Evaluation of C-prenylating enzymes for the heterologous

biosynthesis of cannabigerolic acid

Zur Erlangung des akademischen Grades eines

Dr. rer. nat.

von der Fakultät Bio- und Chemieingenieurwesen
der Technischen Universität Dortmund
genehmigte Dissertation

vorgelegt von

M.Sc. Sara Friederike Degenhardt

aus

Mainz

Tag der mündlichen Prüfung: 09.10.2018

1. Gutachter: Prof. Dr. Dr. h.c. Oliver Kayser

2. Gutachter: Prof. Dr. Heribert Warzecha

Dortmund 2018

Acknowledgements

Thank you to **Jan Schonert** (CCB, TU Dortmund, Germany) for performing NMR measurements and **Nico Zammarelli** for help with GPP synthesis.

I am grateful to my colleagues from the Biozentrum. I enjoyed fruitful discussions, mental support and the great time we spent together in the lab and after work, especially with **Bastian Zirpel**, **Laura Kohnen**, **Evamaria Gruchattka**, **Nouman Almofti**, **Angela Sester**, **Jenny Schwarz**, **Anna Tippelt**, **Friederike Ullrich**, **Julia Schachtsiek, Kathleen Pamplanyil** and **Torsten Arndt**.

Last but not least I want to especially thank my **family**, **friends** and "**Sniffles**" for motivation, encouragement and the best support I could wish for.

Table of contents

Abstract

Aromatic prenyltransferases catalyze the prenylation of an electron-rich aromatic compound by a prenyl diphosphate by transfer a prenyl residue to carbon, nitrogen or an oxygen atom of the aromatic acceptor molecule. The bioactivity of the resulting products is often increased, compared to the non-prenylated substrates. Prenylation of aromatic compounds is important for the diversification of secondary metabolites like phytocannabinoids. The membrane-bound cannabigerolic acid synthase from *Cannabis sativa* catalyzes the prenylation of olivetolic acid (OA) with geranyl diphosphate (GPP) to form cannabigerolic acid (CBGA). CBGA is the central precursor of the main pharmacologically active compounds of *C. sativa:* Δ^9-tetrahydrocannabinol (THC) and cannabidiol (CBD). A biotechnological production using heterologous yeast presents a promising alternative to meet the increasing pharmaceutical demand and interest in these cannabinoids. In this regard, this study aimed for the identification of potential aromatic prenyltransferases able to produce CBGA and their heterologous expression.

Three different membrane-bound aromatic prenyltransferases, isolated from *C. sativa,* were expressed in *Saccharomyces cerevisiae*, but were not able to produce CBGA. Furthermore, the soluble aromatic prenyltransferase NphB, first isolated from *Streptomyces* sp. strain CL190, was investigated as an alternative. Beside the desired product CBGA, NphB forms a major side-product, 2-*O*-geranyl olivetolic acid. Therefore, NphB variants were tested in a low-throughput screening system in order to improve the product specificity towards CBGA and to enable higher specific activities. Additionally, the simultaneous functional expression of *nphB* and the final enzyme of the cannabinoid pathway, tetrahydrocannabinolic acid synthase (THCAS), in *S. cerevisiae* is presented.

This study presented for the first time that the soluble aromatic prenyltransferase NphB produces CBGA from precursors OA and GPP. The obtained results therefore promote the generation of a cannabinoid producing yeast for a biotechnological approach in order to face the increasing demand of the phytocannabinoids CBD and THC in the medical sector.

Zusammenfassung

Aromatische Prenyltransferasen biokatalysieren die Prenylierung eines Elektronen-reichen Aromaten mit Hilfe eines Prenyldiphosphates durch die Übertragung des Prenylrestes auf ein Kohlenstoff-, Stickstoff- oder Sauerstoffatom des aromatischen Akzetormoleküls. Prenylierte Substanzen weisen oft eine höhere Bioaktivität als die entsprechenden nicht-prenylierten Moleküle auf. Die Prenylierung von Aromaten ist wichtig für die Vielfalt von Sekundärmetaboliten wie Phytocannabinoiden. Die Membran-gebundene Cannabigerolsäure Synthase (CBGAS) aus *Cannabis sativa* biokatalysiert die Bildung von Cannabigerolsäure (CBGA) durch die Prenylierung von Olivetolsäure (OA) mit Geranyldiphosphat (GPP). CBGA ist das zentrale Intermediat für die pharmakologisch aktiven Substanzen Δ^9-Tetrahydrocannabinol (THC) und Cannabidiol (CBD) aus *C. sativa*. Eine biotechnologische Produktion mit Hilfe von heterologen Wirtszellen wie Hefe stellt eine vielversprechende Alternative dar, um dem steigenden pharmazeutischen Bedarf und Interesse an diesen Cannabinoiden gerecht zu werden. Im Hinblick darauf sollten im Rahmen dieser Arbeit potentielle aromatische Prenyltransferasen identifiziert und heterolog exprimiert werden, die in der Lage sind CBGA zu produzieren.

Drei unterschiedliche Membran-gebundene aromatische Prenyltransferasen aus *C. sativa* wurden in *Saccharomyces cerevisiae* exprimiert, es konnte aber keine CBGA Produktion nachgewiesen werden. Als Alternative zur CBGAS wurde die lösliche aromatische Prenyltransferase NphB aus *Streptomyces* sp. strain CL190 in Betracht gezogen. NphB bildet hauptsächlich 2-*O*-geranyl Olivetolsäure und CBGA nur als Nebenprodukt. Mit Hilfe eines Screening Systems sind verschiedene NphB Mutanten im Hinblick auf verbesserte Produktspezifitäten und höhere spezifische Aktivitäten untersucht worden. Des Weiteren wird die funktionale Expression von *nphB* sowie die co-Expression von *nphB* und dem finalen Enzym des Cannabinoid Biosyntheseweges, Tetrahydrocannabinolsäure Synthase (THCAS), in *S. cerevisiae* präsentiert.

Die vorliegende Arbeit zeigt zum ersten Mal, dass die lösliche aromatische Prenyltransferase NphB die Bildung von CBGA aus OA und GPP biokatalysiert. Die gezeigten Ergebnisse unterstützen die Entwicklung einer Cannabinoid-produzierenden Hefe für die biotechnologische Produktion der pharmazeutisch bedeutsamen Phytocannabinoide CBD und THC.

Chapter 1

Introduction

Parts of chapter 1.2 were published in:

Degenhardt, F., Stehle, F., Kayser, O., 2017. The biosynthesis of cannabinoids, in: Preedy, V. (Ed.), Handbook of Cannabis and Related Pathologies: Biology, Pharmacology, Diagnosis, and Treatment. Elsevier B.V., pp. 13–23.

1.1. Preface

Cannabis sativa L. is an important medical plant which belongs to the family of Cannabaceae. It is an annual dioecious plant that is also commonly referred to as: grass, hemp, Marijuana and weed (Fischedick et al., 2010; Sirikantaramas et al., 2007). The plant contains secondary metabolites that are called phytocannabinoids and belong to the chemical class of terpenophenolics. Phytocannabinoids are plant-derived natural compounds that bind to the human cannabinoid receptors CB_1 and CB_2 or share chemical similarity with cannabinoids or both (Gertsch et al., 2010; Sirikantaramas et al., 2007, 2005; Taura et al., 2007b). The main pharmacologically active compounds of *C. sativa* are Δ^9-tetrahydrocannabinol (Δ^9-THC) and cannabidiol (CBD). Medical *Cannabis* is used for the treatment of several diseases and symptoms like chronic neuropathic pain, nausea and vomiting associated with chemotherapy, inflammation, symptom relief of those affected by multiple sclerosis and glaucoma (Burstein, 2015; Maida and Daeninck, 2016; Novack, 2016; Schrot and Hubbard, 2016).

In planta, Δ^9-THC and CBD occur mainly in their acidic forms cannabidiolic acid (CBDA) and Δ^9-tetrahydrocannabinolic acid (Δ^9-THCA) (Taura et al., 2007b, 2007c). Chemical synthesis of natural products can be difficult. For example, in respect to the synthesis of THC the position of the double bond (Δ^8-THC vs. Δ^9-THC) and the *cis-trans* stereochemistry at the cyclohexene ring have to be taken in account. Additionally, GMP-grade organic solvents have to be used in pharmaceutical industry, which makes organic synthesis of phytochemicals very expensive and environmentally unfriendly (Chang and Keasling, 2006; Grodowska and Parczewski, 2010; Trost and Dogra, 2007). Another possibility to gain plant secondary metabolites like CBDA and THCA is isolating them from their natural source which is often inefficient and low yielding, because they accumulate only at rather low concentrations within the plant. Furthermore, the application of pesticides, large areas for plant cultivation and the use of organic solvents are required (Chang and Keasling, 2006; Moses et al., 2013). Additionally, *Cannabis* cultivation is illegal in many countries.

In order to cover the increasing demand for cannabinoids in the medical sector and to circumvent legal and environmental issues relating to *Cannabis* cultivation, synthetic biology should be taken into consideration. The genes of the biosynthetic pathway required for the production of e.g. THCA are transferred from *C. sativa* to a fermentable microbial host like *Escherichia coli* or *Saccharomyces cerevisiae* (Figure 1) (Awan et al., 2016; Chang and Keasling, 2006; Mitchell, 2011). For the expression of proteins involved in the biosynthesis of THCA yeast is favored because the microbial host and the natural source of the genes are both eukaryotes. In addition, THCA synthase possesses eight possible N-glycosylation sites and a disulfide-bond which might be important for functional expression, thus posttranslational modifications are necessary (Sirikantaramas et al., 2004). In 2015, Zirpel et al. showed that under the chosen conditions, functional expression in *E. coli* cells is not possible.
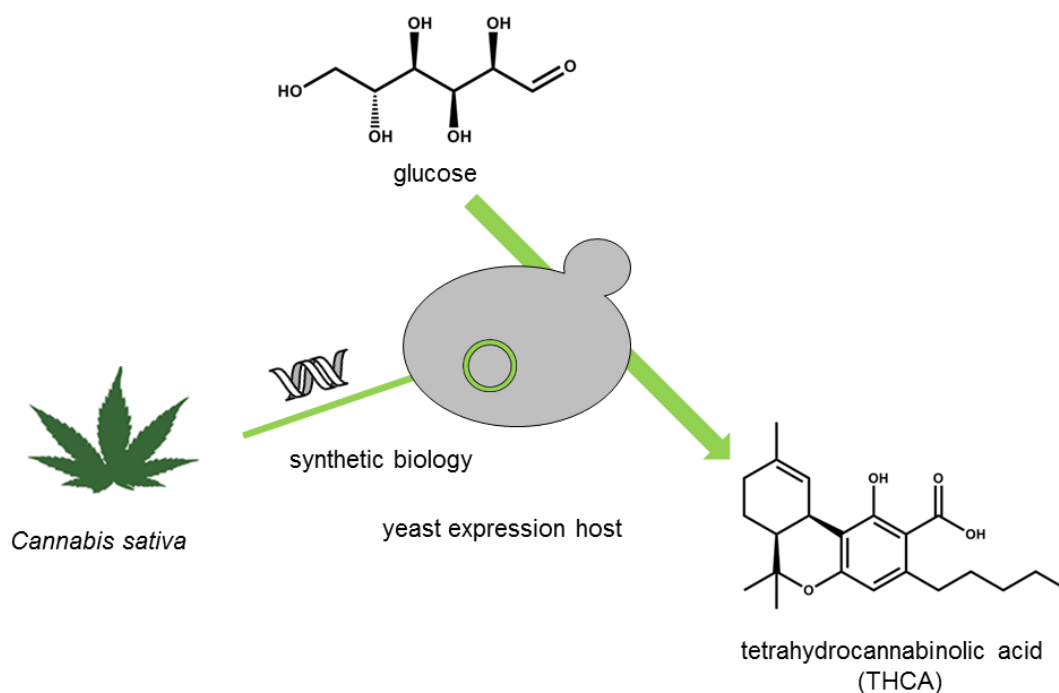
Figure 1: Transfer of the cannabinoid biosynthetic pathway from *C. sativa* into yeast. Modified from Zirpel, Degenhardt et al. (2017).The genes required for the biosynthetic pathway of tetrahydrocannabinolic acid (THCA) are transferred from *C. sativa* into yeast. The yeast cells are cultivated in the presence of a cheap carbon source such as glucose and THCA is synthesized in the heterologous yeast.

This thesis will cover the expression of cannabigerolic acid synthase (CBGAS) candidates from *C. sativa* in *S. cerevisiae*, and the expression of a soluble CBGAS alternative, namely NphB from *Streptomyces* sp. strain CL190, in *S. cerevisiae* and *E. coli*. Additionally, mutational studies of NphB were performed in *E. coli*. Finally, the simultaneous expression of NphB and tetrahydrocannabinolic acid synthase simultaneously is investigated.

1.2. The biosynthesis of cannabinoids

For a long time, cannabinoids were defined as secondary metabolites of *Cannabis sativa* L. which interact with the human cannabinoid receptors (CB_1 and CB_2). Since secondary metabolites from other plants were reported to interact with the human cannabinoid receptors as well, a new definition had to be established. Hence, phytocannabinoids are defined as any plant-derived natural compound which shares chemical similarity with cannabinoids or binds to the human cannabinoid receptors or both (Gertsch et al., 2010; Sirikantaramas et al., 2007, 2005; Taura et al., 2007b).

In 1994, Toyota et al. isolated the bibenzyl cannabinoid perrottetinene from the japanese liverwort *Radula perrottetti*. Besides the phytocannabinoid perrottetinene, the New Zealand liverwort *Radula marginata* contains a cannabinoid type bibenzyl named perrottetinenic acid (Toyota et al., 2002). Both compounds show structural similarity to Δ^9-tetrahydrocannabinolic acid (Δ^9-THCA) or Δ^9-tetrahydrocannabinol (Δ^9-THC), respectively (Figure 2). Another example of a phytocannabinoid is the sesquiterpene β-caryophyllene which binds to the CB_2 receptor (Figure 2). The compound is found

throughout the essential oils of plants like black pepper (*Piper nigrum*), basil (Ocimum spp.) and *C. sativa*. Thus, according to the definition, *C. sativa* biosynthesizes phytocannabinoids with at least two completely different chemical scaffolds (Gertsch et al., 2010, 2008; Hendriks et al., 1975).
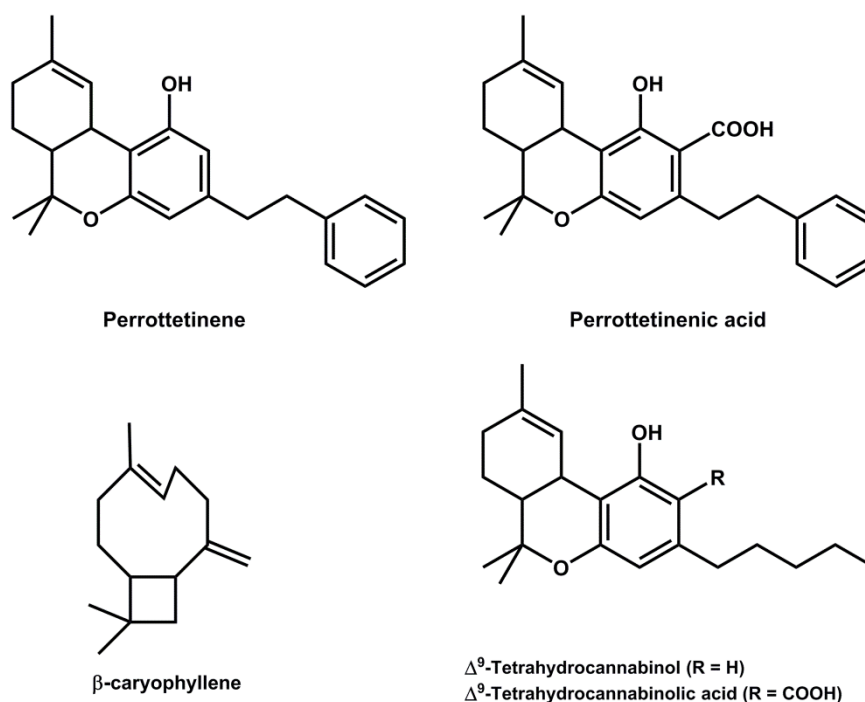
Figure 2: Constitutional formulas of four different phytocannabinoids.Perrottetinen, perrottetinenic acid, β-caryophyllene and tetrahydrocannabinol / tetrahydrocannabinolic acid.

Although more than 120 phytocannabinoids are known in *Cannabis*, only three enzymes, namely cannabidiolic acid synthase (CBDAS), cannabichromenic acid synthase (CBCAS) and tetrahydrocannabinolic acid synthase (THCAS), are involved in cannabinoid biosynthesis (Morales et al., 2017; Taura et al., 2007b). "The resulting acidic cannabinoids are the most abundant cannabinoids accumulating in *Cannabis*. The neutral and psychoactive forms are the results of non-enzymatic decarboxylation during storage, heat or sunlight; explaining the heating of plant material (i.e. smoking or baking) during *Cannabis* consumption (Fischedick et al., 2010; Taura et al., 2007b). Thus, the broad diversity of the different cannabinoids is mainly due to non-enzymatic transformation or degradation of both acidic and neutral cannabinoids by the effects of light (UV irradiation) and auto-oxidation (Crombie et al., 1968; Razdan et al., 1972). It is still unclear, if all these forms are present in living plants as natural or artefacts, due to storage and sample preparation (ElSohly and Slade, 2005)." (Degenhardt et al., 2017).
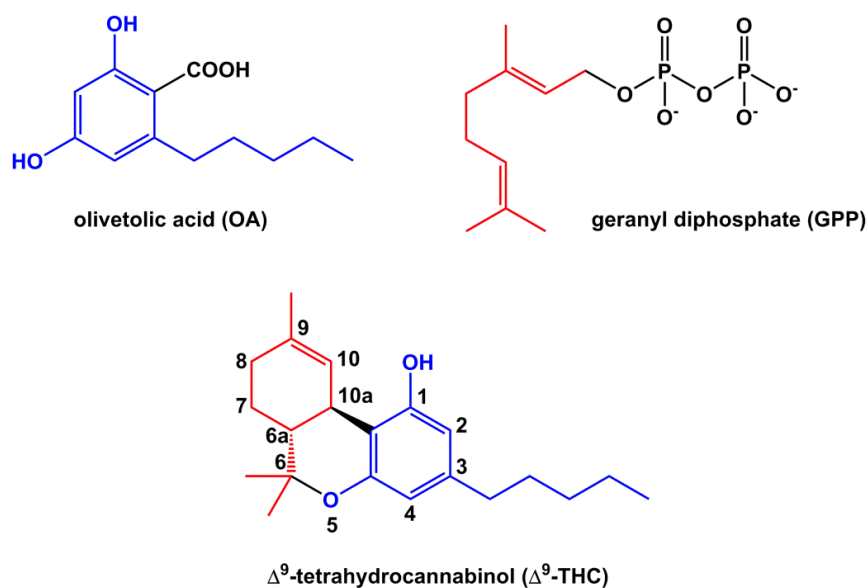
Figure 3: Olivetolic acid, geranyl diphosphate and the general structure of cannabinoids. Cannabinoids consist of two parts: (i) a diphenol part, carrying an alkyl chain (blue), and (ii) a cyclic monoterpene part (red). The dibenzopyran-numbering system is used (Degenhardt et al., 2017).

The phytocannabinoids of the THC-type in *C. sativa* are composed of two parts: (i) an alkylresorcinol moiety derived from olivetolic acid (OA) and (ii) a monoterpene moiety derived from geranyl diphosphate (GPP) (Figure 3).

1.2.1. Polyketide pathway towards olivetolic acid

"The origin of hexanoate in trichomes has not been elucidated so far. Suzuki et al. (2003) showed that the side-chain moiety of alkylresorcinols is formed by fatty acid units, but it remains unclear if the moiety is the result of biosynthesis or degradation of fatty acids. Studies regarding the incorporation of ^{13}C-labels into cannabinoids indicate that hexanoate is synthesized from acetyl-CoA as a starter unit, and five molecules of malonyl-CoA. These building blocks are precursors of the fatty acid biosynthesis (Fellermeier and Zenk, 1998).

Based on this, two pathways are feasibly possible, after analysis of a cDNA / EST library generated from female flowers (glands) of *C. sativa*. First, the hexanoyl residue could be obtained by an early termination of the fatty acid biosynthesis. Subsequently, the hexanoyl moiety of the resulting hexanoyl-ACP would be cleaved by a thioesterase or transferred to CoA by an ACP-CoA transacylase. Finally, acyl-CoA synthetase would catalyze the conversion of the obtained n-hexanol to hexanoyl-CoA (Marks et al., 2009). Second, hexanol could be produced from the breakdown of C18 unsaturated fatty acids via the lipoxygenase pathway (Marks et al., 2009; Stout et al., 2012). Nevertheless, further studies are necessary to clarify the origin of the hexanol moiety.

Hexanoyl-CoA is a medium-chain fatty acyl-CoA that can be detected in high amounts in *Cannabis* flowers (Stout et al., 2012). It is synthesized by an acyl-activating enzyme (AAE) called hexanoyl-CoA synthetase (Marks et al., 2009; Page and Stout, 2013). AAEs can use short-, medium-, long- as well

as very long-chain fatty acids as carboxylic acid substrates. Two novel enzymes were identified, *C. sativa* hexanoyl-CoA synthetase 1 (CsHCS1; CsAAE1) and *C. sativa* hexanoyl-CoA synthetase 2 (CsHCS2; CsAAE3) that are capable of producing hexanoyl-CoA using hexanoate and CoA as substrates (Page and Stout, 2013; Stout et al., 2012)." (Degenhardt et al., 2017). Based on transcript levels CsHCS1 is the most abundant AAE in trichomes. Additionally, the expression of the enzyme is trichome-specific. In comparison to CsHCS1 (GenBank[TM] Accession No. AFD33345), CsHCS2 (GenBank[TM] Accession No. AFD33347) exhibits lower transcript levels and is present in all tissues (Page and Stout, 2013; Stout et al., 2012). Published data indicate "that **CsHCS1** is the enzyme involved in the biosynthesis of cannabinoids: (1) it is the most abundant AAE in trichomes, (2) it is highly specific for short-chain fatty acyl-CoA, particularly hexanoate (K_M value in the nM range), and (3) it is localized in the cytosol as also suggested for the olivetol synthase (see below). In contrast, CsHCS2 is localized in the peroxisomes and accepts a broad range of substrates, while showing a K_M value for hexanoate in the mM range (Page and Stout, 2013; Stout et al., 2012)." (Degenhardt et al., 2017).

"The alkylresorcinol moiety of cannabinoids is derived from OA, the product of polyketide synthases (PKSs) that catalyze the aldol condensation of hexanoyl-CoA with three molecules of malonyl-CoA (Fellermeier and Zenk, 1998; Raharjo et al., 2004) (Figure 5). The second precursor malonyl-CoA is predominantly derived from acetyl-CoA by carboxylation. The ATP-dependent reaction is catalyzed by an acetyl-CoA carboxylase (EC 6.4.1.2). The enzyme utilizes the first step in the fatty acid biosynthesis (Chen et al., 2011; Konishi et al., 1996)." (Degenhardt et al., 2017). Taura et al. (2009) published a plant type III PKS, named **olivetol synthase** (OLS; GenBank[TM] Accession No. AB164375) that is found in flowers and rapidly expanding leaves of *C. sativa*. Results obtained by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and size-exclusion chromatography indicate OLS to be a homodimeric protein. It is noteworthy, that OLS produces olivetol, triketide pyrone and tetraketide pyrone, but not OA (Figure 4) (Taura et al., 2009).
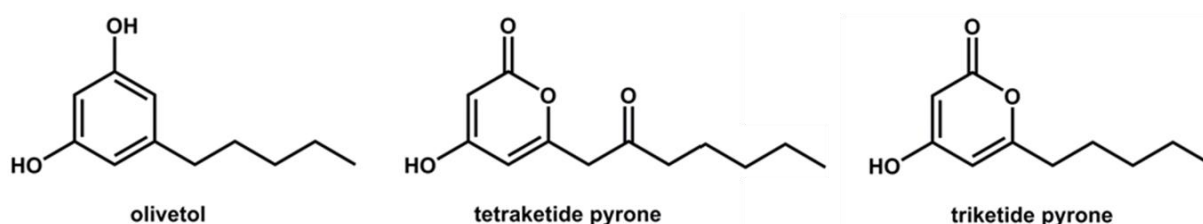
Figure 4: Chemical structures of olivetol, tetraketide pyrone and triketide.Olivetol synthase converts three molecules of malonyl-CoA and one molecule of n-hexanoyl-CoA to olivetol and the side-products tetraketide pyrone and triketide pyrone (Taura et al., 2009).

Gagne et al. (2012) isolated a polyketide cyclase gene from the glandular trichomes of femal *Cannabis* flowers which they named **olivetolic acid cyclase** (OAC). "Finally, using both OLS and OAC with hexanoyl-CoA and malonyl-CoA in one assay, the formation of OA, pentyldiacetic acid (triketide pyrone) and hexanoyltriacetic acid lactone (HTLA; tetraketide pyrone) could be demonstrated

(Page and Gagne, 2013) (Figure 5). It is assumed that OLS catalyzes the formation of an intermediate that is subsequently converted into OA by OAC (Gagne et al., 2012; Taguchi et al., 2008)." (Degenhardt et al., 2017). Furthermore, Taguchi et al. (2008) postulate HTLA, the structural isomer of OA, to be the precursor of OA. It is noteworthy, that until now no free OA has been detected in *Cannabis* plant material. In 2016, Yang et al. solved the crystal structure of OAC at a resolution of 1.32 Å (PDB 5B08 and 5B09).
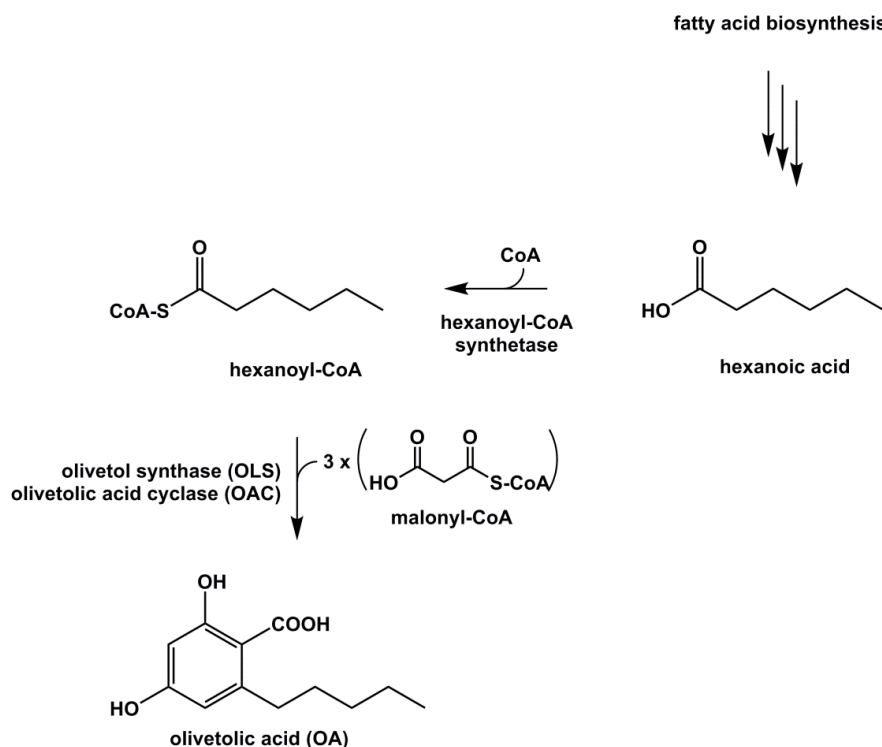
Figure 5: Biosynthesis of olivetolic acid (OA). Modified from Degenhardt et al. (2017). OA is biosynthesized by the polyketide pathway that starts with hexanoic acid. Hexanoyl-CoA synthetase uses hexanoate and CoA to form hexanoyl-CoA. Olivetol synthase and olivetolic acid cyclase catalyze the formation of OA from three molecules of malonyl-CoA and one molecule of hexanoyl-CoA (Gagne et al., 2012; Taura et al., 2009).

1.2.2. Biosynthesis of geranyl diphosphate

"The monoterpene moiety of cannabinoids (Figure 6) is derived from GPP. Its precursors, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), are predominantly (> 98%) biosynthesized via the 2C-methyl-D-erythritol-4-phosphate (MEP) pathway (also termed nonmevalonate pathway or 1-deoxy-D-xylulose-5-phosphate (DOXP) pathway) (Fellermeier et al., 2001)." (Degenhardt et al., 2017). The MEP pathway is built up of seven steps and starts with the linkage of pyruvate and glyceraldehyde-3-phosphate (Tholl, 2015). The results of Fellermeier et al. (2001) "are supported by Marks et al. (2009). They isolated RNA from the glands of a tetrahydrocannabinolic acid (THCA)-producing *Cannabis* strain and generated a cDNA library. After sequencing they were able to identify all but one enzyme involved in the MEP pathway. Additionally, Stout et al. (2012) found high expression of MEP pathway genes in *Cannabis* flowers. Furthermore, in higher plants the MEP

pathway, mainly involved in secondary metabolism, is localized in plastids (described in detail elsewhere, for example, Eisenreich et al. (2004) or Hunter (2007), whereas the mevalonate (MVA) pathway, predominantly contributing to primary metabolism, is localized in the cytosol (Eisenreich et al., 2004; Hunter, 2007)." (Degenhardt et al., 2017). In plants the compartmental separation of the MEP and MVA pathways is not absolute. Isoprenoid intermediates like IPP that is common to both pathways, can be exchanged bidirectional over the plastid membranes. Additionally, there is evidence that prenyl diphosphates like farneyl diphosphate (FPP) or GPP are transported from the plastids to the cytosol (Bick and Lange, 2003; Eisenreich et al., 2004; Tholl, 2015; Wölwer-Rieck et al., 2014).

"The head-to-tail condensation of IPP and DMAPP to form GPP is catalyzed by **geranyl diphosphate synthase** (Figure 6) (Burke et al., 1999)." (Degenhardt et al., 2017). The GPP synthase of *C. sativa* is still unknown.
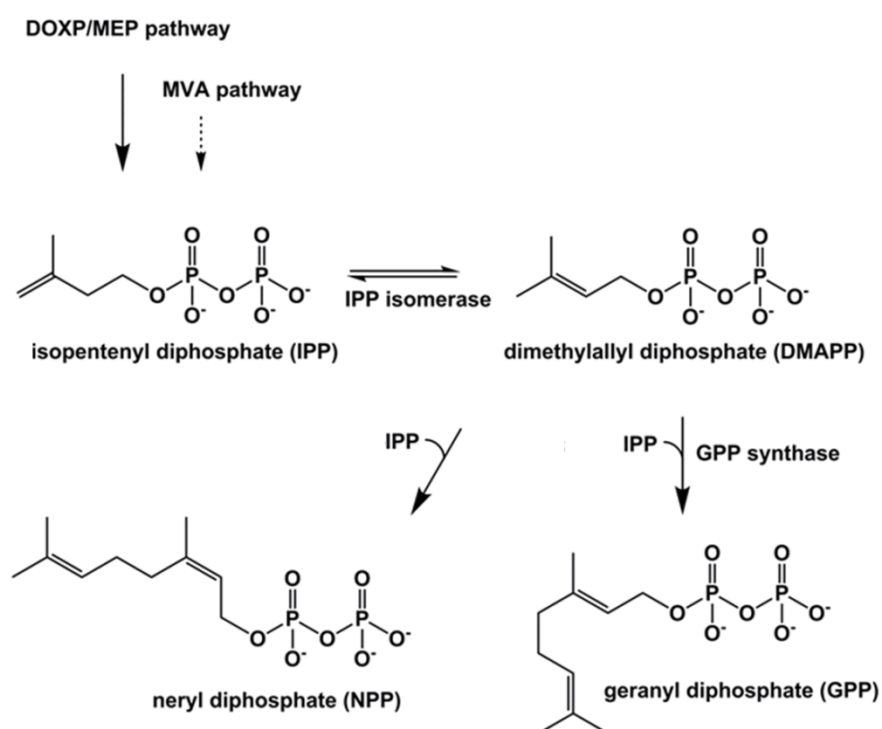
Figure 6: The biosynthesis of geranyl diphosphate (GPP). Modified from Degenhardt et al. (2017). The isoprenoid precursors IPP and DMAPP are mainly biosynthesized in the MEP pathway. In plants the compartmental separation of MEP and MVA pathway is not absolute. The head-to-tail condensation of IPP and DMAPP to form GPP is catalyzed by geranyl diphosphate synthase. DOXP, 1-deoxy-D-xylulose-5-phosphate; MEP, 2C-methyl-D-erythritol-4-phosphate; MVA, mevalonate. (Burke et al., 1999; Fellermeier and Zenk, 1998).

1.2.3. Biosynthesis of cannabinoids from GPP and OA

"**Cannabigerolic acid synthase** (CBGAS) or geranylpyrophosphate:olivetolate geranyltransferase (GOT) predominantly catalyzes the C-prenylation of OA by GPP to form cannabigerolic acid (CBGA) (Figure 5)" (Degenhardt et al., 2017), the first step in cannabinoid production in *C. sativa*. Since different cyclization of the prenyl moiety leads to THCA or its isomers cannabichromenic acid (CBCA) and cannabidiolic acid (CBDA), CBGA is supposed to be the central precursor for cannabinoid

biosynthesis (Fellermeier and Zenk, 1998; Page and Boubakir, 2014; Sirikantaramas et al., 2007). CBGAS is described in more detail in chapter 1.3.1.1.

Cannabidiolic acid synthase (CBDAS), cannabichromenic acid synthase (CBCAS) and tetrahydrocannabinolic acid synthase (THCAS) predominantly catalyze the stereoselective, oxidative cyclization of the monoterpene moiety of CBGA to form CBDA, CBCA or THCA, respectively (Figure 7). Furthermore, they also convert cannabinerolic acid (CBNRA), the *cis* isomer of CBGA, but with a lower specificity (Morimoto et al., 1998; Shoyama et al., 2012; Sirikantaramas et al., 2007; Taura et al., 1996, 1995). The CBDAS and THCAS reactions are oxygen-dependent, producing hydrogen peroxide in equimolar amounts of products, i.e. CBDA or THCA (Sirikantaramas et al., 2004; Taura et al., 2007c). "Remarkably, the CBCAS reaction is oxygen independent, and can be inhibited by hydrogen peroxide. Thus, the enzyme seems not to be an oxygenase or a peroxidase (Morimoto et al., 1998)." (Degenhardt et al., 2017). This assumption should be considered with caution since the same conclusion were made for CBDAS (Taura et al., 1996) and THCAS (Taura et al., 1995) before they were characterized in more detail (Sirikantaramas et al., 2004; Taura et al., 2007c). According to Sirikantaramas et al. (2007) THCA is biosynthesized and stored in the storage cavities of the glandular trichomes of *Cannabis* plants.

"Since no enzymatic activity was detectable using the neutral cannabinoid cannabigerol (CBG), the carboxyl group in CBGA seems to be essential for the enzymatic reactions catalyzed by THCAS, CBDAS and CBCAS (Morimoto et al., 1997; Taura et al., 1996, 1995)." (Degenhardt et al., 2017). Additionally, CBGAS (chapter 1.3.1.1) is not able to catalyze the prenylation of olivetol by GPP (Fellermeier and Zenk, 1998).
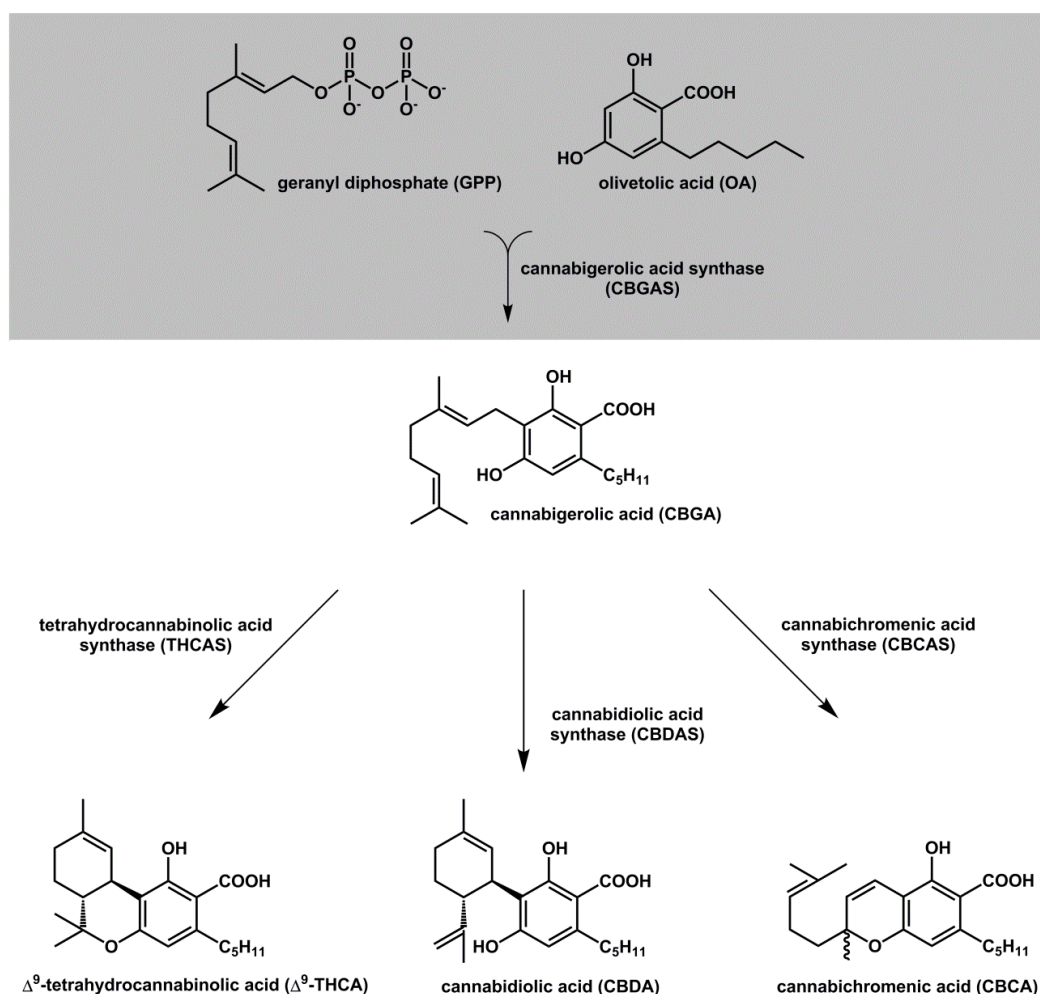
Figure 7: Biosynthesis of cannabidiolic acid (CBDA), cannabichromenic acid (CBCA) and tetrahydrocannabinolic acid (THCA). Modified from Degenhardt et al. (2017). Cannabigerolic acid synthase catalyzes the prenylation of OA with GPP to form CBGA, the central precursor for cannabinoid biosynthesis. CBGA is converted by cannabidiolic acid synthase (CBDAS), cannabichromenic acid synthase (CBCAS) and tetrahydrocannabinolic acid synthase (THCAS) to form CBDA, CBCA or THCA, respectively (Morimoto et al 1998; Shoyama et al 2012; Sirikantaramas 2007; Taura et al 1995; Taura et al 1996).

The **thcas** gene encodes a 545-amino acid polypeptide chain (GenBank[TM] accession number AB057805), containing a 28 amino acid long N-terminal signal peptide which is cleaved in the mature THCAS (Sirikantaramas et al., 2004). The enzyme extracted from a Mexican *Cannabis* strain "has a theoretical molecular mass of 59 kDa. An actual mass of about 75 kDa was detected using SDS-PAGE (Taura et al., 1995)." (Degenhardt et al., 2017). This could be explained by post-translational modifications, since X-ray crystallography of THCAS (PDB ID: 3VTE; resolution: 2.75 Å) confirmed six Asn glycosylation sites (Shoyama et al., 2012). Furthermore, THCAS deglycosylated by EndoH treatment after functional expression indeed showed a molecular mass of 59 kDa and remained fully active (Taura et al., 2007c). "THCAS is a monomeric enzyme with the highest enzyme activity between pH 5.5 and pH 6.0 (Taura et al., 1995)." (Degenhardt et al., 2017). Zirpel et al. (2018a) published a THCAS pH optimum of 4.5. "Sequence comparison identified similarities to the berberine bridge enzyme (BBE) of *Eschscholzia californica* (Shoyama et al., 2012). BBE belongs to the family of

oxidoreductases and has a bi-covalently bound flavine adenine dinucleotide (FAD) (Kutchan and Dittrich, 1995). The THCAS amino acid sequence revealed a flavinylation consensus sequence (Arg[110]-Ser-Gly-Gly-His[114]) in which His[114] is probably the FAD-binding site (Sirikantaramas et al., 2004). This could be confirmed by X-ray crystallography (PDB ID: 3VTE) at a resolution of 2.75 Å. Besides His[114], a second residue, Cys[176], could be identified to be covalently bound to the FAD (Shoyama et al., 2012)." (Degenhardt et al., 2017). A possible catalytic reaction meachanism of THCAS, assigning Tyr[484] a central role, was proposed by Sirikantaramas and Taura (2017). "Nevertheless, since the crystal structure was published without substrate analoga, further studies are necessary to verify the suggested mechanism." (Degenhardt et al., 2017). Interestingly, Zirpel et al. (2018b) showed that THCAS converts CBGA in a pH-dependent manner amongst others to THCA, CBDA and CBCA.

The *cbdas* wildtype gene "encodes a 544-amino acid polypeptide (GenBank[TM] Accession No. AB292682). According to Taura et al. (2007c), processed CBDAS consists of 517 amino acids following cleavage of the 28 amino acid long signal peptide. The mature CBDAS has a theoretical molecular mass of 59 kDa. An actual mass of about 74 kDa was detected by SDS-PAGE that is possibly caused by posttranslational glycosylation of seven Asn residues (Taura et al., 2007a, 1996). CBDAS is a monomeric enzyme with a pH optimum of 5.0 (Taura et al., 1996)." (Degenhardt et al., 2017), which is slightly higher than a pH of 4.5 as reported by Zirpel et al. (2018b). "A comparison between THCAS and CBDAS revealed a sequence similarity of 84% (Taura et al., 2007c). Like THCAS, CBDAS is a flavinated enzyme in which His[114] and Cys[176] are most likely the FAD-binding sites. Since CBDAS exhibits structural and biochemical properties related to those of THCAS, it is probable that the reaction mechanism of CBDAS is similar to that of THCAS (Taura et al., 2007c)." (Degenhardt et al., 2017). Interestingly, Zirpel et al. (2018b) showed that CBDAS, like THCAS, converts CBGA in a pH-dependent manner into THCA, CBDA and CBCA.

The **CBCAS** protein was purified to apparent homogeneity from young leaves of a CBDA strain (Morimoto et al., 1998). The CBCA sequence is not available in public databases, but in 2015 Page and Stout filed a patent containing a coding sequence. THCAS and CBCAS share a nucleotide similarity of 96%. Results of Morimoto et al. (1998) indicate CBCAS to be a homodimeric protein with a molecular mass of 136 kDa, consisting of two identical 71 kDa subunits. Additionally, a maximum activity was determined at pH 6.5. "According to kinetic data, CBCAS has a higher affinity for CBGA than THCAS and CBDAS. CBCA and its neutral form CBC are both racemic. Studies of Morimoto et al. (1997) suggested that both enantiomers of CBCA are formed by a CBCAS catalyzed reaction with a molar ratio of 5:1. But it is still unknown which of the two isomers is the major product (Gaoni and Mechoulam, 1971; Morimoto et al., 1997; Taura et al., 2007b)." (Degenhardt et al., 2017).

Enzymes involved in the biosynthesis of phytocannabinoids in *C. sativa* as well as their GenBank[TM] accession numbers and publications in which they were described for the first time are summarized in Table 1. Non-enzymatically and enzymatically derived products from CBGA in *C. sativa* are summarized in Figure 8.

Table 1: Enyzmes involved in phytocannabinoid biosynthesis in *C. sativa.* The table lists the enzymes, the used abbreviations, the corresponding accession numbers, the PDB numbers of available crystal structures and the publications in which the enzymes were described for the first time. Modified from Degenhardt et al. (2017). [1] GenBank[TM], [2] PDB, [3] patent number

Enzyme		**Accession no.**	**Structural data** [2]	**EC no.**	**Reference**
Acyl-activating enzyme 1 (hexanoyl-CoA synthethase)	AAE1 (CsAAE1)	AFD33345.1 [1]	-	6.2.1.-	(Stout et al., 2012)
Olivetol synthase	OLS	AB164375 [1]	-	2.3.1.206	(Taura et al., 2009)
Olivetolic acid cyclase	OAC	AFN42527.1 [1]	5B08 (OAC apo) 5B09 (OAC-OA binary complex)	4.4.1.26	(Gagne et al., 2012; Yang et al., 2016)
Cannabigerolic acid synthase	CBGAS	US 8884100 B2 [3] WO 2011017798 A1 [3]	-	2.5.1.102	(Fellermeier and Zenk, 1998; Page and Boubakir, 2014)
Cannabichromenic acid synthase	CBCAS	US 20170211049 A1 [3] WO 2015196275 A1 [3]	-	1.3.3.-	(Morimoto et al., 1998; Page and Stout, 2015)
Cannabidiolic acid synthase	CBDAS	AB292682 [1]	-	1.21.3.8	(Taura et al., 1996)
Tetrahydrocannabinolic acid synthase	THCAS	AB057805 [1]	3VTE	1.21.3.7	(Sirikantaramas et al., 2004)
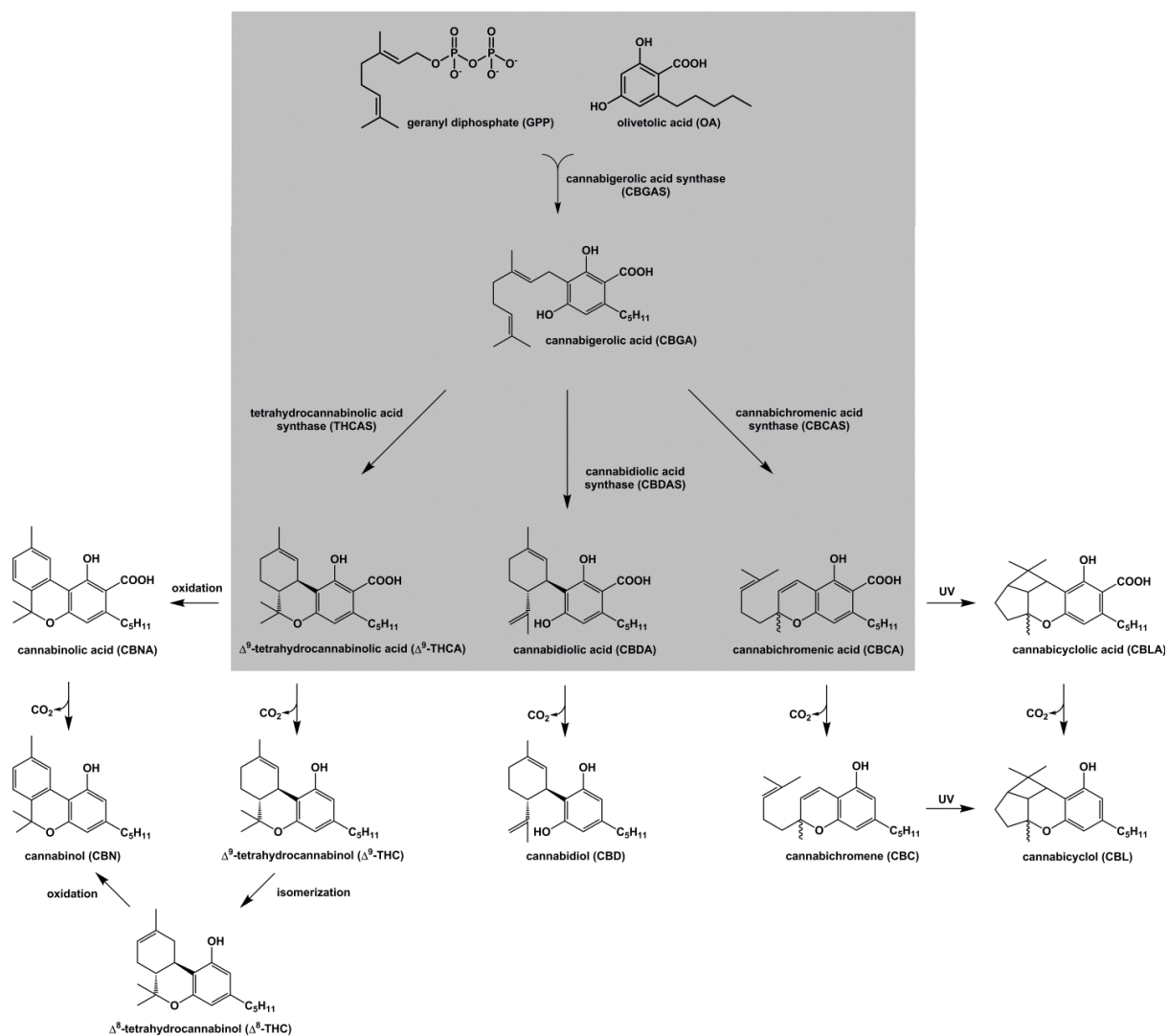
Figure 8: Non-enzymatic and enzymatic catalyzed conversion of CBGA. Modified from Degenhardt et al. (2017). CBGAS catalyzes the C-prenylation of OA with GPP to form CBGA that is converted to THCA, CBDA and CBCA by THCAS, CBDAS or CBCAS, respectively. The enzymatically catalyzed reactions are highlighted in grey (Crombie et al., 1968; Fellermeier and Zenk, 1998; Morimoto et al., 1997; Shoyama et al., 1968; Sirikantaramas et al., 2005; Taura et al., 2007c).

1.3.   Aromatic prenyltransferases

The term prenylation is defined as the addition of a hydrophobic side chain (prenyl group) to a protein or a chemical compound. The size of the prenyl group ranges from relatively small C5 units, including dimethylallyl or isopentenyl groups, to larger C10 (geranyl moiety) or C15 units (farnesyl group) (Heide, 2009; Liang et al., 2002). Each prenylation reaction starts with the cleavage of the isoprenyl diphosphate C-OP bond resulting in a diphosphate anion and a prenyl cation. The cation can, among others, react by proton loss, resulting in an ocimene, or by addition to nucleophiles or carbon nucleophiles (Brandt et al., 2009). Ocimenes are monoterpenes which are for example detectable in the resin of *C. sativa* variety Finola (Booth et al., 2017). Prenylation often leads to the formation of compounds possessing biological activities that their non-prenylated precursors do not exhibit. Due to this, prenylation contributes strongly to the diversification of aromatic natural products such as coumarins, polyketides, and flavonoids (Fellermeier and Zenk, 1998; Steffan et al., 2009; Yazaki et al., 2009). The prenyl residue is provided by the MVA / isoprenoid or the MEP pathway, while the aromatic moiety is derived from the polyketide or shikimate pathway (Yazaki et al., 2009). The diversification of aromatic natural products is amongst others caused by the different chain lengths of the prenyl residue, prenylation of the aromatic substrate at various positions and cyclization of the prenyl moiety. Prenylated products can possess medical desirable biological activities like anti-bacterial, anti-inflammatory, anti-tumor and anti-viral properties (Sasaki et al., 2008; Yang et al., 2012; Yazaki et al., 2009).

Prenyl-converting enzymes are classified into isomerases/hydrolases, terpene synthases (cyclases), transferases, and enzymes which do not belong to the first three groups like epoxidases and hydrogenases. An example for a **prenyl isomerase** is isopentenyl diphosphate isomerase which catalyzes the isomerization of DMAPP from IPP and vice versa (Sharma et al., 2010). Prenyl diphosphatases (EC 3.1.7.1) belong to the family of **hydrolases** and catalyze the hydrolysis of prenyl diphosphate to diphosphate and prenol (H-$[CH_2CCH_3=CHCH_2]_n$-OH) (Schomburg and Salzmann, 1991). **Terpene synthases** like monoterpene (e.g. limonene synthase) or sesquiterpene synthases (e.g. γ-humulene synthase) are responsible for terpene diversification in plants (Brandt et al., 2009; Degenhardt et al., 2009).

The group of **transferases** can be subdivided into (i) isoprenyldiphosphate synthases (IPPS), (ii) protein prenyltransferases, and (iii) aromatic prenyltransferases (Brandt et al., 2009; Bräuer et al., 2008; Liang et al., 2002; Tello et al., 2008).

Isoprenyldiphosphate synthases (IPPS) catalyze the fusion of IPP with allylic diphosphate substrates by head-to-head or head-to-tail condensation, resulting in chain elongation of the latter one. Based on the stereochemical outcome of the products IPPSs can be subdivided into *cis*-isoprenyldiphosphate synthase (e.g. hexaprenyldiphosphate synthase in yeast which is involved in ubiquinone biosynthesis) and *trans*-isoprenyldiphosphate synthase (e.g. geranyldiphosphate synthase (GPPS)). It is noteworthy that even though the enzyme-catalyzed reactions are magnesium dependent, only *trans*-IPPSs contain an aspartate-rich motif (DDxxD) (Brandt et al., 2009; Liang et al., 2002; Marakasova et al., 2013).

Protein prenyltransferases catalyze posttranslational modifications of proteins, which are important for protein-protein interactions and membrane attachment, and can be classified into two groups. Protein farnesyltransferase (FTase) and protein geranylgeranyltransferase I (GGTase-I) belong to the CaaX prenyltransferases, whereas protein geranylgeranyltransferase II (GGTase-II), also known as Rab geranylgeranyltransferase, is called a non-CaaX prenyltransferase (C - cysteine, a - aliphatic amino acid, x - any amino acid). The enzymes of the former group transfer a farnesyl group (C15) or a geranylgeranyl group (C20) to the cysteine residue of proteins containing a CaaX motif at their C-terminus. In the process, the X of the CaaX box influences whether a C15 or a C20 group is added to the protein. Non-CaaX prenyltransferases recognize a different prenylation motif (CC or CxC) and transfer C20 groups to both cysteines of the motif containing protein (Benetka et al., 2006; Casey and Seabra, 1996; Marakasova et al., 2013; Zhang and Casey, 1996).

Aromatic prenyltransferases catalyze the prenylation of an electron-rich aromatic acceptor by a prenyl diphosphate as they transfer a prenyl residue to C, N, or O atoms of aromatic acceptor molecules. The prenyl group can be a dimethylallyl (C5), a geranyl (C10) or a farnesyl (C15) residue (Brandt et al., 2009; Tello et al., 2008). The bioactivity of the resulting products is often increased, compared with the non-prenylated substrates, probably due to higher lipophilicity which results in increased affinity for membranes (Botta et al., 2005). Aromatic prenyltransferases can be found in bacteria, fungi and plants and cause a huge diversity of secondary metabolites in these organisms (Brandt et al., 2009; Saleh et al., 2009). In many cases aromatic prenyltransferases catalyze the rate limiting step in the respective biosynthetic pathway (Heide, 2009). Based on their biochemical properties and sequences they can be classified into two groups: (i) membrane-bound aromatic prenyltransferases, which contain characteristic aspartate-rich motifs and require divalent metal ions for their catalytic activity and (ii) soluble aromatic prenyltransferases, which do not require divalent cations for enzymatic activity and do not possess aspartate-rich motifs in their sequence (Winkelblech et al., 2015).

As this thesis focuses on aromatic prenyltransferases the groups of **membrane-bound aromatic prenyltransferases** as well as of **soluble aromatic prenyltransferases** are described in more detail in the following chapters (chapter 1.3.1 and chapter 1.3.2).

1.3.1.   Membrane-bound aromatic prenyltransferases

Membrane-bound aromatic prenyltransferases are involved in the biosynthesis of primary (e.g. ubiquinone) and secondary metabolites (e.g. CBGA, CBDA, THCA). Enzymes of this group require divalent cations like Mg^{2+} for their catalytic activity and their amino acid sequences possess conserved amino acid sequences responsible for prenyl diphosphate recognition: DDxxDxxxD, DQxxDxxxD, DQxxExxxD, NDxxDxxxD, NQxxDxxxD or NQxxExxxD (Heide, 2009; Marks et al., 2009; Sasaki et al., 2008; Yazaki et al., 2009). An example for a membrane-bound aromatic prenyltransferase are *p*-hydroxybenzoate:polyprenyl transferases (PPTs) (Ohara et al., 2004). This enzyme is required in the

biosynthetic pathway of ubiquinone (UQ) which is present in all membranes and cells. UQ has several functions, e.g. as an electron carrier in the mitochondrial respiratory chain, the oxidation of sulfide (yeast), the introduction of disulfide bonds (bacteria) and the regulation of the physicochemical properties of membranes. PPT catalyzes the C-prenylation of *p*-hydroxybenzoic acid (PHB) with an all-*trans* polyprenyl group (Figure 9) (Forsgren et al., 2004; Melzer and Heide, 1994; Ohara et al., 2004; Turunen et al., 2004).
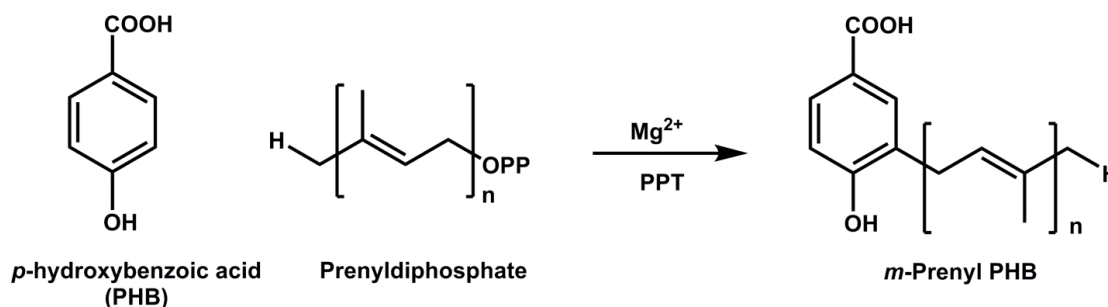
Figure 9: *p*-hydroxybenzoate:polyprenyl transferase (PPT). Modified from Ohara et al. (2004). PPT catalyzes the C-prenylation of *p*-hydroxybenzoate by an all-*trans* polyprenyl group.

Membrane-bound aromatic prenyltransferases of plants can be divided into two groups: (i) PHB prenyltransferases which are localized in the membrane of mitochondria, and (ii) homogentisate prenyltransferases that are located in plastids (Cheng and Li, 2014; Ohara et al., 2006; Yazaki et al., 2009).

In 2002, Yazaki et al. reported the expression of a plant **PHB prenyltransferase**, geranyldiphosphate:4-hydroxybenzoate 3-geranyltransferase (LePGT1) for the first time. Here it was isolated from the medicinal plant *Lithospermum erythrorhizon*. LePGT1 is involved in shikonin/naphthoquinone biosynthesis and possesses strict substrate specificity for GPP. Due to its localization within the endoplasmatic reticulum, LePGT1 presents an exception in the group of PHB prenyltransferases (Yazaki et al., 2009, 2002). In 2013, Ohara et al. reported the purification to homogeneity of a membrane-bound aromatic prenyltransferase (LePGT1) for the first time. They were able to produce the enzyme in its native form in a non-plant host organism via gene expression without its N-terminal mitochondrial targeting sequence (Ohara et al., 2013).

An example for a plant membrane-bound aromatic prenyltransferase that belongs to the group of homogentisate prenyltransferases is presented in chapter 1.3.1.1.

1.3.1.1. CBGA synthase

The geranylpyrophosphate:olivetolate geranyltransferase (GOT) or CBGA synthase (CBGAS) predominantly catalyzes the C-alkylation of OA with GPP to form CBGA, the first step in cannabinoid production in *C. sativa* (chapter 1.2) (Fellermeier and Zenk, 1998; Page and Boubakir, 2014; Sirikantaramas et al., 2007). Fellermeier and Zenk detected the enzyme in extracts from young rapidly

expanding leaves of *C. sativa*, which contain the later enzymes of the THCA biosynthetic pathway (Fellermeier and Zenk, 1998; Morimoto et al., 1997; Taura, 1995). Although there is evidence that CBGAS, like other aromatic prenyltransferases, is a membrane-bound enzyme (Ohara et al., 2009; Yamamoto et al., 1997), Fellermeier and Zenk (1998) detected CBGAS activity in the soluble fraction of the crude extract, but not in the isolated particulate fractions. They detected two products, identified by mass spectrometry (MS) measurements as CBGA and its cis-isomer cannabinerolic acid (CBNRA; Fellermeier and Zenk (1998) used CBNA instead). CBNRA formation may result from the C-prenylation of olivetolic acid by neryl diphosphate (NPP) that is obtained by the availability of a geranyldiphosphate isomerase in the crude enzyme extract. In the presence of NPP the enzyme-catalyzed reaction yields a CBGA/CBNRA ratio of 1:1, while a ratio of 2:1 is obtained in the presence of GPP. Additionally, Fellermeier and Zenk (1998) showed that CBGAS catalyzes a Mg^{2+}-dependent reaction and does not accept olivetol as an aromatic substrate. Although enzyme activity was detected in the soluble fraction, a membrane-bound activity could not be excluded completely.

In 1998, Zuurbier et al. published the presence of two soluble plant aromatic prenyltransferases involved in the biosynthesis of bitter acids in *Humulus lupulus* (hop). It is noteworthy, that these are the only two soluble plant aromatic prenyltransferases described so far and it has not been possible to isolate the corresponding genes/enzymes or obtain sequence information to this date. All known sequences of plant aromatic prenyltransferases encode membrane-bound proteins (Sasaki et al., 2008; Stec and Li, 2012; Tello et al., 2008; Winkelblech et al., 2015; Yamamoto et al., 2000, 1997; Zhao et al., 2003).

Interestingly, Li et al. (2015) showed that two membrane-bound aromatic prenyltransferases catalyze the prenlyation steps in the bitter acid pathway. Additionally, sequence informations of the two enzymes were provided (GenBank[TM] accession numbers: KM222441 and KM222442). In summary, there are inconsistent reports as to whether the enzymatic steps of the bitter acid pathway of *H. lupulus* are catalyzed by soluble (Zuurbier et al., 1998) or membrane-bound prenyltransferases (Li et al., 2015). Similar scientific disagreements exist on the features of the CBGAS. Beside the publication of Fellermeier et al. (1998) which suggest CBGAS to be a soluble aromatic prenyltransferase, an issued patent of Page and Boubakir (2014) indicates CBGAS to be a membrane-bound aromatic prenyltransferase (CsPT1).

The gene sequence of CsPT1 shows similarity to homogentisate phytyltransferase VTE2-2 (Page and Boubakir, 2014). VTE2-2 is a membrane-bound prenyltransferase, which is involved in tocopherol biosynthesis and catalyzes the condensation of homogentisic acid with phytyl diphosphate. Tocopherols (vitamin E) are exclusively synthesized by photosynthetic organisms and function as antioxidants e.g. in biological membranes (Collakova and DellaPenna, 2001; Venkatesh et al., 2006). The CsPT1 coding sequence consists of 1,188 bp (GenBank[TM] accession number: AJN57774.1) and is mainly expressed in young leaves, flowers and glandular trichomes isolated from flowers of *Cannabis* plants. Page and Boubakir (2014) claim to have achieved functional expression of the recombinant prenyltransferase in *S. cerevisiae* cells as well as in *Spodoptera frugiperda Sf9* insect

cells and detected enzyme activity in the microsomal fraction by MS measurements. CsPT1 assayed with OA and GPP resulted in the formation of CBGA (major product; 3-geranyl olivetolic acid) and 5-geranyl olivetolic acid (minor product). Additionally, the authors claim that the enzyme only accepts GPP as prenyl donor, but uses different aromatic substrates as aromatic acceptor (OA, olivetol, resveratrol, naringenin, and phlorisovalerophenone; Figure 10). The CsPT1 catalyzed reaction is dependent on divalent cations, whereas the highest CBGA amount was obtained in the presence of Mg^{2+} ions (Page and Boubakir, 2014).
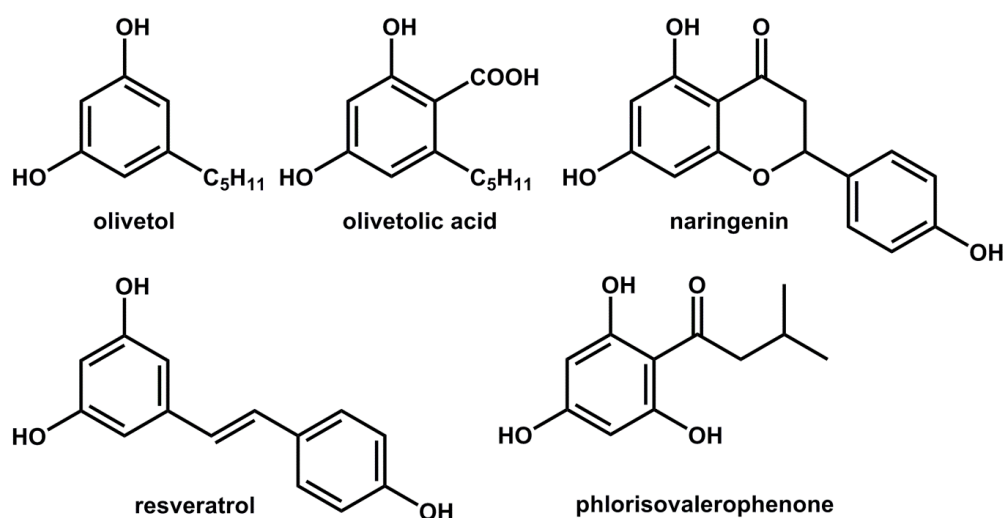
Figure 10: Aromatic acceptor substrates of CsPT1. According to Page and Boubakir (2014) CsPT1 only accepts GPP as prenyl donor, but various compounds as aromatic acceptor molecules: Naringenin, olivetol, olivetolic acid, phlorisovalerophenone and resveratrol.

As the patent of Page and Boubakir (2014) and the publication of Fellermeier and Zenk (1998) are the only two scientific reports dealing with CBGAS, it remains unclear whether the enzyme is a soluble or a membrane-bound aromatic prenyltransferase.

1.3.2. Soluble, aromatic prenyltransferases

Soluble aromatic prenyltransferases – in contrast to membrane-bound aromatic prenyltransferases – do not contain aspartate-rich motifs responsible for prenyl diphosphate binding ((N/D)DxxxD). Furthermore, they are active in the absence of divalent metal ions, with NphB representing an exemption (Heide, 2009; Kuzuyama et al., 2005; Saleh et al., 2009; Steffan et al., 2009). Soluble aromatic prenyltransferaes can be subdivided into two groups: (i) Members of the dimethylallyltryptophan synthase (DMATS) family, and (ii) members of the CloQ/NphB group.

