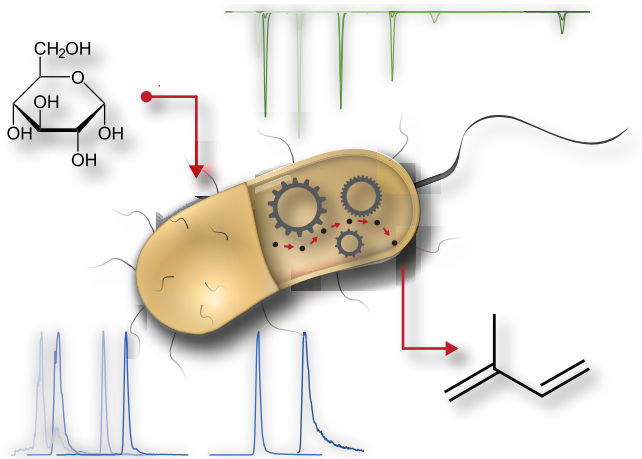
APPLIED MICROBIOLOGY

Volume 9

Daniel Volke

Rational engineering of the methylerythritol 4-phosphate (MEP) pathway for terpenoid production through metabolic control analysis



iAMB Institute of Applied Microbiology | ABBt Aachen Biology and Biotechnology | RWTH AACHEN UNIVERSITY

Edition Wissenschaft Apprimus

“Rational engineering of the methylerythritol 4-phosphate (MEP) pathway for terpenoid production through metabolic control analysis”

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Master of Science Daniel Volke

aus Darmstadt

Berichter*:* Univ. Prof. Dr. Rainer Fischer

Univ. Prof. Dr.-Ing. Lars M. Blank

Tag der mündlichen Prüfung*:* 24.11.2017

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über https://portal.dnb.de abrufbar.

Daniel Volke:

Rational engineering of the methylerythritol 4-phosphate (MEP) pathway for terpenoid production through metabolic control analysis

1. Auflage, 2018

Gedruckt auf holz- und säurefreiem Papier, 100% chlorfrei gebleicht.

Apprimus Verlag, Aachen, 2018
Wissenschaftsverlag des Instituts für Industriekommunikation und Fachmedien
an der RWTH Aachen
Steinbachstr. 25, 52074 Aachen
Internet: www.apprimus-verlag.de, E-Mail: info@apprimus-verlag.de

ISBN 978-3-86359-650-7

D 82 (Diss. RWTH Aachen University, 2017)

Table of Contents

Summary

Due to an increasing demand for sustainable chemicals, new strategies have to be explored to produce these resources at a low-cost and with low environmental impact. Microbially produced terpenoids are a promising source for a broad spectrum of valuable compounds and industrial precursors. The 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway produces the precursor for terpenoid biosynthesis, but previous metabolic engineering attempts focusing on applying the MEP pathway in the production of terpenoids have met with limited success.

Heterologous expression of an isoprene synthase (*ispS*) gene in *Escherichia coli* enabled the strain to produce isoprene. Metabolic characterization revealed the accumulation of isopentenyl pyrophosphate (IPP) in this strain, which led to growth impairment. Through coexpression of the native isopentenyl pyrophosphate isomerase (*idi*), accumulation of IPP was neglected and normal growth was restored, concomitantly with higher isoprene production. To further increase isoprene production, the module expressing *ispS* and *idi* was combined with a second module encoding different combinations of native MEP pathway genes. A module encoding *dxs*, *ispD*, *ispG* and *ispH* showed the highest increase in isoprene production. All engineered *E. coli* strains displayed elevated levels of 2-C-methyl-D-erythritol 2,4-cyclopyrophosphate (MEcPP). IspG, an enzyme functioning with an iron-sulfur cluster, metabolizes MEcPP. As overexpression of ispG did not to reduce the MEcPP concentration, coexpression of proteins responsible for the maturation and reduction of the iron-sulfur cluster in IspG were coexpressed, but did not lead to lower MEcPP concentration or higher isoprene production. Additional expression of further MEP pathway genes (*dxr*, *IspE*, *IspF*) did not lead to improved isoprene titers.

