Biocatalytic Lactone Carbene C–H, B–H, and N–H Insertion Reactions Enabled by Directed Evolution

Thesis by
Andrew Z. Zhou

In Partial Fulfillment of the Requirements for the
Degree of
Bachelor of Science (Chemistry)

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2021
Defended June 4

Andrew Z. Zhou
ORCID: 0000-0002-1763-7353

ACKNOWLEDGEMENTS

I would like to express my gratitude to Professor Frances Arnold for the opportunity to spend all four years of college working in her research group. It was an honor to contribute to such fascinating research, and it has certainly been one of the experiences that has influenced me most. She is an amazing role model and leader.

I was lucky to be mentored by Dr. Kai Chen (now a post-doc at Berkeley in the Doudna lab) and Dr. Zhen Liu. I worked with Kai beginning in my freshman year, and I am so grateful that I was able to start my scientific training with him. Kai was so generous with his time in teaching me lab techniques, revising my writing (including this thesis), answering all sorts of questions, and providing research guidance. His work ethic and brilliance in the lab astound me, and he is one of my main role models. I am also thankful I had the opportunity to work with Zhen. He is kind, thoughtful, and an incredible scientist. Working with him on the N–H insertion project was a great experience. Zhen has also provided valuable feedback on this thesis and other written works.

I would also like to sincerely thank Dr. Sabine Brinkmann-Chen. Sabine taught me lab techniques, helped me navigate some intricacies of research, watched out for my lab safety, and provided research guidance. Her ability to manage more than 20 students and post-docs is truly admirable. She also has had the unenviable responsibility of reviewing and editing multiple drafts of this thesis and other written works. Her comments and revisions were invaluable to this work and to my development as a writer.

Lastly, I would like to express my deepest gratitude to my parents, Dr. Zhen Li and Zhiqing Zhou. I would not be here without their unconditional love and support.

This work has been supported in part by the Shih Family SURF Fellowship, the Stauffer Trust, and the Caltech SURF program.

ABSTRACT

Enzymes are biological catalysts, and they accelerate reactions by lowering the activation barrier. In nature, enzymes have been optimized by natural selection and possess precise three-dimensional active sites. With these active sites, they can typically catalyze reactions with high efficiency and specificity. Compared to traditional catalysts, enzymes are generally more environmentally friendly, and they can catalyze reactions in water and at ambient temperature and pressure. However, native enzymes are usually only well suited for a restricted range of substrates and are limited in the types of reactions they perform. The Arnold lab has recently focused on endowing enzymes with the ability to catalyze new-to-nature reactions through directed evolution. Here, we present a set of enzymes engineered for the ability to insert a lactone carbene into B–H, C–H, and N–H bonds with high yield and enantioselectivity. B–H insertion is achieved by engineered cytochrome *c* enzymes, while N–H and C–H insertions are achieved by engineered cytochrome P450 enzymes. With this work, we expand nature's toolbox for lactone insertion chemistry. Since lactones are highly bioactive, these engineered enzymes could be powerful tools in the synthesis of a range of pharmaceuticals and natural product targets.

PUBLISHED CONTENT AND CONTRIBUTIONS

(1) Liu, Z.; Calvo-Tussell, C.; Zhou, A. Z.; Chen, K.; Garcia-Borras, M.; Arnold, F. H. *ChemRxiv* **2021**,
(Preprint) A.Z.Z participated in the evaluation of enzymes with added FAD domains, the substrate scope study, the synthesis of standards and the large scale enzymatic reactions.
https://doi.org/10.26434/chemrxiv.14452158.v1.

(2) Zhou, A. Z.; Chen, K.; Arnold, F. H. *ACS Catal.* **2020**, *10*, 5393–5398,
A.Z.Z participated in the evolution of the enzyme, the evaluation of the substrate scope, the large scale enzymatic reactions, and the synthesis of standards and substrates.
https://doi.org/10.1021/acscatal.0c01349.

(3) Chen, K.; Huang, X.; Zhang, S. Q.; Zhou, A. Z.; Kan, S. B. J.; Hong, X.; Arnold, F. H. *Synlett* **2019**, *30*, 378–382,
A.Z.Z participated in the evaluation of enzyme scope and the synthesis of standards and substrates.
https://doi.org/10.1055/s-0037-1611662.

TABLE OF CONTENTS

Acknowledgements . . . iii
Abstract . . . iv
Published Content and Contributions . . . v
Table of Contents . . . v
Chapter I: Introduction . . . 1
1.1 Directed Evolution of Enzymes . . . 1
1.2 The Evolution of Enzymatic Carbene Transferases . . . 2
1.3 Brief Background on Lactones . . . 5
Chapter II: Lactone Carbene B–H Insertion . . . 8
2.1 Project Inspiration . . . 8
2.2 Previously Conducted Enzyme Optimization . . . 9
2.3 Substrate Scope . . . 11
2.4 Conclusion . . . 13
Chapter III: Lactone Carbene C–H Insertion . . . 14
3.1 Conception and Basis for Reactivity . . . 14
3.2 Identifying the Best Variant . . . 15
3.3 Evolution of a Biocatalytic Platform . . . 16
3.4 Substrate Scope . . . 18
3.5 Potential for N–H Insertion . . . 20
3.6 Potential Applications in Pharmacuetical Synthesis . . . 21
3.7 Conclusion . . . 23
Chapter IV: Lactone Carbene N–H Insertion . . . 24
4.1 Identifying the Best Variant . . . 26
4.2 Substrate Scope . . . 28
4.3 Large-Scale Reactions . . . 31
4.4 Conclusion . . . 32
Chapter V: Conclusions and Future Directions . . . 33
Chapter VI: Supporting Information . . . 34
6.1 Main Methods . . . 34
Bibliography . . . 39

Chapter 1

INTRODUCTION

1.1 Directed Evolution of Enzymes

Directed evolution emerged in the 1990s as a way to engineer proteins with enhanced natural and new-to-nature properties.[1] To improve the natural functions of an enzyme, directed evolution has been used to introduce and increase thermo- or solvent stability, improve activity on natural substrates, and expand substrate specificity/scope. In recent years, directed evolution has also been used to endow proteins with the ability to perform entirely new reactions. Developed primarily by Frances Arnold at Caltech in response to the difficulty of predicting protein function based on its sequence, directed evolution is a robust and reliable way to tailor complex proteins for a specific function.[2] To some extent, directed evolution mimics natural selection in an experimental setting. Researchers create gene libraries of variants of a "parent" protein, express these variants, screen for a desired function, and select the one with the highest activity. The gene coding for the best variant is used as the "parent' in a subsequent round. This iterative process is repeated until the desired functionality is achieved (Figure 1.1).

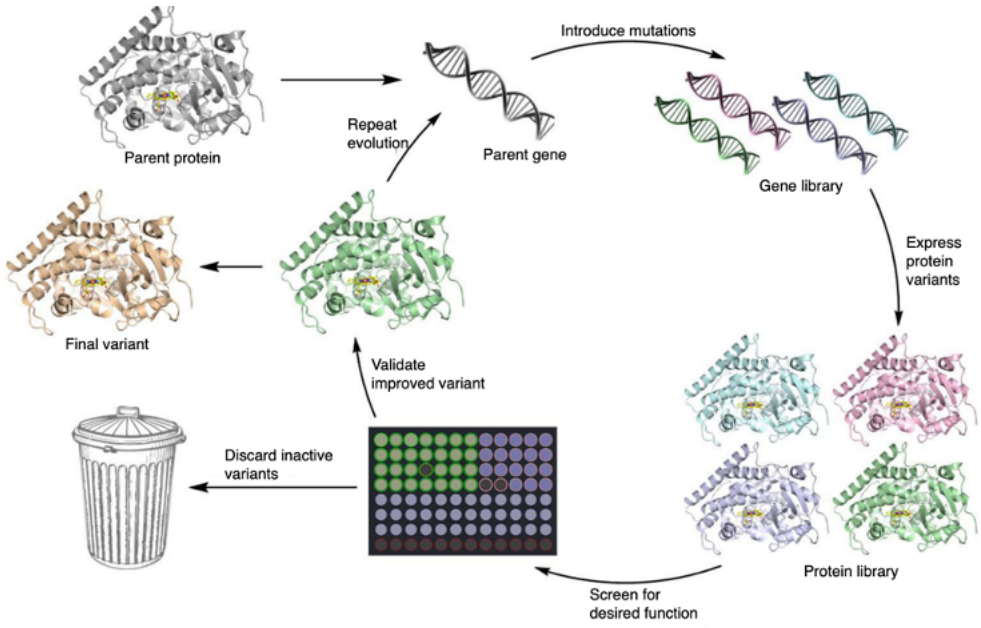


Figure 1.1: An overview of the process of directed evolution. Through an iterative process of mutagenesis and screening, proteins can be optimized for desired functions.

Directed evolution has been applied to the engineering of a diverse range of proteins including antibodies,[3] biotechnology tools such as CRISPR-Cas proteins for genome editing,[4] genetic circuits[5] and biocatalysts.[6] Biocatalysts are the focus of the research in the Arnold lab. Enzymes have risen to prominence in performing chemical transformations because they can have high efficiency and exquisite selectivity, enabled by their tunable three-dimensional active sites, while reducing the environmental hazards produced. In addition, they are genetically encoded and can be expressed in microbes such as *Escherichia coli*. They rely on earth-abundant metals such as iron or zinc, and they can conduct reactions at ambient temperature and pressure.[7] Due to these advantages, among others, enzymes have been slowly taking over some of the niche occupied by traditional transition metal catalysts.[8,9] Since naturally occurring enzymes are typically limited in scope, there is need for researchers to engineer variants for diverse applications.

1.2 The Evolution of Enzymatic Carbene Transferases

In 1999, the Arnold lab started engineering cytochromes P450, which naturally perform oxidative chemistry such as hydroxylation.[10] Researchers in the group focused on expanding the scope of the oxidation chemistry of cytochromes P450 to novel substrates including steroids,[11] small alkanes such as ethane,[12] and alkenes (via

epoxidation).[13] They also worked on engineering other properties of the enzymes, such as their tolerance to organic co-solvents.[14] In the last decade, researchers in the Arnold group took a chemomimetic approach and discovered the ability of engineered cytochromes P450 to perform carbene and nitrene transfer reactions, which proceed through iron-carbenoid and iron-nitrenoid intermediates.[15] This resulted in the introduction of a wide range of new-to-nature chemistry catalyzed by cytochromes P450. Some remarkable, recently developed carbene transfer reactions performed by engineered P450s include stereodivergent cyclopropanation,[16] bicyclobutane formation,[17] cyclopropenation of internal alkynes,[18] and enantioselective carbene C–H insertion[19] (Figure 1.2).

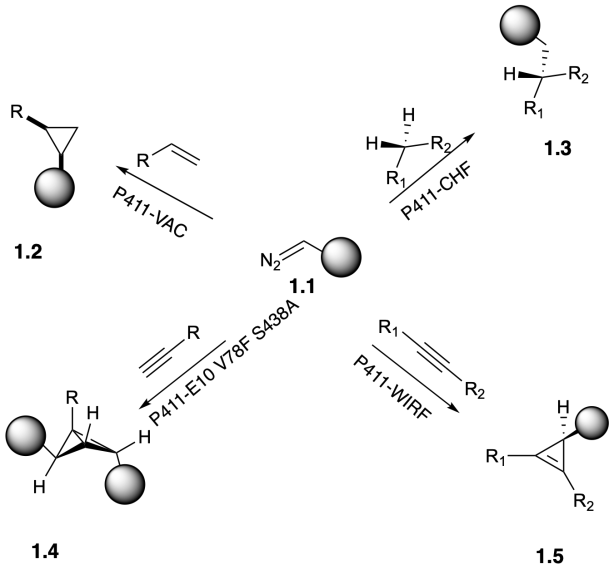


Figure 1.2: A selection of carbene-transfer reactions catalyzed by cytochromes P450 developed by the Arnold lab. These enzymes furnish cyclopropanes from alkenes (**1.2**), bicyclobutanes from alkynes (**1.4**), cyclopropenes from internal alkynes (**1.5**), and new C–C bonds from C–H bonds (**1.3**).

It is posited that these new-to-nature reactions proceed through an iron-carbenoid intermediate instead of the naturally occurring iron-oxo intermediate (Figure 1.3).

