Matthias Fink, BSc

Engineering of a cupin from Thermotoga maritima with oxidative alkene-cleavage activity

MASTER'S THESIS

to achieve the university degree of

Diplom-Ingenieur

Master's degree programme: Biotechnology

submitted to

Graz University of Technology

Supervisor

Univ.-Prof. Dipl.-Ing. Dr.techn. Helmut Schwab

Institute of Molecular Biotechnology

Second Supervisor
Dipl.-Ing. Dr.nat.techn. Kerstin Steiner

Graz, March 2015

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present master's thesis dissertation.

_____
Date

_____
Signature

Abstract

Das cupin TM1459 aus dem Stamm *Thermotoga maritima* wurde durch ortsspezifische Mutagenese und Sättigungsmutagenese der Aminosäuren des aktiven Zentrums gezielt verändert, um die vorliegende Aktivität der oxidativen Spaltung von Kohlenstoffdoppelbindungen zu erhöhen. Die im Zuge der Sättigungsmutagenese mithilfe von NNK Primern erstellten Genbibliotheken, wurden mit einem speziell für diese Aufgabe entwickelten High-Throughput-Screening Verfahren untersucht. Enzymvarianten die dabei eine verbesserte Aktivität zeigten, wurden in Reaktionen im 1 mL Maßstab auf einen gesteigerten Umsatz überprüft. Dabei konnten Enzymvarianten mit verbessertem Umsatz sowie erweitertem Substratspektrum mit Erfolg isoliert werden.

Abstract

The cupin TM1459 from *Thermotoga maritima* was engineered by site directed and site saturation mutagenesis of active site amino acids to increase its activity in the oxidative cleavage of C=C double bonds. Site saturation libraries were generated with NNK primers and screened by a specifically developed photometric high throughput assay for the detection of the ketone product of the reaction. The conversion of different substrates in biphasic, 1 mL scale bioconversion reactions was determined by GC. Variants with improved conversion and extended substrate spectrum could successfully be isolated.

Acknowledgments

Table of Contents

1 Introduction

Classical chemical catalysis involves several severe drawbacks compared to biocatalysis. These drawbacks include the use of hazardous chemicals, the need for harsh reaction conditions and problems with selectivity in certain reactions. All these disadvantages also apply to oxidative alkene cleavage by chemical catalysis. This reaction yields ketones and aldehydes as cleavage products and is today mostly accomplished by ozonolysis. In this process, harsh reaction conditions of -78 °C, special equipment as well as reducing agents in molar amounts are used, imposing safety- and environmental concerns. Moreover, the danger of explosions especially in large-scale processes makes this method unfavorable. These circumstances caused emerging scientific effort to substitute classical catalysis with biocatalysis for these reactions.[1] There is already a considerable pool of enzymes known to catalyze oxidative alkene cleavage reactions, comprising heme-dependent enzymes with oxygen or hydrogen peroxide as oxidant as well as iron and non-iron metal-dependent enzymes, with all groups containing variants able to cleave aromatic and aliphatic double bonds.[1]

Because many metabolites in nature contain ketone- or aldehyde groups, oxidative alkene cleavage can be a potent tool in the synthesis of such compounds. One of many examples for the synthesis of a natural compound that could benefit from biocatalytic oxidative alkene cleavage is the synthesis of vanillin, the most popular flavor in the world. Today, the vast majority of vanillin originates from chemical synthesis, utilizing the petrochemical guaiacol as raw material.[2] Much effort has been put into the biocatalytic synthesis of vanillin in the past few years[3], on the one hand to make the synthesis process more environmentally sustainable and on the other hand to produce vanillin that can be labeled as "natural". The regulations of whether or not a flavor may be labeled as natural are slightly different in the EU and the US[4], but an enzymatic or microbial transformation of a compound originating from a living organism, forming a flavor that is identical to the natural compound, can be labeled as "natural" according to both legislations. There are already commercially available "natural" vanillin products produced by microbial fermentation from (i) ferulic acid (Solvay S.A.), (ii) glucose (Evolva Holding S.A.), (iii) curcumin (De Monchy Aromatics Ltd.) and (vi) eugenol (Mane S.A.). Given its low price, especially the de novo biosynthesis of vanillin from glucose seems to be a promising approach. This biosynthesis is done in *Saccharomyces cerevisiae*, starting with 3-dehydroshikimic acid, an intermediate in the shikimic acid pathway which is naturally occurring in plants and most microorganisms. 3-dehydroshikimic acid is then transformed into protochatechuic acid by a 3-dehydroshikimic acid dehydratase from

Podospora pauciseta. An *O*-methyltransferase from *Homo sapiens* then transforms protocatechuic acid into vanillic acid, which is in turn reduced to vanillin by an aromatic carboxylic acid reductase from a bacterium from the *Nocardia* genus.[5] The introduction of this novel pathway makes the yeast used in this process a genetically modified organism, which raises issues with customer perception of the hereby produced "natural" vanillin.[6] A similar approach was also shown in *E. coli*.[7] The production from ferulic acid, eugenol and isoeugenol was already shown in multiple hosts (GMO and non-GMO), employing different pathways[3]. Although they are generally more expensive than glucose, these substances can be effectively extracted from plants and are thus relatively inexpensive natural substrates for vanillin biosynthesis. It has been reported recently, that the enzyme *VpVAN* transforms ferulic acid to vanillin in the pods of the vanilla plant *Vanilla planifolia* in a single step reaction following a retro-aldol reaction mechanism.[8] Moreover, isoeugenol may be transformed into vanillin by a single oxidative alkene cleavage of the double bond adjacent to the aromatic ring. Enzymes that catalyze this reaction are known as isoeugenol monooxygenases e.g. *Iem* from *Pseudomonas nitroreducens* Jin1[9] and *Iso* from *Pseudomonas putida* IE27[10] but also HRP (Horse Radish Peroxidase) has been found to catalyze this reaction under the right conditions.[11] Further enzymes that may be able to produce vanillin through alkene cleavage are the myeloperoxidase (MPO) and the peroxidase from *Coprinus cinerus* (CiP) which catalyze the $C_\alpha = C_\beta$ double bond cleavage of various substituted styrene precursors[12] and the lignin peroxidase (LiP) which transformed phenolic, protected vanillin precursors like *o*-ethylisoeugenol to the corresponding benzaldehyde derivatives.[13,14] All three enzymes are heme-dependent and require hydrogen peroxide as oxidant. There is also a manganese dependent enzyme from *Trametes hirsuta* FCC047 which was found to cleave the double bond adjacent to the aromatic ring of several styrene derivatives.[15]

Recently, the manganese dependent enzyme TM1459 from *Thermotoga maritima* was found to catalyze the oxidative alkene cleavage of C=C double bonds adjacent to the aromatic ring of several styrene derivatives, with *tert*-butyl hydroperoxide as oxidizing agent[16]. However, the enzyme showed unfavorably high substrate specificity with only *p*-Cl-α-methylstyrene and α-methylstyrene showing significant conversions. This limited substrate spectrum makes this enzyme rather disadvantageous as a biocatalytic tool in its wild type form. However, it may be possible to improve TM1459 by enzyme engineering, making it a potent biocatalyst for the synthesis of vanillin and other valuable compounds.

1.1 Aims of this thesis

The aims of the experiments associated with this thesis were the increase of the activity and the enhancement of the substrate spectrum of TM1459 (from now on called Cupin 2 in this thesis) through enzyme engineering. To be able to quickly test Cupin 2 variants for their activity, a high throughput assay was developed specifically for the detection of the ketone products formed in the reaction. For the screening, α-methylstyrene was chosen as a model substrate for Cupin 2, which catalyzes the oxidative cleavage of this substrate to acetophenone and formaldehyde, as shown in Figure 1.
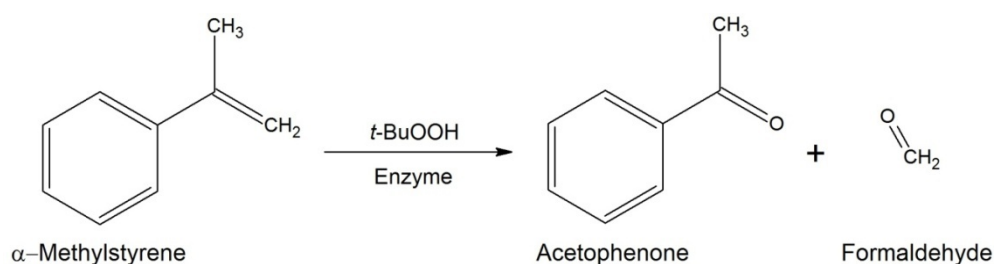
Figure 1: Model reaction chosen as the basis for the screening assay. α-Methylstyrene is oxidatively cleaved by Cupin 2 in the presence of *tert*-butyl hydroperoxide (*t*-BuOOH) as oxidizing agent, forming the two reaction products acetophenone and formaldehyde.

The engineering consisted of three consecutive rounds of mutagenesis. The first round involved site directed mutagenesis of all metal binding site and active site amino acids. In the second round, site saturation libraries of all active site amino acids were generated and screened using the newly developed assay. The third and last round of mutagenesis consisted of the combination of different mutations that turned out to be promising in the first two rounds. Selected variants were then used in 1 mL scale bioconversion reactions with different substrates and the conversion was measured by gas chromatography (GC).

2 Materials and methods

2.1 Target protein

The protein to be engineered in the following experiments was the manganese-dependent cupin TM1459 from the thermophilic bacterium *Thermotoga maritima*. It consists of 114 amino acids and has a mass of 13.1 kDa. The gene, in a codon optimized version for *E. coli,* was available in a cryo-culture of *E. coli* TOP10F' carrying the gene cloned into the pET-26b(+)-vector. The plasmid was isolated and used as template for the mutagenesis experiments. Figure 2 shows the 3D-structure of the enzyme in its dimeric form as well as a detailed view on its active site. The nucleotide- and amino acid sequences as well as additional information about Cupin 2 can be found in Appendix C.
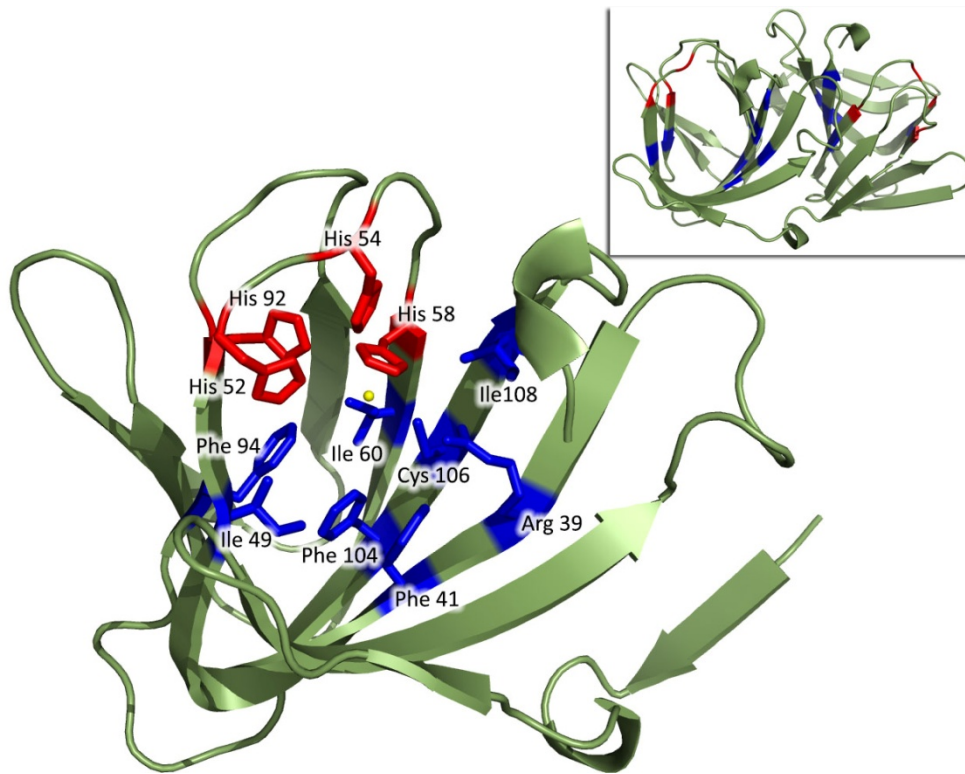
Figure 2: Cupin 2 in its active, homodimeric form (upper right) as well as a detailed view on active site (blue), metal binding site (red) and the manganese ion (yellow).

2.2 Vector

The vector used in all experiments was the pET-26b(+) expression vector. The genes were cloned into the multiple cloning site between the HindIII and NdeI restriction sites, where it stands under transcriptional control of the T7 promoter and terminator. The vector construct with the cloned Cupin 2 gene is shown in Figure 3. The vector was obtained from Merck KGaA (Darmstadt, Germany).
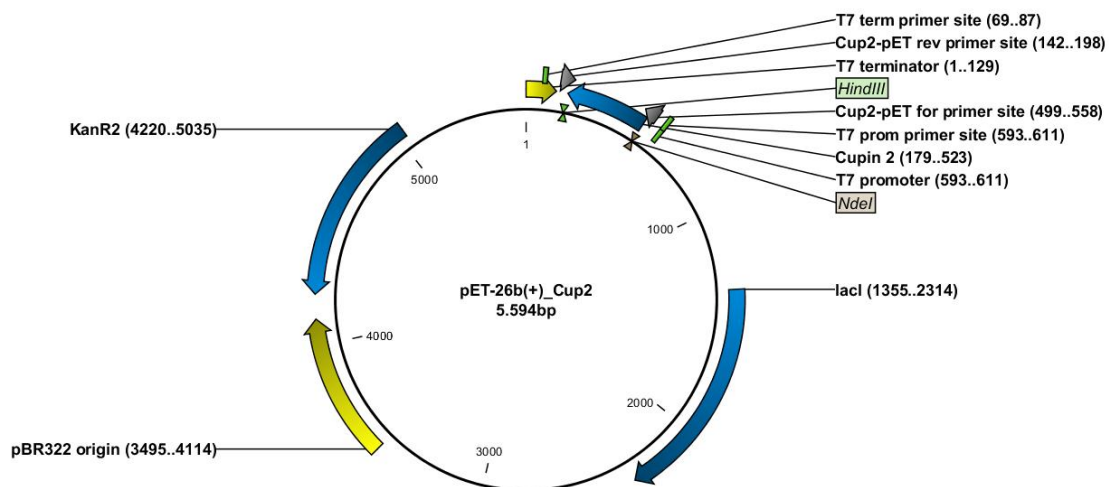
Figure 3: Vector map of the vector construct used in this thesis. The construct consists of the pET-26b(+) expression vector with Cupin 2 cloned into its multiple cloning site, using the HindIII and NdeI restriction sites.

2.3 Bacterial strains

E. coli TOP10F' was used as host in all mutation experiments. For expression of the proteins, *E. coli* BL21 (DE3) Gold was used. This strain carries an isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible T7 polymerase which enables the induced expression of the enzyme. The bacterial strains were obtained from Life Technologies (Carlsbad, CA, USA).

2.4 Cultivation

For agar plates, agar with lysogeny-broth (LB-Agar (Lennox)) with 40 µg/mL of kanamycin was used. For cultivation, the plates were incubated at 37 °C, for storage, the plates were sealed with Parafilm® (Bemis Company Inc., Oshkosh, WI, USA) and stored at 4 °C. For cultivation in liquid media, LB-Broth (Lennox) with 40 µg/mL of kanamycin was used.

As from now, agar plates with 40 µg/mL kanamycin will be referred to as "LB-Kan plates" and LB-Broth with 40 µg/mL kanamycin will be referred to as "LB-Kan broth" in this thesis. Media and kanamycin were obtained from Carl Roth GmbH + Co. KG (Karlsruhe, Germany).

2.5 Generation of Cupin 2 mutants

2.5.1 Mutations and primers

In the initial site directed mutagenesis round, all active site and metal binding site amino acids as well as tryptophan 56 near the active site were mutated. All four histidines involved in metal binding were individually mutated to alanines to verify their role in metal binding. In the active site, isoleucines were mutated to alanines, and phenylalanines were mutated to valines as well as alanines in order to verify the effect of smaller amino acids at these positions. Arginine 39 was mutated to alanine as well as histidine, in order to verify the effects of the positive charge and the size of the residue. Cysteine 106 was mutated to alanine as well as serine, to verify the effects of the thiol group and the size of the residue. Tryptophan 56, which is located near the active site, was mutated to alanine in order to verify the effects of the size of the side chain. Table 1 shows a summary of the amino acid exchanges introduced in the first round of mutagenesis.

Table 1: Amino acid exchanges introduced in the initial round of site directed mutagenesis.

AA position	location	mutated to...
His52	metal binding site	Ala
His54	metal binding site	Ala
His58	metal binding site	Ala
His92	metal binding site	Ala
Arg39	active site	Ala, His
Phe41	active site	Ala, Val
Ile49	active site	Ala
Trp56	near active site	Ala
Ile60	active site	Ala
Phe94	active site	Ala, Val
Phe104	active site	Ala, Val
Cys106	active site	Ala, Ser
Ile108	active site	Ala

In the second mutagenesis round, site saturation libraries of all active site amino acids and of the near active site amino acid tryptophan 56 were generated and screened using the novel screening assay (see below).

All primer sequences that were used for the mutation experiments are listed in Appendix A. The genes coding for the C-terminally truncated variants were generated using a single polymerase chain reaction (PCR) run, utilizing the synCup2(pET26)_for forward primer along with the corresponding reverse primer, followed by Gibson-cloning (see below) to integrate the fragments into the vector. The genes for the three multi-muteins without C-terminal truncation were ordered as gBlocks, which were amplified by PCR (see below). Freeze dried gBlocks were dissolved in 20 µL H_2O ("Fresenius") and 1 µL of a 1:20 dilution was used as template for the PCR reaction. All other site directed and site saturation mutants were generated using an overlap extension PCR protocol, followed by ligation or Gibson-cloning (see below). For the inserts used with the ligation protocol, the outer primers synCupin2(Nde)_for and T7term_rev were used. For the inserts later cloned via Gibson-cloning, the outer primers synCup2(pET26)_for and synCup2(pET26)_rev were used.

All primers and gBlocks were obtained from Integrated DNA Technologies Inc. (Coralville, IA, USA).

2.5.2 PCR

If not stated otherwise, PCR reactions were performed using following standard reaction mix and temperature program.

PCR reaction mix:

- 35.5 µL H$_2$O ("Fresenius")
- 10 µL Phusion buffer HF (5x)
- 1 µL forward primer (10 pmol/µL)
- 1 µL reverse primer (10 pmol/µL)
- 1 µL dNTPs (10 mM)
- 1 µL template (typically 1 ng plasmid DNA)
- 0.5 µL Phusion High Fidelity DNA Polymerase (2 U/µL)

Temperature program:

- 2 min 98 °C
    - 30 sec 95 °C ⎤
    - 30 sec 55 °C ⎬ 30 cycles
    - 15 sec 72 °C ⎦
- 7 min 72 °C
- keep on 4 °C

2.5.2.1 Overlap extension PCR

The overlap extension PCR protocol consists of two consecutive PCR reactions. In the first reaction, two fragments with overlapping ends are produced in two separate PCR reactions. By using a combination of an outer primer with one of the mutagenesis primers in each reaction, partly overlapping fragments are produced that both carry the desired mutation in the overlapping area. For the single-mutants, a pET-26b(+) plasmid carrying the desired wild-type Cupin 2 gene was used as template. For the double-mutants and triple-mutants, a plasmid with a previously obtained single-mutant or double-mutant respectively, carrying one or two of the desired mutations, was used as template. PCR reactions were performed with 1 ng template vector, using the standard protocol given above. The PCR products were then isolated using a preparative agarose gel and gel clean-up (see below). The isolated products were used as templates for the second PCR reaction, which used the standard temperature program (see above) and following reaction mix:

- 35.5 µL H_2O ("Fresenius")
- 10 µL Phusion buffer HF (5 x)
- 1 µL outer primer forward (10 pmol/µL)
- 1 µL outer primer reverse (10 pmol/µL)
- 2.5 µL PCR product forward
- 2.5 µL PCR product reverse
- 1 µL dNTPs (10 mM)
- 0.5 µL Phusion High Fidelity DNA Polymerase

For the first round of mutagenesis, the ligation protocol was used to assemble the expression vector. Because of the inferior quantity of transformants using this protocol, the Gibson-cloning protocol was used for library generation as well as for all further mutation experiments.

Phusion High Fidelity DNA Polymerase and the corresponding buffer as well as deoxynucleoside triphosphates (dNTPs) were obtained from Thermo Fisher Scientific Inc. (Waltham, MA, USA).

2.5.3 Ligation

The whole reaction mix originating from the overlap extension PCR was purified using a commercial kit (see below) The resulting cleaned-up PCR product was digested with HindIII and NdeI at 37 °C over night using following reaction mix:

- whole mixture (~43 µL/~1-2 µg) purified PCR product
- 5 µL CutSmart buffer (10 x)
- 0.5 µL NdeI (20 U/µL)
- 0.5 µL HindIII HF (20 U/µL)

The pET-26b(+) vector was digested using the following reaction mix:

- x µL pET-26b(+) (5-6 µg)
- 10 µL CutSmart buffer (10 x)
- 2.5 µL HindIII HF (20 U/µL)
- 2.5 µL NdeI (20 U/µL)
- x µL H_2O ("Fresenius"; to 100 µL total volume)

After over-night incubation, the digested vector was isolated using a preparative agarose gel followed by a gel clean-up (see below).

For the ligation of the digested gene with the vector, a vector/insert ratio of 1:3 was chosen, which adds up to following reaction mix:

- x µL (20 ng) insert (depending on DNA concentration)
- x µL (100 ng) digested vector (depending on DNA concentration)
- 2 µL T4 Ligase buffer (10x)
- 1 µL T4 Ligase (400 U/µL)
- x µL H_2O ("Fresenius"; to a total volume of 20 µL)

The ligation reaction mixture was incubated at 22 °C for 90 minutes, inactivated for 10 minutes at 64 °C and finally desalted. The whole desalted product (~15 µL) was transformed into electrocompetent *E. coli* TOP10F' (see below). For cultivation, the whole transformation broth was plated on LB-Kan plates.

All enzymes and corresponding buffers were obtained from New England Biolabs Inc. (Ipswich, MA, USA).

2.5.4 Gibson-cloning

For the Gibson-cloning protocol, the whole reaction mix from the overlap extension PCR reaction was purified using a preparative agarose gel followed by a gel clean-up (see below). 30-50 ng of the purified insert were then mixed with digested and purified vector in a molar ratio of 1:1, resulting in 5 µL of DNA solution, which was then added to a 15 µL aliquot of Gibson master mix. The Gibson master mix (1200 µL) consisted of the following components:

- 320 µL 5x ISO reaction buffer
- 0.64 µL T5 exonuclease (10 U/µL)
- 20 µL Phusion High Fidelity DNA Polymerase (2 U/µL)
- 160 µL Taq DNA Ligase (40 U/µL)
- 699.36 µL H_2O ("Fresenius")

The 5x isothermal (ISO) reaction buffer consists of the following components:

- 25 % (w/v) Polyethylene Glycol (PEG)-8000
- 50 mM Tris/Cl pH 7.5
- 50 mM $MgCl_2$

- 50 mM (Dithiothreitol) DTT
- 1 mM Deoxyadenosine triphosphate (dATP)
- 1 mM Deoxycytidine triphosphate (dCTP)
- 1 mM Deoxyguanosine triphosphate (dGTP)
- 1 mM Deoxythymidine triphosphate (dTTP)
- 5 mM Nicotinamide adenine dinucleotide (NAD)
- in H_2O ("Fresenius")

The reaction mix (20 µL) was incubated at 50 °C for 60 minutes. For library generation, the mix was desalted and an equivalent of 1 ng DNA was transformed into *E. coli* TOP10F', while for simple mutation experiments, 2.5 µL of the mix were directly transformed into *E. coli* TOP10F' without prior desalting.

T5 exonuclease was obtained from Biozym Scientific GmbH (Hessisch Oldendorf, Germany), Phusion Polymerase and dNTPs were obtained from Thermo Fisher Scientific Inc., Taq DNA Ligase was obtained from New England Biolabs Inc., PEG-8000 and DTT were obtained from Carl Roth GmbH, Tris/Cl and NAD were obtained from Sigma-Aldrich (St. Louis, MO, USA). $MgCl_2$ was obtained from Merck KGaA.

2.5.5 Gel and reaction clean-up

The purification of fragments from agarose gels, PCR products and products from restriction digests was done using the Wizard® SV Gel and PCR Clean-Up System, which was obtained from Promega Corporation (Fitchburg, WI. USA).

Protocol:

Product Preparation

- For gel clean-up: Excise gel slice, place into 1.5 mL microreaction tube and add 10 µL Membrane Binding Solution per 10 mg of gel slice. Dissolve gel by shaking at 1200 rpm and 75 °C in a thermomixer.
- For reaction clean-up: Add H_2O ("Fresenius") to the reaction mixture to get a volume of 100 µL and add 100 µL Membrane Binding Solution.

Binding of DNA

- Insert SV Minicolumn into collection tube and transfer the prepared mixture to the Minicolumn. Incubate at room temperature for 1 minute.
- Centrifuge the assembly for 1 minute and discard the flow through

Washing

- Add 700 µL Membrane Wash Solution, centrifuge for 1 minute and discard the flow through. Add 500 µL Membrane Wash Solution, centrifuge for 5 minutes and discard the flow through.
- Centrifuge for 1 minute with empty collection tube.