Subsequently, metabolic control analysis was applied to investigate the influence of enzyme activity on intermediate concentrations and fluxes in the MEP pathway. For this purpose, libraries of *E. coli* mutants with altered expression of the MEP pathway genes were generated by randomization of their ribosome binding sites (RBS) through recombineering. The effect of the altered RBS on the enzyme concentration was evaluated through quantitative targeted proteomics. After transformation of the *dxs* expression library with the isoprene production module, containing *ispS* and *idi*, the intermediate concentrations and fluxes in the MEP pathway revealed a strong flux control coefficient of Dxs towards isoprene (0.41) at wild type expression level. Dxs showed less control at higher expression levels. The MEcPP concentration changed linearly with Dxs concentration at lower expression levels of *dxs*, whereas a sharp increase in MEcPP concentration was observed at higher expression levels. This behavior was not observed for any other intermediate. The high concentration caused increased efflux of MEcPP. It is concluded that IspG is operating close to its maximum reaction rate at higher *dxs* expression levels and therefore is limiting the flux through the MEP pathway.

Zusammenfassung

Im Zuge erhöhter Nachfrage nach umweltfreundlichen und nachhaltigen Chemikalien, müssen neue Prozesse zu deren Herstellung etabliert werden. Mikrobiell produzierte Terpenoide sind eine vielversprechende Quelle für eine Vielzahl an wertvollen und ökonomisch relevanten Verbindungen. Die Vorläufermoleküle des Terpenoidbiosynthesewegs werden in vielen Organismen durch den 2C-Methyl-D-erythritol-4-phosphat (MEP)-Weg bereitgestellt. Bisherige *metabolic engineering* Versuche zur Produktion hoher Mengen an Terpenoiden mit Hilfe des MEP Biosynthesewegs hatten wenig Erfolg.

Die Expression einer Isoprensynthase in *Escherichia coli* ermöglichte die mikrobielle Produktion von Isopren. Eine metabolische Charakterisierung des Stammes zeigte die Akkumulation von Isopentenyl-pyrophosphat. Durch die Koexpression der Isopentenylpyrophosphat-Isomerase (*idi*) konnte die Akkumulation verhindert werden, welches zu einer höheren Isoprenproduktion und gesteigerter Wachstumsrate führte. Anschließend wurden weitere Gene des MEP-Weges zusätzlich überexpremiert. Die höchste Steigerung wurde mit Hilfe eines Clusters, welches zusätzlich die Gene *dxs*, *ispD*, *ispG* und *ispH* beinhaltet, erzielt. Alle konstruierten Stämme wiesen stark erhöhte 2-C-Methyl-D-erythritol-2,4-cyclopyrophosphat (MEcPP)-Konzentrationen auf. Eine Überexpression des Gens *ispG*, welches MEcPP metabolisiert, hatte keinen Einfluss auf die Konzentration. Daher wurde die Koexpression mehrerer Proteine, welche mit der Aktivität des Eisen-Schwefel-Clusters in IspG in Zusammenhang stehen, zur Steigerung der katalytischen Aktivität von IspG getestet. Doch eine Steigerung der Isoprenproduktion oder Reduktion der MEcPP-Konzentration wurde nicht erreicht. Auch die Überexpression weiterer Gene des MEP-Weges führte zu keiner Steigerung der Isoprenproduktion.

