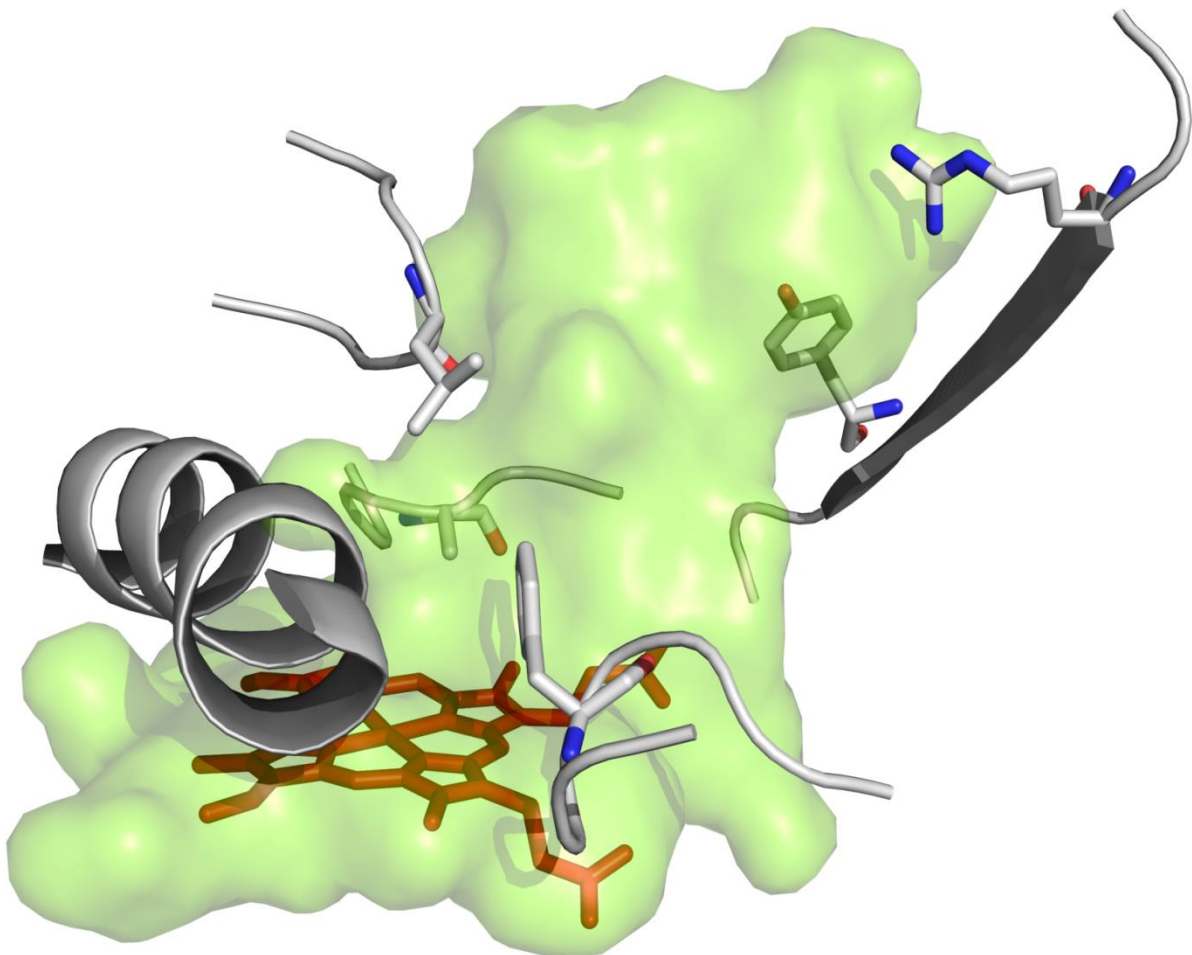


**Novel route to vanillin –
an enzyme-catalyzed multi-step cascade
synthesis**

Neuer Weg zu Vanillin –
eine enzymkatalysierte mehrstufige Kaskadensynthese



Tobias Klaus

Institut für Technische Biochemie

Universität Stuttgart

**Novel route to vanillin –
an enzyme-catalyzed multi-step cascade synthesis**

**Neuer Weg zu Vanillin –
eine enzymkatalysierte mehrstufige Kaskadensynthese**

Von der Fakultät Energie-, Verfahrens- und Biotechnik der Universität
Stuttgart zur Erlangung der Würde eines Doktors der
Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

Vorgelegt von

Tobias Klaus

aus

Hechingen, Deutschland

Hauptberichter: Prof. Dr. Bernhard Hauer
Mitberichter: Prof. Dr. Karl-Heinrich Engesser
Tag der mündlichen Prüfung: 18.10.2016

Institut für Technische Biochemie der Universität Stuttgart

2016

Cover:

Substrate channel and heme access region of CYP102A1 (modified from PDB ID: 1BU7, chain A), which participates in the enzyme-catalyzed multi-step cascade synthesis of vanillin described in this thesis. The image was generated by PyMOL.

Erklärung über die Eigenständigkeit der Dissertation

Hiermit versichere ich, Tobias Klaus, dass ich die vorliegende Arbeit mit dem Titel „Neuer Weg zu Vanillin – eine enzymkatalysierte mehrstufige Kaskadensynthese“ selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Aus fremden Quellen entnommene Passagen und Gedanken sind als solche kenntlich gemacht. Des Weiteren bestätige ich ausdrücklich, dass die hier vorliegende Dissertation nicht in gleicher oder ähnlicher Form bei einer anderen Institution zur Erlangung eines akademischen Grades eingereicht wurde.

Declaration of Authorship

I, Tobias Klaus, hereby declare that the present thesis entitled „Novel route to vanillin – an enzyme-catalyzed multi-step cascade synthesis” is the result of my own work, that all sources used or quoted have been indicated, and that I have not used any illegitimate means. I further declare that I have not submitted this thesis for a degree in some form or another.

Name: Tobias Klaus

Ort und Datum / Place and date: Rottweil, 13.06.2016

Für meine Familie und in Angedenken an meinen Bruder Martin

Acknowledgments

Viele Personen haben dazu beigetragen, dass diese Dissertation entstehen konnte. An dieser Stelle möchte ich mich bei allen, die maßgeblich daran beteiligt waren, ganz herzlich bedanken.

Mein ganz besonderer Dank gilt meinem Doktorvater Prof. Dr. Bernhard Hauer. Er hat mir die Möglichkeit gegeben am Institut für Technische Biochemie (ITB) der Universität Stuttgart in einer hervorragenden Forschungsumgebung dieses spannende und herausfordernde Projekt bearbeiten zu können. Auch für die hervorragende Betreuung sowie viele interessante und hilfreiche Diskussionen und Ratschläge möchte ich mich bedanken.

Herrn Prof. Dr. Karl-Heinrich Engesser und Herrn Prof. Dr. Ralf Takors danke ich für die freundliche Übernahme der Funktionen als Mitberichter und Prüfungsvorsitzender.

Des Weiteren gilt mein Dank insbesondere Dr. Bettina Nestl für ihre ausgezeichnete Betreuung, die vielen wertvollen Ratschläge und die zahlreichen Korrekturen vieler, im Laufe meiner Dissertation entstandener, Texte, Poster, Präsentationen und nicht zuletzt auch meiner Dissertation. Ich danke ihr für all ihre Antworten auf meine Fragen in zahlreichen Gesprächen und auch für die stets sehr angenehme und freundschaftliche Atmosphäre während unserer Zusammenarbeit am ITB.

Meinen Kollegen aus der Bioinformatik-Gruppe des ITB, Dr. Alexander Seifert und Dr. Łukasz Gricman danke ich für ihre Unterstützung bei bioinformatischen Fragestellungen, wie der Durchführung von MD-Simulationen und der Erstellung eines Homologiemodells. Auch möchte ich Dr. Alexander Seifert für die freundliche Zurverfügungstellung zahlreicher Varianten von CYP102A1 danken.

Tim Häbe, Lea Kolb und Sebastian Wagner danke ich für ihre Mitarbeit an Teilprojekten meiner Dissertation im Rahmen ihrer Diplomarbeit bzw. Forschungspraktika am ITB.

Acknowledgments

Nicht nur fachlich, sondern auch persönlich hat mich meine Zeit am ITB sehr bereichert. Dazu haben neben bereits erwähnten Personen noch viele mehr beigetragen. Hiervon möchte ich Dr. Bernd Nebel für seine Unterstützung in vielen Fragestellungen der instrumentellen Analytik, sowie für viele auflockernde Momente besonders danken. Dr. Sumire Honda Malca, Dr. Daniel Scheps und Dr. Martin Weissenborn danke ich für die tolle Zusammenarbeit im P450-Team. Vielen Dank allen lieben Kolleginnen und Kollegen der Biokatalyse und des gesamten ITB für diese prägende und schöne Zeit am Institut und abseits davon.

Dr. Bettina Nestl, Dr. Martin Weissenborn und Monika Digeser danke ich für das Korrekturlesen meiner Dissertation oder Teilen davon.

Für die finanzielle Förderung meines Dissertationsprojektes möchte ich sowohl dem BMBF als auch dem EU-Forschungsprojekt BIONEXGEN Dankeschön sagen.

Mein letzter und zugleich größter Dank gilt meiner Familie, insbesondere meinem Vater und meiner Freundin Moni, für ihre stete Unterstützung während all der Jahre meines Studiums und meiner Dissertation. Auch in schwierigen Zeiten haben sie mir immer zur Seite gestanden, mir den Rücken frei gehalten, mich aufgemuntert und mir letztlich die Kraft gegeben diesen Weg erfolgreich zu gehen. Danke, dass es euch gibt !

Table of contents

Acknowledgments	7
Table of contents	9
Abbreviations	12
Abstract	16
Zusammenfassung	18
1 Introduction	20
1.1 Biocatalysis.....	20
1.2 Aromatic hydroxylation.....	21
1.3 Flavoproteins	24
1.3.1 Flavoprotein monooxygenases.....	25
1.3.2 Flavoprotein oxidases.....	28
1.3.3 <i>m</i> -Hydroxybenzoate hydroxylase MobA from <i>Comamonas testosteroni</i>	29
1.3.4 Vanillyl-alcohol oxidase (VAO) from <i>Penicillium simplicissimum</i>	31
1.4 Cytochrome P450 monooxygenases.....	33
1.4.1 Nomenclature and functions	33
1.4.2 Catalytic reaction mechanism.....	34
1.4.3 Classification	36
1.4.4 CYP102A1 from <i>Bacillus megaterium</i>	38
1.4.5 CYP116B3 from <i>Rhodococcus ruber</i>	39
1.5 Vanillin – an aromatic compound of interest.....	41
1.6 Aim of the project.....	42
2 Materials and methods	44
2.1 Materials	44
2.1.1 Genes, proteins, vectors, primers and strains.....	44
2.1.2 Media, buffers and solutions.....	44
2.1.3 Chemicals and cloning enzymes	44
2.2 Methods.....	45
2.2.1 General methods.....	45
2.2.2 Preparation of competent <i>E. coli</i> cells.....	45
2.2.3 Transformation of competent <i>E. coli</i> cells	46
2.2.4 Agarose gel electrophoresis	47
2.2.5 Site-directed mutagenesis	47
2.2.6 Generation of a focused mutant library of CYP116B3.....	48

Table of contents

2.2.6.1	Homology modelling of CYP116B3	48
2.2.6.2	Identification of mutation sites	48
2.2.7	Protein expression and purification	48
2.2.7.1	General expression and purification protocol	48
2.2.7.2	Expression and purification of MobA	50
2.2.7.3	Expression and purification of CYP116B3	50
2.2.7.4	Expression of CYP102A1	52
2.2.7.5	Expression of VAO	52
2.2.8	SDS-PAGE analysis	52
2.2.9	Quantification of proteins	53
2.2.9.1	Total protein concentration	53
2.2.9.2	Cytochrome P450 monooxygenase concentration	53
2.2.10	Investigation of enzyme activity	54
2.2.10.1	Investigation of MobA activity	54
2.2.10.1.1	Spectro-photometrical analysis	54
2.2.10.1.2	Determination of enzyme activity via NADPH depletion	54
2.2.10.2	Investigation of CYP116B3 activity (7-ethoxycoumarin assay)	54
2.2.10.3	Investigation of VAO activity	55
2.2.11	<i>In vitro</i> biotransformations	55
2.2.11.1	Single-enzyme catalyzed reactions	55
2.2.11.2	Bi-enzymatic one-pot cascade reactions	56
2.2.11.2.1	Combination of two CYP102A1 variants	56
2.2.11.2.2	Combination of CYP102A1 and VAO	57
2.2.12	<i>In vivo</i> biotransformations	57
2.2.13	Product analysis	58
2.2.13.1	Sample treatment	58
2.2.13.2	HPLC-analysis	58
2.2.13.3	GC-FID-analysis	59
2.2.13.4	GC-MS-analysis	60
2.2.14	Molecular dynamics simulations	60
3	Results	61
3.1	Enzyme selection, preparation and application in initial biotransformations	61
3.1.1	Preparation of the <i>m</i> -hydroxybenzoate hydroxylase MobA	61
3.1.1.1	Site-directed mutagenesis, expression and purification of MobA	61
3.1.1.2	Investigation of MobA activity	63
3.1.1.2.1	Spectro-photometrical analysis	63
3.1.1.2.2	Determination of enzyme activity via NADPH depletion	64
3.1.2	Preparation of the cytochrome P450 monooxygenase CYP116B3	65
3.1.2.1	Expression and purification of CYP116B3	65
3.1.2.2	Determination of activity of CYP116B3	69
3.1.3	Preparation of the cytochrome P450 monooxygenase CYP102A1	70

Table of contents

3.1.4	Initial biotransformations with selected enzymes and enzyme variants.....	70
3.2	Expansion of the spectrum of enzymes and variants for further biotransformations ..	74
3.2.1	Generation of a focused mutant library of CYP116B3.....	74
3.2.1.1	Homology modelling of CYP116B3	74
3.2.1.2	Selection of mutation sites and generation of CYP116B3 variants.....	77
3.2.1.3	CYP116B3 variants expression and determination of enzyme activity.....	78
3.2.1.4	<i>In vitro</i> biotransformations with the focused CYP116B3 mutant library	82
3.2.2	Generation and application of further CYP102A1 variants	85
3.2.2.1	Screening of a set of triple mutant variants of CYP102A1	85
3.2.2.2	Molecular dynamics simulations of CYP102A1 with 4-methylguaiacol.....	87
3.2.2.3	Biotransformations with R47L/Y51F-variants of CYP102A1	89
3.2.3	Investigation of the vanillyl alcohol oxidase VAO from <i>P. simplicissimum</i>	92
3.2.3.1	Site-directed mutagenesis, expression and investigation of activity of VAO	93
3.3	<i>In vitro</i> one-pot cascade synthesis of vanillin	95
3.4	<i>In vivo</i> one-pot cascade synthesis of vanillin	98
4	Discussion	102
4.1	Enzyme selection, preparation and application in initial biotransformations.....	103
4.2	Expansion of the spectrum of enzymes and variants for further biotransformation .	107
4.3	<i>In vitro</i> one-pot cascade synthesis of vanillin	112
4.4	<i>In vivo</i> one-pot cascade synthesis of vanillin	113
5	Conclusion and outlook.....	115
6	References.....	118
7	Supplementary material.....	134
7.1	Genes, proteins, vectors, primers and strains	134
7.1.1	Genes.....	134
7.1.2	Proteins	138
7.1.3	Vectors	140
7.1.4	Primers.....	146
7.1.5	Strains	149
7.2	Supplementary tables	150
7.2.1	Media, buffers and solutions.....	150
7.2.2	<i>In vitro</i> biotransformations	154
7.3	Supplementary figures	155
7.3.1	Protein expression.....	155
7.3.2	Molecular dynamics simulations	157
7.3.3	GC-MS analysis.....	159

Abbreviations

x g	gravitational acceleration
°C	degree Celsius
µg	microgram
µl	microlitre
µm	micrometre
µM	micromolar
µmol	micromol
A	absorption
aa	amino acid
ADH	alcohol dehydrogenase
al	after induction
5-ALA	5-aminolevulinic acid
Amp	ampicillin
APS	ammonium persulfate
<i>B. megaterium</i>	<i>Bacillus megaterium</i>
BLAST	basic local alignment search tool
bp	base pair
BVMO	Baeyer-Villiger monooxygenase
ca.	circa
<i>C. testosteroni</i>	<i>Comamonas testosteroni</i>
CO	carbon monoxide
CPR	cytochrome P450 reductase
cww	cell wet weight
CYP	cytochrome P450 monooxygenase
CYP102A1	cytochrome P450 monooxygenase from <i>Bacillus megaterium</i> ATCC 14581, also referred to as P450BM3
CYP116B3	cytochrome P450 monooxygenase from <i>Rhodococcus ruber</i> DSM 44319
CYPED	Cytochrome P450 Engineering Database
Da	dalton
DAD	diode array detector
dH ₂ O	distilled water

Abbreviations

DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
7-EC	7-ethoxycoumarin
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
<i>e.g.</i>	<i>exempli gratia</i>
<i>et al.</i>	<i>et alii</i>
FAD	flavin adenine dinucleotide
FID	flame ionization detector
FMN	flavin mononucleotide
FPMO	flavoprotein monooxygenase
FPO	flavoprotein oxidase
g	gram
G6P	glucose-6-phosphate
G6PDH	glucose-6-phosphate dehydrogenase
GC	gas chromatography
GC-FID	gas chromatography coupled to flame ionization detector
GC-MS	gas chromatography coupled to mass spectrometer
h	hour
7-HC	7-hydroxycoumarin
HPLC	high-performance liquid chromatography
IF	insoluble fraction (after cell lysis)
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl- β -D-thiogalactopyranoside
ITB	Institute of Technical Biochemistry
Kan	kanamycin
kb	kilobase
k_{cat}	turnover number
kDa	kilodalton
kg	kilogram
K_M	Michaelis constant
l	litre
LB	lysogeny (Luria-Bertani) broth
M	molar

Abbreviations

mA	milliampere
MD	molecular dynamics
mg	milligram
MHBH	<i>m</i> -hydroxybenzoate hydroxylase
min	minute
ml	millilitre
mM	millimolar
MobA	<i>m</i> -hydroxybenzoate hydroxylase from <i>Comamonas testosteroni</i> GZ39
MOPS	3-(<i>N</i> -Morpholino)propanesulfonic acid
MS	mass spectrometer
MWCO	molecular weight cutoff
NADH	β -nicotinamide adenine dinucleotide (reduced)
NADP ⁺	β -nicotinamide adenine dinucleotide phosphate (oxidized)
NADPH	β -nicotinamide adenine dinucleotide phosphate (reduced)
NAD(P) ⁺	β -nicotinamide adenine dinucleotide (phosphate) (oxidized)
NAD(P)H	β -nicotinamide adenine dinucleotide (phosphate) (reduced)
NCBI	National Center for Biotechnology Information
ng	nanogram
nm	nanometre
nmol	nanomol
OD ₆₀₀	optical density measured at 600 nm
O/N	overnight
<i>P. simplicissimum</i>	<i>Penicillium simplicissimum</i>
P450	cytochrome P450 monooxygenase
PCR	polymerase chain reaction
PFOR	phthalate-family oxygenase reductase
pH	<i>pondus Hydrogenii</i>
PHBH	<i>p</i> -hydroxybenzoate 3-hydroxylase
pl	prior to induction
PMSF	phenylmethanesulfonyl fluoride
<i>R. ruber</i>	<i>Rhodococcus ruber</i>
rpm	revolutions per minute
RT	room temperature

Abbreviations

s	second
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SF	soluble fraction (after cell lysis)
S/N	supernatant
SOC	super optimal catabolite repression medium
TAE	buffer solution containing Tris, acetic acid and EDTA
TEMED	<i>N,N,N',N'</i> -tetramethylethylenediamine
Tfb	transformation buffer
TMO	toluene monooxygenase
Tris	tris-(hydroxymethyl)-aminomethane
U	unit
UV	ultraviolet
V	volt
VAO	vanillyl alcohol oxidase from <i>Penicillium simplicissimum</i> CBS 170.90
wt	wild type
XMO	xylene monooxygenase

Abstract

The selective hydroxylation of aromatic compounds is one of the most challenging chemical reactions. As an alternative to traditional chemical catalysis, biocatalysis emerged during the past decades. Hence, in the present work, a number of biocatalysts was investigated with regard to the realization of a novel synthesis route to the valuable aromatic compound vanillin, starting from the simple low-cost aromatic substrate 3-methylanisole via the intermediate products 3-methoxybenzyl alcohol or 4-methylguaiacol and via vanillyl alcohol, as an example of consecutive enzyme-catalyzed oxidation reactions accomplished in a multi-enzymatic three-step cascade reaction.

For this reason a preselected set of enzymes, namely the *m*-hydroxybenzoate hydroxylase MobA from *Comamonas testosteroni* GZ39 and the cytochrome P450 monooxygenases CYP116B3 from *Rhodococcus ruber* DSM 44319 and CYP102A1 from *Bacillus megaterium* ATCC 14581, was investigated towards the selective hydroxylation of the substrate 3-methylanisole. Beside the wild type enzymes, a variant of MobA, which was created by rational protein design, and an existing focused minimal mutant library of CYP102A1 were applied in initial biotransformation reactions, combined with an efficient cofactor recycling system. Though the wild type enzymes of CYP116B3 and CYP102A1 displayed only a basic level of activity towards 3-methylanisole, highly increased activity was detected for many of the CYP102A1 variants with a maximum of 59% total conversion for the double mutant F87V/A328L. With 3-methoxybenzyl alcohol and 4-methylguaiacol both intermediate compounds of the intended cascade synthesis were generated, though 4-methoxy-2-methylphenol was the main product in most of the reactions. However, none of the so far investigated variants accepted any of the intermediate compounds as substrate.

As CYP116B3 was a good candidate for further protein engineering approaches, as a basic level of activity towards the substrate of interest was already present in the wild type enzyme, a focused mutant library of 20 single mutant variants of CYP116B3 was created based on literature, sequence and structure information in order to improve the enzymes activity and selectivity towards conversion of the substrate 3-methylanisole and in order to find variants for the conversion of 3-methoxybenzyl alcohol and/or 4-methylguaiacol. Therefore a homology model of the

Abstract

monooxygenase domain of CYP116B3 was generated. Though, compared to the wild type, variants with up to almost six time increased activity towards the model substrate 7-ethoxycoumarin were found, total activity towards 3-methylanisole was still much lower compared to the best CYP102A1 variants. In addition, none of the variants displayed appropriate conversion of the intermediate compounds 3-methoxybenzyl alcohol and 4-methylguaiacol. Moreover, additional mutation in the literature known amino acid position 437 of CYP102A1 variant F87V/A328L revealed no benefit towards conversion of any of the substrates, too.

Molecular dynamics simulations of a CYP102A1 variant with 4-methylguaiacol as substrate revealed the bottleneck in the conversion of this compound. 4-Methylguaiacol was shown to be stabilized at the entrance of the substrate access channel by the polar amino acid residues R47 and Y51. Replacement of these residues by the hydrophobic residues leucine and phenylalanine, respectively, resulted in successful conversion of 4-methylguaiacol to vanillyl alcohol, the precursor of vanillin in the intended cascade synthesis. Though, as the yield of vanillyl alcohol synthesized from 4-methylguaiacol with CYP102A1 variants was rather low, a vanillyl alcohol oxidase from *Penicillium simplicissimum* and rationally designed variants thereof, described in literature, were investigated. As a result, not only vanillyl alcohol but also 4-methylguaiacol was converted in high yield to vanillin. Finally, a combination of the best 4-methylguaiacol producing variant, CYP102A1 variant A328L, with the best 4-methylguaiacol converting variant, VAO variant F454Y, in one reaction system both *in vitro* and *in vivo* yielded vanillin from 3-methylanisole with a maximal product formation of 2.0% and 1.1% vanillin, respectively.

We demonstrated as a proof-of-principle the establishment of the proposed multi-enzymatic three-step cascade reaction pathway. Though further optimizations concerning increase of enzyme activity and improvement of enzyme selectivity are required, the above mentioned exemplary synthesis of vanillin illustrates the capability of biocatalysis.

Zusammenfassung

Die selektive Hydroxylierung aromatischer Verbindungen ist eine der herausforderndsten chemischen Reaktionen. Als eine Alternative zur traditionellen chemischen Katalyse hat sich die Biokatalyse während der vergangenen Jahrzehnte entwickelt. In der hier vorliegenden Arbeit wurden verschiedene Biokatalysatoren hinsichtlich der Verwirklichung einer neuen Syntheseroute zu der wertvollen aromatischen Verbindung Vanillin untersucht. Dies sollte, ausgehend von dem einfachen und preiswerten Substrat 3-Methylanisol, über die Zwischenprodukte 3-Methoxybenzylalkohol oder 4-Methylguaiacol und über Vanillylalkohol, als Beispiel für aufeinanderfolgende enzymkatalysierte Oxidationsreaktionen in einer multienzymatischen dreistufigen Kaskadenreaktion bewerkstelligt werden.

Aus diesem Grund wurde eine Auswahl an Enzymen, und zwar die *m*-Hydroxybenzoathydroxylase MobA aus *Comamonas testosteroni* GZ39 und die Cytochrom P450 Monooxygenasen CYP116B3 aus *Rhodococcus ruber* DSM 44319 und CYP102A1 aus *Bacillus megaterium* ATCC 14581, hinsichtlich der selektiven Hydroxylierung des Substrates 3-Methylanisol untersucht. Neben den Wildtyp-Enzymen wurde eine Variante von MobA, die durch rationales Proteindesign erstellt wurde, und eine bestehende fokussierte Minimalmutantenbibliothek von CYP102A1 kombiniert mit einem effizienten Cofaktor-Regenerierungssystem in anfänglichen Biotransformationsreaktionen angewendet. Obwohl die Wildtyp-Enzyme von CYP116B3 und CYP102A1 nur eine geringe Grundaktivität gegenüber 3-Methylanisol zeigten, wurde für viele der CYP102A1-Varianten eine sehr viel höhere Aktivität gemessen, mit einem Maximum von 59% Gesamtumsetzung für die Doppelmutante F87V/A328L. Mit 3-Methoxybenzylalkohol und 4-Methylguaiacol wurden beide Zwischenprodukte der angestrebten Kaskadensynthese hergestellt. Allerdings war 4-Methoxy-2-methylphenol das Hauptprodukt in den meisten der Reaktionen. Indes akzeptierte keine der bis hierhin untersuchten Varianten irgendeines der Zwischenprodukte als Substrat.

Da CYP116B3, aufgrund seiner bereits im Wildtyp-Enzym vorhandenen grundsätzlichen Aktivität gegenüber dem Substrat, ein guter Kandidat für weitere Protein Engineering Ansätze war, wurde basierend auf Literatur-, Sequenz- und Strukturinformationen eine fokussierte Mutantenbibliothek von CYP116B3 erstellt. Dies geschah mit der Absicht die Enzymaktivität und -selektivität hinsichtlich der

Zusammenfassung

Umsetzung des Substrates 3-Methylanisol zu verbessern, und mit dem Ziel Varianten für die Umsetzung von 3-Methoxybenzylalkohol und/oder 4-Methylguaiacol zu finden. Deshalb wurde ein Homologiemodell der Monooxygenase-Domäne von CYP116B3 erstellt. Obwohl Varianten gefunden wurden, die im Vergleich zum Wildtyp eine bis nahezu sechsfach höhere Aktivität gegenüber dem Modells substrat 7-Ethoxycumarin gezeigt haben, war die Gesamtaktivität gegenüber 3-Methylanisol immer noch viel niedriger verglichen mit der besten CYP102A1-Variante. Außerdem zeigte keine der Varianten eine angemessene Umsetzung der Zwischenprodukte 3-Methoxybenzylalkohol und 4-Methylguaiacol. Auch zusätzliche Mutationen an der literaturbekannten Aminosäureposition 437 der CYP102A1-Variante F87V/A328L zeigten keinen Nutzen hinsichtlich der Umsetzung eines der Substrate.

Molekulardynamiksimulationen einer CYP102A1-Variante mit 4-Methylguaiacol als Substrat enthüllten den Engpass bei der Umsetzung dieser Verbindung. Es wurde gezeigt, dass 4-Methylguaiacol am Eingang des Substratzugangskanals durch die polaren Aminosäurereste R47 und Y51 stabilisiert wurde. Ein entsprechender Austausch dieser Aminosäuren durch die hydrophoben Aminosäuren Leucin und Phenylalanin hatte die erfolgreiche Umsetzung von 4-Methylguaiacol zu Vanillylalkohol zur Folge, welches die Vorstufe von Vanillin in der geplanten Kaskadensynthese war. Da die Ausbeute an Vanillylalkohol, welcher durch CYP102A1-Varianten aus 4-Methylguaiacol erhalten wurde, eher gering war, wurde eine Vanillylalkoholoxidase aus *Penicillium simplicissimum* untersucht, ebenso wie Varianten davon, die durch rationales Proteindesign hergestellt wurden und in der Literatur beschrieben sind. Als Ergebnis wurde nicht nur Vanillylalkohol, sondern auch 4-Methylguaiacol in hohem Maße zu Vanillin umgesetzt.

Schließlich hat eine Kombination der besten 4-Methylguaiacol produzierenden Variante, CYP102A1-Variante A328L, zusammen mit der besten 4-Methylguaiacol umsetzenden Variante, VAO-Variante F454Y, in einem Reaktionssystem sowohl *in vitro* als auch *in vivo* Vanillin aus 3-Methylanisol erzeugt, mit einer entsprechenden maximalen Produktbildung von 2,0% beziehungsweise 1,1% Vanillin.

Somit konnte die angestrebte multienzymatische dreistufige Kaskadenreaktion als grundsätzliches Konzept erfolgreich etabliert werden. Trotz dessen, dass weitere Optimierungen hinsichtlich der Erhöhung der Enzymaktivität und der Verbesserung der Enzymselektivität erforderlich sind, verdeutlicht die oben erwähnte exemplarische Synthese von Vanillin das Leistungsvermögen der Biokatalyse.

1 Introduction

1.1 Biocatalysis

While for many years enzymes as “green” biocatalysts were not of interest for industrial applications, identification of the first three-dimensional enzyme structure by X-ray analysis, giving important insight into how biocatalysts work, was a breakthrough for biocatalysis in the late 1950s. Emerging genetic engineering technologies in the 1980s were a further key step towards application of biocatalysts in industrial biotechnology. In 2012 Bornscheuer and co-workers described three waves of biocatalysis. The first wave started more than one century ago when scientists recognized that components of living cells could be applied to useful chemical transformations. The limited stability of the biocatalyst was the main challenge to be overcome by enzyme immobilization. The second wave took place in the 1980s and 1990s, when initial protein engineering technologies could be employed for extension of the substrate range of enzymes allowing the synthesis of unusual synthetic intermediates. Hence, biocatalysis could be applied in the synthesis of pharmaceutical compounds and fine chemicals. However, in addition to the stability, the optimization of the biocatalyst for non-natural substrates was another challenge. The third wave of biocatalysis, which takes place at the present, started in the mid and late 1990s by the development of directed evolution methods. Today, a better knowledge of proteins and an increased number of available directed evolution methods allows making major changes in enzyme properties by enzyme engineering, supported by advances in bioinformatics.¹ Hence, within the last couple of years, both in laboratory and in industrial scale, biocatalysis could be established as a practical and environmentally friendly alternative to traditional metalcatalysis and organocatalysis in chemical synthesis due to scientific and technological advances.^{1,2} Currently, there are many examples for biocatalysis applied in industrial production processes, for example the production of the antimalarial drug precursor artemisinic acid in engineered yeast³ or the production of the valuable aromatic flavor compound vanillin in a microbiological process.⁴

Major advantages of biocatalysis compared to chemocatalysis are the high selectivity, activity and environmental friendliness of enzymatic biocatalysts.⁵ Their

high selectivities simplify reaction work-ups and provide product in higher yields. Moreover, biocatalytic processes are safe as they typically run at ambient temperature, atmospheric pressure and neutral pH.¹ Due to the high regio- and enantioselectivity, biocatalysts are applied in the synthesis of fine chemicals, including pharmaceuticals, agrochemicals, and their intermediates.⁵ With biocatalysts the number of production steps, waste generation and thus fabrication costs could be reduced in some industrial processes, for example the synthesis of cortisone with the whole-cell biocatalyst *Rhizopus arrhizus*.^{6,7} The enzymatic production of the bulk chemical acrylamide is another example for biocatalysis applied in industry, taking advantage of the low *E* factor ($\text{kg}_{\text{waste}}/\text{kg}_{\text{product}}$) and ambient reaction conditions.⁸

1.2 Aromatic hydroxylation

Some of the major aims in chemistry are the improvement of atoms economies, of chemical steps and of process operations in chemical syntheses.⁹ Thereby, one of the most sought-after reactions both in laboratory and on an industrial scale is the catalytic aerobic C-H oxidation.¹⁰ Here, the activation of an inert C-H bond, the functionalization of the activated organic fragment in a way that allows catalyst turnover, and the realization of these steps with controlled and high chemo-, regio-, and stereoselectivities under environmentally friendly and safe conditions, are some of the main challenges.¹¹⁻¹³ Especially the activation of the thermodynamically strong and kinetically inert C-H bonds in hydrocarbons is difficult to achieve.^{14, 15} This chemical inertness originates from the fact that the constituent atoms of alkanes are held together by strong and localized C–C and C–H bonds. Therefore the molecules have neither empty orbitals of low energy nor filled orbitals of high energy that could participate in a chemical reaction.¹⁶ Though with the usage of transition metal complexes and the development of various methods for catalytic functionalization the activation of C-H bonds has been shown with many substrates, particularly arenes,¹⁷ the achievement of suitable selectivities, especially with functionally complex compounds, still remains challenging.⁹

Concerning the intended oxidation of C-H bonds in aromatic compounds, there are basically three ways described: i) heterogeneous catalysis with e.g. Pd, Re or Cu catalysts, ii) homogeneous catalysis with catalysts like a catecholate-Fe complex, and iii) biocatalysis employing enzymes as catalysts.¹⁰

Introduction

In this context, a variety of enzymatic mechanisms exist in nature that can accomplish C-H bond functionalization in aqueous solution at ambient temperature and pressure.¹⁸ Basically, all of these enzymatic mechanisms involve the action of cofactors (in many cases metals) in the active site of the enzyme. The positioning of these cofactors plays an important role in controlling the substrate scope of the enzyme and the selectivity of the reaction. Two large classes of reactivities can be distinguished by involvement of functionalization of either aromatic or aliphatic C-H bonds. A unique feature of aromatic substrates is that the activation of their C-H bonds can happen via reaction of their π -electrons.¹⁹

Enzymes can catalyze a variety of C-X bond formations (X = C, O, N, S, halogen, etc.). Thereby, the reactivities of the catalytic enzymes depend on the type of the employed cofactor. According to Lewis and co-workers,⁹ enzymes can be grouped depending on the employed cofactors into flavoproteins (cofactor: flavin mononucleotide (FMN), flavin adenine dinucleotide (FAD)), Ado[•] enzymes (cofactor: 5'-deoxyadenosyl radical (Ado[•])), mononuclear non-heme metalloenzymes, metalloenzymes with carboxylate-bridged diiron centers, binuclear copper monooxygenases, and heme enzymes (cofactor: heme). Concerning the challenging hydroxylation of aromatic substrates, the metalloenzymes with carboxylate-bridged diiron centers, the binuclear copper monooxygenases, the flavoproteins and the heme enzymes are of special interest.

Belonging to the group of metalloenzymes with carboxylate-bridged diiron centers, bacterial multicomponent monooxygenases (BMMs) use residue-ligated iron centers for hydroxylation of unactivated C-H bonds.²⁰ Amongst others, phenol hydroxylases and alkene/arene monooxygenases were identified to belong to the BMMs.²¹ Toluene monooxygenases (TMO) from *Pseudomonas mendocina* and *Ralstonia pickettii*, belonging to the class of alkene/arene monooxygenases, are well known for their ability to hydroxylate toluene to *p*-cresol with high regioselectivity and thus a preference for aromatic ring hydroxylation,²² whereas xylene monooxygenases (XMO) preferentially perform benzylic hydroxylations (Figure 1.1).²³

Introduction

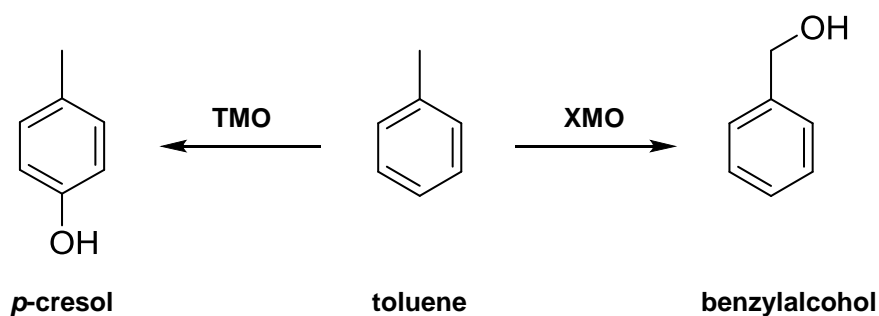


Figure 1.1: Oxidation of toluene catalyzed by toluene monooxygenase (TMO) (aromatic hydroxylation) and xylene monooxygenase (XMO) (benzylic hydroxylation). Graphic adapted from Lewis *et al.* 2011.⁹

Flavoproteins are interesting enzymes, as the flavin cofactors are known for their ability to hydroxylate the C-H bonds of electron-rich (hetero)arenes.^{24, 25} A key step in the hydroxylation of C-H bonds of a variety of compounds is that one of the oxygen atoms of the dioxygen is inserted into the C-H bond of the substrate (monooxygenation), while the second oxygen atom is reduced to water.²⁶ In addition to C-H bond hydroxylations, flavoprotein monooxygenases (FPMOs) catalyze many other oxidation reactions like epoxidations, Baeyer-Villiger oxidations, and amine oxidations.²⁴ Two subclasses of FPMOs, defined by differences in sequence and structure (e.g. number of domains, type of employed cofactors), have aromatic hydroxylase activity.²⁷ As an example, the *p*-hydroxybenzoate 3-hydroxylase is one of the most investigated enzymes of the FPMO family.²⁸ Other examples of FPMOs are known to catalyze the hydroxylation of phenol (e.g. the phenol 2-monooxygenase) or halogenated phenol derivatives.²⁹ However, due to their electrophilic aromatic substitution-based mechanism, FPMOs have a rather small substrate scope, limited to electron-rich aromatic substrates.⁹

The group of the heme enzymes might be the most prominent group of enzymes known to catalyze the functionalization of C-H bonds.³⁰ This functionalization can proceed via complex formation of the heme cofactor either with peroxides or with dioxygen, depending on the respective proximal ligand and active site residues of the enzyme.³¹ The reactive intermediate, known as compound I,³² reacts with the substrates in different ways, characteristic for the different heme enzyme subfamilies (peroxidases, haloperoxidases, and cytochrome P450 monooxygenases (CYPs, P450s)). For example, P450-catalyzed hydroxylation of sp^2 and sp^3 C-H bonds can

take place by direct oxo transfer, electrophilic aromatic substitution, or H atom abstraction.³³

For each of the different enzyme groups described above, there are examples for applications of these enzymes in various syntheses, some of them are already applied in industrial processes,⁹ for example in the production of fine chemicals.³⁴⁻³⁶ However, though these enzymes are often active on a range of substrates with high chemo-, regio- and stereoselectivity, there are also some challenges like insufficient activity, recombinant enzyme expression and purification, cofactor supply, organic solvent and oxygen tolerance, stability and substrate scope that have to be overcome and improved towards a broader practical application of the biocatalysts on an industrial scale.³⁷⁻⁴¹ Another advantage of biocatalysts compared to chemocatalysts is that, in many cases, enzymes can be evolved in the lab (improved activity, altered regioselectivity, ...) by means of protein engineering (directed evolution, rational protein design) and metabolic engineering of production hosts allows the combination of enzymes from different organisms.^{9, 42-45}

1.3 Flavoproteins

Flavoproteins or flavoenzymes are oxidoreductases that catalyze a variety of chemical reactions, having respectively diverse biological functions.⁴⁶ According to their name, flavoproteins contain flavin nucleotides, either flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD), as a cofactor. The FMN is derived from riboflavin via phosphorylation and the FAD is derived from FMN by adenylation.²⁷ Only a few flavoproteins contain a covalently bound FMN or FAD, whereas the majority of flavoproteins contain a non-covalently bound flavin as a prosthetic group.⁴⁷ Because of their ability to catalyze not only one-electron but also two-electron transfer reactions, flavoproteins play a central role in aerobic metabolism.⁴⁸ Moreover, due to their broad reaction spectrum, their high regio- and stereoselectivity, and their ability to catalyze oxidation reactions under environmentally friendly and mild conditions, flavoproteins are important enzymes for the mineralization of phenolic compounds. Hence, they are interesting enzymes for the biocatalytic production of valuable compounds in the pharmaceutical, fine chemical and food industries.^{46, 49}

Flavoproteins comprise two different groups of enzymes, the flavoprotein monooxygenases (FPMOs) and the flavoprotein oxidases (FPOs). FPMOs can be distinguished from FPOs by their mode of reaction. While FPMOs incorporate one oxygen atom into a substrate by previous activation and cleavage of molecular oxygen using electrons from NAD(P)H, flavoprotein oxidases react with their substrates using only molecular oxygen without employment of external cofactors.⁴⁹

1.3.1 Flavoprotein monooxygenases

Flavoprotein monooxygenases (FPMOs) are a very diverse enzyme family, known for their ability not only to catalyze regioselective C-H bond hydroxylations but also a variety of other oxygenation reactions like epoxidations, Baeyer-Villiger oxidations, and enantioselective sulfoxidations (Figure 1.2).²⁴ Hence, they play a role in biological processes like drug detoxification, biodegradation of aromatic compounds, and biosynthesis of antibiotics and siderophores, to mention only a few.⁵⁰ As their name already says, FPMOs catalyze monooxygenation reactions by introduction of a single oxygen atom of molecular oxygen into the substrate, while the second oxygen atom is concomitantly reduced to water.⁴⁶ Most FPMOs are external FPMOs (EC 1.14.13), as they consume NAD(P)H as reduced coenzyme during reduction of their flavin cofactor.²⁴ Internal FPMOs (EC. 1.13.12), where the flavin is reduced by the substrate itself, are very rare and will therefore not be further mentioned, subsequently.