Elution

- Transfer Minicolumn to clean 1.5 mL microreaction tube and wait for 2 minutes for residual ethanol to evaporate.
- Add 50 µL H_2O ("Fresenius"), or 30-40 µL for products showing weaker bands on the agarose gel, and incubate at room temperature for 2 minutes.
- Centrifuge for 2 minutes to elute DNA

All centrifugation steps were carried out at 16,000x g. DNA solutions were stored at 4 °C for short periods and -20 °C for longer periods.

2.5.6 Preparation of electrocompetent cells

The electrocompetent *E. coli* TOP10F' and BL21 (DE3) Gold cells that were used in all transformation steps were prepared using the following protocol:

50 mL of 2xTY broth were inoculated with a single colony of the corresponding strain and incubated at 37 °C over night. 2 L shaking flasks containing 400 mL of 2xTY broth were then inoculated with 4 mL of this overnight culture and incubated at 37 °C for 2.5 hours at 120 rpm in an orbital shaker. At an OD_{600} of 0.8, the cultures were cooled on ice for 30 minutes and were then centrifuged for 10 minutes at 3,000x g. The resulting pellet was suspended in buffer (0.1 mM HEPES, pH 7) and the suspension was centrifuged for 10 minutes at 4,000x g. The resulting pellet was again resuspended in buffer and the centrifugation tube was filled ½ with glycerol solution (10 % in deionized water). The mixture was centrifuged for 20 minutes at 4,500x g and the resulting pellet was again resuspended in glycerol solution. The centrifugation tubes were then filled 2/3 with glycerol solution and centrifuged for 15 minutes at 5,000x g. The resulting pellet was resuspended in 4 mL glycerol solution. The suspension was aliquoted in pre-cooled 1.5 ml microreaction tubes and frozen at -80 °C.

All centrifugation steps were carried out at 4 °C, the cells were kept on ice during the whole procedure and buffer and glycerol solution were pre-cooled to 4 °C. All materials and

substances were sterilized before use. 2xTY broth, glycerol and HEPES buffer were obtained from Carl Roth GmbH.

2.5.7 Transformation

All desalting steps prior to transformation were performed using membrane filters (0.025 µm) obtained from Merck-Millipore (Merck KGaA) on H_2O ("Fresenius") for 15 minutes using a volume of 10-20 µL of reaction mix.

Transformations were performed using program "Ec2" on a "MicroPulser" electroporation device from Bio-Rad Laboratories Inc. (Hercules, CA, USA). For the transformation of Gibson-cloning products and ligation products, 80 µL of electrocompetent *E. coli* TOP10F' cells were mixed with 5 µL (ligation) or 2.5 µL (Gibson-cloning) of desalted reaction mixture. For the transformation of isolated plasmids, 1 µL of isolated plasmid (in H_2O) was transformed into 40 µL of electrocompetent cells. After the transformation, the cells were suspended in 1 mL SOC broth, regenerated for 30 minutes at 37 °C and 700 rpm in a thermomixer and plated on LB-Kan plates in aliquots of 50 and 100 µL (if not stated otherwise).

SOC Broth:

- 20 g Trypton
- 0.58 g NaCl
- 5 g Yeast Extract
- 2 g $MgCl_2$
- 0.18 g KCl
- 2.46 g $MgSO_4$
- 3.46 g Glucose
- 1 L H_2O deionized

$MgCl_2$ was obtained from Merck KGaA. All other media components were obtained from Carl Roth GmbH.

2.5.8 Colony PCR

A colony PCR was used to verify the correct insertion and length of the mutated gene prior to sequencing. As primers, sequences located in the T7 promoter (forward primer) and T7 terminator (reverse primer) regions were used. For this method, colonies were picked directly from the transformation plates using a sterile toothpick. A part of the cell material was then

transferred to a separate LB-Kan plate, creating a "master plate", and the rest was suspended in the prepared PCR mixture.

Colony PCR reaction mixture:

- 20.5 µL H$_2$O ("Fresenius")
- 2.5 µL DreamTaq buffer (10 x)
- 0.5 µL dNTPs (10 mM)
- 0.5 µL T7 primer forward (10 pmol/µL)
- 0.5 µL T7 primer reverse (10 pmol/µL)
- 0.125 µL DreamTaq DNA Polymerase (5 U/µL)

PCR-program:

- 10 min 95 °C
  - o 30 sec 95 °C
  - o 30 sec 53 °C } 25 cycles
  - o 1 min 72 °C
- 7 min 72 °C
- keep on 4 °C

The resulting PCR products were then analyzed using an agarose gel electrophoresis (see below). DreamTaq DNA Polymerase, DreamTaq Buffer and NTPs were obtained from Thermo Fisher Scientific Inc.

2.5.9 Agarose gel electrophoresis

Prior to electrophoresis, DNA Loading Dye Solution (6x) was added to the samples. For the agarose gel electrophoresis, gels with 1 % agarose were used. Analytical gels used for the colony PCRs were run at 120 Volts for 45 minutes. Preparative gels for DNA purification were run at 90 Volts for 90 minutes. For fragments smaller than 200 bp, preparative gels with 2 % agarose were used. Those gels were run at 75 Volts for 2 hours. Per gel (200 mL), one drop (~50 µL) of ethidium bromide (1 mg/ml) was added. TAE-Buffer was used as running buffer for all gels. As standard, 5 µL O'GeneRuler DNA Ladder Mix (0.1 µg/µL) was used. Biozym LE Agarose was obtained from Biozym Scientific GmbH. DNA Ladder and Loading Dye were obtained from Thermo Fisher Scientific Inc.

TAE-Buffer (50x):

- 242 g Tris(hydroxymethyl)aminomethane (Tris)

- 57.1 mL Acetic Acid

- 18.6 g Ethylenediaminetetraacetic acid (EDTA)

- fill with H_2O (deionized) up to 1 L

2.5.10 Plasmid isolation

For plasmid isolations, cell material was streaked on LB-Kan plates and incubated at 37 °C over night. The cell material of half an agar plate was used for one plasmid isolation. All plasmid isolations were done using the GeneJET Plasmid-Miniprep Kit from Thermo Fisher Scientific Inc. with slightly modified protocol:

- Resuspend cell material thoroughly in 250 µL of Resuspension Solution

- Add 250 µL Lysis Solution and mix by inverting 4-6 times until solution gets slightly clear

- Add 350 µL of Neutralization Solution and mix by inverting 4-6 times

- Centrifuge for 10 minutes

- Transfer the supernatant to the GeneJET spin column

- Centrifuge for 1 minute and discard flow through

- Add 500 µL Wash Solution and centrifuge for 30 seconds

- Discard flow through, add another 500 µL Wash Solution and centrifuge for 1 minute

- Discard flow through and centrifuge for another 2 minutes

- Transfer column into fresh 1.5 mL microreaction tube and wait 2 minutes for the residual ethanol to evaporate

- Add 50 µL (or 30-40 µL for low amount of cell material) H_2O ("Fresenius") and incubate for 2 minutes

- Centrifuge for 2 minutes to elute the plasmid DNA

All centrifugation steps were performed at 16,000x g. DNA solutions were stored at 4 °C for short periods and -20 °C for longer periods.

2.5.11 DNA concentration measurement

All DNA concentration measurements were carried out on a NanoDrop 2000 Spectrophotometer from Thermo Fisher Scientific Inc. A volume of 2 µL of DNA solution was used to measure the concentrations.

2.5.12 Sequencing

Isolated plasmids were sequenced using the overnight sanger-sequencing service of Microsynth AG Vienna.

2.6 Library generation and screening

For the generation of the site saturation mutagenesis libraries, primers with NNK codons on the position corresponding to the intended amino acid positions were used in the overlap extension PCR protocol, in order to generate inserts with random codons on these amino acid positions. The inserts were assembled with the pET-26b(+) vector backbone by Gibson-cloning. 2.5 µL of the desalted Gibson-cloning reaction mix were then transformed into 80 µL of electrocompetent *E. coli* TOP10F' cells. After a regeneration time of 60 minutes, an aliquot of 50 µL of the transformation broth was plated on LB-Kan plates. 40 µg/mL kanamycin was added to the rest of the cell suspension and it was incubated at 37 °C for another 3 h. The whole suspension was then used to inoculate an overnight culture of 5 mL LB-Kan broth. 4 mL of this over-night culture were then used for a plasmid isolation to obtain a mixed plasmid preparation. After an over-night incubation of the agar plate that was plated after transformation, 5 colonies were selected and re-streaked for plasmid isolation. Plasmids were sent to sequencing for the evaluation of the library quality. Of the mixed plasmid preparation, an equivalent of 1 ng DNA was transformed into 40 µL of electrocompetent *E. coli* BL-21 (DE3) Gold cells. After regeneration, aliquots of 50 µL and 100 µL of the transformation mix were plated on LB-Kan plates. Of these plates, 176 colonies per library were cultivated using a deepwell plate fermentation and were later screened using the novel screening assay (see below), while 5 additional colonies were re-streaked for plasmid isolation. The plasmids were sent to sequencing for further evaluation of the library quality. After screening, the clones with the highest increase in absorption compared to the wild type were selected for re-screening, which was done by performing another deepwell plate fermentation and consecutive screening of the selected library hits in octuplicates. Inoculation of the selected clones was done from cryo-plates of the library. Variants that showed a positive result in the re-screening were streaked for plasmid isolation and the plasmids were sent to sequencing.

A library size of 176 clones along with the number of possible codons for a single position, which is 32 for NNK codons, results in an oversampling factor of 5.5 and a probability of 99.6 % for a particular codon to occur the library. These values were calculated using Equation 1 which was derived from the work of Reetz et al[17].

$$O_f = \frac{T}{V} = -\ln(1 - P_i)$$

It has to be noted, however, that the aforementioned equation does not respect the unavoidable bias that is created by an uneven distribution of codons in the primer mix as well as the different binding affinities of different primer variants. This means that a certain oversampling factor is obligatory for every library in order to compensate for this bias.

2.7 Deepwell plate fermentation

For an overnight culture, deepwell plates were filled with 750 µL LB-Kan broth. Inoculation was done with sterile toothpicks either from single colonies on agar plates or from cryo-cultures. The overnight cultures (ONC) were incubated at 300 rpm and 37 °C in an orbital shaker. For the main culture, deepwell plates were filled with 750 µL LB-Kan broth with 100 µM $MnCl_2$ and inoculated by transferring 25 µL per well from the ONC to the main culture. The main culture was again incubated at 300 rpm and 37 °C. After 6 hours, protein expression was induced by adding IPTG, pre-dissolved in 50 µL of LB-Kan broth with 100 µM $MnCl_2$, to a final IPTG concentration of 0.1 mM. The incubation was then carried on at 25 °C and 300 rpm for another 17 hours, before the cells were harvested by centrifugation at 3,200x g for 15 minutes. The supernatant was decanted and the deepwell plates containing the pellets were frozen at -20 °C as an initial step in cell lysis. For the deepwell plates carrying the libraries, cryo-cultures of the ONC plates were prepared by adding 400 µL of sterile 50 % glycerol per well and freezing the plates at -20 °C.

2.7.1 Deepwell plate cell lysis

300 µL of lysis buffer per well were transferred directly onto the frozen pellets and the plate was vortexed until the pellet was resuspended. Then the plates were shaken at 1,000 rpm for 1 hour at room temperature. After that, the plates were centrifuged for 15 minutes at 3,200x g. The supernatant was then used for the screening assay. The composition of the lysis buffer was as follows:

- 280 µg/mL Lysozyme from chicken egg white (~70,000 U/mg)
- 4 U/mL Benzonase® Nuclease (purity >90 %, 250 U/µL)
- 0.2 % Triton X-100
- in NaPi 50 mM pH 8

Lysozyme was obtained from Fluka (Sigma-Aldrich), Benzonase® was obtained from Merck KGaA.

2.8 Vanillin assay for ketone detection

The assay developed in the course of this thesis relies on the aldol condensation reaction of vanillin with ketones and is widely based on the work of Amlathe and Gupta, 1990[18]. The reaction forms a yellow color complex, which enables a reliable detection of acetophenone, the ketone product produced in the model reaction. The proposed detection reaction is shown in Figure 4.
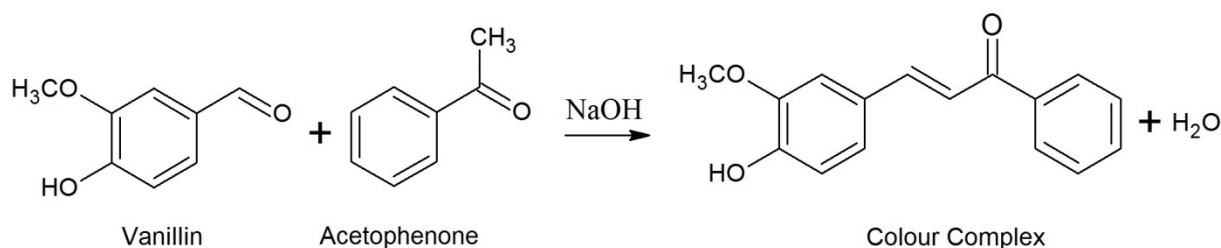

Figure 4: Proposed detection reaction used in the screening assay. In this aldol condensation reaction, vanillin and acetophenone are forming a yellow color complex. NaOH is needed for the activation of vanillin.

After deepwell plate fermentation and cell lysis, 130 µL of lysate per well were transferred from a deepwell plate to a microtiter plate. 50 µL of substrate mix per well were added to the lysate, the plate was sealed with adhesive film and incubated for 2 hours at 37 °C. After incubation, 50 µL of detection mix per well were added, the plate was sealed again and incubated at 37 °C to promote the formation of the color complex. After 30 minutes at 37 °C, the plates were ready for the photometric detection that was done by measuring the absorption at 442 nm in a plate reader. The screening of the variants from the first round of mutagenesis and foregoing experiments in assay development were done on a FLUOstar Omega plate reader (BMG Labtech GmbH, Ortenberg, Germany). The screening of libraries and further variants was done on an EON plate reader (BioTek Instruments GmbH, Bad Friedrichshall, Germany).

The composition of substrate mix and detection mix was as follows:

Substrate mix:

- 46 % Ethanol
- 36 mM α-Methylstyrene (final concentration in reaction 10 mM)
- 90 mM tert-Butyl hydroperoxide (final concentration in reaction 25 mM)
- 0.2 % Triton X-100
- in NaPi 50 mM pH 8 buffer

Detecion mix:

- 2 M NaOH
- 3.5 % (w/v) Vanillin
- 10 % Ethanol

For the preparation of the detection mix, 350 mg/mL vanillin was dissolved in ethanol. 1 part of the vanillin solution was then mixed with 9 parts of 2.2 M NaOH and used immediately. For the preparation of the substrate mix, α-methylstyrene and *tert*-butyl hydroperoxide (70 % solution in water, "Luperox[®]") were first dissolved in ethanol, and buffer as well as Triton X-100 was added afterwards.

NaOH and Triton X-100 were obtained from Carl Roth GmbH, α-methylstyrene, acetophenone (reaction product for assay development), *tert*-butyl hydroperoxide and vanillin were obtained from Sigma-Aldrich. Ethanol 95-97% (V/V) "AnalaR NORMAPUR[®]" was obtained from VWR International LLC (Radnor, PA, USA)

2.9   Protein expression

The correct expression of the newly generated enzyme variants was verified by drawing samples from the deepwell plate cell lysis and analyzing them in a sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The enzymes for the bioconversion reactions were produced in shaking flask scale fermentations and were also analyzed by SDS-PAGE.

2.9.1   Shaking flask fermentation

Protein expression was done using *E. coli* BL21 (DE3) Gold carrying the pET-26b(+) expression vector with the corresponding mutated Cupin 2 gene. An overnight culture of 50 mL LB-Kan broth in 100 mL shaking flasks was inoculated from a cryo culture or with a single colony from an agar plate and then incubated at 37 °C over night in an orbital shaker at 120 rpm. For the main culture, baffled 1 L shaking flasks containing 400 mL of LB-Kan broth with 100 μM $MnCl_2$ were inoculated to an OD_{600} of 0.1 and incubated at 37 °C in an orbital shaker at 120 rpm. When an OD_{600} of 0.8 was reached, protein expression was induced by adding IPTG to a final concentration of 0.1 mM. After induction, the temperature was reduced to 25 °C. After another 20 hours of shaking, the cells were finally harvested by centrifugation and disrupted by sonication (see below). For long time storage, cryo-cultures were produced using 1 mL of ONC, adding 0.5 mL of 50 % glycerol and freezing the mixture.

2.9.2 Cell disruption by sonication

For the cell disruption by sonication, the whole cell broth (~400 mL) was first centrifuged for 15 minutes at 4 °C and 5,000x g in a JA-10 rotor and the pellet was transferred into a pre-weighed 50 mL plastic tube. The weighed pellet was then resuspended in 5 volumes (or at least 25 mL) of cold 50 mM NaPi (pH 7) and disrupted by sonication with a Branson Sonifier S-250 for 6 minutes at 80 % duty cycle and 70 % output control. The sonication was carried out on ice. The cell lysate was then centrifuged for one hour at 50,000x g in a JA 25.50 rotor and the supernatant was used as crude cell lysate.

2.9.3 Bradford protein assay

All protein concentrations were measured using the Bradford protein assay[19]. The calibration curve was done with BSA concentrations of 12, 10.5, 9, 7.5, 6, 4.5, 3 and 1.5 µg/mL. The Bradford reagent was diluted with deionized water at a ratio of 1:5. The lysate sample was diluted 1:50 with deionized water, 50 µL of the sample were added to 950 µL of diluted Bradford reagent and incubated for 10 minutes. Then the absorption at 595 nm was measured in a photometer. The concentrations were automatically calculated from the calibration curve by the photometer's software. All samples were measured in duplicates.

The Bradford solution was obtained from Bio-Rad Laboratories Inc. BSA (Albumin Fraction V) was obtained from Carl Roth GmbH. All absorption measurements were done using a Cary Series UV-Vis Spectrophotometer from Agilent Technologies (Santa Clara, CA, USA).

2.9.4 SDS-PAGE

The soluble protein fraction was diluted and used directly. For lysate samples with known protein concentration (i.e. samples originating from shaking flask fermentation), 5 µg of total protein was used in a single SDS-PAGE sample while for lysates originating from deep-well plate fermentation, 2.5 µL of lysate was used for a single SDS-PAGE sample. The pellet was resuspended in 10 volumes of 6 M urea, centrifuged for 5 minutes at 20,000x g and finally diluted 1:2 for SDS-PAGE analysis. For sample preparation, 15 µL samples of the soluble or pellet fraction, containing 1.5 µL NuPAGE Sample Reducing Agent (10x), were mixed with 5 µL Novex NuPAGE LDS Sample Buffer (4x) and denatured for 10 minutes at 95 °C. 10 µL of this mixture were then loaded on 10 or 15 slot Novex NuPAGE 4-12 % Bis-Tris gels. As running buffer, Novex NuPAGE MES SDS
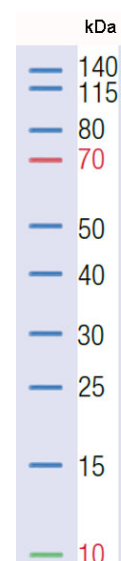

Figure 5: PageRuler Prestained protein ladder as used in all SDS gels in this thesis.

Running Buffer was used. As standard, PageRuler Prestained Protein Ladder was used (see Figure 5). The gels were run with the predefined NuPAGE gel program in an XCell SureLock™ Mini-Cell electrophoresis system. All Novex NuPAGE materials including gel chambers and power supplies were obtained from Life Technologies. The PageRuler Prestained Protein Ladder was obtained from Thermo Fisher Scientific Inc.

The gels were stained for 15 minutes in following staining solution:

- 500 mL Ethanol
- 75 mL Acetic Acid
- 2.5g Coomassie Brilliant Blue G250
- 425 mL H_2O (deionized)

Destaining was performed using the following destaining solution:

- 200 mL Ethanol
- 75 mL Acetic Acid
- 725 mL H_2O (deionized)

The destaining solution was changed every 20 minutes until the destaining was complete.

2.10 Determination of conversions

2.10.1 Heat purification

In the course of this thesis, a simple and effective heat purification protocol for Cupin 2 was established. All enzyme preparations used in the bioconversion reactions were purified using this protocol. Heat purification was done in 2 mL microreaction tubes, filled with 2 mL crude cell lysate. The tubes were incubated for 10 minutes at 75 °C and 600 rpm in a thermomixer (comfort, Eppendorf, Hamburg, Germany) and then centrifuged for 5 minutes at 20,000x g to clear the lysate of precipitated protein. To yield the desired protein concentration, the supernatant was diluted with buffer (50 mM NaPi pH 7). Diluted, purified lysate was directly used in the bioconversion reactions.

2.10.2 Bioconversion reactions

To compare the conversions of the new enzyme variants, bioconversion reactions in 1 mL scale were conducted. The enzymes were produced through cultivation and protein expression in shaking flasks, followed by cell disruption by sonication and heat purification. In order to remove unbound metals, samples of the metal binding site muteins were desalted on a PD-10

column (GE Healthcare Life Sciences, Buckinghamshire, UK) prior to the bioconversion reaction. All other enzymes were used directly after heat purification. The reaction conditions for the bioconversion reactions were derived from Hajnal et al[16]. The reactions were done using a biphasic reaction system in 2 mL microreaction tubes. 1 mg of heat purified enzyme in 850 µL phosphate buffer (NaPi 50 mM pH 7) was supplemented with 150 µL of ethyl acetate containing pre-dissolved substrate to yield a final substrate concentration of 50 mM. The reactions were started by adding *tert*-butyl hydroperoxide to a final concentration of 135 mM and starting the agitation. The tubes were shaken in a thermomixer at 1,000 rpm and 30 °C for 14 hours. To stop the reactions, the tubes were cooled on ice and remaining *tert*-butyl hydroperoxide was quenched by adding 50-100 mg of sodium bisulfite. In the substrate screening experiment, no quenching was done to the samples that were expected to produce aldehydes as reaction products (i.e. the substrates isoeugenol, *trans*-anethole and 2-methyl-1-phenyl-1-propene), because aldehydes are known to form insoluble bisulfite adducts with sodium bisulfite. Afterwards, the product was extracted at first with 350 µL and a second time with 500 µL ethyl acetate. The ethyl acetate used for extraction was premixed with 20 mM of n-decanol as internal standard for the subsequent GC measurement. The combined organic phases were dried over Na_2SO_4 and used for the GC and GC-MS analysis.

Sodium bisulfite, n-decanol, ethyl acetate (99.9 %), the substrates α-methylstyrene (99 %), *trans*-anethole (99 %), 2-methyl-1-phenyl-1-propene (99 %) and isoeugenol (98 %) as well as the reference products *p*-anisaldehyde (98 %), benzaldehyde (99 %) and acetophenone (99 %) were obtained from Sigma-Aldrich. The substrate *p*-Cl-acetophenone (97 %) and the reference product *p*-Cl-α-methylstyrene (95 %) were obtained from Lancaster (Alfa Aesar, Ward Hill, MA, USA).

2.10.3 GC(MS)

GC analysis was performed using an Agilent 7890B GC system and a HP-5 19091J-413 column (Agilent Technologies). The oven program for the GC is given in Table 2.

Table 2: Temperature profile used for GC analysis

	Rate (°C/min)	Value (°C)	Hold Time (min)	Run Time (min)
Initial		80	0.5	0.5
Ramp 1	10	160	0	8.5
Ramp 2	20	280	2	16.5

To quantify the conversions, the peak areas of the product peaks were normalized with the peak area of the internal standard (n-decanol) and compared to a standard curve. Conversions were calculated as percentage of the theoretical maximum yield. Standard curves for substrates and products consisted of at least four samples measured in duplicates and ranging from concentrations of 5 mM to 50 mM. The substrate concentrations after the reaction could not be used to calculate the conversion because of a non-enzymatic substrate loss caused by polymerization of the styrene compound.

The GC-MS analysis was carried out on a 7890A GC System (Agilent Technologies), equipped with 5975C inert XL MSD mass spectrometer (Agilent Technologies). For the GC-MS analysis, different oven programs were used. All 4 temperature programs, along with the samples they were used for, are listed in Table 3.

Table 3: Temperature programs used in the GC-MS analysis. All samples were first measured with program 1 and additionally with other programs if necessary (as indicated).		Rate (°C/min)	Value (°C)	Hold Time (min)	Run Time (min)
Program 1 (Standard, all samples)	Initial		100	0.5	0.5
	Ramp 1	10	300	0	20.5
Program 2 (isoeugenol)	Initial		200	0.5	0.5
	Ramp 1	5	300	2	22.5
Program 3 (2-methyl-1-phenyl-1-propene, t-anethole)	Initial		80	1	1
	Ramp 1	20	300	1	13
Program 4 (indole)	Initial		100	0.5	0.5
	Ramp 1	15	300	1	15.833

2.11 Colorimetric detection of tryptophan and indole oxidation

The detection was done in microtiter plates with one well containing 0.2 mg of heat purified enzyme, 5 mM indole or tryptophan as substrate and 5 mM or 27 mM of *tert*-butyl hydroperoxide as oxidizing agent in a total reaction volume of 200 μL NaPi 50 mM pH 8 as reaction buffer. Detection was done in a plate reader, following the absorption over the lapse of 30 minutes. Tryptophan oxidation was measured at a wavelength of 500 nm and a temperature of 25 °C, indole oxidation was measured at a wavelength of 520 nm and 400 nm at a temperature of 29 °C. In the course of the measurement of the indole oxidation, the temperature decreased from 29.2 °C in the beginning, to 27.4 °C after 30 minutes.

D/L-tryptophan and indole were obtained from Sigma-Aldrich. Detection was done on an EON plate reader (BioTek Instruments GmbH).

2.12 Chemicals

Fine chemicals were obtained from Carl Roth GmbH if not stated otherwise. Distilled nuclease-free water (Aqua Bidest. "Fresenius") was obtained from Fresenius Kabi AG (Bad Homburg vor der Höhe, Germany).