Darauffolgend wurde der MEP-Weg in *E. coli* mit Hilfe von *metabolic control analysis* charakterisiert. Es wurden Expressionsbibliotheken der Gene des MEP-Wegs durch Randomisierung ihrer Ribosomen-bindestellen mittels *recombineering* erstellt. Diese wurden hinsichtlich der Enzym- und Intermediat-konzentrationen und des metabolischen Flusses untersucht. Hierfür wurde *targeted proteomics* und *metabolomics* genutzt. Dies zeigte eine starke Kontrolle durch Dxs mit einem Fluss-Kontroll-Koeffizienten von 0,41 auf den Fluss zu Isopren in Stämmen, welche sowohl die Isoprensynthase als auch *idi* stark exprimieren. Diese Kontrolle war jedoch bei höheren *dxs* Expressionen reduziert. Während die Konzentration an MEcPP in *dxs* unterexpremierenden Stämmen linear mir der Konzentration an Dxs zunahm, zeigte sich eine rapide Zunahme der MEcPP-Konzentration, bei erhöhter *dxs* Expression. Daraus ließ sich schlussfolgern, dass IspG bei erhöhter Dxs Aktivität nahe seiner maximalen Reaktionsgeschwindigkeit arbeitet, daher sind höhere Flüsse nur mit Steigerung der katalytischen Aktivität von IspG zu erreichen. Die Saturierung von IspG durch sein Substrat führte außerdem zu einem hohen Export an MEcPP.

Abbreviations

%(w/w)	weight per weight percentage
A_{600}	Absorption at 600 nm wavelength
ACN	Acetonitrile
ADP	Adenosine diphosphate
APCI	Atmospheric pressure chemical ionization
ATP	Adenosine triphosphate
AU	Arbitrary Units
BCC	β-cyclocitral
CCCP	Carbonyl cyanide 3-chlorophenylhydrazone
CMP	Cytidine monophosphate
CTP	Cytidine triphosphate
Cys	Cysteine
Da	Dalton (atomic mass unit)
DMAPP	Dimethylallyl pyrophosphate
DX	1-Deoxy-D-xylulose
DXP	1-Deoxy-D-xylulose 5-phosphate
Dxr	1-Deoxy-D-xylulose 5-phosphate reductoisomerase
Dxs	1-Deoxy-D-xylulose 5-phosphate synthase
EtOH	Ethanol
$FldA_{ox}$	Flavodoxin oxidized
$FldA_{red}$	Flavodoxin reduced
Fsr	Fosmidomycin resistance protein
GAP	Glyceraldehyde 3-phosphate
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GC-MS/MS	Gas chromatography coupled tandem mass spectrometry
HMBPP	(E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate
Idi	Isopentenyl pyrophosphate (delta) isomerase
IPP	Isopentenyl pyrophosphate
IPTG	Isopropyl-β-D-thiogalactopyranosid
IspD	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
IspE	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase
IspF	4-Diphosphocytidyl-2-C-methyl-D-erythritol kinase
IspG	HMBPP synthase
IspH	HMBPP reductase

IspS	Isoprene synthase
KDPG	2-Keto-3-desoxy-phosphogluconat
LC-MS/MS	Liquid chromatography coupled tandem mass spectrometry
m/z	Mass-to-charge ratio
MAHMPPP	4-amino-5-hydroxymethyl-2-methylpyrimidine-pyrophosphate
MCA	Metabolic control analysis
ME-CDP	4-diphosphocytidyl-2-C-methylerythritol
MEcPP	2-C-methyl-D-erythritol 2,4-cyclopyrophosphate
MeOH	Methanol
MEP	2-C-methylerythritol 4-phosphate
MEP-CDP	4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate
MFS	Major facilitator superfamily
MHThiP	5-methy-4-(β-hydroxyethyl) thiazole phosphate
MRM	Multiple reaction monitoring
NAD(P)$^+$	Nicotinamide adenine dinucleotide (phosphate) oxidized
NAD(P)H/H$^+$	Nicotinamide adenine dinucleotide (phosphate) reduced
OD_{600}	Optical density at 600 nm
PCR	Polymerase chain reaction
PGI	Phosphoglucose isomerase
PLP	Pyridoxal 5-phosphate
PMP	Pyridoxamine 5-phosphate
PN	Pyridoxine
PNP	Pyridoxine 5´-phosphate
PNT	Membrane bound NAD(P)+ transhydrogenase
PPIase	FKBP-type peptidyl-prolyl cis-trans isomerase/rotamase
Psig	Pound-force per square inch gauge
Pyr	Pyruvate
Q1	First Quadrupole mass analyzer
Q3	Second Quadrupole mass analyzer
RBS	Ribosomal binding site
RPM	Revolution per minute
RuBisCO	Ribulose 1,5-bisphosphate carboxylase/oxygenase
SD	Standard deviation
SLIC	Sequence- and ligation-independent cloning
T_M	Melting point

