UNIVERZITA KARLOVA

Přírodovědecká fakulta
Katedra Biochemie

Studijní program: Biochemie

Studijní obor: Biochemie

Helena Smrčková

Terpene discovery combining *in silico* and molecular biology approaches

Objevování molekul terpenů pomocí *in silico* a molekulárně-biologických metod

BAKALÁŘSKÁ PRÁCE

Školitel: Mgr. Tomáš Pluskal, Ph.D.
Konzultant: doc. RNDr. Olga Heidingsfeld, CSc.

Praha, 2023

I affirm this thesis was elaborated independently under the supervision of Mgr. Tomáš Pluskal, Ph.D., that all sources used were cited properly and that any part of the thesis has not been submitted for obtaining another degree.

(Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.)

May 22, 2023, Prague .....

Acknowledgements

I would like to express my gratitude to my supervisor, Mgr. Tomáš Pluskal, Ph.D. for an amazing opportunity to join his research group, long-term guidance and patience and to my consultant doc. RNDr. Olga Heidingsfeld CSc. for her help and support. I would also like to thank all my colleagues at the Institute of Organic Chemistry and Biochemistry for their guidance and for creating a friendly, creative, and innovative working environment. Namely, I would like to thank Mgr. Téo Hebra, Ph.D. and Mgr. Lana Mutabdžija for teaching me new techniques of molecular biology and Bc. Tereza Čalounová for teaching me her *in silico* methodology. I would also like to thank my parents and partner for their endless support during my studies.

Abstract

Terpenoids are the largest class of natural products with remarkable chemical and structural diversity, making them a significant source of compounds for drug discovery. Terpenoids are used as food flavours, therapeutics, cosmetics, hormones or fuel. The starting scaffolds for terpenoid biosynthesis are condensed and, in some cases, cyclic carbohydrates synthesized from linear isoprenyl pyrophosphate precursors by terpene synthases. The enzyme family of terpene synthases is extensive and possesses well-characterized motifs and function-determining domains that can be used to search for homologous proteins with unknown catalytical function that might produce new compounds. Thanks to a bioinformatic pipeline developed in the Pluskal Lab in the Institute of Organic Chemistry and Biochemistry CAS, candidate terpene synthases with the potential to produce new terpene scaffolds were mined *in silico* from large protein sequence repositories. I expressed nine selected proteins originating from plants, bacteria, and fungi in the engineered budding yeast strains JWY501 and ZX178-08 that overproduce isoprenyl pyrophosphates. I then analysed the resulting yeast cell culture extracts using gas chromatography- and liquid chromatography-mass spectrometry. Five out of nine proteins exhibited terpene synthase activity and produced either di- or sesquiterpenes. Two sesquiterpenes produced by a fungal enzyme (UniProt id M2QMG2) were found to be potentially novel compounds, although further experiments are required to elucidate their molecular structures.

Key words

Terpenoids, terpenes, terpene synthases, heterologous expression, engineered yeast

Abstrakt

Terpenoidy představují rozsáhlou a chemicky i strukturně rozmanitou skupinu metabolitů, což z nich činí významný zdroj potenciálních terapeutik. Terpenoidy jsou běžně využívány v potravinářství, farmacii, kosmetice či jako paliva. Výchozí látky pro biosyntézu terpenů jsou kondenzované a v některých případech cyklické uhlovodíky syntetizované z lineárních isoprenyl pyrofosfátů enzymy terpensynthasami. Tato rozsáhlá proteinová rodina sdílí společné motivy a domény určující funkci, které lze použít jako matrici pro hledání homologů mezi proteiny s dosud neznámou funkcí. Potenciál necharakterizovaných proteinů produkovat dosud nepopsané terpenové struktury byl hodnocen *in silico* pomocí metody vyvinuté v laboratoři Mgr.Tomáše Pluskala, Ph.D. na Ústavu organické chemie a biochemie Akademie věd ČR. Devět proteinů pocházejících z rostlin, bakterií a hub bylo podrobeno heterologní expresi v upravených kvasinkových kmenech JWY501 a ZX178-08, které nadměrně produkují nezbytné isoprenyl pyrofosfátové prekurzory. Výsledné extrakty kvasinkových kultur byly analyzovány pomocí plynové chromatografie-hmotnostní spektrometrie a kapalinové chromatografie-hmotnostní spektrometrie, přičemž pět z devíti proteinů vykazovalo terpensynthasovou aktivitu. Bylo zjištěno, že dva seskviterpeny produkované houbovým enzymem (UniProt ID: M2QMG2) jsou potenciálně nové sloučeniny, i když k objasnění jejich struktur jsou zapotřebí další experimenty.

Klíčová slova

Terpenoidy, terpeny, terpensynthasy, heterologní exprese, modifikované kvasinky

List of abbreviations

(v/v)	volume per volume
(w/v)	weight per volume
(w/w)	weight per weight
ADP	adenosine diphosphate
ATP	adenosine triphosphate
bp	base pairs
cDNA	complementary deoxyribonucleic acid
CoA	coenzyme A
CTP	cytidine triphosphate
EDTA	ethylenediaminetetraacetic acid
EI	electron ionization
ESI	electrospray ionization
GC-MS	gas chromatography coupled with mass spectrometry
LC-MS	liquid chromatography coupled with mass spectrometry
MEP	2-C-methyl-D-erythriol 4-phosphate pathway
MoClo	modular cloning
MVA	mevalonic acid
NCBI	National Center of Biotechnology Information
NIST	National Institute of Standards and Technology
OD_{600}	optical density measured at 600 nm
PCR	polymerase chain reaction
PEG	polyethylene glycol
PP	pyrophosphate
PTFE	polytetrafluorethylene
RNA	ribonucleic acid
RPM	rotations per minute
S.O.C.	super optimal broth with catabolite repression
SCE	synthetic complete media for ZX178-08 yeast strain
SDS	sodium dodecyl sulphate
TAE	Tris base, acetic acid and EDTA buffer
TPS	terpene synthase
UDP	uridine diphosphate
YNB	yeast nitrogen base media
YPD	yeast extract, peptone and dextrose media

Table of contents

1.Introduction

1.1 Terpenes and terpenoids

1.1.1 General properties

Natural products refer to chemical compounds synthesized by living organisms. Part of those compounds are classified as specialized metabolites since they are formed as a result of enzymatically catalysed transformation of primary metabolites, such as amino acids, sugars, and vitamins. Among the largest class of secondary metabolites are terpene derivatives, terpenoids. Terpenoids are considered the most extensive class of natural products due to their extraordinary chemical and structural variability, which ultimately renders them a substantial reservoir of compounds for drug discovery[1].

Terpenes are composed of five-carbon isoprene units (C_5H_8 building blocks, also called hemiterpenes; Figure 1) that can be combined in various ways to yield a diverse array of structures. Terpenes are simple hydrocarbons, while terpenoids represent a modified class of terpenes that may contain different functional groups. These modifications can also include the addition of heteroatoms such as oxygen[2], sulphur, or nitrogen, or the rearrangement of the carbon atoms in the molecule[3]. Terpenoids can be divided into several subgroups, including monoterpenes (typically $C_{10}H_{16}$), sesquiterpenes (typically $C_{15}H_{24}$), diterpenes (typically $C_{20}H_{32}$), sesterpenes (typically $C_{25}H_{40}$), triterpenes (typically $C_{30}H_{48}$) and so on, depending on their carbon unit composition (Figure 1).

Typical examples of monoterpenes are limonene, α-pinene and β-pinene (Figure 1). Limonene is a colourless liquid hydrocarbon commonly found in the rinds of citrus fruits. Limonene is used as a flavour and fragrance ingredient in cosmetics thanks to its pleasant scent and low toxicity[4]. Two isomers of pinene (α-pinene and β-pinene) are colourless liquids with a distinct pine-like aroma and can be found in the oils of pine and other species of conifers, or in rosemary oil. Pinene is used to make camphor, fragrances or insectisides[5].

A cyclic sesquiterpene β-caryophyllene (Figure 1) is naturally found in many plants, such as black pepper, cloves, and cannabis[6]. Apart from being used as a flavouring agent, β-caryophyllene has been extensively studied for its anti-inflammatory[7] and analgesic[8] effects.

A cyclic diterpene cembrene A (Figure 1) is commonly found in corals of the genus *Nephthea*[9]. Cembrene A exhibits potent cytotoxicity against human breast cancer cells by inducing apoptosis and cell cycle arrest[10]. Additionally, cembrene A

shows inhibitory effect on the production of inflammatory cytokines in a mouse model of acute inflammation[11].

Haslenes are highly branched polyunsaturated sesterpenes produced by planktonic diatoms such as *Haslea ostrearia* and *Rhizosolenia setigera* and can be used as indicators of presence of specific diatoms in oceanic sedimenting particles[12].

Squalene is a linear triterpene naturally occurring in shark liver and olive oil[13]. Squalene is used in cosmetics as a moisturizes, adjuvants in vaccines[14], and for developments of cholesterol-lowering drugs[15].

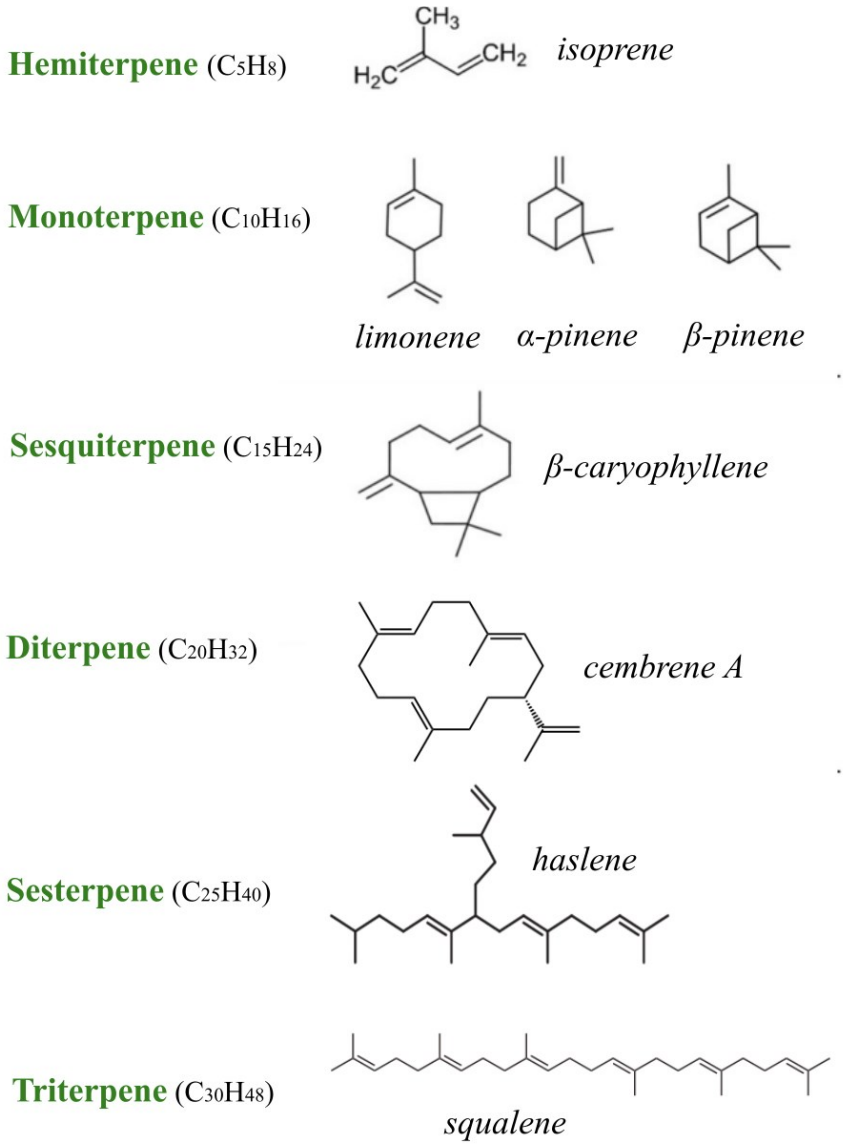

Figure 1. *Classification of terpenes based on the number of isoprene units. Terpenes are defined by the general molecular formula $C_{5n}H_{8n}$, wherein n denotes the quantity of isoprene units present. Hemiterpenes (C_5H_8) correspond to n=1, monoterpenes ($C_{10}H_{16}$) correspond to n=2, and so forth. Figure is adapted from[16].*

Terpenoids play a diverse role in various fields, such as food flavours, drugs, cosmetics, hormones and fuel[2]. Some terpenoids (terpene derivatives) such as taxanes exhibit biological activity and are used for the treatment of human diseases such as cancer.

For example, menthol (Figure 2) is a monoterpenoid that is obtained in significant quantities from peppermint (*Mentha piperita*) as a farm crop. Menthol acts as a cooling agent, helping with itchy and inflamed skin [17].

The second example is artemisinin (Figure 2), an anti-malarial sesquiterpenoid derived from annual wormwood (*Artemisia annua*). Aside from its remarkable clinical effects against malaria, experimental research also indicates that drugs from the artemisinin family demonstrate potent anti-inflammatory properties by modulating both innate and adaptive immunity [18].

The third example is abietic acid and related diterpene resin acids (Figure 2), which are used as a biological feedstock for the chemical industry and are sourced from conifers (Pinaceae) through a resin collection process that has been in use for over a century. Additionally, abietic acid shows inhibitory activity against testosterone 5α-reductase in male rats and could be used as a treatment for androgene-dependent diseases in the future[19].

The fourth example is paclitaxel, also marketed as Taxol (Figure 2), a high-value diterpenoid-derived anti-cancer drug originally sourced from the bark of the Pacific yew tree (*Taxus brevifolia*)[20]. While terpenes and terpenoids have many potential health benefits, some may also be harmful in relatively low doses. For example, α-thujone, a terpenoid found in wormwood, is highly toxic and can cause seizures and other adverse effects when ingested[21].

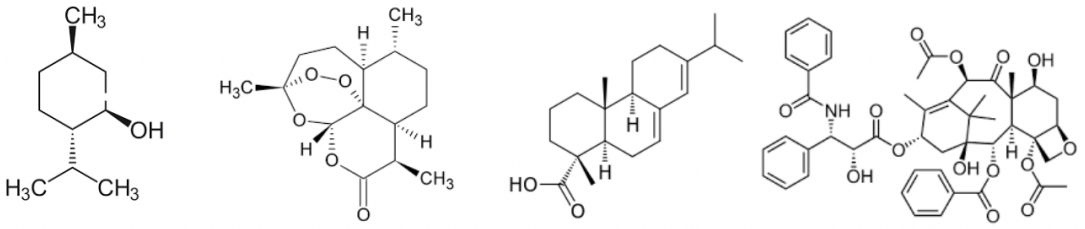

***Figure 2**. Chemical structures of the monoterpenoid (C_{10}) menthol; the sesquiterpenoid (C_{15}) artemisinin; the diterpene resin acid (C_{20}) abietic acid and the diterpenoid (C_{20}) paclitaxel.*

1.1.2 Biosynthesis

Terpenes and terpenoids are natural products that are produced by a variety of organisms, including plants, animals, fungi, and bacteria[22]. Terpenes originate from either the mevalonate (MVA) pathway or 2-C-methyl-D-erythriol 4-phosphate (MEP) pathway[23]. The MVA and MEP pathways are two distinct non-homologous strategies utilized by nature to biosynthesize five-carbon isoprene units. The MVA pathway is the principal route of isopentenyl pyrophosphate (isopentenyl-PP) and dimethylallyl pyrophosphate (dimethylallyl-PP) synthesis in eukaryotes and archaea[24,25], while the MEP pathway is characteristic of bacteria[26]. Nonetheless, there are some exceptions to this general principle. Firstly, certain bacteria possess the MVA pathway[27,28,29]. This has been typically attributed to the acquisition of horizontal gene transfer from archaeal or eukaryotic donors[30]. Secondly, in addition to the MVA pathway, plants, other photosynthetic eukaryotes, and apicomplexa (a group of parasitic protists derived from a photosynthetic ancestor) have been found to contain the MEP pathway. All these eukaryotic groups carry plastids (either photosynthetic or highly derived in the case of apicomplexa), indicating that they acquired the MEP pathway through the transfer of genes from the original cyanobacterial endosymbiont that gave rise to the plastids[26].

1.1.2.1 Mevalonate pathway

The MVA pathway constitutes a crucial metabolic pathway that facilitates the production of bioactive molecules essential for various cellular processes. This cytosolic pathway is responsible for converting mevalonate into precursors of sterol isoprenoids such as cholesterol or ergosterol. Additionally, the pathway generates various nonsterol isoprenoids and their precursors. The MVA pathway is thus of great importance for the maintenance of cellular homeostasis and overall organismal health.

The MVA pathway (Figure 3, page 14) starts with the synthesis of 3-hydroxy-3-methylglutaryl-CoA from acetyl-CoA, which involves the intermediary of acetoacetylCoA[31]. The conversion of 3-hydroxy-3-methylglutaryl-CoA to mevalonic acid is catalysed by 3-hydroxy-3-methylglutaryl-CoA reductase, which is the rate-limiting enzyme of the mevalonate pathway[32]. Mevalonate kinase is the second crucial enzyme of the pathway after 3-hydroxy-3-methylglutaryl-CoA reductase, responsible for the phosphorylation of mevalonic acid to phosphomevalonate. Mevalonate kinase activity is controlled through feedback inhibition by intermediates in the isoprenoid/cholesterol pathway, namely geranyl pyrophosphate (geranyl-PP), farnesyl pyrophosphate (farnesyl-PP), and geranylgeranyl pyrophosphate (geranyl geranyl-PP)[33]. The subsequent reversible step is catalysed by phosphomevalonate kinase, which converts mevalonate 5-phosphate and ATP to mevalonate 5-diphosphate and ADP. The last step is a conversion of mevalonate pyrophosphate to isopentenyl-PP catalysed

by a mevalonate pyrophosphate decarboxylase. Isopentenyl-PP can be isomerised to dimethylallyl-PP by isopentenyl pyrophosphate isomerase[34].

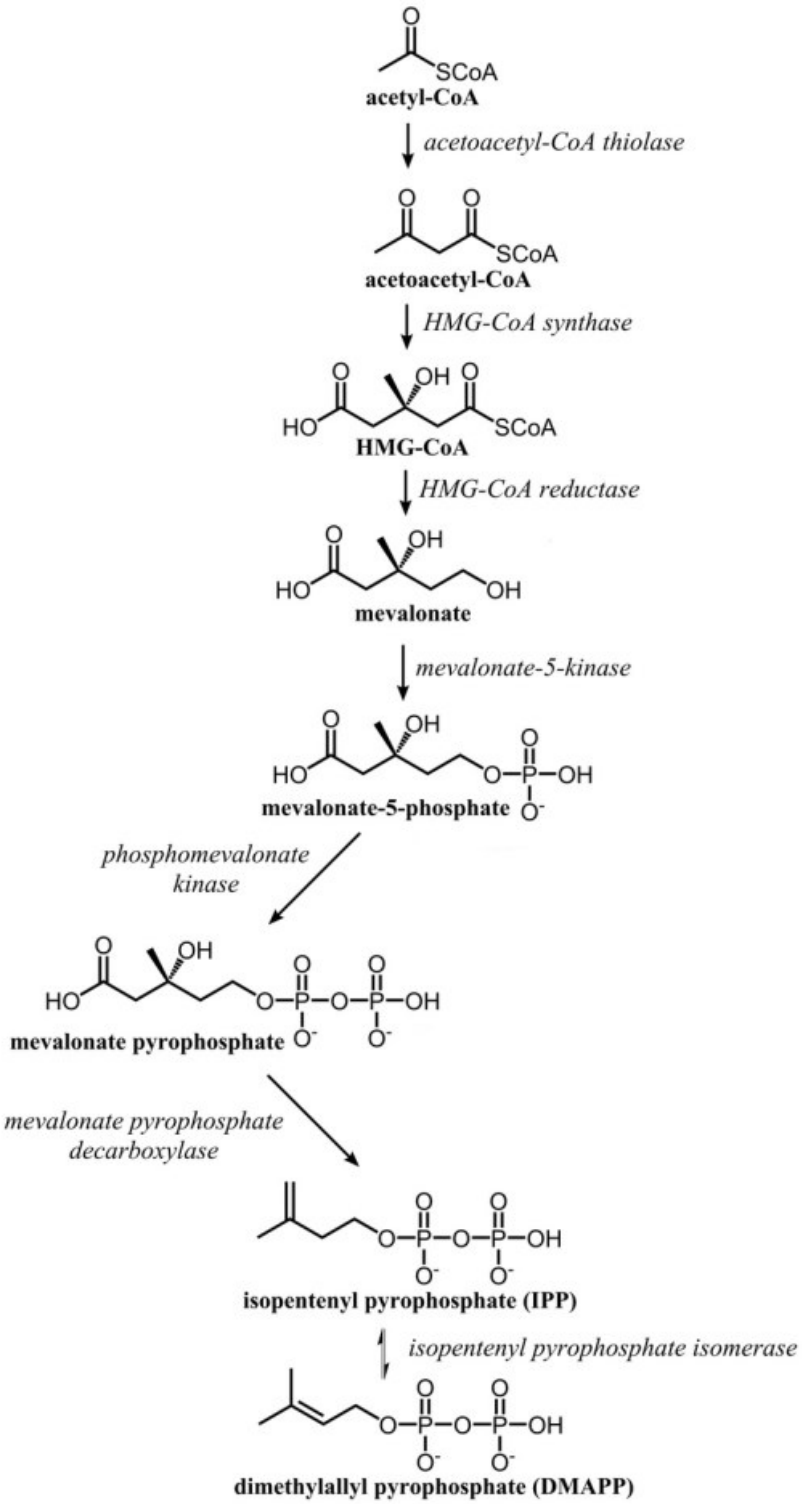

***Figure 3**. The mevalonate pathway in eukaryotes. Abbreviations: HMG-CoA, 3-hydroxy-3-methylglutaryl-coenzyme A. Figure is adapted from [35].*

1.1.2.2 C-methyl-D-erythriol 4-phosphate pathway

The MVA pathway was discovered in 1950s and was thought to be the only pathway for the biosynthesis of isopentenyl-PP and dimethylallyl-PP for decades[36]. However, Rohmer et al. found an alternative pathway (Figure 4) in several bacteria and in plastids [37,38]. The process begins with a thiamine diphosphate dependent condensation reaction between D-glyceraldehyde 3-phosphate (2) and pyruvate (1), resulting in the formation of 1-deoxy-D-xylulose-5-phosphate (3). This product is then isomerized to MEP (4) by the enzyme 1-deoxy-D-xylulose-5-phosphate reductoisomerase (IspC). The coupling of MEP and CTP is subsequently catalysed by methylerythritol cytidyl diphosphate synthetase (IspD), leading to the formation of methylerythritol cytidyl diphosphate (5), which is then phosphorylated by the ATP-dependent enzyme IspE at the C2 hydroxyl group, leading to the formation of 4-diphosphocytidyl-2-C-methyl-D-erythritol-2-phosphate (6). IspF cyclizes this compound to 2-C-methyl-D-erythritol-2,4-cyclodiphosphate (7). The ring opening of the cyclic pyrophosphate and the C3-reductive dehydration of 2-C-methyl-D-erythritol-2,4-cyclodiphosphate to 4-hydroxy-3-methyl-butenyl 1-diphosphate (8) are catalysed by IspG. The final step of the MEP pathway, which converts 4-hydroxy-3-methyl-butenyl 1-diphosphate to both isopentenyl-PP (9) and dimethylallyl-PP (10), is catalysed by IspH. Unlike the MVA pathway, the MEP pathway does not require isopentenyl pyrophosphate isomerase (converts isopentenyl-PP to dimethylallyl-PP and vice versa) in many organisms [39].

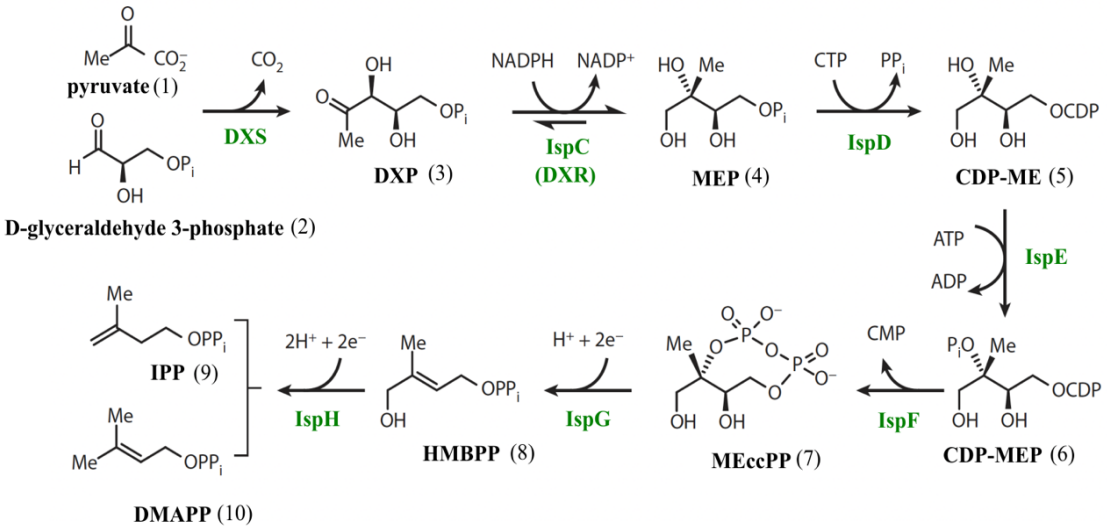

***Figure 4**. The MEP pathway in plastids, algae and eubacteria. Abbreviations: CDP-ME, methylerythritol cytidyl diphosphate; DMAPP, dimethylallyl pyrophosphate; DXP, 1-deoxy-D-xylulose 5-phosphate; DXR, DXP reductoisomerase; DXS, DXPsynthase; HMBPP, 4-hydroxy-3-methyl-butenyl 1-diphosphate; IDI, IPP:DMAPP isomerase; IPP, isopentenyl pyrophosphate; MEcPP, 2-C-methyl-D-erythritol-2,4-cyclodiphosphate. Figure is adapted from [39].*

1.1.2.3 Prenyl transferases

The isoprenoid biosynthesis pathway relies on the catalytic activity of prenyl transferases, which facilitate the condensation of isopentenyl-PP with an allylic co-substrate, such as dimethylallyl-PP, geranyl-PP or farnesyl-PP. The resulting products of this reaction are geranyl-PP, farnesyl-PP, or geranyl geranyl-PP (depending on substrates; Figure 5)[40], which are utilized as linear precursors for various isoprenoid-derived metabolites, including terpenes, sterols, retinoids, dolichols, ubiquinones, prenylated proteins, and chlorophylls. The key examples of prenyl transferase enzymes include geranyl-PP synthase, farnesyl-PP synthase, and geranyl geranyl-PP synthase[41].