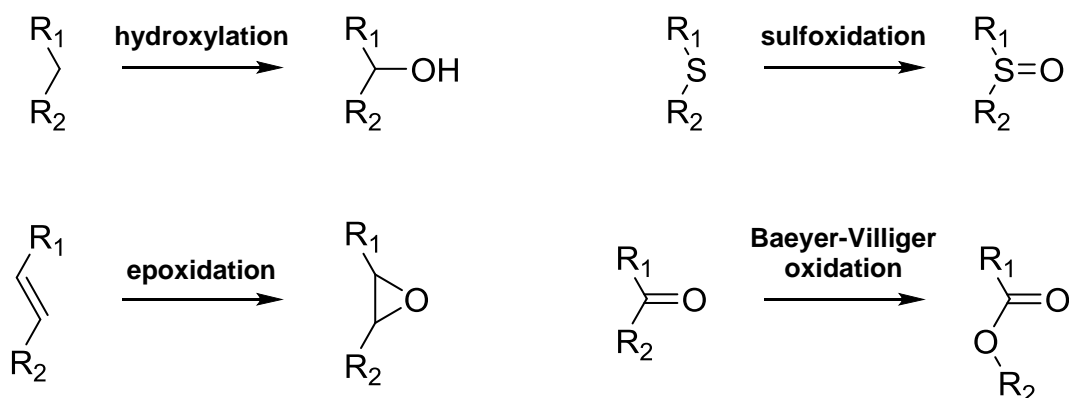


Figure 1.2: A selection of reactions catalyzed by flavoprotein monooxygenases. Graphic adapted from van Berkel *et al.* 2006.²⁴

Introduction

Based on available sequential and structural information, external FPMOs can be classified into six subclasses (A-F).^{24, 46}

Subclass A: These FPMOs are encoded by a single gene and contain a tightly bound FAD cofactor. They are dependent on NAD(P)H as coenzyme, and typically react with aromatic compounds containing an activating hydroxyl or amino group as substrates, whereat NAD(P)⁺ is released immediately after flavin reduction. Structurally they are composed of one dinucleotide binding domain (Rossmann fold) which binds the cofactor FAD. Subclass A FPMO members usually are very regioselective and display a narrow substrate specificity with a subtle coenzyme, oxygen and substrate recognition mechanism.^{24, 46} They are reported to be involved in the microbial degradation of aromatic compounds by hydroxylation of the aromatic ring in *o*- or *p*-position.⁴⁹ The electrophilic attack on the aromatic ring of the substrate takes place by the C_(4a)-hydroperoxyflavin as oxygenating flavin species. A specialty of this subclass of FPMOs is that the isoalloxazine ring of the FAD prosthetic group can move during catalysis between different positions. Hence, there is an *open*, a *closed (in)* and an *out* conformation possible. Further details concerning these conformational changes are given by Ballou and co-workers⁵⁰ and will not be explained in detail here. One of the most prominent members of subclass A is the *p*-hydroxybenzoate 3-hydroxylase (also called PHBH, 4-hydroxybenzoate 3-monooxygenase, or 4-hydroxybenzoate 3-hydroxylase).²⁸ Another interesting member, whose crystal structure was dissolved in 2006,⁵¹ is the 3-hydroxybenzoate 4-hydroxylase from *Comamonas testosteroni*. This enzyme was reported to have a large tunnel that connects the substrate-binding pocket to the protein surface, thus enabling substrate and oxygen transport. This special feature is not present in homologous enzymes like the 4-hydroxybenzoate 3-hydroxylase, though the protein fold of the catalytic domains as well as the active-site architecture are similar.⁵¹

Subclass B: Members of this subclass of FPMOs are, as the subclass A enzymes, encoded by a single gene and contain a tightly bound FAD cofactor. As a difference, they are specific for NADPH as coenzyme and the NADPH/NADP⁺ is kept bound during catalysis. Moreover, these enzymes are composed of two dinucleotide binding domains (Rossmann fold) enabling the binding of cofactor FAD and coenzyme NADPH.^{24, 46} Subclass B can be further differentiated into three subfamilies: flavin-

Introduction

containing monooxygenases, microbial *N*-hydroxylating monooxygenases, and Baeyer-Villiger monooxygenases (Type I BVMOs).⁵²

In contrast to the subclasses A and B that are single-component enzymes as they consist of a single polypeptide chain encoded by a single gene, the subclasses C-F comprise FPMOs that are multi-component enzymes, thus generally consisting of two different polypeptide chains with different functions: a reductase component for the flavin reduction and an oxygenase component for oxidation of the substrate with molecular oxygen. They employ flavin as a coenzyme rather than a cofactor.²⁴ With respect to the coenzyme use and specificity as well as concerning sequence homology, there is a large variety between the subclasses C-F.^{24, 53} Further information about the subclasses C-F is given by van Berkel and co-workers.²⁴

Concerning the catalytic reaction mechanism (Figure 1.3), in general FPMO-catalyzed reactions can be described by three steps: First, the flavin is reduced by NAD(P)H; Second, the reduced flavin reacts with molecular oxygen to provide the reactive intermediate, the C_(4a)-(hydro)peroxyflavin (peroxyflavin for electrophilic substrates and hydroperoxyflavin for nucleophilic substrates); Third, the substrate is bound, oriented and activated for its oxygenation by the C_(4a)-(hydro)peroxyflavin.⁵⁰ However, in detail, the differentiation between single-component and two-component systems goes along with a different kind of reaction mechanism during substrate oxygenation. For the single-component FPMOs, the isoalloxazine ring of the flavin is “flexible” (movement of several angstroms), allowing protein rearrangements as mentioned before (see Subclass A), thus yielding in a way multiple active sites.⁵⁰ For the two-component enzymes, this is not possible. Here, first a transfer of the reducing equivalents from NAD(P)H to a reductase-bound flavin takes place, followed by a transfer of the reduced flavin to the oxygenase component. Then the peroxygenated flavin is formed by reaction with molecular oxygen, finally initiating the monooxygenation reaction of the substrate.²⁴ More detailed information on the catalytic reaction mechanism of single-component and two-component systems is given by Ballou *et al.*⁵⁰

Introduction

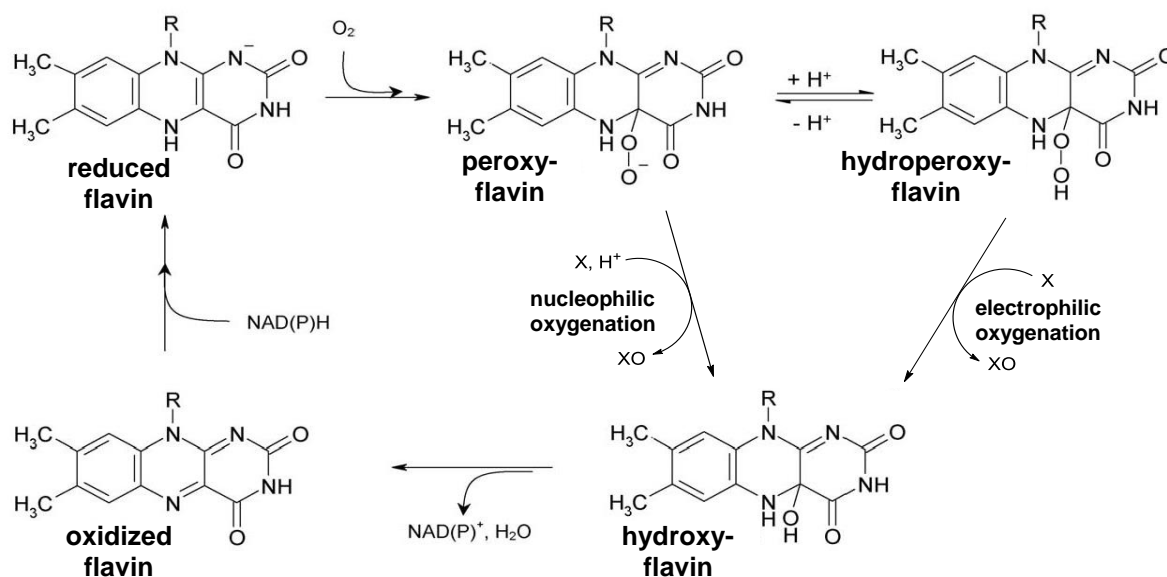


Figure 1.3: General catalytic mechanism of oxygenation reactions catalyzed by external flavoprotein monooxygenases. Graphic adapted from van Berkel *et al.* 2006.²⁴

1.3.2 Flavoprotein oxidases

Flavoprotein oxidases (FPOs) are interesting enzymes for biocatalytical applications, as they are not dependent on expensive external cofactors by simply using molecular oxygen as a cheap and clean oxidant.⁴⁹ They can introduce side-chain modifications in *p*-substituted phenols and many of them display activity with amines or (unnatural) amino acids. Hence, due to this substrate spectrum and due to their high regio- and enantioselectivity they are interesting enzymes for the production of flavors, fragrances and pure enantiomers in chiral organic syntheses.^{49, 54}

FPO-catalyzed substrate conversion consists of two half-reactions. These are the reduction of the flavin cofactor by the substrate and the following re-oxidation of the reduced flavin by molecular oxygen.⁵⁵ The molecular oxygen is concomitantly reduced to hydrogen peroxide.⁴⁶ Hence, the flavin cofactor is necessary for substrate oxidation, though it is not directly involved in the monooxygenation reaction. The oxygen atom used for substrate oxidation originates from water.²⁴

One of the most investigated and best characterized flavoprotein oxidases is the vanillyl-alcohol oxidase (VAO, EC 1.1.3.38).⁵⁶

1.3.3 *m*-Hydroxybenzoate hydroxylase MobA from *Comamonas testosteroni*

The *m*-hydroxybenzoate hydroxylase (MHBH) (3-hydroxybenzoate 4-monooxygenase (EC 1.14.13.23))^{24, 51} from *Comamonas testosteroni* is a single-component FPMO encoded by a single gene,⁵¹ classified as a member of subclass A of the flavoprotein monooxygenases.²⁴ It was reported to be an FAD-dependent monooxygenase, catalyzing the conversion of 3-hydroxybenzoate (*m*-hydroxybenzoate) to 3,4-dihydroxybenzoate (protocatechuate), requiring equimolar amounts of NADPH and molecular oxygen (Figure 1.4 A). The optimal pH in Tris-HCl buffer is 7.3 and the relative molecular weight, determined by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), is 71.0 kDa. Generally, the enzyme displays a strict substrate specificity under experimental conditions. Nevertheless, a few other substrates, like 2,5-dihydroxybenzoate (gentisate) and 2,3-dihydroxybenzoate, were converted with relative activities of 17.0% and 31.7%, respectively, compared to the activity with 3-hydroxybenzoate.⁵⁷ In 1998 MHBH was crystallized and further characterized concerning its absorption profiles and its reaction with several inhibitors.⁵⁸ Hiromoto and co-workers did investigations on the primary structure of MHBH, determined its crystal structure in the Michaelis complex form with its substrate and analyzed the Xe-derivative structure.⁵¹ They determined the nucleotide sequence of the MHBH-gene, named *mobA*, comprising 1917 nucleotides which encode a protein consisting of 639 amino acids with a calculated molecular mass of 70.0 kDa. Highest similarity of the deduced amino acid sequence was found to the sequence of a phenol hydroxylase from *Trichosporon cutaneum* with 35% identity. From the determination of the crystal structure the enzyme, subsequently called MobA, was shown to form a physiologically active homodimer with crystallographic 2-fold symmetry. Each subunit of the homodimer consists of three α/β domains. The catalytic domains I and II comprise a substrate binding pocket (active site) whereas the C-terminal domain III is involved in oligomerization. Structural similarities between MobA, PHBH, and the phenol hydroxylase, concerning the protein fold of the catalytic domains and the active-site architecture, including the FAD and substrate-binding site, strengthen the assumption, that members of the class of flavoprotein aromatic hydroxylases share a similar mechanism of catalytic substrate hydroxylation.⁵⁹ In MobA the substrate-binding pocket is located at the bottom of a large tunnel in the protein interior that

Introduction

connects this pocket to the surface. The two entrances of the tunnel are open to the solvent region and separated on the protein surface by hydrophobic amino acid residues. However, in the protein interior they are building a common large space. This tunnel enables the substrate transport in MobA. Both the environment of the tunnel interior, characterized by distinctly divided hydrophilic and hydrophobic regions, and the size of the tunnel entrance are assumed to play an important role in enzymatic substrate selection by amphiphilic nature and molecular size of the substrate. In addition, results of the Xe-derivative structure analysis of the protein indicate the existence of a highly hydrophobic pocket, a putative oxygen-binding site, in direct proximity to the substrate-binding site. Moreover, the tunnel is suggested to be involved in oxygen transport.⁵¹

Previous mentioned investigations with the MHBH from *C. testosteroni* have been carried out with the strain KH122-3s.^{51, 57, 58} In 2008 Chang and Zylstra examined the substrate range of the MHBH MobA from the *C. testosteroni* strain GZ39, which is 98% identical to MobA from *C. testosteroni* KH122-3s.⁶⁰ As implied by the results of Hiromoto and co-workers it might be difficult to modify the substrate specificity of MobA by rational protein engineering of the substrate-binding pocket, a directed evolution approach was performed by error-prone PCR (polymerase chain reaction) mediated *in vitro* mutagenesis of MobA followed by *in vivo* screening for mutants with an altered substrate specificity. In this way numerous mutants were found, one of these, the MobA variant with the single V257A substitution, was identified to be able to transform phenol to catechol (Figure 1.4 B). This was the first example of a monooxygenase known for its conversion of phenolic acid substrates that can successfully hydroxylate phenol. Moreover, some other substituted phenols were also reported to be converted with enhanced efficiency by the MobA variant V257A.⁶⁰

Introduction

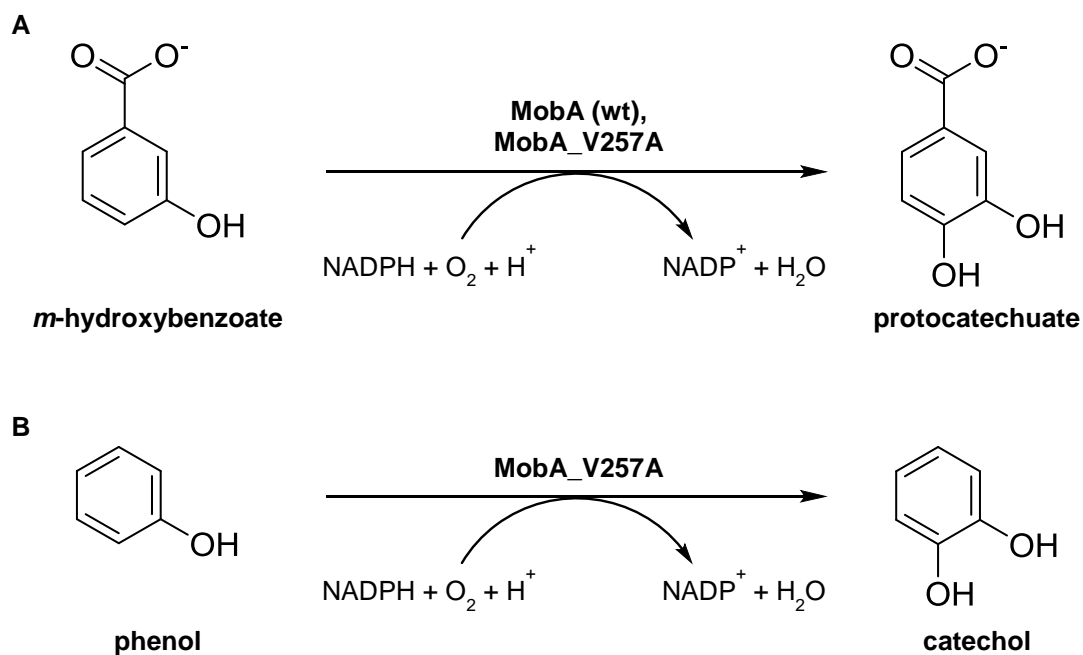


Figure 1.4: Reactions catalyzed by the wild type (wt) of the MHBH MobA (**A**) and by its variant V257A (**A** and **B**). Graphic adapted from Chang and Zylstra 2008.⁶⁰

1.3.4 Vanillyl-alcohol oxidase (VAO) from *Penicillium simplicissimum*

The vanillyl-alcohol oxidase (VAO) flavoprotein family comprises a group of oxidoreductases sharing a conserved FAD-binding domain⁶¹ and favouring the covalent binding of the cofactor FAD to this domain. This covalent linkage of the flavin cofactor to the apoprotein, taking place via a dual covalent binding of the isoalloxazine ring of the flavin,⁶² is an autocatalytic process depending on the primary folding of the polypeptide chain.^{47, 63}

Members of the VAO flavoprotein family are involved in a variety of metabolic processes in nature. A few of these members were summarized by Leferink and co-workers, arranged according to their distinct FAD-binding mode into following groups: histidyl-FAD enzymes, cysteinyl-histidyl FAD enzymes, other covalent linkages, and non-covalent flavoenzymes.⁶⁴

The vanillyl-alcohol oxidase (VAO, EC 1.1.3.38) from *Penicillium simplicissimum*^{56, 65} is one of the most investigated and best characterized flavoprotein oxidases. This enzyme is a stable homooctamer of 520 kDa.⁶⁵ Stability of this octamer is suggested to derive from dimer-dimer interactions created by a specific loop region.⁶⁴ Each subunit contains a FAD cofactor which is covalently linked to the N3 atom of a

histidine (His422) of the apoprotein via the C8 α -position of the isoalloxazine ring of the flavin.^{65, 66} The standard reaction catalyzed by VAO is the conversion of vanillyl alcohol to vanillin in the presence of molecular oxygen,^{67, 68} which is necessary for flavin re-oxidation.²⁴ In addition, VAO is active with a variety of specifically phenolic compounds⁶⁷⁻⁶⁹ and it is known for its ability to enantioselectively hydroxylate various alkylphenols.^{70, 71} However, with 4-alkylphenols having an aliphatic side-chain longer than seven carbon atoms, no catalytic activity was observed⁶⁸ due to the limited size of the VAO active site.⁶⁶ Some VAO-mediated reactions, which are oxidations, deaminations, demethylations, dehydrogenations and hydroxylations, are even of interest for industrial applications like the synthesis of natural vanillin, 4-hydroxybenzaldehyde, coniferyl alcohol and enantiomeric pure phenol derivatives.⁷²

The VAO protein is structurally different compared to the flavoprotein aromatic hydroxylases,⁶¹ resulting in a differing catalytic reaction mechanism compared to FPMOs. This mechanism has been studied intensively by stopped flow kinetic studies with 4-(methoxymethyl)phenol as a substrate. As a first step of the reaction cycle a hydride is transferred from the C α -atom of the substrate to the N5-atom of the flavin. This results in the formation of a binary complex between reduced enzyme and the *p*-quinone methide product intermediate. Next, the reduced flavin is re-oxidized by molecular oxygen with the concomitant formation of hydrogen peroxide. Finally, the *p*-quinone methide intermediate reacts with water in the active site of the enzyme yielding the final products (Figure 1.5).⁷³

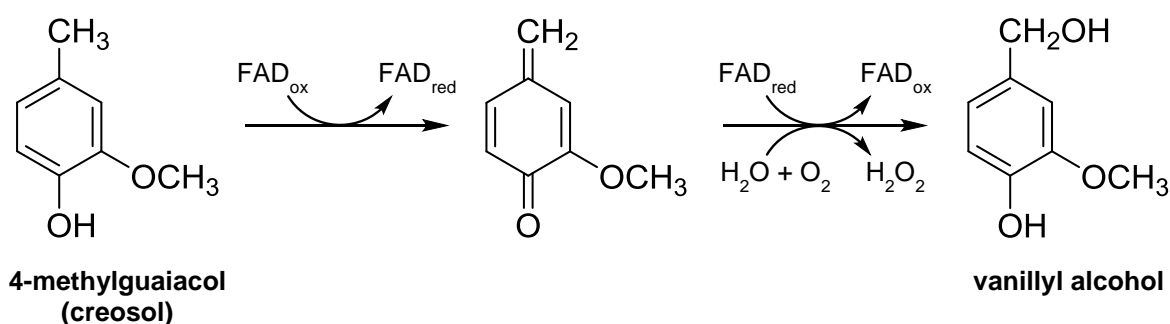


Figure 1.5: Catalytic reaction cycle of the flavoprotein oxidase VAO with the exemplary substrate 4-methylguaiacol (creosol). Graphic adapted from van den Heuvel *et al.* 2004.⁶⁹

There are several amino acid residues in the VAO protein that are known to play important roles for the catalytic function of the enzyme. Clearly, His422, which is

necessary for the covalent binding of the cofactor FAD, is one of these amino acids. Moreover, the presence of an activating nucleophile in the FAD domain, which is His61, was reported to be necessary for the covalent binding of the cofactor FAD to this His422 of the apoprotein.⁷⁴ Another key residue, which was shown to be crucial for efficient redox catalysis, is Asp170. The bulkiness of the side-chain of this amino acid is suggested to regulate the efficiency of the hydroxylation of 4-alkylphenols.⁷¹ In various studies, other amino residues of VAO were addressed by means of rational protein design/site-directed mutagenesis yielding different effects like inversion of the stereospecificity of a VAO-catalyzed reaction⁷⁵ or an increased catalytic efficiency (k_{cat}/K_M) towards the production of vanillin from the precursor creosol (2-methoxy-4-methylphenol, 4-methylguaiacol).⁶⁹

1.4 Cytochrome P450 monooxygenases

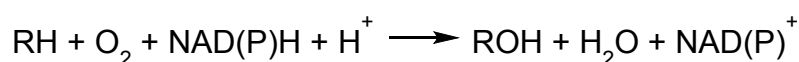
1.4.1 Nomenclature and functions

Cytochrome P450 monooxygenases (CYPs, P450s) (EC 1.14.x.x) comprise one of the largest superfamilies of soluble and membrane-bound heme-thiolate proteins. These enzymes can be found in all domains of life including plants, animals, mammals, fungi, and bacteria.⁷⁶ The terminology **P450** describes the spectral properties of the *b*-type heme-containing reddish pigmented enzymes (**P** = pigment) which show a typical absorption maximum at **450** nm in their reduced carbon monoxide bound form.⁷⁷ Classification of the P450s according to their origin and sequence identity was introduced to simplify the identification of the increasing number of P450s.⁷⁸⁻⁸² According to the “Cytochrome P450 Homepage” (<http://drnelson.uthsc.edu/CytochromeP450.html>),⁸¹ there are more than 20.000 P450 sequences classified (updated in July 2013). These sequences are available in several databases such as the “Cytochrome P450 Engineering Database” (CYPED), which is part of the BioCatNet, an internet database that integrates information on sequence, structure and function of P450s.⁸³ However, there are many more sequences known which have not been classified yet as a result of the increasing speed of identification of new P450 sequences.⁸⁴ As an example, in November 2015 the CYPED database contained more than 52.000 P450 sequences.⁸³

Cytochrome P450 monooxygenases catalyze a variety of different reactions including aliphatic and aromatic hydroxylations, epoxidations, peroxidations, sulfoxidations, aromatizations, cleavage of carbon-carbon bonds, dealkylations (*N*-, *O*-, and *S*-), deaminations, dehalogenations, desulfurations, oxidative phenolic coupling, *N*-oxide reductions and other complex reactions.^{30, 39, 85-90} Moreover, their spectrum of reactions has been expanded to non-natural reactions by protein engineering and evolution.⁹¹⁻⁹⁵ P450s play central roles in living organisms in the metabolism of drugs and are involved in the biosynthesis of numerous important compounds like antibiotics and antifungals, hormones, steroids and many more.^{86, 96-99} In addition to the above mentioned versatility of P450s they are known for their ability to catalyze regio- and stereoselective oxygenations of organic compounds.¹⁰⁰ Hence they are of great interest for a potential industrial usage in the biotechnology sector as, compared to chemical syntheses, they provide a “greener” route to the production of various high value compounds.^{39, 101, 102}

1.4.2 Catalytic reaction mechanism

The most prominent reactions catalyzed by P450s are monooxygenase reactions. Here a single oxygen atom from molecular oxygen is introduced into the substrate compound while the other oxygen atom is reduced to water⁸⁸ (Scheme 1.1).



Scheme 1.1: General monooxygenase reaction catalyzed by P450s.

For most P450-catalyzed reactions an interaction of the enzyme with one or more redox partners is required, which deliver the redox equivalents (electrons in form of hydrid ions) via electron transfer chains to the covalently cysteine-bound heme in the catalytic site of the monooxygenase domain (see section 1.4.3).¹⁰³ Commonly these electrons are supplied to the redox partners from the reduced pyridine nucleotide coenzymes NADH or NADPH.¹⁰⁴

The catalytic reaction mechanism of P450s can be divided into several successive steps (Figure 1.6). First the substrate enters the active site and binds to the P450s ferric ion of the heme moiety with concomitant dissociation of a water molecule (1, 2).

This goes along with a shift in the spin state of the heme iron from low-spin to high-spin resulting in a positive shift in the reduction potential.¹⁰⁵ This allows the reduction of the heme ferric iron (Fe^{III}) (2) to the ferrous iron (Fe^{II}) (3) by a first one-electron transfer from the redox partner. Next the reduced ferrous (Fe^{II}) heme iron binds molecular oxygen building the ferric-superoxy-species (4). The subsequent delivery of a second electron (second one-electron reduction) leads to the formation of a negatively charged ferric-peroxo-intermediate (5a). Protonation of this intermediate results in a ferric-hydroperoxy-complex (5b). A successive second protonation, which occurs at the distal oxygen atom, results in cleavage of the dioxygen bond of the ferric-hydroperoxy-complex. This is accompanied by the release of a water molecule and the formation of the ferryl-oxo-complex, the reactive intermediate in P450s called Compound I (6).

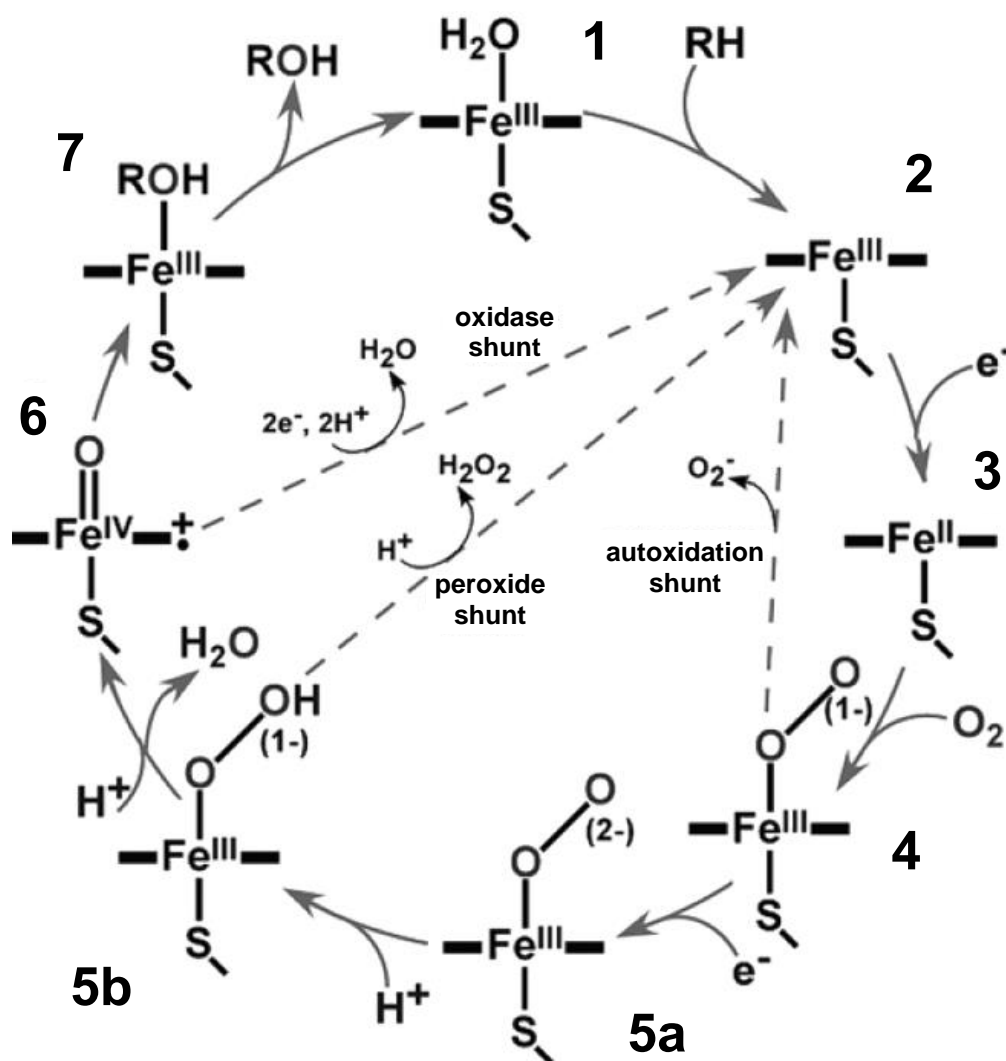


Figure 1.6: Catalytic reaction cycle of cytochrome P450 monooxygenases. Graphic adapted from Denisov *et al.* 2005.¹⁰⁴

Finally the oxygenation (frequently hydroxylation) of the substrate by the activated oxygen atom of Compound I takes place (7) and the oxygenated substrate is subsequently released.^{104, 106}

In addition to the above mentioned catalytic reaction steps, the catalytic reaction cycle of P450s can be interrupted by three so-called uncoupling pathways: the autoxidation shunt, the peroxide shunt, and the oxidase shunt. The autoxidation shunt happens if the delivery of the second electron to the ferric-superoxy-complex (4) is disturbed, resulting in the return of the enzyme to the resting state and the formation of superoxide. The peroxide shunt occurs due to a collapse of the ferric-hydroperoxy-complex (5b) with the concomitant release of hydrogen peroxide and thus resulting in the unproductive two-electron reduction of oxygen. The oxidase shunt means that instead of the oxygenation of the substrate molecule by the reactive ferryl-oxo-complex (6, Compound I) a molecule of water is built.¹⁰⁴ These uncoupling pathways may occur as a result of delayed electron delivery or protonation or by the missing or inappropriate positioning of the substrate in the active site for oxidative attack and they all have in common that reducing equivalents from the coenzyme NAD(P)H are consumed without the concomitant formation of product.¹⁰⁶

1.4.3 Classification

Usually the electrons necessary for P450-catalyzed reactions are transferred from the coenzymes NAD(P)H to the active site of the monooxygenase domain via flavoproteins and/or iron-sulfur redox proteins. Today, based on the topology of these redox partners ten classes of P450 systems differing in their electron transfer chain can be distinguished.¹⁰⁷ In the traditional classification of P450 systems only two main classes were differentiated: the class I system, separated into a bacterial and a mitochondrial system, and the class II microsomal system. The class I systems comprise three soluble components. While in the bacterial class I system (Figure 1.7 A) the electrons from NAD(P)H are transferred by a NAD(P)H-dependent FAD-containing flavoprotein reductase via an iron-sulfur protein (ferredoxin [2Fe-2S]) to the P450 enzyme,¹⁰⁸ in the mitochondrial class I system (Figure 1.7 B) the ferredoxin is the only soluble electron transfer component of the mitochondrial matrix. Here the reductase and the P450 component are membrane-associated and membrane-

Introduction

bound to the inner mitochondrial membrane, respectively.¹⁰⁷ P450_{cam} from *Pseudomonas putida* was described as the prototype of bacterial P450 systems.^{109, 110} In the class II microsomal system (Figure 1.7 C) usually a membrane-anchored P450 receives electrons from a membrane-anchored NADPH-dependent cytochrome P450 reductase (CPR), which has both a FAD and a FMN cofactor.^{107, 108, 111} However, there are a few exceptions from the class II system described above, for example CYP105A3 from *Streptomyces carbophilus* which is a soluble P450 system.¹¹²

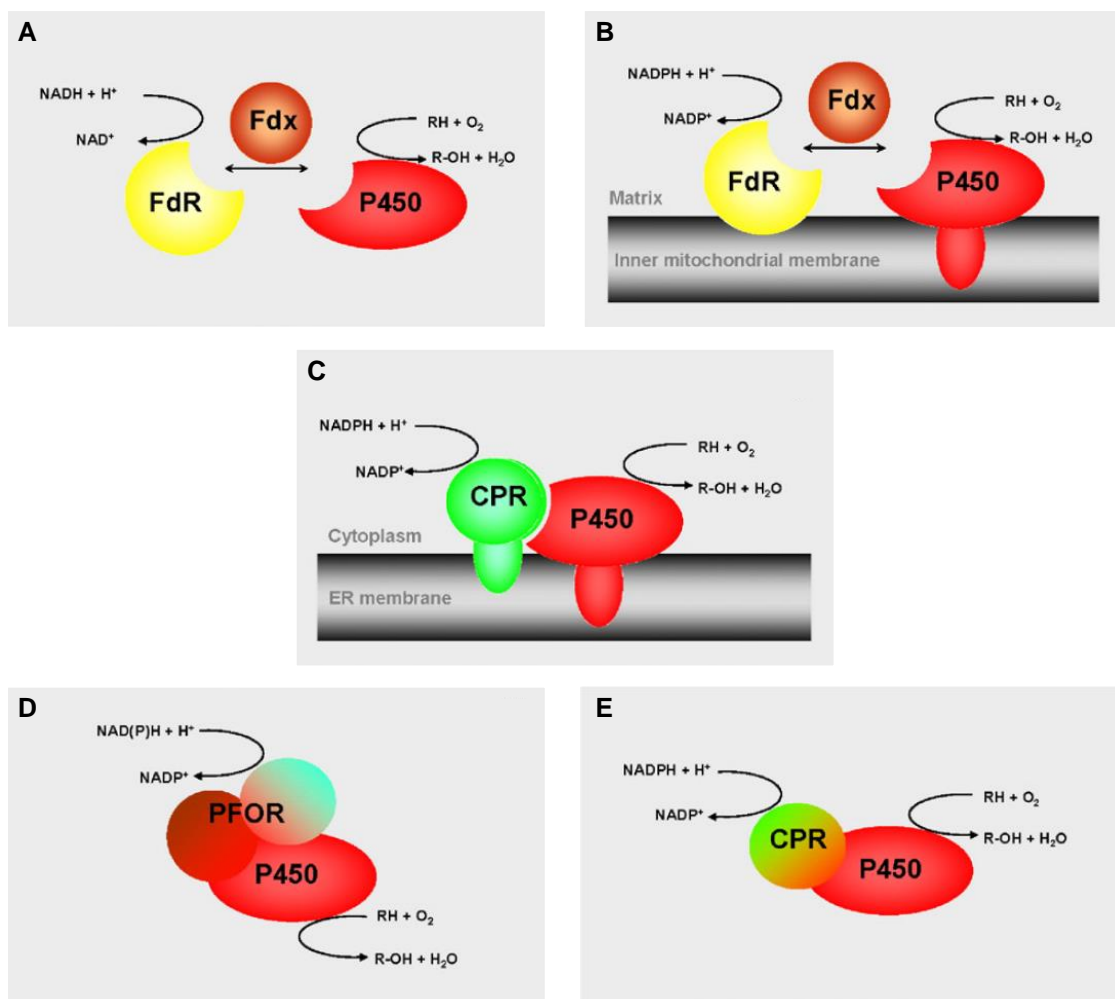


Figure 1.7: Schematic organization of P450 systems differing in their type of employed electron transfer chain. **A** Class I, bacterial system. **B** Class I, mitochondrial system. **C** Class II, microsomal system. **D** Class VII, bacterial [PFOR]–[P450] fusion system. **E** Class VIII, bacterial [CPR]–[P450] fusion system. Graphic adapted from Hannemann *et al.* 2007.¹⁰⁷

Today there are some more P450 redox systems known displaying different topological arrangements of their redox partners compared to those covered by the

traditional classification. As already mentioned, ten classes can be distinguished.¹⁰⁷ Two of those, class VII (Figure 1.7 D) and class VIII (Figure 1.7 E), are of special interest as they comprise natural fusion systems. In CYP116B2 (P450RhF) from *Rhodococcus* sp., an exemplary member of the class VII bacterial [PFOR]-[P450] fusion system, the N-terminal P450 domain is fused to the C-terminal PFOR (phthalate-family oxygenase reductase) domain by a short linker region of 16 amino acids, with the reductase component containing a FMN-binding domain, a NADH-binding domain and a [2Fe-2S] ferredoxin domain.¹¹³ As the most investigated example of the class VIII bacterial [CPR]-[P450] fusion system, CYP102A1 (P450BM3, EC 1.14.14.1) from *Bacillus megaterium* can be mentioned. Its heme-containing P450 domain is fused to the FAD- and FMN-containing diflavin reductase domain (CPR) by a short protein linker.¹¹⁴ The other classes of P450 systems, not mentioned here, are described in detail by Hannemann and co-workers.¹⁰⁷

1.4.4 CYP102A1 from *Bacillus megaterium*

The cytosolic fatty acid hydroxylase CYP102A1 (P450 BM3, EC 1.14.14.1) from *Bacillus megaterium* belongs to the class VIII of cytochrome P450 systems. It is a natural self-sufficient fusion enzyme consisting of a FAD- and FMN-containing diflavin reductase domain (CPR) which is fused to the heme-containing domain via a short protein linker^{114, 115} requiring only NADPH and molecular oxygen to function.¹¹⁶ CYP102A1 has a molecular weight of 119 kDa¹¹⁴ and is a single polypeptide which is fully soluble.¹¹⁷ Since its identification more than forty years ago^{116, 118-120} it has been extensively studied,¹²¹ not least due to the fact that it is the P450 enzyme with the highest known activity of all investigated P450 systems so far.¹²²⁻¹²⁴ CYP102A1 catalyzes not only the hydroxylation of saturated straight-chain substrates with pentadecanoic acid as the preferred substrate, but also the hydroxylation of other substrates like unsaturated fatty acids.¹¹⁶ Thus, CYP102A1 was originally found to be a fatty acid hydroxylase with a preference for medium- to long-chain fatty acids (from lauric acid (C12) to octadecanoic acid (C18)), which were hydroxylated exclusively at sub-terminal positions.^{115, 116, 121, 122} In addition to saturated and unsaturated fatty acids, CYP102A1 is known to accept fatty amides and alcohols,¹²⁵ hydroxylated fatty acids¹²⁶⁻¹²⁸ and ω -oxo fatty acids¹²⁹ as substrates. Within a period of many years in the past, protein engineering techniques like rational protein design and directed

evolution were applied to broaden the substrate profile of CYP102A1. Hence, a variety of CYP102A1 variants was generated with mutations at numerous amino acid residues aiming at the oxidation of non-natural substrates like terpenes, alkanes and aromatic compounds making these enzyme variants suitable for applications in commercial-scale industrial processes, for example the synthesis of fine chemicals.^{39, 121} Moreover, new reactions catalyzed by CYP102A1 variants were established by means of enzyme engineering. Here, the carbene transfer to olefins resulting in highly diastereo- and enantioselective cyclopropanation reactions, which were catalyzed by engineered variants of CYP102A1, can be mentioned as an interesting example.¹³⁰

Two of the most addressed mutation sites in CYP102A1 are the amino acid residues F87 and A328.¹²¹ These two amino acids are located opposite to each other in high proximity to the heme group of the active site and are therefore assumed to be key residues playing important roles concerning the positioning of the substrate to the activated heme oxygen, thus influencing activity, specificity and selectivity of the enzymes catalytic reactions.^{121, 131} In a recent study both residues have been mutated by means of rational protein design to generate a focused minimal and highly enriched mutant library. Therefore the five hydrophobic amino acids alanine, leucine, isoleucine, phenylalanine, and valine were combined to a total of 24 variants in addition to the wild type and were successfully screened for the conversion of various terpene substrates (geranylacetone, nerylacetone, (4*R*)-limonene, (+)-valencene). Eleven variants converted at least one of the substrates displaying distinct differences in the regio- or stereoselectivity compared to the wild type enzyme.¹³² Further variants of CYP102A1, generated based on the results of a preliminary screening of the above mentioned focused library, were successfully screened for hydroxylation activity towards cyclic and acyclic alkanes (*n*-octane, cyclooctane, cyclodecane and cyclododecane).¹³³ Hence the generation of variants of CYP102A1 is a powerful tool towards the application of CYP102A1 in biotransformations of non-natural substrates with improved activities and selectivities.

1.4.5 CYP116B3 from *Rhodococcus ruber*

Members of the CYP116B subfamily belong to the class VII of P450 redox systems. As mentioned above, these enzymes are natural P450-redox partner fusion enzymes

Introduction

consisting of a P450 component linked to a reductase component of the PFOR-type (phthalate family oxygenase reductase) containing a FMN-binding domain, a [2Fe-2S] ferredoxin domain and a NAD(P)H-binding domain.^{113, 134-136} Same as CYP102A1 from *B. megaterium* the CYP116B-type P450-redox partner fusion enzymes have all functionally necessary protein partners covalently linked together and are thus catalytically self-sufficient requiring only the cofactor NAD(P)H as an electron donor and a substrate for oxidative catalysis.¹³⁷ Examples for the above mentioned CYP116B-type enzymes are CYP116B1 from *Cupriavidus metallidurans*¹³⁸ and CYP116B2 (P450RhF) from *Rhodococcus* sp. NCIMB 9784.^{113, 139} Another example of a natural self-sufficient P450 belonging to this CYP116B-type subfamily of fusion enzymes is CYP116B3 from *Rhodococcus ruber* DSM 44319. The gene of this P450 enzyme was identified by directional genome walking using PCR,¹⁴⁰ and the respective protein was found to be built of 771 amino acids with a sequence identity of 93% compared to P450RhF from *Rhodococcus* sp. NCIMB 9784.¹³⁹ CYP116B3 was characterized and described as being composed of a ferredoxin domain and a FMN-containing NADPH-dependent flavoprotein reductase domain fused to a heme-containing P450 domain. It was recombinantly expressed in *Escherichia coli* with a N-terminal His-tag, purified by immobilized metal affinity chromatography (IMAC) and an apparent molecular mass of 89 kDa was detected by SDS-PAGE analysis. With the fluorescent substrate 7-ethoxycoumarin (7-EC), shown previously to be successfully converted to 7-hydroxycoumarin (7-HC) by *O*-dealkylation with the highly homologous P450RhF,¹³⁹ the oxidation activity was investigated. In addition to 7-EC, the CYP116B3-catalyzed conversion of various monocyclic and polycyclic aromatic hydrocarbon substrates with and without alkyl side chains (acenaphthene, fluorene, indene, naphthalene, toluene, *m*-xylene and ethyl benzene) was investigated. While some of these substrates, for example naphthalene, were monohydroxylated at the ring, the alkyl aromatic substrates were hydroxylated exclusively at the side chains. This alkyl chain hydroxylation activity towards aromatic substrates makes CYP116B3 an interesting enzyme for biocatalytic applications, though, compared to the conversion of 7-EC, the activity towards aromatic hydrocarbons was much lower.¹⁴⁰ In order to increase the dealkylation activity of CYP116B3 a site-specific mutagenesis approach was combined with directed evolution to generate 7800 mutant variants of CYP116B3 followed by high-throughput screening towards the model substrate 7-EC. Finally, after four rounds of

directed evolution, a CYP116B3 variant with five amino acid substitutions (A86T/T91S/A109F/I179F/I267L) was found with a 240-fold increased activity compared to the wild type enzyme.¹⁴¹ Hence, CYP116B3 is an interesting enzyme candidate for the conversion of aromatic compounds.

1.5 Vanillin – an aromatic compound of interest

Vanillin (4-hydroxy-3-methoxybenzaldehyde; CAS no. 121-33-5) is one of the most popular and widely used flavor compounds in the world. Besides being extensively applied as a flavor ingredient in the food and beverages industries,^{142, 143} it is applied as a fragrance component in perfumes and cosmetics¹⁴⁴ and as an intermediate in the chemical and pharmaceutical industries.¹⁴⁵ Originally, natural vanillin is extracted from the orchids *Vanilla planifolia*, *Vanilla tahitiensis*, and *Vanilla pompon*.^{143, 146} The vanilla extract comprises more than 200 components. Though vanillin is the most characteristic flavor component of vanilla, its concentration is only 1-2% in cured vanilla pods.¹⁴⁷ Pure vanillin is a white crystalline powder with a low solubility in water at ambient temperature.^{142, 143} More than 15000 tons of vanillin were produced and sold in 2010. However, the production of vanillin, which is naturally obtained from vanilla pods, is a very time consuming and expensive process (1 kg of vanillin produced per 500 kg of vanilla pods).¹⁴⁷ In addition, a number of factors like the weather conditions as well as plant diseases influence the market price for vanillin extracted from vanilla pods. Thus it varies between about 1200 US\$/kg and more than 4000 US\$/kg.¹⁴⁵ Due to the above mentioned restricted availability and high price and due to the steadily increasing global demand for vanillin, less than 1% of vanillin is produced from vanilla pods worldwide today.¹⁴⁸ Nowadays, the majority of vanillin is produced synthetically, reducing the market price below 15 US\$/kg, via alternative processes like the chemical synthesis from eugenol, guaiacol, or lignin as starting materials or via biotechnology-based approaches starting from substances such as ferulic acid, eugenol, isoeugenol, glucose, lignin, phenolic stilbenes, vanillic acid or aromatic amino acids obtained from renewable feedstocks and employing various production systems like fungi, bacteria, yeasts, plant cells, or genetically engineered production strains. Compared to the environmentally rather unfriendly chemical synthesis of vanillin, biotechnology-based approaches have several advantages, for example the reduced production of waste.^{142, 143} As an example of

such a biotechnological process, the synthesis of vanillin from isoeugenol with a yield of 32.5 g/l employing a *Bacillus fusiformis* strain can be mentioned.¹⁴⁹ Moreover, the biotechnology-based microbial synthesis of vanillin from natural precursors yields “natural” vanillin according to the current European¹⁵⁰ and US food legislation.¹⁵¹

1.6 Aim of the project

In the scope of the present project, the selective hydroxylation of aromatic compounds, which is one of the most challenging chemical reactions, was addressed enzymatically. In order to establish a novel route to the valuable aromatic vanillin, as an example of consecutive enzyme-catalyzed oxidation reactions, a selection of enzymes was investigated. According to the proposed reaction pathway (Figure 1.8), vanillin was intended to be synthesized by conversion of the simple aromatic substrate 3-methylanisole via the intermediate products 3-methoxybenzyl alcohol, 4-methylguaiacol and vanillyl alcohol. For this purpose, the following tasks were defined:

- Preparation (expression, purification, determination of activity) of a selection of enzymes and application in *in vitro* biotransformation reactions, combined with efficient cofactor recycling, for the selective oxidation of both the aromatic substrate and intermediate compounds of the vanillin synthesis pathway.
- Improvement of activity and selectivity of the biocatalysts, concerning the hydroxylation of the aromatic target substrates, by means of (rational) protein design.
- Combination of the most suitable enzymes and enzyme variants in one reaction system (both *in vitro* and *in vivo*) for the final cascade synthesis of vanillin.

Introduction

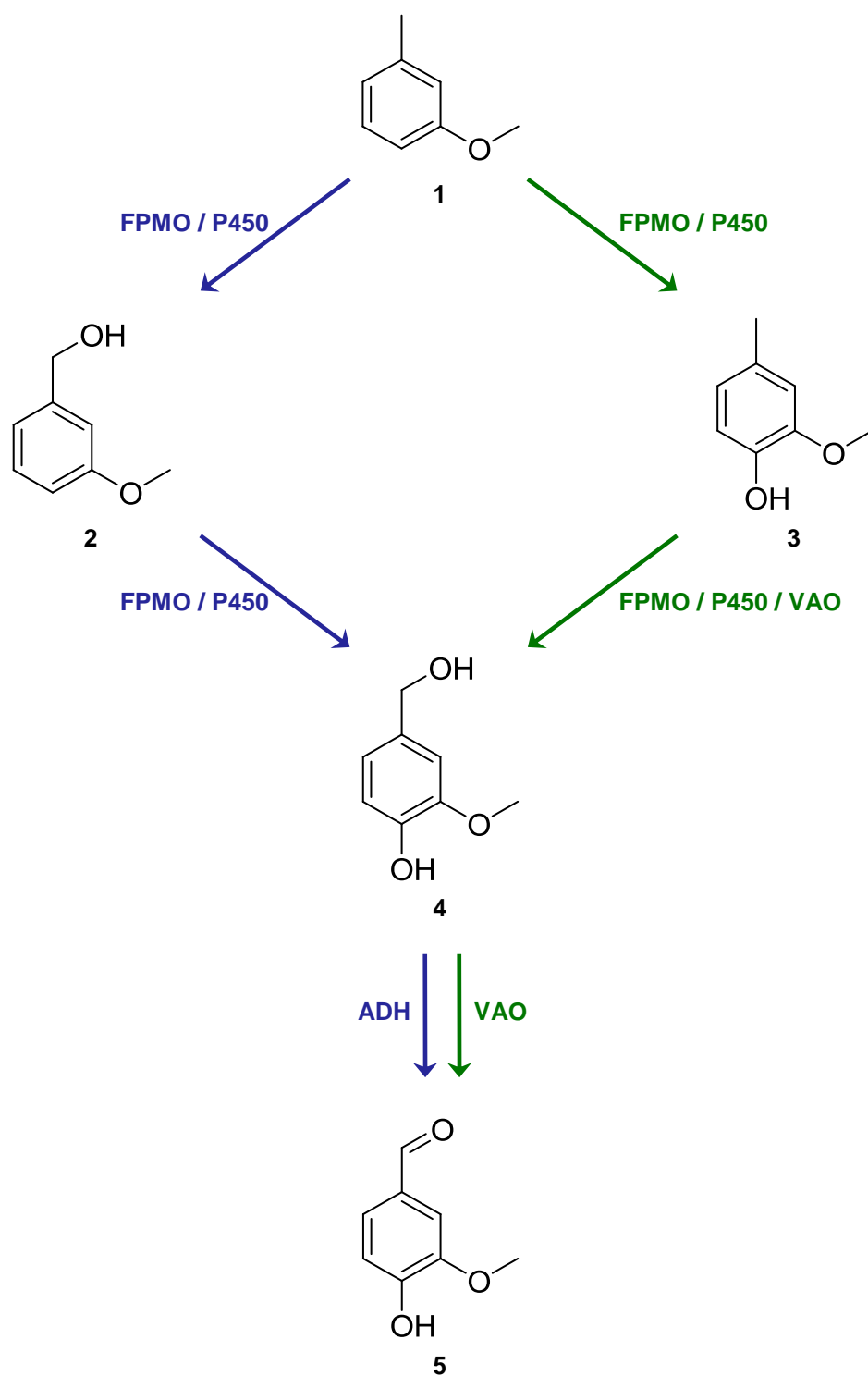


Figure 1.8: Proposed reaction pathway for the synthesis of vanillin. Two different enzyme-catalyzed cascade reaction routes are conceivable, starting with the substrate 3-methylanisole (1). Via the intermediate products 3-methoxybenzyl alcohol (2) or 4-methylguaiaicol (3), the vanillin precursor vanillyl alcohol (4) could be produced and further converted to vanillin (5) either by combination of a flavoprotein monooxygenase (FPMO) and/or a cytochrome P450 monooxygenase (P450) and an alcohol dehydrogenase (ADH) (**blue route**) or alternatively by combination of a FPMO and/or a P450 with a vanillyl alcohol oxidase (VAO) (**green route**).

2 Materials and methods

2.1 Materials

2.1.1 Genes, proteins, vectors, primers and strains

Genes, proteins, vectors, primers and strains used in this work are listed in supplementary material section 7.1.

The genes *mobA* from *Comamonas testosteroni* GZ39, *cyp116B3* from *Rhodococcus ruber* DSM 44319 and *vaoA* from *Penicillium simplicissimum* CBS 170.90 were obtained as synthetic genes (GeneArt[®] Gene Synthesis, Life Technologies GmbH, Darmstadt, Germany), codon-optimized for the heterologous expression in *Escherichia coli*. Plasmid pET-11a harbored *mobA* (integrated between the restriction sites *NdeI* and *BamHI* with a C-terminal His₆-tag), pET-24a(+) contained *cyp116B3* (integrated between the restriction sites *NdeI* and *EcoRI*, with a N-terminal His₆-tag) and pMK-RQ comprised *vaoA* (integrated between the restriction sites *NdeI* and *XhoI*). Plasmids harboring *cyp102A1* wild type gene (from *Bacillus megaterium* ATCC 14581) and some variants thereof (integrated in pET-28a(+) and pET-22b(+)) between the restriction sites *BamHI* and *EcoRI*) were provided by Dr. Alexander Seifert (Institute of Technical Biochemistry (ITB), University of Stuttgart, Stuttgart, Germany) as indicated in Table 7.2 (supplementary material).

2.1.2 Media, buffers and solutions

Media, buffers and solutions used in this work are described in detail in supplementary material section 7.2.1.

2.1.3 Chemicals and cloning enzymes

All chemicals were of analytical-reagent grade or higher quality. Tryptone from casein was purchased from Fluka (Buchs, Switzerland). Yeast extract and vanillin were bought from Roth (Karlsruhe, Germany). β -NADPH tetrasodium salt was obtained

from Acros Organics (Geel, Belgium). Glucose-6-phosphate dehydrogenase from *Saccharomyces cerevisiae* was from Roche Diagnostics (Mannheim, Germany). 3-Methylanisole, 4-methylguaiacol, 3-methoxybenzyl alcohol, vanillyl alcohol, vanillin and 3-methoxy-5-methylphenol were purchased from Alfa Aesar (Karlsruhe, Germany). 2-Methoxy-6-methylphenol was from Frinton Laboratories (Hainesport, NJ, USA). D-Glucose-6-phosphate disodium salt hydrate, *m*-cresol, 4-methoxy-2-methylphenol, methylhydroquinone and all other chemicals, solvents, and buffer components were purchased from Sigma-Aldrich (Schnelldorf, Germany). Endonucleases (*Bam*HI, *Dpn*I, *Eco*RI, *Nde*I and *Xho*I), *Pfu* DNA Polymerase, T4 DNA ligase and the respective necessary buffers were obtained from Fermentas (now Fisher Scientific, Schwerte, Germany).

2.2 Methods

2.2.1 General methods

General molecular biological and microbiological experiments were carried out by standard methods.¹⁵² Plasmid isolation was performed with the Zyppy™ Plasmid Miniprep Kit (Zymo Research, Irvine, CA, USA) following the manufacturer's protocol. Purification of DNA fragments was carried out by agarose gel electrophoresis followed by using the QIAquick Gel Extraction Kit (QIAGEN, Hilden, Germany) according to the manufacturer's protocol. DNA concentrations were determined with the NanoDrop™ 1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).

2.2.2 Preparation of competent *E. coli* cells

For the preparation of chemically competent cells 0.5 ml of an overnight (O/N) culture, which was grown from a single colony and contained the desired *Escherichia coli* (*E. coli*) strain (DH5 α or BL21(DE3)), were used for inoculation of 50 ml of sterile lysogeny broth (LB) medium, followed by incubation at 37°C and 180 rpm until reaching an optical density at a wavelength of 600 nm (OD₆₀₀) of 0.4 – 0.6. After keeping the cell culture on ice for 15 min, the cells were harvested by centrifugation at 4°C and 3200 × g for 10 min. The cells were resuspended in 20 ml of ice-cold Tfb I

buffer, kept on ice for 15 min and centrifuged again as described above. The supernatant (S/N) was discarded and the cell pellet was resuspended in 2 ml of ice-cold Tfb II buffer and incubated on ice for 15 min. Aliquots of 200 μ l in sterile Eppendorf tubes were immediately frozen in liquid nitrogen and stored at -80°C until use.

