Technische Universität München

Fakultät für Chemie | Lehrstuhl für synthetische Biotechnologie

New process strategies for the sustainable production and purification of functional terpenoids

MAX HIRTE

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs

genehmigten Dissertation.

Vorsitzender: Prof. Dr.-Ing. Kai-Olaf Hinrichsen

Prüfende der Dissertation: 1. Prof. Dr. Thomas Brück

                          2. Prof. Dr.-Ing. Andreas Kremling

                          3. Prof. Dr. Rainer Buchholz

Die Dissertation wurde am 19.03.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 01.07.2019 angenommen

Acknowledgement

Abstract

The environmental change and depleting fossil resources demand for novel and alternative approaches for the production of functional molecules. To date, most of the applied active ingredients in pharmaceuticals, cosmetics, food, home care products, etc. are natural molecules or related derivatives. Their generation is primarily conducted by chemical synthesis, because natural occurrence is often limited or natural sourcing is too expensive. However, sustainability, reducing CO_2 emission and independency from fossil resources are modern needs of our growing society. Therefore, replacing traditional, environmentally doubtful production routes for chemicals and generating bio-based, circular refineries is a particular strategy to protect our environment.

With the help of modern biotechnological methodologies, genetic information can be extracted from natural sources and transferred to heterologous hosts. These engineered bio-refineries can consequently convert renewable resources to chemical building blocks, enzymes or active ingredients. Such industrial biotechnology approaches open up economically efficient and sustainable production possibilities. In that context, strain development, fermentation process development and downstream processing are particular requirements in order to achieve a green and holistic process. For some products like proteases this process is well developed and theses enzymes are widely applied in commercial products. However, industrial biotechnology approaches are underrated for the commercial production of complex, natural products that rely on multiple biosynthetic steps. Especially the high complexity and elaborate strain development are obstacles that lowers industrial related research activities, but have to be overcome in the future.

Meanwhile, many tools and software are available that facilitate the laboratory work and allows for fast, safe and reliable genetic modification of microbial strains. Digitization (in terms of big data and artificial intelligence) will also affectively improve and support strain development. Accordingly, many big companies and start-ups start focusing on reliable and sustainable production routes for functional molecules by industrial biotechnology approaches.

Copying and activating natural biosynthesis in heterologous production hosts can provide for efficient access to complex molecules. This is possible in the case of terpenoids that have diverse and approved functionalities of commercial interest. All terpenoids rely on the same precursors that are subsequently processed by further enzymes. These processing steps are generally highly specific for the production

of a single terpenoid. Accordingly, once an efficient microbial platform for the production of the universal terpene precursor is built, it can be applied for the production of all kinds of terpenoids.

Therefore, one aim of the presented work was the generation of a bacterial terpene production platform. This platform was successively optimized by modulating enzymes actively involved in terpene biosynthesis. In order to improve the screening efficiency, a read-out system that encompasses production of the red coloured terpene lycopene was developed. Consequently, the optimized terpene production platform was used to produce diterpenes, evaluate terpene synthase enzyme mutants and to characterize unknown terpene synthase activities.

Another particular aim of this study was the holistic and sustainable process development for terpene production. In this context, I investigated the possibility to convert low value renewable resources to Taxadiene, the particular precursor of Taxol. A stable and reproducible fermentation process was developed that was also approved for the production of other terpenoids (Cylooctat-9-en-7-ol, Cubebol, Copaene, Verticillene, Aphidicolan-16-β-ol). Terpene capture and purification processes were major challenges I was confronted with, because emulsion forming and solubility issues of derived extracts kept terpene recovery low. Therefore, new process strategies for capture by stepwise extraction and liquid/liquid chromatography, as an alternative purification approach to HPLC, were investigated and evaluated for the Taxadiene and Cembratrienol isolation.

Three enzymes from the Basidiomycota *C. puteana* (annotated as putative terpene synthases) were characterized. Two of three enzymes were functional sesquiterpene synthases and allowed for the specific production of the refreshing agent Cubebol and α-Copaene (patented).

In parallel, I studied the structure function relationship of terpene synthases by virtual protein modelling and molecular dynamics studies. These results indicated structural insights of terpene synthases and allowed for the specific generation of enzyme mutants capable producing alternative terpenoid macrocycles (Apdicolan-16-β-ol synthase, Cubebol synthase, Taxadiene synthase).

List of related articles

I. Katarina Kemper, Max Hirte, Markus Reinbold, Monika Fuchs, and Thomas Brück. **"Opportunities and Challenges for the Sustainable Production of Structurally Complex Diterpenoids in Recombinant Microbial Systems**." *Beilstein Journal of Organic Chemistry* 13 (2017): 845-54. http://dx.doi.org/10.3762/bjoc.13.85.7

II. Max Hirte, Nicolas Meese, Michael Mertz, Monika Fuchs, and Thomas B. Brück. **"Insights into the Bifunctional Aphidicolan-16-ß-Ol Synthase through Rapid Biomolecular Modeling Approaches**." *Frontiers in Chemistry* 6, no. 101 (2018-April-10 2018). http://dx.doi.org/10.3389/fchem.2018.00101.

III. Wolfgang Mischko, Max Hirte, Simon Roehrer, Hannes Engelhardt, Norbert Mehlmer, Mirjana Minceva, and Thomas Bruck. **"Modular Biomanufacturing for a Sustainable Production of Terpenoid-Based Insect Deterrents."** *Green Chemistry* (2018). http://dx.doi.org/10.1039/C8GC00434J.

IV. Max Hirte., Wolfgang Mischko, Katarina Kemper, Simon Röhrer, Claudia Huber, Monika Fuchs, Wolfgang Eisenreich, Mirjana Minceva, and Thomas B. Brück. **"From Microbial Upcycling to Biology-Oriented Synthesis: Combining Whole-Cell Production and Chemo-Enzymatic Functionalization for Sustainable Taxanoid Delivery."** *Green Chemistry* 20, no. 23 (2018): 5374-84. http://dx.doi.org/10.1039/C8GC03126F.

V. Wolfgang Mischko, Max Hirte, Monika Fuchs, Norbert Mehlmer, and Thomas B. Brück. **"Identification of Sesquiterpene Synthases from the Basidiomycota Coniophora Puteana for the Efficient and Highly Selective B-Copaene and Cubebol Production in E. Coli."** *Microbial Cell Factories* 17, no. 1 (October 22 2018): 164. http://dx.doi.org/10.1186/s12934-018-1010-z.

Parts of this thesis have been presented at conferences:

I.   **2nd European Conference on Natural Products**

Dechema Biotechnology, Frankfurt

06.09.2015 - 09.09.2015

II.   **Industrial Biotechnology Forum 2016**

Technische Universität München

14.03.2016 - 15.03.2016

Related Press Releases

I.   **"SysBioTerp – Biotechnologische Produktionsplattform für strukturell vereinfachte Taxol-Derivate nutzt biogene Reststoffe"** BioEconomy Cluster, [online]

Available at:
http://www.bioeconomy.de/sysbioterp-biotechnologische-produktionsplattform-fuer-strukturell-vereinfachte-taxol-derivate-nutzt-biogene-reststoffe/ (February 20, 2019)

Content

Part I

Scientific Background

1

General Introduction

Bio-economy

The environmental change leads to rethinking of the future based economy. Specifically, the depletion of fossil resources, climate change and raising world population demands for sustainable and innovative strategies[1, 2].The Bio-economy strategy 2030 is a chance to innovate in sustainability under ecological and economic constraints and complies with the modern needs of a growing society[3-5]. The generation of processes for bio-based materials and compounds that are derived from renewable resources is a particular focus that can be achieved by biotechnological concepts[6-9]. In this context, native or modified microorganisms are used to convert and/or degrade biomass thereby generating valuable molecules. Introducing the required biosynthetic genetic information into a suitable microbial host allows for targeted molecule or enzyme production. Advantageous, microorganisms applied are usually fast growing, engineerable and non-hazardous[7, 10]. Consequently, biotechnological production processes are expanding across industrial sectors (see Figure 1).
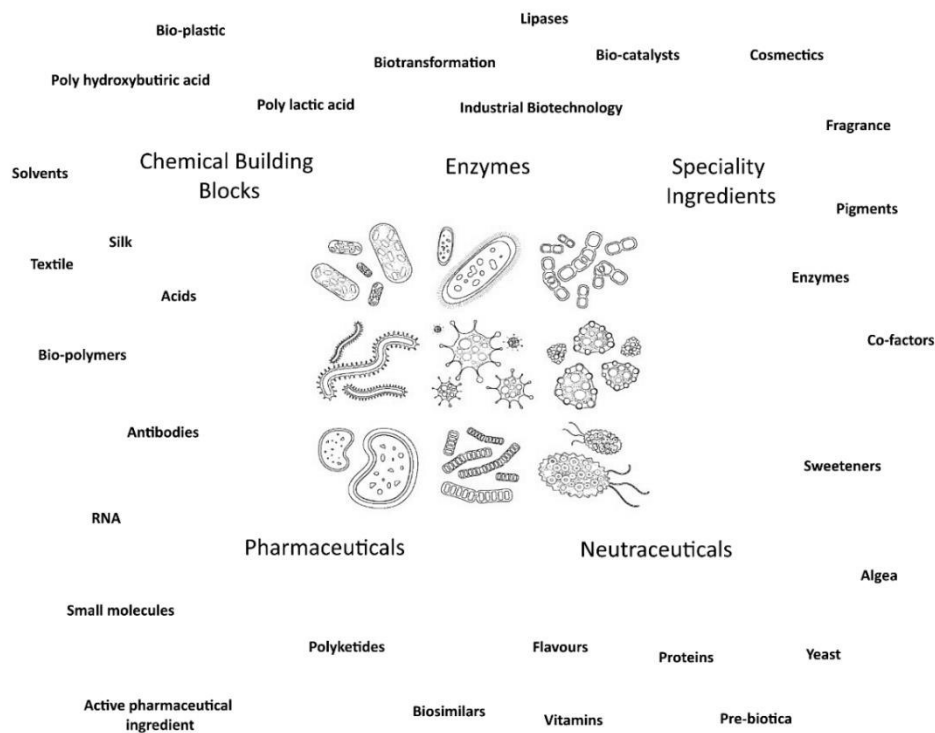
Figure 1: Biotechnological derived products

SysBioTerp

E:Bio - Innovation Competition Systems Biology (E:Bio) is a particular topic of the Bio-economy strategy 2030, which is financially supported by the federal ministry of education and research. The E:Bio initiative focused on the identification of innovative systems biology solutions for the predominant problems of our society[11]. The funding program was initiated helping to secure global food and nutrition as well as contributing to modern health care strategies. In this context, the health care research activity of the SysBioTerp program (see Figure 2) was financially supported with more than 3.7 million Euros.
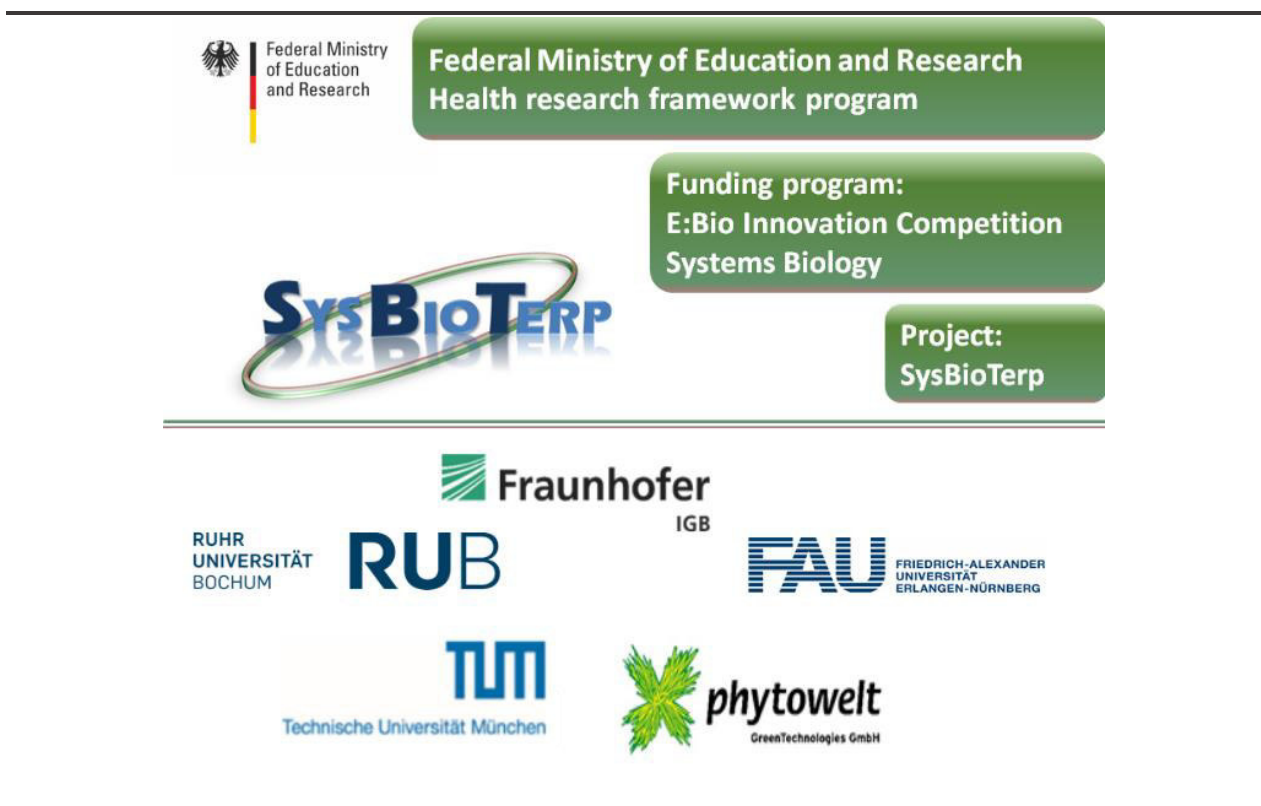
Figure 2: Project funding

The focus of this research collaboration was the development of a bacterial strain that efficiently produces structural minimized, biological active taxanoids from low value, renewable raw materials (see Figure 3).
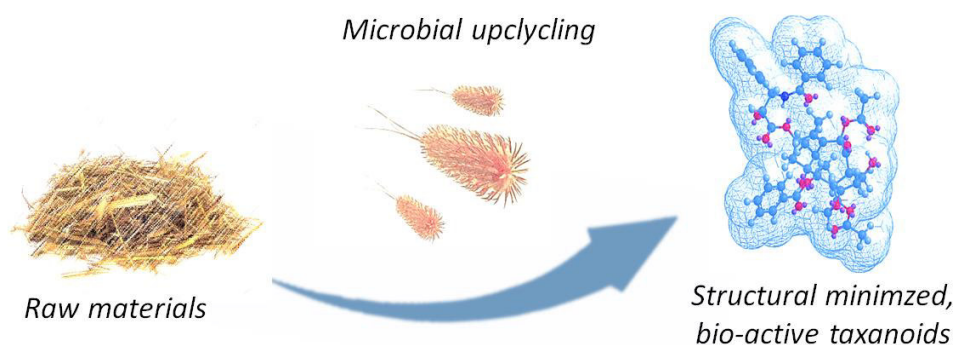
Figure 3: Principle idea of the project SysBioTerp

In terms of E:Bio, a particular project part was the mathematically based modulation and optimization of metabolic fluxes inside the cell. In order to achieve a balanced bio-refinery that efficiently produce taxanoid molecules under ecologically and economically constraints, a consortium of eight specialty divisions was acquired. The project was led and coordinated by Prof. Thomas Brück and Dr. Monika Fuchs both, from the Werner Siemens-Chair of Synthetic Biotechnology at the TU Munich. Moreover, the TU Munich contributed to the consortium, expertise in metabolic engineering, strain development and process development (Werner Siemens Chair of Synthetic Biotechnology), up- and downstream processing (Bioseparation engineering Group), metabolomics and structure elucidation (Chair of Biochemistry) and systems biology (Specialty Division for Systems Biotechnology). Other partners actively supported the project in terms of identification of enzymes that modify and "bio-activate" the core structure (Ruhr-Universität-Bochum: Microbial Biotechnology; Fraunhofer Institute for Interfacial Engineering and Biotechnology; PhytoWelt Green Technologies GmbH). Last not least, the Department of Chemical and Biological Engineering of the Friedrich Alexander University in Erlangen developed and conducted biological activity screening assays. The consortium and their respective tasks are highlighted in Figure 4.
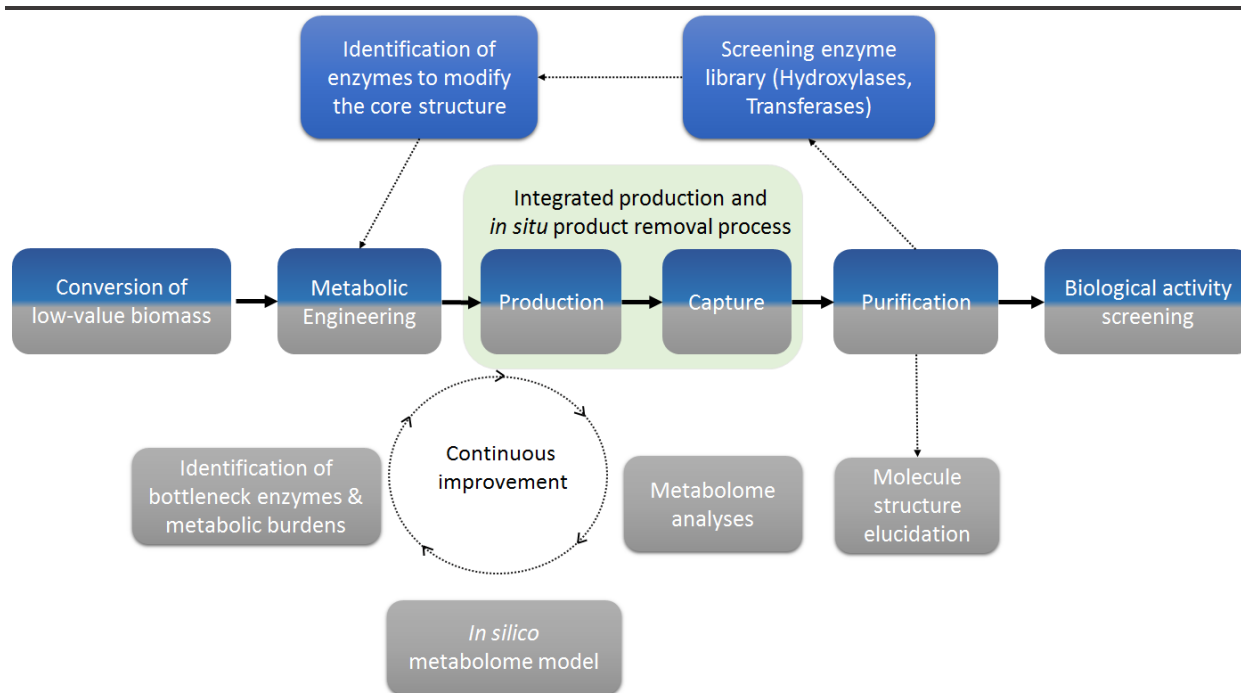
Figure 4: Project process flow

Aim of this work

The world of terpenoids is fascinating. Thousands of functions and possible applications awaken considerable attention from various industries. Terpenes will and do contribute to satisfy modern needs of our society, help protecting the environment and support us to live healthier and longer. Nevertheless, their generally low natural abundance demands for new environmentally compatible sourcing strategies. One particular strategy is the biotechnological production of these natural compounds. Thanks to advances in genetic engineering and in process engineering the basis of industrial terpene production is set. However, genetic engineers are driving microorganism to maximum performance, thereby neglecting its applicability in an industrial setting. Therefore, new, universally applicable and holistic process strategies have to be developed encompassing production, capture and purification processes. Furthermore, improved methods to introduce functional groups to the terpene skeleton have to be developed. To that end, this PhD thesis describes and demonstrates new process strategies for the sustainable production and purification of functional terpenoids.

2

Theoretical Background and Methodologies

Terpenoids

The class of terpenoids is the largest natural product family group found in nature[12, 13]. These compounds, which are often structurally complex, exhibit diverse physical, chemical and/or biological properties and functions[14-16]. Due their specific characteristics they are often used as a flavor, fragrance, cosmetic, food and/or pharmaceutical ingredient[17-19]. Interestingly, terpenoids are all derived from the same universal building blocks, namely dimethylallyl diphosphate (DMADP) and isopentenyl diphosphate (IDP)[20]. These isoprene units are naturally synthesized by two evolutionarily distinct pathways. More specifically, the mevalonate pathway, predominantly found in eukaryotes and archaea, converts acetyl-CoA over the central intermediate mevalonate in five enzymatic reactions to IDP[21, 22]. In contrast, the general prokaryotic bio-synthesis of terpenes relies on seven enzymatically catalyzed steps that convert Pyruvate and Glyceraldehyde 3-phosphate to IDP and DMADP[21, 22]. This pathway's central metabolite is the 1-deoxy D-xylulose 5-phosphate which gives this pathway its name (see Figure 5).
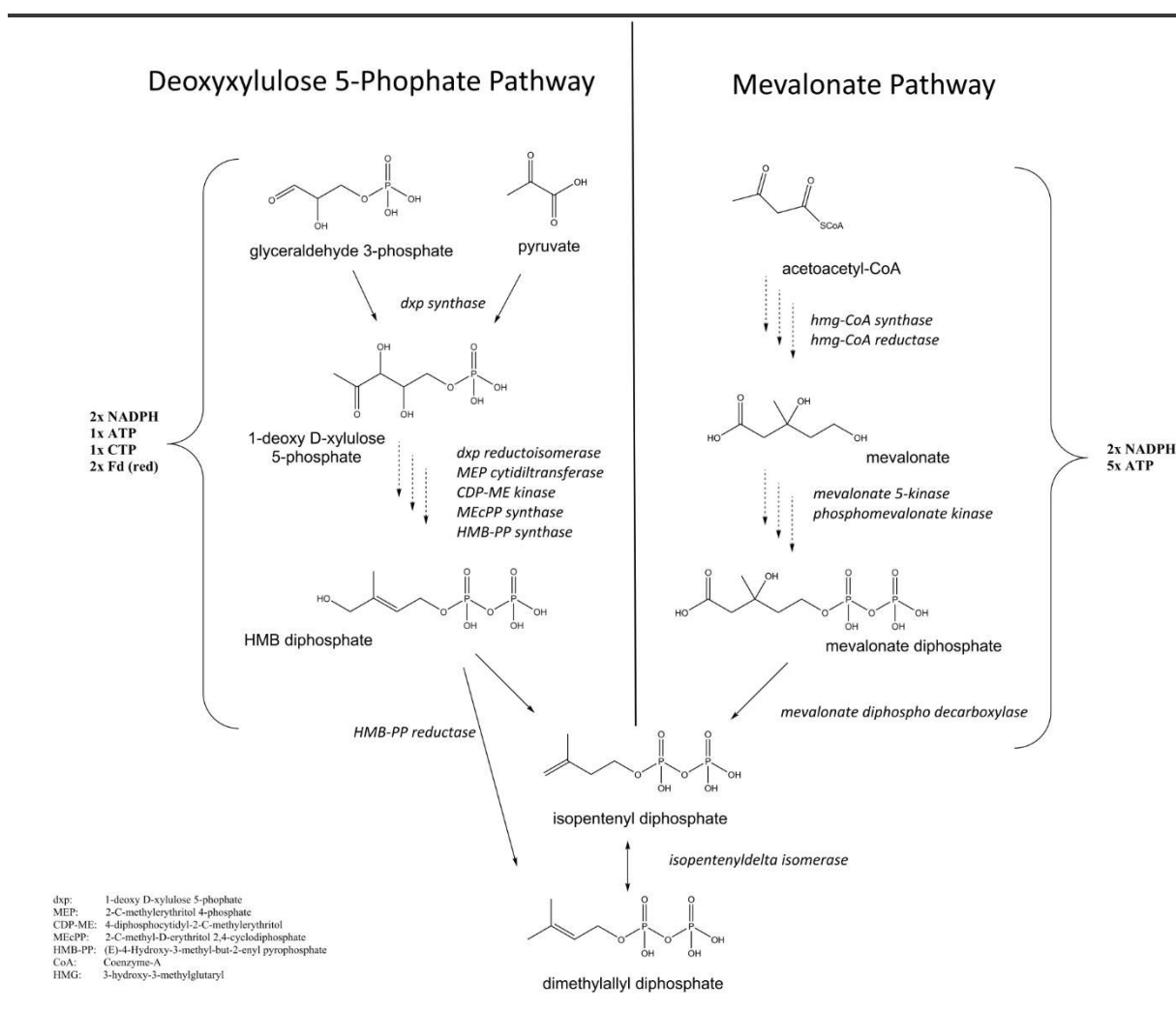
Figure 5: Biosynthesis of terpenes

In all kingdoms, the isopentenyl diphosphate isomerases enzyme family catalyzes the isomerization reaction from IDP to DMADP. This step is essential for the synthesis of longer allylic chains by prenyl transferases that successively adds up to four IDP molecules to DMADP btw. geranyl diphosphate, farnesyl diphosphate (FDP), geranylgeranyl diphosphate (GGDP), etc. by a head to tail condensation. The general biosynthesis strategy for tri- and tetraterpenoids relies on a head to head condensation reaction of two molecules FDP and GGDP, respectively. The number of isoprene units merged, classify terpenoids into hemi (1), mono (2), sesqui (3), di (4), sester (5), tri (6), tetra (8) and poly (8+) terpenoids (see Figure 6)[20].
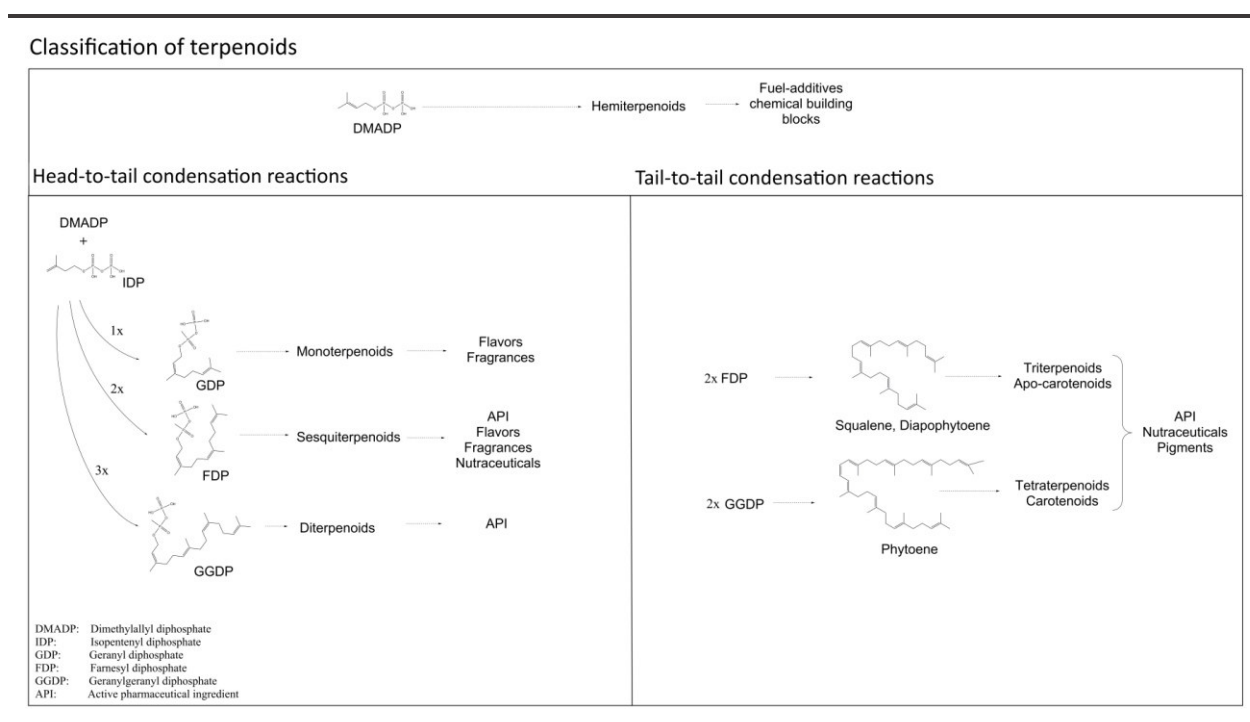
Figure 6: Classification of terpenoids

Terpenoids diversity is derived from the cyclization process of the linear allylic chains. These reactions are catalyzed by terpenoid synthases also known as terpenoid cyclases[20]. Due to changes in carbon bonding, hybridization and stereochemistry during the enzyme's cyclization reaction cascade, they belong to the most complex reactions naturally occurring[23, 24]. Terpene cyclases are classified into two main classes[20]. Class I terpenoid synthases coordinate the substrate (linear or pre-cyclized allylic chains) by a Mg^{2+} triade that allows for cleavage of the pyrophosphate group generating a highly reactive carbon cation and inorganic pyrophosphate. In contrast, class II terpenoid synthase reaction are dependent from an aspartic acid side chains that enables protonation of the terminal carbon-carbon bond of the linear allylic chains providing for a cyclized compound with the pyrophosphate group retained.

Terpenoid synthase protein structures are assembled by three evolutionarily conserved domains (α,β,γ – see Figure 7)[20]. The α-domain catalyzes the class I reactions. Class II reactions are conducted at the interface of the β- and γ-domain. Interestingly, both reactions can be conducted by a single, bifunctional enzyme that contains all three domains (e.g. Abietadiene synthase, Aphidicolan-16-β-ol synthase)[25, 26].
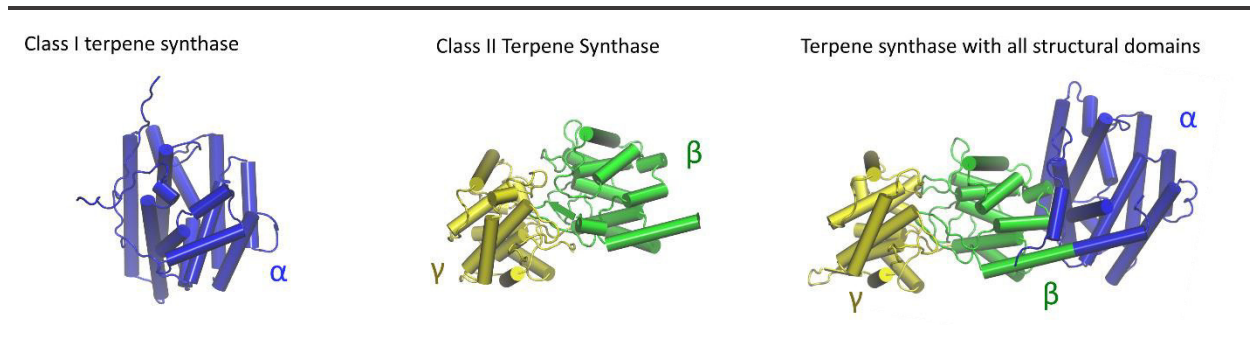
Figure 7: Classification of terpene synthases

In nature, terpenes are often further modified by the introduction of functional groups, which is often required to gain a biological function[12, 27, 28]. The hydrophobic, rigid and specific shape of a terpenoids' three-dimensional structure in combination with a functional group, provides for specific interaction with various targets[12, 29].

Taxol and derivatives

In the late 1960s scientist isolated a compound from the bark of the pacific yew tree that exhibits anti-cancer properties towards leukemia, carcinosarcoma, and lung cancer cells[30]. Subsequently, they demonstrated that this compound binds to beta tubulin thereby stopping its depolymerization[31, 32]. Based on the tree's Latin name *taxus brevifolia* they called the compound Taxol. However, the low natural abundance of Taxol and the slow growth of the tree required alternative, sustainable sourcing strategies[33]. Initially, *de novo* chemical synthesis approaches were developed, but the synthesis procedure comprised more than 40 steps[34]. A breakthrough was achieved by the development of a semi-synthesis approach that utilizes 10-Deacetylbaccatin III, which can be isolated from the needles of the European yew (*taxus baccata*), to fabricate Taxol in fewer synthesis steps. However, both chemical approaches generated tons of toxic waste as byproduct[33]. The first sustainable Taxol production process was established by utilizing plant cells cultures[35]. Nonetheless, slow growth, costly purification steps and necessity of metrics of waters for the production, still demands for alternative production strategies[33].

To that end, scientists intensively researched on the elucidation of the biosynthetic pathway to use this information for biotechnological attempts[36, 37]. In the 90ies the first key step was discovered unraveling the initial step of Taxol biosynthesis[38]. The first significant step is the cyclization of GGDP to Taxadiene (TD), a three cyclic diterpene. TD's molecular structure is still apparent in Taxol (see Figure 8). On the way from TD to Taxol several cytochrome p450 monooxygenase (CYP), acetylase and transferase steps are required. Interestingly, some of the enzymatically catalyzed reactions or spontaneous rearrangements have not been elucidated and understood, yet[36]. The initial TD CYP hydroxylation processing reaction already produces significant limitations in heterologous bio-synthesis imitation[39]. The enzyme generates *in planta* 5-alpha-hydroxy taxadiene (5HTD)[36]. However, heterologous expression of this enzyme and evaluating its function *in vivo* as well as in *in vitro*, revealed that the annotated intermediate 5HDT is only produced as a byproduct in minor amounts[40-42]. Accordingly, complete elucidation of the natural Taxol biosynthesis and its utilization for heterologous, biotechnological attempts is pending.
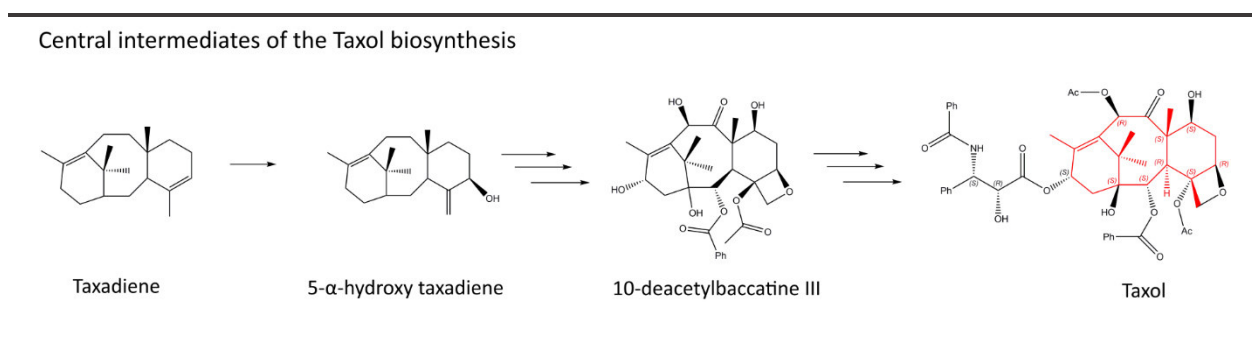
Figure 8: Central metabolites of the Taxol biosynthesis

Aphidicolin – Aphidicolan-16-β-ol

Aphidicolin (AC) is a tetracyclic diterpenoid with four hydroxyl groups[43]. This bio-active substance has been firstly isolated in 1970ies from the fungus *Cephalosporium aphidicolia* and still to date, AC has been exclusively found in the fungal kingdom[44, 45]. It is a potent inhibitor of eukaryotic DNA polymerase α and it is often used as cell synchronization agent[46-48]. Moreover, anti-viral and anti-tumoricidal activities have been recently demonstrated for AC[49, 50]. Hence, this diterpenoid is a highly interesting lead structure for further drug development.

Notably, the biosynthetic cluster was identified in the fungus *Phoma betae* and the biosynthesis elucidated (see Figure 9)[51]. AC is synthesized naturally from GGDP by a diterpene class II and class I cyclization reaction producing Aphidicolan-16-β-ol[52]. This reaction is conducted by the

bifunctional Aphidicolan-16-β-ol synthase (*acs*). Subsequently, Aphidicolan-16-β-ol. is hydroxylated by two CYPs (pbP450-1, pbP450-2) reactions generating AC[51].
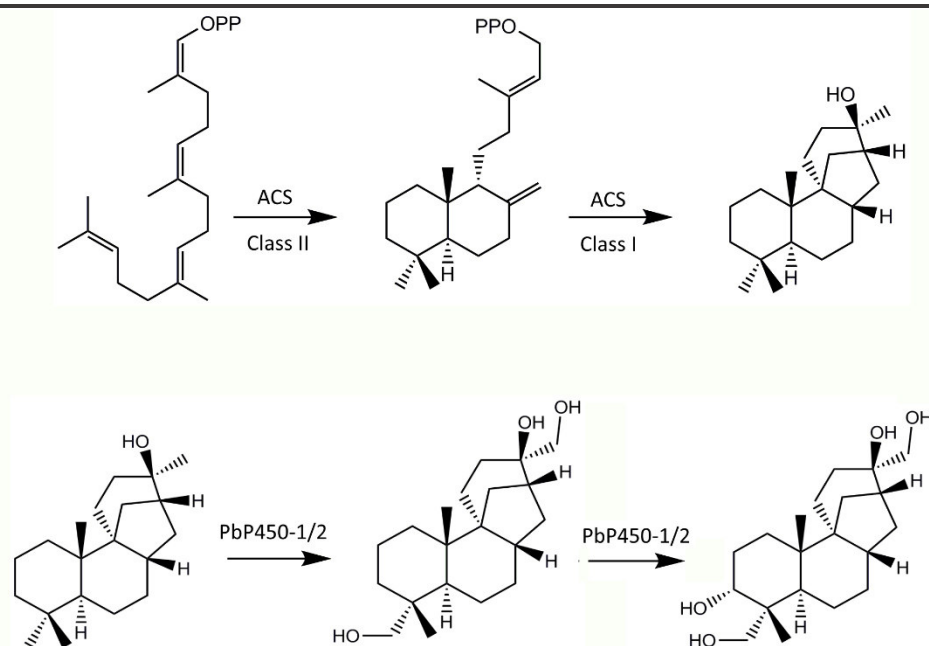
Figure 9: Biosynthesis of Aphidicolin

Cyclooctatin – Cyclooctat-9-en7-ol

The tricyclic diterpene Cyclooctatin (CO) was firstly isolated from the broth of *Streptomyces melanosporofaciens* while screening for anti-inflammatory compounds[53]. Its activity relies on the reversible inhibition to lysophospholipases thereby blocking the hydrolyses of fatty acids. The biosynthesis of CO starts with the cyclization of GGDP to the 5-8-5 fused ring system cyclooctat-9-en-7-ol. This reaction is catalyzed by the cylooctat-9-en-7-ol synthase[54, 55]. Structure function analyses of cylooctat-9-en-7-ol synthase allowed for the identification of catalytic relevant amino acid side chain in the active site cleft. In this regard, specific amino acid substitutions lead to the production of abortion products within the cyclization cascade (see Figure 10)[54]. Consequently, a production system for the targeted generation of the highly valuable diterpenes cembren A and dollabelatriene was created.
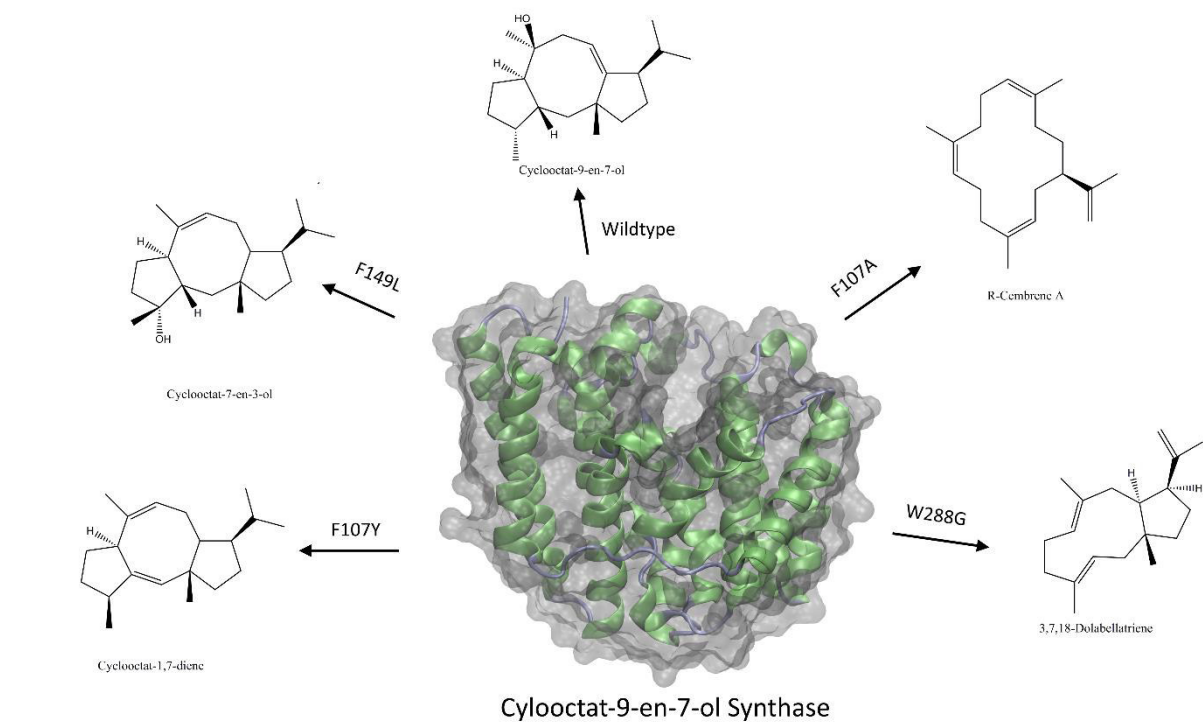
Figure 10: Cyclooctat-9-en-7-ol synthase product portfolio

As depicted in Figure 11, Cylooctat-9-en-7-ol further reacts in the natural host by two successive hydroxylation steps to CO. The CYPs CotB3 and CotB4 were identified to conduct this hydroxylation in the native host[56]. The natural producer is inefficient in terms of growth and product titer. Therefore, alternative biotechnological production routes have been investigated recently[57]. However, the P450 hydroxylation still remains a limiting step for commercial CO production.
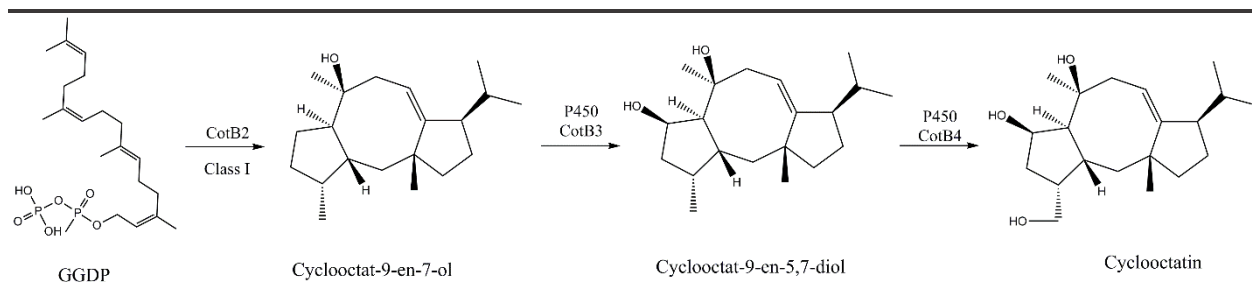
Figure 11: Biosynthesis of Cyclooctatin

(Diapo)-Carotenoids

(Diapo)-Carotenoids are widely distributed in nature and exhibit interesting pigment properties[58, 59]. The carotenoid molecular structure exhibits a polyene chromophore which provides for a specific compound coloration. For example, well known carotenoids are Lycopene and β-carotene that give tomatoes and carrots their typical color. Moreover, the high desaturation in carotenoids molecular structure allows for excellent anti-oxidative action[60]. In this regard, carotenoids and their partly collourless derivatives have received considerable attention from food, pharmaceutical, and cosmetic industry and their biotechnological production is in the scope of various companies[61, 62].