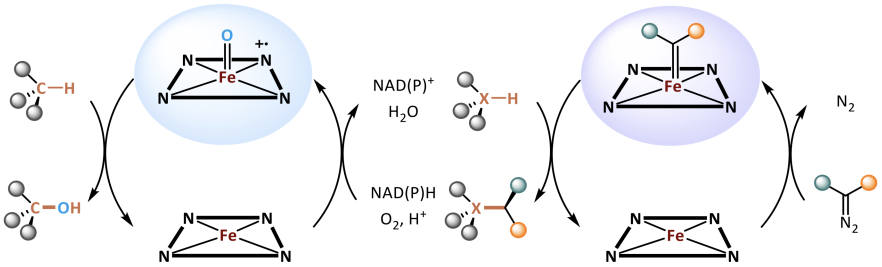


Figure 1.3: An iron-oxo intermediate enables the natural oxidation chemistry of heme proteins (left). The posited intermediate for heme-dependent carbene transfer reactions replaces the iron-oxo intermediate with an iron-carbenoid intermediate (right), enabling the introduction of carbon into organic reactants.

A key development that enabled these oxidative enzymes to perform carbene transfer reactions was the mutation of the axial ligand from cysteine to serine (Figure 1.4), resulting in a new enzyme, which was named cytochrome P411[20] (due to the change in the peak of the Soret band of CO-bound protein from 450 nm to 411 nm). This axial serine ligation has enabled a more diverse array of chemistries. In the natural oxygenation chemistry, the iron is in a low spin Fe(III), which transitions to the high spin state when the substrate binds. At this point, the reduction potential shifts where it can be endogenously reduced to Fe(II), at which point oxygen can bind. Among other effects, the serine axial ligand increases the reduction potential of the low spin state, allowing the heme cofactor to be reduced by endogenous NADPH and disfavoring oxygenation chemistry.

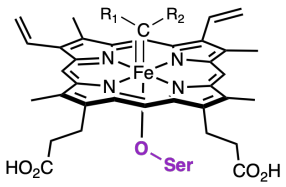


Figure 1.4: The porphyrin complex of a P411 variant, with the defining serine heme ligation. This axial ligand promotes carbene transfer reactions by increasing the reduction potential of the low spin state, among other effects.

The Arnold lab has also developed carbene transfer reactions using cytochrome *c* enzymes. These enzymes are found in prokaryotes and are a part of the electron transfer chain in the mitochondria of eukaryotes. Using cytochrome *c* enzymes derived from *Rhodothermus marinus* the Arnold lab has focused on carbene insertion reactions with less explored heteroatoms such as boron[21] and silicon[22] (Figure

1.5). The lab began working with engineered cytochrome *c* enzymes when a wild-type cytochrome *c* was found to catalyze an Si–H insertion reaction with high enantiomeric excess. This was intriguing because cytochrome *c* does not possess a known catalytic function in nature.[23] Furthermore, the cytochrome *c* from *Rhodothermus marinus* was chosen because of the thermophilic nature of this organism. This protein is quite thermostable, with a T_m of 106 ± 3°C.[24] A thermostable engineering starting point will be able to tolerate more destabilizing mutations than a more unstable parent protein.[25]

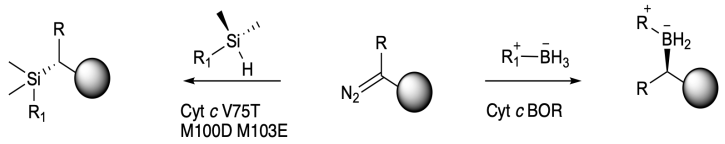


Figure 1.5: A selection of carbene transfer reactions catalyzed by engineered cytochrome *c* enzymes. These engineered enzymes form C–Si (**1.7**) and C–B (**1.8**) bonds, introducing nature to the ability to produce organosilanes and organoboranes.

The Fasan lab has developed myoglobin-based carbene transferases to perform — among other reactions — intramolecular cyclopropanation[26] and S–H insertion.[27] Meanwhile, the Hartwig lab has used artificial metalloenzymes with an iridium-porphyrin cofactor to perform C–H insertion reactions.[28]

In all, the Arnold lab, and others,[29] have been able to demonstrate that enzymes can act as carbene transferases. The lab maintains an interest in further developing and expanding the platform of carbene transferases because of their general synthetic utility.

1.3 Brief Background on Lactones

This thesis will focus on reactions with lactones. Lactones are cyclic esters. Six- and five-member lactones, known as δ-lactones and γ-lactones respectively, are the most common due to ring strain minimization.[30] Synthesis of lactones are rather well described. Aside from the simpler lactonization of hydroxy esters, a variety of methods have emerged. These include iodolactonization,[31] Yamaguchi lactonization,[32] Baeyer-Villiger oxidation,[33] and ring closing metathesis.[34] The extent that the synthesis of lactones has been developed highlights the utility of these compounds. In this work, we do not focus on the formation of lactones, rather on intermolecular reactions of lactones.

One attractive property of lactones, particularly γ- and δ-lactones, is their bioactivity.[35] Compounds such as damsin, ambrosin, and cnicin have shown antimicrobial properties. Other lactones have anti-inflammatory properties. Spironolactone and Eplerenone (marketed under Inspra®) are common antihypertensive pharmaceuticals. Lactone-bearing compounds have also shown efficacy against cancer. Camptothecin and its derivatives are treatments for breast, colon, and lung cancers, among others. Etoposide has been used in leukemia, lung cancer, and testicular cancer treatment. Costunolide and related compounds have been employed against brain cancers and leukemias. Outside of medicine, lactone-bearing compounds have been used in perfumes (δ-dodecalactone and δ-decalactone), flavoring agents for food (5-ethyl-γ-lactone and cis-3-methyl-4-octanolide), and insecticides (flupyradifurone). Figure 1.6 summarizes some applications of lactone bearing compounds.

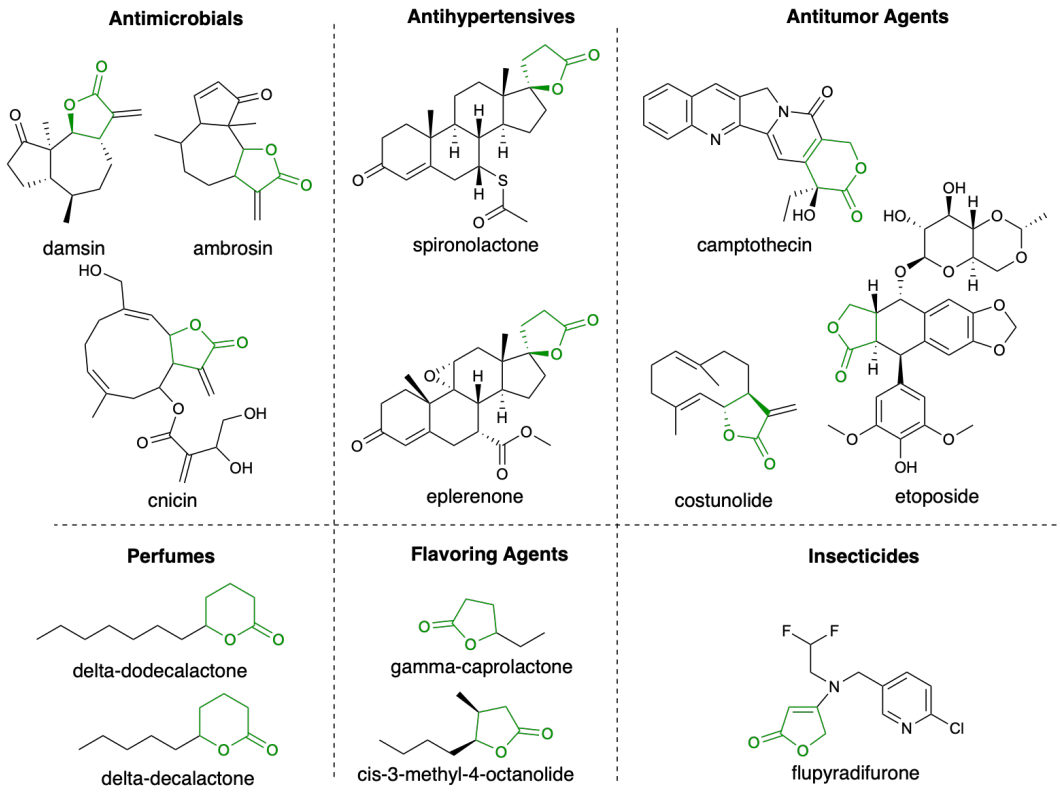


Figure 1.6: A selection of lactone bearing compounds and their applications. Most of these contain stereocenters, underscoring the importance of enantioselective methods.

Lactones are prochiral and most of the compounds in Figure 1.6 bear stereocenters. Thus, in the synthesis of lactone containing pharmaceuticals, it is critical to apply enantioselective methods to maximize potency and avoid unwanted side effects.[36]

In addition to these discovered lactone-bearing compounds, it is likely that there are plenty of uncharacterized compounds that have the potential to be medicinally or industrially important.

Chapter 2

LACTONE CARBENE B–H INSERTION

2.1 Project Inspiration

Synthesis of chiral organoboranes is of interest to synthetic and medicinal chemists. Organoboranes serve as useful intermediates in pharmaceutical syntheses in part because of the variety of methods developed for manipulation of these compounds. The most notable of these is Suzuki coupling.[37] In addition, organoboranes have shown bioactivity as antibacterial, antidiabetic, anticoagulant, antiviral, and even antitumor agents.[38] Transition metal catalyzed reactions for chiral organoborane formation through carbene insertion chemistry have been developed, primarily using rhodium catalysts.[39]

Discussed in Chapter 1, the Arnold lab has previously developed *Rhodothermus marinus* cytochrome *c* variants to insert methyl-EDA into B–H bonds. We wanted to expand the capability for laboratory-evolved enzymes to synthesize chiral organoboranes with lactones. In addition to the increased steric bulk, these cyclic carbenes have completely different geometric configuration in the active site, with a small dihedral angle d(Fe-C-C-O)($< 13°$). Subsequently, their electronic states are quite distinct from the states of linear carbenes[1]. Expanding the enzymatic scope to these lactones would be another tour-de-force of the capabilities of evolved enzymes. Furthermore, lactones are generally incompatible with the aforementioned rhodium catalysts due to β-hydride elimination,[40] necessitating new methods. We investigated five-, six-, and seven-membered lactones for this chemistry. As our model reaction, we set out to evolve an enzyme to catalyze the reaction between an N-heterocyclic carbene (NHC) protected borane (**2.1**) and the 5-membered lactone (**2.2**) with high yield and enantioselectivity (Figure 2.1).

[1]These conclusions were found via a computational density functinal theory (DFT) study conducted by collaborators on this work, Shuo-Qing Zhang and Dr. Xin Hong

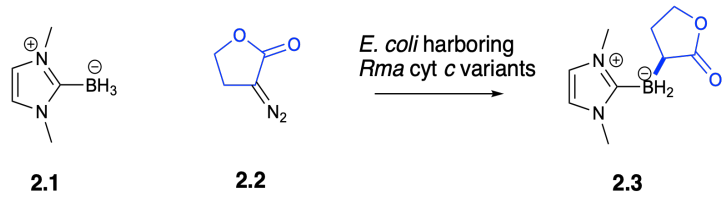


Figure 2.1: The model reaction for the expansion for carbene B–H insertion chemistry to cyclic carbene precursors. Reactions were set in M9-N, under anaerobic conditions, with 10 mM substrate loading.

2.2 Previously Conducted Enzyme Optimization

Other members of this project, Dr. Kai Chen, Dr. Xiongyi Huang, and Dr. S.B. Jennifer Kan began by screening a range of *Rma* cytochrome *c* variants, including those used to perform the previously developed borylation reactions in the Arnold lab, for the reaction between **2.1** and **2.2**. Fascinatingly, they found that wild-type cytochrome *c* could catalyze this reaction with 960 TTN, but with insufficient enantiomeric excess (34%). They then continued the search for variants that would improve the selectivity of this enzyme. They uncovered an M100D mutation, which was beneficial for both the B–H and Si–H insertion chemistry. While it improved the activity to 1280 TTN, it did not affect the selectivity. Site-directed mutagenesis at residue V75 revealed a V75R mutation which improved the total turnovers to 1640 TTN and the enantioselectivity to 87%. An M103V mutation increased the selectivity, though at the cost of the yield, resulting in a variant with 90% *ee*, and 1270 TTN. To further improve the enantioselectivity, they then considered conducting site-saturation mutagenesis at sites within 10 Å of the iron center in the wild-type *Rma* cyt *c* (Figure 2.2).[2]

[2]This figure is reproduced from the following published work:
Chen, K.; Huang, X.; Zhang, S-Q.; Zhou, A.Z.; Kan, S.B.J.; Hong, X.; Arnold, F.H. Engineered Cytochrome c-Catalyzed Lactone-Carbene B–H Insertion, *Synlett* **2019**, *30*, 378-382.

