# Identification, isolation and functional characterization of prenyltransferases in *Cannabis sativa* L.

Zur Erlangung des akademischen Grades eines

**Dr. rer. nat.**

von der Fakultät Bio- und Chemieingenieurwesen

der Technischen Universität Dortmund

Dissertation

vorgelegt von

**M. Sc. Kathleen Pamplaniyil**

aus

Viersen

**Dortmund 2016**

## Acknowledgments

I would like to thank all people who supported me throughout the years of my Ph.D. thesis.

First of all, I would like to express my gratitude towards **Prof. Dr. Oliver Kayser** (Chair of Technical Biochemistry, TU Dortmund) for his support and supervision throughout the time of my thesis. The discussions with him broadened my scientific knowledge and the work in his chair was instructive.

Furthermore, I want to thank **Dr. Felix Stehle** for guidance and support during the last two years of my thesis. The interesting discussions about the course of my experiments were inspirational.

In addition my sincere thanks are given to **Prof. Dr. Jörn Kalinowski** (Research group Microbial genomics and biotechnology, Center for Biotechnology) and **Oliver Rupp** (Chair Bioinformatics and Systems Biology, University Giessen) for their generous support during the in silico analysis.

Moreover, I want to express my appreciation to **Prof. Dr. Heribert Warzecha** (Department of Biology, TU Darmstadt) and **Jascha Volk** (Department of Biology, TU Darmstadt) who supported me with their knowledge and equipment during the transient expression in tobacco plants.

I am very thankful for the financial support of the **Ministry of Innovation, Science and Research** of the German Federal State of North Rhine Westphalia (NRW) and **TU Dortmund.**

Special thanks go to my colleagues especially **Verena, Schütz, Evamaria Gruchattka, Friederike Ullrich, Friederike Degenhardt, Bastian Zirpel** and **Laura Kohne**n for their contribution in scientific discussions and the good time we shared together besides the work in the lab.

I am deeply grateful to **my parents** who were encouraging and ideal examples throughout my whole lifetime and studies. My heartfelt thanks go also to the rest of **my family, my friends** and **Torsten** for their consistent encouragement, advice and mental support.

# Table of contents

# Table of abbreviations

| | |
|---|---|
| CBC | Cannabichromene |
| CBD | Cannabidiol |
| CBGA | Cannabigerolic acid |
| CBNRA | Cannabinerolic acid |
| *E. coli* | Escherichia coli |
| EST | expressed sequence tag |
| GFP | Green fluorescent protein |
| GHB | Geranylhydroxybenzoic acid |
| GOT | Geranylpyrophosphate:olivetolate geranyltransferase |
| GPP | Geranyldiphosphate |
| 4-PHB | 4-hydroxybenzoic acid; parahydroxybenzoic acid |
| OA | Olivetolic acid |
| OH | Hydroxy- |
| PT(s) | Prenyltransferase(s) |

# 1. Abstract

Prenylations contribute to the diversity of secondary metabolites. Prenylated compounds possess antibacterial and anti-inflammatory potential. Enzymes catalyzing the prenylation step by establishing new C-C or C-O bonds are prenyltransferases. One of these enzymes is involved in the biosynthesis of the psychoactive ingredient THC from Cannabis. Olivetolic acid was prenylated with geranyldiphosphate to Cannabigerolic acid, the precursor of THCA. The aim of this study was the identification of potential prenyltransferases able to catalyze that step.

The *in silico* analysis of genetic sequences was the first step to find new members for the family of aromatic, membrane-bound prenyltransferases. The search comprised the characterization of aromatic prenyltransferases based on their genetic sequence using web-based tools. Finding new enzymes of this family in DNA libraries was realized by establishing a specific mathematic algorithm, which distinguished between the different groups of prenyltransferases based on sequential differences. The mathematical search resulted in five putative prenyltransferases. Evolutionary analysis revealed the ancestral relation and classification into different groups. Based on these *in silico* results, certain cDNA sequences were used for the *in vivo* experiments.

Hosts for the *in vivo* analysis of the putative enzymes were *Nicotiana benthamiana* as well as *Saccharomyces cerevisiae* Y05416 cells. The transient expression in tobacco was executed with the innovative magnICON® system. For the expression in yeast cells two strategies for heterologous expression were performed: (i) cloning of cDNA sequences with its native (plant-derived) signal peptide (ii) cloning of cDNA sequences with the signal peptide from an enzyme complex from yeast.

One of the putative plant prenyltransferases was expressed actively in *Saccharomyces cerevisiae* Y05416. This result demonstrates that the established expression system can be used for expression of aromatic plant prenyltransferases in yeast as host cells under specific conditions.

The results of this thesis contribute to the successful expression of membrane-bound enzymes deriving from plants in *Saccharomyces cerevisiae* and enlarge the knowledge of the enzyme family of prenyltransferases. The thesis also illustrates different hosts for heterologous expression and the diversity of cloning strategies.

# Zusammenfassung

Prenylierte Substanzen tragen zur Vielfalt der Sekundärmetabolite bei. Prenylierte Metabolite besitzen entzündungshemmende und antibakterielle Eigenschaften. Die Enzyme, die die Prenylierung durch Bildung einer neuen C-C oder C-O Bindung katalysieren sind Prenyltransferasen. Eines dieser Enzyme ist an der Biosynthese des psychoaktiven Wirkstoffs THC aus Cannabis beteiligt. Olivetolsäure wird mit Geranyldiphosphat bei dieser Reaktion zu Cannabigerolsäure prenyliert, dem Ausgangsstoff von THCA. Das Ziel dieser Arbeit war die Identifikation von Enzymen die potentiell diese Reaktion katalysieren können.

Eine *in silico* Analyse war der erste Schritt, um neue Mitglieder der Familie der aromatischen, membrangebundenen Prenyltransferasen zu finden. Die Suche beinhaltete die Charakterisierung unterschiedlicher Gruppen von Prenyltransferasen bezüglich Unterschiede in ihrer genetischen Sequenz mithilfe von webbasierten Software-Programmen. Die Suche nach neuen Prenyltransferasen in DNA-Bibliotheken wurde durch einen spezifisch dafür entwickelten mathematischen Algorithmus verwirklicht, welcher die verschiedenen Gruppen von Prenyltransferasen anhand des Aufbaus ihrer genetischen Sequenz unterschied. Die Berechnung ergab fünf mögliche neue aromatische Prenyltransferasen. Die evolutionäre Verwandtschaft anhand der genetischen Sequenzen ermöglichte eine Einteilung in verschiedene Gruppen. Anhand der *in silico* berechneten Ergebnisse wurden die cDNA Sequenzen für *in vivo* Experimente bestimmt.

Wirtsorganismen für *in vivo* Experimente der möglichen Enzyme waren die Pflanze *Nicotiana benthamiana* und die Hefe *Saccharomyces cerevisiae* Y05416. Für die transiente Expression in Tabakpflanzen wurde das innovative Verfahren magnICON® verwendet. Für die heterologe Expression in Hefe wurden zwei Strategien verfolgt: (i) die Klonierung der cDNA Sequenzen mit dem nativen (pflanzeneigenen) Signalpeptid, (ii) die Klonierung der cDNA Sequenzen mit dem Signalpeptid des F1F0 ATP Synthase Enzymkomplexes aus Hefezellen.

Eines der möglichen Pflanzenenzyme wurde aktiv in *Saccharomyces cerevisiae* Y05416 exprimiert. Dieses Ergebnis zeigt, dass das verwendete Expressionssystem für die Expression von aromatischen Prenyltransferasen aus Pflanzen geeignet ist und Hefezellen als Plattformorganismus zur Expression von Pflanzengenen unter bestimmten Bedingungen genutzt werden kann.

Die gesamte Arbeit leistet einen Beitrag zur erfolgreichen Expression von Pflanzengenen in *Saccharomyces cerevisiae* und erweitert die Kenntnisse über die Familie der membran-gebundene Prenyltransferasen. Die Arbeit zeigt ebenfalls die Vielfalt an Klonierungsmethoden und Wirts-organismen auf.

# 2. Introduction

## 2.1 *Cannabis sativa* L., a plant with diverse qualities

*Cannabis sativa* L. is an annual dioecious plant and belongs to the family of Cannabaceae (Fischedick et al., 2010). It has been mostly used as intoxicant but also in medicinal therapy (Fischedick et al., 2010). More and more countries are legalizing small amounts of Cannabis for personal or medicinal usage or legalizing it completely. For example, Cannabis was legalized in four states of the USA until April 2015 and 14 further states have been decriminalized Cannabis ("http://www.pewresearch.org," 2016).

There are two morphological types of Cannabis plants: *Cannabis sativa* L. and *Cannabis indica* L. *Cannabis sativa* L. is tall and has more branches than the other plant type. The whole plant has a sparsely look concerning the leaves and the whole shape compared to *Cannabis indica* plants. This leads to a better exchange of air between the single branches and between neighboring plants, which results in a better, healthier growth and lesser fungal infestation. Most of this species are fiber type Cannabis plants. *Cannabis indica* L. is short, has broader leaves, thicker stems, darker green leaves and the type is generally more bushy and compact (Fischedick et al., 2010). It is used for medicinal and recreactional purposes (Fischedick et al., 2010). These two types are mostly used for hybridization experiments. There is a third subtype called *Cannabis ruderalis* (Hillig and Mahlberg, 2004). It has low levels of Cannabinoids. This subtype appears in Russia, central Europe and central Asia and grows mostly at roadsides or fallow arable land. The most prominent property of this subtype is the ability of auto-flowering independent from the photo period.

All types or subtypes of Cannabis have so called glandular trichomes. These hairy structure can be found on the surface of the leaves and especially on the surface of the flower buds. Trichomes are created from the epidermis of the plant and have different functions such as the defense against microorganisms, herbivores and insects as well as the temperature reduction of leaves and the homeostasis of ions (Schilmiller et al., 2008; Wagner, 1991). Trichomes have an important role in the biosynthesis and storage of secondary metabolites. Genes for the production of flavonoids and glucosinolates were found in a specific sort of trichomes (Dai et al., 2010). Furthermore, terpenoids and phenylpropanoids are established and stored in the oily parts of the trichomes (Wagner et al., 2004). A variety of prenylated compounds are also accumulated in these parts, because of their higher lipophilicity based on the prenyl group (Tsurumaru et al., 2012). Due to this characteristic, prenylated compounds have a stronger affinity for the binding to the cell membrane, which results in a higher bioavailability (Tahara and Ibrahim, 1995).

Glandular trichomes of *Cannabis sativa* L. can be divided in capitate-stalked, bulbous and capitate-sessile (Andre et al., 2016; Happyana et al., 2013). The capitate-stalked trichomes were identified as storage and place of biosynthesis of secondary metabolites (Dai et al., 2010; Kim and Mahlberg, 1997; Marks et al., 2009; Sirikantaramas et al., 2005). The disc cells, which are located in the head of the capitate-stalked trichomes, are supposed to be the location of the Cannabinoid biosynthesis (Kim and Mahlberg, 2003, 1997). This was proved by the presence of cDNAs, which were encoding for almost every step in the biosynthesis of the primary metabolites, which lead to the biosynthesis of $\Delta^9$-THC (Marks et al., 2009). The oily cavity surrounding the disc cells is suggested as the storage area of secondary metabolites (Kim and Mahlberg, 1997; Wagner et al., 2004). Besides Cannabis other plants for example *Humulus lupulus* L. (hop) possess these hairy structures on their surface as well. The glandular trichomes from hop are called lupulin glands and contain a variety of prenylated compounds such as humulone and lupulone as well as xanthohumol (Tsurumaru et al., 2012).

For medicinal effects the most important components in the plants are THC and Cannabidiol. These two Cannabinoids have the most important practical and verifiable effects during therapy with Cannabis. The content of THC in different plant varieties increased in the last years extremely. Nowadays plants (Bedrocan®, var. sativa) with up to 19% THC (in plant dry weight) are available (Ministry of Health, Welfare and Sports of the Netherlands, 2011). In contrast to that, the amount of CBD is running against zero. The plant material, which was used in the experiments of the present thesis was supplied by Bedrocan BV (Inc.). The company is licensed and contracted by the Dutch government. The plants are grown under Good Agricultural Practice (Fischedick et al., 2010). Other standardized varieties, which are available from Bedrocan BV (Inc.) are Bedrobinol® (Cannabis var. sativa, 12%THC), Bediol® (var. sativa, 6% THC and 7.5% CBD) and Bedica® (var. indica, 14% THC) (Ministry of Health, Welfare and Sports of the Netherlands, 2011; van Amsterdam et al., 2015).

## 2.2 Cannabinoids, endocannabinoids and their receptors

Besides its role as intoxicant, Cannabis is also known as medicinal active plant. Between the 1940s-1960s a group of terpenophenolic compounds was found in an extraction mixture of the plant (Fischedick et al., 2010). This group of compounds was referred to as Cannabinoids. Until then these terpenophenolics were only found in Cannabis. In the 1990s the discovery of the Cannabinoid receptors renewed the definition of Cannabinoids into compounds, which bind to these G-protein coupled Cannabinoid receptors (CBD1 or CBD2) (Burstein and Zurier, 2009; Fischedick et al., 2010). The discovery and functional characterization will be explained later in this chapter. The most famous secondary metabolite of *Cannabis sativa* L. is $\Delta^9$-THC ($\Delta^9$-Tetrahydrocannabinol), one of the psychoactive ingredients of the plant.

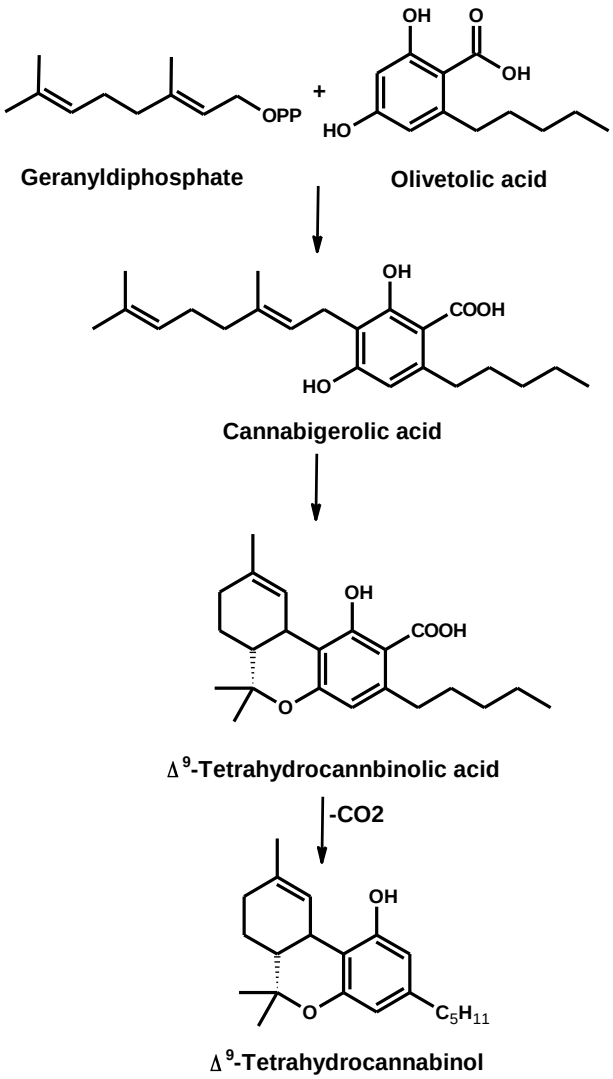



Figure 1: Biosynthetic pathway of $\Delta^9$-THC (Sirikantaramas et al., 2005)

Figure 1 illustrates the biosynthetic pathway of $\Delta^9$-THC. The precursor of the main psychoactive compound is Cannabigerolic acid (CBGA). This component is the result of the prenylation between olivetolic acid (OA) and geranyldiphosphate (GPP). The reaction is catalyzed by a membrane-bound aromatic prenyltransferase named geranylpyrophosphate:olivetolate geranyltransferase (GOT) (Fellermeier and Zenk, 1998). In further research a DNA sequence was published and patent-registered (Page and Boubakir, 2011). This enzymatic step is the main issue of the present thesis. The subsequent cyclization of CBGA results into $\Delta^9$-THCA, which is decarboxylated to $\Delta^9$-THC by heating or drying of the plant after the harvest (Hillig and Mahlberg, 2004).

CBGA is also required to build up two other Cannabinoids: CBDA (Cannabidiolic acid; Cannabidiol) and CBCA (Cannabichromenic acid; Cannabichromene) (Hillig and Mahlberg, 2004). These two secondary metabolites are isomers of THCA with different ring structures (Morimoto et al., 1998). The cyclization of CBGA to CBCA is performed by an Oxidoreductase. CBDA is not only generated

from CBGA but also from CBNRA (Cannabinerolic acid, Cannabinerol), which is a z-isomer of CBGA. The conversion into CBDA is an Oxidocyclization (Taura et al., 1996). Based on the lower substrate specificity of the CBDA synthase to CBNRA compared to CBGA, it is suggested that CBDA is primarily synthesized from CBGA (Taura et al., 1996). CBDA Synthase shows lower $K_{cat}$ and higher $K_M$ values for CBNRA than for CBGA (Taura et al., 1996). CBN results from the degradation of THC and is therefore found in aged Cannabis plants (Andre et al., 2016). This cannabinoid affects the cell of the immune system (Andre et al., 2016).

THC is the most prominent Cannabinoid amongst the others. It is publicly known as the psychoactive ingredient in marijuana. But it has also beneficial effects as medicine such as anti-inflammatory, anti-cancer, analgesic, muscle relaxant, neuro-antioxidative activity (De Petrocellis et al., 2011). The other very frequently appearing compound is Cannabidiol. This non-psychoactive Cannabinoid is the most abundant cannabinoid in the fibre-type plants (Andre et al., 2016; Burstein and Zurier, 2009). CBD reduces intestinal inflammations in experiments with mice and exerts anti-inflammatory, anti-anxiety, anti-nausea and anti-psychotic activity (Burstein, 2015; Burstein and Zurier, 2009). It also reduces undesirable effects of THC such as sedation, psychotropic effects and tachycardia (Burstein and Zurier, 2009). As therapeutic agent it shows potential in the treatment of epilepsy, neurodegenerative diseases, schizophrenia, multiple sclerosis and affective disorders (Hill et al., 2012). Cannabigerolic acid, the product of the prenylation step in the biosynthetic pathway, is not that familiar in the field of research. A slightly bit more is known about the decarboxylated form cannabigerol (CBG). It is a non-psychotropic ingredient in Cannabis plants and exerts antibacterial activity (Appendino et al., 2008). CBG is also involved in the reduction of intraocular pressure, which results in beneficial effects in the treatment of glaucoma (Colasanti, 1990; Colasanti et al., 1984). Cannabigerol is also used in the treatment against the inflammatory bowel disease, which is an incurable disease of the abdomen (Borrelli et al., 2013). Cannabichromene is another cannabinoid, which has to be mentioned. This component has anti-inflammatory, sedative, analgesic, anti-bacterial and anti-fungal properties (Andre et al., 2016; Delong et al., 2010; Eisohly et al., 1982).

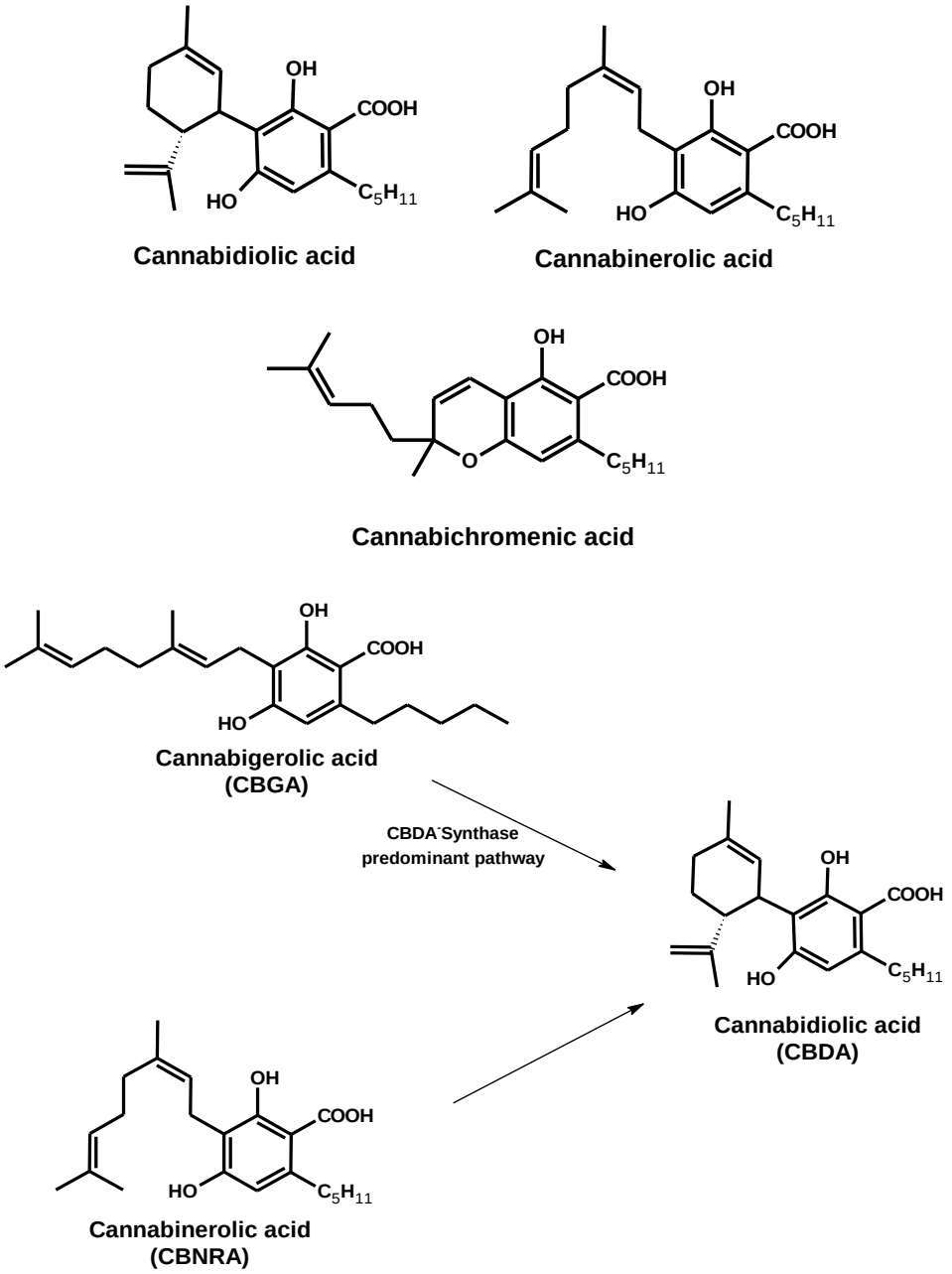



Figure 2: Chemical structure of Cannabinoids (CBDA, CBNRA, CBCA) and CBDA biosynthesis (Maione et al., 2011; Taura et al., 1996)

Besides the Cannabinoids extracted from Cannabis plants, Cannabinoid-like compounds were also found in different species of liverwort. These components have structural similarity to THC and other Cannabis components. In 1994 a bibenzyl Cannabinoid was found for the first time in ether extracts of the liverwort *Radula perrottetii* (Toyota et al., 1994). 2002 the perrottetinenic acid was isolated from *Radula marginata* with the already discovered perrottetinene and two new bibenzyls (Toyota et al., 2002).

There are also other natural occurring compounds, which activate the cannabinoid receptors and are derivatives of long-chain fatty acids. These compounds are summarized as endocannabinoids and are

produced autonomously by every organism (Burstein and Zurier, 2009). These substances belong to the super family of eicosanoids and are generated from lipid precursors (Burstein and Zurier, 2009). Prominent members of this group of components are shown in Figure 3.

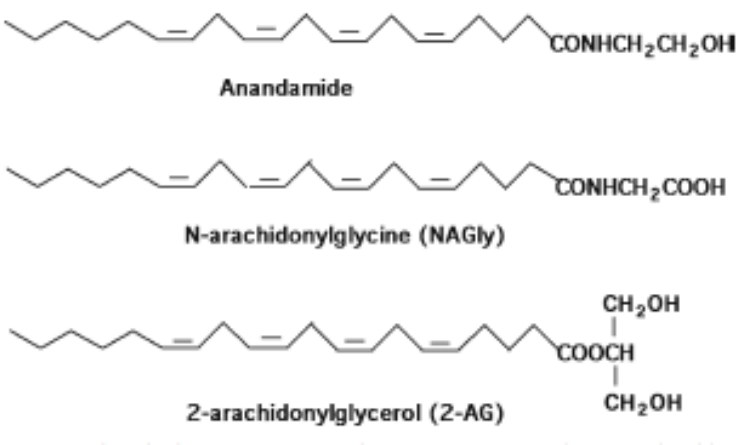



Figure 3: Prominent endocannabinoids (Burstein and Zurier, 2009)

Endocannabinoids are released from neurons (Burstein and Zurier, 2009). They are able to activate the cannabinoid receptors in the same or adjacent cells (Burstein and Zurier, 2009). The metabolism of the endocannabinoids can be performed rapidly by an enzyme called fatty acid amide hydrolase (FAAH) based on a specific serine hydrolysis (Burstein and Zurier, 2009; Deutsch et al., 2000; Giuffrida et al., 2001). There are two other enzymes which are also able to metabolize endocannabinoids: monoglyceride lipase and N-acylethanolamine hydrolyzing acid amidase (Ueda et al., 2005). Different levels of endocannabinoids were determined due to gender and hormonal cycle (Bradshaw and Walker, 2005).

In the animal kingdom Endocannabinoids have a wide range of function (Burstein and Zurier, 2009). For example the control of sensorimotor and motivational aspects of behavior (Giuffrida and Piomelli, 2000), the regulation of multiple physiological and pathological processes in the reproductive system in humans (Schuel et al., 2002), effects on cognition (Lichtman et al., 2002), effects on the cardiovascular system (Kunos et al., 2000), their role in the immunological adaption occurring during the early stages of pregnancy (Maccarrone et al., 2002) and in pain regulation (Walker and Huang, 2002), inflammatory pain, gut and bone physiology and cancer (Chiurchiù et al., 2014; Hanus, 2009).

The discovery of the cannabinoid receptors was based on the observation of the decrease of cyclic AMP accumulation in neuronally derived cells by $\Delta^9$-THC. The response suggested that the presence of specific receptors for THC (Howlett, 1984). A G-protein coupled receptor was first discovered in cells of the rat brain (Devane et al., 1988). The high affinity and stereoselectivity towards different

analgesic cannabinoids were demonstrated. The optimal pH value was determined between 7 and 8 and the specific binding was removed at a reaction temperature of 60°C (Devane et al., 1988). In the beginning of the 90s the cDNA of Cannabinoid receptors from rat and human brains were expressed in immortalized cell lines (Gerard et al., 1991; Matsuda et al., 1990). The translated protein sequence contained 472 amino acids (Matsuda et al., 1990). It was confirmed, that the protein belongs to the family of G-protein coupled receptors by numerous highly conserved residues amongst this group of receptors. Additionally seven hydrophobic domains and several potential glycosylation sites were detected (Matsuda et al., 1990). The receptor was called CB1 and showed 97.3% sequence homology comparing the amino acid sequence between the rat and human protein (Matsuda et al., 1990).

A second Cannabinoid receptor was detected in 1993 (Munro et al., 1993). The cDNA of the receptor was identified in the human promyelocytic leukaemic cell line HL60. Initially six cDNA clones were identified as G-protein coupled receptors by amplification during PCR with degenerated primers. One sequence showed 48% sequence similarity to the already identified Cannabinoid receptor CB1 and was subsequently named CB2 (Howlett et al., 2002; Munro et al., 1993). Homologues of this CB2 can be found in mouse, rat and human tissues (Gerard et al., 1990, 1991). While the Cannabinoid receptor CB1 is located in brain and neuronal cells (central receptor), the peripheral CB2 receptor is found in immune cells (Howlett et al., 2002). Both cannabinoid receptors are associated to adenylyl cyclases and mitogen-activated protein kinases through the G proteins (Hanus, 2009). CB1 receptors are additionally connected to several calcium and potassium channels (Hanus, 2009).

However, $\Delta^9$-THC acts as agonist for the Cannabis receptors CB1 and CB2, but shows higher affinity towards CB1 receptors (Andre et al., 2016; van Amsterdam et al., 2015). Besides its positive effects mentioned above, THC can also have side effects like anxiety, immunosuppression and cholinergic effects (Andre et al., 2016). CBD, the most abundant non-psychoactive compound, reduces these side effects (Englund et al., 2012). CBD does not bind to CB1 or CB2 and is a functional antagonist to the endocannabinoid receptor GPR55 (Burstein, 2015). But the detailed mode of action of CBD is mostly unknown (van Amsterdam et al., 2015). Moreover CBD increases the receptor-mediated effects of the endocannabinoid anandamide by inhibition of the degradation and the following reuptake (van Amsterdam et al., 2015). CBG is supposed to bind to CBD2, but it is not known yet if it acts as agonist or antagonist (Borrelli et al., 2013). CBC acts as inhibitor for the anandamide uptake (De Petrocellis et al., 2011). CBN is also an agonist for the cannabinoid receptors CB1 and CB2, but comparing to THC the binding affinity is lower (Andre et al., 2016).

Besides CB1 and CB2 there exists another receptor, which accepts several natural but also synthetic endocannabinoids (Petitet et al., 2006). The receptor GPR55 was identified and cloned in 1999 (Sawzdargo et al., 1999; Wagner et al., 1999). In contrast to the other two receptors this

endocannabinoid receptor cannot be activated by plant cannabinoids (Hanus, 2009). Nevertheless, the inactivation of this receptor can be performed by the CB1 antagonist SR141716A and Cannabidiol (Chiurchiù et al., 2014; Hanus, 2009). GPR55 was identified from an expressed sequence tags (EST) database, consists of 319 amino acids and possesses the seven transmembrane segments, which represents a structural motif of the G protein-coupled receptor family (Sawzdargo et al., 1999). The receptor is localized in many tissues such as brain, spleen, bone, vascular and endothelial cells (Chiurchiù et al., 2014; Hanus, 2009). GPR55 was supposed to have a prominent role in inflammatory pain and the pathophysiology of the gut (Schicho and Storr, 2012; Yang et al., 2015).

Due to the low sequence homology and agonist binding specificity between GPR55 and CB1/CB2, it is controversially discussed if GPR55 is a third cannabinoid receptor (Liu et al., 2015). For example, the experiments of Kapur et al. in 2009 illustrate that AM251, SR141716A (cannabinoid antagonists) and lysophosphatidylinositol (LPI, no cannabinoid receptor ligand) are binding to GPR55 as agonists (Kapur et al., 2009). Another synthetic cannabinoid agonist (CP55,940) is acting as antagonist/partial agonist. Kapur et al. suggest that GPR55 at the most is an atypical cannabinoid responder (Kapur et al., 2009). Certainly there has to be more research in this area especially including more agonists and specific ligands for cannabinoid receptors (Liu et al., 2015).

2010 six other orphan GPCR receptors (GPR18, GPR34, GPR119, GPR160, GPR183 and P2RY10) were discovered using a robust quantitative PCR array in metastases and benign nevi cells (Qin et al., 2010). They were significantly differentially expressed between the two cell types. Further experiments in yeast and melanoma cells revealed that GPR18 is constitutively active and inhibits apoptosis (Qin et al., 2010). Further research concerning GPR18 was performed by Takenouchi et al. in 2012. In this study the effects and mode of action of N-arachidonoyl glycine (NAGly) were investigated (Takenouchi et al., 2012). Functionally it possesses analgesic effects, anti-inflammatory activity, the ability to inhibit cell proliferation and to mobilize calcium ions (Burstein, 2008). NAGly leads to the activation of the MAPKs (mitogen-activated protein kinase) and the mitochondrial apoptotic pathways. In the study the effects of NAGly administration on macrophages and the affected pathways was examined. NAGly was supposed to bind to the new orphan GPR18 receptor (Takenouchi et al., 2012). The knock out of this protein resulted in a decrease of the apoptotic effects of NAGly. In further experiments GPR18 expression was increased in mouse macrophages. Addition of NAGly led to the increase of apoptosis comparing to cells with normal GPR18 level. The results of the experiments insinuate that the receptor GPR18 might be involved in the NAGly-induced apoptosis and therefore could count as a cannabinoid receptors (Takenouchi et al., 2012).

Further investigations of GPR18 focused on the function of the receptor in HEK cells (Finlay et al., 2016). It was possible to heterologously express the protein intracellular stable. Whereas the cell

surface expression was performed in transiently transfected cells in which the overall expression was higher. The results of the experiments indicate, that the receptor undergoes a rapid membrane trafficking as a constitutive active receptor (Finlay et al., 2016). This property is a very interesting part of investigation of GPR18 and subsequent effects of GPR18 in the cell physiology (Finlay et al., 2016).

Cannabinoids cannot only be administered as dried flowers or leaves by smoking. There are several drugs available which contain different cannabinoids or combination of different cannabinoids. One example of a cannabinoid containing drug is Sativex® (GW Pharmaceuticals, UK). This cannabis extract is applied as oromucosal spray containing THC and CBD (1:1) as active components (Ben Amar, 2006; Fernandez, 2016). Based on the extraction, there are also byproducts present. It is used effectively against moderate to severe spasticity occuring during multiple sclerosis disease (Fernandez, 2016). Sativex® is applied if other medications did not lead to the desired effect. Cannador® (Society for Clinical Research, Germany) is another Cannabis extract based drug. The disparity between Sativex® and Cannador® is the ratio of THC:CBD. In Cannador® it adds up to 2:1 (Hazekamp and Grotenhermen, 2010). It is administered as capsule and is reducing muscle stiffness, spasm and pain during multiple sclerosis therapy. Marinol® (Solvay Pharmaceuticals, Belgium) is synthetic THC dissolved in sesame oil. The drug is applied as a capsule (Hazekamp and Grotenhermen, 2010). It is used as appetite stimulate to gain body weight during the treatment of anorexia and the therapy of AIDS (Hazekamp and Grotenhermen, 2010; Pharmacology, 2004). It can also be used against nausea and vomiting during cancer therapy (Fischedick et al., 2010; Hazekamp and Grotenhermen, 2010). Another synthetic THC drug is Nabilone (Valeant Pharmaceuticals International, USA). It contains a THC analogue and is marketed in some countries as Cesamet®. It is also prescribed against nausea and vomiting, which are a side effects during cancer therapy and after medical surgeries (Fernandez, 2016; Hazekamp and Grotenhermen, 2010).

## 2.3 Prenyltransferases

The biosynthetic pathway of $\Delta^9$-Tetrahydrocannabinol ($\Delta^9$-THC) is shown in Figure 1 (Sirikantaramas et al., 2005). The precursor of $\Delta^9$-THC is Cannabigerolic acid (CBGA). This compound is synthesized by an alkylation of olivetolic acid (OA) with geranyldiphosphate (GPP) (Figure 1). The reaction is catalyzed by a prenyltransferase (PT), which forms a new C-C bond between the aromatic acceptor and the prenyl donor molecule (Fellermeier and Zenk, 1998).

Prenyltransferases (PTs) can be divided into different families. PTs contribute to the diversity of chemical structures and biological activities of secondary metabolites in nature like flavonoids, coumarins, xanthons, phloroglucinols, polyketides (Dhillonz and Brown, 1976; Fellermeier and Zenk,

1998; Hamerski et al., 1990; Sasaki et al., 2008; Sirikantaramas et al., 2005; Yamamoto et al., 2000; Zhao et al., 2003; Zuurbier et al., 1998) and can be found in animals, bacteria and plants (Ohara et al., 2009). The prenylated products can possess antiviral, antibacterial and anticanceral activities (Sirikantaramas et al., 2005; Wang et al., 2014).

The enzyme catalyzing the prenylation of OA and GPP belongs to the family of membrane-bound aromatic prenyltransferases (Fellermeier and Zenk, 1998). Members of this family possess a conserved aspartate-rich motif called UbiA-motif and bind $Mg^{2+}$ ions during the reaction (Heide, 2009). This enzyme will be described later in this chapter in detail.

Another family is the protein prenyltransferase, which is relevant in posttranslational modification of proteins in eukaryotes and can be addressed as drug target (Heide, 2009). A member of this family is for example Ras farnesyltransferase (Long et al., 2002). This enzyme possesses a conserved CAAX-motif and binds $Zn^{2+}$ during the reaction.

Two classes of isopentenyltransferases belong also to the family of PTs. These enzymes are responsible for the cytokinin synthesis (ATP/ADP and tRNA isopentenyltransferase) (Miyawaki et al., 2006).

Additional PT-families are chlorophyll synthases in plants, which catalyze the biosynthesis of bacteriochlorophyll *a* and enzymes, which are involved in the biosynthesis of archeal lipids (Geranylgeranylglyceryl phosphate prenyltransferase) (Payandeh et al., 2006).

A large family of PTs comprises the aromatic prenyltransferases which can be furthermore divided into subfamilies. In the following three of the subfamilies will be introduced: the parahydroxybenzoic acid (PHB/UbiA) prenyltransferases, the ABBA family and the fungal indole PT-family.

Members of the subfamily of parahydroxybenzoic acid PTs, which is also known as UbiA family, are involved in the biosynthesis of ubiquinones (Kawamukai, 2002; Young et al., 1972), plastoquinones (Sadre et al., 2006; Swiezewska et al., 1993), menaquinones (Schledz et al., 2001; Suvarna et al., 1998), bacterial cell wall precursors (Huang et al., 2005) and vitamin E (Collakova and Dellapenna, 2001; Sirikantaramas et al., 2005). During the $Mg^{2+}$-dependent reaction the prenyl moiety from the donor molecule is added to an aromatic acceptor molecule. Due to this reaction both substrates are connected via an emerging C-C bond. The reaction is illustrated in Figure 4.

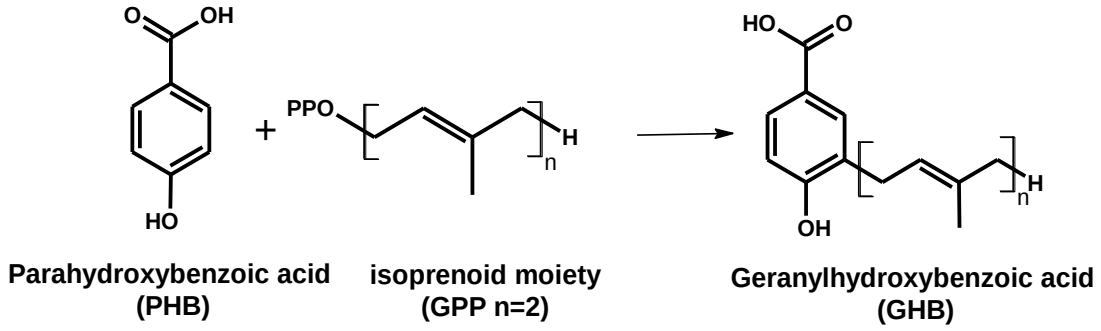



Figure 4: Representative reaction of UbiA prenyltransferases (Bräuer et al., 2008)

The family of PHB-PTs is named after the para-hydroxybenzoic acid prenyltransferase, which is involved in the synthesis of ubiquinone (Ashby et al., 1992; Forsgren et al., 2004; Ohara et al., 2006; Okada et al., 2004; Siebert et al., 1992; Suzuki et al., 1994). Ubiquinone is involved as a redox carrier in the respiratory pathway and cell wall biosynthesis. Members of this subfamily have been identified in eukaryotes and prokaryotes (Ashby et al., 1992; Forsgren et al., 2004; Ohara et al., 2009; Okada et al., 2004; Siebert et al., 1992; Suzuki et al., 1994). The enzymes show a broad acceptance for prenyl diphosphates with different chain lengths and have one aspartate-rich conserved region (NDXXDXXXD) (Huang et al., 2014; Ohara et al., 2013). In general the nature of the substrates differ considerably (Huang et al., 2014).

In *Escherichia coli* this enzyme is called p-hydroxybenzoic acid oligoprenyltransferase (UbiA). This enzyme consists of 290 amino acids and is involved in the synthesis of quinones and bactoprenol (Young et al., 1972). The homologue of this gene in *Saccharomyces cerevisiae* is Coq2. The prenyltransferase consists of 372 amino acids and is involved in the biosynthesis of ubiquinone (Forsgren et al., 2004). The protein possesses a mitochondrial signal sequence at the N-Terminus. The PT-facilitated reaction is $Mg^{2+}$ dependent or similar cations like $Mn^{2+}$, $Co^{2+}$ or $Ni^{2+}$ are necessary (Heide, 2009). The enzymes of this class are integral membrane proteins and possess several transmembrane helices (Forsgren et al., 2004; Heide, 2009). The proteins accept short chain isoprenoid diphosphates except DMAPP and *cis*-prenyl diphosphates (Heide, 2009).

Bräuer et al. published a 3D model of UbiA (Bräuer et al., 2008). Based on the proposed model, the enzyme possesses a deep lipophilic pocket, which can be responsible for the acceptance of the different isoprenoid chain lengths. The diphosphate group of the prenyl donor is bound through Arg137 (R137) and stabilized by the $Mg^{2+}$ ion (Bräuer et al., 2008). The cation itself is stabilized by Asp71 and 75 (D71, D75). On the other side the Arg137 forms salt bridges with two aspartates (D191, D195) (Bräuer et al., 2008). The second substrate, *p*-hydroxybenzoic acid, is on the one side bound through Arg72 in the referring protein model. On the opposite side the phenolic group generates hydrogen bonds to the Asp191 (Bräuer et al., 2008). Furthermore there are several proton transfers

suggested between the phosphate groups, the isoprenoid moiety and the *meta*-position of PHB during the final addition of the prenyl moiety (Bräuer et al., 2008). The active site of UbiA, the participating amino acids, the two substrates (green: PHB; orange and violet: GPP) and the divalent cation are shown in Figure 5.

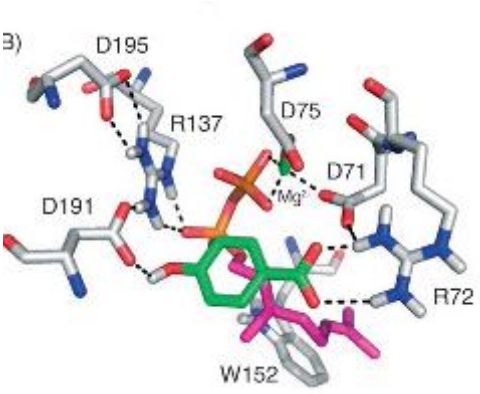



Figure 5: Structural model of the active site of UbiA (Bräuer et al., 2008); green: PHB; orange and violet: GPP

In 2014 the 3D-structure of an UBIAD1 homolog from *Archaeglobus fulgidus* (AfubiA) was solved by X-ray crystallography (Huang et al., 2014). The results of this group support the findings of Bräuer et al., that the divalent cation and the isoprenoid moiety are bound in a chamber inside the enzyme protected from the surrounding solvent. Due to mutation studies, which were also carried out in this publication, the demonstrated results can be transferred to other members of the UbiA family (Huang et al., 2014).

The diversity of secondary metabolites in plants is highly regarded to the presents of prenyltransferases. Plant prenyltransferases from the UbiA family were first functionally expressed in 2002 (Yazaki et al., 2002). In the study the amino acid sequences of prenyltransferases involved in the ubiquinone biosynthesis were used in a PCR reaction against the cDNA of *Lithospermum erythrorhizon* (Yazaki et al., 2002). It was possible to isolate two cDNA sequences and express the proteins in *Saccharomyces cerevisiae* cells with Coq2 deficiency. The enzymes were named LePGT-1 and LePGT-2 (Yazaki et al., 2002). Both proteins showed specificity against the isoprenoid geranyldiphosphate (GPP) in contrast to the other UbiA enzymes (Yazaki et al., 2002). Furthermore, the enzymes are located in the endoplasmic reticulum differently to the other enzymes of the UbiA family (Yazaki et al., 2002). However, both proteins possess the common motif NDxxDxxxD and are depending on $Mg^{2+}$ (Yazaki et al., 2002). The reactions of LePGT-1 and LePGT-2 are identical to that of Coq2 (Figure 4), despite the difference in the localizations (Heide, 2009).

According to these first findings in plants analogues of these genes were searched in *Cannabis*. However, a Geranylpyrophosphate:olivetolate geranyltransferase (GOT) was identified in young leaves of the *Cannabis sativa* L. variety Skunk (Fellermeier and Zenk, 1998). Two products were detected after the enzyme assay with crude plant extract, which were not discovered in the control sample using heat denatured protein (Fellermeier and Zenk, 1998). The major product shows exactly the same spectrum as synthesized CBGA (Fellermeier and Zenk, 1998). The second product was identified as CBNRA, which is the z-isomer of CBGA as mentioned above. It was suggested that a geranylpyrophosphate isomerase was present in the extract, which produces nerylpyrophosphate, which subsequently results in CBNRA (Fellermeier and Zenk, 1998). The product ratio between CBGA and CBNRA yields 2:1 after the addition of GPP to the enzyme assay. The enzyme shows specificity towards olivetolic acid, but not towards the chain lengths of the isoprenoid moiety. The protein does also not accept olivetol as substrate. The heterologous expression of this gene was not described. But in further research a DNA sequence was published and patent-registered (Page and Boubakir, 2011).

A very similar plant prenyltransferase to GOT was discovered in *Humulus lupulus* L., which is generally known as hop. It is used for the bitter flavor in beer. The plant just as well as *Cannabis sativa* L. belongs to the family of Cannabaceae. Both plants are related to each other and possess glandular trichomes (Wagner, 1991). A cDNA library was built up from the lupulin glands (Tsurumaru et al., 2010). The aromatic prenyltransferase named HlPT-1 was found amongst the ESTs. The sequence contains the aspartate-rich motif, the transmembrane helices and a transit peptide (Tsurumaru et al., 2010). The GFP-fused protein was analyzed and localized in the plastids of the plant cells (Tsurumaru et al., 2010). It was possible to express HlPT-1 in baculovirus infected insect cells. During the $Mg^{2+}$-dependent reaction a dimethylallyl residue is transferred to phloroglucinol derivatives to form humulone and lupulone derivatives (Tsurumaru et al., 2012). HlPT-1 possesses a broad substrate spectrum and the optimal pH value was determined around pH 7.0 (Tsurumaru et al., 2012). The expression in insect cells does not reach high amounts of the intended protein, but the heterologous expression of membrane-bound proteins is easier in these cells.

Besides the membrane-bound PTs, there are also soluble PTs, which were identified in bacteria and fungi (Jost et al., 2010). The soluble PTs can be divided into three groups: the ABBA family, the bacterial and fungal indole PTs.

Members of the ABBA family are named after their novel α/β-barrel folds in their structure and are primarily found in bacteria (Heide, 2009; Jost et al., 2010). Examples of this group of PTs are CloQ (4-hydroxyphenylpyruvate: dimethylallyl transferase) and NphB (hydroxynaphthalene PT), which were found in different *Streptomyces* strains (Kuzuyama et al., 2005; Pojer et al., 2003). Both enzymes do

not need $Mg^{2+}$ ions or possess the typical aspartate-rich PT motif (Pojer et al., 2003). Furthermore, CloQ does not accept 4-hydroxybenzoic acid as substrate. Even though the enzymes accept a wide variety of substrates due to their expanded central cavity, which consists of 10 antiparallel β-strands surrounded by 10 α-helices (Heide, 2009; Yang et al., 2012). NphB for example accepts flavonoids, isoflavonoids and olivetol and catalyzes C- as well as O-prenylations (Kumano et al., 2008).

Another subfamily of soluble aromatic prenyltrasferases are fungal indole PTs (Heide, 2009). In general the enzymes prenylate indole moieties and produce different alkaloids (Tsai et al., 1995). One example of a fungal indole PT is the 4-dimethylallyl tryptophan synthase (DMATS) from *Claviceps purpurea*, which is involved in the biosynthesis of ergot alkaloids (Tsai et al., 1995). Members of this PT-family do not possess the typical (N/D)XXD motif and are not dependent on the presence of $Mg^{2+}$ ions (Heide, 2009). The PTs catalyze regular and reverse C- and N-prenylations. These properties are similar to those of the PTs from the ABBA family, but the primary sequence of these two PT family members differ from each other (Heide, 2009).

One example for a bacterial indole prenyltransferase is IptA. The enzyme was discovered while a homology search of PT sequences from known enzymes against the *Streptomyces sp.* SN-593 genome. The reaction catalyzed by IptA is the transfer of a dimethylallyl group to the C6 position in the indole ring (Takahashi et al., 2010). The protein shows a broad acceptance spectrum towards indole derivatives. In conclusion, IptA can be categorized as indole PT, even though phylogenetic analysis shows differences between bacterial and fungal indole PTs (Takahashi et al., 2010).

## 2.4 Bioinformatical sequence analysis

Due to the fact that Cannabinoids are of medicinal relevance, the synthesis of these terpenoids is an interesting field of research. Since the chemical synthesis is expensive and complicated, because several steps are necessary, the alternative way could be the heterologous expression of the involving enzymes (Bräuer et al., 2008). The successful heterologous expression of the THCA-Synthase in *Pichia pastoris* was reported by Taura et al. (Taura et al., 2007). As mentioned above the precursor CBGA is biologically synthesized by a membrane-bound aromatic prenyltransferase (Fellermeier and Zenk, 1998). The DNA sequence of the enzyme is published (Page and Boubakir, 2011). The characteristic of several transmembrane α-helices is the biggest hurdle of a high-yield, heterologous expression and subsequent purification (Ohara et al., 2013).

One possibility to solve the problem of the heterologous expression with this specific genetic sequence, could be the search for alternative DNA sequences in different DNA libraries of Cannabis plants, which are genetically related to the enzyme and may be expressed easier than the actual

protein. For the discovery of new gene sequences of enzymes, bioinformatic tools and modular systems are very important (Sharma and Sarkar, 2013).

The field of biochemistry has seen a growing importance of informatic methods over the last decades. The whole genome shotgun (WGS) sequencing introduced the automated sequencing, which is an important step towards sequencing of large amounts of sequences (Bekel et al., 2009). Simultaneously to the publications of genome sequencing projects, the first software programs to analyze certain aspects of the sequences were published. Examples for these kind of single analyzing tools are a base-calling and quality clipping program or assembly of the sequences (Ewing et al., 1998; Gordon et al., 1998). The further development resulted in the first modular or tool systems like the Staden package, which combined the assembly, editing and fishing of genomic sequences and resulted in more efficient data analysis (Bekel et al., 2009; Staden et al., 2000).

A further step of development of the sequencing was the appearance of expressed sequence tags (ESTs). These comparably small sequences are derived from cDNA. cDNA is produced from mRNA. The number of genetic copies can differ due to external conditions like temperature, exposure to light source, pH value or else (Shcheglov et al., 2007). The normalization of a library is a solution to solve the problem of this inequality of gene copies and to identify enzymes even on low level of expressed genes (Shcheglov et al., 2007).