These ubiquitous pigments are naturally synthesized by many organisms particularly in plants, fungi and bacteria. Commonly, they act as light harvesting, photoprotection agent or as a phytohormone precursors[59]. Carotenoids are derived from two molecules FDP or GGDP merged in a head to head condensation leading to diapo-phytoene or phytoene, respectively (see Figure 6). Further modification by desaturases, hydroxylases, transferases and/or lyases lead to the myriads of carotenoids and their derivatives found in nature (see Figure 12).
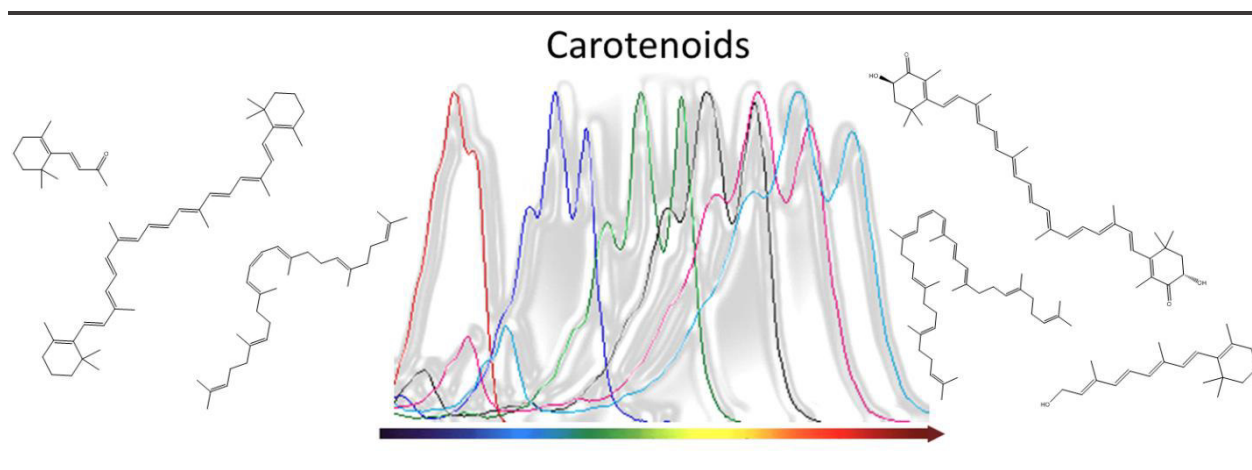
Figure 12: Diversity of carotenoids and its derivatives

Lycopene has been used as a read-out system to improve heterologous terpene production through specific pathway optimization[63, 64]. More specifically, it is highly suitable for the optimization of diterpene production platforms, as both terpenoid classes rely on the GGDP precursor.

Microbial Terpene Production

Terpenoids are already used in various industries because of their valuable physical, biological or chemical characteristics[22]. However, natural occurrence is in most of the cases limited. Therefore, evaluation und targeted engineering of heterologous hosts is a sustainable and attractive alternative to natural sourcing[65, 66].

Terpenoids all rely on the same building blocks IPP and DMAPP which is commonly bio-synthesized by two metabolic pathways in nature. The generation of production platforms able to produce high amounts of these precursors provide for commercial terpene production[67].

The most common used production hosts are *E. coli* and *S. cerevisiae* reaching terpene titers of up to 27,4 g/L [68] and 23 g/l [69], respectively. However, other production systems have also great potential to be applied in commercial production (*Pseudomonas putida, Bacillus subtilis, Yarrowia lipolytica*)[70-72].

Hence, choosing the perfect host is not trivial as growth, atom efficiency or enzyme solubility/functionality/ modification (e.g. glycosylation) are all impacting the final terpene yield[66]. In this accordance, some terpenoids can be effectively generated by one host and are hardly to produce in other production hosts. Taxadiene for example, the first key intermediate in the Taxol biosynthesis, can be hardly generated in *S. cerevisiae* (8.7 mg/L) [73], but good yields were reported in engineered *E. coli* (1 g/L) [74]. To that end, we chose to focus on the production of functionalized Taxanoids in *E. coli*.

Evaluation of Promoter Systems

Gene expression in all organism is conducted by polymerases. These enzymes bind to a promoter sequences on the DNA strand which subsequently results in the translation of the DNA to generate mRNA required for protein biosynthesis. Promoter regions are often regulated by short DNA sequences just in front of the promoter sequence (repressor or activator regions). One famous system is the lac-operon. For heterologous enzyme expression in *E. coli* this repressor is often combined with a promoter sequence recognized by the viral T7 RNA polymerase[75]. The T7 polymerase is a very powerful enzyme for fast and extensive protein-biosynthesis. However, strong overexpression also results in side effects like the formation of inclusion bodies, strong depletion of central metabolites or initiation of the dying phase[9, 76, 77].

Terpene biosynthesis by the DXP-pathway involves 7 enzymes[21]. Overexpression of all enzymes would require extensive energy resources. Interestingly, the pathway is regulated by its intermediates. Accordingly, overexpression of the first enzyme dxs, results in the upregulation of the subsequent

enzymes within this pathway[74]. An additional overexpression of prenyl transferases and terpene synthases provides for diverse possibilities to generate terpene molecules. Still, this requires relevant energy resources. For terpene production in *E. coli* this results in a depletion of the central metabolites Pyruvate and Glyceraldehyde 3 phosphate, that directly affects viability of the cell. In the case of terpene whole cell biosynthesis by a T7 lac-operon regulated expression, the productivity of the host is limited. Even if applied in the culture's stationary phase, timely restricted production, equal inducer distribution and difficult repeatability limits commercial application[9]. Unfortunately, the initiation of expression by addition of IPTG cannot be regulated neither stopped and re-initiated.

In order to overcome the "all or nothing" strategy, other promoter systems (e.g. polymerases) can be used. The inducible XylS/Pm system from *Pseudomonas sp.* allows for gene expression strength regulation by varying inducer concentration or inducer molecule itself[78, 79]. Another strategy next to inducible systems is the application of native polymerases and its promoter sequence. These promoters are not dependent from addition of an inducer molecule. They are often activated by central metabolites or genes regulating the internal metabolism. For example, the Lac-I promoter sequence is accessible to the polymerase in the presence of Lactose[80]. Moreover, modulation of the expression strength is possible by genetic modification of the promoter sequence. This modification affects the polymerase binding affinity[81].

Generation of Optimized Polycistronic Operons for Terpene Production

Terpene biosynthesis is dependent on a cascade of enzymatic reactions. If the secondary metabolism is activated by the overexpression of the first cascade's enzyme dxs, the other DXP-pathway enzymes will automatically be up- regulated by intermediates or activated regulatory elements[74]. Nevertheless, some of these enzymes still display bottlenecks. Their individual overexpression lead to increasing terpene titers[82]. However, expressing several proteins at once commonly require several plasmids and energy resources. Accordingly, the metabolic burden increases and different selection markers are required to prevent plasmid loss.

BioBrick cloning is a well-established method to produce polycistronic operons[83]. This helps to combine several genes in a row under one promoter sequence. Accordingly, the number of plasmids required can be reduced. Biobrick follows the strategy of compatible sticky ends that are released by restriction enzymes recognizing different nucleotide patterns. This allows for continuous cloning having the possibility to add one gene after another. The length of the operon is limited due to plasmid size restriction. The produced polycistronic operon can be easily transferred into an expression vector afterwards. This has already been demonstrated for the production of the tetraterpene beta-carotene[84]. The focus in this study was the modulation of the ribosomal binding sites to identify a

metabolically optimized production platform for terpene production. In this study they used a T7 promoter systems. In this study we adapted the Biobrick cloning to generate a library using a constitutive promoter system that do not require induction. This allowed for direct, visual observation of the productivity of constructed polycistronic operons (see Figure 13).
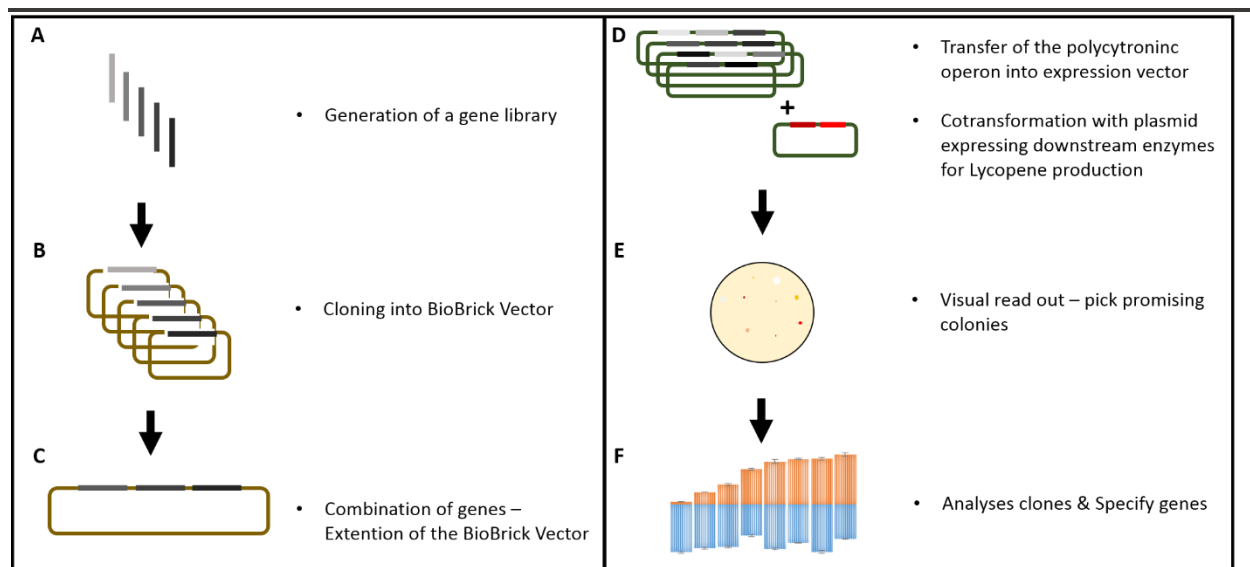
Figure 13: Adapted BioBrick cloning for terpene production optimization

Fermentative Production of Terpenoids

Fermentation describes the controlled cultivation of microbial or eukaryotic cells. An obvious advantage compared to shake flask experiments for instance is the regulation of pH and the dissolved oxygen inside the vessel. This allows for stable cultivation conditions that return reproducible results and higher cell densities. In the case of terpene production, microbial fermentation systems are most commonly applied. Yeast or bacteria can achieve high cell densities and they grow fast. The growth rate can be determined according to:

$$\mu = \frac{1}{c_x} \frac{dc_x}{dt}$$

and

$$\mu = \mu_{max} \frac{S}{K_S + S}$$

Substrate uptake and product production rate can be calculated by the following equations.

$$q_s = \frac{1}{c_x}\frac{ds}{dt}$$

$$q_p = \frac{1}{c_x}\frac{dp}{dt}$$

The coefficient Y_{XS} represents the share of grams substrate that are required to generate 1 g of biomass. Y_{PS} represents the analogues share of product and substrate, and can be considered as atom efficiency grade.

$$Y_{XS} = -\frac{\Delta X}{\Delta S}$$

$$Y_{PS} = -\frac{\Delta P}{\Delta S}$$

In fed-batch fermentation processes the control of carbon sources allows for strategic feeding to enhance the biomass and product yield. An immediate increase for dissolved oxygen inside the reactor is an indication for complete consumption of the primary carbon source. Another indication for complete primary carbon source conversion is a pH increase that originates from Acetate uptake by the microorganism. In this context, a feeding protocol was developed in VBA scripting that takes the signal of pH increase into account to initiate carbon feeding. This protocol provided for an autonomous and stable running fermentation process.

In parallel, the batch media, feed composition and process control were optimized to generate a reproducible, stable and efficient fermentation process. Other fermentation strategies (e.g. *in situ* terpene extraction) were investigated due to commonly adherent problems for microbial fermentation process for terpene production as listed:

1. Distribution/ Location of the product

Terpenes core structures are hydrophobic. The molecule easily diffuses through the cell membrane according to the concentration gradient. However, when the terpene concentration increases relevant amounts of product are incorporated or attached to the cell membrane due to their physico-chemical properties[9].

2. Functionalization of the terpene

Diffusion, especially of the initial product released by terpene synthases, limits an efficient functionalization of P450 catalyzed hydroxylation inside the cell[41].

3. Toxicity and Inhibition

Terpenes physico-chemical properties often provide for a biological function. Hence, a possible interaction with the product and the cell viability should be investigated[8].

4. Depletion and Stability

During fermentation decreasing product yield is observable. These can be traced back to depletion by aeration[9]. Especially, monoterpenes and sesquiterpenes have low boiling point. Strong aeration will subsequently lead to a product depletion[85]. Therefore, aeration should be adequate, even if diterpenes or higher terpenes are produced.

Some terpene products are also quite instable due to oxygen level, temperature, light etc. and their structure can be rearranged to more stable derivatives[85].

Capture and Purification of fermentatively derived Terpenes

Evaluation of different Capture strategies

Extraction of terpenes were performed by adsorption and absorption studies. Both strategies were applied for post-cultivation batches and during the fermentation process (*in situ*). Figure 14 gives an overview of the tested scenarios in this study.
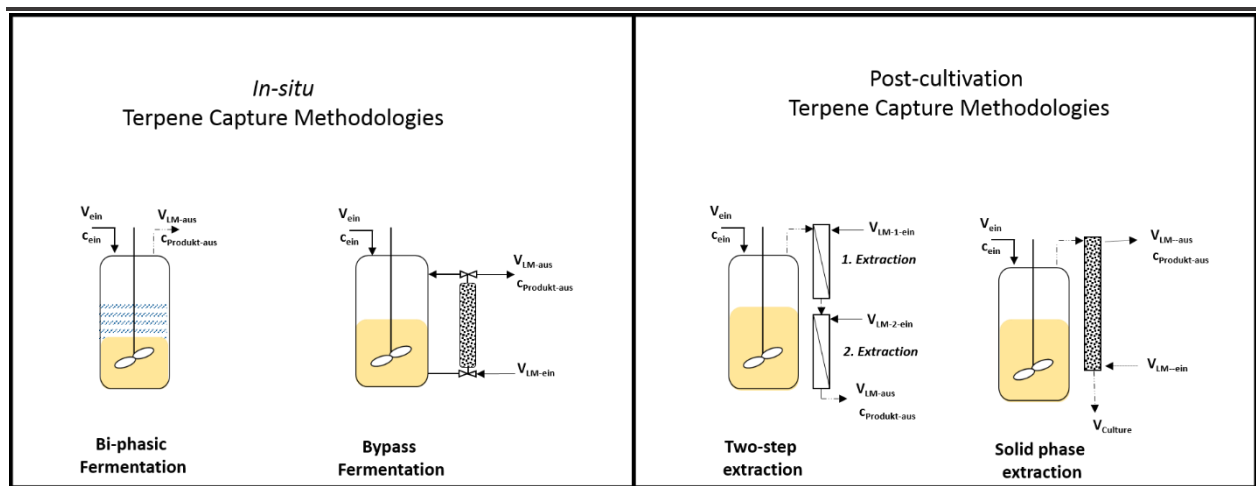
Figure 14: Different capture methodologies for fermentatively derived terpenes

In general adsorption returns a more specific binding of the target molecule, which results in reduced complexity of the extract for purification. However, the costs for adsorption material is high compared to solvents and unfortunately, the regeneration and re-use is limited due to reduced binding affinity[8, 9]. Another limitation is the handling, cleaning and dispose of the adsorption material after treatment with the bacterial broth.

Post-cultivation liquid/liquid extraction was optimized in order to reduce emulsion formation at the interface. Therefore, a two-step extraction protocol was devised. In a first step an ethanol ethyl acetate mixture breaks the cells and precipitate most of the proteins. After centrifugation of the cell debris, the liquid is mixed with hexane. This results in a phase separation and an extract with relatively low complexity[9].

In situ liquid/liquid extraction can be easily conducted by addition of a solvent that is hardly miscible with the culture medium (bi-phasic fermentation). However, a solvent must be identified which is not-toxic to the cells neither to the environment, can absorb the product and does not form stable emulsions. Moreover, the solvent and the product must be separable afterwards.

Evaluation of different Purification strategies

Chromatographic separation is a common method for the isolation and purification of chemical and biological molecules. More specifically, liquid chromatography (LC) is one of the standard methodologies applied in research and industry. The isolation is based on the interaction of the target molecule with the stationary phase. Moreover, variation of the mobile phase can enhance or reduce this interaction with the stationary phase providing for the possibility to isolate the compound in high purity[86].

Liquid/liquid chromatography (LLC) uses a liquid phase as mobile as well as stationary phase[87]. Accordingly, combining mobile and stationary phase must generate a bi-phasic system (e.g. water and benzine). Similar to the classical LC, the interaction of the target molecule with the liquid stationary phase allows for the compound's isolation. Additionally, in LLC an interaction with the mobile phase is also required. To that end, changing mobile and stationary phase is also possible for the separation process.

In the case of Hexane and water one can imagine that a very hydrophobic molecule will not solve or interact with water (e.g. dodecane). However, if you take Ethanol, it will solve in water as well as in the Hexane phase. A number to specify the compound is the log-p value, which represents the partition coefficient of the molecule in an n-octanal-water bi-phasic system. This number is correlated to the hydrophobicity of the compound and it can be theoretically determined or even estimated by computational simulation calculation (e.g. in COSMO-RS)[88].

The hexane-water-ethanol mixture is a good example to demonstrate the behavior of ternary mixtures. Those mixtures can be displayed in triangle diagram showing the biphasic system (section II see figure 15) and mono-phasic system (section I see figure 15) for different shares of hexane, water and ethanol.

To separate a molecule by LLC the distribution of the target molecule should be equal in both phases in the biphasic systems.
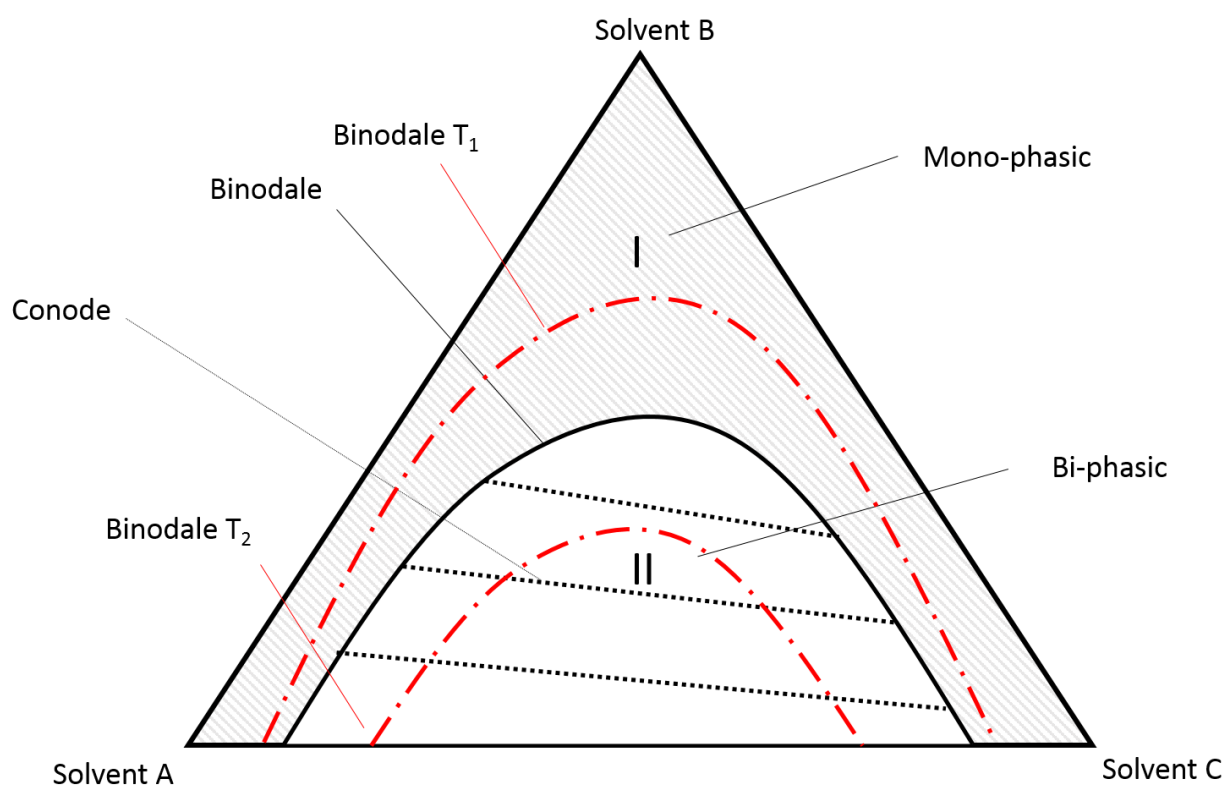
Figure 15: Example of a ternary mixture displayed in a triangle diagram

The composition of a ternary system can be varied to change the polarity and hydrophobicity of both phases. The characteristic of a biphasic system slightly changes along a conode. However, other variations of the mixture affect the distribution of the target molecule. The binodal curve represents the border of the biphasic zone. Mixtures close to this line are most probably instable. Especially, because many ternary systems are also affected by their temperature (see figure 15 red dotted lines).

A parameter to determine the stability of a biphasic system is the settling time. It represents the time that is required for two phases to completely separate after shaking. Long settling times lead to complicated separation processes and the risk of phase leakage during the separation process. Another impacting factor for long settling times can be the target molecule itself or the mixture applied. Molecules that interact or stay at the interface of both liquids dramatically change the behavior of the ternary mixture.

Nevertheless, once a suitable and stable biphasic system is identified this separation technology provide for high throughput rates (compared to HPLC) and the possibility to scale up the process

easily. The recovery rate of the molecules is close to 100% and the process conditions are mild (temp, pressure, etc.)[8, 9]. Figure 16 shows a general process flow of LLC method development.

Biphasic systems

Determination of the distribution coefficient

Preparation of the machine – Equilibrium state

Load sample and separate target compound

t [min]
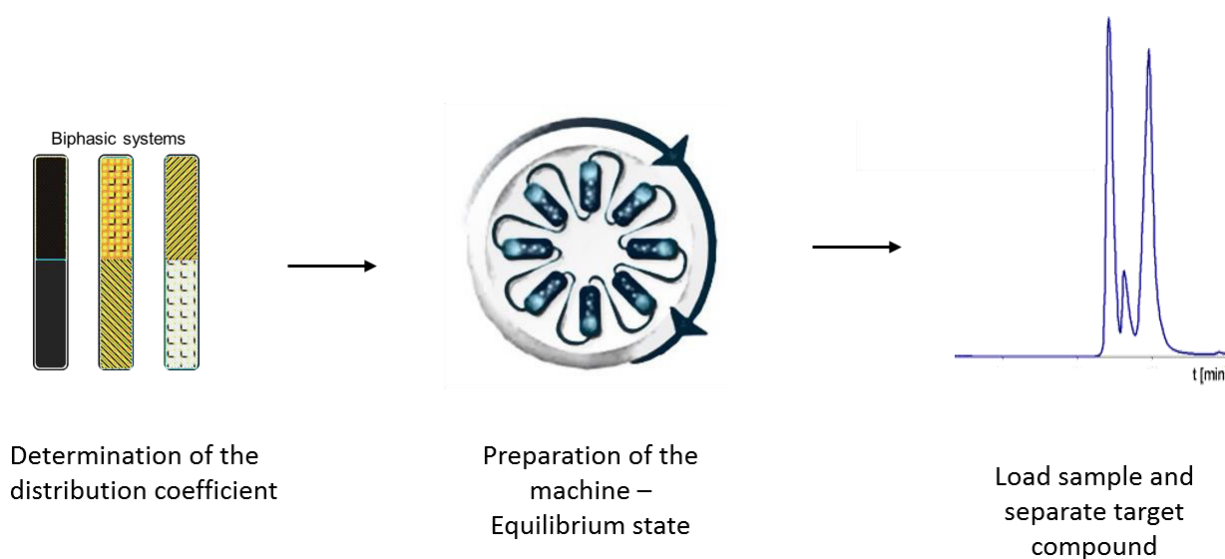
Figure 16: Process Flow for LLC separation for terpene purification

There are different procedures/machines how to apply LLC. Counter-current-LLC pumps both phases in different direction, similar to a counter-current extraction. Centrifugal partitioning chromatography pumps the mobile phase through the stationary phase, which is held in position by centrifugal force.

Cytochrome p450 Monooxygenases – Reductases

Cytochrome P450 oxygenase (CYP) is the largest and most versatile enzyme superfamily[89]. They are particularly needed for biosynthetic functions and xenobiotic oxygenations[90]. They integrate atmospheric oxygen into carbon hydrogen bonds that provides for the possibility of further biochemical processing. In order to achieve this reaction, oxygen initially binds to a ferrous atom that is centrally coordinated by enzyme bound heme[91, 92]. The subsequent reductive activation process, produces the highly reactive ferryl-oxo species, as depicted in Figure 13. This transition state generally attacks the substrates non-activated carbon hydrogen bond and substitute a hydrogen by a hydroxyl function.
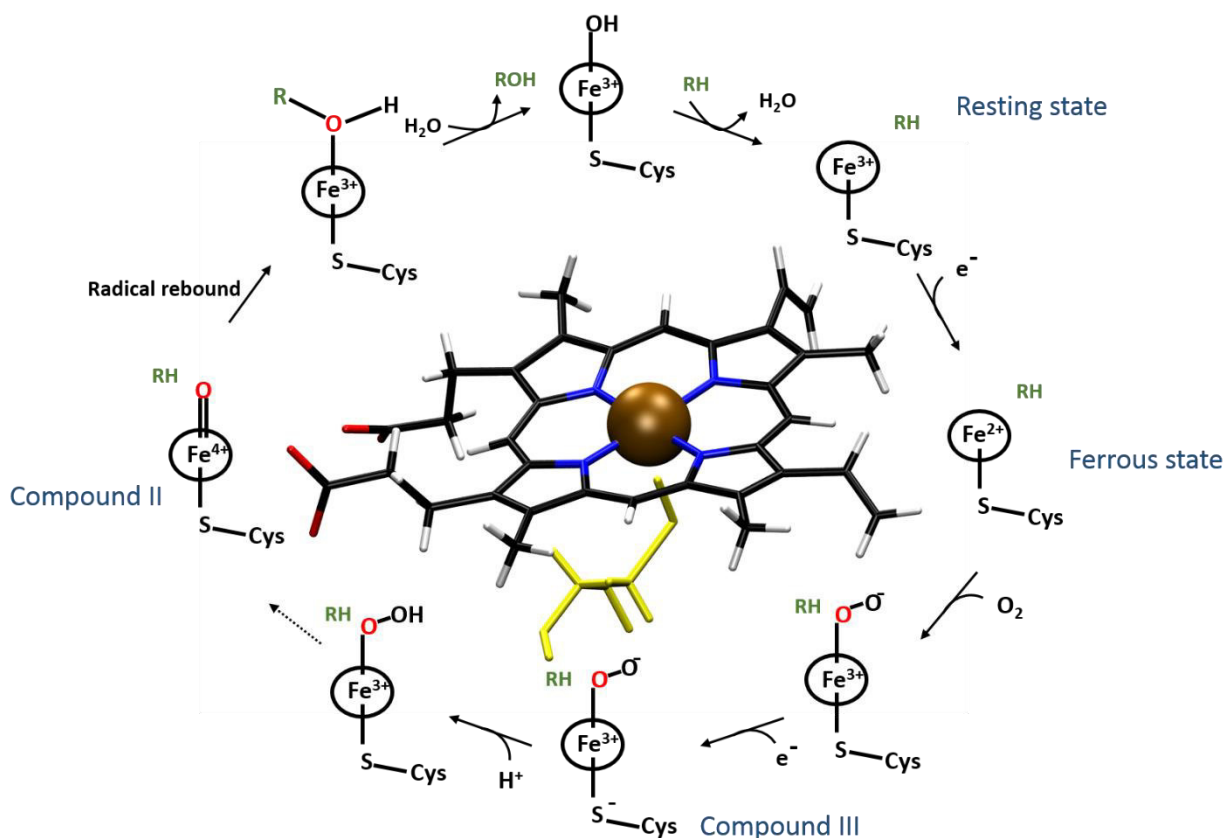
Figure 17:Process of CYP catalyzed hydroxylation reactions

Electrons for the ferryl-oxo generation are usually provided by the cofactors NADH or NADPH but cannot be directly obtained from these cofactors. Thus, redox partner systems are engaged to facilitate the transport of the electrons to heme[93]. CYPs generally receive reducing equivalents either by the interaction of a flavoprotein and an iron sulfur-protein or by a cytochrome p450 reductase (CPR)[93]. Both systems have evolved in eukaryotes and prokaryotes. The eukaryotic mitochondrial and microsomal CYP-systems are mostly membrane associated or covalently bound by a membrane anchor. In contrast, bacterial CYP-systems are generally cytosolic and soluble. CPRs are structurally conserved and can be arbitrarily applied for the electron transfer of CPR dependent CYP systems (Figure 14)[42]. Although, flavoproteins and iron-sulfur proteins are also structurally similar across kingdoms, they cannot substitute each other[94].
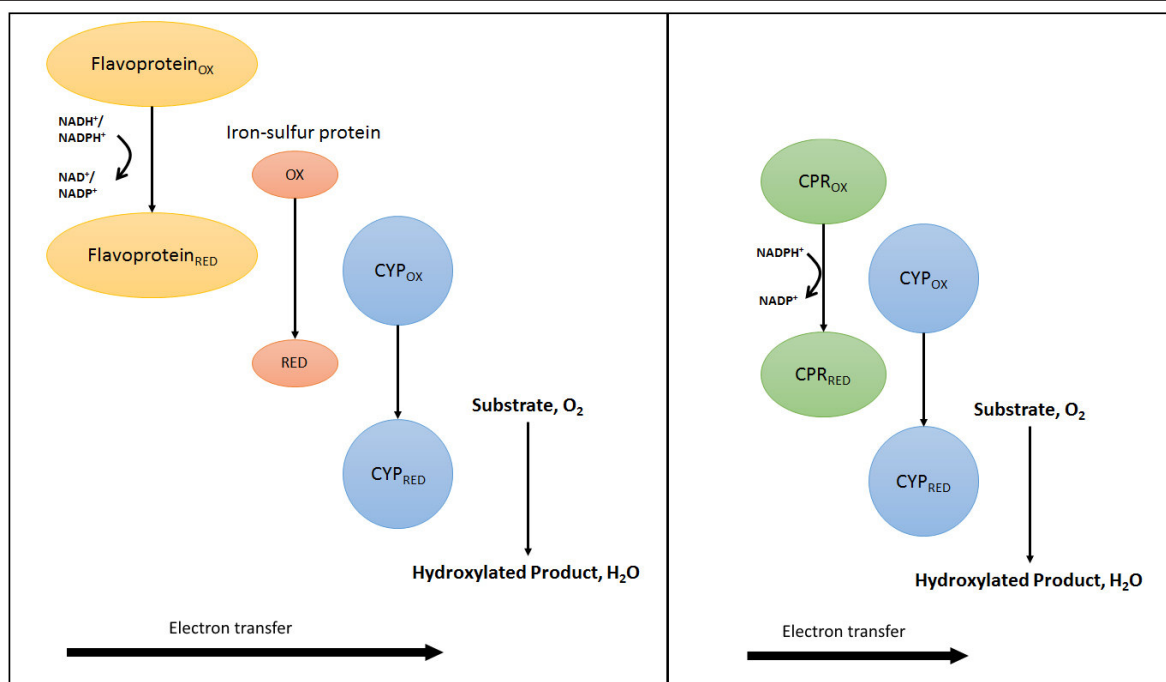
Figure 18: General electron transfer in P450 catalyzed reactions

CYP hydroxylation reaction are often the limiting step in heterologous whole cell biosynthesis approaches, because they rely on cofactors and are dependent from the efficiency of the electron transfer[95]. The enzyme from the *bacillus megaterium* P450 BM3 contains both, a mammal-like CPR and a CYP domain in a single enzyme[96]. This chimeric enzymes allows for improved electron transfer and accordingly, served as a model for the development of CYP-CPR fusion enzymes[97]. Major limitations, however, lie in the individual evaluation of fusion site (N- or C-terminal), linker sequence and truncation for every construct. Moreover, functional and sufficient expression in whole cell bio-refineries is challenging due to the big protein size and low solubility[95, 97].

Introduction of Functional Groups to the Terpene Skeleton

Statistically, introduction of functional groups to the high dense carbon terpene core structure is likely to result in a biological activity[12, 98, 99]. One possibility is the introduction of oxo-functionalities. Sustainable approaches encompass the specific hydroxylation by P450 mono-oxygenase or lipase mediated epoxidation reaction[90, 100].

Lipases catalyzes the reaction of an acid to a per-oxo acids in the presence of H_2O_2. This per-oxo acid attacks the olefinic bonds releasing an epoxide and the original acid. In this context, a methodology was developed to epoxidate TD[9]. Compared to commonly applied epoxidation reactions that are initiated by perchloric acid, this strategy represents a mild and green alternative.

P40 Monooxygenases are commonly highly specific. Whole cell biosynthesis for the production of hydroxylated terpenes have been demonstrated. However, limitations for industrial application are often related due to low turnover[41, 42]. This originates from several sources:

- Low enzyme solubility

- Low electron transfer rate (CYP-CPR)

- Substrate already outside the cell

- High metabolic pressure

- Toxic products

Alternative strategies in whole cell biosynthesis like the implementation of mutated *Bacillus megaterium* BM3 variants (natural fused p450 enzyme with reductase domain) to overcome electron transfer limitation was demonstrated successfully[101]. However, mutation of this enzyme lowers product specificity by widening substrate acceptance.

Therefore, new P450 hydroxylation strategies were investigated. The idea to in-vitro hydroxylate a terpene skeleton by P450 monooxygenase is favorable due to low reaction volumes, high turnover possibility and easy extraction of the target compound[102]. In order to improve electron transfer rate and subsequently hydroxylation efficiency, different reductase systems were screened. One approach was the generation of fusion enzymes, however low solubility of the complex remained a major issue. Alternatively, a bacterial reductase system in combination with a bacterial mono-oxygenase was investigated. This combination lead to high turnover rates of the substrate. Subsequently, this reductase system was also screened for its possibility to transfer electron to plant P450 mono-oxygenase. The successful approach led to further development encompassing the generation of a cloning vector for a P450-CYP library generation (CYP-Swinger).

CYP-Swinger: New strategies for the efficient oxo-functionalization of diterpenes by new-to nature oxygenase-reductase combinations

Oxo-functionalities are very common and widely distributed in nature. These reactions are primarily catalyzed by CYPs that make up the biggest enzyme super family found in nature[89]. CYPs are particularly needed for biosynthetic functions and for xenobiotic oxygenation reactions[90, 103]. They integrate atmospheric oxygen into carbon hydrogen bonds that provides for the possibility of further biochemical processing. Interestingly, over 93% of the listed terpenoids in the Dictionary of Natural Products contain more than two oxo-functionalities[12]. The high carbon density in combination with

the oxo-functionalization generates unique molecular shapes and structure that can provide for a biological function[104, 105].

However, hydroxylation processes in biotechnological whole cell approaches still remains challenging, especially in bacterial fermentation processes[41, 95]. Most of CYPs are membrane bound and truncation of the respective membrane anchor is required to express the protein soluble. However, these shortening can affect protein folding and its activity.

