Identification of genes involved in the biosynthesis of lignans in *Linum flavum*

Dissertation

Zur

Erlangung des Doktorgrades

der Naturwissenschaften

(Dr. rer. nat.)

dem

Fachbereich der Pharmazie

der Philipps-Universität Marburg

vorgelegt von

Thanh Son Ta

aus Gialai/Vietnam

Marburg/Lahn 2019

Erstgutachter: Prof. Dr. Maike Petersen

Zweitgutachter: Prof. Dr. Andreas Heine

Eingereicht am 24.04.2019

Tag der mündlichen Prüfung am 06.06.2019

Hochschulkennziffer: 1180

ERKLÄRUNG

Ich versichere, dass ich meine Dissertation

„Identification of genes involved in the biosynthesis of lignans in *Linum flavum*“

selbständig ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen bedient habe. Alle vollständig oder sinngemäß übernommenen Zitate sind als solche gekennzeichnet.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, den 24.04.2019

Thanh Son Ta

Acknowledgements

After three years of exciting research and joyful moments, I have reached the end of my PhD journey.

Hereby I would like to thank my PhD supervisor Prof. Dr. Maike Petersen for her support during this thesis. For three years, I made many mistakes and each time, she was always willing to lend a helping hand to me. I really appreciate her encouragement and advice throughout this research and her big smile will be the memory I will never forget.

I would like to express my gratitude to FAZIT-Stiftung for funding scholarship during my PhD and helping me to pursue my dream of doing scientific research.

I am very grateful to Prof. Dr. Andreas Heine for being the co-supervisor of my thesis.

Furthermore, I would like to express my sincere appreciation to the current and former colleagues in the Petersen working group for their support and help, including Elke Bauerbach, Dr. Lennart Poppe, Dr. Agus Chahyadi, Julia Wohl, Tobias Busch, Olga Haag, Sandra Dietzler, Lucien Ernst, Dr. Jennifer Robinson, Dr. Victoria Werner, Anne Jahn.

I would also like to express my gratitude to the employees of the Institute of Pharmaceutical Biology and Biotechnology Marburg and the former and current colleagues of the Li working group for the good companionship and joyful atmosphere.

Many thanks to my Vietnamese friends in Germany. Nearly ten years of joy and sadness, we always have each other and overcome many challenges. Our brotherhood makes this country feels like home.

Special thanks go to my parents, my brothers and my sisters for encouraging and supporting me in pursuing my scientific goals and developing my potential.

Finally, I want to address my appreciation to my wife, Thi Kieu Loan Do. You are the best gift that God has given to me. The patience and perseverance that you give me will be the driving force for me to strive. I am lucky to have you with me on the road ahead and I am sure that a bright future awaits our family.

Publications

Thanh Son T., Petersen M. (2018): Identification of genes involved in the biosynthesis of lignans in *Linum flavum*. Meeting of the section "Natural Products", Deutsche Botanische Gesellschaft, Burg Warberg (Oral)

Thanh Son T., Petersen M. (2018): Identification of genes involved in the biosynthesis of lignans in *Linum flavum*. Seminar of Pharmaceutical Biology and Biotechnology Institute Marburg, Marburg (Oral)

Thanh Son T., Petersen M. (2017): Identification of genes of deoxypodophyllotoxin 6-hydroxylase and deoxypodophyllotoxin 7-hydroxylase in *Linum flavum.* International Plant Science Conference, Botanikertagung, Kiel (Poster)

Thanh Son T., Petersen M. (2017): Identification of genes of deoxypodophyllotoxin 6-hydroxylase and deoxypodophyllotoxin 7-hydroxylase in *Linum flavum*. Seminar of Pharmaceutical Biology and Biotechnology Institute Marburg, Marburg (Oral)

I. Table of Content

II. Abbreviations

Measures and units

aa = amino acid	bp = base pair(s)	U = units (enzyme activity)
M = molar, mol l^{-1}	OD_{600} = optical density at 600 nm	rpm = revolutions per minute
kDa = kilo-Dalton	Tm = melting point in °C	

Chemicals

ABTS = 2,2'-azino-di-(3-ethylbenzthiazoline-6-sulfonic acid)
AcOH = acetic acid
APS = ammonium persulphate
BCIP = 5-bromo-4-chloro-3-indolyl phosphate
DTT = dithiothreitol
DOP = deoxypodophyllotoxin
EDTA = ethylenediaminetetraacetic acid
EtOH = ethanol
EtOAc = ethyl acetate
KPi = potassium phosphate buffer
LARI = lariciresinol
MATAI = matairesinol
MeOH = methanol
6-MPTOX = 6-methoxypodophyllotoxin
NBT = nitro blue tetrazolium chloride
PINO = pinoresinol
PTOX = podophyllotoxin
SDS = sodium dodecyl sulphate
SECO = secoisolariciresinol
TEMED = tetramethylethylenediamine
TRIS = tris(hydroxymethyl)aminomethane

Species names and abbreviations

E. coli = *Escherichia coli*, Enterobacteriaceae
Fi = *Forsythia x intermedia*, forsythia, Oleaceae
La = *Linum album*, Linaceae
Lf = *Linum flavum*, Linaceae
Lu = *Linum usitatissimum*, flax, Linaceae
Lp = *Linum perenne*, Linaceae
S. cerevisiae = *Saccharomyces cerevisiae*, Saccharomycetaceae
Tp = *Thuja plicata*, western red cedar, Cupressaceae

Genes and proteins

3H = 4-coumaroylshikimate 3-hydroxylase
4CL = hydroxycinnamic acid CoA ligase
BSA = bovine serum albumin
C4H = cinnamate 4-hydroxylase
CAD = cinnamyl alcohol dehydrogenase
CCoAOMT = caffeoyl-CoA *O*-methyltransferase
CCR = cinnamoyl CoA:NADP oxidoreductase
COMT = caffeic acid *O*-methyltransferase
CPR = cytochrome P450 reductase
CYP = cytochrome P450
DIR = dirigent protein
DOP6H = deoxypodophyllotoxin 6-hydoxylase
DOP7H = deoxypodophyllotoxin 7-hydoxylase
HCT = shikimate *O*-hydroxycinnamoyltransferase
PAL = phenylalanine ammonia-lyase
PLR = pinoresinol-lariciresinol reductase
SDH = secoisolariciresinol dehydrogenase

Nucleic acids and nucleotides

A = adenine C = cytosine G = guanine T = thymine U = uracil
cDNA = complementary DNA
dNTP = deoxynucleotide triphosphate
mRNA = messenger RNA
gDNA = genomic DNA

Proteinogenic amino acids

A = alanine (Ala)	C = cysteine (Cys)	D = aspartic acid (Asp)
E = glutamic acid (Glu)	F = phenylalanine (Phe)	G = glycine (Gly)
H = histidine (His)	I = isoleucine (Ile)	K = lysine (Lys)
L = leucine (Leu)	M = methionine (Met)	N = asparagine (Asn)
P = proline (Pro)	Q = glutamine (Gln)	R = arginine (Arg)
S = serine (Ser)	T = threonine (Thr)	V = valine (Val)
W = tryptophan (Trp)	Y = tyrosine (Tyr)	

III. Introduction

1. Lignans – Occurrence and general structure

Lignans belong to the group of polyphenolic substances and are derived from L-phenylalanine. Hydroxycinnamyl alcohols termed monolignols, usually coniferyl alcohol, are precursors for the biosynthesis of lignans. Two molecules of these phenylpropanes dimerize stereospecifically via a C-C linkage. If the bond is formed via the C8-C8' atoms of the side chains of monolignols, the compounds are called lignans. If the monolignols are linked in a different way, they belong to neolignans. In addition, norlignans lack the C9 or the C9' atom or a methoxy group on the aromatic ring. Lignans occur almost exclusively in plants. They are found in more than 55 plant families of both gymnosperms and angiosperms, as well as in mosses and ferns (Lewis and Davin, 1999). Juniperaceae, Cupressaceae, some *Linum* species (Linaceae), *Podophyllum* species (Berberidaceae), Polygalaceae, Apiaceae, Pinaceae, Hernandaceae, Euphorbiaceae, and Ranunculaceae contain significant amounts of lignans (Imbert, 1998).

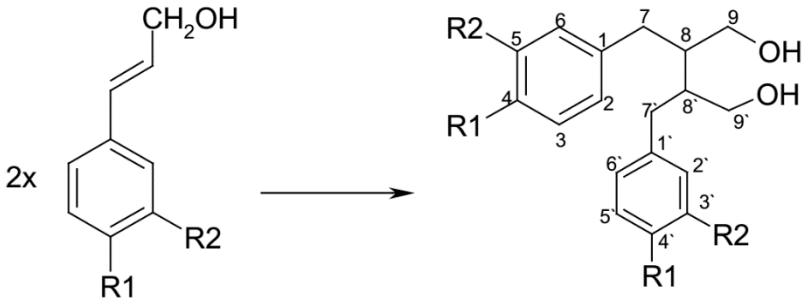
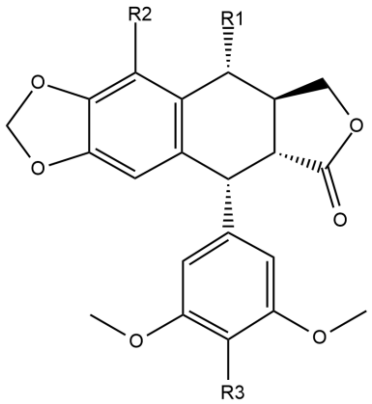


Figure 1: Lignan backbone with 8-8' linkage of two monolignols

2. Biological activity of lignans

Lignans show a wide range of activities in plants, mainly in the defense against pathogens and herbivores. There are antifungal, antiviral and antibacterial lignans (Deyama and Nishibe, 2010). Hence, lignans are interesting for application as medicinal products in humans. Secoisolariciresinol and matairesinol, the main lignans from linseed, are converted in the intestine by bacteria to the hormone-like structures enterolactone and enterodiol, which have a protective effect against hormone-dependent cancer, such as breast cancer. This protective effect has been demonstrated in clinical studies (Adlercreutz, 1999; Cho et al., 1999).

Podophyllotoxin (PTOX) is an aryltetralin-type lignan (Figure 2). It is strongly cytotoxic and was first isolated from roots and rhizomes of plants of the genus *Podophyllum*, whose representatives are found in North America (*P. peltatum*) and the Himalayan region (*P. hexandrum*, synonym *P. emodi*). Both species are known as traditional medicinal and/or poisonous plants and have been used for a variety of purposes. For example, the natives of North America used aqueous extracts from *P. peltatum* as laxans, cathartics and anthelmintics. Among the European immigrants, *Podophyllum* extracts were used as emetics, cathartics, and cholagogum (Imbert, 1998; Lloyd, 1910).



	R1	R2	R3
Podophyllotoxin	OH	H	OCH_3
6-Methoxypodophyllotoxin	OH	OCH_3	OCH_3
Deoxypodophyllotoxin	H	H	OCH_3
α-Peltatin	H	OH	OH
β-Peltatin	H	OH	OCH_3
ß-Peltatin A methyl ether	H	OCH_3	OCH_3

Figure 2: Aryltetralin lignan derivatives

Podophyllin, a resinous extract from Podophyllum roots and rhizomes, has been used against genital warts (*Condyloma acuminata*) in America since 1850 (Lloyd, 1910). In the 40s and 50s of the 20th century, PTOX was identified and isolated as an effective substance in podophyllin. PTOX binds to the α/β tubulin dimer and inhibits the construction of microtubules, thereby prevents mitotis (Canel et al., 2000). However, PTOX is too toxic for use as a cytostatic agent and is thus limited to external application. The semisynthetic derivatives Etoposide and Teniposide (Fig. 3) are formed by demethylation on the C4 atom of ring E, epimerization on the C4 atom of ring C, and by substitution at the OH groups. These derivatives are less toxic

but equally effective (Canel et al., 2000). Interestingly, Etoposide and Teniposide have a completely different mechanism of action compared to PTOX. They are inhibitors of topoisomerase II and thus prevent DNA replication. Today Etoposide, Etopophos® and Teniposide are used as cytostatics against most hormone-dependent types of cancer (leukaemia, ovarian, breast, pancreatic and lung cancer) and non-Hodgkin's lymphome. Etopophos® is Etoposide phosphate, a prodrug with better water solubility that is converted by alkaline phosphatase to Etoposide.

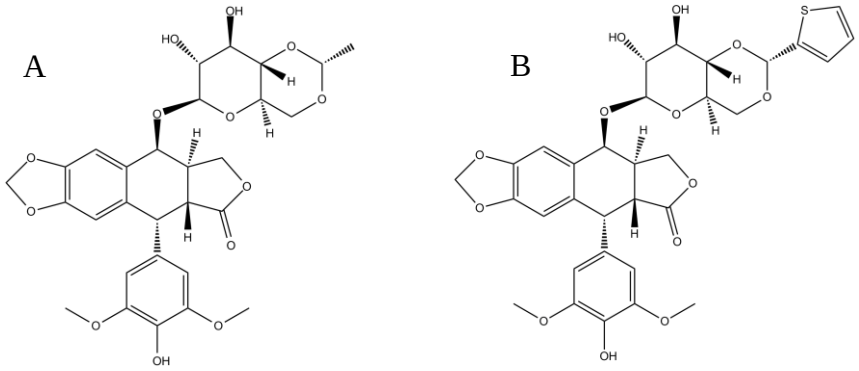


Figure 3: Podophyllotoxin derivatives with anticancer effect: (A) Etoposide, (B) Teniposide

3. Lignans in *Linum* and in plant cell cultures

3.1 Lignans in *Linum*

The genus *Linum* of the family Linaceae comprises about 230 representatives (Van Uden et al., 1994). Based on morphological and phytochemical data, the genus can be divided into different sections. There are different publications on the substructure of the genus *Linum*. Here, the genus *Linum* is divided into five sections according to Davis (1970) and Ockendon and Walters (1968). These sections are Linum, Syllinum, Dasylinum, Linastrum and Cathartolinum. Particularly interesting for this work are the representatives of the section Syllinum, since there are many species containing lignans of the PTOX-type. Within this section, there is a further division into three groups. The first group is perennial, has white flowers and produces mainly PTOX, whereas the second group contains perennial plants with yellow flowers which produce predominantly 6-MPTOX, and third group is annual, has yellow and homostylous flowers (Mohagheghzadeh et al., 2003; Broomhead et al.,1990; Weiss et al., 1975).

3.2 *Linum flavum* - description and distribution

Linum flavum (golden flax, yellow flax) is a species in the family Linaceae, section Syllinum. The plants are growing perennially with semi-evergreen leaves and five-petalled, yellow flowers. *L. flavum* has the chromosome number 2n = 30 (Erich, 2001). It prefers calcareous and nitrogen-poor, warm sites in a sunny to semi-shaded position. The species is common in Central and South-eastern Europe up to Central Russia in high altitude. The occurrence in Germany on the Swabian Alb and in the Illertal form the western edge of the area of *Linum flavum*. The plants are strongly endangered, only a few hundred exist here (Simon et al., 2002).

3.3 Lignans in plant cell cultures

Cell cultures can be obtained from seeds which are germinated on solid medium under sterile conditions or from plant material collected from nature after treatment with sterilizing agents. Callus formation can be obtained on hormone-containing culture media. Callus cells are mainly undifferentiated cells that can be cultivated on a solid medium containing macro- and micronutrients as well as a carbon source and suitable hormone concentrations (usually auxins and cytokinins) over several years and can serve as the starting culture for cell suspension cultures (Empt et al., 2000; Seidel et al., 2002; Smollny et al., 1998). A major disadvantage in the work with cell suspension cultures is their possible genetic instability in comparison to callus cultures, particularly regarding secondary metabolite production (Alfermann and Petersen, 1993; Deus-Neumann and Zenk, 1984). These changes are presumably due to the modification of the genetic material. In addition to the changes in the number of chromosomes, DNA methylations, genomic rearrangements and point mutations have been observed (Bayliss, 1973; Phillips et al., 1994; Sunderland, 1977).

Different species were used to produce PTOX and similar cytotoxic lignans in cell cultures. In *Linum* spec., the largest amount of PTOX (28 mg per litre after 11 days) was found in a suspension culture of *Linum album* (Smollny et al., 1998; Empt et al., 2000). 6-MPTOX was detected in the largest amount (121 mg per litre) in suspension cultures of *Linum flavum* (Berlin et al., 1986).

The investigations carried out in this study were made with suspension cultures of *Linum flavum*.

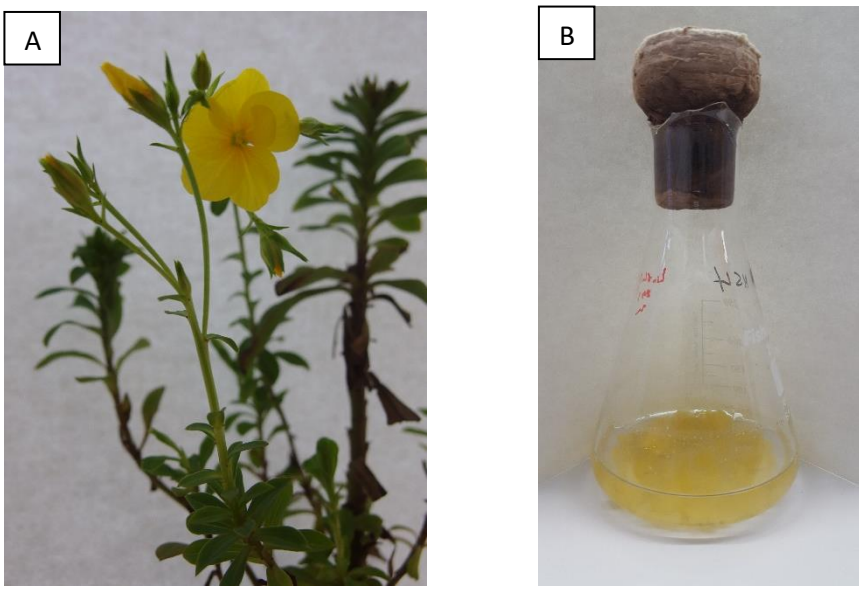


Figure 4: *Linum flavum* (A) and cell suspension culture of *L. flavum* (B)

4. Biosynthesis of lignans

4.1 General phenylpropanoid pathway

Phenylpropanoids are generally referred to as compounds which consist of a phenolic group with a bound C3 side chain and are derived from phenylalanine or tyrosine (Heldt, 1999). Since the formation of coniferyl alcohol is decisive for the synthesis of the aryltetralin lignans, only this biosynthetic route is described (Fraser and Chapple, 2011) (Fig. 5). Phenylalanine is non-oxidatively converted into trans-cinnamic acid by phenylalanine ammonia-lyase (PAL, E.C. 4.3.1.5). PAL is a stress-inducible, soluble homotetrameric protein with subunits between 77 and 83 kDa. Cinnamic acid 4-hydroxylase (C4H, EC 1.14.13.11) is a cytochrome P450 enzyme that introduces a hydroxyl group in para position of trans-cinnamic acid. The resulting *p*-coumaric acid can be converted to *p*-coumaroyl-CoA by hydroxycinnamic acid CoA ligase (4CL, E.C. 6.2.1.12) and *p*-coumaroyl-CoA then transformed to *p*-coumaroyl shikimic acid by shikimate *O*-hydroxycinnamoyltransferase (HCT, EC 2.3.1.133). The conversion of *p*-coumaroyl shikimic acid to caffeoyl shikimic acid is catalysed by the *p*-coumaroyl ester 3-hydroxylase (C3H, E.C. 1.14.13.36). Caffeoyl shikimic acid is converted into caffeoyl-CoA and shikimic acid by HCT and caffeoyl-CoA can be methylated into feruloyl-CoA by caffeoyl-CoA OMT (CCoAOMT, E.C. 2.1.1.104). Cinnamoyl-CoA:NADP oxidoreductase (CCR, E.C. 1.2.1.44) can convert caffeoyl-CoA or feruloyl-CoA into the corresponding cinnamic aldehydes, which are converted into the corresponding cinnamyl alcohol derivatives by

cinnamyl alcohol dehydrogenase (CAD, E.C. 1.1.1.195). Aromatic hydroxyl groups can be methylated by caffeic acid *O*-methyltransferase (COMT, E.C. 2.1.1.68).

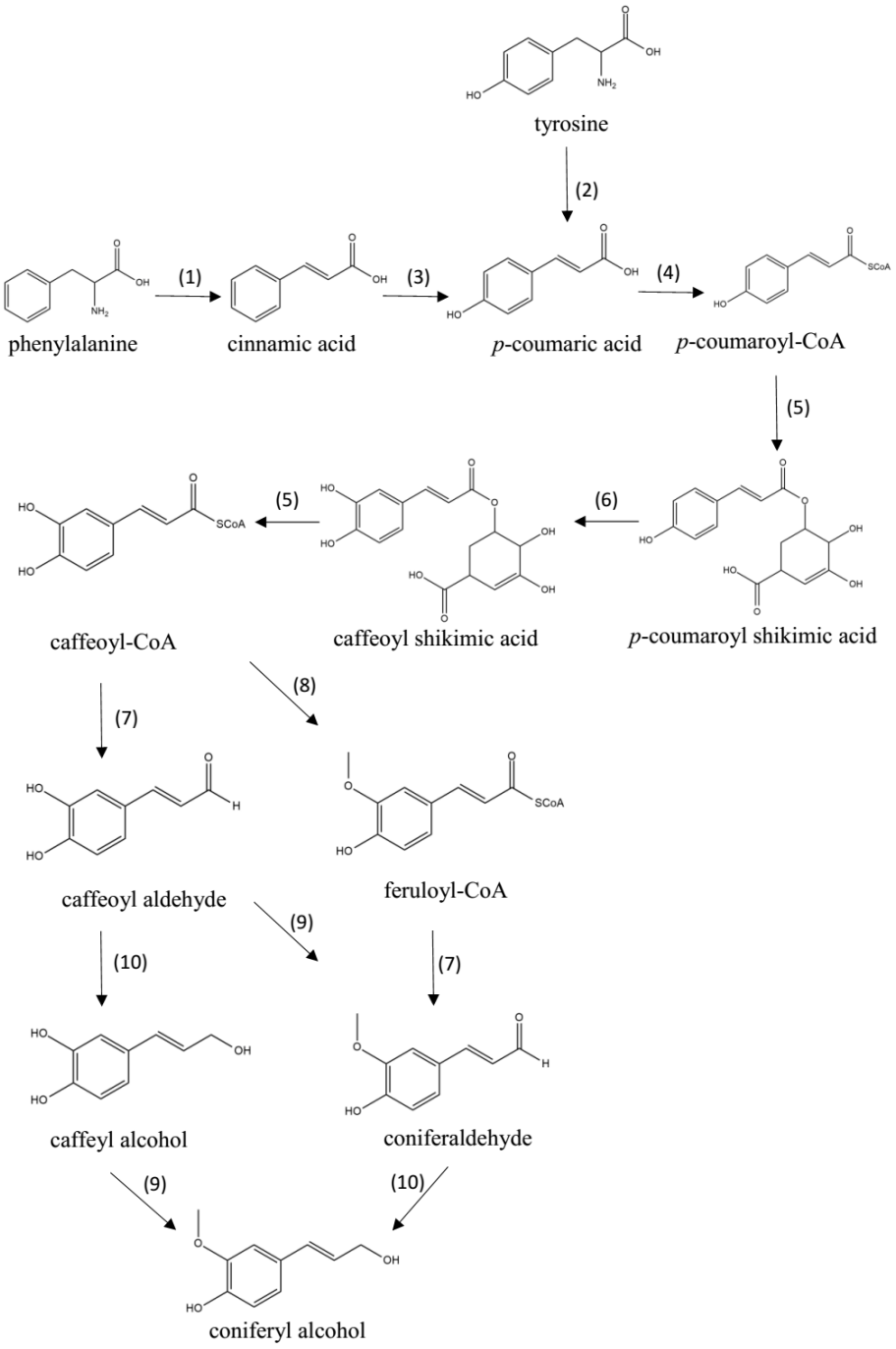


Figure 5: Main biosynthetic pathway to coniferyl alcohol

1: phenylalanine ammonia-lyase (PAL); 2: tyrosine ammonia-lyase; 3: cinnamic acid 4-hydroxylase (C4H); 4: hydroxycinnamic acid CoA ligase (4CL); 5: shikimate *O*-hydroxycinnamoyltransferase (HCT); 6: *p*-coumaroyl shikimic acid 3-hydroxylase (C3H); 7: cinnamoyl-CoA:NADP oxidoreductase (CCR); 8: caffeoyl-CoA *O*-methyltransferase (CCoAOMT); 9: caffeic acid *O*-methyltransferase (COMT); 10: cinnamyl alcohol dehydrogenase (CAD)

4.2 Early stages of lignan biosynthesis - from coniferyl alcohol to matairesinol

In the first step of this stage, two molecules of coniferyl alcohol are linked to each other stereospecifically resulting in pinoresinol (PINO). It is generally assumed that this reaction occurs via radical intermediates and the mechanism of the pinoresinol synthase resembles that of a laccase. Stereospecificity is ensured by the so-called dirigent protein, which itself has no enzymatic activity, but determines the stereochemistry of the product (Davin et al., 1997; Davin and Lewis, 2000). Since in *Forsythia* spec., *Linum perenne* and *Linum album* (+)-PINO is found, while in *Linum usitatissimum* (-)-PINO, there are presumably different dirigent proteins in different species (Kuhlmann, 2004).

Pinoresinol-lariciresinol reductase (PLR) catalyses the conversion of PINO via lariciresinol (LARI) into secoisolariciresinol (SECO) in the dependence of NADPH. These reactions are also stereospecific and different isoforms have been found for this enzyme in different plants, each of which leads to either (+)- or (-)-SECO. The cloning and crystallisation of PLR showed a relationship to isoflavone reductases (Chu et al., 1993, Dinkova-Kostova et al., 1996, Min et al., 2003) (see III.7).

The NAD-dependent secoisolariciresinol dehydrogenase (SDH) forms the lactone ring between C9 and C9' of SECO to produce matairesinol (MATAI). Secoisolariciresinol dehydrogenase was purified from *Forsythia intermedia* and *Podophyllum peltatum* and heterologously expressed in bacteria (Xia et al., 2001). The reaction of SDH had previously been demonstrated in cell-free extracts of *F. intermedia* (Umezawa et al., 1991).

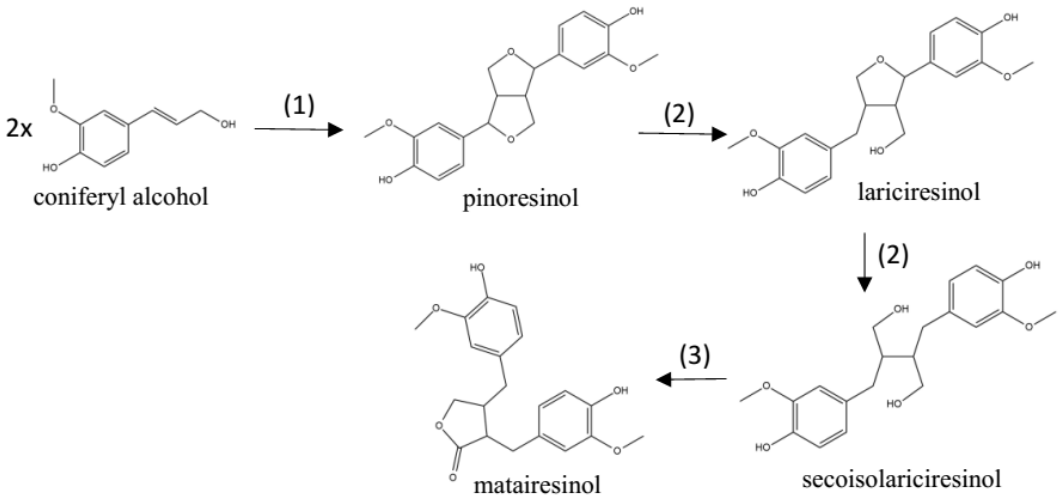


Figure 6: Lignan biosynthesis - from coniferyl alcohol to matairesinol

(1) pinoresinol synthase; (2) pinoresinol-lariciresinol reductase, (3) secoisolariciresinol dehydrogenase

4.3 Lignan biosynthetic pathway downstream of matairesinol - different models and hypotheses

In contrast to the formation of MATAI, the further biosynthesis of PTOX and derivatives such as 6-MPTOX is not fully understood. To clarify the reaction sequence, different hypotheses were used (Fig. 7):

Podophyllum spec.: Feeding experiments with radioactive precursors have shown that MATAI is the common precursor for the 4'-*O*-Methyl series (DOP, β-peltatin, PTOX) as well as the 4'-demethyl series (4'-demethyl-DOP, α-peltatin, 4'-demethyl-podophyllotoxin) (Broomhead et al., 1991). At the stage of the C2-C7'-cyclolignans such as DOP, these two series were no longer interleaved (Jackson and Dewick, 1984). As a direct precursor of α- and β-peltatin, 4'-demethyl DOP and DOP in *P. peltatum* and *P. hexandrum* were confirmed (Kamil and Dewick, 1986). In 2013, Marques et al. (2013) have identified two genes for pluviatolide synthases (CYP719A23 and CYP719A24) after sequencing the transcriptome of *P. hexandrum* and *P. peltatum*. These cytochrome P450s use (-)-matairesinol and form the methylenedioxy bridge thus establishing the A-ring of (-)-pluviatolide and further derived lignans. In 2015, by coupling transcriptome mining with combinatorial expression of candidate enzymes in tobacco, Lau and Sattely (2015) have discovered other six enzymes to complete the biosynthetic pathway to (-)-4′-desmethylepipodophyllotoxin in *Podophyllum hexandrum* (mayapple), including an oxoglutarate-dependent dioxygenase that closes the core cyclohexane ring of the aryltetralin scaffold, two *O*-methyltransferases and three cytochrome P450 enzymes (Fig. 8).

Anthriscus sylvestris: The biosynthesis of yatein was developed from MATAI in *Anthriscus sylvestris* (Sakakibara et al., 2003). For these studies, *A. sylvestris* plants were fed with ^{13}C-labeled phenylalanine. The hydroxylation and subsequent methylation on the pendant aromatic ring took place first, followed by the methylation of the OH group at C4', and finally the formation of the methylenedioxy bridge on the second benzene ring between C4 and C5. However, biotransformation experiments with suspension cultures showed that PTOX was formed from DOP, but not from yatein (Koulman et al., 2003).

Linum spec.: Biotransformation experiments with suspension cultures of *Linum flavum* have shown the transformation of DOP and β-peltatin into 6-MPTOX and 6-MPTOX glucoside (Van Uden et al., 1995; Van Uden et al., 1997). In the same cultures, PTOX was transformed to PTOX-β-D-glucoside instead of 6-MPTOX glucoside, although this is the mainly formed lignan (Van Uden et al., 1992). These experiments suggest that DOP in *Linum flavum* could be

the branching point in the biosynthetic pathways to PTOX and 6-MPTOX. The hydroxylation at position 7 of DOP to PTOX catalysed by deoxypodophyllotoxin 7-hydroxylase still needs to be characterized. On the way to 6-MPTOX, hydroxylation at position 6 of DOP is catalysed by deoxypodophyllotoxin 6-hydroxylase (DOP6H), which was characterised in *L. flavum* as a cytochrome P450 enzyme (Molog et al., 2001). This metabolic step results in the formation of ß-peltatin. This compound is converted to ß-peltatin A methyl ether (PAM) by ß-peltatin 6-*O*-methyltransferase. This enzyme was first characterised in 2003 in *L. nodiflorum* (Kranz and Petersen, 2003). The enzyme for the last hydroxylation step to form 6-MPTOX (ß-peltatin A-methyl ether 7-hydroxylase) is not known yet. In cell cultures of *Linum album*, the conversion of DOP to PTOX has also been shown by biotransformation experiments (Seidel et al., 2002; Empt et al., 2000).

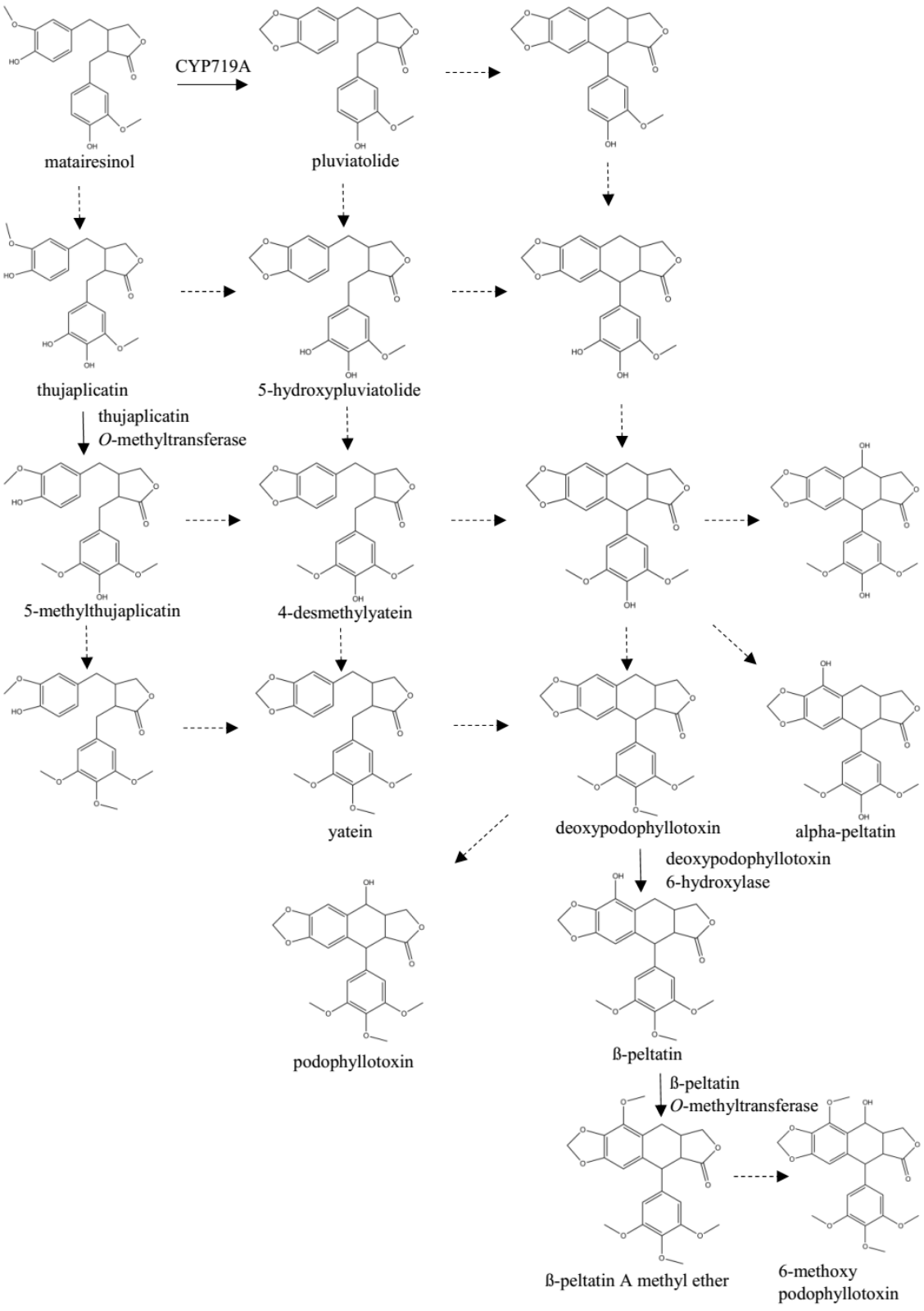


Figure 7: Overview of late stages of lignan biosynthesis (Robinson, 2018).

Known reactions are represented by continuous arrows and unknown with dashed arrows.

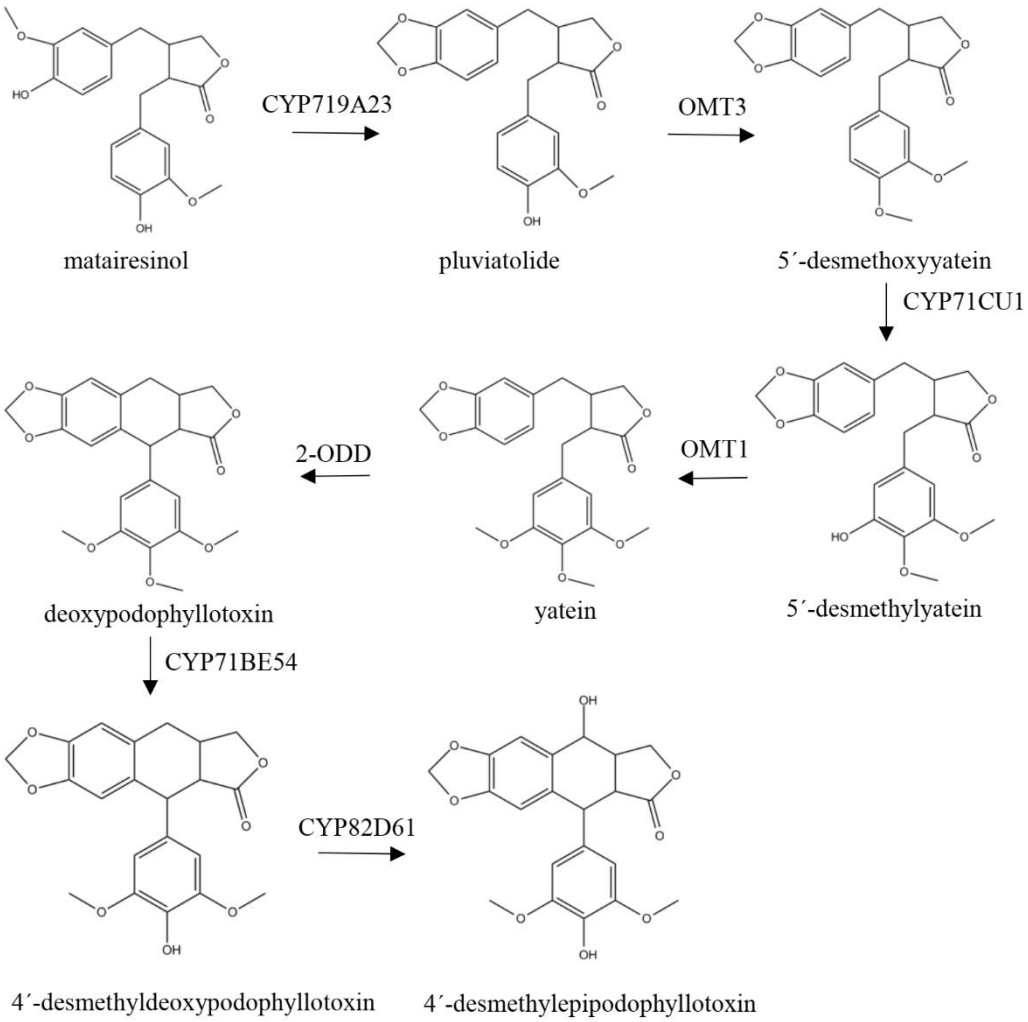


Figure 8: Six enzymes in the biosynthetic pathway to (-)-4'-desmethylepipodophyllotoxin in *Podophyllum hexandrum* (Lau and Sattely, 2015)

5. Cytochrome P450 systems in plants

Cytochromes P450 (CYP; E.C. 1.14.13., 1.14.14., 1.14.15.) are referred to as monooxygenases, as well as mixed function oxygenases. When CYPs are reduced and complexed with carbon monoxide, the enzymes have a spectrophotometric peak at the wavelength 450 nm (Kleinig and Mayer, 1999; Omura and Sato, 1964). The reactions catalysed by CYPs are complex electron transfers, which take place over several protein components.

An iron-protoporphyrin IX (heme chromophore type b), that is attached to a highly-conserved cysteine, is the recipient of the electrons in the CYP protein (Fig. 9). The first 17-29 amino acids of CYPs in the *N*-terminus are hydrophobic and serve to anchor the protein in the ER membrane. CYPs are named and classified according to their amino acid sequence in families and subfamilies. Sequence homologies over 40% are characterised as family, over 55% as subfamily and over 97% as allelic variants (Nelson et al., 1996; Werck-Reichhart et al., 2002).

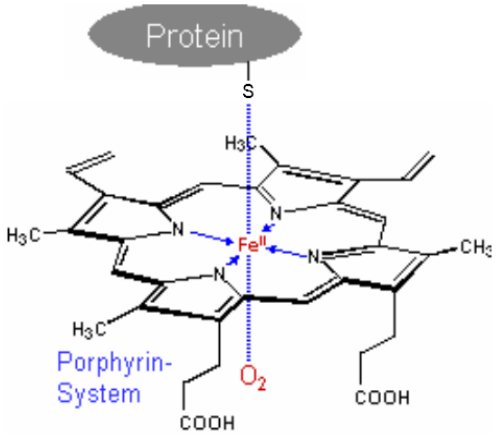


Figure 9: Iron-protoporphyrin IX (copied from Gasteiger and Schunk, 2003)

Iron-protoporphyrin IX consists of four linked pyrrole rings that complex an iron ion. The iron is bound to a cysteine residue of the apoprotein and oxygen by two further ligands at the fifth and sixth coordination sites.

The sequence identity within the plants' CYPs (Mw 45-65 kDa) is extremely low (<20%). The conserved sequence motifs of CYPs are shown in Fig. 10. The "hinge" region consisting of a "cluster" of basic and proline-rich amino acids [consensus sequence (P/I)PGPx(G/P)xP] is followed by the I helix, the "ERR" triad and the heme binding region (Durst and Nelson, 1995; Schuler, 1996; Werck-Reichhart et al., 2002). The I helix encodes the oxygen binding and activation site [consensus sequence (A/G)Gx(E/D)T(T/S)]. The "ERR" triad [consensus sequence ExxR.....R] presumably assists in the stabilisation and positioning of the heme in the binding pocket. The heme-binding region [consensus sequence FxxGxRxCxG] contains the conserved cysteine for binding the iron of protoporphyrin (Werck-Reichhart and Feyereisen, 2000).

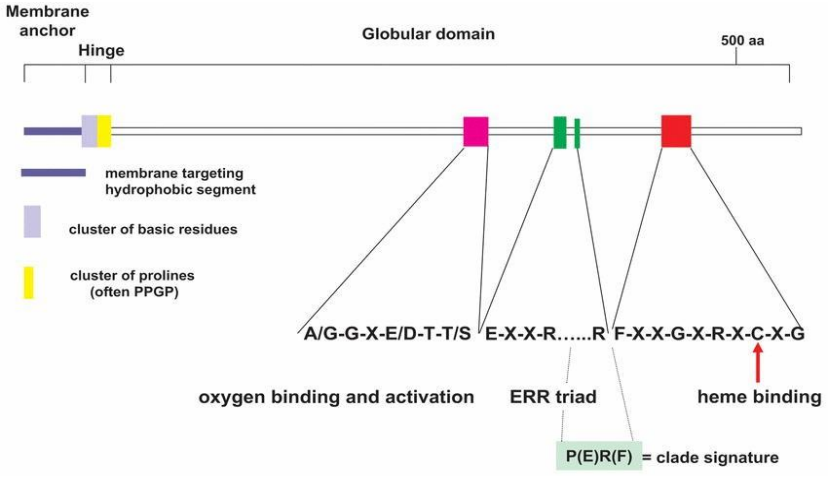


Figure 10: Conserved sequence motifs in CYPs (copied from Werck-Reichhart et al., 2002)

The catalytic reaction cycle of CYP is described in Fig. 11 (Meunier et al., 2004). In the resting state (I), the iron is present as a Fe^{3+} "low-spin" complex. This is converted into the "high-spin" state (II) by binding the substrate to Fe^{3+} and reduction to Fe^{2+}. The missing electron is supplied by NADPH via the NADPH:cytochrome P450 reductase (III). The binding of molecular oxygen leads to the formation of a CYP dioxygen complex (IV) which is activated by a second reduction equivalent and becomes a Peroxo-Fe^{2+} (VI). Protonation and cleavage of the O-O bond releases a molecule of water and leaves the reactive Fe^{3+}-O complex (VII). This complex attacks radically the bound substrate and transfers its O-radical by taking over an H-radical of the substrate and thus forms the alcohol group.

A simplified reaction scheme is the following:

$$RH + O_2 + NADPH + H^+ \rightarrow ROH + H_2O + NADP^+$$

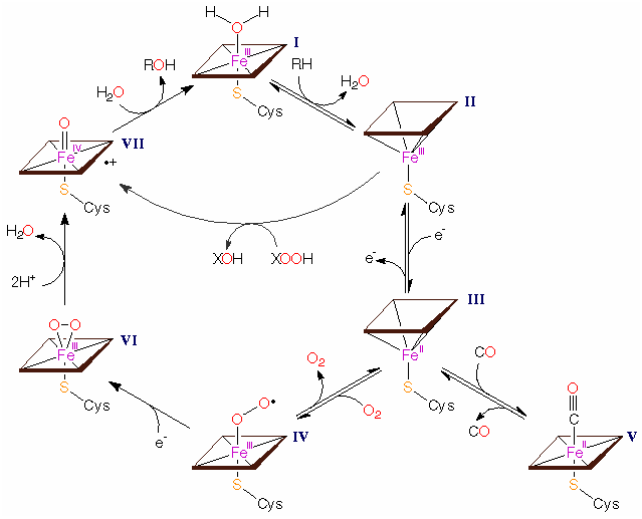


Figure 11: Catalytic reaction cycle of CYPs (copied from http://metallo.scripps.edu/promise/P450.html)

In addition to the "classical" hydroxylations, many different reactions can be catalysed by cytochrome P450-dependent enzymes, such as isomerisation, dimerisation, epoxidation, dealkylation and decarboxylation, oxidation of nitrogen and sulphur, dehalogenation and deamination (Schuler and Werck-Reichhart, 2003; Halkier, 1996).

Cytochrome P450 enzymes are involved in many plant biosynthetic pathways such as phenylpropan metabolism, the biosynthesis of alkaloids, terpenoids, glucosinolates, fatty acids, flavonoids, isoflavonoids (Humphreys and Chapple, 2000) and the detoxification of xenobiotics such as herbicides (Bolwell et al., 1994; Durst, 1988).

The great variety of the described cytochrome P450-catalysed reactions makes it clear that many oxidative steps of lignan biosynthesis in *Linum* species might be P450-dependent. A publication of Molog (2001) has shown that the C6-hydroxylation of DOP to β-peltatin in cell cultures of *Linum album* and *Linum flavum* is catalysed by a cytochrome P450 enzyme (DOP6H). Furthermore, studies with a suspension culture of *Linum album* suggested the participation of a cytochrome P450 oxygenase (DOP7H) in the formation of PTOX from DOP (Henges, 1999).

6. Cytochrome P450 reductase in plants

NADPH:cytochrome P450 reductase (CPR, EC 1.6.2.4) is located in the endoplasmic reticulum (Williams and Kamin, 1962). CPR was isolated for the first time from yeast and annotated as cytochrome c reductase based on its ability to reduce cytochrome c as artificial substrate (Haas et al., 1940). CPR contains flavin adenine dinucleotide (FAD) and flavin mononucleotide (FMN) (Benveniste et al., 1991) and transfers electrons from NADPH via FAD and FMN to the prosthetic heme group of the CYP protein (Porter, 2004).

CPR harbours a FMN-binding domain in the *N*-terminal and a NADPH/FAD-binding domain in the *C*-terminal domain. A membrane-spanning anchor anchoring the protein in the endoplasmic reticulum is formed by 50-60 hydrophobic amino acid residues in the *N*-terminus (Bonina et al., 2005). Ro et al. (2002) suggested differentiating CPR into two classes depending on their *N*-terminal membrane anchoring sequences. Members of class I present short *N*-terminal ends with appr. 50 amino acids, whereas class II show an extended *N*-terminal end with appr. 80 amino acids.

FMN-containing flavodoxin (Fld) is a small soluble electron carrier protein which participates in many redox reactions. Reversible electron transfer between NADP(H) and Fld is catalysed by a monomeric FAD-containing ferredoxin-NADP$^+$ reductase (FNR). FNRs are present in photosynthetic as well as heterotrophic organisms (Kenneth et al., 2010). The FNR domain present in CPR is derived from the plant-type FNRs (Aliverti et al., 2008). The fusion of genes encoding Fld and FNR resulted in the FAD and FMN-binding domains of CPR (Fig. 12) (Porter and Kasper, 1986).

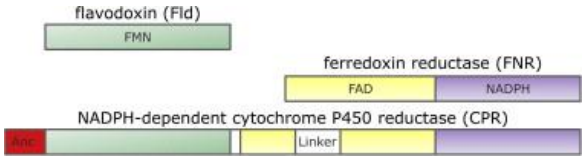


Figure 12: Molecular evolution of NADPH-cytochrome P450 oxidoreductase (CPR) (copied from Kenneth et al., 2010)

In 1997, Wang et al. identified conserved cofactor- and substrate-binding regions in the crystallised CPR from rat liver. The FMN-binding domain is located at the *C*-terminal side of the β-strands (see Fig. 13). The isoalloxazine ring of FAD lies at the boundary between the FAD- and NADP(H)-binding domains, and the interface between the FAD-binding domain and the connecting domain contains the other part of FAD.

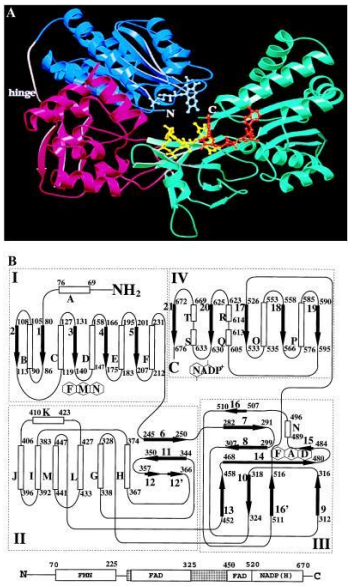


Figure 13: Overall polypeptide fold and topology diagram for CPR (copied from Wang et al., 1997)

A: The FMN-binding domain is represented in blue, the FAD- and NADP(H)-domains are shown in green, and the connecting domain in red. The cofactor FMN is represented in light blue, FAD in yellow, and NADP$^+$ in orange. The "hinge" region is shown in pink.
B: Topology diagram of the CPR protein. The domain arrangement in the CPR structure is shown in a linear diagram at the bottom.

7. Bifunctional pinoresinol-lariciresinol reductase with different stereospecificities

Most lignans are chiral compounds and only one enantiomer can be found in each plant or organ. The enantiomeric purity appears to be determined at various levels in lignan biosynthesis. The binding of the two achiral coniferyl alcohol molecules with the help of the dirigent protein leads to enantiomerically pure (+)-PINO in *Forsythia intermedia* (Davin and Lewis, 2003). In contrast, the enantiomeric purity is achieved at the level of MATAI in *Wikstroemia sikokiana* (Umezawa et al., 2003). Interestingly, opposite lignan enantiomers can be found in different plants or organs. Enzyme preparations of flowers of *Arctium lappa* catalyse the formation of (+)-PINO, (+)-LARI and (-)-SECO, while enzyme preparations from maturing seeds of this plant species catalyse the formation of the opposite enantiomers (Suzuki et al., 2002). Seeds of *Linum usitatissimum* contain pure (+)-SECO diglucoside, whereas *Linum album* accumulates pure (-)-PODO, which should have (-)-SECO as a precursor (Davin and Lewis, 2003; Petersen and Alfermann, 2001).

The enantiospecificity and diastereomeric preferences of pinoresinol-lariciresinol reductase were first investigated by Katayama et al. (1992) when the (+)- and (-)-enantiomers of PINO were incubated with *Forsythia intermedia* cell-free extracts. In the presence of NADPH, PINO was converted preferably into (+)-LARI and (-)-SECO. Incubation with (±)-LARI revealed that only the (+)-antipode was converted to (-)-SECO. This result shows the existence of a bifunctional enantiospecific pinoresinol-lariciresinol reductase (PLR) in the soluble protein extract of *F. intermedia*. The isolation of a cDNA encoding a PLR of *F. intermedia* (PLR-Fi1) and its heterologous expression showed the same enantiospecificity as for the crude extract (Dinkova-Kostova et al., 1996).

In 1999, Fujita et al. reported the presence of cDNAs corresponding to two stereochemically distinct PLR classes in a single plant species, *Thuja plicata*. Four cDNAs were grouped into two different classes of PLRs. In the first class PLR-Tp1 had high similarities with PLR-Tp3 and in the second class PLR-Tp2 showed high similarities to PLR-Tp4. Heterologously expressed PLR-Tp1 reduces (-)-PINO to (+)-SECO. On the other hand, the transformation of (±)-PINO with recombinant PLR-Tp2 led to the accumulation of both (+)- and (-)-LARI, in which only the (+)-LARI was converted to (-)-SECO. (-)-LARI was not further converted to (+)-SECO. Thus, *T. plicata* PLRs can reduce both the (+) and (-) enantiomers of PINO, but are highly enantiospecific with regard to (+)-LARI.

The enantiospecificity of a recombinant PLR from a cell suspension culture of *Linum album* (PLR-La1) has been reported by Heimendahl et al. (2005). It reduces (+)-PINO to (-)-SECO via (+)-LARI. In addition, Heimendahl et al. (2005) cloned a cDNA encoding PLR from a cell suspension culture of *L. usitatissimum* (PLR-Lu1). The recombinant protein PLR-Lu1 converts (-)-PINO to (+)-SECO.

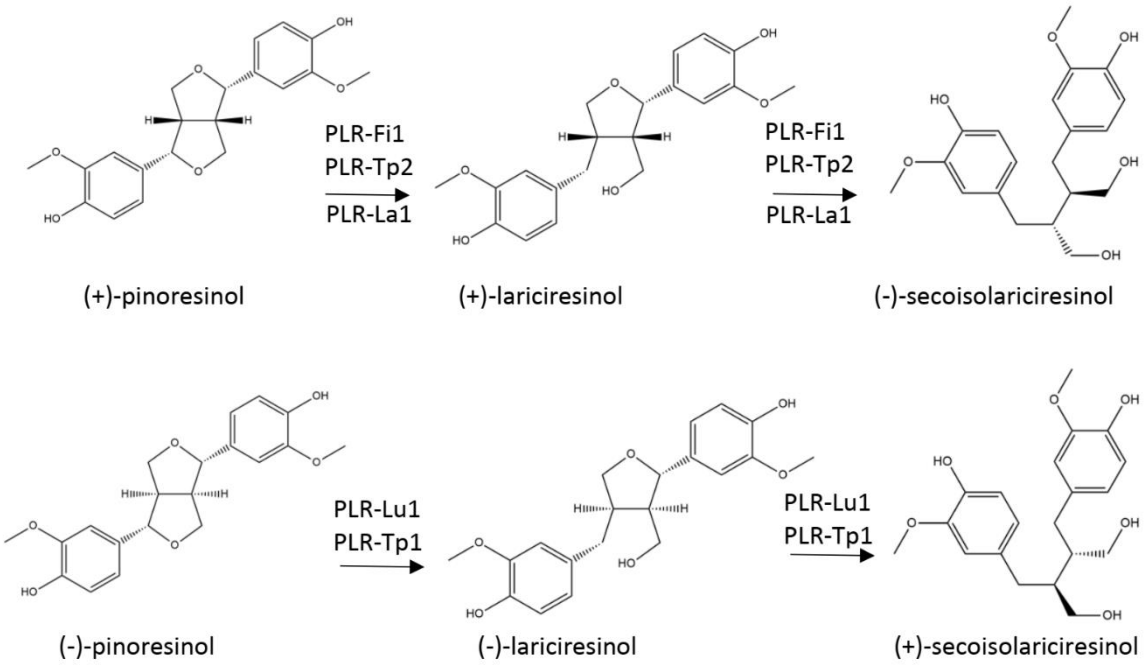


Figure 14: Different bifunctional PLRs with different stereospecificities

Hydride transfer by PLR is highly stereospecific. In partially purified PLR from *F. intermedia*, Chu et al. (1993) and Dinkova-Kostova et al. (1996) have shown that PLR abstracts the *4pro-R* hydrogen from NADPH and the incoming hydride occupies the *Pro-R* position at C-7' in LARI and at C-7/C-7' in SECO (Fig. 15).

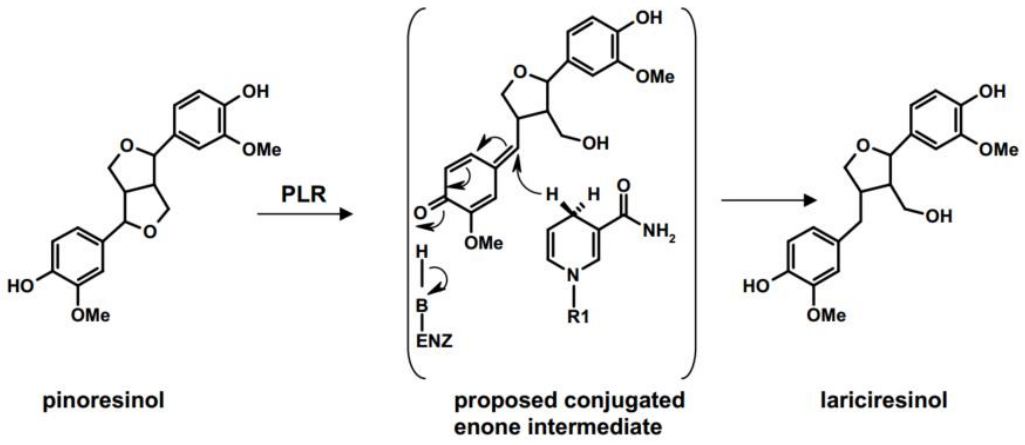


Figure 15: Mechanism of hydride transfer by PLR (copied from Fujita et al., 1999)

8. Secoisolariciresinol dehydrogenase (SDH)

Secoisolariciresinol dehydrogenase (SDH, EC 1.1.1.331) is an oxidoreductase involved in lignan biosynthesis. SDH catalyses the stereospecific conversion of SECO to MATAI via a lactol intermediate. The enzymatic activity of SDH has been identified in *F. intermedia* and *P. peltatum* (Xia et al., 2001) and classified into the enzyme family of short-chain dehydrogenases/reductases (SDRs). The SDR family was established in 1981 when the members were only a prokaryotic ribitol dehydrogenase and an insect alcohol dehydrogenase (Jörnvall et al., 1981). Since then, the SDR family has grown enormously and currently around 47000 members including species variants are known (Kallberg et al., 2010).

The SDRs can be divided into two large families, "classical" with appr. 250 amino acids and "extended" with appr. 350 amino acids. The classical SDRs have single-domain subunits that catalyze NAD(P)(H)-dependent oxidation/reduction reactions. The cosubstrate is bound at the *N*-terminal part, while the substrate binding is at the *C*-terminal part. The classical SDRs have a TGXXX[AG]XG cofactor binding motif and a YXXXK active site motif, with the Tyr residue of the active site motif serving as the critical catalytic residue. In addition to the Tyr and the Lys, there is often an upstream Ser and/or an Asn contributing to the active site. Extended SDRs have additional elements in the *C*-terminal region and typically have a TGXXGXXG cofactor binding motif (Jörnvall et al., 1995).

In the crystal structure SDH exists as a homotetramer (Moinuddin et al., 2006). Based on homology comparisons with other SDRs, SDH shows a conserved catalytic triad (Ser, Tyr and Lys). Analysis of the SDH X-ray structure, site-directed mutagenesis, and NMR spectroscopic data conducted by Moinuddin et al. (2006) have led to the delineation of the catalytic mechanism of SDH, including the role of the conserved catalytic triad (Ser, Tyr and Lys) (see Fig. 16).

Structural data for SDH (Fig. 16A) showed that several water molecules form a hydrogen-bonded network with the hydroxyl, quaternary ammonium, and phenolic groups of the highly conserved catalytic triad residues. The binding of NAD^+ releases the bound water molecules and increases the reaction entropy. Binding of NAD^+ to Lys promotes the deprotonation of the phenolic Tyr group, thereby lowering its pKa (Fig. 16B). Hydrogen bonding to the Ser hydroxyl group further stabilises the phenolate anion. The Tyr phenolate group serves as a general base in the deprotonation of substrates, thus facilitating hydride transfer during SDH catalysis. Deprotonation of the bound (-)-SECO is followed by intramolecular cyclisation/

hydride transfer to give the intermediate lactol (Fig. 16C). The last step is the release of the resulting neutral NADH and lactol from the active site (Fig. 16D). Analogously, the subsequent conversion of the lactol intermediate to (-)-MATAI involves the binding of a second molecule of NAD^+, repeating the catalytic process (Figs. 16E and 16F), hence generating a second molecule of NADH and the final product (-)-MATAI.