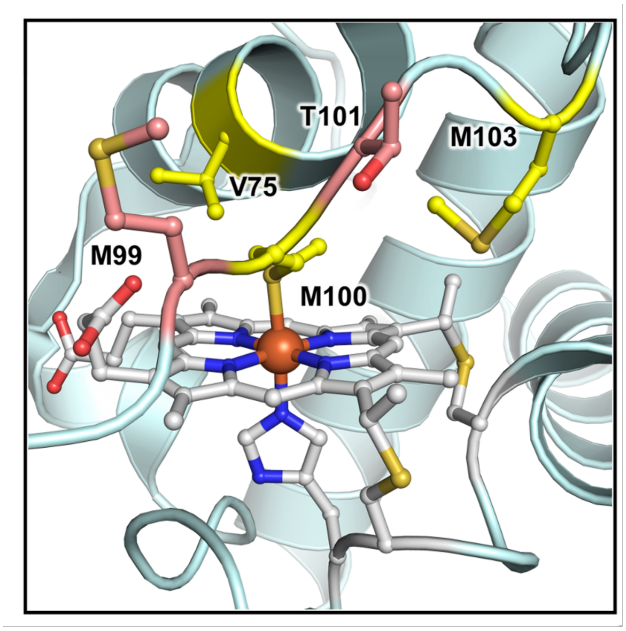


Figure 2.2: The active site of wild-type Rma cyt *c*, which was used to guide the choice of future residues to mutate. The selected residues are highlighted. (PDB: 3CP5)[24]

They formed a double site-saturation mutagenesis library via the 22-codon trick at positions M99 and T101,[41] since these sites reside on the front loop, which is suspected to be important in the overall active-site structure. Indeed, they identified a double mutant (M99Q T101Y), which improved the enantioselectivity even further (93% *ee*), though this did drop the activity (970 TTN). (Figure 2.3). This led to the final variant, **BOR$^{\mathbf{Lac}}$**, which had five mutations from the wild-type *Rma* cyt *c* (V75R M99Q M100D T101Y M103V).

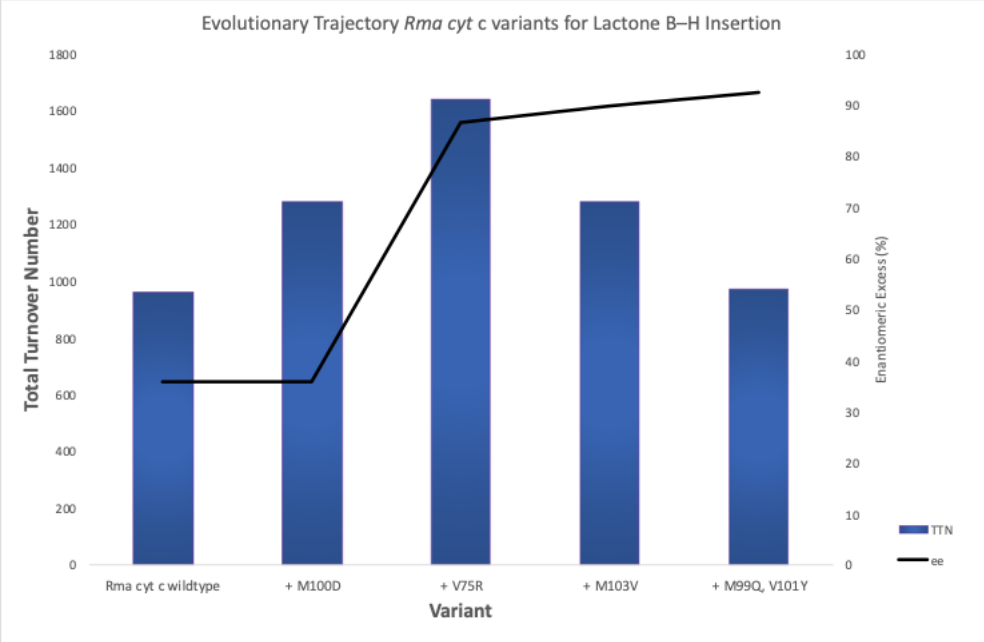


Figure 2.3: The TTN (bars) and enantiomeric excess (line) of variants that we identified in our lactone carbene B–H insertion lineage. Data collected by co-authors in the Arnold lab.

2.3 Substrate Scope

At this point, I joined the project. We examined the substrate scope of our final variant **BOR**$^{\mathbf{Lac}}$. First, we experimented with different substituents on the stabilizing NHC group (Figure 2.4) in reactions with **2.2**. We found that variant **BOR**$^{\mathbf{Lac}}$ could achieve good yield and selectivity with electron withdrawing (**2.4b**), alkyl (**2.4a**, **2.4c**, **2.4d**), and bulky (**2.4d**) substituents on the NHC group.

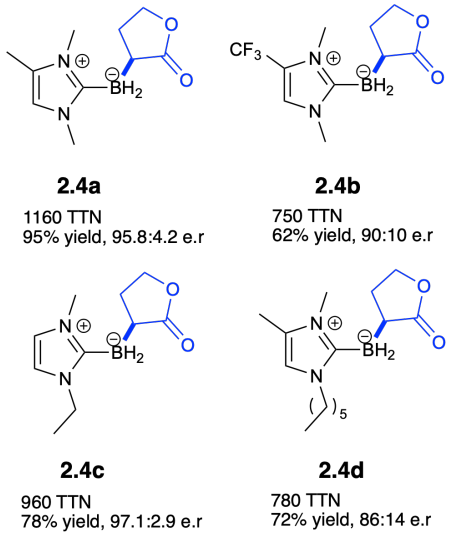


Figure 2.4: The scope of variant **BOR^{Lac}** with 5-membered lactones and NHC-borane derivatives.

We then wanted to challenge our enzyme with 6-member lactones. Despite the increased steric bulk, we found that **BOR^{Lac}** could accept this substrate with high yield (up to 98%) and selectivity (up to 92% *ee*) (Figure 2.5). As with the 5-membered lactone, the enzyme could tolerate electron withdrawing (**2.5c**) and ethyl (**2.5b**) substituents.

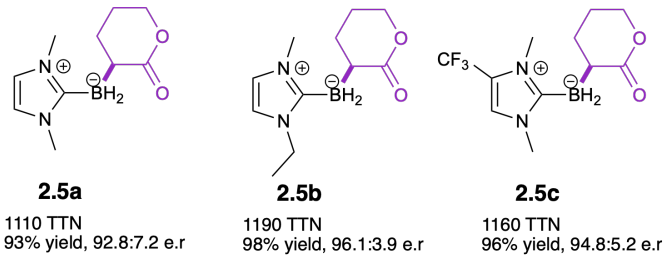
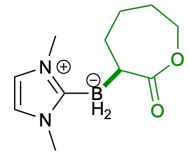


Figure 2.5: The scope of variant **BOR^{Lac}** with 6-membered lactones and NHC-borane derivatives.

To further test variant **BOR^{Lac}**, we assessed 7-membered lactone derivatives, which, despite bearing a ring being just one carbon larger than the 6-membered lactone, could only yield a product with less than 50 TTN (Figure 2.6). However, we speculate that with further evolution, we would be able to craft an active site with a different electronic environment, which would allow for efficient formation of the 7-membered lactone product, **2.6**.



2.6

<50 TTN
<10% yield, n.d e.r

Figure 2.6: The efficacy of variant **BOR**$^{\mathbf{Lac}}$ with 7-membered lactones is limited.

2.4 Conclusion

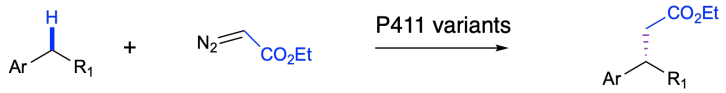
We developed an enzymatic variant, *Rma* cyt *c* **BOR**$^{\mathbf{Lac}}$, which is capable of inserting lactones into NHC stabilized B–H bonds. **BOR**$^{\mathbf{Lac}}$ is compatible with 5- and 6-membered lactones and NHC stabilized boranes with electron-donating and electron-withdrawing substituents. Overall, this project is an expansion of previous work in the Arnold lab to structurally more complex substrates and further demonstrates the applicability of evolved enzymes in performing challenging carbene transfer reactions enantioselectively. To the best of our knowledge, this method is the first demonstration of an intermolecular carbene transfer reaction for the synthesis of chiral organoboranes bearing bioactive lactone motifs.[42] Furthermore, we demonstrate that hemeproteins can circumvent problematic side pathways such as β-hydride migration. This provides a basis for studies on future enzymatic lactone insertion reactions.

Chapter 3

LACTONE CARBENE C–H INSERTION

3.1 Conception and Basis for Reactivity

In 2019, the Arnold lab made a breakthrough in repurposing cytochrome P411s by engineering an enzyme that could perform enantioselective carbene C–H insertion, as discussed in Chapter 1.2. In this reaction, the Arnold lab used an engineered cytochrome P411 to insert a diazoacetate, **3.2**, into C–H bonds with high enantioselectivity and total turnovers (Figure 3.1).



3.1 **3.2** **3.3**

Figure 3.1: Previously developed carbene C–H insertion reaction by the Arnold lab using an evolved cytochrome P450 variant. This is the first instance of carbene C–H insertion using an earth-abundant iron center. The enzyme was optimized for linear carbene precursors.

To further expand this chemistry, we set out to train cytochrome P411s to selectively catalyze carbene C–H insertion using cyclic carbene precursors such as lactones. The lactone motif adds a stereocenter to the product at the α-position, opening up the possibility of creating products with contiguous α- and β-chiral centers. This makes stereoselective synthesis an interesting and challenging endeavor. Furthermore, the lactone is a privileged motif in pharmaceuticals, making the ability to manipulate it potentially relevant to medicinal chemists.[43]

There are transition-metal catalysts that can perform similar carbene insertion chemistries,[44,45] such as dirhodium[46] and iridium[47] catalysts. However, in addition to being less sustainable and sometimes limited in selectivity, transition metal catalysts have shown few demonstrations of C–H insertion of acceptor-only carbenes, particularly those bearing α-alkyl substituents due to competition with β-hydride migration.[40] We found success in circumventing this β-hydride migration pathway in our previously completed lactone B–H insertion project. Thus, we hypothesized that using a cytochrome P450 variant we would be able to avoid side reactions

and form chiral products, including those with contiguous chiral centers, with high stereoselectivity (Figure 2.2)

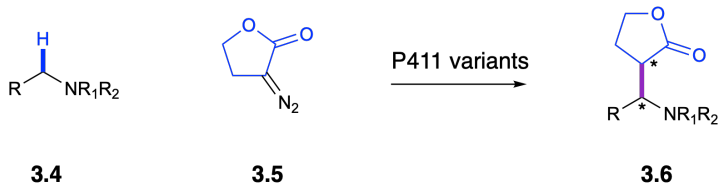


Figure 3.2: A generalized scheme of the developed enzymatic reactions. The proposed enzyme would be able to exert stereo control on two contiguous chiral centers while circumventing side with β-hydride migration pathways.

3.2 Identifying the Best Variant

We began our studies by screening a diverse library of cytochrome P411 and cytochrome *c* variants for activity in performing α–amino C–H insertion using 4,*N*,*N* trimethylaniline (**3.6**) and a γ-lactone carbene precursor (**3.5**) (Figure 3.3). Reactions were conducted under anaerobic conditions in *E. coli* expressing the P411 variants suspended in M9-N. Substrates were loaded to a final concentration of 10 mM.

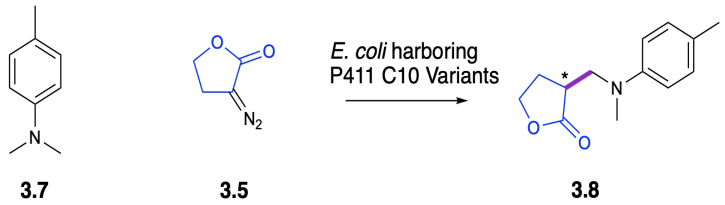


Figure 3.3: The initial screening reaction identified cytochrome P411-C10 as the most promising variant to begin the evolution, though the initial activity and selectivity were quite modest. TTN was characterized by protein concentration and reverse phase HPLC, while *ee* was determined by chiral HPLC.

We found that cytochrome P411-C10 — a robust and versatile variant for carbene C–H insertion reactions was the best performing variant, though it could only complete the α-amino C–H insertion reaction with rather low TTN and enantiomeric excess (115 TTN, 47% *ee*). The resulting product from this reaction bears a β-amino lactone motif, which is the core of pharmaceutical targets such as sesquiterpene lactone amino derivatives.

