we had. In particular, Sarah always asks questions that inspire me to think deeply about my research and proposals in a broader context. Her impressive knowledge and insights in synthetic chemistry have also been a great source of guidance to me. Theo's inorganic chemistry class might be the best class I have ever taken at Caltech. It was the interests I have developed in the class which have made me always passionate about research on metalloenzyme mechanisms. I would also like to thank Lu for all her advice, which motivated me to think about how I can make my enzymes useful to biologists. Looking back, I wish I have interacted with them more and took the opportunities to learn more from each of them.

I am grateful to my mentors and colleagues from the Arnold lab. Back when I was a rotation student, Xiongyi saw the potential in me and took me on as his mentee, even though I did not have any previous expertise in either synthetic chemistry or enzymology. The interactions between us have had a tremendous impact on my career. I also want to express my strong gratitude to Sabine. She is more than a lab manager to many of us; Sabine always makes the effort to ensure everyone in the lab is in a good state and points out every single grammar mistake I make in my writings. I would like to thank Nick Porter for teaching me about crystallography; Edwin Alfonzo, Kai Chen, Anuvab Das, David Miller, Runze Mao, Zhen Liu, Yang Yang, Kelly Zhang for the discussions about chemistry; Noah Dunham and Ben Levin for their insights in enzymology; and many others for their help. The Arnold lab is a great place with many creative and friendly people; I feel very blessed to be part of it.

Life in graduate school can be tough, especially in the era of Covid-19. I am very thankful that I have been surrounded by many good friends, particularly, Ailiena, Alex, Ella, Felix, Jiajun, Kadina, Molly, Morgan, Patrick, Shucong, Shuoyan, Tina, Yue, Yuxing, Yujia, and many others. I feel quite lucky to have met my partner, Timothy Smith, who has been a strong support and great companion to me on this journey. Lastly, I would like to thank my mom. Even though home seems to be so far away with all the travel restrictions, she can always find a way to ease my mind and remind me that I am loved and supported.

ABSTRACT

Synthetic methods to selectively convert C–H bonds, a prevalent motif in organic molecules, into functionalities can significantly accelerate the syntheses and derivatization of molecules. In the past decade, many enzymatic catalysts have emerged as greener, more selective, and more versatile alternatives to small-molecule catalysts for selective C–H functionalization reactions. In nature, enzymes only catalyze a limited set of C–H functionalization reactions that are useful for chemical synthesis, as the overwhelming majority of known C–H functionalization enzymes in nature are hydroxylases. The diversity of the enzymatic reaction scope needs to be substantially expanded to make them broadly useful to synthetic chemists. This thesis will describe new enzymes, which we repurposed from one of the most prevalent C–H hydroxylases in nature, cytochromes P450, which are now able to catalyze new-to-nature C–H alkylation reactions via selective carbene transfer. Given the central role of C–C bond forming reactions in building and elaborating the carbon skeleton of organic molecules, these transformations are of high interest in many fields of research, such as medicinal chemistry and material chemistry. In Chapter 1, I review a recent surge in newly identified enzymes, repurposed enzymes, and artificial metalloenzymes which can catalyze selective C–H functionalization reactions. Chapter 2 details the development of a panel of enantiodivergent α-amino C(sp^3)−H fluoroalkylases. Using directed evolution, the carbene transferases can install fluoroalkyl groups onto these C–H bonds with high activity (4,070 total turnovers, TTN) and selectivity (>99% ee). Notably, complementary regioselectivity can be achieved using an alternative enzyme, P411-PFA-(*S*). In Chapter 3, I report the first carbene transferase, P411-ACHF, which can transfer an α-cyanocarbene to arene C–H bonds of *N*-substituted benzenes. Chemodivergent C(sp^2)–H and C(sp^3)–H functionalization can be achieved using P411-ACHF and P411-PFA. Additionally, structural studies revealed an unprecedented backbone carbonyl flip within the long *I*-helix of P411-PFA, which may suggest how these enzymes have evolved to bind and activate diazo compounds for carbene transfer reactions. In Chapter 4, I discuss the efforts I took toward stabilizing an interesting but unstable P450, CYP3A4. This enzyme exhibits large active site volume and high substrate promiscuity and therefore can be a great candidate to develop late-stage carbene and nitrene transferases. I adopted consensus sequence

mutagenesis and predicted five mutations which have the potential to have the strongest beneficial effects on improving CYP3A4's thermostability. In summary, this thesis work addresses the urgent need for expansion of the current enzymatic C–H functionalization reaction scope and the development of more sustainable and selective C–H functionalization catalysts which can be synthetically useful.

PUBLISHED CONTENT AND CONTRIBUTIONS

(1) **Zhang, J.** †; Huang, X. †; Zhang, R. K.; Arnold, F. H. Enantiodivergent α-amino C–H fluoroalkylation catalyzed by engineered cytochrome P450s. *Journal of the American Chemical Society* **2019**, 141, 9798–9802. DOI: 10.1021/jacs.9b04344. (†indicates equal contribution by authors)

J.Z. and X.H. designed the overall research with F.H.A. providing guidance. J.Z. and X. H. designed and conducted the initial screening of haem proteins; J.Z. performed the directed evolution experiments. J.Z., X.H., and R.K.Z. designed and performed the substrate scope studies. J.Z., X.H., and F.H.A. wrote the manuscript with input from R.K.Z.

(2) **Zhang, J.**; Maggiolo, A. O.; Alfonzo, E.; Mao, R.; Porter, N. J.; Abney, N.; Arnold, F. H. Chemodivergent C(sp^3)–H and C(sp^2)–H cyanomethylation using engineered carbene transferases. *To be submitted.*

J.Z. designed the overall research with F.H.A. providing guidance. J.Z. designed and conducted the initial screening of heme proteins; J.Z. and N.A. performed the directed evolution experiments. J.Z., E.A., and R.M. designed and performed the substrate scope studies and analysis. A.O.M. obtained the X-ray crystal structure of the engineered proteins with N.J.P. providing help. J.Z. and F.H.A. wrote the manuscript with input from all authors.

(3) **Zhang, J.**; Arnold, F. H.; Selective C–H functionalization using enzymes: where are we now and what might be next. *Manuscript in preparation*.

J.Z. and F.H.A. wrote the review article.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

Å	angstrom(s)
BM3	cytochrome P450 from Bacillus megaterium (CYP102A1)
CO	carbon monoxide
CYP	cytochrome P450 monooxygenase identifier
DMSO	dimethyl sulfoxide
ee	enantiomeric excess
e.r.	enantiomeric ratio
Et	ethyl
EtOH	ethanol
GC	gas chromatography
GC-MS	gas chromatography with mass spectrometry
GOx	glucose oxidase
h	hours
HB_{Amp}	Hyperbroth medium with 0.1 mg/mL ampicillin
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid buffer
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
KPi	potassium phosphate buffer
LB_{Amp}	Luria-Bernati medium with 0.1 mg/mL ampicillin
LC-MS	liquid chromatography with mass spectrometry
M9-N	M9 minimal media without nitrogen source added
Me	methyl
NOD	nitric oxide dioxygenase
OD_{600}	optical density of a cell suspension measured at a wavelength of 600 nm
P450	cysteine-ligated cytochrome P450 monooxygenase
P411	serine-ligated cytochrome P450 enzyme
PDB	Protein Data Bank
RT (or r.t.)	room temperature
T_R	retention time
TTN	total turnover number

Chapter 1

EDITING MOLECULES WITH ENZYMES: AN INTRODUCTION TO ENZYMATIC SELECTIVE C–H FUNCTIONALIZATION

Abstract

Selective C–H functionalization is one of the most attractive yet challenging reactions in synthetic chemistry. Achieving catalyst-enabled high selectivity is an enduring challenge for small-molecule catalysts due to the often-indistinguishable nature of different C–H bonds present in a molecule. Over the past decades, enzymatic catalysts have emerged to become promising alternatives for such transformations; Significant efforts in genome mining, protein engineering, and creating artificial metalloenzymes has yielded a more diverse set of "C–H functionalases" which have great potential to be synthetically useful. In this chapter, I summarize the progress that has been made in this area of research in the last decade. In **Section 1.2**, I survey novel enzymes whose native activities is to catalyze a C–H functionalization reactions. Besides the overwhelming majority, C–H hydroxylases, many C–H halogenases and C–H alkylases which were either recently identified from genome mining research or obtained by protein engineering will be discussed. In **Section 1.3**, I review the non-biological activities found in naturally existing enzymes, namely, carbene- and nitrene-transfer reactions, and efforts in directed evolution to optimize these promiscuous enzymes to become proficient non-natural catalysts. Developments in homogeneous catalysis and protein engineering have also led to the recent renaissance of research to create artificial metalloenzymes (ArMs). In **Section 1.4**, I discuss recent research progress of using ArMs for selective C–H functionalization. Throughout this chapter, I hope to provide an overview of current enzymatic methods to achieve selective C–H functionalization, and how it is different from a decades ago. At the end of the chapter, I give a perspective of what I hope will come next: A more diverse scope of enzymatic transformations for selective C–H functionalization is needed. Research strategies to turn an interesting enzymes into a useful

enzyme also need to be capitalized to aid the implementation of C–H functionalization enzymes into practical organic synthesis.

1.1 Overview

C–H functionalization, the direct and selective conversion of a C–H bond to a functional group, is a highly desired transformation in synthetic chemistry. Many decades of the research developments have been made to achieve highly selective C–H functionalization reactions.[1] The ultimate goal is to directly "edit" complex molecules with any functional groups of interest by using a panel of selective C–H functionalization catalysts. While the concept and utility of selective C–H functionalization are well recognized in the synthetic chemistry community and its related fields, we still have a long way to go from where we are now to what we ideally could achieve. Catalyst-enabled high selectivity becomes more and more difficult with increasing complexity of the target molecule.[2] Currently, many catalysts are developed to functionalize one or a few C–H bonds of a selected substrate. Despite the promising utility of having such a panel of catalysts with orthogonal selectivity to enable divergent C–H functionalization of the same substrate, ligand design to achieve such processes is a problem with no obvious solution: Synthetic methods are not often viable to introduce delicate ligand modifications (e.g., appending a small alkyl group to the ligands via selective C–H alkylation). Moreover, beneficial changes made in such ligands are not always additive.

Enzymes, Nature's biocatalysts, are great catalysts for many synthetically useful transformations.[3,4] Over the past two decades, multiple successful synthesis routes with an engineered enzyme or a cascade of enzymes have been reported.[5–10] The protein scaffolds of enzymes can be viewed as sophisticated "ligands". The activity of an enzyme is often enabled by its cofactor or a few catalytic residues, while the protein scaffold itself confers the selectivity control. Since enzymes are genetically encoded, well-established protein engineering strategies such as directed evolution can be employed to accumulate beneficial

mutations which improve the desired activity and/or selectivity and result in proficient catalysts.

Despite this rapidly growing interest in using biocatalysts for practical syntheses, there is still a notable scarcity of "C–H functionalases" in these processes, which could be attributed to a lack of enzymatic C–H functionalization reactions. In 2011, Lewis et al. reviewed enzymes which catalyze C–H bond functionalization in nature and those which have been used, directly or engineered, in chemical syntheses.[11] The overwhelming majority of these potentially useful C–H functionalases were C–H hydroxylases, which restricts the scope of enzymatic transformations to a narrow set of reactions compared to the vast reaction scope invented by synthetic chemists.

To close the gap between what human chemists have invented and what has been mined from Nature (so far), we need to expand the enzymatic toolbox for selective C–H functionalization. During the past ten years, significant progress has been toward achieving this goal. We have witnessed a burgeoning of new enzymatic transformations, either discovered in nature or engineered in the laboratory, which introduce functionalities other than –OH to a C–H bond. The advancements made in sequencing technology assisted in the identification of novel C–H functionalization enzymes.[12] Genome mining of microbial and plant enzymes which are engaged in natural product biosynthetic pathways have resulted in the discovery of multiple interesting C–H functionalization enzymes. Many of them catalyze C–H functionalization reactions which had been unknown or difficult to implement previously. For instance, a small number of α-ketoglutarate-dependent enzymes (e.g., WelO5, BesD) were found to catalyze selective C(sp^3)–H chlorination of stand-alone substrates.[13–15] These findings have greatly expanded the potential synthetic utilities of this class of enzymes, since the early identified enzymes of this category strictly require substrate chaperone,[16–18] which presented a challenge to engineer these enzymes to halogenate non-native substrates and achieve high turnover numbers. An alternative approach to mining the wealth of enzymes found in Nature is to transfer abiological, man-made C–H functionalization reactions to enzymatic systems. This can either be done by repurposing natural enzymes or using artificial metalloenzymes,[19–21] and promising progress has been

reported. For example, the Arnold lab have reported using engineered heme proteins harboring their native iron-heme cofactors to catalyze carbene- and nitrene-transfer reactions.[22–25] Similarly, carbene- and nitrene C–H insertion reactions have also been reported using artificial heme proteins harboring an iridium-porphyrin cofactor by Hartwig and colleagues.[26–28]

In this chapter, I aim to give an overview of the state-of-art of biocatalysis with enzymes, whose native or laboratory-repurposed functions are selective C–H functionalization reactions. In **Section 1.2**, I will review key enzymes which have been identified or engineered to catalyze C–H functionalization reactions in nature. I will survey case studies on using C–H hydroxylases to aid chemoenzymatic total synthesis and late-stage functionalization of complex molecules. Recent advances on enzymatic C–H halogenation and alkylation will also be discussed. I will address the significant progress made in chemomimetic biocatalysis which further expanded the limited reaction scope of native "C–H functionalases". I will cover the topic of exploiting new-to-nature C–H functionalization activity of enzymes in **Section 1.3** and the use of artificial metalloenzymes in **Section 1.4**. Throughout this chapter, I hope to analyze the successful examples, and provide my perspective on what have made these enzymes of more promising potentials for practical syntheses than the others; Many important features of a potentially useful enzyme, such as expression level, cofactor availability, and protein stability could be oversighted throughout target selection or protein engineering. At the end, I will give my perspective on what the future has in store for this field of research.

1.2 Native enzymes which catalyze selective C–H functionalization reactions

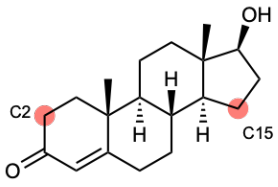
1.2.1 Late-stage C–H hydroxylation using enzymes

Among native C–H functionalization enzymes, C–H hydroxylases have a central position.[11] In fact, C–H hydroxylases might be the only class of C–H functionalization enzymes that has demonstrated synthetic utilities that are comparable or even superior to small-molecule catalysts. Employing C–H hydroxylases for chemical syntheses dates back

to the early 2000s,[29] and since then, extensive research efforts have been devoted to this field of study. A comprehensive review by Hollman and colleagues has summarized the recent development and synthetic utilities of various enzymatic oxyfunctionalization reactions.[30] Here, I will only highlight a small set of examples with the focus on late-stage functionalization and chemoenzymatic synthesis. I will also touch on key factors which can contribute to or detract from an enzyme's potential synthetic utility but are often overlooked by enzymologists, namely protein stability and expression level. I hope to summarize what we have learned from these successful case studies and how these insights may guide future development of other classes of C–H functionalases to become synthetically useful enzymes.

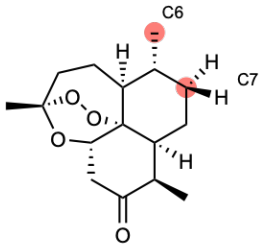
The ultimate goal for developing selective C–H functionalization methods is to enable the late-stage functionalization (LSF) of large, complex molecules.[2,31] In this context, protein catalysts have many advantages compared to small-molecule catalysts. The appeal is that divergent synthesis is more achievable in biocatalysis. This is exemplified by the natural evolution of enzymes: Billions of years of evolution have resulted in a myriad of enzyme families which consist of members with closely related protein sequences. Proteins of an enzyme family can often perform a specific type of transformation on very few common molecular scaffolds, but exert different selectivity controls to yield distinct natural products. For example, terpenes are a class of natural products with more than 30,000 members.[32,33] Such diversity is not solely contributed by their varied scaffolds but also to divergent oxidation patterns.[34] Divergent C–H hydroxylation can also be achieved using laboratory-evolved enzymes. For example, *Bacillus megaterium* P450-BM3 (CYP102A1) has been engineered by our lab to obtain drug metabolites with different hydroxylation patterns. The Reetz lab,[35] Fasan lab,[36] and many others have reported the late-stage, divergent C–H hydroxylation of important bioactive molecules such as testosterone, progesterone, and artemisinin using engineered P450s (**Figure 1-1**). More recently, Renata and colleagues have adopted a panel of P450s and α-ketoglutarate-dependent hydroxylases in the divergent chemoenzymatic syntheses of complex diterpenes.[34] These examples highlight the promising use of enzymes for the LSF of complex molecules. We envision that with the development of other types of C–H functionalization enzymes, enzymatic LSF can greatly facilitate

medicinal chemistry by providing pure constitutional isomers for structure-activity relationship (SAR) studies of pharmaceutical candidates.[31]



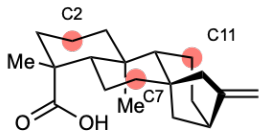
Testosterone

Enzyme variant	Conversion	Hydroxylation selectivity (C2:C15:other products)
KSA-1	79%	97:3:0
KSA-14	85%	3:96:1



Artemisinin

Enzyme variant	Hydroxylation selectivity (C7(*S*):C7(*R*):C6)
IV-H4	100:0:0
II-H10	0:100:0
X-E12	4:2:94



Polycyclic diterpenes

Enzyme variant	Hydroxylation activity
PtmO6 (C7 hydroxylase)	390 TTN
PtmO5-RhFRed (C11 hydroxylase)	88% conversion
BM3 MERO1 M177A (C2 hydroxylase)	51% conversion

Figure 1-1. Selected examples of divergent late-stage C–H hydroxylation catalyzed by engineered or native enzymes.[35,36,34] Red dot denotes positions which reported enzymes can selectively functionalize.

Despite the great diversity of C–H hydroxylases in nature, there are many catalytically interesting enzymes in this category which are not suitable candidates for practical syntheses. For instance, human metabolizing P450, CYP3A4, catalyzes C–H hydroxylation of an impressive array of substrates, and it alone is responsible for 20% of xenobiotic metabolism in human body.[37] However, the potential synthetic utility of this enzyme does not match its extraordinary catalytic profile. One the main issues is the unstable nature of this protein. CYP3A4 is membrane-bound and has a $^{60}T_{50}$ (the temperature at which half of the protein remains folded after 60 min) of only ~40 °C.[38] Research focused on

improving the stability of enzymes has a pivotal role in biocatalysis. Another important factor to consider is the protein expression level. Enzymes which exhibit better protein expression levels are often more favorable for practical synthesis. For instance, in a number of industrial processes, α-ketoglutarate-dependent hydroxylases are more favorable than P450s. Enzyme exploratory research focused on obtaining more robust analogous enzymes, either through bioinformatic analysis or by using protein engineering, can have a strong impact on the synthetic utility of the protein catalysts.

1.2.2 Selective C–H halogenases

The identification and engineering of C–H halogenases might be one of the fastest emerging areas of research in the field of enzymatic chemistry over the past ten years. Given the synthetic utilities and medicinal relevance of halogen-containing molecules, methods to selectively introduce halogen or halogen-like groups into C–H bonds are of great importance. Synthetically useful selective C–H halogenases identified in nature mainly fall into two catagories: flavin-dependent C–H halogenases and α-ketoglutarate-dependent C–H halogenases. Here, I will not include the heme-dependent or vanadium-dependent haloperoxidases. Despite rare cases,[39–42] it is commonly believed that enzyme-controlled selectivity is challenging to achieve in these systems, since the mechanism of halogenation involves the generation of an electrophilic X^+ species, which often diffuses out of the active site to enable the subsequent halogenation reaction. The catalytic profiles of some haloperoxidases are in fact quite impressive, but the lack of selectivity control can be a substantial challenge when engineering these enzymes to selectively halogenate a C–H bond in complex, non-native substrates.

The flavin-dependent halogenases (FDHs) are among some of the most well-studied selective C–H halogenases, and they catalyze the introduction of halogens to an aromatic C–H bond. The commonly accepted mechanism also involves the generation of an electrophilic X^+ species (**Figure 1-2A**):[43,44] the reduced $FADH_2$ first reacts with O_2 to generate an FAD(C_{4a})-OOH intermediate, which subsequently oxidizes X^- into hypohalous acid (HOX). In contrast to haloperoxidases, the HOX species in flavin-dependent halogenases is captured

by a side-chain lysine in the active site, either through hydrogen-bonding or the formation of a chloroamine intermediate. Early studies using structure-based rational protein engineering approaches have set the precedence that the site selectivity[45] and substrate specificity[46,47] of these enzymes can be changed (**Figure 1-2B**). Protein engineering campaigns have greatly accelerated the expansion of the synthetic utilities of FDHs. A pioneering study by Lewis and colleagues demonstrated that tryptophan 7-chlorinase RebH can be stabilized using directed evolution.[57] Moreover, directed evolution enabled higher efficiency in the screening and analysis of sequence-function relationships.[51,53] This resulted in a panel of flavin-dependent halogenases which are active on larger substrates[48] and have non-native site-selectivity.[50] A number of studies have also demonstrated FDHs' utility in chemoenzymatic synthesis.[49,52,54] Selective installation of the halogen offers a great opportunity to further elaborate the molecular scaffold via well-developed coupling reactions without the use of directing groups. Besides laboratory engineering, identifying analogous enzymes with novel activity is an important approach to introduce diversity in the currently existing enzymatic C–H functionalization scope. Previously identified FDHs can only chlorinate or brominate an aromatic C–H bond. Recent studies which have found FDHs that can iodinate a C–H bond have expanded the scope of these enzymes (**Figure 1-2C**).[55,56]

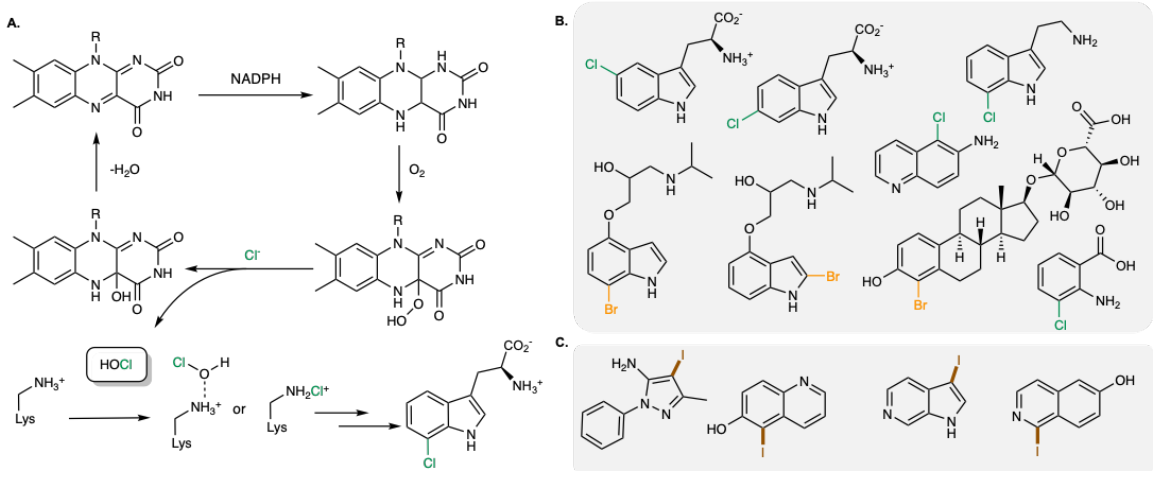


Figure 1-2. Arene C–H halogenation using FDHs. A) Reaction mechanism (RebH is used as an example)[43,44]; B) Selected substrate scope of engineered FDHs[45–54]; C) Selected substrate scope of recently identified FDHs which can iodinate arene C–H bonds.[55,56]

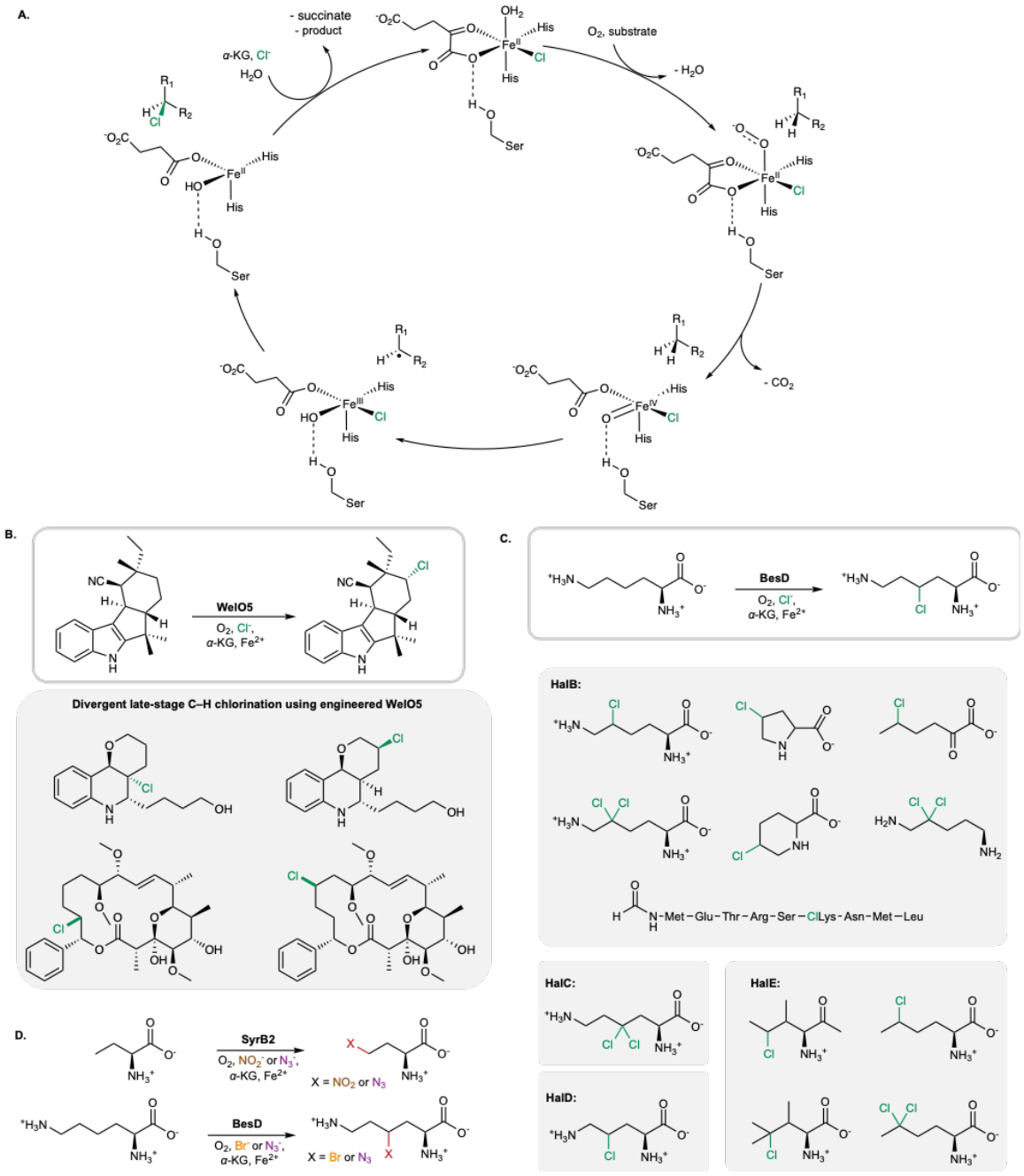


Figure 1-3. C–H halogenation using *α*-ketoglutarate-dependent enzymes (radical C–H halogenases). **A)** Reaction mechanism[58]; **B)** The native and engineered activity of WelO5[13,59–61]; **C)** Substrate scope of BesD and related C–H halogenases[14,15]; **D)** Introducing halogen-like groups into C–H bond using radical halogenases[15,18].

The another interesting class of C–H halogenases is the α-ketoglutarate-dependent enzyme family. In contrast to FDHs, these enzymes often halogenate an aliphatic C–H bond. α-ketoglutarate-dependent halogenases can activate an inert C(sp^3)–H bond in a similar fashion as the α-ketoglutarate-dependent hydroxylases, but the substrate radical preferably rebounds to a halogen radical instead of the hydroxyl radical to form the C–X bond (**Figure 1-3A**).[58] The initial discovery of this activity among the α-ketoglutarate-dependent enzymes dates back to 2005.[16,17] However, the synthetic utility of the early identified enzymes is significantly limited – these enzymes only halogenate substrates that are appended to a carrier protein, which makes it a challenging protein engineering task to change their substrate specificity and achieve high turnover numbers. It was not until the identification of radical C–H halogenases which can act on stand-alone substrates that the avid research interest toward this class of enzymes was ignited. In 2014, the Liu lab discovered an unusual member of this family, WelO5 from *Hapalosiphon welwitschii* UTEX B1830. WelO5 was identified in the welwitindolinone (*wel*) biosynthetic gene cluster and characterized as the first reported α-ketoglutarate-dependent halogenase which can act on stand-alone substrates (**Figure 1-3B**).[13] Separately, Chang and coworkers identified an unexpected α-ketoglutarate-dependent chlorinase in 2019 (**Figure 1-4C**).[14] BesD, from the L-β-ethynylserine biosynthetic pathway, is also a radical chlorinase which works on stand-alone substrates. Subsequent bioinformatic analyses revealed a panel of radical halogenases (HalB-E) which can catalyze selective C–H halogenation on stand-alone amino acid substrates.[15] The identification of these enzymes has sparked significant interest in the protein engineering community. For example, recent studies by Buller and coworkers have demonstrated that WelO5 can be engineered to catalyze late-stage chlorination of non-native substrates in a highly selective fashion (Figure 1-3B).[59–61]

The synthetic utilities of these enzymes do not end with the introduction of halogens into the C–H bond (**Figure 1-3D**). Bollinger and colleagues have shown that SyrB2 can also introduce halogen-like groups such as $–N_3$ and $–NO_2$ via selective C–H functionalization.[18] Such activities have also been demonstrated in carrier-protein-independent radical halogenases.[1] It is within the realm of possibility that directed evolution can further improve

these non-native functions of radical C–H halogenases to aid selective introduction of these important functional groups into molecules of interest.

1.2.3 C–C bond-forming reaction catalyzed by C–H functionalization enzymes

Despite the importance and abundance of C–C bond-forming, C–H functionalization enzymes in nature, there is still a limited number of such enzymes which are synthetically useful. Many C–H alkylation enzymes found in nature belong to the radical (*S*-adenosyl methionine) SAM enzyme family.[62] The burgeoning number of newly identified members in this enzyme family is still disproportionate to the ones which have been used for synthesis.[63] This is due to the unstable and oxygen sensitive nature of these enzymes, which also makes it a challenging task to change the substrate preference of radical SAM enzymes using protein engineering approaches. In rare cases, radical SAM enzymes have been implemented for chemical synthesis either by catalyzing the native reaction,[64] or by changing the alkyl groups they can introduce using analogous SAM cofactors.[65] To our knowledge, there are no literature reports of the use and optimization of any radical SAM enzymes to catalyze C–H alkylation on a non-native substrate.

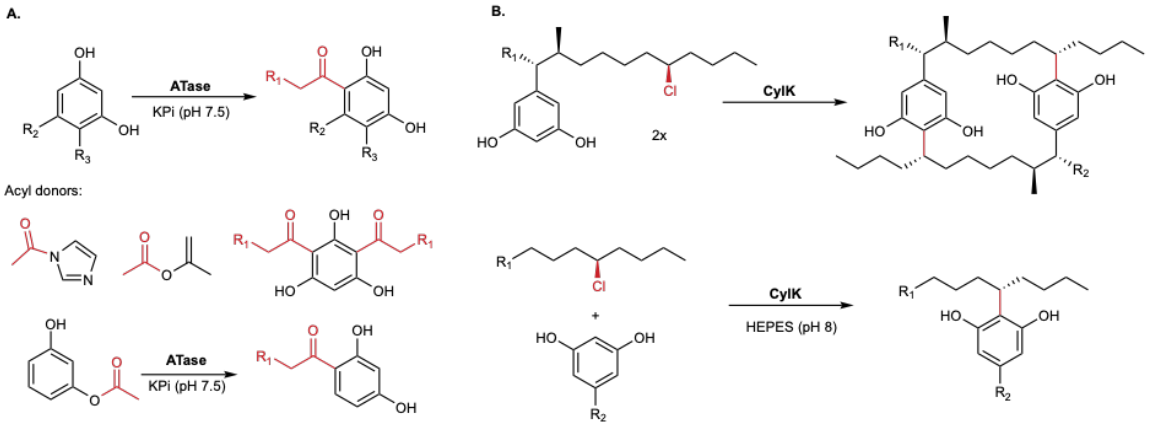


Figure 1-4. Friedel-Crafts acylase[66] and alkylase[67,68] identified in nature.

On the other hand, recent characterization of Friedel-Craft alkylases have expanded potentially engineerable enzymes which can alkylate an arene C–H bond. It is worth noting that enzymatic functionalization of an arene C–H bond rarely proceeds through a classical

C–H activation process, which often involves the metalation of the arene C–H bond. Therefore, I include enzymes that formally form a C–C bond from a C–H bond here. In 2017, Kroutil, Gruber, and coworkers reported an acyltransferase which can catalyze the *C*-acylation of non-natural phenolic substrates.[66] In the same year, the Balskus lab reported a Friedel-Craft alkylase, CylK, which can alkylate activated phenolic arenes with alkyl halide in the cylindrocyclophane biosynthetic pathway.[67] A subsequent study revealed that wild-type CylK is able to catalyze the alkylation of the 2-position in resorcinols with a diverse range of alkyl halide and substituted resorcinols.[68] These studies highlighted the potential of engineering and using these enzymes for site-selective alkylation of activated arenes.

1.3 New-to-nature transformations enabled by laboratory-repurposed enzymes

Another way to expand the scope of enzymatic C–H functionalization reactions is to repurpose existing natural enzymes for new-to-nature transformations. This strategy takes advantage of inherent promiscuity in cofactor-dependent enzymes and uses them to catalyze synthetically useful transformations which were originally established with similar small-molecule catalysts.[20,21] While native enzymes are often inefficient at affording such nonbiological activity with useful yield and kinetic profiles, directed evolution can be used to achieve higher activity and selectivity in enable these new-to-nature transformations. In this section, I will summarize recent advances adopting this strategy.

In the past decades, chemists have designed numerous bio-inspired catalysts to enable synthetically useful transformations. The heme-mimicking metalloporphyrin systems are prominent examples.[71,72] The key intermediate which enables P450s' oxyfunctionalization activity is the high-valent Fe(IV)-oxo radical cation species (P450 Compound I, **Figure 1-5A**).[72,73] Taking inspiration from P450 compound I, many analogous non-biological group-transfer reactions, such as carbene- or nitrene-transfer, have been achieved using various metalloporphyrin systems.[69,74,75] Notably, such activities have also been captured in native heme proteins (**Figure 1-5B**). In 1985, Dawson, Breslow, and coworkers reported the, to our knowledge, first example of enzyme-catalyzed nitrene-transfer reactions.[70] In this study, rabbit liver microsomal P450-LM3 and LM4 were shown

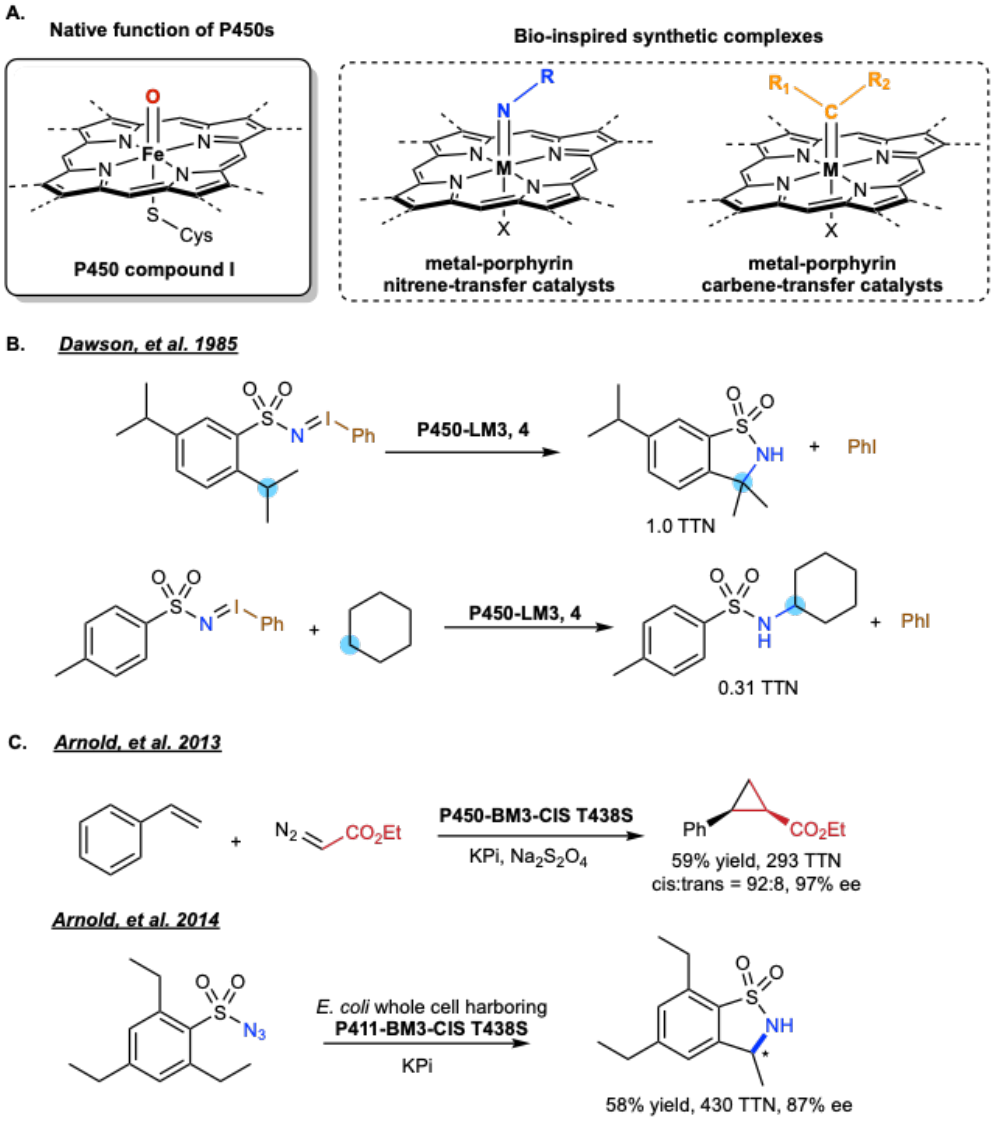


Figure 1-5. Bio-inspired homogeneous catalysis and chemomimetic biocatalysis. **A)** Structures of P450 compound I and the analogous synthetic metalloporphyrin complexes for group transfer chemistry (showing nitrene- and carbene-transfer intermediate)[69]; **B)** Seminal research which demonstrated synthetically important yet non-biological transformations can be achieved using enzyme existing in nature[70]; **C)** Abiological activities of naturally existing enzymes can also be optimized using directed evolution.[22,23]

to catalyze an intra- and intermolecular C–H amidation reaction, mimicking P450's native hydroxylation activity but with very limited catalytic proficiency (turnover numbers ≤ 1.0). While native P450s performed poorly in these transformations, this was a strong indication that the new-to-nature transformation can in fact be achieved using enzymes existing in nature. In 2013, a pioneering study from the Arnold lab demonstrated that the abiological activities of native enzymes are also evolvable (**Figure 1-5C**). In this study, we disclosed the first sets of engineered P450s carbene transferases which catalyze asymmetric olefin cyclopropanation reactions.[22] And soon after, intramolecular nitrene insertion into benzylic C–H bonds was also achieved using engineered P450s with the axial-cysteine-to-serine mutation (P411s).[23,76] Shortly afterwards, Fasan and coworkers reported a similar intramolecular C–H amination activity using engineered P450s[77] and engineered and artificial myoglobins.[78]

Synthetic strategies to build C–N and C–C bonds are among some of the most important bond-forming reactions to build and elaborate molecular scaffolds. Even though there is a scarcity of enzymatic systems that can afford C–N or C–C bonds via C–H functionalization, the Arnold lab and others have demonstrated that a number of heme-containing proteins and α-ketoglutarate-dependent enzymes can be repurposed to catalyze new-to-nature C–N and C–C bond-forming reactions via selective functionalization (**Figure 1-6**). We have also demonstrated that highly selective, intermolecular C(sp^3)–H amination and C(sp^3)–H alkylation can be achieved using engineered P450s which harbor their native iron-heme co-factor.[24,25,79] Such strategies can also be applied to the selective functionalization of an arene C–H bond (**Figure 1-6C**). Fasan and colleagues have shown that engineered myoglobins can catalyze carbene transfer to the 3-position of unprotected indole, using ethyl diazoacetate (EDA) as the carbene precursor.[80] We have also reported that engineered P450s can afford highly regio- and stereoselective functionalization of indoles and pyrroles.[81,82] Beyond highly activated heteroaromatics, I will disclose in Chapter 3 the first set of arene C–H alkylases which are active on substituted benzenes.

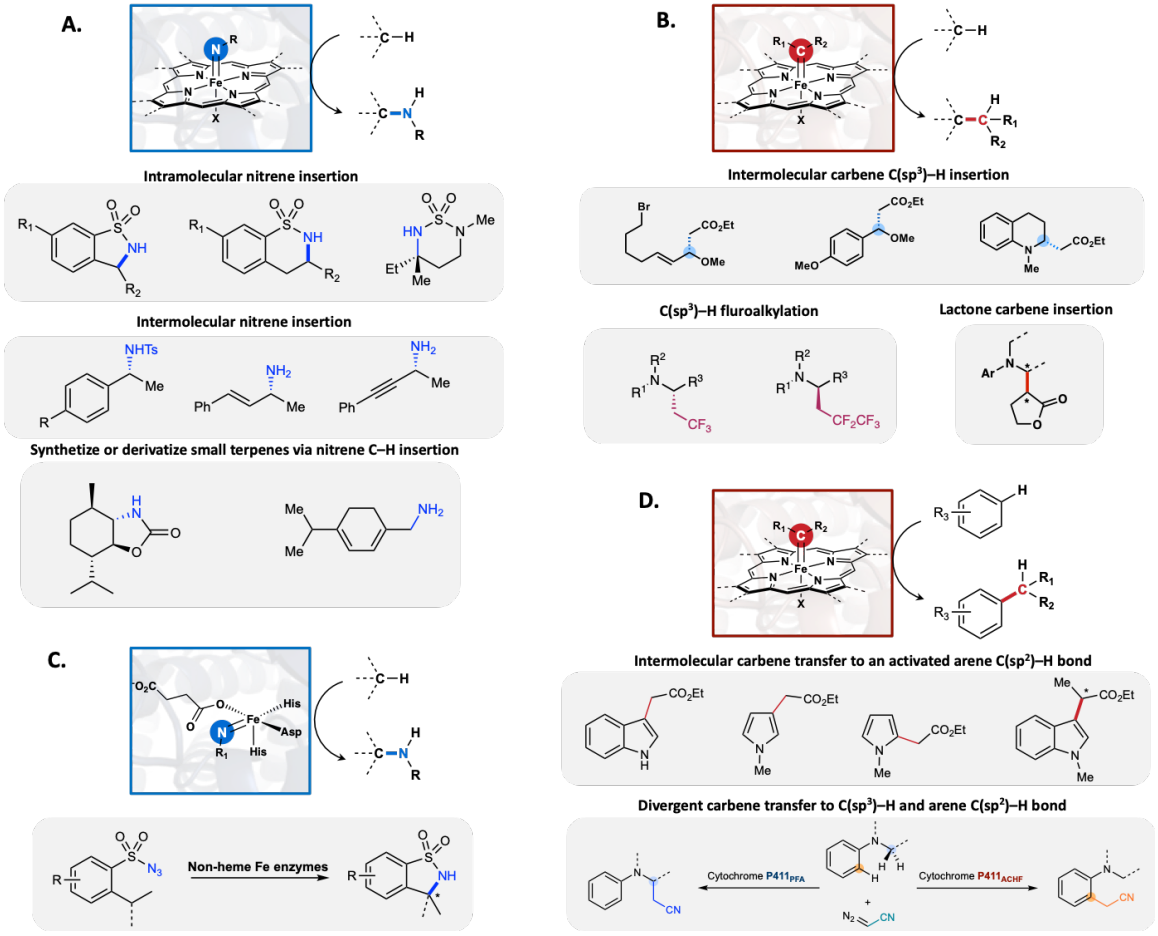


Figure 1-6. Engineered nitrene- and carbene-transferases for selective C–H amination and alkylation reactions. **A)** C(sp^3)–H nitrene insertion using engineered heme proteins[23,79,83–85]; **B)** C(sp^3)–H carbene insertion using engineered heme proteins[25,86,87]; **C)** C(sp^3)–H nitrene insertion using engineered non-heme Fe proteins[88,89]; **D)** C(sp^2)–H alkylation using heme-dependent carbene transferases[80,81].

Given that P450s have a central role in xenobiotic metabolism,[37] we surmised these strategies may offer a promising opportunity to elaborate molecules with medicinal important functionalities at pharmacological-relevant positions. We have demonstrated that a later version of primary aminases can directly and selectively append an $–NH_2$ group to small diterpene substrates (**Figure 1-6A**).[79] Fasan and colleagues applied the enzymatic intramolecular C–H amination reaction to the synthesis of *trans*- and *cis*-oxazolidinone.[85]

For C–C bond forming reactions, a number of natural products have been synthesized using chemoenzymatic approaches.[25] Moreover, medicinally important fluoroalkyl groups (Chapter 2) and lactone groups can also be selectively installed by engineered carbene transferases (**Figure 1-6B**).[86,87]

Unlocking and optimizing these abiological activities are not limited to P450s. In 2019, Goldberg et al. demonstrated that non-heme mononuclear iron enzymes can also be repurposed to catalyze C–H amination reactions via nitrene-transfer reaction (**Figure 1-6D**).[88] Such activity was also later reported by Fasan and colleagues.[89] Interestingly, even though we began our study from an α-ketoglutarate-dependent enzyme, *Ps*EFE (an ethylene forming enzyme from *Pseudomonas savastanoi*), the wild-type enzyme showed significantly improved nitrene-transfer activity when ligands other than the native α-ketoglutarate ligand are used.

These laboratory-repurposed enzymes have shown great potential to become synthetically useful biocatalysts. And yet, there are still a number of limitations preventing them from being broadly applied for practical synthesis: First of all, unlike cytochromes P450 or α-ketoglutarate-dependent enzymes, current nitrene- and carbene-transferases can rarely functionalize unactivated C(sp^3)–H bonds (BDE > 100 kcal/mol) and C(sp^2)–H bonds (alkylbenzene). Unactivated C–H amination was only achieved in intramolecular reactions.[83,84] To achieve selective amination and alkylation of unactivated C–H bonds, additional directed evolution is needed to increase the activity of heme-nitrene and carbene species as well as to remove potential steric clashes between the protein and these intermediates; the heme-nitrene and carbene species are more sterically demanding than P450 compound I, which may prevent the engineered enzymes from performing abiological reactions on larger substrates or substrates with stronger C–H bonds. This is also evidenced by the fact that these enzymes are yet to afford the abiological late-stage functionalization of large, complex molecules, even though late-stage hydroxylation has been extensively pursued using P450-BM3. Moreover, the use of azide and diazo compounds as nitrene and carbene precursors is not ideal. The potential for explosive hazard makes it challenging to adopt both chemocatalytic and biocatalytic carbene- and nitrene-transfer reactions for

industrial processes. The use of alternative nitrene precursors is promising,[79,90,91] but both carbene transfer reactions are still impeded by the lack of non-explosive carbene precursors. Identification and characterization of unusual diazo-forming enzymes have suggested a different approach to circumvent this issue.[92] A viable, alternative way to perform the biocatalytic reaction without using large quantities of the explosive precursor could be an enzymatic cascade: a suitable diazo-forming enzyme transiently generates the diazo compound while the appropriate carbene transferase carries out the C–H alkylation reaction may reduce the potential of explosive hazard.

1.4 Selective C–H functionalization catalyzed by artificial metalloproteins

The creation of artificial metalloenzymes (ArMs) represents another emerging approach to expand the limited reaction scope of native enzymes.[93] This approach takes advantage of both small-molecule catalysis and biocatalysis: The artificial cofactors are often well-documented organometallic complexes which can catalyze synthetically important transformations as the protein scaffold of ArMs can potentially provide more desirable selectivity control. Conceivably, ArMs can be ideal candidates to fill the gap between synthetic chemistry and biocatalysis–a great number of chemist-invented, synthetically useful transformations cannot be readily transferred into naturally existing and laboratory-evolved protein. This is evidenced by the creation of ArMs which can catalyze Suzuki cross-coupling reaction,[94] Heck reaction,[95] and olefin metathesis, reactions that have yet to find their counterparts in Nature.

Aided by the promising development in both the field of organometallic chemistry and protein engineering, over the past two decades, there has been a surge of ArMs which have been developed for C–H functionalization reactions. For example, based on the seminal work done by Whiteside and coworkers using the avidin-biotin binding system to create ArMs,[96] Ward, Rovis, and colleagues have created ArMs which can afford the Rh(III)-catalyzed C–H activation/annulation reactions (**Figure 1-7A**).[97,98] In this example, the Rh(III) complexes were conjugated to biotin and subsequently incorporated into streptavidin (Sav).[99] Modifications to streptavidin have been shown to improve the enantioselectivity of

the transformation. Substitution of the native enzyme's cofactor with a synthetic cofactor can also be an attractive strategy to create C–H functionalization ArMs. Hartwig and colleagues have created a panel of Ir-substituted ArMs by replacing the heme cofactor in myoglobins and P450s with a set of Ir-porphyrin complexes (**Figure 1-7B**).[26,27] Similar to the laboratory-repurposed carbene- and nitrene-transferases, which harbor their native Fe-heme cofactor (discussed in **Section 1.3**), the resulting ArMs in these studies can catalyze selective intra- and intermolecular C–H carbene insertion reactions as well as intramolecular C–H amination.[28,100] The introduction of ncAAs, which can subsequently be elaborated to incorporate metals or organometallic complexes, is also a widely used strategy to create ArMs. For example, the Roefles lab has reported the creation of a Cu(II)-based ArM which can catalyze Friedel-Crafts alkylation of indoles using this approach (**Figure 1-7C**).[101]

Conceptually, ArMs indeed have some great advantages over native enzymes or laboratory-repurposed enzymes regarding the diversity of C–H functionalization strategies they can employ. However, there are many practical limitations preventing ArMs being broadly adopted for chemical synthesis.[93] First of all, many synthetic complexes are not highly compatible with the more complex biological environment. For instance, numerous metabolites produced by cells (e.g. glutathione) can bind to the transition-metal complexes and poison the catalysts. Secondly, unlike the cofactors adopted by nature, whose catalytic activity is often insignificant unless they are properly incorporated into the protein scaffolds, the synthetic cofactors of ArMs are usually self-competent for their activities. This can lead to significant background reactions if excess cofactor is not removed. As a result, the reaction selectivity established by protein-bound cofactor can be diminished by the unbound portion. Moreover, artificial cofactors are not genetically encoded, which complicates the production and protein engineering of ArMs. The development of ArMs is still in its infant stage. Despite the promising features of ArMs, to our knowledge, there are no ArMs being adopted for practical synthesis and extensive research efforts are needed to solve the problems associated with engineering or implementation of ArMs.

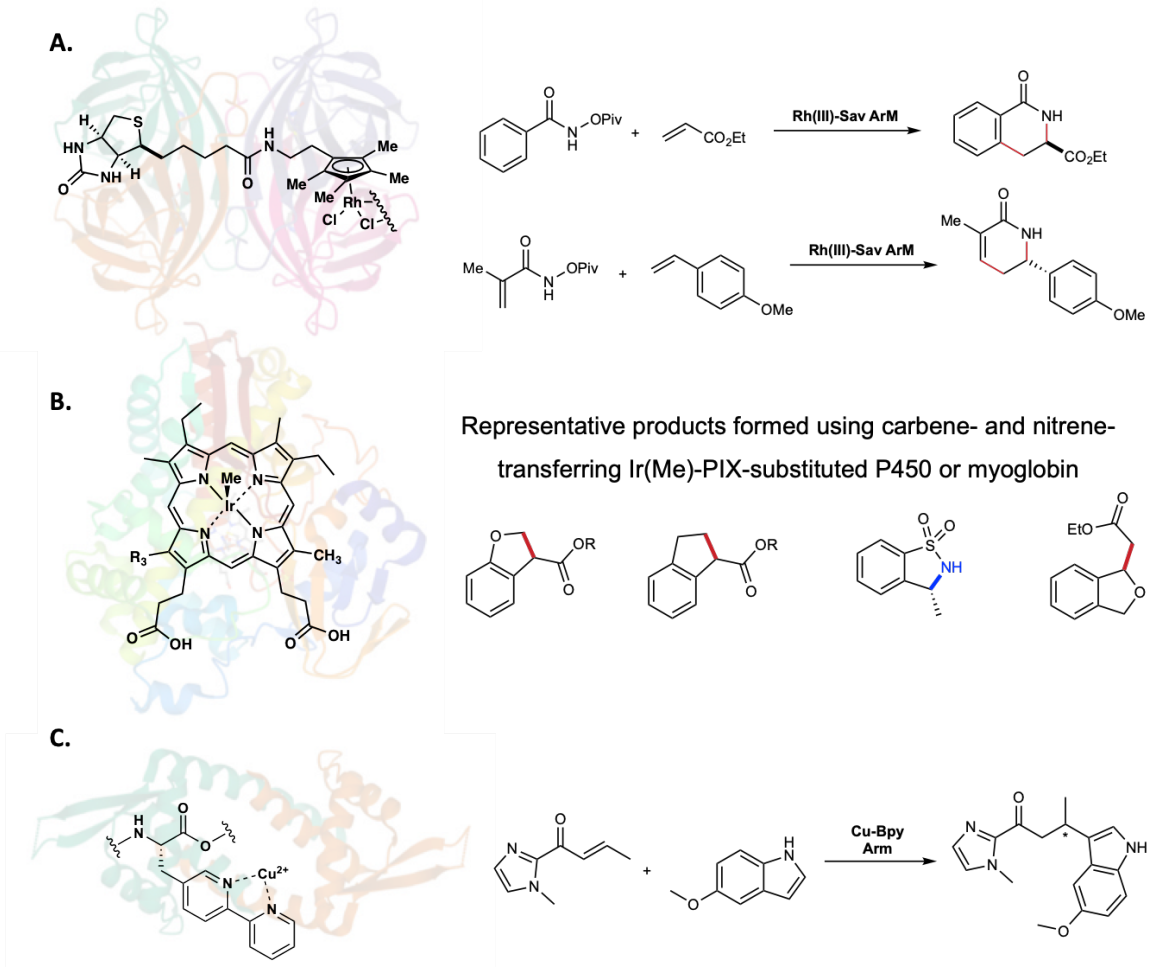


Figure 1-7. Representative ArMs which catalyze selective C–H functionalization reactions. **A)** Ward,[97] Rovis,[98] and colleagues adopted streptavidin scaffold for ArM-catalyzed C–H activation/ annulation reactions. Biotinylated Rh(III) complexes were used as the artificial cofactor for these transformations; **B)** Harwig[26,27] and Fasan[78] have replaced the native iron-heme cofactor in heme proteins with other metalloporphyrins such as Ir(Me)-PIX cofactor to catalyze carbene and nitrene C–H insertion reaction. Notably, the native heme proteins are capable of catalyzing similar reactions (see **Section 1.3**); **C)** Noncanonical amino acid can be introduced to proteins and create binding sites for metal or synthetic cofactor. As an example, Roefles and colleagues[101] have introduced a noncanonical amino acid with bipyridine side chain, which can bind to Cu^{2+} to enable ArM-catalyzed Friedel-Crafts reactions.

1.5 Conclusions and outlook

In the past, the large majority of synthetically useful C–H functionalization enzymes used to be hydroxylases. Today, the paradigm has significantly shifted due to the identification, repurposing, and design of a variety of enzymes that can catalyze diverse selective C–H functionalization reactions. I predict that many of these enzymes, such as aromatic and aliphatic C–H halogenases, who have great synthetic potential, can be further developed and implemented in practical synthesis. Some newer classes of C–H functionalization enzymes, for example laboratory-repurposed carbene- and nitrene-transferases as well as ArMs, still face certain practical limitations (e.g., the compatibility issues of the reagents used in these reactions). I believe that with more research efforts dedicated to solving these issues, many of these enzymes will also become catalysts which can surpass small-molecule catalysts developed for similar transformations.

Despite the significant progress that has been achieved over the past decade, the current enzymatic toolbox for C–H functionalization is still quite limited compared to homogeneous catalysis. Many synthetically useful reactions, such as C–H fluorination, cyanation, methylation, trifluoromethylation, are still largely underexplored. Exploration of these reaction types with biocatalysts will require continuous efforts in all of the three aforementioned aspects: 1) Genome mining of novel enzymatic transformation from nature is still an important approach to further expand the known enzyme toolbox; 2) To expand the diversity of transformations by repurposing native enzymes' non-natural activities, significantly increase research efforts is needed to expand this methodology to other classes of enzymes. However, many important and potentially useful cofactor-dependent enzymes faces practical protein engineering hinderances. For example, cobalamine-dependent enzymes cannot be heterologously expressed in common microorganism hosts. These limitations call for research efforts from multiple fields; 3) As for ArMs, during my literature review for this chapter, I could not find any current ArMs that have been adopted for industrial synthesis. A main reason is that current ArMs did not shown substantial advantages over related small-molecule catalysts or evolved biocatalysts. In addition, the preparation and engineering of ArMs can be complicated. If ArM development were to capitalize on

efforts made toward C–H functionalization reactions which are not likely to be afforded by native enzymes, their implementation could be accelerated. Other than increasing the diversity of enzymatic transformations, extensive efforts are also needed to turn known interesting enzymes into useful enzymes. Currently, exploratory genome-mining research is mainly driven by identifying unprecedented biosynthetic pathways and the enzymes involved therein. It could be worthwhile to use these tools to search for analogous enzymes to known important enzymes but more robust to adopt for protein engineering and chemical synthesis.

The utilization of biocatalysts has just begun to revolutionize chemical synthesis. It is reasonable to believe that with the advancements in protein engineering strategies such as machine learning, an even brighter future for using enzymes to edit molecules is yet to come.

1.6 References

1. Davies, H. M. L., Bois, J. D. & Yu, J.-Q. C–H Functionalization in organic synthesis. *Chem. Soc. Rev.* **40**, 1855–1856 (2011).

2. Börgel, J. & Ritter, T. Late-stage functionalization. *Chem* **6**, 1877–1887 (2020).

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

4. Fryszkowska, A. & Devine, P. N. Biocatalysis in drug discovery and development. *Curr. Opin. Chem. Biol.* **55**, 151–160 (2020).

5. Wu, S., Snajdrova, R., Moore, J. C., Baldenius, K. & Bornscheuer, U. T. Biocatalysis: Enzymatic synthesis for industrial applications. *Angew. Chem. Int. Ed.* **60**, 88–119 (2021).

6. Savile, C. K. *et al.* Biocatalytic Asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329**, 305–309 (2010).

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

8. Schober, M. *et al.* Chiral synthesis of LSD1 inhibitor GSK2879552 enabled by directed evolution of an imine reductase. *Nat. Catal.* **2**, 909–915 (2019).

9. Luo, X. *et al.* Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* **567**, 123–126 (2019).

10. McIntosh, J. A. *et al.* A kinase-cGAS cascade to synthesize a therapeutic STING activator. *Nature* **603**, 439–444 (2022).

11. Lewis, J. C., Coelho, P. S. & Arnold, F. H. Enzymatic functionalization of carbon–hydrogen bonds. *Chem. Soc. Rev* **40**, 2003–2021 (2011).

12. Cacho, R. A., Tang, Y. & Chooi, Y.-H. Next-generation sequencing approach for connecting secondary metabolites to biosynthetic gene clusters in fungi. *Front. Microbiol.* **5**, (2015).

13. Hillwig, M. L. & Liu, X. A new family of iron-dependent halogenases acts on freestanding substrates. *Nat. Chem. Biol.* **10**, 921–923 (2014).

14. Marchand, J. A. *et al.* Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature* **567**, 420–424 (2019).

15. Neugebauer, M. E. *et al.* A family of radical halogenases for the engineering of amino-acid-based products. *Nat. Chem. Biol.* **15**, 1009–1016 (2019).

16. Vaillancourt, F. H., Yeh, E., Vosburg, D. A., O'Connor, S. E. & Walsh, C. T. Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* **436**, 1191–1194 (2005).

17. Vaillancourt, F. H., Yin, J. & Walsh, C. T. SyrB2 in syringomycin E biosynthesis is a nonheme FeII α-ketoglutarate- and O2-dependent halogenase. *Proc. Natl. Acad. Sci.* **102**, 10111–10116 (2005).

18. Matthews, M. L. *et al.* Direct nitration and azidation of aliphatic carbons by an iron-dependent halogenase. *Nat. Chem. Biol.* **10**, 209–215 (2014).