ESTs represent the transcripts under certain experimental conditions (Bekel et al., 2009). The first large-scale publication of ESTs was performed in the early 90s (Adams et al., 1991). The frequency of occurrence of ESTs is correlated with transcription activities on the corresponding genes under certain conditions (Bekel et al., 2009). A huge number of EST datasets are stored at NCBI and TIGR institute, which are also publicly available (Quackenbush et al., 2000). Various publicly available platforms were established, in which ESTs can be analyzed, annotated and classified in gene ontology (Bekel et al., 2009). In this mass of sequences it is one of the most important qualities to identify the function of a corresponding gene and protein.

In the present study two bioinformatics platforms were mainly used: First of all “The Sequence Analysis and Management System (SAMS)” and secondly GenDBE (Bekel et al., 2009; Meyer et al., 2003). Features of SAMS are the automatic recognition of the integrated data type, processing and storing of the huge amount of sequence data (Bekel et al., 2009). Capabilities of the platform are among others the evaluation of the whole genome shotgun sequence quality, clustering and annotation of ESTs and data processing from high-throughput sequencing (Bekel et al., 2009).

GenDB is an open-source platform and was developed for the annotation of prokaryotic genomes (Meyer et al., 2003). It can be used for manual or automatic annotation of genetic sequences. It is a

modular system and can be linked to other analyzing systems via application programmers interface (API) (Meyer et al., 2003). GenDB is a tool package allowing the identification, classification and annotation of genes ("https://www.uni-giessen.de/fbz/fb08/Inst/bioinformatik/software/gendb," 2016). The system can be used for the first steps of the annotation of genetic sequences ("https://www.uni-giessen.de/fbz/fb08/Inst/bioinformatik/software/gendb," 2016). With GenDB it is possible to identify the function by using different prediction tools like for example BLAST, InterProScan or HMMER 2/3. The classification of the gene is performed by EC numbers or categorizing it into gene ontology categories. The further development of the system tools enables the integration of eukaryotic genomes to analyze for example the genomes of sugar beet or the Chinese hamster (Rupp et al., 2014). GenDB(E) is a continuing development of the prokaryotic version of a web-based interface. This allows an easy accessibility to the modular system ("https://www.uni-giessen.de/fbz/fb08/Inst/bioinformatik/software/gendb," 2016).

These two platforms were used for the first overview of the EST dataset from the cDNA library of *Cannabis sativa* L. var. Bediol. The template for the homology search was the genetic sequence of GOT (Fellermeier and Zenk, 1998; Page and Boubakir, 2011). Similar to the study of Gao et al. a normalized cDNA library was generated and functionally annotated (Gao et al., 2011). For further discovery the conserved motif of UbiA PTs was implemented in bioinformatics tools by programming a PT-search.

Similar studies were already performed for example by Akashi et al. to detect new prenyl-transferases, which are involved in the biosynthesis of secondary metabolites of soybean. These metabolites are important in the resistance mechanism of the plant against diseases (Akashi et al., 2008). Another study dealt with monoterpene synthases from *Citrus limon*, which represent new members of the tpsb family and are involved in the synthesis of limonene, β-pinene and γ-terpinene (Lücker et al., 2002). These genes were functionally expressed in *E. coli* cells due to the exclusion of the plastid-signal sequence (Lücker et al., 2002).

Other publications comprise the establishing and assembly of cDNA libraries and their overall functional characterization without the detailed analysis for new members of specific enzyme families (Ling et al., 2007; Long et al., 2012). These studies can be good starting points for the analysis of specific protein groups and subsequent expression, which are responsible for different biosynthesis pathways for secondary metabolites in plants or other organisms.

# 3. Materials and Methods

## 3.1 Molecular biology

### 3.1.1 Organisms

#### *3.1.1.1 Bacteria*

Table 1: Overview of used bacterial strains

| Organism | Genotype | Source/reference |
|---|---|---|
| ***Agrobacterium tumefaciens* GV3101** | C58 (rif R) Ti pMP90 (pTiC58DT-DNA) (gentR/strepR) Nopaline | Department of Biology, TU Darmstadt (Prof. Dr. Heribert Warzecha); (Koncz and Schell, 1986) |
| ***Agrobacterium tumefaciens* LBA4404** | TiAch5 (rif R) Ti pAL4404 (specR/strepR) Octopine | Department of Biology, TU Darmstadt (Prof. Dr. Heribert Warzecha); (Hoekema et al., 1983) |
| ***Escherichia coli* DH5α** | F– Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *rel*A1 *end*A1 *hsd*R17($r_K^-$ $m_K^+$) *pho*A *sup*E44 λ– *thi*-1*gyr*A96 | Department of Technical Biochemistry, TU Dortmund; (Taylor et al., 1993) |

The antibiotics rifampicin and gentamycin were added to LB medium when using GV3101 cells. For usage of LBA4404 cells rifampicin and streptomycin were added to LB medium. Both strains contain a genomic rifampicin resistance. Additionally, both strains carry Ti helper plasmids with the other antibiotics' resistance cassette (Table 5), respectively.

#### *3.1.1.2 Yeasts*

Table 2: Overview of used yeast strains

| Organism | Genotype | Source/reference |
|---|---|---|
| ***Saccharomyces cerevisiae* Y05416** | MATa; ura3Δ0; leu2Δ0; his3Δ1; met15Δ0; YNR041c::kanMX4 | Euroscarf |
| **CEN.PK2-1C** | MATa; ura3-52 MAL2-8C SUC2 | Euroscarf/(Entian and Kötter, 2007) |

#### *3.1.1.3 Plants*

Table 3: Overview of used plants

| Plant | Source |
|---|---|
| ***Cannabis sativa* L. var Bediol** | Bedrocan BV, Groningen, Netherlands |
| ***Nicotiana benthamiana*** | Department of Biology, TU Darmstadt (Prof. Dr. Heribert Warzecha) |

The plant material used in this project was provided by Bedrocan BV (Groningen, The Netherlands). This licensed company is producing standardized cannabis plant material under Good Agricultural Practice (GAP) conditions (Fischedick et al., 2010). The plants are supplied to patients with prescription via pharmacies in agreement with the Dutch government (Fischedick et al., 2010; "The Office of Medicinal Cannabis," 2016).

### 3.1.2 Vectors

#### *3.1.2.1 Modification of pGEM-T easy*

Cloning vector pGEM-T easy (Promega GmbH, Mannheim, Germany) was used in the subsequent experiments and provided by the working group of Prof. Frank Schulz (MPI, Dortmund, Germany) (Figure 26). For multiplication and insertion of two additional cloning sites, the vector was amplified via PCR (PCR conditions are described in chapter 3.1.7).

The purification after gel electrophoresis was performed with the PCR and Gel purification Kit according to the manufacturer's protocol (GE Healthcare, Solingen, Germany). Afterwards the created vector was digested with the restriction enzyme XbaI (New Englands Biolabs, Frankfurt am Main, Germany) and afterwards purified again (PCR and Gel purification Kit, GE Healthcare, Solingen, Germany). The cohesive ends were linked with the Fast Link™ DNA Ligation Kit (Biozym Scientific GmbH, Hessisch Oldendorf, Germany) as described in the manufacturer's protocol.

Table 4: Overview of origin and modification of used vectors

| Vectors | Modification | Source |
|---|---|---|
| **pGEM-T easy** | import of one XbaI and two XcmI restriction sites in the multiple cloning site (**Fehler! Verweisquelle konnte nicht gefunden werden.**) | MPI, Dortmund, Germany (Prof. Frank Schulz) |
| **pDionysos1.1** | Originally pYES2 vector; additional Leu2d marker added | Department of Technical Biochemistry, TU Dortmund |

Table 5: Overview of magnICON® provectors (Icon genetics GmbH, Halle/Saale, Germany) provided by Prof. Dr. Heribert Warzecha (Department of Biology, TU Darmstadt) (Marillonnet et al., 2005)

| Vector | Function | Antibiotic resistance |
|---|---|---|
| **pICH11599** | 3'-module; carrying a multiple cloning site | Ampicillin |
| **pICH7410** | 3'-module; carrying GFP | Carbenicillin |
| **pICH15879** | 5'-module; directing the protein of 3'-module to the cytosol | Carbenicillin |
| **pICH17620** | 5'-module; directing the protein of 3'-module to the apoplast | Carbenicillin |
| **pICH20030** | 5'-module; directing the protein of 3'-module to the chloroplast | Carbenicillin |
| **pICH10881** | Recombinase module | Kanamycin |

### 3.1.3 Competent cells

Competent *Escherichia coli* DH5α cells and *Agrobacterium tumefaciens* strains (GV3101 and LBA4404) were produced as described in literature (Inoue et al., 1990). The cells were stored at -80°C.

Competent *Saccharomyces cerevisiae* Y05416 cells were produced as described in literature (Gietz and Schiestl, 2007).

### 3.1.4 Antibiotics

Table 6: Antibiotics and used standard concentrations

| Antiobiotic resistance | Final concentration (µg/ml) |
|---|---|
| **Ampicillin** | 100 |
| **Kanamycin** | 50 |
| **Carbenicillin** | 50 |
| **Rifampicin** | 50 in YEB medium<br>25 in LB medium |
| **Gentamycin** | 50 |
| **Streptomycin** | 50 |

### 3.1.5 Buffers

Table 7: 50X TAE buffer for preparation of agarose gels for electrophoresis

| Component | Concentration |
|---|---|
| **Tris** | 2 M |
| **EDTA (100 ml, pH 8, 5.0 M)** | 50 mM |
| **Acetic acid (100%)** | 52 ml |
| **Water** | 1 L |

Table 8: Lysis buffer (Yazaki et al., 2002)

| Component | Concentration |
|---|---|
| **Tris** | 20mM (pH7.5) |
| **Sorbitol** | 0.6M |
| **Phenylmethylsulfonylfluoride (PMSF)** | 1mM |
| **Dithiothreitol (DTT)** | 10mM |

Table 9: ICON buffer used for infiltration (Marillonnet et al., 2004)

| Component | Concentration |
|---|---|
| **MES (2-[*N*-morpholino]ethanesulfonic acid)** | 10 mM (pH 5,5) |
| **$MgSO_4$** | 10 mM |
| **Acetosyrignon** | 0.1 mM |

### 3.1.6 Media

Table 10: Composition of LB- and SOC-Medium

| Medium | Component | Amount |
|---|---|---|
| **LB** | Yeast extract | 5g |
| | Tryptone | 10g |
| | NaCl | 10g |
| | Add 900 ml $H_20$; adjust to pH 7 with 10M NaOH; adjust to 1L final volume; sterilize by autoclaving and store at room temperature | |
| **LB Agar** | For Agar-plates add 1,5% Agar-Agar before autoclaving | |
| **SOC** | Tryptone | 20g |
| | Yeast extract | 5g |
| | NaCl | 0.5g |
| | KCl (1M) | 2.5ml |
| | In 900ml $H_20$, adjust to pH 7 with 10M NaOH, add $H_2O$ to 970ml; sterilize by autoclaving and store at room temperature; before use add 10ml sterile $MgCl_2$ and 20ml sterile 1M Glucose | |

Table 11: Composition of YEB medium for cultivation of *Agrobacterium tumefaciens*

| Component | Amount (g) |
|---|---|
| **Beef extract** | 5 |
| **Yeast extract** | 1 |
| **Peptone** | 5 |
| **Sucrose** | 5 |
| **$MgCl_2$** | 0.5 |
| **Water** | Ad 1L |

Table 12: Yeast minimal medium (synthetic dropout medium) and yeast complex medium

| Medium | Component | Amount |
|---|---|---|
| **Minimal medium (-Ura, -Leu)** | Yeast nitrogen basis medium (without amino acids) | 6.7g |
| | Amino acid mixture (Table 41) | 1.546g |
| | Glucose | 4% |
| | Agar | 25g |
| | | Ad 1L Water |
| **Complex medium** | Yeast extract | 10g |
| | Peptone | 20g |
| | Glucose | 2% |
| | Agar | 20g |
| | | Ad 1L Water |

### 3.1.7 PCR

Table 13: PCR for amplification of cloning vector pGEM-T easy with Platinum® PCR Supermix High Fidelity (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

| Step | Temperature | Time | Cycles |
|---|---|---|---|
| **Initial denaturation** | 94°C | 2:00 | 1 |
| **Denaturation** | 94°C | 0:30 | 30 |
| **Annealing** | 55°C | 0:30 | |
| **Elongation** | 68°C | 3:15 | |
| **Final extension** | 68°C | 10:00 | 1 |

Table 14: Primers for construction of cloning vector pGEM-T easy

| Primer | Sequence 5'→3' |
|---|---|
| **pGEM-T easy fw** | GCTGGTGCGCCTGCTCTAGAGCATCCGATTCCGATGCAGACCACGATATCACTGGATCACTAGTGAATTCGCGGCCGCCTG |
| **pGEM-T easy rev** | GCTCTATCTAGAATCTCGTTGTACCGATGAACCCATGATATCACTGGATCGAATTCCCGCGGCCGCCAT |
| **M13 fw** | AACGACGGCCAGTGAATTG |
| **M13 rev** | TACGCCAAGCTATTTAGGTGAC |

Table 15: PCR reaction for expression profile of the putative prenyltransferases from *Cannabis sativa* L. spp.; Taq DNA Polymerase Master Mix Red (1.1x with 1.5mM $MgCl_2$; Ampliqon, Odense, Denmark)

| PCR step | Temperature (°C) | Time (min:sec) | Cycles |
|---|---|---|---|
| **Initial denaturation** | 95 | 2:00 | 1 |
| **Denaturation** | 95 | 0:30 | 40 |
| **Annealing** | From 53 to 67 | 0:30 | |
| **Elongation** | 72 | 1:00 | |
| **Final extension** | 72 | 5:00 | 1 |

Experiments concerning the expression profile of the in silico identified PTs was performed with Taq DNA Polymerase Master Mix Red (1.1x with 1.5mM $MgCl_2$; Ampliqon, Odense, Denmark) (Table 15). Volumes were used according to the manufacturer's protocol. The PCR products were analyzed by agarose gel electrophoresis.

Colony PCR was also performed with Taq DNA Polymerase Master Mix Red (1.1x with 1.5mM $MgCl_2$; Ampliqon, Odense, Denmark) (Table 15). The annealing temperature was set to 50°C and the elongation step lasted 2minutes.

Table 16: PCR reaction for PT1, PT2 and PT3 from cDNA; Platinum® PCR Supermix High Fidelity (Thermo Fisher Scientific, Waltham, Massachusetts, USA)

| Step | Temperature | Time | Cycles |
|---|---|---|---|
| **Initial denaturation** | 94°C | 2:00 | 1 |
| **Denaturation** | 94°C | 0:30 | 40 |
| **Annealing** | 55°C | 0:30 | |
| **Elongation** | 68°C | 2:00 | |
| **Final extension** | 68°C | 10:00 | 1 |

The PCR reaction was performed for transient expression in *Nicotiana benthamiana* according to the manufacturer's protocol with Platinum® PCR Supermix High Fidelity (Thermo Fisher Scientific, Waltham, Massachusetts, USA). Temperatures, duration and cycles are displayed in Table 16 and primer sequences in Table 17.

Table 17: Primer pairs for amplification of PT1, PT2 and PT3 from trichome cDNA with and without signal peptides, respectively

| Primer | Sequence 5'→3' |
|---|---|
| **NcoI PT1 with SP fw** | GTATCCATGGGACTCTCATCAGTTTGTA |
| **NcoI PT1 without SP fw** | TGATCCATGGCAGCTACTACAAATCAAACTGAG |
| **PT1 BamHI rev** | GAATGGATCCTTATATGAAAACATATACTAAATATTCAGCATAATAAA |
| **NcoI PT2 with SP fw** | GATTCCATGGAGCTCTCATCAATTTGTAAC |
| **NcoI PT2 without SP fw** | TTGACCATGGCGACAGATACTGCAAATCAAACT |
| **PT2 BamHI rev** | GGTAGGATCCCTAAATGAAAACATATACTAAATATTCAGCATAAT |
| **NcoI PT3 with SP fw** | TGTACCATGGGACTCTCATTAGTTTGTACCT |
| **NcoI PT3 without SP fw** | ACAGCCATGGCAGGTAGCGATCAAATTGAAG |
| **PT3 BamHI rev** | ACTGGGATCCTTATATAAATACATATACAAAGTATTCAGCATAATATAGC |

Table 18: Primer pairs for Q5® High Fidelity PCR (New England Biolabs GmbH, Frankfurt am Main, Germany) with integrated signal peptide from F1F0 ATP Synthase

| Primer | Sequence 5'→3' |
|---|---|
| **PT6 ATpase fw** | ctcactatagggaatattaatggttttgccaagactatatactgctacatcccgtgctgcttttaaagcagccaaaCTCGGTT ACGTCAGCCAC |
| **PT6 ATPase rev** | ccgagctcggtaccaTCAAGACGATATTCGTGCTAAAAC |
| **pDio incl. ATPase SP fw** | TGGTACCGAGCTCGGATC |
| **pDio incl. ATPase SP rev** | TTTGGCTGCTTTAAAAGCAG |
| **pDio incl. ATPase SP for PT3 fw** | tgcttttaaagcagccaaaGCAGGTAGCGATCAAATTG |
| **pDio incl. ATPase SP for PT3 rev** | ccgagctcggtaccaTTATATAAATACATATACAAAGTATTCAGCATAATATAG |

Table 19: PCR reaction with Q5® High fidelity (New England Biolabs GmbH, Frankfurt am Main, Germany)

| Component | Volume |
|---|---|
| **Q5 high fidelity 2XMasterMix** | 12.5 µl |
| **Forward Primer (10 µM)** | 1.25µl |
| **Reverse Primer (10 µM)** | 1.25µl |
| **Template DNA (<1000 ng)** | Dependent on concentration |
| **$H_2O$ (nuclease-free)** | Ad 20µl |

Table 20: PCR reaction conditions with Q5® High Fidelity reaction mixture (New England Biolabs GmbH, Frankfurt am Main, Germany)

| Step | T (°C) | T (min:sec) |
|---|---|---|
| **Initial Denaturation** | 98°C | 0:30 |
| **35X** | 98°C | 0:10 |
| | 56°C | 0:30 |
| | 72°C | 1:30 |
| **Final extension** | 72°C | 2:00 |
| **Hold** | 4°C | ∞ |

Table 18, Table 19 and Table 20 combined the PCR primer sequences, reaction volumes and reaction conditions for heterologous expression in *Saccharomyces cerevisiae* Y05416. PCR products are afterwards used for homologous recombination or Gibson Assembly®.

### 3.1.8 Agarose gel electrophoresis and purification

Analysis of PCR and RNA products were performed by agarose gel electrophoresis.

For DNA analysis 1g Agarose (Biozym Scientific GmbH, Hessisch Oldendorf, Germany) was added to 100ml of 1XTAE buffer (Table 7). Run at 130V for 30 minutes.

For RNA analysis 2 g Agarose (Biozym Scientific GmbH, Hessisch Oldendorf, Germany) was added to 100 ml of 1XTAE buffer (Table 7). The electrophoresis run at 130V for 30 minutes.

PCR products were purified with the PCR and Gel purification Kit (GE Healthcare, Solingen, Germany) as described in the manufacturer's protocol after agarose gel electrophoresis.

## 3.1.9 Cloning and ligation

### *3.1.9.1 TA-cloning*

Table 21: TA-cloning (Sigma-Aldrich, St. Louis, USA) into pGEM-T easy (Promega GmbH, Mannheim, Germany)

| Component | Volume (µl) |
|---|---|
| **Rapid Ligation Buffer (2X)** | 10 µl |
| **ATP (10 mM)** | 4 µl |
| **T4 DNA Ligase (600 U/µl)** | 1 µl |
| **Linearized vector (1-10 ng/µl)** | X µl |
| **Insert DNA (1-10 ng/µl)[1]** | Y µl |
| **Water, nuclease-free** | **Ad 20µl** |

[1]insert DNA was added in excess

The ligation mixture was incubated at 4°C overnight. Afterwards a transformation into highly competent *Escherichia coli* DH5α cells by heat-shock was performed (chapter 3.1.9.3).

### *3.1.9.2 Ligation for transient expression*

After TA-cloning of PTs into pGEM-T easy, colony PCR was performed (Primers Table 17, PCR conditions Table 16). Positive clones were cultivated in 5ml LB medium with Ampicillin. Plasmid isolation was performed with peqGOLD Plasmid Miniprep Kit (peqlab Biotechnologie GmbH, Erlangen, Germany) as described in the manufacturer's protocol. The concentration was determined by Qubit (Invitrogen Life Technologies, Darmstadt, Germany). Plasmids were sent for sequencing with the according sequencing primers (Table 14, M13 fw and M13 rev). Plasmids of sequentially correct clones were restricted with the endonucleases NcoI and BamHI (New England Biolabs GmbH, Frankfurt am Main, Germany). The reaction mixture is shown exemplary in Table 22 for PT1. The mixture was incubated for 1hour at 37°C. The insert was purified by Gel purification (PCR and Gel purification Kit, GE Healthcare, Solingen, Germany).

Table 22: Enzymatic restriction reaction for pGEM-T easy vector (Promega GmbH, Mannheim, Germany) with NcoI and BamHI Q5

(New England Biolabs GmbH, Frankfurt am Main, Germany)

| Component | Volume |
|---|---|
| **NE 2.1 Buffer (5X)** | 5 µl |
| **NcoI** | 1 µl |
| **BamHI** | 1 µl |
| **DNA (1 µg)** | X µl |
| **$H_20$** | **Ad 50 µl** |

The ligation into pICH11599 was performed with the Fast-Link™ DNA Ligation Kit (Biozym Scientific GmbH, Hessisch Oldendorf, Germany) and a representative reaction mixture is shown in Table 23. Depending on the intensities of the bands after agarose gel electrophoresis, the ratio was set to 1:8 in this reaction. The reaction volumes are shown exemplary in Table 22 for PT1. The mixture was incubated for 1hour at 37°C. The insert was purified by Gel purification.

Table 23: Ligation reaction into pICH11599 (Icon genetics GmbH, Halle/Saale, Germany); Fast-Link™ DNA Ligation Kit (Biozym Scientific GmbH, Hessisch Oldendorf, Germany)

| Component | Volume |
|---|---|
| **Fast-Link Ligation Buffer (10X)** | 1,5 µl |
| **ATP (10 mM)** | 1,5 µl |
| **T4 DNA Ligase (2 U/µl)** | 1 µl |
| **Linearized vector[1]** | X µl |
| **Insert DNA** | Y µl |
| **$H_2O$, sterile** | Ad 15µl |
| **Reaction conditions** | **5 minutes incubation at room temperature; termination: 15 minutes at 70°C** |

[1]molar ratio 1:2 (vector:insert)

### *3.1.9.3 Transformation via Heat-shock* (Inoue et al., 1990)

Competent cells, which were stored at -80°C, were thawed on ice, 2µl of the ligation mix were added and incubated on ice for 30 minutes. The mixture was heated to 42°C for 45 seconds and placed on ice for 2 minutes. 500µl of SOC medium were added and 300µl were plated on LB-Agar petri dishes with Ampicillin (Table 6, Table 10). Dishes were incubated at 37°C overnight. Developing clones were picked for overnight cultures in liquid LB-medium with Ampicillin at 37°C. Plasmids were isolated with the peqGOLD Plasmid Miniprep Kit (peqlab Biotechnologie GmbH, Erlangen, Germany) as described in the protocol. The clones were sent for sequencing (Table 14).

### *3.1.9.4 Gibson Assembly®*

Primers and PCR used for PT6 and PT3 for Gibson Assembly® (New England Biolabs GmbH, Frankfurt am Main, Germany) are shown in Table 18, Table 19 and Table 20 (Gibson, 2011; Gibson et al., 2009). The Gibson Assembly® was performed as described below in Table 24. The mixture was incubated for 60 minutes at 50°C. After 15, 30 and 60 minutes 3µl of the reaction were withdrawn and a transformation into *E. coli* DH5α was performed as described above (heat-shock). Success of the procedure was evaluated via colony PCR and agarose gel electrophoresis (Table 15). The positive clones after gel electrophoresis were sent for sequencing.

Table 24: Gibson Assembly® reaction mixture for ligation into expression vector pDionysos1.1

| **Component** | **Volume** |
|---|---|
| **Gibson Assembly Master Mix (2X)** | 10 µl |
| **pDionysos1.1 (50-100 ng)** | X µl |
| **PCR product (3 fold excess)** | Y µl |
| **Water** | Ad 20 µl |

### *3.1.9.5 Homologous recombination*

Primers used for PCR with Taq DNA Master Mix Red (1.1x with 1.5mM $MgCl_2$; Ampliqon, Odense, Denmark) and the Platinum® PCR Supermix High Fidelity (Thermo Fisher Scientific, Waltham, Massachusetts, USA) are displayed in Table 25 (Table 15, Table 16). The annealing temperatures were set to 68°C for Coq2, 61°C for PT3, 62°C for PT6 with the native signal peptides and 72°C for PT3 with the F1F0-signal peptide. The PCR products were purified after agarose gel electrophoresis with the PCR and Gel purification kit. The homologous recombination was performed as described in literature (Gietz and Schiestl, 2007). The incubation at 42°C was carried out for 40 minutes. 100µl were plated on the SD medium with uracil deficiency (Table 12 and Table 41). The plates were incubated at 30°C. Developed clones were cultivated overnight in 5ml of uracil deficient SD medium. Plasmid isolation was performed with the E.Z.N.A.® Yeast Plasmid Mini Kit (Omega Bio-Tek, Norcross, USA) as described in the manufacturer's protocol. Subsequently the plasmids were transformed into *E. coli* DH5α cells (described above; heat-shock). A colony PCR and an agarose gel electrophoresis were performed. Positive clones after the gel electrophoresis were sent for sequencing.

Table 25: Primer sequences for homologous recombination of Coq2, PT3, PT6 with their corresponding native signal peptides; and primer sequence for PT3 with the signal peptide of the F1F0 enzyme complex

| Primer | Sequence 5'→3' |
|---|---|
| **Coq2 fw** | CTGTAATACGACTCACTATAGGGAATATTAAAAAAAAATGGCTTTTATTTGGCAGAGAAAGAGTATTTT |
| **Coq2 rev** | TACTAGTGGATCCGAGCTCGGTACCAAGCTCTACAAGAATCCAAACAGTCTCAAGAT |
| **PT3 with SP fw** | CTGTAATACGACTCACTATAGGGAATATTAAAAAAAAATGGCTGGACTCTCATTAGTTTGTACCTTTTC |
| **PT3 with SP rev** | TACTAGTGGATCCGAGCTCGGTACCAAGCTTTATATAAATACATATACAAAGTATTCAGCATAATATAGC |
| **PT6 with SP fw** | CTGTAATACGACTCACTATAGGGAATATTACAAAAAAATGGCTGCTATTGCGTTGTG |
| **PT6 with SP rev** | TACTAGTGGATCCGAGCTCGGTACCAAGCTTCAAGACGATATTCGTGCTAAA |
| **PT3 ATpase fw** | ctcactatagggaatattaatggttttgccaagactatatactgctacatcccgtgctgcttttaaagcagccaaaGCAGGTAGCGATCAAATTG |

### 3.1.10 Ethanol precipitation

Sodium acetate solution (10% of the sample volume, 3M, pH 5.2) was added to a DNA sample and mixed. Afterwards 2.5 sample volumes of cold 100% ethanol were added to the mixture and again mixed well. After an incubation on ice for 20 minutes the solution was centrifuged at 4°C at maximum speed for 15 minutes. The supernatant was discarded. 1ml of cold 70% ethanol was added and vortexed, followed by another centrifugation at 4°C for 10 minutes at maximum speed. The supernatant was again discarded. The pellet was air-dried and subsequently resuspended in 11µl of water.

## 3.2 Construction of cDNA library from trichomes of *Cannabis sativa* L. var. Bediol

The isolation of trichomes was performed as described in literature (Yerger *et al*., 1992). The subsequent RNA isolation was performed with the RNeasy Plant Mini Kit from Qiagen (Qiagen GmbH, Hilden, Germany) according to the manufacturer's protocol. The concentration of the RNA sample was measured (82ng/µl) and sent to LGC Genomics (LGC Genomics GmbH, Berlin, Germany) to set up a full-length optimized and normalized cDNA library. The complete RAW data was delivered as CD-ROM.

## 3.3 Bioinformatic analysis of DNA libraries of *Cannabis sativa* L. spp.

The bioinformatic analysis process was based on four libraries: (i) the normalized cDNA library from *Cannabis sativa* L. var. Bediol described above, (ii) a whole genome shotgun from *Cannabis sativa* that was freely available from "medicinal genomics" (Medicinal Genomics Corporation, 2015) and two freely available libraries of the varieties (iii) Purple Kush and (iv) Finola (van Bakel et al., 2011) (Table 26).

Table 26: Datasets used for bioinformatic analysis; CSA: cDNA library from *Cannabis sativa* L. var. Bediol, PK: library from *Cannabis sativa* L. var. Purple Kush, Finola: library from *Cannabis sativa* L. Finola, MPGR: whole genome shotgun from *Cannabis sativa* L.

| Assembly | Isogroups | Isotigs | Contigs |
|---|---|---|---|
| **CSA** | 11300 | 12892 | 14695 |
| **PK** | 40224 | 40224 | 40224 |
| **Finola** | 47671 | 47671 | 47671 |
| **MPGR** | 69556 | 69556 | 69556 |

18 prenyltransferases from different host plants were selected to serve as templates in the search for new putative PTs but also to get an overview of the structural properties of this class of enzymes. The selected proteins are shown in Table 27 and their amino acid sequences are listed in the supplementary information (Table 42). The protein named PT1 is an amino acid sequence which was used in the lab before. This sequence is very similar to the patent-registered prenyltransferase from *Cannabis sativa* L. (PT1p), but lacks one amino acid and carries some additional point mutations (Figure 46).

Table 27: Prenyltransferases from different host plants for sequential analysis

| Number | Enzyme | Host plant |
|---|---|---|
| **1** | Protoheme IX farnesyltransferase (Cox10) | *Arabidopsis thaliana* |
| **2** | Prenyltransferase | *Arabidopsis lyrata* |
| **3** | Protein prenyltransferase | *Arabidopsis lyrata* |
| **4** | cis-prenyltransferase | *Arabidopsis thaliana* |
| **5** | Polyprenyltransferase | *Arabidopsis thaliana* |
| **6** | PT1 | *Cannabis sativa* L. |
| **7** | PT1p | *Cannabis sativa* L. |
| **8** | Aromatic prenyltransferase | *Epimedium acuminatum* |
| **9** | Flavonoids prenyltransferase | *Sophora flavescens* |
| **10** | *cis*-prenyltransferase | *Hevea brasiliensis* |
| **11** | Aromatic prenyltransferase | *Humulus lupulus* |
| **12** | *cis*-prenyltransferase | *Lilium longiflorum* |
| **13** | Protein prenyltransferase (Os11g0483950) | *Oryza sativa* |
| **14** | Prenyltransferase | *Populus trichocarpa* |
| **15** | Prenyltransferase | *Sophora flavescens* |
| **16** | Protoheme IX farnesyltransferase | *Ricinus communis* |
| **17** | Flavonoid prenyltransferase | *Sophora flavescens* |
| **18** | Prenyltransferase | *Zea mays* |

The analysis tool for transmembrane hidden markov model (TMHMM) was used for the prediction of transmembrane helices (Eddy, 1998, 2004; Gromiha et al., 2009; Liu et al., 2002; Pedersen and Hein, 2003). TMHMM additionally predicts the starting and ending point of the transmembrane helix in the amino acid sequence.

SignalP and ChloroP were used for the prediction of signal peptides for different locations in the plant cell and signal sequences for the chloroplasts within the amino acid sequence of the corresponding protein, respectively (Emanuelsson et al., 2007, 1999).

For the identification of putative prenyltransferases in the *C. sativa* transcripts Newbler versin 2.6 (Margulies et al., 2005), InterProScan (Zdobnov and Apweiler, 2001), BLAST (Altschul et al., 1997) and mafft (Katoh et al., 2002) were used.

### 3.3.1 Phylogenetic analysis

The resulting DNA sequences were analyzed by using the Molecular Evolutionary Genetics Analysis (MEGA) software (Tamura et al., 2011). It was used for building phylogenetic trees and sequence alignments to get information about the evolutionary relation between the putative enzymes and already known PTs. With the software MEGA it was possible to analyze the maximum likelihood of the trees, which is comparable with other software packages with these properties in consideration of efficiency and accuracy (Tamura et al., 2011). MEGA software is freely available.

## 3.4 Expression profile of *in silico* identified prenyltransferases from *Cannabis sativa* L. spp.

*Cannabis sativa* L. var. Bediol total RNA was isolated from stem, branches, leaves, flowers and trichomes. The plant material was collected and ground in liquid nitrogen. Afterwards the RNA was isolated as described above with the RNeasy Plant Mini Kit from Qiagen (Qiagen GmbH, Hilden, Germany). The RNA isolation of the leave tissue was performed as described by Vicient and Delseny (Vicient and Delseny, 1999).

The cDNA synthesis was carried out with the MMLV Reverse Transcriptase 1$^{st}$-strand cDNA synthesis Kit (Biozym Scientific GmbH, Hessisch Oldendorf, Germany). The composition of the synthesis reaction is shown in Table 28 and described in the manufacturer's protocol.

Table 28: cDNA synthesis reaction**;** MMLV Reverse Transcriptase 1$^{st}$-strand cDNA synthesis Kit (Biozym Scientific GmbH, Hessisch Oldendorf, Germany)

| **Solution** | **Volume used in the reaction (µl)** |
|---|---|
| **Total RNA sample** | X (up to 1µg) |
| **Oligo(dT)$_{21}$ Primer** | 2 |
| **MML RT 10X reaction buffer** | 2 |
| **100mM DTT** | 2 |
| **dNTP PreMix** | 2 |
| **RiboGuard RNase Inhibitor** | 0.5 |
| **MMLV Transcriptase** | 1 |
| **RNase-free water** | **Ad 20 µl** |

The reaction conditions of the PCR for amplification of prenyltransferases from cDNA are described in chapter 3.1.7. The primers corresponding to the respective enzymes are shown in the

supplementary information (Table 37). The enzyme glyceraldehyde 3-phosphate dehydrogenase (GAPDH) served as positive control (Table 39). The PCR products were analyzed by agarose gel electrophoresis. The annealing temperatures were set to 55°C for PT1, 67°C for PT2 and PT3, 56°C for PT4, 58°C for PT5 and gapdh, 53°C for PT6.

## 3.5 Transient expression in *Nicotiana benthamiana*

PT1, PT2 and PT3 were amplified from trichome cDNA of *Cannabis sativa* L. var. Bediol (week 8). PCR, cloning and ligation conditions are described in chapter 3.1.7, 3.1.9.

### 3.5.1 Cultivation

The ICON vectors carrying the putative prenyltransferases were transformed into *Agrobacterium tumefaciens* GV3101 and LBA4404 cells (chapter 3.1.1.1). The transformation was performed as described above by heat shock (chapter 3.1.9.3).

After two days of incubation at 30°C colonies of transformed *A. tumefaciens* GV3101 cells were growing on the plates. Colony PCR was performed as describe above for *E. coli* DH5α cells (Table 15). The PCR products were analyzed by agarose gel electrophoresis.

### 3.5.2 Vacuum infiltration

For the transfection of the constructs into young *Nicotiana benthamiana* plants (six weeks old), three main cultures were set up. For each transfection *A. tumefaciens* cultures carrying a 3'-module, a 5'-module and a recombinase module were cultivated independently and afterwards mixed for the bacterial suspension used in the transfection.

A preculture of *Agrobacterium tumefaciens* GV3101 cells carrying the appropiate ICON vector were set up overnight in 4ml YEB medium (Table 11) with corresponding antibiotics (Table 6).

PT1+SP stands for the patent-registered prenyltransferase with its natural signal peptide (Page and Boubakir, 2011). PT2+SP and PT3+SP stand for the *in silico* detected putative prenyltransferases PT2 respectively PT3 with their native signal peptide. PT2-SP stands for the putative prenyltransferase PT2 without its native signal peptide. The lengths of the signal peptides were determined by the software tool ChloroP (Emanuelsson et al., 1999).

Table 29: magnICON® module combination used in the infection of *Nicotiana benthamiana*; +SP: prenyltransferase with its native signal peptide; -SP: prenyltransferase without signal peptide

| Experiment | 5' Module | 3' Module | Recombinase Module |
|---|---|---|---|
| **PT1+SP** | pICH15879 (cytosol) | pICH11599_PT1+SP | pICH10881 |
| **PT2+SP** | pICH15879 (cytosol) | pICH11599_PT2+SP | pICH10881 |
| **PT3+SP** | pICH15879 (cytosol) | pICH11599_PT3+SP | pICH10881 |
| **PT2-SP** | pICH15879 (cytosol) | pICH11599_PT2-SP | pICH10881 |
| **PT2-SP** | pICH17620 (apoplast) | pICH11599_PT2-SP | pICH10881 |
| **PT2-SP** | pICH20030 (chloroplast) | pICH11599_PT2-SP | pICH10881 |
| **pICH7410 GFP (control)** | pICH15879 (cytosol) | pICH7410 | pICH10881 |

The 4ml-overnight culture was incubated for 16-20 hours at 28°C. 0.5ml of these cultures were added to the corresponding main culture (100ml YEB with the according antibiotics depending on the used vectors). The main culture was incubated for another 16-20 hours at 28°C. Afterwards the cultures were transferred to 50ml Falcon tubes and centrifuged at 5000xg for 10 minutes. The supernatant was discarded. The bacterial pellet was resuspended in 1L of ICON buffer (Table 9). The bacterial suspension was transferred into a beaker. The plant was turned upside-down into the beaker and an airtight dome was put over the whole setup. Vacuum to -0.7 bar was generated inside the dome by a pump and released.

### 3.5.3 Activity assay

Parts of the tobacco leaves (about 220 mg) were transferred into a glass mortar tube. 500 µl of Tris-Acetate (50 mM, pH7.5) were added and the mixture was homogenized with a glass pestle. 100 µl of the plant mix was transferred to a HPLC vial. 1 µl of olivetolic acid solution (100 mM in DMSO) and 10 µl of GPP solution (10 mM dissolved in Tris-Acetate buffer, 50 mM, pH7.5) were added. The whole mixture was shaken at 200 rpm and incubated at 30 °C for 2 hours in the dark. The reaction was terminated by adding 200 µl of acetonitrile with 5% (v/v) of formic acid. The reaction tube was placed on ice for 30 minutes for protein precipitation. Afterwards the mixture was centrifuged at 16,000 xg for 30 minutes at room temperature. 150 µl of the supernatant were transferred into another HPLC vial and analyzed by HPLC-ESI-MS (Table 30) (Fellermeier and Zenk, 1998).

## 3.6 Heterologous expression in *Saccharomyces cerevisiae* Y05416

PT3, PT6 and Coq2 were cloned and ligated as described above (3.1.7, 3.1.8, 3.1.9). PT3 and PT6 were expressed heterologously in *S. cerevisiae* Y05416 (Table 2). pDionysos1.1 served as control: (i) empty (ii) with Coq2.

PT3 and PT6 were cloned with their native signal peptide as well as with an extrinsic signal peptide derived from the enzyme complex F1F0 ATP synthase from yeast (Takeda et al., 1985).

### 3.6.1 Cultivation and activity assay

Precultures were set up on SD medium lacking leucine and uracil (Table 12 and Table 41). The main cultures (complex medium with 4% galactose, Table 12) were inoculated to $OD_{600}$ 0.5, incubated at 30 °C and shaken at 200 rpm. Samples were taken after 12, 24, 48, 72, 96 and 120 hours. The samples were centrifuged (10,000 x g, 4°C, 5 minutes) and the supernatant discarded (Yazaki et al., 2002). Each sample pellet was resuspended in 1.5 ml of lysis buffer (Table 8). The reaction tubes were filled with glass beads (average diameter of 0.75-1mm) and vortexed 5 to 8 times for 3 minutes at 4°C. The number of repeats was dependent on the disruption rate of the cells, which was observed and estimated under the microscope. When 60-80% of the cells were disrupted, the cell suspension was centrifuged (10,000 x g, 4°C, 5 minutes). The supernatant was transferred to an ultracentrifugation tube and balanced with 5 ml Tris buffer (100mM, pH 7.5). The solution was centrifuged at 100,000 x g at 4°C for one hour (Yazaki et al., 2002). The supernatant was discarded for the enzyme assay and the pellet resuspended in 200µl of Tris buffer (100mM, pH7.5, Table 8) (Yazaki et al., 2002).

The activity assays for Coq2 and PT6 were performed similarly to the description of Heide *et al*. (Heide *et al*., 1989). $MgCl_2$ (10mM), PHB (1mM) and GPP (5mM) were added to 80µl of the resuspended enzyme solution and the mixture was incubated for one hour at 30°C. The reaction was terminated by adding twice the volume (200µl) of acetonitrile. To complete the precipitation of proteins, the solutions were incubated at 4°C overnight. Afterwards the mixture was centrifuged (16,000 x g, room temperature, 15 minutes). 100µl of the supernatant were transferred into an HPLC vial for analysis by HPLC-ESI-MS (Table 31 and Table 32).

The activity assays for the constructs of PT3 were performed according to Fellermeier *et al*. and Page *et al*. (Fellermeier and Zenk, 1998; Page and Boubakir, 2011). The substrates used were olivetolic acid (1 mM) and GPP (5 mM). Furthermore the whole reaction was incubated for 2 hours and shaken at 200 rpm as described above for the plant extracts.

## 3.7 Analytical methods

After extraction of the tobacco leaves and the enzyme assay with OA and GPP, the samples were analyzed with an isocratic HPLC method with a coupled mass spectrometer. The method is shown in Table 30.

Table 30: HPLC-MS method and conditions for CBGA detection; positive mode; isocratic method with 80% solvent A and 20% solvent B.

| **Device** | **Property** |
|---|---|
| **Column** | Poroshell 120 (2.7 µm), C18 |
| **Solvent A** | Acetonitrile |
| **Solvent B** | $H_2O$ + 0.1% formic acid |
| **Flow rate** | 0.7 ml/min |
| **Gradient** | isocratic |
| **Ionization** | Electrospray ionization |
| **Mass range** | 90 – 700 m/z |

To detect the products after the enzyme assays from heterologous expression in Saccharomyces cerevisiae Y05416, the porcessed samples were analyzed by HPLC-ESI-TOF-MS (Table 31 and Table 32).

Table 31: Gradient HPLC method for detection of CBGA and GHB; flow rate: 0.8ml/min; maximum pressure 380bar

| **Time (min)** | **$H_2O$+0.1% formic acid (%)** | **Acetonitrile (%)** |
|---|---|---|
| **0.00** | 90 | 10 |
| **2.50** | 20 | 80 |
| **3.50** | 20 | 80 |
| **4.00** | 90 | 10 |
| **8.00** | 90 | 10 |

Table 32: Gradient HPLC-ESI-TOF-MS method for detection of CBGA and GHB; used in positive mode; solvent A 20% and solvent B 80%

| Device | Property |
|---|---|
| **Column** | Nucleosil 100mm, C18 |
| **Solvent A** | $H_2O$ + 0.1% formic acid |
| **Solvent B** | Acetonitrile |
| **Flow rate** | 0.8 ml/min |
| **Gradient** | gradient |
| **Ionization** | Electrospray ionization |
| **Mass range** | 90-700 m/z |

# 4. Results and discussion

The aim of this thesis was the identification of new aromatic prenyltransferases in the genome of *Cannabis sativa* L.. This was necessary because the prenyltransferase proposed by Page and Boubakir involved in the biosynthesis of $\Delta^9$-THC could not be expressed in yeast cells actively. For identification of putative prenyltransferases an *in silico* analysis was performed. Based on phylogenetic analysis and the expression profile of the proteins identified *in silico* the *in vivo* studies comprising the transient expression in *Nicotiana benthamiana* and the heterologous expression in *Saccharomyces cerevisiae* Y05416 were performed.

## 4.1 Bioinformatical analysis of DNA libraries of *Cannabis sativa* L. ssp.

The goal of the bioinformatic analysis of DNA libraries of Cannabis plants was to identify new genetic sequences of aromatic prenyltransferases which could serve as alternatives to the geranylpyrophosphate:olivetolate geranyltransferase (GOT) and the patent-registered sequence in the step of prenylation of olivetolic acid (Fellermeier and Zenk, 1998; Page and Boubakir, 2011). In long term-view alternatives to this PT could serve as proteins to produce medicinal relevant products.

### 4.1.1 Construction and characterization of a cDNA library of *Cannabis sativa* L. Bediol

It is hypothesized that the trichomes of *C. sativa* are the location of both the production and the storage of plant cannabinoids (Dai et al., 2010; Tahara and Ibrahim, 1995; Tsurumaru et al., 2012; Wagner et al., 2004). Prenyltransferases taking part in the synthesis of plant cannabinoids should therefore be present in the transcriptome of the trichomes. Therefore, a cDNA library of *Cannabis sativa* L. Bediol trichomes was constructed. (chapter 3.2). The cDNA library consisted of 11300 isogroups, 12892 isotigs and 14695 contigs.

For an analysis of the generated library the software tools GenDBE and SAMS were used. The cDNA sequences were clustered into different metabolic pathways. The classification of DNA sequences into different functional classes is based on the comparison and approximation to genes with known function (Somerville and Somerville, 1999). This sequential similarity, is assumed to lead to functional similarity (Arabidopsis Genome Initiative, 2000; Somerville and Somerville, 1999). This method of comparison does not result in specific function of the proteins, but gives a hint in which metabolic pathways different genes are involved (Somerville and Somerville, 1999). Further experimental data has to be generated to specify the protein function (Somerville and Somerville, 1999). The pie chart of the classification calculated by GenDBE and SAMS is illustrated in Figure 6.

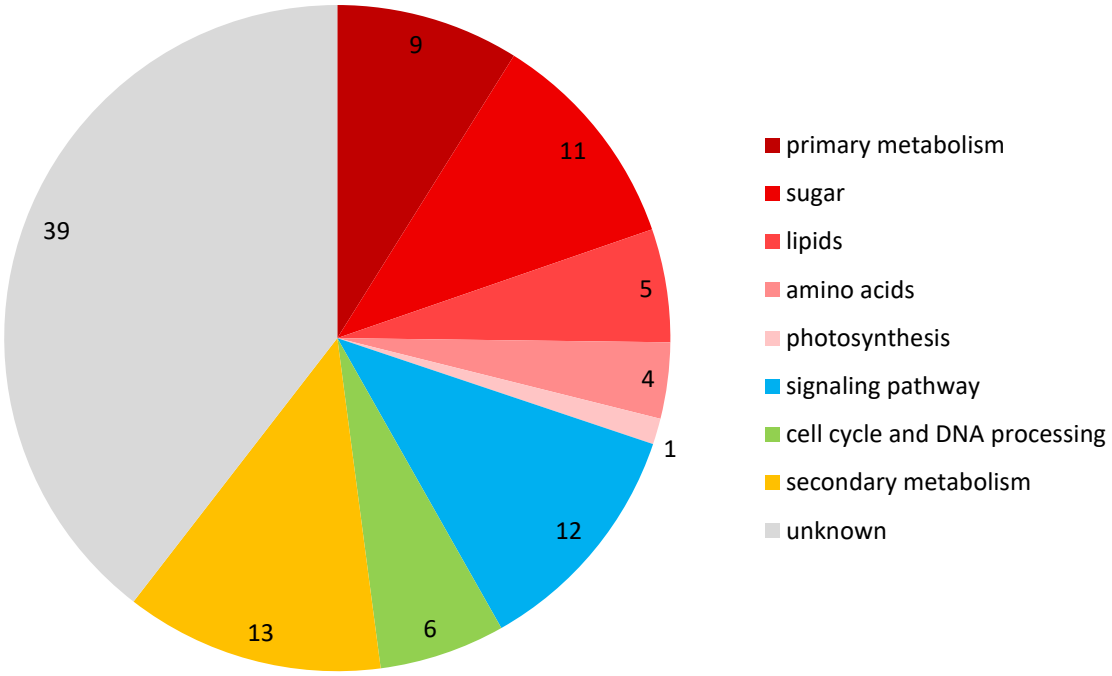



Figure 6: Categorization of cDNA sequences from the *Cannabis sativa* L. var. Bediol trichome cDNA library into different metabolic pathways (in %)

To evaluate the categorization of the cDNA trichome library of *Cannabis sativa* L. Bediol the data was compared to other gene classification of plants. In *Arabidopsis thaliana* the proportion of uncharacterized proteins is about 30% (Arabidopsis Genome Initiative, 2000; Mayer et al., 1999). Within the trichome cDNA library this proportion is about 40%. In the publication of Goff *et al.* the unknown proteins of *Arabidopsis thaliana* were compared to those of *Oryza sativa* as well as other organisms. The comparison to rice showed overlaps in unknown proteins and therefore it was suggested that these proteins are deriving from plant-specific genes (Goff et al., 2002).

The main categories of cDNA sequences from Cannabis are the primary and secondary metabolism as well as the signaling pathways and the synthesis of sugars (Figure 6). Another explanation of the high percentage involved in the signaling pathways is the direction of proteins to certain destinations within higher eukaryotes to operate in specific locations. The Cannabis trichomes are known locations for the synthesis and storage of secondary metabolites acting for example as defense compounds (Dai et al., 2010; Wagner et al., 2004).

### 4.1.2 Data base for Prenyltransferase identification

To provide a wider data base for identification of putative prenyltransferases further sequences from various varieties of *C. sativa* were incorporated in the search:

- CSA: normalized cDNA library from trichomes of *Cannabis sativa* L. var. Bediol (description chapter 3.2); established for PT identification
- PK: transcriptome of *Cannabis sativa* L. Purple kush; reportedly "indica" genetic background (van Bakel et al., 2011); explore cannabinoid biosynthesis in more detail
- FN: transcriptome of *Cannabis sativa* L. var. Finola (van Bakel et al., 2011); established for comparison with transcriptome of purple kush; explore cannabinoid biosynthesis in more detail
- MPGRT: whole genome sequencing (Medicinal Genomics Corporation, 2015); established for optimizing breeding concerning productivity and medicinal aspects

The genetic sequences were uploaded to the platforms SAMS ("The Sequence Analysis and Management System") and GenDBE.