3.3 Evolution of a Biocatalytic Platform

Launching off this starting point, we conducted site-saturation mutagenesis of the starting P411 variant (P411-C10) with the model reaction. Harboring the enzymes in *E. coli*, we targeted residues known to be proximal to the active site[48] using the 22-codon trick to generate site-saturation libraries[41](Figure 3.4)[1]

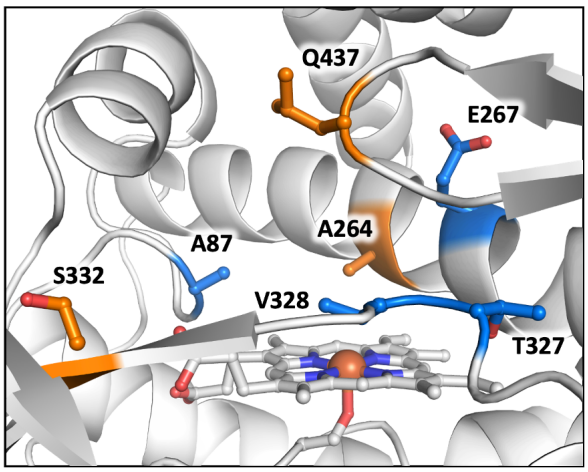


Figure 3.4: Active-site residues that were targeted by site-saturation mutagenesis. This was based off a P411-E10 variant (pdb: 5UCW)[48] with homology to the starting P411 C10 variant. Residues highlighted in blue indicate residues that had more of an effect on the *ee*, while residues highlighted in orange had a greater impact on TTN (See Figure 3.5).

We found that the first two mutations, both residing on active-site loops (T327V and Q437L, respectively) improved the total turnover number 9-fold, and an S332A mutation increased the activity to over 1200 TTN (Figure 3.5). However, the enantioselectivity did not increase and it even decreased slightly from 47% to approximately 40% *ee*. While enantioselectivity is usually correlated to activity due to the optimization for the active site for a specific substrate or class of substrates, we hypothesized that this might not be the case for our model substrate, **3.7** because it may bind with different orientations due to its symmetric nature.

[1]This figure is reproduced from the following published work:
Zhou, A.Z.; Chen, K.; Arnold, F.H. Enzymatic Lactone-Carbene C–H Insertion to Build Contiguous Chiral Centers, *ACS Catal.* **2020**, *10*, 5398-5398

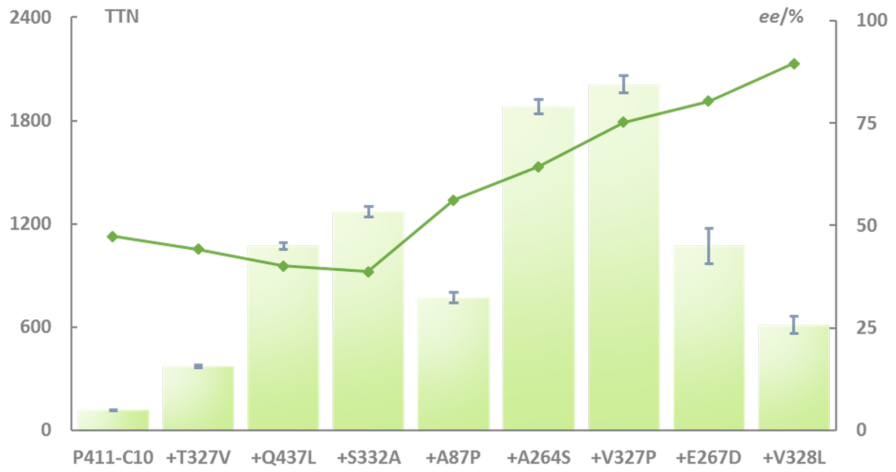


Figure 3.5: The evolutionary lineage for the reaction between 4,*N*,*N*-trimethylaniline (**3.7**) and a γ-lactone carbene precursor (**3.8**). Total turnover number is shown by bars, and the enantioselectivity is shown by the lines. Reactions were conducted in quadruplicate and monitored by reverse phase HPLC and chiral HPLC analysis. **L7** catalyzed the reaction with up to 2010 TTN, and **L9** could catalyze the reaction with up to 90% *ee*.

To evolve for selectivity, we decided to focus on residues that were previously shown to increase enantioselectivity and selected variants based on enantioselectivity rather than total turnover numbers. Instead of conducting TTN-based screens, we used chiral HPLC to test for variants that improved selectivity only. After identifying these variants, we then tested the activity in a validation experiment to ensure that the mutation was not significantly inactivating. Once we shifted to this strategy, we were able to increase the enantioselectivity, though sometimes this was at odds with the TTN. Mutation A87P increased the enantioselectivity from 39% to 56% *ee*, though it dropped the total turnover number. Mutating A264 to a serine resulted in a significant boost in both enantioselectivity and TTN. When re-targeting site 327, a mutation from valine to proline (V327) resulted in a modest boost in selectivity (from 64.5 to 75% *ee*) while maintaining high activity. Interestingly, a proline mutation was not beneficial in the initial screening at site 327 (which resulted in variant **L2**), demonstrating the important effect of epistasis in directed evolution.[49] Mutations E264D and V328L resulted in the final variant (**L9**), which had 90% *ee*, but reduced activity (590 TTN). Fascinatingly, from this step in the evolution, we identified another variant (**L10**), which resulted from a V328R mutation and showed the opposite enantio-preference (-68% *ee*) compared to other enzymes in the lineage. We hypothesize that this might arise due to different carbene orientations on the

heme center, as a leucine residue might interact more strongly with the alkyl groups of the lactone, while an arginine residue might interact with the ester group. While the precise mechanism of this reaction is not known, the promising result from the **L10** variant indicates the tunability of the enzyme for synthesis of diverse products.

3.4 Substrate Scope

To continue the study, we assessed the ability of our enzymes to accept diverse substrates beyond **3.7**. We focused on variants **L6–L10** since they showed the highest enantiomeric excess or TTN. We began by evaluating other primary C–H bonds (Figure 3.6a). We tested a range of substituents on the aromatic ring including electron donating (**3.9a**, **3.9c**, **3.9d**) and electron withdrawing (**3.9b**) groups. We also tested ortho, meta, and para positions. Consistently, **L9** and **L10** gave the opposite stereo preference. In addition, **L9** consistently was more selective than **L6** and **L7**, though this was not a surprise because **L9** was the most selective variant in reactions with the model substrate. The performance of the enzymes showed some degree of substrate dependency. For instance, **L10** had higher activity than **L9** for **3.9e**, even though this was not the case for the model substrate. In addition, **L6** achieved the highest TTN out any of these reactions (**3.9a**, 2040 TTN), but also frequently performed worse than other enzymes in selectivity (**3.9c**, **3.9e**).

We then noticed that for **3.10**, **L9** and **L10** demonstrated a shifting regiospecificity (Figure 3.6b). **L9** heavily favored the primary insertion product (99:1 r:r), but **L10** only moderately favored the primary C–H insertion product (71:29 r:r). Not only did this excite us because it once again showed the capability for our enzymes to be tuned for diverse selectivities, but it also provided hope that our existing lineage could catalyze reactions of secondary C–H bonds. This would enable the construction of contiguous chiral centers. Moreover, we hypothesize that with further evolution, we could favor the synthesis of **3.10b** over **3.10a.**

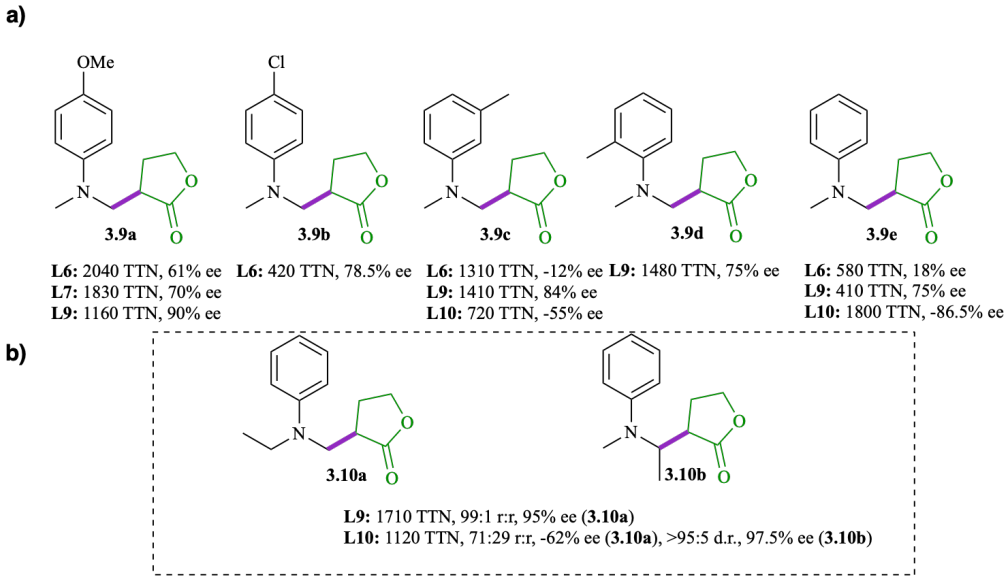
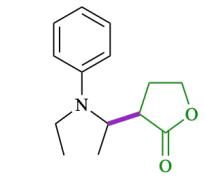
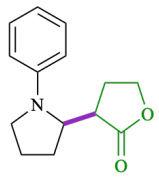


Figure 3.6: **(a)** The activity and selectivity of enzymatic variants for lactone carbene C–H insertion into primary C–H bonds. **(b)** Substrates with both primary and secondary C–H bonds revealed different regioselectivity for **L9** vs **L10**, opening up the possibility for performing carbene insertion into secondary C–H bonds.

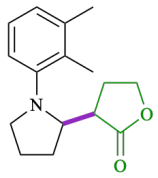
Exploring the potential of the enzymes for insertion into secondary C–H bonds (Figure 3.7), we evaluated pyrrolidine, azetidine, and dialkyl-aniline derivatives. Despite training the enzymes only on primary C–H bonds, these enzymes had very favorable properties with regards to both activity (up to 4,000 TTN, **3.11c/L9**) and enantioselectivity (up to 99% *ee*, **3.11c/L9**) in these reactions. This inspires confidence that we did not over-optimize our enzyme to our model substrate, **3.7**. In general, the diastereoselectivity and enantioselectivity were good to high. In a fascinating case, **3.11a**, the same major diastereomer was formed with both enzymes (**L9**: 94:6 d:r, **L10**: 99:1 d:r), but the opposite enantiomer was formed (**L9**: 81% *ee*, **L10**: -94.5% *ee*). This trend was observed for other compounds including **3.11b**. In addition, we found that these enzymes were not too adversely affected by electron-withdrawing (**3.11e**) or electron-donating (**3.11d**) substituents, as selectivity and activity remained acceptably high. To test the limits of our enzymes, we tried structurally and chemically distinct substrates such as **3.11f**. With a higher C–H bond dissociation energy and a unique structure, we expected our enzymes to struggle to obtain favorable selectivities and activities. Indeed, while diastereoselectivity was high (95:5 d:r), the enantiomeric excess (13%) and activity (140 TTN) were low.



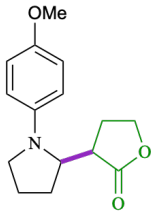
3.11a
L6: 600 TTN, 78:22 d.r., 41% ee/-30% ee
L9: 850 TTN, 94:6 d.r., 81% ee/-15% ee
L10: 740 TTN, 99:1 d:r., -94.5% ee/-19% ee



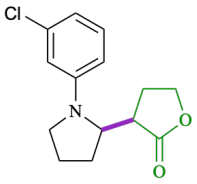
3.11b
L6: 2550 TTN, 52:48 d.r., 94.5% ee/-24% ee
L9: 1430 TTN, 88:12 d.r., 86.5% ee/-73% ee
L10: 1490 TTN, 96:4 d:r., -72% ee/-30% ee



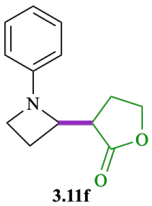
3.11c
L7: 2400 TTN, 22:78 d.r., 96.5% ee/86% ee
L8: 2730 TTN, 18:82 d.r., 96.5% ee/91% ee
L9: 4000 TTN, 38:62 d:r., 99% ee/85% ee



3.11d
L6: 960 TTN, 52:48 d.r., 93% ee/-26% ee
L7: 1260 TTN, 58:42 d.r., 98% ee/-52% ee
L9: 210 TTN, 80:20 d:r., 98% ee/78.5% ee



3.11e
L6: 960 TTN, 34:66 d.r., 95% ee/-27% ee
L8: 860 TTN, 41:59 d.r., 96% ee/7% ee
L9: 1460 TTN, 88:12 d:r., 91% ee/14% ee



3.11f
L10: 140 TTN, 95:5 d:r., 13% ee

Figure 3.7: The activity and selectivity (diastereo- and enantio-) of enzymatic variants for lactone carbene C–H insertion into secondary C–H bonds.