3 Results

3.1 Assay development

In this thesis, several site directed mutants and site saturation libraries of Cupin 2 were generated. In order to quickly screen the hereby produced enzyme variants, a high throughput screening assay for the detection of the alkene cleavage products was developed.

The first attempt was a direct UV-detection of acetophenone, the main product of the model reaction. The absorption spectra of acetophenone and the corresponding substrate α-methylstyrene were measured. The measured absorption maxima were at a wavelength of 245 nm for acetophenone and 239 nm for α-methylstyrene. The absorption spectra are shown in Figure 6. The close proximity between the two absorption maxima makes it impossible to distinguish between product and substrate, which makes an assay based on UV absorption impossible in this case.
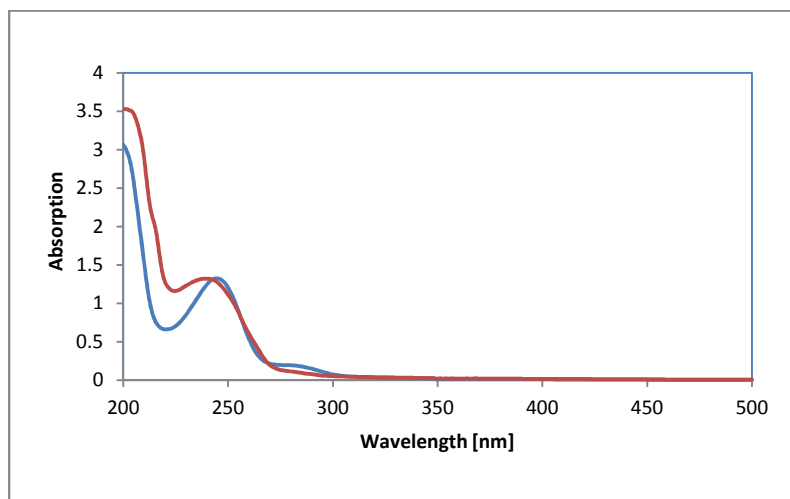

Figure 6: Absorption spectra of α-methylstyrene (red) and acetophenone (blue) at a concentration of 0.1 mM.

The second attempt was a colorimetric detection based on the aldol condensation of ketones with vanillin (described for acetone by Amlathe and Gupta, 1990[18]). The reaction was found to also work with the model products acetophenone and p-Cl-acetophenone, yielding a yellow colored complex, while the corresponding alkene substrates showed no color-formation. The

27

absorption maximum of the colored complex with acetophenone was measured at 442 nm. The full spectrum is shown in Figure 7. A picture of the color development of different acetophenone concentrations over time is shown in Figure 8. The absorption values of the microtiter plate shown in this figure were measured in a plate reader and the results are shown in Figure 9. Figure 10 shows the linear increase of absorption values with increasing acetophenone concentrations in the same microtiter plate after an incubation time of 35 min. For this experiment, 150 µL of a solution consisting of different concentrations of acetophenone (product to be detected), different concentrations of α-methylstyrene (substrate), 25 mM *tert*-butyl hydroperoxide (oxidizing agent later used in the reaction), *E. coli* lysate (to simulate the background, 1 mg/mL final protein concentration) and buffer (NaPi 50 mM, pH 7) were mixed with 40 µL of detection mix (5 M NaOH, 2.5 % vanillin, 50 % Ethanol). Because this detection mix tended to precipitate, further optimizations of the composition was needed.
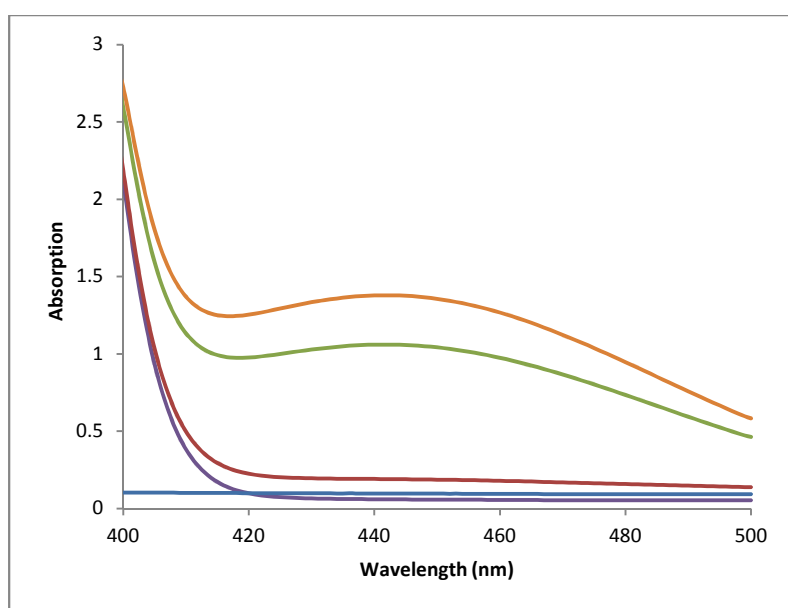

Figure 7: Absorption spectrum of the color-complex formed by acetophenone (10 mM) and vanillin at different points in time (red: 1 min, green: 22 min, orange: 30 min). For reference, a sample without acetophenone (purple) and a sample of acetophenone without detection mix (blue) were measured along with the samples. The detection mix consisted of 0.25% vanillin in 1 M NaOH.
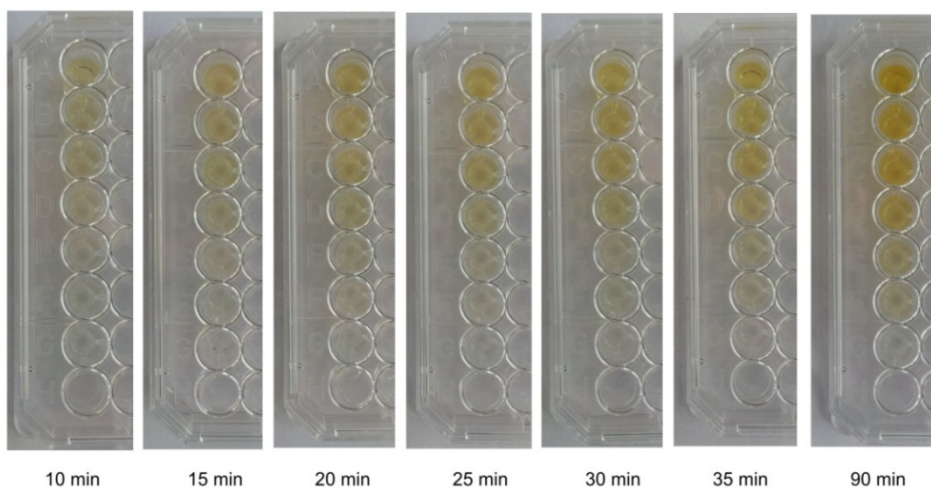
Figure 8: Color development of the acetophenone-vanillin reaction with different acetophenone concentrations. Concentrations (from top to bottom): 10 mM, 8 mM, 6 mM, 4 mM, 2 mM, 1 mM, 0 mM. Reaction conditions: 4.5 µL *tert*-butyl hydroperoxide 70 % solution in water (final concentration 25 mM), 0-15 µL 0.1 M acetophenone stock (final concentration 0-10 mM), 0-15 µL 0.1 M α-methylstyrene stock (final concentration 0-10 mM), 15 µL *E. coli* BL-21 (DE3) Gold lysate with overexpressed Cupin 2 (final protein concentration 1 mg/mL) and 115.5 µL NaPi 50 mM pH 7.0 were mixed, yielding 150 µL of simulated Cupin 2 reaction mix per well. 40 µL of detection mix consisting of 5 M NaOH, 2.5 % vanillin and 50 % ethanol were added to start the color reaction.

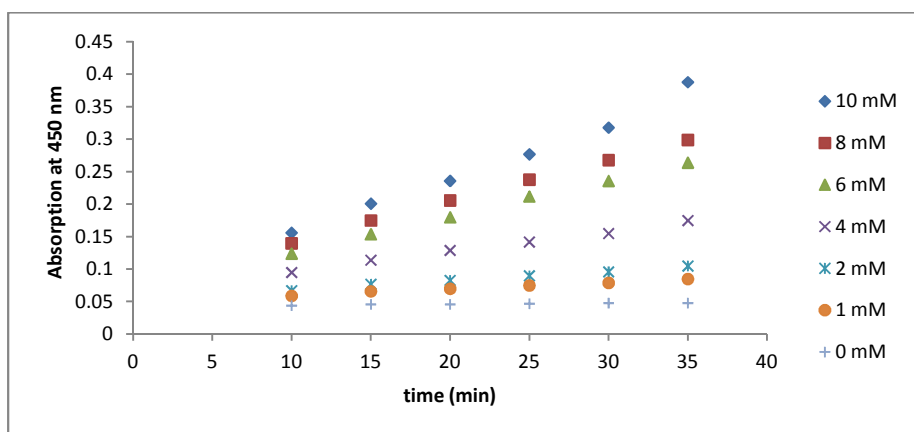**Figure 9: Color development of the color complex formed by different acetophenone concentrations with vanillin over time. Absorption was measured at 450 nm in a microtiter plate. For detailed reaction conditions see Figure 8.**
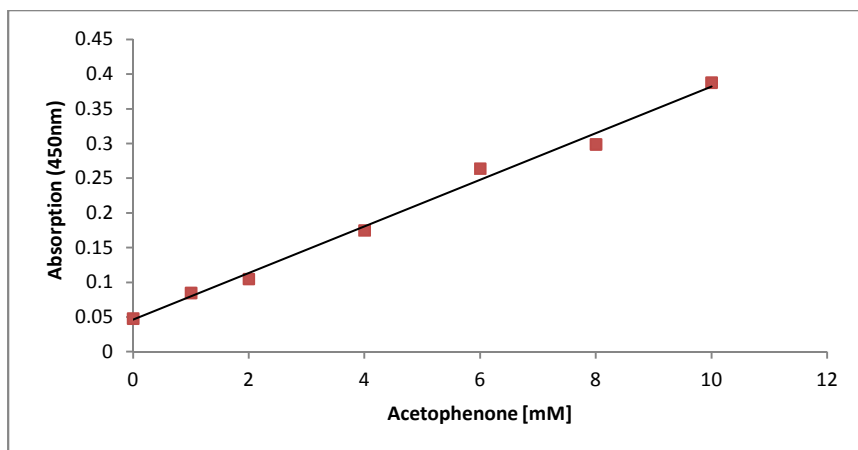

Figure 10: Absorption of samples with acetophenone concentrations from 0-10 mM measured at 450 nm, 35 min after addition of detection mix. For detailed reaction conditions see Figure 8.

29

The further optimized final detection mix consisted of 1 volume of vanillin stock (350 mg/mL in ethanol) and 9 volumes of 2.2 M NaOH. Detection of different acetophenone concentrations with the final detection mix is shown in Figure 11.
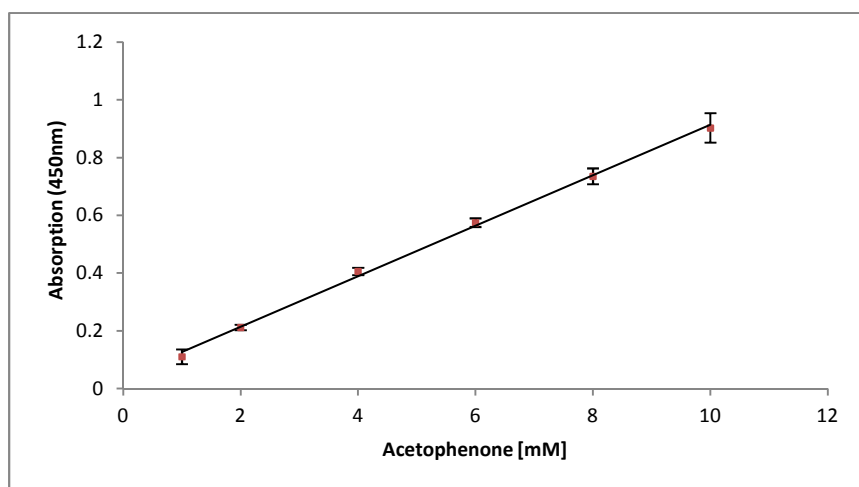

Figure 11: Different acetophenone concentrations detected with the final detection mix. Every point was measured in triplicates. Black bars represent the standard deviation. Reaction conditions: in a microtiter plate, 150 µL of NaPi (50 mM pH 7) per well containing acetophenone concentrations of 1-10 mM were mixed with 50 µL of the final detection mix and incubated for 30 minutes at room temperature. Absorption was measured at 450 nm.

In the observed concentration range of 0–10 mM of acetophenone, the assay showed great linearity and reproducibility. The final assay was done in microtiter plates with 130 µL lysate from a deepwell plate fermentation of *E. coli* BL-21 (DE3) Gold, overexpressing Cupin 2 and derived muteins. For the detailed procedure of the final assay see the "Materials and Methods" section.

3.2 First mutagenesis round - site directed mutagenesis

In order to verify their role in enzyme activity, amino acids of the active site and metal binding site were exchanged through site directed mutagenesis, as described in the materials and methods section. For a detailed view of the active site and metal binding site, see Figure 2 (section 2.1). All variants obtained by site-directed mutagenesis were cultivated in deepwell plates and screened by the vanillin assay. This was done using one column (8 wells) per variant, resulting in cultivation and screening in biological octuplicates. The screening results are shown in Figure 12.
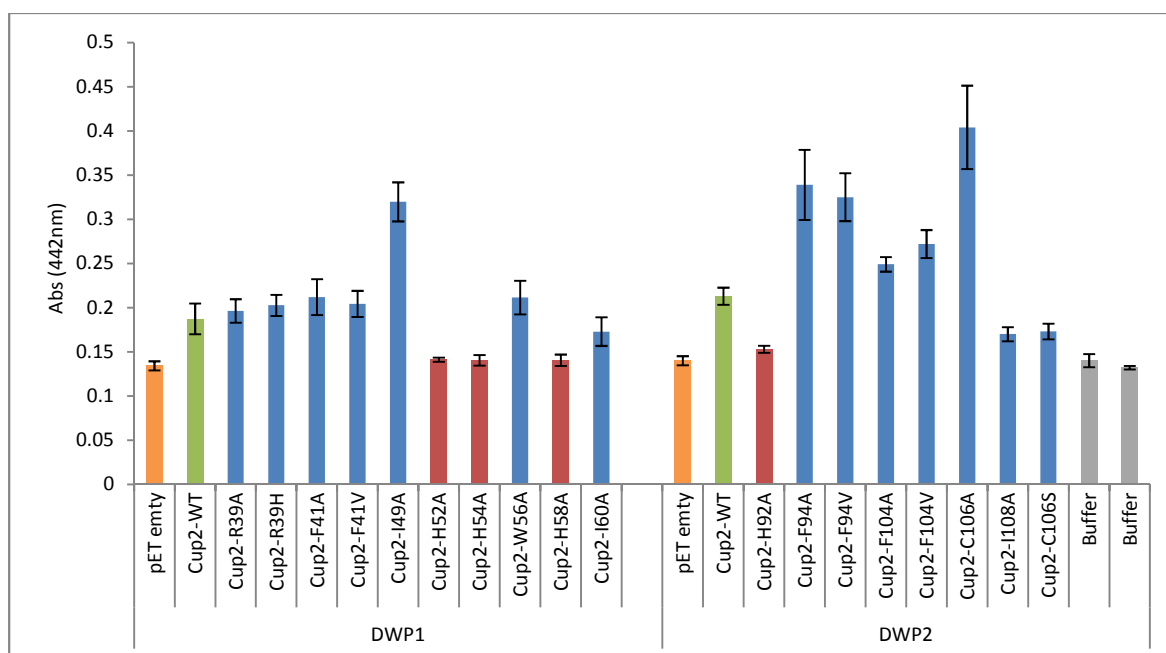
Figure 12: Summary of the screening results of the site directed muteins of the first mutagenesis round. Orange bars: empty vector control; Green bars: wild type control; Red bars: variants with an amino acid exchange in the metal binding site; Grey bars: buffer control. Values are given as mean value of biological octuplicates with standard deviation. Screening was done using the vanillin assay. For detailed reaction conditions, see "Materials and Methods" section 2.8.

It is clearly visible, that all four mutations of the metal binding site amino acids resulted in absorption rates comparable to the empty vector. This indicates a total lack of activity presumably caused by the loss of the manganese binding capabilities of the enzyme. Most of the other variants were on the same level as the wild type, but there was also a set of variants that showed a significant increase in activity, namely Cupin 2-I49A, -F94A, -F94V, -F104A, -F104V and -C106A.

It has to be considered that the crude lysates were directly taken from a deepwell plate cell lysis without prior normalization, so increased acetophenone concentrations may have been caused by higher expression levels instead of higher enzyme activities. In order to verify the potential differences in cell growth and overexpression, all enzyme variants were analyzed via SDS-PAGE. Figure 13 shows the results of the soluble and insoluble protein fractions of the first round of site directed mutagenesis, with lysate samples drawn from the re-screening deepwell plate fermentation. It is visible on the gels, that all variants were highly overexpressed at a comparable level. Moreover, no significant amounts of enzyme were found in the insoluble pellet fraction, indicating a stable enzyme and no formation of inclusion bodies. According to these results, the enzyme concentrations were on the same level for all variants, so the activity increase was most likely caused by increased activity of the newly generated muteins. All variants with an amino acid exchange at positions F41 and F104

showed unexpected double bands whose origin could not be determined. The empty vector control also showed a weaker but significant band at the height of Cupin 2. The only explanation for that is a contamination of this sample with one of the samples overexpressing Cupin 2.
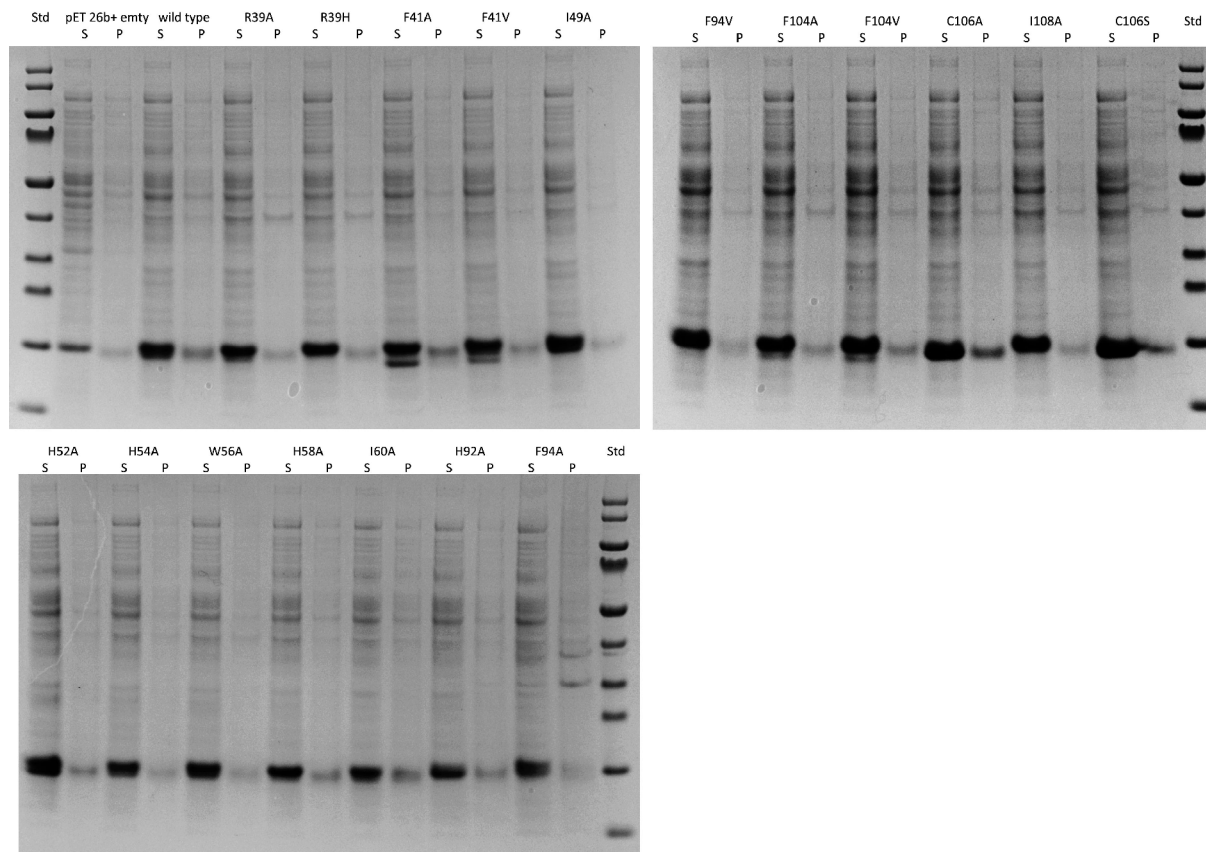

Figure 13: SDS-PAGE gels showing protein expression of all variants from the first mutagenesis round. Samples were derived from deepwell plate scale fermentation. Std: PageRuler Prestained Protein Ladder; S: supernatant; P: pellet.

3.3   Second mutagenesis round - site saturation mutagenesis

Site saturation libraries of all active site amino acids were generated in order to screen for variants with increased activity. For details about the library generation, see "Materials and Methods" sections 2.5 and 2.6

3.3.1   Verification of library quality

The sequencing results of five clones which were sequenced right after the transformation of each library into *E. coli* TOP10F' showed ambiguous results. Some samples clearly indicated the presence of a plasmid mix of different mutants and some could not be sequenced at all. Because of that, these sequencing results did not provide a proper insight into the codon distribution of the libraries and the verification of the library quality was solely based on the sequencing results gained from the clones that were sequenced after the transformation into *E.*

coli BL-21 (DE3) Gold. A detailed evaluation of the codon distribution in the site saturation libraries is given in Table 4.

Table 4: Summary of the codon distribution in the five plasmids per library that were sequenced after transformation in *E. coli* BL21 (DE3) Gold. Results are given as amino acids in one letter code followed by the respective codon.

R39X (WT:CGT)		F41X (WT:TTT)		I49X (WT:ATT)		W56X (WT:TGG)		I60X (WT:ATT)	
	R (CGT)		S (AGT)		I (ATT)		P (CCG)		L (TTG)
	P (CCG)		M (ATG)		H (CAT)		R (CGT)		P (CCT)
	S (TCG)		R (CGT)		S (TCT)		C (TGT)		R (CGT)
	S (AGT)		V (GTT)		R (CGG)		Stop(TAG)		C (TGT)
	S (AGT)		P (CCT)		D (GAT)		S (AGT)		1 failed
F94X (WT:TTT)		F104X (WT:TTT)		C106X (WT:TGT)		I108X (WT:TGT)			
	A (GCG)		A (GCT)		T (ACG)		L (CTG)		
	I (ATT)		L (CTT)		E (GAG)		S (TCG)		
	G (GGG)		C (TGT)		D (GAT)		P (CCG)		
	R (CGG)		T (ACT)		G (GGT)		T (AGG)		
	K (AGG)		W (TGG)		P (CCG)		1 failed		

The results indicated that the library at position R39 may be biased with serine while all other libraries showed an even distribution of codons. It was assumed that the high oversampling factor of the libraries could widely compensate for the bias and so library R39X was used for screening in spite of the allegedly uneven codon distribution.

3.3.2 Library screening

All nine site saturation libraries were screened for acetophenone production using the vanillin assay. An example of a microtiter plate after detection is shown in Figure 14. Differences in color intensity were clearly visible even to the naked eye. The corresponding absorption values for the same plate are shown in Table 5.
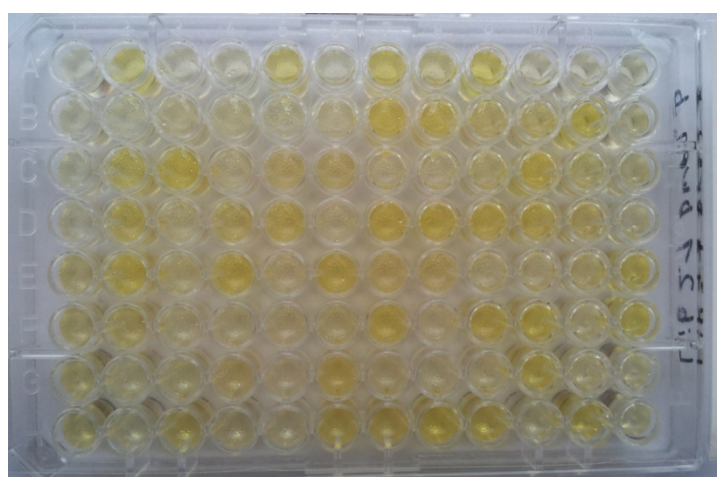
Figure 14: Example for a microtiter plate from library screening. The first column consisted of four negative controls (wells 1A-1D), containing lysate of *E. coli* BL-21 (DE3) Gold carrying the empty expression vector, and four positive controls (wells 1E-1H), containing *E. coli* BL-21 (DE3) Gold expressing the wild type enzyme. The picture shows the plate 35 minutes after addition of the detection solution. Table 5 shows the absorption values of the same plate generated by a plate reader. Screening was done using the vanillin assay. For detailed reaction conditions, see "Materials and Methods" section 2.8.

Table 5: Example of absorption values of a single microtiter plate in library screening, measured in a plate reader at 442 nm. The first column contained four negative and four positive controls (see Figure 14). Screening was done using the vanillin assay. For detailed reaction conditions, see "Materials and Methods" section 2.8. A photo of the same plate is also shown in Figure 14.