19. Arnold, F. H. Directed evolution: Bringing new chemistry to life. *Angew. Chem. Int. Ed.* **57**, 4143–4148 (2017).

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

21. Prier, C. K. & Arnold, F. H. Chemomimetic biocatalysis: Exploiting the synthetic potential of cofactor-dependent enzymes to create new catalysts. *J. Am. Chem. Soc.* **137**, 13992–14006 (2015).

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

23. McIntosh, J. A. *et al.* Enantioselective intramolecular C–H amination catalyzed by engineered cytochrome P450 enzymes in vitro and in vivo. *Angew. Chem. Int. Ed.* **52**, 9309–9312 (2013).

24. Prier, C. K., Zhang, R. K., Buller, A. R., Brinkmann-Chen, S. & Arnold, F. H. Enantioselective, intermolecular benzylic C–H amination catalysed by an engineered iron-haem enzyme. *Nat. Chem.* **9**, 629–634 (2017).

25. Zhang, R. K. *et al.* Enzymatic assembly of carbon–carbon bonds via iron-catalysed sp^3 C–H functionalization. *Nature* **565**, 67–72 (2019).

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

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

28. Gu, Y., Natoli, S. N., Liu, Z., Clark, D. S. & Hartwig, J. F. Site-selective functionalization of (sp^3)C−H bonds catalyzed by artificial metalloenzymes containing an iridium-porphyrin cofactor. *Angew. Chem. Int. Ed.* **58**, 13954–13960 (2019).

29. Cirino, P. C. & Arnold, F. H. Protein engineering of oxygenases for biocatalysis. *Curr. Opin. Chem. Biol.* **6**, 130–135 (2002).

30. Dong, J. *et al.* Biocatalytic Oxidation reactions: A chemist's perspective. *Angew. Chem. Int. Ed.* **57**, 9238–9261 (2018).

31. Cernak, T., Dykstra, K. D., Tyagarajan, S., Vachal, P. & Krska, S. W. The medicinal chemist's toolbox for late stage functionalization of drug-like molecules. *Chem. Soc. Rev.* **45**, 546–576 (2016).

32. Christianson, D. W. Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* **106**, 3412–3442 (2006).

33. Bicas, J. L., Dionísio, A. P. & Pastore, G. M. Bio-oxidation of terpenes: An approach for the flavor industry. *Chem. Rev.* **109**, 4518–4531 (2009).

34. Zhang, X. *et al.* Divergent synthesis of complex diterpenes through a hybrid oxidative approach. *Science* **369**, 799–806 (2020).

35. Kille, S., Zilly, F. E., Acevedo, J. P. & Reetz, M. T. Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nat. Chem.* **3**, 738–743 (2011).

36. Zhang, K., Shafer, B. M., Demars, M. D., Stern, H. A. & Fasan, R. Controlled oxidation of remote sp^3 C–H bonds in artemisinin via P450 catalysts with fine-tuned regio- and stereoselectivity. *J. Am. Chem. Soc.* **134**, 18695–18704 (2012).

37. Rendic, S. & Guengerich, F. P. Survey of human oxidoreductases and cytochrome P450 enzymes involved in the metabolism of xenobiotic and natural chemicals. *Chem. Res. Toxicol.* **28**, 38–42 (2015).

38. Gumulya, Y. *et al.* Engineering highly functional thermostable proteins using ancestral sequence reconstruction. *Nat. Catal.* **1**, 878–888 (2018).

39. Marquez, L. A. & Dunford, H. B. Chlorination of taurine by myeloperoxidase. Kinetic evidence for an enzyme-bound intermediate. *J. Biol. Chem.* **269**, 7950–7956 (1994).

40. Carter-Franklin, J. N. & Butler, A vanadium bromoperoxidase-catalyzed biosynthesis of halogenated marine natural products. *J. Am. Chem. Soc.* **126**, 15060–15066 (2004).

41. Bernhardt, P., Okino, T., Winter, J. M., Miyanaga, A. & Moore, B. S. A stereoselective vanadium-dependent chloroperoxidase in bacterial antibiotic biosynthesis. *J. Am. Chem. Soc.* **133**, 4268–4270 (2011).

42. Diethelm, S., Teufel, R., Kaysser, L. & Moore, B. S. A multitasking vanadium-dependent chloroperoxidase as an inspiration for the chemical synthesis of the merochlorins. *Angew. Chem. Int. Ed.* **53**, 11023–11026 (2014).

43. Yeh, E., Blasiak, L. C., Koglin, A., Drennan, C. L. & Walsh, C. T. Chlorination by a long-lived intermediate in the mechanism of flavin-dependent halogenases. *Biochemistry* **46**, 1284–1292 (2007).

44. Andorfer, M. C. & Lewis, J. C. Understanding and improving the activity of flavin-dependent halogenases via random and targeted mutagenesis. *Annu. Rev. Biochem.* **87**, 159–185 (2018).

45. Lang, A. *et al.* Changing the regioselectivity of the tryptophan 7-halogenase PrnA by site-directed mutagenesis. *Angew. Chem. Int. Ed.* **50**, 2951–2953 (2011).

46. Glenn, W. S., Nims, E. & O'Connor, S. E. Reengineering a tryptophan halogenase to preferentially chlorinate a direct alkaloid precursor. *J. Am. Chem. Soc.* **133**, 19346–19349 (2011).

47. Shepherd, S. A. *et al.* Extending the biocatalytic scope of regiocomplementary flavin-dependent halogenase enzymes. *Chem. Sci.* **6**, 3454–3460 (2015).

48. Payne, J. T., Poor, C. B. & Lewis, J. C. Directed evolution of RebH for site-selective halogenation of large biologically active molecules. *Angew. Chem. Int. Ed.* **54**, 4226–4230 (2015).

49. Durak, L. J., Payne, J. T. & Lewis, J. C. Late-stage diversification of biologically active molecules via chemoenzymatic C–H functionalization. *ACS Catal.* **6**, 1451–1454 (2016).

50. Andorfer, M. C., Park, H. J., Vergara-Coll, J. & Lewis, J. C. Directed evolution of RebH for catalyst-controlled halogenation of indole C–H bonds. *Chem. Sci.* **7**, 3720–3729 (2016).

51. Andorfer, M. C. *et al.* Understanding flavin-dependent halogenase reactivity via substrate activity profiling. *ACS Catal.* **7**, 1897–1904 (2017).

52. Payne, J. T. *et al.* Enantioselective desymmetrization of methylenedianilines via enzyme-catalyzed remote halogenation. *J. Am. Chem. Soc.* **140**, 546–549 (2018).

53. Fisher, B. F., Snodgrass, H. M., Jones, K. A., Andorfer, M. C. & Lewis, J. C. Site-selective C–H halogenation using flavin-dependent halogenases identified via family-wide activity profiling. *ACS Cent. Sci.* **5**, 1844–1856 (2019).

54. Craven, E. J. *et al.* Programmable late-stage C−H bond functionalization enabled by integration of enzymes with chemocatalysis. *Nat. Catal.* **4**, 385–394 (2021).

55. Mori, S., Pang, A. H., Thamban Chandrika, N., Garneau-Tsodikova, S. & Tsodikov, O. V. Unusual substrate and halide versatility of phenolic halogenase PltM. *Nat. Commun.* **10**, 1255 (2019).

56. Gkotsi, D. S. *et al.* A marine viral halogenase that iodinates diverse substrates. *Nat. Chem.* **11**, 1091–1097 (2019).

57. Poor, C. B., Andorfer, M. C. & Lewis, J. C. Improving the stability and catalyst lifetime of the halogenase RebH by directed evolution. *ChemBioChem* **15**, 1286–1289 (2014).

58. Mitchell, A. J. *et al.* Structural basis for halogenation by iron- and 2-oxo-glutarate-dependent enzyme WelO5. *Nat. Chem. Biol.* **12**, 636–640 (2016).

59. Hayashi, T. *et al.* Evolved aliphatic halogenases enable regiocomplementary C−H functionalization of a pharmaceutically relevant compound. *Angew. Chem. Int. Ed.* **58**, 18535–18539 (2019).

60. Voss, M., Honda Malca, S. & Buller, R. Exploring the biocatalytic potential of Fe/α-ketoglutarate-dependent halogenases. *Chem. – Eur. J.* **26**, 7336–7345 (2020).

61. Büchler, J. *et al.* Algorithm-aided engineering of aliphatic halogenase WelO5* for the asymmetric late-stage functionalization of soraphens. *Nat. Commun.* **13**, 371 (2022).

62. Broderick, J. B., Duffus, B. R., Duschene, K. S. & Shepard, E. M. Radical *S*-adenosylmethionine enzymes. *Chem. Rev.* **114**, 4229–4317 (2014).

63. Benjdia, A., Balty, C. & Berteau, O. Radical SAM enzymes in the biosynthesis of ribosomally synthesized and post-translationally modified peptides (RiPPs). *Front. Chem.* **5**, (2017).

64. Cameron, D. C., Altaras, N. E., Hoffman, M. L. & Shaw, A. J. Metabolic engineering of propanediol pathways. *Biotechnol. Prog.* **14**, 116–125 (1998).

65. McKean, I. J. W. *et al. S*-adenosyl methionine cofactor modifications enhance the biocatalytic repertoire of small molecule *C*-alkylation. *Angew. Chem. Int. Ed.* **58**, 17583–17588 (2019).

66. Schmidt, N. G. *et al.* Biocatalytic Friedel–Crafts acylation and Fries reaction. *Angew. Chem. Int. Ed.* **56**, 7615–7619 (2017).

67. Nakamura, H., Schultz, E. E. & Balskus, E. P. A new strategy for aromatic ring alkylation in cylindrocyclophane biosynthesis. *Nat. Chem. Biol.* **13**, 916–921 (2017).

68. Schultz, E. E., Braffman, N. R., Luescher, M. U., Hager, H. H. & Balskus, E. P. Biocatalytic Friedel–Crafts alkylation using a promiscuous biosynthetic enzyme. *Angew. Chem. Int. Ed.* **58**, 3151–3155 (2019).

69. Che, C.-M., Lo, V. K.-Y., Zhou, C.-Y. & Huang, J.-S. Selective functionalisation of saturated C–H bonds with metalloporphyrin catalysts. *Chem. Soc. Rev.* **40**, 1950–1975 (2011).

70. Svastits, E. W., Dawson, J. H., Breslow, R. & Gellman, S. H. Functionalized nitrogen atom transfer catalyzed by cytochrome P-450. *J. Am. Chem. Soc.* **107**, 6427–6428 (1985).

71. Groves, J. T. & Van der Puy, M. Stereospecific aliphatic hydroxylation by an iron-based oxidant. *J. Am. Chem. Soc.* **96**, 5274–5275 (1974).

72. Huang, X. & Groves, J. T. Oxygen activation and radical transformations in heme proteins and metalloporphyrins. *Chem. Rev.* **118**, 2491–2553 (2018).

73. Poulos, T. L. Heme enzyme structure and function. *Chem. Rev.* **114**, 3919–3962 (2014).

74. Doyle, M. P., Duffy, R., Ratnikov, M. & Zhou, L. Catalytic carbene insertion into C−H bonds. *Chem. Rev.* **110**, 704–724 (2010).

75. Davies, H. M. L. & Manning, J. R. Catalytic C–H functionalization by metal carbenoid and nitrenoid insertion. *Nature* **451**, 417–424 (2008).

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

77. Singh, R., Bordeaux, M. & Fasan, R. P450-catalyzed intramolecular sp^3 C–H amination with arylsulfonyl azide substrates. *ACS Catal.* **4**, 546–552 (2014).

78. Bordeaux, M., Singh, R. & Fasan, R. Intramolecular C(sp^3)H amination of arylsulfonyl azides with engineered and artificial myoglobin-based catalysts. *Bioorg. Med. Chem.* **22**, 5697–5704 (2014).

79. Jia, Z.-J., Gao, S. & Arnold, F. H. Enzymatic primary amination of benzylic and allylic C(sp^3)–H bonds. *J. Am. Chem. Soc.* **142**, 10279–10283 (2020).

80. Vargas, D. A., Tinoco, A., Tyagi, V. & Fasan, R. Myoglobin-catalyzed C−H functionalization of unprotected indoles. *Angew. Chem. Int. Ed.* **57**, 9911–9915 (2018).

81. Brandenberg, O. F., Chen, K. & Arnold, F. H. Directed evolution of a cytochrome P450 carbene transferase for selective functionalization of cyclic compounds. *J. Am. Chem. Soc.* **141**, 8989–8995 (2019).

82. Brandenberg, O. F., Miller, D. C., Markel, U., Ould Chaib, A. & Arnold, F. H. Engineering chemoselectivity in hemoprotein-catalyzed indole amidation. *ACS Catal.* **9**, 8271–8275 (2019).

83. Hyster, T. K., Farwell, C. C., Buller, A. R., McIntosh, J. A. & Arnold, F. H. Enzyme-controlled nitrogen-atom transfer enables regiodivergent C–H amination. *J. Am. Chem. Soc.* **136**, 15505–15508 (2014).

84. Yang, Y., Cho, I., Qi, X., Liu, P. & Arnold, F. H. An enzymatic platform for the asymmetric amination of primary, secondary and tertiary C(sp^3)–H bonds. *Nat. Chem.* **11**, 987–993 (2019).

85. Singh, R., Kolev, J. N., Sutera, P. A. & Fasan, R. Enzymatic C(sp^3)-H amination: P450-catalyzed conversion of carbonazidates into oxazolidinones. *ACS Catal.* **5**, 1685–1691 (2015).

86. Zhang, J., Huang, X., Zhang, R. K. & Arnold, F. H. Enantiodivergent *α*-amino C–H fluoroalkylation catalyzed by engineered cytochrome P450s. *J. Am. Chem. Soc.* **141**, 9798–9802 (2019).

87. Zhou, A. Z., Chen, K. & Arnold, F. H. Enzymatic lactone-carbene C–H insertion to build contiguous chiral centers. *ACS Catal.* **10**, 5393–5398 (2020).

88. Goldberg, N. W., Knight, A. M., Zhang, R. K. & Arnold, F. H. Nitrene transfer Catalyzed by a non-heme iron enzyme and enhanced by non-native small-molecule ligands. *J. Am. Chem. Soc.* **141**, 19585–19588 (2019).

89. Vila, M. A., Steck, V., Rodriguez Giordano, S., Carrera, I. & Fasan, R. C−H amination via nitrene transfer catalyzed by mononuclear non-heme iron-dependent enzymes. *ChemBioChem* **21**, 1981–1987 (2020).

90. Liu, Z. *et al.* An Enzymatic platform for primary amination of 1-aryl-2-alkyl alkynes. *J. Am. Chem. Soc.* **144**, 80–85 (2022).

91. Athavale, S. V. *et al.* Biocatalytic, Intermolecular C−H bond functionalization for the synthesis of enantioenriched amides. *Angew. Chem. Int. Ed.* **60**, 24864–24869 (2021).

92. Waldman, A. J. & Balskus, E. P. Discovery of a diazo-forming enzyme in cremeomycin biosynthesis. *J. Org. Chem.* **83**, 7539–7546 (2018).

93. Schwizer, F. *et al.* Artificial metalloenzymes: Reaction scope and optimization strategies. *Chem. Rev.* **118**, 142–231 (2018).

94. Abe, S. *et al.* Control of the coordination structure of organometallic palladium complexes in an apo-ferritin cage. *J. Am. Chem. Soc.* **130**, 10512–10514 (2008).

95. Filice, M. *et al.* Preparation of an immobilized lipase-palladium artificial metalloenzyme as catalyst in the Heck reaction: Role of the solid phase. *Adv. Synth. Catal.* **357**, 2687–2696 (2015).

96. Wilson, M. E. & Whitesides, G. M. Conversion of a protein to a homogeneous asymmetric hydrogenation catalyst by site-specific modification with a diphosphinerhodium(I) moiety. *J. Am. Chem. Soc.* **100**, 306–307 (1978).

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

98. Hassan, I. S. *et al.* Asymmetric *δ*-lactam synthesis with a monomeric streptavidin artificial metalloenzyme. *J. Am. Chem. Soc.* **141**, 4815–4819 (2019).

99. Pierron, J. *et al.* Artificial metalloenzymes for asymmetric allylic alkylation on the basis of the biotin–avidin technology. *Angew. Chem. Int. Ed.* **47**, 701–705 (2008).

100. Dydio, P., Key, H. M., Hayashi, H., Clark, D. S. & Hartwig, J. F. Chemoselective, enzymatic C–H bond amination catalyzed by a cytochrome P450 containing an Ir(Me)-PIX cofactor. *J. Am. Chem. Soc.* **139**, 1750–1753 (2017).

101. Drienovská, I., Rioz-Martínez, A., Draksharapu, A. & Roelfes, G. Novel artificial metalloenzymes by *in vivo* incorporation of metal-binding unnatural amino acids. *Chem. Sci.* **6**, 770–776 (2014).

Chapter 2

ENANTIODIVERGENT α-AMINO C–H FLUOROALKYLATION CATALYZED BY ENGINEERED CYTOCHROMES P450

Material for this chapter appears in: **Zhang, J.**;† Huang, X.;† Zhang, R. K.; Arnold, F. H. Enantiodivergent α-amino C–H fluoroalkylation catalyzed by engineered cytochrome P450s. *J. Am. Chem. Soc.*, **2019**, 141, 25, 9798-9802. DOI: 10.1021/jacs.9b04344. (†: J. Z. and X. H. contributed equally)

J.Z. and X.H. designed the overall research with F.H.A. providing guidance. **J.Z.** and X. H. designed and conducted the initial screening of haem proteins; **J.Z.** performed the directed evolution experiments. **J.Z.**, X.H., and R.K.Z. designed and performed the substrate scope studies. **J.Z.**, X.H., and F.H.A. wrote the manuscript with input from R.K.Z.

Abstract

The introduction of fluoroalkyl groups into organic compounds can significantly alter pharmacological characteristics. One enabling but underexplored approach for the installation of fluoroalkyl groups is selective C(sp^3)−H functionalization due to the ubiquity of C−H bonds in organic molecules. We have engineered heme enzymes that can insert fluoroalkyl carbene intermediates into α-amino C(sp^3)−H bonds and enable enantiodivergent synthesis of fluoroalkyl-containing molecules. Using directed evolution, we engineered cytochrome P450 enzymes to catalyze this abiological reaction under mild conditions with total turnovers (TTN) up to 4070 and enantiomeric excess (ee) up to 99%. The iron-heme catalyst is fully genetically encoded and configurable by directed evolution so that just a few mutations to the enzyme completely inverted product enantioselectivity. These catalysts provide a powerful method for synthesis of chiral organofluorine molecules that is currently not possible with small-molecule catalysts.

2.1 Introduction

Fluoroalkyl groups are important bioisosteres in medicinal chemistry that can enhance the metabolic stability, lipophilicity, and bioavailability of drug molecules.[1–4] Conversion of C–H bonds into carbon-fluoroalkyl bonds represents one of the most appealing strategies for fluoroalkyl group incorporation.[5–9] Such methods are of high atom economy and provide efficient ways to obtain new organofluorine molecules via late-stage functionalization of complex bioactive molecules.[10,11] Despite the synthetic appeal of this strategy, however, enantioselective C(sp^3)–H fluoroalkylation reactions are noticeably lacking. Major obstacles to development of transition-metal catalyzed C–H fluoroalkylation reactions include the inherent challenges associated with carbon-fluoroalkyl bond cross-coupling pathways, such as slow oxidative addition of fluoroalkyl nucleophiles[12,13] and facile fluoride elimination of organometallic species.[13–16]

A strategy involving insertion of fluoroalkyl carbene intermediates into C(sp^3)–H bonds could potentially circumvent these challenges. Although metal fluoroalkylcarbene intermediates have been utilized for a number of carbene transfer reactions,[15,17–23] their applications for C–H functionalization have rarely been explored.[24] Transition metal catalysts, including those based on rhodium,[25–29] iridium,[30,31] copper,[32] iron,[33–36] and other metals,[37–41] have been shown to catalyze carbene insertion into C(sp^3)–H bonds. Intermolecular stereoselective reactions, however, are typically constrained to dirhodium-based catalysts with carbene precursors bearing both electron-donating and electron-withdrawing substituents at the carbene carbon (referred to as donor-acceptor carbene reagents).[25,37] The electron-donating group is required to attenuate the high reactivity of dirhodium-carbene intermediates and offers better stereocontrol of the C–H functionalization/C–C bond forming step.[42] Catalysts that can use acceptor-only-type perfluorodiazoalkanes as carbene precursors for direct, enantioselective C(sp^3)–H fluoroalkylation have not been reported.

Our group recently disclosed iron-heme enzymes derived from cytochromes P450 that catalyze abiological carbene C–H bond insertion reactions using several acceptor-only

diazo compounds.[43] Building on this effort, we now show that engineered cytochrome P450 enzymes can adopt C–H fluoroalkylation activity with high efficiency and enantioselectivity, achieving direct C–H fluoroalkylation of substrates that contain *α*-amino C–H bonds. Given the high prevalence of amines in pharmaceuticals, this simple biocatalytic method provides an efficient route to molecular diversification through selective C–H functionalization.

2.2 Results and Discussion

2.2.1 Identification of heme proteins with promiscuous fluoroalkylation activity

To identify a suitable starting point for directed evolution of a C–H fluoroalkylation enzyme, we first challenged a panel of 14 heme proteins in clarified *Escherichia coli* lysate with *N*-phenyl pyrrolidine (**1a**) and 2,2,2-trifluoro-1-diazoethane (**2**) as model substrates under anaerobic conditions (**Table A-1**). Several proteins, including *Rhodothermus marinus* cyt *c* (*Rma* cyt *c*), engineered *Rma* NOD, and wild-type $P450_{BM3}$ from *Bacillus megaterium*, exhibited trace catalytic activities. Reactions with only the heme cofactor (iron protoporphyrin IX) as the catalyst also delivered trace amounts of product **3a**. Several serine-ligated cytochromes P450 (P411s),[44] however, exhibited promising initial activity for the target trifluoroethylation reaction. The highest activity (1,250 TTN) was obtained with P411-CH-C8. This P411ΔFAD variant, which comprises the heme and FMN but not the FAD domain of $P450_{BM3}$, was originally engineered for carbene C–H insertion with ethyl diazoacetate (EDA).[43]

2.2.2 Directed evolution of an *R*-selective fluoroalkylase

Although P411-CH-C8 exhibited high activity for the fluoroalkylation reaction between **1a** and **2**, the enantioselectivity of the resulting product **3a** was poor (12% ee, (*S*)-enantiomer, **Table A-1**). We therefore used directed evolution to increase enantioselectivity (**Figure 2-1**). We first targeted several amino acid residues in the distal heme pocket for site-saturation mutagenesis and screened for variants with improved enantioselectivity. Many of the sites selected for mutagenesis were previously shown to

affect activity and selectivity in abiological carbene- and nitrene-transfer reactions (**Table A-2**). Although none of the mutants tested showed improvement in forming (*S*)-**3a**, we discovered that a T327V mutation inverted enantioselectivity and yielded (*R*)-**3a** with 28% ee. With the T327V mutant (P411-FA-B3) as the new parent, further rounds of site-saturation mutagenesis and recombination of beneficial mutations yielded variant FA-E3 with five mutations (T327V, E70T, L177M, R226T, and Y330V) compared to P411-CH-C8 (**Figure A-2**). This variant exhibited 88% ee for (*R*)-**3a**.

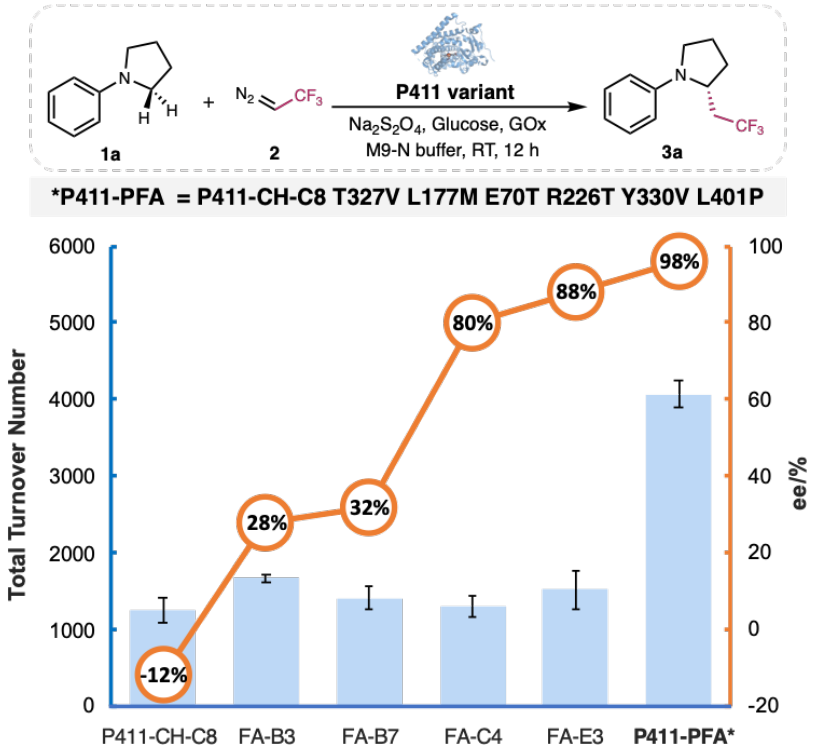


Figure 2-1. Directed evolution of P411 catalysts for C(sp^3)–H fluoroalkylation reaction. Experiments were per-formed using clarified *E. coli* lysate overexpressing the P411 variants, 10 mM **1a**, 20 mM **2**, 25 mM D-glucose, 5 mg/mL sodium dithionite, GOx and 5 vol% EtOH in M9-N buffer at room temperature under anaerobic condition for 12 h.

We next surveyed residues in the enzyme's proximal loop; residues in this region play an important role in regulating the oxidation activity of cytochromes P450.[45–49] Our lab and others have shown that mutations in this region also affect abiological carbene and

nitrene transfer reactivities, mainly by tuning the electron-donating properties of the heme proximal axial ligand.[44,50] With FA-E3 as the parent, site-saturation mutagenesis on proximal loop residues and screening revealed the L401P mutation that further improved activity to 4,070 TTN and enantioselectivity to 98% ee (**Figure 2-1**). This final variant, named P411-PFA, contains six mutations from P411-CH-C8 (T327V, E70T, L177M, R226T, Y330V, and L401P).

2.2.3 Substrate scope studies of P411-PFA

With this laboratory-evolved C–H fluoroalkylation enzyme in hand, we then explored its performance on a diverse set of substrates. As shown in **Figure 2-2**, P411-PFA could install a trifluoroethyl group onto various *N*-aryl pyrrolidine substrates by directly activating the *α*-amino C–H bonds. High activity and enantioselectivity were achieved for pyrrolidines containing a variety of *N*-aryl and *N*-heteroaryl substituents. A range of functional groups including methoxy, halogen, ketone, and aldehyde were well tolerated. The tolerance to the *p*-methoxylphenyl group (PMP) would enable the facile synthesis of other *N*-substituted pyrrolidines bearing a trifluoroethyl stereogenic center, as PMP is a well-established protecting group for the nitro-gen atom and can be removed under mild conditions.[51] Furthermore, given the compatibility of our method with reactive functional groups like halogens and aldehydes, the application could be broadened further by harnessing these functionalities as reaction handles to access a diverse range of structural motifs through well-established cross-coupling and condensation reactions. This enzymatic approach opens possibilities to access a broad range of chiral trifluoroethylated pyrrolidines, whose current construction methods require stepwise, successive radical cross coupling chemistry that is time-consuming and not enantioselective.[52]

In addition to pyrrolidine-type substrates, this enzymatic method could also functionalize *N*,*N*-dialkyl anilines, which is another structural motif prevalent in pharmaceuticals. The enzyme is highly selective toward *α*-amino C–H bonds. For instance, in compound **3k**, the *N*-methyl is activated exclusively in the presence of weaker benzylic C–H bonds. The preference for C–H bonds at α-amino positions over those at the benzylic

and OMe (**3b**) positions might arise from the strong electron-donating properties of nitrogen, which makes α-amino C–H bonds react more favorably with the electrophilic iron-carbene intermediates.[25,53] To demonstrate synthetic utility further, we performed this enzymatic reaction on preparative scale, where it proceeded smoothly and afforded the chiral trifluoroethylated compound **3a** with 64% isolated yield and 98% ee (116 mg). We obtained the crystal structure of compound **3a**, and the absolute configuration of the trifluoroethylated chiral center was determined to be *R*.

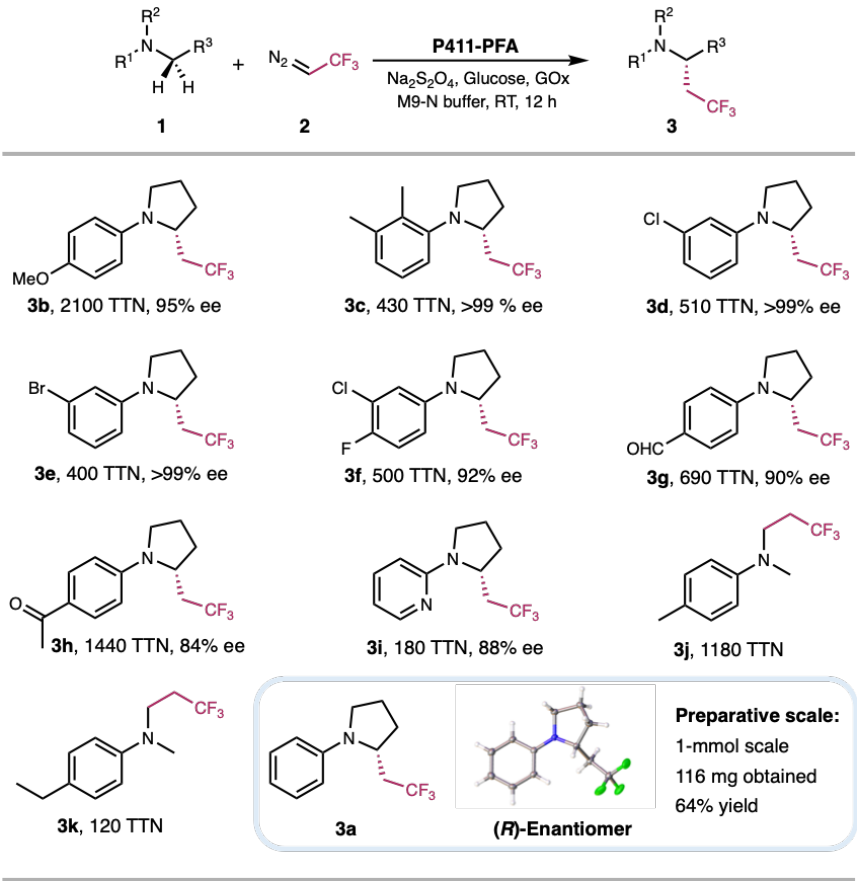


Figure 2-2. Substrate scope of P411-PFA-catalyzed C–H trifluoroethylation reaction. Absolute configuration of **3a** was determined by X-ray crystallography.

2.2.4 Enantiodivergent fluoroalkylation enabled by P411-PFA and P411-PFA-(*S*)

Alternate stereoisomers of a bioactive molecule can have drastically different biological effects and need to be evaluated individually during drug candidate screening.[54] This necessitates the synthesis of all possible stereoisomers of a given molecule, preferably via stereo-divergent asymmetric catalysis.[55,56] Thus, we developed an enzyme catalyst that could perform the targeted C–H trifluoroethylation with enantioselectivity opposite to that of P411-PFA. To find a suitable starting point for evolving an enzyme that exhibits reversed enantioselectivity, we first evaluated the catalytic performance on various substrates of all the variants along the evolution of P411-PFA. Early variant FA-B7 exhibited moderate reversed enantioselectivity (24% ee for the (*S*)-enantiomer) for functionalization of aldehyde-substituted *N*-aryl pyrrolidine substrate **1g**. Further examination of variants derived from FA-B7 led to the discovery of a quadruple mutant of FA-B7 (T70E, V327T, G74P, Q437L, termed P411-PFA-(*S*)) that catalyzes the formation

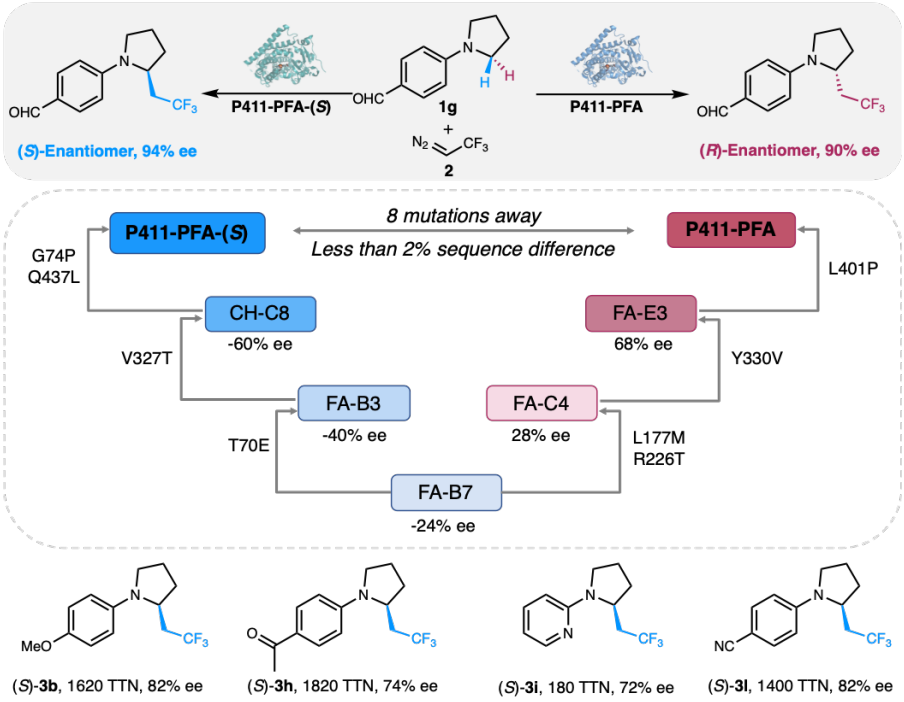


Figure 2-3. Enantiodivergent C(sp^3)–H trifluoroethylation catalyzed by P411-PFA and P411-PFA-(*S*) and substrate scope of P411-PFA-(*S*).

of (*S*)-**3g** with 94% ee (**Figure 2-3**, **Figure A-4**). P411-PFA-(*S*) is a general catalyst for synthesis of the (*S*)-enantiomer of trifluoroethylated pyrrolidines, as demonstrated by its high activity and moderate-to-high (*S*)-enantioselectivity toward a variety of *N*-aryl and *N*-heteroaryl pyrrolidine substrates (**Figure 2-3**). These results further highlight the facile configurability of the enzymatic system for delivering diverse chiral organofluorine molecules.

2.2.5 Enantioselective pentafluoropropylation catalyzed by P411-PFA

Another advantage of this chemistry is its ability to install other fluoroalkyl groups via the same C–H functionalization process. As a proof of concept, we challenged the protein catalysts with 2,2,3,3,3-pentafluoro-1-diazopropane **4** as the carbene precursor. As shown in **Figure 2-4**, P411-PFA can use **4** to introduce pentafluoropropyl groups into the α-amino C–H bonds of both acyclic and cyclic amine substrates with excellent activity and

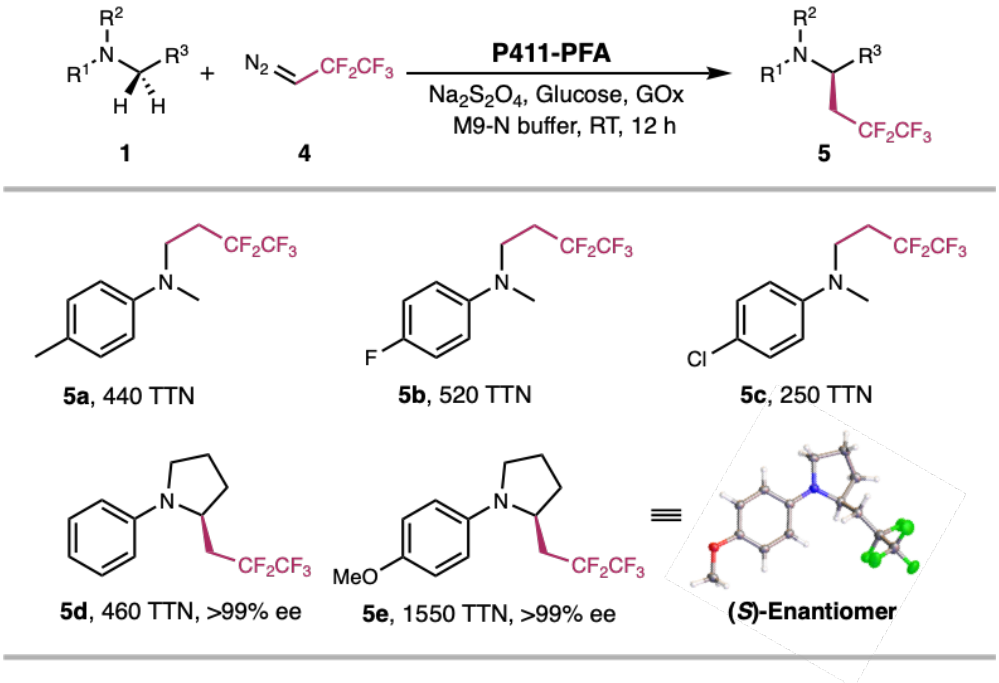


Figure 2-4. Substrate scope of P411-PFA-catalyzed C–H pentafluoropropylation. Absolute configuration of **5e** was determined by X-ray crystallography.

enantioselectivity. We successfully obtained 59 mg of the enzymatic product **5e**. Intriguingly, subsequent X-ray crystallographic analysis showed that the pentafluoropropylation products obtained by P411-PFA exhibited an opposite absolute configuration to that of the trifluoroethylation ones. Although further investigation is needed to fully elucidate the origin of this inversion of absolute configuration, a potential cause is a conformational change of the corresponding fluoroalkylated heme-carbene intermediates, which alters the orientation of the fluoroalkyl groups and reverts the configuration of the prochiral face accessed by the substrates for C–H bond activation. This hypothesis is supported by the fact that carbene intermediates in heme proteins can adopt different conformations depending on their structural properties.[22,23]

2.3 Conclusion

In summary, we have developed a catalytic platform for insertion of fluoroalkyl-substituted carbenes into C(sp^3)–H bonds with high activity and enantioselectivity under mild conditions. With directed evolution, the enantioselectivity of the enzymes can be tuned to achieve enantiodivergent synthesis of organofluorine compounds by this versatile carbene C–H insertion process. This work provides a powerful new approach for addition of fluorine-containing structural motifs prevalent in pharmaceuticals and further expands the reaction scope of new-to-nature enzymatic C–H alkylation. We envision that the enzymes developed in this research will open up new avenues for synthesis of fluorinated bioactive molecules.

2.4 Experimental Methods

2.4.1 General

Unless otherwise noted, all chemicals and reagents were obtained from commercial suppliers (Sigma-Aldrich, VWR, Alfa Aesar, Acros, etc.) and used without further purification. Silica gel chromatography was carried out using AMD Silica Gel 60, 230-400 mesh. ^{1}H and ^{13}C NMR spectra were recorded on a Bruker Prodigy 400 MHz instrument (400 MHz for ^{1}H and 101 MHz for ^{13}C NMR) or a Varian 300 MHz Spectrometer (300

MHz for ^{1}H NMR). Chemical shifts (δ) are reported in ppm downfield from tetramethylsilane, using the solvent resonance as the internal standard (^{1}H NMR: δ = 7.26, ^{13}C NMR: δ = 77.36 for $CDCl_3$). ^{19}F NMR data were collected on a Varian 300 MHz spectrometer (282 MHz for ^{19}F NMR). Data for ^{1}H NMR are reported as follows: chemical shift (δ ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, p = pentet, sext = sextet, m = multiplet, dd = doublet of doublets, dt = doublet of triplets, ddd = doublet of doublet of doublets), coupling constant (Hz), integration. Sonication was performed using a Qsonica Q500 sonicator. High-resolution mass spectra were obtained at the California Institute of Technology Mass Spectral Facility.

E. coli cells were grown using Luria-Bertani medium or HyperBroth (AthenaES) with 100 μg/mL ampicillin (LB_{amp} or HB_{amp}). Primer sequences are available upon request. T5 exonuclease, Phusion polymerase, and *Taq* ligase were purchased from New England Biolabs (NEB, Ipswich, MA). M9-N minimal medium (abbreviated as M9-N buffer; pH 7.4) was used as a buffering system for whole cells, lysates, and purified proteins, unless otherwise specified. M9-N buffer was used without a carbon source; it contains 47.7 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 8.6 mM NaCl, 2.0 mM $MgSO_4$, and 0.1 mM $CaCl_2$.

2.4.2 Chromatography

Gas chromatography-mass spectrometry (GC-MS) analyses were carried out using Shimadzu GCMS-QP2010SE system and J&W HP-5ms column. Analytical chiral HPLC was conducted using either an Agilent 1200 series instrument with *n*-hexane and isopropanol as the mobile phase. Enantiomers were separated using one of the following chiral columns: Chiralpak IA (4.6 mm × 25 cm), Chiralpak IC (4.6 mm × 25 cm), Chiralpak AD (4.6 mm × 25 cm), Chiralcel OJ-H (4.6 mm × 25 cm). Chiral GC was performed on an Agilent 6850 GC with FID detector using a Astec-CHIRALDEX G-TA column (30.0 m × 0.25 mm) at 1.0 mL/min He carrier gas flow.

2.4.3 Cloning and site-saturation mutagenesis

The genes encoding all enzymes described in this study were cloned using Gibson assembly[57] into vector pET22b(+) (Novagen) between restriction sites *Nde*I and *Xho*I in frame with a *C*-terminal 6xHis-tag. Site-saturation mutagenesis was performed using the "22c-trick" method.[58] The PCR products were digested with *Dpn*I, gel purified, and ligated using Gibson Mix™.[57] The ligation mixture was used to directly transform electrocompetent *E. coli* strain *E. cloni* BL21 (DE3) cells (Lucigen).

2.4.4 Expression of P450 and P411 variants in 96-well plates

Single colonies from LB_{amp} agar plates were picked using sterile toothpicks and cultured in deep-well 96-well plates containing LB_{amp} (400 μL/well) at 37 °C, 80% humidity and 250 rpm shaking overnight. Subsequently, HB_{amp} (1080 μL/well) in a deep-well plate was inoculated with an aliquot (120 μL/well) of these overnight cultures and allowed to shake for 3 hours at 37 °C, 80% humidity and 250 rpm. The plates were then cooled on ice for 30 minutes and the cultures were induced with 0.5 mM isopropyl *β*-D-1-thiogalactopyranoside (IPTG) and 1.0 mM 5-aminolevulinic acid (ALA) (final concentrations). Expression were then conducted at 20 °C, 230 rpm for 18–20 hours.

2.4.5 Plate reaction screening in whole-cell format

E. coli cells harboring P411 variants in deep-well 96-well plates were pelleted (3,500 × g, 5 min, 4 °C) and resuspended in M9-N buffer (375 μL) by gentle vortexing. The 96-well plates were then transferred to an anaerobic chamber. To deep-well plates of cell suspensions were added the *N*-phenylpyrrolidine substrate **1a** (10 μL/well, 400 mM in EtOH) and 2,2,2-trifluoro-1-diazoethane **2** (15 μL per well, 600 mM in EtOH). During the addition, the stock solution of 2,2,2-trifluoro-1-diazoethane **2** is kept on an ice-salt bath (-20 °C) to minimize evaporation. The plates were sealed with aluminum sealing foil immediately after the addition and shaken in the anaerobic chamber at room temperature and 700 rpm. After 18–24 hours, the seals were removed and 610 μL 1:1 ethylacetate/hexanes solution containing 0.66 mM 1,2,3-trimethoxybenzene internal

standard was added. The plates were tightly sealed with silicone mats, vigorously vortexed, and centrifuged (5,000 × g, 5 min) to completely separate the organic and aqueous layers. The organic layer of each well was then transferred to individual vials equipped with autosampler vial inserts and analyzed by GC-MS. To prepare samples for enantioselectivity determination via normal-phase chiral HPLC, the organic solution in the vial was removed under reduced pressure and the resulting crude product was re-dissolved in hexanes.

2.4.6 Expression of P411 variants and cell lysate preparation

E. coli (*E. cloni* BL21(DE3)) cells carrying plasmid encoding the appropriate P411 variant were grown overnight in 5 mL LB_{amp}. Preculture (3 mL) was used to inoculate 27 mL of HB_{amp} in a 125-mL Erlenmeyer flask; this culture was incubated at 37 °C, 230 rpm for 2.5 hours. The culture was then cooled on ice for 30 min and induced with 0.5 mM IPTG and 1.0 mM ALA (final concentrations). Expression was conducted at 20 °C, 150 rpm for 16–18 hours. Subsequently, the *E. coli* cells were pelleted by centrifugation (3,000 × g, 5 min, 4 °C). Media was removed and the resulting cell pellet was resuspended in M9-N buffer to OD_{600} = 30. The cell suspension was then lysed by sonication using a Qsonica Q500 sonicator equipped with a microtip (2 minutes, 1 second on, 1 second off, 30% amplitude); samples were kept on wet ice for this process. The resulting lysed solution was centrifuged (20,000 × g, 10 min, 4 °C) to remove cell debris. The supernatant (clarified lysate) was separated from the pellet and kept on ice until use.

2.4.7 Hemochrome assay for the determination of heme protein concentration.

The concentration of heme protein in the clarified lysate was determined by the hemochrome assay.[59] Briefly, 500 µL of the lysate was added to a cuvette and mixed with 500 µL of solution I (0.2 M NaOH, 40% (v/v) pyridine, 500 µM potassium ferricyanide). The UV-Vis spectrum (380-620 nm) of the oxidized Fe^{III} state was recorded immediately. Sodium dithionite (10 µL of 0.5 M solution in 0.5 M NaOH) was added and the UV-Vis spectrum of the reduced Fe^{II} state was recorded immediately. The pyridine

hemochromagen concentration was determined using its Q bands, with the following extinction coefficients: P450s and globins: 34.7 mM^{-1} cm^{-1} at 557 nm; cytochromes *c*: 30.27 mM^{-1} cm^{-1} at 550 nm.

2.4.8 Biotransformation using clarified *E. coli* lysate.

The lysate was placed in a sealed vial on ice and the headspace of the vial was purged with a stream of argon for at least 30 minutes. Enzymatic reactions were then set up in anaerobic chamber. To a 2-mL vial were added lysate (315 μL), a GOx oxygen depletion solution (20 μL of stock solution containing 14,000 U/mL catalase and 1,000 U/mL glucose oxidase in M9-N buffer), D-glucose (20 μL of 250 mM stock solution in M9-N buffer), $Na_2S_2O_4$ (20 μL of 20 mM solution in M9-N), tertiary amine substrate (10 μL of 400 mM stock solution in EtOH) in the listed order. The perfluorodiazoalkane solution (15 μL of 600 mM stock solution in EtOH) kept in an ice-salt bath (-20 °C) was added at last. The reaction vials were then capped and shaken in the anaerobic chamber at room temperature and 700 rpm for 12 hours. After the completion of the reaction, 610 μL 1:1 ethylacetate/hexanes solution containing 0.66 mM 1,2,3-trimethoxybenzene internal standard was added to the vial. The resulting mixture was transferred to a 1.5-mL microcentrifuge tube, vigorously vortexed, and centrifuged (20,000 × g, 5 minutes) to completely separate the organic and aqueous layers. The organic layer was transferred to a vial equipped with an autosampler vial insert and analyzed by GC-MS. To prepare samples for enantioselectivity determination via normal-phase chiral HPLC, the organic solution in the vial was removed under reduced pressure and the resulting crude product was redissolved in hexanes.

2.4.9 Enzymatic preparative synthesis

Clarified lysate of *E. coli* (78.75 mL) harboring the P411-PFA variant was prepared as described in **Section 2.4.6**. The lysate was placed in a sealed vial on ice and the headspace was purged with a stream of argon for 30 minutes. To the reaction vial were added a GOx oxygen depletion solution (5 mL of stock solution containing 14,000 U/mL

catalase and 1,000 U/mL glucose oxidase in M9-N buffer), D-glucose (5 mL of 250 mM stock solution in M9-N buffer), $Na_2S_2O_4$ (5 mL of 20 mM solution in M9-N), and substrate **1** (2.5 mL of 400 mM stock solution in EtOH). 2,2,2-trifluoro-1-diazoethane **2** (3.75 mL of 600 mM stock solution in EtOH) kept in an ice-salt bath was added at last (-20 °C). The reaction vial was immediately capped and sealed with parafilm, removed from anaerobic chamber, and shaken at room temperature at 250 rpm for 12 hours. The reaction solution was then extracted with 100 mL 1:1 hexane/ethylacetate for three times. The combined organic layer was then washed with brine, dried over anhydrous $MgSO_4$, concentrated, and purified by flash chromatography.

2.5 References

1. Meanwell, N. A. Fluorine and fluorinated motifs in the design and application of bioisosteres for drug design. *J. Med. Chem.* **61**, 5822–5880 (2018).

2. Mueller, K., Faeh, C. & Diederich, F. Fluorine in pharmaceuticals: Looking beyond intuition. *Science* **317**, 1881–1886 (2007).

3. Zhou, Y. *et al.* Next generation of fluorine-containing pharmaceuticals, compounds currently in phase II–III clinical trials of major pharmaceutical companies: New structural trends and therapeutic areas. *Chem. Rev.* **116**, 422–518 (2016).

4. Purser, S., Moore, P. R., Swallow, S. & Gouverneur, V. Fluorine in medicinal chemistry. *Chem. Soc. Rev.* **37**, 320–330 (2008).

5. Fujiwara, Y. *et al.* Practical and innate carbon–hydrogen functionalization of heterocycles. *Nature* **492**, 95–99 (2012).

6. Liang, T., Neumann, C. N. & Ritter, T. Introduction of fluorine and fluorine-containing functional groups. *Angew. Chem. Int. Ed.* **52**, 8214–8264 (2013).

7. Ma, J.-A. & Cahard, D. Asymmetric fluorination, trifluoromethylation, and perfluoroalkylation reactions. *Chem. Rev.* **104**, 6119–6146 (2004).

8. McAtee, R. C., Beatty, J. W., McAtee, C. C. & Stephenson, C. R. J. Radical chlorodifluoromethylation: Providing a motif for (hetero)arene diversification. *Org. Lett.* **20**, 3491–3495 (2018).

9. Beatty, J. W. *et al.* Photochemical perfluoroalkylation with pyridine *N*-oxides: Mechanistic insights and performance on a kilogram scale. *Chem* **1**, 456–472 (2016).

10. Yamaguchi, J., Yamaguchi, A. D. & Itami, K. C–H bond functionalization: Emerging synthetic tools for natural products and pharmaceuticals. *Angew. Chem. Int. Ed.* **51**, 8960–9009 (2012).

11. Hartwig, J. F. Evolution of C–H bond functionalization from methane to methodology. *J. Am. Chem. Soc.* **138**, 2–24 (2016).

12. Culkin, D. A. & Hartwig, J. F. Carbon−carbon bond-forming reductive elimination from arylpalladium complexes containing functionalized alkyl groups. Influence of ligand steric and electronic properties on structure, stability, and reactivity. *Organometallics* **23**, 3398–3416 (2004).

13. Zhao, Y. & Hu, J. Palladium-catalyzed 2,2,2-trifluoroethylation of organoboronic acids and esters. *Angew. Chem. Int. Ed.* **51**, 1033–1036 (2012).

14. Uneyama, K., Katagiri, T. & Amii, H. α-Trifluoromethylated carbanion synthons. *Acc. Chem. Res.* **41**, 817–829 (2008).

15. Argintaru, O. A., Ryu, D., Aron, I. & Molander, G. A. Synthesis and applications of α-trifluoromethylated alkylboron compounds. *Angew. Chem. Int. Ed.* **52**, 13656–13660 (2013).

16. Ferguson, D. M., Bour, J. R., Canty, A. J., Kampf, J. W. & Sanford, M. S. Stoichiometric and catalytic aryl–perfluoroalkyl coupling at tri-*tert*-butylphosphine palladium(II) complexes. *J. Am. Chem. Soc.* **139**, 11662–11665 (2017).

17. Morandi, B. & Carreira, E. M. Iron-catalyzed cyclopropanation with trifluoroethylamine hydrochloride and olefins in aqueous media: In situ generation of trifluoromethyl diazomethane. *Angew. Chem. Int. Ed.* **49**, 938–941 (2010).

18. Morandi, B. & Carreira, E. M. Synthesis of trifluoroethyl-substituted ketones from aldehydes and cyclohexanones. *Angew. Chem. Int. Ed.* **50**, 9085–9088 (2011).

19. Hyde, S. *et al.* Copper-catalyzed insertion into heteroatom–hydrogen bonds with trifluorodiazoalkanes. *Angew. Chem. Int. Ed.* **55**, 3785–3789 (2016).

20. Luo, H., Wu, G., Zhang, Y. & Wang, J. Silver(I)-catalyzed *N*-trifluoroethylation of anilines and *O*-trifluoroethylation of amides with 2,2,2-trifluorodiazoethane. *Angew. Chem. Int. Ed.* **54**, 14503–14507 (2015).

21. Liu, C.-B. *et al.* A facile parallel synthesis of trifluoroethyl-substituted alkynes. *Angew. Chem. Int. Ed.* **51**, 6227–6230 (2012).

22. Tinoco, A., Steck, V., Tyagi, V. & Fasan, R. Highly Diastereo- and enantioselective synthesis of trifluoromethyl-substituted cyclopropanes via myoglobin-catalyzed transfer of trifluoromethylcarbene. *J. Am. Chem. Soc.* **139**, 5293–5296 (2017).

23. Huang, X. *et al.* A biocatalytic platform for synthesis of chiral *α*-trifluoromethylated organoborons. *ACS Cent. Sci.* **5**, 270–276 (2019).

24. Duan, Y., Lin, J.-H., Xiao, J.-C. & Gu, Y.-C. Fe-catalyzed insertion of fluoromethylcarbenes generated from sulfonium salts into X–H bonds (X = Si, C, P). *Org. Chem. Front.* **4**, 1917–1920 (2017).

25. Davies, H. M. L. & Morton, D. Guiding principles for site selective and stereoselective intermolecular C–H functionalization by donor/acceptor rhodium carbenes. *Chem. Soc. Rev.* **40**, 1857 (2011).

26. Liao, K., Negretti, S., Musaev, D. G., Bacsa, J. & Davies, H. M. L. Site-selective and stereoselective functionalization of unactivated C–H bonds. *Nature* **533**, 230–234 (2016).

27. Liao, K. *et al.* Site-selective and stereoselective functionalization of non-activated tertiary C–H bonds. *Nature* **551**, 609–613 (2017).

28. Liao, K. *et al.* Design of catalysts for site-selective and enantioselective functionalization of non-activated primary C–H bonds. *Nat. Chem.* **10**, 1048–1055 (2018).

29. Fu, J., Ren, Z., Bacsa, J., Musaev, D. G. & Davies, H. M. L. Desymmetrization of cyclohexanes by site- and stereoselective C–H functionalization. *Nature* **564**, 395–399 (2018).

30. Suematsu, H. & Katsuki, T. Iridium(III) catalyzed diastereo- and enantioselective C−H bond functionalization. *J. Am. Chem. Soc.* **131**, 14218–14219 (2009).

31. Weldy, N. M. *et al.* Iridium(III)-*bis*(imidazolinyl)phenyl catalysts for enantioselective C–H functionalization with ethyl diazoacetate. *Chem. Sci.* **7**, 3142–3146 (2016).

32. Díaz-Requejo, M. M., Belderraín, T. R., Nicasio, M. C., Trofimenko, S. & Pérez, P. J. Intermolecular copper-catalyzed carbon−hydrogen bond activation via carbene insertion. *J. Am. Chem. Soc.* **124**, 896–897 (2002).

33. Griffin, J. R., Wendell, C. I., Garwin, J. A. & White, M. C. Catalytic C(sp^3)–H alkylation via an iron carbene intermediate. *J. Am. Chem. Soc.* **139**, 13624–13627 (2017).

34. Che, C.-M., Lo, V. K.-Y., Zhou, C.-Y. & Huang, J.-S. Selective functionalisation of saturated C–H bonds with metalloporphyrin catalysts. *Chem. Soc. Rev.* **40**, 1950–1975 (2011).

35. Zhu, S.-F. & Zhou, Q.-L. Iron-catalyzed transformations of diazo compounds. *Natl. Sci. Rev.* **1**, 580–603 (2014).

36. Li, Y., Huang, J.-S., Zhou, Z.-Y. & Che, C.-M. Isolation and X-ray crystal structure of an unusual biscarbene metal complex and its reactivity toward cyclopropanation and allylic C−H insertion of unfunctionalized alkenes. *J. Am. Chem. Soc.* **123**, 4843–4844 (2001).

37. Doyle, M. P., Duffy, R., Ratnikov, M. & Zhou, L. Catalytic carbene insertion into C−H bonds. *Chem. Rev.* **110**, 704–724 (2010).

38. Wang, Y., Wen, X., Cui, X. & Zhang, X. P. Enantioselective radical cyclization for construction of 5-membered ring structures by metalloradical C–H alkylation. *J. Am. Chem. Soc.* **140**, 4792–4796 (2018).

39. Caballero, A. *et al.* Silver-catalyzed C-C bond formation between methane and ethyl diazoacetate in supercritical CO_2. *Science* **332**, 835–838 (2011).

40. Gutiérrez-Bonet, Á., Juliá-Hernández, F., de Luis, B. & Martin, R. Pd-catalyzed C(sp^3)–H functionalization/carbenoid insertion: All-carbon quaternary centers via multiple C–C bond formation. *J. Am. Chem. Soc.* **138**, 6384–6387 (2016).

41. Reddy, A. R., Zhou, C.-Y., Guo, Z., Wei, J. & Che, C.-M. Ruthenium–porphyrin-catalyzed diastereoselective intramolecular alkyl carbene insertion into C–H bonds of alkyl diazomethanes generated in situ from *N*-tosylhydrazones. *Angew. Chem. Int. Ed.* **53**, 14175–14180 (2014).

42. Hansen, J., Autschbach, J. & Davies, H. M. L. Computational study on the selectivity of donor/acceptor-substituted rhodium carbenoids. *J. Org. Chem.* **74**, 6555–6563 (2009).

43. Zhang, R. K. *et al.* Enzymatic assembly of carbon–carbon bonds via iron-catalysed sp^3 C–H functionalization. *Nature* **565**, 67–72 (2019).

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

45. Poulos, T. L. Heme enzyme structure and function. *Chem. Rev.* **114**, 3919–3962 (2014).

46. Krest, C. M. *et al.* Significantly shorter Fe–S bond in cytochrome P450-I is consistent with greater reactivity relative to chloroperoxidase. *Nat. Chem.* **7**, 696–702 (2015).

47. Hammer, S. C. *et al.* Anti-Markovnikov alkene oxidation by metal-oxo–mediated enzyme catalysis. *Science* **358**, 215–218 (2017).

48. Onderko, E. L., Silakov, A., Yosca, T. H. & Green, M. T. Characterization of a selenocysteine-ligated P450 compound I reveals direct link between electron donation and reactivity. *Nat. Chem.* **9**, 623–628 (2017).

49. Vatsis, K. P., Peng, H.-M. & Coon, M. J. Replacement of active-site cysteine-436 by serine converts cytochrome P450 2B4 into an NADPH oxidase with negligible monooxygenase activity. *J. Inorg. Biochem.* **91**, 542–553 (2002).

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

51. Giera, D. S., Sickert, M. & Schneider, C. A straightforward synthesis of (*S*)-Anabasine via the catalytic, enantio-selective vinylogous Mukaiyama-Mannich reaction. *Synthesis* **2009**, 3797–3802 (2009).

52. Kawamura, S., Egami, H. & Sodeoka, M. Aminotrifluoromethylation of olefins via cyclic amine formation: Mechanistic study and application to synthesis of trifluoromethylated pyrrolidines. *J. Am. Chem. Soc.* **137**, 4865–4873 (2015).

53. Salamone, M. & Bietti, M. Tuning reactivity and selectivity in hydrogen atom transfer from aliphatic C–H bonds to alkoxyl radicals: Role of structural and medium effects. *Acc. Chem. Res.* **48**, 2895–2903 (2015).

54. Shi, S.-L., Wong, Z. L. & Buchwald, S. L. Copper-catalysed enantioselective stereodivergent synthesis of amino alcohols. *Nature* **532**, 353–356 (2016).

55. Krautwald, S. & Carreira, E. M. Stereodivergence in asymmetric catalysis. *J. Am. Chem. Soc.* **139**, 5627–5639 (2017).

56. Knight, A. M. *et al.* Diverse engineered heme proteins enable stereodivergent cyclopropanation of unactivated alkenes. *ACS Cent. Sci.* **4**, 372–377 (2018).

