Austrian Marshall Plan Foundation

Final Report

"Engineering a Biocatalyst for Carbenoid C–H Insertion"

conducted at the

California Institute of Technology (Caltech), Pasadena

Supervisors (Caltech)

Prof. Frances Arnold

Dr. Sabine Brinkmann-Chen

Ruijie Kelly Zhang, BSc

Supervisor (TU Wien)

Prof. Marko Mihovilovic

by Lena Wohlschlager, BSc

February 20th, 2017

Content

1 Introduction

1.1 C–H Insertion – A Form of C–H Functionalization

When a synthetic chemist looks at an organic molecule he will probably recognize its functional groups first, since most standard approaches in organic synthesis rely on the introduction and modification of functional groups. When it comes to complex molecules this often means multi-step synthesis, which is time consuming, little atom efficient and requires a lot of effort. Moreover, we are limited to those transformations that are possible with the pre-existing groups in available reagents.

We could enable access to a variety of completely new molecules, if we developed methods to perform reactions selectively on ubiquitous C–H bonds. The capability of creating new C–C bonds (or C–O or C–N bonds, for that matter) could lead to a paradigm shift in the way synthetic routes are designed.[1,2]

A promising approach and therefor of high interest in research is the insertion reaction of metal carbenoids.[3,4,5] Other than in C–H activation, where the metal inserts into a C–H bond and thereby activates the substrate for further functionalization, C–H insertion involves the decomposition of a carbene precursor leading to the formation of a metal carbenoid intermediate, which then inserts into the C–H bond of the substrate.

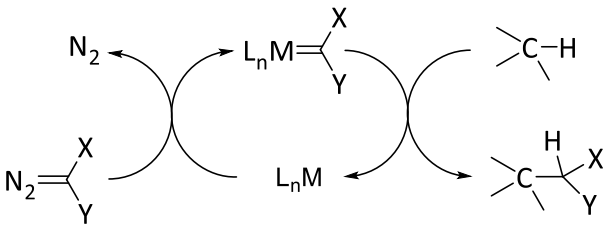


Figure 1: C-H insertion

The major challenge this chemistry faces is the design of carbene reagents that are reactive enough to cleave relatively strong C–H bonds but at the same time feature satisfactory selectivity. Especially the later involves difficulties, since usually many C–H bonds are present in one organic molecules.[6]

This is the reason why first accomplishments in this field were attained with intramolecular reactions. The best results were acquired with catalysts based on copper or rhodium.[4]

More recently, breakthroughs in the investigation of intermolecular reactions were achieved with the discovery that the reactivity of the metal carbenoid can be modulated by altering the structure of the carbene. While early studies with ethyl diazoacetate – an acceptor-only carbene precursor – showed poor selectivity[7], it was later discovered that carbenoids with a donor ligand (such as an aryl or vinyl group) perform a highly regioselective insertion reaction.[1,6]

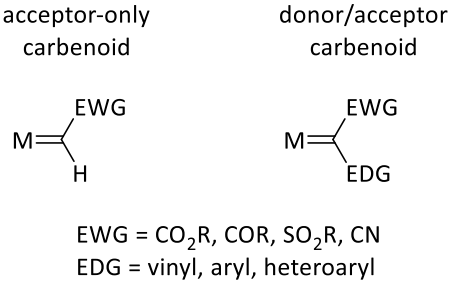


Figure 2: Classification of carbenoids

Therefore, most carbene precursor used these days contain one electron-withdrawing group for reactivity and one electron-donating group for selectivity.[3] But again, this means we are constrained in terms of what groups we add to the starting compound.

Still, most organometallic catalysts developed for carbenoid insertion reactions contain the precious metal rhodium. A few examples haven been published with copper-based catalysts, and even less report the use of an iron complex.[1,3,4,8]

We believe that with the use of iron-dependent enzymes as catalysts, we can circumvent the use of expensive and precious transition metals and, furthermore, are able to reach a higher regio-, stereo- and enantioselectivity.

1.2 Enzymes as Catalysts

Making use of enzymes in chemical transformations is not a new idea. One of the first reports of an enzyme applied in organic synthesis was published by Rosenthaler in 1908, where he used a plant extract to synthesize (R)-mandelonitrile from benzaldehyde and hydrogen cyanide.[9]

Of course, todays approach of utilizing proteins is a lot different. The development of protein engineering technologies in the 1980s and 1990s provided us with tools to modify enzymes and ever since, ongoing improvements in this field increase our possibilities of creating biocatalysts with desired features and properties.[10]

Especially, since in recent years research has focused on improving organic synthesis in terms of ecological aspects and tried to make progress in the direction of 'green' chemistry, biocatalysis is a promising attempt in meeting demands like less chemical waste, mild reaction conditions and biodegradability.[11,12]

Most enzymes have their source in mild environments (20 – 40°C, pH 5 – 8) and therefore work under these ambient conditions.[12] In general, they work best in water, but various enzymes show a great tolerance towards organic solvents.[13] Since a majority of the enzymes origin from similar milieus they are compatible with each other. That makes it possible to perform several biocatalytic transformations in a one-pot reaction cascade.[14] Also, enzymes are not limited to their natural role, but show a high tolerance for non-natural substrates. Sometimes enzymes are even more efficient compared to chemical catalysts and can be used in lower concentrations than their organometallic equivalent.
It is the enzymes' task to work on a specific type of functional group, leaving other sensitive parts of the molecule unchanged. Moreover, enzymes have the ability to distinguish between functional groups which are chemically identical but are placed in different environments within the same molecule. Therefore, enzymes exhibit a high chemo- and regioselectivity. Due to the fact that all enzymes are made from L-amino acids they are all chiral catalysts and show high selectivity for one possible enantiomeric product. The drawback here is that the stereo-preference cannot simply be inverted by changing the chiral catalyst, since amino acids are not available in the D-configuration.[12]

The options on how to use enzymes are manifold. It can start from a single enzyme performing a single specific reaction *in vitro*, but also lead to a multi-enzymatic cascade connected to the metabolic system of a living cell. The advantages of performing *in vitro* reactions with isolated enzymes are that only one reaction is catalyzed, with side reactions merely avoided, that the reaction parameters (temperature, pH, organic solvent concentration, etc.) are easy to control and that the work up is simple and straightforward. On the downside, the desired enzymes have to be isolated and stabilized beforehand, and lots of enzymes need co-enzymes or co-factors, which are expensive and require the implementation of a complex *in situ* regeneration system. Also, not all enzymes appear to be stable under *in vitro* conditions.[15]

This is where you can make use of the features of *in vivo* biotransformation in whole cell systems. The advantages lie in the cells ability to synthesize its needed precursor compounds, co-factors, etc. itself and from reasonably priced medium. Cofactor recycling is automatically performed by the cell. The downside of this approach is that the complexity of the cellular systems complicates the control and adaption of the synthetic procedure. Side reactions are more likely to occur, and a lower substrate concentration tolerance as well as a low tolerance of organic solvent may lead to lower productivity.[12]

As mentioned before, the great progresses in molecular biology and protein engineering enables us now, to modify and develop proteins with desired properties. We are not limited to the enzymes found in their wild type form in nature, but have the opportunity to tune and adapt their functions to our specifications.

1.3 Directed Evolution

It was around 20 years ago that Frances H. Arnold pioneered with her work on developing a new innovative molecular biology method called directed evolution or DE.[10] It revolutionized the way enzymes could be modified since it has proven to be a successful method without the need of extensive knowledge of sequence, structure and function of a protein which is inevitable for rational engineering approaches.[16]

Directed evolution basically mimics Darwinian evolution in a test tube. The broad manifoldness of today's existing enzymes descends from thousands of years of evolution from ancestral proteins. With DE the scientist tries to imitate nature's way of evolving enzymes towards improved properties like activity, stability, selectivity, acceptance for non-natural substrate or tolerance of organic solvents.

Somewhat simplified, the principle of directed evolution can be divided into two parts:
a) creating a library of variant enzymes by mutagenesis of the parent enzyme and
b) screening of the library to identify those variants that show improvement in the desired feature. This is usually repeated several times (the scientist speaks of 'generations') – each time with the best variant of the round before – to get a further enhanced enzyme.[17]

For a successful application of directed evolution a few requirements have to be fulfilled. It is essential to find a suitable microbial host for expression of the protein as well as a rapid but sufficient screening method.[16]

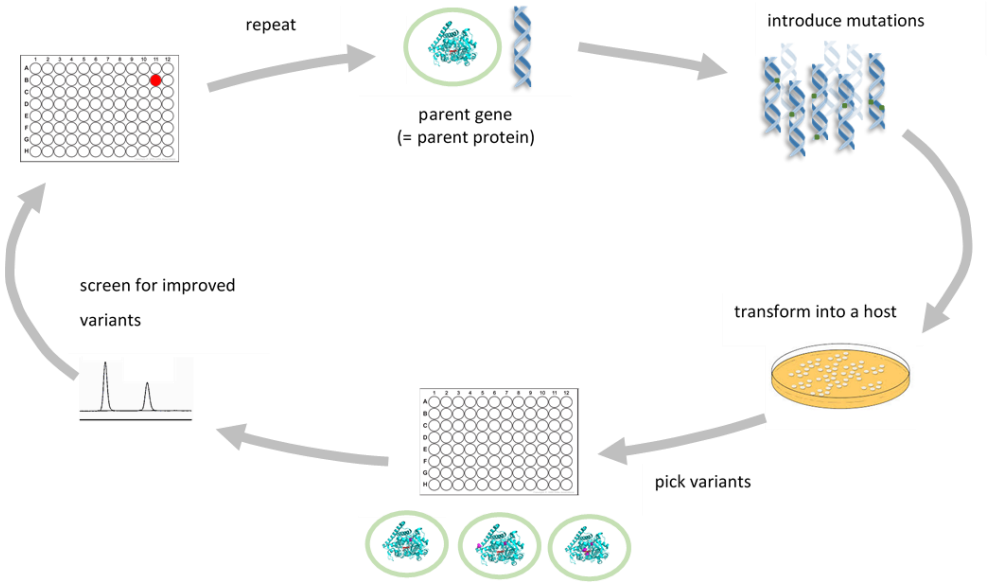


Figure 3: Scheme of directed evolution

Often, the first challenge is to find an appropriate starting point, meaning an enzyme that can be changed toward our desired functionality. This can be an exceptional tricky task when the goal is to design an enzyme for catalyzing a reaction not found in nature. A good strategy is to look for similarities between an already existing biocatalytic transformation and the requested non-natural reaction or identify enzymes known for binding similar substrates.[17]

Once a parent enzyme is assigned, the suitable way to introduce mutations into the gene has to be chosen.

The most common technique to generate random mutations is probably error-prone PCR (epPCR, Figure 4 A)). In this approach PCR amplification is carried out with *Taq* polymerase (a thermostable DNA polymerase with an average error rate of $8*10^{-6}$ mutations/bp[18]) under suboptimal conditions caused by varying the dNTP or $MgCl_2$ concentration or addition of $MnCl_2$.The implementation of this procedure is rather straightforward and easy but the drawbacks of this method are not negligible. Redundancy of the genetic code as well as mutational biases caused by the polymerase prevents diversity at the protein level and since usually the mutation rate is kept low, the probability to introduce mutations close to each other is low as well. Also, to diminish the chance of missing beneficial mutations a large number of mutants have to be screened, which demands a lot of effort.