Interestingly, both groups belong to the family of ABBA prenyltransferases that exhibit a α-β-β-α architecture (Bonitz et al., 2011; Tello et al., 2008; Winkelblech et al., 2015). Enzymes of this family contain a common structural feature, a β-barrel fold, which was observed in X-ray structures of NphB for the first time. The so-called PT barrel possesses a (ααββ)$_4$ – (αββ) – α nomenclature and is

composed of ten anti-parallel β-strands forming a cylindrical β-sheet. The central core is solvent-filled and contains the active center of the enzyme. The β-sheets are surrounded by a ring of α-helices which are exposed to the solvent (Figure 11) (Bonitz et al., 2011; Kuzuyama et al., 2005).

The PT barrel is, as known so far, unique to members of the ABBA prenyltransferase family and the large solvent-filled core enables diverse prenylations of various aromatic acceptor molecules (Saleh et al., 2009).

N-terminus   N-terminus   C-terminus
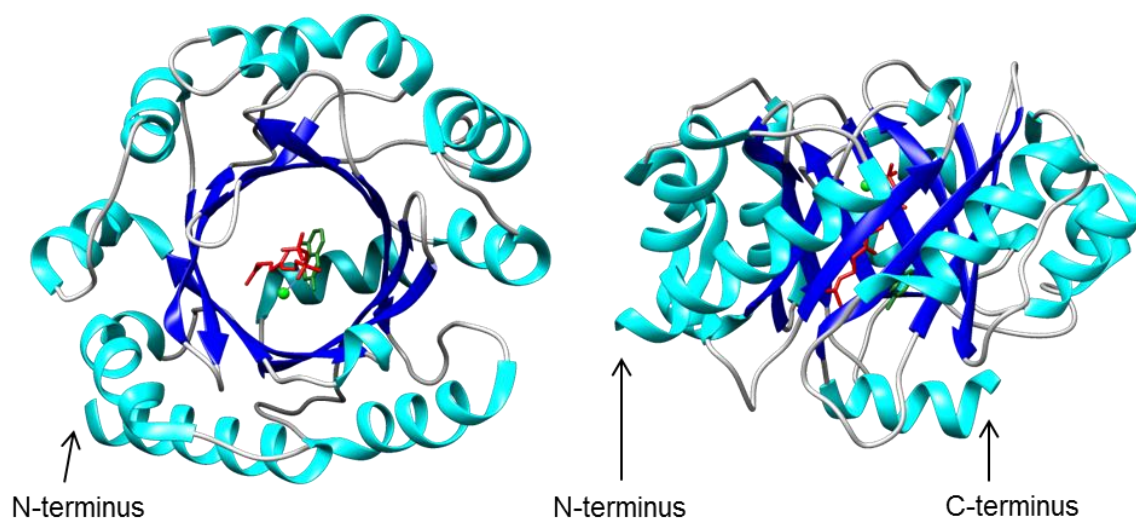
Figure 11: PT barrel of NphB. The PT barrel consists of a solvent-filled core and is built up of ten antiparallel β-sheets. The central core contains the active center of the enzyme. The β-sheets are surrounded by a ring of nine α-helices which are exposed to the solvent. The image shows the crystal structure of NphB complexed with a magnesium ion (light green), 1,6-dihydroxynapthalene (red) and geranyl thiolodiphosphate (dark green) (PDB 1ZB6). The picture on the left shows the enzyme from the top, the right image depicts the lateral view.

Members of the **DMATS** family are mainly found in *Ascomycetes* and catalyze the prenylation of indole derivatives, including tryptophane and tyrosine. Most members of this group use DMAPP as a prenyl donor. Examples for this group are fungal indole prenyltransferases, e.g. dimethylallyl tryptophan synthase, which are specific for DMAPP as a prenyl donor and show substrate flexibility towards their aromatic substrate. The prenyltransferase is involved in the biosynthesis of prenylated indole alkaloids by prenylation of the indole rings, e.g. from L-tryptophan, at different positions (Li, 2010; Stec and Li, 2012; Steffan et al., 2009; Tsai et al., 1995; Winkelblech et al., 2015). The DMATS family is also known as the DMATS/Cymd family since indole prenyltransferases are identified from bacteria as well (Bonitz et al., 2011).

Prenyltransferases of the **NphB/CloQ** group, also known as phenol/phenazine prenyltransferases, occur in bacteria and fungi. The enzyme CloQ was first identified in 2003 from *Streptomyces roseochromogenes*. The prenyltransferase catalyzes the prenylation of 4-hydroxyphenylpyruvate using DMAPP as a prenyl donor and is essential for the biosynthesis of the antibiotic clorobiocin in *S. roseochromogenes* (Pojer et al., 2003). The aromatic prenyltransferase NphB is described in more detail in chapter 1.3.2.1.

1.3.2.1. <u>NphB</u>

NphB (Orf2) is a soluble, aromatic prenyltransferase that was first isolated from *Streptomyces* sp. strain CL190 by Kuzuyama et al. (2005) while studying the gene cluster responsible for naphterpin biosynthesis. The gene (GenBank[TM] accession number: AB187169) encodes a 307-amino acid monomeric protein. *In vivo*, the enzyme is involved in naphterpin biosynthesis and transfers a geranyl group to a 1,3,6,8-tetrahydroxynapthalene-derived polyketide. Nevertheless, the natural, physiological substrate of this prenyltransferase is still unknown (Kumano et al., 2008; Kuzuyama et al., 2005). *In vitro*, the enzyme catalyzes carbon-carbon and carbon-oxygen based prenylation of hydroxyl containing aromatic acceptors of microbial, plant and synthetic origin like 1,3-dihydroxynaphthalene (1,3-DHN), 1,6-DHN, 2,7-DHN, resveratrol, flaviolin, naringenin and olivetol (Kumano et al., 2008; Kuzuyama et al., 2005). Besides 1,6-DHN all aromatic acceptor molecules were geranylated in *ortho* position to a hydroxyl group. Prenylation of 1,6-DHN was detectable in *ortho* and at a very low level in *para* position to a OH-group. NphB accepts GPP and to a lesser degree FPP, but no DMAPP as prenyl donor. Activity assays with 1,3-DHN, 1,6-DHN and 2,7-DHN indicate relaxed substrate specificity for small aromatic molecules (Figure 12) (Kumano et al., 2010; Kuzuyama et al., 2005; Tello et al., 2008).
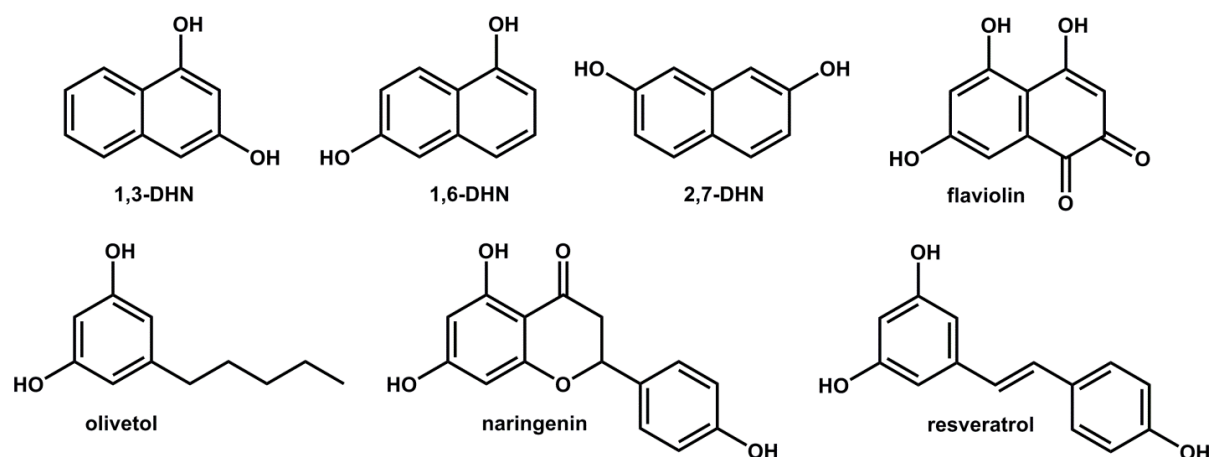
Figure 12: Aromatic acceptor substrates of NphB. NphB exhibits a relaxed substrate specificity for small aromatic molecules by accepting 1,3-dihydroxynapthalene (DHN), 1,6-DHN, 2,7-DHN, flaviolin, naringenin, olivetol and resveratrol as substrates (Kumano et al., 2008; Kuzuyama et al., 2005).

Kuzuyama et al. (2005) crystallized NphB in the presence of different prenyl donor and prenyl acceptor molecules, resulting in four X-ray crystal structures (Table 2).

Co-crystallization of NphB with Mg^{2+}, an aromatic acceptor (1,6-DHN or flaviolin) and geranyl thiolodiphosphate (GSPP) as a prenyl donor indicate a large pocket at the site of aromatic substrate binding, which might explain the relaxed substrate specificity for aromatic acceptors (Koehl, 2005; Kumano et al., 2008; Kuzuyama et al., 2005). The catalytic mechanism for the NphB catalyzed prenylation of aromatic substrates is most likely an electrophilic aromatic substitution (Friedel-Crafts like alkylation). In a first step, GPP binds to the active site of the enzyme, which concurs with the observation that no NphB crystal was obtained when co-crystallizing NphB with an aromatic acceptor

in the absence of a bound GPP/GSPP. The diphosphate is orientated in the binding pocket by diphosphate-cation interactions with the magnesium ion. After binding of the aromatic acceptor, formation of a carbocation through diphosphate loss at the GPP molecule occurs. The carbocation attacks the aromatic ring of the acceptor molecule (electrophilic attack) and an σ-complex is formed followed by proton loss for neutralization of the aromatic substrate (Figure 13) (Haug-Schifferdecker et al., 2010; Kumano et al., 2008; Kuzuyama et al., 2005).

Table 2: Crystal structures of NphB. 1,6-DHN – 1,6-dihydroxynapthalene, GPP – geranyl diphosphate, GSPP - geranyl S-thiolo diphosphate, TAPS - N-[tris(hydroxymethyl)methyl]-3-aminopropanesulfonic acid (Kuzuyama et al., 2005).

PDB entry	NphB complexed with	Resolution [Å]
1ZDY	TAPS	1.44
1ZCW	Magnesium, GPP	2.25
1ZDW	Magnesium, GSPP, flaviolin	2.02
1ZB6	Magnesium, GSPP, 1,6-DHN	1.95
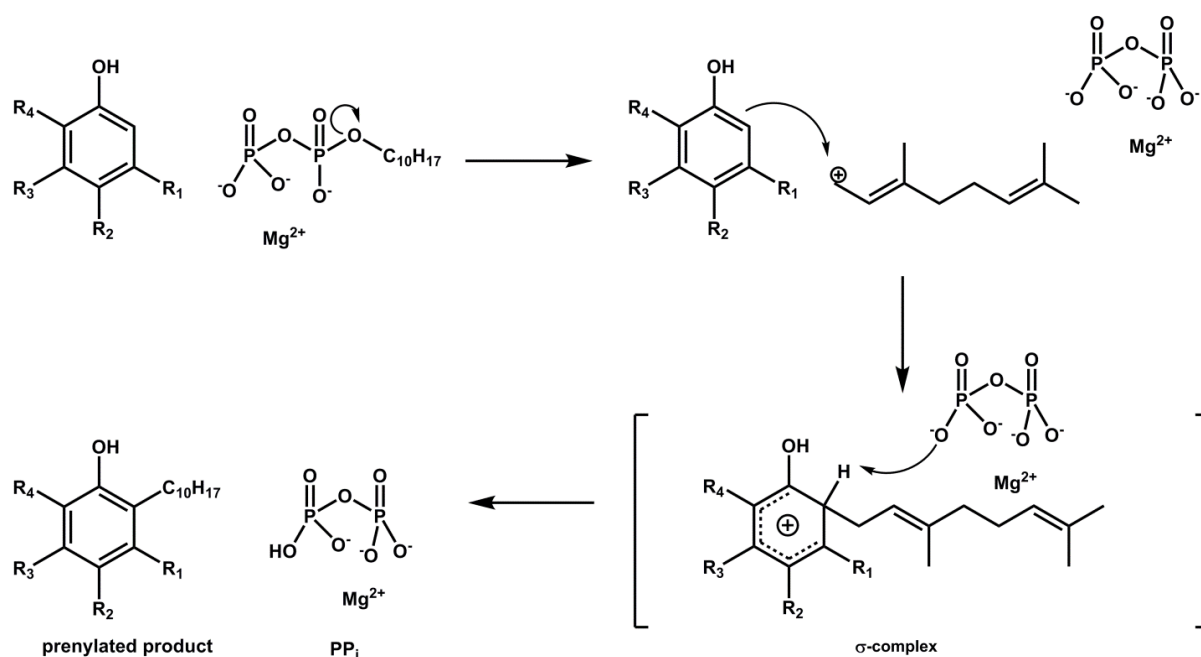
Figure 13: Possible mechanism for a NphB catalyzed reaction. Modified from Yang et al. (2012). The catalytic mechanism for the NphB catalyzed prenylation of aromatic substrates is most likely an electrophilic aromatic substitution (Friedel-Crafts like alkylation). First GPP binds to the active site of the enzyme, close to the Mg^{2+} ion which is orientated by four water molecules and the carboxyl group of Asp^{62}. The diphosphate group of GPP is coordinated in the binding pocket by diphosphate-cation interactions between Mg^{2+} and the negative charges of the diphosphate. After binding of the aromatic acceptor formation of a carbocation through diphosphate loss at GPP occurs. The carbocation attacks the aromatic acceptor (electrophilic attack) on the aromatic ring and a σ-complex is formed followed by proton loss for neutralization of the aromatic substrate (Kumano et al., 2008; Kuzuyama et al., 2005; Yang et al., 2012).

As described in chapter 1.3 the group of prenyl-converting enzymes is composed of different enzyme subgroups that are summarized in Figure 14
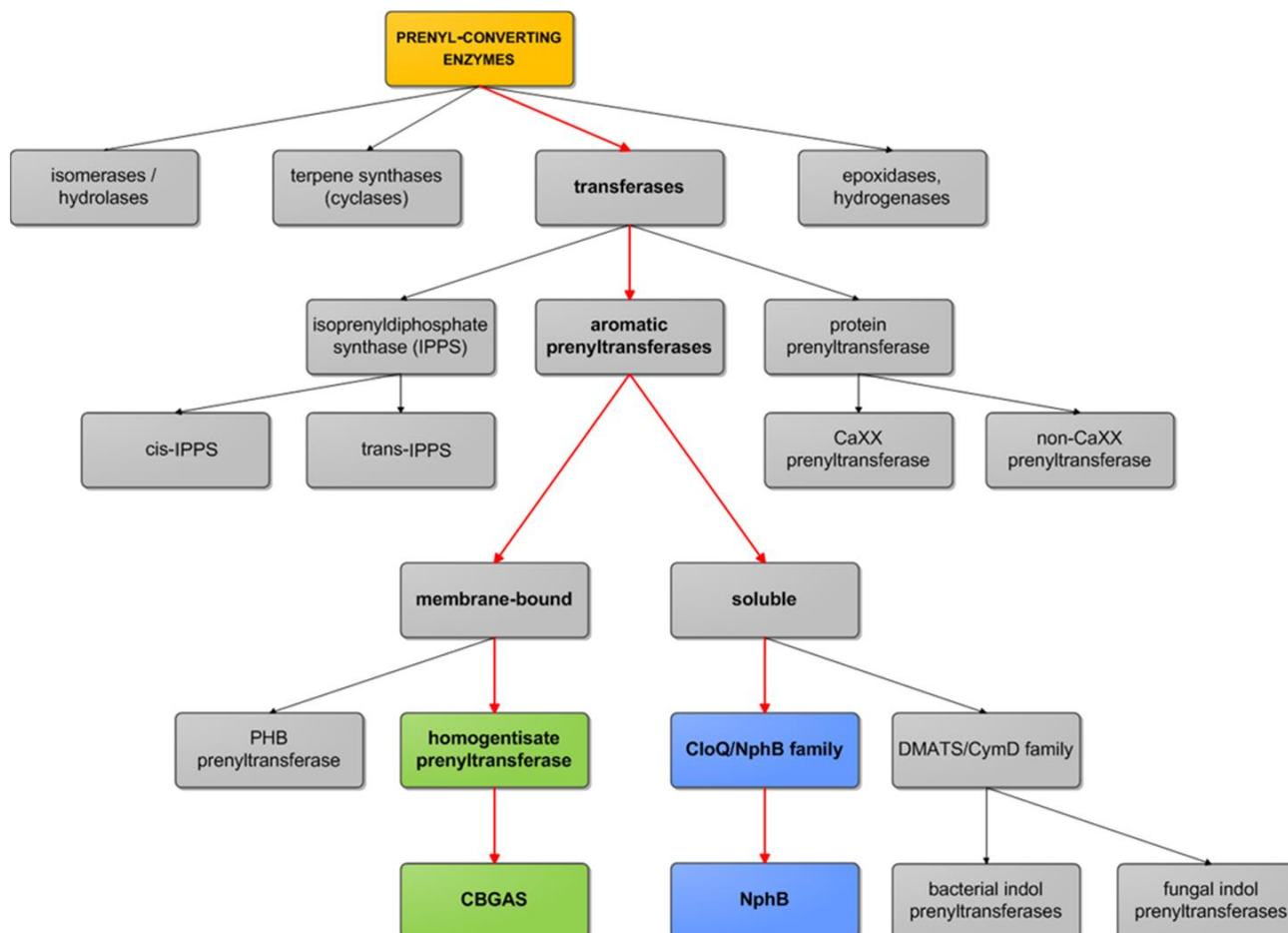
Figure 14: Prenyl-converting enzymes. The group of prenyl-converting enzymes is composed of different subgroups which are presented here. Membrane-bound aromatic prenyltransferases of plant origin and soluble aromatic prenyltransferases can be subdivided into two groups which are highlighted in green or blue, respectively. The paths of CBGAS and NphB are highlighted in red since this thesis focuses on these two enzymes (Bonitz et al., 2011; Brandt et al., 2009; Bräuer et al., 2008; Cheng and Li, 2014; Degenhardt et al., 2009; Fellermeier and Zenk, 1998; Liang et al., 2002; Ohara et al., 2004; Page and Boubakir, 2014; Tello et al., 2008; Winkelblech et al., 2015; Yazaki et al., 2009).

1.4. Scope of this thesis

Facing the increasing demand of Δ^9-THC and CBD in the medical sector, the biotechnological production of phytocannabinoids of medical interest could be, amongst other, be an environmentally friendly and cost-effective alternative to its chemical synthesis and plant extraction. For this purpose, all enzymes involved in the biosynthesis of cannabinoids in *C. sativa* have to be identified and transferred into a microbial host like *S. cerevisiae*. Yeasts, like the organism the genes of interest stem from, are eukaryotics and are able to perform posttranslational modifications. This might be important for some of the enzymes involved in the biosynthetic pathway. Additionally, *S. cerevisiae* is suitable for the use in industrial projects because it can be cultivated in the presence of high ethanol and sugar concentrains as well as low pH values. The yeast is classified as generally regarded as safe (GRAS) by the US Food and Drug Administration (FDA).

The overall goal is the generation of heterologous yeast able to produce cannabinoids. An intermediate target and aim of this study is to accomplish the enzyme-catalyzed prenylation of olivetolic acid by geranyl diphosphate to form cannabigerolic acid (CBGA). *In planta*, this C-prenylation is catalyzed by cannabigerolic acid synthase (CBGAS; CsPT1; geranylpyrophosphate:olivetolate geranyltransferase). A coding sequence of CsPT1 is published in a granted US patent (Page and Boubakir, 2014). Until now, expression experiments with the membrane-bound aromatic prenyltransferase CsPT1 have not resulted in functional expression within our department. In contrast to the US patent, Fellermeier and Zenk (1998) published that they were able to detect CBGA formation in the soluble fraction of the crude extract obtained from *Cannabis* leaves, but not in the membrane fraction. Nonetheless, they could not exclude membrane-bound activity of the CBGAS. Hence, the aims of this thesis are:

I.   To express CsPT1 in *S. cerevisiae* on a large scale using an expression protocol which has been used for functional expression of a plasma membrane protein from *Arabidopsis thaliana* before (Nissen group; DENDRITE, Aarhus University, Denmark). In addition, two other potential CBGASs, identified by Pamplaniyil (2017), will be expressed and tested in an activity assay using the isolated membrane fractions.

II.   To investigate a soluble aromatic prenyltransferase as an alternative enzyme for CBGA production. If an appropriate enzyme is identified, mutational studies in respect to the attainment of higher enzyme specificity for CBGA formation might be necessary. As expression host *Escherichia coli* will be used.

III.   To express a soluble aromatic prenytransferase catalyzing the prenylation of olivetolic acid with geranyl diphosphate in *S. cerevisiae* and optimize the expression level.

IV.   To co-express an aromatic prenyltransferase and tetrahydrocannabinolic acid synthase in *S. cerevisiae* in order to test whether it is possible to combine the last two steps of the THCA biosynthetic pathway.

Chapter 2

Material and methods

2.1. Materials

Chemicals were purchased from Carl Roth (Karlsruhe, Germany), Invitrogen (Karlsruhe, Germany), Merck (Darmstadt, Germany), Sigma-Aldrich (Darmstadt, Germany) and VWR (Darmstadt, Germany) if not stated otherwise.

Cannabigerolic acid was purchased from Taros Chemicals GmbH & Co. KG (Dortmund, Germany). Olivetolic acid was purchased from Santa Cruz Biotechnology, Inc. (Heidelberg, Germany). Δ^9-tetrahydrocannabinolic acid was purchased from THC Pharm GmbH (Frankfurt am Main, Germany). All compounds were checked for identity by ^{1}H-NMR.

Geranyl diphosphate was synthesized in-house according to Woodside et al. (1988). The synthesized GPP was analyzed by ^{1}H-NMR and ^{13}C-NMR.

Plasmid isolation from *E. coli* was performed using the NucleoSpin® Plasmid (NoLid) Kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany). PCR products and agarose gel pieces were purified using the NucleoSpin® Gel and PCR Clean-Up Kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany).

Restriction enzymes were obtained from New England BioLabs® GmbH (NEB; Frankfurt am Main, Germany).

2.2. Methods

The method chapter is divided into three sub-chapters:

I. Expression of membrane-bound prenyltransferases in *S. cerevisiae* (chapter 2.2.1)

II. Expression of *nphB* in *E. coli* (chapter 2.2.2)

III. Expression of *nphB* and *thcas* in *S. cerevisiae* (chapter 2.2.3)

2.2.1. Expression of membrane-bound prenyltransferases in *S. cerevisiae*

2.2.1.1. <u>Strains used in this study</u>

E. coli DH5α served as plasmid amplification host in all experiments. Cells were cultured at 200 rpm and 37 °C in LB medium (10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 10 g L^{-1} sodium chloride, pH 7.0 (20 g L^{-1} agar for solid medium)) containing the appropriate antibiotic if necessary.

Yeast strain *S. cerevisiae* W303-1A *Δpep4 Δgal4* was used as background in the expression experiments. Yeast cells were cultured at 30 °C and 200 rpm if not stated otherwise. *S. cerevisiae* strains without a plasmid were cultured in YPD medium (20 g L^{-1} peptone, 10 g L^{-1} yeast extract, 20 g L^{-1} glucose (20 g L^{-1} agar for solid medium)). Yeast strains carrying a plasmid were cultivated in synthetic mineral salt medium without leucine (6.7 g L^{-1} YNB without amino acids, 1.6 g L^{-1} drop out

supplements without leucine, 20 g L^{-1} glucose (20 g L^{-1} agar for solid medium)). The composition of the drop out supplements is given in the supplementary information (Table S 1).

Yeast strain W303-1A contains an ybp1-1 mutation (I7L, F328V, K343E, N571D). The yap-1-binding protein (ybp1-1) is involved in the oxidative stress response of *S. cerevisiae* and protects yeast cells against peroxide stress. The ybp1-1 mutation causes an increased peroxide sensitivity and thus an increased sensitivity to oxidative stress (Veal et al. 2003). Additionally, the yeast strain has a mutated adenine biosynthesis causing a pink phenotype of the cells under starvation conditions (Kokina et al., 2014; Ugolini and Bruschi, 1996a).

Yeast strain *S. cerevisiae* W303-1A *Δpep4 Δgal4* was transformed with plasmids containing different membrane-bound aromatic prenyltransferase coding sequences and used for expression of the corresponding enzymes.

Table 3 contains all relevant genotypes and applications of the strains used in this study. A table containing the genotypes is given in the supplementary information (Table S 2).

Table 3: Relevant genotypes and applications of the yeast strains used for the expression of membrane-bound aromatic prenyltransferases. + yEGFP – expressed with C-terminal yEGFP as expression reporter, w/o yEGFP – expressed without C-terminal yEGFP as expression reporter, yEGFP – yeast enhanced green fluorescent protein.

Strain	Relevant genotype / transformed strain	Transformed plasmid	Reference
SC-BY4742	*S. cerevisiae* BY4742	-	EUROSCARF (Y100000) (Baker Brachmann et al., 1998)
SC-gal4	*S. cerevisiae* W303-1A Δ*pep4* Δg*al4*	-	kind gift of Prof. Poul Nissen, Aarhus University, Denmark
SC-gal4_CsPT1+	SC-gal4	pDio-CsPT1 + yEGFP	This study
SC-gal4_CsPT1-	SC-gal4	pDio-CsPT1 w/o yEGFP	This study
SC-gal4_CsPT2+	SC-gal4	pDio-CsPT2 + yEGFP	This study
SC-gal4_CsPT2-	SC-gal4	pDio-CsPT2 w/o yEGFP	This study
SC-gal4_CsPT3+	SC-gal4	pDio-CsPT3 + yEGFP	This study
SC-gal4_CsPT3-	SC-gal4	pDio-CsPT3 w/o yEGFP	This study
SC-gal4_yCoq2SP-CsPT1+	SC-gal4	pDio-yCoq2SP-CsPT1 + yEGFP	This study
SC-gal4_yCoq2SP-CsPT1-	SC-gal4	pDio-yCoq2SP-CsPT1 w/o yEGFP	This study
SC-gal4_yCoq2SP-CsPT2+	SC-gal4	pDio-yCoq2SP-CsPT2 + yEGFP	This study
SC-gal4_yCoq2SP-CsPT2-	SC-gal4	pDio-yCoq2SP-CsPT2 w/o yEGFP	This study
SC-gal4_yCoq2SP-CsPT3+	SC-gal4	pDio-yCoq2SP-CsPT3 + yEGFP	This study
SC-gal4_yCoq2SP-CsPT3-	SC-gal4	pDio-yCoq2SP-CsPT3 w/o yEGFP	This study

2.2.1.2. <u>Plasmid construction and transformation</u>

Plasmids used and generated in this study are given in Table 4. Amino acid sequences of CsPT1, CsPT2, CsPT3 (Table S 3) and the yCoq2 signal peptide (Table S 4), primers (Table S 5) and vector maps of the plasmids pDDGFP-2 (Figure S 1) and pDionysos (Figure S 2) are given in the supplementary information.

All plasmids were constructed either by restriction / ligation, by Gibson Assembly (Gibson et al., 2009) or by homologous recombination. The yeast expression vector pDioynsos was used as background (Stehle et al., 2008).

Restriction / ligation

The insert and vector fragments were digested with the desired restriction enzymes, purified and ligated using Quick Ligation[TM] Kit (NEB, Frankfurt am Main, Germany). *E. coli* DH5α transformation was performed using a standard protocol and heat shock transformation (Inoue et al., 1990). Plasmids were isolated for verification via polymerase chain reaction (PCR) (Red-Taq DNA Polymerase 1,1X MasterMix; VWR, Darmstadt, Germany) and sequencing (Seqlab, Göttingen, Germany).

Gibson cloning

For Gibson Assembly the required DNA insert and vector fragments containing 25 bp overlaps were obtained by PCR using Q5[®] High-Fidelity 2X Master Mix (NEB, Frankfurt am Main, Germany). Vector fragments were digested with *DpnI* and purified. DNA insert fragments were purified by gel extraction. Vector and insert fragments were added to a 2X Gibson Assembly Master Mix (NEB, Frankfurt am Main, Germany) and Gibson Assembly was performed according to the manufacturer. The plasmids were transformed into *E. coli* DH5α cells using a standard protocol and heat shock transformation (Inoue et al., 1990). Plasmids were isolated for verification via PCR and sequencing.

Homologous recombination

For homologous recombination the desired vector was linearized by restriction digestion, followed by incubation with calf intestinal alkaline phosphatase (CIAP; Invitrogen, Carlsbad, California, USA) and purification. The insert fragment was obtained by PCR using primers with 30 – 50 bp homologous to the vector and Q5[®] High-Fidelity 2X Master Mix. The PCR product was purified by gel extraction. *S. cerevisiae* BY4742 cells were transformed with linearized vector and the insert fragment with a ratio of 1:10 using LiAc/ss carrier DNA/PEG method (Gietz and Schiestl, 2007), plated on synthetic mineral salt medium without uracil agar plates (6.7 g L^{-1} YNB without amino acids, 1.9 g L^{-1} drop out supplements without uracil, 20 g L^{-1} glucose, 20 g L^{-1} agar) and incubated at 30 °C for 2 - 3 days. The plasmid was isolated from yeast using a yeast modified protocol of the NucleoSpin[®] Plasmid (NoLid) Kit. Plasmids were verified via PCR and sequencing.

The coding sequence of the yCoq2 signal peptide (yCoq2SP; Table S 4) was cloned in N-terminal direction of CsPT1, CsPT2 and CsPT3 by mutagenesis PCR using primers that contained the yCoq2SP coding sequence. The primer design based on Liu and Naismith (2008). For PCR Q5® High-Fidelity 2X Master Mix was applied. The PCR products (plasmids) were digested with *Dpn*I and transformed into chemically competent *E. coli* DH5α using a standard protocol and heat shock transformation (Inoue et al., 1990). Plasmids were isolated and sent for sequencing.

Plasmids with the correct sequence were transformed into frozen competent *S. cerevisiae* cells using LiAc/ss carrier DNA/PEG method (Gietz and Schiestl, 2007) and plated on synthetic mineral salt medium without uracil agar plates (6.7 g L^{-1} YNB without amino acids, 1.9 g L^{-1} drop out supplements without uracil, 20 g L^{-1} glucose, 20 g L^{-1} agar) and incubated at 30 °C for 2 - 3 days. A few colonies from that plate were resuspended in dH$_2$O and plated on synthetic mineral salt medium without leucine (6.7 g L^{-1} YNB without amino acids, 1.6 g L^{-1} drop out supplements without leucine, 20 g L^{-1} glucose, 20 g L^{-1} agar) and incubated at 30 °C for 2 - 3 days.

Table 4: Plasmids used for the expression of membrane-bound aromatic prenyltransferases. + yEGFP – expressed with C-terminal yEGFP as expression reporter, w/o yEGFP – expressed without C-terminal yEGFP as expression reporter, yEGFP – yeast enhanced green fluorescent protein.

Plasmid	Resistance / auxotrophy marker	Expressed coding sequence	Reference
pDDGFP-2	Amp, Ura	-	(Drew et al., 2008)
pDionysos	Amp, Ura, Leu2d	-	(Stehle et al., 2008)
pDio + yEGFP	Amp, Ura, Leu2d	-	This study
pDio-CsPT1 + yEGFP	Amp, Ura, Leu2d	P$_{GAL1}$: CsPT1 – yEGFP	This study
pDio-CsPT1 w/o yEGFP	Amp, Ura, Leu2d	P$_{GAL1}$: CsPT1	This study
pDio-CsPT2 + yEGFP	Amp, Ura, Leu2d	P$_{GAL1}$: CsPT2 – yEGFP	This study
pDio-CsPT2 w/o yEGFP	Amp, Ura, Leu2d	P$_{GAL1}$: CsPT2	This study
pDio-CsPT3 + yEGFP	Amp, Ura, Leu2d	P$_{GAL1}$: CsPT3 – yEGFP	This study
pDio-CsPT3 w/o yEGFP	Amp, Ura, Leu2d	P$_{GAL1}$: CsPT3	This study
pDio-yCoq2SP-CsPT1 + yEGFP	Amp, Ura, Leu2d	P$_{GAL1}$: yCoq2SP - CsPT1 – yEGFP	This study

Table 4 continued

Plasmid	Resistance / auxotrophy marker	Expressed coding sequence	Reference
pDio-yCoq2SP-CsPT1 w/o yEGFP	Amp, Ura, Leu2d	P_{GAL1}: yCoq2SP-CsPT1	This study
pDio-yCoq2SP-CsPT2 + yEGFP	Amp, Ura, Leu2d	P_{GAL1}: yCoq2SP – CsPT2 – yEGFP	This study
pDio-yCoq2SP-CsPT2 w/o yEGFP	Amp, Ura, Leu2d	P_{GAL1}: yCoq2SP-CsPT2	This study
pDio-yCoq2SP-CsPT3 + yEGFP	Amp, Ura, Leu2d	P_{GAL1}: yCoq2SP – CsPT3 – yEGFP	This study
pDio-yCoq2SP-CsPT3 w/o yEGFP-	Amp, Ura, Leu2d	P_{GAL1}: yCoq2SP-CsPT3	This study

2.2.1.3. Expression conditions

The expression of the membrane-bound prenyltransferases was performed using shaking flasks with three baffles. The volume of the cell cultures was kept at 10% of the total flask volume. The cultivation of the yeast cells containing the plasmids with the relevant prenyltransferase gene was performed in two precultures and one main culture. For the first precultures, synthetic mineral salt medium without leucine (6.7 g L^{-1} YNB without amino acids, 1.6 g L^{-1} drop out supplements without leucine, 20 g L^{-1} glucose) was inoculated and the cells were incubated for 24 h at 30 °C and 200 rpm. The second preculture, using synthetic mineral salt medium without leucine, was inoculated utilizing the first preculture and incubated for 12 h at 30 °C and 200 rpm. The cells were harvested (2000 x g, 5 min, 20 °C), resuspended in dH_2O and used for inoculation of the main culture.

In the main culture *S. cerevisiae Δpep4 Δgal4* cells carrying the desired plasmid were grown in complex medium (20 g L^{-1} yeast extract, 20 g L^{-1} Bacto^{TM} Peptone, 2.8% (v/v) ethanol and 10 g L^{-1} glucose). The main culture was inoculated with a final OD_{600} of 0.05 and cultivated at 28 °C and 120 rpm for 36 h. Protein expression was induced by addition of 2% galactose, followed by cultivation at 18 °C and 120 rpm for 12 h. Subsequently, induction was repeated by addition of galactose. Cells were cultivated at 18 °C and 120 rpm as long as the fluorescence intensity (FI) / OD_{600} ratio (chapter 2.2.1.4) was increasing.

Cells were harvested by centrifugation (2500 x g, 10 min, 4 °C), washed with cold dH_2O and used for membrane isolation (chapter 2.2.1.5).

2.2.1.4. <u>Measurement of the whole-cell fluorescence</u>

The fluorescence intensity (FI) of whole yeast cells was determined using a FLUOstar Omega (BMG Labtech, Ortenberg, Germany) microplate reader (software: Omega version 1.30).

2 x 5 mL of each cell culture were centrifuged (2000 x g, 5 min, 4 °C), the pellets were washed with dH_2O and resuspended in 800 µL dH_2O. PureGrade^{TM} 96-well plates (polystyrene, transparent bottom, black; BRAND, Wertheim, Germany) were applied for the determination. The samples were measured in duplicates. The excitation occurred at a wavelength of 485 nm, the emission was detected at 520 nm. The obtained fluorescence intensities were normalized by the OD_{600} of the cell culture in order to correct for variations in cell number. If the FI / OD_{600} ratio decreased the experiment was stopped because there is more free green fluorescent protein (GFP) present in the cell than GFP-fusion protein.

2.2.1.5. <u>Membrane isolation</u>

The harvested yeast cells (chapter 2.2.1.3) were resuspended in 1 mL GTEB buffer / g cell wet weight (50 mM Tris-HCl buffer pH 7.5, 20% (w/v) glycerol, 5 mM β-mercaptoethanol, 1 mM phenylmethylsulfonyl fluoride) and 1 g of glass beads (diameter: 0.4 - 0.6 mm) / mL cell suspension were added. The samples were vortexed twice for 15 min each at full speed and at 4 °C, followed by centrifugation of the cell suspension (2000 x g, 10 min, 4 °C). Cell debris and uncracked cells were removed by centrifugation of the supernatant (4000 x g, 20 min, 4 °C). Subsequently, yeast membranes were isolated by ultracentrifugation (204,526 x g, 90 min, 4 °C). The supernatant was discarded and the isolated membranes were homogenized in 1 mL GTEB buffer / g membranes using a glass homogenizator. The homogenized membranes were used as enzyme solution for activity assays (chapter 2.2.1.6).

2.2.1.6. <u>Membrane protein activity assay</u>

The membrane protein activity assays were performed using homogenized membranes (chapter 2.2.1.5). Isolated membranes from non-induced yeast cell cultures were used as controls.

The activity assays of CsPT1, CsPT2 and CsPT3 were performed similarly to the assay protocols published by Fellermeier and Zenk (1998) and Page and Boubakir (2014). Each assay contained 5 mM magnesium chloride, 1 mM geranyl diphosphate, 1 mM prenyl acceptor and 84 µL homogenized membranes. The enzymatic activity of the different prenyltransferases was tested using olivetolic acid, olivetol, *p*-hydroxybenzoic acid and resveratrol as prenyl acceptors. Each test was run in duplicates. The samples were incubated for 4 h at 37 °C and 1100 rpm. Activity assays were stopped by the addition of 0.6 M hydrochloric acid, followed by extraction with ethyl acetate. The ethyl acetate layers were evaporated, resuspended in methanol, filtered (0.4 µm, nylon) and analyzed by liquid chromatography – mass spectrometry (LC-MS) (chapter 2.2.1.7).

2.2.1.7. <u>LC-MS</u>

Separation of compounds from *S. cerevisiae* extracts was performed on a Poroshell 120 EC-C18 2.7 µm column (Agilent Technologies, Waldbronn, Germany). Activity assays applying OA or olivetol as a prenyl acceptor were analyzed using an isocratic method (0.7 mL min^{-1}, 40 °C, 35% (v/v) H$_2$O with 0.1% (v/v) formic acid (FA), 65% (v/v) acetonitrile (AcN)). Resveratrol containing assays were analyzed in a gradient elution mode (0.4 mL min^{-1}, 35 °C, 50% (v/v) AcN (A) and 50% (v/v) H$_2$O with 0.1% (v/v) FA (B) → 100% A and 0% B → 50% A and 50% B). Activity assays performed with PHB were also analyzed using a gradient elution mode (0.4 mL min^{-1}, 35 °C, 90% (v/v) AcN (A) and 10% (v/v) H$_2$O with 0.1% (v/v) FA (B) → 20% A and 80% B → 90% A and 10% B). The identity of all compounds was confirmed by comparing mass spectra of each sample with coeluting compounds analyzed by Bruker compact^{TM} ESI-Q-TOF (Bruker, Bremen, Germany) using positive ionization mode.

2.2.2. Expression of *nphB* in *Escherichia coli*

2.2.2.1. <u>Strains used in this study</u>

E. coli DH5α was used to amplify plasmids. Cells were cultured at 37 °C and 200 rpm in LB medium (10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 10 g L^{-1} sodium chloride, pH 7.0) containing the appropriate antibiotic if necessary.

Expression experiments were performed using different *E. coli* strains (Table 5). Cells were cultured either in LB medium or in TB medium (12 g L^{-1} tryptone, 24 g L^{-1} yeast extract, 5 g L^{-1} (w/v) glycerol, 17 mM potassium phosphate monobasic, 17 mM dibasic potassium phosphate) containing the appropriate antibiotic.

Table 5: Genotypes and applications of the used *E. coli* strains.

Strain	Genotype	Application	Reference
E. coli DH5α	F$^-$ endA1, hsdR17(rk$^-$ mk$^+$) supE44, thi-l λ−recAl gyrA96 relA1 φ80ΔlacAm15	Plasmid amplification	Life Technologies; Darmstadt, Germany
E. coli BL21 (DE3)	fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHlo ΔEcoRI-B int::(lacl::PlacUV5::T7 gene1) i21 Δnin5	Expression of *nphB*	NEB; Frankfurt am Main, Germany
E. coli BL21 Gold (DE3)	E. coli B F$^-$ ompT hsdS (r$_B^-$ m$_B^-$) dcm$^+$ TetR gal λ (DE3) endA The	Expression of *nphB*	Agilent Technologies; Santa Clara, CA, USA
E. coli BL21 Gold (DE3) pLysS	E. coli B F$^-$ ompT hsdS (r$_B^-$ m$_B^-$) dcm$^+$ TetR gal λ (DE3) endA Hte [pLysS CamR]	Expression of *nphB*	Agilent Technologies; Santa Clara, CA, USA
E. coli OverExpress^{TM} C41 (DE3) pLysS	F$^-$ ompT hsdS$_B$ (r$_B^-$ m$_B^-$) gal dcm (DE3) pLysS (CmR)	Expression of *nphB*	Lucigen$^®$; Middleton, WI, USA
E. coli OverExpress^{TM} C41 (DE3) pLysS	F$^-$ ompT hsdS$_B$ (r$_B^-$ m$_B^-$) gal dcm (DE3) pLysS (CmR)	Expression of *nphB*	Lucigen$^®$; Middleton, WI, USA
E. coli Rosetta^{TM} 2 (DE3)	F$^-$ ompT hsdS$_B$(r$_B^-$ m$_B^-$) gal dcm (DE3) pRARE2 (CamR)	Expression of *nphB*	Merck Millipore; Darmstadt, Germany
E. coli Rosetta^{TM} 2 (DE3) pLysS	F$^-$ ompT hsdS$_B$(r$_B^-$ m$_B^-$) gal dcm (DE3) pLysSRARE2 (CamR)	Expression of *nphB*	Merck Millipore; Darmstadt, Germany

2.2.2.2. <u>Plasmid construction and transformation</u>

Plasmids used and constructed in this study are shown in Table 6. The amino acid sequence of NphB wt (Table S 10) and the used primer sequences (Table S 8, Table S 9) are given in the supplementary information.

Table 6: List of plasmids used for the expression of *nphB* in *E. coli*.

Plasmid	Resistance	Expressed coding sequence	Reference
pET32a(+)	Amp	-	Merck Millipore; Darmstadt, Germany
pET32a-NphB	Amp	P_{T7}: trxA – His$_6$ tag – NphB	This study
pET32a-NphB-GFP	Amp	P_{T7}: trxA – His$_6$ tag – NphB – GFP	This study
pET32a-T126A	Amp	P_{T7}: trxA – His$_6$ tag – NphB (T126A)	This study
pET32a-T126G	Amp	P_{T7}: trxA – His$_6$ tag – NphB (T126G)	This study
pET32a-T126V	Amp	P_{T7}: trxA – His$_6$ tag – NphB (T126V)	This study
pET32a-D127A	Amp	P_{T7}: trxA – His$_6$ tag – NphB (D127A)	This study
pET32a-Q161N	Amp	P_{T7}: trxA – His$_6$ tag – NphB (Q161N)	This study
pET32a-M162A	Amp	P_{T7}: trxA – His$_6$ tag – NphB (M162A)	This study
pET32a-M162K	Amp	P_{T7}: trxA – His$_6$ tag – NphB (M162K)	This study
pET32a-M162N	Amp	P_{T7}: trxA – His$_6$ tag – NphB (M162N)	This study
pET32a-M162W	Amp	P_{T7}: trxA – His$_6$ tag – NphB (M162W)	This study
pET32a-Y175A	Amp	P_{T7}: trxA – His$_6$ tag – NphB (Y175A)	This study
pET32a-Y175N	Amp	P_{T7}: trxA – His$_6$ tag – NphB (Y175N)	This study
pET32a-F213A	Amp	P_{T7}: trxA – His$_6$ tag – NphB (F213A)	This study
pET32a-S214D	Amp	P_{T7}: trxA – His$_6$ tag – NphB (S214D)	This study
pET32a-S214N	Amp	P_{T7}: trxA – His$_6$ tag – NphB (S214N)	This study
pET32a-V271N	Amp	P_{T7}: trxA – His$_6$ tag – NphB (V271N)	This study
pET32a-T126V/Q161A	Amp	P_{T7}: trxA – His$_6$ tag – NphB (T126V/Q161AA)	This study
pET32a-T126V/Q161E	Amp	P_{T7}: trxA – His$_6$ tag – NphB (T126V/Q161E)	This study

Table 6 continued

Plasmid	Resistance	Expressed coding sequence	Reference
pET32a-T126V/Q161N	Amp	P_{T7}: trxA – His$_6$ tag – NphB (T126V/Q161N)	This study
pET32a-M162A/Y175A	Amp	P_{T7}: trxA – His$_6$ tag – NphB (M162A/Y175A)	This study
pET32a-M162A/V271N	Amp	P_{T7}: trxA – His$_6$ tag – NphB (M162N/V271N)	This study

Expression plasmids were generated by restriction / ligation or Gibson Assembly. All expression plasmids are based on the vector pET32a.

Restriction / ligation

NphB was cloned into the vector pET32a by restriction / ligation using the restriction sites *Bam*HI and *Sac*I. The desired restriction recognition sites at the C- and N-terminus of the DNA insert were obtained by PCR using Q5® High-Fidelity 2X Master Mix (NEB, Frankfurt am Main, Germany) and primers containing the appropriate recognition site sequences. The PCR product was purified by agarose gel extraction. Digested plasmids were purified by agarose gel isolation. DNA ligation was performed using a Quick Ligation™ Kit (NEB, Frankfurt am Main, Germany). *E. coli* DH5α transformation was executed using a standard protocol and heat shock transformation (Inoue et al., 1990). Plasmids were isolated for verification via PCR (Red-Taq DNA Polymerase 1,1X MasterMix; VWR, Darmstadt, Germany) and sequencing (Seqlab, Göttingen, Germany).

Gibson Assembly

The DNA insert and vector fragments required for Gibson cloning (Gibson et al., 2009) were obtained by PCR using Q5® High-Fidelity 2X Master Mix (NEB, Frankfurt am Main, Germany) and primers with 25 bp overlaps. Vector fragments were digested with *Dpn*I and purified. DNA insert fragments were purified by agarose gel extraction. Vector and insert fragments were applied in an equimolar ratio and added to 2X Gibson Assembly Master Mix (NEB) and Gibson Assembly was performed according to the manufacturer. *E. coli* DH5α transformation was performed using a standard protocol and heat shock transformation (Inoue et al., 1990). Plasmids were isolated for verification via PCR and sequencing.

Site-directed mutagenesis

nphB mutants were obtained by mutagenesis PCR using primers containing the desired mutation. For primer design two different strategies were used: With the first strategy the primers had a length of about 45 bp with the desired mutation in the middle of the sequence, a GC content of approximately 40% and a melting temperature of $\geq$ 78 °C. The primer design of the second strategy based on Liu and Naismith (2008). For mutagenesis PCR AccuPrime Pfx Super Mix (Invitrogen, Karlsruhe, Germany) or Q5® High-Fidelity 2X Master Mix were applied. The PCR products (plasmids) were digested with *DpnI* and transformed in *E. coli* DH5α using a standard protocol and heat shock transformation (Inoue et al., 1990). Plasmids were isolated and sent for sequencing.

Plasmids with the correct sequence were transformed into cells of the desired *E. coli* strain using a standard protocol and heat shock transformation (Inoue et al., 1990).

2.2.2.3. **Expression of *nphB* in shaking flasks**

The expression of *nphB* in different *E. coli* strains was performed with two precultures and one main culture.

Colonies of the *E. coli* strain transformed with the desired plasmid were cultivated in LB medium (10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 10 g L^{-1} sodium chloride, pH 7.0), containing 200 µg mL^{-1} ampicillin, for 16 – 18 h at 200 rpm and 37 °C. The first preculture was used for the inoculation of the second preculture which was incubated at 37 °C and 200 rpm. At an OD_{600} of 2.5 the cells were harvested, washed with dH_2O and used for the inoculation of the main culture. 330 mL LB medium, containing 200 µg mL^{-1} ampicillin, in 1 L shaking flasks were inoculated with OD_{600} of 0.2. Cultures were incubated at 37 °C and 200 rpm until an OD_{600} of 0.6 – 0.7 was reached. Protein expression was induced by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), followed by cultivation for 17 h at 25 °C and 160 rpm. After taking samples for SDS-PAGE (chapter 2.2.2.5) cells were harvested (4000 x g, 10 min, 4 °C), washed with PBS (137 mM sodium chloride, 27 mM potassium chloride, 10 mM sodium phosphate dibasic, 1.8 mM potassium phosphate monobasic, pH 7.4) and stored at -20 °C.

Cell lysis was performed by sonication (BRANSON Digital Sonifier® Cell Disruptor; Fisher Scientific, Pittsburgh, USA). Cell pellets were resuspended in 3 mL assay buffer (50 mM Tris-HCl buffer pH 7.5, 10% (w/v) glycerol, 100 mM sodium chloride) per gram cell wet weight and lysed twice in 6 s intervals with 10% amplitude over a total time of 60 s. The cell suspensions were centrifuged (40,000 x g, 15 min, 4 °C) and cell lysate supernatants were applied for protein NphB activity assays (5 mM magnesium chloride, 5 mM prenyl acceptor (OA, olivetol or 1,6-DHN), 5 mM GPP, 37 °C, 1100 rpm, 4 h) or protein purification (chapter 2.2.2.3).

2.2.2.4. **Expression of *nphB* in 48-well microplates**

Colonies of *E. coli* BL21 (DE3) transformed with the desired plasmid were cultivated overnight in LB medium (10 g L^{-1} tryptone, 5 g L^{-1} yeast extract, 10 g L^{-1} sodium chloride, pH 7.0), containing 400 µg mL^{-1} ampicillin, at 200 rpm and 37 °C. 800 µL of each culture were centrifuged (16,100 x g, 5 min, 20 °C) and resuspended in 800 µL LB medium containing 400 µg mL^{-1} ampicillin. The cell suspensions were transferred to a sterile 48-round-well Biolector® plate (m2p-labs, Baesweiler, Germany). Cultures were incubated using the Biolector® (m2p-labs, Baesweiler, Germany) at 1200 rpm and 37 °C for 2 h. 40 µL of each preculture was transferred by an attached pipetting robot Sias Xantus (Tecan Group Ltd., Männedorf, Switzerland) to 760 µL TB medium (12 g L^{-1} tryptone, 24 g L^{-1} yeast extract, 0.4% (w/v) glycerol, 0.17 M dipotassium hydrogen phosphate, 0.72 M potassium dihydrogen phosphate) containing 400 µg mL^{-1} ampicillin. Cell cultures were incubated for another 2 h at 1200 rpm and 37 °C using the Biolector®. Subsequently, the temperature was cooled to 25 °C and the incubation at 1200 rpm was continued for 17 h. The protein expression was induced by the addition of 1 mM IPTG at OD_{600} of 1.2. Cells were harvested by centrifugation and the pellet was resuspended in 500 µL assay buffer (50 mM Tris-HCl buffer pH 7.5, 10% (w/v) glycerol, 100 mM sodium chloride, 1 mM phenylmethylsulfonyl fluoride).

Cells were lysed twice by sonication in 0.1 s intervals with 40% amplitude and a total time of 30 s. The cell suspensions were centrifuged (16,100 x g, 10 min, 4 °C) and cell lysate supernatants were applied for NphB activity assays (5 mM magnesium chloride, 5 mM OA, 0.5 mM GPP, 37 °C, 1100 rpm, 15 min).

2.2.2.5. **SDS-PAGE**

SDS-PAGE samples were centrifuged (16,100 x g, 5 min, 4 °C) and the pellets were resuspended in 100 µL 2x SDS sample buffer (150 mM Tris-HCl buffer pH 6.8, 1.2% sodium dodecyl sulfate, 30% (w/v) glycerol, 2.15 M β-mercaptoethanol) followed by 10 min incubation at 96 °C.

The gel was run with a Bio-Rad Mini-Protean® Tetra System (Bio-Rad, München, Germany). The gel consisted of a 4% stacking gel and a 12.5% separating gel. PageRuler™ Plus Prestained Protein Ladder (Fisher Scientific, Schwerte Germany) was used as a marker and SDS-PAGE was carried out according to Laemmli (1970).

2.2.2.6. **Immobilized metal ion affinity chromatography (IMAC)**

Cell lysate supernatant (chapter 2.2.2.3) was incubated on ice for 30 min with *DNase*I (NEB) with a final concentration of 10 µg mL^{-1}. The crude enzyme extract was centrifuged (40,000 x g, 15 min, 4 °C) and the supernatant was loaded, with a flow rate of 1 - 2 mL min^{-1}, onto a 1 mL HisTrap™ FF crude column (GE Healthcare, Freiburg, Germany), equilibrated with binding buffer (20 mM Tris-HCl pH 7.4, 500 mM sodium chloride, 30 mM imidazole, 1 mM dithiothreitol (DTT)). After washing the column with 10 column volumes (CV) binding buffer, 1 mL elution buffer (50 mM Tris-HCl pH 7.5, 500 mM sodium chloride, 250 mM imidazole, 1 mM DTT) was loaded onto the column and the flow through was

collected. After incubation for 30 min at 4 °C four times 1 mL elution buffer were applied for eluting NphB. The column was cleaned with 12 mL elution buffer, 10 CV dH_2O, 10 CV 20% ethanol and stored 4 °C: The success of protein purification was checked by SDS-PAGE (chapter 2.2.2.5).

Based on the results obtained by a NanoDrop 2000 Spectrophotometer (peqlab; Erlangen, Germany) and SDS-PAGE (chapter 2.2.2.5) fractions containing the highest amounts of purified NphB were loaded onto a 0.5 mL Zeba^{Tm} Spin Desalting Column (Thermo Fisher Scientific, Schwerte Germany) with a molecular weight cut-off (MWCO) of 7 K in order to change the buffer. For higher samples volumes PD-10 desaltings columns (Sephadex G-25 resin; GE Healthcare Life Sciences, Freiburg, Germany) were used. The buffer exchange was performed according to the manufactureres.

Purified NphB was stored in assay buffer (50 mM Tris-HCl buffer pH 7.5, 10% (w/v) glycerol, 100 mM sodium chloride) and used in a NphB activity assay (5 mM magnesium chloride, 5 mM prenyl acceptor, 5 mM GPP, 37 °C, 1100 rpm, 4 h) or for enterokinase treatment (chapter 2.2.2.7).

2.2.2.7. Removal of the TrxA-His$_6$ tag by enterokinase

After buffer exchange (chapter 2.2.2.6) purified NphB was incubated with 0.001% (w/w) enterokinase light chain (NEB) for 24 h at 23 °C. NphB and the removed TrxA-His$_6$ tag were separated by IMAC (chapter 2.2.2.6). Buffer exchange was performed using Zebra Spin desalting column (cut-off: 7,000). The enterokinase was inhibited by the addition of trypsin inhibitor from *Glycine max* (Sigma-Aldrich Chemie GmbH, Steinheim, Germany) in a ratio of 1:1 (w/w). The enzyme solution was concentrated using a Vivaspin 500 column (10,000 MWCO PES; Sartorius Stedim Plastics GmbH, Göttingen, Germany) and applied in a NphB activity assay (5 mM magnesium chloride, 5 mM prenyl acceptor, 2 mM GPP, 37 °C, 40 min).

2.2.2.8. LC-MS, LC-DAD

Activity assays were stopped by the addition of 2.9 assay-volumes acetonitrile (AcN) and 0.1 assay-volumes formic acid (FA), followed by an incubation on ice for 20 min. Supernatants were filtered (0.45 μm, nylon) after centrifugation (20,000 x g, 25 min, 20 °C) and analyzed by LC-DAD/MS (Agilent 1260 Infinity HPLC, Waldbronn, Germany). Separation of compounds was performed on a Poroshell 120 EC-C18 2.7 μm column (Agilent Technologies, Waldbronn, Germany) or on a Nucleosil$^®$ 100-5 EC-C18 5.0 μm column (Macherey-Nagel, Düren, Germany).

Activity assays of *nphB* expressed in *E. coli* cells cultivated in 48-well-plates were analyzed using an isocratic method (0.7 mL min^{-1}, 35 °C, 39% (v/v) H_2O with 0.1% (v/v) FA, 61% (v/v) AcN). Quantification of products was based on integrated peak areas of UV chromatograms at 225 nm. A standard curve for CBGA was generated and it was used to estimate the concentrations of CBGA and 2-*O*-GOA.

Activity assays of NphB incubated with 1,6-DHN as prenyl acceptor were analyzed in a gradient elution mode (0.4 mL min^{-1}, 35 °C, 50% (v/v) AcN (A) and 50% (v/v) H_2O with 0.1% (v/v) FA (B) → 100% A and 0% B → 50% A and 50%B). Quantification of products was based on integrated peaks of

UV chromatograms at 250 nm. A standard curve for 1,6-DHN was generated and used to estimate the concentrations of monogeranylated 1,6-DHN. Standard substances of geranylated 1,6-DHN were not available commercially.

Activity assays of *nphB* expressed in *E. coli* cells cultivated in shaking flasks using olivetol or OA as prenyl acceptor were analyzed using an isocratic method (0.7 mL min^{-1}, 35 °C, 25% (v/v) H_2O with 0.1% (v/v) FA, 75% (v/v) AcN). Quantification of products was based on integrated peaks of UV chromatograms at 225 nm. A standard curve for CBGA was generated and used to estimate the concentrations of CBGA and 2-*O*-GOA. A standard curve for olivetol was generated and was used to estimate the concentrations of geranylated olivetol. Standard substances of geranylated olivetol were not available.

2.2.2.9. LC-MS/MS

Data were collected using a Poroshell 120 EC-C18 2.1 x 100 mm 2.7 µm column (Agilent Technologies, Waldbronn, Germany) and an isocratic method (0.8 mL min^{-1}, 40 °C, 35% (v/v) H_2O with 0.1% FA, 65% (v/v) AcN). The tandem mass spectra were obtained by a Bruker compact^{TM} ESI-Q-TOF (Bruker, Bremen, Germany) using positive electrospray (ESI) mode with collision energy of 28 eV (Zirpel, Degenhardt et al., 2017).

2.2.2.10. Preparative RP-HPLC

The isolation of 2-*O*-GOA was executed by extraction of NphB assay samples with ethyl acetate. The extraction was performed on a Nucleodor C18 HTec 5 µm (250 x 10 mm) column (Macherey Nagel, Düren, Germany) using an isocratic gradient (4 mL min^{-1}, 40 °C, 35% (v/v) H_2O, 65% (v/v) AcN). The 2-*O*-GOA peak was detected at 225 nm and fractionated with a FC-1 Dynamax fraction collector (Zinsser Analytik, Frankfurt am Main, Germany). The sample was dried under vacuum, followed by lyophilisation. The obtained product was analyzed by ^{1}H-NMR measurement (Zirpel, Degenhardt et al., 2017).

2.2.2.11. ^{1}H-NMR measurement

CBGA and THCA were dissolved in CDCl$_3$. 10 mg of in-house synthesized GPP were dissolved in D$_2$O with a few drops of NaOD to keep the diphosphate stable during measurement. OA was dissolved in d6-DMSO. Samples were measured on a Bruker AV 500 Avance III HD NMR (Prodigy) (Bruker, Karlsruhe, Germany). About 500 µg of isolated 2-*O*-GOA (chapter 2.2.2.10) were dissolved in 500 µL CDCl$_3$ and data were collected on a Bruker AV 700 Avance III HD NMR (cryo probe) (Bruker, Karlsruhe, Germany). All solvents used had an isotopic purity of 99.8 atom % D (Zirpel, Degenhardt et al., 2017).

2.2.3. Expression of *nphB* and *thcas* in *Saccharomyces cerevisiae*

2.2.3.1. <u>Strains used in this study</u>

E. coli DH5α was used as plasmid amplification host. Cells were cultured at 37 °C and 200 rpm in LB medium (10 g L^{-1} peptone, 5 g L^{-1} yeast extract, 10 g L^{-1} sodium chloride, pH 7.0 (20 g L^{-1} agar for solid medium)) containing the appropriate antibiotic if necessary.

Protein expression experiments were performed using *S. cerevisiae Δpep4 Δgal1 Δgal80*. Yeast strains without plasmids were cultivated in YPD medium (20 g L^{-1} peptone, 10 g L^{-1} yeast extract, 20 g L^{-1} glucose (20 g L^{-1} agar for solid medium)) at 30 °C and 200 rpm. Yeast strains carrying a plasmid were cultivated in synthetic mineral salt medium without leucine (6.7 g L^{-1} YNB without amino acids, 1.6 g L^{-1} drop out supplements without leucine, 20 g L^{-1} glucose (20 g L^{-1} agar)). The composition of the drop out supplements is given in the supplementary information (Table S 1).

Table 7 and Table 8 contain all relevant genotypes and applications of the strains used in this study. The genotypes of the utilized yeast strains are given in the supplementary information (Table S 18).

Table 7: List of *S. cerevisiae* strains used for the expression of *nphB*.

Strain	Relevant genotype/transformed strain	Transformed plasmid(s)	Reference
SC2	*S. cerevisiae* CEN.PK2-1C *Δpep4Δgal1Δgal80Δmig1*	-	M. Sc. Julia Schachtsiek (TU Dortmund)
SC2_eN	*SC2*	pDionysos_eN	This study
SC2_TeN	*SC2*	pDionysos_TeN	This study
SC2_eNx	*SC2*	pDionysos_eNx	This study
SC2_TeNx	*SC2*	pDionysos_TeNx	This study
SC2_N*	*SC2*	pDionysos_NphB*	This study
SC2_ N(arg)*	*SC2*	pDionysos_ N(arg)*	This study
SC2_ N	*SC2*	pDionysos_ NphB	This study
SC2_ N(arg)	*SC2*	pDionysos_NphB(arg)	This study

Table 8: List of *S. cerevisiae* strains used for the expression of *nphB* and *thcas* simulanteously. (Zirpel, Degenhardt et al., 2017)

Strain	Relevant genotype/transformed strain	Transformed plasmid(s)	Reference
SC	*Saccharomyces cerevisiae* CEN.PK2-1C *Δpep4Δgal1Δgal80*	-	This study
SC_EV	SC	pDionysos	This study
SC_N	SC	pDionysos_NphB	This study
SC_T	SC	pDionysos_THCAS	This study
SC_TT2AN	SC	pDionysos TT2AN	This study
SC_T2	SC	pDio2_THCAS	This study
SC_NT	SC	pDio2_THCAS-NphB	This study

2.2.3.2. <u>Plasmid construction and transformation</u>

Plasmids used and generated in this study are shown in Table 10. Protein sequences (Table S 17), primer sequences (Table S 14, Table S 19) and plasmid maps (Figure S 2, Figure S 12) are given in the supplementary information.

A synthetic gene sequence of *nphB* (GenBank[TM] accession number: AB187169) was optimized for expression in *S. cerevisiae* and ordered from GeneArt (Life Technologies, Regensburg, Germany).

A *thcas* (GenBank[TM] accession number: AB057805) coding sequence without N-terminal native signal peptide (84 bp) was codon usage optimized for *S. cerevisiae* and was obtained from GeneArt. The synthetic gene sequence contains an N-terminal signal peptide from *S. cerevisiae* vacuolar proteinase A (GenBank[TM] accession number: CAA78020) (Zirpel et al., 2015).

Upstream from the coding sequences a 5'-untranslated region (UTR) consensus sequence from S. cerevisiae was inserted via PCR primers in all constructs. The different plasmids (Table 9, Table 10) were cloned by restriction / ligation, Gibson Assembly or site-directed mutagenesis.

<u>Restriction / ligation</u>

The desired restriction recognition sites at the C- and N-terminus of the NphB DNA insert were obtained by PCR using Q5[®] High-Fidelity 2X Master Mix (NEB, Frankfurt am Main, Germany) and primers containing the appropriate recognition site sequences (Table S 14). The PCR products and the digested plasmids were purified by agarose gel extraction. DNA ligation was performed using a Quick Ligation[TM] Kit (NEB, Frankfurt am Main, Germany). *E. coli* DH5α transformation was executed

using a standard protocol and heat shock transformation (Inoue et al., 1990). Plasmids were isolated for verification via PCR (Red-Taq DNA Polymerase 1,1X MasterMix; VWR, Darmstadt, Germany) and sequencing (Seqlab, Göttingen, Germany).

<u>Gibson Assembly</u>

The DNA insert and vector fragments required for Gibson cloning (Gibson et al., 2009) were obtained by PCR using Q5$^®$ High-Fidelity 2X Master Mix and primers with 25 bp overlaps (Table S 14, Table S 19). Vector fragments were digested with *DpnI* and purified. DNA insert fragments were purified by agarose gel extraction. Vector and insert fragments were used in an equimolar ratio and added to 2X Gibson Assembly Master Mix (NEB) and Gibson Assembly was performed according to the manufacturer. *E. coli* DH5α transformation was performed using a standard protocol and heat shock transformation (Inoue et al., 1990). Plasmids were isolated for verification via PCR and sequencing.

Coding sequences of nphB and thcas of pDionysos_TT2AN (Table 10) were separated by a T2A sequence (Table S 17) published by Beekwilder et al. (2014). The T2A sequence was integrated into the vector with the PCR primers listed in Table S 19 using Gibson Assembly.

The bidirectional Gal10/Gal1 promoter system was obtained from the plasmid pESC-URA (Table 10). The pGal10-TADH1 cassette was amplified by PCR and inserted into pDionysos. The coding sequences of *nphB* (under the control of the GAL1 promoter) and *thcas* (under control of pGAL10) were integrated into pDio2 (Table 10, Figure S 12) by Gibson Assembly.

<u>Site-directed mutagenesis</u>

The mutated GAL1 promoter with deleted mig1 binding site (Figure S 11) was constructed by mutagenesis PCR. The primer sequences were designed according to Liu and Naismith (2008) and are given in the supplementary information (Table S 15). For mutagenesis Q5$^®$ High-Fidelity 2X Master Mix was applied. The PCR products (plasmids) were digested with *DpnI* and transformed in *E. coli* DH5α using a standard protocol and heat shock transformation (Inoue et al., 1990). Plasmids were isolated and sent for sequencing.