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

58. Kille, S. *et al.* Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* **2**, 83–92 (2013).

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

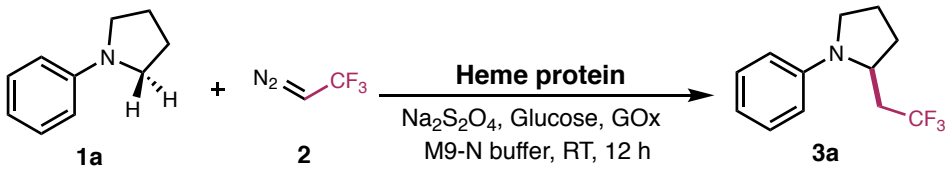
A p p e n d i x A

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

Material for this chapter appears in **Zhang, J.**[†]; Huang, X.[†]; Zhang, R. K.; Arnold, F. H. Enantiodivergent α-amino C–H fluoroalkylation catalyzed by engineered cytochrome P450s. *J. Am. Chem. Soc.*, **2019**, 141, 25, 9798-9802. DOI: 10.1021/jacs.9b04344. ([†]Denotes equal contribution) This work was performed in collaboration with all authors.

A.1 Supporting Table A-1 through A-4

Table A-1. Initial activity screening with heme and heme proteins.[a]



Catalyst	TTN	ee/%	Catalyst	TTN	ee/%
Blank M9-N buffer	NR	-	*A. pernix* protoglobin (*Ap*Pgb) W59G Y60Q	NR	-
Hemin[b]	~ **2**	n.d.	*Ap*Pgb W59G Y60A	NR	-
R. marinus cytochrome c (*Rma* cyt c)	trace	n.d.	CYP119	trace	n.d.
Rma cyt c V75T M100D M103E	trace	n.d.	CYP119 C319S	NR	-
H. thermophilus cyt c	NR	-	$P450_{BM3}$	trace	n.d.
R. globiformis Cyt c WT	NR	-	$P411_{P-4}$ A82L	**40 ± 10**	n.d.
R. marinus nitric oxide dioxygenase *(Rma* NOD) Y32G	trace	n.d.	$P411_{CHA}$	**30 ± 10**	n.d.
Rma NOD Q52V	trace	n.d.	**P411-CH-C8**	**1250 ± 170**	**-12**[c]

[a] Experiments were performed using clarified *E. coli* lysate according to the protocol described in **Section 2.4**. NR – no product was detected; n.d. – not determined.

[b] Experiments with hemin were performed using 5 μM hemin, 10 mM **1a**, 20 mM diazo substrate **2**, 10 mM $Na_2S_2O_4$.

[c] Enantiomeric excess compared to (*R*)-**3a**. The absolute configuration of (*R*)-**3a** was determined by X-ray crystallography (**Section A.8**).

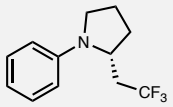
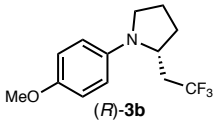
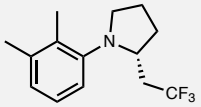
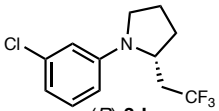
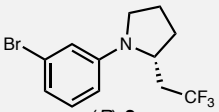
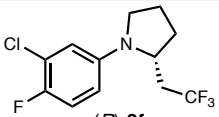
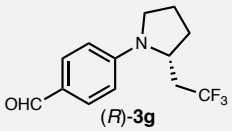
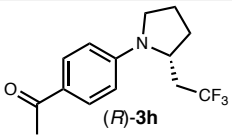
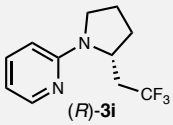
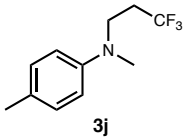
Table A-2. Summary of directed evolution for enantioselective C–H fluoroalkylation.

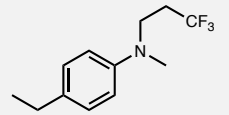
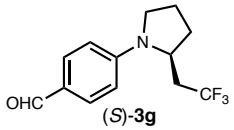
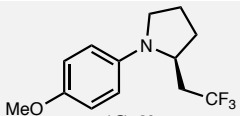
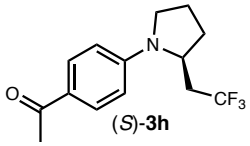
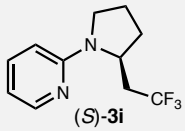
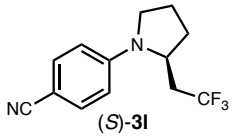
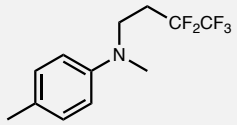
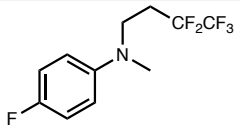
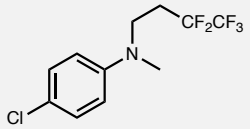
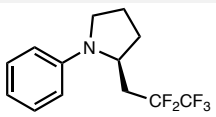
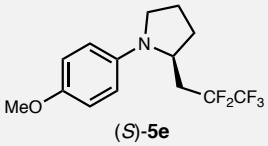
Gen.	Parent	Beneficial mutation identified	TTN	ee*
1	**P411-CH-C8**	T327V	1660 ± 40	28%
2	**P411-FA-B3** (P411-CH-C8 T327V)	E70T	1400 ± 150	32%
3	**P411-FA-B7** (P411-FA-B3 E70T)	L177M, R226T	1300 ± 130	80%
4	**P411-FA-C4** (P411-FA-B7 L177M R226T)	Y330V	1510 ± 240	88%
5	**P411-FA-E3** (P411-FA-C4 Y330V)	L401P	4070 ± 170	98%

* Enantiomeric excess (ee) of (*R*)-**3a**.

Experiments were performed using clarified *E. coli* lysate harboring the corresponding protein prepared according to the protocol described in **Section 2.4**. Reactions were performed in triplicate. TTNs reported are the average of three experiments. Total turnovers (TTN) were defined as the amount of product divided by the total amount of expressed cytochrome P411 protein as determined by the hemochrome assay. Enantiomeric excess (ee) of **3a** was determined by normal phase chiral HPLC as mentioned in **Section A.7**. The absolute configuration of **3a** was determined by X-ray crystallography (**Section A.8**).

Table A-3. Summary of product formation of P411-catalyzed enantioselective C–H fluoroalkylation.

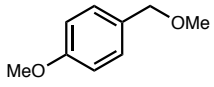
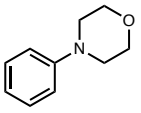
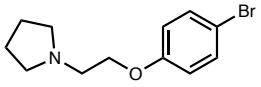
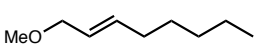
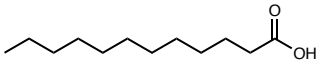
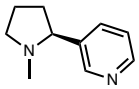
Product formed	Variant used	TTN	ee
 (*R*)-**3a**	P411-PFA	4070 ± 170	98%
 (*R*)-**3b**	P411-PFA	2100 ± 240	95%
 (*R*)-**3c**	P411-PFA	430 ± 40	>99%
 (*R*)-**3d**	P411-PFA	510 ± 70	>99%
 (*R*)-**3e**	P411-PFA	400 ± 60	>99%
 (*R*)-**3f**	P411-PFA	500 ± 20	92%
 (*R*)-**3g**	P411-PFA	690 ± 80	90%
 (*R*)-**3h**	P411-PFA	1440 ± 230	84%
 (*R*)-**3i**	P411-PFA	180 ± 10	88%
 **3j**	P411-PFA	1180 ± 80	-

 **3k**	P411-PFA	120 ± 10	-
 (*S*)-**3g**	P411-PFA-(*S*)	830 ± 50	94%*
 (*S*)-**3b**	P411-PFA-(*S*)	1620 ± 110	82%*
 (*S*)-**3h**	P411-PFA-(*S*)	1820 ± 70	74%*
 (*S*)-**3i**	P411-PFA-(*S*)	180 ± 20	72%*
 (*S*)-**3l**	P411-PFA-(*S*)	1400 ± 70	82%*
 **5a**	P411-PFA	440 ± 80	-
 **5b**	P411-PFA	520 ± 60	-
 **5c**	P411-PFA	250 ± 20	-
 (*S*)-**5d**	P411-PFA	460 ± 50	>99%*
 (*S*)-**5e**	P411-PFA	1550 ± 80	>99%*

* Enantiomeric excess (ee) of the (*S*)-enantiomer.

Experiments were performed using clarified *E. coli* lysate harboring the corresponding protein prepared according to the protocol described in **Section 2.4**. Reactions were performed in triplicate. TTNs reported are the average of three experiments. Total turnovers (TTN) were defined as the amount of product divided by the total amount of expressed cytochrome P411 protein as determined by the hemochrome assay. Enantiomeric excess (ee) of the products were determined by normal phase chiral HPLC or chiral GC as mentioned in **Section A.7.**

Table A-4. Additional substrates tested for P411-catalyzed C–H fluoroalkylation.

Benzylic C–H bond		**Other *N*-aryl amine**		***α*-Amino C–H bonds of aliphatic amine**	
					
1-methoxy-4-(methoxymethyl)benzene		*N*-phenylmorpholine		1-(2-(4-bromophenoxy)ethyl) pyrrolidine	
With **P411-PFA**	**40 ± 2 TTN**	With **P411-PFA**	Trace	With **P411-PFA**	NR
With variant **P411-FA-B7**	**440 ± 10 TTN**	With variant **P411-FA-C4 T226R Y330E**	**P/I* = 0.11**	With variant **P411-FA-E3**	**P/I* = 0.09**
Allylic C–H bond		**Native substrate of $P450_{BM3}$**		**Other**	
					
(*E*)-1-methoxyoct-2-ene		Lauric acid		(*S*)-nicotine	
With **P411-PFA**	NR	With **P411-PFA**	NR	With **P411-PFA**	NR

* P/I: Product versus Internal standard ratio. Product formation was analyzed by GC-MS only. The identity of the product was not confirmed by comparison with chemically synthesized reference compounds or through isolation and characterization. These preliminary results are noteworthy, but should not be used alone for drawing conclusions.

Experiments were performed using *E. coli* whole cells harboring the corresponding protein at OD_{600} = 30. Reactions were performed in triplicate. TTNs reported are the average of three experiments. Total turnovers (TTN) were defined as the amount of product divided by the total amount of expressed cytochrome P411 protein as determined by the hemochrome assay.

A.2 Supporting Figure A-1 through A-4

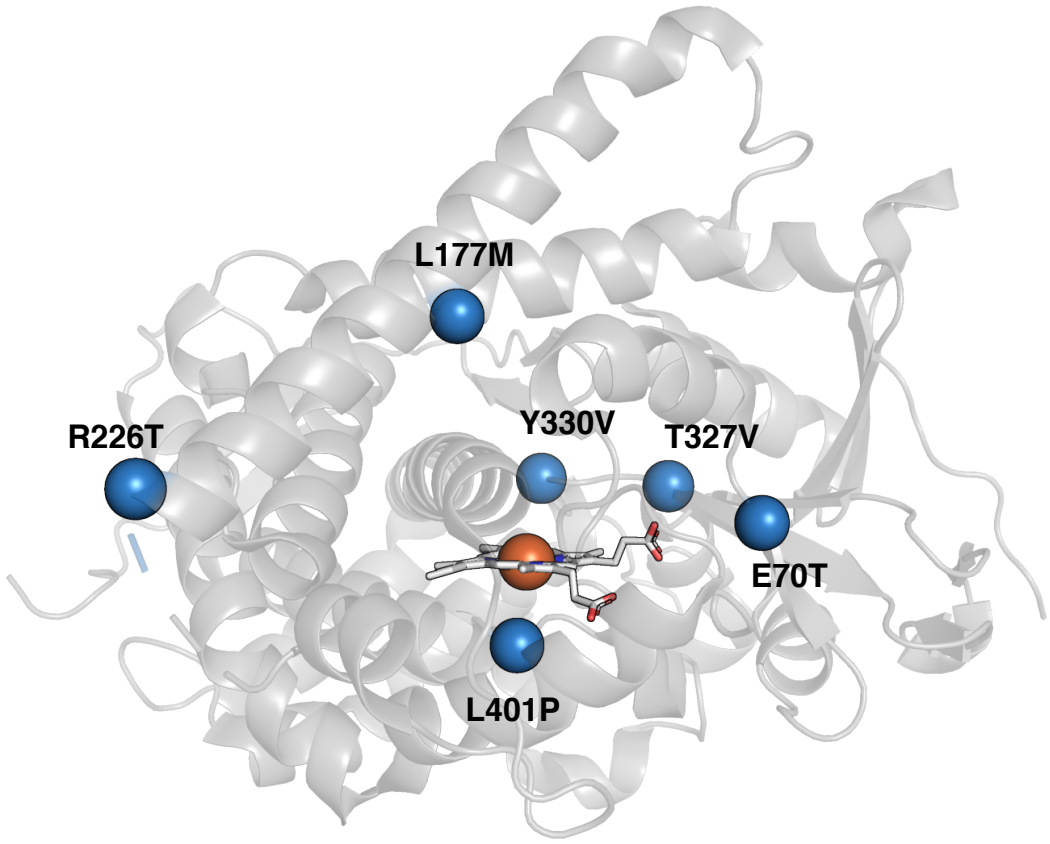
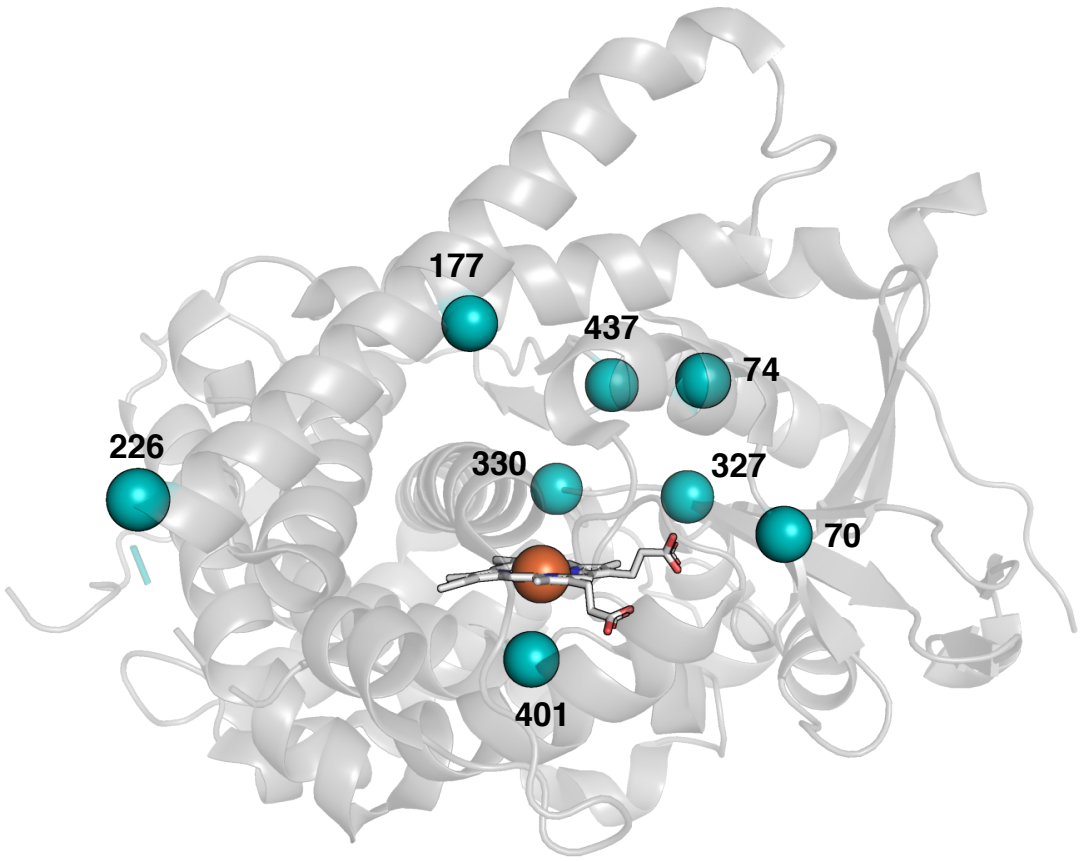


Figure A-1. Structural visualization of amino acid positions which were mutated in the directed evolution trajectory of P411-CH-C8 to P411-PFA.



Position	70	74	177	226	327	330	401	437
P411-PFA	Thr	Gly	Met	Thr	Val	Val	Pro	Gln
P411-PFA-(*S*)	Glu	Pro	Leu	Arg	Thr	Tyr	Leu	Leu

Figure A-2. Structural visualization of amino acid differences between P411-PFA-(*S*) and P411-PFA.

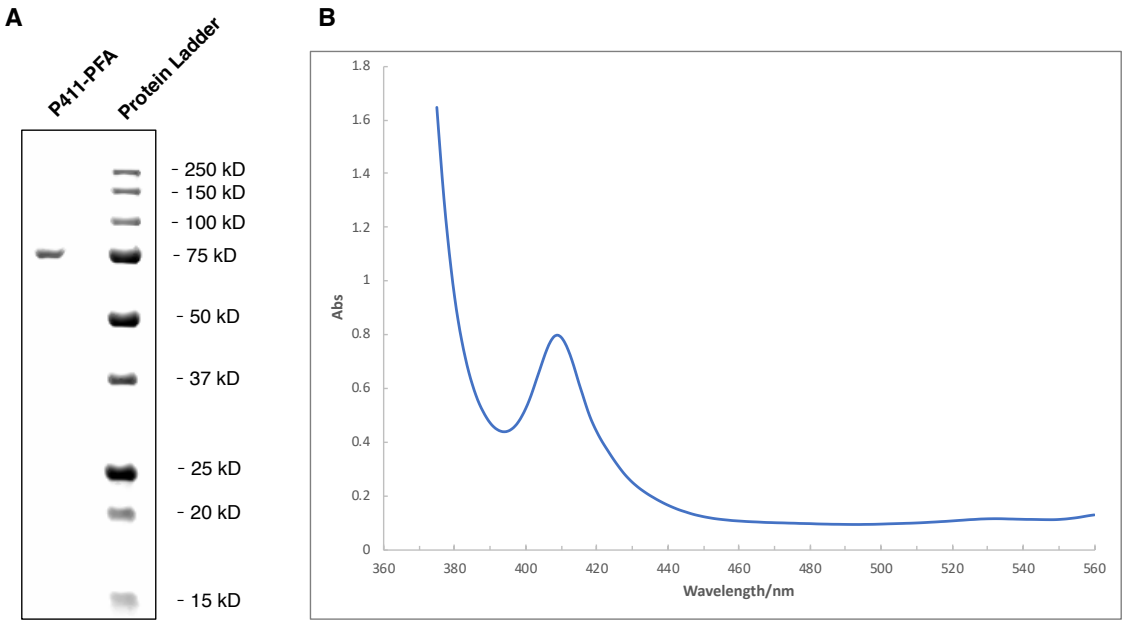
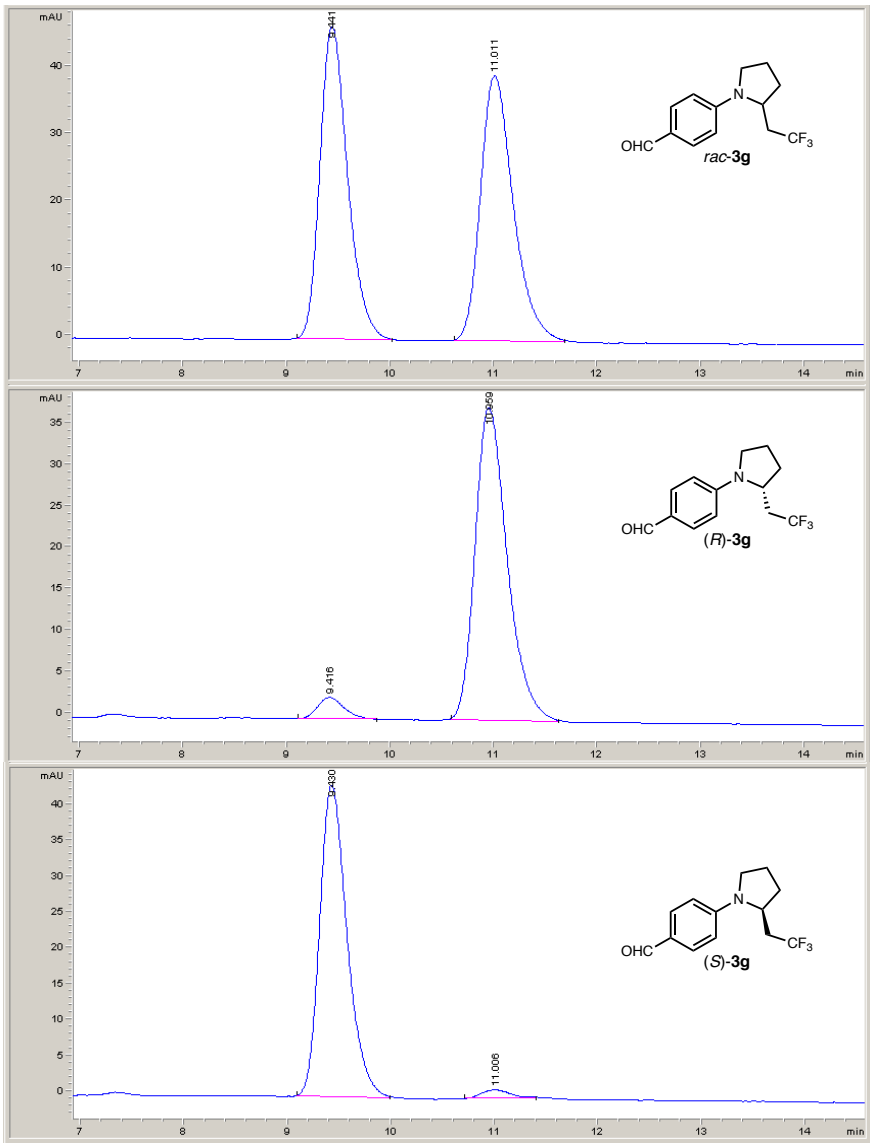


Figure A-3. Characterization of P411-PFA protein. The final variant P411-PFA was purified by using a His-trap column. (**A**) SDS-PAGE gel of P411-PFA. The variant contains an engineered heme domain and the FMN domain of $P450_{BM3}$ (expected size, ~76 kDa); (**B**) UV-visible absorbance spectra of carbon monoxide-bound ferrous P411-PFA. The purified protein was suspended in KPi buffer (0.1 mM, pH = 8.0) with excess amount of $Na_2S_2O_4$. The solution was then purged with CO gas and the absorbance spectrum from 375 nm to 560 nm was immediately collected by a Shimadzu UV-1800 UV-Vis spectrophotometer. The maximum absorbance around 411 nm is consistent with the feature of a typical cytochrome P411 variant.[1]



rac-**3g**			Chiralpak IC (4.6 × 250 nm, 5µm), 30% IPA/hexanes, 1.5 mL/ min, 235 nm		
Retention time (min)	Area (mAu)	Area%			
9.441	869	49.9			
11.011	874	50.1			
(*R*)-**3g** from P411-PFA			(*S*)-**3g** from P411-PFA-(*S*)		
Retention time (min)	Area (mAu)	Area%	Retention time (min)	Area (mAu)	Area%
9.416	46	5.4	9.430	799	97.2
10.959	820	94.6	11.006	23	2.8

Figure A-4. Chiral HPLC analysis of product **3g** obtained by P411-PFA and P411-PFA-(*S*).

A.3 Nucleotide and Amino Acid Sequences

The genes encoding the heme proteins shown below were cloned using Gibson assembly[2] into vector pET-22b(+) (Novagen) between restriction sites *Nde*I and *Xho*I in frame with a *C*-terminal 6xHis-tag.

DNA and amino acid sequences of **P411-CH-C8**:

```
ATGACAATTAAAGAAATGCCTCAGCCAAAAACGTTTGGAGAGCTTAAAAATTTACCGTTATTAAACACAGA
TAAACCGGTTCAAGCTTTGATGAAAATTGCGGATGAATTAGGAGAAATCTTTAAATTCGAGGCGCCTGGTC
GTGTAACGCGCTACTTATCAAGTCAGCGTCTAATTAAAGAAGCATGCGATGAATCACGCTTTGATAAAGAG
TTAAGTCAAGGTCTGAAATTTCTGCGTGATTTTCTTGGAGACGGGTTAGCCACAAGCTGGACGCATGAAAA
AAATTGGAAAAAAGCGCATAATATCTTACTTCCAAGCTTTAGTCAGCAGGCAATGAAAGGCTATCATGCGA
GTATGGTCGATATCGCCGTGCAGCTTGTTCAAAAGTGGGAGCGTCTAAATGCAGATGAGCATATTGAAGTA
TCGGAAGACATGACACGTTTAACGCTTGATACAATTGGTCTTTGCGGCTTTAACTATCGCTTTAACAGCTT
TTACCGAGATCAGCCTCATCCATTTATTATAAGTCTGGTCCGTGCACTGGATGAAGTAATGAACAAGCTGC
AGCGAGCAAATCCAGACGACCCAGCTTATGATGAAAACAAGCGCCAGTTTCAAGAAGATATCAAGGTGATG
AACGACCTAGTAGATAAAATTATTGCAGATCGCAAAGCAAGGGGTGAACAAAGCGATGATTTATTAACGCA
GATGCTAAACGGAAAAGATCCAGAAACGGGTGAGCCGCTTGATGACGGGAACATTCGCTATCAAATTATTA
CATTCTTATATGCGGGAGTTGAAGGTACAAGTGGTCTTTTATCATTTGCGCTGTATTTCTTAGTGAAAAAT
CCACATGTATTACAAAAAGTAGCAGAAGAAGCAGCACGAGTTCTAGTAGATCCTGTTCCAAGCTACAAACA
AGTCAAACAGCTTAAATATGTCGGCATGGTCTTAAACGAAGCGCTGCGCTTATGGCCAACGGTTCCTTATT
TTTCCCTATATGCAAAAGAAGATACGGTGCTTGGAGGAGAATATCCTTTAGAAAAAGGCGACGAAGTAATG
GTTCTGATTCCTCAGCTTCACCGTGATAAAACAGTTTGGGGAGACGATGTGGAGGAGTTCCGTCCAGAGCG
TTTTGAAAATCCAAGTGCGATTCCGCAGCATGCGTTTAAACCGTTTGGAAACGGTCAGCGTGCGTCTCTGG
GTCAGCAGTTCGCTCTTCATGAAGCAACGCTGGTACTTGGTATGATGCTAAAACACTTTGACTTTGAAGAT
CATACAAACTACGAGCTCGATATTAAAGAACTGCAGACGTTAAAACCTAAAGGCTTTGTGGTAAAAGCAAA
ATCGAAAAAAATTCCGCTTGGCGGTATTCCTTCACCTAGCACTGAACAGTCTGCTAAAAAAGTACGCAAAA
AGGCAGAAAACGCTCATAATACGCCGCTGCTTGTGCTATACGGTTCAAATATGGGTACCGCTGAAGGAACG
GCGCGTGATTTAGCAGATATTGCAATGAGCAAAGGATTTGCACCGCAGGTCGCAACGCTTGATTCACACGC
CGGAAATCTTCCGCGCGAAGGAGCTGTATTAATTGTAACGGCGTCTTATAACGGTCATCCGCCTGATAACG
CAAAGCAATTTGTCGACTGGTTAGACCAAGCGTCTGCTGATGAAGTAAAAGGCGTTCGCTACTCCGTATTT
GGATGCGGCGATAAAAACTGGGCTACTACGTATCAAAAAGTGCCTGCTTTTATCGATGAAACGCTTGCCGC
TAAAGGGGCAGAAAACATCGCTGACCGCGGTGAAGCAGATGCAAGCGACGACTTTGAAGGCACATATGAAG
AATGGCGTGAACATATGTGGAGTGACGTAGCAGCCTACTTTAACCTCGACATTGAAAACAGTGAAGATAAT
AAATCTACTCTTTCACTTCAATTTGTCGACAGCGCCGCGGATATGCCGCTTGCGAAAATGCACGGTGCGTT
TTCAACGCTCGAGCACCACCACCACCACCACTGA
```

```
MTIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADELGEIFKFEAPGRVTRYLSSQRLIKEACDESRFDKE
LSQGLKFLRDFLGDGLATSWTHEKNWKKAHNILLPSFSQQAMKGYHASMVDIAVQLVQKWERLNADEHIEV
SEDMTRLTLDTIGLCGFNYRFNSFYRDQPHPFIISLVRALDEVMNKLQRANPDDPAYDENKRQFQEDIKVM
NDLVDKIIADRKARGEQSDDLLTQMLNGKDPETGEPLDDGNIRYQIITFLYAGVEGTSGLLSFALYFLVKN
PHVLQKVAEEAARVLVDPVPSYKQVKQLKYVGMVLNEALRLWPTVPYFSLYAKEDTVLGGEYPLEKGDEVM
VLIPQLHRDKTVWGDDVEEFRPERFENPSAIPQHAFKPFGNGQRASLGQQFALHEATLVLGMMLKHFDFED
HTNYELDIKELQTLKPKGFVVKAKSKKIPLGGIPSPSTEQSAKKVRKKAENAHNTPLLVLYGSNMGTAEGT
ARDLADIAMSKGFAPQVATLDSHAGNLPREGAVLIVTASYNGHPPDNAKQFVDWLDQASADEVKGVRYSVF
GCGDKNWATTYQKVPAFIDETLAAKGAENIADRGEADASDDFEGTYEEWREHMWSDVAAYFNLDIENSEDN
KSTLSLQFVDSAADMPLAKMHGAFSTLEHHHHHH*
```

DNA and amino acid sequences of **P411-PFA**:

ATGACAATTAAAGAAATGCCTCAGCCAAAAACGTTTGGAGAGCTTAAAAATTTACCGTTATTAAACACAGA
TAAACCGGTTCAAGCTTTGATGAAAATTGCGGATGAATTAGGAGAAATCTTTAAATTCGAGGCGCCTGGTC
GTGTAACGCGCTACTTATCAAGTCAGCGTCTAATTAAAGAAGCATGCGATGAATCACGCTTTGATAAAACA
TTAAGTCAAGGTCTGAAATTTCTGCGTGATTTTCTTGGAGACGGGTTAGCCACAAGCTGGACGCATGAAAA
AAATTGGAAAAAAGCGCATAATATCTTACTTCCAAGCTTTAGTCAGCAGGCAATGAAAGGCTATCATGCGA
GTATGGTCGATATCGCCGTGCAGCTTGTTCAAAAGTGGGAGCGTCTAAATGCAGATGAGCATATTGAAGTA
TCGGAAGACATGACACGTTTAACGCTTGATACAATTGGTCTTTGCGGCTTTAACTATCGCTTTAACAGCTT
TTACCGAGATCAGCCTCATCCATTTATTATAAGTATGGTCCGTGCACTGGATGAAGTAATGAACAAGCTGC
AGCGAGCAAATCCAGACGACCCAGCTTATGATGAAAACAAGCGCCAGTTTCAAGAAGATATCAAGGTGATG
AACGACCTAGTAGATAAAATTATTGCAGATCGCAAAGCAACAGGTGAACAAAGCGATGATTTATTAACGCA
GATGCTAAACGGAAAAGATCCAGAAACGGGTGAGCCGCTTGATGACGGGAACATTCGCTATCAAATTATTA
CATTCTTATATGCGGGAGTTGAAGGTACAAGTGGTCTTTTATCATTTGCGCTGTATTTCTTAGTGAAAAAT
CCACATGTATTACAAAAAGTAGCAGAAGAAGCAGCACGAGTTCTAGTAGATCCTGTTCCAAGCTACAAACA
AGTCAAACAGCTTAAATATGTCGGCATGGTCTTAAACGAAGCGCTGCGCTTATGGCCAGTGGTTCCTGTGT
TTTCCCTATATGCAAAAGAAGATACGGTGCTTGGAGGAGAATATCCTTTAGAAAAAGGCGACGAAGTAATG
GTTCTGATTCCTCAGCTTCACCGTGATAAAACAGTTTGGGGAGACGATGTGGAGGAGTTCCGTCCAGAGCG
TTTTGAAAATCCAAGTGCGATTCCGCAGCATGCGTTTAAACCGTTTGGAAACGGTCAGCGTGCGTCTCCGG
GTCAGCAGTTCGCTCTTCATGAAGCAACGCTGGTACTTGGTATGATGCTAAAACACTTTGACTTTGAAGAT
CATACAAACTACGAGCTCGATATTAAAGAACTGCAGACGTTAAAACCTAAAGGCTTTGTGGTAAAAGCAAA
ATCGAAAAAAATTCCGCTTGGCGGTATTCCTTCACCTAGCACTGAACAGTCTGCTAAAAAAGTACGCAAAA
AGGCAGAAAACGCTCATAATACGCCGCTGCTTGTGCTATACGGTTCAAATATGGGTACCGCTGAAGGAACG
GCGCGTGATTTAGCAGATATTGCAATGAGCAAAGGATTTGCACCGCAGGTCGCAACGCTTGATTCACACGC
CGGAAATCTTCCGCGCGAAGGAGCTGTATTAATTGTAACGGCGTCTTATAACGGTCATCCGCCTGATAACG
CAAAGCAATTTGTCGACTGGTTAGACCAAGCGTCTGCTGATGAAGTAAAAGGCGTTCGCTACTCCGTATTT
GGATGCGGCGATAAAAACTGGGCTACTACGTATCAAAAAGTGCCTGCTTTTATCGATGAAACGCTTGCCGC
TAAAGGGGCAGAAAACATCGCTGACCGCGGTGAAGCAGATGCAAGCGACGACTTTGAAGGCACATATGAAG
AATGGCGTGAACATATGTGGAGTGACGTAGCAGCCTACTTTAACCTCGACATTGAAAACAGTGAAGATAAT
AAATCTACTCTTTCACTTCAATTTGTCGACAGCGCCGCGGATATGCCGCTTGCGAAAATGCACGGTGCGTT
TTCAACGCTCGAGCACCACCACCACCACCACTGA

MTIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADELGEIFKFEAPGRVTRYLSSQRLIKEACDESRFDKT
LSQGLKFLRDFLGDGLATSWTHEKNWKKAHNILLPSFSQQAMKGYHASMVDIAVQLVQKWERLNADEHIEV
SEDMTRLTLDTIGLCGFNYRFNSFYRDQPHPFIISMVRALDEVMNKLQRANPDDPAYDENKRQFQEDIKVM
NDLVDKIIADRKATGEQSDDLLTQMLNGKDPETGEPLDDGNIRYQIITFLYAGVEGTSGLLSFALYFLVKN
PHVLQKVAEEAARVLVDPVPSYKQVKQLKYVGMVLNEALRLWPVVPVFSLYAKEDTVLGGEYPLEKGDEVM
VLIPQLHRDKTVWGDDVEEFRPERFENPSAIPQHAFKPFGNGQRASPGQQFALHEATLVLGMMLKHFDFED
HTNYELDIKELQTLKPKGFVVKAKSKKIPLGGIPSPSTEQSAKKVRKKAENAHNTPLLVLYGSNMGTAEGT
ARDLADIAMSKGFAPQVATLDSHAGNLPREGAVLIVTASYNGHPPDNAKQFVDWLDQASADEVKGVRYSVF
GCGDKNWATTYQKVPAFIDETLAAKGAENIADRGEADASDDFEGTYEEWREHMWSDVAAYFNLDIENSEDN
KSTLSLQFVDSAADMPLAKMHGAFSTLEHHHHHH*

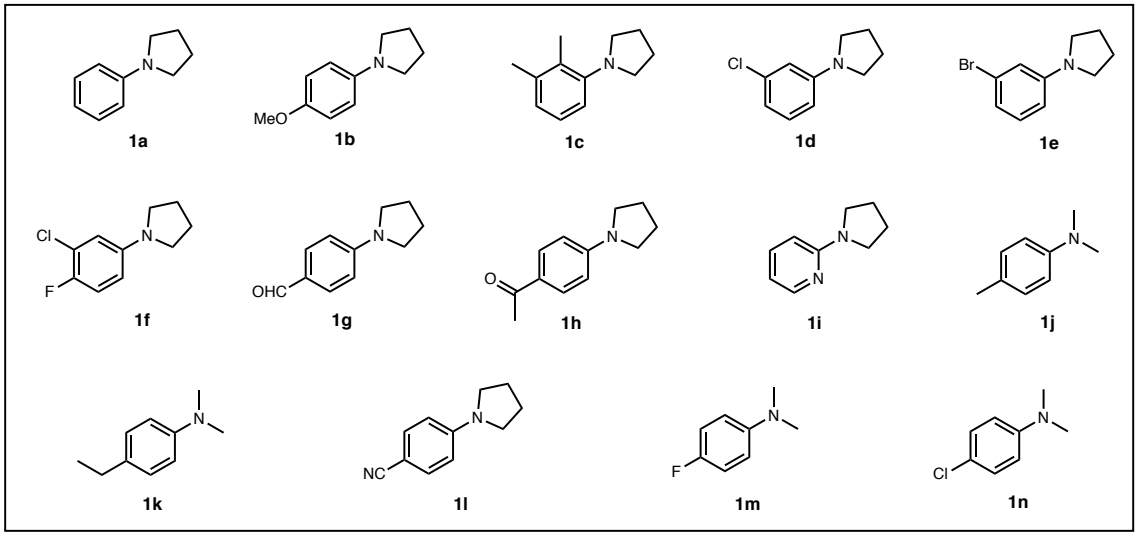
DNA and amino acid sequences of **P411-PFA-(*S*)**:

ATGACAATTAAAGAAATGCCTCAGCCAAAAACGTTTGGAGAGCTTAAAAATTTACCGTTATTAAACACAGA
TAAACCGGTTCAAGCTTTGATGAAAATTGCGGATGAATTAGGAGAAATCTTTAAATTCGAGGCGCCTGGTC
GTGTAACGCGCTACTTATCAAGTCAGCGTCTAATTAAAGAAGCATGCGATGAATCACGCTTTGATAAAGAG
TTAAGTCAACCGCTGAAATTTCTGCGTGATTTTCTTGGAGACGGGTTAGCCACAAGCTGGACGCATGAAAA
AAATTGGAAAAAAGCGCATAATATCTTACTTCCAAGCTTTAGTCAGCAGGCAATGAAAGGCTATCATGCGA
GTATGGTCGATATCGCCGTGCAGCTTGTTCAAAAGTGGGAGCGTCTAAATGCAGATGAGCATATTGAAGTA
TCGGAAGACATGACACGTTTAACGCTTGATACAATTGGTCTTTGCGGCTTTAACTATCGCTTTAACAGCTT
TTACCGAGATCAGCCTCATCCATTTATTATAAGTCTGGTCCGTGCACTGGATGAAGTAATGAACAAGCTGC

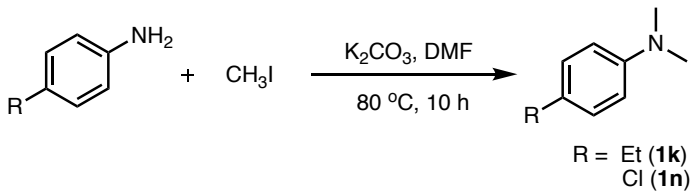
AGCGAGCAAATCCAGACGACCCAGCTTATGATGAAAACAAGCGCCAGTTTCAAGAAGATATCAAGGTGATG
AACGACCTAGTAGATAAAATTATTGCAGATCGCAAAGCAAGGGGTGAACAAAGCGATGATTTATTAACGCA
GATGCTAAACGGAAAAGATCCAGAAACGGGTGAGCCGCTTGATGACGGGAACATTCGCTATCAAATTATTA
CATTCTTATATGCGGGAGTTGAAGGTACAAGTGGTCTTTTATCATTTGCGCTGTATTTCTTAGTGAAAAAT
CCACATGTATTACAAAAAGTAGCAGAAGAAGCAGCACGAGTTCTAGTAGATCCTGTTCCAAGCTACAAACA
AGTCAAACAGCTTAAATATGTCGGCATGGTCTTAAACGAAGCGCTGCGCTTATGGCCAACGGTTCCTTATT
TTTCCCTATATGCAAAAGAAGATACGGTGCTTGGAGGAGAATATCCTTTAGAAAAAGGCGACGAAGTAATG
GTTCTGATTCCTCAGCTTCACCGTGATAAAACAGTTTGGGGAGACGATGTGGAGGAGTTCCGTCCAGAGCG
TTTTGAAAATCCAAGTGCGATTCCGCAGCATGCGTTTAAACCGTTTGGAAACGGTCAGCGTGCGTCTCTGG
GTCAGCAGTTCGCTCTTCATGAAGCAACGCTGGTACTTGGTATGATGCTAAAACACTTTGACTTTGAAGAT
CATACAAACTACGAGCTCGATATTAAAGAACTGCTTACGTTAAAACCTAAAGGCTTTGTGGTAAAAGCAAA
ATCGAAAAAAATTCCGCTTGGCGGTATTCCTTCACCTAGCACTGAACAGTCTGCTAAAAAAGTACGCAAAA
AGGCAGAAAACGCTCATAATACGCCGCTGCTTGTGCTATACGGTTCAAATATGGGTACCGCTGAAGGAACG
GCGCGTGATTTAGCAGATATTGCAATGAGCAAAGGATTTGCACCGCAGGTCGCAACGCTTGATTCACACGC
CGGAAATCTTCCGCGCGAAGGAGCTGTATTAATTGTAACGGCGTCTTATAACGGTCATCCGCCTGATAACG
CAAAGCAATTTGTCGACTGGTTAGACCAAGCGTCTGCTGATGAAGTAAAAGGCGTTCGCTACTCCGTATTT
GGATGCGGCGATAAAAACTGGGCTACTACGTATCAAAAAGTGCCTGCTTTTATCGATGAAACGCTTGCCGC
TAAAGGGGCAGAAAACATCGCTGACCGCGGTGAAGCAGATGCAAGCGACGACTTTGAAGGCACATATGAAG
AATGGCGTGAACATATGTGGAGTGACGTAGCAGCCTACTTTAACCTCGACATTGAAAACAGTGAAGATAAT
AAATCTACTCTTTCACTTCAATTTGTCGACAGCGCCGCGGATATGCCGCTTGCGAAAATGCACGGTGCGTT
TTCAACGCTCGAGCACCACCACCACCACCACTGA

MTIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADELGEIFKFEAPGRVTRYLSSQRLIKEACDESRFDKE
LSQPLKFLRDFLGDGLATSWTHEKNWKKAHNILLPSFSQQAMKGYHASMVDIAVQLVQKWERLNADEHIEV
SEDMTRLTLDTIGLCGFNYRFNSFYRDQPHPFIISLVRALDEVMNKLQRANPDDPAYDENKRQFQEDIKVM
NDLVDKIIADRKARGEQSDDLLTQMLNGKDPETGEPLDDGNIRYQIITFLYAGVEGTSGLLSFALYFLVKN
PHVLQKVAEEAARVLVDPVPSYKQVKQLKYVGMVLNEALRLWPTVPYFSLYAKEDTVLGGEYPLEKGDEVM
VLIPQLHRDKTVWGDDVEEFRPERFENPSAIPQHAFKPFGNGQRASLGQQFALHEATLVLGMMLKHFDFED
HTNYELDIKELLTLKPKGFVVKAKSKKIPLGGIPSPSTEQSAKKVRKKAENAHNTPLLVLYGSNMGTAEGT
ARDLADIAMSKGFAPQVATLDSHAGNLPREGAVLIVTASYNGHPPDNAKQFVDWLDQASADEVKGVRYSVF
GCGDKNWATTYQKVPAFIDETLAAKGAENIADRGEADASDDFEGTYEEWREHMWSDVAAYFNLDIENSEDN
KSTLSLQFVDSAADMPLAKMHGAFSTLEHHHHHH*

A.4 Substrate Syntheses and Characterizations[a]

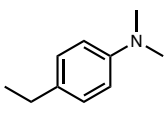


[a] **1a-1j**, **1i**, **1m** were obtained from commercial sources and were used as received.



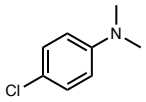
N, *N*-dimethyl anilines **1k** and **1n** were prepared according to a modified procedure reported by Shi[3] and McNally[4]. Namely, a mixture of aniline (5.0 mmol), iodomethane (11.0 mmol), and K_2CO_3 (11.0 mmol) in DMF (20 mL) was stirred at 80 °C for 10 h. The reaction was then cooled and diluted with EtOAc (100 mL) and H_2O (100 mL). The layers were separated, and the organic layer was washed with brine (20 mL), dried over $MgSO_4$, filtered, concentrated, and purified by flash chromatography.

4-ethyl-*N*,*N*-dimethylaniline (1k)



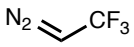
This compound is known.[5] **^{1}H NMR** (300 MHz, $CDCl_3$) δ 7.11 (d, J = 8.6 Hz, 2H), 6.74 (d, J = 8.6 Hz, 2H), 2.93 (s, 6H), 2.59 (q, J = 7.6 Hz, 2H), 1.23 (t, J = 7.6 Hz, 3H).

4-chloro-*N*,*N*-dimethylaniline (1n)



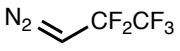
This compound is known.[6] **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.17 (d, *J* = 9.1 Hz, 2H), 6.64 (d, *J* = 9.1 Hz, 2H), 2.93 (s, 6H).

2-diazo-1,1,1-trifluoroethane (2)



The stock solution of diazo compounds **2** was prepared using a modified procedure reported by Ma et al.[7] Namely, a solution of 610 mg $NaNO_2$ in 1 mL of water was added slowly to a vigorously stirring solution of 1.08 g of trifluoroethylamine hydrochloride in 2 mL water at room temperature. The rapidly evolved yellow gas was carefully bubbled via PTFE tubing into 7 mL ethanol placed in a 10-mL vial and chilled on an ice-salt bath. After the cease of the gas bubbling (around 2–3 hours), the vial containing the 2-diazo-1,1,1-trifluoroethane ethanol solution was carefully removed from the ice-salt bath and stored in -20 °C freezer. The concentration of the stock solution was measured by ^{19}F NMR with fluorobenzene as internal standard.

3-diazo-1,1,1,2,2-pentafluoropropane (4)

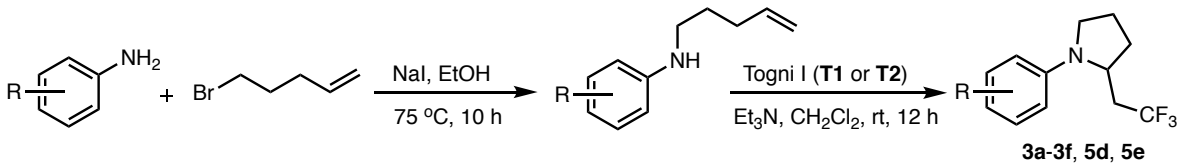


The stock solution of diazo compounds **4** was prepared by reacting $NaNO_2$ with 2,2,2,3,3-pentafluoropropylamine hydrochloride followed a procedure similar to the preparation of **2**. A slow stream of Ar was used to facilitate the transfer of **4** into the ethanol trapping solution. The concentration of the stock solution was measured by ^{19}F NMR with fluorobenzene as internal standard.

CAUTION: Diazo compounds are toxic and potentially explosive and should be handled with care in a well-ventilated hood.

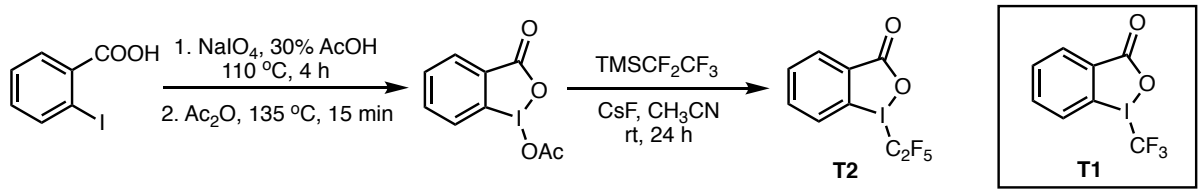
A.5 Syntheses and Characterizations of Reference Compounds

General procedure A:



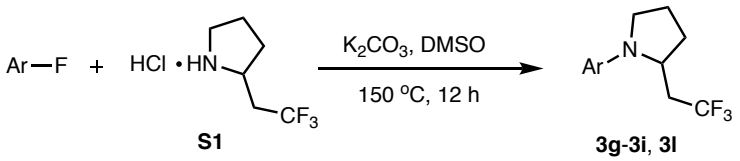
The synthesis of alkenylamine precursors was achieved with a procedure reported by Shen et al.[8]: To a suspension of NaI (0.1 mmol) in EtOH (25 mL), aniline (5 mmol), 5-bromopentene (3 mmol) was added, and the reaction was refluxed at 75 °C for 10 h. The resulting reaction mixture were concentrated and purified by flash chromatography.

Racemic reference compounds **3a-3f**, **5d**, **5e** were synthesized from the corresponding alkenylamines following the procedure reported by Kawamura et al.[9] Namely, a reaction vial charged with CuI (0.01 mmol) and Togni reagent I (0.30 mmol) was degassed and backfilled with argon for three times. A solution of Et_3N (0.04 mmol) in CH_2Cl_2 (0.5 mL) was then added, and the resulting suspension was stirred at room temperature. After 30 min, alkenylamine (0.20 mmol) was added and the reaction mixture was subsequently stirred for 12 h at room temperature. The crude reaction mixture was diluted with 5 mL CH_2Cl_2 and filtered. The collected filtrate was concentrated and purified by chromatography.



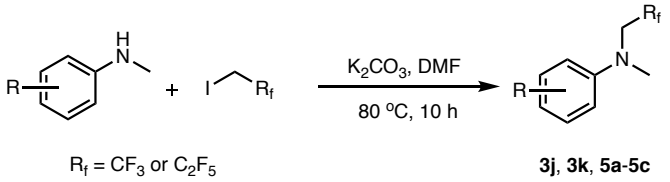
Togni reagent I (**T1**) is obtained from Sigma Aldrich and was used as received. Pentafluoroethyl Togni Reagent (**T2**) was synthesized according to the procedure reported by Li et al.[10]

General procedure B:



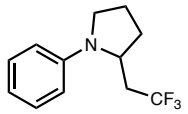
Reference compounds **3g-3i**, **3l** were prepared from the S_NAr reaction between aryl fluorides and 2-(2,2,2-trifluoroethyl)pyrrolidine with the following protocol. To a suspension of K_2CO_3 (0.40 mmol, 55.2 mg) in DMSO (1.0 mL) was added 2-(2,2,2-trifluoroethyl)pyrrolidine hydrochloride **S1** (0.30 mmol, 57.0 mg, obtained from Enamine Ltd.). The reaction was degassed and backfilled with argon for three times. The appropriate aryl fluoride (0.40 mmol) was then added, and the reaction was heated at 150 °C for 12 h. The reaction was cooled and thoroughly mixed with EtOAc (10 mL) and H_2O (10 mL). The layers were separated, and the aqueous layer was further extracted with EtOAc (3 x 20 mL). The organic layers were combined, washed with brine (10 mL), dried over $MgSO_4$, filtered, concentrated. The obtained crude product was further purified by flash chromatography.

General Procedure C:



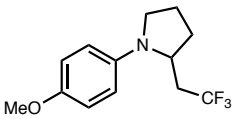
Reference compounds **3j**, **3k**, **5a-5c** were prepared from the *N*-fluoroalkylation of *N*-methylanilines with the following protocol. To a suspension of K_2CO_3 (1.1 mmol, 152 mg) in DMF (2.0 mL) was added the appropriate aniline (1.0 mmol). The reaction was degassed and backfilled with argon for three times. Polyfluoroalkyl iodide (0.40 mmol) was then added, and the reaction was heated at 150 °C for 12 h. The reaction was cooled to room temperature and thoroughly mixed with EtOAc (10 mL) and H_2O (10 mL). The layers were separated, and the aqueous layer was further extracted with EtOAc (3 x 20 mL). The organic layers were combined, washed with brine (10 mL), dried over $MgSO_4$, filtered, and concentrated. The obtained crude product was further purified by flash chromatography.

1-phenyl-2-(2,2,2-trifluoroethyl)pyrrolidine (3a)



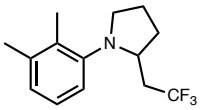
This compound is known.[9] **^{1}H NMR** (400 MHz, $CDCl_3$) δ 7.35 – 7.18 (m, 2H), 6.72 (tt, *J* = 7.2, 1.1 Hz, 1H), 6.59 (dt, *J* = 7.2, 1.1 Hz, 2H), 4.04 (ddt, *J* = 10.5, 6.9, 1.7 Hz, 1H), 3.44 (ddd, *J* = 8.7, 7.1, 3.3 Hz, 1H), 3.17 (td, *J* = 8.6, 6.5 Hz, 1H), 2.62 – 2.44 (m, 1H), 1.97 – 2.14 (m, 5H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 146.2, 129.5, 126.3 (q, *J* = 277.7 Hz),116.4, 111.9, 52.6, 47.8, 36.7 (q, *J* = 25.4 Hz), 31.1, 23.0; **^{19}F NMR** (282 MHz, $CDCl_{3v}$) δ -63.9 (t, *J* = 11.2 Hz).

1-(4-methoxyphenyl)-2-(2,2,2-trifluoroethyl)pyrrolidine (3b)



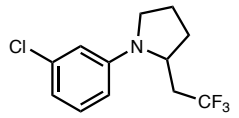
This compound is known.[9] **^{1}H NMR** (400 MHz, $CDCl_3$) δ 6.92 – 6.83 (m, 2H), 6.55 (d, *J* = 8.5 Hz, 2H), 3.94 (dd, *J* = 10.5, 7.0 Hz, 1H), 3.76 (s, 3H), 3.41 (m, 1H), 3.13 (m, 1H), 2.63 – 2.40 (m, 1H), 2.21 – 1.95 (m, 5H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 151.3, 141.1, 126.4 (q, *J* = 277.9 Hz), 115.3, 112.9, 55.9, 53.0, 48.6, 37.0 (q, *J* = 24.9 Hz), 31.3, 23.2; **^{19}F NMR** (282 MHz, $CDCl_3$) δ -63.9 (t, *J* = 11.2 Hz).

1-(2,3-dimethylphenyl)-2-(2,2,2-trifluoroethyl)pyrrolidine (3c)



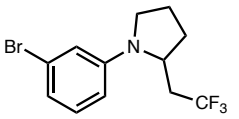
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.08 (t, *J* = 7.7 Hz, 1H), 6.96 – 6.86 (m, 2H), 3.71 (dddd, *J* = 10.4, 8.5, 6.5, 2.6 Hz, 1H), 3.45 (ddd, *J* = 9.2, 7.6, 5.8 Hz, 1H), 2.64 (ddd, *J* = 9.2, 8.4, 5.9 Hz, 1H), 2.30 – 2.37 (m, 2H), 2.28 (s, 3H), 2.18 (s, 3H), 2.02 – 1.81 (m, 3H), 1.78 – 1.68 (m, 1H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 147.6, 138.5, 133.1, 127.1 (q, *J* = 277.2 Hz), 126.2, 125.3, 117.2, 55.0 (q, *J* = 2.8 Hz), 54.1, 38.4 (q, *J* = 26.3 Hz), 32.1, 23.7, 21.0, 14.8; **^{19}F NMR** (282 MHz, $CDCl_3$) δ -63.8 (t, *J* = 11.4 Hz); **HRMS** (ESI) *m/z* $[M + H]^+$ calcd for $C_{14}H_{19}NF_3^+$: 258.1470, found: 258.1453.

1-(3-chlorophenyl)-2-(2,2,2-trifluoroethyl)pyrrolidine (3d)



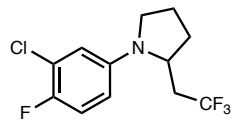
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.15 (t, *J* = 8.1 Hz, 1H), 6.68 (ddd, *J* = 7.9, 2.0, 0.9 Hz, 1H), 6.55 (t, *J* = 2.2 Hz, 1H), 6.45 (ddd, *J* = 8.4, 2.5, 0.8 Hz, 1H), 4.01 (ddt, *J* = 10.3, 6.7, 1.8 Hz, 1H), 3.51 – 3.34 (m, 1H), 3.26 – 3.06 (m, 1H), 2.61 – 2.35 (m, 1H), 2.23 – 1.89 (m, 5H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 147.5, 135.7, 130.7, 126.5 (q, *J* = 277.7 Hz), 116.6, 112.2, 110.3, 53.0 (q, *J* = 3.1 Hz), 48.2, 36.9 (q, *J* = 25.7 Hz), 31.4, 23.3; **^{19}F NMR** (282 MHz, $CDCl_3$) δ -63.9 (t, *J* = 11.1 Hz); **HRMS** (FAB) *m/z* $[M]^+$• calcd for $C_{12}H_{13}NClF_3^+$: 263.0689, found: 263.0691.

1-(3-bromophenyl)-2-(2,2,2-trifluoroethyl)pyrrolidine (3e)



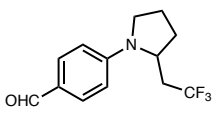
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.09 (dd, *J* = 8.3, 7.9 Hz, 1H), 6.83 (ddd, *J* = 7.9, 1.8, 0.9 Hz, 1H), 6.71 (t, *J* = 2.2 Hz, 1H), 6.49 (ddd, *J* = 8.4, 2.5, 0.9 Hz, 1H), 4.09 – 3.95 (m, 1H), 3.41 (ddt, *J* = 8.8, 5.4, 1.7 Hz, 1H), 3.21 – 3.11 (m, 1H), 2.61 – 2.35 (m, 1H), 2.24 – 1.89 (m, 5H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 147.6, 131.0, 126.5 (q, *J* = 277.8 Hz), 124.0, 119.5, 115.1, 110.8, 53.0 (q, *J* = 3.0 Hz), 48.2, 36.9 (q, *J* = 25.7 Hz), 31.4, 23.3; **^{19}F NMR** (282 MHz, $CDCl_3$) δ -63.8 (t, J = 11.1 Hz); **HRMS** (FAB) *m/z* $[M]^+$• calcd for $C_{12}H_{13}NBrF_3^+$: 307.0183, found: 307.0177.

1-(3-chloro-4-fluorophenyl)-2-(2,2,2-trifluoroethyl)pyrrolidine (3f)



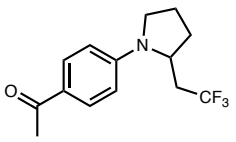
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.02 (t, *J* = 8.9 Hz, 1H), 6.54 (dd, *J* = 6.0, 3.0 Hz, 1H), 6.38 (dt, *J* = 9.1, 3.3 Hz, 1H), 3.94 (ddt, *J* = 10.3, 6.8, 1.7 Hz, 1H), 3.47 – 3.35 (m, 1H), 3.18 – 3.06 (m, 1H), 2.56 – 2.32 (m, 1H), 2.20 – 1.98 (m, 5H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 150.7 (d, *J* = 237.3 Hz), 143.6 (d, *J* = 1.8 Hz), 126.5 (q, *J* = 277.6 Hz), 121.7 (d, *J* = 18.3 Hz), 117.3 (d, *J* = 21.7 Hz), 113.3, 111.0 (d, *J* = 6.2 Hz), 53.4 (q, *J* = 3.0 Hz), 48.7, 37.0 (q, *J* = 25.7 Hz), 31.6, 23.4; **^{19}F NMR** (282 MHz, $CDCl_3$) δ -63.9 (t, *J* = 11.0 Hz), -132.7; **HRMS** (FAB) *m/z* $[M]^+$• calcd for $C_{12}H_{12}NClF_4^+$: 281.0594, found: 281.0583

4-(2-(2,2,2-trifluoroethyl)pyrrolidin-1-yl)benzaldehyde (3g)



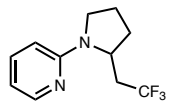
^{1}H NMR (400 MHz, $CDCl_3$) δ 9.75 (s, 1H), 7.76 (d, *J* = 8.8 Hz, 2H), 6.61 (d, *J* = 8.8 Hz, 2H), 4.16 (ddt, *J* = 10.4, 3.9, 1.9 Hz, 1H), 3.52 (ddd, *J* = 10.9, 6.8, 2.1 Hz, 1H), 3.37 – 3.21 (m, 1H), 2.42 – 2.55 (m, 1H), 2.22 – 1.96 (m, 5H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 190.7, 150.8, 132.7, 126.3 (q, *J* = 277.7 Hz), 126.1, 111.9, 53.1 (q, *J* = 3.0 Hz), 48.1, 36.6 (q, *J* = 26.0 Hz), 31.2, 23.1; **^{19}F NMR** (282 MHz, $CDCl_3$) δ -63.9 (t, *J* = 10.9 Hz); **HRMS** (FAB) *m/z* $[M + H]^+$ calcd for $C_{13}H_{15}ONF_3^+$: 258.1106, found: 258.1109.

1-(4-(2-(2,2,2-trifluoroethyl)pyrrolidin-1-yl)phenyl)ethan-1-one (3h)



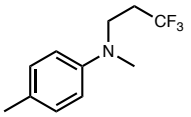
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.90 (d, *J* = 9.0 Hz, 2H), 6.56 (d, *J* = 9.0 Hz, 2H), 4.14 (ddt, *J* = 10.4, 5.3, 1.9 Hz, 1H), 3.51 (ddd, *J* = 8.4, 7.2, 3.4 Hz, 1H), 3.34 – 3.17 (m, 1H), 2.60 – 2.37 (m, 1H), 2.51 (s, 3H), 2.22 – 1.89 (m, 5H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 196.7, 149.7, 131.2, 126.3 (q, *J* = 277.7 Hz), 126.3, 111.4, 53.0 (q, *J* = 3.0 Hz), 48.0, 36.7 (q, *J* = 26.0 Hz), 31.3, 26.4, 23.1; **^{19}F NMR** (282 MHz, $CDCl_3$) δ -63.9 (t, *J* = 10.9 Hz); **HRMS** (FAB) *m/z* $[M + H]^+$ calcd for $C_{14}H_{17}ONF_3^+$: 272.1262, found: 272.1267.