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.144	0.492	0.15	0.159	0.502	0.143	0.584	0.326	0.577	0.163	0.148	0.156
B	0.143	0.141	0.202	0.225	0.153	0.167	0.644	0.538	0.304	0.307	0.61	0.222
C	0.144	0.482	0.616	0.17	0.314	0.317	0.154	0.158	0.217	0.578	0.273	0.264
D	0.145	0.47	0.169	0.387	0.51	0.171	0.528	0.603	0.529	0.494	0.334	0.178
E	0.289	0.557	0.21	0.574	0.18	0.534	0.321	0.307	0.162	0.16	0.255	0.492
F	0.293	0.327	0.175	0.203	0.215	0.253	0.522	0.175	0.492	0.572	0.169	0.524
G	0.284	0.177	0.22	0.298	0.175	0.454	0.213	0.166	0.176	0.53	0.167	0.18
H	0.281	0.163	0.403	0.226	0.173	0.514	0.439	0.579	0.463	0.159	0.458	0.175

The variants with the highest increase in absorption compared to the wild type were re-screened in septuplicates. The variants with the highest gain in absorption in the re-screening were in turn sent to sequencing. The sequencing results for all libraries are summarized in Table 6. Each library resulted in at least one amino acid exchange with improved activity. In several libraries, the best variants were identified multiple times.

Table 6: Overview of amino acid substitutions and the corresponding codons of all sequenced library hits.

Amino acid positions (with corresponding wild type codon)

Amino Acid	R39X (WT:CGT)	F41X (WT:TTT)	I49X (WT:ATT)	W56X (WT:TGG)	I60X (WT:ATT)	F94X (WT:TTT)	F104X (WT:TTT)	C106X (WT:TGT)	I108X (WT:ATT)	possible codons:
Alanine			GCG (2x)	GCT		GCT, GCG (x4)				GCG, GCT
Arginine				AGG (2x) CGG					CGT AGG	CGT, CGG, AGG, AGT
Asparagine		AAT		AAT						AAT
Aspartic Acid	GAT			GAT						GAT
Cysteine				TGT	TGT (4x)					TGT
Glutamic Acid	GAG (3x)			GAG (2x)				GAG		GAG
Glutamine								CAG (2x)		CAG
Glycine		GGT	GGG (3x) GGT (3x)	GGG		GGG				GGG GGT
Histidine	CAT									CAT
Isoleucine										ATT
Leucine		CTT				TTG CTG	TTG (3x) CTG (3x) CTT (3x)	TTG (2x) CTG (5x) CTT	CTG CTT	CTT CTG TTG
Lysine										AAG
Methionine										ATG
Phenylalanine										TTT
Proline	CCT (9x) CCG (3x)								CCT (x9) CCG (x3)	CCT CCG
Serine		AGT TCT	TCT (3x) AGT (2x)	AGT		TCT (2x) TCG (2x)				AGT, TCT, TCG
Threonine						ACT (x2) ACG (x3)				ACT ACG
Tryptophan	TGG									TGG
Tyrosine										TAT
Valine								GTT (x3) GTG		GTT GTG

Amino acid substitutions and corresponding codons

There were also additional mutations inserted by random PCR errors that were selected in the screening. One of the emerging variants showed a truncation at the C-terminus, which was caused by a deletion in the codon at amino acid position 106 that induced a frameshift which in turn introduced a stop codon at amino acid position 107. The hereby generated variants Cupin 2-C106S-L107* and Cupin 2-C106G-L107* (* for STOP codon) showed a significantly increased absorption in the re-screening experiments. Two variants with two amino acid exchanges found in library screening are Cupin 2-A33V-R39H and Cupin 2-W56A-L105M, both also showed increased activity in the re-screening.

After the hits were sequenced, the increase in absorption for every sequenced clone was calculated from the absorption data of the re-screening plates. Table 7 shows the increase in absorption at 442 nm compared to the wild type. The corresponding SDS-PAGE gels are shown in Figure 15. It has to be taken into account that these values are prone to fluctuations caused by differences in expression levels and cell growth between different wells and different variants. Moreover, it has to be noted that the number of samples for every clone was different because some mutations appeared multiple times in the re-screening.

Table 7: Increase in absorption at 442 nm of all library hits according to re-screening. Given is the percentage increase of all sequenced variants (mean value of septuplicates on one screening plate) compared to the wild type (mean value of septuplicates on the same screening plate). Screening was done using the vanillin assay. For detailed reaction conditions, see "Materials and Methods" section 2.8.

amino acid exchange	% increase (±standard deviation)	amino acid exchange	% increase (±standard deviation)	amino acid exchange	% increase (±standard deviation)
I49A	120(±9) %	F94S	45(±9) %	W56N	27(±1) %
I49S	92(±16) %	I108P	43(±12) %	I108L	26(±8) %
R39P	84(±27) %	F94G	43(±12) %	W56G	25(±2) %
I49G	84(±16) %	W56E	41(±2) %	W56R	21(±3) %
R39D	65(±11) %	I108R	40(±14) %	F41S	19(±9) %
R39W	64(±5) %	F104L	36(±14) %	W56C	18(±7) %
C106V	59(±25) %	F94T	35(±8) %	F41G	13(±8) %
C106L	59(±21) %	F94L	34(±7) %	W56S	13(±2) %
F94A	50(±9) %	C106Q	31(±16) %	F41N	12(±10) %
W56D	48(±6) %	F41L	28(±8) %		
R39E	47(±17) %	I60C	27(±6) %		

Figure 15 shows SDS-PAGE gels of the soluble fractions of variants with one amino acid exchange from the first round of mutagenesis, with lysate samples drawn from the re-screening deepwell plate fermentations. Again, all variants showed a high overexpression but the overexpression level as well as the overall protein concentration per lane were vastly differing between different variants. Some mutations seemed to alter the protein in a way that caused a slight shift in the position of its band. For example, the band of Cupin 2-F94A

35

appeared higher than the band of Cupin 2-F94L. Another unexpected observation was the clearly visible double band of Cupin 2-I108R.
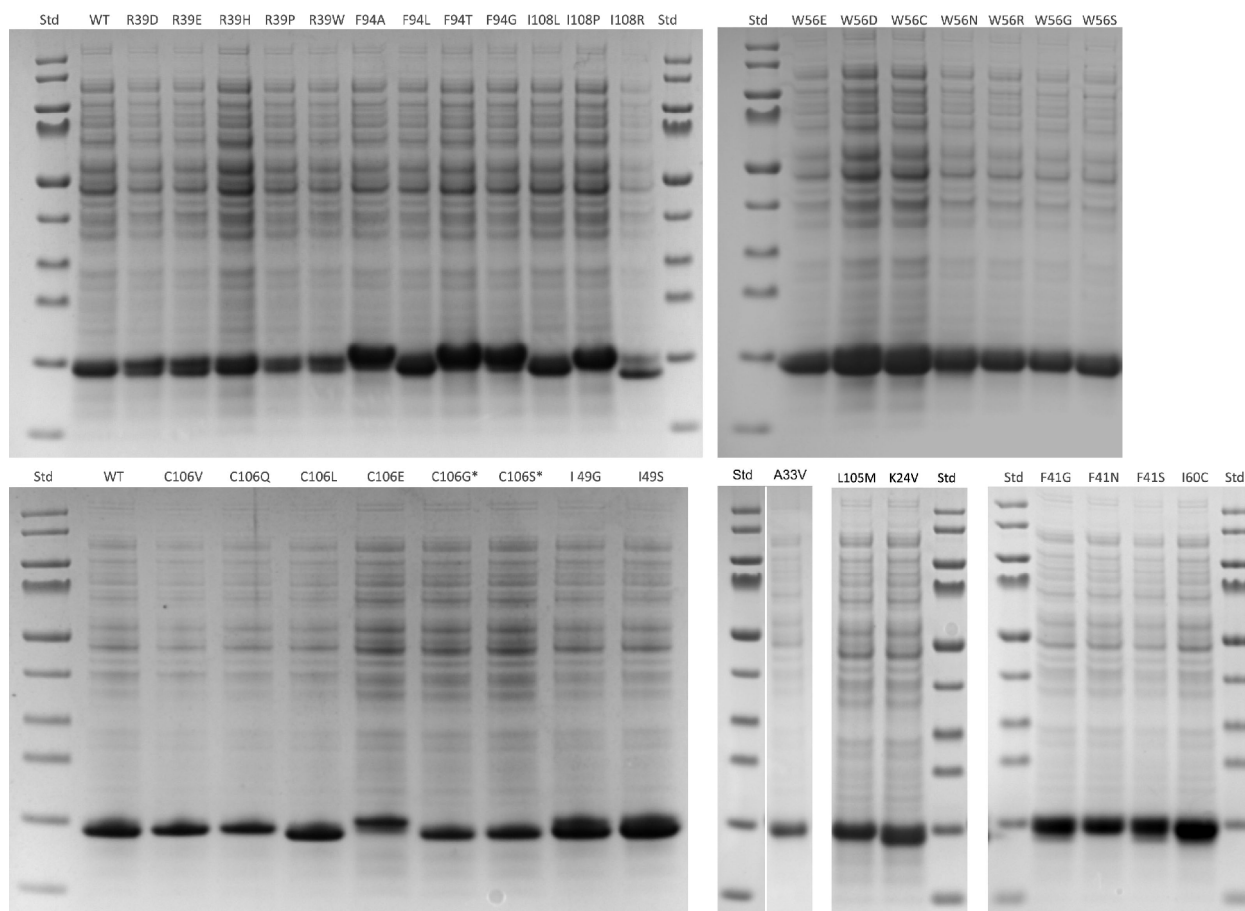
Figure 15: SDS-PAGE gels showing the protein expression of the most promising variants from library screening. Samples were derived from deepwell plate scale fermentation. Std: PageRuler Prestained Protein Ladder. *: mutation to stop-codon at amino acid position 107 (resulting in a C-terminally truncated enzyme).

3.4 Third mutagenesis round – combination of different mutations

Based on the results of the library screening, new variants with one, two or three amino acid exchanges, as well as C-terminally truncated variants and multi-muteins were generated (for details, see "Materials and Methods" section 2.5) and screened using the vanillin assay. Table 8 shows a list of all variants generated in this last round of mutagenesis.

Table 8: List of all variants generated in course of the third mutagenesis round. Asterisks represent mutations to a stop-codon.

One amino acid exchange	C-terminus variants
L105M	P109*
A33V	L107*
Two amino acid exchanges	C106V_L107*
I49A_R39P	I108R_P109*
I49A_F94A	C106V_I108R_P109*
I49A_C106V	C106Q_L107*
I49A_W56D	L105M_C106Q_L107*
I49A_F104L	L105M_C106Q_L107M_I108*
I49A_I108R	L105M_C106V_L107*
F94A_R39P	**Multiple amino acid exchanges ("multi-muteins")**
F94A_C106V	R39P_F41L_I49A_W56D_F94A_F104L_C106V_I108P
F94A_W56D	R39P_I49A_W56D_F94A_C106V_I108R
F94A_F104L	R39P_I49A_F94A_C106V
F94A_I108R	I49A_F94A_C106V_L107*
F94G_I49A	I49A_F94A_L105M_C106Q_L107*
Three amino acid exchanges	I49A_F94A_L105M_C106Q_L107M_I108*
I49A_F94A_I108P	I49A_F94A_L105M_C106V_L107*
I49A_F94A_L105M	I49A_F94A_C106Q_L107*

For the variants with one amino acid exchange, vanillin screening showed no significant increase in activity for Cupin 2-A33V and a significant increase of activity for Cupin 2-L105M (+31±6 % compared to the wild type).

All variants with two amino acid exchanges except for Cupin 2-I49A-I108R resulted in increased activity compared to the wild type, but only Cupin 2-I49A-F94A also showed a significant increase compared to the corresponding variant with single amino acid exchange. A graphical overview of the absorption values of variants with two amino acid exchanges compared to the Cupin 2-I49A and Cupin 2-F94A variants is shown in Figure 16. The corresponding SDS-PAGE gels are shown in Figure 18.
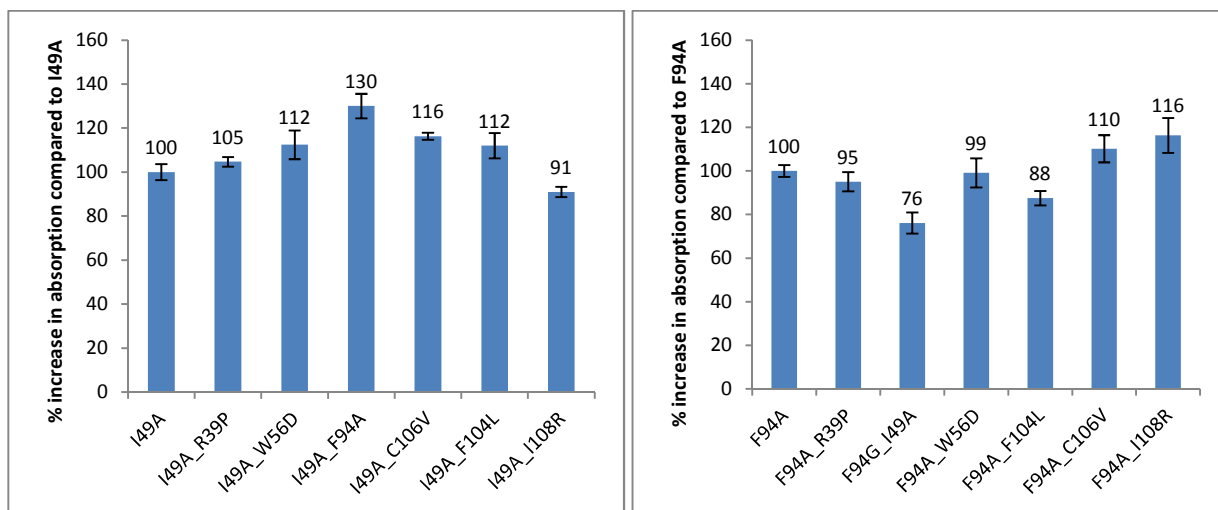
The enzyme variants that were C-terminally truncated at position L107 and P109 both showed an increased activity compared to the wild type. Cupin 2-C106V, which was C-terminally truncated at position L107, showed the best activity among all C-terminally truncated enzyme variants. The combination of I108R and a truncation at position P109 also increased the activity compared to both underlying variants and the addition of a C106V amino acid exchange (resulting in the Cupin 2-C106V-I108R-P109* variant) further increased activity. The screening results of all C-terminally truncated variants are shown in Figure 17, the corresponding SDS-PAGE gels are shown in Figure 18.

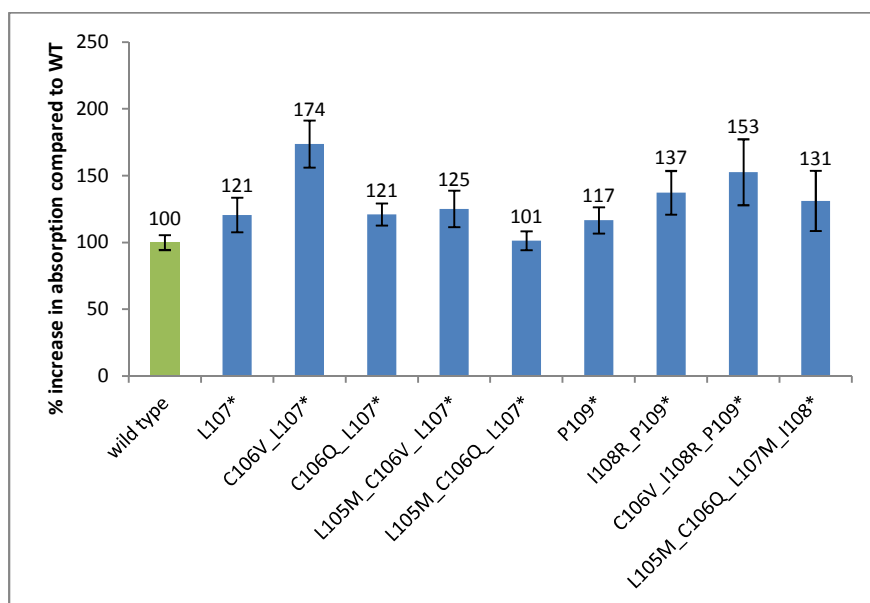**Figure 17: Activity increase of C-terminally truncated variants compared to the wild type. Values are given in percentage increase compared to the wild type. *: mutation to stop-codon (resulting in a C-terminally truncated enzyme). Measurements originate from microtiter plate vanillin assay screening in septuplicates. For detailed reaction conditions, see "Materials and Methods" section 2.8.**

The variant Cupin 2-I49A-F94A-I108P showed decreased activity compared to the wild type. The observation of high amounts of insoluble protein during cell lysis (data not shown) indicates that the lowered activity was caused by a lower concentration of soluble enzyme. The low enzyme concentration is clearly visible in the SDS-PAGE gels (see Figure 18). Cupin 2-I49A-F94A-L105M showed an increase in activity compared to the wild type but a slight decrease compared to the underlying Cupin 2-I49A-F94A. Multi-mutein "Multi1" (Cupin 2-R39P-F41L-I49A-W56D-F94A-F104L-C106V-I108P) showed activity at the same level as the wild type while all further multi-muteins showed no detectable activity at all. Again, high amounts of insoluble protein was observed during cell lysis (data not shown), indicating that the lack of activity is caused by a low concentration of soluble enzyme. The low concentration of soluble enzyme was again confirmed by an SDS-PAGE analysis (see Figure 18).

Figure 18 shows SDS-PAGE gels of the soluble fractions of lysates overexpressing enzyme variants from the third round of mutagenesis. On these gels, different overexpression levels for the different combinations can be observed. Of the proteins with two and three amino acid exchanges, especially Cupin 2-I49G-F94A and -I49A-F94A-I108P showed a much weaker band. Also the expression level of all multi-muteins seemed to be impaired. Furthermore, the double band previously observed for Cupin 2-I108R also appeared in most combined variants containing the I108R or I108P amino acid exchange. The only exception was the multi-mutein "Multi2" (Cupin 2-R39P-I49A-W56D-F94A-C106V-I108R) that only showed a single band. It can also be observed, that variants containing the F94A amino acid exchange were generally stronger expressed compared to those containing the I49A amino acid exchange. This can be seen by comparing Cupin 2-I49A-R39P with Cupin 2-F94A-R39P, Cupin 2-I49A-C106V with Cupin 2-F94A-C106V, Cupin 2-I49A-W56D with Cupin 2-F94A-W56D, Cupin 2-I49A-F104L with Cupin 2-F94A-F104L or Cupin 2-I49A-I108R with Cupin 2-F94A-I108R.
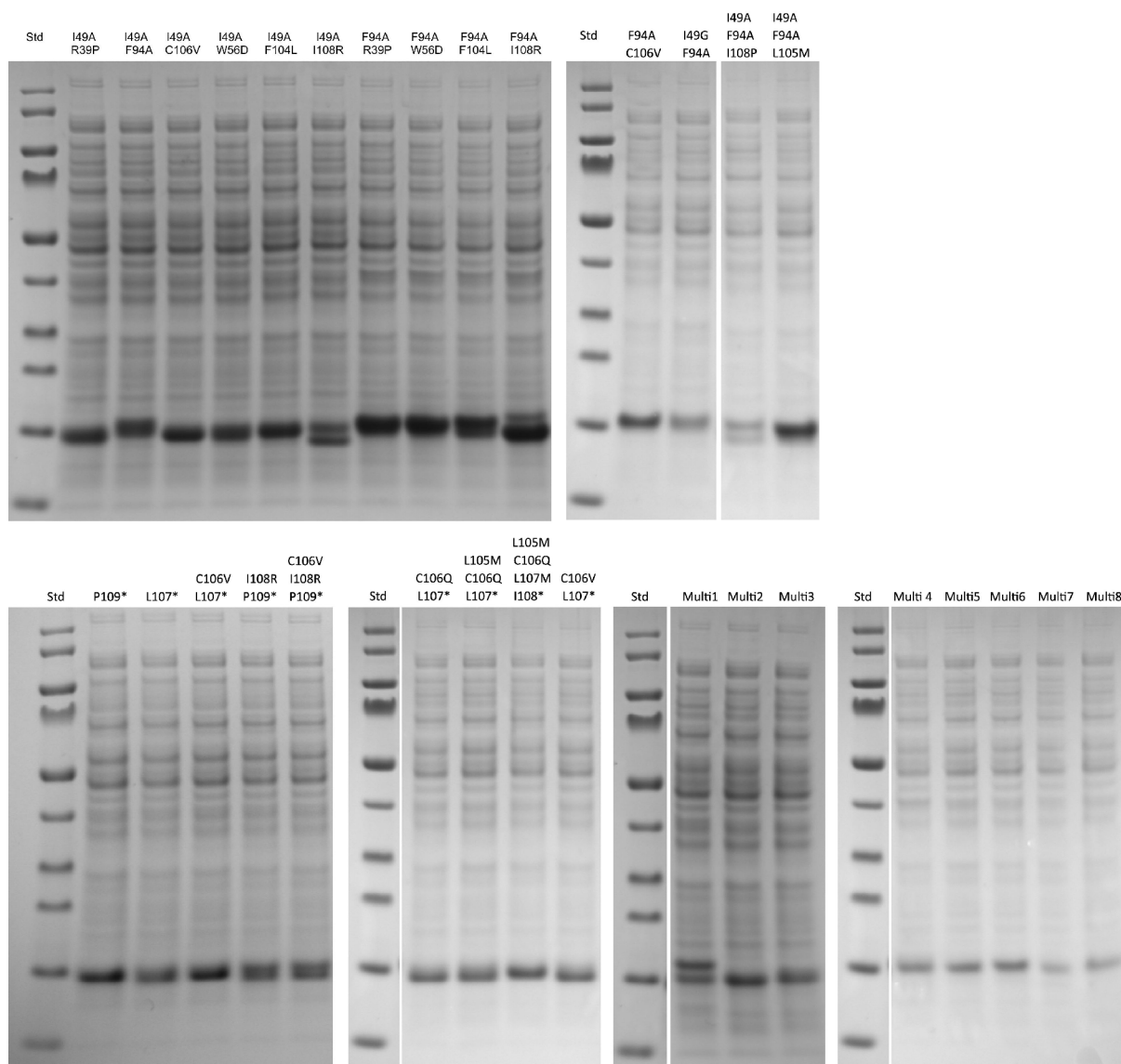
Figure 18: SDS-PAGE gels showing the protein expression of variants with multiple amino acid exchanges as well as C-terminally truncated variants. Samples were derived from deep-well plate scale fermentation. *: mutation to stop-codon (resulting in a C-terminally truncated enzyme) Std: PageRuler Prestained Protein Ladder. Multi-muteins:
Multi1: Cupin 2-R39P-F41L-I49A-W56D-F94A-F104L-C106V-I108P
Multi2: Cupin 2-R39P-I49A-W56D-F94A-C106V-I108R
Multi3: Cupin 2-R39P-I49A-F94A-C106V
Multi4: Cupin 2-I49A-F94A-C106V-L107*
Multi5: Cupin 2-I49A-F94A-L105M-C106Q-L107*
Multi6: Cupin 2-I49A-F94A-L105M-C106Q-L107M-I108*
Multi7: Cupin 2-I49A-F94A-L105M-C106V-L107*
Multi8: Cupin 2-I49A-F94A-C106Q-L107*

The most promising Cupin 2 variants from the screening experiments were produced in larger scale in shaking flasks fermentations. The generated lysates were heat purified and used in the bioconversion reactions.

3.5 Protein purification by heat denaturation

Cupin 2 originates from the thermophilic bacterium *Thermotoga maritima,* which results in a very high heat stability of this protein. Therefore, crude lysate originating from *E. coli* overexpressing Cupin 2 can be purified by heat denaturation of the *E. coli* protein background. Figure 19 shows the outcome of a purification experiment at 70 °C and 75 °C over a period of 25 min. Precipitated protein was removed by centrifugation (5 min, 20,000x g). It is clearly visible from this picture, that Cupin 2 was stable under these harsh conditions, while most of the *E. coli* protein degraded and precipitated.
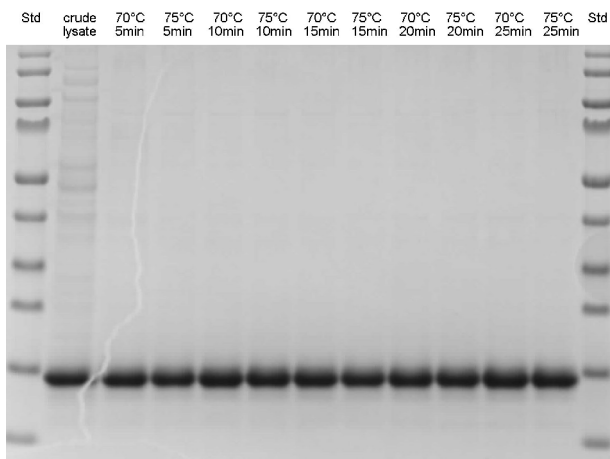
Figure 19: SDS-PAGE gel of a heat purification experiment of Cupin 2 crude lysate (12 mg/mL total protein) at 70 °C and 75 °C at several points in time over a period of 25 minutes. After the denaturation time, the samples were centrifuged for 5 minutes at 20,000x g and the supernatant was used for this SDS-PAGE. The samples were diluted 1:25. Std: PageRuler Prestained Protein Ladder.

