Forschungsberichte   Nr. 25

Laura Drummond

2021

Investigation of Non-Canonical Terpene Biosynthesis

Investigation of Non-Canonical Terpene Biosynthesis

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich 15 – Biowissenschaften
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Laura Drummond
aus Curitiba, Brasilien

Frankfurt am Main (2021)

(D 30)
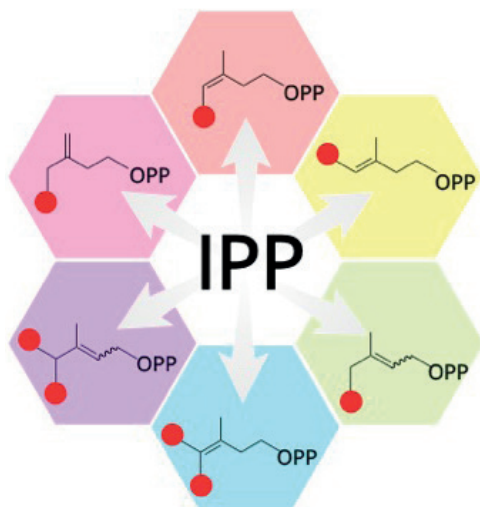
Schriftenreihe des DECHEMA-Forschungsinstituts

Band 25

Laura Drummond

Investigation of Non-Canonical Terpene Biosynthesis

Publications:

Expanding the Isoprenoid Building Block Repertoire with an IPP Methyltransferase from Streptomyces monomycini
Laura Drummond, Max J. Kschowak, Jürgen Breitenbach, Hendrik Wolff, Yi-Ming Shi, Jens Schrader, Helge B. Bode, Gerhard Sandmann, and Markus Buchhaupt
ACS Synthetic Biology **2019** *8* (6), 1303-1313
DOI: 10.1021/acssynbio.8b00525

Microbial Cell Factories for the Production of Terpenoid Flavor and Fragrance Compounds
Florence M. Schempp, Laura Drummond, Markus Buchhaupt, and Jens Schrader
Journal of Agricultural and Food Chemistry **2018** *66* (10), 2247-2258
DOI: 10.1021/acs.jafc.7b00473

Patent application:

Markus Buchhaupt, Max Kschowak, Laura Drummond, Jens Schrader. EP3425051 - METHOD FOR THE PRODUCTION OF ISOPRENOIDS BY HETEROLOGOUS EXPRESSION OF A PROTEIN IN A MICROORGANISM. Application Number 17179874, European Patent Office, 2017.

Acknowledgements

Table of Contents

List of Abbreviations

2MB	2-Methylenebornane
2MIB	2-Methylisoborneol
AACT	Acetoacetyl-CoA-Thiolase
ATP	Adenosine Triphosphate
DMAPP	Dimethylallyl Pyrophosphate
FPP	Farnesyl Pyrophosphate
G3P	Glyceraldehyde-3-phosphate
GC	Gas Chromatography
GGPP	Geranyl Geranyl Pyrophosphate
GPP	Geranyl Pyrophosphate
HDR	HMBPP reductase
HMBPP	4-Hydroxy-3-methylbut-2-enyl-diphosphate
HMG-CoA	3-Hydroxy-3-methylglutaryl-CoA
HMGR	HMG-CoA-Reductase
HMGS	HMG-CoA-Synthase
IDI	IPP/DMAPP-Isomerase
IPK	Isopentenyl Phosphate Kinase
IPP	Isopentenyl Pyrophosphate
JH	Juvenile Hormone
LC	Liquid Chromatography
MAT	Methionine Adenosyl Transferase
MCP	MEP Cytidylyltransferase
MDS	MecPP-Synthase
MEcPP	2-C-Methyl-D-erythritol 2,4-cyclodiphosphate
MEP	2-C-Methyl-D-erythritol-4-phosphate
MK	Mevalonate Kinase
MS	Mass Spectrometry
MT	Methyltransferase
MTAN	5'-Methylthioadenosine Nucleosidase
MVA	Mevalonate
MVAP	5-Phosphomevalonate, (R)-MVA-5-phosphate
MVAPP	5-Diphosphomevalonate
PDMS	Polydimethylsiloxane
PMD	mevalonate diphosphate decarboxylase
PMK	5-Phosphomevalonatekinase
PMVD	5-phosphomevalonate decarboxylase
PT	Prenyltransferase
SAM	S-Adenosylmethionine

SBSE	Stir Bar Sorptive Extraction
SPME	Solid Phase Micro Extraction
TD	Thermodesorption

List of Figures

List of Tables

Zusammenfassung

Terpene sind eine der größten und vielfältigsten Klassen von Naturprodukten, die von Organismen aus allen Reichen des Lebens produziert werden und wichtige Anwendungen in der Pharma-, Aroma- und Duftstoffindustrie haben. Bekannte Beispiele für Terpene sind die Pharmazeutika Artemisinin und Taxol, die Aroma- und Duftstoffe Menthol, Santalol und Sclareol, das Strukturmaterial Polyisopren und die Biokraftstoff-Vorstufe Farnesen.

Terpenproduktion für kommerzielle Zwecke beruht jedoch traditionell auf der Extraktion aus Pflanzen, was zu Schwankungen im Preis und in der Verfügbarkeit führt. Desweiteren werden die natürlichen Ressourcen weltweit immer knapper und angesichts der Komplexität der Terpenchemie, stellt die chemische Synthese meist keine nachhaltige oder praktikable Lösung dar. Vor diesem Hintergrund bietet die mikrobielle Synthese in Verbindung mit den Prinzipien der synthetischen Biologie eine Herstellungsmöglichkeit, die die oben genannten Probleme der traditionellen Terpen-Produktionsmethoden umgeht. Die gentechnische Veränderung von Mikroben zur Expression von Stoffwechselwegen und Schlüsselenzymen ermöglicht die Produktion von Verbindungen aus einfachen Zuckern, ohne die Erzeugung von toxischen Abfällen, die bei der chemischen Synthese üblich sind.

Die Biosynthese aller Terpene basiert auf den universellen C5-Vorstufen Isopentenylpyrophosphat (IPP) und Dimethylallylpyrophosphat (DMAPP), aus denen weitere Prenylpyrophosphat-Vorstufen gebildet werden. Dieser biogenetische Weg ist in der "Isopren-Regel" von Otto Wallach (Wallach 1887) und Leopold Ružička (Ruzicka 1953) beschrieben, die postuliert, dass alle Terpene aus C5-Kohlenstoffbausteinen aufgebaut sind. Ein oder mehrere IPP-Moleküle werden mit DMAPP kondensiert und bilden größere Prenylpyrophosphate mit wachsender Anzahl von Kohlenstoffatomen. Verschiedene Terpen-Synthasen katalysieren die Umsetzung der Substrate DMAPP, GPP (Geranylpyrophosphat, C10), FPP (Farnesylpyrophosphat, C15) oder GGPP (Geranylgeranylpyrophosphat, C20) in

Terpenprodukte. Nach Ružička sind die einzigen bekannten terpenabgeleiteten Naturprodukte mit einer Anzahl von Kohlenstoffatomen, die mit der Isopren-Regel nicht vereinbar sind, Moleküle, die Abbaustufen durchlaufen, die zu einer Verkürzung führen, z. B. Ionon, oder nachfolgend modifiziert werden, z.B. durch Methylierungen und Acetylierungen. Die Allgemeingültigkeit dieser Regel wurde jedoch vor einigen Jahren von drei Gruppen (Dickschat et al. 2007; Komatsu et al. 2008; C. M. Wang and Cane 2008) widerlegt, die den Biosyntheseweg von 2-Methylisoborneol in Myxobakterien und Streptomyceten aufklärten. Im Falle von 2-Methylisoborneol wandelt die entsprechende Monoterpen-Synthase (2-MIB-Synthase) 2-Methyl-GPP, das durch eine spezifische Methyltransferase gebildet wird, direkt in ein C11-Produkt um.

Der Nachweis der Existenz einer Prenylpyrophosphat-Methyltransferase erweiterte die möglichen Produkte für die Terpen-Biosynthese in starkem Maße, da durch die Methylierung von Terpenvorläufern die Anzahl der Strukturen, die gebildet werden können vervielfacht werden. Die modulare Charakteristik der Terpensynthese erleichtert die Übertragung von Modifikationen, die in ihre Bausteine eingeführt wurden. In der vorliegenden Arbeit wurde die Hypothese aufgestellt, dass weitere Prenylpyrophosphat-Methyltransferasen existieren, mit denen die biosynthetische Kapazität des Terpen-Biosynthesewegs erweitert und neue Terpenstrukturen mit unkonventioneller Anzahl von Kohlenstoffatomen ermöglicht werden. Die Diversifizierung der Strukturen dieser Verbindungsklasse durch die Generierung neuer, unkonventioneller Moleküle kann die Entdeckung neuer biologischer Funktionen für die Pharma-, Geschmacks- und Duftstoffindustrie ermöglichen.

Die Verwendung von Methyltransferasen, die Prenylpyrophosphat-Vorstufen von Terpenen modifizieren, war der Ausgangspunkt dieser Arbeit. Es wurde eine BLAST-Suche in der NCBI-Datenbank mit der GPP-Methyltransferase-Proteinsequenz als Query durchgeführt und entsprechende Homologe hinsichtlich ihrer flankierenden Gene analysiert. Eine putative SAM-abhängige Methyltransferase aus *Streptomyces monomycini* wurde ausgewählt und heterolog exprimiert. Es wurden *in vitro* und *in vivo* Experimente durchgeführt und die Funktion dieses Enzyms aufgeklärt. Hierbei wurde festgestellt, dass die neu entdeckte Methyltransferase IPP methyliert und dabei verschiedene C6- und C7-Prenylpyrophosphate bildet, die als Bausteine für die

XX

Synthese von unkonventionellen Terpenoidverbindungen verwendet werden können. *In vitro*-Experimente mit dem gereinigten Enzym, dem Substrat IPP und dem Cofaktor SAM ergaben mono-methylierte (C6) und di-methylierte (C7) Produkte. Die heterologe Expression der IPP-Methyltransferase in *E. coli* zusammen mit Enzymen aus dem Mevalonatweg ermöglichte zudem die Produktion und Identifizierung noch längerer Terpene, die methyliertes IPP oder DMAPP in ihrer Struktur enthielten und C11-, C12-, C16- und C17-Verbindungen bildeten. Außerdem wurde der Wildtyp-Stamm von *S. monomycini* kultiviert und seine Metabolite mittels GC-MS und LC-MS analysiert.

Um neue Methoden zur Erzeugung nicht-konventioneller Terpene zu identifizieren wurden in dieser Arbeit die Auswirkungen der Cofaktor-Manipulation untersucht. Für eine breitere Diversifizierung der Produkte durch Cofaktor-Modifikation wurden Analoga von S-Adenosyl-Methionin (SAM) mit dem Enzym Methionin-Adenosyl-Transferase, auch bekannt als SAM-Synthetase, erzeugt. Methionin, Ethionin und L-Methionin-(Methyl-^{13}C) wurden als Vorstufen verwendet und die Bildung von SAM-Analoga, wie S-Adenosyl-Ethionin (SAE) und ^{13}C-SAM, mit LC-MS nachgewiesen. Im Falle von L-Methionin-(Methyl-^{13}C) wurde der modifizierte Cofaktor ^{13}C-SAM erfolgreich in *in vitro* Assays mit einer Methyltransferase zur Bildung von Terpenen mit Modifikationen, z. B. Isotopen-markiertem Methyl-Isoprenol, eingesetzt. Die Anwendung der SAM-Synthetase *in vitro* führt zur Bildung von SAM-Analoga, die von der IPP-Methyltransferase zur Alkylierung von IPP genutzt werden können. Die Promiskuität dieser Methyltransferase bezüglich der übertragenen Gruppe von ihrem Cofaktor lässt vermuten, dass auch andere funktionelle Gruppen in ähnlicher Weise genutzt werden könnten. Die Übertragung verschiedener Gruppen durch diese Methyltransferase könnte in Zukunft für die Herstellung einer noch größeren Bandbreite von modifizierten Isoprenoiden genutzt werden.

Die intrinsische Promiskuität von Enzymen, die zur Terpen-Biosynthese-Maschinerie gehören, wie z. B. Terpen-Synthasen und Prenyl-Transferasen, wurde in früheren Arbeiten genutzt, um das Produktspektrum durch die Verwendung nicht-natürlicher Substrate zu erweitern. Ein synthetisches fluoriertes FPP-Analogon wurde von der 5-Epi-aristolochen-Synthase unter *in vitro*-Bedingungen als Substrat verwendet, was zur Bildung eines fluorierten Produkt führte (Faraldos et al. 2007). Analog dazu wurde gezeigt, dass eine Farnesyltransferase, die an der Proteinprenylierung beteiligt ist,

verschiedene FPP-Analoga akzeptiert (Subramanian et al. 2012). Mehre Studien mit synthetischen Prenylpyrophosphat-Analoga zeigten die Promiskuität von Prenyltransferasen, sowohl hinsichtlich der Substrate als auch der Produkte (Tanetoshi Koyama et al. 1980; Nagaki et al. 2012). Die Promiskuität solcher Enzyme kann genutzt werden, um eine breitere synthetische Produktvielfalt des Terpen-Biosynthesewegs zu erreichen, indem unnatürliche oder unkonventionelle Vorstufen verwendet werden. In der vorliegenden Arbeit wurde die Promiskuität der pflanzlichen Terpensynthasen untersucht, um eine Vielzahl von C11-Verbindungen zu erzeugen. Dazu wurden *in vitro* Assays mit dem nicht-kanonischen C11-Prenylpyrophosphat 2-Methyl-GPP als Substrat durchgeführt und die gebildeten Verbindungen mit GC-MS nachgewiesen. Der C11-Vorläufer wird bekanntermaßen von Mikroorganismen produziert, sein Vorkommen wurde jedoch nie in Pflanzen beschrieben. Obwohl 2-Methyl-GPP kein natürliches Substrat für eine der verwendeten Terpen-Synthasen ist, setzten die meisten der Pflanzenenzyme das Prenylpyrophosphat erfolgreich in eine Reihe von C11-Terpenprodukten um. Es wurde gezeigt, dass nicht-kanonische Prenylüyrophosphate als nicht-natürliche Substrate von verschiedenen Terpen-Synthasen verwendet werden können, um Produkte mit einer ungewöhnlichen Anzahl von Kohlenstoffen zu bilden.

Die Ergebnisse zeigen eine Vielzahl von Techniken, mit denen der kanonische Weg der Terpenbildung erweitert werden kann. Es wurde gezeigt, dass eine IPP-Methyltransferase eine Vielzahl von neuen C6- und C7-Prenylpyrophosphaten produzieren kann. Durch die Verwendung von methylierten Versionen des universellen Vorläufers IPP in der Terpen-produzierenden Maschinerie, sind viele andere nicht-kanonische Terpenstrukturen möglich. Die Anwendbarkeit dieser neuen Bausteine wurde durch Experimente *in vivo* gezeigt, bei denen die Promiskuität der nativen Prenyltransferasen von *E. coli* die Bildung längerer nicht-kanonischer Vorstufen ermöglichte, die eine ungewöhnliche Anzahl von Kohlenstoffatomen enthalten. Dieses neue Konzept der Methylierung von Terpenvorstufen erweitert die vor Jahrzehnten postulierte Isopren-Regel, mit der die Terpenbiosynthese immer noch als sequentielle Addition von 5-Kohlenstoffatom-Einheiten beschrieben wird, erheblich. Die in dieser Arbeit, aber auch in neueren Arbeiten anderer Forschungsgruppen beschriebenen Ergebnisse (Von Reuss et al. 2018; Ozaki et al. 2018; Ignea et al. 2018; Eiben et al. 2019), sind eine Erweiterung dieses Konzepts.

Methyltransferasen wurden hier im Zusammenhang mit mikrobiellen Zellfabriken und *in vitro* Enzymreaktionen verwendet und könnten auf andere Organismen übertragen werden, die Stoffwechselwege mit Prenyldiphosphaten als Metaboliten beinhalten. Es wurde auch gezeigt, dass die Methyltransferase selbst verschiedene Kofaktor-Analoga akzeptieren kann, wodurch andere Alkylgruppen auf ihr Substrat übertragenwerden können und so die Möglichkeiten der Produktspektren diversifiziert werden. Darüber hinaus wurde die Promiskuität der Terpensynthasen gegenüber methylierten Substraten demonstriert, wobei ein C11-Prenylpyrophosphat in eine Vielzahl von C11-Terpenen umgewandelt wurde. Die Anwendung dieser Methoden in etablierten Produktionssystemen könnte zur Produktion neuer Substanzen führen, mit Anwendungen in allen Bereichen, in denen Terpene eine Rolle spielen. Diese Techniken beweisen darum die Anwendbarkeit von Methyltransferasen, Prenyltransferasen, Methionin-Adenosyl-Transferasen und Terpen-Synthasen in verschiedenen Systemen, für die Produktion von unkonventionellen Terpenen. Die kombinierten Ergebnisse eröffnen einen Präzedenzfall für Designer-Moleküle, die auf diese Weise gebildet werden können, und ein breites Spektrum an neuen Terpenen, das darauf wartet, entdeckt, getestet und in verschiedenen Bereichen eingesetzt zu werden. Die Suche nach neuen Molekülen mit pharmakologischen Wirkungen ist in einer Welt, in der die Resistenz von Krankheitserregern gegen Antibiotika zunimmt, von größter Bedeutung. Auch die Aromen- und Duftstoffindustrie könnte von neuen Aromamolekülen mit neuen, noch unentdeckten Eigenschaften profitieren, die in das Portfolio aufgenommen werden könnten.

Abstract

Terpenes are one of the largest and most diverse class of natural products, produced by organisms from all kingdoms of life and with important applications in the pharma, flavor and fragrance industries. Well-known examples of terpenes are the pharmaceuticals artemisinin and taxol, the flavor and fragrance compounds menthol, santalol and sclareol, the structural material polyisoprene and the biofuel precursor farnesene.

The biosynthesis of all terpenes is based on universal C5 precursors isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), from which further prenyl pyrophosphate precursors are formed. This biogenetic path is described in the "isoprene rule" from Otto Wallach (Wallach 1887) and Leopold Ružička (Ruzicka 1953), that postulates all terpenes to be made up out of C5 carbon building blocks. Different terpene synthases therefore use DMAPP, GPP (geranyl pyrophosphate, C10), FPP (farnesyl pyrophosphate, C15) or GGPP (geranylgeranyl pyrophosphate, C20) as substrates to catalyze their conversion into terpene products. According to Ružička, the only known terpene-derived natural products with a number of carbon atoms incompatible with the isoprene rule are molecules that undergo degradation steps leading to a shortening of the primary terpene synthase products, e.g. ionone, or subsequent methylations, acetylations or other modifications. The general validity of this rule was however disproven some years ago by three different groups, that all clarified the biosynthetic pathway of 2-methylisoborneol in myxobacteria and streptomycetes. In case of 2-methylisoborneol, the respective monoterpene synthase (2-MIB synthase) converts 2-methyl-GPP, which is formed by a specific methyltransferase, directly into a C11 product.

The use of methyltransferases that modify prenyl pyrophosphate precursors of terpenes was the starting point of this work. A BLAST search on the NCBI database was made using GPP methyltransferase protein sequence as query, and homologues were analyzed regarding the flanking genes. A methyltransferase from *Streptomyces monomycini* was selected and expressed heterologously. *In vitro* and *in vivo*

experiments were performed, and the function of this enzyme was clarified. The newly discovered methyltransferase was found to methylate IPP forming a variety of C6 and C7 compounds, some of which were demonstrated to be incorporated into longer methylated prenyl pyrophosphates, when the protein was expressed *in vivo*. Moreover, the wild type *S. monomycini* strain was cultivated and had its compounds analyzed by GC-MS and LC-MS.

Aiming at a wider diversification of products through cofactor modification, SAM analogs were created with the enzyme methionine adenosyl transferase, which synthesizes SAM. Methionine, ethionine and L-methionine-(methyl-^{13}C) were used as precursors, and the formation of SAM analogs was detected with GC-MS. In the case of L-methionine-(methyl-^{13}C), the modified cofactor was successfully used in *in vitro* assays with a methyltransferase to form terpenes with modifications, e.g. isotopically labeled methyl-isoprenol.

The promiscuity of plant terpene synthases was explored to create a variety of C11 compounds *in vitro*. For that, 2-methyl-GPP was provided as substrate and the formed compounds were detected with GC-MS. Even though the C11 precursor is not the natural substrate for any of the terpene synthases used, most of the active enzymes successfully converted the prenyl pyrophosphate precursor into a series of C11 terpene products.

The methods and results presented in this work offer a variety of ways to modify prenyl pyrophosphate precursors for the creation of new terpene molecules. The application of these methodologies in well-established production systems could lead to the production of new substances, with applications in the industrial fields of pharmaceuticals, flavors and fragrances, and biofuels.

1 Introduction

Natural products comprise chemical compounds produced by living organisms, many of which are extracted for human use. These natural products are synthesized by activation of secondary metabolism, meaning that molecules derived from the primary metabolism, like sugars and amino acids among others, are further processed to produce them. Terpenes are a class of naturally produced molecules which originate biologically from the same C5 precursors and are present in virtually all forms of life. Terpenes and its derivates comprise one of the largest and most diverse class of secondary metabolites, reaching over 80.000 different structures and accounting for almost one third of the compounds contained in the Dictionary of Natural Products (http://dnp.chemnetbase.com) (Christianson 2017). Besides being widely distributed across the kingdoms of life regarding their natural production by living organisms as abovementioned, they are very present in our daily lives in consumer products. The aromas of lemons (limonene), of grapefruits (nootkatone), and of mints (menthol) are just a few examples of how mankind's routine is permeated by this class of molecules. They are among the most structurally varying and versatile classes of chemical compounds, and present therefore a broad assortment of biological functions, ranging from semiochemicals involved in ecological interactions until crucial pharmacological agents against diseases like malaria and cancer.

Terpenes, isoprenoids and terpenoids are names given for this group of metabolites widely spread and with very diverse structures derived from 5-carbon containing isoprene-like precursors. The following thesis focuses on methods to diversify terpene chemical structures by changing its basic building blocks. As introductory information, the present section will concentrate on explaining what criteria are used to classify terpene molecules, where they can be found, and provide examples of functions found in nature. Furthermore, the specifics of terpene biosynthesis will be described, including the pathways to non-canonical terpenes and enzymes directly involved with their production. At last, production of terpenes using microbes and the motivation for this thesis will close the introductory section.

1.1   Terpenes, Terpenoids, Isoprenoids

The structural diversity of terpenes, terpenoids and isoprenoids relies on innumerous ways of condensing, rearranging and modifying the basic structure of an isoprene molecule (2-methylbuta-1,3-diene, C5). Isoprene is not the precursor of these compounds, but its C5 structure is enclosed in virtually all terpene structures. The reason for that is that the universal precursors of this class of molecules are C5 isoprene-like compounds, the specifics of this subject will be described in detail later on in the biosynthesis session. The fact that isoprene was identified within the structures of terpenes led to the name isoprenoid, which basically means "isoprene-like". The IUPAC defines isoprenoids as compounds considered derived from isoprene, meaning that in isoprenoids the skeleton of an isoprene can generally be identified within the molecule in a repeated way. The isoprenoid class includes hydrocarbons and also oxygenated derivatives. Terpenoids, as referred by the IUPAC, encompass natural products and related compounds, also derived from isoprene units like the isoprenoid counterpart. Terpenoids usually contain oxygen in its functional groups and like terpenes, are subdivided in groups according to the number of carbon atoms in their structure. The definition of terpenes by the IUPAC, classifies terpenes as hydrocarbon molecules biologically originated, that have carbon skeletons derived from isoprene. The terpene class is also subdivided into subclasses, like the terpenoids, according to the number of isoprene units contained in the molecule structure. The definition of terpenes is very similar to the descriptions of isoprenoid and terpenoid compounds, according to IUPAC (Nič et al. 2009), making them virtually undistinguishable with almost redundant definitions. Basically, the definition puts terpenes as simple hydrocarbons while terpenoids and isoprenoids contain other functional groups. However, this division according to functional groups or heteroatoms does not add any classificatory or organizational value to the present work. The definitions found generally in reference sources like encyclopedias and dictionaries are often controversial, posting assumptions such as that terpenes are only hydrocarbons, when oxygen atoms are extensively found in common terpene alcohols like geraniol. For this reason and to avoid confusion about the different meanings, in this work the terms isoprenoid and terpenoid will be used as synonyms, as well as the dated but more widely known term terpenes. Terpene, terpenoid and

2

isoprenoid are hereby defined as natural compounds derived from 5-carbon precursors, which will be described in the biosynthesis section.

1.2   Origin and History

The first reported use of terpenes by humans goes back as far as ancient Egypt, when essential oils from plants were used in ceremonies and as medicines. The use of mint oil as herbal medicine in the Egyptian culture is registered in Ebers papyrus dated from 1.550 B.C. (Elansary and Ashmawy 2013). Egyptians used to grow plants specifically for therapeutical use, and even though this civilization if reference on recorded history, due to the fact that information was written down and could be recovered afterwards by contemporary scientists, there is plenty of evidence showing that convergently, Chinese, Hindis and later on Romans were also developing the art of essential oil usage (Guenther 1949). Especially in India, the extraction and usage of oils from plants for therapies integrated the traditional Ayurveda system of medicine, which came into existence around 5000 years ago. Ayurveda still is an important part of the hindu healthcare system, and it is considered the foundation of ancient medical science in India (Kapoor 2017). It focuses on the importance of diet for promotion of health and on the use of herbal compounds - many of them rich in terpene compounds - for therapeutic formulas used in prevention and treatment of diseases (Govindarajan, Vijayakumar, and Pushpangadan 2005). Extraction of oils from odorous plants like rosemary and sage has been a common practice for centuries, either for their effects in healing or for their perfumery usage.

Turpentine, a resin extracted from conifer trees that has a high content of terpenes, was used as medicine in medieval times and became a very popular solvent in the beginning of the 19th century (Saeidnia 2014). The name "terpen" was coined by Kekulé to denominate hydrocarbons originated from turpentine, hence the name "terpene" in English (Hillier and Lathe 2019). Originally, turpentine was obtained from the terebinth tree (*Pistacea terebinthus*), and the Greek designations "terebinthos" for the tree and "terebinthine" for the extracted resin developed into the name "turpentine" (Saeidnia 2014). Today turpentine oil is a byproduct from the pulp wood industry, and is commonly used as solvent, as a source of alpha- and beta-pinene and to produce

the commodity chemical isoprene (B. L. Davis, Goldblatt, and Palkin 1946; Herty and Graham 1914). By the time that the term terpene was created, the existence of isoprene and its connection with turpentine was already acknowledged. In 1884 when William Tilden distilled isoprene from turpentine, the chemical formula C5H8 was already known for at least 20 years (Tilden 1884; Hillier and Lathe 2019). But it was only with the work of Otto Wallach in 1887 that the position of isoprene was made central to terpene chemistry. The findings of the german chemist during his lifetime work on terpene biochemistry would then be consolidated in the next century by Leopold Ružička in his work on the biogenesis of terpenes, stipulating the isoprene rule. With the development of science and technology in the last century, and the rise of biochemistry, new methods for the discovery of their intimate structures have come to light, and thus the study of terpenes has become a more (bio) chemical discipline.

1.3   Isoprene rule

The isoprene rule states that terpenes are composed of 5-carbon-atom units and therefore their precursors contain a number of carbon atoms that is always a multiple of five. Otto Wallach first conceived this idea after his extensive work on terpenes, even though the actual identity of terpene precursors was at that time unknown. During his studies, he noticed that terpene structures enclosed 5-carbon units in a repeated way, which permitted the classification of such molecules in subclasses conforming to the number of units they encompassed (Wallach 1887). He identified the basic unit as isoprene (C5H10, Figure 1) and enunciates the isoprene rule as the synthesis of terpenes being dependent on the use of C5 isoprene units. Thus, terpenes containing one isoprene unit, and therefore 5 carbon atoms, are hemiterpenes; two isoprene units and thus 10 carbon atoms form monoterpenes (Figure 2); C15 molecules with three isoprene units are sesquiterpenes. Within the rule, all terpenes should contain a number of carbon atoms that would be a multiple of five. On the next century, the rule was extended to encompass diterpenes (C20 molecules containing four isoprene units), triterpenes (C30), tetraterpenes (C40) and also terpenes with an uneven number of carbon atoms.
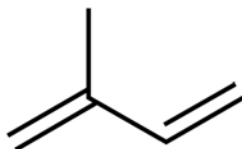
Figure 1. Structure of isoprene.

More than 60 years after the publication of Otto Wallach's work, Leopold Ružička extended the isoprene concept in his biogenetic isoprene rule, which postulated the biochemical synthesis of terpenoids from C5 intermediates but also included the possibility of formation of non-canonical terpenes. These non-canonical terpenes with a number of carbon atoms different than a multiple of five would be derivatives from degradation, enzymatic cleavage, or other modifications of terpenes (Ruzicka 1953). Examples of non-canonical terpenoids are carotenoid cleavage products like the aroma compound from violet, β-ionone ($C_{13}H_{20}O$), the plant signal molecule 4,8-dimethylnona-1,3,7-triene (DMNT, $C_{11}H_{18}$) derived from P450-catalyzed cleavage of (E)-nerolidol, or the earthy flavor substance geosmin, which is produced by fragmentation of one of the primary bacterial geosmin synthase products germacrenediol during the terpene synthase reaction (Cane et al. 2006; J. Jiang, He, and Cane 2007; C. M. Wang and Cane 2008).
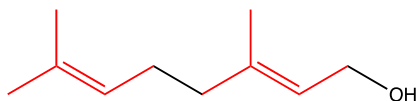
Figure 2. Structure of geraniol highlighting the isoprene units (red).

1.4   Natural occurrence and functions

Ubiquitous in nature, terpenes are produced by organisms belonging to all kingdoms of life, having key roles on the survival of individuals as well as communities (Gershenzon and Dudareva 2007; Hillier and Lathe 2019). Even though they are most commonly associated to aroma compounds originated from plants, organisms like fungi, bacteria and mammals also produce terpenes, including humans (e.g. cholesterol and steroid hormones), adding up to an immense variety of structures and functions (Figure 3). Isoprenoids are one of the oldest compounds classes produced by living organisms, having been found in the form of hopanoids in geological deposits dated from 2,5 billion years ago (Summons et al. 1999; Brocks et al. 1999). The presence of this type of compounds in billion-year-old sediments as fossilized sterols, points to the relevance of this class of compounds as ancient lipids with fundamental importance in the development of cellular life on earth (Ourisson and Albrecht 1992; Melendez, Grice, and Schwark 2013). It has been suggested that terpenes were so crucial for the establishment of biological life in our planet that they might be referred to as the fourth molecular basis of metabolism, together with DNA, RNA and proteins (Hillier and Lathe 2019). Isoprenoids form the backbone of cell membranes in archaea, while in bacteria and eukaryotes this backbone is made of fatty acids. It has been discussed, that a common ancestor from the abovementioned kingdoms might have had lipids and isoprenoids in their structure. As isoprenoid content increases, it stabilizes membranes in hydrothermal conditions, the type of environment where early life has most probably begun, but it brings an energetically high cost when it is used as membrane building block. Differences in microenvironments may have selected different membrane compositions, creating the structural dichotomy we observe today between archaea and the rest of the cells (Jordan, Nee, and Lane 2019).

On a molecular level, isoprenoids play an important role as anchors for proteins, allowing them to stick to a lipid membrane. Prenylated proteins consist in proteins with an isoprenoid polymer (mostly a farnesyl group) attached to them, which creates a hydrophobic tail allowing them to interact with the cell membrane. The presence of prenylated proteins is a fundamental feature of eukaryotic organisms, involved in growth and differentiation of cells, as well as playing a role in cell morphology and signaling. Disruptions on this delicate system are the cause of diseases such as

various types of cancer. The Ras protein, for example, can have a farnesyl group transferred to it, allowing the interaction of this protein with the cell membrane which triggers the expression of genes related to cell differentiation and growth. When the prenylated Ras protein is overactive it ends up leading to cancer, therefore this protein is a target molecule for the development of drugs for cancer treatment and prevention (Casey and Seabra 1996; Stein et al. 2015; Goodsell 1999). Besides their function in prenylating proteins, isoprenoids participate in other mechanisms with very important biological functions, like the production of signaling sterol hormones, regulation of processes such as electron transport (ubiquinones) and post-translation control, synthesis of glycoproteins (dolichols) and marking proteins for degradation. Moreover, the making of membranes from cells and organelles depends on the production of terpenoids (Hunter 2007). Ubiquinones, also called Coenzyme Q, are poly-isoprenoids found in all domains of life. In eukaryotic cells, they are positioned within the membrane bilayer of organelles, acting as indispensable part of the electron transport system that integrates the respiration machinery. They also have a role in the protection of lipoproteins, mitochondrial membrane and DNA from oxidation (Ernster and Dallner 1995). Another important membrane-forming terpenoid is cholesterol, the most common sterol molecule found in animals, present in all animal cells as structural component. Due to its tetracyclic ring (Figure 3) this membrane component is responsible for the structure and fluidity of the cell bilayer across a range of temperatures, as well as a modulator of membrane function (Ohvo-Rekilä et al. 2002; Yeagle 1991). Cholesterol is also involved in cell signaling, transmission of electrical pulses in neuronal cells (the myelin sheath is rich in cholesterol), gating of ion channels from neurotransmitter receptors, and activation of hormone receptors (Levitan, Singh, and Rosenhouse-Dantsker 2014; Incardona and Eaton 2000). As a precursor, cholesterol features in the biosynthesis of all steroid hormones like cortisol, estrogen and testosterone (Figure 3), as well as in the synthesis of vitamin D (Hanukoglu 1992; Payne and Hales 2004).

Isoprene, the prototypical terpenoid molecule, is biologically produced and released in the air in large amounts, around 1 million tons of carbon per year, being the most abundant volatile hydrocarbon of natural origin present in the atmosphere (McGenity, Crombie, and Murrell 2018; Murrell, McGenity, and Crombie 2020). Isoprene accounts for 40% all volatile organic compounds emitted by plants and is the most abundant

hydrocarbon within human breath, being measured and quantifiable in movie theaters and football stadiums (Stönner and Williams 2016; Gelmont, Stein, and Mead 1981). More than 90% of the global isoprene originates from chloroplasts of terrestrial plants, and the yearly amount of isoprene emitted is comparable to the global emissions of methane, having also influence on global climate due to its high reactivity (Murrell, McGenity, and Crombie 2020).
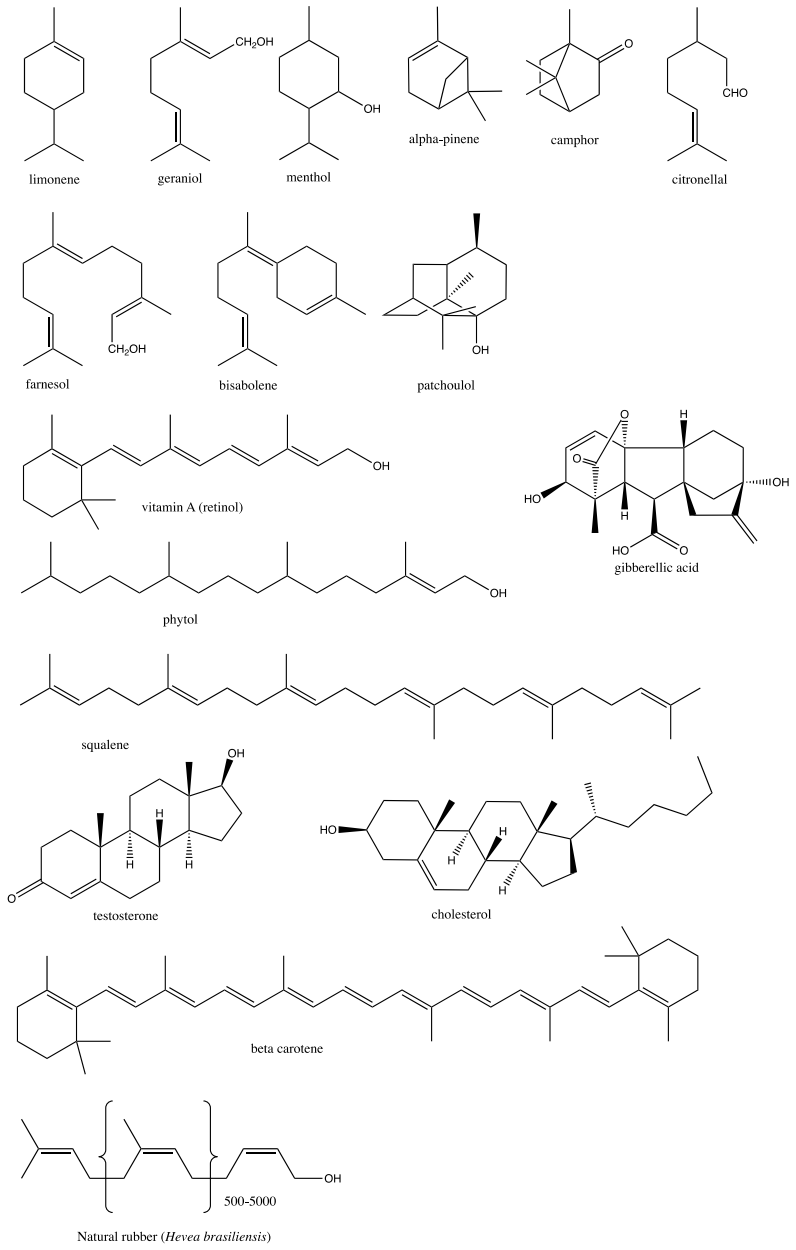
Figure 3. Diversity of terpene structures.

The ecological function of terpenes is described for some cases. In nature, these complex molecules are often regarded as communication signals for different kinds of organisms, functioning as semiochemicals and intermediating the exchange of information between species through specific signals detailed in their intricated structure. For example, the aggregation pheromone of the harlequin bug *Murgantia histrionica* is a sesquiterpene produced by a tissue-specific terpene synthase (Lancaster et al. 2018). In plants, the terpenes from secondary metabolism play an important ecological role as repellents and attractants of insects, by which the plant can call out or dismiss the presence of a beneficial or harmful insect (Dicke and Van Loon 2000). In fact, the decoration of terpenoid structures by particular enzymes, whose expression is under transcriptional control (Tholl 2006), shows that the presence of these molecules in such specific type of interaction is not a case of chance. Millions of years of coevolution between insects and plants led to a myriad of chemical signaling between them (Heil and Karban 2010) that resembles warfare – or a very intricated chemical language. The specific effects of terpenes on organisms that have evolved with them is an indication that the variation of their structures bears a signaling role, resembling language patterns used for communication and with great evolutionary advantages, evidenced by their complexity and omnipresence.

In microbes, it was fairly recently discovered that terpenes are produced and released, and that they have signaling functions involved in communication between microorganisms. For example, when the rhizobacterium *Serratia plymuthica* is exposed to volatiles from fungi, it releases the sesquiterpene sodorifen, in a specific response to the presence of another microbe (Schmidt et al. 2016). Because of the very little amounts in which microbial terpenes are produced, and because of the high level of technology needed to access information regarding these terpenes (molecular biology, biochemical, genomic, etc.), only now this knowledge is being gathered by researchers. In fact, the broad spectrum of terpene synthases in the genomes of bacteria, and their intrinsic capability of producing this class of compounds, and also to modify terpenes produced by plants, is the focus of recent research. On a detailed study using bacterial genome databases (Yamada et al. 2015), 262 different sequences were attributed as putative terpene synthases, and further analysis revealed the structures of some of their products (Figure 4) and biochemical functions, as summarized in review (Dickschat 2016). Terpene synthases are widely spread in

10

bacteria, roles ranging from the synthesis of known bacterial compounds like geosmin and 2-methyl-isoborneol, the production of compounds usually associated with plants, such as linalool, germacrene and cineol, to terpene cyclases with still unknown function. More importantly, among the molecules synthesized by bacterial terpene cyclases activity are antibiotics, like pentalenolactone, albaflavenone and terpentecin (Dickschat 2016).

2-methylisoborneol   geosmin   germacradienol   pentalenene
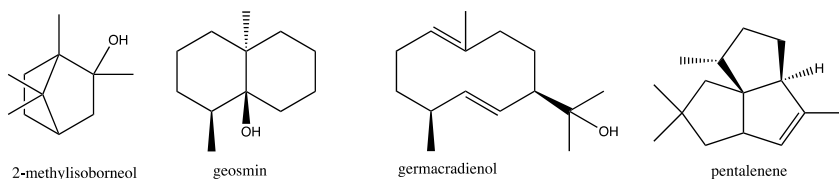
Figure 4. Structures of some terpenes produced by bacteria.

1.5   Economic importance

Commercially, terpenes are prominent substances in the industry, used mainly as fragrances for cosmetics or aromas for food preparations, but also playing an important role as main components for pharmaceuticals. The two most known medicines derived from this class of compounds are artemisinin, a sesquiterpene lactone used for treating malaria caused by *Plasmodium falciparum*; and taxol (paclitaxel), one of the most used chemotherapeutic drugs, effective against seven of the major forms of cancer, including types which are usually resistant to chemotherapy (Rowinsky and Donehower 1995). In the year 2000, taxol turned out to be the best-selling cancer drug of all time, with 1.6 billion dollars' worth of annual sales (Sofias et al. 2017).

As fragrances and flavors, the terpenoid class is no less significant – the terpene market numbers are notorious, reaching more than 500 million US dollars in 2017 and expected to grow in the upcoming years (Wu and Maravelias 2018). The influences

these chemicals exert in human psyche are also used for marketing purposes. Human memory and emotions are bound to odors (Herz 2016; Herz and Cupchik 1992), and here the structural specificity of each compound plays a part on the complexity of our perception and therefore our ability to remember determinate things. There is a new market for using aromas to influence customer's decisions, a section called Scent Marketing, in which companies are hired to produce tailor-made scents for eliciting specific responses from the public (Serras 2019).

In the industry of flavors and fragrances, terpenes have many different representatives. As monoterpenoids one can cite menthol (mint flavor), geraniol (flowery aroma), limonene (present in citric fruits like orange, lime, lemon) and pinene (pine tree aroma). They have characteristically strong odor and are very volatile. Also, small differences in their molecular configuration cause very distinct aroma impressions, making terpenes a rich source of perfumery ingredients. An example for this kind of specificity translated into function is the aroma impression of two stereoisomers of alpha-pinene – (+ and -) – which can be easily discriminated by humans even though the structures are very similar (Bentley 2006). Among sesquiterpenes the most famous aroma compounds are patchoulol (patchouli scent used in perfumery and cosmetics), santalol (from sandalwood oil, also widely used in perfumery), nootkatone (aroma of grapefruit), and valencene (one of the components of citrus aroma). Other sesquiterpenes widely present in consumer products are humulene and caryophyllene, both found in hops plants and involved in the aroma composition of hoppy beers. Besides terpenes present naturally in biological materials, there are commercially available pure compounds that are extracted from natural resources and purified, or chemically synthesized, before being used in pure form or in mixtures of calculated fractions. It has been discussed, that natural mixtures of terpenes would be more effective than pure compounds in the case of medicines (Neighbors 2018), but that is also the case for aromas, and that is actually the strategy used by companies that sell natural extracts like sandalwood oil as basis for their products. On the other hand, professional fragrance developers are responsible for selecting the right amounts of each isolated compound to be added in a blend in order to elicit the desired impression. Besides the obvious function of fragrances in the pleasant experience of smelling a perfume, another well-known application of these

fragrances and oils is aromatherapy, the use of aromas to promote health and well-being.

In the field of medicine, terpenes are considered chemo-preventatives, meaning that they have determinate effects on body organs, leading to the prevention of cancer. Some interesting compounds and their effects are described in detail in a recent review (Neighbors 2018). Among them, the monoterpene menthol was shown to have anti-cancer effects on human prostate, bladder and skin tumor cells. Carvacrol, another monoterpene, was reportedly demonstrated to have chemoprevention properties in cells from colon, breast, and liver cancer cells, among others. These examples are regarding the use of pure compounds, but many synergistic effects also occur, in which one compound enhances the effect of the other, when administered simultaneously. The studies regarding synergistic effects are however fewer than the ones using single compound compositions (Neighbors 2018).

Besides the innumerous applications mentioned above, terpenes can be generated through fermentative bioconversion of sugars coming from plant biomass, originating energy dense compounds. Efforts have been made to develop technologies for the production of terpenoids as drop-in biofuels, like bisabolane and farnesane (Phelan et al. 2017), and new molecules have been prospected for application in a near future (C. L. Liu et al. 2018). The use of terpenes as components for complex materials is exemplified by polyisoprene. Another good example of isoprene polymer is natural rubber, produced out of tree resin (Figure 3). Also, the possibility of creating new terpene-based polymers by synthetic means has been explored recently (Farhat et al. 2019).

As components of food products, terpenes are used not only for their flavor but also as coloring agents and nutrients. The most widely known nutraceuticals (functional foods) are carotenoids like astaxanthin, zeaxanthin and beta-carotene, as well as their derivatives, such as apocarotenoids. Carotenoids are pigments naturally produced with coloration properties (red, orange and yellow) found in leaves, fruits, vegetables, and many other plant derivates, as well as in algae, fungi and bacteria (Zhang 2018). For example, the colors of autumn leaves result from the refraction of light by molecules that are products of carotenoid degradation. Beta-ionone, a highly priced

and beloved aroma compound with violet flower impression, is a volatile molecule originated from carotenoid cleavage. When added to food, beverage or cosmetic products, they bring cytoprotective and antioxidant properties. The demand is so high for this type of compound, that the global market value for carotenoid compounds is expected to reach US \$1.53 billion by 2021, growing 3.9% per year (Zhang 2018; Barredo et al. 2017).

1.6 Biosynthesis

1.6.1 Biosynthesis of terpene precursors

The carbon backbones of terpenes originate from two building blocks, isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). These two universal terpene precursors can be produced through two different metabolic pathways that have compounds derived from glycolysis as starting molecules. The mevalonate pathway occurs in higher eukaryotes like animals and fungi, the cytoplasm of plants and other phototrophic organisms, archaea and some eubacteria. The non-mevalonate pathway, also called the 2-*C*-methyl-D-erythritol-4-phosphate (MEP) pathway or the 1-deoxy-D-xylulose 5-phosphate pathway (DOXP), can be found in most bacteria, in plastids of phototrophic organisms, and some protozoans like the parasite *Plasmodium*. The mevalonate pathway is responsible for forming isopentenyl-5-diphosphate (IPP) starting from acetyl-CoA, while the MEP pathway uses pyruvate and glyceraldehyde-3-phosphate as starting compounds to form the same end product (IPP) and an additional one, dimethylallyl pyrophosphate (DMAPP), which is an isomer of IPP.

The naturally occurring terpene biosynthetical pathway in *E. coli* is the MEP pathway. This pathway is a relatively recent discovery, with the first proof of its existence being published in the early 90's (reviewed by Rohdich *et al.*, 1999). Before that, for decades it was believed that the mevalonate pathway (MVA) was the only existent route to the production of terpene precursors.

1.6.1.1 Mevalonate Pathway

The mevalonate pathway consists of seven enzymes, starting out of acetyl-CoA from glycolysis. It begins with a reaction catalyzed by a thiolase (Acetoacetyl-CoA thiolase, AACT), where two molecules of acetyl-CoA are condensed into one acetoacetyl-CoA. Then another acetyl-CoA is added to this intermediate to form 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) by the enzyme HMG-CoA synthase (HMGS). HMG-CoA is then converted to (R)-mevalonate (MVA) by the enzyme HMG-CoA-reductase (HMGR), in an NADPH-dependent step. Two phosphorylation reactions, both ATP-dependent, follow this step. First the (R)-MVA is converted to (R)-MVA-5-phosphate (MVAP) by mevalonate kinase (MK), then the enzyme diphospho- mevalonate kinase (PMK) adds another phosphate group to it, yielding (R)-MVA-5-diphosphate (MVAPP). From this last intermediate, IPP is formed by decarboxylation, catalyzed by the enzyme mevalonate diphosphate decarboxylase (PMD). At last, IPP isomerase (IDI) is responsible for forming DMAPP and maintaining equilibrium between the two isomers (Figure 5). The last steps described, from mevalonate to IPP, are carried out in a different way in archaea. In these organisms, mevalonate is also phosphorylated by MVK but then, instead receiving another phosphate group from PMK, it is first converted to isopentenyl phosphate (IP) by 5-phosphomevalonate decarboxylase (PMVD). Further conversion of IP to IPP is catalyzed by the enzyme isopentenyl phosphate kinase (IPK) (Kuzuyama 2002; Lombard and Moreira 2011).
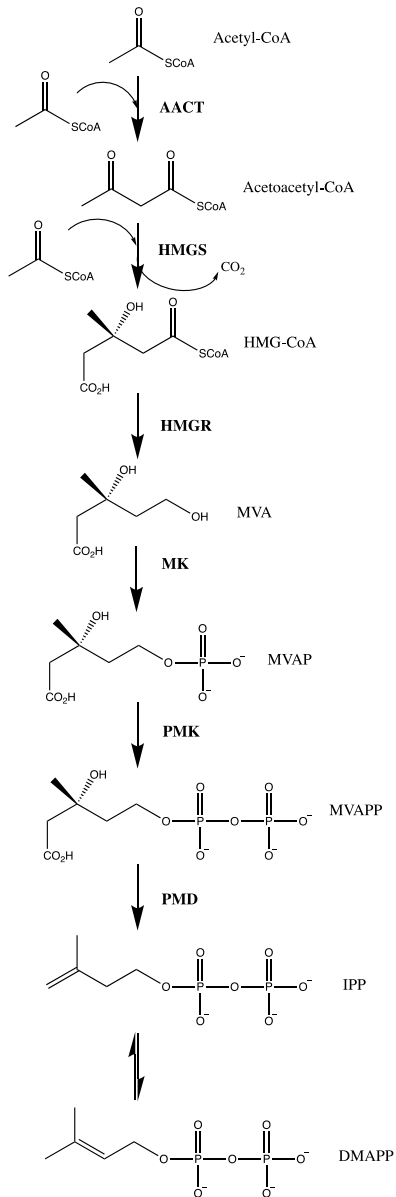
Figure 5. Biosynthesis of IPP and DMAPP via the mevalonate pathway. Acronyms are described in the text.