2.2.3 Transformation of competent *E. coli* cells

Competent *E. coli* cells were thawed on ice and 10 μ l of competent cells (DH5 α or BL21(DE3)) were mixed under sterile conditions with 1 μ l of plasmid containing a target gene insert for retransformation. Alternatively, for transformation after ligation, 100 μ l of competent cells (DH5 α) were mixed with 10 μ l of ligation mixture. After incubation on ice for 30 min a heat shock treatment at 42°C for 45 s was performed, followed by incubation on ice for 2 min. Depending on the volume of the sample, 80 μ l or 800 μ l of sterile SOC medium were added, respectively, and the cells were incubated at 37°C and 500 rpm for 1 h. For re-transformation, 100 μ l of the transformed cells were directly spread onto LB agar plates containing the respective antibiotic and incubated at 37°C overnight. For transformation after ligation the cells were harvested by centrifugation at room temperature (RT) at $11000 \times g$ for 1 min. 810 μ l of the S/N were discarded by pipetting and the cell pellet was resuspended in the residual S/N volume. The complete resuspension was spread onto LB agar plates containing the respective antibiotic and incubated at 37°C overnight. Sterile DEPC-treated water instead of plasmid was used as negative control for the re-transformation. For transformation after ligation a ligation mixture with sterile DEPC-treated water instead of a gene insert was applied.

Single cell colonies of the transformation plates were used for inoculation of 5 ml sterile LB medium containing the respective antibiotic and incubated at 37°C and 180 rpm overnight. Plasmid isolation from the O/N cultures was performed with the ZippyTM Plasmid Miniprep Kit (Zymo Research, Irvine, CA, USA) as recommended by the manufacturer. Glycerol stocks of cells were prepared by mixing of equal volumes of the O/N culture and a 50% glycerol solution. Aliquots of 100 μ l in sterile tubes were stored at -80°C until use.

2.2.4 Agarose gel electrophoresis

Agarose gels with a concentration of 1% (1% agarose in TAE buffer) were cast using a gel comb for the loading of small (5-10 μ l) or larger (25-50 μ l) sample volumes for analytical or preparative agarose gel electrophoresis, respectively. For the detection of the DNA under UV light the agarose gel was supplemented with GelRed™ (Biotium, Hayward, CA, USA) in a 1x concentration. As a reference standard, 1 μ l of GeneRuler™ 1 kb DNA ladder from Fermentas (now Fisher Scientific, Schwerte, Germany) was mixed with 1 μ l of DNA sample loading buffer and deionized water to a final volume of 5 μ l and loaded onto the gel. DNA samples were mixed with DNA sample loading buffer (ratio of loading buffer to DNA sample of 1:5) and loaded onto the gel. The gel was run in 1x TAE buffer at 120 V for 40 min. DNA bands of interest were excised from the gel under UV light and extracted using the QIAquick Gel Extraction Kit (QIAGEN, Hilden, Germany) as recommended by the manufacturer.

2.2.5 Site-directed mutagenesis

Site-directed mutations were introduced according to the QuikChange standard protocol using the oligonucleotides indicated in supplementary material section 7.1.4 (Table 7.5). Successful mutagenesis was verified by automated DNA-sequencing (GATC Biotech, Konstanz, Germany).

MobA: The variant V257A of MobA was mutated using the plasmid pET-11a_MobA.

CYP102A1: Variants of CYP102A1 were generated using plasmids pET-22b(+) and pET-28a(+) harboring the gene of the CYP102A1 wild type or respective mutant (in the case of multiple mutations). Some vector constructs were provided, as indicated in Table 7.2 (supplementary material; respective ITB No. tagged with *).

VAO: The variants F454Y and T505S of VAO were created using the plasmid pET-22b(+)_VAO.

2.2.6 Generation of a focused mutant library of CYP116B3

2.2.6.1 Homology modelling of CYP116B3

In cooperation with the bioinformatics group of the ITB (University of Stuttgart, Stuttgart, Germany), a homology model of CYP116B3 was generated by Dr. Łukasz Gricman. First, the CYP116B3 wild type amino acid sequence was used for a blastp search (protein-protein BLAST) against the Protein Data Bank proteins on the NCBI BLAST Homepage.^{153, 154} One sequence was chosen and aligned to the sequence of the monooxygenase domain of CYP116B3 employing a standard numbering scheme for class I P450 sequences based on structural information.¹⁵⁵ On the basis of this alignment, a homology model of the CYP116B3 monooxygenase domain was generated with SWISS-MODEL.¹⁵⁶⁻¹⁶⁰

2.2.6.2 Identification of mutation sites

Based on the homology model of the CYP116B3 monooxygenase domain, sequential information and literature information,¹⁴¹ five amino acid positions (A109, V316, A317, A318, F420) located close to the active site were chosen for site-directed mutagenesis (performed as mentioned in section 2.2.5 with the respective primers (Table 7.4, supplementary material)). Each amino acid at the selected positions was replaced by the hydrophobic amino acids alanine (Ala, A), phenylalanine (Phe, F), isoleucine (Ile, I), leucine (Leu, L) and valine (Val, V), respectively.

2.2.7 Protein expression and purification

2.2.7.1 General expression and purification protocol

General expression protocol: For protein expression, the particular vector construct was transformed into competent *E. coli* BL21(DE3) cells and transformants were selected on LB agar plates with ampicillin (100 µg ml⁻¹ for pET-11a and pET-22b(+)) or kanamycin (30 µg ml⁻¹ for pET-24a(+) and pET-28a(+)) vector constructs, respectively. An overnight (O/N) culture of 5 ml LB medium containing

Materials and methods

the adequate antibiotic was grown from a single colony at 37°C with shaking at 180 rpm. 400 ml LB medium supplemented with antibiotic was inoculated with 2 ml O/N culture and grown at 37°C and shaking at 180 rpm. At an OD₆₀₀ of 0.7-0.8 expression was induced by supplementation of the culture with inducer isopropyl- β -D-thiogalactopyranoside (IPTG) and grown for a defined time at reduced temperature and shaking speed (exact values are mentioned in sections 2.2.7.2 – 2.2.7.5). Cells were harvested by centrifugation (10800 \times g, 20 min, 4°C) and resuspended in buffer solution (50 mM Tris-HCl buffer pH 7.5 containing 500 mM NaCl, 5% glycerol, 0.1 mM phenylmethylsulfonyl fluoride (PMSF)). Cells were disrupted by sonication on ice (6 x 1 min, interspaced by 30 sec), cell debris was removed by centrifugation (37000 \times g, 30 min, 4°C), and the soluble protein fraction was recovered, filtered through a 0.45 μ m sterile filter and aliquots were stored at -20°C. Samples of protein expression were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

Protein purification: Protein purification was performed from the soluble protein fraction by Ni-based immobilized metal affinity chromatography (IMAC) using a 5 ml HisTrap FF column (GE Healthcare, Freiburg, Germany) with an ÄKTA protein purification system (GE Healthcare, Freiburg, Germany), taking advantage of the His₆-tag. The soluble protein fraction was diluted with equilibration buffer (50 mM Tris-HCl buffer pH 7.5 containing 500 mM NaCl, 5% glycerol, 0.1 mM PMSF) and applied to the column which was pre-equilibrated with the same buffer. Non-specifically bound proteins were removed from the column in washing steps with equilibration buffer containing low imidazole concentrations. Specifically bound protein was eluted with equilibration buffer containing a higher imidazole concentration (exact concentrations are mentioned in section 2.2.7.2 and 2.2.7.3). Fractions containing protein of interest were pooled, concentrated via ultrafiltration with an Amicon[®] Ultra 4 ml filter (Merck Millipore, Darmstadt, Germany) with a molecular weight cutoff (MWCO) of 10 kDa and dialyzed against 50 mM Tris-HCl buffer pH 7.5 containing 5% glycerol and 0.1 mM PMSF using a Spectra/Por[®] dialysis membrane (Spectrum Laboratories, Compton, CA, USA) with a MWCO of 12-14 kDa. Aliquots of the purified protein were stored at -20°C. Samples of protein purification were analyzed by SDS-PAGE.

2.2.7.2 Expression and purification of MobA

Expression study: In an initial expression study, the expression of MobA wild type (pET-11a_MobA) was performed at 20°C, 25°C and 30°C with 0.1 mM, 0.5 mM and 1.0 mM IPTG for expression induction and incubation for up to 96 h with shaking at 180 rpm (expression according to the protocol mentioned in section 2.2.7.1).

Expression for *in vitro* biotransformations: Expression of MobA wild type and variant V257A for *in vitro* biotransformations was performed after induction with 0.5 mM IPTG at 20°C and 140 rpm O/N following the protocol mentioned in section 2.2.7.1.

Purification of MobA: During protein purification of MobA wild type enzyme, target protein was eluted from the IMAC-column with equilibration buffer containing 100 mM imidazole, after washing steps with 20-50 mM imidazole.

2.2.7.3 Expression and purification of CYP116B3

Expression study: In an initial expression study, the expression of CYP116B3 wild type (pET-24a(+)_CYP116B3) was investigated at 20°C, 25°C and 30°C after induction with 0.1 mM, 0.5 mM and 1.0 mM IPTG and incubation for up to 48 h with shaking at 180 rpm (expression according to the protocol mentioned in section 2.2.7.1).

Cloning and expression of His-tag variants of CYP116B3 wild type: His-tag variants of CYP116B3 were constructed by PCR amplification (PCR protocol with *Pfu* DNA Polymerase: i) 95°C for 4 min; ii) 30 cycles of 95°C for 1 min, 72°C±4°C for 1.5 min, 72°C for 5 min; iii) 72°C for 5 min) based on pET-24a(+)_CYP116B3 using the oligonucleotides mentioned in Table 7.4 (supplementary material). The resulting fragments and empty pET-24a(+) vector were digested with the restriction enzymes *NdeI* and *EcoRI* at 37°C for 5 h and purified by agarose gel electrophoresis (section 2.2.4). Fragments and linearized vector were ligated in a vector to insert ratio of 1:3 with T4 DNA ligase and T4 ligation buffer. The ligation was performed at RT for 1 h. After transformation and plasmid amplification in *E. coli* DH5α, the plasmids were isolated and the sequence of the resulting DNA constructs, encoding for CYP116B3

Materials and methods

with a N-terminal MetGlySerSer (MGSS)-sequence prior to the His₆-tag (pET-24a(+)_CYP116B3_Var1), with a C-terminal His₆-tag (pET-24a(+)_CYP116B3_Var2) and without a His-tag (pET-24a(+)_CYP116B3_Var3), was verified by automated DNA-sequencing (GATC Biotech, Konstanz, Germany). Expression of the His-tag variants of CYP116B3 was performed following the general expression protocol (section 2.2.7.1) with 0.1 mM IPTG induction, co-addition of 0.5 mM 5-aminolevulinic acid (5-ALA) and 0.1 mM FeSO₄ and incubation at 20°C and 140 rpm O/N.

Cloning and expression of CYP116B3 on pET-22b(+): For expression of CYP116B3_Var1 in *E. coli* BL21(DE3) using a pET-22b(+) vector system, the *cyp116B3_Var1* gene was cloned from pET-24a(+) into pET-22b(+) by digestion of pET-24a(+)_CYP116B3_Var1 and pET-22b(+) with the restriction enzymes *Nde*I and *Eco*RI at 37°C for 5 h. Insert-DNA and linearized vector were purified by agarose gel electrophoresis, ligated in a vector to insert ratio of 1:3 with T4 DNA ligase and T4 ligation buffer (1 h incubation at RT) and transformed into *E. coli* DH5α for plasmid amplification. The plasmids were isolated and the sequence of the pET-22b(+)_CYP116B3_Var1 construct was verified by automated DNA-sequencing (GATC Biotech, Konstanz, Germany). Expression of this vector system, according to the general expression protocol (section 2.2.7.1), was induced with 0.4 mM IPTG (co-addition of 0.5 mM 5-ALA and 0.1 mM FeSO₄) and incubated at 20°C and 140 rpm O/N.

Expression for *in vitro* biotransformations: Expression of CYP116B3 wild type and variants for *in vitro* biotransformations was performed in *E. coli* BL21(DE3) with 0.1 mM IPTG induction and co-addition of 0.1-0.5 mM 5-ALA and 0.1 mM FeSO₄, following the protocol mentioned in section 2.2.7.1. After induction, cell cultures were incubated at 20°C and 140 rpm O/N.

Purification of CYP116B3: During protein purification of CYP116B3 wild type enzyme, target protein was eluted from the IMAC-column with equilibration buffer containing 50 mM imidazole, after a washing step with 20 mM imidazole.

2.2.7.4 Expression of CYP102A1

Heterologous expression of CYP102A1 wild type and mutants in *E. coli* BL21(DE3) was done as reported previously,¹⁶¹ similar to the expression protocol mentioned in section 2.2.7.1.

2.2.7.5 Expression of VAO

Cloning of *vaoA* into pET-22b(+): The *vaoA* gene was obtained as pMK-RQ_VAO vector construct. As pET-22b(+) was chosen for expression of VAO on a T7 promoter-based vector, pMK-RQ_VAO and pET-22b(+) were digested with the restriction enzymes *NdeI* and *XhoI* at 37°C for 5 h. Insert and linearized vector were purified and ligated in a vector to insert ratio of 1:3 with T4 DNA ligase (Promega, Madison, WI, USA) and T4 ligation buffer. The ligation was performed at RT for 1 h. After transformation into *E. coli* DH5 α for plasmid amplification, the plasmids were isolated and the sequence of the pET-22b(+)_VAO construct was verified by automated DNA-sequencing (GATC Biotech, Konstanz, Germany).

Expression of VAO: Wild type and variants of VAO were expressed in *E. coli* BL21(DE3) induced by addition of 0.5 mM IPTG (final concentration) to the expression culture and incubation at 20°C and 140 rpm O/N, following the general expression protocol (section 2.2.7.1).

2.2.8 SDS-PAGE analysis

SDS-PAGE was performed for analysis of samples of protein expression and protein purification. Therefore SDS-PAGE gels consisting of a stacking gel with 4% acrylamide and a resolving gel containing 12.5% acrylamide were prepared. As a standard 5 μ l of PageRuler™ Unstained Protein Ladder from Fermentas (now Fisher Scientific, Schwerte, Germany) were applied to the gel.

For whole cell samples taken prior to induction (pl) and after induction (al) as well as for resuspended insoluble fractions (IF; pellets after cell lysis by sonication and after centrifugation), the OD₆₀₀ was measured. The samples were adjusted to an OD₆₀₀ value of 0.25, centrifuged for 1 min at 11000 \times g and the supernatant was discarded.

The pellets were resuspended in 50 μl of SDS sample loading buffer and heated at 95°C for 10 min for protein denaturation. For comparison of protein content of the different samples, equal volumes of 7.5 μl to 15 μl of each of the normalized samples were applied to SDS-PAGE.

For samples of soluble protein fractions (SF; cell lysates after cell lysis by sonication and after centrifugation) and purified protein the protein concentration was determined by the Bradford protein assay¹⁶² (section 2.2.9.1). The samples were diluted with 50 mM Tris-HCl buffer pH 7.5 to a concentration of 2 $\mu\text{g}/\mu\text{l}$, mixed with SDS sample loading buffer to a final concentration of 1 $\mu\text{g}/\mu\text{l}$ and heated at 95°C for 10 min. For comparison of protein content of the different samples, equal amounts of protein of 5 μg to 10 μg (volumes of 5 μl to 10 μl , respectively) of each of the samples were applied to SDS-PAGE. The electrophoresis was run in running buffer for 15 min at 12 mA followed by 1 h at 25 mA for a single gel. After separation the proteins were visualized with a staining solution containing Coomassie[®] Brilliant Blue R-250 dye. Unspecific staining of the gels was removed by washing several times with a destaining solution. Finally, the gels were documented on a Quantum ST 4 system (Vilber Lourmat, Eberhardzell, Germany).

Components of buffers and solutions used for SDS-PAGE are described in detail in Table 7.10 (supplementary material).

2.2.9 Quantification of proteins

2.2.9.1 Total protein concentration

The Bradford protein assay¹⁶² was applied for the determination of the total protein concentration in soluble protein fractions. Bovine serum albumin was used to establish standard curves (concentration range 0-10 $\mu\text{g}/\text{ml}$).

2.2.9.2 Cytochrome P450 monooxygenase concentration

Concentrations of P450 enzymes were determined by the carbon monoxide (CO) differential spectral assay as described previously^{77, 163} using an extinction coefficient of 91 $\text{mM}^{-1} \text{cm}^{-1}$.

2.2.10 Investigation of enzyme activity

2.2.10.1 Investigation of MobA activity

2.2.10.1.1 Spectro-photometrical analysis

Purified MobA wild type enzyme (diluted in a ratio of 1:5 with 50 mM Tris-HCl buffer pH 7.5) and soluble protein fraction (lysate) of MobA wild type and variant V257A were applied in a quartz cuvette for spectro-photometrical measurement (Ultrospec 3000, Pharmacia Biotech, Uppsala, Sweden) within a wavelength range of 300-600 nm. 50 mM Tris-HCl buffer pH 7.5 and lysate of empty vector expression of pET-11a were used as reference and negative control, respectively.

2.2.10.1.2 Determination of enzyme activity via NADPH depletion

MobA enzyme was applied in biotransformation reactions with the substrate 3-hydroxybenzoic acid for determination of activity via NADPH depletion. Reaction mixtures contained 1 mM 3-hydroxybenzoic acid and 0.5 mg/ml or 1 μ M of MobA lysate or purified protein, respectively, in a final volume of 1 ml with 50 mM Tris-HCl buffer pH 7.5. After incubation at RT for 5 min the reactions were started by addition of NADPH at a final concentration of 0.5 mM and the NADPH depletion was measured spectro-photometrically over a period of 40 min at an absorption wavelength of 340 nm. Lysate of pET-11a empty vector expression was used as negative control and a sample containing all components except protein was employed as reference.

2.2.10.2 Investigation of CYP116B3 activity (7-ethoxycoumarin assay)

In order to determine the activity of CYP116B3 wild type enzyme and variants thereof, an assay based on the selective deethylation of 7-ethoxycoumarin yielding 7-hydroxycoumarin was applied. This assay was described previously for CYP116B3^{140, 141} and adapted here. For this reason, 50 mM Tris-HCl buffer pH 7.5 was mixed with 7-ethoxycoumarin and CYP116B3 lysate at a final concentration of 1 mM and 0.5 μ M, respectively, in a final volume of 200 μ l in 96-well plates. After

incubation at RT for 5 min, reactions were started by addition of NADPH at a final concentration of 0.4 mM. The fluorescence intensity of the reaction product 7-hydroxycoumarin was immediately measured with a multiwell plate reader (FLUOstar Omega, BMG LABTECH, Ortenberg, Germany) (excitation at 405 nm, emission at 460 nm). The measurements were carried out at 30°C for 20 cycles with intervals of 90 s and intermediate moderate shaking. Lysate of empty vector expression was used as negative control. 50 mM Tris-HCl buffer pH 7.5 containing 1 mM 7-ethoxycoumarin and 0.4 mM NADPH was employed as reference. All measurements were performed in triplicate. For calculation of the enzyme activity, a series of standard solutions consisting of 10-100 µM 7-hydroxycoumarin in 50 mM Tris-HCl buffer pH 7.5 was measured at 30°C in triplicate. 50 mM Tris-HCl buffer pH 7.5 was used as reference.

2.2.10.3 Investigation of VAO activity

The activity of VAO wild type and variants F454Y and T505S was checked by *in vitro* biotransformation of 4-methylguaiacol and vanillyl alcohol according to the protocol mentioned in section 2.2.7.1. However, the activity concerning the conversion of 4-methylguaiacol was investigated in 50 mM Tris-HCl buffer of both pH 7.5 and pH 10. All measurements were performed in triplicate with lysate of empty vector expression (pET-22b(+)) as negative control.

2.2.11 *In vitro* biotransformations

2.2.11.1 Single-enzyme catalyzed reactions

Standard reaction conditions: Biotransformations were performed using a final volume of 1 ml in 50 mM Tris-HCl buffer pH 7.5, containing the glucose-6-phosphate (G6P) and glucose-6-phosphate dehydrogenase (G6PDH) system (1 mM MgCl₂, 5 mM G6P and 5 U/ml G6PDH) for cofactor regeneration. The substrates 3-methylanisole, 4-methylguaiacol, 3-methoxybenzyl alcohol, vanillyl alcohol and other compounds were added at a final concentration of 0.5 mM from a 25 mM stock solution in dimethyl sulfoxide (DMSO). Enzyme was added at a defined concentration and the reaction was started by addition of 0.2 mM NADPH. Samples were incubated

at 30°C and 180 rpm for 2 h. Lysates of empty vector expression (pET-11a, pET-22b(+), pET-24a(+), or pET-28a(+)) in *E. coli* BL21(DE3) were used as a negative control for the *in vitro* biotransformations, respectively. All experiments were performed in triplicate unless otherwise noted. Concentrations of employed enzymes and deviations from the standard reaction conditions are mentioned below for the respective enzymes.

MobA: Lysate of MobA and empty vector was added at a final lysate protein concentration of 1 mg/ml in the reaction mixture; purified MobA was added at a final concentration of 1 µM. Biotransformations were performed only as single measurements.

CYP116B3: Lysate and purified enzyme of CYP116B3 was added at a final concentration of 0.5 µM, except as noted otherwise. Empty vector lysate was added with an appropriate volume as negative control.

CYP102A1: Lysate of CYP102A1 was added at a final concentration of 1 µM for conversion of 3-methylanisole and 0.5 µM for conversion of other substrates. Empty vector lysate was added with an appropriate volume as negative control.

VAO: Lysate of VAO and empty vector was added at a final total protein concentration of 1 mg/ml in the reaction mixture. No NADPH cofactor and no components of the cofactor regeneration system were employed.

2.2.11.2 Bi-enzymatic one-pot cascade reactions

2.2.11.2.1 Combination of two CYP102A1 variants

One-pot cascade reactions with a combination of the CYP102A1 variants A328L and R47L/Y51F/F87V/A328V were performed as mentioned in section 2.2.11.1, with the exception that each of the two CYP102A1 enzyme variants (lysate) were used at a final concentration of 1 µM. The substrate 3-methylanisole was added at a final concentration of 2 mM from a 100 mM stock solution in DMSO. As a variation, the second enzyme variant was added delayed after 1 h (total reaction time of 2 h).

Moreover, in order to investigate a possible negative influence of reactive oxygen species on the reactions, the addition of catalase (1200 U/ml final concentration) and/or superoxide dismutase (SOD) (100 U/ml final concentration) was investigated for a total reaction time of 2 h. All experiments were performed in triplicate, unless otherwise noted, with lysate of empty vector expression as negative control.

2.2.11.2.2 Combination of CYP102A1 and VAO

One-pot cascade reactions with a combination of the two enzyme variants CYP102A1_A328L and VAO_F454Y were performed as mentioned above (section 2.2.11.2.1), except that the lysate of VAO was added at a final VAO-lysate protein concentration of 1 mg/ml and the co-addition of SOD was not investigated. As a variation, the second enzyme variant was added after 1 h (total reaction time of 2 h) or 4 h (total reaction time of 8 h). All experiments were performed in triplicate, unless otherwise noted, with lysate of empty vector expression (pET-22b(+) and pET-28a(+)) as negative control.

2.2.12 *In vivo* biotransformations

Preparation of the cells: Competent *E. coli* BL21(DE3) cells were transformed with the plasmids pET-28a(+)_CYP102A1_A328L and pET-22b(+)_VAO_F454Y and protein expression was performed according to the protocol mentioned previously (section 2.2.7.1). However, the LB medium was supplemented with both kanamycin (30 $\mu\text{g ml}^{-1}$) and ampicillin (100 $\mu\text{g ml}^{-1}$) and protein expression was induced by addition of 1.0 mM IPTG, 0.1 mM 5-ALA and 0.1 mM FeSO_4 . After incubation at 20°C and 140 rpm for 18 h, cells were harvested by centrifugation (2700 \times g, 20 min, 4°C), washed twice with 50 mM Tris-HCl buffer pH 7.5 and resuspended in the biotransformation medium.

Biotransformations: *In vivo* biotransformations were carried out in 100 ml shake flasks with a volume of 30 ml biotransformation medium consisting of fresh cell suspension (50 g cell wet weight (cww) l^{-1}) in 50 mM Tris-HCl buffer pH 7.5 with 1% (w/v) glycerol, 0.4% (w/v) D-glucose, 30 $\mu\text{g ml}^{-1}$ kanamycin and 100 $\mu\text{g ml}^{-1}$ ampicillin. Biotransformations started by addition of the substrate 3-methylanisole

(from a 200 mM stock solution in dimethyl sulfoxide) at a final concentration of 4 mM. Reactions were run at 30°C and 180 rpm for 24 h. Cells were fed with a glycerol/glucose mixture (1% (w/v) glycerol, 0.4% (w/v) D-glucose) after 12 h. Samples were collected at time points 0, 15 and 30 min and 1, 2, 3, 4, 6, 8, 12 and 24 h after substrate addition and prepared for GC-FID analysis as described subsequently (section 2.2.9). *E. coli* BL21(DE3) cells transformed with empty vectors (pET-22b(+)) and pET-28a(+)) were used as a negative control for the *in vivo* biotransformations. All experiments were performed in triplicate.

2.2.13 Product analysis

2.2.13.1 Sample treatment

General method: Substrate conversion was stopped with 20 µl concentrated HCl. For exact determination of substrate conversion and product distribution, extraction from the aqueous reaction mixture was performed with 0.25 ml ethyl acetate containing the internal standard benzaldehyde (0.1 mM) or catechol (0.5 mM) for gas chromatography (GC)-analysis or high-performance liquid chromatography (HPLC)-analysis, respectively. After vigorous mixing for 2 min and centrifugation at 20200 × g for 5 min, 100 µl of organic phase was transferred to a vial with insert and capped.

Primary CYP102A1 library screening: For primary screening of the focused CYP102A1 library by HPLC, substrate and formed products were extracted from the aqueous reaction mixture by addition of 0.5 g of K₂CO₃ and 0.5 ml of a 40/60 (% v/v) mixture of *n*-hexane/2-propanol, vigorous mixing for 2 min and centrifugation at 20200 × g for 3 min. 100 µl of organic phase was combined with 100 µl of *n*-hexane in a vial with insert and capped.

2.2.13.2 HPLC-analysis

General method: Samples were analyzed by normal phase HPLC on a 1200 Series HPLC from Agilent Technologies (Waldbronn, Germany), equipped with a Luna Silica 5 µm column (particle size: 5 µm, pore size: 100 Å, length: 150 mm, inner diameter: 4.6 mm) from Phenomenex (Aschaffenburg, Germany). The mobile phase consisted

of 2-propanol with 0.2 M formic acid (A), 2-propanol (B) and *n*-hexane (C). Concentration of A was kept constant at 5% throughout the measurement. Concentration of B gradually increased from 0% to 25% after 4 min, was then instantly raised to 45% and kept constant for 8 min, followed by a decrease to 0% within 3 min for regeneration of the starting conditions for the next measurement. Concentration of C complemented the mobile phase solvents to 100%, respectively. The temperature of the column was kept constant at 30°C at a flow rate of 0.8 ml/min. For exact determination of 4-methylguaiacol conversion and product distribution, 15 µl of extracted sample were injected and the substrate and products were detected at a wavelength of 280 nm. To obtain straight-line calibration plots (determination of the internal response factor) a series of standard solutions consisting of 0.01-0.5 mM reference substances in 50 mM Tris-HCl buffer pH 7.5 were extracted with ethyl acetate containing internal standard (0.5 mM catechol) and analyzed by HPLC.

Primary CYP102A1 library screening: For primary screening of the focused CYP102A1 library, 40 µl of extracted sample were injected to the HPLC and the substrate and products were detected with a diode array detector (DAD) at a wavelength of 272 nm.

2.2.13.3 GC-FID-analysis

Samples were analyzed by GC coupled to a flame ionization detector (FID) on a GC-FID GC-2010 from Shimadzu (Kyoto, Japan) equipped with a DB-5 capillary column (length: 30 m, internal diameter: 0.25 mm, film thickness: 0.25 µm) from Agilent Technologies (Waldbronn, Germany) and with hydrogen as carrier gas (flow rate: 0.88 ml min⁻¹, linear velocity: 30 cm s⁻¹). The injector and detector temperatures were set at 200°C and 330°C, respectively. The temperature program of the column oven was as follows: 100°C for 3 min, raised to 190°C at a rate of 10°C min⁻¹, raised to 320°C at a rate of 75°C min⁻¹ and held isotherm for 3 min. Reaction products were identified by the characteristic retention time of analyzed reference substances. For determination of conversion and product distribution, a series of standard solutions consisting of 0.01-0.5 mM reference substances in 50 mM Tris-HCl buffer pH 7.5 were extracted with ethyl acetate containing internal standard (0.1 mM

benzaldehyde) and analyzed by GC-FID to obtain straight-line calibration plots (determination of the internal response factor).

2.2.13.4 GC-MS-analysis

Samples were analyzed by GC coupled to a mass spectrometer (MS) on a GC-2010 with GCMS-QP2010 from Shimadzu (Kyoto, Japan) equipped with a DB-5MS capillary column (length: 30 m, internal diameter: 0.25 mm, film thickness: 0.25 μm) from Agilent Technologies (Waldbronn, Germany) and with helium as carrier gas (flow rate: 4.3 ml min⁻¹, linear velocity: 30 cm s⁻¹). The injector temperature was set at 250°C. Mass spectra were collected using electrospray ionization (70 eV). The temperature program of the column oven was as follows: 80°C for 1 min, raised to 140°C at a rate of 60°C min⁻¹ and held isotherm for 3 min, raised to 160°C at a rate of 20°C min⁻¹ and held isotherm for 5 min, raised to 320°C at a rate of 75°C min⁻¹ and held isotherm for 2 min. Reaction products were identified by the characteristic retention time of analyzed reference substances and by comparison of the respective fragmentation patterns.

2.2.14 Molecular dynamics simulations

Molecular dynamics (MD) simulations of CYP102A1 with 4-methylguaiacol were performed in cooperation with the bioinformatics group of the ITB (University of Stuttgart, Stuttgart, Germany) by Dr. Alexander Seifert.

3 Results

3.1 Enzyme selection, preparation and application in initial biotransformations

The *m*-hydroxybenzoate hydroxylase MobA from *Comamonas testosteroni* GZ39 and the cytochrome P450 monooxygenases CYP116B3 from *Rhodococcus ruber* DSM 44319 and CYP102A1 from *Bacillus megaterium* ATCC 14581 as well as variants thereof were selected as catalysts for the hydroxylation of aromatic compounds (Figure 1.8, introduction).

3.1.1 Preparation of the *m*-hydroxybenzoate hydroxylase MobA

3.1.1.1 Site-directed mutagenesis, expression and purification of MobA

The gene of the *m*-hydroxybenzoate hydroxylase MobA was obtained as a synthetic gene (GeneArt[®] Gene Synthesis, Life Technologies GmbH, Darmstadt, Germany) with a C-terminal His₆-tag, codon-optimized for the heterologous expression in *E. coli*, and integrated in the vector pET-11a.

As in literature the MobA single mutant variant V257A was reported to broaden the substrate range of the enzyme, hydroxylating for example phenol and resorcinol,⁶⁰ this variant was generated by site-directed mutagenesis based on the pET-11a_MobA vector construct. The sequence of the variant was verified by automated DNA-sequencing (GATC Biotech, Konstanz, Germany).

In an initial expression study, the expression of MobA wild type in *E. coli* BL21(DE3) was investigated at 20, 25 and 30°C after induction with 0.1, 0.5 or 1 mM IPTG. According to the SDS-PAGE analysis of the expression samples (Figure 7.2, supplementary material), MobA, with a calculated molecular weight of 71.3 kDa (calculated from the amino acid sequence using the CloneManager (Sci-Ed Software, Cary, NC, USA) software), was detected for all investigated expression conditions with an apparent molecular weight of about 70 kDa. Apparently, induction with 0.5 mM IPTG and incubation at 20°C for 15-18 h was sufficient for a high yield of

Results

expressed MobA protein. Thus, these conditions were applied for subsequent expression of MobA wild type enzyme and variant V257A.

Furthermore, after cell harvest and cell disruption by sonication, the MobA wild type enzyme was purified from the soluble protein fraction by means of immobilized metal affinity chromatography (IMAC), taking advantage of the C-terminal His₆-tag. After elution of the protein with 100 mM imidazole, MobA-containing fractions (identified by the yellow color due to the presence of the FAD cofactor) were pooled, concentrated via ultrafiltration and dialyzed. The total protein concentration of the purified enzyme was determined to be about 17 mg/ml, which corresponds to a yield of 37 mg purified protein per litre cell culture. Samples of the expression of MobA wild type and variant V257A as well as the purified wild type enzyme were analyzed by SDS-PAGE (Figure 3.1).

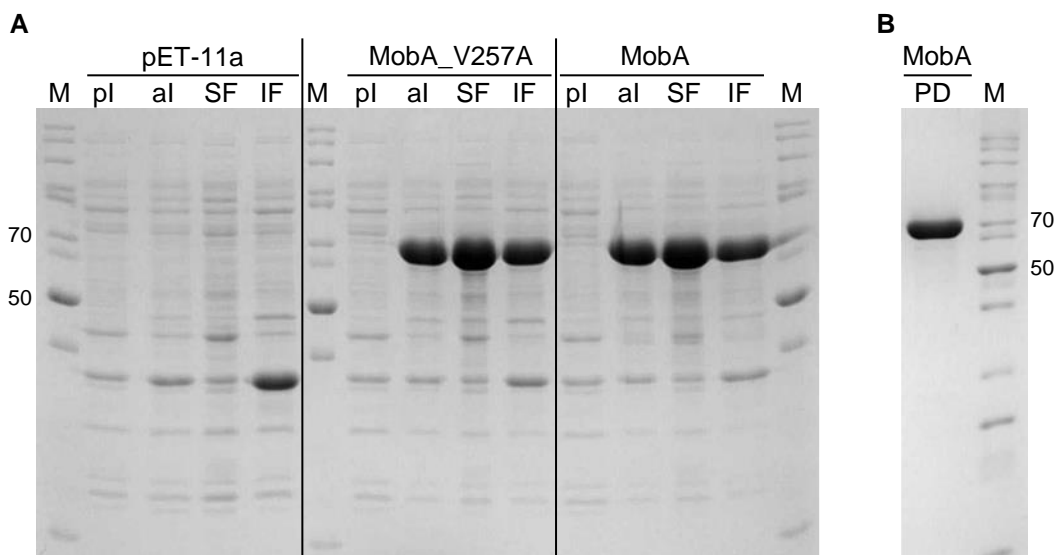


Figure 3.1: SDS-PAGE analysis of the expression (A) and purification (B) of MobA. (A) Expression of MobA wild type and variant V257A was investigated for induction with 0.5 mM IPTG and incubation at 20°C with shaking at 140 rpm O/N. pET-11a empty vector expression was used as a negative control. (B) Purified MobA wild type enzyme after IMAC. M, marker (unstained protein ladder (10-200 kDa) – respective molecular weight is indicated); pl, whole cells prior to induction; al, whole cells after induction and incubation O/N; SF, soluble protein fraction; IF, insoluble fraction after cell lysis; PD, purified and dialyzed protein.

According to the SDS-PAGE analysis, both MobA wild type and variant V257A were expressed and obtained in high yield in the soluble protein fraction after cell lysis. However, distinct amounts of MobA were still detected in the insoluble fraction. The purified MobA wild type enzyme was pure, as only minor protein impurities were faintly visible on the gel (Figure 3.1 (B)).

Results

To enable a calculation of the MobA concentration after purification, it was assumed that the purification yielded completely pure protein. With the molecular weight value of 71.3 kDa, a concentration of 240 μ M MobA was calculated in the purified and dialyzed protein solution.

3.1.1.2 Investigation of MobA activity

3.1.1.2.1 Spectro-photometrical analysis

In order to get a first indication of activity of MobA, both the purified wild type enzyme and the soluble protein fraction of wild type and variant V257A were analyzed spectro-photometrically within a wavelength range of 300-600 nm (Figure 3.2).

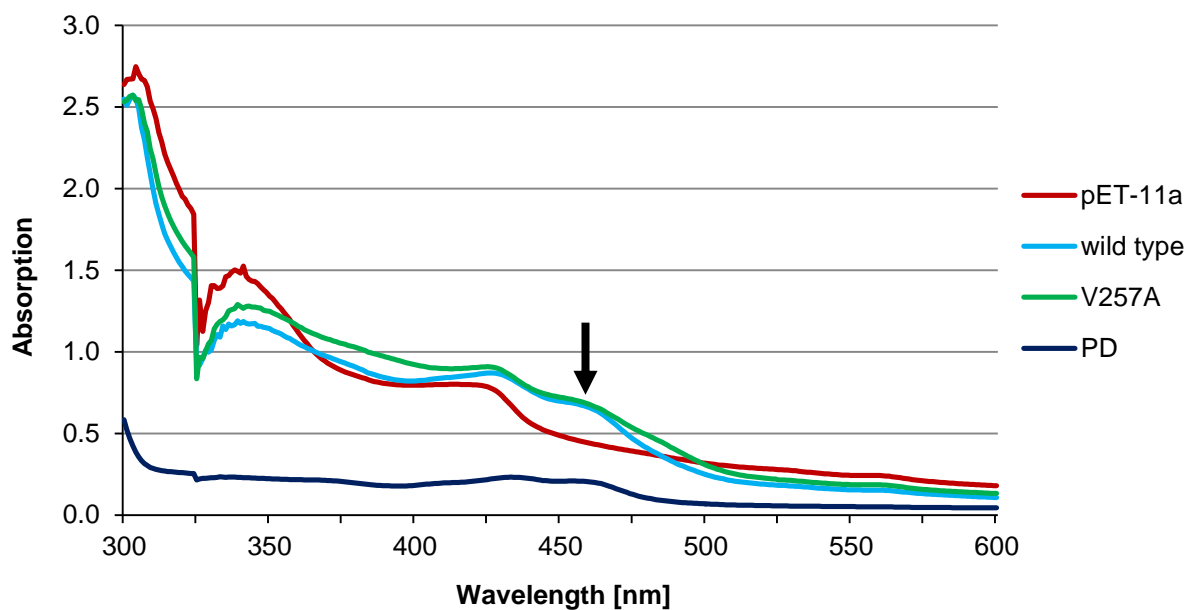


Figure 3.2: Spectro-photometrical analysis of MobA. 50 mM Tris-HCl buffer pH 7.5 was used as reference. Lysate of empty pET-11a vector expression (red) was applied as a negative control. Soluble protein fraction of MobA wild type (light blue) and variant V257A (green) were used in concentrated form. Purified and dialyzed MobA wild type enzyme (PD, dark blue) was measured in a 1:5 dilution with 50 mM Tris-HCl buffer pH 7.5. The black arrow indicates a specific peak, present only in the samples with MobA.

All samples, except the negative control (pET-11a) showed a specific peak at a wavelength of about 455 to 460 nm. As flavin has an absorption maximum at 455 nm,⁵⁸ this is a first indication for the presence of active MobA enzyme.

3.1.1.2.2 Determination of enzyme activity via NADPH depletion

To further investigate the activity of MobA, biotransformation reactions with concomitant detection of NADPH depletion at a wavelength of 340 nm were set up. Therefore 3-hydroxybenzoate was employed as a substrate, due to the fact that *m*-hydroxybenzoate is the preferred substrate for MobA.⁶⁰ As a result, protocatechuate is produced under consumption of molecular oxygen and cofactor NADPH (Figure 3.3).

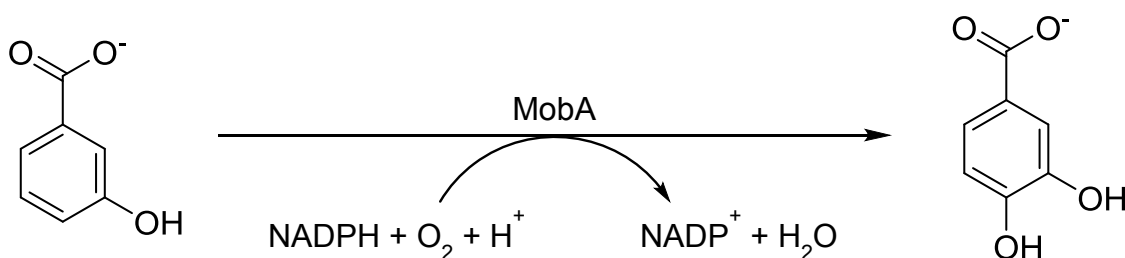


Figure 3.3: MobA-catalyzed conversion of *m*-hydroxybenzoate to protocatechuate. Graphic adapted from Chang and Zylstra 2008.⁶⁰

According to the spectro-photometrical analysis of the NADPH depletion with MobA (Figure 3.4), both lysate of the variant V257A and lysate as well as purified MobA wild type displayed activity towards the model substrate. Though the measurement was started immediately after addition of NADPH to all of the reaction mixtures, the NADPH concentration in the samples with lysate of wild type and V257A enzyme was already at the beginning of the detection only half of the value of the negative control. Only for the purified MobA, the strong initial decline in NADPH concentration could be detected. In the negative control (lysate of pET-11a empty vector expression) only a minor and unspecific depletion of NADPH was noticed. In all reactions with MobA enzyme, NADPH was consumed completely after about 32 min, though the absorption value did not reach a value of zero. Hence, the wild type enzyme and the variant V257A were definitely active in the soluble protein fraction. Further, the wild type enzyme was active after purification and dialysis.

Results

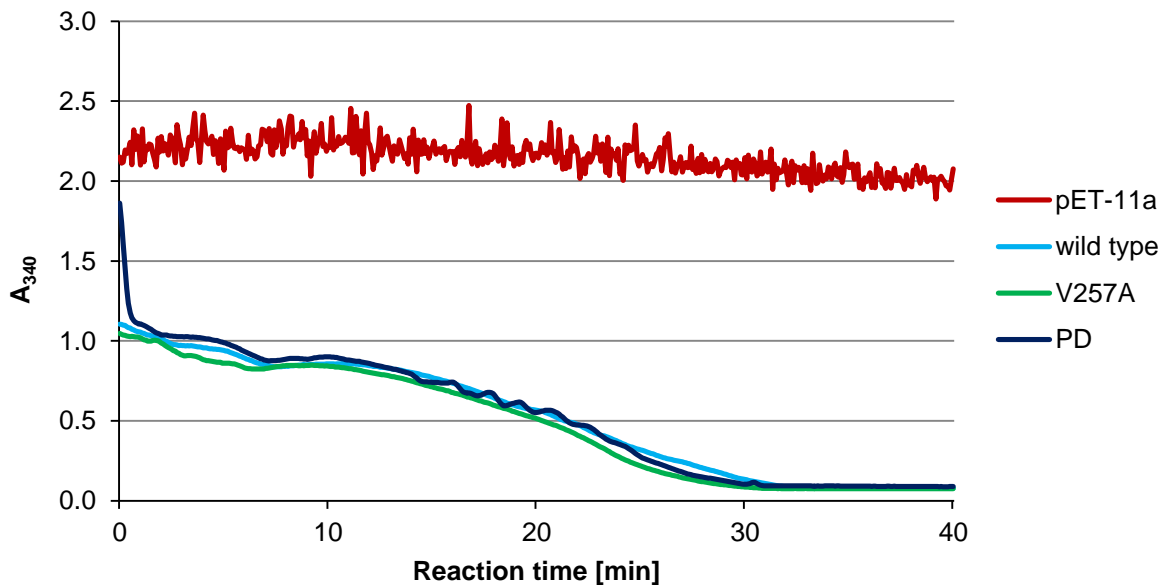


Figure 3.4: Spectro-photometrical analysis of MobA activity via NADPH depletion detection. A sample without protein was used as reference. Lysate of empty pET-11a vector expression (red) was applied as a negative control. Soluble protein fraction of MobA wild type (light blue) and variant V257A (green) was employed at a final total protein concentration of 0.5 mg ml^{-1} . Purified and dialyzed MobA wild type enzyme (PD, dark blue) were assayed at a final concentration of $1 \text{ }\mu\text{M}$.

Summarizing these preliminary results, based on a codon-optimized gene for the *m*-hydroxybenzoate hydroxylase MobA, the wild type enzyme could be successfully expressed and purified in an active form. Furthermore, a single mutant variant V257A was generated by site-directed mutagenesis and expressed in *E. coli* with proven activity.

3.1.2 Preparation of the cytochrome P450 monooxygenase CYP116B3

3.1.2.1 Expression and purification of CYP116B3

The gene of the cytochrome P450 monooxygenase CYP116B3, integrated into the vector pET-24a(+), was obtained as a synthetic gene (GeneArt[®] Gene Synthesis, Life Technologies GmbH, Darmstadt, Germany) with a N-terminal His₆-tag, codon-optimized for the heterologous expression in *E. coli*.

Initial investigations on the expression of CYP116B3 were performed in *E. coli* BL21(DE3). After induction with various concentrations of IPTG (0.1 mM, 0.5 mM and 1 mM) the cell cultures were incubated at different temperatures (20°C, 25°C and 30°C) and samples (whole cells) for SDS-PAGE analysis were taken at several time

Results

points. According to the results of the SDS-PAGE analysis (Figure 7.3, supplementary material), expression of CYP116B3 with a calculated molecular weight of 86.1 kDa (calculated from the amino acid sequence using the CloneManager (Sci-Ed Software, Cary, NC, USA) software) was observed for all investigated expression conditions. The apparent molecular weight of the His₆-tagged CYP116B3 was detected to be around 85 kDa. Induction with 0.1 mM IPTG and incubation at 20°C for 15-18 h was sufficient for a reasonable expression of CYP116B3. Hence, these conditions were applied for subsequent CYP116B3 expressions.

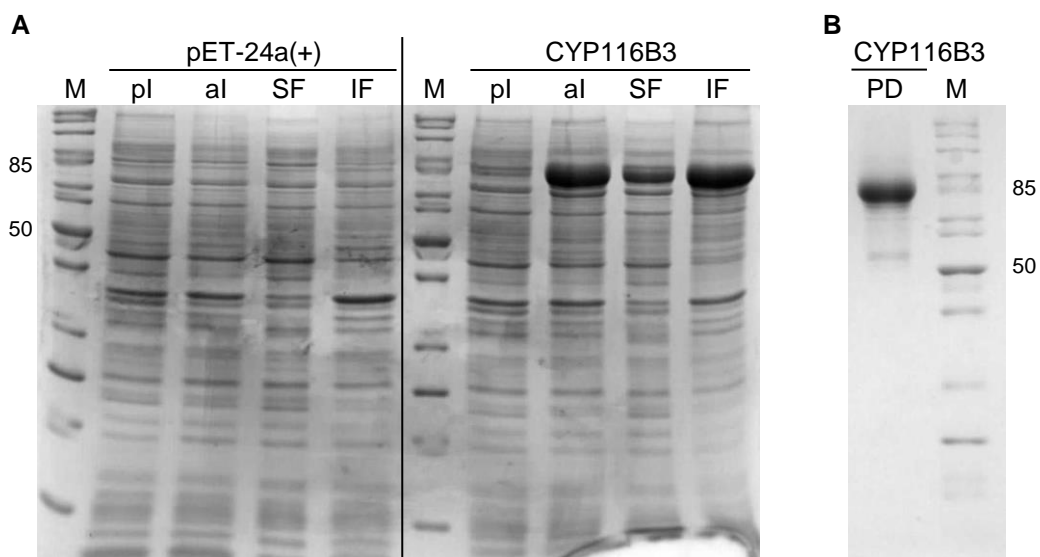


Figure 3.5: SDS-PAGE analysis of the expression (A) and purification (B) of CYP116B3. (A) Expression of CYP116B3 was investigated for induction with 0.1 mM IPTG and incubation at 20°C with shaking at 140 rpm O/N. pET-24a(+) empty vector expression was used as a negative control. (B) Purified CYP116B3 after IMAC. M, marker (unstained protein ladder (10-200 kDa) – respective molecular weight is indicated); pl, whole cells prior to induction; al, whole cells after induction and incubation O/N; SF, soluble protein fraction; IF, insoluble fraction after cell lysis; PD, purified and dialyzed protein.

In the course of the IMAC-purification of CYP116B3 from the soluble protein fraction of harvested and disrupted cells, the protein was eluted with 50 mM imidazole. CYP116B3-containing fractions (identified by the red color due to the presence of the heme) were pooled, concentrated via ultrafiltration and dialyzed. A more detailed SDS-PAGE analysis of the CYP116B3 expression and purification revealed that most of the CYP116B3 was found in the insoluble fraction after cell lysis (Figure 3.5 A). However, the low amount of target enzyme present in the soluble protein fraction was successfully purified, though there was still some contamination with other proteins

Results

(molecular weight of 50 to 85 kDa) after purification (Figure 3.5 B). The concentration of purified CYP116B3 was 9.9 μM , as determined by the CO differential spectral assay.