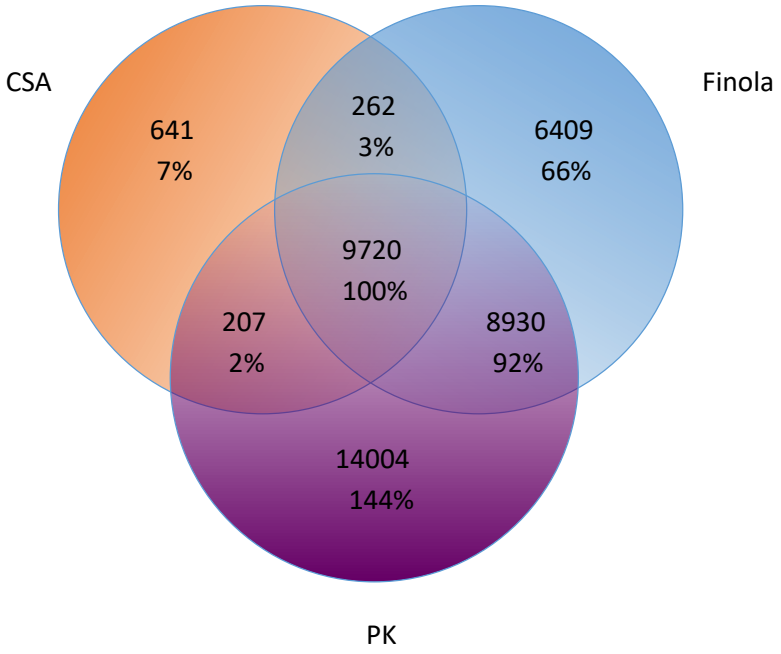



Figure 7: Relations in isotigs of three transcriptome libraries: *Cannabis sativa* L. var. Finola (Finola), var. Purple Kush (PK) and var. Bediol (CSA). The percentages are normalized to the isotigs that are present in all three libraries.

The Venn diagram (Figure 7) calculated by SAMS ("The Sequence Analysis and Management System") displays clusters of three datasets from genetic sequences of Cannabis. The varieties are *Cannabis sativa* L. var. Bediol (orange), *Cannabis sativa* L. var. Finola (blue) and *Cannabis sativa* L. var. Purple

Kush (purple). The clusters combine sequences of the data sets if the alignment of both sets give a coverage of 50% of the shorter gene sequence. In total there are 69252 sequences, 10910 from variety Bediol, 25401 from Finola, 32941 from purple kush. The low number of gene sequences in the trichome cDNA library reflects the usage of solely trichomes to extract total RNA. The ratio of isotigs of CSA also contained in the other two libraries is 94 %, showing that the trichome transcriptome indeed represents a part of the total plant transcriptome. The other libraries contain also gene sequences from leaves, stems and flowers. The gene sequences in this cluster represent the trichome-specific genes for all three plant varieties. The number of sequences is interesting in accordance to their percental proportion of all sequences in the data set. The 9729 cluster-genes represent 89% of the trichome-cDNA, about 30% of the purple kush and about 40% of the Finola library.

The trichome-specific genes in all three Cannabis varieties appear to be highly conserved. This is shown by the small proportions of trichome-specific genes not shared by all three varieties. The differences in trichome-specific genes between the varieties could be caused by the difference in genetic strain background. Purple Kush and Bediol are both bred for high cannabinoid content, even though Purple Kush possesses an indica heritage. Finola in contrast was bred for fibre production.

Another explanation for differences in transcriptome could be different or no normalization strategies during the creation of the cDNA libraries.

The cluster resulting from overlaps of Finola and Purple Kush but not CSA are most likely gene sequences from leaves, stems and flowers of the plant. The overlapping regions between Purple Kush and Finola represent about 57 % and 74 %, respectively. This indicates that both varieties share a common gene pool, but show a relatively large proportion of independent transcripts considering they are both of the variety *Cannabis sativa*. 42% of the PK transcripts are exclusive for the library of Purple Kush. Likewise 25% of the Finola transcripts are exclusive for the Finola library. These exclusive genes could represent the difference between plants producing high amounts of THC and those producing low amounts of the psychoactive ingredient. A further possible explanation is that Purple Kush is a hypothesized sativa-indica hybrid (van Bakel et al., 2011).

A first comparison of the PT1 sequence with the generated trichome transcriptome library CSA was performed using the software tools SAMS and GenDBE. Figure 8 shows a screenshot of a section of the analysis of the sequence of PT1. The results show, that there are some gene sequences with similarity to PT1 in the data sets uploaded to the platforms.

Further investigations of prenyltransferases were performed using SAMS and GenDBE as well as other freely available web-based tools such as TMHMM, SignalP and ChloroP. For that the sequence

of PT1 was used. PT1 is sequentially very similar to the patent-registered aromatic prenyltransferase from *Cannabis sativa* L. (Page and Boubakir, 2011). Three amino acids are exchanged against others and one additional amino acids exists in PT1 comparing to PT1p (Table 33). Since the sequence of PT1 was previously used in the department, this was the sequence of choice for the upcoming experiments.

Table 33: Differences in protein sequence between PT1 and PT1p

| **Position of amino acid** | **PT1p** | **PT1** |
|---|---|---|
| **Position 32** | - | Cysteine |
| **Position 149** | Valine | Isoleucine |
| **Position 223** | Threonine | Isoleucine |
| **Position 256** | Histidine | Tyrosine |

| Name | Group | Assembly | Length | Annotations | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Name | GeneProduct | ECNumber | COG | GO | Confidence | Region |
| CSAI_isotig00533 | CSAG_isogroup00122 | CSA | 774 | - | K12501 homogentisate solanesyltransferase | - | X | 4659,16021 | Metanor 16.03.2012 11:46:12 **Hypothetical protein** | cDNA Importe 12.03.2012 13 |
| CSAI_isotig00534 | CSAG_isogroup00122 | CSA | 761 | - | K12501 homogentisate solanesyltransferase | - | X | 4659,16021 | Metanor 16.03.2012 11:46:12 **Hypothetical protein** | cDNA Importe 12.03.2012 13 |
| CSAI_isotig00535 | CSAG_isogroup00122 | CSA | 513 | - | hypothetical protein | - | X | 4659,16021 | Metanor 16.03.2012 11:46:12 **Hypothetical protein** | cDNA Importe 12.03.2012 13 |
| CSAI_isotig00536 | CSAG_isogroup00122 | CSA | 500 | - | hypothetical protein | - | X | 4659,16021 | Metanor 16.03.2012 11:46:12 **Hypothetical protein** | cDNA Importe 12.03.2012 13 |
| CSAI_isotig04272 | CSAG_isogroup02683 | CSA | 1454 | G4 | K04040 chlorophyll synthase | 2.5.1.62 | X | 16021,43787 | Metanor 16.03.2012 12:28:10 **High confidence in function and specificity** | cDNA Importe 12.03.2012 13 |
| CSAI_isotig04415 | CSAG_isogroup02826 | CSA | 1431 | CBGA Synthase | CBGA | - | X | 4659,16021 | Oliver_Kayser 13.04.2012 18:33:21 **Hypothetical protein** | cDNA Importe 12.03.2012 14 |
| CSAI_isotig04471 | CSAG_isogroup02882 | CSA | 1418 | - | K02257 protoheme IX farnesyltransferase | 2.5.1.- | H | 8496,16021,48034 | Metanor 16.03.2012 12:22:48 **High confidence in function and specificity** | cDNA Importe 12.03.2012 13 |
| CSAI_isotig05491 | CSAG_isogroup03902 | CSA | 1265 | HPT1 | hypothetical protein | - | X | 4659,16021 | Metanor 16.03.2012 12:15:59 **Hypothetical protein** | cDNA Importe 12.03.2012 13 |
| CSAI_isotig06979 | CSAG_isogroup05390 | CSA | 1100 | HST1 | K12501 homogentisate solanesyltransferase | - | X | 4659,16021 | Metanor 16.03.2012 12:35:30 **Hypothetical protein** | cDNA Importe 12.03.2012 14 |

Figure 8: Analysis of PT1 in "The Sequence Analysis and Management System (SAMS)" The first two columns represent the name of the respective isotig and the isogroup. In the following column the dataset where the isotig can be found is mentioned ("assembly"). The next columns show the length of the isotig and the automatic annotations, which were executed based on the comparison of the cDNA sequence to known sequence, e.g. the gene product, the EC number, the gene ontology, confidence calculations.

GenDBE and SAMS represent helpful software tools due to their broad capabilities in sequence analysis. Together a complete and mostly automated processing of raw genetic data into a quality-controlled data base with identified, annotated and clustered genes with linked functions (E.C. number, catalyzed reaction) is possible (Bekel et al., 2009; "https://www.uni-giessen.de/fbz/fb08/Inst/bioinformatik/software/gendb," 2016; Meyer et al., 2003). SAMS and

GenDBE help researchers getting first insights into the data sets and the composition of gene sequences.

Both platforms are auxiliary in comparing sequences and annotating. Since the final goal of the bioinformatic part of this thesis was to find alternatives to a specific aromatic prenyltransferase and express these in microorganisms, it was necessary to moreover use a taylored computational method to find new putative enzymes.

### 4.1.3 Identification of prenyltransferase characteristics

Some specific requirements were necessary to set up such a mathematical calculation. 18 prenyltransferases from different host plants were selected to serve as templates in the search for new putative PTs but also to get an overview of the structural properties of this class of enzymes (Table 34). These enzymes derived from different families of prenyltransferases, which are involved for example in the biosynthesis of tocopherol or flavonoids or ubiquinones. Sequences can be found in Table 42 in the supplementary information. This list does not only contain aromatic prenyltransferases, but also cis-PTs and protein-PTs, to set up a proper discrimination of all kinds of PTs.

Table 34: Prenyltransferases from various host plants for sequential analysis

| Number | Enzyme | Host plant |
|---|---|---|
| **1** | Protoheme IX farnesyltransferase (Cox10) | *Arabidopsis thaliana* |
| **2** | Prenyltransferase | *Arabidopsis lyrata* |
| **3** | Protein prenyltransferase | *Arabidopsis lyrata* |
| **4** | cis-prenyltransferase | *Arabidopsis thaliana* |
| **5** | Polyprenyltransferase | *Arabidopsis thaliana* |
| **6** | PT1 | *Cannabis sativa* L. |
| **7** | PT1p | *Cannabis sativa* L. |
| **8** | Aromatic prenyltransferase | *Epimedium acuminatum* |
| **9** | Flavonoids prenyltransferase | *Sophora flavescens* |
| **10** | *cis*-prenyltransferase | *Hevea brasiliensis* |
| **11** | Aromatic prenyltransferase | *Humulus lupulus* |
| **12** | *cis*-prenyltransferase | *Lilium longiflorum* |
| **13** | Protein prenyltransferase (Os11g0483950) | *Oryza sativa* |
| **14** | Prenyltransferase | *Populus trichocarpa* |
| **15** | Prenyltransferase | *Sophora flavescens* |
| **16** | Protoheme IX farnesyltransferase | *Ricinus communis* |
| **17** | Flavonoid prenyltransferase | *Sophora flavescens* |
| **18** | Prenyltransferase | *Zea mays* |

The proteins were analyzed for common properties that could serve for the identification of novel and yet unrecognized prenyltransferases. The properties examined were transmembrane helices, motifs, signal peptides and specific pattern architecture with web-based tools, i.e. TMHMM, TargetP, SignalP, ChloroP and cDART to detect similarities.

#### *4.1.3.1 Transmembrane helices*

For the analysis of transmembrane helix structure the web-tool TMHMM provided by the Technical University of Denmark was used. TMHMM provides a prediction of transmembrane helices based on a hidden markov model factoring in the hydrophobicity, charge bias, helix length and alternation between loops with cytoplasmic and non-cytoplasmic orientation (Jones et al., 1994; Krogh et al., 2001; Sonnhammer et al., 1998).

```
# 6 Length: 395
# 6 Number of predicted TMHs:  6
# 6 Exp number of AAs in TMHs: 136.52951
# 6 Exp number, first 60 AAs:  0.00264
# 6 Total prob of N-in:        0.49200
6        TMHMM2.0        inside       1   141
6        TMHMM2.0        TMhelix    142   164
6        TMHMM2.0        outside    165   183
6        TMHMM2.0        TMhelix    184   203
6        TMHMM2.0        inside     204   209
6        TMHMM2.0        TMhelix    210   232
6        TMHMM2.0        outside    233   235
6        TMHMM2.0        TMhelix    236   255
6        TMHMM2.0        inside     256   313
6        TMHMM2.0        TMhelix    314   336
6        TMHMM2.0        outside    337   339
6        TMHMM2.0        TMhelix    340   359
6        TMHMM2.0        inside     360   395
```

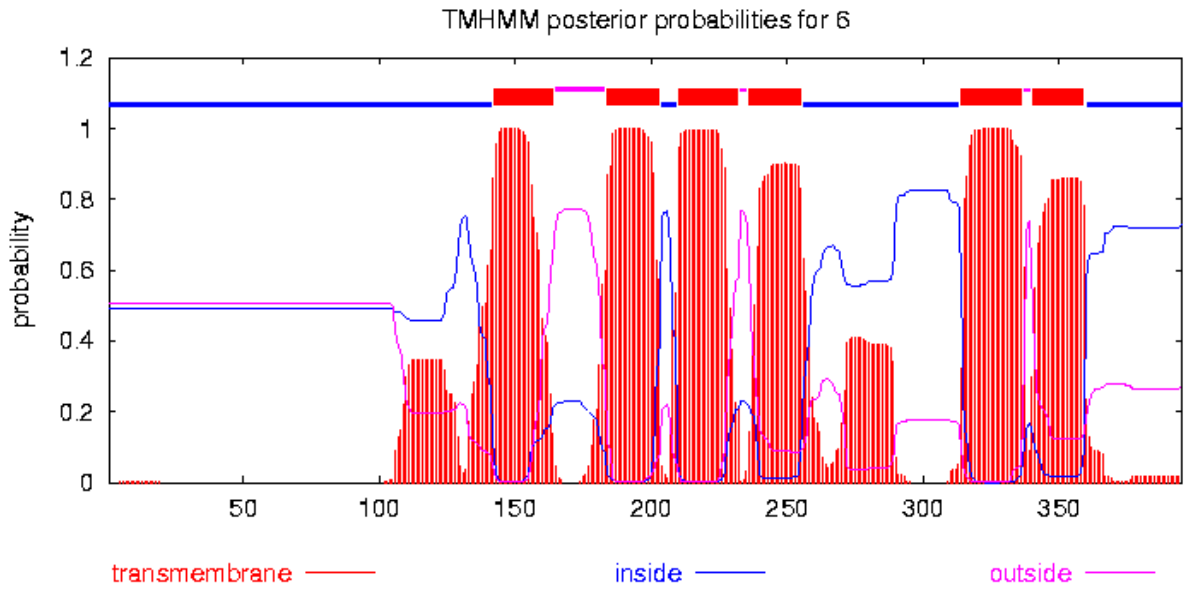



Figure 9: TMHMM analysis for PT1 (http://www.cbs.dtu.dk/services/TMHMM)

Figure 9 shows the calculated probabilities for transmembrane helices in PT1 within the amino acid sequence. The results are shown exemplary for all other 18 prenyltransferases (Table 42).

For PT1 the predicted number of transmembrane helices is determined to be six, which is a rather high value and hence very likely to be a transmembrane protein. The number of "expected amino acids in the TMHs" strengthens this prediction based on a high number of amino acids in the TMHs (cut-off value is 18) (Krogh et al., 2001).

Within the protein sequence there are two regions with some uncertainty of being predicted as TMHs or not. These regions last from positions around 110 to 140 and from 260 to 300. In the first region the uncertainty is also reflected by the orientation of the loop (inside or outside). But the calculated probability is not high enough to classify these regions as TMHs and the orientation of

both regions for being intracellular are higher than extracellular. The topology of the N- as well as the C-Terminus are both inside, which is often preferred by transmembrane proteins (Krogh et al., 2001).

Since the activity of the geranylpyrophosphate:olivetolate geranyltransferase (GOT) was demonstrated by experiments in microsomal fractions and leave extracts, it is very likely that the protein possesses transmembrane helices and therefore is membrane-bound (Fellermeier and Zenk, 1998; Page and Boubakir, 2011). Hence the flavonoid prenyltransferase (number 9 of the 18 prenyltransferases; Table 42) from *Sophora flavescens* was annotated as soluble protein, it was not integrated into the search.

The analysis of all enzymes (Table 42) was performed to get an overview of the quantity of transmembrane helices. A common number or structure of transmembrane helices could have been a good characteristic to be implemented into a mathematical search method, which was used in the query for new prenyltransferases in gene sequences of *Cannabis* in the 4 different DNA data sets. All results concerning the number of TMHs of the various PTs are displayed in Table 35. Since the number of TMHs varies from 6 to 9, a common structure related to TMHs could be excluded. Six of the tested prenyltransferases are predicted to have no TMH at all, which can be explained by their classification. Within those six enzymes the group of protein prenyltransferases and cis-PTs are represented. Both classes have soluble enzymes within their family.

#### *4.1.3.2 Signal peptides*

Another property of membrane bound enzymes is their signal peptide. This is another characteristic which can be implemented in a probabilistic mathematical search method. The presence of signal peptides was analyzed by using the freely available web tools TargetP, ChloroP and SignalP.

Results for the query in TargetP and ChloroP are displayed in Figure 10.

**A**

CENTERFO RBIOLOGI CALSEQU ENCEANA LYSIS CBS

**TargetP 1.1 Server - prediction results**

Technical University of Denmark

```
### targetp v1.1 prediction results ###############################
Number of query sequences:  17
Cleavage site predictions included.
Using PLANT networks.
```

| Name | Len | cTP | mTP | SP | other | Loc | RC | TPlen |
|---|---|---|---|---|---|---|---|---|
| 1 | 431 | 0.569 | 0.564 | 0.006 | 0.013 | C | 5 | 57 |
| 2 | 385 | 0.873 | 0.035 | 0.004 | 0.040 | C | 1 | 71 |
| 3 | 424 | 0.057 | 0.086 | 0.273 | 0.840 | _ | 3 | - |
| 4 | 253 | 0.044 | 0.410 | 0.091 | 0.482 | _ | 5 | - |
| 5 | 429 | 0.364 | 0.852 | 0.018 | 0.008 | M | 3 | 20 |
| 6 | 395 | 0.248 | 0.088 | 0.015 | 0.352 | _ | 5 | - |
| 7 | 395 | 0.316 | 0.069 | 0.020 | 0.228 | C | 5 | 75 |
| 8 | 407 | 0.159 | 0.714 | 0.014 | 0.122 | M | 3 | 38 |
| 10 | 296 | 0.009 | 0.901 | 0.032 | 0.373 | M | 3 | 79 |
| 11 | 411 | 0.923 | 0.012 | 0.013 | 0.162 | C | 2 | 91 |
| 12 | 308 | 0.113 | 0.420 | 0.068 | 0.468 | _ | 5 | - |
| 13 | 478 | 0.082 | 0.757 | 0.003 | 0.163 | M | 3 | 106 |
| 14 | 335 | 0.189 | 0.058 | 0.327 | 0.151 | S | 5 | 31 |
| 15 | 407 | 0.410 | 0.514 | 0.063 | 0.138 | M | 5 | 39 |
| 16 | 333 | 0.100 | 0.079 | 0.228 | 0.344 | _ | 5 | - |
| 17 | 407 | 0.410 | 0.514 | 0.063 | 0.138 | M | 5 | 39 |
| 18 | 425 | 0.606 | 0.095 | 0.205 | 0.043 | C | 3 | 41 |
| cutoff | | 0.000 | 0.000 | 0.000 | 0.000 | | | |

**B**

CENTERFO RBIOLOGI CALSEQU ENCEANA LYSIS CBS

**ChloroP 1.1 Server - prediction results**

Technical University of Denmark

```
### chlorop v1.1 prediction results ##########################
Number of query sequences: 17
```

| Name | Length | Score | cTP | CS-score | cTP-length |
|---|---|---|---|---|---|
| 1 | 431 | 0.545 | Y | 4.751 | 57 |
| 2 | 385 | 0.551 | Y | 0.882 | 71 |
| 3 | 424 | 0.427 | - | -2.317 | 5 |
| 4 | 253 | 0.448 | - | 3.496 | 27 |
| 5 | 429 | 0.545 | Y | -0.722 | 40 |
| 6 | 395 | 0.518 | Y | 8.008 | 75 |
| 7 | 395 | 0.510 | Y | 8.008 | 75 |
| 8 | 407 | 0.525 | Y | 2.426 | 37 |
| 10 | 296 | 0.435 | - | -3.230 | 4 |
| 11 | 411 | 0.563 | Y | 5.337 | 91 |
| 12 | 308 | 0.531 | Y | -2.400 | 50 |
| 13 | 478 | 0.489 | - | 7.900 | 38 |
| 14 | 335 | 0.525 | Y | 5.437 | 35 |
| 15 | 407 | 0.520 | Y | 2.154 | 53 |
| 16 | 333 | 0.520 | Y | 4.419 | 89 |
| 17 | 407 | 0.520 | Y | 2.154 | 53 |
| 18 | 425 | 0.517 | Y | 4.106 | 41 |

Figure 10: Analysis for various prenyltransferase protein sequences (Table 42); **A**: Query in prediction tool TargetP: length of protein sequence; probabilities for cTP (chloroplast transit peptide), mTP (mitochondrial target peptide), SP (secretory pathway, signal peptide), other (any other location); prediction of location; reliability class, 1 indicates strongest and 5 weakest prediction; predicted presequence length; **B**: Query in prediction tool ChloroP: length of protein sequence; the score covers the region between 0.4 to 0.6; scores above 0.5 indicate a potential cTP, scores above 0.55 indicate a likely cTP; cTP: yes (Y) or no (-); CS-score: cleavage site score; cTP length in amino acids.

In TargetP it is possible to discriminate between cTPs (chloroplastid transit peptides), mTPs (mitochondrial transit peptides) and other SPs (signal peptides). This prediction tool has overlapping similarities with ChloroP. One result of TargetP is represented by the reliability class. This value ranges from 1 to 5 and stands for the probability for the existence of a signal peptide based on the highest and second highest output score. The lower the RC value, the more probable is the prediction made by the calculations of TargetP.

The prediction platform “ChloroP” served as tool for a more specific sequence analysis whether the amino acid sequences of the 18 plant prenyltransferases carried a chloroplastid target protein or not. ChloroP is based on a neural network and predicts also the length of the chloropastid peptide. Neural networks are learning systems, which generalize from examples these networks were incorporated with (Emanuelsson et al., 2007). With ChloroP it is also possible to predict the cleavage sites by using a scoring matrix derived by an automatic motif-finding algorithm (Emanuelsson et al., 1999). The score for this MEME scoring matrix is represented by the CS-score.

The mathematical calculations of ChloroP for all plant prenyltransferases are shown in Table 35. The characteristics of chloroplast transit peptides are: an N-Terminus beginning with a MA combination and comprising altogether 10 amino acids, which are free of charge and ending with an glycine or proline residue; followed by a domain with a high number of hydroxylated amino acids especially serines and threonines, but free from acidic amino acids; a C-Terminus containing a lot of arginines,

which possibly form a β-strand (Emanuelsson et al., 2007). The score in ChloroP can range between 0.4 to 0.6 and stands for the likelihood, that there is a chloroplast transit peptide within the protein sequence. Scores higher than 0.5 indicate the presence of a chloroplast transit peptide. The CS-score represents the prediction of the cleavage site. The length of the peptide is deduced from this result (Emanuelsson et al., 2007).

The development of SignalP started from a method which was only based on neural networks. As a further step Hidden Markov models were implemented to the computation and as a third step the composition of all amino acid residues was considered (Emanuelsson et al., 2007). The various scores are calculated from the arrangement of the amino acids in the mature protein and indicate the presence or absence of a signal peptide. The C-score stands for the raw cleavage site score. If a cleavage site exists within the protein this value is highest immediately behind the cleavage site and therefore equal to the starting point of the mature protein (Emanuelsson et al., 2007). The S-score (signal peptide score) differs between enzymes with signal peptide and without. This score is high at all position before the cleavage site. Both scores are combined to the Y-score for precise determination of the cleavage site. It is calculated from the height of C-score graph and slope of the S-score graph (Emanuelsson et al., 2007). The discrimination score (D-score) differentiates between SPs and non-SPs. For that it takes the average of the mean S-score and the maximum Y-score into consideration. The mean S-score is calculated from the N-Terminus to the highest Y-score position (Emanuelsson et al., 2007).

A representative prediction result for one of the plant prenyltransferases (PT1p) is shown in the supplementary information (Figure 45). Since the prediction of all PTs resulted in “no SP” the remaining graphs are not shown. The range of values from 0.109 to 0.244 are too low to be identified as signal peptide.

The inconsistency in the results of SignalP and ChloroP could arise from the fact that ChloroP implements besides the neural networks, weight matrices into the mathematical method, which determines the specific position of the cleavage site of the transit peptide (Emanuelsson et al., 2007).

Two prenyltransferases are predicted to have no TMHs and no target peptide. These two enzymes are classified into the group of cis- and protein prenyltransferases, which are known to possess soluble proteins within their families. Three more PTs are predicted to have no TMHs. But these enzymes are predicted to possess either a mitochondrial target peptide or a chloroplast transit peptide.

As displayed in Table 35 the tool SignalP did not identify a single signal peptide for any of the 18 PTs, which is contradictory to 16 of the enzymes when comparing to the results of TargetP and ChloroP.

However, since the results for all prenyltransferases are contradictory in accordance to the presence of a signal peptide, its length and position within the amino acid sequence, this characteristic is not considered to be helpful to implement into a probabilistic mathematical calculation to search for PT candidates in the 4 genetic sequence data sets of *Cannabis*.

Table 35: Identification of prenyltransferases (Table 34) characteristics; TargetP: cTP (chloroplast transit peptide), mTP (mitochondrial targeting peptide), secret. P (secretory pathway, sequence contains signal peptide), - (any other location); SignalP: no SP (sequence contains no signal peptide), SP (sequence contains signal peptide); ChloroP: cTP; number of predicted TMHS (transmembrane helices)

| Enzyme | TargetP | SignalP | ChloroP | Number of predicted TMHs |
|---|---|---|---|---|
| **Protoheme IX farnesyltransferase (Cox10)** | cTP | No SP | cTP | 9 |
| **Prenyltransferase** | cTP | No SP | cTP | 6 |
| **Protein prenyltransferase** | - | No SP | - | 0 |
| **cis-prenyltransferase** | - | No SP | - | 0 |
| **Polyprenyltransferase** | mTP | No SP | cTP | 6 |
| **PT1** | - | No SP | cTP | 6 |
| **PT1p** | cTP | No SP | cTP | 6 |
| **Aromatic prenyltransferase** | mTP | No SP | cTP | 7 |
| ***cis*-prenyltransferase** | mTP | No SP | - | 0 |
| **Aromatic prenyltransferase** | cTP | No SP | cTP | 8 |
| ***cis*-prenyltransferase** | - | No SP | cTP | 0 |
| **Protein prenyltransferase (Os11g0483950)** | mTP | No SP | - | 0 |
| **Prenyltransferase** | secret. P | No SP | cTP | 7 |
| **Prenyltransferase** | mTP | No SP | cTP | 9 |
| **Protoheme IX farnesyltransferase** | - | No SP | cTP | 8 |
| **Flavonoid prenyltransferase** | mTP | No SP | cTP | 9 |
| **Prenyltransferase** | cTP | No SP | cTP | 0 |

#### *4.1.3.3 Domain architecture*

Another possible common characteristic for a search tool was to find a specific architecture of domains within in the list of prenyltransferases. It is already known that aromatic prenyltransferases carry a common UbiA-motif. But for specific bioinformatic research it is important to know, if there are any other domains which form an architecture with this motif. The analysis was performed with

the web-based tool cDART (conserved domain architecture retrieval tool). cDART bases its search on an algorithm which uses several domain definitions for detection of an architecture and therefore uses protein domain profiles. Besides queries of domain architectures within a group of proteins, it is also possible to query a single protein and subsequently create a group with proteins of the same domain architecture (Geer et al., 2002). If a specific domain architecture exists, these could be implemented to search for alternatives for the already known membrane bound aromatic plant prenyltransferases.

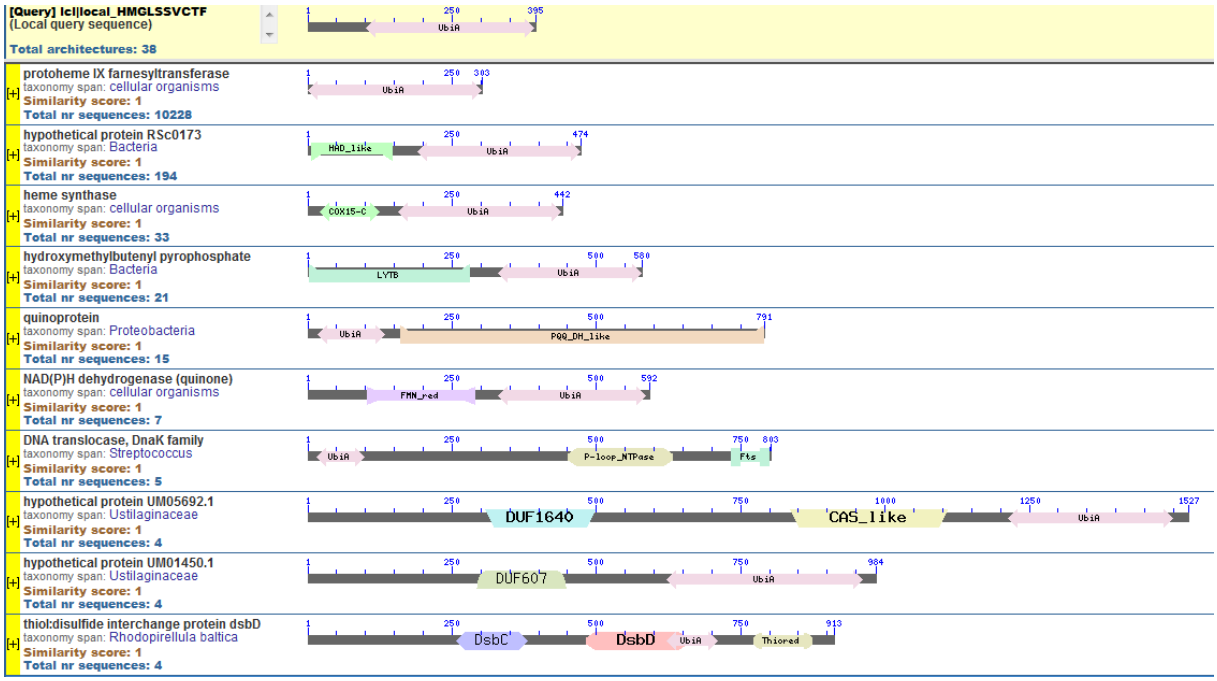



Figure 11:: cDART analysis of prenyltransferases

The search in cDART did not give a unique domain architecture of different motifs within the protein sequences. The only common pattern is the already identified UbiA prenyltransferase family signature pattern. The pattern is also deposited in the database of Pfam (PF01040; Figure 47). Thus searching for a motif architecture is also not an option for implementation to the mathematical calculation.

The query for all these different characteristics of prenyltransferases was performed to specify the mathematical method, which was used in the following to find new putative aromatic prenyltransferases in the 4 data sets of Cannabis.

### *4.1.3.4 Search method for aromatic prenyltransferases*

The search for common characteristics of aromatic prenyltransferases resulted in the confirmation of the UbiA motif as common. Apart from that it can be stated that transmembrane helices as well as

signal peptides are common, but not necessary for aromatic prenyltransferases. Also, no common domain architecture exists for these enzymes. Therefore the UbiA motif with the integrated $Mg^{2+}$ binding motif (IPR000537) was implemented into the mathematical method as search characteristic.

The sequences of the publicly available annotated transcripts were extracted. Additionally, RNA-seq reads were assembled using the Newbler version 2.6 (Margulies et al., 2005) assembler in cDNA mode, creating 12892 isotigs.

433 potential plant prenyltransferase protein sequences were downloaded from the UniProt database (Boutet et al., 2007). The sequences were searched against the InterPro database (Zdobnov and Apweiler, 2001) using InterProScan. Based on the results the proteins were classified into three classes: “UbiA Protohaem IX farnesyltransferase” if the InterPro domains IPR000537 and IPR006369 were found on the protein, “UbiA” if only the IPR000537 domain was found and “No UbiA” if the IPR000537 domain was missing. 238 sequences could be assigned to class “UbiA”, which were further used to identify potential prenyltransferase proteins in the *C. sativa* datasets.

All *C. sativa* transcript sequences were translated to amino-acid sequences for all six reading frames. The 238 UniProt sequences were blasted (Altschul et al., 1997) against the amino-acid translations. All *C. sativa* sequences with blast-hits with an e-value smaller than $10^{-5}$ and completely covering the UniProt query sequence were selected. In total 35 potential complete potential prenyltransferase proteins without an in-frame stop-codon were found. For each of these sequences multiple alignments with all the matching UniProt protein sequences were created using the mafft tool (Katoh et al., 2002) and manually checked. Out of the 35, six sequences were selected for further processing.

This part of the thesis was performed at the Center for Biotechnology (CeBiTec) in Bielefeld.

Table 36: *In silico* identified putative aromatic prenyltransferases in *Cannabis sativa* L. spp.

| Prenyltransferase | Purple Kush | Finola | MPGRT | Trichome cDNA library |
|---|---|---|---|---|
| **PT1** | PK28436.1 | FN08295.1 | MPGRT_csa_locus_9717_iso_1_len_1296 | CSAI_isotig04415 |
| **PT2** | PK29226.1 | | MPGRT_csa_locus_5486_iso_1_len_407 | |
| **PT3** | PK15523.1 | FN24824.1 | MPGRT_csa_locus_4168_iso_1_len_1266 | CSAI_isotig07002 |
| **PT4** | PK17697.1 | FN16821.1, FN27235.1 | | CSAI_isotig00533, CSAI_isotig00535, CSAI_isotig11658 |
| **PT5** | PK11068.1 | FN01550.1 | MPGRT_csa_locus_9487_iso_1_len_1391 | CSAI_isotig06979 |
| **PT6** | PK04673.1 | FN24104.1, FN36096.1 | MPGRT_csa_locus_10855_iso_1_len_1454, MPGRT_csa_locus_10855_iso_2_len_1448 | CSAI_isotig10460, CSAI_isotig11013 |

Table 36 shows the five putative membrane-bound aromatic prenyltransferases resulting from all four data sets of *Cannabis sativa* L. spp.. As explained earlier in this chapter, PT1 is very similar to the patent-registered aromatic prenyltransferases from Cannabis (Table 33). The five enzymes identified *in silico* show sequence similarities to this protein. Some of the putative enzymes can be found in all four data sets and some are only found in particular sets. To evaluate the evolutionary relationship between these putative aromatic PTs and the already known prenyltransferases, a phylogenetic tree was generated using MEGA7 software (Figure 12). The tree distinguishes between enzymes involved in the biosynthesis of ubiquinone, shikonin, tocopherol, flavonoid, plastoquinone, chlorophyll, tocotrienol. Three of the five new proteins belong to the known families of ubiquinone prenyltransferases (PT6) and plastoquinone prenyltransferases (PT4&5). PT2 and PT3 are assembled in a group with PT1 (PT1p) and the aromatic prenyltransferase from *Humulus lupulus*. This indicates that these enzymes have a close ancestral lineage, suggesting, that PT2 and PT3 are likely to be promising candidates to serve as alternatives for PT1 if they catalyze the same reaction and possibly can be expressed easier than PT1 in heterologous microorganisms.

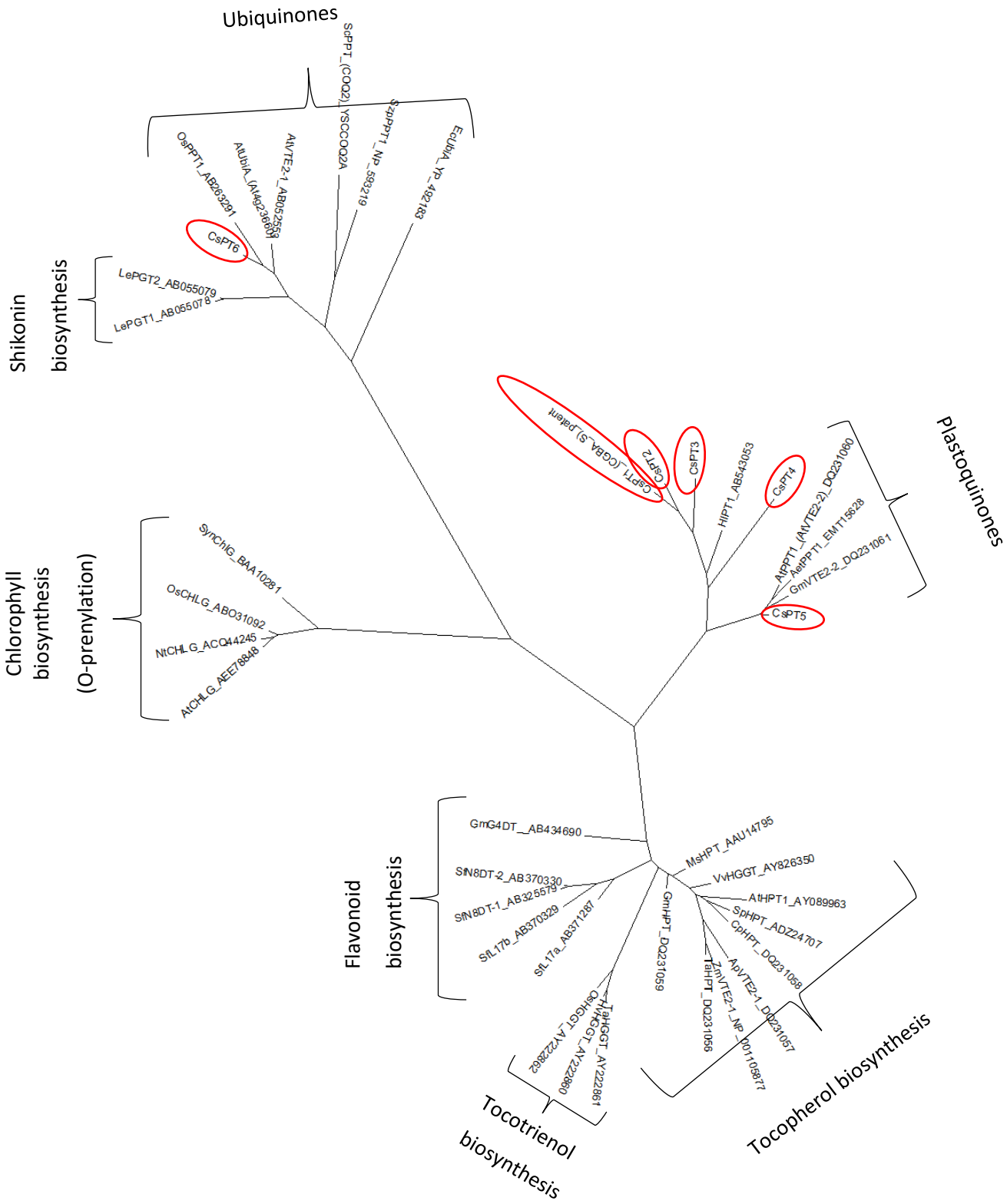



Figure 12: The tree with the highest log likelihood is shown. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join. The analysis involved 39 amino acid sequences. Evolutionary analyses were conducted in MEGA5 (Jones et al., 1992; Tamura et al., 2011).

## 4.2 Expression profile of prenyltransferases identified *in silico* from *Cannabis sativa* L. spp.

A prenyltransferase involved in the production of cannabinoids should be expressed specifically in the place of biosynthesis, that is in the trichomes and flowers but not or less in other parts of the plant. Therefore an expression profile of the prenyltransferases identified *in silico* was generated. Access to Cannabis is restricted, therefore only the variety Bediol (week 8) was available and, the expression profile was determined in this variety. This variety was already used for the generation of the CSA transcriptome library. The expression profile of the prenyltransferases identified *in silico* illustrates the dispersion of the putative prenyltransferases within *Cannabis sativa* L. var. Bediol.

The expression profile was realized by checking the presence of the PTs by PCR. The primers were designed so that the length of the PCR products was set to around 700 bp.

### 4.2.1 Development of a specific detection method

As a result of high sequence similarities, cross creations between PT1, 2 and 3 occurred, which means that for instance the primer pair configured for PT1 also annealed to the cDNA sequence of PT2 and PT3. To prevent this cross reactions, the annealing temperatures during the PCR were optimized for these 3 putative prenyltransferases and the evidence of successful annealing of the correct enzyme/primer-pair was determined by using PT templates previously ligated into the modified cloning vector pGEM-T easy (Figure 26).

One of the resulting agarose gels is shown in Figure 13. It represents the PCR products with cloning vector pGEM-T easy carrying PT3 as template. The primer pairs of PT1, PT2 and PT3 are alternating among each other from lane to lane beginning with the primers for PT1, followed by the primers for PT2 and finalizing with the primers for PT3. The annealing temperatures range from 55°C to 70°C and are shown at the bottom of the gel. The agarose gels in which PT1 and PT2 were used as templates, are shown in the supplementary information (Figure 48).

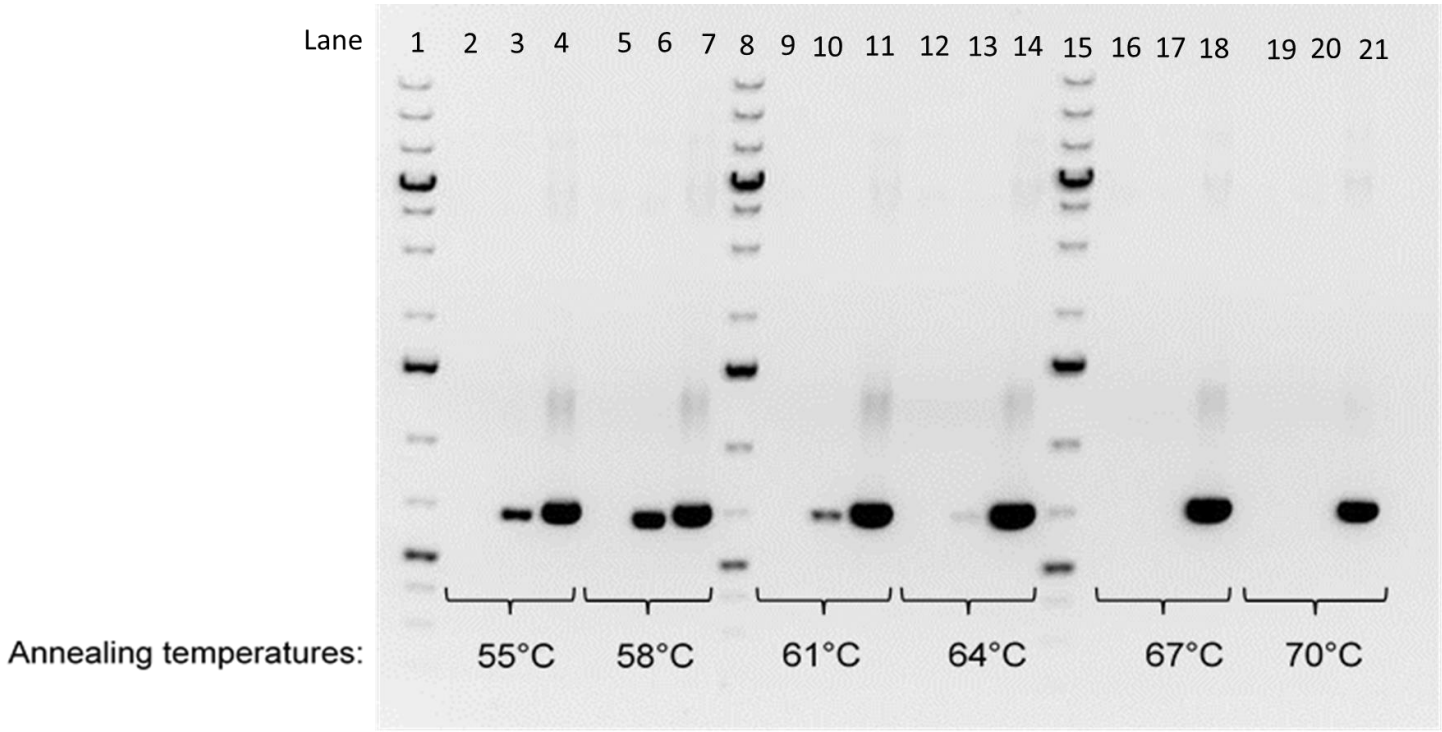



Figure 13: Optimization of annealing temperature for primer/enzyme pair detecting PT3; template used was cloning vector pGEM-T easy carrying PT3; primer pairs alternated so that for each annealing temperature each primer pair was tested; primer pairs for PT1 (lanes 2, 5, 9, 12, 16, 19), PT2 (lanes 3, 6, 10, 13, 17, 20) and PT3 (lanes 4, 7, 11, 14, 18, 21) were used; Lane 1, 8, 15: DNA Ladder (1kb, Thermo Scientific, Waltham, Massachusetts, USA); annealing temperatures are displayed at the bottom of the gel and were set from 55°C to 70°C.

To determine the correct annealing temperature, the temperature during PCR was raised until only the primer designed to bind was actually binding and thus producing a PCR product. The primer pair designed for detecting PT1 was no longer binding to the PT2 template as well as the PT3 template at a temperature of 55 °C (Figure 13 and Figure 48). Hence this was the specific annealing temperature for the PT1 primer pair to anneal to its corresponding cDNA sequence during a PCR reaction. The specific annealing temperature for verification of PT2 was determined to 67 °C. When looking at the agarose gels there was still a very prominent band appearing at 64 °C in case of using the PT1 template (Figure 48) and a slight band in case of using the PT3 template (Figure 13). At 67 °C there were neither cross reactions with the PT1 template nor with the PT3 template. The PCR reaction using the PT3 detecting primers gave also some weak bands at a temperature of 64 °C (Figure 48). When raising the temperature by 3 °C to 67 °C, these bands disappeared and the primer pair was specifically detecting PT3.

It can be concluded that the annealing temperatures had to be set to 55 °C for specific detection of PT1 and to 67 °C for PT2 respectively PT3 during the PCR reactions for the expression profile of the putative PTs.

### 4.2.2 Determination of expression profile

The ubiquitous enzyme Glyceraldehyde-3-phosphate dehydrogenase (gapdh) was used as positive control for the expression profile of all *in silico* detected PTs. GAPDH is a central carbon metabolism enzyme and is therefore present in any living tissue of the plant. It therefore is a good indicator for the success of the RNA isolation and subsequent cDNA synthesis. Figure 14 shows excerpts of the agarose gels representing the PCR products of the six putative PTs and the positive control in various tissues of *Cannabis sativa* L. var. Bediol (week 8). The original agarose gels can be found in the supplementary information (Figure 49-Figure 55).

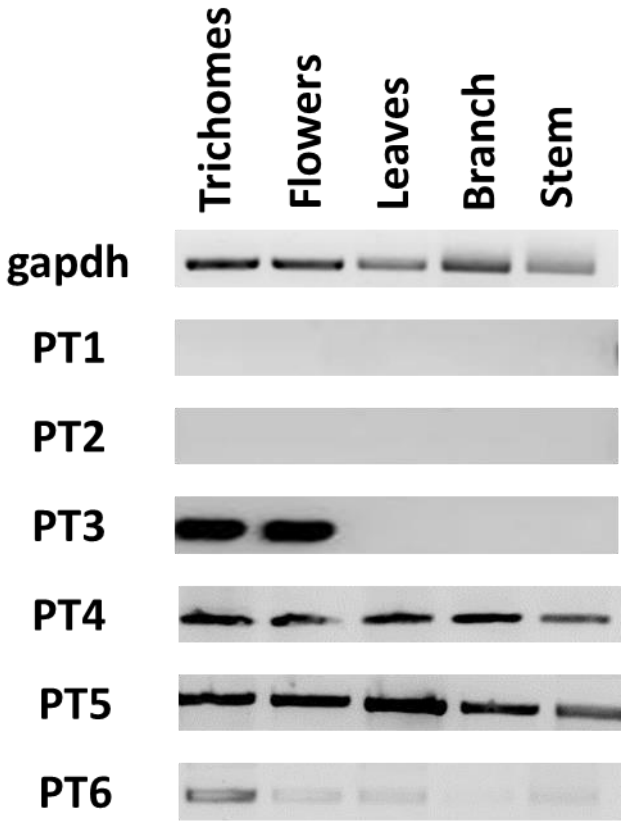



Figure 14: Excerpts of agarose gels from RNA extraction with subsequent cDNA synthesis and PCR with primer pairs for gapdh, PT1, 2, 3, 4, 5 and 6 (Table 37); gapdh serves as positive control; RNA was extracted from different tissues of *Cannabis sativa* L. var. Bediol; original agarose gels are resided in the supplementary information (Figure 49- Figure 55).

The volumes loaded on all gels for the various PT products were chosen so that the gapdh band was approximately of the same strength for all plant tissues. The PCR products from all putative prenyltransferases were designed to have a length of about 700bp due to better comparability. One exception was made for PT6, because the design of a stable primer pair was not accomplishable for that product length. Hence this PCR product has a length of about 500bp.

As expected the positive control gapdh is present in all tissues of *Cannabis sativa* L. var. Bediol.

PT1 and PT2 are not present at all in this *Cannabis* variety. Concerning PT2 this is in line with the transcriptome data set as this cDNA sequence was not found in the dataset of the trichome cDNA library of this variety (Table 36). Probably this enzyme is not universal to all *Cannabis* varieties.

The apparent absence of PT1 in the plant could be explained by the normalization of the trichome cDNA library. Due to the normalization weakly transcribed genes are detectable in the cDNA library, while during the expression profile there was no such step of adjusting different transcription levels.

It is possible that the actual transcription level in the plant trichomes of PT1 is very low and therefore the volume loaded on the gel is not large enough to see a band. Weak transcription of PT1 would argue against the role as a key enzyme of cannabinoid synthesis.

PT3 was detected in trichomes and flowers of the plant, which hints at a specific location in the plant. The high intensities of the two appearing bands indicate a high amount of PT3 in trichomes and flowers. So the specific location, the high presence and the phylogenetic analysis strongly suggest that the protein has a specific function and that it could serve as promising candidate for the Geranylpyrophosphate:olivetolate geranyltransferase (GOT).

PT4 and PT5 were phylogenetically grouped into the family of PTs which are involved in the biosynthesis of plastoquinones. Plastoquinones are electron acceptors during the electron transport chain as part of the photosynthesis. This fact would hint at an omnipresence and a high amount of transcription of these genes in every tissues within the plant, which was reassured by the results displayed in the gels. Having a closer look at the whole agarose gel (Figure 53 and Figure 54), there are a lot of smaller by-products visible after the PCR, which are also present in the control reaction in which only RNA template (from stem tissue) was added but no reverse transcriptase during the cDNA synthesis but also the PCR reaction with the according primers. An explanation could be the presence of genomic DNA, which interferes with the primer pairs of PT4 and PT5.