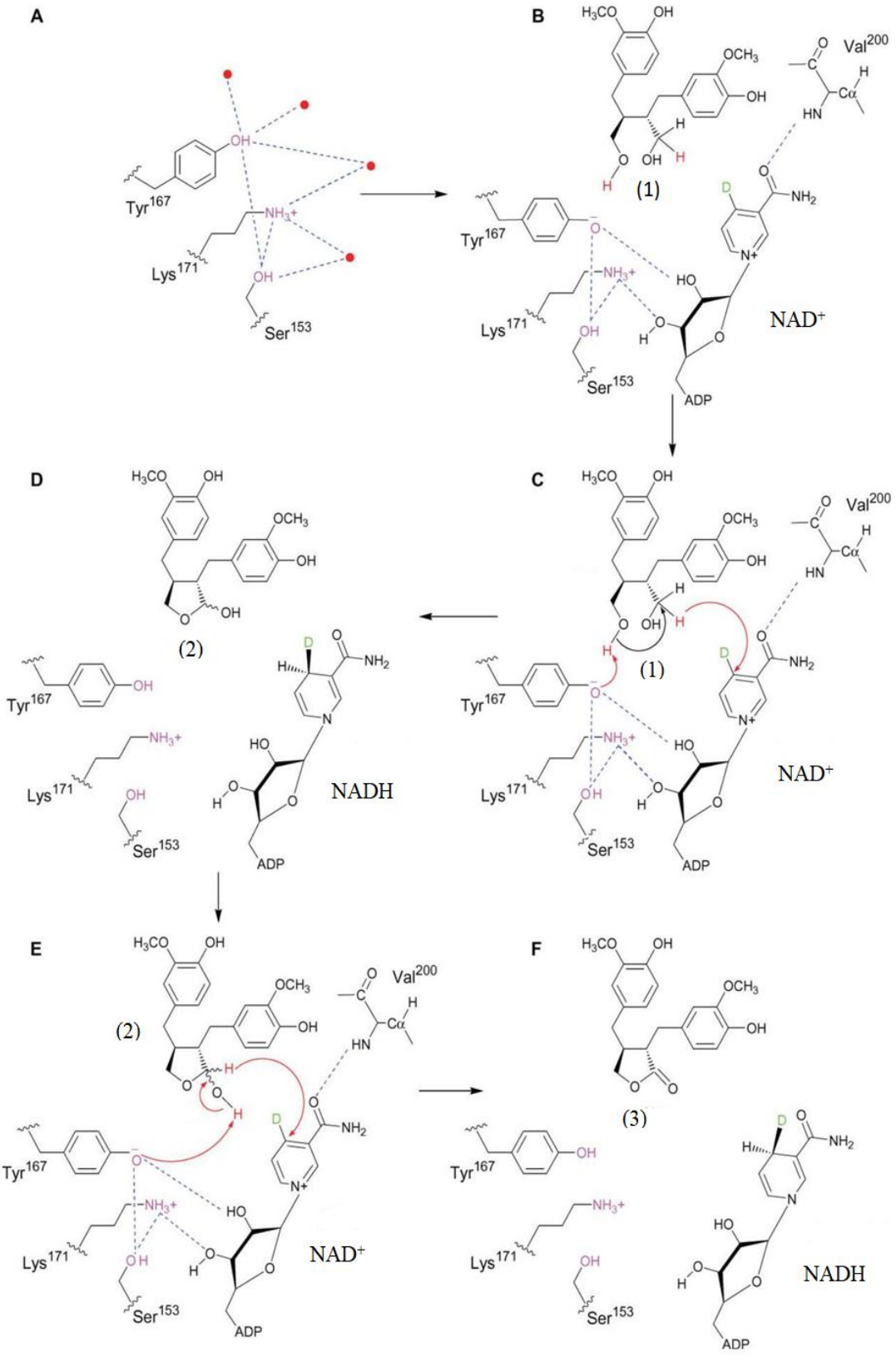


Figure 16: Proposed catalytic mechanism of SDH (taken from Moinuddin et al., 2006)

(1): (-)-secoisolariciresinol; (2): lactol intermediate; (3): (-)-matairesinol

9. Objective

The aryltetralin lignan podophyllotoxin (PTOX) and its semisynthetic derivates, e.g. etoposide and teniposide, play an important role in medicine. They are cytotoxic by binding to DNA/topoisomerase II complexes and thus induce DNA strand breaks. Since the biosynthetic capacity of PTOX in plants is comparatively low to produce pharmaceutically important active ingredients, attempts are made to improve these by targeted interventions or artificial imitation of the synthetic pathway. However, this is only possible if the complex relationships in the biosynthesis of each substance are known. Cell cultures of *Linum flavum* accumulate considerable amounts of 6-MPTOX and traces of PTOX. Therefore, these cell cultures can serve as suitable systems for the elucidation of the biosynthesis of aryltetralin lignans. The aim of this work was to gain insight into biosynthetic pathways to podophyllotoxin-type lignans in *Linum flavum*. Of particular interest are the roles of pinoresinol-lariciresinol reductase (PLR), secoisolariciresinol dehydrogenase (SDH), deoxypodophyllotoxin 6-hydroxylase (DOP6H) and deoxypodophyllotoxin 7-hydroxylase (DOP7H). In addition, mutagenesis of the enzyme PLR from *Linum flavum* was carried out to study protein structure-function relationships of PLR. Furthermore, experiments were made to identify NADPH:cytochrome P450 reductase (CPR), which is essential for cytochrome P450-dependent reactions, to which the above-mentioned enzymes DOP6H and DOP7H potentially belong.

IV. Material

1. List of chemicals

Product	Company
1-naphthaleneacetic acid (NAA)	Duchefa
2,2'-azino-bis(3-ethylbenzothiazoline-6-sulphonic acid) (ABTS)	Sigma
5-bromo-4-chloro-3-indolyl phosphate (BCIP)	Roth
5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal)	Roth
7-hydroxysecoisolariciresinol	Gift from Dr. Patrik Eklund
acetic acid, glacial	Roth
acetone	Roth
acrylamide/bisacrylamide (30%, 37.5:1)	Roth
agar-agar	Cero
agarose	Biozym / Roth
ammonium iron (II) sulfate	Merck
ammonium nitrate	Roth
ammonium persulphate (APS)	Sigma
ammonium sulphate	Roth
ampicillin	Roth
benzylaminopurine	Sigma
boric acid	Roth
bovine serum albumin (BSA)	Roth
bromophenol blue	Merck
calcium chloride dihydrate	Roth
cetyltrimethylammonium bromide (CTAB)	Roth
chloroform	Roth
cobalt (II) chloride	Merck
Coomassie Brilliant Blue G250	Fluka
Coomassie Brilliant Blue R250	Fluka
copper (II) sulfate pentahydrate	Fluka

Product	Company
D-(+)-galactose	Acros Organics / Roth
D-(+)-glucose	Roth
diethyl ether	Roth
dimethylformamide (DMF)	Merck
dipotassium hydrogen phosphate	Roth
disodium ethylenediaminetetraacetate dihydrate (EDTA-Na_2)	Roth
dithiothreitol (DTT)	Roth
dNTPs (dATP, dCTP, dGTP, dTTP)	Fermentas
D-sorbitol	Fluka
ethanol	Roth
ethidium bromide	AppliChem
ethyl acetate	Roth
fish sperm DNA (carrier DNA)	Serva
formic acid (98%)	Roth
glycerol	Roth
glycine	Merck
guanidine thiocyanate	Roth
guanidine-HCl	Roth
hydrochloric acid (37%)	Roth
indole-3-acetic acid (IAA)	Duchefa
iron(II) sulphate heptahydrate	Fluka
isopropyl-β-D-thiogalactopyranoside (IPTG)	Roth
L-adenine	Roth
L-arginine	Roth
L-aspartic acid	Roth
lauryl sarcosine	Sigma

Product	Company
L-cysteine	Roth
5-Aminolevulinic acid	Roth
L-histidine	Roth
L-isoleucine	Roth
lithium acetate	Sigma
L-leucine	Roth
L-lysine	Serva
L-methionine	Roth
L-phenylalanine	Roth
L-proline	Roth
L-serine	Roth
L-threonine	Roth
L-tryptophan	Roth
L-tyrosine	Fluka / Merck
L-valine	Roth
magnesium chloride hexahydrate	Roth
magnesium sulfate heptahydrate	Merck
manganese (II) sulfate pentahydrate	Duchefa
matairesinol	Lab's collection
methanol	Fisher Scientific
myo-inositol	Sigma / Roth
naphthalenacetic acid	Duchefa
nicotinamide-adenine-dinucleotide phosphate, reduced (NADPH)	Roth
Nicotinamide-adenine-dinucleotide, oxidized (NAD)	Biomol
nitro-blue tetrazolium chloride (NBT)	Roth
N-Z-Amine®, casein hydrolysate	Sigma
phenol (citrate buffer saturated)	Sigma
phenol/chloroform (1:1)	Roth

Product	Company
phenylmethylsulfonyl fluoride (PMSF)	Roth
phosphoric acid (85%)	Roth
pinoresinol	Sigma
Polyclar 10	ISP
polyethylene glycol 4000 (PEG)	Roth
polyvinylpyrrolidone MW 40000	Sigma
potassium acetate	Acros
potassium dihydrogen phosphate	Roth
potassium hydroxide	Merck
potassium iodide	Merck
potassium nitrate	Roth
secoisolariciresinol	Sigma
sodium chloride	Roth
sodium dodecyl sulphate (SDS)	Roth
sodium hydroxide	Merck
sodium molybdate dihydrate	Fluka
ß-peltatin	Lab's collection
ß-peltatin A methyl ether	Lab's collection
sucrose	Aldi Nord
tetrabutylammonium hydrogen sulphate	Sigma
tetracycline	Sigma
tetramethylethylenediamine (TEMED)	Roth
thiamine hydrochloride	Roth
tris(hydroxymethyl)-aminomethane (TRIS)	Roth
tryptone/peptone	Roth
tween 20	Sigma
yatein	Lab's collection
yeast extract	Roth
yeast nitrogen base	Conda
zinc (II) sulfate heptahydrate	Merck
α-peltatin	Lab's collection

All chemicals were of p.a. or purest available quality.

2. Reagents and kits

Product	Company
T4 DNA Ligase (5 U µl^{-1})	Fermentas
goat anti-Mouse IgG Fc	Fisher Scientific
cytochrome c (horse heart)	Fluka
GeneRuler™ 1 kb DNA Ladder	Fisher Scientific
GeneRuler™ DNA Ladder Mix	Fisher Scientific
GoTaq® Flexi DNA Polymerase Kit (5 U µl^{-1})	Promega
Ni-NTA His-Bind® Superflow™	Novagen
NucleoSpin®-Extract II Kit	Macherey-Nagel
PageRuler™ Protein Ladder	Fermentas
PCR Cloning kit	Qiagen

Product	Company
PD-10 Columns Sephadex G-25M	GE Healthcare
Phusion® Polymerase (2 U µl^{-1})	NEB
Pierce™ 6x-His Epitope Tag Antibody (HIS.H8)	Fisher Scientific
Qiaprep® Spin Miniprep Kit	Qiagen
restriction enzymes: XbaI, EcoRI, HindIII, NdeI, NotI, XhoI	Fermentas / Fisher Scientific
Revert Aid First Strand cDNA Synthesis Kit	Fisher Scientific
RNase H (5 U µl^{-1})	Fermentas
Roti®-Mark Standard	Roth
Roti®-Mark TRICOLOR Protein marker, prestained	Roth

3. Instruments

Instrument	Product	Manufacturer/Distributor
autoclaves	Systec VX-150	Systec GmbH
	AL02-02-100	Advantage–Lab
benchtop homogeniser	Minilys®	Bertin Instruments
Bunsen burner	Flammy S	Schütt
cell culture shakers	Certomat SII	B. Braun Biotech.
	RS-306	Infors AG
	TR-150	
centrifuges	Biofuge 17RS	Heraeus Sepatech
	Fresco 17	
	Pico 17	
	Centrifuge 5415D	Eppendorf
	Sorvall RC6+	Thermo Scientific
water purifier	OmniaPure	Stakpure GmbH
electroporation apparatus	Agagel Mini Biometra	Biomed Anaytik GmbH
	MultiSUB Mini	Cleaver Scientific
homogeniser	Ultra Turrax T25 Basic	IKA
freeze dryer	Christ L1	B. Braun Biotech

Instrument	Product	Manufacturer/Distributor
freezer	C585 Innova	New Brunswick Scientific
gel documentation systems	FAS-Digi	Nippon Genetics
HPLC columns	Hypersil Hypurity™ Elite	Thermo Scientific
	Chiralcel OD-H	Daicel
HPLC systems	L-4000UV Detector	Merck/Hitachi
	L-6200A Intelligent Pump	
	D-2500 Chromator Integrator	
ice machine	AF 80	Scotsman
magnetic stirrer	MR 3001	Heidolph Instr.
PCR thermocycler	Eppendorf Mastercycler gradient	Eppendorf
	MyCycler	Bio-Rad
pH-electrode	Accumed Basic	Fisher Scientific
photometer	BioPhotometer	Eppendorf
	Specord 200 plus	Analytik Jena
rocking platform	Duomax 1030	Heidolph Instr.
rotary evaporator	Rotavapor RE120	Büchi
scales	H64	Mettler
	PT 310	Sartorius
	EW	Kern
shaking incubator	Ecotron	Infors HT
	10X 400	Gallenkamp
laminar flow bench	Gelaire Laminar Air Flow Class 100	Gelman Instrument
thermomixer	Thermomixer Comfort	Eppendorf
ultrasonic bath	Sonorex Super RK 510 H	Bandelin
ultrasonic processor	UP 200S	Dr. Hielscher
vacuum centrifuge	Univapo 100 H	UniEquip
	RVC 2-18 CDplus	Christ
vacuum pump	MZ 2C NT	Vacuubrand
voltage controller	E835	Consort
	E143	
	EV2310	
	EV3020	
mixer	Vortex-Genie 2	Scientific Industries
	Vortex Mixer	VELP Scientifica
water bath	Thermomix ME	B. Braun Biotech.

4. Genotypes of laboratory strains

Information on the following genotypes was taken from the respective handbooks of the bacterial strains and yeast strain.

E. coli EZ (Qiagen)

E. coli str. [F'::Tn10 (Tc^r) $proA^+B^+$ $lacI^qZ\Delta M15$] recA1 end A1 hsdR17 (r_{K12}^- m_{K12}^+) lac glnV44 thi-1 gyrA96 relA1

SoluBL21™ Competent *E. coli* (Amsbio)

E. coli str. F- ompT $hsdS_B$ (r_B- m_B-) gal dcm (DE3)†

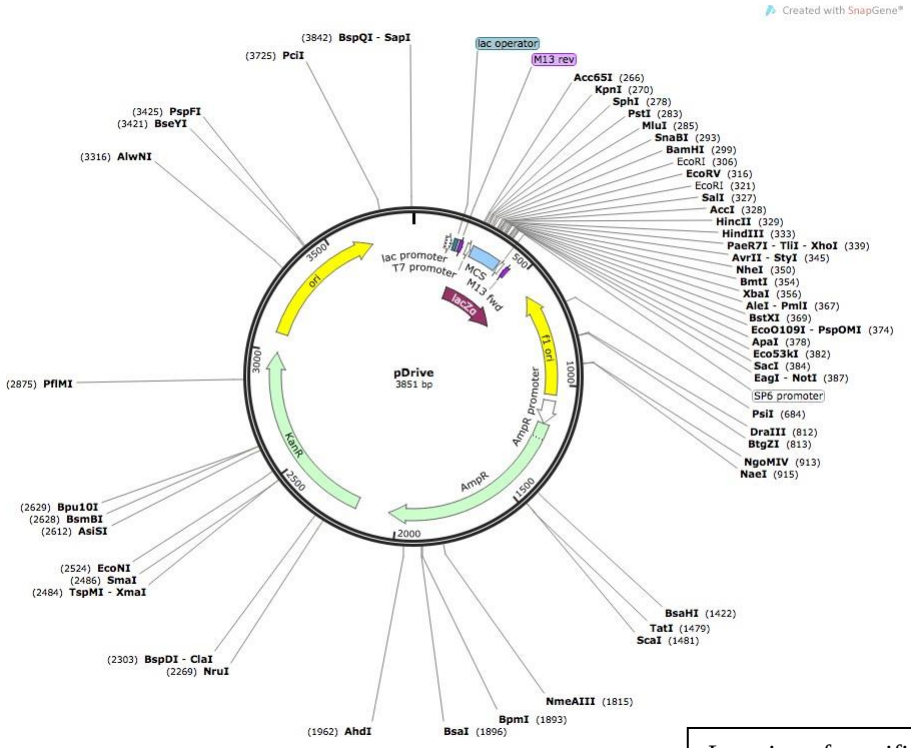
Saccharomyces cerevisiae InvSc1 (Invitrogen)

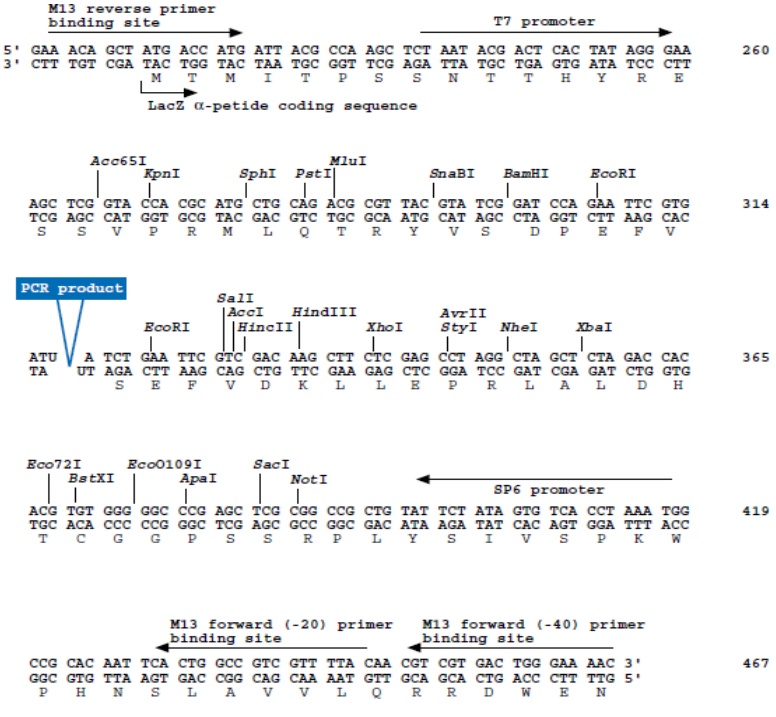
S. cerevisiae str. MATa his3D1 leu2 trp1-289 ura3-52 MAT his3D1 leu2 trp1-289 ura3-52

5. Vector sequences, maps and features

All information on the following vector maps are taken from the manufacturer's manuals.

5.1 pDrive (Qiagen)





Location of specific vector features

Vector size (bp): 3851 Multiple cloning site: 266–393 LacZ α-peptide: 216–593

T7 RNA polymerase promoter: 239–258 T7 transcription start: 256

SP6 RNA polymerase promoter: 398–417 SP6 transcription start: 400

Ampicillin resistance gene: 1175–2032 Kanamycin resistance gene: 2181–2993 pUC origin: 3668

Phage f1 origin: 588–1043
Primer binding sites:

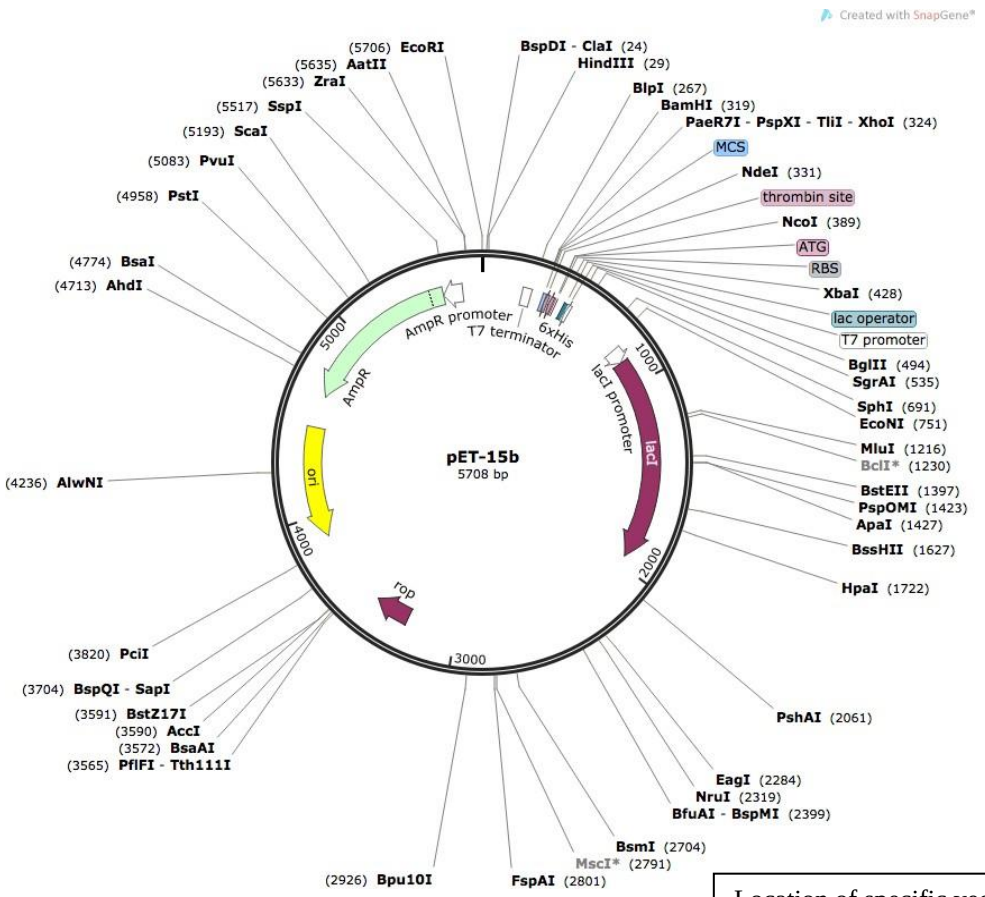
M13 forward (–20): 431–447

M13 forward (–40): 451–467

M13 reverse: 209–224

T7 promoter primer: 239–258
SP6 promoter primer: 400–418

5.2 pET-15b (Novagen)



Location of specific vector features

T7 promoter: 463-479

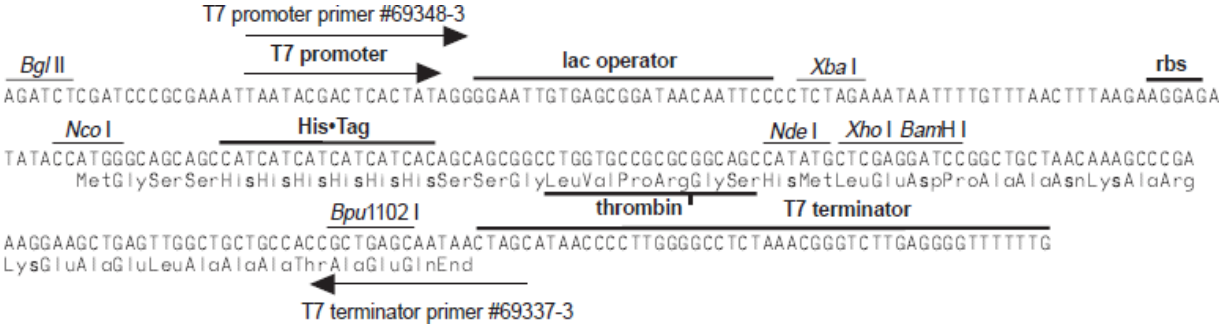
T7 transcription start: 452

His-Tag coding sequence: 362-380

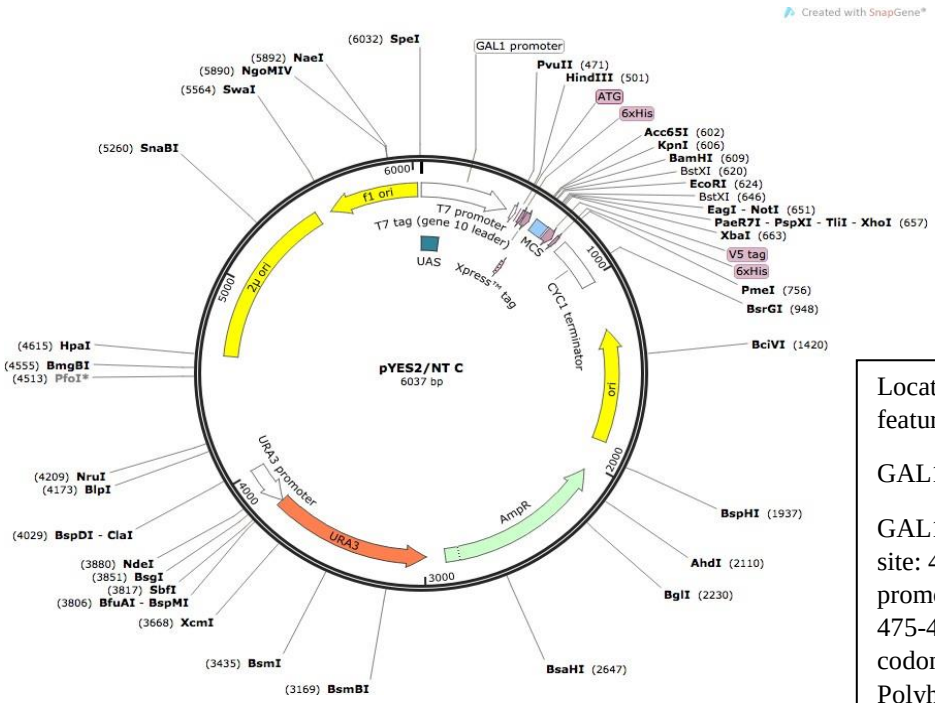
Multiple cloning sites (Nde I - BamH I): 319-335 T7 terminator: 213-259

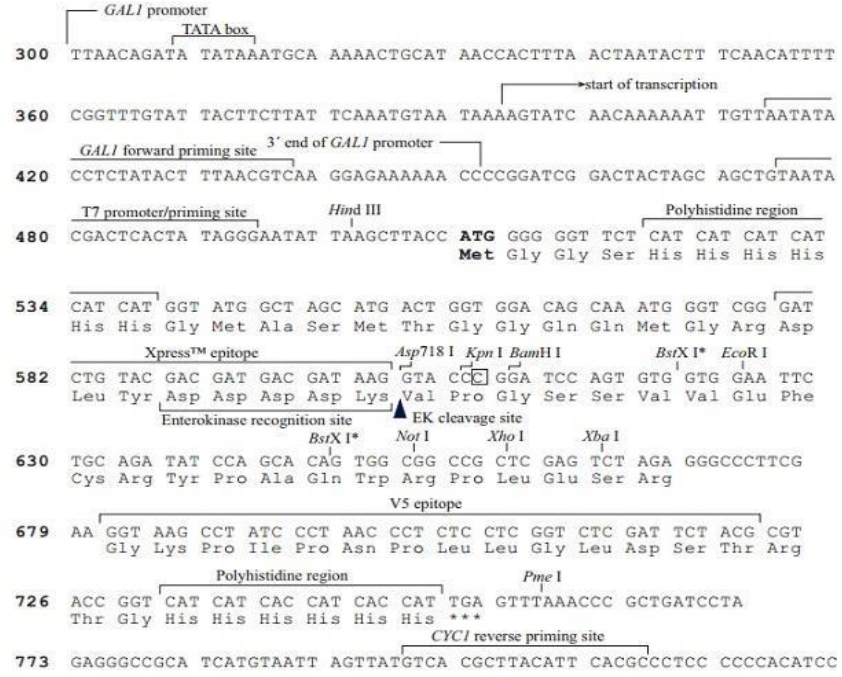
lacI coding sequence: (866-1945) pBR322 origin: 3882

bla coding sequence: 4643-5500



5.3 pYes2/NT C (Invitrogen)





Location of specific vector features

GAL1 promoter: 1-451

GAL1 forward priming site: 414-437 T7 promoter/priming site: 475-494 ATG initiation codon: 510-512 Polyhistidin (6xHis) region: 522-539 Xpress™ epitope: 579-602

Enterokinase (EK) recognition site: 588-602 Multiple cloning site: 602-669

V5 epitope: 682-723

Polyhistidine (6xHis) region: 733-750

CYC1 transcription termination signal: 783-1036 CYC1 reverse priming site: 800-818

pUC origin site: 1220-1893 Ampicillin resistance gene: 2038-2898 (complementary)

URA3 gene: 2916-4023 (complementary) 2µ origin: 4027-5498

f1 origin: 5566-6021 (complementary)

6. Primer list

6.1 Primers for CYP candidates

Name	Sequence (5'–3')	T_m [°C]	Restriction site	Comment
CYP-11511-f	ATGGATTTCTTCACTTCTCTCT	54.7		Full-length primer
CYP-11511-r	TTATGTCTAACATATATCGATCATTC	55.3		Full-length primer
CYP-11862-f	ATGGATTCTCTCTTTGCTTCAATTG	58.1		Full-length primer
CYP-11862-r	TTAAACATAAGCATCGTGAGACAATC	58.5		Full-length primer
CYP-2114-f	ATGGAGCTCCTCCAAATGTTACCTG	63		Full-length primer
CYP-2114-r	CTAAACGGTTGGTACAGGGTTGC	62.4		Full-length primer
CYP-2227-f	ATGGAATGCTCCTACTCCCAATTC	61		Full-length primer
CYP-2227-r	CTAGTGGTATGGGGTTGGAATCAAT	61.3		Full-length primer
CYP-2405-f	ATGTTCATAAGGCCAAGTCCCAA	58.9		Full-length primer
CYP-2405-r	TCATCCATAAACTTCAGGAGCCAA	59.3		Full-length primer
CYP-2408-f	ATGGCCGCCTCGCTCACCT	63.1		Full-length primer
CYP-2408-r	CTAATTTGCAACAGCCTCTAACATTTCAG	62.4		Full-length primer
CYP-2702-f	ATGACTCTAATGGAACTAGCACTAG	59.7		Full-length primer
CYP-2702-r	TCATAGTTTCAAGGCATTAGCATCATA	58.9		Full-length primer
CYP-27263-f	ATGGCCGACAAGTACGGC	58.2		Full-length primer
CYP-27263-r	TTAGCCGTACAAATGAGCTGG	57.9		Full-length primer
CYP-31728-f	ATGGAGCTTCTTCAACTACTC	55.9		Full-length primer
CYP-31728-r	TTAAGCAATGACAGGAACTAATGA	55.9		Full-length primer
CYP-3458-f	ATGGAGAGGAATATCAGAGCTTTCT	59.7		Full-length primer
CYP-3458-r	TCAAGCTGCCATGCCATCGT	59.4		Full-length primer
CYP-38991-f	ATGGATATCATCATCTCCCACC	58.4		Full-length primer
CYP-38991-r	CTAAAGTACTCCATACAACTCGG	58.9		Full-length primer
CYP-4152-f	ATGGCGGCCGGGAGGGAT	62.8		Full-length primer
CYP-4152-r	CTATGTACATGCCACGGGGATAAG	62.7		Full-length primer

CYP-4471-f	ATGCCTTCACTACTTATCTACCTT	58.6		Full-length primer
CYP-4471-r	TTATAAAAACTTTGTAGCTACTAGACATAG	57.6		Full-length primer
CYP-5627-f	ATGGATCTGTTCCTTCCATCCCT	60.6		Full-length primer
CYP-5627-r	TTATTGGTAGAGCCTCCAAGGCAA	61		Full-length primer
CYP-74047-f	ATGGATTCCATAGCTCTACCC	57.9		Full-length primer
CYP-74047-r	TCAAGAAATTATTGGTGGAGGATAG	58		Full-length primer
CYP-9893-f	ATGGATTGGATCAGTCAATTTGGC	59.3		Full-length primer
CYP-9893-r	TCAGAAGAGATTTGGCAGCAGC	60.3		Full-length primer
CYP-11511-Hin-f	AT**AAGCTT**ACCATGGATTTCTTCACTTCTCTCTC	57.1	HindIII	Full-length primer
CYP-11511-Xba-r	AT**TCTAGA**TTCATACAGATGCGGTGGC	56.7	XbaI	Full-length primer
CYP-11862-Hin-f	AT**AAGCTT**ACCATGGATTCTCTCTTTGCTTCAATTGC	60.1	HindIII	Full-length primer
CYP-11862-Xba-r	AT**TCTAGA**AACATAAGCATCGTGAGACAATCG	59.3	XbaI	Full-length primer
CYP-2408-Hin-f	AT**AAGCTT**ACCATGGCCGCCTCGCTCACCT	63.1	HindIII	Full-length primer
CYP-2408-Xba-r	AT**TCTAGA**ATTTGCAACAGCCTCTAACATTTCAGAAG	62.4	XbaI	Full-length primer
CYP-27263-Hin-f	AT**AAGCTT**ACCATGGCCGACAAGTACGGCTCC	63.7	HindIII	Full-length primer
CYP-27263-Xba-r	AT**TCTAGA**GCCGTACAAATGAGCTGGAAGCC	64.2	XbaI	Full-length primer
CYP-4471-Hin-f	AT**AAGCTT**ACCATGCCTTCACTACTTATCTAC	54	XbaI	Full-length primer
CYP-4471-Xba-r	AT**TCTAGA**TAAAAACTTTGTAGCTACTAGACA	54.2	HindIII	Full-length primer
CYP-2114-Hind-f	AT**AAGCTT**ACCATGGAGCTCCTCCAAATGTTACC	60.6	HindIII	Primer for fusion-PCR
CYP-2114-f2	AGGCTTTCCTTCTGGACATTTTCTTAGCCG	72.3		Primer for fusion-PCR
CYP-2114-r1	CGGCTAAGAAAATGTCCAGAAGGAAAGCCT	72.3		Primer for fusion-PCR
CYP-2114-Xba-r	AT**TCTAGA**AACGGTTGGTACAGGGTTGC	59.4	XbaI	Primer for fusion-PCR
CYP-2227-Hind-f	AT**AAGCTT**ACCATGGAATGCTCCTACTCCCAATT	58.9	HindIII	Primer for fusion-PCR
CYP-2227-f2	AGGCCGTCATTCTCGATATATTTATTGCTG	68.5		Primer for fusion-PCR
CYP-2227-r1	CAGCAATAAATATATCGAGAATGACGGCCT	68.5		Primer for fusion-PCR

CYP-2227-Xba-r	AT**TCTAGA**GTGGTATGGGGTTGGAATCAATTT	59.3	XbaI	Primer for fusion-PCR
CYP-3458-Hind-f	AT**AAGCTT**ACCATGGAGAGGAATATCAGAGCTTTC	59.3	HindIII	Primer for fusion-PCR
CYP-3458-f2	AAGCGGTCACTTTGGAACTGTTCATAGCTG	72.5		Primer for fusion-PCR
CYP-3458-r1	CAGCTATGAACAGTTCCAAAGTGACCGCTT	72.5		Primer for fusion-PCR
CYP-3458-Xba-r	AT**TCTAGA**AGCTGCCATGCCATCGTATAAAAT	59.3	XbaI	Primer for fusion-PCR
CYP-38991-Hind-f	AT**AAGCTT**ACCATGGATATCATCATCTCCCACC	58.4	HindIII	Primer for fusion-PCR
CYP-38991-f2	CAGGCTACCGCCATGTCTTTGATCGTGGCG	78		Primer for fusion-PCR
CYP-38991-r1	CGCCACGATCAAAGACATGGCGGTAGCCTG	78		Primer for fusion-PCR
CYP-38991-Xba-r	AT**TCTAGA**AAGTACTCCATACAACTCGGGA	58.4	XbaI	Primer for fusion-PCR
CYP-5627-Hind-f	AT**AAGCTT**ACCATGGATCTGTTCCTTCCATCCC	60.3	HindIII	Primer for fusion-PCR
CYP-5627-f2	CGATCATGGTACCGTTCACGCTTTATAATT	68		Primer for fusion-PCR
CYP-5627-r1	AATTATAAAGCGTGAACGGTACCATGATCG	68		Primer for fusion-PCR
CYP-5627-Xba-r	AT**TCTAGA**TTGGTAGAGCCTCCAAGGCAA	59.8	XbaI	Primer for fusion-PCR
CYP-74047-Hind-f	AT**AAGCTT**ACCATGGATTCCATAGCTCTACCC	57.9	HindIII	Primer for fusion-PCR
CYP-74047-f2	CAAAGCGGTTATTGGGGATGTGTTTATTGC	70		Primer for fusion-PCR
CYP-74047-r1	GCAATAAACACATCCCCAATAACCGCTTTG	70		Primer for fusion-PCR
CYP-74047-Xba-r	AT**TCTAGA**AGAAATTATTGGTGGAGGATAGC	57.1	XbaI	Primer for fusion-PCR
CYP-11511-Xba-rn	AT**TCTAGA**CTATTCATACAGATGCGGTGGC	56.7	XbaI	Full-length reverse primer with stop codon
CYP-11862-Xba-rn	AT**TCTAGA**CTAAACATAAGCATCGTGAGACAATCG	59.3	XbaI	Full-length reverse primer with stop codon
CYP-2114-Xba-rn	AT**TCTAGA**CTAAACGGTTGGTACAGGGTTGC	59.4	XbaI	Full-length reverse primer with stop codon
CYP-2227-Xba-rn	AT**TCTAGA**CTAGTGGTATGGGGTTGGAATCAATTT	59.3	XbaI	Full-length reverse primer with stop codon
CYP-2408-Xba-rn	AT**TCTAGA**CTAATTTGCAACAGCCTCTAACATTTCAGAAG	62.4	XbaI	Full-length reverse primer with stop codon

CYP-27263-Xba-rn	AT**TCTAGA**CTAGCCGTACAAATGAGCTGGAAGCC	64.2	XbaI	Full-length reverse primer with stop codon
CYP-3458-Xba-rn	AT**TCTAGA**CTAAGCTGCCATGCCATCGTATAAAAT	59.3	XbaI	Full-length reverse primer with stop codon
CYP-38991-Xba-rn	AT**TCTAGA**CTAAAGTACTCCATACAACTCGGGA	58.4	XbaI	Full-length reverse primer with stop codon
CYP-4471-Xba-rn	AT**TCTAGA**CTATAAAAACTTTGTAGCTACTAGACA	54.2	XbaI	Full-length reverse primer with stop codon
CYP-5627-Xba-rn	AT**TCTAGA**CTATTGGTAGAGCCTCCAAGGCAA	59.8	XbaI	Full-length reverse primer with stop codon
CYP-74047-Xba-rn	AT**TCTAGA**CTAAGAAATTATTGGTGGAGGATAGC	57.1	XbaI	Full-length reverse primer with stop codon

Restriction sites are written in bold letters. Underlined nucleotides stand for the part of the primer corresponding to the sequence.

6.2 Primers for CPR candidates

Name	Sequence (5'–3')	T_m [°C]	Restriction site	Comment
CPR-4753-f	ATGGACTCGTCCTCGTCTG	58.8		Full-length primer
CPR-4753-r	TTACCAAACGTCGCGCAGG	58.8		Full-length primer
CPR-5254-f	ATGGACTCGCCGTCTTCGT	58.8		Full-length primer
CPR-5254-r	TTACCATACGTCACGAAGGTAC	58.4		Full-length primer
CPR-5729-f	ATGAGTTCCAGCGGTCCGG	61		Full-length primer
CPR-5729-r	TCACCATACATCTCTAAGATATCGCC	61.6		Full-length primer
CPR-66401-f	ATGAGTTCCAGCGGTCTGGA	59.4		Full-length primer
CPR-66401-r	TCACCATACATCTCTAAGATACCG	59.3		Full-length primer
CPR-4753-Not-f	AT**GCGGCCGC**TCATGGACTCGTCCTCGTCTG	58.8	NotI	Full-length primer
CPR-4753-Xba-r	TA**TCTAGA**TTACCAAACGTCGCGCAGG	58.8	XbaI	Full-length primer
CPR-66401-Kpn-f	AT**GGTACC**CATGAGTTCCAGCGGTCTGGA	59.3	KpnI	Full-length primer
CPR-66401-Xba-r	TA**TCTAGA**TCACCATACATCTCTAAGATACCG	59.4	XbaI	Full-length primer

Restriction sites are written in bold letters. Underlined nucleotides stand for the part of the primer corresponding to the sequence.

6.3 Primers for SDH candidates

Name	Sequence (5'–3')	T_m [°C]	Restriction site	Comment
SD-28880-f	ATGGGCTCTGATATCTGTGCAC	60.3		Full-length primer
SD-28880-r	TCATGCTTCAGCGGAAGCGC	61.4		Full-length primer
SD-36067-f	ATGGCGGCAGCAGCACCAG	63.1		Full-length primer
SD-36067-r	TCATGAAGATCCATCTTCCGGGTAATT	61.9		Full-length primer
SD-5591-f	ATGGCCGCCACTAACTTCGTTT	60.3		Full-length primer
SD-5591-r	TCAAGCGAACCTCAAATCATCAAAG	59.7		Full-length primer
SD-73995-f	ATGTCTTCTTCCTACCAGCCGG	62.1		Full-length primer
SD-73995-r	TCAAGCACTCTTCAAACTACAACCG	61.3		Full-length primer
SD-7665-f	ATGTCGAAGCTGGTCATGGGTTCT	62.7		Full-length primer
SD-7665-r	TCACTGGCTATTCCTTGGAGGCAA	62.7		Full-length primer
SD-28880-Nde-f	ATA**CATATG**GGCTCTGATATCTGTGCACCCT	64.6	NdeI	Primer for fusion-PCR
SD-28880-1	GGCTCTGATATCTGTGCACCCTCTGTCACCAAGACGCTAGAAGGCAAAGT	82		Primer for fusion-PCR
SD-28880-f2	CTCTGTCACCAAGACGCTAGAAGGCAAAGT	73.3		Primer for fusion-PCR
SD-28880-Xho-r	TA**CTCGAG**TCATGCTTCAGCGGAAGCGCCAAA	64.4	XhoI	Primer for fusion-PCR
SD-36067-Nde-f	ATA**CATATG**GCGGCAGCAGCACCAGCTTCCTTTAT	69.5	NdeI	Primer for fusion-PCR
SD-36067-1	CAGCACCAGCTTCCTTTATCTCCTCCGTCGCCAGAAGGCTGGAAGGCAAA	84		Primer for fusion-PCR
SD-36067-f2	TCCTCCGTCGCCAGAAGGCTGGAAGGCAAA	81		Primer for fusion-PCR
SD-36067-Xho-r	TA**CTCGAG**TCATGAAGATCCATCTTCCGGGTAATTGAAGATC	67.1	XhoI	Primer for fusion-PCR

SD-5591-Nde-f	ATA**CATATG**GCCGCCACTAACTTCGTTTCTTCCATCATAAAAAGGCTGGA	57.1	NdeI	Primer for fusion-PCR
SD-5591-f1	TCTTCCATCATAAAAAGGCTGGAAGGCAAA	70		Primer for fusion-PCR
SD-5591-r1	CTCCCCGCCGGTACCATCACCCTGGCGGCG	87		Primer for fusion-PCR
SD-5591-f2	CGCCGCCAGGGTGATGGTACCGGCGGGGAG	87		Primer for fusion-PCR
SD-5591-Xho-r	TA**CTCGAG**TCAAGCGAACCTCAAATCATCAAA	57.6	XhoI	Primer for fusion-PCR
SD-73995-Nde-f	ATA**CATATG**TCTTCTTCCTACCAGCCGGCGGCGGCGAGTTCTCACAGACT	67.8	NdeI	Primer for fusion-PCR
SD-73995-f1	GCGGCGGCGAGTTCTCACAGACTAGAAGGCAAGGTGGCAGT	84		Primer for fusion-PCR
SD-73995-Xho-r	TA**CTCGAG**TCAAGCACTCTTCAAACTACAACCGCCG	66.6	XhoI	Primer for fusion-PCR
SD-7665-Nde-f	ATA**CATATG**ATGTCGAAGCTGGTCATGGGTTCTTCCAGAG	59.8	NdeI	Primer for fusion-PCR
SD-7665-f1	TGGTCATGGGTTCTTCCAGAGATCTGGCGTT	76		Primer for fusion-PCR
SD-7665-r1	AGCTTCCTGCTGGTTTCATTGCCACTTGCA	76		Primer for fusion-PCR
SD-7665-f2	TGCAAGTGGCAATGAAACCAGCAGGAAGCT	76		Primer for fusion-PCR
SD-7665-Xho-r	TA**CTCGAG**TCACTGGCTATTCCTTGGAGG	59.1	XhoI	Primer for fusion-PCR
SD-28880-Hin-f	ATA**AAGCTT**ACCATGGGCTCTGATATCTGTGCACC	62.4	HindIII	Full-length primer
SD-28880-Xba-r	AT**TCTAGA**TGCTTCAGCGGAAGCGCCAAA	61.8	XbaI	Full-length primer
SD-36067-Hin-f	ATA**AAGCTT**ACCATGGCGGCAGCAGCACCAGCTT	65.8	HindIII	Full-length primer
SD-36067-Xba-r	AT**TCTAGA**TGAAGATCCATCTTCCGGGTAATTGAAG	63.7	XbaI	Full-length primer
SD-5591-Hin-f	ATA**AAGCTT**ACCATGGCCGCCACTAACTTCGTTTCTT	63	HindIII	Full-length primer
SD-5591-Xba-r	AT**TCTAGA**AGCGAACCTCAAATCATCAAAGATCG	61.6	XbaI	Full-length primer
SD-73995-Hin-f	ATA**AAGCTT**ACCATGTCTTCTTCCTACCAGCCGGC	64.2	HindIII	Full-length primer
SD-73995-Xba-r	AT**TCTAGA**AGCACTCTTCAAACTACAACCGCCG	64.6	XbaI	Full-length primer
SD-7665-Hin-f	ATA**AAGCTT**ACCATGTCGAAGCTGGTCATGGGTTCTT	63	HindIII	Full-length primer
SD-7665-Xba-r	AT**TCTAGA**CTGGCTATTCCTTGGAGGCAACC	64.2	XbaI	Full-length primer

SD-28880-Xba-rn	AT**TCTAGA**CTATGCTTCAGCGGAAGCGCCAAA	62.8	XbaI	Full-length primer
SD-36067-Xba-rn	AT**TCTAGA**CTATGAAGATCCATCTTCCGGGTAATTGAAG	63.9	XbaI	Full-length primer
SD-5591-Xba-rn	AT**TCTAGA**CTAAGCGAACCTCAAATCATCAAAGATCG	62.1	XbaI	Full-length primer
SD-73995-Xba-rn	AT**TCTAGA**CTAAGCACTCTTCAAACTACAACCGCCG	64.8	XbaI	Full-length primer
SD-7665-Xba-rn	AT**TCTAGA**CTACTGGCTATTCCTTGGAGGCAACC	64.8	XbaI	Full-length primer

Restriction sites are written in bold letters. Underlined nucleotides stand for the part of the primer corresponding to the sequence.

6.4 Primers for PLR candidates

Name	Sequence (5'–3')	T_m [°C]	Restriction site	Comment
PLR-10318-Nde-f	**ATACAT**ATGGGTTCCCTGGGGAAAG	58.8	NdeI	Full-length primer
PLR-10318-Xho-r	TA**CTCGAG**CTAGATGTAACGCTTTAGATATTCCT	58.5	XhoI	Full-length primer
PLR-G280Y-f	CGTGATAATAGTGCGTCAGATAAACTTGCTGTGCATAATTTGCTTCTCTC	79.4		Primer for site-directed mutagenesis
PLR-G280Y-r	GAGAGAAGCAAATTATGCACAGCAAGTTTATCTGACGCACTATTATCACG	79.4		Primer for site-directed mutagenesis
PLR-Y284G-f	CTCGTAACACACGTGATAACCGTGCGTCAGTCCAACTTGC	79.1		Primer for site-directed mutagenesis
PLR-Y284G-r	GCAAGTTGGACTGACGCACGGTTATCACGTGTGTTACGAG	79.1		Primer for site-directed mutagenesis

Restriction sites are written in bold letters. Underlined nucleotides stand for the part of the primer corresponding to the sequence.

7. Culture media

7.1 MS-Lf

MS-Lf medium is a modification of the MS medium according to Murashige & Skoog (1962).

MS macro elements (25x):

KNO_3	190 g/l
$MgSO_4$ x 7 H_2O	37 g/l
KH_2PO_4 x H_2O	17 g/l
$CaCl_2$ x 2 H_2O	44 g/l
NH_4NO_3	165 g/l

This solution is autoclaved after preparation.

MS microelements (100x):

H_3BO_3	620 mg/l
$ZnSO_4$ x 7 H_2O	860 mg/l
$MnSO_4$ x H_2O	1690 mg/l
KI	83 mg/l
Na_2MoO_4 x H_2O	25 mg/l
$CuSO_4$ x 5 H_2O	2.5 mg/l
$CoCl_2$ x 6 H_2O	2.5 mg/l
$FeSO_4$ x 7 H_2O	2.78 g/l
Na_2 x EDTA	3.37 g/l

This solution is autoclaved after preparation.

Vitamin solution:

Thiamine dichloride 1 mg/ml; pyridoxine x HCl and nicotinic acid each 5 mg/ml

1-Naphthalenacetic acid stock solution (NAA):

20 mg NAA were dissolved in an Eppendorf tube in 1 ml pure ethanol and added to 80 ml water. The solution was then made up to exactly 100 ml in a volumetric flask. The stock solution was stored at 4 °C.

Indole-3-acetic acid solution (IAA):

50 mg IAA were initially pre-dissolved in 1 ml pure ethanol, added to 80 ml water and then made up to 100 ml. The stock solution was stored at 4 °C.

Benzylaminopurine solution (BAP):

20 mg BAP were dissolved in an Eppendorf tube in 1 ml warm 0.5 M HCl and then added with stirring to about 80 ml water. It was then made up to 100 ml in a volumetric flask. The stock solution was stored at 4 °C.

Composition of 1 l MS-Lf:

Solution of macroelements	40 ml
Solution of microelements	10 ml
Sucrose	40 g
Myo-inositol	100 mg
Glycine solution (200 mg/100 ml)	1 ml
vitamin solution	1 ml
IAA solution	2 ml
BAP solution	5 ml

The pH of the medium was adjusted to 5.9 with 0.5 M KOH and the medium was autoclaved after preparation.

7.2 Lysogeny Broth (LB)

LB is the standard medium for the cultivation of bacteria such as *Escherichia coli* (Bertani, 1951).

Composition of 1 l LB:

Tryptone / peptone	10 g
Yeast extract	5 g
NaCl	10 g
± Agar	10 g

The solution was adjusted with 0.5 M NaOH to pH 7.0 and autoclaved. 1% agar was added additionally to produce solid medium for plates.

7.3 Super Optimal broth with catabolite repression (SOC)

SOC medium is a nutritious bacterial growth medium that is adjusted from the LB medium. It provides a higher transformation efficiency. Hence, it was used to regenerate bacteria after heat shock (Hanahan, 1983).

Composition of 1 l SOC:

Tryptone / peptone	20 g
Yeast extract	5 g
NaCl	0.5 g

The solution was adjusted with 1 M NaOH to pH 7.0 and autoclaved. Then sterile solutions of 1 M glucose (20 ml), 1 M KCl (2.5 ml), 1 M $MgCl_2$ (10 ml) were added.

7.4 Yeast extract Peptone Dextrose medium (YPD)

This specially adapted medium was used for *S. cerevisiae* cell lines for simple cultivation purposes. It aimed to increase the cell mass of precultures.

Composition of 1 l YPD:

Tryptone / peptone	20 g
Yeast extract	10 g
Glucose	20 g
± Agar	20 g

Agarose was added only to make solid medium for plates. In order to prevent possible heat-induced degradation processes, the glucose solution was filter-sterilised and added after autoclaving. YPD media and plates were stored at 4 °C in the dark.

7.5 SC and SC^+ medium

SC^+ is a variant of Synthetic Complete Minimal Defined Medium (SC). SCD / SCG / SC^+G (synthetic glucose or galactose) medium was used for expression of recombinant *S. cerevisiae* strains. The SCD was used mainly for the accumulation of biomass in precultures and the galactose-containing SCG / SC^+G medium for induction.

The amino acid and nucleobase composition for media omit uracil explicitly was used to produce selective medium for growing pYes2/NT C transformants. The following amino acid and nucleobase mixture was used:

1 g each: adenine, arginine, cysteine, leucine, lysine, threonine, tryptophan
0.5 g each: aspartic acid, histidine, isoleucine, methionine, phenylalanine, proline, serine, tyrosine, valine

Composition of 1 l SCD:

Yeast Nitrogen Base	6.7 g
Glucose	20 g
Amino acid mix	1.15 g
± Agar	20 g

Composition of 1 l SCG:

Yeast Nitrogen Base	6.7 g
Galactose	20 g
Amino acid mix	1.15 g
± Agar	20 g

Composition of 1 l $SC^{+}G$:

Yeast Nitrogen Base	6.7 g
Galactose	20 g
Amino acid mix	1.15 g
± Agar	20 g
Ammonium iron (II) sulfate	28 mg
5-Aminolevulinic acid	75 mg

Agar was added only to make solid medium for plates. 100 ml of 200 g/L filter-sterilised glucose or galactose were added later to 900 ml of the pre-cooled medium. Plates and media were stored at 4 °C.

8. Buffers and solutions

Uses	Buffers and solutions	Components	Concentration
Agarose gel electrophoresis	1x TAE buffer	TRIS	40 mM
		Acetic acid	20 mM
		Na_2-EDTA	1 mM
	6x loading buffer	Bromophenol blue	0.03%
		Xylene cyanol	0.03%
		Glycerol	60%
		EDTA	60 mM
Genomic DNA extraction	2x CTAB	Cetyltrimethylammonium-bromide	2% w/v

		TRIS-HCl, pH 8	100 mM
		EDTA	20 mM
		NaCl	1.4 M
		Polyvinylpyrrolidone MW 40000	1% w/v
	10x CTAB	Cetyltrimethylammonium-bromide	10% w/v
		NaCl	0.7 M
	CTAB precipitation buffer	Cetyltrimethylammonium-bromide	1% w/v
		TRIS-HCl, pH 8	50 mM
		EDTA	10 mM
	High Salt TE	TRIS-HCl, pH 8	10 mM
		EDTA	1 mM
		NaCl	1 M
Isolation of microsomes	Microsome buffer 1 (for cytochrome P450)	TRIS-HCl, pH 7.5	0.1 M
		Dithiothreitol	1 mM
		Diethyldithiocarbamate	1 mM
	Microsome buffer 2 (for cytochrome P450 reductase)	TRIS-HCl, pH 7.6	0.5 M
		Dithiothreitol	1 mM
		Diethyldithiocarbamate	1 mM
Isolation of recombinant proteins	Breaking buffer	KH_2PO_4/K_2HPO_4 buffer, pH 7.4	50 mM
		Glycerol	5%
		Phenylmethylsulfonyl fluoride	1 mM
	His-tag binding buffer	KH_2PO_4/K_2HPO_4 buffer, pH 8.0	50 mM
		NaCl	300 mM
		Imidazole	10 mM
	Wash buffer	KH_2PO_4/K_2HPO_4 buffer, pH 8.0	50 mM
		NaCl	300 mM
		Imidazole	20 mM
	Elution buffer	KH_2PO_4/K_2HPO_4 buffer, pH 8.0	50 mM
		NaCl	300 mM
		Imidazole	250 mM
Plasmid preparation	Buffer P1	TRIS-HCl, pH 8.0	50 mM
		EDTA	10 mM
		RNase A	100 μg/ml
	Buffer P2	NaOH	200 mM
		SDS	1%
	Buffer N3	Guanidine-HCl	4.2 M
		Potassium acetate, pH 4.8	0.9 M

	Buffer PB	Guanidine-HCl	5 M
		Isopropanol	30%
	Buffer PE	TRIS-HCl, pH 7.5	10 mM
		Ethanol	80%
RNA extraction	Solution D	Guanidinium thiocyanate solution	4 M
		Citrate buffer, pH 7.0	25 mM
		Lauryl sarcosine	0.50%
SDS-PAGE	SDS running buffer	Glycine	192 mM
		TRIS-HCl, pH 8.3	25 mM
		SDS	0.10%
Transformation of yeast	TE buffer	TRIS-HCl	10 mM
		Na_2-EDTA, pH 8.0	1 mM
Western Blot	TBS buffer	TRIS-HCl, pH 7.4	10 mM
		NaCl	0.90%
	TBS-T buffer	TRIS-HCl, pH 7.4	10 mM
		NaCl	0.90%
		Tween 20	0.05%
	Washing buffer	KH_2PO_4/K_2HPO_4 buffer, pH 8.0	50 mM
		NaCl	300 mM
		Imidazole	20 mM
	Western Blot transfer buffer	TRIS-HCl, pH 8.3	25 mM
		Glycine	192 mM
		MeOH	20%
	Western Blot substrate buffer	TRIS-HCl, pH 9.5	100 mM
		NaCl	100 mM
		$MgCl_2$	5 mM

9. Bioinformatic tools

Bioinformatic tool	Web address
Blastp	http://blast.ncbi.nlm.nih.gov
Blastx	http://blast.ncbi.nlm.nih.gov
Clustal Omega	https://www.ebi.ac.uk/Tools/msa/clustalo/
ExPASy translate	https://web.expasy.org/translate/
ExPASy-Compute pI/Mw	https://web.expasy.org/compute_pi/
Oligo Calc	https://www.eurofinsgenomics.eu/
ORF Finder	http://www.bioinformatics.org
PSI- and PHI-BLAST	https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
Phyre^2 Investigator	http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index

V. Methods

1. Cultivation of suspension cultures

As a basis for this investigation, *Linum flavum* was available as a suspension culture. The cultures were inoculated every two weeks under the sterile bench into fresh MS-Lf medium. 5 g of cells were transferred with a perforated spoon into 50 ml medium. The cultures were shaken on an orbital shaker (150 rpm) at 25 °C to supply the cells with O_2.

2. Molecular biology

2.1 Genomic DNA (gDNA) extraction

The isolation of genomic DNA was carried out according to a modified protocol of Rogers and Bendich (1985).

The cells were separated from the medium via a Büchner funnel and then pulverised in liquid nitrogen. 200 mg of this powder was mixed in a 1.5 ml reaction tube with 300 µl pre-warmed (65 °C) 2x CTAB and incubated at 65 °C for 10-30 minutes, then chilled on ice for 5-30 minutes. After the addition of 300 µl chloroform and vigorous shaking, the cells were centrifuged at 16000 g for 5 minutes. The aqueous supernatant from the extraction was pipetted into a new tube with 30 µl pre-warmed (65 °C) 10x CTAB, mixed with 300 µl chloroform and again centrifuged for 5 minutes at 16000 g. After that, 250 µl of the aqueous supernatant was mixed in a new tube with 250 µl CTAB precipitation buffer and centrifuged for 5 minutes at 16000 g and the liquid discarded. The pellet was gently resuspended in 200 µl High Salt TE, then mixed with 400 µl 96% ethanol and precipitated at -20 °C for 15 minutes. After centrifugation for 15 minutes at 16000 g, the pellet was washed with 70% ethanol and air-dried. The gDNA was re-dissolved in 20-50 µl water and remaining RNA digested with 0.75 µl RNase H for 15-30 minutes at 37 °C. Subsequently, the concentration and purity of the DNA were measured photometrically.

2.2 RNA extraction

Total RNA from fresh leaves and suspension cultures of *Linum flavum* was prepared by using a phenol-chloroform extraction according to Chomczynski and Sacchi (1987).

All materials required for RNA extraction were autoclaved twice or heated to 200 °C for 2 hours. The cells were separated from the medium via a Büchner funnel and the cells or leaves were frozen in liquid nitrogen and then finely ground in a mortar. All material was kept cold to prevent any enzymatic activity that could destroy the RNA. About 50 mg of this powder was mixed with 500 µl Solution D and incubated at room temperature. The solution was mixed gently after addition of 50 µM 2 M sodium acetate pH 4. Thereafter, 500 µl phenol saturated with citrate buffer pH 2 and 100 µl ice-cold chloroform were added. The reaction tube was mixed for 10 seconds and incubated at 4 °C for 15 minutes. Subsequently, the tube was centrifuged for 15 minutes at 12000 g and 4 °C. 400 µl of the supernatant was transferred to a new tube and mixed with the same volume of ice-cold isopropanol. The RNA was precipitated by incubation at -20 °C for at least 20 minutes. Afterwards, the samples were centrifuged for 10 minutes at 12000 g and 4 °C. The supernatant was then removed and the pellet was washed twice with ethanol (70% and 100%). The pellet was dried at 37 °C and dissolved with 20 µl water at 50 °C. The amount of RNA and the purity were measured photometrically, the integrity was monitored on an agarose gel. The RNA was kept at -20 °C. Samples containing a high amount (> 500 ng μl^{-1}) of pure RNA (ratio 260/280 nm > 1.6) were selected for further experiments.

2.3 cDNA synthesis

The cDNA was synthesized by using the RevertAid First Strand cDNA Synthesis Kit.

1 µg RNA and 1 µl Oligo$(dT)_{18}$ primers were incubated at 65 °C for 5 min with sterile H_2O in a total volume of 12 µl and then cooled on ice. Subsequently, 4 µl of the 5x reaction buffer, 1 µl Ribolock RNase inhibitor and 2 µl 10 mM dNTP mix were added. The mixture was incubated for 5 minutes at 37 °C and 1 µl RevertAid reverse transcriptase was added. Thereafter, the mixture was incubated at 42 °C for one hour. The reaction was stopped by heating for 5 min at 70 °C. The cDNA was stored at -20 °C.

2.4 PCR

2.4.1 Standard PCR with GoTaq® polymerase

The standard PCR with GoTaq® polymerase was used to amplify the candidate-genes from cDNA or gDNA of *Linum flavum*. The primers were designed based on the sequences of

candidate-RNA in *Linum flavum*, which is available on the website http://www.medplantrnaseq.org/. The standard PCR had a total volume of 25 µl and was composed as below:

Template (cDNA or gDNA)	1.0 µl
GoTaq® buffer (5x)	5.0 µl
dNTP mix (10 mM)	0.5 µl
$MgCl_2$ (25 mM)	3.0 µl
Primer forward (100 µM)	0.5 µl
Primer reverse (100 µM)	0.5 µl
GoTaq® polymerase	0.1 µl
Dist. H_2O	14.4 µl

The PCR cycles were the following:

1st cycle:	94 °C, 120 sec / Annealing temperature, 60 sec / 70 °C, 90 sec
2nd - 39th cycle:	94 °C, 30 sec / Annealing temperature, 60 sec / 70 °C, 90 sec
40th cycle:	94 °C, 60 sec / Annealing temperature, 60 sec / 70 °C, 10 min
End:	cooling at 4 °C.

Annealing temperature was usually chosen at T_m - 5 °C. T_m (the melting point of primers) was calculated with a bioinformatic tool on the Eurofins® website.