1.6.1.2   MEP Pathway

The MEP pathway starts out of pyruvate and glyceraldehyde-3-phosphate (G3P), which are condensed forming 1-deoxy-D-xylulose-5-phosphate (DXP). The enzyme catalyzing this reaction is 1-deoxy-D-xylulose-5-phosphate synthase (DXS), which uses the cofactor thiamine pyrophosphate. Then, DXP is converted to 2*C*-methyl-D-erythritol-4-phosphate (MEP) in an NADPH-dependent reaction catalyzed by the enzyme 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR). The following step is the formation of 4-diphosphocytidyl-2*C*-methyl-D-erythritol (CDP-ME) from MEP and cytidyl-triphosphate (CTP), catalyzed by the enzyme 4-diphosphocytidyl-2*C*-methyl-D-erythritol cytidylyltransferase (CMS). Then, the ATP-dependent enzyme 4-diphosphocytidyl-2*C*-methyl-D-erythritol kinase (CMK) catalyzes the formation of 4-diphosphocytidyl-2*C*-methyl-D-erythritol-2-phosphate (CDP-ME2P). This substrate is then used by the enzyme 2C-methyl-D-erythritol-2,4-cyclodiphosphate synthase (MCS) to produce 2*C*-methyl-D-erythritol-2,4-cyclodiphosphate (MECP). On the last steps of the pathway, MECP is reduced to 2-methyl-2-(*E*)-butenyl diphosphate (HMB-PP) by the 1-Hydroxy-2-methyl-2-(*E*)-butenyl-4-diphosphate synthase (HDS). HMB-PP is used by the enzyme 4-hydroxy-3-methyl-2-(*E*)-butenyl-4-diphosphate reductase (HDR), with the help of NADPH, to form IPP and DMAPP (Figure 6). Even though these two precursors are produced, an IPP isomerase (IDI) also is present in most of the organisms that contain the MEP pathway, converting IPP to DMAPP and vice versa. However, while in the MVA pathway IDI is essential for interconversion of IPP and DMAPP allowing the formation of the latter, in the MEP pathway the enzyme IDI is only essential for some organisms under certain conditions, because both compounds are already formed by the enzyme HDR. IDI regulates the IPP/DMAPP ratio, depending on the amounts made available by HDR and the demand for each compound, determined by environmental, physiological and biological conditions (Rohdich et al. 2001; Eisenreich et al. 2004).
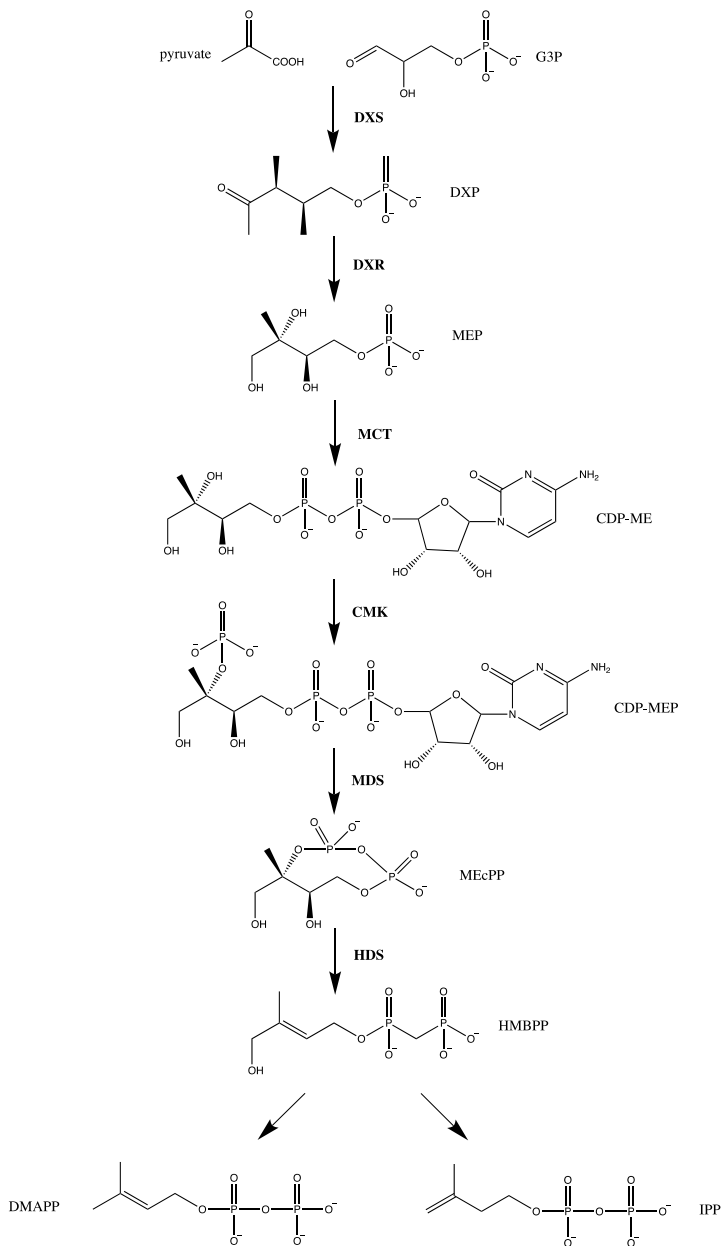
Figure 6. The MEP pathway leading to the formation of IPP and DMAPP.

1.6.2 Terpene biosynthesis

Independently from the organism and the pathways present within the cells, the biosynthesis of terpenes always begins with IPP and its isomer DMAPP. IPP, a nucleophile due to the terminal vinyl group, condensates with the electrophile DMAPP in a stereospecific way. In this reaction, catalyzed by a prenyltransferase, DMAPP has its pyrophosphate ion eliminated, creating a positive charge on C1 of DMAPP. The condensation step takes place with the attack of positively-charged C1 from DMAPP by the C3-C4 double bond from IPP, with stereospecific removal of the pro-R proton at position 2 (Figure 7). From the condensation of IPP and DMAPP, both containing 5 carbon atoms in their structure, geranyl pyrophosphate (GPP) is formed, containing 10 carbon atoms. IPP can then be successively added to prenyl pyrophosphates, forming even larger precursor molecules as depicted in Figure 8, always containing a multiple of five carbon atoms and a pyrophosphate group. These precursor molecules are then used by terpene synthases as substrates to form different types and sizes of terpenes, with varied properties. GPP as a 10-carbon precursor, originates monoterpenes. FPP with 15 carbon atoms, originates sesquiterpenes, GGPP with 20 carbon atoms originates diterpenes, and so on. Triterpenes are the ones with 30 carbon atoms in their structures, for example steroids. Tetraterpenes like carotenoids have 40 carbon atoms in their structure. The further condensations of 5-carbon precursors into longer prenyl diphosphates are always catalyzed by prenyltransferases, also called prenyl diphosphate synthases or named according to the compound that is produced (e.g. GPP synthase, FPP synthase etc.).
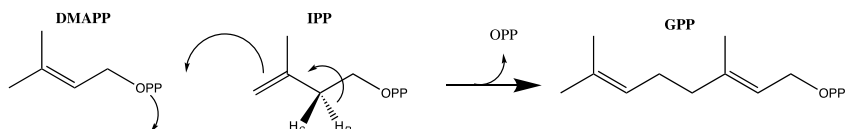
Figure 7. Scheme of DMAPP and IPP condensation leading to GPP formation, catalyzed by a *trans*-prenyltransferase enzyme. The C3-C4 double bond in IPP attacks the positively-charged C1 of DMAPP, and specific abstraction of the pro-R proton leads to the establishment of a double bond in the GPP product.
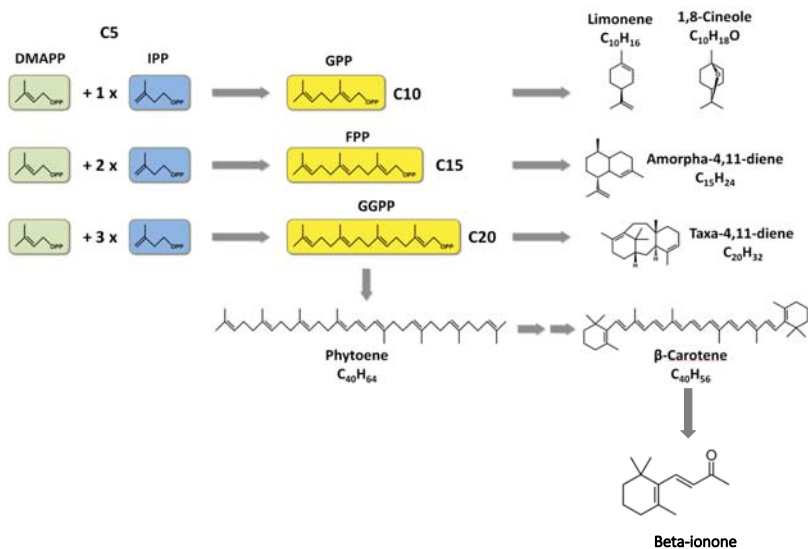
Figure 8. Terpene biosynthesis starting from prenyl diphosphate precursors IPP and DMAPP. Formation of C10, C15 and C20 intermediates is catalyzed by prenyl transferases. GPP, FPP and GGPP are substrates for terpene synthases, and degradation of beta-carotene by cleaving enzymes yields beta-ionone. (Drummond et al. 2019)

Prenyltransferases are enzymes that transfer an allylic prenyl group to the acceptor molecule. This acceptor molecule can be IPP, as it is in the case of terpene biosynthesis, or other molecules such as quinone intermediates or proteins. Regarding the function of these enzymes, they are transferases that transfer a prenyl functional group. Regarding their classification, they are intermolecular transferases (EC 5.4.99). The mechanism for the FPP synthase was studied by Cornforth et al. because of its role in squalene synthesis and cholesterol levels regulation (Cornforth et al. 1966; Popják and Cornforth 1966). Afterwards, many different prenyl transferases were described, and the mechanisms by which they elongate allylic chains of prenyldiphosphates also vary. They can be classified into two major groups, according to the geometry of the products formed: (*E*)- or *trans*- prenyltransferases and (*Z*)- or *cis*- prenyltransferases (C Dale Poulter and Rilling 1978; C D Poulter, Satterwhite, and Rilling 1976; Ito et al. 1987; Kobayashi et al. 1985). For example, the abovementioned condensation reaction of IPP and DMAPP is catalyzed by a *trans*-

20

prenyltransferase, yielding the 10-carbon (*E*)-prenyldiphosphate GPP (Figure 7). The prenyltransferases that catalyze the formation of terpene precursors GPP, FPP and GGPP are all *trans*-prenyltransferases. As example of *cis*-prenyltransferase one can cite the enzyme that catalyzes poly-isoprene formation in rubber trees. The difference in the reactions catalyzed by *trans*- and *cis*-prenyltransferases is the prochirality of the eliminated proton from position 2 of IPP. In a *trans*-prenyltransferase, the eliminated proton is *pro*-R, and for *cis*-prenyltransferases the eliminated proton is *pro*-S. This fundamental difference in catalytic function between *trans*- and *cis*-prenyltransferases reflects differences in their catalytic centers, which can be proven by molecular analysis: the two types of transferases have very different amino acid and nucleotide sequences. In fact, crystal structure of determination shows that 3D structures of cis- and *trans*-prenyltransferases differ completely (S. Takahashi and Koyama 2006; Shimizu, Koyama, and Ogura 1998). *Trans*-prenyltransferases possess aspartate-rich conserved regions involved in catalytic function and substrate binding (T Koyama 1999; K. C. Wang and Ohnuma 2000). In fact, aspartate-rich motifs are a characteristic of other enzymes taking part of isoprenoid biosynthesis, like IPP isomerases and terpene synthases. However, this region is not found in *cis*-prenyltransferases, suggesting a distinct evolutionary origin regarding this protein family (S. Takahashi and Koyama 2006). The specificity of *trans*-prenyltransferases also involves the size of their substrates and products, especially for the linear ones GPP, FPP and GGPP. The length of the chain is controlled by bulky residues located close to the first aspartate-rich motif, and they serve as a "floor", limiting the size of the catalytic pocket and thus the size of its product (K. C. Wang and Ohnuma 2000). In terpene synthesis, after prenyldiphosphates precursors are produced by the either MEP or MVA pathway (in the case of DMAPP) or specific prenyltransferases (in the case of longer prenyl diphosphates), they are further converted by terpene synthases into the final products.

Terpene synthases (TS) are enzymes that have prenyl diphosphates as substrates and yield terpene structures as products. The enormous diversity of terpene structures can be partly attributed to the synthases, which belong to a superfamily of enzymes spread into many taxa. Current reviews on the topic enumerate 12 different subfamilies of terpene synthases but regarding the mechanism they can be divided into two types (Rudolf and Chang 2020; Tholl 2006). TS often have a promiscuous

activity which enables them to produce more than one compound and to accept more than one substrate. While *in vitro* studies show the acceptance of various substrates, the enzyme's reactions within the cell depend on the substrate pool that is available (Tholl 2015). Some terpene synthases, however, are very selective regarding their substrate (Degenhardt, Kollner, and Gershenzon 2009). Most of the terpene synthases described belong to type 1, characterized by its reaction mechanism involving subtraction of the diphosphate moiety, which then ionizes the prenyl chain producing a carbocation intermediate. The reactive ionized carbon chain then undergoes different reactions like hydride shifts, cyclizations, and rearrangements until finally a termination by loss of a proton or a water addition takes place (E. M. Davis and Croteau 2000). Terpene synthases from type 2 do not act on the diphosphate but rather form a tertiary carbocation through protonation of a double bond of an alkene or epoxide. Both types of terpene synthases abovementioned require Mg^{2+} for catalysis and both have mechanisms for carbocation stabilization, rearrangement and quenching. The carbocation rearrangement can include alkene cyclizations, hydride shifts, alkyl shifts (also methyl shifts), ring contractions, ring expansions and proton transfers. The reaction is finally completed with the quenching of the final carbocation, which can happen through deprotonation of a neighbor carbon forming an alkene or hydroxyl capture from the solvent of the molecule itself, forming an alcohol or ether function (E. M. Davis and Croteau 2000). Terpene synthases also possess an aspartate-rich motif, as it was described for prenyltransferases above. The aspartate-rich motif in type 1 terpene synthases is responsible for the binding of the Mg^{2+} cluster for diphosphate abstraction (Christianson 2017).

The products of catalysis by terpene synthases sometimes are subjected to further modifications, made by decorating enzymes, e. g. cytochrome P450 oxygenases, which increase even more the number of different terpene structures (Bathe and Tissier 2019). P450s have a very important role in the production of paclitaxel out of GGPP, catalyzing 8 of the ca. 20 biosynthetical steps. Reactions performed by P450s include epoxidation, hydroxylation, dehydrogenation, isomerization, C-C bond cleavage and even dehalogenation, deamination and oxidation of heteroatom (Sono et al. 1996). Additionally, other decorating enzymes can perform modifications such as acylation, glycosylation, peroxidation and methylation (Chandran, Kealey, and Reeves 2011).

1.6.3 Non-canonical Terpenes

The isoprene rule (Wallach 1887; Ruzicka 1953) states that terpenes are formed by isoprene units and therefore have chemical structures with a number of carbon atoms that is equal to a multiple of five. However, ten years ago three different groups disproved the idea that terpene precursors always contain multiples of five carbon atoms in their structures. During the biosynthesis of the off-flavor molecule 2-methyl-isoborneol, a terpene that contains 11 carbon atoms in its structure, GPP is converted into 2-methyl-GPP by a GPP methyltransferase (GPPMT) (Dickschat et al. 2007; Komatsu et al. 2008; C. M. Wang and Cane 2008). This methyltransferase transfers one methyl group to carbon 2 of its substrate GPP, forming the first non-conventional prenyl pyrophosphate described, the C11 prenylpyrophosphate 2-methyl-GPP (Figure 9). The C11 prenyl pyrophosphate derived from the reaction catalyzed by this GPP methyltransferase is then used by different terpene synthases to produce C11 terpene compounds (Komatsu et al. 2008; Chou, Ikeda, and Cane 2011), in a manner that defies the isoprene rule.
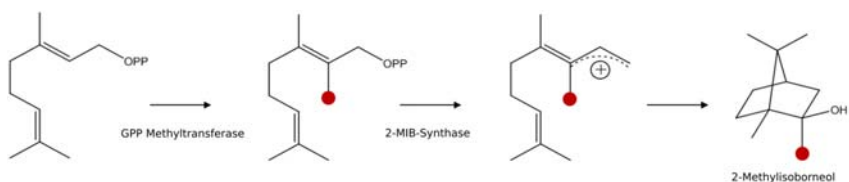
Figure 9. Synthesis of 2-methyl-isoborneol, via methylation of GPP by GPP methyltransferase, generation of 2-methyl-GPP and conversion to 2-MIB by 2-MIB-synthase (modified from Wang and Cane, 2008).

Terpenoids produced by microorganisms, especially volatile compounds, were not that strongly investigated due to the inability to enrich them and obtain quantities sufficient for extraction and identification. More recently, with the possibility of heterologous expression of terpene synthase genes and whole biosynthetic pathways, the largely undiscovered terpenome of microorganisms has also become a subject of research (Cane and Ikeda 2012; Yamada et al. 2015). The world of terpenes was vastly explored in plants, but how terpene biosynthesis works, which products are

available in bacteria, and what roles do they play in their ecology and regulation is still a universe to be unveiled. The first microbial terpenes to be studied were geosmin and 2-methyl-isoborneol, due to the off-flavor they give to water from reservoirs when cyanobacteria populations overgrow (Izaguirre et al. 1982), causing problems in water supply systems. Geosmin is a product of sesquiterpene degradation, 2-methyl-isoborneol is a C11 compound produced by actinomycetes, cyanobacteria and myxobacteria, and these two compounds have a very low detection threshold (Schulz and Dickschat 2007). The fact that these compounds caused very undesirable effects on the smell of the water led to the investigation of their biochemistry and biosynthesis mechanisms in detail.

After the identification of the enzymes participating on the synthesis of 2-methyl-isoborneol and the description of a methyltransferase that uses GPP as substrate for the formation of a C11 prenyldiphosphate, some groups studied the crystal structure of this GPP methyltransferase. Köksal and colleagues elucidated the structure of GPPMT from *Streptomyces coelicolor* in 2012 (Köksal et al. 2012). Ariyawutthiphan and colleagues also published the analysis of the structure of the GPPMT from *Streptomyces lasaliensis* (Ariyawutthiphan et al. 2012; 2011) after identifying the genes involved in 2-methyl-isoborneol formation (Komatsu et al. 2008). In both crystallography studies, the quaternary structure of the crystallized GPPMT protein showed an hexameric geometry, suggesting the formation of a hexamer, as a triad of dimers (210 kDa in total, 35 kDa each monomer). GPPMT is an S-adenosyl-methionine (SAM) dependent methyltransferase with 300 residues. It has a defined core region formed by a seven-stranded beta sheet and five alpha helices, composing a Rossmann-fold (Figure 10) which is typical for SAM-dependent methyltransferases and it is related to the binding of the nucleotide portion of SAM (Kozbial and Mushegian 2005). Besides the Rossmann-fold core, the protein also entails three alpha-helices at the N-terminus, one extra alpha helix after the 5th beta strand and two alpha helices inserted between the 6th and 7th beta strands. The active site entails Mg2+ ions which coordinate to oxygen atoms from the phosphate groups of GPP. The SAM-binding site occurs in the N-terminal region of the beta-sheet, composed by residues from beta-strands 1, 2 and 3. The GPP binding site however has low similarity those of with other methyltransferases, which is expected since the recognition has to be specific for each enzyme's substrate (Martin and McMillan 2002). The SAM-binding and GPP-

binding cavities of the enzyme are continuous, and the binding of GPP to the active site, in the presence of magnesium ion, elicits a conformational change that allows the acceptance of SAM in the proper position. After accommodation of GPP and SAM into the active site, the methyl group from SAM interacts with the alkene chain and forms a C-C bond at C2 position of GPP, generating a carbocation at position C3. The carbocation formed alters the destiny of the carbons posteriorly, when hydride shifts and alkyl transfers take place in reactions catalyzed by terpene synthases (Komatsu et al. 2008; C. M. Wang and Cane 2008; J. Wang et al. 2011; Chou, Ikeda, and Cane 2011). The carbocation is stabilized via cation-π interaction between the positively charged carbon species and an aromatic side chain (F222 in *S. coelicolor* GPPMT), until deprotonation of C2 and quenching of the reactive species (Ariyawutthiphan et al. 2012; Köksal et al. 2012).

The identification of GPPMTs and elucidation of their mechanisms and structures paved the way for new strategies of terpene modification, with the addition of a methyl group in a step before the final processing by terpene synthases.
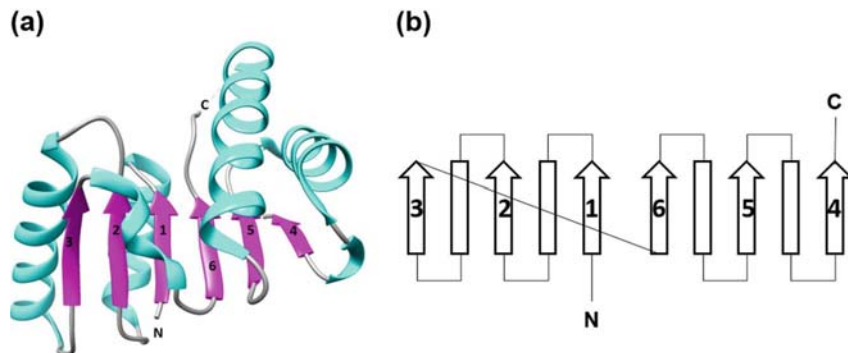
Figure 10. The Rossmann fold. a) Structure showing beta strands numbered 1-6 (purple) and alpha helices (blue). b) Schematics showing sequential organization of beta strands (Shin and Kihara 2019).

1.7   Methyltransferases and cofactor synthesis

Methylation is a modification widely distributed in nature that occurs in many different metabolic pathways. The transformation of molecules through methylation has functions in mediating cell signaling and in the production of particular metabolites. Methylation reactions are most frequently catalyzed by methyltransferases (MT) that use the co-substrate SAM as electron-deficient methyl donor. There are many described methyltransferases with very different functions in primary and secondary metabolism, ranging from pigments and membrane components to signaling molecules and compounds for defense mechanisms (reviewed in Liscombe, Louie and Noel, 2012) . Methyltransferase enzymes also receive a lot of attention for their role on epigenetic control, since they methylate DNA, RNA and also proteins, regulating their activity (Murn and Shi 2017; Biggar and Li 2015; Michalak et al. 2019). The methylation of such molecules, and also of hormones and neurotransmitters, modifies their physico-chemical properties and therefore their biological effects, thus there is interest in knowledge about their mode of action for the development of medicines and therapies, for example for cancer treatment (Goll and Bestor 2005; Kim, Lee, and Lee 2020; Wagner and Jung 2012).

Regarding the official classification system for enzymes, methyltransferases belong to EC 2.1.1.-, which means they are transferases (2.) which transfer one-carbon groups (2.1.) and that this one-carbon group is specifically a methyl group (2.1.1.). There are many superfamilies of methyltransferases recognized according to the protein's fold structure, and here the focus will be on the Class I family. Methyltransferases from Class I are enzymes containing a Rossman fold (a motif that occurs in proteins which bind nucleotides) in their structures, for binding of the cofactor S-Adenosyl-methionine (SAM). The mechanism of reaction of SAM-methyltransferases can be generally described as an S_N2 nucleophilic attack. In the methyl transfer reaction, the sulfur atom from SAM acts as nucleophile, transferring the methyl group to the substrate by creating a substrate-methyl bond and breaking the methyl-SAM bond almost simultaneously (Martin and McMillan 2002). MTs can be also classified according to the atom in the substrate that is targeted for the methyl group transfer. For example, the methyl group can be transferred to an oxygen (O-MT), nitrogen (N-MT) carbon (C-

MT) or sulfur atom (S-MT), among others (those four being the most common ones within this class) (Liscombe, Louie, and Noel 2012).

Another way of separating or classifying methyltransferases is according to their substrate molecule specificity, for example protein, DNA/RNA, lipid or small molecule natural products. There are also methyltransferases that do not depend on SAM for the donation of a methyl group (non-SAM-dependent), but the vast majority of the methyltransferases are SAM-dependent. SAM is a sulfonium molecule and one of the most abundant cofactors in living organisms.
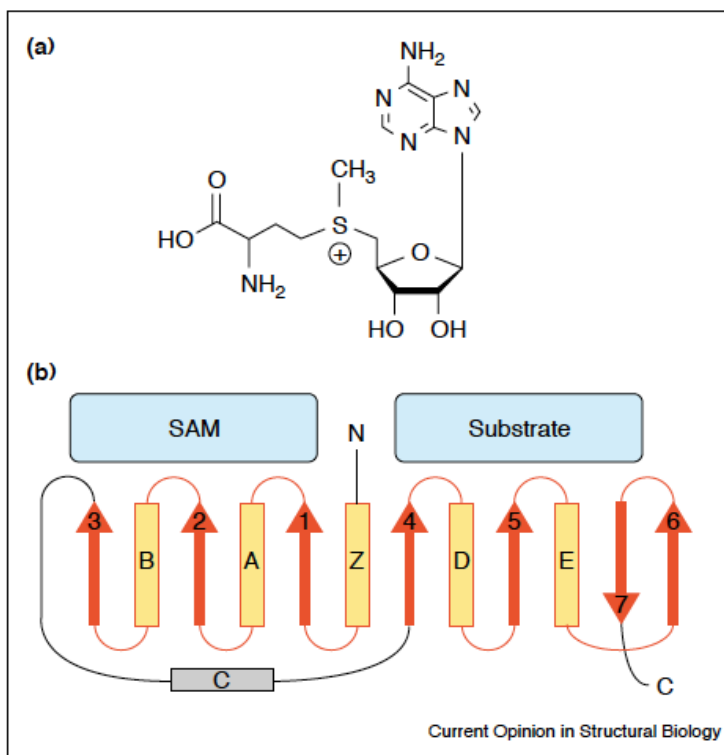
Figure 11. SAM and SAM-MT fold (Martin and McMillan, 2002).

Methyltransferases possess a few conserved domains, which are usually related to the region of the enzyme that binds to the cofactor SAM and occur in enzymes belonging to organisms from diverse kingdoms. However, especially among natural products methyltransferases, the area of the molecule responsible for binding the substrate varies immensely (Martin and McMillan 2002) (Figure 11). Every natural product methyltransferase that had its crystal structure elucidated so far had the same position in the protein for the binding of SAM, even if the SAM-binding residues varied among their sequences. Actually, when comparing the sequences of methyltransferases in general, it was already shown that they could be divided in as many as 15 different categories, when only sequence homology and function are considered (Kozbial and Mushegian 2005). The motif 1 in the structure of MT is usually conserved and contains a glycine-rich region (GxGxG) which is involved with the binding of SAM, specifically to its carboxy-propyl part. The acidic residue present in motif 2 is responsible for binding with the ribosyl portion of the SAM molecule (Kagan and Clarke 1994). The glycine-rich region GxGxG from motif 1 is considered a signature sequence, present in almost all SAM-dependent MTs, and for that reason this motif is frequently used for bioinformatic search of new MTs (Liscombe, Louie, and Noel 2012). Even when one of the glycines is found to have been substituted in a sequence, the substitution it usually for an amino acid also with small side chain (Kozbial and Mushegian 2005; Jansson et al. 2005).

The co-substrate of SAM-dependent methyltransferases, S-Adenosyl methionine, is a cofactor abundant in all cells. It has not only the function of transferring methyl groups to C, N, S and O atoms, but also as a donor of 5'-deoxyadenosyl radicals for isomerizations and other radical-initiated reactions and as a participant in polyamine biosynthesis. SAM is a sulfonium molecule derived from the condensation of methionine and ATP, a reaction catalyzed by the enzyme Methionine Adenosyl Transferase (S-adenosyl methionine synthase, EC 2.5.1.6, MAT). MATs are the only known producers of SAM and since SAM is so essential for reactions in the basic metabolism, MAT has been suggested as a housekeeping gene for analysis on phylogeny, possessing well conserved sequences are across species. The formation of SAM catalyzed by MAT is a two-step process; firstly, in an S_N2 reaction SAM is formed out of methionine and ATP's adenosine moiety; secondly there is the cleavage

of the triphosphate group and release of the products, yielding also pyrophosphate and orthophosphate (Figure 12).
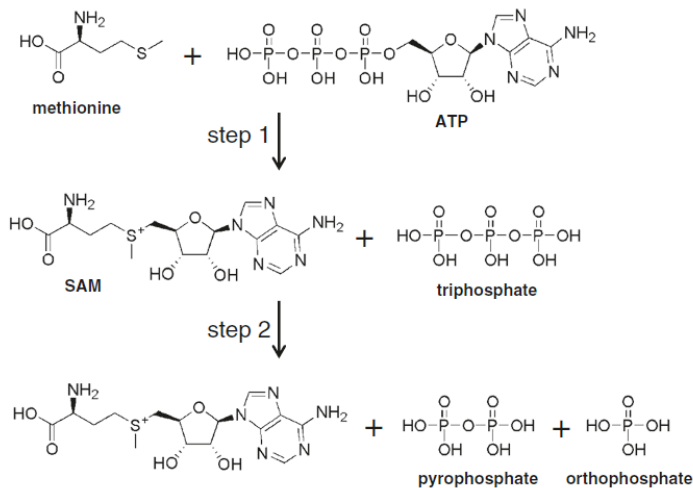
Figure 12. MAT-catalysed reaction. SAM is formed from ATP and methionine (Schlesier et al. 2013).

Methyltransferases are called this way because they take a methyl group contained, in most of the cases, in a methyl group donor like s-adenosyl methionine, and transfer to the substrate, or methyl group acceptor. However, for many of the methyltransferases it is true that the enzyme does not discriminate which group is actually transferred to the substrate. MTs accept also SAM analogues with different alkyl groups. That means that if a SAM analog is offered as donor, with a different group in its structure to be transferred (a group different than methyl), the methyltransferase will transfer to its substrate something different than a methyl group. This has been described before for more than one methyltransferase and several types of groups to be transferred (Siegrist et al. 2017; Singh et al. 2014; Sadler et al. 2017; Sommer-kamann et al. 2017). The MT acceptance of SAM analogues expands the repertoire of groups that can be transferred through catalysis by these enzymes, enabling an even higher structural diversity of their products *in vitro*.

The example given by the biosynthesis of 2-methyl-isoborneol, in which a noncanonical prenyl diphosphate is generated by methylation prior to the terpene synthase catalysis step, offers new possibilities to diversify terpene structures. By modifying terpene precursors with methyl group addition, the biosynthetic pathway incorporates a new module in which new carbocations are generated, influencing key steps on carbon backbone formation.

Inside cells and living organisms, metabolic methylation usually occurs in a very selective way regarding stereochemistry and substrate recognition. Methyl groups have the power of changing the biological effects of determinate molecules in the organism, increase lipophilicity and enhance hydrophobic interactions. The natural methylation systems already existing in nature offer immense possibilities of exploration for the development of methods for small molecule production, and the introduction of methyl groups in molecules has been used as strategy for several research groups involved with the discovery and development of new drugs. A single methyl group is capable of drastically altering the nature, potency and duration of the pharmacological effect. By inserting a methyl group in the right place of a molecule with biological activity, its metabolism can be allosterically blocked and thus increase the molecule's half-life. The specific aspects that can be modified from the addition of a methyl group involve the molecule's conformation, solubility, bioavailability, pharmacokinetics and electronic factors, sometimes increasing the potency of the substance by several orders of magnitude (Schönherr and Cernak 2013; Barreiro, Kümmerle, and Fraga 2011). This effect has been denominated "magic methyl effect", and since then some sectors of research, especially the pharmaceutical industry, have been showing great interest for the selective, enzyme-mediated methylation of goal compounds (Kuntz et al. 2016; Schönherr and Cernak 2013). These MT enzymes are known for their activity on methylating a broad variety of compounds, not only natural but also synthetic, which are required for the production of fine chemicals with applications in the pharmaceutical, agricultural and flavor & fragrance industries, as well as for the production of biofuels (Klimašauskas and Weinhold 2007; Xia Yang et al. 2016). For example, the difference between the alkaloids caffeine, theobromine and theophylline lays in the number and position of methyl groups only, and their effects differ majorly, ranging from strong stimulant (caffeine) to no excitatory activity (theophylline), among other effects. The development of analogues to omeprazol, an

30

important anti-ulcer drug, relied majorly on the redistribution of the molecule's methyl groups (reviewed by Barreiro, Kümmerle and Fraga, 2011). In terpenes, the addition of a methyl group on its precursors inserts a new point of modification for the downstream products, with the potential of creating chemical functions that were not present before.

1.8   Microbial Production of Terpenes

Terpenes are produced not only in plants but also by animals and microbes. Most of the time, they are extracted from plants for human use. However, terpene extraction from plants is – in many cases - not efficient, as large amounts of raw material have to be consumed in order to obtain small amounts of pure compound. Due to their complex structure and variety of isomers, chemical synthesis of these compounds is expensive and demanding, which makes it difficult to separate specific configurations of same size molecules to obtain a pure sample, and thus microbial synthesis appears as the best alternative for production. Microbial production of terpenes using selected enzymes for their conversion brings advantages, not only because it is a cheaper and more sustainable platform for production of a chemical compound, but also because of the achievable selectivity regarding the stereochemistry of these very structurally complex molecules.

The production of terpenes by microbes aiming at market levels is already a reality. Companies like Amyris, Evolva, Firmenich, Cargill and Isobionics (recently acquired by BASF) have been producing and selling flavor and fragrance terpenes for years. The most prominent compounds marketed by these companies include the aroma compounds valencene, nootkatone, sclareol, santalol, and also the compound mixtures agarwood oil and patchouli oil. Farnesene, a drop-in biofuel precursor, is produced in Brazil on a plant built by Amyris and posteriorly acquired by the Dutch company DSM. The method employed in this production plant is based on engineered yeast strains that use glucose from sugar cane to yield a variety of terpenes, including farnesene.

The use of microbes for terpene production depends on the engineering of metabolic pathways, which can be naturally occurring in the organism or heterologously expressed. The specific way by which the engineering is made varies, and many different methodologies have been developed in the last decades, reviewed with detail in previous publications (Kampranis and Makris 2012; Dien 2013; Prather and Martin 2008; Lee, Mattanovich, and Villaverde 2012; Vickers et al. 2014; Kirby et al. 2016; Immethun et al. 2013). In general, after the selection of a suitable host for production, focus is given on protein and genetic engineering to resolve different issues from the pathway itself, like precursor supply, cofactor necessity and reaction competitors, in order to enhance the flux through the pathway. Changing the levels of gene expression may also lead to higher protein and terpene production. In a similar way, engineering the protein can incite interesting effects like higher activity, product and substrate diversity or specificity (depending on the desired result), stereospecificity, among others. Since different reactions take place on terpene biosynthesis in a modular way, the possibilities regarding protein engineering are innumerous, and how they should be modified or screened depends on the goal of the study. For example, since patchouli oil consists in a mixture of compounds, for this product it can be interesting to have a few enzymes with several side products (Deguerry et al. 2006). In the case of stevia, a pure compound, the interest lays on a decreased variety of product formation by the enzyme (Olsson et al. 2016). In general, because terpene synthases naturally exhibit promiscuity they are suitable candidates for protein evolution. Small changes in the structures of terpene synthases reportedly generate novel catalytic properties for terpene synthases (Tholl 2015). Moreover, other factors like intermediate and product toxicity must also be addressed and can be circumvented by approaches like *in situ* product removal and directed evolution towards resilience, for example.

Up to now, the organisms most commonly employed for terpene production are baker's yeast (*Saccharomyces cerevisiae*) and *Escherichia coli*, mainly due to the fact that they are well characterized and easily cultivated in the laboratory (Kampranis and Makris 2012). The search for features that include the possibility of using alternative carbon sources and a natural resistance to product toxicity, among others, pushes research to investigate new microbial hosts. Interesting approaches, like the production of humulene from CO_2 by *Cupriavidus necator*, or engineering

Methylobacterium extorquens to produce alpha-humulene out of methanol, configure as good examples of alternative carbon source together with alternative host (Krieg et al. 2018; Sonntag et al. 2015). In addition to these two species, also *Pseudomonas putida* has been used as host due to favorable characteristics such as high intracellular acetyl-CoA levels (Agrawal, Kotasthane, and Kushwah 2015). However, the majority of studies are still performed in optimized model organisms, either by engineering the pathway native to the host or by heterologous expression of an introduced pathway.

One important factor that can be a setback for microbial biosynthesis of terpenes, if not duly addressed, is the intracellular availability of precursors for terpene production. In a normal cell the pool of precursors is usually very low, and the flux towards the production of these substrates is insufficient for industrial production. Therefore, metabolic engineering strategies need to be developed to increase the pool of precursors for a high yield of final compounds (Q. Wang, Quan, and Xiao 2019).

Terpene biosynthetic pathways are under strict regulation in living organisms, therefore to maximize precursor and final product yields, years of research have been employed on understanding the enzymes that control the fluxes within the metabolism (Hemmerlin, Harwood, and Bach 2012; Banerjee and Sharkey 2014). Regarding the mevalonate pathway, the enzymes IDI, HMGR, HMGS, MVK and PMK are known to be rate-limiting, whereas in the MEP pathway the limitation lies in the enzymes DXS, DXR and HDR (Q. Wang, Quan, and Xiao 2019; Cordoba, Salmi, and León 2009). These rate-limiting steps are the so-called bottlenecks of terpene biosynthesis, one of the biggest setbacks for engineering the biochemical production of terpenoids in microbes. In many cases, the enzyme is inhibited by its downstream intermediates, being shut down or reducing its activity enormously when higher levels of product are achieved. For example, regarding the mevalonate pathway, it has been known since the 80's that the reaction of mevalonate formation, catalyzed by the enzyme HMGR, is a key regulatory step for terpene production. This enzyme was extensively studied also because of its role on regulating cholesterol production, where it has been shown that there is feedback inhibition of HMGR due to accumulation of end-product sterols (Bach 1986). In that sense, efforts have been made to increase the expression of this enzyme in living systems, with the goal of increasing also the final terpene output (Polakowski, Stahl, and Lang 1998; Donald, Hampton, and Fritz 1997). Since then,

many techniques were applied in different organisms on an attempt of overcoming this issue for efficient terpene biosynthesis, mostly by overexpressing this specific enzyme (reviewed by Wang, Quan and Xiao, 2019), but also by changing its cofactor requirements from NADPH to NADH (Chen et al. 2014; Z. Jiang et al. 2015). For the yeast *Saccharomyces cerevisiae* there are numerous examples of engineering the MVA for a better performance and increased precursor supply. For example, a truncated version of HMGR was not inhibited by FPP accumulation (Polakowski, Stahl, and Lang 1998; Donald, Hampton, and Fritz 1997), and the modification of this enzyme led to higher levels of santalene production (Scalcinati, Knuf, et al. 2012; Scalcinati, Partow, et al. 2012). In a similar way, overexpression of a truncated version of HMGR enhanced amorphadiene and taxadiene yields in *S. cerevisiae* (Ro et al. 2006; Engels, Dahm, and Jennewein 2008). For *E. coli* the most common strategy is to introduce the mevalonate pathway completely, a method which overcomes the cell's regulation systems regarding the enzymes generating a precursor pool (Zurbriggen, Kirst, and Melis 2012) and that was used for sclareol production for example (Schalk et al. 2012).

1.9   Motivation

The extraction of natural products from natural sources is in most cases so inefficient that it becomes unfeasible or extremely rare and costly, like ambergris from whales (the price per kilo goes way over expensive goods like gold and truffles). Also, variations on the seasonal availability of natural sources for some of the aromatic terpenes has caused shortages and steep rises on price, like the popular patchouli and sandalwood oils. In an era of global natural resources ramping scarcities, and face to the structural complexity of terpene compounds, microbial synthesis associated with synthetic biology principles come in hand for a sustainable, green and cheap production of fine chemicals like flavors, fragrances, and remedies. By genetic engineering of microbes, it is possible to express complete pathways and key enzymes for the production of specific compounds or blends, or even engineer the genome of the host for enhanced production of precursors, for example.

Terpenes are the largest group of natural compounds and their applications in fragrance and pharmacy industries are well established, ranging from flavor ingredients to anti-cancer drugs. Terpene synthesis follows a very specific modular concept, with the addition of successive 5 carbon units to a prenyl chain. This modular characteristic offers a limitation that can be overcome when the very building blocks are modified. This work starts from the idea that if different molecules, other than the canonical building blocks, can be used as starting material for the terpene biosynthetical pathway, the structural diversity of terpenoids can be greatly expanded. With the diversification of the structural universe of terpene molecules, the new non-conventional molecules can add up to the discovery of new biological functions for the pharma, flavor and fragrance industries. The recent discovery of the existence of methyltransferases that modify the prenyl pyrophosphate precursors opens an avenue for exploration of the promising sphere of non-canonical terpenes.

Following up on the use of methyltransferases to expand the possibilities of terpenes structures through the synthesis of non-canonical terpenes, and on the lookout for enzymes capable of generating new building blocks, a search on the genetic database was made to identify putative enzymes involved with the modification of prenyl diphosphates. The first the aim of this work was to explore the methyltransferase

collection in the microbial genome database, on the search for other prenyl pyrophosphate methyltransferases with similar activities to the above-mentioned GPP methyltransferase.

Towards new methods of increasing the pool of possible non-conventional terpene precursors this work aimed at manipulating the MAT enzyme reaction to produce SAM analogs and use them as substrate for the production of different compounds. The use of SAM analogs together with methyltransferases could lead to differential alkylation of substrates, which in turn expands even more the possibilities of chemical structures.

The intrinsic substrate promiscuity of terpene synthases and transferases has been already exploited to expand their product spectrum by the use of non-natural, chemically synthesized prenyl pyrophosphate substrates (Miller, Yu, and Allemann 2007; Faraldos et al. 2007; Vedula et al. 2007; Subramanian et al. 2012; Rising et al. 2015; Demiray et al. 2017). A fluorinated synthetic FPP was used as substrate for 5-epi-aristolochene synthase *in vitro*, yielding fluorinated product and also altering the reactivity of the enzyme's catalytical center (Faraldos et al. 2007). Several FPP analogues were shown to be accepted by a farnesyl transferase involved in protein prenylation, in a study searching for enzymatic reaction inhibitors (Subramanian et al. 2012). When focusing on prenyl transferases, studies have been made with prenyl pyrophosphates analogues showing high promiscuity of these enzymes towards a variety of uncommon substrates and products (Tanetoshi Koyama et al. 1980; Nagaki et al. 2012). The use of synthetic isotopic-labeled substrates also revealed promiscuity of the isopentenyl pyrophosphate isomerase, besides showing new catalytic effects and the emergence of new products and irreversible reactions (Tanetoshi Koyama, Katsuki, and Ogura 1983). A much broader synthetic capability of the terpenoid biosynthesis pathway can be explored by taking advantage of the demonstrated high substrate promiscuity of prenyl transferases, isomerases and terpene synthases. Making use of the proven promiscuity of the enzymes that compose the terpene biosynthetical pathway, it is hypothesized that the inclusion of methyl groups into early precursors generates final products with additional methyl groups, due to the processing by natural enzymes downstream of the pathway without discrimination between canonical and non-canonical substrates. This can potentially open a new field in which strategies can be developed to create several versions of terpene precursors,

and then introduce them into the regular terpene biosynthetical pathway for the generation of new molecules. In this sense, the production of unconventional terpenes out of methylated precursors should be tested on *in vitro* conditions using different terpene synthases. The hypothesis to be tested was whether plant terpene synthases, which usually do not encounter non-canonical substrates, would be able to transform them into uncommon terpenes with new structures.

2 Materials and Methods

2.1 Materials

2.1.1 Chemicals

The compounds 5-methyl-isoprenol, 4-methyl-prenol, 4,4-dimethyl-prenol, 4,4-dimethyl-isoprenol, 4-methyl-geraniol, 8-methyl-geraniol, 4-methyl-farnesol, 6-methyl-farnesol, 2,4-dimethyl-prenol, 2-methyl-isoprenol, and 2-methyl-prenol were custom synthesized by Enamine (Latvia). Compound identity was confirmed with NMR analysis performed by the manufacturer upon synthesis. Compounds (*E*)-4-methyl-isoprenol and (*Z*)-4-methyl-isoprenol were custom synthesized by Akos (Germany) and had the identity confirmed by NMR analysis performed at the University of Frankfurt by Yi-Ming Shi (^{1}H, ^{13}C, DEPT, HSQC, HMBC, ^{1}H-^{1}H COSY and ROESY NMR in DMSO-d_6). 2-methyl-geranylpyrophosphate was synthesized in the University of Bonn and provided by Prof. Jeroen Dickschat. All the other chemicals were purchased from Carl Roth GmbH or Sigma-Aldrich (Table 1).

Table 1. Chemicals used and their manufacturers

Chemical	Manufacturer
5-methyl-isoprenol	Enamine
4-methyl-prenol	Enamine
4,4-dimethyl-prenol	Enamine
4,4-dimethyl-isoprenol	Enamine
4-methyl-geraniol	Enamine
8-methyl-geraniol	Enamine
4-methyl-farnesol	Enamine
6-methyl-farnesol	Enamine
(*E*)-4-methyl-isoprenol	Akos
(*Z*)-4-methyl-isoprenol	Akos
2,4-dimethyl-prenol	Enamine
2-methyl-isoprenol	Enamine

2-methyl-prenol	Enamine
2-methyl-geranylpyrophosphate (2-m-GPP)	Provided by Prof. Dickschat (University of Bonn)
Acid Phosphatase	Sigma-Aldrich
Ampicilin	Carl Roth GmbH
DNAse	Sigma-Aldrich
DTT	Sigma-Aldrich
Glycerol	Carl Roth GmbH
GPP (geranyl pyrophosphate)	Sigma-Aldrich
HEPES	Carl Roth GmbH
Imidazol	Carl Roth GmbH
IPP (isopentenyl pyrophosphate)	Sigma-Aldrich
Kanamycin	Carl Roth GmbH
$MgCl_2$	Carl Roth GmbH
Na_2HPO_4	Carl Roth GmbH
NaCL	Carl Roth GmbH
NaH_2PO_4	Carl Roth GmbH
PIPES	Carl Roth GmbH
SigmaFAST Protease Inhibitor cocktail tablet	Sigma-Aldrich
RNAse	Sigma-Aldrich
S-adenosyl-methionine	Sigma-Aldrich
Tris	Carl Roth GmbH
Tryptone	Carl Roth GmbH
Yeast extract	Carl Roth GmbH

2.1.2 Media

Table 2. Media used for cell cultures

Medium	Composition
LB	10 g/L Tryptone 5 g/L Yeast extract 10 g/L NaCl
2x YT	16 g/L Tryptone 10 g/L Yeast extract

	5 g/L NaCl pH 7.0
GYM	4.0 g/L Glucose, 4.0 g/L Yeast extract, 10 g/L Malt extract, pH 7.2
TB	12 g/L Tryptone 24 g/L Yeast extract 4 mL/L Glycerol 0.17 M KH_2PO_4, 0.072 M KH_2PO_4

2.1.3 Buffers and solutions

Table 3. Buffers used for assays and electrophoresis

Buffer	Composition
Lysis buffer	50 mM $NaH_2PO_4 * 2H_2O$ 300 mM NaCl 10 mM Imidazol 5 mM DTT pH 7.4
Wash buffer	50 mM $NaH_2PO_4 * 2H_2O$ 300 mM NaCl 30 mM Imidazol 5 mM DTT pH 7.4
Elution buffer	50 mM $NaH_2PO_4 * 2H_2O$ 300 mM NaCl 250 mM Imidazol 5 mM DTT pH 7.4
PIPES buffer	50 mM PIPES 100 mM NaCl 15 mM $MgCl_2$ 5 mM DTT 20 % Glycerol pH 6.7
TRIS buffer	50 mM Tris 5 mM DTT 100 mM NaCl 20 % Glycerol pH 8.0
HEPES buffer	50 mM HEPES 10 mM $MgCl_2 * 6H_2O$ 20 mM $MnCl_2$ 100 mM KCl 1 mM DTT

	10 % Glycerol
	pH 7.0
Acetate buffer	0.07 M Sodium Acetate
	0.03 M Acetic Acid
	pH 5.0
SDS electrophoresis buffer (10x)	30.3 g/L Tris
	144.2 g/L Glycin
	10 g/L SDS
SDS staining buffer	136 g/L Ammonium sulfate
	24 mL/L Phosphoric acid (85 %)
	272 mL/L Methanol
	0.4 g/L Coomassie blue G 250
SDS de-staining buffer	45% (v/v) Ethanol
	10 % (v/v) Acetic acid
TAE buffer	1 mM EDTA
	20 mM Acetic Acid
	40 mM Tris

2.2 Microorganisms, expression plasmids and primers

2.2.1 Microorganisms

Streptomyces monomycini wild type (Gause and Terekhova 1986 emend. Nouioui et al. 2018, type strain (Nouioui et al. 2018)) was obtained from Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) and cultivated for analysis of volatile compounds.

Different strains of *E. coli* were employed according to the type of experiment to be performed. The strains and their specific genomic notations are described on Table 4. *E. coli* BL21 (DE3) was used primarily for protein expression with T7 plasmids. *E. coli* DH5 alpha was used routinely for transformation of plasmids into the cells. *E. coli* MG1655 was used when transforming plasmids for *de novo* production of terpenes.