TP	Thiamine phosphate
TPP	Thiamine pyrophosphate
Tyr	Tyrosine
U-^{13}C-Glucose	Uniformed ^{13}C labeled glucose
UdhA	Soluble NAD(P)+ transhydrogenase
V	Volt
v/v	Volume fraction

I. Introduction

I.1 Terpenoids

Terpenoids are synthesized by all living cells (Matsumi *et al.*, 2011) and play essential roles in several biological functions. They are crucial in photosynthesis, as hormones and as protective agents against oxidative stress and predators (Wink, 2004). Furthermore, they can be found in the cell membrane as integrative components and as electron transporters (Sacchettini and Poulter, 1997). These are just some of the many tasks these ubiquitous molecules fulfill. Through their structural diversity, terpenoids constitute one of the biggest groups of natural products comprising more than 60,000 known compounds (Koksal *et al.*, 2011). Many colors in nature originate from terpenoids, like the red of the tomato, the orange of autumn leaves and the yellow of egg yolk. They are not only pleasant to the eye, but also enrich our world with flavors and fragrances, like in hops and lavender.

Despite their chemical diversity, all terpenoids (also known as isoprenoids (Gershenzon and Dudareva, 2007)) originate from only two common precursors, namely isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) (Figure 1) (McGarvey and Croteau, 1995). This also leads to the so called “isoprene rule” (also known as C5 rule) (Kuzuyama and Seto, 2003), which means that the isoprene structural motif is the repeated unit. The isoprene structure results from the elimination of pyrophosphate from each of the precursors. Although terpenoids are built from DMAPP and IPP, which contain five carbon atoms (C5), the nomenclature is historically based on C10 units, owing to the fact that C10 was believed to be the smallest unit (Ashour *et al.*, 2010). Therefore, terpenoids with a C10 backbone are called monoterpenoids, with C20 backbone diterpenoids and so on. The later found terpenoids with an uneven number of isoprene numbers are called hemiterpenoids (from greek half) if they contain one isoprene unit, sesquiterpenoids (from greek one and a half) if they contain three isoprene units.

The simplest terpenoids, the hemiterpenoids, are synthesized directly through modifications from IPP or DMAPP (Figure 1). Monoterpenoids arise from geranyl pyrophosphate, which is formed by the condensation of IPP and DMAPP. Di-, sesqui-, tri-, tetra-, and polyterpenoids are generated through the consecutive condensation of shorter diphosphate esters (Figure 1). The phosphorylated intermediates DMAPP, IPP, geranyl pyrophosphate, farnesyl pyrophosphate and geranylgeranyl pyrophosphate are common precursors in the terpenoid biosynthesis and lie at multiple branching points (Sacchettini and Poulter, 1997) (Figure 1). These key intermediates are also the prenyl donors for the prenylation of various aliphatic and aromatic acceptors of both high and low molecular weight substances including proteins and nucleic acids (Winkelblech *et al.*, 2015). The prenylation increases

the hydrophobicity of these compounds and in general allows them to reside in the membrane, like ubiquinone.