Plasmids with the correct sequence were transformed into frozen competent *S. cerevisiae* cells using LiAc / ss carrier DNA / PEG method (Gietz and Schiestl, 2007) and plated on synthetic mineral salt medium without uracil agar plates (6.7 g L^{-1} YNB without amino acids, 1.9 g L^{-1} drop out supplements without uracil, 20 g L^{-1} glucose, 20 g L^{-1} agar) and incubated at 30 °C for 2 - 3 days. A few colonies from that plate were resuspended in dH$_2$O and plated on synthetic mineral salt medium without leucine (6.7 g L^{-1} YNB without amino acids, 1.6 g L^{-1} drop out supplements without leucine, 20 g L^{-1} glucose, 20 g L^{-1} agar) and incubated at 30 °C. The composition of the drop out supplements is given in the supplementary information (Table S 1).

Table 9: List of plasmids used for the expression of *nphB* in *S. cerevisiae*.

Plasmid	Resistance/auxotrophy marker	Expressed coding sequence	Reference
pUG6	Ampicillin, G418	loxP-P_{TEF1}-kanMX-P_{TEF1}-loxP	EUROSCARF (P30114); (Güldener et al., 1996)
pSH47	Ampicillin, uracil	P_{GAL1}: Cre	EUROSCARF (P30119); (Güldener et al., 1996)
pDionysos_eN	Ampicillin, uracil, leucine	P_{GAL1}: eNphB	This study
pDionysos_TeN	Ampicillin, uracil, leucine	P_{GAL1}: TeNphB	This study
pDionysos_eNx	Ampicillin, uracil, leucine	P_{GAL1}: eNphBx	This study
pDionysos_TeNx	Ampicillin, uracil, leucine	P_{GAL1}: TeNphBx	This study
pDionysos_N*	Ampicillin, uracil, leucine	P_{GAL1}: NphB	This study
pDionysos_N(arg)*	Ampicillin, uracil, leucine	P_{GAL1}: NphB(arg)	This study
pDionysos_N	Ampicillin, uracil, leucine	P_{GAL1} (mut): NphB	This study (Zirpel, Degenhardt et al., 2017)
pDionysos_N(arg)	Ampicillin, uracil, leucine	P_{GAL1} (mut): NphB (arg)	This study

Table 10: List of plasmids used for the expression of *nphB* and *thcas* simultaneously. (Zirpel, Degenhardt et al., 2017)

Plasmid	Resistance/auxotrophy marker	Expressed coding sequence	Reference
pDionysos	Ampicillin, uracil, leucine	-	(Stehle et al., 2008)
pESC-URA	Ampicillin, uracil	-	Agilent Technologies, Waldbronn, Germany
pDio2	Ampicillin, uracil, leucine	-	This study
pDionysos_THCAS	Ampicillin, uracil, leucine	P_{GAL1}: THCAS	(Zirpel et al., 2015)
pDionysos_TT2AN	Ampicillin, uracil, leucine	P_{GAL1}: THCAS T2A NphB	This study
pDionysos_NphB	Ampicillin, uracil, leucine	P_{GAL1}: NphB	This study
pDio2_THCAS	Ampicillin, uracil, leucine	P_{GAL10}: THCAS	This study
pDio2_THCAS-NphB	Ampicillin, uracil, leucine	P_{GAL10}: THCAS P_{GAL1}: NphB	This study

2.2.3.3. <u>Strain development of *S. cerevisiae Δpep4 Δgal1 Δgal80*</u>

The disruption of the transcriptional regulator GAL80 (YML051W) on the genome of *S. cerevisiae Δpep4 Δgal1* was performed using a gene disruption cassette described by Güldener et al. (1996). The used primers are shown in Table S 19.

The loxP-P_{TEF-1}-kanMX-P_{TEF-1}-loxP cassette was amplified by PCR (Q5® High-Fidelity 2X Master Mix) using pUG6 as DNA template and specific primers (Table S 16). The PCR product was purified by agarose gel extraction and 2 µg of the purified PCR product were transformed into *S. cerevisiae Δpep4 Δgal1* using the LiAc / ss carrier DNA / PEG method (Gietz and Schiestl, 2007). Cells carrying the *gal80* disruption were selected based on their resistance to G418 (Geneticin®). Correct integration was validated by PCR (Red Taq DNA Polymerase 2X Master Mix) using isolated yeast genomic DNA as a template (MasterPure™ Yeast DNA Purification Kit; epicentre, Lucigen Corporation, Middleton, USA) and sequencing with primers C-PCR_Gal80_fw and C-PCR_Gal80_rv. The strain was transformed with pSH47 using LiAc / ss carrier DNA / PEG method (Gietz and Schiestl, 2007). Cre recombinase expression was induced in an overnight culture with 2% galactose for 2 h. Subsequently, cells were selected on agar plates with and without G418 because efficient *Cre* expression leads to sensitivity to G418 again. The loss of the kanMX cassette was verified by PCR and sequencing using primers C-PCR_Gal80_fw and C-PCR_Gal80_rv. Cells were streaked on YNB agar plates containing all amino acids and 3 g L^{-1} 5-fluoroorotic acid to select for those strains that have lost the pSH47 plasmid (Boeke et al., 1984). The newly generated yeast strain carrying the *gal80* gene disruption

(*S. cerevisiae* **pep4::loxP gal1::loxP gal80::loxP**) was transformed with plasmids using LiAc / ss carrier DNA / PEG method (Gietz and Schiestl, 2007).

2.2.3.4. <u>Membrane permeability of olivetolic acid and geranyl diphosphate</u>

The cultivation of SC_N (Table 8) was performed with two precultures followed by one main culture. For the first preculture synthetic mineral salt medium without leucine (6.7 g L^{-1} YNB without amino acids, 1.6 g L^{-1} drop out supplements without leucine, 20 g L^{-1} fructose) was inoculated and the cells were incubated for 24 h at 30 °C and 200 rpm. The second preculture was inoculated using the first preculture and incubated at 30 °C and 200 rpm for 12 h. For the main culture 50 mL complex medium (20 g L^{-1} yeast extract, 40 g L^{-1} peptone, 80 mg L^{-1} adenine hemisulfate, 40 g L^{-1} fructose, 5 g L^{-1} galactose, 100 mM potassium citrate buffer pH 5.5) in 500 mL baffled flasks were inoculated with OD_{600} of 0.5. Cultures were incubated at 15 °C and 200 rpm for 72 h.

Cell culture volumes correlating with an OD_{600} of 500 were harvested (2,000 x g, 15 min, 4 °C) and supernatants were discarded. Cells were resuspended in 2 mL assay buffer (50 mM Tris-HCl buffer pH 7.5, 10% (w/v) glycerol, 100 mM sodium chloride). The cell suspension was divided into two 1 mL aliquots. One aliquot was supplemented with 3 mM GPP and 3 mM OA. After 4 h incubation at 1400 rpm and 37 °C cells were harvested (2,000 x g, 10 min, 4 °C) and used for ethyl acetate extraction. The extracts were dried under vacuum, resuspended in AcN, filtrated (0.45 µm, nylon) and analyzed by LC-MS (chapter 2.2.3.6). The second aliquot was used for cell lysis. 500 µL cell suspension were transferred to 0.5 mL tubes which were subsequently filled up with glass beads (diameter: 0.4 – 0.6 mm). Cells were lysed by vortexing at maximum speed for 30 min at 4 °C. Cell lysates were centrifuged (16.100 x g, 5 min, 4 °C) and supernatants were used for NphB activity assays (5 mM magnesium chloride, 1 mM GPP, 5 mM OA, 37 °C, 1100 rpm, 4 h) (Zirpel, Degenhardt et al., 2017).

2.2.3.5. <u>Expression of *nphB* and *nphB/thcas* in *S. cerevisiae*</u>

The expression of yeast cells transformed with the desired plasmid was performed with two precultures followed by one main culture. For the first preculture synthetic mineral salt medium without leucine (6.7 g L^{-1} YNB without amino acids, 1.6 g L^{-1} drop out supplements without leucine, 20 g L^{-1} fructose) was inoculated and the cells were incubated for 24 h at 30 °C and 200 rpm. The second preculture was inoculated using the first preculture and incubated at 30 °C and 200 rpm for 12 h. 100 mL complex medium (20 g L^{-1} yeast extract, 40 g L^{-1} peptone, 80 mg L^{-1} adenine hemisulfate, 40 g L^{-1} fructose, 5 g L^{-1} galactose, 100 mM potassium citrate buffer pH 5.5) in 1 L baffled flasks were inoculated with OD_{600} of 0.5 for main culture. Cultures were incubated at 15 °C and 200 rpm. Samples were taken every 24 h for measurement of OD_{600} and activity assays.

Cells culture volumes correlating with an OD_{600} of 125 were harvested (2,000 x g, 4 °C, 10 min) and resuspended in 500 µL assay buffer (50 mM Tris-HCl buffer pH 7.5, 10% (w/v) glycerol, 100 mM sodium chloride). Cell suspension was transferred in 0.5 mL tubes and filled with glass beads

(diameter: 0.4 - 0.6 mm). Cells were lysed by vortexing at maximum speed and 4 °C for 30 min. The cell lysate was centrifuged (16,100 x g, 4 °C, 10 min) and the resulting supernatant was used for the determination of NphB activity (5 mM magnesium chloride, 1 mM geranyl diphosphate, 1 mM olivetolic acid, 4 h, 37 °C, 1100 rpm) (Zirpel, Degenhardt et al., 2017).

2.2.3.6. <u>LC-DAD, LC-ESI-MS</u>

Activity assays were stopped by addition of 0.1 assay-volumes formic acid (FA) and 2.9 assay-volumes acetonitrile (AcN) followed by incubation on ice for 15 min. Samples were centrifuged (16,100 x g, 4 °C, 20 min), supernatants were filtered (0.45 m, nylon) and analyzed by LC-DAD (Agilent 1260 Infinity HPLC, Waldbronn, Germany). The separation of the compounds was performed on a EC 100/2 Nucleoshell RP18 2.7 m column (Macherey Nagel, Düren, Germany) or a Poroshell 120 EC-C18 2.1 x 100 mm 2.7 μm column (Agilent Technologies, Waldbronn, Germany) (0.8 mL min^{-1}, 40 °C, 35% (v/v) H_2O with 0.1% FA, 65% (v/v) AcN). Quantification of the products was based on integrated peak areas of the UV chromatograms at 225 nm. Standard curves were generated for CBGA and THCA. The CBGA standard curve was used to estimate the concentration of CBGA and 2-*O*-GOA The identity of each product was confirmed by comparing mass spectra of each sample with coeluting standards analyzed by Bruker compact^{TM} ESI-Q-TOF (Bruker, Bremen, Germany) using positive ionization mode (Zirpel, Degenhardt et al., 2017).

Chapter 3

Results and discussion

Parts of chapter 3.3.2 and chapter 3.3.3 were published in:

Zirpel, B.[1], **Degenhardt, F.**[1], Martin, C., Kayser, O., & Stehle, F. (2017). Engineering yeasts as platform organisms for cannabinoid biosynthesis. Journal of Biotechnology, 259, 204–212.

[1] Authors contributed equally.

The "Results and discussion" chapter is divided into three sub-chapters:

I. Expression of membrane-bound prenyltransferases in *S. cerevisiae* (chapter 3.1)

II. Expression of *nphB* in *E. coli* (chapter 3.2)

III. Expression of *nphB* and *thcas* in *S. cerevisiae* (chapter 3.3)

3.1. Expression of membrane-bound prenyltransferases in *S. cerevisiae*

The coding sequence of the aromatic prenyltransferase CsPT1 has been used for expression experiments in yeasts and CBGAS activity tests in our department. So far, we have not been able to detect any CBGA production. On this account, potential cannabigerolic acid synthases CsPT2 and CsPT3 were identified by *in silico* analysis (Pamplaniyil, 2017). These enzymes could serve as alternatives for CsPT1. The amino acid sequences of CsPT1, CsPT2 and CsPT3 are given in the supplementary information (Table S 3) and an alignment of the three sequences is displayed in Figure 15.

Figure 15: Amino acid sequence alignment of CsPT1, CsPT2 and CsPT3. The alignment was generated using the software Clone Manager 9 Professional Edition (Sci-Ed Software, Denver, USA). The sequences of CsPT1, CsPT2 and CsPT3 were obtained from Pamplaniyil (2017). Matching amino acids within the alignment are coloured in red.

The amino acid sequences of CsPT1 and CsPT2 share 72% of their sequences, while the sequence alignment of CsPT1 and CsPT3 resulted in a sequence identity of 61%. The same value was obtained for the alignment of the amino sequences of CsPT2 and CsPT3. The three sequences are all of approximately the same length (CsPT1: 394 amino acids (AA), CsPT2: 397 AA, CsPT3: 398 AA).

3.1.1. Comparison of CsPT1, CsPT2, CsPT3 and CsPT1P

In 2014, a patent protecting the coding sequence of CBGAS isolated from *C. sativa* was granted (Page and Boubakir, 2014). The protein sequence of this aromatic prenyltransferase, from now on called CsPT1P, is very similar to the one of CsPT1 used in our department. An alignment of the protein sequences CsPT1P, CsPT1, CsPT2 and CsPT3 is shown in Figure S 3. The sequence alignment of CsPT1 and CsPT1P is depicted in Figure 16.
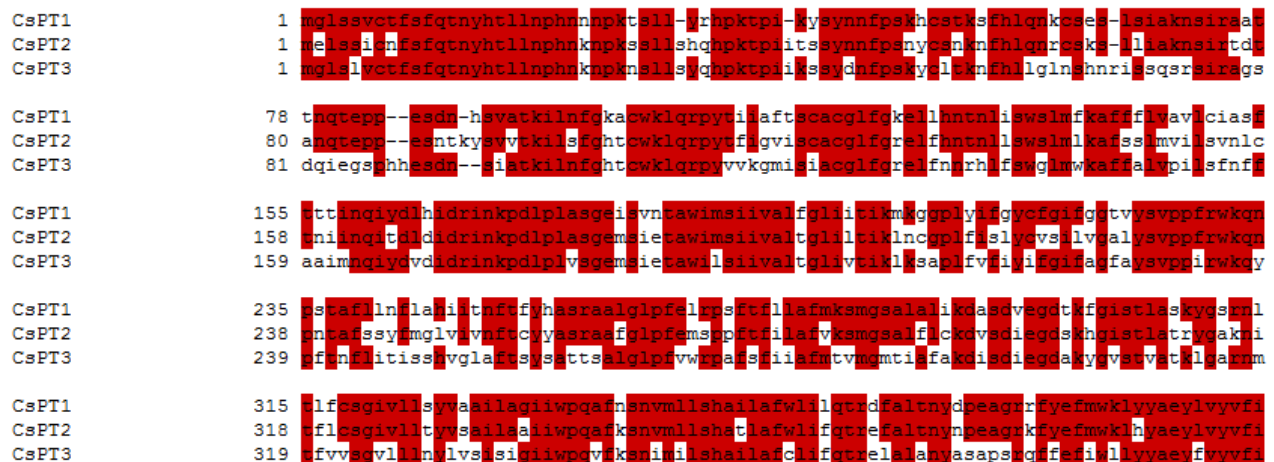

Figure 16: Amino acid sequence alignment of CsPT1P and CsPT1. The alignment was generated using the software Clone Manager 9 Professional Edition (Sci-Ed Software, Denver, USA). The sequence of CsPT1 was obtained from Pamplaniyil (2017). The sequence of CsPT1P was published by Page and Boubakir (2014). Matching amino acids within the alignment are coloured in red.

The alignment of CsPT1P and CsPT1 shows a sequence percentage identity of 98 % (391/395 identical), with a difference of three AA (I149V, I223T, Y256H). Additionally, the sequence of CsPT1P is one amino acid longer than the sequence of CsPT1 (395 AA vs. 394 AA). The additional amino acid, a cysteine, is located at position 32 of CsPT1P. The differences between the amino acids sequences of CsPT1P and CsPT1 can be explained by the fact that even though both coding sequences were isolated from *C. sativa*, different plant varieties were used. For the following expression experiments the coding sequence of CsPT1 was used.

The amino acid sequences of CsPT1P and CsPT2 share a sequence percentage identity of 73%, while the sequence alignment of CsPT1P and CsPT3 resulted in an identity of 61% (Figure S 3).

3.1.2. Sequence analysis

Prior to the expression experiments, the properties of CsPT1, CsPT2 and CsPT3 were investigated using web-based tools like ChloroP, TargetP, and TMHMM. The subcellular localization of the three enzymes is important for understanding their function and for designing an appropriate cloning and expression strategy.

3.1.2.1. <u>Prediction of transmembrane helices</u>

CBGAS belongs to the group of aromatic prenyltransferases since it catalyzes the transfer of a geranyl moiety (GPP) to an aromatic acceptor molecule (OA) (chapter 1.3.1.1). The group of aromatic

prenyltransferases can be subdivided into membrane-bound and soluble enzymes (cf. chapter 1.3). It is important to know which group of aromatic prenyltransferases CsPT1, CsPT2 and CsPT3 belong to as their expression protocols differ and membrane-bound aromatic prenyltransferases require divalent cations for their catalytic activity (cf. chapter 1.3.1).

Therefore, the protein sequences of CsPT1, CsPT2 and CsPT3 were analyzed using the web-tool TMHMM Server v. 2.0 (http://www.cbs.dtu.dk/services/TMHMM/; 03.11.2017). This tool predicts transmembrane helices (TMHs) within protein sequences and is able to discriminate between membrane and soluble proteins with high accuracy. The method is based on a Hidden Markov Model (HMM) considering alternation of cytoplasmic and non-cytoplasmic loops, charge bias, hydrophobicity and helix length (Krogh et al., 2001; Sonnhammer et al., 1998). The TMHMM analysis results of CsPT1, CsPT2 and CsPT3 are shown in Figure 17, Figure S 4 and Figure S 5 (supplementary information).
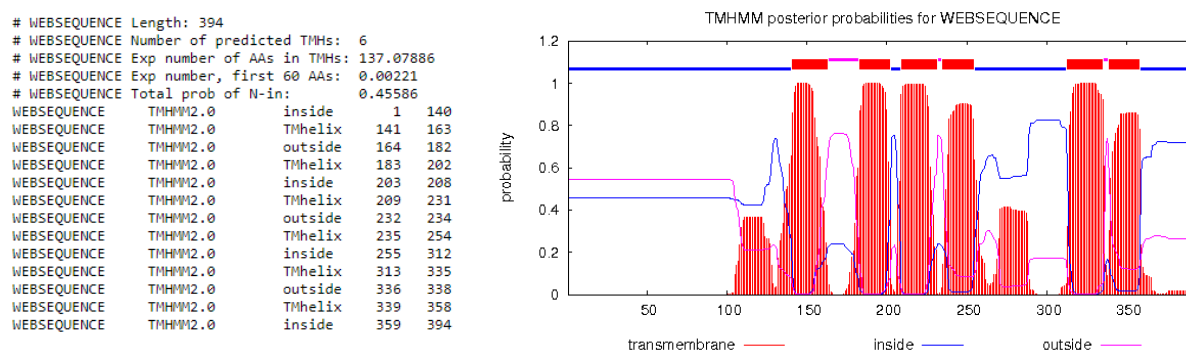
Figure 17: TMHMM analysis of CsPT1. The results show the length of the submitted sequence and the number of predicted transmembrane helices (TMHs). Additionally, the number of amino acids expected to be part of the TMHs and the expected number of amino acids in TMHs in the first 60 amino acids is given. The latter value is important since it gives a hint whether the predicted TMH could be a signal peptide. The total probability of N-in gives information about the probability of the N-terminus being on the cytoplasmic side of the membrane.

The TMHMM web-tool predicted six TMHs within the amino acid sequence of CsPT1 (Figure 17). According to Krogh et al. (2001) the shortest TMHs are built up of approximately 18 amino acids. The six predicted TMHs of CsPT1 contain 20 - 23 residues. Additionally, the enzyme is most likely a helical transmembrane protein since the expected number of TMHs is larger than one and the expected number of amino acids in the TMHs is high (139.4 AA). Since no TMHs were predicted in the first 60 amino acids of CsPT1, no signal peptide was misleadingly classified as TMH. The total probability of the N-terminus of CsPT1 being on the cytoplasmic side of the membrane was calculated at 0.456. The TMHMM analysis of CsPT1 showed two possible weak TMHs, which were not taken into account for the calculation of total number of predicted TMHs (residues 100 to 125 and 275 to 300, respectively) as the calculated probability was not high enough to classify them as TMH. When comparing the plot and the prediction at the top of the plot (Figure 17) it has to be considered that the prediction shows the overall most probable structure while the plot represents probabilities for each amino acid. Although the the total probability that the N-terminus of the protein is localized to the cytoplasmic side

of the membrane is not significant (> 0.5) for CsPT1 (0.456), the assumption that both the N- and C-terminus of CsPT1 are on the cytoplasmic side of the membrane might be valid because Krogh et al. (2001) observed that this orientation dominates in almost all organism.

TMHMM analysis results of CsPT2 and CsPT3 are summarized in Table 11 and the corresponding outputs are given in the supplementary information (Figure S 4, Figure S 5).

Table 11: Results of the TMHMM analysis obtained for CsPT2 and CsPT3. The analysis was performed using the online-tool TMHMM Server v. 2.0. The results show the number of predicted transmembrane helices (TMHs), the number of amino acids expected to be part of the TMHs and the expected number of amino acids in TMHs within the first 60 amino acids. The total probability of N-in gives information on the probability of the N-terminus being on the cytoplasmic side of the membrane. The corresponding images are depicted in the supplementary informations (Figure S 4:).

	CsPT2	CsPT3
Predicted TMH	4	8
TMH in the first 60 AA	0.00055	0.007
Total probability of N-in	0.41891	0.64989
Weak TMH	4	1
Localization of the C-terminus	outside	inside

TMHMM analysis of CsPT2 resulted in the prediction of four TMHs and four possible weak TMHs that were not taken into account for the calculation of the total number of TMHs (Table 11; Figure S 4). No signal peptide was misleadingly identified as a TMH. The value of the total probability of the N-terminus being on the cytoplasmic side of the membrane is not significant (0.420). The web-tool predicts the N- and the C-terminus of the enzyme to be outside (Figure S 4), i.e. they are not localized to the cytoplasm of the cell, but to the inside of an organell (see below).

The aromatic prenyltransferase CsPT3 contains eight predicted TMHs and one possible weak THM that was not included into the calculation of the total number of TMHs (Table 11; Figure S 5). No N-terminal targeting sequence was misleadingly identified as a TMH and the C-terminus of the enzyme is most probably localized on the inside of the cell (0.650). On this account, the N- and the C-terminus of the enzyme are most probable localized on the inside of the cell.

The used terms "inside" (on the cytoplasmic side) and "outside" (non-cytoplasmic side) should be viewed with caution since the "positive inside rule" (von Heijne, 1992, 1986) was postulated for the prokaryote *E. coli*, which does not have membrane-enclosed subcellular compartments like eukaryotes. Thus, predictions of the N-terminus being on the cyctoplasmic side make sense for prokaryotic membrane-bound proteins, but not for eukaryotic ones. In the case of eukaryotic membrane-bound enzymes the term "inside" means within the cytoplasm and the term "outside" describes the enzyme being on the inside of an organelle.

With this in mind the C-termini of CsPT1 and CsPT3 are localized in the cytoplasm and the C-terminus of CsPT2 is most probably located inside of an organelle. Nevertheless, predictions of the localization of specific protein segments have to be handeld with care because without experimental data it is difficult to say whether a potential helical structure completely crosses a membrane.

As mentioned in chapter 1.3.1 membrane-bound aromatic prenyltransferases contain two aspartate-rich motifs (motif I and motif II). Motif I is responsible for prenyl diphosphate binding and motif II might be responsible for binding of the aromatic acceptor molecule (Heide, 2009; Li et al., 2015; Sasaki et al., 2008; Stec and Li, 2012). Sequences analysis of CsPT1, CsPT2 and CsPT3 showed two asparate-rich motifs with the consensus sequences NQxxDxxxD (motif I) and KDxxDxEGD (motif II), respectively, in which 'x' could be any amino acid (cf. Table S 3).

Based on the presence of two aspartate-rich motifs and the results of the TMHMM analyzes CsPT1, CsPT2 and CsPT3 belong to the group of membrane-bound enzymes (Figure 14) with different numbers of helical structures. With this in mind, an expression strategy for membrane proteins had to be developed which is presented in chapter 3.1.3. Additionally, it has to be kept in mind that the activity assays have to be perfomed in the presence of divalent cations.

3.1.2.2. Prediction of N-terminal targeting sequences

As mentioned in chapter 1.3.1 membrane-bound aromatic prenyltransferase with plant origin can be divided into two groups: (i) homogentisate prenyltransferases that are located in the plastids and (ii) PHB prenyltransferases that are localized in the inner membrane of mitochondria (Yazaki et al., 2009). CBGAS is involved in the biosynthesis of cannabinoids that belong to the group of secondary metabolites in *C. sativa* (chapter 1.2). In higher plants the prenyl moiety (GPP) of CBGA is derived from the MEP pathway that is localized in the plastids (chapter 1.2.2), which would support that CBGAS is a plastidic enzyme. Thus, the amino acid sequence might contain a N-terminal signal sequence for targeting the enzyme to the plastids. Additionally, knowledge about the possible existence of a N-terminal targeting sequence is important because on the one hand this sequence might be used for targeting within the expression host organism or on the other hand it could be obstructive. Yeast cells do not contain plastids which are characteristic to algae and plant cells. Loddenkötter et al. (1993) showed that the chloroplast triose phosphate-phosphate translocator isolated from spinach was localized to the rough endoplasmatic reticulum and/or the mitochondria if expressing the enzyme without transit peptide in *S. cerevisiae*.
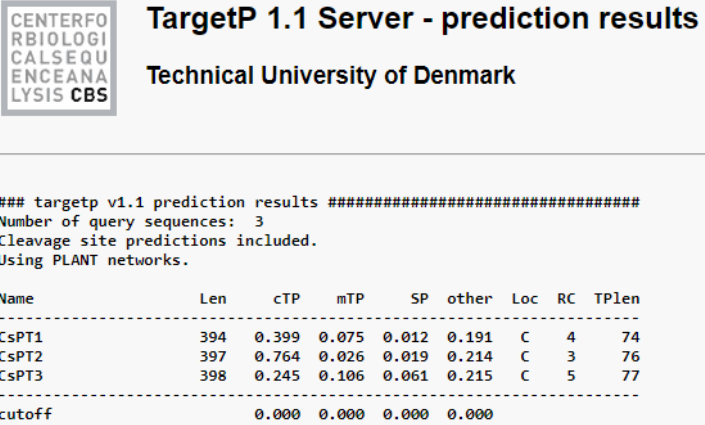
Targeting sequences are short peptides at the N-terminus of a newly synthesized protein that target the completed or nascent protein from the cytosol to cellular membranes, for secretion outside of the cell or to certain organelles inside of the cell (e.g. endoplasmic reticulum (ER), mitochondria, and chloroplasts). Subsequently, they are cleaved off, e.g. by peptidases. The major N-terminal targeting sequences (N-terminal presequence) can be devided into three groups: (i) secretory pathway signal peptides (SP) that target proteins to the secretory pathway through the ER, (ii) chloroplast transit peptide (cTP) that are responsible for targeting proteins to the chloroplasts and (iii) mitochondrial

targeting peptides (mTP) that target proteins to the mitochondria (Emanuelsson et al., 2000; Kapp et al., 2009; Martoglio and Dobberstein, 1998; Small et al., 2004).

In order to elucidate whether CsPT1, CsPT2 and CsPT3 indeed contain an N-terminal signal peptide, their amino acid sequences were analyzed using the online-tools ChloroP, TargetP and Predotar.

The web-tool TargetP 1.1 (http://www.cbs.dtu.dk/services/TargetP/; 03.11.2017) predicts whether an eukaryotic protein is targeted to the secretory pathway (SP), to the mitochondria (mTP) or to the chloroplasts (cTP) (Emanuelsson et al., 2007, 2000; Nielsen et al., 1997). Additionally, cleavage sites of proteins containing potential N-terminal presequences are predicted. The result for CsPT1, CsPT2 and CsPT3 are shown in Figure 18.

The results of TargetP show that all three proteins analyzed contain a N-terminal targeting sequence with a length of 74 to 77 AA and are targeted to the chloroplasts. Each predicted protein localization is assigned to a reliability class (RC) value that ranges from 1 to 5 and specifies how reliable the respective prediction is. The lower the RC value, the stronger the obtained prediction (Emanuelsson et al., 2007). The analysis of CsPT1, CsPT2 and CsPT3 resulted in RC values of 3 to 5, thus the predictions of a N-terminal cTP should be considered with care.


```
### targetp v1.1 prediction results ################################
Number of query sequences:  3
Cleavage site predictions included.
Using PLANT networks.

Name          Len    cTP    mTP    SP   other  Loc  RC  TPlen
----------------------------------------------------------------
CsPT1         394   0.399  0.075  0.012  0.191   C   4    74
CsPT2         397   0.764  0.026  0.019  0.214   C   3    76
CsPT3         398   0.245  0.106  0.061  0.215   C   5    77
----------------------------------------------------------------
cutoff               0.000  0.000  0.000  0.000
```

Figure 18: TargetP results for CsPT1, CsPT2 and CsPT3. The length of the analyzed sequence (Len) is given as well as the probability that the protein is targeted to the chloroplasts (cTP), the mitochondria (mTP), the secretory pathway (SP) or any other location (other). The most probable localization is given in the column Loc (C – chloroplast, M – mitochondrion, S – secretory pathway, _ – any other location). The reliability class (RC) indicates how strong the prediction is and the lower the RC value the safer the prediction of protein localization. In the column "TPlen" the length of the predicted N-terminal presequence is given.

If TargetP predictes a cTP, Emanuelsson et al. (2007) recommend the use of the online-tool ChloroP in order to check how trustworthy the results are regarding localization and length of the targeting sequence obtained by TargetP. To validate our findings, we reanalyzed the amino acid sequences of CsPT1, CsPT2 and CsPT3 using the online-tool ChloroP to predict cTPs and possible cleavage sites (http://www.cbs.dtu.dk/services/ChloroP/; 03.11.2017) (Emanuelsson et al., 1999). The results are displayed in Figure 19.
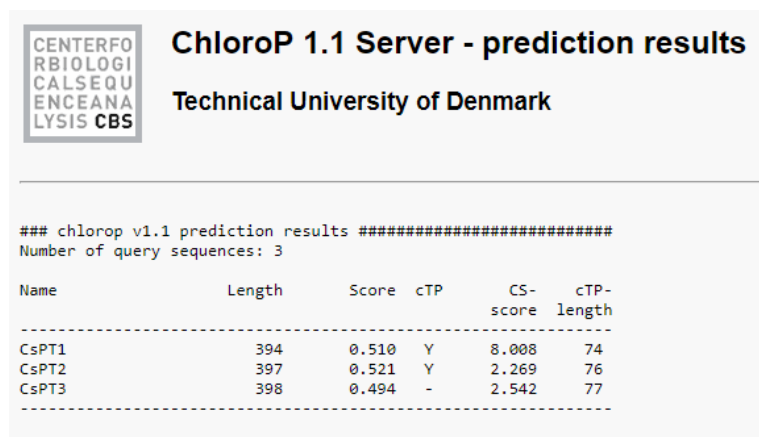
Figure 19: ChloroP results for CsPT1, CsPT2 and CsPT3. The amino acid sequences were analyzed using the online-tool ChloroP 1.1. The results display the length of the analyzed amino acid sequence and whether the uploaded sequence contains a cTP (Y) or not (-). This prediction is based on the score which usually accept values between 0.4 and 0.6. The CS (cleavage site)-score is used for the prediction of the cleavage site and thereby for determining the length of the predicted cTP.

Based on the results obtained by ChloroP (Figure 19) the amino acid sequences of CsPT1 and CsPT2 contain a N-terminal cTP with a length of 74 AA or 76 AA, respectively. According to the web-tool CsPT3 does not contain a N-terminal cTP, nevertheless the length of a possible targeting sequence was calculated. Here it should be noted that the length of a cTP is calculated by ChloroP even though the absence of cTP was predicted (Emanuelsson et al., 1999). According to Emanuelsson et al. (2007) the presence of a N-terminal cTP is predicted (Y) if the score value is above 0.5 (cut-off). Additionally, predictions are classified as reliable (quite strong) when the score value is above 0.55. This means that the predicitions of possible cTP's for CsPT1 and CsPT2 shoud be considered with care as the score values (Figure 19) are smaller than 0.55. Additionally, the score value of CsPT3 is close to 0.5 (0.494) and thus the probability of a present cTP should not be excluded completely.

Since the results obtained by TargetP and ChloroP were not classified as very reliable we decided to use a third web-tool for the identification of a possible N-terminal targeting sequence. The online-tool Predotar v1.04 (https://urgi.versailles.inra.fr/predotar/; 10.11.2017) is able to predict putative mitochondrial, plastid or ER targeting sequences (Small et al., 2004). The results obtained for the analysis of CsPT1, CsPT2 and CsPT3 are displayed in Figure 20.

Sequence	Mitochondrial	Plastid	ER	Elsewhere	Prediction
CsPT1	0,01	0,62	0,00	0,38	plastid
CsPT2	0,01	0,34	0,00	0,65	possibly plastid
CsPT3	0,02	0,85	0,02	0,14	plastid

Figure 20: Predotar results for CsPT1, CsPT2 and CsPT3. The amino acid sequences were analyzed using the online-tool Predotar v1.04. The results display for each enzyme a probability as to presence of a putative mitochondrial, plastid or ER targeting sequence. Additionally, the probability of no targeting sequence and the predicted localization of the enzyme are given.

According to the Predotar results, all three sequences contain a N-terminal plastid targeting sequence (Figure 20). The values of the estimated probabilities can range between 0 and 1. The higher this

value the more reliable the putative N-terminal targeting sequence (Small et al., 2004). It has to be mentioned that according to Small et al. (2004) probability values of 0.2 may indicate the presence of a targeting sequence because the value is greater than the "background". On this account the result for CsPT2 should be considered with care.

Table 12: Summary - prediction of N-terminal targeting sequences. The predicted length of the putative targeting sequences is given in brackets. cTP – chloroplast transit peptide.

	TargetP[1]	ChloroP[2]	Predotar[3]
CsPT1	chloroplasts (74 bp)	cTP (74 bp)	plastid
CsPT2	chloroplasts (76 bp)	cTP (76 bp)	possibly plastid
CsPT3	chloroplasts (77 bp)	-	plastid

The results obtained by TargetP, ChloroP and Predotar are summarized in Table 12. Sequence analysis of CsPT1, CsPT2 and CsPT3 indicated putative N-terminal chloroplast transit peptides. The predicted length of the cTP of CsPT1 points out that the missing amino acid, compared with the amino acid sequence of CsPT1P (cf. chapter 3.1.1), is located in the cTP sequence. On this account, the missing amino acid does probably not affect enzyme activity.

3.1.3. Expression strategy

In the following chapter the strategy for the expression of the membrane-bound proteins CsPT1, CsPT2 and CsPT3 in *S. cerevisiae* is described. First, it is discussed whether the coding sequences should be expressed with or without predicted cTP and if the codon usage should be optimized for expression in *S. cerevisiae* (chapter 3.1.3.1). In chapter 3.1.3.2 the design of the used yeast expression vector is presented.

3.1.3.1. <u>Necessity of cTP for recombinant expression of plant membrane-bound proteins in *S. cerevisiae*</u>

To decide whether CsPT1, CsPT2 and CsPT3 should be produced recombinantly in *S. cerevisiae* using the predicted N-terminal cTP's, four recent case-studies regarding the heterologous expression of membrane-bound, plant prenyltransferases were taken into account.

In 2012, Shen et al. expressed a membrane-bound isoflavonoid-specific prenyltransferase (LaPT1; GenBank[TM] accession number: AER35706) isolated from white lupin (*Lupinus albus*) in *S. cerevisiae* W303-A1. LaPT1 has a predicted cTP with a length of 17 AA (Δ17 LaPT1). According to their results, yeast microsomal extracts containing the full length protein or variant Δ17 LaPT1, respectively, showed similar activity levels and apparent K_M values (Shen et al., 2012).

There are three other publications dealing with the expression of plant, membrane-bound aromatic prenyltransferases, with and without predicted N-terminal transit peptide, in *S. cerevisiae*. In 2008, Akashi et al. isolated the gylcinol 4-dimethylallyltransferase (G4DT; GenBank[TM] accession number: AB434690) from young soybean (*Glycine max*) seedlings that is involved in glyceollin biosynthesis. G4DT has a predicted transit peptide sequence with a length of 44 AA (Δ44 G4DT). According to their results yeast microsomal extracts containing the full length protein did not show detectable activity. In contrast to this, product formation was detectable when using microsomal fractions containing Δ44 G4DT (Akashi et al., 2008). Similar results were obtained by Munakata et al. (2014) who isolated a coumarin-specific membrane-bound aromatic prenyltransferase from lemon (*Citrus limon*). The enzyme ClPT1 (GenBank accession number: AB813876) has a putative N-terminal transit peptide with a length of 36 AA (Δ36 ClPT1). Activity assays using yeast microsomal extracts containing the full length protein or variant Δ36 ClPT1, respectively, resulted in the same enzymatic reaction products, but the activity levels were much higher for Δ36 ClPT1 than for ClPT1 wt (Munakata et al., 2014). In 2015, Li et al. heterologously co-expressed two aromatic prenyltransferase (HlPT1L, HlPT2; GenBank[TM] accession numbers: KM222441 and KM222442) from hop (*Humulus lupulus*) trichomes that are involved in aromatic prenylations of the β-bitter acid pathway. Both enzymes contain predicted N-terminal plastid signal peptides, which were removed in order to test their effect on the prenyltransferase activity. According to their results yeast strains producing the truncated proteins formed much higher amounts of bitter acid than yeast strains with full-length proteins (Li et al., 2015).

The four presented case-studies show that the successful expression of membrane-bound aromatic prenyltransferases in *S. cerevisiae* does not require the usage of a putative native plant signal peptide. Furthermore, removal of the N-terminal transit peptide sequence might even increase the recombinant protein production (Akashi et al., 2008) or the enzyme activity level (Li et al., 2015; Munakata et al., 2014). Therefore, it was decided to perform the expression of CsPT1, CsPT2 and CsPT3 using the native *C. sativa* genes without the N-terminal targeting sequences that were predicted by ChloroP, Predotar and TargetP (chapter 3.1.2.2).

3.1.3.2. Vector design

The expression of the membrane-bound aromatic prenyltransferases in *S. cerevisiae* is performed using a modified pDionysos expression vector. The plasmid contains *URA3* and *LEU2d* genes as selection markers (Stehle et al., 2008). The *LEU2* gene is under control of a truncated version of the LEU2 promoter (LEU2d) with a decreased transcription rate (< 5% compared to the wildtype LEU2 promoter). When LEU2 auxotrophic cells are grown in leucine-deficient medium, they are forced to maintain the plasmid at a high copy number to counteract the reduced transcription rate of the LEU2d promoter (Erhart and Hollenberg, 1983). The increase in copy number can result in a higher expression level of functional membrane proteins. Parker et al. (2014) showed that the introduction of *LEU2d* led to an up to four times higher expression level of several membrane transporters.

A C-terminal yeast-enhanced green fluorescent protein (yEGFP) was used as an expression reporter and was detected by measuring the whole-cell fluorescence. Green fluorescent protein (GFP) is a

suitable marker for the detection of gene expression since is does not need any cofactors specific to the yellyfish *Aequorea victoria* from which the sequence was initially isolated. The wildtype GFP was optimized for the expression in *Candida albicans,* but it can be used in *S. cerevisiae* as well. The codon-usage optimized synthetic GFP (yEGFP) contains two mutations near the chromophore (S65G, S72A) that cause a 75-fold increase of fluorescence as well as higher solubility of the protein in *E. coli* and an up to 40-fold increase of fluorescence in *S. cerevisiae* (Cormack et al., 1997; Drew et al., 2008). Additionally, GFP containing the two mutations (S65G, S72A) matures much faster from the non-fluorescent to the fluorescent protein when expressed in bacterial cells. If this effect occurs in other expression hosts as well, then gene expression would be detectable relatively contemporaneous after induction (Cormack et al., 1996). The yEGFP protein absorbs light at a maximum of 488 nm and emits light in terms of fluorescence at a maximum of 512 nm (Drew et al., 2008).

Membrane protein expression experiments in *E. coli* using GFP as an expression reporter showed that GFP only exerts fluorescence upon insertion of the N-terminal membrane protein into the membrane (Drew et al., 2001). Additionally, Drew et al. (2008) published that the expression of a downstream membrane protein is less efficient than expression of an upstream yEGFP and thus, measurable fluorescence would not be longer a reliable reporter of membrane-integrated expression. On this account it was decided to fuse yEGFP to the C-terminus of CsPT1, CsPT2 and CsPT3.

Downstream of the C-terminal yEGFP a His_8-affinity purification tag is located. Expression reporter and purification tag can be removed by using Tobacco Etch Virus (TEV) protease if necessary. Up- and downstream of the TEV – yEGFP – His_8 fragment *Sal*I recognition sites were introduced for an easy excision of the *yEGFP - His_8* sequence and thus enable recombinant membrane protein expression with and without expression reporter.

The vector pDio + yEGFP was designed using Gibson Assembly (chapter 2.2.1.2) applying the high copy plasmid pDionysos (Stehle et al., 2008) as template. The TEV – yEGFP – His_8 fragment was amplified by PCR from the vector pDDGFP-2 (Drew et al., 2008). The used primers are shown in Table S 5 and a vector map of the generated plasmid is displayed in Figure 21.

CsPT1, CsPT2 and CsPT3 were cloned into the vector by restriction / ligation reaction using restriction enzymes *Bam*HI and *Sph*I, by Gibson Assembly or by homologous recombination. A 5'-UTR consensus sequence from *S. cerevisiae* was cloned into the 5'-untranslated region N-terminal of the prenyltransferase coding sequences (chapter 2.2.1.2).

The coding sequences of CsPT1, CsPT2 and CsPT3 were flanked by a strong *GAL1* promoter and a *CYC1* terminator. Gene expression was induced by the addition of galactose during *S. cerevisiae* cultivation.
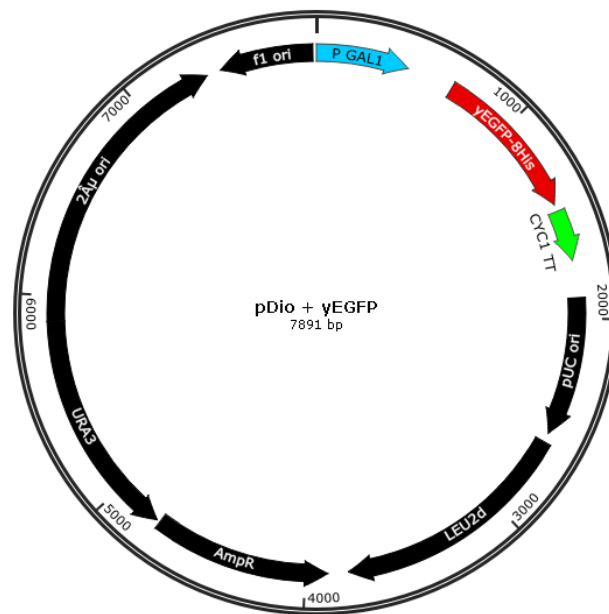
Figure 21: Map of the expression vector pDio + yEGFP. The plasmid uses the vector pDionysos (Stehle et al., 2008) as background and is modified by a yEGFP – His$_8$ fragment obtained from the vector pDDGFP-2 (Drew et al., 2008). The vector map was generated using SnapGene software (GLS Biotech, Chicago, USA).

3.1.4. Expression of CsPT1, CsPT2 and CsPT3

The coding sequences of CsPT1 and CsPT2 without predicted cTP (chapter 3.1.3.1) were amplified from cDNA by PCR. The cDNA was generated using template RNA isolated from the *C. sativa* variety Bedrocan (Bedrocan International B.V., Veendam, Netherlands) by MSc Julia Schachtsiek. The DNA fragments were cloned into the vector pDio + yEGFP (Figure 21) by homologous recombination (chapter 2.2.1.2). The coding sequence of CsPT3 without signal peptide was obtained by PCR using a vector generated by Pamplaniyil (2017) and cloned into the vector pDio + yEGFP by restriction / ligation (chapter 2.2.1.2).

In a second approach it was tested whether the use of the yCoq2 signal peptide affects the expression of the three aromatic prenyltransferase. In 2004, Ohara et al. showed that the overexpression of the membrane-bound aromatic prenyltransferase *p*-HB-polyprenyltransferase in *S. cerevisiae*, encoded by the yeast gene *Coq2* (yCoq2), led to an increase of UQ6. Thus, functional expression of *yCoq2* in *S. cerevisiae* is possible. The prenyltransferase gene contains a sequence coding for a N-terminal mTP. Similar to peptides with a cTP, peptides with a mTP are synthesized in the cytosol and are targeted to their destination after the translation without entering the secretory pathway (Glover and Lindsaytt, 1992). The length of the sequence coding for the yCoq2 mTP (yCoq2SP) was predicted by TargetP and the mTP was inserted into the pDio vectors upstream of coding sequences of *CsPT1*, *CsPT2* and *CsPT3* without cTP by mutagenesis PCR.

All constructs were generated with and without C-terminal yEGFP in order to exclude functional expression problems associated with the expression reporter, even though Drew et al. (2001) published that they did not observe any noteworthy differences when expressing membrane-bound

proteins with and without C-terminal expression reporter. The plasmids without yEGFP were obtained by restriction digestion with *Xba*I followed by ligation.

The plasmids were transformed into chemically competent SC-gal4 cells using standard LiAc method (Gietz and Schiestl, 2007). The cultivations of SC_gal4-CsPT1+, SC_gal4-CsPT2+ and SC_gal4-CsPT3+ (Table 3) as well as SC_gal4-yCoq2SP-CsPT1+, SC_gal4-yCoq2SP-CsPT2+ and SC_gal4-yCoq2SP-CsPT3+ (Table 3) were performed in two precultures and one main culture (chapter 2.2.1.3). After induction with galactose, the highest expression level was identified by measuring the whole-cell fluorescence (chapter 2.2.1.4). Different galactose concentrations were used for induction since this might affect expression levels of membrane-bound enzymes (personal communication: Joseph Lyons, Department of Molecular Biology and Genetics - DANDRITE, Aarhus University, Denmark). Cultivation conditions that resulted in the highest expression levels were repeated, cells were harvested and cell membranes were isolated (chapter 2.2.1.5). Activity assays were performed with GPP as prenyl donor and with four different prenyl acceptors: OA, olivetol, resveratrol and PHB (chapter 2.2.1.6). OA is the aromatic substrate, which is prenylated by CBGAS using GPP as donor resulting in the production of CBGA, the substrate of THCAS. Page and Boubakir (2014) showed that CBGAS additionally accepts olivetol and resveratrol as aromatic substrates. Since the function of CsPT2 and CsPT3 is unknown, PHB, the substrate of yCoq2, was tested as aromatic acceptor as well. 4-hydroxybenzoate (HB) polyprenyltransferase (yCoq2) catalyzes the prenylation of PHB with GPP and is involved in ubiquitin biosynthesis (Yazaki et al., 2002). The activity assay products were analyzed by LC-MS (chapter 2.2.1.7).

In chapter 3.1.4.1 the results of CsPT1, CsPT2 and CsPT3 expressed without N-terminal cTP are presented. The outcomes of the expression experiments of the three membrane-bound aromatic prenyltransferases fused to a yCoq2 targeting peptide at their N-terminus are shown in chapter 3.1.4.2. The results of both expression approaches are discussed in chapter 3.1.5.

3.1.4.1. <u>Expression of truncated CsPT1, CsPT2 and CsPT3</u>

S. cerevisiae Δpep4 Δgal4 cells carrying plasmids with the coding sequences of CsPT1, CsPT2 or CsPT3 (Table 4) were cultivated as described in chapter 2.2.1.3. After addition of galactose as inducer of protein expression the whole-cell fluorescence (chapter 2.2.1.4) and optical density (OD_{600}) were measured every three hours until the fluorescene intensity (FI) to OD_{600} ratio decreased.

The measured fluorescence values were normalized on the OD_{600}. The FI / OD_{600} ratio was normalized on the FI / OD_{600} of non-induced cultivations of the respective strains. This is necessary due to the yeast's autofluorescence which can amongst others be caused by flavins like flavin mononucleotide (FMN), FAD and their precursor vitamin B_2 (riboflavin). The latter two compounds show a maximal emission at 530 nm. In addition, flavoproteins and flavin coenzymes are potential sources of autofluorescence. Other fluorescent coenzymes, besides FAD and FMN, which are important for all living cells are the reduced forms of nicotinamide adenine dinucleotide (NAD^+) and nicotinamide adenine dinucleotide phosphate ($NADP^+$) (Billinton and Knight, 2001). Furthermore, the used yeast strain SC-gal4 has a mutated adenine biosynthesis (Table S 2; ade2-1) that causes the

accumulation of the metabolic intermediate phosphoribosylaminoimidazole, which is formed during purine biosynthesis and results in a pink phenotype of the cells under starvation conditions. This red intermediate can interfere with GFP detection and thus cause autofluorescence (Billinton and Knight, 2001; Kokina et al., 2014; Ugolini and Bruschi, 1996b).

The results of the whole-cell fluorescence measurements during expression of truncated CsPT1, CsPT2 and CsPT3 C-terminally fused to yEGFP (= SC-gal4_CsPT1+, SC_gal4-CsPT2+ and SC_gal4- CsPT3+) using different galactose concentrations as inducer of protein expression are depicted in Figure 22.
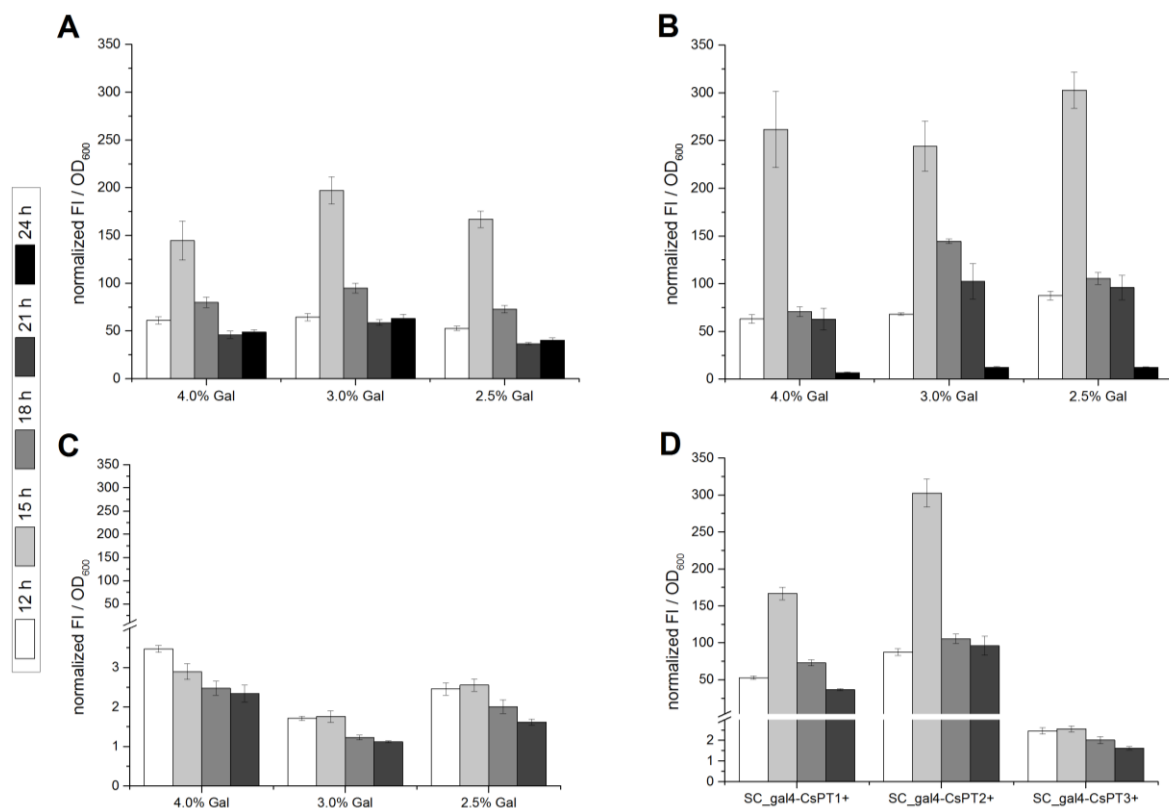
Figure 22: Results of the whole-cell fluorescence measurements of SC_gal4-CsPT1+, SC_gal4-CsPT2+ and SC_gal4-CsPT3+. The cell cultures contained 4% galactose (4.0% Gal), 3% galactose (3.0% Gal) or 2.5% galactose (2.5% Gal) as inducer of protein expression. The fluorescence intensities (FI) were determined in duplicates and normalized on the OD_{600}. The FI / OD_{600} values of the induced cell cultures were normalized on the FI / OD_{600} of the non-induced cell cultures of the respective strain. The measurements of OD_{600} and FI started 12 h after induction and were repeated every 3 h until the FI / OD_{600} ration decreased. (A): SC_gal4-CsPT1+, (B): SC_gal4-CsPT2+, (C): SC_gal4-CsPT3+, (D): Summary of normalized FI / OD_{600} values of SC_gal4-CsPT1+, SC_gal4-CsPT2+ and SC_gal4-CsPT3+ cell cultures containing 2.5% galactose. Data points are calculated from biological triplicates, each analyzed in technical duplicates

The results shown in Figure 22 indicate that expression of *CsPT1+*, *CsPT2+* and *CsPT3+* was successful because according to Drew et al. (2008) the C-terminal yEGFP only folds and becomes fluorescent if the upstream protein is expressed and integrated into a membrane. The expression levels of the different prenyltransferases were not significantly affected by the use of different inducer concentrations. Therefore, we decided to use galactose with a final concentration of 2.5% (Figure

22 D) as inducer for the upcoming experiments. The highest expression levels of CsPT1+ and *CsPT2+* were detectable 15 h after induction. The expression of *CsPT3+* was less successful than the expression of *CsPT1+* and *CsPT2+* because the FI / OD$_{600}$ values obtained were only up to 4-times higher than the values of the corresponding non-induced yeast cell culture (Figure 22 C). On the contrary, the FI / OD$_{600}$ values of CsPT1+ and CsPT2+ increased compared to their corresponding control strains up to 200-fold (Figure 22 A) and 300-fold higher (Figure 22 B), respectively. For the chosen cultivation conditions, the highest expression level of *CsPT3+* was obtained 12 h after induction. With this in mind, we decided to harvest CsPT1 and CsPT2 with and without C-terminal yEGFP 15 h after induction and CsPT3 12 h after induction in all further experiments (Table 13).

Table 13: Final galactose concentration and time of harvest for expression of *CsPT1+*, *CsPT2+* and *CsPT3+* in SC_gal4.

	Final galactose concentration [%]	Time of harvest after first induction [h]
SC_gal4-CsPT1	2.5	15
SC_gal4-CsPT2	2.5	15
SC_gal4-CsPT3	2.5	12

Until now, it was only predicted by the web-tool TMHMM (chapter 3.1.2.1) that the coding sequences of CsPT1, CsPT2 and CsPT3 belong to the group of membrane-bound enzymes. On this account, the fluorescence intensities of the different enzymes without predicted cTP were measured during the process of membrane isolation (chapter 2.2.1.5) in order to see whether the GFP fluorescence is present within the membrane fraction or the supernatant. The detected FI values were normalized on the FI value of the final membrane fraction (P2). As control the membranes of the corresponding non-induced cell culture were isolated as well. The obtained results are depicted in Figure 23.

The obtained results (Figure 23) indicate that CsPT1, CsPT2 and CsPT3 were integrated into the yeast membrane during expression. The expressed prenyltransferase-yEGFP fusions were detectable in the heavy membranes (centrifugation at 23,000 x g) and in the light membranes (centrifugation at 204,000 x g). The chosen ultracentrifugation conditions seem to be sufficient enough since only low FI values are detectable in the supernatants (SN1, SN2; Figure 23). The results of the control strains indicate that the autofluorescence of the yeast cells is caused by soluble and membrane fractions. In order to harvest as much yeast membrane fraction as possible for enzyme activity assays we decided to perform membrane isolation without splitting in heavy and light membranes. Additionally, we decided to use only one ultracentrifugation step instead of two because it is unknown whether washing the isolated membranes with potassium chloride containing GTEB buffer in order to remove peripheral membrane proteins is affecting the folding of CsPT1, CsPT2 and CsPT3 negatively.
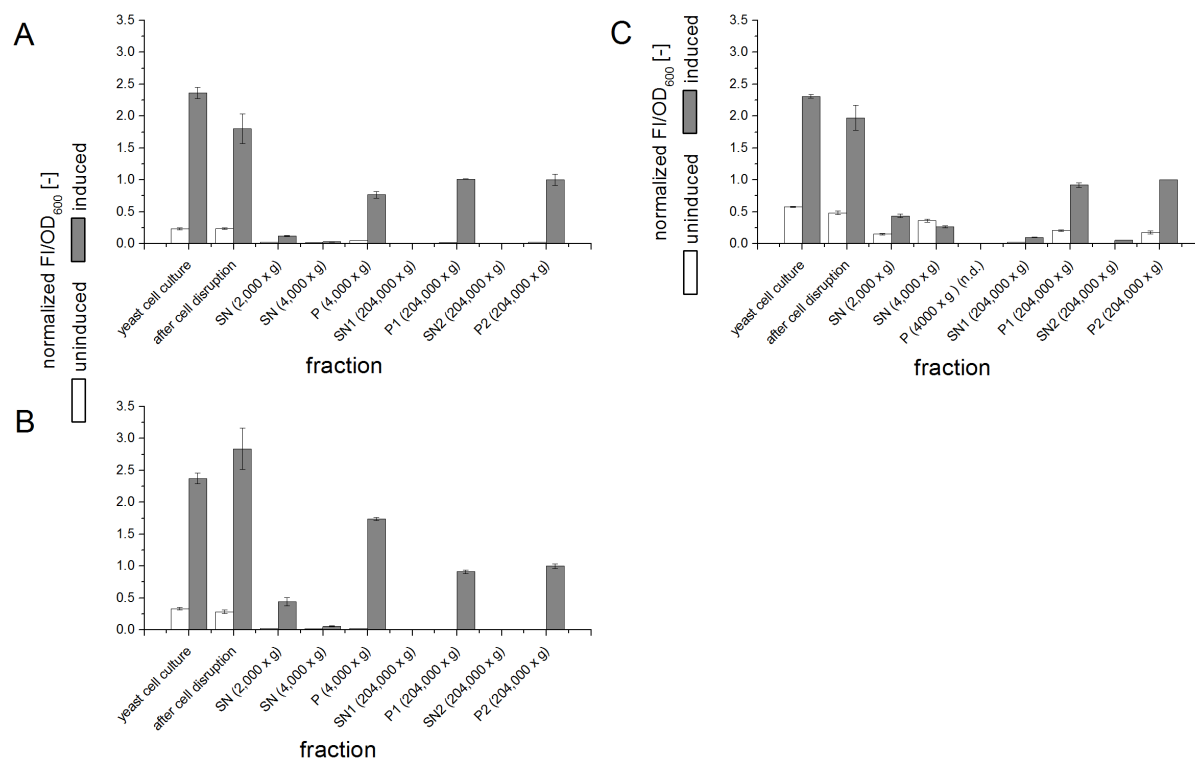
Figure 23: FI values detected during membrane isolation of SC_gal4-CsPT1+, SC_gal4-CsPT2+ and SC_gal4-CsPT3+. The detected FI values were normalized to the FI value of the final membrane fraction (P2) of the induced cell culture. The membranes of the corresponding non-induced cell cultures were isolated as controls. Yeast cell culture: Cells harvested after cultivation, after cell disruption: cells disrupted with glass beads, SN (4,000 x g): supernatant after centrifugation at 4,000 x g (removal of uncracked cells and cell debris); SN / P (23,000 x g): supernatant / pellet after centrifugation at 23,000 x g (the pellet contains the heavy membranes), SN / P (204,000 x g): supernatant / pellet after centrifugation at 204,000 x g (the pellet contains the light membranes). (A) SC_gal4-CsPT1+, (B) SC_gal4-CsPT2+, (C) SC_gal4-CsPT3+; the fluorescence intensity of P (4000 x g) was not determined (n.d.). Data points are calculated from biological duplicates, each analyzed in technical duplicates

Fluorescence measurements during membrane isolation (Figure 23) indicate that CsPT1, CsPT2 and CsPT3 belong to the group of membrane-bound enzymes. On this account, expressions of the three enzymes with (CsPTx+) and without (CsPTx-) C-terminal yEGFP were repeated using the conditions that yielded the highest fluorescence (Figure 22, Table 13). Subsequently, the yeast membranes were isolated and utilized for activity assay using OA, olivetol, resveratrol and PHB as aromatic substrates (chapter 2.2.1.6). The results of the activity assay were analyzed by LC-MS (chapter 2.2.1.7) and extracted ion chromatograms (EIC's) are depicted in Figure 24.
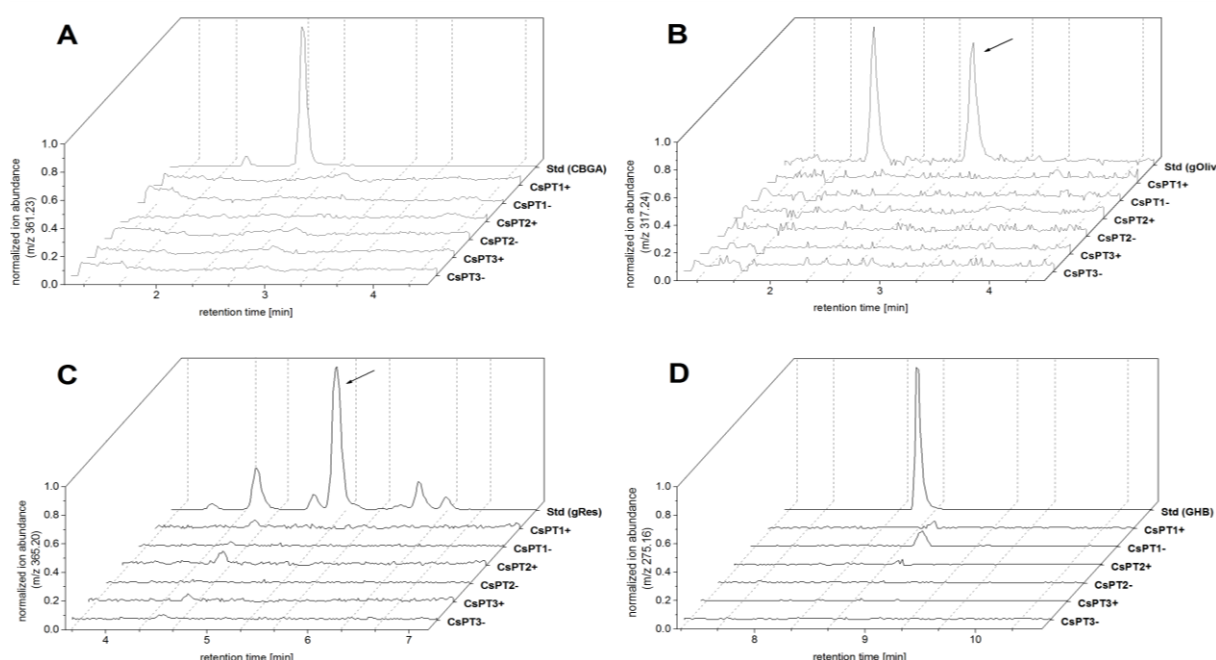
Figure 24: LC-MS results of CsPT1, CsPT2 and CsPT3. The isolated yeast membranes were incubated with 1 mM GPP and 1 mM of the respective aromatic substrate. (A) Extracted ion chromatograms (EIC) of *m/z* 361.23 (CBGA) of *CsPT1*, *CsPT2* and *CsPT3* expressed with and without C-terminal expression reporter. Std (CBGA): CBGA standard substance. (B) EICs of *m/z* 317.24 (CBG) of *CsPT1*, *CsPT2* and *CsPT3* expressed with and without C-terminal expression reporter. Std (gOliv): Olivetol geranylated by NphB (chapter 3.2.3), first peak: 2-geranyl olivetol, second peak: CBG (marked with an arrow). (C) EICs of *m/z* 365.20 (geranylated resveratrol) of *CsPT1*, *CsPT2* and *CsPT3* expressed with and without C-terminal expression reporter. Std (gRes): Resveratrol geranylated by NphB (marked with an arrow). (D) EICs of *m/z* 275.16 (GHB) of *CsPT1*, *CsPT2* and *CsPT3* expressed with and without C-terminal expression reporter. Std (GHB): PHB geranylated by yCoq2 (Pamplaniyil, 2017).

The results shown in Figure 24 indicate that the respective isolates were not able to catalyze the prenylation of OA with GPP to form CBGA since no peak was detectable at *m/z* 361.23. Additionally, no enzyme activity was detectable using olivetol (*m/z* 317.24), resveratrol (*m/z* 365.20) and PHB (*m/z* 275.16) as prenyl acceptor molecules. Since CsPT1, CsPT2 and CsPT3 did not catalyze mono-prenylation, LC-MS results were also analyzed with regard to di-prenylated compounds, but no enzyme activity was detectable as well (results not shown).

Even though the activity assays performed with OA, resveratrol, olivetol and PHB as prenyl acceptors and GPP as prenyl donor did not yield product formation detectable by LC-MS, measurable whole-cell fluorescence indicates that the *CsPT1+*, *CsPT2+* and *CsPT3+* were expressed and the enzymes were integrated into a *S. cerevisiae* membrane (organelles, plasma membrane). According to Drew et al (2008) the C-terminal expression reporter yEGFP only folds and becomes fluorescent if the upstream membrane protein was integrated into the membrane. In this context it has to be kept in mind that even if the protein is membrane-integrated and fluorescence is measureable, a functional expression is not ensured (Drew et al., 2008). On this account, missing enzyme activity (discussed in more detail in chapter 3.1.5) can be caused *inter alia* by non-functional expression under the chosen cultivation conditions or by the fact that none of the tested coding sequences belong to CBGAS.

Since the functional expressions of *CsPT1*, *CsPT2* and *CsPT3* without predicted cTP were not successful in *S. cerevisiae* it was decided to test a N-terminal signal peptide with yeast origin (yCoq2SP). The results are shown in chapter 3.1.4.2.