Table 4. Bacterial strains used in this study

Name	Description	Source
Streptomyces monomycini	Gause and Terekhova 1986 emend. Nouioui et al. 2018, (Nouioui et al. 2018) type strain DSM 41801	DSMZ

E. coli DH5α	F-, *Φ80dlacZΔM15*, Δ(*lacZYA-argF*)*U169*, *deoR*, *recA1*, *endA1,hsdR17(rK-mK+)*, *phoA*, *supE44*, λ-, *thi-1*	ATCC
E. coli MG1655 (DE3)	F-, λ-, ilvG-, *rfb-50*, *rph-1*, Δ*endA*, Δ*recA*	DSMZ
E. coli BL21 (DE3)	F– *ompT gal dcm lon hsdSB(rB–mB–)* λ(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*+]K-12(λS)	Novagen

2.2.2 Plasmids

The constructions of plasmids were done via Gibson assembly (Gibson et al. 2009) or standard restriction-cloning. Ribosomal binding sites (RBS) were designed with the RBS calculator v1.1 (Espah Borujeni, Channarasappa, and Salis 2014; Salis, Mirsky, and Voigt 2009). All plasmids used are listed in Table 5.

pMK-17 (map in Appendix) including the gene of the IPPMT (accession number WP_033037353) and a new RBS with the sequence TAAGATTAAATAAGGAGGTTACCA and a maximal translation initiation rate (TIR) of 399634 was constructed via Gibson assembly. The insert was amplified from the genomic DNA of *S. monomycini* (DSM No. 41801) via PCR with the primers mk34 and mk36. The backbone was amplified from pETDuet-1 via PCR with the primers mk32 and mk33.

For the construction of pLD-03 (map in Appendix) the antibiotic resistance gene of pJBEIc-6409 (map in Appendix) was exchanged to a kanamycin resistance gene via Gibson Assembly. The insert was amplified via PCR from pET28a(+) with the primers mk42 and mk44. The backbone was amplified via PCR from pJBEI-6409 with the primers mk20 and mk23. To delete the genes of the limonene-synthase and GPP-synthase from the assembled product, a PCR with the primers mk8 and mk31, a restriction digest with BamHI and self-ligation was done.

For the construction of pMK-18 (map in Appendix), the sequence of the insert containing the three genes was amplified from the genomic DNA of *S. monomycini* (DSM No. 41801) via PCR with the primers mk34 and mk35. The backbone was amplified from pETDuet-1 via PCR with the primers mk32 and mk33.

For the construction of pLD-04, the terpene synthase gene was deleted from the operon in pMK-18 via PCR with the primers mk48 and mk32. That enabled the amplification of the sequences of genes IPPMT and prenyltransferase genes. The PCR product was digested with *Eco*RI and then ligated with Quick Ligase.

For construction of pMK-03, the sequence of GPP-MT gene (sequence ID NC_003888.3) from *S. coelicolor* genomic DNA was amplified with primers mk12 and mk13. The backbone pETDuet-1 was amplified with primers mk10 and mk11 and then the products were combined via Gibson assembly.

For construction of pMK-23 the gene of GPP-MT (sequence ID NC_003888.3) was amplified via PCR with the primers mk61 and mk60, and pMK-RQ-Opc (GeneArt) as template. PCR product and the vector pET28a(+) were digested with *Bam*HI and *Nde*I. Products of digestion were then ligated.

For construction of pMK-24 the gene of the IPPMT (accession number WP_033037353) was amplified via PCR with the primers mk52 and mk53 and pMK-17 as template. PCR-product and the vector pET28a(+) were digested with *Nde*I and *Bam*HI. Products of digestion were then ligated.

For construction of pMK-22 the gene of *S. monomycini* terpene synthase (accession number WP_030019073.1) was amplified via PCR with the primers mk56 and mk57 and pMK-17 as template. PCR-product and the vector pET28a(+) were digested with *Nde*I and *Bam*HI. Products of digestion were then ligated.

For construction of pMK-25 the gene of *S. monomycini* prenyltransferase (accession number WP_030019072.1) was amplified via PCR with the primers mk54 and mk55 and pMK-17 as template. PCR-product and the vector pET28a(+) were digested with *Nde*I and *Bam*HI. Products of digestion were then ligated.

For construction of pLD-05, first the fragments containing the genes of IPPMT and GPP-MT were obtained from plasmids pMK-03 and pMK-18. For that, both plasmids were digested with *Xba*I and *Eco*RI-HF and had their fragments separated by gel electrophoresis. The fragments were extracted and ligated.

For the construction of pLD-06, the nucleotides encoding tryptophan at position 194 of the IPPMT were exchanged for the ones encoding glycine. Custom synthesis of a plasmid containing the base-exchanged gene was synthesized (GeneArt), using pET28 as vector.

For the construction of pLD-07, the nucleotides encoding glycine at position 202 of the GPP-MT were exchanged for the ones encoding tryptophan. Custom synthesis of a plasmid containing the base-exchanged gene was synthesized (GeneArt), using pET28 as vector.

Table 5. Plasmids used

Name	Description (origin of replication, antibiotic marker, promoter and genes)	Expressed proteins	Reference
pET28a (+)	pBR322, Kan^r, P$_{T7lac}$	-	Novagen No. 69864
pET28b (+)	pBR322, Kan^r, P$_{T7lac}$	-	Novagen No. 69865
pET29a	pBR322, Kan^r, P$_{T7lac}$	-	Novagen No. 69871
pETDuet-1	colE1, Amp^r, P$_{T7}$	-	Novagen No. 71146
pMK-RQ-Opc	colE1, Kan^r, P$_{T7}$, GPPMT, MIBsc	for *E. coli* codon optimized genes of GPPMT (NP_631739.1) and 2-Methyl-isoborneol-synthase from *Streptomyces coelicolor* with designed rbs	This study
pJBEI-6409	p15A, Cm^r, P$_{lacUV5}$, P$_{trc}$ *atoB, mvaS, mvaA, ERG12, ERG18, MVD1, idi, trGPPS, trLS*	mevalonate pathway proteins, IPP-isomerase, GPP-synthase, limonene synthase	(Alonso-Gutierrez et al. 2013)
pLD-01	pJBEI-6409 with Kan^r instead of Cam^r	mevalonate pathway proteins, IPP-isomerase, GPP-synthase, limonene synthase	This study
pLD-03	pLD-01, trLS and trGPPS deleted	mevalonate pathway proteins, IPP-isomerase	This study

pLD-04	pETDuet-1, IPPMT, pt	IPPMT (accession number WP_033037353), PT (accession number WP_03001907.1)	This study
pLD-06	pET28a (+), nucleotide-exchanged (526...528) IPPMT	IPPMT W194G	This study
pLD-07	pET28a (+), base-exchanged (477...479) GPPMT	GPPMT G202W	This study
pMK-17	pETDuet-1, IPPMT	IPPMT from *S. monomycini*, (accession number WP_033037353)	This study
pMK-18	pETDuet-1, IPPMT, pt, ts	IPPMT operon from *S. monomycini*, (MT, PT and TS; accession numbers WP_033037353, WP_030019072.1 and WP_030019073.1)	This study
pMK-22	pET28a (+), ts, N-6xHis	Terpene synthase from *S. monomycini*, (accession number WP_030019073.1), His-Tag	This study
pMK-23	pET28a (+), GPPMT, N-6xHis	GPPMT from *S. coelicolor*, (accession number NP_631739.1), His-Tag	This study
pMK-24	pET28a (+), IPPMT, N-6xHis	IPPMT from *S. monomycini*, (accession number WP_033037353), His-Tag	This study
pMK-25	pET28a (+), pt, N-6xHis	Prenyltransferase from *S. monomycini*, (accession number WP_030019072.1), His-Tag,	This study
pET28b-Gw74M8FL	pET28b, VvGwaPhe, N-6xHis	(+)-α-Phellandrene synthase from *Vitis vinifera* (accession number HM807382), His-Tag	Martin et al. 2010
pET28b-PN20M4FL	pET28b, VvPNaPin1, N-6xHis	(+)-α-Pinene synthase from *Vitis vinifera* (accession number HM807383), His-Tag	Martin et al. 2010
pET28b-O05J11F	pET28b, VvCSbOciM, N-6xHis	Ocymene/Myrcene synthase from *Vitis vinifera* (accession number HM807387, His-Tag	Martin et al. 2010
pET28b-PN09M12	pET28b, VvPNRLin, N-6xHis	(3*R*)-Linalool synthase from *Vitis vinifera*	Martin et al. 2010

		(accession number HM807390), His-Tag	
pET28b-PN25M6FL-1	pET28b, VvPNLinNer1, N-6xHis	(3S)-Linalool/(E)-Nerolidol synthase from *Vitis vinifera* (accession number HM807391), His-Tag	Martin et al. 2010
pET28b-Gw63RR	pET28b, VvGwGer, N-6xHis	Geraniol synthase from *Vitis vinifera* (accession number HM807398), His-Tag	Martin et al. 2010
pET28b-HQ336798.1	pET28b, PsTPS-3Car, N-6xHis	Carene synthase from *Picea sitchensis* (accession number HQ336799), His-Tag	Hall et al. 2011
pET29a-LaLIMS	pET29a, LaLIMS, N-6xHis	Limonene synthase from *Lavandula angustifolia* (accession number DQ263740), His-Tag	Landman et al. 2007
pET28a-MAT	pET28a, TkMAT	Methionine adenosyl transferase from *Thermococcus kodakarensis*	Schlesier et al. 2013
pET28a-MTAN	pET28a, EcMTAN	Methylthioadenosine/SAH nucleosidase from *E. coli*	Siegrist et al. 2015

2.2.3 Primers

Table 6. Primers used (overhangs in italic characters)

Name	Sequence
mk8	CCTAAGGATCCAAACTCGAGTAA
mk10	TTAGTGTCGTCTTAACTTTAGACTAGTCCCGGGTGGGGAATTGTTACCGCTC
mk11	ACCCTCGAGTCTGGTAAAGA
mk12	*AGTCTAAAGTTAAGACGACACTAAGGAGGATATAG*ATGACCACCGAAACCACCAC
mk13	*TCTTTACCAGACTCGAGGG*TTTACACACGATCTGCGGCAA
mk20	GAGCTCGCTTGGACTCC
mk23	*TTAATTAA*GACGTCGGTGCCTAATGAGT
mk31	*TATAGGATCC*GTGCAGTCGGCGAA
mk32	AGGATCCGAATTCGAGCTC
mk33	*TGGTAACCTCCTTATTTAATCTTA*TCTAGAGGGGAATTGTTATCC

mk34	*TAACAATTCCCCTCTAGATAAGATTAA ATAAGGAGGTTACCA*ATGTCCTCTGA GCCAACTGCTG
mk35	*AGCTCGAATTCGGATCCT*TCATGGCT GGGACGGGCT
mk36	*AGCTCGAATTCGGATCCT*TCAGTGGG CGGCGGCGCC
mk42	*ACAGGAGTCCAAGCGAGCTC*AGGTG GCACTTTTCGGGG
mk44	*GGCACCGACGTCTTAATTAAA*ACAAT AAAACTGTCTGCTTACATAAACAG
mk48	ATGAATTC*TGTCAGGCCCGCCTTC*
mk52	AGTCCATATGTCCTCTGAGCCAACTG
mk53	AATTCGGATCCTTCAGTGGG
mk54	AGTCCAT*ATGACCACACCCGATGCG*
mk55	*AGCT*GGATCC*ATGTCAGGCCCGCCT TC*
mk56	*AGCT*CAT*ATGCCCCGGAACATCCCA*
mk57	ATTCGGATCCTTCATGGCTG
mk60	ACGTCATATGACCACCGAAACCACCA C
mk61	AGCT*GGATCC*TTACACACGATCTGCG GCA

2.3 Equipment and Kits

Table 7. Equipment used in the study

Name	Description	Supplier
GC-MS 2	GCMS-QP2010 SE	Shimadzu (Kyoto, Japan)
GC-MS 3	GCMS-QP2020 Autosampler AOC-6000	Shimadzu (Kyoto, Japan)
LC-MS	LCMS-8040	Shimadzu (Kyoto, Japan)
SPME fiber A, manual injection, used in GC-MS 2	SPME fiber assembly 85 µm, Carboxen/Polydimethyl-siloxane (CAR/PDMS), Phase: Polydimethylsiloxane (PDMS)	Supelco (Bellefonte, Pennsylvania, USA)

SPME fiber B, automatic injection, used in GC-MS 3	SPME fiber assembly 30 µm, Polydimethylsiloxane (FIB-P-30/10, PAL systems)	CTC analytics (Zwingen, Switzerland)
GERSTEL SBSE Twister	10 x 0.5 mm Polydimethylsiloxane (PDMS)	Gerstel (Mühlheim, Germany)
Centrifuge	Centrifuge 5415 R	Eppendorf AG (Hamburg, Germany)
Incubator	HT Ecotron, HT Minitron	Infors (Bottmingen/Basel, Switzerland)
PCR Thermocycler	T3000 Thermocycler	Biometra GmbH (Göttingen, Germany)
Centrifuge	Centrifuge 5415 R	Eppendorf AG (Hamburg, Germany)
Transilluminator	BioDoc Analyze Ti5	Biometra (Göttingen, Germany)
Spectrophotometer	NanoDrop 2000c	Thermo Fisher Scientific (Wilmington, Delaware, USA)
TECAN Plate reader	TECAN Infinite M200	TECAN (Männedorf, Switzerland)

Table 8. Kits used in the study

Kit / enzyme	Manufacturer	Use
GeneJET Plasmid Miniprep Kit	Thermo Fisher Scientific (Waltham, USA)	Plasmid DNA extraction
DNA Clean & Concentrator	ZYMO Research	DNA purification up to 5 µg
Pierce BCA Protein Assay kit	Thermo Scientific	Protein detection and quantitation
His-Pur™ Ni-NTA Spin Columns, 3mL resin bed	Thermo Scientific	Protein purification
SAM510: SAM Methyltransferase Assay	G-Biosciences	Monitor SAM-dependent methyltransferase activity
Slide-a-Lyzer Dialysis Cassetes 3-12mL	Thermo Scientific	Protein buffer dialysis
Pre-diluted protein assay standards: Bovine Serum Albumine (BSA) set	Thermo Scientific	Protein quantification
Gel DNA extraction kit	Macherey Nagel	DNA purification from gel
Mini-PROTEAN TGX precast gels	BioRad	SDS-PAGE

Q5 High Fidelity Polymerase	New England Biolabs	Polymerase Chain Reaction (PCR)
T4 DNA Ligase	New England Biolabs	Ligation

2.4 Software and Databases

Cloning was planned with the help of SnapGene software. GC-MS data were analyzed with GCSolution (Shimadzu). LC-MS data were analyzed with LabSolutions (Shimazu). For compound identification, NIST14, NIST14s, NIST17, NIST17s libraries were used, as well as comparison with mass spectra obtained from measurements using reference compounds, stored in in-house constructed unconventional terpene libraries.

2.5 Molecular Biology Methods

2.5.1 PCR

Polymerase chain reactions were performed using the abovementioned DNA plasmids/fragments at 1 ng/μL as template and Q5 High Fidelity Polymerase. The mixture also contained dNTPs, reverse and forward primers and Q5 buffer in the concentrations recommended by the manufacturer.

Thermocycling conditions were 98 °C for 30 seconds, then 35 cycles of the following: 98 °C for 10 seconds, 30 seconds at the annealing temperature specific for the pair of primers used, and 72 °C at a time defined by the length of the fragment to be produced (30 sec/kb); followed by a final extension phase of 72 °C for 2 minutes.

DNA fragments were amplified through polymerase chain reaction using Q5 High-fidelity polymerase.

2.5.2 Restriction Digest

Restriction digests were performed using specific restriction enzymes from NEB (New England Biolabs) according to the region to be cleaved, either to check the identity of DNA fragments and plasmids or to create sticky ends to help on ligation reactions.

The mixtures contained one or two enzymes, a buffer specific for the enzymes, and the DNA to be digested. Time and temperature of incubation were selected according to the manufacturer's instructions.

2.5.3 DNA ligation

Ligation was performed using T4 DNA Ligase (NEB). The mixture contained the DNA fragments to be ligated, ligase at 1 μL/20 μL and ligase buffer. Reaction conditions were set following the manufacturer's instructions.

2.5.4 Transformation

Transformation of plasmid DNA into *E. coli* was made using protocol provided by manufacturer of chemically competent DH5 alpha cells (New England Biolabs).

2.5.5 Plasmid DNA extraction

A pre-culture containing 5 mL LB medium with specific antibiotics and cells scraped from cryo-cultures was incubated overnight at 37 °C. On the next day, cells were collected by centrifugation at 8000 g for 30 minutes and plasmid extraction was done using GeneJET Plasmid Miniprep Kit (Thermo Fisher), according to the manufacturer's indications.

2.5.6 Agarose gel electrophoresis

DNA samples had their sizes confirmed via electrophoresis in agarose gels. For that, TAE buffer solution containing 1 % agarose was prepared and the gel was cast into appropriate chambers. The DNA samples were mixed with loading buffer and loaded into the gel wells. The runs were made under 100 V current in TAE buffer.

2.6   Biochemical methods

2.6.1 Protein expression, extraction and purification

A pre-culture containing 5 mL LB medium with specific antibiotics and cells scraped from cryo-cultures was incubated overnight at 37 °C. LB pre-cultures were used to inoculate 400 ml TB medium main cultures. After cultivation for 2 h at 37 °C (OD_{600} of 0.6), IPTG at 100 µM was added for induction. After cultivation at 18 °C, 110 rpm

overnight, cells were harvested by centrifugation for 20 minutes at 4000 g and 4 °C and then re-suspended in 15 ml lysis buffer. Ultrasound sonication was performed on ice with 20 % amplitude (0.5 seconds pulse, 1 second pause, 4 minutes pulse time), followed by digestion with RNAse (final concentration 10 mg/ml) and DNAse (final concentration 5 mg/ml) on ice for 15 min. After centrifugation for 20 minutes at 4000 g and 4 °C, enzyme purification took place using His-Pur Ni-NTA Spin Columns (Thermo Fischer Scientific) according to the manufacturer's recommendations using buffers described above. Purified protein size was confirmed through SDS-Page.

2.6.2 Protein quantification through BCA test

Extracted and purified proteins were quantified with the use of BCA test according to the manufacturer's instructions. Bovine serum albumin was used as standard and the samples were measured with TECAN microtiter plate reader.

2.6.3 Polyacrylamide Gel Electrophoresis

Protein samples were separated and analyzed via electrophoresis in polyacrylamide gel. Samples were mixed with SDS-PAGE loading buffer and inserted at 20 µL per well in precast gels. The runs were made under 200 V current on SDS-gel buffer.

2.7   *In vitro* assays

2.7.1 *In vitro* methyltransferase incubations

Purified protein (IPPMT or GPPMT) at a final concentration of 25 µM was added to 1 mL PIPES buffer containing 60 µM of IPP, GPP, FPP or DMAPP and 120 µM SAM. The enzymatic reaction was incubated at 30 °C for 20 h or 40 h. Then 1 ml phosphatase solution (acid phosphatase 7.5 mg/ml in 0.1 M sodium acetate, pH 5.0) was added and the reaction was incubated at 30 °C for 2 h. Following the reaction, volatiles in the headspace were extracted with SPME fiber for 10 minutes at 40 °C.

2.7.2 *In vitro* TKMAT/MTAN incubations

Purified TKMAT at 0.5 mg/mL (11 µM) and purified MTAN at 0.01 mg/mL (3.8 µM) were added to 1 mL TRIS buffer containing ATP at 240 µM and methionine, ^{13}C-methionine or ethionine at 240 µM. The enzymatic reaction was incubated at 30 ℃ for 20 h. 100 µL of each sample was collected and frozen for LC-MS analysis.

2.7.3 *In vitro* TKMAT/MTAN-MT incubations

Purified TKMAT at 0.5 mg/mL (11 µM) and purified MTAN at 0.01 mg/mL (3.8 µM) were added to 1 mL TRIS buffer containing ATP at 240 µM and methionine, ^{13}C-methionine or ethionine at 240µM. The enzymatic reaction was incubated at 30 ℃ for 20 h. Purified methyltransferase at 25 µM (IPPMTor GPPMT) and 60 µM IPP or GPP were added to the mixture. The enzymatic reaction was incubated at 30 ℃ for another 20 h. The volatiles in the headspace were extracted with an SPME fiber for 20 minutes at 40 ℃.

2.7.4 *In vitro* incubations with plant terpene synthases

To verify the activity of terpene synthases from plants towards unconventional prenyl pyrophosphates, eight different enzymes from plants were selected, of which   six from *Vitis vinifera* (wine plant), one from *Lavandula angustifolia* (lavender) and one from *Picea sitchensis*. The enzymes belonging to *Vitis vinifera* were (+)-α-Pinene Synthase, (3*R*)-Linalool Synthase, (*E*)-β-ocimene/myrcene synthase, (+)-α-phellandrene synthase, (3*S*)-linalool/(*E*)-nerolidol synthase and geraniol synthase. From *L. angustifolia* was analyzed the activity of (*R*)-limonene synthase, and for *Picea sitchensis* was chosen the enzyme (+)-carene synthase. Plasmids used were pET28b-Gw74M8FL, pET28b-PN20M4FL, pET28b-O05J11F, pET28b-PN09M12, pET28b-PN25M6FL-1, pET28b-Gw63RR, pET28b-HQ336798.1 and pET29a-LaLIMS. The enzymes were expressed in *E. coli* BL21 (DE3), purified and then used on *in vitro* incubations with GPP, its natural substrate, and with 2-methyl-GPP, the unconventional C11 compound used as substrate by bacterial terpene synthases in the formation of non-canonical C11 compounds like 2-methylisoborneol and 2-methylenebornane.

For the incubations it was used HEPES buffer and purified enzyme, in a concentration that was in all cases higher than 0.5 mg/mL. GPP (46 µM) or 2-methyl-GPP (55 mM),

was used as substrate, in a total assay volume of 1 mL, inside 20 mL closed vials with septa. Assays were incubated at 30 °C for 3-4 hours with constant stirring. After the incubation period, the headspace of the assays was explored with a 30 µm thickness PDMS-SPME fiber (FIB-P-30/10; PAL systems) at 40 °C and volatiles were analyzed with GC-MS-QP2020 with AOC-6000 Autosampler (Shimadzu).

2.8   *In vivo* production of terpenes

2.8.1 Production of alkylated isoprenoids

The different *E. coli* strains were transformed with the combinations of plasmids described in page 43. Main cultures with 15 ml 2xYT medium in 100 ml baffled shake flasks were inoculated from LB pre-culture to an OD_{600} of 0.1. Where indicated, ^{13}C-methyl-L-methionine or L-methionine was added to the medium at 3 g/L. After cultivation for approximately 120 minutes at 37 °C (when OD_{600} = 1.0), IPTG at 100 µM was added for induction of gene expression. Induced cultures had terpenes extracted after 24 hours of incubation at 30 °C and 180 rpm, by exposing the SPME fiber A on the headspace for 30 minutes or SBSE devices in the medium during the cultivation.

2.8.2 Investigation of volatile production by *S. monomycini*

S. monomycini 15 mL main cultures were inoculated with cellular material from GYM-agar plates. Liquid cultures were incubated at 28 °C and shaken at 220 rpm. Analysis was made between days four and six of the cultivation cycle, with exposure of SPME fiber A at the culture headspace for 24 h. Solid cultivation was performed on slant agar tubes, prepared by filling agar in a tilted culture tube.

2.9   Analytical methods

2.9.1 GC-MS

Volatile compounds in the headspace of each assay and culture were analyzed after extraction with an 85 µm SPME fiber A composed of PDMS and Carboxen. The SPME fiber was exposed in the headspace of each assay and culture for the determinate

times (specific for each condition) and then inserted into the injection port of a GC-MS-QP2010 (Shimadzu) containing a DB-5 (5 %-phenyl)-methylpolysiloxane column with 30 m length and 0.25 mm thickness. For enzyme assays with plant terpene synthases and 2-methyl-GPP as substrate, SPME fiber B was used with automatic injection in GC-MS-QP2010 AOC6000 (Shimadzu) containing a DB-5 (5 %-phenyl)-methylpolysiloxane column. SBSE magnetic bars were placed in a glass tube (liner) and desorbed at the Thermodesorption Unit (TD-30) of GC-MS3.

Measurements were conducted as follows: helium as carrier gas, splitless injections at 250°C, 1 minute sampling time, and column flow of 1.1 ml/min. The column temperature was programmed as follows: 40 °C for 1.5 minute, 10 °C/min until 250 °C and then 20 °C/min until 300 °C.

Compounds were identified via comparison of mass spectra and retention indexes (RI) to the ones of reference substances or mass spectra of the NIST mass spectral libraries NIST14, NIST14s, NIST17, NIST17s and RIs published elsewhere (Adams 2007).

2.9.2 LC-MS

Samples from *in vitro* experiments were prepared for LC-MS analysis by precipitating the proteins with ice-cold acetonitrile at 3:1 proportion with sample material. The mixture was frozen, thawed and centrifuged (12000 g, 3 min). The supernatant was retrieved, filtered and used for subsequent analyses. Separation was achieved with a Luna Omega Polar C18 HPLC column (100mm length, 2.5 mm diameter, 1.6 µm particle size, 100 Å pore size, Phenomenex), using a binary gradient method (Solvent A: acetonitrile / Solvent B: 2% acetate). Gradient parameters were as follows: 0–4,5 min: 4% B; 4.5–5.0 min: 4% - 95% B; 5.0–7.0 min: 95% B; 7.0–7.5 min: 95% - 4% B; 7.5-12.0 min: 4 % B. Detection was performed with a tandem mass spectrometry detector with an ESI ion source (Shimadzu LCMS-8040) in multiple reaction monitoring (MRM) mode. Peak areas were normalized to internal standards and absolute quantification of metabolites was achieved with linear calibration curves of the standards.

Samples from *S. monomycini* liquid cultures had 1 mL of a liquid culture (4 days cultivated) collected in a microcentrifuge tube. After centrifugation at 16,000 g for 5 minutes, the supernatant was transferred to chromatography vials and used for analysis. The LC-MS entailed a high precision liquid chromatography (HPLC) system equipped with a triple quadrupole mass spectrometer and an APCI (atmospheric pressure chemical ionization) ion source. A Luna 3 Phenyl-Hexyl 100 x 2.00 mm (Phenomenex) column was used with oven at 30 ºC. For the HPLC separation method, a binary mobile phase gradient was used. The mobile phase components were water with 0.1 % formic acid (eluent A) and methanol (eluent B), in the following gradient program: 0-5 min 5 % B, 5.0-7.0 min 98 % B, 7.5-10.0 min 5 % B. Sample analysis were performed using an injection volume of 5 µL.

2.10   Bioinformatics Methods

2.10.1 BLAST

The protein sequence of the *S. coelicolor* GPP methyltransferase (NP_631739.1) was used as query for a BLAST search using the NCBI non-redundant protein database. The operon structures of resulting methyltransferase were analyzed regarding their operon structure and selected based in its novelty (or lack of resemblance to the already know operon of 2-MIB synthesis), which yielded the *S. monomycini* terpene biosynthesis operon shown in Results.

2.10.2 Protein model generation

A protein model of the prenyl pyrophosphate methyltransferase of *S. monomycini* was built with SWISS-MODEL (ProMod Version 3.70) (Biasini et al. 2014; Arnold et al. 2006; Benkert, Biasini, and Schwede 2011). The crystal structure of the GPP-methyltransferase of *S. coelicolor* (PDP ID: 3vc2.1.A) was used as template. Figures of crystal and model structure were created using UCSF Chimera (Pettersen et al. 2004) .

3 Results

3.1   Identification of novel methyltransferase coding regions by sequence comparison

The first step of this research was to look for enzymes capable of modifying terpene precursors. For that, the protein sequence of the GPP methyltransferase from *Streptomyces coelicolor* (NP_631739.1, SEQ01 from Appendix) was used as query for a BLAST search in the NCBI non-redundant protein database. Primarily the search yields more than 100 homologous sequences with similarities higher than 85% to the query. The highly similar sequences are organized in typical 2-methylisoborneol operons, containing a GPP-methyltransferase followed by a C11-terpene synthase. When sequences with lower similarity are analyzed, there are visible differences regarding the operon organization, suggesting that products different than C11 terpenes could be produced. That was the case of the sequence of Lon23, a putative IPPMT (BAF98640.1, SEQ02 from Appendix) from *Streptomyces argenteolus*, which reportedly participates on the synthesis of longestin by catalyzing the methylation of IPP (Hayashi et al. 2007). The C6 prenyl pyrophosphate formed is proposed to be condensed with one more C6 unit and two C5 IPP units to a dimethyl-GGPP (C22), that is an intermediate for the biosynthesis of the tetraterpenoid derivative longestin (Figure 13).
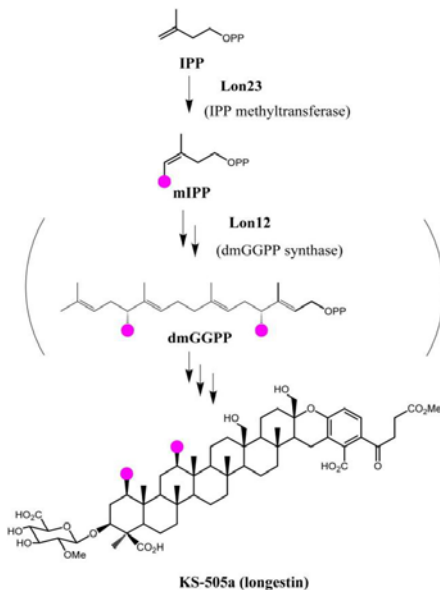
Figure 13. Synthesis of longestin, involving methylation of IPP catalyzed by Lon23 (Hayashi et al. 2007).

Lon23 is a SAM-dependent methyltransferase that has 39% amino acid identity with GPPMT from *S. coelicolor*. In a previous study, the analysis of an alignment between GPPMTs from *Streptomyces lasaliensis*, *S. coelicolor* and Lon23 from *S. argenteolus* infers structural similarities. Residues involved with recognition of the pyrophosphate moiety of GPP in both GPPMTs are also present in Lon23 (Ariyawutthiphan et al. 2012). The amino acid sequence of Lon23 was then used on a new BLASTp search for protein homologues, aiming to find similar methyltransferases with new catalytic functions, different than GPPMT and similar to an IPPMT. The database search for proteins with high similarity to the IPP methyltransferase Lon23 identified many bacterial genes encoding putative methyltransferases. One of the encountered homologue sequences belonged to the genome of *Streptomyces monomycini* Gauze and Terekhova 1986 emend. Nouioui et *al.* 2018, an actinomyces species first

described in 1986 (Reimer et al. 2019). The alignment between the query sequence and the putative methyltransferase from *S. monomycini* is depicted in Figure 14.
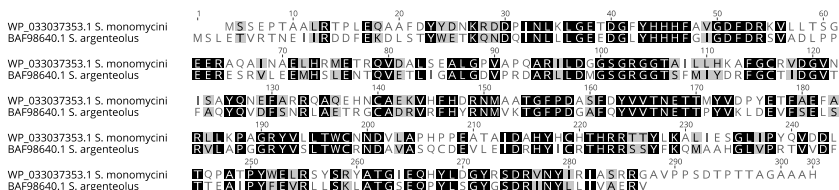
Figure 14. Clustal-omega alignment of IPPMT amino acid sequences from *S. monomycini* and *S. argenteolus*. Identical residues are marked in black, similar residues are marked in grey.

The amino acid sequence of the methyltransferase from *S. argenteolus* (BAF98640.1) shares 49% similarity with the methyltransferase of *S. monomycini* (WP_033037353.1), with 140 residues being identical and 144 divergent residues Since there are no structural information for the protein from *S. argenteolus*, no speculations could be made about their structural similarities. The flanking genes of these homologous sequences were also analyzed, and the presence of sequences encoding proteins involved in the MVA or MEP pathways indicate operons for terpene biosynthesis. Other sequences that indicate involvement in terpene biosynthesis are terpene synthases and prenyl transferases, as well as P450 oxidoreductases. The operons that contain only a methyltransferase gene, a monoterpene synthase gene and a regulator protein, most probably are part of pathways for the biosynthesis of a mixture of C11 terpenes including 2-MIB, like the one from *S. coelicolor*. Few operons seemed to encode pathways to products with complexities comparable to longestin from *S. argenteolus*, in which the methyltransferase might also methylate IPP, which is condensed to diterpene- or tetraterpene-like precursor pyrophosphates. In very few cases the operon encodes a methyltransferase, a prenyltransferase and a terpene synthase, which was the case of the MT of *Streptomyces monomycini*. The operon analysis of selected homologous prenyl pyrophosphate methyl transferase sequences is depicted on Figure 15.
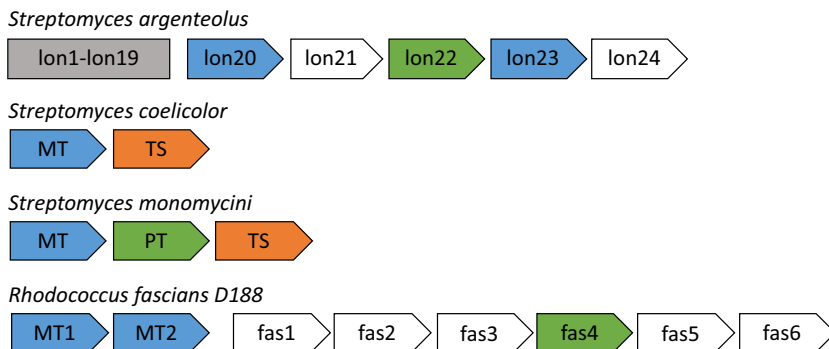
Figure 15. Analysis of operons which contain genes encoding prenyl pyrophosphate methyltransferases. Blue boxes highlight prenylpyrophosphate methyltransferases (lon20, lon23, MT). Green boxes highlight prenyltransferases (lon22, PT, fas4). Grey box represents 19 genes (lon1-lon19) upstream from the lon23 methyltransferase of *S. argenteolus*. Orange boxes highlight terpene synthases (TS).

The operon of *S. argenteolus* contains 24 genes, including the abovementioned methyltransferase (lon23) which was used for BLAST analysis. For the purpose of clarity and to highlight the methyltransferase genes, the first 19 genes upstream are represented as a grey box (Figure 15). The shown part of the operon also highlights another MT involved in the synthesis of longestin (lon20) and a prenyltransferase (lon22). For *S. coelicolor*, the methyltransferase gene (MT) is followed directly by a gene encoding terpene synthase (TS) which converts 2-methyl-GPP into a variety of C11 terpenes. In the case of *R. fascians*, two methyl transferases participate on the biosynthesis of cytokinins, and the respective genes are located side by side in the operon. In this case, MT1 catalyzes the methylation of IPP into an unknown product, and MT2 catalyzes two reactions: the methylation of IPP into 4-methyl-DMAPP and conversion of MT1's unknown product into 3-ethyl-2-penten-1-ol (Radhika et al. 2015). The operon of *R. fascians* also includes genes encoding a putative cytochrome P450 monooxygenase (fas1) and a prenyltransferase (fas4). The operon structure of a probable non-canonical terpenoid biosynthesis pathway was detected in *S. monomycini*, which contains a prenyltransferase-encoding gene flanking the MT sequence, followed by a putative terpene synthase gene.

The sequence of GPPMT from *S. coelicolor* was used as starting point of this study due to the fact it is the best described pyrophosphate methyl transferase, entailing descriptions of catalytic activity and protein structure. Even though the activities of other new methyltransferases were described during the course of this thesis, the GPPMT is still the only one that had its crystal structure elucidated, which brings important information to this study. An alignment made using Clustal Omega with free end gaps, between this GPPMT and *S. monomycini'*s MT sequences is shown below (Figure 16).
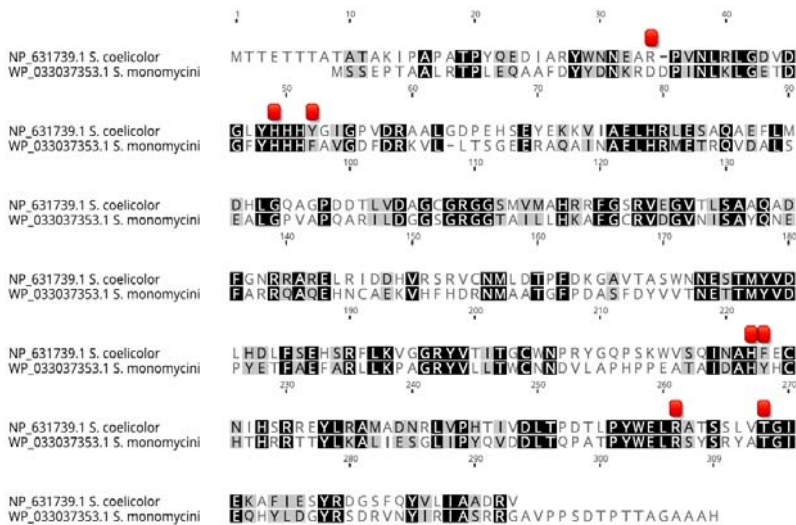
Figure 16. Amino acid sequence alignment of GPPMT from *S. coelicolor* and putative MT from *S. monomycini*. Analysis was performed using Geneious clustal global alignment with free end gaps, and Blosum62 as cost matrix. Identical residues are marked in black, similar residues are marked in grey. Red squares indicate residues involved in pyrophosphate moiety interaction.

Aligning the selected sequences an identity of 38% between the residues was found. Residues involved in the recognition of the pyrophosphate moiety on the GPPMT appear to be conserved, namely R34, H49, H221, R260 and T267 (Figure 16). The residues involved in interaction with substrate which are not identical are still similar between the sequences, namely Y51 (F44 in *S. monomycini*) and F222 (F214 in its

counterpart). The conserved residues indicate an important role on substrate recognition, which was probably maintained throughout the evolution of these sequences.

Aiming to explore the diversity of sequences coding for prenyldiphosphate methyltransferases in the database, a search was made using *S. monomycini's* MT sequence as query. A search for homologue amino acid sequences was made on the NCBI database an alignment was performed with the resulting sequences. A selected section of the alignment is shown on Figure 17. The yellow background highlights amino acids corresponding to the W194 position of the methyltransferase from *S. monomycini,* which will be discussed in a later section.

In the genome of *S. monomycini*, the presence of other terpene synthases was investigated. For that, the hypothetical protein TS from the operon as query (WP_030019073.1). Besides this terpene synthase found in the operon selected for this study, containing the methyltransferase, the genome of *Streptomyces monomycini* contains two other terpene synthase sequences, which could be involved with the synthesis of conventional terpenes. The protein with reference sequence WP_078624463.1 is annotated as germacrendiol/geosmin synthase, and it has 23% similarity with the query sequence. The locus of the gene is flanked by sequences involved with isoprenoid biosynthesis. The other sequence annotated as cyclase (WP_030017636.1), has only 24% similarity with germacrendiol synthase, indicating a possible difference in function.

Figure 17. Partial sequence comparison of the *S. monomycini* methyltransferase and putative prenyl pyrophosphate methyltransferases from different actinobacteria. The protein sequences were identified with a BLAST analysis using the *S. monomycini* methyltransferase as query (WP_033037353.1). The amino acids marked with a yellow background correspond to the W194 position of the methyltransferase from *S. monomycini*. (Drummond et al. 2019).

3.2 Protein homology modeling of prenylpyrophosphate methyltransferases from *S. monomycini* and *S. coelicolor*

With the finding of a promising methyltransferase sequence, the next step taken was to use *in silico* techniques to learn about its protein structure. In order to estimate the possible functions of the proposed isoprenoid pyrophosphate methyltransferase of *S. monomycini* a homology model was created, in addition to the sequence analysis. The analysis was made together with MSc. Max Lepple (geb. Kschowak).

As mentioned in the previous chapter, the sequence analysis of the proposed isoprenoid diphosphate methyltransferase of *S. monomycini* shows a sequence identity of 38% to the GPP-MT from *S. coelicolor* with a coverage of 91%. A homology model of the proposed isoprenoid diphosphate methyltransferase of *S. monomycini* (QMEAN4 = -3.46) based on the X-ray crystal structure of the GPPMT of *S. coelicolor* (PDB ID: 3VC2) reveals that the primary structure allows the model to form the Rossmann fold specific for family 1 of SAM-dependent methyltransferases.

The residues of the active site which are involved in the catalytic mechanism of the GPPMT are similar in the model (Figure 18), which corroborates the results found in the alignment. Equivalent to the GPPMT amino acids R34, H49, H221, R260 and T267, which are responsible for metal coordination and hydrogen bonds with the substrate diphosphate group (Köksal et al. 2012), the proposed methyltransferase of *S. monomycini* possesses at the corresponding positions D27, H42, H213 R252 and T259 similar or same residues. The cation-π interaction between the F222 side chain of the GPP-MT and the carbocation intermediate of the substrate can be carried out by the also aromatic side chain Y214 of the proposed methyltransferase of *S. monomycini*, which is at the equivalent position (Figure 18).

The comparison of the two active sites shows that the substrate pocket of the proposed isoprenoid diphosphate methyltransferase is much smaller due to the large residue of W194 which protrudes into the pocket of the active center. At the equivalent position of the GPP-MT of *S. coelicolor* is the relatively small amino acid G202. These facts, together with the deduction obtained from aligning the two sequences, support

the assumption that the enzyme is a methyltransferase and suggest that a smaller C5-diphosphate, presumably IPP or DMAPP is the substrate.
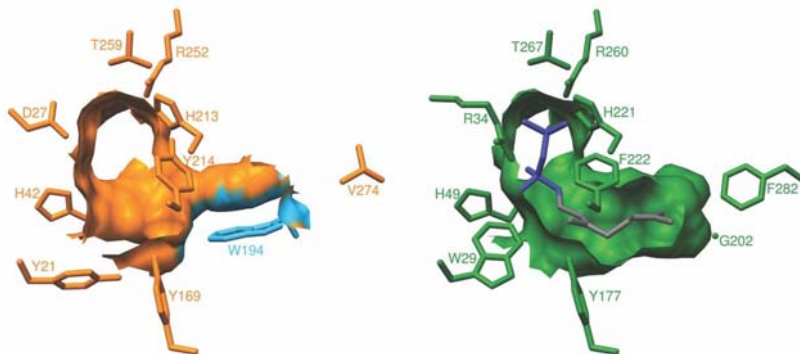
Figure 18. Comparison of substrate binding pockets. Isoprenoid diphosphate methyltransferases from *S. coelicolor* (right image, containing the substrate GPP) and *S. monomycini* (left image) are shown. The *S. coelicolor* enzyme image is based on an X-ray crystal structure with the GPP molecule colored in grey and blue (pyrophosphate). The *S. monomycini* enzyme image is based on a model structure prepared by comparison of the two amino acid sequences. The light blue tryptophan (W194) clearly leads to a much smaller active site pocket if compared with the GPP methyltransferase (Drummond et al. 2019).

3.3 *In vitro* assays with the prenylpyrophosphate methyltransferase from *S. monomycini*

To characterize the function of the methyltransferase encoded within the proposed *S. monomycini* terpenoid biosynthesis operon, *in vitro* assays were performed with the respective protein purified after overexpression in *E. coli*. Separate *in vitro* reactions were performed, using prenyl pyrophosphates with different alkyl chains. The compounds IPP, DMAPP, GPP and FPP were added at 60 μM to a solution that contained enzyme at a final concentration of 25 μM in suitable buffer, together with the cofactor SAM at 120 μM. After incubation at 30 ^{o}C for 20 – 40 hours, phosphatase solution was added to enable the cleavage of phosphate groups, forming terpene alcohols which are volatile. The headspace of each assay was explored with an adsorbing fiber and the compounds were separated and analyzed with GC-MS.

In accordance with the relatively small substrate binding pocket determined in the protein structure model (Figure 18), no 2-methyl-geraniol or other C11 alcohols could be detected, if GPP was used. Furthermore, the reaction with DMAPP didn't yield the expected 2-methyl-DMAPP-derived product 2-methyl-prenol or any other methylated products. Incubation with FPP yielded only farnesol, product of phosphate cleavage by the phosphatase, and not any C16 compounds as it would be expected as product of methylation. The substrate of this methyltransferase was found to be IPP, which contains a terminal double bond in contrast to GPP, DMAPP and FPP. As negative control, the same compounds used for the reaction with IPP were incubated under identical conditions but without the addition of enzyme, to exclude the possibility of spontaneous reaction between the substrates. The incubation of IPP with SAM without enzyme yielded isoprenol at RT=3.5 minutes (Figure 19, bottom chromatogram line), as single product of cleavage of phosphate groups from IPP. When compared with the negative control, no isoprenol was found when the methyltransferase was used, but more than one different IPP conversion products were detected after dephosphorylation (Figure 19, upper chromatogram line). At least two peaks were detected when the methyltransferase was incubated for 24 h with IPP and SAM, at RT=5.25 min (**A**) and RT=5.5 min (**B**), showing that incubation with the enzyme yields products different than the negative control without enzyme.
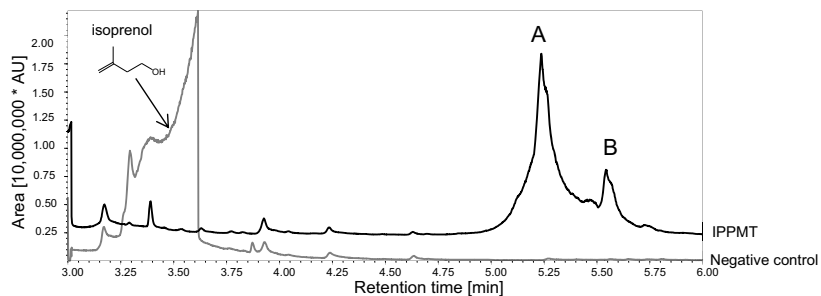
Figure 19. Product analysis after *in vitro* incubation of IPP and SAM. Substrates were incubated with MT and buffer (black line) or only buffer (negative control, gray line).

To explore the catalytical capability of the methyltransferase, longer incubation periods were applied to the reaction system. Incubations of 20 and 40 hours resulted in the detection of different isoprenoid alcohols, as shown in Figure 19. The identification of compounds was made by comparing the mass spectra and retention time of the compounds detected with the ones from synthesized reference compounds. The mass spectra comparisons are depicted in Figure 21 a-e. Four different C6 compounds and two C7 compounds were structurally identified. For clarity, they were named according to the position in the molecule where the methyl group was added, and correspondent IUPAC names are shown in Table 9.
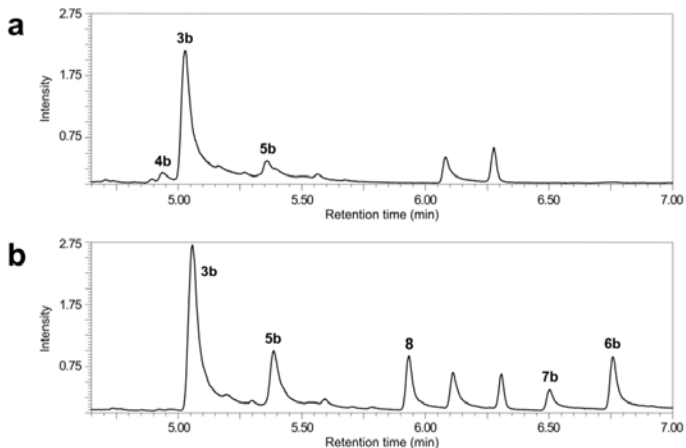
Figure 20. Product analysis after *in vitro* incubation of *S. monomycini* MT with IPP and SAM. Volatile products of the reaction were analyzed via SPME-GC-MS after an incubation time of 20 h (a) and 40 h (b) as described in the methods section. The MT reaction was followed by a phosphatase treatment to dephosphorylate the prenyl pyrophosphate compounds. The chromatograms are representative of three independent experiments. Numbered peaks correspond to identified compounds: **4b**) (*Z*)-4-methyl-isoprenol; **3b**) (*E*)-4-methyl-isoprenol; **5b**) 4-methyl-prenol; **7b**) 4,4-dimethyl-prenol; **6b**) 4,4-dimethyl-isoprenol; **8**) C7 compound of unknown structure. Products were identified by comparison of the obtained mass spectra with mass spectra and retention time of reference compounds (Figure 21) (Drummond et al. 2019).
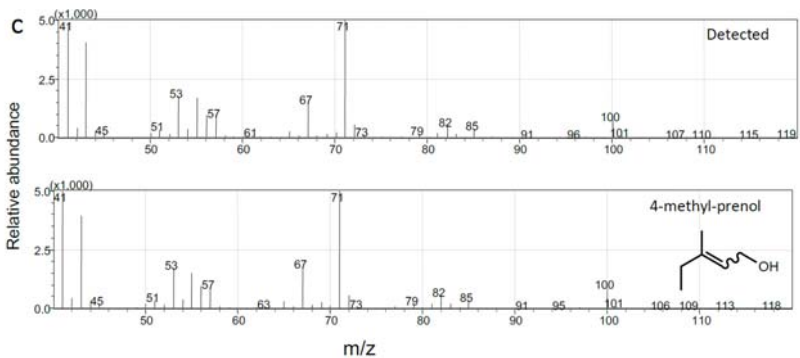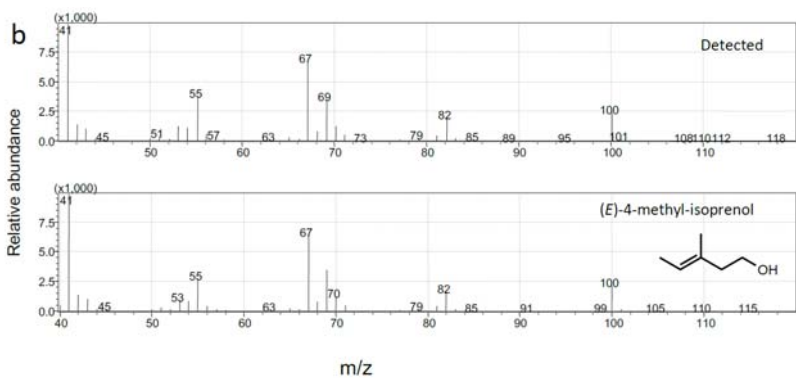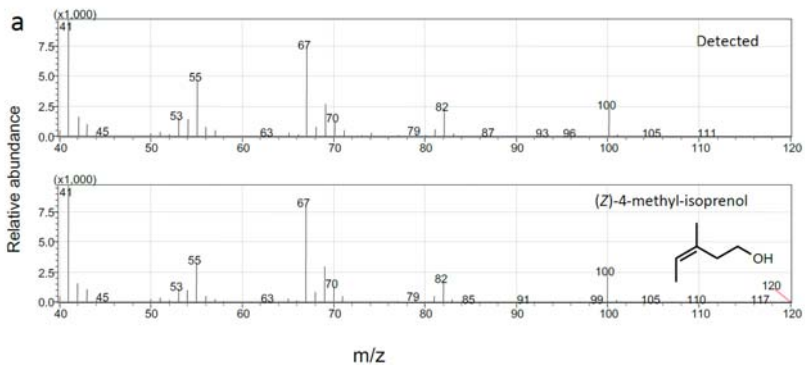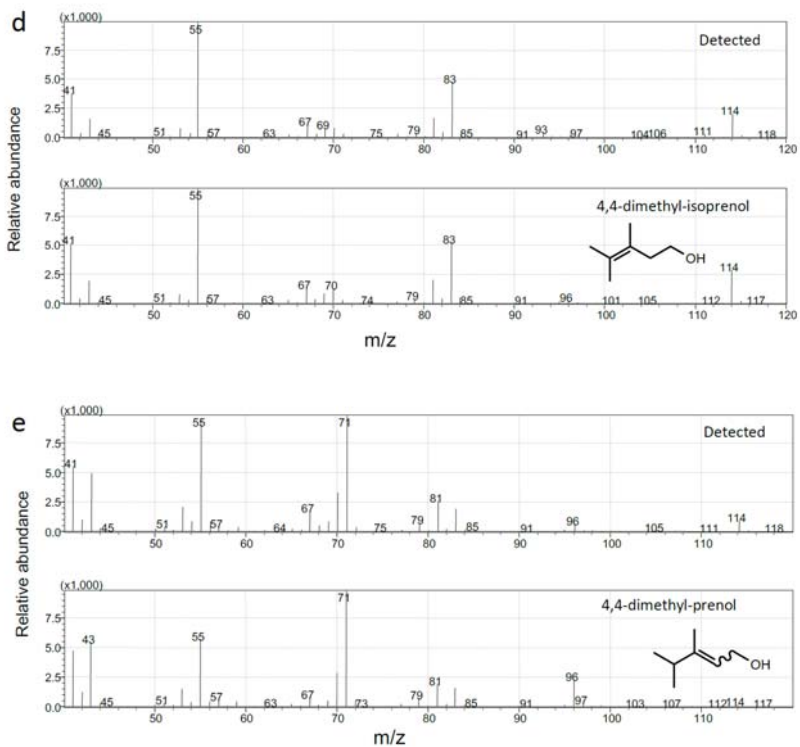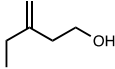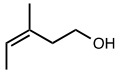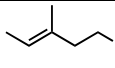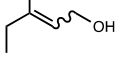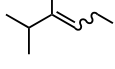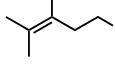
Figure 21. Comparison of mass spectra of detected (upper) and reference (lower) compounds. a) Compound **4b** and reference compound (*Z*)-4-methyl-isoprenol. b) Compound **3b** and reference compound (*E*)-4-methyl-isoprenol. c) Compound **5b** and reference compound 4-methyl-prenol. d) Compound **6b** and reference compound 4,4-dimethyl-isoprenol. e) Compound **7b** and reference compound 4,4-dimethyl-prenol.