The gel reveals that the putative prenyltransferases PT4, PT5 and PT6 were present in all tissues of the plant just as gapdh. Based on the phylogenetic analysis PT6 was classified as ubiquinone producing enzyme. The approximate amount of PT6 was weak, but noticeable. The result indicates, that this enzyme is not commonly transcribed in the tissues like stem and branches and thus the biosynthesis of ubiquinones is not executed in these tissues in high amounts. It is possible that these parts are not favorable for the production of ubiquinones. Another explanation could be that the RNA was already degraded due to the age of the plant or during the storage at -80°C. A more detailed look at the whole gel (Figure 55) detecting PT6 reveals that there are also impurities, which derive from the RNA sample. But since the band is occurring at around 1500bp, it does not interfere with the PT6 PCR product band at around 500bp.

Summing up the findings of the expression profile PT3 is the most likely candidate for the Geranylpyrophosphate:olivetolate geranyltransferase (GOT) function in *Cannabis sativa* L. var Bediol since it shows a high sequence similarity with the patented sequence PT1p and is exclusively located in tissues associated with cannabinoid biosynthesis. PT1 and PT2 were not present in this variety.

PT4, PT5 and PT6 were identified in varying amounts in the whole plant and are likely to have a function that is not related to cannabinoid production.

## 4.3 Transient expression in *Nicotiana benthamiana*

To determine the function of the various detected prenyltransferases and to verify the theoretical results, the enzymes identified *in silico* were heterologously expressed. Therefore, two hosts were used: *Nicotiana benthamiana* and *Saccharomyces cerevisiae* Y05416. The plant host *Nicotiana benthamiana* was chosen because the heterologous expression of a plant prenyltransferase in another plant as host was considered as reasonable. The transient expression system magnICON® was chosen for the experiments with the prenyltransferases PT1, PT2 and PT3. PT1 was utilized for the verification of the aromatic prenyltransferases identified in *Cannabis* (Page and Boubakir, 2011), while PT2 and PT3 were chosen as promising alternatives to the aromatic prenyltransferases due to phylogenetic analysis and localization by the expression profile of the enzymes identified *in silico*.

The genes of interest were cloned into the pICH11599 vector, a 3'-module. For locating the gene of interest into certain compartments of the plant cells, a 5'-module is used. Three different locations were used: the chloroplast, the apoplast and the cytosol. A third module represented the recombinase module. This served to fuse the 3'-module and 5'-module on mRNA level into one mRNA. A GFP-construct served as positive control (Table 5). *Agrobacterium tumefaciens* cell cultures were cultivated with the various modules and afterwards a cell suspension with the respective cells were used for infiltration of whole tobacco plant.

### 4.3.1 Cloning and ligation

The transient expression with magnICON® system is fast, results in high expression levels, is versatile and an open modular system. magnICON® was developed by Icon genetics (Halle/Saale, Germany). The reaction exploits various advantage of different techniques. Plants as expression hosts are a simple-to-use and self-contained eukaryotic system (Marillonnet et al., 2005). The efficiency of plant transfection by *Agrobacterium* is unrivaled by any other system. Utilizing the deconstructed viral vectors guarantees high expression levels and quick spreading within the plant. With this so called 'magnifection' process, it is possible to produce high quantities of recombinant proteins (Gleba et al., 2007, 2005; Marillonnet et al., 2005).

First of all, the putative prenyltransferases were cloned into the cloning vector pGEM-T easy. The vector was provided by the working group of Prof. Dr. Frank Schulz. It was multiplied and modified by inserting the restriction sites XbaI and XcmI and replicated by PCR (section 3.5). The XcmI sites were designed to produce a T-overhang. An agarose gel can be found in the supplementary information (Figure 56).

The plasmid after XcmI restriction was used in the further ligation step. Due to its T-overhangs, it could be used in a TA-cloning reaction with PT1, 2 and 3. The inserts were amplified from RNA and subsequent cDNA synthesis from *Cannabis sativa* L. Bediol with primers containing restriction sites for NcoI and BamHI. To amplify the whole prenyltransferases different primers were used than during expression profiling. The fact that it was possible to amplify PT1 and PT2 from Bediol cDNA but they were not detectable during expression profiling shows the low concentration of the transcripts of these enzymes. This could hint on PT1 and PT2 playing a minor role in the plant metabolism compared to PT3.

After amplification and gel purification the amplified putative prenyltransferases were concentrated by ethanol precipitation due to low concentration. To check if the precipitation worked out and to estimate the volumes needed for ligation another agarose gel was run (Figure 57). TA-cloning was performed with the modified pGEM-T easy and the various prenyltransferases and transformation into E. coli DH5α. Integration was checked by colony PCR (Figure 58). Positive clones were sent for sequencing.

Sequentially correct clones were cut with NcoI and BamHI and ligated into the expression vector pICH11599. Subsequently the constructs were transformed into *Agrobacterium tumefaciens* LBA4404 and GV3101. Agar plates with LBA4404 did not show any growth. Hence for the following experiments GV3101 was used exclusively.

### 4.3.2 Transfection

Vacuum infiltration was used for plant transfection with a bacterial suspension. For each infiltration set-up three main cultures were cultivated and mixed in the ICON buffer, always comprising a 3'-, 5'- and recombinase module. Transfection of the young *Nicotiana benthamiana* plants by vacuum infiltration is illustrated exemplarily in Figure 15.

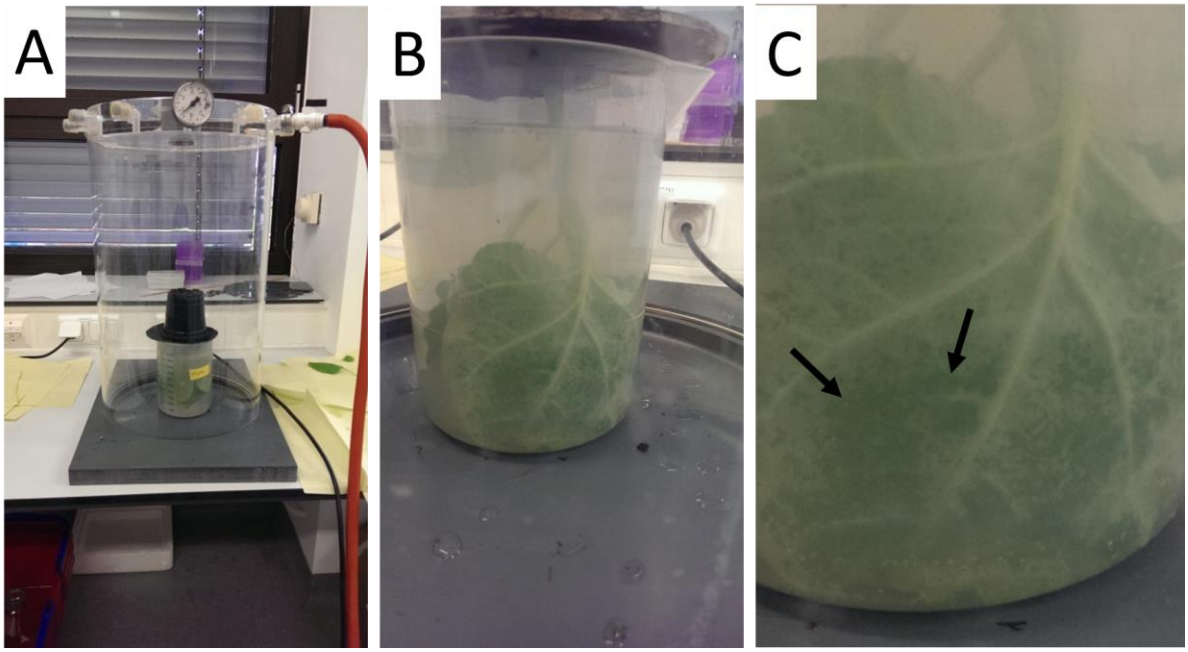



Figure 15: Vacuum infiltration set-up of young *Nicotiana benthamiana* plants; A: overhead dome for vacuum infiltration; B: beaker with *A. tumefaciens* GV3101 cells resuspended in ICON buffer; turned and submerged *Nicotiana benthamiana* plant; C: arrows indicating the areas where bacterial solution already infiltrated the leaf (darker areas).

Vacuum up to -0.7 bar was generated under an overhead cavity by a vacuum pump (Figure 15A). When the vacuum is released, the bacterial suspension infiltrates the leaves of the plant due to opening the stomata on the leaves (Figure 15B and C).

After magnifection plants were grown at room temperature with natural light. The plants grew well for over a week without apparent signs of infection until they withered and died within 24 hours. Consequently plant leaves were sampled over the course of several days.

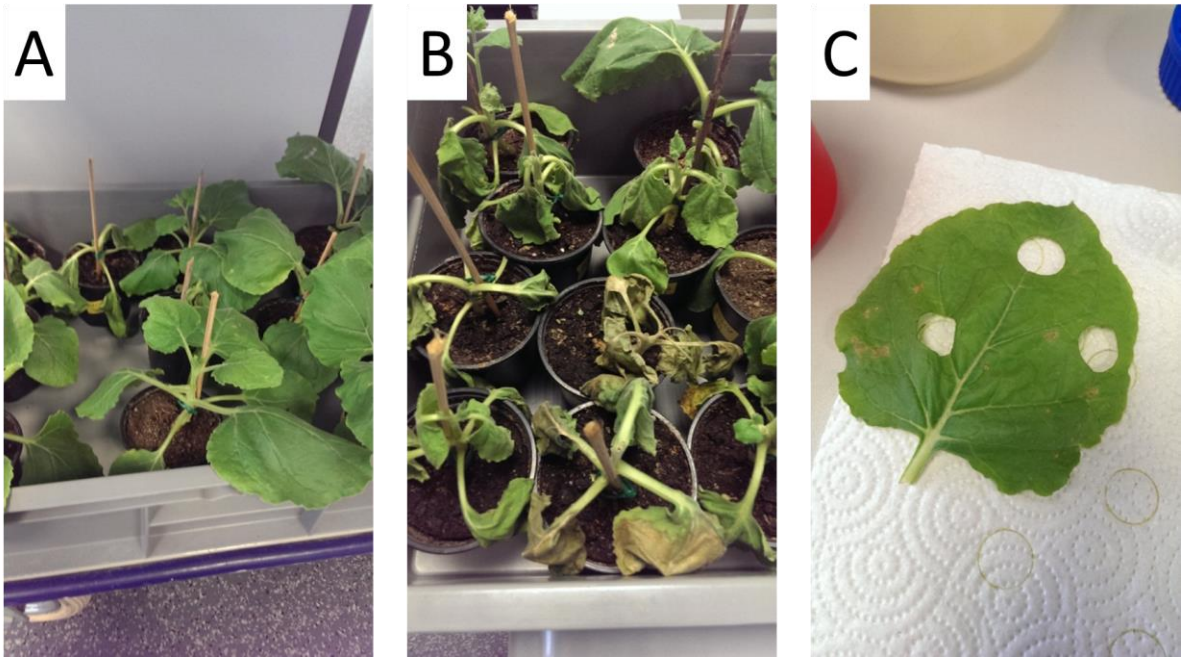



Figure 16: Infected *Nicotiana benthamiana* plants; A: 8 days after infiltration; B: 9 days after infiltration; C: Leaf from *N. benthamiana* after sampling for activity assay.

Figure 16A and B shows the *N. benthamiana* plants after 8 and 9 days after infection with *A. tumefaciens* GV3101. Figure 16C shows an exemplary leaf after sampling for the activity assay for CBGA. The success of the transfection of the *Nicotiana benthamiana* plants was monitored by using a GFP construct alternatively to the putative prenyltransferases. Some of the resulting GFP-expressing-plants are displayed in Figure 17.

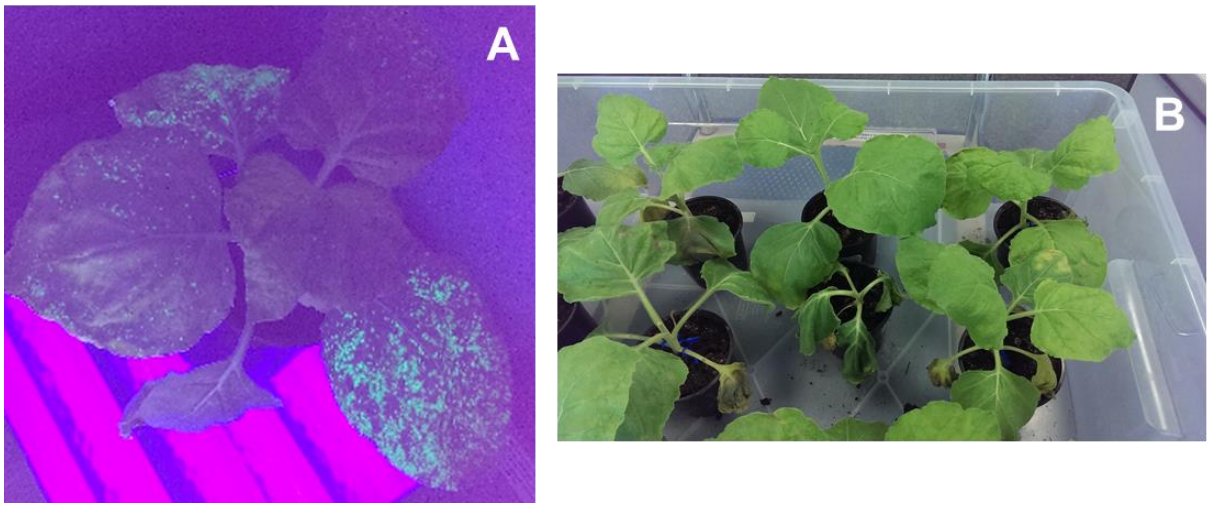



Figure 17: *Nicotiana benthamiana* transfected with *Agrobacterium tumefaciens* GV3101 carrying the MagnICON system with GFP; incubation 7 days.

To assess the quality of the 'magnifection' system, GFP expression in the leaves was monitored. The fluorescence was documented with the gel documentation system under UV light. The fluorescence was present in parts of the plants and appeared like sprinkles on the leaves instead of large areas as described in literature (Marillonnet et al., 2005, 2004). The incubation of the bacterial suspensions for two or four hours prior to vacuum infiltration of the plants did not have any effect on the strength or distribution of fluorescence. The variation of the pH between pH 4 and pH 5 in the MES buffer did also not result in an increase of fluorescence in the tobacco plants. An influence however was found in the choice of plants appearing subjectively healthy at the time of transfection. This optimization yielded an extensive and evenly spread fluorescence (Figure 18).

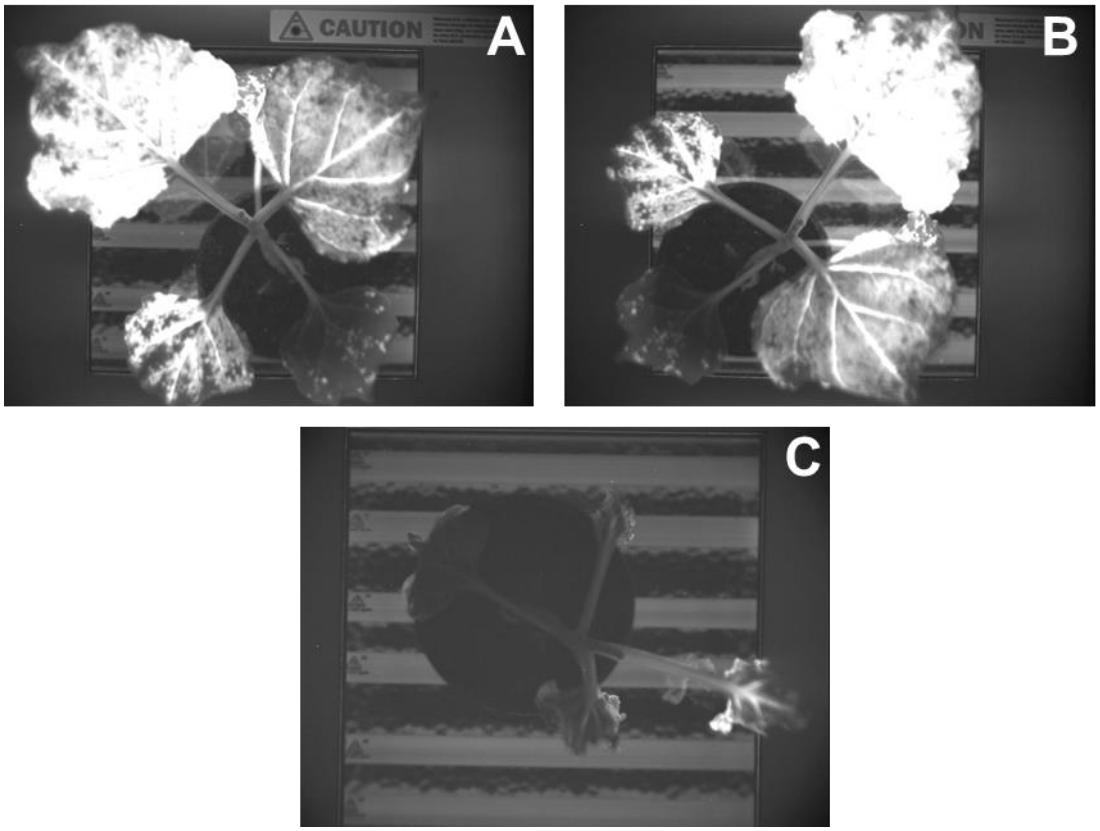



Figure 18: *Nicotiana benthamiana*; infection with *Agrobacterium tumefaciens* with GFP construct; A: incubation 8 days; B: incubation 9 days; C: incubation 10 days.

### 4.3.3 Activity assay and analysis

Sampling was performed after 8, 9 and 10 days. Experience shows that this incubation period is appropriate for transient expression in tobacco plants. The activity assay procedure for PT1 was

performed, but no conversion to CBGA was observable in the MS data. A representative chromatogram is illustrated in Figure 19.

As was discovered well after the end of my lab work the olivetolic acid that was used for these experiments was wrongly synthesized and was 2,6-Dihydroxy-4-pentylbenzoic acid instead of 2,4-Dihydroxy-6-pentylbenzoic acid. Therefore, CBGA synthesis was not feasible. It is therefore not possible to say if the PTs were indeed actively expressed or not. Nevertheless, the constructs carrying the PTs identified *in silico*, PT1, PT2 and PT3, were infiltrated, extracted and the activity assay for aromatic PTs from *Cannabis* was performed.

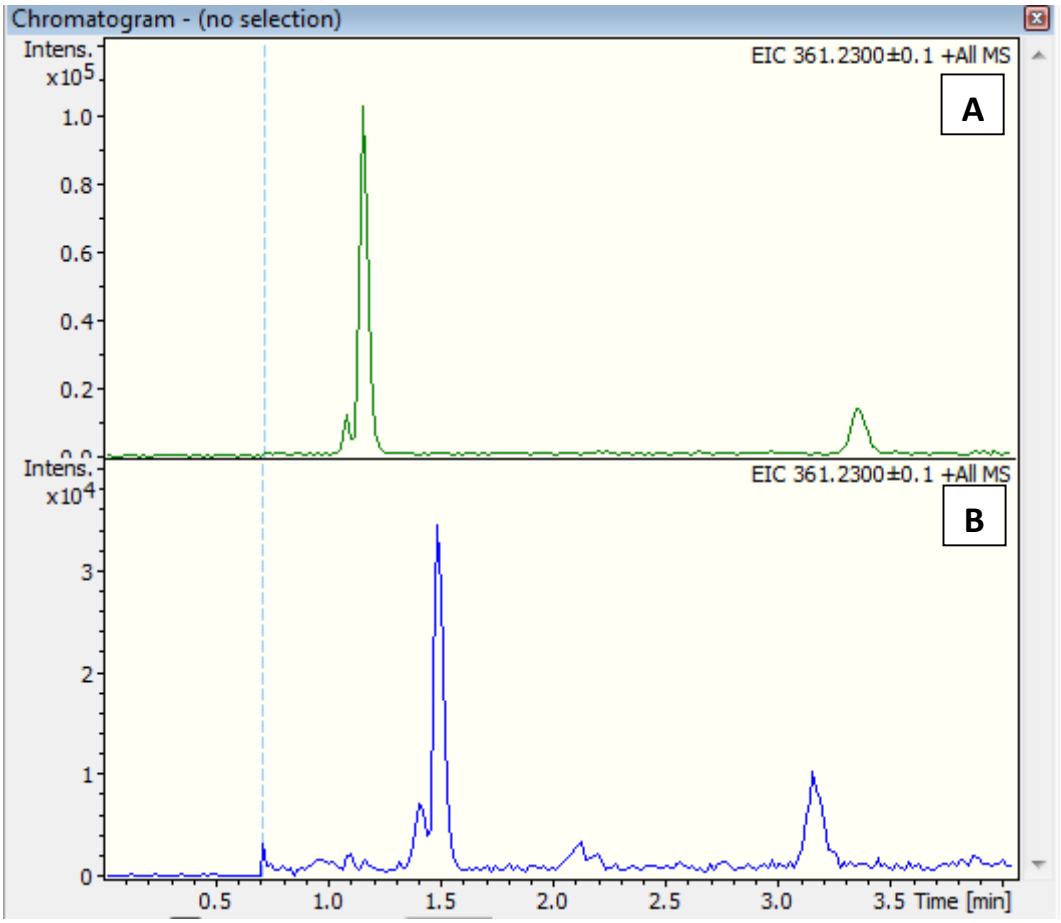



Figure 19: Extracted ion chromatogram for CBGA (m/z=361,23±0.1); A: EIC of CBGA standard (100µM), retention time of CBGA 3.4 minutes; the peak at 1.2 minutes is THCA+2 present in the same standard; B: EIC of extracted plant parts infected by *A. tumefaciens* GV3101 with PT2 with native signal peptide.

Figure 19 shows a combined standard solution of CBGA and THCA (Figure A) and a solution after extracting from leave tissue of a tobacco plant and performing the enzyme assay for PTs from Cannabis (Figure B).

## 4.4 Heterologous expression in *Saccharomyces cerevisiae* Y05416

Additionally to the expression in a plant host the recombinant expression of *Cannabis* prenyltransferases in yeast was attempted. The putative prenyltransferases PT6 and PT3 were chosen to be expressed in *Saccharomyces cerevisiae* Y05416. This yeast strain was used due to being a Coq2 deletion mutant, which represents the UbiA prenyltransferase in yeast cells. PT6 is classified as UbiA prenyltransferase in Cannabis due to phylogenetic analysis and PT3 is the most promising candidate for a GOT prenyltransferase.

To simultaneously widen the capabilities of established methods at the lab two cloning strategies were performed in parallel for the expressions in yeast cells: Gibson Assembly® and homologous recombination. Besides the native plant signal peptides of the prenyltransferases a yeast signal peptide from the enzyme complex F1F0 ATP synthase was used.

### 4.4.1 Cloning via Gibson Assembly

For the cloning strategy using the Gibson Assembly® Kit, the signal peptide of the enzyme complex F1F0 ATP synthase of yeast was implemented in the forward-primer sequence for PT6 during Q5® High Fidelity PCR (Table 19). The signal peptide originates from the ATP2 gene (Table 40) and directs the enzyme complex to the inner membrane of the mitochondria (Takeda et al., 1985).

The primer sequences are shown in Table 18. The further intention was to reconstruct the whole plasmid including the signal peptide of the enzyme complex F1F0 ATP synthase. The aim of this strategy was to receive a plasmid carrying the yeast-own signal peptide for simple insertion of the gene of interest. For the reconstruction of the plasmid, two other primer sequences were designed, which are also represented in Table 18. At last PT3 was amplified from the cloning vector pGEM-T easy-PT3 and inserted into pDionysos carrying the yeast signal peptide (Table 18).

Since the ligation into the cloning vector was not completed for PT6 it was amplified from cDNA of leaves.

### 4.4.2 Cloning via homologous recombination

For the cloning strategy using homologous recombination the putative prenyltransferases PT6 and PT3 with their native signal peptides as well as Coq2 were ligated into pDionysos1.1. Coq2 was amplified from genomic DNA from *S. cerevisiae* CEN.PK2-1C (Table 2). Homologous recombination was performed with *S. cerevisiae* Y05416, which was plated on SD medium lacking uracil and leucine. Colonies developed after 5 days of incubation at 30°C. For plasmid multiplication *E. coli* DH5α was used. The successful integration of the correct sequence was verified by sequencing.

Since both cloning strategies were carried out simultaneously and the establishment of the constructs by homologous recombination was faster (considering sequentially correct clones) and cheaper comparing to Gibson Assembly® (New England Biolabs GmbH, Frankfurt am Main, Germany), the construct of PT3 with the signal peptide of F1F0 enzyme complex was also generated via homologous recombination.

### 4.4.3 Cultivation of PT6 and Coq2 in *Saccharomyces cerevisiae* Y05416

The evaluate the capabilities of the expression system the yeast UbiA prenyltransferase Coq2 was expressed in a Coq2 deletion variant. The UbiA activitiy of the membrane fraction was compared to the activity of the same deletion variant without inserted Coq2. As a control the experiment was also performed with the restored strain but without GPP (substrate).

The strain with restored Coq2 activity grew to a higher optical density than the deletion strain under identical conditions (see Figure 20). This could be due to the crucial role of Coq2 in providing ubiquinone, an electron transporter in the respiratory pathway.

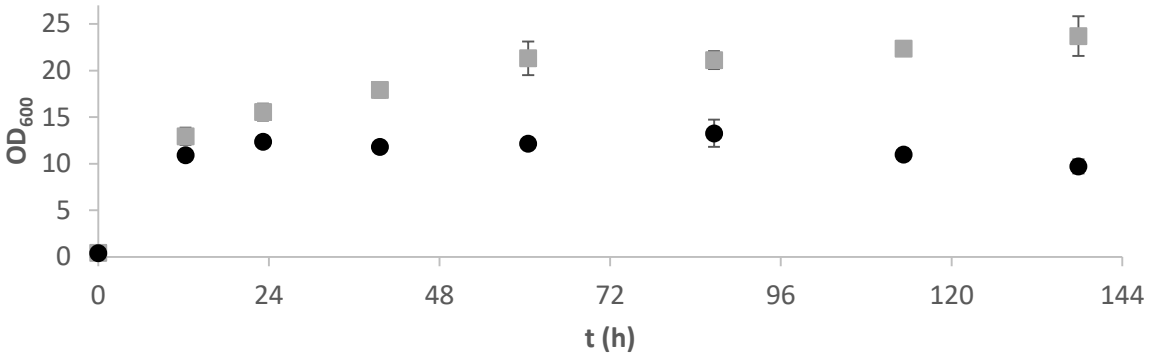



Figure 20: Growth curves: *Saccharomyces cerevisiae* Y05416 pDio-Coq2 (grey rectangle), *S. cerevisiae* Y05416 pDio (black dot); cultivation temperature: 30°C; YPD medium with 4% galactose.

The UbiA activity of the membrane fractions of the knock out strain with and without restored Coq2 was assayed (Figure 21). The data is normalized to the background activity measured when assaying the knock out strain with the empty vector (1 arbitrary unit). *Saccharomyces cerevisiae* Y05416 pDio-Coq2 shows strong heterologous activity after 23 hours of cultivation (approx. 100x background activity). This shows that the system is suitable for the expression of aromatic membrane-bound prenyltransferases.

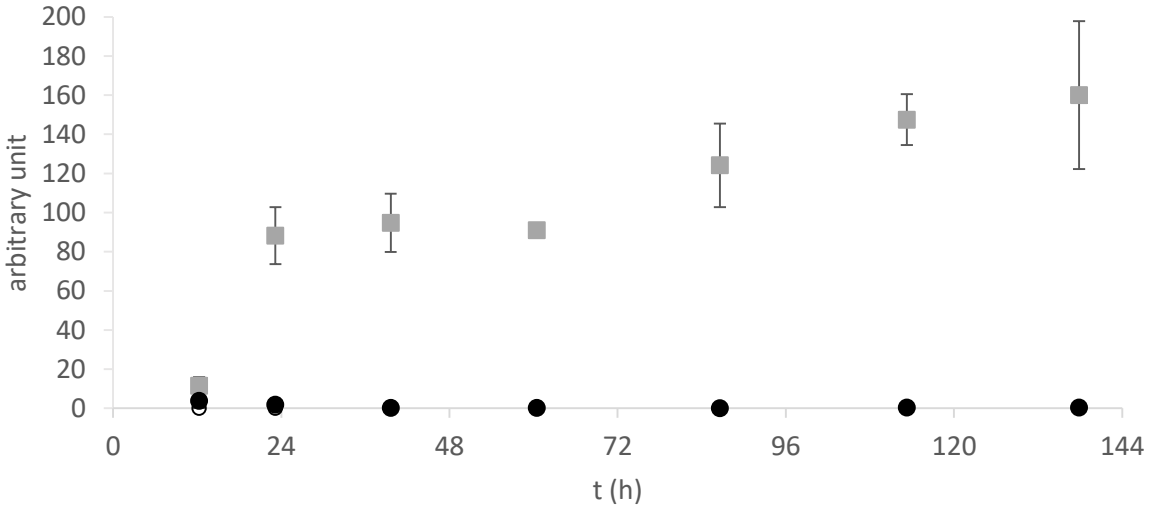



Figure 21: GHB production assay of S. cerevisiae Y05416 membrane fraction after cultivation; grey rectangles: pDio-Coq2 with GPP; black dots: pDio with GPP; white dots: pDio-Coq2 without GPP; white dots are almost entirely covered by black dots; one arbitrary unit (a.u.) represents the average measured background activity of cultivation with empty vector pDio; cultivation temperature: 30°C; YPD medium with 4% galactose.

Additional experiments with Coq2-enzyme solution after membrane fraction purification were performed. For this the protein solution was frozen at -18°C overnight with and without glycerol (50%, v/v). Both samples showed no significant difference in activity to the unfrozen material (Figure 67). This shows that the enzyme solution can be stored in the freezer at least overnight after protein purification.

Since Y05416 pDio-Coq2 confirmed Coq2 activity, the whole expression system was applied to the expression of the plant PTs PT3 and PT6 in yeast cells. PT6 was identified to be in the group of UbiA PTs and therefore a promising candidate to show the success of the heterologous expression system for plant proteins in yeast cells. PCR reaction was performed with cDNA as template (Table 15, Table 25, Figure 61). Homologous recombination was performed with the native signal peptides and Gibson Assembly with the signal peptide from F1F0 enzyme complex from yeast (section 4.4). Positive clones of PT6 showed one mutation in the signal peptide and three mutations in the mature protein, which can be explained by using different varieties of plants when identifying the *in silico* (Purple kush) sequence and in practice (Bediol). The mutations were as follows: V88I (signal peptide), D241S, D242S, H314R (mature protein).

PT6 was expressed with its native plant signal peptide as well as with a yeast signal peptide from the enzyme complex F1F0 ATP synthase (Takeda et al., 1985). The yeast signal peptide directs the enzyme to the inner membrane of mitochondria in *Saccharomyces*.

For this ligation the Gibson Assembly Kit was tested. Aliquots of the Gibson Assembly reaction were terminated after 15, 30 and 60 minutes and the success was verified on an agarose gel after transformation into *E. coli* DH5α cells. Approximately 20% of developing colonies carried an insert after 15 and 30 minutes of incubation, 40% carried an insert after 60 minutes. Sequencing showed, however, that none of the inserts were sequentially correct. In the end homologous recombination was used to build this construct.

Experiences with the heterologous expression of THCA-Synthase revealed that the expression was higher at a cultivation temperature of 18°C instead of 30°C (personal communication). On account of this the cultivations of *S. cerevisiae* Y05416 pDio and pDio-Coq2 were repeated at 18 °C. The production of GHB was again successfully demonstrated after performing the enzyme assay

(supplementary information

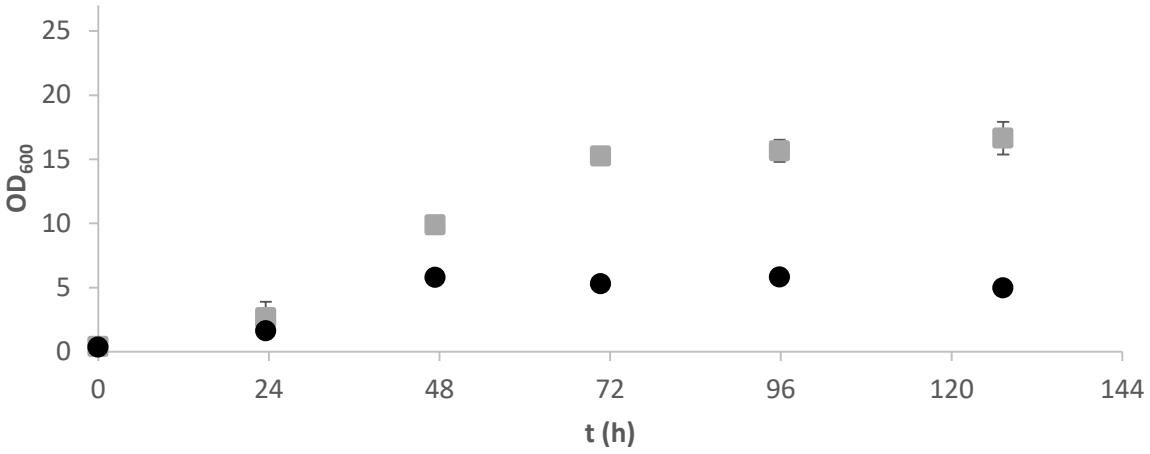



Figure 62 and Figure 63). Other possible advantages could lie in the slower growth of the host cell and therefore a larger time span for folding of the plant enzyme into the correct topology or to process posttranslational modifications.

To show the suitability of the expression system for aromatic plant PTs, PT6 was chosen for expression because based on the phylogenetic analysis it should be able to restore UbiA-functionality to the deletion strain like Coq2. PT6 was therefore expressed with a yeast signal peptide and with its native plant signal peptide.

*S. cerevisiae* Y05416 pDio-PT6ySP are the cell cultures with the signal peptide from the yeast enzyme complex F1F0 ATP Synthase and pDio-PT6nSP are the cell cultures with the native signal peptide. For comparison the growth curves of all four construct (*S. cerevisiae* Y05416 pDio, pDio-Coq2, pDio-PT6ySP, pDio-PT6nSP) at 18°C are shown in Figure 22.

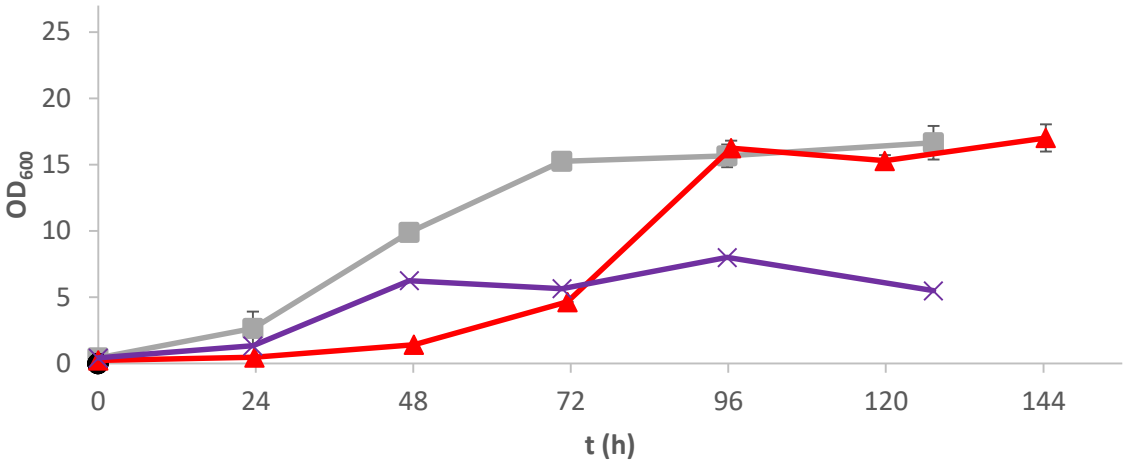



Figure 22: Growth curves: *S. cerevisiae* Y0541 with 4 different plasmids: pDio (black dots); pDio-Coq2 (grey rectangles); pDio-PT6ySP (red triangle; PT6 with signal peptide from yeast protein); pDio-PT6nSP (violet X; PT6 with native plant signal peptide); cultivation temperature: 18°C; YPD medium with 4% galactose.

The growth of *S. cerevisiae* Y05416 pDio-PT6nSP is comparable to the deletion strain carrying the empty vector. Apparently, while the expression of Coq2 in the deletion strain was able to increase biomass yield, the plant enzyme with the native plant signal peptide was not able to do the same. Comparing the growth curves of *S. cerevisiae* Y05416 pDio-Coq2 and pDio-PT6ySP shows approximately the same biomass yield. This indicates that the plant UbiA from *Cannabis* is able to replace the naturally occurring yeast UbiA. A prolonged lag phase is apparent for the culture with plant UbiA, the growth rate between these two cultures, however, appears comparable.

Comparing the UbiA activity in the membrane faction of all strains offers an explanation for the different growth behavior (Figure 23).

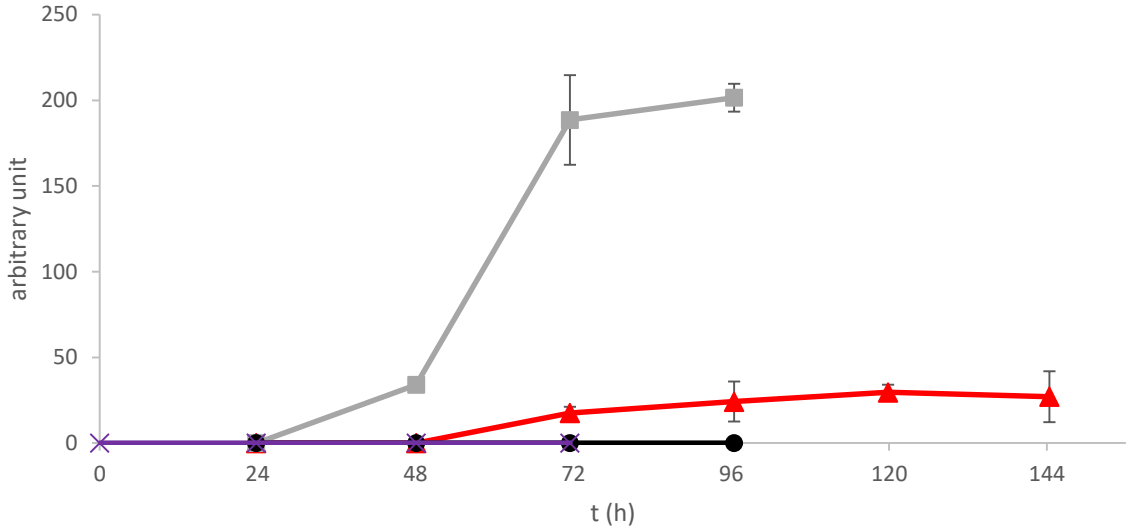



Figure 23: GHB production assay of *S. cerevisiae* Y05416 membrane fraction after cultivation; grey rectangles: pDio-Coq2; black dots: pDio; red triangles: pDio-PT6ySP; violet X: pDio-PT6nSP; cultivation temperature: 18°C; YPD medium with 4% galactose; MS data at t=127 h were not measured.

PT6 with its native signal peptide shows no UbiA activity comparable with the empty vector. UbiA activity was detected when using *S. cerevisiae* Y05416 pDio-PT6ySP. In comparison to the native yeast UbiA Coq2, however, the activity of PT6ySP is approximately one eighth. Reasons for the lower conversion rate could be a lower expression level of the plant enzyme in yeast cells or a lower activity of the expressed enzymes.

Even though the plant enzyme is expressed in lower amounts in the host cells, the amount of biomass indicates that the respiratory pathway is used for aerobic growth and therefore benefits the bio mass development. The prolonged lag-phase observed during cell growth could be due to the lower UbiA activity of PT6 compared to Coq2. Until sufficient ubiquinone is accumulated to support full utilization of the respiratory pathway quick growth is inhibited. Once enough ubiquinone has been produced growth rates are comparable.

The suitability of the expression system *S. cerevisiae* Y05416 pDio for the expression of aromatic plant prenyltransferases has been demonstrated.

### 4.4.4 Cultivation of PT3 in *Saccharomyces cerevisiae* Y05416

Due to phylogenetic analysis and the results from the expression profile, PT3 was considered as the most promising candidate as Geranylpyrophosphate:olivetolate geranyltransferase (GOT). Therefore it was expressed in *Saccharomyces cerevisiae* Y05416.

For the amplification via PCR the construct in the cloning vector pGEM-T easy was used, which already had been constructed for the transient expression in *Nicotiana benthamiana*. Analogous to the PT6 constructs PT3 was cloned into pDionysos with its native plant signal peptide and with the signal peptide from yeast F1F0 ATP synthase enzyme complex by homologous recombination. The cells were cultivated in complex medium with 4 % galactose at 18 °C. Afterwards the cells were disrupted and the membrane fraction was purified.

An activity assay for both PT3 constructs was performed with OA and GPP as substrates. The growth curve of *S. cerevisiae* Y05416 pDio-PT3ySP (PT3 with signal peptide from yeast protein) was plotted in comparison to the cell cultures with pDio-PT3nSP (PT3 with native plant signal peptide) (Figure 24).

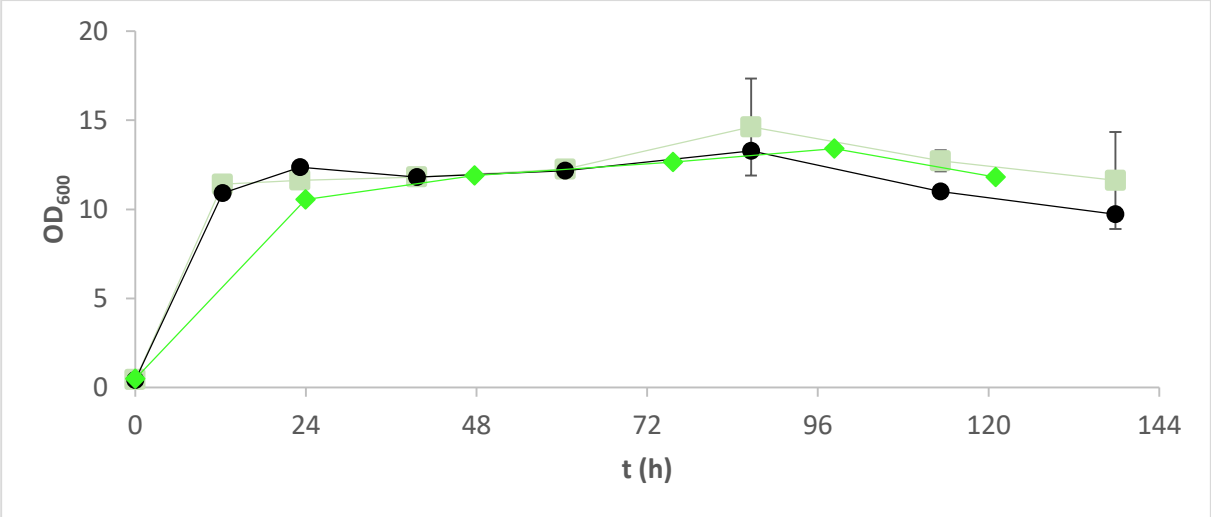



Figure 24: Growth curves *S. cerevisiae* Y05416: pDio-PT3nSP (light green rectangles); pDio-PT3ySP (grass green diamonds; PT3 with signal peptide from yeast protein); pDio (black dots); cultivation temperature: 18 °C; YPD medium with 4 % galactose.

Comparison of the growth curves of *S. cerevisiae* Y05416 pDio-PT3ySP and pDio-PT3nSP show similar growth to the empty vector. Hence there is no discernable metabolic stress through the expression of PT3.

Analysis of the enzyme assays did not show GOT activity. Because no activity could be detected the experiment was repeated in *S. cerevisiae* CEN.PK. Unfortunately still no activity was detected. As

already described in chapter 4.3.3 the used substrate olivetolic acid was wrongly synthesized. Therefore no statement can be made. The experiments should be repeated with the correct substrate.

# 5. Outlook

Further experiments should comprise the trouble-shooting of the PT3 expression. The expression with the extrinsic signal peptide was not working for PT3 while the expression worked for PT6 under the same expression conditions. After the laboratory work for this thesis it was discovered that the used olivetolic acid had the wrong structure. One of the hydroxyl groups was added to the wrong position. Hence the expression experiments should be repeated for PT3 with the extrinsic signal peptide.

Furthermore, the other prenyltransferases (PT1 respectively PT1p, PT2, PT4, PT5) should be expressed in *Saccharomyces cerevisiae* Y05416 as well. Since most of the remaining prenyltransferases could be classified in groups with known function, it should be easy to find out what substrates can be used. The heterologous expression of PT2 should also be performed. Since PT2 could not be categorized, various different substrates should be tested on this enzyme.

Additionally, I would recommend to investigate the usage of different carbon sources. Since Y05416 is not a Δgal strain, it uses the inducer also as carbon source. Hence it is possible that another sugar, like fructose, could be preferred to be used as carbon source and galactose is therefore always present in high concentration to enable induction of the system.

# 6. References


BAdams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., Mccombie, W.R., Venter, J.C., 1991. Complementary Sequencing : Expressed Sequence Tags and Human Genome Project 49, 1651–1656.

Akashi, T., Sasaki, K., Aoki, T., Ayabe, S., 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. doi:10.1104/pp.108.123679

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402. doi:10.1093/nar/25.17.3389

Andre, C.M., Hausman, J.-F., Guerriero, G., 2016. Cannabis sativa: the plant of the thousand and one molecules. Front. Plant Sci. 7, 1–17. doi:10.3389/fpls.2016.00019

Appendino, G., Gibbons, S., Giana, A., Pagani, A., Grassi, G., Stavri, M., Smith, E., Rahman, M.M., 2008. Antibacterial cannabinoids from Cannabis sativa: A structure-activity study. J. Nat. Prod. 71, 1427–1430. doi:10.1021/np8002673

Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408, 796–815. doi:10.1038/35048692

Ashby, M.N., Kutsunai, S.Y., Ackerman, S., Tzagoloff, A., Edwards, P.A., 1992. COQ2 is a candidate for the structural gene encoding para-hydroxybenzoate:polyprenyltransferase. J. Biol. Chem. 267, 4128–4136.

Bekel, T., Henckel, K., K??ster, H., Meyer, F., Mittard Runte, V., Neuweger, H., Paarmann, D., Rupp, O., Zakrzewski, M., P??hler, A., Stoye, J., Goesmann, A., 2009. The Sequence Analysis and Management System - SAMS-2.0: Data management and sequence analysis adapted to changing requirements from traditional sanger sequencing to ultrafast sequencing technologies. J. Biotechnol. 140, 3–12. doi:10.1016/j.jbiotec.2009.01.006

Ben Amar, M., 2006. Cannabinoids in medicine: A review of their therapeutic potential. J. Ethnopharmacol. 105, 1–25. doi:10.1016/j.jep.2006.02.001

Borrelli, F., Fasolino, I., Romano, B., Capasso, R., Maiello, F., Coppola, D., Orlando, P., Battista, G., Pagano, E., Di Marzo, V., Izzo, A.A., 2013. Beneficial effect of the non-psychotropic plant cannabinoid cannabigerol on experimental inflammatory bowel disease. Biochem. Pharmacol. 85, 1306–1316. doi:10.1016/j.bcp.2013.01.017

Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bairoch, A., 2007. UniProtKB/Swiss-Prot. Methods Mol. Biol. 406, 89–112.

Bradshaw, H.B., Walker, J.M., 2005. The expanding field of cannabimimetic and related lipid mediators. Br. J. Pharmacol. 144, 459–65. doi:10.1038/sj.bjp.0706093

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

Burstein, S., 2015. Cannabidiol (CBD) and its analogs: A review of their effects on inflammation. Bioorganic Med. Chem. 23, 1377–1385. doi:10.1016/j.bmc.2015.01.059

Burstein, S., 2008. The elmiric acids: Biologically active anandamide analogs. Neuropharmacology 55, 1259–1264. doi:10.1016/j.neuropharm.2007.11.011

Burstein, S.H., Zurier, R.B., 2009. Cannabinoids, endocannabinoids, and related analogs in inflammation. AAPS J. 11, 109–19. doi:10.1208/s12248-009-9084-5

Chiurchiù, V., Lanuti, M., De Bardi, M., Battistini, L., Maccarrone, M., 2014. The differential characterization of GPR55 receptor in human peripheral blood reveals a distinctive expression in monocytes and NK cells and a proinflammatory role in these innate cells. Int. Immunol. 27, 153–160. doi:10.1093/intimm/dxu097

Colasanti, B.K., 1990. A comparison of the ocular and central effects of delta 9-tetrahydrocannabinol and cannabigerol. J. Ocul. Pharmacol. 6, 259–269.

Colasanti, B.K., Craig, C.R., Allara, R.D., 1984. Intraocular pressure, ocular toxicity and neurotoxicity after administration of cannabinol or cannabigerol. Exp. Eye Res. 39, 251–259. doi:10.1016/0014-4835(84)90013-7

Collakova, E., Dellapenna, D., 2001. Isolation and Functional Analysis of Homogentisate Phytyltransferase from Synechocystis sp . PCC 6803. Plant Physiol. 127, 1113–1124. doi:10.1104/pp.010421.1

Dai, X., Wang, G., Yang, D.S., Tang, Y., Broun, P., Marks, M.D., Sumner, L.W., Dixon, R.A., Zhao, P.X., 2010. TrichOME: A Comparative Omics Database for Plant Trichomes. Plant Physiol. 152, 44–54. doi:10.1104/pp.109.145813

De Petrocellis, L., Ligresti, A., Moriello, A.S., Allar??, M., Bisogno, T., Petrosino, S., Stott, C.G., Di Marzo, V., 2011. Effects of cannabinoids and cannabinoid-enriched Cannabis extracts on TRP channels and endocannabinoid metabolic enzymes. Br. J. Pharmacol. 163, 1479–1494. doi:10.1111/bph.2011.163.issue-7

Delong, G.T., Wolf, C.E., Poklis, A., Lichtman, A.H., 2010. Pharmacological evaluation of the natural constituent of Cannabis sativa, Cannabichromene and its modulation by delta-9-THC. Drug Alcohol Depend. 112, 126–133. doi:10.1016/j.drugalcdep.2010.05.019.Pharmacological

Deutsch, D.G., Ueda, N., Yamamoto, S., 2000. The fatty acid amide hydrolase (FAAH). Chem. Phys. Lipids 108, 107–121. doi:10.1016/S0009-3084(00)00190-0

Devane, W.A., Dysarz III, F.A., Johnson, M.R., Melvin, L.S., Howlett, A.C., 1988. Determination and Characterization of a Cannabinoid Receptor in Rat Brain. Mol. Pharmacol. 34, 605–613.

Dhillonz, D.S., Brown, S.A., 1976. and Characterization of Dimethylallylpyrophosphate : Umbelliferone Dimethylallyltransferase from Ruta graveolensl. Arch. Biochem. Biophys.