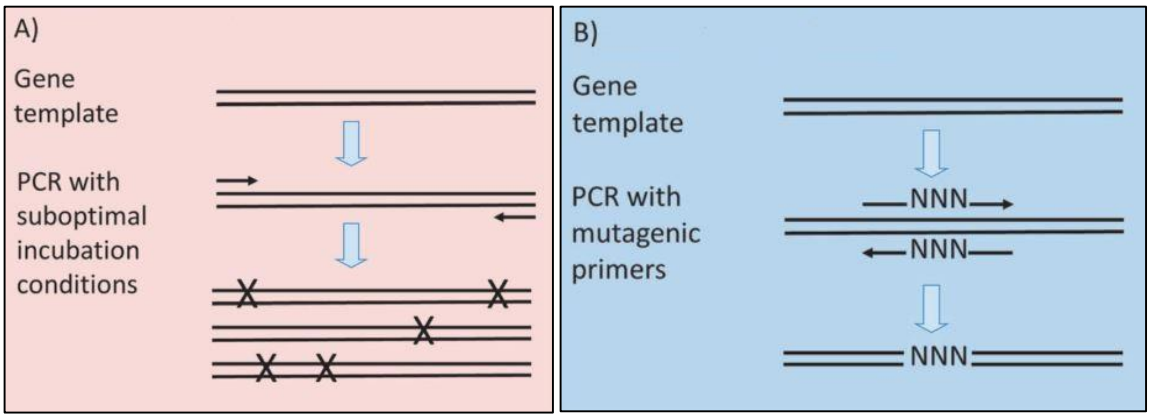


Figure 4: Methods for introducing mutations: A) error prone PCR; B) site-saturation mutagenesis[19]

Another approach is to target specific sites via site-saturation mutagenesis (SSM, Figure 4 B)). For this technique the sequence of the gene has to be known. PCR primers are designed that contain the desired mutation flanked by sequences complementary to the target sequence. Compared to site-directed mutagenesis, where one amino acid is changed for another specific amino acid, site-saturation mutagenesis substitutes the targeted position with all other 19 amino acids resulting in many different variants to explore which new residue is the most favorable.[19,20] This is accomplished by using mixed base codons. To reduce the amino acid bias caused by the genetic code redundancy (*e.g.* serine is encoded by six codons, while methionine is only encoded by one codon) and with that the number of variants that have to be sampled to get a high probability of screening at least one variant with each of the 20 amino acids, we applied the "22-codon trick"[21]. In this approach three different oligonucleotides are used as primers for PCR: two carrying degenerated codons (NDT –12 codons coding for 12 amino acids, VHG – 9 codons coding for 7 amino acids[a]) and one containing TGG (one codon). This decreases the redundancy of the genetic code to a codon to amino acid ratio of 22:20. For screening, only 66 colonies have to be picked to cover 95% of the variants.

Although site-saturation mutagenesis allows very precise and specific mutations, the constrained number of residues that can be efficiently targeted at a time is adverse.

There are other methods (*e.g.* insertion/deletion, recombination techniques)[19,20] used to create libraries with mutated variants but since they have not been applied in this work, they

[a] N = A/C/T/G; D = no C; V = no T; H = no G

will not be discussed in further detail. It is the scientists task to evaluate all the options and decide which one applies best to the demand of the experiment.

Screening is usually performed in microtiter plates. Ideally, UV/Vis-absorbance or fluorescence based screening assays can be developed for the observed reaction since those allow fast and simple high-throughput screening. If that is not an option, it has to be recoursed to analysis methods like HPLC or GC. No matter which technology is used, the goal is always a fast and accurate screening performance to identify the best improved variant.[19]

1.4 Cytochrome P450

1.4.1 Overview

The enzyme studied and engineered in this work belongs to the P450 cytochrome (CYP) superfamily. This type of proteins has been explored for almost 60 years and also the research group around Frances H. Arnold has been working with it for nearly 10 years.[22,23]

P450s are heme-b dependent monooxygenases and found in almost every organism. Their name origins from the characteristic absorption at 450 nm observed when carbon monoxide is bound to the Fe-ion of the heme group.[24]

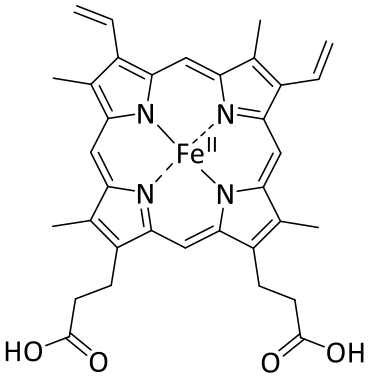


Figure 5: Heme-b cofactor

All P450 enzymes studied show a highly maintained tertiary structure, but only one amino acid is totally conserved in all the wild types– the cysteine residue that ligates to the iron center of the heme co-factor and therewith binds the prosthetic group to the polypeptide chain. The rest of the amino acid sequence can vary a lot (several P450s have less than 15% identity) being the reason for different substrate selectivity and redox partner interactions.[25,26]

Cytochrome P450 enzymes play several physiological roles. They are, for example, known for breaking down xenobiotics (*e.g.* drugs) but are also involved in the synthesis of steroid hormones in mammals and antibiotics in bacteria and fungi. [23,25,26] P450s show a broad substrate scope, including fatty acids, alcohols and aliphatic as well as aromatic hydrocarbons.[24]

The major reaction performed by these proteins is the oxidation of an organic substrate to the corresponding alcohol. Formally, they cleave molecular oxygen and insert into a C–H bond of the substrate to give the oxygenated product and water as a side product. The catalytic reaction cycle is illustrated in Figure 6. The four equatorial coordination sites of the central iron ion are occupied by the nitrogen atoms of the porphyrin ring. One axial position binds to the polypeptide chain through a thiolate formed with cysteine (not shown in figure), while the other faces the active site of the enzyme. This is where the reaction is taking place.

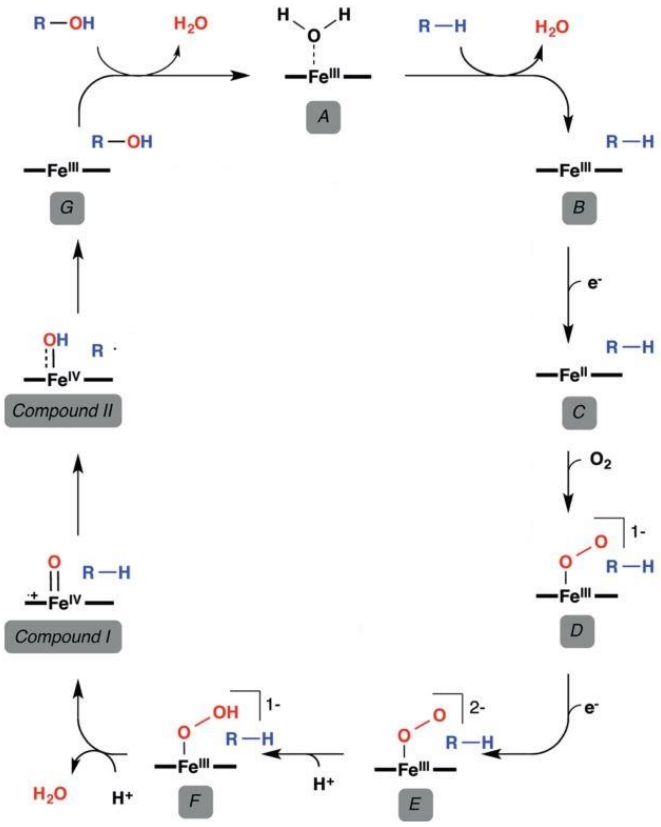


Figure 6: Cytochrome P450 monooxygenation mechanism[27]

In the resting state (Figure 6, **A**), the active site contains a water molecule and the iron is in the ferric state (oxidation number +3). When an organic substrate emerges it displaces the

water ligand (**B**)and causes the iron to change from the low-spin to the high-spin state where it is then easily reduced via a single electron transfer from the reductase domain. In the ferrous (oxidation number +2) state (**C**) it can coordinate molecular oxygen (**D**) and by another reducing step a ferric iron-peroxo species (**E**) is formed. Protonation (**F**) and subsequent elimination of water gives **Compound I**, a highly reactive Fe(IV)-oxo intermediate. It abstracts a hydrogen atom by cleaving a C–H bond providing an organic radical and **Compound II**. The two intermediates undergo a "rebound" process which yields the monooxygenated product and returns the catalyst to the ferric iron low-spin state.[25,27]

The reductase domain and the mechanism for shuttling electrons to the heme domain is not the same in every cytochrome P450. Type I P450s have a two-protein redox chain to transfer electrons from NADH (bacterial) or NADPH (mitochondrial) to the active site: FAD reductase →$(Fe - S)_2$ protein→heme. Type II proteins (microsomal) possess a single FAD/FMN domain that delivers electrons from NADPH.[28,29]

However, this classification is not a clear distinction which becomes apparent when having a closer look at the cytochrome $P450_{BM3}$.

1.4.2 $P450_{BM3}$ (CYP102A1)

The enzyme I have been working with during my research originates from *Bacillus megaterium*. It was the third cytochrome P450 to be isolated from *B. megaterium*, thus the index BM3.[30]

The whole protein consists of 1048 amino acids differentiated into three structural domains: the heme domain with amino acids 1–470, the FMN (flavin mononucleotide) domain with residues 471–664 and the FAD (flavin dinucleotide) domain with residues 665–1048.[29]

It is the first bacterial P450 discovered to possess a microsomal (type II) reductase. $P450_{BM3}$ is a fatty acid hydroxylase (its natural substrates assumed to be branched-chain saturated fatty acids[25]) that displays strong similarities to eukaryotic fatty acid hydroxylases (CYP4A). Its redox system is a mammal-like NADPH-dependent cytochrome P450 reductase (CPR) comprising the two prosthetic flavin groups FAD and FMN in equimolar ratio.

But these are not the only features making $P450_{BM3}$ an interesting target for research. An outstanding characteristic is that the reductase domain is fused to the C-terminal of the heme domain resulting in a single 119-kDa polypeptide chain (for comparison, most other P450s are 40–60 kDa in size and the reductase proteins form an independent amino acid chain). This

unique property seems to be the reason for its high electron transfer rate and superior catalytic activity (highest activity level found in all P450 fatty acid hydroxylases).

Moreover, $P450_{BM3}$ can easily be applied to the *Escherichia coli* recombinant expression system enabling to produce large amounts of the enzyme. As most bacterial P450s it is soluble in water which makes it easy to handle, overexpress, purify and crystallize.

For this reasons, $P450_{BM3}$ has become one of the most studied enzymes of the cytochrome family as well as an attractive research target for biochemical and biotechnological applications.[24,26,29]

1.4.3 Cytochrome P450s in the Arnold Group

As mentioned above, the research group of Frances H. Arnold has a long history of working with cytochrome P450.[23] Only a few years ago they have started to explore the P450 enzyme's capability of performing non-natural reactions and this field has been intensely studied ever since. [17,31,32]

It is the electronic similarity between an oxene and a carbene/nitrene that rose the idea that a cytochrome P450 may be able to catalyze carbene- and nitrene-transfer reactions.