3.1.4.2. Expression of CsPT1, CsPT2 and CsPT3 with a N-terminal yCoq2 signal peptide

S. cerevisiae Δpep4 Δgal4 cells carrying pDio-yCoq2SP-CsPT1+, pDio-yCoq2SP-CsPT2+ or pDio-yCoq2Sp-CsPT3+ (Table 3, cf. chapter 3.1.4) were cultivated in two precultures and one main culture (chapter 2.2.1.3). After addition of galactose as inducer of protein expression the whole-cell fluorescence (chapter 2.2.1.4) and OD_{600} were measured every three hours until the FI to OD_{600} ratio decreased. The measured fluorescence values were normalized on the OD_{600}. The FI / OD_{600} ratio was normalized on the FI / OD_{600} of non-induced cultivations of the respective strains, which is necessary due to the yeast autofluorescence that can amongst others be caused by vitamin B_2, FMN and FAD (chapter 3.1.4.1).
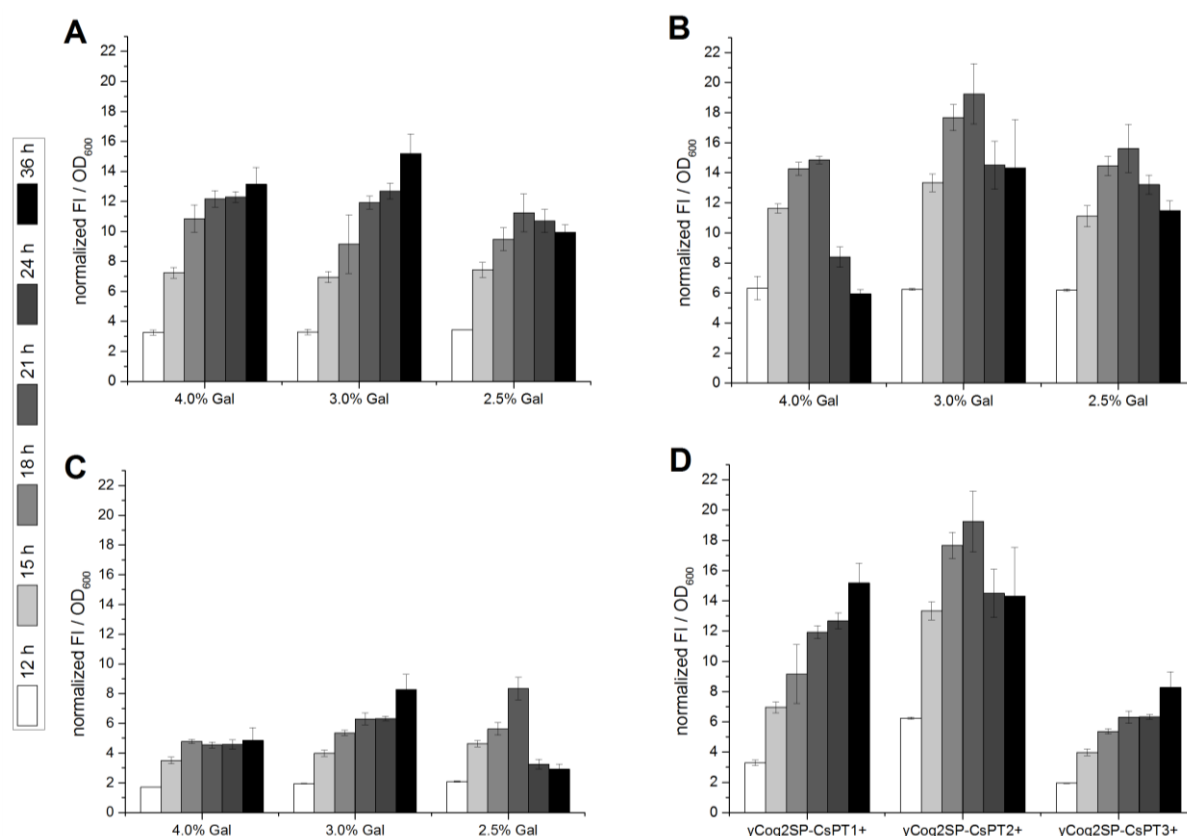
Figure 25: Results of the whole-cell fluorescence measurements of SC_gal4-yCoq2SP-CsPT1+, SC_gal4-yCoq2SP-CsPT2+ and SC_gal4-yCoq2SP-CsPT3+. The cell cultures contained 4% galactose (4.0% Gal), 3% galactose (3.0% Gal) or 2.5% galactose (2.5% Gal) as inducer of protein expression. The fluorescence intensities (FI) were determined in duplicates and normalized on the OD_{600}. The FI / OD_{600} values of the induced cell cultures were normalized on the FI / OD_{600} of the non-induced cell cultures of the respective strain. The measurements of OD_{600} and FI started 12 h after induction and were repeated every 3 h until the FI / OD_{600} ration decreased. (A) SC_gal4-yCoq2SP-CsPT1+, (B) SC_gal4-yCoq2SP-CsPT2+, (C) SC_gal4-yCoq2SP-CsPT3+, (D) Summary of normalized FI / OD_{600} values of SC_gal4-yCoq2SP-CsPT1+, SC_gal4-yCoq2SP-CsPT2+ and SC_gal4-yCoq2SP-CsPT3+ cell cultures containing 3.0% galactose. Data points are calculated from biological triplicates, each analyzed in technical duplicates

The results of the whole-cell fluorescence measurements obtained during the expressions of *yCoq2SP-CsPT1+*, *yCoq2SP-CsPT2+* and *yCoq2SP-CsPT3+* using different galactose concentrations as inducer of gene expression are depicted in Figure 25.

As an increased FI / OD_{600} ratio, compared to the control strains, was detectable, the expression of the three enzymes was considered successful. As mentioned before, the C-terminal yEGFP is only fluorescent if the upstream protein is expressed and integrated into a membrane (Drew et al., 2008). The expression levels of the different prenyltransferases were not significantly affected by the use of different inducer concentrations. On this account, it was decided to use galactose with a final concentration of 3.0% (Figure 25 D) as inducer for the subsequent experiments. The highest expression levels of *yCoq2SP-CsPT1+* and *yCoq2SP-CsPT2+* were obtained 36 h or 21 h after induction, respectively. The expression of *yCoq2SP-CsPT3+* was less successful than the expression of the *yCoq2SP-CsPT1+* and *yCoq2SP-CsPT2+* as the FI / OD_{600} values obtained were only up to 6-times higher than the values of the corresponding non-induced yeast cell culture (Figure 25 C). In contrast to that the FI / OD_{600} values of yCoq2SP-CsPT1+ increased up to 14-fold (Figure 25 A) and the calculated FI / OD_{600} values of yCoq2SP-CsPT2+ were up to 18-times higher (Figure 25 B) than their corresponding control strains. For the chosen cultivation conditions, the highest expression level of *yCoq2SP-CsPT3+* was obtained 36 h after induction. Therefore, it was decided, for the following experiments, to harvest the three prenyltransferase-yEGFP fusions as well as the enzymes without C-terminal yEGFP 36 h (CsPT1, CsPT3) or 21 h (CsPT2) after induction. Since the normalized FI / OD_{600} ratio was still increasing within the tested time frame, the maximal expression time could not be detected for SC_gal4-yCoq2SP-CsPT1+ (Figure 25 A) and SC_gal4-yCoq2SP-CsPT3+ (Figure 25 C; 3.0% galactose).

The results obtained by whole-cell fluorescence measurements indicate that *yCoq2SP-CsPT1+*, *yCoq2SP-CsPT2+* and *yCoq2SP-CsPT3+* were successfully expressed and integrated into a yeast membrane. On this account, expressions of the three enzymes with (yCoq2SP-CsPTx+) and without (yCoq2SP-CsPTx-) C-terminal yEGFP were repeated using the conditions that yielded the highest FI / OD_{600} ratio (Table 14). Subsequently, the yeast membranes were isolated and utilized for activity assay using OA, olivetol, resveratrol and PHB as aromatic substrates (chapter 2.2.1.5, chapter 2.2.1.6). The results of the activity assays were analyzed by LC-MS (chapter 2.2.1.7) and EIC's are displayed in Figure 26.

Table 14: Final galactose concentration and time of harvest for expression of *yCoq2SP-CsPT1+*, *yCoq2SP-CsPT2+* and *yCoq2SP-CsPT3+* in SC_gal4.

	Final galactose concentration [%]	Time of harvest after first induction [h]
SC_gal4-CsPT1	3.0	36
SC_gal4-CsPT2	3.0	21
SC_gal4-CsPT3	3.0	36

The results shown in Figure 26 indicate that *yCoq2SP-CsPT1*, *yCoq2SP-CsPT2* and *yCoq2SP-CsPT3* expressed with and without C-terminal yEGFP are not able to catalyze the prenylation of OA with GPP to form CBGA since no peak was detectable at *m/z* 361.23. Additionally, no enzyme activity was detectable using olivetol (*m/z* 317.24), resveratrol (*m/z* 365.20) and PHB (*m/z* 275.16) as prenyl acceptor. Since yCoq2SP-CsPT1, yCoq2SP-CsPT2 and yCoq2SP-CsPT3 did not catalyze mono-prenylation, LC-MS results were analyzed with regard to di-prenylated compounds, but no enzyme activity was detectable either (results not shown).

Even though the expression of CsPT1, CsPT2 and CsPT3 with a N-terminal mTP with yCoq2 origin, instead of a predicted N-terminal cTP, did not yield product formation, detectable whole-cell fluorescence indicates that the three genes were expressed and the enzymes were integrated into a *S. cerevisiae* membrane (organelles, plasma membrane; cf. chapter 3.1.4.1).

A comparison of the whole-cell fluorescence measurements of the enzymes with and without N-terminal presequence (this chapter and chapter 3.1.4.1) indicates that the expression levels of CsPT1 and CsPT2 (Figure 22) are significantly higher than the expression levels of yCoq2SP-CsPT1 and yCoq2-CsPT2 (Figure 25) which aligns with the results depicted by Akashi et al. (2008), Munakata et al. (2014) and Li et al. (2015) who reported higher activities when expressing constructs without cTP (cf. chapter 3.1.3.1). In contrast to that the expression levels of CsPT3 (Figure 22) are decreased in comparison to the expression levels of yCoq2SP-CsPT3 (Figure 25). The lowest expression levels were obtained for CsPT3 and yCoq2SP-CsPT3.

Possible reasons for missing enzyme activity, under the chosen expression and activity assay conditions, might be the fact that none of the tested coding sequences belong to CBGAS or non-functional expression. This will be discussed in more detail in chapter 3.1.5.
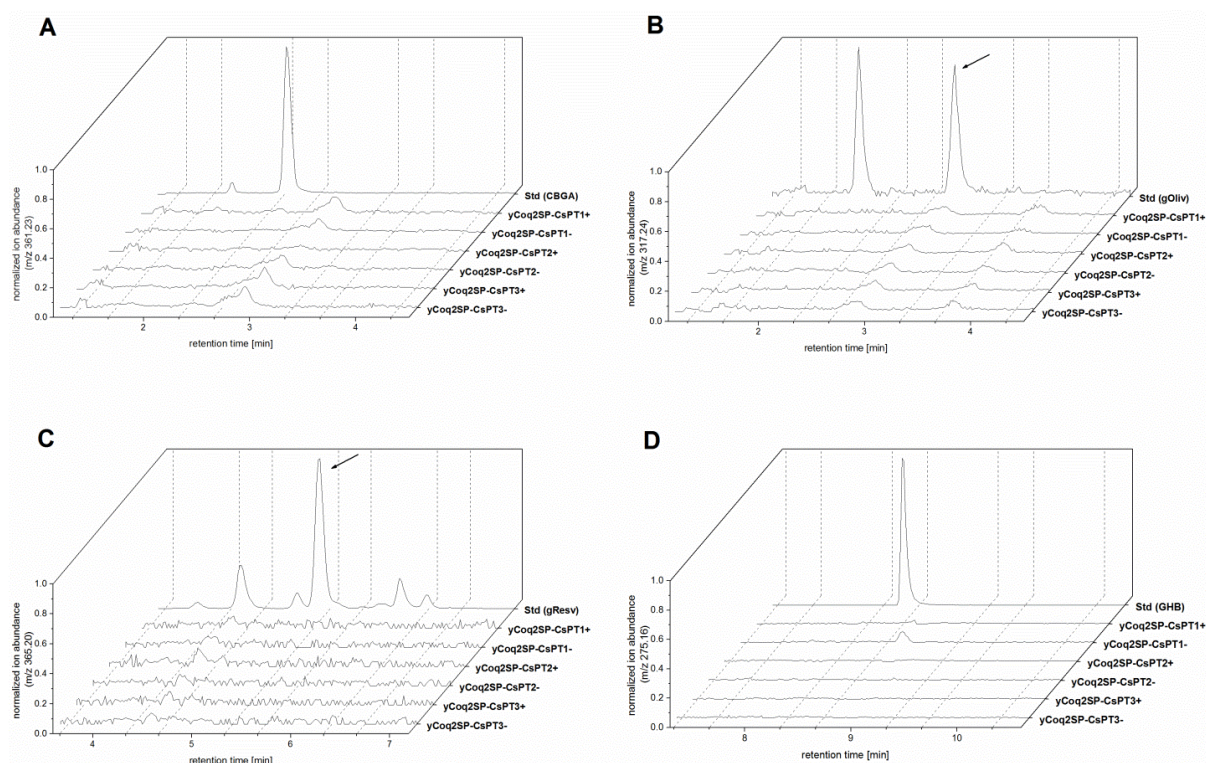
Figure 26: LC-MS analysis yCoq2SP-CsPT1, yCoq2SP-CsPT2 and Coq2SP-CsPT3. The isolated yeast membranes were incubated with 1 mM GPP and 1 mM of the respective aromatic substrate. (A) Extracted ion chromatograms (EIC) of *m/z* 361.23 (CBGA) of yCoq2SP-CsPT1, yCoq2SP-CsPT2 and yCoq2SP-CsPT3 expressed with and without C-terminal expression reporter. Std (CBGA): CBGA standard substance. (B) EICs of *m/z* 317.24 (CBG) of yCoq2SP-CsPT1, yCoq2SP-CsPT2 and yCoq2SP-CsPT3 expressed with and without C-terminal expression reporter. Std (gOliv): Olivetol geranylated by NphB (chapter 3.2.3), first peak: 2-geranyl olivetol, second peak: CBG (marked with an arrow). (C) EICs of *m/z* 365.20 (geranylated resveratrol) of yCoq2SP-CsPT1, yCoq2SP-CsPT2 and yCoq2SP-CsPT3 expressed with and without C-terminal expression reporter. Std (gRes): Resveratrol geranylated by NphB (marked with an arrow). (D) EICs of *m/z* 275.16 (GHB) of yCoq2SP-CsPT1, yCoq2SP-CsPT2 and yCoq2SP-CsPT3 expressed with and without C-terminal expression reporter. Std (GHB): PHB geranylated by yCoq2 (Pamplaniyil, 2017).

3.1.5. Discussion

The enzyme activity assays of CsPT1, CsPT2 and CsPT3 expressed without their predicted cTPs (Figure 24), as well as the activity assays of prenyltransferases with a N-terminal yCoq2SP (Figure 26) did not result in product formations detectable by LC-MS. The absent conversion of the substrates to geranylated products might *inter alia* be caused by non-functional expression or by the fact that none of the used coding sequences belong to the CBGAS.

Although detectable whole-cell fluorescence measurements indicate that the yEGFP-fusion proteins CsPT1, CsPT2 and CsPT3, each without N-terminal putative cTP (Figure 22) and with N-terminal yCoq2SP (Figure 25), were expressed and integrated into the membrane, it has to be kept in mind that detectable yEGFP fluorescence is only an indicator for membrane integration of the expressed protein and does not necessarily support assessment of the functional expression of the desired protein. Nevertheless, the use of yEGFP as an expression reporter helps to speed up the process of finding conditions for functional expression (Drew et al., 2008).

One possible explanation for non-functional expression might be the fact that the used gene sequences did not encode for a functional protein. There are many publications available that recommend the use of codon-usage optimized coding sequences in order to increase heterologous gene expression in the non-native host organism (e.g. Kaishima et al., 2016; Lanza et al., 2014; Sheff and Thorn, 2004). In 2015, Li et al. screened whether codon optimization of the membrane-bound aromatic prenyltransferases *HlPT1L* and *HlPT2*, which are involved in bitter acid biosynthesis in *H. lupulus,* would lead to a higher production of bitter acid. On this account, they tested yeast codon-optimized gene sequences (with N-terminal signal peptide). Interestingly, yeast strains containing the yeast optimized sequences produced even less prenylated product than yeast strains harboring genes sequences of hop origin (Li et al., 2015). Additionally, amongst others Munakata et al. (2014) and Akashi et al. (2008) showed that is possible to express membrane-bound aromatic prenyltransferase with plant origin functionally in *S. cerevisiae* even though the gene sequences were not codon-usage optimized. On this account, we decided to use CsPT1, CsPT2 and CsPT3 coding sequences with *C. sativa* origin for expression experiments. Maybe CsPT1, CsPT2 and CsPT3 can be expressed functional in *S. cerevisiae* if codon-usage optimized coding sequences would be used instead.

Besides the incorrect codon-usage of the used coding sequences, the non-functional expression could also be caused by the N-terminal truncation of the gene sequences. Analysis of the amino acid sequences of CsPT1, CsPT2 and CsPT3 with the web-tools ChloroP, Predotar and TargetP (chapter 3.1.2.2) indicated the presence of N-terminal cTPs. The obtained score values of ChloroP (Figure 19) for the three enzymes were below 0.55 and thus these predictions have to be considered carefully (Emanuelsson et al., 2007). Additionally, Shen et al. (2012) showed that microsomal fractions containing a prenyltransferase without predicted cTP (ChloroP) resulted in similar enzyme activity levels as microsomal fractions containing full length protein. Six times higher activity levels, compared to the wildtype enzyme, were only achieved when truncating the N-terminus significantly more than the length predicted by ChloroP. This observation is supported by Emanuelsson et al. (1999) who published that in a set of 62 proteins annotated in SWISS-PROT 54 proteins contained a longer cTP than predicted by ChloroP. 3 enzymes out of 62 possessed a shorter cTP than predicted and for only five proteins the cTP length was predicted correctly.

A comparison of the whole-cell fluorescence measurements of SC-gal4_CsPT3+ and SC-Gal4_yCoq2SP-CsPT3+ indicates that the expression levels of *CsPT3* without cTP (Figure 22) are decreased in comparison to the expression levels of *CsPT3* with a N-terminal mTP (yCoq2SP) (Figure 25). TargetP and Predotar predicted the presence of a cTP, even though the prediction of TargetP was not classified as reliable (Figure 18, Figure 20). In addition, ChloroP predicted the absence of a N-terminal cTP, but the prediction had to be considered carefully (Figure 19). Therefore, we decided to truncate the first 77 AA of CsPT3, the length of the targeting sequence predicted by ChloroP and TargetP. A comparison of the FI / OD$_{600}$ values of CsPT1, CsPT2 and CsPT3 indicates that the expression of *CsPT3* was less successful than the expression of the other two prenyltransferases (Figure 22, Figure 25). Taking the experimental results and the results of the cTP prediction tool into

account it might be possible that either CsPT3 does not contain a N-terminal targeting sequence at all, or the cTP length, predicted by ChloroP and TargetP was incorrect.

Taking the predictions of ChloroP and the published results of Emanuelsson et al. (1999) as well as Shen et al. (2012) into account, functional expression of *CsPT1*, *CsPT2* and *CsPT3* without predicted cTP might be possible if different truncations of the N-terminus would be tested. Assuming that the predicted size of the cTP was incorrect and the truncation led to non-functional expression, then the use of the yCoq2 signal peptide would most likely also cause non-functional expression because (i) the sequences of two targeting peptides would be mixed (if cTP is longer than predicted) or (ii) the nascent prenyltransferase sequence is shorter than needed for functionality (if cTP is shorter than predicted). Additionally, expression experiments of *CsPT1 wt*, *CsPT2 wt* and *CsPT3 wt* should be performed in *S. cerevisiae* in order to exclude non-functional expression due to a truncated N-terminus. In the context of truncation of the N-terminus of the prenyltransferase, it should be remembered that this could cause a change in membrane topology of the membrane-bound aromatic prenyltransferase.

Assuming that functional expression of the used CsPT1, CsPT2 and CsPT3 coding sequences is possible, the lack of enzymatic activity might be caused by using a heterologous expression host possessing the wrong membrane lipid composition.

In a first experiment, the coding sequences of CsPT1, CsPT2 and CsPT3 were expressed without predicted N-terminal targeting sequence. Due to their hydrophobic characteristics the three enzymes can integrate into each membrane of the yeast cell, including organelles (Figure 27). In a second experiment, the gene sequences of the three prenyltransferases were expressed with an N-terminal yCoq2 signal peptide targeting the enzymes to the yeast mitochondria.
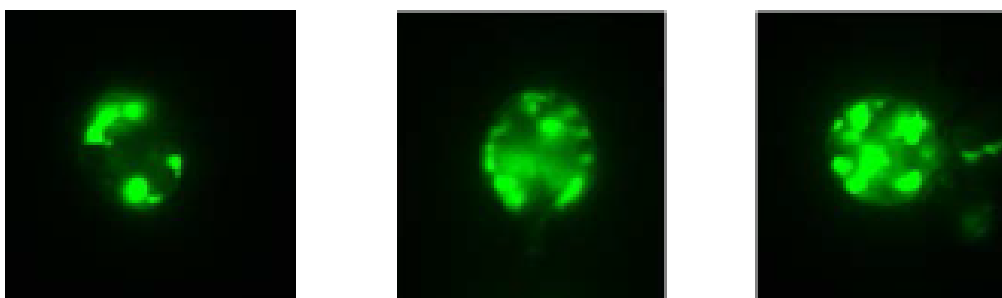
Figure 27: Fluorescence microscopy images of *CsPT1+* expressed in *S. cerevisiae*. Inverse microscopy; ocular: 40x, objective lens: 100x. Left image: possible localization of the enzyme in the endoplasmatic reticulum; image in the center: possible localization of the enzyme in the lipid rafts of the plasma membrane; right image: localization of the enzyme in several different membranes of the yeast cell / yeast organells [a].

The plasma membranes as well as the membranes of mitochondria, vacuoles and the endoplasmic reticulum of *S. cerevisiae* mostly contain phospholipids like phosphatidylethanolamine, phosphatidylcholine and phosphatidylinositol. Additionally, the yeast plasma membrane possesses a high content of phosphatidylserine, sphingolipids and sterols like ergosterol. In the mitochondrial membranes, mostly in the inner membranes, a high content of cardiolipin can be detected (Horvath

and Daum, 2013; Mårtensson et al., 2017). The membranes of chloroplasts however consist to a high degree of galactolipids like mono-galactosyldiacylglycerol and di-galactosyldiacylglycerol instead of phospholipids. Additionally, they do not contain any sterols or sphingolipids (Horvath and Daum, 2013; Schmid and Ohlrogge, 2008). A comparison of the different lipid compositions indicates that the plasma membrane and the organelle membranes of the chosen expression host *S. cerevisiae* might contain a lipid composition unsuitable for functional expression of *CsPT1*, *CsPT2* and *CsPT3*. On this account, the use of a plant expression host should be taken into consideration for identifying CBGAS coding sequence, as - besides a possibly more favourable lipid composition - the wildtype enzymes could be targeted to their natural acceptor membrane. Investigating different yeast strains with more favorable membrane lipid compositions for enzymes of plant origin might present a further approach.

As mentioned in chapter 3.1.1 the amino acid sequences of CsPT1 and CsPT1P (Page and Boubakir, 2014) share a percentage sequence identity of 98%. Assuming CBGAS to be the central branch-point intermediate for the biosynthesis of C5-cannabinoid acids (Degenhardt et al., 2017), one could expect the aromatic prenyltransferase catalyzing CBGA formation to be highly specific for OA and GPP. Page and Boubakir (2014) calculated K_M values of CsPT1P for OA (60 mM) and GPP (150 mM). Here, however, it is not made clear which expression host was used and if purified enzyme was used for the determination of K_M. In Table 15 apparent K_M values of different membrane-bound aromatic prenyltransferases, which are involved in secondary metabolism, are displayed. Compared with the K_M values presented in Table 15 the values published by Page and Boubakir (2014) seem to be high for specialized enzymes. Additionally, the apparent K_M value of CBGAS for GPP determinded by Fellermeier and Zenk (1998) using *C. sativa* crude extract is 75-times smaller than the one calculated by Page and Boubakir (2014) (Table 15). Results published by Bar-Even et al. (2011) indicate a median K_M value of about 35 µM for enzymes of the secondary metabolism. Therefore, the K_M values of the CBGAS are more likely in the µM range than in the high mM range as published Page and Boubakir (2014). However, results of Page and Boubakir (2014) indicate that phlorisovalerophenone (PIVP) and not OA is the preferred aromatic substrate of CsPT1P. In hop PIVP formation is catalyzed by phlorisovalerophenone synthase (VPS) using isovaleryl-CoA and three molecules of malonyl-CoA. VPS is a soluble aromatic polyketide synthase localized in the cone glandular hairs of *H. lupulus* (Paniego et al., 1999; Zuurbier et al., 1998). In 2004, Raharjo et al. isolated a type III polyketide synthase (PKS) mRNA from leaves of a female *C. indica* dominant hybride (80 % indica, 20 % sativa) that catalyzes amongst others the formation of PIVP (Flores-Sanchez and Verpoorte, 2008; Raharjo et al., 2004). It is noteworthy, that no free OA or PIVP has been detected in *Cannabis* trichomes until now. Even though CsPT1P is able to catalyze the prenylation of OA with GPP to form CBGA, the calculated K_M values published by Page and Boubakir (2014) indicate that OA might not be the natural substrate of this enzyme. Assuming CBGAS to be highly specific, CsPT1P and thus CsPT1 (98% percentage sequence identity; chapter 3.1.1) are probably not the enzymes sought after. Nevertheless, it cannot be excluded that CBGAS, like other plant enzymes involved in the secondary metabolism, exhibit a broad substrate spectrum (Bar-Even et al., 2011).

Table 15: Apparent K_M values of membrane-bound aromatic prenyltransferases. The presented enzymes are involved in secondary metabolism. [1]It is unknown, whether *S. cerevisiae* or *Sf9* cells were used as expression host.

Enzyme	GenBank[TM] accession number	Expression host	K_M value [mM], (substrate)	Reference
CBGAS	-	*C. sativa*	2 (GPP)	(Fellermeier and Zenk, 1998)
	AJN57774	*S. cerevisiae* or *Sf9* cells[1]	150 (GPP)[1]	(Page and Boubakir, 2014)
G4DT	AB434690	*S. cerevisiae*	0.105 (DMAPP)	(Akashi et al., 2008)
LaPT1	JN228254	*S. cerevisiae*	0.055 (DMAPP)	(Shen et al., 2012)
LePGT1	AB055078	*S. cerevisiae*	0.010 (PHB) 0.005 (GPP)	(Yazaki et al., 2002)
SfN8DT-1	AB325579	*S. cerevisiae*	0.106 (DMAPP)	(Sasaki et al., 2008)

As mentioned in chapter 1.3.1.1, contrary to Page and Boubakir (2014), Fellermeier and Zenk (1998) were not able to detect CBGAS activity in the isolated particulate fractions, but in the soluble fraction obtained from young expanding hemp leaves. Hence, it might possible that CBGAS is not a membrane-bound aromatic prenyltransferase, but a soluble one. In contrast to membrane-bound aromatic prenyltransferase, soluble enzymes do not contain a characteristic aspartate-rich motif for prenyl diphosphate binding through divalent cations (Saleh et al., 2009; Steffan et al., 2009; Tello et al., 2008; Winkelblech et al., 2015). Thus, the motif based search in *C. sativa* cDNA libraries, performed by Pamplaniyil (2017) and basis for these expression experiments (chapter 3.1), could not lead to success.

Similar scientific disagreement on the subject of whether an enzyme is soluble or membrane-bound, exist on the features of two aromatic prenyltransferases involved in the bitter acid pathway in *H. lupulus* (chapter 1.3.1.1). According to Zuurbier et al. (1998) the prenylation steps in the biosynthesis of bitter acid are catalyzed by soluble aromatic prenyltransferases, while Li et al. (2015) reported that these two aromatic prenyltransferases are membrane-bound. In summary, there are inconsistent reports for *C. sativa* and *H. lupulus* on enzymatic steps being catalyzed by soluble or membrane-bound prenyltransferases. *C. sativa* and *H. lupulus* belong to the family of Cannabaceae and the biosynthesis of cannabinoids as well as the formation of bitter acids is localized in the trichomes of the corresponding plant (Degenhardt et al., 2017; Li et al., 2015; Zuurbier et al., 1998). Up to now, all plant aromatic prenyltransferases are membrane-bound, except the ones described by Zuurbier et al. (1998) (Sasaki et al., 2008; Tello et al., 2008; Winkelblech et al., 2015; Yamamoto et al., 2000; Zhao et al., 2003). However, it can not be excluded that CBGAS is a soluble enzyme. Based on this, CBGA formation catalyzed by CsPT1, CsPT2 or CsPT3 would not be possible since these

enzymes are membrane-bound. Therefore, the search for CBGAS should not only be performed using bioinformatics methods like looking for e.g. specific motifs, but also experimental methods e.g. performing activity assays with microsomal and soluble fractions obtained from *C. sativa* flowers / trichomes to investigate the localization of the prenyltransferase.

The online Basic Local Alignment Search Tool (BLAST[®], https://blast.ncbi.nlm.nih.gov/Blast.cgi; 01.03.2018) searches the chosen sequence database for local alignments (most similar regions) to the input sequence (Altschul, 1997; Altschul et al., 1990; Johnson et al., 2008). The BLAST[®] analysis of the protein sequences of CsPT1, CsPT2 and CsPT3 yielded the same results when only taking the first three entries of the results table into account (Table 16, Table S 6, Table S 7).

Table 16: Results of BLAST[®] for CsPT1. The results shown are limited to the first three entry sequences because for the other potential prenyltransferases no experimental data is available. The Chosen Search Set contained the following settings: Database: Non-redundant protein sequence (nr); organism: -; exclude: models (XM/XP), uncultured/environmental sample sequence; algorithm: blastp (protein-protein BLAST). Description: name of matched database sequence; Max score: highest alignment score; Total score: total alignment scores; Query cover: percentage overlap of input sequence and sequence of the protein mentioned in the description; E value: lowest Expect value; Ident: highest percent identity; Accession: UniProtKB/Swiss-Prot accession number.

Description	Max Score	Total Score	Query cover	E value	Ident	Accession
2-acylphloroglucinol 4-prenyltransferase	357	357	100%	7e-118	49%	A0A0B5A051.1
2-acylphloroglucinol 4-prenyltransferase, chloroplastic	353	353	100%	3e-116	48%	E5RP65.1
2-acyl-4-prenylphloroglucinol 6-prenyltransferase, chlorplastic	301	301	100%	9e-96	45%	A0A0B4ZTQ2.1

CsPT1, CsPT2 and CsPT3 show the highest sequence identities with three aromatic prenyltransferase of hop origin (43% – 50% sequence percentage identity). The membrane-bound aromatic prenyltransferases HIPT1L, HIPT-1 and HIPT2 (GenBank[TM] accession numbers: A0A0B5A051, E5RP65, A0A0B4ZTQ2) possess a N-terminal cTP and are involved in the bitter acid production in *H. lupulus* by catalyzing the prenylation of acylphloroglucinols using DMAPP as prenyl donor (Li et al., 2015; Tsurumaru et al., 2012, 2010). The amino acid sequences of the aromatic prenyltransferases with *H. lupulus* origin as well as CsPT1, CsPT2 and CsPT3 contain the same aspartate-rich motif (NQIxDxxID) responsible for prenyl diphosphate binding (Heide, 2009; Li et al., 2015; Sasaki et al., 2008; Stec and Li, 2012). Therefore, it might be possible that CsPT1, CsPT2 and CsPT3 did not prenylate OA, olivetol, resveratrol or PHB because the wrong prenyl donor (GPP instead of DMAPP) was applied. Additionally, they possess the same second aspartate-rich motif (KDxSDxxGD) which might be responsible for binding of the aromatic substrate (Table S 3) (Li et al., 2015; Stec and Li, 2012). Taking into account that the enzymes of *C. sativa* and *H. lupulus* share the same aspartate-rich

motifs, it might be possible that the missing enzyme activity was caused due to the fact that neither CsPT1 nor CsPT2 or CsPT3 are CBGAS. Based on activity assays with yeast microsomes containing the expressed aromatic prenyltransferases, the enzymes of hop origin accept acylphloroglucinols as aromatic acceptors (Figure 28) and most likely DMAPP (see above) as prenyl donor (Li et al., 2015; Tsurumaru et al., 2012, 2010). In 2015, Li et al. published that they were only able to detect bitter acid production if they co-expressed HIPT1L and HIPT2. Additionally, they showed that a metabolon was formed by interactions between the two enzymes. Keeping this in mind, production of CBGA might only be possible if CBGAS is co-expressed with a so far unknown enzyme in order to form a prenyltransferase complex.
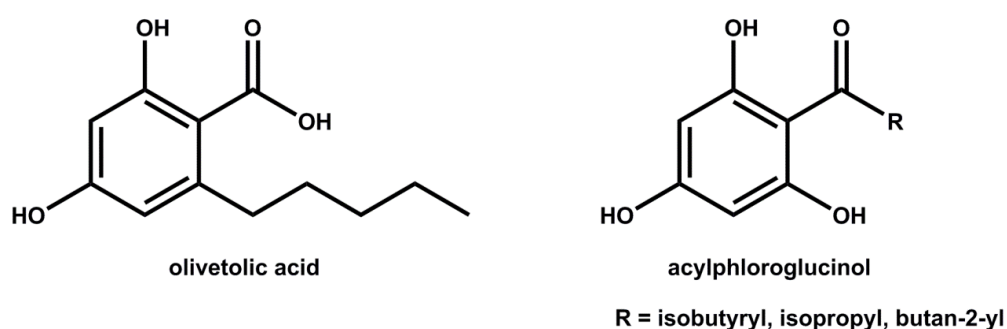
Figure 28: Structural formulas of olivetolic acid and acylphloroglucinol. HIPT-1 accepts phlorisovalerophenone (PIVP), phloisobutyrophenone (PIBP) and phlormethylbutanophenone (PMBP) as substrate (Li et al., 2015). PIVP – R = isobutyryl, PIBP – R = isopropyl, PMBP – R = butan-2-yl.

It is noteworthy, that the amino acid sequence of CsPT1P (GenBank[TM] accession number: AJN57774) cannot be found in "non-redundant protein sequences (nr)" database, but in the NCBI database of "patented protein sequences(pat)". An alignment of the three prenyltransferases with other enzymes of *Cannabis* origin is only possible to a limited extent because only parts of the *Cannabis* genome are annotated to the NCBI database (Oh et al., 2016; van Bakel et al., 2011).

3.2.   Expression of *nphB* in *Escherichia coli*

Assuming CBGAS to be an integral membrane protein (chapter 1.3.1.1), this could be challenging in terms of generating a cannabinoid producing yeast. Besides incorrect organell localization (yeast does not contain chloroplasts; chapter 3.1.5), low expression levels of catalytically slow enzymes of secondary metabolism present a problem (Zirpel, Degenhardt et al., 2017). Additionally, functional expression of CBGAS was not successful so far (chapter 3.1.4.1, chapter 3.1.4.2). In order to circumvent these issues, we decided to replace the membrane-bound aromatic prenyltransferase by a soluble enzyme.

According to Kuzuyama et al. (2005) the soluble aromatic prenyltransferase NphB (chapter 1.3.2.1), first isolated from *Streptomyces* sp. strain CL190, catalyzes the prenylation of OA using GPP as prenyl donor. On this account, NphB is a promising alternative for the CBGAS in terms of CBGA formation. However, it should be noted that Kuzuyama et al. (2005) evaluated the enzymatic assays by thin layer chromatography (TLC) and did not analyze the obtained prenylation products by nuclear magnetic resonance spectroscopy (NMR) or MS. Thus, it remains unclear whether CBGA was formed or if OA was geranylated at other positions than position C3 (Figure 29). As mentioned before (chapter 1.3.2.1) NphB catalyzes carbon-carbon and carbon-oxygen based geranylation of hydroxyl-containing aromatic acceptor molecules in para- or ortho-position of a hydroxyl group (Kumano et al., 2008; Kuzuyama et al., 2005).
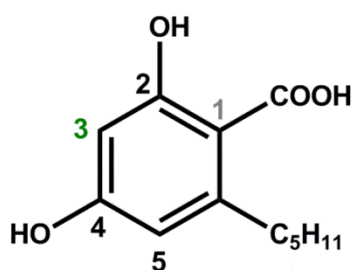

Figure 29: Possible geranylation sites of OA. The soluble, aromatic prenyltransferase NphB is able to catalyze carbon-carbon and carbon-oxygen based geranylation of hydroxyl containing aromatic acceptor molecules. Possible C-C prenylations can take place at positions C3 (CBGA) and C5. C-O based geranylations are possible at position 2-*O* and 4-*O*. Modified from Zirpel, Degenhardt et al. (2017).

In the following chapter the expression of *nphB* in *E. coli* is presented (chapter 3.2.1). In addition, the results of the activity assays that were perfomed with 1,6-DHN, OA and olivetol as aromatic acceptor substrates are shown (chapter 3.2.2, chapter 3.2.5). At the end of this chapter the mutational studies of *nphB* variants are presented (chapter 3.2.5, chapter 3.2.6).

3.2.1.   Screening for an *E. coli* strain with the highest *nphB* expression level

The *E. coli* codon-optimized coding sequence of NphB was cloned into the expression vector pET32a, followed by transformation of pET32a-NphB into different *E. coli* strains in order to screen for the

highest *nphB* expression level. The used *E. coli* strains and their special features regarding protein expression are displayed in Table S 11. The expression experiments were perfomed in LB medium and the expression was induced by the addition of isopropyl-β-D-1-thiogalactopyranoside (IPTG) (chapter 2.2.2.3). The *nphB* expression levels were analyzed by SDS-PAGE (chapter 2.2.2.5).

At the N-terminus *nphB* contains a TrxA tag and a His$_6$ tag that can be cleaved off by enterokinase treatment (Table S 13). *trxA* encodes for the thiol-disulfide oxidoreductase thioredoxin-1 (GenBank^{TM} accession number: BAE77517.1) (Schultz et al., 1999). The TrxA tag belongs to the group of solubility tags and is used in order to enhance expression of soluble proteins in *E. coli* (LaVallie et al., 2000). Results of Dyson et al. (2004) indicate that TrxA is more effective as solubility enhancing tag when fusing it to the N-terminus of the target protein. The His$_6$ tag can be used as purification tag (Malhotra, 2009). The aromatic prenyltransferase NphB (GenBank^{TM} accession number: AB187169) has a calculated molecular weight of 33.74 kDa. Including the N-terminal fused solubility enhancer and purification tags, the NphB fusion protein has a molecular mass of about 51.1 kDa. The SDS-PAGEs in Figure 30, Figure S 6 and Figure S 7 display the results of *nphB* expression in different *E. coli* strains.
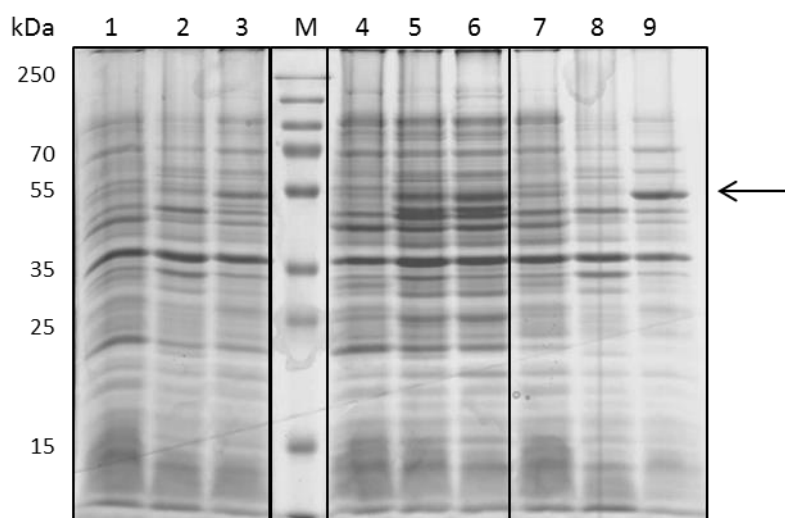

Figure 30: Coomassie stained SDS gel of *nphB* expressed in different *E. coli* strains. 1: Rosetta 2 (DE3) pLysS 0 h, not induced; 2: Rosetta 2 (DE3) pLysS 17 h, not induced; 3: Rosetta 2 (DE3) pLysS 17 h, induced; M: Thermo Scientific PageRuler Plus Prestained Protein Ladder; 4: OverExpress^{TM} C43 (DE3) pLyS 0 h, not induced; 5: OverExpress^{TM} C43 pLysS 17 h, not induced; 6: OverExpress^{TM} C43 (DE3) pLysS 17 h, induced; 7: BL21 (DE3) 0 h, not induced; 8: BL21 (DE3) 17 h, not induced; 9: BL21 (DE3) 17 h, induced. The size of the expected protein band is marked with an arrow.

The expression of *nphB* in *E. coli* is possible, even though different expression levels were obtained using various strains. The highest expression level was observed when expressing *nphB* in *E. coli* BL21 (DE3) (Figure 30). Therefore, it was decided to perform all further *nphB* expression experiments using the expression protocol presented in chapter 2.2.2.3 and *E. coli* BL21 (DE) as expression host. Interestingly, the non-induced samples of *E. coli* OverExpress^{TM} C43 (DE3) pLysS (Figure 30), *E. coli* Rosetta 2 (DE3) and *E. coli* OverExpress^{TM} C41 (DE3) pLysS (Figure S 6) show a protein band at about 51 kDa as well when the expression experiment is terminated (17 h after induction) (cf. chapter

3.2.7). These samples were used as control since the expression vector pET32a is based on a T7 RNA polymerase system which is known to be leaky (cf. chapter 3.2.7).

3.2.2. NphB activity in the presence of 1,6-DHN

In chapter 3.2.1 we showed that expression of *nphB* in *E. coli* is generally possible. In order to test for functional expression, cells were harvested 17 h after induction and lysed. The cell lysate supernatant was supplemented with magnesium, GPP and 1,6-DHN, a known aromatic substrate of NphB (Kumano et al., 2008; Kuzuyama et al., 2005) (chapter 2.2.2.3). The activity assays were analyzed by LC-MS (chapter 2.2.2.6) and an extracted ion chromatogram (EIC) with *m/z* 297.18 is shown in Figure 31.
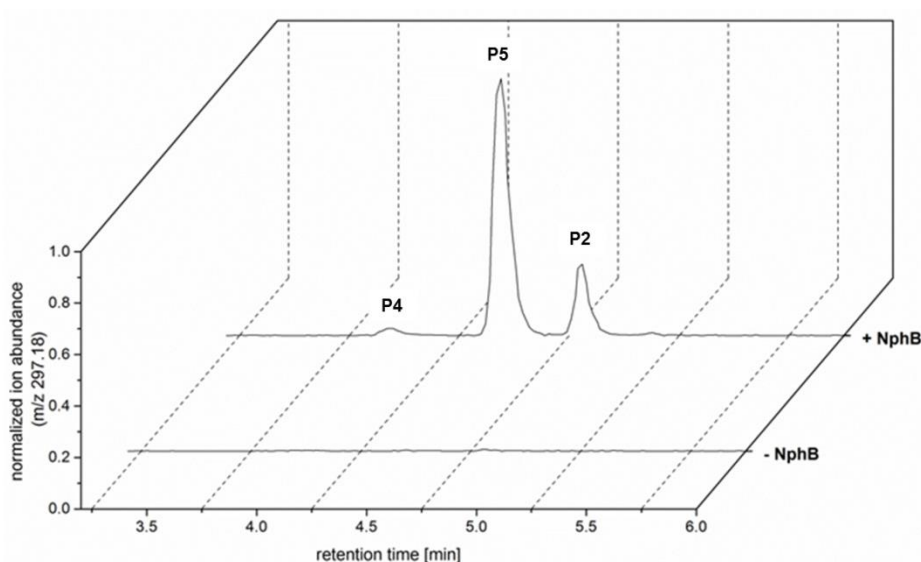

Figure 31: LC-MS analysis of geranylated 1,6-DHN produced with cell lysates of *E. coli* expressing *nphB*. Expression cultures were grown at 37 °C until $OD_{600} = 0.6$ was reached. After addition of 1 mM IPTG the cell cultures were cultivated at 25 °C for 17 h. *E. coli* lysate supernatants were incubated with magnesium, GPP and 1,6-DHN for 4 h at 37 °C. Extracted ion chromatograms (EIC) of *m/z* 297.18 of the assay products (+ NphB) and the control (- NphB) are shown. Standard substances were not available commercially.

The NphB catalyzed geranylation of 1,6-DHN with GPP as prenyl donor yielded one major product and two minor products each showing the mass of mono-geranylated 1,6-DHN (*m/z* 297.18; Figure 31 and Figure 32). According to Kumano et al. (2008) the dominant product peak belongs to 5-geranyl 1,6-DHN (P5), the second most abundant product is 2-geranyl 1,6-DHN (P2) and the third peak is caused by 4-geranyl 1,6-DHN (P4). Activity assays without enzyme did not result in product formation with a *m/z* 297.18.

According to Cui et al. (2007) 1,6-DHN exhibits substantial conformational fluctuation on the active center of NphB. The crystal structure of the prenyltransferase complexed with magnesium, GSPP and 1,6-DHN (PDB 1ZB6) shows the most favoured position of the aromatic acceptor molecule in the binding site, resulting in the formation of P5. Computational simulations indicate that the two most

abundant prenylation products P2 and P5 show different orientations of 1,6-DHN in the binding state. NphB has higher affinity towards P5 as the formation of P2 requires the aromatic acceptor to rotate within the binding site of the enzyme in order to reduce steric hindrance and to reorientate itself relative to the formed geranyl carbocation. The prenylation products P4 and P5 share the same orientation of the aromatic acceptor molecule in the binding site of NphB, but P5 has a lower free energy barrier for carbocation formation (12.6 kcal mol^{-1}) than P4 (14.1 kcal mol^{-1}) resulting in a more favourable production of P5 (Cui et al., 2007; Yang et al., 2012). Interestingly, the yields of the different geranylated products seem to be dependent on the incubation time of the activity assay (Cui et al., 2007). Nevertheless, the mechanism of NphB catalyzed geranylation of 1,6-DHN is not fully understood yet and the presented explanations are based on computer simulations. In Table 17 kinetic data for the formation of geranylated 1,6-DHN, published by Kumano et al. (2008), are presented.

Table 17: K_M values, k_{cat} values and free energy barrier for the geranylation of 1,6-DHN by NphB. K_M – Michaelis constant, k_{cat} – turnover number, k_{cat} / K_M – catalytic efficieny of NphB (Kumano et al., 2008).

Product	K_M [mM]	k_{cat} [s^{-1}] x 10^3	k_{cat} / K_M [M^{-1} s^{-1}]
2-geranyl 1,6-DHN (P2)	0.51 ± 0.08	1.10 ± 0.10	2.20
4-geranyl 1,6-DHN (P4)	1.05 ± 0.17	0.32 ± 0.02	0.30
5-geranyl 1,6-DHN (P5)	0.54 ± 0.07	4.20 ± 0.20	7.70
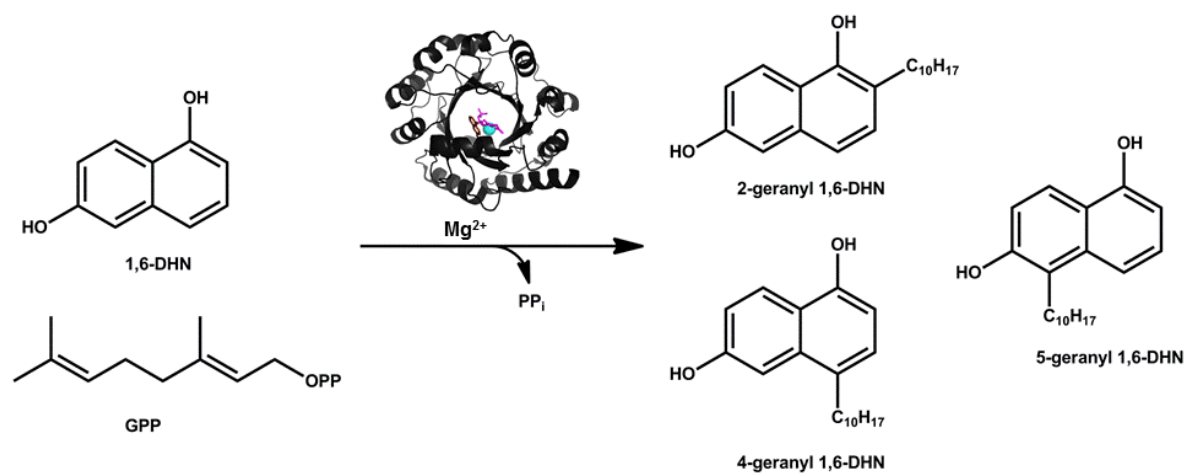
Figure 32: NphB catalyzed geranylation of 1,6-DHN. Ternary NphB complex containing Mg^{2+} (turquoise), GSPP (pink) and 1,6-DHN (brown) (PDB 1ZB6). Modified from Kumano et al. (2008) and Kuzuyama et al. (2005).

The results presented in Figure 31 indicate that functional expression of *nphB* in *E. coli* is possible.

NphB was purified from cell lysate by immobilized metal ion chromatography (IMAC; chapter 2.2.2.6). After buffer exchange the temperature optimum of NphB was determined using 1,6-DHN and GPP as substrates. The results depicted in Figure 33 indicate that the enzymatic activity is optimal over a range of 30 to 40 °C, with a maximum activity at a temperature of 37 °C.

Figure 33: Temperature profil of NphB. Purified NphB was supplemented with 5 mM magnesium chloride, 2 mM GPP and 5 mM 1,6-DHN. The activity assays were incubated at the relevant temperatures for 1 h. Product formation was analyzed by LC-MS and LC-DAD. The specific activities were calculated from technical triplicates.

3.2.3. NphB activity in the presence of olivetol and olivetolic acid
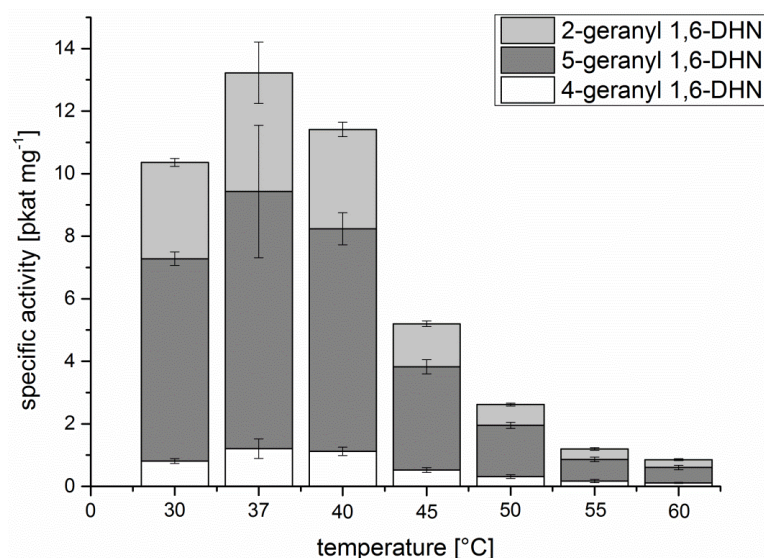
As mentioned before Kuzuyama et al. (2005) published that NphB catalyzes the geranylation of OA with GPP, but this observation is based on TLC. Thus, it remained unclear whether CBGA or a differently prenylated OA was formed. Beside the geranylation of OA, the prenylation of olivetol was tested as well as these two aromatic substrates only differ in a carboxyl group at position C2 (Figure 29). In case of NphB being able to catalyze the formation of the neutral cannabinoid cannbigerol (CBG) but not CBGA, a starting point for protein engineering of NphB would be given. Additionally, Kuzuyama et al. (2008) showed by TLC that NphB catalyzes the prenylation of olivetol.

E. coli cells expressing *nphB* were harvested, lysed and cell lysate supernantants were supplemented with magnesium, GPP and olivetol (chapter 2.2.2.3) The assays were analyzed by LC-MS (chapter 2.2.2.6) and an EIC representing *m/z* 317.24 (geranylated olivetol) is shown in Figure 34.

Using olivetol and GPP as substrates for NphB yielded one major product and one minor product. Kumano et al. (2008) showed that NphB geranylates the plant polyketide olivetol at positions C2 and C4 (Figure 35). The common name of 4-geranyl olivetol is CBG. Since no CBG standard substance was available, the product peaks in the EIC (Figure 34) were evaluated using k_{cat}/K_M values published by Kumano et al. (2008). Thereby CBG was identified as the minor product ($k_{cat}/K_M = 0.052$ M^{-1} s^{-1}) and 2-geranyl olivetol ($k_{cat}/K_M = 0.067$ M^{-1} s^{-1}) as the major product.

Figure 34: LC-MS analysis of geranylated olivetol produced with cell lysates of *E. coli* expressing *nphB*. Expression cultures were grown at 37 °C until $OD_{600} = 0.6$ was reached. After addition of 1 mM IPTG the cell cultures were cultivated at 25 °C for 17 h. The *E. coli* lysate supernatants were incubated with GPP and olivetol for 4 h at 37 °C. Extracted ion chromatograms (EIC) of *m/z* 317.24 of the assay products (+ NphB) and the control (- NphB) are shown.

Figure 35: NphB catalyzed prenylation of olivetol with GPP. Ternary NphB complex containing Mg^{2+}, GSPP and 1,6-DHN (PDB 1ZB6); pink: GSPP, brown: 1,6-DHN, turquoise: magnesium ion.

Since NphB catalyzes the prenylation of olivetol with GPP (Figure 34), it was tested whether the aromatic prenyltransferase accepts OA as aromatic substrate as well. Cell lysate expressing *nphB* was hence used for an activity assay with magnesium, GPP and OA (chapter 2.2.2.3). The assays were analyzed by LC-MS (chapter 2.2.2.6) and an EIC with *m/z* 361.23 is shown in Figure 36.

LC-MS analysis of the activity assays with OA and GPP showed the formation of two different products with *m/z* 361.23. Coeluting standard compound of CBGA was used for identification of the first peak as CBGA indicating that NphB is able to geranylate OA to form CBGA. The second peak has the same *m/z* like CBGA but a different tandem mass (MS^2) spectrum (Figure 37) suggesting that NphB geranylates OA additionally at another position than C3. Kuzuyama et al. (2005) and Kumano et

al. (2008) showed that NphB does not only catalyze carbon-carbon but also carbon-oxygen based geranylation of many hydroxyl-containing aromatic acceptors. Possible geranylation sites of OA are shown in Figure 29.
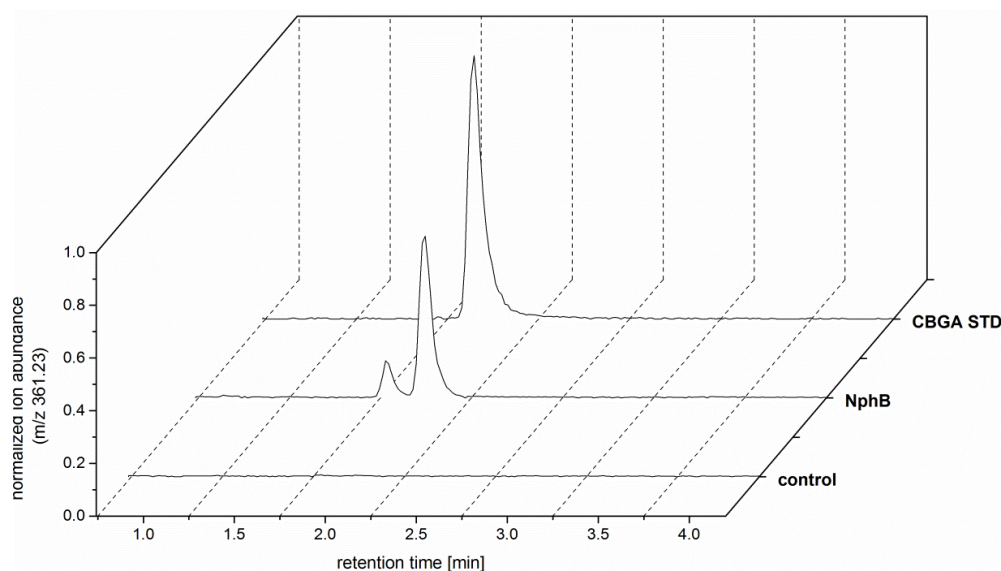
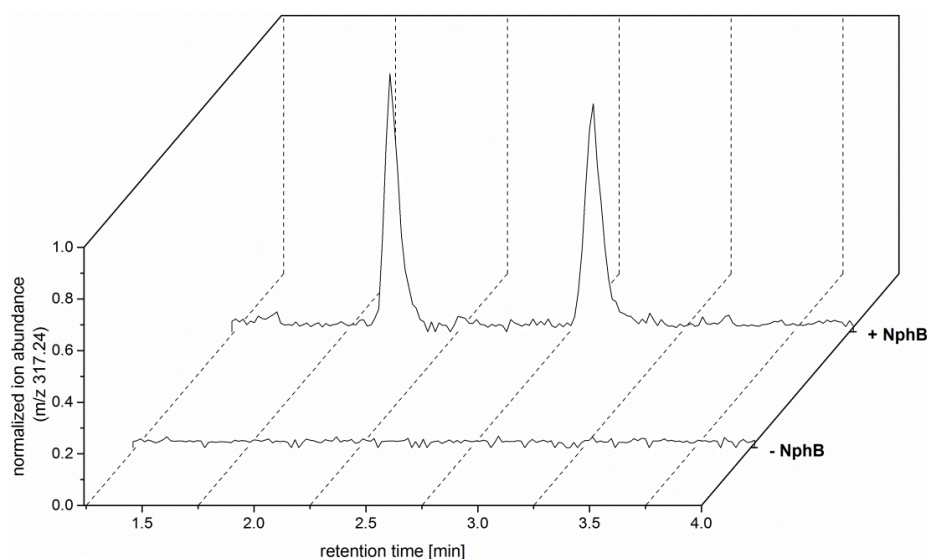**Figure 36:** LC-MS analysis of geranylated OA produced with cell lysates of *E. coli* expressing *nphB*. Expression cultures were grown at 37 °C until OD_{600} = 0.6 was reached. After addition of 1 mM IPTG the cell cultures were cultivated at 25 °C for 17 h. The *E. coli* lysate supernatants were incubated with magnesium, GPP and OA for 4 h at 37 °C. Extracted ion chromatograms (EIC) of *m/z* 361.23 of the assay products (+ NphB), a control sample (- NphB) and CBGA standard (CBGA STD) are shown.

The peak of the unknown compound was isolated by RP-LC (chapter 2.2.2.10) and analyzed by NMR (chapter 2.2.2.11). The obtained results (Table S 12) indicate that NphB preferentially geranylates OA at position 2-*O* resulting in 2-*O*-geranyl olivetolic acid (2-*O*-GOA; Figure 38).

NphB is able to catalyze the prenylation of OA with GPP to from CBGA, but simultaneously a major side-product formation (~ 85 %), 2-*O*-GOA, is detectable (Figure 38). Thus, the formation of 2-*O*-GOA presents a bottleneck in CBGA biosynthesis since most of the substrates OA and GPP are used for production of the side-product and not for the biosynthesis of the target substance. Rational protein design based on the published crystal structures of NphB (PDB 1ZCW, PDB 1ZDW, PDB 1ZB6) together with substrate docking and homology modelling might enable the change of NphB product specificity towards CBGA (chapter 3.2.5, chapter 3.2.6) (Zirpel, Degenhardt et al., 2017).
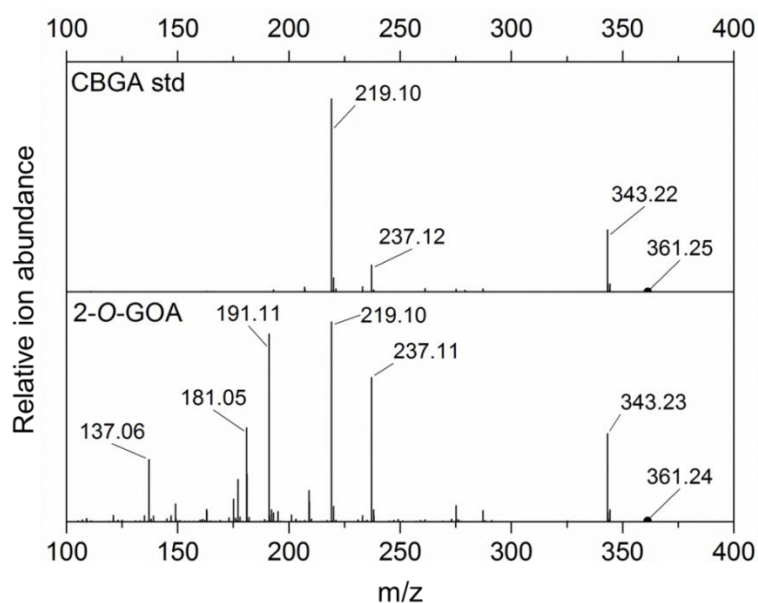
Figure 37: Tandem mass (MS[2]) spectra of the standard compound CBGA and the side-product 2-O-GOA. Modified from Zirpel, Degenhardt et al. (2017).
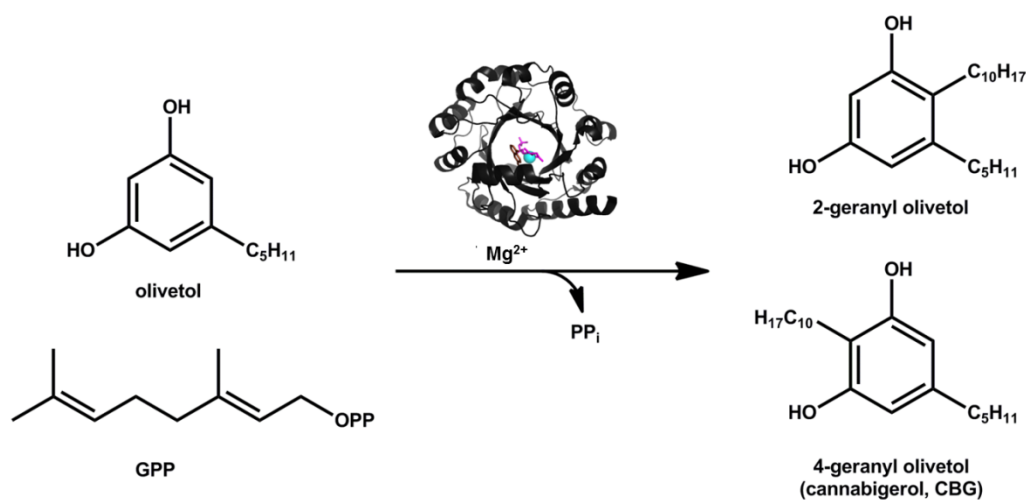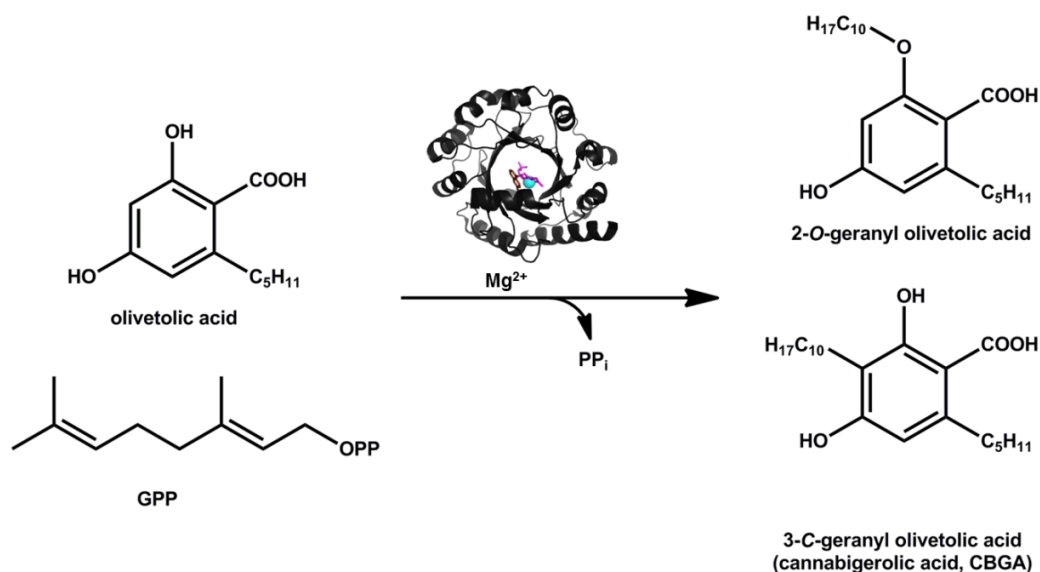

Figure 38: NphB catalyzed prenylation of OA with GPP. Ternary NphB complex containing Mg^{2+}, GSPP and 1,6-DHN (PDB 1ZB6); pink: GSPP, brown: 1,6-DHN, turquoise: magnesium ion.

3.2.4. Influence of the N-terminal TrxA tag

As mentioned in chapter 3.2.1, NphB contains a TrxA tag and a His$_6$ tag at its N-terminus. The AA sequence of the whole N-terminal tag is depicted in Table S 13 and has a calculated molecular weight of 17.34 kDa. A crystal structure of TrxA is shown in Figure 46. In order to test if the N-terminal tag affects NphB activity the target enzyme was purified by IMAC (chapter 2.2.2.6) and treated with enterokinase (chapter 2.2.2.7). Subsequent, NphB activity assays with magnesium, GPP and 1,6-DHN were performed and product formation was analyzed by LC-MS/DAD (chapter 2.2.2.8).

In Figure 39 A the coomassie stained SDS gel obtained after incubation of NphB in the presence of enterokinase is depicted. NphB without N-terminal tag (Table S 13) has a calculated molecular weight of 33.74 kDa (- tag) and a molecular mass of about 51.1 kDa when fused to the N-terminal solubilization tag (+ tag). Enterokinase light chain has a calculated mass of 31 kDa, but it should not be visible on the SDS gel because it was applied in a concentration of 0.001% (w/w) (chapter 2.2.2.7). The band between 10 and 15 kDa might represent the removed TrxA-His$_6$ tag (17 kDa). The enzyme solution was purified by IMAC (chapter 2.2.2.7, chapter 2.2.2.6) and the obtained NphB was incubated with magnesium, GPP and 1,6-DHN. The estimated specific activities are depicted in Figure 39 B. The obtained results indicate that the N-terminal tag seems to affect NphB enzyme activity since higher specific activities were obtained when using NphB without tag (NphB w/o tag). NphB with and without N-terminal solubilization tag exhibit a specific activity of approximately 0.8 nkat mg^{-1} and 2.1 nkat mg^{-1}, respectively for the formation of the major product 5-geranyl-1,6-DHN (Figure 39).