Eddy, S., 1998. Profile hidden Markov models. Bioinformatics 14, 755–763. doi:btb114 [pii]

Eddy, S.R., 2004. What is a hidden Markov model? Nat. Biotechnol. 22, 1315–1316. doi:10.1038/nbt1004-1315

Eisohly, H.N., Turner, C.E., Clark, a M., Eisohly, M. a, 1982. Synthesis and antimicrobial activities of certain cannabichromene and cannabigerol related compounds. J. Pharm. Sci. 71, 1319–1323. doi:10.1002/jps.2600711204

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. doi:10.1038/nprot.2007.131

Emanuelsson, O., Nielsen, H., von Heijne, G., 1999. ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Sci. 8, 978–984.

doi:10.1110/ps.8.5.978

Englund, A., Stone, J.M., Morrison, P.D., 2012. Cannabis in the Arm: What can we learn from intravenous Cannabinoid studies? Curr. Pharm. Des. 18, 4906–4914.

Entian, K.-D., Kötter, P., 2007. 25 Yeast Genetic Strain and Plasmid Collections. Methods Microbiol. 36, 629–666.

Ewing, B., Hillier, L., Wendl, M., Green, P., 1998. Base-calling of automated sequencer traces usingPhred. I. Accuracy assessment. Genome Res. 175–185. doi:10.1101/gr.8.3.175

Fellermeier, M., Zenk, M.H., 1998. Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol. FEBS Lett. 427, 283–285. doi:10.1016/S0014-5793(98)00450-5

Fernandez, O., 2016. THC:CBD in Daily Practice: Available Data from UK, Germany and Spain. Eur. Neurol. 75, 1–3. doi:10.1159/000444234

Finlay, D.B., Joseph, W.R., Grimsey, N.L., Glass, M., 2016. GPR18 undergoes a high degree of constitutive trafficking but is unresponsive to N-Arachidonoyl Glycine. PeerJ 4, e1835. doi:10.7717/peerj.1835

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. doi:10.1016/j.phytochem.2010.10.001

Forsgren, M., Attersand, A., Lake, S., Grünler, J., Swiezewska, 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. 526, 519–526.

Gao, Z.M., Li, C.L., Peng, Z.H., 2011. Generation and analysis of expressed sequence tags from a normalized cDNA library of young leaf from Ma bamboo (Dendrocalamus latiflorus Munro). Plant Cell Rep. 30, 2045–2057. doi:10.1007/s00299-011-1112-0

Geer, L.Y., Domrachev, M., Lipman, D.J., Bryant, S.H., 2002. CDART: Protein homology by domain architecture. Genome Res. 12, 1619–1623. doi:10.1101/gr.278202

Gerard, C., Mollereau, C., Vassart, G., Parmentier, M., 1990. Nucleotide cDNA sequence of a human cannabinoid receptor. Nucleic Acids Res. 18, 7142.

Gerard, C.M., Mollereau, C., Vassart, G., Parmentier, M., 1991. Molecular cloning of a human cannabinoid receptor which is also expressed in testis. Biochem. J. 279, 129–134.

Gibson, D.G., 2011. Enzymatic assembly of overlapping DNA fragments, in: Methods in Enzymology. pp. 349–361. doi:10.1016/B978-0-12-385120-8.00015-2

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C. a, Smith, H.O., Iii, C.A.H., America, N., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–5. doi:10.1038/nmeth.1318

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. doi:10.1038/nprot.2007.17

Giuffrida, A., Beltramo, M., Piomelli, D., 2001. Mechanisms of endocannabinoid inactivation: biochemistry and pharmacology. J. Pharmacol. Exp. Ther. 298, 7–14.

Giuffrida, A., Piomelli, D., 2000. The endocannabinoid system: A physiological perspective on its role

in psychomotor control. Chem. Phys. Lipids 108, 151–158. doi:10.1016/S0009-3084(00)00193-6

Gleba, Y., Klimyuk, V., Marillonnet, S., 2007. Viral vectors for the expression of proteins in plants. Curr. Opin. Biotechnol. 18, 134–141. doi:10.1016/j.copbio.2007.03.002

Gleba, Y., Klimyuk, V., Marillonnet, S., 2005. Magnifection - A new platform for expressing recombinant vaccines in plants. Vaccine 23, 2042–2048. doi:10.1016/j.vaccine.2005.01.006

Goff, S.A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchinson, D., Martin, C., Katagiri, F., Lange, B.M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J., Miguel, T., Paszkowski, U., Zhang, S., Colbert, M., Sun, W., Chen, L., Cooper, B., Park, S., Wood, T.C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Mary, M.R., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A., Briggs, S., 2002. A Draft Sequence of the Rice Genome ( Oryza sativa L . ssp .). Science (80-. ). 296, 92–100. doi:10.1126/science.1068037

Gordon, D., Abajian, C., Green, P., 1998. Consed: A graphical tool for sequence finishing. Genome Res. 8, 195–202. doi:10.1101/gr.8.3.195

Gromiha, M.M., Yabuki, Y., Suresh, M.X., Thangakani, A.M., Suwa, M., Fukui, K., 2009. TMFunction: Database for functional residues in membrane proteins. Nucleic Acids Res. 37, 201–204. doi:10.1093/nar/gkn672

Hamerski, D., Schmitt, D., Matern, U., 1990. Induction of two prenyltransferases for the accumulation of coumarin phytoalexins in elicitor-treated Ammi majus cell suspension cultures. Phytochemistry 29, 1131–1135. doi:10.1016/0031-9422(90)85417-E

Hanus, L.O., 2009. Pharmacological and Therapeutic Secrets of Plant and Brain ( Endo ) Cannabinoids. Med. Res. Rev. 29, 213–271. doi:10.1002/med

Happyana, N., Agnolet, S., Muntendam, R., Van Dam, A., Schneider, B., Kayser, O., 2013. Analysis of cannabinoids in laser-microdissected trichomes of medicinal Cannabis sativa using LCMS and cryogenic NMR. Phytochemistry 87, 51–59. doi:10.1016/j.phytochem.2012.11.001

Hazekamp, A., Grotenhermen, F., 2010. Review on clinical studies with cannabis and cannabinoids 2005-2009. Mult. Scler. 5, 1–21.

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

Heide, L., Nishioka, N., Fukui, H., Tabata, M., 1989. Enzymatic regulation of shikonin biosynthesis in Lithospermum erythrorhizon cell cultures. Phytochemistry 28, 1873–1877. doi:10.1016/S0031-9422(00)97877-4

Hill, A.J., Williams, C.M., Whalley, B.J., Stephens, G.J., 2012. Phytocannabinoids as novel therapeutic agents in CNS disorders. Pharmacol. Ther. 133, 79–97. doi:10.1016/j.pharmthera.2011.09.002

Hillig, K.W., Mahlberg, P.G., 2004. A chemotaxonomic analysis of cannabinoid variation in Cannabis (Cannabaceae). Am. J. Bot. 91, 966–975. doi:10.3732/ajb.91.6.966

Hoekema, a., Hirsch, P.R., Hooykaas, P.J.J., Schilperoort, R. a., 1983. A binary plant vector strategy based on separation of vir- and T-region of the Agrobacterium tumefaciens Ti-plasmid. Nature. doi:10.1038/303179a0

Howlett, A.C., 1984. Inhibition of Neuroblastoma Adenylate Cyclase by Cannabinoid and Nantradol compounds. Life Sci. 35, 1803–1810.

Howlett, A.C., Barth, F., Bonner, T.I., Cabral, G., Casellas, P., Devane, W.A., Felder, C.C., Herkenham, M., Mackie, K., Martin, B.R., Mechoulam, R., Pertwee, R.G., 2002. International Union of Pharmacology . XXVII . Classification of Cannabinoid Receptors. Pharmacol. Rev. 54, 161–202.

http://www.pewresearch.org [WWW Document], 2016.

https://www.uni-giessen.de/fbz/fb08/Inst/bioinformatik/software/gendb [WWW Document], 2016.

Huang, H., Levin, E.J., Liu, S., Bai, Y., Lockless, S.W., Zhou, M., 2014. Structure of a membrane-embedded prenyltransferase homologous to UBIAD1. PLoS Biol. 12, e1001911. doi:10.1371/journal.pbio.1001911

Huang, H., Scherman, M.S., Haeze, W.D., Vereecke, D., Holsters, M., Crick, D.C., Mcneil, M.R., 2005. Identification and Active Expression of the Mycobacterium tuberculosis Gene Encoding 5-Phospho- ␣ - D -ribose-1-diphosphate : Decaprenyl-phosphate 5-Phosphoribosyltransferase , the First Enzyme Committed to Decaprenylphosphoryl- D -arabinose Synthesis *. J. Biol. Chem. 280, 24539–24543. doi:10.1074/jbc.M504068200

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

Jones, D., Taylor, W., Thornton, J., 1992. the Rapid Generation of Mutation Data Matrices From Protein Sequences. Comput. Appl. Biosci. 8, 275–282.

Jones, D.T., Taylor, W.R., Thornton, J.M., 1994. A Model Recognition Approach to the Prediction of All-Helical Membrane Protein Structure and Topology. Biochemistry 33, 3038–3049. doi:10.1021/bi00176a037

Jost, M., Zocher, G., Tarcz, S., Matuschek, M., Xie, X., 2010. Structure - Function Analysis of an Enzymatic Prenyl Transfer Specificity. J. Am. Chem. Soc. 17849–17858. doi:10.1021/ja106817c

Kapur, A., Zhao, P., Sharir, H., Bai, Y., Caron, M.G., Barak, L.S., Abood, M.E., 2009. Atypical responsiveness of the orphan receptor GPR55 to cannabinoid ligands. J. Biol. Chem. 284, 29817–29827. doi:10.1074/jbc.M109.050187

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066. doi:10.1093/nar/gkf436

Kawamukai, M., 2002. REVIEW and Novel Roles of Ubiquinone. Rev. Lit. Arts Am. 94, 511–517.

Kim, E.S., Mahlberg, P.G., 2003. Secretory vesicle formation in the secretory cavity of glandular trichomes of Cannabis sativa L. (Cannabaceae). Mol. Cells 15, 387–395. doi:10.1142/S0578563403000816

Kim, E.S., Mahlberg, P.G., 1997. Plastid development in disc cells of glandular trichomes of Cannabis (Cannabaceae). Mol. Cells 7, 352.

Koncz, C., Schell, J., 1986. The promoter of TL-DNA gene 5 controls the tissue-specific expression of chimaeric genes carried by a novel type of Agrobacterium binary vector. MGG Mol. Gen. Genet. 204, 383–396. doi:10.1007/BF00331014

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. doi: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. Bioorganic Med. Chem. 16, 8117–8126. doi:10.1016/j.bmc.2008.07.052

Kunos, G., Járai, Z., Varga, K., Liu, J., Wang, L., Wagner, J. a, 2000. Cardiovascular effects of endocannabinoids--the plot thickens. Prostaglandins Other Lipid Mediat. 61, 71–84. doi:10.1016/S0090-6980(00)00056-3

Kuzuyama, T., Noel, J.P., Richard, S.B., 2005. Structural basis for the promiscuous biosynthetic prenylation of aromatic natural products. Nature 435, 983–7. doi:10.1038/nature03668

Lichtman, A.H., Varvel, S.A., Martin, B.R., 2002. Endocannabinoids in cognition and dependence. Prostaglandins. Leukot. Essent. Fatty Acids 66, 269–285. doi:10.1054/plef.2001.0351

Ling, P., Wang, M., Chen, X., Campbell, K.G., 2007. Construction and characterization of a full-length cDNA library for the wheat stripe rust pathogen (Puccinia striiformis f. sp. tritici). BMC Genomics 8, 145. doi:10.1186/1471-2164-8-145

Liu, B., Song, S., Jones, P.M., Persaud, S.J., 2015. GPR55: From orphan to metabolic regulator? Pharmacol. Ther. 145, 35–42. doi:10.1016/j.pharmthera.2014.06.007

Liu, Y., Engelman, D.M., Gerstein, M., 2002. Genomic analysis of membrane protein families: abundance and conserved motifs. Genome Biol. 3, research0054. doi:10.1186/gb-2002-3-10-research0054

Long, S.B., Casey, P.J., Beese, L.S., 2002. Reaction path of protein farnesyltransferase at atomic resolution. Nature 419, 645–650. doi:10.1038/nature00986

Long, S.H., Deng, X., Wang, Y.F., Li, X., Qiao, R.Q., Qiu, C.S., Guo, Y., Hao, D.M., Jia, W.Q., Chen, X.B., 2012. Analysis of 2,297 expressed sequence tags (ESTs) from a cDNA library of flax (Linum ustitatissimum L.) bark tissue. Mol. Biol. Rep. 39, 6289–6296. doi:10.1007/s11033-012-1450-1

Lücker, J., El Tamer, M.K., Schwab, W., Verstappen, F.W.A., Van Der Plas, L.H.W., Bouwmeester, H.J., Verhoeven, H.A., 2002. Monoterpene biosynthesis in lemon (Citrus limon) cDNA isolation and functional analysis of four monoterpene synthases. Eur. J. Biochem. 269, 3160–3171. doi:10.1046/j.1432-1033.2002.02985.x

Maccarrone, M., Falciglia, K., Di Rienzo, M., Finazzi-Agrò, a, 2002. Endocannabinoids, hormone-cytokine networks and human fertility. Prostaglandins. Leukot. Essent. Fatty Acids 66, 309–17. doi:10.1054/plef.2001.0354

Maione, S., Piscitelli, F., Gatta, L., Vita, D., Petrocellis, L. De, Palazzo, E., de Novellis, V., Di Marzo, V., 2011. cannabinoids modulate the descending pathway of antinociception in anaesthetized rats through several mechanisms of action. Br. J. Pharmacol. 162, 584–596. doi:10.1111/j.1476-5381.2010.01063.x

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y., Chen, Z., Dewell, B., Du, L., Fierro, J.M., Gomes, X. V, Goodwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Maria, L.I., Jarvie, T.P., Jirage, K.B., Kim, J., Knight, J.R., Lanza, R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Kenton, L., Lu, H., Makhijani, V.B., Mcdade, K.E., Mckenna, M.P., Myers, W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Greg, A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Richard, F., Rothberg, J.M., 2005. Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. Nat. Biotechnol. 437, 376–380. doi:10.1038/nature03959.Copyright

Marillonnet, S., Giritch, A., Gils, M., Kandzia, R., Klimyuk, V., Gleba, Y., 2004. In planta engineering of viral RNA replicons: efficient assembly by recombination of DNA modules delivered by Agrobacterium. Proc. Natl. Acad. Sci. U. S. A. 101, 6852–7. doi:10.1073/pnas.0400149101

Marillonnet, S., Thoeringer, C., Kandzia, R., Klimyuk, V., Gleba, Y., 2005. Systemic Agrobacterium tumefaciens-mediated transfection of viral replicons for efficient transient expression in plants. Nat. Biotechnol. 23, 718–723. doi:10.1038/nbt1094

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 Delta9-tetrahydrocannabinol biosynthesis in Cannabis sativa. J. Exp. Bot. 60, 3715–3726. doi:10.1093/jxb/erp210

Matsuda, L.A., Lolait, S.J., Brownstein, M.J., Young, A.C., Bonner, T.I., 1990. Structure of a Cannabinoid receptor and functional expression of the cloned cDNA. Nature 346, 561–564.

Mayer, K., Schüller, C., Wambutt, R., Murphy, G., Volckaert, G., Pohl, T., Düsterhöft, a, Stiekema, W., Entian, K.D., Terryn, N., Harris, B., Ansorge, W., Brandt, P., Grivell, L., Rieger, M., Weichselgartner, M., de Simone, V., Obermaier, B., Mache, R., Müller, M., Kreis, M., Delseny, M., Puigdomenech, P., Watson, M., Schmidtheini, T., Reichert, B., Portatelle, D., Perez-Alonso, M., Boutry, M., Bancroft, I., Vos, P., Hoheisel, J., Zimmermann, W., Wedler, H., Ridley, P., Langham, S. a, McCullagh, B., Bilham, L., Robben, J., Van der Schueren, J., Grymonprez, B., Chuang, Y.J., Vandenbussche, F., Braeken, M., Weltjens, I., Voet, M., Bastiaens, I., Aert, R., Defoor, E., Weitzenegger, T., Bothe, G., Ramsperger, U., Hilbert, H., Braun, M., Holzer, E., Brandt, a, Peters, S., van Staveren, M., Dirske, W., Mooijman, P., Klein Lankhorst, R., Rose, M., Hauf, J., Kötter, P., Berneiser, S., Hempel, S., Feldpausch, M., Lamberth, S., Van den Daele, H., De Keyser, a, Buysshaert, C., Gielen, J., Villarroel, R., De Clercq, R., Van Montagu, M., Rogers, J., Cronin, a, Quail, M., Bray-Allen, S., Clark, L., Doggett, J., Hall, S., Kay, M., Lennard, N., McLay, K., Mayes, R., Pettett, a, Rajandream, M. a, Lyne, M., Benes, V., Rechmann, S., Borkova, D., Blöcker, H., Scharfe, M., Grimm, M., Löhnert, T.H., Dose, S., de Haan, M., Maarse, a, Schäfer, M., Müller-Auer, S., Gabel, C., Fuchs, M., Fartmann, B., Granderath, K., Dauner, D., Herzl, a, Neumann, S., Argiriou, a, Vitale, D., Liguori, R., Piravandi, E., Massenet, O., Quigley, F., Clabauld, G., Mündlein, a, Felber, R., Schnabl, S., Hiller, R., Schmidt, W., Lecharny, a, Aubourg, S., Chefdor, F., Cooke, R., Berger, C., Montfort, a, Casacuberta, E., Gibbons, T., Weber, N., Vandenbol, M., Bargues, M., Terol, J., Torres, a, Perez-Perez, a, Purnelle, B., Bent, E., Johnson, S., Tacon, D., Jesse, T., Heijnen, L., Schwarz, S., Scholler, P., Heber, S., Francs, P., Bielke, C., Frishman, D., Haase, D., Lemcke, K., Mewes, H.W., Stocker, S., Zaccaria, P., Bevan, M., Wilson, R.K., de la Bastide, M., Habermann, K., Parnell, L., Dedhia, N., Gnoj, L., Schutz, K., Huang, E., Spiegel, L., Sehkon, M., Murray, J., Sheet, P., Cordes, M., Abu-Threideh, J., Stoneking, T., Kalicki, J., Graves, T., Harmon, G., Edwards, J., Latreille, P., Courtney, L., Cloud, J., Abbott, a, Scott, K., Johnson, D., Minx, P., Bentley, D., Fulton, B., Miller, N., Greco, T., Kemp, K., Kramer, J., Fulton, L., Mardis, E., Dante, M., Pepin, K., Hillier, L., Nelson, J., Spieth, J., Ryan, E., Andrews, S., Geisel, C., Layman, D., Du, H., Ali, J., Berghoff, a, Jones, K., Drone, K., Cotton, M., Joshu, C., Antonoiu, B., Zidanic, M., Strong, C., Sun, H., Lamar, B., Yordan, C., Ma, P., Zhong, J., Preston, R., Vil, D., Shekher, M., Matero, a, Shah, R., Swaby, I.K., O'Shaughnessy, a, Rodriguez, M., Hoffmann, J., Till, S., Granat, S., Shohdy, N., Hasegawa, a, Hameed, a, Lodhi, M., Johnson, a, Chen, E., Marra, M., Martienssen, R., McCombie, W.R., 1999. Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana. Nature 402, 769–777. doi:10.1038/47134

Medicinal Genomics Corporation, 2015. www.medicinalgenomics.com [WWW Document].

Meyer, F., Goesmann, A., McHardy, A.C., Bartels, D., Bekel, T., Clausen, J., Kalinowski, J., Linke, B., Rupp, O., Giegerich, R., P??hler, A., 2003. GenDB - An open source genome annotation system for prokaryote genomes. Nucleic Acids Res. 31, 2187–2195. doi:10.1093/nar/gkg312

Ministry of Health, Welfare and Sports of the Netherlands, 1p., 2011. Dutch Association for Legal Cannabis and its Constituents as Medicine [WWW Document].

Miyawaki, K., Tarkowski, P., Matsumoto-Kitano, M., Kato, T., Sato, S., Tarkowska, D., Tabata, S.,

Sandberg, G., Kakimoto, T., 2006. Roles of Arabidopsis ATP/ADP isopentenyltransferases and tRNA isopentenyltransferases in cytokinin biosynthesis. Proc. Natl. Acad. Sci. U. S. A. 103, 16598–603. doi:10.1073/pnas.0603522103

Morimoto, S., Komatsu, K., Taura, F., Shoyama, Y., 1998. PURIFICATION AND CHARACTERIZATION OF CANNABICHROMENIC ACID SYNTHASE FROM CANNABIS SATIVA. Phytochemistry 49, 1525–1529.

Munro, S., Thomas, K.L., Abu-Shaar, M., 1993. Molecular characterization of a peripheral receptor for cannabinoids. Nature 365, 61–65.

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. 280, 2572–80. doi: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, 231–41. doi:10.1042/BJ20081968

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. 47, 581–590. doi:10.1093/pcp/pcj025

Okada, K., Ohara, K., Yazaki, K., Nozaki, K., Uchida, N., Kawamukai, M., Nojiri, H., Yamane, H., 2004. The AtPPT1 gene encoding 4-hydroxybenzoate polyprenyl diphosphate transferase in ubiquinone biosynthesis is required for embryo development in Arabidopsis thaliana. Plant Mol. Biol. 55, 567–577. doi:10.1007/s11103-004-1298-4

Page, J., Boubakir, Z., 2011. Aromatic Prenyltransferase from Cannabis. PCT/CA2010/001222.

Payandeh, J., Fujihashi, M., Gillon, W., Pai, E.F., 2006. The crystal structure of (S)-3-O-geranylgeranylglyceryl phosphate synthase reveals an ancient fold for an ancient enzyme. J. Biol. Chem. 281, 6070–6078. doi:10.1074/jbc.M509377200

Pedersen, J.S., Hein, J., 2003. Gene finding with a hidden Markov model of genome structure and evolution. Bioinformatics 19, 219.

Petitet, F., Donlan, M., Michel, A., 2006. GPR55 as a new cannabinoid receptor: Still a long way to prove it. Chem. Biol. Drug Des. 67, 252–253. doi:10.1111/j.1747-0285.2006.00370.x

Pharmacology, C., 2004. Marinol (Dronabinol).

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

Qin, Y., Verdegaal, E.M.E., Siderius, M., Bebelman, J.P., Smit, M.J., Leurs, R., Willemze, R., Tensen, C.P., Osanto, S., 2010. Quantitative expression profiling of G-protein-coupled receptors (GPCRs) in metastatic melanoma: The constitutively active orphan GPCR GPR18 as novel drug target. Pigment Cell Melanoma Res. 24, 207–218. doi:10.1111/j.1755-148X.2010.00781.x

Quackenbush, J., Liang, F., Holt, I., Pertea, G., Upton, J., 2000. The TIGR gene indices: reconstruction and representation of expressed gene sequences. Nucleic Acids Res. 28, 141–5. doi:10.1093/nar/28.1.141

Rupp, O., Becker, J., Brinkrolf, K., Timmermann, C., Borth, N., Pühler, A., Noll, T., Goesmann, A., 2014. Construction of a public CHO cell line transcript database using versatile bioinformatics analysis pipelines. PLoS One 9. doi:10.1371/journal.pone.0085568

Sadre, R., Gruber, J., Frentzen, M., 2006. Characterization of homogentisate prenyltransferases involved in plastoquinone-9 and tocochromanol biosynthesis. FEBS Lett. 580, 5357–5362. doi:10.1016/j.febslet.2006.09.002

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. doi:10.1104/pp.107.110544

Sawzdargo, M., Nguyen, T., Lee, D.K., Lynch, K.R., Cheng, R., Heng, H.H.Q., George, S.R., O'Dowd, B.F., 1999. Identification and cloning of three novel human G protein-coupled receptor genes GPR52, ??GPR53 and GPR55: GPR55 is extensively expressed in human brain. Mol. Brain Res. 64, 193–198. doi:10.1016/S0169-328X(98)00277-0

Schicho, R., Storr, M., 2012. A potential role for GPR55 in gastrointestinal functions. Curr. Opin. Pharmacol. 12, 653–658. doi:10.1016/j.coph.2012.09.009

Schilmiller, A.L., Last, R.L., Pichersky, E., 2008. Harnessing plant trichome biochemistry for the production of useful compounds. Plant J. 54, 702–711. doi:10.1111/j.1365-313X.2008.03432.x

Schledz, M., Seidler, A., Beyer, P., Neuhaus, G., Gmbh, G.P., Freiburg, D.-, 2001. A novel phytyltransferase from Synechocystis sp . PCC 6803 involved in tocopherol biosynthesis. FEBS 499, 15–20.

Schuel, H., Burkman, L.J., Lippes, J., Crickard, K., Forester, E., Piomelli, D., Giuffrida, A., 2002. N-Acylethanolamines in human reproductive fluids. Chem. Phys. Lipids 121, 211–227. doi:10.1016/S0009-3084(02)00158-5

Sharma, V., Sarkar, I.N., 2013. Bioinformatics opportunities for identification and study of medicinal plants. Brief. Bioinform. 14, 238–250. doi:10.1093/bib/bbs021

Shcheglov, A.S., Zhulidov, P.A., Bogdanova, E.A., Shagin, D.A., 2007. Normalization of cDNA libraries. Nucleic Acids Hybrid. Mod. Appl. 97–124. doi:10.1007/978-1-4020-6040-3_5

Siebert, M., Bechthold, A., Melzer, M., May, U., Berger, U., Schr??der, G., Schr??der, J., Severin, K., Heide, L., 1992. Ubiquinone biosynthesis Cloning of the genes coding for chorismate pyruvate-lyase and 4-hydroxybenzoate octaprenyl transferase from Escherichia coli. FEBS Lett. 307, 347–350. doi:10.1016/0014-5793(92)80710-X

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, 1578–1582. doi:10.1093/pcp/pci166

Somerville, C., Somerville, S., 1999. Plant functional genomics. Science 285, 380–383. doi:10.1007/978-1-4939-2444-8

Sonnhammer, E.L., von Heijne, G., Krogh, A., 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proceedings 6, 175–182.

Staden, R., Beal, K.F., Bonfield, J.K., 2000. The Staden package, 1998. Methods Mol. Biol. 132, 115–130.

Suvarna, K., Stevenson, D., Meganathan, R., Hudspeth, M.E.S., 1998. Menaquinone ( Vitamin K 2 ) Biosynthesis : Localization and Characterization of the menA Gene from Escherichia coli. J. Bacteriol. 180, 2782–2787.

Suzuki, K., Ueda, M., Yuasa, M., Nakagawa, T., Kawamukai, M., Matsuda, H., 1994. Evidence that Escherichia coli ubiA product is a functional homolog of yeast COQ2, and the regulation of ubiA

gene expression. Biosci. Biotechnol. Biochem. 58, 1814–9. doi:10.1271/bbb.58.1814

Swiezewska, E., Dallner, G., Andersson, B., Ernster, L., 1993. Biosynthesis of ubiquinone and plastoquinone in the endoplasmic reticulum-Golgi membranes of spinach leaves. J. Biol. Chem. 268, 1494–1499.

Tahara, S., Ibrahim, R.K., 1995. Prenylated isoflavonoids-An update. Phytochemistry 38, 1073–1094. doi:10.1016/0031-9422(94)00788-U

Takahashi, S., Takagi, H., Toyoda, A., Uramoto, M., Nogawa, T., Ueki, M., Sakaki, Y., Osada, H., 2010. Biochemical characterization of a novel indole prenyltransferase from streptomyces sp. SN-593. J. Bacteriol. 192, 2839–2851. doi:10.1128/JB.01557-09

Takeda, M., Vassarotti, A., Douglas, M.G., 1985. Nuclear genes coding the yeast mitochondrial adenosine triphosphatase complex. Primary sequence analysis of ATP2 encoding the F1-ATPase ??-subunit precursor. J. Biol. Chem. 260, 15458–15465.

Takenouchi, R., Inoue, K., Kambe, Y., Miyata, A., 2012. N-arachidonoyl glycine induces macrophage apoptosis via GPR18. Biochem. Biophys. Res. Commun. 418, 366–371. doi:10.1016/j.bbrc.2012.01.027

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739. doi:10.1093/molbev/msr121

Taura, F., Dono, E., Sirikantaramas, S., Yoshimura, K., Shoyama, Y., Morimoto, S., 2007. Production of Delta1-tetrahydrocannabinolic acid by the biosynthetic enzyme secreted from transgenic Pichia pastoris. Biochem. Biophys. Res. Commun. 361, 675–680. doi: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.

Taylor, R.G., Walker, D.C., Mclnnes, R.R., 1993. E.coli host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. Nucleic Acids Res. 21, 1677–1678.

The Office of Medicinal Cannabis [WWW Document], 2016. . https://www.cannabisbureau.nl/.

Toyota, M., Kinugawa, T., Asakawa, Y., 1994. Bibenzyl Cannabinoid and Bisbibenzyl derivative from the Liverwort Radula Perrottetii. Phytochemistry 37, 859–862.

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.

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. Biochem. Biophys. Res. Commun. doi: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. doi: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. doi:10.1016/j.bbrc.2011.11.125

Ueda, N., Tsuboi, K., Lambert, D.M., 2005. A second N-acylethanolamine hydrolase in mammalian tissues. Neuropharmacology 48, 1079–1085. doi:10.1016/j.neuropharm.2004.12.017

van Amsterdam, J., Brunt, T., van den Brink, W., 2015. The adverse health effects of synthetic cannabinoids with emphasis on psychosis-like effects. J. Psychopharmacol. 29, 254–263. doi:10.1177/0269881114565142

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 The draft genome and transcriptome of Cannabis sativa. Genome Biol. 12, 1–17.

Vicient, C.M., Delseny, M., 1999. Isolation of total RNA from Arabidopsis thaliana seeds. Anal. Biochem. 268, 412–413.

Wagner, G.J., 1991. Secreting glandular trichomes: more than just hairs. Plant Physiol. 96, 675–679. doi:10.1104/pp.96.3.675

Wagner, G.J., Wang, E., Shepherd, R.W., 2004. New approaches for studying and exploiting an old protuberance, the plant trichome. Ann. Bot. 93, 3–11. doi:10.1093/aob/mch011

Wagner, J.A., Varga, K., Jarai, Z., Kunos, G., 1999. Mesenteric vasodilation mediated by endothelial anandamide receptors. Hypertension 33, 429–434. doi:10.1161/01.HYP.33.1.429

Walker, J.M., Huang, S.M., 2002. Endocannabinoids in pain modulation. Prostaglandins Leukot. Essent. Fat. Acids 66, 235–242. doi:10.1054/plef.2001.0361\rS0952327801903618 [pii]

Wang, R., Chen, R., Li, J., Liu, X., Xie, K., Chen, D., Yin, Y., Tao, X., Xie, D., Zou, J., Yang, L., Dai, J., 2014. Molecular Characterization and Phylogenetic Analysis of Two Novel Regio-specific Flavonoid Prenyltransferases from Morus alba and Cudrania tricuspidata. J. Biol. Chem. 289, 35815–35825. doi:10.1074/jbc.M114.608265

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

Yang, H., Zhou, J., Lehmann, C., 2015. GPR55 – a putative “type 3” cannabinoid receptor in inflammation. J. Basic Clin. Physiol. Pharmacol. 0, 1–6. doi:10.1515/jbcpp-2015-0080

Yang, Y., Miao, Y., Wang, B., Cui, G., Merz, K.M., 2012. Catalytic mechanism of aromatic prenylation by NphB. Biochemistry 51, 2606–2618. doi:10.1021/bi201800m

Yazaki, K., Kunihisa, M., Fujisaki, T., Sato, F., 2002. Geranyl diphosphate:4-hydroxybenzoate geranyltransferase from Lithospermum erythrorhizon. Cloning and characterization of a key enzyme in shikonin biosynthesis. J. Biol. Chem. 277, 6240–6246. doi:10.1074/jbc.M106387200

Yerger, E.H., Grazzini, R.A., Hesk, D., Cox-foster, D.L., Craig, R., Mumma, R., Graduate, I., 1992. A Rapid Method for Isolating Glandular Trichomes. Plant Physiol. 99, 1–7.

Young, I.G., Leppik, R.A., Hamilton, J.A., Gibson, F., 1972. Biochemical and Genetic Studies on Ubiquinone Biosynthesis in Escherichia coli K-12 : 4- Hydroxybenzoate Octaprenyltransferase. J. Bacteriol. 110, 18–25.

Zdobnov, E.M., Apweiler, R., 2001. InterProScan - an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17, 847–848. doi:10.1093/bioinformatics/17.9.847

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, 1306–13. doi:10.1104/pp.103.025213

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. doi:10.1016/S0031-9422(98)00179-4

# 7. Supplementary information

Table 37: Primer sequence used for expression profile of putative prenyltransferases

| **Primer** | **Sequence 5'→ 3'** |
|---|---|
| gapdh fw (control) | AACGGCCACTGAAGTACCAC |
| gapdh rev (control) | AGCCTTTGCTGCACCAGTAG |
| PT1 fw | CCTCATTATTATATCGAC |
| PT1 rev | GCCAGGAAATTGAG |
| PT2 fw | CAAACTGAGCCTCCTGAATC |
| PT2 rev | GATATGCCATGCTTGGAGTC |
| PT3 fw | AGGTCCATTAGGGCAGGTAG |
| PT3 rev | CCATATTTGGCGTCGCCTTC |
| PT4 fw | TGGGAGTAATGGCACATCTG |
| PT4 rev | TTGCGGTCACCTTCTATGTC |
| PT5 fw | AAATTCCATCCGGGCATGTG |
| PT5 rev | TGGCGATGACCAAAGCAAAC |
| PT6 fw | ATGAAGAGGTTGACATTTTG |
| PT6 rev | TCAAGACGATATTCGTGCTAAA |

Table 38: Amino acid sequences of the *in silico* identified prenyltransferases in the transcriptome of *Cannabis* ssp. with signal peptide

| **Enzyme** | **Amino acid sequence** |
|---|---|
| PT1p (Page and Boubakir, 2011) | MGLSSVCTFSFQTNYHTLLNPHNNNPKTSLLYRHPKTPIKYSYNNFPSKHCSTKSFHLQNKCS ESLSIAKNSIRAATTNQTEPPESDNHSVATKILNFGKACWKLQRPYTIIAFTSCACGLFGKELL HNTNLISWSLMFKAFFFLVAVLCIASFTTTINQIYDLHIDRINKPDLPLASGEISVNTAWIMSII VALFGLIITIKMKGGPLYIFGYCFGIFGGTVYSVPPFRWKQNPSTAFLLNFLAHIITNFTFYHAS RAALGLPFELRPSFTFLLAFMKSMGSALALIKDASDVEGDTKFGISTLASKYGSRNLTLFCSGI VLLSYVAAILAGIIWPQAFNSNVMLLSHAILAFWLILQTRDFALTNYDPEAGRRFYEFMWKL YYAEYLVYVFI |
| PT1 | MGLSSVCTFSFQTNYHTLLNPHNNNPKTSLLCYRHPKTPIKYSYNNFPSKHCSTKSFHLQNK CSESLSIAKNSIRAATTNQTEPPESDNHSVATKILNFGKACWKLQRPYTIIAFTSCACGLFGKE LLHNTNLISWSLMFKAFFFLVAILCIASFTTTINQIYDLHIDRINKPDLPLASGEISVNTAWIMSI IVALFGLIITIKMKGGPLYIFGYCFGIFGGIVYSVPPFRWKQNPSTAFLLNFLAHIITNFTFYYAS RAALGLPFELRPSFTFLLAFMKSMGSALALIKDASDVEGDTKFGISTLASKYGSRNLTLFCSGI VLLSYVAAILAGIIWPQAFNSNVMLLSHAILAFWLILQTRDFALTNYDPEAGRRFYEFMWKL YYAEYLVYVFI |

| | |
|---|---|
| PT2 | MELSSICNFSFQTNYHTLLNPHNKNPKSSLLSHQHPKTPIITSSYNNFPSNYCSNKNFHLQNR CSKSLLIAKNSIRTDTANQTEPPESNTKYSVVTKILSFGHTCWKLQRPYTFIGVISCACGLFGRE LFHNTNLLSWSLMLKAFSSLMVILSVNLCTNIINQITDLDIDRINKPDLPLASGEMSIETAWI MSIIVALTGLILTIKLNCGPLFISLYCVSILVGALYSVPPFRWKQNPNTAFSSYFMGLVIVNFTC YYASRAAFGLPFEMSPPFTFILAFVKSMGSALFLCKDVSDIEGDSKHGISTLATRYGAKNITFL CSGIVLLTYVSAILAAIIWPQAFKSNVMLLSHATLAFWLIFQTREFALTNYNPEAGRKFYEFM WKLHYAEYLVYVFI |
| PT3 | MGLSLVCTFSFQTNYHTLLNPHNKNPKNSLLSYQHPKTPIIKSSYDNFPSKYCLTKNFHLLGL NSHNRISSQSRSIRAGSDQIEGSPHHESDNSIATKILNFGHTCWKLQRPYVVKGMISIACGLF GRELFNNRHLFSWGLMWKAFFALVPILSFNFFAAIMNQIYDVDIDRINKPDLPLVSGEMSIE TAWILSIIVALTGLIVTIKLKSAPLFVFIYIFGIFAGFAYSVPPIRWKQYPFTNFLITISSHVGLAFT SYSATTSALGLPFVWRPAFSFIIAFMTVMGMTIAFAKDISDIEGDAKYGVSTVATKLGARNM TFVVSGVLLLNYLVSISIGIIWPQVFKSNIMILSHAILAFCLIFQTRELALANYASAPSRQFFEFI WLLYYAEYFVYVFI |
| PT4 | MVFSSVCSFPSSLGTNFKLVPRSNFKASSSHYHEINNFINNKPIKFSYFSSRLYCSAKPIVHREN KFTKSFSLSHLQRKSSIKAHGEIEADGSNGTSEFNVMKSGNAIWRFVRPYAAKGVLFNSAA MFAKELVGNLNLFSWPLMFKILSFTLVILCIFVSTSGINQIYDLDIDRLNKPNLPVASGEISVEL AWLLTIVCTISGLTLTIITNSGPFFPFLYSASIFFGFLYSAPPFRWKKNPFTACFCNVMLYVGTS VGVYYACKASLGLPANWSPAFCLLFWFISLLSIPISIAKDLSDIEGDRKFGIITFSTKFGAKPIAYI CHGLMLLNYVSVMAAAIIWPQFFNSSVILLSHAFMAIWVLYQAWILEKSNYATETCQKYYIF LWIIFSLEHAFYLFM |
| PT5 | MELSLSLGGPTIFPRYRASYTSTKLTTHFSNFPSKFSTKNFHQTLSFYGPTRGSKSLLNTHQW RNSIRACAEAGAAGSNPVLNKVSDFRDACWRFLRPHTIRGTTLGSIALVARALIENPNLIKW SLLLKAFSGLLALICGNGYIVGINQIYDIGIDKVNKPYLPIAAGDLSVQSAWYLVILFAVAGLLT VGFNFGPFITSLYCLGLVLGTIYSVPPFRMKRFPVAAFLIIATVRGFLLNFGVYYATRAALGLTF EWSSAVAFITTFVTLFALVIAITKDLPDVEGDRKFQISTFATKLGVRNIAYLGSGLLLLNYIGAIA AAIYMPQAFKRNLMLPIHTILALSLVFQAWVLEQANYTKEAIAGFYRFIWNLFYVEYIIFPFI |
| PT6 | MAIALWLPRISRSTTRRFLKPSSSLTLFSVSHSHNYIVTSNRSPIPRLFTVPNQSHGREWVSVS EVRLGYVSHISTAGKSDENRSRDAQVADVSWIDLYLPRQIHPYVRLARLDKPIGTWLLAWP CMWSISLAANPGHLPDIKMMTLFGCGALLLRGAGCTINDLLDRDIDTMVERTKLRPVASGII TPFQGICFLGFQLLLGLGILLQLNNYSRILGASSLLLVFSYPLMKRLTFWPQAYLGLTFNWGAL LGWAAVKGNIDPAIVLPLYASGVFWTLVYDTIYAHQDKEDDVRVGIKSTALRFGDLTKQWN MGFGAACISSLALSGYNAEIGWPFYASLVAASGQLAWQISTVDLSSRDDCNKKFVSNKWF GAIIFSGIVLARISS |

Table 39: Nucleotide sequences of the *in silico* identified prenyltransferases in the transcriptome of *Cannabis* spp. with signal peptide

| **Enzyme** | **Nucleotide sequence** |
|---|---|
| PT1 | ATGGGACTCTCATCAGTTTGTACCTTTTCATTTCAAACTAATTACCATACTTTAT<br>TAAATCCTCACAATAATAATCCCAAAACCTCATTATTATATCGACACCCCAAAA<br>CACCAATTAAATACTCTTACAATAATTTTCCCTCTAAACATTGCTCCACCAAGA<br>GTTTTCATCTACAAAACAAATGCTCAGAATCATTATCAATCGCAAAAAATTCCA |

| | |
|---|---|
| | TTAGGGCAGCTACTACAAATCAAACTGAGCCTCCAGAATCTGATAATCATTCA<br>GTAGCAACTAAAATTTTAAACTTTGGGAAGGCATGTTGGAAGCTTCAAAGAC<br>CATATACAATCATAGCATTTACTTCATGCGCTTGTGGATTGTTTGGGAAAGAG<br>TTGTTGCATAACACAAATTTAATAAGTTGGTCTCTCATGTTCAAGGCATTCTTT<br>TTTTTGGTGGCTGTATTATGCATTGCTTCTTTTACAACTACCATCAATCAGATTT<br>ACGATCTTCACATTGACAGAATAAACAAGCCTGATCTACCACTAGCTTCAGGG<br>GAAATATCAGTAAACACAGCTTGGATTATGAGCATAATTGTGGCACTGTTTGG<br>ATTGATAATAACTATAAAAATGAAGGGTGGACCACTCTATATATTTGGCTACT<br>GTTTTGGTATTTTTGGTGGGACTGTCTATTCTGTTCCACCATTTAGATGGAAGC<br>AAAATCCTTCCACTGCATTTCTTCTCAATTTCCTGGCCCATATTATTACAAATTT<br>CACATTTTATCATGCCAGCAGAGCAGCTCTTGGCCTACCATTTGAGTTGAGGC<br>CTTCTTTTACTTTCCTGCTAGCATTTATGAAATCAATGGGTTCAGCTTTAGCTTT<br>AATCAAAGATGCTTCAGACGTTGAAGGCGACACTAAATTTGGCATATCAACCT<br>TGGCAAGTAAATATGGTTCCAGAAACTTGACATTATTTTGTTCTGGAATTGTT<br>CTCCTATCCTATGTGGCTGCTATACTTGCTGGGATTATCTGGCCCCAGGCTTTC<br>AACAGTAACGTAATGTTACTTTCTCATGCAATCTTAGCATTTTGGTTAATCCTC<br>CAGACTCGAGATTTTGCGTTAACAAATTATGACCCGGAAGCAGGCAGAAGAT<br>TTTACGAGTTCATGTGGAAGCTTTATTATGCTGAATATTTAGTATATGTTTTCA<br>TATAA |
| PT1p (Page and Boubakir, 2011) | ATGGGACTCTCATCAGTTTGTACCTTTTCATTTCAAACTAATTACCATACTTTAT<br>TAAATCCTCACAATAATAATCCCAAAACCTCATTATTATGTTATCGACACCCCA<br>AAACACCAATTAAATACTCTTACAATAATTTTCCCTCTAAACATTGCTCCACCA<br>AGAGTTTTCATCTACAAAACAAATGCTCAGAATCATTATCAATCGCAAAAAAT<br>TCCATTAGGGCAGCTACTACAAATCAAACTGAGCCTCCAGAATCTGATAATCA<br>TTCAGTAGCAACTAAAATTTTAAACTTTGGGAAGGCATGTTGGAAACTTCAAA<br>GACCATATACAATCATAGCATTTACTTCATGCGCTTGTGGATTGTTTGGGAAA<br>GAGTTGTTGCATAACACAAATTTAATAAGTTGGTCTCTGATGTTCAAGGCATT<br>CTTTTTTTTGGTGGCTGTATTATGCATTGCTTCTTTTACAACTACCATCAATCAG<br>ATTTACGATCTTCACATTGACAGAATAAACAAGCCTGATCTACCACTAGCTTCA<br>GGGGAAATATCAGTAAACACAGCTTGGATTATGAGCATAATTGTGGCACTGT<br>TTGGATTGATAATAACTATAAAAATGAAGGGTGGACCACTCTATATATTTGGC<br>TACTGTTTTGGTATTTTTGGTGGGATTGTCTATTCTGTTCCACCATTTAGATGG<br>AAGCAAAATCCTTCCACTGCATTTCTTCTCAATTTCCTGGCCCATATTATTACAA |

| | |
|---|---|
| | ATTTCACATTTTATTATGCCAGCAGAGCAGCTCTTGGCCTACCATTTGAGTTGA<br>GGCCTTCTTTTACTTTCCTGCTAGCATTTATGAAATCAATGGGTTCAGCTTTGG<br>CTTTAATCAAAGATGCTTCAGACGTTGAAGGCGACACTAAATTTGGCATATCA<br>ACCTTGGCAAGTAAATATGGTTCCAGAAACTTGACATTATTTTGTTCTGGAATT<br>GTTCTCCTATCCTATGTGGCTGCTATACTTGCTGGGATTATCTGGCCCCAGGCT<br>TTCAACAGTAACGTAATGTTACTTTCTCATGCAATCTTAGCATTTTGGTTAATC<br>CTCCAGACTCGAGATTTTGCGTTAACAAATTACGACCCGGAAGCAGGCAGAA<br>GATTTTACGAGTTCATGTGGAAGCTTTATTATGCTGAATATTTAGTATATGTTT<br>TCATATAA |
| PT2 | ATGGAGCTCTCATCAATTTGTAACTTTTCATTTCAAACTAATTACCATACTTTAT<br>TAAACCCTCATAATAAGAATCCCAAGAGCTCATTATTATCTCATCAACACCCCA<br>AAACACCAATAATTACATCCTCTTATAATAATTTTCCCTCTAACTATTGCTCCAA<br>CAAGAACTTTCATCTACAAAACAGATGCTCAAAATCATTATTAATTGCAAAAA<br>ATTCCATTAGGACAGATACTGCAAATCAAACTGAGCCTCCTGAATCTAATACT<br>AAATATTCAGTAGTAACTAAAATTTTAAGCTTTGGGCACACATGTTGGAAACT<br>TCAAAGACCGTATACATTCATAGGAGTTATCTCATGCGCTTGTGGTTTGTTTG<br>GGAGAGAGTTGTTTCATAACACAAATTTATTGAGTTGGTCTCTGATGTTGAAG<br>GCATTCTCTTCTTTGATGGTTATTTTGAGCGTTAATTTGTGTACTAATATCATTA<br>ATCAGATTACCGATTTGGACATCGACAGGATAAACAAGCCTGATCTACCACTA<br>GCTTCAGGGGAAATGTCAATTGAAACAGCTTGGATTATGAGCATAATTGTGG<br>CCCTAACTGGGTTGATACTAACTATAAAATTGAACTGTGGACCACTCTTTATTT<br>CTCTGTATTGTGTTAGTATTTTGGTTGGAGCTCTCTATTCTGTTCCACCTTTTAG<br>ATGGAAGCAAAATCCTAATACCGCATTTTCCAGTTATTTTATGGGTCTGGTGA<br>TCGTAAATTTCACATGTTATTATGCTAGCAGAGCAGCTTTTGGCCTACCATTTG<br>AGATGAGTCCTCCTTTTACTTTCATCCTAGCCTTTGTTAAATCAATGGGATCAG<br>CTTTATTTTTATGTAAAGATGTTTCAGACATTGAAGGCGACTCCAAGCATGGC<br>ATATCAACCTTGGCAACCAGATATGGTGCTAAAAACATAACATTTCTTTGTTCT<br>GGAATTGTTCTTCTAACCTATGTATCTGCTATACTTGCTGCCATTATTTGGCCA<br>CAGGCTTTCAAGAGTAACGTAATGTTACTTTCACATGCAACCTTAGCATTTTG<br>GTTAATTTTTCAGACTCGAGAGTTTGCTCTAACAAATTACAACCCGGAAGCAG<br>GCAGAAAATTCTACGAGTTCATGTGGAAGCTCCATTATGCTGAATATTTAGTA<br>TATGTTTTCATTTAG |

| | |
|---|---|
| PT3 | ATGGGACTCTCATTAGTTTGTACCTTTTCATTTCAAACTAATTATCATACTTTAT<br>TAAACCCTCATAATAAGAATCCCAAAAACTCATTATTATCTTATCAACACCCCA<br>AAACACCAATAATTAAATCCTCTTATGATAATTTTCCCTCTAAATATTGCTTAAC<br>CAAGAACTTTCATTTACTTGGACTCAATTCACACAACAGAATAAGCTCACAAT<br>CAAGGTCCATTAGGGCAGGTAGCGATCAAATTGAAGGTTCTCCTCATCATGA<br>ATCTGATAATTCAATAGCAACTAAAATTTTAAATTTTGGACATACTTGTTGGAA<br>ACTTCAAAGACCATATGTAGTAAAAGGGATGATTTCAATCGCTTGTGGTTTGT<br>TTGGGAGAGAGTTGTTCAATAACAGACATTTATTCAGTTGGGGTTTGATGTG<br>GAAGGCATTCTTTGCTTTGGTGCCTATATTGTCCTTCAATTTCTTTGCAGCAAT<br>CATGAATCAAATTTACGATGTGGACATCGACAGGATAAACAAGCCTGATCTA<br>CCACTAGTTTCAGGGGAAATGTCAATTGAAACAGCTTGGATTTTGAGCATAAT<br>TGTGGCACTAACTGGGTTGATAGTAACTATAAAATTGAAATCTGCACCACTTT<br>TTGTTTTCATTTACATTTTTGGTATATTTGCTGGGTTTGCCTATTCTGTTCCACC<br>AATTAGATGGAAGCAATATCCTTTTACCAATTTTCTAATTACCATATCGAGTCA<br>TGTGGGCTTAGCTTTCACATCATATTCTGCAACCACATCAGCTCTTGGTTTACC<br>ATTTGTGTGGAGGCCTGCTTTTAGTTTCATCATAGCATTCATGACAGTTATGG<br>GTATGACTATTGCTTTTGCCAAAGATATTTCAGATATTGAAGGCGACGCCAAA<br>TATGGGGTATCAACTGTTGCAACCAAATTAGGTGCTAGGAACATGACATTTGT<br>TGTTTCTGGAGTTCTTCTTCTAAACTACTTGGTTTCTATATCTATTGGGATAATT<br>TGGCCTCAGGTTTTCAAGAGTAACATAATGATACTTTCTCATGCAATCTTAGCA<br>TTTTGCTTAATCTTCCAGACTCGTGAGCTTGCTCTAGCAAATTACGCCTCGGCG<br>CCAAGCAGACAATTCTTCGAGTTTATCTGGTTGCTATATTATGCTGAATACTTT<br>GTATATGTATTTATATAA |
| PT4 | ATGGTGTTCTCATCAGTTTGTAGTTTTCCATCCTCCCTTGGAACTAATTTTAAAT<br>TAGTTCCTCGTAGTAATTTTAAGGCATCATCTTCTCATTATCATGAAATAAATA<br>ATTTTATTAATAATAAACCAATTAAATTCTCATATTTTTCTTCAAGACTATATTG<br>CTCTGCCAAACCAATTGTACACAGAGAAAACAAATTCACAAAATCATTTTCAC<br>TCAGCCACCTCCAAAGGAAAAGCTCCATAAAGGCACATGGTGAAATTGAAGC<br>TGATGGGAGTAATGGCACATCTGAATTTAATGTAATGAAAAGTGGAAACGCA<br>ATTTGGAGATTTGTAAGGCCATATGCAGCCAAGGGAGTATTGTTTAACTCTGC<br>TGCTATGTTTGCAAAAGAGTTGGTGGGGAACCTAAATCTATTTAGTTGGCCTT<br>TGATGTTTAAGATACTCTCTTTTACATTGGTTATTTTATGCATTTTTGTAAGTAC<br>AAGTGGCATCAATCAAATTTATGATCTCGACATCGACAGGTTAAACAAACCTA |