Table 9. IUPAC nomenclature, names and chemical structures of the compounds produced by incubation of the *S. monomycini* MT with IPP and SAM

IUPAC-name	Compound name	Chemical Structure
3-methylbut-3-en-1-ol	5-methyl-isoprenol	
(*Z*)-3-methyl-3-penten-1-ol	(*Z*)-4-methyl- isoprenol	
(*E*)-3-methyl-3-penten-1-ol	(*E*)-4-methyl-isoprenol	
3-methyl-2-penten-1-ol	4-methyl-prenol	
3,4-dimethyl-2-penten-1-ol	4,4-dimethyl-prenol	
3,4-dimethyl-3-penten-1-ol	4,4-dimethyl-isoprenol	

The mass spectrum of (*Z*)-4-methyl-isoprenol (4b) is characterized by m/z of 41 (base peak), 53, 55, 67, 69 and 82, besides the molecular ion of m/z 100 (Figure 21 a). The signal at m/z 82 probably represents loss of water by the molecule. The retention index for this compound is 856. This compound is a product of methylation of IPP at carbon 4, with posterior dephosphorylation. The terminal double bond of IPP is kept at the same position observed in the substrate.

The main product of the methyltransferase-catalyzed reaction using IPP and SAM as substrates, (*E*)-4-methyl-isoprenol, has a mass spectrum very similar to its (*Z*)- isomer (Figure 21 b). The main difference here would be the relative abundance of signal intensity of each specific fragment, but the fragments are the same observed for (*Z*)-4-methyl-isoprenol: base peak at m/z 41, then signals at m/z 53, 55, 67, 69 and 82 (loss of water), and finally the molecular ion at m/z 100. The retention index of 862 differs from the isomer's in +6 points.

The mass spectrum of compound 4-methyl-prenol shows a different profile than the previous isomers (Figure 21 c). Even though the molecular ion at m/z 100 is present, as well as the fragment at m/z 82 representing the loss of water, a pronounced signal at 71 evidences the change in the position of the double bond in this particular molecule. The loss of m/z 29 (100-71) could represent the breakage of terminal alcohol function, since the ion of m/z 71 is probably $C_5H_{11}^+$. Additionally, signal with m/z 41 shows 50% of abundance, and m/z 53, 55 and 67 appear in smaller percentages. The retention index for this compound is 884, a difference of +22 from the main product (*E*)-4-methyl-isoprenol.

The retention index for 4,4-dimethyl-isoprenol is 941. In the mass spectrum of the compound (Figure 21 d), the molecular ion of 114 is present, as well as fragment ions of m/z 41, 55, 67, 70, 81 and 83. This compound is product of a double methylation of IPP on carbon 4. The double bond from the substrate IPP remains at the same position in this case.

The mass spectrum of compound 4,4-dimethyl-prenol (Figure 21 e) differs in fragmentation pattern from the one of its isomer 4,4-dimethyl-isoprenol. The molecular ion of m/z 114 appears with a low intensity signal, and then m/z 71 shows the highest abundance relative to the other signals. The appearance of this ion in such high intensity resembles the pattern of the compound 4-methyl-prenol, which also showed the base peak at m/z 71. The similarity probably lays on the fact that both compounds have the double bond closer to the alcohol function, provoking a breakage of the terminal group forming the ion $C_5H_{11}^+$. Signals at m/z 41, 43, 53, 55 and 83 also appear. The retention index for this double methylated compound is 926.

The reaction of IPP with SAM, catalyzed by the methyltransferase and finalized with dephosphorylation, yielded six different, recognizable peaks, with (*E*)-4-methyl-isoprenol as prominent main product (mean peak area of 86,6 ± 3 % of the total peak area of assigned products, see also Table 10). In addition, five minor products with relative lower product peak areas could be identified. The two diastereomers 4-methyl-prenol and (*Z*)-4-methyl-isoprenol represent methylated isoprenol stereoisomers with different configurations. But surprisingly, also dimethylated isoprenoid compounds were produced, namely 4,4-dimethyl isoprenol and 4,4-dimethyl prenol. These

compounds were found after longer times of incubation (40 h), together with a reduction in the average peak area of the single methylated compounds.

Besides the abovementioned identified compounds, one compound was detected but could not be identified. The peak number 8 probably corresponds to another double methylated compound with unknown structure. The mass spectrum for this unknown compound is shown in Figure 22.
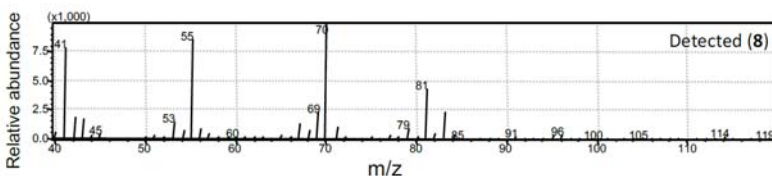
Figure 22. Mass spectrum of compound **8** of unknown structure.

Peak number 8, which appears at 5.9 minutes retention time, is apparently related to the methyltransferase activity. Even though its structure could not be clarified, the mass spectrum (Figure 22) shows similarities with the ones from the other two double methylated compounds that were identified. Similar to 4,4-dimethyl-isoprenol as well as to 4,4-dimethyl-prenol, the unknown compound shows conspicuous signals at m/z 41, 55, 67, 70, 81 and the molecular ion of m/z 114. The strong signal at m/z 41 indicates a terminal double bond. It is reasonable to assume that this compound of unknown structure is also a double methylated isoprenoid, probably an isomer of the two C7 compounds described in this work.

It is important to notice that the cleavage of diphosphate groups can also occur spontaneously in the presence of water, which cannot be discarded in this case also. To control that, an SPME fiber was used to explore the headspace of the reaction before the application of a phosphatase solution, which showed no peaks or only traces of the expected compounds (not shown).

Table 10. Conversion of IPP into mono- and di-methylated IPP derivatives by the *S. monomycini* IPP methyltransferase *in vitro* using SAM as cofactor. Products were identified by comparison with reference compounds via GC-MS and product amounts are given as relative peak area ratios.

Prenyl pyrophosphate	Dephosphorylated form	Relative product amounts	
		20 h	40 h
IPP	Isoprenol	0	0
DMAPP	Prenol	0	0
(*E*)-4-methyl-IPP	(*E*)-4-methyl-isoprenol	86.6 ± 8.4	71.8 ± 6.2
(*Z*)-4-methyl-IPP	(*Z*)-4-methyl-isoprenol	2.8 ± 1.2	0
4-methyl-DMAPP	4-methyl-prenol	10.6 ± 7.5	17.0 ± 0.5
4,4-dimethyl-IPP	4,4-dimethyl-isoprenol	0	8.2 ± 4.2
4,4-dimethyl-DMAPP	4,4-dimethyl-prenol	0	3.0 ± 1.7

After 20 h of incubation, (*E*)-4-methyl-isoprenol is the most abundant product, followed by 4-methyl-prenol and (*Z*)-4-methyl-isoprenol (Table 10). The chromatograms also show that there is complete conversion of IPP into methylated alcohol products, with no isoprenol peaks (Figure 19).

The two forms of 4-methyl-isoprenol appear very close together, with the *Z* form coming in a slightly earlier retention time than the *E* form. 4-methyl-prenol follows the two methylated isoprenoids in a later retention time, directly at 5.4 minutes. After 40 h of incubation, three additional peaks were identified while one of the peaks visible at 20 h of incubation disappeared. The first peak to appear at RT=5.0 minutes is one of the two diastereomers of a methylated IPP, in this case (*E*)-4-methyl-isoprenol, since (*Z*)-4-methyl-isoprenol does not appear anymore after a longer incubation time and 4-methyl-prenol follows at 5.4 minutes.

3.4 *In vitro* incubations with methyltransferase mutants

The first chapters from the results part showed how the amino acid sequence and three-dimensional structure of the methyl transferase from *S. monomycini* were analyzed, especially in comparison with the GPPMT from *S. coelicolor*. It was hypothesized that differences in the amino acid sequences led to differences in the tertiary structure of both enzymes, resulting in distinct catalytic activities. The *in vitro* assays showed that indeed the activity of both enzymes is significantly discrepant, regarding their substrates and products, but still bearing the function of methyl transfer. In order to test if our hypothesis about the role of one specific amino acid from IPPMT on reducing the substrate pocket and causing the substrate specificity found for this enzyme, mutations were inserted on IPPMT and GPPMT sequences. One specific amino acid from IPPMT, namely W194, was exchanged into a G. One specific amino acid for GPPMT, namely G202, was exchanged into a W. In this way, the amino acid exchanges were expected to recreate the substrate pocket of GPPMT, in IPPMT, and vice versa, by exchanging the natural amino acid of one, for the corresponding amino acid (regarding the position) of the other. Plasmids were designed containing the mutants with a his-tag and transferred separately into *E. coli* BL21(DE3). The plasmids used were pLD-6 and pLD-7. After expression and purification, the enzyme was used for incubation with IPP or GPP to check if the substrate specificity was altered by the mutation inserted. Reaction conditions were the same used for incubations with the natural methyltransferases, described above and here as follows: PIPES buffer, 25µM enzyme, 60 µM IPP or GPP, 120 µM SAM, 1 mL reaction volume in 4 mL vials, incubation at 30 °C overnight. After the incubation period with methyltransferase, 1 mL of phosphatase solution (7,5 mg acid phosphatase in acetate buffer) was added for another 2 hours of incubation at 30 °C stirring. Volatiles present in the headspace were extracted with an SPME fiber for 30 minutes at 40 °C and analyzed with GC-MS. The purification of the IPPMT mutant revealed to be unfeasible, as no protein was present already in the cell lysate, and the incubations with the purified GPPMT did not provide any proof to the working hypothesis. The volatiles produced by incubation with IPP were only isoprenol, and the ones produced by incubation with GPP were only geraniol, without any additional methylation.

3.5   *In vivo* experiments

3.5.1   Heterologous expression of the operon from *S. monomycini*

After characterizing the IPPMT reaction *in vitro*, enzyme production in a live host was conducted, and *E. coli* was chosen for heterologous expression. The aim was to observe the function of the expressed protein *in vivo*, as well as setting ground for the establishment of a platform for *in vivo* non-canonical terpene production. For that, the operon from S. *monomycini*, in which the protein sequence of the methyltransferase is encoded, was selected for the expression, detection and analysis of isoprenoid products.

The starting point to generate strains for non-canonical terpene production was the creation of an *E. coli* strain with high levels of terpene precursors. For that, the plasmid pJBEI-6409 was used as base for the creation of further vectors for terpene precursor production. The plasmid pJBEI-6409 was originally developed by researchers from the Joint Bioenergy Institute (thus the acronym pJBEI) in Emeryville-CA for limonene production out of glucose, in engineered *E. coli* (Alonso-Gutierrez et al. 2013). The plasmid contains, besides structural components such as promotors, terminators and antibiotic resistance, 7 enzymes belonging to the mevalonate pathway, from different organisms. The first part of the pathway, entailed in an operon in the plasmid, contains genes for converting acetyl-CoA into mevalonate: AtoB from *E. coli*, and HMGS and HMGR from *Staphylococcus aureus* codon-optimized for *E. coli*. The second operon contains genes for converting mevalonate into isoprenoid precursors IPP and DMAPP: MK, PMK and PMD from *S. cerevisiae* and IDI from *E. coli*. Finally, truncated versions of GPP synthase (GPPS) from *Abies grandis* and limonene synthase (LS) from *Mentha spicata*, both codon-optimized for *E. coli,* complete the plasmid. In order to use this engineered plasmid for overproduction of IPP and DMAPP, a deletion of the genes GPPS and LS was performed, leaving all the genes for converting acetyl-CoA into the precursors IPP and DMAPP. In that way, the cells could use their basic metabolism to provide the substrate acetyl-CoA, and the corresponding enzymes from the mevalonate pathway could then produce the necessary prenyl diphosphate precursors.

To investigate the products of the biosynthetic pathway encoded by the genes from *S. monomycini* in an *in vivo* system, a plasmid containing the operon previously described (see Results section 3.1, operon with methyltransferase, prenyltransferase and terpene synthase) was transformed into *E. coli* for heterologous expression and production of terpenes. The plasmid was transformed into *E. coli* strains overexpressing genes from the mevalonate pathway, so that precursor availability would not be a limitation for terpene production. Cells containing two plasmids, one for the overproduction of precursors IPP and DMAPP, and another one containing the operon from *S. monomycini*, were grown in complex media and then had the produced volatiles extracted and analyzed by HS-SPME-GC-MS. As negative control, cells containing the mevalonate pathway and an empty plasmid (the same pET-Duet vector used for expression of the operon from *S. monomycini*, but with no genes inserted) were created. Their volatile compounds profile was then compared to the one from the cells containing the operon. The analysis of volatiles that were present in the headspace of cultures revealed a series of compounds, which were not produced when only the mevalonate pathway was overexpressed (Figure 23). The volatile compound profile was then analyzed regarding the mass spectra.

Mass spectra of peaks derived from test samples revealed various compounds with no match in the NIST database (less than 80% similarity with the matches after search). Especially in early retention times (around 5 minutes), several peaks, with m/z values of 100 and 114, were detected. At 5.3 minutes, a clear peak with mass of 100 was found, corresponding to (*E*)-4-methyl-isoprenol. Great part of terpene derivatives detectable in the headspace of *in vivo* samples are volatile alcohol terpenoids, due to the presence of phosphatases in the intracellular environment and their cleavage of pyrophosphate groups generating alcohol functions.

The operon from *S. monomycini* contains three genes, coding for three different enzymes. Therefore, in this preliminary analysis it was not possible to identify the functions of each enzyme on the production of the volatiles found. To isolate the different volatiles, two other strains were created based on the one with the whole operon. Two additional strains were created, lacking the terpene synthase from the operon, and lacking the prenyltransferase and the terpene synthase, with only the

methyltransferase in it. To test the effects of each enzyme from the operon in the products formed *in vivo*, different strains were transformed with plasmids containing parts of the operon, and their volatile spectra were compared. All the strains contained two plasmids: one with the enzymes from the mevalonate pathway, and the other plasmid with one or more enzymes from the selected operon from *S. monomycini*. The comparison between volatile profiles from strains containing genes coding different enzymes from the operon revealed that they do not differ qualitatively. The same compound peaks are present in the three strains with slight differences on peak area for some of the compounds (Figure 23). However, for peaks with extremely low area, the presence could not be detected when only the methyltransferase was expressed. The chromatograms of strains containing the whole operon did not differ from chromatograms of strains containing only MT and PT. Therefore, it was not considered the influence the terpene synthase on product formation, at least qualitatively. The presence of the prenyltransferase altered the area of peaks, but not the type of compounds produced. Only the methyltransferase was considered having an effect on non-canonical product formation for those strains, posteriorly, experiments were performed *in vivo* using only the strains expressing the methyltransferase together with the mevalonate pathway.
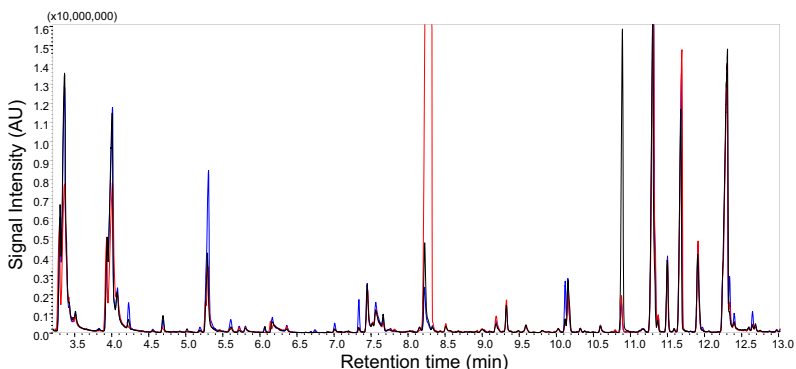
Figure 23. Total ion chromatogram of volatile extract from culture headspace of *E. coli* expressing different sections of S*. monomycini* operon**.** *E. coli* cells were expressing mevalonate pathway with an additional plasmid containing the whole operon (blue), methyltransferase and prenyltransferase (red), or only the methyltransferase (black). Red peak at RT=8.25 minutes is a limonene contamination.

Analysis of products derived from the expression of *Streptomyces monomycini* IPPMT revealed different volatile terpene alcohols, especially in early retention times, ranging from 4.5 to 7.5 minutes (Figure 24). As depicted on Figure 23 and Figure 24 the very early peaks appearing at 3.0 and at 4.0 minutes correspond to isoprenol and prenol, which are the alcohol forms of IPP and DMAPP respectively and show relatively high areas. This result reflects the excess of terpene precursors that was being produced, which was intended so that terpene production is not limited by substrate availability. The peaks also appear in the negative control, which was expected since the strain contained also the mevalonate pathway, but an empty plasmid instead of the plasmid containing the *S. monomycini* IPPMT (Figure 24).
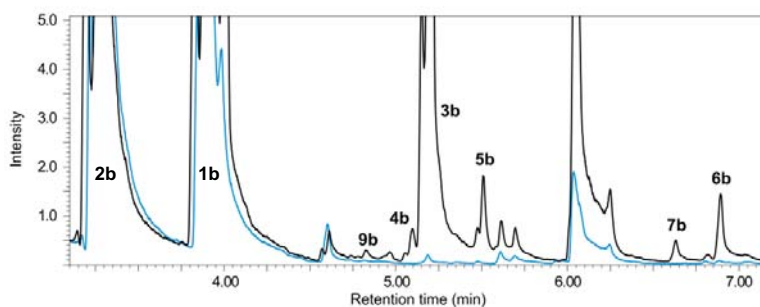
Figure 24. *In vivo* synthesis of non-canonical prenyl pyrophosphates. Production of non-canonical C6 and C7 prenyl alcohols by *E. coli* cells expressing the *S. monomycini* IPPMT. Blue line refers to empty vector control. The alcohol derivatives of respective prenyl pyrophosphates were detected by headspace-SPME-GC-MS analyses and products were identified by comparison of the obtained mass spectra with mass spectra and retention time of reference compounds. **2b**) prenol; **1b**) isoprenol; **9b**) 5-methyl-isoprenol; **4b**) (*Z*)-4-methyl-isoprenol; **3b**) (*E*)-4-methyl-isoprenol; **5b**) 4-methyl-prenol; **7b**) 4,4-dimethyl-prenol; **6b**) 4,4-dimethyl-isoprenol (Drummond et al. 2019).

The volatile profile of *E. coli* cells expressing the *S. monomycini* IPPMT contained the compounds identified during *in vitro* reactions using the methyltransferase. The compounds (*E*)-4-methyl-isoprenol, (*Z*)-4-methyl-isoprenol and 4-methyl-prenol were detected at retention times around 5 minutes, as well as the double methylated compounds 4,4-dimethyl-isoprenol and 4,4-dimethyl-prenol. Additionally, at RT=4.8

minutes, the compound 5-methyl-isoprenol was detected and identified. Mass spectra comparison is shown on Figure 25.
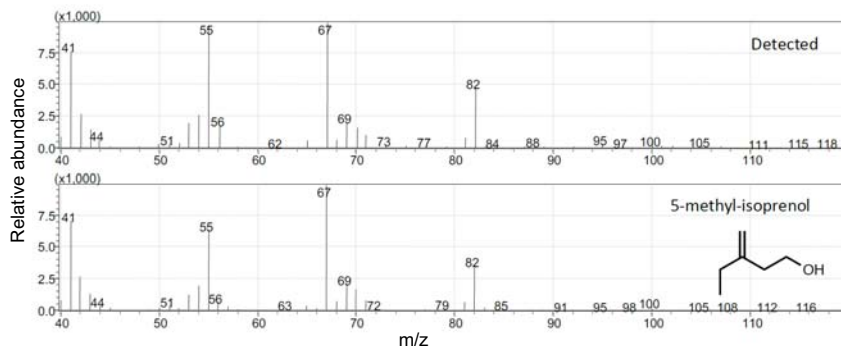
Figure 25. Comparison of mass spectra of compound **9b** detected and the reference compound 5-methyl-isoprenol.

The mass spectrum of 5-methyl-isoprenol is characterized by a base peak of m/z 67, as well as m/z 41, 55, 82 and the molecular ion of m/z 100. The signal at m/z 82 (M-18) indicates loss of water from the molecule, and m/z 67 (M-33) indicates loss of H_2O and CH_3. The detected ion at m/z 41 probably refers to ion $CH_2=CH-C^+$ and indicates a terminal double bond. This compound is a product of methylation of IPP at carbon 4 by the methyltransferase, which results from deprotonation at carbon 5 of the carbocation intermediate.

3.5.2 Isotope labeling of methylated compounds and identification of long-chain methyl-terpenes

The use of stable isotopes in labeling techniques together with mass spectroscopy allows the determination of molecular weights, nature of functional groups, double bonds and chemical formulas. This is especially helpful for the analysis of unknown compounds, since it provides structural information otherwise not obtainable, for example for the identification of methylated compounds. The growth of *E. coli* is

80

reportedly not affected by the addition of nutrients labeled with isotopes ^{13}C, ^{15}N and ^{2}H in the media (Marley, Lu, and Bracken 2001), unless deuterated water is used, which then can hinder growth due to deceleration of enzyme-mediated metabolic reactions (Hoopes et al. 2015).

In order to identify which compounds in the chromatogram were products of methylation by activity of methyltransferases, feeding experiments were performed with labeled methionine. The isotopically labeled amino acid L-methionine-(methyl-^{13}C) contains a ^{13}C isotope on its methyl function, the same one that is transferred to S-adenosyl-methionine by the enzyme MAT. When labeled methionine is used by MAT to produce SAM, the product is labeled SAM, or ^{13}C-SAM, containing a ^{13}C isotope on its methyl function, which further transferred to other molecules by action of methyl transferases. This means that the initial ^{13}C isotope present in L-methionine-(methyl-^{13}C) is the same one that ends up in products of methylation by SAM-dependent methyltransferases. By feeding bacterial cultures with isotopically labeled methionine in excess, it is possible to observe the integration of the labeled atom in metabolites formed by these organisms. This happens because when an amino acid is in excess in the medium, the cells tend to take it up even though they are able to produce it, as means of saving energy for the production of such amino acids. There is always a residual production of the unlabeled amino acid by the bacteria, but the intake amount and consequent yield of labeled products adds up to the majority of detected compounds, which is enough for detection. In other words, an excess of labeled methionine in the medium leads to an intracellular pool of labeled methionine, which in sufficient for producing ^{13}C-SAM and labeled methyl groups in the products of interest.

For the labeling of methylated compounds in the *in vivo* experiments, cells were allowed to grow in medium enriched with L-methionine-(methyl-^{13}C), provided at a concentration of 3 g/L of medium. The control cultures were fed unlabeled methionine at the same concentration of 3 g/L, so that they would also have an excess of the amino acid, therefore exclude possible effects of nutrient addition on the comparison between treatments. Experimental work with labeled amino acids was performed together with MSc. Rupa Bhattarai, which concluded her master thesis within this project.

This technique was employed to identify products containing additional methyl groups. During *in vivo* expression experiments, different compounds were produced in a broad range of retention times, a large percentage of which had no correspondent mass spectra in the NIST database. Not only the previously identified molecules of low molecular weight were produced, but also peaks at later retention times were detected, indicating the production of methylated terpenes with longer prenyl chains. However, many of the compounds produced were not specific to the methyltransferase activity (e. g. geraniol and other similar C10 compounds, product of condensation of IPP and DMAPP by *E. coli*'s native prenyl transferases), thus the use of isotope labeling technique helped to identify the compounds which were indeed a product of methylation by the enzyme in focus. Another extraction method was also used, in which the compounds present in the liquid medium were adsorbed by a PDMS-coated magnetic stirrer and posteriorly directly desorbed thermally into a GC-MS device. The extraction of solutes within the aqueous media allowed the access to larger compounds, which were less volatile and therefore not captured by the headspace SPME fiber extraction. The shift of +1 in the mass spectra of labeled compounds indicated methylation. Also, in some cases the signals (at mass spectra) revealed shifts of +2, indicating two methyl groups added. A chromatogram corresponding to SBSE extraction of medium containing *E. coli* cells expressing the *S. monomycini* IPPMT is depicted in Figure 26.
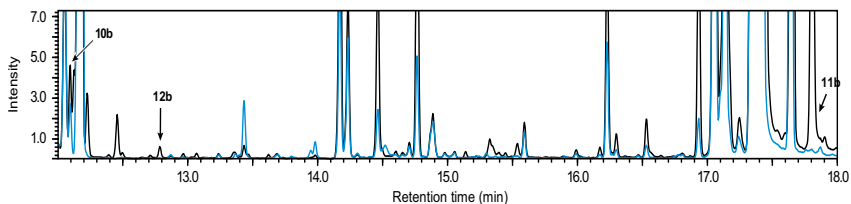
Figure 26. Identification of C11, and C16 isoprenoid intermediates in cultures of *E. coli* cells expressing the *S. monomycini* IPP methyltransferase. Total ion chromatogram of in vivo SBSE-extracted medium sample of strain *E. coli* expressing the mevalonate pathway and IPP methyltransferase. Arrows indicate peaks corresponding to compounds 4-methyl-geraniol (**10b**); 8-methyl-geraniol (**12b**); 4-methyl-farnesol (**11b**) (Drummond et al. 2019).

The labeling of methylated compounds with non-radioactive isotopes allowed the detection of C11 and C16 compounds in the chromatograms, enabling the identification of additional non-canonical terpene structures. The C11 compounds 4-methyl-geraniol and 8-methyl-geraniol, as well as the C16 alcohol 4-methyl-farnesol were identified through comparison of detected mass spectra and retention times with the ones from reference compounds (Figure 27-Figure 29).
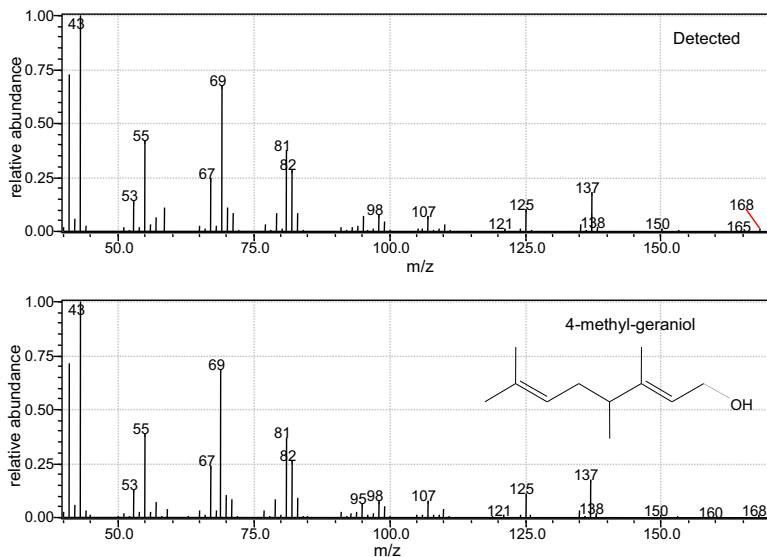
Figure 27. Comparison of mass spectra of **10b** and the reference compound 4-methyl-geraniol.

The fragmentation spectrum of 4-methyl-geraniol revealed fragment ions m/z 43, m/z 55, m/z 69, m/z 81, m/z 137 and finally the molecular ion m/z 168 (Figure 27). The fragment at m/z 150 indicates loss of water (M-18). The retention index for this compound is 1272. The additional methyl group on carbon 4, probably originated from the condensation of 4-methyl-IPP with DMAPP, creates a chiral center. There are four different possible configurations for this molecule, considering *E* or *Z*, *R* and *S* configurations. Chromatograms from compound collections of *in vivo* experiments revealed different peaks associated with isomers from 4-methyl-geraniol.
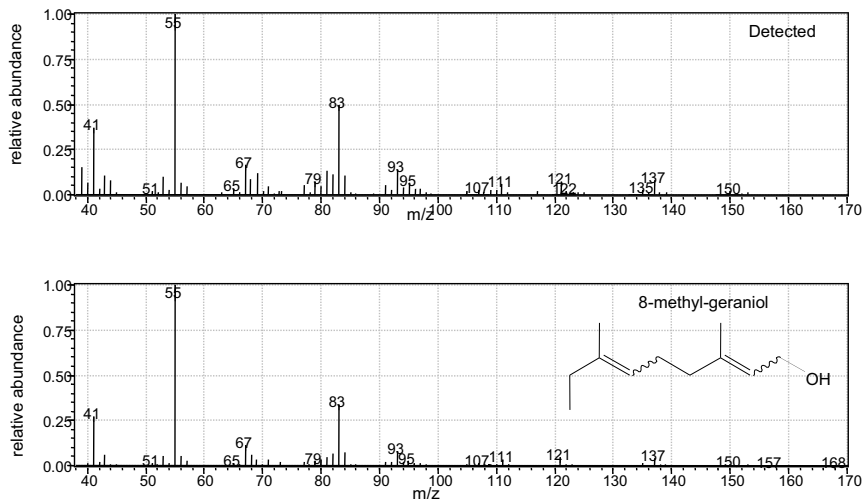
Figure 28. Comparison of mass spectra of **12b** and the reference compound 8-methyl-geraniol.

The C11 compound 8-methyl-geraniol has a mass spectrum which differs substantially from that of 4-methyl-geraniol. The spectrum shows a base peak at m/z 55, clear signals at m/z 83 and m/z 41 and the relative abundance of all other fragments is very low (Figure 28). In comparison, the mass spectrum of 4-methyl-geraniol depicts several clusters of fragments with similar m/z, spaced by 14 (CH_2), which is common for alkenes. Terminal alkenes form m/z 41, fragment that is seen in the lower section of this spectrum. The retention index for 8-methyl-geraniol is 1321. This retention time is higher than that of 4-methyl-geraniol and the two compounds showed good separation in the chromatograms with the methods used, with no overlapping.
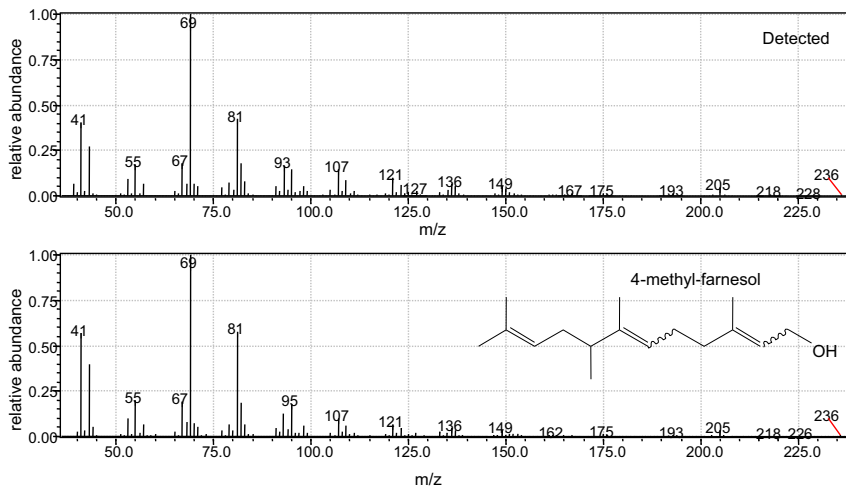
Figure 29. Comparison of mass spectra of **11b** and the reference compound 4-methyl-farnesol.

Similarly to the fragment spectrum of 4-methyl-geraniol, the mass spectrum of 4-methyl-farnesol reveals strong signals at m/z 55, m/z 68 and m/z 81(Figure 29). The molecular ion is present in low amounts at m/z 236 and it is possible to see a small signal at m/z 93 (M-43) corresponding to the loss of $C_3H_7^+$. The retention index for this compound is 1816, and longer chain terpenes like farnesol (C15) are released at much later retention times. At least two different peaks with similar mass spectra were detected at different retention times of the chromatogram during *in vivo* experiments, indicating the presence of different isomers in the mixture. The introduction of the additional methyl group in this compound, which most probably originates from the condensation of 4-methyl-IPP with GPP catalyzed by endogenous prenyltransferase from *E. coli*, also creates a chiral center.

Besides C11 and C16 compounds containing one methyl group, C12 and C17 terpenoids bearing two additional methyl groups were detected with the help of isotope labeling. The chromatogram below (Figure 30) shows peaks corresponding to single- and double methylated terpene alcohols of longer prenyl chain.
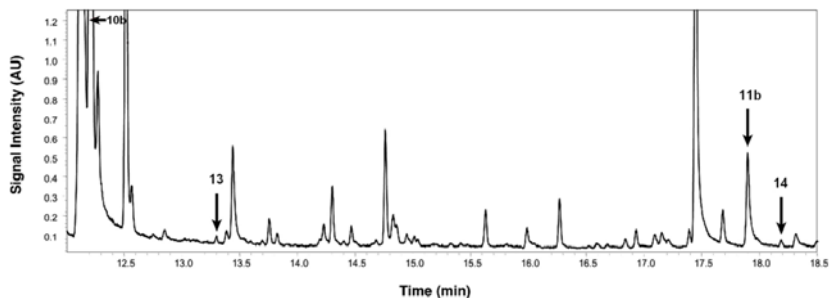
Figure 30. Identification of C12 and C17 isoprenoids in cultures of *E. coli* cells expressing the *S. monomycini* IPP methyltransferase. Total ion chromatogram of *in vivo* SBSE medium extract sample of *E. coli* containing plasmids expressing the mevalonate pathway and IPP methyltransferase. Results were compared with an experiment in which the cultures were fed with an excess of L-methionine-(methyl-^{13}C). Relative mass spectra shifts between the two chromatograms were observed and indicate the MT-catalyzed introduction of methyl groups. **10b**) 4-methyl-geraniol; **13**) C12 compound of unknown structure; **11b**) 4-methyl-farnesol; **14**) C17 compound of unknown structure. The mass spectra of compounds **13** and **14** are shown in Figure 31 and Figure 32 (Drummond et al. 2019).
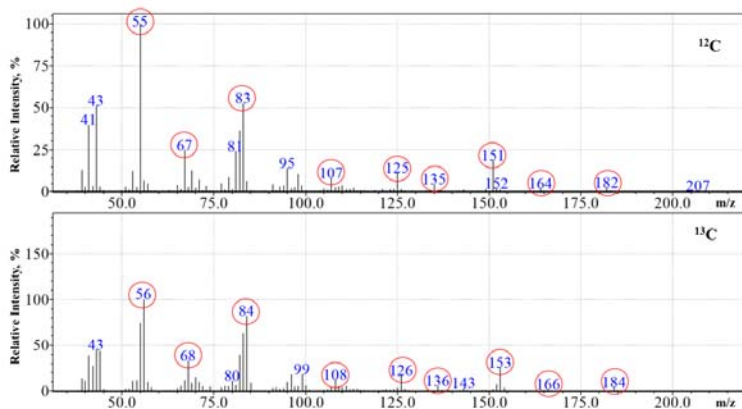
Figure 31. Mass spectra of **13** (C12 isoprenoid alcohol) from cultures of *E. coli* + pLD-03 + pMK-17 grown with L-methionine (upper MS) or L-methionine-(methyl-^{13}C) (lower MS). Red circles mark the fragments which contain the transferred methyl group(s).

The mass spectrum of compound **13** has resemblance with the one from 8-methyl-geraniol. The general ratio between fragments m/z 41, m/z 55, m/z 67 and m/z 83 is the same for both spectra, and the presence of m/z 41 can indicate a terminal methyl group, since terminal alkenes readily form this fragment (Figure 31). The molecular ion of m/z 182 is detected and a shift of +2 is seen in the total mass and also in larger fragments like m/z 164 (M-18, loss of water) and m/z 151. It is probably a C12 compound, a double methylated terpenoid, and with only mass spectral data it is impossible to affirm the position of the methyl groups. The similarity of this mass spectrum with the one from 4,4-dimethyl-isoprenol, also containing m/z 55 at 100% and m/z 83 at 50%, gives a hint about the origin of the two methyl groups present in this C12 molecule.
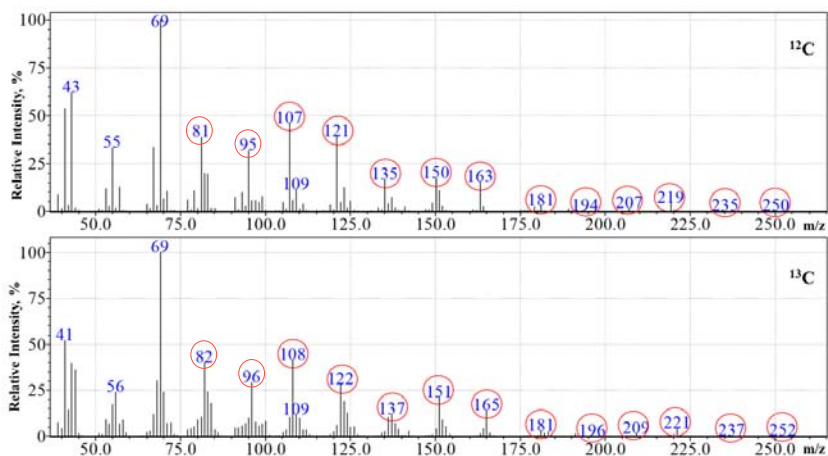
Figure 32. Mass spectra of **14** (C17 isoprenoid alcohol) from cultures of *E. coli* + pLD-03 + pMK-17 grown with L-methionine (upper MS) or L-methionine-(methyl-^{13}C) (lower MS). Red circles mark the fragments which contain the transferred methyl group(s).

The mass spectrum of **14** indicates that the compound is a C17 terpene alcohol bearing two additional methyl groups. The molecular ion is seen at m/z 250 which agrees with the molecular mass of such a compound (Figure 32). Shifts of +2 are seen for larger fragments, like m/z 235, m/z 219, m/z 207, m/z 194, m/z 163 and m/z 135. The ratios between smaller fragments like m/z 55, m/z 69 and m/z 81 resembles the

mass spectra of 4-methyl-geraniol and especially 4-methyl-farnesol. A list of the compounds found with additional methylations in the culture medium of *in vivo* experiments is depicted in Table 11.

Table 11. Overview of compounds detected through isotope labeling *in vivo* experiments with *E. coli* strains expressing the mevalonate pathway and IPPMT from *S. monomycini*. Compounds were extracted from culture medium with SBSE during cultivation for 24h and desorbed into GC-MS for analysis. Mass spectra of peaks from cultures containing labeled and non-labeled methionine were compared. Compounds that had fragments showing mass shifts are listed. Shift of +1 indicate one methyl group in the molecule, shift of +2 indicate two methyl groups in the molecule.

RT (min)	Compound Name	^{13}C shift
6.136	**7b** (4,4-dimethyl-prenol)	+2
6.905	**6b** (4,4-dimethyl-isoprenol)	+2
8.260	**15** (C11)	+1
10.449	**16** (C11)	+1
12.126	**10b** (4-methyl-geraniol)	+1
12.457	**17** (C11)	+1
12.779	**12b** (8-methyl-geraniol)	+1
13.347	**13** (C12)	+2
13.687	**18** (C12)	+2
13.730	**19** (C11)	+1
14.087	**20** (C12)	+2
17.827	**11b 1** (4-methyl-farnesol isomer 1)	+1
18.055	**11b 2** (4-methyl-farnesol isomer 2)	+1
18.103	**11b 3** (4-methyl-farnesol isomer 3)	+1
18.189	**14** (C17)	+2
18.417	**21** (C16)	+1
18.494	**22** (C16)	+1
18.769	**23** (C16)	+1
18.808	**24** (C16)	+1

The use of isotope labeling helped on the detection and identification of longer terpene molecules containing methyl groups. These methyl groups originated from the condensation of methylated precursors 4-methyl-IPP and 4-methyl-DMAPP with IPP and DMAPP, catalyzed by the IPPMT. The specific substrates and pathways leading to each one of the identified products will be tackled in the Discussion.

Besides the C6 and C7 compounds already detected during *in vitro* and *in vivo* experiments, six C11 compounds were detected, from which two of them had their structures elucidated (4-methyl-geraniol and 8-methyl-geraniol). Three peaks corresponding to compounds with supposedly 12 carbon atoms were detected, and their chemical structures remain unclear. One of the peaks with later retention times most probably belongs to a terpene alcohol with 17 carbon atoms in its structure and therefore two methyl groups added. Seven peaks in the chromatogram corresponded to terpene alcohols containing 16 carbon atoms in their structures, from which one chemical structure was elucidated (4-methyl-farnesol). Three different peaks, with similar mass spectra and slightly different retention times, were associated with 4-methyl-farnesol. The GC-MS analysis of the reference compound 4-methyl-farnesol also produced a chromatogram containing more than one peak, with four peaks in slightly different retention times and similar mass spectra. It is assumed therefore that the different peaks, observed during *in vivo* experiments, with high similarity to 4-methyl-farnesol, represent different isomers of this compound.

3.5.3 Investigation of in vivo production of terpenes by wild strain *Streptomyces monomycini*

Streptomyces monomycini, the source of the gene of the methyltransferase that was used in this study, is a species belonging to the phylum Actinobacteria. They are soil organisms known for their capability of producing antibiotics and odor metabolites like geosmin and 2-methyl-isoborneol. Aiming to determine the participation of the methyltransferase on the production of terpenoid compounds, the wild type strain of *Streptomyces monomycini* was cultivated in complex solid medium and had its volatiles extracted and subjected to analysis by HS-SPME-GC-MS.

.
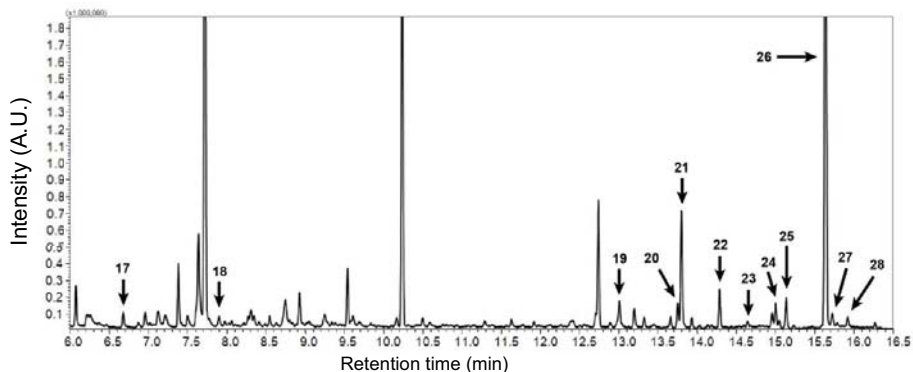
Figure 33. Terpenoids produced by *S. monomycini*. Volatile terpenoids produced by *S. monomycini* grown for 4 days at 28 °C cultivated on GYM agar were identified by headspace-SPME-GC-MS analysis. alpha-pinene (**17**), 3-carene (**18**); delta-elemene (**19**); (-)-beta-elemene (**20**); 2,3-Dihydro-γ-ionone (**21**); gamma-elemene (**22**); *cis*-beta-guaiene (**23**); delta-selinene (**24**); alpha-bulnesene (**25**); (4aR,8aS)-4a-methyl-1-methylene-7-(propan-2-ylidene)-decahydronaphtalene (**26**); selina-3,7(11)-diene (**27**); germacrene (**28**) (Drummond et al. 2019).

The identification of compounds detected by GC-MS analysis revealed that besides compound **21** which is a C13 terpene, all terpenoids are canonical, bearing a multiple of 5 carbon atoms in their structures (Figure 33). Compound 2,3-Dihydro-γ-ionone (**21**) is probably derived from carotenoid cleavage, since *S. monomycini*'s genome bears coding sequences for enzymes related to carotenoid biosynthesis as well as carotenoid degradation. The protein WP_030022897.1 (sequence on Appendix) carotenoid oxygenase encoded in the genome of *S. monomycini* is a marker enzyme for the carotenoid biosynthetical pathway, indicating the presence of carotenoid synthesis pathway within this organism. Therefore, no compounds bearing unconventional prenyl chains were detected in this experimental setup.

The results abovementioned regarded the headspace analysis from solid *S. monomycini* cultures (in agar medium). When liquid cultures were analyzed by the same methods, the detection of volatiles displayed a different chromatographic profile.

The number of compounds produced by the organism cultivated in liquid medium was much lower when compared to the ones from cultivation in solid media. The product profile also differed qualitatively in a comparison between the two cultivation methods. At least one C11 compound, peak 1 with similarity to 2-methyl-linalool (Figure 34), could be detected in the headspace of liquid cultures. Even though the similarity of this compound with the reference compound 2-methyl-linalool was as high as 87%, it is not enough to positively identify its chemical structure. However, sufficient signs from the mass spectrum allow the affirmative identification as a terpenoid. The high similarity with a C11 compound furthermore points to a methylated monoterpene of unknown structure.
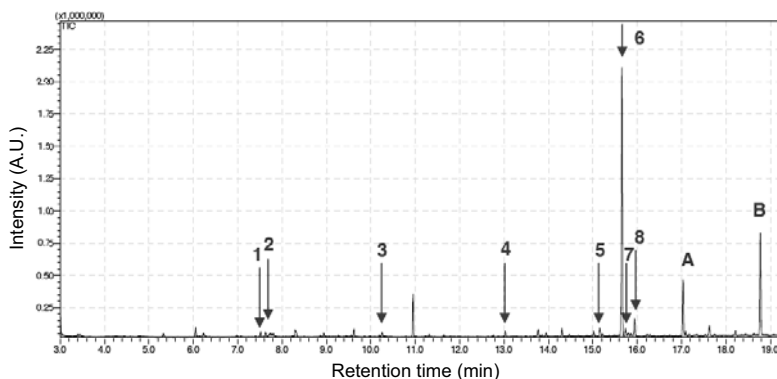
Figure 34. Volatile terpenoids produced by *S. monomycini* cultivated on GYM agar were identified by headspace-SPME-GC-MS analysis. Peaks 1, 2, 4, 5, 6, 7 and 8 have mass spectra with characteristic fragments from terpenoids (see Figure 35). A and B correspond to siloxanes.