A further experiment showed that the enzyme stays soluble up to temperatures of 80 °C. An SDS-PAGE of this experiment is shown in Figure 20. It is clearly visible in the samples which were heated for 20 minutes, that Cupin 2 started to degrade at a temperature of 85 °C.
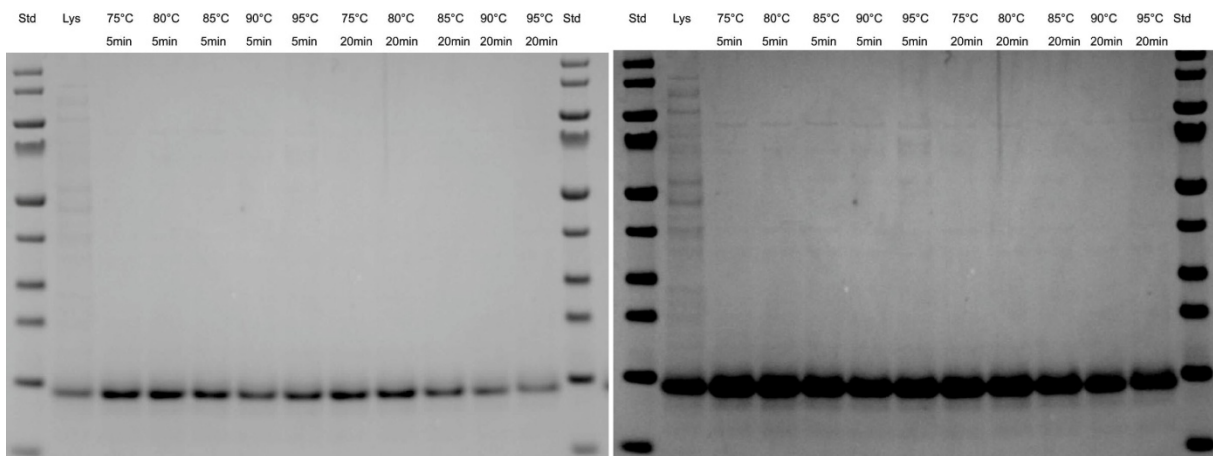
Figure 20: Heat purification of Cupin 2 lysate (12 mg/mL total protein) at different temperatures for 5 minutes and 20 minutes. Left picture: Lower contrast to visualize the fading of the bands starting at 85 °C. Right picture: Higher contrast to visualize the remainders of the *E. coli* protein background. After the denaturation time, the samples were centrifuged for 10 minutes at 20,000x g and the supernatant was used for this SDS-PAGE. The samples were diluted 1:25. Std: PageRuler Prestained Protein Ladder. Lys: Crude lysate prior to purification.

The same samples were tested for their catalytic activity via vanillin assay. As shown in Figure 21, the activity decreased at samples that showed a lowered enzyme concentration after treatment with temperatures over 80 °C, but stayed at the same level for the samples treated with 75 °C or 80 °C.
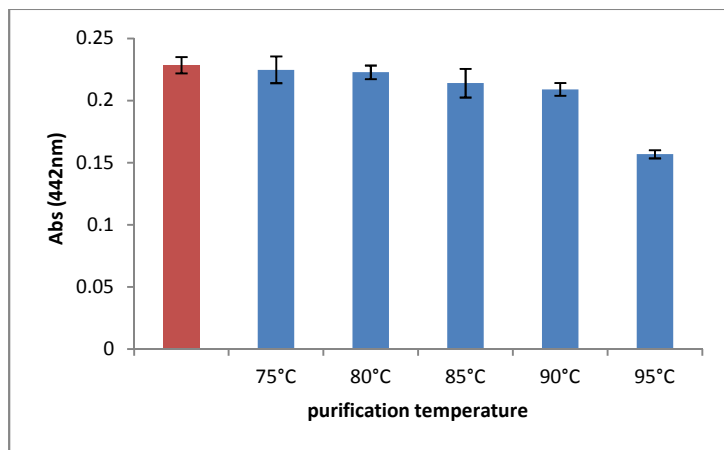
Figure 21: Activity determination of Cupin 2 lysate after 20 min of heat purification at temperatures ranging from 75 °C to 95 °C (blue bars) compared to the crude lysate (red bar). The same volume of lysate was used for all samples. Screening was done using the vanillin assay. For detailed reaction conditions, see "Materials and Methods" section 2.8. Deviation: Standard deviation of quadruplicates.

Prior to the bioconversion reactions, lysates of the variants with one amino acid exchange that were produced in shaking flask fermentation were heat purified, normalized to 1 mg/mL protein concentration and again screened by vanillin assay. The results are shown in Figure 22.
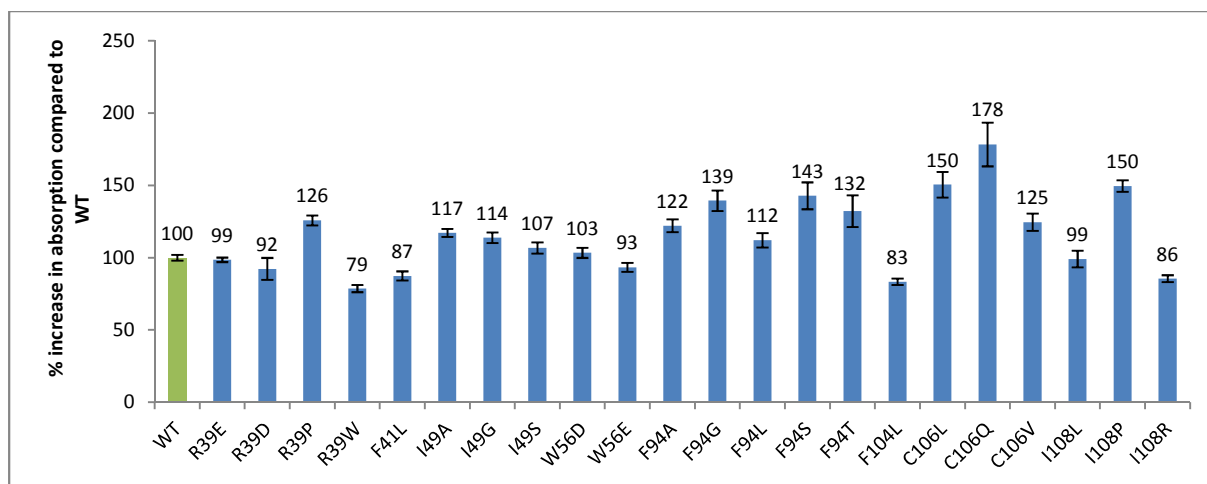
Figure 22: Activity of variants with one amino acid exchange relative to the wild type. All samples were heat purified and normalized to 1 mg/mL protein concentration. Measurements originate from microtiter plate vanillin assay screening in septuplicates. For detailed reaction conditions, see "Materials and Methods" sections 2.8, for purification conditions see section 2.10.1.

The absorption values show that not all results from the re-screening experiments could be reproduced with normalized reaction conditions. This may be due to differences in enzyme concentration in the re-screening experiments or due to a loss of specific activity during heat

denaturation. To verify the existence of such an activity loss, the activity of crude and heat purified lysate was compared by measuring a second set of reactions along with the heat purified samples. In this second set, the purified lysate was substituted with the same volume of crude lysate, generating samples with the same dilution rate but different total protein concentrations. The reached values for the variants were ranging from 73 % to 100 % of remaining activity. Detailed results are shown in Table 9.

Table 9: Comparison of variants with one amino acid exchange concerning their activity left after heat purification. Remaining activity after purification given in % of the activity before the purification. For detailed purification conditions, see "Materials and Methods" section 2.10.1.

amino acid exchange	(%)	amino acid exchange	(%)
WT	94	F94G	88
R39E	81	F94L	75
R39D	91	F94S	93
R39P	82	F94T	86
R39W	87	F104L	75
F41L	86	C106L	93
I49A	73	C106Q	100
I49G	73	C106V	95
I49S	79	I108L	93
W56D	86	I108P	86
W56E	79	I108R	101
F94A	85		

3.6   Bioconversion reactions

In order to quantify the increase in activity of the generated muteins, bioconversion reactions in 1 mL scale were conducted and the conversions were measured by GC. The crude lysates originating from shaking flask fermentations were purified using the heat purification protocol described in the "Materials and Methods" section 2.10.1.

3.6.1   Purification of the lysates

All lysates that were used in the bioconversion reactions were before purified using the heat purification protocol. To verify their stability, SDS-PAGE gels comparing the same volumes of crude and purified lysates were performed. Figure 23 shows the gels of the variants with a single amino exchange that were selected for the bioconversion reactions. There was no significant difference in band intensity of Cupin 2, which means that no significant decrease in structural stability was caused by the mutations.
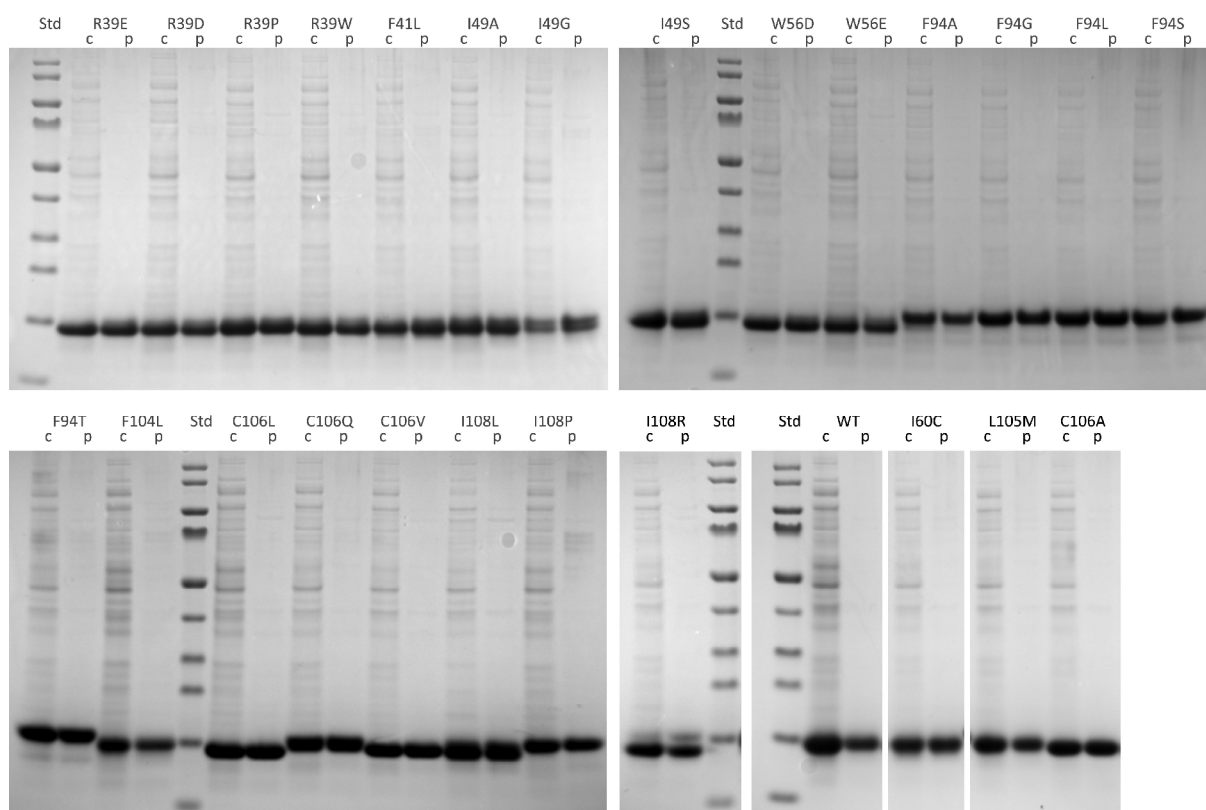
Figure 23: SDS-PAGE gels of selected variants with one amino acid exchange, comparing crude lysate (c) and purified lysate (p). For the crude lysate samples, a volume corresponding to 5 µg of total protein was taken from lysate originating from shaking flask fermentation. After heat purification, the same volume of purified lysate was used as comparison. Std: PageRuler Prestained Protein Ladder. For detailed purification conditions, see "Materials and Methods" section 2.10.1.

Figure 24 shows the SDS-PAGE gels of variants with one, two or three amino acid exchanges as well as C-terminally truncated variants that were selected for the bioconversion reaction. All tested variants with one or two amino acid exchanges as well as the C-terminally truncated Cupin 2-C106Q-L107* seemed to have an unimpaired heat stability while all other truncated enzyme variants showed a significant decrease of band intensity after purification, indicating a lowered heat stability. Because of the lowered heat stability of the C-terminally truncated variant Cupin 2-L105M-C106Q-L107*, the concentration after purification was not sufficient for it to be used in a bioconversion reaction.

Figure 24: SDS-PAGE gels of selected variants with one, two or three amino acid exchanges as well as C-terminally trunc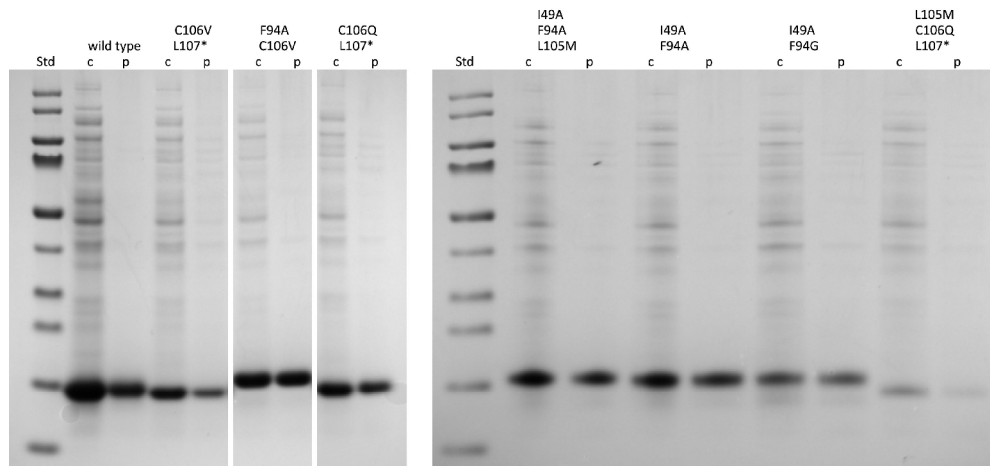ated variants, comparing crude lysate (c) and purified lysate (p). For the crude lysate samples, a volume corresponding to 5 µg of total protein was taken from lysate originating from shaking flask fermentation. After heat purification, the same volume of purified lysate was used as comparison. *: mutation to stop-codon (resulting in a C-terminally truncated enzyme). Std: PageRuler Prestained Protein Ladder. For detailed purification conditions, see "Materials and Methods" section 2.10.1.

3.6.2 Cleavage of α-methylstyrene by Cupin 2 variants

The substrate used in the bioconversion reactions was α-methylstyrene. All samples were measured in triplicates. Because of problems with the internal standard (see discussion) and the standard curves, distinct conversions could not be given for all samples. But to be able to compare the reactions either way, all conversions are additionally given as values relative to the wild type enzyme. Figure 25 shows the conversions of a set of variants with one amino acid exchange found in library screening. Especially the variants Cupin 2-C106L and -C106Q resulted in greatly improved conversions with a twofold increase compared to the wild type enzyme. Figure 26 shows the absolute conversions of some of these variants.
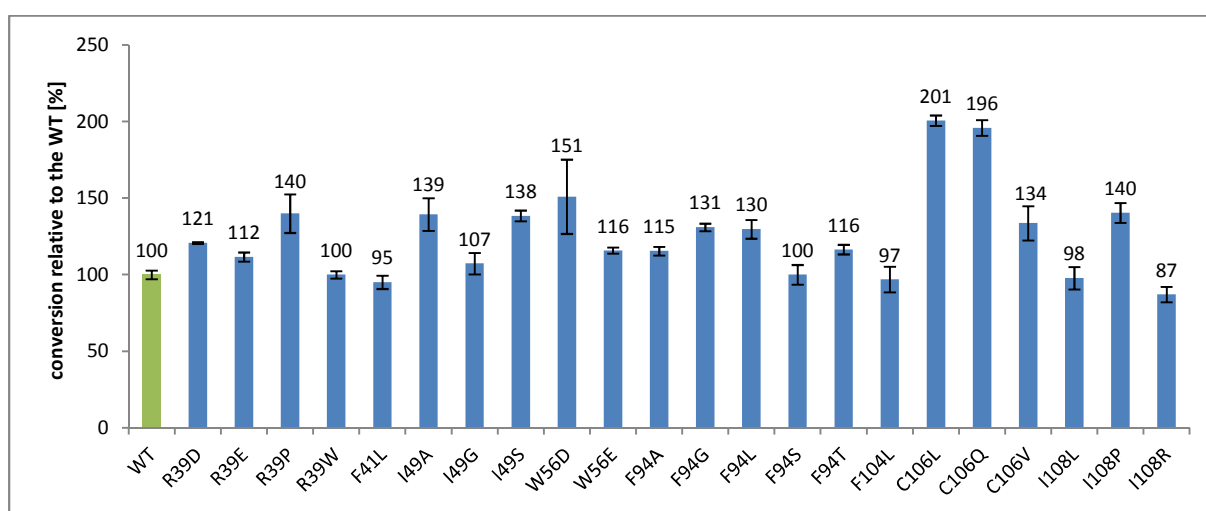

Figure 25: Relative conversions of a set of variants with one amino acid exchange compared to the wild type. Conversion measured by GC. Deviation: Standard deviation of triplicates. For detailed reaction conditions see "Materials and Methods" section 2.10.2, for detailed GC conditions see section 2.10.3.

45
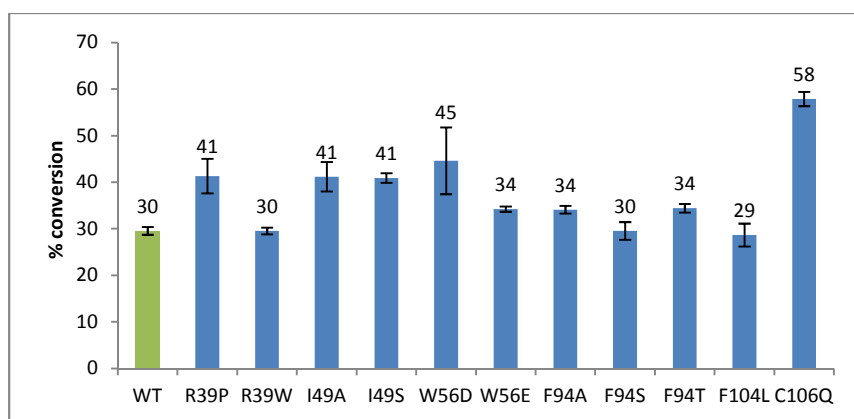
Figure 26: Absolute conversions of one batch of variants with one amino acid exchange measured by GC. Deviation: Standard deviation of triplicates. For detailed reaction conditions see "Materials and Methods" section 2.10.2, for detailed GC conditions see section 2.10.3.

Because the absolute conversions of several variants could not be assessed due to the previously mentioned problems, the best hits from the first GC measurement round were again used in bioconversion reactions and measured again. The results relative to the wild type are shown in Figure 27-A, the results of the absolute conversions are given in Figure 27-B.
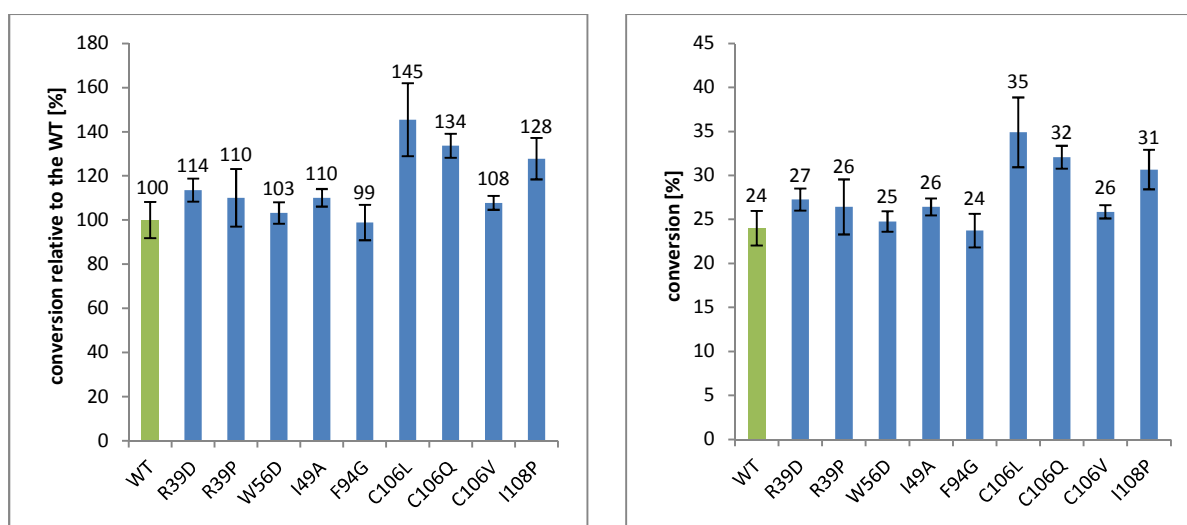
Figure 27-A (left) & Figure 27-B (right): Conversions relative to wild type (left) and absolute conversions (right) of the repeated bioconversion reactions of the most promising variants. Deviation: Standard deviation of triplicates. For detailed reaction conditions see "Materials and Methods" section 2.10.2, for detailed GC conditions see section 2.10.3.

Unfortunately, the results from the first measurement round could not be reproduced. The conversions were overall lower in the repetition of the bioconversion reactions.

Figure 28-A shows another batch of bioconversion reactions measured by GC. Especially the C-terminally truncated Cupin 2-C106V-L107* and Cupin 2-I49A-F94A showed greatly increased conversions with a gain of 74 % and 79 % compared to the wild type enzyme. Cupin 2-I49A-F94A-L105M showed that the addition of another mutation to Cupin 2-I49A-F94A had a detrimental effect on the conversions. Furthermore, Cupin 2-I49A-F94A showed

46

a significantly better conversion than Cupin 2-I49A-F94G. Cupin 2-L105M, which was found by a random PCR error during library screening, showed a much increased conversion, which came as a surprise given the fact that this amino acid is not part of the active site. Figure 28-B shows the absolute conversions acquired for the same batch of reactions.
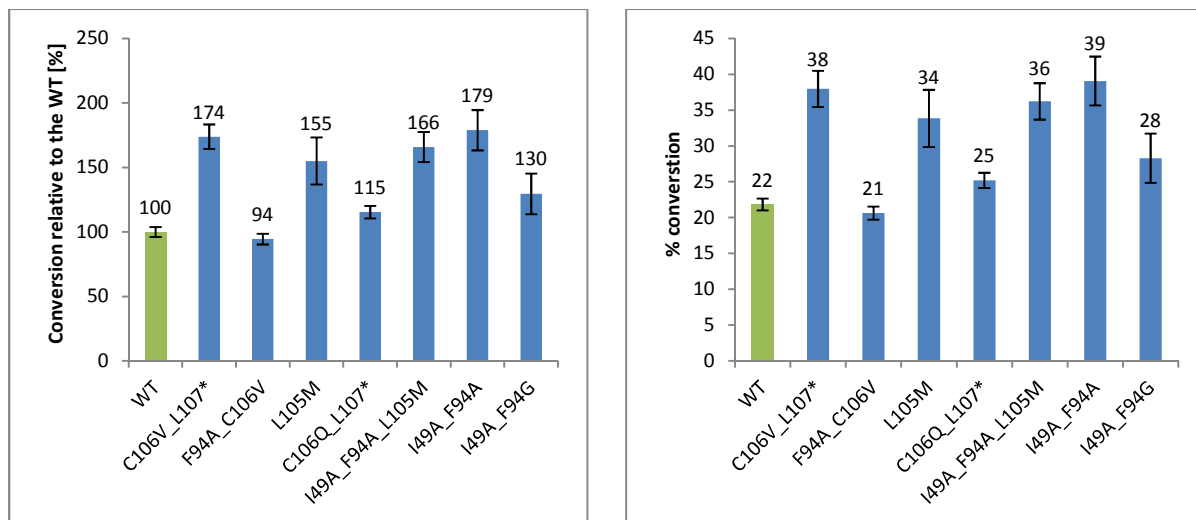
Figure 28-A (left) & Figure 28-B (right): Relative conversions compared to the wild type (left) and absolute conversions (right) of bioconversion reactions of the most promising variants with two or three amino acid exchanges, variants with C-terminal truncation as well as Cupin 2-L105M. Samples measured by GC. Deviation: Standard deviation of triplicates. For detailed reaction conditions see "Materials and Methods" section 2.10.2, for detailed GC conditions see section 2.10.3.

Because of the variations in the conversions, the experiment was again repeated with the most promising variants. The outcome given as relative conversions compared to the wild type is shown in Figure 29.
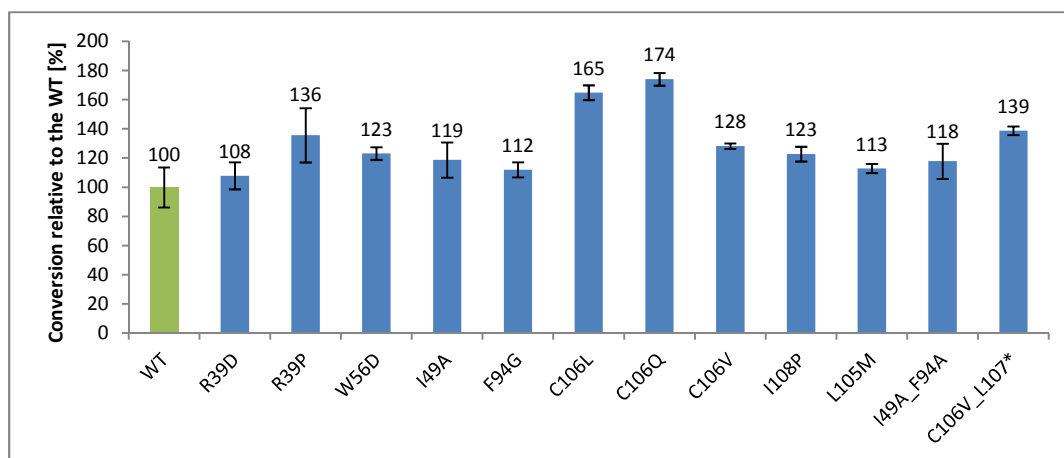
Figure 29: Relative conversions of bioconversion reactions of the most promising variants, measured by GC. Values given in percent compared to the wild type enzyme. Deviation: Standard deviation of triplicates. For detailed reaction conditions see "Materials and Methods" section 2.10.2, for detailed GC conditions see section 2.10.3.