The most comprehensively studied pathway, with respect to the mechanisms and structural attributes of enzyme catalysts, entails the "head-to-tail" (1'-4; as illustrated in Figure 6, page 17) fusion of two isopentenyl-PP molecules using enzymes such as farnesyl-PP synthase. The synthesis of geranyl-PP as an intermediary compound ensues, with geranyl-PP then serving as a partner in the subsequent elongation cycle involving another isopentenyl-PP molecule, ultimately producing farnesyl-PP (Figure 5). Using a different prenyl transferase, elongation may continue and result in geranyl geranyl-PP formation (Figure 5). Subsequently, eight additional naturally occurring isoprenoid coupling patterns were identified (refer to Figure 6)[42, 43].

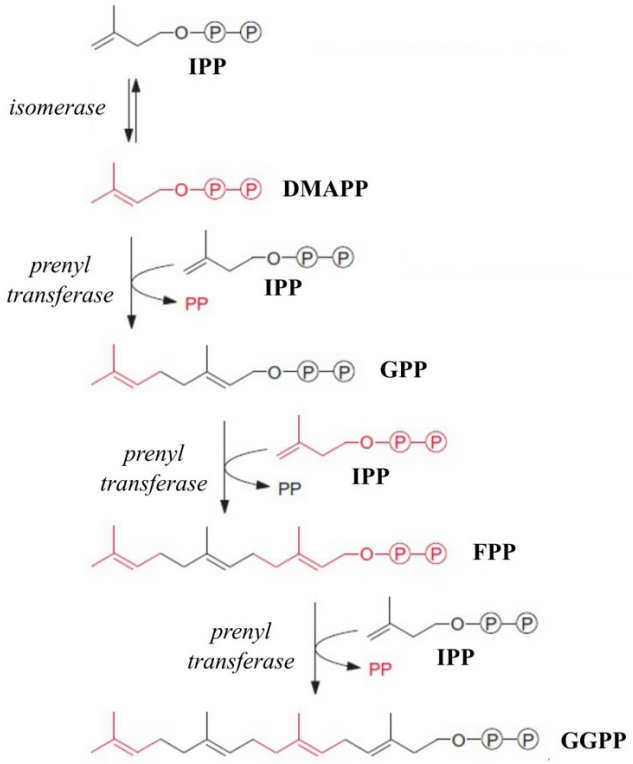

Figure 5. *General scheme of production of terpene precursors via linear elongation with assistance of prenyl transferases. Abbreviations: DMAPP, dimethylallyl pyrophosphate; FPP, farnesyl pyrophosphate; GPP, geranyl pyrophosphate; GGPP, geranyl geranyl pyrophosphate; IPP, isopentenyl pyrophosphate. Figure is adapted from [44].*

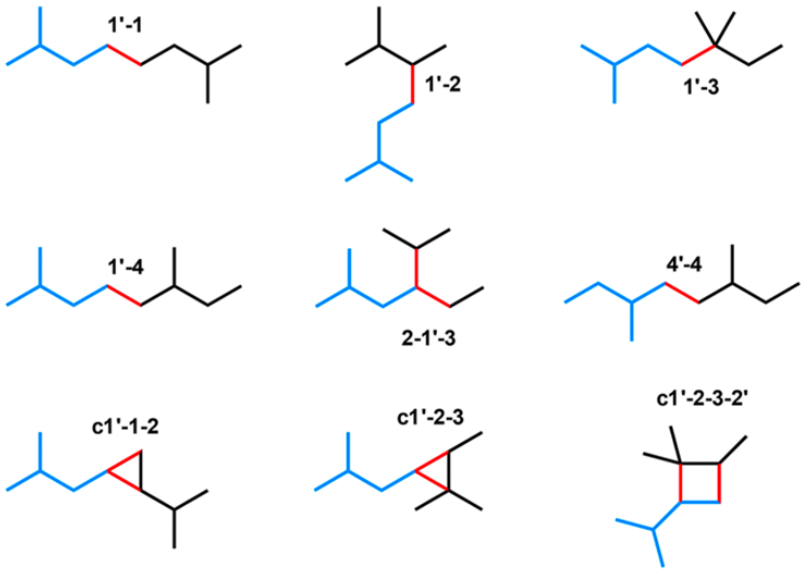

Figure 6. *General scheme of naturally occurring ways of incorporation an isopentenyl-PP molecule to a pre-existing isoprenoid scaffold in assistance of prenyl transferases. Numbers 1'-4'represent carbon atoms of the main chain of blue molecules, numbers 1-4 represent carbon atoms of the main chain of black molecules. Red lines represent bonds between these two molecules. Figure was reproduced from*[43].

1.2 Terpene synthases

As indicated in chapter 1.1.2.3 about prenyl transferases, the structural complexity of terpenes is in contrast with their relatively simple biosynthetic origins. Specifically, linear, achiral C_{5n} isoprenoid diphosphates (n = 1, 2, 3, etc.) are produced through head-to-tail coupling reactions of 5-carbon precursors[45,46] which are subsequently subjected to cyclization reactions leading to the formation of a multitude of products that typically contain multiple fused rings and stereocenters[47,48,49,50]. These reactions are catalysed by enzymes known as terpene synthases (TPSs) or terpenoid synthases, with those that catalyse cyclization reactions being referred to as terpenoid cyclases.

Terpenoid cyclization reactions are considered the most intricate reactions found in nature, with more than half of the substrate carbon atoms, on average, undergoing changes in bonding, hybridization, and stereochemistry during the multi-step cyclization cascade. Additionally, since cyclic terpenoids are generally unable to be generated from linear precursors in the absence of an enzyme, the rate enhancement resulting from the catalytic activity of a terpenoid cyclase over the uncatalysed rate is immeasurably large. Alternatively, two isoprenoid allylic groups have the potential to engage in non-conventional coupling reactions, including but not limited to cyclobutanation, branching, or cyclopropanation reactions. Such reactions lead to the formation of alternative carbon frameworks[43].

1.2.1 Classification

The classification of TPSs is useful for understanding their biochemical and physiological roles in plants and other organisms. It can also provide insights into the evolution of terpene biosynthesis and the diversification of natural product chemistry[51].

1.2.1.1 Sequence-based classification

This classification is based on the amino acid sequences of TPSs. The TPSs family can be divided into several subfamilies based on the presence of conserved motifs in the amino acid sequences. For example, the TPS-a subfamily has a conserved aspartate-rich motif, while the TPS-b subfamily has a conserved DDXXD motif, where D represents aspartic acid and X any amino acid, giving some flexibility to the specific sequence of the motif. This classification scheme was first proposed by Degenhardt et al.[22] and has been used in many subsequent studies. Jiang et al.[51] suggest that TPSs might have originated from isoprenyl diphosphate synthase genes. Chen et al.[52] were able to identify 7 TPS subfamilies among plant kingdom: TPS-a, TPS-b, TPS-c, TPS-d, TPS-e/f (these enzymes produce both sesqui- and diterpenes), TPS-g and TPS-h. These classes differ in substrate preferences and conserved motifs.

1.2.1.2 Substrate-based classification

This classification is based on the substrate specificity of TPSs. TPSs can be divided into several classes based on the type of isoprenyl diphosphate substrate they use. For example, monoterpene synthases use geranyl-PP as a substrate, while sesquiterpene synthases use farnesyl-PP as a substrate and diterpene synthases utilize geranyl geranyl-PP as a substrate[43].

1.2.2 Catalytic activity of terpene synthases

The catalytic process of terpene synthases often involves a sequence of several carbocation intermediates that determine the reaction pathway. Due to their high reactivity, these intermediates have the potential to inactivate the enzyme by alkylation. To manage and manipulate these intermediates effectively, terpene synthases usually contain a nonpolar active site pocket[43].

Nonetheless, occasional polar groups or solvent molecules may be present in the active site, as evidenced by observations in bornyl diphosphate synthase from *Salvia officinalis*[53] and aristolochene synthase from *Aspergillus terreus*[54]. These nucleophiles, however, cannot quench the carbocation intermediate unless they are positioned in a suitable orientation. Studies of terpenoid cyclase reaction mechanisms have shown that tertiary carbocation intermediates are more common than secondary carbocation intermediates. Furthermore, electron delocalization influences and directs the cyclization

trajectory of many carbocation intermediates in terpenoid cyclization cascades[55]. Weakly polar interactions, such as charge-charge, charge-dipole, and charge-quadrupole interactions between carbocation intermediates and suitably oriented amino acid side chains, can stabilize carbocation intermediates in the active site of a terpenoid cyclase.

Carbocation stabilization can be achieved through cation-π interactions with the aromatic side chains of phenylalanine, tyrosine, and tryptophan[56,57,58] (Figure 7). This strategy of electrostatic stabilization of multiple transition states in the terpenoid cyclization cascade reduces the risk of alkylation[43].

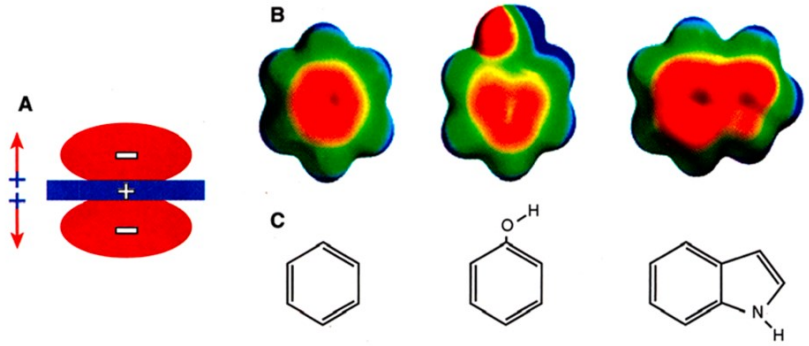

***Figure 7**. (A) The aromatic ring possesses an electronic quadrupole configuration, which results in an absence of net charge and net dipole. (B and C) Electrostatic surface potentials of the aromatic side chains in phenylalanine, tyrosine, and tryptophan indicate that these residues bear significant partial negative charge on their faces. This property enables the aromatic residues to stabilize carbocation intermediates in terpenoid cyclase mechanisms by engaging in quadrupole−charge or cation−π interactions. Red color indicates increased electron density, green and blue indicate low electron density area. Figure was reproduced from[56].*

Terpenoid cyclases are generally classified into two primary classes, depending on their approach to forming the initial carbocation. Class I terpenoid cyclases employ a trinuclear metal cluster to trigger the ionization of an isoprenoid diphosphate substrate, producing an allylic cation and inorganic pyrophosphate. In contrast, class II terpenoid cyclases rely on a general acid to protonate the terminal carbon-carbon double bond of an isoprenoid substrate to produce a tertiary carbocation[59].

The active site of class I cyclases, such as bacterial pentalenene synthase[60] or plant epi-aristolochene synthase[61], is in the centre of an α-helical bundle designated as the "α fold" (Figure 8, page 20). On the other hand, the active site of class II cyclases, such as squalene-hopene cyclase[62], is situated at the interface of two α-helical domains referred to as "β" and "γ" (Figure 8). Notably, the crystal structure of taxadiene synthase has revealed that a class I terpenoid cyclase may exist as a fusion of αβγ domains, where the α domain is"c"talyt"c"lly active, and the βγ domain is an inactive vestige (Figure 8)[63].

Despite some exceptions, these classes help to distinguish the general strategies utilized by terpenoid cyclases during initial carbocation formation[43].

Abietadiene synthase from the grand fir was the first bifunctional αβγ terpenoid synthase to reveal the structure of a catalytically active class I and class II active sites that catalyse tandem diterpene cyclization reactions[64]. The domain architectures presented in Figure 8 represent the context for the majority of terpenoid cyclization reactions in biology. Tertiary structures of each terpenoid cyclase consist of a core domain architecture of α, αβ, βγ, or αβγ. Additionally, αα domain architecture is also observed in certain bifunctional terpenoid synthases[43].

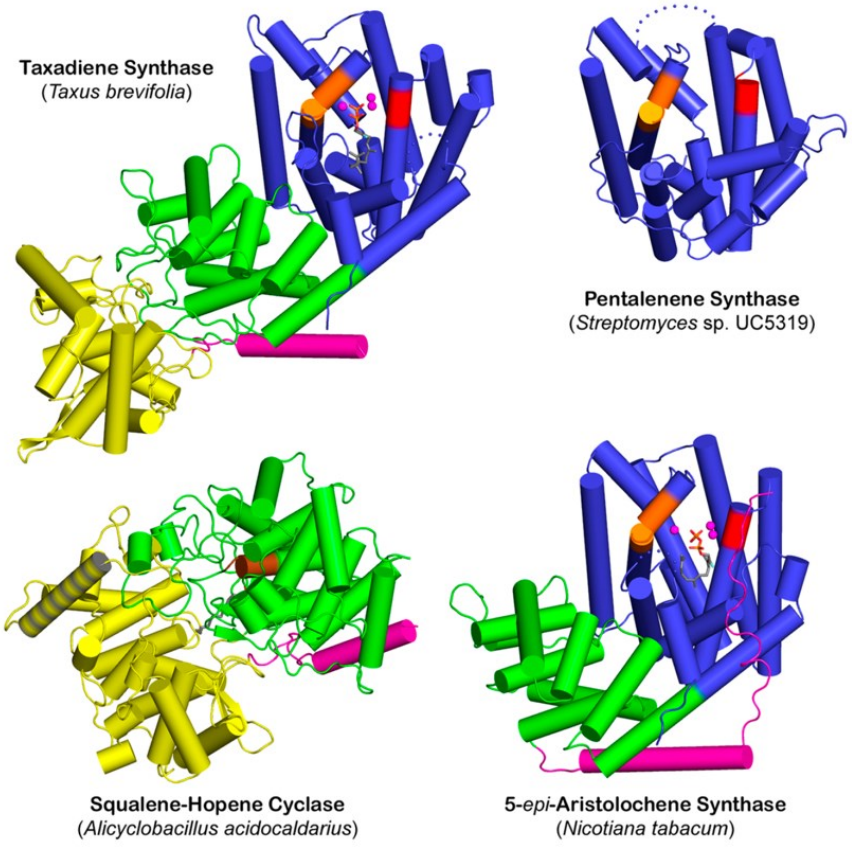

Figure 8. *Terpenoid synthases domains. α, β, and γ domains are depicted in blue, green, and yellow colours, respectively. The N-terminal helix of the β domain is magenta. The α domain corresponds to the class I terpenoid synthase fold and is present as a single domain in bacterial pentalenene synthase, in tobacco epi-aristolochene synthase, and in taxadiene synthase from the Pacific yew. In each α domain, metal-binding motifs, in red and orange colours, coordinate to a trinuclear metal cluster. The βγ domain assembly represents the class II terpenoid synthase fold, The aspartic acid motif, in brown, triggers the class II cyclization reaction. γ domain of squalene-hopene cyclase comprises a membrane-anchoring helix, depicted in stippled yellow-gray, which is not present in the γ domain of taxadiene synthase. This structural diversity in terpenoid synthase domains offers insights into their catalytic mechanisms and provides potential targets for the development of new drugs or biocatalysts. Figure was reproduced from[63].*

1.2.3 Class I Monoterpene cyclases

The C_{10} isoprenoid geranyl-PP (Figure 5, page 16) represents the shortest cyclization substrate in the terpenoid biosynthesis pathway. The cascade for monoterpene cyclization typically proceeds through the formation of the α-terpinyl carbocation (Figure 9). The subsequent reactions of this crucial intermediate lead to the formation of structurally and stereochemically diverse carbon skeletons, which are determined by the conformation of the intermediate molecule. This conformation is dictated by the three-dimensional contour of the enzyme active site, which serves as a catalytic template, ensuring that only conformations leading to the formation of the desired product(s) are adopted by the substrate and intermediates. Thus, the active site contour exhibits a product-like characteristic, especially in high-fidelity cyclases, to ensure the precise generation of a specific product.

Furthermore, the role of water in geranyl-PP cyclization chemistry is noteworthy: the final carbocation intermediate in some monoterpene cyclization cascades is terminated by the addition of a solvent molecule. Therefore, strict control of this solvent molecule is necessary to avoid premature termination of carbocation intermediates in the cyclization cascade. These features are evident in the structural and functional studies of monoterpene cyclases that have produced crystal structures thus far[43]. Examples of class I monoterpene cyclases are (+)-bornyl diphosphate synthase, limonene synthase or cineol synthase.

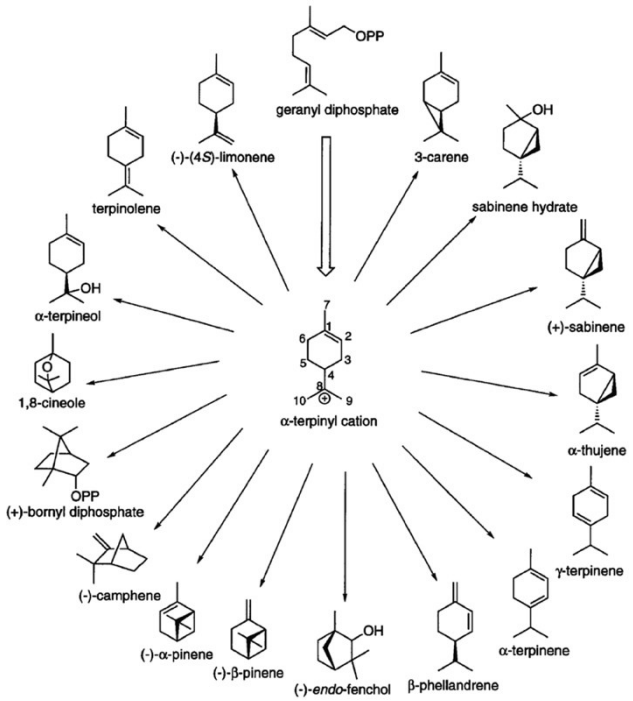

Figure 9. *The formation of α-terpinyl cation can be achieved through the cyclization of GPP by creating a bond between C1 and C6. This reaction subsequently leads to the generation of a wide range of cyclization products through further reactions of the cation. Figure was reproduced from[49].*

1.2.3.1 (+)-Bornyl diphosphate synthase

The catalytic mechanism of (+)-bornyl diphosphate synthase, a homodimeric enzyme found in *Salvia officinalis,* catalyses the formation of the monoterpene (+)-bornyl diphosphate. The enzyme is characterized by an unusual reaction sequence involving dissociation and rebinding of the substrate diphosphate group to carbocation intermediates during geranyl-PP cyclization cascade (Figure 10).

The enzyme's crystal structure, resolved by X-ray crystallography, shows that it consists of two 64-kD subunits assembled in antiparallel fashion (Figure 11, page 23)[64]. The active site of the enzyme plays a critical role in enforcing the left-handed helical conformation of geranyl-PP, required for the generation of (+)-bornyl diphosphate via the (4R)-α-terpinyl cation. The binding of the substrate diphosphate group triggers a conformational change from an open to a closed active site conformation, which ensures that highly reactive carbocation intermediates in catalysis are protected from bulk solvent (Figure 12, page 23). This conformational change is common to all class I terpenoid synthases[43].

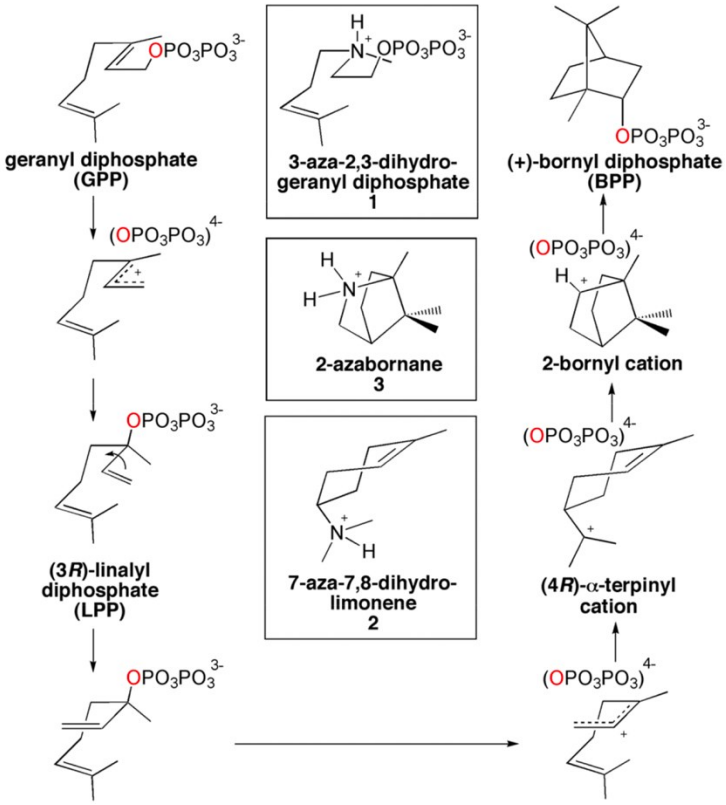

Figure 10*. The geranyl-PP cyclization cascade, which is facilitated by the catalytic activity of (+)-bornyl diphosphate synthase. To gain a deeper understanding of the reaction mechanism, positional isotope exchange experiments were conducted. The experimental results reveal that the oxygen atom of the prenyl diphosphate ester substrate (highlighted in red) is identical to that of the product. Additionally, the study employs aza analogues of carbocation intermediates, which are presented in boxed figures. Figure was reproduced from*[65].

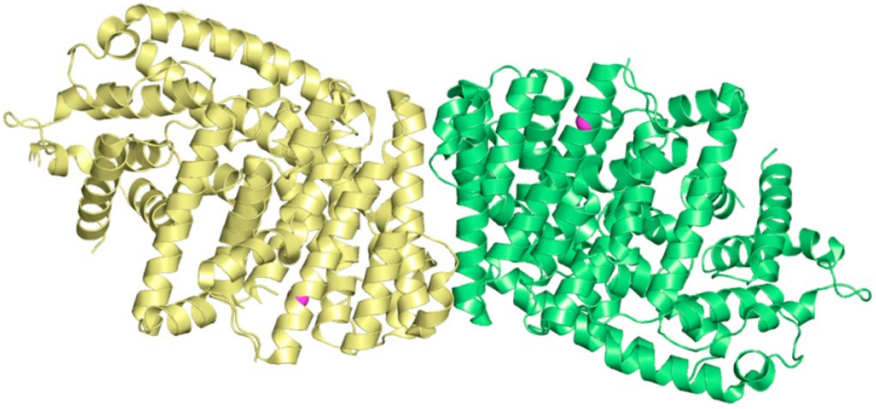

Figure 11. *The (+)-bornyl diphosphate synthase exhibits a dimeric quaternary structure with a βα:αβ arrangement, wherein the dimer interface boasts a buried surface area of over 1000 Å2. Notably, the active sites within each dimer are positioned in an antiparallel configuration, as evidenced by the presence of bound Mg^{2+} ions (magenta) located at the opening of each active site. Figure was adapted from*[66].

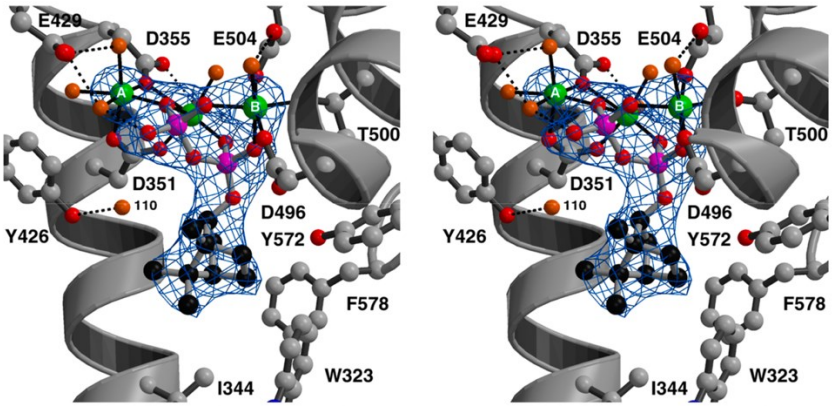

Figure 12. *The stereoview of the active site of (+)-bornyl diphosphate synthase displays the metal (green A, B and C) coordination and hydrogen bond interactions through solid and dashed black lines, respectively. (+)-bornyl (black) diphosphate (magenta) is formed as the product of (+)-bornyl diphosphate synthase. Notably, the active site of the enzyme-ligand complexes, including the one formed with the monoterpene product, consistently captures water molecule no. 110. Figure was reproduced from*[67].

1.2.4 Class I Sesquiterpene cyclases

In comparison to geranyl-PP cyclization reactions, which typically follow an ionization-recombination-reionization sequence to facilitate the formation of C1−C6 bonds and the α-terpinyl cation intermediate (as depicted in Figure 9), farnesyl-PP cyclization reactions can lead to the formation of C1−C6, C1−C7, C1−C10, and C1−C11 bonds, depending on the carbon−carbon double bond that reacts with the initially formed allylic carbocation (as illustrated in Figure 13, page 24)[68].