Agarose gel electrophoresis (see V.2.5) was used to analyse the product of the PCR. Products of the expected size were cut out and purified via gel extraction (as described in V.2.6).

2.4.2 PCR with Phusion® High-Fidelity DNA polymerase

The PCR was performed with a Phusion® High-Fidelity DNA polymerase to avoid replication errors due to proofreading activity. PCR reactions were composed as follows:

HF buffer + $MgCl_2$ (5x)	5 µl
dNTP Mix (10 mM)	1 µl
Primer forward (100 µM)	0.5 µl
Primer reverse (100 µM)	0.5 µl
Template	1 µl
Phusion® High-Fidelity DNA polymerase	0.25 µl
Dist. H_2O	*ad* 25 µl

The PCR cycles were the following:

1st cycle:	98 °C, 30 sec / Annealing temperature, 30 sec / 72 °C, 30 sec
2nd - 39th cycle:	98 °C, 10 sec / Annealing temperature, 30 sec / 72 °C, 30 sec
40th cycle:	98 °C, 10 sec / Annealing temperature, 30 sec / 72 °C, 10 min
End:	cooling at 4 °C.

The PCR products were analysed by agarose gel electrophoresis (see V.2.5). The products of expected size were cut out and purified by gel extraction (as described in V.2.6).

2.4.3 Fusion-PCR

Fusion-PCR was used for the *in vitro* removal of introns from genomic DNA to generate a full-length continuously coding sequence when the genes of interest could not be amplified successfully from complementary DNA. Prior to designing the primers, it is important to define the boundaries for each exon and intron. The two outer (sense and antisense) primers contain restriction enzyme recognition sites, which are important for the ligation into an expression plasmid. The internal primers are constructed to span the fusion point between two adjacent exon fragments. All internal primers were 30 nucleotides in length to ensure that primers bind specifically to the target region of gDNA and include an exon. The first half of internal forward primers were identical to the last 15 nucleotides of an upstream exon and the second half corresponded to the first 15 nucleotides of a downstream exon. The inverse internal primers were the reverse complement sequences of the respective forward primers.

A first-round amplification is used to generate different exon fragments. The PCR was performed with a Phusion® High-Fidelity DNA polymerase to avoid replication errors. PCR reactions were composed as follows:

HF buffer + $MgCl_2$ (5x)	5 µl
dNTP Mix (10 mM)	1 µl
Primer forward (100 µM)	0.5 µl
Primer reverse (100 µM)	0.5 µl
Template	1 µl
Phusion® High-Fidelity DNA Polymerase	0.25 µl
Dist. H_2O	ad 25 µl

The PCR cycles were the following:

1st cycle:	98 °C, 30 sec / Annealing temperature, 30 sec / 72 °C, 30 sec
2nd - 39th cycle:	98 °C, 10 sec / Annealing temperature, 30 sec / 72 °C, 30 sec
40th cycle:	98 °C, 10 sec / Annealing temperature, 30 sec / 72 °C, 10 min
End:	cooling at 4 °C.

The PCR products were separated by agarose gel electrophoresis (see V.2.5). The products of expected size were cut out and purified via gel extraction (see V.2.6). After purification and determination of product concentrations, the first-round PCR products were diluted to the same concentration and used as a template in a second-round amplification. A second-round of fusion-PCR used only the outermost sense and antisense oligonucleotide primers. PCRs were set up to contain the following components:

HF buffer + $MgCl_2$ (5x)	5 µl
dNTP Mix (10 mM)	1 µl
Primer forward (100 µM)	0.5 µl
Primer reverse (100 µM)	0.5 µl
Template 1 (exon 1)	0.5 µl
Template 2 (exon 2)	0.5 µl
Phusion® High-Fidelity DNA Polymerase	0.25 µl
Dist. H_2O	*ad* 25 µl

The second round of fusion-PCR was set up in an identical protocol to that employed for the first-round amplification.

The PCR products were separated by agarose gel electrophoresis (see V.2.5). The distinct PCR products were excised and purified from agarose gel (see V.2.6). After purification, the products were ligated into pDrive (as described in V.2.7.1) and sequenced before being used in further experiments to determine the function of the candidate genes.

2.4.4 Colony-PCR

Colony-PCR is a convenient method for confirmation of the successful transformation of yeast expression cell lines. The colony that was later used for expression was picked from the agar plate and suspended in 100 µl 200 mM LiAc, 1% SDS solution and incubated for 5 minutes at 70 °C. 300 µl 96% EtOH were added and mixed vigorously for further purification. Afterwards, the preparation was centrifuged at 15000 g for 3 min and the resulting pellet was washed in 70% EtOH. After drying, it was resuspended in 50 µl H_2O and centrifuged for 15 s at 15000 g.

The supernatant was then used as template for a standard PCR reaction with specific primers for the respective candidate genes. The verified colonies were inoculated in 2 ml SCD medium in test tubes for 36 hours at 30 °C and mixed with glycerol (ratio 1:1) then stored at -70 °C for further expression analysis.

2.5 Agarose gel electrophoresis

0.7% agarose gels were used for the separation of DNA and RNA as a matrix. 1.4 g agarose was dissolved in 200 ml 1x TAE buffer by heating in a microwave oven. Approximately 50 ml were filled into a suitable gel tray with a slot-forming comb. 0.7 µl 1% ethidium bromide solution was added to the gel and mixed well. The samples were mixed with 6x loading buffer and applied to the slots of the gel. The size marker was the GeneRuler DNA Ladder Mix from Thermo Fisher Scientific with a size range of 100 to 10,000 bp. The RNA or DNA mixtures were separated with 220 V. The run was terminated when the lower dye band (bromophenol blue) had reached 2/3 of the gel. The photographic documentation was done digitally using blue-green LED light and an amber filter.

2.6 Purification of DNA fragments from agarose gels

The bands containing a DNA fragment with the correct calculated length were excised completely from the gel after gel electrophoresis. A NucleoSpin® Extract II kit was used for gel extraction and purification. 200 µl NT binding buffer per 100 mg of the gel was added to the gel and heated to 50 °C with shaking until liquefaction. The solution was completely centrifuged through a NucleoSpin® silica filter at 11000 g for one minute. The flow-through was discarded. The silica filter was washed with 700 µl buffer NT3. Subsequently, the silica filter was dried by centrifugation at 11000 g for 2 minutes. Elution was carried out with 15-50 µl water and incubation for 2-5 minutes at room temperature. The eluate was collected in reaction tubes by centrifugation of the silica filter at 11000 g for one minute. The DNA concentration of the sample was measured photometrically.

2.7 Ligation

2.7.1 UA-ligation

UA- or TA-ligation is a simple way to incorporate inserts into vectors. It is based on the attachment of an adenine at the end of the PCR reaction of DNA polymerase type A to the finished strand. This adenine can interact with an overhanging thymine or uracil of the linear plasmid, hence facilitating the incorporation of the insert into the vector. The UA-ligation was performed with Qiagen's PCR cloning kit. 2 µl insert and 0.5 µl pDrive cloning vector were pipetted to 2.5 µl 2x ligation mix in a total volume of 5 µl and incubated at 4 °C overnight or at 19 °C for at least 2 hours. The ligation was then heated to 65 °C for 10-15 minutes, cooled and used for transformation.

2.7.2 T4-ligation

T4-ligation was performed with T4 DNA ligase and DNA fragments, which were previously cut by a restriction endonuclease. The ligase occurs naturally in the bacteriophage T4. The DNA fragment and plasmid were digested separately with the same restriction enzymes. Different restriction endonucleases used for the two restriction sites lead to a directional incorporation of the insert. The ligation mixture had a total volume of 10 µl.

Plasmid	1 µl
T4 ligase buffer (10x)	1 µl
T4 DNA ligase	1 µl
DNA fragment solution	7 µl

The mixture was incubated for 20 hours at 4 °C. Subsequently, the reaction was stopped at 65 °C for 10 minutes.

2.8 Restriction enzyme digest

Standard digestion was performed in 20 µl volume. The pipetting scheme described here was used:

10x buffer	2 µl
Restriction enzyme 1	1 µl
Restriction enzyme 2 (optional)	1 µl
Water	10 µl

DNA solution	*ad* 20 µl

The choice of final buffer concentration was dependent on the enzyme(s) used. It was aimed to maximise enzyme activity and exclude nonspecific digestion (star activity). The choice of enzyme followed the restriction sites in the gene or vector. The samples were incubated for 3 hours at 37 °C. The digested DNA fragments and plasmid were analysed on an agarose gel, then purified on silica columns.

When the confirmation of the presence of an insert in pDrive vector was needed, only restriction enzyme EcoRI was used and the total volume was cut in a half. A brief microwave heating of the preparation (30 seconds) was sufficient to detect bands after agarose gel electrophoresis.

2.9 Site-directed mutagenesis

In vitro site-directed mutagenesis was done according to instruction of Stratagene's QuikChange® Site-Directed Mutagenesis Kit. It is a technique for carrying out gene modification and was used to make point mutations to switch amino acids. The site-directed mutagenesis method is performed using PfuTurbo® DNA polymerase and a temperature cycler. PfuTurbo® DNA polymerase replicates both plasmid strands with high fidelity and without displacing the mutant oligonucleotide primers. The reaction was prepared as indicated below:

10x reaction buffer	5 µl
dsDNA template (5-50 ng)	1 µl
Oligonucleotide primer 1 (125 ng)	1 µl
Oligonucleotide primer 2 (125 ng)	1 µl
dNTP mix (10 mM)	1 µl
PfuTurbo® DNA polymerase (2.5 U/µl)	1 µl
H_2O	*ad* 50 µl

The temperature cycles were the following:

1st cycle:	95 °C, 30 sec
2nd - 18th cycle:	95 °C, 30 sec / 55 °C, 60 sec / 68 °C, 8 min
End:	cooling at 4 °C.

The oligonucleotide primers, each complementary to opposite strands of the vector, are extended during temperature cycling by PfuTurbo® DNA polymerase. Incorporation of the oligonucleotide primers generates a mutated plasmid containing staggered nicks. Following temperature cycling, 1 µl of the Dpn I (10 U/µl) was added directly to each amplification product and incubated at 37°C for 1 hour. The DpnI endonuclease is specific for methylated and hemimethylated DNA and is used to digest the parental DNA template and to select for mutation-containing synthesized DNA.

3. Genetic engineering

3.1 Chemically competent *E. coli* cells

E. coli strains EZ from Qiagen and SoluBL21 from AmsBio were used for the genetic engineering work. The EZ and SoluBL21 strains were propagated and made competent in the laboratory. EZ cells served to increase the number of plasmids and SoluBL21 to express proteins.

Cells of the old batch are inoculated by means of a sterile toothpick in a test tube with 2 ml liquid LB medium with 2 µl tetracycline (12.5 mg/ml tetracycline in ethanol). The test tubes are closed with a lid and shaken overnight at 37°C and 220 rpm (overnight culture). The next morning, from a 250 ml Erlenmeyer flask with ~90 ml sterile LB + 90 µl tetracycline, 0.75 ml of solution was taken and used as a reference for the determination of optical density value at 600 nm (OD_{600}). 2 ml of the overnight culture are added into the Erlenmeyer flask and incubated at 37 °C, 220 rpm until an OD_{600} of 0.6 is reached. When the required OD_{600} is reached, the suspensions are centrifuged in two sterile 50 ml tubes for 10 minutes at 4 °C and 3000 g. The supernatant was discarded and the pellet was resuspended in 10 ml cold sterile 100 mM $CaCl_2$ solution. The suspensions were combined and centrifuged for 12 minutes at 4 °C, 2500 g. Again, the supernatant was then discarded and the pellet resuspended in 10 ml cold sterile 100 mM $CaCl_2$ solution. The solution is incubated on ice for 20 minutes and recentrifuged (12 minutes, 4 °C, 2500 g). Finally, the supernatant was discarded and the cells were resuspended in 2 ml 100 mM sterile $CaCl_2$ containing 15% glycerol. The bacteria were distributed to 150 µl in sterile 1.5 ml reaction tubes, pre-frozen in liquid nitrogen and stored at -80 °C.

3.2 Transformation of *E. coli* by heat shock

Prior to the transformation, the LB agar plates were prepared. 100 ml solid LB medium was heated by microwave and mixed with 500 µl ampicillin (20 mg/ml in water, sterile-filtered), For transformation of *E.coli* EZ, 160 µl X-gal (50 mg/ml in dimethylformamide) and 50 µl IPTG (0.1 M in water, sterile-filtered), then poured into Petri-dishes. The plates containing approximately 7.5 ml of the mixture were cooled in a sterile bench.

The competent cells were thawed on ice shortly before use. 70 µl of the bacterial suspension were pipetted into the reaction tubes with the vector and incubated on ice for 30 minutes. The cells were heat-shocked in a water bath at 42 °C for 90 seconds, followed by cooling on ice for 1-2 minutes. 150 µl SOC medium was added to the preparation. The cells were regenerated in a Thermomixer® at 37 °C for 45 minutes. Subsequently, the mixture was spread on LB agar plates with a heat-sterilised Drigalski scoopula and cultured at 37 °C overnight. Clones containing pDrive could be selected via blue-white screening for integrated DNA fragments.

3.3 Transformation of yeast strain INVScI

The transformation of *S. cerevisiae* strains was carried out using the lithium acetate method according to Gietz and Schiestl (2002). Carrier DNA (fish sperm) in TE buffer (2 mg/ml) was dissolved by stirring at 4 °C for 1-2 hours. Prior to transformation, carrier DNA was heated to 99 °C for 5 minutes and stored on ice. The plasmid DNA was diluted with dist. H_2O to a concentration of 50 ng/µl. An amount of yeast cells (appr. 50 µl) was picked from a fresh YPD plate and transferred into a tube with 1 ml dist. H_2O. The suspension was centrifuged at 3000 g for 30 seconds and the supernatant was then removed. The following was pipetted onto the cell pellet:

Polyethyleneglycol (50 % w/v)	240 µl
1 M lithium acetate	36 µl
Carrier DNA (2 mg/ml)	50 µl
Plasmid (50 ng/µl)	34 µl

The samples were incubated for 3 h at 42 °C, then centrifuged at 16000 g and resuspended in 200 µl H_2O. In the final step, the preparation was spread on SCD plates and incubated for 3-4 days at 30 °C. When the *S. cerevisiae* cells were successfully transformed, colonies were streaked onto fresh SCD plates and checked for correct insertion by colony PCR.

3.4 Overnight cultures

The grown *E. coli* colonies were propagated in overnight cultures. For this purpose, 20 µl ampicillin (20 mg/ml in water, sterile-filtered) were mixed with 4 ml LB in test tubes. A bacterial colony was picked using sterile toothpicks and placed in a tube. The tubes were cultured at 37 °C and 220 rpm overnight. Thereafter, plasmids were extracted from the cells by plasmid preparation (see 3.5).

3.5 Plasmid preparation

Plasmids were extracted from bacteria using Qiagen's plasmid preparation method.

3 ml bacterial suspension of the overnight cultures were centrifuged in two steps in 1.5 ml reaction tubes at 3000 g for 3 minutes. The supernatant was completely discarded. The cells were resuspended with 250 µl buffer P1, lysed with 250 µl buffer P2 and neutralised with 350 µl buffer N3. The solution was centrifuged for 10 minutes at 16000 g. Thereafter, the supernatants were pipetted to QIAprep® spin columns and centrifuged for 1 min at 16000 g. The flow-through was then discarded. 500 µl buffer PB and 750 µl PE were sequentially centrifuged through the column for 30 seconds at 11000 g. The columns were dried by centrifugation for 1 minute (11000 g) and 40 µl water pipetted onto the column material. After incubation at room temperature for 2-5 minutes, the columns were centrifuged for 1 min at 11000 g and the flow-through containing the plasmid-DNA was collected in 1.5 ml test tubes. The columns were washed with 500 µl water, centrifuged and stored for the next use.

3.6 Sequencing and preparation of glycerol stocks

The vectors obtained from plasmid preparation were subjected to a restriction digest in a total volume of 10 µl and the product was analysed by gel electrophoresis. The plasmid solutions were adjusted to a concentration of 80-100 ng/µl with a total volume of 12 µl in new reaction tubes and sent to Seqlab® for sequencing. From colonies with correct sequences, glycerol stocks were established. 425 µl liquid bacteria culture (overnight cultures) was mixed with 75 µl sterile glycerol. The mixture was frozen in liquid N_2 and stored at -80°C.

3.7 Expression of recombinant proteins

3.7.1 *E. coli* SoluBL21 strains

Vectors containing the open reading frame in the correct reading direction were used for the synthesis of heterologous proteins. These were introduced into *E. coli* SoluBL21 by heat-shock transformation. 2 ml LB with 10 µl ampicillin (20 mg/ml) was inoculated with a small aliquot of a frozen glycerol stock and incubated overnight at 37 °C and 220 rpm. The next day, 0.75 ml sample was taken from 100 ml liquid LB medium supplemented with 100 µg/ml ampicillin for the photometric reference. The overnight culture was added to the remaining LB and cultured at 37 °C, 220 rpm to an OD_{600} of 0.5. Induction of protein expression was carried out with 100 µl 1 M IPTG (final concentration 1 mM) and the cells were incubated for 5 h at 28 °C, 220 rpm. Afterwards, the cells were centrifuged for 5 min at 5000 g in 50 ml tubes and the supernatant discarded. The cells were frozen in liquid N_2 and stored at -80 °C until further homogenisation and purification steps.

3.7.2 *S. cerevisiae* InvSc1 strain

Colonies of the INVSc1 strain containing the correct pYES2/NT C construct were inoculated and grown for 48 h at 30 °C in 10 ml SCD medium as preculture. These were further transferred to 140 ml SCG or $SC^{+}G$ medium in 1 L baffled flasks and shaken at 220 rpm and 30 °C for 24 h to induce expression. The next day, the cells were collected by centrifugation at 1500 g for 5 minutes at 4 °C.

For the expression analysis, 1 ml of the expression culture was transferred to a 1.5 ml tube after 0, 4, 8, 12 and 24 h. The OD_{600} of the expression cultures was measured photometrically. Cell pellets were collected by brief centrifugation and analysed for protein expression by SDS-PAGE (see V.4.6).

4. Enzymology

4.1 Isolation of microsomes

4.1.1 Isolation of microsomes from plant cell cultures

Cells of *Linum flavum* were cultured at 26 °C and permanent light on the shaker in MS-Lf medium. They were harvested with the help of a water jet pump through a Büchner funnel with filter paper. The fresh weight was determined and about 20% of this weight of Polyclar 10 was added. This is intended to bind phenolic substances and thus prevent the activity of protein-inactivating phenolases. The mixture was ground in an ice-cooled mortar along with 1 ml microsome buffer 1 or 2 (see IV.8) per gram fresh weight. To separate the solid cell components, the batch was then centrifuged for 20 minutes at 4 °C and 8000 g. 1 M $MgCl_2$ solution was added dropwise to the supernatant up to a final concentration of 50 mM. This preparation was stirred for at least 20 minutes on ice. After a 20 min centrifugation at 48000 g, the supernatant was discarded and the microsomes were washed with 1 ml buffer and resuspended in up to 2 ml of the same buffer. Forming of membrane vesicles was facilitated by homogenisation of the suspension by a Potter-Elvehjem glass homogenizer. Up to about 6 ml buffer was added before the suspension was used for determination of enzyme activities.

4.1.2 Isolation of microsomes from yeast cells

The yeast cells of 200 ml culture were successively transferred to 50 ml tubes and centrifuged for 5 minutes at 1500 g. After discarding the supernatant, they were resuspended in 27 ml microsome buffer 1 or 2 (see IV.8) and incubated for 5 min at room temperature. This was followed by another centrifugation step for 5 min at 1500 g. The supernatant was discarded. All subsequent steps were performed at 4 °C or on ice. The pellet was completely resuspended in 2.5 ml ice-cold microsome buffer. Disintegration was conducted by vigorous shaking in a benchtop homogeniser (Minilys®) at 4000 rpm for 30 seconds and then cooling on ice for 30 seconds with 4 repetitions. Afterwards, the tubes were washed four times with 5 ml microsome buffer each and the respective supernatants collected in a fresh tube. To remove the cell debris and glass beads from the extract, another centrifugation step was carried out at 8000 g for 10 min. 1 M $MgCl_2$ solution was added slowly to the supernatant up to a final concentration of 50 mM and stirred for 10 minutes. The preparation was centrifuged for 20 minutes at 48000 g, the supernatant was discarded and the microsomes were washed with 1 ml microsome buffer and resuspended in 2 ml of the same buffer. A Potter-Elvehjem glass homogenizer was used to facilitate the formation of membrane vesicles. The microsomes were used to determine the activities of heterologously expressed proteins.

4.2 Isolation of recombinant proteins

4.2.1 *E. coli* strains (genetically modified)

Proteins were synthesized by in *E. coli* SoluBL21 (see V.3.7.1). Afterwards, the bacterial cultures were successively transferred to 50 ml tubes and centrifuged at 4 °C for 5 minutes and 3000 g. After discarding the supernatants, the pellet was frozen in liquid N_2 and stored at -80 °C. Thereafter, the cells were mixed with 0.1 M KH_2PO_4/K_2HPO_4 buffer pH 7.5 (4 ml per 1 g pellet) and thawed on ice for 30 minutes. The cells were homogenised three times with an ultrasonic processor on ice. The intensity setting was 100%, the amplitude 0.3 for 60 seconds each. The suspension was then centrifuged for 10 minutes at 4 °C and 6000 g. The supernatant containing heterologous proteins was used for purification by metal chelate chromatography.

4.2.2 *Saccharomyces cerevisiae* strain (genetically modified)

Proteins were synthesised in *S. cerevisiae* InvSc1 for 24 hours. The cultures were transferred successively into 50 ml tubes and centrifuged at 4 °C for 5 min at 1500 g. The cell pellet was frozen in liquid N_2 and stored at -80 °C. For disintegration, the cells were resuspended in 1 ml breaking buffer per 1 g cell pellet, thawed on ice and centrifuged for 5 min at 4 °C, 1500 g. Subsequently, the supernatant was discarded and the pellet was resuspended in the breaking buffer (1 ml/g) and the same volume of acid-washed glass beads (~ 0.5 mm ø) added. Disintegration was done by vigorous shaking in a benchtop homogeniser (Minilys®) at 4000 rpm for 30 seconds, followed by cooling on ice for 30 seconds and 2-3 repetitions. Thereafter, the suspensions were centrifuged for 10 min at 4 °C, 13000 g. The supernatant containing heterologous protein was transferred to fresh tubes.

4.3 Purification of histidine-tagged proteins

A fusion protein formed with an artificially added histidine hexamer can be purified by affinity chromatography. The histidine tag can be located either at the *N*- or *C*-terminal end of the target.

1 ml of the suspended nickel-NTA column material was pipetted into plastic columns and washed several times with water. 1 ml of His-tag binding buffer (see IV.8) was added to the

column material and the nickel-NTA equilibrated for 30 min and then drained. 500 mM KPi, 3 M NaCl, 100 mM imidazole pH 8.0 in a 1:10 ratio was mixed with the protein sample and the column material. The column was then closed, covered in ice, and gently rocked for one hour. Thereafter, the liquid was drained and 2-3 drops were taken for the SDS-PAGE. The column material was then washed six times with 2 ml washing buffer (see IV.8). Aliquots of the washing fractions were collected for further SDS-PAGE analysis. Thereafter, the bound protein was eluted three times with 1 ml elution buffer each (see IV.8). These fractions were combined and desalted via PD-10 columns (see V.4.4).

4.4 Desalting via PD-10 columns

The desalting of samples was done with PD-10 columns based on the manufacturer's instructions (GE Healthcare). Depending on expressed proteins, 25 ml of the specific buffer that was used later in the enzyme assays were used to equilibrate the PD-10 columns, followed by addition of 2.5 ml protein solution and discarding the flow-through. Thereafter, the PD-10 column was eluted with 3.5 ml buffer and the eluate was collected. The columns were then washed with 25 ml water and stored in the refrigerator until further use. The protein fraction was aliquoted and frozen at -80 °C.

4.5 Determination of protein concentration

The method according to Bradford (1976) was used to determine the protein concentration in the individual samples. This is a colorimetric method in which the dye Coomassie Brilliant Blue G-250 is bound to the proteins in an acidic environment. The dye binds non-specifically to cationic and nonpolar, hydrophobic side chains of the proteins. This shifts the absorption maximum of the dye from 465 to 595 nm (Stoscheck, 1990). At this wavelength, the absorption of the samples can be measured. When the absorption is compared to bovine serum albumin (BSA, fraction V) standard (adjusted to 1 mg protein/ml) and a reference with buffer instead of protein solution, the protein concentration can be calculated.

The Bradford reagent is composed as follows:

Coomassie Brilliant Blue G-250	100 mg
96% Ethanol	50 ml
85% Phosphoric acid	100 ml
Dist. water	850 ml

This solution must be filtered twice to remove the suspended particles.

The assay for determining the protein concentration is made up as follows:

	10 µl buffer as a reference (usually 0.1 M Tris-HCl pH 7.5)
or	10 µl BSA standard 1 mg/ml
or	10 µl crude protein (also diluted)
	2 ml Bradford reagent

These preparations were incubated for 15 minutes at room temperature and then measured at 595 nm against the reference in the photometer.

4.6 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was performed according to Lämmli (1970). The proteins are denatured and form a negatively charged protein-SDS complex. This complex moves differently in an electric field depending on the molecular weight.

The discontinuous SDS gel consisted of a (lower) separating gel and a (top) collecting gel. The polymerization takes place by a radical reaction mechanism. The polymerization initiator ammonium persulphate (APS), together with tetramethylethylenediamine (TEMED) as a catalyst, leads to the polymerization of the acrylamide/bisacrylamide (37.5 : 1) mixture.

Pipetting scheme for two separating gels:

TRIS-HCl-buffer pH 8.8, 1.5 M	2.50 ml
(Bis-)/Acrylamide (30 %)	4.10 ml
Dist. H_2O	2.9 ml
SDS (10 %)	400 µl
TEMED	16 µl
APS (10 %)	70 µl

After filling the gel between the glass plates, the surface was overlaid with 400 µl n-butanol. This was completely removed before adding the stacking gel.

Pipetting scheme for two stacking gels:

TRIS-HCl-buffer pH 6.8, 0.5 M	1.25 ml
(Bis-)/Acrylamide (30 %)	0.75 ml
Dist. H_2O	2.80 ml
SDS (10 %)	200 µl
TEMED	10 µl
APS (10 %)	40 µl

The stacking gel was layered over the separating gel and a comb for pockets added. After complete polymerisation, the comb was removed and the glass plates containing the gel were fixed in the electrophoresis chamber and the buffer tank was filled with running buffer (see IV.8). 15 µl of the protein solutions was mixed with 5 µl 4x Roti®-Load sample buffer, heated at 95 °C for 10 minutes. The samples and size markers as standard were applied to the gel pockets, then run for 1.5 h at 150 V, 100 mA. After the blue band of bromophenol blue crossed the bottom edge of the gel, the run was terminated.

After the run, the collection gel was removed and the separation gel stained with 0.3 mM Coomassie-Brilliant Blue R-250 in MeOH:AcOH:H_2O (4.5:1:4.5; v/v) for 0.5-1 h. The subsequent two-fold destaining (1-2 hour each) was carried out in the same solvent mixture without Coomassie blue until the optimal result was achieved.

4.7 Western Blot

After successful SDS-PAGE, the presence of a specific protein extracted from cells can be detected by Western blotting. The SDS-PAGE separating gel was detached from the chamber and the stacking gel and equilibrated in transfer buffer (see IV.8) for 10-30 min. The membrane (PVDF, Immobilion™-P from Millipore; 0.45 µm pore size) was wetted in MeOH for 15 seconds, then washed in dist. H_2O for 2 min, followed by equilibration in transfer buffer for 5 min. Pre-cut filter paper was briefly put into transfer buffer. The single layers were assembled in the correct order: cathode (-) - 6 filter papers - SDS-gel - PVDF-membrane - 6 filter papers - anode (+) and the package installed in a semi-dry blotting system. A constant current, depending on the membrane size (length*width*2 mA), was applied for blotting for about 90 min. After blotting, the membrane was carefully removed and analysed by immunodetection.

The membrane was washed three times with TBS buffer (see IV.8) for 5 min. The membrane was blocked in TBS-T buffer with 5% milk powder for 1.5 h, then washed twice in TBS-T (see IV.8) for 5 min. Thereafter, the membrane was incubated with the first antibody (Pierce™6x-

His Epitope Tag Antibody), diluted 1:10000 in TBS-T with 1% milk powder for 16 hours at 4 °C. Subsequently, the membrane was washed eight times for 5 min in TBS-T, followed by an incubation with the second antibody (anti-mouse IgG alkaline phosphatase, diluted 1:20000 in TBS-T) for 1 h at 25 °C. The membrane was washed five times in TBS-T for 5 min. Finally, the membrane was equilibrated in substrate buffer (see IV.8) for 5 min and then dyed in 10 ml substrate buffer mixed with 80 µl BCIP (20 mg/ml in 100% dimethylformamide) and 60 µl NBT (50 mg/ml in 70% dimethylformamide) for 30 min. The reaction was stopped by washing the membrane in dist. H_2O. The dyed membranes were photographed and digitalised for later evaluation.

4.8 Enzyme activity assays

4.8.1 Cytochrome P450 reductase (CPR)

The CPR activity in yeast microsome preparations was determined according to Urban et al. (1990, 1994). The standard assay at room temperature contained:

50 mM KCN (always freshly prepared)	10 µl
2 mM cytochrome c	15 µl
Microsomal protein	5-50 µl
6 mM NADPH	25 µl
0.5 M TRIS-HCl pH 7.6	*ad* 1 ml

The reduction of cytochrome c was monitored photometrically at 550 nm. The basal slope before addition of NADPH was recorded for several minutes, then NADPH was added and the absorbance at 550 nm was further recorded. From the difference of basal slope and NADPH-dependent slope, the specific enzyme activity of the CPR was calculated as follows:

$$\text{Specific activity} = \Delta A_{550} * \varepsilon_{550} * V * t^{-1} * p^{-1}$$

with ΔA_{550}: absorption change at 550 nm
ε_{550}: differential absorption coefficient of cytochrome c (ε_{550} = 21000 M^{-1} cm^{-1})
V: reaction volume (ml)
t: reaction time (s)
p: amount of microsomal protein (g)

4.8.2 Deoxypodophyllotoxin 6-hydroxylase and deoxypodophyllotoxin 7-hydroxylase

The assay components were pipetted into 1.5 ml test tubes as described below:

0.1 M TRIS-HCl buffer pH 7.5	8.75 µl
0.1 M DTT	2.5 µl
0.1 mM DIECA	6.25 µl
25 mM NADPH	2.5 µl
Microsomal protein	225 µl
2.5 mM DOP	5 µl

The reaction was started with addition of microsomal protein, immediate shaking and incubating at 25 °C and 600 rpm in an Eppendorf Thermomixer® for 60 min. Addition of 25 µl 6 N HCl and shaking stopped the reaction. Assays directly stopped with 6 N HCl and put on ice were used as zero control samples. Afterwards, the products were extracted twice with 600 µl EtOAc each. The combined EtOAc extracts were evaporated in a centrifugal evaporator (SpeedVac®) and the dry residues redissolved in 100 µl methanol and centrifuged at 16000 g for 5 min prior to HPLC analysis.

4.8.3 Pinoresinol-lariciresinol reductase

The following procedure modified according to Fujita et al. (1999) was used. The standard assay mixture (250 µl) consisted of:

Purified His-tagged protein	25 µl
5 mM racemic pinoresinol	10 µl
25 mM NADPH	25 µl
0.1 M KPi buffer pH 7.1	*ad* 250 µl

Buffer, purified His-tagged protein and racemic pinoresinol were preincubated for 15 min at 30 °C prior to assay initiation by addition of NADPH. After 30 minutes, the assays were stopped by adding 600 µl EtOAc and mixed. Afterwards, the products were extracted twice with 600 µl EtOAc each. The combined EtOAc extracts were evaporated in a centrifugal evaporator (SpeedVac®). The dry residues were re-dissolved in 100 µl methanol and centrifuged at 16000 g for 5 min prior to HPLC analysis.

4.8.4 Secoisolariciresinol dehydrogenase

The enzyme assay modified according to Xia et al. (2000) was used. Each reaction was prepared with a total volume of 250 µl that contained:

Purified His-tagged protein	20 µl
1 M DTT	1 µl
50 mM NAD	20 µl
5 mM SECO	8 µl
20 mM TRIS-HCl buffer pH 8.8	*ad* 250 µl

The reaction was incubated at 30 °C for 2 h. Each sample was extracted twice with 600 µl EtOAc, the combined EtOAc extracts were dried in a centrifugal evaporator (SpeedVac®) and the residue reconstituted in 100 µl methanol for HPLC analyses.

4.8.5 *In vivo* biotransformation enzyme assays

Colonies of the INVSc1 strain containing the correct pYES2/NT C construct were inoculated and grown for 48 h at 30 °C in 5 ml SCD medium as pre-culture. These were further transferred to 100 ml induction medium (SCG or $SC^{+}G$) in 500 ml baffled flasks and shaken at 220 rpm and 30 °C, 24 h to induce expression. Heterologous proteins were synthesised in *S. cerevisiae* InvSc1 for 24 hours. The next day, the cultures were transferred successively into 50 ml reaction tubes and centrifuged at 1500 g for 5 minutes at 4 °C. Thereafter, the cell pellet was redissolved in 10 ml fresh induction medium and 50 µl substrate SECO or DOP in methanol (2 mg/ml) added, then shaken at 220 rpm, 30 °C for 24 h. Each culture was extracted twice with 20 ml EtOAc, the combined EtOAc extracts were dried in a centrifugal evaporator (SpeedVac) and the residue reconstituted in 1 ml methanol for HPLC analyses.

4.9 High-performance liquid chromatography (HPLC)

HPLC is an important method of separating and purifying substances in the laboratory. Many different HPLC systems can be used: analytical, semi-quantitative or quantitative HPLCs, normal phase (NP) or reverse phase (RP). The HPLC analyses were performed with a C18-RP column (Hypersil Hypurity™ Elite, 4.6 mm ø, 250 mm length with 20 mm precolumn) or a chiral column (Chiralcel OD-H) under isocratic condition. The composition of the mobile

phase remained stable throughout the time. All eluents for HPLC with C18-RP column were supplemented with 0.01% phosphoric acid. 20 µl of the samples were applied to the column. HPLC parameters for the analysis of metabolites and enzyme activities are listed below:

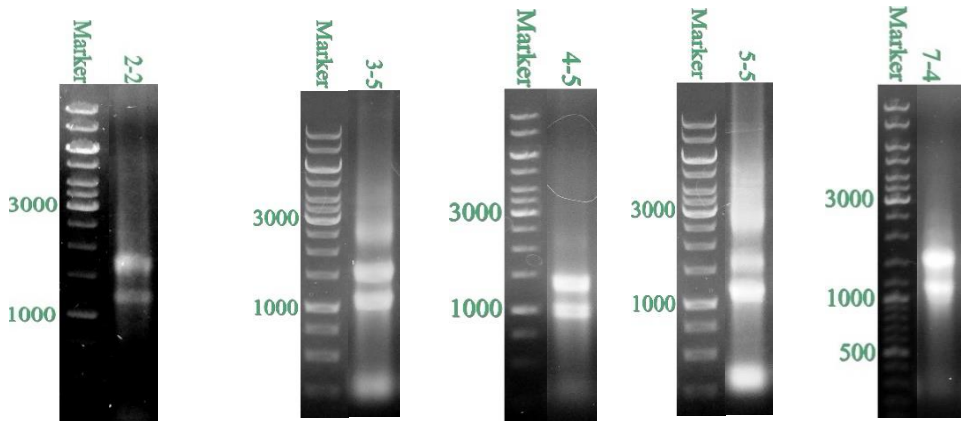
Substances	**Column**	**Wavelength (nm)**	**Mobile phase**	**Flow rate (ml/min)**
Deoxypodophyllotoxin	Hypersil Hypurity™ Elite	280	water:acetonitrile 60:40	1
α-Peltatin	Hypersil Hypurity™ Elite	280	water:acetonitrile 60:40	1
ß-Peltain	Hypersil Hypurity™ Elite	280	water:acetonitrile 60:40	1
Yatein	Hypersil Hypurity™ Elite	280	water:acetonitrile 60:40	1
Secoisolariciresinol	Hypersil Hypurity™ Elite	280	water:acetonitrile 75:25	1.5
Matairesinol	Hypersil Hypurity™ Elite	280	water:acetonitrile 75:25	1.5
Pinoresinol	Hypersil Hypurity™ Elite	280	water:acetonitrile 75:25	1.5
Secoisolariciresinol	Chiralcel OD-H	280	ethanol:n-hexane 50:50	0.6
Pinoresinol	Chiralcel OD-H	280	ethanol:n-hexane 50:50	0.6

VI. Results and discussion

1. Preliminary work

1.1 RNA extraction

The isolation of RNA from cells of a *Linum flavum* suspension culture from days 2, 3, 4, 5, and 7 of the cultivation period was carried out according to the method of Chomczynski and Sacchi (1987) (see III.2.3). After isolation, the quality of total RNA preparations was examined by electrophoresis. On the agarose gel, the 18S and 28S RNA bands were very prominent. In addition, the amount of RNA and its purity were determined photometrically. Many RNA-samples had high quality with the A_{260}/A_{280} ratio in the range of 1.8-2.0. The RNA-samples with the best quality and highest concentration from each batch (2.2; 3.5; 4.5; 5.5; 7.4) were used to synthesise cDNAs (see III.2.3), which were used as templates for amplification of candidate genes by PCR (see III.2.4.1).



Sample	RNA concentration (ng/µl)	A_{260}/A_{280}
day 2 – sample 2	886	2.00
day 3 – sample 5	1241	1.89
day 4 – sample 5	1128	1.76
day 5 – sample 5	800	1.87
day 7 – sample 4	1331	1.73

Figure 17: Agarose gels and photometric results of the best RNA extraction samples from cells of a *Linum flavum* suspension culture of the second, third, fourth, fifth and seventh culture day

1.2 Genomic DNA (gDNA) extraction

The extraction of gDNA from the cells of the seventh-day suspension culture of *Linum flavum* was carried out according to the method of Rogers and Bendich (1985; see IV.2.1). After extraction, the amount of gDNA and purity were determined photometrically. Many gDNA-samples had high-quality with a A_{260}/A_{280} ratio of 1.8-2.0. In addition, the quality of gDNA samples was examined by electrophoresis. The gDNA-sample 2 with the best quality (A_{260}/A_{280} ratio 1.78) and the highest concentration (1139 ng/µl) was used as a template for amplification of candidate genes by PCR (see III.2.4.1).

2. Project 1: Identification and characterisation of a NADPH:cytochrome P450 reductase

2.1 Cytochrome P450 reductase candidates

The transcriptome and the corresponding protein sequences of *Linum flavum* were obtained from the database of the Project 1KP (https://onekp.com/). In 2004, Kuhlmann had identified a partial nucleotide sequence of CPR of *Linum flavum* with a length of 975 nucleotides by using RACE-PCR. This sequence was used as a reference sequence to search for CPR candidates of *Linum flavum* with help of the bioinformatics tool Blastx.

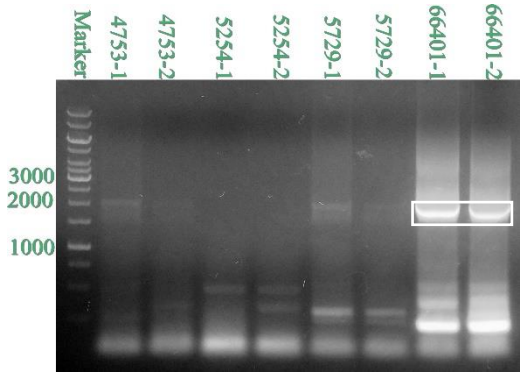
The four candidates with the best results from Blastx are shown below:

Contig	Score	Query cover	E-value	Identity
4753	404	100%	7e-141	70%
5729	587	100%	0.0	94%
66401	610	100%	0.0	98%
5254	401	100%	7e-140	70%

2.2 Amplification of candidates from cDNA and sequencing

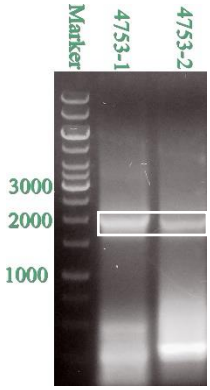
The open reading frames (ORF) of the candidates were found from the transcriptome database by means of the bioinformatic tool ORF Finder. In order to amplify the candidate sequences from cDNA by PCR, the sequences (20-30 nucleotides) located at the beginning and the end of ORFs were used to design primers with the help of the bioinformatic tool Oligo Calc (see IV.6.2).

Using primers and standard PCR conditions (see III.2.4.1) with cDNA as template, two CPR candidates 4753 and 66401 were successfully amplified. The sequences of candidate 4753 and 66401 have a length of appr. 2100 bp. The individually numbered lanes in Fig. 18 and 19 represent PCR products which were formed at the different annealing-temperatures.



Sample	Annealing temperature (°C)
4753-1	54.3
4753-2	56.1
5254-1	54.3
5254-2	56.1
5729-1	56.1
5729-2	57.8
66401-1	54.3
66401-2	56.1

Figure 18: PCR amplification of CPR candidates 4753, 5254, 5729 and 66401. The successful amplification of 66401 is marked with a box.



Sample	Annealing temperature (°C)
4753-1	54.3
4753-2	56.1

Figure 19: PCR amplification of CPR candidate 4753. The successful amplification is marked with a box.

The bands of PCR products with appropriate lengths were excised from the agarose gel and the DNA isolated (see III.2.6). The candidate PCR products were ligated into pDrive vectors (see V.2.7.1) and subsequently introduced into competent *E. coli* EZ cells by heat shock (see III.3.2). The transformed bacterial colonies were selected and grown in overnight cultures at 37° C. The bacterial plasmids were then isolated using a Miniprep kit (see III.3.5) and sent to Seqlab® for verification of the sequences.

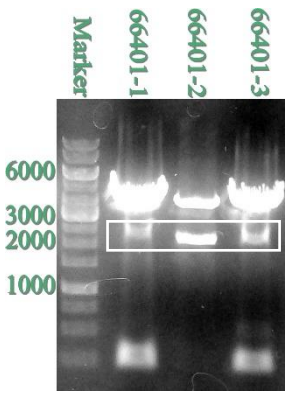


Figure 20: Agarose gel of CPR candidate 66401 in pDrive (colony 1-3)
Plasmids were isolated from *E. coli* EZ and cut by the restriction enzyme EcoRI. Bands of CPR candidate 66401 are marked with a box.

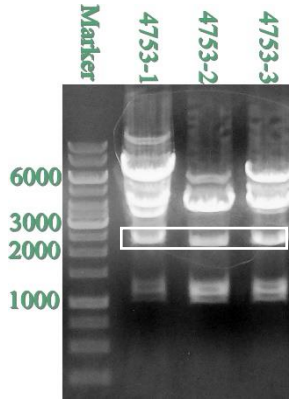


Figure 21: Agarose gel of CPR candidate 4753 in pDrive (colony 1-3)
Plasmids were isolated from *E. coli* EZ and cut by the restriction enzyme EcoRI. The sequence of candidate 4753 contains an internal EcoRI restriction site from 1017 to 1022. Therefore, multiple bands appeared on the agarose gel. The full-length bands for 4753 are marked with boxes.

The ORF of CPR-candidate 66401 comprises 2082 nucleotides and shows high identity (95%) to the sequence of contig 66401 in transcriptome of *L. flavum*. The ORF of CPR-candidate 66401 is shown below:

ATGAGTTCCAGCGGTCTGGAATTGGTGCGGTTCGTGGAGTCGATCCTCGGAGTCTCCCTCAGCGGCGTCTCGTCG
GACCTCGTGGTACTGGTACTCACTACTTCTCTCGCTATTATTGTGGGATTGCTGGTGTTCTTGTGGAAGAAATCG
TCGGATCGGAGTAAAGAGGTGAAGCCGGTGGTGATAACGAAGCCGTATTCGCTCAAGGAGGAAGACGAGGACGAG
TCCGATGCGTTAGCTGGCAAGACTAAGCTCACCATCCTTTACGGTACTCAGACTGGGACTGCTGAAGGTTTCGCT
AAGGCTTTAGCTGAGGAGATCAAGGCAAGATATGAAAAGGCTGCCGTCAAAGTTGTTGACTTGGATGATTATGCT
ATGGATGATGATCAGTATGAAGAAAAATTGAAAAAGGAGTCTTTGACGTTTTTTACGGTGGCTACTTATGGTGAT
GGGGAACCGACGGATAATGCTGCAAGGTTTTATAAGTGGTTCACAGAGGAAAATGAAAGAGGCGTTTGGCTTCCA
AAGCTCAATTTTGGCGTTTTTGGCTTGGGTAATCGTCAATATGAACATTTTAATAAGATTGCAGTTGTCCTTGAT
GAAGAACTTTCTAAGCAAGGTGGTAAGCGTCTCGTGCCAGTTGGTTTAGGTGATGATGATCAATGCATTGAGGAT
GACTTCACTGCATGGAAAGAATTACTCTGGCCAGAGTTGGATCAATTACTCAGAGATGAAGATGATGTGAATACT
CCATCTACACCATATATGGCAGCTATACAAGAATACCGTTTAGTTATTCATGATCCATCTGTAACATCTTATGAG
GACAATTTTGGAAACTTGGCAAATGGCAATACTTCTTTTGACATTCACCATCCGTGCAGAGTCAATGTTGCTGTC
CAAAGAGAGCTCCACTCACCAGAGTCCGACAGATCTTGCTTGCATCTGGAGTTTGATATTTCGGGCACTCCTATT
GCATATGAAACTGGTGATCATATTGGTGTTTATGCTGAGAATTGTGAGGACACTGTTGAAGAAGCAGGGAAGTTA
TTGGATCAACCCTTGGATTTGTTGTTCTCTATTCATTCAGATAAAGAAGATGGGACTCCCCTAGGAGGTTCATTG
GCACCTCCCTTCCCAGGTCCTTGCACTCTTCGTTTGGCACTTTCACGTTATGCTGACCTCTTGAACTCTCCAAAG
AAGTCTGCTTTGGTTGCCTTGGCTGCTCATGCAAGTGAACCAAGTGAGGCAGAGAGGCTTCAATTTTTGTCTTCG
CCGCAGGGGAAGGACGAATATTCACAATGGGTTGTTGCTAGTCAAAGAAGTCTCCTTGAGGTAATGGCCGAGTTC
CCATCTGCAAAACCTCCTCTTGGCGTATTTTTTGCAGCAGTGGCTCCTCGACTGCAGCCTCGTTACTATTCAATC
TCGTCTTCTCCTAGATTTGCTCCCAATAGAGTCCATGTAACCTGCGCTGTAGTAAACGAGCCAACACCAACTGGT
AGAATCCATAAGGGAGTGTGTTCAACCTGGATGAGGCACGCTGTTCCTCTCGAGAAAAGCCGAGATTGTAGCTGG
GCTCCCGTTTTTATCAGGACCTCTAACTTCAAGTTACCAGCCGATCCTTCGACTCCTATTATCATGGTGGGACCG
GGTACTGGATTGGCGCCTTTCAGGGGTTTTCTTCAGGAAAGAATGGCTCTGAAAGAGCAGGGCCGTCAGCTTGGC
ACTGCTTTGCTGTTCTTTGGATGCAGAAACCGACGGATGGATNTTACATACGAGGAGGAGCTAAATAAATTTGCG
GAACAAGGAGCGCTCACTGAACTCATTGTTGCCTTCTCAAGGGAGGGCCCAAAGAAGGAGTATGTTCAACATAAG
ATGACAGATAAAGCTGCAGAAATATGGAGCATAATTTCTCAAGGCGGATACTTTTATGTGTGTGGTGATGCCAAA
GGGATGGCAAGAGATGTTCATCACGCTNTGCACAACATTGTCCAGGAGCAGGGAGGGTTGGATTCGTCGAAGACG
GAGTCTACAGTGAAAAAATTGCAGATGGACGGGCGGTATCTTAGAGATGTATGG**TGA**

The ORF of CPR-candidate 4753 comprises 2151 nucleotides and shows high identity (99%) to the sequence of contig 4753 in transcriptome of *L. flavum*. The ORF of CPR-candidate 4753 is shown below:

ATGGACTCGTCCTCGTCTGCTTCGTCGTCGGGATCTGCGTCGGCGATCGATCTGATGGCGGCAATTATCCGGAAG
AAGTTCGCCGACCCGTCCAACGCGTCGGCGGAGTTGGGAGGGTCGGCGATGGAGGTGACTTCGCTGATTCTGGAC
AACAGGGAGCTTGTTATGATTCTGACGACCTCGATTGCGGTGCTGATCGGGTGCGTGGTGGTGCTGGTGTGGAGA
AGATCGGGATTGCAGAAGCGGAGGAGTTACGAGCCTCCGACGCCTTTGGTTGTGAAGAAGAAGGAGGAGGTTGTG
GTGGACGATGGGAAGAAGAAGGTCACCTTGTTCTTCGGTACTCAGACTGGTACCGCTGAAGGATTCGCCAAGGCT
CTGGCCGAGGAGGCGAAAGCGCGGTATGAGAAAGCTGTATTCAAAGTTGTTGACATTGATGATTATGCCGGGGAT
GATGATGAATACGAGGAGAAGTTAAAGAAAGAGTCCATTGCTTTCTTCTTCCTTGCTACATATGGAGATGGTGAG
CCGACTGATAATGCAGCAAGATTCTACAAGTGGTTTACCGAGGGTAAAGAAGAGAGAGGAGAATGGCTTAAGATG
AAGTATGGAGTGTTCGGCCTTGGCAACAAGCAATATGAACACTTCAATAAGATTGCTAAGGAGGTGGATGATCTT
CTGGCCGAGCAAGGTGGCAAGCGCCTTGTTCCTGTGGGTCTAGGAGATGATGATCAATGCATGGAAGACGACTTC
TCTGCGTGGAAAGAGTTGGTATGGCCAGAGCTCGATGGATTGCTTCTTGACGAGGATGATCAAGCATCTGCTACC
ACCCCTTACACTGCTGCGGTGTTGGAATATCGGGTTGTGTTCTATGACTCCACTGATGCACCAGTAGAAGACAAA
AGCTGGAGTGGCGCCAATGGCCATACTGTCTATGATGCTCAACATCCTTGCCGGTCCAATGTTGCCGTGAGGAAA
GAGCTTCACACTCCTCTTTCTGACCGATCTTGCACTCATTTGGAATTCGACATTGCTGGCACTGGACTCTCATAT
GAAACTGGGGATCATGTTGGTGTGTACTCTGAGAATGTCGAGGAGGTTGTAGAAGAGGCACTGCAGTTGTTAGGT
TTGTCGCCCGACACTTACTTCTCCATCCATACCGATAAAGAGGACGGCACACCACTGAGTGGTAGCTCATTGACG
ACCCCATTCCCACCGTGCACCTTGAGAACAGCTCTGACTCGATATGCCGATCTTTTGAATGCACCCAAAAAGTCT
GCTTTGCTTGCTTTAGCAGCTCATGCTACTGATCCCACCGAAGCCGATAGATTAAGGTATCTTGCATCACCTGCT
GGCAAGGATGAATATGCGAAATGGGTAGTTGCAGAACAGAGAAGTCTCCTCGAGGTCATGGCTGCATTTCCCTCG
GCAAAGCCCCCACTTGGTGTATTCTTTGCCGGAGTTGCTCCTCGCTTGCAGTCTCGATTCTATTCTATCTCCTCA
TCGCCTAAGATGGCACCATCTAGAATCCACGTGACGTGTGCTCTGGTGCTTGAGAAAACTCCGGGAGGACGGCTT
CACAAGGGAGTCTGCTCAACTTGGATGAGGAATGCTGTTCCTATGGAGAAGAGCCAAGACTGCAGTTGGGCACCC
ATTTTCGTTCGGCAATCAAATTTCAGACTTCCTGCAGATACTAAAGTTCCCATCATAATGGTTGGCCCTGGGACC
GGACTGGCTCCTTTCCGCGGCTTCCTGCAGGAAAGACTAGCACTCAAAGAATCTGGAGCTGAACTTGGGCCCTCT
ATCTTCTTCTTCGGTTGCAGAAACCGTAAGACGGATTTCATCTACGAGGACGAACTCAACAACTTTGTCGAAGCT
GGTGCACTCTCTGAGCTGAACGTTGCTTTCTCGAGGGAAGGAGCTACCAAAGAGTACGTGCAACATAGAATGGTG
CAGAAGGCTTCGGAGATCTGGAACCTGATCTCTCAGGGAGCTTATGTATACGTATGTGGCGATGCCAAAGGCATG
GCCAGAGACGTTCACCGAACTCTCCACACCATTGCAATCGAACAGGGCTCGCTCGACTCATCCAAGGCCGAGAGC
TTTGTGAAGAACCTGCAAATGAGTGGTCGGTACCTGCGCGACGTTTGG**TAA**

The internal restriction site GAATTC for the restriction enzyme EcoRI is underlined. The start codon and the stop codon are written in bold letters.

The amino acid sequences of the two candidates were translated from cDNA with ExPASy translate and the molecular weight (MW) of protein was calculated by the ExPASy-Compute pI/Mw tool (see IV.9).

The amino acid sequence of CPR-candidate 66401 with a MW of appr. 77.1 kDa is shown below:

MSSSGLELVRFVESILGVSLSGVSSDLVVLVLTTSLAIIVGLLVFLWKKSSDRSKEVKPVVITKPYSLKEEDEDE
SDALAGKTKLTILYGTQTGTAEGFAKALAEEIKARYEKAAVKVVDLDDYAMDDDQYEEKLKKESLTFFTVATYGD
GEPTDNAARFYKWFTEENERGVWLPKLNFGVFGLGNRQYEHFNKIAVVLDEELSKQGGKRLVPVGLGDDDQCIED
DFTAWKELLWPELDQLLRDEDDVNTPSTPYMAAIQEYRLVIHDPSVTSYEDNFGNLANGNTSFDIHHPCRVNVAV
QRELHSPESDRSCLHLEFDISGTPIAYETGDHIGVYAENCEDTVEEAGKLLDQPLDLLFSIHSDKEDGTPLGGSL
APPFPGPCTLRLALSRYADLLNSPKKSALVALAAHASEPSEAERLQFLSSPQGKDEYSQWVVASQRSLLEVMAEF
PSAKPPLGVFFAAVAPRLQPRYYSISSSPRFAPNRVHVTCAVVNEPTPTGRIHKGVCSTWMRHAVPLEKSRDCSW

APVFIRTSNFKLPADPSTPIIMVGPGTGLAPFRGFLQERMALKEQGRQLGTALLFFGCRNRRMDXTYEEELNKFA
EQGALTELIVAFSREGPKKEYVQHKMTDKAAEIWSIISQGGYFYVCGDAKGMARDVHHAXHNIVQEQGGLDSSKT
ESTVKKLQMDGRYLRDVW

The amino acid sequence of CPR-candidate 4753 with a MW of appr. 78 kDa is shown below:

MDSSSSASSSGSASAIDLMAAIIRKKFADPSNASAELGGSAMEVTSLILDNRELVMILTTSIAVLIGCVVVLVWR
RSGLQKRRSYEPPTPLVVKKKEEVVVDDGKKKVTLFFGTQTGTAEGFAKALAEEAKARYEKAVFKVVDIDDYAGD
DDEYEEKLKKESIAFFFLATYGDGEPTDNAARFYKWFTEGKEERGEWLKMKYGVFGLGNKQYEHFNKIAKEVDDL
LAEQGGKRLVPVGLGDDDQCMEDDFSAWKELVWPELDGLLLDEDDQASATTPYTAAVLEYRVVFYDSTDAPVEDK
SWSGANGHTVYDAQHPCRSNVAVRKELHTPLSDRSCTHLEFDIAGTGLSYETGDHVGVYSENVEEVVEEALQLLG
LSPDTYFSIHTDKEDGTPLSGSSLTTPFPPCTLRTALTRYADLLNAPKKSALLALAAHATDPTEADRLRYLASPA
GKDEYAKWVVAEQRSLLEVMAAFPSAKPPLGVFFAGVAPRLQSRFYSISSSPKMAPSRIHVTCALVLEKTPGGRL
HKGVCSTWMRNAVPMEKSQDCSWAPIFVRQSNFRLPADTKVPIIMVGPGTGLAPFRGFLQERLALKESGAELGPS
IFFFGCRNRKTDFIYEDELNNFVEAGALSELNVAFSREGATKEYVQHRMVQKASEIWNLISQGAYVYVCGDAKGM
ARDVHRTLHTIAIEQGSLDSSKAESFVKNLQMSGRYLRDVW

Based on the PSI- and PHI-BLAST searches against the conserved domain database (CDD) of the National Center for Biotechnology Information, the obtained translated cDNAs were confirmed as cytochrome P450 reductase. Candidate 4753 showed high homology to the CPR genes from several plants, including *Theobroma cacao* (GenBank ID: EOY31887.1; 83% identity), *Azadirachta indica* (GenBank ID: AIG15452.1; 81% identity) and *Gossypium hirsutum* (GenBank ID: NP_001314398.1; 80% identity). Candidate 66401 also showed high homology to the CPR genes from *Theobroma cacao* (79% identity), *Azadirachta indica* (79% identity) and *Gossypium hirsutum* (79% identity).

2.3 Heterologous expression of CPR-candidate proteins

Since CPR is associated with the membrane of the endoplasmic reticulum, *S. cerevisiae* strain INVScI was used for the heterologous expression of CPR using the vector pYes/NTC. The yeast strain INVScI, a common fast-growing *S. cerevisiae* strain, was transformed with the CPR candidate genes by using the lithium acetate method (see III.3.3). The transformed yeast cells are transferred onto SC-U plates and incubated for 3 days at 30 °C. SC-U is a synthetic minimal medium for yeast containing glucose as the sole carbon source without uracil. The yeast cells without pYes2/NTC cannot grow on the SC-U medium. The successful transformations were confirmed by colony PCR (Fig. 22 and 23) (see III.2.4.4).

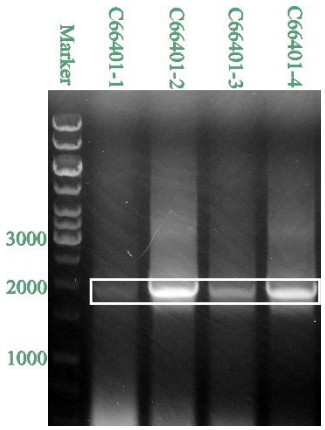


Figure 22: Agarose gel of colony PCR of CPR candidate 66401 to show the transformation of *Saccharomyces cerevisiae* INVScI (colony 1-4)
Candidate genes are marked with a box.

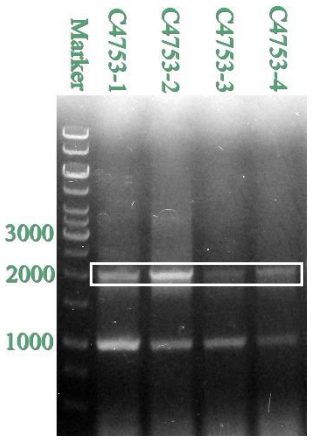


Figure 23: Agarose gel of colony PCR of CPR candidate 4753 to show the transformation of *Saccharomyces cerevisiae* INVScI (colony 1-4)
Candidate genes are marked with a box.

Heterologous expression of the CPR candidate-proteins was performed as described in III.3.7.2. Harvesting yeast cells from expression cultures in SCG after 4, 8, 12, 24, 48, 72, 96 hours aimed at finding out the optimal induction time for CPR proteins. The molecular weight including His-Tag of CPR-candidate 4753 and 66401 is appr. 81.3 kDa and 80.4 kDa, respectively.

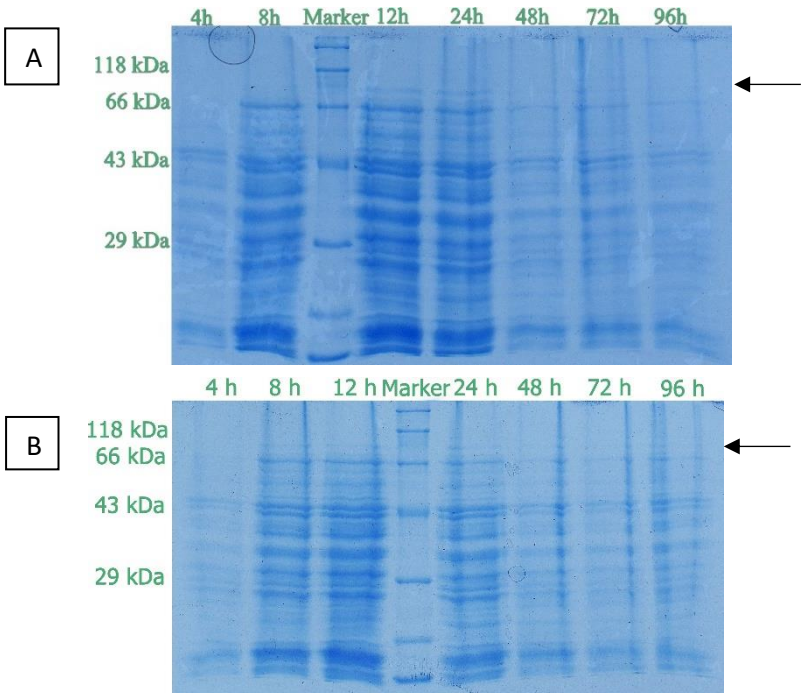


Figure 24: SDS-PAGE of CPR candidate 4753 (A) and 66401 (B) after different time intervals of induction with galactose. The bands of heterologously synthesised CPR-proteins are marked.