3.6.3 Influence of amino acid exchanges in the metal binding site on the conversion

Figure 30 shows the conversion of variants with amino acid exchanges in the metal binding site. It is clearly visible from this diagram that Cupin 2-H58A shows much higher activity compared to other metal binding site variants that showed close to no activity at all. It was shown by metal analysis that Cupin 2-H58A in contrast to the other variants contained significant amounts of manganese (data not shown), which explains the remaining activity.
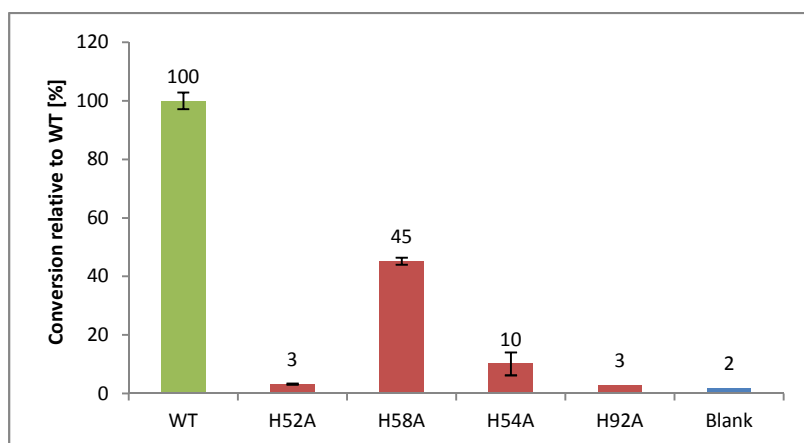
Figure 30: Conversions of all variants with an amino acid exchange in their metal binding site, given relative to the wild type. Conversion measured by GC. Blank: Reaction mix without enzyme. Deviation: Standard deviation of triplicates. For detailed reaction conditions see "Materials and Methods" section 2.10.2, for detailed GC conditions see section 2.10.3.

3.6.4 Progress curve of the reaction

To compare the reaction velocity of different variants in the bioconversion reactions, samples for 5 points in time were measured for the wild type as well as Cupin 2-C106Q, -I49A-F94A and -C106V-L107*. All samples were measured in duplicates. The hereby generated progress curve of the bioconversion reaction over the first 2 hours is shown in Figure 31. It is clearly visible from this data that the new variants were significantly faster than the wild type enzyme. After 2 hours, Cupin 2-C106V_L107* showed a conversion of 35 % while the wild type showed a conversion of 20 %. Additionally, the flattening of the curves shows a reduction of the enzyme activity over the course of the reaction, which could indicate an inactivation of the enzyme.
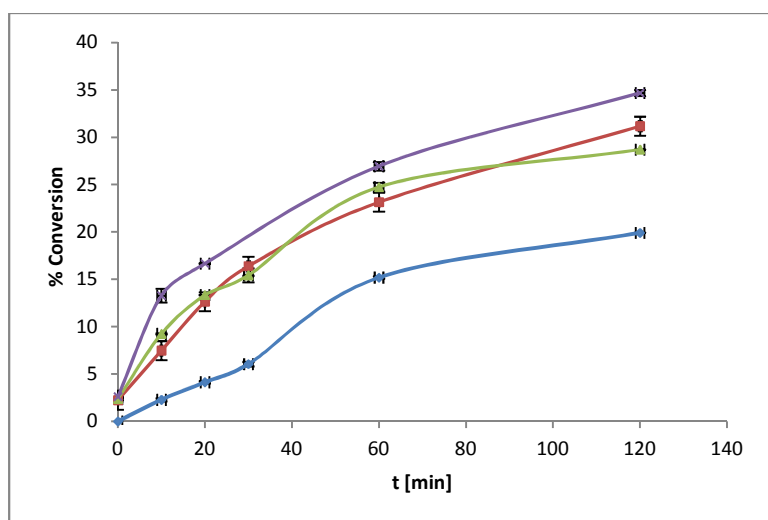
Figure 31: Conversion rates of the wild type (blue) and Cupin 2-C106Q (red), -I49A-F94A (green) and -C106V-L107* (purple) over the first 2 hours of the bioconversion reaction. Samples measured by GC. Deviation: Standard deviation of duplicates. For detailed reaction conditions see "Materials and Methods" section 2.10.2, for detailed GC conditions see section 2.10.3.

Figure 32 shows the loss of substrate caused by side-reactions and the extraction process. Non-enzymatic polymerization of the styrene derivatives is thought to be the main cause for substrate loss during the reaction. The loss was calculated by subtracting the molar product concentration and the molar concentration of the remaining substrate from the 50 mM of substrate initially present in the reaction. The data shows that for all samples, about 25 mM of the substrate was lost during the reaction. Interestingly, substrate loss seemed to be significant only in the first 30 minutes and evened out at a substrate concentration of about 25 mM, which indicates that the loss was caused by high substrate- or *tert*-butyl hydroperoxide concentrations.
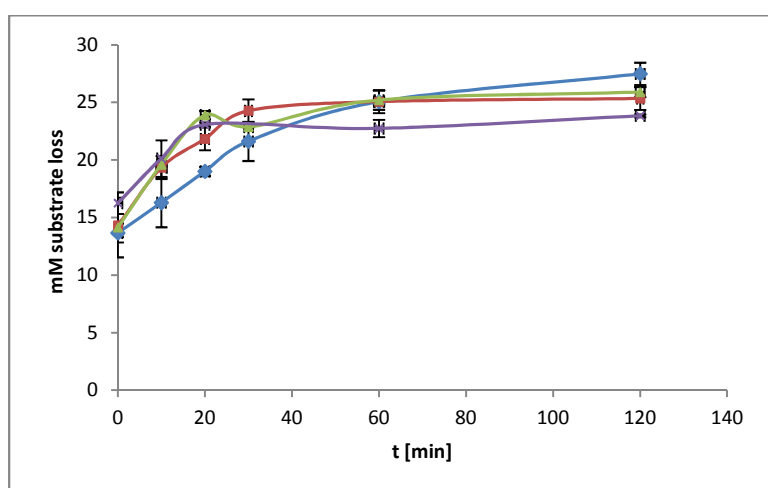
Figure 32: Substrate loss in the samples of the wild type (blue) and the Cupin 2-C106Q (red), -I49A-F94A (green) and -C106V-L107* (purple) due to enzymatic and non-enzymatic side-reactions in the bioconversion reactions over time. Deviation: Standard deviation of duplicates. For detailed reaction conditions see "Materials and Methods" section 2.10.2, for detailed GC conditions see section 2.10.3.

49

3.6.5 Modification of reaction conditions

To evaluate the potential for further optimizations of the reaction conditions, bioconversion reactions with different enzyme and *tert*-butyl hydroperoxide concentrations were performed. The change of the conversion and the recovery in these experiments is shown in Figure 33. When the concentration of *tert*-butyl hydroperoxide was doubled from 135 mM to 270 mM, the conversion as well as the recovery increased significantly. The same applies to the reactions comparing 1 mg and 2 mg of wild type enzyme, where the reaction with the doubled amount of enzyme shows a significantly increased conversion and recovery.
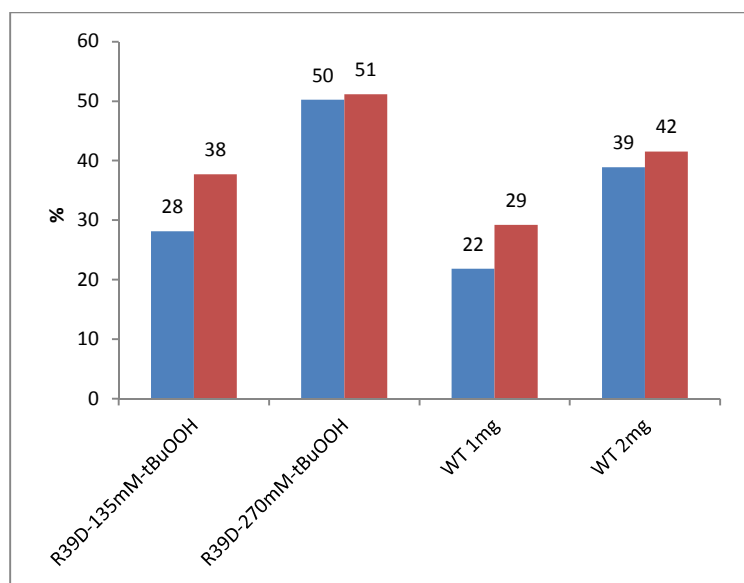
Figure 33: Conversions (blue) and recovery (red) of bioconversion reactions with altered reaction conditions. Shown is the comparison of a reaction of Cupin 2-R39D with 135 mM and 270 mM *tert*-butyl hydroperoxide (tBuOOH) and a comparison of a reaction with 1 mg and 2 mg of wild type enzyme (WT). The recovery is the percentage of the initial substrate concentration that could be recovered as substrate and product after the reaction. For detailed reaction conditions see "Materials and Methods" section 2.10.2, for detailed GC conditions see section 2.10.3.

3.6.6 Substrate screening

To evaluate the effect of the amino acid exchanges on the conversion of other substrates, molecules similar to α-methylstyrene were selected and tested as substrates in bioconversion reactions. For these reactions, Cupin 2-C106Q, -R39P, -I49A-F94A and -C106V-L107* were compared to the wild type enzyme.

Figure 34 shows the different substrates along with the corresponding products that were expected from the cleavage reaction.
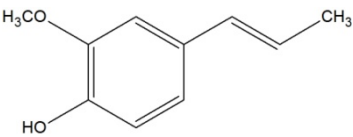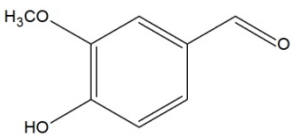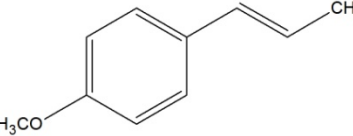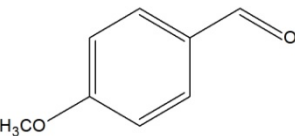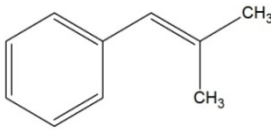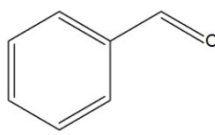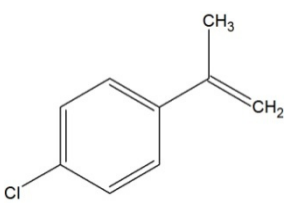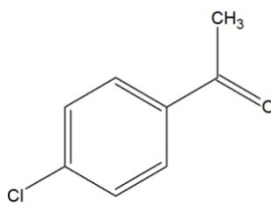
Figure 34: Alternative substrates used in the bioconversion reactions along with the expected product.

In the isoeugenol reactions, traces of vanillin could be detected in the reactions with Cupin 2-C106V-L107*, -C106Q and -I49A-F94A. There was also an unexpected product that was produced by all variants, but in significantly higher amounts by Cupin 2-C106V-L107*. In a GC-MS analysis, this product was later identified as dehydrodiisoeugenol (see Figure 36). The increase of conversion in the reaction forming this unexpected product compared to the wild type is shown in Figure 35. It has to be noted that the wild type indeed also produced dehydrodiisoeugenol, but only in barely detectable traces. Although the production of dehydrodiisoegenol by Cupin 2-C106V_L107* was greatly increased compared to the wild type, the absolute conversion estimated by comparing the peak areas, was still very low.
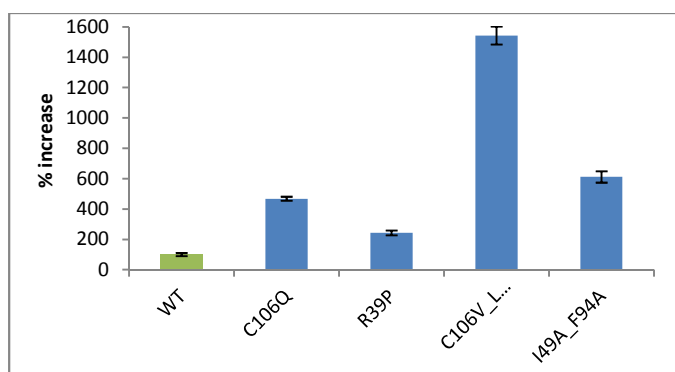
Figure 35: Increase of conversion in the reaction of isoeugenol to dehydrodiisoeugenol. Values are given as percentage increase of the product's peak area normalized with the internal standard and relative to the wild type (WT). Deviation: Standard deviation of duplicates. For detailed reaction conditions see "Materials and Methods" section 2.10.2, for detailed GC conditions see section 2.10.3.
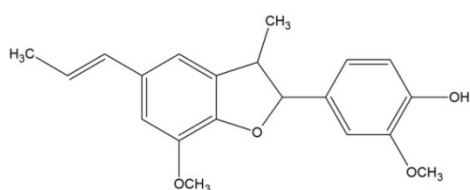
Figure 36: Structure of dehydrodiisoeugenol, the product of the Cupin 2 catalyzed reaction of two isoeugenol molecules.

The conversion from *trans*-anethole to *p*-anisaldehyde was 11 % with insignificant differences between the different enzyme variants. 2-Methyl-1-phenyl-1-propene was converted to benzaldehyde with conversions ranging from 2-4.6 % (see Figure 37-A). The difference in the conversions in this case was so low, that it may also have been caused by slight differences in enzyme concentration rather than by improved enzyme activities. *p*-Cl-α-methylstyrene was converted to *p*-Cl-acetophenone at conversions ranging from 38 % (wild type) to 66 % (Cupin 2-C106V-L107*). The conversions of all enzyme variants are shown in Figure 37-A for 2-methyl-1-phenyl-1-propene and in Figure 37-B for *p*-Cl-α-methylstyrene.
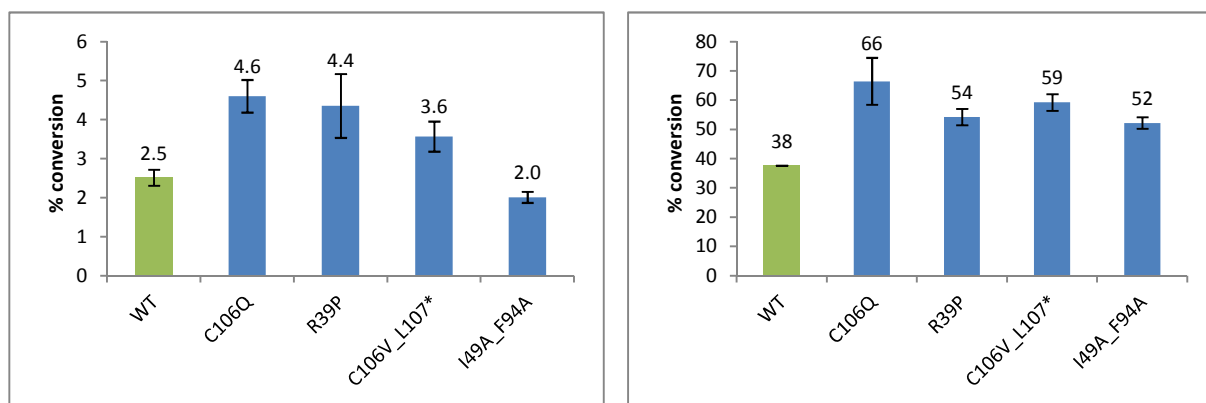
Figure 37-A (left) & Figure 37-B (right): Comparison of the conversions of the wild type enzyme with Cupin 2-C106Q, -R39P, -C106V-L107* and -I49A-F94A in the bioconversion of 2-methyl-1-phenyl-1-propene to benzaldehyde (left) and of *p*-Cl-α-methylstyrene to *p*-Cl-acetophenone (right). Deviation: Standard deviation of duplicates. For detailed reaction conditions see "Materials and Methods" section 2.10.2, for detailed GC conditions see section 2.10.3.

For the substrates *trans*-anethole and 2-methyl-1-phenyl-1-propene, traces of several by-products with oxidations on the double bond and the terminal CH_3 groups of the alkyl chain were identified. For *trans*-anethole, products with hydroxylation and carbonylation at the α-position, as well as carbonylation on the β-position were identified. For the 2-methyl-1-phenyl-1-propene, carbonylation and hydroxylation at the α-position as well as on the terminal CH_3 were identified. Structural formulas and IUPAC names of these by-products are given in Figure 38. For the substrates *p*-Cl-α-methylstyrene and α-methylstyrene, traces of epoxide by-products were identified (see Figure 39).

trans-anethole oxidation by-products:

trans-anethole (substrate) — 1-(4-methoxyphenyl)propan-1-ol — 1-(4-methoxyphenyl)propan-1-one — 1-(4-methoxyphenyl)propan-2-one

2-methyl-1-phenyl-1-propene oxidation by-products:

2-methyl-1-phenyl-1-propene (substrate) — 2-methyl-1-phenylprop-2-en-1-ol — 2-methyl-1-phenylprop-2-en-1-one — 2-methyl-3-phenylprop-2-en-1-ol

2-methyl-3-phenylprop-2-enal
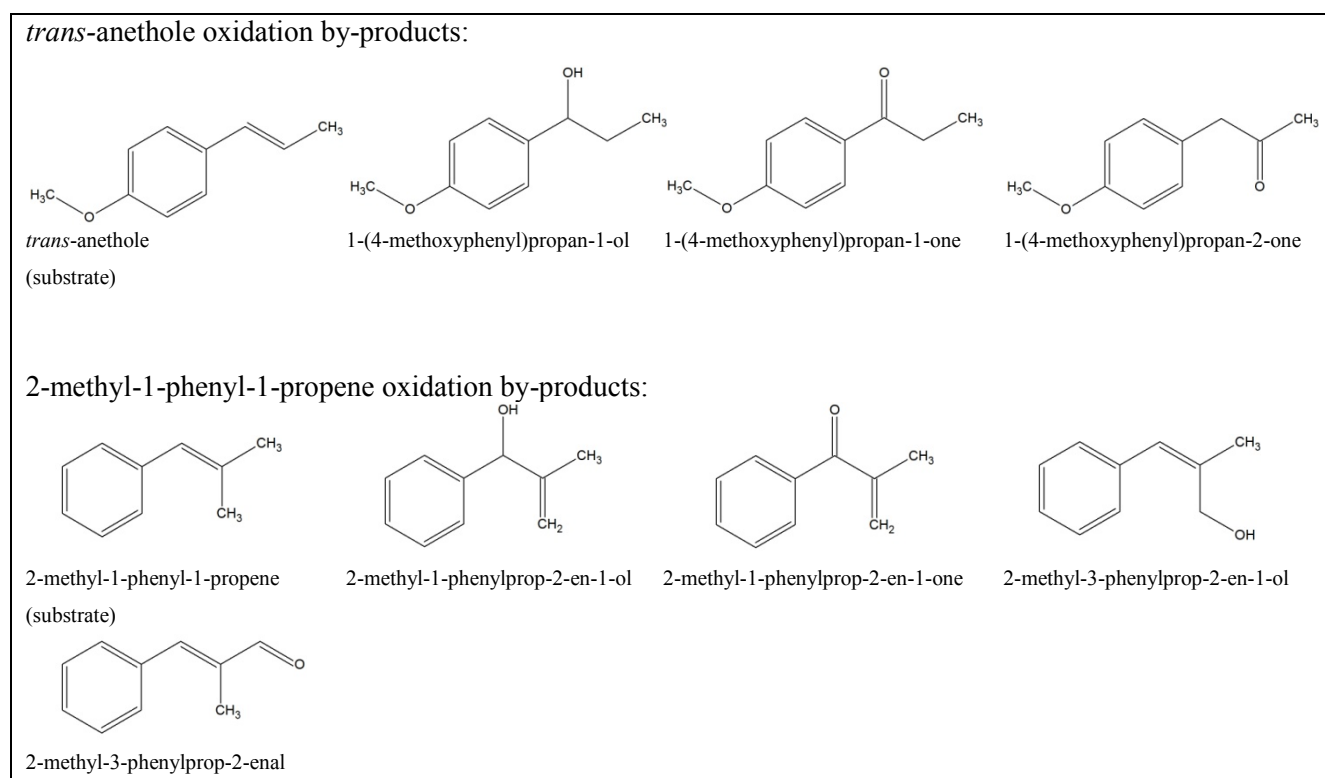
Figure 38: Carbonylated and hydroxylated by-products identified in the *trans*-anethole and 2-methyl-1-phenyl-1-propene reactions given as structural formulas along with IUPAC names. For detailed reaction conditions see "Materials and Methods" section 2.10.2, for detailed GC conditions see section 2.10.3.

Substrate:
R=H: α-methylstyrene
R=Cl: p-Cl-α-methylstyrene

Epoxide product:
R=H: 2-methyl-2-phenyloxirane
R=Cl: 2-(4-chlorophenyl)-2-methyloxirane

Figure 39: Epoxide by-products formed in the bioconversion reactions of α-methylstyrene and p-Cl-α-methylstyrene. For detailed reaction conditions see "Materials and Methods" section 2.10.2, for detailed GC conditions see section 2.10.3.

3.7 Tryptophan oxidation activity

In reaction mixes of the bioconversion reactions as well as the microtiter plate based screening reactions, the formation of a red color was observed. This unexpected phenomenon seemed to correlate with enzyme activity, i.e. variants with a higher activity showed a stronger color formation in their reaction broth. Moreover, it was observed that the color formation was stronger in reactions conducted with crude *E. coli* lysate than in lysates purified by heat purification. A theory that the red color originates from the enzymatic oxidation of tryptophan was proven by preparing a microreaction tube containing 1.7 mg heat purified Cupin 2-I108P variant, 3.9 μL *tert*-butyl hydroperoxide (27 mM) and 5 mg tryptophan in 1 mL NaPi buffer (50 mM, pH 8). After 2 minutes, a significant difference in color compared to a control reaction without added tryptophan was visible. Control reactions without enzyme or without *tert*-butyl hydroperoxide as oxidizing agent showed no sign of color formation.

Figure 40: 1.5 mL microreaction tubes containing 1.7 mg of heat purified Cupin 2-I108P and 3.9 μL *tert*-butyl hydroperoxide (27 mM) in 1 mL NaPi buffer (50 mM, pH 8). Left: reaction mix with 5 mg tryptophan. Right: control reaction without tryptophan. The picture was taken after 30 minutes at room temperature.

Figure 40 shows the color formation with added tryptophan compared to the control reaction without tryptophan.

A similar color formation was found to occur with indole instead of tryptophan. Figure 41 shows the structure of both molecules.

tryptophan   indole

Figure 41: Structures of the two newly discovered substrates, tryptophan and indole.

The absorption maximum of the red color caused by tryptophan oxidation was measured at 500 nm. Interestingly, the color in the indole oxidation experiments was dependent on the concentration of *tert*-butyl hydroperoxide. At an equimolar ratio of indole and *tert*-butyl hydroperoxide (5 mM / 5 mM), a red color with a maximum at 520 nm was observed, while at

54

an excess of the oxidizing agent (5 mM / 27 mM), a yellow color with an absorption maximum at about 400 nm was formed. This is most likely due to the formation of different reaction products.

To compare and quantify the enzyme variants in terms of indole and tryptophan oxidation, the color development was recorded in a plate reader. All variants with one amino acid exchange were tested for tryptophan and indole oxidation activity using this method. The increase in absorption over the first minute and the absorption after 10 minutes of reaction relative to the wild type enzyme are shown in Figure 42 for the tryptophan oxidation and in Figure 43 for indole oxidation with an excess of *tert*-butyl hydroperoxide.

Figure 42: Tryptophan oxidation activity of variants with one amino acid exchange. Shown is the increase in absorption at 500 nm over the first minute of reaction (dA/dt first minute, blue) as well as the absorption after 10 minutes (red). All values are given in percentage relative to the wild type enzyme. Samples were measured in triplicates. Reaction conditions: 5 mM tryptophan, 27 mM *tert*-butyl hydroperoxide. For further reaction conditions see section 2.11.

Especially Cupin 2-R39P, -C106Q and variants with an amino acid exchange at position F94 showed a significantly increased activity in this assay. It is also notable, that Cupin 2-C106L, -F94S and -W56D showed a greatly increased activity at the beginning of the reaction but did not reach a significantly higher level of absorption than the wild type after 10 minutes of reaction. Similar results were obtained with indole as substrate, where the same variants showed a significant increase in reaction speed.