Another limitation in heterologous CYP expression lies in the necessity of a reducing partner[93]. The reducing partner provides electron equivalents for the generation of the highly reactive ferryl-oxo species inside the CYP's cofactor heme. There exist several different CYP-reductase systems, however, the common limitation in whole cell production systems is the efficient and fast electron transfer between these partners. Fusion enzymes consisting of a CYP and a CPR connected by a linker sequence, were generated and functionally tested in previous work[106]. This approach was successful demonstrated and re-constructed for the hydroxylation of TD[107]. Experiments revealed that this fusion construction setting (e.g. fusion site, linker sequence) cannot be equivalently applied for other CYP hydroxylation approaches (e.g. hydroxylation of cembratrienol). However, we demonstrated that the CPR from *T. baccata* can be substituted by a CPR from *Roseus catharanthus* resulting in improved TD hydroxylation[107]. However, TD turnover to hydroxylated taxanoids remains below 10% (w/w) in fermentation experiments. Especially the low protein expression, due to the big protein size (124 kDa), and co-factors dependency were identified as particular bottlenecks for low turnover rates[107]. Therefore, we set out to develop new hydroxylation strategies to improve this limiting step.

Inspired by the heterologous CO biosynthesis previously described [57], we abandoned the fusion enzyme approach and developed a system, independently expressing the CYP and CPR[107]. This plasmid was designed with specific enzyme restriction sites that allows for simple and fast replacement as well as further addition of CYPs and CPR sequences (see Figure 15).
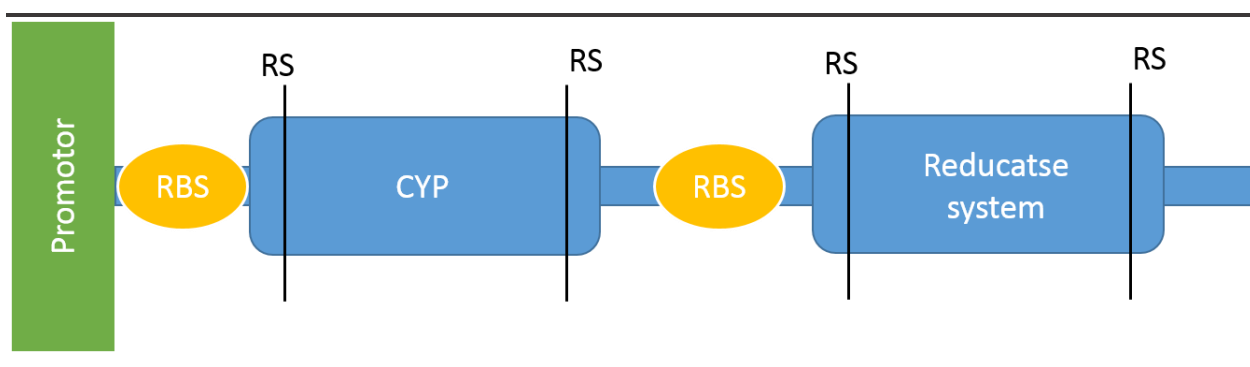
Figure 19: CYP-Swinger Operon (RBS = Ribosomale binding site; RS = restriction site)

The system was successfully tested by successful hydroxylation of TD[107]. Furthermore, we assembled the genetic sequences responsible for CO biosynthesis (CotB3, CotB4, Afr-Afx) and cloned them into the newly designed plasmid. This construct was functional as well, and we were able to produce Cyclooctat-9-en 5,7-diol[107]. Subsequently, we changed oxygenase and reductase partners in order to investigate if an electron transfer across different reductase systems is possible (see Figure 16).
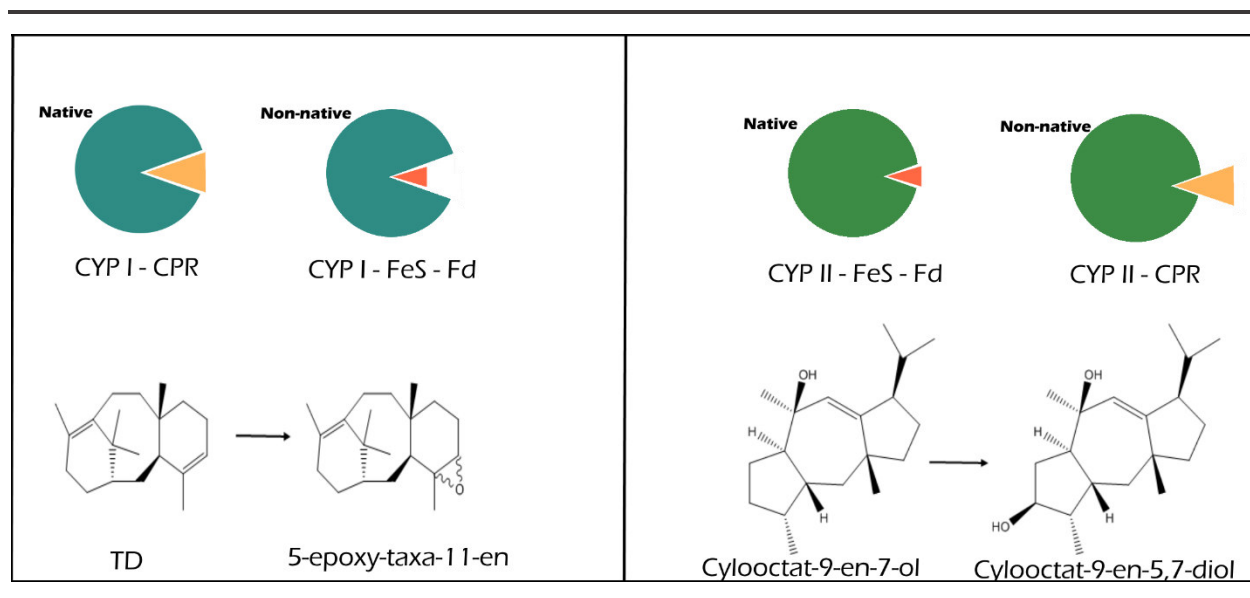
Figure 20: Experimental approach to test functionality of native and non-native CYP-reductase systems hydroxylating TD and Cylooctat-9-en-7-ol

Indeed, we were able to show similar activity for TD hydroxylation using the flavin dependent reductase system (Afr-Afx). Furthermore, implementation of CPR from *R. catharanthus,* for cylooctat-9-en-7-ol hydroxylation, was functional as well. Hence, we assumed that the electron transfer took place also in non-native CYP-reductase systems tested[107]. Even though electron transfer systems were changed, we assume that the non-native reducing equivalents were able to generate the ferryl-oxo species in the CYP, respectively. This transferability has not been demonstrated to date.

Moreover, we screened the CYP-reductase systems for their activity on non-native terpenoid scaffolds (TD, Cyclooctat-9-en-7-ol, Dolabellatriene, Verticilene – see Figure 17)[107].
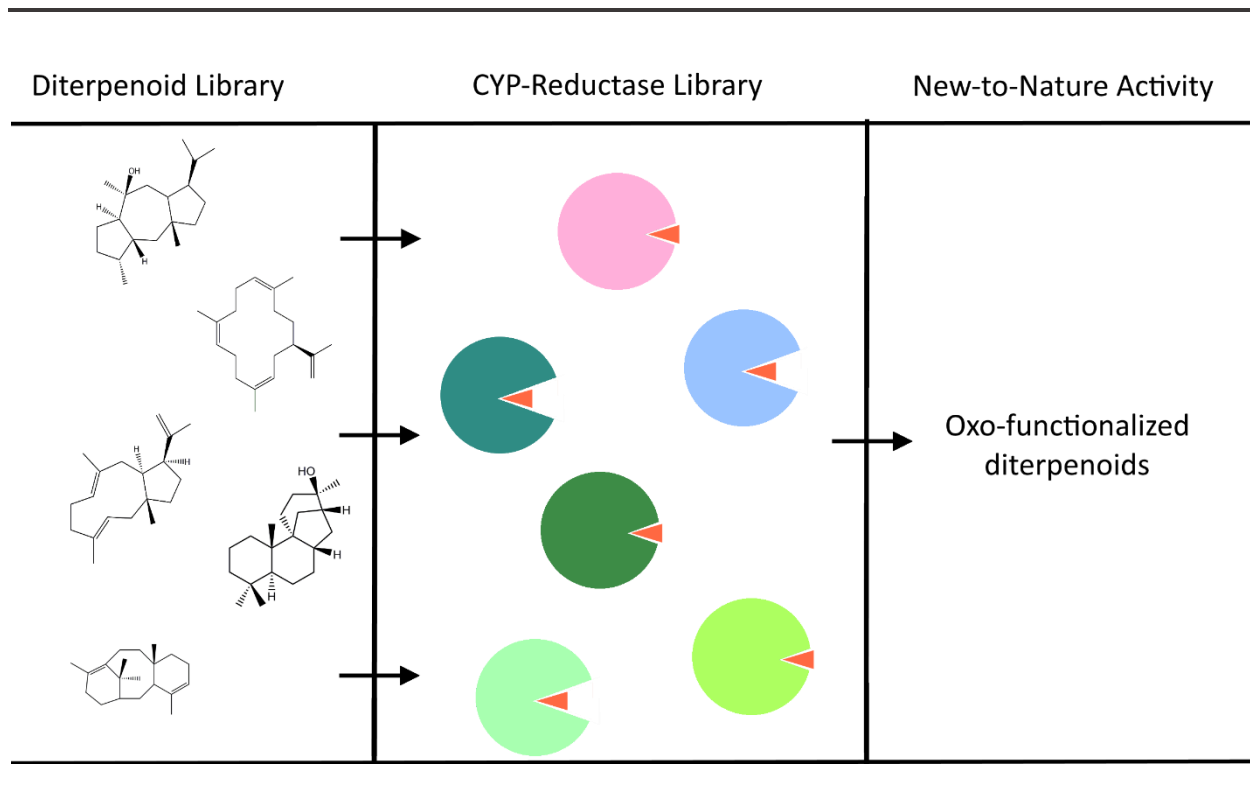
Figure 21: New-to-Nature experimental setting to identify CYP activity for non-native substrates

These new-to-nature combinations revealed measurable activity of CotB3 on all targets screened[107]. However, amounts produced were too low to structurally elucidate these compounds. To further improve and facilitate the screening, we investigated if it is possible to conduct a biotransformation of purified terpenes with engineered bacteria expressing CYP and reductase system[108]. In this context experiments, encompassing different temperatures, time points, solubilizing agents, etc. were conducted to identify an efficient biotransformation hydroxylation setting.

During the experiments we examined that the hydroxylation reaction primarily takes place in the stationary phase (Best performance was with stored, over-night cells at 4°C)[108]. Unfortunately, the designed plasmid encompasses constitutive promoters that expressed proteins during cell growth. Hence, we believe that changing promoter system to strong and inducible will increase the biotransformation rate substantially and allow for further improvement of the CYP-Swinger-System.

Protein Modelling

Homology Model and Docking – Non- expert level

In silico prediction of proteins is a common tool used in research. Modelling is predominantly used to identify binding sites and to identify lead structures that bind to a specific target (structure/activity

relationship-studies). Many tools have been developed to examine and predict the motion of proteins to improve computational methodologies in order to reduce experimental work[109].

Therefore, great interest lies in solving protein structures of pharmaceutical targets by crystallographic studies. Only few researchers have focused on solving terpene synthases' crystal structures so far as they do not represent a target for pharmaceutical applications[20]. However, crystal structure can also be used to study and understand reaction mechanism[55].

Interestingly, terpene synthase tertiary structure is highly conserved similar even though their sequence identity is often below 20%. Terpene synthases are identified by their conserved motifs DXDD and NSE plus DXXDD/E and are accordingly classified. Their structure consists of 13 alpha helices in a barrel structure[20]. This provides for good protein homology models, which can be solely constructed based on their primary sequence information. However, to date the functionality neither the product outcome can be forecasted.

Homology modelling can also be applied for the elucidation of terpene synthase reaction mechanisms[23, 55]. Docking the substrate into the active site of the proteins allows for rough estimation of amino acids that interact with the substrate during the cyclization reaction. Targeted amino acid substitution can stop the mechanism and an abortion product is released. This facilitates the evaluation of carbo cation intermediates on the way to final product. This is another strategy compared to labelling of the substrate. Additionally, protein mutants can provide for the production of commonly inaccessible terpene products.

There are many tools available to perform a protein modelling task. However, the vast variety and partly incompatible programs can easily stop all enthusiasm to continue the research in *in silico* protein modelling[26]. Prediction of a protein homology model is accessible by online tools (Expasy, RaptorX, Zen Lab). In this study RaptorX was used that provides a protein structure usually within 72 hours. The protein is predicted using crystal structure already available at the protein data base. However, cofactors like metals, which are catalytically relevant in terpene synthase reactions, are not yet included in the model. Hence, it is important to add these cofactors prior to any function-structure analyses.

The easiest way to implement the metal ions in the model is by the tool modeler [110]. For this tool a reference structure is required that already has the cofactors co-crystalized. A closely related reference structure to the expected homology model's activity, will return good results.

Chimera (UCSF) is a good program to start with [111]. Once the initial model is predicted and loaded in the software, the reference structure can be added as well, and subsequently be aligned. Starting the

add-on Modeler return a refined structure based on the reference structure and include the cofactors ad-hoc.

The next step is preparation of the enzyme for docking. Chimera has a ready to use tool that adds hydrogens, charge, etc. in a few clicks (AutoDock Prep)[112].

Afterwards the substrate must be prepared as well. Prior, the substrate must be produced and the molecule's charges annotated. There are several possibilities like:

-        Draw the chemical structure (ChemDraw, Avogadro, etc.)[113]

-        Or download the 3D structure (e.g. PubMed)

-        Or extract the ligand from a crystal structure (e.g. Chimera)

Afterwards, a geometrical optimization of the substrate should be performed. This can be done again by several tools (Chem3D, Avogadro, etc.). The substrate can subsequently be loaded into Chimera and should be saved as a mol2 file.

Implantation of the add-on AutoDock Vina to the Chimera software environment facilitates the docking procedure. After the tool has started, receptor and ligand can be defined. A box can be drawn around the active site that represents the docking zone for the substrate. After docking calculation and evaluation of the output, good results can be saved individually. The process flow is depicted in figure 22.
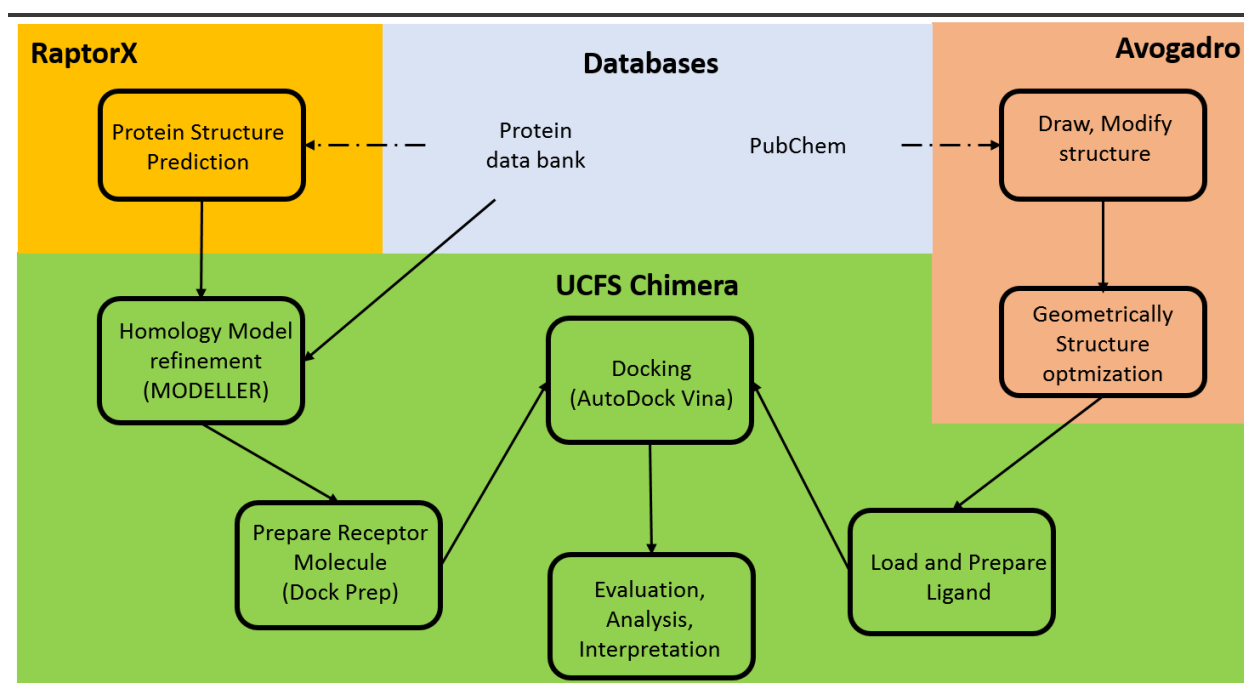
Figure 22: Process Flow for terpene synthase structure function analysis

Molecular Dynamics – Expert Level

In order to perform molecular dynamics studies, the docked enzyme substrate complex must be separated again into a receptor molecule and the ligand (saved as separate files (mol2 und pdb)).

The ligand atoms must be specified prior to simulation. This can be done by the software Yasara. After specification safe the ligand as pdb file. The online service paramchem allows for parametrization of the produced ligand[114]. The outcome is a str file that can be divided into a parameter (par) and topology file (top) by perl.exe running stream_split.pl (see appendix).

The receptor can be loaded in VMD. The commands in the appendix provides for generation of psf and pdb file for the receptor and the ligand, respectively.

After restarting VMD and opening the tool psf merge, ligand and receptor can be combined in a single file returning receptor-ligand.psf and receptor-ligand.pdb. Addition of a water sphere is conducted by another command listed in the appendix.

Subsequently, simulation.config, which includes all parameters necessary to run the simulation, allow for namd2 molecular dynamic simulation. Therefore, Windows CMD must be started and the directory changed to the folder containing Namd2.exe. The command "namd2.exe simulation.conf > simulation.log &" returns a dcd file and a log file. The dcd file can be loaded in VMD together with the respective pdb file. The receptor-ligand complex at different time points of the simulation are now presented (so called frames). Figure 23 demonstrates the process flow to run a molecular dynamic simulation of a protein-ligand complex.
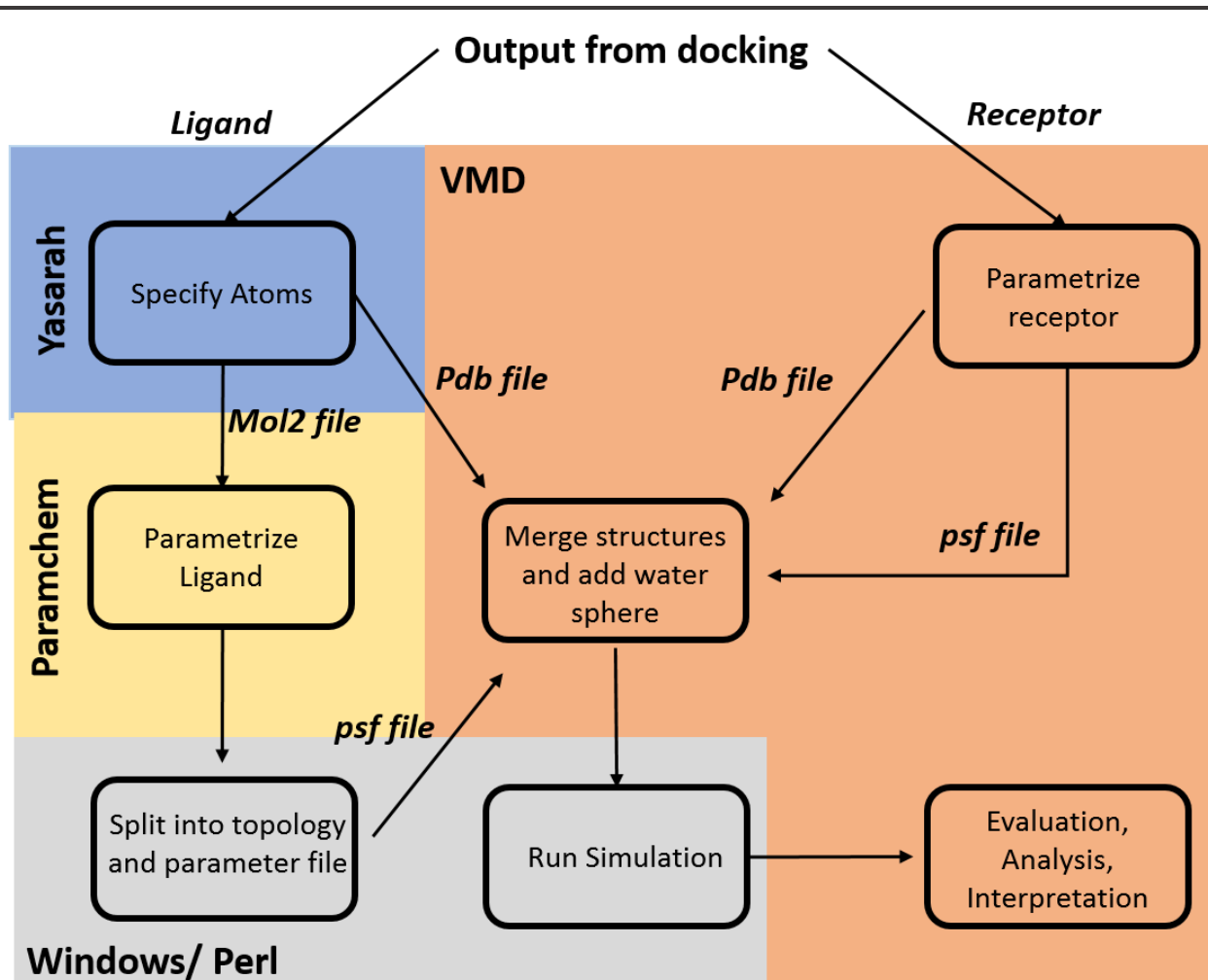
Figure 23: Process Flow for molecular dynamics studies of a protein-ligand complex

Part II

Publications

1

Insights into the Bifunctional Aphidicolan-16-ß-ol Synthase through Rapid Biomolecular Modeling Approaches

Synopsis

Virtual enzyme design and *in silico*-based structure function analyses of proteins are helpful tools for enzyme engineers. However, the numerous tools that are available, strongly complicates the choice which software to use and how to use it. Surely, there are powerful tools available that fulfill all requirements needed in a single program. However, they are very expensive and the complexity of ubiquitous software often requires solid expertise. In order to access *in silico* modeling for non-experts using free of charge software, we developed a strategy and a process flow for the structure function analyses of terpenoid synthases. Our initial goal was protein prediction, based on the primary protein sequence solely, homology model refinement and active site cleft investigations. In this context, a process flow for reliable *in silico* protein modeling was developed. Generated protein models were subsequently evaluated and protein mutants produced to experimentally prove our *in silico* results. The paper "Insights into the Bifunctional Aphidicolan-16-ß-ol Synthase through Rapid Biomolecular Modeling Approaches" describes the optimized procedure to examine a diterpenoid synthase *in silico*.

Initially, the bifunctional enzyme was predicted from the primary sequence. The predicted model was missing essential co-factors and the structure was built upon on an open protein crystal structure conformation. Terpene apo-protein crystal structures that are in open, catalytic inactive conformation represent the majority of published terpene synthases. The cyclization process of terpene synthases is initiated by a conformational change of the enzyme after the substrate is bound to cofactors and the enzyme's active site cleft. In this "closed conformation" the cascade of hydride shifts, carbon bond rearrangements and generation of cyclic molecule structures takes place. To that end, the predicted ACS homology model was refined in order to obtain a closed conformation enzyme. The refinement method included introduction and optimized positioning of cofactors generating a reliable model for further docking studies. CDP was docked in the refined homology model. Our findings were further compared to computational expensive molecular dynamics studies (NAMD) that revealed that our new protein structure function analyses approach produces similar results requiring less time (computational and preparation) and molecular dynamics experience (e.g. scripting, protein and ligand parametrization).

The protein-ligand complex was investigated and targeted alteration of the enzyme's active site cavity conducted. We demonstrated that the reaction from syn-CDP can be aborted prematurely providing access to the primaradiene and labdanoid diterpenoids (Y658, D661). This method provides for insights into molecular conformational changes of ACS and allows for a rapid and precise prediction of relevant amino acids involved in the cyclization reaction cascade.

Authors contribution

The research "Insights into the Bifunctional Aphidicolan-16-ß-ol Synthase through Rapid Biomolecular Modeling Approaches" was initiated by myself and supervised by Monika Fuchs and Thomas Brück. My work included the process development of the described protein structure function analyses approach and the evaluation of homology modeling, refinement and docking studies by molecular dynamics studies. I was supported by Nicolas Meese and Michael Mertz in the generation of protein mutants that were consequently screened in experimental settings. Fermentative production, analyses and structural elucidation of the terpenoids produced was conducted by myself.

Insights Into the Bifunctional Aphidicolan-16-ß-ol Synthase Through Rapid Biomolecular Modeling Approaches

Max Hirte, Nicolas Meese, Michael Mertz, Monika Fuchs* and Thomas B. Brück*

Werner Siemens Chair of Synthetic Biotechnology, Department of Chemistry, Technical University of Munich, Munich, Germany

Diterpene synthases catalyze complex, multi-step C-C coupling reactions thereby converting the universal, aliphatic precursor geranylgeranyl diphosphate into diverse olefinic macrocylces that form the basis for the structural diversity of the diterpene natural product family. Since catalytically relevant crystal structures of diterpene synthases are scarce, homology based biomolecular modeling techniques offer an alternative route to study the enzyme's reaction mechanism. However, precise identification of catalytically relevant amino acids is challenging since these models require careful preparation and refinement techniques prior to substrate docking studies. Targeted amino acid substitutions in this protein class can initiate premature quenching of the carbocation centered reaction cascade. The structural characterization of those alternative cyclization products allows for elucidation of the cyclization reaction cascade and provides a new source for complex macrocyclic synthons. In this study, new insights into structure and function of the fungal, bifunctional Aphidicolan-16-ß-ol synthase were achieved using a simplified biomolecular modeling strategy. The applied refinement methodologies could rapidly generate a reliable protein-ligand complex, which provides for an accurate *in silico* identification of catalytically relevant amino acids. Guided by our modeling data, ACS mutations lead to the identification of the catalytically relevant ACS amino acid network I626, T657, Y658, A786, F789, and Y923. Moreover, the ACS amino acid substitutions Y658L and D661A resulted in a premature termination of the cyclization reaction cascade *en-route* from syn-copalyl diphosphate to Aphidicolan-16-ß-ol. Both ACS mutants generated the diterpene macrocycle syn-copalol and a minor, non-hydroxylated labdane related diterpene, respectively. Our biomolecular modeling and mutational studies suggest that the ACS substrate cyclization occurs in a spatially restricted location of the enzyme's active site and that the geranylgeranyl diphosphate derived pyrophosphate moiety remains in the ACS active site thereby directing the cyclization process. Our cumulative data confirm that amino acids constituting the G-loop of diterpene synthases are involved in the open to the closed, catalytically active enzyme conformation. This study demonstrates that a simple and rapid biomolecular modeling procedure can predict catalytically relevant amino acids. The approach reduces computational and experimental screening efforts for diterpene synthase structure-function analyses.
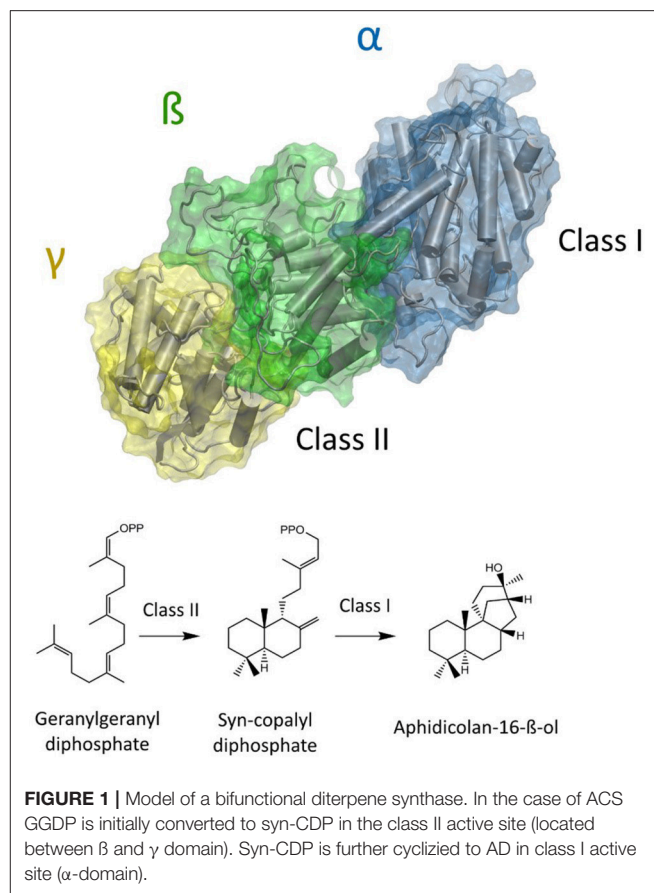
Keywords: homology modeling, aphidicolin, diterpene, diterpene synthase, homology model refinement

INTRODUCTION

With more than 50,000 different molecules known to date terpenes are the greatest natural occurring product family found in organisms from bacteria to fungi, mammals, and plants. They are all derived from the isoprene units' dimethylallyl diphosphate and isopentenyl diphosphate. Condensation reactions of this molecules lead to the formation of different length phosphorylated linear terpenes, serving as substrate for terpene synthases. This enzyme family carry out highly stereo complex C-C coupling reactions, resulting in structurally complex macrocycles that contribute to the structural and functional diversity of terpenes (Christianson, 2017). Diterpenes are derived from the linear aliphatic precursor geranylgeranyl diphosphate (GGDP) being cyclized by diterpene synthases. More specifically, diterpene synthases are classified into class I and class II enzymes based on the structural presence of the conserved motifs DDXD or DDXXD/E and NSE/DTE, respectively. While class II reactions perform a protonation initiated cyclization reaction to generate phosphorylated bicyclic structures, class I reactions are initiated by hydrolyses of the GGDP pyrophosphate moiety that is coordinated by a Mg^{2+}-triad thereby generating mono- or poly-cyclic structures.

The natural product Aphidicolin, initially isolated from the fungus *Cephalosporium aphidicola,* is a hydroxylated, tetracyclic diterpenoid that exhibits a broad range of biological activities and applications (Brundret et al., 1972; Dalziel et al., 1973). More specifically, it is a potent inhibitor of the eukaryotic DNA α-polymerase with a commercial application as a cell synchronization agent. The compound is in pharmaceutical development due anti-tumor, anti-viral, and anti-leishmanial activity (Ikegami et al., 1978; Pedrali-Noy et al., 1980; Kayser et al., 2001; Edwards et al., 2013; Starczewska et al., 2016). Recently, other organisms including the fungus *Nigrospora sphaerica* and the pathogenic fungus *Phoma betae* have been identified as natural Aphidicolin producers. Current data suggests that Aphidicolin biosynthesis is exclusive to fungal metabolism and that natural sources for Aphidicolin are limited (Starratt and Loschiavo, 1974; Fujii et al., 2011; Lopes and Pupo, 2011). Nevertheless, elucidation of the responsible Aphidicolin biosynthetic gene cluster in *P. betae* allowed for the identification of a bifunctional diterpene synthase that contains both a functional class I and class II domain (Oikawa et al., 2001). The Aphidicolan-16-ß-ol synthase (ACS) generates the stereo-chemically demanding Aphidicolan-16-ß-ol (AD)—core structure of Aphidicolin—structure via a two-step reaction as depicted in **Figure 1** (Oikawa et al., 2002).

Initially, GGDP is rearranged in the class II active site cleft by protonation to the bicyclic syn-copalyl diphosphate (syn-CDP). Subsequently, syn-CDP is elaborated to AD in the class I active site (Adams and Bu'Lock, 1975; Oikawa et al., 2002). As depicted in **Figure 2** the cyclization mechanism in the class I active site, initiated by the hydrolysis of the pyrophosphate group, results

---

Abbreviations: ACS, Aphidicolan-16-β-ol synthase; AD, Aphidicolan-16-β-ol; GGDP, geranylgeranyl diphosphate; LRS, labdane related diterpene synthase; syn-CDP, syn-copalyl diphosphate.

FIGURE 1 | Model of a bifunctional diterpene synthase. In the case of ACS GGDP is initially converted to syn-CDP in the class II active site (located between ß and γ domain). Syn-CDP is further cyclizied to AD in class I active site (α-domain).

in 8-ß-pimaradienyl carbocation formation. A subsequent attack of the vinyl group, bridging the C ring, directly undergoes a Wagner-Meerwein rearrangement and results in the formation of the aphidicolenyl carbocation. Eventually, this cation is quenched by water thereby generating AD.

Terpene cyclization mechanisms are conventionally elucidated by radio labeling of protons and carbons (Dickschat, 2017). This substrate specific labeling provides for identification of unusual hydride shifts and rearrangements. Alternatively, the enzyme's cyclization mechanisms can be probed by altering amino acids, trying to terminate the reaction cascade at a specific transition state (Morrone et al., 2008; Janke et al., 2014; Schrepfer et al., 2016; Jia et al., 2017). Therefore, random mutagenesis can be performed but the screening effort for this methodology is elaborate without an efficient high throughput screening options (Lauchli et al., 2013). Biomolecular modeling allows for the rational identification and *in silico* modulation of amino acid networks that are involved in complex reaction cascades (Pemberton et al., 2015; Schrepfer et al., 2016; Christianson, 2017; Escorcia et al., 2018). This methodology provides for a knowledge based approach of enzyme mutagenesis and screening. Nevertheless, a particular challenge for this strategy is based on the missing structural information for most terpene synthases. However, as their structural elements and domains are highly conserved (Christianson, 2017), homology

FIGURE 2 | Proposed cyclization mechanism of the ACS class I reaction (Oikawa et al., 2002).
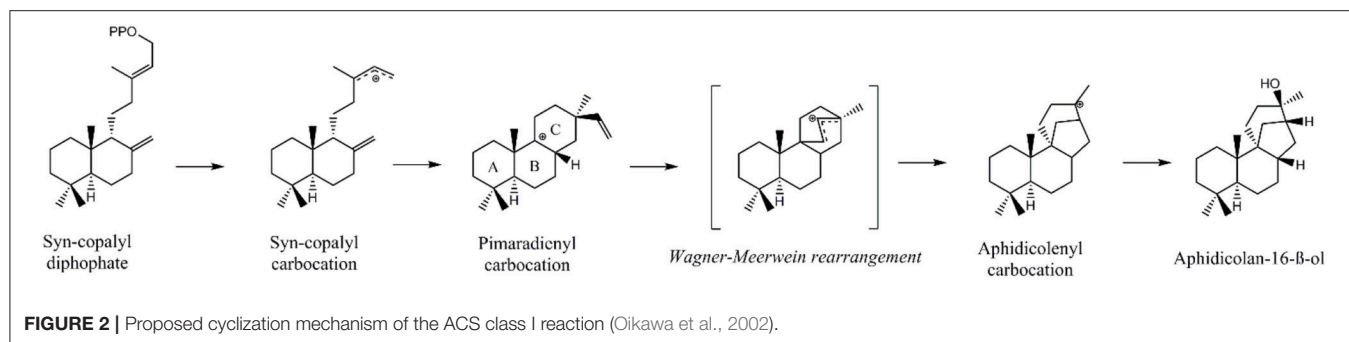
modeling is a potential route to identify catalytically relevant amino acids despite the low primary sequence identities in this enzyme family (Xu and Li, 2003). Unfortunately, most available crystal structures of terpene synthase are deposited in the open apo-enzyme configuration that is catalytically inactive. This open enzyme conformation presents an additional obstacle when catalytically relevant amino acids have to be identified *in silico*. At present, only two diterpene synthase structures have been reported in the closed, catalytically active form (Liu et al., 2014; Serrano-Posada et al., 2015). Therefore, automated homology modeling approaches will almost always result in catalytically non-relevant open enzyme configuration. Moreover, while prediction tools can place large cofactors (i.e., FAD, NADH, Heme) correctly in the apo-protein framework, ligand-metal interactions are difficult to predict because of the multiple coordination geometries and the lack of sufficiently accurate force field parameters (Khandelwal et al., 2005). Hence, structure function predictions that depend on the interplay between the amino acids of the protein framework with small metal ions cannot be conducted solely by application of automated software tools. In this context, a rational combination of structural information by superposition and extraction of cofactors is performed to prepare the protein structure for docking studies. Nevertheless, this approach often neglects reliable positioning of the cofactor coordinating amino acids. Additionally, falsely predicted positioning of amino acid side chains in the active site cleft can lead to invalid interpretation of a homology model based protein-ligand complex. To improve this situation, this study elucidated rapid and simple methodologies to refine diterpene homology models for docking studies thereby allowing for reliable structure-function predictions. In this context, an ACS class I homology model of the α-domain was predicted from the primary sequence. Subsequently, these models were compared to catalytically relevant closed terpene synthases structures. The location of metals was refined and fitted against specifically selected structural templates and multiple docking studies were carried out and validated. Our *in silico* results were experimentally evaluated by ACS mutagenesis studies. This lead to an identification of essential amino acid residue sidechains that are necessary for retaining the enzymes activity. Additionally, we detected amino acid substitutions that abort the catalytic reaction cascade *en- route* from syn-CDP to AD. Structural analyses and elucidation of these compounds

revealed the formation of syn-copalol and a labdane related, non-hydroxylated diterpene by the ACS mutants Y658L and D661A. Our approach of a protein homology model based structure function analysis can be easily adapted for other terpene synthases. This methodology allows for rapid and simple analysis of the catalytically relevant amino acid network that help studying complex reaction cascades and developing new biocatalysts.

MATERIALS AND METHODS

Materials and Chemicals

All genes used were synthesized by Life technologies GmbH and the codon usage was optimized for *E. coli* if not stated otherwise. Primers were obtained from Eurofins Genomics GmbH. Strains and plasmids were obtained from Merck KGaA. All chemicals used were obtained at highest purity from Roth chemicals or Applichem GmbH. Enzymes were purchased from Thermo Fisher Scientific.

Software and Web-Tools

RaptorX was applied for homology modeling studies (http://raptorx.uchicago.edu; Källberg et al., 2012). The initial predicted structure was analyzed and further modified in the environment of UCSF Chimera software package (Pettersen et al., 2004; http://www.cgl.ucsf.edu/chimera). Comparative modeling by spatial restraints was performed by MODELLER (Eswar et al., 2006), and all substrate docking studies performed by AutoDock Vina (Trott and Olson, 2010; http://vina.scripps.edu). Chemical structures were drawn by PerkinElmer ChemBioDraw Ultra (http://www.cambridgesoft.com). For ligand preparation the Avogadro (Hanwell et al., 2012; https://avogadro.cc/) software package was used. A syn-CDP toppar stream file was generated by CHARMM General Force Field program version 1.0.0 for use with CGenFF version 3.0.1 (https://cgenff.paramchem.org; Vanommeslaeghe et al., 2010, 2012; Vanommeslaeghe and MacKerell, 2012). Two ns molecular dynamic studies of the docked ACS model B in a water sphere have been performed under CHARMM general force field by NAMD (Phillips et al., 2005; http://www.ks.uiuc.edu/Research/namd/). NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign. For high resolution pictures the protein was prepared by Visual Molecular Dynamics (http://www.ks.

uiuc.edu/Research/vmd/; Humphrey et al., 1996) and rendered by Tachyon implemented in the VMD software package (Stone, 1998).

Docking

Ligand structures were downloaded from https://pubchem.ncbi.nlm.nih.gov/ available and geometrically optimized by 500 steps of steepest descent under MMFF94 force field parameters included in Avogadro. Protein structures were prepared by Dock Prep, which is part of the Chimera software environment. The AMBER force field (AMBERff14SB) was applied to the receptor while Gasteiger charges were added to the ligand and co-factors. As recently reported, docking can be improved by assigning partial charges to metal ions (Hu and Shelver, 2003). In this context, Mg-ion charges were set to $+1$. Syn-CDP charge was set to -3. Docking was performed by AutoDock Vina using standard parameters. Docking poses were chosen based on a structural comparison to the pyrophosphate group that is co-crystallized in pdb 5A0J (see Figure S1B). The chosen pose was furthermore validated by re-dock approaches. Therefore, the predicted syn-CDP pose was *de novo* geometrically optimized by 500 steps of steepest descent under MMFF94 force field parameters included in Avogadro software environment prior to docking repetition (see Figure S1A).