The bands of the expected products on the SDS-PAGE gels are rather weak. However, faint bands of appr. 80 kDa appeared on the lanes 24h of both gels and, based on the colour intensity, 12h- and 24h-lanes have the highest concentration of the putative CPR-protein. Therefore, the yeast cells containing candidate genes in pYes2/NTC were induced with galactose and incubated for 24 hour to express the protein for the next experiments.

Since the CPR-protein integrates into the endoplasmic reticulum, microsomes of the transformed yeast cells were used for determining the activity of the candidate proteins. The transformed yeast cells were induced 24 hours with galactose and the microsomes were isolated as described in III.4.1.2. The microsome preparations of the CPR-candidates and negative controls (microsomes from yeast cells containing the empty vector pYes2/NTC) were separated by SDS-PAGE (see III.4.6) and then transferred to PVDF-membranes by Western-blotting (see III.4.7). A poly-histidine tag for purification already present on the vector was bound to the *N*-terminal end of the proteins and the expressed protein was detected with anti-His-Tag antibodies.

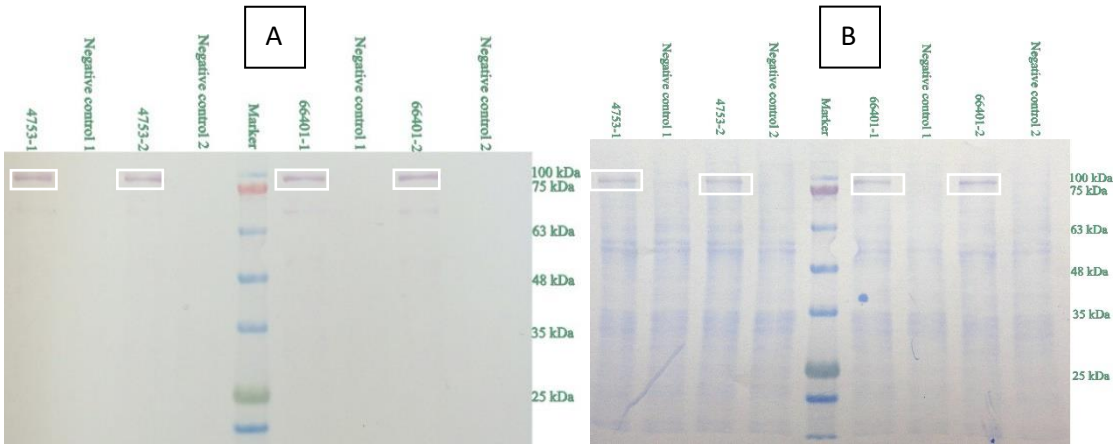


Figure 25: Western Blot of CPR-candidates 4753 and 66401
A: Detection was done with anti-His-Tag antibody and secondary antibodies coupled to alkaline phosphatase using the NBT/BCIP colour reaction.
B: The membrane after NBT/BCIP colour reaction was dyed with Coomassie Brilliant Blue G-250.
Each CPR-candidate was heterologously expressed twice under the same conditions. His-tagged protein bands of CPR-candidates 4753 and 66401 with appropriate molecular weight are marked with boxes.

Gene-specific bands of candidate-proteins in the expected size ranges appeared on the Western-blot membrane (Fig. 25A). There were no bands in the same size ranges in the lanes of negative control samples. In order to check the specificity of the Western-blot, the membranes coloured with NBT/BCIP were dyed with Coomassie Brilliant Blue G-250 by the same colour reaction as for SDS-PAGE gels (see V.4.6) (Fig. 25B). Coomassie staining visualised a multitude of

bands of different sizes. The anti-His-tag antibodies show high affinity for the heterologously expressed proteins and detected successfully and selectively the CPR-proteins. The result of the Western blot showed that the transformed yeast cells produced two different CPR proteins successfully after induction with galactose.

2.4 Functional identification of CPR-candidates 66401 and 4753

In order to determine the activity of the two CPR candidates, an *in vitro* enzyme assay was performed as described in III.4.8.1. CPR activity was determined by measuring its NADPH-dependent cytochrome c reductase activity at room temperature. The recombinant proteins were prepared as microsomes. The time-dependent absorbance change of cytochrome c was monitored at 550 nm prior and after addition of NADPH. A parallel assay with membranes from yeast cells harbouring the empty pYes2/NTC vector was performed as a negative control.

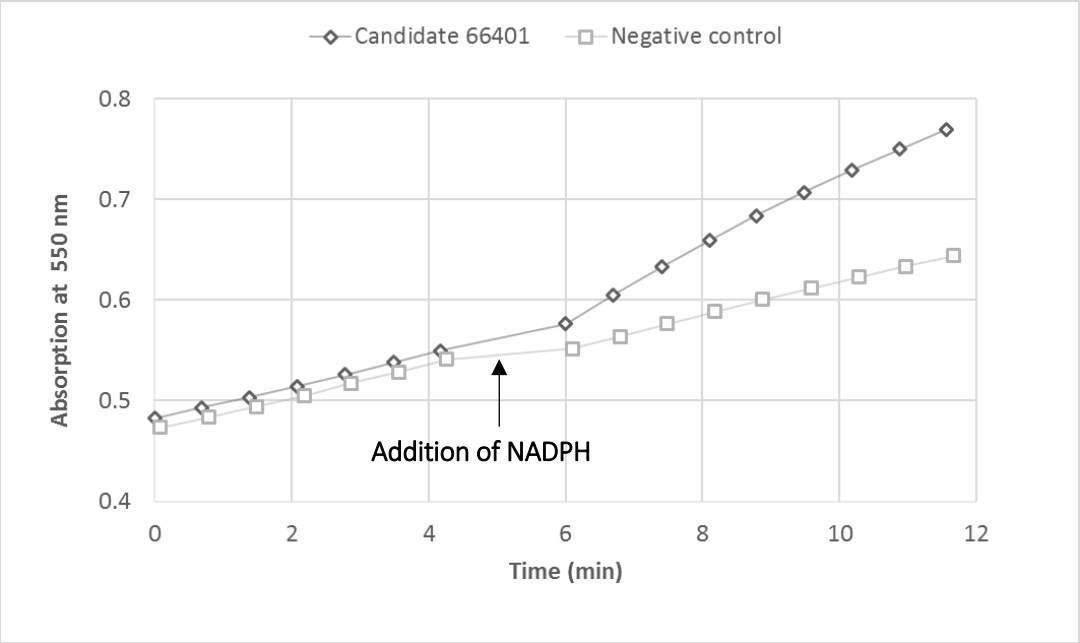


Figure 26: Photometrical enzyme assay of CPR-candidate 66401
The reduction of cytochrome c was initiated by the addition of 0.1 M NADPH at the fifth minute.

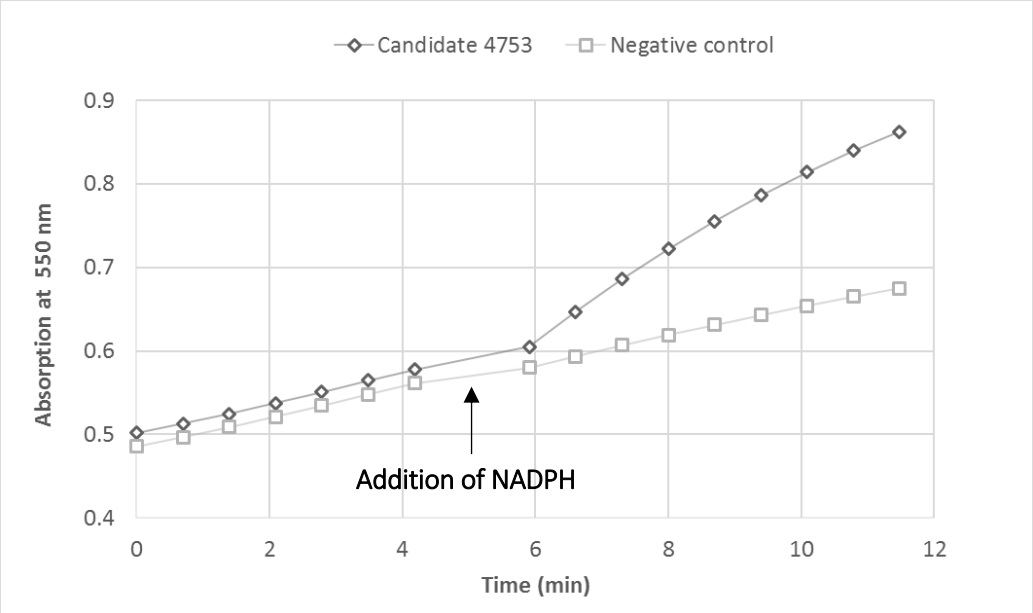


Figure 27: Photometrical enzyme assay of CPR-candidate 4753
The reduction of cytochrome c was initiated by the addition of 0.1 M NADPH at the fifth minute.

Time course measurements revealed a clear reductase activity towards cytochrome c with an explicitly increasing absorption at 550 nm in comparison to the negative control. These results confirm the activity of the CPR-candidates 66401 and 4753 of *L. flavum* in protein preparations from transformed yeast and thus are identified as LfCPR 66401 and LfCPR 4753. The LfCPRs utilized NADPH as electron donor for reducing cytochrome c. LfCPR 66401 and 4753 show a significant increase of specific activity, 35-fold and 33-fold higher than the negative control, respectively (Fig. 28). LfCPR 66401 possessed a slightly higher specific activity than LfCPR 4753.

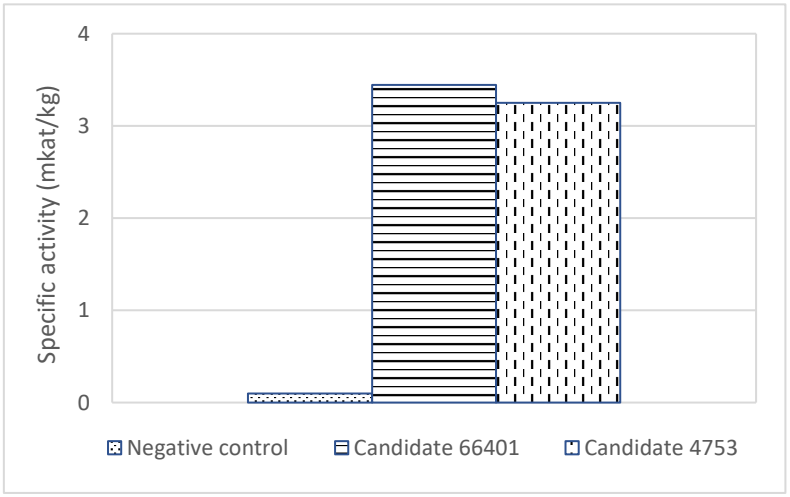


Figure 28: Specific activities of heterolously expressed LfCPR 66401 and 4753
The data represent mean values of four replicate assays (±s.d.).

2.5 Enzyme kinetics of LfCPR 66401 and 4753

2.5.1 K_m-values for cytochrome c

In order to calculate the K_m-values for CPRs, kinetic experiments were performed at a fixed concentration of 0.2 mM NADPH with varying cytochrome c concentrations from 10 to 180 µM. The K_m-value and V_{max} were calculated by linearization of the substrate saturation curves (Fig. 28 and 30) according to Hanes-Woolf (Fig. 30 and Fig. 32). The apparent K_m-value for cytochrome c of LfCPR 66401 was 8.15 ± 0.3 µM while the value for LfCPR 4753 was 15.6 ± 0.35 µM. Based on these K_m-values, LfCPR 66401 has a much higher the affinity towards cytochrome c than LfCPR 4753. Both LfCPRs displayed a V_{max} of 4.9 mkat/kg (Fig. 30 and Fig. 32) under these reaction conditions.

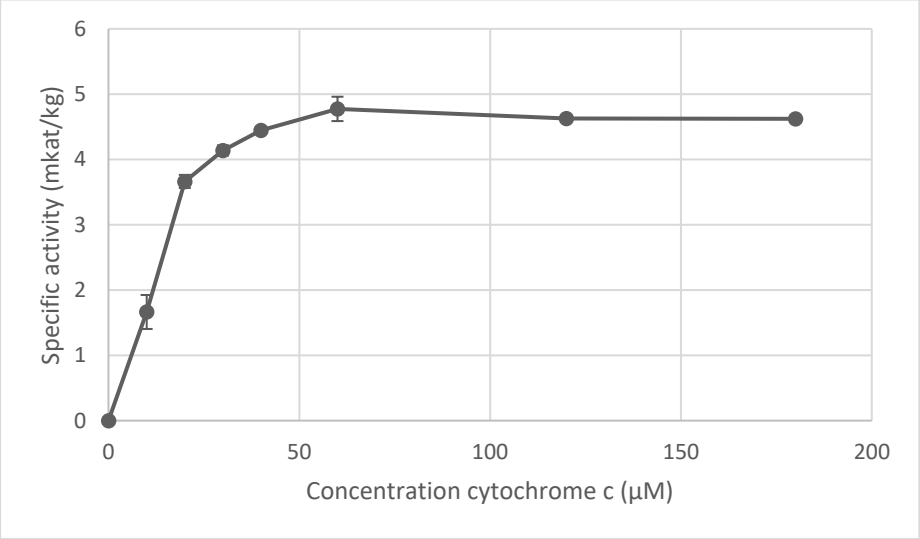


Figure 29: Dependence of the specific activity of LfCPR 66401 on the cytochrome c concentration
The data represent mean values of four replicate assays (±s.d.).

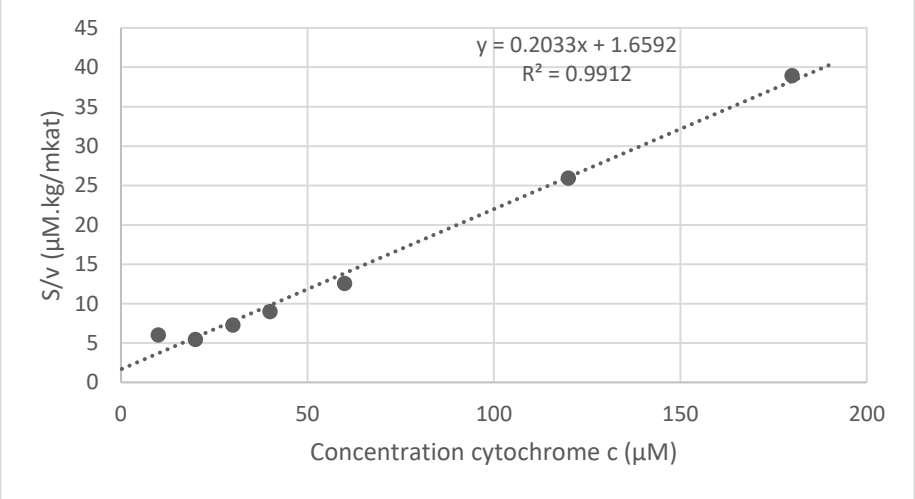


Figure 30: Linearisation of the data from Fig. 29 according to Hanes-Woolf

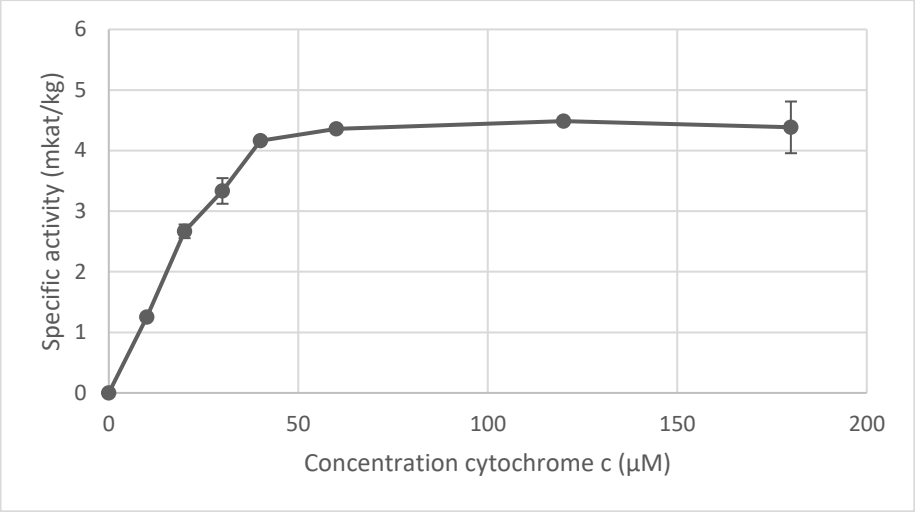


Figure 31: Dependence of the specific activity of LfCPR 4753 on the cytochrome c concentration
The data represent mean values of four replicate assays (±s.d.).

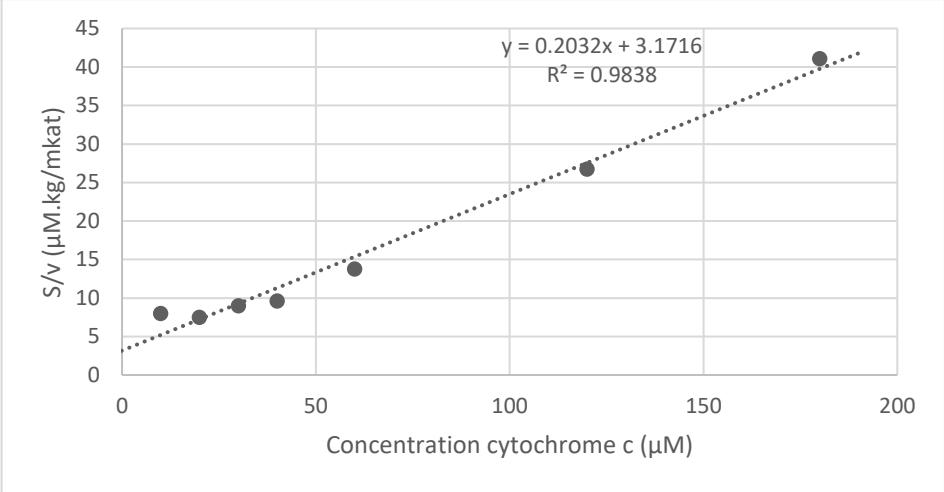


Figure 32: Linearisation of the data from Fig. 31 according to Hanes-Woolf

2.5.2 K_m-values for NADPH

CPRs use NADPH as electron donor to reduce cytochrome c (as artificial electron acceptor). In order to calculate the K_m-values and V_{max} of the LfCPRs, enzyme assays were performed at a fixed concentration of 75 µM cytochrome c with varying NADPH concentrations from 6 to 1200 µM. The K_m-values and V_{max} were calculated by linearization of the substrate saturation curves (Fig. 33 and 35) according to Hanes-Woolf (Fig. 34 and Fig. 36). For LfCPR 66401, this results in a K_m-value for NADPH of 29.6 ± 0.8 µM and V_{max} of 4.38 mkat/kg. For LfCPR 4753, the K_m-value for NADPH is 45.2 ± 0.7 µM and V_{max} is 6.4 mkat/kg. These data show that LfCPR 66401 has a much higher affinity towards NADPH than LfCPR 4753.

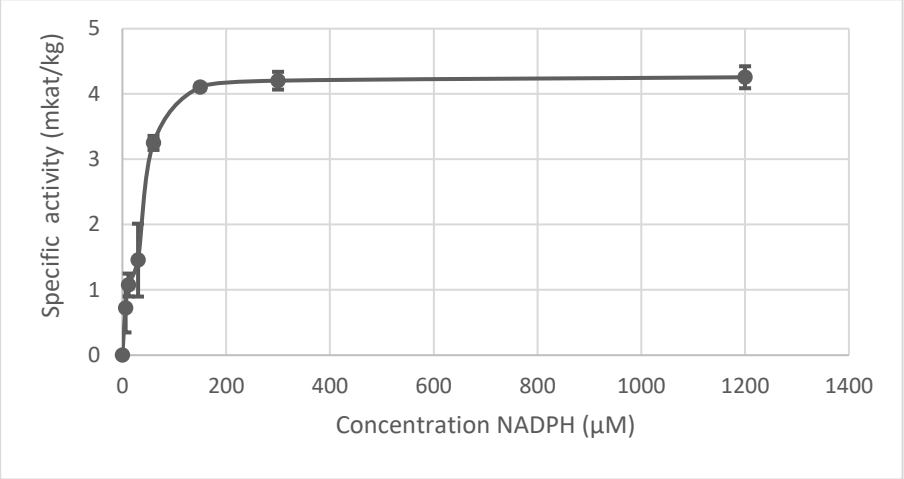


Figure 33: Dependence of the specific activity of LfCPR 66401 on the NADPH concentration
The data represent mean values of four replicate assays (±s.d.).

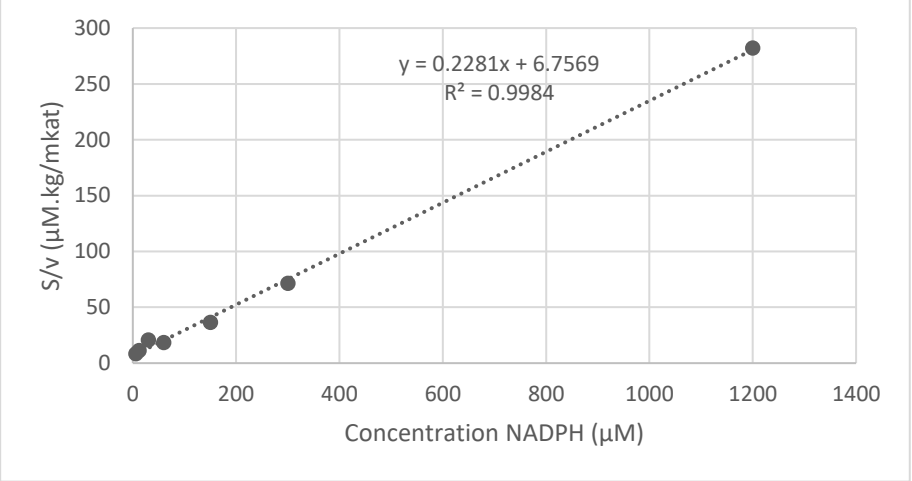


Figure 34: Linearisation of the data from Fig. 33 according to Hanes-Woolf

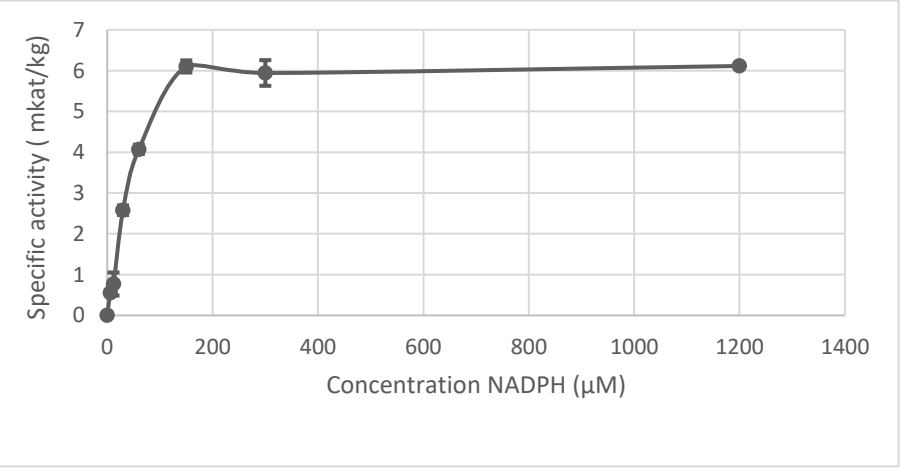


Figure 35: Dependence of the specific activity of LfCPR 4753 on NADPH concentration
The data represent mean values of four replicate assays (±s.d.).

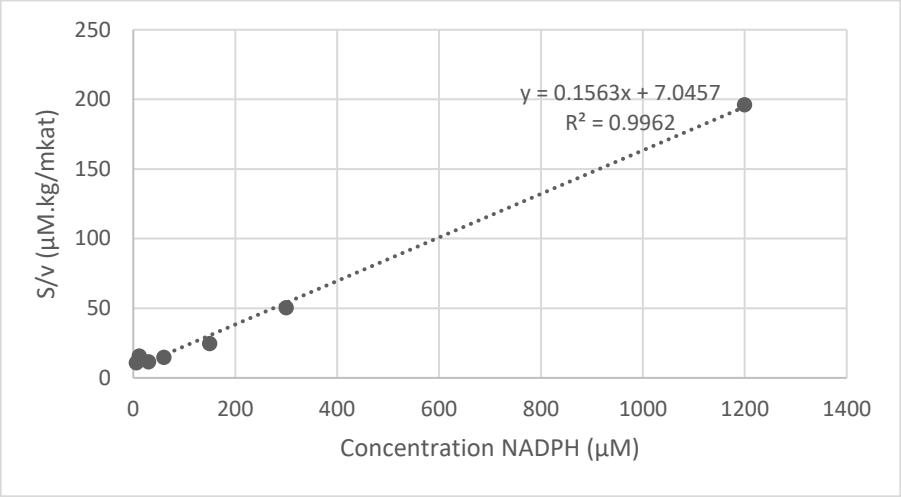


Figure 36: Linearisation of the data from Fig. 35 according to Hanes-Woolf

2.6 Comparison of CPR-sequences from different plants

The alignment of the amino acid sequences of other two CPRs from *Linum flavum* and other plant CPRs including *Theobroma cacao* (GenBank ID: EOY31887.1), *Azadirachta indica* (GenBank ID: AIG15452.1) and *Gossypium hirsutum* (GenBank ID: NP_001314398.1) was performed using Clustal Omega and the result is shown below:

```
66401-ORF       -------------MSSSGLDVVRLV-------ESVLGLSLGG---------ASSDVVVLV
4753-ORF        MD---SSSSASSSGSASAIDLMAAIIRKKFADPSNASAESGGSAMEVTSLILDNREFVMI
A. indica       MQSSSSSASSSNTMKVSPFDLMSAIIKGKT-DPSNV--SSSGSGLEVASIVLENKEFVMI
T. cacao        ----MEPSSSSGSIKVSPLDLMSAIIKGKM-DPSNA--SY-DSAAEVATMLLENREFVMI
G. hirsutum     ----MDSSSSSSSSGPSPLDLMSALVKAKM-DPSNA--SS-DSAAQVTTVLFENREFVMI
                                * :*::  :       *    .  .          ..  .*::

                membrane-anchoring
66401-ORF       LTTSLAVIVGLV-AFFWKRSADRSKEAKPLVITKPYSQKEEEDESEALTGKTKLTIFYGT
4753-ORF        LTTSIAVLIGCVVVLVWRRSGSQKRRNYEP-PTPLVV--KKKEEVVVDDGKKKVTLFFGT
A. indica       LTTSIAVLIGCVVVLIWRRSSSQKPKKIEP--LKPLV--VKEPEVEVDDGKQKITIFFGT
T. cacao        LTTSIAVLIGCVVVLVWRRSGSQKPKQIEP--PKPLI--VKEPELEVDDGKKKVTIFFGT
G. hirsutum     LTTSIAVLIGCVVILIWRRSASQKPKQIQL-PLKPSI--IKEPELEVDDGKKKVTILFGT
                ****:**::* *  :.*:**..:. .             :: *  .  ** *:*:::**

                 FMN binding                P450 binding
66401-ORF       QTGTAEGFAKALAEEIKARYEKAAVKVVDLDDYAMDDDQYEEKLKKETLVCFMVATYGDG
4753-ORF        QTGTAEGFAKALAEEAKARYEKAVFKVVDIDDYAGDDDEYEEKLKKESIAFFFLATYGDG
A. indica       QTGTAEGFAKALADEAKARYDKAIFKVVDLDDYAADDEEYEEKLKKESIAFFFLATYGDG
T. cacao        QTGTAEGFAKALAEEAKARYDKATFKVVDLDDYAVDDDEYEEKMKTESLAFFFLATYGDG
G. hirsutum     QTGTAEGFAKALVEEAKARYEKATFNIVDLDDYAADDEEYEEKMKKDNLAFFFLATYGDG
                ************.:* ****:** .:::**:**** **::****:*.:.:. *::******

                                                      FMN binding
66401-ORF       EPTDNAARFYKWFTEENE-RGVWLPKLNFGVFGLGNRQYEHFNKIAVVLDEELCKQGGKR
4753-ORF        EPTDNAARFYKWFTEGKEERGEWL-KMKYGVFGLGNKQYEHFNKIAKEVDDLLAEQGGKR
A. indica       EPTDNAARFYKWFTEGKE-RGEWLQNLKYGVFGLGNRQYEHFNKIAKVVDDVLAEQGGKR
T. cacao        EPTDNAARFYKWFTEGKE-RGEWLQNMKYGVFGLGNRQYEHFNKVAKVVDDLLTEQGAKR
G. hirsutum     EPTDNAARFYKWFTEGKE-RGEWLQNMKYGIFGLGNKQYEHFNKVAKVVDELLTEQGAKR
                *************** :* ** ** ::::*:*****:*******:*  :*: * :**.**
```

```
                    Cytochrome c binding
66401-ORF       LVPVGLGDDDQCIEDDFTAWKELLWPELDQLLRDEDDVNTPSTPYMAAIQEYRLVIHDPS
4753-ORF        LVPVGLGDDDQCMEDDFSAWKELVWPELDGLLLDEDDQASATTPYTAAVLEYRVVFYDST
A. indica       LVPVGLGDDDQCIEDDFSSWRELVWPELDKLLRDDDDPTTVSTPYTAAISEYRVVFYDSA
T. cacao        LVPVGLGDDDQCIEDDFTAWRELVWPELDQLLRDEDDATTVSTPYTAAVLEYRVLFYDPA
G. hirsutum     IVPLGLGDDDQCIEDDFTAWRELVWPELDQLLRDEDDA-TVSTPYTAAVLEYRVVFYDPA
                :**:********:****::*:**:***** ** *:**  : :*** **: ***::::* :

                                                          FAD binding
66401-ORF       VTSYEDNFGNLANGNTSFDIHHPCRVNVAVQRELHLAESDRSCMHLEFDILGAPIVYETG
4753-ORF        DAPVEDKSWSGANGHTVYDAQHPCRSNVAVRKELHTPLSDRSCTHLEFDIAGTGLSYETG
A. indica       DASIGENNWSNANGHAVYDAQHPCRSNVAVRKELHTPLSDRSCTHLEFEIAGTGLMYETG
T. cacao        DAPVEDENRSNANGHTIYDAQHPCRSNVAVRKELHTSASDRSCTHLEFDIAGTGLSYETG
G. hirsutum     DAPLEDKNWSNANGHATYDAQHPCRSNVAVRKELHAPESDRSCTHLEFDIAGTGLSYETG
                :   ::  . ***:: :* :**** ****::***   ***** ****:* *: : ****

66401-ORF       DHVGVYAENCEDTVEEAGKLLDQPLDLLFSIHSDKEDGTPLGGSLAPPFPGPCTLRLALS
4753-ORF        DHVGVYSENVEEVVEEALQLLGLSPDTYFSIHTDKEDGTPLSGSSLTTPFPPCTLRTALT
A. indica       DHVGVFCENLTETVEEALSLLGLSPDTYFSVHTDKEDGTPLGGSSLPPPFPPCSLRTALA
T. cacao        DHVGVYCENLDEVVEEALTLLGLSPDTYFSVHTDKEDGTPLGGSSLPPPFPPCTLRTALA
G. hirsutum     DHVGVYCENLDEVVDEALSLLGLSPDTYFSVHTDKEDGTPLGGSSLPSSFPPCTLRTALA
                *****:.**  :.*:**  **.   *  **:*:********.**       **:** **:

66401-ORF       LYADLLNSPRKSALVALAAHASEPSEAERLRVLSSPEGKDEYSQWVVASQRSLLEVMAEF
4753-ORF        RYADLLNAPKKSALLALAAHATDPTEADRLRYLASPAGKDEYAKWVVAEQRSLLEVMAAF
A. indica       QYADLLSSPKKSALLALAAHASDPTEADRLRHLASPAGKDEYAQWIVASQRSLLEVMAQF
T. cacao        RYADLLSSPKKSALVALAAHASDPTEADRLRHLASPAGKDEYAQWMVASQRSLLEVMAEF
G. hirsutum     RYADLLSSPKKAALLALAAHASDPTEADRLRHLASPAGKDEYAQWIVANQRSLLEVMAEF
                *****.:*:*:**:******::*:**:*** *:** *****::*:**.********* *

                                        FAD binding
66401-ORF       PSAKPPLGVFFAAVAPRLQPRYYSISSSPRFAPNRVHVTCAVVNEPTPTGRIHKGVCSTW
4753-ORF        PSAKPPLSVFFAGVAPRLQPRFYSISSSPKMAPSRIHVTCALVLEKTPGGRLHKGVCSTW
A. indica       PSAKPPLGVFFAAVAPRLQPRYYSISSSPRVAPSRIHVTCALVYEKTPTGRIHKGVCSTW
T. cacao        PSAKPPLGVFFAAVAPRLQPRYYSISSSPRMALSRIHVTCALVYEKTLTGRIHKGICSTW
G. hirsutum     PSAKPPLGVFFAAVAPRLQPRYYSISSSPRLAPSRIHVTCALVYEKTPTGRIHKGVCSTW
                *******.****.********:*******:.* .*:*****:* * *  **:***:****

                      NADPH binding
66401-ORF       MKHAVPLEKSRDCSWAPVFIRTSNFKLPADPSTPIIMVGPGTGLAPFRGFLQERMALKEQ
4753-ORF        MRNAVPMEKSQDCSWAPIFVRQSNFRLPADTKVPIIMVGPGTGLAPFRGFLQERLALKES
A. indica       MKNCVPMEKSNDCGWAPIFVRQSNFRLPADPKVPVIMIGPGTGLAPFRGFLQERFALKEA
T. cacao        MKNSVPMEKSHDCSWAPIFVRQSNFKLPLDTKVPIIMIGPGTGLAPFRGFLQERLALKEA
G. hirsutum     MKNAVSSGKSDDCGWAPIFVRQSNFKLPSDTKVPIIMIGPGTGLAPFRGFLQERLALKEA
                *::.*   ** **.***:*:* ***:** * ..*:**:****************:****

66401-ORF       GRQLGTALLFFGCRNRRMDFIYEEELNKFVEQGALSELIVAFSREGPQKEYVQHKMADKA
4753-ORF        GAELGPSIFFFGCRNRKTDFIYEDELNNFVEAGALSELNVAFSREGATKEYVQHRMMQKA
A. indica       GAELGPSVLFFGCRNRQMDYIYEDELNNFVQSGALSELVVAFSRQGPTKEYVQHKMMEKA
T. cacao        GAELGPSVLFFGCRNRKMDYIYEDELNNFVNGGALSEIVVAFSREGPTKEYVQHKMMQKA
G. hirsutum     GAELGPSVLFFGCRNRKMDFIYEDELNNFVNSGALSELVVAFSREGPTKEYVQHKMMEKA
                * :** :::*******: *:***:***:**: *****: *****:*  ******:* :**
```

```
                          NADPH binding
66401-ORF       AEIWSIISQGGYFYVCGDAKGMARDVHRTLHNIVQEQGGLDSSKTESMVKKLQMDGRYLR
4753-ORF        SDLWNLISQGAYVYVCGDAKGMARDVHRTLHTIAIEQGSLDSSKAESFVKNLQMSGRYLR
A. indica       SDIWNMISQGGYLYVCGDAKGMARDVHRTLHTIVQEQGSVDSSKAESIVKNLQMTGRYLR
T. cacao        SDIWNMISQGGYLYVCGDAKGMAKDVHRTLHTILQEQGCLDSSKAESMVKNLQMTGRYLR
G. hirsutum     KDIWDMISQGGYLYVCGDAKGMARDVHRALHTIFQEQGSLDSSKAESMVKNLQMSGRYLR
                 ::*.:****.*.**********:****:**.*  *** :****:**:**:*** *****


66401-ORF       DVW
4753-ORF        DVW
A. indica       DVW
T. cacao        DVW
G. hirsutum     DVW
                ***
```

The typical motifs of the P450-, NADPH-, FMN-, FAD-binding and membrane-anchoring domains are shaded in grey and marked (Eberle et al., 2008; Ro et al., 2002). Gaps were inserted to maximise the homology. The alignment profile confirmed that CPR candidates 4753 and 66401 are similar to other CPRs, had two binding sites each for FMN, FAD, and NADPH and one binding site for cytochrome P450. LfCPR 66401 and LfCPR 4753 both contain the conserved acidic amino acid sequence DDDQC(I/M)EDD, which interacts with cytochrome c and P450s, localized near the FMN binding domain. There are membrane anchor regions at the *N*-terminus of each CPR in order to ensure the membrane attachment and the interaction between CPR and P450 monooxygenases. LfCPR 66401 contains a short amino acid sequence and LfCPR 4753 has an extended anchoring sequence at the *N*-terminus. According to Ro et al. (2002), LfCPR 66401 and LfCPR 4753 are classified into class I and II, respectively. CPRs belonging to class I are transcribed constitutively, whereas CPRs of class II are expressed under stress or elicited by wounding or elicitors (Rana et al., 2013).

In plants, multiple CPR isoforms can exist. Three CPRs have been found in *Helianthus tuberosus* and *Populus jackii* (Benveniste et al., 1991; Ro et al., 2002). Two different CPR genes have been characterised in *Withania somnifera* (Rana et al., 2013), *Petroselinum crispum* (Koopmann and Hahlbrock 1997) and *Arabidopsis thaliana* (Mizutani and Ohta 1998; Urban et al., 1997). During biotic and abiotic stress or differential expression at various stages of plant development, the demand of electron supply becomes higher than normal and the presence of multiple CPRs in plants might answer this demand (Ro et al., 2002; Rana et al., 2013). The diversity of P450s that catalyse reactions for the synthesis of primary and secondary metabolites might be also the reason for the occurrence of multiple CPRs in plants (Werck-Reichhart et al., 2000; Ro et al., 2003).

3. Project 2: Identification of genes encoding DOP6H and DOP7H

3.1 DOP6H und DOP7H candidates

In 2001, deoxypdophyllotoxin 6-hydroxylase (DOP6H) was characterised in protein extracts from *Linum flavum* suspension cultures as cytochrome P450 by Molog et al. (2001). In cell cultures of *Linum album*, DOP is converted into PTOX and there are several experiments suggesting that deoxypdophyllotoxin 7-hydroxylase (DOP7H) is also cytochrome P450-dependent (Fig. 37) (Henges, 1999; Kuhlmann, 2004). However, since PTOX formation by DOP7H was not blocked by CYP-inhibitors in suspension cultures of *L. album*, this hydroxylase is suggested by Federolf (2007) not as a cytochrome P450 monooxygenase.

Recently, CYP71BE54 and CYP82D61 in *Podophyllum hexandrum* were successfully identified by Lau and Sattely (2015) (Fig. 38). CYP71BE54 uses DOP as a substrate like DOP6H and DOP7H, however, it demethylates DOP to (–)-4'-desmethyl-deoxypodophyllotoxin. On the other hand, CYP82D61 hydroxylates on C7 of DOP like DOP7H. Therefore, the obtained amino acid sequences of CYP71BE54 and CYP82D61 from *Podophyllum hexandrum* were used as reference sequences to search for DOP6H and DOP7H candidates in the protein database of *Linum flavum* with the help of the bioinformatics tool Blastp.

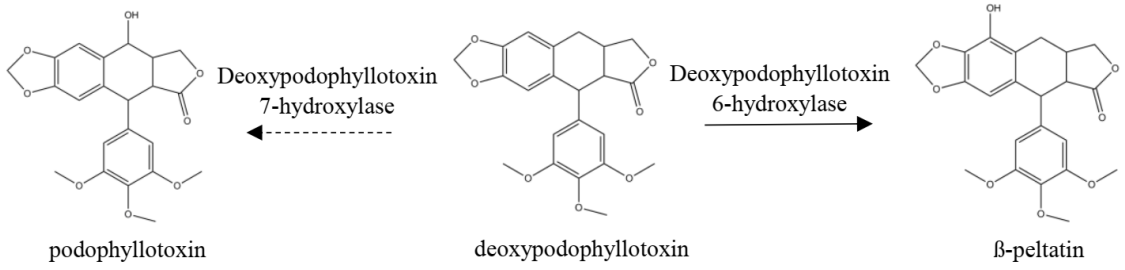


Figure 37: C6- and C7- hydroxylations of DOP
Known reactions are shown by a continuous arrow and unknown with a dashed arrow.

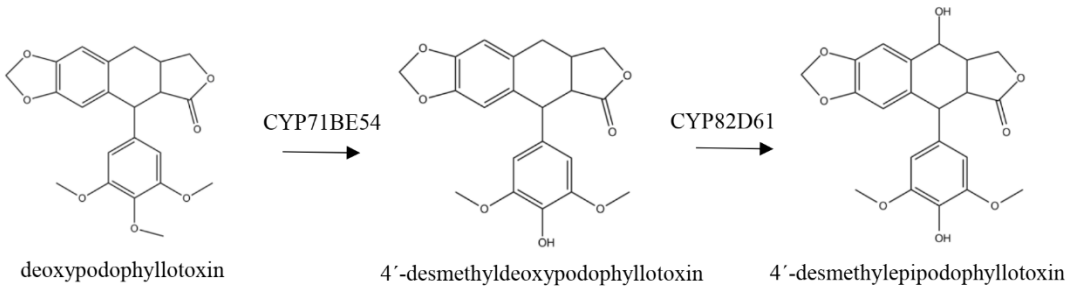


Figure 38: CYP71BE54 and CYP82D61 in the putative biosynthetic pathway to (-)-4'-desmethylepipodophyllotoxin in *Podophyllum hexandrum* (Lau and Sattely, 2015)

The candidates, that have best Blastp-results with CYP82D61, are listed below:

Contig	Score	Query cover	E-value	Identity
27263	450	86%	1e-158	50%
11511	449	92%	5e-157	47%
38991	427	96%	1e-148	45%
11862	449	100%	2e-157	44%
5627	419	98%	4e-145	44%
4471	279	92%	2e-91	35%
2408	251	89%	3e-80	34%
2227	278	90%	6e-91	34%
2114	250	91%	7e-80	33%
3458	234	90%	4e-97	32%
74047	254	89%	2e-81	31%

The candidates, that have best Blastp-results with CYP71BE54, are listed below:

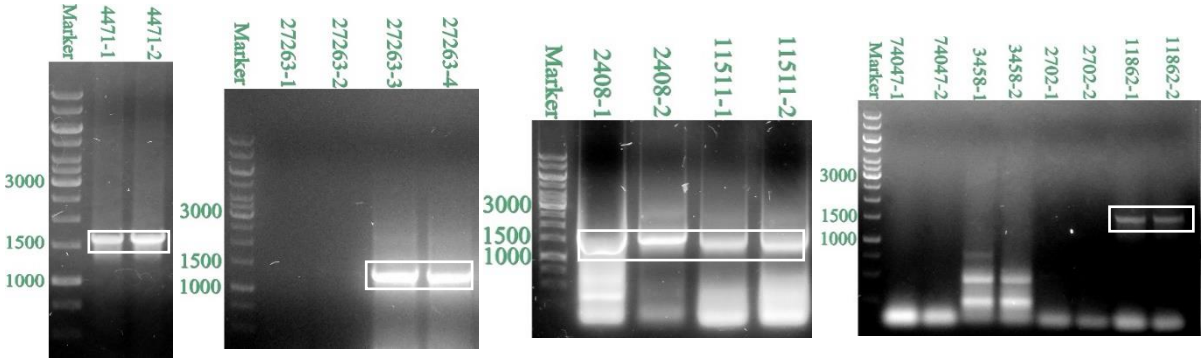
Contig	Score	Query cover	E-value	Identity
2227	550	90%	0.0	56%
2408	506	89%	2e-179	54%
74047	482	98%	2e-170	49%
2114	496	97%	1e-175	48%
3458	452	96%	2e-158	46%
4471	413	91%	7e-144	45%
11511	257	89%	2e-82	33%
11862	264	98%	9e-86	32%
38991	257	95%	1e-82	32%
27263	222	88%	2e-70	31%
5627	217	95%	4e-67	28%

3.2 Generating full-length sequences of CYP candidate genes

3.2.1 Amplification of candidates from cDNA and sequencing

The open reading frame (ORF) of the candidates were identified in the transcriptome database of *Linum flavum* by means of ORF Finder. In order to amplify the candidate sequences from cDNA through standard PCR (see III.2.4.1), the sequences (20-30 nucleotides) located in the beginning and the end of ORFs were used to design primers (see IV.6.1) with the help of the bioinformatic tool Oligo Calc.

As shown in Fig. 39, only five CYP-candidates 11511, 11862, 2408, 27263 and 4471 were successfully amplified in PCR experiments by using full-length primers with cDNA as template. Many attempts to amplify the other six candidates (2114, 2227, 2702, 3458, 38991, 5627) failed. Candidate sequences containing no intron have a length of appr. 1500 bp.



Sample	Annealing temperature (°C)
4471-1	58.9
4471-2	60.1
27263-1	56.1
27263-2	57.5
27263-3	58.9
27263-4	60.1
2408-1	58.1
2408-2	59.8
11511-1	50.3
11511-2	51.7

Sample	Annealing temperature (°C)
74047-1	53
74047-2	54.5
3458-1	53
3458-2	54.5
2702-1	54.5
2702-2	55.9
11862-1	54.5
11862-2	55.9

Figure 39: Agarose gels of PCR-products of candidates 4471, 27263, 2408, 11511 and 11862
The individual numbered lanes represent PCR-products which are formed under the indicated annealing-temperatures. PCR-products are marked with boxes.

The ORF of CYP-candidate 11511 comprises 1611 bp and shows 97.8% identity to the sequence of contig 11511 in transcriptome of *L. flavum*. The ORF of CYP-candidate 11511 is shown below:

ATGGATTTCTTCACTTCTCTCTCCACCGCTTCTCTCTCCACCGTCGCCATAGCCACCGCCTCCCTCGCCTTCCTC
TACTTCCTCTTCCGCACAATACTCTCACCGGCCGGCGACTCCAAACCCCATGCACAACAACAGCAACAAGCACCG
CCGCCGGTGGAGGCCAGCGGCGCGTGGCCCATCATCGGCCACCTCCACCTCCTCGCCGGACCTCAACCGGCGCAC
ATCGTCCTCGGCCAGATGGCCGAGCACCACGGCCCCATCTTCACCATCCGGATGGGAGTCCACCGCGCCGTCGTC
GTCAGCAACCGTGAAACCGCCAAAGAATGCTTAACCACCCACGACCGGGTCTTCGCCGATCGACCCGCCACCCTC
GCCATGGACCTTCTCGGGTACCGCCGCTCCATGCTCGGGTTCAGCCCGTACGGCGCTTACTGGCGGCAGATCCGG
AAGATCGTCACGCTGGAGCTCCTCTCCTCCCACCGCCTCGAAACTCTCAAACACGTCAGGGAATCGGAGATCCAC
GTGGCAACCAAGGAGCTCTACGCGGCGGCCGGAGGGGGAAACGGCGTCGTAGAGATGGGTACTTGGTTCGGGGAG
ATAACGTTGAACGTGATCTTGAAGATGATCGTGGGGAAGAACGTCGGGTACTTTTCGGGCGGAGGAGAGGGAGTG
AAGCTGAGGAAGTTGTTGAAGGATTTTTTCGAGCTGGTCGGGTGGTTTCTGGTGACCGACGGACTGCCGTTTCTG

CGGTGGCTGGACGTCGGAGGGTCCGAGAAGGCGATGAGGAAGACGGCGAAGGAGTTGGACGTTGTGGTCGGAGAG
TGGATGAGGGAGCACGAAGGGAAGCGAGGTTCCGCCGGCGGCGGGAAGAAGGCGGCGGCGGAGGAGGATTTTATG
GATGTGATTATGAACGTCGTCGGCGGTGAGGAAGGCGGCGGCGTCGTGGGTCGGGATTCTGAAACGGTCAACAAA
GCTACTTGTTTGGCTCTCATCTTAGCCTCTTCAGACACGACCACCGTGACAATGACGTGGCTCATGGCTCTCCTC
GTCAACCACCCCGACGTCCTAAAAAAAGCTCAGACCGAGCTAGACGACGTCGTCGGCAAACTCCGGCGAGTTCAA
GACTCCGATCTCCATAACTTACACTACCTCAAAGCCATCATCAAAGAAACCCTCCGGCTCTACCCGGCCGCACCT
CTCTCCATCCCCCACAAGTCCACGGAGGACTGCACCGTCGCCGGCCGGTTCGTCCCCAAAGGCACCCAACTCATC
GTCAACATATCGAAGATCCAGCGCGACCCGGGTGCCTGGTCGGACCCGGACGAGTTTCGACCCGATCGGTTCGTT
ACCACGCACAAGGACGTCGACTTCCGGGGTCAATATTTCGAGCTCATACCGTTCGGTAGCGGAAGGAGGATGTGT
CCCGGAGTGACTTTCGCGATGCAGGTCATGCAGTTGACGGTGGCGACGTGGTTGCATGGGTTCGATTTCAAGAGG
ACGAGTGACGAGATGGTCGATATGACGGAGGGCGTAGGGCTGACGAATCCGAGAGCGGAGCCTCTCGAAGTTGTG
TTGTCTCCTCGCCTGCCACCGCATCTGTATGAA**TGA**

The ORF of CYP-candidate 11862 comprises 1545 bp and shows 98.2% identity to the sequence of contig 11862 in transcriptome of *L. flavum*. The ORF of CYP-candidate 11862 is shown below:

ATGGATTCTCTCTTTGCTTCAATTGCTCTGGCCGTCTCCTTCCTCTGTCTCGCCGGCTTCTTCCTCACCGACGGC
CTCCGGCGCCTCGGCTCCGACAAGCAAAAGCGATATCCACCGGAACCAGCCAACAGATGGCCAGTCATCGGCCAT
CTCCCGGTAATGGCCGGATCGGAGCTCCCTCACCGAGCACTCGCCGCCTTGGCCGACAAATACGGCCCCATCTTC
ACCCTCCGCCTCGGCTTCAACCGCGTAGTGATCGTCAGCAGCCCGGAGATCGCCAAGGAGCTCTTCACGACGAAA
GACACCGTCGCGCTCTCCCGCCCCAGAATGATCGCCACCGAGACCTTCTCCTACGACTTCGCCGTCTTCGGGATG
GGTCCGCACGGAGACTACTGGCGGGAGACGAGGAAGATCGCCGTCCTCGAGCTTCTCTCCAACCGCCGGCTGGAT
CTCCTCCGGCGGTTCAAGGAATCCGAAGTCCATAACTCGATCAACGACCTCTACAAGCAAATTGGGGGTAAATCG
GAGATCGACGTGAACAAATGGATCGGCGATTTGAGCATGAACCTCATGCTGAAGCTAGTTATAGGGAAGAGCTCC
TCCTCTGCTTCCGGCGGCTCTGCTATGATGGTCAAGTTCCAGATTGCGGTGAGGGAGCTTTTCCACTTAGCCGGC
CTGTCTCTGGTCGGCGACGCGCTTCCTTTTCTCCGGTGGATGGATATCGGCGGGCACGAGAAGGCGATGAAAAGA
GCTGCGAAAGCGATGGATGGGTATCTTCAGGAGTGGCTGAATGAGCACAAGGAAAACAGGGGAGGGGAGAAAACA
GAGCATGATTTCATCGACGTTATGCTGTCGCATCTCGACGGCCAGTCTCTCCACGGCTACGATTCCGATACCATC
ATCAAAGCAATGTGTCTGGCCATGGTTGCAGGAGGAATGGATACGACCAACGCGACAGTCATGTGGGCGATTGCT
CTTATAATGAACAACCGTGACGTGCGGCTCAAGGCACAGAACGAGCTCGACGACGTCGCCTCGAGGGAAAGGCTC
GTGACGGAGGCGGACATCGGAAAGTTACCGTACCTCCAAGCCGTCGTCAAAGAGTCGATGAGGATGTACCCGGCG
ACGCCGCTCCTGGCGGCGCGTGAATTCGTTACAGACCGTAAGGTGGGCGGGTACGACGTCGAGAAGGGTACGTGG
CTGATGTTCAACGCATGGAGGGTTCAGAACGACCCGAACGTGTGGCCCGACCCGACGAAATTCGACCCGGAGAGG
TTCCTGACGTCGGAGTTTCGAGACGTCGACGTGAAGGGGCAGAGCTTCGAGCTGTTCCCGTTCTCGAGCGGGAGG
AGGTCTTGTCCCGGGATGAATCTGGGACTGCAGCTGGTGCATTTGGCGGTGGCGAACTTTCTGCACGCGTTCGAG
GTTTCGCCGGCGGGGGATGCGCCGGTGGATATGAAGGAGAGCTTTGGGATGACGAATATGAAAGCGACGCCGCTT
TCGGTCGTGGCTTCGCCGCGATTGTCTCACGATGCTTATGTT**TAA**

The ORF of CYP-candidate 2408 comprises 1605 bp and shows 99.9% identity to the sequence of contig 2408 in transcriptome of *L. flavum*. The ORF of CYP-candidate 2408 is shown below:

ATGGCCGCCTCGCTCACCTTCTCCCCCCTGAAATCAACATTCCCCGGCGGCAACAACCCCCCGCCGCCGTCACTC
CGCCTCCACCGCATCCCAACCAGAATCTCCTGTCAAAACCAGAAACCCCAATCCGAATCCACCACTAAAAAGCAG
CTCCCTCCGGGACCGCCAAAACTCCCCCTCATCGGAAACCTCCACAACCTCGTCGGCGCGCTACCACACCAAGCC
CTCCGAAACCTCGCCACCGAATACGGCCCTTTGATCCACCTTCAGCTAGGGGAAATCTCCGCCGCCGTCGTCTCC
AACCCGGCGATCGCCCAAGAGATAATGAAAACCCACGACCTGAACTTCGCCGACCGACCCTGCATCTTCGCCGCC
GAGATCGCCACGTGGGGGGGCCAGGACATCGCCTTCTCCCCGCACGGCGAGTACTGGAAGCAAATGAAGCGGATC
TCGCTGACGGAGCTTCTGGGTCCCCGGAAAACCCAGTCTTTCCGCGGGATCAGAGAAACAGAGGTCGCCGGAATG
ATCGAGTCGGTGAGGAAGTCGGTCGGGAAACCGGTCAATGTAACCGAGAAGGTTCTCAGCCTGACGAATACGATC
ACTTGTAAGACCGCGTTTGGTTACCAGTGTATGGATCAGGAGGAGTTCGTCGGGCTGATGAACGGGGCAGTTCAG

GCGGCGGGGGGGTTTAACATCGCCGATTTGTACCCGTCGTTGGGGTTTCTTCAGGGGCTTACGGGCATGAAATCG
GAGCTTCTGAGGATCCGAAACGGGTTGGATCGGATCTTCGATAAGATCATTAAGCAGCACGAGGAGAAGAGAGCT
AACGGCGATGTGGAATTGGACGACGAAGATCTGATGGATGTTCTTCTCCGGCTTCAGGGGAGTGGCGGCTTCAAG
TGTCCCATTACCAGCACCAACATCAAAGCCGTTCTTGTGGATCTTTTTATAGCTGGAACCGACACGTCATCGACG
ACCGTAGAATGGGCGATGTCGGAAATGATGAAGAACCCACGAGTGATGCAGAAGGCTCAGGCCGAGGTACGAGAA
GCCATGAGAGGCAAGTCAGTAGTCACCGAAGCGGATATTCAAAACCTACCGTACTTGAGCTCGATAATCAAAGAA
ACTTTCAGACTACATCCACCAGCCCCGCTGCTCCTTCCGAGGGAAAGCAAAGCGAACTGCGAGGTTGCAGGGTAC
GAGATTCCGAAGAAGACGAAAGTGATTGTGAACGCCTGGGCGATCGGGAGGGATCCAGGAACTTGGAAGGATCCC
GAGACTTTCATTCCGGAGAGGTTCGACGGAAGCGAGATTGACTTCAAAGGGATGCATTTCGAGCTGATCCCGTTC
GGTGCCGGAAGGAGGATATGTCCCGGGATCGCATTCGGGATGGCCAACGTCGAGCTCCCGATCGCTCAGTTGCTC
TACTATTTCAACTGGGAGCTCCCCGAAGGGGTCGCCGTCGAGGATTTCGACATGGAAGAGTCGTTTGCAGCCACC
ATGGGACGGAAGAACCCATTGTTTCTGGTTCCCAGAGAGTTCAATGGTCAACCGGAGGCTGTGAATGGAAGTTCT
TCTGAAATGTTAGAGGCTGTTGCAAAT**TAG**

The ORF of CYP-candidate 27263 comprises 1386 bp and shows 100% identity to the sequence of contig 27263 in transcriptome of *L. flavum*. The ORF of CYP-candidate 27263 is shown below:

ATGGCCGACAAGTACGGCTCCATATTCACCGTCCGAATGGGGAGGCAACGAGCCCTCGTCGTCAGCTCCTGGCCG
ATCGCCAAAGAGTGCTTGGTCACCGTCGACAAGCCGTTCCTCAACCGGCCGAAGAATCTCGCCGCCAAGATCATG
ACCTACGACATAGCCATGTTCGGGTTCAGCCCTTACGGCCACTACTGGCGGGAGATGCGGAAGCTCGTCTCCCTC
GAGCTCCTCTCGGCTCAACGGCTCCGGACGTTCGCTCCCGTCCGGGAGTCGGAGGTCCGGGCAGCCACGAAAGCG
TTGTACGATCTCTGCGACGGCTCCTCCGTCGCGGTGGAGATGAAGCGGTGGTTCGCCGACGTGACGTTGAATGTG
ATAATGAGGATCATCGCCGGGAAGACGATCGGGTATGGTTCGGTCGTCGCCAGAGACGGTGGCGGCGGGGCGATG
GTCGGGTGGCAGAAGTCGTTGAGGGATTTCTTCGACTTGTTGGGGAAGTTCACGGTGGCGGATGCGGTGCCGGCG
TTGAGGTGGTTGGACGTCGGAGGGTACGAGAAGGAGATGAGGAAGACGGCGAAGGAGCTGGATACGGTGGTCGGA
GGATGGTTGGAGGAGCATAAGAGGAAGAGGGTGGTGGCTGCCGCCGGCAGCGGGAAGGTGGAGGAGGATTTCATG
GACGTGATTCTTAACGTTATTGGTGATTCCGGTGGGATCGACGGCCGTGATTCCGACGTCATCAATAAAGCTACT
TGCTTGTCTCTAACCCTCGGCGGCTCCGACACGACGACGGTGACGATGTCGTGGGCGCTCTCCTTACTAGTAAAC
CACCGGAGCGTCTTGACGACGGTTCAAGAAGAGCTCGACCGCTGCGTCGGCCGGGAGAGACAAGTCCAGGACTCC
GACCTCCACAACCTGACCTACCTGAAAGCCGTCGTGAAAGAGACGCTCCGCCTCTACCCGGCGGCGCCGGTGATA
ATGCGCGACGATTCCGACGACGACGCCGTCATCGACGGCCGCCACGTGGGCAAGGGGACGAGGCTCCTCGTCAAC
CTCTTCAAGCTCCAGCGCGACCCGGCGGTGTGGTCCCCGGATCCGGACGAGTTCCGGCCGGAGAGGTTTCTGGAA
GGGCGGCACAAGGACGCCGACGTGAAGGGGCAGAGCTTCGAACTGATACCGTTCAGCAGCGGCCGGAGGATGTGC
CCCGGCGCGTCGTTCGCTCTCCAGGTTATGTACTTGACTCTGGCGACGGTCCTACAAGGGTTTGAGGTCACGACG
GCGGACGGCGGGCCGGTGGATATGACGGACGGCGCCGGGTTCACTAATCTGAGAGCTACTCCGTTGGAGGTTTTA
CTGTCTCCCCGGCTTCCAGCTCATTTGTACGGC**TAA**

The ORF of CYP-candidate 4471 comprises 1515 bp and shows 100% identity to the sequence of contig 4471 in transcriptome of *L. flavum*. The ORF of CYP-candidate 4471 is shown below:

ATGCCTTCACTACTTATCTACCTTCTCCTAGCTCTCCCTCCCGTCCTACTATTATTCATTTTCCAAAAGCATAAG
AAGAAGAATTCAAAGAGTGCTTCGTCTCTCCCACCTGGCCCTCGGGGACTTCCCTTGATAGGGAACTTGTTCCAG
CTTGATCCCTCAGCCCCTCATCGCTATCTATGGCAACTTTCTCAAACATACGGACCCTTATTATCTCTGAGGCTG
GGCCGTGTTCAGGCTGTCGTAGTTTCCACCGCCAAAATGGCTCAGGAAGTCATGAAAACTCAAGATCTTGTCTTC
TGCAACAGGCCACAAGCAATGGGCACAACAAAGTTGTCGTACAACCGCGTGGACTTGGCATTTGCACCCTATGAT
GCTTACTGGAGAGAGATGAGAAAAATATGTATGGTATATCTCTTTAACTCCAATAGAACTCAAAGCTTTGGTCCC
ATCAGAGAATACGAGGTCTCCCAAATGGTTGGGAAGGCTTCGAGATCTTCATTATTAGCTTCTCAGCCTTTTAAC
TTGTCAGAGGCCATGATGTCTCTGACAACTACAATTATATGTCGTGTGGCTTTCGGCAAGAGGTACGAGGAGGAG
GGAACAGAGAGAAGCAGATTCCAGTCTATGTTGAATGAAACTCAAGCCATGTTTACGAGTTTCTTTTTCTTGGAT
CATTTCCCCATTTTAGGTTTCATCGATAGGCTTACTGGGTTGACCAATCGTCTCGAGAAGAATTTCATTGAGTTT
GATGCTTTTTACCAAGAGATAATTGATGAGCATCTTGACCCAAAAAGAGTAAAGCCCGAGCAGGAAGACATTCTC
GACGTCTTGCTTCAAATATGGAAGGATCGGTCTTTCAAAGTCCAACTCTCATTCGACCACATCAAAGCCCTTCTT

ATGAACATATTTGTTGGAGGATCGGACACAAGTGCAGCTACCGTTGTTTGGTCGATGACATATCTAATGAAGAAC
CCTCTGATGATGGGAAAAGTTCAGAAAGAAGTTAGGGATGTAGTTGGGAATAAAGGTTTCGTAAATGAAACAGAC
ATACAACAGTTACCTTACTTGAAAGCTGTGATTAAAGAGACGATGAGATTGCAACCTACGGCTCCGTTACTGCTC
CCAAGGGAATCAATTGAGAAATGCAACTTAGGTGGGTACGAAATACCAGCCAAAACTGTAGTTCATGTGAATGCA
TGGGCAATCGGAAGGGATCCTGAAGCCTGGGGCGAGAACCCAGAAGAGTTCAAGCCAGAAAGGTTCATAGGAAAG
TGTATCGATGTGAAAGGCATGGACTATGAGCTAATACCATTTGGAGCAGGAAGAAGAATATGTCCTGGGATTCAC
ATGGGGCTTGTAACGGTGGAACTTTCACTAGCTAATTTGCTTTATGCCTTCGACTGGAGGATGCCAGCCGGAGTA
AAAAGCGAAGATTTGGACATGGATGTTCTACCAGGTCTCACGATGCACAAGAAAAATGCACTATGTCTAGTAGCT
ACAAAGTTTTTA**TAA**

The start codon and the stop codon are written in bold letters. The amino acid sequences of the candidates were translated from cDNA with ExPASy translate.