Especially the high increase in reaction velocity experienced for some variants is remarkable in this experiment. In the tryptophan oxidation assay for example, the fastest variants showed a 10-fold increase within the first minute compared to the wild type. For the indole oxidation

with 27 mM *tert*-butyl hydroperoxide, the fastest variant (C106Q) showed an even higher reaction velocity, with a 21-fold increase compared to the wild type.
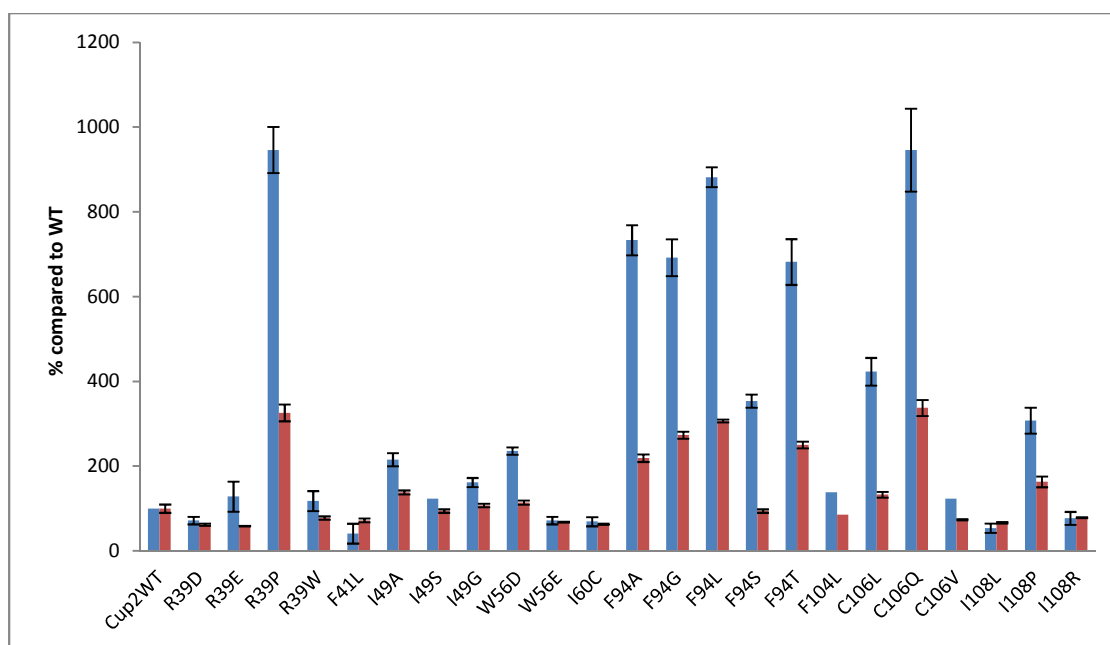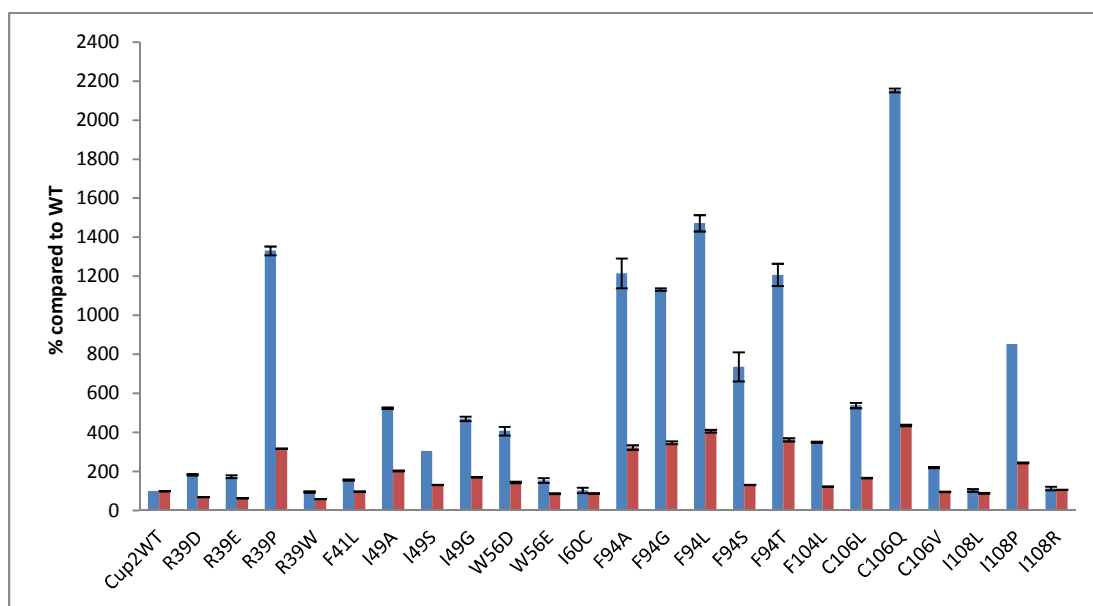

Figure 43: Indole oxidation activity of variants with one amino acid exchange. Shown is the increase in absorption at 400 nm over the first minute of reaction (dA/dt first minute, blue) as well as the absorption after 10 minutes (red). All values are given in percent relative to the wild type enzyme. Samples were measured in triplicates. Reaction conditions: 5 mM indole, 27 mM *tert*-butyl hydroperoxide. For further reaction conditions see section 2.11.

The increase of activity compared to the wild type was even higher at samples with an equimolar ratio of indole and *tert*-butyl hydroperoxide. The best variant (C106Q) showed a 230-fold increase in this experiment (see Figure 44).
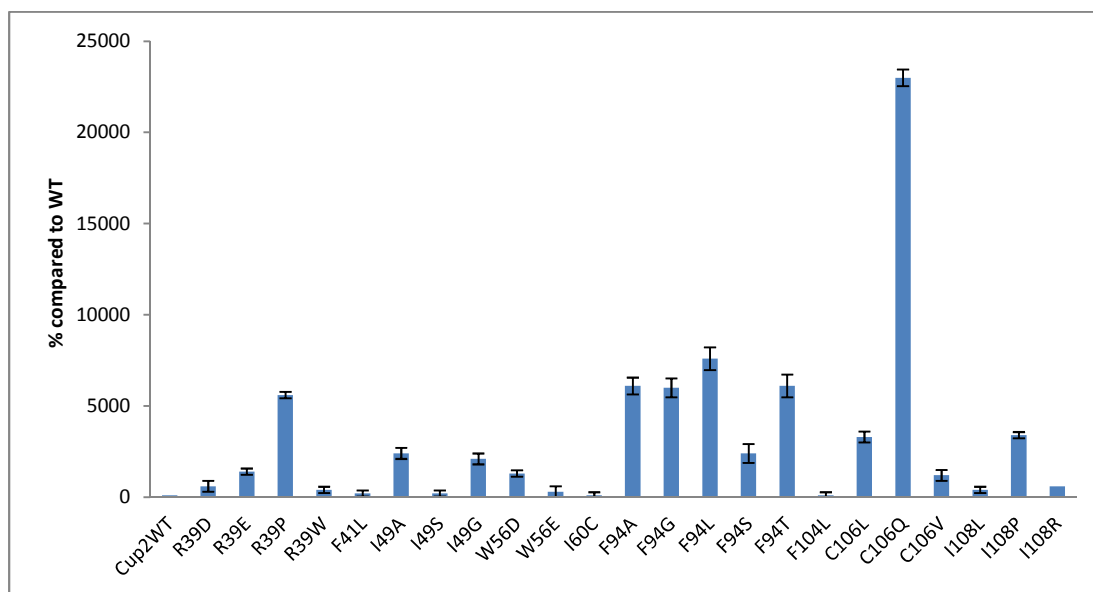
Figure 44: Indole oxidation activity of variants with one amino acid exchange compared to the wild type enzyme. Shown is the increase in absorption at 520 nm over the first minute of reaction (dA/dt first minute). All values are given in percent relative to the wild type enzyme. Samples were measured in triplicates. Reaction conditions: 5 mM indole, 5 mM *tert*-butyl hydroperoxide. For further reaction conditions see section 2.11.

From the progress curves of the indole reactions, it was visible that the reaction with the fastest variants reaches a plateau after about 10 minutes (

Figure 45).
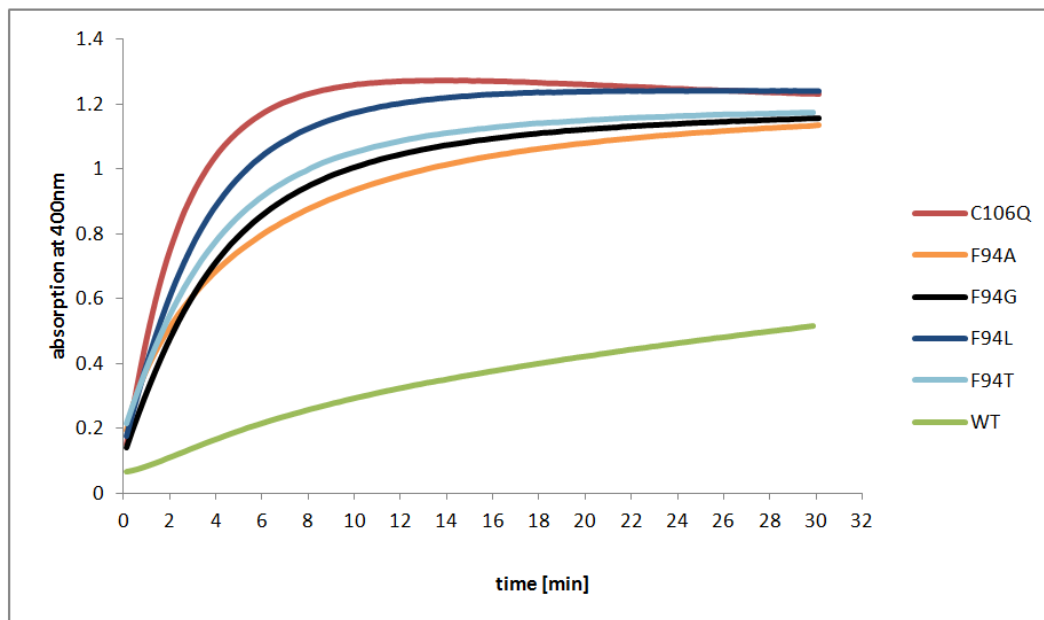
Figure 45: Increase of absorption at 400 nm measured for the indole oxidation reaction with 27 mM *tert*-butyl hydroperoxide. Shown are the fastest variants along with the wild type enzyme. Samples were measured in triplicates. For further reaction conditions see section 2.11.

In the reaction with tryptophan, however, no plateau is reached after 27 minutes (Figure 46).
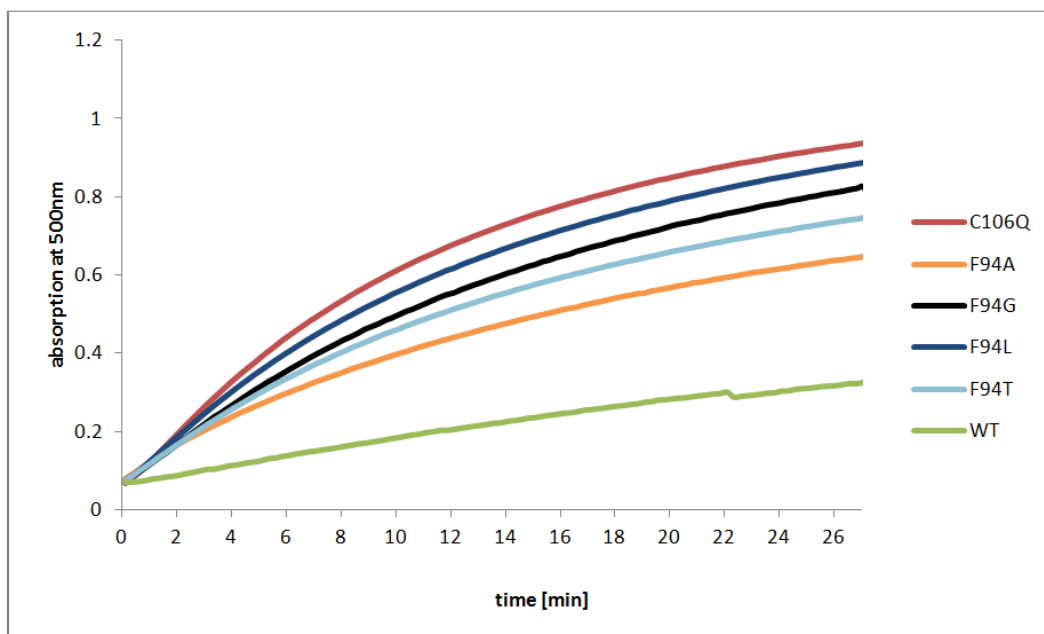
Figure 46: Increase of absorption at 500 nm measured for the tryptophan oxidation reaction. Shown are the fastest variants along with the wild type enzyme. Samples were measured in triplicates. For further reaction conditions see section 2.11.

To identify the oxidation products of tryptophan and indole, over-night bioconversion reactions in 1 mL scale were conducted. For indole, a standard reaction in a biphasic system with 50 mM substrate and 135 mM *tert*-butyl hydroperoxide was chosen. For the tryptophan, a single phase system with 10 mM tryptophan and 27 mM *tert*-butyl hydroperoxide was chosen because of the insolubility of tryptophan in organic solvents. In both cases, 1 mg of heat purified Cupin 2-C106Q was used as biocatalyst. The compounds of the reaction were extracted with ethyl acetate and measured by GC-MS. The red color of the indole reaction was almost completely located in the organic phase after the reaction, which made the extraction of this compound very effective. Unfortunately, the colored tryptophan oxidation products seemed to be close to insoluble in ethyl acetate and were therefore barely extractable. The two main reaction products of the indole oxidation could be identified as isatin and oxindole by a GC-MS analysis. For isatin, a absorption maximum at 418 nm was reported in literature[20]. This value would correlate with the peak at about 400 nm that was observed in the photometric determination of the indole oxidatioin. Small traces of n-formyl-2-aminobenzaldehyde that were also found in the reaction mix indicate a second, ring-cleaving background reaction. Structures of the identified products are show in Figure 47.

oxindole   **isatin**   **n-formyl-2-aminobenzaldehyde**
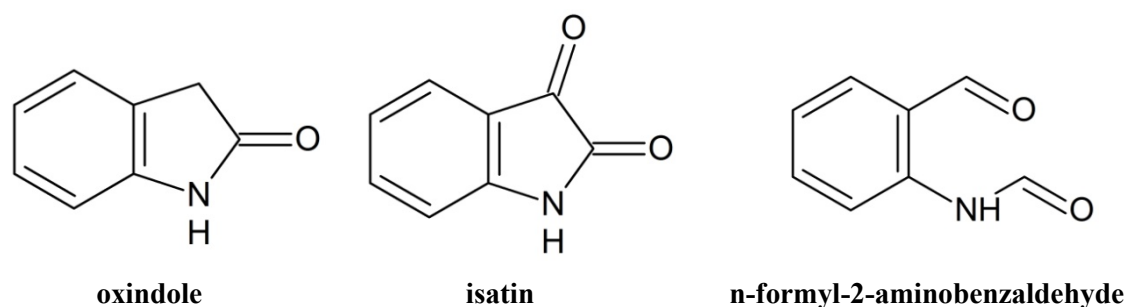
Figure 47: Structure of oxindole, isatin and n-formyl-2-aminobenzaldehyde, the indole oxidation products identified by GC-MS. For detailed GC-MS conditions see "Materials and Methods" section 2.10.3

The conversion of indole to isatin and oxindole was not quantified, but a comparison of the peak areas of substrate and products indicate a very low product concentration and therefore a marginal conversion (see Figure 48).
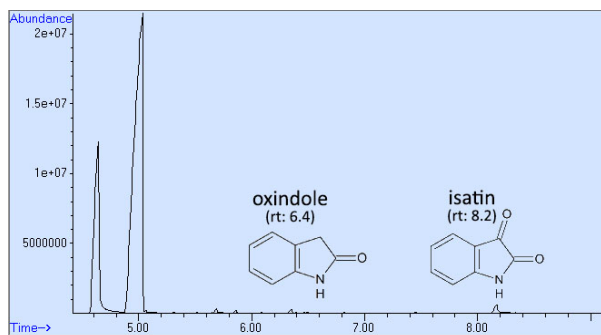
Figure 48: Peaks of the Cupin 2 catalyzed reaction producing isatin and oxindole from indole. Retention time 4.6 min: n-decanol (internal standard); Retention time 5 min: indole; Retention time 6.4 min: oxidole; Retention time 8.2 min: isatin. For detailed GC-MS conditions see "Materials and Methods" section 2.10.3

4 Discussion

4.1 Assay

The high throughput assay for acetophenone detection developed in the course of this thesis turned out to be a very reliable screening method. In the library screening, no false positive results could be discovered in the re-screening process of the library hits. The assay results are also well reproducible when the given periods for reaction and detection are exactly followed. As an alternative to the endpoint method, acetophenone concentrations can also be measured by recording the slope of the absorption increase at the beginning of the reaction. However, no significant increase in precision could be observed compared to the endpoint method. Despite the reproducibility of the assay, the re-screening experiments showed vastly differing absorption values between the biological septuplicates. These significant deviations only occurred when lysate was used without foregoing normalization to protein concentration, as it was the case when lysate from deepwell plate fermentations was directly used for screening. Using the heat purified lysate with normalized protein concentration showed significantly less deviations, indicating that the issue does not originate from the detection but rather from differences in cell growth or cell disruption efficiency between different wells of the deepwell plates. This means that the assay itself could potentially be used to quantify acetophenone concentrations and therefore enzyme activity of Cupin 2, but the enzyme concentration has to be normalized beforehand. Another reason why a normalization to enzyme concentration seems to be critical is that the protein expression experiments show differing overexpression levels in different variants. Because the activity screening by vanillin assay was done without prior normalization, the values for the activity increase acquired using this method should to be taken with a pinch of salt.

This detection method should theoretically be able to detect all ketones but was only tested on acetophenone, *p*-Cl-acetophenone (data not shown) and acetone[18].

4.2 Heat purification

Heat purification was proven to be a very effective method to cheaply and quickly purify large amounts of crude lysate samples. The small amounts of *E. coli* lysate protein that remained soluble after the heat treatment had no significant negative effects on the bioconversion reactions. The data from the SDS-PAGE gels after heat purification experiments indicate that the stability of the enzyme was not impaired by any of the variants with one or two amino acid exchanges. It has to be noted though, that because of the very high initial intensity of the bands, small discrepancies in band intensity may not be visible

from the produced gels. C-terminally truncated versions of the enzyme, however, appear to be more sensitive to the high temperatures, indicated by significantly weaker bands on heat purified samples. An activity assay after heat purification showed that the activity of some variants is decreased after heat purification. The reduction of the heat purification time from 10 to 5 minutes may be beneficial for such variants while still offering a decent purity.

4.3 Substrate spectrum

Beside the substrates with high similarity to the model substrate α-methylstyrene (i.e. styrene derivatives) the enzyme is also able to oxidize tryptophan and indole, yielding red colored compounds in both cases. The reaction product that forms the red color in the oxidation of indole was identified as isatin, while the corresponding compound of the tryptophan oxidation could not be identified. The presence of mainly isatin and oxindole and only traces of n-formyl-2-aminobenzaldehyde as oxidation products shows that the predominant reaction is the mere oxidation, rather than the oxidative cleavage of the indole ring system that is known from indole 2,3-dioxygenases[21]. Because of the structural similarities of indole and tryptophan, it may be assumed that the oxidation reaction of tryptophan yields the corresponding products dioxindolylalanine and oxindolylalanine (ring oxidation) as well as traces of N-formylkynurenine (ring cleavage). Because of the low solubility of tryptophan oxidation products in ethyl acetate, which is caused by the additional amino acid backbone, it is possible that these products were present in the reaction mixture, but were lost during extraction. Additionally, the GC setup was not ideal for such substances with comparatively low hydrophobicity. An experimental setup comprising a high-performance liquid chromatography (HPLC) and an optimized extraction protocol would most likely improve the chances of identifying these products. The colorimetric detection of indole and tryptophan oxidation activity revealed that the activity of most variants was significantly more increased with these substrates than with any styrene derivatives. This may be explained by the larger size of indole and tryptophan, making the enlargement of the active site that is presumably caused by most amino acid exchanges more prominent. Especially the 230-fold activity increase in the indole oxidation reaction, which occurred under certain reaction conditions, is notable and may be subject to further studies. The conversion of indole in the biphasic system was not quantified, but the comparison of the peak areas suggests a very low conversion.

Given the fact that the enzyme oxidizes free tryptophan, it could also occur that it is oxidizing its own tryptophan residue, which is located at the surface near the active site. Beside the enzymatic oxidation of tryptophan, the presence of an oxidizing agent in the reaction broth

may also promote non-enzymatic oxidation of histidines, methionines, cysteines and tryptophans. When samples were drawn from a mix of Cupin 2 incubated with *tert*-butyl hydroperoxide and analyzed via SDS-PAGE, the samples that were incubated for a longer period of time showed weaker bands. This was not only observed for the bands of the enzyme but also for the bands of the remaining *E. coli* lysate protein and indicates changes of the proteins by oxidation. The effects of the oxidation of these amino acids on the activity and stability of the enzyme are not known, but they could potentially render the enzyme inactive. This potential enzyme inactivation through oxidation processes may be the cause for the decreasing activity over longer periods of time that was shown in the time course experiment. One observation that promotes this theory is that the doubling of the enzyme concentration led to a significant increase in the conversion, which is assumed to be caused by a faster initial conversion speed and by the fact that active enzyme is present for a longer period of time.

Methods to prevent oxidative damage to proteins are for example the addition of β-mercaptoethanol, dithiothreitol or TCEP [Tris(2-carboxyethyl)phosphin] to the reaction mix, but the effects of these substances on the reaction are yet to be evaluated. Especially the effects of β-mercaptoethanol may be problematic for the reactions, because this reagent is known to react with aldehydes and ketones to give the corresponding oxathiolanes. Another solution may be the substitution of easily oxidized amino acids by site directed mutagenesis, but the effects on overall structural stability and enzyme activity are mostly unpredictable. Lowering the concentration of *tert*-butyl hydroperoxide would most likely also reduce the oxidative damage to the enzyme, but this would also have a detrimental effect on the conversion. The indole oxidation assay showed that the variant Cupin 2-C106Q showed a 230-fold increase in indole oxidizing activity compared to the wild type in conditions with just 5 mM *tert*-butyl hydroperoxide. If this can also be applied to other substrates, this mutant may hold the key for reaction conditions with drastically lowered oxidizing agent concentrations, which would most likely be beneficial for enzyme stability. However, this activity increase at lower concentrations of the oxidizing agent may be caused by the requirement of just one oxygen atom for the oxidation of indole to oxindole and therefore may not be applicable to alkene cleavage reactions, where the need for two oxygen atoms makes the presence of an excess of oxidizing agent favorable.[16]

The bioconversion reactions conducted with the alternative substrates showed that the activity towards *p*-Cl-α-methylstyrene and 2-methyl-1-phenyl-1-propene was increased in the tested variants with Cupin 2-C106Q being the most active variant in both cases. For the substrate

trans-anethole, the activity was not altered by the mutations. The activity towards isoeugenol was greatly increased in Cupin 2-C106V-L107*, but instead of the expected product vanillin, which could only be detected in traces, dehydrodiisoeugenol was produced as main product. This was particularly unexpected, because the underlying reaction was an oxidative carbon bond formation instead of the expected oxidative carbon bond cleavage. This type of reaction has never been observed for this enzyme before. The reaction mechanism behind the formation of dehydrodiisoeugenol from two molecules of isoeugenol is called oxidative coupling and the same reaction is known to be catalyzed enzymatically by peroxidase-H_2O_2 systems[22] or laccases[23], chemically by inorganic oxidants such as Ag_2O, hexacyanoferrates or $FeCl_3$[24] as well as photo-chemically by di-*tert*-butyl peroxide or chloranil along with UV radiation[25]. Given the various already established ways to synthesize dehydrodiisoeugenol and the possibility to extract the reagent from plants[26], this newly discovered activity is of minor importance for an industrial biocatalytic synthesis of dehydrodiisoeugenol.

4.4 Bioconversion reactions

It was not possible to get reproducible data in the bioconversion experiments. To get an idea of the reproducibility, a triplicate of the wild type enzyme was always measured along with each batch of reactions, which at least enabled a calculation of the conversion compared to the wild type. The GC results clearly show that the vast majority of variants show superior activity compared to the wild type enzyme measured in the same batch. The variants with the highest increase in activity were the variants with one amino acid exchange Cupin 2-C106L, -C106Q and -I108P, the variant with two amino acid exchanges Cupin 2-I49A-F94A as well as the C-terminally truncated variant Cupin 2-C106V-L107*. This was the case in every batch of reactions. However, the distinctive values for the conversion were varying too much between the different batches, which made a correct quantification of the conversion impossible. The problems in reproducibility may be caused by a suboptimal experimental setup. For every batch of reactions, a new aliquot of the crude lysate was thawed, heat purified and its protein concentration was quantified. Although the conditions for thawing, heat purification and Bradford assay were always the same, slight, disregarded differences in the conditions may have caused variations in enzyme activity and concentration. The standard curves for the GC were also not always measured in the same run as the samples, which may have also lead to inaccuracies. There were also major problems with the internal standard that was added to the reactions. It was discovered that the n-decanol got partly esterified with ethyl acetate in the presence of the quenching agent sodium bisulfite, forming decyl acetate. To be able to

quantify the reactions anyway, the peak of the decyl acetate was also quantified, the peak area multiplied by an experimentally determined proportionality factor of 0.9 and finally added to the peak area of the n-decanol, forming an adapted value for the internal standard. This, of course, added another source of inaccuracy to the GC measurements. All these issues with the bioconversion experiments lead to the conclusion that the experimental setup needs to be reworked in order to generate significant and reproducible results. There also seems to be room for improvement in the reaction conditions of the biphasic system. It was shown that the doubling of the *tert*-butyl hydroperoxide concentration led to an increase of the conversion from 22 % to 50% and also improved the recovery rate significantly, indicating the need for further optimizations of the reaction conditions. The cause for these imperfect reaction conditions is probably that they were initially optimized for a single-phase micro-emulsion system and were adopted for the biphasic system without any adaptation.

4.5 Mutations and enzyme structure

The results of the mutation experiments show that especially the C-terminus of the enzyme is important for its activity. Mutations of the cysteine at position 106 or the isoleucine at position 108 as well as the C-terminal truncation at position 107 or 109 are greatly improving enzyme activity. The C-terminus of Cupin 2 starts with a loop at position 109 that twists the following short α-helical structure towards the protein's centre, narrowing the cavity that leads to the active site (see Figure 49). It is assumed that mutations at positions 106 and 108 relocate the terminal α-helical structure and thereby widen the channel to the active site. This hypothesis is supported by the fact that the truncation of the C-terminus at positions 107 or 109 that completely removes the terminal helix induces an increase in activity which is comparable to the increase in activity caused by mutations at positions 106 and 108. The protein structure with the C-terminal area is shown in Figure 49.