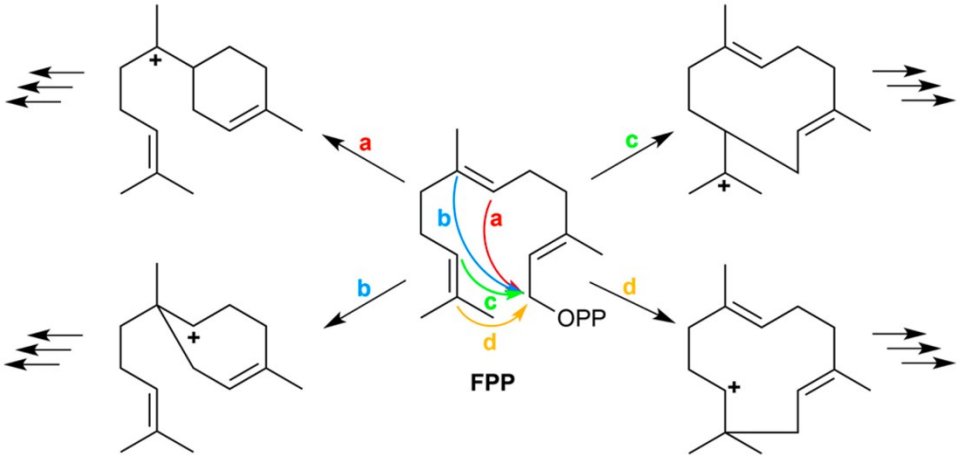

Figure 13. *The potential pathways for the formation of carbon-carbon bonds during the cyclization reactions of sesquiterpenes (where OPP refers to diphosphate). Figure was reproduced from*[43].

Furthermore, the resulting carbocation intermediate can undergo additional cyclization reactions, hydride transfers, methyl migrations, and so on before the reaction sequence terminates. As a result, sesquiterpene cyclization reactions produce increasingly diverse arrays of carbon skeletons when compared to monoterpene cyclization reactions, owing to the larger substrate's additional carbon−carbon bond-forming pathways.

The α domain of an enzyme wit' overall α or αβ domain architecture provides a template for the catalysis by a sesquiterpene cyclase, ensuring that the substrate farnesyl-PP and the subsequently formed intermediates adopt only conformations that lead to the correct product(s) formation. Consequently, active site contours are product-like, especially for high-fidelity cyclases, to ensure specific product generation[43]. Examples of class I sesquiterpene synthases are α-bisabolene synthase, trichodiene synthase or selinadiene synthase.

1.2.4.1 α-Bisabolene synthase

This high-fidelity cyclase utilizes farnesyl-PP as a substrate and generates more than 92% α-bisabolene (Figure 14, A, page 25). The crystal structure of unliganded α-bisabolene synthase with a resolution of 2.14 Å reveals the αβγ domain architecture, in which the class I active site is located in the α domain. The βγ domains have no known catalytic function, except for the N-terminal polypeptide segment that helps cap the active site upon substrate binding. Crystal structures of complexes with substrate analogues and bisphosphonate inhibitors, such as alendronate, pamidronate, and etidronate, reveal molecular recognition of the diphosphate moiety by 3 Mg^{2+} ions complexed by the aspartate-rich and DTE metal-binding motifs in the active site (Figure 14, B, page 25)[69].

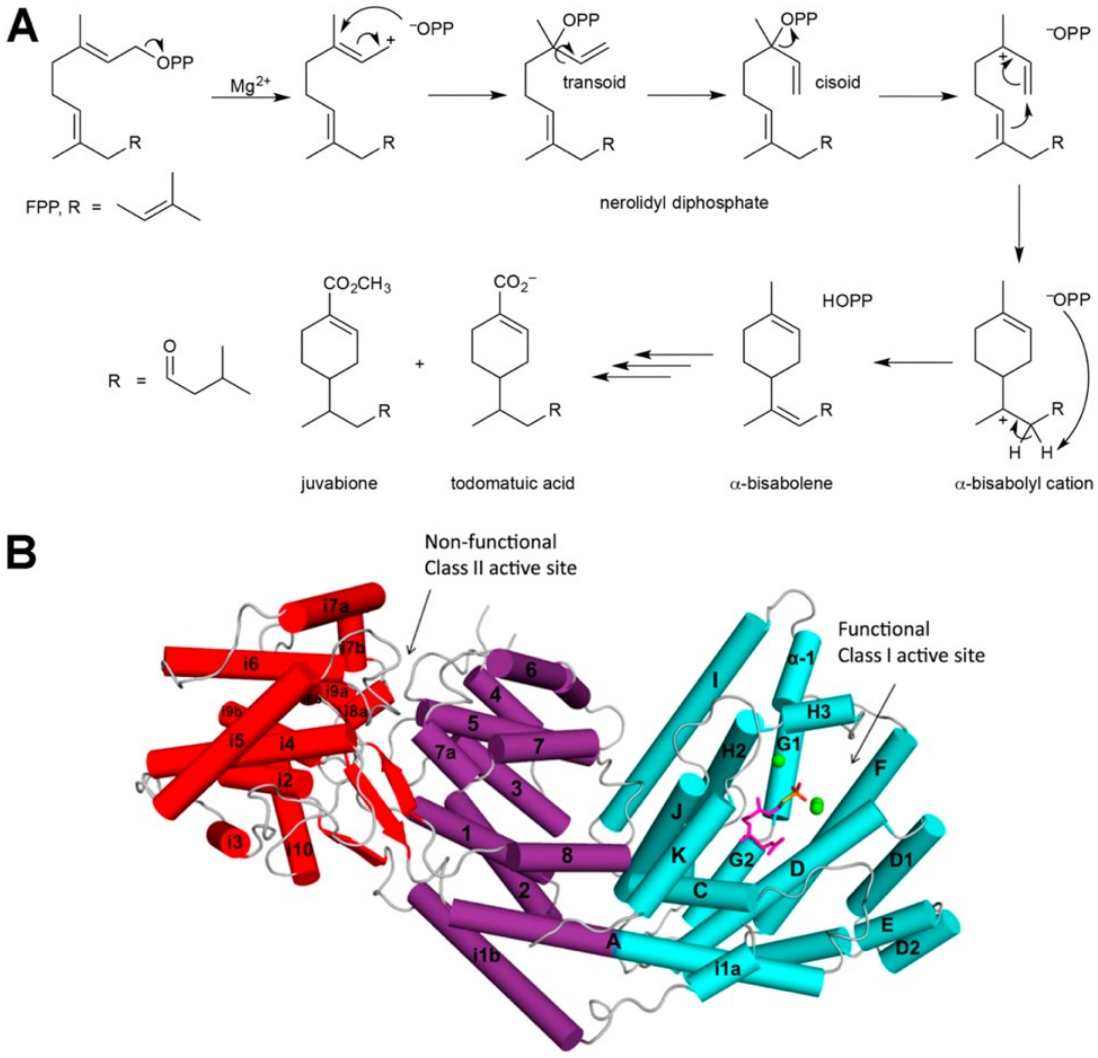

Figure 14. *(A) The α-bisabolene synthase employs a distinctive ionization-recombination-reionization sequence to facilitate the formation of C1−C6 bond, thereby initiating the catalytic mechanism. The resulting cyclization product, α-bisabolene, serves as a biosynthetic precursor for juvabione and todomatuic acid, both of which act as analogs of insect juvenile hormones. (B) The crystal structure of α-bisabolene synthase reveals a tripartite αβγ domain architecture. The α domain (cyan) is a functional class I terpenoid cyclase active site, whereas the βγ domains (purple and red, respectively) are vestigial and lack any known catalytic function. Figure was reproduced from*[70].

1.2.5 Class I Diterpene cyclases

The initial cyclization of geranylgeranyl-PP is facilitated by metal-dependent ionization and departure of the pyrophosphate group, leading to various bond-forming reactions such as C1−C6, C1−C7, C1−C10, C1−C11, C1−C14, and C1−C15, depending on the specific carbon-carbon double bonds involved. In addition, a C3−C8 class I

cyclization reaction can occur with a bicyclic diterpene diphosphate substrate generated in a class II cyclization reaction with geranylgeranyl-PP.

Further cyclization reactions, hydride transfers, and methyl migrations may also take place before the reaction terminates by proton elimination or the addition of a solvent molecule. Compared to sesquiterpene cyclization reactions, diterpene cyclization reactions yield even more diverse carbon skeletons due to the additional carbon-carbon bond-forming trajectories available to the larger substrate.

The α domain of an enzyme with overall α or αβγ domain architecture serves as an active site contour, acting as a catalyst for diterpene cyclases, ensuring that substrate geranylgeranyl-PP and subsequently formed intermediates are guided through the correct sequence of reactions leading to the correct product(s)[43]. Examples of class I diterpene synthases are taxadiene synthase and *ent*-kaurene synthase.

1.2.5.1 Taxadiene synthase

Taxadiene synthase is responsible for the first committed step of paclitaxel biosynthesis, generating the tricyclic hydrocarbon skeleton of paclitaxel with structural and stereochemical precision. Initially isolated from the Pacific yew[71], taxadiene synthase was later cloned and expressed in *E. coli* as an 862-residue full-length protein[72]. The full-length protein includes a plastidial targeting sequence at the N-terminus, and a construct lacking this sequence was prepared for the overexpression of a pseudomature form of the enzyme for functional analysis[73]. While the major cyclization product of recombinant pseudomature taxadiene synthase is taxa-4(5),11(2)-diene, minor side products taxa-4(20),11(12)-diene (∼5%) and verticillene (∼1%) are also generated.

The chemical mechanism of geranylgeranyl-PP cyclization catalysed by taxadiene synthase has been extensively studied, using isotopically labelled substrates and fluorinated substrate analogues. Metal-triggered ionization of the substrate diphosphate group enables C1−C14 bond formation with inversion of configuration at C1, possibly in concert with C10−C15 bond formation, to yield the verticillyl cation intermediate[74]. An unusual intramolecular proton transfer between C11 and the re face of C7 followed by transannular ring closure yields the taxenyl cation, which undergoes proton elimination to yield taxadiene (Figure 15, page 27)[75].

The crystal structures revealed αβγ domain architecture, with the class I active site residing in the α domain, while the βγ domains have no known catalytic function, other than perhaps to provide the N-terminal polypeptide segment to help cap the active site upon substrate binding (Figure 16, page 28). The N-terminus of the plastidial targeting sequence and an additional 27 residues were deleted from the taxadiene synthase construct that yielded crystals, attenuating but not obliterating catalytic activity[63].

Taxadiene synthase has been utilized for taxadiene overproduction, e.g., in *E. coli*[76] or yeast (*Saccharomyces cerevisiae*)[77]. Metabolic engineering approaches

have been explored for the biosynthesis of taxadiene and other paclitaxel biosynthetic intermediates with the aim of providing a more economical and renewable source of the drug through synthetic biology. Recent studies show that the combination of mutagenesis and pathway engineering yields an alternative approach for paclitaxel generation through the alternative cyclization product taxa-4(20)-11(12)-diene, which is a minor product generated by the wild-type enzyme. These studies suggest a promising future for the manufacturing of paclitaxel through synthetic biology, given the high demand for paclitaxel in cancer chemotherapy[78].

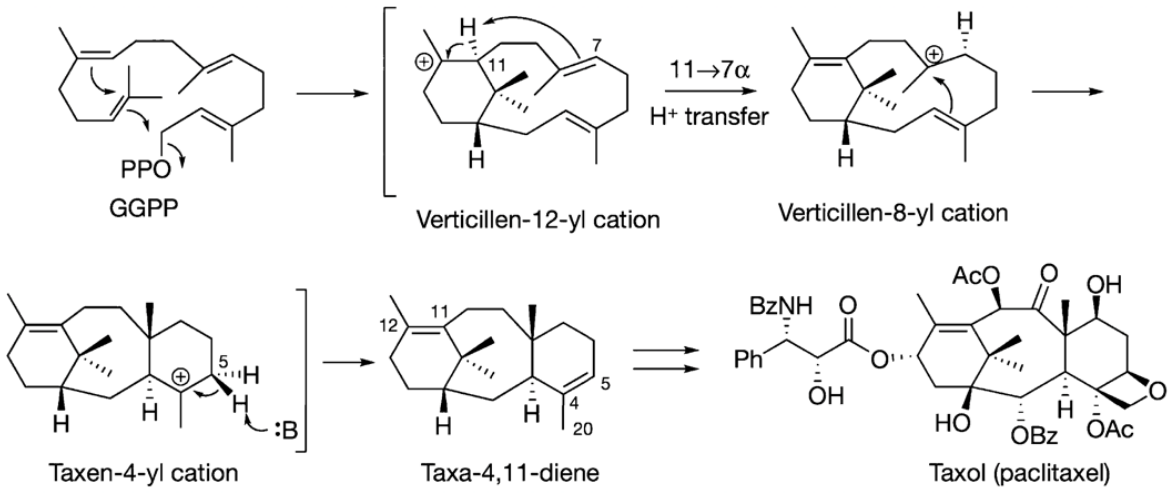

Figure 15. *The catalytic mechanism underlying taxadiene synthesis. Taxadiene is also an intermediate compound undergoing a series of biosynthetic modifications, culminating in the production of Taxol. Figure was reproduced from*[63].

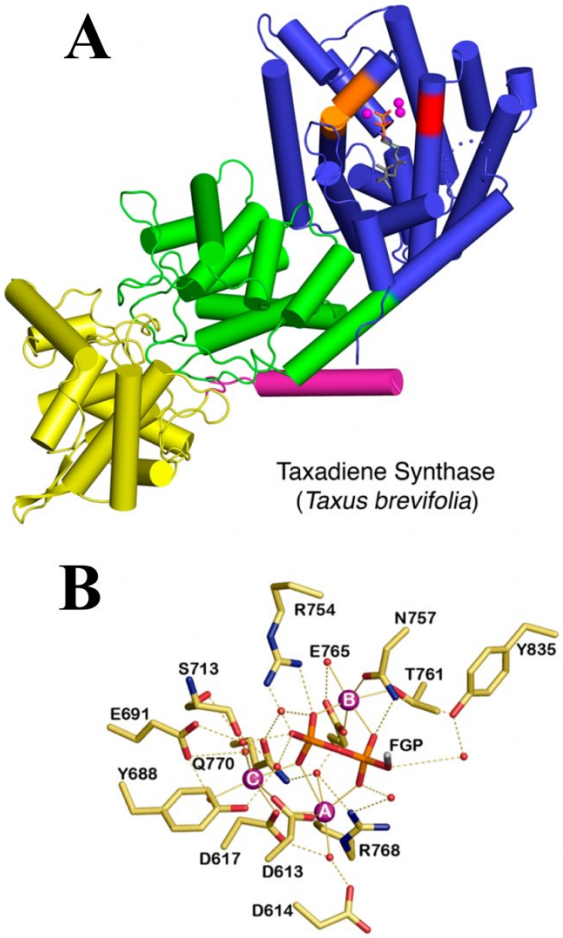

Figure 16. *(A) Taxadiene synthase exhibits αβγ domain architecture (α domain, blue; β domain, green; γ domain, yellow; N-terminus, magenta). The active site of taxadiene synthase is located in the α domain, where 3 Mg2+ ions (magenta spheres) stabilize the diphosphate group of a bound substrate analogue (stick figure). (B) Stereoview of the Mg2+ 3 cluster in the active site of taxadiene synthase. The isoprenoid moiety of substrate analogue 2-fluorogeranylgeranyl diphosphate (FGP) is truncated for clarity. Figure was reproduced from*[79] *and adjusted.*

1.2.6 Class II Diterpene cyclases

Class II diterpene cyclases initiate the catalytic process through the protonation of the terminal π bond of geranylgeranyl-PP, which occurs concurrently with the formation of the C10−C15 and C6−C11 bonds. This process results in the generation of a bicyclic tertiary carbocation intermediate. Similar to cyclization reactions facilitated by class I terpenoid cyclases, this carbocation intermediate can undergo subsequent reactions, such as hydride transfers and methyl migrations, before the reaction is terminated by proton elimination or the addition of a solvent molecule.

Thus far, two crystal structures of monofunctional class II diterpene cyclases have been resolved. These enzymes are ent-copalyl diphosphate synthases, with one originating from a plant and the other from a bacterium. Both enzymes employ identical chemical strategies for geranylgeranyl-PP activation and cyclization[43].

1.2.6.1 ent-Copalyl diphosphate synthase

The labdane-related diterpenoids are a group of natural products with over 7000 members found in plants, fungi, and bacteria[80]. In plants, *ent*-copalyl diphosphate is a labdane diterpenoid intermediate in the production of gibberellin phytohormones, which are important for plant growth and development[81]. The enzyme *ent*-copalyl diphosphate synthase converts geranylgeranyl-PP into *ent*-copalyl diphosphate (Figure 17).

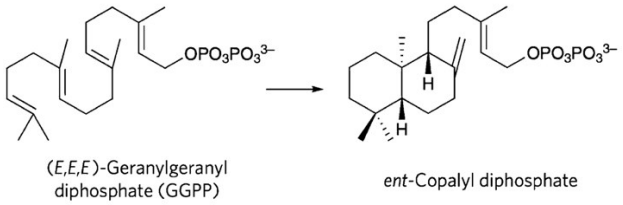

Figure 17. *Conversion of geranylgeranyl-PP to ent-copalyl diphosphate catalysed by ent-copalyl diphosphate synthase. Figure was reproduced from[79] and adjusted.*

The crystal structure of this enzyme has been studied in complex with substrate and product analogues. The active site cavity of the enzyme is hydrophobic, and its three-dimensional shape is defined by aromatic and aliphatic residues (Figure 18). The cyclization cascade is triggered by the protonation of C14 of geranylgeranyl-PP, which is facilitated by the anti-oriented proton of D379, and the cyclization cascade is terminated by proton elimination. The H263F and H263Y mutants of the enzyme reprogram the cyclization cascade leading to the formation of an alternative diterpene product, kolavenyl diphosphate, and this residue plays a critical role in directing the termination step of a class II terpenoid cyclase[82].

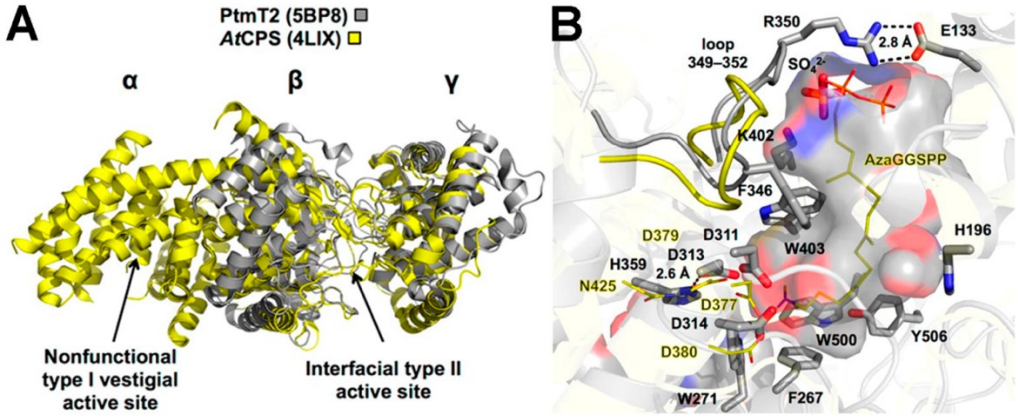

Figure 18. *(A) depicts the superposition of ent-copalyl diphosphate synthase from Arabidopsis thaliana, characterized by an αβγ domain architecture represented in yellow, and Streptomyces platensis CB00739, characterized by a βγ domain architecture represented in gray. (B) shows the superposition of the active sites of both enzymes. Notably, in the plant enzyme, N425 forms a hydrogen bond with general acid D379, while in the bacterial enzyme, H359 forms a hydrogen bond with general acid D313. Figure was reproduced from[83].*

1.2.7 Bifunctional terpenoid cyclases

The biosynthesis of terpenoids offers occasional instances of reaction sequences that entail the tandem cyclization of diterpenes or coupling-cyclization reactions, catalysed in two distinct active sites present on the same protein. Such systems may provide a catalytic advantage through a proximity effect, referred to as "proximity channelling," even in the absence of a direct channel between active sites[84]. This proximity channelling would only be advantageous if the active sites were appropriately oriented against each other. In an oligomeric protein, a catalytic advantage may result from "cluster channelling." Although the probability of a product of one reaction serving as the substrate for the second reaction in a bifunctional enzyme might be low if the active sites are not appropriately oriented, the probability of the second reaction being catalysed by one of the multiple subunits in an oligomeric assembly or aggregate may be higher (Figure 19)[85].

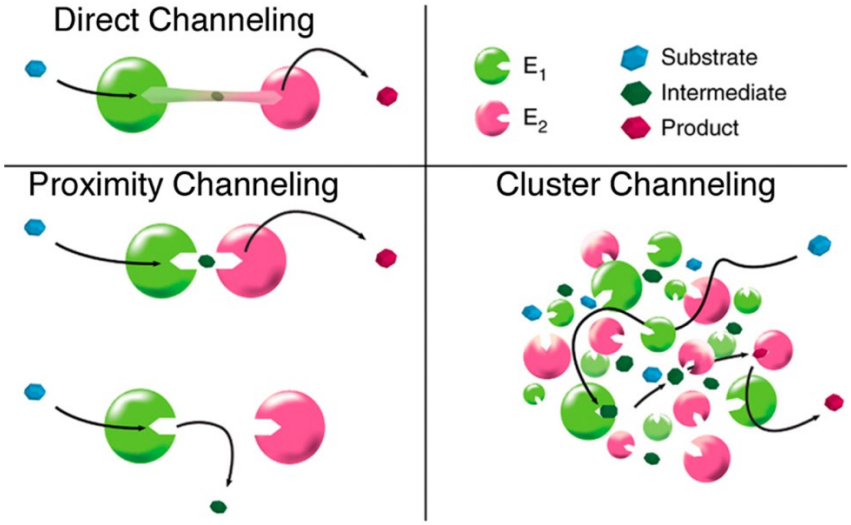

Figure 19. *In the context of enzymology, the present discussion pertains to the catalytic activities of hypothetical enzymes denoted as E1 and E2. These enzymes are responsible for carrying out successive biosynthetic reactions, wherein the initial substrate is converted by E1 into an intermediate product. This intermediate product subsequently acts as a substrate for E2, leading to the production of the final desired product. The active sites of E1 and E2 can be physically linked by a direct channel, or alternatively arranged in a manner that may promote or inhibit proximity channelling. It is noteworthy that if E1 and E2 form a multienzyme oligomer or aggregation, cluster channelling can augment the biosynthetic flux of the final product. Specifically, the intermediate product synthesized by E1 can diffuse into multiple nearby E2 active sites, resulting in an increase in the rate of the second reaction in the biosynthetic sequence. Figure was reproduced from*[85].

1.2.7.1 Fusicoccadiene synthase

The diterpene glucoside fusicoccin A is generated by the fungus *Phomopsis amygdali*[86] and has antitumor activity due to its ability to modulate protein-protein interactions involving the 14-3-3 protein and inhibit focal adhesion kinase[87]. The first step in the biosynthesis of fusicoccin A is coupling of one molecule of dimethylallyl-PP with three isopentenyl-PP molecules catalysed by fusicoccadiene synthase to produce geranylgeranyl-PP. The subsequent cyclization of geranylgeranyl-PP forms fusicoccadiene and is also catalysed by fusicoccadiene synthase (Figure 20)[88].

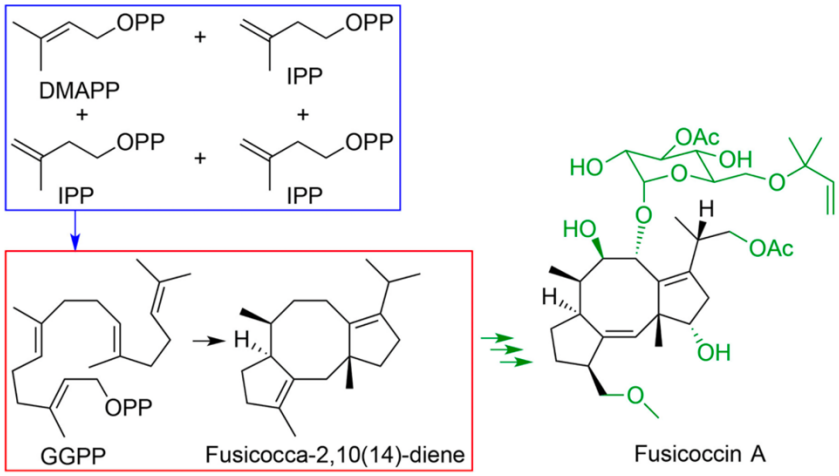

Figure 20. *The C-terminal α domain of fusicoccadiene synthase serves as a catalyst for the coupling of dimethylallyl-PP with three molecules of isopentenyl-PP, as indicated by the blue box. The resulting molecule of geranylgeranyl-PP is subsequently cyclized in the N-terminal α domain to yield fusicoccadiene, denoted by the red box. Furthermore, fusicoccin A is produced through additional biosynthetic modifications, as highlighted by the green modifications. Figure was reproduced from*[89].

The crystal structures of the individual N-terminal and C-terminal α domain constructs of fusicoccadiene synthase were determined, revealing substrate-induced conformational changes that enclose the active site (Figure 21, A and B, page 32)[89]. The hexameric quaternary structure of the geranylgeranyl-PP synthase domain is similar to that of human geranylgeranyl-PP synthase[90]. The crystal structure of the N-terminal cyclase domain complexed with 3 Mg^{2+} ions and pamidronate reveals a monomeric enzyme in which the active site is fully enclosed and sequestered from bulk solvent (Figure 21, C and D, page 32) . The full-length fusicoccadiene synthase is a hexamer in solution, and the hexameric structure of the geranylgeranyl-PP synthase domain was used as a starting point to interpret a low-resolution molecular envelope of full-length fusicoccadiene synthase determined by small-angle X-ray scattering. The model of hexameric fusicoccadiene synthase provides a plausible starting point for studying the catalytic importance of quaternary structure in this system, specifically the possibility of cluster channel[91].