93
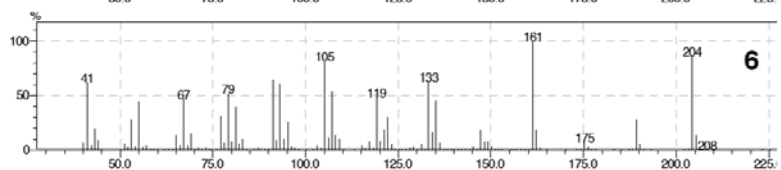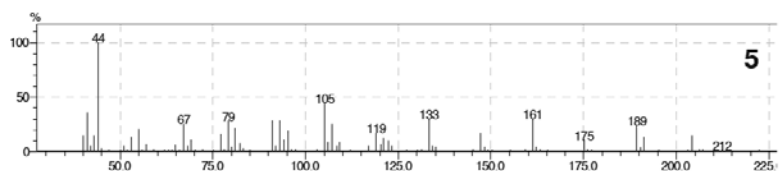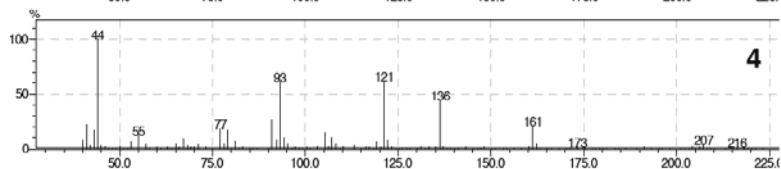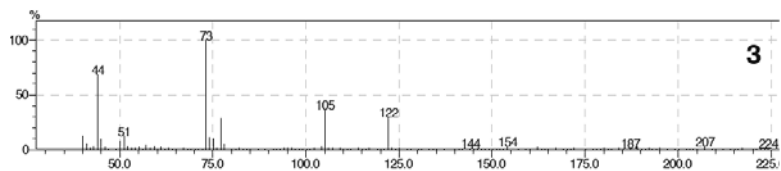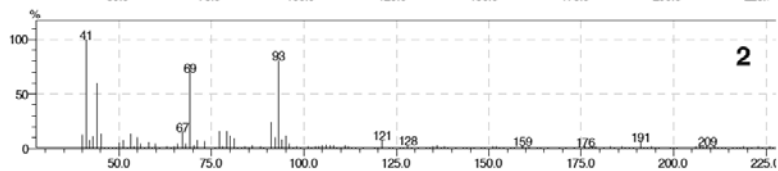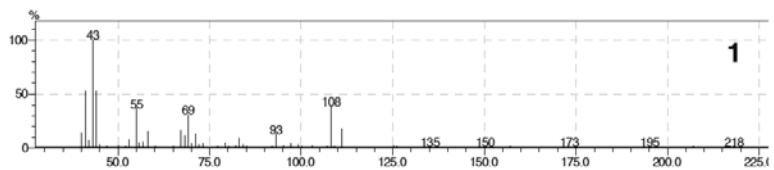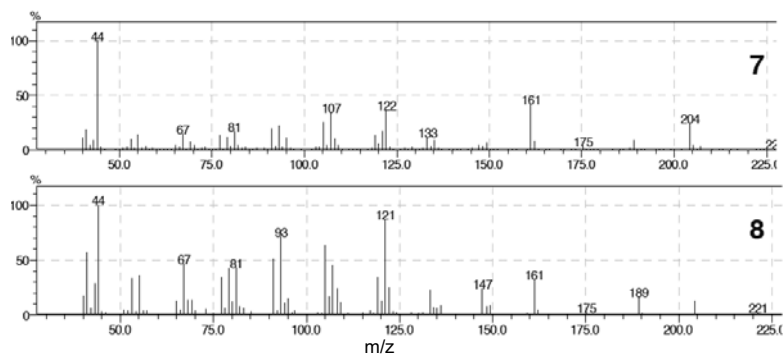
Figure 35. Mass spectra of compounds corresponding to the peaks shown in the chromatogram in Figure 34. Relative abundance of fragments m/z is shown in percentage of the base peak. Volatiles were collected from the headspace of liquid culture of *S. monomycini*.

Mass spectrum of peak 1 shows fragments of low intensity at m/z 150 and m/z 135, which are in accordance with a methylated monoterpene (C11). The similarity with 2-methyl-linalool is of 87% which is not high enough for clear identification of the compound but is a strong indication of a C11 terpene. The spectrum of peak 2 had similarity to beta-myrcene (73%), but some of its m/z values, specifically m/z 67, m/z 69 and m/z 93 have intensities similar to the monoterpene citronellal. Peak 3 had no high similarity with any mass spectra from the database (lower than 70% similarities were discarded). For peak 4 spectrum the best matches were 4-terpinenyl acetate and 4-carene, both with 83% similarity. However, because of the disparities in the retention index, which is later for this compound than the two monoterpenoid matches, could be plausible that this compound is in fact a sesquiterpenoid. The spectrum of peak 5 has 80% similarity with gamma-selinene, a sesquiterpene, and the fragments m/z 189, m/z 161, m/z 119 and m/z 67 are also characteristic for C15 terpenoids. The highest peak in the chromatogram, peak 6, has a mass spectrum with high similarity to γ-selinene (93%) and patchoulene (91%). Even though the retention index does not match the ones from these two compounds, there are other indications that this compound is a sesquiterpenoid, such as the fragments m/z 161, m/z 147, m/z 119, m/z 93 and m/z 67. The mass spectrum of peak 7 resembles the one from peak 6 regarding its m/z values 204 and 161, a similarity which could indicate a sesquiterpenoid; and similarities with γ-selinene (79%) and selina-3,7-diene (78%) were found. Regarding

peak 8, the best hits for its mass spectrum were γ-elemene (88%) and γ-selinene (85%), and m/z values 161, 147, 93 and 67 could also indicate a sesquiterpenoid.

The analysis of *S. monomycini* cultures were done together with MSc. Ina Huth who collaborated with this research project during her master thesis. Using the LC-MS in SIM mode for detection of pre-selected values typical to terpene fragmentation patterns, a screening was made focusing on C10, C11, C15 and C16 compounds. For the detection of monoterpenoids, the selected ions for monitoring were fragments with m/z values 151, 153, 154 and 157, considering the common masses of monoterpene alcohols, aldehydes and ketones, plus one proton due to the positive ionization. For C11 compounds the m/z values were 165, 167, 169 and 171. For sesquiterpenes the m/z values were 215, 217, 233, 235 and 237. For C16 compounds the m/z values were 229, 231, 233, 235 and 237. The chromatograms generated with this method did not show any signals for C10 or C11 compounds. Small peaks corresponding to compounds eluted at different retention times can be seen at the m/z selection for C10 compounds, and they could be related to the presence of C10 isoprenoids (Figure 36). As for C11 compounds, no signal was detected for any of the selected ions (Figure 37). For C15 sesquiterpenes, peaks were detected at m/z 219 and m/z 215 (Figure 38). The results indicated the presence of C16 compounds by the identification of fragments with m/z 229 (Figure 39), although exact structures or molecular masses could not be confirmed.
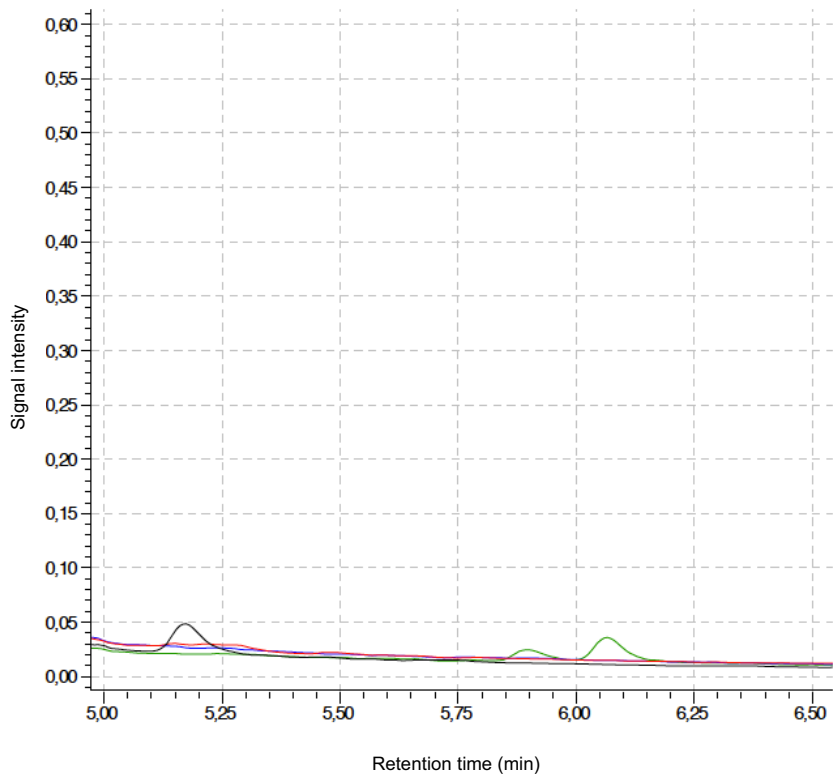
Figure 36. LC-MS SIM analysis of the supernatant from *S. monomycini* liquid culture, m/z values were chosen to detect monoterpenoids. Black: m/z 151, red: m/z 153, blue: m/z 155, green: m/z 157.
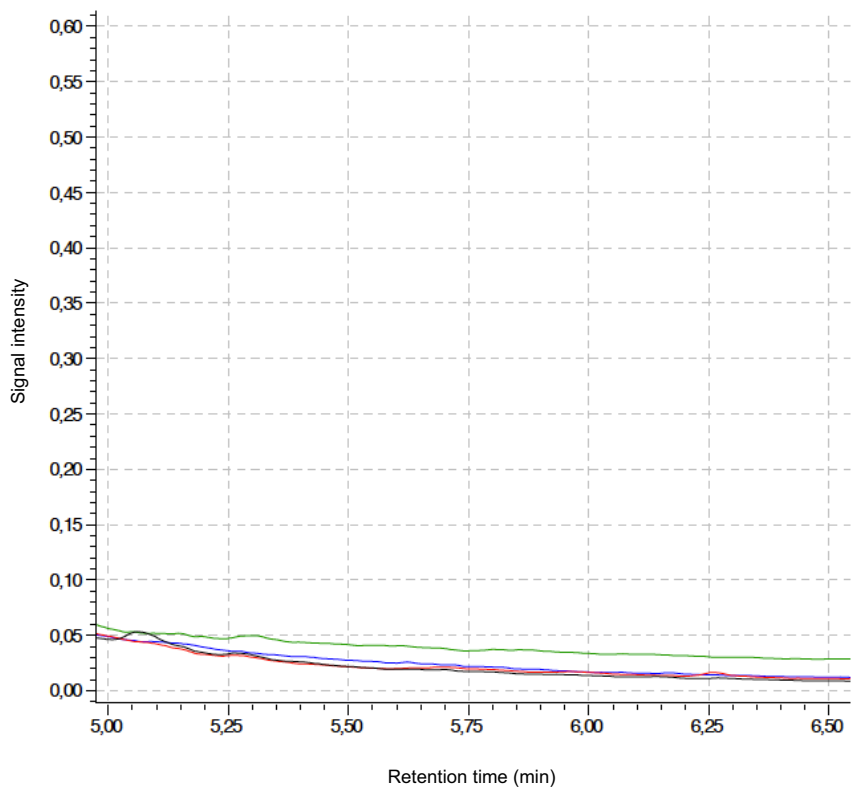
Figure 37. LC-MS SIM analysis of the supernatant from *S. monomycini* liquid culture, m/z values were selected for monitoring methylated monoterpenoids. Black: m/z 165, red: m/z 167, blue: m/z 169, green: m/z 171.
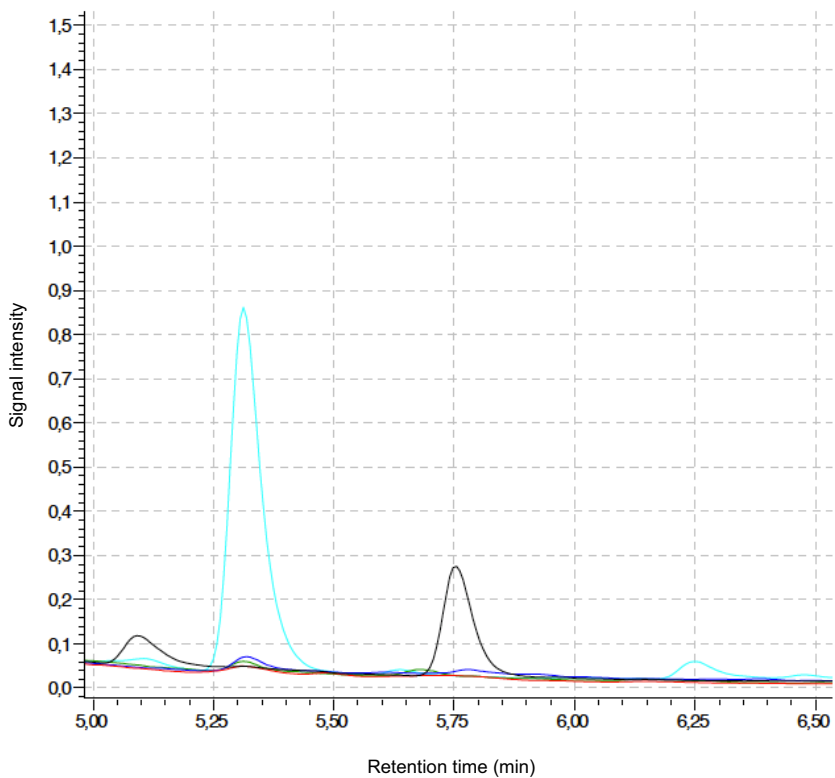
Figure 38. LC-MS SIM analysis of the supernatant from *S. monomycini* liquid culture, m/z values were selected for monitoring sesquiterpenoids. Black: m/z 215, red: m/z 217, blue: m/z 219, green: m/z 221, turquoise: m/z 223.
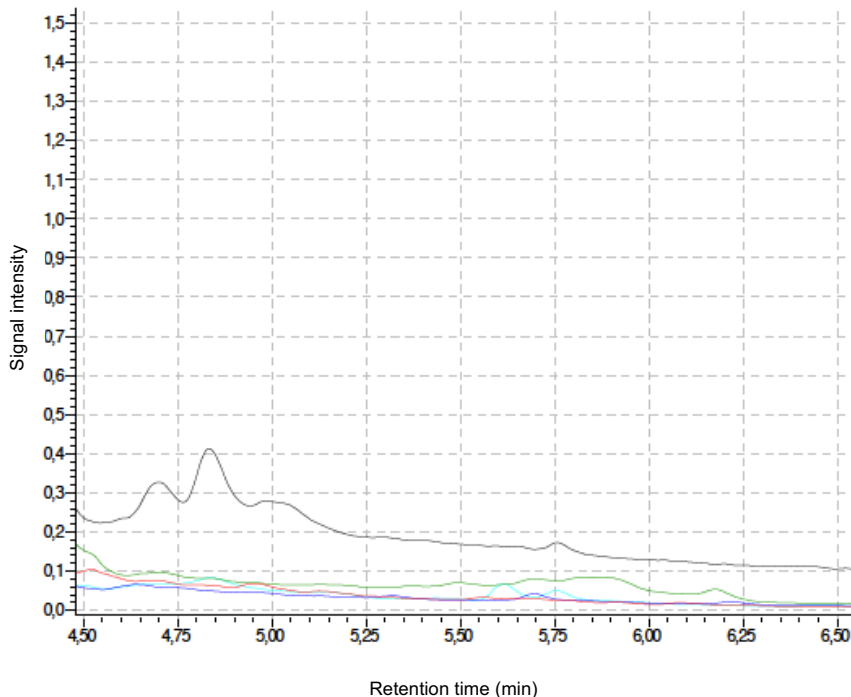
Figure 39. LC-MS SIM analysis of the supernatant of liquid culture from *S. monomycini*. The m/z values were selected for monitoring methylated sesquiterpenoids. Black: m/z 229, red: m/z 231, blue: m/z 233, green: m/z 235, turquoise: m/z 237.

3.6   Enzymatic assays with IPPMT and *in situ* SAM and SAM analogs formation

To test the substrate specificity of the IPPMT regarding the transferred side group of SAM, the cofactor SAM and SAM analogs were produced *in situ* using different substrates and the enzyme methionine adenosyl transferase (MAT) as catalyst. Recapitulating, MAT catalyzes the addition of one adenosyl group from ATP to methionine generating S-adenosyl-methionine (SAM). SAM can then be employed as cofactor in SAM-methyltransferase reactions as methyl group donor (Figure 40).
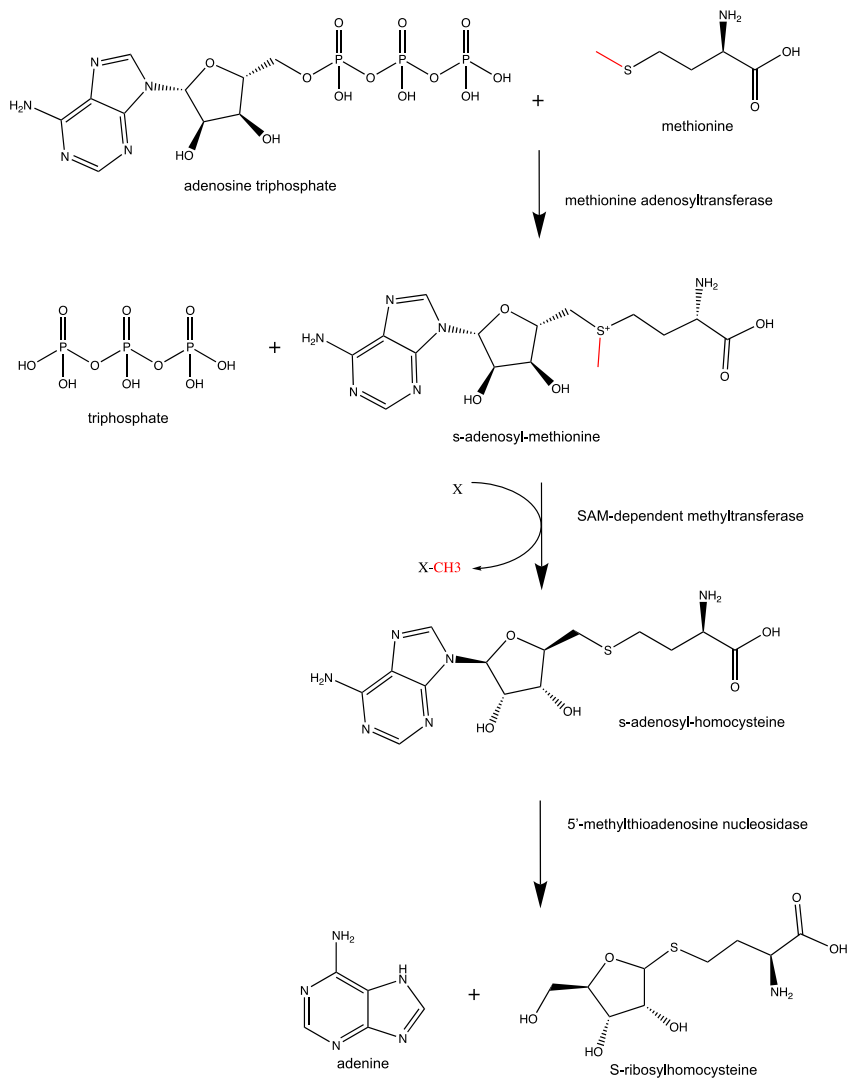
Figure 40. SAM formation from ATP and methionine catalyzed by the enzyme methionine adenosyl transferase. SAM is used for transfer of methyl group to substrate (X), catalyzed by a SAM-dependent methyltransferase, yielding s-adenosyl-homocysteine and methylated product (X-CH$_3$). The byproduct SAH is then cleaved by MTAN yielding adenine and S-ribosylhomocysteine.

Since the accumulation of the byproduct S-adenosyl-homocysteine (SAH) can inhibit the methyltransferase reaction, on the *in vitro* IPPMT assays with *in situ* cofactor formation by MAT, an SAH-degrading enzyme, 5'-methylthioadenosine nucleosidase (MTAN) was used concomitantly. MTAN catalyzes the cleavage of the glycosidic bond in S-adenosylhomocysteine to adenine and S-ribosylhomocysteine (Figure 40).

3.6.1 *In situ* SAM production with TKMAT

The *in situ* formation of SAM is an attractive alternative for enzymatic methylation of substrates. Even though ATP is unstable, methionine is a very cheap substrate, produced in bulk amounts and sold by reasonable prices per kilo. The low prices and availability of the two MAT substrates make the use of this enzyme an interesting substitute to the use of SAM in methylation reactions. In a system where ATP can be provided by enzymatic reaction or regeneration, the substitution of stoichiometric amounts of SAM by its *in situ* formation is a much more viable option for upscaled processes. In this sense, efforts have been made to create a biomimetic SAM cycle, in which SAM is created and regenerated *in situ*, with only methionine and adenosine being added in stoichiometric amounts (Mordhorst et al. 2017).

It is known that MAT can also accept other substrates instead of methionine, like for example ethionine, and that the production of SAM derivatives can be used as a tool for differential alkylation of substrates (Sommer-kamann et al. 2017; Mordhorst et al. 2017). The product of the reaction catalyzed by MAT when ethionine and ATP are used as substrates, is SAE, or S-adenosylethionine. In this case, if SAE is used by a methyltransferase, instead of having a methyl group being transferred to the substrate, an ethyl group will be the added as rest group to the molecule (Figure 41). In a similar fashion, if L-methionine-(methyl-^{13}C) is offered as substrate with ATP, MAT converts it to ^{13}C-SAM, which can then be used for ^{13}C-labeled methylation of substrates (Figure 41).
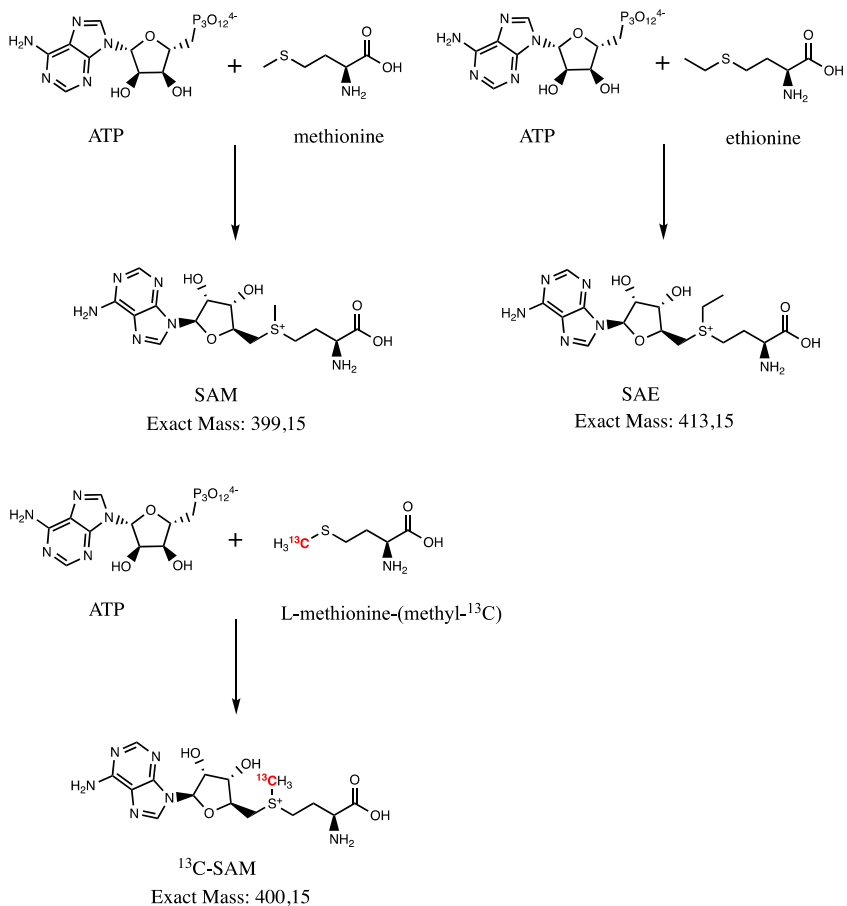
Figure 41. *In vitro* formation of SAM analogs by MAT. The group to be transferred by the methyl transferase, attached to the sulfur atom, depends on the substrate used and can be either e.g. CH_3 (methyl), CH_2-CH_3 (ethyl), or $^{13}CH_3$ (^{13}C-methyl).

3.6.2 *In vitro* formation of unconventional terpenes using SAM analogs

The addition of the enzyme MAT into methyltransferase reaction systems brings the possibility of creating SAM analogs for the transfer of different groups to the substrate. Aiming at the development of a combined one-pot reaction of SAM generation and methyl terpene formation, MAT from *Thermococcus kodakarensis* (TKMAT) and MTAN (5'-Methylthioadenosine Nucleosidase) enzymes were added to the *in vitro* MT reaction. The enzyme MTAN had the function of degrading the inhibitory byproduct SAH. The reactions were performed in a similar fashion to the methyltransferase assays described above, but this time, instead of SAM, the substrates ATP and either L-methionine, L-methionine-(methyl-^{13}C) or ethionine were added. The prenyl diphosphate substrate IPP remained the same and reaction conditions of time and temperature were also comparable (30 ℃, 24 h incubation). A schematic overview of the reaction with the three enzymes is depicted in Figure 42. Phosphatase solution was added at the end of incubation period for cleavage of phosphate groups and release of volatile compounds. The volatile compounds in the headspace of the assay were extracted with SPME and analyzed via GC-MS.
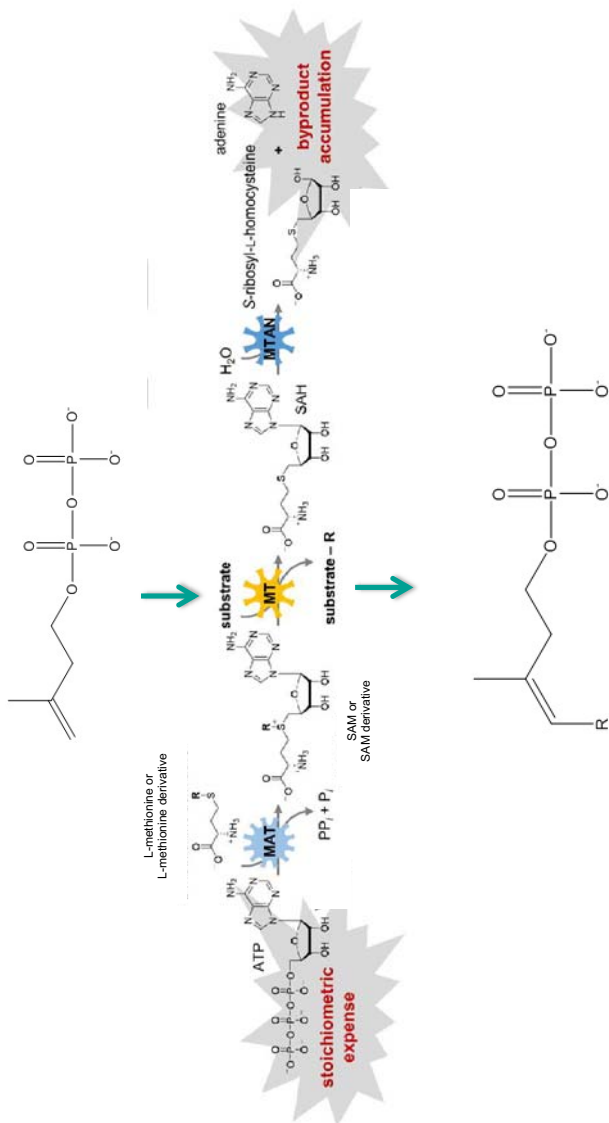
Figure 42. One-pot reaction with three enzymes, MAT, MT and MTAN, for in situ SAM formation, substrate methylation and SAH degradation (adapted from Mordhorst et al., 2017).

In parallel, incubation of IPPMT with IPP and SAM was performed, as a positive control to confirm the functionality if the IPPMT under the conditions used here. The peak at 5.55 minutes retention time was confirmed to be corresponding to the compound (*E*)-4-methyl-isoprenol. Mass spectra of the observed peaks are comparable for positive control and SAM formation (Figure 44 a and b), showing equivalence of product 4-methyl-isoprenol. Not only (*E*)-4-methyl-IPP was formed when methionine was provided as substrate, but also ^{13}C-labeled (*E*)-4-methyl-IPP was formed when ^{13}C-labeled methionine was available for TKMAT, which was evidenced by the peak in red at 5.55 minutes (Figure 43), with mass spectra corresponding to [13]C-(*E*)-4-methyl-isoprenol (Figure 44 c). By observing the mass spectrum of this compound, it is clear the resemblance with the one from (*E*)-4-methyl-isoprenol (Figure 44 a and b).
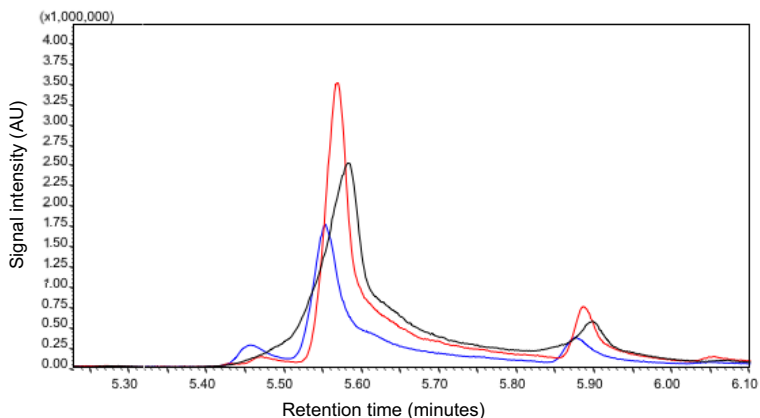
Figure 43. *In vitro* production of methylated terpenes with *in situ* SAM formation. Total ion chromatogram of HS-SPME-GC/MS of volatiles from the headspace of *in vitro* assays using IPPMT and *in situ* SAM formation with TKMAT. IPP was used as substrate, as well as ATP and L-methionine (blue) or L-methionine-(methyl-^{13}C) (red). In black, the positive control was an incubation of IPP and SAM with IPPMT.
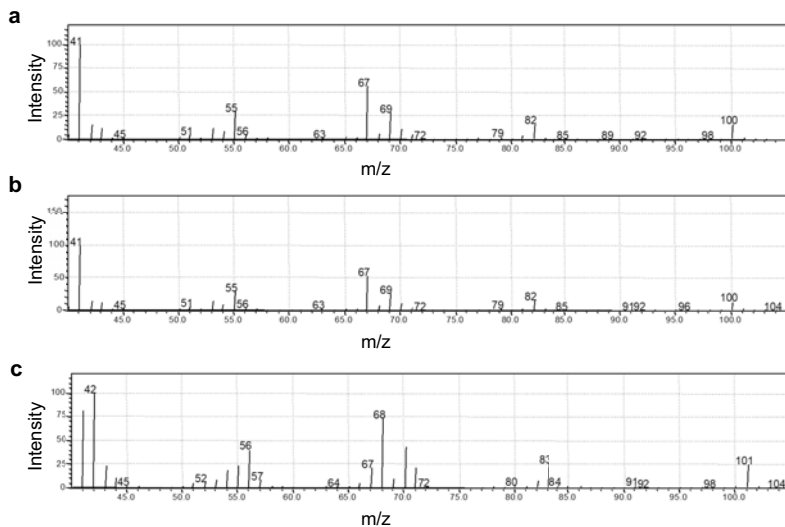
Figure 44. Mass spectra corresponding to the peaks shown on the chromatograms from Figure 43. a) (*E*)-4-methyl-isoprenol, formed by positive control reaction with IPP + SAM and IPPMT (black chromatogram); b) (*E*)-4-methyl-isoprenol, formed by one-pot reaction with ATP, methionine, TKMAT and IPPMT (blue chromatogram); c) ^{13}C-(*E*)-4-methyl-isoprenol, formed by one-pot reaction with ATP, L-methionine-(methyl-^{13}C), TKMAT and IPPMT (red chromatogram).

By comparing the mass spectra of the same peak at RT=5.55 minutes from the assay using normal L-methionine with the one using L-methionine-(methyl-^{13}C), it was possible to observe a shift of +1 in most of the fragments, as shown in Figure 44. The shift of +1 in the mass spectrum confirms the transfer of the ^{13}C carbon from labeled methionine to the final product, showing activity of TKMAT and the methyltransferase in the one-pot reaction. The shift of +1 present in the molecular ion of m/z 101 and in the fragments m/z 83, m/z 68, m/z 56 and m/z 42 evidence the presence of a ^{13}C-labeled methyl group in the compound, in comparison with the mass spectrum from 4-methyl-isoprenol which shows molecular ion of m/z 100 and fragments m/z 82, m/z 67, m/z 67, m/z 55 and m/z 41.

The addition of an ethyl group to IPP by the methyltransferase, instead of methyl, would in theory create a C7 prenyldiphosphate, which would originate a C7 terpene

106

alcohol after the addition of phosphatase solution at the end of the reaction cascade. In the case of a double ethylation, however, a C9 terpene alcohol could be formed. The presence of compounds with the respective masses of 114 and 142 associated with C7 and C9 alcohols could not be confirmed by analyzing the corresponding mass spectra from the chromatogram. In fact, for experiments using ethionine as substrate, a peak for isoprenol was detected (not shown), showing that the substrate IPP was not converted by the methyltransferase, but rather dephosphorylated in its original form. Therefore, it was assumed that ethylated isoprenoids were not produced with this particular experimental setup.

To confirm the production of SAM analogs by MAT, the *in vitro* reactions were analyzed by LC-MS. When performing the *in vitro* assays, samples were taken from the mixture before addition of the methyltransferase and before the addition of phosphatase solutions. The samples were frozen in acetonitrile for protein precipitation and posteriorly analyzed by LC-MS, in order to enable the detection of non-volatile compounds.

3.6.3 SAM and SAM analogs detection

LC-MS analysis was applied to detect substrates and products of the reactions catalyzed by the enzyme MAT from *Thermococcus kodakarensis*. Compounds present in the supernatant of *in vitro* assays were subjected to separation in the mobile phase of a liquid chromatography system, and then ionized for identification. Measurements were performed by selecting ions with masses specific for the compounds of interest (exact mass +1 due to ionization process) and subjecting them to fragmentation to analyze signals of the predominant species of the fragmentation pattern. The Multiple Reaction Monitoring (MRM) mode allowed the analysis and quantification of parent ions and the formed fragments specific for each target molecule. For each compound, particular fragment masses were analyzed, together with the m/z corresponding to molecular ion. In some cases, the same m/z was common for different molecules, which is understandable since their structures are similar (Figure 45). However, *in vitro* assays containing methionine, ethionine or ^{13}C-methionine as substrates were performed separately, thus excluding the possibility

that different SAM analogs could be produced simultaneously within one reaction pot. Preliminary tests were performed with LC-MS scanning mode to determine the fragments detectable within the samples. The selected analyzed species for each one of the molecules are listed in Table 12.

Chemical Formula: $C_{10}H_{12}N_5O_3^{\bullet}$
Exact Mass: 250.09

Chemical Formula: $C_5H_{11}NO_2S^{\bullet+}$
Exact Mass: 149.05
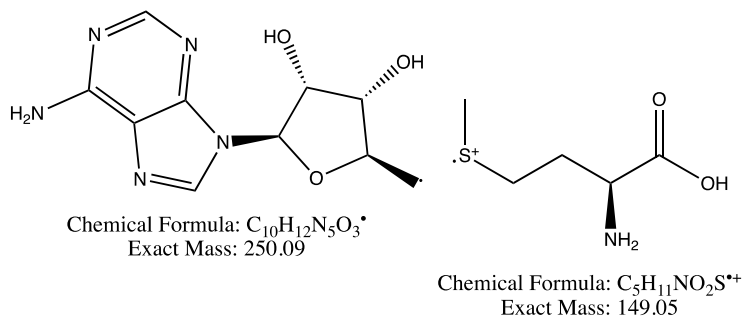
Figure 45. Fragmentation of S-adenosyl-methionine. The fragmentation on the sulfur atom generates a fragment with mass of 250 (left) that is common to SAM, SAE and ^{13}C-SAM due to the similarity in their structures.

Table 12. Multiple Reaction Monitoring (MRM) selection of fragments detected by LC-MS analysis of *in vitro* assays using TKMAT with ATP and different substrates (L-methionine, L-methionine-(methyl-^{13}C) or ethionine) for the formation of SAM and SAM analogs.

Compound	Precursor (m/z)	Fragment 1	Fragment 2	Fragment 3	Mode
ATP	508,00	136,10	409,95	427,90	positive
Methionine	149,95	56,05	104,05	61,05	positive
[13]C-Methionine	150,95	56,05	105,05	61,95	positive
Ethionine	163,95	75,05	56,10	147,05	positive
SAM	400,15	251,05	250,05	136,10	positive
[13]C-SAM	401,15	251,05	250,05	136,10	positive
SAE	414,15	251,05	250,05	136,10	positive

The analysis of compounds through LC-MS revealed the presence of not only SAM but also of SAM analogs in the respective assays using different substrates. Chromatograms, showing time of elution and peak intensities, depicted the relative abundance of fragments of each compound present in a given sample (Figures 46, 47 and 48).

When methionine and ATP were supplied as substrates for TKMAT, the formation of SAM could be detected (Figure 46). In the chromatogram, this is evidenced by the peaks with m/z 251, m/z 250.05, and m/z 136.10. The presence of methionine, which was not converted to SAM, could also be detected through fragments m/z 56.05, m/z 104.05, and m/z 61.05. The fact that the peaks corresponding to methionine have higher intensity than the ones corresponding to SAM, shows an incomplete conversion by the enzyme, which could have been limited by ATP depletion and will be discussed in a later chapter. The quantification of methionine and SAM was enabled by a calibration curve with reference compounds, and the quantified peaks allowed calculation of conversion percentages (Table 13).
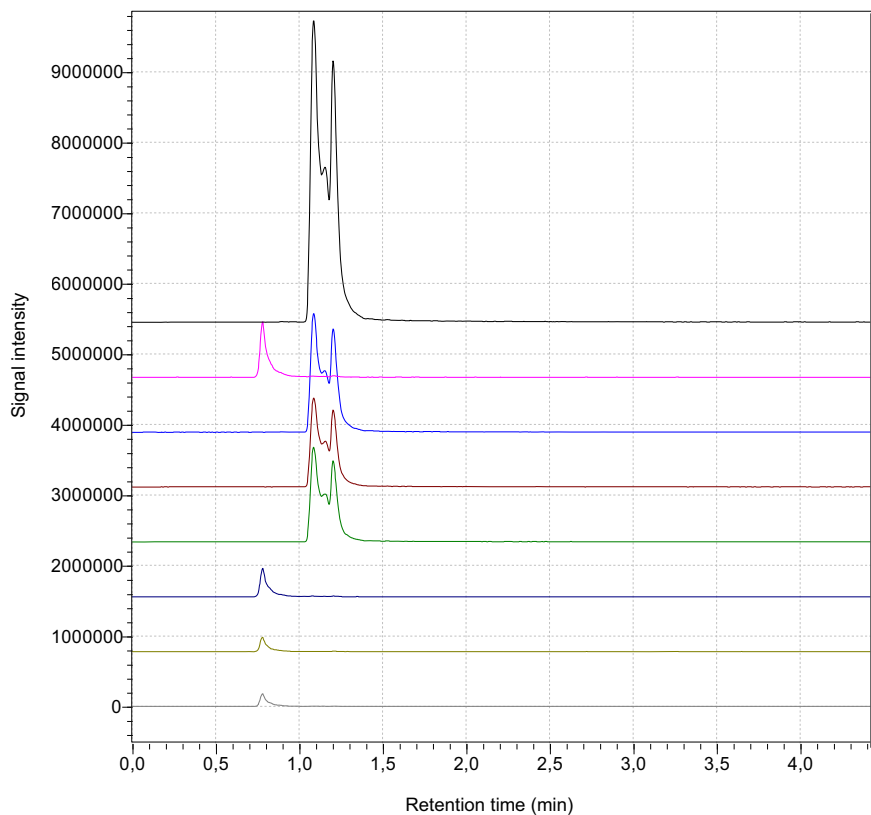
Figure 46. MRM analyses of the supernatant of *in vitro* SAM generation assay using TKMAT, methionine and ATP. For top to bottom, the lines correspond to different fragments detected. Black: TIC methionine, Pink: TIC SAM, Blue: m/z 56.05, Red: m/z 104.05, Green: m/z 61.05, Dark Blue: m/z 251, Light Green: m/z 250.05, Grey: m/z 136.10

Table 13. Concentrations of methionine and SAM in samples from *in vitro* assays (n=2) with TKMAT. Conversion percentage was calculated from quantified substrates and products.

Methionine concentration [µM]	SAM concentration [µM]	Initial methionine concentration [µM]	Conversion [%]
781	479	1259	38
686	535	1221	44

The table shows solutes concentrations obtained by LC-MS analysis of the *in vitro* assays using TKMAT for *in situ* SAM formation, with incubation at 30 °C for 24 h. Each line of the column represents a different sample from separate experiments under the same conditions. The methionine concentration was also measured, allowing the calculation of conversion percentage.

For the detection of compounds with LC-MS, two approaches were used depending on the existence of reference compounds for the analyte or not. For the compounds ATP, methionine, ethionine, L-methionine-(methyl-^{13}C) and SAM, calibration curves enabled the exact quantification of the analyte in the sample. For ^{13}C-SAM and SAE, only the MRM detection of fixed fragments was possible, without quantification of the analyte, because standards for the SAM analogs were not available. The table shows the ions that were selected for analysis for each compound in the samples from *in vitro* assays. For SAM production, a conversion rate of 38 % was calculated, and a total concentration of 478 µM was found in the sample. Considering that the IPP concentration in the sample was 60 µM, the SAM concentration in the assay was 8 times the one from the substrate, which is more than enough for methylations to occur.
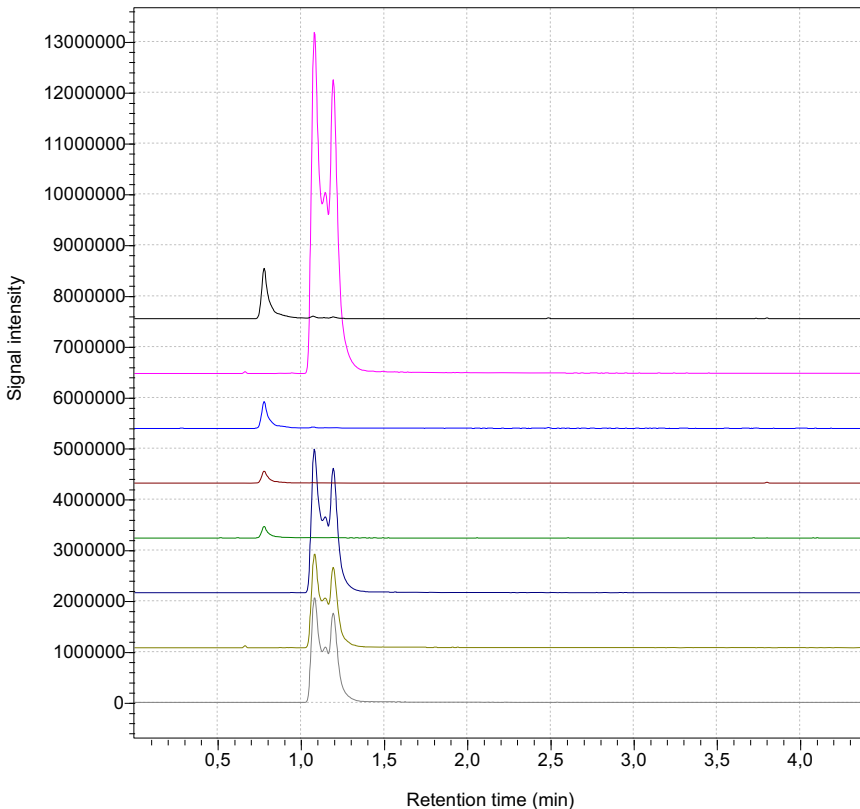
Figure 47. MRM analyses of the supernatant of *in vitro* ^{13}C-SAM generation assay using TKMAT, L-methionine-(methyl-^{13}C) and ATP. From top to bottom, the lines correspond to different fragments detected. Black: TIC ^{13}C-SAM, Pink: TIC ^{13}C-methionine, Blue: m/z 251.05, Red: m/z 250.05, Green: m/z 136.10, Dark Blue: m/z 56.05, Light Green: m/z 105.05, Grey: m/z 61.95

Providing L-methionine-(methyl-^{13}C) and ATP as substrates for the reaction with TKMAT led to the formation of ^{13}C-labeled SAM. The chromatogram in

Figure 47 shows peaks corresponding to L-methionine-(methyl-^{13}C) of m/z 56.05, m/z 101.05 and m/z 61.95. The presence of ^{13}C-SAM in the mixture was evidenced by peaks at m/z 251.05, m/z 250.05 and m/z 136.10. In this case, similarly to SAM

formation, the conversion of the substrates into the product was not complete, what can be seen by differences in the intensity of the peaks. ^{13}C-methionine TIC peak has more than 6 times the intensity of the TIC for ^{13}C-SAM. Since there is no reference compound for ^{13}C-SAM, the quantification through a calibration curve was not possible.
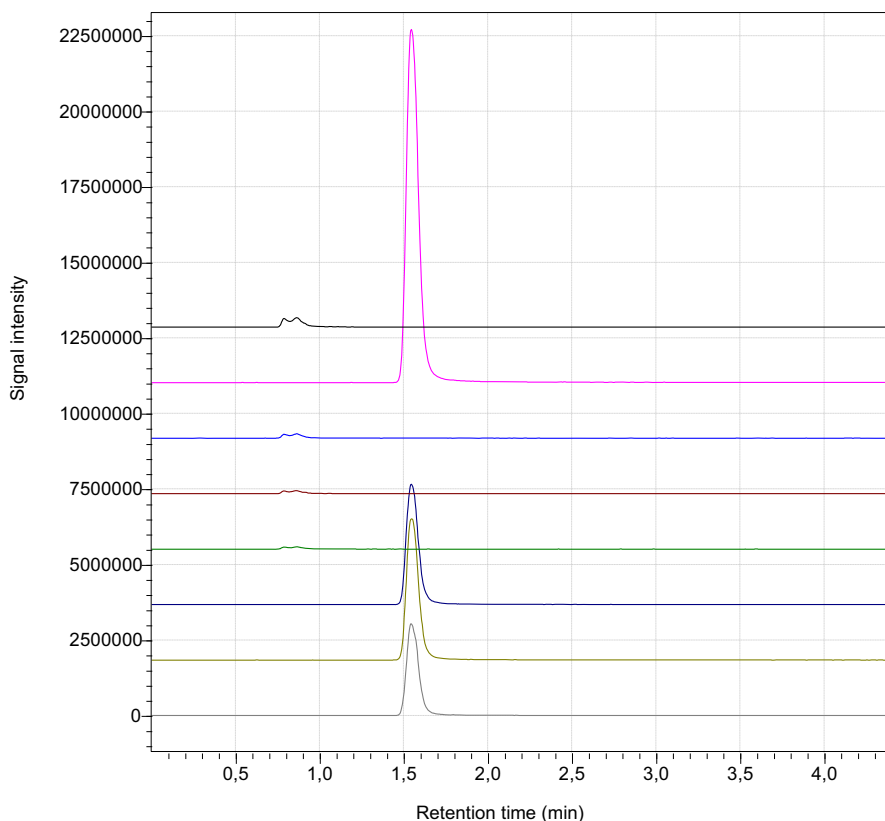
Figure 48. MRM analyses of the supernatant of *in vitro* SAE generation assay using TKMAT, ethionine and ATP. From top to bottom, the lines correspond to different fragments detected. Black: TIC SAE, Pink: TIC ethionine, Blue: m/z 251.05, Red: m/z 250.05, Green: m/z 136.10, Dark Blue: m/z 75.05, Light Green: m/z 56.10, Grey: m/z 147.05

When ethionine and ATP were provided as substrates for TKMAT, S-adenosyl-ethionine (SAE) was formed. The presence of SAE was evidenced by fragments of m/z 251.05, m/z 250.05, and m/z 136.10 (

Figure 48). Ethionine was also detected, with peaks of high intensity for m/z 75.05, m/z 56.10, m/z 147.05. Once again, the conversion of substrates into product was not complete, but in this case, the disparity between TIC of ethionine and TIC of SAE is even higher than the ratios found for the other reactions. The TIC intensity for ethionine is around 70 times higher than the one for SAE. The inexistence of SAE as a reference compound makes it impossible to quantify the amount of solute in the sample, but the relative peak areas give a general idea of how much substrate was converted to product.

With the use of LC-MS analysis, it was possible to confirm the formation of SAM and SAM analogs by TKMAT out of ATP and L-methionine, L-methionine-(methyl-^{13}C) and ethionine.

3.7   Production of non-canonical terpenes using plant terpene synthases

Terpenes are mainly produced by plants. This fact is translated into a vast myriad of terpene synthases known for organisms belonging to this kingdom. However, the methyl-transferase-mediated production of non-canonical terpenes, in a similar fashion to the one from bacteria described in this work with the formation of methyl-prenyl diphosphates, is so far unknown for plants. Reportedly some terpene synthases seem to be promiscuous regarding their substrate and originate a variety of products, or a few main products and several side products. In order to explore the promiscuity of this class of enzymes in plants and expand the possibilities of non-canonical terpene structures formation, the molecule 2-methyl-GPP was used in *in vitro* assays with different terpene synthases from plants, and the volatile products were collected and analyzed via GC-MS. The production of the non-canonical intermediate 2-methyl-GPP and its usage by terpene synthesis for the production of C11 terpenes like 2-methyl-isoborneol and 2-methylene-bornane has been well described for at least *Streptomyces coelicolor*. The chosen terpene synthases are strictly monoterpene

synthases, converting the C10 precursor GPP into C10 terpenes. As a control, we performed the same assays with the natural substrate of the terpene synthases, GPP.

The selected terpene synthases, from *Vitis vinifera, Picea sitchensis* and from *Lavandula angustifolia*, had their genes expressed heterologously in *E. coli* strains, then extracted and purified, and finally used in *in vitro* assays with the above mentioned unconventional prenyl diphosphate. The experiments were conducted together with the student Clara Börnsen during her practical studies within the framework of this thesis.