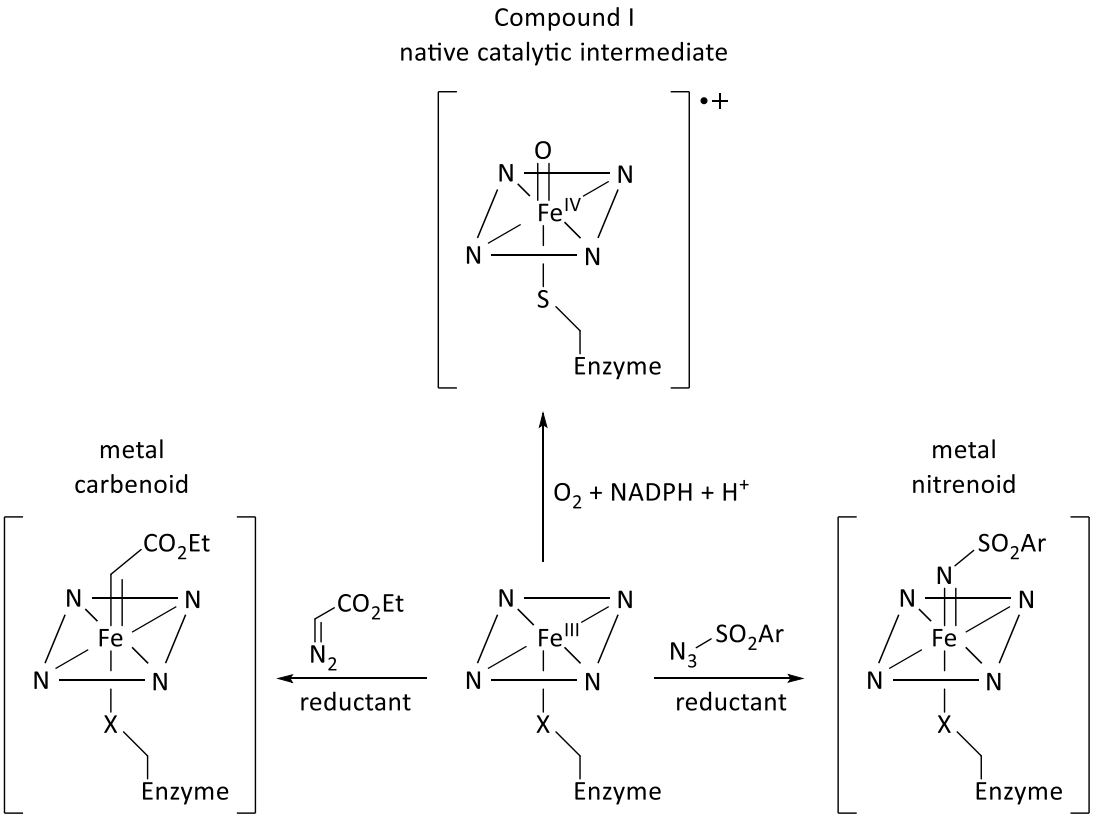


Figure 7: Similarity between oxo-intermediate of the natural reaction mechanism and metal carbenoid/nitrenoid – proposed intermediates for non-natural reactions[31]

As a first reaction, the cyclopropanation of styrene with ethyl diazoacetate (EDA) was tested with wild-type $P450_{BM3}$ by Coelho *et al.* in 2013.[33] Traces of the desired product were observed and further studies with several cytochrome P450s, other hemoproteins as well as hemin (free iron-heme) were carried out.[33,34] The results showed that cyclopropanation can be catalyzed with various enzymes but varied diastereo- und enantioselectivies were observed. This led to the assumption that the reaction is taking place in the active site of the enzyme and that the enzyme's amino acid sequence and structure influences the stereochemistry of the product. Therefore, P450s appeared to be promising starting enzymes to be engineered toward new non-natural reactions.

During the first rounds of directed evolution on $P450_{BM3}$ to improve the activity as well as the selectivity of the cyclopropanation reaction, it was discovered that changing the axial cysteine ligand (that binds the heme group to the polypeptide chain) to a serine[35] residue (C400S) or to a histidine[36] (C400H) residue were the most beneficial mutations. It caused the characteristic peak of the CO-bound enzyme at 450 nm to shift to 411 nm (for serine ligated P450s; histidine ligation shifted to a slightly longer wavelengths). Variants with the C400S mutation have been labeled "P411".

The serine ligation also altered the redox potential of the heme cofactor. The iron could be reduced to Fe^{2+} under cellular conditions enabling the reaction to be performed *in vivo*. The P411s could now be expressed in *E. coli* and the reaction directly carried out in the whole cells.[35]

Further investigations expanded the substrate scope for cyclopropanation and gained access to the synthesis of precursors to derivatives of the antidepressant levomilnacipran. The turnovers and selectivities achieved with the new variants are comparable to the best rhodium-based organometallic catalysts [36,37]

Of course, cyclopropanation was not the only reaction of interest. In the last three years, various publications showed the successful engineering of variants for intramolecular C–H amination[38,39], sulfimidation[40,41] and aziridination[42] via nitrenoid insertion (azides were used as precursor). Also, a successful carbenoid C–H insertion into N–H bonds (using ethyl diazoacetate as carbene precursor) was reported by the Arnold group.[43]

1.5 Aim of this Project

We believe that the range of opportunities for new non-natural reactions catalyzed by cytochrome P450 enzymes is far from exhausted. The goal is to explore all the possibilities P450s (or P411s) have to offer.

After the successful development of protein catalysts for various nitrene transfer reactions and also for cyclopropanation and carbenoid C–H insertion into N–H bonds (see chapter 1.4.3) the focus now lies on the formation of new intermolecular C–C bonds.

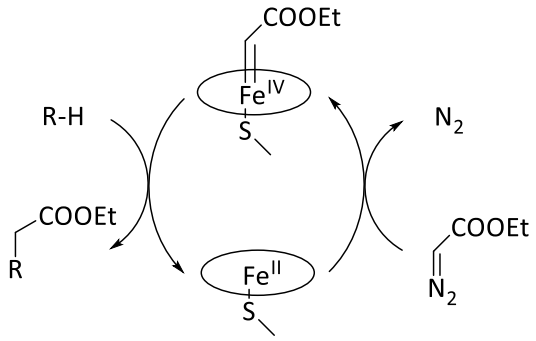


Figure 8: Enzyme-catalyzed intermolecular carbenoid C–H insertion

The aim of this project is to engineer an enzyme of the cytochrome P450 family using directed evolution methods to develop a highly active biocatalyst for selective intermolecular carbene transfer reactions.

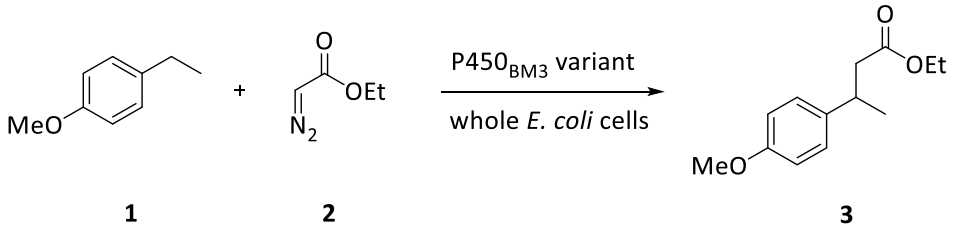
2 Results and Discussion

Unless otherwise mentioned, screening of mutagenesis libraries and rescreen of potential hits was always performed in whole E. coli cells in aqueous M9-N media (OD_{600} = 45 in rescreen reactions) under anaerobic conditions with a substrate concentration of 10 mM each (they were added to the reaction as solution in ethanol, resulting in 5% ethanol co-solvent). It was analyzed via GC and HPLC measurements.

2.1 Previous Work on this Project

Before I joined the project, my supervisor Ruijie Kelly Zhang has tested various $P450_{BM3}$ variants that have been engineered in the group for previous projects for catalytic activity toward our model reaction (Table 1).

As a screening model the reaction between 4-ethylanisole (**1**) and ethyl diazoacetate (**2**) was chosen (Scheme 1).



1 **2** **3**

Scheme 1: Carbenoid C–H insertion reaction with 4-ethylanisole and ethyl diazoacetate

Aromatic compounds have a rather weak C–H bond (~85 kcal/mol[44]) at the benzylic position so this molecule seemed to be a good model substrate for this reaction.

EDA is used as the carbene precursor. The poor selectivity of EDA due to its high reactivity narrowed its use as a carbene precursor in organometallic-catalyzed reactions (see chapter 1.1). A whole new range of products would be accessible if selective carbenoid C–H insertion reactions were feasible with acceptor-only carbenoids as well.

The variant "P_0-A82L" originating from one of the nitrene transfer projects[41] showed traces (0.13% yield) of the desired product **3**. Other variants related to "P_0-A82L" but differing by one amino acid residue at the 78 or 263 position were also found to catalyze the reaction at trace

levels (Table 1). All these variants have the C400S mutation of the axial ligand and are therefore technically "P411s".

Table 1: Tested $P450_{BM3}$ variants for carbenoid C–H insertion.

Variant	Mutations relative to wild-type $P450_{BM3}$	Yield[b] [%]
P-I263F	V78A, F87V, P142S, T175I, A184V, S226R, H236Q, E252G, I263F, T268A, A290V, L353V, I366V, C400S, T438S, E442K	N.D.
$P411_{BM3}$-H2-A-10	L75A, V78A, F87V, P142S, T175I, L181A, A184V, S226R, H236Q, E252G, T268A, A290V, L353V, I366V, C400S, E442K	N.D.
"P_0-A82L"	V78A, A82L, F87A, P142S, T175I, A184V, S226R, H236Q, E252G, I263F, T268G, A290V, A328V, L353V, I366V, C400S, T438S, E442K.	0.1%
"P_0-A82L" C400H	V78A, A82L, F87A, P142S, T175I, A184V, S226R, H236Q, E252G, I263F, T268G, A290V, A328V, L353V, I366V, C400H, T438S, E442K.	N.D.
"P_0-A82L" A78T	V78T, A82L, F87A, P142S, T175I, A184V, S226R, H236Q, E252G, I263F, T268G, A290V, A328V, L353V, I366V, C400S, T438S, E442K.	0.1%
"P_0-A82L" A78C	V78C, A82L, F87A, P142S, T175I, A184V, S226R, H236Q, E252G, I263F, T268G, A290V, A328V, L353V, I366V, C400S, T438S, E442K.	0.2%
"P_0-A82L" F263Y	V78A, A82L, F87A, P142S, T175I, A184V, S226R, H236Q, E252G, I263Y, T268G, A290V, A328V, L353V, I366V, C400S, T438S, E442K.	0.2%

"P_0-A82L" F263Y was selected as starting point for directed evolution. It contains the following mutations from wild-type cytochrome $P450_{BM3}$: V78A, A82L, F87A, P142S, T175I, A184V, S226R, H236Q, E252G, I263Y, T268G, A290V, A328V, L353V, I366V, C400S, T438S, E442K. This catalyst was then further improved by iterative rounds of site-saturation mutagenesis and screening for C–H insertion activity on substrate **1**. Three beneficial mutations were found: A78L, T327I and A74G (Figure 9).