The amino acid sequence of CYP-candidate 11511 is as follows:

MDFFTSLSTASLSTVAIATASLAFLYFLFRTILSPAGDSKPHAQQQQQAPPPVEASGAWPIIGHLHLLAGPQPAH
IVLGQMAEHHGPIFTIRMGVHRAVVVSNRETAKECLTTHDRVFADRPATLAMDLLGYRRSMLGFSPYGAYWRQIR
KIVTLELLSSHRLETLKHVRESEIHVATKELYAAAGGGNGVVEMGTWFGEITLNVILKMIVGKNVGYFSGGGEGV
KLRKLLKDFFELVGWFLVTDGLPFLRWLDVGGSEKAMRKTAKELDVVVGEWMREHEGKRGSAGGGKKAAAEEDFM
DVIMNVVGGEEGGGVVGRDSETVNKATCLALILASSDTTTVTMTWLMALLVNHPDVLKKAQTELDDVVGKLRRVQ
DSDLHNLHYLKAIIKETLRLYPAAPLSIPHKSTEDCTVAGRFVPKGTQLIVNISKIQRDPGAWSDPDEFRPDRFV
TTHKDVDFRGQYFELIPFGSGRRMCPGVTFAMQVMQLTVATWLHGFDFKRTSDEMVDMTEGVGLTNPRAEPLEVV
LSPRLPPHLYE

The amino acid sequence of CYP-candidate 11862 is as follows:

MDSLFASIALAVSFLCLAGFFLTDGLRRLGSDKQKRYPPEPANRWPVIGHLPVMAGSELPHRALAALADKYGPIF
TLRLGFNRVVIVSSPEIAKELFTTKDTVALSRPRMIATETFSYDFAVFGMGPHGDYWRETRKIAVLELLSNRRLD
LLRRFKESEVHNSINDLYKQIGGKSEIDVNKWIGDLSMNLMLKLVIGKSSSSASGGSAMMVKFQIAVRELFHLAG
LSLVGDALPFLRWMDIGGHEKAMKRAAKAMDGYLQEWLNEHKENRGGEKTEHDFIDVMLSHLDGQSLHGYDSDTI
IKAMCLAMVAGGMDTTNATVMWAIALIMNNRDVRLKAQNELDDVASRERLVTEADIGKLPYLQAVVKESMRMYPA
TPLLAAREFVTDRKVGGYDVEKGTWLMFNAWRVQNDPNVWPDPTKFDPERFLTSEFRDVDVKGQSFELFPFSSGR
RSCPGMNLGLQLVHLAVANFLHAFEVSPAGDAPVDMKESFGMTNMKATPLSVVASPRLSHDAYV

The amino acid sequence of CYP-candidate 2408 is as follows:

MAASLTFSPLKSTFPGGNNPPPPSLRLHRIPTRISCQNQKPQSESTTKKQLPPGPPKLPLIGNLHNLVGALPHQA
LRNLATEYGPLIHLQLGEISAAVVSNPAIAQEIMKTHDLNFADRPCIFAAEIATWGGQDIAFSPHGEYWKQMKRI
SLTELLGPRKTQSFRGIRETEVAGMIESVRKSVGKPVNVTEKVLSLTNTITCKTAFGYQCMDQEEFVGLMNGAVQ
AAGGFNIADLYPSLGFLQGLTGMKSELLRIRNGLDRIFDKIIKQHEEKRANGDVELDDEDLMDVLLRLQGSGGFK
CPITSTNIKAVLVDLFIAGTDTSSTTVEWAMSEMMKNPRVMQKAQAEVREAMRGKSVVTEADIQNLPYLSSIIKE
TFRLHPPAPLLLPRESKANCEVAGYEIPKKTKVIVNAWAIGRDPGTWKDPETFIPERFDGSEIDFKGMHFELIPF
GAGRRICPGIAFGMANVELPIAQLLYYFNWELPEGVAVEDFDMEESFAATMGRKNPLFLVPREFNGQPEAVNGSS
SEMLEAVAN

The amino acid sequence of CYP-candidate 27263 is as follows:

MADKYGSIFTVRMGRQRALVVSSWPIAKECLVTVDKPFLNRPKNLAAKIMTYDIAMFGFSPYGHYWREMRKLVSL
ELLSAQRLRTFAPVRESEVRAATKALYDLCDGSSVAVEMKRWFADVTLNVIMRIIAGKTIGYGSVVARDGGGGAM
VGWQKSLRDFFDLLGKFTVADAVPALRWLDVGGYEKEMRKTAKELDTVVGGWLEEHKRKRVVAAAGSGKVEEDFM
DVILNVIGDSGGIDGRDSDVINKATCLSLTLGGSDTTTVTMSWALSLLVNHRSVLTTVQEELDRCVGRERQVQDS
DLHNLTYLKAVVKETLRLYPAAPVIMRDDSDDDAVIDGRHVGKGTRLLVNLFKLQRDPAVWSPDPDEFRPERFLE
GRHKDADVKGQSFELIPFSSGRRMCPGASFALQVMYLTLATVLQGFEVTTADGGPVDMTDGAGFTNLRATPLEVL
LSPRLPAHLYG

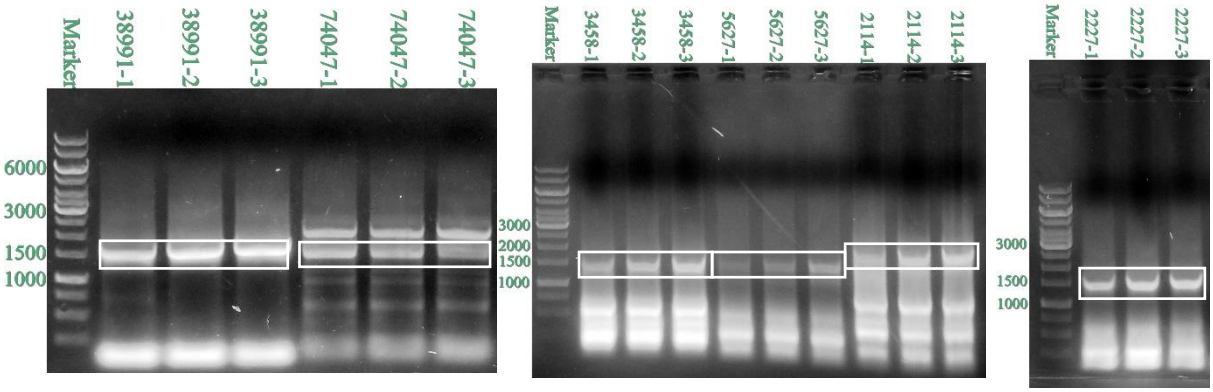
The amino acid sequence of CYP-candidate 4471 is as follows:

```
MPSLLIYLLLALPPVLLLFIFQKHKKKNSKSASSLPPGPRGLPLIGNLFQLDPSAPHRYLWQLSQTYGPLLSLRL
GRVQAVVVSTAKMAQEVMKTQDLVFCNRPQAMGTTKLSYNRVDLAFAPYDAYWREMRKICMVYLFNSNRTQSFGP
IREYEVSQMVGKASRSSLLASQPFNLSEAMMSLTTTIICRVAFGKRYEEEGTERSRFQSMLNETQAMFTSFFFLD
HFPILGFIDRLTGLTNRLEKNFIEFDAFYQEIIDEHLDPKRVKPEQEDILDVLLQIWKDRSFKVQLSFDHIKALL
MNIFVGGSDTSAATVVWSMTYLMKNPLMMGKVQKEVRDVVGNKGFVNETDIQQLPYLKAVIKETMRLQPTAPLLL
PRESIEKCNLGGYEIPAKTVVHVNAWAIGRDPEAWGENPEEFKPERFIGKCIDVKGMDYELIPFGAGRRICPGIH
MGLVTVELSLANLLYAFDWRMPAGVKSEDLDMDVLPGLTMHKKNALCLVATKFL
```

3.2.2 Generating full-length sequences of CYP-candidates from gDNA

3.2.2.1 Amplification of CYP-candidates from gDNA and sequencing

Standard PCR was performed firstly with cDNA as template and corresponding primers of the identified CYP-candidates in order to amplify these sequences. However, only five candidates were successfully amplified. gDNA was used as alternative template for standard PCR experiments (see III.2.4.1). As shown in Fig. 40, additional six CYP-candidates were successfully amplified.



Sample	Annealing temperature (°C)
38991-1	51.2
38991-2	52.7
38991-3	54
74047-1	51.2
74047-2	52.7
74047-3	54

Sample	Annealing temperature (°C)
3458-1	51.2
3458-2	52.7
3458-3	54
5627-1	51.2
5627-2	52.7
5627-3	54

Sample	Annealing temperature (°C)
2114-1	54
2114-2	55.5
2114-3	58.1
2227-1	54
2227-2	55.5
2227-3	58.1

Figure 40: PCR of CYP-candidates 38991, 74047, 3458, 5627, 2114 and 2227 with gDNA as template
The individually numbered lanes represent PCR-products which are formed at the indicated annealing temperature. PCR-products are marked with boxes.

The gDNA-sequence of CYP-candidate 2114 comprises 1991 bp incuding one intron (underlined) with 434 bp and is shown below:

ATGGAGCTCCTCCAAATGTTACCTGCAATCCCACTCCCATCCCTTCTCCTCACCATCGCCACCATCTTCTTAATC
TCAATCATCATCTCCAAAAAAACCTCCCAACGCAAAAATGGGTCGTCCAACAACCTCCCGCCTGCACCCTGGAAG
CTACCGGTGCTCGGACACCTCCACCACTTCTTCACCAGCCCCGAGCCCCCACACCGCCGCCTCCGCCAGCTGGCC
CAAACCCACGGCGACGTCATGCAGCTGGAGCTCGGCGAGATCAGCCACGTCATCATTACTTCTGCTGAAGCAGCC
AAGGAAGTGATGCGGACCCACGACATCAAGTTCGCCCAGCGTCCCTTCCATCCTCACCAGGCCAAGATCATGTAC
GGCGGCATCAACCTCATCCACTCCCCCTACGGCGAGTACTGGCGCCAGCTCCGCCGCATTGCCACCCTCGAGCTC
CTCACCGCCAAGCGCGTGGAGGCTCTCCGTCGCGTGAGGGAGCCTGAGGTCCAGGCCCTCATGAGGACCATCGCG
AAGCAACGGCCCGGAGAGGTGGTGGACCTGAGCAAGGaTTTTGTTCGGTCTGTCGTACAGCATCATCTCGAAGGC
GACGTTTGGCGACGTGTCGTCGGAGCAGGAGGAGTTCATAGCGATAGCGGAGGCGCTGGTGAGGCACGGCGGTGG
GTTCGGGCTGTCGTTTTTGTTCCCTTCGAGTGGGTTGGTGCAGAAGTTATTTGGGGTGAAGGAGTGGCTGGATAA
GATGCATGAAGGGACCGACAAGTTGGCTGAGAGCATCATTAAGCAGCATAGGGCGAGGAGAGCTGTGACTAAGAA
ATCCGAGGATGAAGAGGATTTGCTGGATGTGTTGCTTAACCTTCAGGAAGATGAGACCACGCTGGGTTTCAACTT
GTCCACCGATTCCATCAAGGCTTTCCTTCTGGACGGTTAGTGGTGTGGGTGTTTGTTTATTTGCCGTTTCGGTTT
TTGTTTTTTTTTTTTNNCAATGGTTCCTTCCTGTCCTTTTCTTTGGTGGGCATGTAGGTAGCAATGGTAAGTGGG
TTCGGCTATAGTGTTATTTCTTTTTAATAATAATAATGGTATTGTTATTTGGATGTTGCATGAAAACACGCATAT
CACAAGTCACGTGCCAGTCAATACAATAAGATATAAATTTTGTTTTGTCGTATATATAACACGTGAAATGTCCAT
GTGAAGTGTCGGTCTTATCATAATTCATAATGCAAGAGGATGATAACAATTAATACTTGTTATATTCAAATTTCG
AGATCCACAAATAATTATAATAAAGAAGAGTGATGCAAAGTAATTGTTTGATCAATAAGTAACATTTATCATTGA
CTCCATGGATTACAGGACATTTTCTTAGCCGGAAGTGAGACTCCATCATCGTTGACCGAATGGGCAATGTCAGAA
TTGATGAGACATCCGGACGTATTACAAAAGGCACAAGCTGAAGTGAGAAAAGTGTTCGGTACAAAGGGAAGGGTC
GATGAAGCTGGGATTCACGAACTAACCTACTTGAAAATGGTTATCAAAGAGACACTACGACTGCATACTCCCGCG
CCACTAGYSCTTCCAAGAGAGTGTAGAGAGGAGTGCAAAATTGGCGGGTACGATATTCCTTTGAAGACAACCGTT
GTTGTGAATGCGTGGGCTATCGCAAGAGATCCTCGTTATTGGGGAAAGGAAGCTGAAAAGTTTATGCCGGAAAGG
TTTTTGAACAACGAAGTCACATTCAGAGGAGGTGATTTTGAGTTTCTTCCATTTGGAGCGGGTAGAAGGATGTGT
CCAGGAATGACGTTCGGATTAGCGGCCGTGGAACTTCCTCTCGCTAATTTGCTATACCATTTTGATTGGAAGCTT
CCAAATGGGGTCGAGCCGGCTAATTTGAATATGGATGAAATATTTGGCATCACTATCAGGAGGAAGAAGCATTTA
GAGTTGATTCCCGTGGCTCGCAACCCTGTACCAACCGTT**TAG**

The gDNA-sequence of CYP-candidate 2227 comprises 1738 bp incuding one intron (underlined) with 211 bp and is shown below:

ATGGAATGCTCCTACTCCCAATTCCTTTCTTTCCCACCCATCTTCGCCTCTCTTCTCTTCCTTGTCTCCTTCCTC
CTCCTCCTCAAACTCCACCGCGGATCAAACAGTTCCGCCGCAGCCCCGCCGCCTCCGGGACCCACTAAGCTGCCC
ATAATAGGAAACCTCCACAACCTCGCCGGCGGCTCCCTACCCCACCACCGCCTCCGGGACCTTGCCACCAAATAC
GGCGGCGTGATGGGGCTTCAGTTAGGCCAAGTCCCCCACTTGGTCATCACCCCCGCAGAAGCCGCCAAAGAAGTA
ACGAAGACCCACGACGTCGTTTTCGCCCAGCGCCCCGCCATGCTCGCCGCCGAGATCATCAGCTACAACTTCACC
GACATCGCTTTCGCCCCCTACGGCGCCTACTGGCGCCAGCTCCGCAAGATCTGCATTCTCGAACTCCTCAGCGCC
AAACGGGTCCAATCTTTCGGGTCGATCCGGGATGAAGAGGTATCGAACGCCGTGACTCGAATCGCTGCTTCGGCT
GGTCGGGAGCTTAATTTCAGCAGGATGCTCTTCTCCCTTACCTACGGGATCGCCGCCAGGGTCACTTTCGGAAAG
AAGTACAATGGGCAGGACGAATTCATTCCGGTGGTGGAAGAAATCACTCGGGTCGCGGGTGGATTCAGCCTCGCC
GACTTGTTCCCTTCGGTAAAACTGCTTCAGGTTGCGAGCGGGATGAGATCCACGCTTCTCAGGCTGAGAGGTGAA
GCTGATCGGATGTTGGGGAGTATAATCGCCGACCATAGAAGCAAGGGTAGTAAGGAAGATTACGCCGGCGGCGAG
GTGGAAGATTTGGTTGATGTCCTTTTGAAGCTTCAGGGTAATGGCGAGCTCGATTTCCCCCTCACCGATGACAAC
ATCAAGGCCGTCATTCTCGTAAGAAATCACCCCCCTCAAAATTTCCATCAGGCCTCAACGAATTTCAAAATACAG
TTATTTCAACTTTTTTTTTTTTTTTGTCAATCCTTTTTTAAAGGGGTGCAACACGAAGACTTCCCAATATTTCAA
CTTTTATTAATATAGGACAAAATATGAGTAGAATCTTAATGGTAAGCTTAAATAGTGTGTTTTTTTTCTTCTTTT
TAAGGATATATTTATTGCTGGGAGTGAGACTTCGTCGACCGCAGTAGAATGGGCGATGTCAGAAATGCTGAAGAA
TCGTAGAGTTCTCCATGAAGCACAAGCAGAGGTGAGGAGAGTTGTCGGAGGAAAAGGGGGAGTCGACGAAACCTT
ACTTCATGAGTTACACTATATGAAATTGGTGATCAAGGAGACGCTAAGACTACATGCTCCAGTGCCTTTGCTACT
TCCAAGAGAATGCGGTGAGGATTGCAAGATCGGTGGGTTTCAAGTAGAGGCGAAGAGCAAGGTGATTGTGAATGT
TTGGGCGTTTGGAAGGGATCCAAAGTACTGGAAGGAACCAGAGGAGTTTCGTCCCGAAAGGTTTATCGATAGTTC
AGTGGATTTCAAGGGGTCGAATTTCGAGTTCTTACCGTTTGGAGCCGGAAGGAGAATGTGTCCAGGAATGGTATT

TGGGATGGCGAATGTTGAGTTGCCGCTGGCCAAATTTTTGTACCATTTTGATTGGGAGCTAGCGGGTGGGATGAA
GGCAGGAGATTTGGATATGGAAGAGTCGTTTGGGGCGACGCTTACAAGAAAAAACGACCTCAAATTGATTCCAAC
CCCATACCAC**TAG**

The gDNA-sequence of CYP-candidate 3458 comprises 1801 bp incuding one intron (underlined) with 253 bp and is shown below:

ATGGAGAGGAATATCAGAGCTTTCTCTTCCTTCGATCTCCCTTTCCTCTTCTCTCTTGCAACCATCCTTGCAGTC
ACCTCATTCATCATCCGAAAGCAGAGGTCTAAACCCGCCGCCGGTTTAAACCTGCCGCCGGGGCCATGGAAGCTA
CCCATCATAGGAAACCTCCACCAGATGCTCGGCGACCAACCCCACCGCCGACTGAGGGACCTGTCCCTCAAATAC
GGCCCCGATGTGATGCAGCTTCAGCTTGGGGAGATCTCCCATATCGTGATCTCAACTCCTGAAGCAGCCAAACTA
GTGATGAAAACCCACGACATAGCTTTCGCTTCGAGGGCTTCTCTCCTGGCCACCAACATACTCTACGACGGTTGC
AAGGACATAGCGTTTGCGCCCTATGGAGAGTACTGGAGACAGATGCGCAAGATCTGCACGTTGGAGCTTTTCAGC
GCTAGAAGGGTTCAGTCTTTTCGGCACATCAGGGAAGAAGAGGTCTCCAACCTCGTGGCCTCTCTTGCTCGTTCT
GCGGGGAAGCCTGTGGATCTCACTCAGATGTTGTTTAAGCAAATGAGCACTGTTACTTCGAGGGCAGCGTTTGGT
AGAGTGCAGCAGCTAAACGATGCTTTTATGGTGGTCCTCGATAATATTTCGGATGTTCTTGCAGGTTTTACAATC
TCTGATCTCTATCCTTCCTTCAAATTCTTGCCTATCATTACTGGGTTCAGAGCTAAGTTAGAGAAGATGCATGAG
GCATCAGATTCGTTGCTAGACCGAATCATCGACGACCATAAATCGAGGAGATCGAGCAAAGGTGGTGATTATAAT
AATGATGAAAAAGAGGATCTTGTTGATGTGCTTTTGAATCTTCATGAAAATCAGAACCTTGGAGTACCCGTTACC
ATGGAAGTTATAAAAGCGGTCACTTTGGTAAGTAGATGTTAGTTCTATTCAAGTTTTGCCTCGTATATGCAAATT
TTAGCTCACTTCAATTCTCTTAGAATGTGTTTGGTTTGGATATGTGTATTTGAAATGCATATGTATTCGAAATCC
TCAATTCTAAATGTCTGCAATTGAAATGCGTGTAATTCAAATCCATTTGTTTGGATATTCATTTGCAACTTTTAG
ATATTTGAGATTTTGACTGAAAAAACTATATGTTGGATTTTGCTTTTGTTGTTAGGAACTGTTCATAGCTGGCAT
TGAGACAACAGTAACGGCGATGGAGTGGACCATGTCTGAAATTATAAAAGACTCAAGAGTACTTCAAAAGGCTCA
AAAAGAAGTCAGACATGTATTCGGCGAGCGTGGAAGTGATATTGACGAAGCAAGTCTTAATCAACTGACGTATTT
GGACATGGTTATTTCGGAAAGTTTGAGATTACACCCTCCGCTTCCATTCATTCCAAGAGAAAATAGGGAGAAGTT
GGAACTTGGTGCATATGATGTCCCCATCAACACCAAAGTTTTTGTGAATGCATGGGCGATCAATCGAGATCCTCA
TCAATGGGTCGATCCAGAGAAATTCTATCCCGAAAGATTCTTGAACTGCTCAACTGATTACAAAGGAAACGATTT
TCAGTTCCTTCCATTTGGTGCCGGAAGAAGGATATGTCCAGGGATGTCTTTCGGAATGACGAATGTTAAACTTTC
CTTGGCGAATTTGCTCTATCACTTTCAGTGGGAGCTTCCATCCGAAATGAAACAAGAAGGCATTGATATGACTGA
ACGTTTTGGAGCGTCACTTAGAGGGAAACATGCTCTTCGTTTGATTCCAATTTTATACGATGGCATGGCAGCT**TG
A**

The gDNA-sequence of CYP-candidate 38991 comprises 1686 bp incuding one intron (underlined) with 99 bp and is shown below:

ATGGATATCATCATCTCCCACCTCCCAGAAATCCTCGTCGCTCTAGCCGCCGCCGTGTTCATCCACCTATGGCGG
AAGCAGCAGCAGCAGAAGAACAGCGGAGGAGTCCCCGAAATCCCCGGCGGCTTGCCATTAGTCGGCCATCTCCAC
CTCCTCGCCCGCGGCAACCAGACCCTCTCCCGGAAACTAGCAGCCTTCTCCGAAAAACACGGACCCGTCTTCACC
ATCCGCTTGGGAGCCAACGCCGCCGTCGTCGTCAGCGACTACGAGTCAATGAAAGAGTGCTTCACCACCAACGAC
CGGACCCTCTCCTTCCGGCCGGACTCCAGCCAGGCCAAAGTCCTCGGCTACAACTACGCAGCCTTCGGGTTCGCT
TCCTACGGGAACTACTGGCGGGACATGAAGAAGATGCTCATGACCGAGCTCCTCTCCGTCCAGCGGATCAAAGCC
CTCCGCGGCGTCCAGATCTTCGAGGTCAATTCCTTAATCCGCGACCTCTACAAACAAAGCAAGGTTGGTGACAAG
TCGATTGTGATCAGCGACGCGCTGCAGAGCTACGTGCTGAACGTCATCACAAGCATGGTCGCGGGGAAGAGGTAC
TTTAACAAGACGGAGGAGGAGGAGGAGGTGTCGTCGACGACGACCTCCGGAGGGAGGCCGATAGGTCAGGTGATG
AGGGAGTTCATGCTCGTCACTGGAACCCTCGTCCCCTCCGATCTGATCCCGATTCTCGGGTGGTTCGGGTTCCAG
GGAGTCGTCAAGGATATGAAGCGGGTGTCGAAGGAGCTCGACGTGATCATGGAGAGCTGGCTGGACGACCACAAG
AGGAAGAAGAAACACCCGACGACGACGAACGGCGGCGCGAATCCGGACCTTATCGACGTCATGCTCTCCGAAATC
AAAGAAGAAGTCGCTTACGGCCACAAGCGCGACGACATCATCAAGGCTACCGCCATGGGTAGGTGGTAGTTTATT
CATTAACATTTTCATTCTCTGAGATGATCTTTGCCAATGGTGATTAATTATGATTATTTGTTGTGATTTATCTTT
GTTACATCTTTGATCGTGGCGGGGTCGGATACGACGTCGATTACGCTGACGTGGATCTTATCCAATCTGCTGAAC
CACGAGCGGGTGATGAAACTGGCTCAGAAAGAGATCGACGACGTGGTCGGGACAGAACAGTGGGTGGACGATTCG
GATATCGAAAAGTTGACGTATTTGGGAGCGGTGATTAAGGAGACGTTGAGAATGTACCCGCCGGGTCCGCTTTCA
GTGCCTCGAGAGGCGTCTGAGGATATCACGATTAGAGGGTACCGCGTTCCGAAAGGTACGCGGTTCTTGGCCAAC
TTTTGGAAGTTGCATAGGGATCCGAAGGTTTGGTCCGACCCGGACGAGTACAGACCGGAGAGGTTCTTGACTGAT

AATGTCAATATGGATATATTCGGTCAGCAGTTCGAGTATCTTCCGTTCGGGTCGGGAAGGAGAGGGTGCCCCGGG
ATCAATTTCGGGATGCAGGTGACCCAATTGACTTTGGCGAGGCTGCTTCAGGGGTTCCATTGGGGTACTCCGGAC
AATAAACCTGTTGATATGACGGAAGGGTTGGGAATTGCGCTTCCGAAGGCTAATCCGCTCGAGGTCGTCCTAACG
CCGCGCCTTTCTCCCGAGTTGTATGGAGTACTT**TAG**

The gDNA-sequence of CYP-candidate 5627 comprises 1716 bp incuding one intron (underlined) with 42 bp and is shown below:

ATGGATCTGTTCCTTCCATCCCTGAGCTCTGCAATTCCAGGAATTGCTGTTGCTCTGTTTCTGTTTCTGTTCATC
ATCGCGAAAAACAGAGCAGGAAAGGGCAAGCGATCAGTTCCCGAAGCAGGGCGATCATGGCCGATCATCGGCCAT
CTCCTTTCGCTGGCCGGAGGAACAGAGCAGCAACTTCCCCACTTGATACTCGCAGAGCTTGCTGACAAATACGGC
CCGATCTTCATGCTCAGGATCGGCATGTTCCCGACGGTGGTCGTCAGCAGCTCGGAGATTGCTAAAGAGCTTTAT
ACTACTCACGACGCCGCCATCTCTTCCCGCCCCAACTTCACCGCTTCTGCGATTTTGGGGAACAACAGCGCCAAT
TTCGGGTTCTCGCCGTACGGCGAGTACTGGCGGATGATGAGGAAGATCACGGCGCTGGAGCTGCTCTCCGGCCGG
AGACTCGACCTTCTGAAGCGCGTCCGGACGTCCGAAGTGGAATCCGCCGTCAAGGATCTGTATATGTCGTGGGAG
AAGGAGAAGGAGAAGCACAGAGAAGAAGAGGGGACGGTCTGTGTTGAGATGAACAAGTGGTTTGGGGATATGAAC
TTGAACGTTATACTCGGGATGATCGCCGGGAAGAGGTACGGCGGCGGCGGCGAGGAGGAGGAGGAGGCGGAGCGG
TGCCGGAAGCTGATGAGAGACTTCTTCCACTACGCAGGGCAGATCGTGGCGAGAGACCTGTTCCCGTTTCTAGGG
TTTCTGGATATCGGAGGGCACGAGAGGACGATGCGGAGACTAAACGCCGATCTGAACGGACTCGCCGAGCAATGG
ATCGAGGAGCATCGCCGTCGCCGTCGTGACGGCGACGGCGAAAAGGAGGAGAAGGGAATGCGGACGACRAAGAAC
GATGAACATCGCGATTTCATCGACGTGTTGCTCTCGGTACTGAAAGATGTCGATCTCGCCGGTTACGATTTGGAC
ACTGTCTGCAAGTCCACGATCATGGTACCGTAGTAGTAGTTGTGAATTTGATTGTCGCGTCGGGATTTATAATTC
ACGCTTTATAATTTCTCGTCGGTTTATTATTTGCAGACTCTGATCGTCGGAGGTACCGACACGACGACGGTGACT
CTGACGTGGGCTCTGTCCCTGATCTTGAACAATCGGAGAACCCTAAAGAGGGTGCAGGAAGAGCTCGACGCAGTC
GTCGGGAGGGAGAGGCTGGTGGATGAATCGGACGTCCACAAGCTTCCATACTTTCAGGCCATGATCAAGGAGGTG
ATGAGGCTGTACCCGGCGGGTCCAATCGGAGGGCCTCGAGAATTCTCTCGAGACTGCTCCGTCGCCGGGTACCAT
GTCCCGGCGGGGACTCGGCTGATAGTGAACCTCTACAAGCTGCAGACCGACCCGAAGGTGTGGTTAGAGCCGATG
GAGTTCAGGCCGGAGAGGTTCCTCGATGCAGAGTGCAAGGATGTTGATGTGAAAGGGCAACACTTCGAGCTCATC
CCGTTCGGTGCTGGAAGAAGGTCGTGCCCTGGAATCAACTTTGGTATGCAGATGATCGGACTCGCCCTCGCGAGC
TTCCTGCAAGCGTTTGAGGTCTCTGTCCCGTCGGATGCACCGGTCATCGACATGACTGCGGTTTCTGGGCTGACG
GTGAGCAAAGCCACCCCGCTCGAAGTTCTGGTTAAGCCACGCTTGCCTTGGAGGCTCTACCAA**TAA**

The gDNA-sequence of CYP-candidate 74047 comprises 1690 bp incuding one intron (underlined) with 91 bp and is shown below:

ATGGATTCCATAGCTCTACCCGTCATTTTCTCCCTCCTAACCATCTTCCTCTTCTCTTTCATCATCCTCCCGCGG
ATCCTTAACCGGAGATCAAAACCCAGACCACCGCCGCCAGAACCATGGAGGTTGCCGATAATCGGCCACATTCAC
CATTTGGCCGGCGGCGGCGGTAGCAGCGCTTTACCCCACCGCCGCCTCAGAGACCTGGCCAAAACTCACGGCCCG
GTGATGCTGCTCCGCCTCGGCGAAATCAACCACGTCATCATCTCCTCCGCCGATGCCGCCAAGCAAGTGTACAAG
ACCCACGACGTCGTTTTCGCGCAGCGGCCCTTCAACCTCGCCGCCAAGATCATCACCTACGACCGCTCCGACATC
GCCCACGCCGCCTACGGCGATTTCTGGAGGCAGCTGCGCAAGCTCTGTACCCTCGAGCTTCTCAGCCCGAAACGG
GTCCTATCGTTCCGGCCAATTCGCGAGGAGGAAGGGTTGAAATTCGTCCGGCGAATTTCCGCCGCCGCGGCGGCG
GAGGGTTCGCGGCCGGTGAATTTATCGAGGCTGATTTTCTCGCTGACGTACGGCGTGACGTCGAGGATCGCGTTC
GGGACGGTGAGGGAGGAGAAGGACGATGAGTATATTCCGGTGGTGGAGGAGATTATGGCGGCGGCGGCGGGGTTC
GGGGTGGCGGATCTGTTCCCGTCGGTGAAATGGCTGGAGAGGGTGAGCGGGATGAAAGCCAAGCTCGAGCGGCTG
CACGGAGTTCAGGATAGGTTGCTGGAGAAGATCATCGCCGATCATCGGGCGAGGAGATCGTCGCCGTCGTCGGCG
GCGGCGGCGGGAGAAAACAGGGGAGGAGGAGCAGGGGAACGATACGACGTCGCAGAGGATCTAGTGGATGCTCTG
TTGAATCTCCAGAGAGACGACGACGACGTCGGATTCAGCTTGACAAATGACAACATCAAAGCGGTTATTGGTAAT
CATTTCTTTCGGCTTCATCCTTTCATAACCCCGTCACAGTTTGGTTGTATGATTTTTTTTTTATTTCCTCGGATT
GTGGTTTTTTCAGGATGTGTTTATTGCCGGGAGCGAGACTTCGTCGACGACGGTGGAATGGGCGATTTCGGAGCT
GATAAGGAATCCGGACGCGATGAGAAGGGCACAAGAAGAAGTGAGACGAGTGTTCGGCGAGAGAGGGAGAGTCGA
CGAGGCGGGGCTGGATCGGTTGAGCTATTTGAAACTGGTGATTAAAGAGACTTTGAGACTGCACACTCCCGCTCC
TCTTTTGGTTCCAAGAGAGAGCAGGGAGGAGTGCGAGGTTGGCGGGTATTTGATACCCGTTAAAACAACTGTGTT
GGTGAATGCGTGGGCGATTGCGAGGGATCCTGGTACGTGGGACGAACCGGAAGAGTTTCGTCCGGAGAGGTTTCT
CGATGGCAAGGTGGATTATAAGGGGACGAATTTCGAGTATCTACCGTTCGGGTCGGGAAGGAGGATGTGTCCCGG

GATAACGTTCGGCGTGGCGAATACGGAGTTACCGCTTGCCAATTTGCTTTTCTATTTTGATTGGAAGCTTCCGGG
TGATTTGAAGGTGGAGAGGTTGGATATGGATGAATCGTTTGGTGTCACGGTGAGACGGAAAAATGACCTGGAAGT
AGTTCCTGTAATTCGCTATCCTCCACCAATAATTTCT**TGA**

In many eukaryotic genes, the coding sequences are interrupted by introns, which are cleaved from pre-mRNAs during splicing. The introns in gDNA sequences of the successfully amplified CYP-candidates were identified by comparing with the ORF-sequences obtained from the transcriptome database. Depending on the position of an intron in coding sequences, introns are classified into three different phase classes: phase 0 (located between two codons), phase I (located between the first and second nucleotide of a codon) or phase II (splitting codons between the second and third nucleotides) (Long et al., 1995). Furthermore, introns spliced by the major spliceosome belong to the U2-type and less abundant introns of the U12-type are processed by the minor spliceosome (Dietrich et al., 1997). The consensus sequences located at the intronboundaries are important for proper recognition and removal of introns. Dinucleotide sequences at these boundaries are relevant for proper splicing and conserved in multiple introns. The vast majority of introns are flanked by GT-AG splice site dinucleotides (Aebi et al., 1986; Dietrich et al., 2005). The often exception to this rule are GC-AG and AT-AC splice sites (Sheth et al., 2006). Many introns that have sequences other than GT-AG, GC-AG or AT-AC at the intronboundaries, so-called non-canonical splice sites, have been reported to be efficiently removed as well (Szafranski et al., 2007).

The intron phase and length of all introns of the six CYP-candidates identified from gDNA are listed below:

CYP-candidates	Intron phase	Length	Start position	Dinucleotide sequences at intron boundaries (5' - 3')
2114	II	434	932	AC-GG
2227	I	211	919	GT-AG
3458	I	253	928	GT-AG
38991	0	99	957	GG-CA
5627	I	42	1006	AG-AT
74047	II	91	971	TA-CA

Introns tend to be more abundant close to the 5' end of genes (Sakurai et al., 2002). However, all CYP-candidates identified here contain only one intron that is located at appr. two third of the sequences, close to the 3' end. In most coding regions, the most common are introns phase

0, followed by introns phase I and the least common type is intron phase II (Tomita et al., 1996). Whereas introns phase I are the most common in proteins with secretory signal peptides (Vibranovski et al., 2006). The intron phases of the sic CYP-sequences identified here are not consistent and the majority is phase I (50%). In addition, just two introns of CYP-candidates 2227 and 3458 have the consensus sequences GT-AG at the intronboundaries. The other four candidates contain non-canonical splice sites.

The amino acid sequences of the candidates were translated from full-length sequences and the molecular weight of the proteins calculated with the respective ExPASy tools.

Candidate	Number of amino acids	Molecular weight (kDa)	Candidate	Number of amino acids	Molecular weight (kDa)
11511	536	59.2	2114	518	58.5
11862	514	57.4	2227	508	56.1
2408	534	59	3458	515	58.8
27263	461	51	38991	528	59.3
4471	504	57.3	5627	557	62.8
74047	532	59.5			

The amino acid sequence of CYP-candidate 2114 is as follows:

```
MELLQMLPAIPLPSLLLTIATIFLISIIISKKTSQRKNGSSNNLPPAPWKLPVLGHLHHFFTSPEPPHRR
LRQLAQTHGDVMQLELGEISHVIITSAEAAKEVMRTHDIKFAQRPFHPHQAKIMYGGINLIHSPYGEYWR
QLRRIATLELLTAKRVEALRRVREPEVQALMRTIAKQRPGEVVDLSKVLFGLSYSIISKATFGDVSSEQE
EFIAIAEALVRHGGGFGLSFLFPSSGLVQKLFGVKEWLDKMHEGTDKLAESIIKQHRARRAVTKKSEDEE
DLLDVLLNLQEDETTLGFNLSTDSIKAFLLDIFLAGSETPSSLTEWAMSELMRHPDVLQKAQAEVRKVFG
TKGRVDEAGIHELTYLKMVIKETLRLHTPAPLVLPRECREECQIGGYDIPLKTTVVVNAWAIARDPRYWG
KEAEKFMPERFLNNEVTFRGGDFEFLPFGAGRRMCPGMTFGLAAVELPLANLLYHFDWKLPHGVEPVNLN
MDEIFGITIRRKKHLELIPVARNPVPTV
```

The amino acid sequence of CYP-candidate 2227 is as follows:

```
MECSYSQFLSFPPIFASLLFLVSFLLLLKLHRRSNSSAAAPPPPGPTKLPIIGNLHNLAGGSLPHHRLRDLATKY
GGVMGLQLGQVPHLVITSAEAAKEVMKTHDVVFAQRPAMLAAEIISYNFTDIAFAPYGAYWRQLRKICILELLSA
KRVQSFGSIRDEEVSNAVTRIAASAGRELNFSRMLFSLTHGIAARVTFGKKYNGQDEFIPVVEEITRVAGGFSLA
DLFPSVKLLQVASGMRSTLLRLRGEADRMLGSIIADHRSKGSTEDSAGGEVEDLVDVLLKLQGNGELDFPLTDDN
IKAVILDIFIAGSETSSTAVEWAMSEMLKNRRVLHEAQAEVRRVFGGKGGVDETLLHELHYMKLVIKETLRLHAP
VPLLLPRECGEDCKIGGFQVEAKSKVVVNVWALGRDPKYWKEPEEFRPERFIDSSVDFKGSNFEFLPFGAGRRMC
PGMVFGMANVELPLAKFLYHFDWELAGGMKAGDLDMEESFGATLTRKNDLKLIPTPYH
```

The amino acid sequence of CYP-candidate 3458 is as follows:

```
MERNIRAFSSFDLPFLFSLATILAVTSFIIRKQRSKPAAGLNLPPGPWKLPIIGNLHQMLGDQPHRRLRDLSLKY
GPDVMQLQLGEISHIVISTPEAAKLVMKTHDIAFASRASLLATNILYDGCKDIAFAPYGEYWRQMRKICTLELFS
ARRVQSFRHIREEEVSNLVASLARSAGKPVDLTQMLFKQMSTVTSRAAFGRVQQLNDAFMVVLDNISDVLAGFTI
SDLYPSFKFLPIITGFRAKLEKMHEASDSLLDRIIDDHKSRRSSKGGDYNNDEKEDLVDVLLNLHENQNLGVPVT
```

MEVIKAVTLELFIAGIETTVTAMEWTMSEIIKDSRVLQKAQKEARHVFGERGSDIDEANLNQLTYLDMVISESLR
LHPPLPFIPRENREKLELGAYDVPINTKVFVNAWAINRDPHQWLDPEKFYPERFLNCSTDYKGNDFQFLPFGAGR
RICPGMSFGMTNVKLSLANLLYHFQWELPSEMKQEGIDMTERFGASLRRKHALRLIPILYDGMAA

The amino acid sequence of CYP-candidate 38991 is as follows:

MDIIISHLPEILVALAAAVFIHLWRKQQQQKNSLNRGGGVPEIPGGLPLVGHLHLLARGNQTLSRKLAAFSEKHG
PVFTIRLGANAAVVVSDYESMKECFTTNDRTLSFRPDSSQAKVLGYNYAAFGFASYGKYWRDMKKMLMTELLSVQ
RIKALRGVQIFEVNSLIRDLYKQSKVGDKSIVISDALQSYVLNVITSMVAGKRYFNKTEEEEVSSTTTSGGRPIG
QVMREFMLVTGTLVPSDLIPILGWFGFQGVIKDMKRVSKELDVIMESWLDDHKKKKHPTTNGGANPDLIDVMLS
EIKEEVAYGHKRDDIIKATAMSLIVAGSDTTSITLTWILSNLLNHERVMKLAQKEIDDVVGTDRWVDDSDIEKLT
YLGAVIKETLRMYPPGPLSVPREASEDITIRGYRVPKGTRFLANFWKLHRDPKVWSDPNEYMPERFLTANANMDI
FGQQFEYLPFGSGRRGCPGINFGMQVTQLTLARLLQGFHWGTPDNKPVDMTEGLGIALPKANPLEVVLTPRLSPE
LYGVL

The amino acid sequence of CYP-candidate 5627 is as follows:

MDLFLPSLSSAIPGIAVALFLSLFIIAKNRAGKGKRSVPEAGRSWPIIGHLLSLAGGTEQQLPHLILAELADKYG
PIFMLRIGMFPTVVVSSSEIAKELYTTHDAAISSRPNFTASAILGNNSANFGFSPYGEYWRMMRKITALELLSGR
RLDLLKRVRTSEVESAVKDLYMSWEKEKEKHREEEGTVRVEMNKWFGDMNLNVILGMIAGKRYGGGGEEEEEAER
CRKLMRDFFHYAGQIVARDLFPFLGFLDIGGHERTMRRLNADLNGLAEQWIEEHRRRRHGGDGEEEEKGMRTTKN
DEHRDFIDVLLSVLKDVDLXGYDLDTVCKSTIMVPLFTLYNFSSVYYLQTLIVGGTDTTTVTLTWALSLILNNRR
TLMRVQEELDAVVGRERLVDESDVHKLPYFQAMIKEVMRLYPAGPIGGPREFSRDCSVAGYHVPAGTRLMNLYKL
QTDPKVWLEPMEFRPERFLDAECKDVDVKGQHFELIPFGAGRRSCPGINFGMQMIGLALASFLQAFEVSVPPDAP
VVDMTAVSGLTVSKATPLEVLVKPRLPWRLYQ

The amino acid sequence of CYP-candidate 74047 is as follows:

MDSIALPAIFSLLTIFLFSLIILPRILNRRSKPRPPPPEPWRLPIIGHIHHLAGGGGSSALPHRRLRDLARTHGP
VMLLRLGEINHVIISSADAAKQVYTTHDVVFAQRPFNLAAKIITYDRSDIAHAAYGDFWRQMRKLCTLELLSPKR
VLSFRPIREEEGLKFVRRISAAVAAEGSRPVNLSRLIFSLTYGVTSRIALGTVREEKDDEYIPVVEEIMAAAAGF
GVADLFPSVKWLERVSGMKAKLERLHGVQDRLLEKIIADHRARRSSPSSSSAAGENRGGGAGERYDVAEDLVDAL
LNLQRDGDDVGFSLTNDNIKAVIGDVFIAGSETSSTTVEWAISELIRNPDVMRTAQEEVRRVFGERGRIDEAGLD
RLSYLKLVIKETLRLHTPAPLLVPRESREECEVGGYLIPVKTTVLVNVWAIARDPGTWDEPEEFRPERFLDGKVD
YKGTNFEYLPFGAGRRICPGITFGVANTELPLANLLFYFDWKLTGDLKVERLDMDESFGVTVRRKNDLEVVPVIH
YPPPIIS

Amino acid sequence alignment was conducted between the ORFs of all eleven candidates by Clustal Omega to search for the conserved domains in the sequences and the results are shown in Fig. 41.

```
4471       ------------------MPSLLIYLLLALPPVLLLFIFQKHKKKN----SKSASSLPP     37
2408       MAASLTFSPLKSTFPGGNNPPPPSLRLHR--I-PTRISCQNQKPQSE----STTKKQLPP     53
3458       MERNIRAFSS------FD-LPF-LFSLATILA-VTSFIIRK--QRSK----PAAGLNLPP     45
2114       MEL-LQMLPA------IP-LPSLLLTIATIFL-ISIIISKKTSQRK-----NGSSNNLPP     46
2227       MEC--SYSQF------LS-FPPI-FASLL-FL-VSFLLLLKLHRRSN----SSAAAPPPP     44
74047      -------MDS------IA-LPAI-FSLLTIFL-FSLIILPR--ILNR----RSKPRPPPP     38
38991      MDIIISH------------LPEILVALAAAVF-IHL--WRKQQQQKN---SLNRGGGVPE     42
11511      MDFFTSLST-----ASLSTVAIATASLAFLYF-LFRTILSPAGDSKPHAQQQQQAPPPVE     54
27263      ------------------------------------------------------------     0
11862      MDSLFAS------------IALAVSFLCLAGF-FLTDGLRRLG-----S--DKQKRYPPE     40
5627       MDLFLPS------------LSSAIPGIAVALF-LSLFIIAKNR-----A--GKGKRSVPE     40

           proline-rich region
4471        GPRGLPLIGNLFQLDPS-----APHRYLWQLSQTYGP-LLSLRLGRVQAVVVSTAKMAQE     91
2408        GPPKLPLIGNLHNLVGA-----LPHQALRNLATEYGP-LIHLQLGEISAAVVSNPAIAQE     107
```

```
3458       GPWKLPIIGNLHQMLGD-----QPHRRLRDLSLKYGPDVMQLQLGEISHIVISTPEAAKL     100
2114       APWKLPVLGHLHHFFTS---PEPPHRRLRQLAQTHGD-VMQLELGEISHVIITSAEAAKE     102
2227       GPTKLPIIGNLHNLAGG----SLPHHRLRDLATKYGG-VMGLQLGQVPHLVITSAEAAKE     99
74047      EPWRLPIIGHIHHLAGGGGSSALPHRRLRDLARTHGP-VMLLRLGEINHVIISSADAAKQ     97
38991      IPGGLPLVGHLHLLARG-N--QTLSRKLAAFSEKHGP-VFTIRLGANAAVVVSDYESMKE     98
11511      ASGAWPIIGHLHLLAGP----QPAHIVLGQMAEHHGP-IFTIRMGVHRAVVVSNRETAKE     109
27263      -----------------------------MADKYGS-IFTVRMGRQRALVVSSWPIAKE     29
11862      PANRWPVIGHLPVMAGS----ELPHRALAALADKYGP-IFTLRLGFNRVVIVSSPEIAKE     95
5627       AGRSWPIIGHLLSLAGG-TEQQLPHLILAELADKYGP-IFMLRIGMFPTVVVSSSEIAKE     98
                                      ::  :*  :: :.:*     :::     :

4471       VMKTQDLVFCNRPQAMGTTKLSYNRVDLAFAPYDAYWREMRKICMVYLFNSNRTQSFGPI     151
2408       IMKTHDLNFADRPCIFAAEIATWGGQDIAFSPHGEYWKQMKRISLTELLGPRKTQSFRGI     167
3458       VMKTHDIAFASRASLLATNILYDGCKDIAFAPYGEYWRQMRKICTLELFSARRVQSFRHI     160
2114       VMRTHDIKFAQRPFHPHQAKIMYGGINLIHSPYGEYWRQLRRIATLELLTAKRVEALRRV     162
2227       VMKTHDVVFAQRPAMLAAEIISYNFTDIAFAPYGAYWRQLRKICILELLSAKRVQSFGSI     159
74047      VYTTHDVVFAQRPFNLAAKIITYDRSDIAHAAYGDFWRQMRKLCTLELLSPKRVLSFRPI     157
38991      CFTTNDRTLSFRPDSSQAKVLGYNYAAFGFASYGKYWRDMKKMLMTELLSVQRIKALRGV     158
11511      CLTTHDRVFADRPATLAMDLLGYRRSMLGFSPYGAYWRQIRKIVTLELLSSHRLETLKHV     169
27263      CLVTVDKPFLNRPKNLAAKIMTYDIAMFGFSPYGHYWREMRKLVSLELLSAQRLRTFAPV     89
11862      LFTTKDTVALSRPRMIATETFSYDFAVFGMGPHGDYWRETRKIAVLELLSNRRLDLLRRF     155
5627       LYTTHDAAISSRPNFTASAILGNNSANFGFSPYGEYWRMMRKITALELLSGRRLDLLKRV     158
              * *     *               :  . :. :*:  :::    *:  .:   :  .

4471       REYEVSQMVGKASRS-S--------LLASQPFNLSEAMMSLTTTIICRVAFGKRYEEE--     200
2408       RETEVAGMIESVR------------KSVGKPVNVTEKVLSLTNTITCKTAFGYQC-MD--     212
3458       REEEVSNLVASLA------------RSAGKPVDLTQMLFKQMSTVTSRAAFGRVQ-QL--     205
2114       REPEVQALMRTIAK-----------QRPGEVVDLSKVLFGLSYSIISKATFGDVS-SE--     208
2227       RDEEVSNAVTRIA------------ASAGRELNFSRMLFSLTHGIAARVTFGKKY-NG--     204
74047      REEEGLKFVRRISAAVA--------AEGSRPVNLSRLIFSLTYGVTSRIALGTVREEK--     207
38991      QIFEVNSLIRDLYKQS---------KVGDKSIVISDALQSYVLNVITSMVAGKRYFNKTE     209
11511      RESEIHVATKELYAAA---------GGGNGVVEMGTWFGEITLNVILKMIVGKNVGYFS-     219
27263      RESEVRAATKALYDLC---------DGSSVAVEMKRWFADVTLNVIMRIIAGKTIGYGSV     140
11862      KESEVHNSINDLYKQI----------GGKSEIDVNKWIGDLSMNLMLKLVIGKSSSSAS-     204
5627       RTSEVESAVKDLYMSWEKEKEKHREEEGTVRVEMNKWFGDMNLNVILGMIAGKRYGGGG-     217
           :  *                              . .   .      :      *

4471       --------GTERSRFQSMLNETQAMFTSFFFLDHFPILGFIDRLTGLTNRLEKNFIEFDA     252
2408       --------Q---EEFVGLMNGAVQAAGGFNIADLYPSLGFLQGLTGMKSELLRIRNGLDR     261
3458       --------N---DAFMVVLDNISDVLAGFTISDLYPSFKFLPIITGFRAKLEKMHEASDS     254
2114       --------Q---EEFIAIAEALVRHGGGFGLSFLFPSSGLVQKLFGVKEWLDKMHEGTDK     257
2227       --------Q---DEFIPVVEEITRVAGGFSLADLFPSVKLLQVASGMRSTLLRLRGEADR     253
74047      --------D---DEYIPVVEEIMAAAAGFGVADLFPSVKWLERVSGMKAKLERLHGVQDR     256
38991      EEEVSSTTTSGGRPIGQVMREFMLVTGTLVPSDLIPILGWFGF-QGVIKDMKRVSKELDV     268
11511      -------GGGEGVKLRKLLKDFFELVGWFLVTDGLPFLRWLDV-GGSEKAMRKTAKELDV     271
27263      VAR--DGGGGAMVGWQKSLRDFFDLLGKFTVADAVPALRWLDV-GGYEKEMRKTAKELDT     197
11862      ------GGSAMMVKFQIAVRELFHLAGLSLVGDALPFLRWMDI-GGHEKAMKRAAKAMDG     257
5627       ------EEEEEAERCRKLMRDFFHYAGQIVARDLFPFLGFLDI-GGHERTMRRLNADLNG     270
                                           *    .    *    : :     :

4471       FYQEIIDEHLDPKRVK------------------PEQEDILDVLLQIWKDR---SFKVQL     291
2408       IFDKIIKQHEEKRANGDVE--------------LDDEDLMDVLLRLQGSG---GFKCPI     303
3458       LLDRIIDDHKSRRSSKGGD------------YNNDEKEDLVDVLLNLHENQ---NLGVPV     299
2114       LAESIIKQHRARRAVTKKS--------------EDEEDLLDVLLNLQEDET--TLGFNL     300
2227       MLGSIIADHRSKGSTEDSA--------G------GEVEDLVDVLLKLQGNG---ELDFPL     296
74047      LLEKIIADHRARRSSPSSSSAAGENRGGGAGERYDVAEDLVDALLNLQRDGD--DVGFSL     314
38991      IMESWLDDHKKKKKHPT--------------TNGGANPDLIDVMLSEIKEE---VAYGHK     311
11511      VVGEWMREHEGKRGSAGGG----K--------KAAAEEDFMDVIMNVVGGEEGGGVVGRD     319
27263      VVGGWLEEHKRKRVVAAAG----S---------GKVEEDFMDVILNVIGD--SGGIDGRD     242
11862      YLQEWLNEHKENRGGEK------------------TEHDFIDVMLSHLDGQ---SLHGYD     296
5627       LAEQWIEEHRRRRHGGDGE----EEEKGMRTTKNDEHRDFIDVLLSVLKDV---DLXGYD     323
                : :*                              *::*.::

                                  oxygen binding domain
4471       SFDHIKALL----------------MNIFVGGSDTSAATVVWSMTYLMKNPLMMGKVQKE     335
2408       TSTNIKAVL----------------VDLFIAGTDTSSTTVEWAMSEMMKNPRVMQKAQAE     347
3458       TMEVIKAVT----------------LELFIAGIETTVTAMEWTMSEIIKDSRVLQKAQKE     343
2114       STDSIKAFL----------------LDIFLAGSETPSSLTEWAMSELMRHPDVLQKAQAE     344
2227       TDDNIKAVI----------------LDIFIAGSETSSTAVEWAMSEMLKNRRVLHEAQAE     340
```

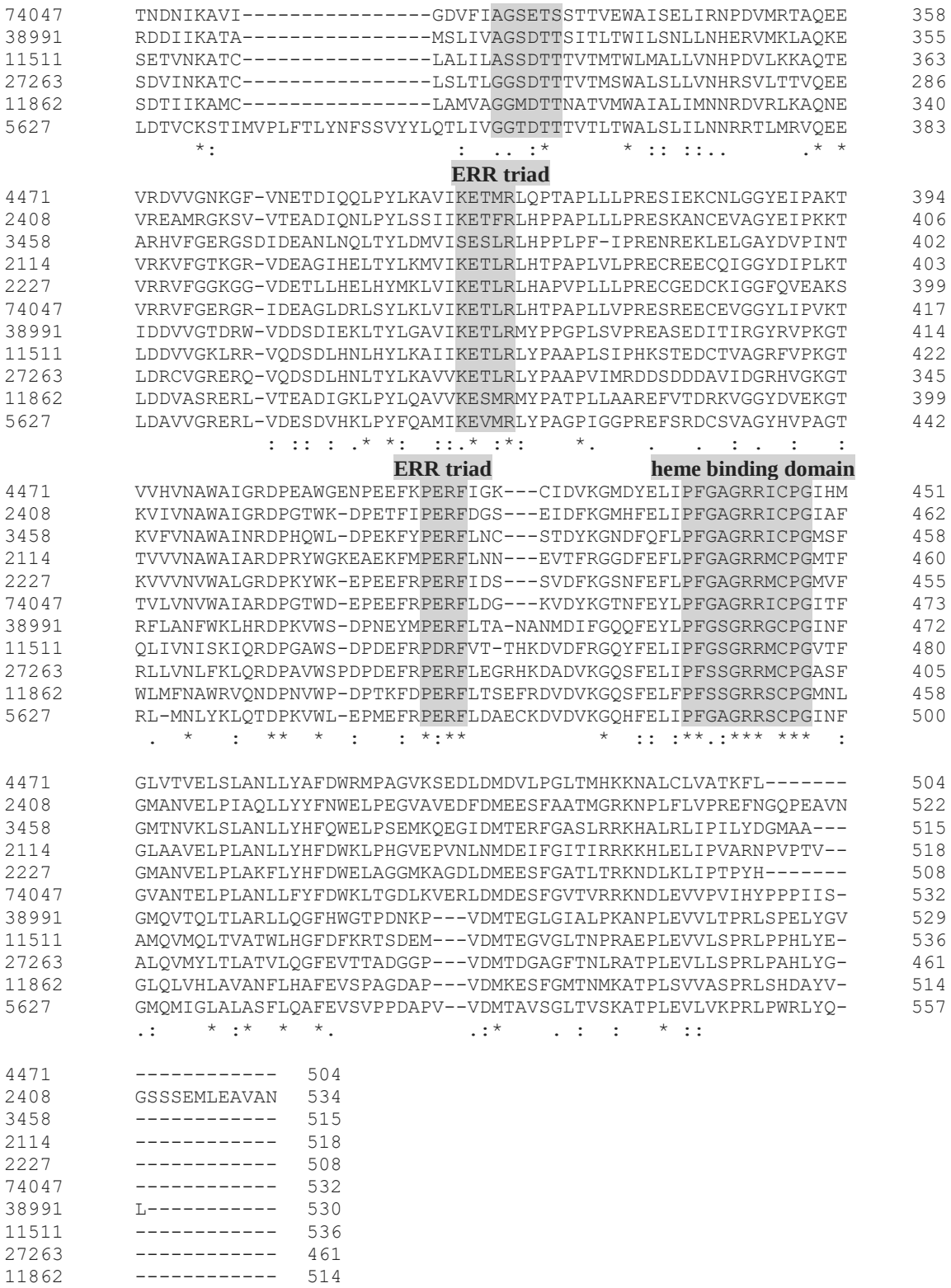


Figure 41: Amino acid sequence alignment of CYP-candidates
Gaps were inserted to maximise the homology. The typical motifs of the proline-rich region, oxygen-binding domain and ERR triad are shaded and marked.

The sequence conservation is relatively low among the CYP-candidates. Nevertheless, the amino acid sequences of CYP-candidates still have several conserved areas. The first 20-30

amino acids at the *N*-terminus are predominantly hydrophobic and form the hydrophobic membrane anchor. Located near the *N*-terminus of seven candidates including 2114, 2227, 2408, 3458, 38991, 4471 and 74047 is the proline-rich region with the consensus sequence (P/I)PGPx(G/P)xP. It provides sufficient flexibility between the membrane anchor and the globular part and helps to stabilise the enzyme (Kemper, 2004). About 60% away from the C-terminus of all sequences is the oxygen-binding domain with the consensus sequence (A/G)Gx(D/E)T(T/S). The sequences of all candidates have the EER triad ExxR.....R containing the glutamine and arginine from the consensus sequence KETLR and the arginine from the consensus sequence PERF. It helps to stabilize and position the heme in the binding pocket (Hasemann et al., 1995). The heme binding region is located about 15% from the C-terminus of all sequences. It is the most conserved domain with the consensus sequence PFGxGRRxCxG, in which the cysteine serves as the fifth axial ligand of heme-iron. The ERR triad, oxygen binding domain and heme binding domain were found in the sequences of all candidates.

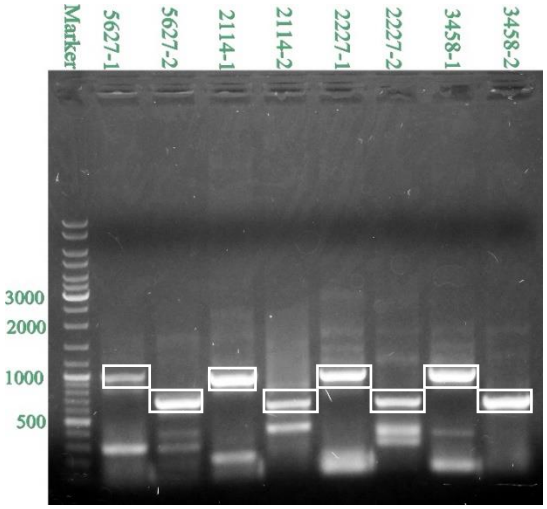
Furthermore, based on the PSI- and PHI-BLAST searches against the conserved domain database (CDD) of the National Center for Biotechnology Information, all eleven CYP-candidates were confirmed belonging to the cytochrome P450 family.

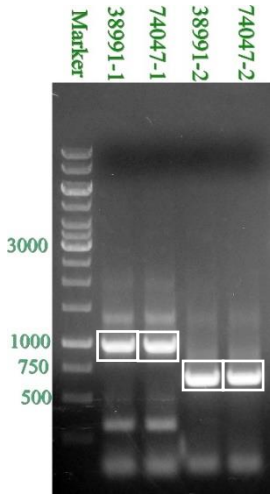
3.2.2.2 Fusion-PCR and verification of full-length sequences

In order to generate full-length sequences with continuous coding capacity, fusion-PCR reactions were performed for the *in-vitro* removal of introns from genomic DNA sequences (see III.2.4.3).

3.2.2.2.1 Exon fragments in the first rounds

Distinct exon fragments were generated in the first round with Phusion® Polymerase to avoid replication errors (see III.2.4.3). Afterwards, the PCR products were analysed by agarose gel electrophoresis. Fig. 42 shows the successful amplification of two exon fragments of each of the CYP-candidates.