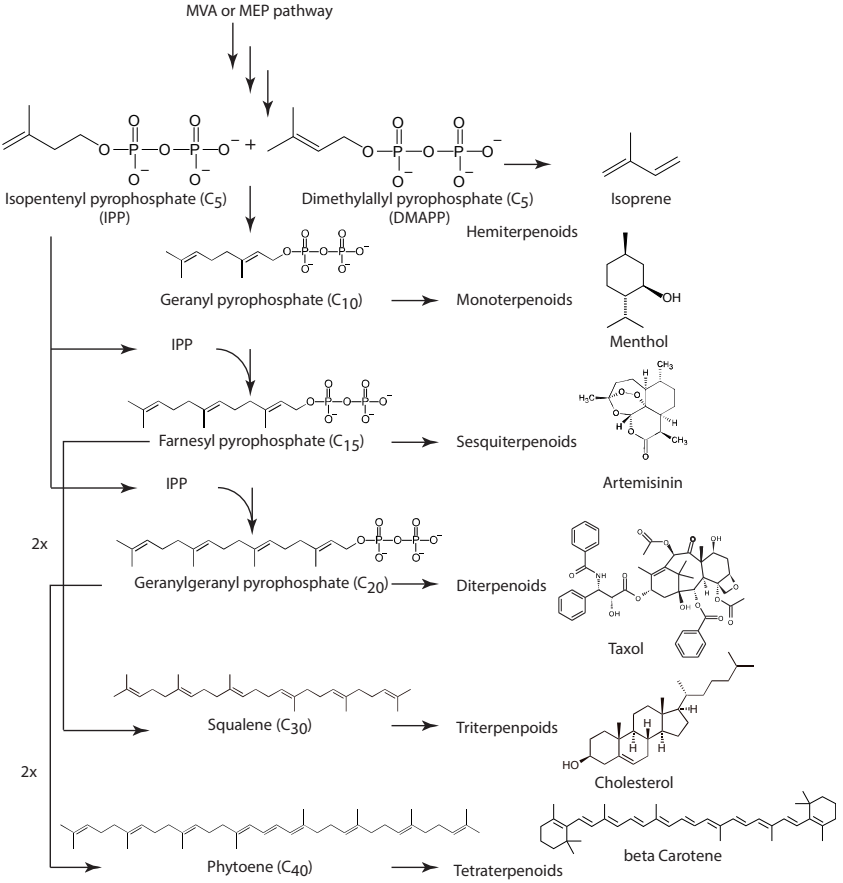


Figure 1| Biosynthesis of terpenoids from the universal precursors IPP and DMAPP. Through the successive condensation of IPP and DMAPP the alkyl chain is extended. The central intermediates are geranyl pyrophosphate, farnesyl pyrophosphate and geranylgeranyl pyrophosphate. These can be modified to get to the monoterpenoids, sesquiterpenoids and diterpenoids. Two molecules of farnesyl pyrophosphate can be condensed to squalene, which is the precursor for the triterpenoids. Two molecules of farnesyl pyrophosphate lead to phytoene, which opens the branch of the tetraterpenoids. Examples of each terpenoid family are shown on the right. Graphic is adapted from Keeling and Jörg (2012) and Suire *et al.* (2000).

After the elongation of the hydrocarbon chain, the terpenes can be subjected to diverse modifications such as cyclisation, oxidation, glycosylation, desaturation and many more (Ashour *et al.*, 2010). These layers of diversification are responsible for the chemical heterogeneity and structural complexity of the terpenoid group and allow terpenoids to play a part in different facets of life. For example, the

shorter terpenoids (hemi-, mono-, and sesquiterpenoids) are often volatile and some are synthesized by plants to communicate with other organisms, either to attract pollinators or in response to herbivory attack (Bouwmeester *et al.*, 1999; Clavijo McCormick *et al.*, 2012). The longer terpenoids (di-, tri- and tetraterpenoids) occur in hydrophilic and hydrophobic forms and exhibit extraordinarily complex structures, containing one, two or even three ring structures. Their functions are highly diverse.