| | |
|---|---|
| | ATTTGCCAGTAGCATCAGGAGAAATTTCAGTTGAATTGGCATGGTTGTTGACT<br>ATAGTTTGTACAATAAGTGGCCTCACATTAACAATTATAACGAACTCAGGGCC<br>ATTCTTCCCTTTTCTCTACTCTGCTAGTATCTTTTTTGGCTTTCTCTATTCTGCTC<br>CTCCATTCAGATGGAAGAAGAATCCTTTTACAGCATGTTTCTGTAATGTTATGT<br>TGTATGTTGGCACAAGCGTTGGTGTCTATTATGCTTGTAAGGCTAGTCTCGGG<br>CTTCCAGCCAACTGGAGCCCTGCTTTTTGTTTGCTCTTTTGGTTTATTTCATTGT<br>TGAGTATACCCATCTCCATTGCAAAAGATCTTTCAGACATAGAAGGTGACCGC<br>AAGTTTGGAATCATAACCTTCTCAACTAAATTTGGAGCAAAACCCATAGCATA<br>TATTTGTCATGGACTCATGCTTCTGAATTACGTGAGTGTTATGGCTGCAGCTA<br>TTATTTGGCCACAGTTTTTCAACAGTAGCGTAATATTGCTTTCTCATGCATTCA<br>TGGCAATTTGGGTATTATATCAGGCTTGGATATTGGAGAAATCAAATTACGCC<br>ACGGAGACGTGCCAAAAATACTATATATTCCTTTGGATAATTTTTTCTCTTGAA<br>CATGCCTTCTATTTGTTCATGTAG |
| PT5 | ATGGAGCTCTCACTCTCCCTTGGAGGACCAACTATATTCCCTCGTTACAGGGC<br>CTCATATACTTCTACCAAGCTCACAACCCACTTCTCCAACTTCCCTTCAAAATTC<br>TCCACCAAAAACTTTCATCAAACACTTAGCTTCTACGGCCCGACCAGGGGCTC<br>AAAATCTTTATTAAACACCCACCAATGGAGAAATTCCATCCGGGCATGTGCTG<br>AAGCTGGAGCTGCTGGATCAAATCCAGTACTTAACAAAGTTTCAGACTTTAGA<br>GATGCTTGCTGGAGATTCCTAAGGCCCCATACAATACGGGGAACCACTCTAG<br>GATCAATTGCTTTGGTTGCAAGAGCATTGATTGAGAACCCAAATTTGATCAAG<br>TGGTCTCTACTGCTAAAGGCATTCTCTGGCCTACTTGCTTTAATATGTGGGAAT<br>GGTTATATTGTTGGCATCAATCAGATTTACGATATCGGTATTGACAAGGTAAA<br>CAAACCTTATTTGCCTATAGCTGCTGGTGACCTTTCAGTTCAATCAGCATGGTA<br>CTTGGTGATACTTTTTGCAGTAGCTGGTTTATTGACTGTCGGATTCAACTTCGG<br>GCCATTCATCACTTCACTCTACTGTCTTGGTCTTGTCCTTGGCACCATCTACTCT<br>GTTCCTCCATTTAGAATGAAGAGATTTCCTGTTGCAGCATTTCTTATAATAGCT<br>ACGGTACGAGGGTTCCTTCTGAATTTCGGTGTATATTATGCCACCAGAGCTGC<br>CCTTGGACTTACATTTGAGTGGAGTTCAGCTGTTGCTTTCATCACTACCTTTGT<br>GACATTGTTTGCTTTGGTCATCGCCATCACAAAAGATCTTCCAGACGTGGAGG<br>GCGACCGCAAATTTCAAATATCAACCTTTGCAACAAAACTTGGAGTTCGAAAC<br>ATAGCATACCTTGGCTCTGGGCTTCTTTTGCTAAATTATATCGGTGCTATAGCG<br>GCTGCAATTTACATGCCTCAGGCTTTCAAGCGTAACTTAATGTTGCCCATTCAC<br>ACAATCTTGGCACTGTCCTTAGTTTTCCAGGCCTGGGTTTTGGAGCAAGCAAA |

| | |
|---|---|
| | TTACACTAAGGAGGCAATAGCAGGTTTCTATAGATTCATATGGAATCTCTTCT<br>ACGTGGAATATATAATATTCCCTTTCATCTAG |
| PT6 | ATGGCTATTGCGTTGTGGCTCCCTCGCATCTCACGTAGTACCACTAGGAGGTT<br>CCTCAAACCCTCTTCCTCACTCACTCTCTTTTCTGTCTCTCATTCTCATAACTATA<br>TAGTCACCTCTAACCGTAGCCCTATCCCTCGCCTCTTTACTGTTCCTAATCAAA<br>GCCATGGCCGGGAATGGGTCTCTGTCTCTGAAGTCAGACTCGGTTACGTCAG<br>CCACATTTCTACCGCCGGTAAAAGCGACGAGAATCGGAGCCGCGATGCCCAG<br>GTAGCTGATGTATCTTGGATTGACTTGTACTTGCCCAGACAGATTCATCCCTA<br>CGTTCGACTTGCTCGGCTTGATAAGCCCATTGGGACATGGTTGCTCGCTTGGC<br>CTTGTATGTGGTCAATTAGTTTGGCAGCAAACCCTGGACATCTTCCTGATATC<br>AAGATGATGACATTATTTGGATGTGGGGCTTTGCTTCTACGGGGTGCTGGAT<br>GTACCATTAACGATCTTCTCGATCGAGATATTGATACAATGGTAGAGCGTACA<br>AAGTTGAGGCCAGTCGCAAGCGGTATCATCACACCCTTTCAAGGAATTTGTTT<br>TCTTGGGTTTCAATTGCTGTTGGGTCTTGGCATTCTTCTTCAACTGAATAATTA<br>TAGTCGGATATTGGGGGCTTCATCTTTGTTGCTAGTCTTTTCCTATCCCCTTAT<br>GAAGAGGTTGACATTTTGGCCTCAAGCCTATCTAGGCTTAACTTTTAATTGGG<br>GAGCTTTATTAGGATGGGCTGCTGTTAAAGGAAACATTGATCCAGCTATAGTT<br>CTCCCCTTGTATGCCTCTGGAGTATTTTGGACTCTAGTATATGACACGATTTAT<br>GCACATCAGGACAAAGAAGATGATGTGAGAGTTGGTATTAAGTCCACGGCTT<br>TGAGATTTGGGGATTTAACTAAGCAGTGGAATATGGGGTTTGGAGCTGCATG<br>CATTAGTAGTCTTGCCCTCAGTGGATATAATGCTGAAATTGGGTGGCCATTTT<br>ATGCATCATTGGTTGCTGCATCTGGACAATTAGCTTGGCAGATATCGACCGTT<br>GACCTTTCATCTCGAGATGATTGCAATAAAAAATTTGTGTCAAACAAGTGGTT<br>TGGTGCTATTATTTTTAGTGGAATTGTTTTAGCACGAATATCGTCTTGA |
| Glyceraldehyde-3-phosphate dehydrogenase (gapdh) | ATGGCGTTCTCTTCTCTGCTCAGATCTACCACCGGAGCTGTTCTGATTGAAGCT<br>TCTAGACCTGAAATCTTTGCGCCTCTTGCTTCTTCCAATCCTTCTAAGGCATCG<br>AGCATTTGTTTCAATCGCAACCCAAACTCTGCAAGATTTCAATCAAGTTTATTT<br>GGCACTGCTGTCCCAAGTGGATCATCTGTTTTACAGGTATGCAATGTCAAGAG<br>TGTCCAACCCATTAAAGCAACGGCCACTGAAGTACCACCAACTGTTCAGAAAT<br>TGAAGACCAGTGGGAACACAAAGGTTGGGATCAATGGTTTTGGTCGGATTG<br>GAAGATTGGTTTTACGAGTAGCAACATTCAGGGATGATATTGACGTAGTGGC<br>AGTGAATGATCCTTTCATTGATGCTAAATACATGGCCTACATGTTCAAATATG<br>ATTCTACTCATGGATTATTCAAAGGAAGCATTAGGGTTGTAGATGATTCAACC |

| | |
|---|---|
| | TTAGAAATCAATGGGAAGCAGATAAAGGTTCTGAGCAAAAGGGACCCTGCT<br>GAGATTCCTTGGGGTGATTATGGAGCTGAGTATGTCGTTGAATCTTCCGGTGT<br>TTTCACAACAATTGACAAGGCTTCAGCACATAAAAAGGGTGGAGCCAAGAAA<br>GTAGTTATATCAGCTCCATCAGCTGATGCACCTATGTTCGTGGTTGGAGTAAA<br>TGAGCAGACATACAAGCCAAACATGGACATTGTTTCCAATGCAAGCTGTACTA<br>CTAACTGTCTTGCTCCTCTAGCTAAGGTGGTTCATGAGGAATTTGGCATTCTT<br>GAAGGTTTAATGACAACGGTTCATGCAACTACAGCAACACAGAAGACCGTTG<br>ATGGCCCATCAATGAAGGATTGGAGAGGAGGCCGTGGAGCTGGACAAAATA<br>TCATTCCTAGTTCTACTGGTGCAGCAAAGGCTGTTGGTAAGGTTCTCCCAGAA<br>CTGAATGGAAAGCTTACTGGAATGGCCTTCCGTGTTCCTACTCCTAATGTCTC<br>AGTGGTAGACTTAACTTGTCGACTTGAGAAGAGTGCTTCTTATGAAGATGTCA<br>AGGCAGCCATTAAGTATGCGTCAGAGGGACCACTTAAAGGCATTCTTGGGTA<br>CACAGATGAAGATGTTGTCTCCAATGATTTGTTGGTGACACAAGGTCAAGTAT<br>TTTCGATGCCAAGGCTGGAATAG |

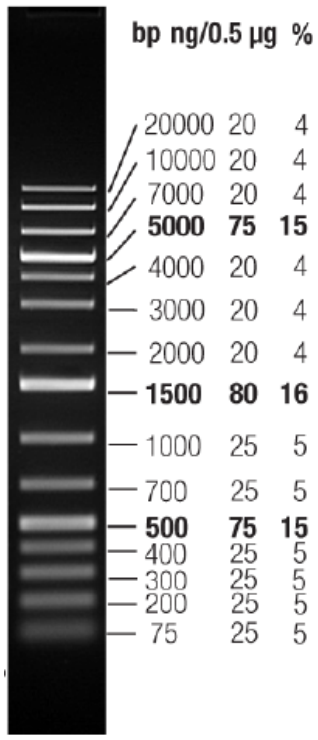



Figure 25: DNA ladder 1kb (Thermo Fisher Scientific, Waltham, Massachusetts, USA)

Table 40: Nucleotide sequences: pGEM-T easy modified; beta-subunit of F1F0 ATP Synthase complex (yellow: signal peptide, turquoise: mature protein)

| **Plasmid** | **Nucleotide sequence** |
|---|---|
| pGEM-T easy (modified) (integrated insert is highlighted) | GGGCGAATTGGGCCCGACGTCGCATGCTCCCGGCCGCCATGGCGGCCGCGG<br>GAATTCGATCCAGTGATATCATGGGTTCATCGGTACAACGAGATTCTAGAGC<br>ATCCGATTCCGATGCAGACCACGATATCACTGGATCACTAGTGAATTCGCGGC<br>CGCCTGCAGGTCGACCATATGGGAGAGCTCCCAACGCGTTGGATGCATAGCT<br>TGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTG<br>TTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGG<br>AAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTA<br>ATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCT<br>GCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCG<br>CTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGC<br>GAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGG<br>GGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGA<br>ACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGAC<br>GAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGA<br>CTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGT<br>TCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCG<br>TGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTT<br>CGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCG<br>CCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCG<br>CCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGC<br>GGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAA<br>CAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTT<br>GGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGT<br>TTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTG<br>ATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGA<br>TTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAA<br>AATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGT<br>TACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCA<br>TCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTT<br>ACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTC<br>CAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTG<br>GTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTA<br>GAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACA<br>GGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTC<br>CCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTT<br>AGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATC<br>ACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAG<br>ATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTA<br>TGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCC<br>ACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGA<br>AAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCG<br>TGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGC<br>AAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGA<br>AATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGG<br>GTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAA<br>ATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGATGCGGTGTGAA |

| | |
|---|---|
| | ATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGAAATTGTAAGCGT<br>TAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAAC<br>CAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGA<br>TAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGT<br>GGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTA<br>CGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACT<br>AAATCGGAACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCC<br>GGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCT<br>AGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCG<br>CGCTTAATGCGCCGCTACAGGGCGCGTCCATTCGCCATTCAGGCTGCGCAACT<br>GTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAA<br>AGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCA<br>GTCACGACGTTGTAAAACGACGGCCAGTGAATTGTAATACGACTCACTATA |
| Yeast (*S.cerevisiae*) nuclear ATP2 gene encoding mitochondrial F-1-ATPase beta-subunit, complete cds | GGTCTCCTCCTCAAGTCAAATTGCTTTCTTTTCTTTCATTTTTTAGTCTGTTGTTA<br>TCCAATTTATACTGAATCTTTGAGAGAAAACAAAAATAAAAAAAAGAATGGTT<br>TTGCCAAGACTATATACTGCTACATCCCGTGCTGCTTTTAAAGCAGCCAAACA<br>ATCCGCTCCGCTTCTATCCACTTCGTGGAAAAGATGTATGGCCTCAGCTGCTC<br>AATCTACTCCAATCACCGGTAAAGTTACCGCTGTCATTGGTGCCATTGTTGAC<br>GTTCATTTTGAACAATCAGAGTTGCCCGCTATTTTGAACGCTTTAGAAATTAAA<br>ACACCTCAAGGTAAGTTGGTTTTGGAAGTTGCTCAACATTTGGGTGAAAACAC<br>TGTCAGAACCATTGCTATGGATGGTACCGAAGGTTTGGTCCGTGGTGAAAAG<br>GTTCTTGACACTGGTGGCCCTATCTCCGTCCCAGTTGGGAGAGAAACTTTAGG<br>GAGAATCATCAACGTTATCGGTGAACCTATTGATGAAAGAGGTCCAATTAAG<br>TCCAAACTAAGAAAGCCAATTCACGCAGACCCTCCTAGTTTTGCAGAACAATC<br>TACTTCGGCTGAAATTTTGGAAACAGGTATCAAAGTCGTCGATCTATTAGCTC<br>CTTATGCCAGAGGTGGTAAGATTGGTCTTTTCGGTGGTGCAGGTGTCGGTAA<br>GACTGTGTTCATTCAAGAATTGATTAACAATATCGCCAAGGCCCATGGTGGTT<br>TTTCCGTTTTCGCCGGTGTTGGTGAAAGGACCAGAGAGGGTAATGACTTGTA<br>CCGTGAAATGAAGGAAACTGGAGTCATTAACTTGGAAGGTGAATCCAAGGTC<br>GCCTTAGTTTTCGGTCAAATGAACGAACCTCCAGGAGCCAGAGCCAGAGTCG<br>CTTTAACTGGTTTGACGATCGCTGAATATTTCAGAGATGAAGAAGGTCAAGA<br>CGTCTTGTTGTTTATCGACAATATCTTTAGATTTACTCAAGCTGGTTCAGAAGT<br>CTCTGCCCTTTTGGGTCGTATTCCATCTGCCGTCGGTTATCAACCAACTTTGGC<br>CACTGATATGGGTCTCTTACAAGAAAGAATTACCACCACAAAGAAGGGTTCT<br>GTCACTTCTGTGCAAGCCGTTTATGTTCCAGCCGATGATTTAACAGATCCGTCT<br>CCGTCCACATCTTTTGCCCATTTGGACGCATCATCCGTCTTGTCAAGAGGTATT<br>TCAGAATTAGGTATTTACCCTGCAGTGGATCCATTGGATTCTAAATCAAGGTT<br>ATTGGATGCCGCCGTTGTCGGTCAAGAACATTATGACGTCGCCTCCAAGGTTC<br>AAGAAACTTTACAGACCTATAAATCTTTACAAGATATCATTGCTATTTTGGGTA<br>TGGATGAATTGTCCGAACAAGATAAACTAACTGTCGAAAGGGCAAGAAAGAT<br>TCAAAGATTCTTATCTCAACCATTTGCTGTCGCCGAAGTCTTTACTGGTATCCC<br>AGGTAAATTAGTGAGATTAAAGGACACCGTTGCCTCGTTCAAAGCCGTTTTG<br>GAAGGTAAATACGATAATATACCAGAACATGCTTTCTATATGGTTGGTGGTAT<br>TGAAGATGTTGTTCGTAAAGCTGAAAAGTTAGCCCGTGAAGCCAACTAGAAG<br>AAATAAAGCTTAAACCAAGGGAAGCAAAATTTGAAATACCGAAGATGAACAA<br>TAAGGATGATGGGAAAAAAAAGAGAATTTTTTTTTTTTGTTTTTCCCTGCTTCC<br>TTCTTGTTTATTGGTATTATTATGTTACGATATTCATTCATTATCCTATTGATAT<br>TTTCTTTATATTCACTAAAAAAAAATTTATTCTATAAGACTGACTATAATTTTTT<br>TTACTCCCAACTGTAAGTAAATAAAGACTCACCTACGCATACATTTTTTATATA |

| | |
|---|---|
| | TACTATAAGATGTAGGATCT |

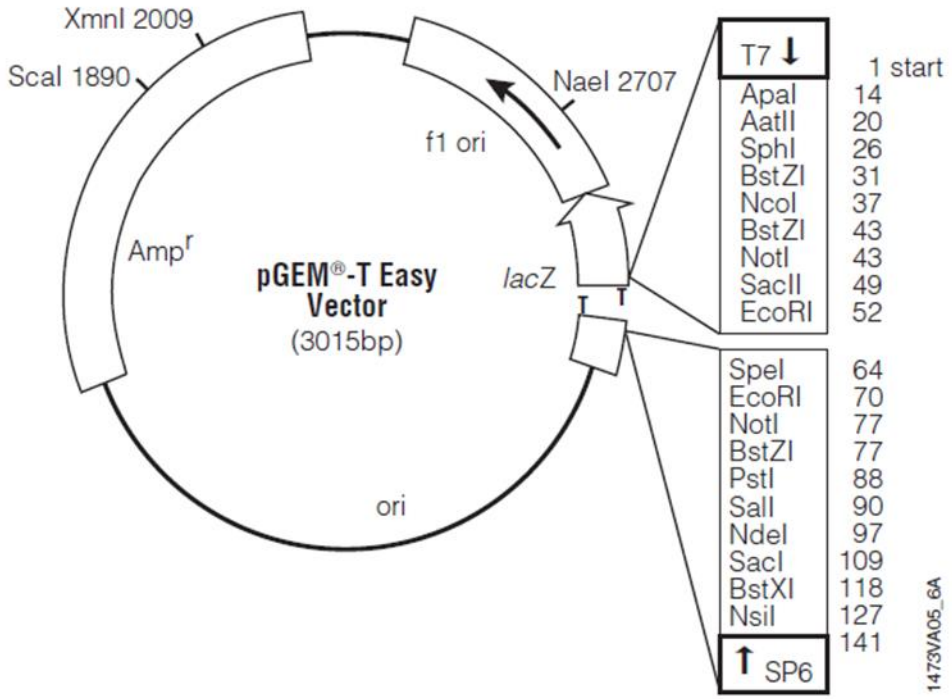



Figure 26: Cloning vector pGEM-T easy (Promega GmbH, Mannheim, Germany)

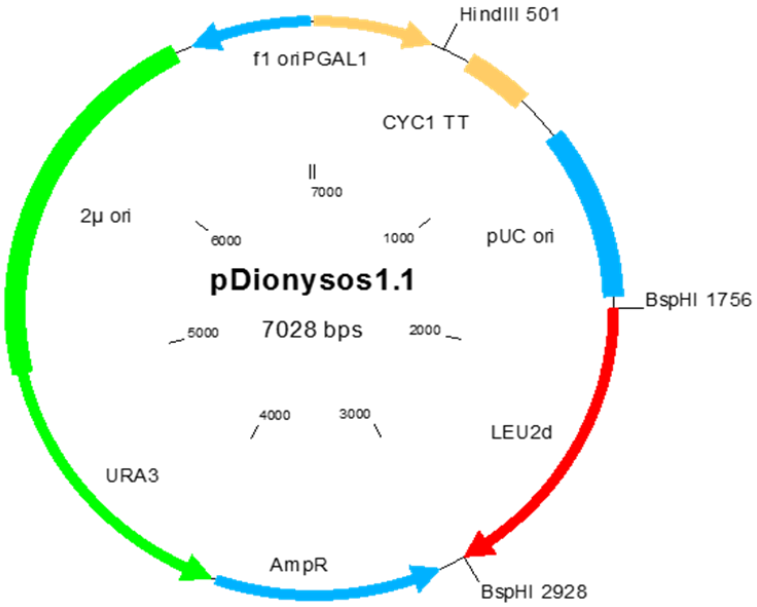



Figure 27: Expression vector pDionysos1.1: pYES2 vector (Thermo Fisher Scientific Life Technologies GmbH, Darmstadt, Germany) modified by an additional Leu2d marker (red).

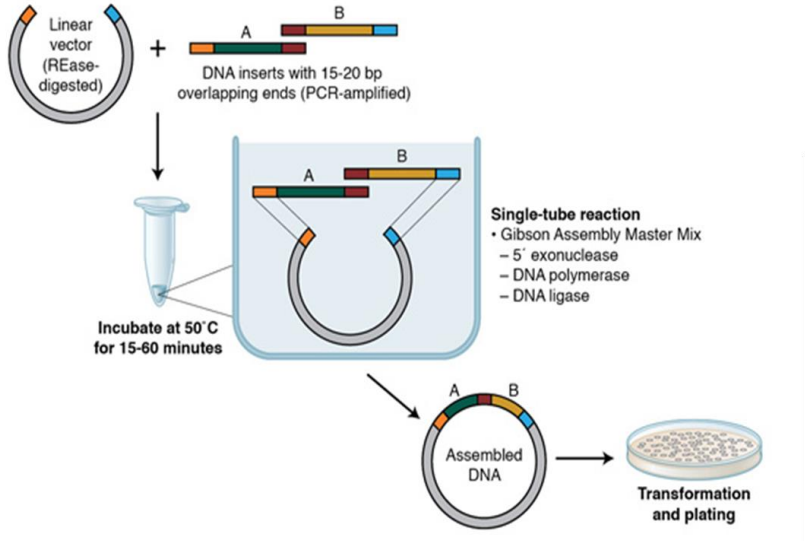



Figure 28: Schematic explanation of Gibson® Assembly (New England Biolabs GmbH, Frankfurt am Main, Germany) (Source: New England Biolabs)

Table 41: Amino acid mixture used in minimal medium (SD medium)

| Component | Concentration (mg/L) |
|---|---|
| **Adenine hemisulfate** | 18 |
| **L-Alanine** | 76 |
| **L-Arginine** | 76 |
| **Asparagine Monohydrate** | 76 |
| **Aspartic acid** | 76 |
| **L-Cysteine** | 76 |
| **L-Glutamic acid** | 76 |
| **L-Glutamine** | 76 |
| **Glycine** | 76 |
| **L-Histidin** | 76 |
| **Meso-Inositol** | 76 |
| **L- Isoleucin** | 76 |
| **L-Leucine** | 380 |
| **L-Lysine Monohydrochloride** | 76 |
| **L-Methionine** | 76 |
| **4-Aminobenzoic acid** | 8 |
| **L-Phenylalanine** | 76 |
| **L-Proline** | 76 |
| **L-Serine** | 76 |
| **L-Threonine** | 76 |
| **L-Tryptophane** | 76 |
| **L-Tyrosine** | 76 |
| **Uracile** | 76 |
| **L-Valine** | 76 |

Table 42: Amino acid sequences of PTs for sequential analysis

| **No.** | Enzyme | **Amino acid sequence** |
|---|---|---|
| **1** | protoheme IX farnesyltransferase (Cox10) (*Arabidopsis thaliana*) | MWRRSVVYRFSSRISVSSSLPNPRLIPWSRELCAVNSFSQPPVSTESTAKLGITGVRSDA NRVFATATAAATATATTGEISSRVAALAGLGHHYARCYWELSKAKLSMLVVATSGTGYIL GTGNAAISFPGLCYTCAGTMMIAASANSLNQIFEISNDSKMKRTMLRPLPSGRISVPHAV AWATIAGASGACLLASKTNMLAAGLASANLVLYAFVYTPLKQLHPINTWVGAVVGAIPPL LGWAAASGQISYNSMILPAALYFWQIPHFMALAHLCRNDYAAGGYKMLSLFDPSGKRIAA VALRNCFYMIPLGFIAYDWGLTSSWFCLESTLLTLAIAATAFSFYRDRTMHKARKMFHAS LLFLPVFMSGLLLHRVSNDNQQQLVEEAGLTNSVSGEVKTQRRKKRVAQPPVAYASAAPF PFLPAPSFYSP |
| **2** | Prenyltransferase | FVWLVASMELSISQSPCVRFSSLAPRFLAASHHHRSSVHLAGKFISLPRDVSFTSLSTSR MRSKFVSTNYRKISIQACSQVGAAGTDPVLDRITRFQNACWRFLRPHTIRGTALGSTALV |

| | | |
|---|---|---|
| | (*Arabidopsis lyrata* subsp. *Lyrata*) | ARALIENTHLIKWSLVLKALSGLLALICGNGYIVGINQIYDIGIDKVNKPYLPIAAGDLS VQSAWLLVIFFAIAGLLVVGFNFGPFITSLYSLGLFLGTIYSVPPLRMKRFPIAAFLIIA TVRGFLLNFGVYHATRAALGLPFQWSAPVAFITSFVTLFALVIAITKDLPDVEGDRKFQI STLATKLGVRNIAFLGSGLLLVNYVSAISLAFYMPQVFRGSLMIPAHLILASCLIFQTWV LEKANYTKEAISGYYRFIWNLFYAEYLLFPFL*LSW*GYA |
| **3** | Protein prenyl-transferase (*Arabidopsis lyrata* subsp. *Lyrata*) | MRQGPCSEEDCFSLLKQFEHVLEADPLIDEVGFIHPSQFNLLDEEAGSANVYQNELQPNN GTSRKFWNQDHKLGISTDILVQLCKDAKHVFLLAFEEYKRHGNACNESQIENFSCSPGTP ESEVMKHSQAVLLISSDFGTAWNARKLILSKKDQLSAFMEELRLSGLILSNSHKSESTWS HRRWIIKMISQRFSTPQEIITKESELVESIGERSKMNYRAWYHRCWLVSYMAIEQVIQEL NKSKRWARLHVADSSCFHYRRRLMLKILESLYVKGGNAYDKSEARKIWKEELDWNKELVE RYVGREALWLHRRFLSLNWIMYFACNDSDVSPEPGESTIMNEEIAIFIDNEIRLVEFSMTVPD TKFEDFQAQALHAAVYMLWLTKKMPELWRMVEEKLGTEKVKCVMSTIAQERPSLLHHLVNV |
| **4** | cis-prenyltrans-ferase (*Arabidopsis thaliana*) | MMTTRNGSLRFFNHTFSFYDEATCVVREEELQRELMPRHVSFILDGNRRWAKRDGLTTAQ GHEAGTKRIIEIAEVCFELGIHTVSAFAFSTENWGRDKFEVKCLMSLFNHYLKSNIQYFQ RKEVRVSVIGNKTKIPESLLKEIHEIEEATKATRISISSWHLVKKSEKGLIREEDVDEAL IERELLTNCSDFPSPDLMIRTSGEQRISNFFLWQLAYTELFYSPVLWPDFDKDKLLEALA SYQGRERRFGCRV |
| **5** | Polyprenyl-transferase (*Arabidopsis thaliana*) | DDVLFLAGVVVALWQQFLSKMAFFGLSRVSRRLLKSSVSVTPSSSSALLQSQHKSLSNPV TTHYTNPFTKCYPSWNDNYQVWSKGRELHQEKFFGVGWNYRLICGMSSSSSVLEGKPKKD DKEKSDGVVVKEASWIDLYLPEEVRGYAKLARLDKPIGTWLLAWPCMWSIALAADPGSLP SFKYMALFGCGALLLRGAGCTINDLLDQDIDTKVDRTKLRPIASGLLTPFQGIGFLGLQL LLGLGILLQLNNYSRVLGASSLLLVFSYPLMKRFTFWPQAFLGLTINWGALLGWTAVKGS IAPSIVLPLYLSGVCWTLVYDTIYAHQLHHHRFLFQLLFHVSCKCIHFDDKEDDVKVGVK STALRFGDNTKLWLTGFGTASIGFLALSGFSADLGWQYYASLAAASGQLGWQIGTADLSS GADCSRKFVSNKWFGAIIFSGVVLGRSFQ |
| **6** | PT1 (*Cannabis sativa* L.) | HMGLSSVCTFSFQTNYHTLLNPHNNNPKTSLLYRHPKTPIKYSYNNFPSKHCSTKSFHLQ NKCSESLSIAKNSIRAATTNQTEPPESDNHSVATKILNFGKACWKLQRPYTIIAFTSCAC GLFGKELLHNTNLISWSLMFKAFFFLVAVLCIASFTTTINQIYDLHIDRINKPDLPLASG EISVNTAWIMSIIVALFGLIITIKMKGGPLYIFGYCFGIFGGTVYSVPPFRWKQNPSTAF LLNFLAHIITNFTFYHASRAALGLPFELRPSFTFLLAFMKSMGSALALIKDASDVEGDTK FGISTLASKYGSRNLTLFCSGIVLLSYVAAILAGIIWPQAFNSNVMLLSHAILAFWLILQ TRDFALTNYDPEAGRRFYEFMWKLYYAEYLVYVFI |
| **7** | PT1p (*Cannabis sativa* L.) (patent-registered) | MGLSSVCTFSFQTNYHTLLNPHNNNPKTSLLCYRHPKTPIKYSYNNFPSKHCSTKSFHLQ NKCSESLSIAKNSIRAATTNQTEPPESDNHSVATKILNFGKACWKLQRPYTIIAFTSCAC GLFGKELLHNTNLISWSLMFKAFFFLVAILCIASFTTTINQIYDLHIDRINKPDLPLASG EISVNTAWIMSIIVALFGLIITIKMKGGPLYIFGYCFGIFGGIVYSVPPFRWKQNPSTAF LLNFLAHIITNFTFYYASRAALGLPFELRPSFTFLLAFMKSMGSALALIKDASDVEGDTK FGISTLASKYGSRNLTLFCSGIVLLSYVAAILAGIIWPQAFNSNVMLLSHAILAFWLILQ TRDFALTNYDPEAGRRFYEFMWKLYYAEYLVYVFI |
| **8** | aromatic prenyltransferase (*Epimedium acuminatum*) | KYPFMDSLLLLGSVSKPCWRLSFSPSAKLSTATRGY HVPIRFANSSAWSTQERRYFGQLQGHLINHHITIDAEKSSFYRRADKKGLVNATSEPPFA SEPESYNPNNFWRSMQSATDAFYRFSRPHTVIGTALSILSVSLLAIERLSDLSPLFFTGL LEAIVAALFMNIYIVGLNQLFDVEIDKVNKPYLPLASGEYSIGTGILIVAAFAVMSFWLG WFVGSGPLLWALSISFILGTAYSINLPLLRWKRFALVAAMCILAVRAVIVQLAFFLHIQT FVYRRPAILTRPLIFATAFMSFFSVVIALFKDIPDIEGDAIFGIRSFTVRLGQKRVFWIC VYLLEMAYGVAVLVGAASPSPLSKLVTVLGHVVLASILWLNAKSVDLTNKTAITSFYMFI WKLFYAEYLLIPLVR*NIWYLSGHVRSSFIFKRILFISFFAGCMLGLRWLYLMNYNIIFLAL |
| **9** | flavonoids prenyltransferase (*Sophora flavescens*) | MGSMLLASFPGASSITTGGSCLRSKRYAKNYYASSYVTTLWHKKGKIQ KEHRAVIFSKHNLKQHYKVNEGGSTSNTSKTCEKKYVVNAISEQSFEYEPQARDPESIWG SVNDALDTFYKFCRPYAMFNVVLGATFKSLVAVEKLSDLSLAFFIGWLQVVVAVICSHIF GVGLNQLCDIEIDKINKPDLPLASGKLSFRNVVIITASSLILGLGFAWIVDSWPLFWTVF ISCMVASAYNVDLPLLRWKKYPVLTAINFIADVAVTRSLGFFLHMQTCVFKRPTTFPRPL IFCTAIVSIYAIVIALFKDIPDMEGDEKFGIQSLSLRLGPKRVFWICVSLLEMTYGVTIL VGATSPILWSKIITVLGHAVLASVLWYHAKSVDLTSNVVLQSFYMFIWKLHTAEYFLIPLFR |
| **10** | *cis*- | HRDLGLRKMEIYTGQRPSVFRIFGKYMRKGLYSILTQGPIPTHLAFIMDGNRRFAKKHKM |

| | | |
|---|---|---|
| | prenyltransferase (*Hevea brasiliensis*) | KEAEGYKAGYLALLRTLTYCYELGVRYVTIYAFSIDNFRRQPREVQCVMNLMMEKIEEIIVE ESIMNAYDVGVRIVGNLNLLDEPIRIAAEKIMRATANNSRFVLLIAVAYSSTDEIVHAVEES SKDKLNSNEVCNNGIEAEQEFKEANGTGNSVIPVQKTESYSGINLADLEKNTYVNPHPDV LIRTSGLSRLSNYLLWQTSNCILYSPFALWPEIGLRHLVWTVMNFQRHHSYLEKHKEYLK |
| **11** | aromatic prenyltransferase (*Humulus lupulus*) | MELSSVSSFSLGTNPFISIPHNNNNLKVSSYCCKSKSRVINSTNSKHCSPNNNNNTSNKT THLLGLYGQSRCLLKPLSFISCNDQRGNSIRASAQIEDRPPESGNLSALTNVKDFVSVCW EYVRPYTAKGVIICSSCLFGRELLENPNLFSRPLIFRALLGMLAILGSCFYTAGINQIFD MDIDRINKPDLPLVSGRISVESAWLLTLSPAIIGFILILKLNSGPLLTSLYCLAILSGTI YSVPPFRWKKNPITAFLCILMIHAGLNFSVYYASRAALGLAFAWSPSFSFITAFITFMTL TLASSKDLSDINGDRKFGVETFATKLGAKNITLLGTGLLLLNYVAAISTAIIWPKAFKSN IMLLSHAILAFSLIFQARELDRTNYTPEACKSFYEFIWILFSAEYVVYLFI |
| **12** | *cis*-prenyltransferase (*Lilium longiflorum*) | MISHELSKWKKNDNQFAPTKFFSNVTSLLRRFFFAVLSVGPMPRHIAFILDGNRRYGK KWKLKEGESHNIGFLTLVRILRYCCEMGVEYVTLYAFSIDNFNRKPNEVQYVMNLIREN TQALVRDLDTVNRLGVRVNFIGRLDLLDGPLREAARTVMKATAGNTRIVLWVCTAYTS TEEIVHGVQGAVEDEWARLRMEGTKREISLEDLEGKMYFERNPDPDILIRTSGETRLSNF VLWQTSFCLLYAPRCLWPDLSLRHLVWAVLLYQRSYAYLEKAKKYKLEVNGQGRSLTPEC MAAFTAASSYIQF |
| **13** | Protein prenyltransferase (Os11g0483950) (*Oryza sativa*) | MAAAAAAAVGARGGIRRRPNLSLLADRCATPRALAAVHAAMLVSGRLADDAFAASRLLA AHAALSPPGAVLRLLASLPCAPNSFMLNITLRALASSPDPASALRFFSLLRRGSGGGGGSYSP GRHTFTFLLKASARLPLRASEQLHALAVRHGLERDAYVANGLVRAYSLAGLVPLARRVFDGLP ERSAVVCTTMVSGYAQNGMHEDAMRSFEEMVGDGIEPHGAALASVLSSCARSGSRGLEM GRRVHELMESRRVTAPVVGAILGTALVDMYAKTGAMEEATAVFDRMPERQTATWNA LITGLAHHGHGEVALATFHRMRRDGVPPNGATLVGVLSAYGCTGRLDEARRVFASMEK DFAVAPTIQHYGCMVDLLGRSGLLTEAEEMIRGMTTCDADTVIWGALLNACKNH GDIDVAERAVQEMLKLDPGNHGVYVVLSNMYAEAGRWQDVDRLRKVMKRARLSKIPG SSTVAGDDS |
| **14** | Prenyltransferase (*Populus trichocarpa*) | MLVVATSGTGYVLGSAGSAVDLAGLCWTCAGTMMVASSANSLNQVFEKNNDAKMKRTRLR PLPSRRMTIPHALTWASSVGVAGTAILACKANMLAAGLGVSNLILYAFVYTPLKQIHPVN TWVGAVVGAIPPLLGWAAASGEVSLNSMILPAALYFWQIPHFMALAYLCRSDYAAGGFKM FSLADASGRRTAAVALRNCLYLVPLGFLACDWGVTSSWFCLESSLITLAISALAFSFYLN RTTQNARRMFHASLLYLPVFMSGLLFHRVSDNQQCLSEDSLGRITELEPVVQGRENLKHG NQETHSAIGKQARAPVAYASAAPFPFLPAPSYTTS |
| **15** | Prenyltransferase (*Sophora flavescens*) | MGFVLPASFPGASSITTGGSCLRSKQYAKNYYASSYVTTLWHKKGKIQKEYCAVIFSRHN LKQHYKVNEGGSTSKECEKKYVVNAISEQSFEYEPQARDPKNIWGSVNDALDTFYKFCRP YAIFSVVLGATFKSLVAVERLSDLSLAFFIGWLQVVVAVICIHIFDVGLNQLCDIEIDKI NKPDLPLASGNLSFRNVVIITASSLILGLGFAWIVGSWPLFWTVFICCMFAAAYNVDLPL LRWKKYPVLTAISFIANVAVTRSLGFFLHMQTCVFKRPTTFPRPLIFCTAIVSIYAIVIA LFKDIPDMEGDEKFGIQSLSLRLGPKRVFWICVSLLEMAYGVTILVGATSPILWSKIITV LGHAILASVLWYHAKSTDLTSNVVLQSFYMFIWKLHTAEYCLIPLFR |
| **16** | protoheme IX farnesyltransferase (*Ricinus communis*) | MLVVATSGTGYVLGSGNAVDFAGLCWTCAGTMMVAASANSLNQVFEINNDSKMKRTRLRP LPSGRLTVPHAVTWASSVGLAGTAILASKANMLAAGLGASNLILYAFIYTPLKQIHPVNT WVGAVVGAIPPLLGWAAASGQVSLNAMILPAALYFWQIPHFMALAYLCRNDYAAGGFKMF SLADASGQRTALVALRNCLYLVPMGFLAYDWGVTSGWFCLESSLLTLAISASAFSFCRDR TMKKAKRMFHASLLYLPVFMSGLLFHRLCDNQAGLAEETELLSSSKAVVRGSGNDDRKTK ERHRSASVQARPPVAYASVAPFPFLPAPSYATP |
| **17** | flavonoid prenyltransferase (*Sophora flavescens*) | QKLKLSCQMDSVLRASFPRVSSISTGGSCWRSKQNAKHYFTSSNVTTSWHKTGKIQK EYCVVMSSRHNLKHHYKVIEGESTCQNSDRKYVVKATFGQPFEYEHQAQDPKRILNSVKN ALDIFYRFSRPYAAIGAALGATSVSFLAVEKLSDLSLAFVIGWLQMVVASFCMNIFNCGL NQLCDVEIDKINKPFLPLASGELSFRTAVLIVASSLIMSFWLAWVEGSWPLFWAFSVSSV LGAAYSVDWPLLRWKKSPVLAAVNILINSAIARPLGYFLHIQTRVFKRPPTFPKPMIFCT AIVSLFFVVIALFKDLSDMEGDEKHGIQSLSLRLGQKRVFWICISLLEMAYGVTILVGAT SPFLWSKISTGLGHAILALVLWFHAKSVDMKSNAALQSFYLFIWKLLWAEYFLIPLFR |
| **18** | Prenyltransferase (*Zea mays*) | FVFCSPRLPSPHPDKGREYRPAPPCLPSPLLLSPRPTPRRPRSRLPQPAGVLDAALPVRR RLIPSLLPQVRSIRMSPGPFFGCAVHFFVASFQIPRLQYPRKKKNVSVSSETITVAEIER QMYISSEKLENAAELLGTAPSRVQVPGLITKQAATARMELPVAEPGTERGEGPLIQCPYC |

DSEAMHKLAQILLPGLAAVCVDGTTGDPFHSPSVVAVDLRREMVDYITQRSESFIADALI
ESEANQEGPDAAVPDDPFEAVSVFMDDFGSTKRNIIGHVSGWLLSDSRDDKIDDLVQEME
ATRFWPLERREAVAETLLKNVDLRTRFHCPEKYESEARLADHKQQCSFRPVACPNDGCRA
RVSVRCARDHDAACPYKPLQCEQGCEKRLLRRDMDRHCVTVCPMRPMKCPFGCDSSFPDR
DLEKHCAQFLRPHLIKVLKAIHKKARSEDELKDLALKLEKYDGDGKLAKALDARPLTNAV
KDLEAKMKAEHSSRRAVLH

```
# 1 Length: 431
# 1 Number of predicted TMHs:  9
# 1 Exp number of AAs in TMHs: 162.43789
# 1 Exp number, first 60 AAs:  0.00506
# 1 Total prob of N-in:        0.91517
1       TMHMM2.0        inside      1   105
1       TMHMM2.0        TMhelix   106   127
1       TMHMM2.0        outside   128   130
1       TMHMM2.0        TMhelix   131   153
1       TMHMM2.0        inside    154   173
1       TMHMM2.0        TMhelix   174   196
1       TMHMM2.0        outside   197   199
1       TMHMM2.0        TMhelix   200   217
1       TMHMM2.0        inside    218   228
1       TMHMM2.0        TMhelix   229   251
1       TMHMM2.0        outside   252   254
1       TMHMM2.0        TMhelix   255   277
1       TMHMM2.0        inside    278   297
1       TMHMM2.0        TMhelix   298   317
1       TMHMM2.0        outside   318   321
1       TMHMM2.0        TMhelix   322   344
1       TMHMM2.0        inside    345   356
1       TMHMM2.0        TMhelix   357   374
1       TMHMM2.0        outside   375   431
```

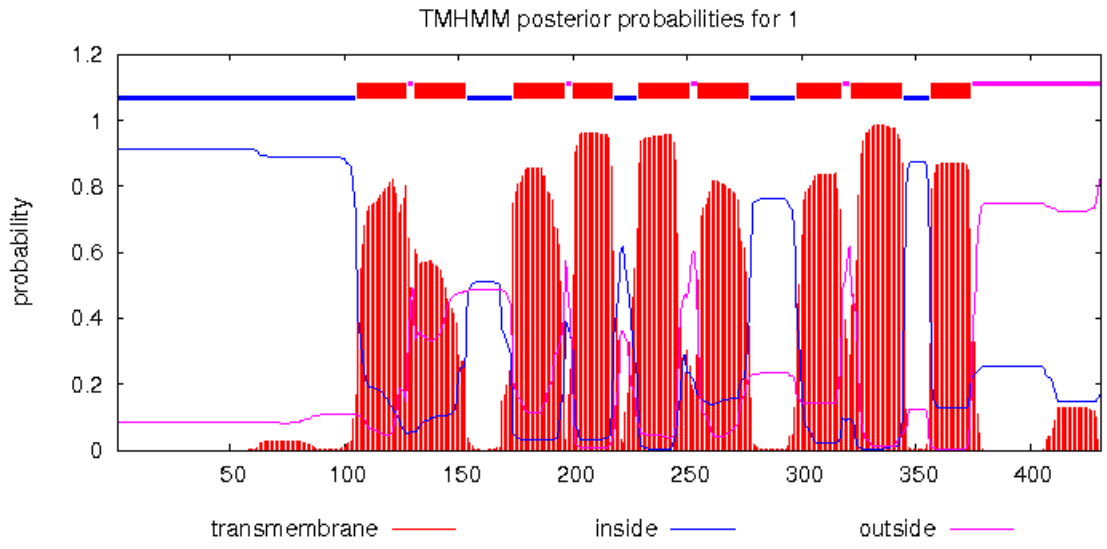



Figure 29: TMHMM analysis for Protoheme IX farnesyltransferase (Cox10) (*Arabidopsis thaliana*) (http://www.cbs.dtu.dk/services/TMHMM)

```
# 2 Length: 385
# 2 Number of predicted TMHs:  6
# 2 Exp number of AAs in TMHs: 158.85532
# 2 Exp number, first 60 AAs:  0.00705
# 2 Total prob of N-in:        0.58007
2       TMHMM2.0        inside      1   134
2       TMHMM2.0        TMhelix   135   157
2       TMHMM2.0        outside   158   176
2       TMHMM2.0        TMhelix   177   211
2       TMHMM2.0        inside    212   223
2       TMHMM2.0        TMhelix   224   246
2       TMHMM2.0        outside   247   255
2       TMHMM2.0        TMhelix   256   278
2       TMHMM2.0        inside    279   304
2       TMHMM2.0        TMhelix   305   327
2       TMHMM2.0        outside   328   336
2       TMHMM2.0        TMhelix   337   354
2       TMHMM2.0        inside    355   385
```

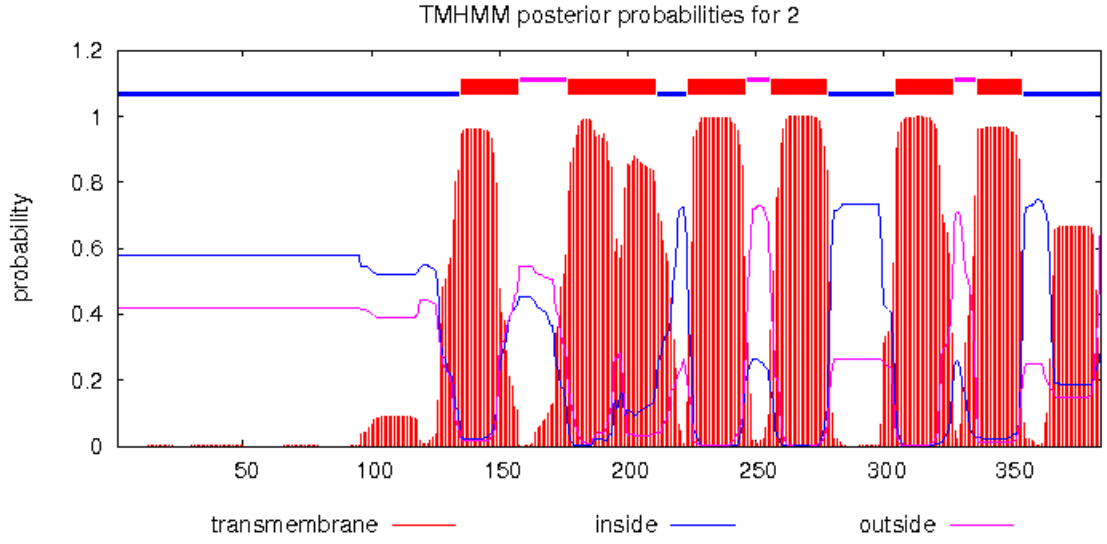



Figure 30: TMHMM analysis for Prenyltransferase (*Arabidopsis lyrata*) (http://www.cbs.dtu.dk/services/TMHMM)

```
# 3 Length: 424
# 3 Number of predicted TMHs:  0
# 3 Exp number of AAs in TMHs: 0.01545
# 3 Exp number, first 60 AAs:  0
# 3 Total prob of N-in:        0.00634
3       TMHMM2.0        outside     1   424
```

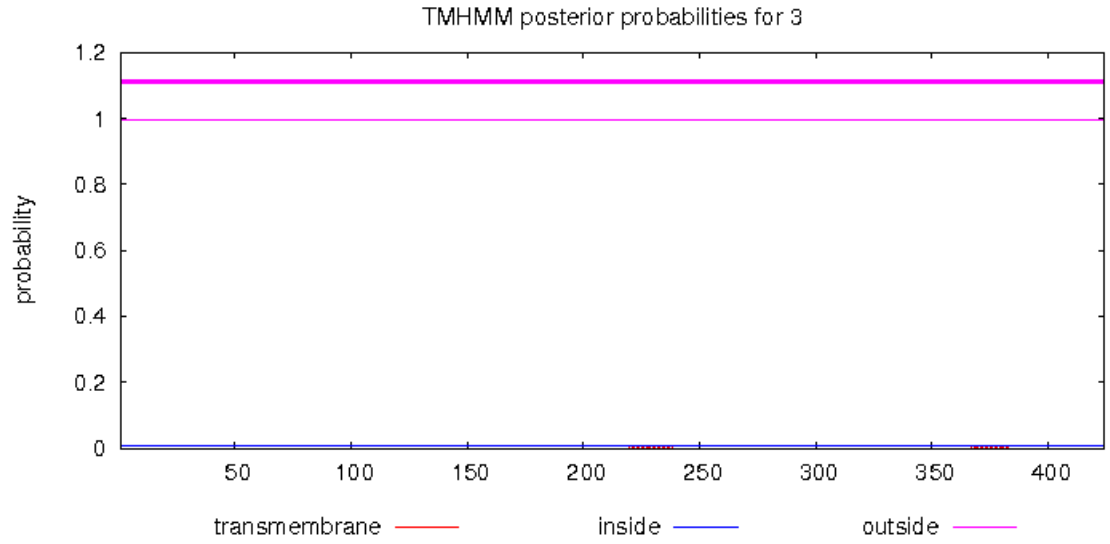



Figure 31: TMHMM analysis for protein prenyltransferase (*Arabidopsis lyrata*) (http://www.cbs.dtu.dk/services/TMHMM)