Model Generation

An initial homology model of the ACS α-domain was predicted by RaptorX starting from the amino acid 565. A model based on the pdb crystal structure 5A0J, referring to a labdane related diterpene synthase, was manually selected for further structure function analyses. In order to prepare the model for docking studies, the coordinating Mg^{2+}-ion triad and water molecules were implemented in the structure by different methods. Model A was generated by structural alignment to 5A0J. Cofactor positions were transferred from the structure template to Model A without any further adjustment prior to docking studies. Model B was created by MODELLER implemented in the Chimera software environment using the 5A0J as template structure. In this model hetero atoms and water molecules in the structure environment were computationally implemented. The pyrophosphate group was removed prior to docking with syn-CDP. Model C was prepared analogously to Model B but prior to refinement by MODELLER, syn-CDP was docked into the template structure 5A0J.

Model Validation

The protein ligand complex of Model B was validated by molecular dynamics studies. Therefore, syn-CDP was initially extracted from Model B and parameterized by CHARMM General Force Field program version 1.0.0 for use with CGenFF version 3.0.1. VMD was used to parameterize the protein and for merging ligand and protein. Subsequently, a water sphere was added around the protein-ligand complex. Two nanoseconds of molecular dynamic studies under CHARMM General Force Field was applied to the protein complex by NAMD. The calculated rmsd of the generated frames was plotted over time (Figure S2). A constant rmsd value was chosen as the criteria for an equilibrated

protein-ligand complex. The last frame obtained was compared to the initial model B (Figure S3).

Plasmids for Diterpene Production

For all cloning procedures *E. coli* HMS 174 (DE3) was used. Clones were cultivated at 37°C in Luria-Bertani (LB) medium. Chloramphenicol (34 μg/L) and Kanamycin (50 μg/L) were added as required. For efficient production of the diterpene AD, *E. coli's* internal 1-deoxy-xylulose-5 phosphate pathway flux was increased by overexpression of deoxy-xylulose 5 phosphate synthase (dxs: GenBank: YP001461602.1), isopentenyl-diphosphate delta isomerase (idi: GenBank: AAC32208.1), and further extended by expressing geranylgeranyl diphosphate synthase (crtE: GenBank: KPA04564.1) and Aphidicolan-16-ß-ol synthase (acs: GenBank: AB049075.1). Therefore, dxs and acs were amplified from original sources by PCR. Polycistronic operons (**Table 1**) were constructed by BioBrick cloning standard (Shetty et al., 2008).

Site directed mutations of acs were generated by PCR. Forward primers were designed exhibiting the respective mutation at the 5′ end while the corresponding reverse primers were phosphorylated at 5′ end (Table S1). PCR products were ligated by T4 Ligase prior to transformation. All amino acid exchanges were confirmed by sequencing.

Production of Diterpenes

All diterpene production experiments were performed in *E. coli* BL 21 (DE3). To investigate the product outcome of ACS mutants, pACYC acs plasmids were co-transformed with pAX dic. Cultivation was performed in minimal media supplemented with 6 g/L yeast extract and 30 g/L glycerol at 25°C. After 60 h the culture was extracted with a mixture of hexane, ethanol and ethyl acetate (1:1:1) (v/v/v) for 1 h. The extract was centrifuged at 10,000 g for 2 min. The upper, organic phase was directly analyzed for diterpene products via GC-MS.

Diterpene Analytics

GC-MS analyses of diterpenes was performed by a Trace GC Ultra with DSQII (Thermo Fisher Scientific). Therefore, 1 μL sample was loaded (Split 1/10) by TriPlus AS onto a SGE BPX5 column (30 m, I.D 0.25 mm, Film 0.25 μm). The initial column temperature was set to 160°C and maintained for 5 min before a temperature gradient at 8°C/min up to 320°C was applied. The final temperature was kept for additional 3 min. MS data were recorded at 70 eV (EI) and m/z (rel. intensity in %) as total ion current (TIC). The recorded m/z range was in between 50 to 650.

NMR spectra were recorded in $CDCl_3$ with an Avance III 500 MHz (Bruker) at 300 K. ^{1}H NMR chemical shifts are given in ppm relative to $CDCl_3$ (δ = 7.26 ppm). The 2D experiments

---

TABLE 1 | Plasmids used for AD production in *E. coli*.

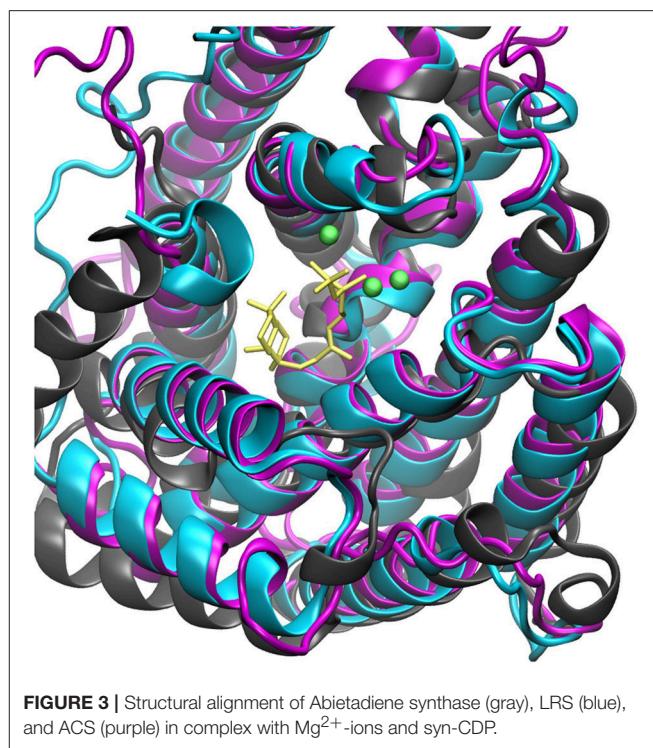| Name	Promotor strength	Genes	Resistance	vector
pAX dic	Weak	dxs, idi, crtE	Kanamycin	pBR322
pACYC acs	Strong	acs	Chloramphenicol	pACYC duet

(HSQC) were performed using standard Bruker pulse sequences and parameters.

RESULTS AND DISCUSSION

Homology Model Refinement

The steady increase in published protein crystal structures provides for an accelerated improvement of computational homology prediction. Especially due to the high structurally conservation of the terpene synthase enzyme families, biomolecular tools can predict structures solely based on the amino acid sequence. In this context, structure prediction of the bifunctional ACS was performed to analyze the highly complex conversion of GGDP via syn-CDP to the tetracyclic AD which is the core structure of the cytostatic compound Aphidicolin. ACS belongs to the diterpene synthase family and we identified three highly structurally conserved domains. The initial conversion from the universal diterpene precursor GGDP to syn-CDP occurs in class II active site, located between the ACS ß- and γ-domain. The subsequent syn-CDP cyclization to AD is then conducted in the class I active site that is positioned in the middle of an α-helical bundle forming the ACS α-domain. Notably, the fungal ACS is structurally highly similar to the previously crystallized plant diterpene synthases Abietadiene (pdb: 3S9V) and Taxadiene synthase (pdb: 3P5R), respectively. Homology prediction based on the full ACS sequence took those two structures into account, but for both, crystals could only be achieved in N-terminal truncated forms. Furthermore, these crystal structures have only been solved in an open conformation that is catalytically inactive. In order to circumvent the consideration of these catalytically inactive templates, only the ACS α-domain sequence was used for homology prediction. A model based on the labdane related diterpene synthases (LRS) (pdb: 5A0J), which is provided in a catalytically active holo-complex (Serrano-Posada et al., 2015), was selected for ACS homology refinement. The structural superposition of Abietadiene (pdb: 3S9V), LRS (pdb: 5A0J), and the ACS model, as depicted in **Figure 3**, explicitly demonstrates that there is a better fit between the ACS model and the LRS crystal structure. While the structural fit between LRS and the ACS model is visually well apparent, we have not calculated an rmsd value qualifier as structural domains that do not constitute the active site region are highly variable.

Co-crystallized cofactors (Mg^{2+}-ion triad) and waters, both provided in the LRS structure, are also involved in the ACS reaction *en-route* from syn-CDP to AD. Therefore, we differentially adapted both, the positions of the Mg^{2+} ions and waters into the ACS models that resulted in the generation of three ACS models (A–C). Model A was prepared by adaptation of cofactor positions from the template structure LRS after structural alignment. Initial evaluation of this model indicated that this un-refined modeling method results in clashes of cofactors positions with amino acids side chains. Generally, in homology prediction the active site's cavity is not reserved for the substrate or cofactors specifically. Therefore, we presume that amino acid sidechains occupy this free space due to applied energy minimization optimizations. This is demonstrated in

FIGURE 3 | Structural alignment of Abietadiene synthase (gray), LRS (blue), and ACS (purple) in complex with Mg^{2+}-ions and syn-CDP.

our docking studies of model A, where ACS amino acid Y658 is preventing syn-CDP to completely access the active site cavity. With the MODELLER package, which is based on comparative protein structure modeling by spatial restraints, a protein structure can be refined based on a template structure. Additionally, hetero-atoms and water molecules can be included directly in the model refinement. This refinement methodology applied to our initial model structure lead to the generation of ACS model B. This model B computationally included the three Mg^{2+}-ions, a pyrophosphate group (conventionally derived by Mg^{2+} based hydrolysis of the phosphorylated substrate [syn-CDP] substrate) and water molecules directly as they are all present in the LRS template structure. Model B provides reliable positioning of the conserved amino acids that constitute the class I diterpene synthase signature DDXXD/E and NSE/DTE motifs in relation to the adapted Mg^{2+}- ions, water and pyrophosphate moieties, respectively. Subsequently, we removed the pyrophosphate group from the model B structure to enable docking with the native syn-CDP substrate. Our docking data indicated that in Model B syn-CDP can completely access the active site's cavity. A specific syn-CDP conformation was selected pointing toward the ACS G-Helix, as this flexible helix is proposed to be involved in terpene cyclization reactions (Yoshikuni et al., 2006; Baer et al., 2014; Jia et al., 2017). This docking pose was validated by multiple re-docking approaches (Figure S1A). Additionally, we validated the pose while the position of the pyrophosphate moiety was compared to the pyrophosphate group co-crystallized in LRS (Figure S1B). Finally, a third approach for structure-function analyses was performed by docking syn-CDP into LRS prior

to ACS refinement with MODELLER. Again, a syn-CDP conformation was chosen with close proximity toward the G-Helix. On the basis of this LRS holo-protein complex, an ACS holo-complex model C was generated. This method provided for a protein model that was refined around the substrate and cofactors. This methodology also provided for a precise specification of amino acids involved in the AD cyclization reaction. For all three models amino acids located within a five Å vicinity to the docked substrate syn-CDP (thereby neglecting the pyrophosphate moiety) were analyzed by mutational studies to elucidate their catalytic relevance (see **Figure 4**, Table S2).
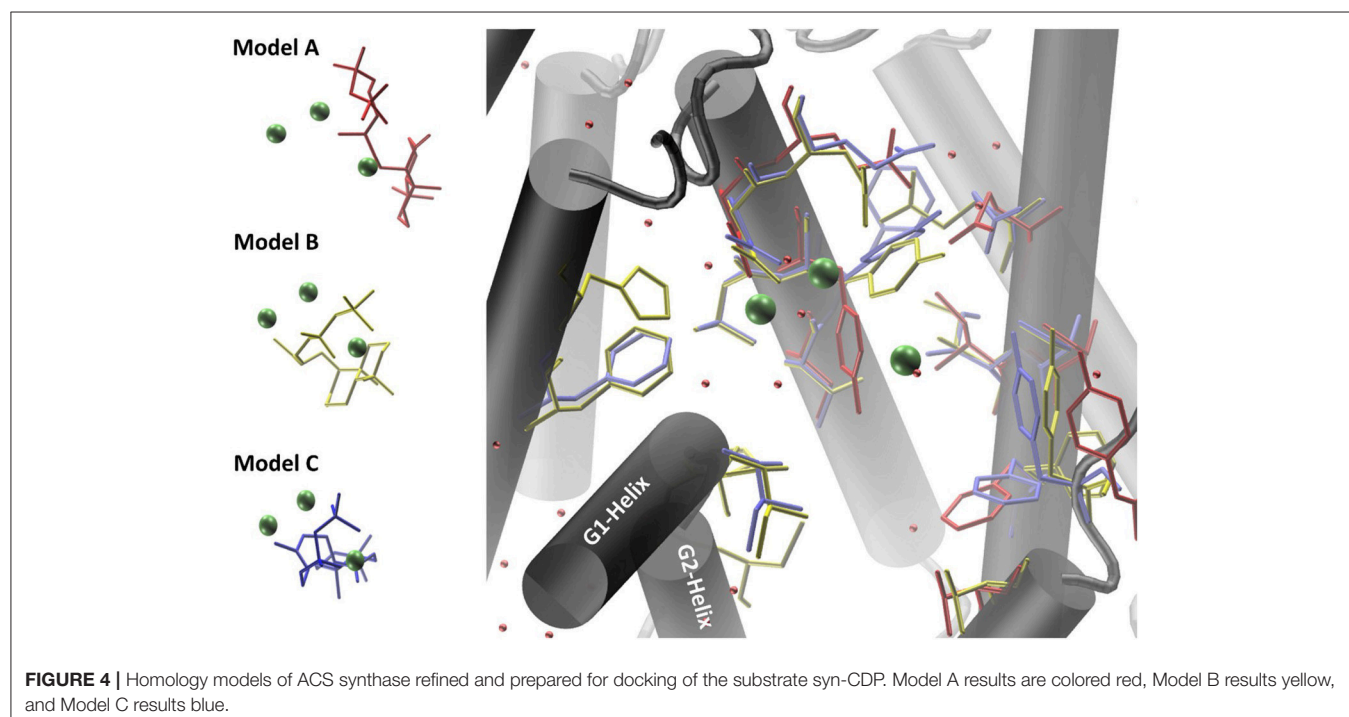
Mutational Validation of Catalytically Relevant ACS Amino Acids

Due to their stereo-chemical diversity, natural diterpene scaffolds are attractive research leads. The enormous stereo-chemical demand of diterpene macrocycles renders them difficult to access via total chemical synthesis approaches. Therefore, biosynthetic routes to generate these complex structures are currently an intense research focus (Dickschat, 2016; Bian et al., 2017; Jones, 2017). The ability to access new diterpene macrocycles via selective alteration of amino acids in diterpene synthases provides for a highly varied accessible chemical space. For the class I cyclooct-9-en-7-ol synthase, which naturally generates a tricyclic fusicoccin type diterpene, amino acid mutations in the vicinity of the active site lead to intermittent abortion of the reaction cascade. Hence, alternative macrocyclic structures, such as the bicyclic dolabellane and the monocyclic cembrane, could be generated thereby elucidating the reaction cascade (Görner et al., 2013; Janke et al., 2014). In this study, insights into the class I reaction of the ACS were achieved

by mutational studies. In that respect, we intended to quench the reaction from syn-CDP to AD at previously proposed transitional states (Adams and Bu'Lock, 1975; Oikawa et al., 2002). Based on the proposed ACS transitional states we presume that syn-labdatriene and syn-copalol (termination product of the syn-copalyl carbocation), stereoisomers of syn-pimaradiene (termination products of the pimaradienyl carbocation), or aphidicolene and stemodene (termination products of the aphidicolenyl carbocation) are potential abortion products (see **Figure 5**).

For an intermittent abortion of the reaction cascade from syn-CDP to AD, we have selected amino acids within a range of five Å to the docked ligands as prime targets for mutagenesis (see **Figure 4**). Preliminary studies revealed that sidechain substitutions encompassing amino acids exchanges that inherently change physico-chemical properties frequently resulted in inactive enzyme variants (Janke et al., 2014; Schrepfer et al., 2016). In this context, we focused on changing the size of the respective amino acid sidechain thereby trying to preserve physico-chemical characteristics. Alternatively, we chose amino acid side chain substitutions that would replace polar groups with similar size amino acids (Table S2).

ACS syn-CDP docking results pointed toward a strong interaction between the decalin core and surrounding hydrophobic sidechains. However, as the decalin structure of syn-CDP remains untouched in further cyclization steps most of the implemented mutations near this particular moiety resulted in inactive (I626A, Y923L, F789L) or wildtype activity variants (F629L, Y658F, C831G, C831T, T920G, Y923F). Based on our modeling results, we also identified specific amino acids located in the ACS G-helix that in other studies have been proposed to be of catalytic relevance (Baer et al., 2014;

FIGURE 4 | Homology models of ACS synthase refined and prepared for docking of the substrate syn-CDP. Model A results are colored red, Model B results yellow, and Model C results blue.

FIGURE 5 | Expected products generated by ACS if the reaction cascade *en-route* from syn-CDP to AD is prematurely terminated.
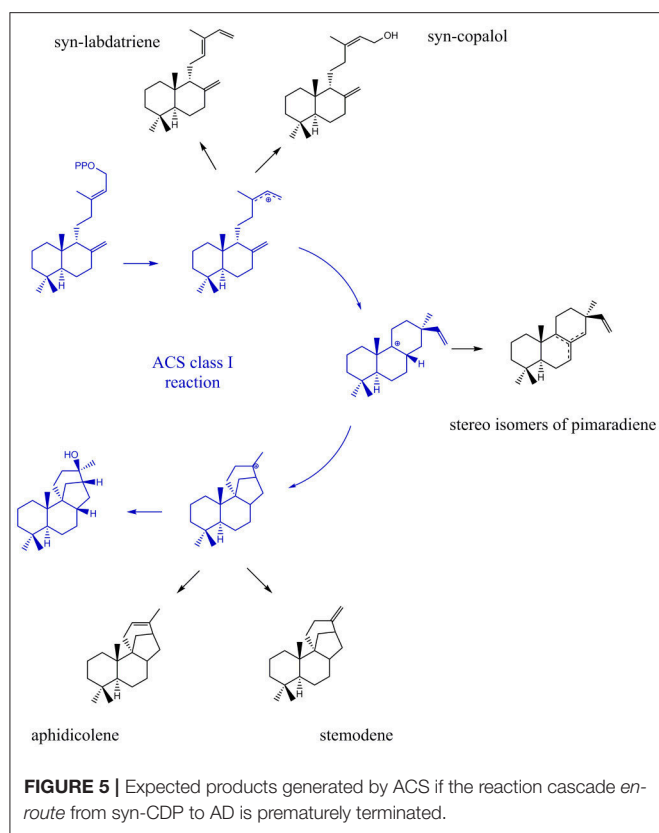
Jia et al., 2017). While mutational changes in the G-Helix of Kaurene synthase like diterpene synthases resulted in alternative product profiles (Jia et al., 2017), our analogous approaches with ACS only provided inactive (A786L, F789L) or wildtype active (A786G, F789Y) variants. Nevertheless, our results support previous findings that propose the G-Helix as an essential flexible motif which is involved in the catalytically relevant structural change from the open to the closed enzyme configuration (Baer et al., 2014).

Only the substitution of ACS Y658L and D661A provided for a varied product outcome. In addition to amino acids that constitute the DXXDD/E and NSE/DTE signature motifs that are responsible for Mg^{2+}-ion coordination, our combined *in silico* and experimental study identified only a few amino acids (see **Figure 6**, colored in pink) capable to terminate activity. Our successful mutations (D661A, Y658L) indicated that the unusual cyclization from syn-CDP to AD proceeds in a spatially restricted area of the active site's cleft. Additionally, our data suggests that the pyrophosphate group remains in the active site and coordinates the reaction cascade. This is in accordance to the recently postulated Taxadiene synthase reaction mechanism (Schrepfer et al., 2016).

ACS Mutants D661A and Y658L

GC-MS analyses of the ACS mutants Y658L and D661A revealed that this mutations lead to the formation of two unknown diterpene products (see **Figure 7**). In contrast to the native AD,

which had a GC retention time of 17.67 min, these new diterpenes had a retention time of 12.79 and 13.46 min, respectively. The latter product with a retention time of 13.46 min, showed a total mass of 290 m/z. Comparison of the MS spectral data suggests that this was a hydroxylated diterpene with a similar structure to syn-copalol (Hoshino et al., 2011). Subsequently, this compound was isolated and structural characterized by NMR (Figures S4, S5). The results are in accordance to previous spectral data for syn-copalol (Yee and Coates, 1992). One plausible explanation for syn-copalol formation is the quenching of the syn-copalyl carbocation intermediate by water in the active site of the enzyme. The other diterpene product with a retention time of 12.79 min had a total mass of 272 m/z indicating that this structure was not-hydroxylated. While we expected the formation of syn-labda-8(17),12E,14-triene, comparison with published MS-spectra revealed significant differences (Morrone et al., 2011). Unfortunately, due to the low amounts produced and purification issues for this highly hydrophobic compound, we could not conduct NMR analysis. However, we presume that this compound is also originated from the syn-copalyl carbocation and that a labdane related diterpene with high structural similarity to syn-labda-8(17),12E,14-triene was generated by the ACS mutants. The newly generated diterpenes are of great interest as copalol derivatives display various biological activities analogous to aphidicolin (Hanson, 2015).

The structural changes (D661A and Y658L) still allowed syn-CDP binding in the active site with subsequent hydrolyses of the pyrophosphate group. The syn-copalyl carbocation was then quenched either by water (release of syn-copalol) or an amino acid side chain (release of non-hydroxylated diterpene). Furthermore, as we did not find other substitution that stopped cyclization at the proposed transitional states and as we could not even detect changes in the byproduct formation of the active mutants, we presume that the ACS cyclization occurs in a spatially restricted area and that the pyrophosphate group remains in the active site, which is in accordance to recent reports (Schrepfer et al., 2016).

Former diterpene centered production processes were limited by low target compound yields. However, optimization of recombinant diterpene production hosts has extensively progressed to provide gram per liter yields (Ajikumar et al., 2010; Schalk et al., 2012). Today, access to novel diterpene lead structures is limited by the effective identification of relevant enzyme systems from large scale genome sequencing projects. Therefore, rational alteration of known terpene synthase product profiles by using a combination of *in silico* prediction and knowledge based mutagenesis studies can allow for a more rapid and targeted expansion of the desired chemical space.

CONCLUSION

A model of ACS synthase was computed that required the application of various methods for model refinement to improve the quality of *in silico* structure function analysis. A model of the
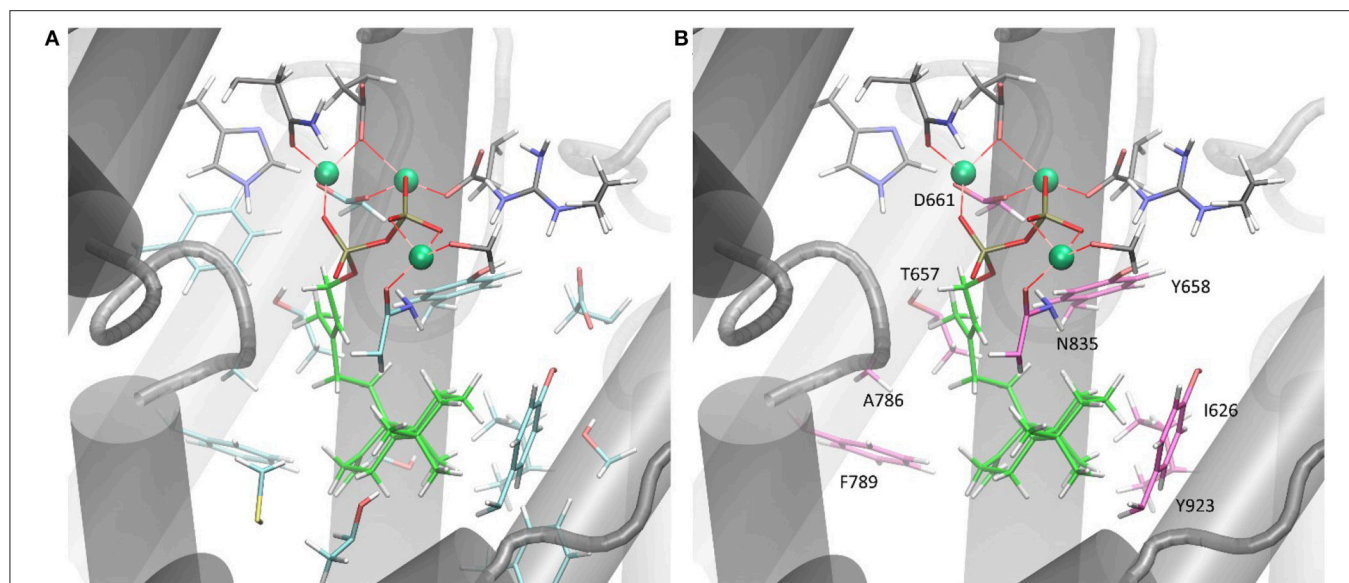
FIGURE 6 | (A) ACS active sites cleft in complex with Mg^{2+} and syn-CDP. Amino acid network within five Å to syn-CDP are displayed. **(B)** ACS active sites cleft in complex with Mg^{2+} and syn-CDP. Substitution of labeled amino acid (displayed in pink) resulted in inactive enzyme versions or mutants with altered product outcome.

FIGURE 7 | Analysis of ACS wildtype and ACS D661A mutant product outcome by GC. The MS-patterns for syn-labdatriene, syn-copalol and AD (from right to left) are presented below.
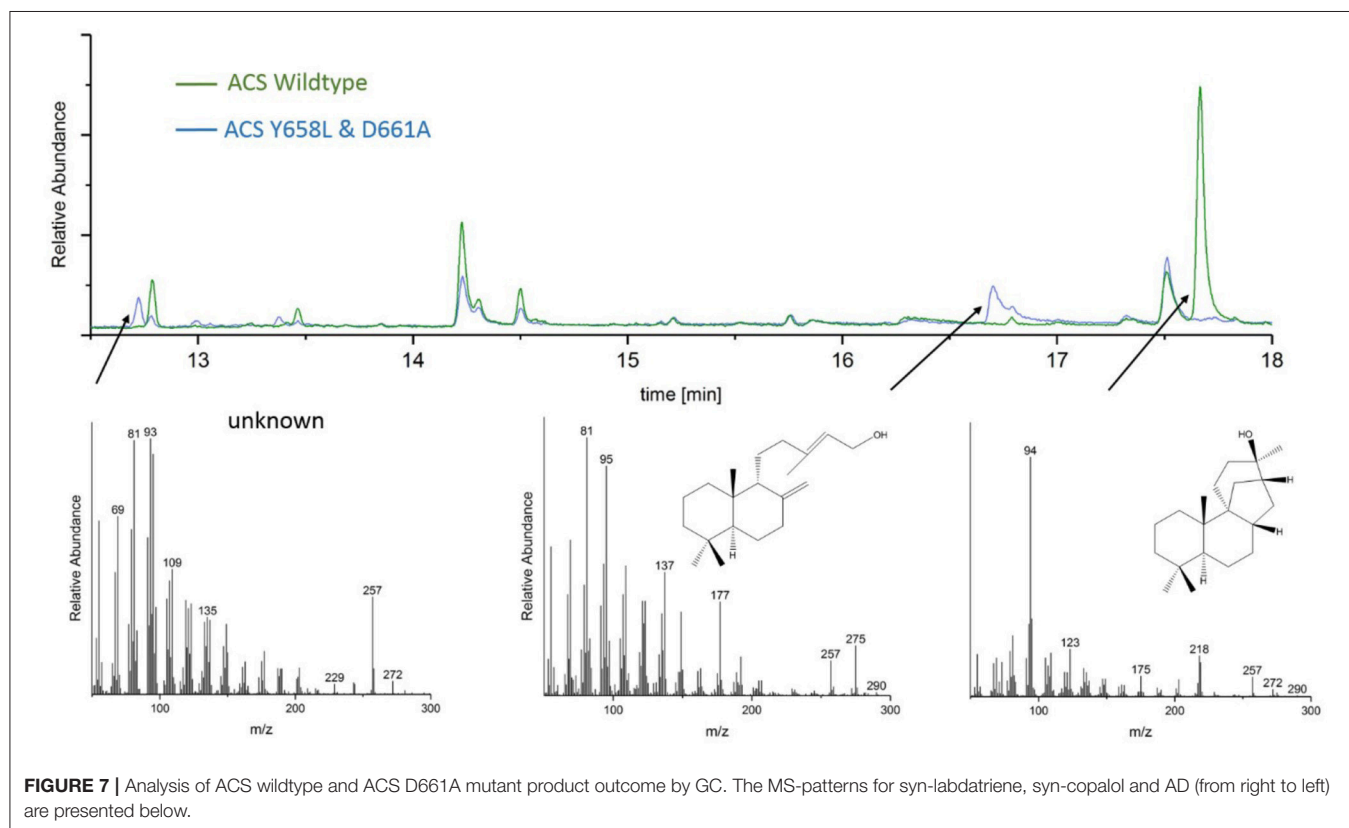
catalytically active, closed ACS α-domain complex was generated. Examination of this model provided for the identification of catalytically active amino acid sidechains. The *in silico* results were confirmed by mutational studies of the ACS. The amino

acid substitutions Y658L and D661A in the vicinity of the ACS active site lead to formation of the alternative cyclization products syn-copalol and a minor labdane related diterpene. Formation of these products were delineated by quenching of

the syn-copalyl carbocation *en-route* to AD. Additional mutants leading to inactive enzyme variants (A786L, F789L) provided insights into catalytically relevant amino acid residues within the G-Helix. The cumulative *in-silico* and experimental data suggests that amino acids constituting the G-loop motif of class I terpene cyclases are involved in the transformation of the open to the closed, catalytically active enzyme conformation. Moreover, as we only obtained a limited number of alternative cyclization products in our mutational screens, we presume that AD formation occurs in a rather confined location of the ACS active site. With respect to our biomolecular modeling approaches, we demonstrated that application of simple and rapid computational methodologies can be employed for prediction and structure function analyses of class I diterpene synthases.

AUTHOR CONTRIBUTIONS

TB and MF supervised this study. MH initiated this study and performed virtual modeling and docking studies. NM and MM conducted mutagenesis experiments and screening under supervision of MH and MF. Data was analyzed by MH, MF, NM, MM, and TB. All figures were created by MH. All authors verified the data, contributed to the manuscript, and approved the final version.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2018.00101/full#supplementary-material

REFERENCES

Adams, M. R., and Bu'Lock, J. D. (1975). Biosynthesis of the diterpene antibiotic, aphidicolin, by radioisotope and ^{13}C nuclear magnetic resonance methods. *J. Chem. Soc. Chem. Commun.* 389–391. doi: 10.1039/c39750000389

Ajikumar, P. K., Xiao, W.-H., Tyo, K. E., Wang, Y., Simeon, F., Leonard, E., et al. (2010). Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. *Science* 330, 70–74. doi: 10.1126/science.1191652

Baer, P., Rabe, P., Fischer, K., Citron, C. A., Klapschinski, T. A., Groll, M., et al. (2014). Induced-fit mechanism in class I terpene cyclases. *Angew. Chem. Int. Edn.* 53, 7652–7656. doi: 10.1002/anie.201403648

Bian, G., Deng, Z., and Liu, T. (2017). Strategies for terpenoid overproduction and new terpenoid discovery. *Curr. Opin. Biotechnol.* 48, 234–241. doi: 10.1016/j.copbio.2017.07.002

Brundret, K. M., Dalziel, W., Hesp, B., Jarvis, J. A. J., and Neidle, S. (1972). X-Ray crystallographic determination of the structure of the antibiotic aphidicolin: a tetracyclic diterpenoid containing a new ring system. *J. Chem. Soc. Chem. Commun.* 1027–1028. doi: 10.1039/c39720001027

Christianson, D. W. (2017). Structural and chemical biology of terpenoid cyclases. *Chem. Rev.* 117, 11570–11648. doi: 10.1021/acs.chemrev.7b00287

Dalziel, W., Hesp, B., Stevenson, K. M., and Jarvis, J. A. J. (1973). The structure and absolute configuration of the antibiotic aphidicolin: a tetracyclic diterpenoid containing a new ring system. *J. Chem. Soc. Perkin Trans.* 1, 2841–2851. doi: 10.1039/p19730002841

Dickschat, J. S. (2016). Bacterial terpene cyclases. *Nat. Prod. Rep.* 33, 87–110. doi: 10.1039/C5NP00102A

Dickschat, J. S. (2017). Modern aspects of isotopic labellings in terpene biosynthesis. *Eur. J. Org. Chem.* 2017, 4872–4882. doi: 10.1002/ejoc.201700482

Edwards, T. G., Helmus, M. J., Koeller, K., Bashkin, J. K., and Fisher, C. (2013). Human papillomavirus episome stability is reduced by aphidicolin and controlled by DNA Damage response pathways. *J. Virol.* 87, 3979–3989. doi: 10.1128/JVI.03473-12

Escorcia, A. M., van Rijn, J. P. M., Cheng, G.-J., Schrepfer, P., Brück, T. B., and Thiel, W. (2018). Molecular dynamics study of taxadiene synthase catalysis. *J. Comput. Chem.* doi: 10.1002/jcc.25184. [Epub ahead of print].

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M.-Y., et al. (2006). Comparative protein structure modeling using modeller. *Curr. Protoc. Bioinformatics* Chapter 5, Unit-5.6. doi: 10.1002/0471250953.bi0506s15

Fujii, R., Minami, A., Tsukagoshi, T., Sato, N., Sahara, T., Ohgiya, S., et al. (2011). Total biosynthesis of diterpene aphidicolin, a specific inhibitor of DNA polymerase α: heterologous expression of four biosynthetic genes in *Aspergillus oryzae*. *Biosci. Biotechnol. Biochem.* 75, 1813–1817. doi: 10.1271/bbb.110366

Görner, C., Häuslein, I., Schrepfer, P., Eisenreich, W., and Brück, T. (2013). Targeted engineering of cyclooctat-9-en-7-ol synthase: a stereospecific access to two new non-natural fusicoccane-type diterpenes. *ChemCatChem* 5, 3289–3298. doi: 10.1002/cctc.201300285

Hanson, J. R. (2015). Diterpenoids of terrestrial origin. *Nat. Prod. Rep.* 32, 76–87. doi: 10.1039/C4NP00108G

Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., and Hutchison, G. R. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* 4:17. doi: 10.1186/1758-2946-4-17

Hoshino, T., Nakano, C., Ootsuka, T., Shinohara, Y., and Hara, T. (2011). Substrate specificity of Rv3378c, an enzyme from *Mycobacterium tuberculosis*, and the inhibitory activity of the bicyclic diterpenoids against macrophage phagocytosis. *Org. Biomol. Chem.* 9, 2156–2165. doi: 10.1039/C0OB00884B

Hu, X., and Shelver, W. H. (2003). Docking studies of matrix metalloproteinase inhibitors: zinc parameter optimization to improve the binding free energy prediction. *J. Mol. Graph. Modell.* 22, 115–126. doi: 10.1016/S1093-3263(03)00153-0

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5

Ikegami, S., Taguchi, T., Ohashi, M., Oguro, M., Nagano, H., and Mano, Y. (1978). Aphidicolin prevents mitotic cell division by interfering with the activity of DNA polymerase-α. *Nature* 275:458. doi: 10.1038/275458a0

Janke, R., Görner, C., Hirte, M., Brueck, T., and Loll, B. (2014). The first structure of a bacterial diterpene cyclase: CotB2. *Acta Crystallogr. D Biol. Crystallogr.* 70, (Pt 6), 1528–1537. doi: 10.1107/S1399004714005513

Jia, M., Zhou, K., Tufts, S., Schulte, S., and Peters, R. J. (2017). A pair of residues that interactively affect diterpene synthase product outcome. *ACS Chem. Biol.* 12, 862–867. doi: 10.1021/acschembio.6b01075

Jones, B. (2017). *Diterpenoids: Types, Functions and Research.* New York, NY: Nova Science Publishers, Incorporated.

Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., et al. (2012). Template-based protein structure modeling using the RaptorX web server. *Nat. Protoc.* 7, 1511–1522. doi: 10.1038/nprot.2012.085

Kayser, O., Kiderlen, A. F., Bertels, S., and Siems, K. (2001). Antileishmanial activities of aphidicolin and its semisynthetic derivatives. *Antimicrob. Agents Chemother.* 45, 288–292. doi: 10.1128/AAC.45.1.288-292.2001

Khandelwal, A., Lukacova, V., Comez, D., Kroll, D. M., Raha, S., and Balaz, S. (2005). A Combination of docking, QM/MM methods, and MD simulation for binding affinity estimation of metalloprotein ligands. *J. Med. Chem.* 48, 5437–5447. doi: 10.1021/jm049050v

Lauchli, R., Rabe, K. S., Kalbarczyk, K. Z., Tata, A., Heel, T., Kitto, R. Z., et al. (2013). High-throughput screening and directed evolution of terpene synthase-catalyzed cylization(). *Angew. Chem. Int. Ed. Engl.* 52, 5571–5574. doi: 10.1002/anie.201301362

Liu, W., Feng, X., Zheng, Y., Huang, C.-H., Nakano, C., Hoshino, T., et al. (2014). Structure, function and inhibition of ent-kaurene synthase from *Bradyrhizobium japonicum*. *Sci. Rep.* 4:6214. doi: 10.1038/srep06214

Lopes, A. A., and Pupo, M. T. (2011). Biosynthesis of aphidicolin proceeds via the mevalonate pathway in the endophytic fungus *Nigrospora sphaerica*. *J. Braz. Chem. Soc.* 22, 80–85. doi: 10.1590/S0103-50532011000100010

Morrone, D., Hillwig, M. L., Mead, M. E., Lowry, L., Fulton, D. B., and Peters, R. J. (2011). Evident and latent plasticity across the rice diterpene synthase family with potential implications for the evolution of diterpenoid metabolism in the cereals. *Biochem. J.* 435, 589–595. doi: 10.1042/BJ20101429

Morrone, D., Xu, M., Fulton, D. B., Determan, M. K., and Peters, R. J. (2008). Increasing complexity of a diterpene synthase reaction with a single residue switch. *J. Am. Chem. Soc.* 130, 5400–5401. doi: 10.1021/ja710524w

Oikawa, H., Nakamura, K., Toshima, H., Toyomasu, T., and Sassa, T. (2002). Proposed mechanism for the reaction catalyzed by a diterpene cyclase, aphidicolan-16β-ol synthase: experimental results on biomimetic cyclization and examination of the cyclization pathway by ab initio calculations. *J. Am. Chem. Soc.* 124, 9145–9153. doi: 10.1021/ja025830m

Oikawa, H., Toyomasu, T., Toshima, H., Ohashi, S., Kawaide, H., Kamiya, Y., et al. (2001). Cloning and functional expression of cDNA encoding aphidicolan-16β-ol synthase: a key enzyme responsible for formation of an unusual diterpene skeleton in biosynthesis of aphidicolin. *J. Am. Chem. Soc.* 123, 5154–5155. doi: 10.1021/ja015747j

Pedrali-Noy, G., Spadari, S., Miller-Faurès, A., Miller, A. O., Kruppa, J., and Koch, G. (1980). Synchronization of HeLa cell cultures by inhibition of DNA polymerase alpha with aphidicolin. *Nucleic Acids Res.* 8, 377–387. doi: 10.1093/nar/8.2.377

Pemberton, R. P., Ho, K. C., and Tantillo, D. J. (2015). Modulation of inherent dynamical tendencies of the bisabolyl cation via preorganization in epi-isozizaene synthase. *Chem. Sci.* 6, 2347–2353. doi: 10.1039/C4SC03782K

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

Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781–1802. doi: 10.1002/jcc.20289

Schalk, M., Pastore, L., Mirata, M. A., Khim, S., Schouwey, M., Deguerry, F., et al. (2012). Toward a biosynthetic route to sclareol and amber odorants. *J. Am. Chem. Soc.* 134, 18900–18903. doi: 10.1021/ja307404u

Schrepfer, P., Buettner, A., Goerner, C., Hertel, M., van Rijn, J., Wallrapp, F., et al. (2016). Identification of amino acid networks governing catalysis in the closed complex of class I terpene synthases. *Proc. Natl. Acad. Sci. U.S.A.* 113, E958–E967. doi: 10.1073/pnas.1519680113

Serrano-Posada, H., Centeno-Leija, S., Rojas-Trejo, S., Stojanoff, V., Rodriguez-Sanoja, R., Rudino-Pinera, E., et al. (2015). Crystallization and X-ray diffraction analysis of a putative bacterial class I labdane-related diterpene synthase. *Acta Crystallogr. Sect. F* 71, 1194–1199. doi: 10.1107/S2053230X15014363

Shetty, R. P., Endy, D., and Knight, T. F. (2008). Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* 2:5. doi: 10.1186/1754-1611-2-5

Starczewska, E., Beyaert, M., Michaux, L., Vekemans, M.-C., Saussoy, P., Bol, V., et al. (2016). Targeting DNA repair with aphidicolin sensitizes primary chronic lymphocytic leukemia cells to purine analogs. *Oncotarget* 7, 38367–38379. doi: 10.18632/oncotarget.9525

Starratt, A. N., and Loschiavo, S. R. (1974). The production of aphidicolin by *Nigrospora sphaerica*. *Can. J. Microbiol.* 20, 416–417. doi: 10.1139/m74-063

Stone, J. E. (1998). *An Efficient Library for Parallel Ray Tracing and Animation: A Thesis Presented to the Faculty of the Graduate School of the University of Missouri-Rolla in Partial Fulfillment of Requirements for the Degree of Master of Science in Computer Science.* Master thesis, University of Missouri-Rolla, Rolla, MO.