As the His-tag might influence the protein's folding, several His-tag variants of CYP116B3 were cloned by PCR (oligonucleotides are listed in Table 7.3, supplementary material), in order to increase the amount of CYP116B3 in the soluble protein fraction after cell lysis. Correct protein folding is assumed to lead to reduced formation of inclusion bodies and consequently to improved accumulation of the enzyme in the soluble fraction. In analogy to the pET-28a(+) vector, where a MetGlySerSer(MGSS)-sequence is attached prior to the N-terminal His-tag, the variant CYP116B3_Var1 was constructed with three additional amino acids (Gly, Ser, Ser) prior to its N-terminal His₆-tag. Thus, the new N-terminal sequence was MGSS-His₆. N-terminal amino acids are known to influence a protein's stability in bacteria, as determined by the N-end rule.¹⁶⁴ A second His-tag variant, CYP116B3_Var2, was constructed with a His₆-tag at the C-terminus instead of the N-terminus. A third variant, CYP116B3_Var3, contained no His-tag at all. The sequences of all His-tag variants were verified by automated DNA-sequencing (GATC Biotech, Konstanz, Germany).

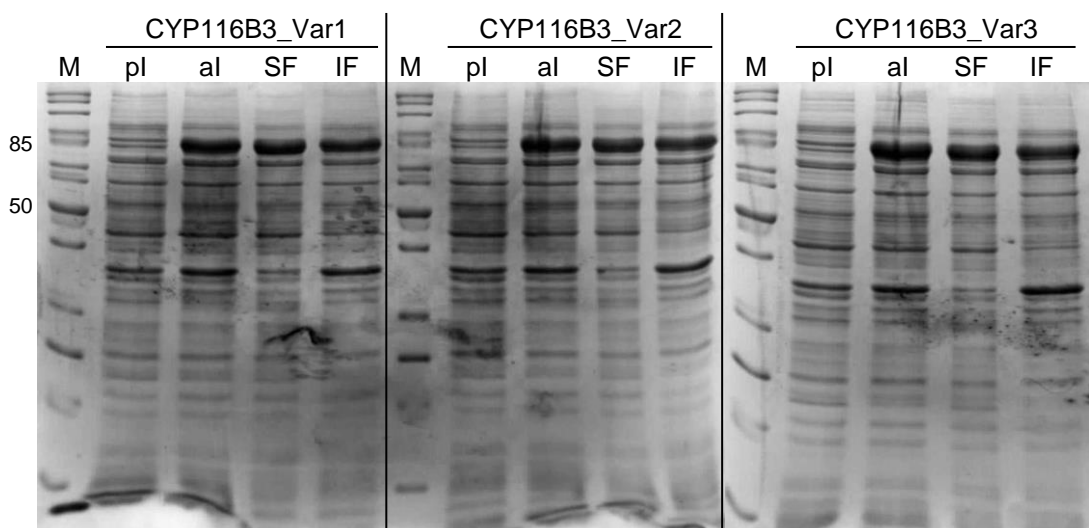


Figure 3.6: SDS-PAGE analysis of the expression of His-tag variants of CYP116B3. CYP116B3_Var1 has a N-terminal MetGlySerSer (MGSS) sequence prior to the His₆-tag. CYP116B3_Var2 has a C-terminal His₆-tag. CYP116B3_Var3 has no His-tag. Expression was investigated for induction with 0.1 mM IPTG and incubation at 20°C with shaking at 140 rpm O/N. M, marker (unstained protein ladder (10-200 kDa) – respective molecular weight is indicated); pl, whole cells prior to induction; al, whole cells after induction and incubation O/N; SF, soluble protein fraction; IF, insoluble fraction after cell lysis.

Results

From the SDS-PAGE analysis of the CYP116B3 His-tag variants expression (Figure 3.6), one might assume a slightly enhanced amount of CYP116B3 in the soluble protein fraction of all three His-tag variants (Var1, Var2, Var3), compared to the initial CYP116B3 variant (Figure 3.5 A). However, still considerable amounts of insoluble proteins of CYP116B3 could be observed. After the specific determination of the CYP116B3 concentration in the soluble protein fraction by CO differential spectral assay, a calculation of the expression of soluble CYP116B3 was performed (Table 3.1). According to these results, a higher amount of CYP116B3 was found in the soluble protein fraction for all newly generated His-tag variants compared to the original CYP116B3 enzyme (N-terminal His₆-tag). The highest amount of soluble CYP116B3 was detected for expression of CYP116B3_Var1 (N-terminal MGSS-His₆-sequence). With a value of 1.42 mg CYP116B3 per g cell wet weight (cww), expression of CYP116B3_Var1 yielded almost double the amount of soluble CYP116B3 compared to the initial CYP116B3 expression, though the total protein concentration was lower.

Table 3.1: Investigation of CYP116B3 His-tag variant expression. The total protein concentration in the soluble protein fraction was determined using the Bradford protein assay.¹⁶² The CO differential spectral assay was applied for determination of the CYP116B3 concentration in the soluble protein fraction as described previously.^{77, 163}

CYP116B3 His-tag variant	c(total protein) [mg (l cell culture)⁻¹]	c(CYP116B3) [mg (l cell culture)⁻¹]	c(CYP116B3) [mg (g_{cww})⁻¹]
CYP116B3 (N-term His ₆ -tag)	229.1	2.62	0.79
CYP116B3_Var1 (N-term MGSS-His ₆ -sequence)	195.9	5.85	1.42
CYP116B3_Var2 (C-term His ₆ -tag)	248.8	4.03	1.04
CYP116B3_Var3 (no His-tag)	240.7	4.29	1.17

3.1.2.2 Determination of activity of CYP116B3

To further elucidate which His-tag variant was most suitable for subsequent expression of CYP116B3, the activity of all His-tag variants, including the initial CYP116B3 in lysate and as purified enzyme, was tested, taking advantage of the known dealkylation activity of CYP116B3 towards 7-ethoxycoumarin. The accompanying formation of 7-hydroxycoumarin was detected by fluorescence spectroscopy (excitation at 405 nm, emission at 460 nm). Then the product formation rate, as a measure of activity, was calculated based on an extinction coefficient derived from a series of 7-hydroxycoumarin standard solutions.

Table 3.2: Determination of CYP116B3 His-tag variant activity. The dealkylation of 7-ethoxycoumarin to 7-hydroxycoumarin was detected by fluorescence spectroscopy. Reaction mixture without protein was used as reference and reaction mixture with lysate of empty vector expression as negative control.

CYP116B3 His-tag variant	Product formation rate [$\mu\text{mol} (\mu\text{mol CYP116B3})^{-1} \text{min}^{-1}$]
CYP116B3 (N-term His ₆ -tag)	0.32
CYP116B3_Var1 (N-term MGSS-His ₆ -sequence)	0.43
CYP116B3_Var2 (C-term His ₆ -tag)	-
CYP116B3_Var3 (no His-tag)	0.41

Reactions were performed in triplicate with a standard deviation of less than 10%.

Highest product formation rate was obtained for the His-tag variant CYP116B3_Var1 with a value of $0.43 \mu\text{mol} (\mu\text{mol CYP116B3})^{-1} \text{min}^{-1}$. This equals an about 34% higher activity compared to the activity of the initial CYP116B3 enzyme. For CYP116B3_Var2 no activity could be determined as no 7-hydroxycoumarin formation was detectable.

Subsequently, CYP116B3_Var1 was purified as mentioned previously (section 2.2.7.3, materials and methods) and the activity was determined to be similar to the activity of the enzyme in lysate (product formation rate of $0.42 \mu\text{mol} (\mu\text{mol CYP116B3})^{-1} \text{min}^{-1}$, data not shown in Table 3.2). Protein purification had no negative effect on activity of CYP116B3.

Results

The His-tag variant CYP116B3_Var1, containing an N-terminal MGSS-His₆-sequence, showed highest expression, was successfully purified and displayed highest activity towards the model substrate 7-ethoxycoumarin. Therefore, this His-tag variant of CYP116B3 was chosen for all subsequent experiments. The application of CYP116B3 for the conversion of aromatic substrates, aiming at the synthesis of vanillin, is described in section 3.1.4.

3.1.3 Preparation of the cytochrome P450 monooxygenase CYP102A1

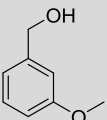
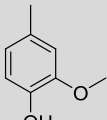
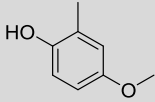
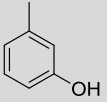
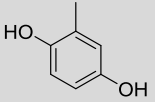
Both the wild type and variants of the cytochrome P450 monooxygenase CYP102A1 from *B. megaterium* were shown to convert various aromatic substrates in previous studies.^{121, 165} Moreover, a minimal and highly enriched CYP102A1 mutant library was designed by rational protein design and successfully applied for the conversion of a variety of terpene substrates¹³² and cyclic and acyclic alkanes.¹³³ This focused CYP102A1 library, which consists of the wild type and 24 single and double mutant variants (combination of the five hydrophobic amino acids Ala, Leu, Ile, Phe and Val in the two positions 87 and 328), was chosen as a promising set of CYP102A1 candidates for the synthesis of vanillin from aromatic compounds in this work. All variants of the focused library, including the wild type, were kindly provided by Dr. Alexander Seifert (Institute of Technical Biochemistry (ITB), University of Stuttgart, Stuttgart, Germany) as indicated in Table 7.2 (supplementary material) and expressed in *E. coli* BL21(DE3) according to an established protocol.¹⁶¹ After cell lysis, the soluble protein fractions were applied for *in vitro* biotransformations of aromatic substrates, particularly regarding the synthesis of vanillin (see section 3.1.4).

3.1.4 Initial biotransformations with selected enzymes and enzyme variants

The *m*-hydroxybenzoate hydroxylase MobA (wild type and variant V257A), the P450 CYP116B3 (wild type) and the P450 CYP102A1 (focused mutant library consisting of wild type and 24 single and double mutant variants) were prepared as mentioned above (section 3.1.1, 3.1.2 and 3.1.3, respectively) and applied in initial *in vitro* biotransformation reactions of aromatic substrates.

Results

Table 3.3: Biotransformation results for the *in vitro* conversion of 3-methylanisole with MobA, CYP116B3 and selected single and double mutant variants of the focused CYP102A1 library.

Enzyme (wild type) variant	Product formation [% of initial substrate]					Total conversion [%] ¹
	 2	 3	 6	 9	 10	
MobA	-	-	-	-	-	-
V257A	-	-	-	-	-	-
CYP116B3	<1	-	<1	1.4 ± 0.1	-	2.1 ± 0.0
CYP102A1	-	-	<1	-	-	<1
A328L	<1	2.7 ± 0.2	26.6 ± 2.5	<1	-	31.0 ± 2.8
F87V	<1	<1	12.5 ± 0.2	<1	-	14.9 ± 0.3
F87V/A328F	2.2 ± 0.2	1.2 ± 0.1	16.5 ± 1.3	7.1 ± 0.4	-	26.9 ± 2.0
F87V/A328I	6.2 ± 0.5	<1	5.6 ± 0.3	4.9 ± 0.3	-	16.9 ± 1.1
F87V/A328L	9.1 ± 0.2	1.6 ± 0.1	23.2 ± 0.4	12.0 ± 0.8	13.5 ± 0.8	59.3 ± 1.8
F87V/A328V	4.9 ± 0.5	<1	23.1 ± 0.9	3.0 ± 0.5	2.1 ± 0.1	34.1 ± 2.0

-, not detected or <0.05%; **2**, 3-methoxybenzyl alcohol; **3**, 4-methylguaiacol; **6**, 4-methoxy-2-methylphenol; **9**, *m*-cresol; **10**, methylhydroquinone. Reactions were performed in triplicate with soluble protein fractions and run at 30°C and 180 rpm for 2 h. MobA was employed at a final lysate protein concentration of 1 mg/ml. CYP116B3 and CYP102A1 were used at a final P450 concentration of 0.5 μM and 1 μM, respectively. Samples were analyzed by GC-FID. Negative controls showed no conversion. ¹Total conversion was calculated from the determined product formations and was not equivalent to the substrate consumption due to the volatility of 3-methylanisole.

In addition to 3-methylanisole (**1**), the primary substrate of the proposed vanillin synthesis pathway (Figure 1.8, introduction), the intermediate products 3-methoxybenzyl alcohol (**2**) and 4-methylguaiacol (**3**) were employed as substrates. Moreover, the vanillin (**5**) precursor vanillyl alcohol (**4**) was used as a substrate to investigate a possible overoxidation reaction.

Results of the 3-methylanisole conversion with soluble protein fractions are shown in Table 3.3. Similarly to MobA and CYP116B3, none of the above mentioned intermediate substrates was converted by any of the variants of the focused CYP102A1 minimal mutant library. However, the wild type of CYP102A1 and 22 of the variants of the focused library were identified to be active with 3-methylanisole in a primary screening (Table 7.13, supplementary material). Only two variants were completely inactive and six variants, including the wild type, exhibited less than 5%

Results

conversion. A 3-methylanisole conversion of more than 25% was detected for six variants, the single mutant variants F87V and A328L and the four double mutant variants F87V/A328F/I/L/V. Thus, five of the six variants with a conversion higher 25%, had a phenylalanine to valine substitution in amino acid position 87. Consequently, these most active six variants and the wild type were studied in more detail for exact determination of the product formation and the total conversion (Table 3.3). Compared to CYP116B3, the wild type of CYP102A1 revealed a total conversion of lower than 1%, with 4-methoxy-2-methylphenol as exclusive product. This is less than half the conversion compared to the CYP116B3 wild type, though twice as much enzyme was used. However, all of the six chosen single and double mutant variants of CYP102A1 displayed a much higher total conversion compared to the CYP102A1 wild type. But the values were lower in comparison to the conversion values of the primary screening (Table 7.13, supplementary material). This was due to the fact that the substrate 3-methylanisole is a volatile compound, which, in contrast to the primary screening by HPLC, has been considered for detailed investigation by GC-FID (Table 3.3).

Apparently, variants with high activity demonstrated only poor selectivity. Hence, not only the anticipated intermediate products 3-methoxybenzyl alcohol (**2**) and 4-methylguaiacol (**3**), but also the byproducts 4-methoxy-2-methylphenol (**6**), *m*-cresol (**9**) and methylhydroquinone (**10**) were generated (Figure 3.7). In addition, product formation of 2-methoxy-6-methylphenol (**8**) was detected with less than 1% for the CYP102A1 variants F87V and F87V/A328F (data not shown in Table 3.3). The formation of 3-methoxy-5-methylphenol (**7**) was not discovered at all.

Five of the selected CYP102A1 variants revealed 4-methoxy-2-methylphenol as main product. Only variant F87V/A328I converted the substrate to a mixture of 3-methoxybenzyl alcohol, 4-methoxy-2-methylphenol and *m*-cresol in comparable amounts. Even though for the single mutant variants F87V and A328L a total conversion of 15% and 31%, respectively, was detected, the corresponding double mutant variant F87V/A328L exhibited the highest determined activity of about 59% total conversion. Furthermore, product formation of 3-methoxybenzyl alcohol and *m*-cresol reached a maximum of about 9% and 12%, respectively, with this variant, whereas the single mutant variant A328L yielded overall highest 4-methylguaiacol formation of 2.7%. Only for the variants F87V/A328L and F87V/A328V overoxidation with methylhydroquinone as product could be observed. Thereby, methyl-

Results

hydroquinone was identified to be synthesized via the aromatic hydroxylation of *m*-cresol. This was elucidated in an investigation of the *in vitro* conversion (standard reaction conditions as described in section 2.2.11.1, materials and methods) of both 4-methoxy-2-methylphenol and *m*-cresol. Whilst 4-methoxy-2-methylphenol was not converted at all, *m*-cresol conversion resulted in methylhydroquinone formation as exclusive product (data not shown).

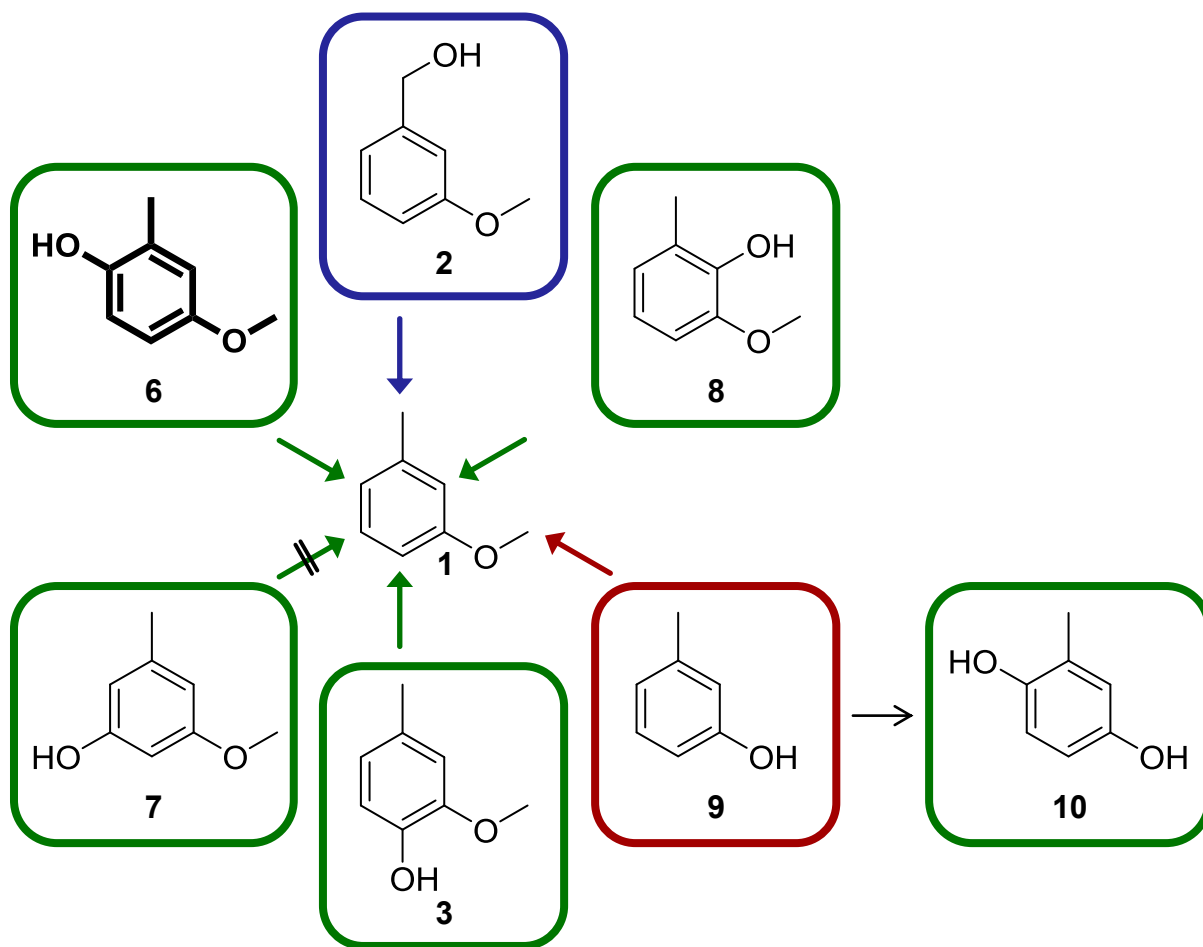


Figure 3.7: Products of the conversion of 3-methylanisole (1) with selected CYP102A1 variants. Three types of reaction were possible (indicated with colored arrows): benzylic hydroxylation (blue), aromatic hydroxylation (green) and O-demethylation (red). The respective reaction product is shown in colored frame. Formation of 3-methoxy-5-methylphenol (7) was not detected. Main product 4-methoxy-2-methylphenol (6) is highlighted in bold. 2, 3-methoxybenzyl alcohol; 3, 4-methylguaiacol; 8, 2-methoxy-6-methylphenol; 9, *m*-cresol; 10, methylhydroquinone.

In summary, these first biotransformations revealed that neither MobA (wild type and variant V257A), nor CYP116B3 or any of the variants of the focused CYP102A1 library displayed activity towards the intermediate substrates 3-methoxybenzyl alcohol, 4-methylguaiacol or vanillyl alcohol. However, in contrast to MobA, both CYP116B3 and CYP102A1 wild type showed marginal conversion of

3-methylanisole. In a screening of the focused minimal mutant library of CYP102A1, several variants were identified with much higher activity compared to the wild type. In addition to the formation of 4-methylguaiacol (**3**), 4-methoxy-2-methylphenol (**6**) and 2-methoxy-6-methylphenol (**8**) by aromatic hydroxylation, 3-methoxybenzyl alcohol was produced by benzylic hydroxylation. Moreover, *O*-demethylation of 3-methylanisole (**1**) generated *m*-cresol (**9**), which was further converted to methylhydroquinone (**10**) by aromatic hydroxylation (Figure 3.7). Thus, three different types of reaction were catalyzed by variants of CYP102A1, demonstrating the catalytic flexibility of this large enzyme family.

3.2 Expansion of the spectrum of enzymes and variants for further biotransformations

As neither the wild type nor the single mutant variant V257A of the *m*-hydroxybenzoate hydroxylase MobA was shown to convert any of the substrates into vanillin or a precursor thereof, a focus was set on CYP116B3 and CYP102A1 in subsequent investigations. In order to realize the vanillin synthesis according to the proposed reaction pathway (Figure 1.8, introduction), both CYP116B3 and CYP102A1 had to be improved towards the conversion of 3-methylanisole and especially towards the intermediate compounds 3-methoxybenzyl alcohol and 4-methylguaiacol. For this purpose, rational protein design was applied. Moreover, for the final step of the intended cascade reaction, the synthesis of vanillin from vanillyl alcohol, a vanillyl alcohol oxidase was tested.

3.2.1 Generation of a focused mutant library of CYP116B3

3.2.1.1 Homology modelling of CYP116B3

The generation of a focused mutant library is an efficient method to expand a protein's substrate spectrum and enhance its activity and/or selectivity, as previously shown for CYP102A1 with the substrate 3-methylanisole (section 3.1.4). Compared to the wild type, which displayed only marginal substrate conversion, many of the variants of the focused mutant library were identified with considerably increased

Results

activity. As the wild type activity of CYP116B3 towards 3-methylanisole was already higher compared to the CYP102A1 wild type, the generation of a focused library of CYP116B3 mutants was assumed to be a promising method to increase the enzymes' activity and possibly expand its substrate scope. However, in contrast to CYP102A1, there is no crystal structure available for CYP116B3. Thus, a homology model of CYP116B3 had to be created. This was done in cooperation with Dr. Łukasz Gricman from the bioinformatics group of the ITB (University of Stuttgart, Stuttgart, Germany). In a first step, the amino acid sequence of the CYP116B3 wild type was employed for a BLAST¹⁵⁴ search against the NCBI "Protein Data Bank proteins (pdb)" database. From the resulting list of proteins, the sequence with the accession number "3CV9_A" was chosen for further processing. This is the sequence of a P450 monooxygenase domain with a low percentage of gaps (5%) in the displayed alignment with the CYP116B3 sequence. In addition, it showed relatively high "Max score" and "Ident" values compared to the other listed protein sequences. The monooxygenase domain sequence of CYP116B3 was aligned to the 3CV9_A sequence by using a standard numbering scheme for class I P450 sequences.¹⁵⁵ According to this standard numbering scheme, the alignment was done based on structural information instead of sequence information. Finally, a homology model of CYP116B3 (Figure 3.8) was generated with SWISS-MODEL,¹⁵⁶⁻¹⁶⁰ based on the alignment result.

Results

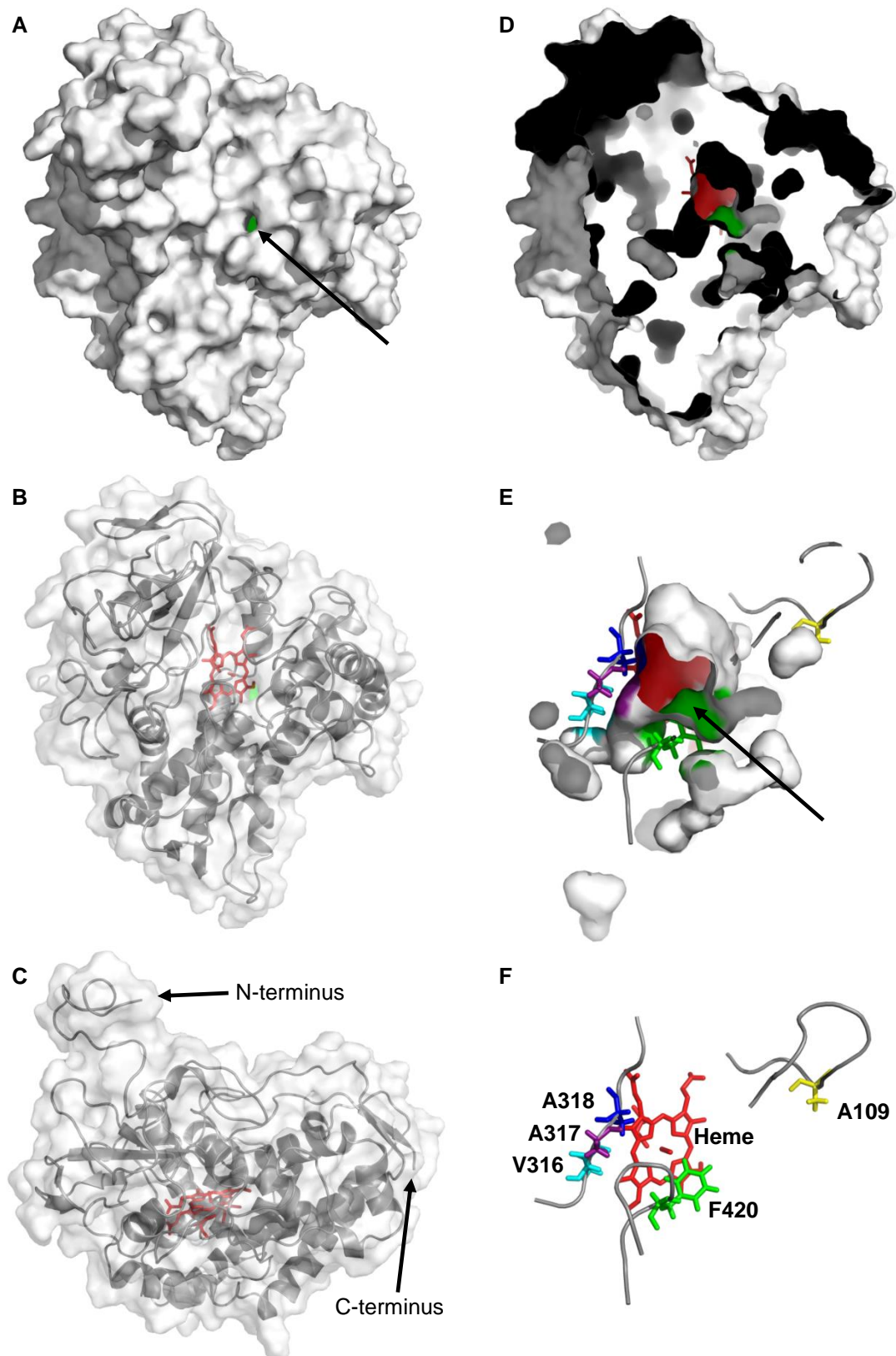


Figure 3.8: Homology model of the monooxygenase domain of CYP116B3. **A** Surface view. Substrate access is indicated by an arrow. **B** Secondary structure view. **C** Rotated view of B to visualize the location of N- and C-terminus of the protein. **D** Scrolled in view into A. **E** Visualization of cavities and pockets in the protein. View of the opened active site cavity. Substrate access is indicated by an arrow. **F** Heme with selected residues for site-directed mutagenesis. **A-F** Heme is colored red. Residues A109, V316, A317, A318 and F420 and corresponding surfaces are colored yellow, cyan, purple, blue and green, respectively. The images were generated by PyMOL.¹⁶⁶

3.2.1.2 Selection of mutation sites and generation of CYP116B3 variants

All subsequently mentioned amino acid positions of CYP116B3 are positions with a numbering referred to the wild type sequence, though the construct CYP116B3_Var1 with an N-terminal MGSS-His₆-sequence was used as a basis for the generation of the focused CYP116B3 mutant library.

Mutation sites for the generation of the focused mutant library of CYP116B3 were selected based on literature as well as sequence and structural information of CYP116B3:

- a) According to the homology model in a previous study, the alanine at position 109 was reported to be located in a small region at the ceiling of the substrate binding pocket, positioned directly above the heme group. Replacement of alanine by phenylalanine resulted in higher activity towards a variety of aromatic substrates.¹⁴¹ Hence, this position was chosen for the generation of a focused CYP116B3 mutant library.
- b) By systematic analysis of numerous crystal structures and a multitude of sequences of P450s, in 98.4% of all sequences the 5th amino acid after the highly conserved ExxR motive was predicted to be a preferentially hydrophobic residue, located close to the heme group and pointing towards the heme center.¹³¹ By analyzing the amino acid sequence of CYP116B3, five ExxR motives were found. Three of those were present in the monooxygenase domain. By alignment of sequences of the CYP116B family using the Cytochrome P450 Engineering Database (CYPED),¹⁶⁷ only two of the five ExxR motives of CYP116B3 were identified to be positioned in active site areas. From these two motives, only one was highly conserved throughout all of the aligned sequences. This was the motive ECLR (amino acid position 308-311). The fifth amino acid after this highly conserved ExxR motive was a valine at position 316 in CYP116B3. Therefore, this position was chosen for rational protein design, in addition to the previously selected position 109.
- c) Based on the structure of the created homology model of CYP116B3, the two amino acid positions 317 and 318, which were both occupied by an alanine in the wild type sequence, were selected for mutagenesis due to their localization at the active site, adjacent to V316. Moreover, the phenylalanine at position

Results

F420, positioned at the entrance to the active site, opposite to V316, A317 and A318, was chosen as further residue.

Though, according to the homology model (Figure 3.8), the residue A109 is shown not to be in direct proximity to the heme (Figure 3.8 E, F), all of the selected five amino acid positions are promising candidates for the generation of a focused library containing variants with enhanced activity and/or selectivity, as they are located on loop regions and thus are expected to be quite flexible.

All variants of the focused CYP116B3 mutant library were generated by site-directed mutagenesis based on the CYP116B3 His-tag variant 1 (CYP116B3_Var1, N-terminal MGSS-His₆-sequence), using the respective oligonucleotides (Table 7.4, supplementary material). CYP116B3_Var1 was chosen due to previous results concerning expression and activity of different CYP116B3 His-tag variants (section 3.1.2). In analogy to the focused CYP102A1 library, the amino acids at the five selected mutation sites in CYP116B3 were replaced by the hydrophobic amino acids alanine (Ala, A), phenylalanine (Phe, F), isoleucine (Ile, I), leucine (Leu, L) and valine (Val, V), respectively. Thus, a CYP116B3 library containing the wild type and 20 single mutant variants was obtained.

3.2.1.3 CYP116B3 variants expression and determination of enzyme activity

The CYP116B3 variants were expressed in *E. coli* BL21(DE3) following the standard expression conditions mentioned in section 2.2.7.3. According to SDS-PAGE analysis (Figure 7.4, supplementary material), all variants were expressed successfully and a reasonable amount of CYP116B3 was recovered in the soluble protein fraction, with some variance. The CYP116B3 concentration of the variants was determined by CO differential spectral analysis (Figure 3.9, Table 3.4).

Results

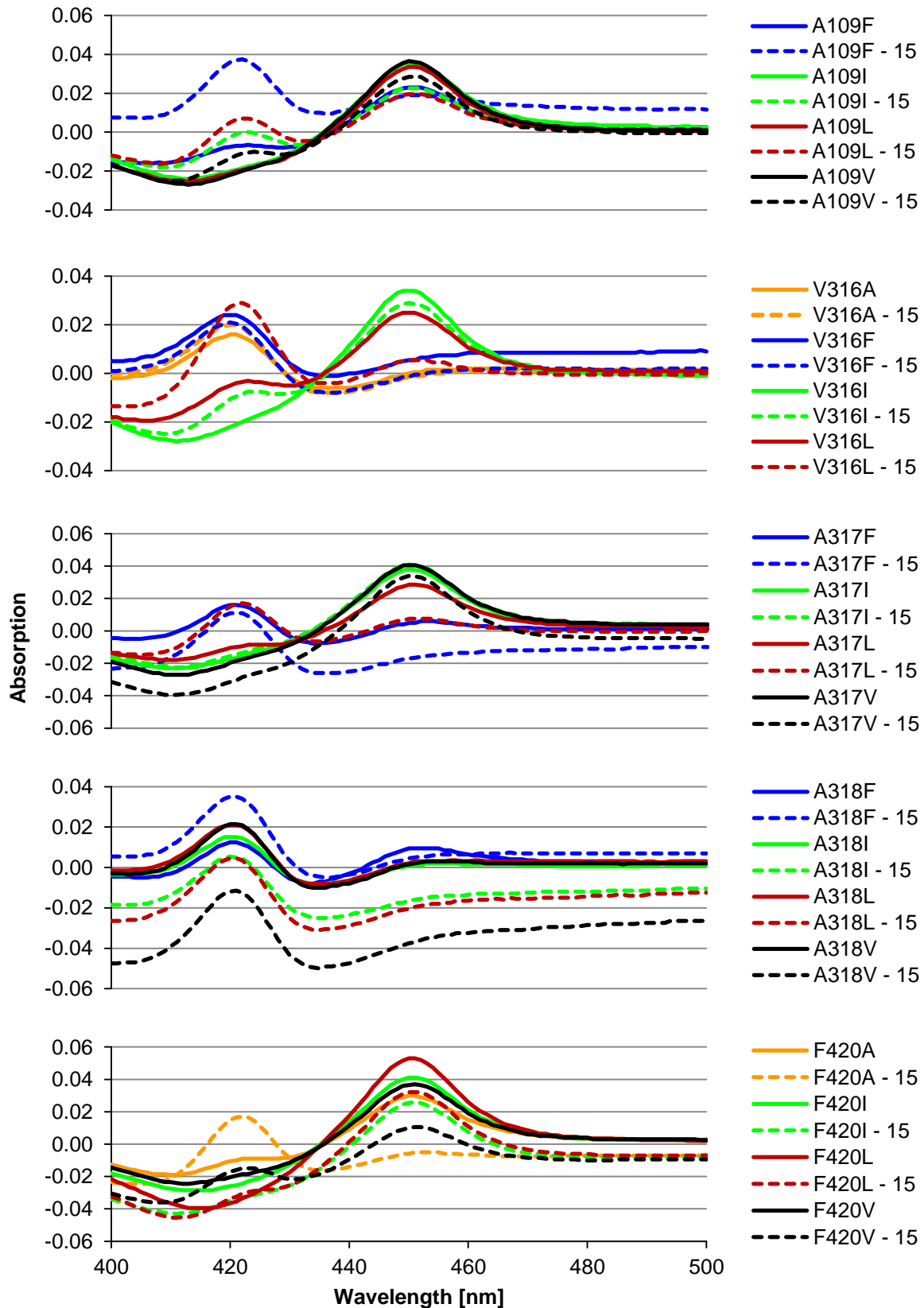


Figure 3.9: CO differential spectral assay of the focused CYP116B3 mutant library variants. Solid lines indicate measurements immediately after bubbling of the samples with CO. Dashed lines indicate measurements of the same samples after 15 min at RT. All measurements were done in duplicate.

Results

The obtained spectra displayed for many of the variants a shoulder or even a peak at a wavelength of 420 nm, especially for the repeated measurements after 15 min. This A_{420} -peak is an indication for inactive enzyme. Hence, variants V316A, V316F and all A318-variants were expected to be rather inactive. As a consequence of the spectra, no concentration of CYP116B3 in the soluble protein fraction could be determined for these variants, except for variant A318F. In contrast, the variants A317I, A317V, F420I and F420L seemed to be relatively stable, according to their spectra.

Table 3.4: CYP116B3 concentration in the soluble protein fraction, expression efficiency and activity of CYP116B3 focused mutant library variants towards 7-ethoxycoumarin.

CYP116B3 variant	c(CYP116B3) [μ M] ^a	c(CYP116B3) [mg (g _{cww}) ⁻¹] ^a	Product formation rate [μ mol (μ mol CYP116B3) ⁻¹ min ⁻¹] ^b
A109F	2.3	0.99	1.81
A109I	3.5	1.51	0.59
A109L	3.6	1.56	0.69
A109V	4.0	1.70	0.41
V316A	-	-	y*
V316F	-	-	n*
V316I	3.9	1.66	0.52
V316L	2.6	1.14	3.95
A317F	0.4	0.19	y*
A317I	4.1	1.75	0.69
A317L	2.9	1.23	1.26
A317V	4.3	1.84	0.81
A318F	0.8	0.33	0.51
A318I	-	-	y*
A318L	-	-	y*
A318V	-	-	y*
F420A	3.0	1.28	0.41
F420I	4.2	1.80	0.49
F420L	5.5	2.37	0.33
F420V	3.7	1.61	0.43

^a-, concentration could not be determined by CO differential spectral assay. ^bReactions were performed in triplicate with a standard deviation of less than 15%. Lysate of empty vector expression was used as a negative control. CYP116B3 was added at a final concentration of 0.5 μ M. *Due to the missing information about CYP116B3 concentration or concentration lower 0.5 μ M, 100 μ l of lysate were added in the reaction mixture for determination of activity. n, no activity; y, displayed activity, but product formation rate could not be determined due to missing information about concentration.

Results

For all variants with a determined value of the CYP116B3 concentration in the lysate, the expression efficiency was calculated subsequently (Table 3.4). Highest expression of 2.37 mg CYP116B3 ($\text{g}_{\text{cww}}^{-1}$) was identified for the variant F420L. Determination of activity was done according to the 7-ethoxycoumarin assay (section 2.2.10.2) by fluorescence spectroscopy (Table 3.4). Corresponding to the calculation of the expression efficiency, only for the variants with a defined CYP116B3 concentration in the lysate, the activity could be determined. For the variants with a very low or undefinable enzyme concentration value, only a qualitative activity analysis was possible.

With respect to the results of the activity determination, the only variant which displayed no activity was the variant V316F. For the variants V316A, A317F, A318I, A318L and A318V, which were expected to be inactive according to the respective CO differential spectra, at least some activity was detected in the fluorescence spectroscopy. Product 7-hydroxycoumarin was generated to some extent. Highest activity was determined for the variant V316L with a product formation rate of $3.95 \mu\text{mol} (\mu\text{mol CYP116B3})^{-1} \text{min}^{-1}$.

Determination of the relative activity of the CYP116B3 variants compared to the wild type enzyme revealed that the activity increased for four of the variants, especially for V316L with almost six times higher activity (Figure 3.10). For two variants (A109L and A317I) activity did not change significantly and for eight variants the activity towards 7-ethoxycoumarin decreased. As mentioned previously, the variants V316A, A317F, A318I, A318L and A318V displayed activity, though the product formation rate could not be determined. Variant V316F was the only variant without any detectable activity.

All variants of the created CYP116B3 library were successfully expressed in *E. coli* BL21(DE3). Both variants with increased and with decreased activity towards the standard substrate 7-ethoxycoumarin were detected as a result of the introduced mutations into CYP116B3. However, in order to investigate the hydroxylation activity and thus applicability of the variants concerning the synthesis of vanillin, the library was screened with the respective aromatic substrates.

Results

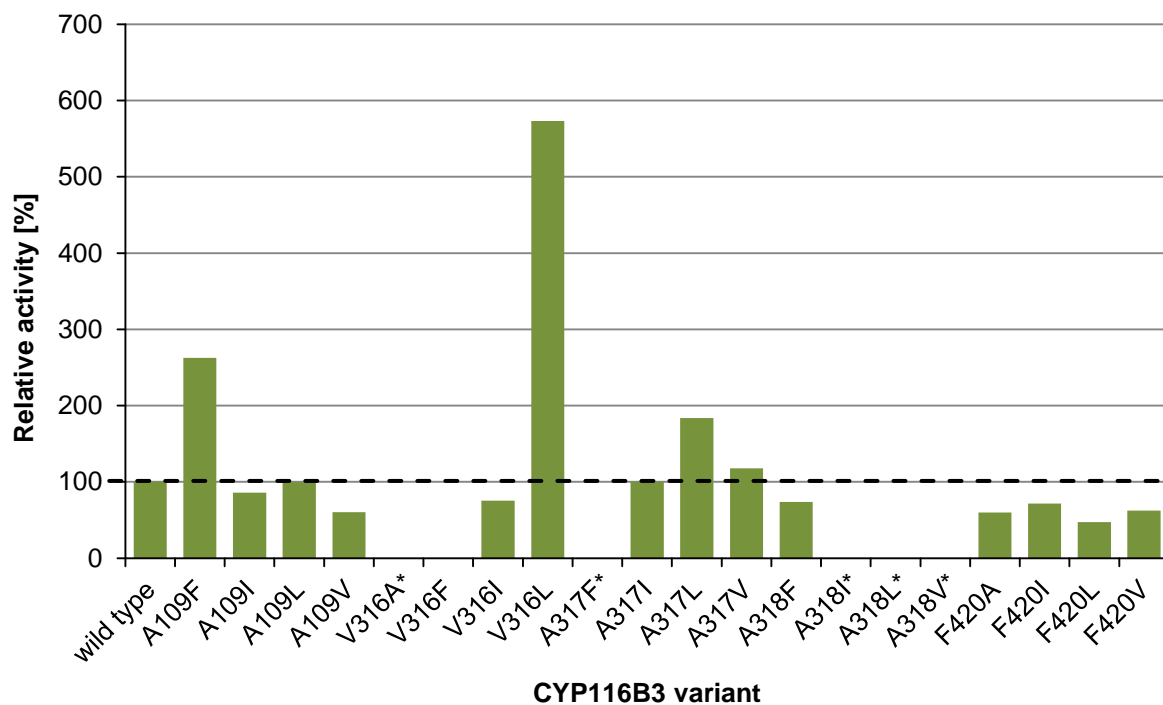


Figure 3.10: Relative activity of CYP116B3 focused mutant library variants compared to the wild type enzyme. Wild type expressed on pET-22b(+) was used as reference. Wild type activity corresponds to 100% and is indicated by a dashed line. *Activity was detected but relative activity could not be determined.

3.2.1.4 *In vitro* biotransformations with the focused CYP116B3 mutant library

Variants of the focused mutant library of CYP102A1 were previously shown to convert the aromatic substrate 3-methylanisole with enhanced activity and changed selectivity (Table 3.3, section 3.1.4). In comparison to the CYP102A1 wild type, the wild type enzyme of CYP116B3 was identified to be more active towards 3-methylanisole. Hence, aiming at the final synthesis of vanillin in accordance with the proposed reaction pathway (Figure 1.8, introduction), the application of the newly created focused mutant library of CYP116B3 in an *in vitro* biotransformation screening for the conversion of aromatic substrates was promising. In addition to 3-methylanisole, the intermediate compounds 3-methoxybenzyl alcohol and 4-methylguaiacol were employed as substrates. Also the conversion of the vanillin precursor vanillyl alcohol was investigated to gain information about overoxidation of this compound.

For the conversion of 3-methylanisole (Table 3.5), only three different products were identified. These are 3-methoxybenzyl alcohol, 4-methoxy-2-methylphenol and *m*-cresol, with a preference for *m*-cresol and thus for the demethylation reaction,

Results

detected for most of the variants. However, two variants, F420I and F420V, displayed a shifted chemoselectivity, as aromatic hydroxylation was the preferred reaction, yielding 4-methoxy-2-methylphenol as main product. In the case of the variant F420I, benzylic hydroxylation with 3-methoxybenzyl alcohol formation was also favored over the demethylation reaction.

Table 3.5: Biotransformation results for the *in vitro* conversion of 3-methylanisole with the focused CYP116B3 mutant library.

CYP116B3 variant	Product formation [% of initial substrate]				Total conversion [%] ¹
	2	3	6	9	
wild type	<1	-	<1	1.4 ± 0.1	2.1 ± 0.0
A109F	-	-	<1	2.2 ± 0.1	3.3 ± 0.2
A109I	<1	-	<1	2.1 ± 0.4	3.8 ± 0.9
A109L	<1	-	<1	2.0 ± 0.3	3.4 ± 0.5
A109V	<1	-	<1	1.4 ± 0.1	2.4 ± 0.3
V316A ^a	<1	-	<1	<1	0.7 ± 0.1
V316F ^a	-	-	-	-	-
V316I	<1	-	<1	1.3 ± 0.1	2.1 ± 0.2
V316L	1.5 ± 0.1	-	1.1 ± 0.2	2.8 ± 0.1	5.4 ± 0.5
A317F ^b	<1	-	<1	1.1 ± 0.1	1.7 ± 0.2
A317I	<1	-	<1	1.1 ± 0.1	2.0 ± 0.2
A317L	<1	-	<1	1.0 ± 0.1	2.0 ± 0.2
A317V	<1	-	<1	1.7 ± 0.1	2.8 ± 0.2
A318F ^b	-	-	-	-	-
A318I ^a	<1	-	<1	<1	0.9 ± 0.1
A318L ^a	<1	-	<1	<1	1.1 ± 0.1
A318V ^a	<1	-	<1	<1	0.6 ± 0.1
F420A	<1	-	<1	<1	0.4 ± 0.1
F420I	1.2 ± 0.2	-	1.3 ± 0.1	<1	2.8 ± 0.3
F420L	<1	-	<1	<1	1.0 ± 0.1
F420V	<1	-	1.0 ± 0.2	<1	2.2 ± 0.4

-, not detected or <0.05%; **2**, 3-methoxybenzyl alcohol; **3**, 4-methylguaiacol; **6**, 4-methoxy-2-methylphenol; **9**, *m*-cresol. Reactions were performed with soluble protein fractions and run at 30°C and 180 rpm for 2 h. CYP116B3 was employed at a final P450 concentration of 0.5 µM. Reactions were performed in triplicate with a standard deviation of less than 15%. Lysate of empty vector expression was used as a negative control. Samples were analyzed by GC-FID. Negative controls showed no conversion. ¹Total conversion was calculated from the determined product formations and was not equivalent to the substrate consumption due to the volatility of 3-methylanisole. ^aDue to the missing information about CYP116B3 concentration, 200 µl of lysate were added to the reaction mixture. ^bDue to the low CYP116B3 concentration in the lysate, the CYP116B3 enzyme was applied at a final concentration of 0.1 µM.

Results

Concerning the synthesis of the vanillin pathway intermediate 4-methylguaiacol, no product formation was detected with any of the CYP116B3 library variants to more than 0.05% of the initial substrate concentration. In a comparison of the total conversion of the variants with the wild type enzyme, only the variants F420A and F420L demonstrated a significant loss of activity. Highest activity of 5.4% total conversion was obtained for the variant V316L. However, for all variants with non-determinable or very low enzyme concentration in the lysate, the determined product formation and conversion values cannot be compared to the wild type or the other variants. Most of these variants were shown to be active towards 3-methylanisole, except the two variants V316F and A318F. V316F was also shown to be inactive with 7-ethoxycoumarin, previously (section 3.2.1.3). Overoxidation of *m*-cresol to methylhydroquinone, detected for some of the CYP102A1 variants in the screening of the focused mutant library (section 3.1.4), was not identified for any of the variants of the CYP116B3 library.

In addition, none of the variants exhibited activity towards 3-methoxybenzyl alcohol. However, towards 4-methylguaiacol as substrate, three variants (V316L, F420I, F420V) were active, producing low amounts of vanillyl alcohol (Table 3.6). Again, highest activity was determined for V316L (1.9% total conversion). The accompanying formation of a very low amount of vanillin (0.11% for V316L) can be attributed to *E. coli* enzymes in the lysate. This is assumed due to the fact that the negative controls with empty vector lysates showed also a basic level of conversion of vanillyl alcohol to vanillin. Both with empty vector lysates and with lysates of all investigated variants a marginal product formation of vanillin from vanillyl alcohol with values between 0.2% and 0.4% was obtained. Only for variant V316L a slightly increased vanillin synthesis was determined, with a maximum of 0.8% total conversion (data not shown in table).

Table 3.6: Biotransformation results for the *in vitro* conversion of 4-methylguaiacol with the focused CYP116B3 mutant library.

CYP116B3 variant	Product formation [% of initial substrate]		Total conversion [%]
	Vanillyl alcohol	Vanillin	
V316L	1.8 ± 0.3	0.11 ± 0.0	1.9 ± 0.0
F420I	0.6 ± 0.1	0.07 ± 0.0	0.7 ± 0.0
F420V	0.6 ± 0.1	0.07 ± 0.0	0.7 ± 0.0

Reactions were performed in triplicate with soluble protein fractions and run at 30°C and 180 rpm for 2 h. CYP116B3 was used at a final concentration of 0.5 µM. Samples were analyzed by HPLC. Negative controls showed no conversion.

Results

Some variants of the generated focused mutant library of CYP116B3 revealed an increased activity and/or changed chemoselectivity towards 3-methylanisole. However, the total conversion values are much lower compared to selected variants of the CYP102A1 mutant library. Moreover, only minor product formation of 3-methoxybenzyl alcohol and no explicit synthesis of 4-methylguaiacol was detected. In contrast to the CYP102A1 variants, three CYP116B3 variants were identified, which were shown to convert 4-methylguaiacol to vanillyl alcohol, though only to a very low extent. Thus, by creation of a focused CYP116B3 mutant library, variants with improved conversion performance were found, although, due to the overall low activity, the applicability for the synthesis of vanillin is not given yet. Further enzyme optimization regarding an increase of the enzymes activity would be necessary first.