2-(2-(2,2,2-trifluoroethyl)pyrrolidin-1-yl)pyridine (3i)



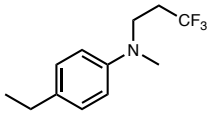
^{1}H NMR (400 MHz, $CDCl_3$) δ 8.17 (ddd, *J* = 5.0, 2.0, 0.9 Hz, 1H), 7.45 (ddd, *J* = 8.9, 7.1, 2.0 Hz, 1H), 6.57 (ddd, *J* = 7.1, 5.0, 0.9 Hz, 1H), 6.37 (dt, *J* = 8.6, 1.0 Hz, 1H), 4.47 – 4.23 (m, 1H), 3.52 (ddd, *J* = 9.8, 6.3, 5.1 Hz, 1H), 3.40 – 3.22 (m, 1H), 2.75 – 2.88 (m, 1H), 2.21 – 1.94 (m, 5H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 156.9, 148.7, 137.5, 126.7 (q, *J* = 277.8 Hz), 112.4, 106.9, 52.5 (q, *J* = 3.2 Hz), 47.4, 36.9 (q, *J* = 25.7 Hz), 30.8, 23.6; **^{19}F NMR** (282 MHz, $CDCl_3$) δ -63.3 (t, *J* = 11.3 Hz); **HRMS** (FAB) *m/z* $[M + H]^+$ calcd for $C_{11}H_{14}N_2F_3^+$: 231.1109, found: 231.1124.

***N*,4-dimethyl-*N*-(3,3,3-trifluoropropyl)aniline (3j)**



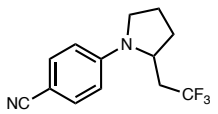
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.10 (dd, *J* = 8.9, 0.8 Hz, 2H), 6.67 (d, *J* = 8.6 Hz, 2H), 3.68 – 3.56 (m, 2H), 2.94 (s, 3H), 2.45 – 2.31 (m, 2H), 2.29 (s, 3H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 146.4, 130.3, 126.9, 126.9 (q, *J* = 276.9 Hz), 113.2, 46.4 (q, *J* = 3.5 Hz), 38.8, 30.8 (q, *J* = 27.1 Hz), 20.5; **^{19}F NMR** (282 MHz, $CDCl_3$) δ -65.2 (t, *J* = 10.9 Hz); **HRMS** (FAB) *m/z* $[M]^+$• calcd for $C_{11}H_{14}NF_3^+$: 217.1078, found: 217.1085.

4-ethyl-*N*-methyl-*N*-(3,3,3-trifluoropropyl)aniline (3k)



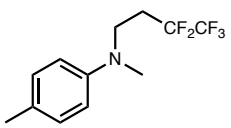
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.12 (dt, *J* = 8.8, 0.7 Hz, 2H), 6.68 (d, *J* = 8.8 Hz, 2H), 3.73 – 3.48 (m, 2H), 2.93 (s, 3H), 2.58 (q, *J* = 7.6 Hz, 2H), 2.45 – 2.26 (m, 2H), 1.22 (t, *J* = 7.6 Hz, 3H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 146.6, 133.5, 129.1, 126.9 (q, *J* = 276.9 Hz), 113.1, 46.4 (q, *J* = 3.5 Hz), 38.9, 30.9 (q, *J* = 27.1 Hz), 28.1, 16.2; **^{19}F NMR** (282 MHz, $CDCl_3$) δ -65.2 (d, *J* = 22.0 Hz); **HRMS** (FAB) *m/z* $[M]^+$• calcd for $C_{12}H_{16}NF_3^+$: 231.1298, found: 231.1326.

4-(2-(2,2,2-trifluoroethyl)pyrrolidin-1-yl)benzonitrile (3l)



^{1}H NMR (400 MHz, $CDCl_3$) δ 7.48 (d, *J* = 9.0 Hz, 2H), 6.54 (d, *J* = 8.9 Hz, 2H), 4.14 – 3.98 (m, 1H), 3.57 – 3.38 (m, 1H), 3.36 – 3.10 (m, 1H), 2.37 – 2.49 (m, 1H), 2.03 – 2.17 (m, 5H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 148.9, 134.1, 126.2 (q, *J* = 277.6 Hz), 120.8, 112.2, 98.5, 53.0 (q, *J* = 2.9 Hz), 48.0, 36.5 (q, *J* = 26.0 Hz), 31.3, 23.1; **^{19}F NMR** (282 MHz, $CDCl_3$) δ -64.0 (t, *J* = 10.9 Hz); **HRMS** (FAB) *m/z* $[M + H]^+$ calcd for $C_{13}H_{14}N_2F_3^+$: 255.1104, found: 255.1103.

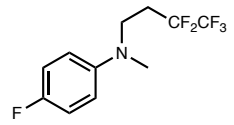
***N*,4-dimethyl-*N*-(3,3,4,4,4-pentafluorobutyl)aniline (5a)**



^{1}H NMR (400 MHz, $CDCl_3$) δ 7.08 (dd, *J* = 8.8, 0.8 Hz, 2H), 6.65 (d, *J* = 8.7 Hz, 2H), 3.88 – 3.50 (m, 2H), 2.92 (s, 3H), 2.34 – 2.16 (m, 2H), 2.27 (s, 3H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 146.3, 130.3, 127.0, 118.1 (tq, *J* = 249.0, 36.2 Hz), 113.2, 45.1(t,

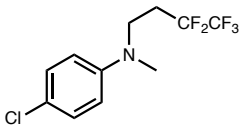
J = 4.2 Hz), 38.8, 27.3 (t, J = 21.6 Hz), 20.6. The ^{13}C resonance corresponds to the -CF_3 group was not well resolved and the signal for this carbon is not reported; **^{19}F NMR** (282 MHz, $CDCl_3$) δ -85.6, -118.3 (t, J = 18.6 Hz); **HRMS** (ESI) *m/z* $[M + H]^+$ calcd for $C_{12}H_{15}NF_5^+$: 268.1125, found: 268.1123.

4-fluoro-*N*-methyl-*N*-(3,3,4,4,4-pentafluorobutyl)aniline (5b)



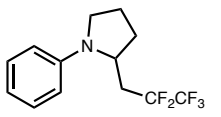
^{1}H NMR (400 MHz, $CDCl_3$) δ 6.97 (dd, J = 9.3, 8.3 Hz, 2H), 6.66 (dd, J = 9.2, 4.3 Hz, 2H), 3.70 – 3.57 (m, 2H), 2.91 (s, 3H), 2.41 – 2.12 (m, 2H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 156.2 (d, J = 236.3 Hz), 145.1 (d, J = 1.9 Hz), 119.3 (qt, J = 285.3, 36.1 Hz), 116.2 (d, J = 22.2 Hz), 115.8 (tq, J = 253.0, 38.2 Hz), 114.2 (d, J = 7.41 Hz), 45.6 (t, J = 4.4 Hz), 39.1, 27.4 (t, J = 21.7 Hz); **^{19}F NMR** (282 MHz, $CDCl_3$) δ -85.6, -118.2 (t, J = 18.4 Hz), -128.1 (tt, J = 8.4, 4.2 Hz); **HRMS** (FAB) *m/z* $[M + H]^+$ calcd for $C_{11}H_{12}NF_6^+$: 272.0874, found: 272.0883.

4-chloro-*N*-methyl-*N*-(3,3,4,4,4-pentafluorobutyl)aniline (5c)



^{1}H NMR (400 MHz, $CDCl_3$) δ 7.20 (d, J = 9.1 Hz, 2H), 6.63 (d, J = 9.1 Hz, 2H), 3.79 – 3.54 (m, 2H), 2.94 (s, 3H), 2.46 – 2.14 (m, 2H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 146.8, 129.6, 122.7, 114.0, 45.1 (t, J = 4.1 Hz), 38.8, 27.5 (t, J = 21.7 Hz). The ^{13}C resonance corresponds to the -CF_2CF_3 group was not well resolved and the signals for these carbons are not reported; **^{19}F NMR** (282 MHz, $CDCl_3$) δ -85.6, -118.3 (t, J = 18.3 Hz); **HRMS** (FAB) *m/z* $[M]^+$• calcd for $C_{11}H_{11}NClF_5^+$: 287.0500, found: 287.0510.

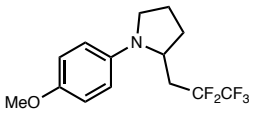
2-(2,2,3,3,3-pentafluoropropyl)-1-phenylpyrrolidine (5d)



^{1}H NMR (300 MHz, $CDCl_3$) δ 7.32 – 7.21 (m, 2H), 6.73 (t, J = 7.3 Hz, 1H), 6.60 (d, J = 7.8 Hz, 2H), 4.23 – 4.11 (m, 1H), 3.56 – 3.36 (m, 1H), 3.18 (td, J = 8.8, 6.7 Hz, 1H), 2.36 – 2.54 (m, 1H), 2.20 – 1.82 (m, 5H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 146.4, 129.9, 116.8, 112.2, 51.9, 48.1, 33.5 (t, J = 20.3 Hz), 32.0, 23.4. The ^{13}C resonance corresponds to the -CF_2CF_3 group was not well resolved and the signals for these carbons

are not reported; **^{19}F NMR** (282 MHz, , $CDCl_3$) δ -85.9, -100.7 – -126.1 (m); **HRMS** (FAB) *m/z* $[M]^+\bullet$ calcd for $C_{13}H_{14}NF_5^+\bullet$: 279.1046, found: 279.1035.

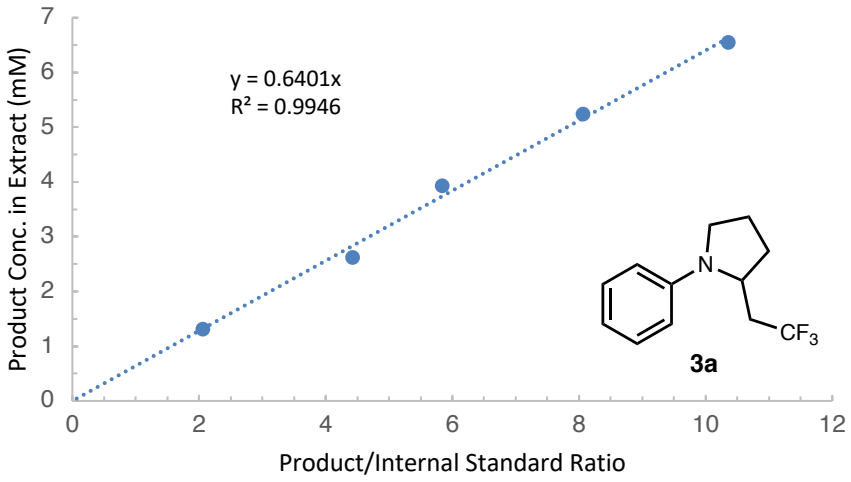
1-(4-methoxyphenyl)-2-(2,2,3,3,3-pentafluoropropyl)pyrrolidine (5e)

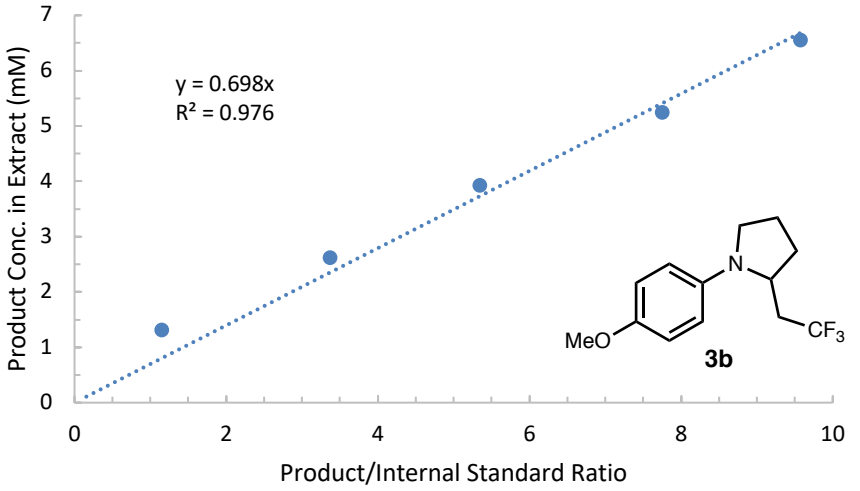


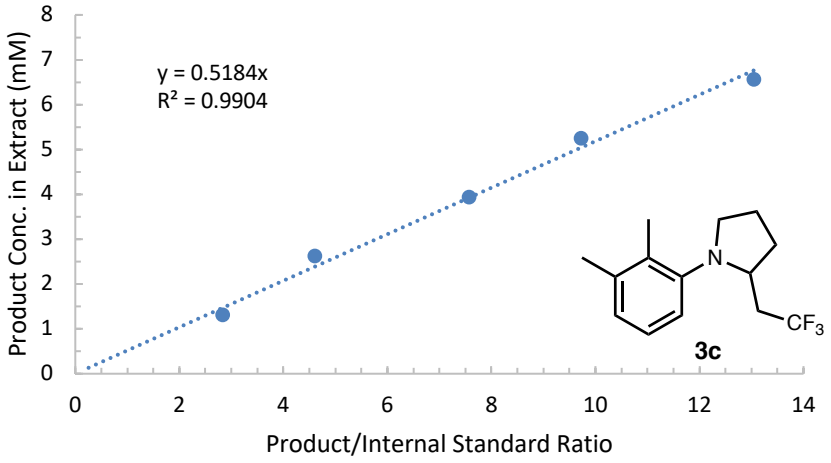
^{1}H NMR (400 MHz, $CDCl_3$) δ 6.88 (d, *J* = 9.0 Hz, 2H), 6.55 (d, *J* = 8.4 Hz, 2H), 4.09 (t, *J* = 8.6 Hz, 1H), 3.77 (s, 3H), 3.47 – 3.35 (m, 2H), 3.13 (q, *J* = 8.0 Hz, 1H), 2.60 – 2.33 (m, 1H), 2.21 – 2.10 (m, 1H), 2.12 – 1.81 (m, 3H); **^{13}C NMR** (101 MHz, $CDCl_3$) δ 151.7, 141.3, 115.6, 113.1, 56.2, 52.3, 48.8, 33.9 (q, *J* = 20.8, 19.6 Hz), 32.2, 23.5. The ^{13}C resonance corresponds to the $-CF_2CF_3$ group was not well resolved and the signals for these carbons are not reported; **^{19}F NMR** (282 MHz, $CDCl_3$) δ 85.9, -117.5 (qdd, *J* = 180.0, 28.8, 9.8 Hz); **HRMS** (FAB) *m/z* $[M]^+\bullet$ calcd for $C_{14}H_{16}ONF_5^+$: 309.1152, found: 309.1162.

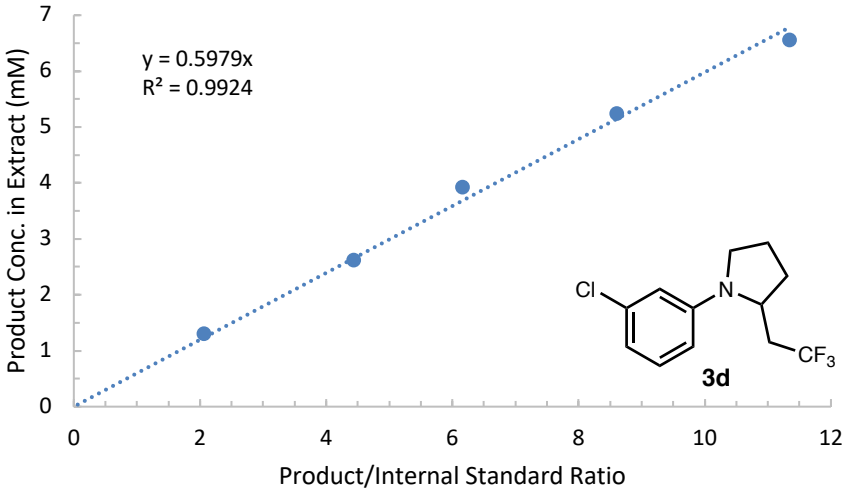
A.6 Product GC-MS Calibration Curves

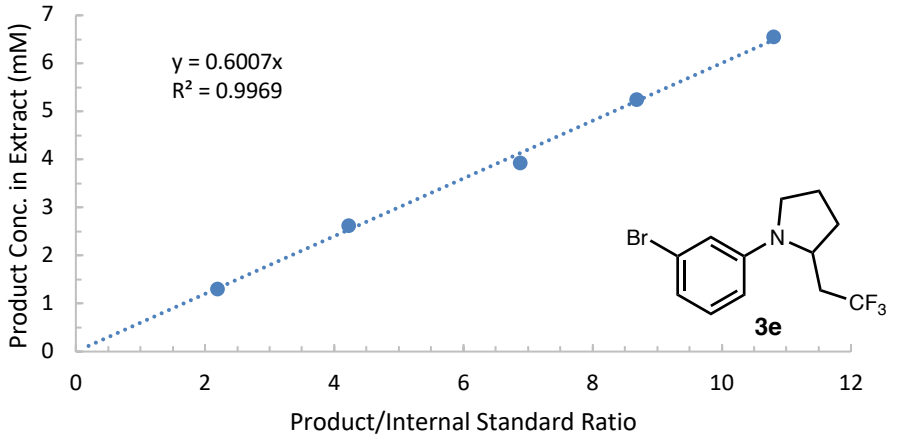
Product formation of **3a-3l**, **5a-5c** in enzymatic reactions was quantified by GC-MS based on standard curves. To determine the standard calibration curves, stock solutions of chemically synthesized authentic products were prepared at various concentrations (0.5–7 mM in 4:6 hexanes/EtOAc) with added internal standard 1,2,3-trimethoxybenzene (0.66 mM final concentration in the stock solutions). All data points represent the average of duplicate runs. The standard curves plot product concentration in mM (y-axis) against the ratio of product area to internal standard area on GC-MS (x-axis). For **5d** and **5e**, 15 analytical scale reactions were combined to measure the product formation and total turnover numbers by ^{19}F NMR with fluorobenzene as the internal standard.

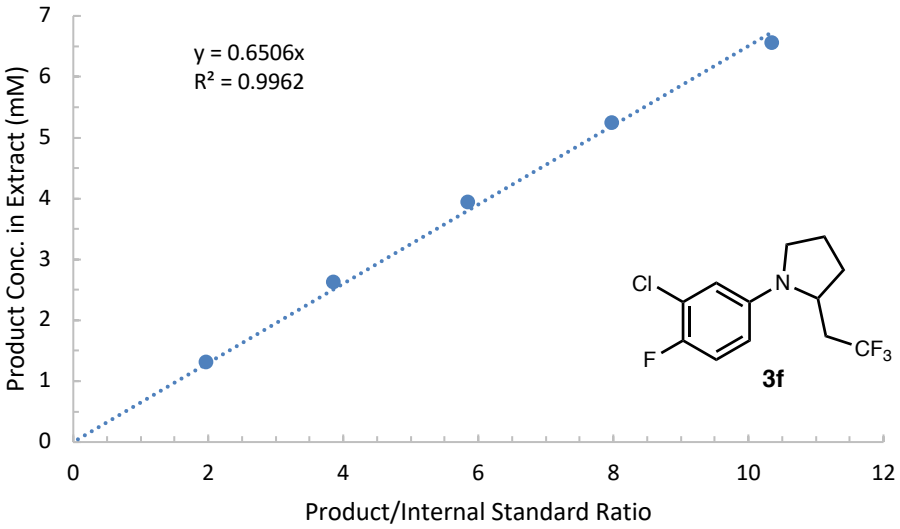


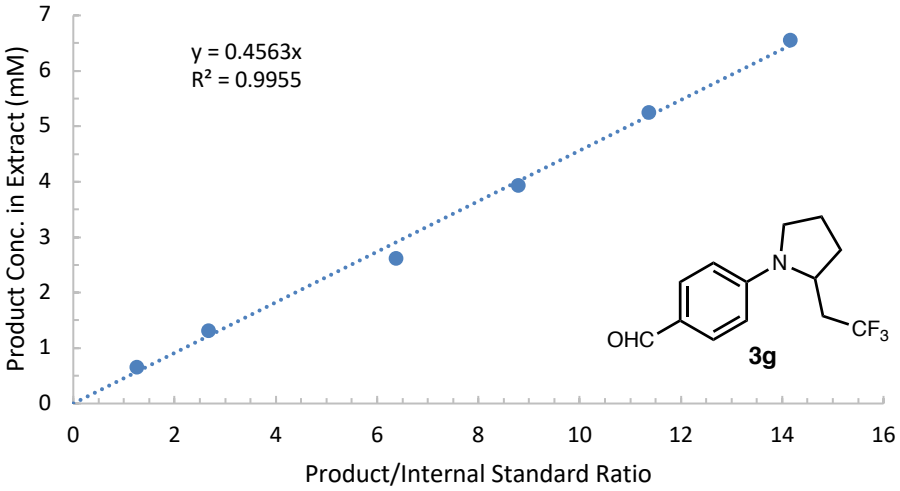

y = 0.698x
R² = 0.976
Product Conc. in Extract (mM)
Product/Internal Standard Ratio
MeO
CF3
3b

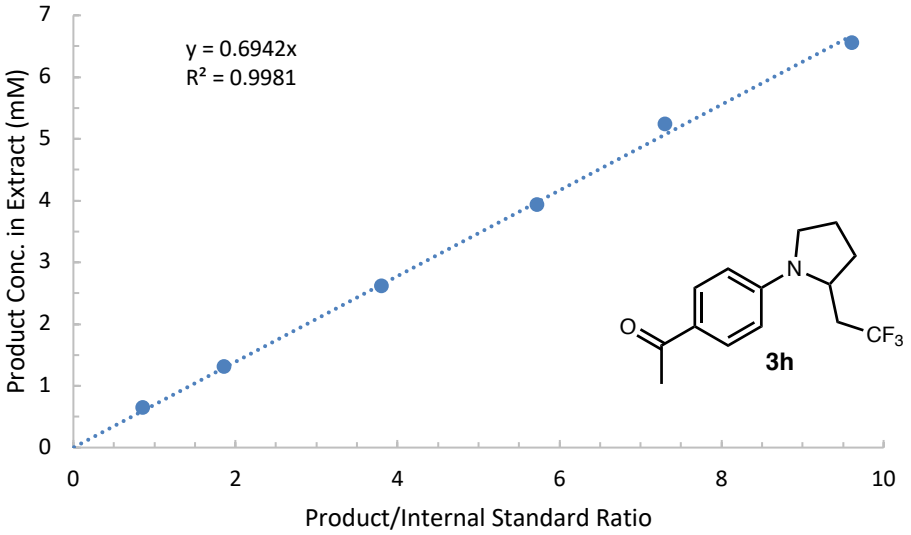

y = 0.5184x
R² = 0.9904
Product Conc. in Extract (mM)
Product/Internal Standard Ratio
CF3
3c

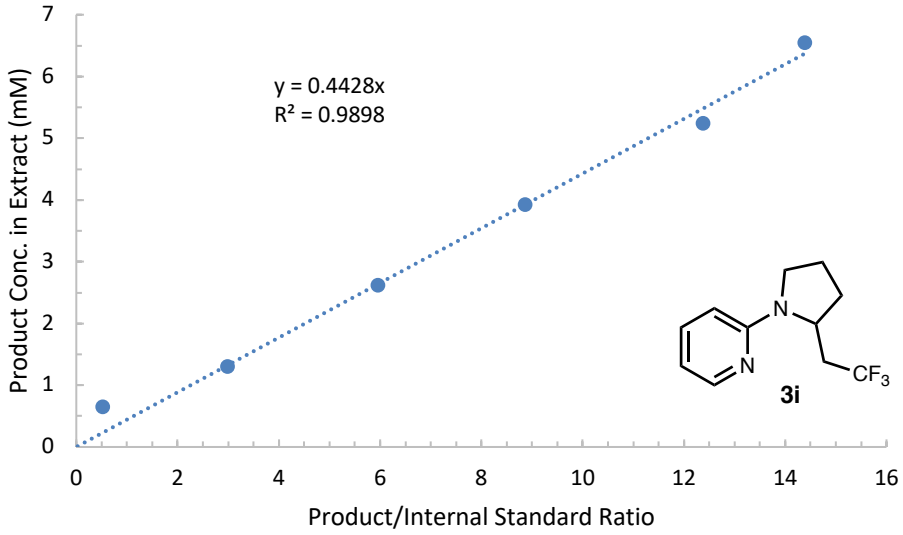

y = 0.5979x
R² = 0.9924
Product Conc. in Extract (mM)
Product/Internal Standard Ratio
Cl
CF3
3d

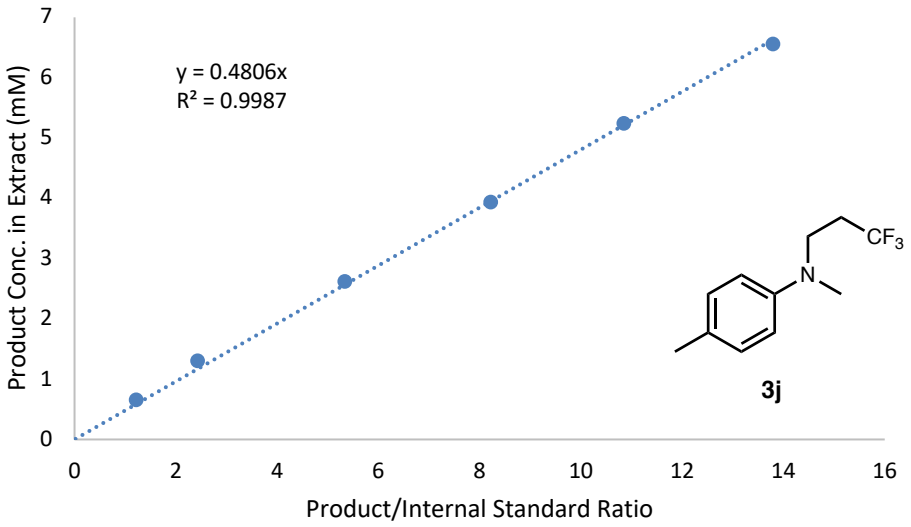

y = 0.6007x
R² = 0.9969
Product Conc. in Extract (mM)
Product/Internal Standard Ratio
Br
N
CF3
3e

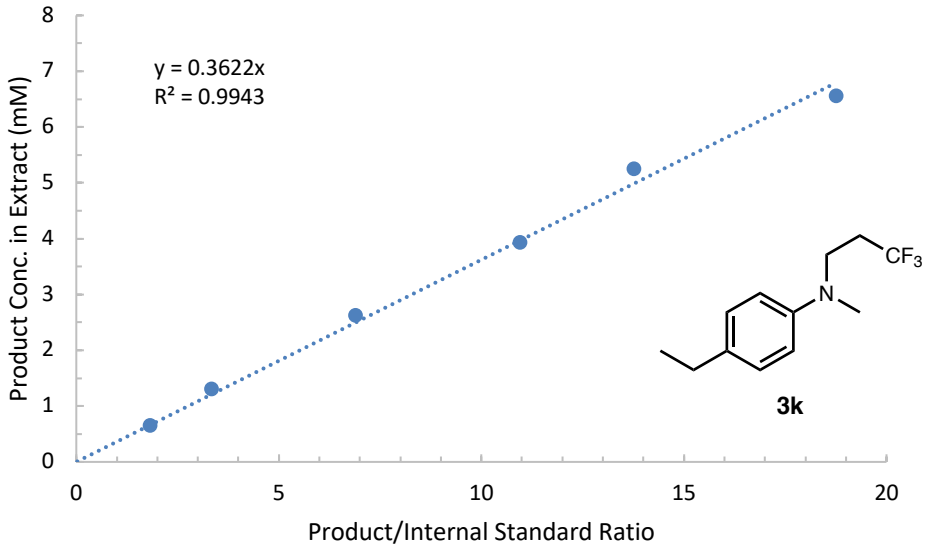

y = 0.6506x
R² = 0.9962
Product Conc. in Extract (mM)
Product/Internal Standard Ratio
Cl
F
N
CF3
3f

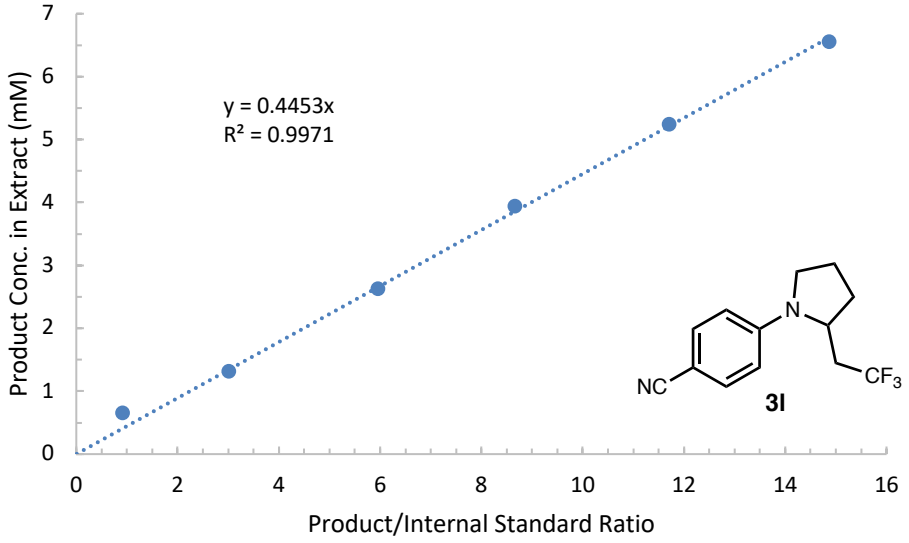

y = 0.4563x
R² = 0.9955
Product Conc. in Extract (mM)
Product/Internal Standard Ratio
OHC
N
CF3
3g

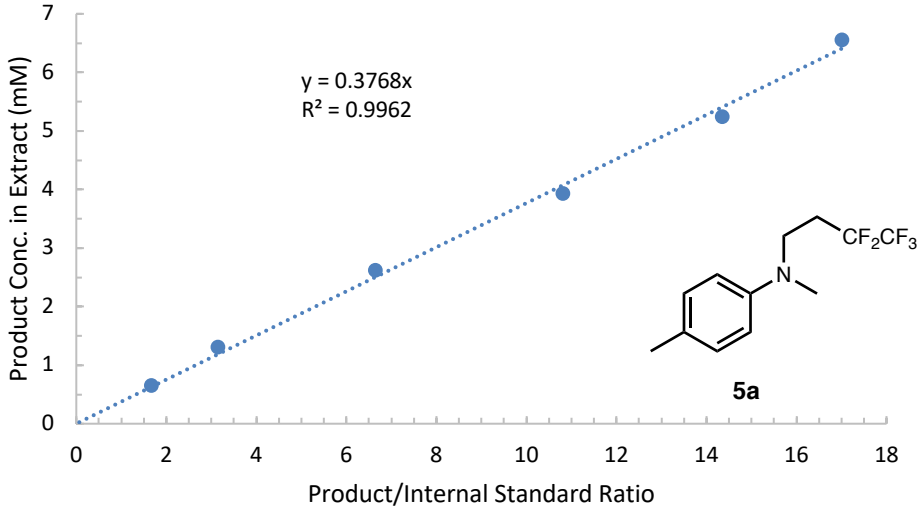

y = 0.6942x
R² = 0.9981
Product Conc. in Extract (mM)
Product/Internal Standard Ratio
0
1
2
3
4
5
6
7
0
2
4
6
8
10
O
N
CF3
3h

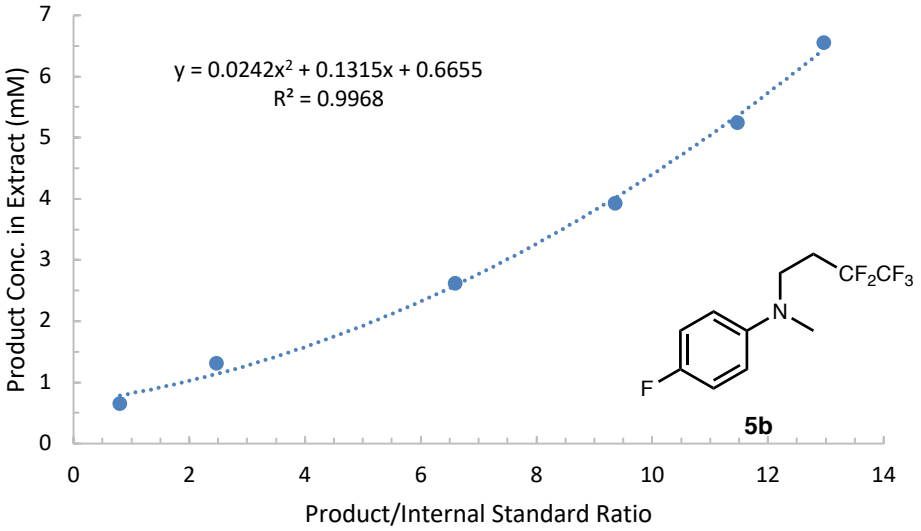

y = 0.4428x
R² = 0.9898
Product Conc. in Extract (mM)
Product/Internal Standard Ratio
0
1
2
3
4
5
6
7
0
2
4
6
8
10
12
14
16
N
N
CF3
3i

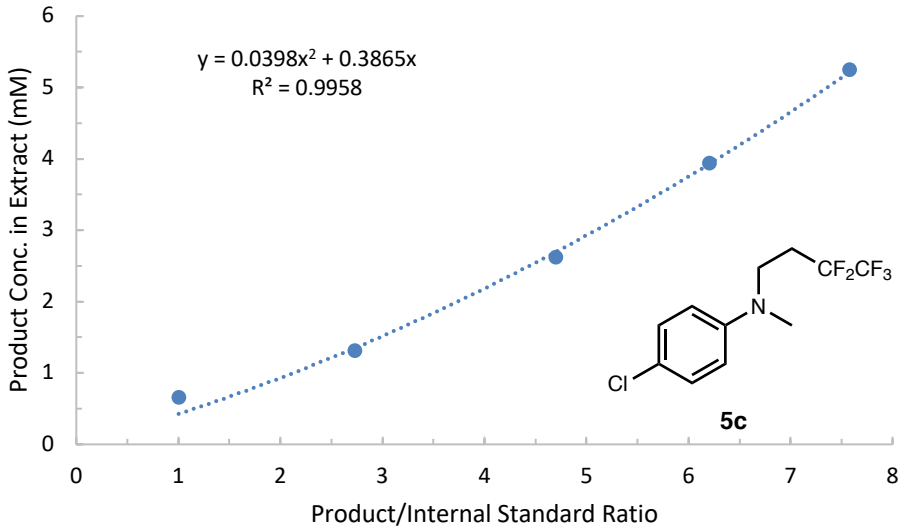

y = 0.4806x
R² = 0.9987
Product Conc. in Extract (mM)
Product/Internal Standard Ratio
0
1
2
3
4
5
6
7
0
2
4
6
8
10
12
14
16
CF3
N
3j


y = 0.3622x
R² = 0.9943
Product Conc. in Extract (mM)
Product/Internal Standard Ratio
0
1
2
3
4
5
6
7
8
0
5
10
15
20
CF3
N
3k


y = 0.4453x
R² = 0.9971
Product Conc. in Extract (mM)
Product/Internal Standard Ratio
0
1
2
3
4
5
6
7
0
2
4
6
8
10
12
14
16
N
NC
CF3
3l

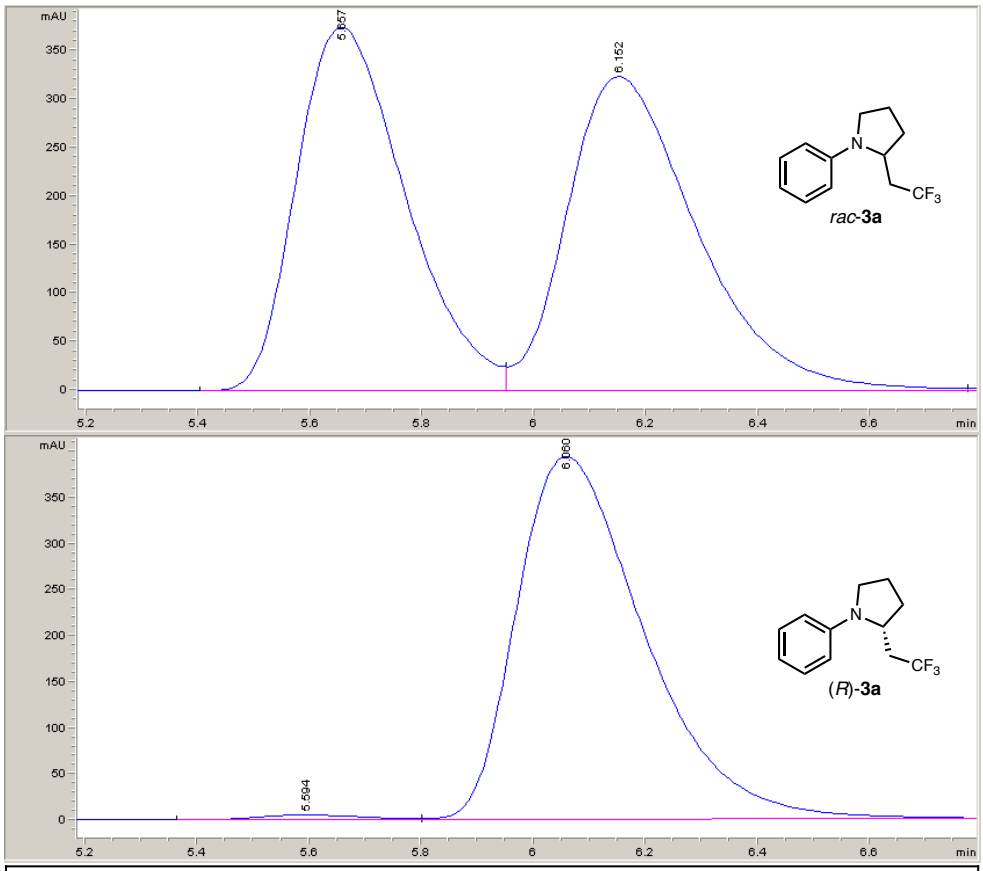

y = 0.3768x
R² = 0.9962
Product Conc. in Extract (mM)
Product/Internal Standard Ratio
0
1
2
3
4
5
6
7
0
2
4
6
8
10
12
14
16
18
N
CF2CF3
5a


y = 0.0242x² + 0.1315x + 0.6655
R² = 0.9968
Product Conc. in Extract (mM)
Product/Internal Standard Ratio
0
1
2
3
4
5
6
7
0
2
4
6
8
10
12
14
CF2CF3
N
F
5b

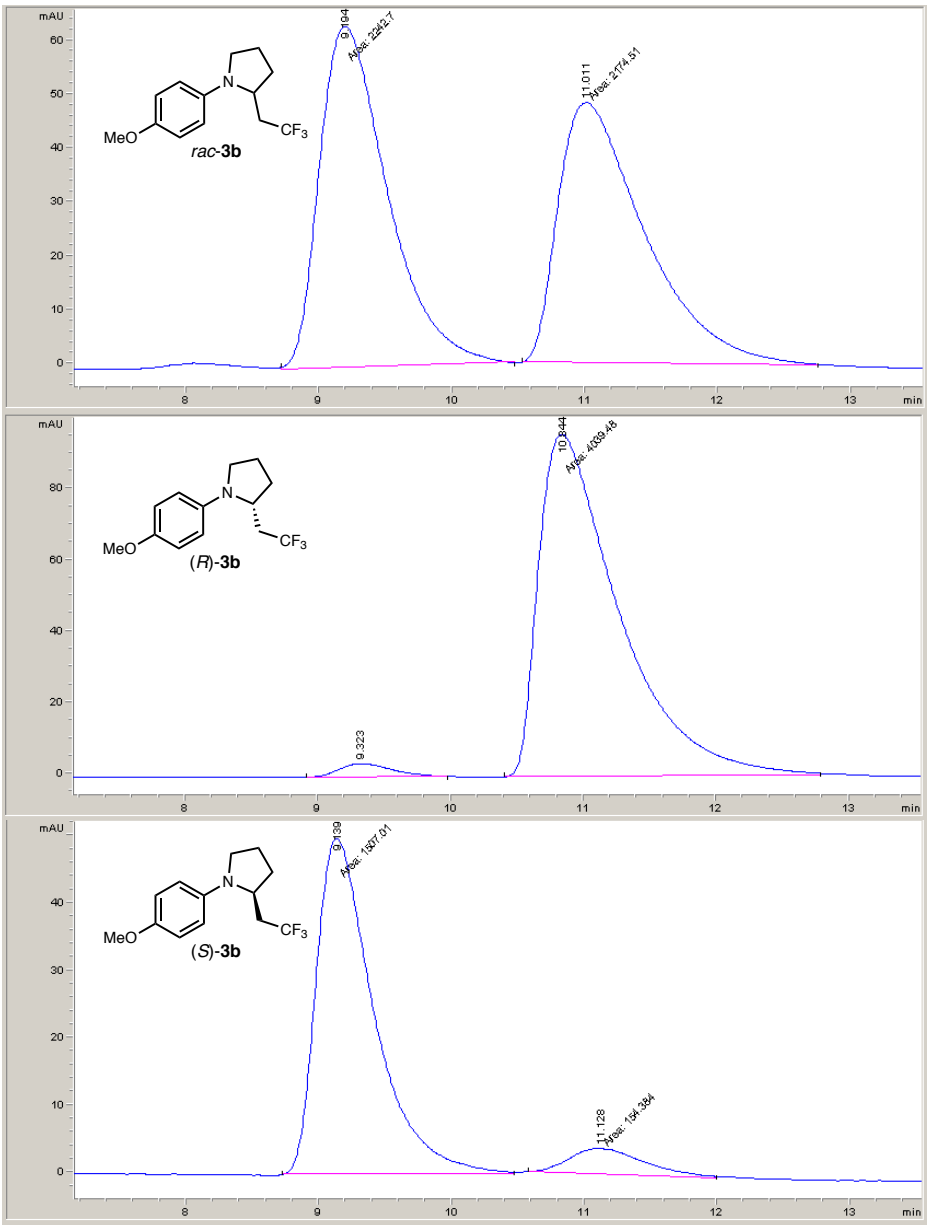

y = 0.0398x² + 0.3865x
R² = 0.9958
Product Conc. in Extract (mM)
Product/Internal Standard Ratio
0
1
2
3
4
5
6
0
1
2
3
4
5
6
7
8
CF2CF3
N
Cl
5c

A.7 Determination of Enantioselectivity

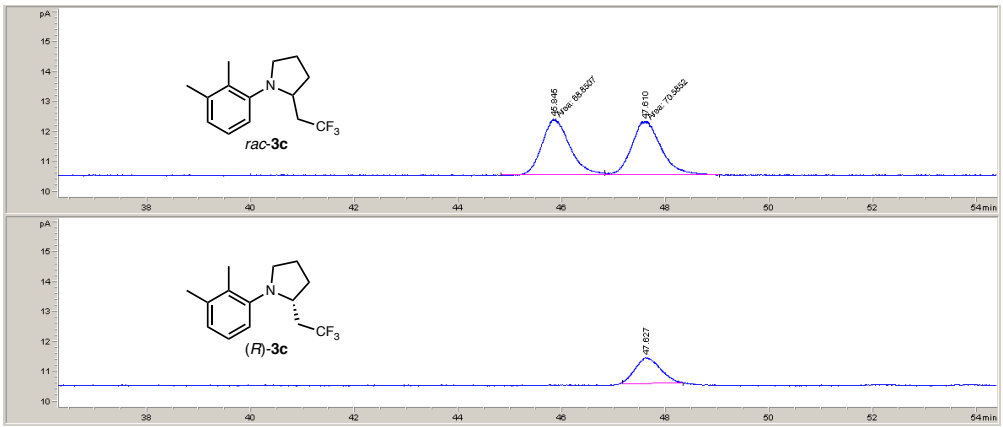
All ee values of enzymatically synthesized products were determined using normal-phase chiral HPLC or chiral GC. The absolute configuration of product **3a** synthesized by P411-PFA variant was determined to be *R* based on the X-ray crystallography data (see **Section A.8** for details). Absolute configurations of other C–H trifluoroethylation products (**3b**-**3i**, **3l**) were inferred by analogy, assuming the facial selectivity of the trifluoroethyl diazo reagents from which these products were made remains the same as that of **3a**. The absolute configuration of product **5e** synthesized by P411-PFA variant were determined to be *S* based on the X-ray crystallography data (see **Section A.8** for details). Absolute configuration of **5d** were inferred by analogy.



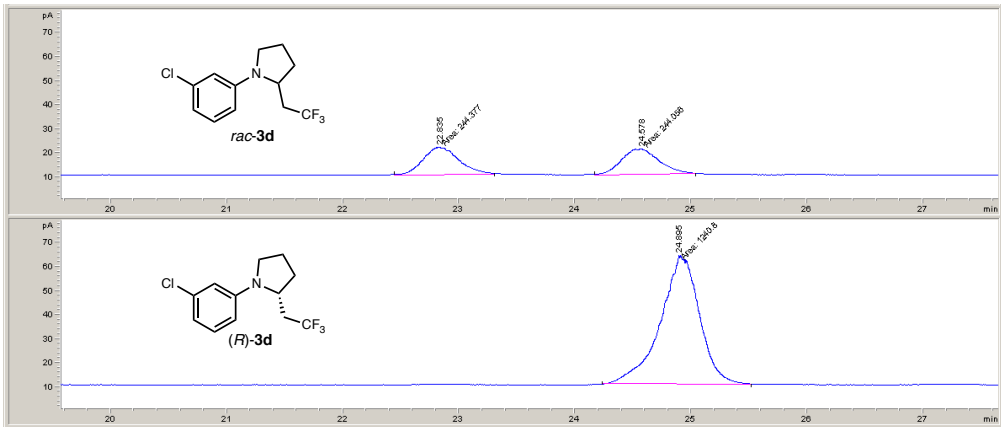
Chiralpak AD (4.6 × 250 nm, 5µm), 100% Hexanes, 1.0 mL/min, 254 nm					
rac-**3a**			(*R*)-**3a** from P411-PFA		
Retention time (min)	Area (mAu)	Area%	Retention time (min)	Area (mAu)	Area%
5.657	5071	49.2	5.594	68	1.1
6.152	5228	50.8	6.060	6253	98.9



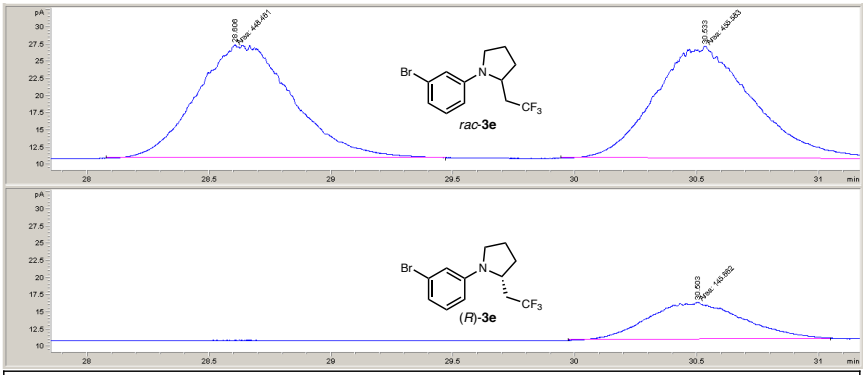
rac-**3b**			Chiralpak IA, 100% Hexanes, 1.5 mL/min, 254 nm		
Retention time (min)	Area (mAu)	Area%			
9.194	2246	50.2			
11.011	2225	49.8			
(*R*)-**3b**			(*S*)-**3b**		
Retention time (min)	Area (mAu)	Area%	Retention time (min)	Area (mAu)	Area%
9.323	108	2.5	9.139	1511	91.0
10.844	4122	97.5	11.128	149	9.0



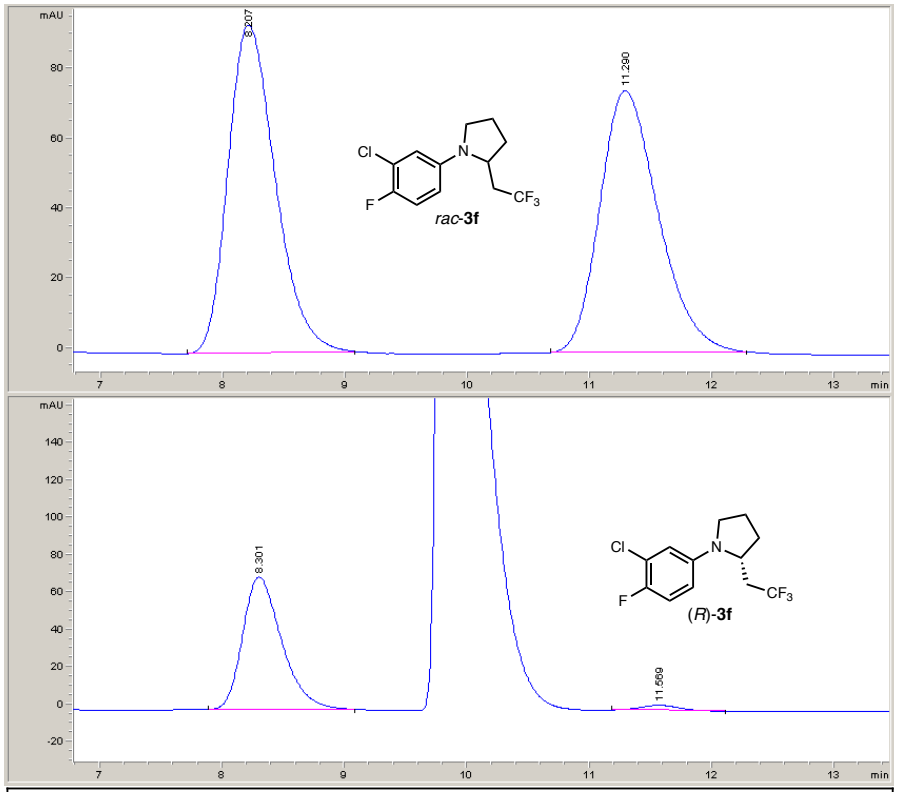
Astec CHIRALDEX G-TA (30 m × 0.25 mm × 0.12 µm), 100 ℃					
rac-**3c**			(*R*)-**3c** from P411-PFA		
Retention time (min)	Area (pA)	Area%	Retention time (min)	Area (pA)	Area%
45.845	68.9	49.4	-	-	-
47.610	70.6	50.6	47.627	55.4	>99



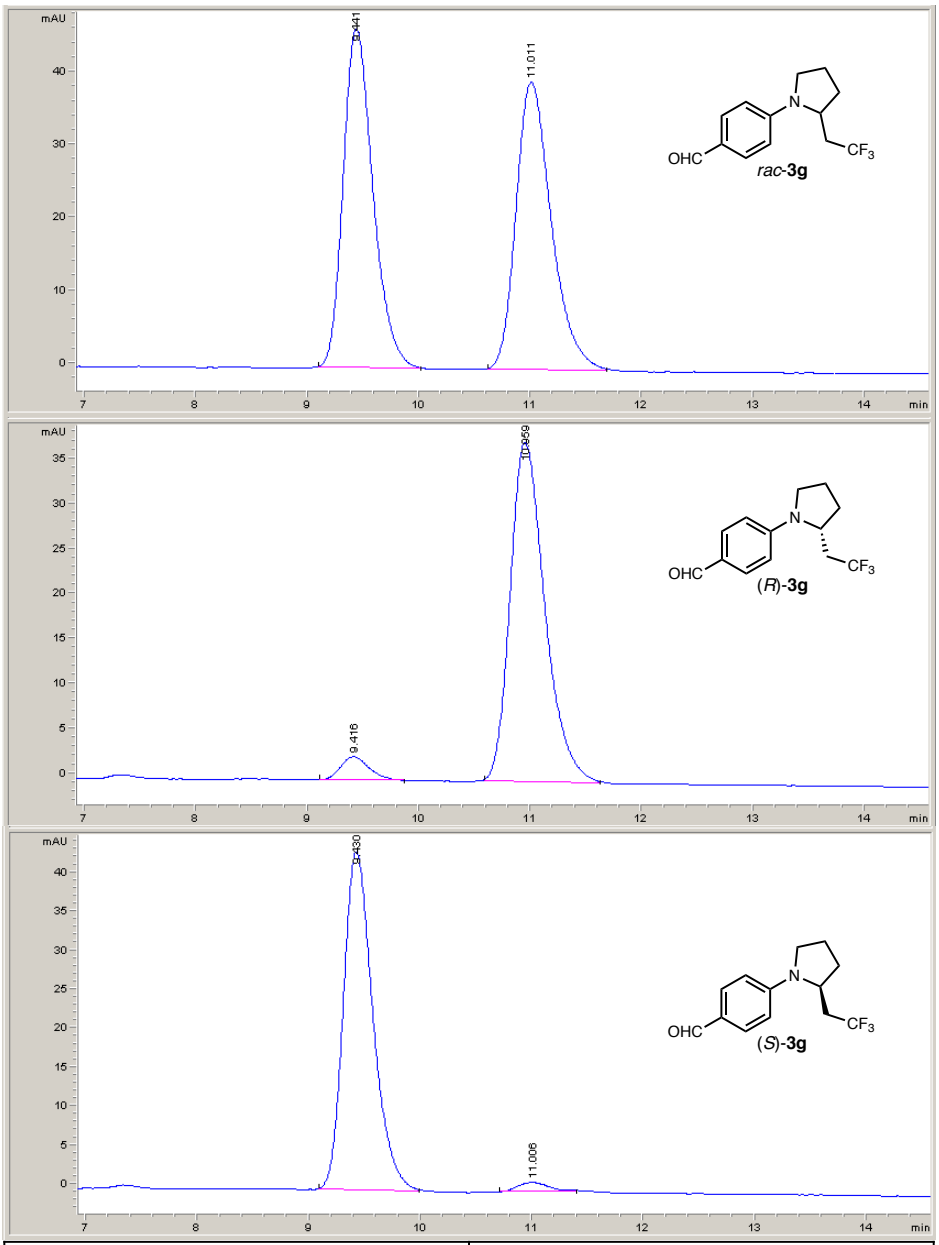
Astec CHIRALDEX G-TA (30 m × 0.25 mm × 0.12 µm), 140 ℃					
rac-**3d**			(*R*)-**3d** from P411-PFA		
Retention time (min)	Area (mAu)	Area%	Retention time (min)	Area (mAu)	Area%
22.835	244.4	50.0	-	-	-
24.578	244.1	50.0	24.895	1240.8	>99



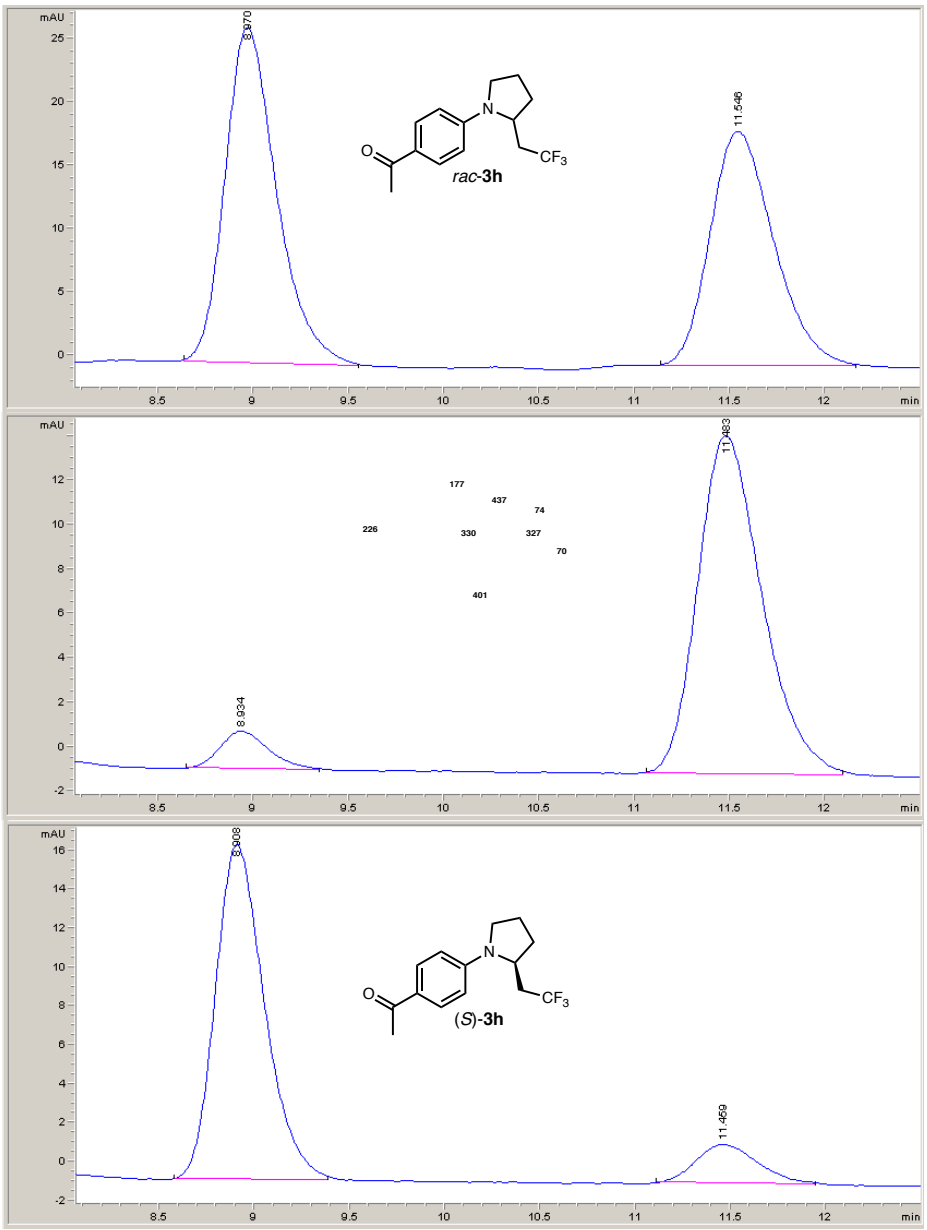
Astec CHIRALDEX G-TA (30 m × 0.25 mm × 0.12 µm), 145 ℃					
rac-**3e**			(*R*)-**3e** from P411-PFA		
Retention time (min)	Area (mAu)	Area%	Retention time (min)	Area (mAu)	Area%
28.606	448	49.6	-	-	-
30.533	456	50.4	30.503	146	>99



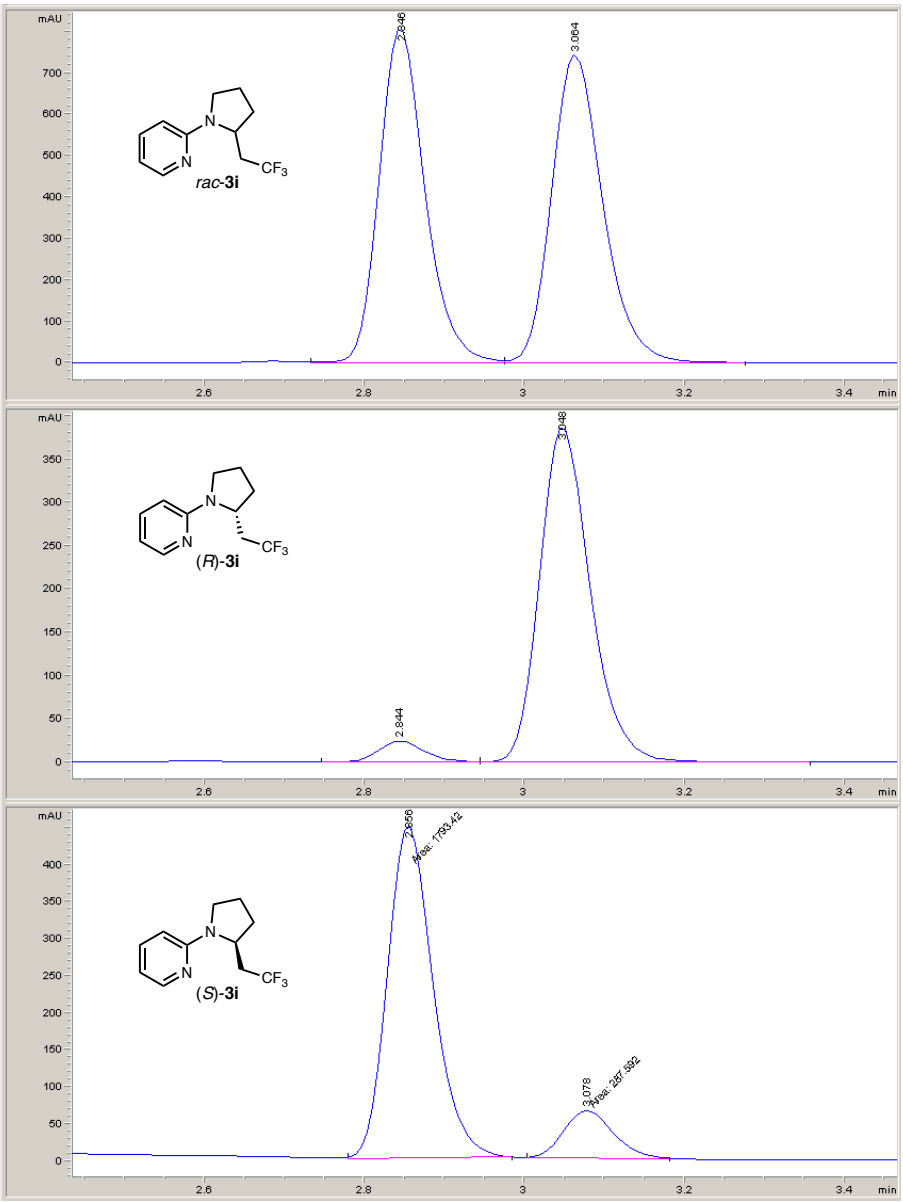
Chiralpak OJ-H (4.6 × 250 nm, 5µm), 100% Hexanes, 1.5 mL/min, 254 nm					
rac-**3f**			(*R*)-**3f** from P411-PFA		
Retention time (min)	Area (mAu)	Area%	Retention time (min)	Area (mAu)	Area%
8.207	2537	50.3	8.301	1579	96.1
11.290	2508	49.7	11.569	64	3.9