Sample	Annealing temperature (°C)
5627-1	54
5627-2	54
2114-1	58.1
2114-2	58.1
2227-1	58.1
2227-2	58.1

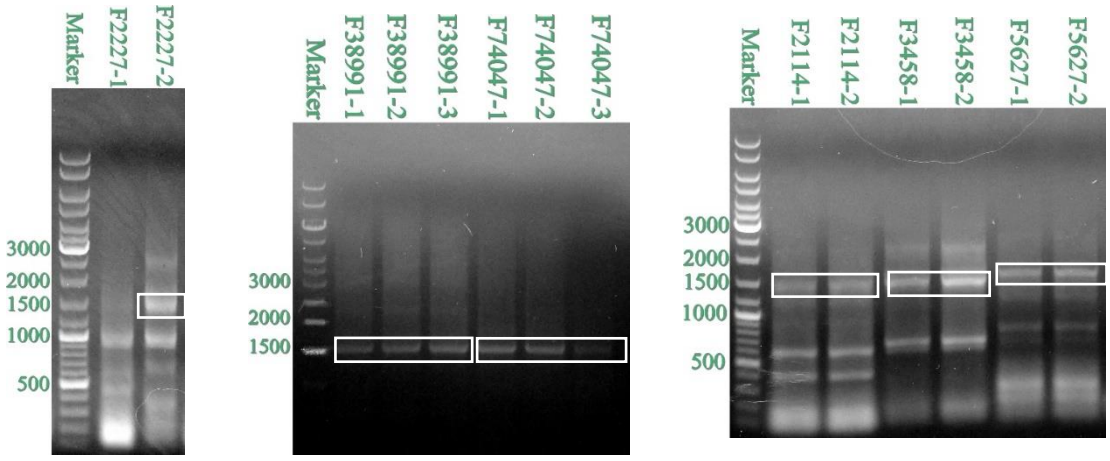
Sample	Annealing temperature (°C)
3458-1	54
3458-2	54
38991-1	54
38991-2	54
74047-1	54
74047-2	54

Figure 42: Two distinct exon-fragments each of candidates 2114, 2227, 3458, 38991, 5627 and 74047
PCR-products are marked with boxes.

The bands of PCR-products with the expected size were cut out and purified via gel extraction (see III.2.6). After purification and determination of product concentrations, the first-round PCR products were diluted to the same concentration and used as templates for the second rounds of fusion-PCR.

3.2.2.2.2 Full-length sequences of CYP-candidates

In the second rounds of fusion-PCR, two outermost primers including specific sequences for restriction enzymes were used to generate the full-length sequences. Distinct exon fragments were fused into the elongated sequences based on the overlapping regions of the respective internal primers.



Sample	Annealing temperature (°C)
2227-1	55.5
2227-2	58.1
38991-1	51.2
38991-2	52.7
38991-3	54
74047-1	51.2
74047-2	52.7
74047-3	54

Sample	Annealing temperature (°C)
2114-1	55.5
2114-2	58.1
3458-1	52.7
3458-1	54
5627-1	52.7
5627-2	54

Figure 43: The second-round fusion-PCR of DOP6H- and DOP7H-candidates
Full-length sequences of candidates are marked with boxes.

3.3 Heterologous expression of CYP-candidate proteins in *Saccharomyces cerevisiae*

3.3.1 Expression of CYP-candidate proteins with His-tag

The yeast clones carrying the complete ORFs in an inducible vector (pYES2/NT C) were used for expression to elucidate the biochemical function of the encoded proteins. There are five different systems available for the expression of eukaryotic cytochrome P450, namely *E. coli*, insect cells, mammalian cells, *Physcomitrella patens* and yeast. Heterologous expression of eukaryotic cytochrome P450 basically involves several problems. First, they are associated with the membrane, and active protein also requires the formation of heme, which must then be non-covalently bound in the polypeptide. Furthermore, CYP enzymes require electron transfer by NADPH:cytochrome P450 reductase (CPR). The CPR is an integral membrane protein and transports the necessary electrons from NADPH via FAD and FMN to CYP. Since *E. coli* does not have its own reductase, a construct of CYP/CPR fusion for expression in *E.*

coli is required. The NADPH:cytochrome P450 reductase of the same plant can be additionally added in the enzyme assays after isolating membranes of CPR-transformed *E. coli* (Bak et al., 1998). In some cases, expression with the *Spodoptera frugiperda* baculovirus system may be required. Jennewein et al. (2001) were able to functionally express taxane 13α-hydroxylase from *Taxus brevifolia* only after expression with this system. This expression leads to expression with the same potency as yeast or *E. coli*. The most common expression is in *Saccharomyces cerevisiae* (Urban et al., 1994). This system is particularly useful because the reductase from yeast provides the electrons for the foreign VYP. In this project, *Saccharomyces cerevisiae* strain INVScI containing only the yeast's own reductase was used for the heterologous expression of CYP-candidate genes. The coding sequences for a polyhistidine-tag for purification and detection of heterologous proteins is already included in the *N*- and *C*-terminus of the pYes2/NTC vector. Since the first 17-29 amino acids in the *N*-terminus of P450 enzymes function as a membrane anchor, the His-tag was attached to the *C*-terminus of the proteins in order to avoid negative effects on the function of the membrane anchor.

The yeast strain INVScI was transformed with the CYP-candidate genes in pYes2/NTC using the lithium acetate method (III.3.3). The transformed yeast cells were plated onto SC-U plates and incubated for 3 days at 30 °C. SC-U is a synthetic minimal medium for yeast containing glucose as the sole carbon source without uracil. Yeast cells without the plasmid pYes2/NTC cannot grow on the SC-U medium. Successful transformations were confirmed by colony PCR (III.2.4.4).

Firstly, the expression of candidate-proteins in *Saccharomyces cerevisiae* was undertaken with SCG medium. As negative control, the yeast cells containing the empty vector pYes2/NTC, were expressed as described in III.3.7.2. The microsome preparations of candidates and negative controls were analysed by SDS-PAGE and Western blot. The Western blot showed very faint bands at the expected size range from 57 kDa to 63 kDa (data not shown). In many reports, 5-aminolevulinic acid, a precursor of the porphyrin synthesis pathway, and iron compounds were added to the medium in order to improve the yields of recombinant proteins (Antipov et al., 2009; Dietzsch et al., 2011). 5-Aminolevulinic acid and iron could support the native protoporphyrin IX formation and the heme biosynthesis pathway in yeast. Therefore, an attempt was made to express the candidate proteins in yeast with $SC^{+}G$ medium (SCG medium with the addition of 5-aminolevulinic acid and ammonium iron (II) sulfate). The microsomes from cell pellets of CYP-candidates and the negative control were then analysed by SDS-PAGE (see III.4.6) and visualised by Western-blotting (III.4.7).

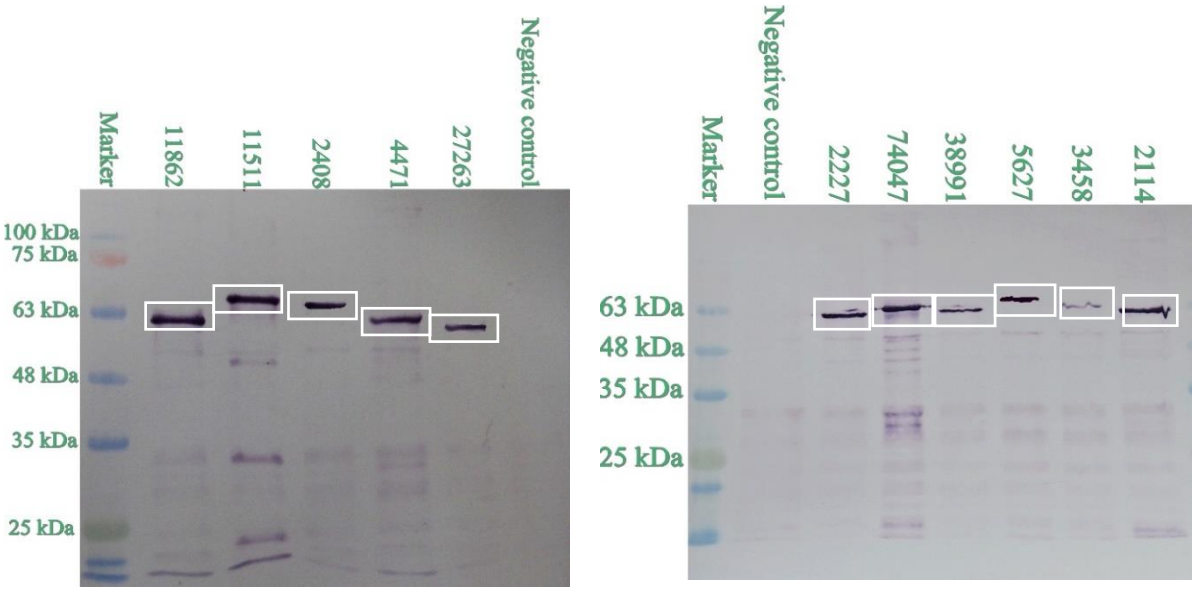
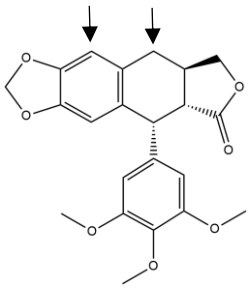


Figure 44: Western Blots of eleven CYP-candidates
Detection was done with anti-His-Tag antibody and secondary antibodies coupled to alkaline phosphatase using the NBT/BCIP color reaction. The molecular weight of the His-Tag is appr. 3.5 kDa. His-tagged protein bands of candidates with appropriate molecular weights are marked with boxes.

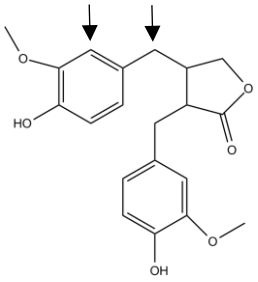
Distinct gene-specific bands of CYP-candidate proteins in the expected size ranges appeared clearly on the membranes of Western blot (Fig. 44). There were no bands in the same size ranges in the lanes of negative control samples. The results of Western blot showed that heterologous proteins with His-tag were produced successfully by the transformed yeast cells after induction with galactose and the use of $SC^{+}G$ medium for the induction yielded better expression of recombinant CYP-proteins.

3.3.2 Enzyme assays with different substrates

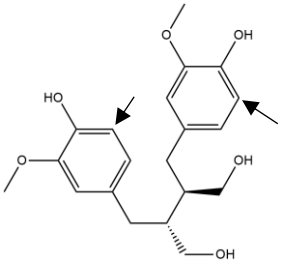
Enzyme assays as described in III.4.8.2 were conducted to test the activity of the CYP-candidate proteins. Besides DOP, different available substrates of the lignan biosynthesis pathway were tested under the same conditions: MATAI, SECO, ß-peltatin, ß-peltatin A methyl ether, yatein, α-peltatin. The substrates tested and the expected position for hydroxylation marked with arrows are listed below:



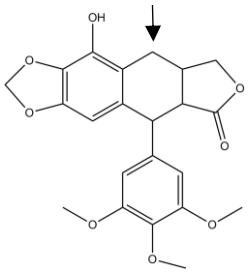
DOP



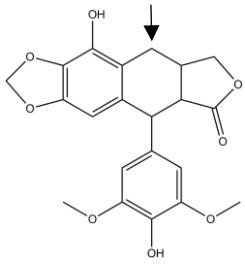
matairesinol



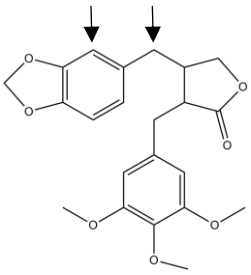
secoisolariciresinol



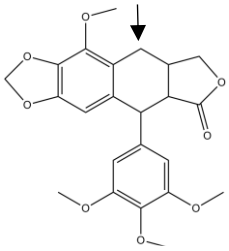
ß-peltatin



α-peltatin



yatein



ß-peltatin A methyl ether

Unfortunately, no product formation could be detected in the HPLC chromatograms and enzyme activity tests with heterologous proteins with His-tag were negative. The ORFs of CYP-candidates most probably are complete and the presence of the heterologously expressed proteins had been confirmed by Western blotting (Fig. 43). Since the microsomes were prepared analogously to the CPR microsomes and the activity of the CPR in the microsomes has been confirmed, a mistake in microsome preparation can be ruled out. In many cases, the His-tag can influence the structure as well as the function of the protein. Hence, the next attempts were made to express candidate proteins without His-tag.

3.3.3 Expression of candidate proteins without His-tag

The *N*- or *C*-terminally attached His-tag may influence the structure as well as the function of proteins. Therefore, CYP-candidate proteins without His-tag were expressed to overcome this possible issue. New full-length primers with a restriction site for XbaI and a stop codon in the reverse primers were designed for PCR-amplification (see IV.6.3). The full-length sequences were amplified with full-length forward primers and new full-length reverse primers by standard PCR-experiments (see III.2.4.1). The successful transformations of the yeast strain INVScI with the constructs of candidate genes in pYes2/NTC were confirmed by colony-PCR (as shown in Fig. 45).

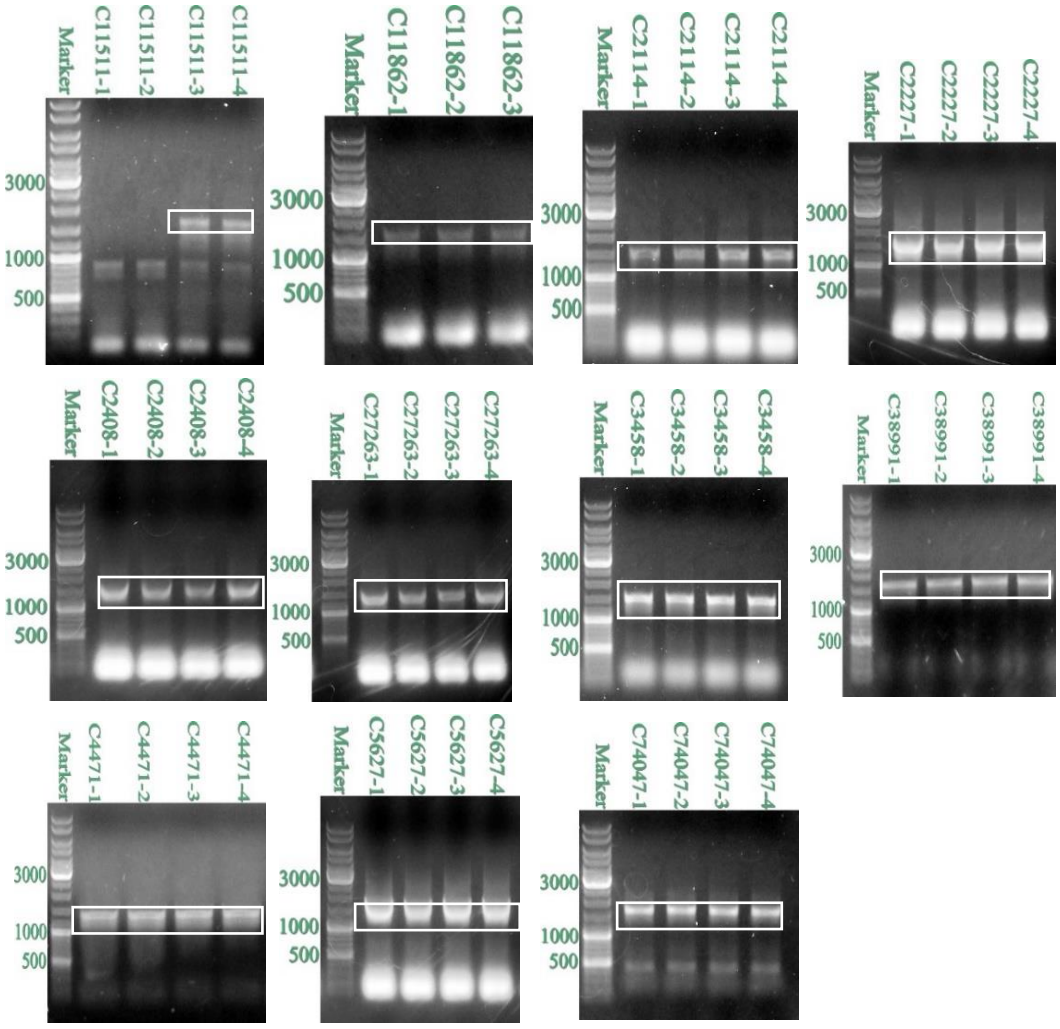


Figure 45: Agarose-gels of colony PCR to show the transformation of yeast INVScI (colony 1-4)
Candidate-genes are marked with boxes.

The transformed yeast cells were cultivated and protein expression performed as described in V.3.7.1. The detection of the heterologously expressed proteins without His-tag by Western

blotting is not possible. It is very difficult to detect the bands of the recombinant CYP-proteins in the microsomes, since yeast cells express also their own CYP enzymes as well as many other proteins of the same size. In many reports, antibodies were designed based on the specific sequences of proteins and Western blots were used to overcome this problem. Regarding the CYP enzymes, the sequence conservation is relatively low within the family, although the general topography and structural fold of CYP are highly conserved. The sequence of proteins from different superfamilies share less than 20% identity (Graham et al., 1999). Only the glutamic acid and the arginine of the ExxR-motif and the heme-binding cysteine are fully conserved among all CYP sequences (Sirim et al., 2010). On the other hand, yeasts as the host cells produce own CYPs. Therefore, the development of specific primary antibodies to detect recombinant CYP-proteins would be very challenging.

In this work, the expression of CYP-protein without His-tag could not be proven. Nevertheless, enzyme assays for CYPs (see III.4.8.2) were performed with microsomes of transformed yeast cells and substrates listed in VI.3.3.2. However, again no product formation could be detected in the HPLC chromatograms and enzyme activity tests were negative.

3.4 *In-vivo* biotransformation enzyme assays

In many reports, the heterologously expressed protein did not show enzymatic activity in *in-vitro* enzyme assays, but function normally in *in-vivo* biotransformation enzyme assays (Kuo et al., 2014). Therefore, in the next attempt *in-vivo* enzyme assays were conducted to test the catalytic activity of recombinant CYP-candidate proteins.

The recombinant proteins of CYP-candidates were expressed and tested in *in vivo* enzyme assays with different substrates (MATAI, SECO, ß-peltatin, yatein, α-peltatin) as described in V.4.8.5. However, no enzymatic activity of the candidate proteins was observed in HPLC chromatograms.

3.5 Concluding remarks

Many potential candidates of cytochrome P450 were found in the transcriptome of *Linum flavum.* Eleven ORFs of CYP were successfully amplified from RNA/cDNA or gDNA of *L. flavum.* Many plant CYPs have been successfully expressed in yeast and showed enzymatic activity in *in vitro* enzyme assays, for example CYP98A14 of *Coleus blumei* (Eberle et al.,

2009) or CYP719A23 and CYP719A24 of *Podophyllum hexandrum* (Marques et al., 2013). However, all attempts to test the activities and possible reactions for the recombinant CYP-candidates with and without His-tag failed.

One reason could be the limited number of lignan compounds for the enzyme assays. Cytochrome P450s of plants are considered to be very substrate-specific unlike cytochrome P450s of animals (Werck-Reichhart et al., 2002). However, only six substrates (MATAI, SECO, ß-peltatin, ß-peltatin A methyl ether, yatein, α-peltatin) were available for enzyme assays. Another aspect for consideration is the extraordinary diversity of CYPs. The first cytochrome P450 was discovered in liver microsomes of rats in 1958 (Klingenberg, 1958). In 1990, it was possible to clone the first plant cytochrome P450 from avocado (Bozak et al., 1990). Since then, a variety of cytochrome P450s have been isolated from plants. There are 273 cytochrome P450 genes in the complete genome of *Arabidopsis thaliana*, which makes P450s one of the largest families of catalytically active proteins in plants. More than 1% of each plant genome is cytochrome P450s (Mizutani, 2012). CYPs are involved in the biosynthesis of various compounds, such as the formation of pigments (anthocyanidins), plant hormones (gibberellins), phenylpropanoids, fatty acids, phytoalexins (terpenes, furanocoumarins, glucosinolates), cyanogenic glycosides, signalling compounds (salicylic acid, jasmonic acid) or alkaloids. The formed substances play an important role in the plant in the pathogen defence, UV protection, membrane formation and pigmentation. From the transcriptome of *L. flavum*, appr. 400 contigs were annotated as CYP by the Blast2GO® program after searching and comparing them with the CYP-sequences in the NCBI database.

Another reason could be the electron supplier system for CYP, since it is known from other plants that the yeast's own reductase and cytochrome b5 may not be sufficient to transfer electrons for plant cytochrome P450s. For example, flavonoid 3',5'-hydroxylase from anthocyanin biosynthesis also requires cytochrome b5 as a cofactor (De Vetten et al., 1999). The similarities of the yeast and plant CPRs may not be sufficient and the homology of the plant with yeast-specific cytochrome b5 is also often insufficient for functional expression (Urban et al., 1990). Furthermore, the majority of newly synthesised plant proteins requires post-translational modification that will finish their tertiary and quaternary structures in order to become functional proteins. The post-translational modification could be vital for protein activity, including function, dynamic interactions with other molecules, localisation and stability (Webster and Thomas, 2012). The CYP-candidate proteins of *L. flavum* were

successfully heterologously expressed in yeast, but the proteins possibly need to undergo some additional modifications and the post-translational modification system of yeast may not be sufficient for plant CYPs.

Furthermore, Shiraishi et al. (2016) suggested that the podophyllotoxin biosynthetic pathway in *Linum flavum* and *Podophyllum hexandrum* are different and originated from convergent evolution. Many enzymes involved in the podophyllotoxin metabolic pathway of *Podophyllum hexandrum* were not found in the transcriptome of *L. flavum*. For example, CYP719A23, which converts matairesinol to pluviatolide was highly expressed in *P. hexandrum* (FPKM (Fragments Per Kilobase Million) > 34) but no CYP719A23-related specific contig with FPKM more than 10 were found in the transcriptome of *L. flavum* (Shiraishi et al., 2016). The feeding experiments by Kamil and Dewick in *Podophyllum* spec. (Kamil and Dewick, 1986) reported a turnover of yatein to podophyllotoxin. Furthermore, an enzyme catalysing the formation of DOP from yatein was successfully identified in *P. hexandrum* (Lau and Sattely, 2015). Nevertheless, it was not possible to detect an enzymatic conversion of yatein to DOP in *Linum album, L. flavum and L. nodiflorum* (Kuhlmann, 2004). Therefore, using CYP71BE54 and CYP82D61 of *Podophyllum hexandrum* to search for the CYP-candidates in *L. flavum* could lead to the wrong candidates. And although eleven candidates were already tested, there are still a large number of other candidates that are likely to be missed during the search.

3.6 Outlook

The expression could be optimized by simultaneous expression of cytochrome P450 and CPR from *L. flavum*. Ro et al. (2002) used the dual yeast expression vector pESC-Leu (Stratagene) with galactose induction for the concomitant expression of C4H and CPR in *Saccharomyces cerevisiae*. In case that the simultaneous expression with CPR is not successful, the co-expression with cytochrome b5 may be necessary.

If the recombinant proteins need special post-translational modifications that could not be done in yeast, the expression in insect or in plant cells could be a solution.

Furthermore, just a limited number of substrates involved in the lignan biosynthesis pathway has been tested in this work. In the elucidation of the function of recombinant proteins, model substrates derived from the physiological substrate may possibly help (Schalk et al., 1997).

4. Project 3: Identification and characterisation of pinoresinol-lariciresinol reductase

4.1 Pinoresinol-lariciresinol reductase (PLR) candidates

The protein sequence of PLR from *Linum usitatissimum* (von Heimendahl et al., 2005) was used as a reference sequence to search for PLR-candidates in the transcriptome database of *Linum flavum* with help of Blastx.

One candidate with convincing similarity was found:

Contig	Score	Query cover	E-value	Identity
10318	544	93%	6e-131	83%

The open reading frame (ORF) of PLR-candidate 10318 was determined by ORF Finder.

The ORF of PLR-candidate 10318 is as below:

ATGGGTTCCCTGGGGAAAGTGATGAATAATGAGATTCCGATTAAGAGCAGCAAAGTGCTGGTGATAGGGGGAACT
GGCTACTTGGGGAAGAGGCTGGTGAAGGCTAGTTTGGATTCGGGCCACGACACCTACGTCATGCATCGCCCGGAG
ATTGGCGTCGACATCGAGAAAGTTCAGCTGCTGCTCTCCTTCAAGATGCAAGGCGCTCATCTCGTCTCCGCCTCC
TTCGACGACCAGCGCAGCCTCGTCGATGCTGTCAAGTTGGTCGACGTCGTCATCTGCGCCATCTCCGGGGTTCAC
ATCCGCAGCCACCAGATCCTTCTTCAGCTCAAGCTCGTCGAAGCCATCAAAGAAGCCGGTAACGTCAAGAGGTTT
GTACCGTCGGAGTTCGGAACAGATCCGGCGAGGATGGAAAACGCGATGGAGCCAGGAAGGATCACATTCGACGAC
AAAATGGTGGTGAGGAGAGCGATAGAGGAAGCTGGGATCCCTTTCACTTACGTCTCTGCTAATTGCTTCGCTGGC
TACTTCCTTGGCGGTCTTTGCCAACCAGGCTACATTCTTCCTTCTAGAGACCATGTTACTTTGCTTGGAGATGGC
GACAAAAAGGGGGTGTACGTGGACGAGGATGATATAGCCGCTTACACATTGAGGGCCATAGACGATCCTCGAACC
CTCAACAAGACGATCTACGTGAAGCCACCTAAGAACGTGTTGTCCCAAAGAGAAGTTGTTGGGATTTGGGAGAAA
TATATCGGCAAAGAGCTCCAAAAGACCATTCTATCCGAGCAAGACTTCCTCGCTACTATGAGAGAAGCAAATTAT
GCACAGCAAGTTGGACTGACGCACTATTATCACGTGTGTTACGAGGGGTGTCTGTCGAATTTCGAGGTTGATGAT
GAGCAGGAAGCCTCCAAGCTCTACCCTGATGTTCACTACACCACCGTCG**AGGAATATCTAAAGCGTTACATCTAG**

The sequences used to design full-length primers are written in bold. For integration into the expression vector pET15b restriction sites for NdeI in the forward primer and XhoI in the reverse primer were added.

4.2 Amplification of PLR-candidate 10318 from cDNA and sequencing

PLR-candidate 10318 was PCR-amplified (see III.2.4.1) by using full-length primers (see IV.6.4) with cDNA as template; the annealing temperature was 54.5 °C. All lanes in Fig. 45 represent PCR products which were formed under the same conditions.

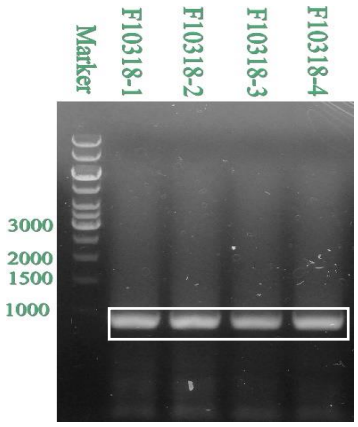


Figure 46: PCR amplification of PLR-candidate 10318 from *Linum flavum* cDNA
PCR-products at appr. 1000 bp are marked with a box.

The PCR products of the appropriate length were excised from the gel (see III.2.6), isolated (see III.2.6), ligated into pDrive vectors (see V.2.7.1) and used for transformation of competent *E. coli* cells (EZ from Qiagen) by heat shock (III.3.2). Plasmid isolation from transformed bacteria and sequencing verified the sequence of a PLR.

The ORF of PLR-candidate 10318 comprises 975 bp and shows 99.8% identity to the sequence of contig 10318 in transcriptome of *L. flavum*. The ORF of CYP-candidate 10318 is shown below:

ATGGGTTCCCTGGGGAAAGTGATGAATAATGAGATTCCGATTAAGAGCAGCAAAGTGCTGGTGATAGGGGGAACT
GGCTACTTGGGGAAGAGGCTGGTGAAGGCTAGTTTGGATTCGGGCCACGACACCTACGTCATGCATCGCCCGGAG
ATTGGCGTCGACATCGAGAAAGTTCAGCTGCTGCTCTCCTTCAAGATGCAAGGCGCTCATCTCGTCTCCGCCTCC
TTCGACGACCAGCGCAGCCTCGTCGATGCTGTCAAGTTGGTCGACGTCGTCATCTGCGCCATCTCCGGGGTTCAC
ATCCGCAGCCACCAGGTCCTTCTTCAGCTCAAGCTCGTCGAAGCCATCAAAGAAGCCGGTAACGTCAAGAGGTTT
GTACCGTCGGAGTTCGGAACAGATCCGGCGAGGATGGAAAACGCGATGGAGCCAGGAAGGATCACATTCGACGAC
AAAATGGTGGTGAGGAGAGCGATAGAGGAAGCTGGGATCCCTTTCACTTACGTCTCTGCTAATTGCTTCGCTGGC
TACTTCCTTGGCGGTCTTTGCCAACCAGGCTACATTCTTCCTTCTAGAGACCATGTTACTTTGCTTGGAGATGGC
GACAAAAAGGGGGTGTACGTGGACGAGGATGATATAGCCGCTTACACATTGAGGGCCATAGACGATCCTCGAACC
CTCAACAAGACGATCTACGTGAAGCCACCTAAGAACGTGTTGTCCCAAAGAGAAGTTGTTGGGATTTGGGAGAAA
TATATCGGCAAAGAGCTCCAAAAGACCATTCTATCCGAGCAAGACTTCCTCGCTACTATGAGAGAAGCAAATTAT
GCACAGCAAGTTGGACTGACGCACTATTATCACGTGTGTTACGAGGGGTGTCTGTCGAATTTCGAGGTTGATGAT
GAGCAGGAAGCCTCCAAGCTCTACCCTGATGCTCACTACACCACCGTCGAGGAATATCTAAAGCGTTACATC**TAG**

The start codon and the stop codon are written in bold letters. The amino acid sequence of the PLR-candidate was translated from the cDNA sequence and the molecular weight of the protein was calculated.

The amino acid sequence of PLR-candidate 10318 with a molecular weight of appr. 36.3 kDa is shown below:

MGSLGKVMNNEIPIKSSKVLVIGGTGYLGKRLVKASLDSGHDTYVMHRPEIGVDIEKVQLLLSFKMQG
AHLVSASFDDQRSLVDAVKLVDVVICAISGVHIRSHQVLLQLKLVEAIKEAGNVKRFVPSEFGTDPAR
MENAMEPGRITFDDKMVVRRAIEEAGIPFTYVSANCFAGYFLGGLCQPGYILPSRDHVTLLGDGDKKG
VYVDEDDIAAYTLRAIDDPRTLNKTIYVKPPKNVLSQREVVGIWEKYIGKELQKTILSEQDFLATMRE
ANYAQQVGLTHYYHVCYEGCLSNFEVDDEQEASKLYPDAHYTTVEEYLKRYI

The obtained cDNA sequence of PLR-candidate 10318 shows high similarity to known PLR-sequences from *Linum corymbulosum* (Accession No. B5KRH5.1) (86.4 %), *Linum usitatissimum* (Accession No. AJ849359) (78 %) and *Forsythia intermedia* (Accession No. U8158) (70.3 %).

4.3 Heterologous expression of PLR candidate 10318 in *E. coli*

After ligation of the full-length sequences of 10318 into the NdeI and XhoI restriction sites of the pET15b vector and transformation of *E. coli* SoluBL21 cells by heat shock the sequence was verified.

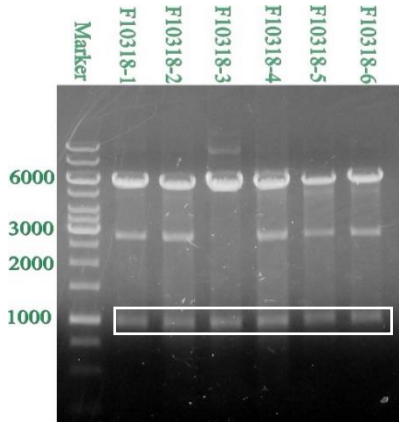


Figure 47: Agarose gel to verify the full-length sequence of the PLR-candidate 10318 in pET-15b (colony 1-6) Plasmids were isolated from *E. coli* SoluBL21 and digested by the restriction enzymes NdeI and XhoI. Candidate bands (appr. 1000 bp) are marked with a box.

Heterologous expression was performed (see III.3.7.1) and the His-tagged protein was purified with a nickel-NTA column (see III.4.3). All fractions of the purification procedure were separated by SDS-PAGE and subjected to Western blotting (see III.4.7). The expressed protein was detected with anti-His-Tag antibody. The molecular weight including His-tag of PLR-candidate 10318 is appr. 38 kDa.

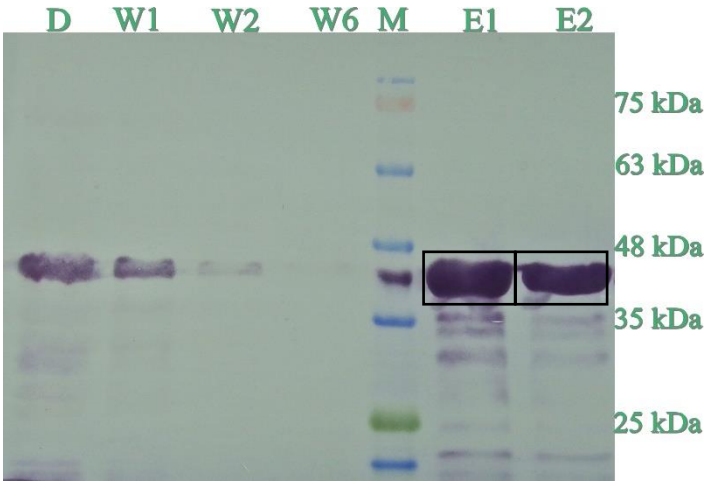


Figure 48: Western Blot of PLR-candidate 10318
M: marker; D: flow through fraction; W1: wash fraction 1; W2: wash fraction 2; W6: wash fraction 6; E1: elution fraction 1; E2: elution fraction 2. His-tagged protein bands of PLR-candidate 10318 with appropriate molecular weight in the elution fractions are marked with boxes.

The result of the Western blot showed that the transformed *E. coli* SoluBL21 cells produced a considerable amount of heterologous protein after induction with IPTG and the purification of the heterologously expressed protein was successful.

4.4 Functional identification of PLR-candidate 10318

The elution fraction after metal chelate chromatography desalted with PD-10 columns (see III.4.4) was used for enzyme assays to determine its catalytic activity. The concentration of the eluted protein was 0.7 mg/ml. The enzyme assay was performed as described in III.4.8.3. After initiation by addition of NADPH, the assays were incubated for different time intervals (0-30 minutes) and extracted twice with 600 µl EtOAc. The products after evaporation of the solvent were dissolved in 100 µl methanol and used for HPLC analysis. HPLC analysis was carried out as described in V.4.9. The eluting compounds were detected at a wavelength of 280 nm. PINO and SECO showed as references significant peaks at 13.6 and 6.3 minutes, respectively (Fig. 49).

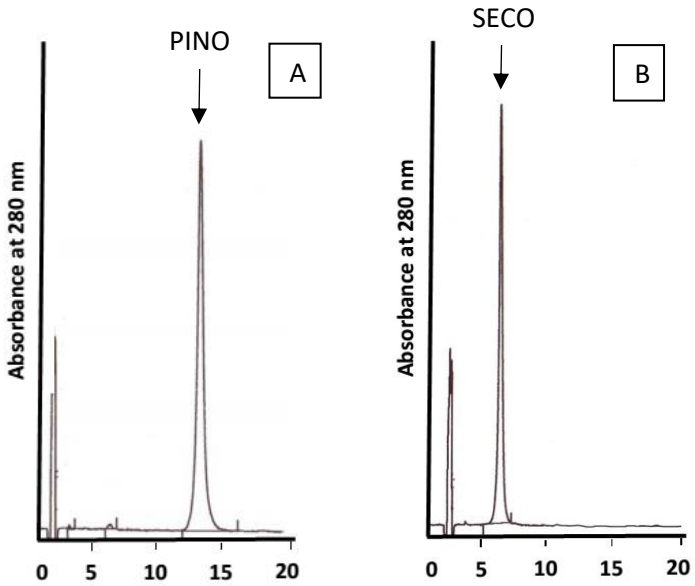


Figure 49: Chromatograms of standards: racemic pinoresinol (A) and secoisolariciresinol (B)
PINO and SECO showed significant peaks at 13.6 and 6.3 minutes, respectively.

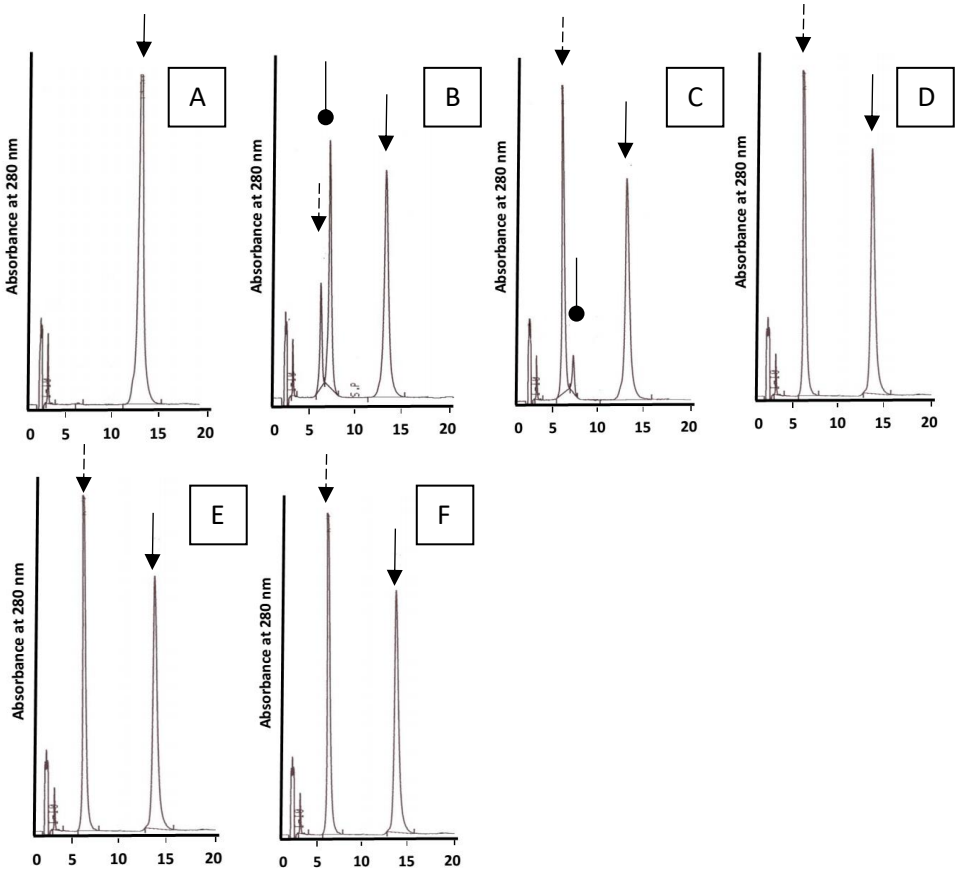


Figure 50: Chromatograms of PLR enzyme assays with different reaction times
Racemic PINO and NADPH were incubated with the heterologously expressed LfPLR. The reaction time is A: 0 min; B: 1 min; C: 2 min; D: 5 min; E: 10 min, F: 30 min. Peaks of PINO, SECO and LARI on the chromatograms are marked with undashed arrows, dashed arrows, and arrows with a round head, respectively.

The reaction product SECO and the remaining PINO were identified by using authentic standards for comparison of retention times. The chromatogram of the enzyme assay with a reaction time of 1 minute (Fig. 50B) showed a high peak for an intermediate and a smaller peak at 6.3 minutes for SECO. After 2 minutes reaction time, the peak for SECO increased and the peak for the intermediate decreased. According to Heimendahl et al. (2005), the intermediate could be identified as LARI. Unfortunately, no standard was available for LARI. Racemic PINO was used in enzyme assays and only one enantiomer of PINO was converted into SECO. Therefore, from 5 to 30 minutes reaction time (Fig. 50), there were two peaks in the chromatograms, one peak for SECO and another one for the remaining enantiomer of PINO.

No activity was detected either in the absence of substrate and cofactor or when the protein was denatured by boiling. Extracts from cells containing the expression vector lacking an insert did not show any catalytic activity as well. The HPLC-results showed that PLR-candidate 10318 was highly active and catalysed the conversion of PINO via an intermediate (presumably LARI) into SECO in the dependence of NADPH.

Furthermore, HPLC with a chiral column (Chiralcel OD-H) was used to analyse the stereospecificity of the enzymatic reaction. The (+)- and (-)-enantiomer of SECO and PINO were identified by comparing with the chiral HPLC-chromatograms of SECO and PINO published by Heimendahl et al. (2005).

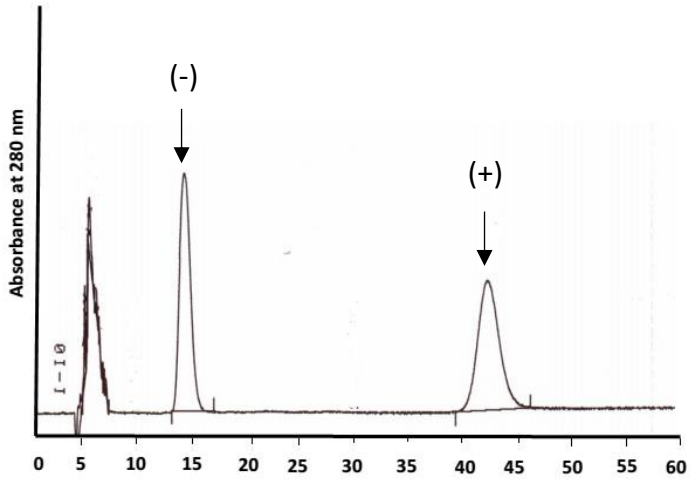


Figure 51: Chromatogram of racemic PINO as standard
The arrows indicate the peaks for (-)-PINO and (+)-PINO at 14.4 and 42.5 minutes, respectively.

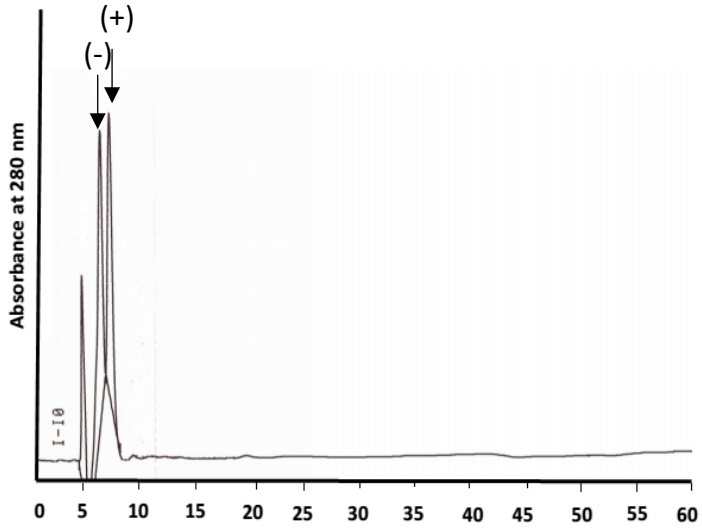


Figure 52: Chromatogram of racemic SECO as standard
The arrows indicate the peaks for (-)-SECO and (+)-SECO at 6.7 and 7.6 minutes, respectively.

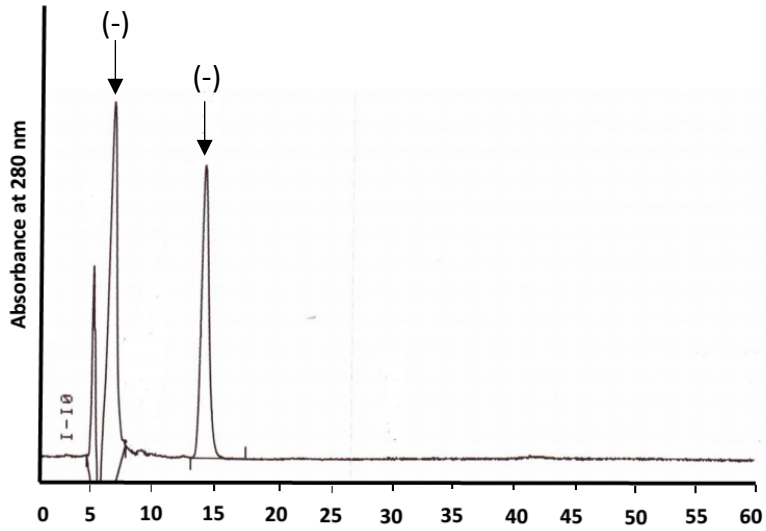


Figure 53: Chromatogram of a PLR enzyme assay (30 minutes)
The arrows indicate the peaks for (-)-SECO and (-)-PINO at 6.7 and 14.4 minutes, respectively

The chiral HPLC showed that the formed product was (-)-SECO and the remaining substrate was (-)-PINO. The reaction from PINO to SECO catalysed by PLR-candidate 10318 is stereospecific, namely only (+)-PINO is used to form (-)-SECO.

4.5 Characterisation of PLR

4.5.1 Time course experiment

A time course experiment was performed with different amounts of purified PLR. The aim was to determine the optimal protein concentration and incubation time for linear product formation.

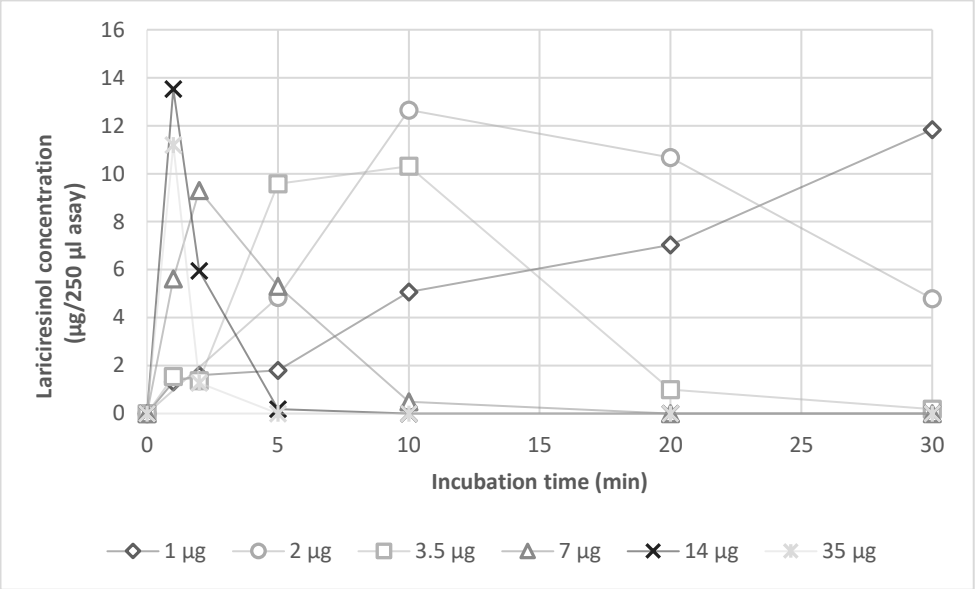


Figure 54: Formation of lariciresinol
Time-dependent formation of LARI from racemic PINO with different amounts of recombinant purified PLR (1 µg, 2 µg, 3.5 µg, 7 µg, 14 µg, 35 µg)

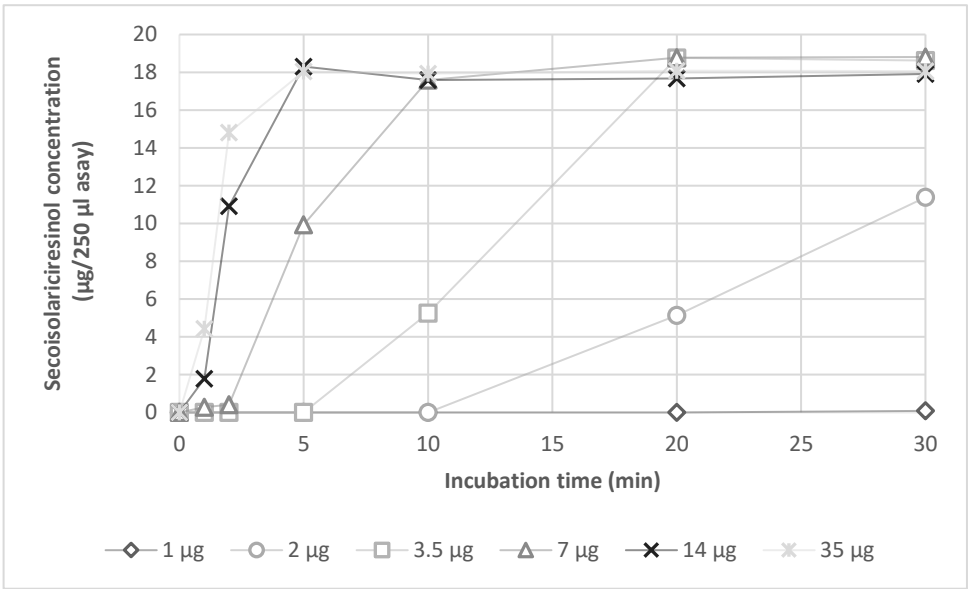


Figure 55: Formation of SECO
Time-dependent formation of SECO from racemic PINO with different amounts of recombinant purified PLR (1 µg, 2 µg, 3.5 µg, 7 µg, 14 µg, 35 µg)

Protein concentrations as low as 1 µg/250 µl assay only formed LARI within 30 minutes. With higher protein concentrations LARI concentrations reached a maximum within 10 minutes. Afterwards, the concentration of SECO increased, while the amount of LARI decreased.

4.5.2 Optimal temperature

The temperature optimum was determined by incubation of enzyme assays at various temperatures between 21 °C and 42 °C.

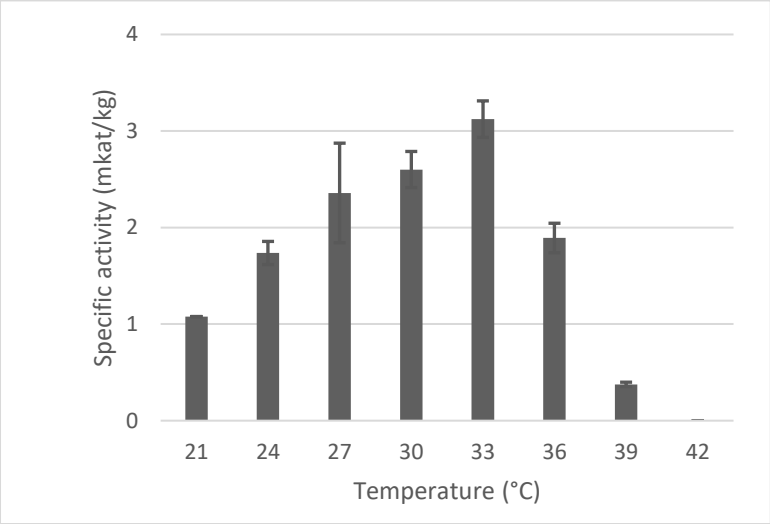


Figure 56: Temperature optimum of PLR
Dependence of the specific PLR activity on the incubation temperature. The data represent mean values of three replicate assays (±s.d.).

Fig. 55 shows an increase of specific activity from 21 °C to a maximum of about 3 mkat/kg at 27-33 °C. At higher temperatures, the specific activity decreases continuously.

4.5.3 pH-optimum of PLR

In order to determine the exact impact of the pH during the reaction, a series of enzyme assays at different pH values was performed. The pH values of the 0.1 mM KPi buffer systems ranged between 5.5 and 9 in steps of 0.5 pH units

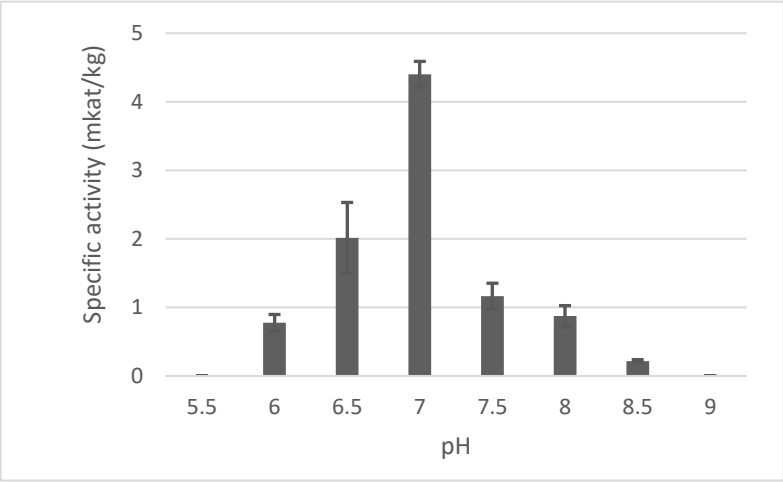


Figure 57: pH-optimum of PLR
Dependence of the specific PLR activity on the pH of the 0.1 M KPi-buffer. The data represent mean values of three replicate assays (±s.d.).

In Fig. 57, PLR of *L. flavum* showed first activity at a pH of 6 and reached a maximum of about 4.4 mkat/kg at pH 7. Under more alkaline conditions the activity dropped significantly and no activity was observed at pH 9.

4.5.4 K_m-value for NADPH

Kinetic experiments were performed at a fixed concentration of purified PLR and racemic PINO with varying NADPH concentrations from 0 to 750 µM. PLR displayed a K_m-value for NADPH of 22.4 ± 1.2 µM and a V_{max} of 4.9 mkat/kg (Fig. 59).

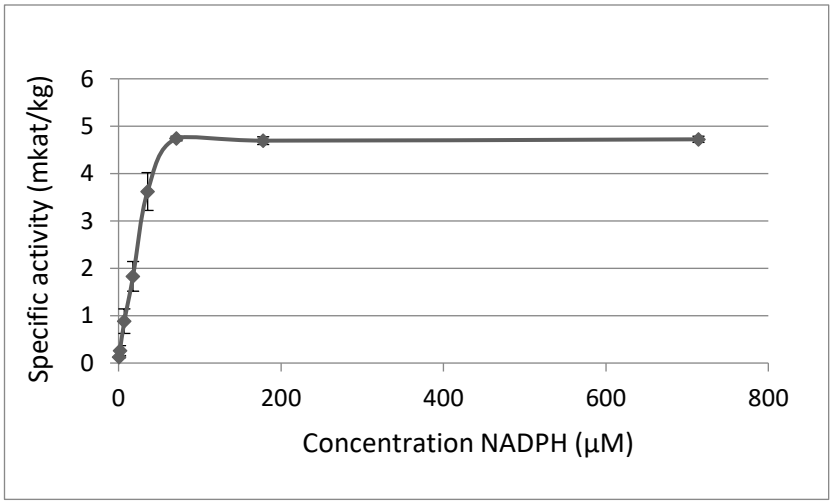


Figure 58: Dependence of the specific PLR activity on the NADPH concentration
The data represent mean values of three replicate assays (±s.d.).

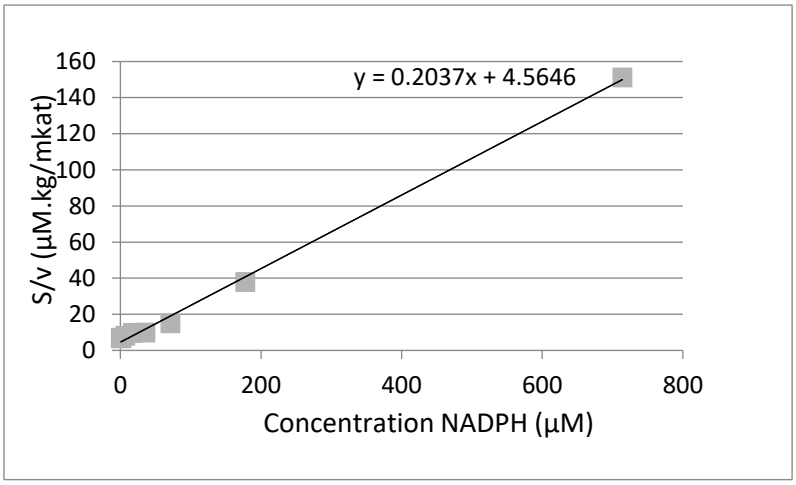


Figure 59: Linearisation of the data from Fig. 58 according to Hanes-Woolf

4.6 Comparison of PLR sequences from different plants

Amino acid sequence alignment was conducted between PLRs with specificity to form (-)-SECO from *Linum flavum* (PLR-Lf), *Thuja plicata* (PLR-Tp2, accession No. AF242504), *Linum perenne* (PLR-Pp, accession No. ABM68630.1), *Forsythia intermedia* (PLR-Fi, accession No. U81158), *Linum album* (PLR-La, accession No. AJ849358), and PLRs with specificity to form (+)-SECO from *Thuja plicata* (PLR-Tp1, accession No. AF242503), *Linum usitatissimum* (PLR-Lu1, accession No. AJ849359) with Clustal Omega (Fig. 60). The conserved sequence motif ‘‘GxxGxxG’’ plays a crucial role in the correct positioning of NAD(P)H (Dinkova-Kostova et al., 1996). More importantly, a lysine residue (Lys^{138} in PLR-Tp1 from *Thuja plicata*) is conserved among all PLRs and is believed to be involved in catalysis as a general base (Min et al., 2003).

```
PLR-Lf      MGSLGKVMNNEIPIK---SSKVLVIGGTGYLGKRLVKASLDSGHDTYVMHRPEIGVDIEK 57
PLR-Tp2     ---------------MEESSRILVVGGTGYIGRRIVKASIALGHPTFILFRKEVVSDVEK 45
PLR-Lp      ----------------MKPCSVLVVGGTGYIGKRIVSASLYLGHDTYVLKRPGTGLDIEK 44
PLR-Fi      ----------------MGKSKVLIIGGTGYLGRRLVKASLAQGHETYILHRPEIGVDIDK 44
PLR-La      MGSLGKVN-NEIPTKSSGGSKVLVIGGTGYLGKRLVKASLDSGHDTYVMHRPEIGVDIEK 59
PLR-Tp1     ---------------MDKKSRVLIVGGTGFIGKRIVKASLALGHPTYVLFRPEALSYIDK 45
PLR-Lu1     ----------------MGRCRVLVVGGTGYIGKRIVKASIEHGHDTYVLKRPETGLDIEK 44
                           . :*::****::*:*:*.**:  ** *::: *      ::*

PLR-Lf      VQLLLSFKMQGAHLVSASFDDQRSLVDAVKLVDVVICAISGVHIRSHQVLLQLKLVEAIK 117
PLR-Tp2     VEMLLSFKKNGAKLLEASFDDHESLVDAVKQVDVVISAVAGNHMRH-HILQQLKLVEAIK 104
PLR-Lp      LQLLLSFKKRGAHLVEASFSDHDSLVRAVRLVDVVICTMSGVHFRSHNILLQLKLVEAIK 104
PLR-Fi      VEMLISFKMQGAHLVSGSFKDFNSLVEAVKLVDVVISAISGVHIRSHQILLQLKLVEAIK 104
PLR-La      VQLLLSFKMQGAHLVSASFDDQRSLVDAVKLVDVVICAISGVHIRSHQILLQLKLVEAIK 119
PLR-Tp1     VQMLISFKQLGAKLLEASLDDHQGLVDVVKQVDVVISAVSGGLVRH-HILDQLKLVEAIK 104
PLR-Lu1     FQLLLSFKKQGAHLVEASFSDHESLVRAVKLVDVVICTVSGAHSR--SLLLQLKLVEAIK 102
            .::*:***  **:*:..*:.*  .** .*: *****.:::*   *   :* *********
```

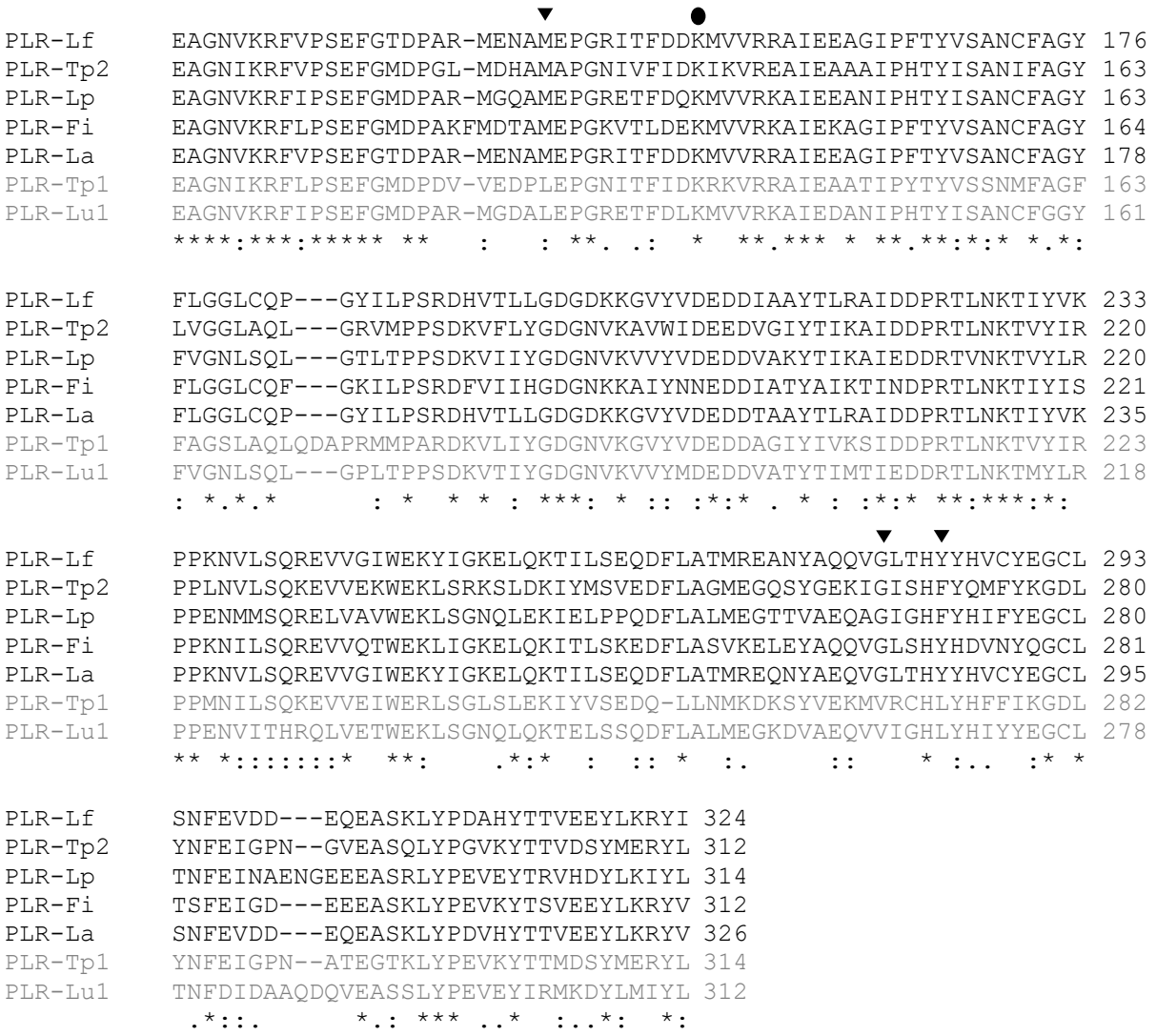


Figure 60: Amino acid sequence alignment of PLRs with specificities to form (-)-SECO (black) and (+)-SECO (grey)

Linum flavum: PLR-Lf, *Thuja plicata*: PLR-Tp2, *Linum perenne*: PLR-Lp, *Forsythia intermedia*: PLR-Fi, *Linum album*: PLR-La, *Thuja plicata*: PLR-Tp1, *Linum usitatissimum*: PLR-Lu1. The conserved sequence motif ''GxxGxxG'' of the NADPH binding domain is marked with a box. Amino acid positions predicted to be involved in stereospecificity by Min et al. (2003) are indicated by triangles and the Lys residue involved in general base catalysis is indicated by a dot (Min et al., 2003).