Trott, O., and Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi: 10.1002/jcc.21334

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., et al. (2010). CHARMM General Force Field (CGenFF): a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* 31, 671–690. doi: 10.1002/jcc.21367

Vanommeslaeghe, K., and MacKerell, A. D. (2012). Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *J. Chem. Inf. Model.* 52, 3144–3154. doi: 10.1021/ci300363c

Vanommeslaeghe, K., Raman, E. P., and MacKerell, A. D. (2012). Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* 52, 3155–3168. doi: 10.1021/ci3003649

Xu, J., and Li, M. (2003). Assessment of RAPTOR's linear programming approach in CAFASP3. *Proteins* 53, 579–584. doi: 10.1002/prot.10531

Yee, N. K. N., and Coates, R. M. (1992). Total synthesis of (+)-9,10-syn- and (+)-9,10-anti-copalol via epoxy trienylsilane cyclizations. *J. Org. Chem.* 57, 4598–4608. doi: 10.1021/jo00043a014

Yoshikuni, Y., Martin, V. J. J., Ferrin, T. E., and Keasling, J. D. (2006). Engineering cotton (+)-δ-cadinene synthase to an altered function: germacrene D-4-ol synthase. *Chem. Biol.* 13, 91–98. doi: 10.1016/j.chembiol.2005.10.016

2

Modular Biomanufacturing for a Sustainable Production of Terpenoid-Based Insect Deterrents

By Wolfgang Mischko, Max Hirte, Simon Roehrer, Hannes Engelhardt, Norbert Mehlmer, Mirjana Minceva, and Thomas B. Brück

Synopsis

Production and degradation of traditionally applied pesticides have a measurable impact on health and environment. Therefore, it is necessary to identify compounds that protect crops under ecological constraints. The research "Modular biomanufacturing for a sustainable production of terpenoid-based insect deterrents" describes a holistic production process encompassing biotechnological production and purification of cembranoid diterpenoids at technical scale as well as successful biological activity screening of α and β-cembratrienol on insect and cancer cell lines.

A polycistronic operon for the production of cembratrienol was constructed and optimized. In order to generate a circular and sustainable production process, wheat bran was used as renewable raw material feedstock for the fermentative production. This low value by-product is produced during the wheat milling procedure, however needs to be hydrolyzed and concentrated prior to usage in a microbial fermentation process setting. A central innovation was the *in situ* extraction of the target molecule during the fermentation process. This was conducted by a by-pass fermentation strategy implementing a continuously flushed column that was filled with a hydrophobic adsorber polymere. In order to maximize the diterpene yield, elution and regeneration of the column was conducted every 30 hours. However, a decrease in the adsorber's capacity was measurable. Eventually, a cembrantrienol yield of almost 80 mg/L was reached exceeding all previously reported biotechnologically obtained titers and that of natural cembratrienol producers.

After the fermentation process all terpene extracts were combined and evaporated, which resulted in an oily resin. It was hardly possible to mix the resin with an acetonitrile water mixture applied in cembratrienol RP-HPLC purification. To that end, we devised a new purification strategy for the isolation of the α- and β-cembratrienol isomers encompassing centrifugal partitioning chromatography (CPC) followed by chiral RP-HPLC. Individual screening of the isomers revealed that α-cembratrienol has higher activity on aphids that are the predominant pest in wheat production. Moreover, both compounds showed significant activity towards MCFS cancer cell lines.

Authors contribution

In this research collaboration I consulted and supported Wolfgang Mischko in analytical steps, as well as in upstream and downstream processing steps. Together we devised ternary, biphasic solvent mixtures that are suitable for diterpene purification in liquid/liquid chromatography.

Modular biomanufacturing for a sustainable production of terpenoid-based insect deterrents†

Wolfgang Mischko, [ID] [a] Max Hirte, [ID] [a] Simon Roehrer,[b] Hannes Engelhardt,[c] Norbert Mehlmer, [ID] [a] Mirjana Minceva [ID] [b] and Thomas Brück [ID] *[a]

Synthetic agricultural insecticides are toxic to many species and accumulate in the environment. Therefore, the development of target-specific and biodegradable insecticides and deterrents is in demand. This study describes an improved and sustainable process for the green production of a biological insect repellent based on the diterpene cembratriene-ol (CBT-ol). This compound is a natural part of the tobacco (*Nicotiana* sp.) plant's defense against insects and thus minimizes damage to the environment. The study reports a new recombinant (*E. coli*) CBT-ol production and purification system. Efficient production was achieved by ribosomal binding site combinatorics using the BioBrick assembly system. These methods generated a metabolically balanced microbial system capable of generating 78.9 ± 2.4 mg L^{-1} CBT-ol in a 50 L bioreactor. Fermentations were entirely carried out on enzymatically generated wheat bran hydrolysate, representing a waste fraction of the grain milling process. The application of this complex and cost-efficient cultivation medium enabled an ecologically and economically sensible production of this high-value insect deterrent. Moreover, an ecologically favorable downstream processing protocol was established, combining adsorptive CBT-ol capture and centrifugal partition chromatography (CPC) followed by HPLC-based isomer separation. This is the first report using CPC to recover recombinant-generated, bioactive terpenes. The methodology enabled 95% CBT-ol recovery and purification in a single CPC step with significantly reduced solvent consumption in comparison to conventional chromatographic methods. *In vivo* and *in vitro* bioactivity studies confirmed the insecticide characteristics but also indicated that CBT-ol shows other bioactivities specifically targeting Gram-positive bacteria.

Introduction

With over 50 000 characterized compounds, terpenoids are the largest and structurally most diverse group of natural products. With respect to the development of a sustainable bioeconomy, terpenoids represent important renewable chemical building blocks with industrial applications in the pharmaceutical (*e.g.*, taxol[1]), cosmetic (*e.g.*, pseudopterosin[2]), food (*e.g.*, β-carotene[3]), fragrance (*e.g.*, nootkatone[4]), and chemical (*e.g.*, limonene[5]) industries. Like most bioactives, terpenoids are secondary metabolites, only found in very low abundance in the natural source.[6] As most industrially relevant terpenoids

are derived from plant material, their extraction poses significant challenges due to the presence of multiple contaminants. To obtain high titers of specific terpenoids, there is an ongoing effort in the scientific community to generate important building blocks by genetic engineering in microbial hosts, such as *E. coli* and *S. cerevisiae*.[7] While these efforts have been successful for some industrially important compounds such as limonene (2.7 g L^{-1})[8,9] and nootkatone (208 mg L^{-1}),[10] most engineering efforts do not generate industrially relevant titers of the desired target compounds. Moreover, reported studies conventionally do not address issues of process scaling and downstream purification of target molecules under economic constraints. The current study addresses issues of sustainable heterologous production and purification at the laboratory and technical scales for a diterpene-type insect repellent from agricultural waste streams targeted at the agrochemical industry.

In the context of the global insecticide market, chemically synthesized active compounds, such as artificial pyrethrum derivatives and most prominently neonicotinoids (*e.g.*, imidacloprid, thiamethoxan and clothianidin; worth US $1.89 billion[11]), are dominant, fast-acting but non-target discriminant products. Neonicotinoids represent the major class of

[a]*Werner Siemens-Chair of Synthetic Biotechnology, Department of Chemistry, Technical University of Munich, D-85748 Garching, Germany. E-mail: brueck@tum.de*
[b]*Biothermodynamics, TUM School of Life Sciences Weihenstephan, Technical University of Munich, D-85354 Freising, Germany*
[c]*Institute of Bioprocess Engineering, Friedrich-Alexander University of Erlangen-Nuernberg, D-91052 Erlangen, Germany*

insecticides that mediate neurotoxic effects in all insects through irreversible binding to nicotinic acetylcholine receptors.[12,13] Therefore, these compounds affect pests as well as beneficial insects such as bees and bumble bees, thereby negatively impacting pollination of agricultural crops and biodiversity in rural regions.[13,14] Chemical insecticides even endanger industrial agricultural activity and the sustainability of a still growing human population. Moreover, these chemicals are poorly biologically degradable,[13,15] which leads to an accumulation in the environment resulting in a negative impact on the biodiversity of terrestrial and aquatic ecosystems.[16] The broad environmental effects of neonicotinoids has recently triggered the European Commission to strictly limit the applicability of these compounds in agricultural activities.[17–19] By contrast, bio-based insecticides are rapidly degraded by terrestrial microbes or light and therefore do not accumulate. In addition, selected natural insecticidal compounds are non-toxic to off-target insects, ensuring only positive effects on agricultural activity and crops yields.

In this context, the class of cembranoid diterpenes, originally reported from cuticular wax of many *Nicotiana* species, has been identified as a promising resource. The two most abundant epimeric tobacco cembranoids are (1S,2E,4S,6R,7E,11E)-2,7,11-cembratriene-4,6-diol and (1S,2E,4R,6R,-7E,11E)-2,7,11-cembratriene-4,6-diol (CBT-diol) with the typical macrocyclic C_{14} backbone.[20,21] These are constituents of the plant's exudates[22] representing a major part of the plant's defense against insects, pathogenic microbes and herbivores.[23] Additionally, several other bioactivities encompassing anti-tumor,[24] antibiotic[25] or neuroprotective[26] properties have been reported. Interestingly, the mono-hydroxylated biosynthetic precursors of these compounds, cembratriene-ol (CBT-ol), could only be detected in trace amounts *in planta*. However, CBT-ol has recently received some attention[20,27] due to its apparent insecticidal activity but data are scanty.

CBT-ol occurs as α- and β-2,7,11-cembratriene-4-ol isomers (α-CBT-ol, also known as thunbergol or isocembrol and β-CBT-ol, equivalent to 4-epiisocembrol) with absolute configurations determined as (1S,2E,4R,7E,11E)- and (1S,2E,4S,7E,11E)-2,7,11-cembratriene-4-ol, respectively.[28,29] The corresponding CBT-ol synthase (CBTS) has been reported from *Nicotiana tabacum* and *Nicotiana sylvestris* with almost identical amino acid sequences. The enzyme is a 58 kDa monomeric[29] class I terpene synthase with the signature aspartate-rich DDXXD and (N,D)DXX(S,T)XXXE motifs, and a magnesium cluster, which is essential for the ionization of an isoprenoid diphosphate group to generate a reactive carbocation intermediate.[30] In nature, the subsequent step in the native pathway is a site-specific oxidation by an also specified cytochrome P450 monooxygenase. The transcriptional silencing of this CYP450 increases the amount of available CBT-ol in the plant exudates and simultaneously enhances the resistance to aphids.[20,31] Other results demonstrate the direct application of purified CBT-ol and its potential as a contact insecticide.[32]

CBT-ol and related cembrene-like compounds derived from the tobacco plant are susceptible to biodegradation and are rapidly catabolized within the maturing plant.[21,33] Additionally, field studies demonstrated the natural degradation process of exogenously applied CBT-ol without an accumulation in the ecosphere.[34] These results contrast those of synthetic neonicotinoids, which are highly toxic and accumulate in soil or water ecosystems.[13,15,16] Therefore, the use of CBT-ol as an ecologically friendly and rapidly biodegradable insecticide represents a radically new, sustainable crop protection strategy, that does not affect beneficial insects populations.

Due to the low concentration *in planta* (0.18% of the leaf dry weight[20]) and the presence of various functionalized cembrene derivatives,[29] the scale-up and purification of the bioactive CBT-ol from plant biomass is technically challenging and associated with high costs. To simplify this process, a more targeted approach is the direct biosynthesis of bioactive cembranoids in an engineered microbial host.[7,35] The recent developments in synthetic biology and bioprocess engineering provide a directed route to obtain tailor-made bioactive compounds, which even improve on cumbersome chemical synthesis routes for structurally complex terpenoids.[36]

The costs and the sustainability of fermentation are still a central obstacle in the economical production of heterologously generated natural products. Utilization of complex biomass hydrolysates derived from agricultural waste streams may provide an alternative toward economically viable production processes. In this respect, wheat bran, a waste stream of the grain processing industry encompassing a production volume of more than 7 million tons per year in the EU[37] may present a formidable resource for the generation of cost-effective fermentation media. Wheat bran predominantly consists of non-starch polysaccharides (up to 45% w/w) but in contrast to other agricultural residues (*e.g.*, straw and wood) it lacks significant amounts of lignin.[37,38] Therefore, wheat bran may be more amenable to enzymatic hydrolysis, which was extensively investigated in this study.

In addition to fermentation media, the recovery of terpenoid target molecules from the fermentation broth is a particular challenge as critical concentrations may affect the production host and the received titers. In this respect, a tailor-made capturing method was evaluated and combined with solid support-free liquid–liquid chromatography, better known as centrifugal partition chromatography (CPC), representing a suitable alternative to conventional separation methods. This flexible technique does not require expensive HPLC column material and provides a high loading capacity in conjunction with efficient sample recovery. The method can be regarded as a more sustainable chromatography method as solvent consumption is significantly reduced compared to the alternatives.[39,40] As this methodology allows for tailor-made biphasic solvent systems, CPC is a rather versatile technology to purify molecules varying in size, polarity or chemical functionality.[39,41] While CPC has been applied in many cases for the purification and separation of bioactive compounds

from plant biomass,[42] this is the first account of the purification of a heterologously produced diterpene in microbial cells.

The total chemical synthesis of the cembrene type macrocyclic core relies on petrochemical educts, which are assembled in at least 8 consecutive reaction steps.[43] This synthesis path involves several intermediate and waste streams. Various toxic, irritant and organ damaging reagents (*e.g.*, oxalyl chloride and triphenylphosphine) and hazardous solvents such as THF, DMF and HMPA are applied. Moreover, the synthesis steps are dependent on the application of several protecting moieties (*e.g.*, TBSCl). As reactions are carried out at temperatures varying between −78 °C to +50 °C, the energy efficiency of the synthesis route is limited.[43,44] An equivalent scenario is reported for the chemical synthesis routes of neonicotinoid type insecticides. The use of hazardous reactants, intermediates and solvents, which are particularly harmful for the environment, is indispensable as well.[45] In contrast, our new biotechnological production process strictly adheres to the 12 principles of green chemistry.[46] Most prominently, we rigorously apply the principles of preventing waste, the use of renewable feedstocks and catalysis as well as the reduction of derivatives.

For the first time, this study presents a holistic production process for a bioactive diterpene insect-deterrent using an engineered *E. coli* production host cultivated on enzymatically generated wheat bran hydrolysate (Fig. 1). In contrast to the chemical multi-step synthesis of the cembrene macrocycle,[43,44] we engineered a consolidated whole cell biocatalyst that enables the single step synthesis in conjunction with a site-specific hydroxylation. Moreover, the biotechnological approaches applied in our study provides a hazardless synthesis route that uses benign chemicals, benign auxiliaries and is designed for energy efficiency. With respect to production host engineering, this study presents a combinatorial ribosome binding site-centered approach for the harmonization of biosynthetic enzyme expression levels. That ultimately allows for an optimized metabolic flux of terpene-relevant intermediates resulting in high target product titers. Additionally, it was demonstrated that the target compound CBT-ol could be purified from crude fermentation broth using a combined adsorptive bypass system in conjunction with a customized CPC methodology. The purified CBT-ol target compound was structurally characterized and respective isomers were tested for their bioactivity in insect deterrent and continuative bioassays. This is the first account of an *E. coli*-based terpene production system cultivated on wheat bran milling waste without accumulating toxic and persisting waste streams. The application of this agricultural residue does not impact agricultural activity *per se* and has no negative influence on food production. We demonstrate that CBT-ol serves as an effective insect deterrent that has no direct toxic effects on the target organisms and may therefore serve in organic farming.

Results and discussion

The new agrochemical biorefinery setting was developed according to subsequent unit operations. One main operation was the construction of a suitable high-performance *E. coli* production system by metabolic engineering.

Establishing cembratriene-ol biosynthesis in *E. coli*

In order to develop the heterologous production and processing of CBT-ol, the codon-optimized open reading frame of the previously reported CBTS (lacking the 52 AA plastid transit peptide) from *Nicotiana sylvestris* (AAS46038.1) was co-expressed with an essential geranylgeranyl pyrophosphate synthetase (GGPPS) gene in *E. coli*. A pETDuet-1 vector carrying two strong T7 promotors and integrated ribosomal binding site (RBS) sequences was used for the maximal expression rate of the enzymes resulting in the pETDuet-GGPPS-CBTS vector. The initial gene construct was transformed into *E. coli* HMS 174 (DE3), which was subsequently cultivated (37 °C, 48 h) in artificial M9 minimal medium. Subsequently, the entire fermentation broth with cell mass was extracted with ethyl acetate. In the resulting crude extract, CBT-ol could be identified *via* Gas Chromatography-Mass Spectrometry (GC-MS) analysis by comparison to a commercial standard and to the respective NIST Database entry. The quantification by Gas Chromatography-Flame Ionization Detector (GC-FID) indicated a CBT-ol concentration of 0.92 mg L^{-1}. The action of the CBTS enzyme on the universal diterpene precursor GGPP typically generates two isomers.[20] However, since α- and β-CBT-ol isomers cannot be separated *via* GC-MS, they result in a single product peak (Fig. 3A).
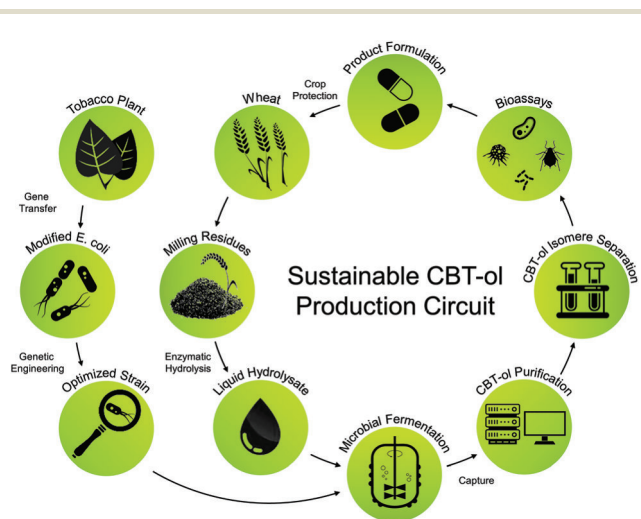
Fig. 1 Graphical presentation of the holistic production process for the sustainable generation of CBT-ol. This process combines the application of an engineered host organism and the circular product generation based on the utilization of the wheat bran waste stream.

Optimization of cembratriene-ol production by designing a new synthetic polycistronic operon using RBS combinatorics

To increase the CBT-ol production efficiency, the entire production system had to be redesigned on a genetic level. In this regard, it was essential to primarily enhance the supply of the universal terpene precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) in the native *E. coli* non-mevalonate (MEP) isoprenoid pathway. To that end, the major enzymatic bottlenecks 1-deoxy-D-xylulose-5-phosphate synthase (dxs) and IPP isomerase (idi) were overexpressed (Fig. 2A). This measure has previously been reported to increase heterologous terpene yields.[47] All four genes were placed under the control of only one inducible tac promoter due to its benefits in controlling a whole operon without the danger of overcharging the cells like the previous T7 promotor.

To reduce the metabolic burden on the microbial host, we opted to generate a single plasmid containing a polycistronic operon, which allowed for the minimization of the required antibiotics selectors, thereby reducing cellular stress.[48] Moreover, heterologous production *via* non-native metabolic pathways are often limited by the relative biosynthetic enzyme expressions. Therefore, the respective enzyme expression levels must be fine-tuned to balance the system and increase target productivity.[49] In this regard, the order of enzyme operators within an operon structure can be altered.[50] Alternatively, their relative expression rates can be harmonized in a combinatorial approach by varying the strength of the respective ribosomal binding sites (RBS).[49] This leads to different possibilities in the design and optimization of metabolic pathways to ensure high enzyme activities and to avoid protein burden or the accumulation of toxic intermediates. In this study, we opted to
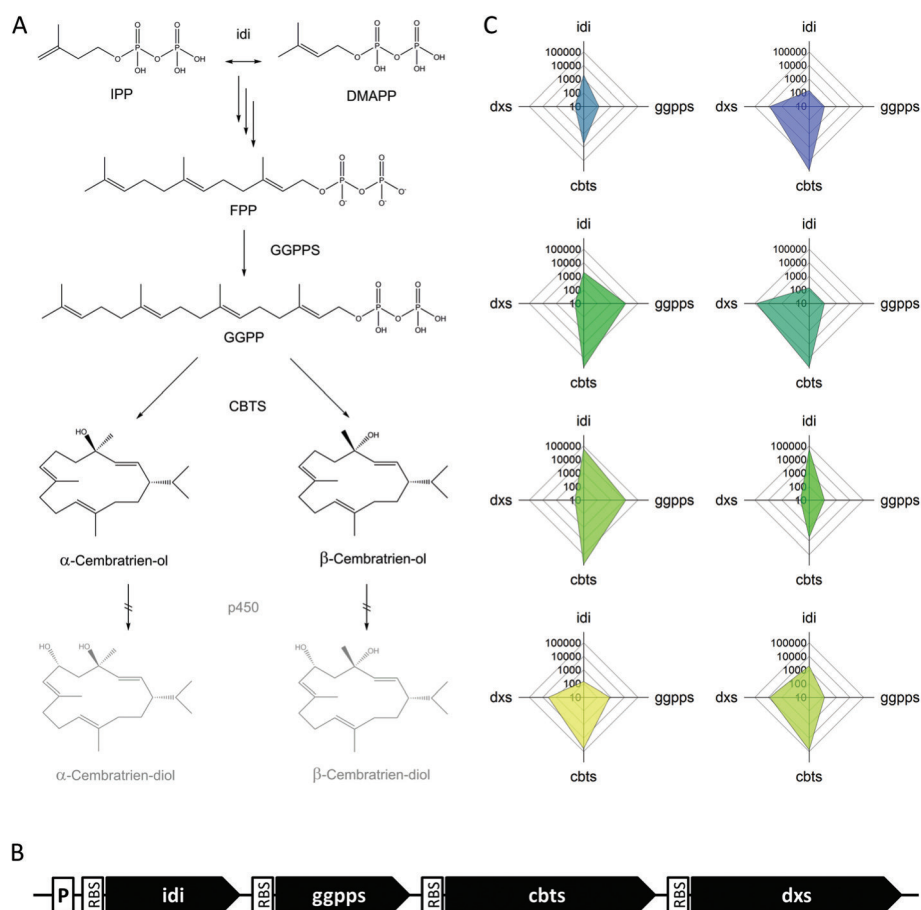
Fig. 2 [A] The metabolic pathway for the production of CBT-ol in *E. coli* consists of the native upstream non-mevalonate (MEP) isoprenoid pathway and a heterologous downstream terpenoid pathway. The structural diversity of all terpenes is derived from two universal isoprenoid C_5 building blocks isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) to form farnesyl diphosphate (FPP, C_{15}) and geranylgeranyl diphosphate (GGPP, C_{20}). GGPP represents the substrate for the cyclization reaction by the CBT-ol synthase (CBTS) providing the two isoforms α- and β-CBT-ol. In the tobacco plant, the further formation of the final α- and β-CBT-diols is catalyzed by a P450 hydroxylase. [B] Design of the synthetic operon for an increased precursor flux and CBT-ol production. Two enzymatic bottlenecks within the upstream MEP pathway (dxs, idi) and the heterologous downstream genes for the GGPP synthase (ggpps) and CBTS were combined in the synthetic operon (idi-ggpps-CBTS-dxs). The pathway is placed under the control of an inducible tac promoter (P). [C] Different exemplary combinations of the operon library illustrated by radar diagrams show RBS-dependent changes in the specific translation rate (au) of all involved genes within the polycistronic operon. The modulation of every enzyme expression level is essential to ensure a balanced pathway function and process efficiency.

leave the order of enzyme operators within the operon structure unaffected, while varying the respective RBS strength for each element in a combinatorial manner. All selected RBS were categorized by their translation initiation rate (au) using the RBS calculator tool[51,52] depending on the associated gene spanning a range from about 50 au up to 500 000 au. Each enzyme operator encoded in the synthetic CBT-ol operon structure was randomly paired with a preselected RBS set (Fig. 2B). We applied the BioBrick cloning strategy designated for the assembly of synthetic operons.[49,53] As a result, a widespread library of different combinations was combinatorially assembled and transformed into *E. coli* HMS 174 (DE3) (Fig. 2C). Subsequently, in excess of 200 clones were screened for CBT-ol production titers. The CBT-ol concentration varied greatly with the changing expression levels within this library. The clone with the highest CBT-ol productivity and constant cell growth was selected for further experiments. Detailed characterization of this clone demonstrated that the optimized translation rate combination for CBT-ol centered biosynthetic genes was: idi (830 au); ggpps (2829 au); CBTS (60 688 au); dxs (167 au). Interestingly, these relatively low translation values indicate that construct stability and production efficiency increase when the cellular burden with heterologous proteins is low. Moreover, the low expression of heterologous proteins also fostered an increased metabolic balance within the cellular system. The new strain carrying the optimized operon (pSB4K5-CBT) was used for further evaluation of the CBT-ol production. Three-day shake flask cultures (50 mL) resulted in one dominant peak in a GC-MS chromatogram (Fig. 3B) with CBT-ol concentrations of up to 13.9 mg L^{-1}. Subsequent, preparative TLC purification resulted in a colorless, highly viscous liquid. This liquid was characterized by nuclear magnetic resonance spectroscopy (NMR) as a mixture of the α- and β-CBT-ol isoforms (Fig. 3D). These isoforms are commonly termed thunbergol and 4-epiisocembrol, respectively.

A comparison of CBT-ol product titers showed that the synthetic CBT-ol operon with harmonized RBS binding sites provided a 15-fold product increase with respect to the primary expression system (0.91 mg L^{-1}). The RBS optimized operon was used for further technical scale-up, purification and bioactivity studies.

Enzymatic wheat bran hydrolysate as alternative nutrition supply

To enable a sustainable production of terpenoids that does not impact land use and agricultural activity, we focused on wheat bran as a feedstock, which constitutes up to 19% of the total grain composition[54] and is considered a major by-product of flour milling. In this study, we had access to this milling residue and devised an enzymatic process for its hydrolysis and use as fermentation medium for *E. coli*. Therefore, the procedure is in line with the concept of generating a sustainable agrochemical from the field for the field. The use of wheat bran hydrolysate as a carbon source has been reported previously in another context.[55] Prior to tailoring the enzyme system for wheat bran hydrolysis, we confirmed the reported biomass composition range of the present wheat bran feedstock: ash 5.2% (3.9–8.10%[56]) (w/w), protein 19.8% (9.60–18%[56]) (w/w) and carbohydrates 65.2% (cellulose 11–13%;[37] hemicellulose 27–31%;[37] starch 18–20%[37]) (w/w). The amounts of lignin and lipids were measured as 5.5% (w/w) and 3.7% (w/w), respectively. Much like other agricultural residues (*e.g.*, straw), wheat bran requires physical pretreatment, a preprocessing step that improves enzyme access to the cellulose.[57,58] In this study, we selected a mild hydrothermal pretreatment (121 °C, 15 min). This methodology did not result in any sugar degradation products and was environmentally and energetically favorable due to the lack of chemical additives (*e.g.*, acid or base).[58] The subsequent enzymatic hydrolysation (combined amylase, cellulase
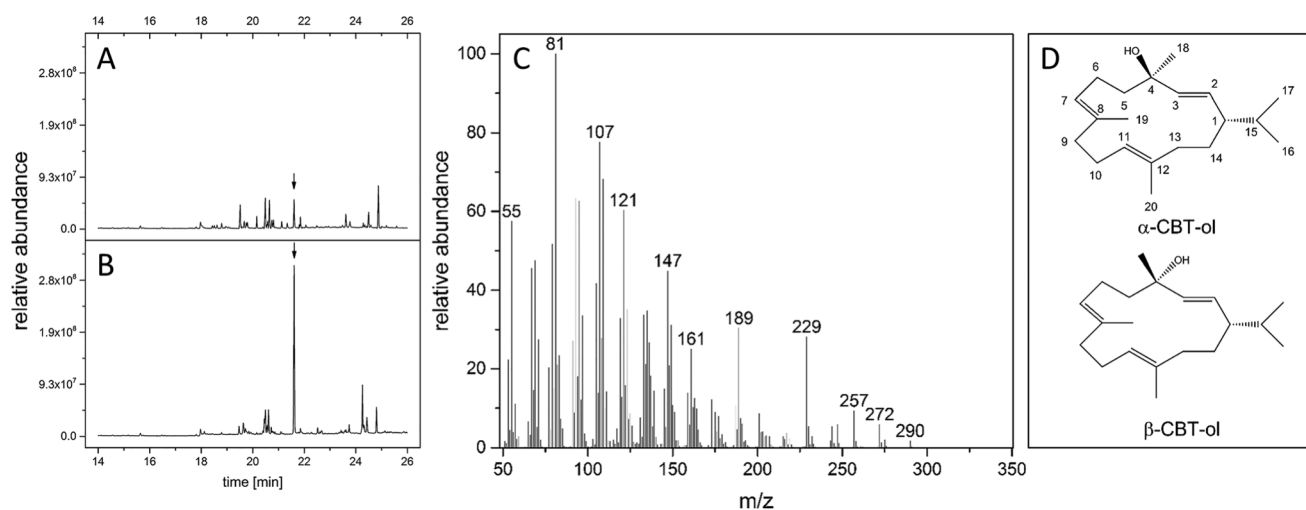
Fig. 3 Gas chromatographic profiles of extracts from *E. coli* cultures transformed with [A] pETDuet-GGPPS-CBTS and [B] the optimized pSB4K5-CBT plasmid showing a dominant product peak at 21.72 min RT. The optimized system presents a significantly higher CBT-ol concentration. [C] The associated mass peaks of the produced CBT-ol, which consists of two isomeric forms with differently oriented hydroxyl groups [D].

and hemicellulase system) of a 100 g L^{-1} wheat bran solution released 33.0 g L^{-1} glucose, 20.5 g L^{-1} xylose and 1.3 g L^{-1} arabinose. The total sugar content of the final wheat bran hydrolysate was 548 mg g^{-1} dry wheat bran, which is equivalent to a 91.7% recovery of accessible sugar from the solid feedstock. For an optimal *E. coli* cultivation medium, we added essential metal salts such as MgSO$_4$, CaCl$_2$, and the M9 mineral salt mixture (see Experimental). *E. coli* cultivated for 24 h in the wheat bran-based hydrolysate medium standardized to 10 g glucose per L showed better growth rates and final cell densities compared to conventional complex media like Luria–Bertani (LB). This is most likely because of the higher amounts of sugars in the wheat bran-based growth medium. A parallel growth experiment with *E. coli* in LB medium supplemented with the same amount of glucose showed approximately equivalent growth rates and final OD$_{600}$. The results indicate that the chemical composition (*e.g.*, low lignin concentration) of wheat bran is particularly favorable to achieve efficient enzymatic hydrolysis with high fermentable sugar titers compared to other agricultural waste streams, such as corn stover[59] or wheat straw.[60] Additionally, it demonstrates the efficiency of the applied enzymatic hydrolysis protocol. The data shows that wheat bran constituents could replace commercial carbon and nitrogen sources, such as purified glucose or yeast extract. Therefore, the generated hydrolysate media can serve as an excellent replacement for commercially available cultivation media in industrial applications.

Technical scale-up with semi-continuous product capturing in 50 L controlled bioreactors

The economic feasibility of our CBT-ol production process is interdependent with end-product toxicity effects. The assays demonstrated that CBT-ol imparts almost no toxicity towards *E. coli* at concentrations of 2 g L^{-1}. Specific cell growth in cultivation media supplemented with increasing amounts of CBT-ol is comparable to the negative control (see ESI†). The lack of toxicity forms the basis for the efficient biotechnological production of CBT-ol *in E. coli*.

We aimed to optimize terpene product concentration by streamlining the entire bioprocess on a technically relevant scale of 50 L in controlled bioreactor systems. We utilized the wheat bran hydrolysate as fermentation medium in conjunction with the optimized CBT-ol production host. To facilitate the process, we additionally devised a comfortable capture method for parallel product removal. With regard to the latter, an adsorption-based bypass system was mounted on the bioreactor outlet, which enabled cyclization of the fermentation broth and semicontinuous depletion of the terpenoid product. In the developed system, the fermentation broth continuously flowed through an external column filled with hydrophobic Amberlite® XAD®-2 beads (Fig. 4A). The system provided for binding of hydrophobic products such as CBT-ol to the hydrophobic beads while the bacterial suspension was recycled into the fermentation process. This methodology reduced the product stress on the *E. coli* system and enabled simplified product recovery from the column material. For process optim-
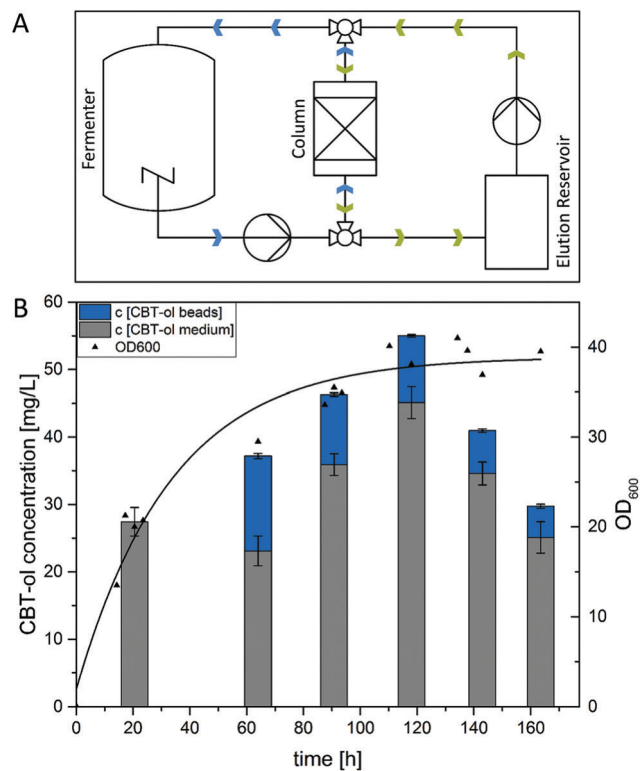
Fig. 4 [A] Sketch of the fermentation set up with the applied bypass system composed of two basic circuits. First circuit (blue arrows) serves to capture the product *via* absorption on the hydrophobic beads in the column. The second circuit (green arrows) is activated for the actual recovery and regeneration process by switching to a solvent which leads to product elution. The three-way valves allow an easy transition between the two circuits. [B] Bacterial growth curve and CBT-ol production rate during a 50 L fermentation with the applied bypass system. Extracted (blue) and remaining (gray) CBT-ol concentrations are shown for every point of measurement. The first capturing step started after 40 h and was then carried out every 24 h. The production dropped after 120 h which is therefore considered as the endpoint of the fermentation followed by continuous product depletion. The error bars represent the mean values ± standard deviation of technical triplicates.

ization, it was essential to provide a sufficient flow rate in the column in order to minimize the residence time of the bacterial suspension within the bypass system as much as possible. This minimal residence time was essential to maintain a critical oxygen and nutrient supply for bacterial growth. The bypass adsorption system presented in this study has the potential to replace extremely time- and solvent-consuming methods such as liquid–liquid extraction, which are conventionally applied for the quantitative recovery of hydrophobic molecules from the reactor volume after fermentation. To maintain operations, we eluted adsorbed products from the hydrophobic resin at 24 h intervals with fresh ethyl acetate to keep capturing efficiency over the duration of the fermentation process.

Within this set-up, the CBT-ol production rate after the induction with IPTG (after 15 h, OD$_{600}$ ~12) showed a constant correlation with the *E. coli* growth rate measured at OD$_{600}$

(Fig. 4B). The presented data reveal that CBT-ol is not accumulated in the stationary phase and that the biosynthesis occurs mainly in the active growth phase. The terpene production rate dropped after 120 h and the overall concentration of CBT-ol in the bioreactor decreased with subsequent capturing steps. At this time, the fermentation process was terminated. Similar results were reported for the biotechnological production of other terpenoid compounds.[61,62] The recombinant CBT-ol biosynthesis is directly coupled and therefore interdependent with the endogenous FPP supply. FPP biosynthesis in *E. coli* is linked to an active cellular metabolism since it is essential for growth related respiratory quinone- and cell wall synthesis.[62,63] Therefore, the recombinant CBT-ol production closely follows exponential growth and ceases in the stationary phase. The cumulative CBT-ol concentration obtained within the 120 h fermentation time was measured to be 78.9 ± 2.4 mg L^{-1}. The specific CBT-ol productivity was 5.17 mg g^{-1} dry cell weight (DCW) and is therefore 2.9 times higher than in natural tobacco biomass ($\sim$0.18 mg g^{-1} dry weight[20]). In the 50 L reaction system, we could therefore generate a total CBT-ol quantity of 3.5 g. Based on this data, the terpene yield corresponded to $\sim$3.4 mg CBT-ol per g glucose. The maximum and average production rates for CBT-ol were 1.48 ± 0.12 mg L^{-1} h^{-1} and 0.65 ± 0.02 mg L^{-1} h^{-1} over 120 h, respectively. With this method, it was possible to semi-continuously isolate more than 42% (w/w) of the generated CBT-ol from a large reactor volume already during the fermentation process and in a manageable time frame. A replacement of the adsorbent further increased the capturing efficiency subsequent to the fermentation. The pooled solvent fractions were concentrated *via* evaporation and subsequently separated with CPC.

To date, recovery of heterologous generated terpenes in almost all comparable studies relies on the *in situ* organic solvent extraction with dodecane or decane.[59,64,65] In comparison to alternative liquid–liquid extraction methods, this methodology prevents product volatilization into the gas phase and allows for significantly higher terpene titers, thus meeting all requirements of most literature studies.[66] However, this terpene recovery procedure is not trivial to implement on a larger scale and prevents simple downstream product recovery, since terpene products and the overlay solvent generally exhibit very similar physical properties. Therefore, an essential focus of this study was to establish a technically and economically scalable method for CBT-ol recovery and purification. In that regard, the developed bypass adsorption system enabled effective removal of toxic metabolites and separation of desired target products. Thereby, optimal *E. coli* growth and product recovery was achieved, while using less solvent compared to conventional methods.