Even though, the specific activity of NphB is affected negatively by the TrxA tag (Figure 39), the solubility tag seems to be important for the functional expression of the enzyme. Until now, expression of *nphB* in the absence of the N-terminal tag has not been yielded expression detectable by SDS-PAGE (data not shown).

io
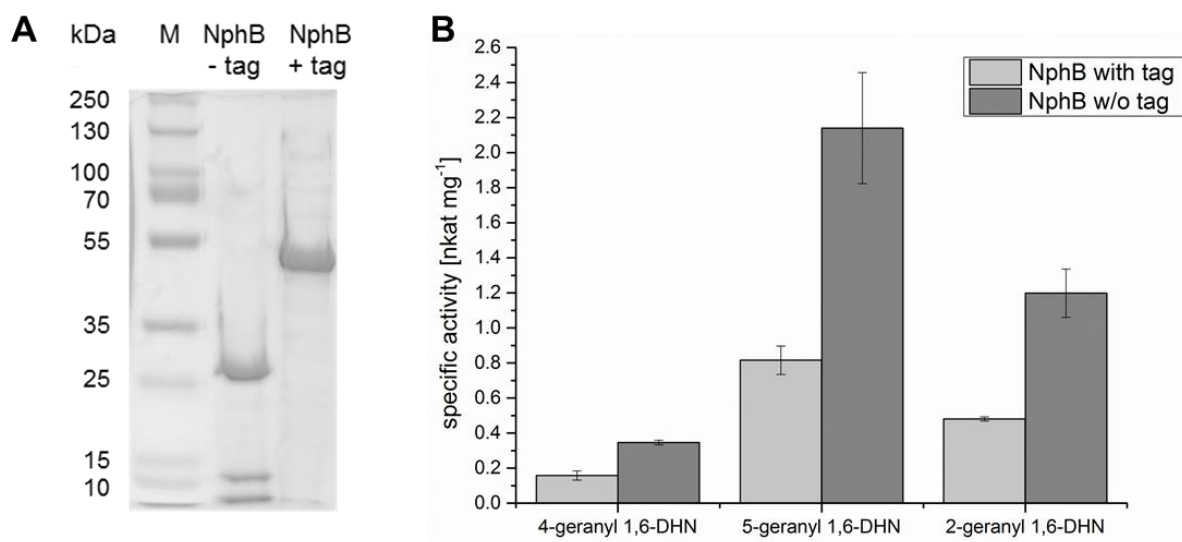
Figure 39: Specific activities of NphB with and without N-terminal TrxA tag. After cultivation NphB containing *E. coli* BL21 (DE3) cells were harvested and lysed. (A) The target protein was purified using IMAC and the N-terminal tag was removed by incubation in the presence of enterokinase. M: Thermo Scientific PageRuler Plus Prestained Protein Ladder, NphB – tag: NphB without N-terminal tag, NphB + tag: NphB with N-terminal TrxA tag. (B) For activity assays NphB with and without N-terminal tag (NphB – tag) were supplemented with magnesium, GPP and 1,6-DHN and were incubated for 40 min at 37 °C. Product formation was analyzed by LC-DAD (λ = 250 nm) and the specific activity was calculated from technical triplicates.

3.2.5. Protein engineering of NphB

Kuzuyama et al. (2005) published four NphB crystal structures each co-crystallized with different compounds (Table 2) and resolutions of 1.44 Å to 2.25 Å. An alignment of the four crystal structures is depicted in Figure 40.

The structural alignment of the four crystal structures available for NphB indicates that the 3D structures of the aromatic prenyltransferase complexed with different chemical compounds have a high similarity. The bottom of the PT barrel is coated with a short C-terminal α-helix (Figure 40 B).
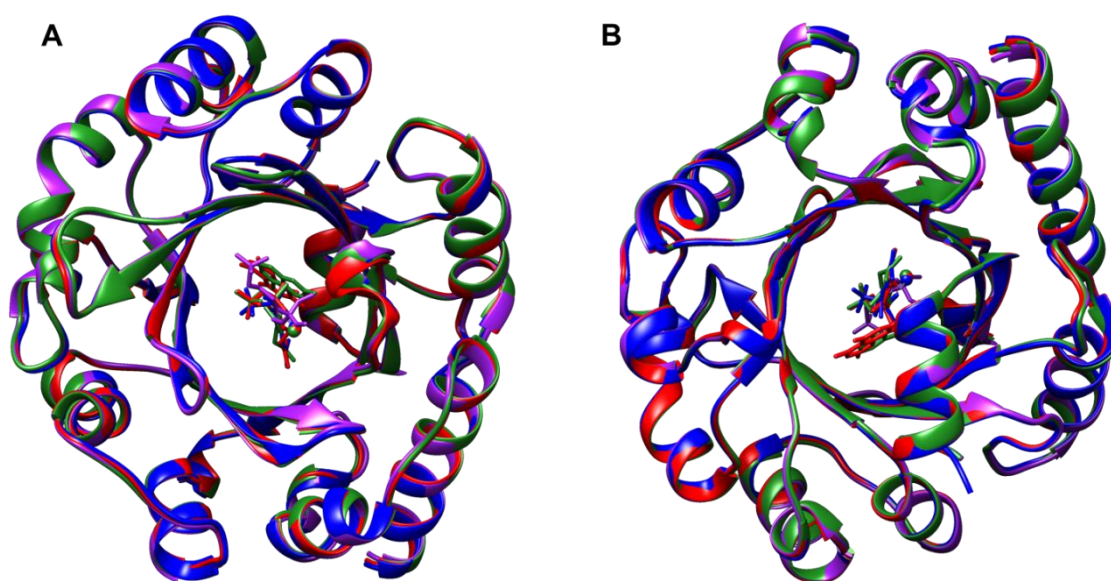
Figure 40: Structural alignment of the four crystal structures available for NphB. Green: NphB complexed with GSPP and 1,6-DHN (PDB: 1ZB6), blue: NphB complexed with GPP (PDB: 1ZCW), red: NphB complexed with GSPP and flaviolin (PDB: 1ZDW), purple: NphB complexed with TAPS (PDB: 1ZDY). A: View from the top, B: View from the bottom. The model was generated using UCSP chimera (Pettersen et al., 2004).
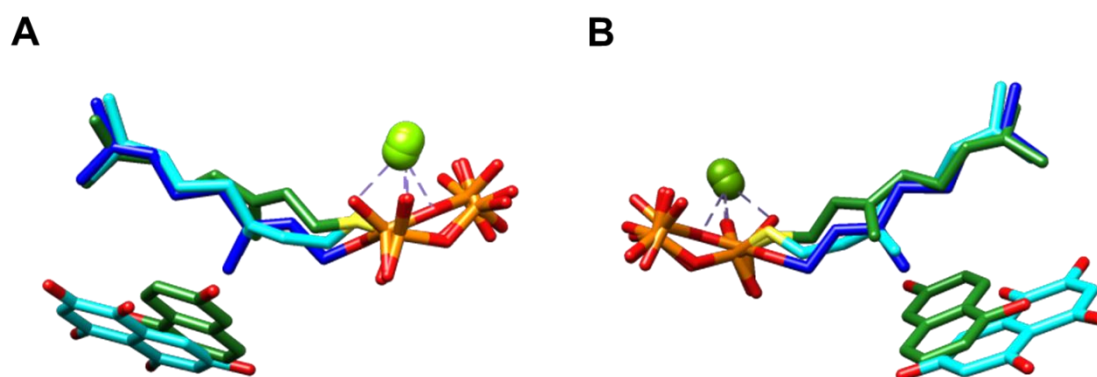
Figure 41: Alignment of different NphB 3D models co-crystallized with the co-factor Mg^{2+} and various prenyl donor as well acceptor molecules. Green: NphB complexed with GSPP and 1,6-DHN (PDB: 1ZB6), cyan: NphB complexed with GSPP and flaviolin (PDB. 1ZDW), blue: NphB complexed with GPP (PDB: 1ZCW). The spheres in both models represent magnesium ions. For a better overview only the co-crystallized substrates, coloured according to their heteroatoms, are depicted but not the crystal structures of the corresponding NphB. Images A and B show different views on the aligned substrates. The model was generated using UCSP chimera (Pettersen et al., 2004).

The alignments of the co-crystallized prenyl donor molecules (GSPP, GPP) and the magnesium ions indicate that the divalent cations and the C_{10}-molecules are coordinated in the same way in the PT-barrel (Figure 41). The diphosphate residues of GSPP and GPP, respectively, are located at the more polar end of the PT-barrel, while the geranyl tails are trackable in the neighbourhood of hydrophobic amino acid residues inside the barrel. The magnesium ion is coordinated in the upper, more hydrophilic part of the barrel and orients the GPP by diphosphate-cation interactions (Figure 40). The aromatic acceptor molecules 1,6-DHN and flaviolin are located in the same plane, but they are arranged differently (Figure 41). It is important to know, that the incubation of NphB with GPP and flaviolin yielded insignificant product amounts that were not further investigated (Kumano et al., 2008; Kuzuyama et al., 2005). However, the prenylation of 1,6-DHN resulted in the formation of three different products (chapter 3.2.2). The 3D model of NphB co-crystallized with magnesium, GSSP and 1,6-DHN represents the conditions in the active site leading to the formation of the major product 5-geranyl-1,6-DHN (Chakravorty and Merz, 2015; Cui et al., 2007; Yang et al., 2012). Structural formulars of GPP, GSPP, 1,6-DHN and flaviolin are depicted in Figure S 10.

Based on the crystal structures available, a 3D model of NphB complexed with OA (Pubchem CID 2826719) and GSPP was generated by performing homology modelling and molecular docking (PD Dr. Wolfgang Brandt, Leibniz Insitute for Plant Biochemistry, Halle (Saale), Germany). An image of the resulting computer 3D model is depicted in Figure 42.

The computer 3D model of NphB complexed with magnesium, GSPP and OA was used to propose amino acid exchanges within the active site of the enzyme that might change the enzyme's specificity towards CBGA. The different mutants were obtained by site-directed mutagenesis (chapter 2.2.2.2) and were tested in a low through-put screening platform (chapter 2.2.2.4). The results are presented in chapter 3.2.6.
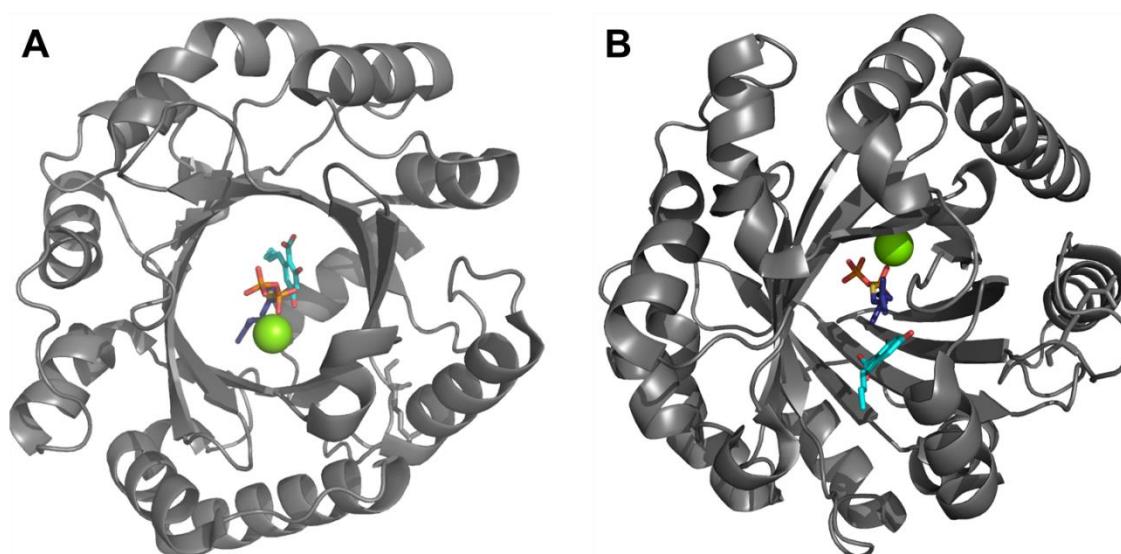

Figure 42: 3D model of NphB complexed with OA and GSPP. Green: Magnesium ion, cyan: olivetolic acid, blue: carbon chain of GSPP, orange: diphosphate of GSPP. A: View from the top, B: View from the bottom. The model was generated by PD Dr. Wolfgang Brandt (Leibniz Insitute for Plant Biochemistry, Halle (Saale), Germany).

3.2.6. Screening of different NphB mutants for high CBGA formation

In order to increase CBGA formation and to decrease production of the side-product 2-O-GOA by changing the product specificity of NphB towards CBGA, rational protein design, more precisely protein re-design, were performed (chapter 3.2.5). In order to screen the different NphB variants, a low through-put screening platform was developed using a NphB-GFP fusion protein and the RoboLector system (chapter 2.2.2.4). The C-terminal GFP was used as reporter of NphB expression levels in order to design a suitable cultivation protocol for the screening platform. The cultivation was performed with two precultures and one main culture (chapter 2.2.2.4).
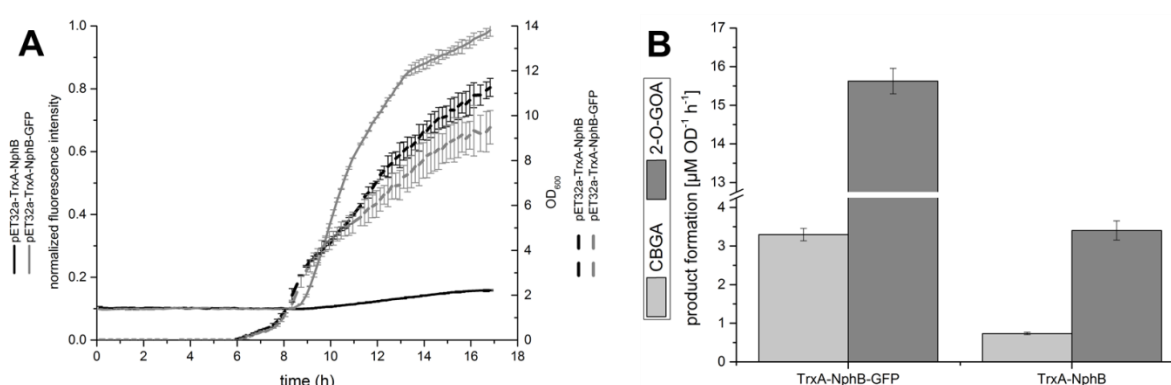
Figure 43: Cultivation of *E. coli* BL21 (DE3) containing the plasmid pET32a-TrxA-NphB-GFP or pET32a-TrxA-NphB. Cultivation was performed with two precultures and one main culture using the RoboLector system. (A) Cultivation profile: At t = 2 h the cell cultures were cooled to 25 °C. Protein expression was induced by addition of 1 mM IPTG at OD_{600} = 1.2. The graph shows the growth curves (OD_{600}) of *E. coli* BL21 (DE3) containing the two different plasmids and the normalized fluorescence intensity detected during the expression experiment. (B) Cell lysate supernatants were incubated with 5 mM magnesium chloride, 5 mM GPP and 5 mM OA at 37°C for 15 min. Assays were analyzed by LC-DAD and product formation was normalized on cell culture OD_{600} and assay incubation time. The values were obtained from biological duplicates.

The results depicted in Figure 43 A indicate that the growth behavior of *E. coli* BL21 (DE3) containing the plasmid pET32a-NphB or pET32a-NphB-GFP, respectively, is almost the same. The detected fluorescence intensities of the NphB-GFP fusion protein (Figure 43 B) show that *nphB* expression is possible using the chosen expression protocol. According to the detected fluorescence the expression of the NphB-GFP fusion protein starts within one hour after induction at OD_{600} = 1.2. Additionally, the results obtained by measuring the GFP fluorescence indicate that the *nphB* expression level was still increasing when the experiment was terminated (Figure 43 A). Therefore, it was decided to extend the expression protocol by four hours. Cell lysates were incubated with magnesium, GPP and OA and product formation was analyzed by LC-DAD (chapter 2.2.2.8). Results depicted in Figure 43 B show that CBGA and 2-O-GOA were formed, indicating that functional expression of *nphB* using the low through-put screening platform and the RoboLector system is possible. Interestingly, NphB activity assays yielded an approximately 4.5-times higher product formation rate of CBGA and 2-O-GOA for NphB fused to GFP than NphB without C-terminal expression reporter (Figure 43 B), whereas the CBGA / 2-O-GOA ratio remained constant. The effect of the GFP-tag was not further investigated since the GFP-fusion protein was only used for establishing a screening protocol (cf. chapter 3.2.7).

The final cultivation protocol presented in chapter 2.2.2.4 was used for cultivation of different NphB variants without C-terminal GFP. The mutants were obtained by site-directed mutagenesis (chapter 2.2.2.2) and cloned into *E. coli* BL21 (DE3). Expression of *nphB* wt was performed in each experiment as control. After cultivation cells were harvested and lysed (chapter 2.2.2.4). Cell lysate supernatants were supplemented with magnesium, GPP and OA and product formation was analyzed by LC-DAD/MS (chapter 2.2.2.6). The results were normalized on the CBGA formation of NphB wt and are presented in Figure 45. The AA residues within the active site of NphB that were used as targets for mutagenesis are depicted in Figure 44.
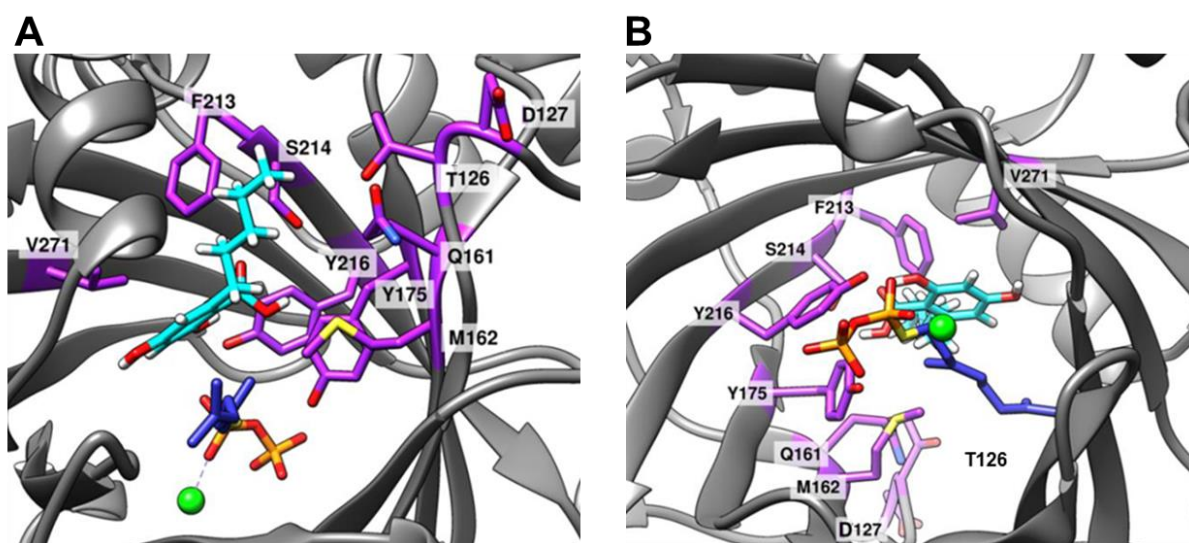
Figure 44: Amino acids within the active site of NphB. Cyan: potential position of olivetolic acid within the active site, orange: diphosphate residue of GSPP, blue: geranyl chain of GSPP, green: magnesium ion, purple: AA residues within the active site as potential targets for mutagenesis. Numbering refers to AA residues of NphB. The graphicl was generated using UCSP chimera (Pettersen et al., 2004).

The results presented in Figure 45 indicate that several NphB variants were enzymatically inactive regarding the formation of CBGA or 2-O-GOA, namely M162N, M162W, Y175A, Y175N, S214D, Y216A, M162A/V271N and M162W/Y175A. The NphB variants T126A, T126G, T126V and D127A exhibited enzyme activities comparable to NphB wt (Figure 45 A). Unfortunately, none of the tested NphB variants yielded significantly higher CBGA formation than the wildtype enzyme. Nevertheless, mutant M162K is less active than NphB wt (~12 % residual activity of NphB wt) but forms CBGA and 2-*O*-GOA in a ratio of 97% to 3%. As mentioned before the wildtype enzyme forms these products in a ratio of 15% to 85% (Zirpel, Degenhardt et al., 2017). Additionally, the mutant screening resulted in four enzyme variants that displayed a significant increase in 2-*O*-GOA formation compared to NphB wt while their CBGA formation was not enhanced, namely D127G, Q161N, F213A (Figure 45 A) and T126V/Q161N (Figure 45 B).
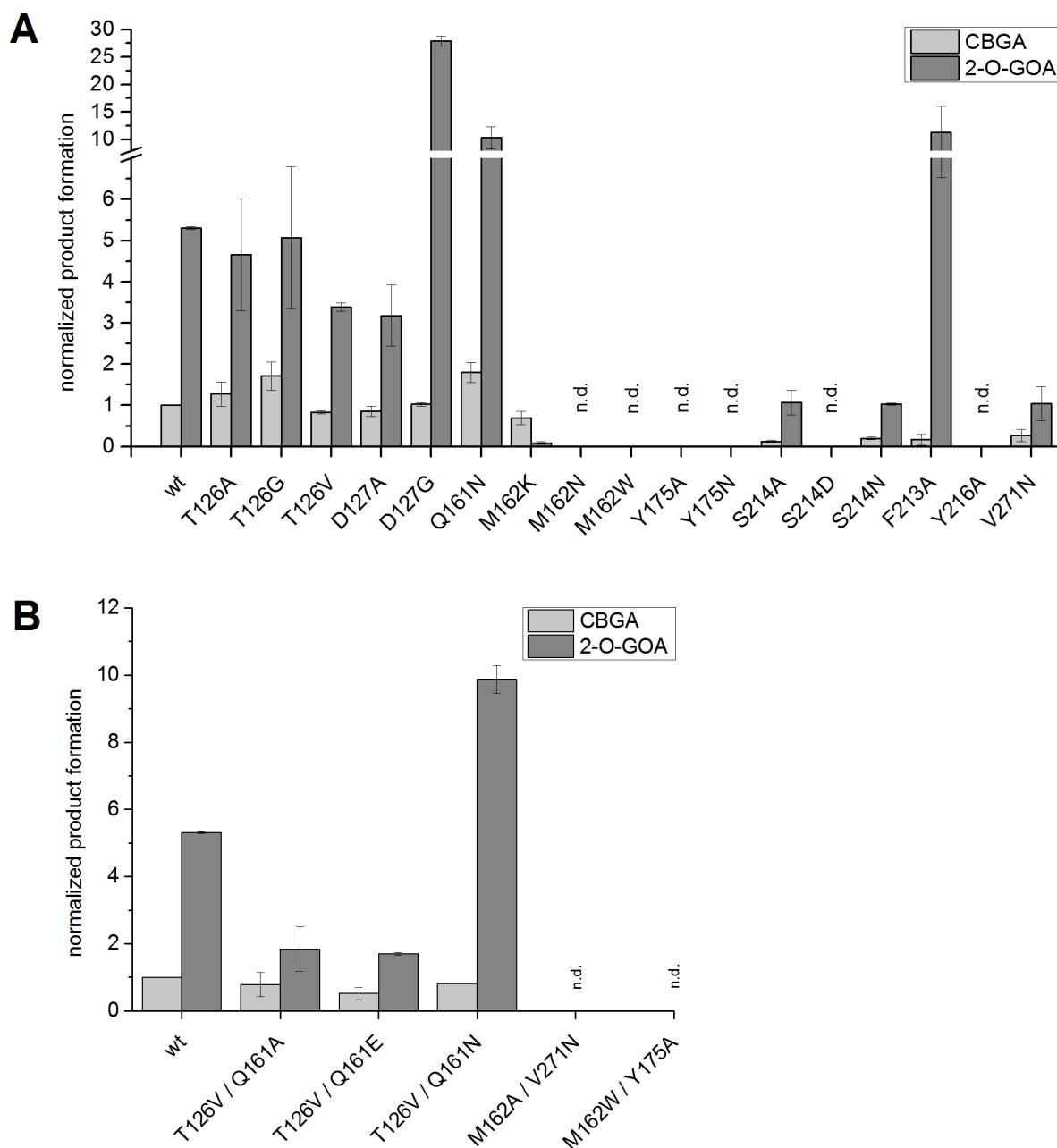
Figure 45: Screening of different NphB mutants expressed by *E. coli* BL21 (DE3). Expression cultures were grown for 2 h at 37 °C and for 19 h at 25 °C in deep-well plates; cell lysate supernatants were supplemented with magnesium, OA and GPP and were incubated for 15 min at 37°C. Product formation was analyzed by LC-DAD/MS and normalized on assay incubation time, cell culture OD_{600} and cell culture volume. The obtained results were normalized on the normalized product formation of NphB wt. A: single mutants, B: double mutants. n.d. no product formation was detectable. Data points are calculated from biological triplicates each analyzed in technical duplicates.

3.2.7. Discussion

Functional expression of the soluble, aromatic prenyltransferase NphB, first isolated from *Streptomyces* sp. strain CL190 (Kuzuyama et al., 2005), in *E. coli* was successful using various expression strains (chapter 3.2.1). The used *E. coli* strains exhibit different special features (Table S

11) that might be important for the expression of different target proteins. Different expression levels were obtained and the highest *nphB* expression level was detectable when using *E. coli* BL21 (DE3) as expression host. This strain is commonly used for high level recombinant protein expression.

As mentioned in chapter 3.2.1 the non-induced samples of *E. coli* OverExpress™ C43 (DE3) pLysS (Figure 30), *E. coli* OverExpress™ C41 (DE3) pLysS and *E. coli* Rosetta 2 (DE3) (Figure S 6) showed a protein band at about 51 kDa when the expression experiment was terminated. This non-induced expression might be caused by the leaky T7 RNA polymerase system. The *lacUV5* promoter is inducible by lactose or the lactose-mimic IPTG that is non-hydrolyzable. Gene expression is detectable even in the absence of the inducer IPTG because of lactose being present in complex medium (peptone). Leaky expression can be reduced by T7 lysozym, encoded by pLysS, that inhibits T7 RNA polymerase or by glucose being present in the medium (carbon-catabolite repression). Inhibition of basal expression might be important when toxic proteins should be expressed (Briand et al., 2016; Grossman et al., 1998). Interestingly, a protein band at about 51 kDa was detectable when using *E. coli* OverExpress™ C43 (DE3) pLysS (Figure 30) or *E. coli* OverExpress™ C41 (DE3) pLysS (Figure S 6) as expression host, even though the cells should carry a pLysS plasmid. It can not be excluded that both cell lines did not carry a pLysS plasmid or that the concentration of the desired antibiotic as selection marker was not sufficient.

The highest *nphB* expression level was obtained using *E. coli* BL21 (DE3) as expression host. Nonetheless, this level was not very high. Beside the presence of plasmid DNA, induction of protein expression by the addition of IPTG might put an additional metabolic load on *E. coli* as a down-regulation of the genes responsible for transcription, translation and energy synthesis were detectable (Haddadin and Harcum, 2005; Kosinski et al., 1992; Zhang et al., 2015). Additionally, the authors published that they obtained higher protein production levels in the absence of the non-hydrolyzable inducer. Since IPTG seems to be somehow toxic for the expression host and might lead to the formation of inclusion bodies, the change of the inducer should be considered. Results published by several research groups show that the use of lactose as inducer can yield higher protein expression levels (Bashir et al., 2015; Fruchtl et al., 2015; Kosinski et al., 1992; Wurm et al., 2016). One disadvantage of lactose is that it serves as inducer and as carbon source. Due to its metaboliziation by *E. coli* re-induction or continuous feeding of lactose as inducer might be necessary. Nevertheless, lactose is much cheaper than IPTG, does not harm the expression host and can result in increasing functional expression. Regardless whether IPTG or lactose are used as inducer of protein expression, the results of *nphB* expression in different *E. coli* strains (Figure 30, Figure S 6, Figure S 7) indicate that use of non-induced samples is necessary in order to evaluate the expression level of NphB. Nevertheless, under the tested expression conditions, the highest *nphB* expression level was obtained using *E. coli* BL21 (DE3) as expression host. Therefore, this strain was used as a chassis strain for the recombinant production of NphB.

However, several aspects of our recombinant expression system for the production of NphB have to be further investigated in the future.

First of all, NphB exhibits an optimum activity over a temperatur range from 30 to 40 °C, with a maximum activity at 37 °C (Figure 33). The activity tests were performed with 1,6-DHN as aromatic acceptor molecule. Additionally, the enzyme contained a TrxA tag at its N-terminus. As shown in Figure 39, the enzymatic activity of NphB seems to be affected by this solubility tag. With regard to the optimization of NphB towards CBGA formation, the determination of the temperature optimum should be repeated with OA as aromatic acceptor molecule and with NphB without N-terminal TrxA tag. Subsequently, the pH optimum of the enzyme and the optimal substrate concentrations of the prenyl donor and acceptor should be investigated in order to prevent toxification. Results published by Xiao et al. (2009) indicate that GPP concentrations higher than 0.2 mM, when using 1,6-DHN as aromatic substrate, might yield in decreasing reaction rates and that the reactions rate is maximal at a GPP concentration of approximately 0.1 mM. In regard to the optimal OA concentration and the generation of a cannabinoid producing microorganism (chapter 1.1) it has to be kept in mind that higher OA concentrations (> 5 mM) might be toxic for the expression host (Zirpel, Degenhardt et al., 2017).

Secondly, expression in *E. coli* was so far only performed using the vector pET32a that carrys the *bla* gene encoding β-lactamase. The enzyme hydrolyses the β-lactam amide bond of β-lactam antibiotics like ampicillin or carbenicillin and thus, *E. coli* cells containing the plasmid are resistant against these antibiotics (Frère, 1995). One disadvantage of β-lactamase is that it is secreted by bacteria, so that the selection marker present in the culture medium is inactivated, resulting in a decreasing selection pressure (Korpimäki et al., 2003). Therefore, the change of the selection marker used for the expression of *nphB* in *E. coli* should be taken into consideration, especially with regard to fermentation. Additionally, higher expression levels might be obtained if the selection pressure is more stable. Possible alternatives for β-lactam antibiotics might be chloramphenicol, kanamycin and tetracycline. An advantage of the latter antibiotic over the former ones is the resistance machinery that is not based on degradation but on inhibition of protein translation and efflux. Therefore, the tetracycline concentration might be stable in the medium over the whole time of cell cultivation allowing for a prolonged, more robust and effective protein production (Korpimäki et al., 2003; Nguyen et al., 2014).

Furthermore, we were only able to functionally express NphB in *E. coli* by using a N-terminal TrxA-tag. However, results depicted in Figure 39 B indicate that the enzymatic activity of NphB is affected by the N-terminal TrxA tag with specific enzyme activities increasing when removing the tag. The solubility tag seems to be important for the functional expression of the enzyme as using the expression vector pET28a did not yield in NphB expression at all (data not shown). TrxA consists of 109 AA (Table S 13) and the crystal structure is depiced in Figure 46. NphB and the solubility tag are separated from each other by a "linker" consisting of 52 AA (Table S 13). The active site of NphB is localized in the PT-barrel consisting of antiparallel β-sheets (Figure 46). This stable tertiary structure is probably not affected by the solubility tag fused to the N-terminus of NphB. Results published by Kumano et al. (2008) suggest that the aromatic acceptor molecule is only able to bind to the enzyme if GPP binds first. On this account there would be only one possible entrance for the aromatic substrate to the active site of NphB that might be blocked from time to time by the flexible N-terminal solubility tag

(Figure 46) resulting in a lower specific activity than NphB without N-terminal tag. The TrxA tag can be removed by digestion of the fusion protein with enterokinase, but this is associated with high cost if the specific protease has to be bought commercially. In 2013, Skala et al. published a protocol for the expression and purification of enterokinase from *E. coli.*
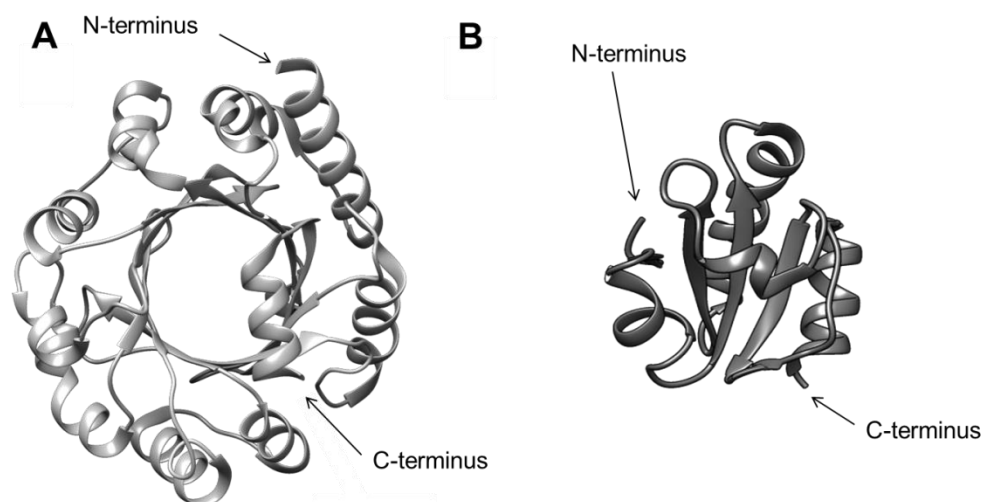
Figure 46: Crystal structures of NphB and TrxA. Left image: 3D model of NphB co-crystallized with magnesium (green), GSPP (magenta) and 1,6-DHN (red) (PDB 1ZB6). The crystal structure displays the amino acids 3 to 303 with a resolution of 1.95 Å. Right image: 3D model of TrxA wildtype (chain A; PDB 2H6X). The crystal structure displays the amino acids 1 to 108 with a resolution of 2.6 Å.

During the establishment of our low-throughput screening system, we also found that NphB activity assays yielded four times higher product formation if NphB was fused to GFP compared to NphB without C-terminal expression reporter (Figure 43). This difference might be caused by an interference of the C-terminal GFP-tag on the folding of NphB and thus on the function / structure of the enzyme, since GFP is a quite bulky protein itself (236 AA; UniProtKB P42212; PDB 1EMA) (Figure 47). However, another explanation for these results might be that the C-terminal GFP protects the NphB C-terminus and thus more functional protein is present within the cells (Lin et al., 2018). Although the effect of the GFP-tag was not further investigated since the GFP-fusion protein was only used for establishing a screening protocol, a strategy to protect NphB from C-terminal protein degradation should be investigated.
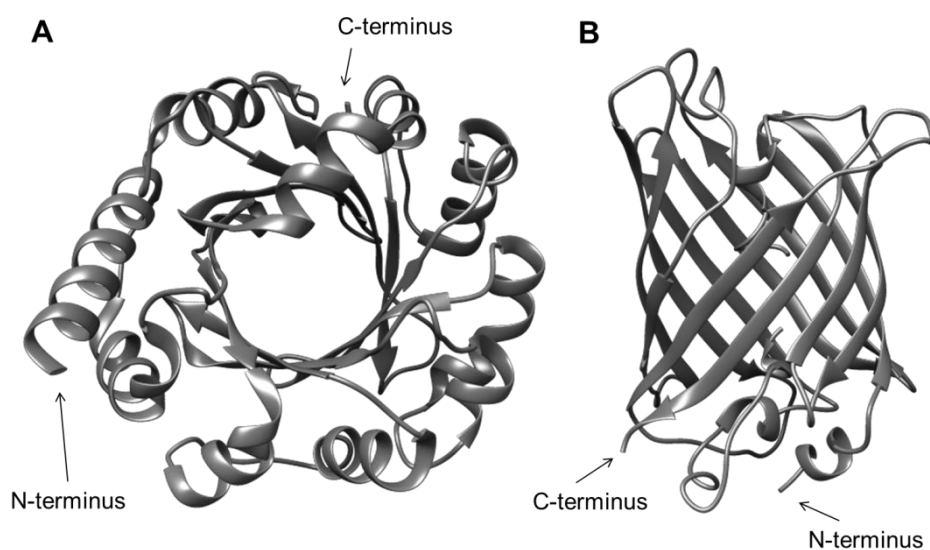
Figure 47: Crystal structures of NphB and GFP.A: NphB (PDB 1ZB6), B: green fluorescent protein (GFP) (PDB 1EMA). The model was generated using UCSP chimera (Pettersen et al., 2004).

Previous studies of Kuzuyama et al. (2005) indicated that NphB catalyzes the prenylation of OA, but product formation was only confirmed by TLC. Here, we could show for the first time that the aromatic prenyltransferase is able to geranylate OA yielding in the formation of CBGA (Figure 36). However, a major side-product formation (~ 85 %) is detectable resulting in the formation of 2-*O*-GOA (Figure 38) (Zirpel, Degenhardt et al., 2017). The obtained results show that NphB catalyzes carbon-carbon and carbon-oxygen based geranlyation of a hydroxyl-containing aromatic acceptor molecule (Kumano et al., 2008; Kuzuyama et al., 2005). Nevertheless, the production of 2-*O*-GOA presents a bottleneck for cannabinoid production in the same host since CBGA is the precursor for THCA formation catalyzed by THCAS (Sirikantaramas et al., 2007, 2004, Taura et al., 2007b, 1995). We therefore pursued a rational protein engineering approach to enhance NphB specificity towards CBGA (chapter 3.2.5, chapter 3.2.6) (Zirpel, Degenhardt et al., 2017). Until now, however, no 3D model of NphB co-crystallized with magnesium, OA and GSPP is available. Thus, a protein structural model was generated (Figure 42) and used as basis for mutational studies of NphB.

With our rational mutagenesis approach we tried to investigate different aspects of the NphB catalyzed prenylation of OA.

Based on the assumptions of Yang et al. (2012) and Kuzuyama et al. (2005), several single mutants of NphB were generated to analyze the influence of amino acid residues on the GPP coordination and binding within the active site (Figure 44: M162K, M162N, M162W, Y175A, Y175N, Y216A). NphB variants Y175A, Y175N and Y216A were inactive in our experiments aligning with the suggestions of Yang et al. (2012) and Kuzuyama et al. (2005) that tyrosines at these positions are involved in GPP binding by hydrogen bond formation as well as stabilizing the carbocation during catalysis. The absent enzyme activity of these variants indicate that these amino acids are indeed important for the positioning of GPP within the active center of NphB. Methionine at position 162 is supposedly involved

in holding the geranyl side chain of GPP in place for catalysis. By altering the amino acid residue at this position, we aimed for a disposition of GPP and subsequently creating more space for OA to rearrange itself within the active site in a position favoring 4-C-geranylation (= CBGA). However, generated NphB variants M162N and M162W showed no detectable enzyme activity. The exchanges of the hydrophobic methionine against the hydrophobic tryptophan (M162W) or polar asparagine (M162N) might have either lead to blocking of the entrance of the active site or mispositioning of GPP, hence preventing correct substrate binding for catalysis. On the contrary, results obtained for the variant M162K supported our intentions as it exerted an altered product specifity towards CBGA production (~89% CBGA/~11% 2-O-GOA). Unfortunately, the product formation by this enzyme was strongly decreased compared to the NphB wt (~12% residual activity). Compared to methionine, lysine owns one CH_2 group more and has a polar amino group. Their spatial needs are similar which might have caused enzyme activity in contrast to the other two M162 variants. An exchange of several amino acids within the active site that rest against the carbon chain of GPP might be necessary to allow for a rearrangement of the substrates.

A second set of amino acid residues were targets for amino acid exchanges in order to optimize positioning of OA for production of CBGA. These approaches aimed to enable spatial orientation of OA within the active site for a 4-C-geranylation either upon holding the carboxy group of OA in place for catalysis (Figure 44; M162K, M162N, M162W, Y175A, Y175N, M162W/Y175A, S214A, S214D,) or upon creating more space for OA to reach close proximity to GPP (Figure 44: T126A, T126G, T126V, D127A, D127G, Y175A, F213A, V271N, T126V/Q161A, T126V/Q161E, T126V/Q161N, M162A/V271N).

The amino acid S214 might facilitate stabilization of the aromatic acceptor molecule OA. Here, we tested three different amino acid exchanges, S214A, S214N and S214D. The former two NphB variants were several fold less active than NphB wt and the latter one did exert no enzymatic activity at all (Figure 45). While alanine needs less space in the active center, asparagine requires more space than serine. Interestingly, both NphB variants exhibit the same product formation. Serine and asparagine belong to the group of polar amino acids with uncharged side chains, while aspartic acid has a negatively charged side chain. Since the latter two amino acids are of similar size, the negative charge of S214D most likely causes non-detectable enzyme activity. While S214N should stabilize OA within the binding site by forming a hydrogen bond with the carboxy group of OA, it was intended that S214D would bring OA closer to GPP and prevent the access to 2-OH at the same time. The obtained results, however, indicate that none of the tested S214 variants help to geranylate OA preferrently at C3 in order to form CBGA (Figure 29).

Another NphB variant that is less active than the wildtype enzyme is V271N. On this account, the exchange of an amino acid with a hydrophobic side chain to an amino acid with a more hydrophilic side chain did not help to bind OA in the binding site of NphB. The enzyme variants T126A, T126G, T126V as well as D127A exhibit similar product formation rates like NphB wt. T126 and D127 are localized in a loop at the bottom of the enzyme (Figure 44, Figure 42). Based on this, it can be assumed that the replacement of aspartic acid and threonine by amino acids with smaller side chains

did not affect the conformation of OA within the active site of NphB. However, variant D127G showed about 6-fold increased 2-O-GOA formation compared to wt enzyme, contradicting the results of the other enzyme variants with exchanges at position 126 and 127. Therefore, this result should be considered with care and examination of variant D127G should be repeated including sequencing of the genetic construct before drawing false conclusions.

NphB variants F213A and Q161N both exhibit a 2-fold increased 2-O-GOA product formation rate compared to NphB wt. However, while the product ratio of CBGA/2-O-GOA is not changed for variant Q161N, variant F213A only produces residual amounts of CBGA (2% CBGA/ 98% 2-O-GOA). The exchange from phenylalanine to alanine might alter the spatial orientation of OA in the binding site to that extent that geranylation of 2-O-GOA is more preferreable than in the wildtype enzyme.

The screening of the five NphB double mutants resulted in two enzymatically inactive variants (M162A/V271N, M162W/Y175A) and one mutant with a significantly increased 2-O-GOA product formation (T126V/Q161N). Unfortunately, compared to the wildtype enzyme, none of the tested NphB variants showed an improved product specifity towards CBGA production while preserving enzymatic activites of NphB wt (Figure 45). However, the CBGA/2-O-GOA ratios of variants T126V/Q161E and T126V/Q161N changed in the direction of CBGA.

In summary, none of the tested NphB variants showed a significant change of the product specificity towards CBGA while preserving at least the enzymatic activities of NphB wt.

Additionally, the results obtained for the amino acids M162, Y175 and Y216 indicate that these amino acids are important for the geranylation of the aromatic acceptor molecule, as suggested by Kuzuyama et al. (2005) and Yang et al. (2012).

Even though some of the NphB variants exhibited a complete lack of enzymatic activity, it can not be excluded that CBGA / 2-O-GOA formation would be detectable using different cultivation conditions. With the used screening system conclusion could only be made about whether functional expression was obtained, but not about the expression level received. Former expression experiments in shaking flasks (data not shown; [b], [c]) indicated that functional expression of *nphB* variants was not always feasible when using the same OD_{600} for induction. Furthermore, the RoboLector system is only able to detect optical densities which are greater than one. For this reason induction of gene expression below that value is not feasible, which might be necessary for expression of some NphB mutants. Therefore, it is possible that functional expression of some of the variants and higher expression levels might be accessible when using a different expression protocol. Additionally, it has to be kept in mind that the used screening system does not enable conlusions about whether changes in the enyzme activity or the expression level yielded shifts in CBGA and 2-O-GOA formation. Induction of the gene expression at a different optical density of the expression culture might affect the expression rate and thus result in different enzyme activities. However, the low throughput screening system was developed in order to identify NphB variants with altered product specifities in the first place and not to

detect enzymes with higher activities. Nevertheless, the established screening system helps to save time and money in order to identify promising NphB variants.

Until now, only a protein structural model of NphB co-crystallized with OA and GPP is available (Figure 42). In order to get more precise informations about the localization of OA and GPP in the binding site of NphB a 3D model obtained by crystallography is necessary. The electron density of 1,6-DHN in the NphB crystal structures available indicate significant conformational fluctuation of the substrate within the active site of NphB (Cui et al., 2007). The 3D model of NphB co-crystallized with GSPP and 1,6DHN (PDB 1ZB6) only shows the orientation of 1,6-DHN within the active site which results in the formation of the major product 5-geranyl-1,6-DHN (Figure 31, Figure 32). In order to obtain the two minor products, 2-geranyl-1,6-DHN and 4-geranyl-1,6-DHN, the aromatic acceptor molecule has to rotate freely (Cui et al., 2007; Yang et al., 2012). Therefore, co-crystallization of NphB wt and for example NphB M162K with OA and GSPP would be useful to analyze the conformation of OA within the active site since the wildtype enzyme produces mainly 2-O-GOA and M162K preferably catalyzes the formation of CBGA. Thus, the two possible conformations of OA within the active site of NphB would be obtained.

3.3. Expression of *nphB* and *thcas* in *Saccharomyces cerevisiae*

As shown in chapter 3.2.3 *nphB* expressed in *E. coli* is able to catalyze the prenylation of OA by GPP to form CBGA. Keeping in mind that the overall goal is the expression of the genes of the whole THCA biosynthetic pathway in yeast, expression of different NphB coding sequences in *S. cerevisiae* cells was tested (chapter 3.3.1). Until now, the NphB activity assays have been performed with cell lysate supernantants or purified enzyme. In order to reduce costs and time the use of whole cells for biocatalysis was investigated (chapter 3.3.2). THCAS catalyzes the final step of THCA biosynthesis by oxidatively cyclizing the monoterpene moiety of CBGA, resulting in the formation of THCA (chapter 1.2.3). Since NphB is able to produce CBGA, we tried to express the soluble aromatic prenyltransferase and THCAS simultaneously in *S. cerevisiae* (chapter 3.3.3).

3.3.1. Test of different NphB coding sequences

Functional expression of *nphB* in *S. cerevisiae* (SC; Table 7) was tested using different NphB coding sequences cloned into the yeast expression vector pDionysos (Stehle et al., 2008) (chapter 2.2.3.2). First different variants of *E. coli* codon usage-optimized sequences were tested: (i) NphB (pDionysos_eN), (ii) NphB with a N-terminal TrxA tag (pDionysos_TeN), (iii) NphB with a codon exchange of Arg211 (pDionysos_eNx), and (iv) NphB with a N-terminal TrxA tag and a codon exchange of Arg211 (pDionysos_TeNx).

In *E. coli* the TrxA tag is used to enhance the solubility of the protein of interest. As mentioned in chapter 3.2.7, the solubility tag has a crucial effect on functional expression of NphB in *E. coli*. The graphical codon usage analyser version 2.0 (http://gcua.schoedl.de; 01.04.2018) (Fuhrmann et al., 2004) was used to analyze the codon usage of the *E. coli* codon usage optimized coding sequence of NphB. The algorithm of the web-tool matches each triplet position against the codon usage table of *S. cerevisiae.* The results (Figure 48) show that the used NphB sequence contains six different codons that do not occur very often in *S. cerevisiae*. The red bar at amino acid position 211 (Arg211) indicates that the corresponding codon is used less than 10%, which means that it occurs only rarely in yeast and thus the responsible tRNAs are probably not very common in *S. cerevisiae*. Amino acids encoded by codons that are used less than 20% are highlighted in green (Arg9, Lys169, Arg228, Arg268, Arg296). The rarely used codon for Arg211 was changed by site-directed mutagenesis to a codon which is more common in *S. cerevisiae* in order to test if the codon-usage is effecting functional expression of *nphB* in yeast. A more frequent codon for Arg was identified by a comparison of the codon usage tables of *E. coli* and *S cerevisiae* (Figure S 13).
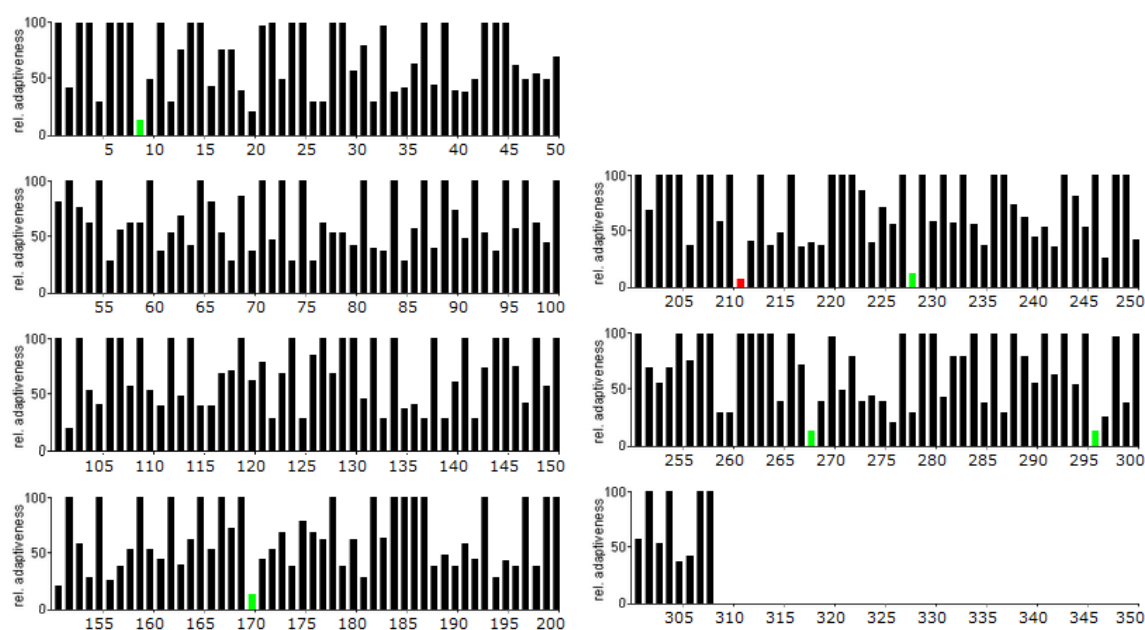
Figure 48: Graphical codon usage analysis of *E. coli* codon usage optimized NphB. The image was obtained using the web-tool "graphical codon usage analyser" version 2.0 (http://gcua.schoedl.de) (Fuhrmann et al., 2004). Codons that are used less than 20% are highlighted in green, codons that are used less than 10% are highlighted in red. The used NphB coding sequence was compared with the codon table of *S. cerevisiae*.

SC2 cells carrying plasmids with the different NphB coding sequences (chapter 2.2.3.2, Table 9) under the control of the inducible GAL1 promoter were cultivated for 48 h at 15 °C (chapter 2.2.3.5). Sampes for activity assays and OD_{600} measurements were taken every 24 h. Formation of CBGA was detected by LC-MS (chapter 2.2.3.6) after supplementing cell lysate supernatants with GPP and OA (chapter 2.2.3.5). Unfortunately, none of the tested NphB sequences yielded the formation of CBGA (*m/z* 361.23; data not shown).

Based on this, different variants of *S. cerevisiae* codon-usage optimized sequences were tested: (i) codon-usage optimization performed by GeneArt (Life Technologies; Regensburg, Germany) (pDionysos_N*), and (ii) codon-usage optimized sequence of (i) without lysines (pDionysos_N(arg)*) which are located on the α-helices surrounding the PT-barrel and thus are exposed to the solvent (Figure 49). All eligible lysines were replaced by arginines (Table S 17) in order to decrease ubiquitination (Rodriguez et al., 2000a; Worthington and Carbonetti, 2007a) that directs NphB to degradation by the proteasome.

SC2_N* and SC2_N(arg)* (chapter 2.2.3.2, Table 9) were cultivated for 48 h at 15 °C. Samples for activity assays and OD_{600} measurements were taken every 24 h. Yeast cell lysates were supplemented with GPP and OA and product formation was analyzed by LC-MS (chapter 2.2.3.6). The results depicted in Figure 50 and Figure S 14 indicate, that *nphB* was functionally expressed in *S. cerevisiae*. As shown in chapter 3.2.3, CBGA and the major side-product 2-*O*-GOA (~ 85 %) were formed. The product formation rate of SC2_N* expressing NphB wt, is approximately two times higher than the product formation rate of SC2_N(arg)*. Nevertheless, both constructs yield product formation within the first 24 h after induction of gene expression.
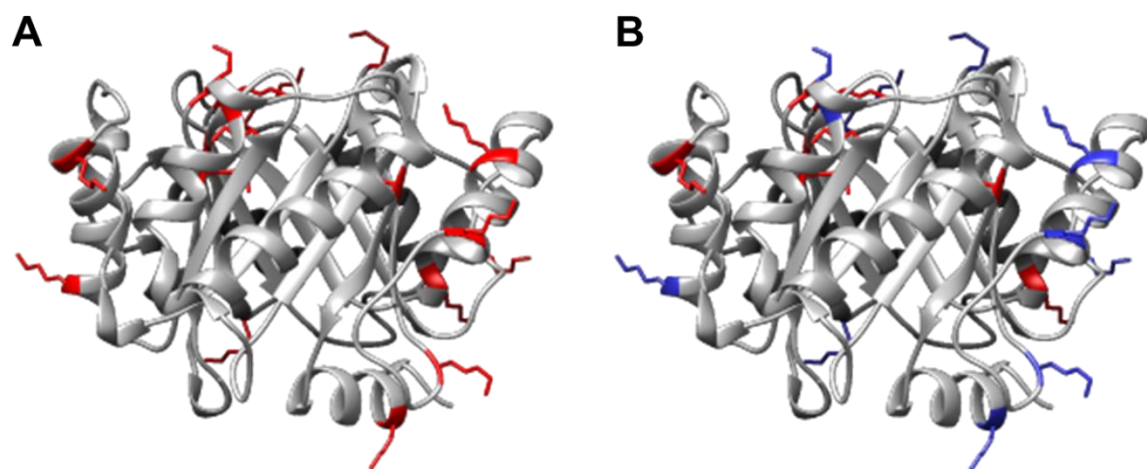
Figure 49: Structural representation of NphB. (A) All lysines present in the amino acid sequence of NphB are labeled in red (pDio_N*). (B) Lysines that were changed to arginine on the surface of NphB (pDio_N(arg)*) are labeled in blue and lysines that were not changed are labeled in red. The model was generated using UCSP chimera (Pettersen et al., 2004).

S. cerevisiae uses fructose and glucose as the preferred carbon sources, with glucose being the primary source of energy. If one of these sugars is available for the cells, the expression of a large number of genes responsible for the utilization of alternative carbon sources like galactose is repressed. This phenomenon is called "carbon catabolite repression" or "catabolite repression". It is worthy to note that catabolite repression occurs more dominantly in the presence of glucose than in the presence of fructose (Gancedo, 1998; Hohmann et al., 1999). One inhibitor of the *GAL* genes in the presence of glucose is the transcription factor MIG1. This zinc finger protein on the one hand represses the activator (GAL4) of the *GAL* genes, and on the other hand binds to the GAL1 promoter to prevent transcription. Only upon glucose concentrations in the culture medium dropping below 0.2%, the uptake of the gene expression inducer galactose by *S. cerevisiae* and subsequently the expression of the *GAL* genes is enabled (Johnston et al., 1994; Kayikci and Nielsen, 2015; Nehlin et al., 1991). In order to reduce the effect of carbon catabolite repression on the expression of *nphB*, we decided to knockout the *mig1* binding sites (upstream repression sequence (URS) A, URS B and URS C; Figure S 11) of the GAL1 promoter in the utilized expression vector pDionysos by site-directed mutagenesis (chapter 2.2.3.2). The resulting plasmids pDionysos_N(arg) and pDionysos_N were transformed into SC2 cells and used for expression experiments. Procedure of the expression cultivation and the activity assays are described above. Product formation was analyzed by LC-MS (chapter 2.2.3.6) and the results are depicted in Figure 50.

The results depicted in Figure 50 indicate that the knockout of the *mig1* binding sites (Figure S 11) of the GAL1 promoter did not effect the product formation rate of SC2_N(arg) and SC2_N in the same way. A comparison of SC2_N(arg)* and SC2_N(arg) shows that the reduction of the carbon catabolite repression did not result in higher expression levels and thus an increased product formation of CBGA and 2-*O*-GOA. In contrast to this, the expression levels of NphB wildtype increased using the mutated

pGAL1 (SC2_N), resulting in a significantly higher CBGA and 2-O-GOA production rate. The highest CBGA production rate of SC2_N was 36 pmol mL^{-1} OD^{-1} h^{-1}, while 128 pmol mL^{-1} OD^{-1} h^{-1} 2-O-GOA were formed.
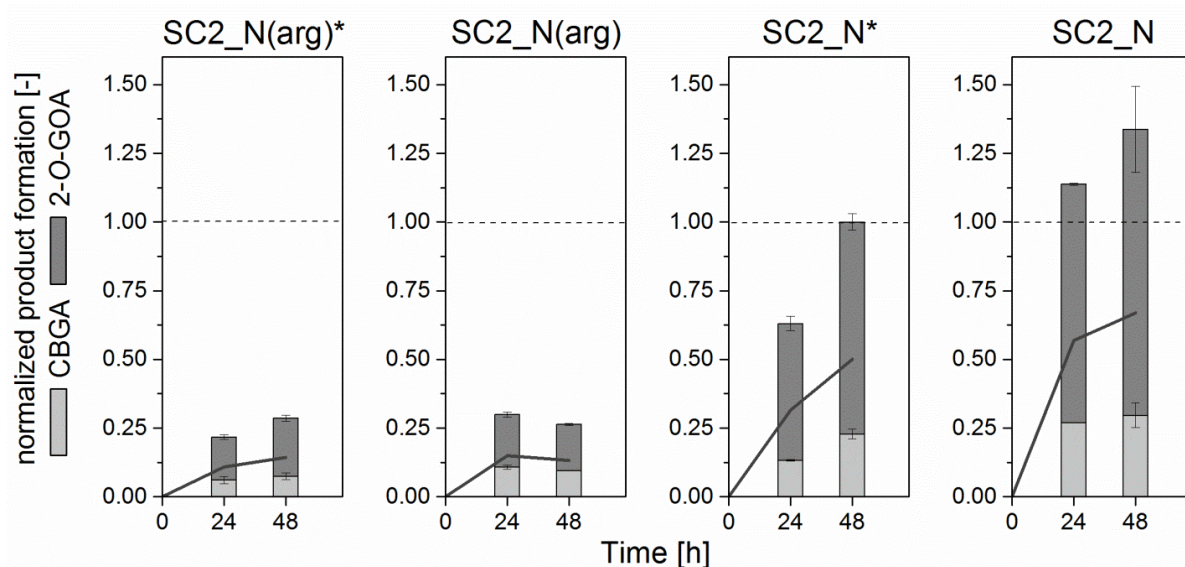
Figure 50: Product formation of SC2 cells expressing yeast codon-usage optimized *nphB*. N(arg) – NphB with amino acid exchanges of lysines, that are exposed to the solvent, by argenines, N – NphB wildtype; N(arg)*, N* - use of the wildtype GAL1 promoter. Line - visualization of the mean value of the overall production of CBGA and 2-O-GOA at a certain time point. Expression cultures were inoculated at OD$_{600}$ of 0.5 and incubated for 48 h at 15 °C and 200 rpm. Cell lysate supernatants were incubated with 5 mM magnesium chloride, 2 mM GPP and 5 mM OA (4 h, 37 °C) to determine 2-O-GOA and CBGA formation. Activity assays were analyzed by LC-MS. Product formations were normalized on cell culture OD$_{600}$, cell culture volume, assay incubation time and the maximum product formation of SC2_N*.

The expression experiments using various NphB coding sequences resulted in the highest product formation rates using SC2_N. Therefore, the vector pDionysos_N was used for further experiments.

3.3.2. Whole cell assay

Until now, NphB activity assays were always performed with yeast cell lysates. In order to simplify activity assays, we tested whether *S. cerevisiae* cells are able to take up OA and GPP from the cultivation medium. Initial tests indicated that OA concentrations of up to 5 mM and a final DMSO concentration of 1% are not toxic to *S. cerevisiae* cells (data not shown; [d]). SC_N cells were cultivated for 72 h in complex medium, before they were harvested and resuspended in assay buffer (chapter 2.2.3.4). The cells were (i) either lysed and applied in an NphB activity assay to test if *nphB* was functionally expressed or (ii) used for a whole cell assay by supplementing 3 mM OA and 3 mM GPP. After incubation the samples were prepared for LC-MS analysis (chapter 2.2.3.4, chapter 2.2.3.6).

The results shown in Figure 51 indicate that *nphB* was functionally expressed since formation of CBGA and 2-O-GOA was detectable using cell lysate supernatant. However, no product formation

with an *m/z* 361.23 was detected in the extracts of the whole cell assay suggesting that there was no uptake of OA and/or GPP or that this uptake was at least insufficient under the tested conditions (Zirpel, Degenhardt et al., 2017).
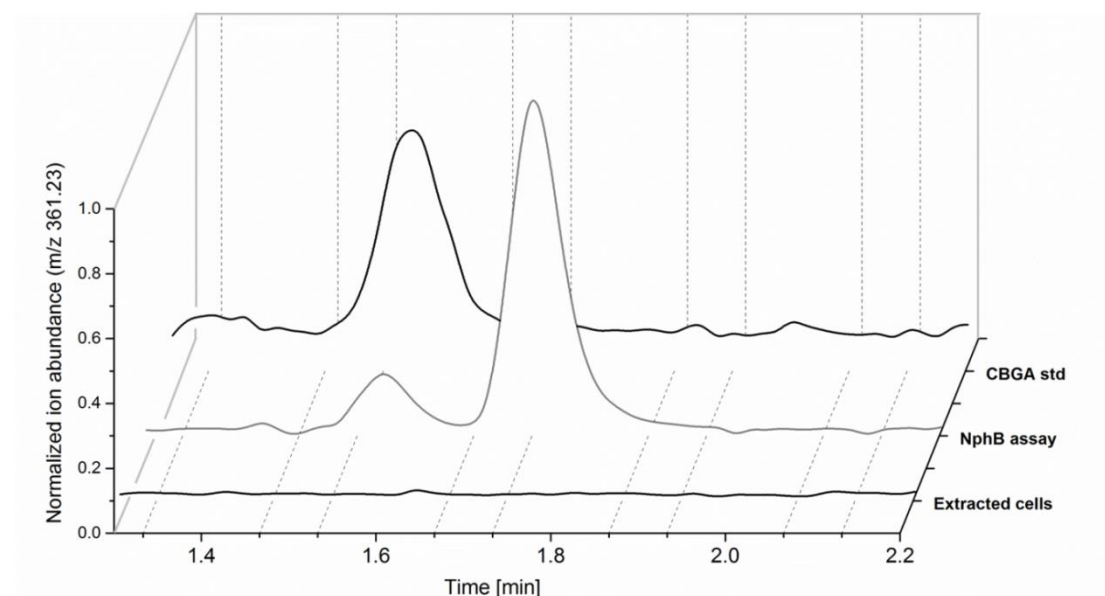
Figure 51: Membrane permeability of *S. cerevisiae* cells for GPP and OA. SC_N expression cultures were inoculated at OD_{600} of 0.5 and incubated for 72 h at 15 °C and 200 rpm in 500 mL baffled shaking flasks with a filling volume of 10%. Cells were harvested, resuspended in assay buffer and either used for cell lysis or for a whole cell assay by supplementing 3 mM GPP and 3 mM OA (4 h, 37 °C). Cell lysate supernatants were incubated with 5 mM magnesium chloride, 1 mM GPP and 1 mM OA (4 h, 37 °C) to determine 2-*O*-GOA and CBGA formation. Cells of the whole cell assay were extracted with ethyl acetate, extracts were vaccum dried and resuspended in AcN. Activity assays and extracts were analyzed by LC-MS. Modified from Zirpel, Degenhardt et al. (2017).

3.3.3. Fusion of *nphB* and *thcas*

As shown in chapter 3.3.1, yeast lysate supernatants expressing *nphB* catalyze the geranylation of OA by GPP to form CBGA, the substrate of the THCAS to form THCA (chapter 1.2.3). Results published by Zirpel et al. (2015) indicate that functional expression of THCAS in *S. cerevisiae* is possible as well. Here, THCAS was targeted to the vacuole using the signal sequence of the vacuolar protease proteinase A (aspartyl protease) which is encoded by *PEP4*, instead of the natural N-terminal signal peptide (Zirpel et al., 2015). Since it is possible to express *nphB* and *thcas* in *S. cerevisiae* separately, we decided to test the expression of both enzymes in *S. cerevisiae* simultaneously. For this, two different approaches were pursued: (i) expression of both enzymes driven by a single promoter, (ii) use of a bidirectional promoter system.

The viral 2A sequences are small peptides that were first identified in the picornaviruses, like the foot-and-mouth disease virus (Beekwilder et al., 2014; Ryan et al., 1991). T2A is a "self-cleaving" peptide identified in *Thosea asigna* virus 2A. The peptides have an average length of 18 – 22 AA and the 2A-mediated cleavage is observable in all eukaryotic cells. The T2A sequence, which has high ribosomal release efficiency, is inserted between the different target genes that should be expressed. After

translation of the upstream gene and the T2A sequence the nascent polypeptide is released by the ribosome which continues the translation of the downstream coding sequence (Beekwilder et al., 2014; Kim et al., 2011). The separation of the coding sequences of *nphB* and *thcas* by a viral T2A sequence enables the expression of both enzymes driven by a single *GAL1* promoter (Beekwilder et al., 2014).