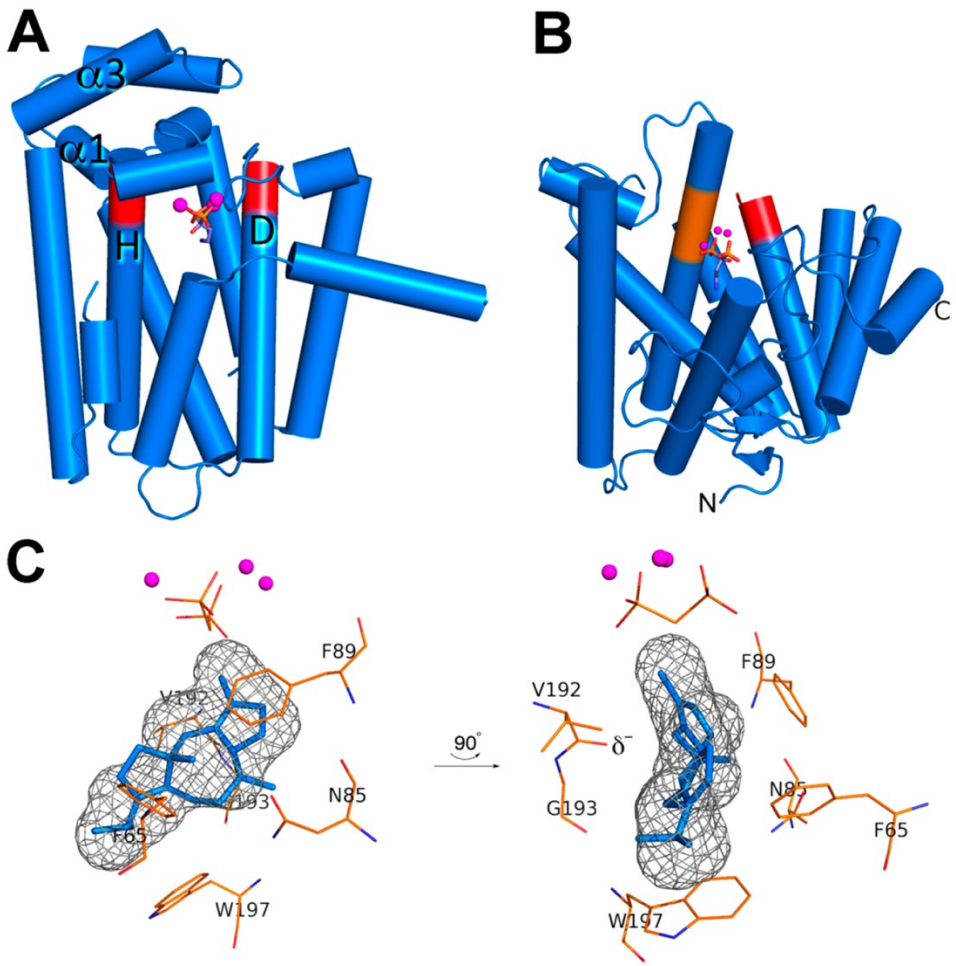

Figure 21. *The crystal structures of the C-terminal geranylgeranyl-PP synthase domain (A) and the N-terminal cyclase domain of fusicoccadiene synthase (B) are presented. (A) Metal-binding motifs comprising aspartate-rich DDXXD motifs on helices D and H (depicted in red) bind to three Co^{2+} ions (represented by magenta spheres) and the bisphosphonate inhibitor pamidronate (represented by a stick figure). This results in the full closure of the active site. (B) DDXXD and NSE metal-binding motifs on helices D and H (depicted in red and orange, respectively) coordinate with three Mg^{2+} ions (represented by magenta spheres) and pamidronate (represented by a stick figure), locking the active site in a fully closed conformation. (C) The molecular surface of the enclosed active site in the N-terminal cyclase domain is depicted in a gray meshwork, into which the product fusicoccadiene is docked. The three-dimensional contour of the active site closely resembles the product and reflects the active site's role as a catalytic template. Figure was reproduced from[89].*

1.2.8 Approaches to search for novel terpene synthases

Many previously discussed terpene synthases such as taxadiene synthase or fusicoccadiene synthase synthetize compounds of significant medicinal value. The discovery of new terpene synthases is important for several reasons. First, it allows for the identification and characterization of novel terpenes, which can have unique and valuable properties.

Second, the identification of new terpene synthases can provide insight into the biosynthesis of terpenes, which can help researchers understand how these

compounds are produced in nature. This knowledge can then be used to engineer new terpene biosynthesis pathways in plants and microbes, which could be used for a variety of applications, such as enhancing plant growth or producing new flavours and fragrances.

Finally, the discovery of new terpene synthases can lead to a better understanding of the evolution of terpene biosynthetic pathways. By studying the diversity of terpene synthases in different organisms, researchers can gain insight into the origins of these pathways and how they have evolved over time. Several different approaches have been utilized in search of novel terpene synthases including homology-based approaches, transcriptomic analysis and functional characterization.

1.2.8.1 Homology-based approaches

These approaches involve searching for homologs of known terpene synthases in genome or transcriptome databases *in silico*. This approach can be used to identify new members of existing terpene synthase families or to discover entirely new families.

Bc. Tereza Čalounová, a bioinformatics student from the Faculty of Natural Sciences, Charles University, and the Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, developed an *in silico* method for identifying homologs of TPSs from protein databases and for selecting potential enzymes for the production of novel scaffolds. This section briefly summarizes her work. However, a detailed account of her approach will be provided in her master thesis, "Mining novel terpene synthases from large-scale repositories" (as yet unpublished), and all source codes are currently available in her Github repository[*].

Her filtering algorithm was based on two assumptions:

1. Uncharacterised proteins (= with unknown function) possessing the same function determining domains as already described terpene synthases (TPSs) are likely to be TPSs.
2. Evolutionary more distant proteins with these domains are more likely to produce a novel compound than proteins showing high sequence similarity to already known TPSs.

To create a database of TPSs, Bc. Čalounová filtered out sequences from the Uniprot, Phytozome 13, 1KP, and NCBI TSA databases based on the presence of domains specific to TPSs (Table 1, page 34). The database comprises both experimentally evaluated enzymes and predicted TPSs based on homology.

[*] https://github.com/pluskal-lab/TPS_mining_v1

Table 1. *Table of protein domains specific to TPSs which were used by Bc. Tereza Čalunová as filtering criteria for homology. Domain accession names in Pfam protein database or InterPro protein database along with information about function of these protein domains are listed below.*

Domain accession names	Function
PF00494	Squalene/phytoene synthase
PF01397	Terpene synthase, N-terminal domain
PF06330	Trichodiene synthase
PF13249	Squalene-hopene cyclase, N-terminal domain
PF13243	Squalene-hopene cyclase, C-terminal domain
PF03936	Terpene synthase, C-terminal domain (metal binding)
PF19086	Terpene synthase, C-terminal domain 2 (metal binding)

The current approach yielded 191,451 TPSs comprising both experimentally evaluated and yet uncharacterised proteins. These TPSs were classified into different classes such as mono-, sesqui-, and diTPSs based on the presence of domains that are typical for each class. Among the vast dataset, identifying the enzymes capable of producing novel terpene scaffolds required further analysis. To achieve this, multiple sequence alignment was conducted using MAFFT 7.490 software[92]. The alignment was then trimmed with trimAl 1.4.1 software[93], and a phylogenetic tree was constructed using FastTree 2.0 software[94]. The web interface of iTOL[95] was used to visualize the resulting tree. Considering the dataset size was significant, Bc. Tereza Čalounová computed the branch distances between uncharacterized proteins and their closest characterized entries (source codes available in her Github repository). Additionally, she investigated the abundance of amino acid lengths among different TPS classes and created density distribution plots to observe the characteristic amino acid sequence and domain architecture of each TPS class.

To enhance the reliability of predictions, Bc. Čalounová devised a "reliability score". The formula of the reliability score was the following:

```
total_score = arch_scores[i] + type_scores[i] + (2*normalized_distances[i]) + completeness_scores[i] + length_scores[i]
```

Each parameter lies within interval from 0 to 1. This score favoured sequences which started with methionine (completeness_score, either 1 if the sequence starts with M or 0 if not), exhibited the typical amino acid length and domain architecture for the given terpene class (arch_scores and length_scores, ranges from 0 to 1), and were positioned further away in the phylogenetic tree from the nearest characterized entry (normalized_distances; see Figure 22). The search for TPSs responsible for producing new compounds was then confined to possible sesquiTPSs and diTPSs, thus proteins

anticipated to be sesquiTPSs or diTPSs were also prioritized (type_scores, either 1 for sesqui- or diTPS, or 0).

The output of this approach was a sheet with all 191,451 candidate proteins rated according to previously mentioned criteria. The complete source codes leading to this output is again available at the Github repository.

***Figure 22**. General scheme of phylogenetical analysis and its contribution to reliability score. Hypothetical proteins a-g with green dots are uncharacterised, branches with orange dots correspond to experimentally evaluated proteins with known functions. The parameter normalized_distances for each hypothetical protein a-g increases with the distance from an unannotated protein to its closest annotated protein. Figure was created in collaboration with Bc. Čalounová and adjusted.*

1.2.8.2 Functional characterization of proteins

This method entails examining candidate genes for their terpene synthase activity, which can be accomplished through either *in vitro* enzyme assays or *in vivo* functional complementation assays conducted in a heterologous system.

For example, Martin et. al. cloned and expressed a candidate gene for (−)-α-terpineol synthase in *E. coli* to produce the recombinant terpene synthase enzyme. The researchers conducted separate *in vitro* assays with geranyl-PP, farnesyl-PP and geranylgeranyl-PP substrates to confirm that the recombinant enzyme could produce (−)-α-terpineol from geranyl-PP. Products were identified by gas chromatography-mass spectrometry (GC-MS) after organic extraction. The authors also characterized the enzyme's kinetic properties, optimal reaction conditions, and substrate specificity[96].

1.2.8.3 Combinatory approaches

Kumar et al. aimed to identify the terpene synthase enzymes responsible for the terpene profiles in *Marchantia polymorpha*[97] . The authors used HMMER search software and conserved protein sequence domains associated with terpene synthases (PF01397, PF03936, and PF06330) to identify 17 putative terpene synthase transcripts. The transcripts were clustered into two groups, microbial terpene synthase-like contigs and diterpene synthases. The study found that microbial terpene synthases had low similarity to functionally characterized terpene synthases, while diterpene synthases had

greater similarity to other plant diterpene synthases. In total, they identified four diterpene synthase genes and 13 microbial terpene synthase-like transcript.

Microbial TPS-like candidate enzymes were encoded by MpMTPSL1 to MpMTPSL9 genes from *M. polymorpha*. The cDNA of these genes was heterologously expressed in *E. coli*, and the crude lysates were screened for substrate preferences in presence of geranyl-PP and its isomer neryl-PP, all trans and all cis farnesyl-PP and geranylgeranyl-PP separately. MpMTPSL3 and MpMTPSL4 demonstrated activity levels capable of converting the unusual all-cis configuration of farnesyl-PP to hydrocarbon products. MpMTPSL2 exhibited an unusual substrate preference for neryl-PP. MpMTPSL6 was found to be the most versatile synthase, with the ability to use both mono- (geranyl-PP) and sesquiterpene (farnesyl-PP) substrates.

Subsequently, functional characterization of putative *M. polymorpha* diterpene synthase candidate genes MpDTPS1, 3, and 4, using an *in vivo* bacterial expression platform. The study was unable to functionally characterize MpDTPS2 due to the inability to recover a full-length cDNA for this gene and observed a frame-shift mutation within the coding sequence. Their bacterial expression system relies on the generation of (E,E,E)-geranylgeranyl-PP by coexpression of a geranylgeranyl-PP synthase, plus one or more of the putative diterpene synthases. Bacterial cultures were then extracted, and the organic extract containing the diterpene synthase product analyzed by GC-MS. The study found that copalol was the only diterpene produced when MpDTPS3 was evaluated, indicating the production of copalyl-PP with dephosphorylation to the primary alcohol by endogenous phosphatases. Coexpression of MpDTPS3 with *ent*-copalyl-PP specific *ent*-kaurene synthase from *A.thaliana* yielded *ent*-kaurene. Coexpression of MpDTPS1 or MpDTPS4 with only geranylgeranyl-PP synthase did not yield any products. However, coexpression of MpDTPS1 or MpDTPS4 with geranylgeranyl-PP synthase and *ent*-copalyl-PP synthase yielded terpenoid products, namely *ent*-kaurene in case of both canidates and 2 other products (*ent*-atiserene and atiseran-16-ol) in case of MpDTPS1[97].

1.3 Terpenoid production using heterologous biosynthesis

As stated in chapter 1.1 Terpenes and terpenoids, terpenes are a significant category of specialized metabolites that find a wide range of applications in food, pharmaceuticals, and biofuels. Traditional methods of terpene production typically rely on plant extraction and chemical conversion. The plant extraction method is resource-intensive and poses challenges in the separation of the target compound from the extractives[98]. The chemical conversion method entails a complex synthesis process that results in severe environmental pollution [99]. In contrast, the microbial synthesis method offers advantages such as sustainability, low production cost, and environmental friendliness, making it a potential approach for achieving efficient terpene production in the future[100].

Microorganisms, particularly *E.coli* and *S.cerevisiae* have been established as advantageous organisms for heterologous biosynthesis of secondary metabolites. They have been genetically manipulated to enhance natural product biosynthetic pathways, which has resulted in the economical production of valuable compounds[101], in addition to the discovery of new ones. The latest progress in synthetic biology and genome engineering has facilitated the creation of novel approaches for constructing, regulating, and refining metabolic pathways in microorganisms for heterologous biosynthesis.

Microorganisms such as *E. coli* and *S.cerevisiae* exhibit terpenoid biosynthetic pathways that operate at low fluxes[101]. As stated in chapter 1.1.2 Biosynthesis, the five-carbon precursors necessary for terpenoid biosynthesis, are generated from MEP pathway in *E. coli* and the MVA pathway in yeast. In *S.cerevisiae*, ergosterol, the final metabolite resulting from these successive condensation reactions, is derived from C_{30} squalene (Figure 1), which is synthesized through the condensation of two farnesyl-PP molecules (Figure 5). The presence of adequate precursor molecules in sufficient quantities within host cells is essential for production of compounds of interest after introducing a TPS gene[102]. However, host organisms need terpene precursor for sustaining their growth and overexpression of TPS leads to depletion of essential metabolites. To solve this problem, several strategies have been established.

1.3.1 Strategies for expression optimization

1.3.1.1 Codon optimization

Codon optimization is the process of designing DNA sequences such that the codons (a sequence of three nucleotides) used to encode amino acids are optimized for the expression system being used. The goal of codon optimization is to increase the production of a protein by ensuring that the codons used to encode the amino acids

match the codon usage bias of the organism in which the protein is being expressed. This can be achieved by replacing the codons from the native sequence with a low frequency in the heterologous host by codons which are frequently used by the target organism. Codon optimization has become an important tool in synthetic biology, where it is used to improve protein expression, increase protein yield, and enhance protein solubility[103].

1.3.1.2 Promoter engineering

Promoter engineering refers to the genetic modification of the promoter region of a gene, which is responsible for regulating the expression of the gene. This can involve changing the sequence or structure of the promoter to enhance or suppress gene expression, depending on the desired outcome.

The strength of a promoter determines the level of gene expression, i.e., how much of a specific protein is produced. One strength of a promoter is its ability to recruit RNA polymerase, which is the enzyme responsible for transcription. Strong promoters recruit RNA polymerase more efficiently, resulting in higher levels of gene expression, which is preferred when expressing an enzyme producing low-toxicity compound. However, such promoter can put a metabolic burden on host cell. Conversely, weak promoters recruit RNA polymerase less efficiently, resulting in lower levels of gene expression[104].

1.3.1.3 Metabolic engineering of the host

Metabolomic engineering is a field of synthetic biology that involves manipulating and optimizing metabolic pathways in microorganisms to produce desired compounds for various applications. This approach typically involves modifying the genetic makeup of the microorganism and optimizing its growth conditions to increase the production of the desired metabolite precursor[105].

Zhuang et al. for example developed a yeast production platform through an impartial genetic selection strategy. In this approach, the yeast strain BY4741 (Figure 23, page 39) was subjected to ethyl methane sulfonate mutagenesis, followed by selection for growth in the presence of nystatin, squalestatin, and exogenous cholesterol. The purpose of this screening was to identify mutant yeast strains that possessed sterol uptake enhancement mutations that support the aerobic uptake of exogenous sterols. Following this step, the selected mutants were screened for high-level accumulation of farnesol, which is an indicator of the accumulation of the key intermediate farnesyl-PP for producing sesquiterpene. To further enhance the farnesyl-PP pool in these mutants, insertional mutations were introduced into the *ERG9* gene, which codes for squalene synthase, in the lines that could accumulate at least 50 mg farnesol per liter. This led to the development of another series of lines that could accumulate farnesol levels above 70 mg/L in small-scale shake cultures. To assess

the utility of these lines as a general production platform for specific terpenes, selected *SUE/erg9* lines were transformed with a vector containing the *Hyoscyamus muticus* premnaspirodiene synthase gene that encodes for a sesquiterpene synthase. The newly generated yeast line ZX178-08 (Figure 23, page 39) exhibited the highest level of premnaspirodiene accumulation, up to 116 mg/L, with FOH levels of 23.6 mg/L. In contrast, the parental line BY4741 produced 10 times less premnaspirodiene (10.94 mg/L) and no detectable farnesol, making strain ZX178-08 a useful tool for heterologous biosynthesis of sesquiterpenes[106].

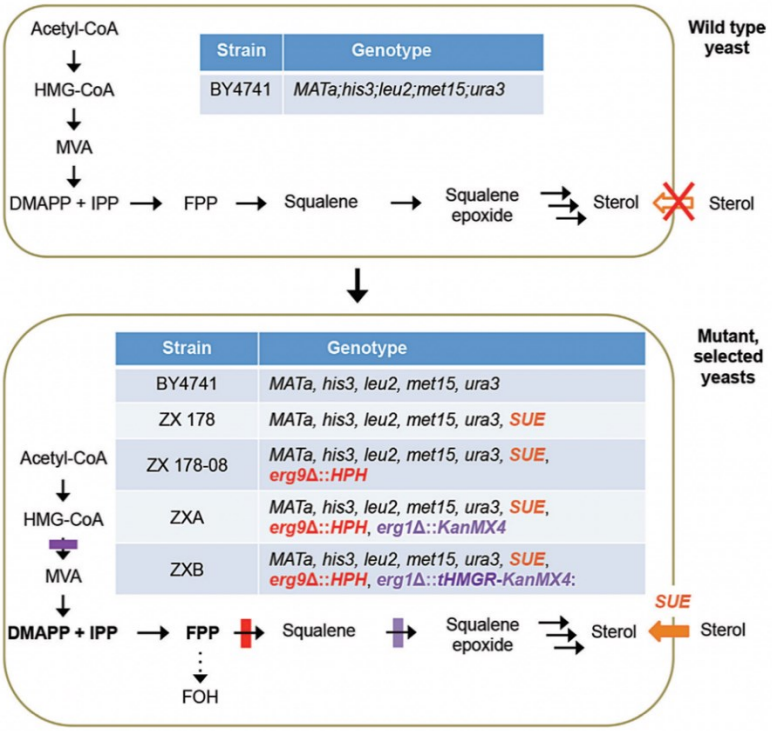

Figure 23. *General scheme of metabolomic engineering starting from wild type yeast strain BY4741 by redirecting its ergosterol bioproduction to endogenous ergosterol uptake and resulting to ZX178-08 strain with increased ability to cumulate an important sesquiterpene synthase substrate - farnesyl-PP. Figure was reproduced from*[107].

Wong et al. describe the engineering of *Saccharomyces cerevisiae* to produce high-titer amounts of lathyrane diterpenoids from sugar. Lathyrane diterpenoids are a class of natural products with potential pharmacological activity, but their production from natural sources is limited. The authors engineered a geranyl geranyl-PP overproducing strain JWY501 by removing ura 3-52 prototrophy from parentals strain GTY116 {*MATa, leu2-3,112::HIS3MX6-GAL1p-ERG19/GAL10p-ERG8;ura3-52::URA3-GAL1p-MvaSA110G/GAL0p-MvaE (codon optimized); his3Δ1::hphMX4-GAL1p-ERG12/GAL10p-IDI1;trp1-289::TRP1_GAL1p-CrtE(X.den)/GAL10p-ERG20;YPRCdelta15::NatMXGAL1p-CrtE(opt)/GAL10p-CrtE)*} with galactose inducible pathway for geranyl geranyl-PP overproduction to decrease metabolomic burden on cells. Engineered strain JWY501 proved itself as a suitable host for enzymes utilizing geranyl geranyl-PP for biosynthesis[107].

Then the authors engineered the yeast to express genes involved in the biosynthesis of lathyrane diterpenoids and optimized the fermentation conditions to achieve high yields. The results showed that the engineered yeast strain was able to produce lathyrane diterpenoids at a high titer, jolkinol C production reaching up to 800mg/L making it a promising platform for the industrial production of these compounds.

1.3.1.4 Removal of signal peptides

Plants direct their terpene production to chloroplasts[108,109,110]. Signal peptides in plant terpene synthases need to be removed prior to expression in heterologous host, because yeast and bacterial cells lack the necessary enzymes to correctly process and modify plant-specific signal peptides, which can result in incorrect localization, folding and activity of the expressed protein[111,112,113]. Screening for presence of signal peptides can be done *in silico* using the TargetP[114] and SignalP[115] webtools.

1.3.1.5 Modular cloning

Heterologous expression of a TPS is done by introducing a plasmid carrying a TPS gene to host cells. This expression vector can be designed using a modular cloning approach. Modular cloning is a technique used in molecular biology to assemble DNA fragments rapidly and efficiently into a desired construct. It involves the use of standardized genetic parts, or "modules," such as promoters, terminators, or selection markers, which can be easily combined in different combinations to generate a wide range of complex DNA constructs[116].

To give an example, Lee et al. engineered a highly characterised yeast toolkit for modular, multipart assembly for *Saccharomyces cerevisiae*[117]. The toolkit is based on the Modular Cloning system and uses type II restriction enzymes BsmBI and BsaI to generate DNA fragments that can be assembled in a precise and flexible manner. Type II restriction enzymes cleave DNA at a specific site near the recognition site and do not require ATP or S-adenosyl methionine for their activity. BsaI recognizes the sequence `GGTCTC` and BsmBI recognizes the sequence `CGTCTC` [117].

The MoClo yeast toolkit consists of a set of 96 part plasmids in *E.coli* DH10β (*str. K-12F⁻ Δ(ara-leu)7697[Δ(rapA'-cra')]Δ(lac)X74[Δ('yahH-mhpE)] duplication(514341-627601)[nmpC-gltI] galK16galE15 e14⁻(icd^{WT} mcrA) φ80dlacZΔM15 recA1 relA1 endA1 Tn10.10 nupG rpsL150(Str^{R}) rph⁺ spoT1 Δ(mrr-hsdRMS-mcrBC) λ⁻ Missense(dnaA glmS glyQ lpxK mreC murA) Nonsense(chiA gatZ fhuA? yigA ygcG) Frameshift(flhC mglA fruB)* strain stocks. Each part plasmid encodes for different Type II restriction enzymes and a "part type" such as promoter or terminator (Figure 24, B, page 41) as well as bacterial resistance marker to enable antibiotic selection. These "part types" possess predetermined flanking overhangs, which guarantee interchangeability of component. An expression vector, also called cassette plasmid in case of single gene assembly (Figure 24, C, page 41), consists of assembly

connectors, promoter, coding sequence, terminator, yeast marker, yeast origin and bacterial marker and origin (Figure 24, B, page 41). Construction of multigene plasmids containing multiple transcriptional units is also possible with this kit (Figure 24, D). Firstly, DNA selected for cloning is introduced to an entry vector using BsmBI, replacing encoded fluorescent protein with a new coding sequence (Figure 24, A). Secondly, part plasmids encoding desired part types are assembled into a cassette plasmid using BsaI. Additionally, cassette plasmids can be assembled into a multigene plasmid using BsmbI reaction.

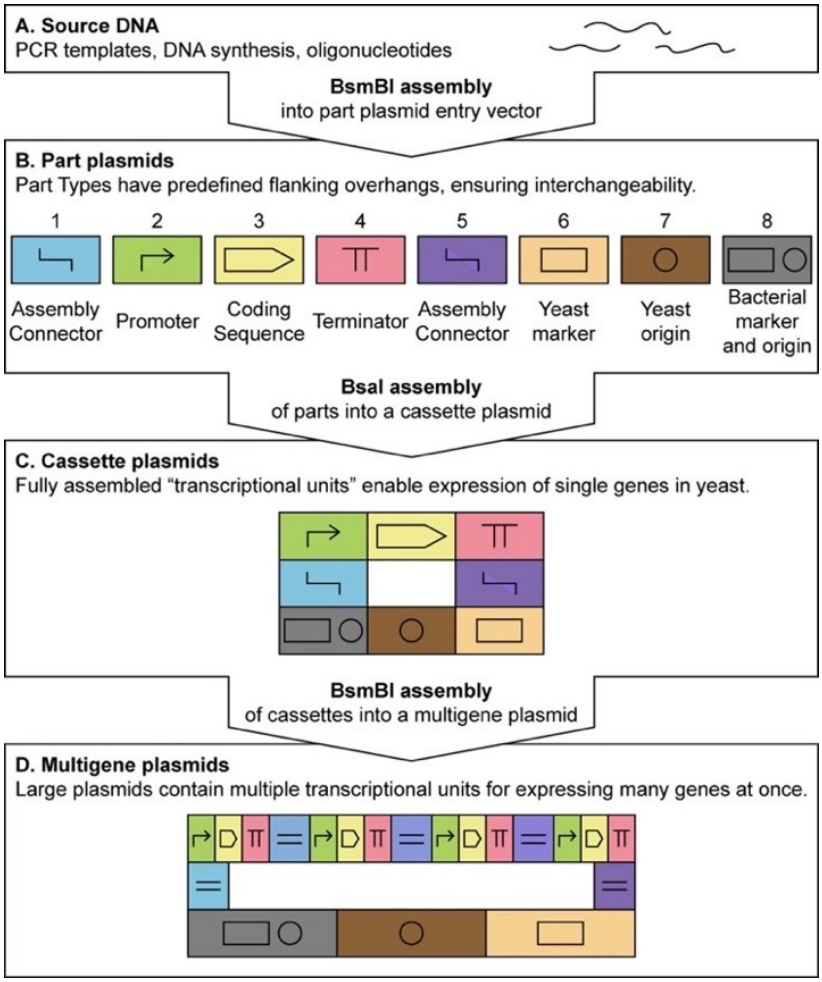

Figure 24. *Hierarchical assembly strategy based on modular cloning. (A) The source DNA can be obtained through PCR, DNA synthesis, or oligonucleotides, and subsequently assembled using BsmBI into a part plasmid entry vector. (B) Part plasmids belonging to a particular type have unique upstream and downstream BsaI-generated overhangs, rendering them interchangeable. At this stage, part plasmids typically confer chloramphenicol resistance in E. coli. (C) One representative part plasmid of each type is assembled using BsaI, resulting in a cassette plasmid that contains a complete transcriptional unit and can be directly transformed into yeast. At this stage, cassette plasmids typically confer ampicillin resistance in E. coli. (D) Cassette plasmids can be further assembled using BsmBI to form a multigene plasmid. The latter contains multiple transcriptional units, whose order is defined by the assembly connector parts employed to flank the individual cassettes. Plasmids at this stage typically confer kanamycin resistance in E. coli. Figure was reproduced from*[117].