3.5 Potential for N–H Insertion

While evaluating the properties of our enzymes for **3.10**, we noticed the formation of a side product, **3.10c** (Figure 3.8). We hypothesized that this might arise from an elimination reaction followed by N–H insertion. If this were the case, it would imply that our engineered enzymes might be able to perform carbene N–H insertion, which will be discussed in the following chapter. We are unsure why **3.10c** is only observed in reactions of this substrate. We speculate it might be caused by the specific configuration of **3.10a**, which might position the eliminated hydrogen in close proximity of a basic residue in the active site.

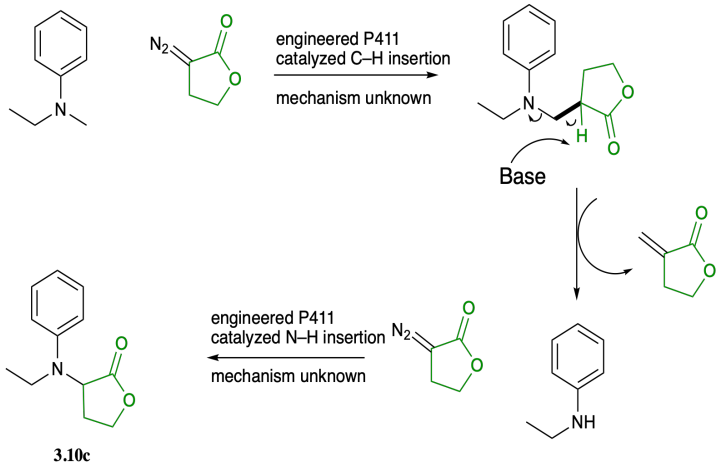
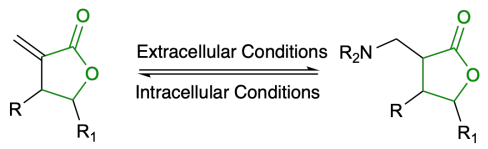


Figure 3.8: A potential mechanism to justify formation of **3.10c**. Here, product from the lactone carbene C–H insertion reaction undergoes elimination. The resulting amine can react to give product **3.10c**.

3.6 Potential Applications in Pharmacuetical Synthesis

A promising potential application of this chemistry is in the construction of amino derivatives of sesquiterpene lactones. Sesquiterpene lactones often possess antitumor or antimicrobial activities and are classified by possessing an α-methylene-γ-lactone motif, which allows them to bind to biological nucleophiles such as thiols and shut down deleterious cells. However, these compounds are often constrained by off-target effects and limited solubility. Researchers have found that masking the α-methylene group with an amine alleviates these problems by improving solubility and preventing off-target electrophilic activity (Figure 3.9). The amino derivatives often are specific for deleterious cells, and once inside these cells, they are converted back to the original bioactive sesquiterpene lactone.



3.12a **3.12b**

Figure 3.9: The conversion of sesquiterpene lactones (**3.12a**) to their amino derivatives (**3.12b**) occurs in the bloodstream, enabling more favorable pharmacological properties. Once inside cells, the compound is reverted to its original bioactive form.

Thus, sesquiterpene lactone amino derivatives are highly valuable prodrugs with incredible potential in medicinal chemistry.[50] Our enzymatic platform could be useful in constructing sesquiterpene lactone amino derivatives sustainably and cost effectively, especially those with aryl substituents on the amine. Figure 3.10 shows some identified targets with the chiral β-amino lactone motif. We imagine that our enzymatic platform could be used to synthesize related molecules. The versatility of our enzymes and broad substrate scope coupled with the tunability of regioselectivity and enantioselectivity could be used to meet the needs of medicinal chemists targeting these amino derivatives of sesquiterpene lactones and other pharmaceutical targets.

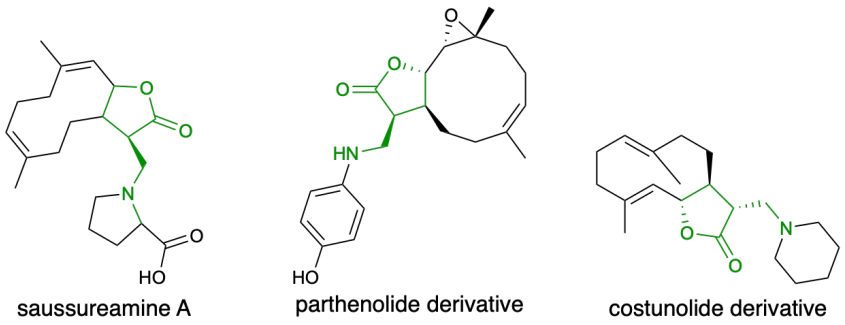


Figure 3.10: Several known sesquiterpene lactone amino derivative prodrugs, with the chiral β-amino lactone motif highlighted.

3.7 Conclusion

Our enzymatic platform based on a series of engineered cytochrome P411 enzymes can catalyze lactone carbene C–H insertion to form new C–C bonds with high efficiency — up to 4000 TTN — and good to excellent enantioselectivity, -94.5% to 99%. Our enzyme circumvented a β-hydride migration pathway that has plagued other catalysts attempting to perform similar chemistry. To our knowledge, this is the first enzymatic synthesis of a chiral β-amino lactone motif. This motif is extremely relevant to synthetic and medicinal chemistry, and our platform could be especially useful for the construction of sesquiterpene lactone amino derivatives, among other potential useful products. We were able to construct products selectively with two contiguous chiral centers, and we found that a single mutation can completely invert both centers. The same single mutation also inverted single chiral centers and was able to affect the enzyme's preference of primary versus secondary C–H bonds. Furthermore, our enzymes have been shown to be quite accepting of diverse substrates. The scope of the enzymes and their tunability enable them to meet the needs of chemists interested in preparing diverse chiral β-amino lactones.

Chapter 4

LACTONE CARBENE N–H INSERTION

Having evolved a robust platform for carbene C–H insertion (Chapter 3), we wanted to expand this chemistry to N–H insertion. We are interested in N–H insertion chemistry for several reasons. Amines, owing primarily to their nucleophilicity are active in biological systems and are found in a variety of pharmaceuticals;[51,52] for this reason, reactions to form C–N bonds are of interest for synthetic and medicinal chemists.[53] However, both enzymatic and transition-metal enabled reactions to enantioselectively perform N–H insertion reactions are limited.

Transition metal catalyzed methods to do such insertion reactions typically use both a transition metal as well as a chiral proton transfer catalyst (PTC) (usually a small molecule) to bias the product to a certain stereo fate. After the metal carbenoid is formed, which is akin to the metal carbenoid discussed in Chapter 3, the amine will act as a nucleophile and attack it, forming an ylide. Protonation is then guided by the PTC to form a stereoselective product (Figure 4.1).[54]

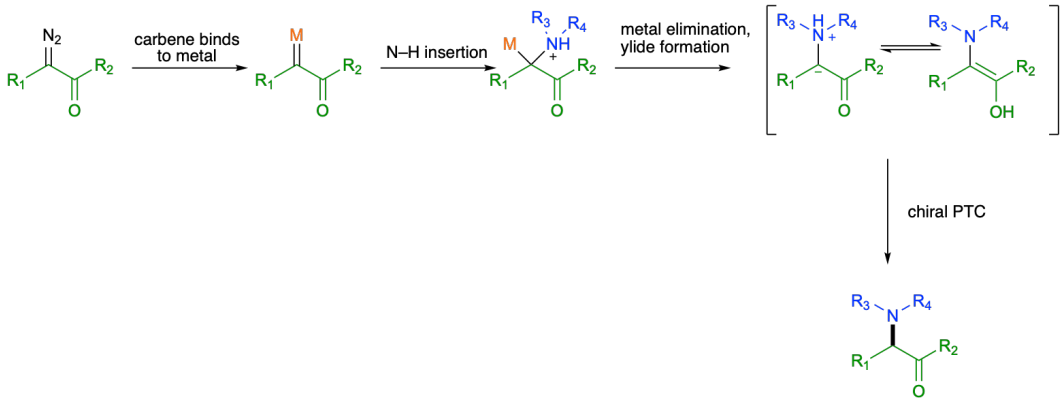


Figure 4.1: The typical strategy to perform enantioselective N–H insertion in transition metal catalyzed reactions. A chiral PTC helps ensure the enantioselective fate. After the carbene binds to the transition metal, the insertion reaction occurs. At this point, the transition metal is eliminated, forming an ylide intermediate. To form an enantiopure product, the protonation must be controlled, which the chiral PTC accomplishes.

Studies on similar reactions within a heme protein have been conducted, and they demonstrated that the enzymatic system behaves similarly.[55] While the insertion reaction is facile — even a free heme complex can catalyze it — the enantioselectivity remains difficult to control. Enzymes reside in protic environments (water) and have protic amino acid residues that surround the iron porphyrin center. These make it difficult to control the proton transfer. Therefore, to achieve these enantioselective N–H insertion reactions, the active site must be carefully tuned to ensure that the proton transfer, either from water or from the amino acids, occurs at the correct orientation and correct timing (Figure 4.2). Indeed, this is quite challenging, and there is only one reported instance of an enzyme achieving enantioselective N–H insertion.[56] In this work, the Fasan group engineered a myoglobin variant to perform this chemistry. However, this was limited in its enantioselectivity (only up to 82%) and restriction to linear carbenes and primary anilines.

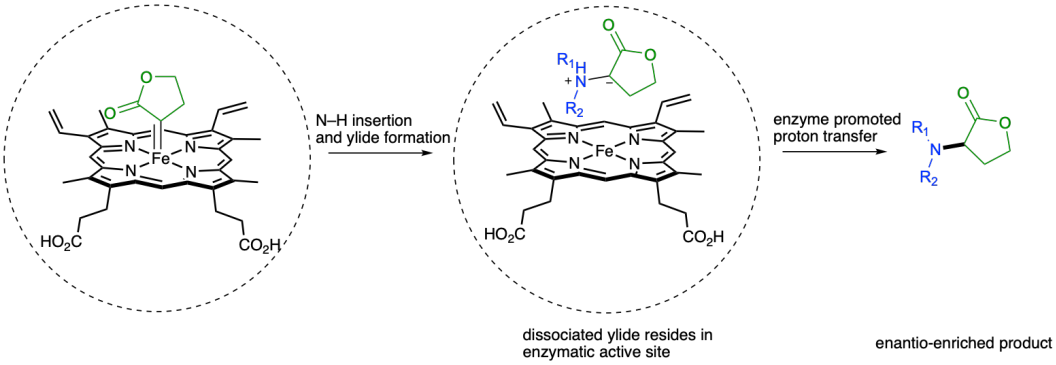


Figure 4.2: The posited pathway for N–H insertion within the enzymatic active site. The dotted circle represents the enzymatic active site. Amino acid residues create an environment than can precisely perform proton transfer to the ylide. The nature of this proton transfer will be further discussed.

Inspired by the observation of trace amounts of an N–H insertion product during the lactone carbene C–H insertion process, we decided to use a lactone carbene precursor for this study. The lactone is also fascinating because of its bioactivity and because traditional transition metal catalysts have been largely unsuccessful with them, in part due to the control needed to prevent unwanted β-H elimination.

4.1 Identifying the Best Variant

To start our study, we screened a reaction of **4.1** and **4.2** catalyzed by hemoprotein variants in whole cell *E. coli* (Figure 4.3). This screening was carried out by Dr. Zhen Liu.

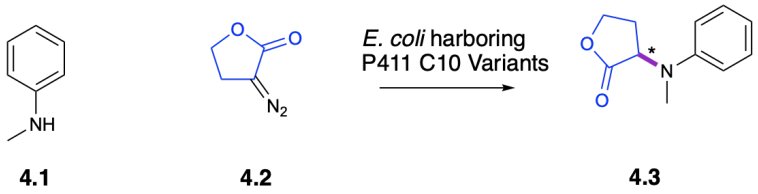


Figure 4.3: The model reaction for carbene N–H insertion reaction. The product, **4.3**, is bioactive. Whole cells expressing hemeprotein variants were suspended in M9-N. Reactions were set in a 96-well plate under anaerobic conditions (10 mM amine, 10 mM carbene precursor, 25 mM D-glucose) and shaken overnight.

Dr. Zhen Liu screened more than 40 enzymes that were previously optimized for carbene and nitrene reactions, including cytochrome P411 and *Rma* cytochrome *c* variants. He assessed enzymes previously evolved for lactone insertion reactions: P411-G8S, P411-L5, P411-L7, and Cyt *c* BOR^{Lac}. This showed that **L7**, an enzyme

from the lactone carbene C–H insertion lineage (Chapter 3) could catalyze this reaction with 81% yield and 94% *ee* (Figure 4.4). Interestingly, there is another instance of divergent enantioselectivity, as **L5** (well C9) and **L7** (well C10) form opposite enantiomers, despite being only two mutations apart. Indeed, the highest performance from enzymes derived from the carbene insertion lineage. Most other enzymes were either limited in yield, selectivity, or both.