[b] N.D. indicated not detected

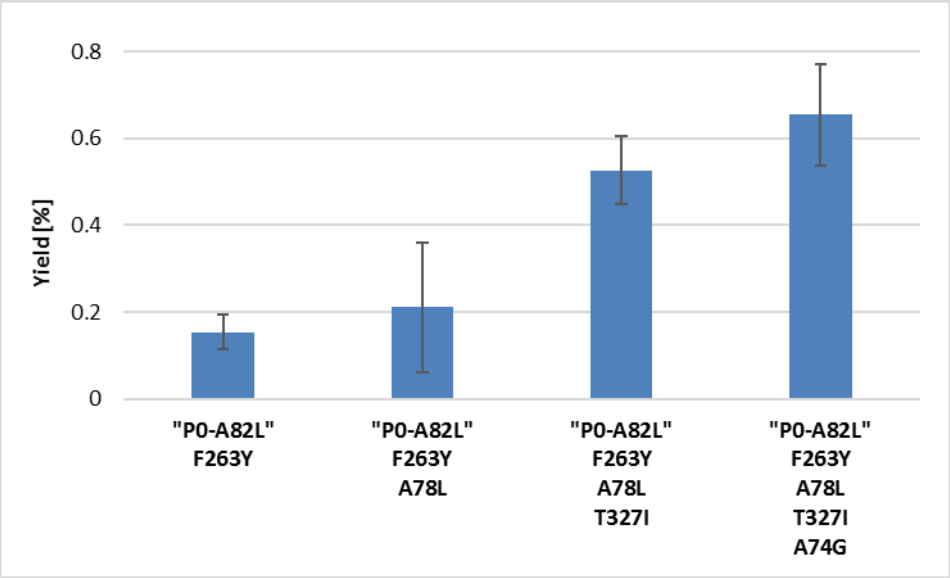
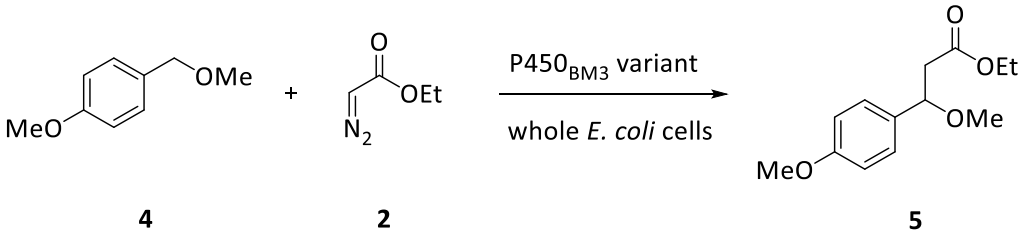


Figure 9: Evolution of "P_0-A82L" for carbenoid C–H insertion

This new variant is now termed "P_1" in this thesis. It has 20 mutations relative to wild-type cytochrome $P450_{BM3}$: A74G, V78L, A82L, F87A, P142S, T175I, A184V, S226R, H236Q, E252G, I263Y, T268G, A290V, T327I, A328V, L353V, I366V, C400S, T438S, E442K.

A second molecule (1-methoxy-4-(methoxymethyl)benzene, **4**) was then identified as a suitable alkane substrate for C–H insertion with ethyl diazoacetate (**2**) (Scheme 2).



Scheme 2: Carbenoid C–H insertion reaction with 1-methoxy-4-(methoxymethyl)benzene and ethyl diazoacetate

In this molecule the benzylic C–H bond is additionally weakened by the adjacent heteroatom which thermodynamically makes it easier to break and insert into the C–H bond. The enzyme showed acceptance for substrate **4** and higher yields were obtained for this reaction.

It was then decided to use the reaction between 1-methoxy-4-(methoxymethyl)benzene (**4**) and EDA (**2**) as the model reaction for screening in the ongoing project.

The aim of my work was now to further enhance variant "P_1" in terms of yield, total turnover number and enantioselectivity.

2.2 Protein Engineering with Directed Evolution Methods

2.2.1 First Round of Site-Saturation Mutagenesis

For the first round of site-saturation mutagenesis we selected three residues in or close to the active site of the parent enzyme "P_1" – H100, F393 and L437. H100 and F393 are both located on the proximal side of the heme cofactor (side of the heme plane where the S400 residue coordinates to the heme; the other side of the heme plane that faces the active site is called "distal"). However, H100 is near the surface of the protein while F393 is close to the heme group and known for having an influence on the redox potential.[45] L437 is distal to the heme and situated in the active site. It already appeared to be a promising site for site-saturation mutagenesis during the first cyclopropanation attempts.[33,37] Farwell *et al.* have reported L437V to be a helpful amino acid change in their aziridination study from 2015.[42]

High-throughput screening gave a few promising results but the rescreen of the potential hits revealed that only L437Q might have been a beneficial mutation. It gave a 1.2-fold enhancement in yield (3.3% yield) relative to "P_1" (2.7% yield) screened at the same time for reaction with the screening substrate **4** (1-methoxy-4-(methoxymethyl)benzene). When examined in biological duplicates it showed that improvement was still within the margin of error, so it was uncertain if L437Q presents a real beneficial mutation (Figure 11). Also, it has to be considered that the overall low yields at this point of the project result in a high error-proneness of the HPLC measurements which increases the margin of error.

When the biotransformation with the new variant was performed with substrate **1** (4-ethylanisole) the results were parent-like (here it was measured on GC due to the even lower yields of this substrate, Figure 12).

However, the mutation did not deteriorate the enzyme's performance so it was considered to be a "neutral" mutation in the worst case.

"P_1" L437Q was decided to be the parent of further engineering anyway.

2.2.2 Truncating the Enzyme

As described in chapter 1.4.1 $P450_{BM3}$ variants contain three parts: the heme domain (amino acids 1–470) and the reductase domain which consists of the FMN and the FAD domain (471–1048). Since we assume that the reaction is taking place in the heme domain, we wanted to examine if truncating the enzymes has any effect on the catalytic activity.

We did experiments with just the heme domain as well as taking amino acids 1–658 and 1–664 (both remove the FAD domain) and also tested if adding exogenous reducing agent (dithionite, ascorbic acid; 1–10 mM) was required (Table 2). The reactions were performed with substrate **4** *in vivo* (Scheme 2).

It showed that the heme domain alone is capable of effecting the biotransformation when dithionite is used as reductant. This emphasized our hypothesis that the chemistry happens at the heme domain.

Also both variants, where the FAD domain was removed, gave an active catalyst. To our pleasant surprise the reaction was effected even without any additional reducing agent. The cell's own reducing conditions seemed sufficient to enable the catalytic cycle.

"P_1" L437Q-FMN665[c], where the protein is truncated at amino acid residue 665, was taken to the next round of site-saturation mutagenesis. We believe that reducing the metabolic loading by using a smaller enzyme can be of advantage and possibly result in better expression at some point.

[c] FMN665 is from now on added to the name to clarify that we are working with the truncated protein

Table 2: Effect of truncating the biocatalyst on intermolecular C-H insertion. Highlights are marked in bold.

Variant	Amino acid composition	Reducing conditions	Yield [%]	TTN[d]
"P_1" L437Q	1–1048 (holo protein)	no reductant	4.87	149
"P_1" L437Q-heme	1–470 (heme domain)	no reductant	1.44	24
		1 mM dithionite	**5.04**	82
		10 mM dithionite	3.42	56
		1mM ascorbic acid	1.06	17
		10 mM ascorbic acid	0.87	14
"P_1" L437Q-FMN659	1–658 (heme + FMN domain)	**no reductant**	**5.55**	185
		1 mM dithionite	5.33	178
		10 mM dithionite	3.32	111
		1mM ascorbic acid	4.76	159
		10 mM ascorbic acid	3.77	126
"P_1" L437Q-FMN665	1–664 (heme + FMN domain)	**no reductant**	**5.74**	187
		1 mM dithionite	5.19	169
		10 mM dithionite	3.18	103
		1mM ascorbic acid	4.80	156
		10 mM ascorbic acid	4.85	158

2.2.3 Second Round of Site-Saturation Mutagenesis

Four site-saturation libraries were created with "P_1" L437Q-FMN665 as the parent enzyme. All four targeted sites are located in the active site of the protein: L78, A87, Y263 and S438[e]. Even though the enzyme already contained mutations at these sites compared to wild-type $P450_{BM3}$, we still chose these residues for mutagenesis because those amino acids might not be most optimal for the carbenoid C–H insertion. The mutations to A87, Y263 and S438 originate from cyclopropanation[33,37] and amination projects[40–42] and might not be the most helpful residues for the substrates and chemistry of the reaction studied in this work. Position 78 has been mutated before in the beginning of this project (chapter 2.1) but sometimes beneficial amino acid changes only occur in association with other new mutations. It can

[d] TTN = total turnover number

[e] amino acids are labeled with the residues of the parent enzyme "P_1" L437Q FMN665; they already contain mutations relative to wild-type $P450_{BM3}$: V78L, F87A, I263Y, T438S

therefore be of advantage to return to an already targeted site in a later stage of the engineering process.

A potential hit in the mutagenesis library of L78 turned out to be a silent mutation (different codon but still encoding for the same amino acid) and only resulted in a slightly better expression of the protein.

A real improvement in yield was found with the mutation S438T. We obtained a 3-fold enhancement in yield (12%, 285 TTN) for substrate **4** and were able to measure enantiomeric excess (ee) via SFC for the first time (9% ee, Figure 11). For the less reactive substrate **1** (4-ethylanisole) a yield of 2.6% (62 TTN) was measured (Figure 12).

It is remarkable that the residue at 438 was changed back to its wild-type amino acid. While serine seemed to be the best residue at this position for cyclopropanation[33] and amination[42,40] reactions, for intermolecular carbenoid C–H insertion the variant with threonine gives better yields.

2.2.4 Random Mutagenesis with Error-Prone PCR

Site-saturation mutagenesis is a good approach for investigating specific amino acid residues. When trying to adapt a biocatalyst for a new non-natural reaction, you usually do not know how the substrates find their way into the enzyme and how they are bound in the active site. That can make it difficult to guess which amino acids are good targets for substitution. With error-prone PCR (epPCR) you have the chance to find beneficial mutations all over the protein and at positions you might not have thought about before.

We carried out error-prone PCR with *Taq* polymerase and with two different $MnCl_2$ concentrations (300 µM– 400 µM) to obtain various error rates (average number of mutations per gene). Too many mutations could destabilize and flaw the protein, while too little mutations decrease the probability of effecting promising amino acid changes.

Also, we performed the random mutagenesis only at the DNA fragment encoding for the heme domain, since the transformation is believed to happen there. As parent we chose the variant from the last site-saturation mutagenesis "P_1" L437Q-FMN665 S438T.

As usual, screening was done in 96-well plates. We picked 88 variants from each error-prone library to be able to generate a retention of function (ROF) curve (Figure 10). We also added three parents, three sterile controls and two negative controls (halohydrin dehalogenase; enzyme without a heme group) to each plate.

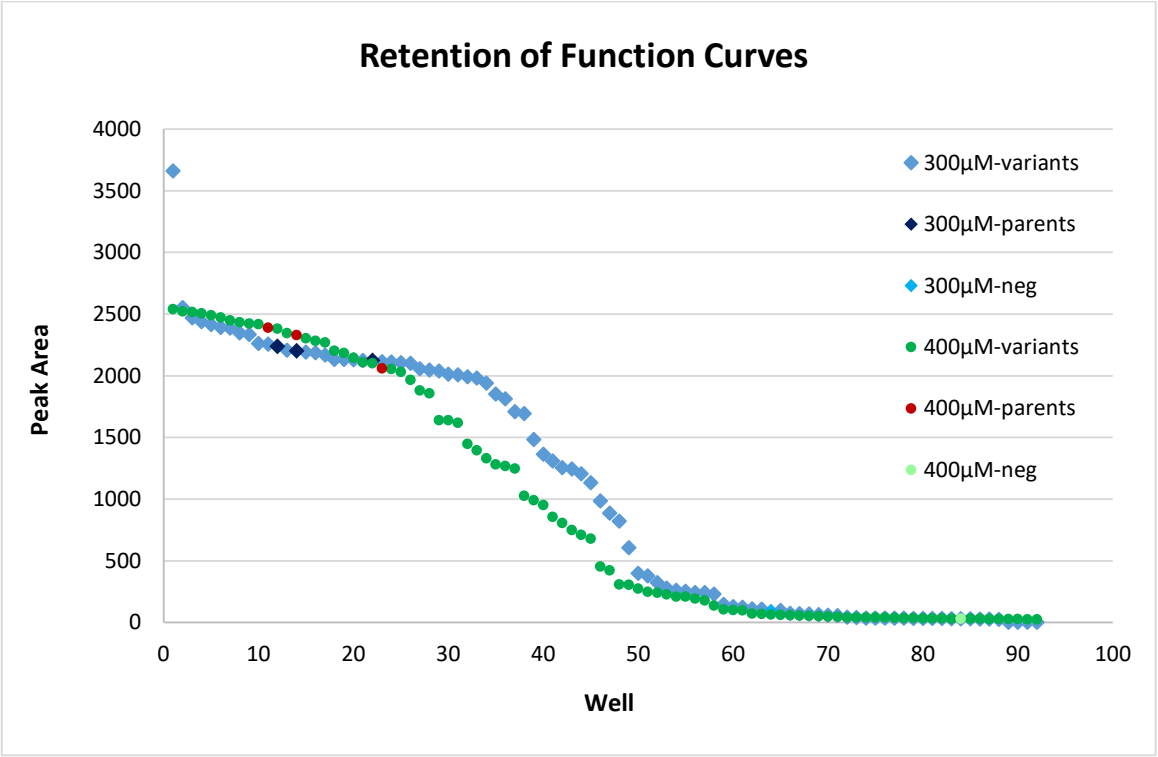
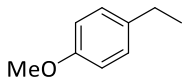
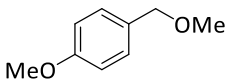


Figure 10: ROF curves of 300 µM (blue, parents in dark blue) and 400 µM (green, parents in red) $MnCl_2$ plate.