---

```
# 4 Length: 253
# 4 Number of predicted TMHs:  0
# 4 Exp number of AAs in TMHs: 1.02362
# 4 Exp number, first 60 AAs:  0.00038
# 4 Total prob of N-in:        0.20191
4       TMHMM2.0       outside      1   253
```

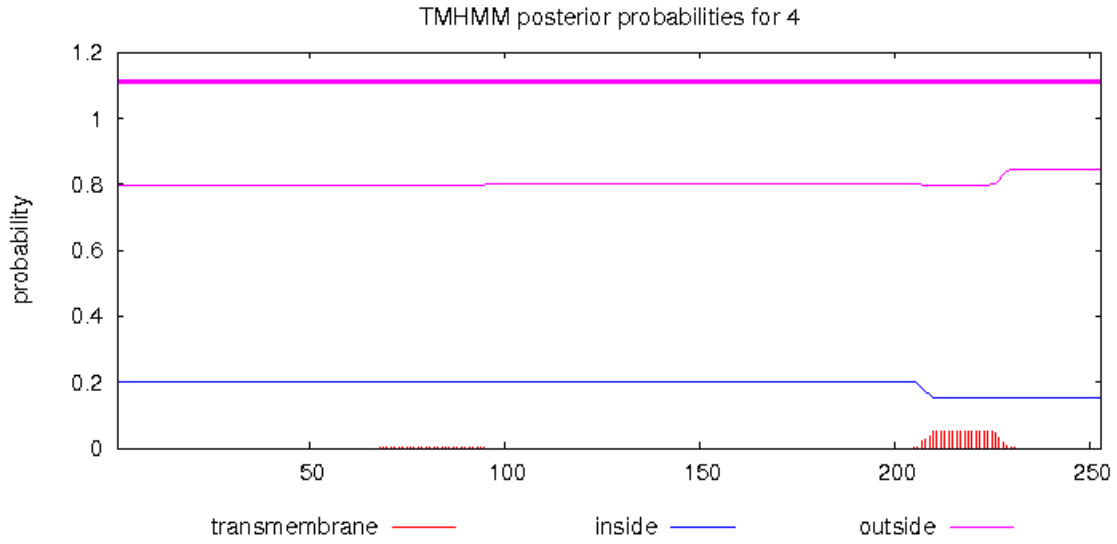



Figure 32: TMHMM analysis for cis-prenyltransferase (*Arabidopsis thaliana*) (http://www.cbs.dtu.dk/services/TMHMM)

---

```
# 5 Length: 429
# 5 Number of predicted TMHs:  6
# 5 Exp number of AAs in TMHs: 118.72501
# 5 Exp number, first 60 AAs:  0.00111
# 5 Total prob of N-in:        0.56564
5       TMHMM2.0       inside       1   136
5       TMHMM2.0       TMhelix    137   154
5       TMHMM2.0       outside    155   163
5       TMHMM2.0       TMhelix    164   186
5       TMHMM2.0       inside     187   205
5       TMHMM2.0       TMhelix    206   228
5       TMHMM2.0       outside    229   254
5       TMHMM2.0       TMhelix    255   277
5       TMHMM2.0       inside     278   283
5       TMHMM2.0       TMhelix    284   306
5       TMHMM2.0       outside    307   351
5       TMHMM2.0       TMhelix    352   371
5       TMHMM2.0       inside     372   429
```

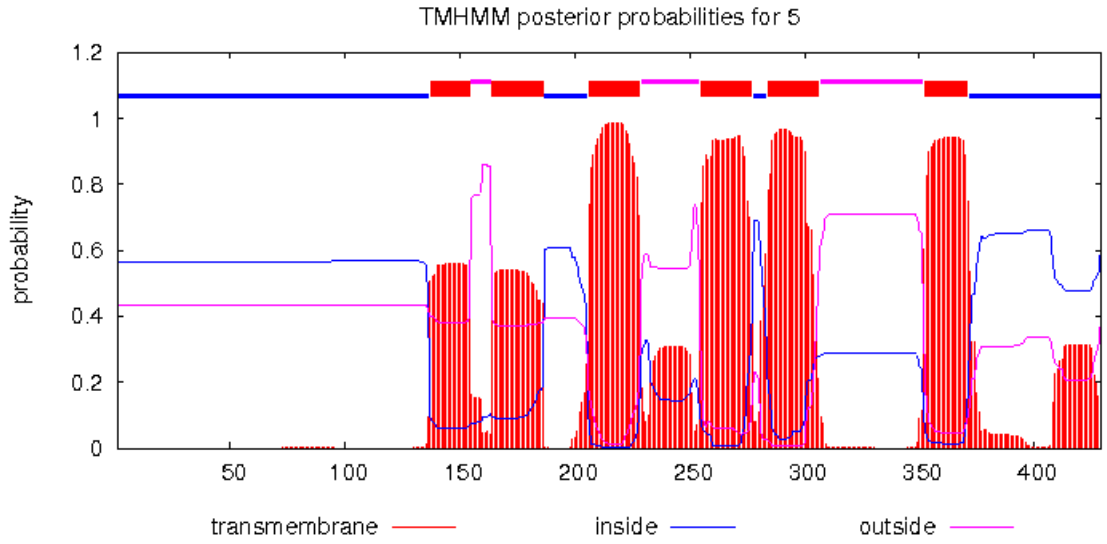



Figure 33: TMHMM analysis for polyprenyltransferase (*Arabidopsis thaliana*) (http://www.cbs.dtu.dk/services/TMHMM)

---

```
# 7 Length: 395
# 7 Number of predicted TMHs:  6
# 7 Exp number of AAs in TMHs: 139.40014
# 7 Exp number, first 60 AAs:  0.00202
# 7 Total prob of N-in:        0.43078
7       TMHMM2.0        inside      1   139
7       TMHMM2.0        TMhelix   140   162
7       TMHMM2.0        outside   163   181
7       TMHMM2.0        TMhelix   182   201
7       TMHMM2.0        inside    202   207
7       TMHMM2.0        TMhelix   208   230
7       TMHMM2.0        outside   231   239
7       TMHMM2.0        TMhelix   240   262
7       TMHMM2.0        inside    263   313
7       TMHMM2.0        TMhelix   314   336
7       TMHMM2.0        outside   337   339
7       TMHMM2.0        TMhelix   340   359
7       TMHMM2.0        inside    360   395
```

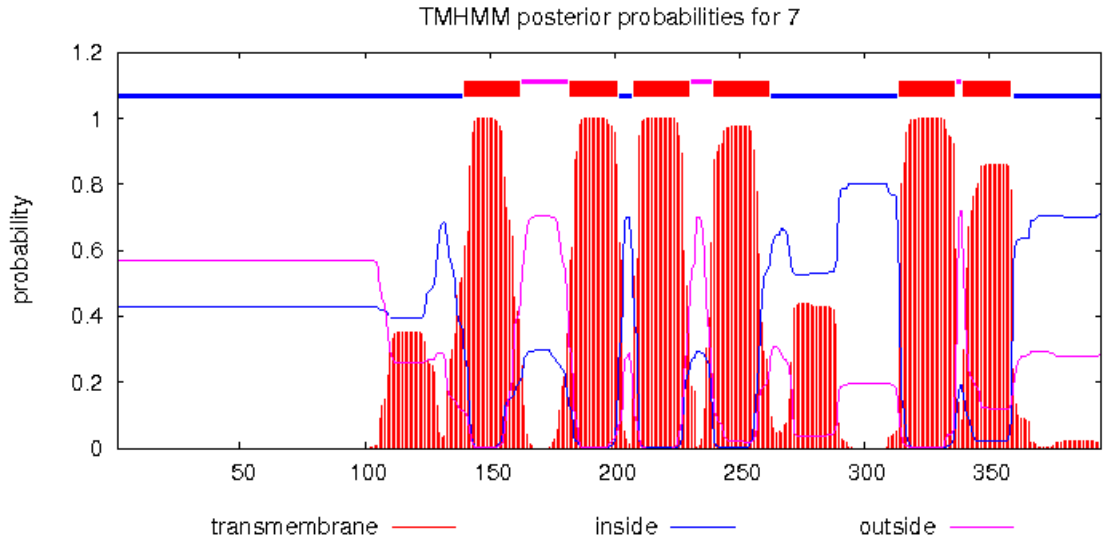



Figure 34: TMHMM analysis for PT1p (*Cannabis sativa* L.) (http://www.cbs.dtu.dk/services/TMHMM)

---

```
# 8 Length: 407
# 8 Number of predicted TMHs:  7
# 8 Exp number of AAs in TMHs: 181.32021
# 8 Exp number, first 60 AAs:  0.01835
# 8 Total prob of N-in:        0.34385
8       TMHMM2.0        outside      1   146
8       TMHMM2.0        TMhelix    147   169
8       TMHMM2.0        inside     170   191
8       TMHMM2.0        TMhelix    192   214
8       TMHMM2.0        outside    215   218
8       TMHMM2.0        TMhelix    219   241
8       TMHMM2.0        inside     242   247
8       TMHMM2.0        TMhelix    248   270
8       TMHMM2.0        outside    271   279
8       TMHMM2.0        TMhelix    280   302
8       TMHMM2.0        inside     303   327
8       TMHMM2.0        TMhelix    328   350
8       TMHMM2.0        outside    351   383
8       TMHMM2.0        TMhelix    384   406
8       TMHMM2.0        inside     407   407
```

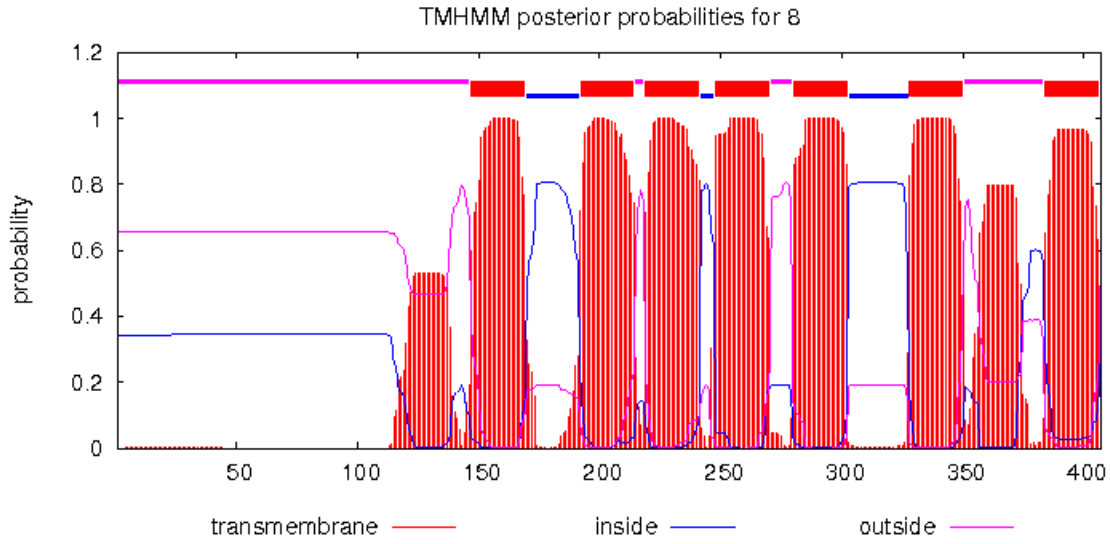



Figure 35: TMHMM analysis for aromatic prenyltransferase (*Epimedium acuminatum*) (http://www.cbs.dtu.dk/services/TMHMM)

```
# 10 Length: 296
# 10 Number of predicted TMHs:  0
# 10 Exp number of AAs in TMHs: 14.24509
# 10 Exp number, first 60 AAs:  8.80427
# 10 Total prob of N-in:        0.42520
10      TMHMM2.0        outside      1   296
```

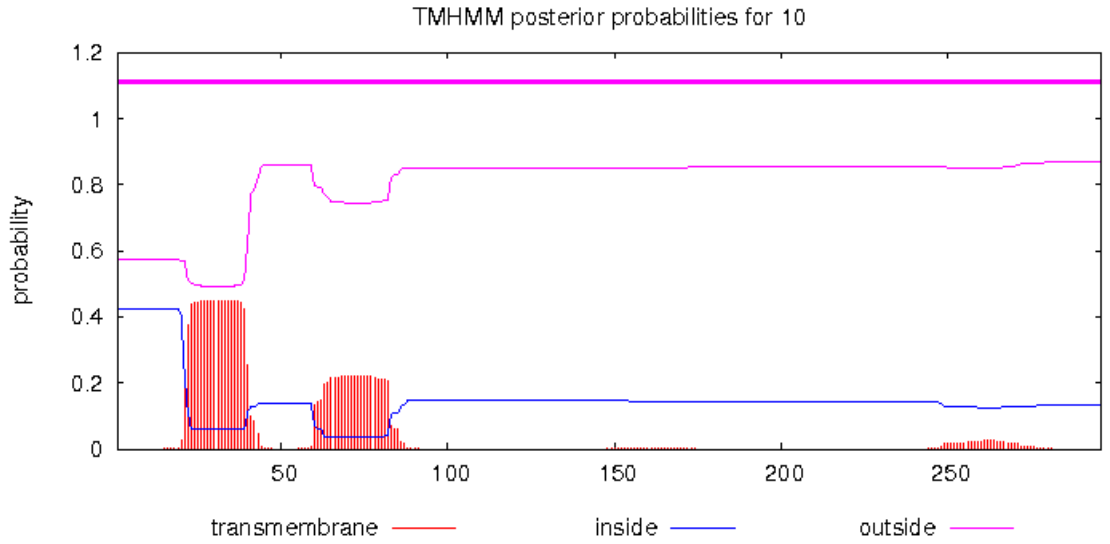



Figure 36: TMHMM analysis for cis-prenyltransferase (*Hevea brasiliensis*) (http://www.cbs.dtu.dk/services/TMHMM)

```
# 11 Length: 411
# 11 Number of predicted TMHs:  8
# 11 Exp number of AAs in TMHs: 158.21045
# 11 Exp number, first 60 AAs:  0.04118
# 11 Total prob of N-in:        0.69557
11      TMHMM2.0        inside      1   152
11      TMHMM2.0        TMhelix   153   175
11      TMHMM2.0        outside   176   196
11      TMHMM2.0        TMhelix   197   219
11      TMHMM2.0        inside    220   225
11      TMHMM2.0        TMhelix   226   248
11      TMHMM2.0        outside   249   251
11      TMHMM2.0        TMhelix   252   274
11      TMHMM2.0        inside    275   280
11      TMHMM2.0        TMhelix   281   303
11      TMHMM2.0        outside   304   330
11      TMHMM2.0        TMhelix   331   353
11      TMHMM2.0        inside    354   359
11      TMHMM2.0        TMhelix   360   377
11      TMHMM2.0        outside   378   391
11      TMHMM2.0        TMhelix   392   410
11      TMHMM2.0        inside    411   411
```

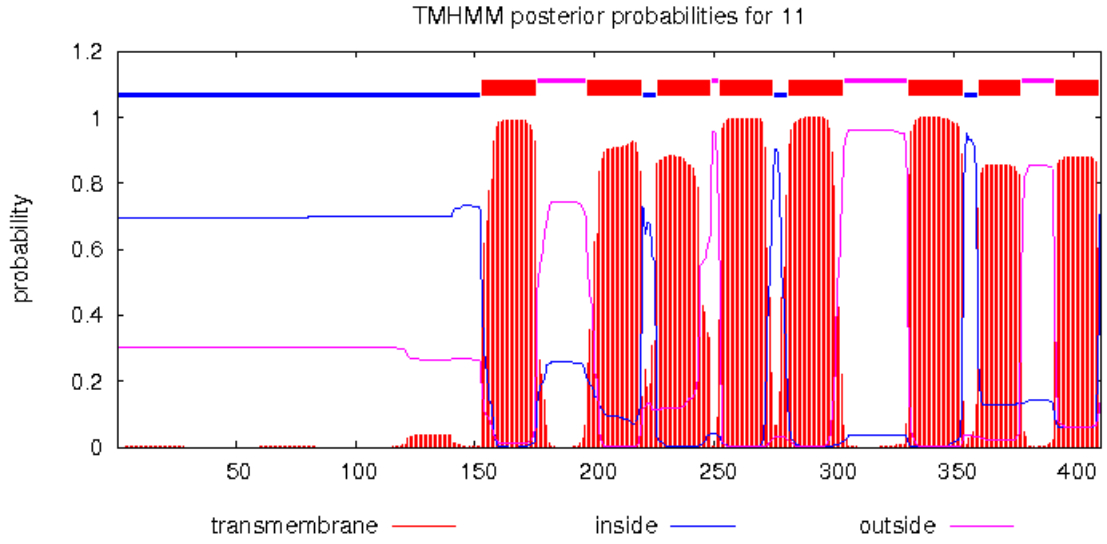



Figure 37: TMHMM analysis for aromatic prenyltransferase (*Humulus lupulus*) (http://www.cbs.dtu.dk/services/TMHMM)

---

```
# 12 Length: 308
# 12 Number of predicted TMHs:  0
# 12 Exp number of AAs in TMHs: 5.77091
# 12 Exp number, first 60 AAs:  5.57655
# 12 Total prob of N-in:        0.24917
12      TMHMM2.0        outside     1   308
```

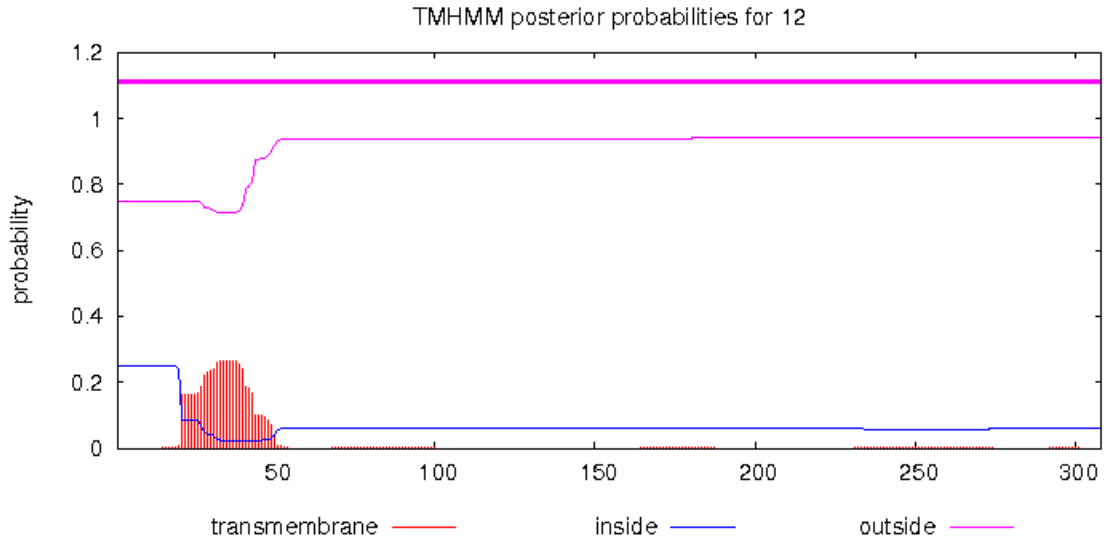



Figure 38: TMHMM analysis for cis-prenyltransferase (*Lilium longiflorum*) (http://www.cbs.dtu.dk/services/TMHMM)

---

```
# 13 Length: 478
# 13 Number of predicted TMHs:  0
# 13 Exp number of AAs in TMHs: 0.293319999999999999
# 13 Exp number, first 60 AAs:  0.05752
# 13 Total prob of N-in:        0.01188
13      TMHMM2.0        outside      1   478
```

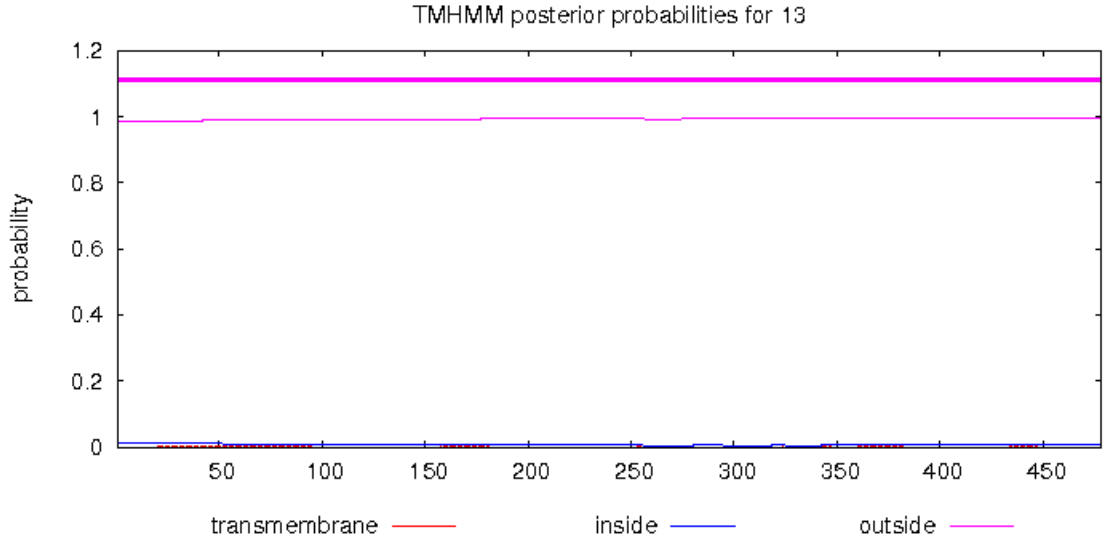



Figure 39: TMHMM analysis for protein prenyltransferase (*Oryza sativa*) (http://www.cbs.dtu.dk/services/TMHMM)

```
# 14 Length: 335
# 14 Number of predicted TMHs:  7
# 14 Exp number of AAs in TMHs: 145.18004
# 14 Exp number, first 60 AAs:  19.51683
# 14 Total prob of N-in:        0.34704
# 14 POSSIBLE N-term signal sequence
14      TMHMM2.0       outside      1    23
14      TMHMM2.0       TMhelix     24    46
14      TMHMM2.0       inside      47    66
14      TMHMM2.0       TMhelix     67    89
14      TMHMM2.0       outside     90    92
14      TMHMM2.0       TMhelix     93   110
14      TMHMM2.0       inside     111   121
14      TMHMM2.0       TMhelix    122   139
14      TMHMM2.0       outside    140   148
14      TMHMM2.0       TMhelix    149   171
14      TMHMM2.0       inside     172   191
14      TMHMM2.0       TMhelix    192   214
14      TMHMM2.0       outside    215   217
14      TMHMM2.0       TMhelix    218   240
14      TMHMM2.0       inside     241   335
```

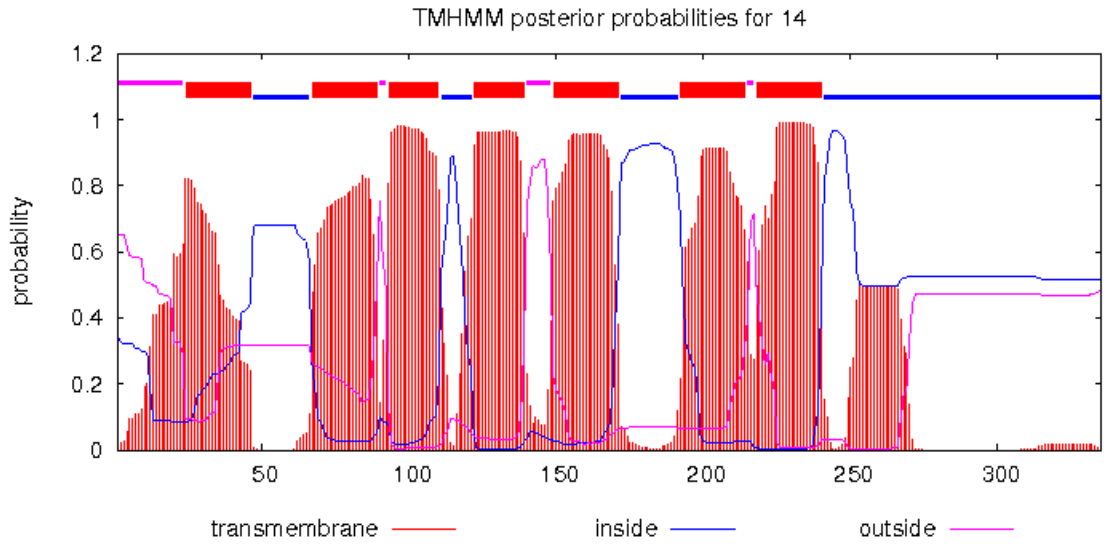



Figure 40: TMHMM analysis for prenyltransferase (*Populus trichocarpa*) (http://www.cbs.dtu.dk/services/TMHMM)

```
# 15 Length: 407
# 15 Number of predicted TMHs:  9
# 15 Exp number of AAs in TMHs: 184.40543
# 15 Exp number, first 60 AAs:  0.00791
# 15 Total prob of N-in:        0.78332
15      TMHMM2.0       inside       1   120
15      TMHMM2.0       TMhelix    121   138
15      TMHMM2.0       outside    139   147
```

```
15      TMHMM2.0        TMhelix    148   170
15      TMHMM2.0        inside     171   191
15      TMHMM2.0        TMhelix    192   214
15      TMHMM2.0        outside    215   218
15      TMHMM2.0        TMhelix    219   241
15      TMHMM2.0        inside     242   247
15      TMHMM2.0        TMhelix    248   270
15      TMHMM2.0        outside    271   284
15      TMHMM2.0        TMhelix    285   302
15      TMHMM2.0        inside     303   327
15      TMHMM2.0        TMhelix    328   347
15      TMHMM2.0        outside    348   351
15      TMHMM2.0        TMhelix    352   374
15      TMHMM2.0        inside     375   386
15      TMHMM2.0        TMhelix    387   406
15      TMHMM2.0        outside    407   407
```

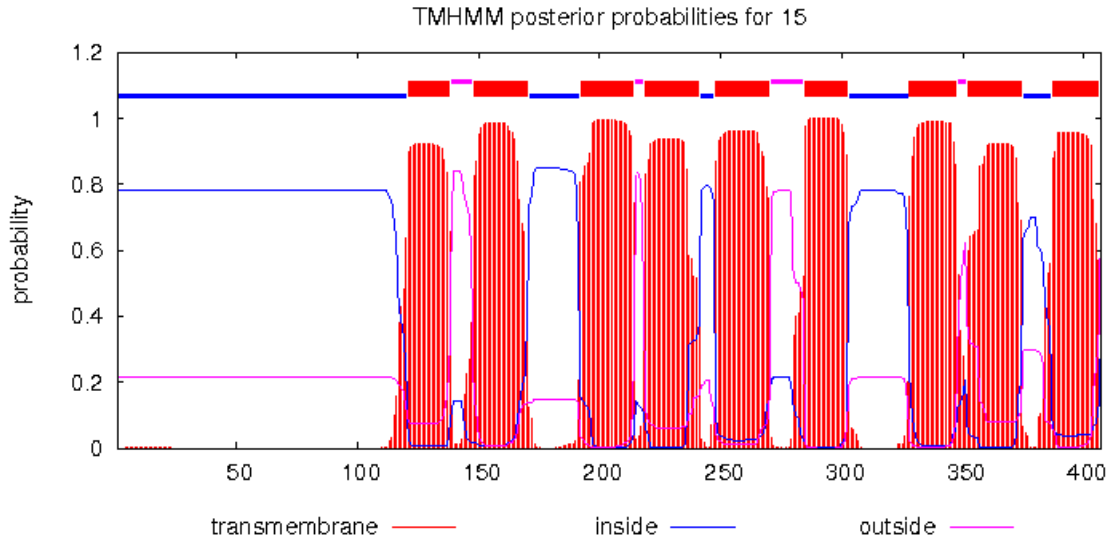



Figure 41: TMHMM analysis for prenyltransferase (*Sophora flavescens*) (http://www.cbs.dtu.dk/services/TMHMM)

---

```
# 16 Length: 333
# 16 Number of predicted TMHs:  8
# 16 Exp number of AAs in TMHs: 150.34255
# 16 Exp number, first 60 AAs:  20.20062
# 16 Total prob of N-in:        0.36984
# 16 POSSIBLE N-term signal sequence
16      TMHMM2.0        outside     1    22
16      TMHMM2.0        TMhelix    23    45
16      TMHMM2.0        inside     46    65
16      TMHMM2.0        TMhelix    66    88
16      TMHMM2.0        outside    89    91
16      TMHMM2.0        TMhelix    92   111
16      TMHMM2.0        inside    112   117
16      TMHMM2.0        TMhelix   118   137
16      TMHMM2.0        outside   138   146
16      TMHMM2.0        TMhelix   147   169
16      TMHMM2.0        inside    170   189
16      TMHMM2.0        TMhelix   190   209
```

```
16      TMHMM2.0        outside    210   213
16      TMHMM2.0        TMhelix    214   236
16      TMHMM2.0        inside     237   248
16      TMHMM2.0        TMhelix    249   266
16      TMHMM2.0        outside    267   333
```

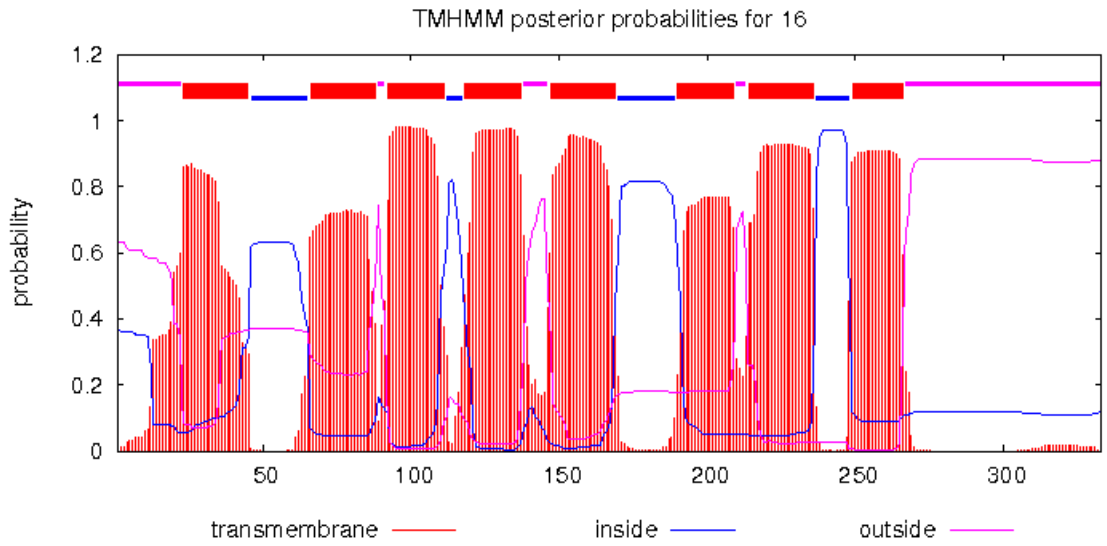



Figure 42: TMHMM analysis for protoheme IX farnesyltransferase (*Ricinus communis*) (http://www.cbs.dtu.dk/services/TMHMM)

---

```
# 17 Length: 407
# 17 Number of predicted TMHs:  9
# 17 Exp number of AAs in TMHs: 184.40543
# 17 Exp number, first 60 AAs:  0.00791
# 17 Total prob of N-in:        0.78332
17      TMHMM2.0        inside       1   120
17      TMHMM2.0        TMhelix    121   138
17      TMHMM2.0        outside    139   147
17      TMHMM2.0        TMhelix    148   170
17      TMHMM2.0        inside     171   191
17      TMHMM2.0        TMhelix    192   214
17      TMHMM2.0        outside    215   218
17      TMHMM2.0        TMhelix    219   241
17      TMHMM2.0        inside     242   247
17      TMHMM2.0        TMhelix    248   270
17      TMHMM2.0        outside    271   284
17      TMHMM2.0        TMhelix    285   302
17      TMHMM2.0        inside     303   327
17      TMHMM2.0        TMhelix    328   347
17      TMHMM2.0        outside    348   351
17      TMHMM2.0        TMhelix    352   374
17      TMHMM2.0        inside     375   386
17      TMHMM2.0        TMhelix    387   406
17      TMHMM2.0        outside    407   407
```

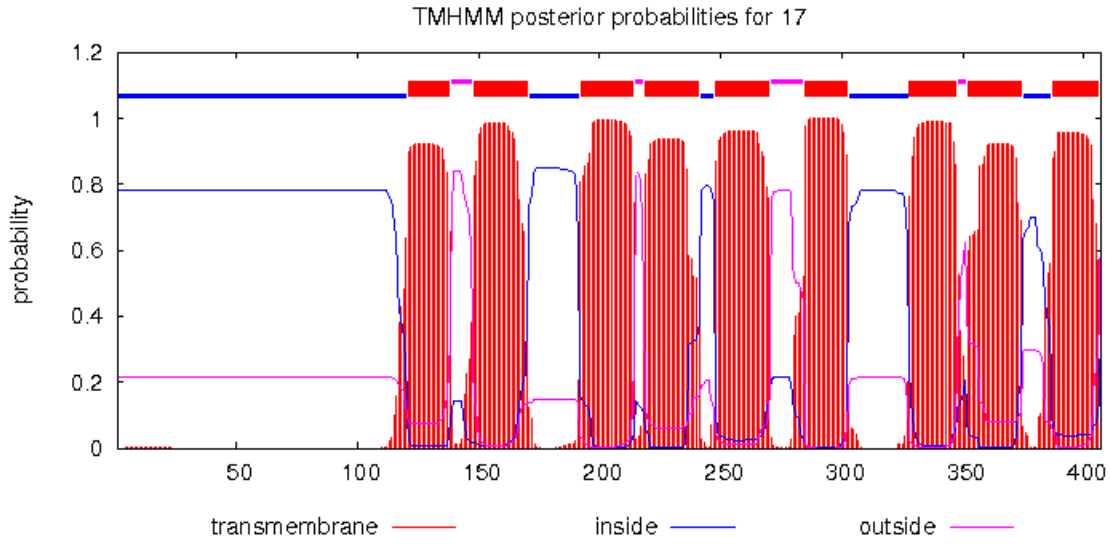



Figure 43: TMHMM analysis for flavonoid prenyltransferase (*Sophora flavescens*) (http://www.cbs.dtu.dk/services/TMHMM)

---

```
# 18 Length: 425
# 18 Number of predicted TMHs:  0
# 18 Exp number of AAs in TMHs: 0.15131
# 18 Exp number, first 60 AAs:  0.15005
# 18 Total prob of N-in:        0.00521
18      TMHMM2.0        outside      1   425
```

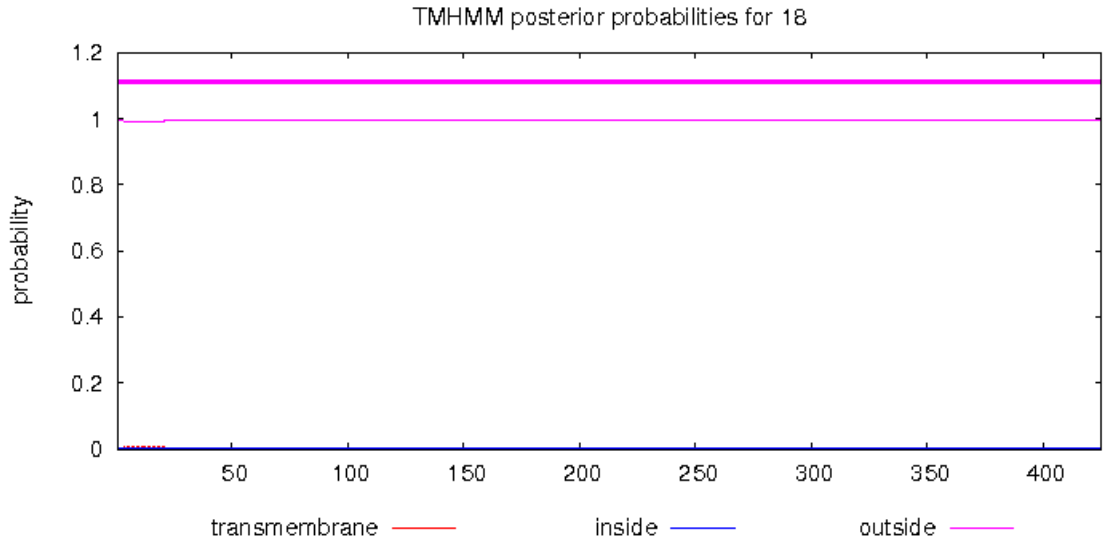
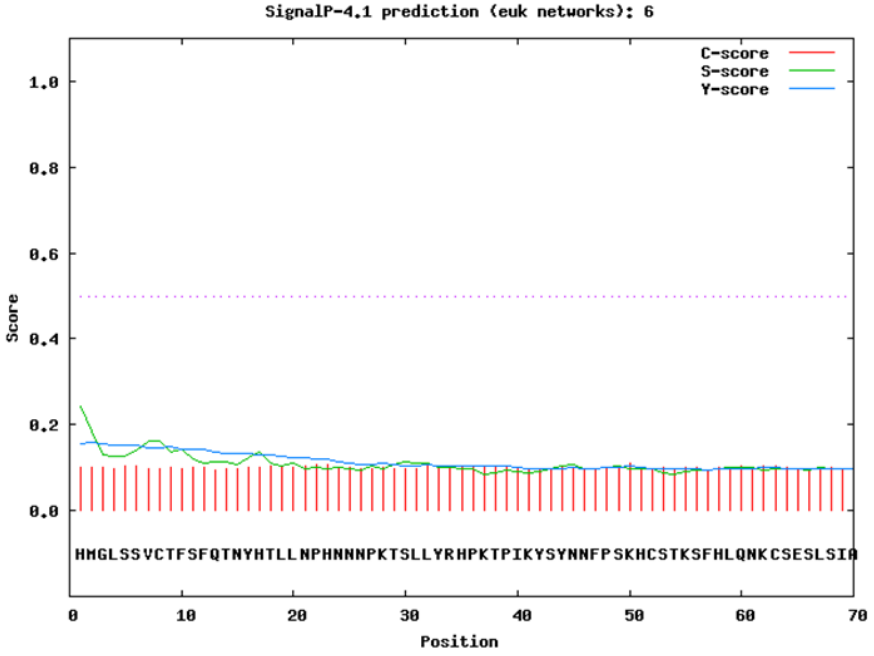



Figure 44: TMHMM analysis for prenyltransferase (*Zea mays*) (http://www.cbs.dtu.dk/services/TMHMM)




| Measure | Position | Value | Cutoff | signal peptide? |
|---|---|---|---|---|
| max. C | 50 | 0.109 | | |
| max. Y | 11 | 0.144 | | |
| max. S | 1 | 0.244 | | |
| mean S | 1-10 | 0.155 | | |
| D | 1-10 | 0.150 | 0.450 | NO |

Name=6 SP='NO' D=0.150 D-cutoff=0.450 Networks=SignalP-noTM

Figure 45: Analysis for various prenyltransferase protein sequences (Table 42); Query in prediction tool SignalP: C-score: raw cleavage site score; S-score: signal peptide score; Y-score: combined cleavage site score; D-score: discrimination score

```
CBGAS_native      1 atgggactctcatcagtttgtacctttttcatttcaaactaattaccatactttattaaatcctcacaataataatcccaaaacctcattattat---atc
CBGAS_patent      1 atgggactctcatcagtttgtacctttttcatttcaaactaattaccatactttattaaatcctcacaataataatcccaaaacctcattattatgttatc

CBGAS_native     98 gacaccccaaaacaccaattaaatactcttacaataatttttccctctaaacattgctccaccaagagttttcatctacaaaacaaatgctcagaatcatt
CBGAS_patent    101 gacaccccaaaacaccaattaaatactcttacaataatttttccctctaaacattgctccaccaagagttttcatctacaaaacaaatgctcagaatcatt

CBGAS_native    198 atcaatcgcaaaaaattccattagggcagctactacaaatcaaactgagcctccagaatctgataatcattcagtagcaactaaaattttaaacttttggg
CBGAS_patent    201 atcaatcgcaaaaaattccattagggcagctactacaaatcaaactgagcctccagaatctgataatcattcagtagcaactaaaattttaaacttttggg

CBGAS_native    298 aaggcatgttggaagcttcaaagaccatatacaatcatagcatttacttcatgcgcttgtggattgtttgggaaagagttgttgcataacacaaatttaa
CBGAS_patent    301 aaggcatgttggaaacttcaaagaccatatacaatcatagcatttacttcatgcgcttgtggattgtttgggaaagagttgttgcataacacaaatttaa

CBGAS_native    398 taagttggtctctcatgttcaaggcattctttttttggtggctgtattatgcattgcttcttttacaactaccatcaatcagatttacgatcttcacat
CBGAS_patent    401 taagttggtctctgatgttcaaggcattctttttttggtggctatattatgcattgcttcttttacaactaccatcaatcagatttacgatcttcacat

CBGAS_native    498 tgacagaataaacaagcctgatctaccactagcttcaggggaaatatcagtaaacacagcttggattatgagcataattgtggcactgtttggattgata
CBGAS_patent    501 tgacagaataaacaagcctgatctaccactagcttcaggggaaatatcagtaaacacagcttggattatgagcataattgtggcactgtttggattgata

CBGAS_native    598 ataactataaaaatgaagggtggaccactctatatatttggctactgttttggtatttttggtgggactgtctattctgttccaccatttagatggaagc
CBGAS_patent    601 ataactataaaaatgaagggtggaccactctatatatttggctactgttttggtatttttggtgggattgtctattctgttccaccatttagatggaagc

CBGAS_native    698 aaaatccttccactgcatttcttctcaatttcctggcccatattattacaaatttcacattttatcatgccagcagagcagctcttggcctaccatttga
CBGAS_patent    701 aaaatccttccactgcatttcttctcaatttcctggcccatattattacaaatttcacattttattatgccagcagagcagctcttggcctaccatttga

CBGAS_native    798 gttgaggccttcttttactttcctgctagcatttatgaaatcaatgggttcagctttagctttaatcaaagatgcttcagacgttgaaggcgacactaaa
CBGAS_patent    801 gttgaggccttcttttactttcctgctagcatttatgaaatcaatgggttcagctttggctttaatcaaagatgcttcagacgttgaaggcgacactaaa

CBGAS_native    898 tttggcatatcaaccttggcaagtaaatatggttccagaaacttgacattattttgttctggaattgttctcctatcctatgtggctgctatacttgctg
CBGAS_patent    901 tttggcatatcaaccttggcaagtaaatatggttccagaaacttgacattattttgttctggaattgttctcctatcctatgtggctgctatacttgctg

CBGAS_native    998 ggattatctggccccaggctttcaacagtaacgtaatgttactttctcatgcaatcttagcattttggttaatcctccagactcgagattttgcgttaac
CBGAS_patent   1001 ggattatctggccccaggctttcaacagtaacgtaatgttactttctcatgcaatcttagcattttggttaatcctccagactcgagattttgcgttaac

CBGAS_native   1098 aaattatgacccggaagcaggcagaagattttacgagttcatgtggaagctttattatgctgaatatttagtatatgttttcatataa
CBGAS_patent   1101 aaattacgacccggaagcaggcagaagattttacgagttcatgtggaagctttattatgctgaatatttagtatatgttttcatataa
```

Figure 46: Multiple alignment of cDNA sequence of CBGA-Synthase from working group and patent-registered prenyltransferase from *Cannabis sativa* L.

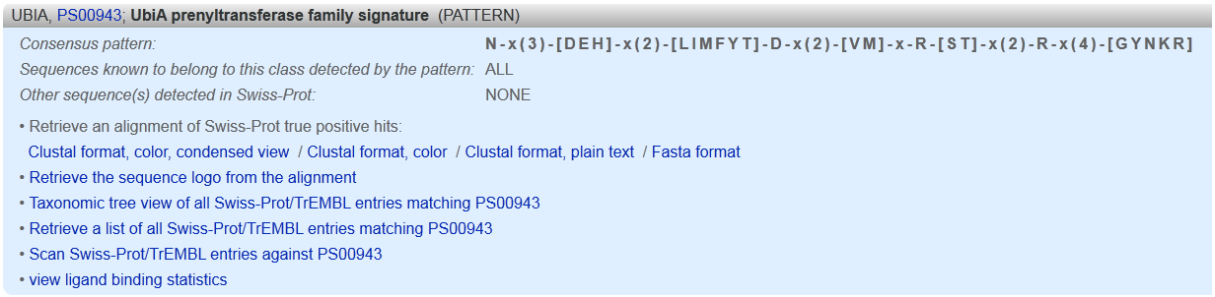



Figure 47: Motif pattern for UbiA prenyltransferases (PF01040)

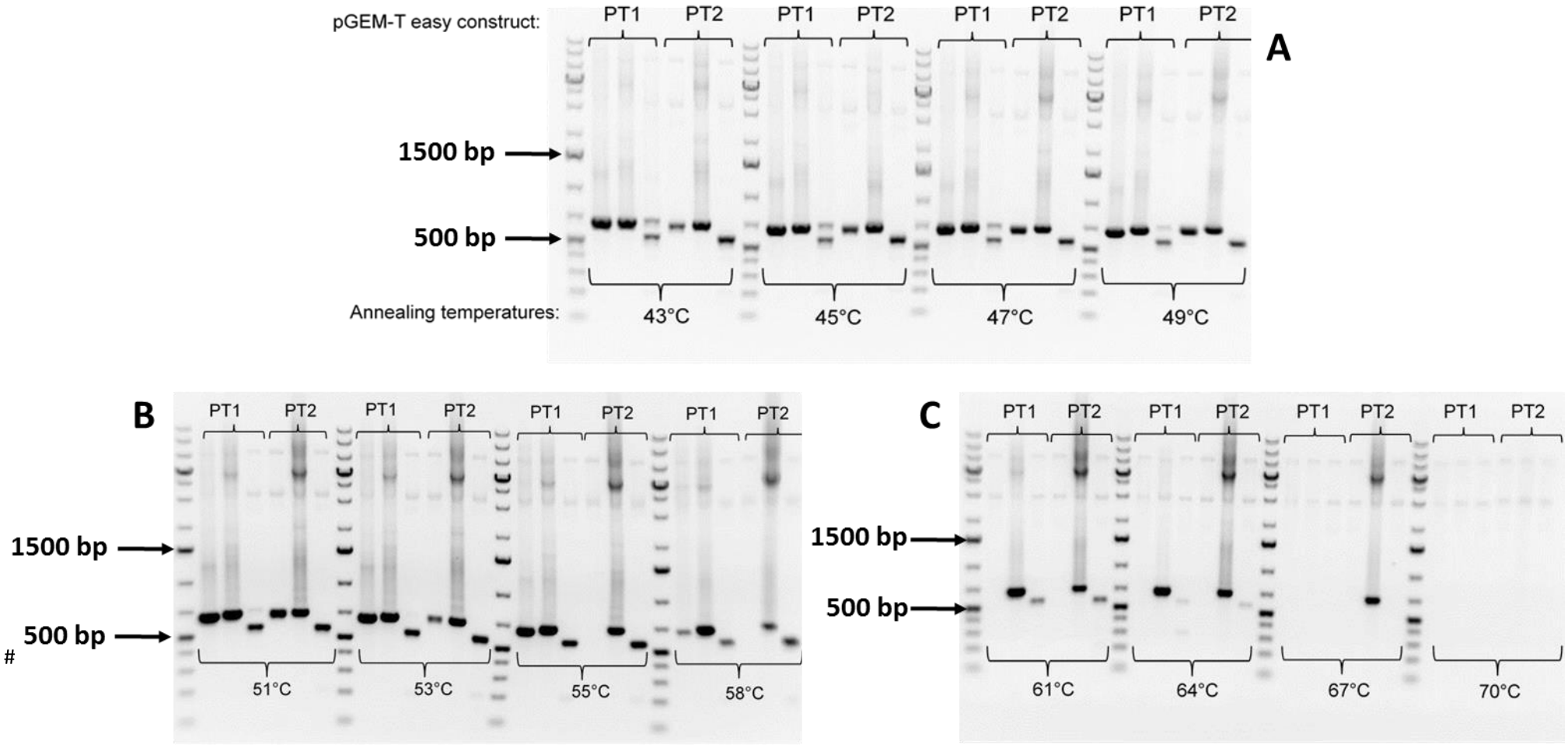



Figure 48: Optimization of annealing temperature for primer/enzyme pairs detecting PT1 and 2; templates used were cloning vector pGEM-T easy carrying PT1 with the native signal peptide or cloning vector carrying PT2 with the native signal peptide; template used are described above the brackets; used primers are alternating between primer pair for PT1 (lane 2, 5, 9, 12, 16, 19 etc., always the left lane), for PT2 (lane 3, 6, 10, 13, 17, etc. always the middle lane) and for PT3 (lane 4, 11, 14, 18, etc. always the right lane); DNA Ladder (1kb, Thermo Scientific, Waltham, Massachusetts, USA) is present in lane 1, 8, 15, etc.; used annealing temperatures range from 43°C-70°C and are described below the brackets.

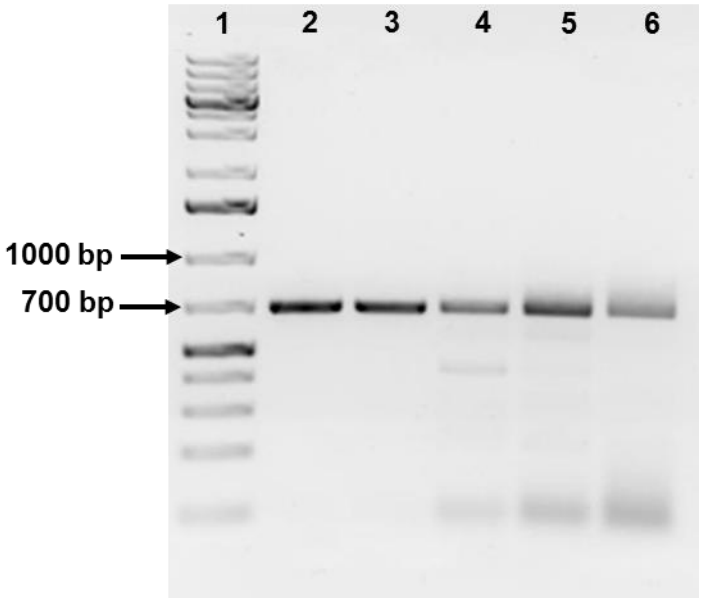



Figure 49: Agarose gel from RNA extraction with subsequent cDNA synthesis and PCR with primer pair for gapdh (Table 37); gapdh serves as positive control; lane 1: DNA ladder, 1kb (Thermo Fisher Scientific, Waltham, Massachusetts, USA) lane 2: trichomes; lane 3: flower tissue; lane 4: Leaves; lane 5: branch tissue; lane 6: stem tissue.