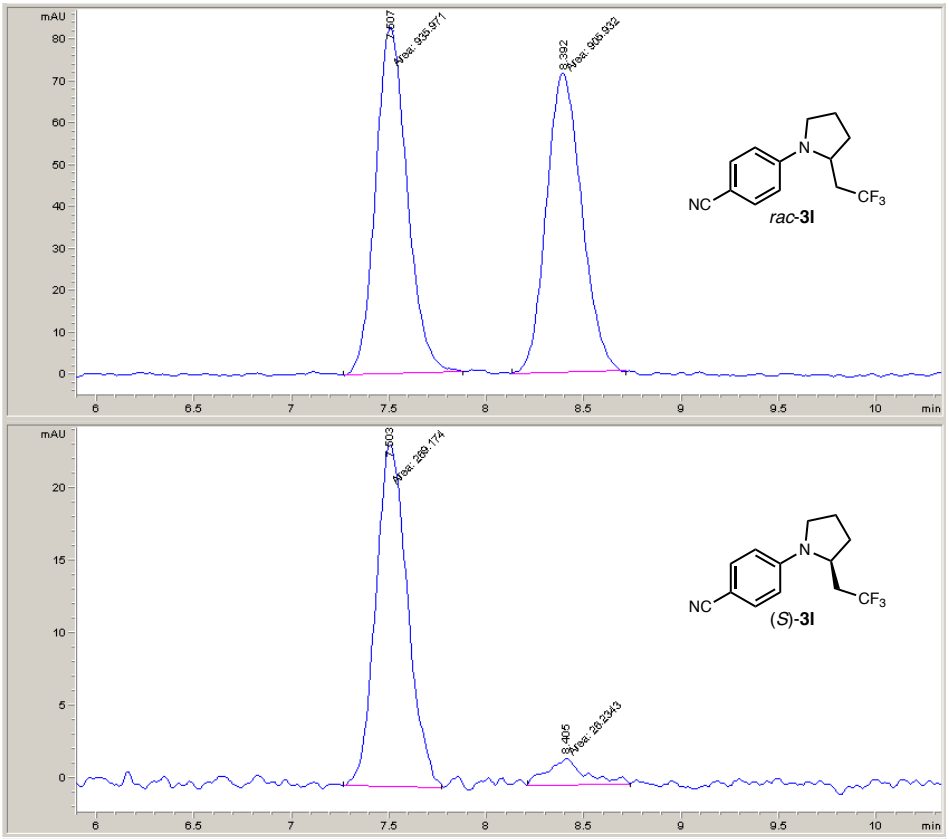
rac-**3g**					
Retention time (min)	Area (mAu)	Area%	Chiralpak IC (4.6 × 250 nm, 5µm), 30% IPA/hexanes, 1.5 mL/ min, 235 nm		
9.441	869	49.9			
11.011	874	50.1			
(*R*)-**3g** from P411-PFA			(*S*)-**3g** from P411-PFA-(*S*)		
Retention time (min)	Area (mAu)	Area%	Retention time (min)	Area (mAu)	Area%
9.416	46	5.4	9.430	799	97.2
10.959	820	94.6	11.006	23	2.8



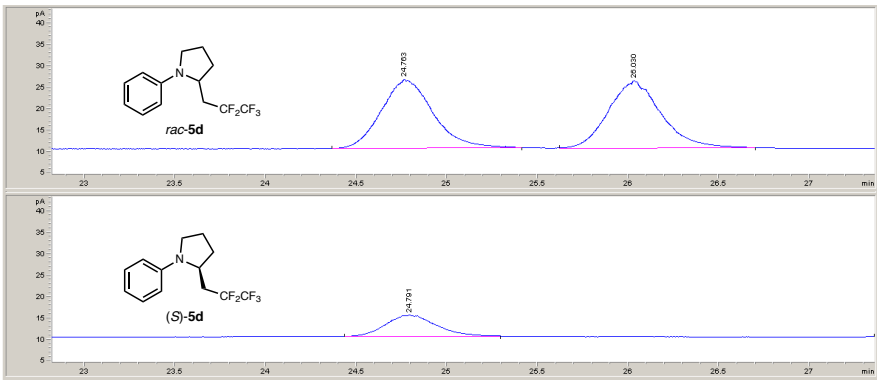
rac-**3h**			Chiralpak IC (4.6 × 250 nm, 5µm), 30% IPA/hexanes, 1.5 mL/ min, 235 nm
Retention time (min)	Area (mAu)	Area%	
8.970	478	52.7	
11.546	430	47.3	
(*R*)-**3h** from P411-PFA			(*S*)-**3h** from P411-PFA-(*S*)
Retention time (min)	Area (mAu)	Area%	Retention time (min) Area (mAu) Area%
8.934	30	7.8	8.908 309 87.1
11.483	357	92.2	11.459 46 12.9



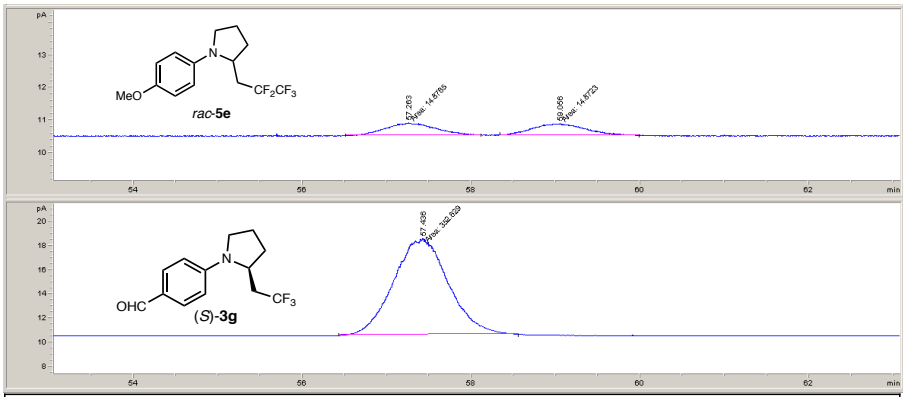
rac-**3i**			Chiralpak IC (4.6 × 250 nm, 5µm), 5% IPA/hexanes, 1.5 mL/min, 254 nm		
Retention time (min)	Area (mAu)	Area%			
2.846	3219	49.6			
3.064	3270	50.4			
(*R*)-**3i** from P411-PFA			(*S*)-**3i** from P411-PFA-(*S*)		
Retention time (min)	Area (mAu)	Area%	Retention time (min)	Area (mAu)	Area%
2.844	25	5.9	2.856	1793	86.2
3.048	389	94.1	3.078	288	13.8



Chiralpak IC (4.6 × 250 nm, 5µm), 30% IPA/hexanes, 1.5 mL/min, 254 nm					
rac-**3l**			(*S*)-**3l** from P411-PFA-(*S*)		
Retention time (min)	Area (mAu)	Area%	Retention time (min)	Area (mAu)	Area%
7.507	936	50.8	7.503	269	91.2
8.392	906	49.2	8.405	26	8.8



Astec CHIRALDEX G-TA (30 m × 0.25 mm × 0.12 μm), 110 ℃					
rac-**5d**			(*S*)-**5d** from P411-PFA		
Retention time (min)	Area (mAu)	Area%	Retention time (min)	Area (mAu)	Area%
24.763	317		24.791	95.6	>99
26.030	314		-	-	-



Astec CHIRALDEX G-TA (30 m × 0.25 mm × 0.12 μm), 120 ℃					
rac-**5e**			(*S*)-**5e** from P411-PFA		
Retention time (min)	Area (mAu)	Area%	Retention time (min)	Area (mAu)	Area%
57.263	14.8765	50.0	57.436	352.629	>99%
59.056	14.8723	50.0	-	-	-

A.8 X-Ray Crystallography and Assignments of Absolute Configuration

Crystal growth

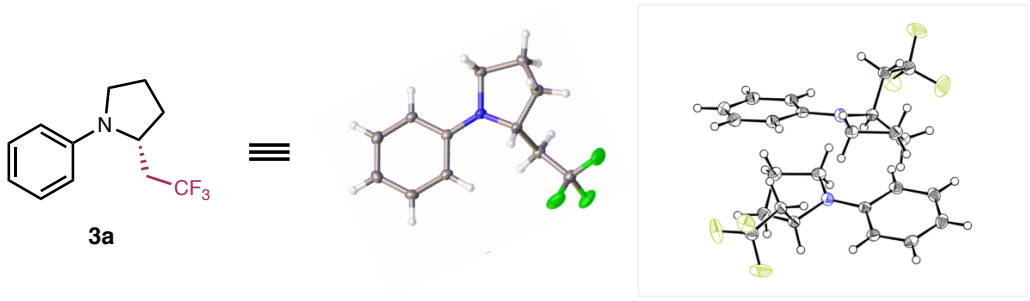
For product **3a**, 10 mg of pure enzymatic product was dissolved in 0.5 mL *n*-hexane in a 2 mL vial. The vial was loosely capped and left undisturbed at ambient temperature. After two days, the solvent was slowly evaporated and white, needle-like crystals were formed. The single crystals of product **5e** were obtained in a similar manner with *n*-pentane as the solvent.

Refinement details

Crystals were mounted on a polyimide MiTeGen loop with STP Oil Treatment and placed under a nitrogen stream. Low temperature (100K) X-ray data were collected with a Bruker AXS D8 VENTURE KAPPA diffractometer running at 50 kV and 1mA (Cu K_α = 1.54178 Å; PHOTON II CPAD detector and Helios focusing multilayer mirror optics). All diffractometer manipulations, including data collection, integration, and scaling were carried out using the Bruker APEX3 software. An absorption correction was applied using SADABS. The space group was determined, and the structure solved by intrinsic phasing using XT. Refinement was full-matrix least squares on F^2 using XL. All non-hydrogen atoms were refined using anisotropic displacement parameters. Hydrogen atoms were placed in idealized positions and refined using a riding model. The isotropic displacement parameters of all hydrogen atoms were fixed at 1.2 times (1.5 times for methyl groups) the U_{eq} value of the bonded atom.

Special refinement details

Both compounds crystallize in the monoclinic space group $P2_1$. For compound **3a**, there are two molecules in the asymmetric unit; for compound **5e**, there is one.



3a

Identification code	v18648
Empirical formula	$C_{12}H_{14}F_3N$
Formula weight	229.24
Temperature/K	100.0
Crystal system	monoclinic
Space group	$P2_1$
a/Å	8.3312(11)
b/Å	17.574(2)
c/Å	8.3790(11)
α/°	90
β/°	116.975(5)
γ/°	90
Volume/Å^3	1093.3(3)
Z	4
ρ_{calc}g/cm^3	1.393
μ/mm^{-1}	1.008
F(000)	480.0
Crystal size/mm^3	0.28 × 0.21 × 0.15
Radiation	CuKα (λ = 1.54178)
2Θ range for data collection/°	10.066 to 161.532
Index ranges	$-10 \leq h \leq 10$, $-22 \leq k \leq 22$, $-10 \leq l \leq 10$
Reflections collected	50963
Independent reflections	4683 [R_{int} = 0.0463, R_{sigma} = 0.0198]
Data/restraints/parameters	4683/1/290
Goodness-of-fit on F^2	1.045
Final R indexes [I>=2σ (I)]	R_1 = 0.0279, wR_2 = 0.0705
Final R indexes [all data]	R_1 = 0.0279, wR_2 = 0.0705
Largest diff. peak/hole / e Å^{-3}	0.21/-0.19
Flack parameter	0.07(2)

Datablock: v18648

Bond precision: C-C = 0.0027 A Wavelength=1.54178

Cell: a=8.3312(11) b=17.574(2) c=8.3790(11)
alpha=90 beta=116.975(5) gamma=90
Temperature: 100 K

	Calculated	Reported
Volume	1093.3(2)	1093.3(3)
Space group	P 21	P 1 21 1
Hall group	P 2yb	P 2yb
Moiety formula	C12 H14 F3 N	C12 H14 F3 N
Sum formula	C12 H14 F3 N	C12 H14 F3 N
Mr	229.24	229.24
Dx,g cm-3	1.393	1.393
Z	4	4
Mu (mm-1)	1.008	1.008
F000	480.0	480.0
F000'	481.80	
h,k,lmax	10,22,10	10,22,10
Nref	4788[2474]	4683
Tmin,Tmax	0.776,0.860	0.785,1.000
Tmin'	0.754	

Correction method= # Reported T Limits: Tmin=0.785 Tmax=1.000
AbsCorr = MULTI-SCAN

Data completeness= 1.89/0.98 Theta(max)= 80.766

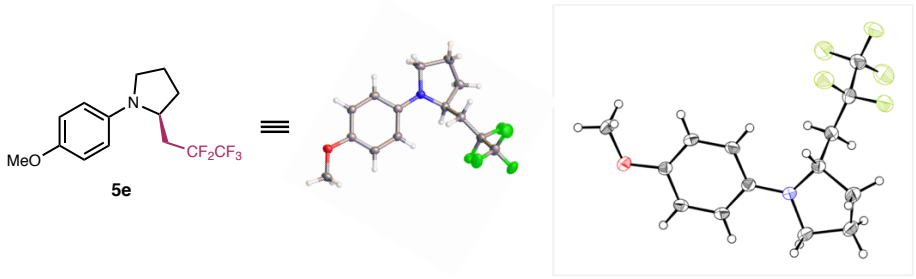
R(reflections)= 0.0279(4669) wR2(reflections)= 0.0705(4683)

S = 1.045 Npar= 290

The following ALERTS were generated. Each ALERT has the format
test-name_ALERT_alert-type_alert-level.
Click on the hyperlinks for more details of the test.

Alert level G

PLAT033 ALERT 4 G	Flack x Value Deviates > 3.0 * sigma from Zero .	0.070 Note
PLAT791 ALERT 4 G	Model has Chirality at C10 (Chiral SPGR)	R Verify
PLAT791 ALERT 4 G	Model has Chirality at C10B (Chiral SPGR)	R Verify
PLAT883 ALERT 1 G	No Info/Value for _atom_sites_solution_primary .	Please Do !
PLAT912 ALERT 4 G	Missing # of FCF Reflections Above STh/L= 0.600	14 Note
PLAT913 ALERT 3 G	Missing # of Very Strong Reflections in FCF	1 Note
PLAT961 ALERT 5 G	Dataset Contains no Negative Intensities	Please Check
PLAT978 ALERT 2 G	Number C-C Bonds with Positive Residual Density.	17 Info



Identification code	v18651
Empirical formula	$C_{14}H_{16}F_5NO$
Formula weight	309.28
Temperature/K	100
Crystal system	monoclinic
Space group	$P2_1$
a/Å	8.7849(11)
b/Å	5.4108(8)
c/Å	14.8752(19)
α/°	90
β/°	98.627(7)
γ/°	90
Volume/Å^3	699.07(16)
Z	2
ρ_{calc}g/cm^3	1.469
μ/mm^{-1}	1.212
F(000)	320.0
Crystal size/mm^3	0.24 × 0.23 × 0.12
Radiation	CuKα (λ = 1.54178)
2Θ range for data collection/°	6.01 to 158.618
Index ranges	$-10 \le h \le 11$, $-6 \le k \le 6$, $-18 \le l \le 18$
Reflections collected	22266
Independent reflections	2842 [R_{int} = 0.0414, R_{sigma} = 0.0219]
Data/restraints/parameters	2842/1/192
Goodness-of-fit on F^2	1.046
Final R indexes [I>=2σ (I)]	R_1 = 0.0276, wR_2 = 0.0694
Final R indexes [all data]	R_1 = 0.0279, wR_2 = 0.0696
Largest diff. peak/hole / e Å^{-3}	0.21/-0.14
Flack parameter	0.15(4)

Datablock: v18651

Bond precision: C-C = 0.0030 A Wavelength=1.54178

Cell: a=8.7849(11) b=5.4108(8) c=14.8752(19)
alpha=90 beta=98.627(7) gamma=90
Temperature: 100 K

	Calculated	Reported
Volume	699.07(16)	699.07(16)
Space group	P 21	P 1 21 1
Hall group	P 2yb	P 2yb
Moiety formula	C14 H16 F5 N O	C14 H16 F5 N O
Sum formula	C14 H16 F5 N O	C14 H16 F5 N O
Mr	309.28	309.28
Dx,g cm-3	1.469	1.469
Z	2	2
Mu (mm-1)	1.212	1.212
F000	320.0	320.0
F000'	321.35	
h,k,lmax	11,6,18	11,6,18
Nref	3035[1682]	2842
Tmin,Tmax	0.785,0.865	0.645,1.000
Tmin'	0.712	

Correction method= # Reported T Limits: Tmin=0.645 Tmax=1.000
AbsCorr = MULTI-SCAN

Data completeness= 1.69/0.94 Theta(max)= 79.309

R(reflections)= 0.0276(2809) wR2(reflections)= 0.0696(2842)

S = 1.046 Npar= 192

The following ALERTS were generated. Each ALERT has the format
test-name_ALERT_alert-type_alert-level.
Click on the hyperlinks for more details of the test.

Alert level C

PLAT911_ALERT_3_C	Missing FCF Refl Between Thmin & STh/L= 0.600	21	Report
PLAT915_ALERT_3_C	No Flack x Check Done: Low Friedel Pair Coverage	90	%

Alert level G

PLAT033_ALERT_4_G	Flack x Value Deviates > 3.0 * sigma from Zero .	0.150	Note
PLAT242_ALERT_2_G	Low 'MainMol' Ueq as Compared to Neighbors of	C14	Check
PLAT791_ALERT_4_G	Model has Chirality at C11 (Chiral SPGR)	S	Verify
PLAT883_ALERT_1_G	No Info/Value for _atom_sites_solution_primary .	Please	Do !
PLAT912_ALERT_4_G	Missing # of FCF Reflections Above STh/L= 0.600	35	Note
PLAT978_ALERT_2_G	Number C-C Bonds with Positive Residual Density.	10	Info

A.9 ^{1}H, ^{13}C, and ^{19}F NMR spectra

^{1}H, ^{13}C, and ^{19}F NMR spectra of the following compounds can be found in the Supporting Information of the published paper.

A.10 References

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

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

3. Shi, R. *et al.* Palladium/copper-catalyzed oxidative C–H alkenylation/*N*-dealkylative carbonylation of tertiary anilines. *Angew. Chem. Int. Ed.* **52**, 10582–10585 (2013).

4. McNally, A., Prier, C. K. & MacMillan, D. W. C. Discovery of an α-amino C–H arylation reaction using the strategy of accelerated serendipity. *Science* **334**, 1114–1117 (2011).

5. Kondolff, I., Doucet, H. & Santelli, M. Palladium−tetraphosphine as catalyst precursor for high-turnover-number Negishi cross-coupling of alkyl- or phenylzinc derivatives with aryl bromides. *Organometallics* **25**, 5219–5222 (2006).

6. Selva, M., Perosa, A., Tundo, P. & Brunelli, D. Selective *N*,*N*-dimethylation of primary aromatic amines with methyl alkyl carbonates in the presence of phosphonium salts. *J. Org. Chem.* **71**, 5770–5773 (2006).

7. Zhang, F.-G., Wei, Y., Yi, Y.-P., Nie, J. & Ma, J.-A. Regioselective cycloaddition of trifluorodiazoethane with electron-deficient allenic esters and ketones: Access to CF_3-substituted pyrazolines and pyrazoles. *Org. Lett.* **16**, 3122–3125 (2014).

8. Shen, H. *et al.* Intramolecular aminocyanation of alkenes promoted by hypervalent iodine. *Org. Chem. Front.* **4**, 1806–1811 (2017).

9. Kawamura, S., Egami, H. & Sodeoka, M. Aminotrifluoromethylation of olefins via cyclic amine formation: Mechanistic study and application to synthesis of trifluoromethylated pyrrolidines. *J. Am. Chem. Soc.* **137**, 4865–4873 (2015).

10. Li, Y. & Studer, A. Transition-metal-free trifluoromethylaminoxylation of alkenes. *Angew. Chem. Int. Ed.* **51**, 8221–8224 (2012).

Chapter 3

CHEMODIVERGENT C(SP^3)–H AND C(SP^2)–H CYANOMETHYLATION USING ENGINEERED CARBENE TRANSFERASES

Material from this chapter will be submitted for peer review as: **Zhang, J.**; Maggiolo, A. O.; Alfonzo, E.; Mao, R.; Porter, N. J.; Abney, N.; Arnold, F. H. Chemodivergent C(sp^3)–H and C(sp^2)–H cyanomethylation using engineered carbene transferases.

J.Z. designed the overall research with F.H.A. providing guidance. J.Z. designed and conducted the initial screening of heme proteins; J.Z. and N.A. performed the directed evolution experiments. J.Z., E.A., and R.M. designed and performed the substrate scope studies and analysis. A.O.M. obtained the X-ray crystal structure of the engineered proteins with N.J.P. providing help. J.Z. and F.H.A. wrote the manuscript with input from all authors.

Abstract

The ubiquity of C–H bonds presents an attractive yet challenging opportunity to elaborate molecules. C–H functionalization methods are typically complicated by the presence of multiple chemically similar and, in some cases indistinguishable, C–H bonds. Unlike chemocatalysts, enzymes can be finely tuned using directed evolution, allowing exacting control over divergent C–H functionalization pathways. Here, we report two complementary carbene C–H transferases derived from the bacterial cytochrome P450 from *Bacillus megaterium*. In parallel, the engineered enzymes can deliver an α-cyanocarbene into the α-amino C(sp^3)–H bonds and the *ortho*-arene C(sp^2)–H bonds of *N*-substituted arenes, offering site-selectivity rarely seen with chemocatalysts. These transformations proceed via different reaction mechanisms, yet only minimal changes to the protein scaffold (eight mutations) were needed to adjust the reaction selectivity. The X-ray crystal structure of P411-PFA reveals a previously unobserved conformation of the protein backbone that alters the shape and electronics in and around the enzyme active site as a result of mutations installed

during directed evolution. Overall, this work demonstrates complementary biocatalysts can be developed to achieve chemodivergent C–H alkylation reactions and highlights the versatility of enzymatic catalysts for divergent organic synthesis.

3.1 Introduction

Given the ubiquity of C–H bonds in organic molecules, advancing selective C–H functionalization methodology can fundamentally simplify chemical synthesis.[1–5] Ideally, precise and divergent molecular scaffold alterations would be achieved by using a panel of selective, distinct catalysts that can functionalize each C–H bond in a molecule, including both C(sp^3)–H and C(sp^2)–H bonds.[6] Methods of this kind are promising in many settings, such as late-stage derivatizations of pharmaceuticals, agrochemicals, and materials.[4,7–9] Notwithstanding the appeal of this approach, design principles and methodologies featuring small-molecule catalysts with complementary selectivity are scarce. Successful examples have exploited directing groups to guide the functionalization of the desired C–H bonds.[10–12] Divergent, highly selective methods that act on desired substrates absent of guiding functional groups are desired.[3]

We consider enzymatic catalysts as promising alternatives to address the unmet selectivity challenges in C–H functionalization reactions. In contrast to synthetic organometallic complexes, catalyst-controlled selectivity can be more readily achieved and reprogrammed in enzymes' active sites through fine-tuned substrate alignments that enable complementary reaction outcomes. This is exemplified by C–H hydroxylases and halogenases with divergent chemo-, regio-, and stereoselectivity either found in nature or engineered using directed evolution.[13–20] Over the past decade, our group and others have further broadened the scope of enzymatic C–H functionalization reactions by introducing abiological transformations originally developed by synthetic chemists to natural enzymes.[21] Representative examples include repurposing heme proteins and non-heme Fe enzymes to selectively alkylate or aminate C–H bonds via carbene- or nitrene-transfer reactions[22–25] and exploiting non-heme Fe enzymes for radical-mediated selective C–H azidation and nitration

reactions.[19,26,27] However, the vast majority of these efforts have solely focused on targeting one specific sp^3-hybridized C–H bond; chemodivergent approaches to functionalizing both $C(sp^3)$–H bonds and aromatic $C(sp^2)$–H bonds are lacking.

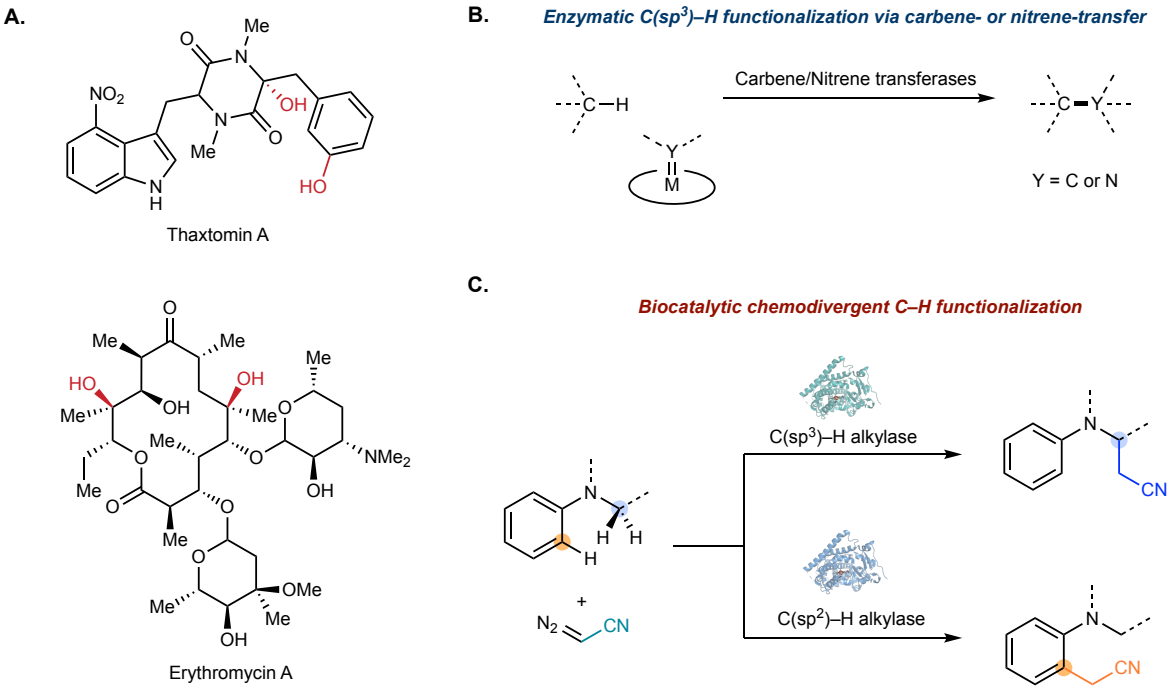


Figure 3-1. Reaction design. A) Enzymatic C–H functionalization reactions are often highly selective and divergent; B) Previous work on abiological C–H functionalization mainly focused on modifying $C(sp^3)$–H bonds; C) The goal of this study was to develop complementary selective C–H alkylases which can distinguish nearby $C(sp^3)$–H and arene $C(sp^2)$–H bonds to enable divergent C–H functionalization with closely related catalysts.

Here, we describe two complementary P450-based carbene transferases which can selectively install a cyanomethyl group to a $C(sp^3)$–H bond or an arene $C(sp^2)$–H bond. Whereas enzymatic and transition-metal catalyzed $C(sp^3)$–H carbene insertion are well documented,[1,28–34] examples of highly selective intermolecular carbene transfer to an arene C–H bond remain rare.[35,36] In small-molecule catalysis, high regioselectivity is challenging to achieve via the postulated Friedel-Crafts-like electrophilic substitution mechanism.[37,38] Site-selective transformations often occur at the least sterically demanding position.[39–41]

Meanwhile, state-of-the-art biocatalytic systems developed by the Fasan lab and our lab have been limited to electron-rich heteroaromatics.[42,43] Inspired by these precedents, we set out to engineer carbene transferases that favor arene C–H functionalization, as complements to the reported "C(sp^3)–H alkylases". This complementary, chemodivergent enzymatic platform can enable straightforward generation of different constitutional isomers which are laborious to make otherwise. Additionally, nitriles and their derivatives (e.g. amide) are well-established functional groups in medicinal chemistry that can be rapidly diversified in complexity building transformations.[44,45] We see this enzymatic platform as a proof of principle and ideal starting point to generate new "C–H cyanomethylases" which can functionalize each one of the C–H bonds in complex bioactive molecules.

3.2 Results and discussion

3.2.1 Initial activity screening and reaction discovery

We commenced this effort by evaluating the biocatalytic C–H carbene transfer reaction with diazoacetonitrile **1** and *N*-phenyl morpholine **2a** (**Figure 3-2**). This transformation relies on the generation and transfer of an α-cyanocarbene intermediate,[46] which has been reported in both chemocatalytic and biocatalytic systems for cyclopropanation, N–H insertion, S–H insertion and indole alkylation reactions.[46–48] General C–H cyanomethylation via catalytic carbene transfer has been elusive. In our initial studies, we examined the enzymatic C(sp^3)–H cyanomethylation reaction with a panel of 82 different variants of an axial serine-ligated cytochrome P450 (P411s, previously engineered for abiological carbene transfer reactions) (**Supplementary Table B-1**). The heme cofactor alone is not an active catalyst for this reaction. Many of the P411 variants, however, catalyzed the carbene insertion into the α-amino C(sp^3)–H bond of **2a**, affording **3a** with moderate yield. Notably, P411-PFA, a carbene transferase previously engineered to catalyze α-amino C(sp^3)–H fluoroalkylation,[30] afforded the α-amino C(sp^3)–H cyanomethylation product with 9% yield.

We were fascinated to observe that the use of diazoacentonitrile **2** as carbene precursor unlocked basal-level *ortho*-arene C–H functionalization activity (**Figure 3-1**,

Supplementary Figure B-1) in a small number of P411 variants (**Supplementary Table B-2**). This activity was not seen in previous studies using diazoacetate (e.g. ethyl diazoacetate, EDA) or perfluorinated diazo compounds (e.g. 2,2,2-triflurodiazoethane). The formation of **4a** is not catalyzed by the heme cofactor, nor is it produced by the cellular background. FA-E3, the penultimate variant from the previous C–H fluoroalkylation lineage, catalyzes the formation of both products at a low level and with the highest **4a**/**3a** ratio (2:3) among the enzymes tested.

3.2.2 Structural studies of P411-PFA

To gain structural insight into the α-amino C(sp^3)–H cyanomethylase that might help guide the directed evolution of an arene C–H cyanomethylase, we obtained the X-ray crystal structure of P411-PFA's heme domain at 1.87-Å resolution. Overall, P411-PFA adopts an architecture that is typical of previously solved structures of cytochromes P450 and P411.[49–51] Compared to variant P-4 A82L A78V F263L, an intermolecular C–H aminase which functionalizes benzylic C–H bonds using tosyl azide as nitrene precursor,[51] P411-PFA contains 13 additional mutations (**Supplementary Figure B-2**), most of which are located near the active site. Relative to this amination catalyst, the substrate pocket volume of P411-PFA shrank by ~7% (**Figure 3-2A**). This is potentially due to the smaller substrates used in its evolution and the precise requirement of steric environment to orient the substrate for carbene insertion at the selected position.

Beyond these similarities, nonetheless, an unprecedented conformational change was observed. (**Figure 3-2**). In the *I*-helix of wild-type $P450_{BM3}$, there is a helical distortion over a single turn between residues 263 and 268. This breaks the standard hydrogen bonding pattern of this α-helix, instead forming a water-mediated interaction between these residues reminiscent of the $i + 5 \rightarrow i$ hydrogen bonding pattern present in a π-helix.[52] Notably, this distortion is expanded in the structure of P411-PFA due to an unusual backbone carbonyl flip present at position 267 to fully disrupt the helical hydrogen bonding network, trapping a water molecule within an expanded coil and breaking the standard *I*-helix into two distinct helices (*I* and *I'*). This breakage results in the accumulation of complementary dipoles on

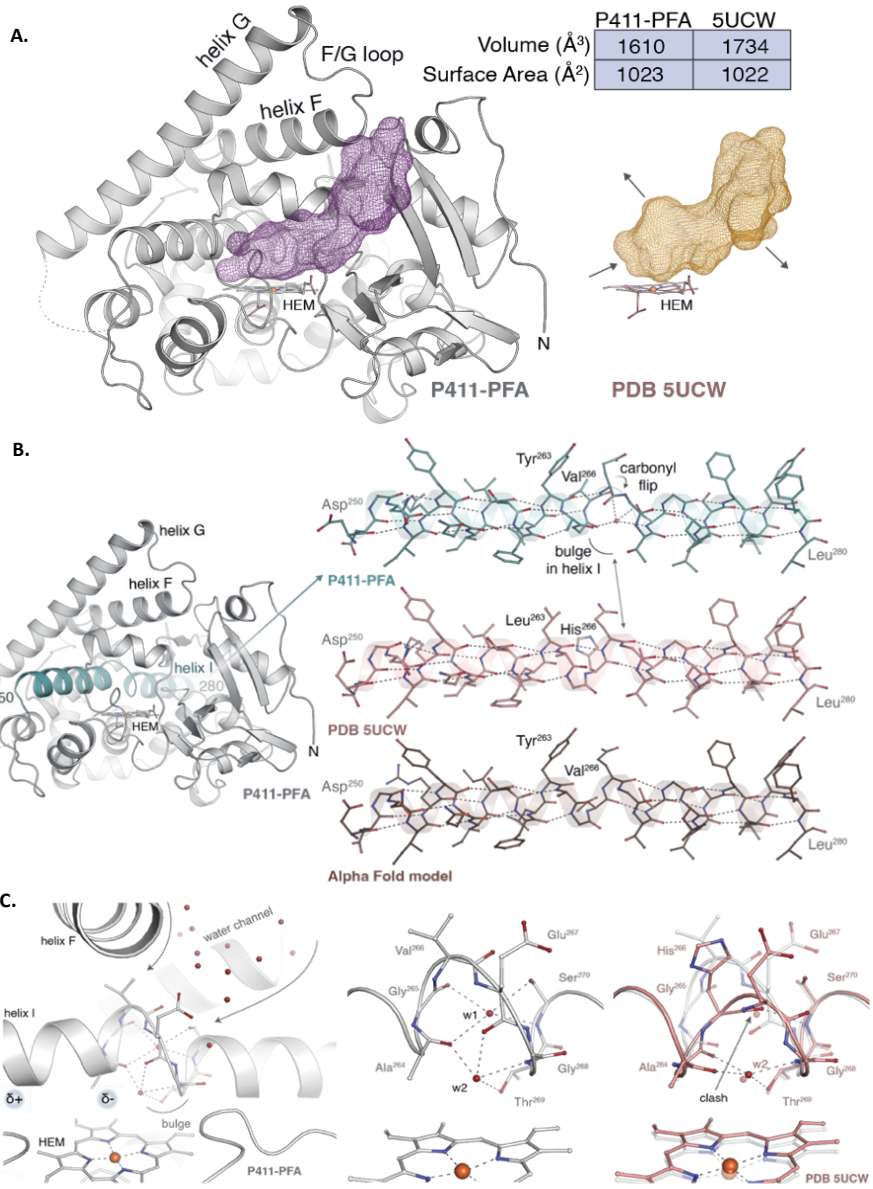


Figure 3-2. Crystallography studies of P411-PFA. **A)** The overall active site comparison of P411-PFA with a closely related C–H aminase, P-4 A82L A78V F263L (PDB ID: 5UCW) (P411-PFA contains 13 additional mutations); **B**) An unusual "backbone carbonyl flip" was observed in the *I*-helix of P411-PFA. As expected, an AlphaFold2-predicted model of P411-PFA fails to capture this structural change; **C**) Close caption showing the backbone carbonyl flip. This unprecedented conformational change creates a strong dipole near the active site, which might make the local electronic environment more suitable for diazo binding and activation.

either side of this expanded coil, altering the electrostatics in the active site (**Figure 3-2C**). Intriguingly, while P411-PFA contains 26 heme-domain mutations compared to wild-type P450-BM3, Glu^{267} had not been changed (**Supplementary Figure B-3**). Mutations L263Y and H266V are present at positions flanking the *I*-helix and likely stabilize the otherwise unfavorable flipped conformation of the Glu^{267} carbonyl. Given that these two mutations reside in the region known to affect O_2 activation in wild-type P450 monooxygenases[53] and enhanced carbene C(sp^3)–H insertion activity,[29] we postulated that this unusual conformational change may play a role in how these carbene transferases bind and activate diazo compounds to generate the metallocarbene species.

3.2.3 Directed evolution of a selective arene C–H carbene transferase

The different C(sp^3)–H and C(sp^2)–H selectivities of P411-PFA and FA-E3 inspired us to engineer a carbene transferase which would favor arene C–H functionalization over α-amino C(sp^3)–H insertion. P411-FA-E3 served as the starting enzyme for sequential rounds of site-saturation mutagenesis and error-prone PCR mutagenesis to obtain enzymes with C(sp^2)–H cyanomethylation activity and selectivity for the arene C–H bond over the nearby alkyl C–H bond (**Figure 3-3A**). To preserve the novel structural feature (helix breakage) found in P411-PFA which might be beneficial for carbene-transfer activity, we targeted amino acid residues which reside on the opposite side of the active site from the disrupted *I*-helix and were previously found to affect abiological carbene- and nitrene-transfer activities for site-saturation mutagenesis (**Figure 3-3B, C**). We found mutations within the proximal active site pocket (A87V, M177Y, W325C, V330C, and M354V) as well as distal ones that affect the chemoselectivity of the protein catalysts (S118Q, G252L).

Furthermore, in comparison to FA-E3, P411-PFA contains an additional L401P mutation in the axial "Cys pocket" (L401P), which disrupted the hydrogen bond between the amide proton of residue 402 and axial serine ligand (**Supplementary Figure B-4**); mutations in this pocket are known to tune the electronic property of heme.[54,49,55] The different selectivities of these two variants encouraged us to investigate other residues in the Cys pocket to improve arene C–H functionalization selectivity; F393W was then identified as a

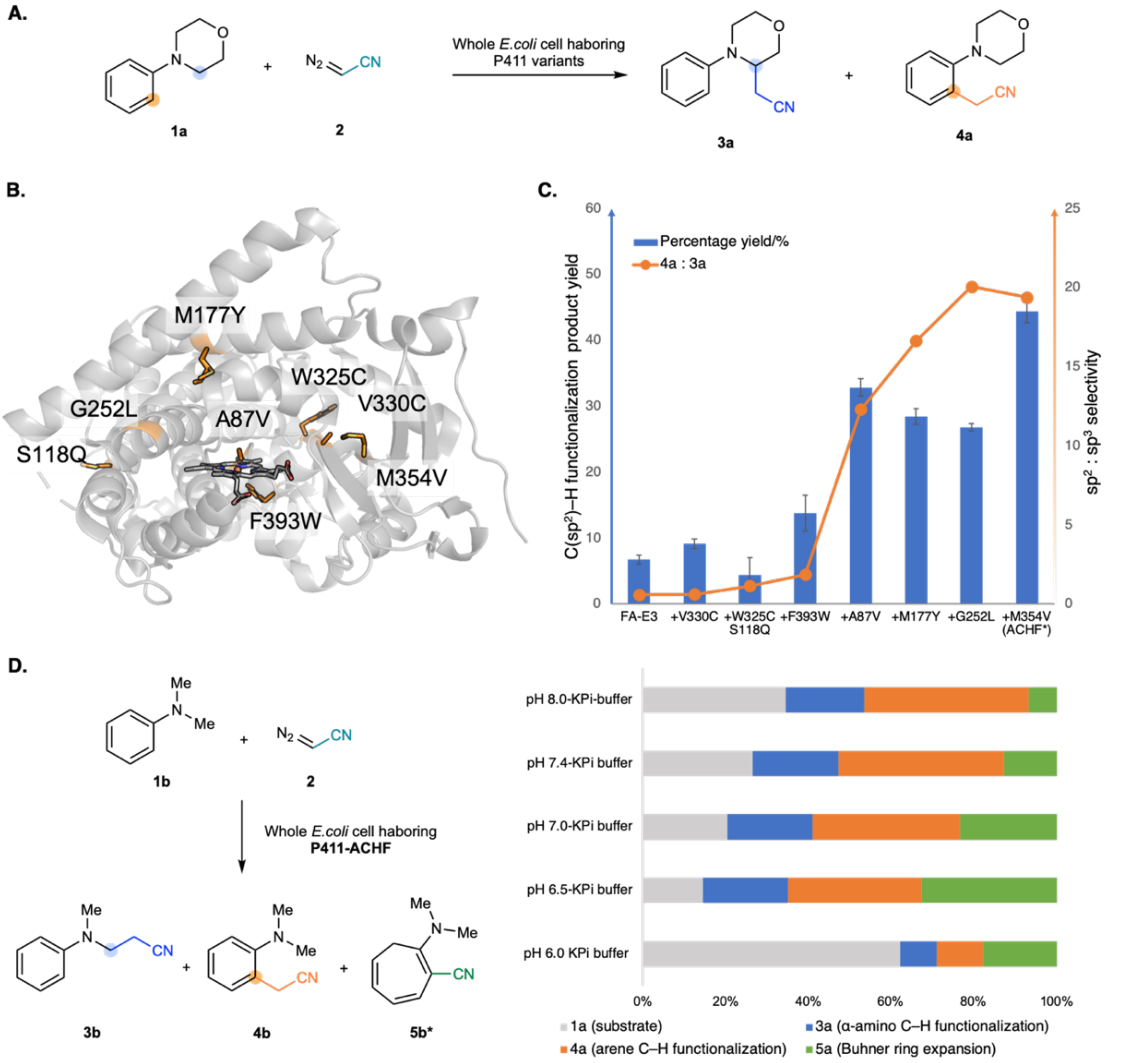


Figure 3-3. Reaction discovery and directed evolution. Reaction conditions: 5mM **1a** or **1b**; 48 mM **2** (9.63 equiv.); *E. coli* whole cell harboring P411 variants (OD_{600} = 30) M9-N aqueous buffer (pH 7.4); 10% v/v EtOH (co-solvent); room temperature; 18–20 h. **A**) Initial activity screening revealed the formation of C(sp^3)–H insertion product **3a** and ortho-arene C–H alkylation product **4a**; **B**) Locations of beneficial mutations are shown in P411-PFA structure; **C**) Directed evolution of a highly selective arene C–H cyanomethylase; **D**) Enzymatic Buchner ring expansion reaction. Basic pH suppresses Buchner product formation (**5b**, *: inseparable mixture of tautomerized isomers).

beneficial mutation from site-saturation mutagenesis of this region. In summary, a total of seven rounds of mutagenesis and screening yielded the final variant P411-ACHF (arene C–H functionalization enzyme, **Supplementary Table B-3**), which contains eight additional mutations relative to P411-PFA (**Figure 3-2B**). Under optimized conditions (**Supplementary Table B-4**), P411-ACHF delivers **4a** in 51% yield and excellent chemoselectivity (**4a**:**3a** = 19:1).

3.2.4 Enzymatic Buchner ring expansion reaction

Similar to many small-molecule arene C–H carbene-transfer catalysts,[37,56,57] P411-ACHF also exhibits (low) Buchner ring expansion reaction activities. We observed the most Buchner product when the enzyme is challenged with **1b** and **2** (**Figure 3-2D**). In that case, P411-ACHF cyclopropanates the most nucleophilic π-bond of **1b**. Subsequently, due to the electron-withdrawing effect of the nitrile group, the α-proton in the cycloheptatriene is highly acidic, and therefore the product can readily tautomerize into **5b** to break the norcaradiene-cycloheptatriene equilibrium. Interestingly, the enzymatic Buchner ring expansion reaction is sensitive to the pH of the reaction buffer (**Figure 3-2D**). We found that mildly basic conditions (pH = 7.4–8.0) minimize the ring expansion side product.

3.2.5 Substrate scope studies of P411-PFA and P411-ACHF

Substrate scope studies were conducted under yield-optimized conditions with the two highly selective but distinct cyanomethylases, P411-PFA and P411-ACHF. As shown in **Figure 3-3**, these enzymes are capable of alkylating *N*-substituted arenes with complementary chemoselectivity. Without any additional protein engineering, the previously engineered C(sp^3)–H fluoroalkylase P411-PFA efficiently installs cyanomethyl groups to α-amino C(sp^3)–H bonds with high enantio-, and chemoselectivity. While P411-PFA functionalizes the C(sp^3)–H bonds of these *N*-substituted arenes (**3a-e**, **Figure 3-3A**), P411-ACHF alkylates their arene *ortho*-C–H bonds (**4a-e**). The chemoselectivity of P411-ACHF is not yet excellent with some substrates, but directed evolution should be able to improve activity and selectivity on any specific substrate. As proof-of-concept, we performed an additional round of directed evolution using **1b** as the new model substrate and identified

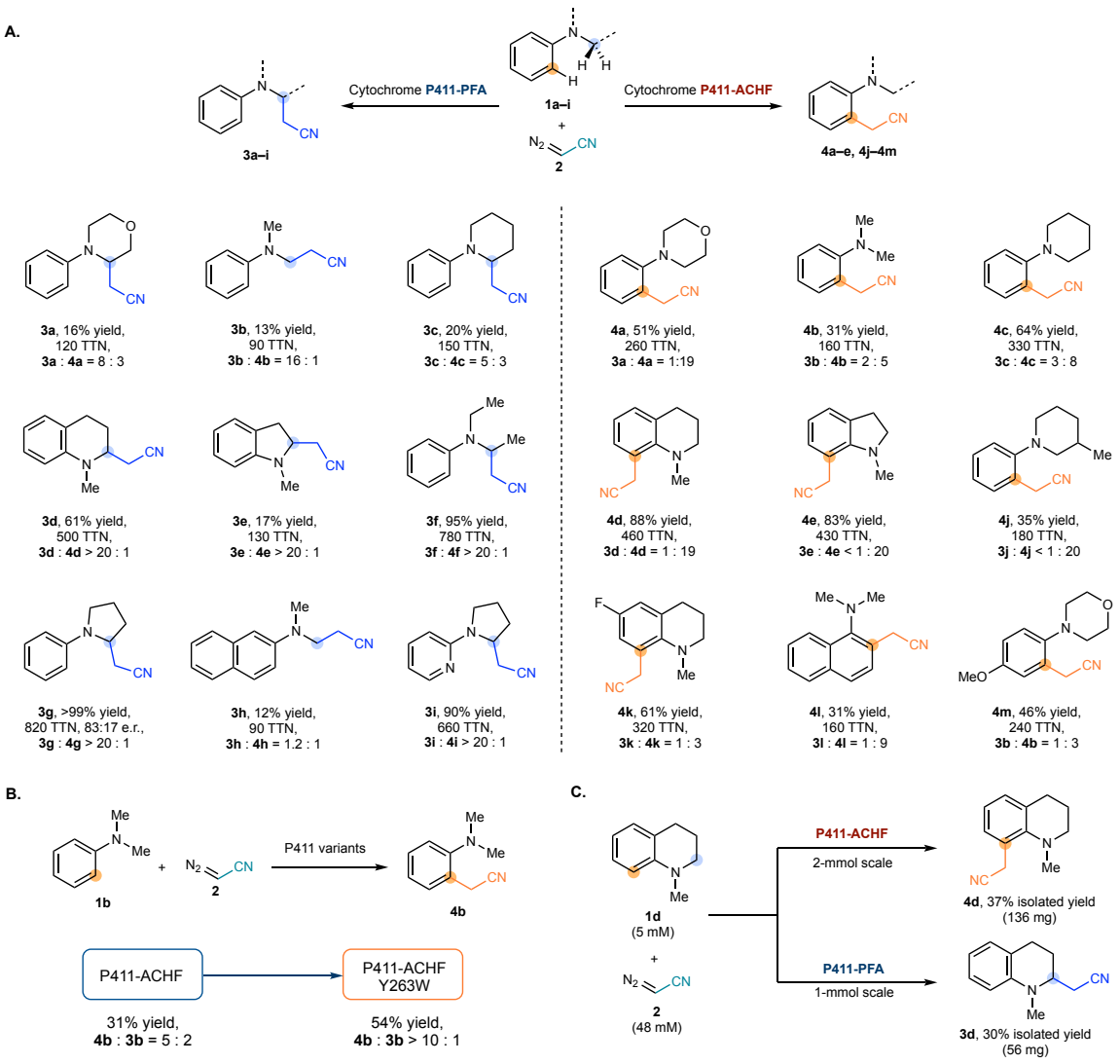


Figure 3-4. Substrate scope studies. **A**) Substrate scope of P411-PFA and P411-ACHF. Reaction conditions: 2.5 mM **1**; 51.6 mM **2** (20.63 equiv.); *E. coli* whole cells harboring P411 variants (OD_{600} = 30) M9-N aqueous buffer (pH 7.4); 10% v/v EtOH (co-solvent); room temperature; 18–20 h; **B**) The arene C–H functionalization activity and selectivity can be further optimized on individual target molecules; C) Chemodivergent, preparative-scale syntheses using P411-PFA and P411-ACHF. Reaction conditions: 5 mM **1d**; 48 mM **2** (9.63 equiv.); *E. coli* whole cells harboring P411 variants (OD_{600} = 30) M9-N aqueous buffer (pH 7.4); 10% v/v EtOH (co-solvent); room temperature; 18–20 h.

Y263W as a beneficial mutation. The optimized variant, P411-ACHF Y263W, with increased steric bulk within the enzyme active site, delivers **4b** with 54% yield and >10:1 **4b**:**3b** ratio (**Figure 3-3B**, **Supplementary Table B-5**). As proof-of-concept to demonstrate the synthetic utilities of these enzymes, we performed chemodivergent derivatization of **1d** using P411-PFA and P411-ACHF on preparative scale, where the enzymes yielded constitutional isomer **3d** and **4d** with 30% (1-mmol scale, 56 mg) and 37% (2-mmol scale, 136 mg) isolated yield, respectively (**Figure 3-3C**).

Substrate recognition within the protein scaffold is important in determining the carbene-transfer selectivity. P411-PFA preferentially functionalizes the secondary α-amino C(sp^3)–H bonds in the presence of other C–H bonds with similar bond-dissociation energies (e.g. benzylic C–H bonds in **1d** and **1e**). Meanwhile, P411-ACHF exhibited exclusive selectivity toward *ortho*-C–H bonds (to nitrogen) over other activated positions (positions *ortho* to –OMe in **4m**). Notably, in transition metal-catalyzed carbene transfer to C(sp^2)–H bonds of *N*-substituted arenes, the kinetically more accessible *para*-positions are favored.[40,41] However, in the biocatalytic transformations reported here, the *para*-position may be sterically occluded by the enzyme, leading to the observed *ortho*-selectivity. The substrate scope studies have revealed that with a small number of mutations, these new-to-nature enzymes can be devised to discriminate nearby C–H bonds. Indeed, precise substrate recognition prevented P411-ACHF from selectively functionalizing the arene C–H bonds in other substrates (e.g. **1g**). But we propose these enzymes could serve as starting points to evolve a set of C(sp^3)–H and C(sp^2)–H alkylases to functionalize every C–H bond in a molecule with high selectivity.

3.3 Conclusion

In this study, we developed two complementary P450-based carbene transferases which selectively functionalize C(sp^3)–H bonds and arene C(sp^2)–H bonds present in the same molecule. These results demonstrate that the divergent, native C–H functionalization selectivity widely exhibited by hydroxylating P450s can be re-established in a non-native enzymatic reaction. Using directed evolution, we have obtained a highly active enzyme

which catalyzes unprecedented C(sp^2)–H carbene transfer to arene substrates which are not activated heteroaromatics (indole, pyrrole, etc.). Structural studies revealed an unusual helix disruption in the active site that was not captured by computational predictions, underscoring the advantages of using directed evolution to accentuate structural features that unlock new-to-nature activities in repurposed enzymes. In summary, we envision these complementary carbene C–H functionalization enzymes can be further engineered to perform more challenging, divergent C–H alkylation reactions, with selectivity unmatched by chemical catalysts.

3.4 Experimental methods

3.4.1 General

Unless otherwise noted, all chemicals and reagents were obtained from commercial suppliers (Sigma-Aldrich, VWR, Alfa Aesar, Acros, Combi Blocks, etc.) and used without further purification. Silica gel chromatography was carried out using AMD Silica Gel 60, 230-400 mesh. ^{1}H and ^{13}C NMR spectra were recorded on a Bruker Prodigy 400 MHz instrument (400 MHz for ^{1}H and 101 MHz for ^{13}C NMR) or a Varian 300 MHz Spectrometer (300 MHz for ^{1}H NMR). Chemical shifts (δ) are reported in ppm downfield from tetramethylsilane, using the solvent resonance as the internal standard (^{1}H NMR: δ = 7.26, ^{13}C NMR: δ = 77.36 for $CDCl_3$). Data for ^{1}H NMR are reported as follows: chemical shift (δ ppm), multiplicity (s = singlet, d = doublet, t = triplet, q = quartet, p = pentet, sext = sextet, m = multiplet, dd = doublet of doublets, dt = doublet of triplets, ddd = doublet of doublet of doublets), coupling constant (Hz), integration. Sonication was performed using a Qsonica Q500 sonicator. High-resolution mass spectra were obtained at the California Institute of Technology Mass Spectral Facility.

Escherichia coli cells were grown using Luria-Bertani medium or HyperBroth (AthenaES) with 100 μg/mL ampicillin (LB_{amp} or HB_{amp}). Primer sequences are available upon request. T5 exonuclease, Phusion polymerase, and *Taq* ligase were purchased from New England Biolabs (NEB, Ipswich, MA). M9-N minimal medium (abbreviated as M9-N

buffer; pH 7.4) or potassium phosphate buffer (abbreviated as KPi) was used as a buffering system for whole cells, lysates, and purified proteins, unless otherwise specified. M9-N buffer was used without a carbon source; it contains 47.7 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 8.6 mM NaCl, 2.0 mM $MgSO_4$, and 0.1 mM $CaCl_2$.

3.4.2 Chromatography

Gas chromatography mass spectrometry (GC-MS) and GC-FID analyses were carried out using a Agilent system and a J&W HP-5MS column. Analytical chiral HPLC was conducted using either an Agilent 1200 series instrument with *n*-hexane and isopropanol as the mobile phase. Enantiomers were separated using one of the following chiral columns: Chiralpak IA (4.6 mm × 25 cm), Chiralpak IB (4.6 mm × 25 cm),

3.4.3 Cloning and site-saturation mutagenesis

The genes encoding all enzymes described in this study were cloned using Gibson assembly into vector pET22b(+) (Novagen) between restriction sites *Nde*I and *Xho*I in frame with a *C*-terminal 6xHis-tag. Site-saturation mutagenesis was performed using the "22c-trick"[58] or "NNK" as degenerative codons. The PCR products were digested with *Dpn*I, gel purified, and ligated using Gibson Mix™.[59] The ligation mixture was used to directly transform electrocompetent *E. coli* strain *E. cloni* BL21 (DE3) cells (Lucigen).

3.4.4 Expression of P450 and P411 variants in 96-well plates

Single colonies from LB_{amp} agar plates were picked using sterile toothpicks and cultured in deep-well 96-well plates containing LB_{amp} (400 µL/well) at 37 °C , 80% humidity and 250 rpm shaking overnight. Subsequently, HB_{amp} (1080 µL/well) in a deep-well plate was inoculated with an aliquot (120 µL/well) of these overnight cultures and allowed to shake for 3 hours at 37 °C , 80% humidity and 250 rpm. The plates were then cooled on ice for 30 minutes and the cultures were induced with 0.5 mM isopropyl *β*-D-1-thiogalactopyranoside (IPTG) and 1.0 mM 5-aminolevulinic acid (ALA) (final concentrations). Expression was then conducted at 20 °C, 230 rpm for 18–20 hours.

3.4.5 Plate reaction screening in whole-cell format

E. coli cells harboring P411 variants in deep-well 96-well plates were pelleted (3,500 × g, 5 min, 4 °C) and resuspended in M9-N buffer (375 μL) by gentle vortexing. The 96-well plates were then transferred to an anaerobic chamber. To deep-well plates of cell suspensions were added the *N*-phenylmorpholine substrate **1a** (5 μL/well, 400 mM in EtOH) and diazoacetonitrile **2** (35 μL per well, 600 mM in EtOH). During the addition, the stock solution of 2 diazoacetonitrile **2** was kept on an ice bath to minimize evaporation. The plates were sealed with aluminum sealing foil immediately after the addition and shaken in the anaerobic chamber at room temperature and 600 rpm. After 18–24 hours, the seals were removed and 610 μL 1:1 ethylacetate/hexanes solution containing 0.66 mM 1,2,3-trimethoxybenzene internal standard was added. The plates were tightly sealed with silicone mats, vigorously vortexed, and centrifuged (5,000 × g, 5 min) to completely separate the organic and aqueous layers. The organic layer of each well was then transferred to individual vials equipped with autosampler vial inserts and analyzed by GC-MS.

3.4.6 Expression of P411 variants

E. coli (*E. cloni* BL21(DE3)) cells carrying plasmid encoding the appropriate P411 variant were grown overnight in 5 mL LB_{amp}. Preculture (1.5 mL) was used to inoculate 28.5 mL of HB_{amp} in a 125 mL Erlenmeyer flask; this culture was incubated at 37 °C, 230 rpm for 2.5 hours. The culture was then cooled on ice for 30 min and induced with 0.5 mM IPTG and 1.0 mM ALA (final concentrations). Expression was conducted at 20 °C, 150 rpm for 16-18 hours. Subsequently, the *E. coli* cells were pelleted by centrifugation (3,000 × g, 5 min, 4 °C). Media was removed and the resulting cell pellet was resuspended in phosphate buffer (M9-N or KPi with varied pH) to OD_{600} = 30.

3.4.7 Hemochrome assay for the determination of heme protein concentration

The concentration of heme protein in the clarified lysate was determined by the hemochrome assay.[60] Briefly, 500 μL of the lysate were added to a cuvette and mixed with 500 μL of solution I (0.2 M NaOH, 40% (v/v) pyridine, 500 μM potassium ferricyanide).

The UV-Vis spectrum (380–600 nm) of the oxidized Fe^{III} state was recorded immediately. Sodium dithionite (10 μL of 0.5 M solution in 0.5 M NaOH) was added and the UV-Vis spectrum of the reduced Fe^{II} state was recorded immediately. The pyridine hemochromagen concentration was determined using its Q bands, with the extinction coefficients of 34.7 mM^{-1} cm^{-1} at 557 nm.

3.4.8 Biotransformation using whole *E. coli* cells

Suspension of *E. coli* (*E. cloni* BL21(DE3)) cells expressing the appropriate heme protein variant in M9-N or KPi buffer (typically OD_{600} = 30) were transferred to a reaction vial. The headspace of the reaction vial was purged with a stream of argon for at least 15 minutes. Enzymatic reactions were then set up in anaerobic chamber. To a 2-mL vial were added degassed suspension of *E. coli* expressing P411 variant (typically OD_{600} = 30, 360 μL), *N*-substituted arene substrate (typically 2.5 or 5 μL of 400 mM stock solution in EtOH). The diazoacetonitrile solution (typically 37.5 or 35 μL of 550 mM stock solution in EtOH. The concentration of diazo solution was measured by ^{1}H-NMR and adjusted to 550 mM) was kept in an ice bath and added last. The final volume of the biotransformation was set to be 400 μL, with 10% vol EtOH. The reaction vials were then capped and shaken in the anaerobic chamber at room temperature and 700 rpm for 18–20 hours. After the completion of the reaction, 610 μL 1:1 ethylacetate/hexanes solution containing 0.66 mM 1,2,3-trimethoxybenzene internal standard was added to the vial. The resulting mixture was transferred to a 1.5-mL microcentrifuge tube, vigorously vortexed, and centrifuged (20,000 × g, 5 minutes) to completely separate the organic and aqueous layers. The organic layer was transferred to a vial equipped with an autosampler vial insert and analyzed by GC-FID.