3.2.2 Generation and application of further CYP102A1 variants

3.2.2.1 Screening of a set of triple mutant variants of CYP102A1

Previous results demonstrated the beneficial effect of single and double mutations on the conversion of 3-methylanisole with CYP102A1 (section 3.1.4, Table 3.3). In order to further increase the activity and selectivity, an additional mutation site was investigated. Therefore, the amino acid position 437, which is occupied by a leucine in the wild type enzyme (L437) (Figure 3.12), was selected based on the results of a previous study. There, the amino acid at position 437, in addition to the positions 87 and 328, was reported to have an effect on both activity and selectivity of the respective CYP102A1 variants by influencing the orientation of substrates to the active site's heme oxygen.¹⁶⁸ Due to the preference of P450 enzymes for hydrophobic amino acids in substrate-interacting positions,¹³¹ the CYP102A1 double mutant variant F87V/A328L, which exhibited the highest total conversion of 3-methylanisole in the initial biotransformations (section 3.1.4, Table 3.4), was combined with the hydrophobic amino acids alanine, phenylalanine, isoleucine and valine at position 437, alternatively to leucine in the wild type. The respective triple mutant variants of CYP102A1 (F87V/A328L/L437A/F/I/V), were available at the ITB (University of Stuttgart, Stuttgart, Germany) and kindly provided by Dr. Alexander Seifert as indicated in Table 7.2 (supplementary material). Soluble protein fractions of the triple mutant variants were applied for *in vitro* biotransformation reactions of the

Results

aromatic substrates 3-methylanisole, 3-methoxybenzyl alcohol, 4-methylguaiacol and vanillyl alcohol.

As for the single and double mutant variants of CYP102A1, no conversion of 3-methoxybenzyl alcohol, 4-methylguaiacol and vanillyl alcohol was detected with any of the triple mutant variants. However, all variants were active towards 3-methylanisole as a substrate (Table 3.7). Compared to the variant F87V/A328L as a reference, the additional mutation of L437 clearly had an impact on both activity and selectivity of the investigated CYP102A1 variants. A considerable decrease in total conversion was determined for the variants with an alanine or phenylalanine instead of the leucine in position 437. In contrast, replacement of leucine by isoleucine and valine did not affect the variant's activity, though the product distribution changed. This is an indication for an altered selectivity of the respective triple mutant variants F87V/A328L/L437I and F87V/A328L/L437V. With these variants, about twice as much 3-methoxybenzyl alcohol was generated, compared to the corresponding double mutant variant F87V/A328L, and overoxidation of *m*-cresol to methylhydroquinone increased marginally. Besides, product formation of 4-methylguaiacol, 4-methoxy-2-methylphenol and *m*-cresol was diminished approximately by half.

Table 3.7: Biotransformation results for the *in vitro* conversion of 3-methylanisole with selected L437 triple mutant variants of CYP102A1. Results of conversion with variant F87V/A328L are shown for comparison.

CYP102A1 variant	Product formation [% of initial substrate]					Total conversion [%] ¹
	2	3	6	9	10	
F87V/A328L	9.1 ± 0.2	1.6 ± 0.1	23.2 ± 0.4	12.0 ± 0.8	13.5 ± 0.8	59.3 ± 1.8
F87V/A328L/L437A	1.6 ± 0.1	<1	3.1 ± 0.2	4.7 ± 0.2	-	9.5 ± 0.5
F87V/A328L/L437F	1.7 ± 0.2	<1	5.7 ± 0.9	7.3 ± 1.0	<1	15.5 ± 2.3
F87V/A328L/L437I	20.0 ± 0.9	<1	14.0 ± 0.7	7.4 ± 1.3	19.3 ± 0.5	61.3 ± 2.9
F87V/A328L/L437V	22.3 ± 2.0	<1	10.5 ± 1.0	7.0 ± 0.8	18.5 ± 2.7	58.9 ± 5.3

-, not detected or <0.05%; **2**, 3-methoxybenzyl alcohol; **3**, 4-methylguaiacol; **6**, 4-methoxy-2-methylphenol; **9**, *m*-cresol; **10**, methylhydroquinone. Reactions were performed in triplicate with soluble protein fractions and run at 30°C and 180 rpm for 2 h. CYP102A1 was used at a final concentration of 1 µM. Samples were analyzed by GC-FID. Negative controls showed no conversion. ¹ Total conversion was calculated from the determined product formations and was not equivalent to the substrate consumption due to the volatility of 3-methylanisole.

This additional mutation in position 437 did influence both activity and selectivity of the investigated conversion of 3-methylanisole. Nevertheless, no activity with any of the other substrates was detected.

3.2.2.2 Molecular dynamics simulations of CYP102A1 with 4-methylguaiacol

Regarding the previous biotransformation results for a selection of single, double and triple mutant variants of CYP102A1 (section 3.1.4 and section 3.2.1.1), 3-methylanisole was accepted as a substrate by nearly all of the investigated variants. Apart from intense byproduct formation, the desired intermediate products 3-methoxybenzyl alcohol and 4-methylguaiacol were produced in varying amounts by means of benzylic and aromatic hydroxylation, respectively. These compounds were needed for further conversion to the vanillin precursor vanillyl alcohol following the proposed vanillin synthesis pathway (Figure 1.8, introduction). The benzylic hydroxylation of 3-methylanisole was clearly favored over the aromatic hydroxylation in *para* position to the methyl group, as, according to the biotransformation results (Table 3.3 and Table 3.7), the maximal product formation of 3-methoxybenzyl alcohol (about 22%) was much higher compared to the maximal product formation of 4-methylguaiacol (about 3%). However, none of these intermediate compounds was further converted to vanillyl alcohol or any other product.

In order to find out why the synthesis of vanillyl alcohol was ineffective with any of the previously in this study employed CYP102A1 variants, molecular dynamics (MD) simulations were carried out by Dr. Alexander Seifert from the bioinformatics group of the ITB (University of Stuttgart, Stuttgart, Germany). Concerning the two alternative routes of the proposed vanillin synthesis pathway (Figure 1.8, introduction) to produce vanillyl alcohol, either via 3-methoxybenzyl alcohol or via 4-methylguaiacol as intermediates, and due to the identified preference of the CYP102A1 variants for benzylic hydroxylation compared to aromatic hydroxylation in *para* position to the methyl group, the route via benzylic hydroxylation of 4-methylguaiacol was assumed to be more promising. Hence, 4-methylguaiacol was chosen as substrate molecule for the MD simulations in combination with the CYP102A1 variant F87V/A328I, which was selected based on product formation results of the primary 3-methylanisole conversion screening of the focused CYP102A1 library (data not shown). In total, 30 repeated 100 ns MD simulations of the enzyme substrate complex were performed to

Results

discover if 4-methylguaiacol binds in the active site of the enzyme in productive orientations. As a result, 4-methylguaiacol was frequently detected in three different regions: close to its starting position (distance of 10 Å to the activated heme oxygen - in five simulations), in proximity to the activated heme oxygen (distance suitable for oxidation - in six simulations) and in the substrate access channel (distance of about 20 Å to the activated heme oxygen - in four simulations) (Figure 3.11).

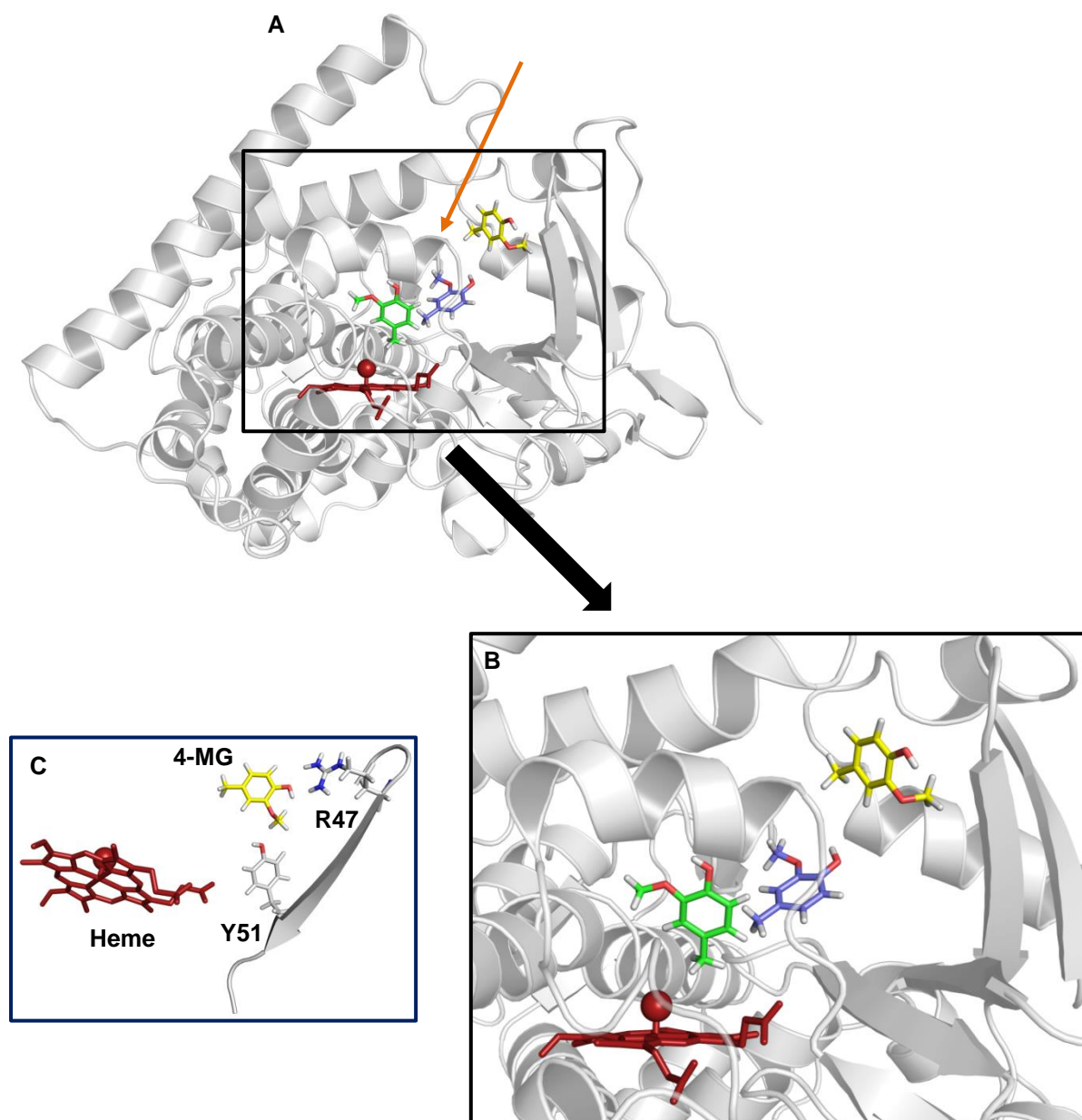


Figure 3.11: MD simulation of CYP102A1 variant F87V/A328I with the substrate 4-methylguaiacol (4-MG). Regions in the active site cavity are shown, where 4-methylguaiacol was frequently detected (**A** and **B** (**B** is zoomed in on **A**)). Heme moiety with oxygen is shown in red. The blue 4-methylguaiacol molecule displays the starting position, the green molecule the position close to the heme center, and in yellow the position in a pocket shaped by the β_1 sheet and the A' helix is illustrated. An orange arrow indicates the direction of the substrate access channel in **A**. **C** is an illustration with a focus on the yellow 4-methylguaiacol molecule located in high distance to the heme, where additionally the substrate channel residues R47 and Y51 are shown as sticks in close proximity to 4-methylguaiacol (rotated view). The images were generated by PyMOL.¹⁶⁶

Results

In principle, 4-methylguaiacol seems to be able to bind in a productive confirmation to the activated heme oxygen of the CYP102A1 variant, as for example in one of the simulations, the 4-methylguaiacol molecule approached the heme within the first 10 ns and exposed its ring methyl group to the activated oxygen for the remaining 90 ns with a distance of less than 4 Å (Figure 7.5, supplementary material). However, in the simulations where 4-methylguaiacol moved away from the active site to a more distant pocket (formed by the β_1 sheet and the A' helix of CYP102A1) in the substrate access channel, the molecule was stabilized by polar interaction/hydrogen bonds with the residues R47 and Y51 (Figure 7.6, supplementary material). This might be a reason for the inactivity of the previously in this study investigated CYP102A1 variants concerning the conversion of 4-methylguaiacol, though 4-methylguaiacol differs from 3-methylanisole, a proven good substrate, by only one hydroxyl group.

3.2.2.3 Biotransformations with R47L/Y51F-variants of CYP102A1

As the MD simulations revealed a stabilizing effect of the polar substrate channel residues R47 and Y51 (Figure 3.11, Figure 3.12) on the substrate 4-methylguaiacol, these residues were replaced by the hydrophobic residues leucine and phenylalanine, respectively, in order to enable the conversion of 4-methylguaiacol by variants of CYP102A1. Therefore, by means of rational protein design using respective oligonucleotides (Table 7.4, supplementary material), the mutations R47L and Y51F were introduced into selected variants of CYP102A1, which displayed the highest biotransformation activity with 3-methylanisole in previous investigations (Table 3.3, section 3.1.4 and Table 3.7, section 3.2.2.1).

Results

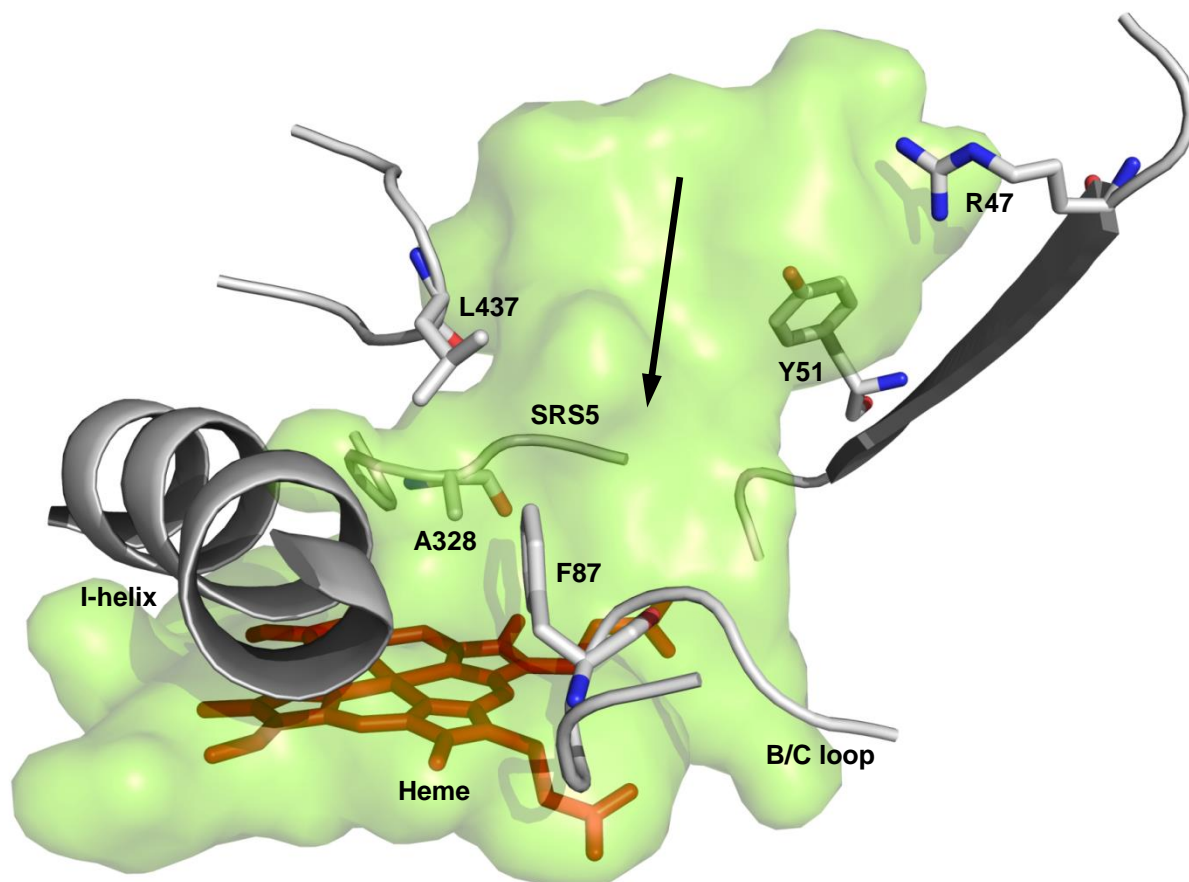


Figure 3.12: Substrate and heme access region in CYP102A1 (PDB ID: 1BU7, chain A). Substrate channel residues R47 and Y51 and active site residues, including the heme group (red), as well as parts of the I-helix, the B/C loop with F87 and the substrate recognition site 5 (SRS5) with residues A328 and L437 (positioned at the entrance to the active site cavity) are depicted. The entrance route of the substrate to the active site is indicated by a black arrow. The image was generated by PyMOL.¹⁶⁶

As a result, a set of double, triple, quadruple and one quintuple mutant variants of CYP102A1 was generated (some variants were already available at the ITB (University of Stuttgart, Stuttgart, Germany) and kindly provided by Dr. Alexander Seifert as indicated in Table 7.2 (supplementary material)) and investigated for conversion of both 4-methylguaiacol and the compounds 3-methylanisole, 3-methoxybenzyl alcohol and vanillyl alcohol.

Six out of the set of eight investigated CYP102A1 variants, carrying the R47L and Y51F mutations, successfully converted 4-methylguaiacol to vanillyl alcohol (Table 3.8). Only the variants R47L/Y51F and R47L/Y51F/A328L were completely inactive with 4-methylguaiacol. The highest production of vanillyl alcohol with a conversion of 4.5% was determined for the variant R47L/Y51F/F87V/A328V. In addition to vanillyl alcohol, an unknown byproduct was detected as a very small peak for all quadruple mutants and the one quintuple mutant variant (data not shown).

Results

Table 3.8: Biotransformation results for the *in vitro* conversion of 4-methylguaiacol with R47L/Y51F-variants of CYP102A1.

CYP102A1 variant	Conversion to vanillyl alcohol [%]
R47L/Y51F	-
R47L/Y51F/A328L	-
R47L/Y51F/F87V	<1
R47L/Y51F/F87V/A328F	1.6 ± 0.1
R47L/Y51F/F87V/A328I	3.0 ± 0.3
R47L/Y51F/F87V/A328L	<1
R47L/Y51F/F87V/A328V	4.5 ± 0.3
R47L/Y51F/F87V/A328L/L437I	3.1 ± 0.6

-, not detected or <0.05%; Reactions were performed in triplicate with soluble protein fractions and run at 30°C and 180 rpm for 2 h. CYP102A1 was used at a final concentration of 0.5 µM. Samples were analyzed by HPLC. Negative controls showed no conversion.

In contrast to 4-methylguaiacol, the substrates 3-methoxybenzyl alcohol and vanillyl alcohol were not converted by any of the R47L/Y51F-variants. However, all variants were active with 3-methylanisole (Table 3.9).

Table 3.9: Biotransformation results for the *in vitro* conversion of 3-methylanisole with R47L/Y51F-variants of CYP102A1.

CYP102A1 variant	Product formation [% of initial substrate]					Total conversion [%] ¹
	2	3	6	9	10	
R47L/Y51F	<1	-	1.5 ± 0.1	-	-	1.6 ± 0.1
R47L/Y51F/A328L	<1	1.2 ± 0.1	12.5 ± 1.3	<1	-	14.8 ± 1.5
R47L/Y51F/ F87V	2.1 ± 0.1	1.6 ± 0.1	35.1 ± 3.1	1.2 ± 0.1	2.0 ± 0.3	42.5 ± 3.7
R47L/Y51F/ F87V/A328F	2.1 ± 0.3	1.3 ± 0.2	26.1 ± 3.5	5.1 ± 0.4	5.8 ± 1.4	40.3 ± 5.8
R47L/Y51F/ F87V/A328I	10.4 ± 1.1	<1	9.5 ± 0.8	6.9 ± 0.5	1.6 ± 0.3	28.8 ± 2.7
R47L/Y51F/ F87V/A328L	4.9 ± 0.5	<1	13.9 ± 0.9	9.7 ± 0.3	7.9 ± 0.9	37.4 ± 2.6
R47L/Y51F/ F87V/A328V	5.0 ± 0.3	<1	24.0 ± 0.8	3.0 ± 0.1	2.1 ± 0.2	34.9 ± 1.4
R47L/Y51F/ F87V/A328L/L437I	13.5 ± 0.4	<1	10.1 ± 0.5	7.7 ± 0.2	17.8 ± 0.8	50.4 ± 1.8

-, not detected or <0.05%; **2**, 3-methoxybenzyl alcohol; **3**, 4-methylguaiacol; **6**, 4-methoxy-2-methylphenol; **9**, *m*-cresol; **10**, methylhydroquinone. Reactions were performed in triplicate with soluble protein fractions and run at 30°C and 180 rpm for 2 h. Samples were analyzed by GC-FID. Negative controls showed no conversion. CYP102A1 was used at a final concentration of 1 µM. ¹Total conversion was calculated from the determined product formations and was not equivalent to the substrate consumption due to the volatility of 3-methylanisole.

Results

Low conversion of 3-methylanisole was detected for the variant R47L/Y51F. For the quadruple variant R47L/Y51F/F87V/A328V, both total conversion and product distribution was comparable to the values of the respective double mutant variant F87V/A328V. All other R47L/Y51F-variants displayed varying activity towards 3-methylanisole, compared to the respective variants without the R47L and Y51F mutations, whereas the product distribution did not change significantly. Only methylhydroquinone, the product of the overoxidation of *m*-cresol, was produced for some more variants as a result of the higher activity. With a total conversion of 50%, the quintuple variant R47L/Y51F/F87V/A328L/L437I demonstrated highest activity, although lower compared to the respective triple mutant variant F87V/A328L/L437I.

The additional R47L and Y51F mutations had different effects on enzyme activity and selectivity with respect to the conversion of 3-methylanisole. However, when 4-methylguaiacol was used as a substrate for the R47L/Y51F-variants of CYP102A1, the formation of vanillyl alcohol was detected for the first time in a CYP102A1-catalyzed reaction. These additionally introduced substrate channel mutations were shown to expand the spectrum of the investigated CYP102A1 variants to 4-methylguaiacol as expected after MD simulation.

3.2.3 Investigation of the vanillyl alcohol oxidase VAO from *P. simplicissimum*

Various variants of CYP116B3 and especially of CYP102A1 were previously generated and identified to catalyze the first two steps of the proposed vanillin synthesis cascade reaction (Figure 1.8, introduction) from 3-methylanisole to 4-methylguaiacol and from 4-methylguaiacol to vanillyl alcohol. In literature, the flavoenzyme vanillyl alcohol oxidase (VAO) from *Penicillium simplicissimum* was reported not only to be highly selective for the conversion of vanillyl alcohol to vanillin, but also to accept 4-methylguaiacol as a substrate.^{72, 169} Thereby, the VAO-catalyzed conversion of 4-methylguaiacol proceeds via vanillyl alcohol directly to vanillin as final product. Hence, VAO was selected for application in the vanillin cascade synthesis reported here.

3.2.3.1 Site-directed mutagenesis, expression and investigation of activity of VAO

The *vaoA* gene from *P. simplicissimum* CBS 170.90 was obtained as a synthetic gene (GeneArt[®] Gene Synthesis, Life Technologies GmbH, Darmstadt, Germany), codon-optimized for the heterologous expression in *E. coli*. From the pMK-RQ_VAO vector construct, the *vaoA* gene was cloned into pET-22b(+) by restriction enzyme digestion with *NdeI* and *XhoI* and ligation with linearized pET-22b(+) vector. The sequence of the pET-22b(+)_VAO construct was verified by automated DNA-sequencing (GATC Biotech, Konstanz, Germany). Moreover, as single mutant variants of VAO, created by random mutagenesis with an up to 40-fold increased catalytic efficiency (k_{cat}/K_M) towards 4-methylguaiacol as substrate were reported in a previous study,⁶⁹ two of these mutations were chosen for investigation concerning their applicability in the vanillin synthesis pathway here. Single mutant variants F454Y and T505S, which were described to have the highest catalytic efficiency at pH 7.5 and pH 10, respectively, were generated by site-directed mutagenesis based on the pET-22b(+)_VAO vector construct using the appropriate oligonucleotides (Table 7.4, supplementary material). The sequence of the variants was verified by automated DNA-sequencing (GATC Biotech, Konstanz, Germany). SDS-PAGE analysis of the expression of both the VAO wild type and the variants F454Y and T505S in *E. coli* BL21(DE3) indicated VAO enzyme, with a calculated molecular weight of 63.0 kDa (calculated from the amino acid sequence using the CloneManager (Sci-Ed Software, Cary, NC, USA) software) and an apparent molecular weight of about 60 kDa, in the soluble protein fraction. The majority of the enzyme seemed to be present in the insoluble fraction (Figure 3.13).

In order to investigate the enzyme activity, wild type and variants of VAO were applied for *in vitro* biotransformation of 4-methylguaiacol at pH 7.5 and pH 10. At both pH-conditions the highest conversion of 4-methylguaiacol was determined for the variant F454Y (Table 3.10). According to the literature, at pH 10 the variant T505S was reported to have the higher catalytic efficiency.⁶⁹ Main product of the 4-methylguaiacol conversion was vanillin for all of the investigated reactions. Only for the wild type enzyme the intermediate product vanillyl alcohol was detected to a very low extent, which is a result of the lower enzyme activity.

Results

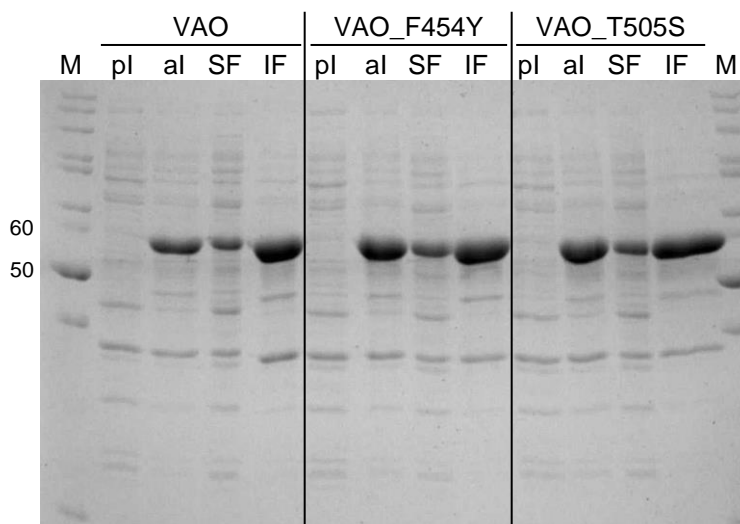


Figure 3.13: SDS-PAGE analysis of the expression of VAO wild type and variants. Expression was investigated for induction with 0.5 mM IPTG and incubation at 20°C with shaking at 140 rpm O/N. M, marker (unstained protein ladder (10-200 kDa) – respective molecular weight is indicated); pl, whole cells prior to induction; al, whole cells after induction and incubation O/N; SF, soluble protein fraction; IF, insoluble fraction after cell lysis.

Table 3.10: Biotransformation results for the *in vitro* conversion of 4-methylguaiacol with VAO wild type and variants F454Y and T505S at pH 7.5 and pH 10.

Enzyme (wild type) variant	Product formation [% of initial substrate]		Total conversion [%]
	Vanillyl alcohol	Vanillin	
pH 7.5			
VAO	<1	53.2 ± 5.5	54.1 ± 5.7
F454Y	-	80.3 ± 3.2	80.3 ± 3.2
T505S	-	69.6 ± 7.5	69.6 ± 7.5
pH 10			
VAO	<1	2.8 ± 0.5	3.6 ± 0.5
F454Y	-	81.3 ± 4.2	81.3 ± 4.2
T505S	-	71.9 ± 5.1	71.9 ± 5.1

Reactions were performed in triplicate with soluble protein fractions and run at 30°C and 180 rpm for 2 h. VAO lysate was used at a final total protein concentration of 1 mg/ml. Samples were analyzed by HPLC. Negative controls showed no conversion.

Compared to the wild type enzyme, activity was clearly increased for both variants F454Y and T505S, with comparable synthesis of vanillin at pH 7.5 and at pH 10, respectively. In contrast, for the wild type enzyme the activity at pH 10 was much lower compared to the activity at pH 7.5.

Results

Concerning the activity towards vanillyl alcohol as substrate, the wild type enzyme and the variants were applied in *in vitro* biotransformation reactions at pH 7.5. As a result, complete conversion of vanillyl alcohol to vanillin was detected both for the wild type and for the variants. Moreover, no byproduct formation or overoxidation of vanillin was identified.

To further explore the applicability of VAO in the proposed vanillin synthesis cascade pathway, possible side-reactions with all compounds, potentially present in the reaction mixture of a one-pot setup, were investigated. These are, next the substrate 3-methylanisole, the intermediate compounds 3-methoxybenzyl alcohol and 4-methylguaiacol as well as the side-products 4-methoxy-2-methylphenol, 2-methoxy-6-methylphenol and *m*-cresol and the overoxidation product methylhydroquinone. Neither by the VAO wild type nor by the variants F454Y or T505S, any of these compounds were accepted as a substrate (data not shown).

All enzymes were successfully expressed and shown to be active (complete conversion of vanillyl alcohol to vanillin at pH 7.5). However, compared to the wild type enzyme, a much higher activity towards 4-methylguaiacol was determined for both variants. Finally, the variant F454Y exhibited highest activity towards 4-methylguaiacol at pH 10, slightly higher compared to its activity at pH 7.5.

3.3 *In vitro* one-pot cascade synthesis of vanillin

In order to synthesize vanillin from 3-methylanisole in a multi-enzymatic three-step cascade reaction, a variety of enzymes and variants thereof was investigated in previous *in vitro* biotransformations concerning the realization of the two possible routes of the proposed vanillin synthesis reaction pathway (Figure 1.8, introduction). As none of the investigated enzymes and variants was able to convert 3-methoxybenzyl alcohol, this route was not considered in the further proceeding. As a consequence, focus was set on the synthesis of vanillin according to the route via 4-methylguaiacol. Enzymes and variants were identified and applied in separate *in vitro* reactions for all three different steps of this route, previously. While most efficient conversion of 3-methylanisole to 4-methylguaiacol was achieved with the CYP102A1 variant A328L, the quadruple CYP102A1 variant R47L/Y51F/F87V/A328V showed highest conversion of 4-methylguaiacol to vanillyl

Results

alcohol. For the third and final reaction step from vanillyl alcohol to vanillin, the well-known VAO variant F454Y was shown to be a suitable catalyst. Moreover, VAO_F454Y was also shown to convert 4-methylguaiacol to vanillin in high yield, via the intermediate compound vanillyl alcohol.⁶⁹

In a first *in vitro* one-pot reaction setup, the two P450 variants CYP102A1_A328L and CYP102A1_R47L/Y51F/F87V/A328V were combined to examine the two-step cascade synthesis of the vanillin precursor vanillyl alcohol from 3-methylanisole. After 2 h reaction time, containing both enzyme variants in the reaction mixture from the start of the reaction, about 25% of the substrate was converted, with 4-methoxy-2-methylphenol as main product (Table 3.11). Even though 4-methylguaiacol was produced to 1.2% of the initial substrate, no concomitant synthesis of vanillyl alcohol was detected. By successive addition of the second enzyme variant (R47L/Y51F/F87V/A328V) after 1 h (half of the reaction time), the conversion was increased to about 30% and formation of 4-methylguaiacol was doubled. However, still no vanillyl alcohol was identified.

Table 3.11: Biotransformation results for the *in vitro* conversion of 3-methylanisole with a combination of enzymes in one reaction system.

Enzymes	Conversion time [h]	Product formation [% of initial substrate]				Total conversion [%] ¹
		2	3	5	6	
CYP102A1_A328L	2	2.2 ± 0.1	1.2 ± 0.1	-	19.3 ± 1.3	24.7 ± 1.6
+ CYP102A1_R47L/Y51F/F87V/A328V	1+1	1.2 ± 0.2	2.4 ± 0.5	-	25.6 ± 4.8	30.3 ± 5.7
	2	<1	<1	<1	19.8 ± 2.7	22.9 ± 3.1
CYP102A1_A328L	1+1	1.0 ± 0.1	<1	1.2 ± 0.1	25.0 ± 2.2	29.0 ± 2.5
+ VAO_F454Y	8	<1	<1	1.2 ± 0.5	20.1 ± 3.0	23.2 ± 3.5
	4+4	1.1 ± 0.1	<1	2.0 ± 0.2	27.1 ± 2.3	31.4 ± 2.7

-, not detected or <0.05%; **2**, 3-methoxybenzyl alcohol; **3**, 4-methylguaiacol; **5**, vanillin; **6**, 4-methoxy-2-methylphenol. Reactions were performed in triplicate with soluble protein fractions lysate and run at 30°C and 180 rpm. Conversion time “1+1” or “4+4” indicate successive addition of the second enzyme after 1 h or 4 h, respectively. Samples were analyzed by GC-FID. Negative controls showed no conversion. ¹Total conversion was calculated from the determined product formations and was not equivalent to the substrate consumption due to the volatility of 3-methylanisole.

As a further variation, both separate and combined addition of catalase and superoxide dismutase (SOD) to the biotransformation mixture was investigated, in order to increase the enzyme's activity by minimizing the negative influence of

Results

possibly generated reactive oxygen species. No significant change in product distribution and total conversion was measured (data not shown).

In a further *in vitro* one-pot reaction setup CYP102A1_A328L was combined with the vanillyl alcohol oxidase variant VAO_F454Y aiming at the direct three-step biocatalytic cascade synthesis of vanillin. The total conversion of 3-methylanisole and the product formation of the main product 4-methoxy-2-methylphenol were comparable to the first reaction setup with two CYP102A1 variants (Table 3.11). Moreover, again an increase in total conversion (from about 23% to 29%) was detected by successive addition of the second enzyme, here the VAO variant F454Y, after half of the reaction time. However, in contrast to the first cascade reaction setup, vanillin was synthesized with 1.2% for the successive addition of the enzymes, whereas the determined production of 4-methylguaiacol was reduced to less than 1%. Formation of the vanillin precursor vanillyl alcohol was not detected for any of the investigated reactions. Consequently, this is an indication that in a first reaction step the CYP102A1 variant A328L catalyzed the conversion of 3-methylanisole to 4-methylguaiacol, which was further converted in two successive reactions via vanillyl alcohol to vanillin, catalyzed by the VAO variant F454Y (Figure 3.14). Thereby, as no vanillyl alcohol was identified, the last step was a complete conversion.

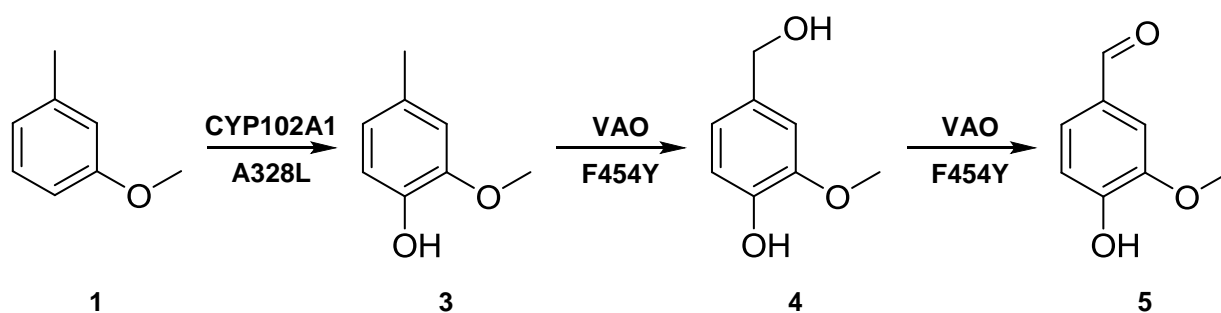


Figure 3.14: Cascade conversion of 3-methylanisole (1) via 4-methylguaiacol (3) and vanillyl alcohol (4) to vanillin (5), catalyzed by the CYP102A1 variant A328L and the VAO variant F454Y.

In order to increase the vanillin synthesis, the total reaction time was extended to 8 h. However, conversion as well as product formation did not change significantly, though the vanillin production increased to 2% in the setup with the successive addition of the enzymes (4+4) (Table 3.11). Furthermore, the addition of catalase to the reaction mixture with VAO was investigated (for 2 h total reaction time), as in the VAO-catalyzed reactions hydrogen peroxide is known to be produced. However, this

Results

resulted in dramatically lower conversion (less than half of the total conversion compared to the reactions without catalase; data not shown). While the side-product *m*-cresol was detected in all reactions in very low amounts, the overoxidation product methylhydroquinone was only identified for the biotransformations with the CYP102A1 variant R47L/Y51F/F87V/A328V (<1%; data not shown).

In conclusion, it was shown that vanillyl alcohol could not be synthesized by a combination of the two CYP102A1 variants A328L and R47L/Y51F/F87V/A328V *in vitro*. However, in contrast to the first reaction setup, vanillin was synthesized by a combination of the CYP102A1 variant A328L and the VAO variant F454Y. Consequently, the second route of the proposed vanillin synthesis cascade reaction was successfully realized *in vitro*, as a proof-of-principle.

3.4 *In vivo* one-pot cascade synthesis of vanillin

Due to the previously shown synthesis of vanillin from 3-methylanisole in a one-pot two-enzymatic three-step cascade conversion *in vitro* (section 3.3), the same system was now investigated for its applicability *in vivo*. Therefore, *E. coli* BL21(DE3), which was chosen as expression host, was transformed with the plasmids pET-28a(+)_CYP102A1_A328L and pET-22b(+)_VAO_F454Y. Expression was investigated by SDS-PAGE (Figure 3.15), showing strong protein bands at around 60 kDa and 120 kDa, indicating successful expression of VAO (63.0 kDa calculated molecular weight) and CYP102A1 (117.8 kDa calculated molecular weight), respectively, with a preference for VAO expression. After protein expression, the cells were harvested, washed with buffer and adjusted in the biotransformation medium to a final cell wet weight of 50 g l⁻¹ for *in vivo* whole-cell conversion of 3-methylanisole. Samples of the *in vivo* biotransformation were taken at several time points for up to 24 h and analyzed by GC-FID.

Results

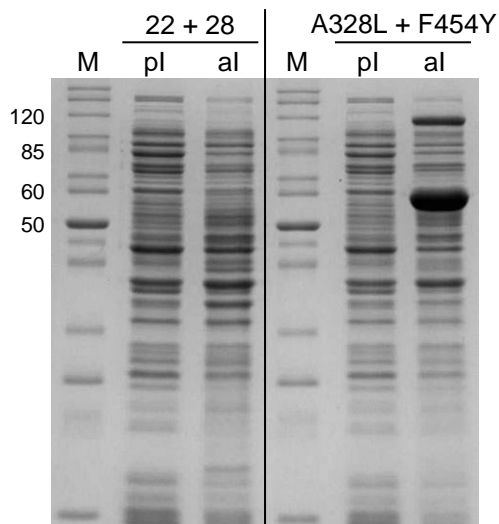


Figure 3.15: SDS-PAGE analysis of the expression for *in vivo* whole-cell biotransformation of 3-methylanisole with a combination of CYP102A1_F454Y and VAO_F454Y. Expression was investigated for induction with 0.1 mM IPTG and incubation at 20°C with shaking at 140 rpm O/N. M, marker (unstained protein ladder (10-200 kDa) – respective molecular weight is indicated); pl, whole cells prior to induction; al, whole cells after induction and incubation O/N.

According to the analysis, 4-methoxy-2-methylphenol was the only conversion product present within the first hour of reaction (Table 3.12). Moreover, this compound was detected to be the main reaction product during all measurements, as determined in previous *in vitro* biotransformations with CYP102A1 enzyme (section 3.1.4 and section 3.2.2). In addition to 4-methoxy-2-methylphenol, the byproducts 3-methoxybenzyl alcohol and the *m*-cresol were detected in low amounts after 3 h. A low product formation of 4-methylguaiacol, the intermediate product in the intended cascade synthesis of vanillin, was determined in the biotransformation samples after 2 h. The vanillin precursor vanillyl alcohol was not detected at all during the *in vivo* biotransformations. However, while the concentration of 3-methoxybenzyl alcohol and *m*-cresol did not change significantly throughout the whole investigation, the product formation of 4-methylguaiacol decreased after 4 h and finally disappeared, while the formation of vanillin increased steadily from 3 h to 12 h reaction time with a maximum of 1.1% after 12 h (Figure 3.16). Total conversion also reached a maximum after 12 h with a value of almost 12%, although the biotransformation medium was supplemented with fresh glycerol and glucose after 12 h.

Results

Table 3.12: Biotransformation results for the *in vivo* conversion of 3-methylanisole with a combination of CYP102A1_A328L and VAO_F454Y in one cellular whole-cell reaction system.

Conversion time [h]	Product formation [% of initial substrate]					Total conversion [%] ¹
	2	3	5	6	9	
0	-	-	-	-	-	-
0.25	-	-	-	<0.5	-	<0.5
0.5	-	-	-	<0.5	-	<0.5
1	-	-	-	<0.5	-	<0.5
2	-	<0.5	-	0.9 ± 0.1	-	1.0 ± 0.1
3	<0.5	<0.5	<0.5	3.6 ± 0.4	<0.5	4.3 ± 0.4
4	<0.5	<0.5	0.5	6.4 ± 0.1	<0.5	7.6 ± 0.2
6	<0.5	<0.5	0.8	8.1 ± 0.2	<0.5	9.7 ± 0.3
8	<0.5	<0.5	0.9 ± 0.1	9.2 ± 0.8	<0.5	10.8 ± 1.0
12	<0.5	-	1.1 ± 0.1	9.9 ± 1.3	<0.5	11.7 ± 1.4
24	<0.5	-	1.1 ± 0.3	9.9 ± 2.4	<0.5	11.7 ± 2.6

-, not detected or <0.05%; **2**, 3-methoxybenzyl alcohol; **3**, 4-methylguaiacol; **5**, vanillin; **6**, 4-methoxy-2-methylphenol; **9**, *m*-cresol. Reactions were performed in triplicate at 30°C and 180 rpm for 24 h. Samples at different time points were analyzed by GC-FID. Negative controls showed no conversion. ¹Total conversion was calculated from the determined product formations and was not equivalent to the substrate consumption due to the volatility of 3-methylanisole.

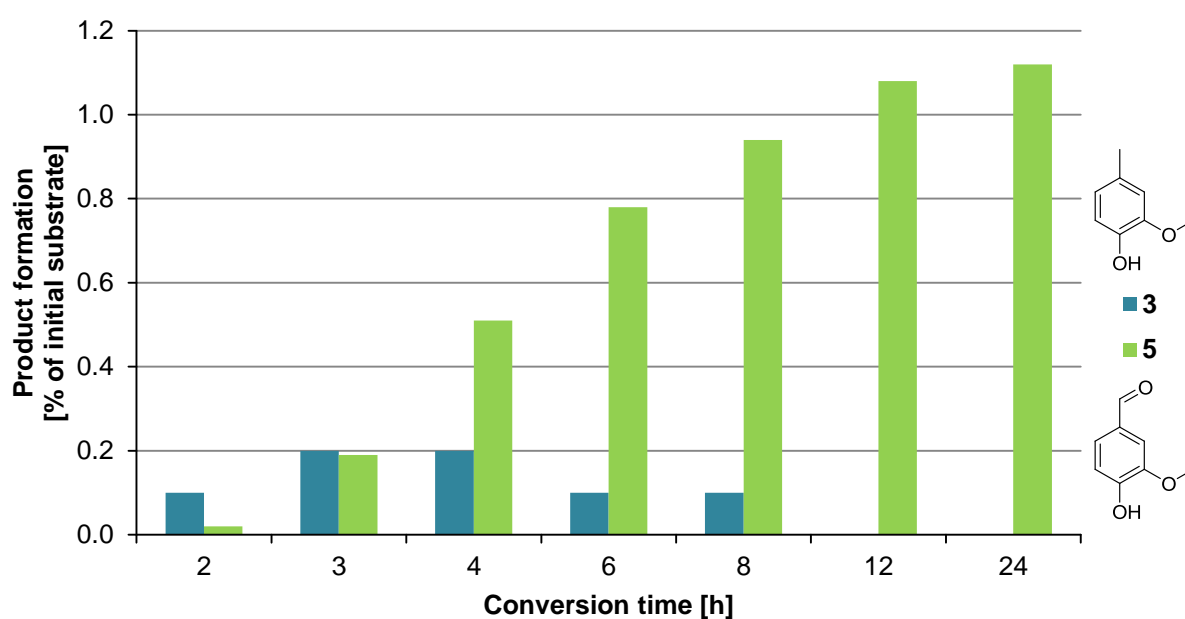


Figure 3.16: Product formation of 4-methylguaiacol (**3**) and vanillin (**5**) during the *in vivo* whole-cell biotransformation of 3-methylanisole.

Results

The results of the *in vivo* whole-cell biotransformation clearly indicate that, as for the *in vitro* one-pot cascade synthesis of vanillin (section 3.3.), 3-methylanisole was converted via 4-methylguaiacol to vanillyl alcohol, which was then completely oxidized to vanillin. Successful vanillin synthesis was also confirmed by GC-MS analysis (Figure 7.7, supplementary material).

Thus, by a combination of two different enzymes, the cytochrome P450 monooxygenase CYP102A1 (variant A328L) and the vanillyl alcohol oxidase VAO (variant F454Y), the synthesis of vanillin was achieved both *in vitro* and *in vivo* according to the proposed cascade reaction pathway (Figure 1.8, introduction).

4 Discussion

The catalytic aerobic oxidation of thermodynamically strong and kinetically inert C-H bonds in hydrocarbons is one of the most sought-after but also one of the most difficult chemical reactions to achieve.^{10, 14, 15, 170} The realization of such a reaction, especially with functionally complex compounds and with controlled and high chemo-, regio-, and stereoselectivities under environmentally friendly and safe conditions, still remains very challenging,^{9, 11-13} though with the usage of transition metal complexes and the development of various methods for catalytic functionalization the activation of C-H bonds has been shown with many substrates, particularly arenes.¹⁷ However, the concomitant formation of product mixtures is still the main challenge for chemocatalysis.

For the intended oxidation of C-H bonds in aromatic compounds, biocatalysis, employing enzymes as catalysts, can be applied as an alternative to chemical syntheses by heterogeneous or homogeneous catalysis.¹⁰ Naturally evolved enzymes that can be found in virtually all habitats in nature, are known to catalyze a variety of reactions, for example the above mentioned challenging C-H bond functionalization in aqueous solution at ambient temperature and pressure.¹⁸ Thereby, cofactors, that are positioned in the active site of the enzyme, are involved in the enzymatic mechanisms and the reactivity of the catalytic enzyme depends on the type of the employed cofactor.^{9, 19} Moreover, the positioning of these cofactors plays an important role in controlling the substrate scope of the enzyme and the selectivity of the reaction.¹⁹

Challenges in biocatalysis, like insufficient enzyme activity, recombinant enzyme expression, enzyme purification, cofactor supply, organic solvent and oxygen tolerance, stability and substrate scope³⁷⁻⁴¹ can be addressed by laboratory evolution techniques like, for example, protein engineering (directed evolution, rational protein design) and metabolic engineering of production hosts, which allows the combination of enzymes from different organisms.⁹ As an example for successful application of laboratory evolution techniques in biocatalysis, the olefin cyclopropanation via carbene transfer catalyzed by engineered P450s can be mentioned.¹³⁰ Thereby, natural enzymes were screened for non-native activity against synthetic reagents, and both activity and stereoselectivities could be improved.

In this project, the establishment of a consecutive enzyme-catalyzed oxidation reaction yielding the valuable and sought-after aromatic compound vanillin as an exemplary product of interest was investigated. In order to realize the establishment of this cascade reaction according to the proposed reaction pathway as shown previously (Figure 1.8, introduction), a selection of enzymes was chosen as catalysts for initial investigations on the hydroxylation of aromatic substrates.

4.1 Enzyme selection, preparation and application in initial biotransformations

As, amongst others, flavoproteins and heme enzymes have been described to be of special interest concerning the challenging hydroxylation of aromatic substrates,⁹ the *m*-hydroxybenzoate hydroxylase MobA from *C. testosteroni* GZ39 and the cytochrome P450 monooxygenases CYP116B3 from *R. ruber* DSM 44319 and CYP102A1 from *B. megaterium* ATCC 14581 as well as variants thereof were chosen as representatives of these enzyme families, respectively, and employed for investigations in initial biotransformation reactions.