One of the key advantages of the MoClo yeast toolkit is its flexibility and efficiency. It can be used to assemble DNA fragments of different sizes and sequences and can be easily customized to meet the specific needs of a particular research project. Additionally, the MoClo yeast toolkit is compatible with a wide range of other genetic tools and techniques, making it a valuable resource for genetic engineering in yeast cells[117].

2. Aims of this thesis

The aims of this thesis are:

- to pick out up to 10 candidates for heterologous expression from a list of uncharacterised proteins, which were assigned terpene synthases *in silico* and exhibit potential for the synthesis of a novel terpene product
- to experimentally evaluate *in silico* prediction using methods of molecular biology involving following steps:
    - selection a suitable host for heterologous expression of TPSs
    - design and preparation of expression constructs
    - heterologous expression of selected protein candidates
    - analysis of potential enzymatic products via GC-MS and LC-MS

3. Material and methods

3.1 Instruments

Analytical scales AS R2 PLUS	RADWAG, Poland
Autoclave	Sanyo, Japan
Automatic pipets	Gilson, USA
Axygen® 24-Well Deep Well Plate	Corning, USA
Centrifuge MiniSpin Plus	Eppendorf, Germany
Centrifuge 5427R	Eppendorf, Germany
Centrifuge 5430R	Eppendorf, Germany
Centrifuge Labnet Prism™ Mini Centrifuge	Labnet International, USA
Electroporator ®	Eppendorf, Germany
Flowbox Eco safe Comfort Plus	Envair, Germany
GC-MS 7890A GC	Agilent Technologies, USA
5975C MS with EI and quadrupole	Agilent Technologies, USA
Hot plate stirrer 10X10" DIGITAL	Benchmark Scientific Inc., USA
Imaging system Azure 600	Azure biosystems, USA
Incubator INCU-Line®	VWR, USA
LC-MS Vanquish™ Flex UHPLC System,	Thermo Fisher Scientific, USA
Orbitrap ID-X Tribrid MS	Thermo Fisher Scientific, USA
Milli-Q® Type 1 Ultrapure Water Systems	MilliporeSigma, Germany
NanoDrop™ UV-Vis Spectrophotometer	Thermo Fisher Scientific, USA
pH meter 3510 Standard Digital pH Meter Kit	Jenway, UK
PowerPac™ Basic Power Supply	Bio-Rad Laboratories, USA
Shaking incubator ThermoMixer C	Eppendorf, Germany
Shaking incubator Innova™ 40R	Eppendorf, Germany
Shaking incubator Model SI4, SHEL LAB	Sheldon Manufacturing, Inc., USA
Spectrophotometer BioSpectrometer®	Eppendorf, Germany
Thermocycler ProFlex PCR System	Thermo Fisher Scientific, USA
Thermoshaker rant-Bio PHMT thermoshaker	Grant Instruments, UK
Vortex BioSan Vortex V-1 plus	Biosan, Germany
250mL Bottle-Top-Filter, 0,2_µm PES-Membrane	VWR, USA

3.2 Chemicals and strains

3.2.1 Chemicals

Agar	Difco Laboratories Ltd., USA
Agarose	Thermo Fischer Scientific, USA
Ampicillin	Sigma-Aldrich, USA
Bacto agar	Difco Laboratories Ltd., USA
Ergosterol	Sigma-Aldrich, USA
Ethanol	Penta chemicals, Czech Republic
Ethyl acetate	Lach-Ner s.r.o, Czech Republic
Galactose	Sigma-Aldrich, USA
GelRed Nucleic Acid Gel Stain, 10,000×	Thermo Fischer Scientific, USA
GeneRuler 1 kb plus DNA ladder	Thermo Fischer Scientific, USA
Glucose	Lach-Ner s.r.o, Czech Republic
Guaiazulene	Tokyo Chemical Industry, Japan
Chlorapmhenicol	Sigma-Aldrich, USA
KCl	Sigma-Aldrich, USA
K_2HPO_4	Sigma-Aldrich, USA
KH_2PO_4	Sigma-Aldrich, USA
KOH	Lachema, Czech Republic
LB media premix	Carl Roth, Germany
Lithium acetate	Sigma-Aldrich, USA
$MgCl_2$	Sigma-Aldrich, USA
$MgSO_4 . 7 H_2O$	Sigma-Aldrich, USA
NaCl	Penta chemicals, Czech Republic
NaOH	Sigma-Aldrich, USA
PEG 3350	Sigma-Aldrich, USA
Peptone	OXOID CZ s.r.o., Czech Republic
Phire master mix	Thermo Fischer Scientific, USA
Salmon sperm DNA	Sigma-Aldrich, USA
SDS	Carl Roth, Germany
Sorbitol	Penta chemicals, Czech Republic
Succinic acid	Penta chemicals, Czech Republic
T4 DNA ligase buffer	New England BioLabs, USA
TAE	Thermo Fischer Scientific, USA
Tryptone	Sigma-Aldrich, USA
Yeast dropout mixture without uracil	Sigma-Aldrich, USA
Yeast nitrogen base without amino acids	Sigma-Aldrich, USA

3.2.2 Solutions and media

LB media: 1 % (w/v) tryptone, 1 % (w/v) NaCl, 0.5 % (w/v) yeast extract

SCE media: 0.6 % (w/v) succinic acid, 0.45 % (w/v) KOH, 2 % (w/v) glucose, 10 % (v/v) YNB(U-), 40 mg/L ergosterol, pH 5.3 – adjusted with HCl

S.O.C. media: 0.5 % (w/v) yeast extract, 2 % (w/v) tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20mM glucose

YNB+D media without uracil (U-): 6.7 % (w/v) yeast nitrogen base without amino acids, 1.4 % (w/v) yeast dropout mixture without uracil 2 % (w/v) glucose

YNB+G+D media without uracil (U-): 6.7 % (w/v) yeast nitrogen base without amino acids, 1.4 % (w/v) yeast dropout mixture without uracil, 2 % (w/v) galactose, 0.2 % (w/v) glucose

YNBE+D media without uracil (U-): 6.7 % (w/v) yeast nitrogen base without amino acids, 1.4 % (w/v) yeast dropout mixture without uracil, 2 % (w/v) glucose, 40 mg/L ergosterol

YPD media: 1 % (w/v) yeast extract, 2 % (w/v) peptone, 2 % (w/v) glucose

YPDE media: 1 % (w/v) yeast extract, 2 % (w/v) peptone, 2 % (w/v) glucose, 40 mg/L ergosterol, pH 5.3 – adjusted with HCl

YPD 2× media: 2 % (w/v) yeast extract, 4 % (w/v) peptone, 4 % (w/v) glucose

YPDE 2× media: 2 % (w/v) yeast extract, 4 % (w/v) peptone, 4 % (w/v) glucose, 40 g/L ergosterol, pH 5.3 – adjusted with HCl

Solid forms of any media contained 2 % (w/v) of agar.

Sorbitol wash buffer for ZX178-08: 22 % (w/v) sorbitol, 70 % (v/v) 0.1M K_2HPO_4, 30 % (v/v) 0.1M KH_2PO_4, pH 7.5 – adjusted with KOH

LC mobile phase A: water with 0.1 % (v/v) formic acid

LC mobile phase B: acetonitrile with 0.1 % formic acid

Lysing solution for yeast colony PCR: 200mM lithium acetate, 1% (w/v) SDS

3.2.3 Enzymes

T4 DNA ligase	New England BioLabs, USA
BsaI-HF®v2	New England BioLabs, USA
BsmBI-v2	New England BioLabs, USA
Phire Hot Start II DNA polymerase	Thermo Fischer Scientific, USA

3.2.4 Strains

E.coli DH10β	Thermo Fischer Scientific, USA
S.cerevisiae JWY501	Ernest Orlando Lawrence Berkeley National Lab, USA
S.cerevisiae ZX178-08	University of Kentucky, USA

Genotypes of both yeast strains are listed in chapter 1.3.1.3 and for bacterial strains in chapter 1.3.1.5

3.2.5 Commercial kits

MoClo-YTK	Addgene, USA
QIAprep Spin Miniprep Kit	Quiagen, Germany

3.3 In silico search for candidates

The objective was to identify up to 10 candidates for heterologous gene expression. Concerning the assumption that diterpenes are more likely to be structurally diverse than sesquiterpenes, the goal was to increase the proportion of diTPS candidates in the final list of proteins for heterologous expression, because we think they hold more potential for producing novel structures. To ensure diversity in the dataset, at least one bacterial, fungal, and plant sequence needed to be present in the final candidate list. Additionally, each candidate had to originate from a different organism. That led to a choice of expressing 9 proteins, 6 of them being diTPS (1 bacterial and 5 from plants) and 3 sesquiTPS (1 bacterial, 1 fungal, 1 from a plant) candidates.

3.4 Design and order of TPS candidate genes

First, all protein sequences picked up for heterologous expression were checked for presence of signal peptides using DeepLoc, SignalP and TargetP tools. To give an example, Figure 25 shows TargetP results for plant protein GIMH01053214.1.p1.

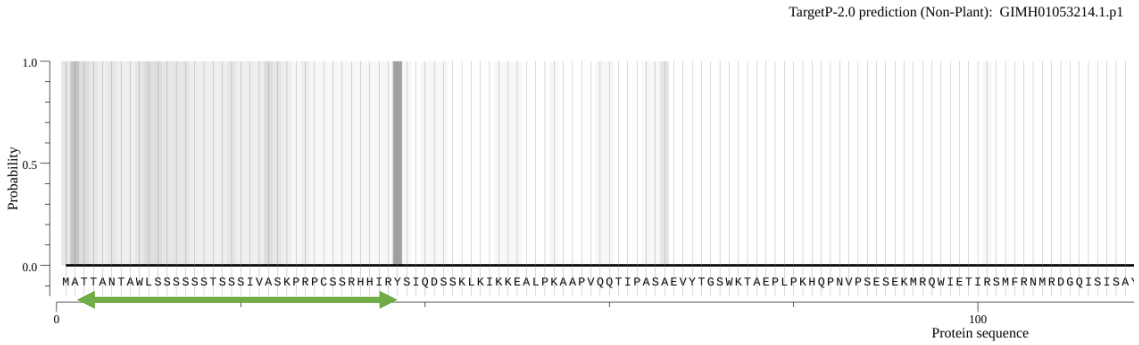

***Figure 25**. This figure shows a TargetP result table for a plant protein GIMH01053214.1.p. Increasing darkness of stripes above each amino acid in sequence indicates higher probability of presence of signal peptide. In this case, amino acids 2(A)-37(Y) were considered signal peptide and excluded from the sequence as shown by a green arrow.*

Second, Thermo Fischer Scientific GeneArt tool was used to convert protein sequence to nucleotide sequence and codon optimize the genes for expression in *Saccharomyces cerevisiae*. Additionally, the coding sequences of all genes were carefully examined for the presence of restriction sites that corresponded with the BsmBI and BsaI enzymes. It was imperative that these enzymes were not present in the protein coding regions since they were intended for use in Golden Gate cloning.

Before placing the order for the synthesis of gene fragments at Twist Bioscience, final adjustments were made, including the inclusion of restriction sites both upstream and downstream of the coding sequences for BsmBI and BsaI, as well as non-coding ends with `ACTAGACAAC` before and after the restriction sites. Due to the limitations imposed by the manufacturer, it was necessary to order six out of the nine genes as two-piece fragments. In total, 15 gene fragments were ordered for this study.

3.5 Designing expression vectors

A comprehensively characterized yeast toolkit, which facilitated the design and assembly of the desired cassette plasmid for ultimate integration with the candidate genes described in chapter 1.3.1.5 Modular cloning was utilized to design and construct expression vectors used in this thesis[117].

The objective of the experiment was to insert each candidate gene into a backbone created using the previously described MoClo-YT kit[117], resulting in nine distinct single gene expression constructs. To begin with, all the part plasmids required for backbone generation were selected (Table 2).

Assembly connectors *ConLS* and *ConRI* were selected to enable future multigene assembly if required. In the final assembly, a *GFP dropout* and its promoter and terminator, were replaced with different TPS candidate genes and their respective promoter and terminator. The yeast selection marker *URA3* was employed since both strains utilized for expression (JWY501 and ZX178-08) are uracil auxotrophs. Yeast origin *2micron* was chosen because of its higher copy numbers compared to the other yeast origins in the kit. Ampicillin resistance (*AmpR-ColE1*) was chosen as the bacterial selection marker.

Table 2. *List of used part plasmids and part types from MoClo YT kit for generating expression vector backbone pTP0027 (part plasmids with white background) and expression vectors pTP0195-pTP0203 carrying TPS genes (all part plasmids except pYTK067). Each part plasmid encodes for a bacterial antibiotic resistance marker to enable selection.*

Part plasmid name	Part type	Part type name	Bacterial marker
pYTK002	assembly connector	*ConLS*	chloramphenicol
pYTK047	coding sequence	*GFP dropout*	chloramphenicol
pYTK067	assembly connector	*ConR1*	chloramphenicol
pYTK074	yeast selection marker	*URA3*	chloramphenicol
pYTK082	yeast origin	*2micron*	chloramphenicol
pYTK083	ampicillin resistance	*AmpR-ColE1*	ampicillin
pYTK009	promoter	*pTDH3*	chloramphenicol
pYTK051	terminator	*tENO1*	chloramphenicol

The final expression vectors contain the *pTDH3* promoter, chosen for having the highest relative strength from all the promoters in the kit, and *tENO1* terminator, which displayed the highest relative expression levels from all the terminators in the kit. The backbone was designated as pTP0027, where "p" represents plasmid, "TP" represents the initials of the research group leader, and "27" indicates the order of the plasmid generated using the MoClo YT kit in the lab. The final expression plasmids, pTP0195-pTP0203, were named using the same system (Figure 26, page 49).

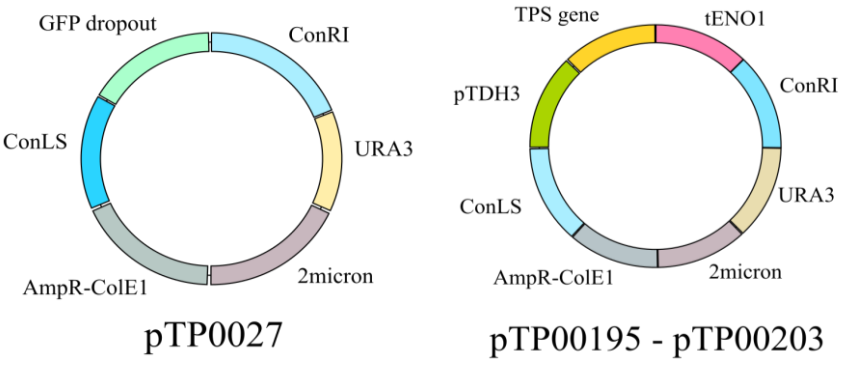

Figure 26. *General scheme of expression plasmid backbone named pTP0027 and final expression plasmids named pTP0195-pTP0203. Plasmid pTP0027 encodes a GFP dropout coding sequence with its own promoter and terminator, two assembly connectors ConLS and ConRI, yeast marker URA3, yeast origin 2micron and bacterial ampicillin resistance marker AmpR-ColEl. Expression vectors pTP0195-pTP0203 possess the same connectors, yeast origin, marker and bacterial resistance marker and coding sequence was replaced with 9 different TPS genes with pTDH3 promotors and tENO1 terminator resulting in 9 different vectors.*

3.6 Preparation of expression vectors

3.6.1 Obtaining part plasmid DNA

Plasmids were preserved as *Escherichia coli* strain DH10β glycerol stocks. Plasmids were amplified from 37°C, 160 RPM shaking in an Innova™ 40R incubator shaker for 24 h in sterilized 50mL Erlenmeyer flasks in 5 mL LB + either chloramphenicol or ampicillin. Plasmids were subsequently extracted using the QIAprep Spin Miniprep Kit to obtain plasmid DNA. A detailed protocol was provided with the kit and covered the following steps: A bacterial overnight culture of 5 mL was pelleted through centrifugation at 6,800 × g for 3 minutes at room temperature using a MiniSpin Plus benchtop centrifuge. Pellets were resuspended in 250 µL of P1 buffer and transferred to 2mL Eppendorf tubes. Next, 250 µL of P2 buffer was added, and the tubes were mixed 6 times by inversion. The lysis reaction was stopped after 5 minutes by adding 350 µL of N3 buffer, followed by mixing through inversion 6 times. The tubes were then centrifuged for 10 minutes at 17,900 × g using a 5427R Eppendorf centrifuge, and 800 µL of supernatant were transferred to a QIAprep 2.0 spin column. The samples were then centrifuged at 17,900 × g for 1 minute, and the flow-through was discarded. The columns were washed with 500 µL of PB buffer, centrifuged at 17,900 × g for 1 minute, and the flow-through was discarded. The columns were then washed with 750 µL of PE buffer, centrifuged at 17,900 × g for 1 minute, and the flow-through was discarded. All samples were then centrifuged for 1 minute at maximum speed (18,213 × g) to remove residual wash buffer. Finally, all columns were placed into clean 1.5mL Eppendorf tubes, and plasmid DNA was eluted using 50 µL of Mili-Q water and 1-minute centrifugation at maximum speed. The concentration of plasmid DNA in the samples was determined using a Nanodrop spectrophotometer.

3.6.2 Expression plasmid backbone preparation reaction

A Golden Gate Assembly Protocol utilizing the NEB Golden Gate Assembly Mix (E1600) developed by New England BioLabs was modified and refined by a collaborator, Mgr. Téo Hebra, Ph.D. The reaction involved six plasmids, namely, pYTK002, pYTK047, pYTK067, pYTK074, pYTK082, and pYTK083. The Golden Gate reaction for 2-5 or more fragments assembly using BsaI was the following: 0.5 µL of each plasmid at 40 nM, 1 µL T4 DNA ligase buffer, 0.5 µL T4 DNA ligase, 0.5 µL BsaI-HF®v2, and MiliQ water up to a final volume of 10 µL. The reaction consisted of 25 cycles at 37 °C for 2 min for cutting and 16 °C for 2 min for ligation. The Golden Gate reaction was concluded with a final cut at 60 °C for 30 min and enzyme denaturation at 80 °C for 10 min. For number of DNA fragments greater than 5, the cutting and ligation duration

were extended to 5 min. Assembled plasmids were either directly transformed or stored at -20°C for later use.

3.6.3 Ligation of gene fragments prior to final assembly

All ordered TPS genes were received in the form of linear fragments, and a total of six candidate genes required ligation. Moreover, all candidate genes necessitated ligation with an entry vector, namely pYTK001 (Figure 27, page 51), which is a storage plasmid and an entry vector for subsequent Golden Gate reactions. The Golden Gate reaction for 2-5 or more fragments assembly using BsmBI was the following: 0.5 µL of each plasmid at 40 nM, 1 µL of T4 DNA ligase buffer, 0.5 µL of T4 DNA ligase, 0.5 µL of BsmBI-v2, and MiliQ water up to a final volume of 10 µL. The reaction consisted of 25 cycles at 42 °C for 2 min for cutting and 16 °C for 2 min for ligation.

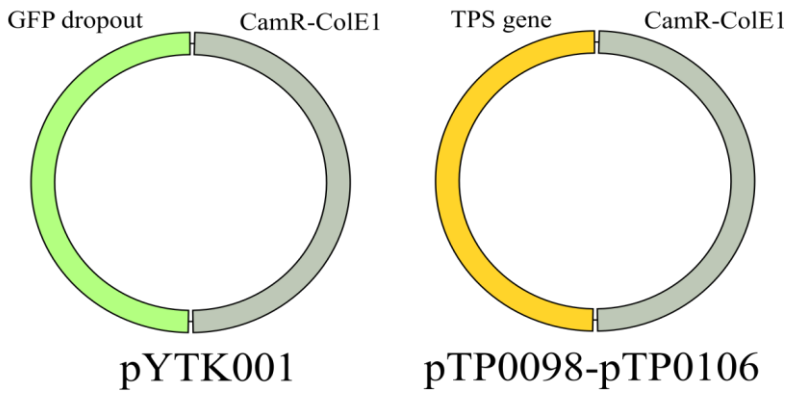

Figure 27. *General scheme of part plasmid pYTK001 containing a GFP dropout and chloramphenicol bacterial resistance marker CamR-ColE1 and storage plasmids named pTP0098-pTP0106 with a GFP dropout replaced with a TPS gene resulting in 9 different vectors.*

The Golden Gate reaction was concluded with a final cut at 60 °C for 30 min and enzyme denaturation at 80 °C for 10 min. For number of DNA fragments greater than 5, the cutting and ligation duration were extended to 5 min. The final products were plasmids pTP0098-pTP0106. Newly assembled plasmid may be transformed directly or preserved at a temperature of -20°C.

3.6.4 Multiplication of pTP0027 and pTP0098-pTP0106 in *E.coli*

3.6.4.1 *E.coli* transformation

The following equipment and chemicals were prepared in advance:

- 1 mL of preheated S.O.C. media for 37°C in 5mL falcon tubes
- precooled electroporation cuvettes on ice (one per plasmid)

The deep-frozen electrocompetent *E.coli* cells in eleven 1.5 mL Eppendorf tubes were placed on ice for a duration of 10 minutes. Subsequently, 2 µL of an assembled plasmid or 2 µL of Mili-Q water as negative control were added to each tube. Next, 45 µL of the resulting cell suspension was transferred to a cuvette. All cuvettes were then inserted into the Eppendorf Eporator® (1700 V, 2 ms pulse length) and 1 mL of S.O.C. media was added to each tube immediately after electroporation, followed by the suspension being thoroughly mixed. Cell suspensions were then transferred back to falcon tubes and placed in a shaking incubator (37°C, 160 RPM) for a duration of 1 hour. The resulting cell suspensions were then centrifuged for 3 minutes at 5,000 × g, and 850 µL of supernatant was discarded from each tube. The pellets were resuspended with the remaining supernatant, and the resulting suspensions were used to inoculate the appropriate selection plates. These plates were then incubated in a preheated 37°C incubator for at least 16 hours.

3.6.4.2 Bacterial colony PCR

LB agar plates supplemented with either chloramphenicol or ampicillin and PCR tubes containing 10 µL of sterile Milli-Q water were prepared in advance for the experiment. Two to four colonies were tested from each transformation plate. Colonies were picked using sterile pinewood toothpicks. Subsequently, a fresh LB plate with appropriate antibiotic was inoculated with the same toothpick with attached colony, labelled and the toothpick was then placed in PCR tube with Mili-Q water. Following this, the toothpick was discarded and the samples in the PCR tubes were heated to 99°C for 10 minutes in a thermocycler to lyse the cells and release DNA into the solution. The samples were then centrifuged at maximum speed on a benchtop minicentrifuge for 1 minute to remove cellular debris. Reaction volume per tube was 10 µL. Reaction mix in each tube contained following components: 5 µL of Phire Master Mix, 0.5 µL of 10µM forward primer, 0.5 µL of 10µM reverse primer and 4 µL of the DNA solution was used as a template. Table 3 illustrates used primers and required annealing temperatures for each pair as well as expected sizes of PCR products. Table 4 illustrates thermocycler setup for the reaction.

Table 3. *E. coli transformants carrying storage plasmids pTP0098-pTP0106 were subjected to colony PCR testing and primer pairs used for each plasmid along with their annealing temperatures and expected product sizes are listed below.*

Plasmid	Forward primer	Reverse primer	Annealing temperature [°C]	Product size [bp]
pTP0098	pug3	TH0033R	63	907
pTP0099	gfq4	TH0033R	60	752
pTP0100	gim4	TH0033R	61	748
pTP0101	gira4	TH0033R	61	720
pTP0102	H5_4	TH0033R	61	657
pTP0103	mapoly4	TH0033R	59	842
pTP0104	gff2	TH0033R	61	958
pTP0105	m2q1	TH0033R	61	725
pTP0106	NZ7_1	TH0033R	56	724

Table 4*. Thermocycling conditions for colony PCR with Phire Hot Start II DNA polymerase. Annealing temperature needs to be adjusted to used pair of primers. Moreover, the elongation time of the third step at 72°C is set according to PCR product size, which is 10 seconds per each 1000 base pairs.*

Phire Hot Start II DNA polymerase	Temperature [°C]	Time
Initial denaturation	98	2 min
30 cycles	98 Annealing temperature specific to primers 72	10 s 10 s 10 s per 1 kb
Final extension	72	2 min

Verification of the PCR product was performed using a 1% agarose gel in Tris-acetate-EDTA (TAE) buffer. Prior to loading the samples, the PCR tubes were centrifuged at maximum speed for 30s to eliminate any condensation that had accumulated on the sides and lid. Next, 9 µL of the PCR reaction were directly pipetted into the designated wells, as the loading dye was already present in the master mix. Electrophoresis was carried out at 120 V for a duration of 30 minutes. The visualization of the gels was accomplished using the Azure 600 imaging system set for ethidium

bromide visualization. A complete list of used primers is available in supplementary data (Supplementary table 1).