With respect to the economic feasibility, most industrially relevant terpene production systems are currently designed for the generation of mono- or sesquiterpenes, such as limonene or nootkatone.[9,10,64,67] Especially in the field of low-cost biofuel production, the focus is exclusively on the less complex biosynthesis of smaller terpenoids.[59,64,67,68] In contrast, most diterpene-producing platforms are still less advanced than established production systems for mono- or sesquiterpenes. Insufficient precursor supply and flux into competing pathways as well as low expression of plant enzymes in microbial hosts are limiting factors that often prohibit high diterpene product titers.[69] Several years were spent optimizing specialized limonene production routes and processes[8] to achieve the current titers of 2.7 g L^{-1}.[9] The same effort in optimization of individual process steps, fermentation parameters and enzyme systems would lead to comparable diterpene titers in the gram scale. However, the target titer is also dependent on the pricing and application of the final product. In that respect, we estimate that titers of 500–1000 mg L^{-1} CBT-ol would be sufficient for industrial exploitation of our process, and we primarily envision to improve the cell densities ($OD_{600} > 150$) in technical scale fermentations.[70,71]

Purification and separation of cembratriene-ol isomers *via* CPC and chiral HPLC column

A major challenge in CBT-ol product purification and characterization is the separation of the respective isomers. In that regard, the concentrated crude extract, obtained in the capturing step, contained 10% (w/w) of racemic CBT-ol as well as other hydrophobic byproducts of the bacterial fermentation. To reduce costs and environmental impact, a purification procedure that provided high target product purity and recovery was favored. Centrifugal partition chromatography (CPC) allows versatile processing of different raw materials paired with reduced solvent consumption at high loading capacities.[40] In order to develop a CPC purification protocol for CBT-ol, we initially screened several biphasic solvent systems to find a system in which the partition coefficient (K) for racemic CBT-ol was within or close to the preferred "sweet spot" range ($0.4 < K < 2.5$). According to literature, this range provides a good compromise between the separation resolution, productivity, and solvent consumption in CPC.[72] In order to reduce experimental effort, the predictive thermodynamic model Conductor-like Screening Model for Realistic Solvation (COSMO-RS) was used as suggested by Hopmann *et al.*[73] For the selection of a suitable system, the partition coefficient K of CBT-ol was predicted in different biphasic solvent systems commonly used in CPC. For this study, only the molecular structures of CBT-ol and considered solvents were needed as input information. Solvent systems with various compositions and different polarities were screened, including hexane/ethyl acetate/methanol/water, heptane/ethyl acetate/methanol/water, butanol/methanol/water, hexane/ethyl acetate/acetonitrile and hexane/ethanol/acetonitrile. Based on the COSMO-RS predicted K-values of CBT-ol, the hexane/ethanol/acetonitrile system was selected to be most suitable. Systems with ethyl acetate and a K value of CBT-ol in the preferred range were excluded in order to avoid any UV-interference with the CBT-ol absorbance spectrum at the point of detection at 210 nm. Subsequently, the hexane/ethanol/acetonitrile system was experimentally evaluated to verify the predicted partition coefficients of CBT-ol and determine the K values of the unknown impurities, which could not be taken

into account in the COSMO-RS screening study due to missing molecular structures. Based on the K values determined by shake flask experiments, an acetonitrile/ethanol/n-hexane system composition of 27.4/1.6/71 (v/v/v) was selected as a suitable system for CBT-ol separation with CPC. The CPC separation was performed in descending mode, using the lower phase of the solvent system as the mobile phase. A stationary phase retention (S_F) of 0.63 was obtained. After the sample injection, only a small stationary phase loss of about 4 mL was detected over the entire runtime. Fractions were collected every 30 s and analyzed by GC-FID. In Fig. 5A, the reconstructed CPC chromatogram is presented. As apparent from the chromatogram, a small peak overlap occurred. Still, sufficient separation of racemic CBT-ol from main impurities was achieved with this system.

As a good compromise between the CBT-ol purity and recovery, the fractions were combined in a way that both exceeded 95% (Fig. 5A). According to the CPC chromatogram, a purity of racemic CBT-ol higher than 99.5% with a 28.3% loss of the target product would be possible. A productivity of 390 mg h^{-1} with a solvent consumption of 1.23 mL mg^{-1} was obtained, calculated based on a mass load of 280 mg per pulse injection and a purity of racemic CBT-ol of 95%.

The subsequent separation of α- and β-CBT-ol isomers was accomplished via a chiral HPLC column at a semi-preparative scale. The chromatogram indicated two clearly separated compounds (Fig. 5B), which could be identified as the respective CBT-ol isomers with NMR analysis (see ESI†). The ^{13}C and ^{1}H (CDCl$_3$) NMR data matched the reported data sets for thunbergol and 4-epiisocembrol, respectively (Table 1). Signature NMR signals like the ^{13}C NMR singlet at δ 72.8 and 73.8 for the C-4 atom connected to the hydroxyl group and the ^{1}H singlet at δ 1.34 and 1.27 (C-18 methyl group) could be confirmed.[28] The peak area integration revealed a relative ratio of 1 : 2.3, α- and β-CBT-ol in the racemic mixture, respectively. These results

Table 1 ^{13}C chemical shifts (CDCl$_3$) and assignments for α- and β-CBT-ol compared to the literature[28]

C	α	Ref.	β	Ref.
1	46.1	45.9	46.3	46.2
2	129.3	129.0	127.1	126.9
3	138.2	138.3	138.9	138.9
4	72.8	72.5	73.9	73.8
5	43.1	43.1	44.2	44.1
6	22.7	22.6	23.6	23.5
7	128.6	128.6	127.9	127.9
8	132.5	132.2	132.7	132.4
9	36.9	36.8	37.0	36.9
10	23.8	23.8	23.8	23.7
11	125.3	125.2	125.0	124.9
12	132.6	132.3	133.0	132.7
13	39.3	39.2	39.2	39.1
14	27.7	27.6	28.1	28.0
15	33.0	33.0	33.2	33.1
16	20.6	20.4	20.8	20.7
17	19.6	19.5	19.5	19.4
18	28.2	28.1	29.5	29.3
19	14.9	14.7	15.0	14.8
20	15.2	15.0	15.3	15.1

contrast with the isomer distribution reported *in planta*, where the α-CBT-ol isomer was dominant.[20]

Assessment of insect repellent activity by a specific aphid colonization test

CBT-ol is reported to exert an insect protective effect *in planta*.[32] Most interestingly, genetic plant modifications leading to higher CBT-ol concentration enhance the insect protective effect.[20] These reports lead to the expectation that recombinantly generated CBT-ol could be applied as an insect repellent on plant material. To test this hypothesis, a two-choice aphid colonization experiment was carried out, comparing CBT-ol-treated plants and untreated controls upon release
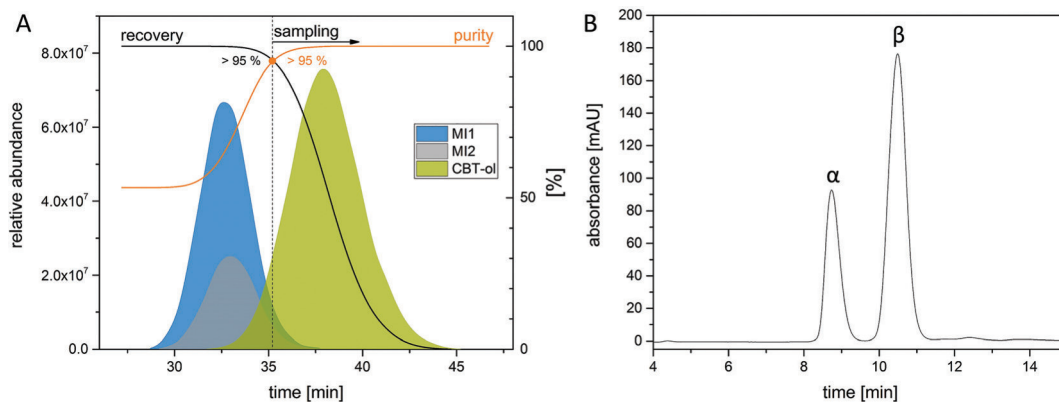
Fig. 5 [A] Chromatogram of a CPC batch separation of racemic CBT-ol from crude extract using the solvent system acetonitrile/ethanol/n-hexane 27.4/1.6/71 (v/v/v) (descending mode: lower phase as mobile phase, 8 mL min^{-1} mobile phase flow rate, 1700 rpm, m_{inj} = 280 mg, V_{inj} = 2 mL). Fraction content was determined and quantified via GC-FID. High purity as well as recovery rates were intended to separate CBT-ol (green peak) from the major impurities (blue and gray peaks). A purity >95% was aspired choosing a specific starting point for the sample fractioning after 35.2 minutes. [B] Separation of the two CBT-ol isomer forms using a chiral Lux® 5 μm amylose-1 LC column 250 × 10 mm (mobile phase: ACN and H$_2$O (7 : 3), 2.2 mL min^{-1} flow rate, m_{inj} = 40 mg, V_{inj} = 2 mL) resulted in a chromatogram with two clear and independent peaks representing the α- and β-CBT-ol (ratio 1 : 2.3).

of live aphids (*Rhopalosiphum padi*). We observed an increased aphid deterrent activity regarding the infestation rate for the CBT-ol-treated plants. This protective effect reached a maximum of 70% within the first 48 h after CBT-ol racemate application (Fig. 6A). We subjected the data set to a Student's *t*-test, which indicated a high significance ($p < 0.005$) with regard to the CBT-ol aphid repellent activity. These results are consistent with the previous reports indicating that increased CBT-ol content in plants results in lower aphid colonization,[20] and potentially acting as a contact insecticide.[32] To gain a more detailed insight into the bioactivity of the two CBT-ol isomers, we investigated the specific aphid deterrent effects for each compound separately. The data initially suggested a slightly higher activity of the α-isomer, but a statistical evaluation did not support a significant difference between the CBT-ol isomer activities. In the control experiments with commercially available plant-derived and pyrethrin-based insecticide,[74] all aphids that initially colonized the assay plants were killed. By contrast, the application of either CBT-ol isomer only induced an avoidance behavior in the applied aphids, which indicates that CBT-ol has a deterrent but no actual toxic activity toward these pests. This suggests that CBT-ol utilization may mimic the biological protective mechanism of the tobacco plant and exert no toxic effects on beneficial insects, such as bees, which would significantly benefit agricultural efficiency *per se*.

Regarding CBT-ol formulation, application and dosage as an insecticidal repellent, further studies have to be carried out. In summary, the data suggest that CBT-ol may play a promising role in future plant protection strategies as a deterrent. Additionally, the application of CBT-ol as a topical insect repellent for human use may be an alternative. In that respect, the development of a cosmetic application may be associated with reduced regulatory demands and could accelerate market entry for this interesting compound.

Extended cembratriene-ol bioactivity studies

Due to previous literature evidence, this study initially focused on the characterization of CBT-ol as an insecticide deterrent. Although various hydroxylated cembrenes have been reported to show neuroprotective[26] and anti-tumor[75] activities, at present no data have been reported for the monohydroxylated CBT-ol in this regard. To elucidate whether the CBT-ol racemate has anti-microbial, anti-tumor or extended anti-insecticidal activities, we conducted an array of *in vitro* toxicity tests with respective cell lines. We applied the purified CBT-ol racemate to nine different cell types including human cancer cells (MCF-7, HeLa), bacteria (*E. coli*, *B. subtilis*, *M. luteus*), yeast (*C. glabrata*, *S. pombe*, *Z. rouxii*) and *Spodoptera frugiperda* insect (*Sf21*) cells. Of the nine tested cell lines, six showed growth inhibition when exposed to CBT-ol. In that respect, the half maximum inhibitory concentration (IC_{50}) could be calculated (Fig. 6B). *S. pombe*, *B. subtilis* and *M. luteus* appeared to be the most sensitive organism to the CBT-ol treatment with the lowest IC_{50} values of 8, 9 and 10 μM, respectively. By contrast, the human breast adenocarcinoma and human cervix carcinoma cell lines MCF-7 and HeLa as well as the insect cell *Sf21* were more resilient toward CBT-ol exposure, with IC_{50} values of 42 μM, 51 μM and 68 μM, respectively. No CBT-ol toxicity effects were observed for *E. coli*, *Z. rouxii* and *C. glabrata*.

The relatively low CBT-ol toxicity toward insect cells may support its observed action as an insect repellent and not as an insecticide. Furthermore, the enhanced CBT-ol sensitivity of Gram-positive bacteria (*B. subtilis* and *M. luteus*) might be attributed to the different cell wall composition compared to Gram-negative bacteria, such as *E. coli*. While the Gram-negative cell wall is rather impermeable to lipophilic molecules,[76] the hydrophobic nature of CBT-ol may be better suited to enter Gram-positive bacterial cells. This selective activity may trigger the development of CBT-ol as a specific antibiotic or topical
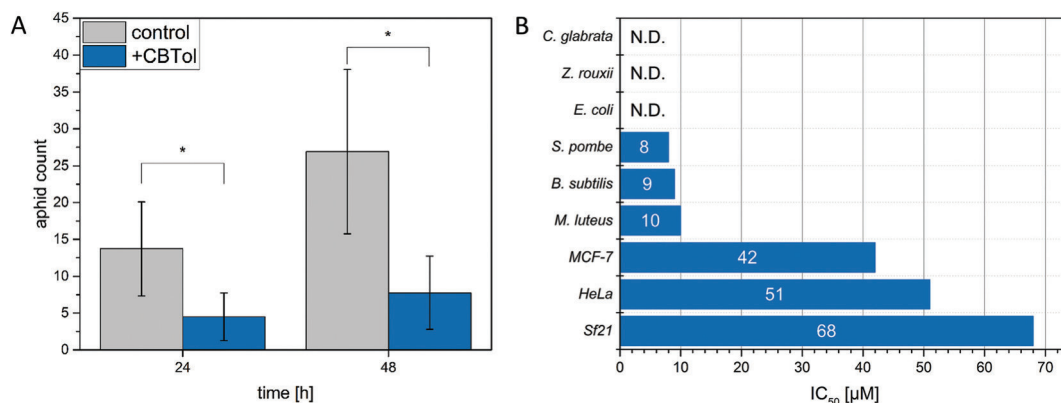
Fig. 6 [A] Biological effect of a 0.25% CBT-ol solution on the colonization behavior of aphids. Treated plants (blue) are compared to a control group (gray), showing that the aphid count was reduced by up to 70%. The results are represented as the mean ± standard deviation of multiple repeated experiments (*$p < 0.005$). [B] Effect of CBT-ol on different cell lines and types represented as the particular IC_{50} values. All experiments were based on a 6-fold approach. The experiments on the cell lines (HeLa; MCF-7; *Sf-21*) were performed two times to calculate the IC_{50} values (N.D.: IC_{50} were not determined due to no measurable activity).

antiseptic preparation against clinically relevant pathogens such as *Bacillus cereus*, *Staphylococcus aureus*, *Streptococcus pneumoniae* and *Listeria monocytogenes*.[77,78]

Experimental

General

Components for cultivation media were obtained from Roth chemicals and Applichem GmbH. Extraction was performed with technical grade ethyl acetate and hexane from Westfalen AG. For all other procedures, the highest purity grade chemicals were used. Acetic acid, acetonitrile, ethyl acetate, ethanol and hexane were obtained from Roth chemicals. $CDCl_3$ was purchased from Sigma Aldrich. Cellic® HTec and Cellic® CTec for enzymatic hydrolysis were obtained from Novozymes and α-amylase (*Bazillus licheniformis*) from Sigma Aldrich. Phusion polymerase for polymerase chain reactions (PCR), restriction enzymes and T4 Ligase were purchased from Thermo Fisher Scientific. DNA preparations kits (Thermo Fisher Scientific) and gel extractions kits (Analytik Jena) were used for DNA preparation.

Cloning

E. coli HMS 174 (DE3) was used for cloning and was cultivated at 37 °C in lysogeny broth (LB) medium. Chloramphenicol (34 μg L^{-1}) and kanamycin (50 μg L^{-1}) were added as required. All primers and genes were synthesized by Eurofins Genomics. Standard protocols were used for PCR and ligation. The RBS calculator software tool[49,51] was used to design and evaluate the relevant ribosomal binding sites (RBS). The following genes were used in this study: isopentenyl pyrophosphate isomerase (idi) from *Haematococcus lacustris* (GenBank: AAC32208.1); geranylgeranyl pyrophosphate synthetase (crtE) from *Pantoea ananatis* (GenBank: ADD79325.1); Terpene Cyclase/Synthase (TPS) from *Nicotiana sylvestris* (GenBank: AAS46038.1); 1-deoxy-D-xylulose-5-phosphate synthase (dxs) from *Escherichia coli* (GenBank: AF035440.1).

All genes were custom synthesized and codon optimized for best possible expression results. All vectors, genes and RBS sequences were verified by sequencing (Eurofins Genomics). The iterative assembly of the synthetic operon receiving a combinatorial library was performed according to the BioBrick principles[53] described by Zelcbuch *et al.*[49] The construction of the synthetic operon was performed in pNiv vectors. The whole operon library was transferred into a pSB4K5 expression vector for production studies. Idi, crtE, dxs, pSB4K5 and pNiv were kindly provided by prof. R. Milo, Weizmann Institute of Science, Israel.

CBT-ol *in vivo* production

75 mL of M9 Minimal Medium supplied with 4 g L^{-1} glucose, 1 mg L^{-1} thiamine and 1 mg L^{-1} biotin containing the required antibiotic in a 250 mL baffled shake flask was inoculated with 1 mL of overnight LB culture of pre-engineered *E. coli* harboring the appropriate plasmid. The cultures were grown at 37 °C at 130 rpm up to an optical density (OD_{600}) of

0.5–0.7 before inducing with IPTG (150 μM) and switching to 22 °C for 2 days.

CBT-ol isolation for analytical purposes

For analytical CBT-ol isolation within the screening process, 30 mL of the selected cultures was lysed by brief sonication. Then, 15 mL of a mixture of ethyl acetate and hexane (1 : 1) was added. The suspension was then strongly shaken for 30 min. A centrifugation step at 8000*g* for 60 s was carried out to separate the phases. A defined volume of the upper organic phase was then sampled and evaporated under a gentle N_2 stream. The residues were resuspended in 1 mL of ethyl acetate and analyzed.

CG-MS and NMR analytics

Extracted terpene products were analyzed by a Trace GC Ultra with DSQII (Thermo Scientific). The sample was loaded by TriPlus AS onto an SGE BPX5 column (30 m, I.D. 0.25 mm, film 0.25 μm). The initial column temperature was set to 50 °C and was maintained for 2.5 min before a temperature gradient of 10 °C min^{-1} to 320 °C was applied. The final temperature was kept for additional 3 min. MS data were recorded at 70 eV (EI) and *m/z* (rel. intensity in %) as total ion current (TIC). The data were collected in full scan mode (*m/z* 50–650). Terpenes were identified by comparison of GC-MS retention times and mass spectra to a commercially available CBT-ol standard (Biomol) and mass spectra data of the NIST Standard Reference Database. Concentrations were quantified correlating the FID peak area to a defined CBT-ol standard of known quantity, ranging from 85 ng to 2.5 mg (see ESI†).

The NMR spectra were recorded in $CDCl_3$ with a Bruker Ascend™ 400 MHz NMR spectrometer. All chemical shifts are relative to $CDCl_3$ at δ = 7.26 (^{1}H-NMR) or $CDCl_3$ at δ = 77.16 (^{13}C-NMR) using the standard δ notation in parts per million.

Wheat bran hydrolysation and media preparation

For the preparation of 1 L of 10% (w/v) biomass hydrolysate solution, 500 mL of 50 mM sodium acetate buffer (pH 5 adjusted with acetic acid) was mixed with 100 g of wheat bran and supplemented with 100 μL of amylase. After incubation for 1 h, the mixture was autoclaved at 121 °C for 15 min. Subsequently, 1% of Cellic® HTec and Cellic® CTec, respectively, were solved in another 470 mL of sodium acetate buffer and centrifuged to separate sediments before sterile filtration with a 0.2 μm filter. The enzyme mixture was added to the wheat bran mash for a final volume of 1 L. After 72 h of hydrolysis at 50 °C and moderate constant shaking, the solid residue was removed using filter paper followed by a sterile filtration step of the liquid phase. For the large scale hydrolysation, 5 kg of wheat bran was added to 35 liters of sodium acetate buffer and transferred to an LP 75 L Bioreactor (Bioengineering). Starch hydrolysis with 5 mL of amylase for one hour at 37 °C was followed by sterilization at 121 °C for 30 minutes with constant stirring. For the enzymatic hydrolysis, 3.0 volume percent (1.5 L) of each enzyme were mixed with the same amount of hydrolysis buffer and pumped directly into the reactor using a

sterile filter (final volume 45 L). The reactor temperature was reduced to 50 °C with constant stirring for 48 hours and a pH of 5. To clarify the hydrolysate, a plate separator was used at 9000 rpm followed by a final sterile filtration step. The final hydrolysate media was based on modified M9 minimal media with a final glucose concentration of 10 g L^{-1} and a pH of 7.

Hydrolysate media

M9 salts mixture	50 mL
60 g Na$_2$HPO$_4$, 30 g KH$_2$PO$_4$,	
5 g NaCl, 10 g NH$_4$Cl, 1 L H$_2$O (pH 7.4)	
1 M MgSO$_4$	250 μL
1 M CaCl$_2$	100 μL
Hydrolysate (10%)	250 mL
Kanamycin monosulfate	50 mg
H$_2$O	Up to 1 L

Fed-batch fermentation including bypass system

The large-scale fermentation was performed in a 75 L Bioreactor (Bioengineering). Initial cultivation volume was set to 32 L using optimized hydrolysate medium (10 g L^{-1} glucose). The pH was set to 7.0 and controlled by 25% ammonium hydroxide and 25% phosphoric acid. We used 2 L of an 8 h culture grown in LB medium for inoculation (final OD$_{600}$ ∼0.1). The feed solution consisted of pure hydrolysate solution (∼40 g L^{-1} glucose). The feed protocol started with 1.5 mL min^{-1} and was increased to 3 mL min^{-1} after 40 h. The total volume of consumed feed solution was 18 L within 120 h. Cultivation temperature was kept constant at 37 °C for the initial growth phase and was reduced to 22 °C after induction with 1 mM IPTG at an OD$_{600}$ of approximately 12. Dissolved oxygen was controlled at 40% air saturation by adjusting the agitation rate (300–530 rpm) at a constant airflow of 10 standard liters per minute of compressed air. Cell density and product output were monitored by periodic sampling.

For the semicontinuous product isolation during the fermentation, a 2 L chromatography column (10 × 24 cm) was applied and filled with 1 kg Amberlite® XAD®-2 beads (Sigma-Aldrich). The whole column was sterilized with 1 M NaOH for three hours and washed with sterile water afterward. The bypass system was mounted to the reactor *via* hose pumps and frits (45 μm pore diameter). The entire reactor volume was constantly pumped through the packed column at a flow rate of 120 mL min^{-1} to prevent blockage. Every 24 hours, the column was washed with 8 L demineralized water followed by a product elution step with 1.5 L ethyl acetate. After a final wash with sterile water (10 L) and sterile air, the column was again incorporated into the fermentation circuit. Ethyl acetate from all elution steps was pooled and evaporated using a rotary evaporator (40 °C, 240 mbar) and the concentrated crude extract was stored at −20 °C.

Terpene purification by CPC

Experiments were performed at room temperature in an SCPC 250 unit from Armen Instrument (now called CPC 250, Gilson Purification SAS, Saint-Avé, France). The unit consists of two single columns connected in series with a total column volume of 182 mL. The maximum achievable rotational speed is 3000 rpm and the whole column can be operated at a maximum pressure drop of 100 bar. The CPC unit was connected to two isocratic preparative HPLC pumps with maximum flowrates of 50 mL min^{-1}, one for filling the column with the stationary phase and the other one for pumping the mobile phase during the separation. The effluent was monitored with a UV detector (ECOM DAD600 2WL 200–600 nm, Prague, Czech Republic) at 210 nm, and a fraction collector (LS 5600, Armen Instrument, France) was used for collecting fractions during the separation. The feed sample was introduced through a six-port manual injection valve. An injection loop of 2 mL was used.

We used the biphasic solvent system acetonitrile/ethanol/ *n*-hexane 27.4/1.6/71 (v/v/v) to separate the concentrated CBT-ol crude extract. The biphasic system was prepared by mixing the respective volume portions of the solvents at room temperature. The mixture was vigorously shaken and equilibrated at room temperature for at least two hours. Afterwards, the two phases were filled in a separation funnel, split and conveyed into two distinct reservoirs. Before use, the phases were degassed in an ultrasonic bath. The lower phase was used as the mobile phase in descending mode. The feed mixture was prepared by dissolving the concentrated CBT-ol crude extract obtained from the bypass adsorptive isolation in the mobile phase (140 mg mL^{-1}). At the beginning of the CPC experiment, the columns were filled with the stationary phase (upper phase) at a flow rate of 40 mL min^{-1}. Afterwards, the rotational speed was set to 1700 rpm and the mobile phase was pumped through the column at 8 mL min^{-1} until no more stationary phase eluted from the column. Then, 2 ml of the feed mixture was injected, the effluent was monitored at a wavelength of 210 nm and fractions were collected every 30 seconds. The CBT-ol-containing fractions were pooled and evaporated under a N$_2$ stream. The purified racemic CBT-ol oil was stored at −20 °C.

Semi-preparative CBT-ol isomer separation

The two isomers α- and β-CBT-ol were separated *via* High-performance liquid chromatography (HPLC: UltiMate™ HPG-3200BX pump, UltiMate™ 3000 AFC and DAD from Thermo Fischer; Jetstream II Plus Column Thermostat; Degasser from Knauer) using a Lux® 5 μm amylose-1, LC column 250 × 10 mm (Phenomenex) and a diode array UV detector. The separation was carried out in an isocratic mode with a mobile phase composed of ACN and H$_2$O (7 : 3) (2.2 mL min^{-1}, 50 °C, 15 min) with active peak fraction collection. The injection volume was 2 mL with a concentration of 20 mg mL^{-1} of CPC-purified racemic CBT-ol oil in ACN and H$_2$O (7 : 3). The chromatogram was recorded at 210 nm wavelength. The corresponding fractions with the separated products were pooled and evaporated under a N$_2$ stream. The pure isomers were stored at −20 °C for further usage.

Bio-insecticide assay

The aphid response to the isolated and purified compound was measured in a two-choice colonization experiment. The compound was applied as 0.25% solutions in ethyl acetate to separated groups of wheat seedlings. The plants were grown in the laboratory at RT in a homemade growing chamber for 7 days on cotton wool. A nebulizer was used for the equal application of the solution. Control seedlings were treated with pure ethyl acetate. The solvent had evaporated completely after 1 hour, and all plants were placed randomly in a test box and the aphids (*Rhopalosiphum padi*) were distributed evenly at a certain distance to all plants. Aphids that settled on the green parts of each group were counted after 24 h and 48 h. The results were statistically analyzed using Student's *t*-test.

Conclusions

We strongly advocate the substitution of non-target discriminant insecticides (*e.g.*, imidacloprid, thiamethoxan and clothianidin) with bio-degradable insect deterrents, such as CBT-ol. The move towards biological insect deterrents represents a fundamental change in crop protection strategies, which would support and expand sustainable agricultural practices. Specifically, we would like to propose that future crop protection methods focus on deterrent rather than lethal active ingredients to protect beneficial and increasingly endangered insect populations, which are essential for pollination of agricultural crops. In this context, the EU commission has recently decided to significantly limit the applicability of toxic insecticides, particularly of the neonicotinamide family in order to enhance protection of bee populations.[17–19]

In this study, we present the heterologous production of the complex diterpene CBT-ol in an engineered *E. coli* production strain. Application of combinatorial RBS adaptation to each biosynthetic bottleneck enzyme provided for an improved protein expression and metabolic flux in the host cell. Moreover, we devised a commercial enzyme system to convert a wheat bran-based milling waste stream into a highly effective *E. coli* fermentation medium. On this waste stream-based growth medium, we obtained CBT-ol titers of 78.9 mg L^{-1} using a technically relevant 50 L bioreactor. To avoid toxic end-product effects on *E. coli* and enable in-process target compound enrichment, we devised an adsorption-based bypass system, which was directly connected to the bioreactor system. We developed a centrifugal partition chromatography (CPC)-based CBT-ol purification strategy that allowed for 95% purity and recovery of the target compound to work up the crude extract eluted from the bypass absorbent. Moreover, this CPC strategy allowed for the reduced use of organic solvents during purification, which rendered the process environmentally benign. To our knowledge, this is the first account of a CPC strategy for the purification of a microbial-generated diterpene. The CBT-ol methodologies disclosed here are entirely scalable

from the laboratory to the technical scale, which provides a direct route for industrial exploitation.[79] The application of chiral HPLC chromatography allowed for the isolation of CBT-ol α- and β-isomers. For the first time, we could demonstrate that the CBT-ol isomer distribution in heterologous *E. coli* systems contrasts the situation in plant systems. Initial aphid avoidance assays indicated that both α- and β-CBT-ol isomers had equivalent effects, demonstrating insect repellent activity. It was confirmed in *in vitro* assays with an insect cell line that CBT-ol was not directly toxic. Therefore, we suggest that CBT-ol has repellent and not direct insecticidal activity. This insect repellent activity may be useful in a cosmetic preparation and is also of particular importance for agricultural applications, as beneficial insects such as bees are not affected by its application. In addition, CBT-ol does not accumulate[34] in soil or aquatic habitats, which contrasts the situation with synthetic insecticides such as neonicotinoids. Most interestingly, we conducted extended cellular toxicity assays including yeast, bacteria and human cancer cell lines and for the first time, we can report that the CBT-ol racemate shows selective toxicity toward Gram-positive bacteria. This observation may trigger further developments of CBT-ol as a selective antibiotic or topical antiseptic toward clinically highly relevant pathogens such as *Staphylococcus aureus*, *Streptococcus pneumoniae* or *Listeria monocytogenes*.[78]

With reference to other reports concerning transfer of laboratory data to industrially applicable processes, starting from a CBT-ol titer of 78.9 mg L^{-1} represents a manageable leap to a commercially viable process in the g L^{-1} scale.[8,80] Currently, genomic integration is frequently reported to enhance both metabolic flux and microbial host productivities in non-terpenoid production systems.[81] However, particularly the integration of plasmid-optimized operons for heterologous terpenoid pathways can result in unpredictable metabolic effects that negatively affect both, protein expression and product titer.[82] Therefore, we would suggest to address other system improvements, such as the evaluation and adaption of different microbial production chassis[70,83] as well as the iterative optimization of the fermentation process towards higher cell densities.[70,71]

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Notes and references

1 D. G. I. Kingston, *Phytochemistry*, 2007, **68**, 1844–1854.

2 S. A. Look, W. Fenical, R. S. Jacobs and J. Clardy, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 6238–6240.

3 R. Edge, D. J. McGarvey and T. G. Truscott, *J. Photochem. Photobiol., B*, 1997, **41**, 189–200.

4 M. A. Fraatz, R. G. Berger and H. Zorn, *Appl. Microbiol. Biotechnol.*, 2009, **83**, 35–41.

5 R. Ciriminna, M. Lomeli-Rodriguez, P. Demma Carà, J. A. Lopez-Sanchez and M. Pagliaro, *Chem. Commun.*, 2014, **50**, 15288–15296.

6 J. Gershenzon and N. Dudareva, *Nat. Chem. Biol.*, 2007, **3**, 408–414.

7 X. Liu, W. Ding and H. Jiang, *Microb. Cell Fact.*, 2017, **16**, 125.

8 E. Jongedijk, K. Cankar, M. Buchhaupt, J. Schrader, H. Bouwmeester and J. Beekwilder, *Appl. Microbiol. Biotechnol.*, 2016, **100**, 2927–2938.

9 C. Willrodt, C. David, S. Cornelissen, B. Bühler, M. K. Julsing and A. Schmid, *Biotechnol. J.*, 2014, **9**, 1000–1012.

10 T. Wriessnegger, P. Augustin, M. Engleder, E. Leitner, M. Müller, I. Kaluzna, M. Schürmann, D. Mink, G. Zellnig, H. Schwab and H. Pichler, *Metab. Eng.*, 2014, **24**, 18–29.

11 N. Simon-Delso, V. Amaral-Rogers, L. P. Belzunces, J. M. Bonmatin, M. Chagnon, C. Downs, L. Furlan, D. W. Gibbons, C. Giorio, V. Girolami, D. Goulson, D. P. Kreutzweiser, C. H. Krupke, M. Liess, E. Long, M. McField, P. Mineau, E. A. D. Mitchell, C. A. Morrissey, D. A. Noome, L. Pisa, J. Settele, J. D. Stark, A. Tapparo, H. van Dyck, J. van Praagh, J. P. van der Sluijs, P. R. Whitehorn and M. Wiemers, *Environ. Sci. Pollut. Res. Int.*, 2015, **22**, 5–34.

12 J. E. Casida and K. A. Durkin, *Annu. Rev. Entomol.*, 2013, **58**, 99–117.

13 D. Goulson and D. Kleijn, *J. Appl. Ecol.*, 2013, **50**, 977–987.

14 V. Girolami, M. Marzaro, L. Vivan, L. Mazzon, C. Giorio, D. Marton and A. Tapparo, *J. Appl. Entomol.*, 2013, **137**, 35–44.

15 A. S. Gunasekara, T. Truong, K. S. Goh, F. Spurlock and R. S. Tjeerdema, *J. Pestic. Sci.*, 2007, **32**, 189–199.

16 M. Chagnon, D. Kreutzweiser, E. A. D. Mitchell, C. A. Morrissey, D. A. Noome and J. P. van der Sluijs, *Environ. Sci. Pollut. Res. Int.*, 2015, **22**, 119–134.

17 D. Butler, EU expected to vote on pesticide ban, *Nature*, 2018, **555**, 150–151.

18 D. Butler, Scientists hail European ban on bee-harming pesticides, *Nature*, 2018, available at: https://www.nature.com/articles/d41586-018-04987-4, accessed 4 May 2018.

19 *Neonicotinoids: risks to bees confirmed | European Food Safety Authority*, available at: https://www.efsa.europa.eu/en/press/news/180228, accessed 4 May 2018.

20 E. Wang, R. Wang, J. DeParasis, J. H. Loughrin, S. Gan and G. J. Wagner, *Nat. Biotechnol.*, 2001, **19**, 371–374.

21 I. Wahlberg and C. R. Enzell, *Beitr. Tabakforsch. Int./Contrib. Tob. Res.*, 1984, **12**, 93–104.

22 C. K. Keene and G. J. Wagner, *Plant Physiol.*, 1985, **79**, 1026–1032. http://www.plantphysiol.org/content/79/4/1026.full.pdf.

23 G. J. Wagner, *Plant Physiol.*, 1991, **96**, 675–679.

24 Y. Saito, H. Takizawa, S. Konishi, D. Yoshida and S. Mizusaki, *Carcinogenesis*, 1985, **6**, 1189–1194.

25 F. Aqil, M. Zahin, K. A. El Sayed, I. Ahmad, K. Y. Orabi and J. M. Arif, *Drug Chem. Toxicol.*, 2011, **34**, 167–179.

26 A. H. Martins, J. Hu, Z. Xu, C. Mu, P. Alvarez, B. D. Ford, K. El Sayed, V. A. Eterovic, P. A. Ferchmin and J. Hao, *Neuroscience*, 2015, **291**, 250–259.

27 N. P. Thao, B. T. T. Luyen, R. Brun, M. Kaiser, P. van Kiem, C. van Minh, T. J. Schmidt, J. S. Kang and Y. H. Kim, *Molecules (Basel, Switzerland)*, 2015, **20**, 12459–12468.

28 I. Wahlberg, I. Wallin, C. Narbonne, T. Nishida and C. R. Enzell, *Acta Chem. Scand.*, 1981, **35b**, 65–68.

29 Z. Guo and G. J. Wagner, *Plant Sci.*, 1995, **110**, 1–10.

30 D. J. Hosfield, Y. Zhang, D. R. Dougan, A. Broun, L. W. Tari, R. V. Swanson and J. Finn, *J. Biol. Chem.*, 2004, **279**, 8526–8529.

31 E. Wang and G. J. Wagner, *Planta*, 2003, **216**, 686–691.

32 R. F. Severson, R. V. W. Eckel, D. M. Jackson, V. A. Sisson and M. G. Stephenson, in *Natural and engineered pest management agents*, ed. P. A. Hedin, J. J. Menn and R. M. Hollingworth, American Chemical Society, Washington, 1994, vol. 551, pp. 172–179.

33 S. Y. Chiang and C. Grunwald, *Phytochemistry*, 1976, **15**, 961–963.

34 D. M. Jackson, R. F. Severson, A. W. Johnson and G. A. Herzog, *J. Chem. Ecol.*, 1986, **12**, 1349–1359.

35 J. Du, Z. Shao and H. Zhao, *J. Ind. Microbiol. Biotechnol.*, 2011, **38**, 873–890.

36 M. C. Y. Chang and J. D. Keasling, *Nat. Chem. Biol.*, 2006, **2**, 674–681.

37 F. Martel, B. Estrine, R. Plantier-Royon, N. Hoffmann and C. Portella, in *Carbohydrates in Sustainable Development I*, ed. A. P. Rauter, P. Vogel and Y. Queneau, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010, vol. 294, pp. 79–115.

38 Z. Merali, S. R. A. Collins, A. Elliston, D. R. Wilson, A. Kasper and K. W. Waldron, *Biotechnol. Biofuels*, 2015, **8**, 23.

39 J. B. Friesen, J. B. McAlpine, S.-N. Chen and G. F. Pauli, *J. Nat. Prod.*, 2015, **78**, 1765–1796.

40 M. Bojczuk, D. Żyżelewicz and P. Hodurek, *J. Sep. Sci.*, 2017, **40**, 1597–1609.

41 A. P. Foucault, *Centrifugal partition chromatography*, Marcel Dekker, New York, 1995, vol. 68.

42 G. F. Pauli, S. M. Pro and J. B. Friesen, *J. Nat. Prod.*, 2008, **71**, 1489–1508.

43 X. Yue and Y. Li, *Synthesis*, 1996, 736–740.

44 F. Zhang, L. Peng, T. Zhang, T. Mei, H. Liu and Y. Li, *Synth. Commun.*, 2003, **33**, 3761–3770.

45 (*a*) H. Uneme, M. Konobe, A. Akayama, T. Yokota and K. Mizuta, Discovery and Development of a Novel Insecticide "Clothianidin", in *R&D Report – SUMITOMO KAGAKU*, 2006, pp. 20–33; (*b*) Y. He, D. Hu, M. Lv, L. Jin, J. Wu, S. Zeng, S. Yang and B. Song, *Chem. Cent. J.*, 2013, **7**,

76; (c) P. Krumlinde, K. Bogár and J.-E. Bäckvall, *J. Org. Chem.*, 2009, **74**, 7407–7410.

46 (a) S. L. Y. Tang, R. L. Smith and M. Poliakoff, *Green Chem.*, 2005, **7**, 761; (b) *Green Chemistry and Catalysis*, ed. R. A. Sheldon, I. W. C. E. Arends and U. Hanefeld, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2007.

47 A. Das, S.-H. Yoon, S.-H. Lee, J.-Y. Kim, D.-K. Oh and S.-W. Kim, *Appl. Microbiol. Biotechnol.*, 2007, **77**, 505–512.

48 W. E. Bentley, N. Mirjalili, D. C. Andersen, R. H. Davis and D. S. Kompala, *Biotechnol. Bioeng.*, 1990, **35**, 668–681.

49 L. Zelcbuch, N. Antonovsky, A. Bar-Even, A. Levin-Karp, U. Barenholz, M. Dayagi, W. Liebermeister, A. Flamholz, E. Noor, S. Amram, A. Brandis, T. Bareia, I. Yofe, H. Jubran and R. Milo, *Nucleic Acids Res.*, 2013, **41**, e98.