One important class of diterpenoids are the gibberellins (Wink, 2004), which act as growth hormones in plants and regulate a multitude of processes including germination and flowering. Regarding triterpenes, the major class are steroids, which can be found as part of the membrane or as signaling molecules. Many steroids and triterpenes are also found in a glycosylated form. The glycosylated compounds are called saponins and are amphiphilic in contrast to the lipophilic non-glycosylated triterpenoids and steroids. This amphiphilic nature can cause perturbation of cell membranes and is responsible for the antimicrobial traits of saponins (Augustin *et al.*, 2011). The ability to produce saponins is widespread in the plant kingdom. Regarding tetraterpenoids, the most important class are carotenoids. These are organic pigments and serve functions for example in photosynthesis to absorb light or as single oxygen quenchers to protect the cells from damage (Keeling and Jörg, 2012).

Further elongation of the terpenoid backbone leads to polyterpenoids. One member of this class is caoutchouc, or natural rubber, which is made from the polymerization of IPP and can reach a mass up to 10^6 Da (Schulze Gronover *et al.*, 2011).

I.2 MVA and MEP pathway- Biosynthesis of DMAPP and IPP through two analogous pathways

Surprisingly, the synthesis of DMAPP and IPP does not occur through one unique pathway, even though all known organisms synthesise terpenoids from the same common precursors IPP and DMAPP. There are two analogous pathways: the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway and the mevalonate (MVA) pathway (Figure 2 and Figure 3).

The MVA pathway was discovered in the 1950s and was the target of intensive research since it is responsible for the biosynthesis of cholesterol in humans. The function and regulation of the pathway had been elucidated in depth, leading to the development of enzyme inhibitors such as statins. Statins inhibit one of the enzymes of the MVA pathway, the HMG-CoA reductase (Figure 2), and consequently act as cholesterol regulators. Michael S. Brown and Joseph L. Goldstein were awarded with the nobel

prize for Physiology or Medicine *"for their discoveries concerning the regulation of cholesterol metabolism"*.

The MVA pathway starts with the formation of the thioester 3-hydroxy-3-methyl-glutaryl-coenzyme A (HMG-CoA) from three molecules of Acetyl-CoA. The reaction is catalyzed by the enzyme *acetoacetyl-CoA thiolase* followed by *HMG-CoA synthase*. The thioester is then hydrolyzed and reduced to mevalonate. After consecutive phosphorylation to mevalonate 5-pyrophosphate, the last step consists of the decarboxylation and dehydration to yield IPP, which can be isomerized to DMAPP.

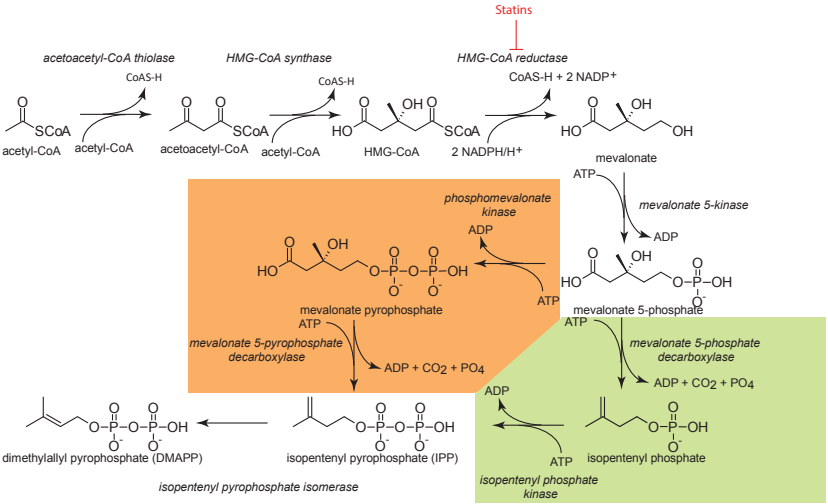


Figure 2| The Mevalonate (MVA) pathway It starts with the transfer of one acetyl unit on acetyl-CoA to yield acetoacetyl-CoA. Mevalonate, the first committed precursor, is formed after the addition of one more acetyl group and a reduction step. Two distinct MVA pathways are fully elucidated. Most organisms synthesize IPP through the decarboxylation of mevalonate pyrophosphate. While archaea use the mevalonate 5-phosphate decarboxylase and isopentenyl phosphate kinase to synthesize IPP from mevalonate 5-phosphate (Vannice *et al.*, 2014). Inhibition of HMG-CoA reductase through statins (shown in red) has high importance for medicine.