The terpene synthases tested were limonene synthase from lavender (*Lavandula angustifolia*); geraniol synthase, phellandrene synthase, pinene synthase, ocimene synthase, linalool synthase and nerolidol synthase from wine grape (*Vitis vinifera*); and carene synthase from Picea (*Picea sitchensis*). From the eight enzymes tested, four of them showed promiscuity by accepting the non-canonical prenyl diphosphate as substrate and forming non-canonical terpene alcohols with various structures. The positive results were found for pinene synthase, carene synthase, geraniol synthase and limonene synthase. As negative control, incubations with the substrate without the terpene synthases were made. The enzymes that did not show products when 2-methyl-GPP was used as substrate were demonstrated to have no activity towards the natural substrate GPP.

For at least three of the tested terpene synthases, the generation of C11 compounds was possible when 2-methyl-GPP was used as substrate. Interestingly, not only methylated versions of compounds known to be produced by plants were formed, like methyl-linalool or methyl-limonene, but also 2-methylene-bornane and 2-methyl-isoborneol were detected, both compounds that are usually associated with bacterial production of C11 compounds.

Even though 2-methyl-GPP is not the natural substrate and is not reported to occur in plants, the investigated terpene synthases also accept 2-methyl-GPP. The non-canonical products of this reaction also differ considerably in structure, when compared with the products of the reaction using GPP as substrate.

By comparing the chromatograms of assays containing 2-methyl-GPP as substrate with the ones containing GPP as substrate, the resulting difference in peak presence, absence or intensity allowed to infer the success of the reactions.
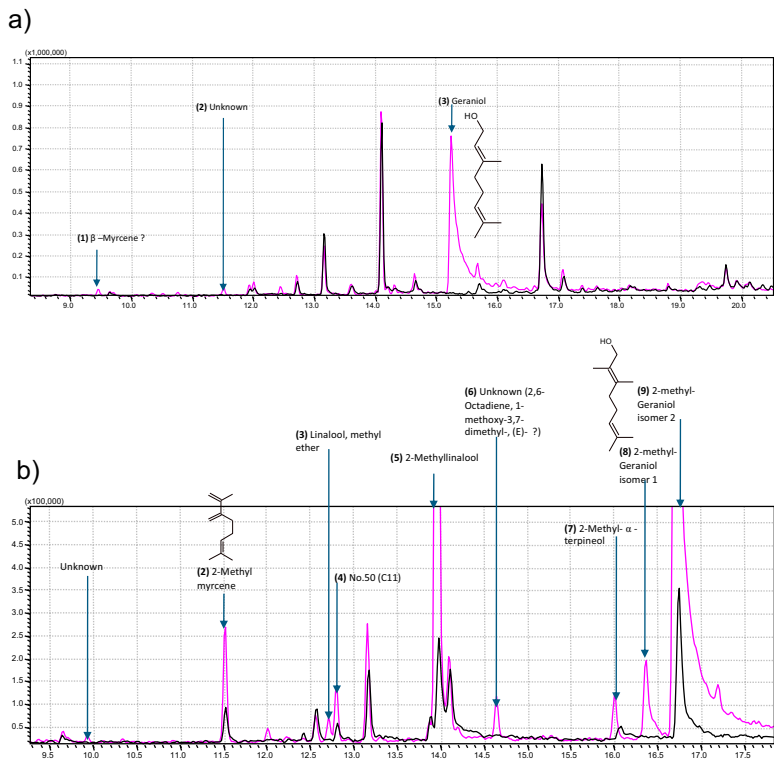
Figure 49. GC-MS analysis of the *in vitro* reactions of geraniol synthase (Ger-TS) with A) GPP and B) 2-methyl-GPP. The chromatograms for the samples (enzyme with GPP or 2-methyl-GPP) are coloured in pink, the respective controls (GPP or 2-methyl-GPP) are coloured in black.

The incubations using geraniol synthase had geraniol as main product when GPP was the substrate (Figure 49 A). In the chromatogram, other smaller peaks can be seen, corresponding to unknown compounds. When 2-methyl-GPP was used as substrate for the enzyme geraniol synthase, the products obtained were diverse. Two peaks with

relatively higher area can be seen, corresponding to 2-methyl-linalool and one of the isoforms of 2-methyl-geraniol. Another seven peaks can be distinguished, 2-methyl-myrcene, linalool methyl ether, 2-methyl-terpineol, the other isoform of 2-methyl-geraniol, and two unidentified compounds. Interestingly, the number of compounds which are detected upon catalysis by the enzyme increase when 2-methyl-GPP is used instead of GPP.

a)
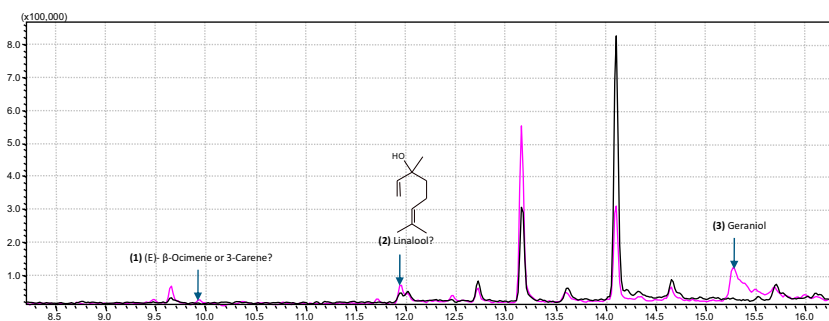
b)
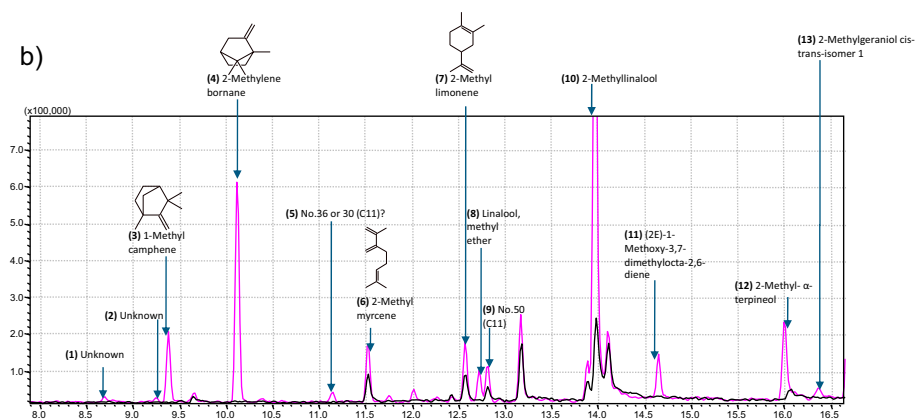
Figure 50. GC-MS analysis of the *in vitro* reactions of limonene synthase (LaLIMS-TS) with A) GPP and B) 2-methyl-GPP as substrates. The chromatograms for the samples (enzyme with GPP or 2-methyl-GPP) are coloured in pink and the respective controls (GPP or 2-methyl-GPP) are coloured in black.

The incubations with limonene synthase yielded unexpected results, since the incubation with its natural substrate GPP resulted in geraniol instead of limonene as product (Figure 50 A). When 2-methyl-GPP was used as substrate, the main product was 2-methyl-linalool, and another eleven peaks could be detected (Figure 50 B). 2-methylene-bornane was the second highest peak, and 1-methyl-camphene, 2-methyl-myrcene, 2-methyl-limonene, linalool methyl ether, (2*E*)-methoxy-3,7-dimethylocta-2,6-diene, 2-methyl-alpha-terpineol, and 2-methyl-geraniol appear as smaller peaks with relatively lower areas, besides two unidentified compounds. Interestingly, when 2-methyl-GPP is used as substrate, the compound 2-methyl-limonene appears, whereas the usage of GPP yielded no limonene.
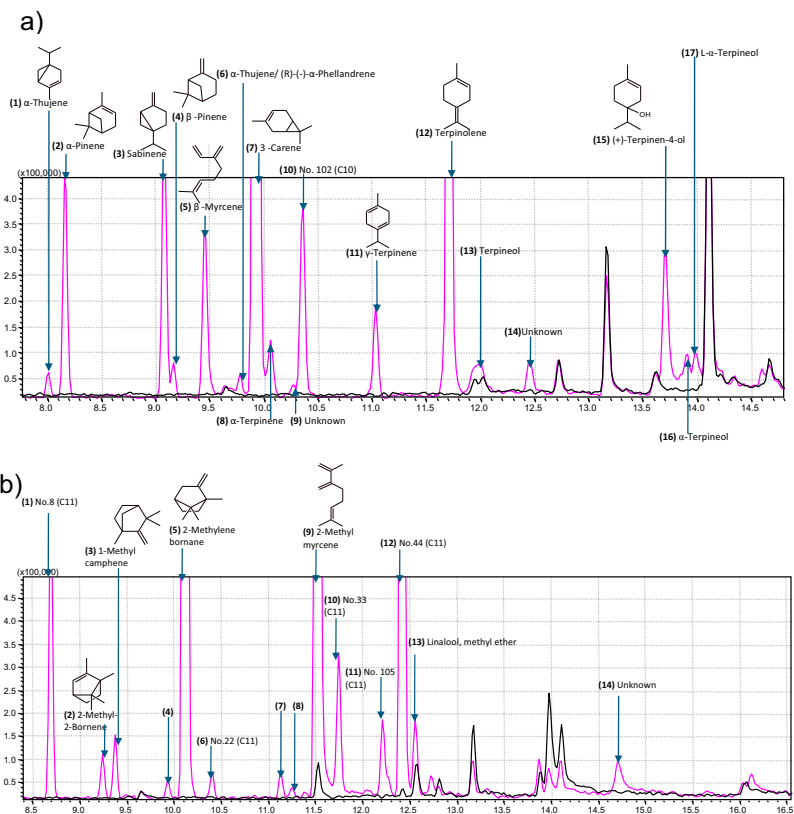
Figure 51. GC-MS analysis of the *in vitro* reactions of (+)-3-carene synthase (Car-TPS) in assay buffer 1 (pH 7) with A) GPP and B) 2-methyl-GPP. The chromatograms for the samples (enzyme with GPP or 2-methyl-GPP) are coloured in pink and the controls (GPP or 2-methyl-GPP) are coloured in black.

The incubation of carene synthase with GPP yielded three main products, (+)-3-carene, terpinolene, and sabinene, agreeing with results already published for this enzyme (Hall et al. 2011). Additionally, 14 minor products were detected (Figure 51 A), including isomers of terpineol and pinene. From the 17 peaks detected, two could not be identified. The incubation with 2-methyl-GPP yielded 14 different C11 terpenes (Figure 51 B). 2-methylene-bornane and 2-methyl-myrcene are among the main products of this reaction, together with a compound of unknown structure (compound No. 44, peak number 12 in the chromatogram of Figure 51 B). 2-methyl-2-bornene, 1-

methyl-camphene and linalool methyl ether were identified as side products of this reaction, and the rest of the compounds could not be structurally determined or identified using the available libraries. Surprisingly, the product blend of this enzyme changes to different peak intensities when a buffer with a lower pH is used (Figure 52).
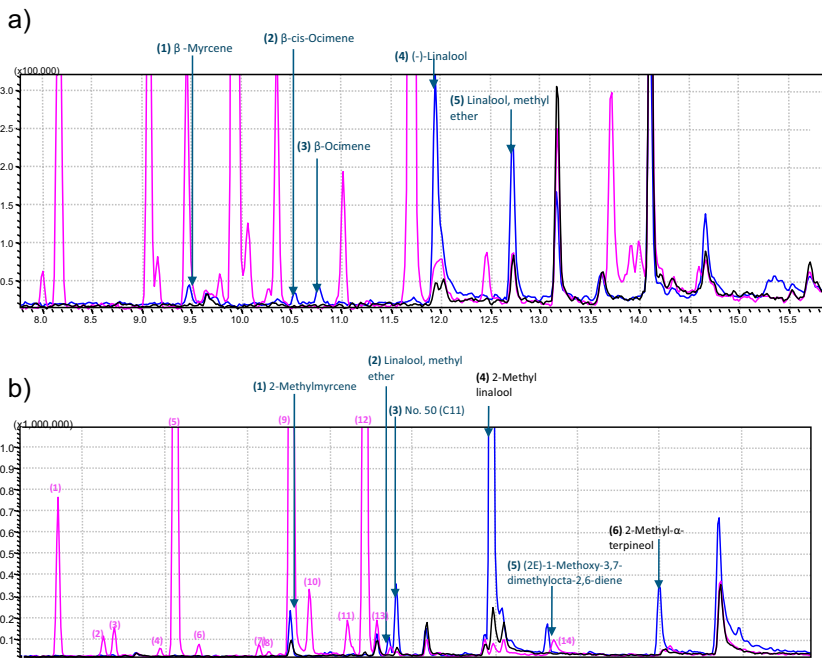
Figure 52. GC-MS analysis of *in vitro* reactions using Car-TS different pHs with A) GPP and B) 2-methyl-GPP as substrate. The chromatograms of the samples (enzyme with GPP or 2-methyl-GPP) are coloured in pink (pH 7) and blue (pH5), the respective controls (GPP or 2-methyl-GPP) are coloured in black.

Terpene formation catalysed by carene synthase yielded a different product mixture when lower pH is applied. When comparing different pHs for the same reaction conditions, compounds show disparity in the peak areas (Figure 52). When GPP was used as substrate, at least 5 compounds showed peak areas which depended on the buffer pH, specially the side products (i.e. compounds which don't belong to the list of

expected, published products for this enzyme). For (-)-linalool, signal intensity and peak area were clearly higher when pH 5 was used, in comparison with pH 7 and control (no enzyme). The chromatograms for in vitro assays using 2-methyl-GPP as substrate also show differences according to the pH used. 2-methyl-myrcene showed higher peak area under pH 7, while 2-methyl-linalool has higher peak area under pH 5. Carene synthase, therefore, showed pH dependency concerning *its in vitro* terpene production under the conditions used in this study.

a)

(1) α-Thujene

(2) α-Pinene

(5) β -Pinene

(8) (R)-(+)-Limonene

(x100,000)

(4) Sabinene ?

(6) β -Myrcene

(10) Linalool

(11) Linalool, methyl ether

(3) Camphene

(7) α-Thujene/ α-Phellandrene

(9) Terpinolene or α-Terpinen

b)

(7) 2-Methyl-2-Bornene /2-Methylenebornane

(5) 2-Methylene bornane

(8) No.30 (C11)

(9) 2-Methyl myrcene

(15) 2-Methyllinalool

(x100,000)

(10) No.33(C11)

(3) 2-Methyl-2-Bornene

(12) 2-Methyl limonene

(16) 2-Methyl- α-terpineol

(4) 1-Methyl camphene

(2) Unknown

(6) No.22 (C11)

(13) No.50 (C11)

(14) No.68 (C11) ?

(1) Unknown

(11) No.44 (C11)

Figure 53. GC-MS analysis of the *in vitro* reactions of pinene synthase (Pin-TPS) with A) GPP and B) 2-methyl-GPP. The chromatograms of the samples (enzyme with GPP or 2-methyl-GPP) are coloured in pink and the respective controls (GPP or 2-methyl-GPP) are coloured in black.

The incubation of pinene synthase with GPP yielded alpha-pinene and (R)-(+)-limonene as main products, with nine other minor peaks as side products (Figure 53

A). The compounds alpha-thujene, camphene, beta-myrcene and linalool are among them. When 2-methyl-GPP was provided as substrate, 2-methyl-myrcene, 2-methylene-bornane and 2-methyl-linalool were conspicuously detected, together with other 13 smaller peaks (Figure 53 B). From these 13, eight are compounds of unknown structure.

1-methyl-alpha-pinene could not be identified but was also not available in our in-house library of reference methyl-terpene compounds. This methylated version of pinene was the expected product of conversion of 2-methyl-GPP by the pinene synthase. The positive identification of methyl-pinene, an undescribed compound, with the NIST database was unsuccessful. In a similar fashion, it was not possible to detect 4-methyl-carene, the theoretical product of conversion of 2-methyl-GPP by carene synthase.

4 Discussion

4.1 Identification of novel methyltransferase coding regions by sequence comparison

The search for prenyl diphosphate modifying enzymes yielded the discovery of an IPP methyltransferase from *Streptomyces monomycini*. This enzyme methylates the universal precursor of terpenes IPP, producing different building blocks that can be further used by prenyl transferases to produce longer terpene precursors with non-canonical structures. The *in vivo* investigation of the operon of *S. monomycini*, and *in vitro* characterization of the reactions performed by the IPPMT opened new horizons for the framework of terpene structure diversity. The characterization of other methyltransferases is a promising approach for the production of unusual terpenes.

While this thesis was being confectioned, new publications came to light involving the discovery of prenyl diphosphate methyltransferases and the use of methylated terpene precursors for the production of non-canonical terpenes (Ignea et al. 2018; Ozaki et al. 2018; Von Reuss et al. 2018; Jameson et al. 2019; Eiben et al. 2019). The fact that this kind of work came to light by different groups almost simultaneously, shows a broadened interest in new terpene molecules and indicates that this type of prenyl diphosphate modifying enzyme is probably widespread in the genomes of prokaryotes. In fact, when searching for sequences on the NCBI database, using the methyltransferase of *S. monomycini* as query, one can already visualize the diversity of methyltransferases present in different genomes (Figure 17) and the potential of exploration of their catalytical properties that is in great part still unveiled.

By comparing the sequences of genes coding for methyltransferases in different organisms, it is possible to verify that there is less than 50% similarity between them. However, similarities can be found in residues which are probably involved with substrate recognition, for example the W194 position of the IPP methyltransferase from *S. monomycini* (Figure 17). A tryptophan residue is also present at the respective position in the protein sequences from *S. argenteolus* and *Rhodococcus fascians*, that

also have been demonstrated to accept IPP as substrate (Radhika et al. 2015; Ozaki et al. 2018). The low sequence similarity therefore does not necessarily translate in a difference of function regarding the enzymatic activity of the encoded protein. In this case, some of the homologous sequences have proven prenylpyrophosphate methyltransferase activity and also use IPP as substrate. The further sequences shown that also have a tryptophan in positions homologous to W194 from *S. monomycini* are speculated to have IPP-methyltransferase activity as well.

Methyltransferases are known to share structural similarities even when the amino acid correspondence between them is low (Liscombe, Louie, and Noel 2012). Even though there is a necessity for binding capability to the same cofactor, SAM, and a commonality on the mechanism for transferring the methyl group, the chemical structures of the proteins occurring in different organisms is largely changeable. The conserved regions for this class of methyltransferases it is considered to be the GxGxG motif (end of first beta-sheet), responsible for binding to SAM carboxy-propyl region, and the highly conserved acidic residue (second beta-sheet), which forms hydrogen bonds with the hydroxy groups of SAM ribose (Liscombe, Louie, and Noel 2012).

4.2 Protein homology modeling of IPPMT from *S. monomycini* and GPPMT from *S. coelicolor*

When comparing the IPPMT sequence with the one from the GPPMT, the identity is less than 40% (106 out of 274), with a total of 58% positives (159 out of 274). The amino acids present in the active site are similar, with a difference regarding the residue from W194 on the IPPMT substrate pocket. The model served as basis for formulating the hypothesis of *S monomycini* MT accepting a different substrate than the GPPMT from *S. coelicolor*, due to the visible difference in pocket size. This hypothesis proved to be true when *in vitro* experiments were performed, which shows how modelling can help in defining experimental setups. For a more detailed reaction mechanism including the functions of amino acid residues on the active site, a crystal structure of the methyltransferase from *S. monomycini* would be neeeded.

4.3   *In vitro* assays with MT from *S. monomycini*

After analyzing the methyltransferase regarding its gene and protein sequence, the enzyme was expressed and tested in *in vitro* assays. The substrate IPP was shown to be converted to a variety of C6 and C7 compounds, with (*E*)-4-methyl-IPP as the main product. The other two C6 products are (*Z*)-4-methyl-IPP and 4-methyl-DMAPP. The C7 compounds are 4,4-dimethyl-IPP and 4,4-dimethyl-DMAPP. Based on the substances found in the headspace of *in vitro* assays using IPP as substrate, and the methyltransferase of *S. monomycini*, it was possible to outline the general mechanism of reaction for this enzyme (Figure 55).

The reaction starts with the addition of a methyl group to carbon 4 of the IPP molecule. The methylation of IPP generates primarily a carbocation intermediate (Figure 54), which can be quenched by different deprotonation paths.
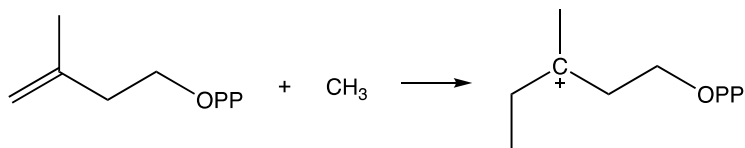
Figure 54. Reaction of IPP methylation, generating a carbocation intermediate.

With this step, a carbocation intermediate is created, which is then quenched by deprotonation. Depending on the source of the proton that stabilizes the carbocation, different conformations of methyl-IPP can be created. With deprotonation on carbon 4, either (*Z*)- or (*E*)-4-methyl-IPP can be formed, depending on which proton is the source. If deprotonation occurs at carbon 2, 4-methyl-DMAPP is formed.

Then, another methylation reaction takes place, done by the same enzyme. It uses one of its methylated products for a second methylation, again on carbon 4. The second methylation reaction gives rise again to a carbocation intermediate, which can be quenched by different deprotonation paths. Depending on the pathway for

stabilization of the reactive carbocation intermediate, different molecules can be formed. With deprotonation on carbon 4, 4,4-dimethyl-IPP is formed. If stabilization of the carbocation intermediate occurs by deprotonation of carbon 2, 4,4-dimethyl-DMAPP is formed.

After 40 h of incubation, three additional peaks were identified while one of the peaks visible at 20 h of incubation disappeared. The disappearance of the peak corresponding to (Z)-4-methyl-isoprenol indicates its consumption as substrate and for a second methylation by the methyltransferase. Since the peak area for this particular compound is very small after 40 h (Table 10), its consumption for the formation of double methylated molecules led to its depletion. The first peak to appear at RT=5.0 minutes is one of the two diastereomers of a methylated IPP, in this case (E)-4-methyl-isoprenol, since (Z)-4-methyl-isoprenol does not appear anymore after a longer incubation time and 4-methyl-prenol follows at 5.4 minutes.
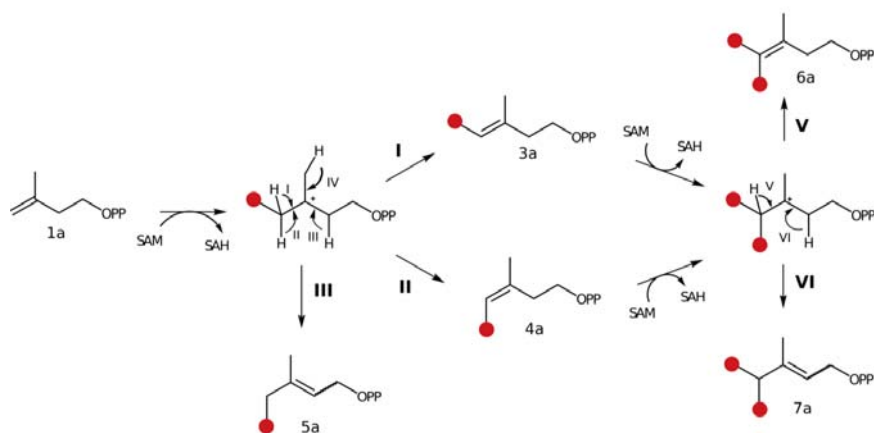
Figure 55. Proposed reaction mechanism for the IPPMT from *S. monomycini*. The proposed reaction pathway of the IPP methyltransferase proceeds via proton abstraction from the initially formed methylated carbocation intermediate, product of IPP (**1a**) methylation, leading to the main product (E)-4-methyl-IPP (**3a**) by deprotonation at carbon atom 4 (deprotonation pathway I). Deprotonation pathway II leads to the formation of (Z)-4-methyl-IPP (**4a**); and deprotonation pathway III generates 4-methyl-DMAPP (**5a**). Subsequent methylation by the methyltransferase forms a double methylated carbocation intermediate, which can be further deprotonated through pathway V forming 4,4-dimethyl-IPP (**6a**); or through pathway

VI forming 4,4-dimethyl-DMAPP (**7a**). Red dots indicate the transferred methyl groups (Drummond et al., 2019).

The methylation of IPP, with the formation of several unconventional methylated IPP and DMAPP derivatives, offers the possibility of including new building blocks in the framework of terpene biosynthesis. The inclusion of new building blocks in terpene biosynthesis enables the formation of a myriad of new compounds, which were inaccessible until now. We together with collaborators furthermore showed the applicability of this technology by expressing the gene in a carotenoid forming system (Drummond et al. 2019), which provided new methylated carotenoids. The applicability in a terpene producing system (*E. coli* overexpressing enzymes from the mevalonate pathway) was also demonstrated, and it will be discussed in the next session.

Another question to be tackled would be the double methylation that was detected in both *in vivo* and *in vitro* experiments. The double methylation of a substrate or use of a methyltransferase's product as substrate for a second methylation has been reported before, for the methyltransferase SgvM from *Streptomyces griseoviridis* (Sommer-kamann et al. 2017). This enzyme belongs to the biosynthetic gene cluster of an antibiotic (viridogrisein), and transfers a methyl group to the compound 4-methyl-2-oxovalerate (alpha-ketoleucine) on position C3, yielding (*R*)-3-methylketoleucine (Figure 56). Afterwards, the latter is transferred to a much larger molecule, yielding the final product viridogrisein. The authors successfully obtained the product by incubating SgvM and its substrate together with TKMAT, ATP, methionine and MTAN, in a system similar to the one used in the present work.
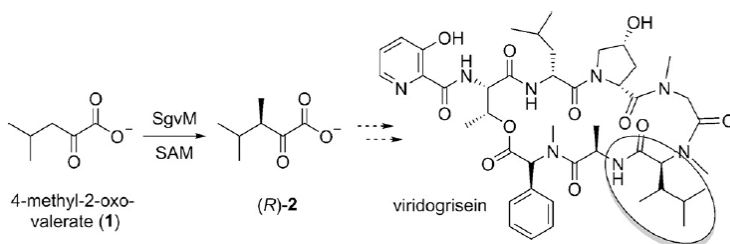
Figure 56. Reaction catalyzed by the methyltransferase from *Streptomyces griseoviridis*. The substrate 4-methyl-2-oxo-valerate is methylated forming (R)-3-

methylketoleucine, which is further incorporated into the antibiotic viridogrisein (adapted from Sommer-Kamann et al., 2017).

Sommer-Kamann and colleagues also performed a series incubations with SgvM, using non-physiological substrates and the protocol used for the *in situ* formation of SAM. Interestingly, when pyruvate was provided as substrate, it is methylated twice by SgvM (Figure 57). The first methylation yields methyl-pyruvate, which is then used as substrate for a second methylation generating di-methyl-pyruvate.
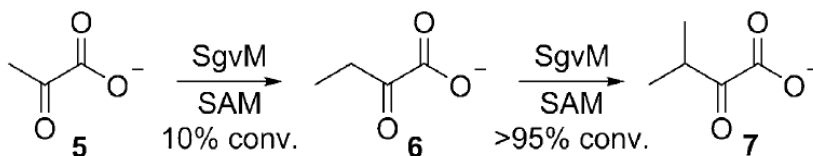
Figure 57. SgvM-catalyzed mono- and dialkylation with pyruvate as substrate (Sommer-kamann et al. 2017).

Longer incubation periods of the IPPMT with IPP and SAM resulted in larger amounts of double methylated compounds. After 40 hours of incubation, peak areas of 4,4-dimethyl-prenol and 4,4-dimethyl-isoprenol increase when compared with 20h of incubation. The isoprenol levels at later stages are undetectable, suggesting that double methylation could be an effect of depletion of the substrate IPP and increase on the concentration of 4-methyl-IPP, which is then accepted by the methyltransferase as substrate for a second methylation. It cannot be discarded however, the possibility that the final product of this methyltransferase reaction is in fact a double methylated C7 compound. To investigate this hypothesis, kinetic studies, both with IPP and with synthetic methyl-IPP, would have to be performed.

The fact that there is still a double bond in end products like 4-methyl-IPP and 4-methyl-DMAPP, allows for the molecules to be subjected to further methylations, and to be further prenylated or elongated, since it does not disturb the inner conditions necessary for a prenyltransferase to react with it. The exact spot where this particular methylation of IPP occurs (the carbon which is methylated) is very fortunate, because

it allows the formation of a chiral center when another prenyl chain is added. 4-methyl-GPP and 4-methyl-FPP, products of the use of 4-methyl-IPP by prenyl transferases (discussed in the next chapter), are both chiral molecules. Interestingly, the use of either (*Z*)- or (*E*)- 4-methyl-IPP as substrate generates specifically (*R*) or (*S*) configurations of the C11 and C16 prenyl diphosphates. This phenomenon was already described using synthetic substrates in assays with prenyl transferases from pig liver (Tanetoshi Koyama et al. 1980; Nagaki et al. 2012). The possibility of enzymatically creating chiral molecules according to the configuration of the substrate offered encompasses an attractive approach for the pharmaceutical industry, which is in constant pursuit of enzymatic technologies for delivering enantiopure products, especially generics (Blaser 2013).

The use of one or more enzymes that can modify prenyl diphosphates increases exponentially the number of structures that can be created. The expression of prenyl diphosphate methyl transferase enzymes together with other terpene-producing systems will probably generate even more unusual terpenes. As proof of concept, one could express two different prenylpyrophosphate methyltransferases in *E. coli* cells overexpressing enzymes from the mevalonate pathway. For example, the expression of IPPMT from *S. monomycini* together with GPPMT in such a system could potentially yield C12 and C13 terpenoids.

The sequence of *S. argenteolus* MT was used as query for a search that yielded the methyltransferase investigated in this study. Regarding its amino acid sequence, the methyltransferase of *S. monomycini* shows remarkable similarity with the methyltransferase from *S. argenteolus* (BAF98640.1), which has recently been attributed the function of IPP methyltransferase (Ozaki *et al.*, 2018). Unfortunately, no structural information is currently available for this protein. This methyltransferase (Lon23) is responsible for the synthesis of methyl-IPP, which is further incorporated at different positions of a dimethyl-GGPP molecule (Hayashi *et al.*, 2007; Ozaki *et al.*, 2018) for the synthesis of longestin. Opposingly to the main product found in this study, the activity of *S. argenteolus* methyltransferase forms strictly (*Z*)-4-methyl-IPP (Figure 58). This methylated IPP is then incorporated into larger molecules by prenyltransferases, a phenomenon also found in the present study and that configures a remarkable feature to be applied for the production of new terpene compounds. All

the enzymes involved in the pathway for longestin formation are encoded in the same operon.
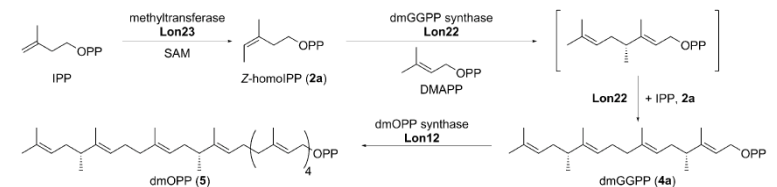
Figure 58. Reaction catalyzed by the methyltransferase from *S. argenteolus* (Lon23). The product (*Z*)-homoIPP is condensed with IPP to form dmGGPP by Lon22. Dimethyloctaprenyl (dmOPP) is an intermediate during the synthesis of longestin (Ozaki *et al.*, 2018).

A comparison between the sequence of the IPPMT from *S. monomycini* and the one from *S. argenteolus*, together with other methyltransferases, was shown in Figure 17 in the first chapter of the present section. Protein sequences with more than 45% identity were aligned and analyzed, especially regarding the amino acid W194 in *S. monomycini* methyltransferase and the equivalent positions in the other protein sequences. Some of the sequences from different species also contain tryptophan in the equivalent position, and at least for *Streptomyces argenteolus* and *Rhodococcus fasciens*, it is known that they code for IPP methyltransferases. For the latter, two different methyltransferases are involved in the synthesis of cytokinins, MT1 and MT2. The substrate IPP is first methylated by MT1, and then the methyl-IPP is methylated a second time by MT2 yielding 4,4-dimethyl-DMAPP, which is then used to form methyl-cytokinin (Figure 59). Interestingly, MT2 also accepts IPP, yet forming 4-methyl-DMAPP, and the product of MT1 reaction was not structurally identified (Radhika et al. 2015). The two methyltransferases involved in the formation of cytokinins in *Rhodococcus fasciens* were proven to be crucial for its virulence (Jameson et al. 2019). In the case of the methyltransferase described on the present study, 4-methyl-DMAPP and 4,4-dimethyl-DMAPP are both formed through activity of a single enzyme.
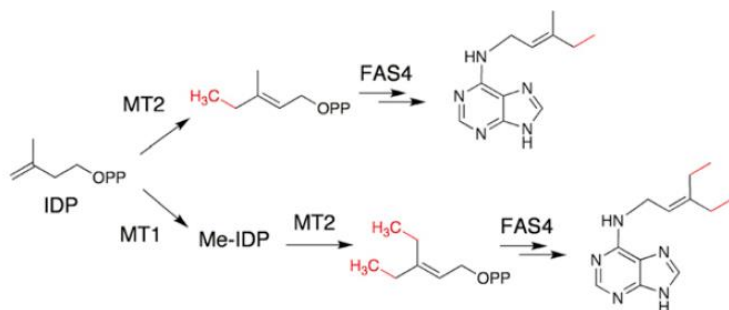
Figure 59. Reactions catalyzed by methyltransferases from *Rhodococcus fasciens*. Isopentenyl diphosphate (IDP) is the substrate for MT1 forming methyl-IDP which is methylated a second time by MT2. IDP is also a substrate for MT2, which converts it to (Z)-4-methyl-isopentenyl diphosphate. The products are then used by the enzyme FAS4 which forms methyl-cytokinins (modified from Radhika *et al.*, 2015).

The recently proven existence of other methyltransferases which also have prenyldiphosphates as substrate shows that this is a field yet to be explored. Most of these methyltransferases come from microorganisms and their biochemical properties and pathways involved in their production are poorly studied, therefore most probably there are many other enzymes with similar activities still to be discovered. Also, the fact that they are produced in very reduced amounts by bacteria, makes it even more difficult to detect these volatile compounds. However, the approach combining data mining, gene selection, plasmid design and heterologous expression in a model organism like *E. coli* or yeast, turns possible the study of these compounds and enzymes.

Supplementary to the pathways described for microorganisms, insects exhibit a very peculiar pathway for the formation of terpenoid compounds. In the case of lepidopterans, the mevalonate pathway can use as starting molecule propionyl-CoA together with acetyl-CoA, which leads to the synthesis of homo-prenylpyrophosphate intermediates homo-IPP and homo-DMAPP and ultimately the formation of juvenile hormones (Figure 60). This pathway was recently engineered in *E. coli* for the production of novel C16 terpenes (Eiben et al. 2019), showing the plasticity of terpene

biosynthesis and exemplifying how genome mining and strain engineering can lead to *in vivo* production of novel compounds.
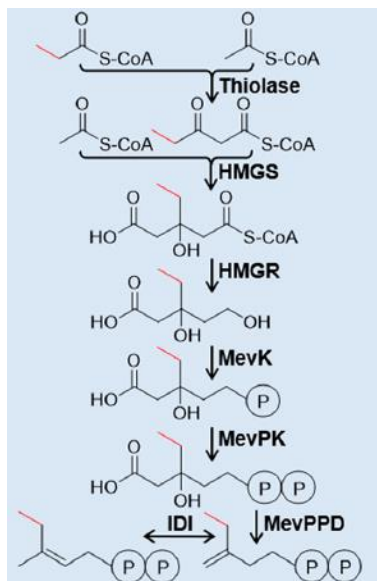
Figure 60. Mevalonate pathway from Lepidoptera, which starts with the condensation of propionyl-CoA and acetyl-CoA, leading to the formation of homo-IPP and homo-DMAPP (adapted from Eiben *et al.*, 2019).

It is possible to assume that there are many other methyltransferases to be discovered, perhaps it is even a common function among actinobacteria. An aspect still uncovered is what the functions of this type of methylated compounds could have for these organisms and within their micro ecosystem. Further studies considering the activation of those genes *in vivo* under different physiological conditions of the cells could be an approach for tackling the question about function. Possibly, there might be a correlation between the activation of methyltransferase genes and non-canonical terpene production with responses to the environment such as nutritional stress or population detection. A study on bacterial communication through terpenes (Schmidt et al. 2017) shines light over this issue, pointing to a direction of study that was not sufficiently explored: how do microorganisms communicate, which compounds are involved in their communication, what triggers the production of signaling molecules,

how is the information received and processed. In this publication, it is shown how the contact with volatiles produced by fungi elicits a very specific response from the bacteria *Serratia plymuthica*, with increase of motility and the production of a non-canonical C16 terpene. The production of this unusual terpene, called sodorifen, involves an FPP methyltransferase, which methylates an cyclizes the C15 prenyl diphosphate FPP originating a C16 intermediate (Figure 61) that is further processed by a terpene synthase (Von Reuss et al. 2018). This FPP methyltransferase constitutes one more example of prenyl diphosphate modification, and the creation of a new module for terpene biosynthesis.
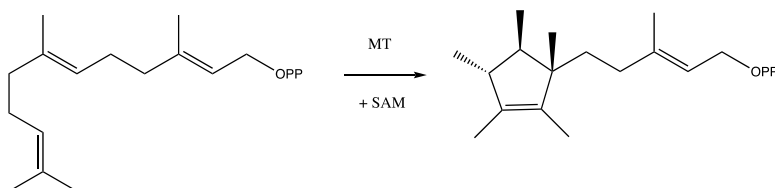
Figure 61. Reaction catalyzed by the FPP methyltransferase from *Serratia plymuthica* (modified from Von Reuss *et al.*, 2018).

Although the proven existence of FPP and GPP methyltransferases is a valuable knowledge for the increase in the number of possible mono- and sesquiterpene structures, having a methyltransferase that methylates directly the universal precursor of terpenes, IPP, allows changes in virtually all terpene structures, from hemiterpenes to terpene polymers with unending chains. The acceptance of methyl-IPP by prenyltransferases as building block for the construction of larger terpene precursors, shown in this work, includes another level of flexibility for the application of the IPPMT.

With further information about the IPPMT's structure and function, protein engineering aiming at modifications in the protein structure could change various features, such as the point of insertion of methyl group in the molecule, substrate promiscuity or specificity, production rate, protein expression levels, and product specificity or diversity, depending on the purpose.

The applications of the technologies here described are very diverse. On the field of pharmaceutical research, the methylation of prenyldiphosphates, which can be incorporated to other molecules, (methyl-prenylation) could increase the efficacy of certain compounds, due to the consequential increase in membrane permeability. Diverse secondary metabolites with added prenyl chains like prenylated flavonoids, coumarins, chalcones and cinnamic acids; are known for a wide variety of biological activities, from cytotoxicity in tumor cells to treatment of climacteric disorders (Chen et al. 2014; Alhassan et al. 2014). Besides the activity of the secondary metabolite itself, the addition of a prenyl chain often leads to an improvement regarding the compound's pharmacological activity. The prenyl side chain is lipophilic, which in turn affects the access of the compound through the lipophilic membrane of the cells. Prenylation is an interesting approach for the development of new drugs, the best example being the compound farnesyl thiosalicylic acid, which is used for treatment of cancers related to Ras proteins mechanism (Alhassan et al. 2014). New technologies are under development in this direction, including engineering prenyltransferases for differential prenylation of target molecules and generation of unnatural novel compounds (Mori et al. 2016). The use of methylated prenyl chains for drug development could be an interesting application for modulation of the activities derived from prenylation, using the technology described in this work. In a similar fashion, prenylpyrophosphate methyltransferases could be used to modify prenyl side chains of flavonoids, conferring new biological, flavor or nutritious activities to these class of high-value compounds (Xiaoman Yang et al. 2015; Mukai 2018).

On the field of flavor and fragrance, new aroma compounds can be formed by addition of methyl groups to precursors of terpenes. Due to the volatile nature and pleasant impressions of mono-, sesqui- and diterpenes, they are quite commonly employed in the composition of fragrances and flavoring agents. The modification of terpene precursors by addition of one or more methyl groups, and then application in preexistent platforms and biosynthetical routes leading to the formation of aroma compounds, offers a very promising approach to diversify the structures of single compounds and general impressions of fragrance blends. Considering the growing markets for terpene oils, aromatherapy, cosmetics, flavors and fragrances in general, the exploration of a new framework based on the modification of basic building blocks

and therefore the whole chain of production of terpenes is a quite attractive gateway for the development of new products.

The identification of compounds was obtained majorly through comparison with custom-synthesized reference compounds, a method that despite being laborious and time-consuming, allowed us to build a library of methylated terpenes which is in constant upgrading and will help future studies in this field. Since it is still not possible to predict fragmentation patterns of candidate compounds, the methodology above-mentioned has to be applied for correct identification of the novel terpenes. Another possibility would be collection, extraction, isolation and characterization of fractions via NMR. Due to the fact that the compounds analyzed in this study are highly volatile and produced in small quantities, volatile capture and extraction techniques, as well as improved production systems are still to be developed with this goal.

4.4   *In vitro* incubations with methyltransferase mutants

The experiments using mutants with exchanged amino acids (GPPMT G202W and IPPMT W194G bearing glycine instead of tryptophan), which aimed to reproduce an exchange of amino acids in equivalent positions from both enzymes, revealed to be unsuccessful. That can indicate an inconsistence of the theoretical model, which arises from the fact that little is known about the *S. monomycini*'s methyltransferase. The reasons for the observed differences in function between the two enzymes in question apparently go beyond one divergent amino acid. For further experiments to be made based on engineering of the IPPMT, more information would have to be made available about this protein's structure. The elucidation of IPPMT crystal structure could be a starting point for many other projects involving the better understanding of this enzyme's function. Having previous work (Köksal et al. 2012; Ariyawutthiphan et al. 2012) on the crystal structure of GPPMT from S. *coelicolor* as starting point will greatly help on the structure elucidation of the IPPMT from *S. monomycini*.

4.5 *In vivo* experiments

4.5.1 Heterologous expression of operon from *S. monomycini*

The expression of IPPMT together with enzymes from the mevalonate pathway in *E. coli* production strains led to the formation of methylated non-canonical terpenes. Besides the C6 and C7 isoprenoids which were also detected in *in vitro* assays, a new C6 compound, 5-methyl-isoprenol, was also identified. This compound is most probably one more product from methylation of IPP on C4, which is formed by deprotonation pathway IV (Figure 55). The fact that this product was only detected during *in vivo* expression can be due to a higher concentration of products overall, when this method was employed.

Even though the function of the methyltransferase from *S. monomycini* was characterized, the final product of the proposed operon could not be identified. The fact that the genes are organized in an operon suggests conjunct regulation and activation, but the expression of the operon in *E. coli* could not reveal its final product. Possible reasons for this result could be insufficient or failed expression of the prenyltransferase and terpene synthase genes from the operon. It is also possible that additional factors from outside of the operon are necessary before the TS reaction, for example if the methyl-pyrophosphates are transferred to another molecule whose production depends of the activation of genes outside of the operon.

4.5.2 Isotope labeling of methylated compounds and identification of long-chain methyl-terpenes

Non-radioactive isotope labeling is very good method for identifying methylated compounds in a mixture, especially through GC-MS. It has been used extensively not only in the identification of terpene compounds, but also to clarify mechanisms of reactions (Dickschat, 2017). The technique allowed us to proceed with the first identifications of double methylated compounds, with the detection of mass spectra with a shift of +2.

The application of the isotope labeling technique together with SBSE extraction of compounds present in the liquid media allowed the recognition of methylated terpenes with longer carbon chains, which were not detected before. More than one factor contributed to the lack of detection of longer compounds in previous experiments. Firstly, the longer compounds are more soluble in water than the smaller molecules that are highly volatile. Therefore, it is easier to extract longer chain terpenes with the twisters which are placed in the medium. Secondly, the use of isotopically labeled methionine allowed us to focus on methylated compounds which could have been overseen before, when there was no clear indication of which compounds among the rich volatile blend could be indeed methylated.

With the validation of the exact peaks containing methyl groups added by the methyltransferase, through comparison of mass spectra of labeling and no labeling treatments for *in vivo* experiments, it was possible to detect long chain terpenoids with not only one, but also with two methyl groups added. The use of this labeling technique led to the identification of three novel terpenes, two with 11 carbon atoms and one with 16 carbon atoms in their structures. The C11 compounds 4-methyl-geraniol (Figure 27) and 8-methyl-geraniol (Figure 28); together with the C16 compound 4-methyl-farnesol (Figure 29) had their structures elucidated. Also, according to the structures formed, it was possible to hypothesize the substrates used by prenyltransferases to form these compounds, or more precisely, which of the methylated prenyl diphosphates were involved in the formation of their structures. Since *in vitro* assays demonstrated that IPP is the only accepted substrate for the methyltransferase, one can safely assume that the methyl groups come from condensed methylated precursors and not from the methylation of GPP or FPP. The condensation of methylated and non-methylated precursors had one pathway possible according to the structures found, which are depicted in Figure 62.
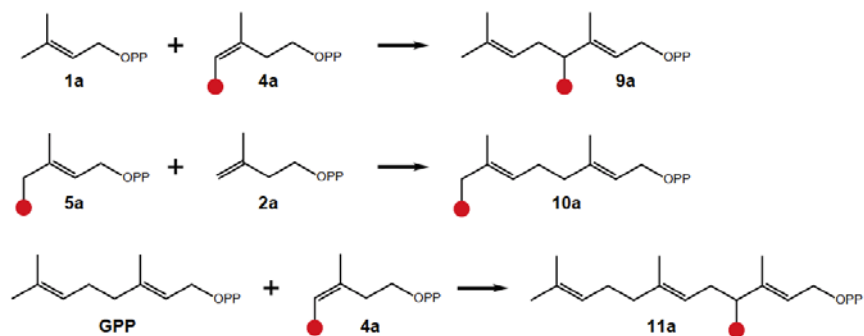
Figure 62. Schematic presentation of proposed biosynthetic pathways towards longer unconventional prenyl diphosphates. 4-methyl-GPP (**10a**), 8-methyl-GPP (**12a**) and 4-methyl-FPP (**11a**) are formed by incorporation of 4-methyl-IPP or 4-methyl-DMAPP during prenyl transferase-catalyzed elongation of the prenyl chain. Red dots indicate the transferred methyl groups (Drummond et al. 2019).

The analysis of products of *in vivo* expression with isotope labeling showed compounds with added methyl groups that were not structurally identified yet. Even though, through the identification of 4-methyl-geraniol, 4-methyl-farnesol and 8-methyl-geraniol, the incorporation of 4-methyl-IPP and 4-methyl-DMAPP into larger terpenes was proven, it is still unknown how the double-methylated C12 and C17 compounds are structurally organized. It was not possible to affirm whether the compounds which contain two methyl groups are formed through incorporation of a double methylated precursor, or if two single methylated precursor building blocks are incorporated in the same molecule.

There is a high probability that enzymes from *E. coli* are taking part on the formation of longer prenyl chains and that in this process they also accept methylated substrates. There are three prenyl transferases in *E. coli*: FPP synthase, octaprenyl synthase and undecaprenyl synthase. The most logical candidate for the catalysis of such reaction is the FPP synthase (IspA) due to the length of its final product (C15), since the other two prenyl transferases abovementioned a responsible for the formation of much longer products, containing in their structures 40 carbon atoms (octaprenyl pyrophosphate) and 55 carbon atoms (undecaprenyl pyrophosphate).

Even though there is more than one possible configuration for 8-methyl-geraniol, in all the *in vivo* experiments where this compound was detected there was only one peak with this specific mass spectrum. This is probably due to the fact that its methylated precursor, 4-methyl-DMAPP, is present only in the E form, and that the FPP synthase from *E. coli* is i a trans-FPP synthase. For the other C11 compound formed 4-methyl-geraniol, in some cases it was possible to observe more than one peak with mass spectra corresponding to it. By analyzing the chromatograms only, it is unclear whether only one configuration of the compound was produced, or only one configuration could be detected, since in the vicinities of the point in retention time where these C11 compounds usually appear, there was always signal noise from other compounds that eluted together from the column. One example is the peak for indole, which elutes almost at the same time as 4-methyl-geraniol and has a much larger peak area, which ended up masking the peak of interest in some of the experiments. However, when analyzing the biosynthesis of these longer unconventional compounds, it is possible to discuss their possibilities for configuration. The best candidate for catalyst of the formation of methyl-GPP and methyl-FPP out of methylated C6 precursors is the *E. coli* FPP synthase (IspA). Since the reaction catalyzed by this FPP synthase is stereo-specific, yielding always *E,E*-FPP, it is logical to assume that the synthesis of methyl-GPP out of methyl-IPP and DMAPP by the same enzyme would also yield *E*-4-methyl-GPP. The stereochemistry of the final molecule in this case would be determined by whether the methyl-IPP molecule was *E* or *Z*. The substrate *E*-4-methyl-IPP would yield *S*-4-methyl-GPP, and *Z*-4-methyl-IPP would yield *R*-4-methyl-GPP, as demonstrated in previous studies with FPP synthase and synthetic substrates (Tanetoshi Koyama et al. 1980; Nagaki et al. 2012). Considering this, the specific stereochemistry of 4-methyl-geraniol formed *in vivo* actually depends on the configuration of precursor available for its synthesis. Since the *E* form of 4-methyl-IPP is the most abundant one, as the main product of IPP methylation by the IPPMT, it is logical to assume that (*E*,*S*)-4-methyl-GPP is the most prominent isoform of 4-methyl-GPP produced.