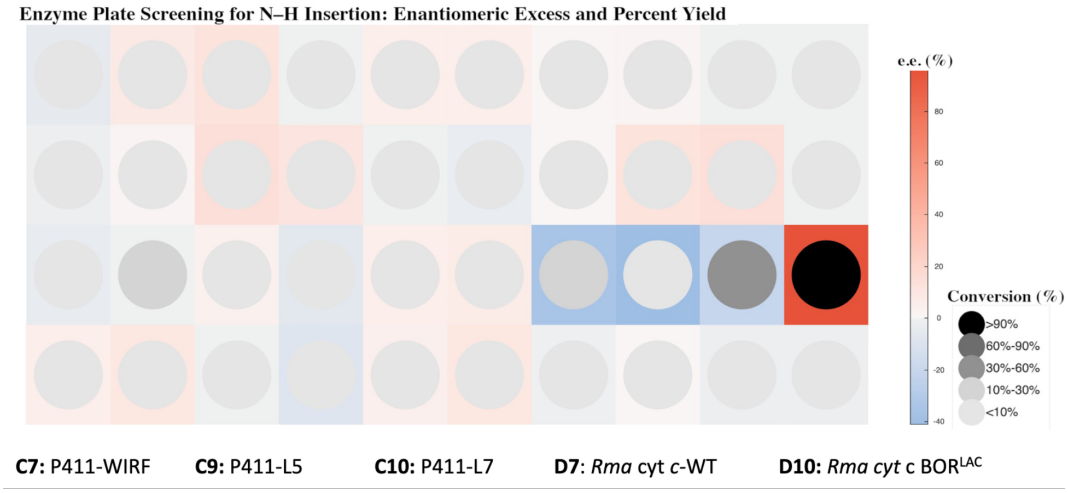


Figure 4.4: A heatmap demonstrating the results of the enzyme screening with some notable wells labeled with the enzymes that reside within. P411-L7 is in well C10 and had the best properties: 81% yield and 94% *ee*. Figure made using Holoviews and Python.

Seeing more evidence of the promise of the C–H insertion lineage in catalyzing this N–H insertion reaction, we screened the entire lineage (Figure 4.5). We hypothesized that to further improve the enantioselectivity, it would be beneficial to improve the stability of the enzyme to prevent the release of free heme as the enzyme degrades. This is important since free heme forms racemic N–H insertion product. To that end, we added back the previously truncated FAD domain to our best variant, (**L7**) to give variant **L7-FL**. Indeed, we found that adding back this domain further improved the selectivity and activity of the enzyme. We hypothesize that the addition of this domain makes the enzyme more thermodynamically favorable because — having been optimized by nature — it has minimal disfavored interactions.

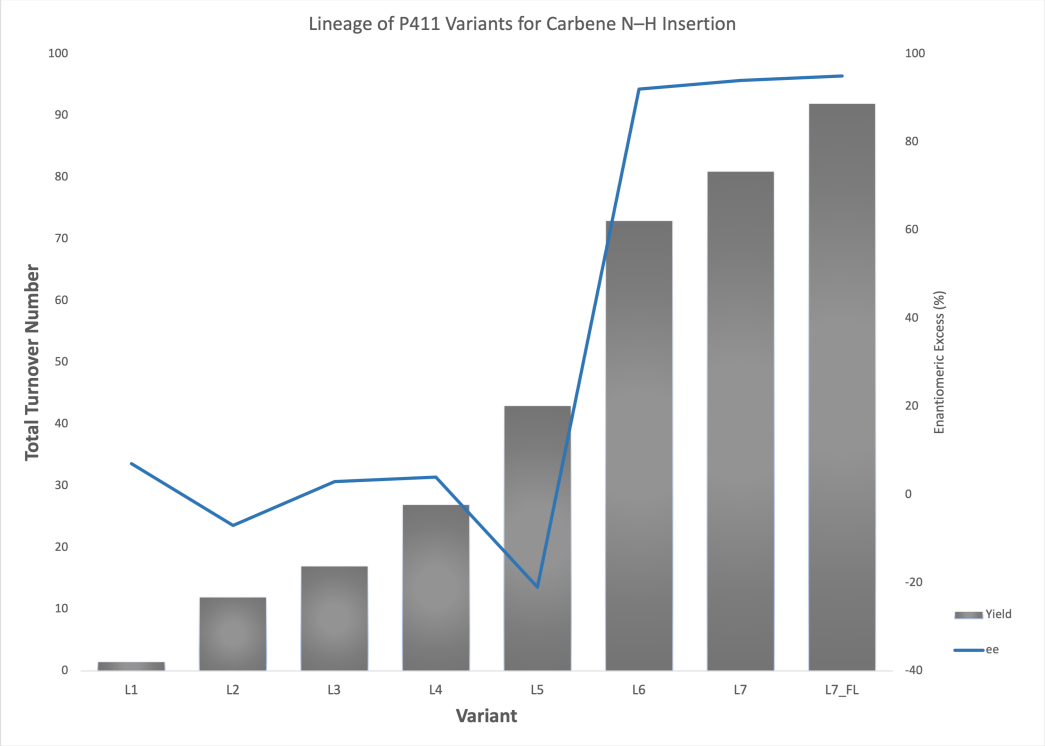


Figure 4.5: The lineage of **L1**-**L7** and their yields (bars) and selectivities (line) in forming **4.3**. **L1**-**L7** have the FAD domain truncated. **L7-FL** has the FAD domain added back to the enzyme, resulting in a slight increase to the yield and selectivity. It appears that the mutation from **L5** to **L6** (A264S) was crucial in improving the enantioselectivity.

It appears that the A264S mutation that is added between **L5** and **L6** is crucial in increasing the yield and selectivity. Based on the crystal structure of P411-E10 (as shown in Chapter 3), this residue resides directly above the active site and is the closest in proximity to the reaction. It is not surprising that this might be promising in promoting this carbene transfer reaction and in guiding the stereoselectivity during the proton transfer step. In an experiment conducted by Dr. Zhen Liu, it was confirmed that the serine mutation was able to much better facilitate high yield and selectivity compared to other amino acids (including those with similar properties such as threonine and cysteine).

4.2 Substrate Scope

Next, we assessed the substrate scope. We used the best enzyme in the initial screening, **L7-FL** and tested a range of substrates including secondary amines, primary amines, and aliphatic amines (Figure 4.6).

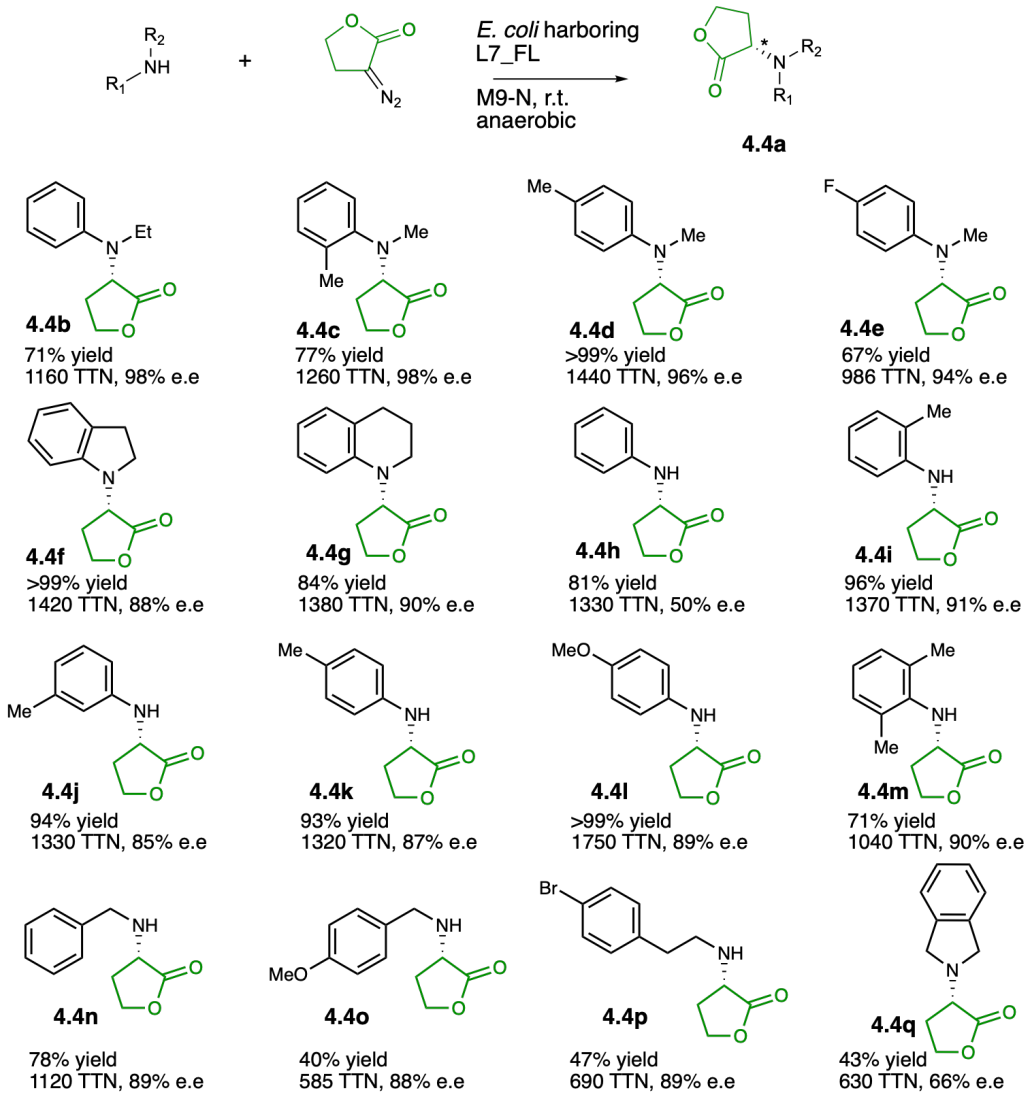


Figure 4.6: The substrate scope of the N–H insertion reaction using **L7-FL**. Reactions were set at 10 mM substrate loading using whole-cell *E. coli*. TTN was determined by protein concentration and HPLC data.

We assessed the yield and selectivity in formation of generalized product **4.4a**. Both secondary anilines (**4.4b**-**4.4g**, **4.4q**), and primary anilines (**4.4h**-**4.4p**) showed high yield and selectivity across the board. Yields were as high as 99% (**4.4d**, **4.4f**), and selectivity of up to 98% *ee* was also observed (**4.4b**, **4.4c**). Substrates with substituents on the ring including fluoro (**4.4e**) and methyl (**4.4c** and **4.4d**) were well tolerated, as were heterocycles (**4.4f**, **4.4g**) despite their bulk. In addition, an ethyl group on the amine (**4.4b**) does not affect the efficacy of the enzyme. Indeed, methyl substituents on the ring, did not negatively affect the selectivity or yield, no matter the position. For instance, **4.4m** is sterically hindered and bears two ortho-methyl

substituents, yet it achieves 1040 TTN and 90% *ee*. Interestingly, the unsubstituted aniline (**4.4h**) had the lowest enantiomeric excess (50% *ee*).

Finally, we wanted to test our enzyme with substrates with amines attached to sp^3 hybridized carbons. Aliphatic amines make enantioselective N–H insertion challenging due to the increased basicity of the substrate, and reactions of this nature were recently demonstrated by copper catalysts in 2019.[57] Our enzymes could catalyze these reactions with both primary (**4.4n-p**) and secondary (**4.4q**) aliphatic amines with high ee and yield. All of the primary amines had selectivity above 85% *ee*, though the secondary amine only displayed 66% *ee*. With this promising start, we predict that with further evolution, we may be able to expand this platform beyond homobenzylic amines and perhaps tolerate aliphatic amines without proximal aromatic systems, though this remains a challenge for biocatalysts.

Curious about the properties of other enzymes in the lineage bearing the A264S, particularly **L6** and **L7**, we then compared the properties of **L6**, **L7**, and **L7-FL** for a subset of substrates. We wanted to assess whether the superiority of **L7-FL** identified in the model reaction held true for other substrates (Figure 4.7).