4.7 Enantiospecific conversion and PLR-mutants

PLRs catalyse the enantiospecific conversions from (+)- and (-)-PINO to (-)- and (+)-SECO, respectively. The protein structure of PLR-Lf was predicted by the bioinformatic tool Phyre[2] (Kelley et al., 2015) and the ligand (+)-PINO docked into the binding pocket of PLR-Lf (Fig. 61). The crystal structure of PLR-Tp1 and the energy-minimized model of PLR-Tp2 obtained by Min et al. (2003) (Fig. 62) and the predicted protein structure of PLR-Lf allowed us to study more closely the structure-function relationship of the enantiospecific enzymatic conversion,

namely PLR-Tp1 with the specificity to utilise (-)-PINO and PLR-Tp2 and PLR-Lf with the specificity to use the opposite enantiomer (+)-PINO. The overall secondary and tertiary structures of PLR-Tp1 and PLR-Tp2 are very similar. The only functionally significant differences are in the substrate and cofactor binding sites. The structures of PLR-Tp1 and PLR-Tp2 show remarkable differences in the local environment of the substrate binding site. In PLR-Tp1, (-)-PINO fits tightly between the side chains of Phe^{164}, Val^{268}, and Leu^{272}, but in PLR-Tp2, favouring the binding of the (+)-enantiomer of PINO, there are Leu^{164}, Gly^{268} and Phe^{272} (Min et al., 2003). The position Phe^{164} in (+)-SECO-forming PLR-Tp1 was found in sequences of many (-)-SECO-forming PLRs as well (e.g Phe^{174} in PLR-Lf or Phe^{165} in PLR-Fi) (see Fig. 60), thus this position might not play a relevant role in determining the enantiospecificity. The position Phe^{272}, an unpolar amino acid, in PLR-Tp2 is not conserved among all (-)-SECO-forming PLRs and can be replaced by tyrosine, a polar amino acid (e.g. Tyr^{284} in PLR-Lf or Tyr^{272} in PLR-Fi). Therefore, not the polarity but the size of amino acid residues in this position might be important for the enantiospecific conversion.

An attempt was made in this work to seek the explanation for the enantiospecific differences of PLRs and to prove whether the varying size of the amino acids in the binding pocket is responsible. Mutagenesis of the coding sequence of PLR from *Linum flavum* was carried out: the small amino acid glycine (G^{280}) was replaced by tyrosine (mutant G280Y) and the amino acid tyrosine (Y^{284}) was changed to glycine (mutant Y284G).

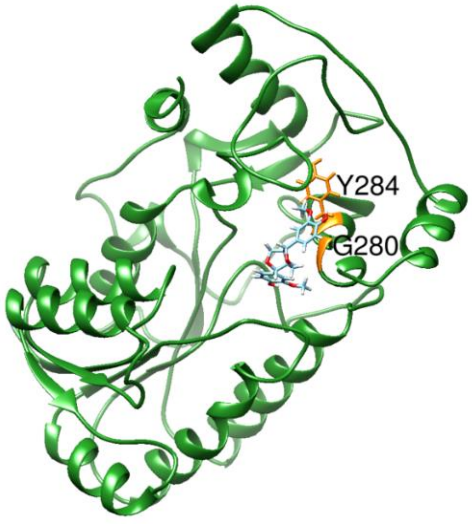


Figure 61: Schematic representations of the predicted protein model of PLR-Lf from *Linum flavum*
The position of the substrate (+)-PINO in the binding pocket and the potential key residues and surfaces that form substrate contact points are highlighted.

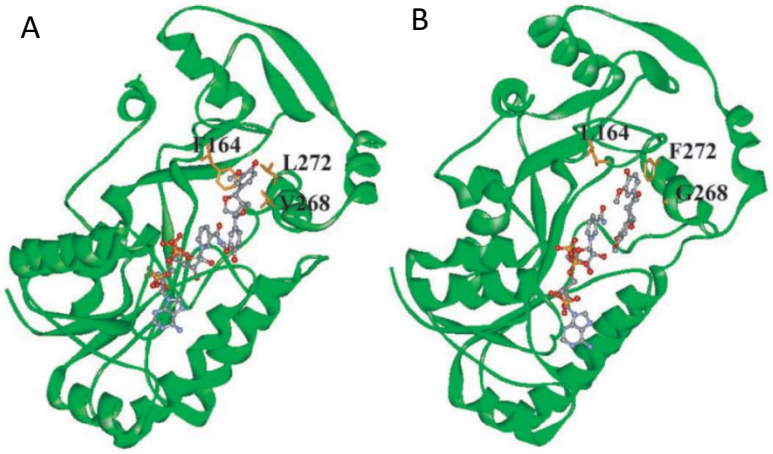


Figure 62: Schematic representation of the crystal structures of PLRs from *Thuja plicata* (Min et al., 2003)
A: Schematic representation of the crystal structure of PLR-Tp1; B: energy minimized model of PLR-Tp2. NADPH, PINO and the potential key amino acids are highlighted.

4.7.1 LfPLR G280Y

In this mutant, the amino acid glycine was replaced by the polar and aromatic amino acid tyrosine (G280Y) using site-directed mutagenesis (see III.2.9). The mutated PLR-Lf inserted into pET15b was used to transform *E. coli* SoluBL21, express and purify the mutant protein. SDS-PAGE (Fig. 63) showed the successful expression of LfPLR-G280Y.

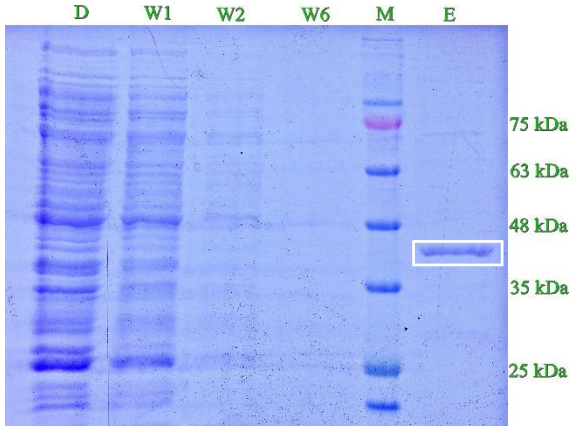


Figure 63: SDS-PAGE of heterologously expressed PLR-Lf mutant G280Y purified by metal chelate chromatography
M: marker; D: flow through fraction; W1: wash fraction 1; W2: wash fraction 2; W6: wash fraction 6; E: elution fraction. The His-tagged protein band of LfPLR-G280Y in the elution fractions is marked with a box.

In the elution fraction, a band in the expected size-range was detected on the SDS-PAGE gel. The result of SDS-PAGE showed that the transformed *E. coli* SoluBL21 cells produced the heterologous protein after induction with IPTG and that the purification of the His-tagged

protein was successful. Then, PLR enzyme assays were performed with His-tagged purified protein as described in III.4.8.3 in order to test the enzymatic activity of this mutant. After initiation by addition of NADPH, the assays were incubated for 2 hours. The substances formed in the enzyme assays were analysed by HPLC. However, no peak for any product (SECO or LARI) was observed on the HPLC chromatogram. Another attempt was made to test the activity of this mutant with a ten-fold higher concentration of PINO and longer incubation time (24 hours). Nevertheless, again no activity of the mutant LfPLR-G280Y was detected and the catalytic activity of the mutated LfPLR-G280Y was totally abolished.

In order to explain why the mutation from G to Y at the position 280 of wild-type LfPLR led to the loss of the enzyme's function, the single amino acid variants (SAVs) were analysed by Phyre^2 Investigator (Kelley et al., 2015). These predictions are made using the SuSPect method (Yates et al., 2014). The mutational analysis graph (Fig. 64) represents the predicted effect of mutations at position 280 in the wild-type LfPLR sequence. In the graph, the 20 possible amino acids are labelled along the x-axis with their one-letter code. The coloured bars indicate the probability that a mutation to the corresponding residue will have some effect on the function of PLR. In position 280 of the LfPLR sequence, the mutation to Y has the highest likelihood to affect the function of the enzyme, similar as R, C, Q, K or F.

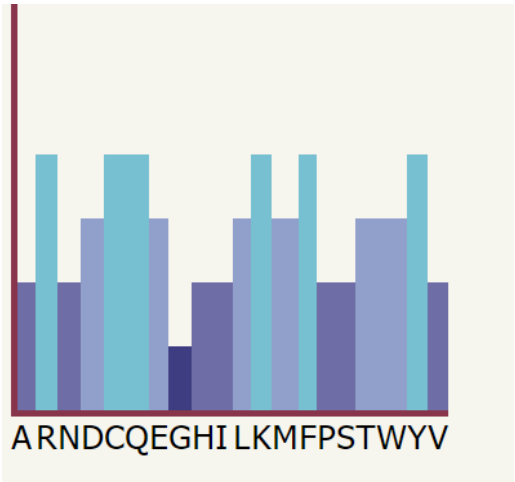


Figure 64: Mutational analysis graph of position 280 of wild-type PLR-Lf performed by Phyre^2 Investigator

Additionally, the protein sequence of LfPLR-G280Y was submitted to Phyre^2 (Kelley et al., 2015) to predict the protein structure. Subsequently, the position of the ligand (+)-PINO in this mutant was predicted by Swiss Dock. The best-ranked position of PINO in the protein structure according to the average full fitness of the elements is shown in Fig. 65 (Grosdidier et al., 2007)

and compared to the wild-type (Fig. 61). When viewing the images depicting the location of (+)-PINO in the binding pockets in both cases, the difference of the positions due to the steric effect of Y280 in the mutant is very clear. Tyr with the phenolic residue is much larger than Gly, thus encroaching on the methoxyphenol position of the ligand. This leads to a big change in the position of the tetrahydrofurofuran ring of the ligand. Hence, in the binding pocket of the mutant, the distance between PINO and the cofactor NADPH could be too long to enable the transfer of the hydride ion from NADPH to the tetrahydrofurofuran ring. That might be the reason why the mutated enzyme is ineffective and cannot catalyse the reaction.

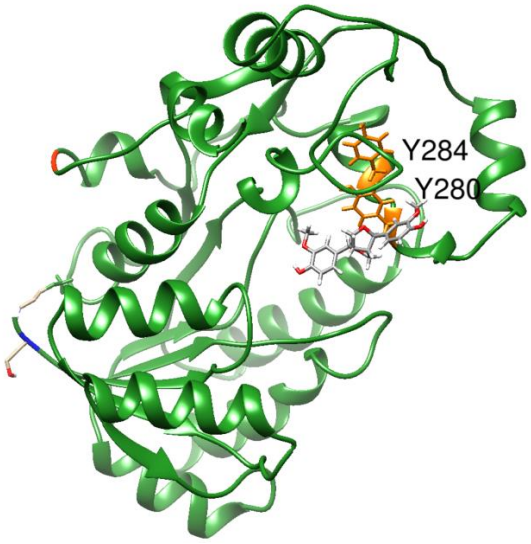


Figure 65: Schematic representations of the predicted protein model of the mutant LfPLR-G280Y
The position of the substrate (+)-PINO in the binding pocket and the potential key residues and surfaces that form substrate contact points are highlighted.

4.7.2 LfPLR-Y284G

In the mutant LfPLR-Y284G, the amino acid tyrosine was replaced by the amino acid glycine (Y284G) using site-directed mutagenesis (see III.2.9). The mutated PLR-Lf inserted into pET15b was used to transform *E. coli* SoluBL21, express and purify the mutated protein. SDS-PAGE (Fig. 66) showed the successful expression of LfPLR-Y284G.

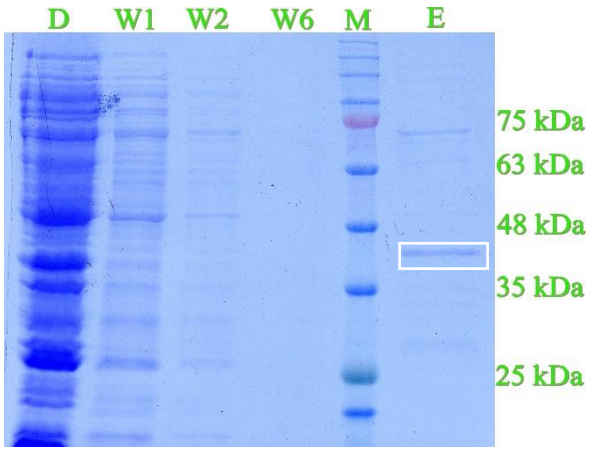


Figure 66: SDS-PAGE of heterologously expressed PLR-Lf mutant Y284G and purification sequence by metal chelate chromatography
M: marker; D: flow through fraction; W1: wash fraction 1; W2: wash fraction 2; W6: wash fraction 6; E: elution fraction. The His-tagged protein band of LfPLR-Y284G in elution fractions is marked with a box.

In the elution fraction, a band in the expected size-range was detected on the SDS-PAGE gel. The result of the SDS-PAGE showed that the transformed *E. coli* SoluBL21 cells produced the heterologous protein after induction with IPTG and that the purification of the His-tagged protein was successful.

Enzyme assays were performed with the His-tagged purified mutant protein as described in III.4.8.3 in order to test the enzymatic activity of this mutant. After initiation by addition of NADPH, the assays were incubated for 2 hours. The enzyme assays were analysed by HPLC. However, no peak for any product (SECO or LARI) was observed on the HPLC chromatogram. Another attempt was made to test the activity of this mutant with a ten-fold higher concentration of PINO and longer incubation time (24 hours). Nevertheless, again no activity of the mutant LfPLR-G280Y was detected and the catalytic activity of the mutated LfPLR-G280Y was totally abolished.

The single amino acid variants (SAVs) at the position 284 of wild-type PLR was analysed by Phyre^2 Investigator (Kelley et al., 2015) in order to elucidate why the mutant LfPLR-Y284G lost its catalytic activity. The mutational analysis graph (Fig. 67) represents the predicted effect of mutations at position 284 in the wild-type PLR sequence. At position 284 of the PLR sequence, it can be seen that the mutation to G has the highest likelihood to affect the function of the enzyme, similar as D.

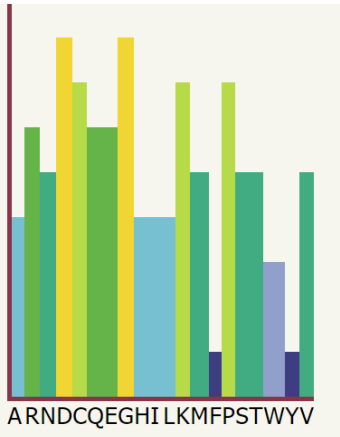


Figure 67: Mutational analysis graph of position 284 of wild-type PLR-Lf performed by Phyre² Investigator

In addition, the protein structure of mutant Y284G and the position of the ligand (+)-PINO were predicted by the same methods used for mutant G280Y. The position of the ligand in the predicted protein structure in the mutant (Fig. 68) was compared to the ligand position in wild-type LfPLR (Fig. 61). When comparing the position of (+)-PINO in the binding pockets in both cases, the difference is obvious. In the binding pockets of wild-type LfPLR, the position of the ligand (+)-PINO is very close to Tyr^{284}, which acts as an anchor point due to the hydrophobic interaction between the Tyr-residue and the methoxyphenol ring of PINO. However, in the binding pockets of mutated LfPLR Y284G, the position of PINO is far from Gly284 and located deeper in the binding pocket. Hence, PINO could push the cofactor NADPH out of the binding pocket and thus cannot receive the hydride ion from the cofactor. That might be the reason why the mutated LfPLR-Y284G lost the catalytic activity.

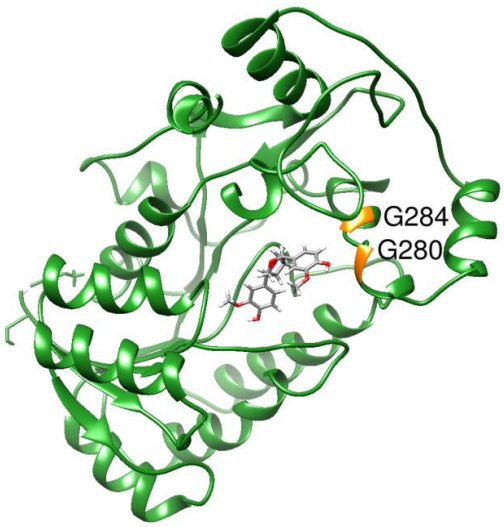


Figure 68: Schematic representations of the predicted protein model of the mutant LfPLR-Y284G
The position of the substrate (+)-PINO in the binding pocket and the potential key residues and surfaces that form substrate contact points are highlighted.

4.8 Concluding remarks

From the transcriptome database of *L. flavum* one contig that potentially encodes the enzyme PLR was found. The ORF of PLR-candidate 10318 was amplified with cDNA from a suspension culture of *L. flavum.* This ORF was heterologously expressed in *E. coli* SoluBL21 and the recombinant protein catalysed the formation of (-)-secoisolariciresinol from (+)-pinoresinol in *in-vitro* enzyme assays. PLR is the first gene encoding an enzyme in lignan biosynthesis successfully identified in *L. flavum.*

Two mutants of PLR-Lf were constructed to investigate the role of the varying size of important amino acids in the binding pocket. The loss of enzymatic activity of mutants LfPLR G280Y and Y284G showed that the size of amino acids at two positions 280 and 284 in the binding pocket are not only important for the enantiospecificity but also crucial for the catalytic activity of the enzyme. Y^{284} and G^{280} might play a vital role in the positioning of ligand in the binding pocket. Further experiments and mutageneses need to be employed to fully understand the steric effect of other amino acids on the stereospecificity of PLR.

5. Project 4: Identification of secoisolariciresinol dehydrogenase

5.1 Secoisolariciresinol dehydrogenase (SDH) candidates

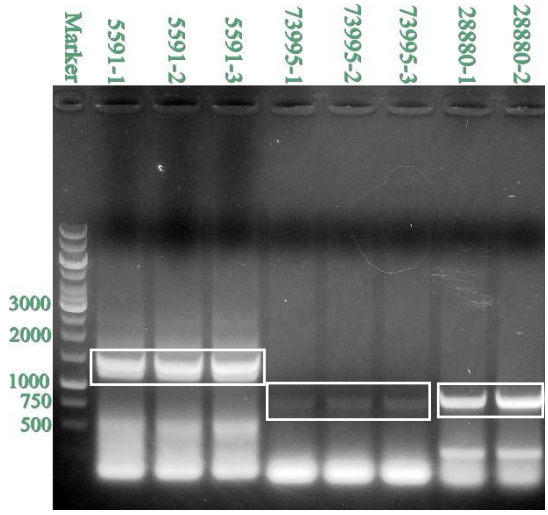
The amino acid sequence of SDH from *Forsythia intermedia* (Xia et al., 2001) was used as a reference gene to search for SDH candidates in the transcriptome database of *Linum flavum* with the help of Blastp.

The five candidates with the best Blastp-results are shown below:

Contig	Score	Query cover	E-value	Identity
36067	393	99%	3e-143	68%
28880	260	90%	5e-91	53%
73995	244	95%	6e-85	51%
5591	239	94%	9e-83	54%
7665	234	90%	5e-80	49%

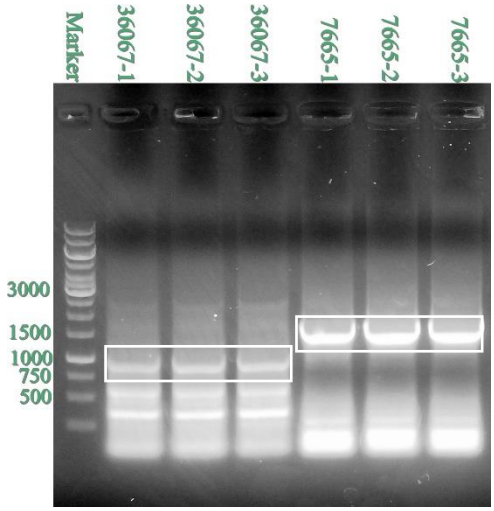
5.2 Amplification of candidates from gDNA and sequencing

The open reading frames (ORF) of the candidates were found in the transcriptome database by ORF Finder and Oligo Calc was used to design the primers (see IV.6.3) based on the sequences (20-30 nucleotides) located at the beginning and the end of the identified ORFs. Many attempts to perform standard PCR reactions (see III.2.4.1) with cDNA as template and the corresponding primers failed. The suitability of the cDNA is dependent on the transcription of the respective gene and therefore highly dependent on external factors, such as cultivation time, medium, etc. The next attempts, therefore, were made with gDNA as a template for standard PCR experiments and the five SDH candidates were successfully amplified.



Sample	Annealing temperature (°C)
5591-1	52.7
5591-2	54
5591-3	55.7
73995-1	54
73995-2	55.7
73995-3	57.2
28880-1	54
28880-2	55.7

Figure 69: PCR of SDH-candidates 5591, 73995 and 28880
PCR-products are marked with boxes.



Sample	Annealing temperature (°C)
36067-1	54
36067-2	55.7
36067-3	57.2
7665-1	54
7665-2	55.7
7665-3	57.2

Figure 70: PCR of SDH-candidates 36067 and 7665
PCR-products are marked with boxes.

The PCR products (Fig. 69 and 70) considerably varied in their sizes from appr. 800 bp to 1600 bp. The PCR products were excised from the agarose gel and isolated from the gels (see III.2.6). The candidate PCR products were ligated into pDrive vectors (see V.2.7.1) and subsequently introduced into competent *E. coli* EZ cells by heat shock (see III.3.2). The transformed bacterial

colonies were selected and grown in overnight cultures at 37° C. The bacterial plasmids were then isolated (see III.3.5) and sent to Seqlab® for verifying the sequences.

The gDNA-sequence of SDH-candidate 28880 comprises 938 bp incuding one intron (underlined) and is shown below:

ATGGGCTCTGATATCTGTGCACCCTCTGTCACCAAGACGTACGTAGGGATTTACACGTTATATTCCACCACCAGT
TAATGTTTTATCGTAATTTACTGGTTATCGATGCATATGTTCATTTTAGTTAGTTGGTTAATACATACTTCTGCA
GGCTAGAAGGCAAAGTGGCGCTCATAACTGGAGGAGCAAGCGGCATAGGCGAGAGCGCTGTCAGGCTGTTCATGC
AACACGGAGCCAAGGCTCTCATCGCCGACCTTCAAGATAAACTCGGCGAATCCCTCGCCACGGAACTCGGTCCAC
CGGAGATTGTCTCCTTCGTCCACTGCAATGTCACTGTAGACTCCGATGTGGGAAACGCTGTGGATACTGCGGTTT
TGAGGTACGGACAGTTGGACATCATGTACAACAACGCTGGAATCGGAGGCAACCTCGACACCACGATTTTAAACT
CCGACAATGACGATTTCAAAAGGGTATTGGAAATAAACCTGTTCGGATCGTTCCCCGGAGCCAAACACGCTGCTA
GAGTGATGATTCCGGCGAGGAAAGGTTGCATACTATTTACGGGAAGCGTGGCGTCGTCAATAAGCGGGGATTTAT
CATACGCTTACAAGGCTTCGAAGCACGCGTTGTTGGGGCTAGCCAATAATTTGACTGTGGAGATGGGGAAGTACG
GGATTAGGGTCAACACTATATCCCCGTATGGAGTGGCCACTCCTATGGTGACACGTGGCCTGCAAATGGATAAGA
AGGCTGTCGAGCAGTTCATGTCGGCAGCGGGGAACCTAAGAGGGGCGATTTTGGAGCCAGAGGATGTTGCCAAGG
CGGCGCTGTACTTGGCTAGCGATGATGCGAAATATGTGAGTGGGCTGAACTTGATTGTTGACGGTGGCCACAATC
ACAACCATCCTCTCTTTGGCGCTTCCGCTGAAGCA**TGA**

The gDNA-sequence of SDH-candidate 36067 comprises 1061 bp incuding one intron (underlined) with 212 bp and is shown below:

ATGGCGGCAGCAGCACCAGCTTCCTTTATCTCCTCCGTCGCCAGAAGGTAACCTCTTACCAGTCCTTTAGTTCTT
GCTTAATTCAATCCTTTCCTCTTTCTCAAATCCTTCCATTTCCAGTCTCTGATCTAGTATACCCTATCACCCCAT
CTCTCTGTTCACTCCCTGGTTTTGTCTTTATTTGATCTGTTCTTGATTGATTCTTCAGTTCGGATATCAACTGAT
CCACCTCGTGTTGATTTCTCTGACATTTCTGCAGGCTGGAAGGCAAAGTAGCGCTAATCACCGGAGGAGCAAGCG
GGATCGGCGAATTCACCGCCAGAGTGTTCGCCCACCACGGCGCCAAAGTCGTGATCGCCGACGTCCAAGACGACC
TTGGCCAATCCGTCGTGGCCGACACCCTGGGCCGCGCCAACTCCACCTACGTCCGCTGCGACGTCACCGAGGAGT
CCCAAATCAAAGACGCCGTCGACAAGGCCGTCTCCACCTACGGCAAGCTCGACATCATGTTCAACAACGCCGGCA
TCGCCGACCCCAACAAGCCCCGCATAGTCGACAACGAGAAGTCCGACTTCGAGCGCGTCCTCAGCGTCAACGTCA
CCGGTGTCTTCCTCGGGATGAAGCACGCGGCGCGGGTGATGATCCCGCGCCGCACGGGGGCCATAATCTCGACCG
CCAGCATCAGCTGCGCCATGGGCGGCGCAGCCTCGCACGCATACTGCTGCTCGAAGCACGCCGTGCTGGGGCTGA
CCCGGAACGCGGCGGTCGAGCTGGGGCAGTTCGGGATCAGGGTGAACTGCCTGGCGCCGTACGCGATGGCGACGC
CTCTGGCGACGAAGTTCGTGGGGTTGCCGGATGAGGAGTTGGAAGGGGCGATGAATGCTCTGGCGAATTTGAAAG
GAGTGACCCTGAAAACAGAGGATGTGGCGAACGCTGCTCTGTTCTTGGCTAGCGACGAAGGTAGGTATGTTAGTG
GGCATAGTTTGGTTGTGGATGGAGGGTTTAGCATTCATAATACTGCGTTTAGGATCTTCAATTACCCGGAAGATG
GATCTTCA**TGA**

The gDNA-sequence of SDH-candidate 5591 comprises 1619 bp incuding two intron (underlined) and is shown below:

ATGGCCGCCACTAACTTCGTTTCTTCCATCATAAAAAGGTTCTCATCACTTTCTTACTAATCTCAACCCTTCACT
ATACTTCACTATGCATTTCCCGATCAAATTTTTCATTAATTCTCAGGCTGGAAGGCAAAGTAGCGCTGATCACCG
GCGGAGCCAGCGGAATCGGCGAGGCCACCGCCAGCCTTTTCGTCCAACACGGCGCCAAAGTCCTAATCGCCGACG
TCCAAGACCAACTCGGCGAATCCTTAGCCCGAAACCTAGGCCCGCCGGACACGGTCTCCTACGTCCGCTGCGATG
TTACCAGTCCTCCCGAAGTCGGGGCCGCCGTGGACGAGGCGGTGTCGAGGTACGGCCAGCTGGACATCATGCACA
ACAACGCCGGCATCGCCGGGGGCCTAGACCAGCGTATCGTTCACTCCGACGACGATGACTTCCGGCGGGTCCTGG
ACGTCAACCTCTTCGGATCCTACCTTGGGGCTAAGCACGCCGCCAGGGTGATGGTAAGTAAATGTTGAGAAGTAG
TTAACAATACTGTAGTTGAAATGGCGTTTTTTATCTTAAAATTTTACGCTTTTATCAAAATAATGTACGTTTTCG

TATAATTTTTTTTACGAACAGATGTAAATTGGATGGATGTTCACCAAAATTGTTCGCTTCTTATCGAAAAAATAT
TAATTTTTTATGTACTCTTTTAAATATAGAGTAAATATAATTATTAAAAGTCTATAATATTATATTTTAAATAAT
ACTTGTTGTTAAAAATATTTCTATCAATTTTATACGATATATAAAATCAAAATTTGTAAAATTATATCCGAATTT
ATTCGGAACAAGTTAGTGGTAAAGTACTATCATGGTTTGAAGACAATGTACCGTTGTTAAAATGATATTAAAATA
AAAATTATATATTCAACCTGTAATGTACCGCAATAATATCATTTAAAACCAAAAAAATCGACTGATATATCGTTG
ATTAATAATTTAACCTTCTTGTACTAATCATATTAAAATTGTCCAACCTTAATATAATACTGTATCATAAGGTAT
CCGTTTTTTACACCTTAACACTTTCATCCATATTTCCATTTATACGATTGCGCTCTTTACTACATTAATAGTAGA
GTTATATGTTCGGTTGAGAAGTGACAGACGACGTAATATATTATATATAACACTTGGTATATACTTTGCCAGGTA
CCGGCGGGGAGAGGCGGGTGCATTCTGTTCACGGCGAGCAGCGCGGCGGTGACGAGCGGGAACATTACGTACGCG
TACAAGGCGTCGAAGAACGGGGTGTTGGGGCTGGCGAACAACCTGTGCGTGGAGCTGGGTCAGTACGGGATCCGG
GTAAACGCGATCTCGCCGTTCACGGTGGTGACTCCGCTGTTGGTGAGTGTGCTGGGAGGCGGCGGGCTGGGGAAG
GGGGAGGTGGAGAGGCTGTGGCGGCGGCGACGAAGCCGAACTTGAAAGGAGCGGTGTTGGAAGCGGCGGATATTG
CCAAGGCGGCGCTGTTTTTGGCGAGCGATGATGCCAAGTACGTGAGTGGGTTGAATTTGGTCGTCGATGGTGGGC
ATAATCACAACCGTCCGATCTTTGATGATTTGAGGTTCGCT**TGA**

The gDNA-sequence of SDH-candidate 73995 comprises 908 bp incuding one intron (underlined) and is shown below:

ATGTCTTCTTCCTACCAGCCGGCGGCGGCGAGTTCTCACTGGTACGGCGAACTGAAACTCTGAAACCCTAAAACC
CTACTCTGTATCGTCTCTGTTCGTTCAGCTTATAAACGATGTCGTTTAATTTGTTTAAGCAGACTAGAAGGCAAG
GTGGCAGTGATTACCGGCGGCGCGAGCGGGATCGGCGCAGCGACGGCCAAGCTCTTCGCCCTCAACGGCGCCAGA
GTCATAATCGCCGACGTCCAATCCGAACTGGGCCGCTCCGTCGCCGAAGAAATCAGCTCCGTATCTCCCCACCCG
GTCACCTACGTGGACTGCGACGTCAGCAAGGAGTCCGACGTGGCGAGCGCGGTGGACGGCGCCGTTTCGGCGTAC
GGGAAGCTGGACATTATGTACAGCAACGCCGGGATCATGGGGAGCTGACCGGGATGGGGATCGCCACGGTGGACA
GCGGCGATCTCGCCAGAGTGTTTGAAGTCAACGTCAAGGGAGCTTTCCACTGCGCCAAGCAGGCGGCGAGGGTGA
TGATACCGGAGAAGAAGGGGACGATTTTGTTCACGGCGAGTAGCGTTACGGCGACGTATGGGAACGCGCCGCACC
CTTATACGGCGTCGAAGCACGCGGTGGTCGGGCTGATGAAGAATTTGACGGTGGAGTTGGGAGGGTATGGGATTC
GGGTGAACTCGATTTCGCCGGACGGGGTGCCGACGCCGATGGCGATGAAGTTGATGGGGATGGACCGGAAGGCGG
TTCAGGAGCTGGGGTCACAGAGGGCGAGCTTGAAGGGGACGGTTCTGGACGAGAACGACGTCGCAGAGGCGGCTA
TGTACCTCGCCGGCGACGAGTCGAAGTTTGTGAGCGGGTTGAACCTCGTGTTGGACGGCGGTTGTAGTTTGAAGA
GTGCT**TGA**

The gDNA-sequence of SDH-candidate 7665 comprises 1626 bp incuding two introns (underlined) and is shown below:

ATGTCGAAGCTGGTCATGGGTTCTTCCAGGTGATTAACGAGAAAGCTAACGTCTTTTCTTTTTTCTTCTCTCTTT
GCTTCTTCTGCATAGTGGTAGACTCTATTAATTATCATATTCCGATGTGATTATGAACAGAGATCTGGCGTTAAA
AGCCCGGTCACCGGCGGCCAACGTTCTTAGACGCTTATTCTCTACCTTACCCGGAGAAGCTGTTGCAAGTGGCAA
TGAAAGCAGGTAAGTGGCTCTTTCCTCTTGTTCAATTTAATTACGTTAGTAATCACCTCCCACCTTTTTCACTGT
ATTATTTACCCTCTGCTTCGTTCTCTCTGTTTTCCCCTTTTCTTCTAGCTGGTAGTCAATTTTCCAGGAACTGGA
GCGGTTAAATTGATTAGGACTGGAAAACTAATCTCATTAGGCAGTAGGGTAGTTCTGACAATGCATGTGTTTTCT
CTCTGTCTCTGCAACTTGCTCGACTCTAATCTAAATCTAAATCTAATCATTAAACAAAACGAATGAGGCCTGTTT
GTCTTATACTATGGGTCTCATATTTCTTGTAATTTTTTATGGGAACTGATTCCATCTTCACTGACTAGTGGCCCC
TCCCTATTTACGTTTGCATGTCCCTGTCGGCTGCCGTCCATGTTGGCACGGGCCCACATTTTTGGGCCCACTCGT
GTCTTTGTCCAATCAAACATATCCGTACTATTGGGCCTTATTTTTAAGCTACGGCCCAGAGTCACAACAAAGAGA
GTCGACCTCGGCCCGGACTTGTTCATATATGTTTGGGCTTTGTAAATACACTTGGCAACGTGATTAACCAGCAGG
AAGCTAGAGGGGAAAGTAGCGGTGATCACGGGAGGAGCAAATGGGATTGGAAGAGAAACCGCCGTCAAATTCATC
GCCAACGGCGCCAAAGTCATAATAGCCGACATCAACAAGGAACTCGGCCTCCAAACAGCGCGCCAGCTAGGCCCC
AACGCATCGTTCATCCCATGTGACGTGGCCAATGAATCCGACGTGTCGGACGCCGTGGACTTTGCAGTGTCCCAG
CACAGCCGCCTCGACATCATGTACAACAATGCTGGCATCGCCTGCAGAACTCCTCCCAGCGTCGCCGACCTCGAC
CTGGCAGCCTTCGATACGGTGATGGGGATAAACGTGAGGGGAGCGCTGGCGGGAGTGAAGCACGCGGCGCGGGTG
ATGATCCCGAGGCAGAGCGGGGCCATCTTGTGCACGGCGAGCGTGACGGGGATCATGGCAGGGATGGCTCAGCTG

ACGTATTCGGTGTCGAAGGCGGGGGTGGTGGCCATCGTGAAATCGGCGGCGGCGGAGCTGTGCCAGCACGGGGTG
AGGGTGAACTGCATCTCCCCGTTCGCCATCCCGACGGCATTCGTGATGGAGGAGCTGCAGCATTACTACGCCGGG
GAGAGGCCGGAGCGGATCATGGAGAACGTGTACGCCATGGGGTCGCTGAAGGGAGGGAGGTGCGAGCCGGGTGAC
GTGGCGAACGCGGCGGTTTATCTGGCGTCGGAGGATGCAAAGTACGTTACGGGGCATAACTTGGTGGTGGATGGT
GGGTTTACGTCGTTCAAGACTCTTGGGTTGCCTCCAAGGAATAGCCAG**TGA**

The introns (underlined) in the gDNA sequences of SDH candidates were identified by comparing the gDNA sequences with the ORF-sequences obtained from the transcriptome database.

The intron phases and lengths of all introns of the five SDH-candidates are listed below:

SDH-candidate	Intron start position	Intron phase	Length	Dinucleotide sequences at intron boundaries (5' - 3')
28880	38	II	113	GT-AG
36067	45	0	212	AG-GC
5591	38	II	83	GT-AG
	420	0	694	GT-AG
73995	38	II	96	CT-AG
7665	29	II	106	GT-AG
	122	II	587	AA-TT

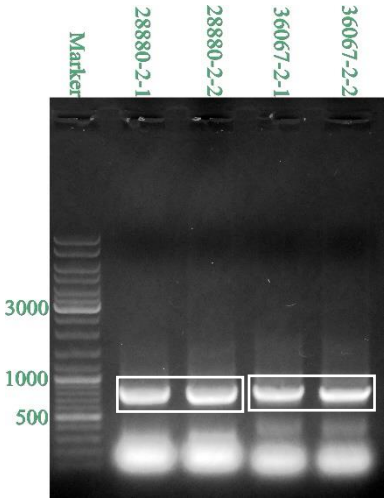
All SDH-candidates contain one or two introns that are located close to the 5'-end of the sequences. Generally, the most common introns are phase 0 introns, followed by phase I introns and the least common are phase II introns (Tomita et al., 1996). In proteins with secretory signal peptides, phase I introns are the most common (Vibranovski et al., 2006). However, the intron phases in the five SDH-candidates is not consistent and the majority is phase II (71%). In addition, four of the seven introns have the consensus sequences GT-AG at the intron boundaries. The other four candidates contain non-canonical splice sites.

5.3 Fusion-PCR and verification of full-length SDH-candidate sequences

Fusion-PCR (see III.2.4.3) was performed for the *in-vitro*-removal of introns from genomic DNA in order to generate a full-length sequence of continuous coding capacity.

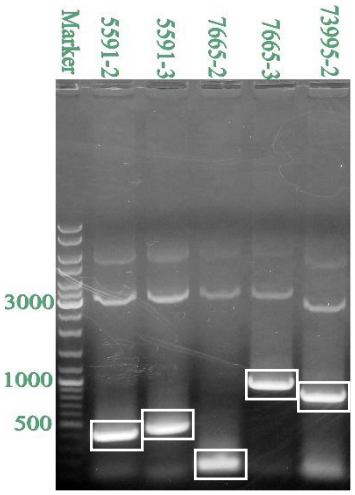
5.3.1 Exon fragments in the first rounds

Distinct exon fragments were generated in the first PCR rounds with Phusion® Polymerase to avoid replication errors (see III.2.4.3). The first exon fragments of all SDH-candidates are shorter than 50 nucleotides and were synthesized directly by Eurofins.



Sample	Annealing temperature (°C)
28880-2-1	61.3
28880-2-2	63.4
36067-2-1	65.2
36067-2-2	67.5

Figure 71: PCR amplification of the second exon fragments of SDH-candidates 28880 and 36067
PCR-products are marked with a box.



Sample	Annealing temperature (°C)
5591-2	55.5
5591-3	55.5
7665-2	56.7
7665-3	56.7
73995-2	62.4

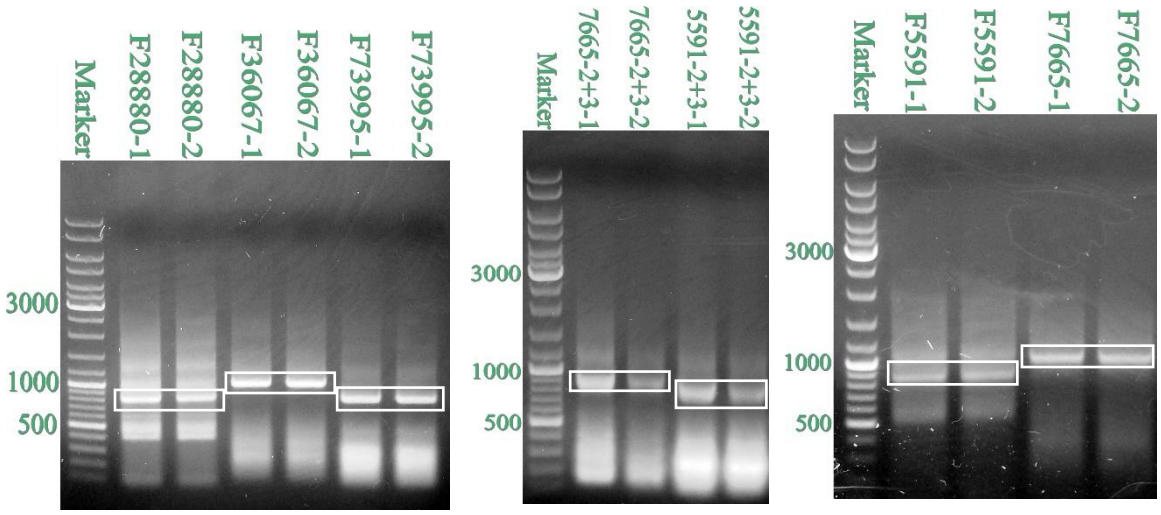
Figure 72: PCR amplification of the second and third exon fragments of SDH-candidates 5591, 7665 and 73995
PCR-products are marked with boxes.

The PCR products of the expected size were cut out, purified via gel extraction. After purification and determination of product concentrations, the first-round PCR products were diluted to the same concentration and used as template in the next rounds.

5.3.2 Full-length sequences of SDH-candidates in the second and third rounds

Candidates 28880, 36067 and 73995 each contain one intron. Hence, only two rounds of PCR were required to finish the generation of full-length sequences. Candidates 5591 and 7665 contain two introns each and three rounds of fusion-PCR were performed to excise all introns

from the sequences. Two outermost primers including specific sequences for restriction enzymes were used in the last rounds of fusion-PCR.



Sample	Annealing temperature (°C)
F28880-1	61.3
F28880-2	63.4
F36067-1	65.2
F36067-2	67.5
F73995-1	61.3
F73995-2	63.4

Figure 73: The second-round fusion-PCR of SDH-candidates 28880, 36067 and 73995
Full-length sequences are marked with boxes.

Sample	Annealing temperature (°C)
7665-2+3-1	56.7
7665-2+3-2	56.7
5591-2+3-1	55.5
5591-2+3-2	55.5

Figure 74: The second-round fusion-PCR of SDH-candidates 7665 and 5591
The second and the third exon-fragments of candidates were amplified together. PCR-products are marked with boxes.

Sample	Annealing temperature (°C)
F5591-1	55.5
F5591-2	55.5
F7665-1	56.7
F7665-2	56.7

Figure 75: The third-round fusion-PCR of SDH-candidates 7665 and 5591
Full-length sequences are marked with boxes.

The full-length PCR products of the SDH-candidates with the appropriate lengths were excised and isolated from the agarose gel. The amplicons were then ligated into the pDrive vector in order to multiply and verify the sequences.

5.3.3 Verification of full-length sequences

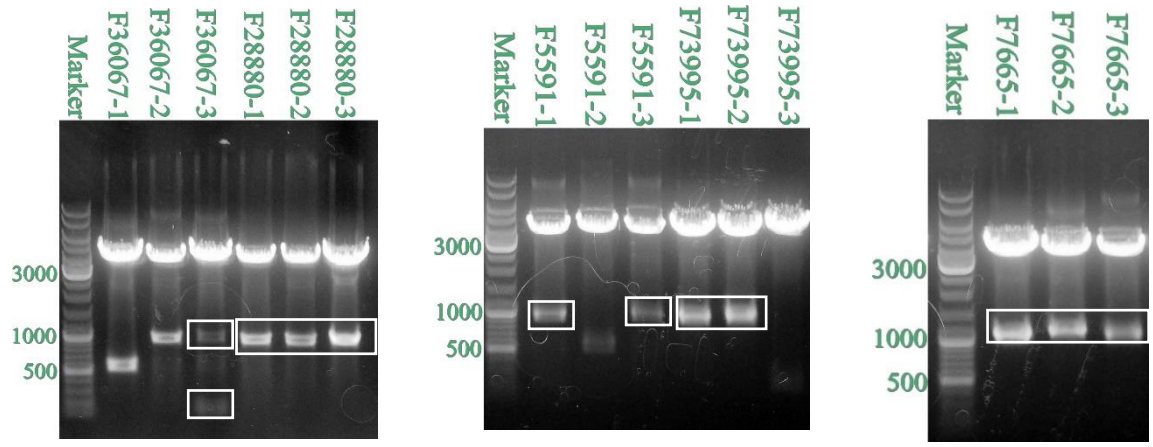


Figure 76: Agarose gels to verify the full-length sequences of SDH-candidates in the vector pDrive (colony 1-3)
Plasmids were isolated from *E. coli* EZ and cut with EcoRI. The sequence of candidate 36067 contains a restriction site for EcoRI. Hence, two bands for 36067 appeared on the agarose gel. Candidate genes are marked with boxes.

The ORF of SDH-candidate 28880 comprises 825 bp and shows 99.6% identity to the sequence of contig 28880 in transcriptome of *L. flavum*. The ORF of CYP-candidate 28880 is as below:

ATGGGCTCTGATATCTGTGCACCCTCTGTCACCAAGACGCTAGAAGGCAAAGTGGCGCTCATAACTGGAGGAGCA
AGCGGCATAGGCGAGAGCGCTGTCAGGCTGTTCATGCAACACGGAGCCAAGGTTCTCATCGCCGACCTTCAAGAT
AAACTCGGCGAATCCCTCGCCACGGAACTCGGTCCACCGGAGATTGTCTCCTTCGTCCACTGCAATGTCACTGTA
GACTCCGATGTGGGAAACGCTGTGGATACTGCGGTTTTGAGGTACGGACAGTTGGACATCATGTACAACAACGCT
GGAATCGGAGGCAACCTCGACACCACGATTTTAAACTCCGACAATGACGATTTCAAAAGGGTATTGGAAATAAAC
CTGTTCGGATCGTTCCTCGGAGCCAAACACGCTGCTAGAGTGATGATTCCGGCGAGGAAAGGTTGCATACTATTT
ACGGGAAGCGTGGCGTCGTCAATAAGCGGGGATTTATCATACGCTTACAAGGCTTCGAAGCACGCGTTGTTGGGG
CTAGCCAATAATTTGACTGTGGAGATGGGGAAGTACGGGATTAGGGTCAACACTATATCCCCGTATGGAGTGGCC
ACTCCTATGGTGACACGTGGCCTGCAAATGGATAAGAAGGCTGTCGAGCAGTTCATGTCGGCAGCGGGGAACCTA
AGAGGGGCGATTTTGGAGCCAGAGGATGTTGCCAAGGCGGCGCTGTACTTGGCTAGCGATGATGCGAAATATGTG
AGTGGGCTGAACTTGATTGTTGACGGTGGCCACAATCACAACCATCCTCTCTTTGGCGCTTCCGNTGAAGCA**TGA**

The ORF of SDH-candidate 36067 comprises 849 bp and shows 96.5% identity to the sequence of contig 36067 in transcriptome of *L. flavum*. The ORF of CYP-candidate 36067 is as below:

ATGGCGGCAGCAGCACCAGCTTCCTTTATCTCCTCCGTCGCCAGAAGGCTGGAAGGCAAAGTAGCGCTAATCACC
GGAGGAGCAAGCGGGATCGGCGAATTCACCGCCAGAGTGTTCGCCCACCACGGCGCCAAAGTCGTGATCGCCGAC
GTCCAAGACGACCTCGGCCAATCCGTCGTGGCCGACACCCTGGGCCGCGCCAACTCCACCTACGTCCGCTGCGAC
GTCACCGAGGAGTCCCAAATCAAAGACGCCGTCGACAAGGCCGTCTCCACCTACGGCAAGCTGGACATCATGTTC
AACAACGCCGGCATCGCCGACCCCAACAAGCCCCGCATAGTCGACAACGAGAAGTCCGACTTCGAGCGCGTCCTC
AGCGTCAACGTCACCGGTGTCTTCCTCGGGATGAAGCACGCGGCGCGGGTGATGATCCCGCGCCGCACGGGGGCC
ATAATCTCGACCGCCAGCATCAGCTGCGCTATGGGCGGCGCAGCCTCGCACGCATACTGTTGCTCGAAGCACGCC
GTGCTGGGGCTGACTCGGAACGCGGCGGTCGAGCTGGGGCAGTTCGGGATCAGGGTGAACTGCCTGGCGCCGTAC
GGAATGGCGACGCCTCTGGCGACGAAGTATGTGGGGCTGCCGGATGAGGAGTTGGAAGGGTCGATGAATGCTCTG
GCGAATTTGAAAGGAGTGACGCTGAAAACAGAGGATGTGGCGAACGCTGCTCTGTTCTTGGCTAGCGATGAAGGT

AGGTATGTTAGTGGGCATAGTTTGGTTGTGGATGGAGGGTTTAGCATTCATAATACTGCGTTTAGGATCTTCAAT
TACCCGGAAGATGGATCTTCA**TGA**

The specific sequence GAATTC for restriction enzyme EcoRI is underlined.

The ORF of SDH-candidate 5591 comprises 843 bp and shows 100% identity to the sequence of contig 5591 in transcriptome of *L. flavum*. The ORF of CYP-candidate 5591 is as below:

ATGGCCGCCACTAACTTCGTTTCTTCCATCATAAAAAGGCTGGAAGGCAAAGTAGCGCTGATCACAGGCGGAGCC
AGCGGAATCGGCGAGGCCACCGCCAGCCTTTTCGTCCAACACGGCGCCAAAGTCCTAATCGCCGACGTCCAAGAC
CAACTCGGCGAATCCTTAGCCCGAAACCTAGGCCCGCCGGACACGGTCTCCTACGTCCGCTGCGATGTTACCAGT
CCTCCCGAAGTCGGGGCCGCCGTGGACGAGGCGGTGTCGAGGTACGGCCAGCTGGACATCATGCACAACAACGCC
GGCATCGCCGGGGGCGTAGACCAGCGTATCGTTCACTCCGACGACGATAACTTCCGGCGGGTCCTGGACGTCAAC
CTCTTCGGATCCTACCTTGGGGCTAAGCACGCCGCCAGGGTGATGGTACCGGCGGGGAGAGGCGGGTGCATTCTG
TTCACGGCGAGCAGCGCGGCGGTGACGAGCGGGAACATTACGTACGCGTACAAGGCGTCGAAGAACGGGGTGTTG
GGGCTGGCGAACAACCTGTGCGTGGAGCTGGGTCAGTACGGGATCCGGGTAAACGCGATCTCGCCGTTCACGGTG
GTGACTCCGCTGTTGGTGAGTGTGCTGGGAGGCGGCGGGCTGGGGAAGGAGGAGGTGGAGAGGCTGGTGGCGGCG
GCGACGAAGCCGAACTTGAAAGGAGCGGTGTTGGAAGCGGCGGATATTGCCAAGGCGGCGCTGTTTTTGGCGAGC
GATGATGCCAAGTACGTGAGTGGGTTGAATTTGGTCGTCGATGGTGGGCATAATCACAACCGTCCGATCTTTGAT
GATTTGAGGTTCGCT**TGA**

The ORF of SDH-candidate 73995 comprises 813 bp and shows 96.7% identity to the sequence of contig 73995 in transcriptome of *L. flavum*. The ORF of CYP-candidate 73995 is as below:

ATGTCTTCTTCCTACCAGCCGGCGGCGGCGAGTTCTCACAGACTAGAAGGCAAGGTGGCAATGATCACCGGCGGC
GCGAGCGGGATCGGCGCAGCGACGGCCAAGCTCTTCGCCCTCAACGGCGCCAGAGTCATAATCGCCGACGTCCAA
TCCGAACTGGGCCGCTCCGTCGCCCAACAGATCAGCTCCGTCGCACCCCACCCGGTCACCTACGTCCACTGCGAC
GTCAGCAACGAGTCCGACGTGGCGAGCGCGGTGGACGGCGCCGTTTCGGCCTACGGCAAGCTGGACATCATGTAC
AGCAACGCCGGGATCATGGGGGAGCTGACGGGGATGGGGATCACCACGGTGGACAGTGGCGATCTCGCCAGAGTG
TTTGAAGTCAACGTCAAGGGAGCTTTCCACTGCGCCAAGCAGGCGGCGAGGGTGATGATACCGGAGAAGAAGGGG
ACGATTTTGTTCACGGCGAGTAGCATAACGGCGACGTATGGGAACGCGCCGCATCCTTATACGGCGTCGAAGCAC
GCGGTGGTCGGGCTGATGAAGAATTTGACGGTGGAGTTGGGAGGGTATGGGATTCGGGTCAACGCGATTTCGCCG
GACACGGTGCCGACGCCGATGGCGATGAACTTGTTGGGGATGGACCGGAAGGCGGTTCAGGAGTTTGGGTCGCAG
AGGGCGAGCTTGAAAGGGACGGTTCTGGACGAGAACGACGTCGCACAGGCGGCTATGTACCTCGCCGGCGACGAG
TCGAAGTTTGTGAGCGGGTTGAACCTCGTGTTGGACGGCGGTTGTAGTTTGAAGAGTGCT**TGA**

The ORF of SDH-candidate 7665 comprises 933 bp and shows 98.1% identity to the sequence of contig 7665 in transcriptome of *L. flavum*. The ORF of CYP-candidate 7665 is as below:

ATGTCGAAGCTGGTCATGGGTTCTTCCAGAGATCTGGCGTTAAAAGCCCGGTCACCGGCGGCCAACGTTCTTAGA
CGCTTATTCTCTACCTTACCCGGAGAAGCTGTTGCAAGTGGCAATGAAACCAGCAGGAAGCTAGAGGGGAAAGTA
GCGGTGATCACGGGAGGAGCAAATGGGATTGGAAGAGAAACCGCCGTCAAATTCATCGCCAACGGCGCCAAAGTC
ATAATAGCCGACATCAACAAGGAACTCGGCCTCCAAACAGCGCGCCAGCTAGGCCCCAACGCATCGTTCATCCCA
TGTGACGTGGCCAATGAATCCGACGTGTCGGACGCCGTGGACTTTGCAGTGTCCCAGCACAGCCGCCTCGACATC
ATGTACAACAATGCTGGCATCGCCTGCAGAACTCCTCCCAGCGTCGCCGACCTCGACCTGGCAGCCTTCGATACG
GTGATGGGGATAAACGTGAGGGGAGCGCTGGCGGGAGTGAAGCACGCGGCGCGGGTGATGATCCCGAGGCAGAGC
GGGGCCATCTTGTGCACGGCGAGCGTGACGGGGATCATGGCAGGGATGGCTCAGCTGACGTATTCGGTGTCGAAG
GCGGGGGTGGTGGCCATCGTGAAATCGGCGGCGGCGGAGCTGTGCCAGCACGGGGTGAGGGTGAACTGCATCTCC
CCGTTCGCCATCCCGACGGCATTCGTGATGGAGGAGCTGCAGCATTACTACGCCGGGGAGAGGCCGGAGCGGATC
ATGGAGAACGTGTACGCCATGGGGTCGCTGAAGGGAGGGAGGTGCGAGCCGGGTGACGTGGCGAACGCGACGGTT
TATCTGGCGTCGGAGGATGCAAAGTACGTTACGGGGCATAACTTGGTGGTGGATGGTGGGTTTACGTCGTTCAAG
ACTCTTGGGTTGCCTCCAAGGAATAGCCAG**TGA**

The start codon and the stop codon are written in bold. The amino acid sequences of the SDH-candidates were translated from ORFs with the help of ExPASy-translate and the molecular weights of the proteins were calculated by the ExPASy-Compute pI/Mw tool.