The MVA pathway is found in animals, several gram-positive bacteria, archaea, algae and the cytosol of higher plants and other photosynthetic organisms (Table 1). In short: The MVA pathway is spread over all kingdoms and habitats. The MVA pathway is especially important for the kingdom of the archaea. In these organisms IPP and DMAPP are used for the synthesis of the membrane (Koga and Morii, 2007). The high demand of IPP and DMAPP in archaea implicates a high flux through the pathway. Indeed, the MVA pathway in archaea lacks feedback inhibition of the mevalonate kinase by GPP, DMAPP, IPP or FPP, which is found in many other organisms (Primak *et al.*, 2011). Furthermore, the pathway shows alternate reactions from mevalonate 5-phosphate to IPP (Vannice *et al.*, 2014) (Figure 2).

Compared to the MVA pathway, the MEP pathway (also known as DXP or non-mevalonate pathway) was discovered much later by Rohmer *et al.* (1993). The first indications appeared in the late 1980s in isotopic labeling experiments (Flesch and Rohmer, 1988). Most gram-negative bacteria synthesize terpenoids via the MEP pathway, with the known exceptions of *Borrelia burgdorferi, Myxococcus fulvus* and *Chloropseudomonas ethylica* (Wilding *et al.*, 2000). The MVA pathway dominates in gram-positive bacteria, but species with the MEP pathway are also known. Interestingly, the MEP pathway is found in the chloroplast of algae and higher plants next to the MVA pathway in the cytosol. The two pathways are not only locally separated in these organisms, but also functionally separated. While the MVA pathway is used for the synthesis of sesquiterpenoids and triterpenoids, the MEP pathway is used for the production of hemi-, mono-, and diterpenoids (Table 1).

Table 1| Occurrence of the biosynthetic routes for DMAPP and IPP. The MVA and MEP pathway are distributed over all kingdoms. While most organisms exclusively use one of the pathways, some organism, like higher plants, inherent both. (Adapted from Steinbüchel (2003))

	Organisms	MVA pathway	MEP pathway
Vertebrates	*Homo sapiens*	+	-
Plants	*Arabidopsis thaliana*	+[a]	+[b]
Chrysophytes	*Ochromonas danica*	+[a]	+[b]
Diatoms	*Phaeodactylum cornutum*	+[a]	+[b]
Green algae	*Chlamydomonas reinhardtii*	-	+
Red algae	*Cyanidium caldarium*	+[a]	+[b]
Eukaryotic microorganisms	*Saccharomyces cerevisiae*	+	-
	Prototheca wickermanii	-	+
Archaea	*Haloferax volcanii* (Bischoff and Rodwell, 1996)	+	-
Prokaryotic microorganism gram-negative	*Escherichia coli*	-	+
	Borrelia burgdorferi, Myxococcus fulvus, Chloropseudomonas ethylica (Wilding *et al.*, 2000)	+	-
	Pseudomonas mevalonii (Wilding *et al.*, 2000)	(+)[c]	+
Prokaryotic microorganism gram-positive	*Streptomyces aeriouvifer* (Seto *et al.*, 1996)	+	+
	Staphylococcus aureus, Streptococcus pneumoniae, *Enterococcus faecalis (Balibar et al., 2009)*	+	-
	Bacillus subtilis	-	+

[a] in cytoplasm; [b] in chloroplast; [c] only partial pathway

The reason for the appearance and coexistence of two independent pathways for the same products is not understood. Furthermore, no pattern has been found in the distribution of these two pathways. More research is needed to understand the advantages and disadvantages of both pathways and therefore their coexistence.