The ROF curves of both libraries looked promising, so more variants of each library were picked and screened (6 more plates of 300 µM $MnCl_2$ library, 2 more plates of 400 µM $MnCl_2$ library).

The variant with the largest peak area (blue data point at ~3600 in Figure 10) in the retention of function curve of the 300 µM $MnCl_2$ library was validated as a 1.2-fold improvement in yield (15%, 290 TTN) with two mutations: P248P (silent) and F331I. Screening of more plates identified a few more mutations with different effect on the yields of the two substrates **1** and **4**. The results are summarized in Table 3.

Table 3: Yields and TTN of P411 variants obtained via epPCR. Highlights are marked in bold.

Variant	Substrate 1 		Substrate 4 	
	Yield [%]	TTN	Yield [%]	TTN
"P_1" L437Q- FMN665 S438T (Parent)	2.6	62	12.0	285
"P_1" L437Q- FMN665 S438T **P248P(sil) F331I**	2.9	52	13.5	246
"P_1" L437Q- FMN665 S438T **F162L**	**3.7**	**125**	14.0	478
"P_1" L437Q- FMN665 S438T **D63N V178A**	2.6	56	**16.9**	**373**
"P_1" L437Q- FMN665 S438T **M118T T365A**	2.9	71	15.2	371

2.2.5 And More Site-Saturation Mutagenesis

In parallel to the time-consuming screening of the error-prone libraries we conducted another round of site-saturation mutagenesis with variant "P_1" L437Q-FMN665 S438T as parent where we targeted six sites: L82[f], L181, L188, A330, F331 and T436.

Amino acid residue 82 and 181 have been proven to be promising sites for mutations in previous projects (moreover, L181 is located near mutation L437Q in the active site).[36,41] F331 was chosen because it was one of the beneficial mutations discovered in the epPCR approach (Table 3). Since L437 and S438 have shown the best results in the previous experiments, T436 was selected for this round. The other residues were picked by looking at the crystal structure (PDB: 4WG2) to find out which amino acids are in the active site.

Amino acid change T436L gave a boost in yield (29%, 760 TTN) and also enhanced the enantiomeric excess to 20% for 1-methoxy-4-(methoxymethyl)benzene (**4**, Figure 11). For the other substrate (**1**) we reached 7.2% yield (186 TTN) and 27% ee (Figure 12).

Since this was the better improvement (2.5-fold) compared to the error-prone results (Table 3), we decided to continue engineering on variant "P_1" L437Q-FMN665 S438T T436L, but to keep the results of the random mutagenesis in mind for potential new residues to target.

[f] 82 already contains the following mutation from wild-type $P450_{BM3}$: A82L

For the final site-saturation mutagenesis libraries we created while I worked on this project we again altered six sites: L52, D63, F162, M177, V178, L439.

D63, F162 and V178 were selected due to the error-prone results (Table 3), the others because they showed proximity to those sites in the crystal structure (PDB: 4WG2).

No mutation was found that increased the yield significantly, but M177L gave a variant with better enantiomeric excess for both substrates (42% ee for substrate **1**, 31% ee for substrate **4**).

The best biocatalyst for carbenoid C–H insertion at this stage was "P_1" L437Q-FMN665 S438T T436L M177L with 22 mutations relative to wild-type $P450_{BM3}$: A74G, V78L, A82L, F87A, P142S, T175I, M177L, A184V, S226R, H236Q, E252G, I263Y, T268G, A290V, T327I, A328V, L353V, I366V, C400S, T436L, L437Q, E442K.

This variant is now referred to as "M177L".

All the results of directed evolution of the biocatalyst are illustrated in Figure 11 and Figure 12.

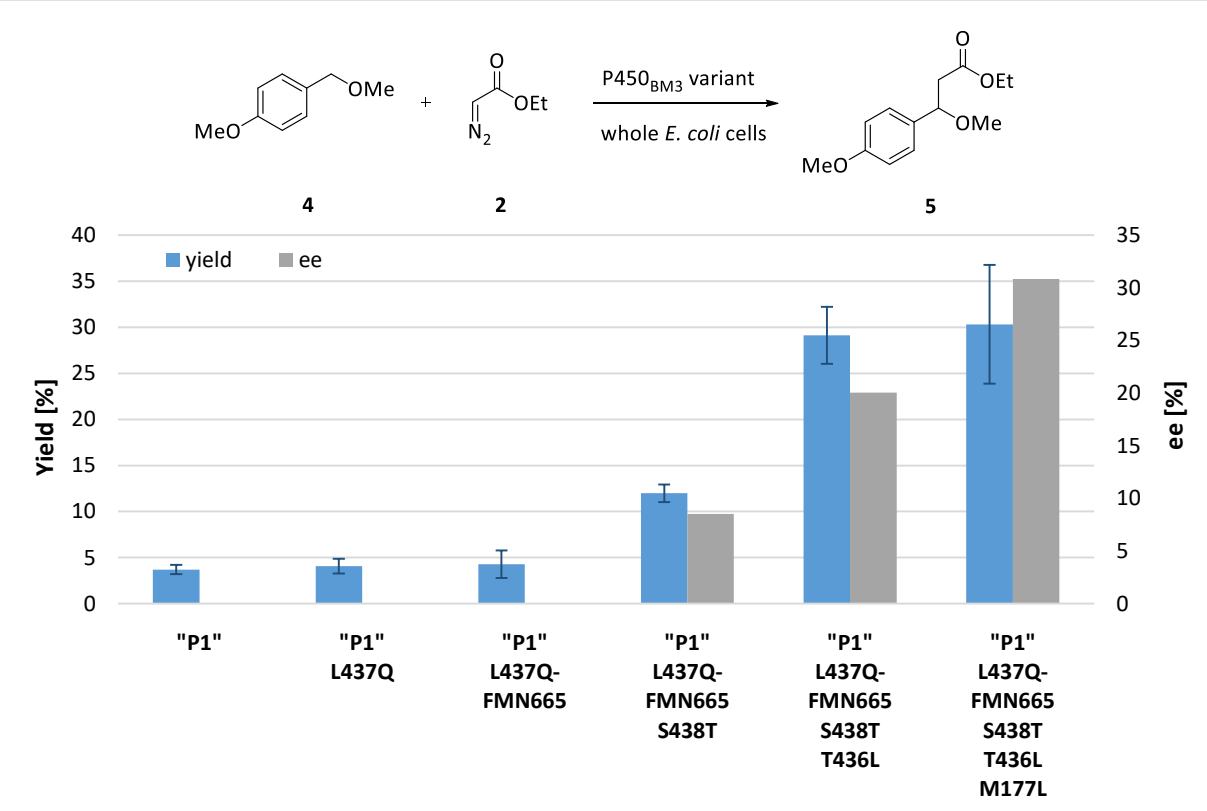


Figure 11: Evolution of a biocatalyst for intermolecular carbenoid C-H insertion of 1-methoxy-4-(methoxymethyl)benzene (4) and ethyl diazoacetate (2)

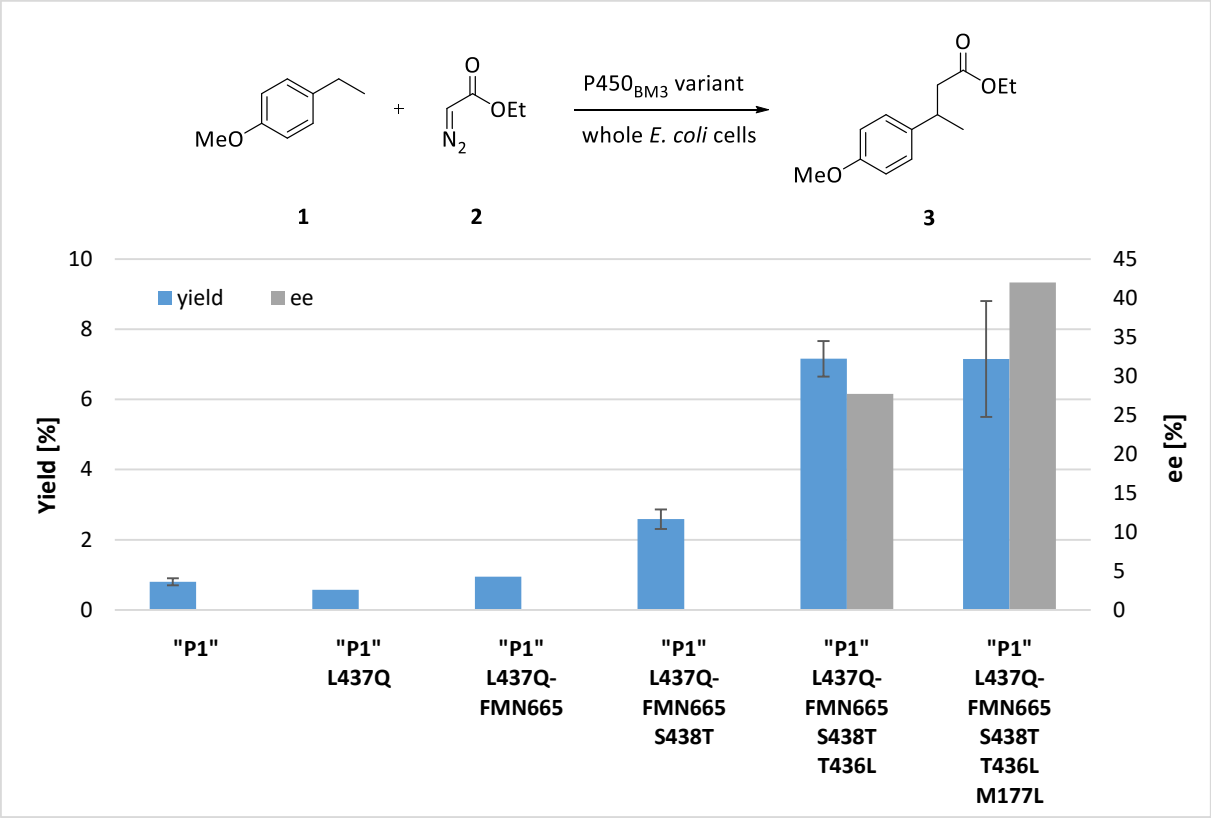


Figure 12: Evolution of a biocatalyst for intermolecular carbenoid C-H insertion of 4-ethylanisole (1) and ethyl diazoacetate (2)

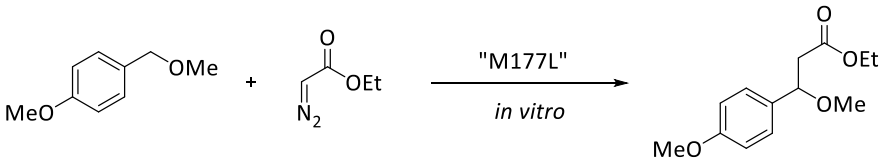
2.3 Determination of the Initial Rate with Purified Enzymes

We wanted to investigate if the introduced mutations enhanced the reaction rate or if the improvements in yield are effected by other factors (*e.g.* stability of the protein, access of substrate to the active site).