3.6.4.3 Growing transformed *E.coli* for plasmid extraction

Nine autoclaved 50mL Erlenmeyer flasks equipped with sterile foil caps were prepared. Subsequently, 5 mL of LB media containing the appropriate antibiotic was added to each flask, followed by inoculation with a colony that tested positive for insert presence. The bacterial cultures were grown overnight under conditions of shaking at 160 RPM and a temperature of 37°C. Thereafter, the cultures were subjected to plasmid extraction again using the QIAprep Spin Miniprep Kit, extraction protocol is shown in chapter 3.6.1 Obtaining part plasmid DNA.

3.6.5 Final assembly of expression plasmids

After acquiring all the necessary part plasmids, the final Golden Gate reaction was performed to obtain expression vectors pTP0195-pTP0203 (Figure 26, page 49). The thermocycler was set up to accommodate up to four fragments, as depicted in chapter 3.6.2 Expression plasmid backbone (pTP0027) preparation reaction. The restriction enzyme BsaI-HF®v2, which was kept at a temperature of -20°C, was added as the last reaction component. Newly assembled plasmids were either directly transformed or stored at -20°C for future use.

Expression constructs were subsequently transformed using the protocol provided in chapter 3.6.4.1 *E.coli* transformation. To confirm successful transformation, bacterial colony PCR was performed using Phire Hot Start II DNA polymerase as described in chapter 3.6.4.2 Bacterial colony PCR with modifications shown in Table 5, page 55. Finally, plasmid extraction was conducted again using the QIAprep Spin Miniprep Kit protocol specified in chapter 3.6.1 Obtaining part plasmid DNA.

Table 5. *Transformants carrying expression plasmids pTP0195-pTP0203 were subjected to colony PCR testing and primer pairs used for each plasmid along with their annealing temperatures and expected product sizes are listed below.*

Plasmid	Forward primer	Reverse primer	Annealing temperature [°C]	Product size [bp]
pTP0195	pug3	HS_R	66	1295
pTP0196	gfq4	HS_R	62	1140
pTP0197	gim4	HS_R	64	1136
pTP0198	gira4	HS_R	63	1108
pTP0199	H5_4	HS_R	63	1045
pTP0200	mapoly4	HS_R	61	1230
pTP0201	gff2	HS_R	60	1346
pTP0202	m2q1	HS_R	64	1113
pTP0203	NZ7_1	HS_R	57	1112

3.7 Yeast transformation

3.7.1 Strain JWY501

High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method[118] was utilized for the transformation of *S. cerevisiae* strain JWY501. A fresh YPD plate was inoculated with deep-frozen JWY501 glycerol stock and incubated in 30°C incubator for 3 days. A single colony was then used to inoculate 5 mL of YPD media in a sterilized and foil-capped 50mL Erlenmeyer flask, which was then placed in shaking incubator set to 200 RPM and 30°C. A bottle of 2× YPD and 250mL culture flask were let to preheat in 30°C incubator. The titter of the pre-production culture (OD_{600}) after 24 h of incubation was determined using spectrophotometer. A starting OD_{600} of production culture was 0.05 in 50 mL of 2× YPD in a pre-warmed culture flask, which was incubated in shaking incubator set to to 200 RPM and 30°C for 6 hours. This duration of incubation allowed for two rounds of cell division, resulting in a four-fold increase in cell numbers. A 1mL sample of carrier DNA was denatured in a boiling water bath for 5 minutes and was immediately chilled on ice. The cells were harvested by centrifugation at 3,000 × g for 5 minutes and subsequently resuspended in 25 mL of sterile Mili-Q water. After centrifugation at 3,000 × g for 5 minutes at 20°C, the supernatant was removed, and the cells were washed with fresh 25 mL

of sterile Mili-Q water. The cells were pelleted in the same manner as the prior wash and resuspended in 1 mL of sterile Mili-Q water. The resulting cell suspension was transferred to a 1.5mL Eppendorf tube, centrifuged for 30 s at 13,000 × g, and the supernatant was discarded. The cells were then resuspended in 1 mL of sterile Mili-Q water, and 100 µL of the resulting suspension was pipetted into 1.5mL Eppendorf tubes, with one tube used for each transformation. The samples were centrifuged at 13,000 × g for 30 s, and the supernatant was removed.

The transformation of a single plasmid was performed by preparing a transformation mix with following components: 240 µL of PEG 3355 (50%(w/v)), 36 µL of 1.0M LiAc, 50 µL of single-stranded carrier DNA (2.0 mg/mL), 5 µL of plasmid DNA (40nM) and 29 µL of sterile Mili-Q water.

Transformation mix (360 µL) was added to each transformation tube and cells were resuspended by vigorous vortexing. A negative control using sterile Mili-Q water in place of plasmid DNA, and a positive control using the pTP0027 backbone were included. All transformation tubes were placed in 42°C water bath and incubated for 1 h. Tubes were then centrifuged for 30 s at 13,000 × g and supernatant was carefully removed. The cells were resuspended in 1 mL of sterile Mili-Q water and vortexed. Subsequently, 20 µL of cell suspension were plated on YNB + D, U- plates (yeast nitrogen base solid media without uracil and with 2% (w/v) glucose) and incubated at 30°C for 4 days.

3.7.2 Strain ZX178-08

A protocol kindly provided by one of the ZX178-08 strain engineers, professor Joseph Chappel, Ph.D. was followed. A fresh YPDE plate was inoculated with deep-frozen ZX178-08 glycerol stock and incubated in 30°C incubator for 3 days. A single colony was then used to inoculate 3 mL of YPDE media in a sterilized 40mL falcon tube, which was then placed in shaking incubator set to 230 RPM and 30°C. The titter of the pre-production culture (OD_{600}) after 24 h of incubation was determined using a spectrophotometer. A starting OD_{600} of production culture was 0.1 in 50 mL of 2× YPDE in a pre-warmed culture flask, which was incubated in shaking incubator set to 230 RPM and 28°C for 6 hours. This duration of incubation allowed for two rounds of cell division, resulting in a four-fold increase in cell numbers. A 1mL sample carrier DNA was denatured in a boiling water bath for 5 minutes and chilled immediately on ice. The cells were transferred into a 50mL falcon tube and subjected to centrifugationat at 3,000 × g for 5 minutes for harvesting. Subsequently, the cells were resuspended in 25 mL of sorbitol wash buffer and centrifuged again at 3,000 × g for 5 minutes at 20°C. After removal of the supernatant, the cells were washed with 1 mL of 100 mM LiAc and centrifuged again in the same manner as the prior wash. Transformation buffer needed for the next step was prepared from following components: 2.4 mL of of PEG 3355 (50%(w/v), 690 µL of sterile

Mili-Q water, 360 µL of 1.0M LiAc and 100 µL of single-stranded carrier DNA (10.0 mg/mL). Following this, the cell suspension was treated with 3,550 µL of filter sterilized (0.2micron PTFE) transformation and was carefully mixed by swirling the tube.

Aliquots of 200 µL of transformation buffer with cells were added to 1.5mL Eppendorf tubes containing 5 µL of appropriate plasmid DNA (40 nM). The samples were mixed by pipetting up and down and were incubated at 42°C for an hour. Negative control with sterile Mili-Q water, instead of plasmid DNA, and positive control with pTP0027 backbone were included. Tubes were then centrifuged for 30 s at 13,000 × g and supernatant was carefully removed. The cells were resuspended in 1 mL of sterile Mili-Q water and vortexed. Following this, 20 µL of cell suspension were plated onto SCE plates and incubated at 30°C for 4 days.

3.8 Yeast colony PCR

YNB, D, U- (for JWY501 colonies) and SCE (for ZX178-08 colonies) plates and PCR tubes with 100 µL of lysing solution (200mM lithium acetate and 1%(w/v) SDS) were prepared in advance. Two to four colonies were tested from each transformation plate. Colonies were picked using sterile pinewood toothpicks. Subsequently, an appropriate selection plate was inoculated with the same toothpick with attached colony, labelled and the toothpick was then placed in PCR tube with lysing solution. Following this, the toothpick was discarded and the samples in the PCR tubes were placed in 70°C water bath for 10 minutes. Then, 300 µL of 96% (v/v) ethanol were added and samples were vortexed. All samples were centrifuged at 15,000 × g for 3 minutes and the supernatant was removed. The pellets were then washed with 500 µL of 70% (v/v) ethanol and centrifuged at 15,000 × g for 1 minute. The supernatant was removed and precipitated DNA was dissolved with 100 µL of sterile Mili-Q water. All tubes were subjected to centrifugation at maximum speed for a duration of 1 minute in a benchtop mini centrifuge to pellet cellular debris.

Yeast colony PCR was performed with Phire Hot Start II DNA polymerase in Phire Master Mix. Reaction volume per tube was 10 µL. Reaction mix in each tube contained following components: 5 µL of Phire Master Mix, 0.5 µL of 10µM forward primer, 0.5 µL of 10µM reverse primer, 3 µL of sterile Mili-Q water and 4 µL of the DNA solution used as a template. Table 5 on page 54 illustrates used primers and required annealing temperatures for each pair as well as expected sizes of PCR products. Table 4, page 52 illustrates thermocycler setup for the reaction.

Verification of the PCR product was performed using a 1% agarose gel in Tris-acetate-EDTA (TAE) buffer. Prior to loading the samples, the PCR tubes were centrifuged at maximum speed for 30s to eliminate any condensation that had accumulated on the sides and lid. Next, 9 µL of the PCR reaction were directly pipetted

into the designated wells, as the loading dye was already present in the master mix. Electrophoresis was carried out at 120 V for a duration of 30 minutes. The visualization of the gels was accomplished using the Azure 600 imaging system set for ethidium bromide visualization.

3.9 Yeast cultivation

In this experiment, a ZX178-08 colony was utilized as a host for one of plasmids pTP0195-pTP0203 or pTP0027, with the latter serving as a negative control. The inoculation was conducted in a sterilized 24-Well Deep Well Plate containing 2 mL of YNBE + D, U- media in each well. Similarly, a single JWY501 colony carrying one of plasmids pTP0195-pTP0203 or pTP0027 negative control, was used to inoculate 2 mL of YNB + D, U- media, in a separate 24-Well Deep Well Plate. The plates were placed in a shaking incubator, adjusted to 230 RPM and 30°C, for 24 hours, and the pre-production culture's optical density was measured using a spectrophotometer.

After the pre-production stage, the production cultures were initiated with a starting OD_{600} of 0.05 in 2 mL of YNBE + D, U- media for ZX178-08 and YNB + G + D, U- media for JWY501. The cultures were inoculated in sterilized Enzyscreen 24-Well Deep Well Plates and subjected to a ThermoMixer C shaking incubator at 800 RPM and 30°C for 48 hours.

3.10 Extraction prior to GC-MS and LC-MS analyses

Guaiazulene ethanolic solution (100µM) was used as internal standard and 1 µL was spiked into 1 mL of media with cells prior extraction. Ethyl acetate was added to spiked media in 1:1 ratio. Samples were then incubated in the thermomixer at 900 RPM for 1 hour and centrifuged at 15,000 × g for 10 minutes. Top organic layer was then collected and concentrated under nitrogen gas. Finally, extracts were resuspended with 100 µL ethyl acetate and transferred to a GC or LC vial.

3.11 GC-MS analysis

GC-MS analyses were performed using a 7890A gas chromatograph coupled with a 5975C mass spectrometer, equipped with electron ionization (EI) and quadrupole analyzer (Agilent Technologies, USA). The samples (1 µl) were injected to split/splitless inlet in split mode (split ratio 10:1). The injector temperature was 250 °C. A DB-1ms fused silica capillary column (30 m × 250 µm; a film thickness of 0.25 µm, J&W Scientific) was used for separation. The carrier gas was helium at a constant flow rate

of 1.0 ml/min. The temperature program was: 40 °C (1 min), then 5 °C per min to 100 °C, followed by 15 °C per min to 230°C. The temperatures of the transfer line, ion source and quadrupole were 320°C, 230°C and 150°C, respectively. EI spectra (70 eV) were recorded from 25 to 500 m/z.

3.12 LC-MS analysis

For LC-MS/MS analysis, samples were resuspended in 100 µL of ethyl acetate. LC-MS analyses were performed using Vanquish™ Flex UHPLC System interfaced to a Orbitrap ID-X Tribrid mass spectrometer, equipped with heated electrospray ionization (ESI). The LC conditions were as follows: column, Waters BEH (Ethylene Bridget Hybrid) C18 50 × 2.1 mm, 1.7 µm; mobile phase, (A) water with 0.1% (v/v) formic acid; (B) acetonitrile with 0.1% (v/v) formic acid; flow rate, 350 µL/min; column oven temperature, 40 °C, injection volume, 1 µL, linear gradient of 5 to 100 % B over 5 min and isocratic at 100% B for 2 min. Electrospray ionization was achieved in positive mode and mass spectrometer parameters were as follows: ion transfer tube temperature, 325 °C, auxiliary gas flow rate 10 L/min, vaporizer temperature 350 °C; sheath gas flow rate, 50 L/min; capillary voltage, 3000 V, MS resolution was 60 000, quadrupole isolation, scan range from m/z 100-1000, RF Lens 45 %, maximum injection time 118 ms. The data-dependent MS/MS events were acquired for the most intense ions from MS scan for a cycle time of 0.6 s, above a threshold of 1×10^5 intensity threshold, with a dynamic exclusion list of 2 s, including the isotopes. Selected precursor ions were fragmented with a fixed normalized HCD collision energy of 35 %, an isolation window of m/z 0.8 and a resolution of 15,000 with a maximum injection time of 80 ms.

3.13 Data analysis

LC-MS raw data files were directly imported into MZmine 3.3.0[119]. Extracted ion chromatograms for compounds of interest (masses corresponding to sesquiterpenes $C_{15}H_{24}$ and diterpenes $C_{20}H_{32}$) were generated using the raw data overview feature and exported as .pdf. GC-MS .dx files were analyzed using OpenLab CDS 2.4. Extracted ion chromatograms for compounds of interest were exported as .csv files and built using the ggplot2 package in R.

4. Results

4.1 Final list of protein candidates

The collaboration with Bc. Čalounová resulted in a list of the most promising candidates for heterologous expression (Table 6) according to her result sheet described in chapter 1.2.8.1 Homology-based approaches. The final list comprises 5 predicted diTPSs and 1 predicted sesquiTPS from plants, 1 predicted diTPS and 1 predicted sesquiTPS from bacteria and 1 predicted fungal sesquiTPS.

Table 6. *List of proteins subjected to heterologous expression in engineered S. cerevisiae. The candidates from A0A5A7PUH7 protein to Mapoly0046s0122.1.p1 protein were predicted as diTPS and proteins GFFR01069293.1.p1 to A0A7Y7XNZ7 as sesquiTPSs by Bc. Čalounová.*

Protein	Organism	Kingdom	Length [amino acids]	Domains present
A0A5A7PUG7	*Striga asiatica*	Plants	721	PF01397, PF03936
GFQG01057826.1.p1	*Amaranthus palmeri*	Plants	831	PF01397, PF03936
GIMH01053214.1.p1	*Selaginella denticulata*	Plants	839	PF01397, PF03936
GIRA01039209.1.p1	*Alternanthera philoxeroides*	Plants	822	PF01397, PF03936
A0A353ZAH5	*Planctomycetaceae bacterium*	Bacteria	779	PF13243
Mapoly0046s0122.1.p	*Marchantia polymorpha*	Plants	879	partial PF13243, PF01397, PF03936
GFFR01069293.1.p1	*Phlox drummondii*	Plants	571	PF01397, PF03936
M2QMG2	*Ceriporiopsis subvermispora*	Fungi	354	PF19086
A0A7Y7XNZ7	*Pseudomonas agarici*	Bacteria	355	PF19086

4.2 GC-MS analysis

A GC-MS analysis of JWY501 yeast extracts showed that 4 out of 9 protein candidates exhibit TPS activity and produce terpenoid products. Two proteins predicted as sesquiTPS, namely fungal M2QMG2 and plant GFFR01069293.1.p1, produce sesquiterpene products. Additionally, two plant proteins predicted as diTPS, namely Mapoly0046s0122.1.p and GIMH01053214.1.p1, produce diterpene products. Figure 28 on page 61 shows extracted ion chromatograms for mass corresponding to a common molecular formula of all sesquiterpenes ($C_{15}H_{24}$). Figure 29 on page 62 shows extracted ion chromatograms for mass corresponding to a common molecular formula of all diterpenes ($C_{20}H_{32}$). The samples "Empty-pESC" are negative controls (extract from yeast strain JWY501 cultures containing empty plasmid with no TPS gene). All peaks unique to each sample are coloured. MS spectra from peaks, with a high confidence spectral match with EI-NIST library are green. MS spectra from peaks with a low confidence spectral match with a NIST library are red. A peak has been highlighted in blue peak as it is the same peak in both $C_{15}H_{24}$ and $C_{20}H_{32}$ extracted ion chromatogram.

Figure 30 on page 63 shows two examples of high confidence matches with NIST spectral library. A high confidence match refers to a situation in which the fragmentation

spectra of a particular peak and its best-scoring spectral match are deemed highly similar in terms of both the fragments present and their respective intensities. Figure 31 on page 63 shows two examples of low confidence spectral matches of two red peaks' spectra from Figure 28 with NIST spectral library. In contrast, a low confidence match denotes a scenario in which the fragmentation spectra of a given peak, and its top-scoring spectral match are deemed dissimilar in terms of both the fragments present and their corresponding intensities, after undergoing visual confirmation.

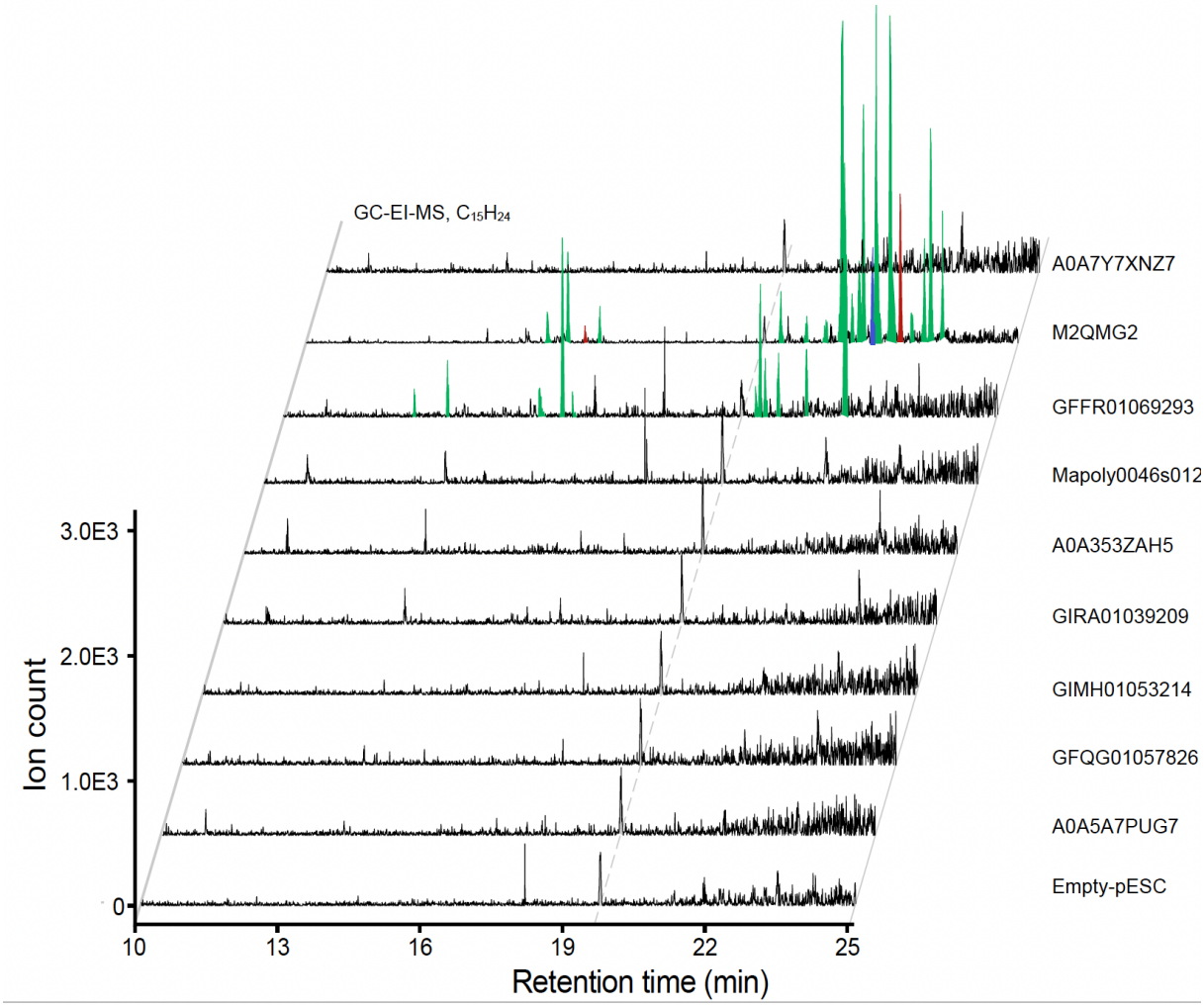

Figure 28. *Extracted ion chromatograms for mass corresponding $C_{15}H_{24}$. Sample "Empty-pESC" is a negative control. All peaks unique to each sample are coloured. Peaks with a high confidence spectral library match are green. Peaks with a low confidence spectral library match are red. A blue peak is thought to be responsible for an artefact blue in Figure 29.*

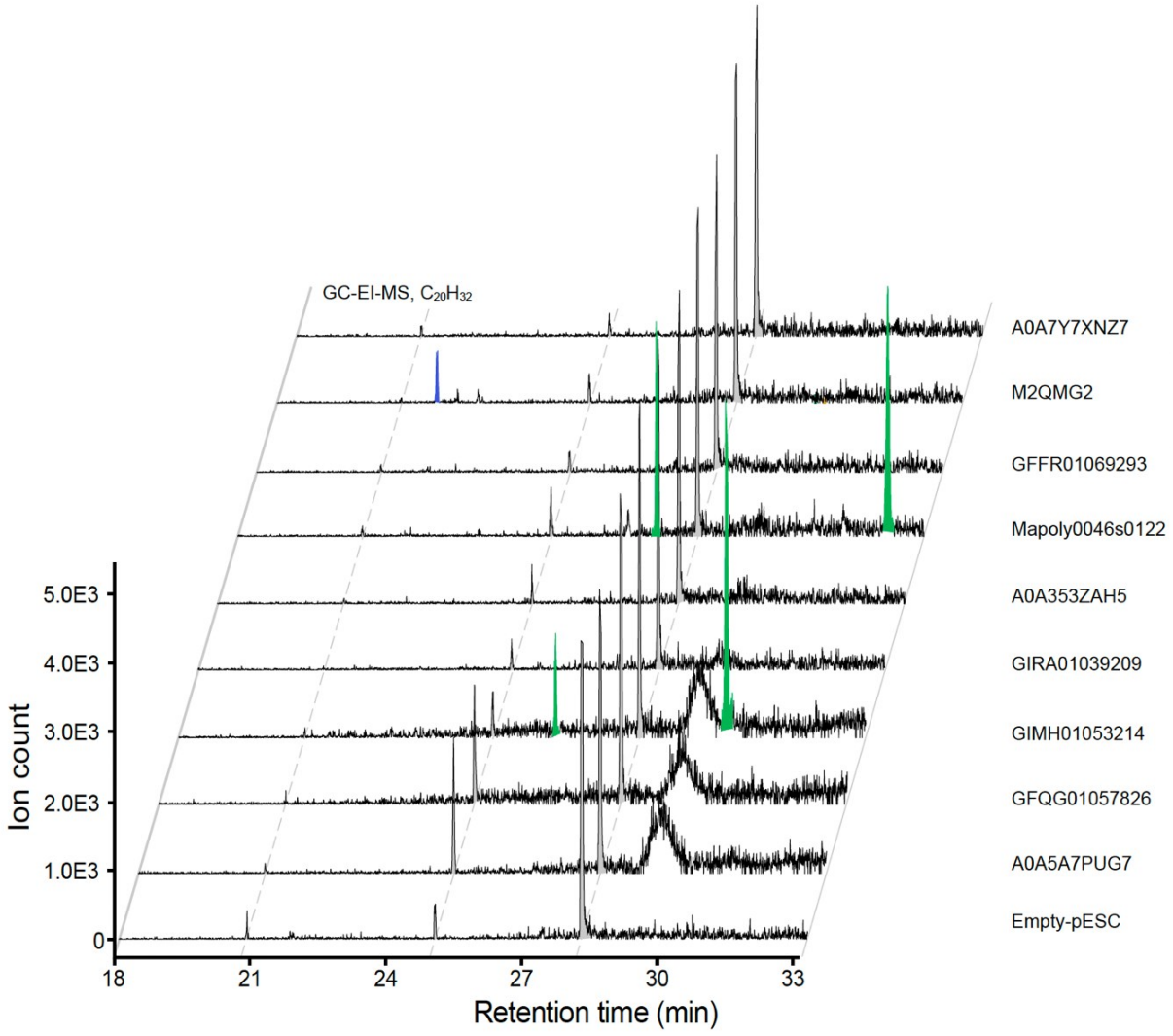

Figure 29 *Extracted ion chromatograms for mass corresponding $C_{20}H_{32}$. Sample "Empty-pESC" is a negative control. All peaks unique to each sample are coloured. Peaks with a high confidence spectral library match are green. Peaks with a low confidence spectral library match are red. A blue peak is thought to be artefact of a blue peak from figure 28.*