The FAD-dependent *m*-hydroxybenzoate hydroxylase MobA from *C. testosteroni* GZ39, a single-component FPMO, was reported to display a generally strict substrate specificity, with the conversion of 3-hydroxybenzoate (*m*-hydroxybenzoate) to 3,4-dihydroxybenzoate under consumption of equimolar amounts of NADPH and molecular oxygen as an exemplary reaction.⁵⁷ In order to alter the substrate specificity of MobA, mutants were reported to have been generated by means of directed evolution followed by *in vivo* screening. One of these mutants was the MobA variant V257A, which was shown to successfully convert phenol to catechol, besides other substituted phenols that were reported to be converted with enhanced efficiency.⁶⁰ Hence, both MobA wild type and variant V257A were chosen for investigation in initial aromatic substrate biotransformations *in vitro*. The variant V257A was generated by site-directed mutagenesis and both wild type enzyme and variant V257A of MobA were expressed in *E. coli* BL21(DE3). High yields of MobA wild type and variant V257A were obtained in the soluble protein fraction after cell lysis, though distinct amounts were still detected in the insoluble fraction. In addition, MobA wild type was purified by IMAC. A first indication for the activity of MobA wild

type and variant V257A, obtained by spectro-photometrical analysis, was confirmed by the investigation of the NADPH depletion in an *in vitro* conversion reaction of *m*-hydroxybenzoate to protocatechuate catalyzed by respective enzyme samples. Nevertheless, neither for MobA wild type nor for variant V257A conversion towards the aromatic compounds was detected in *in vitro* biotransformation reactions. This is in accordance with the known generally strict substrate specificity of MobA which was reported previously.⁵⁷ The reason for this narrow substrate range could be the topology and the properties of the enzymes substrate channel and active site cavity. The environment of the tunnel interior was reported to be characterized by distinctly divided hydrophilic and hydrophobic regions. Additionally, the size of the tunnel entrance is assumed to play an important role in enzymatic substrate selection by the amphiphilic nature and molecular size of the substrate.⁵¹ Finally, as there was no initial activity indicated with the substrates of interest, no further investigations of MobA were undertaken at this point.

The cytochrome P450 monooxygenases CYP116B3 from *R. ruber* DSM 44319 was chosen as candidate for initial biotransformation reactions towards the realization of the proposed vanillin synthesis pathway (Figure 1.8, introduction) as this enzyme was shown previously to catalyze the conversion of various monocyclic and polycyclic aromatic hydrocarbon substrates with and without alkyl side chains (acenaphthene, fluorene, indene, naphthalene, toluene, *m*-xylene and ethyl benzene).¹⁴⁰ Besides, CYP116B3 was reported to be a self-sufficient fusion enzyme being composed of a ferredoxin domain and a FMN-containing NADPH-dependent flavoprotein reductase domain fused to a heme-containing P450 domain requiring only the cofactor NADPH as an electron donor and a substrate for oxidative catalysis,^{137, 140} which makes this enzyme relatively comfortable to handle.

CYP116B3 with a N-terminal His₆-tag was recombinantly expressed in *E. coli* BL21(DE3) and purified by IMAC. In order to increase the amount of soluble CYP116B3 by reduction of inclusion body formation, His-tag variants of this enzyme were created as the His-tag might influence the folding of the protein. According to the subsequent determination of CYP116B3 concentration in the soluble protein fraction after expression, the amount of soluble CYP116B3 was increased for all new His-tag variants, especially for CYP116B3_Var1, which has a N-terminal MGSS-His₆-sequence. This effect of an increased amount of soluble CYP116B3 can be attributed

to the introduced additional three amino acids (Gly, Ser, Ser) prior to the original N-terminal His₆-tag. An explanation for this effect could be given by the N-end rule,¹⁶⁴ which states that N-terminal amino acids are known to influence a protein's stability in bacteria. Determination of activity of CYP116B3 by fluorescence spectroscopic analysis of the conversion of 7-ethoxycoumarin to 7-hydroxycoumarin and calculation of the product formation rate demonstrated CYP116B3_Var1 to be the most active His-tag variant of CYP116B3, whereas CYP116B3_Var2 (C-terminal His₆-tag) was completely inactive. These results clearly confirm the significant influence of the His-tag on the enzymes properties.

As the His-tag variant CYP116B3_Var1 displayed best expression, was successfully purified and showed highest activity towards the model substrate 7-ethoxycoumarin, it was applied as basis for subsequent investigations concerning the conversion of aromatic substrates of interest aiming at the synthesis of vanillin. In initial CYP116B3-catalyzed *in vitro* biotransformation reactions of the respective aromatic substrates, only for the substrate 3-methylanisole a low total conversion of 2.1% was determined with *m*-cresol as main product and 3-methoxybenzyl alcohol and 4-methoxy-2-methylphenol as byproducts. Hence, though the total conversion was rather low, CYP116B3 did not show an explicit selectivity for any of the possible products. Nevertheless, a slight preference for demethylation over aromatic or aliphatic hydroxylation was detected. This is in accordance with the results reported by Liu and co-workers,¹⁴⁰ where dealkylation is preferred over aromatic hydroxylation and aliphatic side-chain oxidation of aromatic substrates is preferred over aromatic ring hydroxylation. Hence, CYP116B3 is a good candidate for further protein engineering approaches, as a basic level of activity towards the substrate of interest is already present in the wild type enzyme.

As a third and alternative enzyme to the above mentioned ones CYP102A1 from *B. megaterium* was investigated in this work. From previous studies both wild type enzyme and some variants of CYP102A1 were known to convert various aromatic substrates.^{121, 165} Hence the members of a focused CYP102A1 mutant library consisting of the wild type and 24 single and double mutant variants, which were selectively replaced by a set of five hydrophobic amino acids (alanine, leucine, isoleucine, phenylalanine, valine) in the positions F87 and A328 by means of rational protein design in a previous study,¹³² were now screened for their applicability in the

in vitro biotransformation of aromatic substrates concerning the synthesis of vanillin according to the proposed vanillin synthesis cascade reaction (Figure 1.8, introduction).

Only with 3-methylanisole as a substrate, conversion was detected while the other substrates were not converted at all. Thereby the marginal conversion of 3-methylanisole by aromatic hydroxylation catalyzed by the CYP102A1 wild type is in accordance with the results reported in a preceding study.¹⁷¹ Single and double mutant variants, especially those having a valine instead of a phenylalanine at position 87, displayed a much higher substrate conversion compared to the wild type. This is assumed to arise from the smaller size of the side-chain of valine compared to the one of phenylalanine, which allows a better access and positioning of the substrate to the activated heme oxygen in the active site. Such a beneficial effect of the F87V mutation on the hydroxylation of aromatic compounds, which resulted in an improved NADPH consumption rate and coupling efficiency, was described by Sulistyningdyah and co-workers in a previous work.¹⁷² The formation of a variety of side-products accompanying the production of the intended intermediate compounds 3-methoxybenzyl alcohol and 4-methylguaiacol is an indication for a rather inefficient positioning of the substrate in the active site, resulting in a relatively low regio- and chemoselectivity of the most active variants. Consequently, oxidation of the aromatic ring occurred in three of the four possible positions with a preference for the *para* position to the methoxy group. This reaction yielded 4-methoxy-2-methylphenol as main product of the 3-methylanisole conversion. In addition to aromatic and benzylic hydroxylation a demethylation reaction of the methoxy group was detected. Thus, the different types of reaction catalyzed by variants of CYP102A1 clearly demonstrate the catalytic flexibility of this enzyme. Solely oxidation of the aromatic ring in *meta* position to the methoxy group did not occur in the investigated *in vitro* biotransformations. This is in conformity with the results of a study, where among other benzene derivatives the conversion of toluene and anisole was investigated. The reported conversion reaction was catalyzed by CYP102A1 wild type with the assistance of short-alkyl-chain perfluorinated carboxylic acids. These molecules serve as 'dummy' substrates as they have a structural similarity to the natural substrate. Partial binding of these molecules to the substrate binding site of CYP102A1 was reported to initiate the activation of molecular oxygen leading to the generation of compound I, while small non-natural substrate molecules can occupy

the residual space in the active site and thus are efficiently hydroxylated. Here, only the *ortho* and *para* hydroxylated products, with a preference for the *ortho* position, were reported.¹⁷³ Steric properties of the active site rather than C-H bond reactivity were indicated to be important factors for the selectivity of aromatic and benzylic hydroxylation.¹⁷⁴

4.2 Expansion of the spectrum of enzymes and variants for further biotransformations

The focus was now set on CYP116B3 and CYP102A1, as the wild types of both enzymes were already shown to accept 3-methylanisole as a substrate. Optimization of these enzymes was necessary to yield higher substrate conversions with 3-methylanisole, to yield more of the intermediate products of the vanillin synthesis cascade reaction and to expand the substrate profile to the intermediate substrates 3-methoxybenzyl alcohol and 4-methylguaiacol. The investigation of a rationally designed focused mutant library of CYP102A1 was already shown to be an efficient tool of enzyme engineering. Previous experimental data indicated that the wild type enzyme of CYP116B3 displays a higher initial activity towards 3-methylanisole compared to the wild type enzyme of CYP102A1 and thus, a focused mutant library of CYP116B3 was generated. Therefore a homology model of the monooxygenase domain of CYP116B3 was created, in cooperation with the bioinformatics group of the ITB (University of Stuttgart, Stuttgart, Germany) by Dr. Łukasz Gricman, based on structural information by employment of a standard numbering scheme for class I P450 sequences. After generation of the CYP116B3 homology model a focused minimal mutant library of CYP116B3 was created in analogy to the focused mutant library of CYP102A1. Therefore the five amino acid positions A109, V316, A317, A318 and F420 were chosen based on information from literature,¹⁴¹ from the protein sequence and from the protein structure according to the homology model. Each of the selected amino acids was replaced by the hydrophobic amino acids alanine, phenylalanine, isoleucine, leucine and valine, respectively. While, according to a previously generated homology model, the alanine at amino acid position 109 was reported to be positioned directly above the heme group in a small region at the ceiling of the substrate binding pocket,¹⁴¹ according to the newly generated homology model it is positioned in rather big distance to the active site. However, its location on

Discussion

a flexible loop region could explain its influence on the enzymes activity reported by Liu and co-workers after replacement of alanine by phenylalanine.¹⁴¹ Valine at position 316, identified as the fifth amino acid after the highly conserved ExxR motive, was assumed to play an important role in the positioning of the substrate to the activated heme oxygen in the active site, as amino acids at this position were reported to be located close to the heme group and pointing towards the heme center.¹³¹ This anticipated central position of V316 is in agreement with the positioning shown in the homology model of CYP116B3. In addition, the residues A317 and A318, located in direct proximity to V316 and thus in central positions in the active site, as well as F420, positioned at the entrance to the enzymes active site opposite to V316, A317 and A318, are located on loop regions according to the homology model.

Subsequent to the generation of the CYP116B3 mutant library, consisting of the wild type and 20 single mutant variants, by site-directed mutagenesis, all CYP116B3 variants were expressed in *E. coli* BL21(DE3) followed by determination of concentration by CO differential spectral analysis. The activity of the enzyme variants was examined with the 7-ethoxycoumarin assay. Only one out of the 20 CYP116B3 variants was completely inactive while the residual 19 variants were shown to be active towards 7-ethoxycoumarin. Subsequently, the relative activity of the CYP116B3 variants compared to the wild type enzyme was determined. While variant V316F was completely inactive, four of the variants displayed an increased relative activity with V316L as the most active one (almost six times higher activity compared to the wild type enzyme). This clearly indicates the importance of the amino acid position 316. However, most of the amino acid exchanges located in the other addressed positions also resulted in altered enzyme activity. Thereby the activity for the A109F variant was increased almost threefold compared to the wild type which is in agreement with the increase in activity towards 7-ethoxycoumarin reported in a former study.¹⁴¹ Interestingly, the mutations of the alanine and phenylalanine in the positions 318 and 420, respectively, led to a decrease in activity in all investigated cases. We assume the amino acid residues in these positions to be involved in the positioning of the substrate to the active site heme oxygen.

Finally, in order to investigate the hydroxylation activity of the CYP116B3 variants, a screening of the focused CYP116B3 library with the respective aromatic substrates concerning the synthesis of vanillin according to the proposed cascade reaction

Discussion

(Figure 1.8, introduction) was performed. With 3-methylanisole as substrate only three different products were found, namely 3-methoxybenzyl alcohol, 4-methoxy-2-methylphenol and *m*-cresol, while most of the variants showed a preference for the demethylation reaction of the methoxy group yielding *m*-cresol as main product. This could be seen as an indication for a rather strict selectivity of CYP116B3 compared to the results obtained with CYP102A1 variants. On the other hand, selectivity of CYP116B3 might decrease with an increasing activity of the enzyme, which was rather low for all variants. Only for V316L, the most active variant, activity compared to the CYP116B3 wild type was increased about 2.5 times to 5.4%, which is still lower compared to selected variants of the CYP102A1 mutant library. An altered chemoselectivity was detected solely for the variants F420V and F420I, with a shift from demethylation towards aromatic hydroxylation and, for F420I, also towards benzylic hydroxylation as the preferred reaction. This clearly indicates that, though the activity could not be increased dramatically for these variants, the amino acid at position 420 and its size determines the catalytic reaction by influencing the enzymes chemoselectivity. Very interestingly three variants were found to successfully convert 4-methylguaiacol to vanillyl alcohol by benzylic hydroxylation of the methyl group, though to a very low extent. These are again the variants V316L, F420I, and F420V mentioned above. Hence, the generation of a focused mutant library of CYP116B3 yielded interesting results for the conversion of aromatic substrates concerning the intended vanillin synthesis reaction.

Regarding CYP102A1, in a previous study additional mutations in position 437 were shown to be of benefit for the conversion of (4*R*)-limonene to perillyl alcohol. This was believed to be induced by the strong influence of the respective amino acid at position 437 on the orientation of the substrate to the heme oxygen in the active site.¹⁶⁸ For this reason, triple mutant variants based on F87V/A328L, exhibiting the highest total conversion of 3-methylanisole in initial biotransformations, combined with an additional amino acid exchange of leucine at position 437 in the wild type to other hydrophobic amino acids (alanine, phenylalanine, isoleucine and valine) were investigated. Hydrophobic amino acids were chosen, as a preference for those in substrate-interacting positions of P450s was reported.¹³¹ After expression, soluble protein fractions of the triple mutant variants were screened for *in vitro* biotransformation of aromatic substrates towards the intended cascade synthesis of

Discussion

vanillin. All variants were found to be active exclusively towards 3-methylanisole while enzymatic activity differed strongly from a dramatic decrease to an almost equal value compared to the activity of the double mutant F87V/A328L. Additionally, a shift in product distribution was determined as the result of an altered chemoselectivity of the triple mutant variants towards a higher preference for benzylic hydroxylation, yielding twice as much 3-methoxybenzyl alcohol as product as before. The altered selectivity is presumed to be a result of the expected change in the orientation of the substrate to the activated heme oxygen.

Due to the fact that 3-methylanisole was successfully converted in *in vitro* biotransformations with CYP102A1 variants, but neither 3-methoxybenzyl alcohol nor 4-methylguaiacol was converted to vanillyl alcohol or even accepted as a substrate for any kind of reaction, this second reaction step of the intended cascade reaction was a bottleneck towards the synthesis of vanillin. However, in principle, both benzylic hydroxylation of the methyl group as well as aromatic hydroxylation in *para* position to the methyl group was detected, with a preference for benzylic hydroxylation, as shown with 3-methylanisole as substrate. Therefore, in order to gain more information about this bottleneck of the cascade reaction, MD simulations were carried out by Dr. Alexander Seifert from the bioinformatics group of the ITB (University of Stuttgart, Stuttgart, Germany). As the route via benzylic hydroxylation of 4-methylguaiacol was thought to be more promising due to the above mentioned preference for benzylic hydroxylation of the CYP102A1 variants, the MD simulations were performed with 4-methylguaiacol as substrate molecule in combination with a CYP102A1 double mutant variant to examine if the substrate binds in the enzymes active site in a productive manner. According to the results obtained by the MD simulations this seems feasible. However, in some of the MD simulations the amino acids R47 and Y51, located in the substrate access channel in a more distant pocket to the active site, were identified to stabilize the 4-methylguaiacol substrate molecule via hydrogen bonds. This could be an explanation for the inactivity of the CYP102A1 variants towards conversion of 4-methylguaiacol, as this molecule is stabilized in the substrate channel impeding access of further substrate molecules to the active site. R47 and Y51 have already been reported in previous studies, to be key residues positioned at the entrance of the substrate channel.¹²¹ Substitution of R47 and Y51 by hydrophobic residues has been shown to result in an increased activity towards various substrates. For example, a six-fold increased activity in the epoxidation of

Discussion

geranylacetone was measured.¹⁷⁵ Hence the polar substrate channel residues R47 and Y51 were replaced by the hydrophobic residues leucine and phenylalanine by means of site-directed mutagenesis in selected single, double and triple mutant variants of CYP102A1, which displayed the highest activity with 3-methylanisole in previous investigations, yielding R47L/Y51F-variants, respectively. Screening of these variants for conversion of 4-methylguaiacol as well as other aromatic (intermediate) substrates of the intended cascade synthesis of vanillin revealed that, as expected, 4-methylguaiacol was successfully converted to vanillyl alcohol by most of the investigated variants. This clearly confirms the successful elimination of 4-methylguaiacol stabilization in the substrate access channel in distance to the active site. As a consequence the substrate spectrum of the respective CYP102A1 variants was expanded.

To address conversion of vanillyl alcohol to vanillin in a cascade, the portfolio of investigated enzymes was expanded to a vanillyl alcohol oxidase from *P. simplicissimum*. This enzyme was reported in literature not only to convert vanillyl alcohol to vanillin with high selectivity but also to catalyze the conversion of 4-methylguaiacol to vanillin via vanillyl alcohol as intermediate product.^{72, 169} Consequently VAO seemed to be an ideal candidate for application in cascade synthesis of vanillin reported in this work. As not only the wild type enzyme but also some variants with intensively increased catalytic efficiency (k_{cat}/K_M) towards 4-methylguaiacol have been reported in literature,⁶⁹ the single mutant variants F454Y and T505S were generated by site-directed mutagenesis based on the VAO wild type gene (*vaoA*) which was employed as a codon-optimized gene for the heterologous expression in *E. coli*. Though the majority of the protein seemed to be present in the insoluble protein fraction after cell lysis, activity of the wild type enzyme and the single mutant variants F454Y and T505S was successfully tested with soluble protein fractions in *in vitro* biotransformations of both vanillyl alcohol and 4-methylguaiacol as substrates. Thereby variant F454Y displayed highest activity towards 4-methylguaiacol at pH 7.5 yielding exclusively vanillin as product. As vanillyl alcohol was even completely converted to vanillin without any byproduct or overoxidation product formation, this demonstrates not only the activity but also the high selectivity of the investigated VAO both for the wild type and for the variants thereof. In addition, none of the investigated aromatic compounds, which were previously reported to be

yielded as byproducts in the conversion of 3-methylanisole by CYP102A1 variants, was accepted as a substrate. All these results indicate that the VAO variant F454Y could be an ideal candidate for application in one-pot cascade reactions for the synthesis of vanillin.

4.3 *In vitro* one-pot cascade synthesis of vanillin

In order to realize the proposed multi-enzymatic three-step cascade reaction for the synthesis of vanillin (Figure 1.8, introduction) a variety of enzymes and variants thereof was investigated previously. Consequently, the best enzyme variants were combined in *in vitro* one-pot cascade reactions for the conversion of 3-methylanisole. Thus, in a first reaction setup the CYP102A1 variants A328L and R47L/Y51F/F87V/A328V catalyzing the first and second step of the cascade reaction, respectively, were combined. Though the successive time-shifted addition of R47L/Y51F/F87V/A328V after half of the reaction time, instead of direct addition of both enzymes from the beginning, was beneficial concerning total substrate conversion (increase from 24.7% to 30.3%) and formation of 4-methylguaiacol (increase from 1.2% to 2.4%), 4-methoxy-2-methylphenol was found as main product and no production of vanillyl alcohol was measured. This is assumed to be a result of the low amount of 4-methylguaiacol that is produced by A328L, in combination with the competitive conversion of 3-methylanisole and 4-methylguaiacol by the second enzyme, with a clear preference for 3-methylanisole conversion. The addition of catalase and superoxide dismutase was investigated in order to reduce the concentration of reactive oxygen species like hydrogen peroxide or superoxide, which could have arisen from inefficient coupling in CYP102A1 and might possibly negatively influence the enzymes activity.

As an alternative, in a second reaction setup the CYP102A1 quadruple variant R47L/Y51F/F87V/A328V was replaced by the VAO variant F454Y in order to directly convert intermediately produced 4-methylguaiacol via vanillyl alcohol to vanillin. With 22.9% total conversion of 3-methylanisole was comparable to the first reaction setup after 2 h and could be increased by successive addition of the VAO variant F454Y after half of the reaction time to 29.0%, too. However, with this setup the formation of 1.2% vanillin was detected, accompanied by a decreased amount of 4-methylguaiacol in the mixture, while vanillyl alcohol was not detected at all due to immediate

and complete conversion to vanillin. These results definitely confirm the realization of the intended *in vitro* one-pot synthesis of vanillin in a multi-enzymatic three-step cascade reaction following the proposed route from 3-methylanisole via 4-methylguaiacol and vanillyl alcohol to vanillin (Figure 1.8, introduction). Highest vanillin formation of 2% was determined in a reaction setup with successive addition of both enzymes and a total conversion of 31.4% after a total reaction time of 8 h.

4.4 *In vivo* one-pot cascade synthesis of vanillin

To demonstrate the proof-of-principle, *E. coli* BL21(DE3) was chosen as host for expression of CYP102A1 variant A328L and VAO variant F454Y using a two-plasmid system and for subsequent whole-cell substrate conversion. Success of the expression of both enzymes was confirmed by SDS-PAGE analysis showing bands at expected molecular weights. Finally, harvested whole cells were applied in a biotransformation medium for *in vivo* conversion of 3-methylanisole. 4-Methoxy-2-methylphenol was the first product coming up and persisted over the whole period of reaction as the main product detected in the samples. This product formation can surely be attributed to the selectivity of the CYP102A1 variant A328L. In contrast, the formation of 4-methylguaiacol increased at the beginning to a low level followed by a decrease until it completely disappeared while the formation of vanillin steadily increased up to a final concentration of 1.1%. Surely the reaction performed by CYP102A1 variant A328L is the limiting step of the cascade reaction. However, other effects like substrate delivery into and product release from the host cell, substrate and product inhibition and/or toxicity, oxygen supply, cofactor availability and regeneration, stability of the P450 enzyme, host metabolic effects, cell viability/host cell death as well as other unknown factors could have had an influence on total conversion, concomitantly on product formation and thus finally on the synthesis of vanillin. Though, compared to the *in vitro* one-pot results, the final product formation of vanillin was lower in the *in vivo* reactions the outcome of the investigations was basically the same.

In summary, as a proof-of-principle, the valuable aromatic flavor compound vanillin was synthesized in a novel *in vivo* two-enzymatic three-step cascade system from the simple and low-cost chemical starting material 3-methylanisole. Though further optimizations are inevitable, the newly designed biocatalytic cascade for the

Discussion

exemplary synthesis of vanillin, realized within this project, demonstrates the capability of biocatalysis.

5 Conclusion and outlook

The selective hydroxylation of aromatic compounds, which is one of the most challenging chemical reactions, was addressed enzymatically within the present project. Therefore as an example of consecutive enzyme-catalyzed oxidation reactions, a multi-enzymatic three-step cascade reaction starting with the simple low-cost aromatic substrate 3-methylanisole was investigated as a novel route to the valuable aromatic compound vanillin.

In a first step a defined set of enzymes had to be selected for initial investigations. In this context, the *m*-hydroxybenzoate hydroxylase MobA from *C. testosteroni* GZ39 and the cytochrome P450 monooxygenases CYP116B3 from *R. ruber* DSM 44319 and CYP102A1 from *B. megaterium* ATCC 14581 as well as variants thereof were chosen as promising enzyme candidates. After successful recombinant expression in *E. coli*, purification of some of the enzymes and determination of activity, the enzymes were applied in *in vitro* biotransformation reactions, combined with efficient cofactor recycling, for the selective oxidation of both aromatic substrate and intermediate compounds of the pursued vanillin synthesis pathway. In addition to the wild type enzymes some variants thereof were created by means of rational protein design and examined in *in vitro* reactions, too. While MobA was not suitable for the intended purpose as it did not show activity with any of the investigated aromatic compounds of interest, CYP116B3 displayed a basic level of activity already with the wild type enzyme. Moreover, by investigation of a focused mutant library of CYP102A1, several single and double mutant variants were identified with a higher activity towards the substrate 3-methylanisole compared to the wild type. Hence, for the synthesis of both target intermediate products, namely 3-methoxybenzyl alcohol and 4-methylguaiacol, enzyme candidates were found, though the byproduct 4-methoxy-2-methylphenol was produced as main product in most of the conversion reactions.

In order to improve both activity and selectivity of the biocatalysts to gain higher yields of intermediate products and subsequent conversion of those to vanillyl alcohol as precursor of vanillin, further enzyme variants of CYP102A1 and CYP116B3 were created by rational protein design. For this reason, for CYP116B3 a focused mutant library based on literature, sequence and structure (homology model) information was generated in analogy to the focused CYP102A1 mutant library. However, none

Conclusion and outlook

of the variants showed a dramatic increase in either activity or selectivity towards the synthesis of the intermediate products 3-methoxybenzyl alcohol and 4-methylguaiacol, nor did any of the new variants display significant conversion of those intermediate compounds to vanillyl alcohol.

Specific and more detailed investigation of the second step of the intended cascade reaction by molecular dynamics simulations, with a defined focus on the conversion of 4-methylguaiacol to vanillyl alcohol, yielded improved variants of CYP102A1, that were successfully applied in respective *in vitro* biotransformation reactions. In addition, rationally designed variants of a vanillyl alcohol oxidase from *P. simplicissimum* were proven to be highly active towards 4-methylguaiacol, accompanied by a strict selectivity yielding exclusively vanillin as final product of interest, by catalysis of both the conversion of 4-methylguaiacol to vanillyl alcohol and the consecutive conversion of vanillyl alcohol to vanillin.

Consequently enzymes and variants thereof were identified for each of the necessary reaction steps of the intended cascade synthesis of vanillin. The most suitable ones, CYP102A1 variant A328L and VAO variant F454Y, were combined in one reaction system both *in vitro* in one-pot biotransformation reactions and *in vivo* in biotransformations with whole cells of *E. coli* as expression and production host. As a proof-of-principle, vanillin was successfully produced in both systems. However, the yield of vanillin was rather low due to the bottleneck of the reaction, which is the sufficient production of the intermediate compound 4-methylguaiacol in the first step of the cascade reaction, to supply the highly active and selective VAO variant with enough substrate.

In order to optimize the cascade, alternative enzymes, which could be better applicable for the addressed task, could be searched for by proteome analysis after screening of a diversity of strains for growth on minimal media containing the substrate of interest as sole carbon source. Moreover, as an alternative approach to rational protein design, a semi-rational approach like iterative saturation mutagenesis or combinatorial active site saturation testing (CASTing) in the identified key residues might be applied, or a random enzyme engineering method like directed evolution could help to find better variants of the chosen enzymes, though in this case, an efficient high throughput screening system would be necessary. These approaches could help to overcome the bottleneck reaction of the established cascade, the

Conclusion and outlook

conversion of 3-methylanisole to 4-methylguaiacol. In addition, enzymes and variants which are able to convert the intermediate compound 3-methoxybenzyl alcohol to vanillyl alcohol could be identified and thus increase the substrate concentration for the VAO-catalyzed reaction step. Furthermore, usage of respectively diverse plasmids to enable selective induction of expression could be beneficial for the regulation of the expression of the different enzymes in the whole-cell systems. Regarding the production of vanillin, identification of a biological system for the synthesis of the substrate 3-methylanisole would be economically advantageous, as the produced vanillin could then be designated as 'natural' vanillin. Finally, the application of a multi-enzymatic cascade reaction approach, as described in this work, could be transferred to alternative substrates and products of interest to illustrate the versatility and capability of biocatalysis.

6 References

1. Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K., Engineering the third wave of biocatalysis. *Nature* 2012, **485**, (7397), 185-94.
2. Moore, J. C.; Bornscheuer, U. T., Editorial overview: biocatalysis and biotransformation: riding the third wave of biocatalysis. *Curr Opin Chem Biol* 2014, **19**, v-vi.
3. Ro, D. K.; Paradise, E. M.; Ouellet, M.; Fisher, K. J.; Newman, K. L.; Ndungu, J. M.; Ho, K. A.; Eachus, R. A.; Ham, T. S.; Kirby, J.; Chang, M. C.; Withers, S. T.; Shiba, Y.; Sarpong, R.; Keasling, J. D., Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 2006, **440**, (7086), 940-3.
4. Muheim, A.; Müller, B.; Münch, T.; Wetli, M., Microbiological process for producing vanillin. *Patent US/2001/6235507 from Givaudan Roure (International) Sa* 2001.
5. Yuryev, R.; Liese, A., Biocatalysis: The Outcast. *ChemCatChem* 2010, **2**, (1), 103-107.
6. Sarett, L. H., Partial synthesis of pregnene-4-triol-17(β), 20(β), 21-dione-3,11 and pregnene-4-diol-17(β), 21-trione-3,11,20 monoacetate. *J. Biol. Chem.* 1946, **162**, 601-31.
7. Sebek, O. K.; Perlman, D., *Microbial Technology, 2nd ed., Vol. 1 (Eds.: H. J. Peppler, D. Perlman)*, Academic Press, New York 1979, 483-96.
8. Rothenberg, G., *Catalysis: Concepts and Green Applications*, Wiley-VCH, Weinheim 2008.
9. Lewis, J. C.; Coelho, P. S.; Arnold, F. H., Enzymatic functionalization of carbon-hydrogen bonds. *Chem Soc Rev* 2011, **40**, (4), 2003-21.
10. Roduner, E.; Kaim, W.; Sarkar, B.; Urlacher, V. B.; Pleiss, J.; Gläser, R.; Einicke, W.-D.; Sprenger, G. A.; Beifuß, U.; Klemm, E.; Liebner, C.; Hieronymus, H.; Hsu, S.-F.; Plietker, B.; Laschat, S., Selective Catalytic Oxidation of C-H Bonds with Molecular Oxygen. *ChemCatChem* 2013, **5**, (1), 82-112.

References

11. Crabtree, R. H., Alkane C–H activation and functionalization with homogeneous transition metal catalysts: a century of progress—a new millennium in prospect. *J. Chem. Soc., Dalton Trans.* 2001, (17), 2437-50.
12. Bergman, R. G., Organometallic chemistry: C-H activation. *Nature* 2007, **446**, (7134), 391-3.
13. Haber, J.; Mlodnicka, T., *J. Mol. Catal.* 1992, **74**, 131-41.
14. Balcells, D.; Clot, E.; Eisenstein, O., C-H bond activation in transition metal species from a computational perspective. *Chem Rev* 2010, **110**, (2), 749-823.
15. Goldman, A. S.; Goldberg, K. I., Activation and Functionalization of C-H Bonds. *ACS Symposium Series 885, ACS, Washington, DC* 2004, 1-45.
16. Labinger, J. A.; Bercaw, J. E., Understanding and exploiting C-H bond activation. *Nature* 2002, **417**, (6888), 507-14.
17. Lewis, J. C.; Bergman, R. G.; Ellman, J. A., Direct functionalization of nitrogen heterocycles via Rh-catalyzed C-H bond activation. *Acc Chem Res* 2008, **41**, (8), 1013-25.
18. Ragsdale, S. W., Metals and their scaffolds to promote difficult enzymatic reactions. *Chem Rev* 2006, **106**, (8), 3317-37.
19. Shilov, A. E.; Shul'pin, G. B., Activation of C-H Bonds by Metal Complexes. *Chem Rev* 1997, **97**, (8), 2879-2932.
20. Lippard, S. J., Hydroxylation of C-H bonds at carboxylate-bridged diiron centres. *Philos Trans A Math Phys Eng Sci* 2005, **363**, (1829), 861-77; discussion 1035-40.
21. Leahy, J. G.; Batchelor, P. J.; Morcomb, S. M., Evolution of the soluble diiron monooxygenases. *FEMS Microbiol Rev* 2003, **27**, (4), 449-79.
22. Mitchell, K. H.; Studts, J. M.; Fox, B. G., Combined participation of hydroxylase active site residues and effector protein binding in a *para* to *ortho* modulation of toluene 4-monooxygenase regioselectivity. *Biochemistry* 2002, **41**, (9), 3176-88.
23. Wubbolts, M. G.; Reuvekamp, P.; Witholt, B., TOL plasmid-specified xylene oxygenase is a wide substrate range monooxygenase capable of olefin epoxidation. *Enzyme Microb Technol* 1994, **16**, (7), 608-15.
24. van Berkel, W. J. H.; Kamerbeek, N. M.; Fraaije, M. W., Flavoprotein monooxygenases, a diverse class of oxidative biocatalysts. *J Biotechnol* 2006, **124**, (4), 670-89.

References

25. Abood, A.; Al-Fahad, A.; Scott, A.; Hosny, A. E.-D. M. S.; Hashem, A. M.; Fattah, A. M. A.; Paul R. Race; Simpson, T. J.; Cox, R. J., Kinetic characterisation of the FAD dependent monooxygenase TropB and investigation of its biotransformation potential. *RSC Adv.* 2015, **5**, (62), 49987-95.
26. Torres Pazmino, D. E.; Winkler, M.; Glieder, A.; Fraaije, M. W., Monooxygenases as biocatalysts: Classification, mechanistic aspects and biotechnological applications. *J Biotechnol* 2010, **146**, (1-2), 9-24.
27. Massey, V., The chemical and biological versatility of riboflavin. *Biochem Soc Trans* 2000, **28**, (4), 283-96.
28. Entsch, B.; van Berkel, W. J. H., Structure and mechanism of *para*-hydroxybenzoate hydroxylase. *Faseb J* 1995, **9**, (7), 476-83.
29. Montersino, S.; Tischler, D.; Gassner, G. T.; van Berkel, W. J. H., Catalytic and structural features of flavoprotein hydroxylases and epoxidases. *Adv. Synth. Catal.* 2011, **353**, (13), 2301-19.
30. Sono, M.; Roach, M. P.; Coulter, E. D.; Dawson, J. H., Heme-containing oxygenases. *Chem Rev* 1996, **96**, (7), 2841-88.
31. Green, M. T., C-H bond activation in heme proteins: the role of thiolate ligation in cytochrome P450. *Curr Opin Chem Biol* 2009, **13**, (1), 84-8.
32. Harris, D. L., High-valent intermediates of heme proteins and model compounds. *Curr Opin Chem Biol* 2001, **5**, (6), 724-35.
33. Groves, J. T., The bioinorganic chemistry of iron in oxygenases and supramolecular assemblies. *Proc Natl Acad Sci U S A* 2003, **100**, (7), 3569-74.
34. Julsing, M. K.; Cornelissen, S.; Buhler, B.; Schmid, A., Heme-iron oxygenases: powerful industrial biocatalysts? *Curr Opin Chem Biol* 2008, **12**, (2), 177-86.
35. Guengerich, F. P., Cytochrome P450 enzymes in the generation of commercial products. *Nat Rev Drug Discov* 2002, **1**, (5), 359-66.
36. van Beilen, J. B.; Duetz, W. A.; Schmid, A.; Witholt, B., Practical issues in the application of oxygenases. *Trends Biotechnol* 2003, **21**, (4), 170-7.
37. Chefson, A.; Auclair, K., Progress towards the easier use of P450 enzymes. *Mol Biosyst* 2006, **2**, (10), 462-9.

References

38. Eiben, S.; Kaysser, L.; Maurer, S.; Kuhnel, K.; Urlacher, V. B.; Schmid, R. D., Preparative use of isolated CYP102 monooxygenases - a critical appraisal. *J Biotechnol* 2006, **124**, (4), 662-9.
39. Urlacher, V. B.; Lutz-Wahl, S.; Schmid, R. D., Microbial P450 enzymes in biotechnology. *Appl Microbiol Biotechnol* 2004, **64**, (3), 317-25.
40. Savile, C. K.; Janey, J. M.; Mundorff, E. C.; Moore, J. C.; Tam, S.; Jarvis, W. R.; Colbeck, J. C.; Krebber, A.; Fleitz, F. J.; Brands, J.; Devine, P. N.; Huisman, G. W.; Hughes, G. J., Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* 2010, **329**, (5989), 305-9.
41. Lundemo, M. T.; Notonier, S.; Striedner, G.; Hauer, B.; Woodley, J. M., Process limitations of a whole-cell P450 catalyzed reaction using a CYP153A-CPR fusion construct expressed in *Escherichia coli*. *Appl Microbiol Biotechnol* 2016, **100**, (3), 1197–1208.
42. Hoffmann, S. M.; Weissenborn, M. J.; Gricman, Ł.; Notonier, S.; Pleiss, J.; Hauer, B., The Impact of Linker Length on P450 Fusion Constructs: Activity, Stability and Coupling. *ChemCatChem* 2016, **8**, (8), 1591–97.
43. Gricman, Ł.; Weissenborn, M. J.; Hoffmann, S. M.; Borlinghaus, N.; Hauer, B.; Pleiss, J., Redox Partner Interaction Sites in Cytochrome P450 Monooxygenases: *In Silico* Analysis and Experimental Validation. *ChemistrySelect* 2016, **1**, (6), 1243–51.
44. Gally, C.; Nestl, B. M.; Hauer, B., Engineering Rieske Non-Heme Iron Oxygenases for the Asymmetric Dihydroxylation of Alkenes. *Angew Chem Int Ed Engl* 2015, **54**, (44), 12952-6.
45. Sigdel, S.; Hui, G.; Smith, T. J.; Murrell, J. C.; Lee, J. K., Molecular dynamics simulation to rationalize regioselective hydroxylation of aromatic substrates by soluble methane monooxygenase. *Bioorg Med Chem Lett* 2015, **25**, (7), 1611-5.
46. Joosten, V.; van Berkel, W. J., Flavoenzymes. *Curr Opin Chem Biol* 2007, **11**, (2), 195-202.
47. Mewies, M.; McIntire, W. S.; Scrutton, N. S., Covalent attachment of flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) to enzymes: the current state of affairs. *Protein Sci* 1998, **7**, (1), 7-20.

References

48. Hefti, M. H.; Vervoort, J.; van Berkel, W. J., Deflavination and reconstitution of flavoproteins. *Eur J Biochem* 2003, **270**, (21), 4227-42.
49. Moonen, M. J. H.; Fraaije, M. W.; Rietjens, I. M. C. M.; Laane, C.; van Berkel, W. J. H., Flavoenzyme-Catalyzed Oxygenations and Oxidations of Phenolic Compounds. *Adv. Synth. Catal.* 2002, **344**, (10), 1023-35.
50. Ballou, D. P.; Entsch, B.; Cole, L. J., Dynamics involved in catalysis by single-component and two-component flavin-dependent aromatic hydroxylases. *Biochem Biophys Res Commun* 2005, **338**, (1), 590-8.
51. Hiromoto, T.; Fujiwara, S.; Hosokawa, K.; Yamaguchi, H., Crystal structure of 3-hydroxybenzoate hydroxylase from *Comamonas testosteroni* has a large tunnel for substrate and oxygen access to the active site. *J Mol Biol* 2006, **364**, (5), 878-96.
52. Fraaije, M. W.; Kamerbeek, N. M.; van Berkel, W. J. H.; Janssen, D. B., Identification of a Baeyer-Villiger monooxygenase sequence motif. *FEBS Lett* 2002, **518**, (1-3), 43-7.
53. Chaiyen, P.; Suadee, C.; Wilairat, P., A novel two-protein component flavoprotein hydroxylase. *Eur J Biochem* 2001, **268**, (21), 5550-61.
54. Fotheringham, I.; Archer, I.; Carr, R.; Speight, R.; Turner, N. J., Preparative deracemization of unnatural amino acids. *Biochem Soc Trans* 2006, **34**, (Pt 2), 287-90.
55. Massey, V., Activation of molecular oxygen by flavins and flavoproteins. *J Biol Chem* 1994, **269**, (36), 22459-62.
56. Fraaije, M. W.; van den Heuvel, R. H.; van Berkel, W. J. H.; Mattevi, A., Covalent flavinylation is essential for efficient redox catalysis in vanillyl-alcohol oxidase. *J Biol Chem* 1999, **274**, (50), 35514-20.
57. Chen, R.; Chaen, H.; Hosokawa, K., Studies on *m*-hydroxybenzoate 4-hydroxylase from *Comamonas testosteroni*. I. Purification and characterization. *Res. Commun. Biochem. Cell Mol. Biol.* 1997, **1**, (4), 304-22.
58. Chen, R.; Oki, H.; Scott, R. P., Jr.; Yamaguchi, H.; Kusunoki, M.; Matsuura, Y.; Chaen, H.; Tsugita, A.; Hosokawa, K., Crystallization and further characterization of *meta*-hydroxybenzoate 4-hydroxylase from *Comamonas testosteroni*. *Res. Commun. Biochem. Cell Mol. Biol.* 1998, **2**, (3&4), 253-74.

References

59. Entsch, B.; Cole, L. J.; Ballou, D. P., Protein dynamics and electrostatics in the function of *p*-hydroxybenzoate hydroxylase. *Arch Biochem Biophys* 2005, **433**, (1), 297-311.
60. Chang, H. K.; Zylstra, G. J., Examination and expansion of the substrate range of *m*-hydroxybenzoate hydroxylase. *Biochem Biophys Res Commun* 2008, **371**, (1), 149-53.
61. Fraaije, M. W.; Van Berkel, W. J.; Benen, J. A.; Visser, J.; Mattevi, A., A novel oxidoreductase family sharing a conserved FAD-binding domain. *Trends Biochem Sci* 1998, **23**, (6), 206-7.
62. Huang, C. H.; Lai, W. L.; Lee, M. H.; Chen, C. J.; Vasella, A.; Tsai, Y. C.; Liaw, S. H., Crystal structure of glucooligosaccharide oxidase from *Acremonium strictum*: a novel flavinylation of 6-S-cysteinyl, 8 α -N1-histidyl FAD. *J Biol Chem* 2005, **280**, (46), 38831-8.
63. Brandsch, R.; Bichler, V., Autoflavinylation of apo6-hydroxy-D-nicotine oxidase. *J Biol Chem* 1991, **266**, (28), 19056-62.
64. Leferink, N. G.; Heuts, D. P.; Fraaije, M. W.; van Berkel, W. J., The growing VAO flavoprotein family. *Arch Biochem Biophys* 2008, **474**, (2), 292-301.
65. de Jong, E.; van Berkel, W. J.; van der Zwan, R. P.; de Bont, J. A., Purification and characterization of vanillyl-alcohol oxidase from *Penicillium simplicissimum*. A novel aromatic alcohol oxidase containing covalently bound FAD. *Eur J Biochem* 1992, **208**, (3), 651-7.
66. Mattevi, A.; Fraaije, M. W.; Mozzarelli, A.; Olivi, L.; Coda, A.; van Berkel, W. J., Crystal structures and inhibitor binding in the octameric flavoenzyme vanillyl-alcohol oxidase: the shape of the active-site cavity controls substrate specificity. *Structure* 1997, **5**, (7), 907-20.
67. Fraaije, M. W.; Veeger, C.; van Berkel, W. J. H., Substrate specificity of flavin-dependent vanillyl-alcohol oxidase from *Penicillium simplicissimum*. Evidence for the production of 4-hydroxycinnamyl alcohols from 4-allylphenols. *Eur J Biochem* 1995, **234**, (1), 271-7.
68. van den Heuvel, R. H. H.; Fraaije, M. W.; Laane, C.; van Berkel, W. J., Regio- and stereospecific conversion of 4-alkylphenols by the covalent flavoprotein vanillyl-alcohol oxidase. *J Bacteriol* 1998, **180**, (21), 5646-51.

References

69. van den Heuvel, R. H. H.; van den Berg, W. A.; Rovida, S.; van Berkel, W. J., Laboratory-evolved vanillyl-alcohol oxidase produces natural vanillin. *J Biol Chem* 2004, **279**, (32), 33492-500.
70. Drijfhout, F. P.; Fraaije, M. W.; Jongejan, H.; van Berkel, W. J.; Franssen, M. C., Enantioselective hydroxylation of 4-alkylphenols by vanillyl alcohol oxidase. *Biotechnol Bioeng* 1998, **59**, (2), 171-7.
71. van den Heuvel, R. H. H.; Fraaije, M. W.; van Berkel, W. J., Direction of the reactivity of vanillyl-alcohol oxidase with 4-alkylphenols. *FEBS Lett* 2000, **481**, (2), 109-12.
72. van den Heuvel, R. H. H.; Fraaije, M. W.; Mattevi, A.; Laane, C.; van Berkel, W. J. H., Vanillyl-alcohol oxidase, a tasteful biocatalyst. *Journal of Molecular Catalysis B: Enzymatic* 2001, **11**, (4), 185-8.
73. Fraaije, M. W.; van Berkel, W. J., Catalytic mechanism of the oxidative demethylation of 4-(methoxymethyl)phenol by vanillyl-alcohol oxidase. Evidence for formation of a *p*-quinone methide intermediate. *J Biol Chem* 1997, **272**, (29), 18111-6.
74. Fraaije, M. W.; van Den Heuvel, R. H.; van Berkel, W. J.; Mattevi, A., Structural analysis of flavinylation in vanillyl-alcohol oxidase. *J Biol Chem* 2000, **275**, (49), 38654-8.
75. van den Heuvel, R. H. H.; Fraaije, M. W.; Ferrer, M.; Mattevi, A.; van Berkel, W. J., Inversion of stereospecificity of vanillyl-alcohol oxidase. *Proc Natl Acad Sci U S A* 2000, **97**, (17), 9455-60.
76. Nelson, D. R., Cytochrome P450 and the individuality of species. *Arch Biochem Biophys* 1999, **369**, (1), 1-10.
77. Omura, T.; Sato, R., The Carbon Monoxide-Binding Pigment of Liver Microsomes. II. Solubilization, Purification, and Properties. *J Biol Chem* 1964, **239**, 2379-85.
78. Nelson, D. R.; Kamataki, T.; Waxman, D. J.; Guengerich, F. P.; Estabrook, R. W.; Feyereisen, R.; Gonzalez, F. J.; Coon, M. J.; Gunsalus, I. C.; Gotoh, O.; Okuda, K.; Nebert, D. W., The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA Cell Biol* 1993, **12**, (1), 1-51.
79. Nebert, D. W.; Nelson, D. R.; Adesnik, M.; Coon, M. J.; Estabrook, R. W.; Gonzalez, F. J.; Guengerich, F. P.; Gunsalus, I. C.; Johnson, E. F.; Kemper,

References

- B.; *et al.*, The P450 superfamily: updated listing of all genes and recommended nomenclature for the chromosomal loci. *DNA* 1989, **8**, (1), 1-13.
80. Nebert, D. W.; Adesnik, M.; Coon, M. J.; Estabrook, R. W.; Gonzalez, F. J.; Guengerich, F. P.; Gunsalus, I. C.; Johnson, E. F.; Kemper, B.; Levin, W.; *et al.*, The P450 gene superfamily: recommended nomenclature. *DNA* 1987, **6**, (1), 1-11.
81. Nelson, D. R., The cytochrome p450 homepage. *Hum Genomics* 2009, **4**, (1), 59-65.
82. Nelson, D. R., Mining databases for cytochrome P450 genes. *Methods Enzymol* 2002, **357**, 3-15.
83. <https://cyped.biocatnet.de>.
84. Nelson, D. R., Progress in tracing the evolutionary paths of cytochrome P450. *Biochim Biophys Acta* 2011, **1814**, (1), 14-8.
85. Guengerich, F. P., Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chem Res Toxicol* 2001, **14**, (6), 611-50.
86. Bernhardt, R., Cytochromes P450 as versatile biocatalysts. *J Biotechnol* 2006, **124**, (1), 128-45.
87. Momoi, K.; Hofmann, U.; Schmid, R. D.; Urlacher, V. B., Reconstitution of beta-carotene hydroxylase activity of thermostable CYP175A1 monooxygenase. *Biochem Biophys Res Commun* 2006, **339**, (1), 331-6.
88. Cryle, M. J.; Stok, J. E.; De Voss, J. J., Reactions catalyzed by bacterial cytochromes P450. *Australian Journal of Chemistry* 2003, **56**, (8), 749-62.
89. Guengerich, F. P., Mechanisms of cytochrome P450 substrate oxidation: MiniReview. *J Biochem Mol Toxicol* 2007, **21**, (4), 163-8.
90. Urlacher, V. B.; Eiben, S., Cytochrome P450 monooxygenases: perspectives for synthetic application. *Trends Biotechnol* 2006, **24**, (7), 324-30.
91. Arnold, F. H., The nature of chemical innovation: new enzymes by evolution. *Q Rev Biophys* 2015, **48**, (4), 404-10.
92. McIntosh, J. A.; Heel, T.; Buller, A. R.; Chio, L.; Arnold, F. H., Structural Adaptability Facilitates Histidine Heme Ligation in a Cytochrome P450. *J Am Chem Soc* 2015, **137**, (43), 13861-5.