The T2A construct was designed with an upstream *thcas* and a downstream *nphB* because it is unknown whether the N-terminal *pep4* signal peptide of THCAS is negatively affected by co-translational cleavage of the T2A peptide sequence and thus an additional proline at its N-terminus. As shown in Figure 52, the co-translational cleavage (ribosome skipping) of the upstream and the downstream peptide occurs at the C-terminus of the T2A sequence between the glycine and the last proline. Therefore, the upstream enzyme carries 19 additional amino acids, while the downstream peptide starts with a proline (Beekwilder et al., 2014; Kim et al., 2011).
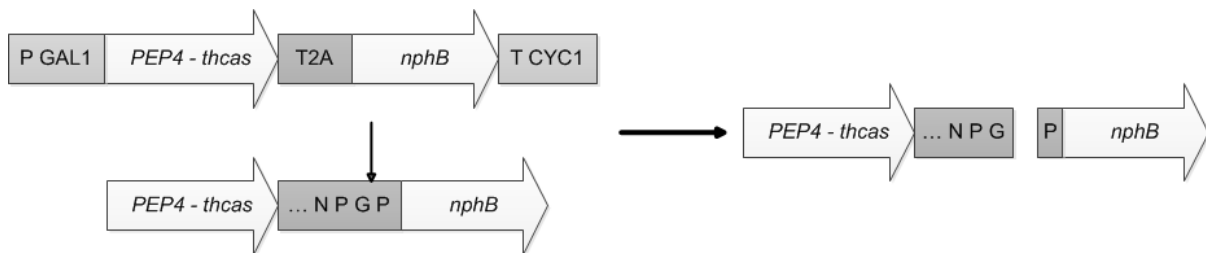
Figure 52: Coding sequences of *thcas* and *nphB* separated by a viral T2A sequence.

The plasmid pDionysos_TT2AN was generated by Gibson Assembly using PCR primers that contained the T2A sequence (chapter 2.2.3.2). SC_TT2AN was cultivated and activity assays were peformed according to chapter 2.2.3.5. Product formation was analyzed by LC-MS (chapter 2.2.3.6) and the results are depicted in Figure 53.

Figure 53: Shaking flask experiments of *S. cerevisiae* cells and analysis of product formation. (A) Comparison of cells expressing only *thcas* (SC_T, (a)), only *nphB* (SC_N, (b)) or *nphB* and *thcas* (SC_TT2AN, (c)), SC_NT, (d)). (B) THC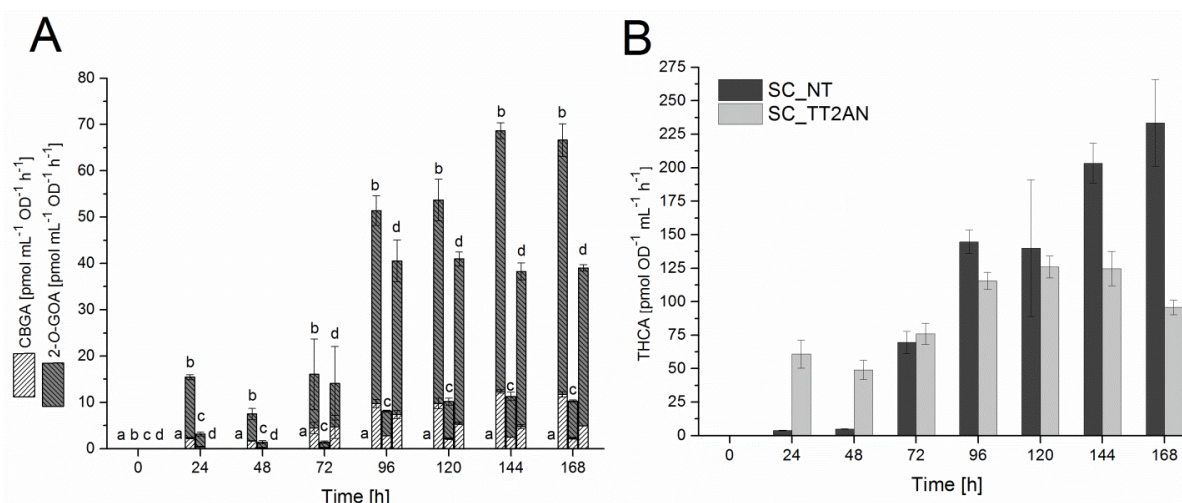AS activity of SC_TT2AN (c) and SC_NT (d). Expression cultures were inoculated at OD_{600} of 0.5 and incubated for 168 h at 15 °C and 200 rpm in baffled shaking flasks with 10% filling volume; cell lysate supernatants were supplemented with 5 mM magnesium chloride, 1 mM OA and 1 mM GPP and incubated at 37 °C for 4 h to determine CBGA, 2-*O*-GOA and THCA formation (A) or supplemented with 0.3 mM CBGA and incubated at 37 °C for 4 h to determine the highest possible THCA production by THCAS under the applied conditions (B). Activity assays were analyzed by LC-MS and LC-DAD and product formation was normalized on cell culture OD_{600}, cell culture volume and assay incubation time. Data points are calculated from biological triplicates, each analyzed in technical duplicates. Modified from Zirpel, Degenhardt et al. (2017).

The results depicted in Figure 53 and Figure 54 show that *nphB* was functionally expressed in the fusion construct as formation of CBGA and 2-*O*-GOA was detectable. The highest CBGA production rate in SC_TT2AN lysates was obtained 96 h after induction (3 pmol mL^{-1} OD^{-1} h^{-1}). However, produced CBGA amounts were not sufficient for THCAS to produce detectable amounts of THCA, even though THCAS was active in these lysates (see Figure 53 B, 96 h). A comparison of SC_TT2AN and SC_N lysates, each supplemented with GPP and OA, indicates that the expression level of separately expressed *nphB* is higher than in the fusion construct. The highest CBGA production rate for SC_N was detected 144 h after induction (12 pmol mL^{-1} OD^{-1} h^{-1}) which is approximately 4-times higher than the highest CBGA production rate obtained for SC_TT2AN (3 pmol mL^{-1} OD^{-1} h^{-1}). For both constructs, the highest 2-*O*-GOA production rates were obtained 144 h after induction (SC_N: 56 pmol mL^{-1} OD-1 h^{-1}, SC_TT2AN: 9 pmol mL^{-1} OD^{-1} h^{-1}). Results of SC_TT2AN lysate supernatants supplemented with CBGA indicate that *thcas* was functionally expressed and the highest THCA production rate was obtained 120 h after induction (126 pmol mL^{-1} OD^{-1} h^{-1}) (Figure 53 B). In contrast to that the highest possible THCA production rate for SC_T was already reached 96 h after induction (427 pmol mL^{-1} OD^{-1} h^{-1}) (Figure 55). This means that the activity levels of both NphB and THCAS are higher if the enyzmes are expressed separately than expressed simultanesously using a single *GAL1* promoter (Zirpel, Degenhardt et al., 2017).
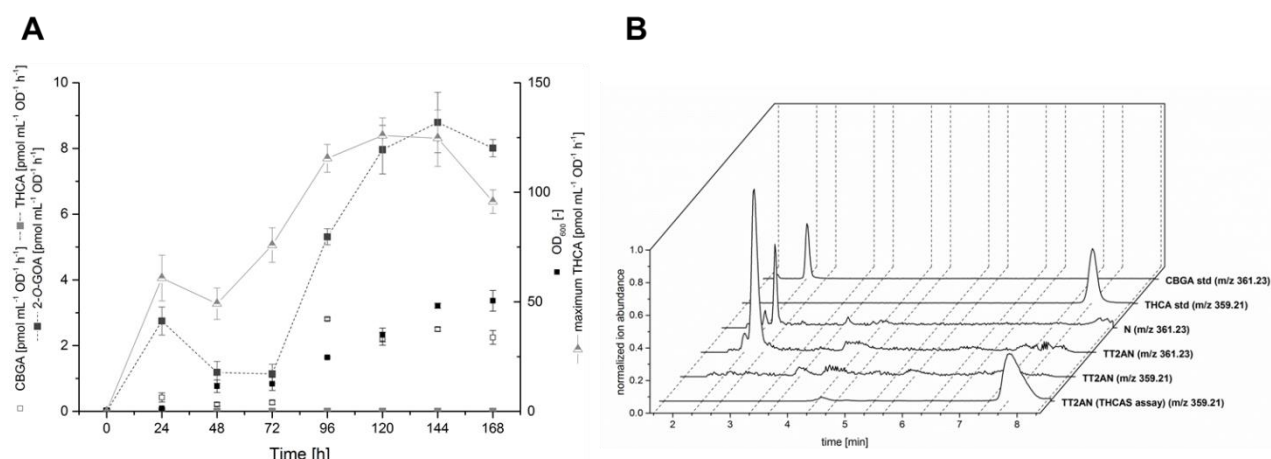
Figure 54: Shaking flask experiments of SC_TT2AN cells and analysis of product formation. Expression cultures were inoculated at OD_{600} of 0.5 and incubated for 168 h at 15 °C and 200 rpm in baffled shaking flasks with 10% filling volume; cell lysate supernatants were either incubated with 0.3 mM CBGA at 37 °C for 4 h to determine the highest possible THCA formation by THCAS under the tested conditions or with 1 mM OA, 1 mM GPP, 5 mM magnesium chloride at 37 °C for 4 h to determine CBGA, 2-*O*-GOA and THCA formation (A). Activity assays were analyzed by LC-MS (B) and LC-DAD and product formation was normalized on cell culture OD_{600}, cell culture volume and assay incubation time. Data points are calculated from biological triplicates each analyzed in technical duplicates.

Since the co-expression of *nphB* and *thcas* driven by a single *pGAL1* promoter did not yield THCA formation, we tested the use of a bidirectional promoter system in order to express *nphB* and *thcas* simultaneously without the utilization of a T2A sequence-facilitated ribosomal skipping. If *S. cerevisiae* is used for the expression of multiple coding sequences without linker sequences like T2A then the application of different promoter und terminator sequences is necessary because eukaryotes like yeast can not translate polycistronic mRNAs. Additionally, yeast possesses a highly active homologous recombination machinery which can cause recombination between two similar or identical sequences and thereby damages in the cloned DNA construct (Siddiqui et al., 2012). We decided to test the bidirectional *GAL10/GAL1* promoter system which is localized on chromosome two of *S. cerevisiae* (Johnston and Davis, 1984; West et al., 1984). pGAL1 and pGAL10 are both classified as strong galactose-induced promoters (Partow et al., 2010).

Zirpel et al. (2015) published that *thcas* is functionally expressed under pGAL1. Therefore, the construct pDio2_THCAS was generated in order to test whether *thcas* expression is possible under pGAL10 as well. SC_T2 cells were cultivated for 168 h at 200 rpm and 15 °C (chapter 2.2.3.5). Samples for activity assays and OD_{600} measurements were taken every 24 h. Cell lysate supernatants were supplemented with CBGA and THCA formation was analyzed by LC-MS and LC-DAD (chapter 2.2.3.6). The results are depicted in Figure 55.
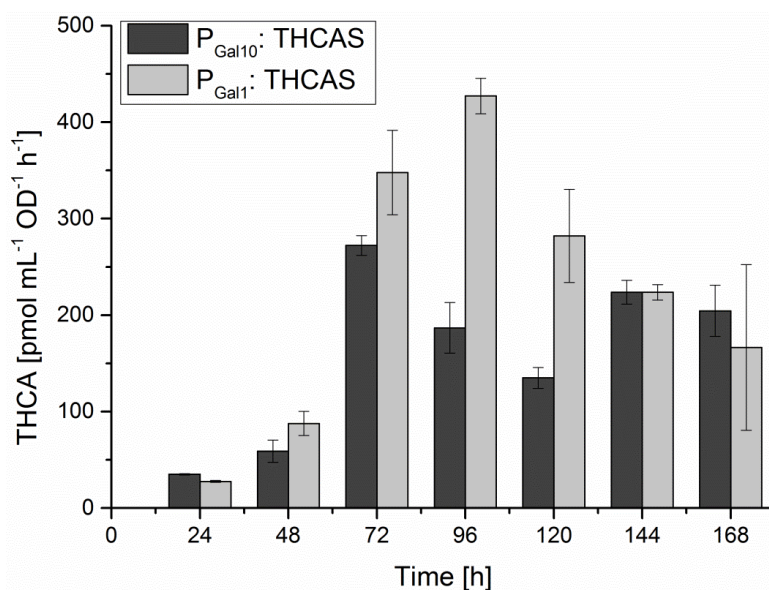
Figure 55: Shaking flask experiments of *S. cerevisiae* cells and analysis of THCA formation. Expression cultures (SC_T and SC_T2) were inoculated at OD_{600} of 0.5 and incubated for 168 h at 15 °C and 200 rpm in baffled shaking flasks with 10 % filling volume; cell lysate supernatants were supplemented with 0.3 mM CBGA and incubated at 37 °C for 4 h. Activity assays were analyzed by LC-MS and LC-DAD and THCA formation was normalized on cell culture OD_{600}, cell culture volume and assay incubation time. Data points are calculated from biological triplicates, each analyzed in technical duplicates. Modified from Zirpel, Degenhardt et al. (2017).

Enzyme activities of SC_T2 lysate supernatants supplemented with CBGA show that functional expression of *thcas* under pGAL10 is possible. The highest THCA production rate was obtained 72 h after induction (272 pmol mL^{-1} OD^{-1} h^{-1}) (Figure 55). A comparison of the THCA production rates of SC_T and SC_T2 indicates that functional expression of *thcas* under control of pGAL10 is less than *thcas* expressed under control of pGAL1. The THCA production rate of SC_T lysate supernatants supplemented with CBGA yielded 348 pmol mL^{-1} OD^{-1} h^{-1} THCA 72 h after induction and the highest THCA production rate was obtained 96 h after induction (427 pmol mL^{-1} OD^{-1} h^{-1}) (Figure 55). On this account, the *GAL1* promoter is more favourable for *thcas* expression than pGAL10. Since the THCA formations rates of SC_T and SC_T2 were both high, we decided to use the stronger *GAL1* promoter for the expression of *nphB* and pGAL10 for *thcas* expression (Zirpel, Degenhardt et al., 2017).

The vector pDio2_NT was cloned and transformed into the yeast strain SC (chapter 2.2.3.2). SC_NT cells were cultivated for 168 h at 200 rpm and 15 °C. Samples for activity assays and OD_{600} measurements were taken every 24 h (chapter 2.2.3.5). Cell lysate supernatants were supplemented with OA and GPP and product formation was analyzed by LC-MS and LC-DAD (chapter 2.2.3.6). The results depicted in Figure 53 and Figure 56 indicate that CBGA and 2-*O*-GOA were formed, but no THCA. The highest CBGA production rate was obtained 96 h after induction (7 pmol mL^{-1} OD^{-1} h^{-1}), which is similar to the CBGA production rate of SC_N at the same time point (10 pmol mL^{-1} OD^{-1} h^{-1}). As described for SC_TT2AN the CBGA production rate does not seem to be sufficient enough for oxidative cyclization by THCAS. The 2-*O*-GOA production rate is constant between 96 h and 168 h after induction (approximately 34 pmol mL^{-1} OD^{-1} h^{-1}). SC_NT lysate supernatants supplemented with CBGA yielded the highest possible THCA production rate 168 h after induction

(233 pmol mL^{-1} OD^{-1} h^{-1}). The results shown in Figure 53 and Figure 57 indicate that the THCA production rate of SC_NT was still increasing 168 h after induction (Zirpel, Degenhardt et al., 2017).
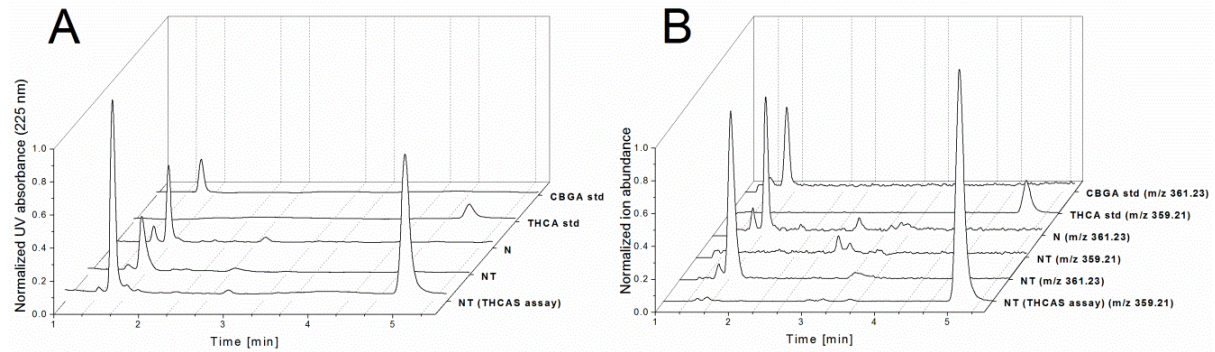
Figure 56: LC-MS analysis of cannabinoids produced with cell lysate supernatants of SC_N and SC_NT. The yeast cell lysate supernatants were incubated with 5 mM magnesium chloride, 1 mM OA and 1 mM GPP and incubated at 37 °C for 4 h to determine CBGA, 2-*O*-GOA and THCA formation or they were supplemented with 0.3 mM CBGA and incubated at 37 °C for 4 h to determine the highest possible THCA formation under the tested conditions (THCA assay). (A) LC-UV chromatograms at 225 nm of CBGA standard, THCA standard and the assay products. (B) Extracted ion chromatograms of *m/z* 361.23 (CBGA, 2-*O*-GOA) and *m/z* 359.21 (THCA) of CBGA standard, THCA standard and the assay products. N – assay of yeast cell lysate supernatants expressing only *nphB*, NT – assay of yeast cell lysate co-expressing *nphB* and *thcas* (SC_NT) (Zirpel, Degenhardt et al., 2017).
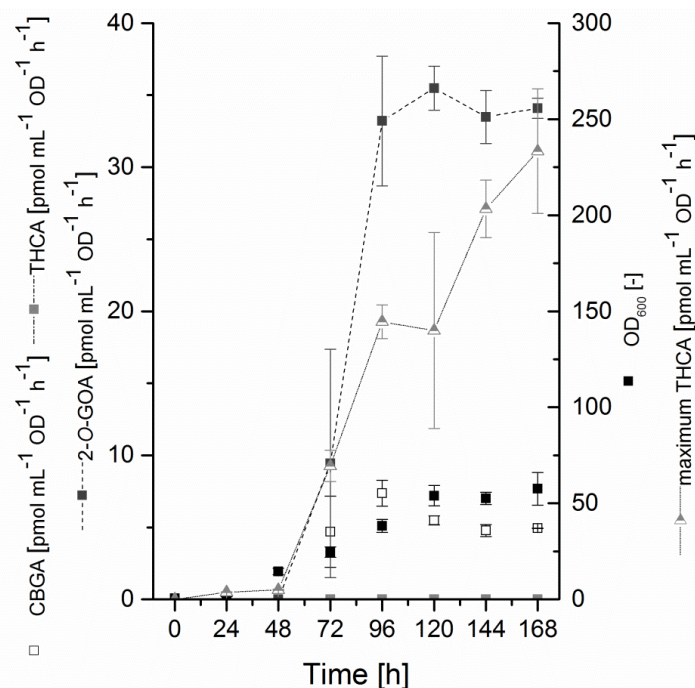
Figure 57: Shaking flask experiments of SC_NT cells and analysis of product formation. Expression cultures were inoculated at OD$_{600}$ of 0.5 and incubated for 168 h at 15 °C and 200 rpm in baffled shaking flasks with 10% filling volume; cell lysate supernatants were either incubated with 0.3 mM CBGA at 37 °C for 4 h to determine the highest possible THCA formation under the tested conditions or with 1 mM OA, 1 mM GPP, 5 mM magnesium chloride at 37 °C for 4 h to determine CBGA, 2-*O*-GOA and THCA formation. Activity assays were analyzed by LC-MS and LC-DAD and product formation was normalized on cell culture OD$_{600}$, cell culture volume and assay incubation time. Data points are calculated from biological triplicates each analyzed in technical duplicates (Zirpel, Degenhardt et al., 2017).

3.3.4. Discussion

In chapter 3.2.3 we showed that NphB is able to catalyze the prenylation of OA by GPP to form CBGA. So far, the aromatic prenyltransferase was only functionally expressed in *E. coli*. With regard to the overall goal of generating a cannabinoid producing yeast, functional expression of *nphB* in *S. cerevisiae* cells is necessary. For that reason we first tried to express *E. coli* codon-usage optimized NphB in SC2 cells. Neither expression with a N-terminal TrxA tag, nor expression without a solubility tag yielded detectable product formation. As mentionend in chapter 3.2.7 expression experiments of *nphB* with and without N-terminal TrxA tag indicated that the tag is necessary for functional expression because expression was only detectable by SDS-PAGE when NphB was fused to TrxA. According to Kimple et al. (2013) and Terpe (2003) this solubility tag is used for the heterologous expression in bacteria. If one aims to enhance recombinant protein expression in eukaryotes like yeast, the solubility tags small ubiquitin-like modifier (SUMO) tag or the glutathione-S-transferase (GST) tag are applicable (Bell et al., 2013; Kimple et al., 2013). Hence, missing enzymatic activity might be caused by non-functional expression, a lack of expression due to use of a solubility tag that is not suitable for yeast and/or applying an *E. coli* codon-usage optimized coding sequence. In a next step we compared the codon-usages of *S. cerevisiae* and *E. coli* and identified a codon, encoding for Arg211, of the *E. coli* codon-usage optimized NphB sequence with low frequency in *S. cerevisiae*, *i.e.* the corresponding t-RNA is rarely in yeast, which could cause low expression levels or a total lack of expression. However, even after exchange of the rare codon with a more common one, no product formation of CBGA or 2-*O*-GOA was detectable. In this regard, functional expression might be obtained if the codons coding for Arg9, Lys169, Arg228, Arg268 and Arg296 were exchanged by more common codons as well. As shown in Figure 48 these amino acids are encoded by codons which are used less than 20%.

Since the expression of the *E. coli* codon-usage optimized NphB sequences did not yield functional expression, we decided to use a yeast-codon usage optimized coding sequence instead (pDionysos_N*). Additionally, we tested a NphB coding sequence without lysines that are located on the α-helices surrounding the PT-barrel of the aromatic prenyltransferase (pDionysos_N(arg)*; Figure 49). The corresponding amino acids were exchanged by arginines in order to decrease ubiquitination that directs NphB to degradation by the proteasome. The results obtained by activity assays indicate that NphB wt forms approximately two to four times more CBGA and 2-*O*-GOA than the enzyme without lysines on its surface (Figure 50). Since none of the amino acids in the active site of the enzyme were changed, the amino acid exchanges might effect the expression or the proper folding of the prenyltransferase. Sokalingam et al. (2012) exchanged 14 lysines on the surface of GFP, which exhibits a β-barrel structure, against arginine. The obtained GFP variant was more stable in the presence of chemical denaturants like ionic detergents or an alkaline pH. In comparison to the wildtype enzyme additional hydrogen-bonds and salt-bridges had been formed. Even though, Sokalingam et al. (2012) were able to enhance protein stability, the folding rate of GFP and its productivity in forms of detectable fluorescence decreased. Furthermore, they showed that most of the target protein was obtained in insoluble form. Besides the formation of new hydrogen bond

interactions and salt bridges, it can not be excluded that one or more of the exchanged lysines on the surface of NphB are essential for functional expression. Maybe functional expression of *nphB(arg)* could be increased if all arginines present on the surface of NphB wt would be exchanged against lysines to re-establish the surface arginine to lysine ratio (Figure 58). Rodriguez et al. (2000) showed that the transcriptional activity of the p53 protein is higher if lysine-to-arginine changes were performed. Addtionally, Worthington and Carbonetti (2007) published that the toxicity (cellular activity) of the pertussis toxin decreased after arginine-to-lysine changes in the amino acid sequence of the enzymatically active A subunit (S1) that lacks of lysine residues. The authors hypothesized that the arginine-to-lysine changes enable ubiquitnation and thus protein degradation by the proteasome. Therefore, it should be evaluated on a case by case basis, which lysines on the surface of NphB are essential for functional expression and which lysines can be changed to arginine in order to decrease protein degradation and thus to increase the amount of enzymatically active NphB.
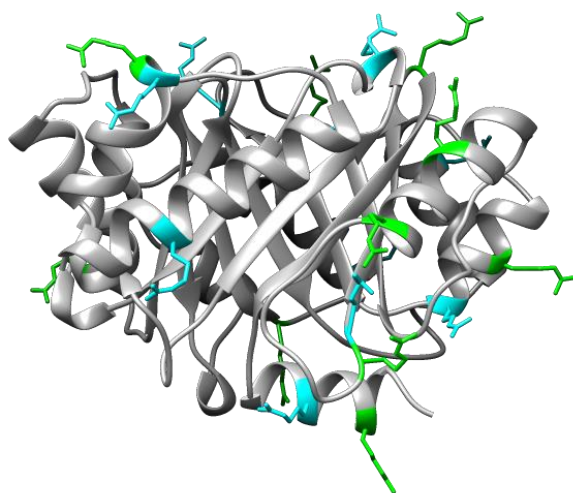
Figure 58: Crystal structure of NphB*.Cyan: Arginine residues of NphB wt. Green: Lysine residues on the surface of NphB that were mutated to arginine in order to decrease ubiquitination that directs NphB for degradation by the proteasome. The model was generated using UCSP chimera (Pettersen et al., 2004). The lysines were exchanged using Dunbrack rotamers (Dunbrack, 2002).

Subsequently, we tried to enhance *nphB* expression levels and thus formation of CBGA and 2-*O*-GOA by reducing the effect of carbon catabolite repression. The transcription factor MIG1 is known to repress the expression of yeast *GAL* genes in the presence of glucose by binding to the *GAL4* and the *GAL1* promoters (Johnston et al., 1994; Kayikci and Nielsen, 2015; Nehlin et al., 1991). Results of Nehlin et al. (1991) showed that expression of GAL1 is repressed 83-fold when cultivating yeast in the presence of 2% glucose and 2% galactose. After disruption of the *mig1* gene a 3.9-fold depression of *GAL1* was detectable. In addition to cultivation in the presence of fructose, we decided to knockout the mig1 binding sites of pGAL1 in order to reduce carbon repression. The results depicted in Figure 50 indicate that the expression of *nphB* wt and thus the product formation rate increased (SC2_N) in comparison to expression under the control of the native pGal1 (SC2_N*). In contrast to that, no effect on the expression of N(arg) was detectable indicating again that the lysine-arginine exchanges might affect protein folding. The highest product formation rate was obtained for SC2_N (CBGA:

36 pmol mL^{-1} OD^{-1} h^{-1}, 2-O-GOA: 128 pmol mL^{-1} OD^{-1} h^{-1}). A comparison of the product formations rates of SC2_N and SC_N (CBGA: 12 pmol mL^{-1} OD^{-1} h^{-1}, 2-O-GOA: 56 pmol mL^{-1} OD^{-1} h^{-1}; Figure 53) indicates that SC2 cells are more suitable for functional expression of *nphB*. Unfortunately, expression of *thcas* in SC2 cells was not successful (data not shown). Therefore, expression of *nphB* and *thcas* simultaneously was performed in SC cells.

Until now, we have performed the NphB activity assays with cell lysate supernatants or purified enzyme. In terms of generating a cannabinoid producing yeast as well as reducing time and costs, the use of whole cells for biocatalysis is inevitable. Therefore, we tried to feed NphB expressing yeast cells with OA and GPP. The results of the whole cell assay indicate that there was no uptake of OA and/or GPP or that this uptake was at least insufficient under the tested conditions (chapter 3.3.2). Nevertheless, *nphB* was functionally expressed since CBGA formation was detectable in SC_N lysate supernatants supplemented with OA and GPP.

In plants GPP is formed within the plastids (MEP pathway) by the condensation of DMAPP and IPP. The reaction is catalyzed by a prenyltransferase, namely GPP synthase (GPPS) (Bouvier et al., 2000; Burke et al., 1999). Even though GPP is produced in the plastids, higher plants possess a cytosolic GPP pool as well as transport of some isoprenoid intermediates over the plastid membranes is possible (Bick and Lange, 2003; Dong et al., 2016; Gutensohn et al., 2013). In contrast to that, *S. cerevisiae* cells do not contain a specific GPPS but they carry a farnesyl diphosphate synthase (FPPS) with GPPS and FPPS activities. The *trans*-prenyltransferase FPPS is encoded by the *ERG20* gene and catalyzes in a first step the head-to-tail condensation of IPP and DMAPP to form GPP and in a second step the condensation of GPP with another IPP molecule to form FPP (Fischer et al., 2011). *In vitro* measurements of FPPS wt activity resulted in the generation of 75% FPP and 25% GPP (Oswald et al., 2007). On this account, the intracellular GPP supply in wildtype yeast is most likely not high enough to use it as precursor for CBGA formation.

The polyketide metabolite OA is naturally not available in yeast. Therefore, the substrate has to be fed by supplementing the yeast cultivation medium with OA. Due to its hydrophobic character it is possible that OA binds to the yeast membrane or is integrated into it, but it is not released into the cytosol of the yeast cells. The whole-cell assay was performed at pH 7.5 (chapter 2.2.3.4). According to Casal et al. (2008) and Jamalzadeh et al. (2012) uptake of carboxylic acids across the plasma membrane is possible at a low pH (pH 3 – 5). Depending on the pH of the aqueous solution and the pKa value, an equilibrium between the undissociated and the dissociated form of the carboxylic acid is established. The uncharged form is favoured at low pH values and might exert lipid-solubility. One disadvantage of passive diffusion over the plasma membrane is the fact that the undissociated carboxylic acid dissociates in the neutral pH of the cytosol and acidifies it which could influence metabolic pathways of the yeast (Casal et al., 2008). Experiments relating the uptake of the dicarboxylic acid fumaric acid in yeast showed that with decreasing pH levels of the medium the uptake rate of the acid was increasing until a maximum level was reached (Jamalzadeh et al., 2012). Based on the results of Casal et al. (2008) and Jamalzadeh et al. (2012), the pH applied in the whole-cell assay might be too high (pH 7.5)

for the uptake of the carboxylic acid OA. Additionally, the use of complex medium instead of assay buffer would be preferable for performing a whole-cell assay since complex medium contains glucose and other supplements which enables the yeast cell to grow and stay viable.

Taking the total biosynthetic pathway for cannabinoid production into consideration, OA feeding will be too expensive for an economically viable industrial process. On this account, inter alia expression of GPPS (Marks et al., 2009) as well as OLS (Taura et al., 2009) and OAC (Gagne et al., 2012) in yeast might enable cost-efficient whole cell catalysis, while feeding hexanoate/malonate (Zirpel, Degenhardt et al., 2017).

The simultaneous expression of *nphb* and *thcas* in *S. cerevisiae* was not successful. Neither the expression of both enzymes driven by a single promoter (pGAL1) nor the use of pGAL10/pGAL1 bidirectional promoter system led to formation of THCA. In this context it is important to remember that the functional expression of both enzymes was possible, even though the CBGA production rates of NphB were low (Figure 53).

In a first approach we tried to express *nphB* and *thcas* under the control of a single pGAL1. The two coding sequences were separated by a T2A sequence. For both enzymes coded by the genetic construct, individual enzyme activities were measurable, but no THCA formation was detectable when supplementing SC_TT2AN lysate supernatants with OA and GPP (chapter 3.3.3). The CBGA production rate was low (3 pmol mL^{-1} OD^{-1} h^{-1}) and thereby most probably not sufficient enough for detectable oxidative cyclization catalyzed by THCAS. The cleavage of the upstream THCAS and the downstream NphB is carried out by ribosome skipping. Thereby the formation of a peptide bond between the C-terminal glycine of 2A and the N-terminal proline of NphB ("2A cleavage site") (Figure 59) is prevented. After releasing the upstream, nascent THCAS from the ribosome, the translation of the downstream peptide can continue (Donnelly et al., 2001b, 2001a; Liu et al., 2017). According to Donnelly et al. (2001b) and Liu et al. (2017) the ribosomal skipping can lead to three different outcomes: (i) successful skipping of THCAS with a C-terminal T2A sequence and NphB with an additional N-terminal proline, (ii) unsuccessful skipping which results in a THCA-NphB fusion protein, and (iii) successful skipping, but discontinued translation of the downstream NphB due to ribosome fall-off. Even though Beekwilder et al. (2014) and Liu et al. (2017) published that the use of a T2A sequence resulted in hardly any fusion product, the low CBGA and THCA production rates of NphB and THCAS could be caused by unsuccessful ribsosomal skipping.
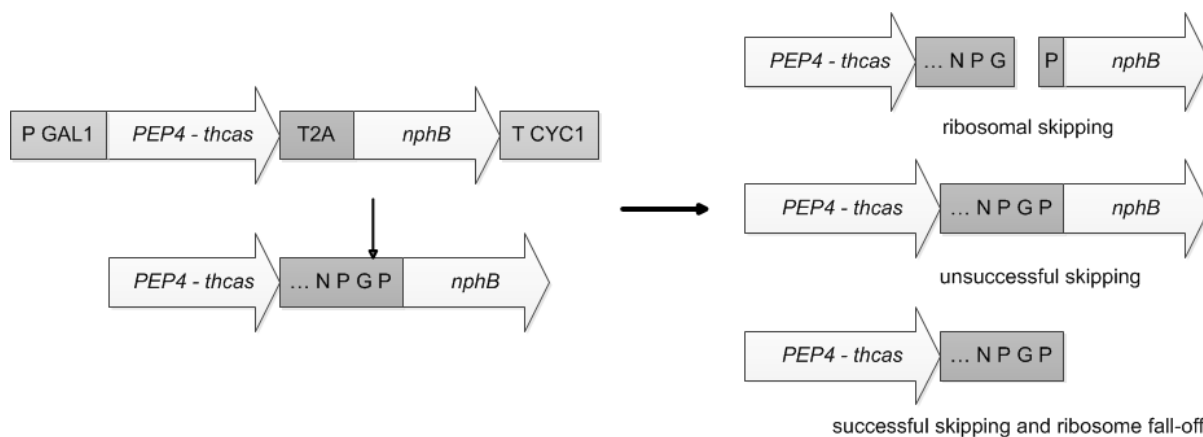
Figure 59: Outcomes of ribosomal skipping. The "self-cleavage" mechanism of the T2A sequence can lead to successful cleavage, unsuccessful cleavage or successful cleavage followed by fall-off of the ribosome and discontinued translation of the downstream sequence. Modified from Liu et al. (2017).

Liu et al. (2017) showed that the expression level of a protein in a bi-cistronic vector is dependent on the position of the enzyme. For that purpose they used constructs with GFP N- or C-terminal of the T2A sequence and analyzed the cells using Western Blot, live fluorescent images and flow cytometry. Their results indicate that the expression level of the gene at the second position is decreased by about 70%, compared with the expression level of the gene at the first position. This can be explained by the fact that the probability of ribsosomal skipping is approximately 30%, while unsuccessful skipping happens with a probability of 10%. Additionally, there is a possibility of 60% that the ribosome falls off and the translation of the downstream protein is not continued (Liu et al., 2017). The oberserved imbalances in translation products are consistent with the results of Donnelly et al. (2001b) who showed that the molar excess of the peptide at the first position is two to five times higher than the one of the peptide at the second position. Additionally, they reported that these imbalances were only detectable in artificial polypeptide systems.

The results published by Donnelly et al. (2001b) and Liu et al. (2017) indicate that the amount of functional expressed *thcas* is higher than the expression of *nphB* because the coding sequence of THCAS was cloned upstream of the T2A sequence (first position). A comparison of the respective production rates of the two enzymes in SC_TT2AN lysate supernatants confirms that the enzymatic activity of THCAS was much higher than the one of NphB (maximum THCA production rate (THCAS): 126 pmol mL^{-1} h^{-1} OD^{-1}, 120 h after induction; maximum CBGA production rate (NphB): 3 pmol mL^{-1} h^{-1} OD^{-1}, 96 h after induction). The highest CBGA production rate of SC_N was detected 144 h after induction (12 pmol mL^{-1} h^{-1} OD^{-1}) which is about 4.4 fold higher than the one in SC_TT2AN lysate supernatants. This might indicate again that the expression level of NphB in SC_TT2AN might be low due to unsuccessful ribosome skipping and/or successful skipping followed by ribosome fall-off (Figure 59). The highest THCA production rate of SC_T was approximately 3.4 times higher (427 pmol mL^{-1} h^{-1} OD^{-1}; 96 h after induction) than the maximum THCA production rate of SC_TT2AN (126 pmol mL^{-1} h^{-1} OD^{-1}; 120 h after induction). Beside unsuccessful ribosome skipping, this difference in enzymatic activity could be caused by the fact that the C-terminus of THCAS was prolonged by 19 AA pertaining to the T2A sequence (Figure 59). THCAS belongs to the BBE-like family (chapter

1.2.3) and the crystal structure (PDB: 3VTE) indicates that the BBE domain at the C-terminus of the enzyme is not part of a stabilized protein secondary structure (α-helix, β-sheet) or stabilized by strong chemical bonds (Sirikantaramas et al., 2007, 2004; Taura et al., 2007b; Zirpel et al., 2018b). Recently, Zirpel et al. (2018b) published that the fusion of GFP to the C-terminus of THCAS resulted in enzyme expression, but no activity was detectable. Based on that they concluded that folding of the THCAS occurred, but the catalytic base of the active center (Y484) was depositioned causing a complete lack of enzymatic activity. Additionally, they showed that step-wise truncation of the C-terminus influenced THCAS activity negatively. Since the tertiary structure of the enzyme seems to be susceptible to changes at its C-terminus, the addition of the T2A sequence might have caused a decreased THCA production rate of SC_TT2AN lysates compared with SC_T lysates.

The low CBGA production rate of NphB in SC_TT2AN might be caused by a false localization of the enzyme due to the N-terminal *PEP4* signal peptide of THCAS. Contrary to that, results of de Felipe et al. (2003a) showed that even if the protein N-terminal to the T2A sequence contains a N-terminal signal peptide (like THCAS) the C-terminal protein stays in the cytosol, unless it contains a targeting sequence as well. Additionally, they demonstrated that peptides with a N-terminal signal sequence are also targeted correctly when they are C-terminal of the T2A sequence (de Felipe et al., 2003). As the additional proline does not seem to have an effect on the targeting, it might be possible to switch the genes up- and downstream of the T2A sequence in pDionysos_TT2AN. According to the results published by Donnelly et al. (2001b) and Liu et al. (2017), the expression level of *nphB* would be increased and the C-terminus of THCAS would not be modified, thus production rates of NphB and THCAS might be increasing.

Since the expression of *nphB* and *thcas* simultaneously driven by a single GAL1 promoter did not lead to the production of THCA, we decided to test the expression of both enyzmes using a pGAL10/pGAL1 bidirectional promoter system. Even though both promoters are classified as strong inducible promoters (Partow et al., 2010), THCA production rates of SC_T and SC_T2 lysate supernatants supplemented with CBGA indicate that pGAL1 (max. THCA production rate: 427 pmol mL^{-1} h^{-1} OD^{-1}, 96 h after induction) is a stronger promoter than pGAL10 (max. THCA production rate: 272 pmol mL^{-1} h^{-1} OD^{-1}, 72 h after induction). These results are supported by Cartwright et al. (1994) and West et al. (1987), but are in contrast to Partow et al. (2010) who showed that pGAL10 is stronger than pGAL1 in shake flask experiments. However, we decided to use pGAL10 for *thcas* expression and pGAL1 for *nphB* expression. The obtained results indicate that *S. cerevisiae* is able to express both enzymes in a functional and active manner, but unfortunately no THCA was detectable in SC_NT lysates supplemented with OA and GPP. Even though, the activity levels of NphB seem to be the rate limiting step under the tested conditions, results obtained by co-expressing *nphB* and *thcas* in *Komagataella phaffii* (formely *Pichia pastoris*) indicate that functional expression of THCAS is most likely the limiting step (see below) (Zirpel, Degenhardt et al., 2017). The expression levels of NphB might be increased when using the "weaker" GAL10 promoter instead of pGAL1.

The results of both approaches tested for the expression of *nphB* and *thcas* simultaneously show that the CBGA productions rates of *nphB* are too low for utilization by THCAS as a substrate. A

comparison of the CBGA production rates of SC_N and SC_NT lysates supplemented with OA and GPP indicate that the low NphB activity levels are predominantly caused by the levels of functional expression of *nphB* itself and not by the co-expression with *thcas* (max. CBGA production rates: 12 pmol mL^{-1} h^{-1} OD^{-1} (SC_N; 144 h after induction), 7 pmol L^{-1} h^{-1} OD^{-1} (SC_NT; 96 h after induction). Under the tested conditions NphB catalyzes a C-prenylation (CBGA) and an O-prenylation (2-*O*-GOA) of OA. The side-product 2-*O*-GOA represents about 85% of the formed prenylation products. If the formation of 2-*O*-GOA is decreased and thereby the generation of CBGA is increased, the CBGA concentration might be high enough for THCAS to oxidatively cyclize CBGA to THCA (Zirpel, Degenhardt et al., 2017). As mentioned before and shown in chapter 3.2.6 a rational protein engineering approach based on the crystal structures of NphB and substrate docking are able to enhance NphB specificity towards CBGA.

Interestingly, co-expression of *nphB* and *thcas* in *K. phaffii* simultaneously yielded THCA formation after supplementing lysates with OA and GPP. As observed for *S. cerevisiae,* 2-*O*-GOA is formed as a major side-product. The best performing clone was able to produce approximately 80 pmol mL^{-1} h^{-1} OD^{-1} THCA. A comparison of the results obtained for the activity assays performed with *S. cerevisiae* or *K. phaffii* lysates, respectively, indicates that in the baker's yeast expression system THCAS might the limiting factor. According to activity measurements, clone C19 of *K. phaffii* formed comparable amounts of CBGA and 2-*O*-GOA as SC_NT (168 h after induction). This CBGA quantity was sufficient enough for THCAS produced in *K. phaffii* to form THCA, while no THCA production was detectable in *S. cerevisiae* lysates (Zirpel, Degenhardt et al., 2017). Nevertheless, even though expression of *thcas* seems to be the bottleneck, enzyme activites of NphB and THCAS have to be enhanced by increasing the level of functional expression and/or the specific enzyme acitivities. Zirpel et al. (2018a, 2018b) published that they were able to increase THCAS activity by protein engineering and co-expression of helper proteins (see below).

The genes of NphB and THCAS were expressed in a high-copy number plasmid using the strong inducible promoters GAL1 and GAL10 which could have led to high expression levels of one or both enzymes. On this account, the availability of chaperones and other factors involved in proper protein folding might be limiting and thus cause unfolded, misfolded or degraded enzymatically inactive peptides. Additionally, overexpression of proteins which have to pass the secretory pathway can cause severe stress in the ER inducing unfolded protein response (UPR) (Spohner et al., 2015). Zirpel et al. (2018a) showed that THCAS activities were enhanced upon co-production of FAD synthase FAD1p, the chaperones CNE1p and Kar2p, the UPR-activator Hac1p and the foldase PDI1p.

Chapter 4

Summary and outlook

4. Conclusion and outlook

4.1. Expression of membrane-bound prenyltransferases in *S. cerevisiae*

CsPT1, CsPT2 and CsPT3 (Pamplaniyil, 2017) belong to the group of membrane-bound aromatic prenyltransferases since each of them features more than one transmembrane helix and two aspartate-rich motifs that are characteristic for this subgroup of aromatic prenyltransferases. The three corresponding genes were expressed in *S. cerevisiae* using C-terminal yEGFP as an expression reporter and pGAL1 as the inducible promoter. The online-tools ChloroP, Predotar and TargetP predicted N-terminal cTP's within the coding sequences of CsPT1, CsPT2 and CsPT3. We decided to express the native *C. sativa* genes encoding for the three enzymes because removal of the N-terminal transit peptide sequences was likely to increase enzyme activity levels or the recombinant protein production (Akashi et al., 2008; Li et al., 2015; Munakata et al., 2014). Additionally, the expression of the three aromatic prenyltransferases in presence of a N-terminal yCoq2 signal peptide was tested. Detectable whole-cell fluorescence during cultivation indicated that *CsPT1*, *CsPT2* and *CsPT3* as well as *yCoq2SP-CsPT1*, *yCoq2SP-CsPT2* and *yCoq2SP-CsPT3* were expressed and integrated into a yeast membrane as yEGFP only folds and becomes fluorescent if the upstream membrane protein was integrated into a membrane (Drew et al., 2008). The yeast membranes were isolated and used for activity assays using GPP as the prenyl donor and OA, olivetol, resveratrol or PHB as the prenyl acceptor. Unfortunately no conversion of the aromatic substrates to their geranylated products was detectable by LC-MS. The missing enzyme activity, under the tested conditions, could amongst others be caused by the fact that none of the tested sequences belong to the CBGAS or by non-functional expression. In this context, it has to be kept in mind that even if the protein is membrane-integrated and fluorescence is detectable a functional expression is not ensured (Drew et al., 2008).

As mentioned previously, up to this date, there are only two publications dealing with this aromatic prenyltransferase. According to Fellermeier et al. (1998) CBGAS is a soluble enzyme, even though membrane-bound activity could not be excluded. On the contrary, Page and Boubakir (2014) published the coding sequence of membrane-bound CBGAS (CsPT1P). Before further expression experiments are performed and since there are contradictory publications regarding the CBGAS, it should be investigated whether CBGAS activity is detectable in the membrane or the soluble fraction of *C. sativa* flowers. If CBGAS is a membrane-bound enzyme, different recombinant expression systems, besides *S. cerevisiae*, should be taken into account in order to find a more favourable lipid composition and maybe higher expression levels: *Sf9* insect cells, mammalian cells, *K. phaffii.* In some cases membrane lipids function as cofactors for enzyme activity. One advantage of *K. phaffii* over *S. cerevisiae* is the fact that they can be cultivated at higher cell densities. Additionally, the change of the strong GAL1 promoter to a weaker promoter should be taken into consideration because this could enable more time for proper folding of the nascent peptide (Bernaudat et al., 2011; Drew et al., 2008; He et al., 2014). In regard to autofluorescence, a yeast strain without an *ade2* mutation should be utilized for expression experiments when yEGFP is used as an expression reporter. According to Kokina et al. (2014) adenine is often depleted faster than glucose in yeast extract-based rich medium. If the adenine in the medium is used up, increasing cell densities are not caused by dividing cells but

by swelling yeast cells. Additionally, long-term survival of the yeast cells is decreasing. Thus, lower protein expression levels and a lower amount of yeast cells containing the target protein might be obtained. In order to set up an expression platform for membrane-bound aromatic prenyltransferases with *C. sativa* origin, the use of CsPT6 as a positive control should be taken into consideration. CsPT6 is a membrane-bound aromatic prenyltransferase that catalyzes the prenylation of PHB with GPP. This *p*-hydroxybenzoate:polyprenyl transferase with *C. sativa* origin has been described recently by Pamplaniyil (2017).

If CBGAS activity is detectable in the solube fraction of *C. sativa* flowers, the protein of interest could be isolated by several purification steps and analyzed by peptide mass fingerprinting. If the aromatic prenyltransferase, however, belongs to the group of integral membrane proteins, *in silico* analysis might be supported by density gradient centrifugation in order to get more information on the localization of the CBGAS. In this regard it has to be kept in mind that CBGAS activity might be dependent on the co-expression of a so far unknown enzyme. As mentioned previously (chapter 3.1.5), Li et al. (2015) showed that they were only able to detect bitter acid production if the aromatic prenyltransferases HIPT1L and HIPT2 were co-expressed and form an enzyme complex.

4.2.   Expression of *nphB* in *E. coli*

With regard to an industrial relevant biotechnological process membrane-bound enzymes like CBGAS are avoided since expression levels are relatively low compared to soluble enyzmes and membrane-bound enzymes are difficult to handle. However, functional expression of CBGAS has not been achieved so far. Therefore, we tried to replace CBGAS with a soluble aromatic prenyltransferase. In this study we showed that functional expression of *nphB*, isolated from *Streptomyces* spec. strain CL190, is possible in *E. coli* BL21 (DE3). The enzymes catalyzes the prenylation of 1,6-DHN resulting in three different mono-geranylated products. In addition, NphB catalyzes the formation of 2-geranyl olivetol and CBG. Unfortunately, CBDAS, CBCAS and THCAS do not accept CBG as substrate (Morimoto et al., 1997; Taura et al., 1996, 1995). In this study, we presented for the first time that NphB is able to prenylate OA with GPP to form CBGA. In this context a major side-product formation (~ 85 %) resulting in 2-*O*-GOA was detectable, which presents an obstacle for microbial cannabinoid biosynthesis. A rational protein engineering might enable increased enzyme activities and enhanced NphB product specificity towards CBGA (Zirpel, Degenhardt et al., 2017).

So far, no protein structural model of NphB co-crystallized with magnesium, GSPP and OA is available. Therefore a structural model was generated using homology modelling and substrate docking. Based on this model, amino acid exchanges were predicted that could enable the improvement of NphB's specificity towards CBGA and maybe higher specific activities. A low throughput screening platform was established in order to cultivate the different NphB variants. Unfortunately, compared to the wildtype enzyme, none of the tested NphB variants showed an improved product specifity towards CBGA production while preserving enzymatic activites of NphB wt. Therefore, additional screenings for NphB variants have to be performed in order increase not only

specific activity of the enzyme but also product specificity towards CBGA formation. Availability of a crystal structure of NphB co-crystallized with magnesium, GPPS and OA would facilitate identification of potential amino acids exchanges in the active center of NphB which could change NphB's specificity towards CBGA. In this regard, the crystallization of the wildtype enzyme and a NphB variant with a higher CBGA formation rate than NphB wt should be taken into consideration as the X-ray structure of NphB co-crystallized with 1,6-DHN suggests significant conformational fluctuations of the aromatic acceptor molecule within the active site of the enzyme (Cui et al., 2007; Yang et al., 2012). As mentioned before, *E. coli* lysate supernatants supplemented with magnesium, 1,6-DHN and GPP resulted in the formation of three different mono-geranylated products. Thus, the crystal structure of NphB wt co-crystallized with OA would display the position of the aromatic molecule that leads to formation of the major product 2-*O*-GOA.

With regard to the development of a cannabinoid producing yeast the pH of the used enzyme activity assay buffer (pH 7.5) should be changed to a more acidic pH as the intracellular pH of *S. cerevisiae* varies significantly (pH 5 – 7), depending on the extracellular pH and the growth phase of the cells (Valli et al., 2005). Thus, the product specificity of NphB would be improved using conditions that prevail in the cytosol of yeast cells. Additionally, a low throughput screening system using *S. cerevisiae* as expression host should be established as the expression levels and enzymatic activities of *nphB* expressed in yeast are important for the overall objective.

4.3.   Expression of *nphB* and *thcas* in *S. cerevisiae*

As mentioned in chapter 1.1 the overall goal is the development of a cannabinoid producing yeast. The yeast *S. cerevisiae* is a well studied and characterized platform organism with a generally recognized as safe (GRAS) status. Various plant pathways leading to the opiods thebaine and hydrocodone (Galanie et al., 2015), the flavonoids naringenin and kaempferol (Rodriguez et al., 2017), the stilbene resveratrol or the antimalarial drug precursor artemisinin (Paddon and Keasling, 2014) could be successfully reconstituted in *S. cerevisiae*. In this context we tried to express *nphB* as part of the THCA biosynthetic pathway, originated in the plant *C. sativa*, in baker's yeast.

For the expression of NphB in *S. cerevisiae* different coding sequences were tested. We were able to show for the first time that functional expression of *nphB* in *S. cerevisiae* is possible. In addition to CBGA, 2-*O*-GOA was formed as the major side-product (~ 85%; cf. chapter 3.2.3). Mutational studies in *E. coli* (chapter 3.2.6) indicate that a rational protein engineering approach is able to increase the specificity of NphB towards CBGA. The highest product formation rate was obtained using a yeast codon-usage optimized coding sequence and a mutated pGAL1 for the expression of *nphB* (CBGA: 36 pmol mL^{-1} OD^{-1} h^{-1}, 2-*O*-GOA: 128 pmol mL^{-1} OD^{-1} h^{-1}). Until now, the mutational studies of NphB were only performed in the *E. coli*. In order to save time and money, a screening system for *S. cerevisiae* should be established.

Results of the whole cell assay indicate that yeast cell lysates have to be tested for activity because intact yeast cells are not able to take up OA and/or GPP or the uptake is at least not sufficient enough

to result in detectable product formation. As discussed in chapter 3.3.4 decreasing the pH of the cultivation medium might lead to an uptake of OA. Furthermore, whole cell biocatalysis might be possible upon expression of OAC and OLS (Gagne et al., 2012; Taura et al., 2009) as well as GPPS (Marks et al., 2009) while feeding hexanoate/malonate (cf. chapter 1.2.1).

The co-expression of *nphB* and *thcas* in *S. cerevisiae* simultaneously was tested using two different approaches. Neither the expression of both enzymes driven by a single GAL1 promoter, nor the use of the bidirectional pGAL10/pGAL1 promoter system of yeast yielded the formation of THCA after supplementing yeast cell lysates with magnesium, GPP and OA. *S. cerevisiae* is able to express the coding sequences of NphB and THCAS in a functional and active manner, but the obtained expression levels are too low to couple enzyme activities in order to generate yeast containing the whole THCA biosynthetic pathway. Interestingly, co-expression of *nphB* and *thcas* in *K. phaffii* yielded THCA formation (Zirpel, Degenhardt et al., 2017). A comparison of the results obtained for the activity assays performed with *S. cerevisiae* or *K. phaffii* lysates, respectively, indicates that in the baker's yeast expression system THCAS might the limiting factor. Clone C19 of *K. phaffii* formed comparable amounts of CBGA and 2-*O*-GOA like SC_NT (168 h after induction). This CBGA quantity was sufficient enough for THCAS expressed in *K. phaffii* to form THCA, while no THCA production was detectable in *S. cerevisiae* lysates (Zirpel, Degenhardt et al., 2017). As mentioned in chapter 1.2.3 eight possible N-glycosylation sites were confirmed for THCAS. One disadvantage of *S. cerevisiae* as host for heterologous protein expression is that N-linked sites are often hyperglycosylated, which might affect enzyme activity. In comparison to baker's yeast, *K. phaffii* uses oligosaccharides with shorter chain lengths and hyperglycosylation occurs less frequently (Darby et al., 2012; Vogl et al., 2013). Nevertheless, even though expression of *thcas* seems to be the bottleneck in the first place, enzyme activites of NphB and THCAS have to be enhanced by increasing the level of functional expression and/or the specific enzyme acitivities.

Zirpel et al. (2018a, 2018b) published that they were able to increase THCAS activity by protein engineering and co-expression of helper proteins FAD1p, CNE1p, Kar2p, PDI1p and Hac1p. FAD1p catalyzes the adenylation of FMN with adenosine triphosphate (ATP) to form FAD (Wu et al., 1995). Results of Taura et al. (2007a) indicate that the activities of the FAD-dependent oxygenase THCAS expressed in *K. phaffii* were limited due to an insuffcent FAD pool in the expression host. CNE1p is a calnexin homologue which acts as a molecular chaperone in the ER. The enzyme is involved in the quality control and proper folding of glycoproteins like THCAS (Xu, 2004). The leucine-zipper protein Hac1p is only detectable in ER-stressed cells as part of the UPR. UPR is activated when unfolded proteins accumulate in the lumen of the ER due to physiological or environmental stress conditions and induces upregulation of protein-folding catalysts and chaperones of the ER (Kawahara et al., 1997; Pal et al., 2006). The Hsp70 chaperone Kar2p is localized to the ER and facilitates the folding and transport of nascent peptides within this organell (Hale et al., 2010). *PDI1* encodes for a protein disulfide isomerase in the ER that catalyzes the formation of disulfide bonds (Farquhar et al., 1991). The coding sequences of the FAD synthetase FAD1p, the calnexin homologue CNE1p, the Hsp70 Kar2p, the protein disulfide isomerase PDI1p and the UPR-activator Hac1p (also known as IRE15p or

ERN4p) are known for *S. cerevisiae*. Therefore, it could be tested whether the overexpression of these helper proteins affects expression of *thcas* positively, as shown for *K. phaffii*. In addition, it should be evaluated on a case by case basis whether overexpression of further helper proteins like chaperones localized in the cytoplasm (Gong et al., 2009) could support co-expression of *nphB* and *thcas* in *S. cerevisiae*. Beside co-expression of helper proteins the use of less strong promoters or a low-copy number plasmid might facilitate higher levels of functionally expressed enzymes since the secretory pathway and the protein folding machinery are less overloaded (Darby et al., 2012).

Taking the total THCA biosynthetic pathway into consideration, the cultivation temperature of *S. cerevisiae* co-expressing *nphB* and *thcas* has to be increased. At the moment 15 °C are applied, but this temperature is not feasible for an industrial relevant biotechnological process. In addition, a NphB variant with a more acidic pH optimum is needed because the intracellular pH of *S. cerevisiae* varies significantly (pH 5 – 7), depending on the extracellular pH and the growth phase of the cells (Valli et al., 2005). The enzyme activity assays were performed at pH 7.5, but preliminary tests indicate that NphB has a more basic pH optimum (pH 8.5 – 9.5; [c], data not shown).

According to Carvalho et al. (2017) and Vogl et al. (2013), *K. phaffii* is more suitable for the heterologous expression of proteins with plant origin since higher protein levels are achievable due to multi-copy integration of the target coding sequence into the yeast genome. Additionally, *K. phaffii* can grow to very high cell densities and is classified as Crabtree negative. According to results published by Zirpel et al. (2015) and Zirpel, Degenhardt et al. (2017) higher levels of functional expression or enzymes with higher activity levels are obtained when using *K. phaffii* instead of *S. cerevisiae* as expression host for *nphB* and *thcas*. Furthermore, co-expression of both enzymes simultaneously in *K. phaffii* yielded the formation of up to 80 pmol mL^{-1} OD^{-1} h^{-1} THCA. In addition, product formations of up to 280 pmol mL^{-1} OD^{-1} h^{-1} were obtained when expressin *nphB* in this yeast (Zirpel, Degenhardt et al., 2017).

Even though *S. cerevisiae* is a well studied organism and extensive molecular biology tools are available, *K. phaffi* should be taken into account as alternative for developing a cannabinoid producing yeast. In the last couple of years several inducible and constitutive promoters have been described for the methylotrophic yeast that are necessary to implement a whole biosynthetic pathway in one organism (Ahmad et al., 2014; Kang et al., 2017; Vogl et al., 2013). Additionally, the Cre-loxP system and the clustered regularly interspaced short palindromic repeats (CRISPR) / Cas9, important for genome engineering (knocking-in, knocking out, site-directed mutagenesis), were successfully established for *K. phaffi* (Kang et al., 2017; Pan et al., 2011; Weninger et al., 2018, 2016). As described for *S. cerevisiae* multiple genes can be expressed under the control of one promoter by separating the coding sequences with 2A sequences (Schwarzhans et al., 2017). Additionally, the use of a bidirectional promoter system was published by Weninger et al. (2016). Even though, no or only a few publications are available regarding acetyl-CoA, GPP and hexanoic acid engineering in *K. phaffi*, the yeast might be a promising alternative for *S. cerevisiae* as chassis organism. In the last couple of years the interest in *K. phaffii* as expression host was increasing. Therefore, investigations regarding *inter alia* the metabolic pathways and metabolic engineering of the yeast will rise as well.

This study presented for the first time that the soluble, aromatic prenyltransferase NphB produces CBGA from precursors OA and GPP, enabling to circumvent expression of the membrane-bound CBGAS from C. *sativa*. Additionally, we showed that *nphB* and the final enzyme of the cannabinoid pathway, THCAS, can both be actively expressed in *S. cerevisiae*. The obtained results therefore promote the generation of a cannabinoid producing yeast for a biotechnological approach in order to face the increasing demand of these secondary metabolites in the medical sector.

Erklärung zur Reproduktion vorab veröffentlichter Inhalte

Teile dieser Arbeit sind bereits von der Autorin veröffentlicht und präsentiert worden:

Kapitel 1	1.2	in Teilen modifiziert aus	**[A]**
Kapitel 3	3.1.5	Figure 27	**[a]**
	3.3.2	in Teilen modifiziert aus	**[B]**
	3.3.3	in Teilen modifiziert aus	**[B]**

Publikationen

[A] Degenhardt, F., Stehle, F., Kayser, O., 2017. The biosynthesis of cannabinoids, in: Preedy, V. (Ed.), Handbook of Cannabis and Related Pathologies: Biology, Pharmacology, Diagnosis, and Treatment. Elsevier B.V., pp. 13–23.

[B] Zirpel, B.[1], **Degenhardt, F.**[1], Martin, C., Kayser, O., Stehle, F., 2017. Engineering yeasts as platform organisms for cannabinoid biosynthesis. J. Biotechnol. 259, 204–212.

[1] Contributed equally to this study.

Studentische Arbeiten

[a] Degenhardt, F., 2013, "Synthetic biology for the production of tetrahydrocannabinolic acid (THCA) from *Cannabis sativa* in *Saccharomyces cerevisiae*", Masterarbeit, TU Dortmund, Dortmund.

[b] Hansmeyer, R., 2015, „Erstellung und Charakterisierung von NphB-Varianten hinsichtlich der Substratspezifität", Bachelorarbeit, TU Dortmund, Dortmund.

[c] Kares, M. L., 2017, "Design and characterization of NphB variants for an optimized cannabigerolic acid production", Masterarbeit, TU Dortmund, Dortmund.

[d] Estrela, J., 2016, "Studies on the uptake of olivetol and olivetolic acid by *Saccharomyces cerevisiae*", Masterarbeit, TU Dortmund, Dortmund

References

Ahmad, M., Hirz, M., Pichler, H., Schwab, H., 2014. Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. Appl. Microbiol. Biotechnol. 98, 5301–5317. https://doi.org/10.1007/s00253-014-5732-5

Akashi, T., Sasaki, K., Aoki, T., Ayabe, S. -i., Yazaki, K., 2008. Molecular Cloning and Characterization of a cDNA for Pterocarpan 4-Dimethylallyltransferase Catalyzing the Key Prenylation Step in the Biosynthesis of Glyceollin, a Soybean Phytoalexin. PLANT Physiol. 149, 683–693. https://doi.org/10.1104/pp.108.123679

Altschul, S., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Awan, A.R., Shaw, W.M., Ellis, T., 2016. Biosynthesis of therapeutic natural products using synthetic biology. Adv. Drug Deliv. Rev. 105, 96–106. https://doi.org/10.1016/j.addr.2016.04.010

Baker Brachmann, C., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., Boeke, J.D., 1998. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast 14, 115–132. https://doi.org/10.1002/(SICI)1097-0061(19980130)14:2<115::AID-YEA204>3.0.CO;2-2

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D.S., Milo, R., 2011. The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. Biochemistry 50, 4402–4410. https://doi.org/10.1021/bi2002289

Bashir, H., Ahmed, N., Zafar, A.U., Khan, M.A., Tahir, S., Khan, M.I., Khan, F., Husnain, T., 2015. Simple procedure applying lactose induction and one step purification for high-yield production of rhCIFN. Biotechnol. Appl. Biochem. https://doi.org/10.1002/bab.1426

Beekwilder, J., van Rossum, H.M., Koopman, F., Sonntag, F., Buchhaupt, M., Schrader, J., Hall, R.D., Bosch, D., Pronk, J.T., van Maris, A.J.A., Daran, J.-M., 2014. Polycistronic expression of a β-carotene biosynthetic pathway in *Saccharomyces cerevisiae* coupled to β-ionone production. J. Biotechnol. 192, 383–392. https://doi.org/10.1016/j.jbiotec.2013.12.016

Bell, M.R., Engleka, M.J., Malik, A., Strickler, J.E., 2013. To fuse or not to fuse: What is your purpose? Protein Sci. 22, 1466–1477. https://doi.org/10.1002/pro.2356

Benetka, W., Koranda, M., Eisenhaber, F., 2006. Protein Prenylation: An (Almost) Comprehensive Overview on Discovery History, Enzymology, and Significance in Physiology and Disease. Monatshefte für Chemie / Chem. Mon. 137, 1241. https://doi.org/10.1007/s00706-006-0534-9

Bernaudat, F., Frelet-Barrand, A., Pochon, N., Dementin, S., Hivin, P., Boutigny, S., Rioux, J.-B., Salvi, D., Seigneurin-Berny, D., Richaud, P., Joyard, J., Pignol, D., Sabaty, M., Desnos, T., Pebay-Peyroula, E., Darrouzet, E., Vernet, T., Rolland, N., 2011. Heterologous Expression of

Membrane Proteins: Choosing the Appropriate Host. PLoS One 6, e29191. https://doi.org/10.1371/journal.pone.0029191

Bick, J.A., Lange, B.M., 2003. Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. Arch. Biochem. Biophys. 415, 146–154. https://doi.org/10.1016/S0003-9861(03)00233-9

Billinton, N., Knight, A.W., 2001. Seeing the Wood through the Trees: A Review of Techniques for Distinguishing Green Fluorescent Protein from Endogenous Autofluorescence. Anal. Biochem. 291, 175–197. https://doi.org/https://doi.org/10.1006/abio.2000.5006

Boeke, J.D., Lacroute, F., Fink, G.R., 1984. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol. Gen. Genet. https://doi.org/10.1007/BF00330984

Bonitz, T., Alva, V., Saleh, O., Lupas, A.N., Heide, L., 2011. Evolutionary relationships of microbial aromatic prenyltransferases. PLoS One 6, e27336. https://doi.org/10.1371/journal.pone.0027336

Booth, J.K., Page, J.E., Bohlmann, J., 2017. Terpene synthases from *Cannabis sativa*. PLoS One 12, e0173911. https://doi.org/10.1371/journal.pone.0173911

Botta, B., Monache, G.D., Menendez, P., Boffi, A., 2005. Novel prenyltransferase enzymes as a tool for flavonoid prenylation. Trends Pharmacol. Sci. 26, 606–608. https://doi.org/10.1016/j.tips.2005.09.012

Bouvier, F., Suire, C., D'Harlingue, A., Backhaus, R.A., Camara, B., 2000. Molecular cloning of geranyl diphosphate synthase and compartmentation of monoterpene synthesis in plant cells. Plant J. 24, 241–252. https://doi.org/10.1046/j.1365-313x.2000.00875.x

Brandt, W., Bräuer, L., Günnewich, N., Kufka, J., Rausch, F., Schulze, D., Schulze, E., Weber, R., Zakharova, S., Wessjohann, L., 2009. Molecular and structural basis of metabolic diversity mediated by prenyldiphosphate converting enzymes. Phytochemistry. https://doi.org/10.1016/j.phytochem.2009.09.001

Bräuer, L., Brandt, W., Schulze, D., Zakharova, S., Wessjohann, L., 2008. A Structural Model of the Membrane-Bound Aromatic Prenyltransferase UbiA from *E. coli*. ChemBioChem 9, 982–992. https://doi.org/10.1002/cbic.200700575

Briand, L., Marcion, G., Kriznik, A., Heydel, J.M., Artur, Y., Garrido, C., Seigneuric, R., Neiers, F., 2016. A self-inducible heterologous protein expression system in *Escherichia coli*. Sci. Rep. 6, 33037. https://doi.org/10.1038/srep33037

Burke, C.C., Wildung, M.R., Croteau, R., 1999. Geranyl diphosphate synthase: Cloning, expression, and characterization of this prenyltransferase as a heterodimer. Proc. Natl. Acad. Sci. U. S. A. 96, 13062–13067.