3.4.9 Enzymatic preparative synthesis

Suspension of *E. coli* (*E. cloni* BL21(DE3)) cells harboring either P411-PFA or P411-ACHF was prepared as described in **Section 3.4.6**. The lysate was placed in a sealed vial on ice and the headspace was purged with a stream of argon for 30 minutes. To the reaction vial were added a GOx oxygen depletion solution (5 mL of stock solution containing 14,000 U/mL catalase and 1,000 U/mL glucose oxidase in M9-N buffer), D-

glucose (5 mL of 250 mM stock solution in M9-N buffer), $Na_2S_2O_4$ (5 mL of 20 mM solution in M9-N), and substrate **1** (2.5 mL of 400 mM stock solution in EtOH). 2,2,2-trifluoro-1-diazoethane **2** (3.75 mL of 600 mM stock solution in EtOH) kept in an ice-salt bath was added at last (-20 °C). The reaction vial was immediately capped and sealed with parafilm, removed from anaerobic chamber, and shaken at room temperature at 250 rpm for 12 hours. The reaction solution was then extracted with 100 mL 1:1 hexane/ethylacetate for three times. The combined organic layer was then washed with brine, dried over anhydrous $MgSO_4$, concentrated, and purified by flash chromatography.

3.5 References

1. Davies, H. M. L. & Manning, J. R. Catalytic C–H functionalization by metal carbenoid and nitrenoid insertion. *Nature* **451**, 417–424 (2008).

2. Davies, H. M. L., Bois, J. D. & Yu, J.-Q. C–H functionalization in organic synthesis. *Chem. Soc. Rev.* **40**, 1855–1856 (2011).

3. Hartwig, J. F. & Larsen, M. A. Undirected, homogeneous C–H bond functionalization: Challenges and opportunities. *ACS Cent. Sci.* **2**, 281–292 (2016).

4. Yamaguchi, J., Yamaguchi, A. D. & Itami, K. C–H bond functionalization: Emerging synthetic tools for natural products and pharmaceuticals. *Angew. Chem. Int. Ed.* **51**, 8960–9009 (2012).

5. Dalton, T., Faber, T. & Glorius, F. C–H activation: Toward sustainability and applications. *ACS Cent. Sci.* **7**, 245–261 (2021).

6. Börgel, J. & Ritter, T. Late-stage functionalization. *Chem* **6**, 1877–1887 (2020).

7. Cernak, T., Dykstra, K. D., Tyagarajan, S., Vachal, P. & Krska, S. W. The medicinal chemist's toolbox for late stage functionalization of drug-like molecules. *Chem. Soc. Rev.* **45**, 546–576 (2016).

8. Guillemard, L., Kaplaneris, N., Ackermann, L. & Johansson, M. J. Late-stage C–H functionalization offers new opportunities in drug discovery. *Nat. Rev. Chem.* **5**, 522–545 (2021).

9. Fazekas, T. J. *et al.* Diversification of aliphatic C–H bonds in small molecules and polyolefins through radical chain transfer. *Science* **375**, 545–550 (2022).

10. Neufeldt, S. R. & Sanford, M. S. Controlling site selectivity in palladium-catalyzed C–H bond functionalization. *Acc. Chem. Res.* **45**, 936–946 (2012).

11. Dai, H.-X., Stepan, A. F., Plummer, M. S., Zhang, Y.-H. & Yu, J.-Q. Divergent C–H functionalizations directed by sulfonamide pharmacophores: Late-stage diversification as a tool for drug discovery. *J. Am. Chem. Soc.* **133**, 7222–7228 (2011).

12. Sambiagio, C. *et al.* A comprehensive overview of directing groups applied in metal-catalysed C–H functionalisation chemistry. *Chem. Soc. Rev.* **47**, 6603–6743 (2018).

13. Greule, A., Stok, J. E., Voss, J. J. D. & Cryle, M. J. Unrivalled diversity: The many roles and reactions of bacterial cytochromes P450 in secondary metabolism. *Nat. Prod. Rep.* **35**, 757–791 (2018).

14. Kille, S., Zilly, F. E., Acevedo, J. P. & Reetz, M. T. Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nat. Chem.* **3**, 738–743 (2011).

15. Zhang, K., Shafer, B. M., Demars, M. D., Stern, H. A. & Fasan, R. Controlled oxidation of remote sp^3 C–H bonds in artemisinin via P450 catalysts with fine-tuned regio- and stereoselectivity. *J. Am. Chem. Soc.* **134**, 18695–18704 (2012).

16. Lukowski, A. L. *et al.* C–H hydroxylation in paralytic shellfish toxin biosynthesis. *J. Am. Chem. Soc.* **140**, 11863–11869 (2018).

17. Zhang, X. *et al.* Divergent synthesis of complex diterpenes through a hybrid oxidative approach. *Science* **369**, 799–806 (2020).

18. Andorfer, M. C., Park, H. J., Vergara-Coll, J. & Lewis, J. C. Directed evolution of RebH for catalyst-controlled halogenation of indole C–H bonds. *Chem. Sci.* **7**, 3720–3729 (2016).

19. Neugebauer, M. E. *et al.* A family of radical halogenases for the engineering of amino-acid-based products. *Nat. Chem. Biol.* **15**, 1009–1016 (2019).

20. Hayashi, T. *et al.* Evolved aliphatic halogenases enable regiocomplementary C−H functionalization of a pharmaceutically relevant compound. *Angew. Chem. Int. Ed.* **58**, 18535–18539 (2019).

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

22. Yang, Y. & Arnold, F. H. Navigating the unnatural reaction space: Directed evolution of heme proteins for selective carbene and nitrene transfer. *Acc. Chem. Res.* **54**, 1209–1225 (2021).

23. Zhang, R. K., Huang, X. & Arnold, F. H. Selective C–H bond functionalization with engineered heme proteins: New tools to generate complexity. *Curr. Opin. Chem. Biol.* **49**, 67–75 (2019).

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

25. Natoli, S. N. & Hartwig, J. F. Noble-metal substitution in hemoproteins: An emerging strategy for abiological catalysis. *Acc. Chem. Res.* **52**, 326–335 (2019).

26. Matthews, M. L. *et al.* Direct nitration and azidation of aliphatic carbons by an iron-dependent halogenase. *Nat. Chem. Biol.* **10**, 209–215 (2014).

27. Rui, J. *et al.* Directed evolution of nonheme iron enzymes to access abiological radical-relay C(sp^3)−H azidation. *Science* **376**, 869–874 (2022).

28. Doyle, M. P., Duffy, R., Ratnikov, M. & Zhou, L. Catalytic carbene insertion into C−H bonds. *Chem. Rev.* **110**, 704–724 (2010).

29. Zhang, R. K. *et al.* Enzymatic assembly of carbon–carbon bonds via iron-catalysed sp^3 C–H functionalization. *Nature* **565**, 67–72 (2019).

30. Zhang, J., Huang, X., Zhang, R. K. & Arnold, F. H. Enantiodivergent α-amino C–H fluoroalkylation catalyzed by engineered cytochrome P450s. *J. Am. Chem. Soc.* **141**, 9798–9802 (2019).

31. Zhou, A. Z., Chen, K. & Arnold, F. H. Enzymatic lactone-carbene C–H insertion to build contiguous chiral centers. *ACS Catal.* **10**, 5393–5398 (2020).

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

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

34. Gu, Y., Natoli, S. N., Liu, Z., Clark, D. S. & Hartwig, J. F. Site-selective functionalization of (sp^3)C−H bonds catalyzed by artificial metalloenzymes containing an iridium-porphyrin cofactor. *Angew. Chem. Int. Ed.* **58**, 13954–13960 (2019).

35. Caballero, A. *et al.* Catalytic functionalization of low reactive C(sp^3)–H and C(sp^2)–H bonds of alkanes and arenes by carbene transfer from diazo compounds. *Dalton Trans* **44**, 20295–20307 (2015).

36. Liu, Z. & Wang, J. Cross-coupling reactions involving metal carbene: From C═C/C–C bond formation to C–H bond functionalization. *J. Org. Chem.* **78**, 10024–10030 (2013).

37. Fructos, M. R. *et al.* A gold catalyst for carbene-transfer reactions from ethyl diazoacetate. *Angew. Chem. Int. Ed.* **44**, 5284–5288 (2005).

38. Conde, A. *et al.* Iron and manganese catalysts for the selective functionalization of arene C(sp^2)−H bonds by carbene insertion. *Angew. Chem. Int. Ed.* **55**, 6530–6534 (2016).

39. Yu, Z. *et al.* Highly site-selective direct C–H bond functionalization of phenols with α-aryl-α-diazoacetates and diazooxindoles via gold catalysis. *J. Am. Chem. Soc.* **136**, 6904–6907 (2014).

40. Xu, B., Li, M.-L., Zuo, X.-D., Zhu, S.-F. & Zhou, Q.-L. Catalytic asymmetric arylation of α-aryl-α-diazoacetates with aniline derivatives. *J. Am. Chem. Soc.* **137**, 8700–8703 (2015).

41. Holmberg-Douglas, N., Onuska, N. P. R. & Nicewicz, D. A. Regioselective arene C−H alkylation enabled by organic photoredox catalysis. *Angew. Chem. Int. Ed.* **59**, 7425–7429 (2020).

42. Vargas, D. A., Tinoco, A., Tyagi, V. & Fasan, R. Myoglobin-catalyzed C−H functionalization of unprotected indoles. *Angew. Chem. Int. Ed.* **57**, 9911–9915 (2018).

43. Brandenberg, O. F., Chen, K. & Arnold, F. H. Directed evolution of a cytochrome P450 carbene transferase for selective functionalization of cyclic compounds. *J. Am. Chem. Soc.* **141**, 8989–8995 (2019).

44. Zhang, W. *et al.* Enantioselective cyanation of benzylic C–H bonds via copper-catalyzed radical relay. *Science* **353**, 1014–1018 (2016).

45. Lennox, A. J. J. *et al.* Electrochemical aminoxyl-mediated α-cyanation of secondary piperidines for pharmaceutical building block diversification. *J. Am. Chem. Soc.* **140**, 11227–11231 (2018).

46. Mykhailiuk, P. K. & Koenigs, R. M. Diazoacetonitrile (N_2CHCN): A long forgotten but valuable reagent for organic synthesis. *Chem. – Eur. J.* **26**, 89–101 (2020).

47. Chandgude, A. L. & Fasan, R. Highly diastereo- and enantioselective synthesis of nitrile-substituted cyclopropanes by myoglobin-mediated carbene transfer catalysis. *Angew. Chem. Int. Ed.* **57**, 15852–15856 (2018).

48. Hock, K. J. *et al.* Tryptamine synthesis by iron porphyrin catalyzed C−H functionalization of indoles with diazoacetonitrile. *Angew. Chem. Int. Ed.* **58**, 3630–3634 (2019).

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

50. Hyster, T. K., Farwell, C. C., Buller, A. R., McIntosh, J. A. & Arnold, F. H. Enzyme-controlled nitrogen-atom transfer enables regiodivergent C–H amination. *J. Am. Chem. Soc.* **136**, 15505–15508 (2014).

51. Prier, C. K., Zhang, R. K., Buller, A. R., Brinkmann-Chen, S. & Arnold, F. H. Enantioselective, intermolecular benzylic C–H amination catalysed by an engineered iron-haem enzyme. *Nat. Chem.* **9**, 629–634 (2017).

52. Cooley, R. B., Arp, D. J. & Karplus, P. A. Evolutionary origin of a secondary structure: π-Helices as cryptic but widespread insertional variations of α-helices that enhance protein functionality. *J. Mol. Biol.* **404**, 232–246 (2010).

53. Hasemann, C. A., Kurumbail, R. G., Boddupalli, S. S., Peterson, J. A. & Deisenhofer, J. Structure and function of cytochromes P450: A comparative analysis of three crystal structures. *Structure* **3**, 41–62 (1995).

54. Ost, T. W. B. *et al.* Structural and spectroscopic analysis of the F393H mutant of flavocytochrome P450 BM3. *Biochemistry* **40**, 13430–13438 (2001).

55. Krest, C. M. *et al.* Significantly shorter Fe–S bond in cytochrome P450-I is consistent with greater reactivity relative to chloroperoxidase. *Nat. Chem.* **7**, 696–702 (2015).

56. Ford, A. *et al.* Modern organic synthesis with α-diazocarbonyl compounds. *Chem. Rev.* **115**, 9981–10080 (2015).

57. Reisman, S. E., Nani, R. R. & Levin, S. Buchner and beyond: Arene cyclopropanation as applied to natural product total synthesis. *Synlett* **2011**, 2437–2442 (2011).

58. Kille, S. *et al.* Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* **2**, 83–92 (2013).

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

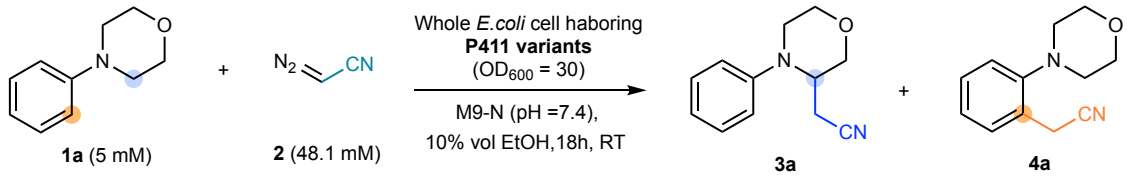
60. Berry, E. A. & Trumpower, B. L. Simultaneous determination of hemes a, b, and c from pyridine hemochrome spectra. *Anal. Biochem.* **161**, 1–15 (1987).

A p p e n d i x B

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

B.1 Supplementary Table B-1 through B-5

Table B-1. Initial activity screening with engineered P411s.[a] (Biotransformation yields of the top 20 P411 variants with α-amino C(sp^3)–H cyanomethylation activity are listed.)



Variant	Yield of **3a**	Yield of **4a**
P411-CH-C3	33%	3.3%
P411-CH-C5	26%	3.3%
P411-CH-C4	26%	2.0%
H10	25%	3.2%
H8	25%	2.9%
H9	25%	2.2%
D12	22%	1.5%
B12	22%	1.1%
D11	15%	1.1%
D10	14%	0.9%
B11	13%	1.0%
G4	12%	1.4%
B4	12%	0.9%

C5	11%	1.3%
E4	11%	1.1%
A11	10%	1.0%
H7	9%	1.4%
P411-PFA	9%	2.6%
H12	9%	0.9%
E6	9%	2.0%
…	…	…
[b] hemin (5 μM) + $Na_2S_2O_4$	NR	NR
[b] hemin (5 mM) + $Na_2S_2O_4$	NR	NR
[b] hemin (10 mM) + $Na_2S_2O_4$	NR	NR
[c] hemin (10 mM)	NR	NR
[d] cellular background	NR	NR

[a] Experiments were performed using whole *E. coli* cell format according to the protocol described in **Section 3.4.4**. NR – no product was detected.

[b] Negative control experiment using free hemin under reduced conditions (Fe(II)) were performed using excess $Na_2S_2O_4$ (20 mM).

[c] Negative control experiment (with free hemin) was also performed without the addition of $Na_2S_2O_4$. Under this condition, the resting oxidation state of hemin in aqueous buffer should be Fe(III)

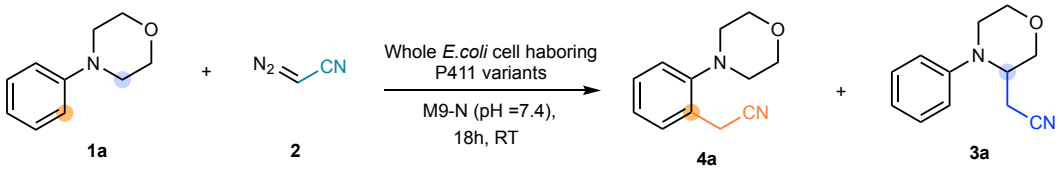
[d] Cellular background control experiment was performed in whole-cell format, using *E. coli* (*E. cloni* BL21(DE3)) cell harboring an engineered tryptophan synthase *β*-subunit (Tm9D8*). The gene of this enzyme was cloned into the same pET22b(+) vector (Novagen) between restriction sites *Nde*I and *Xho*I. The protein expression protocol for this experiment follows the standard P450 expression conditions as described in **Section 3.4.4**.

Table B-2. Identification of *ortho*-arene C–H cyanomethylation activity. (Variants with top 10 highest activity are listed.)

P411 variant	Yield of **4a**	Yield of **3a**	**4a** : **3a**
P411-FA-E3	**3.9%**	**6.6%**	**~ 3 : 5**
A12	3.3%	33%	1 : 10
B2	3.3%	26%	~ 1 : 8
H10	3.2%	25%	~ 1 : 8
H8	2.9%	25%	~ 1 : 9
P411-PFA	2.6%	9.1%	~ 3 : 10
H9	2.1%	25%	~ 1 : 11
B1	2.0%	26%	~ 1 : 13
E6	2.0%	8.7%	~ 2 : 9
E5	1.9%	6.3%	3 : 10
…	…	…	…

P411-FA-E3 exhibited the highest yield and selectivity of delivering **4a**. Therefore, it was chosen as the starting variant for the directed evolution experiment to develop highly active and selective arene C–H alkyases.

Table B-3. Directed evolution of an *ortho*-selective arene C–H cyanomethylase.[a]



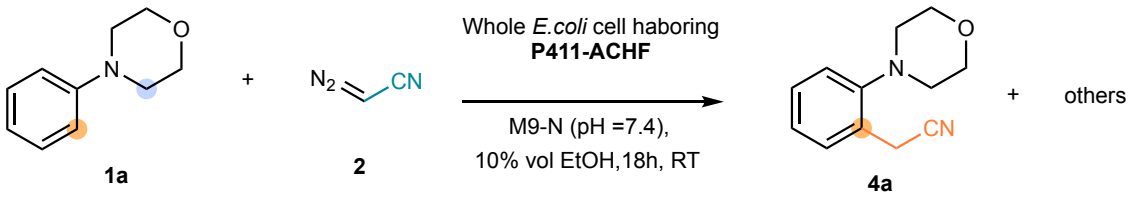
Generations	Variant name	Mutation (s)	Yield of **4a**	Selectivity (**4a** : **3a**)
-1	P411-PFA	-	2.6% ± 0.7%	3 : 10
0	P411-FA-E3	P401L	6.7 ± 0.7%	3 : 5
1	"V330C" (Gen1)	V330C	9.2 ± 0.7%	3 : 5
2	"S118Q, W325C" (Gen2)	S118Q, W325C	5.9 ± 0.1%	3 : 2
3	"F393W" (Gen3)	F393W	14 ± 3%	1.9 : 1
4	"A87V" (Gen4)	A87V	33 ± 1%	12 : 1
5	"M177Y" (Gen5)	M177Y	28 ± 1%	17 : 1
6	"G252L" (Gen6)	G252L	27 ± 1%	20 : 1
7	**P411-ACHF** [b]	**M354V**	**44 ± 2%**	**19 : 1**
-	P411-ACHF Y263W	Y263W	-	-

Experiments were performed using suspension of *E. coli* cells harboring the corresponding protein prepared according to the protocol described in **Section 3.4.4–3.4.6**. Reactions were performed in triplicate. Yields reported are the average of three experiments.

[a] Reaction condition: OD_{600} of *E. coli* cell suspension = 30; 5 mM **1a**; 48.1 mM **2** (9.6 equiv.).

[b] Final variant.

Table B-4. Reaction engineering to optimize the yield of **4a**.



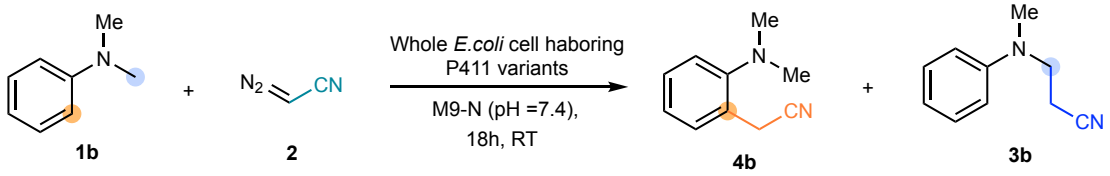
Entry	OD_{600} of cell suspension	Concentration of **1a**	Concentration of **2** (equiv.)	Yield/%
1[a]	30	5 mM	48.1 mM (9.6 equiv.)	44 ± 2
2[b]	30	2.5 mM	51.6 mM (20.6 equiv.)	51 ± 5
3	30	10 mM	41.3 mM (4.1 equiv.)	31 ± 1
4	10	5 mM	48.1 mM (9.6 equiv.)	34 ± 1
5	20	5 mM	48.1 mM (9.6 equiv.)	45 ± 1
6	45	5 mM	48.1 mM (9.6 equiv.)	38 ± 4
7	60	5 mM	48.1 mM (9.6 equiv.)	35 ± 4

Experiments were performed using suspension of *E. coli* cells harboring the corresponding protein prepared according to the protocol described in **Section 3.4.6**. Reactions were performed in triplicate. Yields reported are the average of three experiments.

[a] Reaction condition adopted for directed evolution experiment.

[b] Optimized reaction condition.

Table B-5. Directed evolution for further optimization of specific target **4b**.[a]



Generations	Variant name	Mutation (s)	Yield of **4b**	Selectivity (**4b** : **3b**)
0	P411-ACHF	-	31% ± 4%	5 : 2
+1	P411-ACHF Y263W	Y263W	54 ± 1%	> 10 : 1

Experiments were performed using suspension of *E. coli* cells harboring the corresponding protein prepared according to the protocol described in **Section 3.4.6**. Reactions were performed in triplicate. Yields reported are the average of three experiments.

[a] Reaction condition: OD_{600} of *E. coli* cell suspension = 30; 5 mM **1b**; 48.1 mM **2** (9.6 equiv.).

B.2 Supplementary Figure B-1 through B4

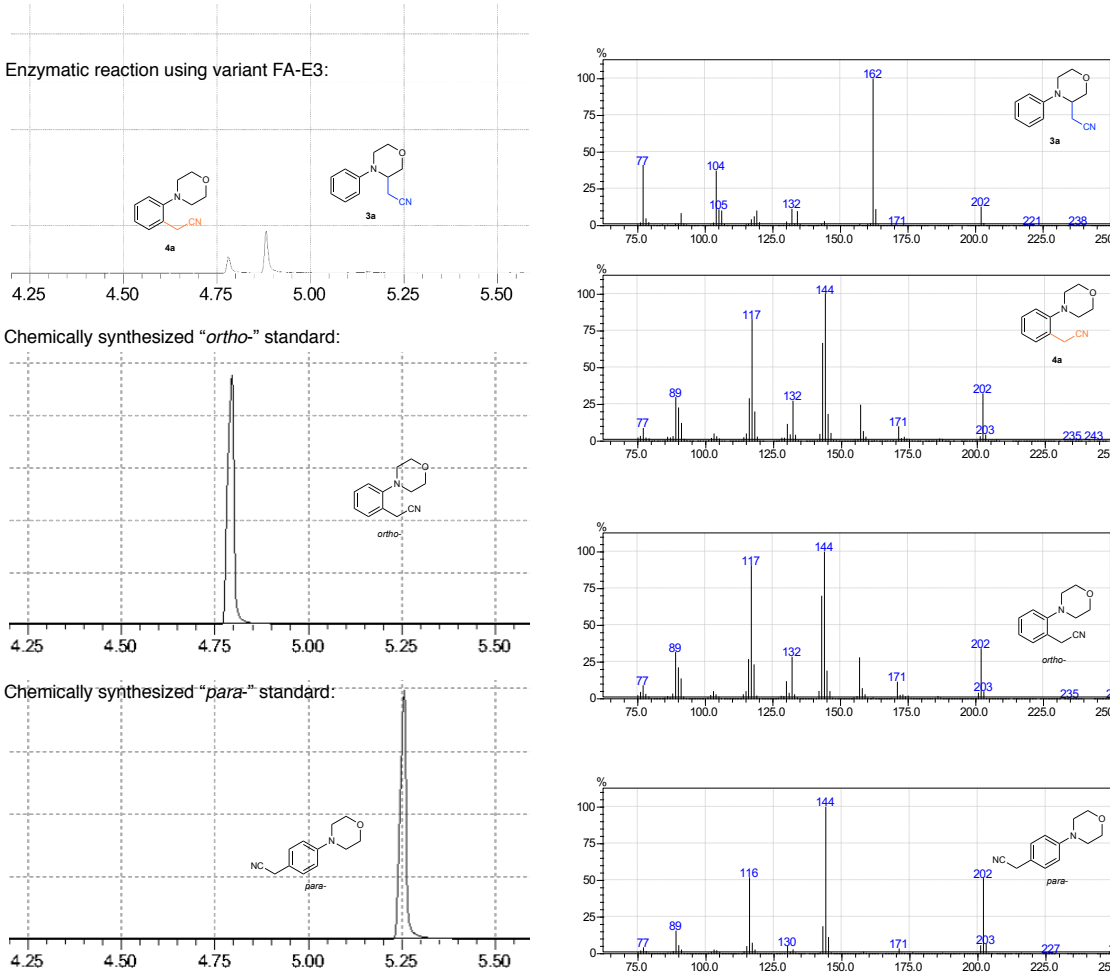


Figure B-1. Confirming the product identity of enzymatic arene C–H alkylation. MS spectrum of the enzymatic arene C–H functionalization product **4a** in comparison to *ortho*- and *para*-substituted product synthesized by chemical methods. The enzymatic arene C–H functionalization product match with the *ortho*-substituted product standard (Both the retention times and the MS fingerprints on GC-MS match).

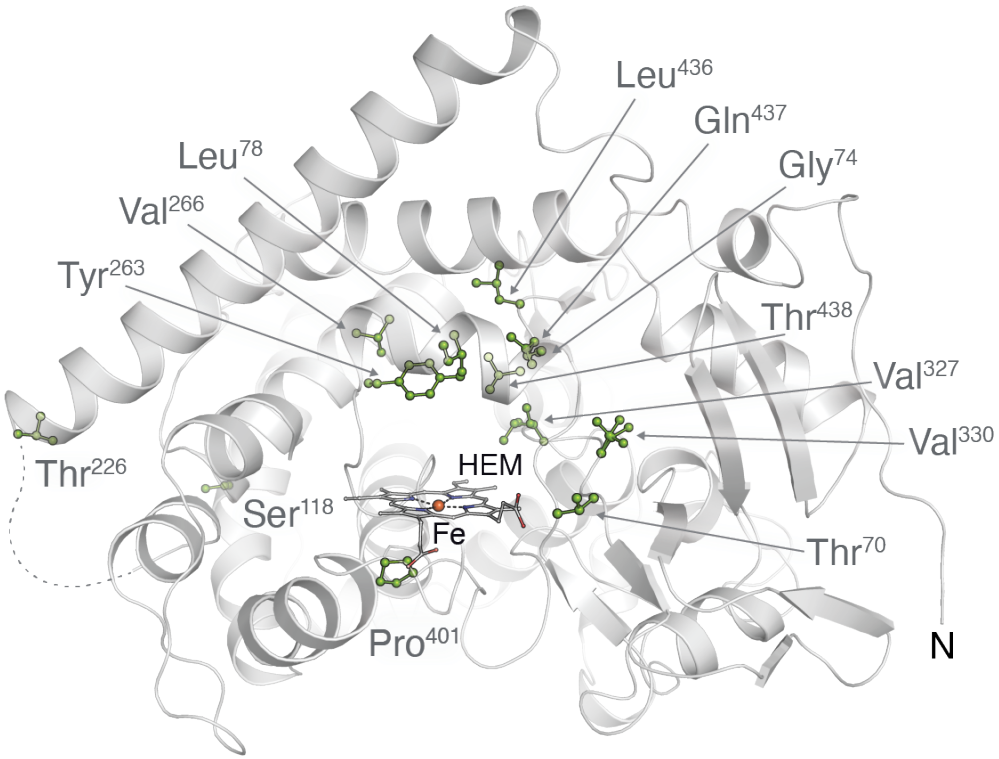


Figure B-2. Mutations in P411-PFA relative to P-4 A82L A78V F263L. In total, P411-PFA contains 13 mutations relative to P-4 A82L A78V F263L, the closely related C–H aminase whose structure was reported (PDB ID: 5UCW).

```
CYP102A1    MTIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADELGEIFKFEAPGRVTRYLSSQRLIK   60
P411-PFA    MTIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADELGEIFKFEAPGRVTRYLSSQRLIK   60
            ************************************************************

CYP102A1    EACDESRFDKNLSQALKFVRDFAGDGLFTSWTHEKNWKKAHNILLPSFSQQAMKGYHAMM   120
P411-PFA    EACDESRFDKTLSQGLKFLRDFLGDGLATSWTHEKNWKKAHNILLPSFSQQAMKGYHASM   120
            **********.***.***:*** **** ****************************** *

CYP102A1    VDIAVQLVQKWERLNADEHIEVPEDMTRLTLDTIGLCGFNYRFNSFYRDQPHPFITSMVR   180
P411-PFA    VDIAVQLVQKWERLNADEHIEVSEDMTRLTLDTIGLCGFNYRFNSFYRDQPHPFIISMVR   180
            ********************* ****************************** ****

CYP102A1    ALDEAMNKLQRANPDDPAYDENKRQFQEDIKVMNDLVDKIIADRKASGEQSDDLLTHMLN   240
P411-PFA    ALDEVMNKLQRANPDDPAYDENKRQFQEDIKVMNDLVDKIIADRKATGEQSDDLLTQMLN   240
            ****.*******************************************:*********:***

CYP102A1    GKDPETGEPLDDENIRYQIITFLIAGHETTSGLLSFALYFLVKNPHVLQKAAEEAARVLV   300
P411-PFA    GKDPETGEPLDDGNIRYQIITFLYAGVEGTSGLLSFALYFLVKNPHVLQKVAEEAARVLV   300
            ************ ********** ** * **********************.*********

CYP102A1    DPVPSYKQVKQLKYVGMVLNEALRLWPTAPAFSLYAKEDTVLGGEYPLEKGDELMVLIPQ   360
P411-PFA    DPVPSYKQVKQLKYVGMVLNEALRLWPVVPVFSLYAKEDTVLGGEYPLEKGDEVMVLIPQ   360
            ***************************..*.**********************:******

CYP102A1    LHRDKTIWGDDVEEFRPERFENPSAIPQHAFKPFGNGQRACIGQQFALHEATLVLGMMLK   420
P411-PFA    LHRDKTVWGDDVEEFRPERFENPSAIPQHAFKPFGNGQRASPGQQFALHEATLVLGMMLK   420
            ******:**********************************. ******************

CYP102A1    HFDFEDHTNYELDIKETLTLKPEGFVVKAKSKKIPLGGIPSPSTLEHHHHHH      472
P411-PFA    HFDFEDHTNYELDIKELQTLKPKGFVVKAKSKKIPLGGIPSPSTLEHHHHHH      472
            ****************  ****:******************************
```

Figure B-3. Sequence comparison between P411-PFA (heme domain) and wild-type $P450_{BM3}$ (CYP102A1, showing the heme domain sequence only). P411-PFA contains 26 mutations relative to wild-type $P450_{BM3}$, however, Glu^{267} (the "flipped" residue; denoted in red box) has not been targeted in previous directed evolution experiment.

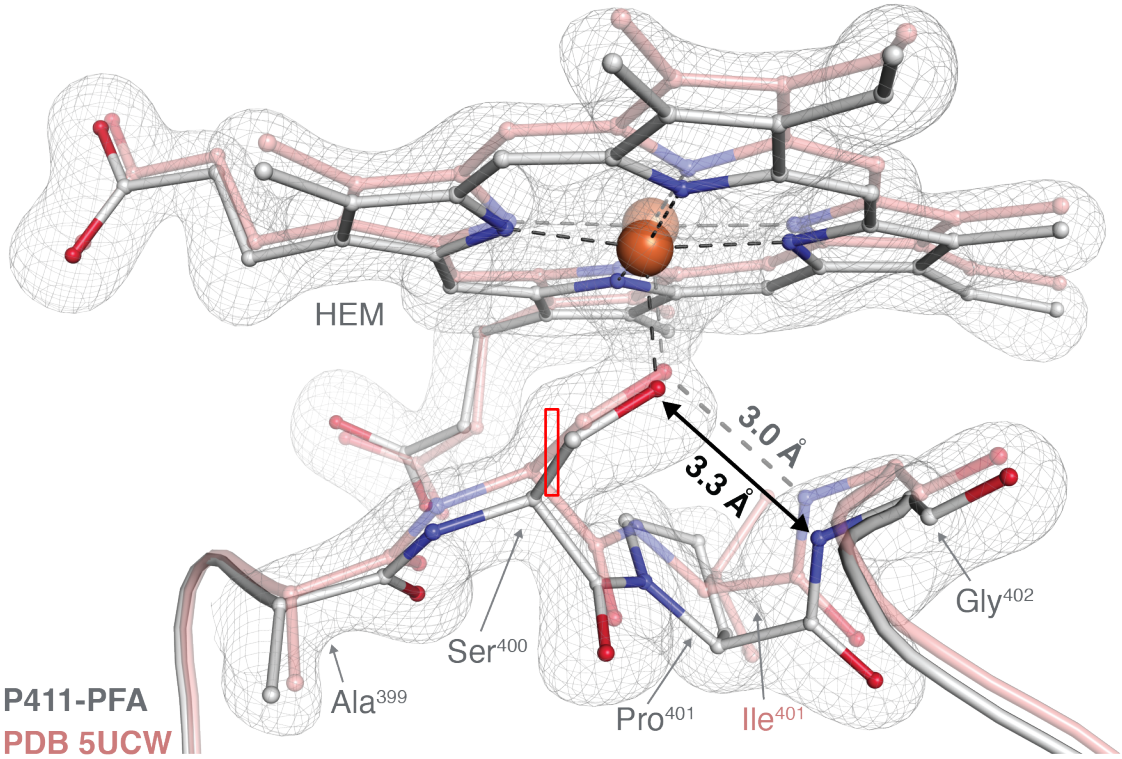


Figure B-4. Structural difference between P411-PFA and P-4 A82L A78V F263L (PDB ID: 5UCW) at the "Cys pocket" region. Residues in this region are highly conserved to affect the electronic properties of heme. Mutation L401P in P411-PFA has disrupted the hydrogen bond between the backbone amide N–H of Gly^{402} to the axial Ser^{400} (an effective hydrogen bond should be < 3.3 Å). This structural change has led to the enzyme activity and selectivity differences between P411-PFA (Pro^{401}) and FA-E3 (Leu^{401}). We postulated that other residues in this region can be targeted to further improve arene C–H functionalization selectivity, and identified F393W as a beneficial mutation.

B.3 Nucleotide and Amino Acid Sequences

The genes encoding the heme proteins shown below were cloned using Gibson assembly[1] into vector pET-22b(+) (Novagen) between restriction sites *Nde*I and *Xho*I in frame with a *C*-terminal 6xHis-tag.

DNA and amino acid sequences of **P411-PFA**:

```
ATGACAATTAAAGAAATGCCTCAGCCAAAAACGTTTGGAGAGCTTAAAAATTTACCGTTATTAAACACAGA
TAAACCGGTTCAAGCTTTGATGAAAATTGCGGATGAATTAGGAGAAATCTTTAAATTCGAGGCGCCTGGTC
GTGTAACGCGCTACTTATCAAGTCAGCGTCTAATTAAAGAAGCATGCGATGAATCACGCTTTGATAAAACA
TTAAGTCAAGGTCTGAAATTTCTGCGTGATTTTCTTGGAGACGGGTTAGCCACAAGCTGGACGCATGAAAA
AAATTGGAAAAAAGCGCATAATATCTTACTTCCAAGCTTTAGTCAGCAGGCAATGAAAGGCTATCATGCGA
GTATGGTCGATATCGCCGTGCAGCTTGTTCAAAAGTGGGAGCGTCTAAATGCAGATGAGCATATTGAAGTA
TCGGAAGACATGACACGTTTAACGCTTGATACAATTGGTCTTTGCGGCTTTAACTATCGCTTTAACAGCTT
TTACCGAGATCAGCCTCATCCATTTATTATAAGTATGGTCCGTGCACTGGATGAAGTAATGAACAAGCTGC
AGCGAGCAAATCCAGACGACCCAGCTTATGATGAAAACAAGCGCCAGTTTCAAGAAGATATCAAGGTGATG
AACGACCTAGTAGATAAAATTATTGCAGATCGCAAAGCAACAGGTGAACAAAGCGATGATTTATTAACGCA
GATGCTAAACGGAAAAGATCCAGAAACGGGTGAGCCGCTTGATGACGGGAACATTCGCTATCAAATTATTA
CATTCTTATATGCGGGAGTTGAAGGTACAAGTGGTCTTTTATCATTTGCGCTGTATTTCTTAGTGAAAAAT
CCACATGTATTACAAAAAGTAGCAGAAGAAGCAGCACGAGTTCTAGTAGATCCTGTTCCAAGCTACAAACA
AGTCAAACAGCTTAAATATGTCGGCATGGTCTTAAACGAAGCGCTGCGCTTATGGCCAGTGGTTCCTGTGT
TTTCCCTATATGCAAAAGAAGATACGGTGCTTGGAGGAGAATATCCTTTAGAAAAAGGCGACGAAGTAATG
GTTCTGATTCCTCAGCTTCACCGTGATAAAACAGTTTGGGGAGACGATGTGGAGGAGTTCCGTCCAGAGCG
TTTTGAAAATCCAAGTGCGATTCCGCAGCATGCGTTTAAACCGTTTGGAAACGGTCAGCGTGCGTCTCCGG
GTCAGCAGTTCGCTCTTCATGAAGCAACGCTGGTACTTGGTATGATGCTAAAACACTTTGACTTTGAAGAT
CATACAAACTACGAGCTCGATATTAAAGAACTGCAGACGTTAAAACCTAAAGGCTTTGTGGTAAAAGCAAA
ATCGAAAAAAATTCCGCTTGGCGGTATTCCTTCACCTAGCACTGAACAGTCTGCTAAAAAAGTACGCAAAA
AGGCAGAAAACGCTCATAATACGCCGCTGCTTGTGCTATACGGTTCAAATATGGGTACCGCTGAAGGAACG
GCGCGTGATTTAGCAGATATTGCAATGAGCAAAGGATTTGCACCGCAGGTCGCAACGCTTGATTCACACGC
CGGAAATCTTCCGCGCGAAGGAGCTGTATTAATTGTAACGGCGTCTTATAACGGTCATCCGCCTGATAACG
CAAAGCAATTTGTCGACTGGTTAGACCAAGCGTCTGCTGATGAAGTAAAAGGCGTTCGCTACTCCGTATTT
GGATGCGGCGATAAAAACTGGGCTACTACGTATCAAAAAGTGCCTGCTTTTATCGATGAAACGCTTGCCGC
TAAAGGGGCAGAAAACATCGCTGACCGCGGTGAAGCAGATGCAAGCGACGACTTTGAAGGCACATATGAAG
AATGGCGTGAACATATGTGGAGTGACGTAGCAGCCTACTTTAACCTCGACATTGAAAACAGTGAAGATAAT
AAATCTACTCTTTCACTTCAATTTGTCGACAGCGCCGCGGATATGCCGCTTGCGAAAATGCACGGTGCGTT
TTCAACGCTCGAGCACCACCACCACCACCACTGA
```

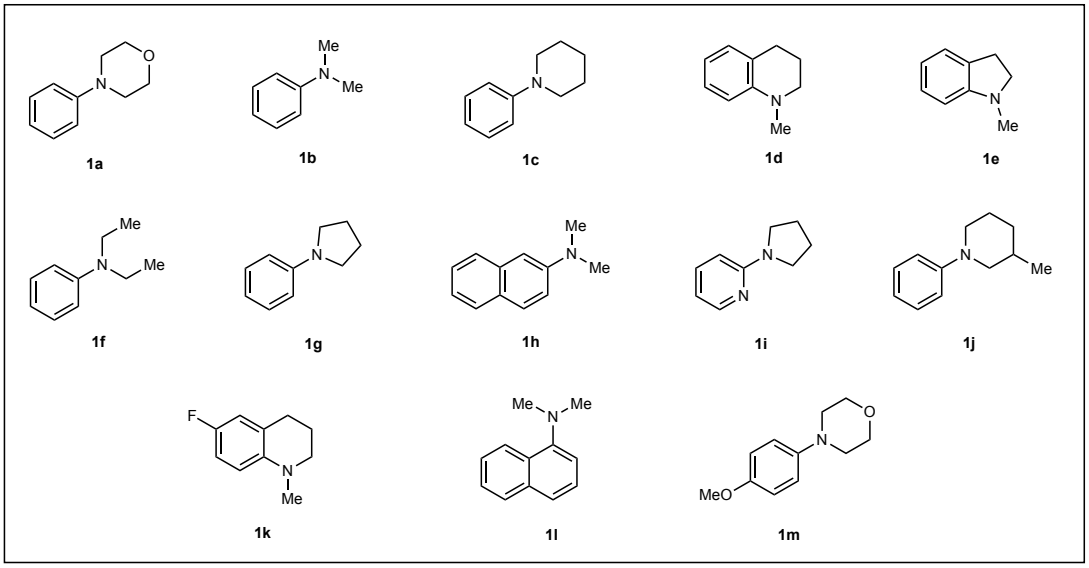
```
MTIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADELGEIFKFEAPGRVTRYLSSQRLIKEACDESRFDKT
LSQGLKFLRDFLGDGLATSWTHEKNWKKAHNILLPSFSQQAMKGYHASMVDIAVQLVQKWERLNADEHIEV
SEDMTRLTLDTIGLCGFNYRFNSFYRDQPHPFIISMVRALDEVMNKLQRANPDDPAYDENKRQFQEDIKVM
NDLVDKIIADRKATGEQSDDLLTQMLNGKDPETGEPLDDGNIRYQIITFLYAGVEGTSGLLSFALYFLVKN
PHVLQKVAEEAARVLVDPVPSYKQVKQLKYVGMVLNEALRLWPVVPVFSLYAKEDTVLGGEYPLEKGDEVM
VLIPQLHRDKTVWGDDVEEFRPERFENPSAIPQHAFKPFGNGQRASPGQQFALHEATLVLGMMLKHFDFED
HTNYELDIKELQTLKPKGFVVKAKSKKIPLGGIPSPSTEQSAKKVRKKAENAHNTPLLVLYGSNMGTAEGT
ARDLADIAMSKGFAPQVATLDSHAGNLPREGAVLIVTASYNGHPPDNAKQFVDWLDQASADEVKGVRYSVF
GCGDKNWATTYQKVPAFIDETLAAKGAENIADRGEADASDDFEGTYEEWREHMWSDVAAYFNLDIENSEDN
KSTLSLQFVDSAADMPLAKMHGAFSTLEHHHHHH*
```

DNA and amino acid sequences of **P411-ACHF**:

ATGACAATTAAAGAAATGCCTCAGCCAAAAACGTTTGGAGAGCTTAAAAATTTACCGTTATTAAACACAGA
TAAACCGGTTCAAGCTTTGATGAAAATTGCGGATGAATTAGGAGAAATCTTTAAATTCGAGGCGCCTGGTC
GTGTAACGCGCTACTTATCAAGTCAGCGTCTAATTAAAGAAGCATGCGATGAATCACGCTTTGATAAAACA
TTAAGTCAAGGTCTGAAATTTCTGCGTGATTTTCTTGGAGACGGGTTAGTGACAAGCTGGACGCATGAAAA
AAATTGGAAAAAAGCGCATAATATCTTACTTCCAAGCTTTAGTCAGCAGGCAATGAAAGGCTATCATGCGC
AGATGGTCGATATCGCCGTGCAGCTTGTTCAAAAGTGGGAGCGTCTAAATGCAGATGAGCATATTGAAGTA
TCGGAAGACATGACACGTTTAACGCTTGATACAATTGGTCTTTGCGGCTTTAACTATCGCTTTAACAGCTT
TTACCGAGATCAGCCTCATCCATTTATTATAAGTTATGTCCGTGCACTGGATGAAGTAATGAACAAGCTGC
AGCGAGCAAATCCAGACGACCCAGCTTATGATGAAAACAAGCGCCAGTTTCAAGAAGATATCAAGGTGATG
AACGACCTAGTAGATAAAATTATTGCAGATCGCAAAGCAACAGGTGAACAAAGCGATGATTTATTAACGCA
GATGCTAAACGGAAAAGATCCAGAAACGGGTGAGCCGCTTGATGACTTGAACATTCGCTATCAAATTATTA
CATTCTTATATGCGGGAGTTGAAGGTACAAGTGGTCTTTTATCATTTGCGCTGTATTTCTTAGTGAAAAAT
CCACATGTATTACAAAAAGTAGCAGAAGAAGCAGCACGAGTTCTAGTAGATCCTGTTCCAAGCTACAAACA
AGTCAAACAGCTTAAATATGTCGGCATGGTCTTAAACGAAGCGCTGCGCTTATGTCCAGTGGTTCCTTGTT
TTTCCCTATATGCAAAAGAAGATACGGTGCTTGGAGGAGAATATCCTTTAGAAAAAGGCGACGAAGTAGTG
GTTCTGATTCCTCAGCTTCACCGTGATAAAACAGTTTGGGGAGACGATGTGGAGGAGTTCCGTCCAGAGCG
TTTTGAAAATCCAAGTGCGATTCCGCAGCATGCGTTTAAACCGTGGGGAAACGGTCAGCGTGCG
TCTCTGGGTCAGCAGTTCGCTCTTCATGAAGCAACGCTGGTACTTGGTATGATGCTAAAACACTTTGACTT
TGAAGATCATACAAACTACGAGCTCGATATTAAAGAACTGCAGACGTTAAAACCTAAAGGCTTTGTGGTAA
AAGCAAAATCGAAAAAAATTCCGCTTGGCGGTATTCCTTCACCTAGCACTGAACAGTCTGCTAAAAAAGTA
CGCAAAAAGGCAGAAAACGCTCATAATACGCCGCTGCTTGTGCTATACGGTTCAAATATGGGTACCGCTGA
AGGAACGGCGCGTGATTTAGCAGATATTGCAATGAGCAAAGGATTTGCACCGCAGGTCGCAACGCTTGATT
CACACGCCGGAAATCTTCCGCGCGAAGGAGCTGTATTAATTGTAACGGCGTCTTATAACGGTCATCCGCCT
GATAACGCAAAGCAATTTGTCGACTGGTTAGACCAAGCGTCTGCTGATGAAGTAAAAGGCGTTCGCTACTC
CGTATTTGGATGCGGCGATAAAAACTGGGCTACTACGTATCAAAAAGTGCCTGCTTTTATCGATGAAACGC
TTGCCGCTAAAGGGGCAGAAAACATCGCTGACCGCGGTGAAGCAGATGCAAGCGACGACTTTGAAGGCACA
TATGAAGAATGGCGTGAACATATGTGGAGTGACGTAGCAGCCTACTTTAACCTCGACATTGAAAACAGTGA
AGATAATAAATCTACTCTTTCACTTCAATTTGTCGACAGCGCCGCGGATATGCCGCTTGCGAAAATGCACG
GTGCGTTTTCAACGCTCGAGCACCACCACCACCACCACTGA

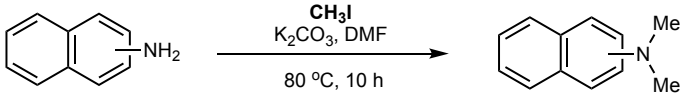
MTIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADELGEIFKFEAPGRVTRYLSSQRLIKEACDESRFDKTLS
QGLKFLRDFLGDGLVTSWTHEKNWKKAHNILLPSFSQQAMKGYHAQMVDIAVQLVQKWERLNADEHIEVSEDM
TRLTLDTIGLCGFNYRFNSFYRDQPHPFIISYVRALDEVMNKLQRANPDDPAYDENKRQFQEDIKVMNDLVDK
IIADRKATGEQSDDLLTQMLNGKDPETGEPLDDLNIRYQIITFLYAGVEGTSGLLSFALYFLVKNPHVLQKVA
EEAARVLVDPVPSYKQVKQLKYVGMVLNEALRLCPVVPCFSLYAKEDTVLGGEYPLEKGDEVVVLIPQLHRDK
TVWGDDVEEFRPERFENPSAIPQHAFKPWGNGQRASLGQQFALHEATLVLGMMLKHFDFEDHTNYELDIKELQ
TLKPKGFVVKAKSKKIPLGGIPSPSTEQSAKKVRKKAENAHNTPLLVLYGSNMGTAEGTARDLADIAMSKGFA
PQVATLDSHAGNLPREGAVLIVTASYNGHPPDNAKQFVDWLDQASADEVKGVRYSVFGCGDKNWATTYQKVPA
FIDETLAAKGAENIADRGEADASDDFEGTYEEWREHMWSDVAAYFNLDIENSEDNKSTLSLQFVDSAADMPLA
KMHGAFSTLEHHHHHH*

B.4. Substrate Syntheses and Characterizations[a]



[a] **1a-1g**, **1m** were obtained from commercial sources and were used as received.

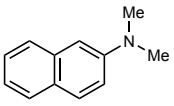
Synthetic Procedure A:



1h, **1l**

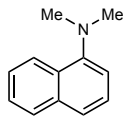
N, *N*-dimethyl naphthylamines **1h** and **1l** were prepared according to a modified procedure reported by Shi[2] and McNally[3]. A mixture of aniline (5.0 mmol), iodomethane (11.0 mmol), and K_2CO_3 (11.0 mmol) in DMF (20 mL) was stirred at 80 °C for 10 h. The reaction was then cooled and diluted with EtOAc (100 mL) and H_2O (100 mL). The layers were separated and the organic layer was washed with brine (20 mL), dried over $MgSO_4$, filtered, concentrated, and purified by flash chromatography.

N,*N*-dimethylnaphthalen-2-amine (1h)



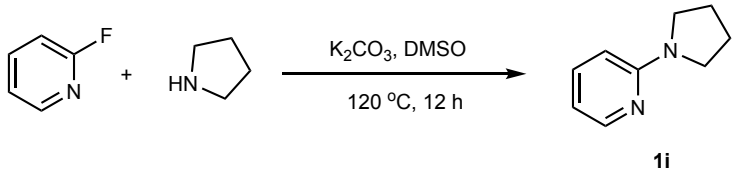
This compound is known.[2] **^{1}H NMR (400 MHz, $CDCl_3$)** δ 7.76 – 7.64 (m, 3H), 7.38 (ddd, *J* = 8.2, 6.8, 1.3 Hz, 1H), 7.25 – 7.16 (m, 2H), 6.98 (broad peak, 1H), 3.06 (s, 6H).

***N*,*N*-dimethylnaphthalen-1-amine (1l)**



This compound is known.[2] **^{1}H NMR (400 MHz, $CDCl_3$)** δ 8.26 (dq, *J* = 7.4, 0.9 Hz, 1H), 7.84 (dd, *J* = 7.6, 1.7 Hz, 1H), 7.50 (dddt, *J* = 15.0, 12.8, 6.8, 1.2 Hz, 3H), 7.41 (ddd, *J* = 8.2, 7.4, 0.8 Hz, 1H), 7.09 (dd, *J* = 7.4, 1.1 Hz, 1H), 2.93 (s, 6H).

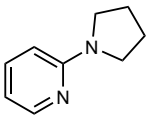
Synthetic Procedure B:



1i

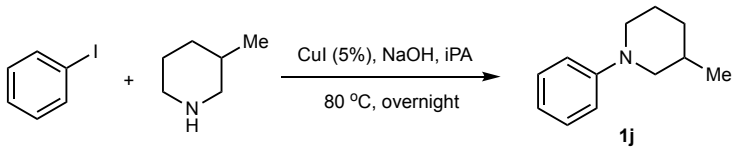
Substrate **1i** was prepared from the S_NAr reaction with the following protocol. To a reaction vial, suspension of K_2CO_3 (760 mg, 5.5 mmol, 1.1 equiv.) in DMSO (10 mL, 0.5 M) was degassed and backfilled with argon for three times. Pyrrolidine (357 mg, 5.0 mmol, 1.0 equiv.) and 2-fluoropyridine (490 mg, 5.0 mmol, 1.0 equiv.) was then added, and the reaction was heated at 120 °C for 12 h. The reaction was then cooled and thoroughly mixed with EtOAc (10 mL) and H_2O (10 mL). The layers were separated and the aqueous layer was further extracted with EtOAc (3 x 20 mL). The organic layers were combined, washed with brine (10 mL), dried over $MgSO_4$, filtered, concentrated. The crude mixture was further purified by flash chromatography.

2-(pyrrolidin-1-yl)pyridine (1i)



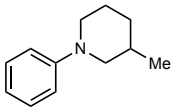
This compound is known.[3] **^{1}H NMR (400 MHz, $CDCl_3$)** δ 8.14 (ddd, *J* = 5.1, 2.0, 0.9 Hz, 1H), 7.41 (ddd, *J* = 8.9, 7.1, 2.0 Hz, 1H), 6.58 – 6.41 (m, 1H), 6.34 (dd, *J* = 8.5, 1.0 Hz, 4H), 3.65 – 3.16 (m, 4H), 2.26 – 1.64 (m, 4H).

Synthetic Procedure C:



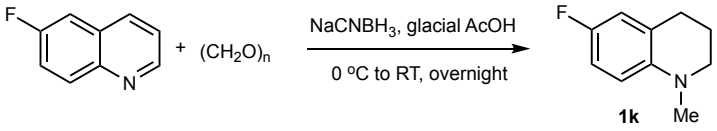
Substrate **1j** was synthesized using the following protocol. A reaction vial was charged with Cu(I) (50 mg, 0.262 mmol, 0.05 equiv), powder NaOH (350 mg, 8.75 mmol, 2 equiv), and 3-methylpiperidine (430 mg, 4.34 mmol, 1 equiv). The vial was sealed and placed under an argon atmosphere. To the vial was added isopropyl alcohol (iPA) (5 mL, 0.8 M) and iodobenzene (600 uL, 5.2 mmol, 1.2 equiv). The reaction was heated to 80 °C and left to stir overnight. The solvent was removed *in vacuo* and reconstituted in ethyl acetate and water. The water layer was extracted three times with ethyl acetate, and the combined organic layers were washed with brine and dried over Na_2SO_4. The solution was reduced *in vacuo* and purified via flash chromatography, affording **1j** as an oil.

3-methyl-1-phenylpiperidine (1j)



This compound is known.[4] **^{1}H NMR (400 MHz, $CDCl_3$)** δ 7.40 – 7.23 (m, 2H), 7.01 (d, *J* = 7.7 Hz, 2H), 6.90 – 6.83 (m, 1H), 3.76 – 3.56 (m, 2H), 2.69 (td, *J* = 11.7, 3.4 Hz, 1H), 2.39 (dd, *J* = 11.9, 10.3 Hz, 1H), 1.96 – 1.65 (m, 4H), 1.10 (m, 1H), 1.01 (d, *J* = 6.5 Hz, 3H).

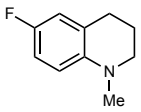
Synthetic Procedure D:



Substrate **1k** was prepared according to a modified procedure reported by Jia et al.[5] An oven-dried flask was charged with 6-fluoroquinoline (5 mmol, 0.83 g, 1.0 equiv.), paraformaldehyde (50 mmol, 1.50 g, 10.0 equiv.), and glacial acidic acid (25 mL, 0.2 M). $NaCNBH_3$ (25 mmol, 1.57 g, 5 equiv.) was then added in batches at 0 °C. The reaction was

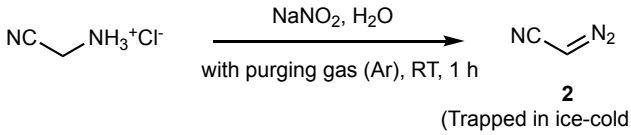
left to stir overnight at room temperature overnight and was quenched at 0 °C by slow addition of concentrated NaOH aqueous solution. The mixture was extracted with EA three times, combined, washed with brine, and dried with Na_2SO_4. The product was purified by flash chromatography to yield a dark red liquid.

6-fluoro-1-methyl-1,2,3,4-tetrahydroquinoline (1k)

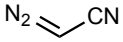


This compound is known.[6] **^{1}H NMR (400 MHz, $CDCl_3$)** ^{1}H NMR (400 MHz, Chloroform-*d*) δ 6.77 (ddd, *J* = 9.0, 8.3, 3.1 Hz, 1H), 6.70 (dd, *J* = 9.0, 3.1 Hz, 1H), 6.51 (dd, *J* = 8.9, 4.7 Hz, 1H), 3.16 (m, 2H), 2.85 (s, 3H), 2.75 (t, *J* = 6.5 Hz, 2H), 2.03 – 1.94 (m, 2H).

Preparation of diazoacetonitrile (2):



CAUTION: Diazo compounds are toxic and potentially explosive and should be handled with care in a well-ventilated hood.

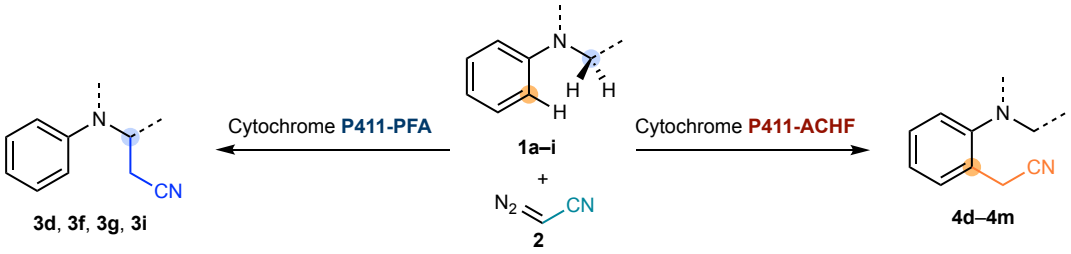


The stock solution of diazo compound **2** was prepared by using a modified procedure reported by Hock et al.[4] To a vigorously stirring solution of 1.3 g (14 mmol) of aminoacetonitrile hydrochloride in 1 mL of water solution of 0.96 g $NaNO_2$ in 2 mL of water was added slowly (1 hour) at room temperature using a syringe pump. The rapidly evolved yellow gas was carefully bubbled (carrier argon gas was used) via PTFE tubing into 6 mL of ethanol placed in a 7.5-mL sealed vial which is chilled on an ice bath. Upon the completion of addition, the aqueous mixture was further bubbled for 30 min to maximize the stock solution concentration, the vial containing the diazo ethanol solution was carefully removed from the ice bath and stored in -20 °C freezer. The substrate may deactivate/polymerize within two days to yield black insoluble mixture, therefore need to be prepared freshly. The concentration of the stock solution was measured

by ^{1}H NMR with non-deuterated DMSO as internal standard. The diazo solution concentration was then adjusted to 550 mM for biotransformations unless noted

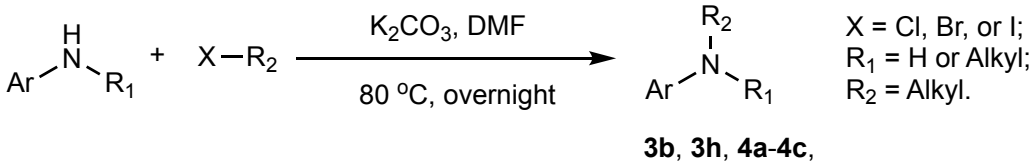
B.5 Syntheses of Reference Compounds

Synthetic Procedure A:



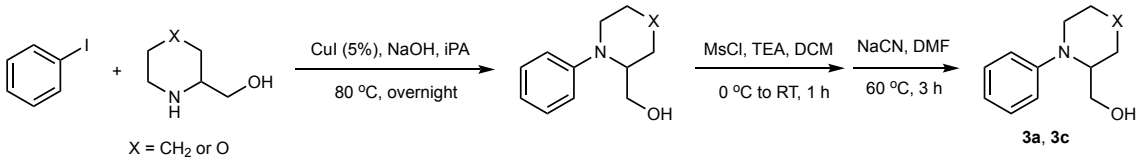
Product standards **3d**, **3f**, **3g**, **3i**, **4d**–**4m** were prepared on 0.5-mmol scales via preparative enzymatic syntheses using *E. coli* whole cell harboring either P411-PFA or P411-ACHF. The biotransformation protocol was described in **Section 3.4.9**.

Synthetic Procedure B:



3b, **3h**, **4a-4c** were prepared according to a modified procedure reported by Shi[2] and McNally[3]. A mixture of arylamines (0.50 mmol, 1.0 equiv.), alkyl halides (0.55 mmol, 1.1 equiv.), and K_2CO_3 (0.55 mmol, 1.1 equiv.) in DMF (1.0 mL, 0.5 M) was stirred at 80 °C for 10 h. The reaction was then cooled and diluted with EtOAc (100 mL) and H_2O (100 mL). The layers were separated and the organic layer was washed with brine (20 mL), dried over $MgSO_4$, filtered, concentrated, and purified by flash chromatography.

Synthetic Procedure C:



Racemic reference compounds **3a** and **3c** were prepared with a three-step protocol as shown above.

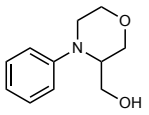
Step 1: A 40-mL dram vial with a magnetic stirring bar was charged with Cu(I) (50 mg, 0.262 mmol, 0.05 equiv.), powder NaOH (350 mg, 8.75 mmol, 2 equiv.), and morpholin-3-ylmethanol (510 mg, 4.34 mmol, 1 equiv.) or piperidin-2-ylmethanol (500 mg, 4.34 mmol, 1 equiv.). The vial was sealed and placed under an argon atmosphere. To the vial was added isopropyl alcohol (iPA) (5 mL, 0.8 M) and iodobenzene (600 uL, 5.2 mmol, 1.2 equiv.) via syringe. The reaction was heated to 80 °C and left to stir overnight. The following day the reaction was reduced *in vacuo* and reconstituted in DCM and water. The mixture was introduced into a separatory funnel and the organic layer was separated. The water layer was extracted twice with DCM and the combined organic layers were washed with brine and dried over $MgSO_4$. The solution was reduced *in vacuo* and purified via flash chromatography, affording (4-phenylmorpholin-3-yl)methanol (522 mg) in 62% yield as a solid, and (1-phenylpiperidin-2-yl)methanol (240 mg) in 29% yield as an oil.