50 H. N. Lim, Y. Lee and R. Hussein, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 10626–10631.

51 A. Espah Borujeni, A. S. Channarasappa and H. M. Salis, *Nucleic Acids Res.*, 2014, **42**, 2646–2659.

52 H. M. Salis, E. A. Mirsky and C. A. Voigt, *Nat. Biotechnol.*, 2009, **27**, 946–950.

53 R. P. Shetty, D. Endy and T. F. Knight Jr., *J. Biol. Eng.*, 2008, **2**, 5.

54 O. O. Onipe, A. I. O. Jideani and D. Beswa, *Int. J. Food Sci. Technol.*, 2015, **50**, 2509–2518.

55 L. Favaro, M. Basaglia, W. H. van Zyl and S. Casella, *Appl. Energy*, 2013, **102**, 170–178.

56 E. Curti, E. Carini, G. Bonacini, G. Tribuzio and E. Vittadini, *J. Cereal Sci.*, 2013, **57**, 325–332.

57 A. Singh and N. R. Bishnoi, *Ind. Crops Prod.*, 2013, **41**, 221–226.

58 R. P. Chandra, R. Bura, W. E. Mabee, A. Berlin, X. Pan and J. N. Saddler, *Adv. Biochem. Eng./Biotechnol.*, 2007, **108**, 67–93.

59 J. Yaegashi, J. Kirby, M. Ito, J. Sun, T. Dutta, M. Mirsiaghi, E. R. Sundstrom, A. Rodriguez, E. Baidoo, D. Tanjore, T. Pray, K. Sale, S. Singh, J. D. Keasling, B. A. Simmons, S. W. Singer, J. K. Magnuson, A. P. Arkin, J. M. Skerker and J. M. Gladden, *Biotechnol. Biofuels*, 2017, **10**, 241.

60 A. M. Lopez-Hidalgo, A. Sánchez and A. de León-Rodríguez, *Fuel*, 2017, **188**, 19–27.

61 S. KAJIWARA, P. D. FRASER, K. KONDO and N. MISAWA, *Biochem. J.*, 1997, **324**, 421–426.

62 V. J. Martin, Y. Yoshikuni and J. D. Keasling, *Biotechnol. Bioeng.*, 2001, **75**, 497–503.

63 K. Saiki, T. Mogi and Y. Anraku, *Biochem. Biophys. Res. Commun.*, 1992, **189**, 1491–1497.

64 M. Frederix, F. Mingardon, M. Hu, N. Sun, T. Pray, S. Singh, B. A. Simmons, J. D. Keasling and A. Mukhopadhyay, *Green Chem.*, 2016, **18**, 4189–4197.

65 J. D. Newman, J. Marshall, M. Chang, F. Nowroozi, E. Paradise, D. Pitera, K. L. Newman and J. D. Keasling, *Biotechnol. Bioeng.*, 2006, **95**, 684–691.

66 X. Tang, R. K. Allemann and T. Wirth, *Eur. J. Org. Chem.*, 2017, **2017**, 414–418.

67 W. Wu, F. Liu and R. W. Davis, *Metab. Eng. Commun.*, 2018, **6**, 13–21.

68 P. P. Peralta-Yahya, M. Ouellet, R. Chan, A. Mukhopadhyay, J. D. Keasling and T. S. Lee, *Nat. Commun.*, 2011, **2**, 483.

69 P. Zerbe and J. Bohlmann, *Trends Biotechnol.*, 2015, **33**, 419–428.

70 H. Tsuruta, C. J. Paddon, D. Eng, J. R. Lenihan, T. Horning, L. C. Anthony, R. Regentin, J. D. Keasling, N. S. Renninger and J. D. Newman, *PLoS One*, 2009, **4**, e4489.

71 S. Menendez-Bravo, J. Roulet, M. Sabatini, S. Comba, R. Dunn, H. Gramajo and A. Arabolaza, *Biotechnol. Biofuels*, 2016, **9**, 215.

72 J. Brent Friesen and G. F. Pauli, *J. Liq. Chromatogr. Relat. Technol.*, 2005, **28**, 2777–2806.

73 (a) E. Hopmann, A. Frey and M. Minceva, *J. Chromatogr. A*, 2012, **1238**, 68–76; (b) E. Hopmann, W. Arlt and M. Minceva, *J. Chromatogr. A*, 2011, **1218**, 242–250.

74 M. Elliott, A. W. Farnham, N. F. Janes, P. H. Needham and D. A. Pulman, *Nature*, 1974, **248**, 710–711.

75 H. Y. Ebrahim, M. M. Mohyeldin, M. M. Hailat and K. A. El Sayed, *Bioorg. Med. Chem.*, 2016, **24**, 5748–5761.

76 P. W. Grosvenor, A. Supriono and D. O. Gray, *J. Ethnopharmacol.*, 1995, **45**, 97–111.

77 E. J. Bottone, *Clin. Microbiol. Rev.*, 2010, **23**, 382–398.

78 S. Tchatalbachev, R. Ghai, H. Hossain and T. Chakraborty, *BMC Microbiol.*, 2010, **10**, 275.

79 K. Kemper, M. Hirte, M. Reinbold, M. Fuchs and T. Brück, *Beilstein J. Org. Chem.*, 2017, **13**, 845–854.

80 P. K. Ajikumar, W.-H. Xiao, K. E. J. Tyo, Y. Wang, F. Simeon, E. Leonard, O. Mucha, T. H. Phon, B. Pfeifer and G. Stephanopoulos, *Science*, 2010, **330**, 70–74.

81 J. A. Englaender, J. A. Jones, B. F. Cress, T. E. Kuhlman, R. J. Linhardt and M. A. G. Koffas, *ACS Synth. Biol.*, 2017, **6**, 710–720.

82 J. Alonso-Gutierrez, D. Koma, Q. Hu, Y. Yang, L. J. G. Chan, C. J. Petzold, P. D. Adams, C. E. Vickers, L. K. Nielsen, J. D. Keasling and T. S. Lee, *Biotechnol. Bioeng.*, 2018, **115**, 1000–1013.

83 (a) M. J. Dunlop, Z. Y. Dossani, H. L. Szmidt, H. C. Chu, T. S. Lee, J. D. Keasling, M. Z. Hadi and A. Mukhopadhyay, *Mol. Syst. Biol.*, 2011, **7**, 487; (b) G. H. Han, S. K. Kim, P. K.-S. Yoon, Y. Kang, B. S. Kim, Y. Fu, B. H. Sung, H. C. Jung, D.-H. Lee, S.-W. Kim and S.-G. Lee, *Microb. Cell Fact.*, 2016, **15**, 185; (c) P. J. Westfall, D. J. Pitera, J. R. Lenihan, D. Eng, F. X. Woolard, R. Regentin, T. Horning, H. Tsuruta, D. J. Melis, A. Owens, S. Fickes, D. Diola, K. R. Benjamin, J. D. Keasling, M. D. Leavell, D. J. McPhee, N. S. Renninger, J. D. Newman and C. J. Paddon, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, E111–E118.

3

From Microbial Upcycling to Biology-Oriented Synthesis: Combining Whole-Cell Production and Chemo-Enzymatic Functionalization for Sustainable Taxanoid Delivery

By Max Hirte, Wolfgang Mischko, Katarina Kemper, Simon Röhrer, Claudia Huber, Monika Fuchs, Wolfgang Eisenreich, Mirjana Minceva, and Thomas B. Brück

Synopsis

The article "From microbial upcycling to biology-oriented synthesis: A systems approach for the delivery of functionalized taxanoids" describes a novel strategy for the biotechnological production and purification of taxanoid diterpenoids.

The generation and optimization strategy for the generation of a terpene production platform using lycopene as read out system is described. One major progress and finding in strain optimization was the substitution of the inducible to constitutive protein expression. This step substantially increased TD titers and was transferable without limitations to any production scale. Using corn steep liquor and glycerol as feedstock in the bacterial fermentation process significantly improved the process ecologically and economically. Moreover, a newly devised two step liquid/liquid extraction facilitates handling and efficiency of the TD capture process from high cell density cultures. Screening and optimization of ternary solvent mixtures was conducted in order to identify a stable, biphasic system that can be applied for TD liquid/liquid purification. A solvent mixture containing acetonitrile, ethanol and hexane in specific proportion provided for high purity and recovery rates in CPC purification. This new strategy provides for industrial approved and time saving downstream processing options and it can be scaled up easily. Furthermore, TD was oxygenated by a lipase mediated epoxidation reaction. Our results demonstrate that *in vitro* catalyzed lipase mediated epoxidation is a great possibility to activate carbon bonds in diterpenes that allow for further functionalization steps. We believe that this so-called biology-oriented synthesis strategy is highly relevant for industrial active ingredient production.

Authors contribution

My responsibilities within the research collaboration of the division of biochemistry, bio thermodynamic and the Werner Siemenes-Chair of synthetic biotechnology, was the general coordination, conception, experimental planning and conduction. More specifically, I conducted strain engineering, fermentation process development, analyses and quantification of taxanoids, as well as optimization of the lipase mediated epoxidation reaction. I was supported by Claudia Huber in the elucidation of the taxa-bis-epoxide structure. Together with Wolfgang Mischko and Simon Röhrer, who was the expert for conducting and maintenance of the CPC purification machine, we screened and identified a suitable, biphasic mixtures for TD separation. In addition, Wolfgang Mischko and Monika Fuchs continuously gave scientific advice and supported the project whenever needed.

From microbial upcycling to biology-oriented synthesis: combining whole-cell production and chemo-enzymatic functionalization for sustainable taxanoid delivery†

M. Hirte, [a] W. Mischko, [a] K. Kemper,[a] S. Röhrer,[b] C. Huber,[c] M. Fuchs, [a] W. Eisenreich, [c] M. Minceva [b] and T. B. Brück *[a]

A holistic bio-process based on the microbial upcycling of low-value feedstocks and leading to chemo-enzymatically derived, functionalized taxanoids was established. The upcycling of biotechnological by-product streams glycerol and corn steep liquor by an engineered *Escherichia coli* strain that constitutively expressed bottleneck enzymes within the MEP-pathway led to the formation of 364.4 ± 10.7 mg L^{-1} taxadiene within 44 h. In contrast to standard inducible systems, our constitutive microbial production system provided concomitant growth and taxadiene formation. This strategy is the basis for subsequent continuous taxadiene production processes, which is favorable under economic constraints. The growth dependent taxadiene production showed improved yields, reproducibility and transferability at any scale compared to inducible taxadiene production platforms examined in this study. Additionally, we developed new taxadiene isolation and purification strategies. To that end, these new downstream processing strategies display efficient alternatives to conventional biphasic, *in situ* extraction and purification procedures. Specifically, we developed a rapid and easy two-step extraction procedure followed by centrifugal partition chromatography purification. This process strategy provided 249.0 ± 11.1 mg L^{-1} taxadiene with 95% purity from *E. coli* high cell-density cultures at a liter scale. To functionalize taxadiene, a mild lipase mediated epoxidation reaction was devised. Optimization of this biology-oriented synthesis strategy led to a quantitative conversion of taxadiene to taxa-4(5),11(12)-bisepoxide with an enantiomeric excess of over 83%. The holistic process strategy afforded in excess of 215 mg taxa-4(5),11(12)-bisepoxide by upcycling the low-value feedstocks glycerol (300 g) and corn steep liquor (25 g).

Introduction

Legislative measures for reducing CO_2 emissions, and price fluctuations in fossil fuel resources, drive the development and industrial deployment of sustainable bio-based processes. These sustainable processes should comply with the 12 principles of green chemistry.[1,2] Consequently, by-product streams, such as glycerol from bio-diesel production or corn steep liquor (CSL) from wet corn milling, represent optimal carbon and nitrogen feedstocks for designing new bio-processes. One

[a]Technical University of Munich, Werner Siemens – Chair of Synthetic Biotechnology, Lichtenbergstraße 4, Garching 85748, Germany. E-mail: brueck@tum.de; Tel: +49 -89-289-13253
[b]Technical University of Munich, TUM School of Life Sciences, Biothermodynamics, Maximus-von-Imhof-Forum 2, 85354 Freising, Germany
[c]Technical University of Munich, Chair of Biochemistry, Lichtenbergstraße 4, Garching 85748, Germany
† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8gc03126f

particularly resourceful option for such a process is the use of metabolically engineered whole-cell biocatalysts, such as *Escherichia coli*, which can convert these residual feedstocks into high-value target compounds.[3–6] Additionally, this "microbial upcycling" processing strategy complies with the aim of a zero waste, circular bio-economy.[7]

The family of diterpenoid natural products is structurally and functionally highly diverse. It encompasses several high-value compounds such as the tumor therapeutic Paclitaxel or the fragrance precursor Sclareol.[8,9] Due to their structural complexity and predominantly low natural abundance, synthetic or semi-synthetic production routes are often applied for their industrial-scale production.[10,11] However, synthesis of these compounds is often associated with low product yields and/or significant generation of toxic waste streams.[12,13] Hence, a biotechnological production route represents an attractive and alternative strategy to chemical synthesis.[14–22] In this respect, engineered microbial platforms provide a technically relevant supply of the universal terpene precursors isopentenyl dipho-

sphate and dimethylallyl diphosphate, which in conjunction with expression of prenyl transferases and diterpene synthases, facilitate the production of diterpenes.[23–25] To date, recovery of these heterologously generated diterpenes has almost exclusively relied on organic solvent-mediated *in situ* extraction mediated by a dodecane or decane overlay.[26–28] In comparison with post-cultivation liquid–liquid extraction, this methodology yields significantly higher achievable diterpene titers that can be obtained from 1 L of cell culture.[29,30] However, the applicability of *in situ* extraction method for the diterpene molecules has to be controlled in terms of further downstream recovery, synthetic work-up options and their associated costs. Additionally, the *in situ* extraction method faces the issue of premature compound extraction if *in vivo* functionalization of the diterpene molecule is to be conducted. Furthermore, such diterpene-centered *in vivo* catalyzed transformations often suffer from low enzyme activity or promiscuity. Moreover, product instability and toxicity effect towards to the microbial production system are observed, leading to low target product yields.[31–34] However, functionalization of the diterpene skeleton is mostly required to convey biological activity.[18,35–37]

Therefore, new routes for efficient diterpene recovery and functionalization have to be identified to enable eco-friendly diterpenoid production at economically relevant scales. One production strategy is the initial biotechnological production of a diterpene scaffold combined with subsequent *in vitro* functionalization. In general, such biology-oriented strategies are highly promising in the identification of new lead compounds.[38–40]

One of the most prominent and structurally complex diterpenoids is the clinically relevant anti-tumor agent Paclitaxel.[41] In this regard, the bio-production of the macrocyclic core, which comprises taxadiene (TD) and its hydroxylated congeners (taxanoids), remains challenging to realize at large scales.[14,27,34] In this study, we report a new holistic approach for the sustainable delivery of promising taxanoid diterpenoids that are targeted at pharmaceutical applications. In this context, we demonstrate the possibility of upcycling by-product streams of industrial origin by utilizing glycerol and CSL for the sustainable production of the diterpene TD. We initially engineered *E. coli* strains to express enzyme complements that are required for TD generation. We identified that constitutive expression of those enzymes improve TD yields substantially. Biphasic, *in situ* extraction by the overlay of dodecane was compared to a two-step post-cultivation liquid–liquid extraction. Subsequently, a new purification method for TD enrichment and recovery from high-cell density cultures was established. This demonstrates that terpenoids such as TD can be efficiently purified from cell broth-derived extracts in a single step using centrifugal partition chromatography (CPC). Additionally, a highly specific biology-oriented synthesis approach for purified TD is reported that allows for specific diterpene scaffold functionalization. More specifically, a mild lipase-mediated epoxidation that follows the Prilezahev reaction was established.[42] This reaction provides mass efficient access to a single bis-epoxilated taxanoid molecule that can be further derivatized by various, synthetic strategies.

Results and discussion

Strain engineering

Heterologous production of diterpenes have been conducted in several organisms such as plants, fungi, yeast and bacteria.[43–45] In that regard, *S. cerevisae* and *E. coli* are widely used because they are well studied, easy to manipulate and fast to cultivate. In this study, we choose *E. coli* as production host, as higher diterpene production titers are achievable in shorter time compared to alternative production systems.[21]
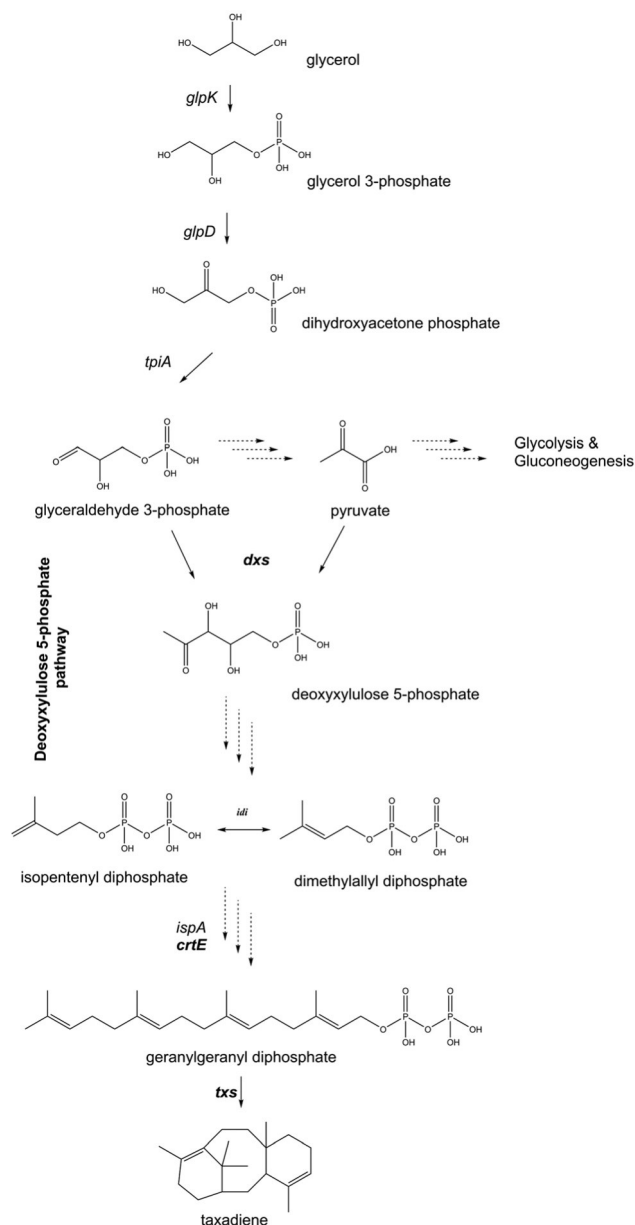
Fig. 1 Biosynthetic pathway from glycerol to TD (overexpressed enzymes are in bold face).

Recent reports have demonstrated that the overexpression of biosynthetic bottleneck enzymes within the deoxyxylulose 5-phosphate pathway (MEP-pathway) provides for significantly improved terpene yields in *E. coli*.[46] In our experimental sets however, the reconstruction of an equivalent inducible synthetic operon (T$_7$-promotor – induction by isopropyl β-D-1-thiogalactopyranoside) resulted in low TD titers in amounts of 8.77 ± 0.5 mg L^{-1} after 24 h (ESI†). This was traced back to reduced cellular viability after induction, which resulted in a short TD production period. Specifically, *dxs* overexpression potentially led to a marked depletion of the central metabolites, pyruvate and glyceraldehyde 3-phosphate, which are the initial precursors of *E. coli*'s terpene biosynthesis (see Fig. 1).

Therefore, we hypothesized that induction-based terpene production limits the potential of whole-cell bio-catalysis systems. To test this hypothesis, we replaced the T$_7$-promoter region by lac-I-derived constitutive promoter systems. Such constitutive expression systems do not rely on an inducer molecule, and continuously synthesize the targeted enzymes. The expression rates of the applied constitutive promoters are much lower compared to T$_7$ promoter systems. Hence, while cells remained in an active growth phase, the metabolic flux from glucose to the desired terpene product remained stable throughout the entire cultivation cycle (see Fig. 2). In addition, switching from an induction-based expression to a constitutive expression system enabled us to efficiently and comparatively screen different synthetic operons for terpene production and productivity (see ESI†). Initially, *txs* was replaced by genes coding for the biosynthesis of the tetraterpene lycopene (*CrtL*, *CrtI*) as both, lycopene as well as TD rely on the diterpene precursor GGDP. Lycopene is a prominent red pigment (*i.e.* tomato and water melon), which is highly suitable for simple visual control of terpene product formation. This system provides for accurate and simple determination of terpene product formation by UV-Vis spectrophotometry and therefore, rapid optimization of heterologous terpene production hosts.

Once a heterologous terpene centered operon has been harmonized, with respect to carbon and redox flux balance as well as product yield using lycopene as a read-out system. To implement a metabolically balance host for the desired terpene production, central genetic elements specific to lycopene production, can simply be exchanged with those coding for taxadiene in our study (see ESI†).[47]

Accordingly, our screening approach identified an operon construct combining the genes *dxs*, *idi* and *crtE* that, if co-transformed with the *ctrL–crtI* operon, yielded amounts of 1.26 ± 0.06 mg$_{Lyc}$ g$_{cdw}$$^{-1}$ lycopene with a growth rate of 0.42 h^{-1} in minimal media. Co-transformation of the *dxs*, *idi* and *crtE* operon with the *txs* coding plasmid yielded significantly improved TD titers with high reproducibility even in shake flask cultures (7.3 ± 0.25 mg$_{TD}$ g$_{cdw}$$^{-1}$). By contrast, our initial, inducible T$_7$-promotor production system yielded only 2.7 ± 0.25 mg$_{TD}$ g$_{cdw}$$^{-1}$ after 48 h. Consequently, we used our constitutive, whole cell expression system for fed-batch fermentation experiments to efficiently generate TD.

Microbial upcycling and TD production

E. coli is capable of metabolizing a wide range of carbon and nitrogen sources.[48,49] Especially low-value feedstocks, like agro-industrial waste streams, are promising sources for the generation of new sustainable processes.[50,51] However, the utilization of alternative feedstocks, specifically with respect to complex nitrogen sources, can lead to significant productivity modulation.[48] In that regard, we investigated TD production in fed-batch fermentations that were supplemented with either yeast extract (YE) or CSL as main nitrogen source in combination with glycerol as main carbon source. Fermentations using YE under low basal protein expression resulted in a growth rate of $\mu_{max} = 0.36 \pm 0.01$ h^{-1} at 30 °C. A TD titer of 274.3 ± 1.0 mg L^{-1} was obtained after 40 h fermentation runtime, and no further changes in TD concentration were detected beyond this time point (see Fig. 3 purple bars). This

Bio-production of terpenes

strategies for whole-cell bio-catalysis

Constitutive expression system
- Stable fermentation conditions
- No inducer required
- Weak expression
- Terpene yield directly correlated to bacterial growth

Inducer

Induction based expression system
- Requires addition of inducer
- Strong expression
- Immediate depletion of energy resources and central metabolites
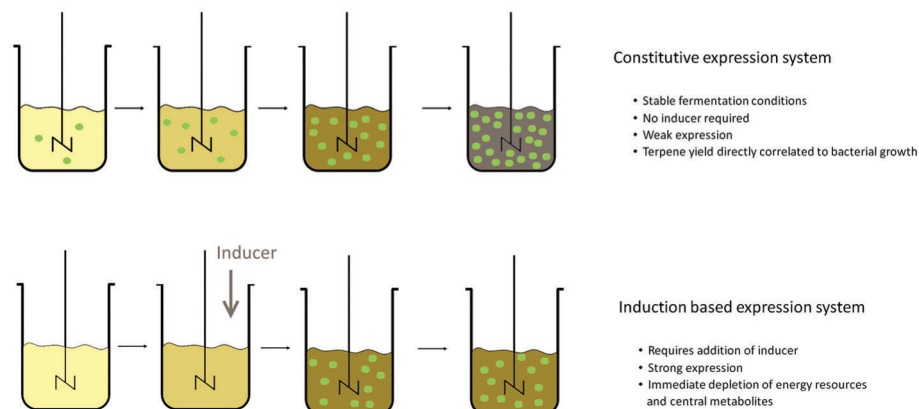
Fig. 2 Constitutive expression and induction–based expression strategy for whole cell biocatalysis of terpenes. Both strategies have been investigated for TD production in this study.
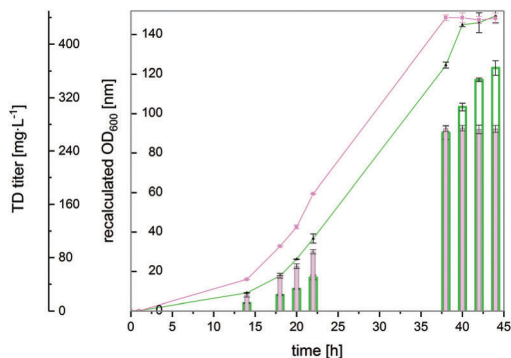
Fig. 3 Time-dependent production of TD in fed-batch fermentations. Fermentation set-ups distinguished between the complex nitrogen sources used (100% YE—purple; mixture of 87.5% CSL and 12.5% YE—green). Lines Represent the calculated OD_{600}; bars represent the TD concentration.

suggests that TD production was directly correlated with bacterial growth. The correlation between cell dry weight (cdw) and TD concentration was calculated and revealed a stable TD amount of 6.65 ± 0.1 mg_{TD} g_{cdw}^{-1}. This correlation was lower than in our shake flask experiment. We hypothesize that the extraction was not complete or TD was actively depleted by aeration or degradation effects in the reaction medium.

In parallel, we investigated whether CSL, which accumulates as a by-product in the wet corn milling process, is a suitable YE replacement for fed-batch TD fermentation processes. Interestingly, the use of CSL as the main nitrogen source resulted in a 32% (w/v) increase in TD production at the end of the fermentation (see Fig. 3 green bar). Although the maximal growth rate ($\mu_{max} = 0.18 \pm 0.01$ h^{-1}) in CSL-based culture was lower than in the YE-based control, a TD concentration of 364.5 ± 10.3 mg L^{-1} was obtained after 44 h. While no active cell growth was observed during the next 3 h, our analyses revealed that the TD concentration increased further during this time. Therefore, we determined the cell dry weight in final samples taken from the fermentation and observed a much lower correlation factor of 0.22 g_{cdw} L^{-1} for an OD_{600} of 1 than during the batch fermentation phase (0.37 g_{cdw} L^{-1}). This effect could be correlated with the high amounts of solid impurities in the fermentation broth derived from the CSL feed solution. In this regard, TD titers were determined in amounts of 9.9 ± 0.6 mg_{TD} g_{cdw}^{-1}.

Additionally, we investigated the economic gain by replacing main parts of YE from the fermentation process by CSL. Interestingly, the cost per liter cultivation (includes batch and feed solution) could be lowered by more than 35% (see ESI†). Altogether, this highlights the benefit of upcycling CSL. Moreover, this approach demonstrates that further medium and fermentation optimization will lead to higher TD titers.

TD capture by *in situ* dodecane overlay and two-step liquid–liquid extraction

Higher alkanes such as decane or dodecane are often used for diterpene *in situ* extraction in biphasic fermentation systems.[26,30]

This diterpene capture process is highly efficient because continuous extraction keeps air-stripping, degradation, and other cultivation-based losses to a minimum.[26,30] However, further processing steps have to be considered and *in situ* extraction strategies have to be evaluated for applicability.

Primarily, the diterpene molecule has to be investigated for its thermal stability, because removing the *in situ* extracting agent is commonly conducted by distillation, which exerts thermal stress on the target compound TD. In parallel, the toxicity and "sustainability" of possible solvents has to agree with the principles of green chemistry. Eventually, in most of the cases a further functionalization step is required to produce biological active diterpenoids. This modification can be carried out *in vivo e.g.* by expressing a cytochrome p450 hydroxylase or *in vitro via* chemically and enzymatically catalyzed reactions respectively. However, if an *in vivo* functionalization step is conducted, premature extraction of the unmodified diterpene molecule drastically lowers titers of functionalized targets when the *in situ* extraction strategy is applied.

Keeping these limitations in mind, we identified dodecane as one promising and suitable *in situ* extraction solvent for diterpenes. TD remained stable under dodecane distillation conditions. Nevertheless, higher alkenes or oils have high boiling points and its or the diterpene molecules separation by distillation would be highly cumbersome and energy intensive. By contrast, lower alkenes like decane, undecane or nonane are toxic and have a negative environmental impact.[52] These extraction reagents can only be applied if they are applied in a minimal fashion or their recovery efficiency exceeds other options. Unfortunately, this cannot be guaranteed in biphasic fermentation systems because solvent loss by aeration or formation of stable emulsions is hardly avoidable. In this regard, we repeated our initial YE and CSL based fermentation process, however, this time using a dodecane overlay of 10% (v/v) fermentation volume, respectively. Interestingly, we did not observe an alteration of the growth rates nor other lasting effects during the fermentation process. This data is in accordance to previous work, which demonstrated that solvents with a $\log P_{OW}$ value above 4 do not interact with *E. coli*'s cell membrane (dodecane $\log P_{OW} = 7$).[53]

As dodecane had no effect on the initial fermentation process, we subsequently analyzed TD formation over time in fed-batch fermentation procedures supplemented either with YE or CSL as main nitrogen source. In the presence of CSL the strong formation of stable emulsion (probably dodecane in combination with free peptides) was observed during the fermentation procedure. This effect prevented an accurate determination of the TD concentration (see ESI†). Accordingly, we did not further investigate the process of biphasic *in situ* extraction in combination with CSL as main nitrogen source as almost none of the applied dodecane could be recovered.

In contrast, we were able to demonstrate the benefit of *in situ* extraction by dodecane within the YE based fermentation (see ESI†). More specifically, the TD concentration per cell dry weight (cdw) increased by more than 40% (w/w), affording

9.4 ± 0.5 mg$_{TD}$ g$_{cdw}{}^{-1}$. After separating the dodecane phase from the aqueous fermentation broth, we evaporated the dodecane phase to prepare TD for further purification. In this regard, we observed that TD exhibits significant stability (170 °C, <50 mbar, >30 min). After evaporation a very low TD loss of 7% (v/v) was determined. Therefore, we presume that TD is a very stable molecule and that the dodecane mediated TD *in situ* extraction process is a suitable extraction procedure depending on the constituents of the fermentation media. Concerning *in situ* extraction production process the depletion of hydrophobic cultivation media components by the overlay is expected.

Nevertheless, recovery of the dodecane solvent turned out to be challenging. Only 70% (v/v) of the initially applied dodecane could be recaptured (Table 1).

Accordingly, the greenness and sustainability of such a biphasic fermentation process is questionable. Particularly, the spent culture after extraction must be treated as special waste as dodecane evaporates much slower than water under normal conditions. The dodecane loss during the extraction and solvent recovery step contributed over 75% to the total costs of the fermentation process. However, the processing time savings using this *in situ* extraction method lowers personal expenses. Therefore, with regard to economic efficiency the dodecane extraction process must be investigated individually by a thorough techno-economic analysis, which is not focus of this manuscript.

In our alternative TD capture strategy, we devised a convenient procedure based on classical liquid–liquid extraction. The efficiency of liquid–liquid extraction is commonly low, and high product recovery requires repeated and elaborate extraction procedures. As fermentatively generated diterpenes are located in both the supernatant as well as in the cell pellet fractions, the extraction of the entire fermentation broth is required to maximize product yields.[30]

To that end, our high cell-density culture broth required special handling procedures to enable efficient extraction. In order to minimize the effort of extraction, we developed a simple two-step procedure. Initially, the entire cell-containing fermentation broth (1 L) was treated with an admixture of ethyl acetate and ethanol (1 L) that led to cell lysis and liberation of intracellularly accumulated TD. Subsequently, the lysed cell debris were removed by centrifugation. Hexane (500 mL) was added to the resulting supernatant, which induced a phase separation and a reduction of the sample volume (750 mL). This step-wise TD extraction circumvented the formation of a highly stable emulsion as observed with dodecane extraction. In our experience this emulsion is formed only when solvent assisted terpene extraction is conducted in a single step from high cell-density fermentation broths. Notably, our two-step extraction protocol resulted in a 71% (w/v) TD recovery at liter scale. This TD yield could potentially be increased even further by repeated hexane extraction. However, the excessive solvent handling does not justify these extended efforts to maximize TD yield at this point. Moreover, the possibility of fractionated solvent recovery improves the sustainability aspect of classical liquid–liquid extraction. Albeit our liquid–liquid extraction is superior due to the minimization of emulsion formation. Furthermore, it is also suitable for thermally labile or functionalized terpenes.

TD purification by CPC

For further TD purification, the extracts obtained, whether by *in situ* dodecane overlay or by post-cultivation liquid–liquid extraction, were evaporated until an oily mixture that predominantly consisted of lipids with a TD content below 10% (w/v) remained. The initial purification strategy involved an isocratic (100% acetonitrile (w/v)) reversed phase HPLC (RP-HPLC) purification method where the oily mixture was mixed with acetonitrile. However, this re-solubilization approach resulted in the formation of immiscible oil drops/resins that incorporated a substantial amount of TD. Only significant dilution with more acetonitrile led to complete dissolution of the mixture. Thus, many feed injections would have been necessary to completely purify the produced TD. The screening of other solvent systems that are classically used in RP-HPLC protocols did not solve this admixing issue. Therefore, we set out to develop an alternative downstream technology, due to this universal issue concerning diterpene purification from microbial-derived extracts. Eventually, we devised a new CPC-based purification protocol inspired by sesquiterpenoid purification from plant material.[54] This liquid–liquid chromatographic separation technology uses two immiscible liquid phases, one serving as a mobile and the other as a stationary phase. The mobile phase is pumped through the stationary phase, whereas the latter is kept in the column by applying a centrifugal force field. The separation of the applied sample is achieved by differential partitioning of the mixture compounds between

Table 1 Evaluation of both strategies investigated in this study for TD extraction from 1 L fermentation broth (*in situ* product removal, two step liquid–liquid extraction)

		Extraction strategies 1 L TD fermentation broth		
		In situ (YE)	Liquid–liquid (YE)	Liquid–liquid (CSL)
Solvents [L] and Recovery [%]	Dodecane	0.1 70	—	—
	Hexane	—	0.5 90	0.5 90
	Ethyl acetate	—	0.5 n.d.	0.5 n.d.
	Ethanol	—	0.5 n.d.	0.5 n.d.
Evaporation	Solvent volume [L]	0.07	0.75	0.75
	Energy required [kJ]	~16	~200	~200
	TD recovery [%]	93	99	99
Results	Expenses solvent	+++	+	+
	Special waste	+++	+	+
	Time	+	++	++
	Produced TD [mg]	~394	~274	~364
	Extracted TD [mg]	~256	~195	~260

the two phases. In order to achieve sufficient separation in terms of resolution, productivity and solvent consumption, a biphasic system in which the partition coefficient (K) of the target compound is in the range of $0.5 < K < 2.5$ is preferred.[55] Initially, various biphasic solvent systems were screened. The predictive thermodynamic model "Conductor-like Screening Model for Realistic Solvation" was used for the pre-selection of a suitable solvent system in order to reduce experimental effort as described by Hopmann *et al.*[56,57] On the basis of the molecular structure of TD, K-values were predicted in commonly used CPC solvent systems, including hexane/ethyl acetate/methanol/water, heptane/ethyl acetate/methanol/water, butanol/methanol/water, hexane/ethyl acetate/acetonitrile, and hexane/ethanol/acetonitrile. Various compositions of these systems were screened in order to cover a wide polarity range. In that regard, a hexane/ethanol/acetonitrile biphasic system was computationally selected as a prime target for TD purification. Subsequently, an experimental evaluation using shaker flask experiments was performed and the respective TD K-values determined for different solvent compositions. An *n*-hexane/ethanol/acetonitrile system with proportions of 4.33/1.0/1.07 (v/v/v) was ultimately selected and applied for TD separation by CPC. The stationary phase retention (SF) determined for this system was 0.56 at a mobile phase flow rate of 8 mL min^{-1}. After sample injection, constant stationary phase loss (2 mL min^{-1}) was detected over the entire runtime. The collected fractions were analyzed by GC-MS. TD eluted between minute 43 and 48, and TD purity and recoveries were determined by GC-flame ionization detection (FID). The final purification process produced TD that was 95% pure with a recovery of 95% (Fig. 4).

This purification process allowed for rapid and efficient isolation of the sample diterpene TD. Specifically, sample dissolution with the lower phase of the biphasic system did not lead to the incorporation of the diterpene molecule into immiscible oil droplets. This significantly raised the process efficiency because of the increased TD mass load in CPC purification compared with our initial RP-HPLC separation

process. Additionally, although irreversible adsorption of the diterpene macrocycle scaffold to solid phases is particularly responsible for losses in RP-HPLC diterpene purification, this does not apply to CPC, resulting in improved recovery. Therefore, we recommend CPC as an alternative to the classically used RP-HPLC purification processes for diterpene purification from high cell density fermentation broth.[22]

Lipase-mediated epoxidation of TD

Most of the terpenoids mentioned in the dictionary of natural products harbor at least two oxygen-containing functional groups.[37] These groups, in conjunction with the high stereo complexity and density of the carbon atoms within diterpenoids, commonly convey biological functions.[18,35,36] Notably, natural diterpene epoxides such as dictyoepoxide, dollabellanone, and calyculone have been found to act as a vasopressin receptor agonist, an anti-protozoal agent, and a cytotoxic agent, respectively.[58–60] Moreover, epoxides can be chemically diversified further into ketones, alcohols (diols, thiols, amino alcohols, *etc.*), and polymers, thereby expanding the chemical space.[61–64] To date, epoxidation of monoterpenes has been mainly investigated because of their abundant availability in plant extracts.[65–69] However, Dolabellatriene, a natural diterpene found in algae and corals, was also recently reported to be a suitable target for a lipase-mediated epoxidation reaction.[70] In contrast to chemical epoxidation protocols, enzymatically mediated reactions are highly specific and can be conducted under milder conditions.[68,69,71]

In this study, we investigated a lipase (*CalB*)-mediated TD epoxidation reaction. In order to slowly release H_2O_2 into the reaction medium, we used environmentally compatible urea-hydrogen peroxide.[68] As depicted in Fig. 5, acetic acid primarily reacts with hydrogen peroxide to form a per-oxo acid. This per-oxo-acetic acid targets the olefinic bonds of TD either from the *re*- or *si* face, which theoretically would results in 4 stereo-isomers. The epoxidation reaction was monitored by time-resolved GC-FID analysis of samples that were regularly drawn from the reaction vessel (0.75 h intervals). Detailed
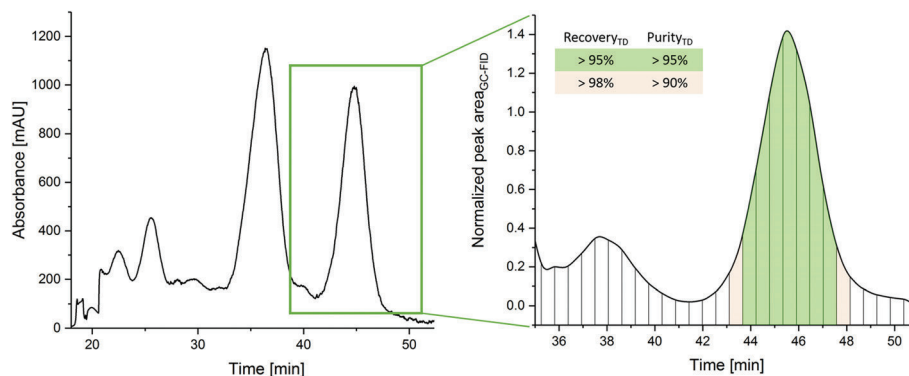
Fig. 4 Left: On-line chromatogram at 210 nm of a centrifugal partition chromatography batch seperation process of the TD-containing crude extract that was obtained from *E. coli* fed-batch fermentation. The solvent system of *n*-hexane/acetonitrile/ethanol/4.33/1.0/1.07 (v/v/v) was applied (descending mode: lower phase as mobile phase, 8 mL min^{-1} mobile phase flow rate, 1700 rpm, cinj TD ≈ 16 mg mL^{-1}, vinj = 2 mL). Right: Off-line chromatogram of the fractions in close proximity to the TD retention time reconstructed by GC-FID analyses.
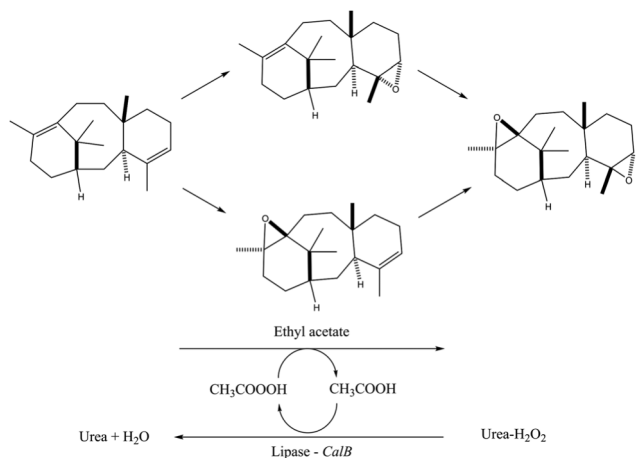
Fig. 5 Reaction mechanism from TD to bis-epoxilated taxanoids by lipase (*CalB*) mediated epoxidation reaction.