The positive effect of the mutations in this area is not only limited to the model substrate α-methylstyrene, but also applies to other substrates, especially on indole and tryptophan. These molecules are larger that α-methylstyrene and therefore especially benefit from a wider channel to the active site.

An alternative explanation for the increased activity of mutants of the cysteine 106 residue is that the cysteine possibly gets oxidized and in some way inhibits the reaction in its oxidized form, which is avoided by the substitution with an amino acid that is not easily oxidized.

The amino acids I49 and F94 also turned out to be important. In both cases, the mutation to the smaller amino acids alanine and glycine improved the activity. The activity increase of

variants with amino acid exchanges at position F94 is particularly high in indole and tryptophan oxidation. This can again be explained by the fact that these larger molecules particularly profit from the slightly larger cavity generated by the exchange of the large amino acid phenylalanine with smaller ones. At position R39, a mutation from arginine to proline was able to greatly increase activity. This was rather surprising because R39 is located in a β-sheet, where a proline most likely compromises the structure. It is, however, supposable that this alteration in the β-sheet structure again widens the active site. This hypothesis conforms to the fact that the activity increase in this case is again higher for the larger substrates indole and tryptophan.
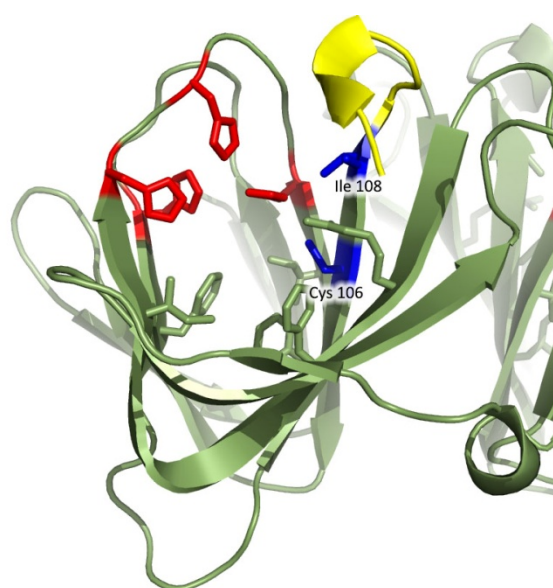
Figure 49: Structure of Cupin 2. Red: metal binding site. Blue: cysteine 106 and isoleucine 108 amino acid residues. Yellow: C-terminal loop and α-helical structure.

The exact origin of the double bands on SDS-PAGE gels that were observed for some variants could not be determined. However, it is assumed that the additional bands were caused by stability differences between different enzyme variants which lead to an incomplete degradation in the course of SDS-PAGE sample preparation in some cases.

4.6 Conclusion

The engineering of Cupin 2 was successful in enabling the production of vanillin from isoeugenol but the yield of this reaction is highly insufficient and the formation of another product in higher amounts is disadvantageous. However, the ability to produce dehydrodiisoeugenol from isoeugenol may shed some further light on the so far unknown catalytic mechanism of Cupin 2.

The overall goal of this thesis was to improve the enzyme activity with the model substrate and to enable the reaction with alternative substrates. Although it was not successfully quantified, there is sufficient evidence for increased activity in several variants to see the first goal as achieved. This activity increase was not only limited to the substrate used in the screening, but also applied for other similar substrates like p-Cl-α-methylstyrene as well as for different newly discovered substrates like indole. The second goal was achieved by the discovery of the newly generated activity with isoeugenol, although the main product of this reaction turned out to be a totally unexpected one. Despite of the improvements achieved by mutagenesis of the enzyme, its low stability under the reaction conditions, which is supposedly caused by oxidative damage, drastically limits its usefulness in biocatalysis. But if this issue can be overcome by optimized reaction conditions or further mutagenesis targeting the stability against oxidation, Cupin 2 may become a potent biocatalytic tool after all.

5 References

1   Rajagopalan, A., Lara, M. & Kroutil, W. Oxidative Alkene Cleavage by Chemical and Enzymatic Methods. *Advanced Synthesis & Catalysis* **355**, 3321-3335, doi:10.1002/adsc.201300882 (2013).

2   Kamlet, J. & Olin, M. Manufacture of vanillin and its homologues. U.S. patent 2640083 (1953).

3   Gallage, N. J. & Moeller, B. L. Vanillin − Bioconversion and Bioengineering of the most popular plant flavour and its de novo biosynthesis in the vanilla orchid. *Molecular Plant*, doi:10.1093/mp/ssu105 (2014).

4   Sabisch, M. & Smith, D. *The Complex Regulatory Landscape for Natural Flavor Ingredients*, <http://www.sigmaaldrich.com> (23. Jannuary 2015).

5   Hansen, E. H. *et al.* De Novo Biosynthesis of Vanillin in Fission Yeast (Schizosaccharomyces pombe) and Baker's Yeast (Saccharomyces cerevisiae). *Applied and Environmental Microbiology* **75**, 2765-2774, doi:10.1128/aem.02681-08 (2009).

6   Pollack, A. *What's That Smell? Exotic Scents Made From Re-engineered Yeast*, <http://www.nytimes.com/2013/10/21/business/whats-that-smell-exotic-scents-made-from-re-engineered-yeast.html> (11.02.2015).

7   Li, K. & Frost, J. W. Synthesis of Vanillin from Glucose. *Journal of the American Chemical Society* **120**, 10545-10546, doi:10.1021/ja9817747 (1998).

8   Gallage, N. J. *et al.* Vanillin formation from ferulic acid in Vanilla planifolia is catalysed by a single enzyme. *Nat Commun* **5**, doi:10.1038/ncomms5037 (2014).

9   Ryu, J.-Y. *et al.* Characterization of an Isoeugenol Monooxygenase (Iem) from Pseudomonas nitroreducens Jin1 That Transforms Isoeugenol to Vanillin. *Bioscience, Biotechnology, and Biochemistry* **77**, 289-294, doi:10.1271/bbb.120715 (2013).

10 Yamada, M., Okada, Y., Yoshida, T. & Nagasawa, T. Purification, characterization and gene cloning of isoeugenol-degrading enzyme from Pseudomonas putida IE27. *Arch Microbiol* **187**, 511-517, doi:10.1007/s00203-007-0218-9 (2007).

11 Mutti, F. G., Lara, M., Kroutil, M. & Kroutil, W. Ostensible Enzyme Promiscuity: Alkene Cleavage by Peroxidases. *Chemistry – A European Journal* **16**, 14142-14148, doi:10.1002/chem.201002265 (2010).

12 Tuynman, A., Spelberg, J. L., Kooter, I. M., Schoemaker, H. E. & Wever, R. Enantioselective Epoxidation and Carbon–Carbon Bond Cleavage Catalyzed by Coprinus cinereus Peroxidase and Myeloperoxidase. *Journal of Biological Chemistry* **275**, 3025-3030, doi:10.1074/jbc.275.5.3025 (2000).

13 ten Have, R., Rietjens, I. M. C. M., Hartmans, S., Swarts, H. J. & Field, J. A. Calculated ionisation potentials determine the oxidation of vanillin precursors by lignin peroxidase. *FEBS Letters* **430**, 390-392, doi:http://dx.doi.org/10.1016/S0014-5793(98)00682-6 (1998).

14 Renganathan, V., Miki, K. & Gold, M. H. Multiple molecular forms of diarylpropane oxygenase, an H2O2-requiring, lignin-degrading enzyme from Phanerochaete chrysosporium. *Archives of Biochemistry and Biophysics* **241**, 304-314, doi:http://dx.doi.org/10.1016/0003-9861(85)90387-X (1985).

15 Lara, M., Mutti, F. G., Glueck, S. M. & Kroutil, W. Biocatalytic Cleavage of Alkenes with O2 and Trametes hirsuta G FCC 047. *European Journal of Organic Chemistry* **2008**, 3668-3672, doi:10.1002/ejoc.200800261 (2008).

16 Hajnal, I., Faber, K., Schwab, H., Hall, M. & Steiner, K. Oxidative Alkene Cleavage Catalyzed by Mn-dependent Cupin TM1459 from *Thermotoga maritima*. (unpublished at the time of writing this thesis).

17   Reetz, M. T., Kahakeaw, D. & Lohmer, R. Addressing the Numbers Problem in Directed Evolution. *ChemBioChem* **9**, 1797-1804, doi:10.1002/cbic.200800298 (2008).

18   Amlathe, S. & Gupta, V. K. Spectrophotometric determination of acetone using vanillin. *Analyst* **115**, 1385-1387, doi:10.1039/an9901501385 (1990).

19   Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **72**, 248-254, doi:http://dx.doi.org/10.1016/0003-2697(76)90527-3 (1976).

20   Sousa, M. M. *et al.* A photochemical study on the blue dye indigo: from solution to ancient Andean textiles. *Photochemical & Photobiological Sciences* **7**, 1353-1359, doi:10.1039/b809578g (2008).

21   Nair, P. M. & Vaidyanathan, C. S. An indole oxidase isolated from the leaves of Tecoma stans. *Biochimica et Biophysica Acta (BBA) - Specialized Section on Enzymological Subjects* **81**, 496-506, doi:http://dx.doi.org/10.1016/0926-6569(64)90134-8 (1964).

22   Chioccara, F. *et al.* Regio- and Diastereo-selective Synthesis of Dimeric Lignans Using Oxidative Coupling. *Acta Chemica Scandinavica* **47**, 610-616, doi:10.3891/acta.chem.scand.47-0610 (1993).

23   Shiba, T., Xiao, L., Miyakoshi, T. & Chen, C. L. Oxidation of isoeugenol and coniferyl alcohol catalyzed by laccases isolated from Rhus vernicifera Stokes and Pycnoporus coccineus. *Journal of Molecular Catalysis B: Enzymatic* **10**, 605-615, doi:http://dx.doi.org/10.1016/S1381-1177(00)00184-3 (2000).

24   Setälä, H.   Page 21 (VTT Technical Research Centre of Finland, 2008).

25   DellaGreca, M. *et al.* Lignans by photo-oxidation of propenyl phenols. *Photochemical & Photobiological Sciences* **7**, 28-32, doi:10.1039/b707933h (2008).

26   Forrest, J. E., Heacock, R. A. & Forrest, T. P. The isolation of dehydrodiisoeugenol from the aril of *Myristica fragrans* Houtt. *Experientia* **29**, 139-139, doi:10.1007/bf01945432 (1973).

Appendix A - Primer Sequences

Table 10: List of all primer sequnces used in this thesis. Blue: Start/stop codon; Yellow: changed nucleotides.

Cupin 2 outer primer	
synCup2(pET26)_for	ATAATTTTGTTTAACTTTAAGAAGGAGATATACAT**ATG**ATTCTGAAACGTGCCTATGATG
synCup2(pET26)_rev	GGTGGTGGTGGTGGTGCTCGAGTGCGGCCGCAAGCTT**TTA**TTCGCCACCTTCTTTCG
synCupin2(Nde)_for	AATCACAT**ATG**ATTCTGAAACGTGCCTATGATGTTACAC
T7 Primer	
T7prom	TAATACGACTCACTATAGG
T7term	GCTAGTTATTGCTCAGCGG
Cupin 2 site directed mutagenesis primer:	
Cup2-H52A_for	GGTCTGATTGATCGT**GCG**AGCCATCCGTGGGAAC
Cup2-H52A_rev	GTTCCCACGGATGGCT**CGC**ACGATCAATCAGACC
Cup2-H54A_for	CTGATTGATCGTCATAGC**GCG**CCGTGGGAACATG
Cup2-H54A_rev	CATGTTCCCACGG**CGC**GCTATGACGATCAATCAG
Cup2-H58A_for	CATAGCCATCCGTGGGAA**GCG**GAAATTTTTGTGCTGAAAG
Cup2-H58A_rev	CTTTCAGCACAAAAATTTC**CGC**TTCCCACGGATGGCTATG
Cup2-H92A_for	GTGGAACCGAATGAAATT**GCG**GGCTTTCGTAATGATACC
Cup2-H92A_rev	GGTATCATTACGAAAGCC**CGC**AATTTCATTCGGTTCCAC
Cup2-R39H_for	CCGAATTTTGTGATGC**A**TCTGTTTACCGTTGAAC
Cup2-R39H_rev	GTTCAACGGTAAACAGA**T**GCATCACAAAATTCGG
Cup2-R39A_for	CACCGAATTTTGTGATG**GCG**CTGTTTACCGTTGAACC
Cup2-R39A_rev	GGTTCAACGGTAAACAG**CGC**CATCACAAAATTCGGTG
Cup2-F41A_for	GAATTTTGTGATGCGTCTG**GCG**ACCGTTGAACCGGGTGGTC
Cup2-F41A_rev	GACCACCCGGTTCAACGGT**CGC**CAGACGCATCACAAAATTC
Cup2-F41V_for	GAATTTTGTGATGCGTCTG**G**TTACCGTTGAACCGGGTG
Cup2-F41V_rev	CACCCGGTTCAACGGTAA**C**CAGACGCATCACAAAATTC
Cup2-I49A_for	GTTGAACCGGGTGGTCTG**GC**TGATCGTCATAGCCATC
Cup2-I49A_rev	GATGGCTATGACGATCA**GC**CAGACCACCCGGTTCAAC
Cup2-W56A_for	GATCGTCATAGCCATCCG**GC**GGAACATGAAATTTTTG
Cup2-W56A_rev	CAAAAATTTCATGTTCC**GC**CGGATGGCTATGACGATC
Cup2-I60A_for	CATCCGTGGGAACATGAA**GCG**TTTGTGCTGAAAGGTAAAC
Cup2-I60A_rev	GTTTACCTTTCAGCACAAA**CGC**TTCATGTTCCCACGGATG
Cup2-F94A_for	CCGAATGAAATTCATGGC**GCG**CGTAATGATACCGATAG
Cup2-F94A_rev	CTATCGGTATCATTACG**CGC**GCCATGAATTTCATTCGG
Cup2-F94V_for	CCGAATGAAATTCATGGC**G**TTCGTAATGATACCGATAG

Cup2-F94V_rev	CTATCGGTATCATTACGAA<mark>C</mark>GCCATGAATTTCATTCGG
Cup2-F104A_for	CCGATAGCGAAGTTGAA<mark>GCG</mark>CTGTGTCTGATTCCG
Cup2-F104A_rev	CGGAATCAGACACAG<mark>CGC</mark>TTCAACTTCGCTATCGG
Cup2-F104V_for	CGATAGCGAAGTTGAA<mark>G</mark>TTCTGTGTCTGATTC
Cup2-F104V_rev	GAATCAGACACAGAA<mark>C</mark>TTCAACTTCGCTATCG
Cup2-C106A_for	GCGAAGTTGAATTTCTG<mark>GCG</mark>CTGATTCCGAAAGAAGG
Cup2-C106A_rev	CCTTCTTTCGGAATCAG<mark>CGC</mark>CAGAAATTCAACTTCGC
Cup2-C106S_for	GATAGCGAAGTTGAATTTCTG<mark>AGC</mark>CTGATTCCGAAAGAAGGTGG
Cup2-C106S_rev	CCACCTTCTTTCGGAATCAG<mark>GCT</mark>CAGAAATTCAACTTCGCTATC
Cup2-I108A_for	GAATTTCTGTGTCTG<mark>GCG</mark>CCGAAAGAAGGTGGCG
Cup2-I108A_rev	CGCCACCTTCTTTCGG<mark>CGC</mark>CAGACACAGAAATTC
Site saturation NNK primer	
Cup2-R39X_for	CCGAATTTTGTGATG<mark>NNK</mark>CTGTTTACCGTTGAAC
Cup2-R39X_rev	GTTCAACGGTAAACAG<mark>MNN</mark>CATCACAAAATTCGG
Cup2-F41X_for	GAATTTTGTGATGCGTCTG<mark>NNK</mark>ACCGTTGAACCGGGTG
Cup2-F41X_rev	CACCCGGTTCAACGGT<mark>MNN</mark>CAGACGCATCACAAAATTC
Cup2-I49X_for	GTTGAACCGGGTGGTCTG<mark>NNK</mark>GATCGTCATAGCCATC
Cup2-I49X_rev	GATGGCTATGACGATC<mark>MNN</mark>CAGACCACCCGGTTCAAC
Cup2-W56X_for	GATCGTCATAGCCATCCG<mark>NNK</mark>GAACATGAAATTTTTG
Cup2-W56X_rev	CAAAAATTTCATGTTC<mark>MNN</mark>CGGATGGCTATGACGATC
Cup2-I60X_for	CATCCGTGGGAACATGAA<mark>NNK</mark>TTTGTGCTGAAAGGTAAAC
Cup2-I60X_rev	GTTTACCTTTCAGCACAAA<mark>MNN</mark>TTCATGTTCCCACGGATG
Cup2-F94X_for	CCGAATGAAATTCATGGC<mark>NNK</mark>CGTAATGATACCGATAG
Cup2-F94X_rev	CTATCGGTATCATTACG<mark>MNN</mark>GCCATGAATTTCATTCGG
Cup2-F104X_for	CGATAGCGAAGTTGAA<mark>NNK</mark>CTGTGTCTGATTC
Cup2-F104X_rev	GAATCAGACACAG<mark>MNN</mark>TTCAACTTCGCTATCG
Cup2-C106X_for	GCGAAGTTGAATTTCTG<mark>NNK</mark>CTGATTCCGAAAGAAGG
Cup2-C106X_rev	CCTTCTTTCGGAATCAG<mark>MNN</mark>CAGAAATTCAACTTCGC
Cup2-I108X_for	GAATTTCTGTGTCTG<mark>NNK</mark>CCGAAAGAAGGTGGCG
Cup2-I108X_rev	CGCCACCTTCTTTCGG<mark>MNN</mark>CAGACACAGAAATTC
Primer for the generation of c-terminally truncated muteins	
Cup2(pET26)_P109*_rev	GTGGTGGTGCTCGAGTGCGGCCGC<u>AAGCTT</u><mark style="background:cyan">TTA</mark>AATCAGACACAGAAATTCAACTTCGC
Cup2(pET26)_I108R_P109*_rev	GTGGTGGTGCTCGAGTGCGGCCGC<u>AAGCTT</u><mark style="background:cyan">TTA</mark><mark>ACG</mark>CAGACACAGAAATTCAACTTCGC
Cup2(pET26)C106V-I108R-P109*rev	GTGGTGGTGCTCGAGTGCGGCCGC<u>AAGCTT</u><mark style="background:cyan">TTA</mark><mark>ACG</mark>CAG<mark>AAC</mark>CAGAAATTCAACTTCGC
Cup2(pET26)_L107*_rev	GTGGTGGTGCTCGAGTGCGGCCGC<u>AAGCTT</u><mark style="background:cyan">TTA</mark>ACACAGAAATTCAACTTCGCTATCGG

Cup2(pET26)_C106V_ L107*_rev	GTGGTGGTGCTCGAGTGCGGCCGCAAGCTTTTAAACCAGAAATTCAACTT CGCTATCGG
Cup2(pET26)C-termE6_rev (C106Q_ L107*_rev)	GTGGTGGTGCTCGAGTGCGGCCGCAAGCTTTTACTGCAGAAATTCAACTT CGCTATCGG
Cup2(pET26)C-termE7_rev (L105M_C106Q_ L107*_rev)	GTGGTGGTGCTCGAGTGCGGCCGCAAGCTTTTACTGCATAAATTCAACTT CGCTATCGG
Cup2(pET26)C-termE8_rev (L105M_C106Q_ L107M_I108*_rev)	GTGGTGGTGCTCGAGTGCGGCCGCAAGCTTTTACATCTGCATAAATTCAA CTTCGCTATCGG
Cup2(pET26)C-termE9_rev (L105M_C106V_ L107*_rev)	GTGGTGGTGCTCGAGTGCGGCCGCAAGCTTTTAAACCATAAATTCAACTT CGCTATCGG
Primers for the last round of site directed mutagenesis	
Cup2-A33V_for	CTGATTGGTCTGAAAGATGTTCCGAATTTTGTGATGCGTC
Cup2-A33V_rev	GACGCATCACAAAATTCGGAACATCTTTCAGACCAATCAG
Cup2-L107M_for	CGAAGTTGAATTTCTGTGTATGATTCCGAAAGAAGGTGG
Cup2-L107M_rev	CCACCTTCTTTCGGAATCATACACAGAAATTCAACTTCG
Cup2-I108P_for	GAATTTCTGTGTCTGCCTCCGAAAGAAGGTGGCG
Cup2-I108P_rev	CGCCACCTTCTTTCGGAGGCAGACACAGAAATTC
Cup2-L105M_for	CCGATAGCGAAGTTGAATTTATGTGTCTGATTCCGAAAGAAG
Cup2-L105M_rev	CTTCTTTCGGAATCAGACACATAAATTCAACTTCGCTATCGG
Cup2-K24V_for	GTTCGTGGTGTTCGTGTACGTGTTCTGATTGGTC
Cup2-K24V_rev	GACCAATCAGAACACGTACACGAACACCACGAAC

71

Appendix B - gBlock Sequences

Changed codons are marked in yellow

Cup2-Multi-1 (R39P F41L I49A W56D F94A F104L C106V I108P)

GATATA<u>CAT**ATG**</u>ATTCTGAAACGTGCCTATGATGTTACACCGCAGAAAATTAGCACCGATAAAGTTCGTGGTGTTCGTAAACGTGTTCTGATTGGTCTGAAAGATGCACCGAATTTTGTGATG<mark>CCT</mark>CTG<mark>CTT</mark>ACCGTTGAACCGGGTGGTCTG<mark>GCG</mark>GATCGTCATAGCCATCCG<mark>GAT</mark>GAACATGAAATTTTTGTGCTGAAAGGTAAACTGACCGTGCTGAAAGAACAGGGTGAAGAAACCGTTGAAGAAGGCTTTTATATTTTTGTGGAACCGAATGAAATTCATGGC<mark>GCG</mark>CGTAATGATACCGATAGCGAAGTTGAA<mark>CTG</mark>CTG<mark>GTT</mark>CTG<mark>CCG</mark>CCGAAAGAAGGTGGCGAA**TAA**<u>AAGCTT</u>GCGATGACT

Cup2-Multi-2 (R39P I49A W56D F94A C106V I108R)

GATATA<u>CAT**ATG**</u>ATTCTGAAACGTGCCTATGATGTTACACCGCAGAAAATTAGCACCGATAAAGTTCGTGGTGTTCGTAAACGTGTTCTGATTGGTCTGAAAGATGCACCGAATTTTGTGATG<mark>CCT</mark>CTGTTTACCGTTGAACCGGGTGGTCTG<mark>GCG</mark>GATCGTCATAGCCATCCG<mark>GAT</mark>GAACATGAAATTTTTGTGCTGAAAGGTAAACTGACCGTGCTGAAAGAACAGGGTGAAGAAACCGTTGAAGAAGGCTTTTATATTTTTGTGGAACCGAATGAAATTCATGGC<mark>GCG</mark>CGTAATGATACCGATAGCGAAGTTGAATTTCTG<mark>GTT</mark>CTG<mark>CGT</mark>CCGAAAGAAGGTGGCGAA**TAA**<u>AAGCTT</u>GCGATGACT

Cup2-Multi-3 (R39P I49A F94A C106V)

GATATA<u>CAT**ATG**</u>ATTCTGAAACGTGCCTATGATGTTACACCGCAGAAAATTAGCACCGATAAAGTTCGTGGTGTTCGTAAACGTGTTCTGATTGGTCTGAAAGATGCACCGAATTTTGTGATG<mark>CCT</mark>CTGTTTACCGTTGAACCGGGTGGTCTG<mark>GCG</mark>GATCGTCATAGCCATCCGTGGGAACATGAAATTTTTGTGCTGAAAGGTAAACTGACCGTGCTGAAAGAACAGGGTGAAGAAACCGTTGAAGAAGGCTTTTATATTTTTGTGGAACCGAATGAAATTCATGGC<mark>GCG</mark>CGTAATGATACCGATAGCGAAGTTGAATTTCTG<mark>GTT</mark>CTGATTCCGAAAGAAGGTGGCGAA**TAA**<u>AAGCTT</u>GCGGATGACT

Appendix C - Cupin 2

Amino Acid Sequence:

MILKRAYDVTPQKISTDKVRGVRKRVLIGLKDAPNFVMRLFTVEPGGLIDRHSHPWE
HEIFVLKGKLTVLKEQGEETVEEGFYIFVEPNEIHGFRNDTDSEVEFLCLIPKEGGE

Nucleotide Sequence (Codon Optimized for E. coli)

ATGATTCTGAAACGTGCCTATGATGTTACACCGCAGAAAATTAGCACCGATAAAG

TTCGTGGTGTTCGTAAACGTGTTCTGATTGGTCTGAAAGATGCACCGAATTTTGTG

ATGCGTCTGTTTACCGTTGAACCGGGTGGTCTGATTGATCGTCATAGCCATCCGTG

GGAACATGAAATTTTTGTGCTGAAAGGTAAACTGACCGTGCTGAAAGAACAGGG

TGAAGAACCGTTGAAGAAGGCTTTTATATTTTTGTGGAACCGAATGAAATTCAT

GGCTTTCGTAATGATACCGATAGCGAAGTTGAATTTCTGTGTCTGATTCCGAAAG

AAGGTGGCGAATAA

Additional Information

Number of amino acids:	114
Number of Nucleotides:	345
Molecular weight:	13109.0
Definition:	hypothetical protein TM1459 [Thermotoga maritima MSB8].
Accession:	NP_229258
PDB-ID:	1VJ2