References

93. Farwell, C. C.; Zhang, R. K.; McIntosh, J. A.; Hyster, T. K.; Arnold, F. H., Enantioselective Enzyme-Catalyzed Aziridination Enabled by Active-Site Evolution of a Cytochrome P450. *ACS Cent Sci* 2015, **1**, (2), 89-93.
94. Furuya, T.; Shitashima, Y.; Kino, K., Alteration of the substrate specificity of cytochrome P450 CYP199A2 by site-directed mutagenesis. *J Biosci Bioeng* 2015, **119**, (1), 47-51.
95. Weissenborn, M. J.; Löw, S. A.; Borlinghaus, N.; Kuhn, M.; Kummer, S.; Rami, F.; Plietker, B.; Hauer, B., Enzyme-Catalyzed Carbonyl Olefination by the *E. coli* Protein YfeX in the Absence of Phosphines. *ChemCatChem* 2016, **8**, (9), 1636–1640.
96. Werck-Reichhart, D.; Feyereisen, R., Cytochromes P450: a success story. *Genome Biol* 2000, **1**, (6), REVIEWS3003.
97. Fjaervik, E.; Zotchev, S. B., Biosynthesis of the polyene macrolide antibiotic nystatin in *Streptomyces noursei*. *Appl Microbiol Biotechnol* 2005, **67**, (4), 436-43.
98. Mendes, M. V.; Anton, N.; Martin, J. F.; Aparicio, J. F., Characterization of the polyene macrolide P450 epoxidase from *Streptomyces natalensis* that converts de-epoxypimaricin into pimaricin. *Biochem J* 2005, **386**, (Pt 1), 57-62.
99. Pylypenko, O.; Schlichting, I., Structural aspects of ligand binding to and electron transfer in bacterial and fungal P450s. *Annu Rev Biochem* 2004, **73**, 991-1018.
100. Roiban, G. D.; Reetz, M. T., Expanding the toolbox of organic chemists: directed evolution of P450 monooxygenases as catalysts in regio- and stereoselective oxidative hydroxylation. *Chem Commun (Camb)* 2015, **51**, (12), 2208-24.
101. Kubo, T.; Peters, M. W.; Meinhold, P.; Arnold, F. H., Enantioselective Epoxidation of Terminal Alkenes to (R)- and (S)-Epoxides by Engineered Cytochromes P450 BM-3. *Chemistry* 2006, **12**, (4), 1216-20.
102. Bell, S. G.; Chen, X.; Sowden, R. J.; Xu, F.; Williams, J. N.; Wong, L. L.; Rao, Z., Molecular recognition in (+)-alpha-pinene oxidation by cytochrome P450cam. *J Am Chem Soc* 2003, **125**, (3), 705-14.
103. Porter, T. D.; Coon, M. J., Cytochrome P-450. Multiplicity of isoforms, substrates, and catalytic and regulatory mechanisms. *J Biol Chem* 1991, **266**, (21), 13469-72.

References

104. Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I., Structure and chemistry of cytochrome P450. *Chem Rev* 2005, **105**, (6), 2253-77.
105. Sligar, S. G., Coupling of spin, substrate, and redox equilibria in cytochrome P450. *Biochemistry* 1976, **15**, (24), 5399-406.
106. Munro, A. W.; Girvan, H. M.; McLean, K. J., Cytochrome P450 - redox partner fusion enzymes. *Biochim Biophys Acta* 2007, **1770**, (3), 345-59.
107. Hannemann, F.; Bichet, A.; Ewen, K. M.; Bernhardt, R., Cytochrome P450 systems - biological variations of electron transport chains. *Biochim Biophys Acta* 2007, **1770**, (3), 330-44.
108. McLean, K. J.; Sabri, M.; Marshall, K. R.; Lawson, R. J.; Lewis, D. G.; Clift, D.; Balding, P. R.; Dunford, A. J.; Warman, A. J.; McVey, J. P.; Quinn, A. M.; Sutcliffe, M. J.; Scrutton, N. S.; Munro, A. W., Biodiversity of cytochrome P450 redox systems. *Biochem Soc Trans* 2005, **33**, (Pt 4), 796-801.
109. Gunsalus, I. C., A soluble methylene hydroxylase system: structure and role of cytochrome P-450 and iron-sulfur protein components. *Hoppe Seylers Z Physiol Chem* 1968, **349**, (11), 1610-3.
110. Katagiri, M.; Ganguli, B. N.; Gunsalus, I. C., A soluble cytochrome P-450 functional in methylene hydroxylation. *J Biol Chem* 1968, **243**, (12), 3543-6.
111. Smith, G. C.; Tew, D. G.; Wolf, C. R., Dissection of NADPH-cytochrome P450 oxidoreductase into distinct functional domains. *Proc Natl Acad Sci U S A* 1994, **91**, (18), 8710-4.
112. Serizawa, N.; Matsuoka, T., A two component-type cytochrome P-450 monooxygenase system in a prokaryote that catalyzes hydroxylation of ML-236B to pravastatin, a tissue-selective inhibitor of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *Biochim Biophys Acta* 1991, **1084**, (1), 35-40.
113. Hunter, D. J.; Roberts, G. A.; Ost, T. W.; White, J. H.; Muller, S.; Turner, N. J.; Flitsch, S. L.; Chapman, S. K., Analysis of the domain properties of the novel cytochrome P450 RhF. *FEBS Lett* 2005, **579**, (10), 2215-20.
114. Narhi, L. O.; Fulco, A. J., Characterization of a catalytically self-sufficient 119,000-dalton cytochrome P-450 monooxygenase induced by barbiturates in *Bacillus megaterium*. *J Biol Chem* 1986, **261**, (16), 7160-9.
115. Narhi, L. O.; Fulco, A. J., Identification and characterization of two functional domains in cytochrome P-450BM-3, a catalytically self-sufficient

References

- monooxygenase induced by barbiturates in *Bacillus megaterium*. *J Biol Chem* 1987, **262**, (14), 6683-90.
116. Miura, Y.; Fulco, A. J., (Omega -2) hydroxylation of fatty acids by a soluble system from *Bacillus megaterium*. *J Biol Chem* 1974, **249**, (6), 1880-8.
117. Ruettinger, R. T.; Wen, L. P.; Fulco, A. J., Coding nucleotide, 5' regulatory, and deduced amino acid sequences of P-450BM-3, a single peptide cytochrome P-450:NADPH-P-450 reductase from *Bacillus megaterium*. *J Biol Chem* 1989, **264**, (19), 10987-95.
118. Hare, R. S.; Fulco, A. J., Carbon monoxide and hydroxymercuribenzoate sensitivity of a fatty acid (omega-2) hydroxylase from *Bacillus megaterium*. *Biochem Biophys Res Commun* 1975, **65**, (2), 665-72.
119. Ho, P. P.; Fulco, A. J., Involvement of a single hydroxylase species in the hydroxylation of palmitate at the omega-1, omega-2 and omega-3 positions by a preparation from *Bacillus megaterium*. *Biochim Biophys Acta* 1976, **431**, (2), 249-56.
120. Matson, R. S.; Hare, R. S.; Fulco, A. J., Characteristics of a cytochrome P-450-dependent fatty acid omega-2 hydroxylase from *Bacillus megaterium*. *Biochim Biophys Acta* 1977, **487**, (3), 487-94.
121. Whitehouse, C. J.; Bell, S. G.; Wong, L. L., P450(BM3) (CYP102A1): connecting the dots. *Chem Soc Rev* 2012, **41**, (3), 1218-60.
122. Fulco, A. J., P450BM-3 and other inducible bacterial P450 cytochromes: biochemistry and regulation. *Annu Rev Pharmacol Toxicol* 1991, **31**, 177-203.
123. Munro, A. W.; Daff, S.; Coggins, J. R.; Lindsay, J. G.; Chapman, S. K., Probing electron transfer in flavocytochrome P-450 BM3 and its component domains. *Eur J Biochem* 1996, **239**, (2), 403-9.
124. Munro, A. W.; Leys, D. G.; McLean, K. J.; Marshall, K. R.; Ost, T. W.; Daff, S.; Miles, C. S.; Chapman, S. K.; Lysek, D. A.; Moser, C. C.; Page, C. C.; Dutton, P. L., P450 BM3: the very model of a modern flavocytochrome. *Trends Biochem Sci* 2002, **27**, (5), 250-7.
125. Miura, Y.; Fulco, A. J., Omega-1, Omega-2 and Omega-3 hydroxylation of long-chain fatty acids, amides and alcohols by a soluble enzyme system from *Bacillus megaterium*. *Biochim Biophys Acta* 1975, **388**, (3), 305-17.

References

126. Matson, R. S.; Stein, R. A.; Fulco, A. J., Hydroxylation of 9-hydroxystearate by a soluble cytochrome P-450 dependent fatty acid hydroxylase from *Bacillus megaterium*. *Biochem Biophys Res Commun* 1980, **97**, (3), 955-61.
127. Matson, R. S.; Fulco, A. J., Hydroxystearates as inhibitors of palmitate hydroxylation catalyzed by the cytochrome P-450 monooxygenase from *Bacillus megaterium*. *Biochem Biophys Res Commun* 1981, **103**, (2), 531-5.
128. Ahmed, F.; Al-Mutairi, E. H.; Avery, K. L.; Cullis, P. M.; Primrose, W. U.; Roberts, G. C. K.; Willis, C. L., An unusual matrix of stereocomplementarity in the hydroxylation of monohydroxy fatty acids catalysed by cytochrome P450 from *Bacillus megaterium* with potential application in biotransformations. *Chem Comm* 1999, (20), 2049-50.
129. Davis, S. C.; Sui, Z.; Peterson, J. A.; Ortiz de Montellano, P. R., Oxidation of omega-oxo fatty acids by cytochrome P450BM-3 (CYP102). *Arch Biochem Biophys* 1996, **328**, (1), 35-42.
130. Coelho, P. S.; Brustad, E. M.; Kannan, A.; Arnold, F. H., Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. *Science* 2013, **339**, (6117), 307-10.
131. Seifert, A.; Pleiss, J., Identification of selectivity-determining residues in cytochrome P450 monooxygenases: a systematic analysis of the substrate recognition site 5. *Proteins* 2009, **74**, (4), 1028-35.
132. Seifert, A.; Vomund, S.; Grohmann, K.; Kriening, S.; Urlacher, V. B.; Laschat, S.; Pleiss, J., Rational design of a minimal and highly enriched CYP102A1 mutant library with improved regio-, stereo- and chemoselectivity. *Chembiochem* 2009, **10**, (5), 853-61.
133. Weber, E.; Seifert, A.; Antonovici, M.; Geinitz, C.; Pleiss, J.; Urlacher, V. B., Screening of a minimal enriched P450 BM3 mutant library for hydroxylation of cyclic and acyclic alkanes. *Chem Commun (Camb)* 2011, **47**, (3), 944-6.
134. Correll, C. C.; Batie, C. J.; Ballou, D. P.; Ludwig, M. L., Phthalate dioxygenase reductase: a modular structure for electron transfer from pyridine nucleotides to [2Fe-2S]. *Science* 1992, **258**, (5088), 1604-10.
135. De Mot, R.; Parret, A. H., A novel class of self-sufficient cytochrome P450 monooxygenases in prokaryotes. *Trends Microbiol* 2002, **10**, (11), 502-8.
136. Roberts, G. A.; Celik, A.; Hunter, D. J.; Ost, T. W.; White, J. H.; Chapman, S. K.; Turner, N. J.; Flitsch, S. L., A self-sufficient cytochrome p450 with a

References

- primary structural organization that includes a flavin domain and a [2Fe-2S] redox center. *J Biol Chem* 2003, **278**, (49), 48914-20.
137. Guengerich, F. P.; Munro, A. W., Unusual cytochrome p450 enzymes and reactions. *J Biol Chem* 2013, **288**, (24), 17065-73.
138. Warman, A. J.; Robinson, J. W.; Luciakova, D.; Lawrence, A. D.; Marshall, K. R.; Warren, M. J.; Cheesman, M. R.; Rigby, S. E.; Munro, A. W.; McLean, K. J., Characterization of *Cupriavidus metallidurans* CYP116B1--a thiocarbamate herbicide oxygenating P450-phthalate dioxygenase reductase fusion protein. *Febs J* 2012, **279**, (9), 1675-93.
139. Roberts, G. A.; Grogan, G.; Greter, A.; Flitsch, S. L.; Turner, N. J., Identification of a new class of cytochrome P450 from a *Rhodococcus* sp. *J Bacteriol* 2002, **184**, (14), 3898-908.
140. Liu, L.; Schmid, R. D.; Urlacher, V. B., Cloning, expression, and characterization of a self-sufficient cytochrome P450 monooxygenase from *Rhodococcus ruber* DSM 44319. *Appl Microbiol Biotechnol* 2006, **72**, (5), 876-82.
141. Liu, L.; Schmid, R. D.; Urlacher, V. B., Engineering cytochrome P450 monooxygenase CYP 116B3 for high dealkylation activity. *Biotechnol Lett* 2010, **32**, (6), 841-5.
142. Gallage, N. J.; Moller, B. L., Vanillin-bioconversion and bioengineering of the most popular plant flavor and its de novo biosynthesis in the vanilla orchid. *Mol Plant* 2015, **8**, (1), 40-57.
143. Kaur, B.; Chakraborty, D., Biotechnological and molecular approaches for vanillin production: a review. *Appl Biochem Biotechnol* 2013, **169**, (4), 1353-72.
144. Priefert, H.; Rabenhorst, J.; Steinbuchel, A., Biotechnological production of vanillin. *Appl Microbiol Biotechnol* 2001, **56**, (3-4), 296-314.
145. Walton, N. J.; Mayer, M. J.; Nabad, A., Vanillin. *Phytochemistry* 2003, **63**, (5), 505-15.
146. Rao, S. R.; Ravishankar, G. A., Vanilla flavour: production by conventional and biotechnological routes. *J. Sci. Food Agric.* 2000, **80**, (3), 289-304.
147. Sinha, A. K.; Sharma, U. K.; Sharma, N., A comprehensive review on vanilla flavor: extraction, isolation and quantification of vanillin and others constituents. *Int J Food Sci Nutr* 2008, **59**, (4), 299-326.

References

148. Walton, N. J.; Narbad, A.; Faulds, C.; Williamson, G., Novel approaches to the biosynthesis of vanillin. *Curr Opin Biotechnol* 2000, **11**, (5), 490-6.
149. Zhao, L. Q.; Sun, Z. H.; Zheng, P.; Zhu, L. L., Biotransformation of isoeugenol to vanillin by a novel strain of *Bacillus fusiformis*. *Biotechnol Lett* 2005, **27**, (19), 1505-9.
150. European Directive 88/388/CEE, J. N. L., 22 June 1988.
151. Krings, U.; Berger, R. G., Biotechnological production of flavours and fragrances. *Appl Microbiol Biotechnol* 1998, **49**, (1), 1-8.
152. Sambrook, J.; Russell, D., *Molecular Cloning: A Laboratory Manual*. Third edition ed.; Cold Spring Harbor Laboratory Press: New York, 2001; p 2344.
153. <http://blast.ncbi.nlm.nih.gov>.
154. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J., Basic local alignment search tool. *J Mol Biol* 1990, **215**, (3), 403-10.
155. Gricman, Ł.; Vogel, C.; Pleiss, J., Conservation analysis of class-specific positions in cytochrome P450 monooxygenases: functional and structural relevance. *Proteins* 2014, **82**, (3), 491-504.
156. <http://swissmodel.expasy.org>.
157. Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T., The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 2006, **22**, (2), 195-201.
158. Kiefer, F.; Arnold, K.; Kunzli, M.; Bordoli, L.; Schwede, T., The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res* 2009, **37**, (Database issue), D387-92.
159. Guex, N.; Peitsch, M. C.; Schwede, T., Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* 2009, **30** Suppl 1, S162-73.
160. Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Cassarino, T. G.; Bertoni, M.; Bordoli, L.; Schwede, T., SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 2014, **42**, (Web Server issue), W252-8.
161. Maurer, S.; Urlacher, V.; Schulze, H.; Schmid, R. D., Immobilisation of P450 BM-3 and an NADP⁺ cofactor recycling system: towards a technical application of heme-containing monooxygenases in fine chemical synthesis. *Adv Synth Catal* 2003, **345**, (6-7), 802-10.

References

162. Bradford, M. M., A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 1976, **72**, (1-2), 248-54.
163. Omura, T.; Sato, R., The Carbon Monoxide-Binding Pigment of Liver Microsomes. I. Evidence for Its Hemoprotein Nature. *J Biol Chem* 1964, **239**, 2370-8.
164. Tobias, J. W.; Shrader, T. E.; Rocap, G.; Varshavsky, A., The N-end rule in bacteria. *Science* 1991, **254**, (5036), 1374-7.
165. Dennig, A.; Lülldorf, N.; Liu, H.; Schwaneberg, U., Regioselective *o*-hydroxylation of monosubstituted benzenes by P450 BM3. *Angew Chem Int Ed Engl* 2013, **52**, (32), 8459-62.
166. DeLano, W. L., The PyMOL Molecular Graphics System. 2002.
167. Fischer, M.; Knoll, M.; Sirim, D.; Wagner, F.; Funke, S.; Pleiss, J., The Cytochrome P450 Engineering Database: a navigation and prediction tool for the cytochrome P450 protein family. *Bioinformatics* 2007, **23**, (15), 2015-7.
168. Seifert, A.; Antonovici, M.; Hauer, B.; Pleiss, J., An efficient route to selective bio-oxidation catalysts: an iterative approach comprising modeling, diversification, and screening, based on CYP102A1. *Chembiochem* 2011, **12**, (9), 1346-51.
169. van den Heuvel, R. H. H.; Fraaije, M. W.; Laane, C.; van Berkel, W. J., Enzymatic synthesis of vanillin. *J Agric Food Chem* 2001, **49**, (6), 2954-8.
170. Gunay, A.; Theopold, K. H., C-H bond activations by metal oxo compounds. *Chem Rev* 2010, **110**, (2), 1060-81.
171. Neufeld, K.; Marienhagen, J.; Schwaneberg, U.; Pietruszka, J., Benzylic hydroxylation of aromatic compounds by P450 BM3. *Green Chem* 2013, **15**, (9), 2408-21.
172. Sulistyaningdyah, W. T.; Ogawa, J.; Li, Q. S.; Maeda, C.; Yano, Y.; Schmid, R. D.; Shimizu, S., Hydroxylation activity of P450 BM-3 mutant F87V towards aromatic compounds and its application to the synthesis of hydroquinone derivatives from phenolic compounds. *Appl Microbiol Biotechnol* 2005, **67**, (4), 556-62.
173. Shoji, O.; Kunimatsu, T.; Kawakami, N.; Watanabe, Y., Highly selective hydroxylation of benzene to phenol by wild-type cytochrome P450BM3

References

- assisted by decoy molecules. *Angew Chem Int Ed Engl* 2013, **52**, (26), 6606-10.
174. Whitehouse, C. J.; Rees, N. H.; Bell, S. G.; Wong, L. L., Dearomatisation of *o*-xylene by P450BM3 (CYP102A1). *Chemistry* 2011, **17**, (24), 6862-8.
175. Watanabe, Y.; Laschat, S.; Budde, M.; Affolter, O.; Shimada, Y.; Urlacher, V. B., Oxidation of acyclic monoterpenes by P450 BM-3 monooxygenase: influence of the substrate E/Z-isomerism on enzyme chemo- and regioselectivity. *Tetrahedron* 2007, **63**, (38), 9413-22.

7 Supplementary material

7.1 Genes, proteins, vectors, primers and strains

7.1.1 Genes

The genes *mobA*, *cyp116B3* and *vaoA* were synthesized (GeneArt[®] Gene Synthesis, Life Technologies GmbH, Darmstadt, Germany), codon-optimized for the heterologous expression in *E. coli*. His-tags are underlined in the wild type gene sequences shown below.

Codon-optimized *mobA* gene from *Comamonas testosteroni* GZ39

```

ATGCAGTTTCATCTGAATGGTTTTTCGTCCGGGTAATCCGCTGATTGCACCGGCAAGTCCGCTGGCACCGGCACAT
ACCGAAGCAGTTCCGAGCCAGGTTGATGTTCTGATTGTTGGTTGTGGTCCGGCAGGCTGACCCCTGGCAGCACAG
CTGGCAGCATTTCGGGATATTCGTACCTGTATTGTGGAACAGAAAGAAGGTCCGATGGAACCTGGGTCAGGCAGAT
GGTATTGCATGTCGTACCATGGAAATGTTTGAAGCATTTGAATTTGCCGATAGCATCCTGAAAGAAGCATGTTGG
ATTAACGATGTGACCTTTTGGAAACCTGATCCGGCAGCAGCCTGGTTCGTATTGCACGTCATGGTTCGTGTTTCAGGAT
ACCGAAGATGGTCTGAGCGAATTTCCGCATGTTATTCTGAATCAGGCACGTTTCATGATCATTATCTGGAACGT
ATGCGTAATAGCCCCGAGCCGTCTGGAACCGCATTATGCACGTCGTGTTCTGGATGTTAAAAATTGATCATGGTGCA
GCAGATTATCCGGTTACCGTTACCCTGGAACGTTGTGATGCAGCACATGCAGGTCAGATTGAAACCGTTCAGGCA
CGTTATGTTGTGGGTTGTGATGGTGCACGTAGCAATGTTTCGTTCGTGCAATTGGTTCGTTCAGCTGGTTCAGTATAGC
GCAAATCAGGCATGGGGTGTATGGATGTTCTGGCAGTTACCGATTTTCCGGATGTTTCGTTATAAAAGTTGCCATT
CAGAGCGAACAGGGTAATGTTCTGATTATTTCCGCGTGAAGGTGGTCATCTGGTTCGTTTTTATGTGAAATGGAT
AAACTGGATGCCGATGAACGTGTTGCAAGCCGTAATATTACCGTTGAACAGCTGATTGCAACCCGACAGCGTGT
CTGCATCCGTATAAACTGGACGTTAAAAATGTTCCGTGGTGGTTCAGTGTATGAAATTGGTTCAGCGTATTTGCGCC
AAATATGATGATGTTGAGATGCAGTTGCAACACCGGATAGTCCGCTGCCTCGTGTTTTTTATTGCCGGTGATGCA
TGTCATACCCATAGCCCCGAAAGCAGGTCAGGGTATGAATTTTAGCATGCAGGATTCATTTAATCTGGGTTGGAAA
CTGGCAGCAGTTCTGCGTAAACAGTGTGCACCGGAACTGCTGCATACCTATAGCAGCGAACGTCAGGTTGTTGCA
CAGCAGCTGATTGATTTTTGATCGTGAATGGGCCAAAATGTTTAGCGATCCGGCAAAAAGAAGGTGGTCAAGGCGGA
GTTGATCCGAAAGAATTTTCAGAAATATTTTCGAACAGCACGGTTCGTTTTTACCGCAGGCGTTGGCACCCATTATGCA
CCGAGCCTGCTGACAGGTCAGGCAAGCCATCAGGCACTGGCAAGCGGTTTTTACCGTTGGTATGCGTTTTTCATAGC
GCACCGGTTGTTTCGTGTTAGTGATGCAAAACCGCTGCAGCTGGGTCATTGTGGTAAAGCAGATGGTTCGTTGGCGT
CTGTATGCATTTGCAGGTCAGAATGATCTGGCACAGCCGAAAGCGGTCTGCTGGCACTGTGTTCGTTTTCTGGAA
AGTGATGCAGCTTACCAGCTGCGTTCGTTTTTACCCGAGCGGTTCAGGATATTGATAGCATTTTTTATGCTGCGTGCC
ATTTTTCCGCAGGCATATACCGAAGTTGCACTGGAAACCCCTGCCTGCCTGCTGCTGCCTCCGAAAGGTCAGCTG
GGTATGATTGATTATGAAAAAGTGTTCAGTCCGGATCTGAAAAATGCAGGCCAGGATATTTTCGAACTGCGTGGT
ATTGATCGTCAGCAGGGTGCAGTGGTTGTTGTTTCGTCCGGATCAGTATGTTGCACAGGTTCTGCCGCTGGGTGAT
CATGCAGCACTGAGCGCATATTTTGAAGCTTTATGCGTGCACATCATCATCACCATAA

```

GenBank ID of not codon-optimized gene sequence: AY450844

Supplementary material

Codon-optimized *cyp116B3* gene from *Rhodococcus ruber* DSM 44319

ATGCATCATCATCATCACCATAGCGCAAGCGTTCCGGCAAGCGCATGTCCGGTTGATCATGCAGCACTGGCAGGC
GGTTGTCCGGTTAGCACCAATGCAGCAGCATTGATCCGTTTGGTCCGGCATATCAGGCAGATCCGGCAGAAAAGC
CTGCGTTGGAGCCGTGATGAAGAACCAGGTTTTTTATAGTCCGGAACCTGGGTTATTGGGTTGTTACCCGTTATGAA
GATGTGAAAGCCGTGTTTCGTGATAATCTGGTTTTTTAGTCCGGCAATTGCCCTGGAAAAAATTACACCGTTAGC
GAAGAAGCAACCAGCAACCCTGGCAGCTTATGATTATGCAATGGCAGCTACCCCTGGTGAATGAAGATGAACCGGCA
CACATGCCTCGTTCGTGACTGATGGATCCGTTTACCCCGAAAGAACTGGCACATCATGAAGCAATGGTTTCGT
CGTCTGACCCGTGAATATGTTGATCGTTTTTGGTTGAAAGCGGTAAGCAGATCTGGTTGATGAAATGCTGTGGGAA
GTTCCGCTGACCGTTGCACTGCATTTTTCTGGGTGTTCCGGAAGAGGATATGGCAACCATGCGTAAATATTCAATT
GCCCATACCGTTAATACCTGGGGTTCGTCCGGCACCAGGAAGAACAGGTTGCAGTTGCAGAAGCAGTTGGTTCGTTTT
TGGCAGTATGCAGGCACCGTTCTGGAAAAAATGCGTCAGGATCCGAGCGGTCATGGTTGGATGCCGTATGGTATT
CGTATGCAGCAGCAGATGCCGGATGTTGTTACCGATAGCTATCTGCATAGCATGATGATGGCAGGTATTGTTGCA
GCACATGAAACCACCGCAAATGCAAGCGCAAATGCATTTAAACTGCTGCTGGAAAAATCGTCCGGTTTGGGAAGAA
ATTTGTGCAGATCCGAGCCTGATTCCGAATGCAGTTGAAGAATGTCTGCGTCATAGCGGTAGCGTTGCAGCATGG
CGTTCGTGTTGCAACCACCGATACCCGTTATTGGTGTGTTGATATTCCGGCAGGCGCAAAACTGCTGGTTGTTAAT
GCAAGCGCCAATCATGATGAACGTCATTTTTGATCGTCCGGATGAATTTGATATTTCGTTCGTCCGAATAGCAGCGAT
CATCTGACCTTTGGTTATGGTAGCCATCAGTGTATGGGTAAAAATCTGGCACGTATGGAAATGCAGATCTTTCTG
GAAGAACTGACCACCCGTCTGCCGCACATGGAACCTGGTTCCGGATCAAGAATTTACCTATCTGCCGAATACCAGC
TTTCGTGGTCCGGATCATGTTTGGGTTTCAGTGGGATCCGCAGGCAAATCCGGAACGTACCGATCCGGCAGTTCTG
CAGCGTCAGCATCCGGTTACCATTGGTGAACCGAGCACCCGTAGCGTTAGCCGTACCGTTACCGTTGAACGTCCTG
GATCGTATTGTTGATGATGTTCTGCGTGTGTTCTGCGTGCACCCGGCAGGTAATGCACTGCCTGCATGGACACCG
GGTGCACATATTGATGTTGATCTGGGTGCACTGAGCCGTAGTATAGCCTGTGTGGTGCACCTGATGCACCGACC
TATGAAATTGCAGTTCTGCTGGATCCGGAAAGCCGTGGTGGTAGCCGTTATGTTTCATGAACAGCTGCGTGTGGT
GGTAGCCTGCGTATTTCGTGGTCCGCGTAATCATTTTTGCACTGGATCCGGATGCAGAACATTATGTTTTTGGTTGCC
GGTGGTATTGGTATTACACCGGTGCTGGCAATGGCAGATCATGCACGTGCTCGTGGTTGGAGCTATGAACTGCAT
TATTGTGGTTCGTAATCGTAGCGGTATGGCATATCTGGAACGTGTTGCAGGTCATGGTGATCGTGCAGCTCTGCAT
GTTAGTGCCGAAGGCACCCGTGTTGATCTGGCAGCACTGCTGGCAACACCCGGTGGAGCGGCACCCAGATTTATGCA
TGTGGTCCGGGTCGTCTGCTGGCAGGTCTGGAAGATGCAAGCCGTGATTGGCCTGATGGTGCATGTTGAA
CATTTTACCAGCAGCCTGACCGCACTGGATCCTGACGTTGAACATGCATTTGATCTGGATCTGCGTGATAGTGGT
CTGACCGTTTCGTGTTGAACCGACCCAGACCGTTCTGGATGCACTGCGTGCAAATAATATTGATGTTCCGAGCGAT
TGCGAAGAGGGTCTGTGTGGTAGCTGTGAAGTTACCGTCTGGAAGGTGAAGTTGATCATCGTGATACCGTTCTG
ACCAAAGCAGAACGTGCAGCAAATCGTCAGATGATGACCTGTTGTAGCCGTGCATGTGGTGTGATCGTCTGACCCCTG
CGTCTGTAA

GenBank ID of not codon-optimized gene sequence: AY957485

Supplementary material

cyp102A1 gene from *Bacillus megaterium* ATCC 14581

ATGACAATTAAGAAATGCCTCAGCCAAAAACGTTTGGAGAGCTTAAAAATTTACCGTTATTAAACACAGATAAA
CCGGTTCAAGCTTTGATGAAAATTGCGGATGAATTAGGAGAAATCTTTAAATTCGAGGCGCTGGTCTGTAAACG
CGTACTTTATCAAGTCAGCGTCTAATTAAGAAGCATGCGATGAATCACGCTTTGATAAAAACTTAAGTCAAGCG
CTTAAATTTGTACGTGATTTTGCAGGAGACGGGTATTTACAAGCTGGACGCATGAAAAAATGGAAAAAGCG
CATAATATCTTACTTCCAAGCTTCAGTCAGCAGGCAATGAAAGGCTATCATGCGATGATGGTCGATATCGCCGTG
CAGCTTGTTCAAAGTGGGAGCGTCTAAATGCAGATGAGCATATTGAAGTACCGGAAGACATGACACGTTTAAACG
CTTGATACAATTGGTCTTTGCGGCTTTAACTATCGCTTTAACAGCTTTTACCAGATCAGCCTCATCCATTTATT
ACAAGTATGGTCCGTGCACTGGATGAAGCAATGAACAAGCTGCAGCGAGCAAATCCAGACGACCCAGCTTATGAT
GAAAACAAGCGCCAGTTTCAAGAAGATATCAAGGTGATGAACGACCTAGTAGATAAAAAATTTATGCGATCGCAAA
GCAAGCGGTGAACAAAGCGATGATTTATTAACGCATATGCTAAACGGAAAAGATCCAGAAAACGGGTGAGCCGCTT
GATGACGAGAACATTGCTATCAAATTATTACATTCTTAATTGCGGGACACGAAAACAAGTGGTCTTTTATCA
TTTGGCTGTATTTCTTAGTGAAAAATCCACATGTATTACAAAAAGCAGCAGAAGAAGCAGCAGAGTTCTAGTA
GATCCTGTTCCAAGCTACAAACAAGTCAAACAGCTTAAATATGTCGGCATGGTCTTAAACGAAGCGCTGCGCTTA
TGGCCAACCTGCTCCTGCGTTTTCCCTATATGCAAAAAGAAGATACGGTGCTTGGAGGAGAATATCCTTTAGAAAA
GGCGACGAACATAATGGTTCTGATTCTCAGCTTACCCTGATAAAAAAATTTGGGGAGACGATGTGGAAGAGTTC
CGTCCAGAGCGTTTTGAAAATCCAAGTGCATTCCGAGCATGCGTTTTAAACCGTTTTGAAAACGGTTCAGCGTGCG
TGTATCGGTGAGCAGTTGCTCTTCATGAAGCAACGCTGGTACTTGGTATGATGCTAAAAACACTTTGACTTTGAA
GATCATAAAAACCTACGAGCTGGATATTAAGAAACTTTAACGTTAAAAACCTGAAGGCTTTGTGGTAAAAAGCAAAA
TCGAAAAAATTCGCTTGGCGGTATTCCTTACCTAGCACTGAACAGTCTGCTAAAAAAGTACGCAAAAAGGCA
GAAAACGCTCATAATACGCCGCTGCTTGTGCTATACGGTTCAAATATGGGAACAGCTGAAGGAACGGCGCGTGT
TTAGCAGATATTGCAATGAGCAAAGGATTTGACCCGAGGTGCAACGCTTGATTCACACGCCGAAATCTTCCG
CCGAAGGAGCTGATTAATGTAACCGGCTTATAACGGTCACTCCGCTGATAACGCAAAGCAATTTGTGCGAC
TGTTTAGACCAAGCGTCTGCTGATGAAGTAAAAGGCGTTGCTACTCCGTATTTGGATGCGGCGATAAAAACTGG
GCTACTACGTATCAAAGTGCCTGCTTTTTATCGATGAAACGCTTGCCGCTAAAAGGGGCAGAAAACATCGCTGAC
CGCGGTGAAGCAGATGCAAGCGACGACTTTGAAGGCACATATGAAGAATGGCGTGAACATATGTGGAGTGACGTA
GCAGCCTACTTTAACCTCGACATTGAAAACAGTGAAGATAATAAATCTACTCTTTCACTTCAATTTGTGCGACAGC
GCCGCGGATATGCCGCTTGCGAAAATGCACGGTGCCTTTTCAACGAACGTCGTAGCAAGCAAAGAACTTCAACAG
CCAGGCAGTGCACGAAGCACGCGACATCTTGAATTTGAACCTTCCAAAAGAAGCTTCTTATCAAGAAGGAGATCAT
TTAGGTGTTATTCCTCGCAACTATGAAGGAATAGTAAACCGTGTAAACAGCAAGGTTGCGCCTAGATGCATCACAG
CAAATCCGTCTGGAAGCAGAAGAAGAAAAATTAGCTCATTTGCCACTCGCTAAAACAGTATCCGTAGAAGAGCTT
CTGCAATACGTGGAGCTTCAAGATCCTGTTACGCGCACGCAGCTTCCGCGCAATGGCTGCTAAAACGGTCTGCCCG
CCGCATAAAGTAGAGCTTGAAGCCTTGCTTGAAGAAGCAAGCTTACAAAAGAACAAGTGTGGCAAAAACGTTTAAACA
ATGCTTGAACCTGCTTGAAAAATACCCGGCGTGTGAAATGAAATTCAGCGAATTTATCGCCCTTCTGCCAAGCATA
CGCCCGCGCTATTACTCGATTTCTTCATCACCTCGTGTGATGAAAAACAAGCAAGCATCACGGTCAGCGTTGTC
TCAGGAGAAGCGTGGAGCGGATATGGAGAATATAAAGGAATTTGCGTCAACTATCTTGCCGAGCTGCAAGAAGGA
GATACGATTACGTGCTTTATTTCCACACCCGAGTCAGAATTTACGCTGCCAAAAGACCCTGAAAACGCCGCTTATC
ATGGTCCGACCGGGAACAGGCGTGCAGCCGTTTAGAGGCTTTGTGCAGGCGCGCAAACAGCTAAAAGAACAAGGA
CAGTCACTTGGAGAAGCACATTTATACTTCGGCTGCCGTTACCTCATGAAGACTATCTGTATCAAGAAGAGCTT
GAAAACGCCCAAAGCGAAGGCATCATTACGCTTACATACCGCTTTTTCTCGCATGCCAAATCAGCCGAAAACATAC
GTTACAGCAGTAATGGAACAAGACGGCAAGAAATGATTGAACCTTCTTGATCAAGGAGCGCACTTCTATATTTGC
GGAGACGGAAGCCAAATGGCACCTGCCGTTGAAGCAACGCTTATGAAAAGCTATGCTGACGTTACCAAGTGAGT
GAAGCAGACGCTCGCTTATGGCTGCAGCAGCTAGAAGAAAAAGGCCGATACGCAAAAAGACGTGTGGGCTGGGTAA

GenBank ID of gene sequence: J04832

Supplementary material

Codon-optimized *vaoA* gene from *Penicillium simplicissimum* CBS 170.90

ATGAGCAAAACCCAAGAATTTTCGTCCGCTGACCCTGCCTCCGAAACTGAGCCTGAGCGATTTTAATGAATTTATC
CAGGATATCATTTCGCATCGTGGGTAGCGAAAATGTTGAAGTTATTAGCAGCAAAGATCAGATTGTGGATGGCAGC
TATATGAAACCGACCCATACCCATGATCCGCATCATGTTATGGATCAGGATTATTTCTGGCAAGCGCAATTGTT
GCACCCGTAATGTTGCAGATGTTTCAGAGCATTGTTGGTCTGGCAAACAAATTTAGCTTTCCGCTGTGGCCGATT
AGCATTGGTCGTAATAGCGGTTATGGTGGTGCAGCACCGCGTGTAGCGGTAGCGTTGTTCTGGATATGGGCAA
AACATGAATCGTGTTCTGGAAGTTAATGTGGAAGGTGCCTATTGTGTTGTTGAACCGGGTGTACCTATCATGAT
CTGCATAATTATCTGGAAGCCAATAACCTGCGTGATAAACTGTGGCTGGATGTTCCGGATCTGGGTGGTGGTAGC
GTTCTGGGTAATGCAGTTGAACGTGGTGGTTGGTTATACCCCGTATGGTGATCATTGGATGATGCATAGCGGTATG
GAAGTTGTGCTGGCAAATGGTGAACCTGCTGCGTACCGGTATGGGTGCACTGCCGGATCCGAAACGTCCGGAAACA
ATGGGTCTGAAACCGGAAGATCAGCCGTGGTCAAAAATTGCACACCTGTTTCCGTATGGTTTTGGTCCGTATATTT
GATGGTCTGTTTAGCCAGAGCAATATGGGTATTGTTACCAAAATTTGGCATTGGCTGATGCCGAATCCGGGTGGT
TATCAGAGCTATCTGATTACCCTGCCGAAAGATGGTGATCTGAAACAGGCAGTTGATATTATCCGTCCGCTGCGT
CTGGGTATGGCACTGCAGAATGTTCCGACCATTTCGTCATATTCTGCTGGATGCCGCAGTTCTGGGTGATAAACGT
AGCTATAGCAGTAAACCGAACCGCTGAGTGATGAAGAACTGGATAAAAATTGCAAAACAGCTGAATCTGGGTGCG
TGGAACTTTTATGGTGCCTGTATGGTCCGGAACCGATTTCGTCGTGTGCTGTGGGAAACCATTAAGATGCATTT
AGCGCAATTCGGGTGTGAAATTTCTATTTTCCGGAAGATACACCGGAAAATTCAGTTCTGCGTGTTCGTGATAAA
ACCATGCAGGGTATTCCGACCTATGATGAACTGAAATGGATTGATTGGCTGCCGAATGGTGCCACCTGTTTTTT
AGCCCGATTGCAAAAGTTAGCGGTGAAGATGCAATGATGCAGTATGCAGTGACCAAAAAACGTTGTCAAGAAGCA
GGTCTGGATTTTATTGGCACCTTTACCGTTGGTATGCGTGAATGCATCATATTGTGTGCATCGTGTTTAACAAA
AAAGATCTGATCCAGAAACGCAAAGTTTCAGTGGCTGATGCGTACCCTGATTGATGATTGTGCAGCAAAATGGTTGG
GGTGAATATCGTACCCATCTGGCATTATGGACCAGATTATGGAAACCTATAACTGGAACAATAGCAGCTTTCTG
CGCTTTAATGAAGTGCTGAAAAATGCCGTTGATCCGAATGGTATTATTGCACCGGGTAAAAGCGGTGTTTGCCG
AGCCAGTATAGCCATGTTACCTGGAAACTGTAA

GenBank ID of not codon-optimized gene sequence (with introns): Y15627

Supplementary material

7.1.2 Proteins

Amino acid sequences of the wild type enzymes are shown below (without His-tags).

MobA *m*-hydroxybenzoate hydroxylase from *C. testosteroni* GZ39

GenBank: AAR25885

MQFHLNGFRPGNPLIAPASPLAPAHEAVPSQVDVLIIVGCGPAGLTLAAQLAAFPDIRTCIVEQKEGPMELGQAD
GIACRTMEMFEAFEFADSILKEACWINDVTFWKPDPAQPGRIARHGRVQDTEGLSEFPHVILNQARVHDHYLER
MRNSPSRLEPHYARRVLDVKIDHGAADYPVTVTLERCDAAHAGQIETVQARYVVGCDGARSNVRRAIGRQLVGD
ANQAWGVMVDLAVTDFPDVRYKVAIQSEQGNVLIIPREGGHLVRFYVEMDKLDADERVASRNIITVEQLIATAQ
LHPYKLDVKNVPWWSVYEIGQRICAKYDDVADAVATPDSPLPRVFIAGDACHTHSPKAGQGMNFSMQDSFN
LQWKLAAVLRKQCAPELLHTYSSERQVVAQQLIDFDREWAKMFSDBAKEGGQGGVDPKEFQKYFEQHGRFTAG
VGTHTYAPSLLTGQASHQALASGFTVGMRFHSAPVVRVSDAKPLQLGHCGKADGRWRLYAFAGQNDLAQPE
SGLLALCRFLESDAASPLRRFTPSGQDIDSIIDLRAIFPQAYTEVALETLPALLLPPKQGLGMIDYEKVFSP
DLKNAGQDIFELRGIDRQQGALVVVRPDQYVAQVPLPLGDHAALSAYFESFMRA

CYP116B3 cytochrome P450 monooxygenase from *R. ruber* DSM 44319

GenBank: AAY17950

MSASVPASACPVDHAALAGGCPVSTNAAAFDPFGPAYQADPAESLRWSRDEEPVFYSPELGYWVVTRYEDV
KAVFRDNLVFSPIALEKITPVSEEAATLARYDYAMARTLVNEDEPAHMPRRRALMDFPTPKELAHHEAMVRRL
TREYVDRFVESGKADLVEDEMLWEVPLTVALHFLGVPPEEDMATMRKYSIAHTVNTWGRPAPEEQVAVAEAV
GRFWQYAGTVLEKMRQDPSGHGWMPYGIIRMQQMPDVVTDVSYLHSMAGIVAAHETTANASANAFKLLLEN
RPVWEEICADPSLIPNAVEECLRHSGSVAAWRRVATTDTRIGDVIDIPAGAKLLVNASANHDERHDFDRP
DEFDIRRPNSSDHLTFGYGSHQCMGKNLARMEMQIFLEELTTRLPHMELVPDQEFYLPNTSFRGPDHV
VWVQWDPQANPERTDPAVLQRQHPVTIGEPSTRSVSRTVTVERLDRIVDDVLRVVLRAPAGNALPAWTP
GAHIDVDLGLSRSQYSLCGAPDAPTYEIAVLDPESRGGSRVYVHEQLRVGGSLRIRGPRNHFALDP
DAEHYVVFVAGGIGITPVLAMADHARARGWSYELHYCGRNRSGMAYLERVAGHGDRAALHVS
AEGTRVDLAALLATPVSGTQIYACGPGRLLAGLEDASRHWPDGALHVEHFTSSLTALDPDVEHAF
DLDRDSGLTVRVEPTQTVLDALRANNIDVPSDCEEGLCGSCEVTVLEGEVDHRDRTLTKAERA
ANRQMMTCCSRACGDRLTLRL

CYP102A1 cytochrome P450 monooxygenase from *B. megaterium* ATCC 14581

GenBank: AAA87602

MTIKEMPQPKTFGELKNLPLLNTDKPVQALMKIADDELGEIFKFEAPGRVTRYLSSQRLIKEACDES
RFDKNLSQALKFVRDFAGDGLFTSWTHEKNWKAHNILLPSFSQQAMKGYHAMMVDIAVQLVQKWERL
NADEHIEVPEDMTRLTLDTIGLCGFNYRFNSFYRDQPHPFITSMVRALDEAMNKLQRANPDDPAYDEN
KRQFQEDIKVMNDLVDKI IADRKASGEQSDDLLTHMLNGKDPETGEPLDDENIRYQIITFLIAGH
ETTSGLLSFALYFLVKNPHVLQKAAEEAARVLVDPVPSYKQVKQLKYVGMVLNEALRLWPTA
PAFSLYAKEDTVLGGEYPLEKGDELMLVLIPLQHRDKTIWGDVVEEFRPERFENPSAIPQHAFK
PFGNGQRACIGQQFALHEATLVLGMMLKHDFDFEDHTNYELDIKETLTLKPEGFVVKAKSKKI
PLGGIPSPSTEQSAKKVRKKAENAHTPLLVLVYGSNMGTAEGTARDLADIAMSKGFAPQVATLDS
HAGNLPREGAVLIVTASYNGHPPDNAKQFVDWLDQASADEVKGVRYSVFVCGGDKNWATTYQKVP
AFIDETLAAKGAENIADRGEADASDDFEGTYEEWREHMWSDVAAYFNLDIENSEDNKSTLSLQFV
DVAADMPLAKMHGAFSTNVVASKELQQPGSARSTRHLEIELPKEASYQEGDHLGVI PRNYEGIV
NRVTARFGLDASQQIRLEAEEKLAHLPLAKTVSVEELLQYVELQDPVTRTQLRAMAAKTVCPPH
KVELEALLEKQAYKEQVLAKRLTMLELLEKYPACEMKFSEFIALLPRI RPRYYSISSSPRVDEK
QASITVSVVSGEAWSGYGEYKGIASNYLAELQEGDTITCFISTPQSEFTLPKDPETPLIMVGP
GTGVAPFRGFVQARKQLKEQGQSLGEAHLVYFGCRSPHEDYLYQEELENAQSEGIITLHTAFSR
MPNQPKTYVQHVMEQDGKKLIELLDQGAHFYICGDSQMAPAVEATLMKSYADVHVQVSEADARLW
LQQLLEEKGRYAKDVWAG

Supplementary material

VAO vanillyl alcohol oxidase from *P. simplicissimum* CBS 170.90

GenBank: CAA75722

MSKTQEFRPLTLPPKLSLSDFNEFIQDIIRIVGSENVEVISSKDQIVDGSYMKPTHHTDHPHHVMDQDYFLASAIV
APRNVADVQSIVGLANKFSFPLWPISIGRNSGYGGAAPRVSGSVVLDMGKNMNRVLEVNVEGAYCVVEPGVTYHD
LHNYLEANNLRDKLWLDVVDLGGGSVLGNAVERGVGYTPYGDHMMHSGMEVVLANGE LLRTGMGALPDPKRPET
MGLKPEDQPWSKIAHLFPYGFPGPYIDGLFSQSNMGIVTKIGIWLMPNPGGYQSYLITLPKDGDLKQAVDIIRPLR
LGMALQNVPTIRHILLDAAVLGDKRSYSSKTEPLSDEELDKIAKQLNLGRWNFYGALYGPEPIRRVLWETIKDAF
SAIPGVKFFYPEDTPENSVLRVRDKTMOGIPTYDELKWDWLPNGAHLFFSPIAKVSGEDAMMQYAVTKKRCQEA
GLDFIGTFTVGMREMHIVCIVFNKKDLIQKRKVQWLMRTLIDDCANGWGEYRTHLAFMDQIMETYNWNNSSFL
RFNEVLKNAVDPNGIIAPGKSGVWPSQYSHVTWKL

7.1.3 Vectors

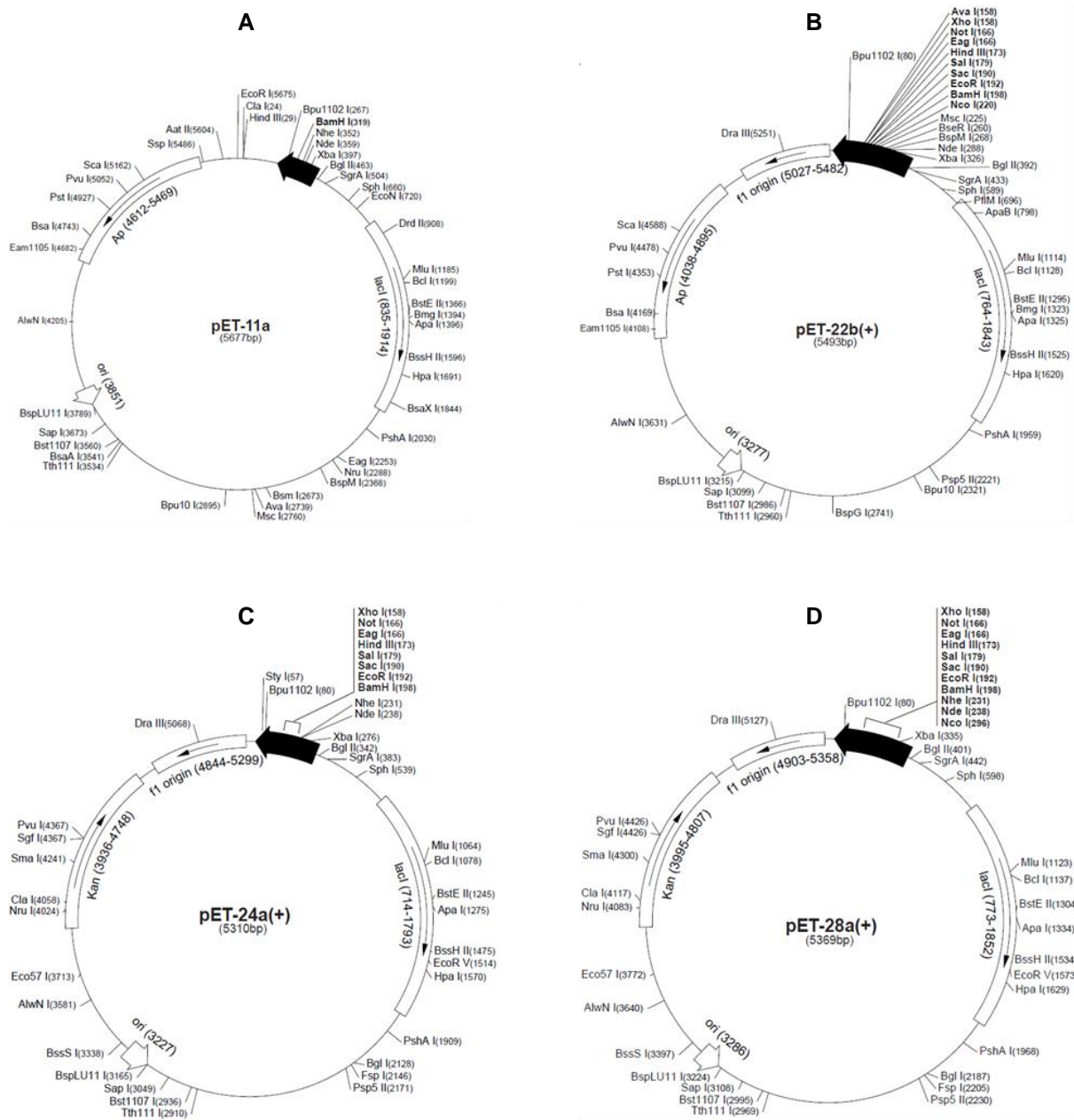


Figure 7.1: Expression vectors (Novagen, Madison, WI, USA) used in this work. (A) pET-11a was used for expression of *mobA*. (B) pET-22b(+) was used for expression of *cyp116B3*, *cyp102A1* and *vao*. (C) pET-24a(+) was used for expression of *cyp116B3*. (D) pET-28a(+) was used for expression of *cyp102A1*.