Step 2: A 10-mL dram vial with a magnetic stirring bar was charged with (4-phenylmorpholin-3-yl)methanol (195 mg, 1 mmol, 1 equiv.) or (1-phenylpiperidin-2-yl)methanol (200 mg, 1 mmol, 1 equiv.) and placed under an argon atmosphere. Thereafter, DCM (5 mL, 0.2M) was added via syringe and the reaction was stirred and cooled to 0 °C (ice/water bath). Triethyl amine (TEA) (280 µL, 2.0 mmol, 2.0 equiv.) and methanesulfonyl chloride (MsCl) (120 µL, 1.5 mmol, 1.5 equiv.) were added dropwise in that order. The reaction was left to warm to room temperature and stirred for 1 hour at that temperature. The reaction was then introduced into a separatory funnel and washed with saturated $NaHCO_3$,

brine, and dried over $MgSO_4$. The resultant oil was used immediately in the following reaction.

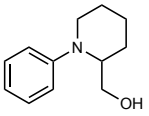
Step 3: A 10-mL dram vial with a magnetic stirring bar and the resultant oil from the previous reaction was charged with DMF (5 mL, 0.2 M) and NaCN (100 mg, 2 mmol, 2 equiv.). The reaction was sealed, blanketed with an argon atmosphere, and heated to 60 °C (heating block) for 3 h. The reaction was cooled to room temperature and diluted with ethyl acetate and water. The organic layer was washed several times with brine, dried over $MgSO_4$, and reduced *in vacuo*. The resultant residue was purified using flash chromatography affording **3a** in 50% yield (100 mg) as an oil and **3c** in 60% yield (120 mg) as an oil.

(4-phenylmorpholin-3-yl)methanol



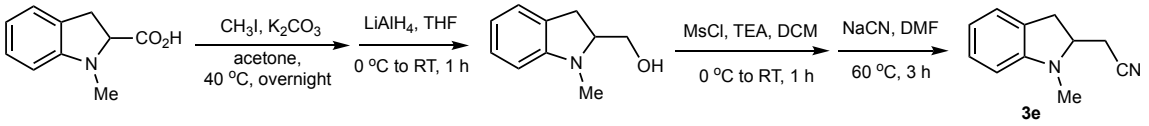
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.33 – 7.16 (m, 2H), 7.03 (d, *J* = 8.1 Hz, 2H), 6.86 (t, *J* = 7.3 Hz, 1H), 3.90 – 3.71 (m, 2H), 3.60 (dd, *J* = 9.4, 6.0 Hz, 1H), 3.47 – 3.30 (m, 1H), 3.10 (dd, *J* = 12.0, 6.6 Hz, 1H), 2.12 – 1.45 (m, 7H); **^{13}C NMR (101 MHz, $CDCl_3$)** δ 129.28, 119.88, 117.93, 60.57, 58.07, 45.69, 25.39, 24.46, 20.64.

(1-phenylpiperidin-2-yl)methanol



^{1}H NMR (400 MHz, $CDCl_3$) δ 7.29 (dd, *J* = 8.7, 7.2 Hz, 2H), 7.04 – 6.84 (m, 3H), 4.04 (dq, *J* = 8.8, 4.3 Hz, 1H), 3.21 (dt, *J* = 12.5, 4.4 Hz, 1H), 2.88 (ddd, *J* = 12.4, 9.4, 3.1 Hz, 1H), 2.54 (dd, *J* = 16.7, 9.8 Hz, 1H), 2.33 (dd, *J* = 16.7, 4.0 Hz, 1H), 2.12 – 1.88 (m, 2H), 1.79 (tt, *J* = 6.3, 2.9 Hz, 1H), 1.74 – 1.51 (m, 3H); **^{13}C NMR (101 MHz, $CDCl_3$)** δ 150.12, 129.59, 121.49, 118.79, 54.21, 46.24, 28.75, 25.56, 19.77, 16.96.

Synthetic Procedure D:



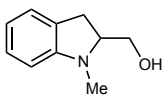
Racemic reference and *S*-isomer of the **3e** was prepared with a four-step protocol as shown above.

Step 1: A 200-mL round bottom flask with a stir bar was charged with indoline-2-carboxylic acid (1.63 g, 10.0 mmol, 1 equiv.) and K_2CO_3 (2.8 g, 20 mmol, 2.0 equiv.). The vial was sealed and placed under an argon atmosphere. Subsequently, acetone (25 mL, 0.4 M) and iodomethane (1.42 g, 623 µL, 10.0 mmol, 1 equiv.) were added via syringe. The reaction was heated to 40 °C and stirred overnight. The reaction was cooled to room temperature and 60 mL of Hex/EtOAc (1:1) was added, at which point a solid was observed to precipitate. The reaction was filtered and the filtrate was reduced in vacuo. The resultant residue was reconstituted in DCM and washed with water and brine, and dried over $MgSO_4$. The solution was reduced in vacuo and the obtained yellow oil was dried overnight.

Step 2: A 200-mL round bottom flask with a magnetic stirring bar was charged with the yellow oil (*vide supra*) and THF (50 mL, 0.2 M). The stirred reaction was cooled to 0 °C (water/ice bath) and $LiAlH_4$ (1.2 g, 30 mmol, 3.0 equiv) was added in one portion. The reaction was left to warm to room temperature and after a 1 h stir period it was cooled to 0 °C and quenched by dropwise addition of acetone. Thereafter, a saturated solution of Rochelle's salt (50 mL) was added and the reaction was left to stir until the gray solution became clear. The resultant mixture was introduced into a separatory funnel and extracted thrice with Et_2O.The combined organic fractions were dried over $MgSO_4$ and reduced in vacuo. The resultant material was purified using flash chromatography to afford 1-methylindolin-2-yl)methanol (826 mg) in 50% yield as a yellow oil.

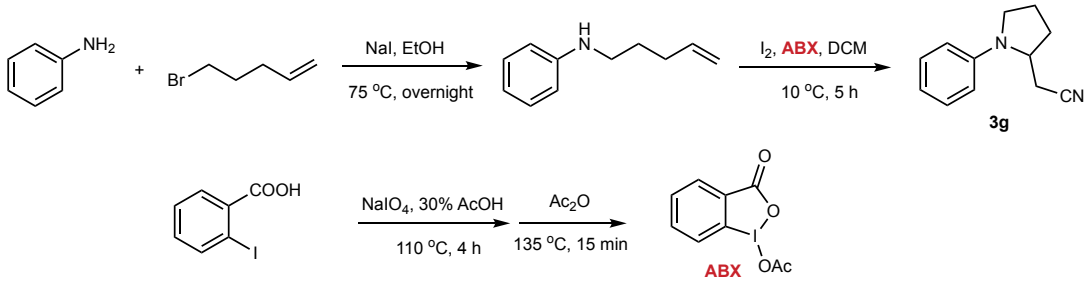
Step 3 and 4: The two subsequent steps to prepare racemic and the *S*-enantiomer of **3e** follow a similar approach as Step 2 and 3 in making **3a** and **3c** (*vide supra*). **3e** was obtained in 60% yield (103 mg) as an oil that solidified after cooling.

(1-methylindolin-2-yl)methanol



^{1}H NMR (400 MHz, $CDCl_3$) δ 7.17 – 7.04 (overlap, 2H), 6.75 (td, *J* = 7.4, 1.0 Hz, 1H), 6.57 (d, *J* = 7.8 Hz, 1H), 3.95 (dd, *J* = 11.5, 3.4 Hz, 1H), 3.68 (dd, *J* = 11.5 Hz, 1H), 3.51 (tdd, *J* = 9.8, 3.4, 2.3 Hz, 1H), 3.10 – 3.01 (overlap, 2H), 2.78 (s, 3H), 2.00 (bs, 1H); **^{13}C NMR (101 MHz, $CDCl_3$)** δ 153.41, 129.38, 127.57, 124.52, 119.23, 108.70, 68.54, 61.13, 35.25, 31.23.

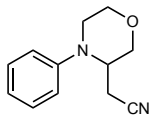
Synthetic Procedure E:



The synthesis of racemic product standard of **3g** was reported by Shen et al.[7]

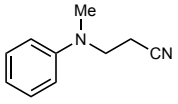
B.6 Characterizations of Standard Products:

2-(4-phenylmorpholin-3-yl)acetonitrile (3a)



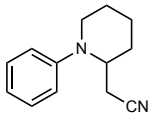
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.32 (dd, *J* = 8.8, 7.5 Hz, 2H), 7.00 – 6.93 (m, 1H), 6.90 (dt, *J* = 7.5, 1.1 Hz, 2H), 4.10 (ddd, *J* = 11.7, 2.1, 1.0 Hz, 1H), 4.07 – 3.97 (m, 2H), 3.91 (ddd, *J* = 11.7, 2.8, 1.6 Hz, 1H), 3.73 (td, *J* = 11.2, 3.4 Hz, 1H), 3.16 (ddd, *J* = 12.3, 3.4, 2.2 Hz, 1H), 3.05 (ddd, *J* = 12.3, 11.0, 3.7 Hz, 1H), 2.85 (dd, *J* = 16.5, 10.4 Hz, 1H), 2.31 (ddd, *J* = 16.5, 3.4, 1.6 Hz, 1H); **^{13}C NMR (101 MHz, $CDCl_3$)** δ 148.27, 129.84, 121.19, 118.35, 116.32, 68.40, 66.88, 53.28, 43.28, 14.17; **HRMS** m/z $[M]^{+\bullet}$ calcd for $C_{12}H_{14}N_2O^+$: 202.1106, found: 202.1112.

3-(methyl(phenyl)amino)propanenitrile (3b)



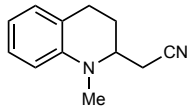
This compound is known.[8] **^{1}H NMR (400 MHz, $CDCl_3$)** δ 7.39 – 7.25 (m, 2H), 6.84 (tt, *J* = 7.2, 1.0 Hz, 1H), 6.77 (dd, *J* = 8.8, 1.1 Hz, 2H), 3.74 (t, *J* = 6.9 Hz, 2H), 3.06 (s, 3H), 2.60 (t, *J* = 6.9 Hz, 2H); **^{13}C NMR (101 MHz, $CDCl_3$)** δ 147.65, 129.63, 118.58, 117.89, 112.74, 49.11, 38.80, 15.31.

2-(1-phenylpiperidin-2-yl)acetonitrile (3c)



This compound is known.[7] **^{1}H NMR (400 MHz, $CDCl_3$)** δ 7.29 (dd, *J* = 8.7, 7.2 Hz, 2H), 7.04 – 6.84 (m, 3H), 4.04 (dq, *J* = 8.8, 4.3 Hz, 1H), 3.21 (dt, *J* = 12.5, 4.4 Hz, 1H), 2.88 (ddd, *J* = 12.4, 9.4, 3.1 Hz, 1H), 2.54 (dd, *J* = 16.7, 9.8 Hz, 1H), 2.33 (dd, *J* = 16.7, 4.0 Hz, 1H), 2.12 – 1.88 (m, 2H), 1.79 (tt, *J* = 6.3, 2.9 Hz, 1H), 1.74 – 1.51 (m, 3H); **^{13}C NMR (101 MHz, $CDCl_3$)** δ 150.12, 129.59, 121.49, 118.79, 118.72, 54.21, 46.24, 28.75, 25.56, 19.77, 16.96.

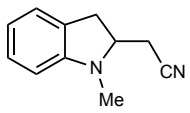
2-(1-methyl-1,2,3,4-tetrahydroquinolin-2-yl)acetonitrile (3d)



^{1}H NMR (400 MHz, $CDCl_3$) δ 7.12 (dddd, *J* = 8.1, 7.3, 1.6, 0.8 Hz, 1H), 7.02 (ddt, *J* = 7.4, 1.9, 1.0 Hz, 1H), 6.69 (td, *J* = 7.3, 1.1 Hz, 1H), 6.63 – 6.51 (m, 1H), 3.75 (dddd, *J* = 8.6, 5.6, 4.4, 3.3 Hz, 1H), 3.00 (s,

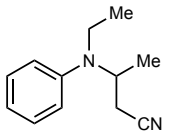
3H), 2.84 – 2.77 (m, 2H), 2.60 (dd, *J* = 5.5, 0.7 Hz, 1H), 2.42 (dd, *J* = 16.8, 8.5 Hz, 1H), 2.20 – 2.01 (m, 2H); **^{13}C NMR (101 MHz, CDCl$_3$)** δ 143.76, 129.21, 127.73, 121.28, 118.31, 117.31, 111.71, 56.24, 38.28, 25.09, 22.92, 19.82; **HRMS** m/z $[M]^{+\bullet}$ calcd for $C_{12}H_{14}N_2^+$: 186.1152, found: 232.1162.

2-(1-methylindolin-2-yl)acetonitrile (3e)



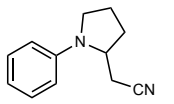
This compound is known.[7] **^{1}H NMR (400 MHz, CDCl$_3$)** δ 7.19 – 7.02 (m, 2H), 6.80 – 6.66 (m, 1H), 6.50 (d, *J* = 7.8 Hz, 1H), 3.69 (tdd, *J* = 8.8, 6.9, 4.4 Hz, 1H), 3.29 (dd, *J* = 15.7, 8.8 Hz, 1H), 2.90 (dd, *J* = 15.7, 8.8 Hz, 1H), 2.80 (s, 3H), 2.76 (dd, *J* = 16.8, 4.4 Hz, 1H), 2.65 (dd, *J* = 16.8, 6.9 Hz, 1H); **^{13}C NMR (101 MHz, CDCl$_3$)** δ 152.26, 128.01, 127.40, 124.45, 119.00, 117.57, 107.87, 63.03, 35.21, 34.52, 22.05.

3-(ethyl(phenyl)amino)butanenitrile (3f)



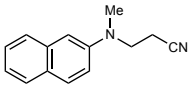
^{1}H NMR (400 MHz, CDCl$_3$) δ 7.30 – 7.19 (m, 2H), 6.81 (d, *J* = 7.7 Hz, 3H), 4.20 (td, *J* = 7.1, 5.6 Hz, 1H), 3.27 (ddt, *J* = 20.2, 14.9, 7.5 Hz, 2H), 2.64 – 2.41 (m, 2H), 1.42 (d, *J* = 6.7 Hz, 3H), 1.16 (t, *J* = 7.0 Hz, 3H); **^{13}C NMR (101 MHz, CDCl$_3$)** δ 147.45, 129.57, 118.56, 118.43, 115.22, 51.51, 38.80, 22.60, 18.10, 14.73; **HRMS** m/z $[M]^{+\bullet}$ calcd for $C_{12}H_{16}N_2^+$: 188.1308, found: 188.1318.

2-(1-phenylpyrrolidin-2-yl)acetonitrile (3g)



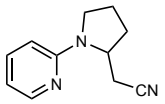
This compound is known.[7] **^{1}H NMR (400 MHz, CDCl$_3$)** δ 7.37 – 7.19 (m, 2H), 6.76 (tt, *J* = 7.2, 1.1 Hz, 1H), 6.59 (dd, *J* = 8.8, 1.1 Hz, 2H), 4.09 (tdd, *J* = 7.2, 3.2, 1.6 Hz, 1H), 3.57 – 3.51 (m, 1H), 3.23 (td, *J* = 8.6, 6.8 Hz, 1H), 2.72 (dd, *J* = 16.8, 3.3 Hz, 1H), 2.44 (dd, *J* = 16.8, 8.7 Hz, 1H), 2.27 – 2.00 (m, 4H); **^{13}C NMR (101 MHz, CDCl$_3$)** δ 146.07, 129.73, 118.35, 117.18, 112.29, 55.35, 48.75, 31.04, 29.88, 23.15, 21.57.

3-(methyl(naphthalen-2-yl)amino)propanenitrile (3h)



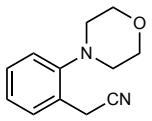
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.80 – 7.66 (m, 3H), 7.42 (ddd, *J* = 8.2, 6.8, 1.3 Hz, 1H), 7.30 – 7.27 (m, 1H), 7.14 (dd, *J* = 9.0, 2.6 Hz, 1H), 7.01 (s, 1H), 3.82 (t, *J* = 7.0 Hz, 2H), 3.14 (s, 3H), 2.64 (t, *J* = 7.0 Hz, 2H); **^{13}C NMR (101 MHz, $CDCl_3$)** δ 134.90, 129.68, 127.69, 126.86, 126.58, 123.29, 118.41, 115.94, 108.24, 49.68, 39.46, 15.48; **HRMS** m/z $[M]^+$• calcd for $C_{14}H_{14}N_2^+$: 210.1152, found: 210.1174.

2-(1-(pyridin-2-yl)pyrrolidin-2-yl)acetonitrile (3i)



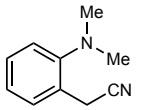
^{1}H NMR (400 MHz, $CDCl_3$) δ 8.14 (ddd, *J* = 5.1, 2.0, 0.9 Hz, 1H), 7.55 – 7.47 (m, 1H), 6.62 (t, *J* = 6.1 Hz, 1H), 6.44 (d, *J* = 8.5 Hz, 1H), 4.46 (s, 1H), 3.59 (s, 1H), 3.41 – 3.24 (m, 1H), 2.90 (d, *J* = 4.5 Hz, 2H), 2.39 – 1.94 (m, 4H); **HRMS** m/z $[M]^+$• calcd for $C_{11}H_{13}N_3O_2^+$: 187.1104, found: 187.1121.

2-(2-morpholinophenyl)acetonitrile (4a)



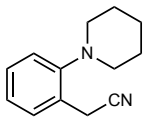
This compound is known.[9] **^{1}H NMR (400 MHz, $CDCl_3$)** δ 7.37 (dd, *J* = 7.6, 1.6 Hz, 1H), 7.29 (td, *J* = 7.7, 1.6 Hz, 1H), 7.17 – 7.08 (m, 2H), 3.81 – 3.78 (m, 4H), 3.77 (s, 2H), 2.84 – 2.80 (m, 4H); **^{13}C NMR (101 MHz, $CDCl_3$)** δ 151.12, 129.90, 129.62, 126.76, 125.51, 121.74, 118.76, 67.54, 53.16, 19.71.

2-(2-(dimethylamino)phenyl)acetonitrile (4b)



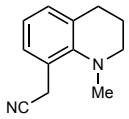
This compound is known.[10] **^{1}H NMR (400 MHz, $CDCl_3$)** δ 7.45 (dd, *J* = 7.7, 1.6 Hz, 1H), 7.32 (td, *J* = 7.7, 1.6 Hz, 1H), 7.19 (dd, *J* = 8.1, 1.3 Hz, 1H), 7.12 (td, *J* = 7.5, 1.3 Hz, 1H), 3.86 (s, 2H), 2.68 (s, 6H); **^{13}C NMR (101 MHz, $CDCl_3$)** δ 152.57, 129.68, 129.26, 125.93, 124.47, 120.59, 119.01, 45.11, 19.66.

2-(2-(piperidin-1-yl)phenyl)acetonitrile (4c)



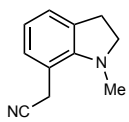
This compound is known.[9] **^{1}H NMR (400 MHz, $CDCl_3$)** δ 7.44 (dd, *J* = 7.7, 1.6 Hz, 1H), 7.31 (td, *J* = 7.7, 1.6 Hz, 1H), 7.18 – 7.10 (m, 2H), 3.84 (s, 2H), 2.84 – 2.79 (m, 4H), 1.73 (m, *J* = 5.4 Hz, 4H), 1.66 – 1.48 (m, 2H); **^{13}C NMR (101 MHz, $CDCl_3$)** δ 152.54, 129.53, 129.24, 126.40, 124.54, 121.20, 119.06, 63.77, 54.31, 26.68, 24.27, 19.51.

2-(1-methyl-1,2,3,4-tetrahydroquinolin-8-yl)acetonitrile (4d)



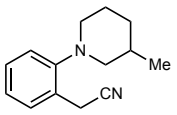
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.28 (d, *J* = 7.7 Hz, 1H), 7.07 – 6.94 (m, 2H), 3.79 (s, 2H), 3.18 – 2.98 (m, 2H), 2.81 (t, *J* = 6.8 Hz, 2H), 2.67 (s, 3H), 1.87 (t, *J* = 5.7 Hz, 2H); **^{13}C NMR (101 MHz, $CDCl_3$)** δ 147.78, 130.36, 129.81, 127.44, 124.32, 122.74, 119.19, 51.63, 43.92, 27.40, 19.77, 16.55; **HRMS** m/z $[M]^+$• calcd for $C_{12}H_{14}N_2^+$: 186.1152, found: 186.1166.

2-(1-methylindolin-7-yl)acetonitrile (4e)



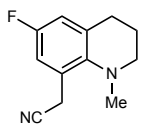
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.09 – 7.00 (m, 1H), 6.75 (t, *J* = 7.5 Hz, 1H), 3.78 (s, 1H), 3.38 (t, *J* = 8.5 Hz, 1H), 2.96 (s, 3H); **^{13}C NMR (101 MHz, $CDCl_3$)** δ 151.01, 132.84, 129.36, 124.74, 119.97, 118.27, 112.07, 57.45, 40.27, 28.82, 21.34; **HRMS** m/z $[M]^+$• calcd for $C_{11}H_{12}N_2^+$: 172.0995, found: 172.1005.

2-(2-(3-methylpiperidin-1-yl)phenyl)acetonitrile (4j)



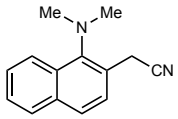
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.44 (dd, *J* = 7.6, 1.6 Hz, 1H), 7.31 (td, *J* = 7.7, 1.6 Hz, 1H), 7.18 – 7.08 (m, 2H), 3.83 (s, 2H), 2.90 (ddt, *J* = 11.2, 5.4, 3.3 Hz, 2H), 2.61 (td, *J* = 11.2, 3.2 Hz, 1H), 2.33 (dd, *J* = 11.4, 9.7 Hz, 1H), 1.92 – 1.63 (m, 4H), 1.04 (dtd, *J* = 13.3, 11.7, 4.7 Hz, 1H), 0.93 (d, *J* = 6.5 Hz, 3H); **^{13}C NMR (101 MHz, $CDCl_3$)** δ 152.39, 129.51, 129.23, 126.40, 124.51, 121.27, 119.05, 61.38, 53.80, 32.83, 31.89, 29.86, 26.13, 19.53; **HRMS** m/z $[M]^+$• calcd for $C_{14}H_{18}N_2^+$: 214.1465, found: 214.1473.

2-(6-fluoro-1-methyl-1,2,3,4-tetrahydroquinolin-8-yl)acetonitrile (4k)



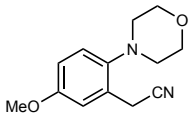
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.01 (dd, *J* = 8.8, 2.9 Hz, 1H), 6.74 (dd, *J* = 8.9, 2.9 Hz, 1H), 3.79 (s, 2H), 3.15 – 2.91 (m, 2H), 2.80 (t, *J* = 6.8 Hz, 2H), 2.61 (s, 3H), 1.87 (qd, *J* = 6.6, 2.7 Hz, 2H); **^{13}C NMR (101 MHz, $CDCl_3$)** δ 158.41 (d, *J* = 242.3 Hz), 143.64, 132.32 (d, *J* = 7.5 Hz), 126.30 (d, *J* = 8.0 Hz), 118.62, 115.84 (d, *J* = 21.1 Hz), 114.24 (d, *J* = 23.6 Hz), 51.52, 43.89, 27.44 (d, *J* = 1.5 Hz), 19.70, 15.95; **^{19}F NMR (282 MHz, $CDCl_3$)** δ -120.02; **HRMS** m/z $[M]^+$• calcd for $C_{12}H_{13}N_2F^+$: 204.1057, found: 204.1070.

2-(1-(dimethylamino)naphthalen-2-yl)acetonitrile (4l)



^{1}H NMR (400 MHz, $CDCl_3$) δ 8.07 – 8.00 (m, 1H), 7.93 – 7.85 (m, 1H), 7.71 (d, J = 8.4 Hz, 1H), 7.58 – 7.42 (m, 3H), 3.96 (s, 2H), 3.07 (s, 6H); **^{13}C NMR (101 MHz, $CDCl_3$)** δ 147.30, 135.28, 132.56, 129.18, 127.14, 126.98, 126.78, 126.19, 126.11, 124.52, 119.03, 43.95, 20.98; **HRMS** m/z $[M]^+$• calcd for $C_{14}H_{14}N_2^+$: 210.1152, found: 210.1168.

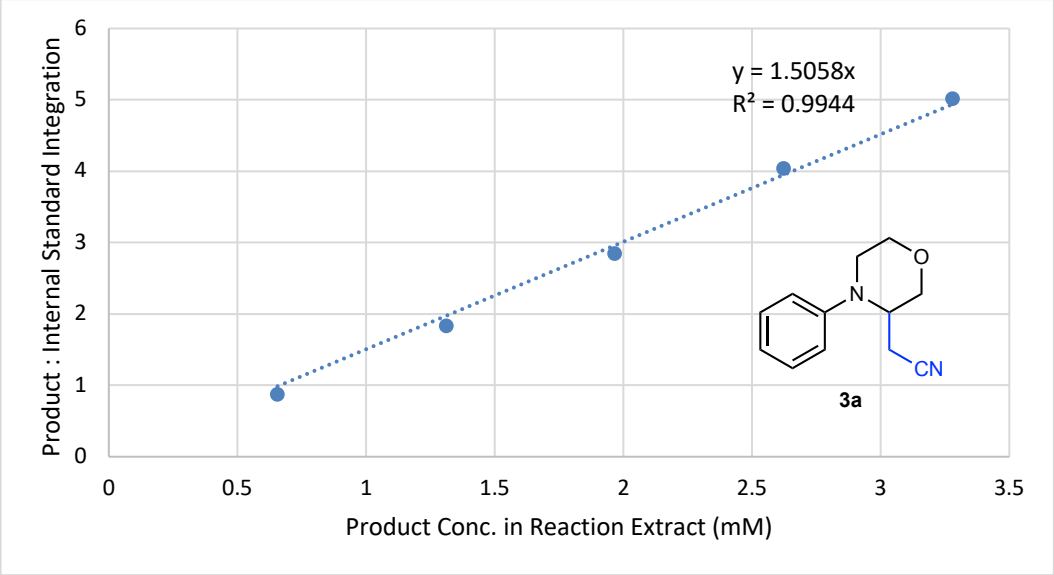
2-(5-methoxy-2-morpholinophenyl)acetonitrile (4m)

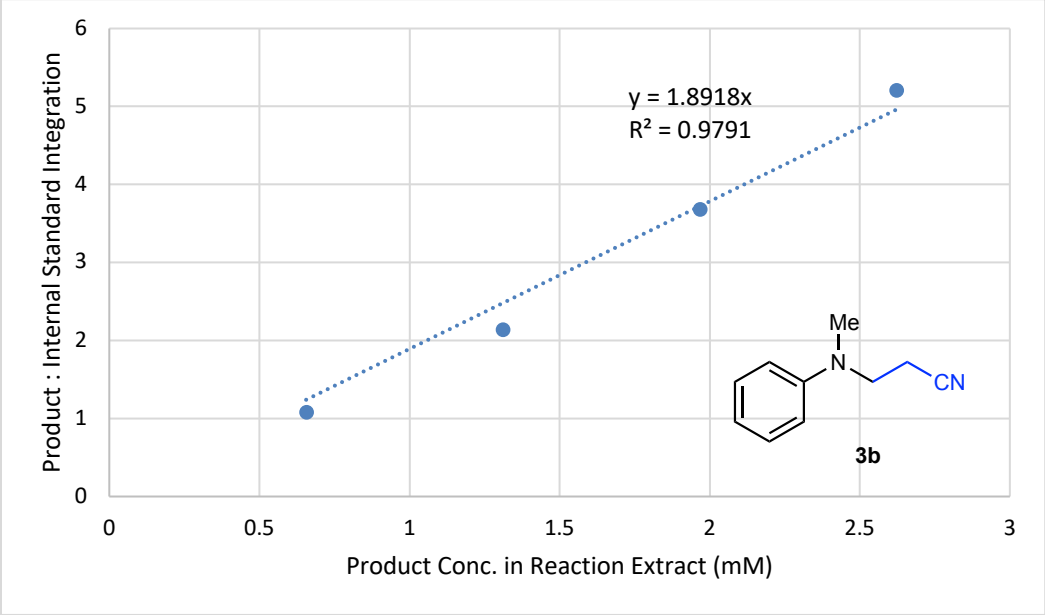


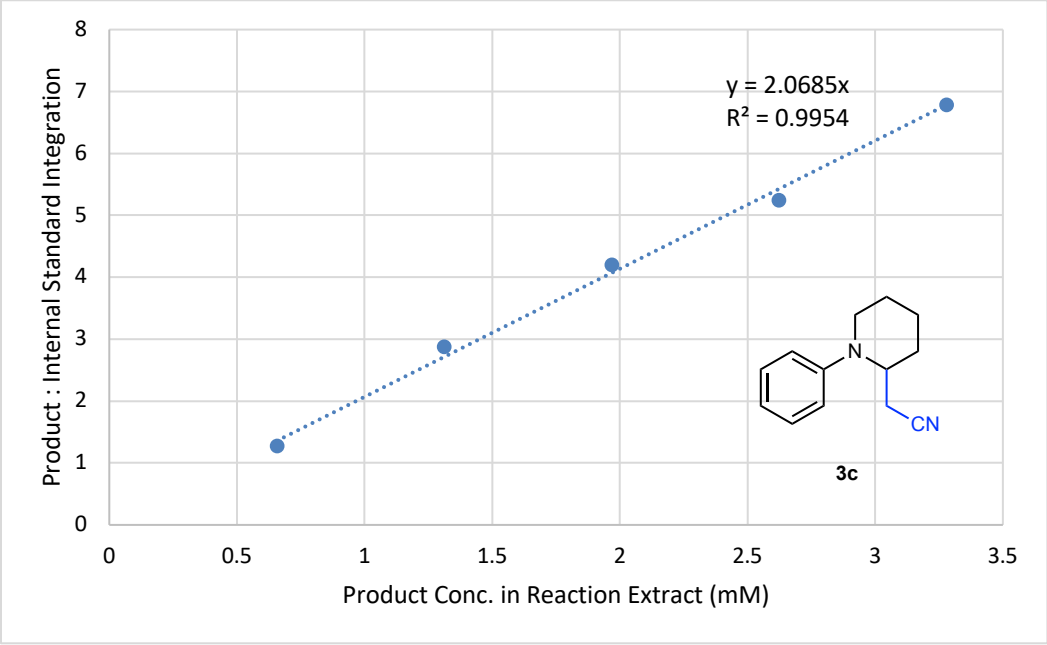
^{1}H NMR (400 MHz, $CDCl_3$) δ 7.17 (d, *J* = 8.6 Hz, 1H), 6.96 (d, *J* = 2.8 Hz, 1H), 6.88 (dd, *J* = 8.7, 2.8 Hz, 1H), 3.93 – 3.66 (m, 6H), 3.81 (s, 3H), 2.83 (t, *J* = 4.3 Hz, 4H); **^{13}C NMR (101 MHz, $CDCl_3$)** δ 157.16, 144.13, 128.31, 123.19, 118.67, 114.98, 114.70, 67.66, 55.73, 53.56, 19.89; **HRMS** m/z $[M]^+$• calcd for $C_{14}H_{14}N_2O_2^+$: 232.1206, found: 232.1225.

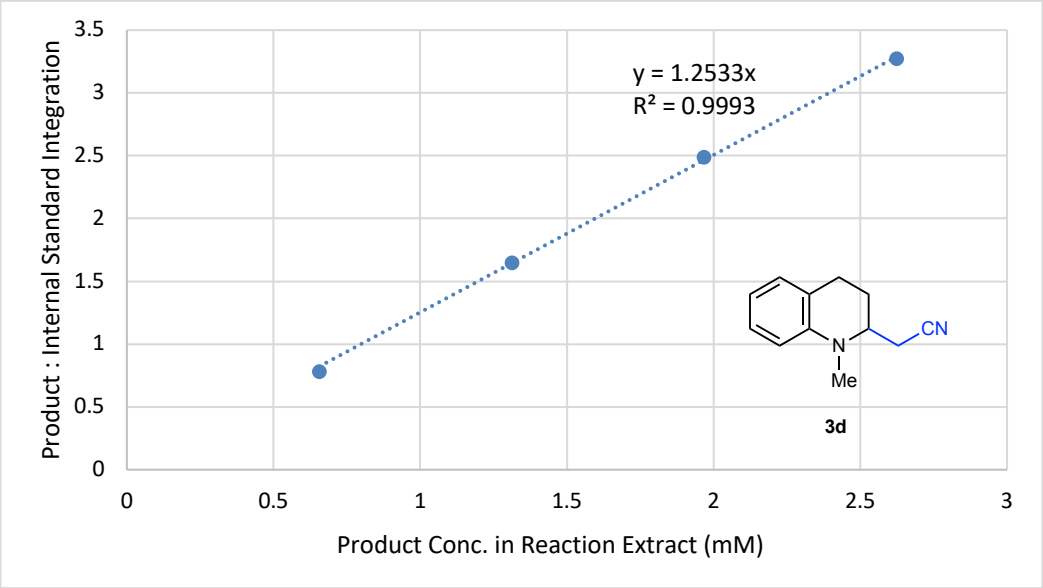
B.7 Product GC-FID Calibration Curves

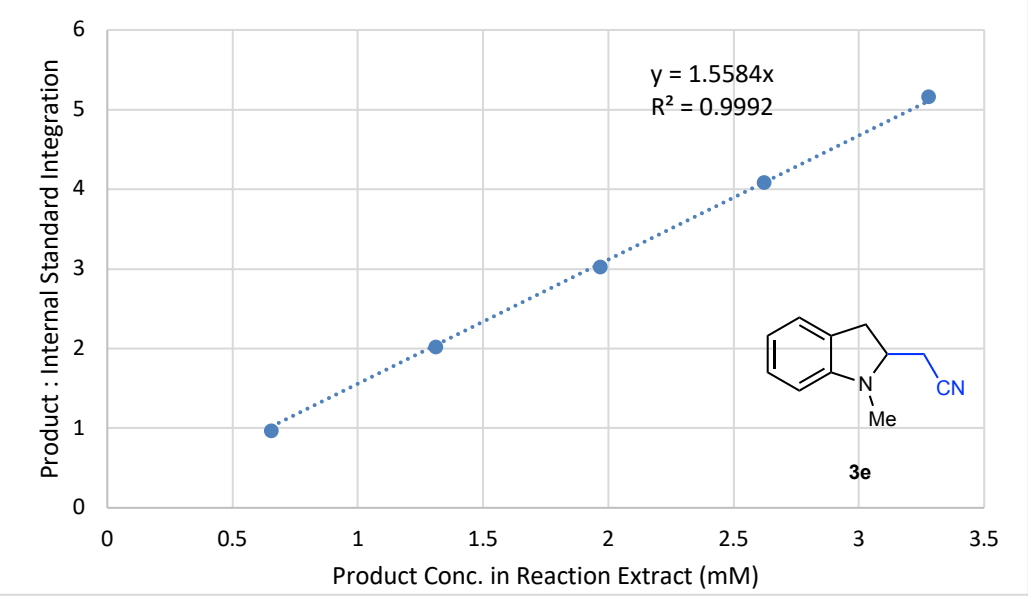
Product formation of **3a-3i**, **4a-4e, 4j-4m** in enzymatic reactions was quantified by GC-FID based on standard curves. To determine the standard calibration curves, stock solutions of chemically synthesized authentic products were prepared at various concentrations (0.5–7 mM in 4:6 hexanes/EtOAc) with added internal standard 1,2,3-trimethoxybenzene (0.66 mM final concentration in the stock solutions). All data points represent the average of duplicate runs. The standard curves plot product concentration in mM (y-axis) against the ratio of product area to internal standard area on GC-MS (x-axis). For **5d** and **5e**, 15 analytical scale reactions were combined to measure the product formation and total turnover numbers by ^{19}F NMR with fluorobenzene as the internal standard.

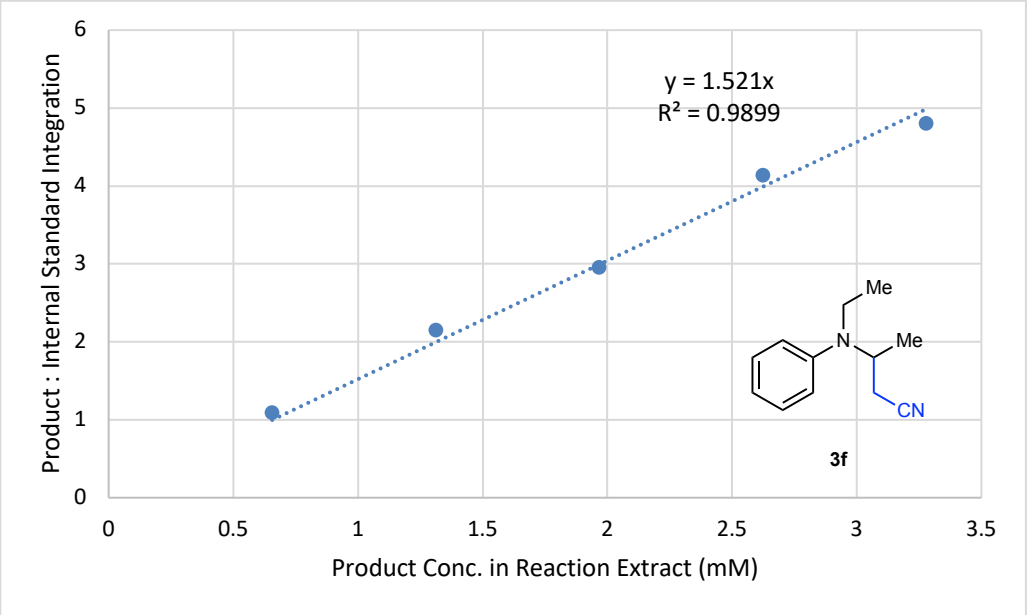


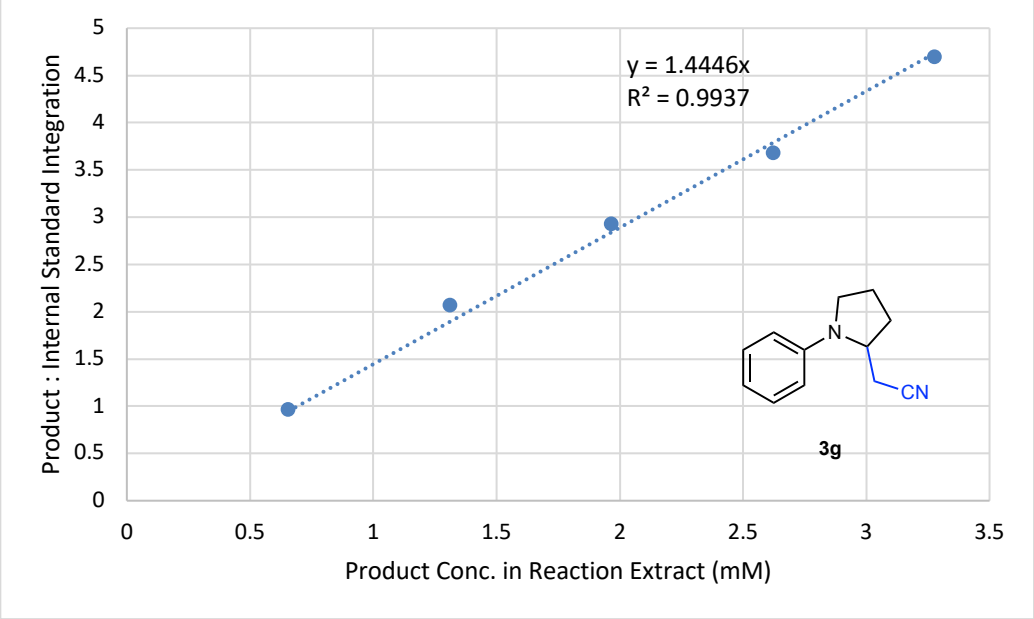

y = 1.8918x
R² = 0.9791
Product : Internal Standard Integration
Product Conc. in Reaction Extract (mM)
Me
N
CN
3b
0
1
2
3
4
5
6
0
0.5
1
1.5
2
2.5
3

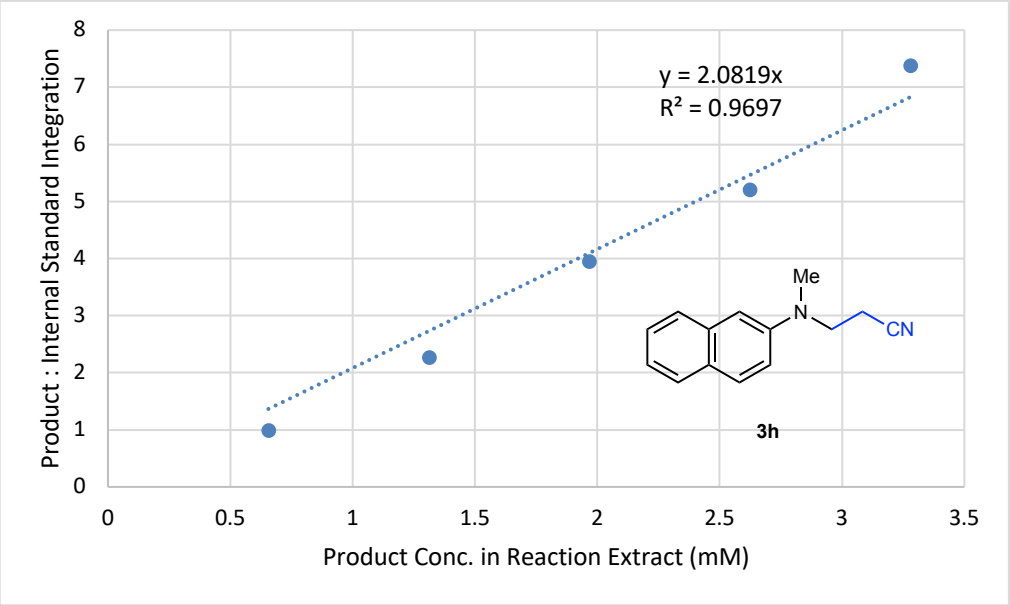

y = 2.0685x
R² = 0.9954
Product : Internal Standard Integration
Product Conc. in Reaction Extract (mM)
N
CN
3c
0
1
2
3
4
5
6
7
8
0
0.5
1
1.5
2
2.5
3
3.5

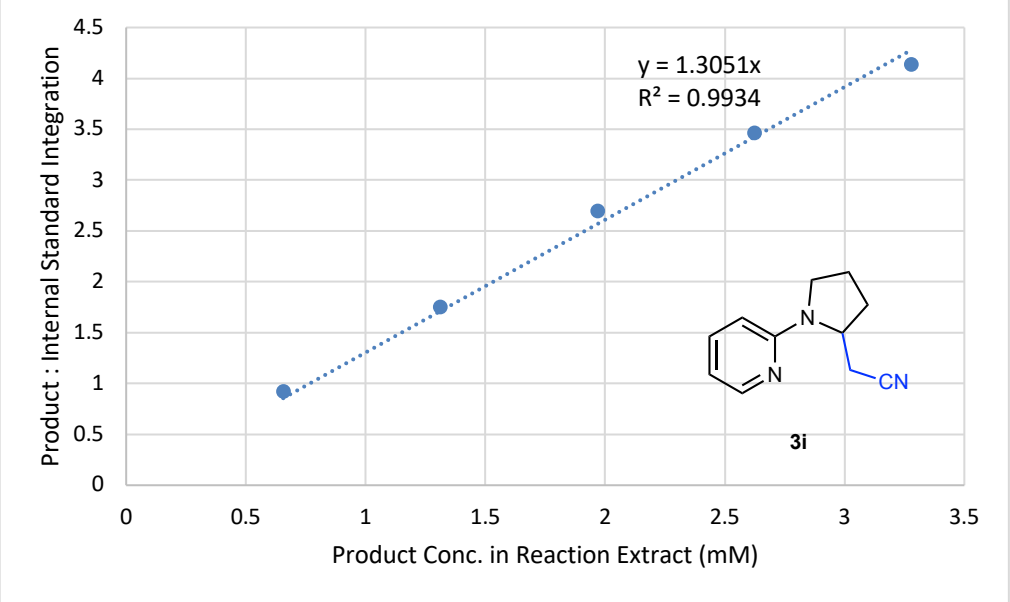

y = 1.2533x
R² = 0.9993
Product : Internal Standard Integration
Product Conc. in Reaction Extract (mM)
0
0.5
1
1.5
2
2.5
3
3.5
CN
N
Me
3d

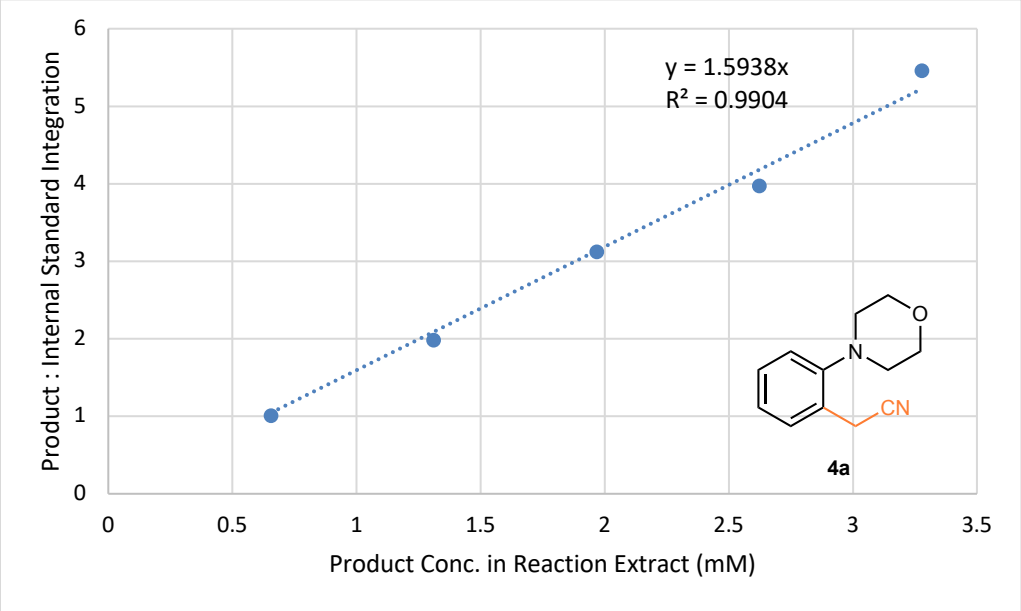

y = 1.5584x
R² = 0.9992
Product : Internal Standard Integration
Product Conc. in Reaction Extract (mM)
0
1
2
3
4
5
6
0.5
1.5
2.5
3.5
CN
N
Me
3e

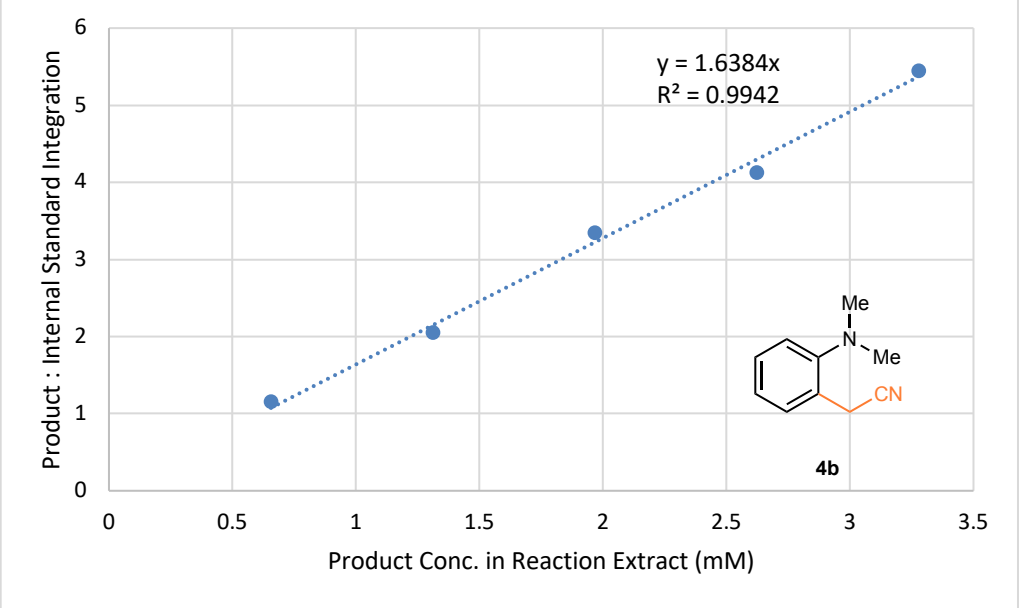

y = 1.521x
R² = 0.9899
Product : Internal Standard Integration
Product Conc. in Reaction Extract (mM)
Me
N
Me
CN
3f

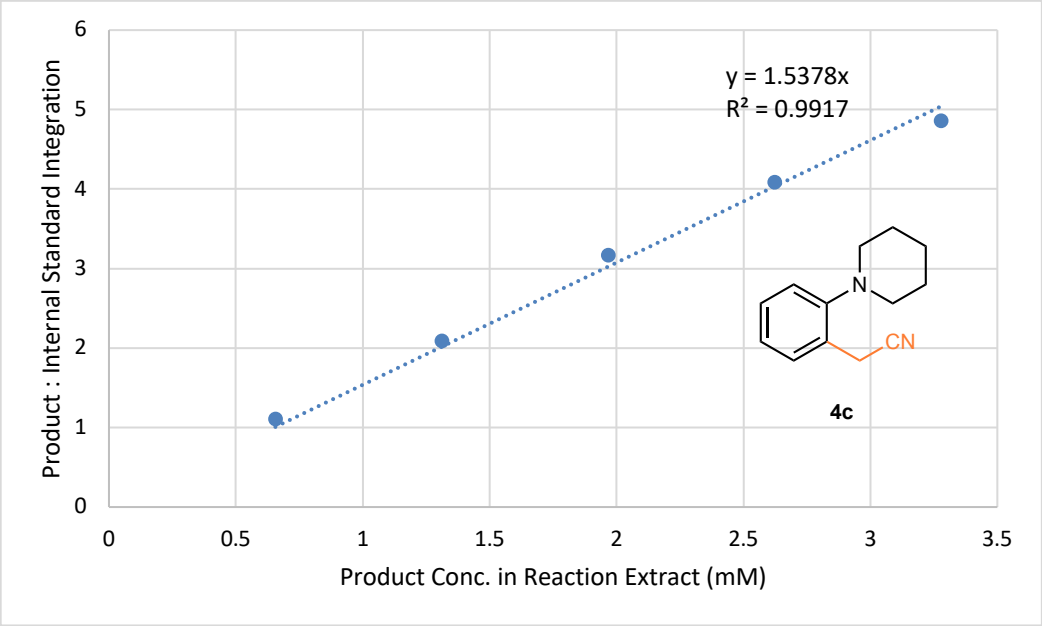

y = 1.4446x
R² = 0.9937
Product : Internal Standard Integration
Product Conc. in Reaction Extract (mM)
N
CN
3g

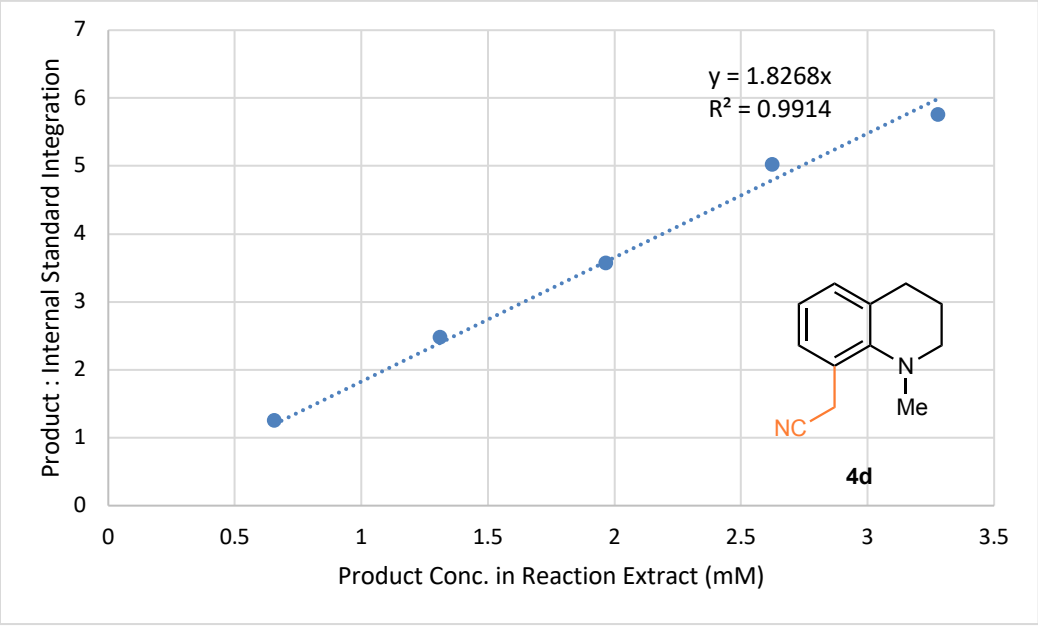

y = 2.0819x
R² = 0.9697
Product : Internal Standard Integration
Product Conc. in Reaction Extract (mM)
Me
N
CN
3h
0
0.5
1
1.5
2
2.5
3
3.5
4
5
6
7
8

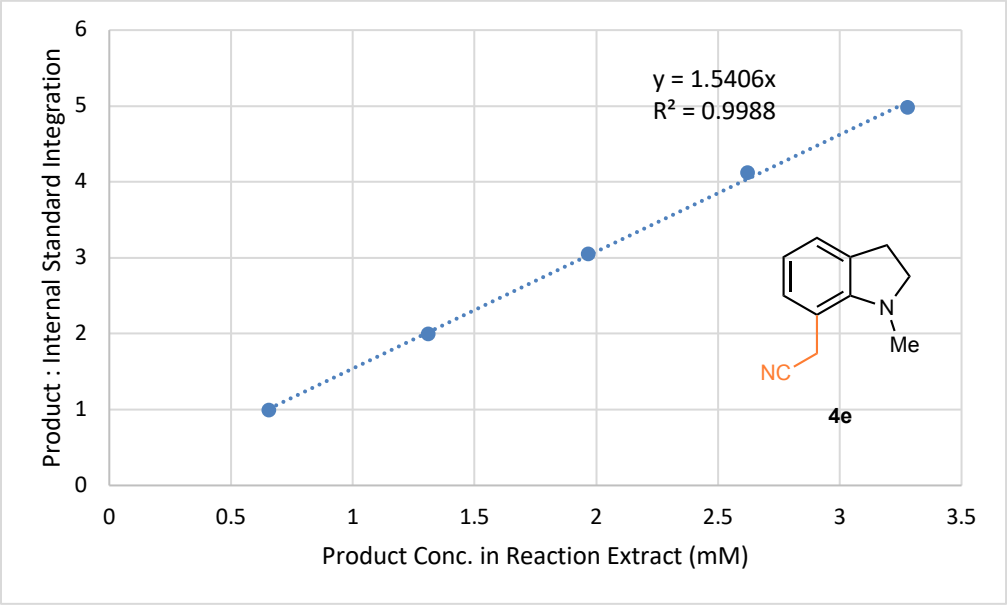

y = 1.3051x
R² = 0.9934
Product : Internal Standard Integration
Product Conc. in Reaction Extract (mM)
N
N
CN
3i
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5

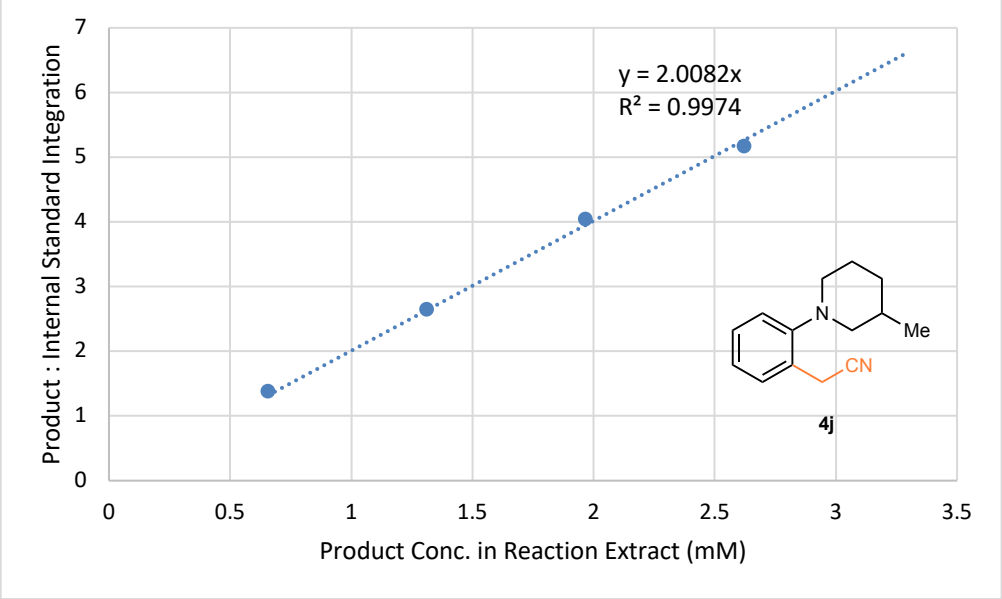

y = 1.5938x
R² = 0.9904
Product : Internal Standard Integration
Product Conc. in Reaction Extract (mM)
0
1
2
3
4
5
6
0
0.5
1
1.5
2
2.5
3
3.5
O
N
CN
4a

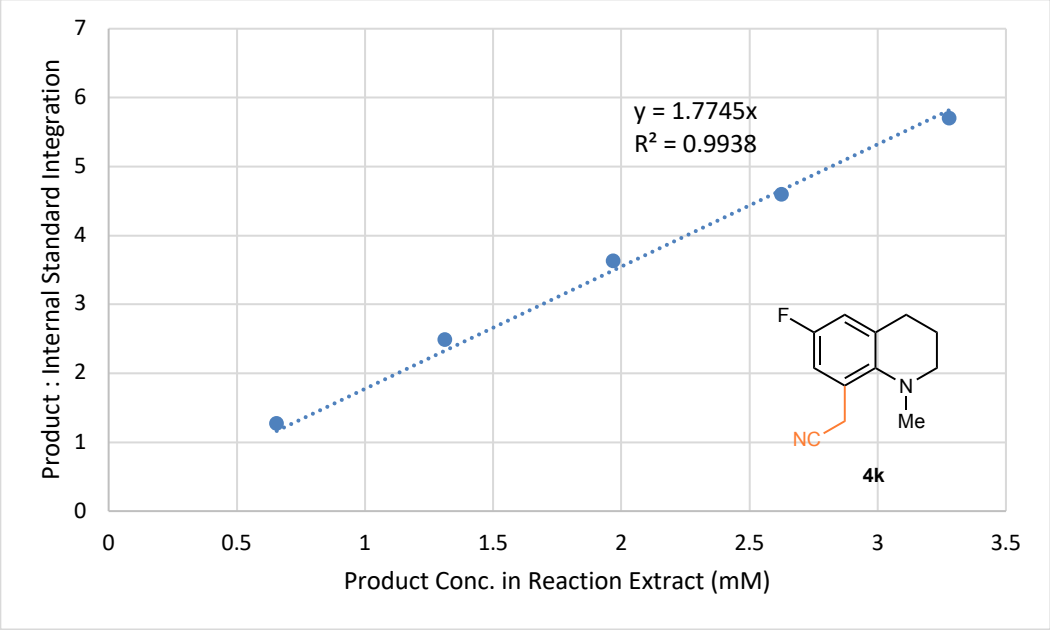

y = 1.6384x
R² = 0.9942
Product : Internal Standard Integration
Product Conc. in Reaction Extract (mM)
0
1
2
3
4
5
6
0
0.5
1
1.5
2
2.5
3
3.5
Me
N
Me
CN
4b

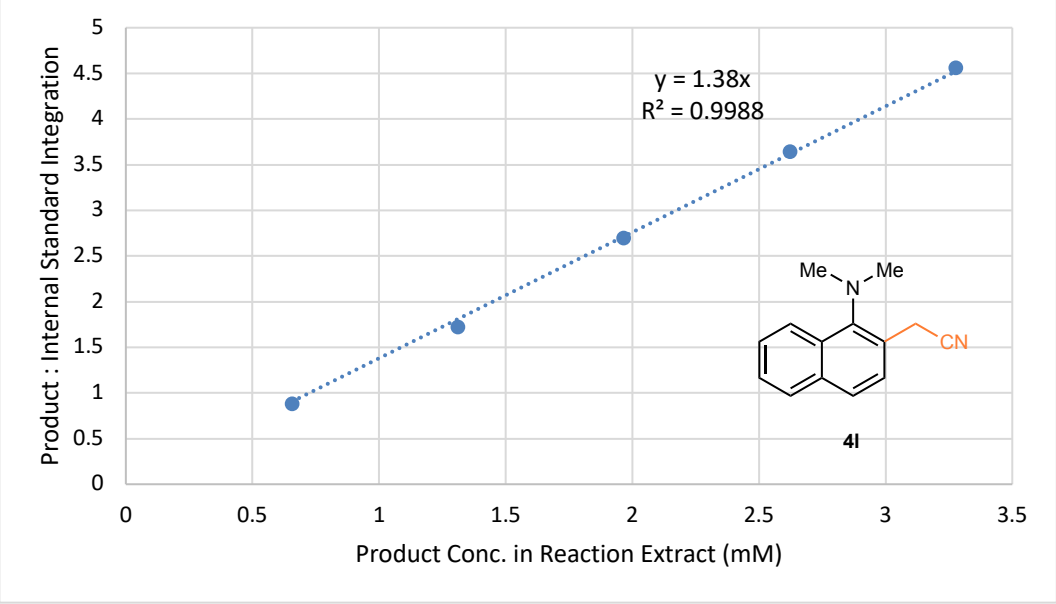

y = 1.5378x
R² = 0.9917
Product : Internal Standard Integration
Product Conc. in Reaction Extract (mM)
0
1
2
3
4
5
6
0
0.5
1
1.5
2
2.5
3
3.5
N
CN
4c

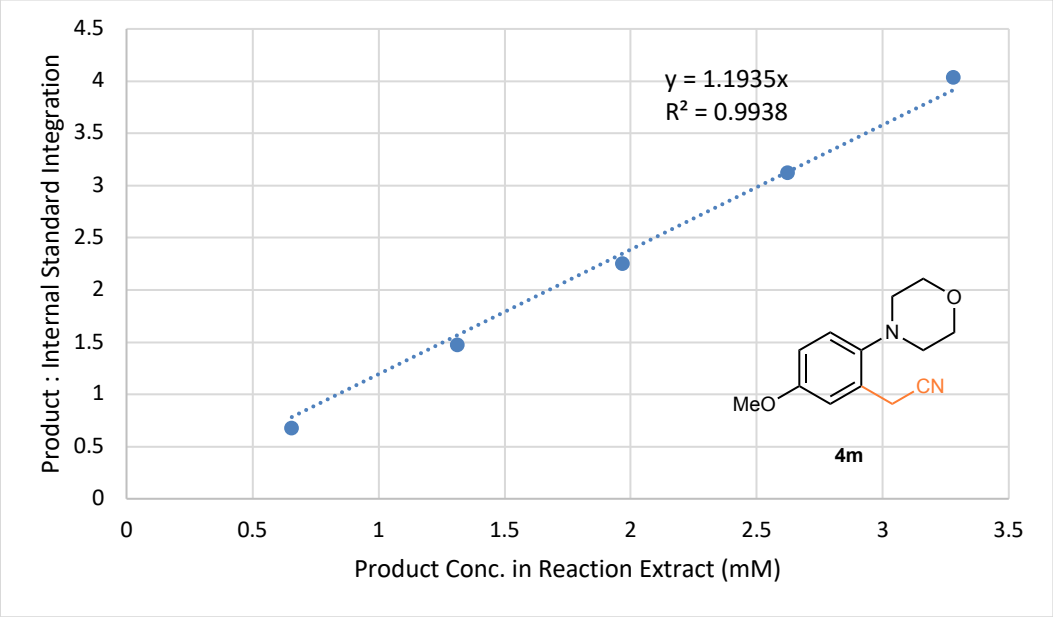

y = 1.8268x
R² = 0.9914
Product : Internal Standard Integration
Product Conc. in Reaction Extract (mM)
0
1
2
3
4
5
6
7
0
0.5
1
1.5
2
2.5
3
3.5
N
Me
NC
4d


y = 1.5406x
R² = 0.9988
Product : Internal Standard Integration
Product Conc. in Reaction Extract (mM)
0
1
2
3
4
5
6
0
0.5
1
1.5
2
2.5
3
3.5
N
Me
NC
4e


y = 2.0082x
R² = 0.9974
Product : Internal Standard Integration
Product Conc. in Reaction Extract (mM)
0
1
2
3
4
5
6
7
0
0.5
1
1.5
2
2.5
3
3.5
N
Me
CN
4j


y = 1.7745x
R² = 0.9938
Product : Internal Standard Integration
Product Conc. in Reaction Extract (mM)
F
N
Me
NC
4k


y = 1.38x
R² = 0.9988
Product : Internal Standard Integration
Product Conc. in Reaction Extract (mM)
Me
N
Me
CN
4l



B.8 ^{1}H and ^{13}C NMR spectra

^{1}H and ^{13}C NMR spectra of the following compounds can be found in the Supporting Information of the paper to be submitted.