Burstein, S., 2015. Cannabidiol (CBD) and its analogs: a review of their effects on inflammation.

Bioorg. Med. Chem. 23, 1377–1385. https://doi.org/10.1016/j.bmc.2015.01.059

Cartwright, C.P., Li, Y., Zhu, Y.-S., Kang, Y.-S., Tipper, D.J., 1994. Use of β-lactamase as a secreted reporter of promoter function in yeast. Yeast 10, 497–508. https://doi.org/10.1002/yea.320100409

Carvalho, Â., Hansen, E.H., Kayser, O., Carlsen, S., Stehle, F., 2017. Designing microorganisms for heterologous biosynthesis of cannabinoids. FEMS Yeast Res. 17, fox037-fox037.

Casal, M., Paiva, S., Queirós, O., Soares-Silva, I., 2008. Transport of carboxylic acids in yeasts. FEMS Microbiol. Rev. 32, 974–994. https://doi.org/10.1111/j.1574-6976.2008.00128.x

Casey, P.J., Seabra, M.C., 1996. Protein Prenyltransferases. J. Biol. Chem. 271, 5289–5292. https://doi.org/10.1074/jbc.271.10.5289

Chakravorty, D.K., Merz, K.M., 2015. Role of Substrate Dynamics in Protein Prenylation Reactions. Acc. Chem. Res. 48, 439–448. https://doi.org/10.1021/ar500321u

Chang, M.C.Y., Keasling, J.D., 2006. Production of isoprenoid pharmaceuticals by engineered microbes. Nat. Chem. Biol. 2, 674–81. https://doi.org/10.1038/nchembio836

Chen, H., Kim, H.U., Weng, H., Browse, J., 2011. Malonyl-CoA Synthetase, Encoded by ACYL ACTIVATING ENZYME13, Is Essential for Growth and Development of Arabidopsis. Plant Cell 23, 2247–2262. https://doi.org/10.1105/tpc.111.086140

Cheng, W., Li, W., 2014. Structural Insights into Ubiquinone Biosynthesis in Membranes. Science (80-.). 343, 878–881. https://doi.org/10.1126/science.1246774

Collakova, E., DellaPenna, D., 2001. Isolation and Functional Analysis of Homogentisate Phytyltransferase from *Synechocystis* sp. PCC 6803 and Arabidopsis. PLANT Physiol. 127, 1113–1124. https://doi.org/10.1104/pp.010421

Cormack, B., Bertram, G., Egerton, M., Gow, N., Falkow, S., Brown, A., 1997. Yeast-enhanced green fluorescent protein (yEGFP): A reporter of gene expression in *Candida albicans*, Microbiology (Reading, England). https://doi.org/10.1099/00221287-143-2-303

Cormack, B.P., Valdivia, R.H., Falkow, S., 1996. FACS-optimized mutants of the green fluorescent protein (GFP). Gene 173, 33–38. https://doi.org/10.1016/0378-1119(95)00685-0

Crombie, L., Ponsford, R., Shani, A., Yagnitinsky, B., Mechoulam, R., 1968. Hashish components. Photochemical production of cannabicyclol from cannabichromene. Tetrahedron Lett. 55, 5771–5772.

Cui, G., Li, X., Merz, K.M., 2007. Understanding the Substrate Selectivity and the Product Regioselectivity of Orf2-Catalyzed Aromatic Prenylations. Biochemistry 46, 1303–1311. https://doi.org/10.1021/bi062076z

Darby, R.A.J., Cartwright, S.P., Dilworth, M. V, Bill, R.M., 2012. Recombinant Protein Production in Yeast, Methods in Molecular Biology. Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-

61779-770-5

de Felipe, P., Hughes, L.E., Ryan, M.D., Brown, J.D., 2003. Co-translational, Intraribosomal Cleavage of Polypeptides by the Foot-and-mouth Disease Virus 2A Peptide. J. Biol. Chem. 278, 11441–11448. https://doi.org/10.1074/jbc.M211644200

Degenhardt, F., Stehle, F., Kayser, O., 2017. The biosynthesis of cannabinoids, in: Preedy, V. (Ed.), Handbook of Cannabis and Related Pathologies: Biology, Pharmacology, Diagnosis, and Treatment. Elsevier B.V., pp. 13–23. https://doi.org/http://dx.doi.org/10.1016/B978-0-12-800756-3.00002-8

Degenhardt, J., Köllner, T.G., Gershenzon, J., 2009. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. Phytochemistry 70, 1621–1637. https://doi.org/10.1016/j.phytochem.2009.07.030

Dong, L., Jongedijk, E., Bouwmeester, H., Van Der Krol, A., 2016. Monoterpene biosynthesis potential of plant subcellular compartments. New Phytol. 209, 679–690. https://doi.org/10.1111/nph.13629

Donnelly, M.L.L., Hughes, L.E., Luke, G., Mendoza, H., ten Dam, E., Gani, D., Ryan, M.D., 2001a. The 'cleavage' activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring '2A-like' sequences. J. Gen. Virol. 82, 1027–1041. https://doi.org/10.1099/0022-1317-82-5-1027

Donnelly, M.L.L., Luke, G., Mehrotra, A., Li, X., Hughes, L.E., Gani, D., Ryan, M.D., 2001b. Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip.' J. Gen. Virol. 82, 1013–1025. https://doi.org/10.1099/0022-1317-82-5-1013

Drew, D., Newstead, S., Sonoda, Y., Kim, H., von Heijne, G., Iwata, S., 2008. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. Nat. Protoc. 3, 784–798.

Drew, D.E., von Heijne, G., Nordlund, P., de Gier, J.-W.L., 2001. Green fluorescent protein as an indicator to monitor membrane protein overexpression in *Escherichia coli*. FEBS Lett. 507, 220–224. https://doi.org/10.1016/S0014-5793(01)02980-5

Dunbrack, R.L., 2002. Rotamer Libraries in the 21st Century. Curr. Opin. Struct. Biol. 12, 431–440. https://doi.org/10.1016/S0959-440X(02)00344-5

Dyson, M.R., Shadbolt, S.P., Vincent, K.J., Perera, R.L., McCafferty, J., 2004. Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. BMC Biotechnol. 4, 32. https://doi.org/10.1186/1472-6750-4-32

Eisenreich, W., Bacher, A., Arigoni, D., Rohdich, F., 2004. Biosynthesis of isoprenoids via the non-mevalonate pathway. Cell. Mol. Life Sci. C. 61, 1401–1426. https://doi.org/10.1007/s00018-004-3381-z

ElSohly, M.A., Slade, D., 2005. Chemical constituents of marijuana: The complex mixture of natural

cannabinoids. Life Sci. 78, 539–548. https://doi.org/10.1016/J.LFS.2005.09.011

Emanuelsson, O., Brunak, S., von Heijne, G., Nielsen, H., 2007. Locating proteins in the cell using TargetP, SignalP and related tools. Nat. Protoc. 2, 953–971. https://doi.org/10.1038/nprot.2007.131

Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G., 2000. Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. J. Mol. Biol. 300, 1005–1016. https://doi.org/10.1006/jmbi.2000.3903

Emanuelsson, O., Nielsen, H., Heijne, G. Von, 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci. 8, 978–984. https://doi.org/10.1110/ps.8.5.978

Erhart, E., Hollenberg, C.P., 1983. The presence of a defective *LEU2* gene on 2 μ DNA recombinant plasmids of *Saccharomyces cerevisiae* is responsible for curing and high copy number. J. Bacteriol. 156, 625–635.

Estrela, J., 2016. Studies on the uptake of olivetol and olivetolic acid by *Saccharomyces cerevisiae*. TU Dortmund.

Farquhar, R., Honey, N., Murant, S.J., Bossier, P., Schultz, L., Montgomery, D., Ellis, R.W., Freedman, R.B., Tuite, M.F., 1991. Protein disulfide isomerase is essential for viability in *Saccharomyces cerevisiae*. Gene 108, 81–89. https://doi.org/10.1016/0378-1119(91)90490-3

Fellermeier, M., Eisenreich, W., Bacher, A., Zenk, M.H., 2001. Biosynthesis of cannabinoids. Eur. J. Biochem. 268, 1596–1604. https://doi.org/10.1046/j.1432-1327.2001.02030.x

Fellermeier, M., Zenk, M.H., 1998. Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol. FEBS Lett. 427, 283–5.

Fischedick, J.T., Hazekamp, A., Erkelens, T., Choi, Y.H., Verpoorte, R., 2010. Metabolic fingerprinting of *Cannabis sativa* L., cannabinoids and terpenoids for chemotaxonomic and drug standardization purposes. Phytochemistry 71, 2058–2073. https://doi.org/10.1016/j.phytochem.2010.10.001

Fischer, M.J.C., Meyer, S., Claudel, P., Bergdoll, M., Karst, F., 2011. Metabolic engineering of monoterpene synthesis in yeast. Biotechnol. Bioeng. 108, 1883–1892. https://doi.org/10.1002/bit.23129

Forsgren, M., Attersand, A., Lake, S., Grünler, J., Swiezeska, E., Dallner, G., Climent, I., 2004. Isolation and functional expression of human *COQ2*, a gene encoding a polyprenyl transferase involved in the synthesis of CoQ. Biochem. J. 382, 519 LP-526.

Frère, J.-M., 1995. Beta-lactamases and bacterial resistance to antibiotics. Mol. Microbiol. 16, 385–395. https://doi.org/10.1111/j.1365-2958.1995.tb02404.x

Fruchtl, M., Sakon, J., Beitle, R., 2015. Expression of a collagen-binding domain fusion protein: Effect of amino acid supplementation, inducer type, and culture conditions. Biotechnol. Prog. 31, 503–

509. https://doi.org/10.1002/btpr.2048

Fuhrmann, M., Hausherr, A., Ferbitz, L., Schödl, T., Heitzer, M., Hegemann, P., 2004. Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene. Plant Mol. Biol. 55, 869–881. https://doi.org/10.1007/s11103-005-2150-1

Gagne, S.J., Stout, J.M., Liu, E., Boubakir, Z., Clark, S.M., Page, J.E., 2012. Identification of olivetolic acid cyclase from *Cannabis sativa* reveals a unique catalytic route to plant polyketides. Proc. Natl. Acad. Sci. 109, 12811–12816. https://doi.org/10.1073/pnas.1200330109

Galanie, S., Thodey, K., Trenchard, I.J., Filsinger Interrante, M., Smolke, C.D., 2015. Complete biosynthesis of opioids in yeast. Science (80-.). 349, 1095–1100. https://doi.org/10.1126/science.aac9373

Gancedo, J.M., 1998. Yeast carbon catabolite repression. Microbiol Mol Biol Rev 62, 334–361.

Gaoni, Y., Mechoulam, R., 1971. Isolation and structure of D1- tetrahydrocannabinol and other neutral cannabinoids from hashish. J. Am. Chem. Soc. 93, 217–224. https://doi.org/10.1021/ja00730a036

Gertsch, J., Leonti, M., Raduner, S., Racz, I., Chen, J.-Z., Xie, X.-Q., Altmann, K.-H., Karsak, M., Zimmer, A., 2008. Beta-caryophyllene is a dietary cannabinoid. Proc. Natl. Acad. Sci. 105, 9099–9104. https://doi.org/10.1073/pnas.0803601105

Gertsch, J., Pertwee, R.G., Di Marzo, V., 2010. Phytocannabinoids beyond the Cannabis plant - do they exist? Br. J. Pharmacol. 160, 523–529. https://doi.org/10.1111/j.1476-5381.2010.00745.x

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Meth 6, 343–345.

Gietz, R.D., Schiestl, R.H., 2007. Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. 2, 1–4. https://doi.org/10.1038/nprot.2007.17

Glover, L.A., Lindsaytt, J.G., 1992. Targeting proteins to mitochondria: a current overview. Biochem. J 284, 609–620.

Gong, Y., Kakihara, Y., Krogan, N., Greenblatt, J., Emili, A., Zhang, Z., Houry, W.A., 2009. An atlas of chaperone–protein interactions in *Saccharomyces cerevisiae*: implications to protein folding pathways in the cell. Mol. Syst. Biol. 5. https://doi.org/10.1038/msb.2009.26

Grodowska, K., Parczewski, A., 2010. Organic solvents in the pharmaceutical industry. Acta Pol. Pharm. 67, 3–12.

Grossman, T.H., Kawasaki, E.S., Punreddy, S.R., Osburne, M.S., 1998. Spontaneous cAMP-dependent derepression of gene expression in stationary phase plays a role in recombinant expression instability. Gene 209, 95–103. https://doi.org/10.1016/S0378-1119(98)00020-1

References

Güldener, U., Heck, S., Fielder, T., Beinhauer, J., Hegemann, J.H., 1996. A new efficient gene disruption cassette for repeated use in budding yeast. Nucleic Acids Res. 24, 2519–2524.

Gutensohn, M., Orlova, I., Nguyen, T.T.H., Davidovich-Rikanati, R., Ferruzzi, M.G., Sitrit, Y., Lewinsohn, E., Pichersky, E., Dudareva, N., 2013. Cytosolic monoterpene biosynthesis is supported by plastid-generated geranyl diphosphate substrate in transgenic tomato fruits. Plant J. 75, 351–363. https://doi.org/10.1111/tpj.12212

Haddadin, F.T., Harcum, S.W., 2005. Transcriptome profiles for high-cell-density recombinant and wild-type *Escherichia coli*. Biotechnol. Bioeng. 90, 127–153. https://doi.org/10.1002/bit.20340

Hale, S.J., Lovell, S.C., de Keyzer, J., Stirling, C.J., 2010. Interactions between Kar2p and Its Nucleotide Exchange Factors Sil1p and Lhs1p Are Mechanistically Distinct. J. Biol. Chem. 285, 21600–21606. https://doi.org/10.1074/jbc.M110.111211

Haug-Schifferdecker, E., Arican, D., Brückner, R., Heide, L., 2010. A new group of aromatic prenyltransferases in fungi, catalyzing a 2,7-dihydroxynaphthalene 3-dimethylallyl-transferase reaction. J. Biol. Chem. 285, 16487–16494. https://doi.org/10.1074/jbc.M110.113720

He, Y., Wang, K., Yan, N., 2014. The recombinant expression systems for structure determination of eukaryotic membrane proteins. Protein Cell 5, 658–672. https://doi.org/10.1007/s13238-014-0086-4

Heide, L., 2009. Prenyl transfer to aromatic substrates: genetics and enzymology. Curr. Opin. Chem. Biol. 13, 171–179. https://doi.org/10.1016/j.cbpa.2009.02.020

Hendriks, H., Malingré, T.M., Batterman, S., Bos, R., 1975. Mono- and sesqui-terpene hydrocarbons of the essential oil of *Cannabis sativa*. Phytochemistry 14, 814–815. https://doi.org/10.1016/0031-9422(75)83045-7

Hohmann, S., Winderickx, J., de Winde, J.H. De, Valckx, D., Cobbaert, P., Luyten, K., Meirsman, C. De, Rarnos, J., Thevelein, J.M., 1999. Novel alleles of yeast hexokinase PII with distinct effects on catalytic activity and catabolite repression of *SUC2*. Microbiology 145, 703–714.

Horvath, S.E., Daum, G., 2013. Lipids of mitochondria. Prog. Lipid Res. 52, 590–614. https://doi.org/10.1016/j.plipres.2013.07.002

Hunter, W.N., 2007. The Non-mevalonate Pathway of Isoprenoid Precursor Biosynthesis. J. Biol. Chem. 282, 21573–21577. https://doi.org/10.1074/jbc.R700005200

Inoue, H., Nojima, H., Okayama, H., 1990. High efficiency transformation of *Escherichia coli* with plasmids. Gene 96, 23–28. https://doi.org/https://doi.org/10.1016/0378-1119(90)90336-P

Jamalzadeh, E., Verheijen, P.J.T., Heijnen, J.J., van Gulik, W.M., 2012. pH-Dependent Uptake of Fumaric Acid in *Saccharomyces cerevisiae* under Anaerobic Conditions. Appl. Environ. Microbiol. 78, 705–716. https://doi.org/10.1128/AEM.05591-11

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., Madden, T.L., 2008. NCBI BLAST: a better web interface. Nucleic Acids Res. 36, W5–W9.

References

https://doi.org/10.1093/nar/gkn201

Johnston, M., Davis, R.W., 1984. Sequences that regulate the divergent GAL1-GAL10 promoter in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 4, 1440–1448. https://doi.org/10.1128/MCB.4.8.1440

Johnston, M., Flick, J.S., Pexton, T., 1994. Multiple mechanisms provide rapid and stringent glucose repression of *GAL* gene expression in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 14, 3834–3841. https://doi.org/10.1128/MCB.14.6.3834.Updated

Kaishima, M., Ishii, J., Matsuno, T., Fukuda, N., Kondo, A., 2016. Expression of varied GFPs in *Saccharomyces cerevisiae*: codon optimization yields stronger than expected expression and fluorescence intensity. Sci. Rep. 6, 35932. https://doi.org/10.1038/srep35932

Kang, Z., Huang, H., Zhang, Y., Du, G., Chen, J., 2017. Recent advances of molecular toolbox construction expand *Pichia pastoris* in synthetic biology applications. World J. Microbiol. Biotechnol. 33, 19. https://doi.org/10.1007/s11274-016-2185-2

Kapp, K., Schrempf, S., Lemberg, M., Dobberstein, B., 2009. Post-targeting functions of signal peptides. Madame Curie Bioscience Database [Internet]. Austin (TX): Landes Bioscience; 2000-2013.

Kawahara, T., Yanagi, H., Yura, T., Mori, K., 1997. Endoplasmic Reticulum Stress-induced mRNA Splicing Permits Synthesis of Transcription Factor Hac1p/Ern4p That Activates the Unfolded Protein Response. Mol. Biol. Cell 8, 1845–1862. https://doi.org/10.1091/mbc.8.10.1845

Kayikci, Ö., Nielsen, J., 2015. Glucose repression in *Saccharomyces cerevisiae*. FEMS Yeast Res. 15, fov068. https://doi.org/10.1093/femsyr/fov068

Kim, J.H., Lee, S.-R., Li, L.-H., Park, H.-J., Park, J.-H., Lee, K.Y., Kim, M.-K., Shin, B.A., Choi, S.-Y., 2011. High Cleavage Efficiency of a 2A Peptide Derived from Porcine Teschovirus-1 in Human Cell Lines, Zebrafish and Mice. PLoS One 6, 1–8. https://doi.org/10.1371/journal.pone.0018556

Kimple, M.E., Brill, A.L., Pasker, R.L., 2013. Overview of Affinity Tags for Protein Purification, in: Current Protocols in Protein Science. John Wiley & Sons, Inc., Hoboken, NJ, USA, p. 9.9.1-9.9.23. https://doi.org/10.1002/0471140864.ps0909s73

Koehl, P., 2005. Relaxed specificity in aromatic prenyltransferases. Nat. Chem. Biol. 1, 71–72. https://doi.org/10.1038/nchembio0705-71

Kokina, A., Kibilds, J., Liepins, J., 2014. Adenine auxotrophy – be aware: some effects of adenine auxotrophy in *Saccharomyces cerevisiae* strain W303-1A. FEMS Yeast Res. 14, 697–707. https://doi.org/10.1111/1567-1364.12154

Konishi, T., Shinohara, K., Yamada, K., Sasaki, Y., 1996. Acetyl-CoA Carboxylase in Higher Plants: Most Plants Other Than Gramineae Have Both the Prokaryotic and the Eukaryotic Forms of This Enzyme. Plant Cell Physiol. 37, 117–122. https://doi.org/10.1093/oxfordjournals.pcp.a028920

Korpimäki, T., Kurittu, J., Karp, M., 2003. Surprisingly fast disappearance of β-lactam selection pressure in cultivation as detected with novel biosensing approaches. J. Microbiol. Methods 53,

37–42. https://doi.org/10.1016/S0167-7012(02)00213-0

Kosinski, M., Rinas, U., Bailey, J., 1992. Isopropyl-beta-d-thiogalactopyranoside influences the metabolism of *Escherichia coli*. Appl. Microbiol. Biotechnol. 36, 782–784. https://doi.org/10.1007/BF00172194

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L.., 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. J. Mol. Biol. 305, 567–580. https://doi.org/10.1006/jmbi.2000.4315

Kumano, T., Richard, S.B., Noel, J.P., Nishiyama, M., Kuzuyama, T., 2008. Chemoenzymatic syntheses of prenylated aromatic small molecules using *Streptomyces* prenyltransferases with relaxed substrate specificities. Bioorg. Med. Chem. 16, 8117–8126. https://doi.org/10.1016/j.bmc.2008.07.052

Kumano, T., Tomita, T., Nishiyama, M., Kuzuyama, T., 2010. Functional characterization of the promiscuous prenyltransferase responsible for furaquinocin biosynthesis: identification of a physiological polyketide substrate and its prenylated reaction products. J. Biol. Chem. 285, 39663–71. https://doi.org/10.1074/jbc.M110.153957

Kutchan, T.M., Dittrich, H., 1995. Characterization and Mechanism of the Berberine Bridge Enzyme, a Covalently Flavinylated Oxidase of Benzophenanthridine Alkaloid Biosynthesis in Plants. J. Biol. Chem. 270, 24475–24481. https://doi.org/10.1074/jbc.270.41.24475

Kuzuyama, T., Noel, J.P., Richard, S.B., 2005. Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products. Nature 435, 983–987. https://doi.org/10.1038/nature03668

Laemmli, U.K., 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature 227, 680–685. https://doi.org/10.1038/227680a0

Lanza, A.M., Curran, K.A., Rey, L.G., Alper, H.S., 2014. A condition-specific codon optimization approach for improved heterologous gene expression in *Saccharomyces cerevisiae*. BMC Syst. Biol. 8, 33. https://doi.org/10.1186/1752-0509-8-33

LaVallie, E.R., Lu, Z., Diblasio-Smith, E., Collins-Racie, L., McCoy, J.M., 2000. [21] Thioredoxin as a fusion partner for production of soluble recombinant proteins in *Escherichia coli*, in: Methods in Enzymology. pp. 322–340. https://doi.org/10.1016/S0076-6879(00)26063-1

Li, H., Ban, Z., Qin, H., Ma, L., King, A.J., Wang, G., 2015. A Heteromeric Membrane-Bound Prenyltransferase Complex from Hop Catalyzes Three Sequential Aromatic Prenylations in the Bitter Acid Pathway. Plant Physiol. 167, 650–659. https://doi.org/10.1104/pp.114.253682

Li, S.-M., 2010. Prenylated indole derivatives from fungi: structure diversity, biological activities, biosynthesis and chemoenzymatic synthesis. Nat. Prod. Rep. 27, 57–78. https://doi.org/10.1039/B909987P

Liang, P.-H., Ko, T.-P., Wang, A.H.-J., 2002. Structure, mechanism and function of prenyltransferases.

Eur. J. Biochem. 269, 3339–3354. https://doi.org/10.1046/j.1432-1033.2002.03014.x

Lin, H.-C., Yeh, C., Chen, Y., Lee, T., Hsieh, P., Rusnac, D. V, Lin, S., Elledge, S.J., Zheng, N., Yen, H.S., 2018. C-Terminal End-Directed Protein Elimination by CRL2 Ubiquitin Ligases. Mol. Cell 70, 602–613.e3. https://doi.org/10.1016/j.molcel.2018.04.006

Liu, H., Naismith, J.H., 2008. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotechnol. 8, 91. https://doi.org/10.1186/1472-6750-8-91

Liu, Z., Chen, O., Wall, J.B.J., Zheng, M., Zhou, Y., Wang, L., Ruth Vaseghi, H., Qian, L., Liu, J., 2017. Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. Sci. Rep. 7, 2193. https://doi.org/10.1038/s41598-017-02460-2

Loddenkötter, B., Kammerer, B., Fischer, K., Flügge, U.I., 1993. Expression of the functional mature chloroplast triose phosphate translocator in yeast internal membranes and purification of the histidine-tagged protein by a single metal-affinity chromatography step. Proc. Natl. Acad. Sci. 90, 2155–2159. https://doi.org/10.1073/pnas.90.6.2155

Maida, V., Daeninck, P.J., 2016. A user's guide to cannabinoid therapies in oncology. Curr. Oncol. Vol 23, No 6 (2016)DO - 10.3747/co.23.3487.

Malhotra, A., 2009. Chapter 16 Tagging for Protein Expression, in: Methods in Enzymology. Elsevier Inc., pp. 239–258. https://doi.org/10.1016/S0076-6879(09)63016-0

Marakasova, E.S., Akhmatova, N.K., Amaya, M., Eisenhaber, B., Eisenhaber, F., van Hoek, M.L., Baranova, A. V, 2013. Prenylation: From bacteria to eukaryotes. Mol. Biol. 47, 622–633. https://doi.org/10.1134/S0026893313050130

Marks, M.D., Tian, L., Wenger, J.P., Omburo, S.N., Soto-Fuentes, W., He, J., Gang, D.R., Weiblen, G.D., Dixon, R. a, 2009. Identification of candidate genes affecting Δ9-tetrahydrocannabinol biosynthesis in *Cannabis sativa*. J. Exp. Bot. 60, 3715–3726. https://doi.org/10.1093/jxb/erp210

Mårtensson, C.U., Doan, K.N., Becker, T., 2017. Effects of lipids on mitochondrial functions. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids 1862, 102–113. https://doi.org/10.1016/J.BBALIP.2016.06.015

Martoglio, B., Dobberstein, B., 1998. Signal sequences: more than just greasy peptides. Trends Cell Biol. 8, 410–415. https://doi.org/10.1016/S0962-8924(98)01360-9

Melzer, M., Heide, L., 1994. Characterization of Polyprenyldiphosphate: 4-Hydroxybenzoate Polyprenyltransferase from *Escherichia coli*. Biochim. Biophys. Acta - Lipids Lipid Metab. 1212, 93–102. https://doi.org/10.1016/0005-2760(94)90193-7

Mitchell, W., 2011. Natural products from synthetic biology. Curr. Opin. Chem. Biol. 15, 505–515. https://doi.org/10.1016/j.cbpa.2011.05.017

Morales, P., Hurst, D.P., Reggio, P.H., 2017. Molecular Targets of the Phytocannabinoids: A Complex Picture, in: Progress in the Chemistry of Organic Natural Products. pp. 103–131.

https://doi.org/10.1007/978-3-319-45541-9_4

Morimoto, S., Komatsu, K., Taura, F., Shoyama, Y., 1998. Purification and characterization of cannabichromenic acid synthase from *Cannabis sativa*. Phytochemistry 49, 1525–1529. https://doi.org/10.1016/S0031-9422(98)00278-7

Morimoto, S., Komatsu, K., Taura, F., Shoyama, Y., 1997. Enzymological Evidence for Cannabichromenic Acid Biosynthesis. J. Nat. Prod. 60, 854–857. https://doi.org/10.1021/np970210y

Moses, T., Pollier, J., Thevelein, J.M., Goossens, A., 2013. Bioengineering of plant (tri)terpenoids: from metabolic engineering of plants to synthetic biology *in vivo* and *in vitro*. New Phytol. 200, 27–43. https://doi.org/10.1111/nph.12325

Munakata, R., Inoue, T., Koeduka, T., Karamat, F., Olry, A., Sugiyama, A., Takanashi, K., Dugrand, A., Froelicher, Y., Tanaka, R., Uto, Y., Hori, H., Azuma, J.-I., Hehn, A., Bourgaud, F., Yazaki, K., 2014. Molecular Cloning and Characterization of a Geranyl Diphosphate-Specific Aromatic Prenyltransferase from Lemon. Plant Physiol. 166, 80–90. https://doi.org/10.1104/pp.114.246892

Nehlin, J.O., Carlberg, M., Ronne, H., 1991. Control of yeast *GAL* genes by MIG1 repressor: a transcriptional cascade in the glucose response. EMBO J. 10, 3373–3377.

Nguyen, F., Starosta, A.L., Arenz, S., Sohmen, D., Dönhöfer, A., Wilson, D.N., 2014. Tetracycline antibiotics and resistance mechanisms. Biol. Chem. 395, 559–575. https://doi.org/10.1515/hsz-2013-0292

Nielsen, H., Engelbrecht, J., Brunak, S., Heijne, G. Von, 1997. A Neural Network Method for Identification of Prokaryotic and Eukaryotic Signal Peptides and Prediction of their Cleavage Sites. Int. J. Neural Syst. 08, 581–599. https://doi.org/10.1142/S0129065797000537

Novack, G.D., 2016. Cannabinoids for treatment of glaucoma. Curr. Opin. Ophthalmol. 27, 146–150. https://doi.org/10.1097/ICU.0000000000000242

Oh, H., Seo, B., Lee, S., Ahn, D.-H., Jo, E., Park, J.-K., Min, G.-S., 2016. Two complete chloroplast genome sequences of *Cannabis sativa* varieties. Mitochondrial DNA Part A 27, 2835–2837. https://doi.org/10.3109/19401736.2015.1053117

Ohara, K., Kokado, Y., Yamamoto, H., Sato, F., Yazaki, K., 2004. Engineering of ubiquinone biosynthesis using the yeast *coq2* gene confers oxidative stress tolerance in transgenic tobacco. Plant J. 40, 734–743. https://doi.org/10.1111/j.1365-313X.2004.02246.x

Ohara, K., Mito, K., Yazaki, K., 2013. Homogeneous purification and characterization of LePGT1 - A membrane-bound aromatic substrate prenyltransferase involved in secondary metabolism of *Lithospermum erythrorhizon*. FEBS J. https://doi.org/10.1111/febs.12239

Ohara, K., Muroya, A., Fukushima, N., Yazaki, K., 2009. Functional characterization of LePGT1, a membrane-bound prenyltransferase involved in the geranylation of *p*-hydroxybenzoic acid. Biochem. J 421, 231–241. https://doi.org/10.1042/BJ20081968

References

Ohara, K., Yamamoto, K., Hamamoto, M., Sasaki, K., Yazaki, K., 2006. Functional characterization of OsPPT1, which encodes *p*-hydroxybenzoate polyprenyltransferase involved in ubiquinone biosynthesis in *Oryza sativa*. Plant Cell Physiol. https://doi.org/10.1093/pcp/pcj025

Oswald, M., Fischer, M., Dirninger, N., Karst, F., 2007. Monoterpenoid biosynthesis in *Saccharomyces cerevisiae*. FEMS Yeast Res. 7, 413–421. https://doi.org/10.1111/j.1567-1364.2006.00172.x

Paddon, C.J., Keasling, J.D., 2014. Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. Nat. Rev. Microbiol. 12, 355–367. https://doi.org/10.1038/nrmicro3240

Page, J.E., Boubakir, Z., 2014. Aromatic prenyltransferase from *Cannabis*. Patent no. US 8884100 B2.

Page, J.E., Gagne, S.J., 2013. Genes and proteins for aromatic polyketide synthesis. Patent no. US 20130067619 A1.

Page, J.E., Stout, J.M., 2015. Cannabichromenic acid synthase from *Cannabis sativa*. Patent no. US 20170211049 A1.

Page, J.E., Stout, J.M., 2013. Genes and proteins for alkanoyl-CoA synthesis. Patent No. WO 2013006953 A1.

Pal, B., Chan, N.C., Helfenbaum, L., Tan, K., Tansey, W.P., Gething, M.-J., 2006. SCFCdc4-mediated Degradation of the Hac1p Transcription Factor Regulates the Unfolded Protein Response in *Saccharomyces cerevisiae*. Mol. Biol. Cell 18, 426–440. https://doi.org/10.1091/mbc.E06-04-0304

Pamplaniyil, K., 2017. Identification, isolation and functional characterization of prenyltransferases in Cannabis sativa L. https://doi.org/10.17877/DE290R-18338

Pan, R., Zhang, J., Shen, W.-L., Tao, Z.-Q., Li, S.-P., Yan, X., 2011. Sequential deletion of *Pichia pastoris* genes by a self-excisable cassette. FEMS Yeast Res. 11, 292–298. https://doi.org/10.1111/j.1567-1364.2011.00716.x

Parker, J.L., Newstead, S., 2014. Method to increase the yield of eukaryotic membrane protein expression in *Saccharomyces cerevisiae* for structural and functional studies. Protein Sci. 23, 1309–1314. https://doi.org/10.1002/pro.2507

Partow, S., Siewers, V., Bjørn, S., Nielsen, J., Maury, J., 2010. Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. Yeast 27, 955–964. https://doi.org/10.1002/yea.1806

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera? A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. https://doi.org/10.1002/jcc.20084

Pojer, F., Wemakor, E., Kammerer, B., Chen, H., Walsh, C.T., Li, S.-M., Heide, L., 2003. CloQ, a prenyltransferase involved in clorobiocin biosynthesis. Proc. Natl. Acad. Sci. 100, 2316–2321.

https://doi.org/10.1073/pnas.0337708100

Raharjo, T.J., Chang, W.-T., Verberne, M.C., Peltenburg-Looman, A.M.G., Linthorst, H.J.M., Verpoorte, R., 2004. Cloning and over-expression of a cDNA encoding a polyketide synthase from *Cannabis sativa*. Plant Physiol. Biochem. 42, 291–297. https://doi.org/10.1016/j.plaphy.2004.02.011

Razdan, R.K., Puttick, A.J., Zitko, B.A., Handrick, G.R., 1972. Hashish VI1: Conversion of (-)-Δ1(6)-Tetrahydrocannabinol to (-)-Δ1(7)-Tetrahydrocannabinol. Stability of (-)-Δ1 and (-)-Δ1(6)-Tetrahydrocannabinols. Experientia 28, 121–122. https://doi.org/10.1007/BF01935704

Rodriguez, A., Strucko, T., Stahlhut, S.G., Kristensen, M., Svenssen, D.K., Forster, J., Nielsen, J., Borodina, I., 2017. Metabolic engineering of yeast for fermentative production of flavonoids. Bioresour. Technol. 245, 1645–1654. https://doi.org/10.1016/j.biortech.2017.06.043

Rodriguez, M.S., Desterro, J.M.P., Lain, S., Lane, D.P., Hay, R.T., 2000a. Multiple C-Terminal Lysine Residues Target p53 for Ubiquitin-Proteasome-Mediated Degradation. Mol. Cell. Biol. 20, 8458–8467. https://doi.org/10.1128/MCB.20.22.8458-8467.2000

Rodriguez, M.S., Desterro, J.M.P., Lain, S., Lane, D.P., Hay, R.T., 2000b. Multiple C-Terminal Lysine Residues Target p53 for Ubiquitin-Proteasome-Mediated Degradation. Mol. Cell. Biol. 20, 8458–8467. https://doi.org/10.1128/MCB.20.22.8458-8467.2000

Ryan, M.D., King, A.M.Q., Thomas, G.P., 1991. Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. J. Gen. Virol. 72, 2727–2732. https://doi.org/10.1099/0022-1317-72-11-2727

Saleh, O., Haagen, Y., Seeger, K., Heide, L., 2009. Prenyl transfer to aromatic substrates in the biosynthesis of aminocoumarins, meroterpenoids and phenazines: The ABBA prenyltransferase family. Phytochemistry 70, 1728–1738. https://doi.org/10.1016/j.phytochem.2009.05.009

Sasaki, K., Mito, K., Ohara, K., Yamamoto, H., Yazaki, K., 2008. Cloning and Characterization of Naringenin 8-Prenyltransferase, a Flavonoid-Specific Prenyltransferase of *Sophora flavescens*. Plant Physiol. 146, 1075–1084. https://doi.org/10.1104/pp.107.110544

Schmid, K.M., Ohlrogge, J.B., 2008. Lipid metabolism in plants, in: Vance Lipoproteins and Membranes (Fifth Edition), J.E.B.T.-B. of L. (Ed.), Biochemistry of Lipids, Lipoproteins and Membranes. Elsevier, San Diego, pp. 97–130. https://doi.org/10.1016/B978-044453219-0.50006-4

Schomburg, D., Salzmann, M., 1991. Prenyl-pyrophosphatase, in: Schomburg, D., Salzmann, M. (Eds.), Enzyme Handbook 3: Class 3: Hydrolases. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 715–717. https://doi.org/10.1007/978-3-642-76463-9_151

Schrot, R.J., Hubbard, J.R., 2016. Cannabinoids: Medical implications. Ann. Med. 48, 128–141. https://doi.org/10.3109/07853890.2016.1145794

Schultz, L.W., Chivers, P.T., Raines, R.T., 1999. The CXXC motif: crystal structure of an active-site

variant of *Escherichia coli* thioredoxin. Acta Crystallogr. Sect. D Biol. Crystallogr. 55, 1533–1538. https://doi.org/10.1107/S0907444999008756

Schwarzhans, J.-P., Luttermann, T., Geier, M., Kalinowski, J., Friehs, K., 2017. Towards systems metabolic engineering in *Pichia pastoris.* Biotechnol. Adv. 35, 681–710. https://doi.org/10.1016/j.biotechadv.2017.07.009

Sharma, N.K., Pan, J.-J., Poulter, C.D., 2010. Type II Isopentenyl Diphosphate Isomerase: Probing the Mechanism with Alkyne/Allene Diphosphate Substrate Analogues. Biochemistry 49, 6228–6233. https://doi.org/10.1021/bi100844e

Sheff, M.A., Thorn, K.S., 2004. Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae*. Yeast 21, 661–670. https://doi.org/10.1002/yea.1130

Shen, G., Huhman, D., Lei, Z., Snyder, J., Sumner, L.W., Dixon, R.A., 2012. Characterization of an Isoflavonoid-Specific Prenyltransferase from *Lupinus albus*. PLANT Physiol. 159, 70–80. https://doi.org/10.1104/pp.112.195271

Shoyama, Y., Fujita, T., Yamauchi, T., Nishioka, I., 1968. Cannabichromenic Acid, a Genuine Substance of Cannabichromene. Chem. Pharm. Bull. (Tokyo). 16, 1157–1158. https://doi.org/10.1248/cpb.16.1157

Shoyama, Y., Tamada, T., Kurihara, K., Takeuchi, A., Taura, F., Arai, S., Blaber, M., Shoyama, Y., Morimoto, S., Kuroki, R., 2012. Structure and Function of Δ1-Tetrahydrocannabinolic Acid (THCA) Synthase, the Enzyme Controlling the Psychoactivity of *Cannabis sativa*. J. Mol. Biol. 423, 96–105. https://doi.org/https://doi.org/10.1016/j.jmb.2012.06.030

Siddiqui, M.S., Thodey, K., Trenchard, I., Smolke, C.D., 2012. Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. FEMS Yeast Res. 12, 144–170. https://doi.org/10.1111/j.1567-1364.2011.00774.x

Sirikantaramas, S., Morimoto, S., Shoyama, Y., Ishikawa, Y., Wada, Y., Shoyama, Y., Taura, F., 2004. The Gene Controlling Marijuana Psychoactivity. J. Biol. Chem. 279, 39767–39774. https://doi.org/10.1074/jbc.M403693200

Sirikantaramas, S., Taura, F., 2017. Cannabinoids: Biosynthesis and Biotechnological Applications, in: Chandra, S., Lata, H., ElSohly, M.A. (Eds.), *Cannabis Sativa* L. - Botany and Biotechnology. Springer International Publishing, Cham, pp. 183–206. https://doi.org/10.1007/978-3-319-54564-6_8

Sirikantaramas, S., Taura, F., Morimoto, S., Shoyama, Y., 2007. Recent advances in *Cannabis sativa* research: Biosynthetic studies and its potential in biotechnology. Curr. Pharm. Biotechnol. 8. https://doi.org/10.2174/138920107781387456

Sirikantaramas, S., Taura, F., Tanaka, Y., Ishikawa, Y., Morimoto, S., Shoyama, Y., 2005. Tetrahydrocannabinolic acid synthase, the enzyme controlling marijuana psychoactivity, is secreted into the storage cavity of the glandular trichomes. Plant Cell Physiol. 46, 1578–82.

https://doi.org/10.1093/pcp/pci166

Skala, W., Goettig, P., Brandstetter, H., 2013. Do-it-yourself histidine-tagged bovine enterokinase: A handy member of the protein engineer's toolbox. J. Biotechnol. 168, 421–425. https://doi.org/10.1016/j.jbiotec.2013.10.022

Small, I., Peeters, N., Legeai, F., Lurin, C., 2004. Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics 4, 1581–1590. https://doi.org/10.1002/pmic.200300776

Sokalingam, S., Raghunathan, G., Soundrarajan, N., Lee, S.-G., 2012. A Study on the Effect of Surface Lysine to Arginine Mutagenesis on Protein Stability and Structure Using Green Fluorescent Protein. PLoS One 7, e40410. https://doi.org/10.1371/journal.pone.0040410

Sonnhammer, E., von Heijne, G., Krogh, A., 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol. 6, 175–182. https://doi.org/9783223

Spohner, S.C., Müller, H., Quitmann, H., Czermak, P., 2015. Expression of enzymes for the usage in food and feed industry with *Pichia pastoris*. J. Biotechnol. 202, 118–134. https://doi.org/10.1016/j.jbiotec.2015.01.027

Stec, E., Li, S.-M., 2012. Mutagenesis and biochemical studies on AuaA confirmed the importance of the two conserved aspartate-rich motifs and suggested difference in the amino acids for substrate binding in membrane-bound prenyltransferases. Arch. Microbiol. 194, 589–595. https://doi.org/10.1007/s00203-012-0795-0

Steffan, N., Grundmann, A., Yin, W.-B., Li, A.K. and S.-M., 2009. Indole Prenyltransferases from Fungi: A New Enzyme Group with High Potential for the Production of Prenylated Indole Derivatives. Curr. Med. Chem. https://doi.org/http://dx.doi.org/10.2174/092986709787002772

Stehle, F., Stubbs, M.T., Strack, D., Milkowski, C., 2008. Heterologous expression of a serine carboxypeptidase-like acyltransferase and characterization of the kinetic mechanism. FEBS J. 275, 775–87. https://doi.org/10.1111/j.1742-4658.2007.06244.x

Stout, J.M., Boubakir, Z., Ambrose, S.J., Purves, R.W., Page, J.E., 2012. The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in *Cannabis sativa* trichomes. Plant J. 71, no-no. https://doi.org/10.1111/j.1365-313X.2012.04949.x

Suzuki, Y., Kurano, M., Esumi, Y., Yamaguchi, I., Doi, Y., 2003. Biosynthesis of 5-alkylresorcinol in rice: incorporation of a putative fatty acid unit in the 5-alkylresorcinol carbon chain. Bioorg. Chem. 31, 437–452. https://doi.org/10.1016/J.BIOORG.2003.08.003

Taguchi, C., Taura, F., Tamada, T., Shoyama, Y., Shoyama, Y., Tanaka, H., Kuroki, R., Morimoto, S., 2008. Crystallization and preliminary X-ray diffraction studies of polyketide synthase-1 (PKS-1) from *Cannabis sativa*. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 64, 217–220. https://doi.org/10.1107/S1744309108003795

Taura, F., 1995. Cannabinerolic acid, a cannabinoid from *Cannabis sativa*. Phytochemistry 39, 457–458. https://doi.org/10.1016/0031-9422(94)00887-Y

Taura, F., Dono, E., Sirikantaramas, S., Yoshimura, K., Shoyama, Y., Morimoto, S., 2007a. Production of Δ1-tetrahydrocannabinolic acid by the biosynthetic enzyme secreted from transgenic *Pichia pastoris*. Biochem. Biophys. Res. Commun. 361, 675–680. https://doi.org/10.1016/j.bbrc.2007.07.079

Taura, F., Morimoto, S., Shoyama, Y., 1996. Purification and Characterization of Cannabidiolic-acid Synthase from *Cannabis sativa* L . J. Biol. Chem. 271, 17411–17416. https://doi.org/10.1074/jbc.271.29.17411

Taura, F., Morimoto, S., Shoyama, Y., Mechoulam, R., 1995. First direct evidence for the mechanism of D1-tetrahydrocannabinolic acid biosynthesis. J. Am. Chem. Soc. 117, 9766–9767. https://doi.org/10.1021/ja00143a024

Taura, F., Sirikantaramas, S., Shoyama, Y., Shoyama, Y., Morimoto, S., 2007b. Phytocannabinoids in *Cannabis sativa*: Recent Studies on Biosynthetic Enzymes. Chem. Biodivers. 4, 1649–1663. https://doi.org/10.1002/cbdv.200790145

Taura, F., Sirikantaramas, S., Shoyama, Y., Yoshikai, K., Shoyama, Y., Morimoto, S., 2007c. Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type *Cannabis sativa*. FEBS Lett. 581, 2929–34. https://doi.org/10.1016/j.febslet.2007.05.043

Taura, F., Tanaka, S., Taguchi, C., Fukamizu, T., Tanaka, H., Shoyama, Y., Morimoto, S., 2009. Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway. FEBS Lett. 583, 2061–6. https://doi.org/10.1016/j.febslet.2009.05.024

Tello, M., Kuzuyama, T., Heide, L., Noel, J.P., Richard, S.B., 2008. The ABBA family of aromatic prenyltransferases: broadening natural product diversity. Cell. Mol. Life Sci. 65, 1459–1463. https://doi.org/10.1007/s00018-008-7579-3

Terpe, K., 2003. Overview of tag protein fusions: From molecular and biochemical fundamentals to commercial systems. Appl. Microbiol. Biotechnol. 60, 523–533. https://doi.org/10.1007/s00253-002-1158-6

Tholl, D., 2015. Biosynthesis and Biological Functions of Terpenoids in Plants BT - Biotechnology of Isoprenoids, in: Schrader, J., Bohlmann, J. (Eds.), . Springer International Publishing, Cham, pp. 63–106. https://doi.org/10.1007/10_2014_295

Toyota, M., Kinugawa, T., Asakawa, Y., 1994. Bibenzyl cannabinoid and bisbibenzyl derivative from the liverwort *Radula perrottetii*. Phytochemistry 37, 859–862. https://doi.org/10.1016/S0031-9422(00)90371-6

Toyota, M., Shimamura, T., Ishii, H., Renner, M., Braggins, J., Asakawa, Y., 2002. New Bibenzyl Cannabinoid from the New Zealand Liverwort *Radula marginata*. Chem. Pharm. Bull. 50, 1390–1392. https://doi.org/10.1248/cpb.50.1390

References

Trost, B.M., Dogra, K., 2007. Synthesis of (-)-Δ 9 - trans -Tetrahydrocannabinol: Stereocontrol via Mo-Catalyzed Asymmetric Allylic Alkylation Reaction. Org. Lett. 9, 861–863. https://doi.org/10.1021/ol063022k

Tsai, H.F., Wang, H., Gebler, J.C., Poulter, C.D., Schardl, C.L., 1995. The *Claviceps purpurea* gene encoding dimethylallyltryptophan synthase, the committed step for ergot alkaloid biosynthesis. https://doi.org/10.1006/bbrc.1995.2599

Tsurumaru, Y., Sasaki, K., Miyawaki, T., Momma, T., Umemoto, N., Yazaki, K., 2010. An aromatic prenyltransferase-like gene HlPT-1 preferentially expressed in lupulin glands of hop. Plant Biotechnol. 27, 199–204. https://doi.org/10.5511/plantbiotechnology.27.199

Tsurumaru, Y., Sasaki, K., Miyawaki, T., Uto, Y., Momma, T., Umemoto, N., Momose, M., Yazaki, K., 2012. HlPT-1, a membrane-bound prenyltransferase responsible for the biosynthesis of bitter acids in hops. Biochem. Biophys. Res. Commun. 417, 393–398. https://doi.org/10.1016/j.bbrc.2011.11.125

Turunen, M., Olsson, J., Dallner, G., 2004. Metabolism and function of coenzyme Q. Biochim. Biophys. Acta - Biomembr. 1660, 171–199. https://doi.org/10.1016/j.bbamem.2003.11.012

Ugolini, S., Bruschi, C. V., 1996a. The red/white colony color assay in the yeast *Saccharomyces cerevisiae*: epistatic growth advantage of white ade8-18, ade2 cells over red ade2 cells. Curr. Genet. 30, 485–492. https://doi.org/10.1007/s002940050160

Ugolini, S., Bruschi, C. V., 1996b. The red/white colony color assay in the yeast *Saccharomyces cerevisiae*: epistatic growth advantage of white ade8-18, ade2 cells over red ade2 cells. Curr. Genet. 30, 485–492. https://doi.org/10.1007/s002940050160

Valli, M., Sauer, M., Branduardi, P., Borth, N., Porro, D., Mattanovich, D., 2005. Intracellular pH Distribution in *Saccharomyces cerevisiae* Cell Populations, Analyzed by Flow Cytometry. Appl. Environ. Microbiol. 71, 1515–1521. https://doi.org/10.1128/AEM.71.3.1515-1521.2005

van Bakel, H., Stout, J.M., Cote, A.G., Tallon, C.M., Sharpe, A.G., Hughes, T.R., Page, J.E., 2011. The draft genome and transcriptome of *Cannabis sativa*. Genome Biol. 12, R102. https://doi.org/10.1186/gb-2011-12-10-r102

Venkatesh, T. V, Karunanandaa, B., Free, D.L., Rottnek, J.M., Baszis, S.R., Valentin, H.E., 2006. Identification and characterization of an Arabidopsis homogentisate phytyltransferase paralog. Planta 223, 1134–1144. https://doi.org/10.1007/s00425-005-0180-1

Vogl, T., Hartner, F.S., Glieder, A., 2013. New opportunities by synthetic biology for biopharmaceutical production in *Pichia pastoris*. Curr. Opin. Biotechnol. 24, 1094–1101. https://doi.org/10.1016/J.COPBIO.2013.02.024

von Heijne, G., 1992. Membrane protein structure prediction. J. Mol. Biol. 225, 487–494. https://doi.org/10.1016/0022-2836(92)90934-C

von Heijne, G., 1986. The distribution of positively charged residues in bacterial inner membrane

proteins correlates with the trans-membrane topology. EMBO J. 5, 3021–3027. https://doi.org/10.1002/j.1460-2075.1986.tb04601.x

Weninger, A., Fischer, J.E., Raschmanová, H., Kniely, C., Vogl, T., Glieder, A., 2018. Expanding the CRISPR/Cas9 toolkit for *Pichia pastoris* with efficient donor integration and alternative resistance markers. J. Cell. Biochem. 119, 3183–3198. https://doi.org/10.1002/jcb.26474

Weninger, A., Hatzl, A., Schmid, C., Vogl, T., Glieder, A., 2016. Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*. J. Biotechnol. 235, 139–149. https://doi.org/10.1016/j.jbiotec.2016.03.027

West, R.W., Chen, S.M., Putz, H., Butler, G., Banerjee, M., 1987. GAL1-GAL10 divergent promoter region of *Saccharomyces cerevisiae* contains negative control elements in addition to functionally separate and possibly overlapping upstream activating sequences. Genes Dev. 1, 1118–1131. https://doi.org/10.1101/gad.1.10.1118

West, R.W., Yocum, R.R., Ptashne, M., 1984. *Saccharomyces cerevisiae* GAL1-GAL10 divergent promoter region: location and function of the upstream activating sequence UASG. Mol. Cell. Biol. 4, 2467–2478. https://doi.org/10.1128/MCB.4.11.2467

Winkelblech, J., Fan, A., Li, S.-M., 2015. Prenyltransferases as key enzymes in primary and secondary metabolism. Appl. Microbiol. Biotechnol. 99, 7379–7397. https://doi.org/10.1007/s00253-015-6811-y

Wölwer-Rieck, U., May, B., Lankes, C., Wüst, M., 2014. Methylerythritol and Mevalonate Pathway Contributions to Biosynthesis of Mono-, Sesqui-, and Diterpenes in Glandular Trichomes and Leaves of *Stevia rebaudiana Bertoni*. J. Agric. Food Chem. 62, 2428–2435. https://doi.org/10.1021/jf500270s

Woodside, A.B., Huang, Z., Poulter, C.D., Woodside, A.B., Huang, Z., Poulter, C.D., 1988. Trisammonium geranyl diphosphate. Org. Synth. 66, 211. https://doi.org/10.15227/orgsyn.066.0211

Worthington, Z.E. V, Carbonetti, N.H., 2007a. Evading the Proteasome: Absence of Lysine Residues Contributes to Pertussis Toxin Activity by Evasion of Proteasome Degradation. Infect. Immun. 75, 2946–2953. https://doi.org/10.1128/IAI.02011-06

Worthington, Z.E. V, Carbonetti, N.H., 2007b. Evading the Proteasome: Absence of Lysine Residues Contributes to Pertussis Toxin Activity by Evasion of Proteasome Degradation. Infect. Immun. 75, 2946–2953. https://doi.org/10.1128/IAI.02011-06

Wu, M., Repetto, B., Glerum, D.M., Tzagoloff, A., 1995. Cloning and characterization of FAD1, the structural gene for flavin adenine dinucleotide synthetase of *Saccharomyces cerevisiae*. Mol. Cell. Biol. 15, 264–271. https://doi.org/10.1128/MCB.15.1.264

Wurm, D.J., Veiter, L., Ulonska, S., Eggenreich, B., Herwig, C., Spadiut, O., 2016. The *E. coli* pET expression system revisited—mechanistic correlation between glucose and lactose uptake. Appl.

Microbiol. Biotechnol. 100, 8721–8729. https://doi.org/10.1007/s00253-016-7620-7

Xiao, Y., Machacek, M., Lee, K., Kuzuyama, T., Liu, P., 2009. Prenyltransferase substrate binding pocket flexibility and its application in isoprenoid profiling. Mol. Biosyst. 5, 913. https://doi.org/10.1039/b902370d

Xu, X., 2004. Expression and Characterization of *Saccharomyces cerevisiae* Cne1p, a Calnexin Homologue. J. Biochem. 135, 615–618. https://doi.org/10.1093/jb/mvh074

Yamamoto, H., Kimata, J., Senda, M., Inoue, K., 1997. Dimethylallyl diphosphate: Kaempferol 8-dimethylallyl transferase in *Epimedium diphyllum* cell suspension cultures. Phytochemistry 44, 23–28. https://doi.org/10.1016/S0031-9422(96)00484-0

Yamamoto, H., Senda, M., Inoue, K., 2000. Flavanone 8-dimethylallyltransferase in *Sophora flavescens* cell suspension cultures. Phytochemistry 54, 649–655. https://doi.org/10.1016/S0031-9422(00)00198-9

Yang, X., Matsui, T., Kodama, T., Mori, T., Zhou, X., Taura, F., Noguchi, H., Abe, I., Morita, H., 2016. Structural basis for olivetolic acid formation by a polyketide cyclase from *Cannabis sativa*. FEBS J. 283, 1088–1106. https://doi.org/10.1111/febs.13654

Yang, Y., Miao, Y., Wang, B., Cui, G., Merz, K.M., 2012. Catalytic Mechanism of Aromatic Prenylation by NphB. Biochemistry 51, 2606–2618. https://doi.org/10.1021/bi201800m

Yazaki, K., Kunihisa, M., Fujisaki, T., Sato, F., 2002. Geranyl Diphosphate:4-Hydroxybenzoate Geranyltransferase from *Lithospermum erythrorhizon*. J. Biol. Chem. 277, 6240–6246. https://doi.org/10.1074/jbc.M106387200

Yazaki, K., Sasaki, K., Tsurumaru, Y., 2009. Prenylation of aromatic compounds, a key diversification of plant secondary metabolites. Phytochemistry 70, 1739–1745. https://doi.org/http://dx.doi.org/10.1016/j.phytochem.2009.08.023

Zhang, F.L., Casey, P.J., 1996. Protein Prenylation: Molecular Mechanisms and Functional Consequences. Annu. Rev. Biochem. 65, 241–269. https://doi.org/10.1146/annurev.bi.65.070196.001325

Zhang, Z., Kuipers, G., Niemiec, Ł., Baumgarten, T., Slotboom, D.J., de Gier, J.-W., Hjelm, A., 2015. High-level production of membrane proteins in *E. coli* BL21(DE3) by omitting the inducer IPTG. Microb. Cell Fact. 14, 142. https://doi.org/10.1186/s12934-015-0328-z

Zhao, P., Inoue, K., Kouno, I., Yamamoto, H., 2003. Characterization of Leachianone G 2"-Dimethylallyltransferase, a Novel Prenyl Side-Chain Elongation Enzyme for the Formation of the Lavandulyl Group of Sophoraflavanone G in *Sophora flavescens Ait.* Cell Suspension Cultures. Plant Physiol. 133, 1306–1313. https://doi.org/10.1104/pp.103.025213

Zirpel, B., Degenhardt, F., Martin, C., Kayser, O., Stehle, F., 2017. Engineering yeasts as platform organisms for cannabinoid biosynthesis. J. Biotechnol. 259, 204–212. https://doi.org/10.1016/j.jbiotec.2017.07.008

Zirpel, B., Degenhardt, F., Zammarelli, C., Wibberg, D., Kalinowski, J., Stehle, F., Kayser, O., 2018a. Optimization of Δ 9 -tetrahydrocannabinolic acid synthase production in *Komagataella phaffii* via post-translational bottleneck identification. J. Biotechnol. 272–273, 40–47. https://doi.org/10.1016/j.jbiotec.2018.03.008

Zirpel, B., Kayser, O., Stehle, F., 2018b. Elucidation of structure-function relationship of THCA and CBDA synthase from *Cannabis sativa* L. J. Biotechnol. 284, 17–26. https://doi.org/10.1016/j.jbiotec.2018.07.031

Zirpel, B., Stehle, F., Kayser, O., 2015. Production of Δ9-tetrahydrocannabinolic acid from cannabigerolic acid by whole cells of *Pichia (Komagataella) pastoris* expressing Δ9-tetrahydrocannabinolic acid synthase from *Cannabis sativa* l. Biotechnol. Lett. 37, 1869–1875. https://doi.org/10.1007/s10529-015-1853-x

Zuurbier, K.W.M., Fung, S.Y., Scheffer, J.J.C., Verpoorte, R., 1998. *In-vitro* prenylation of aromatic intermediates in the biosynthesis of bitter acids in *Humulus lupulus*. Phytochemistry 49, 2315–2322. https://doi.org/10.1016/S0031-9422(98)00179-4

Supplementary information

List of abbreviations

2-O-GOA	2-*O*-geranyl olivetolic acid
5-FOA	5-fluoroorotic acid
AA	amino acid
AAE	acyl-activating enzyme
AcN	acetonitrile
ACP	acyl carrier protein
APS	ammonium persulfate
ATP	adenosine triphosphate
BBE	berberine-bridge enzyme
Bp	base pairs
CBC	cannabichromene
CBCA	cannabichromenic acid
CBCAS	cannabichromenic acid synthase
CBD	cannabidiol
CBDA	cannabidiolic acid
CBDAS	cannabichromenic acid
CBG	cannabigerol
CBGA	cannabigerolic acid
CBGAS	cannabigerolic acid synthase
CBNA	cannabinolic acid
CBNRA	cannabinerolic acid (*cis*-CBGA)
CDCl3	deuterated chloroform
cDNA	complementary DNA
cf.	Confer
CIAP	calf intestinal alkaline phosphatase
CoA	coenzyme A
CsHCS1, CsAAE1	*C. sativa* hexanoyl-CoA synthetase 1
CsHCS2, CsAAE3	*C. sativa* hexanoyl-CoA synthetase 2

CsPT1P	sequence of CsPT1 patented by Page and Boubakir (2014)
cTP	chloroplast transit peptide
CV	column volume
DAD	diode array detector
DHN	dihydroxynapthalene
DMAPP	dimethylallyl diphosphate
DMATS	dimethylallytryptophan synthase
DMSO	dimethyl sulfoxide
DOXP	1-deoxy-D-xylulose-5-phosphate
DTT	dithiothreitol
e.g.	lat.exempli gratia; for example
EIC	extracted ion chromatogram
ER	endoplasmic reticulum
ESI	electrospray ionization
EST	expressed sequence tag
FA	formic acid
FAD	flavine adenine dinucleotide
FI	fluorescence intensity
FMN	flavin mononucleotide
FPP	farnesyl diphosphate
FPPS	farnesyl diphosphate synthase
Fw	Forward
GFP	green fluorescent protein
GHB	geranyl hydroxybenzoate
GMP	good manufacturing practice
GOT	geranylpyrophosphate:olivetolate geranyltransferase
GPP	geranyl diphosphate
GPPS	Geranyl diphosphate synthase
GRAS	generally recognized as safe

GSPP	geranyl thiolodiphosphate
HMM	Hidden Markov model
HPLC	high performance liquid chromatoraphy
HTLA	hexanoyltriacetic acid lactone
i.e.	latin: id est; that is to say
IMAC	immobilized metal ion affinity chromatography
IPP	isopentenyl diphosphate
IPPS	isoprenyl diphosphate synthase
IPTG	isopropyl β-D-thiogalactopyranosid
k_{cat}	turnover rate
K_M	Michaelis constant
LC	liquid chromatography
LePGT1	geranyldiphosphate:4-hydroxybenzoate 3-geranyltransferase
LiAc	lithium acetate
m/z	mass-to-charge ratio
MEP pathway	2C-methyl-D-erythritol-4-phosphate pathway
MEV pathway	mevalonate pathway
MS	mass spectrometer / -metry
MS^2	tandem mass spectrometry
mTP	mitochondrial targeting peptides
MWCO	molecular weight cut-off
NAD	nicotinamide adenine dinucleotide
NADP	nicotinamide adenine dinucleotide phosphate
NMR	nuclear magnetic resonance spectroscopy
NPP	neryl diphosphate
OA	olivetolic acid
OAC	olivetolic acid cyclase
OD	optical density
OLS	olivetol synthase

PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
PHB	*p*-hydroxybenzoic acid
PKS	polyketide synthase
PP_i	inorganic diphosphate
PPT	*p*-hydroxybenzoate:polyprenyl transferase
PT	prenyltransferase
RP-LC	reverse-phase liquid chromatography
Rpm	rounds per minute
Rv	reverse
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SP	secretory pathway signal peptide
ss carrier DNA	single stranded carrier DNA
TEV	Tobacco Etch Virus
THC	tetrahydrocannabinol
THCA	tetrahydrocannabinolic acid
THCAS	tetrahydrocannabinolic acid synthase
TLC	thin layer chromatography
TMH	transmembrane helix
TOF	time of flight
UPR	unfolded protein response
UQ	ubiquitin
URS	upstream repression sequence
UTR	untranslated region
UV	ultraviolet
v/v	volume per volume
vs.	versus
w/v	weight per volume

w/w	weight per weight
wt	wildtype
yCoq2SP	yeast Coq2 signal peptide
yEGFP	yeast enhanced green fluorescent protein
YNB	yeast nitrogen base
YPD	yeast, peptone, dextrose

According to the IUB-IUPAC recommendations amino acid sequences were written in single-letter codes.

Expression of membrane-bound prenyltransferases in *S. cerevisiae*

Table S 1: Composition of yeast mineral salt medium drop out supplements. The supplements do not contain leucine (380 mg L^{-1}) and uracil (76 mg L^{-1}).

Component	Concentration [mg L^{-1}]
Adenine hemisulfate	18
Alanine	76
Arginine hydrochloride	76
Asparagine monohydrate	76
Aspartic acid	76
Cysteine hydrochloride monohydrate	76
Glutamic acid monosodium salt	76
Glutamine	76
Glycine	76
Histidine	76
myo-Inositol	76
Isoleucine	76
Lysine monohydrochloride	76
Methionine	76
p-Aminobenzoic acid potassium salt	8
Phenylalanine	76
Proline	76
Serine	76
Threonine	76
Tryptophan	76
Tyrosine disodium salt	76
Valine	76

Table S 2: Genotypes of the strains used for the expression of membrane-bound prenyltransferases.

Organism	Strain	Genotype	Reference / source
E. coli	DH5α	*F⁻ endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK⁻ mK⁺), λ⁻*	Invitrogen (Karlsruhe, Germany)
S. cerevisiae	W303-1A *Δpep4 Δgal4*	*MATa; pep4::kanMX; gal4::kanMX; leu2-3,112; trp1-1; can1-100; ura3-1; ade2-1; his3-11,15*	Kind gift of Prof. Poul Nissen, PUMPkin, Aarhus University, Denmark
S. cerevisiae	BY4742	*MATalpha; his3D1; leu2D0; ura3D0*	EUROSCARF (Y100000) (Baker Brachmann et al., 1998)

Table S 3: Amino acid sequences of the aromatic prenyltransferases used in this study. Predicted targeting sequences are marked in bold. The conserved amino acid sequence NQxxDxxxD responsible for prenyl diphosphate recognition in Mg²⁺-dependent prenyltransferase and the second aspartate-rich motif KDxxDxEGD is marked underlined and bold. [1] Sequence obtained from Pamplaniyil (2017), [2] Sequence obtained from Page and Boubakir (2014).