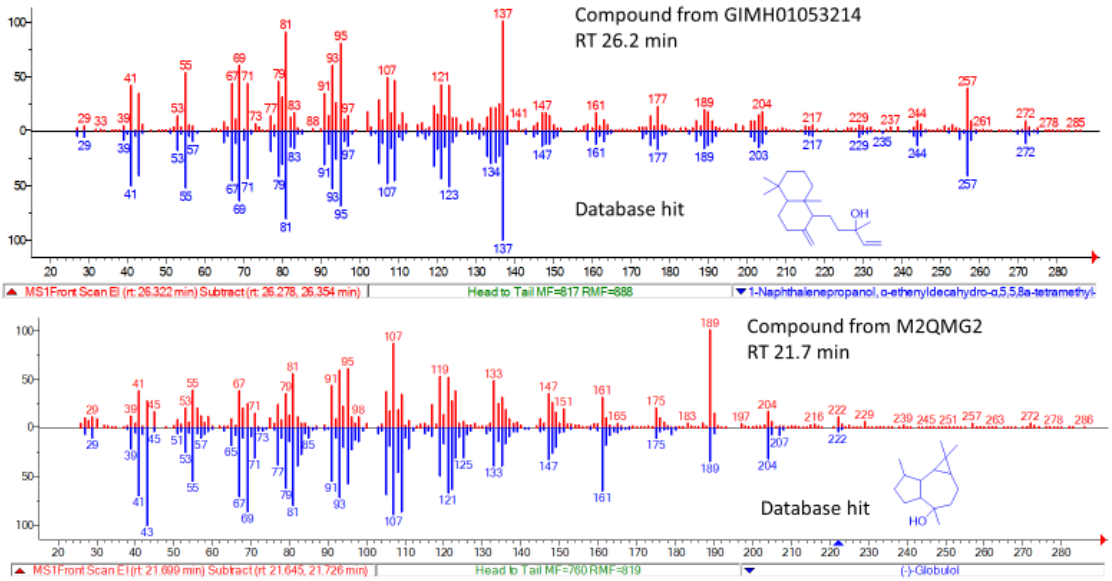

Figure 30. *Two examples of high confidence spectral matches of a "green" peak with retention time of 26.2 min from sample GIMH01053214 and another "green" peak with retention time 21.7 min from sample M2QMG2 with NIST spectral library. Both peaks' spectra exhibit majority of fragments and their intensities being the same as their spectral match.*

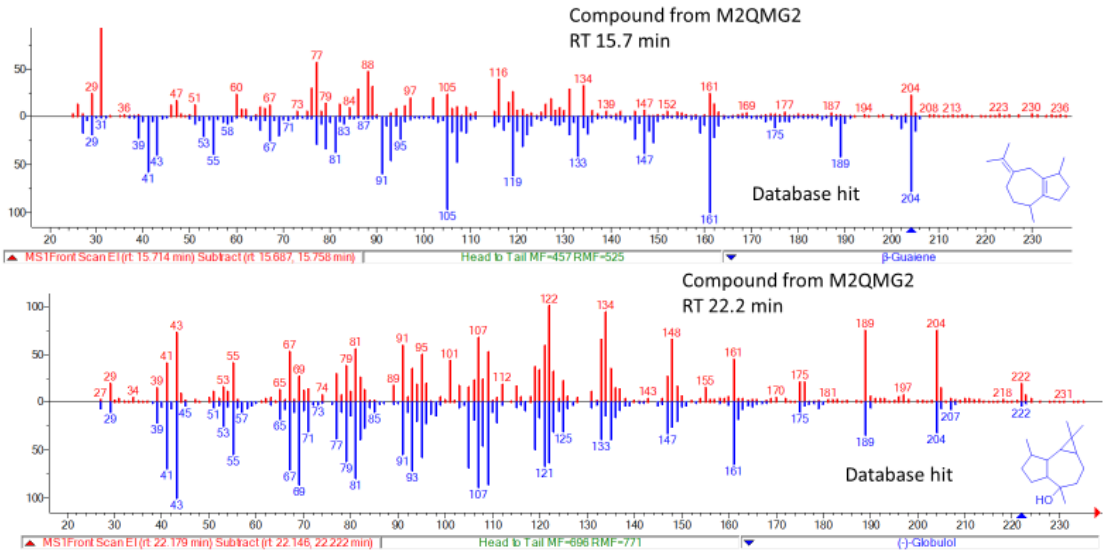

Figure 31. *Two examples of low confidence spectral matches of both "red" peaks sample M2QMG2 with retention times of 15.7 min and 22.2 min, respectively, with NIST spectral library. Both peaks' spectra exhibit majority of fragments and their intensities being the same as their spectral match.*

4.3 LC-MS analysis

An LC-MS analysis of JWY501 yeast extracts showed that 5 out of 9 protein candidates exhibit a TPS activity by producing terpenoid compounds. Three proteins predicted as sesquiTPS, namely fungal M2QMG2, plant GFFR01069293.1.p1 and bacterial A0A7Y7XNZ7, produce sesquiterpene products. Additionally, two plant proteins predicted as diTPS, namely Mapoly0046s0122.1.p and GIMH01053214.1.p1, produce diterpene products. Figure 32 shows extracted ion chromatograms for mass corresponding to a common molecular formula of all sesquiterpenes ($C_{15}H_{24}$). Figure 33 on page 65 shows extracted ion chromatograms for mass corresponding to a common molecular formula of all diterpenes ($C_{20}H_{32}$). The samples "Empty-pESC" are negative controls. All peaks unique to each sample are coloured. Peak intensities are normalized to the most intense peak.

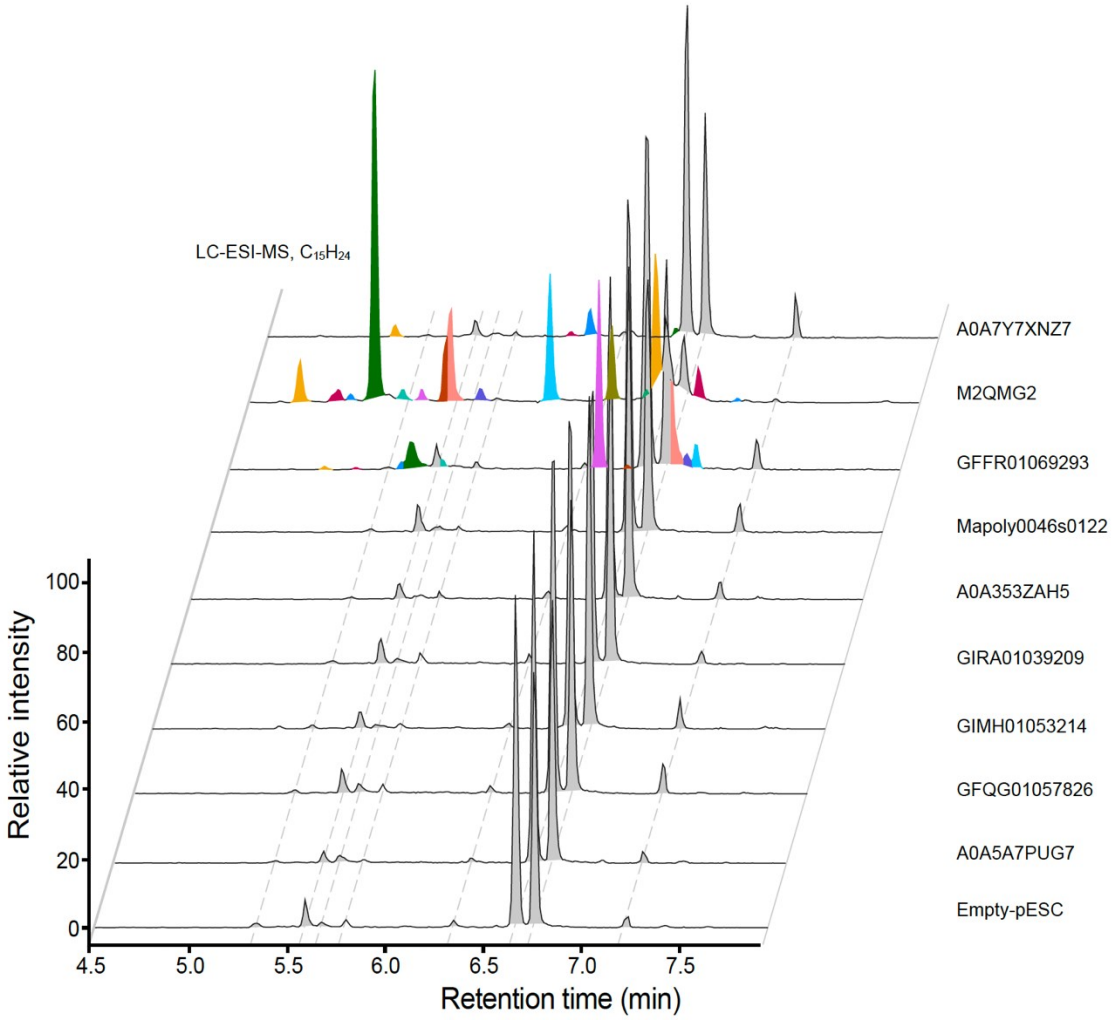

Figure 32. *Extracted chromatograms for mass corresponding $C_{15}H_{24}$. Sample "Empty-pESC" is a negative control. All peaks unique to each sample are coloured. Peak intensities are normalized to the most intense peak.*

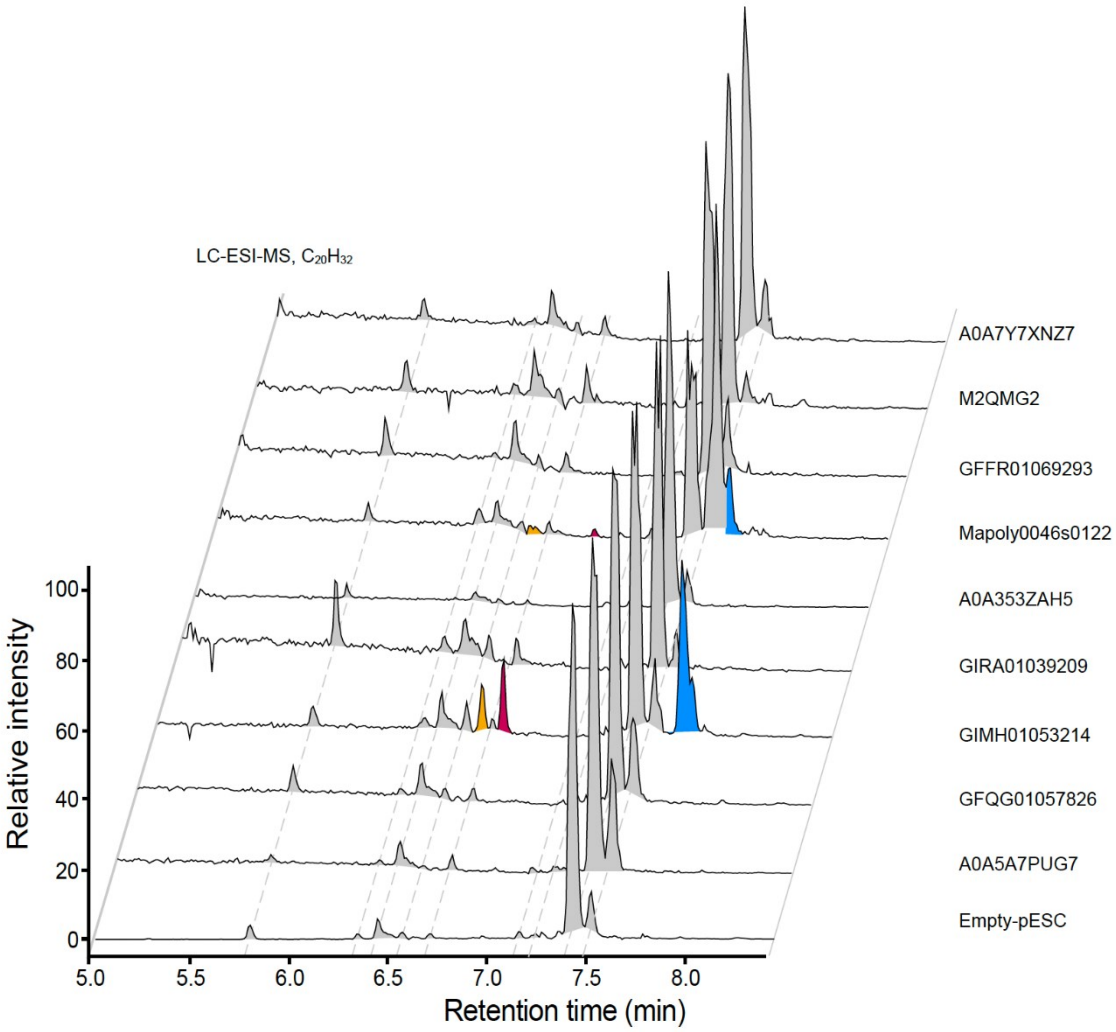

Figure 33. *Extracted chromatograms for mass corresponding $C_{20}H_{32}$. Sample "Empty-pESC" is a negative control. All peaks unique to each sample are coloured. Peak intensities are normalized to the most intense peak.*

5.Discussion

5.1 GC-MS and LC-MS analysis

GC-MS is generally considered more suitable for the detection of non-polar volatile compounds such as terpenes, albeit with lower sensitivity. On the other hand, LC-MS exhibits higher sensitivity yet is typically regarded as suitable for non-volatile and larger molecules making it less suitable for detecting terpenes.

In this thesis, GC-MS analysis was employed as a preliminary approach to detect any terpene products. Subsequently, to confirm the GC-MS results, LC-MS analysis was conducted. Additionally, due to higher sensitivity, LC-MS might detect low abundance compounds, not detectable in GC-MS. This approach showed to be useful since additional LC-MS analysis not only did confirm terpene products in all JWY501 samples that showed terpene products in GC-MS analysis, but also showed that protein A0A7Y7XNZ7 had TPS activity as well, by producing sesquiterpene products. Important note is that GC-MS and LC-MS fragmentation spectra are incomparable due to different ionization methods. Moreover, there is no spectral library including enough terpene for LC-MS fragmentation spectra such as a GC-EI-NIST library commonly used for GC-MS spectra, therefore compounds could not be identified using LC-ESI-MS only. LC-MS analysis was conducted purely for the purpose of confirmation of any terpene products in samples.

Interestingly, a geranylgeranyl-PP overproducing strain JWY501 showed to be a suitable host for both sesqui- and diTPSs even though sesquiTPSs primarily use farnesyl-PP as a substrate. Culturing JWY501 transformants yielded a wide range of terpene products in 5 different samples. However, cultures of ZX178-08 were lagging in growth behind JWY501 cultures, exhibited increased susceptibility to dying and provided spectra with a lot of noise due to the addition of ergosterol in the media. Therefore, only chromatograms of JWY501 samples are shown.

MS spectra of all peaks unique to each sample were visually examined for the presence of a molecular ion, which was necessary to be present as it proves compound's molecular formula as a sesqui- or diterpene. With regards to potential novelty among the detected terpene compounds, most of the GC-MS spectra demonstrated high spectral similarities to known compounds from the GC-EI-NIST library. This finding suggests that the likelihood of these compounds possessing a yet unreported structure is low (see Figure 30 on page 63). However, when the spectra of two red peaks from M2QMG2 sample (Figure 28 on page 61) were compared to the spectral database, no convincing spectral hits were obtained meaning that even the best scoring match resembled inconsistencies in number of shared fragments and their intensities. This suggests that these two peaks may potenially be novel compounds (see Figure 31 on page 63).

Nonetheless, it is important to perform future nuclear magnetic resonance (NMR) structure assessments to confirm this possibility.

The blue peak observed in Figure 29 on page 62 as a diterpene is suspected to be an artifact, and to be a noise signal from the spectra of a sesquiterpene (blue marked peak in Figure 28). This is supported by their same retention time, as well as the high similarity between their spectra and a sesquiterpene reference spectrum.

6. Conclusions

To conclude, 9 proteins predicted as either sesqui- or diterpene synthases originating from plants, bacteria and fungi were heterologously expressed and screened for terpene products. With that intention, 9 optimized vectors were prepared for heterologous expression of terpene synthases in yeast. To increase terpene production, 2 budding yeast strains engineered to overproduce sesqui- and diterpene precursors were used as hosts, namely ZX178-08 and JWY501, respectively. A GC-MS and LC-MS screening for sesqui- or diterpene products in culture extracts were performed and:

- 5 out of 9 proteins exhibited terpene synthase activity
- 2 compounds produced by a fungal enzyme M2QMG2 are likely to be novel

7. Bibliography

1. Zeng, T., Liu, Z., Liu, H., He, W., Tang, X., Xie, L., & Wu, R. (2019). Exploring Chemical and Biological Space of Terpenoids. *Journal of Chemical Information and Modeling*, *59*(9), 3667-3678. https://doi.org/10.1021/acs.jcim.9b00443
2. Perveen, S. (2018). Introductory Chapter: Terpenes and Terpenoids. *IntechOpen EBooks*. https://doi.org/10.5772/intechopen.79683
3. Bruce, S. (2022). Secondary Metabolites from Natural Products. *IntechOpen EBooks*. https://doi.org/10.5772/intechopen.102222
4. National Center for Biotechnology Information (2023). PubChem Compound Summary for CID 22311, Limonene. Retrieved April 18, 2023 from https://pubchem.ncbi.nlm.nih.gov/compound/Limonene.
5. National Center for Biotechnology Information (2023). PubChem Compound Summary for CID 440968, (-)-alpha-Pinene. Retrieved April 18, 2023 from https://pubchem.ncbi.nlm.nih.gov/compound/1S_-_-_-alpha-Pinene.
6. Fidyt, K., Fiedorowicz, A., Strządała, L., & Szumny, A. (2016). *β*-caryophyllene and*β*-caryophyllene oxide-natural compounds of anticancer and analgesic properties. *Cancer Medicine*, *5*(10), 3007–3017. https://doi.org/10.1002/cam4.816
7. Medeiros, R., Passos, G. F., Vitor, C. E., Koepp, J., Mazzuco, T. L., Pianowski, L. F., Campos, M. M., & Calixto, J. B. (2007). Effect of two active compounds obtained from the essential oil of *Cordia verbenacea* on the acute inflammatory

responses elicited by LPS in the rat paw. *British Journal of Pharmacology*, *151*(5), 618–627. https://doi.org/10.1038/sj.bjp.0707270

8. Klauke, A., Racz, I., Pradier, B., Markert, A., Zimmer, A., Gertsch, J., & Zimmer, A. (2014). The cannabinoid CB2 receptor-selective phytocannabinoid beta-caryophyllene exerts analgesic effects in mouse models of inflammatory and neuropathic pain. *European Neuropsychopharmacology*, *24*(4), 608–620. https://doi.org/10.1016/j.euroneuro.2013.10.008
9. Vanderah, D. J., Rutledge, N., Schmitz, F. J., & Ciereszko, L. S. (1978). Marine natural products: cembrene-A and cembrene-C from a soft coral, Nephthea species. *Journal of Organic Chemistry*, *43*(8), 1614–1616. https://doi.org/10.1021/jo00402a040
10. Zhang, C., Li, J., Su, J., Liang, Y., Yang, X., Zheng, K., & Zeng, L. (2006). Cytotoxic Diterpenoids from the Soft Coral *Sarcophyton crassocaule*. *Journal of Natural Products*, *69*(10), 1476–1480. https://doi.org/10.1021/np050499g
11. Peng, C., Huang, C., Sheu, J., Hwang, T., & Sheu, J. (2020). Anti-Inflammatory Cembranoids from a Formosa Soft Coral Sarcophyton cherbonnieri. *Marine Drugs*, *18*(11), 573. https://doi.org/10.3390/md18110573
12. Rowland, S. K., Allard, W., Belt, S. T., Massé, G., Robert, J., Blackburn, S., Frampton, D. M. F., Revill, A. T., & Volkman, J. K. (2001). Factors influencing the distributions of polyunsaturated terpenoids in the diatom, Rhizosolenia setigera. *Phytochemistry*, *58*(5), 717–728. https://doi.org/10.1016/s0031-9422(01)00318-1
13. Kim, S., & Karadeniz, F. (2012). Biological Importance and Applications of Squalene and Squalane. *Elsevier eBooks*, 223–233. https://doi.org/10.1016/b978-0-12-416003-3.00014-7
14. Mohan, T., Verma, P., & Rao, D. N. (2013). Novel adjuvants & delivery vehicles for vaccines development: a road ahead. *The Indian journal of medical research*, *138*(5), 779–795.
15. Ramírez-Torres, A., Gabás, C., & Barranquero, C. (2011). *Squalene: Current Knowledge and Potential Therapeutical Uses*. Nova Science Publishers.
16. Weston-Green, K., Clunas, H., & Naranjo, C. J. (2021). A Review of the Potential Use of Pinene and Linalool as Terpene-Based Medicines for Brain Health: Discovering Novel Therapeutics in the Flavours and Fragrances of Cannabis. *Frontiers in Psychiatry*, *12*. https://doi.org/10.3389/fpsyt.2021.583211
17. Bromma, B., Scharein, E., Darsow, U., & Ring, J. (1995). Effects of menthol and cold on histamine-induced itch and skin reactions in man. *Neuroscience Letters*, *187*(3), 157–160. https://doi.org/10.1016/0304-3940(95)11362-z

18. Hou, L., & Huang, H. (2016). Immune suppressive properties of artemisinin family drugs. *Pharmacology & Therapeutics, 166*, 123–127. https://doi.org/10.1016/j.pharmthera.2016.07.002
19. Roh, S., Park, M. S., & Kim, Y. (2010). Abietic Acid from Resina Pini of Pinus Species as a Testosterone 5.ALPHA.-Reductase Inhibitor. *Journal of Health Science*, *56*(4), 451–455. https://doi.org/10.1248/jhs.56.451
20. Bohlmann, J., & Keeling, C. I. (2008). Terpenoid biomaterials. *Plant Journal*, *54*(4), 656–669. https://doi.org/10.1111/j.1365-313x.2008.03449.x
21. National Toxicology Program (2011). Toxicology and carcinogenesis studies of alpha,beta-thujone (CAS No. 76231-76-0) in F344/N rats and B6C3F1 mice (gavage studies). *National Toxicology Program technical report series*, (570), 1–260.
22. Degenhardt, J., Köllner, T. G., & Gershenzon, J. (2009). Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. *Phytochemistry*, *70*(15–16), 1621–1637. https://doi.org/10.1016/j.phytochem.2009.07.030
23. Ludwiczuk, A., Skalicka-Woźniak, K., & Georgiev, M. I. (2017). Terpenoids. *Elsevier eBooks*, 233–266. https://doi.org/10.1016/b978-0-12-802104-0.00011-1
24. Kuzuyama, T. (2002). Mevalonate and Nonmevalonate Pathways for the Biosynthesis of Isoprene Units. *Bioscience, Biotechnology, and Biochemistry*, *66*(8), 1619–1627. https://doi.org/10.1271/bbb.66.1619
25. Boucher, Y., Kamekura, M., & Doolittle, W. F. (2004). Origins and evolution of isoprenoid lipid biosynthesis in archaea. *Molecular Microbiology*, *52*(2), 515–527. https://doi.org/10.1111/j.1365-2958.2004.03992.x
26. Lange, B. M., Rujan, T., Martin, W., & Croteau, R. (2000). Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(24), 13172–13177. https://doi.org/10.1073/pnas.240454797
27. Bochar, D. A., Stauffacher, C. V., & Rodwell, V. W. (1999). Sequence Comparisons Reveal Two Classes of 3-Hydroxy-3-methylglutaryl Coenzyme A Reductase. *Molecular Genetics and Metabolism*, *66*(2), 122–127. https://doi.org/10.1006/mgme.1998.2786
28. Romanowski, M. H., Bonanno, J. B., & Burley, S. K. (2002). Crystal structure of theStreptococcus pneumoniae phosphomevalonate kinase, a member of the GHMP kinase superfamily. *Proteins*, *47*(4), 568–571. https://doi.org/10.1002/prot.10118
29. Voynova, N. E., RíOs, S. E., & Miziorko, H. M. (2004). *Staphylococcus aureus* Mevalonate Kinase: Isolation and Characterization of an Enzyme of the

Isoprenoid Biosynthetic Pathway. *Journal of Bacteriology*, *186*(1), 61–67. https://doi.org/10.1128/jb.186.1.61-67.2004

30. Boucher, Y., & Doolittle, W. F. (2000). The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Molecular Microbiology*, *37*(4), 703–716. https://doi.org/10.1046/j.1365-2958.2000.02004.x
31. Zhao, L., Chang, W., Xiao, Y., Liu, H., & Liu, P. (2013). Methylerythritol Phosphate Pathway of Isoprenoid Biosynthesis. *Annual Review of Biochemistry*, *82*(1), 497–530. https://doi.org/10.1146/annurev-biochem-052010-100934
32. Buhaescu, I., & Izzedine, H. (2007). Mevalonate pathway: A review of clinical and therapeutical implications. *Clinical Biochemistry*, *40*(9–10), 575–584. https://doi.org/10.1016/j.clinbiochem.2007.03.016
33. Goldstein, J. L., & Brown, M. E. (1990). Regulation of the mevalonate pathway. *Nature*, *343*(6257), 425–430. https://doi.org/10.1038/343425a0
34. Hinson, D. D., Chambliss, K. L., Toth, M. J., Tanaka, R. D., & Gibson, K. (1997). Post-translational regulation of mevalonate kinase by intermediates of the cholesterol and nonsterol isoprene biosynthetic pathways. *Journal of Lipid Research*, *38*(11), 2216–2223. https://doi.org/10.1016/s0022-2275(20)34935-x
35. https://en.wikipedia.org/wiki/Mevalonate_pathway
36. Bloch, K., and Lynen, F. (1952). The site of synthesis of cholesterol in the animal body. Proc. Natl. Acad. Sci. U.S.A. 38, 415–423
37. Rohmer, M. (1999). The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants†. *Natural Product Reports*, *16*(5), 565–574. https://doi.org/10.1039/a709175c
38. Rohmer, M. (2007). Diversity in isoprene unit biosynthesis: The methylerythritol phosphate pathway in bacteria and plastids. *Pure and Applied Chemistry*, *79*(4), 739–751. https://doi.org/10.1351/pac200779040739
39. Zhao, L., Chang, W., Xiao, Y., Liu, H., & Liu, P. (2013b). Methylerythritol Phosphate Pathway of Isoprenoid Biosynthesis. *Annual Review of Biochemistry*, *82*(1), 497–530. https://doi.org/10.1146/annurev-biochem-052010-100934
40. Vandermoten, S., Haubruge, E., & Cusson, M. (2009). New insights into short-chain prenyltransferases: structural features, evolutionary history and potential for selective inhibition. *Cellular and Molecular Life Sciences*, *66*(23), 3685–3695. https://doi.org/10.1007/s00018-009-0100-9
41. Zhang, Y., & Li, Z. (2012). FUNCTIONAL ANALYSIS AND MOLECULAR DOCKING IDENTIFY TWO ACTIVE SHORT-CHAIN PRENYLTRANSFERASES IN THE GREEN PEACH APHID, Myzus persicae.