The amino acid sequence of candidate 28880 with a molecular weight of appr. 28.7 kDa is shown below:

MGSDICAPSVTKTLEGKVALITGGASGIGESAVRLFMQHGAKVLIADLQDKLGESLATELGPPEIVSF
VHCNVTVDSDVGNAVDTAVLRYGQLDIMYNNAGIGGNLDTTILNSDNDDFKRVLEINLFGSFLGAKHA
ARVMIPARKGCILFTGSVASSISGDLSYAYKASKHALLGLANNLTVEMGKYGIRVNTISPYGVATPMV
TRGLQMDKKAVEQFMSAAGNLRGAILEPEDVAKAALYLASDDAKYVSGLNLIVDGGHNHNHPLFGASX
EA

The amino acid sequence of candidate 36067 with a molecular weight of appr. 29.6 kDa is shown below:

MAAAAPASFISSVARRLEGKVALITGGASGIGEFTARVFAHHGAKVVIADVQDDLGQSVVADTLGRAN
STYVRCDVTEESQIKDAVDKAVSTYGKLDIMFNNAGIADPNKPRIVDNEKSDFERVLSVNVTGVFLGM
KHAARVMIPRRTGAIISTASISCAMGGAASHAYCCSKHAVLGLTRNAAVELGQFGIRVNCLAPYGMAT
PLATKYVGLPDEELEGSMNALANLKGVTLKTEDVANAALFLASDEGRYVSGHSLVVDGGFSIHNTAFR
IFNYPEDGSS

The amino acid sequence of candidate 5591 with a molecular weight of appr. 29 kDa is shown below:

MAATNFVSSIIKRLEGKVALITGGASGIGEATASLFVQHGAKVLIADVQDQLGESLARNLGPPDTVSY
VRCDVTSPPEVGAAVDEAVSRYGQLDIMHNNAGIAGGVDQRIVHSDDDNFRRVLDVNLFGSYLGAKHA
ARVMVPAGRGGCILFTASSAAVTSGNITYAYKASKNGVLGLANNLCVELGQYGIRVNAISPFTVVTPL
LVSVLGGGGLGKEEVERLVAAATKPNLKGAVLEAADIAKAALFLASDDAKYVSGLNLVVDGGHNHNRP
IFDDLRFA

The amino acid sequence of candidate 73995 with a molecular weight of appr. 27.9 kDa is shown below:

MSSSYQPAAASSHRLEGKVAMITGGASGIGAATAKLFALNGARVIIADVQSELGRSVAQQISSVAPHP
VTYVHCDVSNESDVASAVDGAVSAYGKLDIMYSNAGIMGELTGMGITTVDSGDLARVFEVNVKGAFHC
AKQAARVMIPEKKGTILFTASSITATYGNAPHPYTASKHAVVGLMKNLTVELGGYGIRVNAISPDTVP
TPMAMNLLGMDRKAVQEFGSQRASLKGTVLDENDVAQAAMYLAGDESKFVSGLNLVLDGGCSLKSA

The amino acid sequence of candidate 7665 with a molecular weight of appr. 32.5 kDa is shown below:

MSKLVMGSSRDLALKARSPAANVLRRLFSTLPGEAVASGNETSRKLEGKVAVITGGANGIGRETAVKF
IANGAKVIIADINKELGLQTARQLGPNASFIPCDVANESDVSDAVDFAVSQHSRLDIMYNNAGIACRT

PPSVADLDLAAFDTVMGINVRGALAGVKHAARVMIPRQSGAILCTASVTGIMAGMAQLTYSVSKAGVV
AIVKSAAAELCQHGVRVNCISPFAIPTAFVMEELQHYYAGERPERIMENVYAMGSLKGGRCEPGDVAN
ATVYLASEDAKYVTGHNLVVDGGFTSFKTLGLPPRNSQ

All five SDH-candidates were confirmed to belong to the short-chain dehydrogenases/reductase family (SDR) based on the PSI- and PHI-BLAST searches against the conserved domain database (CDD) of the National Center for Biotechnology Information. The obtained amino acid sequenses of the SDH-candidates were compared to already known SDH sequences from *Forsythia x intermedia* (accession No. AAK38665.1) and *Podophyllum peltatum* (accession No. AAK38664.1) by Blastp. All five SDH-candidates of *L. flavum* have appr. 50% identity with SDH from *Forsythia x intermedia* and *Podophyllum peltatum.* Especially, candidate 36067 showed high homology to SDH from *Forsythia x intermedia* (68%). Furthermore, the alignment of the amino acid sequences of these SDH-candidates from *Linum flavum* and SDH from *Forsythia x intermedia* and *Podophyllum peltatum* was performed using Clustal Omega and the result is shown below:

```
SDH-Pp      --------------------------------MGS-TSTPDSSTNRLQDKVAIITGGAGGI    28
SDH-Fi      --------------------------------MAATSQVLTAIARRLEGKVALITGGASGI    29
28880       --------------------------------MGS-DICAPSVTKTLEGKVALITGGASGI    28
36067       -------------------------------MAAAAPASFISSVARRLEGKVALITGGASGI    31
5591        --------------------------------MAA-TNFVSSIIKRLEGKVALITGGASGI    28
73995       --------------------------------MSSSYQPAAASSHRLEGKVAMITGGASGI    29
7665        MSKLVMGSSRDLALKARSPAANVLRRLFSTLPGEAVASGNETSRKLEGKVAVITGGANGI    60
                                            .          . *:.***:*****.**

SDH-Pp      GETTAKLFVRYGAKVVIADIADDHGQKVCNNIG--SPDVISFVHCDVTKDEDVRNLVDTT    86
SDH-Fi      GETTAKLFSQHGAKVAIADVQDELGHSVV-EAI--GTSNSTYIHCDVTNEDGVKNAVDNT    86
28880       GESAVRLFMQHGAKVLIADLQDKLGESLATELG--PPEIVSFVHCNVTVDSDVGNAVDTA    86
36067       GEFTARVFAHHGAKVVIADVQDDLGQSVVADTL--GRANSTYVRCDVTEESQIKDAVDKA    89
5591        GEATASLFVQHGAKVLIADVQDQLGESLARNLG--PPDTVSYVRCDVTSPPEVGAAVDEA    86
73995       GAATAKLFALNGARVIIADVQSELGRSVAQQISSVAPHPVTYVHCDVSNESDVASAVDGA    89
7665        GRETAVKFIANGAKVIIADINKELGLQTARQLG----PNASFIPCDVANESDVSDAVDFA    116
            *  :.  *   **:* ***: .. * .   :         ::: *:*:    :   ** :

SDH-Pp      IAKHGKLDIMFGNVGVLSTTP-YSILEAGNEDFKRVMDINVYGAFLVAKHAARVMIPA-K    144
SDH-Fi      VSTYGKLDIMFSNAGISDPNR-PRIIDNEKADFERVFSVNVTGVFLCMKHAARVMIPA-R    144
28880       VLRYGQLDIMYNNAGIGGNLD-TTILNSDNDDFKRVLEINLFGSFLGAKHAARVMIPA-R    144
36067       VSTYGKLDIMFNNAGIADPNK-PRIVDNEKSDFERVLSVNVTGVFLGMKHAARVMIPR-R    147
5591        VSRYGQLDIMHNNAGIAGGVD-QRIVHSDDDNFRRVLDVNLFGSYLGAKHAARVMVPAGR    145
73995       VSAYGKLDIMYSNAGIMGELTGMGITTVDSGDLARVFEVNVKGAFHCAKQAARVMIPE-K    148
7665        VSQHSRLDIMYNNAGIACRTP-PSVADLDLAAFDTVMGINVRGALAGVKHAARVMIPR-Q    174
            :  :.:****..*.*:        :       :  *: :*: *     *:*****:*  :

SDH-Pp      KGSIVFTASISSFTAGEGVSHVYTATKHAVLGLTTSLCTELGEYGIRVNCVSPYIVASPL    204
SDH-Fi      SGNIISTASLSSTMGG-GSSHAYCGSKHAVLGLTRNLAVELGQFGIRVNCLSPFGLPTAL    203
28880       KGCILFTGSVASSISG-DLSYAYKASKHALLGLANNLTVEMGKYGIRVNTISPYGVATPM    203
36067       TGAIISTASISCAMGG-AASHAYCCSKHAVLGLTRNAAVELGQFGIRVNCLAPYGMATPL    206
5591        GGCILFTASSAAVTSG-NITYAYKASKNGVLGLANNLCVELGQYGIRVNAISPFTVVTPL    204
73995       KGTILFTASSITATYG-NAPHPYTASKHAVVGLMKNLTVELGGYGIRVNAISPDTVPTPM    207
7665        SGAILCTASVTGIMAG-MAQLTYSVSKAGVVAIVKSAAAELCQHGVRVNCISPFAIPTAF    233
```

```
            * *: *.*      *      *  :* .::.:  .  .*:  .*:*** ::*  : : :

SDH-Pp      LTDVFGVD---SSRVEELA--HQAANLKGTLLRAEDVADAVAYLAGDESKYVSGLNLVID    259
SDH-Fi      GKKFSGIK--NEEEFENV--INFAGNLKGPKFNVEDVANAALYLASDEAKYVSGHNLFID    259
28880       VTRGL---QMDKKAVEQFMS--AAGNLRGAILEPEDVAKAALYLASDDAKYVSGLNLIVD    258
36067       ATKYVGLP--DE-ELEGS--MNALANLKGVTLKTEDVANAALFLASDEGRYVSGHSLVVD    261
5591        LVSVLGGGGLGKEEVERLVAAATKPNLKGAVLEAADIAKAALFLASDDAKYVSGLNLVVD    264
73995       AMNLLGM---DRKAVQEFGS--QRASLKGTVLDENDVAQAAMYLAGDESKFVSGLNLVLD    262
7665        VMEELQHYYAGERPERIMENVYAMGSLKGGRCEPGDVANATVYLASEDAKYVTGHNLVVD    293
                        .         .*:*      *:*.*. :**.::.::*:* .*.:*

SDH-Pp      GGYTRTNPAFPTALKHGLA-- 278
SDH-Fi      GGFSVCNSVIKVFQYPDS--- 277
28880       GGHNHNHPLFGASXEA----- 274
36067       GGFSIHNTAFRIFNYPEDGSS 282
5591        GGHNHNRPIFDDLRFA----- 280
73995       GGCSLKSA------------- 270
7665        GGFTSFKTLG----LPPRNSQ 310
            ** .
```

Conserved amino acids associated with the NAD-binding motif are shown in boldface type (Xia et al., 2001). All five SDH-candidates contain the catalytic triad Serine - Tyrosine - Lysine of SDH (Moinuddin et al., 2006). The catalytic triad residues are shaded in grey.

5.4 Heterologous expression of SDH-candidate proteins

5.4.1 Expression in *E. coli* as prokaryotic cell line

Initially, in order to express large protein quantities in a short time, the discovered sequences were expressed in *E. coli* SoluBL21 cell lines. The full-length sequences of all SDH-candidates were ligated into the NdeI and XhoI restriction sites of the pET15b vector (see V.2.7.2) and then transferred into *E. coli* SoluBL21 cells by heat shock (see III.3.2). Heterologous expression of SDH-candidate proteins in *E. coli* SoluBL21 was performed as described in III.3.7.1. Subsequently, the recombinant proteins were purified by metal-chelate chromatography (see III.4.3). All fractions from the purification of the recombinant proteins were analysed on SDS-PAGE gels (see III.4.6).

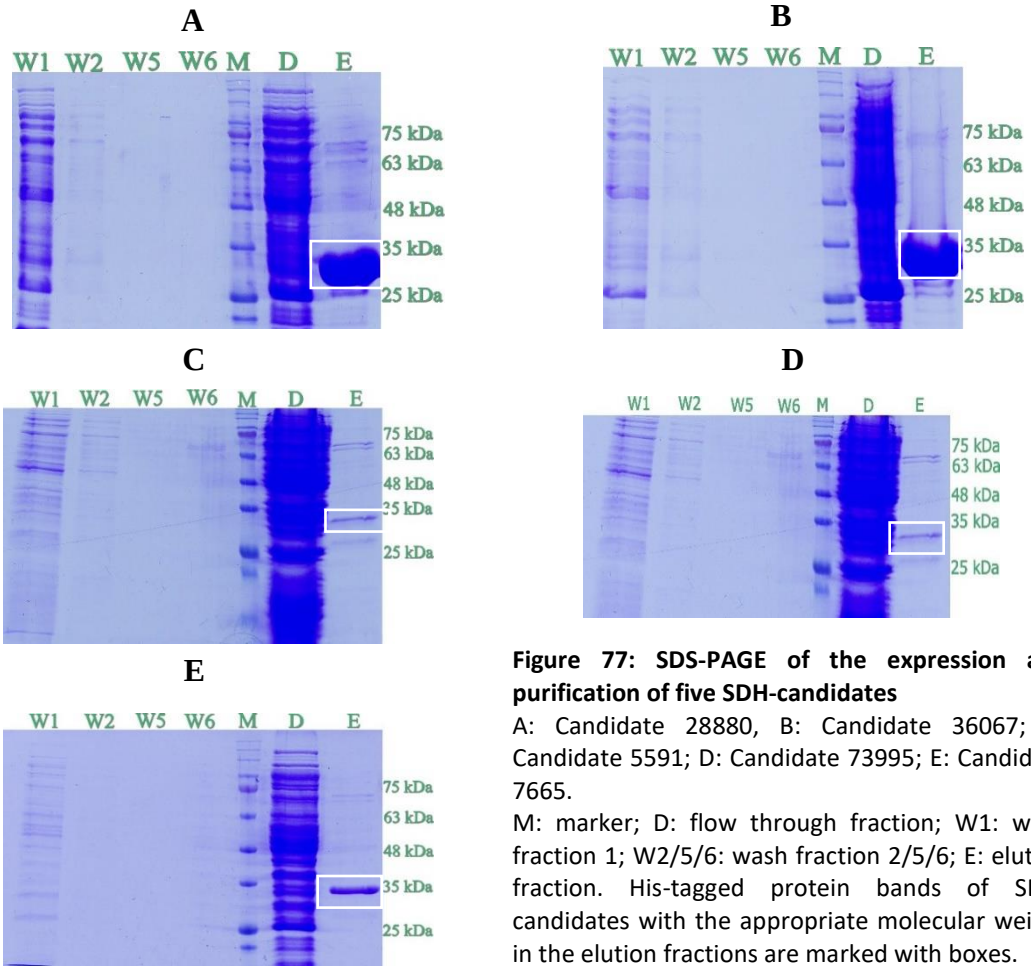


Figure 77: SDS-PAGE of the expression and purification of five SDH-candidates
A: Candidate 28880, B: Candidate 36067; C: Candidate 5591; D: Candidate 73995; E: Candidate 7665.
M: marker; D: flow through fraction; W1: wash fraction 1; W2/5/6: wash fraction 2/5/6; E: elution fraction. His-tagged protein bands of SDH-candidates with the appropriate molecular weight in the elution fractions are marked with boxes.

The molecular weight including His-Tag of SDH-candidate 28880, 36067, 5591, 73995 and 7665 is appr. 30.8 kDa, 31.7 kDa, 31.1 kDa, 30 kDa and 34.6 kDa, respectively. As can be seen in Fig. 77, gene-specific bands in the expected size ranges appeared in elution fractions on all SDS-PAGE gels. The result of SDS-PAGE showed that the transformed *E. coli* SoluBL21 cells carrying all SDH-candidates produced significant amounts of proteins after induction with IPTG and the purifications of heterologous proteins were successful.

The activities of the heterologous putative SDH-proteins were tested in enzyme assays modified according to Xia et al. (2000) (see V.4.8.4). However, no product formation (matairesinol) could be detected using SECO and NAD^+ as substrate in the HPLC chromatograms and enzyme activity tests with heterologously expressed proteins were negative.

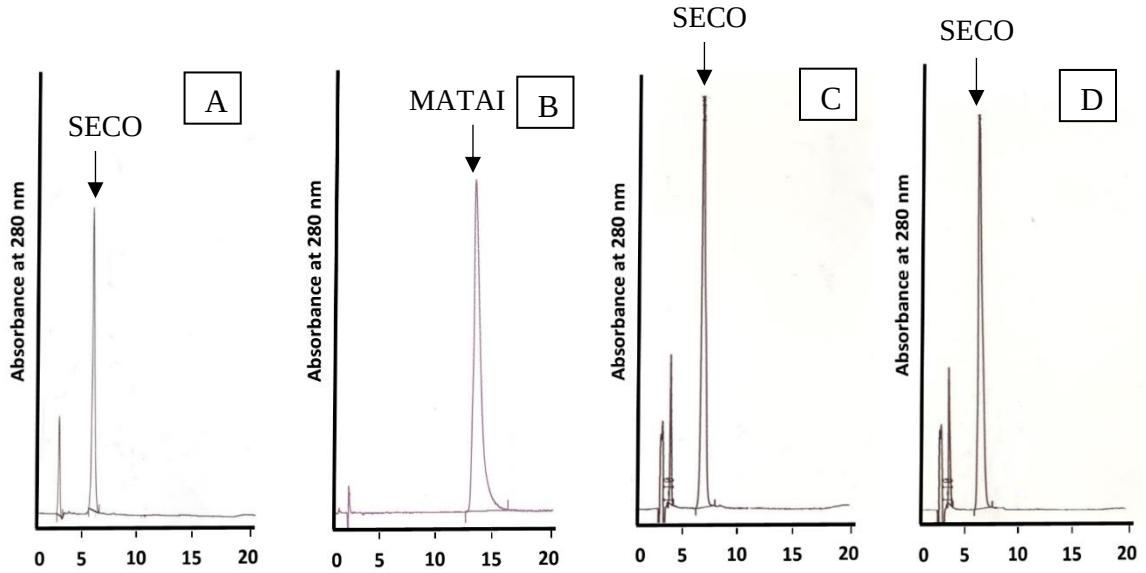


Figure 78: Chromatograms of racemic secoisolariciresinol (A), racemic matairesinol (B) and enzyme assays with different reaction times (C and D)
Racemic SECO and NAD^+ were incubated with the heterologously expressed protein of SDH-candidate 28880. SECO and MATAI showed significant peaks at 6.3 minutes and 14 minutes, respectively. The reaction time is C: 0 min; D: 2 hours.

5.4.2 Expression in *Saccharomyces cerevisiae* INVScI

5.4.2.1 Heterologous expression of candidate proteins with His-tag

Post-translational modifications could probably be responsible for the correct structure of SDH-proteins and are major problems for *E. coli*-based expression systems because of its limited ability for post-translational processing of proteins. Hence, the heterologous expression was performed in the eukaryotic cell line *Saccharomyces cerevisiae* INVScI to overcome this possible issue.

The full-length sequences of all SDH-candidates were amplified with new full-length primers by standard PCR reactions (see III.2.4.1). The full-length sequences were ligated into the HindIII and XbaI restriction sites of the pYes2/NTC vector (see V.2.7.2). *Saccharomyces cerevisiae* INVScI was transformed with the respective constructs (see III.3.3).

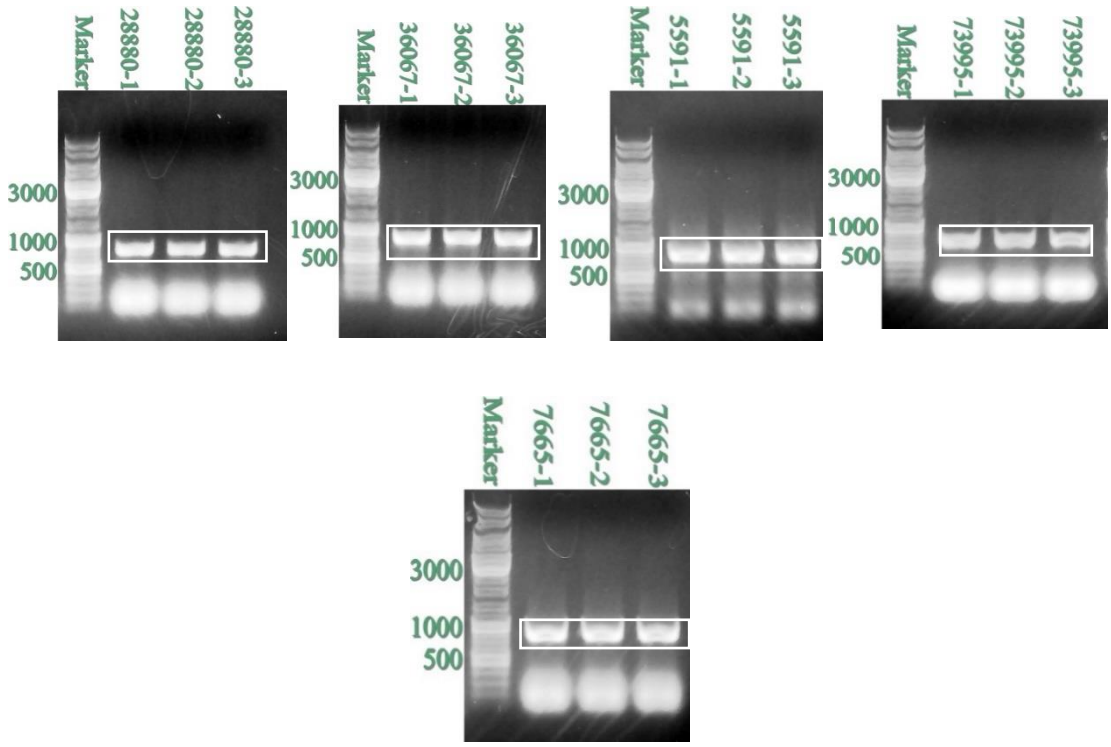


Figure 79: Agarose gels of colony PCR to verify the transformation of *Saccharomyces cerevisiae* INVScI with the plasmids pYES2/NTC carrying SDH candidate sequences (colony 1-3)
Bands of the candidate genes are marked with boxes.

Heterologous expression of candidate proteins was undertaken as described in III.3.7.2. As a negative control, the yeast cells containing the empty vector pYes2/NTC were expressed under the same condition. The crude protein from cell pellets of the SDH-candidates, negative controls and marker were separated by SDS-PAGE and then transferred to PVDF-membranes by Western blotting (see III.4.7). The His-Tag present on the vector was attached to the *C*-terminal end of the sequences facilitating the immunodetection of the expressed proteins with the anti-His-Tag antibody.

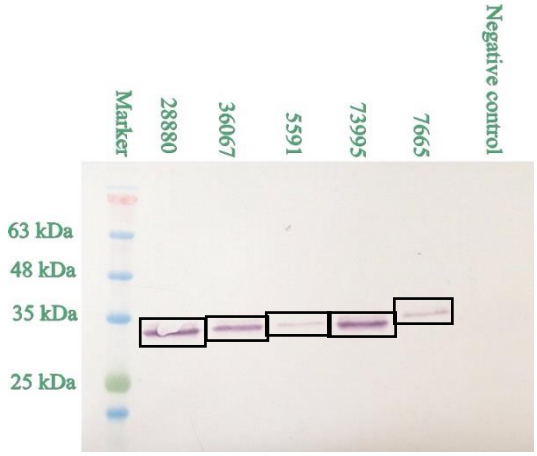


Figure 80: Western blot of SDH-candidates expressed in yeast

Detection was done with anti-His-Tag antibody and secondary antibodies coupled to alkaline phosphatase using the NBT/BCIP color reaction. The molecular weight of the attached His-Tag is appr. 3 kDa. His-tagged protein bands of SDH-candidates with appropriate molecular weights are marked with boxes.

In Fig. 80, significant specific bands of all candidate-proteins in the expected size range between 32 and 36 kDa appeared on the membrane. There were no bands in the same size in the lane of the negative control. The result of the Western blot showed that the transformed yeast cells of all five SDH-candidates produced the respective heterologous proteins.

The activity of heterologously expressed proteins was tested in enzyme assays as described in III.4.8.4. Furthermore, 20 mM TRIS/HCl buffer with different pH-values from 6 to 10 as well as 7-hydroxysecoisolariciresinol as alternative substrate were used in enzyme assays to test the enzymatic activity of the expressed proteins. However, no product formation could be detected in the HPLC chromatograms.

5.4.2.2 Heterologous expression of candidate proteins without His-tag

The *N*- or *C*-terminally attached His-tag may influence the structure as well as the function of proteins. Hence, the next attempts were conducted to express SDH-candidate proteins without His-tag. Full-length primers with a restriction site for XbaI and a stop codon in the reverse primers were designed for PCR-amplification (see IV.6.3). The full-length sequences were amplified with full-length forward primers and new full-length reverse primers in standard PCR reactions (see III.2.4.1). The successful transformations of the yeast strain INVScI with

the constructs of the candidate genes in pYes2/NTC were confirmed by colony-PCR (as shown in Fig. 81).

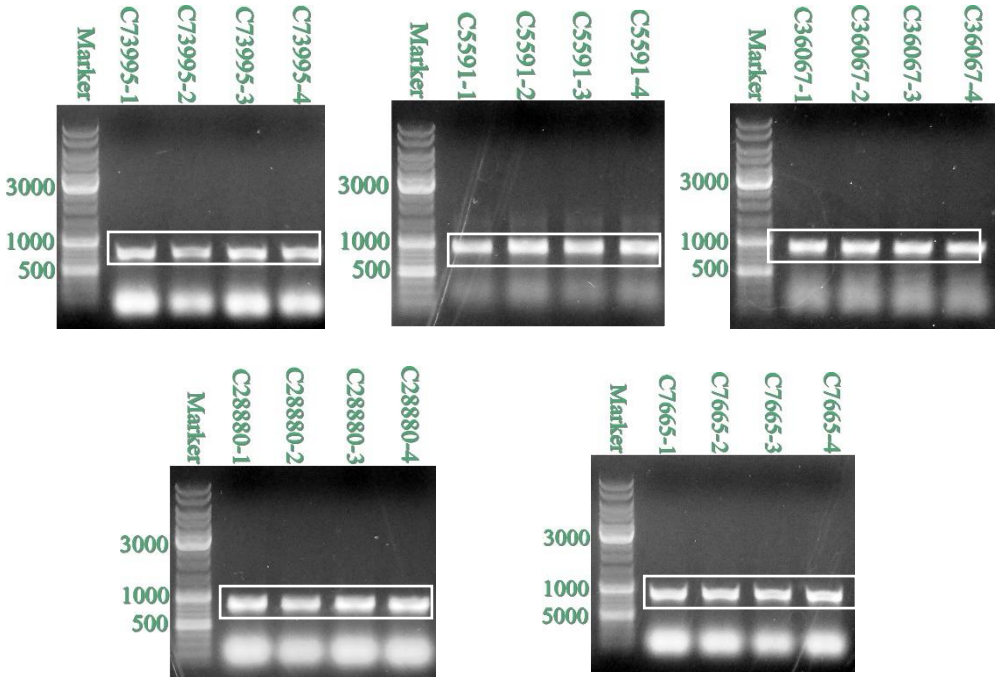


Figure 81: Agarose gels of colony PCR to verify the transformation of yeast INVScI (colony 1-4)
Candidate genes are marked with boxes.

At this stage, the detection of expressed SDH-candidate proteins without His-tag via anti-His-tag antibodies was not possible. The activity of potentially heterologously expressed proteins was still tested in enzyme assays with SECO as well as 7-hydroxysecoisolariciresinol as described in III.4.8.4. Additionally, 20 mM TRIS/HCl buffer with different pH-values from 6-10 were used to test the enzymatic activity of the expressed proteins. Unfortunately, no additional peak of the formed product was detected in the HPLC-chromatograms.

5.5 Concluding remarks and outlooks

SDH was firstly identified in *Forsythia intermedia* and *Podophyllum peltatum* by Xia et al. (2001). Since then, only a few researchers published information about the enzymatic activity of SDH. In 2006, the crystal structure of SDH of *Podophyllum peltatum* was reported by Moinuddin et al. (2006). The enzyme SDH is a homotetramer and consists of an α/β single domain monomer that contains seven parallel β-strands flanked by eight α-helices on both sides. The catalytic triad Ser^{153}, Tyr^{167} and Lys^{171} of SDH of *Podophyllum peltatum* was determined. The order of binding, and a catalytic mechanism for the enantiospecific conversion

of SECO into MATAI has also been proposed. Recently, SDH of the endophytic fungus *Phialocephala podophylla* was identified by Arneaud and Porter (2015). Nonetheless, the sequence of SDH of *Phialocephala podophylla* is 99% identical to the sequence of SDH of *Podophyllum peltatum* and differs only in six nucleotides.

From the transcriptome of *L. flavum*, five potential candidates of secoisolariciresinol dehydrogenase were obtained. All five SDH-candidates were successfully amplified from gDNA of *L. flavum* and the open reading frames were generated by fusion-PCR. Subsequently, recombinant SDH-candidate proteins were successfully expressed in *E. coli* and yeast. However, all attempts to test the activity and possible reaction of recombinant proteins with and without His-tag failed. The main reason could be the post-translational modification that will finish their tertiary and quaternary structures in order to become functional. The post-translational modification system of *E. coli* and yeast is possibly not sufficient to finish the modifications of SDH-proteins of plant. The expression in insect or in plant cells could be a solution for this problem.

VII. Summary

Lignans are phenolic specialized plant metabolites derived from the amino acids phenylalanine or tyrosine. Lignans are widely distributed throughout the plant kingdom and have multiple biological and pharmacological activities. An example of a medicinally used substance is the aryltetralin lignan podophyllotoxin and its derivatives that have cytotoxic activities. *Podophyllum* spec. are currently the major source of podophyllotoxin. However, aryltetralin lignans can also be detected in some flax species (*Linum* spec.). The biosynthesis of aryltetralin lignans is divided into the early and late steps. While the early reaction steps have been completely identified, most enzymes and reactions as well as their encoding genes are unknown regarding the late steps. In the present work, studies on the biosynthesis of lignans such as podophyllotoxin or 6-methoxypodophyllotoxin in suspension cultures of *L. flavum* are described. The roles of pinoresinol-lariciresinol reductase, secoisolariciresinol dehydrogenase, deoxypodophyllotoxin 6-hydroxylase and deoxypodophyllotoxin 7-hydroxylase are of particular interest. Deoxypodophyllotoxin 6-hydroxylase was already characterised in *Linum flavum* and belongs to the cytochrome P450 family. Hence, the identification of NADPH:cytochrome P450 reductase, which is essential for cytochrome P450-dependent reactions, was also an objective in this work.

The basis for the search of lignan biosynthetic genes was the transcriptome database of the 1KP project (https://onekp.com/). In the beginning, RNA and gDNA were isolated from suspension cultures of *L. flavum.* From cDNA two NADPH:cytochrome P450 reductases (CPR) were successfully identified, heterologously expressed and characterised. In order to determine the apparent K_m-values, enzyme assays were performed with varying concentrations of NADPH and cytochrome c. Both CPRs 66401 and 4753 showed high catalytic activity towards cytochrome c with K_m-values of 8.15 ± 0.3 µM and 15.6 ± 0.35 µM, respectively. Towards NADPH as an electron donor, CPR 66401 showed a K_m-value of 29.6 ± 0.8 µM and CPR 4753 had a K_m-value of 45.2 ± 0.7 µM. Two binding sites each for FMN, FAD, and NADPH and one binding site for cytochrome P450 were identified in the sequences of the two CPRs. Based on the membrane anchor regions at the *N*-terminus the NADPH:cytochrome P450 reductases 66401 and 4753 are classified into class I and II, respectively, according to Ro et al. (2002).

Pinoresinol-lariciresinol reductase is a bifunctional enzyme that catalyses the formation of secoisolariciresinol from pinoresinol via lariciresinol. From the transcriptome of *L. flavum*, one contig with the highest similarity to pinoresinol-lariciresinol reductase of *Forsythia intermedia*

was found. This contig was successfully amplified from the RNA/cDNA of suspension cultures of *L. flavum*, expressed in *E. coli* and subsequently characterised as pinoresinol-lariciresinol reductase. Pinoresinol-lariciresinol reductase of *L. flavum* catalyses stereospecifically the conversion of (+)-pinoresinol via (+)-lariciresinol to (-)-secoisolariciresinol. Pinoresinol-lariciresinol reductase of *L. flavum* displayed a K_m-value for NADPH of 22.4 ± 1.2 µM and has the highest catalytic activity at pH 7 and at a temperature of 27 °C to 33 °C. In order to investigate the role of the varying size of important amino acids in the binding pocket, two mutants of pinoresinol-lariciresinol reductase of *L. flavum*, G280Y and Y284G, were constructed. The loss of enzymatic activity of both mutants LfPLR G280Y and Y284G showed the crucial role of the two positions G^{280} and Y^{284} for the catalytic activity of this enzyme.

In addition, multiple candidates of cytochrome P450 and secoisolariciresinol dehydrogenase were obtained from the transcriptome of *L. flavum.* Eleven open reading frames of cytochrome P450 and five open reading frames of secoisolariciresinol dehydrogenase (SDH) were successfully amplified from RNA/cDNA or gDNA of *L. flavum*. The cytochrome P450s were successfully expressed in *Saccharomyces cerevisiae* and SDH in *E. coli* and yeast. The conserved domains for cytochrome P450 and secoisolariciresinol dehydrogenase were observed in the sequences of each candidate. Nevertheless, the recombinant proteins were not catalytically active with the applied putative substrates. Maybe a post-translational modification is essential for these enzymes and bacteria and yeast are not sufficient to conduct the required modifications. A more suitable expression system, for example in plant or insect cultures, could be a solution for this problem. The search for candidates should be refined to look for other candidates for cytochrome P450 and secoisolariciresinol dehydrogenase as well.

VIII. Zusammenfassung

Lignane sind phenolische Verbindungen des pflanzenspezifischen Stoffwechsels, die von den Aminosäuren Phenylalanin oder Tyrosin abgeleitet sind. Lignane sind im Pflanzenreich weit verbreitet und haben zahlreiche biologische und pharmakologische Aktivitäten. Ein Beispiel für eine medizinisch verwendete Substanz ist das Aryltetralin-Lignan Podophyllotoxin und seine Derivate, die cytotoxische Aktivitäten aufweisen. *Podophyllum* spec. sind derzeit die Hauptquelle für Podophyllotoxin. Aryltetralin-Lignane können jedoch auch in einigen Flachsarten nachgewiesen werden (*Linum* spec.). Die Biosynthese von Aryltetralin-Lignanen ist in die frühen und späten Schritte unterteilt. Während die frühen Reaktionsschritte vollständig identifiziert wurden, sind die meisten Enzyme und Reaktionen sowie ihre kodierenden Gene hinsichtlich der späten Schritte unbekannt. In der vorliegenden Arbeit werden Untersuchungen zur Biosynthese von Lignanen wie Podophyllotoxin oder 6-Methoxypodophyllotoxin in Suspensionskulturen von *Linum flavum* beschrieben. Die Rollen der Pinoresinol-Lariciresinol Reduktase, der Secoisolariciresinol Dehydrogenase, der Desoxypodophyllotoxin 6-Hydroxylase und der Desoxypodophyllotoxin 7-Hydroxylase sind von besonderem Interesse. Desoxypodophyllotoxin 6-Hydroxylase wurde bereits in *Linum flavum* untersucht und gehört zur Cytochrom P450 Familie. Daher war die Identifizierung der NADPH:Cytochrom P450 Reduktase, die für Cytochrom-P450-abhängige Reaktionen unerlässlich ist, auch ein Ziel dieser Arbeit.

Als Basis für die Suche nach Genen der Lignan-Biosythese diente die Transkriptom-Datenbank des 1KP Projekts (https://onekp.com/). Zu Beginn wurden RNA und gDNA aus Suspensionskulturen von *Linum flavum* isoliert. Mithilfe der cDNA wurden zwei NADPH:Cytochrom P450 Reduktasen erfolgreich identifiziert, heterolog exprimiert und charakterisiert. Zur Bestimmung der apparenten K_m-Werte wurden Enzymtests mit unterschiedlichen Konzentrationen von NADPH und Cytochrom c durchgeführt. Beide NADPH:Cytochrom P450 Reduktasen 66401 und 4753 zeigten eine hohe enzymatische Aktivität mit Cytochrom c mit K_m-Werten von 8,15 ± 0,3 µM bzw. 15,6 ± 0,35 µM. Gegenüber NADPH als Elektronendonor zeigte NADPH:Cytochrom P450 Reduktase 66401 einen K_m-Wert von 29,6 ± 0,8 µM und NADPH:Cytochrom P450 Reduktase 4753 einen K_m-Wert von 45,2 ± 0,7 µM. In den Sequenzen der beiden NADPH:Cytochrom P450 Reduktasen wurden jeweils zwei Bindungsstellen für FMN, FAD und NADPH und eine Bindungsstelle für Cytochrom P450 identifiziert. Basierend auf den Membranankerregionen am *N*-Terminus

werden NADPH:Cytochrom P450 Reduktase 66401 und 4753 gemäß Ro et al. (2002) in Klasse I bzw. II klassifiziert.

Pinoresinol-Lariciresinol Reduktase (PLR) ist ein bifunktionelles Enzym, das die Bildung von Secoisolariciresinol aus Pinoresinol über Lariciresinol katalysiert. Aus dem Transkriptom von *Linum flavum* wurde ein Contig mit der höchsten Ähnlichkeit mit Pinoresinol-Lariciresinol Reduktase von *Forsythia intermedia* gefunden. Dieses Contig wurde erfolgreich aus der RNA/cDNA von Suspensionskulturen von *Linum flavum* isoliert, in *E. coli* exprimiert und anschließend als Pinoresinol-Lariciresinol Reduktase charakterisiert. Die Pinoresinol-Lariciresinol Reduktase von *Linum flavum* katalysiert die stereospezifische Umwandlung von (+)-Pinoresinol über (+)-Lariciresinol in (-)-Secoisolariciresinol. PLR zeigte einen K_m-Wert für NADPH von 22,4 ± 1,2 µM und hatte die höchste katalytische Aktivität bei pH 7 und bei einer Temperatur von 27 °C bis 33 °C. Um die Rolle der Größe wichtiger Aminosäuren in der Bindungstasche zu untersuchen, wurden zwei Mutanten von PLR, G280Y und Y284G, konstruiert. Der Verlust der enzymatischen Aktivitär beider Mutanten der PLR, G280Y und Y284G, zeigte die entscheidende Rolle der beiden Positionen G280 und Y284 für die katalytische Aktivität dieses Enzyms.

Darüber hinaus wurden mehrere Kandidaten für Cytochrom P450 (CYP) und Secoisolariciresinol Dehydrogenase (SDH) aus dem Transkriptom von *Linum flavum* erhalten. Elf offene Leserahmen von CYP und fünf offene Leserahmen von SDH wurden erfolgreich aus RNA/cDNA oder gDNA von *Linum flavum* amplifiziert. CYP wurde erfolgreich in *Saccharomyces cerevisiae* und SDH in *E. coli* und Hefe exprimiert. Die konservierten Domänen für Cytochrom P450 und Secoisolariciresinol Dehydrogenase wurden in den Sequenzen jedes Kandidaten beobachtet. Trotzdem waren die rekombinanten Proteine mit den getesteten mutmaßlichen Substraten nicht katalytisch aktiv. Möglicherweise ist eine posttranslationale Modifikation für diese Enzyme unerlässlich und die verwendeten Bakterien und Hefen reichen nicht aus, um die erforderlichen Modifikationen durchzuführen. Ein besser geeignetes Expressionssystem, zum Beispiel in Pflanzen- oder Insektenkulturen, könnte eine Lösung für dieses Problem sein. Die Suche nach Kandidaten sollte verfeinert werden, um auch nach anderen Kandidaten für Cytochrom P450 und Secoisolariciresinol Dehydrogenase zu suchen.

IX. References

Adlercreutz H. (1999): Phytoestrogens. State of the art. Environ. Toxicol. Phar. 7, 201-207

Aebi M., Hornig H., Padgett R.A., Reiser J., Weissmann C. (1986): Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. Cell 47, 555-565

Alfermann A.W., Petersen M. (1993): Plant cell cultures. Biological Fundamentals, Sahm H. (ed.), VCH Verlagsgesellschaft, 577-614

Aliverti A., Pandini V., Pennati A., de Rosa M., Zanetti G. (2008): Structural and functional diversity of ferredoxin-NADP(+) reductases. Arch. Biochem. Biophys. 474, 283-291

Antipov E., Cho A.E., Klibanov A.M. (2009): How a single-point mutation in horseradish peroxidase markedly enhances enantioselectivity. J. Am. Chem. Soc. 131, 11155-11160

Arneaud S.L.B., Porter J.R. (2015): Investigation and expression of the secoisolariciresinol dehydrogenase gene involved in podophyllotoxin biosynthesis. Mol. Biotechnol. 57, 961-973

Bak S., Kahn R. A., Nielsen H. L., Möller B. L., Halkier B. A. (1998): Cloning of three A-type cytochromes P450, CYP71E1, CYP98, and CYP99 from *Sorghum bicolor* (L.) Moench by a PCR approach and identification by expression in *Escherichia coli* of CYP71E1 as a multifunctional cytochrome P450 in the biosynthesis of the cyanogenic glucoside dhurrin. Plant. Mol. Biol. 36, 393-405

Bayliss M.W. (1973): Origin of chromosome number variation in cultured plant cells. Nature 246, 529-530

Benveniste I., Lesot A., Hasenfratz G. (1991): Multiple forms of NADPH-cytochrome P450 reductase in higher plants. Biochem. Biophys. Res. Commun. 177, 105-112

Berlin J., Wray V., Mollenschott C., Sasse F. (1986): Formation of β-peltatin-A-methylether and coniferin by root cultures of *Linum flavum*. J. Nat. Prod. 49, 435-439

Bertani G. (1951): Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. J. Bacteriol. 62, 293-300

Bolwell G.P., Bozak K., Zimmerlin A. (1994): Plant cytochrome P450. Phytochem. 37, 1491-1506

Bonina F.P., Puglia C., Cimino F. (2005): Oxidative stress in handball players: effect of supplementation with a red orange extract. Nutr. Res. 25, 917-924

Bozak K.R., Yu H., Sirevag R., Christoffersen R.E. (1990): Sequence analysis of ripening-related cytochrome P-450 cDNAs from avocado fruit. Proc. Natl. Acad. Sci USA 87, 3904-3908

Bradford M.M. (1976): A rapid and sensitive method for the quantication of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248-254

Broomhead A. J., Dewick P. M. (1990): Aryltetralin lignans from *Linum flavum* and *Linum capitatum*. Phytochem. 29, 3839-3844

Broomhead A. J., Rahman M. M. A., Dewick P. M., Jackson D. E., Lucas J. A. (1991): Matairesinol as a precursor of *Podophyllum* lignans. Phytochem. 30, 14 89-1492

Canel C., Moraes R. M., Dayan F. E., Ferreira D. (2000): Molecules of interest: Podophyllotoxin. Phytochem. 54, 115-120

Cho J. Y., Kim A.R., Yoo E. S., Baik K. U., Park M. H. (1999): Immunomodulatory effect of arctigenin, a lignan compound, on tumour necrosis factor-alpha and nitric oxide production, and lymphocyte proliferation. J. Pharm. Pharmacol. 51, 1267-1273

Chomczynski P., Sacchi N. (1987): Single-Step Method of RNA Isolation by Acid Guanidinium Thiocyanate-Phenol-Chloroform Extraction. Anal. Biochem. 162, 156-159

Chu A., Dinkova A., Davin L.B., Bedgar D. L., Lewis N. G. (1993): Stereospecificity of (+)-pinoresinol and (+)-lariciresinol reductases from *Forsythia intermedia*. J. Biol. Chem. 268, 27026-27033

Coorssen J.R., Blank P.S., Albertorio F. (2002): Quantitative femto- to attomole immunodetection of regulated secretory vesicle proteins critical to exocytosis. Anal. Biochem. 307(1), 54-62

Davin L. B., Lewis N. G. (2000): Dirigent proteins and dirigent sites explain the mystery of specificity of radical precursor coupling in lignan and lignin biosynthesis. Plant. Physiol. 123, 453-462

Davin L.B., Lewis N.G. (2003): An historical perspective on lignan biosynthesis: Monolignol, allylphenol and hydroxycinnamic acid coupling and downstream metabolism. Phytochem. Rev. 2, 257-288

Davin L.B., Wang H.-B., Crowell A.L., Bedgar D.L., Martin D.M., Sarkanen S., Lewis N.G. (1997): Stereoselective bimolecular phenoxy radical coupling by an auxiliary (dirigent) protein without an active center. Science 275, 362-266

Davis P. H. (1970): *Linum*. In: Davis P. H. (ed), Flora of Turkey and East Aegean Islands, Vol. 2, Edinburgh University Press, Edinburgh

De Vetten N., Horst T.J., van Schaik H.P., De Boer A., Mol J., Koes R. (1999): A cytochrome b5 is required for full activity of the flavonoid 3', 5'-hydroxylase, a cytochrome P450 involved in the formation of blue flower colors. Proc. Natl. Acad. Sci. USA 96, 778-783

Deus-Neumann B., Zenk M.H. (1984): Instability of indole alkaloid production in *Catharanthus roseus* cell suspension cultures. Planta. Med. 50, 427-431

Deyama T., Nishibe S. (2010): Pharmacological properties of lignans. In: Heitner C, Dimmel DR, Schmidt JA (eds) Lignin and lignans, CRC Press, Boca Raton, 585-629

Dietrich R.C., Fuller J.D., Padgett R.A. (2005): A mutational analysis of U12-dependent splice site dinucleotides. RNA 11, 1430-1440.

Dietrich R.C., Incorvaia R., Padgett R.A. (1997): Terminal intron dinucleotide sequences do not distinguish between U2-and U12-dependent introns. Mol. Cell 1, 151-160

Dinkova-Kostova A. T., Gang D. R., Davin L. B., Bedgar D. L., Chu A., Lewis N. G. (1996): (+)-Pinoresinol/(+)-lariciresinol reductase from *Forsythia intermedia*. Protein purification, cDNA cloning, heterologous expression and comparison to isoflavone reductase. J. Biol. Chem. 271, 29473-29482

Durst F. (1988): Biochemistry and physiology of Plant Cytochrome P450. In: Frontiers of Biotransformation 4, Microbial and Plant Cytochrome P450: Biochemical Characteristics, Genetic Engineering and Practical Purification, Academic Verlag, Berlin, 192-210

Durst F., Nelson D.R. (1995): Diversity and evolution of plant P450 and P450-reductases. Drug Metabol. Drug Interact. 12, 189-268

Eberle D., Ullmann P., Werck-Reichhart D., Petersen M. (2009): cDNA cloning and functional characterisation of CYP98A14 and NADPH:cytochrome P450 reductase from *Coleus blumei* involved in rosmarinic acid biosynthesis. Plant Mol. Biol. 69, 239-253

Empt U., Alfermann A.W., Pras N., Petersen M. (2000): The use of plant cell cultures for the production of podophyllotoxin and related lignans. J. Appl. Bot. 74, 145-150

Erich O. (2001): Pflanzensoziologische Exkursionsflora für Deutschland und angrenzende Gebiete, Verlag Eugen Ulmer, Stuttgart, 631

Federolf K., Alfermann A.W., Fuss E. (2007): Aryltetralin-lignan formation in two different cell suspension cultures of *Linum album*: deoxypodophyllotoxin 6-hydroxylase, a key enzyme for the formation of 6-methoxypodophyllotoxin. Phytochem. 68(10), 1397-406

Fraser C.M., Chapple C. (2011): The phenylpropanoid pathway in *Arabidopsis*. The arabidopsis book (9), e0152

Fujita M., Gang D.R., Davin L.B., Lewis N.G. (1999): Recombinant pinoresinol-lariciresinol reductase from Western Red Cedar (*Thuja plicata*) catalyze opposite enantiospecific conversions. J. Biol. Chem. 274, 618-627

Gasteiger J., Schunk A., (2003): CCC Univ. Erlangen, GMT BMBF-Leitprojekt Vernetztes Studium - Chemie. www2.chemie.uni-erlangen.de/projects/vsc/chem

Gietz R.D., Schiestl R.H. (2007): High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. 2, 31-34

Graham S.E., Peterson J.A. (1999): How similar are P450s and what can their differences teach us? Arch. Biochem. Biophys. 369, 24-29.

Grosdidier A., Zoete V., Michielin O. (2007): Docking of small molecules into protein active sites with a multiobjective evolutionary optimization. Proteins 67(4), 1010-1025

Groves J.T., Han Y.-Z. (1995): Models and mechanism of cytochrome P450 action. In: Cytochrome P450 - Structure, Mechanism and Biochemistry (Hrsg.: P.R. Ortiz de Montellano), Plenum Press, New York, 3-48

Haas E., Horecker B.L., Hogness T.R. (1940): The enzymatic reduction of cytochrome c. J. biol. Chem. 136, 747-774

Halkier B.A. (1996): Catalytic reactivities and structure/function relationships of cytochrome P450 enzymes. Phytochem. 43, 1-21

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

Heimendahl C.B., Schafer K.M., Eklund P., Sjoholm R., Schmidt T.J. and Fuss E. (2005): Pinoresinol-lariciresinol reductases with different stereospecificity from *Linum album* and *Linum usitatissimum.* Phytochem. 66 (11), 1254-1263

Heldt H.W. (1999): Pflanzenbiochemie, Heidelberg, Berlin, Sektrum Akad. Verlag

Henges A. (1999): Biosynthese und Kompartimentierung von Lignanen in Zellkulturen von *Linum album*. Dissertation, Heinrich-Heine-Universität Düsseldorf

Humphreys J.M. and Chapple C. (2002): Rewriting the lignin roadmap. Current Opinion in Plant Biology 5, 224-229

Imbert T. F. (1998): Discovery of podophyllotoxins. Biochem. 80, 207-222

Jackson D. E., Dewick P. M. (1984): Aryltetralin lignans from *Podophyllum hexandrum* and *Podophyllum peltatum*. Phytochem. 23, 1147-1152

Jennewein S., Rithner C.D., Williams R.M., Croteau R. B. (2001): Taxol biosynthesis: Taxane 13α-hydroxylase is a cytochrome P450-dependent monooxygenase. Proc. Natl. Acad. Sci. USA 98, 13595-13600.

Jörnvall H., Persson B., Krook M., Atrian S., Gonzalez-Duarte R., Jeffery J., Ghosh D. (1995): Short-chain dehydrogenases/reductases (SDR). Biochem. 34, 6003-6013

Jörnvall H., Persson M., Jeffery J. (1981): Alcohol and polyol dehydrogenases are both divided into two protein types, and structural properties cross-relate the different enzyme activities within each type. Proc. Natl. Acad. Sci. USA 78, 4226-4230

Kallberg Y., Oppermann U., Persson B (2010): Classification of the short-chain dehydrogenase/reductase superfamily using hidden Markov models. The FEBS J., 277(10): 2375-86

Kamil W. M., Dewick P. M. (1986): Biosynthesis of *Podophyllum* lignans. 3. Biosynthesis of the lignans alpha-peltatin and beta-peltatin. Phytochem. 25, 2089-2092

Kamil W.M., Dewick P.M. (1986): Biosynthesis of *Podophyllum* lignans. Biosynthesis of the lignans alpha-peltatin and beta-peltatin. Phytochem. 25, 2089-2092

Katayama T., Davin L.B., Lewis N.G. (1992): An extraordinary accumulation of (-)-pinoresinol in cell free extracts of *Forsythia intermedia*: evidence for enantiospecific reduction of (+)-pinoresinol. Phytochem. 31, 3875-3881

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015): The Phyre2 web portal for protein modeling, prediction and analysis. Nature Prot. 10, 845-858

Kemper B. (2004): Structural basis for the role in protein folding of conserved proline-rich regions in cytochromes P450. Toxicol. Appl. Phar. 199, 305-315

Kenneth J. B., Lindberg M. (2010): Plant NADPH-cytochrome P450 oxidoreductases. Phytochem. 71, 132-141

Kleinig H. und Mayer U. (1999): Zellbiologie, G. Fischer Verlag

Klingenberg M. (1958): Pigments of rat liver microsomes. Arch. Biochem. Biophys. 75, 376-386.

Koopmann E, Hahlbrock K (1997): Differentially regulated NADPH:cytochrome P450 oxidoreductases in parsley. Proc. Natl. Acad. Sci. USA 94, 14954-14959

Koulman A. (2003): Podophyllotoxin. Dissertation, Rijksuniversiteit Groningen, Niederlande

Koulman A., Bos R., Medarde M., Pras N., Quax W.J. (2001): A fast and simple GC MS method for lignan profiling in *Anthriscus sylvestris* and biosynthetically related plant species. Planta Medica 67, 858-862.

Kranz K., Petersen M., (2003): β-peltatin 6-O-methyltransferase from cell suspension cultures of *Linum nodiflorum*. Phytochem. 64, 453-458

Kuhlmann S. (2004): Untersuchungen zur Rolle von Cytochrom P450-Enzymen in der Biosynthese von Aryltetralin-Lignanen in Zellkulturen von *Linum* spec. Dissertation, Philipps-Universität Marburg

Kuo H.J., Wei Z.Y., Lu P.C., Huang P.L., Lee K.T. (2014): Bioconversion of pinoresinol into matairesinol by use of recombinant *Escherichia coli.* Appl. Environ. Microbiol. 80(9), 2687-2692

Lau W., Sattely E.S. (2015): Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. Science 349, 1224

Lewis N. G., Davin L. B. (1999): Lignans: biosynthesis and function. Compr. Nat. Prod. Chem., 639-712

Lloyd J. U. (1910): The eclectic alkaloids, resins, resinoids, oleo-resins and concentrated principles. Bulletin of the Lloyd Library of Botany, Pharmacy and Materia Medica, Pharmacy series No.2. Cincinnati, Ohio, USA

Long M., Rosenberg C., Gilbert W. (1995). Intron phase correlations and the evolution of the intron/exon structure of genes. Proc. Natl. Acad. Sci. USA 92(26), 12495-12499

Marques J., Kim K.W., Lee C., Costa M., May G., Crow J., Davin L., Lewis G.N. (2012): Next Generation Sequencing in Predicting Gene Function in Podophyllotoxin Biosynthesis. J. Biol. Chem. 288(1), 466-79.

Marques J.V., Kim KW., Lee C., Costa M.A., May G.D., Crow J.A., Davin L.B., Lewis N.G. (2013): Next generation sequencing in predicting gene function in podophyllotoxin biosynthesis. J. Biol. Chem. 288, 466-479

Meunier B., de Visser SP, Shaik S. (2004): Mechanism of oxidation reactions catalyzed by cytochrome p450 enzymes. Chem. Rev. 104(9), 3947-80

Min T., Kasahara H., Bedgar D. L., Youn B., Lawrence P. K., Gang D. R., Halls S. C., Park H., Hilsenbeck J. L., Davin L. B., Lewis N. G., Kang C. (2003): Crystal structures of pinoresinol-lariciresinol and phenylcoumaran benzylic ether reductases and their relationship to isoflavone reductases. J. Biol. Chem. 278 (50), 50714-50723

Mizutani M, Ohta D (1998): Two isoforms of NADPH:cytochrome P450 reductase in *Arabidopsis thaliana*. Gene structure, heterologous expression in insect cells, and differential regulation. Plant. Physiol. 116, 357-367

Mizutani M. (2012): Impacts of diversification of cytochrome P450 on plant metabolism. Biol Pharm. Bull. 35(6), 824-32.

Mohagheghzadeh A. (2003): Evaluation of lignans in *Linum persicum.* Phytochem. Rev., in press

Moinuddin S. G., Youn B., Bedgar D. L., Costa M. A., Helms G. L., Kang C. (2006): Secoisolariciresinol dehydrogenase: Mode of catalysis and stereospecificity of hydride transfer in *Podophyllum peltatum*. Org. Biomol. Chem. 4, 808-816

Molog G.A., Empt U., Kuhlmann S., Van Uden W., Pras N., Alfermann A.W., Petersen M. (2001): Deoxypodophyllotoxin 6-hydroxylase, a cytochrome P450 monooxygenase from cell cultures of *Linum flavum* involved in the biosynthesis of cytotoxic lignans. Planta 214, 288-294

Murphy R.M., Lamb G.D. (2013): Important considerations for protein analyses using antibody based techniques: down-sizing Western blotting up-sizes outcomes. J. Phys. 591, 5823-5831.

Nelson D.R., Koymans L., Kamataki T., Stegeman J.J., Feyereisen R., Waxman D.., Waterman M.R., Gotoh O., Coon M.J., Estabrook R.W., Gunsalus I.C., Nebert D.W., (1996): P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes and nomenclature. Pharmacogentics 6, 1-42

Ockendon D. J., Walters S. M. (1968): *Linum*. In: Tutin T. G., Heywood V. H., Burges N. A., Moorre D. M., Valentine D. H., Walters S. M., Webb D. A. (eds), Cambridge University Press. Flora Europaea 2, 206-211

Omura T., Sato R., (1964): The carbon monoxide pigment of liver microsomes. I Evidence for its hemoprotein nature. J. Biol. Chem. 239, 2370-2378

Petersen M., Alfermann A.W. (2001): The production of cytotoxic lignans by plant cell cultures, Appl. Microbiol. Biotechnol. 55, 135-142

Phillips R.L., Kaeppler S.M., Olhoft P. (1994): Genetic instability of plant tissue cultures: breakdown of normal controls, Proc. Natl. Acad. Sci. USA 91, 5222-5226

Porter T.D., Kasper C.B. (1986): NADPH-cytochrome P-450 oxidoreductase: flavin mononucleotide and flavin adenine dinucleotide domains evolved from different flavoproteins. Biochem. 25, 1682-1687

Porter TD (2004): Jud Coon: 35 years of P450 research, a synopsis of P450 history. Drug. Metab. Dispos. 32, 1-6

Rana S., Lattoo S.K., Dhar N., Razdan S., Bhat W.W., Dhar R.S, Vishwakarma R. (2013): NADPH-Cytochrome P450 Reductase: Molecular Cloning and Functional Characterization of Two Paralogs from *Withania somnifera* (L.) Dunal. PLoS. ONE 8(2): e57068

Ro D.K., Ehlting J., Douglas C.J. (2002): Cloning, functional expression, and subcellular localization of multiple NADPH-cytochrome P450 reductases from hybrid poplar. Plant Physiol. 130, 1837-1851

Robinson J. (2018): Untersuchungen von Methyltransferasen aus *Linum*-arten. Dissertation, Philipps-Universität Marburg

Rogers S.O., Bendich A.J. (1985): Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. Plant Mol. Biol. 5 (2), 69-76

Sakakibara N., Suzuki S., Umezawa T., Shimada M. (2003): Biosynthesis of yatein in *Anthriscus sylvestris*. Org. Biomol. Chem. 1 (14), 2474-2485

Sakurai A., Fujimori S., Kochiwa H., Kitamura-Abe S., Washio T., Saito R. (2002): On biased distribution of introns in various eukaryotes. Gene 300, 89-95

Schuler M.A. (1996): Plant cytochrome P450 monooxygenases, Crit. Rev. Plant Sci. 15, 235-284

Schuler M.A., Werck-Reichhart D. (2003): Functional genomics of P450s. Annual Review of Plant Biol. 54, 629-667

Seidel V., Windhovel J., Eaton G., Alfermann A.W., Arroo R.R., Medarde M., Petersen M., Woolley J.G. (2002): Biosynthesis of podophyllotoxin in *Linum album* cell cultures. Planta 215, 1031-1039

Sheth N., Roca X., Hastings M.L., Roeder T., Krainer A.R., Sachidanandam R. (2006) Comprehensive splice-site analysis using comparative genomics. Nucleic Acids Res. 34, 3955-3967.

Shiraishi A., Murata J., Matsumoto E., Matsubara S., Ono E., Satake H. (2016): DeNovo Transcriptomes of *Forsythia koreana* Using a Novel Assembly Method: Insight into Tissue- and Species-Specific Expression of Lignan Biosynthesis-Related Gene. PLoS. ONE 11(10), e0164805

Simon H., Jelitto L., Schacht W. (2002): Die Freiland-Schmuckstauden. Handbuch und Lexikon der Gartenstauden, 5. Auflage, Eugen Ulmer, Stuttgart

Sirim D., Widmann M., Wagner F., Pleiss J. (2010): Prediction and analysis of the modular structure of cytochrome P450 monooxygenases. BMC Struct. Biol. 10:34

Smollny T., Wichers H., Kalenberg S., Shahsavari A., Petersen M., Alfermann A.W. (1998): Accumulation of podophyllotoxin and related lignans in cell suspension cultures of *Linum album.* Phytochem. 48, 975-979

Stoscheck C.M. (1990): Quantitation of Protein. Methods in Enzymology 182, 50-69

Sunderland N. (1977): Nuclear cytology, in: Plant Tissue and Cell Culture (Street H.E., Ed.). Oxford: Blackwell Scientific Publ., 177-205

Szafranski K., Schindler S., Taudien S., Hiller M., Huse K., Jahn N., Schreiber S., Backofen R., Platzer M. (2007): Violating the splicing rules: TG dinucleotides function as alternative 3' splice sites in U2-dependent introns. Genome Biol. 8, R154.

Tomita M., Shimizu N., Brutlag D.L. (1996): Introns and Reading Frames: Correlation Between Splicing Sites and Their Codon Positions. Mol Biol Evol. 13, 1219-1223

Umezawa T. (2003): Diversity in lignan biosynthesis. Phytochem. Rev. 2, 371-390

Umezawa T., Davin L. B., Lewis N. G. (1991): Formation of lignans (-)-secoisolariciresinol and (-)-matairesinol with *Forsythia intermedia* cell-free extracts. J. Biol. Chem. 266, 10210-10217

Urban P, Mignotte C, Kazmeier M et al (1997): Cloning, yeast expression, and characterization of the coupling of two distantly related *Arabidopsis thaliana* NADPH-cytochrome P450 reductases with P450 CYP73A5. J. Biol. Chem. 272, 19176-19186

Urban P., Cullin C., Pompon D. (1990): Maximizing the expression of mammalien cytochrome P450 monooxygenase activities in yeast cells. Biochimie 72, 463-472

Urban P., Werck-reichhart D., Teutsch H.G., Durst F., Regnier S., Kazmeier M., Pompon D. (1994): Characterization of recombinant plant cinnamate 4-hydroxylase produced in yeast. Kinetic and spectral properties of the major plant P450 of the phenylpropanoid pathway. Eur. J. Biochem. 222, 843-850

Van Uden W., Bos J.A., Boeke G.M., Woerdenbag H.J., Pras N. (1997): The large-scale isolation of deoxypodophyllotoxin from rhizomes of *Anthriscus sylvestris* followed by its bioconversion into 5-Methoxypodophyllotoxin betaD-glucoside by cell-cultures of *Linum flavum*. J. Nat. Prod. 60, 401-403

Van Uden W., Bouma A.S., Waker J.F.B., Middel O., Wichers H.J., Dewaard P., Woerdenbag H.J., Kellogg R.M., Pras N. (1995): The production of podophyllotoxin and its 5-methoxy derivative through bioconversion of cyclodextrin-complexed desoxypodophyllotoxin by plant cell-cultures. Plant Cell Tiss. Org. 42, 73-79

Van Uden W., Homan B., Woerdenbag H.J., Pras N., Malingre T.M., Wichers H.J., Harkes M. (1992): Isolation, purification, and cytotoxicity of 5-methoxypodophyllotoxin, a lignan from a root culture of *Linum flavum*. J. Nat. Prod. 55, 102-110

Van Uden W., Pras N., Woerdenbag H. J. (1994): *Linum* species (Flax): In vivo and in vitro accumulation of lignans and other metabolites, Springer Verlag, Berlin. Biotechnology in agriculture and forestry 26, 219-244

Vibranovski M.D., Sakabe N.J., de Souza S.J. (2006): A possible role of exon-shuffling in the evolution of signal peptides of human proteins. FEBS Lett. 580, 1621-1624.

Wang M., Roberts D.L., Paschke R., Shea T.M., Masters B.S., Kim J.J. (1997): Three-dimensional structure of NADPH-cytochrome P450 reductase: prototype for FMN- and FAD-containing enzymes Proc. Natl. Acad. Sci. USA 94, 8411-8416

Webster D.E., Thomas M.C. (2012): Post-translational modification of plant-made foreign proteins; glycosylation and beyond. Biotechnol. Adv. 30, 410-418

Weiss S. G., Tin-Wa M., Perdue R. E., Farnsworth N. R. (1975): Potential anticancer agents II: antitumor and cytotoxic lignans from *Linum album* (Linaceae). J. Pharm. Sci. 64, 95-98

Werck-Reichhart D., Bark S., Paquette S. (2002): Cytochrome P450. In: The *Arabidopsis* Book. American Society of Plant Biologists

Werck-Reichhart D., Feyereisen R. (2000): Cytochrome P450: a success story. Genome Biol. 1, 3003.1- 3003.9

Williams Ch. Jr., Kamin H. (1962): Microsomal triphosphopyridine nucleotide-cytochrome c reductase of liver. J. Biol. Chem. 237, 587-95

Xia Z. Q., Costa M. A., Pelissier H. C., Davin L. B., Lewis N. G. (2001): Secoisolariciresinol dehydrogenase purification, cloning and functional expression. Implications for human health protection. J. Biol. Chem. 276, 12614-12623

Yates C. M., Filippis I., Kelley LA, Sternberg J.E. (2014): SuSPect: Enhanced Prediction of Single Amino Acid Variant (SAV) Phenotype Using Network Features. J. Mol. Biol. 426(14), 2692-2701

Curriculum vitae

Personal data

Full name: Thanh Son Ta

Place of birth: Kon Tum - Viet Nam

Date of birth: 03-05-1990

Nationality: Vietnamese

Educational background

2005-2008:	Senior high school Hung Vuong, Gia Lai - Viet Nam
2008-2009:	Faculty of Pharmacy, University of Medicine and Pharmacy, Ho Chi Minh City - Viet Nam
2009-2010:	Faculty of German language, University Hanoi, Ha Noi - Viet Nam
2010-2011:	Studienkolleg Mittelhessen, Marburg - Germany Graduation: High school diploma for foreign students
2011-2015:	Faculty of Pharmacy, Philipps-Universität Marburg, Marburg - Germany Graduation: 2nd State Examination
2016-2019:	PhD at Institute for Pharmaceutical Biology and Biotechnology, Philipps-Universität Marburg, Marburg - Germany