For that purpose, we isolated two variants: the best biocatalyst to this point, “M177L”, and for comparison “P_1” L437Q-FMN665, the first truncated enzyme of this study. The proteins were purified by immobilized metal affinity chromatography, where the enzymes coordinate to nickel ions in the column via a 6xHis-tag and are later eluted with an imidazole buffer.

The suitable reducing conditions for the biotransformation *in vitro* were determined by overnight experiments with variant “M177L” on screening substrate (**4**).

Table 4: Determination of reaction conditions for in vitro biotransformation. Highlights are marked in bold. Carbenoid C–H insertion performed with purified "M177L" variant. Reactions were performed with purified protein (0.1 mol%) in 0.1 M potassium phosphate buffer (pH=8.0) under anaerobic conditions. Substrate concentrations were 10 mM alkane and 10 mM ethyl diazoacetate; substrates were added to the reaction as solutions in ethanol, giving 5% ethanol co-solvent.



4 **2** **5**

Variant	Reducing conditions	Yield [%]	TTN
"M177L"	no reductant	6.4	54
	1 mM NADPH	17.8	149
	10 mM NADPH	**33.5**	**281**
	1 mM dithionite	18.4	154
	5 mM dithionite	25.1	211

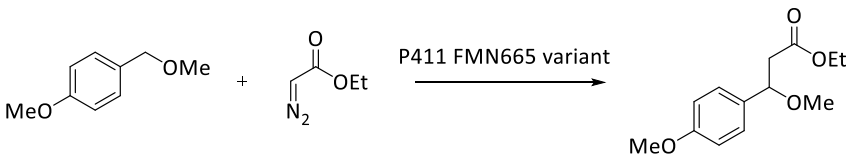
The reactions performed with 10 mM NADPH gave the best results (Table 4) and therefore, those reducing conditions were applied in the following experiments.

At first we wanted to ascertain if the results of the reactions performed *in vitro* are comparable to those obtained with biotransformation in whole cells. To this end, overnight reactions with both purified variants were carried out and compared to the results collected during evolution of the biocatalyst (Table 5).

Especially, variant "M177L" showed similar behavior both *in vivo* and in isolated form. The yield is around 30% as well as the enantiomeric excess.

Figure 13 depicts the performance of the two variants *in vitro*.

Table 5: Comparison of transformation in vivo and vitro. In vitro reactions were performed with purified protein (0.1 mol%) in 0.1 M potassium phosphate buffer (pH=8.0) containing 10 mM NADPH under anaerobic conditions. Substrate concentrations were 10 mM alkane and 10 mM ethyl diazoacetate; substrates were added to the reaction as solutions in ethanol, giving 5% ethanol co-solvent. Data for in vivo reactions was taken from the results obtained during evolution.



4 **2** **5**

Variant	*in vivo*			*in vitro*		
	Yield [%]	**TTN**	**ee [%]**	**Yield [%]**	**TTN**	**ee [%]**
"P_1" 437Q-FMN665	4.3	136	N.D.	2.9	29	N.D.
"M177L"	30.3	583	30.9	28.6	267	30

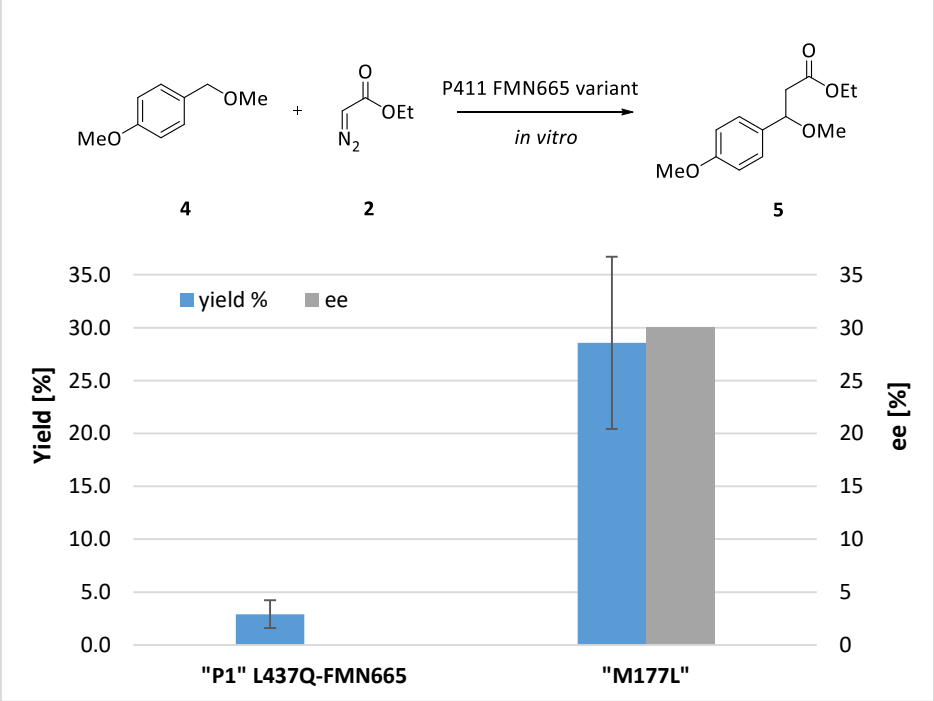


Figure 13: Intermolecular carbenoid C–H insertion in vitro. Reactions were performed as in Table 5.

For the determination of the initial rate, transformations were set up the same way as for the overnight reactions, but the reactions were aborted at definite time points (2, 4, 6, 8 and 10 minutes). The initial rate can be calculated from the slope of the linear regression through the data points (Figure 14). Normalizing to the concentration of the enzyme (0.01 mM) gives the initial reaction rate in [min^{-1}] (Table 6).

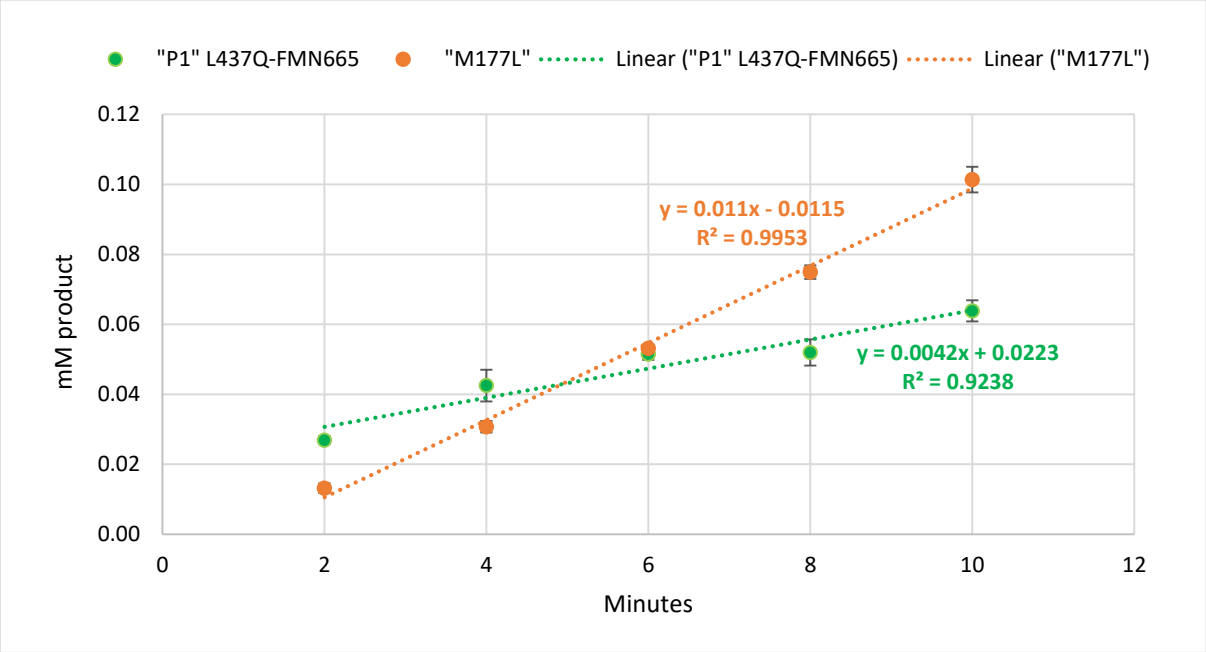


Figure 14: Determination of the initial rate of P411 variants "P_1" L437Q-FMN665 and "M177L" for the reaction between 1methoxy-4-(methoxymethyl)benzene (4) and ethyl diazoacetate (2). Reactions were performed as in Table 5.

Table 6: Initial rates of "P_1" L437Q-FMN665 and "M177L" for the reaction between 1methoxy-4-(methoxymethyl)benzene (4) and ethyl diazoacetate (2). Reactions were performed as in Table 5.

Variant	Initial rates [min^{-1}]
"P_1" L437Q-FMN665	0.42
"M177L"	1.1

The enhancement in initial rate is only 2.6-fold. It cannot be the only reason for the improvement in yield. It is possible, that the beneficial mutations have increased the enzyme's stability and therefore caused a longer lifetime, or that the enzyme is now more likely to catalyze the desired intermolecular carbenoid C–H insertion instead of the side reaction (forming the EDA dimers).

3 Summary and Outlook

Directed Evolution of cytochrome $P450_{BM3}$ resulted in an enzyme that is able to perform a non-natural intermolecular carbenoid insertion into a C–H bond *in vivo* and *in vitro*. As a carbene precursor ethyl diazoacetate – an acceptor-only precursor – was utilized.

The best biocatalyst at this stage was “P411” variant "P_1" L437Q-FMN665 S438T T436L M177L with 22 mutations relative to wild-type $P450_{BM3}$: A74G, V78L, A82L, F87A, P142S, T175I, M177L, A184V, S226R, H236Q, E252G, I263Y, T268G, A290V, T327I, A328V, L353V, I366V, C400S, T436L, L437Q, E442K.

Two substrates have been tested so far with this biocatalyst. For 1-methoxy-4-(methoxymethyl)benzene (**4**) a yield of 30% (580 TTN *in vivo*) was achieved, and also an enantiomeric excess of 30%. The less reactive substrate 4-ethylanisole (**1**) reached 7.2% yield (186 TTN) and 42% enantiomeric excess.

Additional protein engineering will be carried out with this variant to further improve the yield as well as the enantioselectivity.

Also, we are interested in expanding the substrate scope for both alkane substrate and carbene precursor to expand the possibilities of synthesizing new molecules.

4 Experimental

4.1 DNA Cloning

pET22b(+) was used as a cloning and expression vector in these experiments. The gene encoding for the P450 variant studied in this work has already been inserted into the vector in previous projects.

The vector carries genes encoding for ß-lactamase to provide ampicillin resistance, the lac operon to ensure expression in the host organism *E. coli* as well as a 6x-His tag for purification.