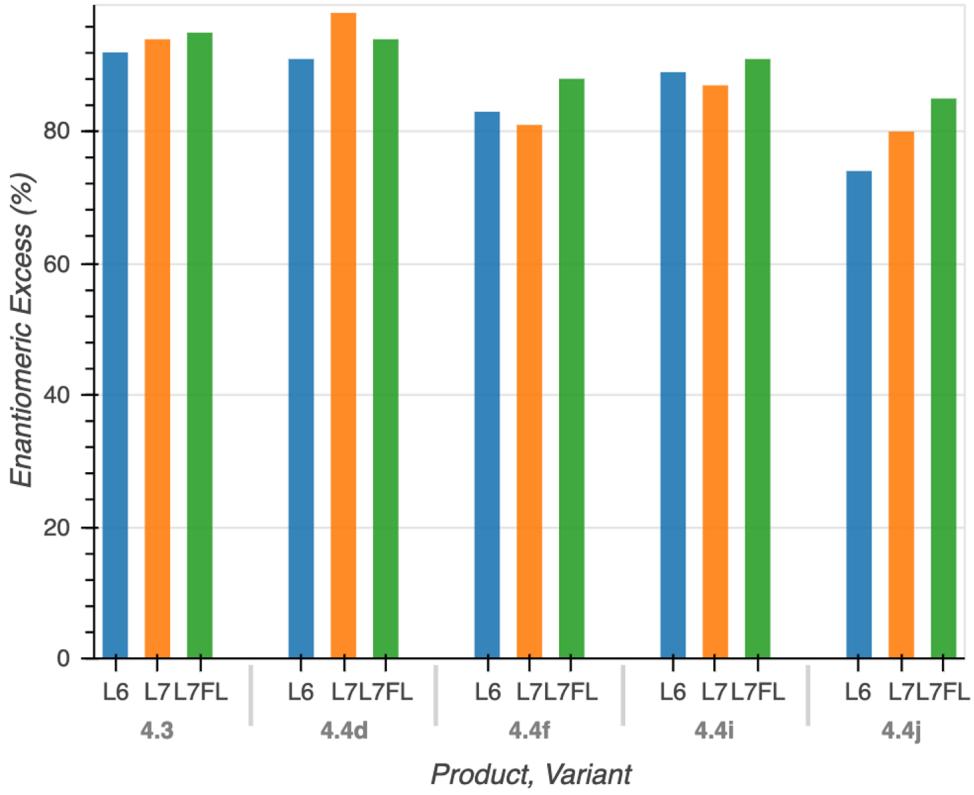


Figure 4.7: A bar graph comparing **L6** (blue), **L7** (orange), and **L7-FL** (green) for five substrates (labels in Figure 4.6). **L7-FL** shows highest *ee* for most substrates tested, including the model substrate. Figure made using Holoviews with a Bokeh backend.

Indeed, **L7-FL** performed with higher enantiomeric excess than **L7** and **L6** for all substrates tested except **4.4d**. We are unsure why the addition of the FAD domain decreases the enantioselectivity of **4.4d** Nevertheless, **L7-FL** appears to have the most favorable overall selectivity profile, so we decided to focus on it and test its limits.

4.3 Large-Scale Reactions

We turned our attention to large scale reactions (Figure 4.8). We synthesized **4.4d** on a gram scale with 97% yield and 96% *ee*. We also synthesized **4.4e** and **4.4l** at a 1-mmol scale with 97% and 86% yield, respectively. We then synthesized 4.4m on a mmol scale, which was then applied to the formal synthesis of (*S*)-ofurace, a fungicide.[58] This demonstrates an application of this technology. By X-ray crystallography, carried out by Dr. Zhen Liu, we determined the (*S*) absolute configuration for these products.

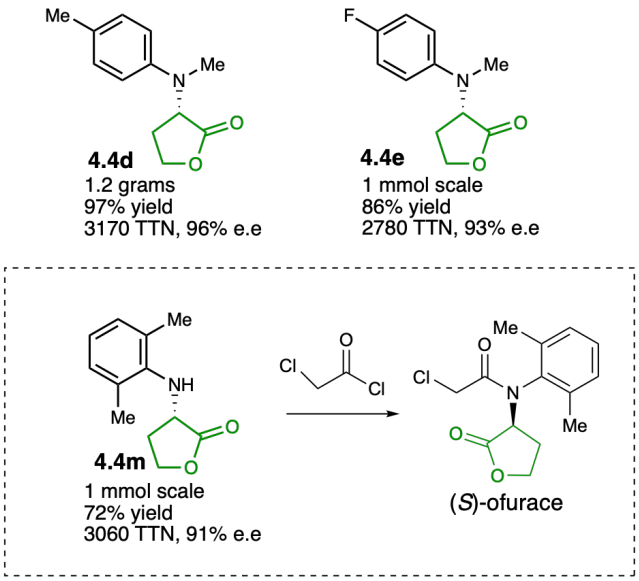


Figure 4.8: The application of our enzymes in large scale synthesis. This demonstrates that our enzymes remain effective on a preparative scale. The products are all formed with more than 2700 TTN, more than 70% yield. The *ee* is consistent with the analytical scale reactions. Product **4.4m** can be extended to the formal synthesis of (*S*)-orfurace, a fungicide.

4.4 Conclusion

While enantioselective N–H insertion is a challenge, we demonstrated that **L7-FL** is able to obtain 95% *ee* and >92% yield for the model reaction. **L7-F**L was tested on diverse substrates, including heterocycles and homobenzylic amines. Across these substrates, **L7-FL** has good to excellent yield (from 40 to >99%) and consistently high enantioselectivity (from 50% to 98% *ee*).Furthermore, **L7-FL** has shown to be scalable with preparative scale reactions. This further expands the growing repertoire of biocatalysts in asymmetric catalysis and could potentially provide an environmentally friendly approach to the synthesis of important chiral α-amino lactones.

Chapter 5

CONCLUSIONS AND FUTURE DIRECTIONS

We have expanded nature's chemical repertoire by creating biocatalytic methods to form bonds between a biologically privileged lactone into second period atoms. We leveraged the precise three-dimensional, active sites of enzymes to achieve high yield and enantioselectivity, even in products with contiguous chiral centers. We hope our enantioselective enzymatic methods for lactone insertion will be used in more environmentally friendly, cost-efficient, and sustainable synthesis of chemical targets.

Furthermore, this set of work has demonstrated an instance where evolved enzymes have shown impressive flexibility. We demonstrated that the same enzymatic lineage (P450-**L**) could catalyze both enantioselective C–H and N–H insertion reactions with excellent yields and selectivity. This goes against the notion that enzymes are only capable of an extremely limited set of reactions. This provides a good outlook for a future where many diverse industrial chemical reactions are achieved by enzymes. To build upon this work, we are interested in expanding our biocatalytic platforms to include more diverse substrates, including lactones of different sizes, and B–H, C–H, and N–H bonds with different electronic environments. We are also interested in better characterizing the precise mechanism of these lactone insertion reactions because this could provide valuable insights into enzymatic processes. Finally, we are searching for specific complex chemical targets that we may apply our enzymes toward, as we continue moving engineered biocatalysts out of the lab and into the world.

Chapter 6

SUPPORTING INFORMATION

6.1 Main Methods

Synthesis of 5-memberd Lactone Carbene Precursor (2.2)

A mixture of sodium azide (4.83g, 4eq), sodium hydroxide solution (80 mL, 3.875 M), hexane (80 mL), and tetrabutylammonium bromide (60 mg, 0.1 eq) were cooled on an ice bath. Trifluoromethanesulfonic anhydride (6.20 mL, 2 eq) was added dropwise in a reaction vesicle exposed to air. After 10 minutes, 2-acetyl-butyrolactone (2 mL, in 80 mL ACN) was added, and the reaction quickly turned yellow. After 30 minutes, the organic layer was extracted using a separatory funnel, dried with sodium sulfate, filtered, and concentrated. The product was then purified by a silica gel column. The purity of the product was confirmed by TLC and the product was vacuumed to yield a yellow-orange solid.

Synthesis of 6-membered Lactone Carbene Precursor

A round bottom flask was dried, put under N_2, and charged with diisopropyl amine (13.41 mmol, 1.88 mL). THF (anhydrous, 40 mL) was added and the flask was placed on ice. When cooled, a solution of n-butyllithium (13.4 mmol, 5.36 mL, 2.5 M, in hexanes) was added. Using a magnetic stir bar, the solution was stirred and cooled to -78 °C. tetrahydro-*2H*-pyran-2-one (11.2 mmol, in 27 mL THF) was added dropwise and the reaction was stirred (15 min, -78°C). Finally, 2,2,2-trifluoroethyl trifluoroacetate (16.8 mmol, 2.4 mL) was added dropwise and the reaction was stirred (-78 °C, 30 minutes). The reaction was quenched with HCl (50 mL, 10%). Following extraction (diethyl ether, 50 mL, three times) and washing of the organic layer (brine), the crude product was dried (Mg_2SO_4) and concentrated under vacuum. In the next step, the product was then dissolved in CH_2Cl_2 and σ-NBSA (14.1 mmol, 3.21 g) and DBU (2.39 mL, 16 mmol) were added. After stirring for 30 minutes (at room temperature), the product was quenched with water (50 mL). The product was extracted (diethyl ether, 40 mL, seven times). Following a wash with brine, the product was dried (Mg_2SO_4) and concentrated by vacuum. The product was purified using a silica column. Product identity was confirmed by TLC.

Synthesis of Authentic Standards (B–H Insertion)

To a solution of NHC-borane (1 equiv.), $Rh_2(OAc)_4$(2 mol%) was added. The diazo (1.2 equiv) was added through a syringe pump (addition over 2 h). After stirring (overnight), the product was concentrated and purified by silica chromatography.

Synthesis of Authentic Standards (N–H Insertion)

α-Bromo-γ-butyrolactone (330 mg, 0.185 mL, 2 mmol, 1 equiv), the aryl amine (2.5 mmol, 1.25 equiv.), and DMF (1.5 mL) were added to a vial. The reaction was stirred for 24 h (at 80 °C).

Saturation Mutagenesis

Single-site saturation mutagenesis libraries were constructed by a 22-codon trick protocol.[41] PCR products were *DpnI* digested. The fragments were purified by gel electrophoresis on either a 1% or 2% agarose gel, depending on the length of the DNA, followed by the Zymoclean™ DNA recovery protocol. A Gibson assembly protocol[59] was then performed on the fragments and the pET22b (+) backbone for ligation. These plasmid libraries were stored at -20°C.

Transformation

After cloning, electrocompetent cells (BL21 E. cloni®, Lucigen) were transformed with the plasmid libraries via electroporation. After shocking the cells, they were allowed to recover in SOC at 37 °C for 45 min. The cells were then spread on agar plates containing LB and ampicillin and allowed to incubate for 16 hours at 37 °C. Single colonies were picked for overnight cultures using sterilized toothpicks.

Glycerol Stock Preparation

Overnight cultures of *E. coli* harboring hemeprotein variants in LB_{Amp} were mixed with 50% glycerol solution (1:1) and stored at -78 °C.

DNA Extraction (Using QIAPrep Kit)

Overnight cultures of the hemeprotein variants were resuspended in buffer P1. Lysis buffer (250 μL) was added followed by neutralization buffer (350 μL). The mixture was then centrifuged (10 mins, 14000 $\times$ g), and the supernatant was added to a spin column. The suspension was centrifuged again (14000 $\times$ g, 1 min), and the flow through was discarded. Wash buffer (750 μL) was added to the column and centrifuged (14000 $\times$ g, 1 min). The flow through was discarded, and the column was transferred to a new vial. Elution buffer (50 μL) was added, and the solution

was centrifuged (14000 × g, 1 min). The plasmid DNA-containing eluate was stored at -20 °C, and an aliquot was sequenced.

Protein Expression (P411 Variants)

Overnight cultures of *E. coli* expressing the evolved P411 variants in LB_{Amp} grown at 37 °C for 15–18 hours (220 rpm) in a 96-well plate. Preculture (50 μL) was added to Hyper Broth supplemented with ampicillin (HB_{Amp}, 950 μL) for inoculation and was then allowed to grow for 2.5 hours (37 °C, 220 rpm, 80% relative humidity). The cells were then induced for expression by addition of 5-aminolevulinic acid (5-ALA, 1.0 mM, final concentration) and isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.5 mM, final concentration) after being cooled on ice for 45 minutes. Protein variants were expressed at 20 °C for 22 h and 220 rpm.

Reaction Setup (Plates)

E. coli in 96-well plates expressing protein variants were pelleted via centrifuge (4500 × g, 5 minutes, 4 °C). The supernatant was decanted, and the cells were resuspended in M9-N buffer (340 μL). D-glucose solution (40 μL, 250 mM) was added to the cell suspension (340 μL). The plates were then brought into an anaerobic chamber, where each substrate (10 μL, 400 mM in ethanol) was added. The plates were allowed to shake in the anaerobic chamber for 12 h.