Archives of Insect Biochemistry and Physiology, *81*(2), 63–76. https://doi.org/10.1002/arch.21032

42. Walsh, C. T. (2007). Revealing Coupling Patterns in Isoprenoid Alkylation Biocatalysis. *ACS Chemical Biology*, *2*(5), 296–298. https://doi.org/10.1021/cb700094s
43. Christianson, D. W. (2017). Structural and Chemical Biology of Terpenoid Cyclases. *Chemical Reviews*, *117*(17), 11570–11648. https://doi.org/10.1021/acs.chemrev.7b00287
44. Heldt, H., & Piechulla, B. (2021). A Large Diversity of Isoprenoids Has Multiple Functions in Plants. *Elsevier eBooks*, 391–410. https://doi.org/10.1016/b978-0-12-818631-2.00017-9
45. Poulter, C. D., & Rilling, H. C. (1978b). The prenyl transfer reaction. Enzymic and mechanistic studies of the 1'-4 coupling reaction in the terpene biosynthetic pathway. *Accounts of Chemical Research*, *11*(8), 307–313. https://doi.org/10.1021/ar50128a004
46. Kellogg, B., & Poulter, C. D. (1997b). Chain elongation in the isoprenoid biosynthetic pathway. *Current Opinion in Chemical Biology*, *1*(4), 570–578. https://doi.org/10.1016/s1367-5931(97)80054-3
47. Croteau, R. (1987). Biosynthesis and catabolism of monoterpenoids. *Chemical Reviews*, *87*(5), 929–954. https://doi.org/10.1021/cr00081a004
48. Cane, D. E. (1990). Enzymic formation of sesquiterpenes. *Chemical Reviews*, *90*(7), 1089–1103. https://doi.org/10.1021/cr00105a002
49. Davis, E. P., & Croteau, R. (2000). Cyclization Enzymes in the Biosynthesis of Monoterpenes, Sesquiterpenes, and Diterpenes. *Topics in Current Chemistry*, 53–95. https://doi.org/10.1007/3-540-48146-x_2
50. Wendt, K. U., Schulz, G. E., Corey, E. J., & Liu, D. R. (2000). Enzyme Mechanisms for Polycyclic Triterpene Formation. *Angewandte Chemie*, *39*(16), 2812–2833. https://doi.org/10.1002/1521-3773(20000818)39:16
51. Jiang, S., Jin, J., Sarojam, R., & Ramachandran, S. (2019). A Comprehensive Survey on the Terpene Synthase Gene Family Provides New Insight into Its Evolutionary Patterns. *Genome Biology and Evolution*, 11(8), 2078–2098. https://doi.org/10.1093/gbe/evz142
52. Chen, F., Tholl, D., Bohlmann, J., & Pichersky, E. (2011). The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant Journal*, *66*(1), 212–229. https://doi.org/10.1111/j.1365-313x.2011.04520.x
53. Whittington, D. A., Wise, M. L., Urbansky, M., Coates, R. M., Croteau, R., & Christianson, D. W. (2002). Bornyl diphosphate synthase: Structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proceedings of the*

National Academy of Sciences of the United States of America, *99*(24), 15375–15380. https://doi.org/10.1073/pnas.232591099

54. Chen, M., Al-Lami, N., Janvier, M., D'Antonio, E. L., Faraldos, J. A., Cane, D. E., Allemann, R. K., & Christianson, D. W. (2013). Mechanistic Insights from the Binding of Substrate and Carbocation Intermediate Analogues to Aristolochene Synthase. *Biochemistry*, *52*(32), 5441–5453. https://doi.org/10.1021/bi400691v
55. Tantillo, D. J. (2010). The carbocation continuum in terpene biosynthesis—where are the secondary cations? *Chemical Society Reviews*, *39*(8), 2847. https://doi.org/10.1039/b917107j
56. Dougherty, D. A. (1996). Cation-π Interactions in Chemistry and Biology: A New View of Benzene, Phe, Tyr, and Trp. *Science*, *271*(5246), 163–168. https://doi.org/10.1126/science.271.5246.163
57. C, J., MA, & Dougherty, D. A. (1997). The Cation−π Interaction. *Chemical Reviews*, *97*(5), 1303–1324. https://doi.org/10.1021/cr9603744
58. Jenson, C., & Jorgensen, W. L. (1997). Computational Investigations of Carbenium Ion Reactions Relevant to Sterol Biosynthesis. *Journal of the American Chemical Society*, *119*(44), 10846–10854. https://doi.org/10.1021/ja9714245
59. Wendt, K. U., & Schulz, G. E. (1998). Isoprenoid biosynthesis: manifold chemistry catalyzed by similar enzymes. *Structure*, *6*(2), 127–133. https://doi.org/10.1016/s0969-2126(98)00015-x
60. Lesburg, C. A., Zhai, G., Cane, D. E., & Christianson, D. W. (1997). Crystal Structure of Pentalenene Synthase: Mechanistic Insights on Terpenoid Cyclization Reactions in Biology. *Science*, *277*(5333), 1820–1824. https://doi.org/10.1126/science.277.5333.1820
61. Starks, C. M., Back, K., Chappell, J., & Noel, J. P. (1997). Structural Basis for Cyclic Terpene Biosynthesis by Tobacco 5-Epi-Aristolochene Synthase. *Science*, *277*(5333), 1815–1820. https://doi.org/10.1126/science.277.5333.1815
62. Wendt, K. U., Poralla, K., & Schulz, G. E. (1997). Structure and Function of a Squalene Cyclase. *Science*, *277*(5333), 1811–1815. https://doi.org/10.1126/science.277.5333.1811
63. Köksal, M., Jin, Y., Coates, R. M., Croteau, R., & Christianson, D. W. (2011). Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature*, *469*(7328), 116–120. https://doi.org/10.1038/nature09628
64. Zhou, K., Gao, Y., Hoy, J. A., Mann, F. O., Honzatko, R. B., & Peters, R. (2012). Insights into Diterpene Cyclization from Structure of Bifunctional Abietadiene Synthase from Abies grandis. *Journal of Biological Chemistry*, *287*(9), 6840–6850. https://doi.org/10.1074/jbc.m111.337592

65. Wise, M. L., Savage, T., Katahira, E., & Croteau, R. (1998). Monoterpene Synthases from Common Sage (Salvia officinalis). *Journal of Biological Chemistry*, *273*(24), 14891–14899. https://doi.org/10.1074/jbc.273.24.14891
66. Whittington, D. A., Wise, M. L., Urbansky, M., Coates, R. M., Croteau, R., & Christianson, D. W. (2002b). Bornyl diphosphate synthase: Structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(24), 15375–15380. https://doi.org/10.1073/pnas.232591099
67. Köksal, M., Zimmer, I., Schnitzler, J., & Christianson, D. W. (2010). Structure of Isoprene Synthase Illuminates the Chemical Mechanism of Teragram Atmospheric Carbon Emission. *Journal of Molecular Biology*, *402*(2), 363–373. https://doi.org/10.1016/j.jmb.2010.07.009
68. Cane, D. E., Yang, G., Xue, Q., & Shim, J. (1995). Trichodiene Synthase. Substrate Specificity and Inhibition. *Biochemistry*, *34*(8), 2471–2479. https://doi.org/10.1021/bi00008a010
69. Li, J. C., Fang, X., Zhou, E., Ruan, J., Yang, C., Wang, L., Miller, D., Faraldos, J. A., Allemann, R. K., Chen, X. R., & Zhang, P. (2013). Rational engineering of plasticity residues of sesquiterpene synthases from *Artemisia annua*: product specificity and catalytic efficiency. *Biochemical Journal*, *451*(3), 417–426. https://doi.org/10.1042/bj20130041
70. McAndrew, R., Peralta-Yahya, P., DeGiovanni, A., Pereira, J. H., Hadi, M. Z., Keasling, J. D., & Adams, P. D. (2011). Structure of a Three-Domain Sesquiterpene Synthase: A Prospective Target for Advanced Biofuels Production. *Structure*, *19*(12), 1876–1884. https://doi.org/10.1016/j.str.2011.09.013
71. Hezari, M., Lewis, N. K., & Croteau, R. (1995). Purification and Characterization of Taxa-4(5),11(12)-diene Synthase from Pacific Yew (Taxus Brevifolia) that Catalyzes the First Committed Step of Taxol Biosynthesis. *Archives of Biochemistry and Biophysics*, *322*(2), 437–444. https://doi.org/10.1006/abbi.1995.1486
72. Wildung, M. R., & Croteau, R. (1996). A cDNA Clone for Taxadiene Synthase, the Diterpene Cyclase That Catalyzes the Committed Step of Taxol Biosynthesis. *Journal of Biological Chemistry*, *271*(16), 9201–9204. https://doi.org/10.1074/jbc.271.16.9201
73. Williams, D. R., Wildung, M. R., Jin, A. Q., Dalal, D., Oliver, J. P., Coates, R. M., & Croteau, R. (2000). Heterologous Expression and Characterization of a "Pseudomature" Form of Taxadiene Synthase Involved in Paclitaxel (Taxol) Biosynthesis and Evaluation of a Potential Intermediate and Inhibitors of the

Multistep Diterpene Cyclization Reaction. *Archives of Biochemistry and Biophysics*, *379*(1), 137–146. https://doi.org/10.1006/abbi.2000.1865

74. Lin, X., Hezari, M., Koepp, A. E., Floss, H. G., & Croteau, R. (1996). Mechanism of Taxadiene Synthase, a Diterpene Cyclase That Catalyzes the First Step of Taxol Biosynthesis in Pacific Yew. *Biochemistry*, *35*(9), 2968–2977. https://doi.org/10.1021/bi9526239
75. Williams, D. R., Carroll, B., Jin, Q., Rithner, C. D., Lenger, S. R., Floss, H. G., Coates, R. M., Williams, R. W., & Croteau, R. (2000). Intramolecular proton transfer in the cyclization of geranylgeranyl diphosphate to the taxadiene precursor of taxol catalyzed by recombinant taxadiene synthase. *Chemistry & Biology*, *7*(12), 969–977. https://doi.org/10.1016/s1074-5521(00)00046-6
76. Ajikumar, P. K., Xiao, W., Tyo, K. E. J., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T. H., Pfeifer, B. A., & Stephanopoulos, G. (2010b). Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in *Escherichia coli*. *Science*, *330*(6000), 70–74. https://doi.org/10.1126/science.1191652
77. Engels, B., Dahm, P., & Jennewein, S. (2008). Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production. *Metabolic Engineering*, *10*(3–4), 201–206. https://doi.org/10.1016/j.ymben.2008.03.001
78. Edgar, S., Li, F., Qiao, K., Weng, J., & Stephanopoulos, G. (2017). Engineering of Taxadiene Synthase for Improved Selectivity and Yield of a Key Taxol Biosynthetic Intermediate. *ACS Synthetic Biology*, *6*(2), 201–205. https://doi.org/10.1021/acssynbio.6b00206
79. Köksal, M., Hu, H., Coates, R. M., Peters, R., & Christianson, D. W. (2011). Structure and mechanism of the diterpene cyclase ent-copalyl diphosphate synthase. *Nature Chemical Biology*, *7*(7), 431–433. https://doi.org/10.1038/nchembio.578
80. Peters, R. (2010). Two rings in them all: The labdane-related diterpenoids. *Natural Product Reports*, *27*(11), 1521. https://doi.org/10.1039/c0np00019a
81. Fleet, C. M., & Sun, T. (2005). A DELLAcate balance: the role of gibberellin in plant morphogenesis. *Current Opinion in Plant Biology*, *8*(1), 77–85. https://doi.org/10.1016/j.pbi.2004.11.015
82. Potter, K. D., Zi, J., Hong, Y. J., Schulte, S., Malchow, B., Tantillo, D. J., & Peters, R. (2016). Blocking Deprotonation with Retention of Aromaticity in a Plant *ent* -Copalyl Diphosphate Synthase Leads to Product Rearrangement. *Angewandte Chemie*, *55*(2), 634–638. https://doi.org/10.1002/anie.201509060
83. Rudolf, J. D., Dong, L., Cao, H., Hatzos-Skintges, C., Osipiuk, J., Endres, M., Chang, C., Ma, M., Babnigg, G., Joachimiak, A., Phillips, G. N., & Shen, B.

(2016). Structure of the *ent*-Copalyl Diphosphate Synthase PtmT2 from *Streptomyces platensis* CB00739, a Bacterial Type II Diterpene Synthase. *Journal of the American Chemical Society*, *138*(34), 10905–10915. https://doi.org/10.1021/jacs.6b04317

84. Bauler, P., Huber, G. L., Leyh, T. S., & McCammon, J. A. (2010). Channeling by Proximity: The Catalytic Advantages of Active Site Colocalization Using Brownian Dynamics. *Journal of Physical Chemistry Letters*, *1*(9), 1332–1335. https://doi.org/10.1021/jz1002007
85. Castellana, M., Wilson, M. Z., Xu, Y., Joshi, P. G., Cristea, I. M., Rabinowitz, J. D., Gitai, Z., & Wingreen, N. S. (2014). Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nature Biotechnology*, *32*(10), 1011–1018. https://doi.org/10.1038/nbt.3018
86. Ballio, A., Chain, E. B., De Leo, P., Erlanger, B. F., Mauri, M., & Tonolo, A. (1964). Fusicoccin: a New Wilting Toxin produced by Fusicoccum amygdali Del. *Nature*, *203*(4942), 297. https://doi.org/10.1038/203297a0
87. De Vries-Van Leeuwen, I. J., Kortekaas-Thijssen, C., Mandouckou, J. a. N., Kas, S., Evidente, A., & De Boer, A. G. (2010). Fusicoccin-A selectively induces apoptosis in tumor cells after interferon-α priming. *Cancer Letters*, *293*(2), 198–206. https://doi.org/10.1016/j.canlet.2010.01.009
88. Toyomasu, T., Tsukahara, M., Kaneko, A., Niida, R., Mitsuhashi, W., Dairi, T., Kato, N., & Sassa, T. (2007). Fusicoccins are biosynthesized by an unusual chimera diterpene synthase in fungi. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(9), 3084–3088. https://doi.org/10.1073/pnas.0608426104
89. Chen, M., Chou, W. K. W., Toyomasu, T., Cane, D. E., & Christianson, D. W. (2016). Structure and Function of Fusicoccadiene Synthase, a Hexameric Bifunctional Diterpene Synthase. *ACS Chemical Biology*, *11*(4), 889–899. https://doi.org/10.1021/acschembio.5b00960
90. Kavanagh, K. L., Dunford, J. V., Bunkoczi, G., Russell, R., & Oppermann, U. (2006). The Crystal Structure of Human Geranylgeranyl Pyrophosphate Synthase Reveals a Novel Hexameric Arrangement and Inhibitory Product Binding. *Journal of Biological Chemistry*, *281*(31), 22004–22012. https://doi.org/10.1074/jbc.m602603200
91. Pemberton, T. A., Chen, M., Harris, G. G., Chou, W. K. W., Duan, L., Köksal, M., Genshaft, A. S., Cane, D. E., & Christianson, D. W. (2017). Exploring the Influence of Domain Architecture on the Catalytic Function of Diterpene Synthases. *Biochemistry*, *56*(14), 2010–2023. https://doi.org/10.1021/acs.biochem.7b00137

92. 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 Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436
93. Capella-Gutierrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348
94. Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLOS ONE*, *5*(3), e9490. https://doi.org/10.1371/journal.pone.0009490
95. Letunic, I., & Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, *44*(W1), W242–W245. https://doi.org/10.1093/nar/gkw290
96. Martin, D. P., & Bohlmann, J. (2004). Identification of Vitis vinifera (−)-α-terpineol synthase by in silico screening of full-length cDNA ESTs and functional characterization of recombinant terpene synthase. *Phytochemistry*, *65*(9), 1223–1229. https://doi.org/10.1016/j.phytochem.2004.03.018
97. Kumar, S., Kempinski, C., Zhuang, X., Norris, A., Mafu, S., Zi, J., Bell, S., Nybo, S. E., Kinison, S., Jiang, Z., Goklany, S., Linscott, K. B., Chen, X., Jia, Q., Chen, F., Bowman, J. L., Babbitt, P. C., Peters, R., Chen, F., & Chappell, J. (2016). Molecular Diversity of Terpene Synthases in the Liverwort Marchantia polymorpha. *The Plant Cell*, tpc.00062.2016. https://doi.org/10.1105/tpc.16.00062
98. Jones, W., & Kinghorn, A. D. (2012). Extraction of Plant Secondary Metabolites. *Humana Press eBooks*, 341–366. https://doi.org/10.1007/978-1-61779-624-1_13
99. Anastas, P. T., Heine, L., & Williamson, T. C. (2000). Green Chemical Syntheses and Processes: Introduction. *Acs Symposium Series*, 1–6. https://doi.org/10.1021/bk-2000-0767.ch001
100. Yang, L., Liu, H., Jin, Y., Liu, J., Deng, L., & Wang, F. (2022). Recent Advances in Multiple Strategies for the Synthesis of Terpenes by Engineered Yeast. *Fermentation*, *8*(11), 615. https://doi.org/10.3390/fermentation8110615
101. Schmidt-Dannert, C. (2014). Designer microbes for biosynthesis. *Current Opinion in Biotechnology*, *29*, 55–61. https://doi.org/10.1016/j.copbio.2014.02.014
102. Luo, Y., Li, B., Liu, D., Zhang, L., Chen, Y., Jia, B., Zeng, B., Zhao, H., & Yuan, Y. (2015). Engineered biosynthesis of natural products in heterologous

hosts. *Chemical Society Reviews*, *44*(15), 5265–5290. https://doi.org/10.1039/c5cs00025d

103. Angov, E. (2011). Codon usage: Nature's roadmap to expression and folding of proteins. *Biotechnology Journal*, *6*(6), 650–659. https://doi.org/10.1002/biot.201000332

104. Blazeck, J., & Lee, S. Y. (2013). Promoter engineering: Recent advances in controlling transcription at the most fundamental level. *Biotechnology Journal*, *8*(1), 46–58. https://doi.org/10.1002/biot.201200120

105. Gosset, G. (2017). Engineering of Microorganisms for the Production of Chemicals and Biofuels from Renewable Resources. In *Springer eBooks*. Springer Nature. https://doi.org/10.1007/978-3-319-51729-2

106. Zhuang, X., & Chappell, J. (2015). Building terpene production platforms in yeast. *Biotechnology and Bioengineering*, *112*(9), 1854–1864. https://doi.org/10.1002/bit.25588

107. Wong, J., De Rond, T., D'Espaux, L., Van Der Horst, C., Dev, I., Rios-Solis, L., Kirby, J. T., Scheller, H. V., & Keasling, J. D. (2018). High-titer production of lathyrane diterpenoids from sugar by engineered Saccharomyces cerevisiae. *Metabolic Engineering*, *45*, 142–148. https://doi.org/10.1016/j.ymben.2017.12.007

108. Beyraghdar Kashkooli, A., Van der Krol, A., & Bouwmeester, H. J. (2018). Terpenoid biosynthesis in plants. *Flavour Sci., Verlag der Technischen Universität Graz*, 3-11.

109. Heldt, H., Heldt, F., & Piechulla, B. (2005b). *Plant Biochemistry* (pp. 431-432)Academic Press.

110. Grumbach, K. H., & Lichtenthaler, H. K. (1982). CHLOROPLAST PIGMENTS and THEIR BIOSYNTHESIS IN RELATION TO LIGHT INTENSITY*. *Photochemistry and Photobiology*, *35*(2), 209–212. https://doi.org/10.1111/j.1751-1097.1982.tb03833.x

111. Siddiqui, M. S., Thodey, K., Trenchard, I., & Smolke, C. D. (2012). Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools. *Fems Yeast Research*, *12*(2), 144–170. https://doi.org/10.1111/j.1567-1364.2011.00774.x

112. Williams, D. R., McGarvey, D. J., Katahira, E., & Croteau, R. (1998). Truncation of Limonene Synthase Preprotein Provides a Fully Active 'Pseudomature' Form of This Monoterpene Cyclase and Reveals the Function of the Amino-Terminal Arginine Pair. *Biochemistry*, *37*(35), 12213–12220. https://doi.org/10.1021/bi980854k

113. Turner, G. C., Gershenzon, J., Nielson, E. E., Froehlich, J. E., & Croteau, R. (1999). Limonene Synthase, the Enzyme Responsible for Monoterpene Biosynthesis in

Peppermint, Is Localized to Leucoplasts of Oil Gland Secretory Cells1. *Plant Physiology*, *120*(3), 879–886. https://doi.org/10.1104/pp.120.3.879

114. Emanuelsson, O., Nielsen, H., Brunak, S., & Von Heijne, G. (2000). Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. *Journal of Molecular Biology*, *300*(4), 1005–1016. https://doi.org/10.1006/jmbi.2000.3903

115. Petersen, T. S., Brunak, S., Von Heijne, G., & Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, *8*(10), 785–786. https://doi.org/10.1038/nmeth.1701

116. Weber, E. J., Engler, C., Gruetzner, R., Werner, S., & Marillonnet, S. (2011). A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLOS ONE*, *6*(2), e16765. https://doi.org/10.1371/journal.pone.0016765

117. Lee, M. D., DeLoache, W. C., Cervantes, B., & Dueber, J. E. (2015). A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synthetic Biology*, *4*(9), 975–986. https://doi.org/10.1021/sb500366v

118. Gietz, R. D., & Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols*, *2*(1), 31–34. https://doi.org/10.1038/nprot.2007.1

119. Schmid, R., Heuckeroth, S., Korf, A., Smirnov, A., Myers, O., Dyrlund, T. S., Bushuiev, R., Murray, K. J., Hoffmann, N., Lu, M., Sarvepalli, A., Zhang, Z., Fleischauer, M., Dührkop, K., Wesner, M., Hoogstra, S. J., Rudt, E., Mokshyna, O., Brungs, C., . . . Pluskal, T. (2023). Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nature Biotechnology*. https://doi.org/10.1038/s41587-023-01690-2

Supplementary data

Supplementary table 1. *A comprehensive inventory of primers is presented for each plasmid utilized in this thesis. The list includes information on the type, sequence, complementary plasmids, and melting temperatures of the primers. The primers were designed and synthesized for both PCR and sequencing applications. A total of 29 forward (F in the table) and 2 reverse (R in the table) primers were designed for the purposes of this thesis. The procurement of these primers was facilitated through Generi Biotech, a biotechnological company based in Hradec Králové, Czech Republic, where the order was placed.*

Primer	Type	Complementary plasmid	Sequence	Tm [°C]
pug1	F	pTP0098, pTP00195	AGCTGAACATTTCGGTGAAG	55
pug2	F	pTP0098, pTP00195	TAGAAAGTGGATCGAAGAGTC	53
pug3	F	pTP0098, pTP00195	TGTGGTCCACAAGAAAGGAC	57
gfq1	F	pTP0099, pTP00196	GTTGGAAATTGCTCAAACCG	56
gfq2	F	pTP0099, pTP00196	TGAACAAAAGGCCTCTCAAG	55
gfq3	F	pTP0099, pTP00196	TGAACGACAAGAACTTGCTGC	58
gfq4	F	pTP0099, pTP00196	GTCTAAAACTGGTGAAATCCC	53
gim1	F	pTP0100, pTP00197	TTTGGAAAACCGGCGAAAAG	57
gim2	F	pTP0100, pTP00197	GTTGATTTGTTTGAGCACTTG	53
gim3	F	pTP0100, pTP00197	AATACCCATTCCATGCCAAC	55
gim4	F	pTP0100, pTP00197	GCTGAAAGATGGGATGCTAC	55
gira1	F	pTP0101, pTP00198	GGTATGGTTGAATTGGCTAG	52
gira2	F	pTP0101, pTP00198	CAACATTGCTTGATTGATGG	53
gira3	F	pTP0101, pTP00198	CATCGATTTGTTGCAATTGG	53
gira4	F	pTP0101, pTP00198	CTTGGTCTAGAACAGGTCAAG	54
H5_1	F	pTP0102, pTP00199	ACCTAAAGCTGGTTTTGGTG	55
H5_2	F	pTP0102, pTP00199	GTCTTGTTGGGTTTAGATAAG	50

Primer	Type	Complementary plasmid	Sequence	Tm [°C]
H5_3	F	pTP0102, pTP00199	TGAAGGTAGAGATATGGTTAC	50
H5_4	F	pTP0102, pTP00199	TGAGAACCAGACATTCTGTG	55
mapoly1	F	pTP0103, pTP00200	CACAATTTCCACAATCATTG	50
mapoly2	F	pTP0103, pTP00200	CGATGATGGTTCTATTTGTTG	52
mapoly3	F	pTP0103, pTP00200	ATTGTCTTATCCAGGTGAAAC	52
mapoly4	F	pTP0103, pTP00200	TAGAATTGCTGCTTCTAAGTG	53
gff1	F	pTP0104, pTP00201	ACACGGTCACAGGATCTCTTC	58
gff2	F	pTP0104, pTP00201	GTTTCCCACGTTAGAGATAG	51
m2q1	F	pTP0105, pTP00202	CGATTTGGTTAGATGCGGTC	56
NZ7_1	F	pTP0106, pTP00203	CATGCTTCTAGATTTGTTTC	49
TH0033_F	F	pYTK001, pTP0098-pTP0106	TTGCTGGCCTTTTGCTCACATG	62
TH0033_R	R	pYTK001, pTP0098-pTP0106	ACCTCAGAACTCCATCTGGATTTGTTCAG	62
HS_R	R	pTP0195-pTP0203	CTTCCTTCTGTTCGGAGATTACCGAATC	61