4.1.1 Preparation of Electro-Competent Cells

5 mL lysogeny broth (LB) were inoculated with cells of the *E. coli* strain BL21(DE3) and incubated overnight at 37°C and 230 rpm. 1 mL of this starter culture was added to 250 mL 2xYT medium in a 1L Erlenmeyer and grown at 37°C and 230 rpm until an OD_{600} of 0.4 – 0.6 was reached (after about 1.5 h). The cells as well as centrifuge buckets were cooled on ice for 30 min. The suspension was centrifuged for 15 min at 4000*g* and 4°C. The supernatant was discarded and the cells washed twice with cold, sterile dH_2O and once with 20 mL cold 10% glycerol. Finally, the cells were resuspended in 2 mL 10% glycerol, aliquoted into Eppendorf tubes (50 µL each) and stored at -80°C.

4.1.2 Site-Saturation Mutagenesis

In this method a polymerase chain reaction (PCR) was performed to introduce site-specific mutations into the gene encoding for the heme domain in cytochrome P411 enzymes. Primers were used that contained base exchanges at the specific site and generated the mutations in the base sequence during amplification. The "22-codon-trick"[21] was applied to overcome the drawbacks of codon redundancy in site-saturation experiments. In this procedure a mixture of three forward primers is used which in combination reduce the codon to amino acid ratio to 22:20 (in comparison to the ratio of 64:20 appearing in nature). Two of the primers carry degenerated codons (NDT – 12 distinct codons; VHG – 9 distinct codons)[g], while the third one (TGG) encodes for a single amino acid.

[g] N = A/C/T/G; D = no C; V = no T; H = no G

This approach amplifies the whole vector. The reverse primer was designed in such a way that it contains overlapping regions with the forward primers. The linear PCR products can then be ligated to form circular vectors again.

The reaction was carried out in 100 µL PCR tubes and contained the components listed in Table 7.

Table 7: Protocol for site-saturation mutagenesis PCR

Component	**Volume [µL]**	**Final concentration**	**Thermocycler program**		
ddH_2O	30.5		98°C	30 s	
Phusion buffer 5x	10	1x	98°C	10 s	30x
DNA template	2	~0.8 µg	X°C[h]	15 s	
dNTP mix (10 mM each)	1	200 µM	72°C	2 min 30 s	
Phusion Polymerase (2000 U/ml)	0.5	1 U	72°C	10 min	
Forward primer mix (10 µM each)[i]	3	0.6 µM	10°C	∞	
everse primer (10 µM)	3	0.6 µM			
DMSO	1	2%			

4.1.3 Random Mutagenesis

A different approach to introduce mutations into the gene of an enzyme is error prone PCR (epPCR). In this method a polymerase is used that does not work as accurate as a high-fidelity polymerase and therefor causes errors in the DNA sequence during amplification. Also Mn^{2+} ions are added to increase the probability of errors.

To ensure mutations only occur in the heme domain of the gene, only the relevant sequence was amplified during epPCR.

Several PCRs with different Mn^{2+} concentrations were performed to obtain different mutation rates.

[h] choose X°C so that it is 3-5°C below the lowest primer T_m of the set of primers
[i] NDT:VHG:TGG = 12:9:1

Table 8: Protocol for epPCR

Component	Volume [µL]	Final concentration	Thermocycler program		
ddH_2O	to 100 µl		95°C	30 s	
Taq buffer 5x	10	1x	95°C	20 s	30x
DNA template	1	~0.8 µg	45°C	20 s	
dNTP mix (10 mM each)	4	400 µM	68°C	1 min 30 s	
$MnCl_2$ (1 mM)	30 – 40	300 – 400 µM	68°C	5 min	
Heme forward primer (10 µM)	2	0.2 µM	10°C	∞	
Heme reverse primer (10 µM)	2	0.2 µM			
Taq Polymerase (5 U/µL)	1.6	8 U			

4.1.4 High-Fidelity PCR

High-fidelity PCR was either performed to amplify the expression vector pET-22b(+) backbone or to cut out desired regions of the studied protein.

Table 9: Protocol for high-fidelity PCR

Component	Volume [µL]	Final concentration	Thermocycler program		
ddH_2O	30.5		98°C	30 s	
Phusion buffer 5x	10	1x	98°C	10 s	30x
DNA template	2	~0.8 µg	57°C	15 s	
dNTP mix (10 mM each)	1	200 µM	72°C	30 s per kb	
Phusion Polymerase (2000 U/ml)	0.5	1 U	72°C	5 min	
Forward primer (10 µM)	3	0.6 µM	10°C	∞	
Reverse primer (10 µM)	3	0.6 µM			
DMSO	1	2%			

4.1.5 Creation of an Expression Construct

After PCR, *Dpn*I was added to the samples (0.8 U/µL) and incubated for 2 h at 37°C to digest the template DNA. *Dpn*I is a restriction enzyme that recognizes and cuts the sequence GATC when A is methylated. This is only the case for the parental strain, while the PCR products remain untouched.

The DNA fragments obtained during PCR were then purified via gel electrophoresis using 1% agarose gel. The agarose solution was prepared in batch (200 – 300 mL) and stored at 60°C. When needed the solution was poured into an electrophoresis chamber and after 30 minutes it had solidified to form the gel. Loading dye containing the fluorescent stain SYBR® Gold was added into the PCR tubes. The gel was loaded with the samples as well as with a 1-Kb DNA ladder and run at 105 – 130 V for 20 – 30 min. The DNA bands were visualized by UV-light and excised with a razor blade. The DNA was extracted using the QIAGEN® QIAquick Gel Extraction Kit.

The Gibson Assembly procedure was applied to connect the insert with the vector backbone or – in case of the site-saturation mutagenesis products – to rebuild the circular form. 2.5 µL DNA (ratio of insert : vector = 1 : 2, if applied) were added to 7 µL Gibson Mix and incubated in a thermomixer for 1h at 50°C. The product could directly be transformed into electro-competent cells.

4.1.6 Transformation of Electro-Competent Cells

50 µL-aliquots of electro-competent BL21(DE3) *E. coli* cells was thawed on ice and 1.5 µL Gibson-assembled DNA was added. The mix was transferred to a cooled 2 mm electroporation cuvette and an electro shock applied using a Gene Pulser Xcell™ electroporation system. The parameters were set to 2.5 kB, 25 µF and 200 Ω.

The shocked cells were immediately recovered with 600 mL SOC medium and incubated at 37°C at 230 rpm for 30 – 45 min.

After that, the transformants were plated out on LB-amp (0.1 mg/mL ampicillin) agar plates (volume and concentration dependent on quality of the cells).

4.2 Library Screening

4.2.1 High-Throughput Reaction Screening in 96-well Plate Format

Libraries have been generated either by site-saturation mutagenesis or error prone PCR and plated out on LB-amp (0.1 mg/mL ampicillin) agar plates to obtain single colonies.

For screening the single colonies were picked and used to inoculate 300 µL LB medium with 0.1 mg/mL ampicillin in 96-deep-well plates. Three wells were inoculated with cells expressing the parent enzyme, another three wells were no-inoculation sterility controls.

The plates were covered with a breathable film and it was grown overnight at 37°C, 220 rpm and 80% humidity.

50 µL of overnight culture was transferred to 1 ml expression medium (Hyper Broth + 0.1 mg/mL ampicillin) in new 96-deep-well plates using a multichannel pipette. The cultures were incubated at 37°C, 220 rpm and 80% humidity for 3 h.

The cells were cooled in an ice bath for 20 min, followed by the induction of expression by adding IPTG to a final concentration of 0.5 mM and 5-aminolevulinic acid to a final concentration of 1.0 mM. Expression was allowed overnight (18 – 20 h) at 20°C and 150 rpm.

The cells were pelleted by spinning the plates at 3000*g* for 5min and resuspended in 20 µL M9-N medium. 20 µL/well GOX solution (14 000 U/mL catalase, 1 000 U/mL glucose oxidase) was added and the plates transferred to an anaerobic chamber.

In the chamber 24 mL argon sparged M9-N/glucose mix (M9-N : 250 mM glucose in M9-N = 6.5 : 1,) were combined with 800 µL 1-methoxy-4-(methoxymethyl)benzene (400 mM in EtOH, final concentration 10 mM) and 800 µL ethyl diazoacetate (400 mM in EtOH, final concentration 10 mM). 320 µL of this solution was pipetted into each well.

The plates were sealed with aluminum foil and shaken in the chamber for 5 – 6 h.

Finally, the seal was removed and 400 µL/well acetonitrile was added and mixed by pipetting up and down with a multichannel pipette. The plates were centrifuged (3000*g*, 5 min) and 250 µL of the supernatant were filtered through 0.2 µm plate filters into shallow-well plates for analysis by HPLC.

4.2.2 Typical Procedure for Small-Scale Bioconversion Under Anaerobic Conditions Using Whole Cells

Variants that showed improved properties in the high-throughput screening were rescreened in duplicates, together with the parent of the library.

The corresponding cells were taken from the pre-culture plate and streaked out on new LB-amp plates. A single colony was picked und grown in 5 mL LB with 0.1 mg/mL ampicillin. The whole starter culture was poured into 45 mL expression medium (Hyper Broth + 0.1 mg/mL ampicillin) in a 125 mL Erlenmeyer flask and incubated for 2.5 h at 37°C and 220 rpm.

After, the flasks were put in ice for 30 min before expression was induced by adding 50 µL of 0.5 M IPTG (final concentration 0.5 mM) and 50 µL of 1.0 M 5-aminolevulinic acid (final concentration 1.0 mM).

The flasks were returned to shaking at 22°C and 130 rpm overnight (18 – 20 h).

The cells were then pelleted (2500g, 5 min) and resuspended to OD_{600} = 45 in M9-N. An aliquot of the cell suspension (2 mL) was centrifuged (2500g, 5 min) and the pellet stored for later determination of the enzyme expression level.

The rest of the cell suspensions was transferred to sealed 6 mL crimp vials and degassed with argon for 30 min as well as 250 mM glucose in M9-N. To a 2 mL crimp vial was added 20 µL of the beforehand mentioned GOX solution and after that everything was moved into the anaerobic chamber.

The components were then added to the 2 mL crimp vial in the following order: 40 µL glucose (250 mM in M9-N, final concentration 25 mM), 320 µL cell suspension, 10 µL alkane substrate (400 mM in EtOH, final concentration 10 mM) and 10 µL ethyl diazoacetate (400 mM in EtOH, final concentration 10 mM).

The crimped vials were moved out of the chamber and put on a shaker at room temperature and 40 rpm for 5 – 6 h.

The work-up procedure was dependent on the used analysis method.

For measuring on GC or GC/MS 10 µL 20 mM trimethoxybenzene (in cyclohexane, final concentration 0.5 mM) was added to the reaction vial. 500 µL of a 1:1 mixture of cyclohexane and ethyl acetate were added as well and the suspension pipetted into an Eppendorf tube. It was vortexed, centrifuged at 14 000 rpm for 5 min and the upper organic layer transferred into a 2 mL screw vial.

HPLC samples were prepared by adding 10 µL 50 mM trimethoxybenzene (in DMSO, final concentration 1.2 mM) to the reaction vial and quenching the reaction with 400 µL acetonitrile. The mixture was transferred to an Eppendorf tube, vortexed and centrifuged (14 000 rpm, 8 min) and the supernatant pipetted into a new 2 mL screw vial.

4.3 *In Vitro* Reactions

4.3.1 Protein Expression and Purification

For purification proteins had to be expressed in larger scale. A 19 mL LB-pre-culture (with 0.1 mg/mL ampicillin) was inoculated with a scratch from a glycerol stock and grown overnight at 37°C and 220 rpm.