Supplementary material

Table 7.1: List of expression vectors (Novagen, Madison, WI, USA) used in this work.

Plasmid (Size)	Sequence landmarks	
pET-11a (5677 bp)	T7 promoter T7 transcription start T7-Tag coding sequence T7 terminator <i>lacI</i> coding sequence pBR322 origin Amp ^R (<i>bla</i>) coding sequence	432-448 431 328-360 213-259 835-1914 3851 4612-5469
pET-22b(+) (5493 bp)	T7 promoter T7 transcription start <i>pelB</i> coding sequence Multiple cloning sites (<i>NcoI</i> – <i>XhoI</i>) His-Tag coding sequence T7 terminator <i>lacI</i> coding sequence pBR322 origin Amp ^R (<i>bla</i>) coding sequence f1 origin	361-377 360 224-289 158-225 140-157 26-72 764-1843 3277 4038-4895 5027-5482
pET-24a(+) (5310 bp)	T7 promoter T7 transcription start T7-Tag coding sequence Multiple cloning sites (<i>Bam</i> HI – <i>Xho</i> I) His-Tag coding sequence T7 terminator <i>lacI</i> coding sequence pBR322 origin Kan ^R coding sequence f1 origin	311-327 310 207-239 158-203 140-157 26-72 714-1793 3227 3936-4748 4844-5299
pET-28a(+) (5369 bp)	T7 promoter T7 transcription start His-Tag coding sequence T7-Tag coding sequence Multiple cloning sites (<i>Bam</i> HI – <i>Xho</i> I) His-Tag coding sequence T7 terminator <i>lacI</i> coding sequence pBR322 origin Kan ^R coding sequence f1 origin	370-386 369 270-287 207-239 158-203 140-157 26-72 773-1852 3286 3995-4807 4903-5358

Supplementary material

Table 7.2: Vector constructs used in this work. All plasmids are available at ITB.

Name	Gene insert / Application	Point mutation	ITB No.
pET-11a	-	-	pITB893
pET-22b(+)	-	-	pITB895
pET-24a(+)	-	-	pITB896
pET-28a(+)	-	-	pITB897
pET-11a_MobA	Codon-optimized <i>m</i> -hydroxybenzoate hydroxylase MobA from <i>C. testosteroni</i> GZ39 with C-terminal His ₆ -tag	-	pITB899
pET-11a_MobA_V257A	Codon-optimized <i>m</i> -hydroxybenzoate hydroxylase MobA from <i>C. testosteroni</i> GZ39 with C-terminal His ₆ -tag	V257A	pITB901
pET-24a(+)_CYP116B3	Codon-optimized cytochrome P450 monooxygenase CYP116B3 from <i>R. ruber</i> DSM 44319 with N-terminal His ₆ -tag	-	pITB903
pET-24a(+)_CYP116B3_Var1	Codon-optimized cytochrome P450 monooxygenase CYP116B3 from <i>R. ruber</i> DSM 44319 with N-terminal MetGlySerSer (MGSS)-sequence prior to His ₆ -tag	-	pITB905
pET-24a(+)_CYP116B3_Var2	Codon-optimized cytochrome P450 monooxygenase CYP116B3 from <i>R. ruber</i> DSM 44319 with C-terminal His ₆ -tag	-	pITB906
pET-24a(+)_CYP116B3_Var3	Codon-optimized cytochrome P450 monooxygenase CYP116B3 from <i>R. ruber</i> DSM 44319 without His ₆ -tag	-	pITB907
pET-24a(+)_CYP116B3_Var1_A109F	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	A109F	pITB908
pET-24a(+)_CYP116B3_Var1_A109I	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	A109I	pITB909
pET-24a(+)_CYP116B3_Var1_A109L	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	A109L	pITB910
pET-24a(+)_CYP116B3_Var1_A109V	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	A109V	pITB911
pET-24a(+)_CYP116B3_Var1_V316A	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	V316A	pITB912
pET-24a(+)_CYP116B3_Var1_V316F	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	V316F	pITB913
pET-24a(+)_CYP116B3_Var1_V316I	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	V316I	pITB914
pET-24a(+)_CYP116B3_Var1_V316L	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	V316L	pITB915

Supplementary material

Table 7.2: Vector constructs used in this work (continued).

Name	Gene insert / Application	Point mutation	ITB No.
pET-22b(+)_CYP116B3_Var1	Codon-optimized cytochrome P450 monooxygenase CYP116B3 from <i>R. ruber</i> DSM 44319 with N-terminal MetGlySerSer (MGSS)-sequence prior to His ₆ -tag	-	pITB916
pET-22b(+)_CYP116B3_Var1_A317F	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	A317F	pITB917
pET-22b(+)_CYP116B3_Var1_A317I	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	A317I	pITB918
pET-22b(+)_CYP116B3_Var1_A317L	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	A317L	pITB919
pET-22b(+)_CYP116B3_Var1_A317V	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	A317V	pITB920
pET-22b(+)_CYP116B3_Var1_A318F	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	A318F	pITB921
pET-22b(+)_CYP116B3_Var1_A318I	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	A318I	pITB922
pET-22b(+)_CYP116B3_Var1_A318L	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	A318L	pITB923
pET-22b(+)_CYP116B3_Var1_A318V	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	A318V	pITB924
pET-22b(+)_CYP116B3_Var1_F420A	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	F420A	pITB925
pET-22b(+)_CYP116B3_Var1_F420I	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	F420I	pITB926
pET-22b(+)_CYP116B3_Var1_F420L	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	F420L	pITB927
pET-22b(+)_CYP116B3_Var1_F420V	Focused mutant library of CYP116B3 (N-terminal MGSS_His ₆ -sequence)	F420V	pITB928
pET-28a(+)_CYP102A1	Cytochrome P450 monooxygenase CYP102A1 from <i>B. megaterium</i> ATCC 14581	-	pITB930*
pET-28a(+)_CYP102A1_A328F	Focused mutant library of CYP102A1	A328F	pITB931*
pET-28a(+)_CYP102A1_A328I	Focused mutant library of CYP102A1	A328I	pITB932*
pET-28a(+)_CYP102A1_A328L	Focused mutant library of CYP102A1	A328L	pITB933*
pET-28a(+)_CYP102A1_A328V	Focused mutant library of CYP102A1	A328V	pITB934*

Supplementary material

Table 7.2: Vector constructs used in this work (continued).

Name	Gene insert / Application	Point mutation	ITB No.
pET-28a(+)_CYP102A1_F87A	Focused mutant library of CYP102A1	F87A	pITB935*
pET-28a(+)_CYP102A1_F87A_A328F	Focused mutant library of CYP102A1	F87A, A328F	pITB936*
pET-28a(+)_CYP102A1_F87A_A328I	Focused mutant library of CYP102A1	F87A, A328I	pITB937*
pET-22b(+)_CYP102A1_F87A_A328L	Focused mutant library of CYP102A1	F87A, A328L	pITB938*
pET-28a(+)_CYP102A1_F87A_A328V	Focused mutant library of CYP102A1	F87A, A328V	pITB939*
pET-28a(+)_CYP102A1_F87I	Focused mutant library of CYP102A1	F87I	pITB940*
pET-28a(+)_CYP102A1_F87I_A328F	Focused mutant library of CYP102A1	F87I, A328F	pITB941*
pET-28a(+)_CYP102A1_F87I_A328I	Focused mutant library of CYP102A1	F87I, A328I	pITB942*
pET-22b(+)_CYP102A1_F87I_A328L	Focused mutant library of CYP102A1	F87I, A328L	pITB943*
pET-28a(+)_CYP102A1_F87I_A328V	Focused mutant library of CYP102A1	F87I, A328V	pITB944*
pET-28a(+)_CYP102A1_F87L	Focused mutant library of CYP102A1	F87L	pITB945*
pET-28a(+)_CYP102A1_F87L_A328F	Focused mutant library of CYP102A1	F87L, A328F	pITB946*
pET-28a(+)_CYP102A1_F87L_A328I	Focused mutant library of CYP102A1	F87L, A328I	pITB947*
pET-22b(+)_CYP102A1_F87L_A328L	Focused mutant library of CYP102A1	F87L, A328L	pITB948*
pET-28a(+)_CYP102A1_F87L_A328V	Focused mutant library of CYP102A1	F87L, A328V	pITB949*
pET-28a(+)_CYP102A1_F87V	Focused mutant library of CYP102A1	F87V	pITB950*
pET-28a(+)_CYP102A1_F87V_A328F	Focused mutant library of CYP102A1	F87V, A328F	pITB951*
pET-28a(+)_CYP102A1_F87V_A328I	Focused mutant library of CYP102A1	F87V, A328I	pITB952*
pET-22b(+)_CYP102A1_F87V_A328L	Focused mutant library of CYP102A1	F87V, A328L	pITB953*
pET-22b(+)_CYP102A1_F87V_A328V	Focused mutant library of CYP102A1	F87V, A328V	pITB954*

Supplementary material

Table 7.2: Vector constructs used in this work (continued).

Name	Gene insert / Application	Point mutation	ITB No.
pET-22b(+)_CYP102A1_F87V_A328L_L437A	Triple mutant of CYP102A1	F87V, A328L, L437A	pITB955*
pET-22b(+)_CYP102A1_F87V_A328L_L437F	Triple mutant of CYP102A1	F87V, A328L, L437F	pITB956*
pET-22b(+)_CYP102A1_F87V_A328L_L437I	Triple mutant of CYP102A1	F87V, A328L, L437I	pITB957*
pET-22b(+)_CYP102A1_F87V_A328L_L437V	Triple mutant of CYP102A1	F87V, A328L, L437V	pITB958*
pET-28a(+)_CYP102A1_R47L_Y51F	Substrate channel mutation variant of CYP102A1	R47L, Y51F	pITB959
pET-28a(+)_CYP102A1_R47L_Y51F_F87V	Substrate channel mutation variant of CYP102A1	R47L, Y51F, F87V	pITB960
pET-28a(+)_CYP102A1_R47L_Y51F_A328L	Substrate channel mutation variant of CYP102A1	R47L, Y51F, A328L	pITB961
pET-28a(+)_CYP102A1_R47L_Y51F_F87V_A328F	Substrate channel mutation variant of CYP102A1	R47L, Y51F, F87V, A328F	pITB962*
pET-28a(+)_CYP102A1_R47L_Y51F_F87V_A328I	Substrate channel mutation variant of CYP102A1	R47L, Y51F, F87V, A328I	pITB963*
pET-28a(+)_CYP102A1_R47L_Y51F_F87V_A328L	Substrate channel mutation variant of CYP102A1	R47L, Y51F, F87V, A328L	pITB964*
pET-28a(+)_CYP102A1_R47L_Y51F_F87V_A328V	Substrate channel mutation variant of CYP102A1	R47L, Y51F, F87V, A328V	pITB965*
pET-22b(+)_CYP102A1_R47L_Y51F_F87V_A328L_L437I	Substrate channel mutation variant of CYP102A1	R47L, Y51F, F87V, A328L, L437I	pITB966
pMK-RQ_VAO	Codon-optimized vanillyl alcohol oxidase VAO from <i>P. simplicissimum</i> CBS 170.90 in GeneArt cloning vector	-	pITB976
pET-22b(+)_VAO	Codon-optimized vanillyl alcohol oxidase VAO from <i>P. simplicissimum</i> CBS 170.90	-	pITB977
pET-22b(+)_VAO_F454Y	Codon-optimized vanillyl alcohol oxidase VAO from <i>P. simplicissimum</i> CBS 170.90	F454Y	pITB980
pET-22b(+)_VAO_T505S	Codon-optimized vanillyl alcohol oxidase VAO from <i>P. simplicissimum</i> CBS 170.90	T505S	pITB983

* Provided by Dr. Alexander Seifert (ITB, University of Stuttgart, Stuttgart, Germany).

7.1.4 Primers

Oligonucleotides were obtained from Metabion (Martinsried, Germany) in salt-free form. The lyophilized oligonucleotides were resuspended in TE buffer pH 8.0 and diluted 1:10 with deionized water prior to use.

Table 7.3: Primers used in this work for PCR amplification of His-tag variants of CYP116B3.

His-tag cloning based on pET-24a(+)_CYP116B3		
Application	Primer	Sequence (5' → 3')
Cloning of CYP116B3 _Var1	F	C CGC <i>catatg</i> <i>GGC AGC AGC</i> <u>CAT CAT CAT CAT CAC CAT</u> AGC
	R	GGT <i>gaattc</i> TTA CAG ACG CAG GGT CAG ACG ATC ACC ACA TGC
Cloning of CYP116B3 _Var2	F	C CGC <i>catatg</i> AGC GCA AGC GTT CCG GCA AGC GC
	R	GGT <i>gaattc</i> TTA <u>ATG GTG ATG ATG ATG ATG</u> CAG ACG CAG GGT CAG ACG ATC
Cloning of CYP116B3 _Var3	F	C CGC <i>catatg</i> AGC GCA AGC GTT CCG GCA AGC GC
	R	GGT <i>gaattc</i> TTA CAG ACG CAG GGT CAG ACG ATC ACC ACA TGC

Underlined, His-tag; italic, MGSS-sequence; lower case, restriction sites (*catatg*, *NdeI*; *gaattc*, *EcoRI*); F, forward; R, reverse.

Supplementary material

Table 7.4: Primers used in this work for site-directed mutagenesis.

Site-directed mutagenesis of MobA from <i>C. testosteroni</i> GZ39		
Mutation	Sequence (5' → 3')	
V257A	F	G AGC GAA CAG GGT AAT GCA CTG ATT ATT CCG CGT GAA GG
	R	CC TTC ACG CGG AAT AAT CAG TGC ATT ACC CTG TTC GCT C
Site-directed mutagenesis of CYP116B3 from <i>R. ruber</i> DSM 44319		
Mutation	Sequence (5' → 3')	
A109F	F	CGT TAT GAT TAT GCA ATG TTT CGT ACC CTG GTG AAT GAA G
	R	C TTC ATT CAC CAG GGT ACG AAA CAT TGC ATA ATC ATA ACG
A109I	F	CGT TAT GAT TAT GCA ATG ATT CGT ACC CTG GTG AAT GAA G
	R	C TTC ATT CAC CAG GGT ACG AAT CAT TGC ATA ATC ATA ACG
A109L	F	CGT TAT GAT TAT GCA ATG CTG CGT ACC CTG GTG AAT GAA G
	R	C TTC ATT CAC CAG GGT ACG CAG CAT TGC ATA ATC ATA ACG
A109V	F	CGT TAT GAT TAT GCA ATG GTT CGT ACC CTG GTG AAT GAA G
	R	C TTC ATT CAC CAG GGT ACG AAC CAT TGC ATA ATC ATA ACG
V316A	F	CGT CAT AGC GGT AGC GCA GCA GCA TGG CGT CGT G
	R	C ACG ACG CCA TGC TGC TGC GCT ACC GCT ATG ACG
V316F	F	CGT CAT AGC GGT AGC TTT GCA GCA TGG CGT CGT G
	R	C ACG ACG CCA TGC TGC AAA GCT ACC GCT ATG ACG
V316I	F	CGT CAT AGC GGT AGC ATT GCA GCA TGG CGT CGT G
	R	C ACG ACG CCA TGC TGC AAT GCT ACC GCT ATG ACG
V316L	F	CGT CAT AGC GGT AGC CTG GCA GCA TGG CGT CGT G
	R	C ACG ACG CCA TGC TGC CAG GCT ACC GCT ATG ACG
A317F	F	CAT AGC GGT AGC GTT TTT GCA TGG CGT CGT GTT G
	R	C AAC ACG ACG CCA TGC AAA AAC GCT ACC GCT ATG
A317I	F	CAT AGC GGT AGC GTT ATT GCA TGG CGT CGT GTT G
	R	C AAC ACG ACG CCA TGC AAT AAC GCT ACC GCT ATG

Supplementary material

Table 7.4: Primers used in this work for site-directed mutagenesis (continued).

Site-directed mutagenesis of CYP116B3 from <i>R. ruber</i> DSM 44319		
Mutation	Sequence (5' → 3')	
A317L	F	CAT AGC GGT AGC GTT CTG GCA TGG CGT CGT GTT G
	R	C AAC ACG ACG CCA TGC CAG AAC GCT ACC GCT ATG
A317V	F	CAT AGC GGT AGC GTT GTT GCA TGG CGT CGT GTT G
	R	C AAC ACG ACG CCA TGC AAC AAC GCT ACC GCT ATG
A318F	F	CAT AGC GGT AGC GTT GCA TTT TGG CGT CGT GTT GCA ACC
	R	GGT TGC AAC ACG ACG CCA AAA TGC AAC GCT ACC GCT ATG
A318I	F	CAT AGC GGT AGC GTT GCA ATT TGG CGT CGT GTT GCA ACC
	R	GGT TGC AAC ACG ACG CCA AAT TGC AAC GCT ACC GCT ATG
A318L	F	CAT AGC GGT AGC GTT GCA CTG TGG CGT CGT GTT GCA ACC
	R	GGT TGC AAC ACG ACG CCA CAG TGC AAC GCT ACC GCT ATG
A318V	F	CAT AGC GGT AGC GTT GCA GTT TGG CGT CGT GTT GCA ACC
	R	GGT TGC AAC ACG ACG CCA AAC TGC AAC GCT ACC GCT ATG
F420A	F	CTG CCG AAT ACC AGC GCA CGT GGT CCG GAT CAT G
	R	C ATG ATC CGG ACC ACG TGC GCT GGT ATT CGG CAG
F420I	F	CTG CCG AAT ACC AGC ATT CGT GGT CCG GAT CAT G
	R	C ATG ATC CGG ACC ACG AAT GCT GGT ATT CGG CAG
F420L	F	CTG CCG AAT ACC AGC CTG CGT GGT CCG GAT CAT G
	R	C ATG ATC CGG ACC ACG CAG GCT GGT ATT CGG CAG
F420V	F	CTG CCG AAT ACC AGC GTT CGT GGT CCG GAT CAT G
	R	C ATG ATC CGG ACC ACG AAC GCT GGT ATT CGG CAG
Site-directed mutagenesis of CYP102A1 from <i>B. megaterium</i> ATCC 14581		
Mutation	Sequence (5' → 3')	
R47L	F	C GAG GCG CCT GGT CTG GTA ACG CGC
	R	GCG CGT TAC CAG ACC AGG CGC CTC G

Supplementary material

Table 7.4: Primers used in this work for site-directed mutagenesis (continued).

Site-directed mutagenesis of CYP102A1 from <i>B. megaterium</i> ATCC 14581		
Mutation	Sequence (5' → 3')	
Y51F	F	CT GGT CGT GTA ACG CGC TTC TTA TCA AGT CAG
	R	CTG ACT TGA TAA GAA GCG CGT TAC ACG ACC AG
F87V	F	GCA GGA GAC GGG TTA GTA ACA AGC TGG ACG CAT G
	R	C ATG CGT CCA GCT TGT TAC TAA CCC GTC TCC TGC
A328L	F	CGC TTA TGG CCA ACT CTG CCT GCG TTT TC
	R	GA AAA CGC AGG CAG AGT TGG CCA TAA GCG
Site-directed mutagenesis of VAO from <i>P. simplicissimum</i> CBS 170.90		
Mutation	Sequence (5' → 3')	
F454Y	F	GAA GCA GGT CTG GAT TAT ATT GGC ACC TTT ACC
	R	GGT AAA GGT GCC AAT ATA ATC CAG ACC TGC TTC
T505S	F	GG GGT GAA TAT CGT AGC CAT CTG GCA TTT ATG
	R	CAT AAA TGC CAG ATG GCT ACG ATA TTC ACC CC

Bold, codon of exchanged amino acid; F, forward; R, reverse.

7.1.5 Strains

E. coli strain DH5 α (formerly Invitrogen now Life Technologies, Darmstadt, Germany) was employed for cloning and plasmid amplification. As host for recombinant gene expression the *E. coli* strain BL21(DE3) (Novagen, Madison, WI, USA) was used.

Table 7.5: Bacterial strains.

Strain	Genotype description	Manufacturer (Country)
<i>Escherichia coli</i> DH5 α	F ⁻ supE44 Δ lacU169 (ϕ 80lacZ Δ M15) hsdR17 recA1 endA1 gyrA96 thi-1relA1	Invitrogen (now Life Technologies, Darmstadt, Germany)
<i>Escherichia coli</i> BL21(DE3)	F ⁻ ompT hsdS _B (r _B ⁻ m _B ⁻) gal dcm (DE3)	Novagen (Madison, WI, USA)

7.2 Supplementary tables

7.2.1 Media, buffers and solutions

Table 7.6: Culture media.

Medium	Component	Amount
LB broth	Yeast extract	5 g
	NaCl	5 g
	Tryptone	10 g
	dH ₂ O	to 1 l
LB agar	pH 7	
	Agar	15 g
SOC	Yeast extract	5 g
	NaCl	0.5 g
	Tryptone	20 g
	250 mM KCl	10 ml
	Glucose	20 mM
	dH ₂ O	to 1 l

Culture media were autoclaved at 121°C for 20 min. For the preparation of media supplemented with antibiotics, stock solutions of ampicillin or kanamycin were prepared, sterilized by filtration using a 0.2 µm sterile filter and added to the medium after autoclaving and cooling to ca. 45°C to yield a final concentration of 100 µg/ml or 30 µg/ml, respectively.

Supplementary material

Table 7.7: Buffers for preparation of competent cells.

Buffer	Component	Amount
Tfb I	Potassium acetate RbCl CaCl ₂ MnCl ₂ · 4 H ₂ O Glycerol (87%) adjust to pH 5.8 with acetic acid dH ₂ O	0.59 g 2.42 g 0.29 g 2.00 g 30 ml to 200 ml
Tfb II	MOPS CaCl ₂ RbCl Glycerol (87%) adjust to pH 6.5 with NaOH dH ₂ O	0.21 g 1.10 g 0.12 g 15 ml to 100 ml

Tfb I and Tfb II buffers were sterilized by filtration using a 0.2 µm sterile filter.

Table 7.8: Buffers for DNA reconstitution.

Buffer	Component	Amount
TE, pH 8.0	1 M Tris-HCl, pH 8.0 EDTA dH ₂ O	1 ml 37.22 mg to 100 ml

Table 7.9: Buffers and solutions for agarose gel electrophoresis.

Buffer	Component	Amount
1% Agarose gel solution	Agarose 1x TAE	5 g to 500 ml
TAE buffer (50x)	Tris Glacial acetic acid 0.5 M EDTA (pH 8.0) dH ₂ O	242 g 57 ml 100 ml to 1 l
DNA sample loading buffer	Urea EDTA Glycerol (87%) Bromophenol blue dH ₂ O	12.01 g 0.21 g 25 ml 50 mg to 1 l

Supplementary material

Table 7.10: Buffers and solutions for SDS-PAGE.

Buffer	Component	Amount
SDS sample loading buffer (2x)	MgCl ₂ SDS Glycerol (87%) Bromophenol blue 2-Mercaptoethanol 1 M Tris-HCl (adjust to pH 6.8 with HCl) dH ₂ O	190 mg 0.8 g 2 ml 2 mg 1 ml 2 ml to 20 ml
4x Lower Tris	Tris SDS adjust to pH 8.8 with HCl dH ₂ O	36.34 g 0.8 g to 200 ml
4x Upper Tris	Tris SDS adjust to pH=6.8 with HCl dH ₂ O	12.11 g 0.8 g to 200 ml
10% APS	APS dH ₂ O	100 mg to 1 m
Resolving gel (12.5% Acrylamide)	Acrylamide (30%) 4x Lower Tris dH ₂ O 10% APS TEMED	3.33 ml 2.00 ml 2.67 ml 40 µl 4 µl
Stacking gel (4% Acrylamide)	Acrylamide (30%) 4x Upper Tris dH ₂ O 10% APS TEMED	0.52 ml 1.00 ml 2.47 ml 40 µl 4 µl
Running buffer	Tris SDS Glycine dH ₂ O	30 g 10 g 144 g to 2 l
Staining solution	Coomassie® Brilliant Blue Methanol Glacial acetic acid dH ₂ O	500 mg 150 ml 50 ml to 500 ml
Destaining solution	Methanol Glacial acetic acid dH ₂ O	300 ml 100 ml to 1000 ml

Supplementary material

Table 7.11: Buffers for IMAC purification.

Medium	Component	Amount
1 M Tris-HCl, pH 7.5	Tris dH ₂ O adjust to pH 7.5 with 37% HCl	121.14 g to 1 l
50 mM Tris-HCl, pH 7.5	1 M Tris-HCl, pH 7.5 dH ₂ O	25 ml to 500 ml
Equilibration buffer → 50 mM Tris-HCl, pH 7.5 → 500 mM NaCl → 5% Glycerol → 0.1 mM PMSF	1 M Tris-HCl, pH 7.5 NaCl Glycerol (87%) 100 mM PMSF dH ₂ O	50 ml 29.22 g 57.5 ml 1 ml to 1 l
Elution buffers	Imidazol Buffer B	20-200 mM to 100 ml

Table 7.12: Buffer for dialysis.

Medium	Component	Amount
Dialysis buffer → 50 mM Tris-HCl, pH 7.5 → 5% Glycerol → 0.1 mM PMSF	1 M Tris-HCl, pH 7.5 Glycerol (87%) 100 mM PMSF dH ₂ O	75 ml 86 ml 1.5 ml to 1.5 l

7.2.2 *In vitro* biotransformations**Table 7.13:** Primary screening for the conversion of 3-methylanisole by all members of the CYP102A1 minimal mutant library.

Amino acid position		87				
		A	F	I	L	V
328	A	++	+ ^a	+	+	+++
	F	++	++	++	+	+++
	I	++	++	++	++	+++
	L	+	+++	++	-	+++
	V	+	++	++	-	+++

^aCYP102A1 wild type -, no conversion +, conversion < 5% ++, conversion < 25% +++, conversion > 25%

Samples were analyzed by HPLC. Negative controls showed no conversion.

7.3 Supplementary figures

7.3.1 Protein expression

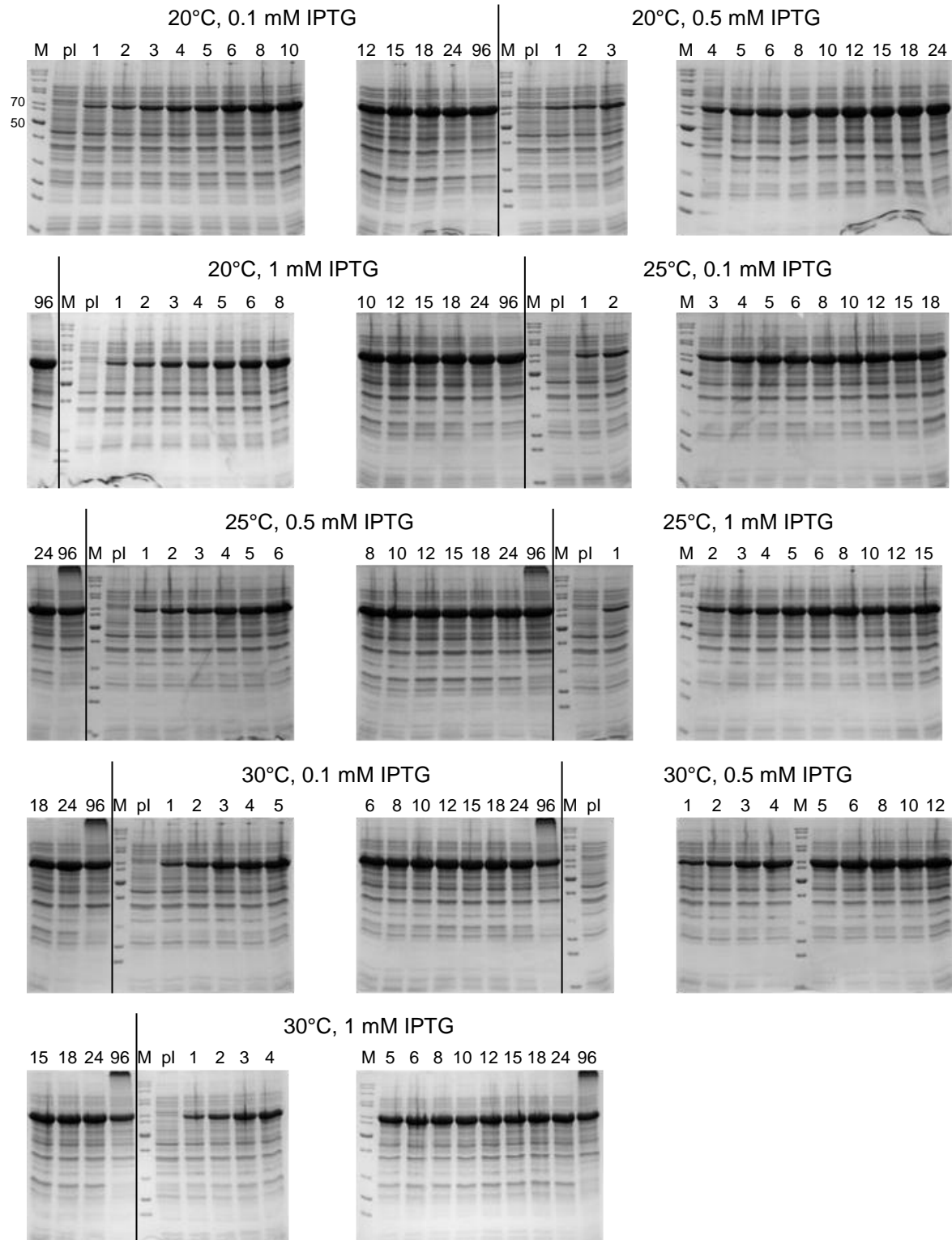


Figure 7.2: SDS-PAGE analysis of the MobA expression study samples. Expression was investigated after induction with 0.1, 0.5 or 1 mM IPTG and incubation at 20, 25 and 30°C with shaking at 180 rpm. M, marker; pl, prior to induction; numbers 1 to 96 indicate the total expression time in hours of samples taken after induction.

Supplementary material

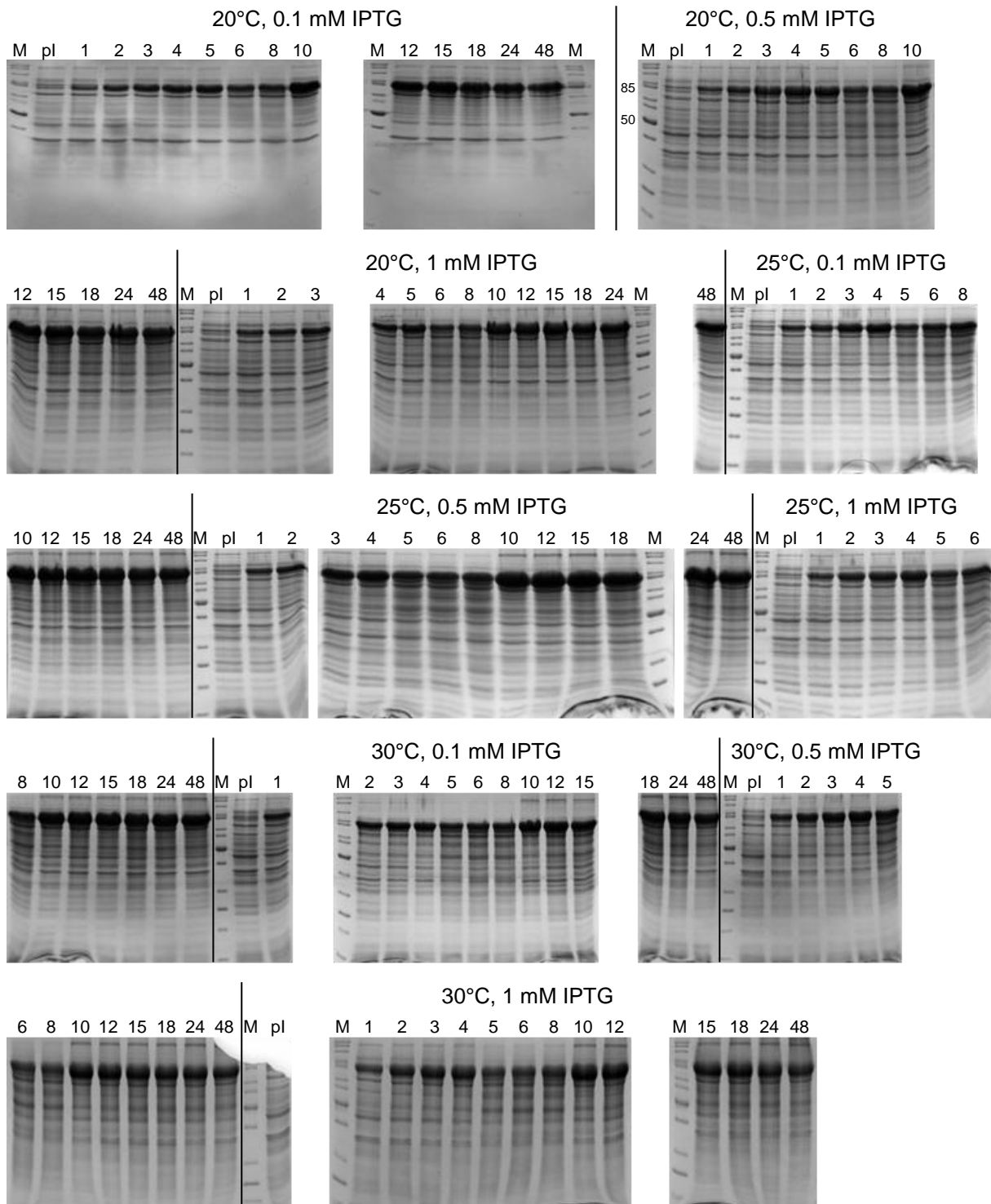


Figure 7.3: SDS-PAGE analysis of the CYP116B3 expression study samples. Expression was investigated after induction with 0.1, 0.5 or 1 mM IPTG and incubation at 20, 25 and 30°C with shaking at 180 rpm. M, marker; pl, prior to induction; numbers 1 to 48 indicate the total expression time in hours of samples taken after induction.

Supplementary material

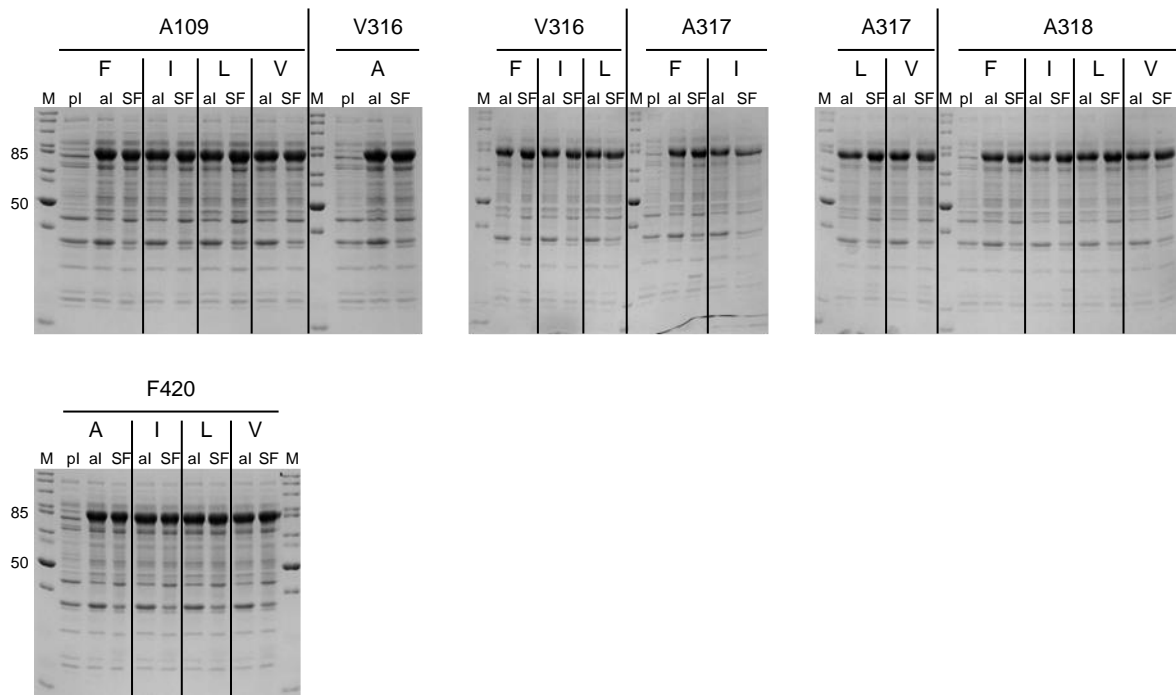


Figure 7.4: SDS-PAGE analysis of the expression of the CYP116B3 focused mutant library variants (A109F/I/L/V, V316A/F/I/L, A317F/I/L/V, A318F/I/L/V, F420A/I/L/V). Expression was investigated after induction with 0.1 mM IPTG and incubation at 20°C with shaking at 140 rpm for 18 h. Whole cell samples (pl and al) were adjusted to an OD_{600} value of 0.25 and an equal volume of 7.5 μ l was loaded onto the gel. For soluble protein fractions, 7.5 μ g of protein was loaded in the respective lanes.

7.3.2 Molecular dynamics simulations

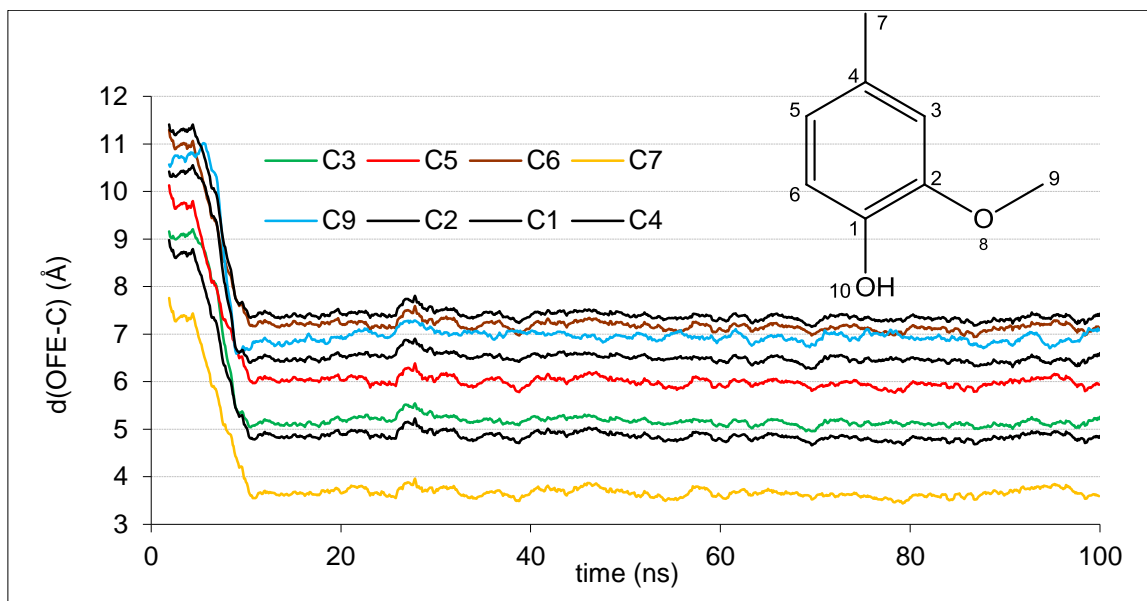


Figure 7.5: Distance of 4-methylguaicol carbon atoms from the activated oxygen of the CYP102A1 heme in a 100 ns MD run. The molecule approaches the heme oxygen to a distance <4 Å (C7) where it stays stable for 90 ns. Lines show a moving average of 20 data points.

Supplementary material

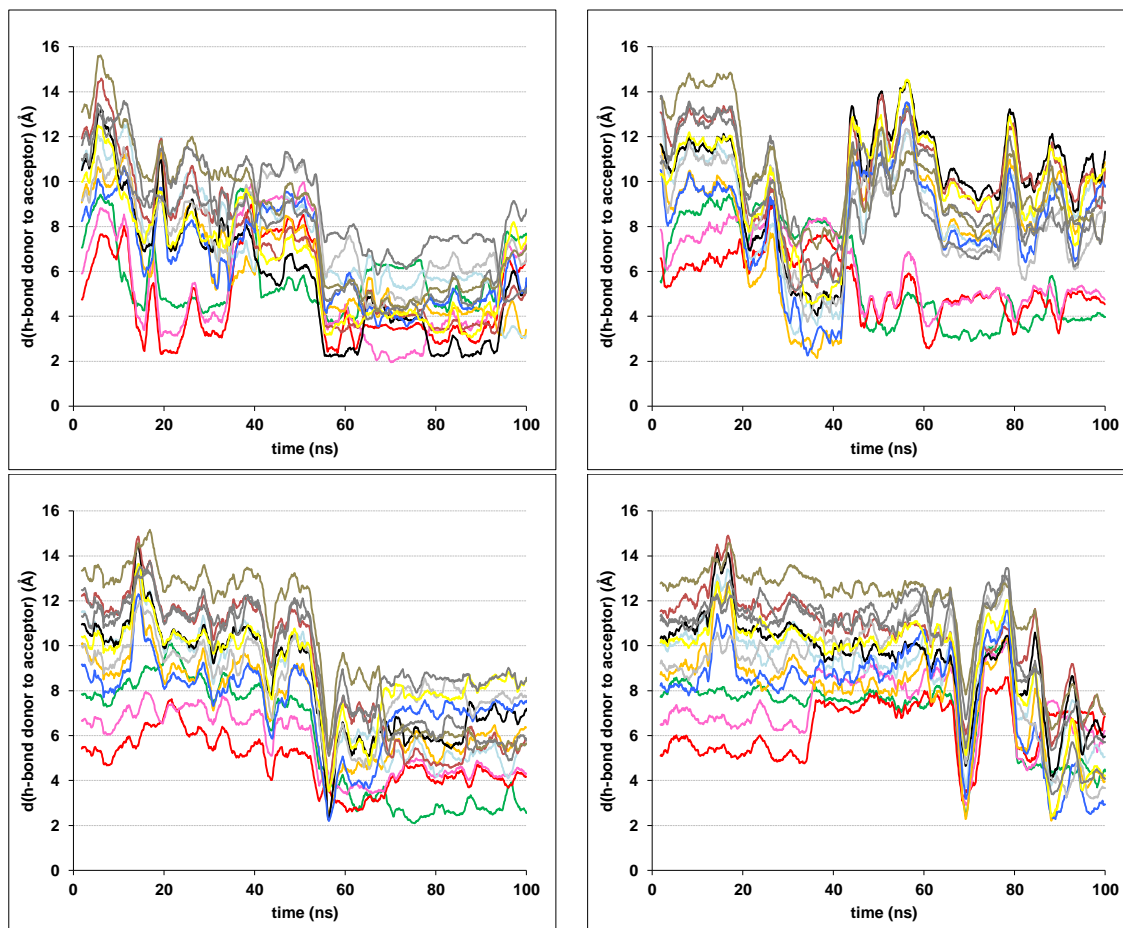


Figure 7.6: Distance between all possible hydrogen bond donors and acceptors of 4-methylguaiacol and the side chain of Y51 (green, red and magenta lines) and R47 (all remaining lines) in 4 out of 15 MD simulations. A hydrogen bond is considered present at a hydrogen bond donor to acceptor distance ≤ 3.5 Å. Lines show a moving average of 20 data points.

7.3.3 GC-MS analysis

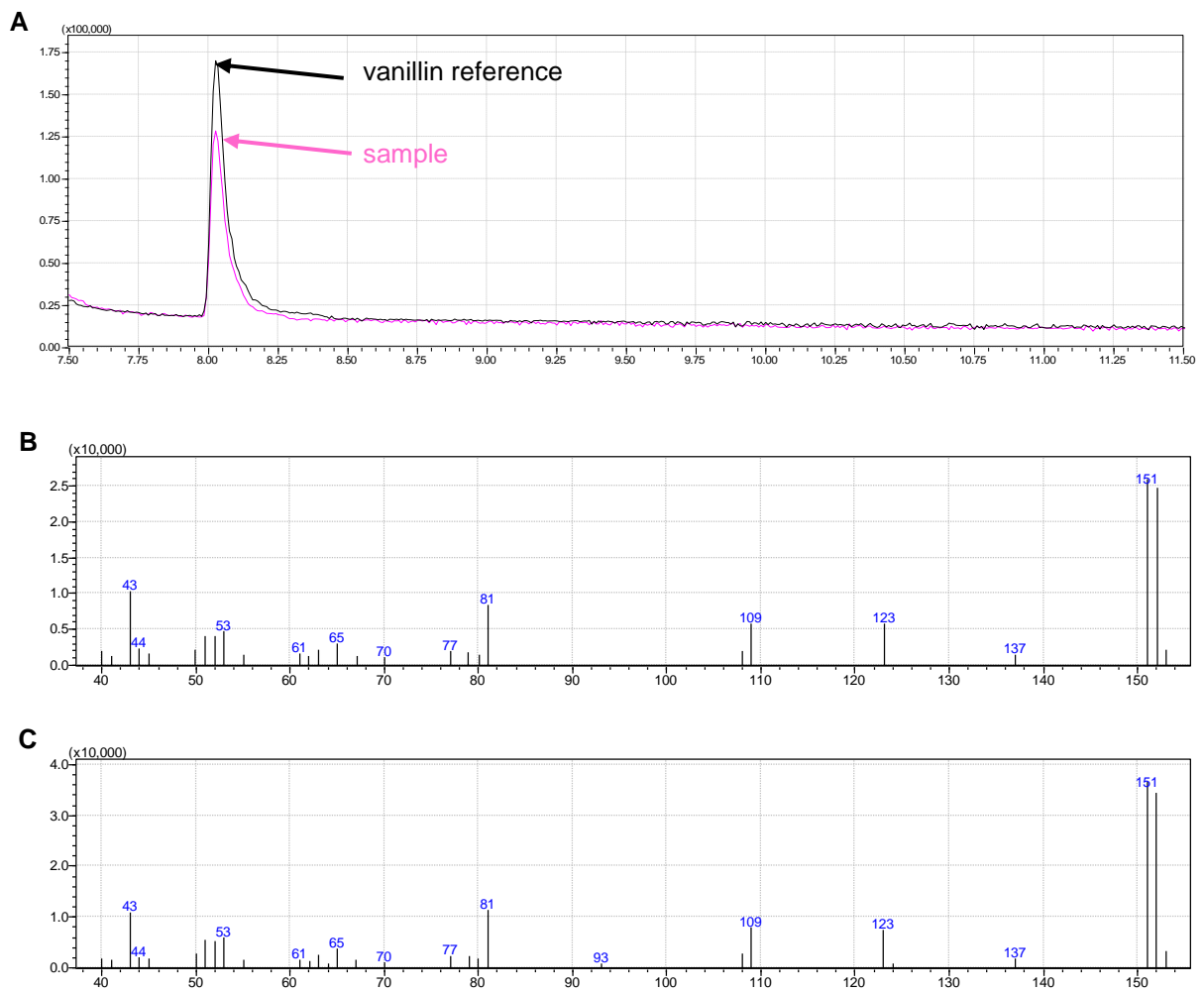


Figure 7.7: Chromatogram (A) and fragmentation pattern (B) of GC-MS analysis of the *in vivo* biotransformation sample after 12 h conversion time. (C) Fragmentation pattern of the GC-MS analysis of a vanillin reference (0.05 mM vanillin).