Reaction Analysis (Reverse Phase HPLC)

Internal standard (20 μL of 20 mM solution of phenethyl acetate, *p*-methyl-anisole, allyl phenyl ether, or 1,3,5-trimethoxy-benzene) and ACN (580 μL) were added to each well in the plate. The internal standard selection is substrate-dependent and is chosen to avoid overlap with any peaks. The plate was vortexed and centrifuged. The supernatant was then filtered through centrifuge filtration. The resulting filtrate was analyzed on the HPLC using a developed method (C18-Kromasil column, 40% water, 60% ACN, 1.2 ml/min) that results in unique and discrete peaks.

Protein Expression (P411, Validation)

Overnight cultures of *E. coli* harboring a desired P411 variant were set up in LB_{Amp} (6 mL) and grown at 37 °C for 15-18 hours (250 rpm). Five mL of preculture were added to HB_{Amp} (45 mL, with ampicillin and D-glucose) and shaken at 37 °C for 2.5 h). The cultures were cooled on ice for 45 minutes and 5-aminolevulinic acid (5-ALA, 1.0 mM final concentration) and β-D-1-thiogalactopyranoside (IPTG,0.5 mM final concentration) were added. Protein variants were expressed at 20 °C for

22 h (140 rpm).

Protein Expression (Cytochrome *c*, Validation)

Overnight cultures of *E. coli* harboring a desired cytochrome *c* variant were set up in LB_{chlor} (5 mL) and grown at 37 °C for 15-18 hours (250 rpm). Preculture (0.5 mL) was added to HB_{chlor} (25 mL, treated with ampicillin and glucose) and shaken (37 °C, 3h). The cultures were cooled on ice for 45 minutes and 5-aminolevulinic acid (5-ALA) and β-D-1-thiogalactopyranoside (IPTG) were added (200 μM and 20 μM final concentration). Protein variants were expressed at 20 °C for 22 h (160 rpm).

Reaction Setup (Validation)

The cultures were centrifuged (4500 × g, 5 minutes, 4 °C). The supernatant was decanted, and the cells were resuspended in M9-N buffer such that an OD_{600}= 60 was achieved. To vials, cell suspension (340 μL) and D-glucose (40 μL, 250 mM) were added. Substrates (10 μL, 400 mM in ethanol) were added to the vials in an anaerobic chamber, then the vials were sealed and allowed to shake for 12 h. Replicates were worked up for HPLC analysis (detailed above) or worked up with 1:1 hexane : ethyl acetate for chiral HPLC analysis.

Reaction Analysis (Chiral HPLC)

1:1 Hexane : ethyl acetate (1 mL) was added to each reaction vial. After cells were vortexed and centrifuged (4500 × g, 5 minutes), the organic phase was transferred to glass vials. The solvent was removed by rotary evaporation, and the dichloromethane (200 μL) was added to each reaction vial. Using synthesized authentic standards, chiral HPLC methods were developed that allowed for discrete peaks for both enantiomers. Using these methods, the enantiomeric excess of the reactions was assessed.

Hemochrome Assay

Aliquots of an OD_{600}=60 cell suspension (3 mL) were added to 15-mL tubes and lysed by sonication (2 minutes, 1 second on/ 1 second off, two times). The lysed cells were then transferred to Eppendorf tubes and centrifuged (20,000 × g,, 4 °C, 10 minutes). The supernatant was transferred to a glass vial. Then, a solution of NaOH/pyridine/$K_3Fe(CN)_6$ (500 μL) was mixed with the lysate (500 μL) and the UV absorbance was assessed (380 nm – 650 nm). Then, a solution of dithionite (5 μL) was added and the UV absorbance was assessed again (reduced spectrum).

Enzymatic Synthesis of Authentic Standards (C–H Insertion)

The desired protein was expressed as described above, using larger quantities (250 μL HB_{Amp}). The cells were centrifuged (4500 $\times$ g, 5 minutes, and 4 °C) and resuspended in M9-N buffer such that an OD_{600}=60 was achieved. To 50-mL Falcon tubes, cell suspension (30 mL) was added followed by D-glucose solution (3 mL). In an anaerobic chamber, substrates (1.25 mL, 400 mM in ethanol) were added. The tubes were shaken in the anaerobic chamber for 12 h. The cells were then treated with 1:1 hexane : ethyl acetate (15 mL) and centrifuged (5000 $\times$ g, 10 minutes, room temperature). The organic phase was separated. This process was repeated four times. The organic phase was dried by magnesium sulfate. Following filtration, the products were isolated by silica gel chromatography.

Calibration Curves

A stock solution of the purified product (500 μL, 100 mM) was prepared. Vials containing 0.6 mL acetonitrile, 0.4 mL H2O, and 20 μL of a desired internal standard were prepared. Incrementing amounts of the stock solution were added to vials (4 μL, 8 μL, 16 μL, 24 μL, 32 μL, 40 μL) in duplicate. These vials were than assayed by HPLC and a calibration curve was developed.

BIBLIOGRAPHY

(1) Chen, K. Q.; Arnold, F. H. *Proc. Natl. Acad. Sci.* **1993**, *90*, 5618–5622.

(2) Arnold, F. H. *ACIE* **2017**, *57*, 4143–4148.

(3) Hawkins, R. E.; Russell, S. J.; Winter, G. *J Mol Bio* **1992**, *226*, 889–896.

(4) Abudayyeh, O. O.; Gootengberg, J. S.; Franklin, B., et al. *Science* **2019**, *365*, 382–386.

(5) Yokobayashi, Y.; Weiss, R.; Arnold, F. H. *Proc. Natl. Acad. Sci.* **2002**, *99*, 16587–16589.

(6) Chen, K.; Arnold, F. H. *Nat. Catal.* **2020**, *3*, 203–213.

(7) Turner, N. J. *Nat. Chem. Biol.* **2009**, *5*, 567–573.

(8) Saville, C. K. et al. *Science* **2010**, *329*, 305–309.

(9) Huffman, M. A. et al. *Science* **2019**, *366*, 1255–1259.

(10) Joo, H.; Lin, Z.; Arnold, F. H. *Chem Biol.* **1999**, *6*, 699–706.

(11) Lewis, J. C. et al. *ChemBioChem* **2010**, *11*, 2502–2505.

(12) Meinhold, P.; Peters, M. W.; Chen, M. M. Y.; Takahashi, K.; Arnold, F. H. *ChemBioChem* **2005**, *6*, 1765–1768.

(13) Farinas, E. T.; Alcalde, M.; Arnold, F. H. *Tetrahedron Lett.* **2003**, *60*, 525–528.

(14) Wong, T. W.; H, A. F.; Schwaneberg, U. *Biotechnol. Bioeng.* **2004**, *85*, 351–358.

(15) Prier, C. K.; Arnold, F. H. *J. Am. Chem. Soc.* **2015**, *137*, 13992–14006.

(16) Knight, A. M.; Kan, S. B. J.; Lewis, R. D.; Brandenberg, O. F.; Chen, K.; Arnold, F. H. *ACS Cent. Sci.* **2018**, *4*, 372–377.

(17) Chen, K.; Huang, X.; Kan, S. B. J.; Zhang, R. K.; Arnold, F. H. *Science* **2018**, *360*, 71–75.

(18) Chen, K.; Arnold, F. H. *J. Am. Chem. Soc.* **2020**, *142*, 6891–6895.

(19) Zhang, R. K.; Chen, K.; Huang, X.; Wohlschlager, L.; Renata, H.; Arnold, F. H. *Nature* **2019**, *565*, 67–72.

(20) Hyster, T. K.; Arnold, F. H. *Isr. J. Chem* **2014**, *55*, 15–20.

(21) Kan, S. B. J.; Gumulya, X. H. Y.; Chen, K.; Arnold, F. H. *Nature* **2017**, *552*, 132–136.

(22) Kan, S. B. J.; Lewis, R. D.; Chen, K.; Arnold, F. H. *Science* **2016**, *354*, 1048–1051.

(23) Kleingardner, J. G.; Bren, K. L. *Acc. Chem. Res.* **2015**, *48*, 1845–1852.

(24) Stelter, M. et al. *Biochemistry* **2008**, *47*, 11953–11963.

(25) Giver, L.; Gershenson, A.; Freskgard, P. O.; Arnold, F. H. *Proc. Natl. Acad. Sci.* **1998**, *95*, 12809–12813.

(26) Chandgude, A. L.; Ren, X.; Fasan, R. *J. Am Chem Soc.* **2019**, *141*, 9145–9150.

(27) Tyagi, V.; Bonn, R. B.; Fasan, R. *Chemical Science* **2015**, *6*, 2488–2494.

(28) Gu, Y.; Natoli, S. N.; Liu, Z.; Clark, D. S.; Hartwig, J. F. *ACIE* **2019**, *58*, 13954–13960.

(29) Liu, Z.; Arnold, F. H. *Curr. Opin. Biotechnol.* **2021**, *69*, 43–51.

(30) D, D. K.; Borys, F.; Brodzka, A.; Ostaszewski, R. *Eur. Jour. Org. Chem.* **2019**, *7*, 1653–1658.

(31) Denmark, S. E.; Burk, M. T. *Proc. Natl. Acad. Sci.* **2010**, *107*, 20655–20660.

(32) Junji, I.; Kuniko, H.; Hiroko, S.; Tsutomu, K.; Masaru, Y. *Bull. Chem. Soc. Jpn.* **1993**, *7*, 1989–1993.

(33) Renz, M.; Meunier, B. *Eur. J. Org. Chem.* **1999**, *1999*, 737–750.

(34) Monfette, S.; Fogg, D. E. *Chem. Rev.* **2009**, *109*, 3783–3816.

(35) Sartori, S. K.; Diaz, M. A. N.; Munoz, G. D. *Tetrahedron* **2021**, *84*, 132001.

(36) Farina, V.; Reeves, J. T.; Senanyake, C. H.; Song, J. J. *Chem. Rev.* **2006**, *107*, 2734–2793.

(37) Paranjeet, P.; Gopal, K.; Surendra, K. *Curr. Org. Synth.* **2017**, *18*, 665–682.

(38) Das, B. C. et al. *Future Med. Chem.* **2013**, *5*, 653–676.

(39) Li, X.; Curran, D. *J. Am. Chem. Soc.* **2013**, *135*, 12076–12081.

(40) Deangelis, A.; Panish, R.; Fox, J. M. *Acc. Chem. Res.* **2016**, *49*, 115–127.

(41) Kille, S. et al. *ACS Synth. Biol.* **2013**, *2*, 82–92.

(42) Otog, N.; Chanthamath, S.; Fujisawa, I.; Iwasa, S. *Eur. Jour. Org. Chem.* **2021**, *10*, 1564–1567.

(43) Yi, L.; Bandu, M.; Desaire, H. *Anal. Chem.* **2005**, *77*, 6655–6663.

(44) Davies, H. M. L.; Pelphrey, P. M., *Organic Reactions*, 2011.

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

(46) Davies, H. M. L.; Jin, Q. *Org. Lett.* **2004**, *6*, 1669–1672.

(47) Suematsu, H.; Katsuki, T. *J. Am. Chem. Soc.* **2009**, *40*, 14218–14219.

(48) Prier, C. K.; Zhang, R. K.; Bueller, A. R.; Brinkmann-Chen, S.; Arnold, F. H. *Nat. Chem.* **2017**, *9*, 629–634.

(49) Starr, T. N.; Thronton, J. W. *Protein Sci.* **2016**, *25*, 1204–1218.

(50) Woods, J. R.; Mo, H.; Bieberich, A. A.; Alavanja, T.; Colby, D. A. *Med. Chem. Comm.* **2012**, *4*, 27–33.

(51) Hili, R.; Yudin, A. K. *Nat. Chem. Biol.* **2006**, *2*, 284–287.

(52) Froidevaux, V.; Negrell, C.; Caillol, S.; Pacault, J. P.; Boutevin, B. *Chem. Rev.* **2016**, *116*, 14181–14224.

(53) Bariwal, J.; der Eycken, E. V. *Chem. Soc. Rev.* **2013**, *42*, 9283.

(54) Ren, Y. Y.; Zhu, S. F.; Zhou, Q. L. *Org. Biomol. Chem.* **2018**, *16*, 3087–3094.

(55) Sharon, D. A.; Mallick, D.; Wang, B.; Shaik, S. *J. Am. Chem. Soc.* **2016**, *138*, 9597–9610.

(56) Streck, V.; Carminati, D. M.; Johnson, N. R.; Fasan, R. *ACS Catal.* **2020**, *10*, 10967–10997.

(57) Li, M. L.; Yu, J. H.; Li, Y. H.; Zhu, S. F.; Zhou, Q. L. *Science* **2019**, *366*, 360–364.

(58) Fisher, D. J.; Hayes, A. L. *Pestic. Sci.* **1982**, *13*, 330–339.

(59) Gibson, D. G.; Young, L.; Chuang, R. Y.; Venter, J. C.; Hutchinson, C. A.; Smith, H. O. *Nat. Methods* **2009**, *6*, 343–345.