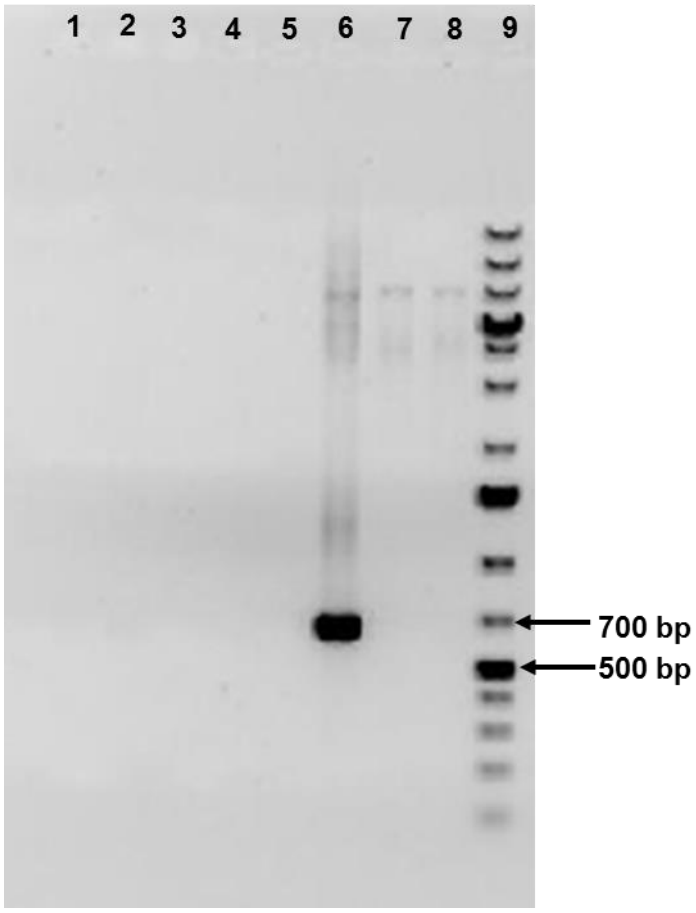



Figure 50: Agarose gel from RNA extraction with subsequent cDNA synthesis and PCR with primer pair for PT1 (Table 37); lane 1: trichomes; lane 2: flower tissue; lane 3: leaves; lane 4: branch tissue;

lane 5: stem tissue; lane 6: pGEM-T easy carrying PT1 with native signal peptide; lane 7: pGEM-T easy carrying PT2 with native signal peptide; lane 8: pGEM-T easy carrying PT3 with native signal peptide; lane 9: DNA ladder, 1kb (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

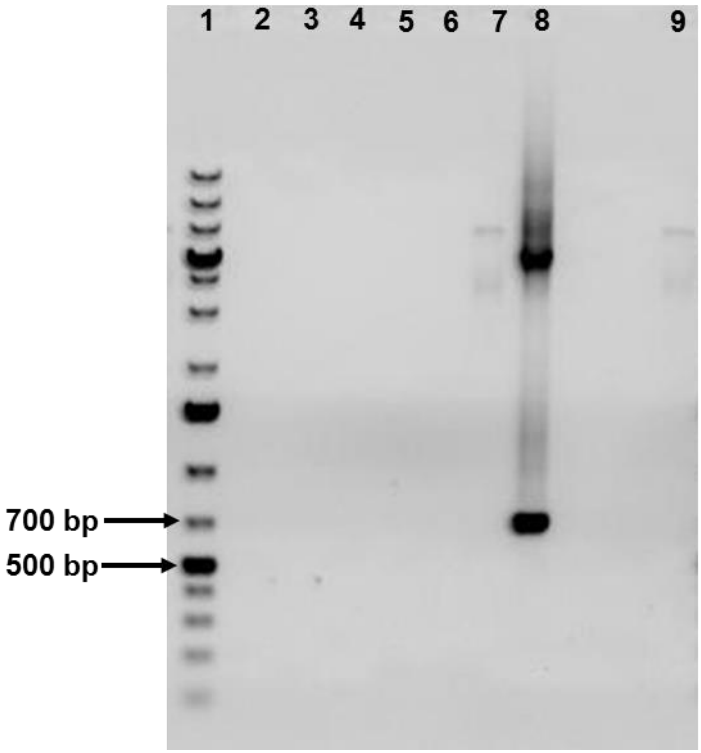



Figure 51: Agarose gel from RNA extraction with subsequent cDNA synthesis and PCR with primer pair for PT2 (Table 37); lane 1: DNA ladder, 1kb (Thermo Fisher Scientific, Waltham, Massachusetts, USA) lane 2: trichomes; lane 3: flower tissue; lane 4: Leaves; lane 5: branch tissue; lane 6: stem tissue; lane 7: pGEM-T easy carrying PT1 with native signal peptide; lane 8: pGEM-T easy carrying PT2 with native signal peptide; lane 9: pGEM-T easy carrying PT3 with native signal peptide.

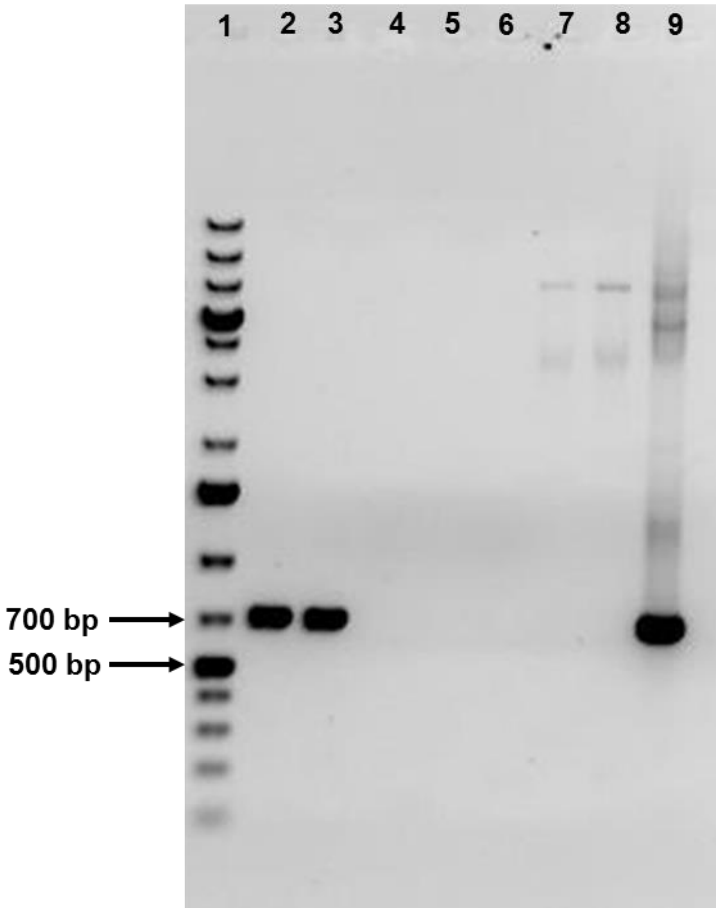



Figure 52: Agarose gel from RNA extraction with subsequent cDNA synthesis and PCR with primer pair for PT3 (Table 37); lane 1: DNA ladder, 1kb (Thermo Fisher Scientific, Waltham, Massachusetts, USA) lane 2: trichomes; lane 3: flower tissue; lane 4: Leaves; lane 5: branch tissue; lane 6: stem tissue; lane 7: pGEM-T easy carrying PT1 with native signal peptide; lane 8: pGEM-T easy carrying PT2 with native signal peptide; lane 9: pGEM-T easy carrying PT3 with native signal peptide.

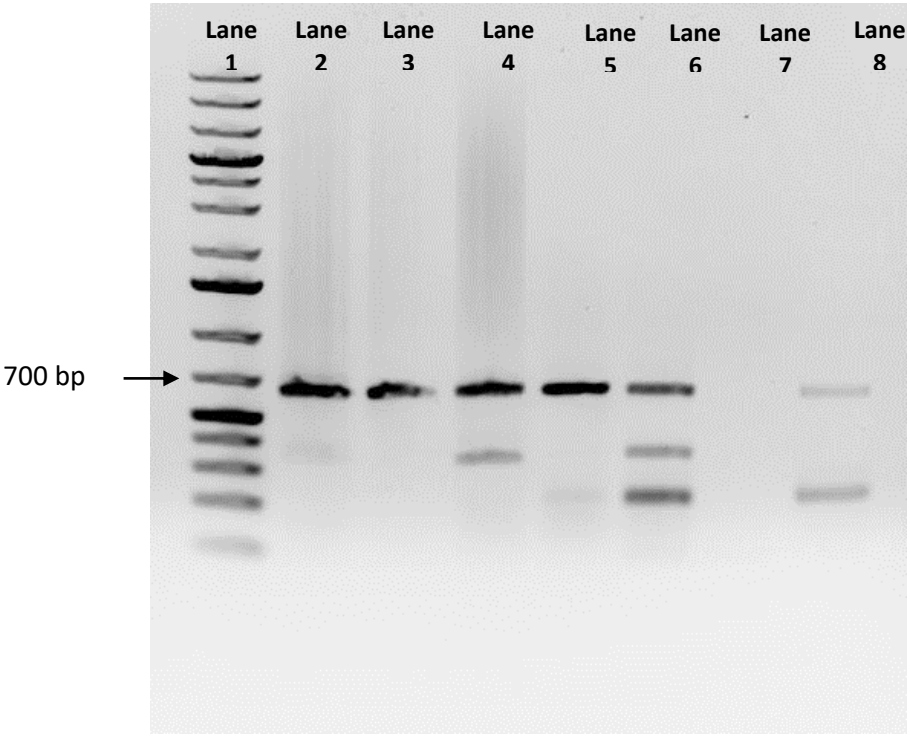



Figure 53: Agarose gel from RNA extraction with subsequent cDNA synthesis and PCR with primer pair for PT4 (Table 37); lane 1: DNA ladder, 1kb (Thermo Fisher Scientific, Waltham, Massachusetts, USA) lane 2: flower tissue; lane 3: trichomes; lane 4: Leaves; lane 5: branch tissue; lane 6: stem tissue; lane 7: control 1, cDNA synthesis without RNA template, lane 8: control 2, cDNA synthesis without Reverse Transcriptase.

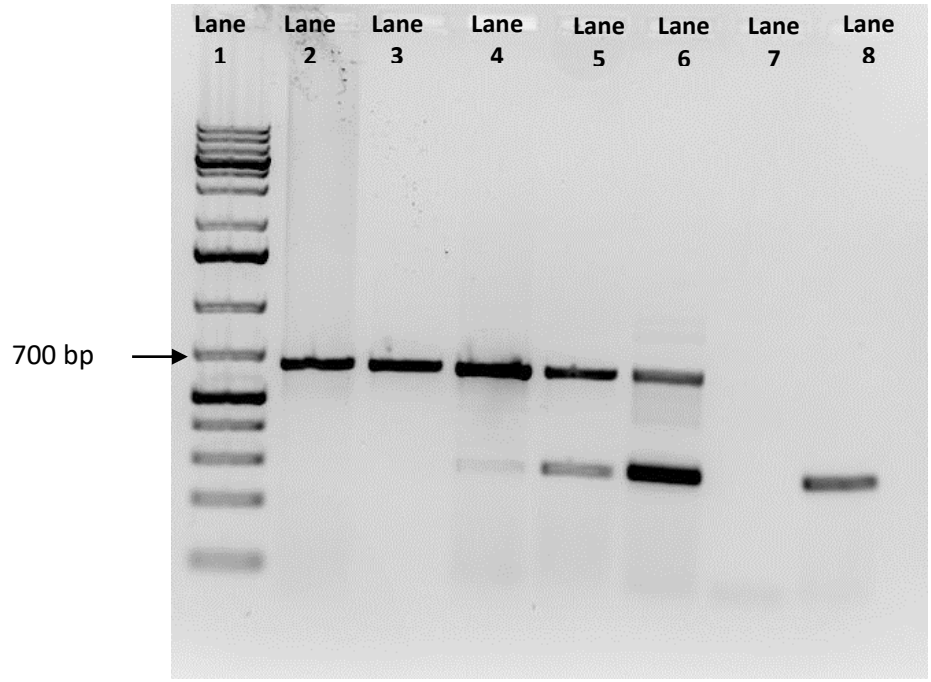



Figure 54: Agarose gel from RNA extraction with subsequent cDNA synthesis and PCR with primer pair for PT5 (Table 37); lane 1: DNA ladder, 1kb (Thermo Fisher Scientific, Waltham, Massachusetts, USA) lane 2: flower tissue; lane 3: trichomes; lane 4: Leaves; lane 5: branch tissue; lane 6: stem tissue; lane 7: control 1, cDNA synthesis without RNA template, lane 8: control 2, cDNA synthesis without Reverse Transcriptase.

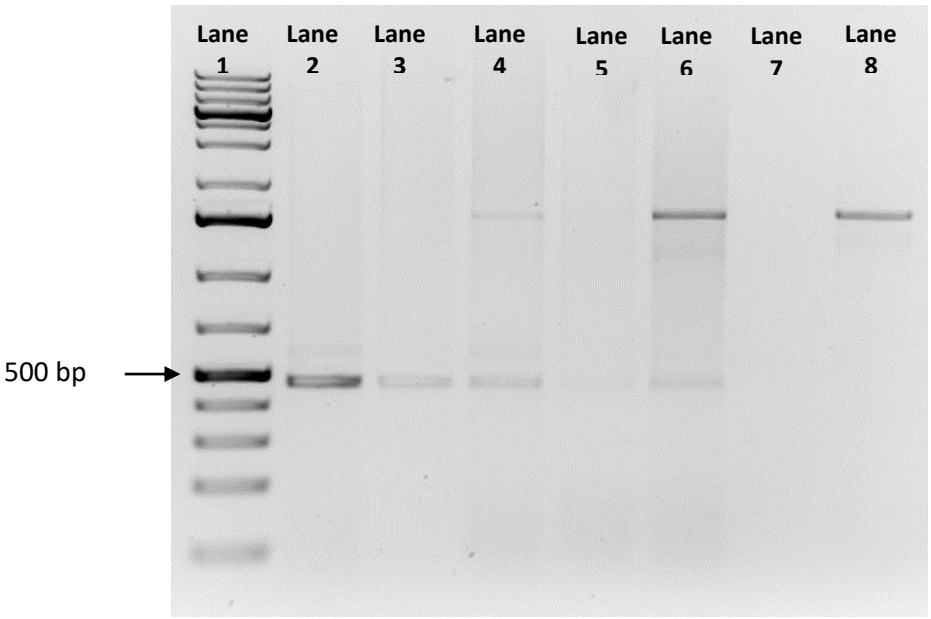



Figure 55: Agarose gel from RNA extraction with subsequent cDNA synthesis and PCR with primer pair for PT6 (Table 37); lane 1: DNA ladder, 1kb (Thermo Fisher Scientific, Waltham, Massachusetts, USA) lane 2: flower tissue; lane 3: trichomes; lane 4: Leaves; lane 5: branch tissue; lane 6: stem tissue; lane 7: control 1, cDNA synthesis without RNA template, lane 8: control 2, cDNA synthesis without Reverse Transcriptase.

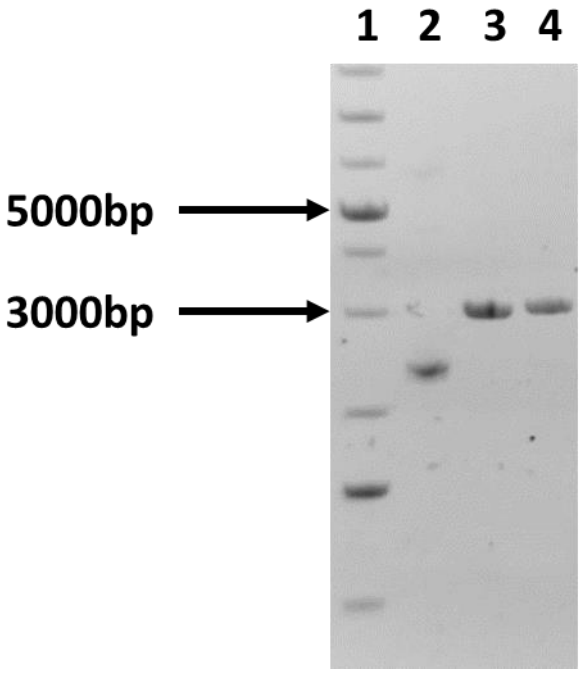



Figure 56: Agarose gel (1%); lane 1: DNA ladder, 1kb (Thermo Fisher Scientific, Waltham, Massachusetts, USA); lane 2: intact, supercoiled, modified pGEM-T easy vector; lane 3: modified pGEM-T easy vector after restriction reaction with endonuclease XcmI (New Englands Biolabs, Frankfurt am Main, Germany); lane 4: modified pGEM-T easy after restriction reaction with endonuclease XbaI (New Englands Biolabs, Frankfurt am Main, Germany).

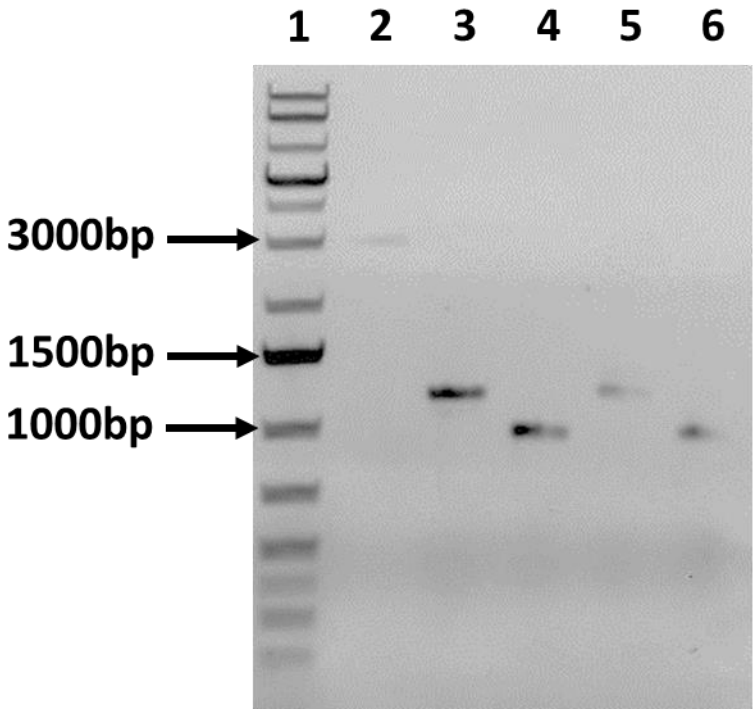



Figure 57: Agarose gel (1%); lane 1: DNA ladder, 1kb (Thermo Fisher Scientific, Waltham, Massachusetts, USA); lane 2: modified cloning vector pGEM-T easy after restriction with XcmI; lane 3:

PT1 with signal peptide; lane 4: PT1 without signal peptide; lane 5: PT2 with signal peptide; lane 5: PT2 without signal peptide; PTs were amplified from plant tissue cDNA and precipitated.

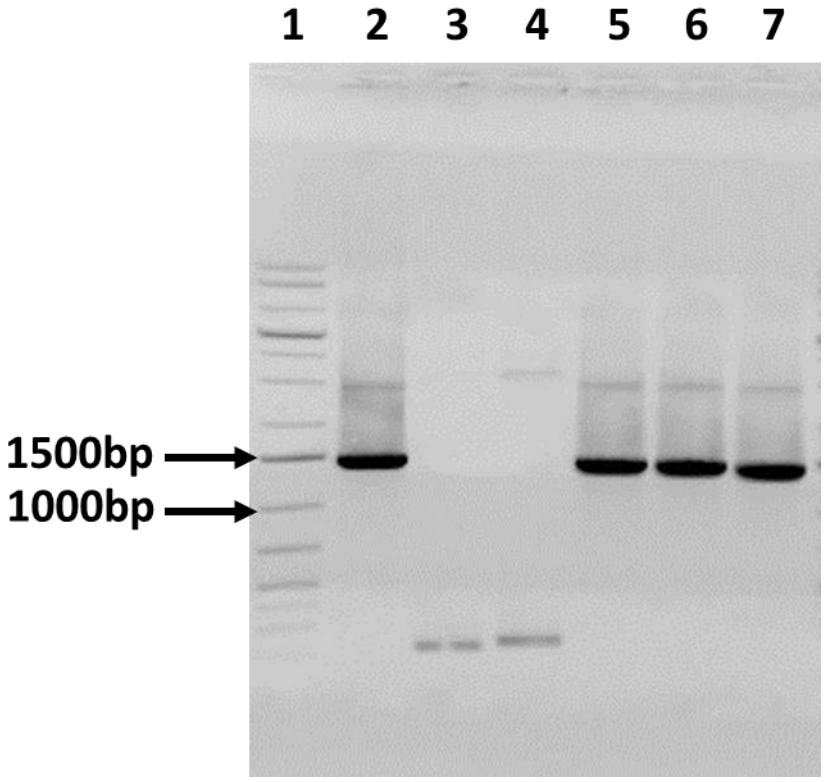



Figure 58: Agarose gel (1%); lane 1: DNA ladder, 1kb (Thermo Fisher Scientific, Waltham, Massachusetts, USA); lane 2-7: colony PCR of modified pGEM-T easy inserted PT1 with native signal peptide.

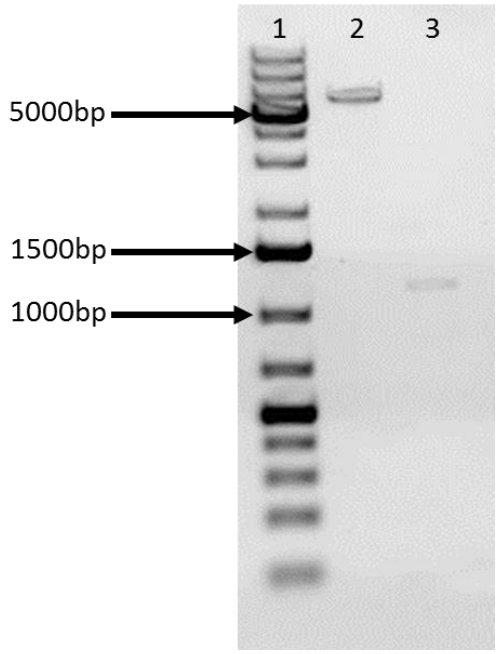



Figure 59: Agarose gel (1%); Lane 1: DNA ladder, 1kb (Thermo Fisher Scientific, Waltham Massachusetts, USA); Lane 2: pICH11599 after digestion with NcoI and BamHI (New England Biolabs GmbH, Frankfurt am Main, Germany); Lane 3: purified PT1 with signal peptide

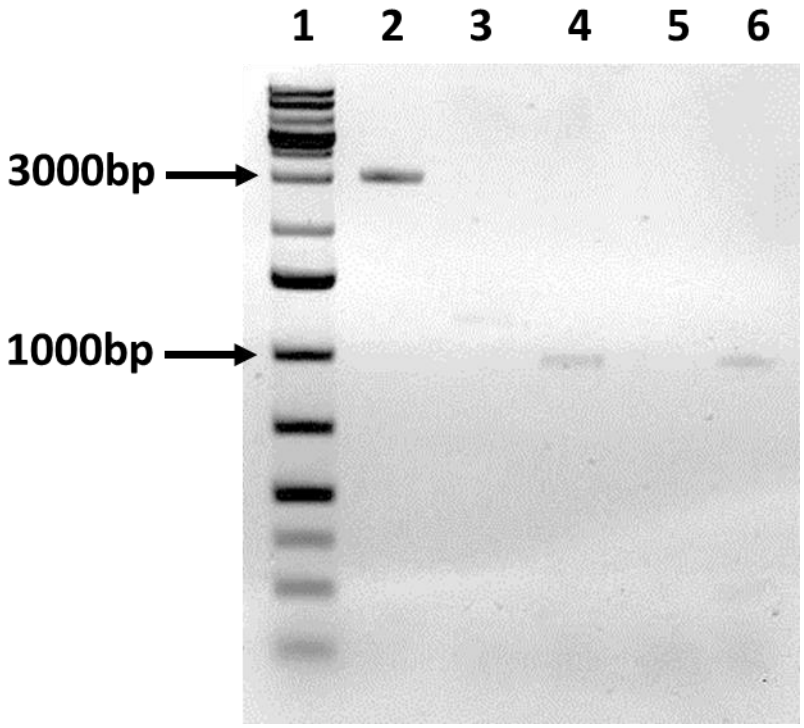



Figure 60: Agarose gel (1%); lane 1: DNA ladder, 1kb (Thermo Fisher Scientific, Waltham, Massachusetts, USA); lane 2: modified pGEM-T easy after restriction reaction with XcmI (New Englands Biolabs, Frankfurt am Main, Germany); lane 3: PCR product after amplification from cDNA

with PT1 primers with native signal peptide and subsequent gel purification; lane 4: PCR product after amplification from cDNA with PT1 primer without native signal peptide and subsequent gel purification; lane 5: PCR product after amplification from cDNA with PT2 primers with native signal peptide and subsequent gel purification; lane 6: PCR product after amplification from cDNA with PT2 primers without native signal peptide and subsequent gel purification (primers listed in Table 17).

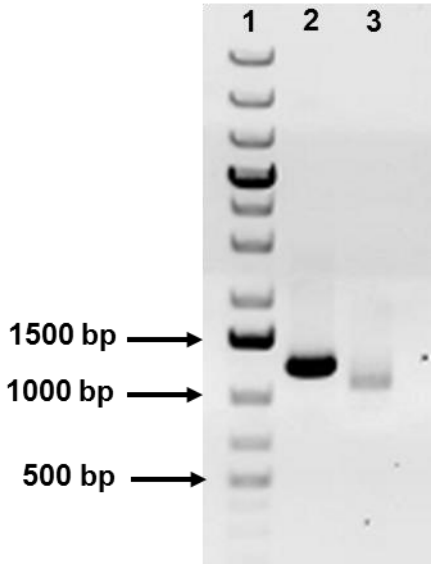



Figure 61: Agarose gel (1%); Lane 1: DNA Ladder 1kb (Thermo Fisher Scientific, Waltham, Massachusetts, USA); lane 2: Coq2 PCR using genomic DNA from CEN.PK2-1C (*MAT***a**; *ura3-52 MAL2-8C SUC2*; Euroscarf) as template; lane 3: PCR for amplification of PT6 with signal peptide from enzyme complex F1F0 ATP synthase using cDNA from *Cannabis sativa* L. Bediol as template.

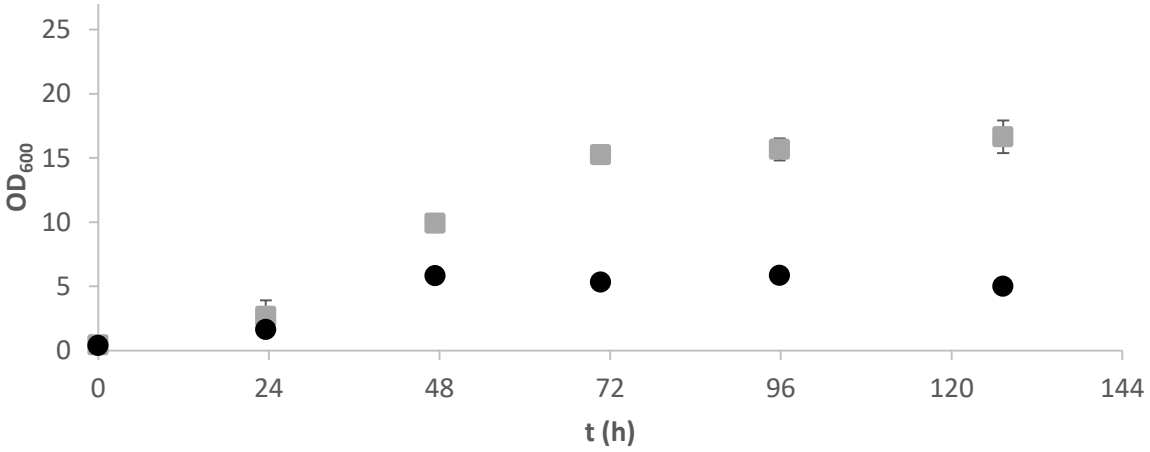



Figure 62: Growth curves: *Saccharomyces cerevisiae* Y05416 pDio-Coq2 (grey rectangles), *S. cerevisiae* Y05416 pDio (black dots); cultivation temperature: 18°C; YPD medium with 4% galactose.

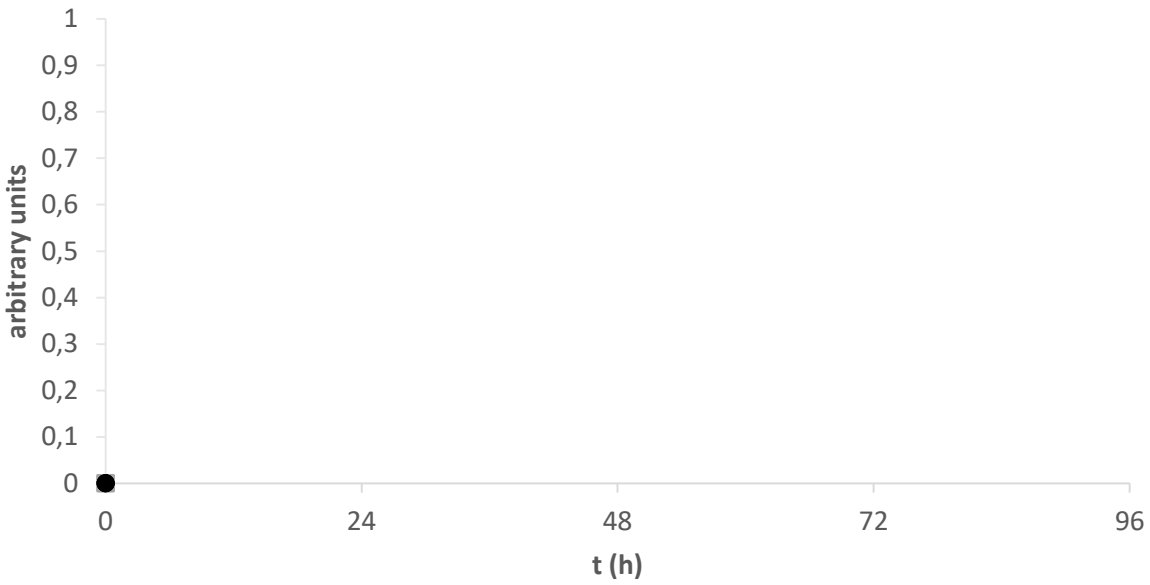



Figure 63: GHB production assay of *S. cerevisiae* Y05416 membrane fraction after cultivation; grey rectangles: pDio-Coq2 with GPP; black dots: pDio; cultivation temperature: 18°C; YPD medium with 4% galactose.

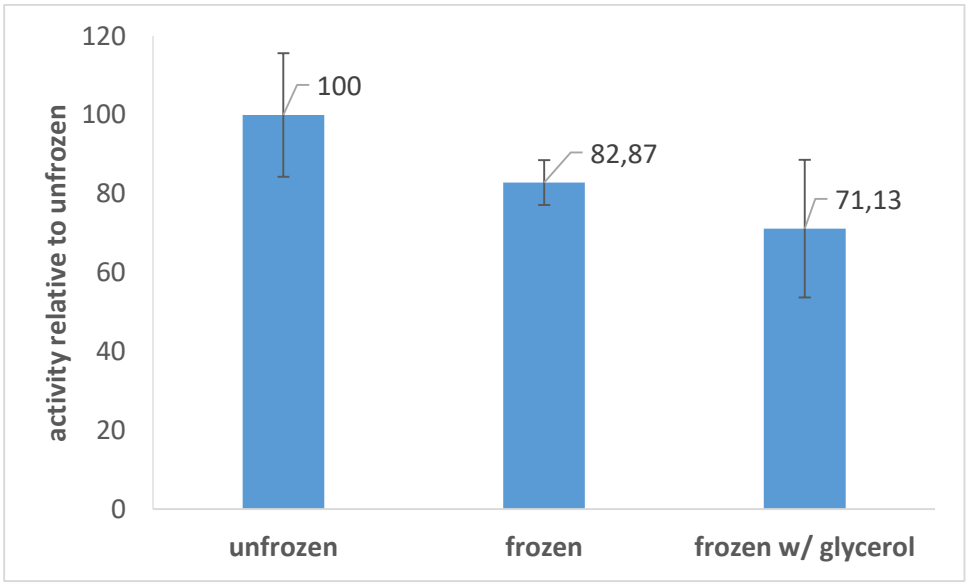



Figure 64: GHB production of Coq2 before and after freezing overnight; the membrane fraction was tested with and without addition of 50% glycerol prior to freezing.

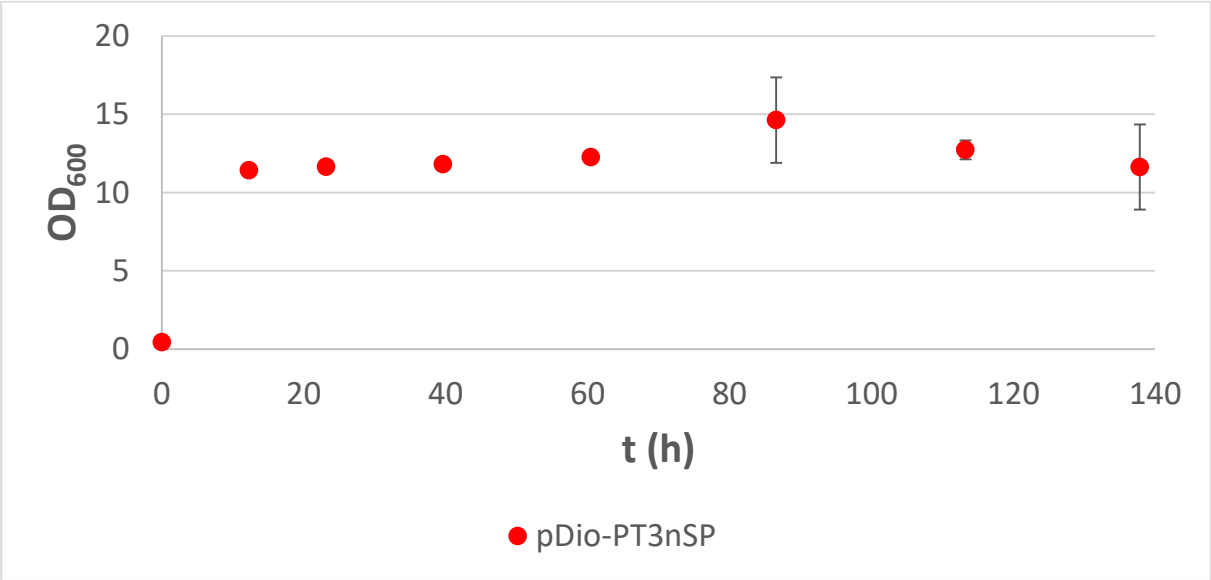



Figure 65: Growth curve: *Saccharomyces cerevisiae* Y05416 pDio-PT3nSP; cultivation temperature: 30°C; YPD medium.

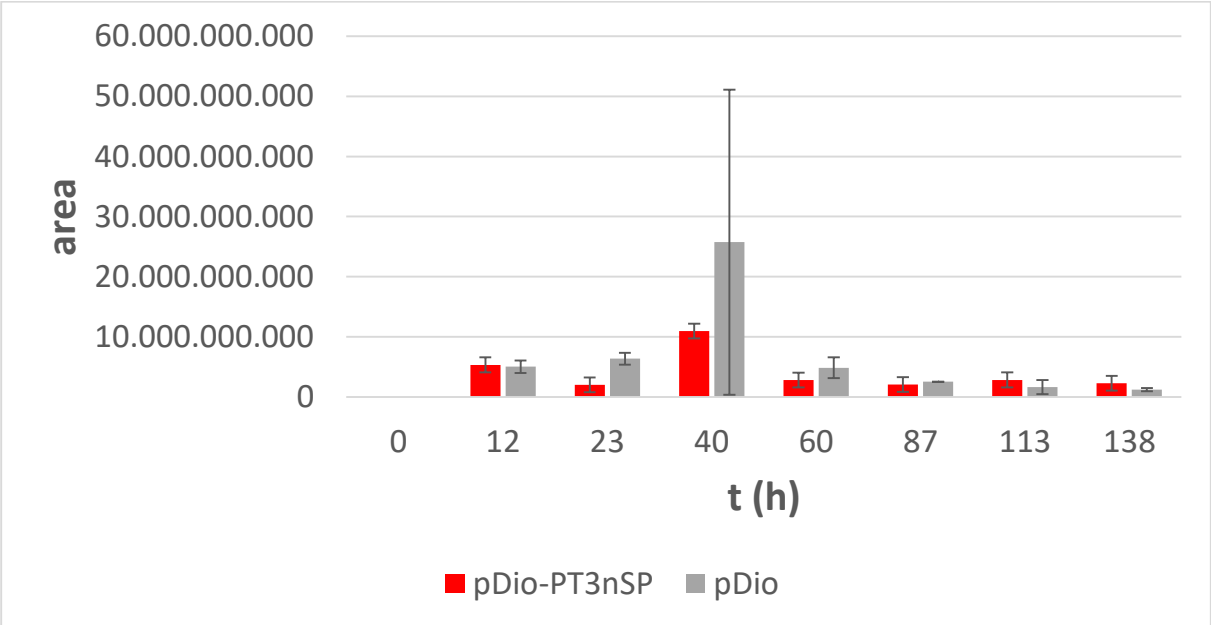



Figure 66: LC-MS peak area after enzyme assay; conversion of OA and GPP to CBGA; cell cultures: *S. cerevisiae* Y05416 pDio-PT3nSP (red); *S. cerevisiae* Y05416 pDio (grey); cultivation temperature: 30°C; YPD medium.

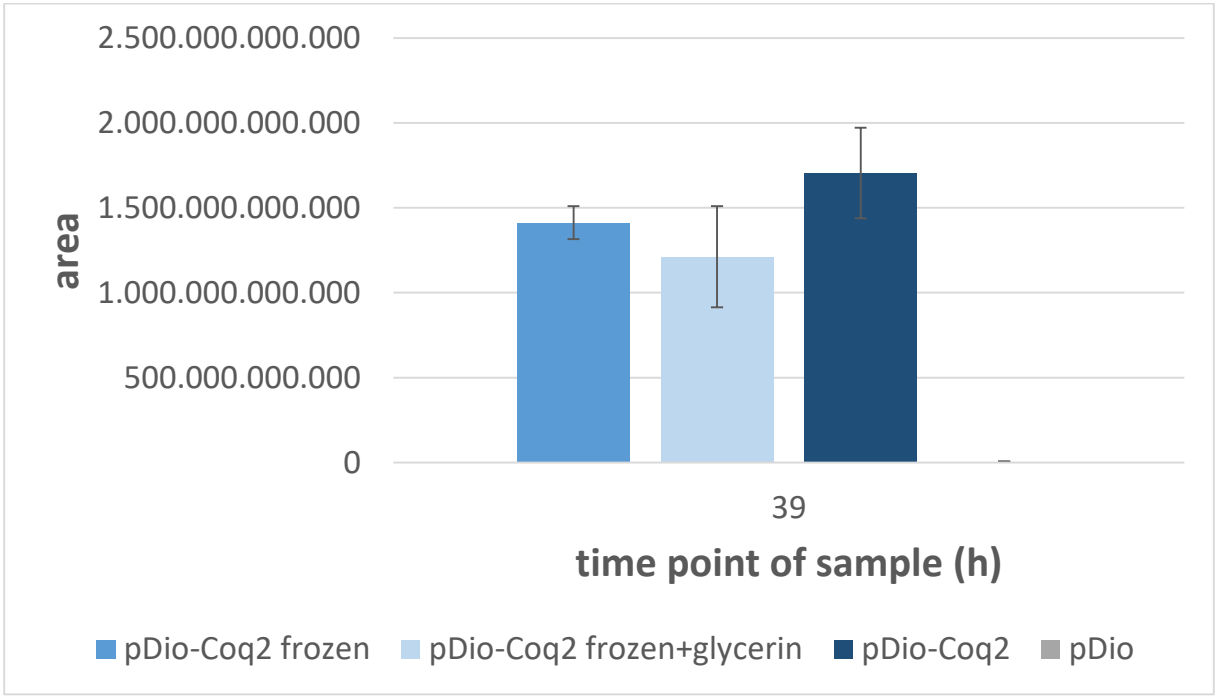



Figure 67: LC-MS peak area after enzyme assay; cell cultures: *S. cerevisiae* Y05416 with pDioCoq2 (all blue columns); *S. cerevisiae* Y05416 with pDio (grey); blue columns are samples which were frozen overnight and frozen& addition of 50% (v/v) glycerin added.

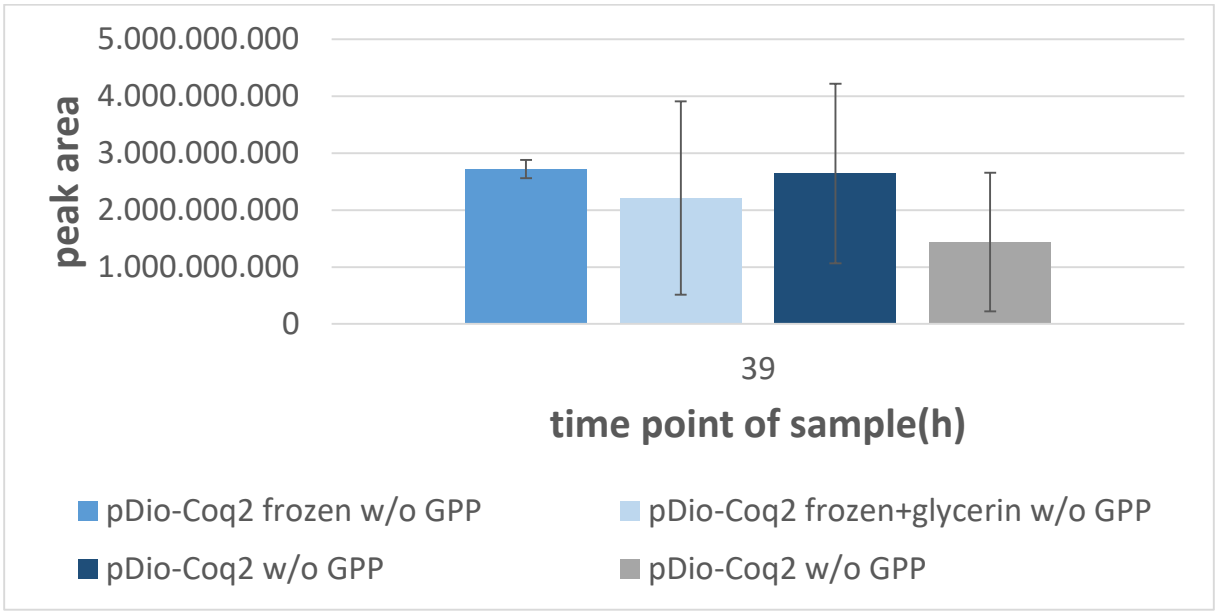



Figure 68: LC-MS peak area after Coq2 enzyme assay without addition of GPP; cell cultures: *S. cerevisiae* Y05416 with pDioCoq2 (all blue columns); *S. cerevisiae* Y05416 with pDio (grey); blue columns are samples which were frozen overnight and frozen& addition of 50% (v/v) glycerin added.

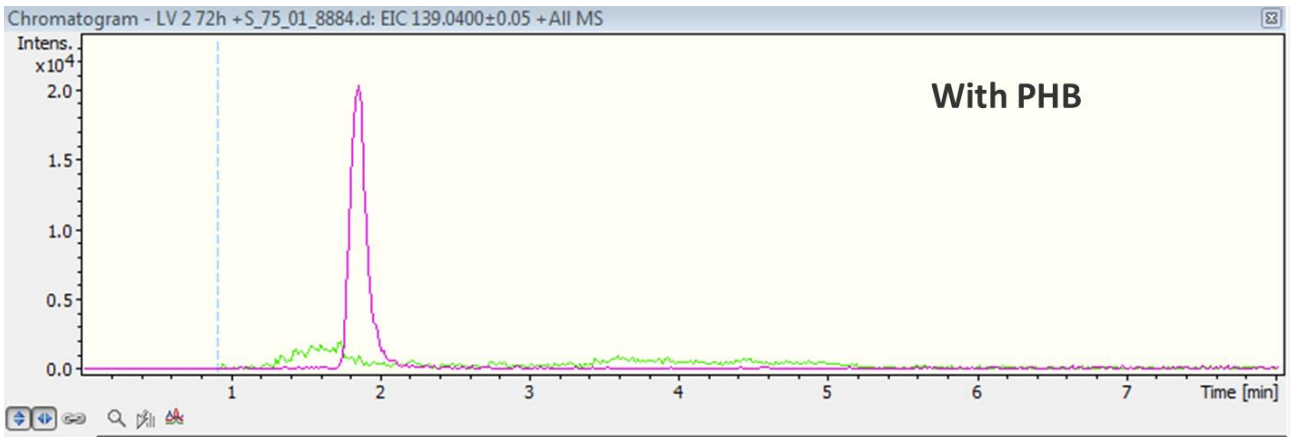



Figure 69: Control expression with *S. cerevisiae* Y05416 pDio; cultivation conditions: temperature 18°C, 200rpm, complex media.

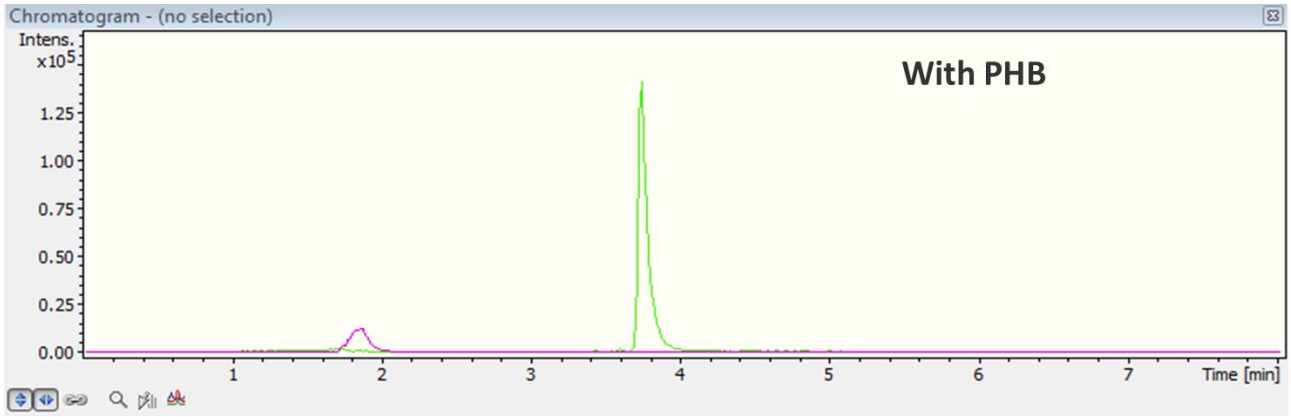



Figure 70: Expression of *S. cerevisiae* Y05416 pDio-Coq2; positive control; cultivation conditions: 18°C, 200rpm, complex media.

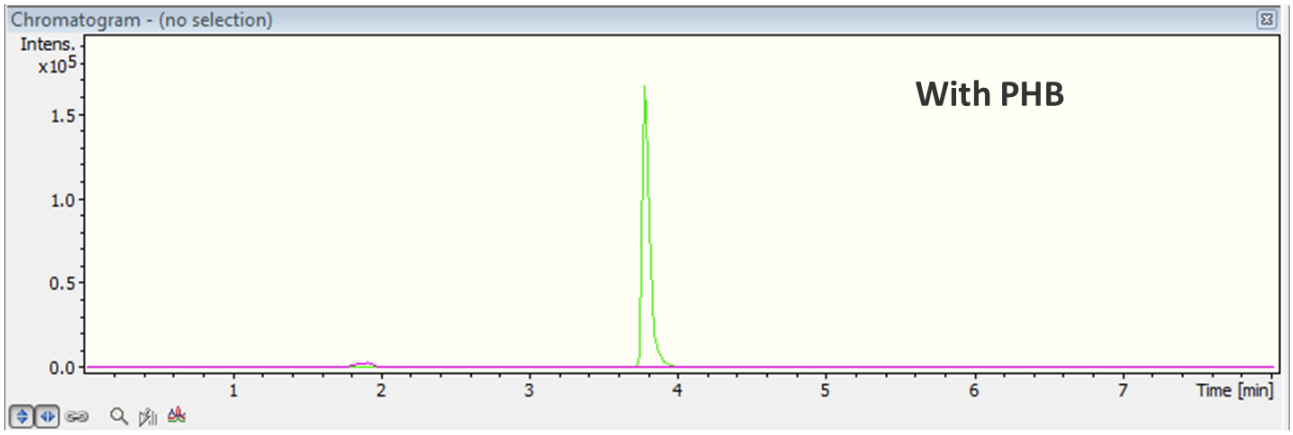



Figure 71: Expression of S. cerevisiae Y05416 pDio-PT6ySP; cultivation for 96h, at 18°C, 200rpm, complex medium.

# Curriculum vitae

## Personal Details

| | |
|---|---|
| Name | Kathleen Pamplaniyil |
| Date of birth | 02.12.1984 |
| Place of birth | Viersen |
| Nationality | German |

## Professional experience

| | |
|---|---|
| 04.2011-03.2015 | **Scientific research assistant**, Chair of Technical Biochemistry, TU Dortmund<br>Research in the context of the Ph.D. thesis, *in silico* analysis of DNA databases, heterologous expression of prenyltransferases in *Saccharomyces* cerevisiae, transient expression in *Nicotiana benthamiana*, breeding of *Cannabis sativa* L. Fedora 17 |
| 01.2011-03.2011 | **Scientific research assistant**, Institute of Environmental Research (INFU), TU Dortmund<br>Degradation and fate of sulfadiazine in soil |

## Education and qualifications

| | |
|---|---|
| Since 04.2011 | **Ph. D. thesis**, Chair of Technical Biochemistry, TU Dortmund<br>"Identification, isolation and functional characterization of prenyltransferases in *Cannabis sativa* L." |
| 10.2008-12.2010 | **Master of Science Chemical Biology**, TU Dortmund<br>Thesis: „Transformation and photo-degradation of the veterinary antibiotic sulfadiazine in soil" |
| 10.2004-09.2008 | **Bachelor of Science Chemical Biology**, TU Dortmund<br>Thesis: „Isolation and characterization of proteins from the sulfur-oxidizing enzyme system from *Paracoccus pantotrophus*" |
| 08.1995-06.2004 | **Abitur**, Gymnasium Liebfrauenschule Mühlhausen, Grefrath |