B.9 References

1. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
2. Li, J. *et al.* Nickel-catalyzed amination of aryl chlorides with amides. *Org. Lett.* **23**, 687–691 (2021).
3. Xie, H. *et al.* Co(II)-catalyzed regioselective pyridine C–H coupling with diazoacetates. *Org. Lett.* **21**, 3427–3430 (2019).

4. Zeng, H., Cao, D., Qiu, Z. & Li, C.-J. Palladium-catalyzed formal cross-coupling of diaryl ethers with amines: Slicing the 4-*O*-5 Linkage in lignin models. *Angew. Chem. Int. Ed.* **57**, 3752–3757 (2018).
5. Jia, W.-L. *et al.* Selective C–H olefination of indolines (C5) and tetrahydroquinolines (C6) by Pd/S,O-ligand catalysis. *Org. Lett.* **21**, 9339–9342 (2019).
6. Wang, H., Huang, Y., Jiang, Q., Dai, X. & Shi, F. Reductive *N*-methylation of quinolines with paraformaldehyde and H_2 for sustainable synthesis of *N*-methyl tetrahydroquinolines. *Chem. Commun.* **55**, 3915–3918 (2019).
7. Shen, H. *et al.* Intramolecular aminocyanation of alkenes promoted by hypervalent iodine. *Org. Chem. Front.* **4**, 1806–1811 (2017).
8. Fu, M.-C., Shang, R., Cheng, W.-M. & Fu, Y. Boron-catalyzed *N*-alkylation of amines using carboxylic acids. *Angew. Chem. Int. Ed.* **54**, 9042–9046 (2015).
9. Verboom, W., Reinhoudt, D. N., Visser, R. & Harkema, S. '*Tert*-amino effect' in heterocyclic synthesis. Formation of *N*-heterocycles by ring-closure reactions of substituted 2-vinyl-*N*,*N*-dialkylanilines. *J. Org. Chem.* **49**, 269–276 (1984).
10. Wu, J. *et al.* Efficient construction of diverse 3-cyanoindoles under novel tandem catalysis. *Chem. Commun.* **56**, 12660–12663 (2020).

Chapter 4

STABILIZING METABOLIZING P450S FOR ABIOLOGICAL LATE-STAGE C–H FUNCTIONALIZATION

Abstract

Stability is an important criterion for synthetically useful biocatalysts. Enzymes with low stability often express poorly in heterologous hosts and have low tolerance to mutations which could be beneficial to the new function but at the same time deleterious to stability. This can prohibit the use of interesting biocatalysts in practical organic synthesis. In the past decades, many experimental and computational strategies have been developed to enhance the stability of proteins. Compared to other computational approaches, such as ancestral sequence reconstruction for example, consensus mutagenesis has the potential to achieve enhanced stability, but with minimal sequence manipulations, preserving the intrinsically promiscuous nature of the enzyme. Here, I describe our goal to transfer the abiological carbene- and nitrene-C–H insertion activity, which we have identified in P450-BM3-derived protein catalysts, into metabolizing P450s such as CYP3A4 to enable abiological late-stage functionalization of large substrates. Since we anticipated that CYP3A4's inherent poor stability would be a limiting factor, we applied consensus sequence mutagenesis algorithms to predict stability-enhancing mutations. Overall, we identified five mutations, which have the potential to improve CYP3A4's stability and subsequently enhance its evolvability.

4.1 Introduction

The appeal to use enzymes for C–H functionalization amplifies in the context of late-stage functionalization (LSF) of complex, large molecules.[1–3] In comparison to small-molecule catalysts, protein catalysts have many advantages: Through numerous non-covalent, corporative protein-substrate interactions, enzymes can often impart unmatched selectivity control. Additionally, interaction networks between amino acid residues and substrates can be altered and fine-tuned by directed evolution. Moreover, enzymatic transformations often proceed under mild conditions, which makes the use of enzymes highly attractive for LSF of delicate natural products. This is exemplified by the enzymatic late-stage hydroxylation of artemisinin (**Figure 4-1**).[4] Fasan and colleagues showed that laboratory-evolved P450-BM3 (CYP102A1, a P450 from *Bacillus megaterium*) variants can catalyze divergent, late-stage C–H hydroxylation of artemisinin with unparalleled high stereo- and site-selectivity that no current small-molecule hydroxylation catalysts can achieve.

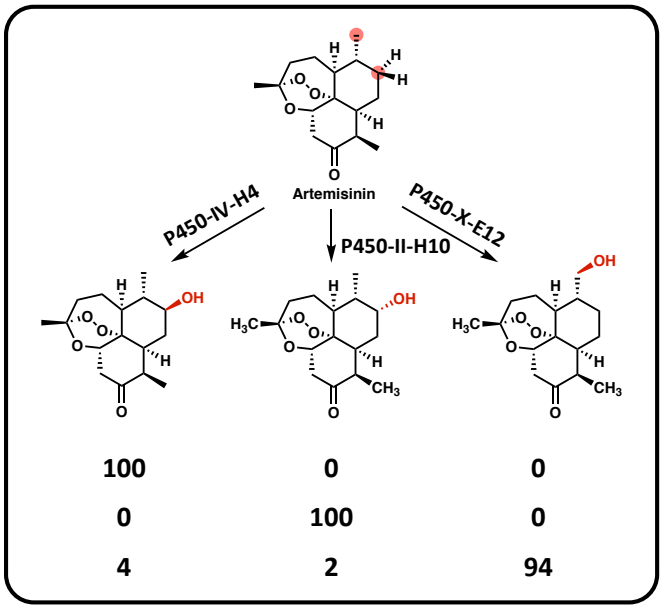


Figure 4-1. Divergent late-stage C–H hydroxylation catalyzed by engineered $P450_{BM3}$. Using a panel of complementary protein catalysts evolved from the same parent enzyme, high site- and stereoselectivity of C–H hydroxylation was achieved.

In the previous chapters, I have discussed that $P450_{BM3}$ can be repurposed to catalyze abiological C–H functionalization via carbene- and nitrene-transfer. However, adopting these engineered enzymes for abiological late-stage C–H functionalization remains a challenge. To date, none of our $P411_{BM3}$ variants, accrued over the course of past evolution efforts, has been shown to catalyze carbene- or nitrene-C–H insertion reactions on large substrates. In nature, $P450_{BM3}$'s active site is tailored to catalyze the hydroxylation of the subterminal positions in long chain fatty acids.[5] The shape and size of its active site is not adapted to fit large and bulky substrates properly. Despite previous studies showing that engineered $P450_{BM3}$ can, in fact, oxidize larger substrates (**Figure 4-2A**),[4,6–8] I hypothesize that the often greater steric demand of nitrene- or carbene-based C–H bond activations compared to oxo-based C–H activation (**Figure 4-2B**) is preventing these new-to-nature activities to be readily applicable for LSF of larger substrates.

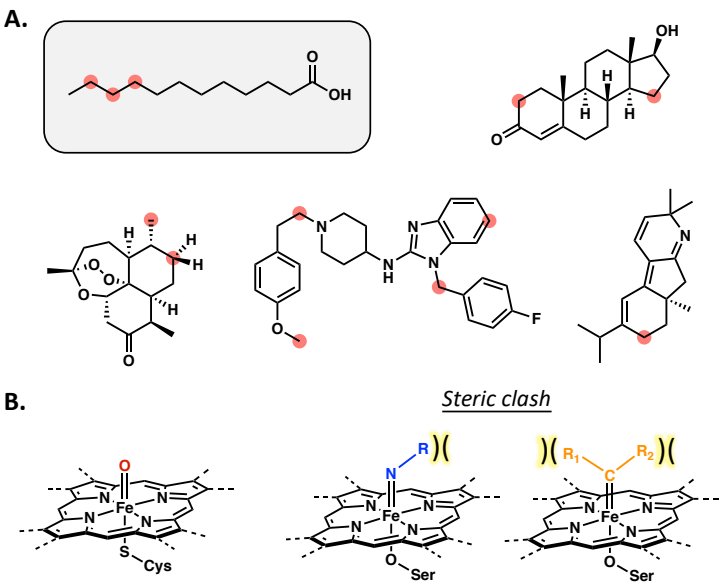


Figure 4-2. Using engineered $P450_{BM3}$ for late-stage C–H functionalization. **A)** Selected examples show the native substrate of $P450_{BM3}$ as well as larger substrates engineered $P450_{BM3}$ can oxidize (Red dots denote positions where hydroxylation can occur). **B)** Compared to the native C–H hydroxylation activity, abiological carbene- and nitrene-transfer activity may be rejected by $P450_{BM3}$ due to steric clashes with the carbene and nitrene intermediates.

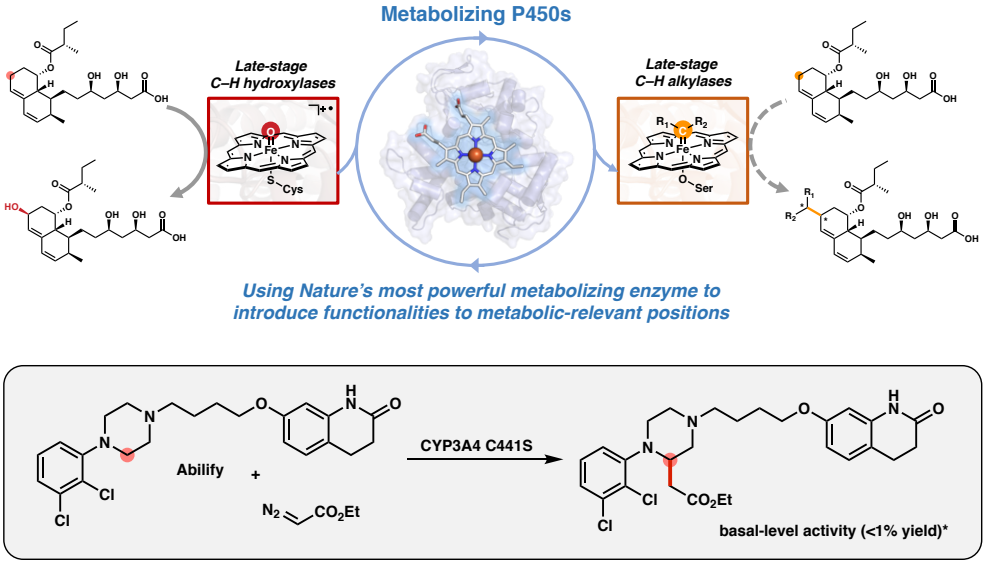


Figure 4-3. Repurposing metabolizing P450s for abiological LSF of complex molecules. (******Product mass of m/z = 534 is observed in LC-MS analysis. However, due to the low level of the activity, product identity was not fully confirmed*).

In contrast to $P450_{BM3}$, many eukaryotic P450s, especially those which are engaged in metabolism, often have substantially large substrate pockets and high scaffold flexibility.[9] For instance, a human P450, CYP3A4, has an active site volume of 1020 Å^3 in its substrate-bound form[10]. It is highly promiscuous to substrates with different geometries. CYP3A4 alone is responsible for 20% of xenobiotic metabolism in human body.[11] A matching reductase is needed for hydroxylation chemistry of P450s, but it is not required for carbene- or nitrene-transfer reactions. I thus propose that enzymes like CYP3A4 are potential candidates for late-stage abiological C–H functionalization. In an initial study, we introduced the axial cysteine-to-serine mutation in CYP3A4 and expressed the new variant in *Escherichia coli*. CYP3A4 C441S has shown promising initial late-stage C–H carbene insertion reaction activity toward Abilify, an anti-depressant drug molecule that contains multiple α-amino C–H bonds (**Figure 4-3**). Notably, C–H bonds of this kind are also often the most labile positions against metabolic clearance (by metabolizing P450s).[12–14] We surmised that systematically repurposing these "molecule-breaking" P450s into late-stage C–H alkylases or aminases can be a powerful strategy to introduce functionalities onto

positions which are of the most pharmacological relevance (e.g., metabolic stability) in drug candidates.

Despite the immense potential of employing metabolizing P450s like CYP3A4 for the new-to-nature LSF of bioactive molecules, a major hurdle associated with using or engineering them for synthetically useful transformations is their poor stability.[15,16] Like many other eukaryotic P450s, CYP3A4 is not thermostable – P450s in the 3A family have an average $^{60}T_{50}$ (the temperature at which half of the protein remains folded after 60 min) of ~40 °C. The lack of stability makes CYP3A4 difficult to use in biocatalysis. Additionally, laboratory evolution of an enzyme can be greatly aided by starting with a stable parent enzyme.[17,18] Mutations that are beneficial toward eliciting the new activity can be destabilizing, and an unstable parent enzyme may not tolerate the introduction of many of these mutations (**Figure 4-4**).

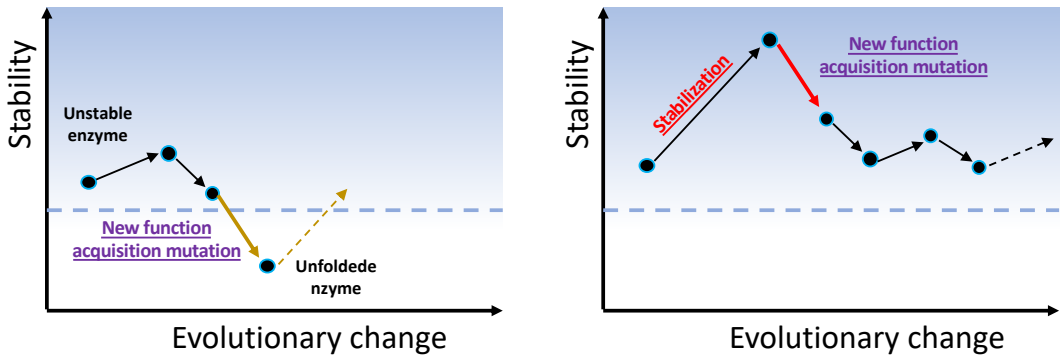


Figure 4-4. Evolution of an unstable enzyme with or without prior stabilization. Stability of an enzyme can significantly contribute to its evolvability towards new functions.

In this chapter, I will discuss efforts to use a computational approach, consensus sequence mutagenesis,[19,20] to reduce experimental burden and predict potential mutations or amino acid positions which may have the most substantial impact on CYP3A4's stability. Using this strategy, we predicted the five most probable mutations/sites which may improve CYP3A4's thermostability. Previous works using different methods to predict stabilizing mutations (ancestral sequence reconstruction) will also be compared and discussed in this

section.[16] Taken together, we hope to lay out a generalizable workflow which can turn interesting yet unstable enzymes into a stable biocatalyst for practical organic synthesis.

4.2 Approach

In this work, we applied consensus sequence design/mutagenesis to improve the stability of CYP3A4. The overarching hypothesis behind this method is that throughout natural evolution, protein families are results of iterative rounds of random mutagenesis, recombination, and selection, a process that tends to eliminate destabilizing mutations.[21] In other words, residues which stabilize a protein may be more conserved than those which do not stabilize. Through multiple-sequence alignment (MSA), we can identify these consensus mutations and then introduce them to the target protein, which may increase its thermostability. We aimed to identify residues of two characteristics in the CYP3A4 sequence: **1**) residues which are uncoupled, since disrupting interacting residue pairs can have detrimental effects on protein stability; and **2**) conserved residues, which are not present in CYP3A4 but prevalent in other P450 homologues. The detailed workflow is explained below.

4.2.1 Construction of a non-redundant sequence database and MSA

The construction of a non-biased MSA is the first and key step to identify consensus mutations which may stabilize the target protein. It is worth noting that some enzymes are better characterized than the other members in the protein family. As a result, the abundance of protein sequences in the database can be biased toward those which are better characterized (e.g., more annotated isoforms; more sequences contain single or few amino acid mutation(s) which are created during structural or functional studies). This may lead to biased conservation assessment which will eventually disrupt consensus prediction. To address this issue, we applied methods to reduce sequence redundancy: 1) We performed the BLAST search to collect sequences from the NCBI non-redundant database; 2) Subsequently, we performed clustering to further reduce protein isoforms which contribute to the sequence redundancy. The clustering step allows us to group sequences which share greater than 90% sequence identities to be categorized into the same cluster. Only one

representative sequence from each cluster was used to construct the non-redundant database for MSA.

4.2.2 Analysis of conservation and covariance

The next step for consensus mutagenesis is to identify potentially stabilizing mutations from the MSA. In this part, we searched for conserved residues without interrupting potential interaction pairs.[19,20,22,23] Besides the highly conserved residues which are critical for protein function, we can also discover other conserved residues that tend to be stabilizing. By mutating the residues of the target protein into the consensus residue at this position, we hoped to obtain variants with improved stability. One potential pitfall using this approach is altering residues that form interaction networks with other residues. The covariance of residues depicts the epistatic relationships between residue pairs from the protein sequence. Without ruling out residues from high covariance pairs, we could break the interaction pairs and consequently destabilize the protein. A stabilizing single amino-acid mutation should be of moderate to high conservation and low covariance. We used relative entropy (RE) and mutual information (MI) to quantitatively analyze the conservation and covariance of each amino acid position in the MSA, respectively. The relative entropy of a given position is calculated from **Eq. 1**,

$$RE = \Sigma p_x \ln \frac{p_x}{f_x} \tag{1}$$

with p_x, the frequency of amino acid x at a specific position, and f_x, the referencing frequency of amino acid x, if the distribution is not biased from the standard codon usage. A high RE value suggests the frequencies of amino acids that appear at a given position deviate from the neutral reference state (conserved), whereas positions which are not conserved should have an RE value near zero.

Mutual information between two positions in the MSA is determined by **Eq. 2**,

$$MI\ (i,j) = \Sigma_i \Sigma_j p_{x,y} \ln \frac{p_{x,y}}{p_x p_y} \tag{2}$$

where $p_{x,y}$ is the frequency of amino acid x and y simultaneously appear at position i and j, p_x is the frequency of amino acid x at position i, and p_y is the frequency of amino acid y at position j. A high MI value suggests residues at position i and j are of high covariance and therefore should not be altered to maintain the enzyme's stability.

4.3 Results and discussion

To obtain a high quality, non-redundant MSA, we first performed a BLAST search to gather protein sequences from the NCBI non-redundant database using the amino acid sequence of CYP3A4 (UniProt accession ID: P08684) as the seed sequence. The top 5,000 protein sequences sharing 30–90% similarity to CYP3A4 were collected. Given that different protein isoforms can lead to phylogenetic bias in the final consensus sequence, we further minimized the sequence redundancy by using clustering. We used CD-HIT to cluster the original database with a sequence identify of 0.9 and *E*-value as 1.0 as set cutoff. This allows sequences that share greater than 90% identity to be grouped into the same cluster. Using CD-HIT, we clustered the 5,000 sequences into 2,146 individual clusters and a representative sequence from each cluster was chosen to construct a new sequence database, with which we subsequently performed MSA. The MSA was generated using Clustal Omega. Some of the sequences used for MSA shared low sequence identities which results in gaps. Columns in the MSA with greater than 50% gaps are not carried on for subsequent calculations, and this results in 456 sequences in the curated MSA.

The next step was to identify uncoupled and conserved positions at which CYP3A4 has a different amino acid compared to the consensus sequence from the MSA. To eliminate coupled residues, we used the Python Biotite package to calculate the MI values. In parallel, we have also adopted a web tool, MISTIC,[24] to help visualize the conservation and covariance of each position in the MSA. The Mutual Information Circos generated from MISTIC illustrates the conservation and covariance at each amino acid position (**Figure 4-5**). We have eliminated positions of residue pairs with calculated MI values ≥ 0.5 from the list. Highly coupled sites which were eliminated from the study are summarized in **Appendix C.3**. These sites were further confirmed using the Mutual Information Circos. Additionally,

residues 2–22 in CYP3A4 belong to the membrane-bound domain and are eliminated experimentally to aid heterologous expression. Therefore, they were also not further considered for consensus mutagenesis.

We next worked on recognizing positions which are highly or moderately conserved. To visualize the conservation level of each position, we adopted Seq2Logo to generate the consensus sequence with Shannon logo (**Figure C-1**, **Appendix C**).[25] The size of the amino acid letter represents its conservation level at the given position) and compared it with the sequence of CYP3A4. As expected, the positions of highest conservation are residues which are critical for P450 function (e.g., the axial cysteine ligand C441; **Figure 4-6A**). CYP3A4's sequence is identical to the consensus sequence at these positions. As we further reduced the conservation cutoff, we found that the CYP3A4 sequence starts to deviate from the consensus sequence at positions of moderate conservation levels (**Figure 4-6B**). We reasoned that these mutations are most likely to improve the stability of CYP3A4. The five mutations falling into this category are V146F, L248F, V349L, F366Y, and M385V. The positions of these five most promising mutations are denoted in the CYP3A4 structure (PDB ID: 1TQN; **Figure 4-7**). All of the five mutations are located near the surface of the protein, which indicates these consensus mutations may be introduced without substantially perturbing CYP3A4's function.

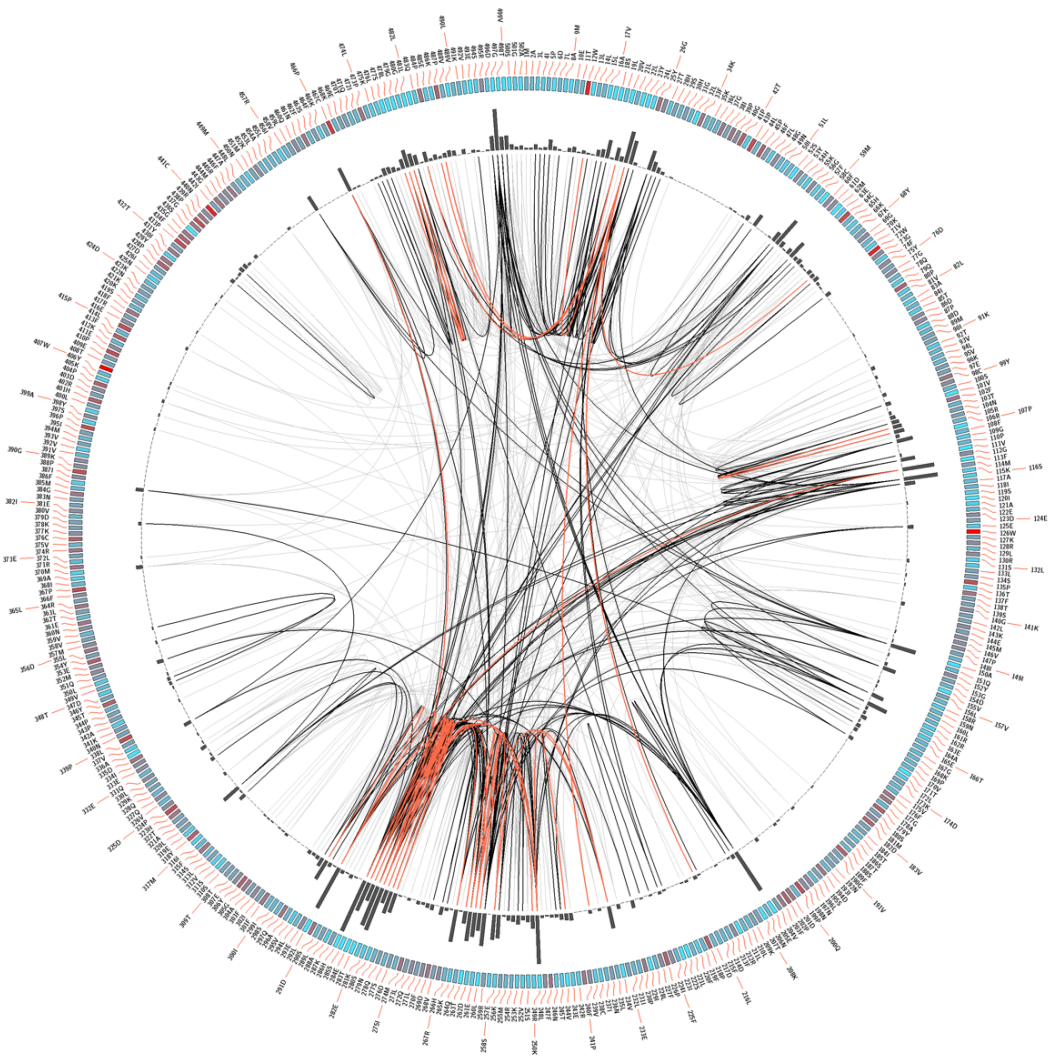


Figure 4-5. The Mutual Information Circos generated from MISTIC indicates the covariance between residues in CYP3A4. The red, black, and grey lines between the residue pairs indicate the pairs have calculated MI values of top 5%, 30%, and bottom 70%. Additionally, the color boxes near each residue indicates the conservation level at these positions (red box indicates the residue is highly conserved, while blue suggest low conservation level or non-conserved residues).

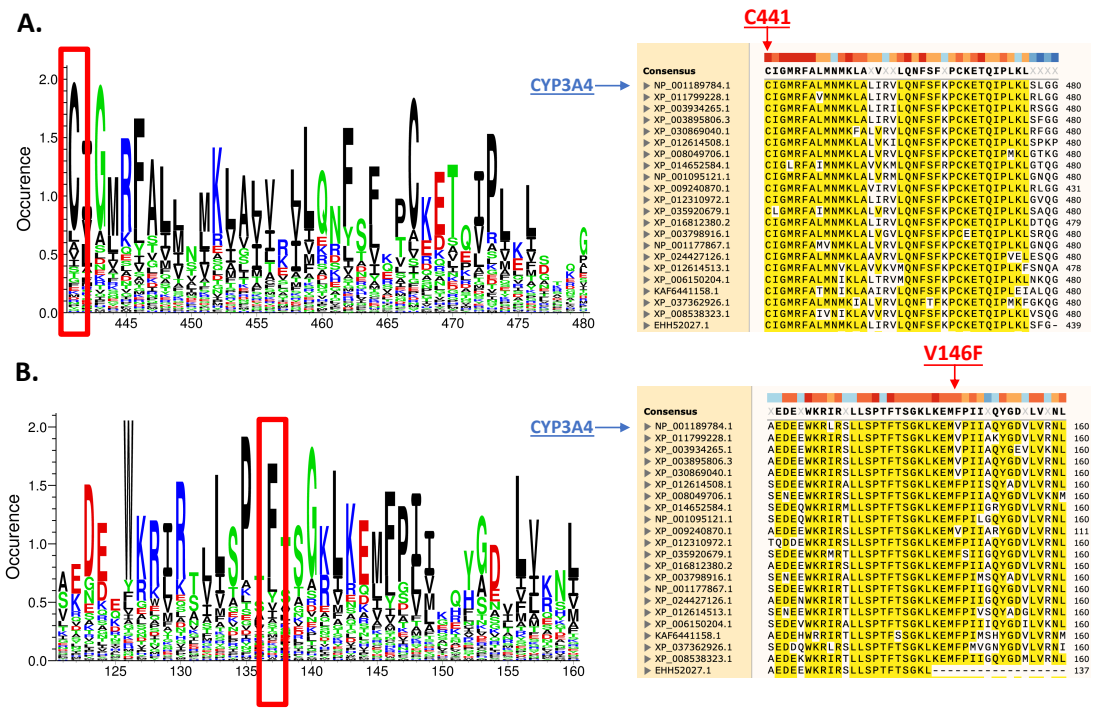


Figure 4-6. Determination of consensus mutations. **A)** Highly conserved residues, such as the axial cysteine C441 is key to P450s function; **B)** Mutations to the consensus sequence at moderately conserved residues, such as position 146, are desired mutations that may stabilize CYP3A4.

Of note is that in a recent study,[16] Gillam and colleagues postulated that ancestral sequence reconstruction can be a powerful tool to stabilize enzymes, based on the assumption that pre-Cambrian enzymes might be more stable than their modern forms.[26] Indeed, they have demonstrated that using this approach, the predicted common ancestor of the P450s from the 3A family, CYP3_N1, exhibits a $^{60}T_{50}$ of 66°C (~30°C higher than CYP3A4). However, CYP3_N1 differs by 140 mutations from CYP3A4, which corresponds to approximately 30% of the residues in CYP3A4. An in-depth statistically significant phylogenetic analysis spanning 1005 enzyme families by Herschlag *et al.*[27] showed that only limited numbers of mutations in each interaction network are required to switch a non-stable enzyme into a thermostable counterpart. Hence, extensive sequence manipulation is not necessary–the majority of the mutations might not be beneficial for its stability. Moreover,

potential allosteric regulation which contributes to CYP3A4's high substrate promiscuity might be lost after excessive mutagenesis. Indeed, all of the five most probable stabilizing mutations we predicted using consensus sequence mutagenesis are also present in CYP3_N1. We surmised that consensus mutagenesis has the advantage of predicting stable variation of unstable proteins with minimal sequence manipulation.

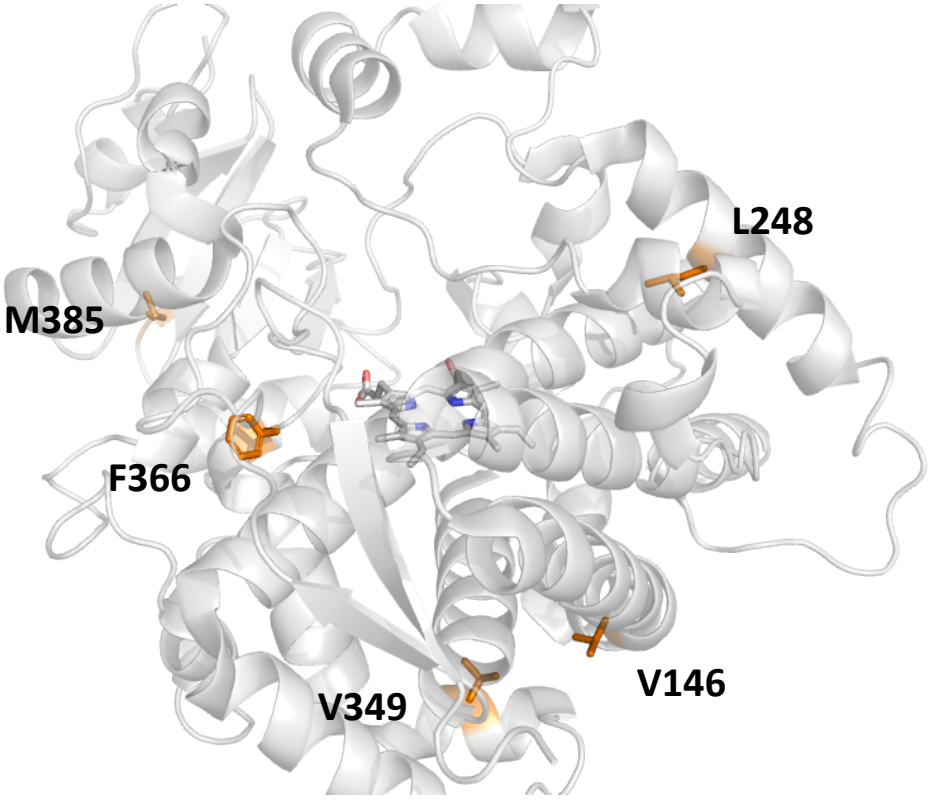


Figure 4-7. Illustration of the resulting five residues in CYP3A4 structure (PDB ID: 1TQN). The five mutations are located near the surface of the protein.

4.4 Conclusion and outlook

In this chapter, we proposed the transfer of the new-to-nature activities identified in bacterial P450-BM3 into larger, more promiscuous P450s such as CYP3A4 for late-stage C–H functionalization. To achieve this, we first needed to lay the groundwork for stabilizing, improving the evolvability, and achieving high expression levels with our candidate P450,

CYP3A4. We applied consensus sequence mutagenesis to predict stabilizing mutations. Our study narrowed down the possibilities and indicated the five most probable mutations which might be beneficial to CYP3A4's stability. These predictions need to be further validated by experiments, but we envisioned that by introducing these mutations directly or by performing site-saturation mutagenesis at these sites can eventually turn CYP3A4 into a thermostable enzyme.

4.5 References

1. Cernak, T., Dykstra, K. D., Tyagarajan, S., Vachal, P. & Krska, S. W. The medicinal chemist's toolbox for late stage functionalization of drug-like molecules. *Chem. Soc. Rev.* **45**, 546–576 (2016).

2. Börgel, J. & Ritter, T. Late-stage functionalization. *Chem* **6**, 1877–1887 (2020).

3. Guillemard, L., Kaplaneris, N., Ackermann, L. & Johansson, M. J. Late-stage C–H functionalization offers new opportunities in drug discovery. *Nat. Rev. Chem.* **5**, 522–545 (2021).

4. Zhang, K., Shafer, B. M., Demars, M. D., Stern, H. A. & Fasan, R. Controlled oxidation of remote sp^3 C–H bonds in artemisinin via P450 catalysts with fine-tuned regio- and stereoselectivity. *J. Am. Chem. Soc.* **134**, 18695–18704 (2012).

5. Whitehouse, C. J. C., Bell, S. G. & Wong, L.-L. $P450_{BM3}$ (CYP102A1): Connecting the dots. *Chem Soc Rev* **41**, 1218–1260 (2012).

6. Sawayama, A. M. *et al.* A panel of cytochrome P450 BM3 variants to produce drug metabolites and diversify lead compounds. *Chem. - Eur. J.* **15**, 11723–11729 (2009).

7. Kille, S., Zilly, F. E., Acevedo, J. P. & Reetz, M. T. Regio- and stereoselectivity of P450-catalysed hydroxylation of steroids controlled by laboratory evolution. *Nat. Chem.* **3**, 738–743 (2011).

8. Loskot, S. A., Romney, D. K., Arnold, F. H. & Stoltz, B. M. Enantioselective total synthesis of nigelladine A via late-stage C–H oxidation enabled by an engineered P450 enzyme. *J. Am. Chem. Soc.* **139**, 10196–10199 (2017).

9. Atkins, W. M. Mechanisms of promiscuity among drug metabolizing enzymes and drug transporters. *FEBS J.* **287**, 1306–1322 (2020).

10. Poulos, T. L. Heme enzyme structure and function. *Chem. Rev.* **114**, 3919–3962 (2014).

11. Rendic, S. & Guengerich, F. P. Survey of human oxidoreductases and cytochrome P450 enzymes involved in the metabolism of xenobiotic and natural chemicals. *Chem. Res. Toxicol.* **28**, 38–42 (2015).

12. Burka, L. T., Guengerich, F. P., Willard, R. J. & Macdonald, T. L. Mechanism of cytochrome P-450 catalysis. Mechanism of *N*-dealkylation and amine oxide deoxygenation. *J. Am. Chem. Soc.* **107**, 2549–2551 (1985).

13. Coleman, P. J. *et al.* Discovery of [(2*R*,5*R*)-5-[(5-fluoropyridin-2-yl)oxy]methyl-2-methylpiperidin-1-yl][5-methyl-2-(pyrimidin-2-yl)phenyl]methanone (MK-6096): A dual orexin receptor antagonist with potent sleep-promoting properties. *ChemMedChem* **7**, 415–424 (2012).

14. Dang, N. L., Hughes, T. B., Miller, G. P. & Swamidass, S. J. Computationally assessing the bioactivation of drugs by *N*-dealkylation. *Chem. Res. Toxicol.* **31**, 68–80 (2018).

15. Gillam, E. M. J., Baba, T., Kim, B. R., Ohmori, S. & Guengerich, F. P. Expression of modified human cytochrome P450 3A4 in *Escherichia coli* and purification and reconstitution of the enzyme. *Arch. Biochem. Biophys.* **305**, 123–131 (1993).

16. Gumulya, Y. *et al.* Engineering highly functional thermostable proteins using ancestral sequence reconstruction. *Nat. Catal.* **1**, 878–888 (2018).

17. Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl. Acad. Sci.* **103**, 5869–5874 (2006).

18. Tokuriki, N. & Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19**, 596–604 (2009).

19. Amin, N. *et al.* Construction of stabilized proteins by combinatorial consensus mutagenesis. *Protein Eng. Des. Sel.* **17**, 787–793 (2004).

20. Sullivan, B. J. *et al.* Stabilizing proteins from sequence statistics: The interplay of conservation and correlation in triosephosphate isomerase stability. *J. Mol. Biol.* **420**, 384–399 (2012).

21. Kimura, M. Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. *Proc. Natl. Acad. Sci.* **88**, 5969–5973 (1991).

22. Porebski, B. T. & Buckle, A. M. Consensus protein design. *Protein Eng. Des. Sel.* **29**, 245–251 (2016).

23. Sternke, M., Tripp, K. W. & Barrick, D. Consensus sequence design as a general strategy to create hyperstable, biologically active proteins. *Proc. Natl. Acad. Sci.* **116**, 11275–11284 (2019).

24. Simonetti, F. L., Teppa, E., Chernomoretz, A., Nielsen, M. & Marino Buslje, C. MISTIC: Mutual information server to infer coevolution. *Nucleic Acids Res.* **41**, W8–W14 (2013).

25. Thomsen, M. C. F. & Nielsen, M. Seq2Logo: A method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res.* **40**, W281–W287 (2012).

26. Gaucher, E. A., Thomson, J. M., Burgan, M. F. & Benner, S. A. Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature* **425**, 285–288 (2003).

27. Pinney, M. M. *et al.* Parallel molecular mechanisms for enzyme temperature adaptation. *Science* **371**, eaay2784 (2021).

Appendix C

SUPPLEMENTARY INFORMATION FOR CHAPTER 4

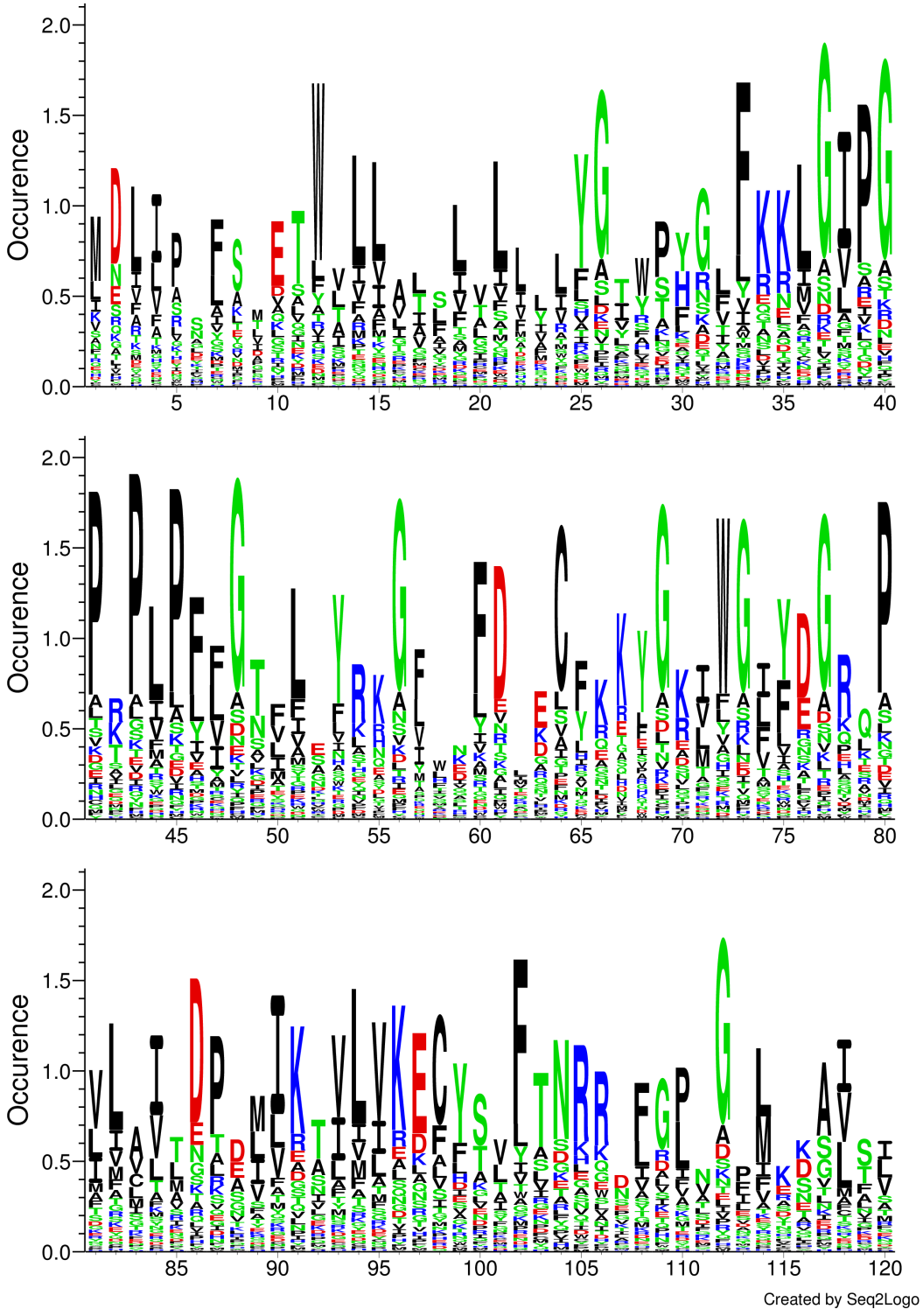
C.1 Amino acid sequence of CYP3A4.

>sp|P08684|CP3A4_HUMAN Cytochrome P450 3A4 OS=Homo sapiens OX=9606 GN=CYP3A4 PE=1 SV=4

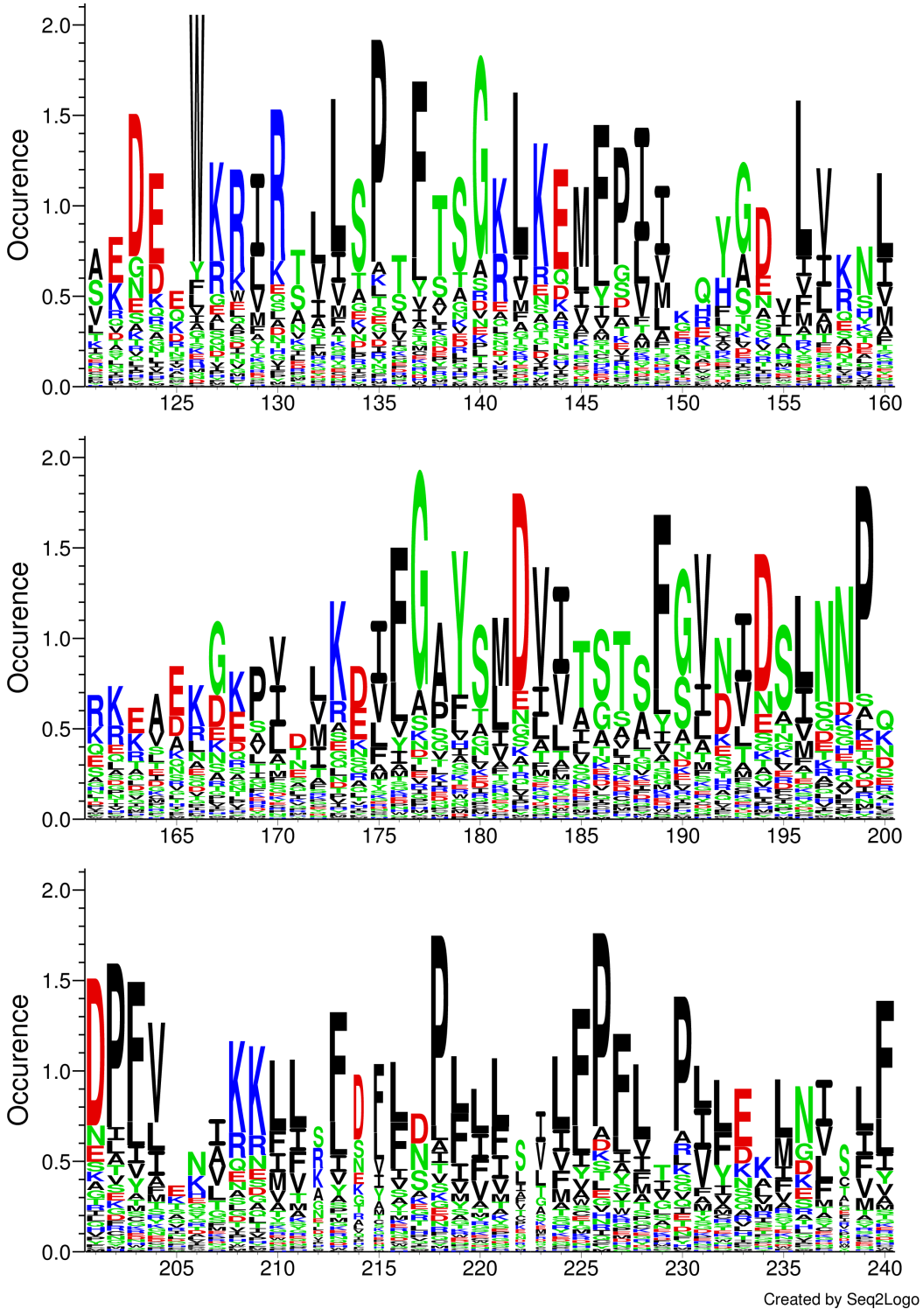
MALIPDLAMETWLLLAVSLVLLYLYGTHSHGLFKKLGIPGPTPLPFLGNILSYHKGFCMF
DMECHKKYGKVWGFYDGQQPVLAITDPDMIKTVLVKECYSVFTNRRPFGPVGFMKSAISI
AEDEEWKRLRSLLSPTFTSGKLKEMVPIIAQYGDVLVRNLRREAETGKPVTLKDVFGAYS
MDVITSTSFGVNIDSLNNPQDPFVENTKKLLRFDFLDPFFLSITVFPFLIPILEVLNICV
FPREVTNFLRKSVKRMKESRLEDTQKHRVDFLQLMIDSQNSKETESHKALSDLELVAQSI
IFIFAGYETTSSVLSFIMYELATHPDVQQKLQEEIDAVLPNKAPPTYDTVLQMEYLDMVV
NETLRLFPIAMRLERVCKKDVEINGMFIPKGVVVMIPSYALHRDPKYWTEPEKFLPERFS
KKNKDNIDPYIYTPFGSGPRNCIGMRFALMNMKLALIRVLQNFSFKPCKETQIPLKLSLG
GLLQPEKPVVLKVESRDGTVSGA

C.2 Visualization of the consensus sequence

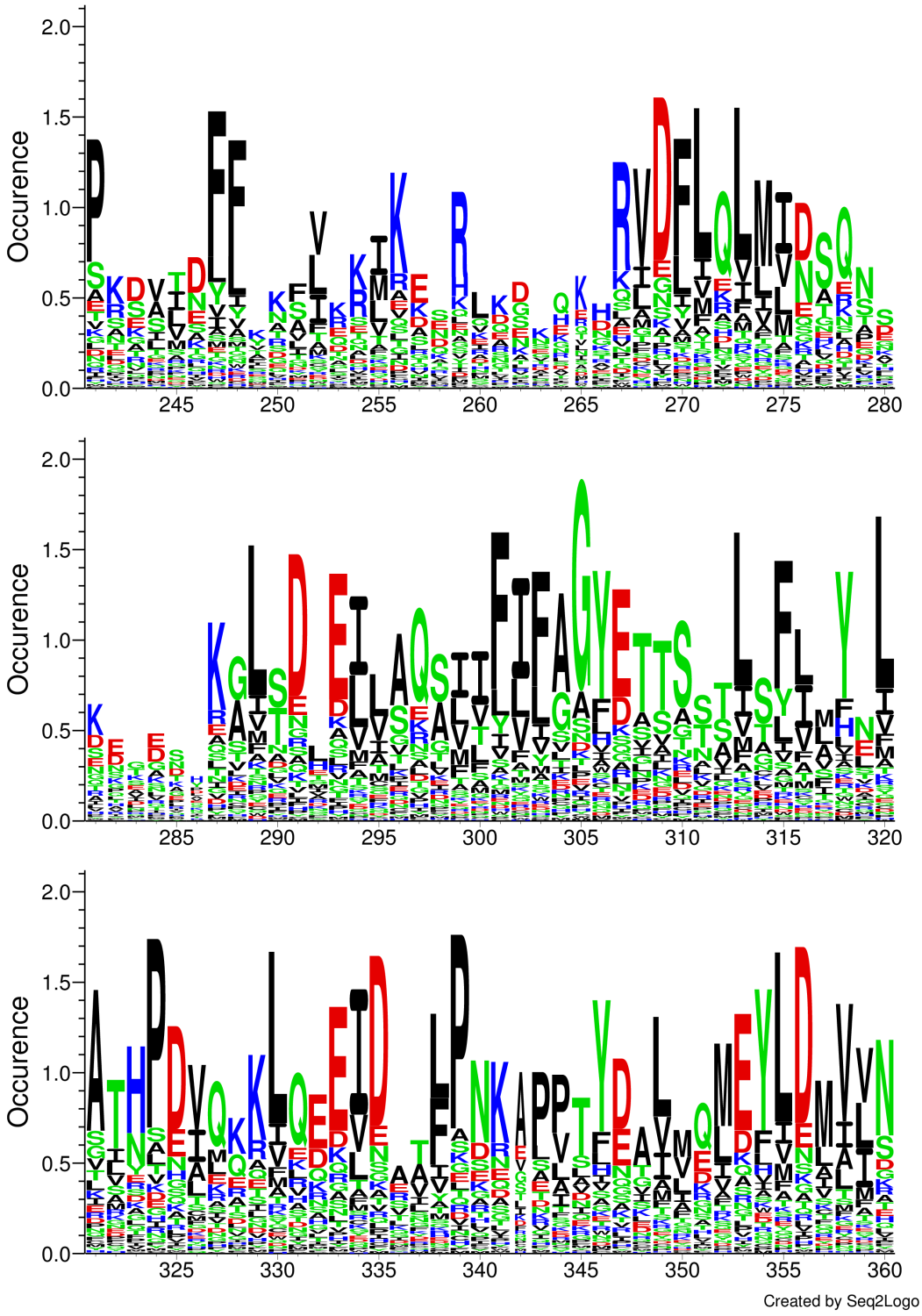
The curated MSA containing 435 homologous sequences (deposited on the Arnold lab internal server) is plotted using Seq2Logo to help visualize the residue conservation at each amino acid position. The result is displayed in Shannon logo. The relative size of the amino acid letters at each position indicate the amino acid occurrence at this position (larger font means higher occurrence), while the cumulative size of each column denotes the conservation level of the given position. The result is shown in **Figure C-1**.


Occurence
2.0
1.5
1.0
0.5
0.0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
Created by Seq2Logo

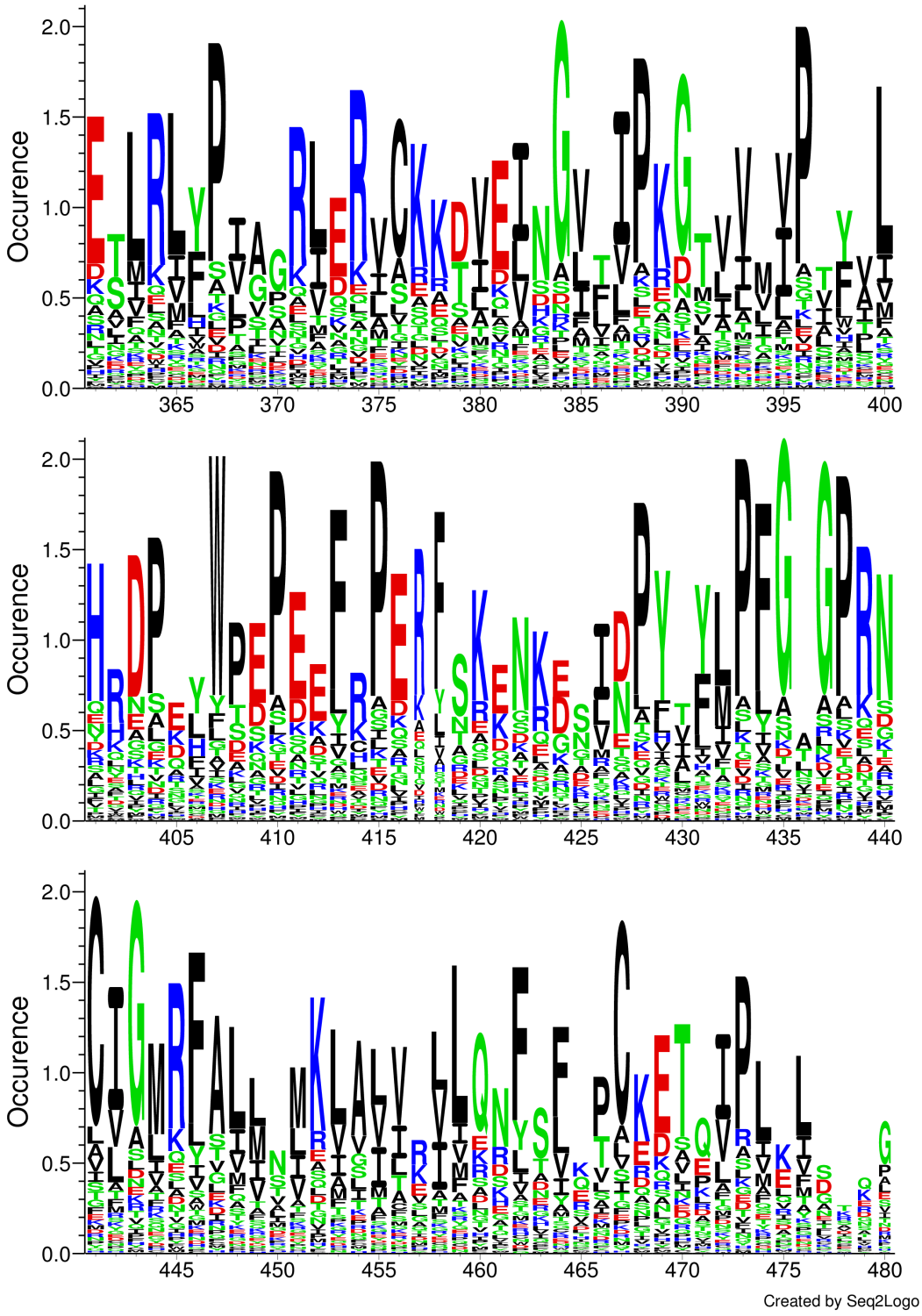
(Continued)


Occurence
2.0
1.5
1.0
0.5
0.0
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
Created by Seq2Logo

(Continued)


Occurence
2.0
1.5
1.0
0.5
0.0
245
250
255
260
265
270
275
280
Occurence
2.0
1.5
1.0
0.5
0.0
285
290
295
300
305
310
315
320
Occurence
2.0
1.5
1.0
0.5
0.0
325
330
335
340
345
350
355
360
Created by Seq2Logo

(Continued)


Occurence
2.0
1.5
1.0
0.5
0.0
365
370
375
380
385
390
395
400
Occurence
2.0
1.5
1.0
0.5
0.0
405
410
415
420
425
430
435
440
Occurence
2.0
1.5
1.0
0.5
0.0
445
450
455
460
465
470
475
480
Created by Seq2Logo

(Continued)

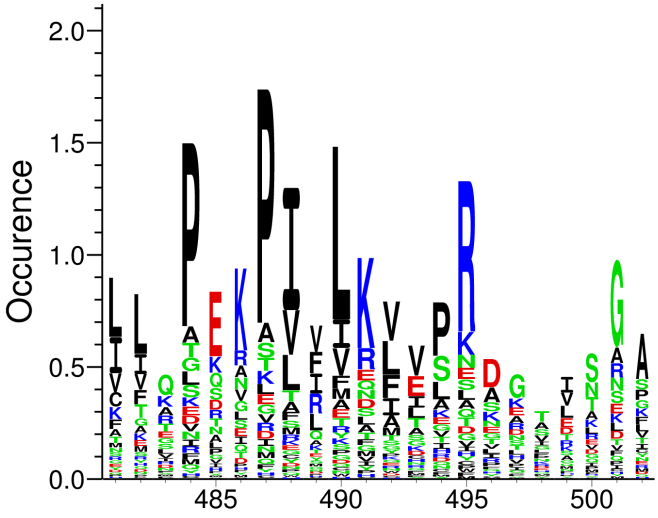


Figure C-1. Visualization of the conservation level of each amino acid position in consensus sequence.

C.3 List of residues that belong to interaction pairs with high MI value

(Position 2-22 are not considered, since it is the membrane-bound domain)

Y68, N104, R105, F113, F215, V234, C238, E243, R249, K250, S258, E261, D262, T263, Q264, H266, N279, S280, K281, T283, E284, S285, H286, L292, V295, K465, S477, L478, G479, L482, E485, K486.

C.4 Other moderately conserved mutations that may stabilize CYP3A4

N49T, Q78R, L129I, T166K, I224L, T348A, V391T, T408P, K412E, L414R, I456V, V488I