Name	Sequence
CsPT1[1]	**MGLSSVCTFSFQTNYHTLLNPHNNNPKTSLLYRHPKTPIKYSYNNFPSKHCSTKSFHLQNKCSE SLSIAKNSIR**AATTNQTEPPESDNHSVATKILNFGKACWKLQRPYTIIAFTSCACGLFGKELLHNTN LISWSLMFKAFFFLVAVLCIASFTTTI**NQ**IY**D**LHI**D**RINKPDLPLASGEISVNTAWIMSIIVALFGLIITIK MKGGPLYIFGYCFGIFGGTVYSVPPFRWKQNPSTAFLLNFLAHIITNFTFYHASRAALGLPFELRPS FTFLLAFMKSMGSALALI**KD**AS**D**V**EGD**TKFGISTLASKYGSRNLTLFCSGIVLLSYVAAILAGIIWPQ AFNSNVMLLSHAILAFWLILQTRDFALTNYDPEAGRRFYEFMWKLYYAEYLVYVFI
CsPT1P[2]	**MGLSSVCTFSFQTNYHTLLNPHNNNPKTSLLCYRHPKTPIKYSYNNFPSKHCSTKSFHLQNKCS ESLSIAKNSIR**AATTNQTEPPESDNHSVATKILNFGKACWKLQRPYTIIAFTSCACGLFGKELLHNT NLISWSLMFKAFFFLVAVLCIASFTTTI**NQ**IY**D**LHI**D**RINKPDLPLASGEISVNTAWIMSIIVALFGLIITI KMKGGPLYIFGYCFGIFGGIVYSVPPFRWKQNPSTAFLLNFLAHIITNFTFYYASRAALGLPFELRP SFTFLLAFMKSMGSALALI**KD**AS**D**V**EGD**TKFGISTLASKYGSRNLTLFCSGIVLLSYVAAILAGIIWP QAFNSNVMLLSHAILAFWLILQTRDFALTNYDPEAGRRFYEFMWKLYYAEYLVYVFI

Table S 3 continued

Name	Sequence
CsPT2[1]	**MELSSICNFSFQTNYHTLLNPHNKNPKSSLLSHQHPKTPIITSSYNNFPSNYCSNKNFHLQNRCS KSLLIAKNSIR**TDTANQTEPPESNTKYSVVTKILSFGHTCWKLQRPYTFIGVISCACGLFGRELFHN TNLLSWSLMLKAFSSLMVILSVNLCTNII**NQ**IT**D**LD**ID**RINKPDLPLASGEMSIETAWIMSIIVALTGLIL TIKLNCGPLFISLYCVSILVGALYSVPPFRWKQNPNTAFSSYFMGLVIVNFTCYYASRAAFGLPFEM SPPFTFILAFVKSMGSALFLC**KD**VS**DIEGD**SKHGISTLATRYGAKNITFLCSGIVLLTYVSAILAAIIWP QAFKSNVMLLSHATLAFWLIFQTREFALTNYNPEAGRKFYEFMWKLHYAEYLVYVFI
CsPT3[1]	**MGLSLVCTFSFQTNYHTLLNPHNKNPKNSLLSYQHPKTPIIKSSYDNFPSKYCLTKNFHLLGLNS HNRISSQSRSIR**AGSDQIEGSPHHESDNSIATKILNFGHTCWKLQRPYVVKGMISIACGLFGRELFN NRHLFSWGLMWKAFFALVPILSFNFFAAIM**NQ**IY**D**VD**ID**RINKPDLPLVSGEMSIETAWILSIIVALTG LIVTIKLKSAPLFVFIYIFGIFAGFAYSVPPIRWKQYPFTNFLITISSHVGLAFTSYSATTSALGLPFVW RPAFSFIIAFMTVMGMTIAFA**KD**IS**DIEGD**AKYGVSTVATKLGARNMTFVVSGVLLLNYLVSISIGIIW PQVFKSNIMILSHAILAFCLIFQTRELALANYASAPSRQFFEFIWLLYYAEYFVYVFI

Table S 4: Amino acid sequences of the *p*-hydroxybenzoate polyprenyltransferase (Coq2) from *S. cerevisiae*. The systematic name of Coq2 is YNR041C. The predicted targeting sequence (mTP) is marked in bold.

Name	Sequence
yCoq2	**MFIWQRKSILLGRSILGSGRVTVAGIIGSSRKRYTSSSSSSS**SPSSKESAPVFTSKELEVARKERL DGLGPFVSRLPKKWIPYAELMRLEKPVGTWLLYLPCSWSILMGAMMQGATLSATAGMLGIFGVG ALVMRGAGCTINDFLDRKLDQRVIRSVERPIASGRVSPRRALVFLGAQTLVGMGVLSLLPAQCWW LGLASLPIVFTYPLFKRFTYYPQAALSACFNWGALLGFPAMGVMSWPTMIPLYLSSYLWCMTYDTI YAHQDKKFDIKAGIKSTALAWGPRTKSIMKAMSASQIALLAVAGLNSGLLWGPGFIGGLGVFAYRL FSMIKKVDLDNPKNCWKYFNANINTGLYFTYALAVDYILRLFGFLACPP
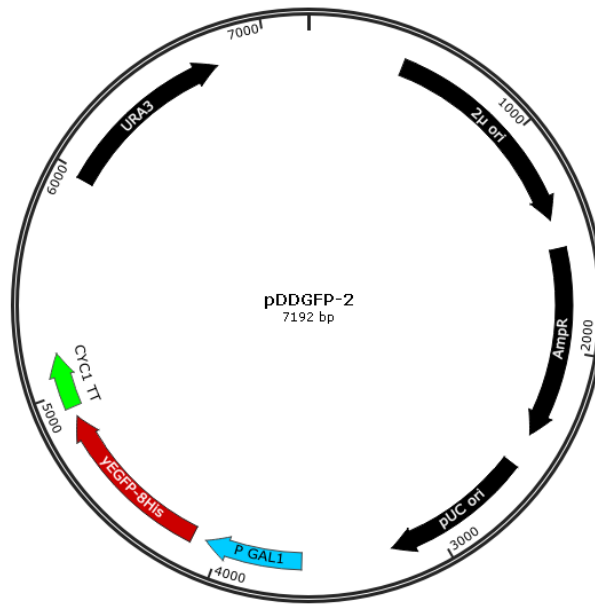
Figure S 1: Expression vector pDDGFP-2.This plasmid was used as PCR template to get the yEGFP-8His-Stop fragment for the modification of the vector pDionysos. The vector pDDGFP-2 was a kind gift of Drew et al. (2008).
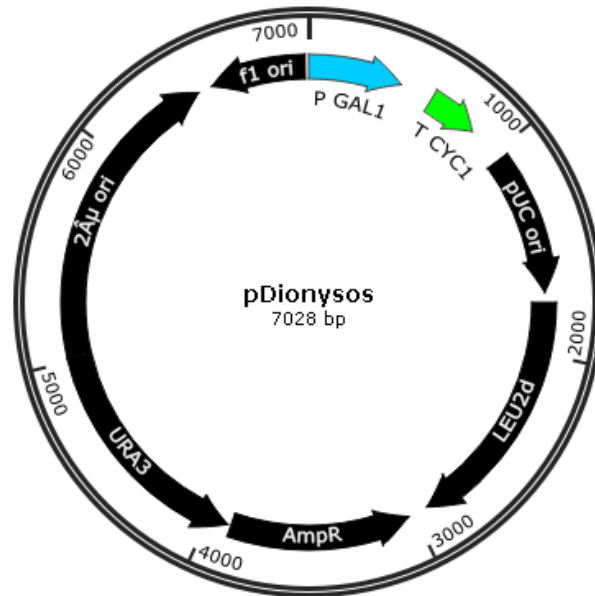

Figure S 2: Expression vector pDionysos. The plasmid uses the pYES2 yeast expression vector (Fisher Scientific GmbH, Schwerte, Germany) as background and is modified by an additional Leu2d gene (Stehle et al., 2008).

Table S 5: Primers used for in this study. Flanking homologous regions are marked in bold, asterisk (*) indicates HPLC purified primers, restriction enzyme recognition sites are underlined, spacer sequences are marked in italic, enterokinase recognition site is marked in bold red, His$_8$ coding sequence is marked in bold blue, 3x stop coding sequence is marked in bold green.

Primer	Sequence (5' → 3')	Application
SFD_pDio+yEGFP _fw*	**ACTGGCGGCCGCTCGAGCATGCA**<u>TCTAGA</u>**GACGACGACGACAAG***G GCAGCAGC***CATCATCACCATCACCAT***AGCAGCGGC*<u>GTCGAC</u>**GAAA ATTTATATTTTCAAGGTCAATTTTCTAAAG**	Generation of pDio + yEGFP
SFD_pDio+yEGFP _rv*	**CATAACTAATTACATGATGCGGCCCT**CTAG**TCACTATTA**<u>TCTAGA</u> <u>GTCGAC</u>**TTAATGATGATGATGATGGTGGTGG**	
SFD_SQ_pDio+yE GFP_fw	GCTGTAATACGACTCACTATAGG	Sequencing
SFD_SQ_pDio+yE GFP_rv	CTATAAAAAAATAAATAGGGACCTAGACTTC	
SFD_BamHI_PT1	CGAT<u>GGATCC</u>ATGGCAGCTACTACAAATCAAACTGAGC	Generation of pDio-CsPT1+
SFD_SphI_PT1	GCGC<u>GCATGC</u>GATATGAAAACATATACTAAATATTCAGCATAATAAA GCTTCC	
SFD_BamHI_PT2	CGAT<u>GGATCC</u>ATGACAGATACTGCAAATCAAACTGAGC	Generation of pDio-CsPT2+
SFD_SphI_PT2	GCGC<u>GCATGC</u>GAAATGAAAACATATACTAAATATTCAGCATAATGGA G	
SFD_BamHI_PT3	CGAT<u>GGATCC</u>ATGGCAGGTAGCGATCAAATTGAAGG	Generation of pDio-CsPT3+
SFD_SphI_PT3	GCGC<u>GCATGC</u>GATATAAATACATATACAAAGTATTCAGCATAATATA GCAACC	
SFD_pDio+CsPT1-SP+Coq2SP_fw	CATTCTGGGCAGCGGCCGAGTTACCGTTGCTGGAATCATTGGAAGC AGCAGAAAGAGATATACGTCTTCTTCCTCCTCCTCCGCAGCTAC TACAAATCAAACTGAGCCTCCAG	Generation of pDio-yCoq2SP-CsPT1+
SFD_pDio-Coq2SP_rv	CAGCAACGGTAACTCGGCCGCTGCCCAGAATGGACCTCCCTAGTAA AATACTCTTTCTCTGCCAAATAAACATTTTTTTGGATCCGAGCTCGGT ACCAAGCTTAATATTC	Generation of pDio-yCoq2SP-CsPT1+, pDio-yCoq2SP-CsPT2+, pDio-yCoq2SP-CsPT3+

Table S 5 continued

Primer	Sequence (5' → 3')	Application
SFD_pDio+CsPT2-SP+Coq2SP_fw	CATTCTGGGCAGCGGCCGAGTTACCGTTGCTGGAATCATTGGAAGC AGCAGAAAGAGATATACGTCTTCTTCCTCCTCCTCCACAGATAC TGCAAATCAAACTGAGCCTCCTGAATC	Generation of pDio-yCoq2SP-CsPT2+
SFD_pDio+CsPT3-SP+Coq2SP_fw	CATTCTGGGCAGCGGCCGAGTTACCGTTGCTGGAATCATTGGAAGC AGCAGAAAGAGATATACGTCTTCTTCCTCCTCCTCCGCAGGTAG CGATCAAATTGAAGGTTCTCCTCATC	Generation of pDio-yCoq2SP-CsPT3+
SQ_pDio-NphB_fw	CATTTTCGGTTTGTATTACTTC	Sequencing
SQ_pDio-NphB_rv	TAAATAGGGACCTAGACTTC	

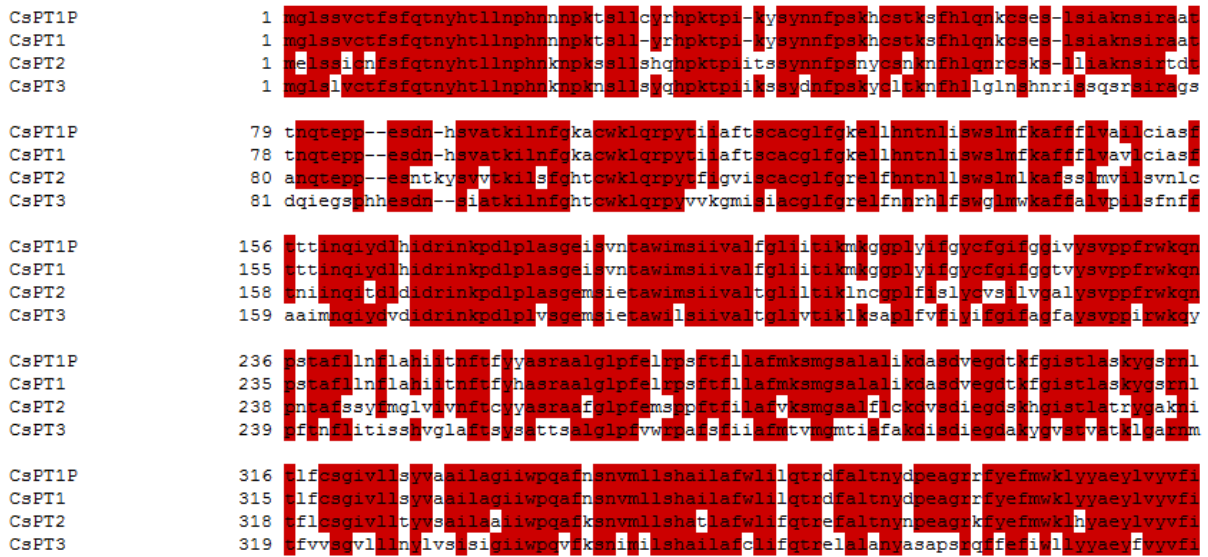**Figure S 3:** Amino acid sequence alignment of CsPT1P, CsPT1, CsPT2 and CsPT3. The alignment was generated using the software Clone Manager 9 Professional Edition (Sci-Ed Software, Denver, USA). The sequences of CsPT1, CsPT2 and CsPT3 were obtained from Pamplaniyil (2017). The sequence of CsPT1P was published by Page and Boubakir (2014). Matching amino acids within the alignment are coloured in red.
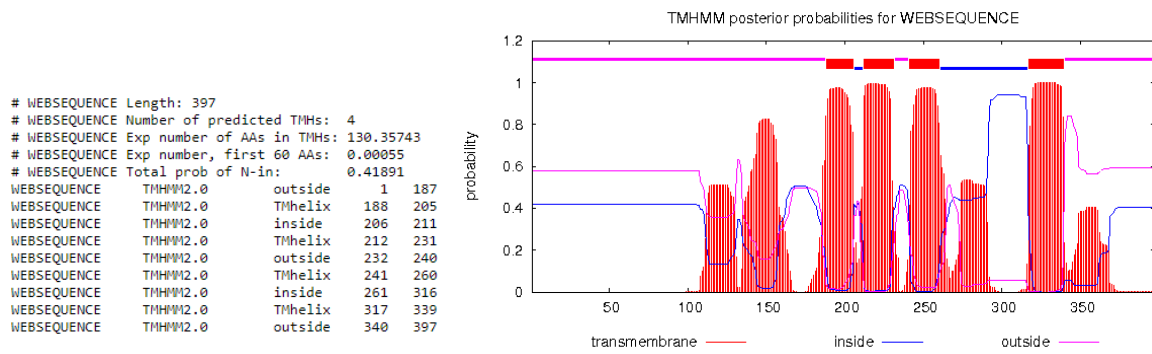
Figure S 4: TMHMM analysis of CsPT2. The results show the length of the submitted sequence and the number of predicted transmembrane helices (TMHs). Additionally, the number of amino acids expected to be part of the TMHs and the expected number of amino acids in TMHs in the first 60 amino acids is given. The latter value is important since it gives a hint whether the predicted TMH could be a signal peptide. The total probability of N-in gives information about the probability of the N terminus to be on the cytoplasmic side of the membrane.

Figure S 5: TMHMM analysis of CsPT3. The results show the length of the submitted sequence and the number of precdicted transmembrane helices (TMHs). Additionally, the number of amino acids expected to be part of the TMHs and the expected number of amino acids in TMHs in the first 60 amino acids is given. The latter value is important since it gives a hint whether the predicted TMH could be a signal peptide. The total probability of N-in gives information about the probability of the N terminus to be on the cytoplasmic side of the membrane.
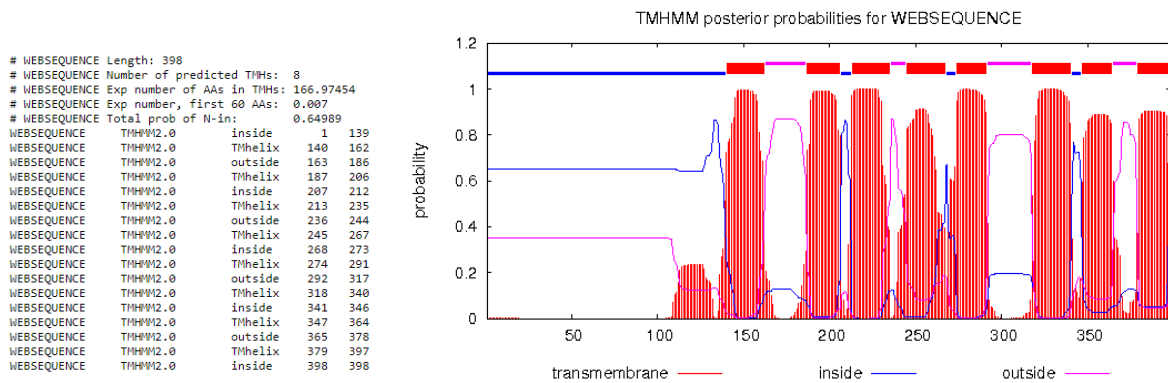
Table S 6: Results of Blast[®] for CsPT2. The results shown are limited to the first three entry sequences because for the other potential prenyltransferases no experimental data is available. The Chosen Search Set contained the following settings: Database: Non-redundant protein sequence (nr); organism: -; exclude: models (XM/XP), uncultured/environmental sample sequence; algorithm: blastp (protein-protein BLAST). Description: name of matched database sequence; Max score: highest alignment score; Total score: total alignment scores; Query cover: percentage overlap of input sequence and sequence of the protein mentioned in the description; E value: lowest Expect value; Ident: highest percent identity; Accession: UniProtKB/Swiss-Prot accession number.

Description	Max Score	Total Score	Query cover	E value	Ident	Accession
2-acylphloroglucinol 4-prenyltransferase	358	358	100%	5e-118	50%	A0A0B5A051.1
2-acylphloroglucinol 4-prenyltransferase, chloroplastic	348	348	100%	3e-114	50%	E5RP65.1
2-acyl-4-prenylphloroglucinol 6-prenyltransferase, chlorplastic	321	321	100%	1e-103	46%	A0A0B4ZTQ2.1

Table S 7: Results of Blast[®] for CsPT3. The results shown are limited to the first three entry sequences because for the other potential prenyltransferases no experimental data is available. The Chosen Search Set contained the following settings: Database: Non-redundant protein sequence (nr); organism: -; exclude: models (XM/XP), uncultured/environmental sample sequence; algorithm: blastp (protein-protein BLAST). Description: name of matched database sequence; Max score: highest alignment score; Total score: total alignment scores; Query cover: percentage overlap of input sequence and sequence of the protein mentioned in the description; E value: lowest Expect value; Ident: highest percent identity; Accession: UniProtKB/Swiss-Prot accession number.

Description	Max Score	Total Score	Query cover	E value	Ident	Accession
2-acylphloroglucinol 4-prenyltransferase	357	357	100%	1e-117	48%	A0A0B5A051.1
2-acylphloroglucinol 4-prenyltransferase, chloroplastic	349	349	100%	1e-114	47%	E5RP65.1
2-acyl-4-prenylphloroglucinol 6-prenyltransferase, chlorplastic	294	294	100%	4e-93	43%	A0A0B4ZTQ2.1

Expression of *nphB* in *Escherichia coli*

Table S 8: Primers used for expression of *nphB* in *E. coli*.

Primer	Sequence (5' → 3')	Application
Nde_NphB_fw	CGTA<u>CATATG</u>AGCGAAGCTGCGGATG	Generation of pET32a-NphB
Bam_NphB_rv	GCAT<u>GGATCC</u>TCATTAATCCTCCAGAGAGTCAAAGG	
SQ_pET32a_neu_fw	CCAAGGGGTTATGCTAGTTA	Sequencing
SQ_pET32a_rv	AAGTGGCGGCAACCAAAGTG	

Table S 9: Primers used for mutational studies of NphB.

Primer	Sequence (5' → 3')
T126A fw	GGCATGTTATCTGCCGGAAAGAACGCGTAGGTTTTCTTGAAGCCGCCAGTCACTTCGC
T126A rev	CGTTCTTTCCGGCAGATAACATGCCAGGAGTTGCGGAACTGAGCGCGATTCC
T126G fw	GGCATGTTATCTCCCGGAAAGAACGCGTAGGTTTTCTTGAAGCCGCCAGTCACTTCGC
T126G rev	CGTTCTTTCCGGGAGATAACATGCCAGGAGTTGCGGAACTGAGCGCGATTCC
T126V fw	TCAAGAAAACCTACGCGTTCTTTCCGGTAGATAACATGCCAGGAGTTGCGGAACTGA
T126V rev	TCAGTTCCGCAACTCCTGGCATGTTATCTACCGGAAAGAACGCGTAGGTTTTCTTGA
D127A fw	AACTCCTGGCATGTTAGCTGTCGGAAAGAACGCGTAGGTTTTCTTGAAGCCGCCAGT
D127A rev	TTTCCGACAGCTAACATGCCAGGAGTTGCGGAACTGAGCGCGATTCCGTCCA
D127G fw	AACTCCTGGCATGTTACCTGTCGGAAAGAACGCGTAGGTTTTCTTGAAGCCGCCAGT
D127G rev	TTTCCGACAGGTAACATGCCAGGAGTTGCGGAACTGAGCGCGATTCCGTCCA
Q161N fw	CATACTCGTCATATTGACTTTGTCCAGCCCATAACGGGCAAAGAGTTCGGCATTCTCT
Q161N rev	TATGGGCTGGACAAAGTCAATATGACGAGTATGGACTATAAGAAACGCCAGGTCAACCTG
Q161A fw	CATACTCGTCATCGCGACTTTGTCCAGCCCATAACGGGCAAAGAGTTCGGCATTCTCT
Q161A rev	TATGGGCTGGACAAAGTCGCGATGACGAGTATGGACTATAAGAAACGCCAGGTCAACCTG
Q161E fw	CATACTCGTCATAGCGACTTTGTCCAGCCCATAACGGGCAAAGAGTTCGGCATTCTCT
Q161E rev	TATGGGCTGGACAAAGTCGCTATGACGAGTATGGACTATAAGAAACGCCAGGTCAACCTG

Table S 9 continued

Primer	Sequence (5' → 3')
M162A fw	GTTGACCTGGCGTTTCTTATAGTCCATACTCGTAGCCTGGACTTTGTCCAGCCCATAA
M162A rev	TTATGGGCTGGACAAAGTCCAGGCTACGAGTATGGACTATAAGAAACGCCAGGTCAAC
M162K fw	TATAGTCCATACTCGTCTTCTGGACTTTGTCCAGCCCATAACGGGCAAAGAGTTC
M162K rev	ACAAAGTCCAGAAGACGAGTATGGACTATAAGAAACGCCAGGTCAACCTGTACTT
M162N fw	TAGTCCATACTCGTATTCTGGACTTTGTCCAGCCCATAACGGGCAAAGAGTTC
M162N rev	ACAAAGTCCAGAATACGAGTATGGACTATAAGAAACGCCAGGTCAACCTGTAC
M162W fw	TAGTCCATACTCGTCCACTGGACTTTGTCCAGCCCATAACGGGCAAAGAGTTC
M162W rev	ACAAAGTCCAGTGGACGAGTATGGACTATAAGAAACGCCAGGTCAACCTGTAC
Y175A fw	GTTCACTGAAGGCCAGGTTGACCTGGCGTTTCTTATAGTCCATACTCGT
Y175A rev	AGGTCAACCTGGCCTTCAGTGAACTGAGTGCGCAAACCTTGGAAGCTGA
Y175N fw	<u>GTTCACTGAA</u>GTTCAGGTTGACCTGGCGTTTCTTATAGTCCATACTCGT
Y175N rev	<u>AGGTCAACCTG</u>AACTTCAGTGAACTGAGTGCGCAAACCTTGGAAGCTGA
S214A fw	CGTGGGATACACTGCAAAGCTCCGTTTG
S214A rev	CAAACGGAGCTTTGCAGTGTATCCCACG
S214D fw	TTCAGCGTGGGATACACGTCAAAGCTCCGTTTGCAAAATTTCAGACCCAATTCGTTTGGAACAT
S214D rev	AAACGGAGCTTTGACGTGTATCCCACGCTGAATTGGGAAACAGGCAAGATCGATCGCTTGT
S214N fw	TTCAGCGTGGGATACACGTTAAAGCTCCGTTTGCAAAATTTCAGACCCAATTCGTTTGGAACAT
S214N rev	AAACGGAGCTTTAACGTGTATCCCACGCTGAATTGGGAAACAGGCAAGATCGATCGCTTGT
F213A fw	GATACACCGAAGCGCTCCGTTTGCAAAATTTCAGACCCAATTCGTTTG
F213A rev	GCAAACGGAGCgctTCGGTGTATCCCACGCTGAATTGGGAAACAG
Y216A fw	CCAATTCAGCGTGGGTGCCACCGAAAAGC
Y216A rev	GCTTTTCGGTGGCACCCACGCTGAATTGG
V271N fw	GCGTAAGGCCGTAATTTAACGTGCGC
V271N rev	GCGCACGTTAAATTACGGCCTTACGC

Table S 10: Amino acid sequence of NphB wt. GenBank accession number: AB187169.

Name	Sequence
NphB wt	MSEAADVERVYAAMEEAAGLLGVACARDKIYPLLSTFQDTLVEGGSVVVFSMASGRHS TELDFSISVPTSHGDPYATVVEKGLFPATGHPVDDLLADTQKHLPVSMFAIDGEVTGGF KKTYAFFPTDNMPGVAELSAIPSMPPAVAENAELFARYGLDKVQMTSMDYKKRQVNLY FSELSAQTLEAESVLALVRELGLHVPNELGLKFCKRSFSVYPTLNWETGKIDRLCFAVIS NDPTLVPSSDEGDIEKFHNYATKAPYAYVGEKRTLVYGLTLSPKEEYYKLGAYYHITDV QRGLLKAFDSLED

Table S 11: *E. coli* strains used for the expression of *nphB*. [1]pLysS strains are used in general if toxic proteins which affect cell growth or cell viability are expressed by suppressing basal T7 RNA polymerase expression prior to induction.

| *E. coli* strain	Special features	Reference
Rosetta 2 (DE3)	contains a plasmid which encodes for tRNAs of seven codons that are rarely used in *E. coli*	Merck Millipore; Darmstadt, Germany
Rosetta 2 (DE3) pLyS[1]	Like Rosetta 2 (DE3) + pLysS	Merck Millipore; Darmstadt, Germany
BL21 Gold (DE3)	lacks two proteases (OmpT and Lon) which could cause degradation of the target protein	Agilent Technologies; Santa Clara, CA, USA
BL21 Gold (DE3) pLysS[1]	Like BL21 Gold (DE3) + pLysS	Agilent Technologies; Santa Clara, CA, USA
OverExpress™ C41 (DE3) pLysS[1]	Used for expression of many recombinant toxic proteins, lacks two proteases (OmpT and Lon)	Lucigen®; Middleton, WI, USA
OverExpress™ C43 (DE3) pLysS[1]	Like OverExpress™ C41 (DE3) pLysS, but resistant to a different toxic protein	Lucigen®; Middleton, WI, USA
BL21 (DE3)	lacks two proteases (OmpT and Lon), standard strain for high level protein expression	NEB; Frankfurt am Main, Germany
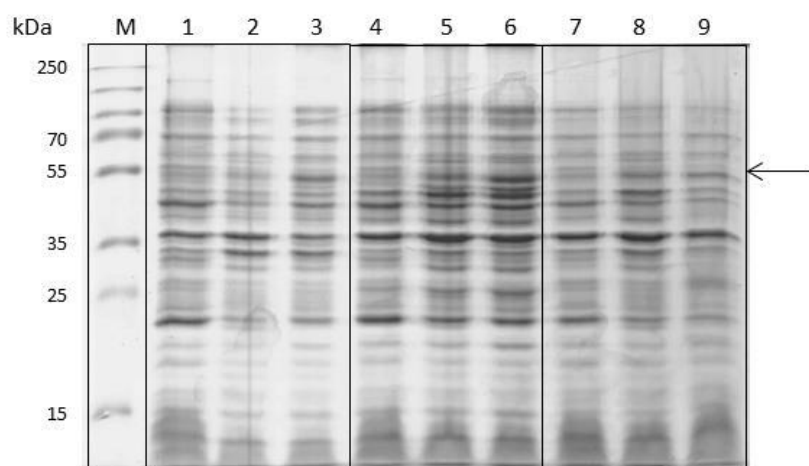
Figure S 6: Coomassie stained SDS gel of *nphB* expression in *E. coli* (2). M: Thermo Scientific PageRuler Plus Prestained Protein Ladder; 1: BL21 Gold (DE3) pLysS 0 h, not induced; 2: BL21 Gold (DE3) pLysS 17 h, induced; 3: BL21 Gold (DE3) pLysS 17 h, induced; 4: OverExpress[TM] C41 (DE3) pLysS 0 h, not induced; 5: OverExpress[TM] C41 (DE3) pLysS 17 h, not induced; 6: OverExpress[TM] C41 (DE3) pLysS 17 h, induced; 7: Rosetta 2 (DE3) 0 h, not induced; 8: Rosetta 2 (DE3) 17 h, not induced; 9: Rosetta 2 (DE3) 17 h, induced. The size of the expected protein band is marked with an arrow.
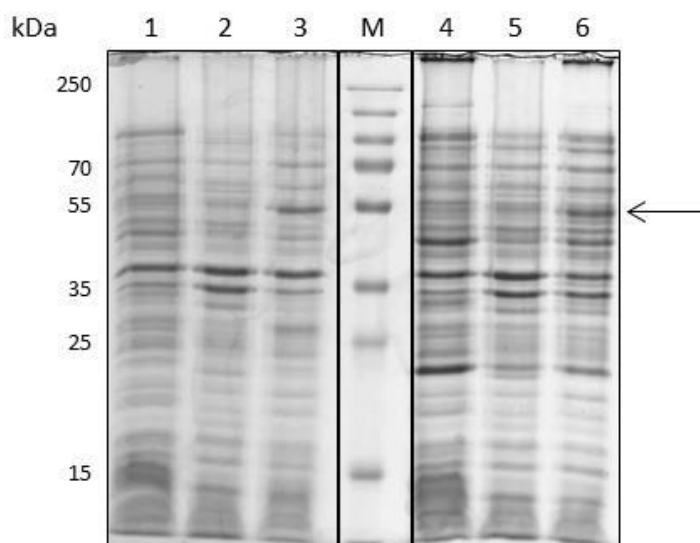

Figure S 7: Coomassie stained SDS gel of *nphB* expression in *E. coli* (3). 1: BL21 Gold (DE3) 0 h, not induced; 2: BL21 Gold (DE3) 17 h, not induced; 3: BL21 Gold (DE3) 17 h, induced; M: Thermo Scientific PageRuler Plus Prestained Protein Ladder; 4: BL21 Gold (DE3) pLysS 0 h, not induced; 5: BL21 Gold (DE3) pLysS 17 h, not induced; 6: BL21 Gold (DE3) pLysS 17 h, induced. The size of the expected protein band is marked with an arrow.

Table S 12: ^{1}H-NMR results of the standard compounds CBGA, OA and THCA as well as of the side-product 2-O-GOA and the in-house synthesized GPP (Zirpel, Degenhardt et al., 2017).

GPP

^{1}H-NMR (500 MHz, D_2O + NaOD) δ: 0.86 ppm (t, 3H, J = 6.8 Hz) 1.28 ppm (m, 4H) 1.49 - 1.46 ppm (m, 2H) 2.74 ppm (t, 2H, $J_{1H,1H}$ = 7.2 Hz) 6.12 ppm (d, 1H, $J_{1H,1H}$ = 2.4 Hz) 6.15 ppm (d, 1H, $J_{1H,1H}$ = 2.4 Hz) 10.05 ppm (broad, 1 H)

OA

^{1}H-NMR (500 MHz, d_6-DMSO) δ: 1.61 ppm (s, 3H, methyl) 1.68 ppm (s, 3H, methyl) 1.70 ppm (s, 3H, methyl) 2.09 ppm (m, 4H, CH_2 at C_4 and C_5) 4.46 ppm (t, 2H, $J_{1H,1H}$ =6.5 Hz, CH_2 at C_1) 5.20 ppm (broad, 1 H, $J_{1H,1H}$ =6.5 Hz, H at C_6) and 5.45 ppm (t, 1 H, $J_{1H,1H}$ =6.5 Hz, H at C_2)

CBGA

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 0.84 ppm (t, 3H, $J_{1H,1H}$= 7.2 Hz) 1.28 ppm (m) 1.53 ppm (s) 1.61 ppm (m) 1.65 ppm (s) 1.75 ppm (s) 2.02 ppm (m) 2.21 ppm (m) 2.82 ppm (t, $J_{1H,1H}$ = 7.2 Hz) 3.38 ppm (d, $J_{1H,1H}$ = 7.0 Hz) 4.99 ppm (t, $J_{1H,1H}$ = 7.0 Hz) 5.52 ppm (t, $J_{1H,1H}$ = 7.0 Hz) 6.22 ppm (s)

2-*O*-GOA

^{1}H-NMR (700 MHz, $CDCl_3$) δ: 0.88 ppm (t) 1.55 ppm (m) 5.35 ppm (m) 6.17 ppm (s) 6.24 ppm (s). The shifts (6.17 and 6.24 ppm) indicate an alpha prenylation at the C2 hydroxyl group.

THCA

^{1}H-NMR (500 MHz, $CDCl_3$) δ: 0.84 ppm (t, $J_{1H,1H}$= 6.4 Hz) 1.09 ppm (s) 1.39 ppm (m) 1.42 ppm (s) 1.65 ppm (s) 1.54 ppm (m) 1.69 ppm (s) 2.21 ppm (m) 1.70 ppm (m) 2.10 ppm (m) 2.78 ppm, 2.90 ppm (m) 3.18 ppm (d, $J_{1H,1H}$ = 10.4 Hz) 6.18 ppm (s) 6.39 ppm (s)

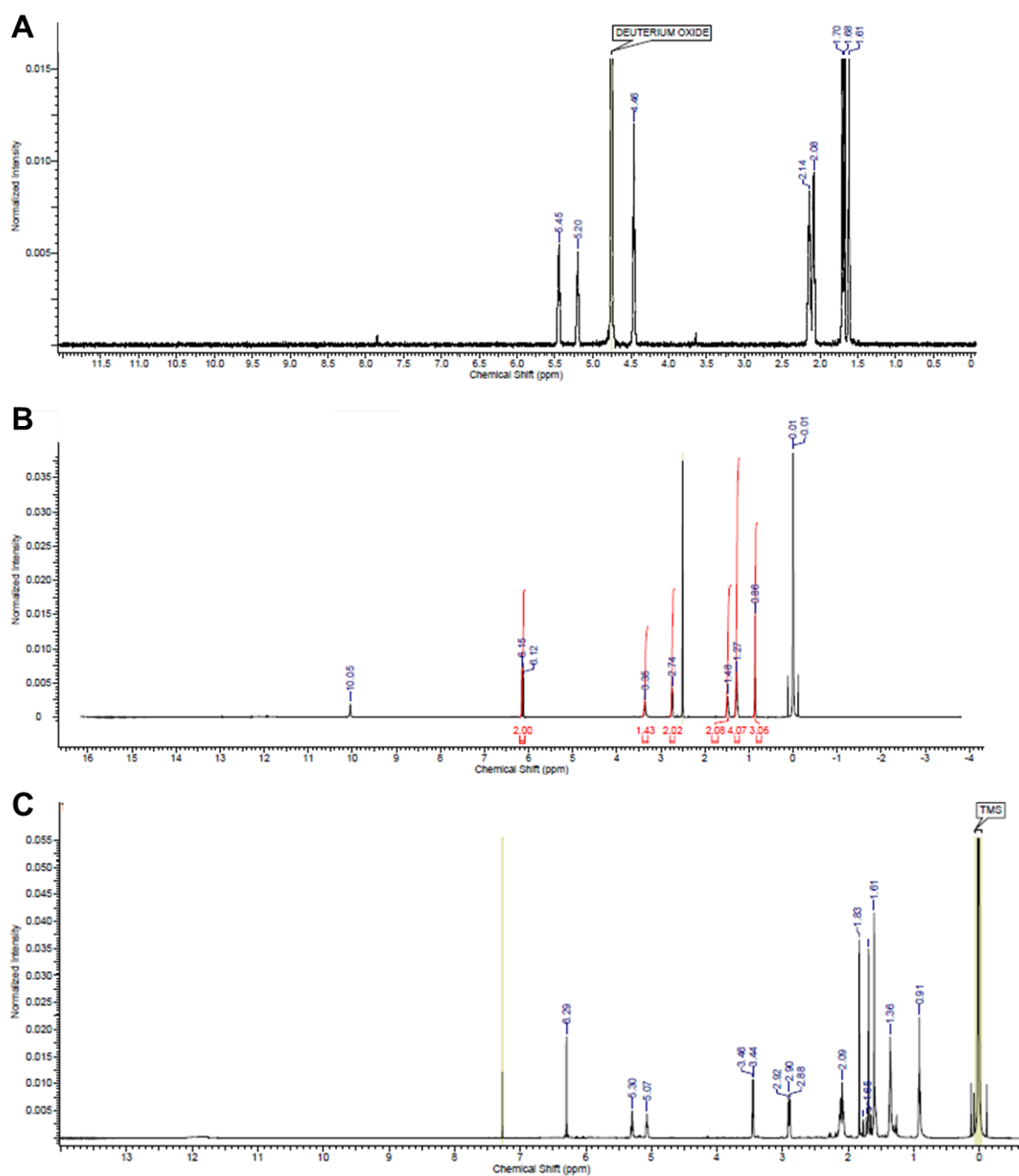**Figure S 8:** ^{1}H-NMR spectra of GPP, OA and CBGA. A: GPP (D_2O, 27 °C, 500 MHz), B: OA (DMSO-d6, 23 °C, 500 MHz), C: CBGA (chloroform-d, 23 °C, 500 MHz). The NMR data were processed using the ACD/NMR Processor Academic Edition software package.

Figure S 9: ^{1}H-NMR spectra of 2-*O*-GOA and THCA. D: 2-*O*-GOA (chloroform-d6, 20 °C, 500 MHz), E: THCA (chloroform-d, 23 °C, 500 MHz). The NMR data were processed using the ACD/NMR Processor Academic Edition software package.
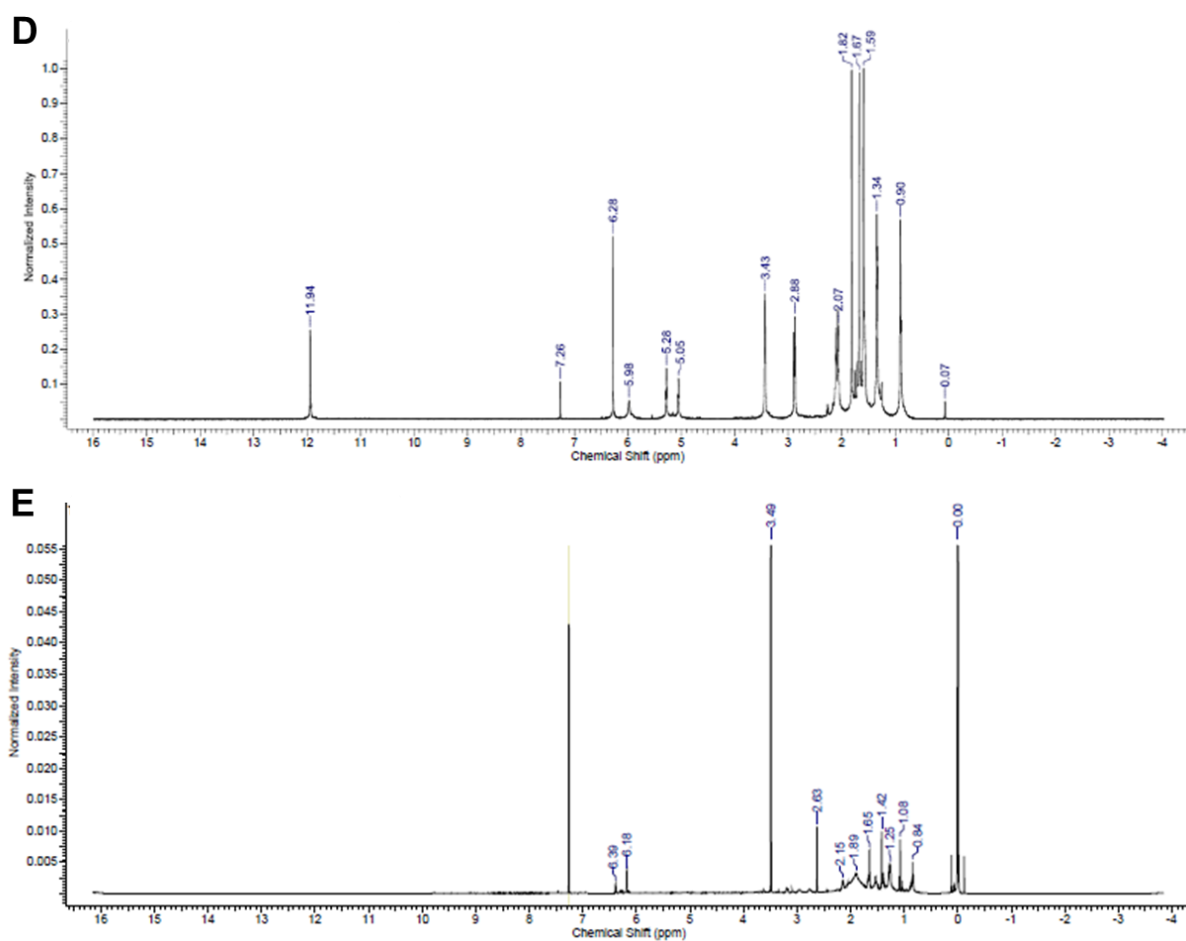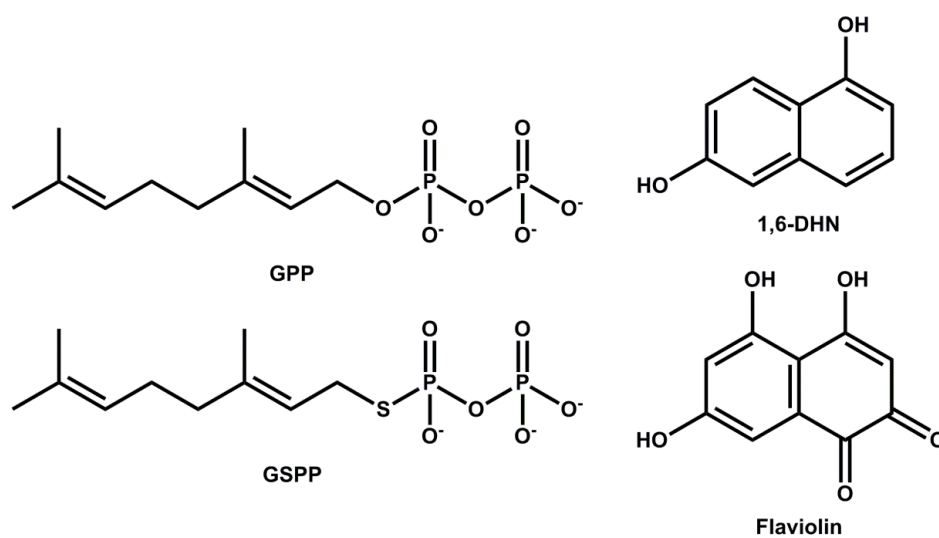

Figure S 10: Structural formulars of GPP, GSPP, 1,6-DHN and flaviolin.

Table S 13: N-terminal tag of NphB cloned into the vector pET32a. Underlined: TrxA tag, bold: His₆ tag, underlined twice: thrombin recognition site, dotted line: S tag, dashed line: enterokinase recognition site.

MSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQNPGTAPKY GIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGSGSGHM**HHHHHH**SSGLVPRGSGMKETAAAK FERQHMDSPDLGTDDDDKAMA

Expression of *nphB* and *thcas* in *S. cerevisiae*

Table S 14: Primers used for the generation of different pDionysos-NphB variant. Recognition sites of the restrictions enzymes are underlined.

Primer	Sequence 5' →3'	Application
SFD_pDio-NphB_fw	GATCAAGCTTAAAAAAATGGCGATGAGCGAAGC	Generation of pDionysos_eN (*Hind*III, *Xba*I)
SFD_pDio-NphB_rv	GATCTCTAGATTAATCCTCCAGAGAGTCAAAGGCTTTC	
NdeI_NphB_fw	CGTACATATGAGCGAAGCTGCGGATGTTG	Colony PCR
HindIII_NphB_rv	GCATAAGCTTTCATTAATCCTCCAGAGAGTCAAAGGCTTTC	
SQ_pDio-NphB_fw	CATTTTCGGTTTGTATTACTTC	Sequencing
SQ_pDio-NphB_rv	CTTCAGATCCAGGGATAAAT	
SFD_pDio-TrxA-NphB_fw	GATCAAGCTTAAAAAAATGAGCGATAAAATTATTCACCTGACTGAC	Generation of pDionysos_TeN (*Hind*III, *Xba*I)
SFD_pDio-NphB_rv	GATCTCTAGATTAATCCTCCAGAGAGTCAAAGGCTTTC	
SQ_pDio+yEGFP_neu_fw	TCGGTTTGTATTACTTCTTATTC	Colony PCR, sequencing
SQ_pDio+yEGFP_rv	CTATAAAAAAATAAATAGGGACCTAGACTTC	
p_TN+_fw	CAAAAGAAGCTTTTCGGTGTATCCCACGCTGAATTGGGAAAC	Site-directed mutagenesis of pDionysos_eN and pDionysos_TeN
p_TN+_rv	ATACACCGAAAAGCTTCTTTTGCAAAATTTCAGACCCAATTCGTTTGGA	

Table S 14 continued

Primer	Sequence 5' →3'	Application
Vf-pDio (Str_codon_yeast)	TCTAGAGGGCCGCATCATGTAATTAGT	
Vr-pDio (Str_codon_yeast)	AAGCTTAATATTCCCTATAGTGAGTCGT	Generation of pDionysos_N
1f-AAAAAA+NphB-o(Str_codon yeast)	CGACTCACTATAGGGAATATTAAGCTTAAAAAAATGTCTGAAGCGGCGGAC	
1r-AAAAAA+NphB-o(Str_codon yeast)	TAATTACATGATGCGGCCCTCTAGATTATCAATCTTCCAAAGAATCAAAAGCCTTCAA	
Vf-pDionysos (Lys-Arg-Str)	TCTAGAGGGCCGCATCATGTAATTAGT	
Vf-pDionysos (Lys-Arg-Str)	AAGCTTAATATTCCCTATAGTGAGTCGT	Generation of pDionysos_Nx
1f-AAAAAA-o(Lys-Arg-Str)	CGACTCACTATAGGGAATATTAAGCTTAAAAAAATGTCTGAAGCGGCGGAC	
1r-AAAAAA-o(Lys-Arg-Str)	TAATTACATGATGCGGCCCTCTAGATTATCAATCTTCCAAAGAATCAAAAGCTCTC	

Table S 15: Primers used for the knockout of the mig1 binding site of pGal1 in pDionysos.

Primer	Sequence 5' →3'	Application
Fw GAL1 Gluc rep free	GCTCCGAACAATAAAGATTCGAAGCGATGATTTTTGATCTATTAA	Generation of mutated *Gal1* promoter in pDionysos
Rv GAL1 Gluc rep free	AGATCAAAATCATCGCTTCGAATCTTTATTGTTCGGAGCAGTGCGGCGCGAG	
FW pDio GAL1 Del Check	AGCCCTCCGAAGGAAGACTC	Colony PCR
RV pDio Gal1 Del Check	AGCTGCTAGTAGTCCGATCC	

Table S 15 continued

Primer	Sequence 5' →3'	Application
SQ_PGal1 in pDio	GAAGGGAAGAAAGCGAAAGG	
SQ_PGal1 in pDio-PGal10-pGal1	CCACCATATACATATCCATATCTAATC	Sequencing

```
 801  tgtatcagtt taatcaccat aatatcgttt tctttgttta gtgcaattaa ttttttcctat tgttacttcg ggccttttttc
      acatagtcaa attagtggta ttatagcaaa agaaacaaat cacgttaatt aaaaaggata acaatgaagc ccggaaaaag

 881  tgttttatga gctattttttt ccgtcatcct tccccagatt ttcagcttca tctccagatt gtgtctacgt aatgcacgcc
      acaaaatact cgataaaaaa ggcagtagga aggggtctaa aagtcgaagt agaggtctaa cacagatgca ttacgtgcgg
                                        >>..SiteI..>>
                                          s  p  d  f
                                                          >>.SiteII..>>
                                                            s  p  d  c

 961  atcattttaa gagaggacag agaagcaagc ctcctgaaag atgaagctac tgtcttctat cgaacaagca tgcgatattt
      tagtaaaatt ctctcctgtc tcttcgttcg gaggactttc tacttcgatg acagaagata gcttgttcgt acgctataaa
                                                  >>..................Gal4....................>
                                                    m  k  l    l  s  s    i  e  q  a    c  d  i

1041  gccgacttaa aaagctcaag tgctccaaag aaaaaccgaa gtgcgccaag tgtctgaaga acaactggga gtgtcgctac
      cggctgaatt tttcgagttc acgaggtttc tttttggctt cacgcggttc acagacttct tgttgaccct cacagcgatg
      >.............................................Gal4.........................................>
        c  r  l    k  k  l  k    c  s  k    e  k  p    k  c  a  k    c  l  k    n  n  w    e  c  r  y
```

Figure S 11: pGal1 with mig 1 binding sites URS A, URS B and URS C.

Table S 16: Primers used for the generation of *S. cerevisiae Δpep4Δgal1Δgal80*. (Zirpel, Degenhardt et al., 2017)

Primer	Sequence (5' → 3)'	Application
HR_Gal80_fw	ATGGACTACAACAAGAGATCTTCGGTCTCAACCGTGCCTAATGCAGCTCCGTACGCTGCAGGTCGACAAC	Generation of the *Gal80* knockout in the genome of *S. cerevisiae* CEN.PK2-1C *Δpep4Δgal1*
HR_Gal80_rv	TTATAAACTATAATGCGAGATATTGCTAACGTTTAATGTGGAGCCCATCAGCATAGGCCACTAGTGGATCTGATAT	
C-PCR_Gal80_fw	CATATCACTGCTGGTCCTTG	colony PCR, sequencing
C-PCR_Gal80_rv	TGGTTGTTTTACATAGATATATACTCAGTATT	

Table S 17: Amino acid sequences of NphB (GenBank[TM] accession number: AB187169), NphB(arg), THCAS (GenBank[TM] accession number: AB057805), proteinase A (GenBank[TM] accession number: CAA78020), T2A linker sequence. Signal peptides predicted by TargetP (http://www.cbs.dtu.dk/services/TargetP/) are marked in bold. [1]Beekwilder et al. (2014).

Name	Sequence
NphB	MSEAADVERVYAAMEEAAGLLGVACARDKIYPLLSTFQDTLVEGGSVVVFSM ASGRHSTELDFSISVPTSHGDPYATVVEKGLFPATGHPVDDLLADTQKHLPVS MFAIDGEVTGGFKKTYAFFPTDNMPGVAELSAIPSMPPAVAENAELFARYGLD KVQMTSMDYKKRQVNLYFSELSAQTLEAESVLALVRELGLHVPNELGLKFCK RSFSVYPTLNWETGKIDRLCFAVISNDPTLVPSSDEGDIEKFHNYATKAPYAY VGEKRTLVYGLTLSPKEEYYKLGAYYHITDVQRGLLKAFDSLED
NphB(arg)	MSEAADVERVYAAMEEAAGLLGVACARRKIYPLLSTFQDTLVEGGSVVVFSM ASGRHSTELDFSISVPTSHGDPYATVVEKGLFPATGHPVDDLLADTQRHLPVS MFAIDGEVTGGFKKTYAFFPTDNMPGVAELSAIPSMPPAVAENAELFARYGLD RVQMTSMDYKKRQVNLYFSELSAQTLEAESVLALVRELGLHVPNELGLRFCK RSFSVYPTLNWETGRIDRLCFAVISNDPTLVPSSDEGDIERFHNYATRAPYAY VGERRTLVYGLTLSPREEYYKLGAYYHITDVQRGLLRAFDSLED
THCAS	**MNCSAFSFWFVCKIIFFFLSFHIQISIA**NPRENFLKCFSKHIPNNVANPKLVYTQ HDQLYMSILNSTIQNLRFISDTTPKPLVIVTPSNNSHIQATILCSKKVGLQIRTRS GGHDAEGMSYISQVPFVVVDLRNMHSIKIDVHSQTAWVEAGATLGEVYYWIN EKNENLSFPGGYCPTVGVGGHFSGGGYGALMRNYGLAADNIIDAHLVNVDG KVLDRKSMGEDLFWAIRGGGGENFGIIAAWKIKLVAVPSKSTIFSVKKNMEIHG LVKLFNKWQNIAYKYDKDLVLMTHFITKNITDNHGKNKTTVHGYFSSIFHGGV DSLVDLMNKSFPELGIKKTDCKEFSWIDTTIFYSGVVNFNTANFKKEILLDRSA GKKTAFSIKLDYVKKPIPETAMVKILEKLYEEDVGAGMYVLYPYGGIMEEISES AIPFPHRAGIMYELWYTASWEKQEDNEKHINWVRSVYNFTTPYVSQNPRLAY LNYRDLDLGKTNHASPNNYTQARIWGEKYFGKNFNRLVKVKTKVDPNNFFRN EQSIPPLPPHHH
Proteinase A	**MFSLKALLPLALLLVSANQVAA**KVHKAKIYKHELSDEMKEVTFEQHLAHLGQ KYLTQFEKANPEVVFSREHPFFTEGGHDVPLTNYLNAQYYTDITLGTPPQNFK VILDTGSSNLWVPSNECGSLACFLHSKYDHEASSSYKANGTEFAIQYGTGSLE GYISQDTLSIGDLTIPKQDFAEATSEPGLTFAFGKFDGILGLGYDTISVDKVVPP FYNAIQQDLLDEKRFAFYLGDTSKDTENGGEATFGGIDESKFKGDITWLPVRR KAYWEVKFEGIGLGDEYAELESHGAAIDTGTSLITLPSGLAEMINAEIGAKKGW TGQYTLDCNTRDNLPDLIFNFNGYNFTIGPYDYTLEVSGSCISAITPMDFPEPV GPLAIVGDAFLRKYYSIYDLGNNAVGLAKAI
T2A (yeast)[1]	RAEGRGSLLTCGDVEENPGP

Table S 18: Genotype of used *S. cerevisiae* strains. (Zirpel, Degenhardt et al., 2017)

Organism	Strain	Genotype	Reference
S. cerevisiae	CEN.PK2-1C *Δpep4Δgal1*	*MATa; gal1::loxP; pep4::loxP; ura3-52; trp1-289; leu2-3,112; his3Δ 1; MAL2-8C; SUC2*	(Zirpel et al., 2015)
S. cerevisiae	CEN.PK2-1C *Δpep4Δgal1Δgal80*	*MAT1; gal1::loxP; pep4::loxP; gal80::loxP; ura3-52; trp1-289; leu2-3,112; his3Δ 1; MAL2-8C; SUC2*	This study (Zirpel, Degenhardt et al., 2017)

Table S 19: List of primers used for expression of *nphB* and *thcas* simultaneously.(Zirpel, Degenhardt et al., 2017)

Primer	Sequence (5' →3')	Application
Vf-pDio-THCAS	TGATCTAGAGGGCCGCATCATGTAATTAGTT ATGTCACGCTTACATTCACG	
Vr-pDio-THCAS	ATGGTGATGTGGAGGCAATGGAGGGATGGA TTGTTCGTTTCTAAAGAAGTTGTTTGGGTCG ACTTTAGTCTTGACTTTGACTAATCTAT	Generation of pDionysos_TT2AN
1f-ORF-2-o(Vec)-NphB (Cod)	CCCTCCATTGCCTCCACATCACCAT<u>AGAGCT GAGGGTAGAGGAAGTCTACTAACATGCGGT GACGTAGAAGAAAACCCTGGTCCT</u>TCTGAAG CGGCGGACGTT	
1r-ORF-2-o(Vec)-NphB (Cod)	TTACATGATGCGGCCCTCTAGATCATCAATC TTCCAAAGAATCAAAAGCCTTCAA	

Table S 19 continued

Primer	Sequence (5' → 3')	Application
Vf-pDio (Str_codon_yeast)	TCTAGAGGGCCGCATCATGTAATTAGT	
Vr-pDio (Str_codon_yeast)	AAGCTTAATATTCCCTATAGTGAGTCGT	
1f-AAAAAA+NphB-o(Str_codon yeast)	CGACTCACTATAGGGAATATTAAGCTTAAAAAAATGTCTGAAGCGGCGGAC	Generation of pDionysos_NphB
1r-AAAAAA+NphB-o(Str_codon yeast)	TAATTACATGATGCGGCCCTCTAGATTATCAATCTTCCAAAGAATCAAAAGCCTTCAA	
pDio-Gal10_fw	AGTACGGATTAGAAGCCGCCGAGCGGGTGA	
pDio-Gal10_rv	AGTGGATCATCCCCACGCGCCCTGT	
Gal10-NphB-ADH1_fw	ACAGGGCGCGTGGGGATGATCCACTGAGCGACCTCATGCTATACC	Generation of pDio2
Gal10-NphB-ADH1_rv	CGCTCGGCGGCTTCTAATCCGTACTTCAATATAGCAATGAGCAGTTAAGCGTA	
Vf_pDio-ADH1	TGGAGACTTGACCAAACCTCTGGCGAAGAAT	
Vr_pDio-ADH1	TGTTGACACTTCTAAATAAGCGAATTTCTTATGATTTATGA	
1f_THCAS (1-)o(pDio2)	ATTCGCTTATTTAGAAGTGTCAACATTAATGATGATGATGGTGATGTGG	Generation of pDio2_THCAS
1r_THCAS (1-)o(pDio2)	CGCCAGAGGTTTGGTCAAGTCTCCAAAAAAATGTCCAGCTTGAAAGCATTAT	

Table S 19 continued

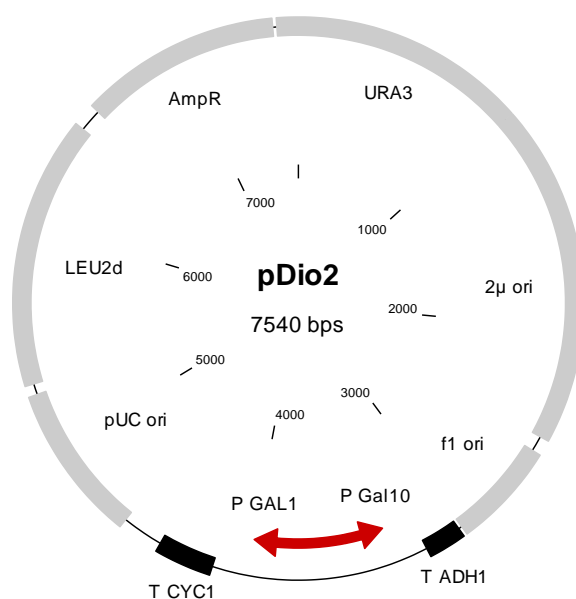| Primer	Sequence (5' →3')	Application
pDio-Gal10_fw	AGTACGGATTAGAAGCCGCCGAGCGGGTGA	
pDio-Gal10_rv	AGTGGATCATCCCCACGCGCCCTGT	
Gal10-NphB-ADH1_fw	ACAGGGCGCGTGGGGATGATCCACTGAGCG ACCTCATGCTATACC	Generation of pDio2_THCAS-NphB
Gal10-NphB-ADH1_rv	CGCTCGGCGGCTTCTAATCCGTACTTCAATA TAGCAATGAGCAGTTAAGCGTA	

Figure S 12: Vector map of pDio2. The plasmid pDionysos uses the pYES2 yeast expression vector (Fisher Scientific GmbH, Schwerte, Germany) as background and is modified by an additional Leu2d gene (Figure S 2). The plasmid pDio2 uses pDionysos as background and is modified by a pGal10-TADH1 cassette.
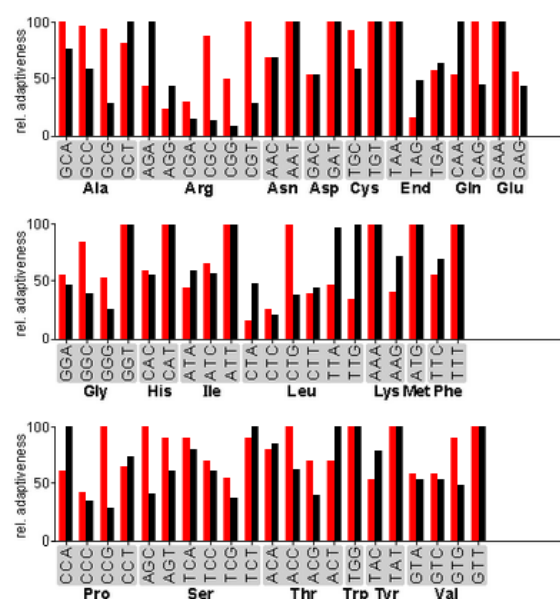
Figure S 13: Comparison of the codon usage tables of *E. coli* and *S. cerevisiae.* The comparison of the two codon usage tables was performed using the graphical codon usage analyzer (http://gcua.schoedl.de) (Fuhrmann et al., 2004). The codons used by *E. coli* are marked in red, while the ones of *S. cerevisiae* are marked in black. The codon usage tables of these two organisms exhibit a mean difference of 22.16 %.
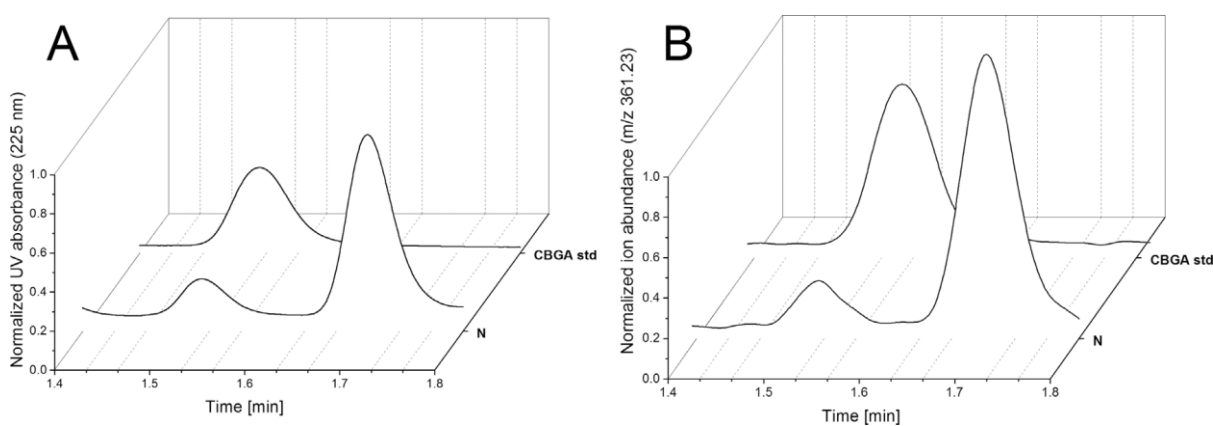

Figure S 14: LC-MS analysis of compounds with *m/z* 361.23 produced with cell lysate supernatants of SC_N. The lysate supernatants were incubated with 5 mM magnesium chloride, 1 mM OA and 1 mM GPP. (A) LC UV chromatograms at 225 nm of the standard compound CBGA and the assay products. (B) Extracted ion chromatogram of *m/z* 361.23 (CBGA) of the standard compound CBGA and the assay products. N – activity assay of SC_N lysates (Zirpel, Degenhardt et al., 2017).

CURRICULUM VITAE

PERSONAL DETAILS

Name	Sara Friederike Degenhardt
Place of birth	Mainz
Nationality	German

EDUCATION, QUALIFICATIONS AND PROFESSIONAL EXPERIENCE

Since 06/2013	**PhD student** Technical Biochemistry, Biochemical and Chemical Engineering (BCI), TU Dortmund University
06/2017 – 12/2017	**Guest scientist** Technical Biochemistry, Biochemical and Chemical Engineering (BCI), TU Dortmund University
05/2013 – 05/2017	**Research associate** Technical Biochemistry, Biochemical and Chemical Engineering (BCI), TU Dortmund University
05/2013 – 05/2016	**Graduate cluster fellow** CLIB-Graduate Cluster "Industrial Biotechnology" (CLIB-GC within CLIB^{2021})
10/2010 – 05/2013	**M.Sc. Chemical Biology** Faculty of Chemistry and Chemical Biology (CCB), TU Dortmund University
10/2005 – 06/2010	**B.Sc. Chemical Biology** Faculty of Chemistry and Chemical Biology (CCB), TU Dortmund University
08/2003 – 07/2005	**Chemical Technical Assistant (CTA)** Sabine Blindow Schule, Hannover
04/2003 – 07/2003	**Studies of Human Medicine** Medizinische Hochschule Hannover
04/2001 – 03/2003	**Studies of Human Medicine** Georg-August Universität Göttingen

SCHOOL EDUCATION

1997 – 2000	**Abitur**
	Internatsgymnasium Pädagogium Bad Sachsa
1993 – 1997	Gymnasium Käthe-Kollwitz-Schule, Hannover

INTERNSHIP

01/2015 – 03/2015	Department of Molecular Biology and Genetics – DANDRITE (Prof. Poul Nissen), Aarhus University, Denmark
06/2009 – 12/2009	Department of Physical Biochemistry (Prof. Dr. Goody), Max Planck Institute of Molecular Physiology, Dortmund
09/2000 – 03/2001	"Freiwilliges Soziales Jahr"
	Henriettenstiftung Neu Bethesda, Hannover

List of publications

- **Degenhardt, F.,** Stehle, F., Kayser, O., 2017. The biosynthesis of cannabinoids, in: Preedy, V. (Ed.), Handbook of Cannabis and Related Pathologies: Biology, Pharmacology, Diagnosis, and Treatment. Elsevier B.V., pp. 13–23.

- Zirpel, B.[1], **Degenhardt, F.[1]**, Martin, C., Kayser, O., Stehle, F., 2017. Engineering yeasts as platform organisms for cannabinoid biosynthesis. J. Biotechnol. 259, 204–212.
  [1] Contributed equally to this study.

- Zirpel, B., **Degenhardt, F.**, Zammarelli, C., Wibberg, D., Kalinowski, J., Stehle, F., Kayser, O., 2018a. Optimization of Δ^9-tetrahydrocannabinolic acid synthase production in *Komagataella phaffii* via post-translational bottleneck identification. J. Biotechnol. 272-273, 40–47.

Poster presentations

- Degenhardt, F., Stehle, F., Kayser, O., 2016, Functional characterization of C-prenylating enyzmes. 9. Tag der Chemie, Fakultät für Chemie und Chemische Biologie, TU Dortmund, Germany

- Degenhardt, F., Stehle, F., Kayser, O., 2016, Functional characterization of C-prenylating enyzmes. 7th Annual CLIB-Graduate Cluster Retreat, Herdecke, Germany

- Degenhardt, F., Stehle, F., Kayser, O., 2015, Functional characterization of C-prenylating enyzmes. 6th Annual CLIB-Graduate Cluster Retreat, Wermelskirchen, Germany

- Degenhardt, F., Stehle, F., Kayser, O., 2014, Functional characterization of C-prenylating enyzmes. 5th Annual CLIB-Graduate Cluster Retreat, Lünen, Germany