The entire starter culture was poured into 450 mL expression medium (Hyper Broth + 0.1 mg/mL ampicillin) in a 1 L Erlenmeyer flask and incubation allowed for 2.5 h at 37°C and 220 rpm. Expression was induced by adding IPTG (final concentration 0.5 mM) and 5-aminolevulinic acid (final concentration 1.0 mM). It was shaken at 22°C and 130 rpm overnight.

The next day the cells were pelleted (5000*g*, 8 min) and stored at -20°C for at least 24h.

The frozen cells were resuspended in His-trap buffer A (4 ml/g wet cell weight) and hemin (30 mg/mL stock solution in 0.1 M NaOH) was added to a final concentration of 1 mg/g wet cell weight as well as some DNase I.

The cells were lysed by sonication (1 min protocol: 1 s off, 1 s on, 30% amplitude) and the suspension centrifuged at 20 000*g* for 20 min at 4°C to remove insoluble material.

The enzymes were purified using a nickel NTA column in an AKTA fast protein liquid chromatography (FPLC) system for affinity chromatography. Proteins were then eluted on a linear gradient from His-trap buffer A to His-trap buffer B (P450 enzymes elute at around 80 mM imidazole). and fractions with high protein concentration were combined and concentrated.

The solution was submitted to three rounds of buffer exchange to 0.1 M KPi pH 8.0 buffer using 10 kDa centrifugal filter units.

The enzymes were aliquoted into PCR tubes, flash-frozen on dry ice and stored at -80°C.

4.3.2 Typical Procedure for Small-Scale Bioconversion Under Anaerobic Conditions Using Purified Enzymes

Reactions with purified enzymes were carried out in 2 mL crimp vials and set up in the anaerobic chamber similar to the whole cell procedure.

NADPH had to be added as reductant and has already been pre-mixed with the reaction buffer (0.1 M KPi pH 8.0) outside the chamber.

20 µL GOX was pipette into the reaction vials still outside the anaerobic chamber, then the following components were added under anaerobic conditions in the order presented: 40 µL glucose (250 mM in 0.1 M KPi pH 8.0, final concentration 25 mM), 300 µL NADPH in reaction buffer (13.3 mM in 0.1 M KPi pH 8.0, final concentration 10 mM), 20 µL protein stock solution (200 µM in 0.1 M KPi pH 8.0, final concentration 10 µM).

The mixture was shaken for 5 min to ensure even distribution of the proteins. Finally, 20 µL of a 1:1 mix of the two substrates (each 400 mM in EtOH, final concentration 10 mM) were added and the vials were sealed.

Reaction was allowed to go overnight at room temperature and 50 rpm.

Work-up was performed according to the procedure described in 0.

4.4 Determination of Enzyme Concentration

Protein concentration of the purified enzymes as well as expression in whole cells were both determined via hemochrome assay.

2 mL aliquots of the OD_{600} = 45 cell suspension for whole cell reaction have been pelleted and frozen at -20°C. The cells were resuspended in 4 mL M9-N buffer and lysed by sonication (1 min protocol: 1 s off, 1 s on, 30% amplitude). The lysate was transferred to Eppendorf tubes and centrifuged (14 000 rpm, 10 min, 4°C) to remove cell debris.

To 735 µL of the lysate 75 µL 1 M NaOH were pipetted and incubated for 5 min. 175 µL pyridine were added and mixed by pipetting up and down. After addition of 15 µL dithionate solution (300 mM in dH_2O) it was incubated for another 10 min. Samples were measured by using UV/Vis. Heme concentrations were determined by using the reported extinction coefficient (ε = 191500 L mol $^{-1}$ cm^{-1})[46].

5 References

1. Davies, H. M. L. & Manning, J. R. Catalytic C-H functionalization by metal carbenoid and nitrenoid insertion. *Nature* **451,** 417–24 (2008).
2. 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).
3. Wu, W., Yang, Z. & You, S. in *Asymmetric Functionalization of C-H Bonds* (ed. You, S.) 1–66 (Royal Society of Chemistry, 2015). doi:10.1039/9781782621966-00001
4. Doyle, M. P., Duffy, R., Ratnikov, M. & Zhou, L. Catalytic Carbene Insertion into C - H Bonds. *Chem. Rev.* 704–724 (2010). doi:10.1021/cr900239n
5. Davies, H. M. L. & Beckwith, R. E. J. Catalytic enantioselective C-H activation by means of metal-carbenoid-induced C-H insertion. *Chem. Rev.* **103,** 2861–2903 (2003).
6. Davies, H. M. L. Recent advances in catalytic enantioselective intermolecular C-H functionalization. *Angew. Chemie - Int. Ed.* **45,** 6422–6425 (2006).
7. Paulissen, R., Hayez, E., Hubert, A. J. & Teyssie, P. Transition metal catalysed reactions of diazocompounds - part III a one-step synthesis of substituted furanes and esters. *Tetrahedron Lett.* **15,** 607–608 (1974).
8. Mbuvi, H. M. & Woo, L. K. Catalytic C-H insertions using iron(III) porphyrin complexes. *Organometallics* **27,** 637–645 (2008).
9. Rosenthaler, L. Durch Enzyme bewirkte asymmetrische Synthese. *Biochem. Z.* **14,** 238–253 (1908).
10. Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485,** 185–194 (2012).
11. Ricca, E., Brucher, B. & Schrittwieser, J. H. Multi-Enzymatic Cascade Reactions: Overview and Perspectives. *Adv. Synth. Catal.* **353,** 2239–2262 (2011).
12. Faber, K. *Biotransformations in Organic Chemistry*. (Springer-Verlag Berlin Heidelberg, 2011). doi:10.1007/978-3-642-17393-6
13. Loughlin, W. A. Biotransformations in organic synthesis. *Bioresour. Technol.* **74,** 49–62 (2000).
14. Simon, R. C., Richter, N., Busto, E. & Kroutil, W. Recent developments of cascade reactions involving omega-transaminases. *ACS Catal.* **4,** 129–143 (2014).

15. Lopez-Gallego, F. & Schmidt-Dannert, C. Multi-enzymatic synthesis. *Curr. Opin. Chem. Biol.* **14,** 174–183 (2010).
16. Kuchner, O. & Arnold, F. H. Directed evolution of enzyme catalysts. *Trends Biotechnol.* **15,** 523–530 (1997).
17. Renata, H., Wang, Z. J. & Arnold, F. H. Expanding the enzyme universe: Accessing non-natural reactions by mechanism-guided directed evolution. *Angew. Chemie - Int. Ed.* **54,** 3351–3367 (2015).
18. Cline, J., Braman, J. C. & Hogrefe, H. H. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res.* **24,** 3546–3551 (1996).
19. Currin, A., Swainston, N., Day, P. J. & Kell, D. B. Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chem. Soc. Rev.* **44,** 1172–239 (2015).
20. Rubin-Pitel, S. B., Cho, C. M.-H., Chen, W. & Zhao, H. in *Bioprocessing for Value-Added Products from Renewable Resources* (ed. Yang, S.-T.) 49–72 (Elsevier Science, 2006).
21. 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).
22. Estabrook, R. W. A passion for P450s (rememberances of the early history of research on cytochrome P450). *Drug Metab. Dispos.* **31,** 1461–1473 (2003).
23. Joo, H., Lin, Z. & Arnold, F. H. Laboratory evolution of peroxide-mediated cytochrome P450 hydroxylation. *Nature* **399,** 670–673 (1999).
24. Roccatano, D. Structure, dynamics, and function of the monooxygenase P450 BM-3: insights from computer simulations studies. *J. Phys. Condens. Matter* **27,** 273102 (2015).
25. Whitehouse, C. J. C., Bell, S. G. & Wong, L.-L. P450 $_{BM3}$ (CYP102A1): connecting the dots. *Chem. Soc. Rev.* **41,** 1218–1260 (2012).
26. Munro, A. W. *et al.* P450 BM3: The very model of a modern flavocytochrome. *Trends Biochem. Sci.* **27,** 250–257 (2002).
27. Hyster, T. K. & Arnold, F. H. P450BM3-axial mutations: A gateway to non-natural reactivity. *Isr. J. Chem.* **55,** 14–20 (2015).
28. 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).

29. Govindaraj, S. & Poulos, T. L. The domain architecture of cytochrome P450BM-3. *J. Biol. Chem.* **272,** 7915–7921 (1997).
30. Fulco, A. J. P450BM-3 and other Inducible Bacterial P450 Cytochromes: Biochemistry and Regulation. *Annu. Rev. Pharmacol. Toxicol.* **31,** 177–203 (1991).
31. Arnold, F. H. The nature of chemical innovation: new enzymes by evolution. *Q. Rev. Biophys.* **48,** 404–410 (2015).
32. McIntosh, J. A., Farwell, C. C. & Arnold, F. H. Expanding P450 catalytic reaction space through evolution and engineering. *Curr. Opin. Chem. Biol.* **19,** 126–134 (2014).
33. Coelho, P. S., Brustad, E. M., Kannan, A. & Arnold, F. H. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* **339,** 307–10 (2013).
34. Heel, T., McIntosh, J. A., Dodani, S. C., Meyerowitz, J. T. & Arnold, F. H. Non-natural olefin cyclopropanation catalyzed by diverse cytochrome P450s and other hemoproteins. *ChemBioChem* **15,** 2556–2562 (2014).
35. 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).
36. Wang, Z. J. *et al.* Improved cyclopropanation activity of histidine-ligated cytochromeP450 enables the enantioselective formal synthesis of levomilnacipran. *Angew. Chemie - Int. Ed.* **53,** 6810–6813 (2014).
37. Renata, H. *et al.* P450-catalyzed asymmetric cyclopropanation of electron-deficient olefins under aerobic conditions. *Catal. Sci. Technol.* **4,** 3640–3643 (2014).
38. McIntosh, J. A. *et al.* Enantioselective intramolecular C-H amination catalyzed by engineered cytochrome P450 enzymes in vitro and in vivo. *Angew. Chemie - Int. Ed.* **52,** 9309–9312 (2013).
39. 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).
40. Farwell, C. C., McIntosh, J. A., Hyster, T. K., Wang, Z. J. & Arnold, F. H. Enantioselective imidation of sulfides via enzyme-catalyzed intermolecular nitrogen-atom transfer. *J. Am. Chem. Soc.* **136,** 8766–8771 (2014).
41. Prier, C. K., Hyster, T. K., Farwell, C. C., Huang, A. & Arnold, F. H. Asymmetric Enzymatic Synthesis of Allylic Amines: A Sigmatropic Rearrangement Strategy. *Angew. Chemie -*

Int. Ed. **55,** 4711–4715 (2016).

42. Farwell, C. C., Zhang, R. K., McIntosh, J. a., Hyster, T. K. & Arnold, F. H. Enantioselective Enzyme-Catalyzed Aziridination Enabled by Active-Site Evolution of a Cytochrome P450. *ACS Cent. Sci.* 150422075525009 (2015). doi:10.1021/acscentsci.5b00056
43. Wang, Z. J., Peck, N. E., Renata, H. & Arnold, F. H. Cytochrome P450-catalyzed insertion of carbenoids into N-H bonds. *Chem. Sci.* **5,** 598–601 (2014).
44. Dean, J. A. in *Lange's Handbook of Chemistry* (ed. Dean, J. A.) 4.46 (1999).
45. Chen, Z., Ost, T. W. B. & Schelvis, J. P. M. Phe393 Mutants of Cytochrome P450 BM3 with Modified Heme Redox Potentials Have Altered Heme Vinyl and Propionate Conformations. *Biochemistry* **43,** 1798–1808 (2004).
46. Berry, E. A. & Trumpower, B. L. Simultaneous determination of hemes a, b, and c from pyridine hemocrome spectra. *Anal. Biochem.* **161,** 1–15 (1987).