In the case of C12 and C17 compounds where two additional methyl groups are present, it is not clear whether the methyl groups come from two different single methylated precursors (like C6 prenyl diphosphates 4-methyl-IPP and 4-methyl-

DMAPP), or from one single precursor bearing two methyl groups (like C7 prenyl diphosphates 4,4-dimethyl-IPP or 4,4-dimethyl-DMAPP). Similarities with mass spectra of compounds with known structure can give a hint for the origin of the two methyl groups present in the C12 and C17 molecules, but it is very difficult to assume any probable position for their insertion. Reliable structure identification would only be possible with clear similarity to mass spectra and retention index from reference compounds, or with compound separation and NMR analysis.

4.5.3 Investigation of *in vivo* production of terpenes by wild strain *Streptomyces monomycini*

Actinomycetes, also called "nature's chemists" are substantially different from commonly used model organisms like *Escherichia coli* and *Saccharomyces cerevisiae*, because they are able to produce a much wider diversity of complex chemical compounds. Bacteria from the genus *Streptomyces*, belonging to the phylum of Gram-positive Actinobacteria, are a class of soil bacteria that grow forming branching filaments and spore chains. Genomes from *Streptomyces* were shown to have 20 to 30 clusters for the production of secondary metabolites (Kuzuyama 2017), and *Streptomyces* species are well-known for their intricate secondary metabolism (van Wezel and McDowall 2011), their key role in agricultural and forest systems, and also for the synthesis of a myriad of antibiotics like streptomycin, erythromycin and chloramphenicol (de Lima Procópio et al. 2012).

In the context of this thesis, a specific operon from *Streptomyces monomycini* was studied by heterologous expression and in vitro assays. Cultivation of the strain complemented the study providing information about the compounds produced naturally by the species. However, there was no clear indication that the compounds detected in the headspace of *S. monomycini* cultures could be derived from the expression of enzymes encoded on the operon of our choice. Many factors might influence this finding. The product of the operon might be non-volatile, belonging to a larger molecule which would not be detectable by the means that we applied (HS-SPME). It is known that terpenes with longer chains are less volatile, and on the heterologous expression experiments it was possible to detect compounds containing 15 carbons or more when SBSE adsorbing devices were used in the medium instead

of SPME fibers on the headspace. For that reason, LC-MS analysis of the culture supernatant was performed and indicated the presence of C15 and C16 compounds. The C16 compounds in the culture supernatant could be products of the IPPMT activity and could even be the final product from the expression of the selected operon. However, similar compounds were detected *in vivo* when only the methyltransferase was heterologously expressed, indicating they are not the final product of the operon, but only products of methyltransferase expression *in vivo*. The final product of the chosen operon remains unknown.

Another possible reason for the absence, on the headspace and liquid phase of *S. monomycini* cultures, of compounds which would be candidates for products of the operon activity can be that the genes in focus are only activated by some kind of stress. It is known for *Streptomyces* species that some genes, especially those ones associated with secondary metabolism, have their expression associated with stress factors like high temperature, nutrient deficiency or the presence of harmful compounds (Salah-Bey, Blanc, and Thompson 1995; Kieser et al. 2000). The production of specialized compounds like terpenes and other signaling molecules is reduced and usually specific to certain conditions, for the exact reason of them functioning as messages to communicate with other organisms on the environment with same-species individuals or even other species (Jones et al. 2019; 2017; Jones and Elliot 2017).

Methylation of natural products linked to a physiological response to stress is a known mechanism described for plants. In response to pathogen infection, plants can produce methylated flavonoids that function as chemical defense against them (C. J. Liu et al. 2006). So far, no studies were conducted on the function of methylated terpenes in bacteria. To my knowledge, no general ecological or signaling function has been attributed to methylated terpene compounds, nor to 2-methylisoborneol or 2-methylenebornane. A possible link between production of methylated compounds and bacterial physiological response to stress is still a hypothesis to be evaluated.

It is important to remember that *Streptomyces* are the most economically important genus of Actinomycetes. They are producers of one of the most famous antibiotics to treat tuberculosis: streptomycin (*S. griseus*), and another 50 antibiotics which account

for more than half of the world's antibiotics. Other examples of antibiotics produce by Streptomyces species are erythromycin (*S. erythraeus*), chloramphenicol (*S. venezuelae*), chlortetracycline (*S. aureofaciens*) and oxytetracycline (*S. rimosis*). Streptomyces have large genomes for bacteria. The sequencing of *S. coelicolor*'s genome revealed astonishing 8,667,507 base pairs and predicted 7,825 genes, which is twice as much genes found in normal free-living bacteria. Due to such a large genome, containing a myriad of transcription factors controlling the expression of genes, they dispose of a diversity of genomic factors that translate into a rich metabolism, source of products which account for most of the discoveries regarding antibiotics and bioactive compounds like antitumorals and immunosuppressors (Manteca and Yagüe 2019). Using a huge set of molecular switches, these soil organisms sense opportunities and threats from the environment (in the soil around them) and react accordingly. Studying Streptomyces is becoming more and more necessary, especially with the increase of reported resistance of microbes to existing antibiotics, which brings the urgent necessity for new antibiotics, a type of compound for which Streptomyces are known to produce. Growing these cryptic creatures in the laboratory to better understand their biology and biochemistry is an important but very difficult approach which is why it is rarely seen in current publications, even though examples do exist (Chater 2016). By isolating cells from the environment and screening for activity, many different species and new compounds can be discovered (Y. Takahashi and Omura 2003), however heterologous expression is still a more efficient way of producing specific compounds in laboratory conditions. In this sense, a metagenomic analysis of environmental samples could provide valuable information on interesting biosynthetical pathways. Methyltransferase conserved motifs or consensus sequences could be used to create signatures to identify and characterize genes encoding new prenylpyrophosphate methyltransferases.

Taking this biosynthetical pathway as an example, also from a Streptomyces species, it is possible to extrapolate that the production of other complex molecules from unknown metabolic pathways might be taking place in *Streptomyces monomycini* as well, and that maybe the discovered IPPMT is an enzyme whose reaction participates on the formation of a much larger molecule, that was not detected on the scope of this study. The methylated products of this reaction, C6 or C7 prenyldiphosphates, may be the substrates for the prenylation of a still unknown complex molecule, maybe even

with interesting biological activities. Three decades ago, the incorporation of a prenyl chain (C10) into a larger molecule was already described for Streptomyces (Shin-ya et al. 1990). It should be considered the possibility that the reactions described in this work for IPPMT may be involved, together with other enzymes not enclosed within the operon, in pathways still to be uncovered.

In the genome of *S. monomycini*, the presence of other terpene synthases was investigated. Since the terpene synthase in the operon selected by us for this study did not produce any detectable products in the different experimental setups used, another search was made in the rest of the genome for suitable terpene synthase sequences, using the hypothetical protein TS from the operon as query (WP_030019073.1). It might be possible that the sequence coding for the enzyme responsible for catalyzing the final reaction, which would generate the product related with the presence of the IPP methyltransferase, is located in another region of the genome, and not organized as an operon. The genome of *Streptomyces monomycini* contains only two other terpene synthase sequences, which have unknown functions. One is the protein with reference sequence WP_078624463.1 is annotated as germacrendiol/geosmin synthase. The other sequence is annotated simply as cyclase (WP_030017636.1). Whether this cyclase accepts non-canonical precursors is yet to be investigated. It is difficult to identify a terpene synthase specific for the formation of methylated terpenes out of methylated prenyl diphosphates, only by its sequence. Even the 2-MIB synthase, responsible for the conversion of 2-methyl-GPP into 2-MIB, also accepts GPP as substrate (C.-M. Wang and Cane 2010). As it was shown in this study, this is also the case for some plant terpene synthases, to accept non-conventional prenyl diphosphates with an uneven number of carbon atoms as substrate. The analysis of reactions using plant terpene synthases to convert GPP and 2-methyl-GPP into terpenes also revealed that for one case, the reaction yielded a larger diversity and number of terpenes when the methylated precursor was provided. Taking these facts into account, further investigation on the function of the terpenes synthase contained in the analyzed operon, as well as the other cyclase found in the genome of *S. monomycini*, should be considered.

4.6 *In situ* cofactor formation

The use of TKMAT, the methionine-adenosyl transferase from Thermococcus kodakarensis, allowed the formation of SAM and SAM analogs *in situ*. When applied to *in vitro* reactions with IPPMT, the formed cofactors were able to transfer the donating group to the final terpene product, except in the case of ethionine. 4-methyl-isoprenol was formed with transfer of the methyl group of SAM formed *in situ*, as well as ^{13}C-methyl-isoprenol as a result of ^{13}C-methyl group donation from ^{13}C-SAM. The formation of compounds containing ^{13}C was proven for reactions with the IPPMT via GC-MS analysis of the volatile products during *in vitro* assays, as well as with assays using GPPMT from *S. coelicolor* (data not shown). Interestingly, no double-methylated compounds were detected for assays with *in situ* SAM formation. A possible explanation could be the slow rates of SAM formation, which do not enable prenyldiphosphate depletion as substrate and therefore, the build-up of methylated product. With still enough substrate to be transformed, it is possible that the methyltransferase prefers IPP than the methylated version.

The LC-MS analysis of *in vitro* assays using the enzyme TKMAT and substrates different than methionine, confirmed the formation of SAM, ^{13}C-SAM and SAE, thus proving that SAM and SAM analogs can be formed by this method. However, in the case of SAE, the ethyl group was not transferred to IPP, and no ethylated compounds were formed. The inefficiency of this particular reaction cannot be explained, but the results give some insight on why the formation of ethylated isoprenoids was not detected by GC-MS. Low amounts of SAE in the reaction vial could have suppressed the activity of the methyltransferase for not reaching a minimum concentration that would initiate the necessary equilibrium for the product formation to occur. Another possibility is that the amounts of ethylated compound were so infinitesimal low that its detection through the used methods was impossible.

TKMAT (MAT from *Thermococcus kodakarensis*) bears one interesting advantage when compared with MAT from *E. coli;* its affinity for ethionine is much higher than the *E. coli* homologue (Schlesier et al. 2013). Even though the detection of ethylated volatile compounds was not possible with GC-MS analysis, I detected the formation of

SAE through LC-MS analysis, which showed successful incorporation of ethionine catalyzed by TKMAT. A probable limiting factor for the reaction is ATP, since it is a molecule which gets degraded very easily in the presence of water, with hydrolysis of the phosphate moieties. Another evidence was the fact that the peaks corresponding to methionine have higher intensity than the ones corresponding to SAM, showing an incomplete conversion by the enzyme, which could have been limited by ATP depletion. ATP is a highly unstable molecule that can be easily degraded by hydrolysis in aqueous media. Analysis by LC-MS of the molecular ion and species associated with ATP showed no signal, indicating no ATP present in the analyzed samples, thus pointing to degradation of the compound. On the other hand, the absence of ethylated IPP could also mean that the methyltransferase is selective in terms of its cofactor. The successful ethylation of substrates using MAT was reported (Mordhorst et al. 2017; Sommer-kamann et al. 2017). However, lower levels of ethylation, when compared to methylation, were described as well (Sadler et al. 2017).

The use of methyltransferases together with *in situ* SAM generation offers an attractive approach for *in vitro* production of regioselective methylated compounds. Additionally, the methylation of compounds with ^{13}C-SAM is a very useful technique because it makes it easier to clarify compound structure through NMR. There is also the possibility of using this system for producing methylated compounds *in vivo*, since the offer of substrate and SAM can be continuous until the carbon source in the medium is depleted. Within living organisms, there is an inherent mechanism for SAM regeneration, which sometimes involves the participation of vitamin B12. However, for *in vivo* set-ups, the production and use of SAE as cofactor is not likely, since the ethylation in living systems was proven lethal. For *in vitro* systems, efforts have been made to create a cycle of SAM complete regeneration (Mordhorst et al. 2017). Even though the complete regeneration of SAM yielded only a few cycles (maximum of 10), the regeneration of ATP and its use for SAM formation, together with the cascade of methylation of the substrate and SAH degradation yielded more than 1000 cycles. The need for compounds in stoichiometric amounts in this case would be restricted to adenosine and methionine, which can be both acquired without great expenses. Within the context of industrial enzymatic production of methylated compounds, the development of SAM *in situ* generation and/or regeneration is an important question to be tackled.

The joint use of methyltransferases and SAM analogs can be applied to differentially alkylate substrates. Alkylation of natural products is a prominent biosynthetic approach to modify naturally produced compounds of economic importance, generating novel products through a process called alkyl randomization or differential alkylation (Huber et al. 2016). Technologies that allow the synthesis and use of SAM analogs have been applied to the production of novel small molecules, besides polyketides, non-ribosomal peptides and coumarins, among others (reviewed by Huber *et al.*, 2016). The use of methyltransferases in these alkylation processes offers access to carbon-carbon bond formation which are sometimes synthetically difficult. In this context, SAM analogs work as powerful chemical biology tools on the expansion of possible structures to be formed through enzymatic synthesis, as well as enablers of new chemical structures to be discovered. The use of the IPP methyltransferase together with SAM synthetase can allow the alkyl randomization of a myriad of molecules, offering new building blocks for the modular terpene biosynthetical pathway and exponentially increasing the number of compounds that can be synthesized.

4.7   Production of non-canonical terpenes using plant terpene synthases

The promiscuous activity of plant terpene synthases was demonstrated in a closed *in vitro* system, with the provision of synthetic 2-methyl-GPP and purified enzyme. For geraniol and limonene synthase the expected products were detected (2-methyl-geraniol and 2-methyl-limonene), but for pinene and carene synthase the expected products were either not produced or not detected. The incubations with limonene synthase yielded unexpected results, since the incubation with its natural substrate GPP resulted in geraniol instead of limonene as product. The reason for this result may lay in an incomplete folding of the protein due to conditions distant from the enzyme's optimum. Geraniol is a product of simple cleavage of the phosphate group from GPP, with no further modification of the molecule. The incorrect folding of the protein could have led to the loss of function in some parts of the active site, leaving out only the phosphate cleaving region. One of the C11 products of the enzyme carene synthase was the known bacterial product 2-methylenebornane, a compound that was never reported as a plant product. The fact that the enzyme carene synthase converts

2-methyl-GPP into 2-methylenebornane through cyclization probably using a 2-methyl-terpinyl cation in a similar fashion as a C11 terpene synthase, is very interesting and the exact mechanism should be investigated in future work. Furthermore, a variety of C11 compounds of unknown structures were also produced. The formation of carbocation intermediates can partially explain the variety of products formed. Because product formation is associated with generating and quenching a carbocation, product ratios and qualities are pH-dependent.

Interestingly, the number of peaks which are synthesized by the enzyme increase when 2-methyl-GPP is used instead of GPP. This could indicate a higher possibility of side product formation when a non-natural substrate is used, or even a higher structural diversity of 2-methyl-GPP processing, when compared to its non-methylated counterpart. An additional methyl group allows a greater diversity of products, especially when subsequent modifications are considered. Further processing of the molecule through e. g., cyclization, hydride shift or hydroxylation, offer even more structural possibilities when an additional methyl group is present. The product spectra will vary even more when changes in pH are made.

A differential terpene production depending on the buffer's pH agrees with the terpene synthase mechanism. The enzymes binding site has its properties changed when the ion concentration in solution varies. When more protons are present, the interaction with its substrate changes, resulting in a different blend of side products. The mechanism of the reaction underlying this variety of non-canonical terpenes which were produced by different plant enzymes when 2-methyl-GPP was used as substrate resides in the mechanism of reaction that is typical for the terpene synthases. The cleavage of pyrophosphates usually leads to the formation of a carbocation intermediate, leading to different possibilities of quenching the carbocation, causing the production of a variety of different compounds from a few intermediates.

The use of plant terpene synthases is an interesting approach to the formation of complex non-canonical terpene structures, with the differential modification of methylated prenyldiphosphates into cyclic methylated terpenes. Recent work from Ignea and collaborators showed how C11 compounds can be formed through the expression of terpene synthases in yeast strains overproducing 2-methyl-GPP (Ignea

149

et al. 2018). The work very elegantly demonstrated the possibility of creating production strains of unusual terpenes, and how to overcome inner regulations from the cell to enable higher titers of target compounds. The expression of different C11 terpene synthases in *E. coli* cells overexpressing the mevalonate pathway and GPPMT was shown to generate many new C11 terpene structures (Kschowak et al. 2018). In this particular work are displayed new compounds which were not described before, including the *in vivo* production of a C16 compound derived from the condensation, by prenyltransferases, of 2-methyl-GPP with IPP.

New properties can be accessed with the addition of a methyl group and posterior transformation (e.g. cyclization) by terpene synthases, and the engineering of terpene synthases could fine tune product formation. For example, new chiral centers can be added, aromatic (smell) or physicochemical properties can be changed or modulated, medicinal properties can be fine-tuned, and the antibiotic properties of these molecules can be accessed and tested. It is an efficient method for creating new C11 terpenes. As discussed above, cellular synthesis of non-canonical terpenes is also possible, enabling the production of high-value designed compounds out of simple sugars, using a single host and yielding pure substances or complex mixtures through protein engineering.

Having several side products for one reaction with one substrate can be advantageous when a mixture of compounds is desired, for example in the fragrance industry. There are cases in which a blend of aromas is responsible for the fragrant impression, which is the effect generated by natural extracts. The extracts from plants, for example, contain a myriad of different components, from whose proportion a complex and pleasant smell emerges. This effect could also be achieved by metabolically engineering organisms to contain enzymes with promiscuous product formation, which could generate a specific blend to be extracted.

If specificity is required, for example in the case of single compound production for flavor and fragrance formulation, or the case of pharmaceutical compounds, the problem of several side products could be circumvented by engineering these proteins aiming at a narrower product spectra or even higher reaction speed rates. The engineering of terpene synthases was already shown to have an effect on product

spectra for C11 terpenes, even with the formation of unknown, possibly new to nature compounds, or at least so far undiscovered (Kschowak et al. 2018).

5 Conclusions

The finding of a methyltransferase that modifies IPP, the universal precursor of all terpenes, was enabled by a database search for homologues to the GPP methyltransferase from *S. coelicolor*. The enzyme from *S. monomycini* was expressed heterologously in *E. coli*, and using its purified fraction on *in vitro* experiments, it was possible to characterize it regarding its enzymatic activity, with the formation of C6 and C7 compounds. A new prenyl diphosphate methyltransferase was discovered, that uses IPP as substrate forming a diversity of methylated precursors for terpene biosynthesis. These new precursors can be used as building blocks for the synthesis of unconventional terpenoid compounds. Whole cell expression together with enzymes from the mevalonate pathway allowed the identification of longer terpenes which contained methylated IPP or DMAPP in their structures, forming C11, C12, C16 and C17 compounds. Using a SAM synthetase on *in vitro* assays, together with the methyltransferase, the formation of methylated and ^{13}C-labeled compounds was achieved. Analysis through LC-MS revealed that the SAM synthetase could produce SAM analogs including S-adenosyl-ethionine, which can be used in the future to ethylate IPP with the achievement of higher rates by optimizing the methodology. Moreover, *in vitro* assays using a non-canonical C11 prenyldiphosphate as substrate for terpene synthases from plants resulted in the formation of a variety of C11 terpenes. It was demonstrated that non-canonical prenyl diphosphates can be used by different terpene synthases to form terpene compounds with uneven number of carbons.

Considering the overall production of methylated compounds, even though quantification of products was not the topic of this work, a follow-up study would most definitely include different extraction methods for quantification of the different compounds produced, especially regarding *in vivo* heterologous expression. For that matter it should be noted that it is possible that the production *in vivo* was not optimal because of lack of phosphatases in the intracellular environment of *E. coli*. The co-expression of different phosphatases increased isoprenol and prenol synthesis in *S. cerevisiae* and in *E. coli* for the production of biofuels (Zheng et al. 2013). A similar

strategy should be taken into consideration for future work involving in vivo production of methylated terpene alcohols.

Particularly for the experiments involving plant terpene synthases, many chemical structures could not be identified and remain as unknown compounds. The lack of specific reference compounds for some of expected products hindered a clear conclusion on whether they were produced or not. For those compounds to be identified in the mixture, reference compounds would have to be synthesized and have their chromatograms as mass spectra compared to the ones obtained experimentally. Alternatively, the compounds could be extracted, isolated and structurally identified via NMR spectroscopy.

The results show a variety of techniques by which the canonical framework of terpene formation can be broadened. With the use of methylated versions of the universal precursor IPP, many other non-canonical terpene structures are possible. This new concept of methylation of terpene precursors greatly expands the isoprene rule which was postulated decades ago but is still used to describe terpene biosynthesis as a sequential addition of 5-carbon-atom units. With the results described in this thesis, and also the ones described in recent work from other research groups, an expansion of this concept is in progress.

The technologies developed offer a myriad of ways by which the chemical structure of terpenes can be significantly diversified. The applications involve all fields in which terpenes are prominent. These techniques also prove the applicability of methyltransferases in different systems, for the production of unconventional terpenes using terpene synthases. Methyl transferases were used here in the context of microbial cell factories and *in vitro* enzyme reactions, and the applications could be expanded to other areas which involve pathways that have prenyl diphosphates as metabolites. The combined results open a precedent for designer molecules to be formed in this way, and a broad spectrum of new terpenes that awaits to be discovered, tested and used in different areas. The methylation of terpene precursors opens a path for the discovery of new molecules, with the incorporation of an additional piece on the modular terpene biosynthesis. The search for new molecules with pharmacological effects is of utmost importance in a world where resistance of

pathogens to antibiotics is on the rise. The flavor and fragrance industry could also benefit from new aroma molecules to be included in portfolios, with new yet undiscovered properties.

6 References

Adams, R P. 2007. *Identification of Essential Oil Components by Gas Chromatography/Mass Spectrometry.* Carol Stream: Allured Publishing Corporation.

Agrawal, Toshy, Anil S Kotasthane, and Renu Kushwah. 2015. "Genotypic and Phenotypic Diversity of Polyhydroxybutyrate (PHB) Producing Pseudomonas Putida Isolates of Chhattisgarh Region and Assessment of Its Phosphate Solubilizing Ability." *3 Biotech* 5 (1): 45–60. https://doi.org/10.1007/s13205-014-0198-9.

Alhassan, Alhassan M., Musa I. Abdullahi, Ahmad Uba, and Abdulrashid Umar. 2014. "Prenylation of Aromatic Secondary Metabolites: A New Frontier for Development of Novel Drugs." *Tropical Journal of Pharmaceutical Research* 13 (2): 307–14. https://doi.org/10.4314/tjpr.v13i2.22.

Alonso-Gutierrez, Jorge, Rossana Chan, Tanveer S. Batth, Paul D. Adams, Jay D. Keasling, Christopher J. Petzold, and Taek Soon Lee. 2013. "Metabolic Engineering of Escherichia Coli for Limonene and Perillyl Alcohol Production." *Metabolic Engineering* 19: 33–41. https://doi.org/10.1016/j.ymben.2013.05.004.

Ariyawutthiphan, Orapin, Toyoyuki Ose, Atsushi Minami, Sandip Sinde, Muneya Tsuda, Yong Gui Gao, Min Yao, Hideaki Oikawa, and Isao Tanaka. 2012. "Structure Analysis of Geranyl Pyrophosphate Methyltransferase and the Proposed Reaction Mechanism of SAM-Dependent C-Methylation." *Acta Crystallographica Section D: Biological Crystallography* 68 (11): 1558–69. https://doi.org/10.1107/S0907444912038486.

Ariyawutthiphan, Orapin, Toyoyuki Ose, Muneya Tsuda, YongGui Gao, Min Yao, Atsushi Minami, Hideaki Oikawa, and Isao Tanaka. 2011. "Crystallization and Preliminary X-Ray Crystallographic Study of a Methyltransferase Involved in 2-Methylisoborneol Biosynthesis in Streptomyces Lasaliensis." *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications* 67 (Pt 3): 417–20. https://doi.org/10.1107/S1744309110051523.

Arnold, Konstantin, Lorenza Bordoli, Jürgen Kopp, and Torsten Schwede. 2006. "The SWISS-MODEL Workspace: A Web-Based Environment for Protein Structure Homology Modelling." *Bioinformatics* 22 (2): 195–201. https://doi.org/10.1093/bioinformatics/bti770.

Bach, Thomas J. 1986. "Hydroxymethylglutaryl-CoA Reductase, a Key Enzyme in Phytosterol Synthesis?" *Lipids* 21 (1): 82–88. https://doi.org/10.1007/BF02534307.

Banerjee, A, and T D Sharkey. 2014. "Methylerythritol 4-Phosphate (MEP) Pathway Metabolic Regulation." *Nat. Prod. Rep.* 31 (8): 1043–55. https://doi.org/10.1039/C3NP70124G.

Barredo, Jose L, Carlos García-Estrada, Katarina Kosalkova, and Carlos Barreiro. 2017. "Biosynthesis of Astaxanthin as a Main Carotenoid in the Heterobasidiomycetous Yeast Xanthophyllomyces Dendrorhous." *Journal of Fungi* . https://doi.org/10.3390/jof3030044.

Barreiro, Eliezer J, Arthur E Kümmerle, and Carlos A M Fraga. 2011. "The Methylation Effect in Medicinal Chemistry." *Chemical Reviews* 111 (9): 5215–46. https://doi.org/10.1021/cr200060g.

Bathe, Ulschan, and Alain Tissier. 2019. "Cytochrome P450 Enzymes: A Driving Force of Plant Diterpene Diversity." *Phytochemistry* 161 (December 2018): 149–62. https://doi.org/10.1016/j.phytochem.2018.12.003.

Benkert, Pascal, Marco Biasini, and Torsten Schwede. 2011. "Toward the Estimation of the Absolute Quality of Individual Protein Structure Models." *Bioinformatics* 27 (3): 343–50. https://doi.org/10.1093/bioinformatics/btq662.

Bentley, Ronald. 2006. "The Nose as a Stereochemist. Enantiomers and Odor." *Chemical Reviews* 106 (9): 4099–4112. https://doi.org/10.1021/cr050049t.

Biasini, Marco, Stefan Bienert, Andrew Waterhouse, Konstantin Arnold, Gabriel Studer, Tobias Schmidt, Florian Kiefer, et al. 2014. "SWISS-MODEL: Modelling Protein Tertiary and Quaternary Structure Using Evolutionary Information." *Nucleic Acids Research* 42 (W1): 252–58. https://doi.org/10.1093/nar/gku340.

Biggar, Kyle K, and Shawn S.-C. Li. 2015. "Non-Histone Protein Methylation as a Regulator of Cellular Signalling and Function." *Nature Reviews Molecular Cell Biology* 16 (1): 5–17. https://doi.org/10.1038/nrm3915.

Blaser, Hans Ulrich. 2013. "Chirality and Its Implications for the Pharmaceutical Industry." *Rendiconti Lincei* 24 (3): 213–16. https://doi.org/10.1007/s12210-012-0220-2.

Brocks, J J, G A Logan, R Buick, and R E Summons. 1999. "Archean Molecular Fossils and the Early Rise of Eukaryotes." *Science (New York, N.Y.)* 285 (5430): 1033–36. https://doi.org/10.1126/science.285.5430.1033.

Cane, David E., Xiaofei He, Seiji Kobayashi, Satoshi Omura, and Haruo Ikeda. 2006. "Geosmin Biosynthesis in Streptomyces Avermitilis. Molecular Cloning, Expression, and Mechanistic Study of the Germacradienol/Geosmin Synthase." *Journal of Antibiotics* 59 (8): 471–79. https://doi.org/10.1038/ja.2006.66.

Cane, David E., and Haruo Ikeda. 2012. "Exploration and Mining of the Bacterial Terpenome." *Accounts of Chemical Research* 45 (3): 463–72. https://doi.org/10.1021/ar200198d.

Casey, Patrick J, and Miguel C Seabra. 1996. "Protein Prenyltransferases." *Journal of Biological Chemistry* 271 (10): 5289–92. https://doi.org/10.1074/jbc.271.10.5289.

Chandran, Sunil S., James T. Kealey, and Christopher D. Reeves. 2011. "Microbial Production of Isoprenoids." *Process Biochemistry* 46 (9): 1703–10.

https://doi.org/10.1016/j.procbio.2011.05.012.

Chater, Keith F. 2016. "Recent Advances in Understanding Streptomyces." *F1000Research* 5 (0): 1–16. https://doi.org/10.12688/f1000research.9534.1.

Chen, Xi, Emmanuel Mukwaya, Man Sau Wong, and Yan Zhang. 2014. "A Systematic Review on Biological Activities of Prenylated Flavonoids." *Pharmaceutical Biology* 52 (5): 655–60. https://doi.org/10.3109/13880209.2013.853809.

Chou, Wayne K. W., Haruo Ikeda, and David E. Cane. 2011. "Cloning and Characterization of Pfl_1841, a 2-Methylenebornane Synthase in Pseudomonas Fluorescens PfO-1." *Tetrahedron* 67 (35): 6627–32. https://doi.org/10.1016/j.tet.2011.05.084.

Christianson, David W. 2017. "Structural and Chemical Biology of Terpenoid Cyclases." *Chemical Reviews* 117 (17): 11570–648. https://doi.org/10.1021/acs.chemrev.7b00287.

Cordoba, Elizabeth, Mari Salmi, and Patricia León. 2009. "Unravelling the Regulatory Mechanisms That Modulate the MEP Pathway in Higher Plants." *Journal of Experimental Botany* 60 (10): 2933–43. https://doi.org/10.1093/jxb/erp190.

Cornforth, J W, Rita H Cornforth, G Popjak, and L Yengoyan. 1966. "Studies on the Biosynthesis of Cholesterol." *Journal of Biological Chemistry* 241 (17): 3970–87. https://doi.org/10.1016/S0003-9861(53)80011-1.

Davis, B L, L A Goldblatt, and S Palkin. 1946. "Production of Isoprene from Turpentine Derivatives." *Industrial & Engineering Chemistry* 38 (1): 53–57. https://doi.org/10.1021/ie50433a028.

Davis, Edward M, and Rodney Croteau. 2000. "Cyclization Enzymes in the Biosynthesis of Monoterpenes, Sesquiterpenes, and Diterpenes." In *Biosynthesis: Aromatic Polyketides, Isoprenoids, Alkaloids*, edited by Finian J Leeper and John C Vederas, 53–95. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-48146-X_2.

Degenhardt, Jorg, Tobias G Kollner, and Jonathan Gershenzon. 2009. "Monoterpene and Sesquiterpene Synthases and the Origin of Terpene Skeletal Diversity in Plants." *Phytochemistry* 70 (15–16): 1621–37. https://doi.org/10.1016/j.phytochem.2009.07.030.

Deguerry, Fabienne, Laurence Pastore, Shuiqin Wu, Anthony Clark, Joseph Chappell, and Michel Schalk. 2006. "The Diverse Sesquiterpene Profile of Patchouli, Pogostemon Cablin, Is Correlated with a Limited Number of Sesquiterpene Synthases." *Archives of Biochemistry and Biophysics* 454 (2): 123–36. https://doi.org/https://doi.org/10.1016/j.abb.2006.08.006.

Demiray, Melodi, Xiaoping Tang, Thomas Wirth, Juan A. Faraldos, and Rudolf K. Allemann. 2017. "An Efficient Chemoenzymatic Synthesis of Dihydroartemisinic Aldehyde." *Angewandte Chemie - International Edition* 56 (15): 4347–50. https://doi.org/10.1002/anie.201609557.

Dicke, M., and J. J.A. Van Loon. 2000. "Multitrophic Effects of Herbivore-Induced Plant Volatiles in an Evolutionary Context." *Entomologia Experimentalis et Applicata*. Springer Netherlands. https://doi.org/10.1046/j.1570-7458.2000.00736.x.

Dickschat, Jeroen S. 2016. "Bacterial Terpene Cyclases." *Natural Product Reports* 33 (1): 87–110. https://doi.org/10.1039/c5np00102a.

Dickschat, Jeroen S., Thorben Nawrath, Verena Thiel, Brigitte Kunze, Rolf Müller, and Stefan Schulz. 2007. "Biosynthesis of the Off-Flavor 2-Methylisoborneol by the Myxobacterium Nannocystis Exedens." *Angewandte Chemie - International Edition* 46 (43): 8287–90. https://doi.org/10.1002/anie.200702496.

Dien, Stephen Van. 2013. "From the First Drop to the First Truckload: Commercialization of Microbial Processes for Renewable Chemicals." *Current Opinion in Biotechnology* 24 (6): 1061–68. https://doi.org/https://doi.org/10.1016/j.copbio.2013.03.002.

Donald, K. Allen G., Randolph Y. Hampton, and Irving B. Fritz. 1997. "Effects of Overproduction of the Catalytic Domain of 3-Hydroxy-3- Methylglutaryl Coenzyme A Reductase on Squalene Synthesis in Saccharomyces Cerevisiae." *Applied and Environmental Microbiology* 63 (9): 3341–44. https://doi.org/10.1128/aem.63.9.3341-3344.1997.

Drummond, Laura, Max J. Kschowak, Jürgen Breitenbach, Hendrik Wolff, Yi Ming Shi, Jens Schrader, Helge B. Bode, Gerhard Sandmann, and Markus Buchhaupt. 2019. "Expanding the Isoprenoid Building Block Repertoire with an IPP Methyltransferase from Streptomyces Monomycini." *ACS Synthetic Biology* 8 (6): 1303–13. https://doi.org/10.1021/acssynbio.8b00525.

Eiben, Christopher B., Tristan de Rond, Clayton Bloszies, Jennifer Gin, Jennifer Chiniquy, Edward E. K. Baidoo, Christopher J. Petzold, Nathan J. Hillson, Oliver Fiehn, and Jay D. Keasling. 2019. "Mevalonate Pathway Promiscuity Enables Noncanonical Terpene Production." Rapid-communication. *ACS Synthetic Biology* 8: 2238–47. https://doi.org/10.1021/acssynbio.9b00230.

Eisenreich, W, A Bacher, D Arigoni, and F Rohdich. 2004. "Biosynthesis of Isoprenoids via the Non-Mevalonate Pathway." *CMLS, Cell. Mol. Life Sci* 61: 1401–26. https://doi.org/10.1007/s00018-004-3381-z.

Elansary, Hosam O, and Nader A Ashmawy. 2013. "Essential Oils of Mint between Benefits and Hazards." *Journal of Essential Oil Bearing Plants* 16 (4): 429–38. https://doi.org/10.1080/0972060X.2013.813279.

Engels, Benedikt, Pia Dahm, and Stefan Jennewein. 2008. "Metabolic Engineering of Taxadiene Biosynthesis in Yeast as a First Step towards Taxol (Paclitaxel) Production." *Metabolic Engineering* 10 (3): 201–6. https://doi.org/https://doi.org/10.1016/j.ymben.2008.03.001.

Ernster, Lars, and Gustav Dallner. 1995. "Biochemical, Physiological and Medical Aspects of Ubiquinone Function." *Biochimica et Biophysica Acta (BBA) -*

Molecular Basis of Disease 1271 (1): 195–204.
https://doi.org/https://doi.org/10.1016/0925-4439(95)00028-3.

Espah Borujeni, Amin, Anirudh S. Channarasappa, and Howard M. Salis. 2014.
"Translation Rate Is Controlled by Coupled Trade-Offs between Site
Accessibility, Selective RNA Unfolding and Sliding at Upstream Standby Sites."
Nucleic Acids Research 42 (4): 2646–59. https://doi.org/10.1093/nar/gkt1139.

Faraldos, Juan A., Yuxin Zhao, Paul E. O'Maille, Joseph P. Noel, and Robert M.
Coates. 2007. "Interception of the Enzymatic Conversion of Farnesyl
Diphosphate to 5-Epi-Aristolochene by Using a Fluoro Substrate Analogue: 1-
Fluorogermacrene A from (2E,6Z)-6-Fluorofarnesyl Diphosphate."
ChemBioChem 8 (15): 1826–33. https://doi.org/10.1002/cbic.200700398.

Farhat, Wissam, Arne Stamm, Maxime Robert-Monpate, Antonino Biundo, and Per
Olof Syrén. 2019. "Biocatalysis for Terpene-Based Polymers." *Zeitschrift Fur
Naturforschung - Section C Journal of Biosciences* 74 (3–4): 91–100.
https://doi.org/10.1515/znc-2018-0199.

Gelmont, David, Robert A Stein, and James F Mead. 1981. "Isoprene — The Main
Hydrocarbon in Human Breath." *Biochemical and Biophysical Research
Communications* 99 (4): 1456–60. https://doi.org/https://doi.org/10.1016/0006-
291X(81)90782-8.

Gershenzon, Jonathan, and Natalia Dudareva. 2007. "The Function of Terpene
Natural Products in the Natural World." *Nature Chemical Biology* 3 (7): 408–14.
https://doi.org/10.1038/nchembio.2007.5.

Gibson, Daniel G., Lei Young, Ray-Yuan Chuang, J. Craig Venter, Clyde A.
Hutchison, and Hamilton O. Smith. 2009. "Enzymatic Assembly of DNA
Molecules up to Several Hundred Kilobases." *Nature Methods* 6 (5): 343–45.
https://doi.org/10.1038/nmeth.1318.

Goll, Mary Grace, and Timothy H. Bestor. 2005. "Eukaryotic Cytosine
Methyltransferases." *Annual Review of Biochemistry* 74 (1): 481–514.
https://doi.org/10.1146/annurev.biochem.74.010904.153721.

Goodsell, David S. 1999. "The Molecular Perspective: The Ras Oncogene." *The
Oncologist* 4 (3): 263–64. https://doi.org/10.1634/theoncologist.4-3-263.

Govindarajan, R., M. Vijayakumar, and P. Pushpangadan. 2005. "Antioxidant
Approach to Disease Management and the Role of 'Rasayana' Herbs of
Ayurveda." *Journal of Ethnopharmacology* 99 (2): 165–78.
https://doi.org/10.1016/j.jep.2005.02.035.

Guenther, Ernest. 1949. "The Essential Oils." *Nature* 163 (4148): 663–663.
https://doi.org/10.1038/163663c0.

Hall, Dawn E., Jeanne A. Robert, Christopher I. Keeling, Dominik Domanski, Alfonso
Lara Quesada, Sharon Jancsik, Michael A. Kuzyk, Britta Hamberger, Christoph
H. Borchers, and Jörg Bohlmann. 2011. "An Integrated Genomic, Proteomic and
Biochemical Analysis of (+)-3-Carene Biosynthesis in Sitka Spruce (Picea

Sitchensis) Genotypes That Are Resistant or Susceptible to White Pine Weevil." *Plant Journal* 65 (6): 936–48. https://doi.org/10.1111/j.1365-313X.2010.04478.x.

Hanukoglu, Israel. 1992. "Steroidogenic Enzymes: Structure, Function, and Role in Regulation of Steroid Hormone Biosynthesis." *The Journal of Steroid Biochemistry and Molecular Biology* 43 (8): 779–804. https://doi.org/https://doi.org/10.1016/0960-0760(92)90307-5.

Hayashi, Yutaka, Hiroyasu Onaka, Nobuya Itoh, Haruo Seto, and Tohru Dairi. 2007. "Cloning of the Gene Cluster Responsible for Biosynthesis of KS-505a (Longestin), a Unique Tetraterpenoid." *Bioscience, Biotechnology, and Biochemistry* 71 (12): 3072–81. https://doi.org/10.1271/bbb.70477.

Heil, Martin, and Richard Karban. 2010. "Explaining Evolution of Plant Communication by Airborne Signals." *Trends in Ecology and Evolution* 25 (3): 137–44. https://doi.org/10.1016/j.tree.2009.09.010.

Hemmerlin, Andréa, John L Harwood, and Thomas J Bach. 2012. "A Raison d'être for Two Distinct Pathways in the Early Steps of Plant Isoprenoid Biosynthesis?" *Progress in Lipid Research* 51 (2): 95–148. https://doi.org/https://doi.org/10.1016/j.plipres.2011.12.001.

Herty, Chas H., and J. O. Graham. 1914. "Isoprene from Commercial Turpentines." *Journal of Industrial & Engineering Chemistry* 6 (10): 803–4. https://doi.org/10.1021/ie50070a007.

Herz, Rachel S. 2016. "The Role of Odor-Evoked Memory in Psychological and Physiological Health." *Brain Sciences* 6 (3). https://doi.org/10.3390/brainsci6030022.

Herz, Rachel S., and Gerald C. Cupchik. 1992. "An Experimental Characterization of Odor-Evoked Memories in Humans." *Chemical Senses* 17 (5): 519–28. https://doi.org/10.1093/chemse/17.5.519.

Hillier, Stephen G., and Richard Lathe. 2019. "Terpenes, Hormones and Life: Isoprene Rule Revisited." *The Journal of Endocrinology* 242 (2): R9–22. https://doi.org/10.1530/JOE-19-0084.

Hoopes, J Todd, Margaret A Elberson, Renae J Preston, Prasad T Reddy, and Zvi Kelman. 2015. "Chapter Two - Protein Labeling in Escherichia Coli with 2H, 13C, and 15N." In *Isotope Labeling of Biomolecules - Labeling Methods*, edited by Zvi B T - Methods in Enzymology Kelman, 565:27–44. Academic Press. https://doi.org/https://doi.org/10.1016/bs.mie.2015.08.023.

Huber, Tyler D, Brooke R Johnson, Jianjun Zhang, and Jon S Thorson. 2016. "AdoMet Analog Synthesis and Utilization : Current State of the Art." *Current Opinion in Biotechnology* 42: 189–97. https://doi.org/10.1016/j.copbio.2016.07.005.

Hunter, William N. 2007. "The Non-Mevalonate Pathway of Isoprenoid Precursor Biosynthesis." *Journal of Biological Chemistry* 282 (30): 21573–77. https://doi.org/10.1074/jbc.R700005200.

Ignea, Codruta, Marianna Pontini, Mohammed S Motawia, Massimo E Maffei, Antonios M Makris, and Sotirios C Kampranis. 2018. "Synthesis of 11-Carbon Terpenoids in Yeast Using Protein and Metabolic Engineering." *Nature Chemical Biology* 14 (12): 1090–98. https://doi.org/10.1038/s41589-018-0166-5.

Immethun, Cheryl M., Allison G. Hoynes-O'Connor, Andrea Balassy, and Tae Seok Moon. 2013. "Microbial Production of Isoprenoids Enabled by Synthetic Biology." *Frontiers in Microbiology* 4 (APR): 1–8. https://doi.org/10.3389/fmicb.2013.00075.

Incardona, John P, and Suzanne Eaton. 2000. "Cholesterol in Signal Transduction." *Current Opinion in Cell Biology* 12 (2): 193–203. https://doi.org/https://doi.org/10.1016/S0955-0674(99)00076-9.

Ito, Michio, Masaki Kobayashi, Tanetoshi Koyama, and Kyozo Ogura. 1987. "Stereochemical Analysis of Prenyltransferase Reactions Leading to (Z)- and (E)-Polyprenyl Chains." *Biochemistry* 26 (15): 4745–50. https://doi.org/10.1021/bi00389a022.

Izaguirre, G., C. J. Hwang, S. W. Krasner, and M. J. McGuire. 1982. "Geosmin and 2-Methylisoborneol from Cyanobacteria in Three Water Supply Systems." *Applied and Environmental Microbiology* 43 (3): 708–14. https://doi.org/10.1128/aem.43.3.708-714.1982.

Jameson, Paula E., Pragathi Dhandapani, Jiancheng Song, Marek Zatloukal, Miroslav Strnad, Mitja N.P. Remus-Emsermann, Rudolf O. Schlechter, and Ondrej Novák. 2019. "The Cytokinin Complex Associated with Rhodococcus Fascians: Which Compounds Are Critical for Virulence?" *Frontiers in Plant Science* 10 (May): 1–13. https://doi.org/10.3389/fpls.2019.00674.

Jansson, Anna, Hanna Koskiniemi, Anna Erola, Jessy Wang, Pekka Mäntsälä, Gunter Schneider, and Jarmo Niemi. 2005. "Aclacinomycin 10-Hydroxylase Is a Novel Substrate-Assisted Hydroxylase Requiring S-Adenosyl-L-Methionine as Cofactor." *The Journal of Biological Chemistry* 280 (5): 3636–44. https://doi.org/10.1074/jbc.M412095200.

Jiang, Jiaoyang, Xiaofei He, and David E. Cane. 2007. "Biosynthesis of the Earthy Odorant Geosmin by a Bifunctional Streptomyces Coelicolor Enzyme." *Nature Chemical Biology* 3 (11): 711–15. https://doi.org/10.1038/nchembio.2007.29.

Jiang, Zuodong, Chase Kempinski, Caroline J Bush, S Eric Nybo, and Joe Chappell. 2015. "Engineering Triterpene and Methylated Triterpene Production in Plants Provides Biochemical and Physiological Insights into Terpene Metabolism." *Plant Physiology* 170 (February): 702–16. https://doi.org/10.1104/pp.15.01548.

Jones, Stephanie E., and Marie A. Elliot. 2017. "Streptomyces Exploration: Competition, Volatile Communication and New Bacterial Behaviours." *Trends in Microbiology* 25 (7): 522–31. https://doi.org/10.1016/j.tim.2017.02.001.

Jones, Stephanie E, Louis Ho, Christiaan A Rees, Jane E Hill, Justin R Nodwell, and Marie A Elliot. 2017. "Streptomyces Exploration Is Triggered by Fungal

Interactions and Volatile Signals." *ELife* 6 (January): e21738.
https://doi.org/10.7554/eLife.21738.

Jones, Stephanie E, Christine A Pham, Matthew P Zambri, Joseph McKillip, Erin E
Carlson, and Marie A Elliot. 2019. "Streptomyces Volatile Compounds Influence
Exploration and Microbial Community Dynamics by Altering Iron Availability."
Edited by Julian E Davies. *MBio* 10 (2). https://doi.org/10.1128/mBio.00171-19.

Jordan, Sean F, Eloise Nee, and Nick Lane. 2019. "Isoprenoids Enhance the
Stability of Fatty Acid Membranes at the Emergence of Life Potentially Leading
to an Early Lipid Divide." *Interface Focus* 9 (6): 20190067.
https://doi.org/10.1098/rsfs.2019.0067.

Kagan, R M, and S Clarke. 1994. "Widespread Occurrence of Three Sequence
Motifs in Diverse S-Adenosylmethionine-Dependent Methyltransferases
Suggests a Common Structure for These Enzymes." *Archives of Biochemistry
and Biophysics* 310 (2): 417–27. https://doi.org/10.1006/abbi.1994.1187.

Kampranis, Sotirios C., and Antonios M. Makris. 2012. "Developing a Yeast Cell
Factory for the Production of Terpenoids." *Computational and Structural
Biotechnology Journal*. Research Network of Computational and Structural
Biotechnology. https://doi.org/10.5936/csbj.201210006.

Kapoor, L. D. 2017. *Handbook of Ayurvedic Medicinal Plants: Herbal Reference
Library*. *Handbook of Ayurvedic Medicinal Plants: Herbal Reference Library*.
CRC Press. https://doi.org/10.1201/9780203719473.

Kieser, Tobias, Mervyn J Bibb, Mark J Buttner, Keith F Chater, David A Hopwood,
and others. 2000. *Practical Streptomyces Genetics*. Vol. 291. John Innes
Foundation Norwich.

Kim, Seung-Su, Min-Ho Lee, and Mi-Ock Lee. 2020. "Histone Methyltransferases
Regulate the Transcriptional Expression of ERα and the Proliferation of
Tamoxifen-Resistant Breast Cancer Cells." *Breast Cancer Research and
Treatment* 180 (1): 45–54. https://doi.org/10.1007/s10549-019-05517-0.

Kirby, James, Kevin L. Dietzel, Gale Wichmann, Rossana Chan, Eugene Antipov,
Nathan Moss, Edward E.K. Baidoo, et al. 2016. "Engineering a Functional 1-
Deoxy-D-Xylulose 5-Phosphate (DXP) Pathway in Saccharomyces Cerevisiae."
Metabolic Engineering 38 (October): 494–503.
https://doi.org/10.1016/j.ymben.2016.10.017.

Klimašauskas, Saulius, and Elmar Weinhold. 2007. "A New Tool for Biotechnology:
AdoMet-Dependent Methyltransferases." *Trends in Biotechnology* 25 (3): 99–
104. https://doi.org/https://doi.org/10.1016/j.tibtech.2007.01.006.

Kobayashi, Masaki, Michio Ito, Tanetoshi Koyama, and Kyozo Ogura. 1985.
"Stereochemistry of the Carbon to Carbon Bond Formation in the Biosynthesis
of Polyprenyl Chains with Z Double Bonds. Studies with Undecaprenyl-
Pyrophosphate Synthetase." *Journal of the American Chemical Society* 107
(15): 4588–89. https://doi.org/10.1021/ja00301a057.

Köksal, Mustafa, Wayne K W Chou, David E. Cane, and David W. Christianson. 2012. "Structure of Geranyl Diphosphate C -Methyltransferase from Streptomyces Coelicolor and Implications for the Mechanism of Isoprenoid Modification." *Biochemistry* 51 (14): 3003–10. https://doi.org/10.1021/bi300109c.

Komatsu, Mamoru, Muneya Tsuda, Satoshi Omura, Hideaki Oikawa, and Haruo Ikeda. 2008. "Identification and Functional Analysis of Genes Controlling Biosynthesis of 2-Methylisoborneol." *Proceedings of the National Academy of Sciences of the United States of America* 105 (21): 7422–27. https://doi.org/10.1073/pnas.0802312105.

Koyama, T. 1999. "Molecular Analysis of Prenyl Chain Elongating Enzymes." *Bioscience, Biotechnology, and Biochemistry* 63 (10): 1671–76. https://doi.org/10.1271/bbb.63.1671.

Koyama, Tanetoshi, Yukio Katsuki, and Kyozo Ogura. 1983. "Studies on Isopentenyl Pyrophosphate Isomerase with Artificial Substrates: Z-E Isomerization of Z-3-Methyl-3-Pentenyl Pyrophosphate." *Bioorganic Chemistry* 12 (1): 58–70. https://doi.org/10.1016/0045-2068(83)90007-X.

Koyama, Tanetoshi, Akio Saito, Kyozo Ogura, and Shuichi Seto. 1980. "Substrate Specificity of Farnesylpyrophosphate Synthetase. Application to Asymmetric Synthesis." *Journal of the American Chemical Society* 102 (10): 3614–18. https://doi.org/10.1021/ja00530a050.

Kozbial, Piotr Z., and Arcady R. Mushegian. 2005. "Natural History of S-Adenosylmethionine-Binding Proteins." *BMC Structural Biology* 5. https://doi.org/10.1186/1472-6807-5-19.

Krieg, Thomas, Anne Sydow, Sonja Faust, Ina Huth, and Dirk Holtmann. 2018. "CO2 to Terpenes: Autotrophic and Electroautotrophic α-Humulene Production with Cupriavidus Necator." *Angewandte Chemie International Edition* 57 (7): 1879–82. https://doi.org/10.1002/anie.201711302.

Kschowak, Max J., Hannah Wortmann, Jeroen S. Dickschat, Jens Schrader, and Markus Buchhaupt. 2018. "Heterologous Expression of 2-Methylisoborneol / 2 Methylenebornane Biosynthesis Genes in Escherichia Coli Yields Novel C11-Terpenes." *Plos One* 13 (4): e0196082. https://doi.org/10.1371/journal.pone.0196082.

Kuntz, Kevin W., John E. Campbell, Heike Keilhack, Roy M. Pollock, Sarah K. Knutson, Margaret Porter-Scott, Victoria M. Richon, et al. 2016. "The Importance of Being Me: Magic Methyls, Methyltransferase Inhibitors, and the Discovery of Tazemetostat." *Journal of Medicinal Chemistry* 59 (4): 1556–64. https://doi.org/10.1021/acs.jmedchem.5b01501.

Kuzuyama, Tomohisa. 2002. "Mevalonate and Nonmevalonate Pathways for the Biosynthesis of Isoprene Units." *Bioscience, Biotechnology, and Biochemistry* 66 (8): 1619–27. https://doi.org/10.1271/bbb.66.1619.

———. 2017. "Biosynthetic Studies on Terpenoids Produced by Streptomyces."

Journal of Antibiotics 70 (7): 811–18. https://doi.org/10.1038/ja.2017.12.

Lancaster, Jason, Ashot Khrimian, Sharon Young, Bryan Lehner, Katrin Luck, Anna Wallingford, Saikat Kumar B. Ghosh, et al. 2018. "De Novo Formation of an Aggregation Pheromone Precursor by an Isoprenyl Diphosphate Synthase-Related Terpene Synthase in the Harlequin Bug." *Proceedings of the National Academy of Sciences of the United States of America* 115 (37): E8634–41. https://doi.org/10.1073/pnas.1800008115.

Lee, SangYup, Diethard Mattanovich, and Antonio Villaverde. 2012. "Systems Metabolic Engineering, Industrial Biotechnology and Microbial Cell Factories." *Microbial Cell Factories* 11 (1): 156. https://doi.org/10.1186/1475-2859-11-156.

Levitan, Irena, Dev Singh, and Avia Rosenhouse-Dantsker. 2014. "Cholesterol Binding to Ion Channels ." *Frontiers in Physiology* . https://www.frontiersin.org/article/10.3389/fphys.2014.00065.

Lima Procópio, Rudi Emerson de, Ingrid Reis da Silva, Mayra Kassawara Martins, João Lúcio de Azevedo, and Janete Magali de Araújo. 2012. "Antibiotics Produced by Streptomyces." *Brazilian Journal of Infectious Diseases* 16 (5): 466–71. https://doi.org/10.1016/j.bjid.2012.08.014.

Liscombe, David K., Gordon V. Louie, and Joseph P. Noel. 2012. "Architectures, Mechanisms and Molecular Evolution of Natural Product Methyltransferases." *Natural Product Reports* 29 (10): 1238–50. https://doi.org/10.1039/c2np20029e.

Liu, Chang Jun, Bettina E. Deavours, Stéphane B. Richard, Jean Luc Ferrer, Jack W. Blount, David Huhman, Richard A. Dixon, and Joseph P. Noel. 2006. "Structural Basis for Dual Functionality of Isoflavonoid O-Methyltransferases in the Evolution of Plant Defense Responses." *Plant Cell* 18 (12): 3656–69. https://doi.org/10.1105/tpc.106.041376.

Liu, Chun Li, Tian Tian, Jorge Alonso-Gutierrez, Brett Garabedian, Shuai Wang, Edward E.K. Baidoo, Veronica Benites, et al. 2018. "Renewable Production of High Density Jet Fuel Precursor Sesquiterpenes from Escherichia Coli." *Biotechnology for Biofuels* 11 (1): 1–15. https://doi.org/10.1186/s13068-018-1272-z.

Lombard, Jonathan, and David Moreira. 2011. "Origins and Early Evolution of the Mevalonate Pathway of Isoprenoid Biosynthesis in the Three Domains of Life." *Molecular Biology and Evolution* 28 (1): 87–99. https://doi.org/10.1093/molbev/msq177.

Manteca, Ángel, and Paula Yagüe. 2019. "Streptomyces as a Source of Antimicrobials: Novel Approaches to Activate Cryptic Secondary Metabolite Pathways." In *Antimicrobials, Antibiotic Resistance, Antibiofilm Strategies and Activity Methods*, edited by Sahra Kırmusaoğlu. IntechOpen. https://doi.org/10.5772/intechopen.81812.

Marley, Jonathan, Min Lu, and Clay Bracken. 2001. "A Method for Efficient Isotopic Labeling of Recombinant Proteins." *Journal of Biomolecular NMR* 20 (1): 71–75.

https://doi.org/10.1023/A:1011254402785.

Martin, Jennifer L, and Fiona M McMillan. 2002. "SAM (Dependent) I AM: The S-Adenosylmethionine-Dependent Methyltransferase Fold." *Current Opinion in Structural Biology* 12 (6): 783–93. https://doi.org/https://doi.org/10.1016/S0959-440X(02)00391-3.

McGenity, Terry J, Andrew T Crombie, and J Colin Murrell. 2018. "Microbial Cycling of Isoprene, the Most Abundantly Produced Biological Volatile Organic Compound on Earth." *The ISME Journal* 12 (4): 931–41. https://doi.org/10.1038/s41396-018-0072-6.

Melendez, I, K Grice, and L Schwark. 2013. "Exceptional Preservation of Palaeozoic Steroids in a Diagenetic Continuum." *Scientific Reports* 3: 2768. https://doi.org/10.1038/srep02768.

Michalak, Ewa M, Marian L Burr, Andrew J Bannister, and Mark A Dawson. 2019. "The Roles of DNA, RNA and Histone Methylation in Ageing and Cancer." *Nature Reviews Molecular Cell Biology* 20 (10): 573–89. https://doi.org/10.1038/s41580-019-0143-1.

Miller, David J., Fanglei Yu, and Rudolf K. Allemann. 2007. "Aristolochene Synthase-Catalyzed Cyclization of 2-Fluorofarnesyl- Diphosphate to 2-Fluorogermacrene A." *ChemBioChem* 8 (15): 1819–25. https://doi.org/10.1002/cbic.200700219.

Mordhorst, Silja, Jutta Siegrist, M Michael, Michael Richter, and Jennifer N Andexer. 2017. "Catalytic Alkylation Using a Cyclic S -Adenosylmethionine Regeneration System." *Angewandte Chemie - International Edition* 56: 4037–41. https://doi.org/10.1002/anie.201611038.

Mori, Takahiro, Lihan Zhang, Takayoshi Awakawa, Shotaro Hoshino, Masahiro Okada, Hiroyuki Morita, and Ikuro Abe. 2016. "Manipulation of Prenylation Reactions by Structure-Based Engineering of Bacterial Indolactam Prenyltransferases." *Nature Communications* 7: 1–11. https://doi.org/10.1038/ncomms10849.

Mukai, Rie. 2018. "Prenylation Enhances the Biological Activity of Dietary Flavonoids by Altering Their Bioavailability." *Bioscience, Biotechnology, and Biochemistry* 82 (2): 207–15. https://doi.org/10.1080/09168451.2017.1415750.

Murn, Jernej, and Yang Shi. 2017. "The Winding Path of Protein Methylation Research: Milestones and New Frontiers." *Nature Reviews Molecular Cell Biology* 18 (8): 517–27. https://doi.org/10.1038/nrm.2017.35.

Murrell, J Colin, Terry J McGenity, and Andrew T Crombie. 2020. "Microbial Metabolism of Isoprene: A Much-Neglected Climate-Active Gas." https://doi.org/https://doi.org/10.1099/mic.0.000931.

Nagaki, Masahiko, Takumi Ichijo, Rikiya Kobashi, Yusuke Yagihashi, Tohru Musashi, Jun Kawakami, Norimasa Ohya, Takeshi Gotoh, and Hiroshi Sagami. 2012. "Substrate Specificities of E- and Z-Farnesyl Diphosphate Synthases with Substrate Analogs." *Journal of Molecular Catalysis B: Enzymatic* 80: 1–6.

https://doi.org/10.1016/j.molcatb.2012.04.006.

Neighbors, Jeffrey D. 2018. "The Mevalonate Pathway and Terpenes: A Diversity of Chemopreventatives." *Current Pharmacology Reports* 4 (2): 157–69. https://doi.org/10.1007/s40495-018-0128-3.

Nič, Miloslav, Jiří Jirát, Bedřich Košata, Aubrey Jenkins, and Alan McNaught, eds. 2009. *IUPAC Compendium of Chemical Terminology*. Research Triangle Park, NC: IUPAC. https://doi.org/10.1351/goldbook.

Nouioui, Imen, Lorena Carro, Marina García-López, Jan P Meier-Kolthoff, Tanja Woyke, Nikos C Kyrpides, Rüdiger Pukall, Hans-Peter Klenk, Michael Goodfellow, and Markus Göker. 2018. "Genome-Based Taxonomic Classification of the Phylum Actinobacteria." *Frontiers in Microbiology* 9: 2007. https://doi.org/10.3389/fmicb.2018.02007.

Ohvo-Rekilä, Henna, Bodil Ramstedt, Petra Leppimäki, and J Peter Slotte. 2002. "Cholesterol Interactions with Phospholipids in Membranes." *Progress in Lipid Research* 41 (1): 66–97. https://doi.org/https://doi.org/10.1016/S0163-7827(01)00020-0.

Olsson, Kim, Simon Carlsen, Angelika Semmler, Ernesto Simón, Michael Dalgaard Mikkelsen, and Birger Lindberg Møller. 2016. "Microbial Production of Next-Generation Stevia Sweeteners." *Microbial Cell Factories* 15 (1): 207. https://doi.org/10.1186/s12934-016-0609-1.

Ourisson, G, and P Albrecht. 1992. "Hopanoids. 1. Geohopanoids: The Most Abundant Natural Products on Earth?" *Accounts of Chemical Research* 25: 398–402. https://doi.org/10.1021/ar00021a003.

Ozaki, Taro, Sandip S. Shinde, Lei Gao, Ryo Okuizumi, Chengwei Liu, Yasushi Ogasawara, Xiaoguang Lei, Tohru Dairi, Atsushi Minami, and Hideaki Oikawa. 2018. "Enzymatic Formation of a Skipped Methyl-Substituted Octaprenyl Side Chain of Longestin (KS-505a): Involvement of Homo-IPP as a Common Extender Unit." *Angewandte Chemie - International Edition* 57 (22): 6629–32. https://doi.org/10.1002/anie.201802116.

Payne, Anita H, and Dale B Hales. 2004. "Overview of Steroidogenic Enzymes in the Pathway from Cholesterol to Active Steroid Hormones." *Endocrine Reviews* 25 (6): 947–70. https://doi.org/10.1210/er.2003-0030.

Pettersen, Eric F., Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. 2004. "UCSF Chimera - A Visualization System for Exploratory Research and Analysis." *Journal of Computational Chemistry* 25 (13): 1605–12. https://doi.org/10.1002/jcc.20084.

Phelan, Ryan M, Daniel Sachs, Shayne J Petkiewicz, Jesus F Barajas, Jacquelyn M Blake-Hedges, Mitchell G Thompson, Amanda Reider Apel, Blake J Rasor, Leonard Katz, and Jay D Keasling. 2017. "Development of Next Generation Synthetic Biology Tools for Use in Streptomyces Venezuelae." *ACS Synthetic*

Biology 6 (1): 159–66. https://doi.org/10.1021/acssynbio.6b00202.

Polakowski, T, U Stahl, and C Lang. 1998. "Overexpression of a Cytosolic Hydroxymethylglutaryl-CoA Reductase Leads to Squalene Accumulation in Yeast." *Applied Microbiology and Biotechnology* 49 (1): 66–71. https://doi.org/10.1007/s002530051138.

Popják, G, and J W Cornforth. 1966. "Substrate Stereochemistry in Squalene Biosynthesis: The First Ciba Medal Lecture." *The Biochemical Journal* 101 (3): 553.b4-568. http://www.ncbi.nlm.nih.gov/pubmed/16742426.

Poulter, C D, D M Satterwhite, and H C Rilling. 1976. "Letter: Prenyltransferase. The Mechanism of the Reaction." *Journal of the American Chemical Society* 98 (11): 3376–77. https://doi.org/10.1021/bi00650a019.

Poulter, C Dale, and Hans C Rilling. 1978. "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): 307–13. https://doi.org/10.1021/ar50128a004.

Prather, Kristala L Jones, and Collin H Martin. 2008. "De Novo Biosynthetic Pathways: Rational Design of Microbial Chemical Factories." *Current Opinion in Biotechnology* 19 (5): 468–74. https://doi.org/https://doi.org/10.1016/j.copbio.2008.07.009.

Radhika, Venkatesan, Nanae Ueda, Yuuri Tsuboi, Mikiko Kojima, Jun Kikuchi, Takuji Kudo, and Hitoshi Sakakibara. 2015. "Methylated Cytokinins from the Phytopathogen *Rhodococcus Fascians* Mimic Plant Hormone Activity." *Plant Physiology* 169 (2): 1118–26. https://doi.org/10.1104/pp.15.00787.

Reimer, Lorenz Christian, Anna Vetcininova, Joaquim Sardà Carbasse, Carola Söhngen, Dorothea Gleim, Christian Ebeling, and Jörg Overmann. 2019. "Bac Dive in 2019: Bacterial Phenotypic Data for High-Throughput Biodiversity Analysis." *Nucleic Acids Research* 47 (D1): D631–36. https://doi.org/10.1093/nar/gky879.

Reuss, Stephan Von, Dajana Domik, Marie Chantal Lemfack, Nancy Magnus, Marco Kai, Teresa Weise, and Birgit Piechulla. 2018. "Sodorifen Biosynthesis in the Rhizobacterium Serratia Plymuthica Involves Methylation and Cyclization of MEP-Derived Farnesyl Pyrophosphate by a SAM-Dependent C-Methyltransferase." Research-article. *Journal of the American Chemical Society* 140 (37): 11855–62. https://doi.org/10.1021/jacs.8b08510.

Rising, Kathleen A., Charisse M. Crenshaw, Hyun Jo Koo, Thangaiah Subramanian, Kareem A.H. Chehade, Courtney Starks, Keith D. Allen, et al. 2015. "Formation of a Novel Macrocyclic Alkaloid from the Unnatural Farnesyl Diphosphate Analogue Anilinogeranyl Diphosphate by 5-Epi-Aristolochene Synthase." *ACS Chemical Biology* 10 (7): 1729–36. https://doi.org/10.1021/acschembio.5b00145.

Ro, Dae-Kyun, Eric M. Paradise, Mario Ouellet, Karl J. Fisher, Karyn L. Newman,

John M. Ndungu, Kimberly A. Ho, et al. 2006. "Production of the Antimalarial Drug Precursor Artemisinic Acid in Engineered Yeast." *Nature* 440 (7086): 940–43. https://doi.org/10.1038/nature04640.

Rohdich, Felix, Klaus Kis, Adelbert Bacher, and Wolfgang Eisenreich. 2001. "The Non-Mevalonate Pathway of Isoprenoids: Genes, Enzymes and Intermediates." *Current Opinion in Chemical Biology* 5 (5): 535–40. https://doi.org/10.1016/S1367-5931(00)00240-4.

Rohdich, Felix, Juraithip Wungsintaweekul, Monika Fellermeier, Silvia Sagner, Stefan Herz, Klaus Kis, Wolfgang Eisenreich, Adelbert Bacher, and Meinhart H Zenk. 1999. "Isoprenoids : YgbP Protein of Escherichia Coli Catalyzes the Formation of 4-Diphosphocytidyl-2-C-Methylerythritol."

Rowinsky, Eric K, and Ross C Donehower. 1995. "Paclitaxel (Taxol)." *New England Journal of Medicine* 332 (15): 1004–14. https://doi.org/10.1056/NEJM199504133321507.

Rudolf, Jeffrey D, and Chin-Yuan Chang. 2020. "Terpene Synthases in Disguise: Enzymology, Structure, and Opportunities of Non-Canonical Terpene Synthases." *Natural Product Reports* 37 (3): 425–63. https://doi.org/10.1039/c9np00051h.

Ruzicka, L. 1953. "The Isoprene Rule and the Biogenesis of Terpenic Compounds." *Experientia* 9 (10): 357–67. https://doi.org/10.1007/BF02167631.

Sadler, Joanna C, Luke D Humphreys, Radka Snajdrova, and Glenn A Burley. 2017. "A Tandem Enzymatic Sp 2 -C-Methylation Process : Coupling in Situ S-Adenosyl- l -Methionine Formation with Methyl Transfer," 992–95. https://doi.org/10.1002/cbic.201700115.

Saeidnia, S. 2014. "Turpentine." In *Encyclopedia of Toxicology*, 860–65. Elsevier. https://doi.org/10.1016/B978-0-12-386454-3.01034-4.

Salah-Bey, Khadidja, Veronique Blanc, and Charles J Thompson. 1995. "Stress-Activated Expression of a Streptomyces Pristinaespiralis Multidrug Resistance Gene (Ptr) in Various Streptomyces Spp. and Escherichia Coli." *Molecular Microbiology* 17 (5): 1001–12. https://doi.org/10.1111/j.1365-2958.1995.mmi_17051001.x.

Salis, Howard M., Ethan A. Mirsky, and Christopher A. Voigt. 2009. "Automated Design of Synthetic Ribosome Binding Sites to Control Protein Expression." *Nature Biotechnology* 27 (10): 946–50. https://doi.org/10.1038/nbt.1568.

Scalcinati, Gionata, Christoph Knuf, Siavash Partow, Yun Chen, Jérôme Maury, Michel Schalk, Laurent Daviet, Jens Nielsen, and Verena Siewers. 2012. "Dynamic Control of Gene Expression in Saccharomyces Cerevisiae Engineered for the Production of Plant Sesquitepene α-Santalene in a Fed-Batch Mode." *Metabolic Engineering* 14 (2): 91–103. https://doi.org/https://doi.org/10.1016/j.ymben.2012.01.007.

Scalcinati, Gionata, Siavash Partow, Verena Siewers, Michel Schalk, Laurent Daviet,

and Jens Nielsen. 2012. "Combined Metabolic Engineering of Precursor and Co-Factor Supply to Increase α-Santalene Production by Saccharomyces Cerevisiae." *Microbial Cell Factories* 11 (1): 117. https://doi.org/10.1186/1475-2859-11-117.

Schalk, Michel, Laurence Pastore, Marco A Mirata, Samretthy Khim, Marina Schouwey, Fabienne Deguerry, Virginia Pineda, Letizia Rocci, and Laurent Daviet. 2012. "Toward a Biosynthetic Route to Sclareol and Amber Odorants." *Journal of the American Chemical Society* 134 (46): 18900–903. https://doi.org/10.1021/ja307404u.

Schlesier, Julia, Jutta Siegrist, Stefan Gerhardt, Annette Erb, Simone Blaesi, Michael Richter, Oliver Einsle, and Jennifer N. Andexer. 2013. "Structural and Functional Characterisation of the Methionine Adenosyltransferase from Thermococcus Kodakarensis." *BMC Structural Biology* 13 (1). https://doi.org/10.1186/1472-6807-13-22.

Schmidt, Ruth, Desalegn W. Etalo, Victor de Jager, Saskia Gerards, Hans Zweers, Wietse de Boer, and Paolina Garbeva. 2016. "Microbial Small Talk: Volatiles in Fungal-Bacterial Interactions." *Frontiers in Microbiology* 6 (JAN): 1–12. https://doi.org/10.3389/fmicb.2015.01495.

Schmidt, Ruth, Victor De Jager, Daniela Zühlke, Christian Wolff, Jörg Bernhardt, Jules Beekwilder, Wilfred Van Ijcken, Frank Sleutels, and Wietse De Boer. 2017. "Fungal Volatile Compounds Induce Production of the Secondary Metabolite Sodorifen in Serratia Plymuthica PRI-2C," no. November 2016: 1–14. https://doi.org/10.1038/s41598-017-00893-3.

Schönherr, Heike, and Tim Cernak. 2013. "Profound Methyl Effects in Drug Discovery and a Call for New C-H Methylation Reactions." *Angewandte Chemie - International Edition* 52 (47): 12256–67. https://doi.org/10.1002/anie.201303207.

Schulz, Stefan, and Jeroen S. Dickschat. 2007. "Bacterial Volatiles: The Smell of Small Organisms." *Natural Product Reports* 24 (4): 814–42. https://doi.org/10.1039/b507392h.

Serras, Leanna. 2019. "Scent Marketing." 2019. https://www.fragrancex.com/blog/scent-marketing/.

Shimizu, Naoto, Tanetoshi Koyama, and Kyozo Ogura. 1998. "Molecular Cloning, Expression, and Purification of Undecaprenyl Diphosphate Synthase: NO SEQUENCE SIMILARITY BETWEEN E- ANDZ-PRENYL DIPHOSPHATE SYNTHASES ." *Journal of Biological Chemistry* 273 (31): 19476–81. https://doi.org/10.1074/jbc.273.31.19476.

Shin-ya, Kazuo, Kazuo Furihata, Yoichi Hayakawa, and Haruo Seto. 1990. "Biosynthetic Studies of Naphterpin, a Terpenoid Metabolite of Streptomyces." *Tetrahedron Letters* 31 (42): 6025–26. https://doi.org/10.1016/S0040-4039(00)98019-5.

Shin, Woong-Hee, and Daisuke Kihara. 2019. "55 Years of the Rossmann Fold." In *Protein Supersecondary Structures*, edited by Alexander E. Kister, 1–13. https://doi.org/10.1007/978-1-4939-9161-7_1.

Siegrist, Jutta, Julia Netzer, Silja Mordhorst, Lukas Karst, Stefan Gerhardt, Oliver Einsle, Michael Richter, and Jennifer N Andexer. 2017. "Functional and Structural Characterisation of a Bacterial O -Methyltransferase and Factors Determining Regioselectivity" 591: 312–21. https://doi.org/10.1002/1873-3468.12530.

Singh, Shanteri, Jianjun Zhang, Tyler D Huber, Manjula Sunkara, Katherine Hurley, Randal D Goff, Guojun Wang, et al. 2014. "Facile Chemoenzymatic Strategies for the Synthesis and Utilization of S -Adenosyl- l -Methionine Analogues ** Angewandte" 40536: 3965–69. https://doi.org/10.1002/anie.201308272.

Sofias, Alexandros Marios, Michael Dunne, Gert Storm, and Christine Allen. 2017. "The Battle of 'Nano' Paclitaxel." *Advanced Drug Delivery Reviews* 122: 20–30. https://doi.org/10.1016/j.addr.2017.02.003.

Sommer-kamann, Christina, Alexander Fries, Silja Mordhorst, Jennifer N Andexer, and M Michael. 2017. "Asymmetric C-Alkylation by the S -Adenosylmethionine-Dependent Methyltransferase SgvM Angewandte," 4033–36. https://doi.org/10.1002/anie.201609375.

Sonntag, Frank, Cora Kroner, Patrice Lubuta, Rémi Peyraud, Angelika Horst, Markus Buchhaupt, and Jens Schrader. 2015. "Engineering Methylobacterium Extorquens for de Novo Synthesis of the Sesquiterpenoid α-Humulene from Methanol." *Metabolic Engineering* 32: 82–94. https://doi.org/https://doi.org/10.1016/j.ymben.2015.09.004.

Sono, Masanori, Mark P Roach, Eric D Coulter, and John H Dawson. 1996. "Heme-Containing Oxygenases." *Chemical Reviews* 96 (7): 2841–88. https://doi.org/10.1021/cr9500500.

Stein, Viktor, Marta H Kubala, Jason Steen, Sean M Grimmond, and Kirill Alexandrov. 2015. "Towards the Systematic Mapping and Engineering of the Protein Prenylation Machinery in Saccharomyces Cerevisiae." *PLOS ONE* 10 (3): e0120716. https://doi.org/10.1371/journal.pone.0120716.

Stönner, Christof, and Jonathan Williams. 2016. "Goals Change Crowd Air Chemistry." *Nature* 535 (7612): 355. https://doi.org/10.1038/535355a.

Subramanian, Thangaiah, June E. Pais, Suxia Liu, Jerry M. Troutman, Yuta Suzuki, Karunai Leela Subramanian, Carol A. Fierke, Douglas A. Andres, and H. Peter Spielmann. 2012. "Farnesyl Diphosphate Analogues with Aryl Moieties Are Efficient Alternate Substrates for Protein Farnesyltransferase." *Biochemistry* 51 (41): 8307–19. https://doi.org/10.1021/bi3011362.

Summons, R E, L L Jahnke, J M Hope, and G A Logan. 1999. "2-Methylhopanoids as Biomarkers for Cyanobacterial Oxygenic Photosynthesis." *Nature* 400 (6744): 554–57. https://doi.org/10.1038/23005.

Takahashi, Seiji, and Tanetoshi Koyama. 2006. "Structure and Function of Cis-Prenyl Chain Elongating Enzymes." *Chemical Record (New York, N.Y.)* 6 (4): 194–205. https://doi.org/10.1002/tcr.20083.

Takahashi, Yoko, and Satoshi Omura. 2003. "Isolation of New Actinomycete Strains for the Screening of New Bioactive Compounds." *Journal of General and Applied Microbiology*. https://doi.org/10.2323/jgam.49.141.

Tholl, Dorothea. 2006. "Terpene Synthases and the Regulation, Diversity and Biological Roles of Terpene Metabolism." *Current Opinion in Plant Biology* 9 (3): 297–304. https://doi.org/10.1016/j.pbi.2006.03.014.

———. 2015. "Biosynthesis and Biological Functions of Terpenoids in Plants." In *Biotechnology of Isoprenoids*, edited by Jens Schrader and Jörg Bohlmann, 63–106. Cham: Springer International Publishing. https://doi.org/10.1007/10_2014_295.

Tilden, William A. 1884. "XLVI.—On the Decomposition of Terpenes by Heat." *J. Chem. Soc.{,} Trans.* 45 (0): 410–20. https://doi.org/10.1039/CT8844500410.

Vedula, L. Sangeetha, Yuxin Zhao, Robert M. Coates, Tanetoshi Koyama, David E. Cane, and David W. Christianson. 2007. "Exploring Biosynthetic Diversity with Trichodiene Synthase." *Archives of Biochemistry and Biophysics* 466 (2): 260–66. https://doi.org/10.1016/j.abb.2007.06.016.

Vickers, Claudia E., MAREIKE BONGERS, QING LIU, THIERRY DELATTE, and HARRO BOUWMEESTER. 2014. "Metabolic Engineering of Volatile Isoprenoids in Plants and Microbes." *Plant, Cell & Environment* 37 (8): 1753–75. https://doi.org/10.1111/pce.12316.

Wagner, Tobias, and Manfred Jung. 2012. "New Lysine Methyltransferase Drug Targets in Cancer." *Nature Biotechnology* 30 (7): 622–23. https://doi.org/10.1038/nbt.2300.

Wallach, Otto. 1887. "Zur Kenntniss Der Terpene Und Ätherischen Oele - 5. Abhandlung." *Justus Liebigs Ann. Chem.* 238: 78–89.

Wang, Chieh-Mei, and David E. Cane. 2010. "Biochemistry and Molecular Genetics of the Biosynthesis of the Earthy Odorant Methylisoborneol in *Streptomyces Coelicolor*." *Journal of the American Chemical Society* 132 (27): 9509–9509. https://doi.org/10.1021/ja104306p.

Wang, Chieh M., and David E. Cane. 2008. "Biochemistry and Molecular Genetics of the Biosynthesis of the Earthy Odorant Methylisoborneol in Streptomyces Coelicolor." *Journal of the American Chemical Society* 130 (28): 8908–9. https://doi.org/10.1021/ja803639g.

Wang, Jianhua, Chunshan Quan, Xue Wang, Pengchao Zhao, and Shengdi Fan. 2011. "Extraction, Purification and Identification of Bacterial Signal Molecules Based on N-Acyl Homoserine Lactones." *Microbial Biotechnology* 4 (4): 479–90. https://doi.org/10.1111/j.1751-7915.2010.00197.x.

Wang, Kevin C, and Shin-ichi Ohnuma. 2000. "Isoprenyl Diphosphate Synthases." *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1529 (1): 33–48. https://doi.org/https://doi.org/10.1016/S1388-1981(00)00136-0.

Wang, Qin, Shu Quan, and Han Xiao. 2019. "Towards Efficient Terpenoid Biosynthesis: Manipulating IPP and DMAPP Supply." *Bioresources and Bioprocessing* 6 (1). https://doi.org/10.1186/s40643-019-0242-z.

Wezel, Gilles P van, and Kenneth J McDowall. 2011. "The Regulation of the Secondary Metabolism of Streptomyces: New Links and Experimental Advances." *Nat. Prod. Rep.* 28 (7): 1311–33. https://doi.org/10.1039/C1NP00003A.

Wu, Wenzhao, and Christos T Maravelias. 2018. "Synthesis and Techno-Economic Assessment of Microbial-Based Processes for Terpenes Production." *Biotechnology for Biofuels* 11 (October): 294. https://doi.org/10.1186/s13068-018-1285-7.

Yamada, Yuuki, Tomohisa Kuzuyama, Mamoru Komatsu, Kazuo Shin-Ya, Satoshi Omura, David E Cane, and Haruo Ikeda. 2015. "Terpene Synthases Are Widely Distributed in Bacteria." *Proceedings of the National Academy of Sciences of the United States of America* 112 (3): 857–62. https://doi.org/10.1073/pnas.1422108112.

Yang, Xia, Komi Nambou, Liujing Wei, and Qiang Hua. 2016. "Heterologous Production of α-Farnesene in Metabolically Engineered Strains of Yarrowia Lipolytica." *Bioresource Technology* 216: 1040–48. https://doi.org/10.1016/j.biortech.2016.06.028.

Yang, Xiaoman, Yueming Jiang, Jiali Yang, Jirui He, Jian Sun, Feng Chen, Mingwei Zhang, and Bao Yang. 2015. "Prenylated Flavonoids, Promising Nutraceuticals with Impressive Biological Activities." *Trends in Food Science & Technology* 44 (1): 93–104. https://doi.org/https://doi.org/10.1016/j.tifs.2015.03.007.

Yeagle, P L. 1991. "Modulation of Membrane Function by Cholesterol." *Biochimie* 73 (10): 1303–10. https://doi.org/https://doi.org/10.1016/0300-9084(91)90093-G.

Zhang, Congqiang. 2018. "Biosynthesis of Carotenoids and Apocarotenoids by Microorganisms and Their Industrial Potential." In *Progress in Carotenoid Research*, Ch. 5. Rijeka: IntechOpen. https://doi.org/10.5772/intechopen.79061.

Zheng, Yanning, Qiang Liu, Lingling Li, Wen Qin, Jianming Yang, Haibo Zhang, Xinglin Jiang, et al. 2013. "Metabolic Engineering of Escherichia Coli for High-Specificity Production of Isoprenol and Prenol as next Generation of Biofuels." *Biotechnology for Biofuels* 6 (1): 57. https://doi.org/10.1186/1754-6834-6-57.

Zurbriggen, Andreas, Henning Kirst, and Anastasios Melis. 2012. "Isoprene Production Via the Mevalonic Acid Pathway in Escherichia Coli (Bacteria)." *Bioenergy Research* 5 (4): 814–28. https://doi.org/10.1007/s12155-012-9192-4.

7 Appendix

SEQ01 methyltransferase [Streptomyces coelicolor A3(2)] (GPPMT)

NCBI Reference Sequence: NP_631739.1

>NP_631739.1 methyltransferase [Streptomyces coelicolor A3(2)]
MTTETTTATATAKIPAPATPYQEDIARYWNNEARPVNLRLGDVDGLYHHHYGIGPVDRAALGDPEHS
EYEKKVIAELHRLESAQAEFLMDHLGQAGPDDTLVDAGCGRGGSMVMAHRRFGSRVEGVTLSAAQ
ADFGNRRARELRIDDHVRSRVCNMLDTPFDKGAVTASWNNESTMYVDLHDLFSEHSRFLKVGGRYV
TITGCWNPRYGQPSKWVSQINAHFECNIHSRREYLRAMADNRLVPHTIVDLTPDTLPYWELRATSSL
VTGIEKAFIESYRDGSFQYVLIAADRV

SEQ02 putative methyl transferase [Streptomyces argenteolus] (Lon23)

GenBank: BAF98640.1

>BAF98640.1 putative methyl transferase [Streptomyces argenteolus]
MSLETVRTNEIIRDDFEKDLSTYWETKQNDQINLLLGEEDGLYHHHFGIGDFDRSVADLPPEERESRV
LEEMHSLENTQVETLIGALGDVPRDARLLDMGSGRGGTSFMIYDRFGCTIDGVTFAQYQVDFSNRLA
ETRGCADRVRFHYRNMVKTGFPDGAFQYVVTNETTPYVKLDEVFSELSRVLAPGGRYVSLTWCRND
AVASQCDEVLEIDRHYICRTHRRSSYFKQMAAHGLVPRTVVDFTTEAIPYFEVRLLSKLATGSEQPYL
SGYGSDRINYLLIVAERV

SEQ03 class I SAM-dependent methyltransferase [Rhodococcus fascians]

NCBI Reference Sequence: WP_037174547.1

>WP_037174547.1 class I SAM-dependent methyltransferase [Rhodococcus fascians]
MISADQYARDDYERELKAHWDAKTTDDINLLLGADDDLYHHHYAIGDFDRSILDSVGEDRENAILREL
HRMENDQVGLILDALGPLPPNSRGMDAGSGRGGTSFRLAGATESRIDGVNFCEHHVAFAEQIARKR
GWDSRVQFHLGNMLQAPFPDRTFDFVVSNETTMYADAYEAMAEFSRLLRRGGRYVMTTWCRNDA
VDPRSDATRQIDEHYVCRMHRRSTYFEAFAANGLIPYHVAQYTHEAMPYWELRNNSKLRTGVEDAF
LSGYSDGSLNYLVIAAERI

SEQ04 methyltransferase domain-containing protein [Rhodococcus fascians]

NCBI Reference Sequence: WP_037174549.1

>WP_037174549.1 methyltransferase domain-containing protein [Rhodococcus fascians]
MPNLDVADLGQHDVQQRRYWDAKKSDDINLLLGTEDGLYHHHYGIGDYDHSVLAAPAELRESLILRE
LHRMESLEINLIVDALGEVSPSSRVMDAGSGRGGTTFTIADRFGCRVDGVNYCAHHVEFAEKLARER
GSSDRVQFHFANMVQTPFEDNTFDYIVSNETTMCVDLGEAFTEFARLLRPGGRYVAVTWCRNDVVA
ERSEASRLIDEEYLCAMHTRSTYFQTLAANGLVPYHVQRYTDEAIPYWDLRNQAALRTGVEEPFLQG
FRERSIDYLVIACERL

SEQ05 class I SAM-dependent methyltransferase [Streptomyces monomycini]

NCBI Reference Sequence: WP_033037353.1

>WP_033037353.1 class I SAM-dependent methyltransferase [Streptomyces monomycini]
MSSEPTAALRTPLEQAAFDYYDNKRDDPINLKLGETDGFYHHHFAVGDFDRKVLLTSGEERAQAINA
ELHRMETRQVDALSEALGPVAPQARILDGGSGRGGTAILLHKAFGCRVDGVNISAYQNEFARRQAQE
HNCAEKVHFHDRNMAATGFPDASFDYVVTNETTMYVDPYETFAEFARLLKPAGRYVLLTWCNNDVL
APHPPEATAIDAHYHCHTHRRTTYLKALIESGLIPYQVDDLTQPATPYWELRSYSRYATGIEQHYLDG
YRSDRVNYIRIASRRGAVPPSDTPTTAGAAAH

SEQ07 polyprenyl synthetase family protein [Streptomyces monomycini] (PT)

NCBI Reference Sequence: WP_030019072.1

>WP_030019072.1 polyprenyl synthetase family protein [Streptomyces monomycini]
MTTPDAAAPPPLPDVILHAREFARRGLEEALARLHPDLARICGYYFGWNDAEGTATNRRGSRTLQAA
MVMLAARATGQDEAAAAPGAAAVELTHNFSLLHDDIADGDEIRRDRPTAWVAFGTGPTLVAGDALFN
EAVRLLAATQAATTAVRVFTDGVAHMIHAWASEPAFDRTDPLDISLDAYLECCRGKGGALLGTGAVL
GTVLCGKPAQDGDLLRRAAHHAGTAWQAVNDLENIWGNAALVGKPGFQDLRLRKHTLPVITAMQSQ
HPHTRHLRQLLAQPAPDDNALQTTADLLEELGGKAATERVAQHHLTQALATLDEASLPKAVHEDLAE
LLQFTVTRRPSPHGRRA

SEQ08 hypothetical protein [Streptomyces monomycini] (TS)

NCBI Reference Sequence: WP_030019073.1

>WP_030019073.1 hypothetical protein [Streptomyces monomycini]
MPRNIPLAMPLPHQPLPSGLEQVREAHVEWLHRHGLLPTEQSTERYLRSAVADIAAYGDPNSEDLLL
SFHVCGWLFLYDDFMDAPTGPAPGHAETVTAELTSMLYRASTPSTPLTSAFADLWRRLCAGMSESW
RLRMAWIWQEVYAGMLAETVNRRLDVTLSYHEHLRVRDMSIGTTLVAAVAERTGGYEVPVSIWPAA
YLSTLRHHVTRHMLLTNDIFSLEKESARADANLVSLAMRERHQTREQALQALKEAADRHIQQLHTRSA
AVNDFCDRLHLPQSARQAVTRHVEALRAWARGAYDWQCTTARYGPEEAACSAPDQPGYLSPSQP

SEQ09 carotenoid oxygenase family protein [Streptomyces monomycini]

NCBI Reference Sequence: WP_030022897.1

>WP_030022897.1 carotenoid oxygenase family protein [Streptomyces monomycini]
MTTTNALGSAAPVSGSGRAPAPHLAGNFAPVTKELTAYELPVTGAIPPELVGWYLRNGPNPRDAASA
HWFFGDGMVHGVRLEGGKAVSYRNRWVRTSTLTDGAQVRGRDGRPDLTAGVANTHVVHHAGRTL
ALVESSFPYELDCSPGHELETVGPYDFDGRLTTAMTAHPKTCPTTGELHFFGYGAPEPPYLTYHRAD
ASGALELSRPVEVAASTMMHDFHLTAGHVVFMDLPLVYDRGTPGMPYVWDPAYGARLGVLRRDDP
HGDVRWFAIDPCYVFHSLNAYDDGSGRIVLYVCRYPALELKSLPTLWRWTIDLASGKVVEEQVDDQS
GEFPRMDDRLAGLDARFGHITRSGGPGAGPTPSALIRYDLHTGTSTRYDFDSGRTPAEAAFAPADDR
PGGPGWLLTYVYDAARDSSDLGILDAEDIAKGPVATVHLPQRVPYGFHGNWIPDPAQ

Plasmid Maps

pMK-17
6278 bp

pMK-18
8393 bp

178
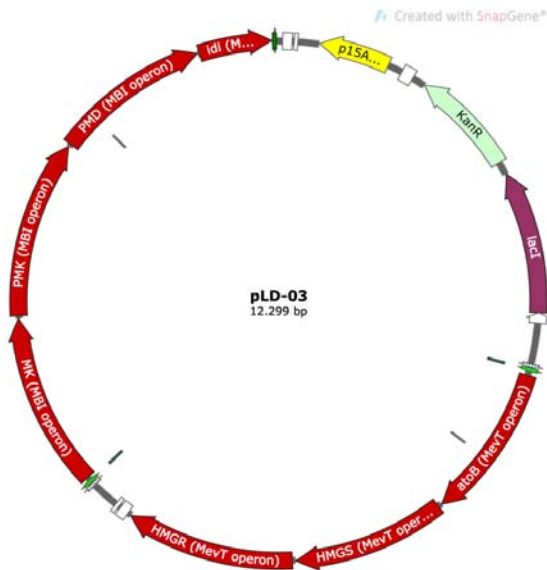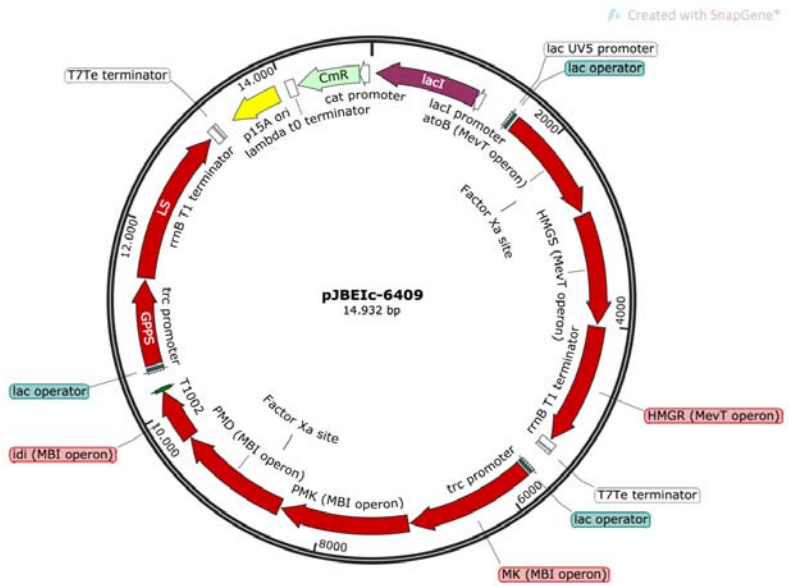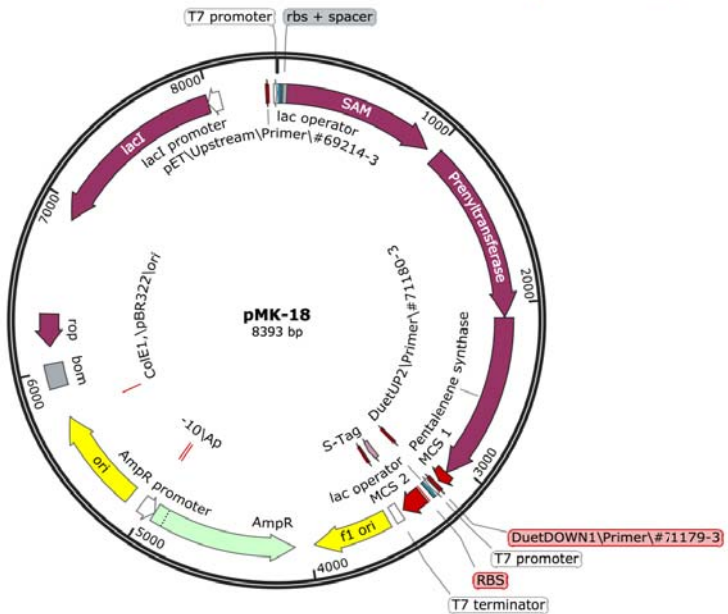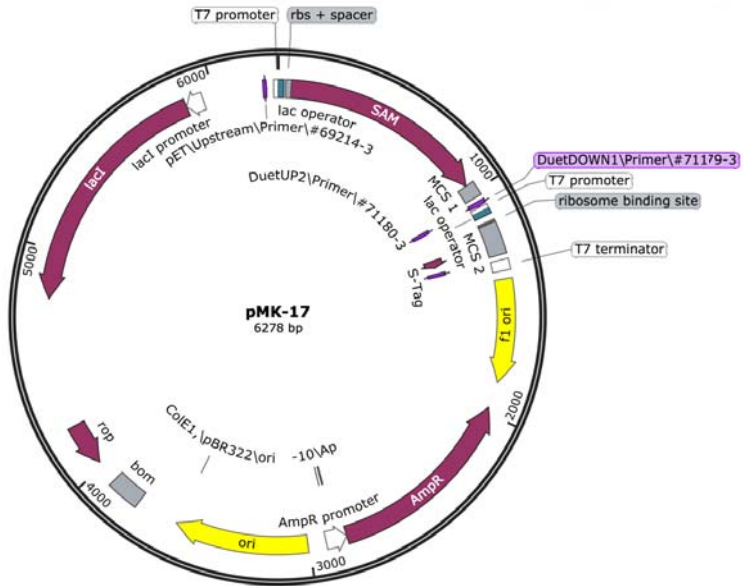
<center>**Curriculum Vitae**</center>

Laura Silveira Drummond Moreira

Born 19/05/1984 in Curitiba, Brazil

Name in Scientific Publications: Drummond, L.; Moreira, L. S. D.

EDUCATION	07.2015 – 12.2020 **PhD in Biological Sciences, Emphasis area: Biotechnology. Goethe University Frankfurt** Investigation of non-canonical terpene biosynthesis
	07.2014 – 09.2014 **Specialization in Biological Laboratory Techniques and Laboratory Safety. University of Campinas**
	03.2008 – 10.2010 **MSc in Entomology, Emphasis area: Chemical Ecology. University of Sao Paulo** Gene expression and volatile compounds induced by herbivory of *Spodoptera frugiperda* (J. E. Smith) (Lepidoptera: Noctuidae) in corn, *Zea mays* L. (Poaceae)
	03.2003 – 10.2007 **BSc in Biological Sciences, Emphasis area: Molecular Biology. University of Sao Paulo** Analysis of conformational structures composed by homopolymeric DNA and RNA
EMPLOYMENT	01.2020 – 12.2021 Scientific Staff, Hochschule Geisenheim University, Geisenheim Germany
	08.2019 - 01.2020 Scientific Research Assistant, Chemical Technology, DECHEMA Research Institute, Frankfurt Germany
	07.2015 – 07.2019 Scientific Researcher, Industrial Biotechnology, DECHEMA Research Institute, Frankfurt Germany
	01.2015 – 05.2015 English Teacher, CNA, Piracicaba Brazil
	11.2012 – 01.2014 Reviewer, 3D Education, Sao Paulo Brazil

11.2011 – 04.2012
English Teacher, Red Balloon, Sao Paulo Brazil

10.2010 – 10.2011
Copyeditor, Meca Publishing House, Sao Paulo Brazil

03.2008 – 11.2010
Master Thesis Student, University of Sao Paulo, Piracicaba Brazil

06.2006 – 10.2007
Scientific Initiation Student, University of Sao Paulo, Sao Paulo Brazil

SCHOLARSHIPS and GRANTS

Research Assistantship – GRADE Academy, Goethe University Frankfurt. 2019
Doctoral Scholarship – CNPq (www.cnpq.br) 2015-2019
Master's Scholarship – FAPESP (www.fapesp.br) 2008-2010
Developing Country Grant – ISCE Annual Meeting (http://www.chemecol.org/), Neuchâtel, Switzerland. 2009
Scientific Initiation Scholarship – FAPESP (www.fapesp.br) 2006-2007

PUBLICATIONS

Drummond, L.; Kschowak, M.; Breitenbach, J.; Wolff, H.; Shi, Y.; Schrader, J.; Bode, H.; Sandmann, G.; Buchhaupt, M. Expanding the isoprenoid building block repertoire with an IPP methyltransferase from *Streptomyces monomycini*. ACS Synthetic Biology, V. 8, Issue 6, pp 1303-1313, 2019

Schempp, F. M.; Drummond, L.; Buchhaupt, M.; Schrader, J.. Microbial cell factories for the production of terpenoid flavor and fragrance compounds. Journal of Agricultural and Food Chemistry, V. 66, Issue 10, pp 2247-2258, 2018

Signoretti, A.G.C.; Peñaflor, M.F.G.V.; Moreira L.S.D.; Noronha, N.C.; Bento J.M.S.. Diurnal and nocturnal herbivore induction on maize elicit different innate response of the fall armyworm parasitoid, *Campoletis flavicincta*. Journal of Pest Science, V. 85, Issue 1, pp 101-107, 2012

Gorab, E.; Amabis, J. M.; Stocker, A.J.; Drummond, L.; Stollar, B.D.. Potential sites of triple-helical nucleic acid formation in chromosomes of *Rhynchosciara* (Diptera: Sciaridae) and *Drosophila melanogaster*. Chromosome Research, V. 17, pp 821-832, 2009

PATENT APPLICATION

Markus Buchhaupt, Max Kschowak, Laura Drummond, Jens Schrader. EP3425051 - METHOD FOR THE PRODUCTION OF ISOPRENOIDS BY HETEROLOGOUS EXPRESSION OF A PROTEIN IN A MICROORGANISM. Application Number 17179874, European Patent Office, 2017.

INTERNATIONAL CONFERENCES

Bioflavour – Biotechnology of Flavours, Fragrances and Functional Ingredients. Frankfurt (Germany), 2018 - Lecture

4th Applied Synthetic Biology in Europe. Toulouse (France), 2018 - Lecture

Crossing Kingdoms: An International Synthetic Biology Symposium. Cambridge (England), 2018 – Poster / Lecture

International Society for Chemical Ecology Annual Meeting Neuchâtel (Switzerland), 2009 – Poster

SUPERVISION OF STUDENTS

10.2018 – 12.2018
Clara Börnsen, Goethe University Frankfurt – Specialization

03.2017 – 09.2017
Rupa Bhattarai, Technical University Dresden – Master Thesis