Fig. 7 Possible routes for expanding the taxanoid chemical space based on our holistic process for TD and taxa-4(5),11(12)-bisepoxide molecules production.
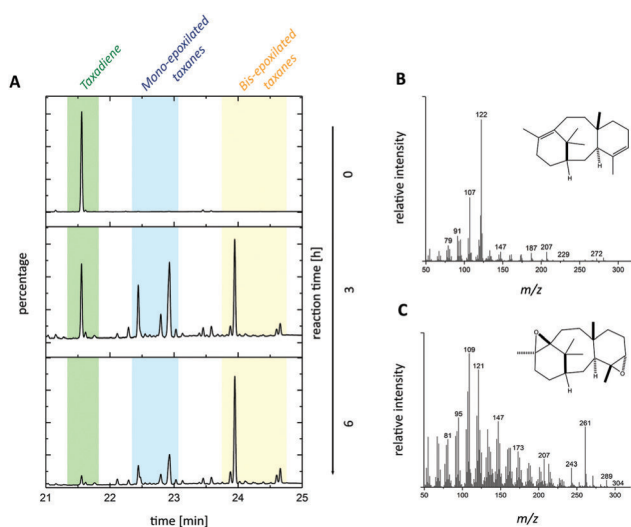

Fig. 6 A: Time-resolved epoxidation reaction from TD (green), over mono-epoxilated taxanoids (blue) to bis-epoxilated taxanoids (yellow). B: MS-spectra of TD. C: MS-spectra of the major product taxa-4(5),11 (12)-bisepoxide.

GC-MS analyses of the sample revealed the formation of mono-epoxides (m/z 288) prior to a second epoxidation step that generated bis-epoxilated taxanoid compounds (m/z 304) (see Fig. 6).

Interestingly, we noticed a predominant formation of a single bis-epoxid species that was isolated by thin layer chromatography and subsequently structurally characterized. The 1D and 2D-NMR spectroscopy data of this compound were consistent with reported chemical shifts of taxa-4(5),11(12)-bisepoxide (ESI†).[71]

These GC-MS and NMR datasets indicate that the complex stereochemistry of TD and this reaction set-up resulted in the generation of taxa-4(5),11(12)-bisepoxide as the major product. Subsequently, we fine-tuned the reaction conditions towards
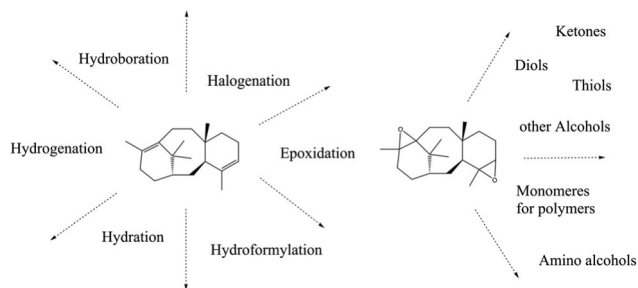
the exclusive production of this molecule. The enantiomeric excess was increased from an initial 58.3% to 83.1% simply by lowering the reaction temperature from 30 °C to 20 °C (ESI†). Under these conditions, TD conversion was complete after 6 h and *in situ* product stability was monitored for a period of another 2 h. TD oxy-functionalization outside the cellular environment can circumvent common issues that are inherent to the *in vivo* targeted taxanoid production. In that regard, intracellular hydroxylation conventionally leads to a dramatic decrease in taxanoid titers due to low enzyme transformation rates and to decreased viability from the enhanced metabolic burden.[27] Moreover, the heterologous expression of the native cytochrome p450 5-α-TD hydroxylase leads to a diverse range of products, which complicates the efficient production of single taxanoids in whole-cell biocatalysis approaches and reduces the yield of target compounds.[72]

In addition, the epoxides generated in this study are highly reactive, and their chemical space can be easily diversified in order to obtain new taxanoid structures with potential pharmacological activities (see Fig. 7).

Conclusions

This study describes a fast and sustainable fermentative production of the structurally complex natural compound TD and its subsequent chemo-enzymatic functionalization. In order to provide practical access to taxanoid molecules, initially different TD extraction methods from *E. coli* high cell-density cultures were evaluated and optimized with a particular focus on further recovery and purification options. Subsequently, a new purification process for diterpenes from biotechnological fermentation broth using CPC was developed that resulted in high TD recovery and purity. The obtained TD was further functionalized *in vitro* by a lipase mediated epoxidation reaction. Most notably, this epoxy-functionalization approach resulted in the highly specific generation of taxa-4(5),11(12)-bisepoxide that can be easily converted to a diverse array of taxanoids using standard chemical reactions (see Fig. 7). This situation contrasts with the commonly applied *in vivo* cytochrome P450 hydroxylation approaches that usually exhibit low reaction rates and the generation of multiple products.

All the process steps involved in TD generation, purification and functionalization can be transferred to commercially relevant scales, with a particular focus on economic and ecological constraints. For the first time, the processes reported here facilitate the ready availability of the TD molecule and its functionalized congeners. The technology platform described in this study enables a biology-oriented synthesis of *ad hoc* designed bioactive taxanoids.

Experimental

Materials

Fermentation media components were obtained at highest purity from Roth chemicals and Applichem GmbH. CSL was purchased from Sigma-Aldrich. Extraction was performed with technical grade ethanol, ethyl acetate, and hexane from Westfalen AG. For all other procedures, highest purity grade chemicals were used. In this context, acetic acid, acetonitrile, chloroform, ethyl acetate, and hexane were obtained from Roth chemicals. Immobilized Lipase B from *Candida Antarctica* (*CalB*), $CDCl_3$, isopropyl-β-D-thiogalactopyranoside and urea-hydrogen peroxide were purchased from Sigma-Aldrich.

Bacterial strains and growth conditions

All strains used were obtained from Merk Millipore. For cloning, *E. coli* HMS 174 (DE3) was used and grown at 37 °C in Luria–Bertani medium. The screening of the constructed lycopene production platform library is described in the ESI.†

E. coli BL21 (DE3) grown in R-Media[73] supplemented with 10 g L^{-1} glycerol instead of sucrose and 12 g L^{-1} YE or 10.5 g L^{-1} CSL in combination with 1.5 g L^{-1} YE was used for TD production at 30 °C (shake flask experiments and fermentation batch medium). Induction of T_7-promoted TD production was initiated at OD_{600} = 0.2 by the addition of 0.1 mM isopropyl-β-D-thiogalactopyranoside. Chloramphenicol (34 μg mL^{-1}) and Kanamycin (50 μg mL^{-1}) were added as required.

Cloning

Genes coding for deoxyxylulose 5 phosphate synthase (*dxs*; GenBank: YP001461602.1), isopentenyl-diphosphate delta isomerase (*idi*; GenBank: AAC32208.1), geranylgeranyl diphosphate synthase (*crtE*; GenBank: KPA04564.1), taxadiene synthase (*txs*; GenBank: AAK83566.1), phytoene desaturase (CrtB; GenBank: AHG94990.1) and Phytoene synthase (CrtL; GenBank: AHG94989.1) were used.[74] The construction of polycistronic operons were conducted by standard BioBrick cloning and adjusted as described previously.[75,76] *Dxs* was amplified from original sources by PCR. The other genes were synthesized by Life Technologies GmbH, and the codon usage was optimized for *E. coli*. Primers were obtained by Eurofins Genomics, and plasmids were obtained from Novagen/Merk Millipore. The Operons were set under constitutive expression by lac-I-derived promoters.[77]

Analytics and TD quantification

Analysis of TD was performed using a Trace GC-MS Ultra with DSQII (Thermo Scientific). One microliter (1/10 split) of the sample was loaded using a TriPlus auto sampler onto an SGE BPX5 column (30 m, I.D 0.25 mm, film 0.25 μm). The injection temperature was 280 °C, and helium was used as the carrier gas. Initial column temperature was set to 50 °C and maintained for 2.5 min before a temperature gradient at 10 °C min^{-1} up to 320 °C was applied. The final temperature was held for an additional 3 min. MS data were recorded at 70 eV (EI) as total ion current, and *m/z* was given as relative intensity (%). The recorded *m/z* range was between 50 and 650. The same GC protocol was followed for FID.

NMR spectra were recorded in $CDCl_3$ with an Avance-III 500 MHz (Bruker) at 300 K. ^{1}H NMR chemical shifts are given in ppm relative to $CDCl_3$ (δ = 7.26 ppm). ^{13}C NMR chemical shifts are given in ppm relative to $CDCl_3$ at δ = 77.16. The 2D experiments (HSQC, HMBC, COSY, and NOESY) were performed using standard Bruker pulse sequences and parameters.

For TD quantification, the purified compound was evaporated to dryness, weighed and dissolved at different concentration in hexane before analyzed by GC-FID. A standard curve of TD concentration over the respective GC-FID peak area was produced, and a correlation factor was calculated (ESI†).

Fermentation

Fermentation was performed in a DASGIP® 1.3 L parallel reactor system (Eppendorf AG). An overnight pre-culture was used for the inoculation of fermenters (OD = 0.1) that differed in the supplemented nitrogen source. The cultivation temperature was kept constant at 30 °C. Initial stirring velocity and airflow was set to 200 rpm and to 0.2 volumes of air per volume of medium per min (vvm), respectively. Dissolved oxygen was kept at 30% by successive increases of the stirrer velocity, the oxygen proportion, and eventually the airflow. A pH value of 7.00 was controlled by the addition of 25% aqueous ammonia. A pH value shift above 7.05 initiated a feed shot of 40 mL. The feed solution consisted of 600 g L^{-1} glycerol, 20 g L^{-1} $MgSO_4·7H_2O$, 2 mg L^{-1} thiamine–HCl, 16 mL 100× trace elements solution (5 g L^{-1} EDTA; 0.83 g L^{-1} $FeCl_3$-6·H_2O; 84 mg L^{-1} $ZnCl_2$, 13 mg L^{-1} $CuCl_2$-2·H_2O, 10 mg L^{-1} $CoCl_2$-2·H_2O, 10 mg L^{-1} H_3BO_3, and 1.6 mg L^{-1} $MnCl_2$-4·H_2O), and 40 g L^{-1} complex nitrogen source (pH = 7.00). Complex nitrogen sources besides 100% YE consisted of 87.5% CSL mixed with 12.5% YE. Samples were taken at different time points to determine the OD_{600} and the TD content.

TD extraction and its preparation for the purification process

In situ extraction was conducted by the addition of 10% (v/v) dodecane after 25 h of cultivation time. TD quantification was conducted by analyzing the TD content in the dodecane phase. Therefore, samples taken were centrifuged for 2 min at 10 000 rpm and 100 μL of the upper phase taken, diluted 1 : 10 with hexane and analyzed by GC-FID.

After the fermentation was terminated the whole fermentation broth was centrifuged for 20 min at 10 000 rpm. The dodecane phase was separated by a separation funnel. The recovered dodecane was completely evaporated prior to re-solubilization.

For the other extraction strategy investigated, samples were regularly taken from the fermenter in the range from 0.5 to 10 mL, depending on the OD_{600}, and diluted as required to a final volume of 10 mL. Thirty milliliters of a solvent mixture that consisted of ethanol, ethyl acetate, and hexane (1 : 1 : 1) (v/v/v) was added and vigorously mixed for 1 h at 30 °C. The solution was centrifuged for 2 min at 10 000 rpm. Subsequently, the upper organic phase was directly analyzed by GC-FID for TD quantification.

For the extraction of the whole cultivation broth, equal amounts of ethyl acetate and ethanol were added to the cell culture in a ratio of 1 : 1. This first extraction step was performed for 12 h at 30 °C on a rotary shaker (60 rpm). Subsequently, the extract was centrifuged for 15 min at 7000 rpm, the supernatant separated from the pelleted cell debris, and hexane added (25% of the supernatant volume). The mixture was shaken for an additional 3 h before the phases were separated using a separation funnel. The upper phase was evaporated until an oily resin remained.

For the re-solubilization and preparation for CPC purification of the oily resin, a biphasic system consisting of n-hexane/ethanol/acetonitrile 4.33/1.0/1.07 (v/v/v) was prepared and separated by a separation funnel. 20 mL of the lower phase was mixed with the TD enriched oil (approximately 5 mL) and subsequently centrifuged for 5 min at 12 000 rpm. The soluble fraction was applied for CPC purification.

Purification of TD

Purification was performed at room temperature in a SCPC 250 unit from Armen Instrument (now called CPC 250, Gilson Purification SAS). Two preparative HPLC pumps (maximal flow rates of 50 mL min^{-1}) were used for filling the column with stationary phase and for pumping the mobile phase during the separation process, respectively. The CPC column volume was 182 mL. The feed sample was introduced through a six port manual injection valve and a 2 mL injection loop was used. The effluent was monitored with a UV-Detector (ECOM DAD600 2WL 200–600 nm) at 210 nm.

The biphasic liquid system, n-hexane/ethanol/acetonitrile 4.33/1.0/1.07 (v/v/v), was prepared at room temperature (22 °C) by mixing the corresponding volume portions of the solvents. The mixture was vigorously shaken and equilibrated for at least 2 h before being split with a separation funnel.

The CPC column was filled with the upper (stationary) phase, and the instrument's rotational speed was set at 1700 rpm. The lower phase (mobile) phase was pumped with a flow rate of 8 mL min^{-1} until no more stationary phase eluted from the column, *i.e.* hydrodynamic equilibrium was achieved. Subsequently, CPC separation was initiated by an injection of 2 mL of sample while a 1700 rpm rotational speed and a mobile phase flow rate of 8 mL min^{-1} in descending mode

were applied. Fractions were collected during the separation process using a fraction collector (LS 5600, Armen Instrument) and analyzed by GC-FID.

Lipase-mediated epoxidation of TD

For the epoxidation reaction 1 mg mL^{-1} TD, 0.1 mg mL^{-1} acetic acid, and 2 mg mL^{-1} immobilized *CalB* were added to 5 mL ethyl acetate. In order to keep temperature constant at 20 °C, 25 °C, or 30 °C, the reaction vessel was placed in a thermoblock (Eppendorf AG). Reactions were initiated by the addition of 2 mg mL^{-1} urea-hydrogen peroxide, which was dissolved in ethanol as a stock solution of 100 mg mL^{-1}, before shaking at 1000 rpm. Samples were taken after different time points, and the reaction monitored over time by GC-MS and GC-FID analysis.

Conflicts of interest

The authors declare no conflict of interests.

Acknowledgements

Notes and references

1 R. A. Sheldon, I. W. Arends and U. Hanefeld, Introduction: Green Chemistry and Catalysis, in *Green Chemistry and Catalysis*, ed. R. A. Sheldon, I. W. Arends and U. Hanefeld, Wiley-VCH Verlag GmbH & Co. KGaA, 2007, pp. 1–47, DOI: 10.1002/9783527611003.ch1.

2 S. L. Y. Tang, R. L. Smith and M. Poliakoff, *Green Chem.*, 2005, **7**, 761–762.

3 S. S. Yazdani and R. Gonzalez, *Curr. Opin. Biotechnol.*, 2007, **18**, 213–219.

4 Q. Ye, X. Li, M. Yan, H. Cao, L. Xu, Y. Zhang, Y. Chen, J. Xiong, P. Ouyang and H. Ying, *Appl. Microbiol. Biotechnol.*, 2010, **87**, 517–525.

5 A. Murarka, Y. Dharmadi, S. S. Yazdani and R. Gonzalez, *Appl. Environ. Microbiol.*, 2008, **74**, 1124–1135.

6 R. W. Liggett and H. Koffler, *Bacteriol. Rev.*, 1948, **12**, 297–311.

7 P. Lacy and J. Rutqvist, *Waste to Wealth: The Circular Economy Advantage*, Palgrave Macmillan, UK, 2016.

8 D. G. I. Kingston, A. A. Molinero and J. M. Rimoldi, in *Fortschritte der Chemie organischer Naturstoffe/Progress in the Chemistry of Organic Natural Products*, ed. W. Herz, G. W. Kirby, R. E. Moore, W. Steglich and C. Tamm, Springer Vienna, Vienna, 1993, pp. 1–165, DOI: 10.1007/978-3-7091-9242-9_1.

9 M. Schalk, L. Pastore, M. A. Mirata, S. Khim, M. Schouwey, F. Deguerry, V. Pineda, L. Rocci and L. Daviet, *J. Am. Chem. Soc.*, 2012, **134**, 18900–18903.

10 L. M. T. Frija, R. F. M. Frade and C. A. M. Afonso, *Chem. Rev.*, 2011, **111**, 4418–4452.

11 D. J. Jansen and R. A. Shenvi, *Future Med. Chem.*, 2014, **6**, 1127–1148.

12 R. Janke, M. Fuchs, M. Christmann, T. Brück and B. Loll, *BIOspektrum*, 2017, **23**, 709–711.

13 P. J. Dunn, A. Wells and M. T. Williams, *Green Chemistry in the Pharmaceutical Industry*, Wiley, 2010.

14 P. K. Ajikumar, W.-H. Xiao, K. E. J. Tyo, Y. Wang, F. Simeon, E. Leonard, O. Mucha, T. H. Phon, B. Pfeifer and G. Stephanopoulos, *Science*, 2010, **330**, 70–74.

15 M. C. Y. Chang and J. D. Keasling, *Nat. Chem. Biol.*, 2006, **2**, 674–681.

16 C. J. Paddon, P. J. Westfall, D. J. Pitera, K. Benjamin, K. Fisher, D. McPhee, M. D. Leavell, A. Tai, A. Main, D. Eng, D. R. Polichuk, K. H. Teoh, D. W. Reed, T. Treynor, J. Lenihan, H. Jiang, M. Fleck, S. Bajad, G. Dang, D. Dengrove, D. Diola, G. Dorin, K. W. Ellens, S. Fickes, J. Galazzo, S. P. Gaucher, T. Geistlinger, R. Henry, M. Hepp, T. Horning, T. Iqbal, L. Kizer, B. Lieu, D. Melis, N. Moss, R. Regentin, S. Secrest, H. Tsuruta, R. Vazquez, L. F. Westblade, L. Xu, M. Yu, Y. Zhang, L. Zhao, J. Lievense, P. S. Covello, J. D. Keasling, K. K. Reiling, N. S. Renninger and J. D. Newman, *Nature*, 2013, **496**, 528–532.

17 J. Yang and Q. Nie, *Microb. Cell Fact.*, 2016, **15**, 74.

18 C. Khosla and J. D. Keasling, *Nat. Rev. Drug Discovery*, 2003, **2**, 1019–1025.

19 E. Leonard, P. K. Ajikumar, K. Thayer, W.-H. Xiao, J. D. Mo, B. Tidor, G. Stephanopoulos and K. L. J. Prather, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 13654–13659.

20 D. Morrone, L. Lowry, M. K. Determan, D. M. Hershey, M. Xu and R. J. Peters, *Appl. Microbiol. Biotechnol.*, 2010, **85**, 1893–1906.

21 K. Kemper, M. Hirte, M. Reinbold, M. Fuchs and T. Brück, *Beilstein J. Org. Chem.*, 2017, **13**, 845–854.

22 W. Mischko, M. Hirte, S. Roehrer, H. Engelhardt, N. Mehlmer, M. Minceva and T. Brück, *Green Chem.*, 2018, **20**, 2637–2650.

23 Y. Gao, R. B. Honzatko and R. J. Peters, *Nat. Prod. Rep.*, 2012, **29**, 1153–1175.

24 Y. Yamada, T. Kuzuyama, M. Komatsu, K. Shin-ya, S. Omura, D. E. Cane and H. Ikeda, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 857–862.

25 G. Bian, Z. Deng and T. Liu, *Curr. Opin. Biotechnol.*, 2017, **48**, 234–241.

26 J. D. Newman, J. Marshall, M. Chang, F. Nowroozi, E. Paradise, D. Pitera, K. L. Newman and J. D. Keasling, *Biotechnol. Bioeng.*, 2006, **95**, 684–691.

27 B. W. Biggs, C. G. Lim, K. Sagliani, S. Shankar, G. Stephanopoulos, M. De Mey and P. K. Ajikumar, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 3209–3214.

28 P. K. Ajikumar, W.-H. Xiao, K. E. Tyo, Y. Wang, F. Simeon, E. Leonard, O. Mucha, T. H. Phon, B. Pfeifer and G. Stephanopoulos, *Science*, 2010, **330**(6000), 70–74.

29 X. Tang, R. K. Allemann and T. Wirth, *Eur. J. Org. Chem.*, 2017, **2017**, 414–418.

30 C. Görner, P. Schrepfer, V. Redai, F. Wallrapp, B. Loll, W. Eisenreich, M. Haslbeck and T. Brück, *Microb. Cell Fact.*, 2016, **15**, 86.

31 C. Görner, P. Schrepfer, V. Redai, F. Wallrapp, B. Loll, W. Eisenreich, M. Haslbeck and T. Brück, *Microb. Cell Fact.*, 2016, **15**, 86.

32 S. Mafu, M. Jia, J. Zi, D. Morrone, Y. Wu, M. Xu, M. L. Hillwig and R. J. Peters, *Proc. Natl. Acad. Sci. U. S. A.*, 2016, **113**, 2526–2531.

33 M. Xu, M. L. Hillwig, M. S. Tiernan and R. J. Peters, *J. Nat. Prod.*, 2017, **80**, 328–333.

34 S. Edgar, K. Zhou, K. Qiao, J. R. King, J. H. Simpson and G. Stephanopoulos, *ACS Chem. Biol.*, 2016, **11**, 460–469.

35 C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Delivery Rev.*, 1997, **23**, 3–25.

36 J. R. King, S. Edgar, K. Qiao and G. Stephanopoulos, *F1000Research*, 2016, **5**, 397.

37 I. Pateraki, A. M. Heskes and B. Hamberger, in *Biotechnology of Isoprenoids*, ed. J. Schrader and J. Bohlmann, Springer International Publishing, Cham, 2015, pp. 107–139, DOI: 10.1007/10_2014_301.

38 H. van Hattum and H. Waldmann, *J. Am. Chem. Soc.*, 2014, **136**, 11853–11859.

39 S. Wetzel, R. S. Bon, K. Kumar and H. Waldmann, *Angew. Chem., Int. Ed.*, 2011, **50**, 10800–10826.

40 M. E. Maier, *Org. Biomol. Chem.*, 2015, **13**, 5302–5343.

41 S. Jennewein and R. Croteau, *Appl. Microbiol. Biotechnol.*, 2001, **57**, 13–19.

42 Z. Wang, in *Comprehensive Organic Name Reactions and Reagents*, John Wiley & Sons, Inc., 2010, DOI: 10.1002/9780470638859.conrr513.

43 K. Brückner and A. Tissier, *Plant Methods*, 2013, **9**, 46.

44 C. E. Vickers, T. C. Williams, B. Peng and J. Cherry, *Curr. Opin. Chem. Biol.*, 2017, **40**, 47–56.

45 J. Wong, L. Rios-Solis and J. D. Keasling, in *Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids: Production of Fuels and Chemicals*, ed. S. Y. Lee, Springer International Publishing, Cham, 2017, pp. 1–24, DOI: 10.1007/978-3-319-31421-1_219-1.

46 L. Z. Yuan, P. E. Rouvière, R. A. LaRossa and W. Suh, *Metab. Eng.*, 2006, **8**, 79–90.

47 H. Alper, K. Miyaoku and G. Stephanopoulos, *Appl. Microbiol. Biotechnol.*, 2006, **72**, 968–974.

48 S. Y. Lee and H. N. Chang, *J. Environ. Polym. Degrad.*, 1994, **2**, 169–176.

49 G. Aidelberg, B. D. Towbin, D. Rothschild, E. Dekel, A. Bren and U. Alon, *BMC Syst. Biol.*, 2014, **8**, 133.

50 S. Khanna, A. Goyal and V. S. Moholkar, *Crit. Rev. Biotechnol.*, 2012, **32**, 235–262.

51 A. A. Koutinas, A. Vlysidis, D. Pleissner, N. Kopsahelis, I. Lopez Garcia, I. K. Kookos, S. Papanikolaou, T. H. Kwan and C. S. K. Lin, *Chem. Soc. Rev.*, 2014, **43**, 2587–2627.

52 E. Bingham, B. Cohrssen and F. A. Patty, *Patty's Toxicology, 6 Volume Set*, Wiley, 2012.

53 R. Aono, H. Kobayashi, K. N. Joblin and K. Horikoshi, *Biosci., Biotechnol., Biochem.*, 1994, **58**, 2009–2014.

54 K. Skalicka-Woźniak and I. Garrard, *Phytochem. Rev.*, 2014, **13**, 547–572.

55 J. Brent Friesen and G. F. Pauli, *J. Liq. Chromatogr. Relat. Technol.*, 2005, **28**, 2777–2806.

56 E. Hopmann, W. Arlt and M. Minceva, *J. Chromatogr., A*, 2011, **1218**, 242–250.

57 E. Hopmann, A. Frey and M. Minceva, *J. Chromatogr., A*, 2012, **1238**, 68–76.

58 X. Wei, A. D. Rodríguez, P. Baran and R. G. Raptis, *J. Nat. Prod.*, 2010, **73**, 925–934.

59 A. D. Patil, D. Berry, D. P. Brooks, M. E. Hemling, N. V. Kumar, M. P. Mitchell, E. H. Ohlstein and J. W. Westley, *Phytochemistry*, 1993, **33**, 1061–1064.

60 X. Wei, K. Nieves and D. Rodríguez Abimael, *Journal*, 2012, **84**, 1847.

61 F. A. Saddique, A. F. Zahoor, S. Faiz, S. A. R. Naqvi, M. Usman and M. Ahmad, *Synth. Commun.*, 2016, **46**, 831–868.

62 O. Hauenstein, M. Reiter, S. Agarwal, B. Rieger and A. Greiner, *Green Chem.*, 2016, **18**, 760–770.

63 C. Wang, L. Luo and H. Yamamoto, *Acc. Chem. Res.*, 2016, **49**, 193–204.

64 Z. Wang, Y.-T. Cui, Z.-B. Xu and J. Qu, *J. Org. Chem.*, 2008, **73**, 2270–2274.

65 J. M. R. da Silva and M. d. G. Nascimento, *Process Biochem.*, 2012, **47**, 517–522.

66 M. A. Moreira and M. G. Nascimento, *Catal. Commun.*, 2007, **8**, 2043–2047.

67 V. Skouridou, H. Stamatis and F. N. Kolisis, *J. Mol. Catal. B: Enzym.*, 2003, **21**, 67–69.

68 E. G. Ankudey, H. F. Olivo and T. L. Peeples, *Green Chem.*, 2006, **8**, 923–926.

69 S. Ranganathan, S. Zeitlhofer and V. Sieber, *Green Chem.*, 2017, **19**, 2576–2586.

70 C. Görner, M. Hirte, S. Huber, P. Schrepfer and T. Brück, *Front. Microbiol.*, 2015, **6**, 1115.

71 N. A. Barton, B. J. Marsh, W. Lewis, N. Narraidoo, G. B. Seymour, R. Fray and C. J. Hayes, *Chem. Sci.*, 2016, **7**, 3102–3107.

72 V. G. Yadav, *J. Mol. Catal. B: Enzym.*, 2014, **110**, 154–164.

73 S. Y. Lee and H. N. Chang, *Biotechnol. Lett.*, 1993, **15**, 971–974.

74 F. Matthäus, M. Ketelhot, M. Gatter and G. Barth, *Appl. Environ. Microbiol.*, 2014, **80**, 1660–1669.

75 A. Che, in Tech. rep., MIT Synthetic Biology Working Group Technical Reports, 2004.

76 R. P. Shetty, D. Endy and T. F. Knight, *J. Biol. Eng.*, 2008, **2**, 5.

77 J. Müller, S. Oehler and B. Müller-Hill, *J. Mol. Biol.*, 1996, **257**, 21–29.

4

Identification of Sesquiterpene Synthases from the Basidiomycota Coniophora Puteana for the Efficient and Highly Selective β-Copaene and Cubebol Production in *E. coli*

Synopsis

To date, Basidiomycota have hardly been investigated for the secondary metabolome, although they are widely applied in traditional medicine approaches. Interestingly, genomic data reveals that plenty of putative terpene synthases are in the Basidiomycota's genetic code. The brown rot fungus *Coniphora puteana* for instance, has more than 20 annotated terpene synthases and none of them has been investigated in experimental settings to date. We selected three putative terpene synthase sequences (*copu 1-3*) from *C. puteana* and subsequently investigated the function in a heterologous expression system. The corresponding genetic information of *copu 1-3* were cloned into a constitutive expressing system encompassing *dxs, idi* and *crtE*. This combination allowed for the screening of mono, sesqui- and diterpene synthase activity of Copu 1-3. Terpene production was evaluated by thin layer chromatography and GC-MS. Extracts of Copu 2 and Copu 3 expressing cultures, both, showed a unique GC-spectra. MS analyses of the GC peaks revealed the formation of sesquiterpenoids. Interestingly, both terpene synthases mainly generated a single product. The products were characterized by MS and NMR analyses. Copu 3 specifically cyclizes FDP to the tricyclic cubebol. This molecule has been isolated prior from pepper and grapefruits. It is of high commercial interest, because it has a strong refreshing effect without taste and it is approved as a safe food supplement by the FDA. Copu 2 specifically produces the tricyclic β-copaene from FDP. To date, this molecule has not been investigated for their biological function in detail due to limited availability. In this study we additionally showed that β-copaene can be completely transformed to α-copaene if β-copaene is stored in chloroform. The alpha isomer of copaene shows significant activity as an insect attractant.

Moreover, all published and characterized sesquiterpene synthase sequences from Basidiomycota were collected and a phylogenetic tree generated. This investigation revealed a correlation between sesquiterpene product and synthase sequence similarity.

A fermentation process was developed utilizing collagen hydrolysate and glycerol, which are both low value and renewable resources. The copu2 expressing culture reached stationary phase after 46 hours (OD_{600} of 130) at 30°C providing 215 mg/L β-copaene. Fermentation expressing copu3 reached an OD_{600} of 182 that in parallel provided for 497 mg/L cubebol. For both molecules these are the highest titers that have been reached in biotechnological attempts. Compared to natural producers the complexity, in terms of sesquiterpenoid by-products, of the obtained extracts is much lower and facilitates further purification steps.

Authors contribution

W. Mischko and me initiated this project. Based on conserved terpene synthase motifs and the proteins' C-terminal domain, we selected promising terpene synthases. In this context, I contributed

to this work performing the bioinformatic tasks, supporting the fermentation process development and supporting in characterization and elucidation of unknown sesquiterpenoids. In parallel, I provided the initial expression system which was subsequently modified by W. Mischko for the screening and production of the terpene synthase activity.

RESEARCH

Identification of sesquiterpene synthases from the Basidiomycota *Coniophora puteana* for the efficient and highly selective β-copaene and cubebol production in *E. coli*

Wolfgang Mischko, Max Hirte, Monika Fuchs, Norbert Mehlmer and Thomas B. Brück*

Abstract

Background: Terpenes are an important and extremely versatile class of secondary metabolites that are commercially used in the pharmaceutical, food and cosmetics sectors. Genome mining of different fungal collections has revealed the genetic basis for a steadily increasing number of putative terpene synthases without any detailed knowledge about their biochemical properties. The analysis and research of this rich genetic source provides a precious basis for the advancing biotechnological production of an almost endless number of valuable natural metabolites.

Results: Three annotated terpene synthases from the little investigated Basidiomycota *Coniophora puteana* were studied in this work. For biochemical characterization, the heterologous expression in *E. coli* was conducted leading to the identification of two sesquiterpene synthases capable of the highly selective generation of β-copaene and cubebol. These compounds are commercially used as food and flavor additives. The new enzymes show the highest reported product selectivity for their main compounds and therefore represent the first exclusive synthases for β-copaene (62% product selectivity) and cubebol (75% product selectivity) generation. In combination with an optimized heterologous microbial production system, we obtained product titers of 215 mg/L β-copaene and 497 mg/L cubebol.

Conclusion: The reported product selectivity and our generated terpene titers exceed all published biotechnological data regarding the production of β-copaene and cubebol. This represents a promising and economic alternative to extraction from natural plant sources and the associated complex product purification.

Keywords: Copaene, Cubebol, Sesquiterpene, Basidiomycota, *Coniophora puteana*, Heterologous expression, Fermentation, Phylogenetic analysis

Background

Filamentous fungi are experts at producing highly complex natural compounds of commercial interest [1]. Fungal-derived polyketides have been the main focus of recent research activities, whereas the identification of terpenoids and their biosynthesis in fungi have received little attention although these compounds represent the most structurally diverse group of natural products [2]. All terpenoids are based on the same basic C_5 isoprene building blocks, dimethylallyl pyrophosphate (DMAPP) and isopentenyl pyrophosphate (IPP), which are consecutively fused by head to tail condensation. Depending on their carbon chain length, these linear phosphorylated alkenes are universal precursors of mono(C_{10})-, sesqui(C_{15})-, di(C_{20})-, sester(C_{25})- or tri(C_{30})-terpenes [3]. The structural diversity within the class of terpenoids results from the complex cyclization of the linear precursors into chemically complex molecules, a reaction catalyzed by the family of terpene synthase (TPS) enzymes. More specifically, sesquiterpene

*Correspondence: brueck@tum.de
Werner Siemens-Chair of Synthetic Biotechnology, Department of Chemistry, Technical University of Munich, 85748 Garching, Germany

Mischko *et al. Microb Cell Fact*   (2018) 17:164

Page 2 of 13

synthases (STPSs) transform the linear C_{15} precursor farnesyl pyrophosphate (FPP) into a variety of different scaffolds, which form the structural core of functionalized, bioactive sesquiterpenoids (STPs) [2]. Many STPs are lead structures in pharmaceutical applications, encompassing anti-cancer [4, 5], anti-inflammatory [6, 7] and antibiotic [8] therapies. STPs also have existing applications in the food and cosmetics industries, where they are used as flavor and fragrance ingredients [9, 10].

As the extraction of these latter compounds from natural sources is often cost-intensive and not suitable to meet market demands [11], effective biotechnological production routes are the focus of development efforts [12]. Due to the rapid progress of modern sequencing techniques, in silico genome mining based on conserved amino acid motifs can be applied to identify putative TPSs [13]. Whole genome projects of different mushroom-forming fungi (Basidiomycota) species represent a largely unexplored source of investigation and extraction of rarely characterized STPSs [14, 15]. In this context, numerous putative TPSs have already been annotated but their catalytic capacities remain to be established [14]. To exploit the biotechnological potential of fungal biosynthetic pathways, subsequent functional expression and characterization of these enzymes is required. This represents the first step in providing a sustainable supply of high-value natural products using microorganisms as cell factories [16].

Based on the available genome data, we were able to select potential TPSs from the Basidiomycota *Coniophora puteana*, which is classified as a common wood rotting fungus [17]. At present, only the enzyme systems involved in *C. puteana*-dependent wood depolymerization have been characterized in detail [18, 19]. However, there are no reports of any other enzymes involved in secondary metabolite production. Therefore, this study focuses on the identification and characterization of three putative *C. puteana*-derived TPSs. We present the functional reconstitution of these enzymes in an *E. coli* whole-cell production system. With respect to designing an effective STP production system, the supply of the FPP precursor needed to be ensured by an adapted co-expression of bottleneck enzymes (DXS, Idi) from the native non-mevalonate pathway (MEP) (Fig. 2a). This optimization measure ensured a directed carbon flux towards STP production. Two of the three identified TPSs from *C. puteana* (Copu1-3) could be expressed functionally in *E. coli*, resulting in a range of sesquiterpene products. The main product of the Copu2 fermentations was the tricyclic β-copaene. By contrast, Copu3 fermentations provided cubebol as the main product, which is approved as a dietary supplement and flavoring agent [9, 20] due to its pronounced cooling effect [21]. Utilizing the new

terpene synthases from *C. puteana* within an optimized production construct provided 215 mg/L β-copaene and 497 mg/L cubebol. The cubebol production titers reported in this study exceeded all other described production systems by a factor of 50.

Results

Identification and characterization of putative terpene synthase genes in C. *puteana*

A Basic Local Alignment Search Tool (BLAST) analysis of fungal genomes with conserved terpene synthases sequences resulted in the identification of a large number of putative terpene synthases (TPSs). However, for the majority of TPS candidates, a biochemical and functional characterization remains to be established. In order to gain insight into their catalytic function, three putative TPSs (Copu1: XP_007772164.1; Copu2: XP_007771895.1; Copu3: XP_007765978.1) from *C. puteana* were selected for cloning and functional characterization. The specific selection was made on the basis of characteristic conserved sequence motifs. Moreover, Copu1-3 showed closely related amino acid (AA) sequences (55–62% similarity). A comparison of the AA sequence of Copu1 and Copu2 with the public database showed <50% similarity to other listed enzymes, covering all biological realms. By contrast, Copu3 showed 65% similarity to putative TPS sequences, which were not functionally characterized. The AA sequences of all three enzymes contained typical sequence motifs common to the TPS family, such as the highly conserved (N/D)D(L/I/V)x(S/T)xxxE (NSE) triad and the aspartate-rich D(D/E)xxD motif, coordinating a trinuclear Mg^{2+} cluster, which is catalytically essential for the initial hydrolysis of the FPP-derived pyrophosphate group [22] (see Additional file 1). A highly conserved arginine residue, indicated as the pyrophosphate sensor, is located 46 positions upstream of the NSE triad. Additionally, the catalytically important RY-dimer, which is involved in the formation of hydrogen bonds to the substrate-derived pyrophosphate, is found 80 AA downstream of the NSE triad and close to the C-terminus [23–25].

Heterologous expression of *C. puteana* TPS genes resulted in the generation of diverse sesquiterpenes in *E. coli*

To study the product profile of Copu1-3, their predicted coding sequences were codon-optimized and synthesized for transfer into *E. coli* expression vectors. For the heterologous expression, an adapted production system based on a single operon with a constitutive promoter was constructed. Reported bottleneck enzymes from *E. coli*'s native non-mevalonate pathway (MEP) were selected for co-expression (DXS; WP_099145004.1 and idi; AAC32208.1) to increase

Mischko *et al. Microb Cell Fact*   (2018) 17:164

Page 3 of 13

the precursor supply and enhance the general isoprenoid production (Fig. 1). The resulting plasmids were transformed into *E. coli* HMS174 (DE3) for recombinant gene expression and subsequent analysis of new, potential terpene compounds. In the first series of experiments, a geranylgeranyl diphosphate synthase (*Pantoea ananatis*; crtE; ADD79325.1) was additionally integrated into the operon to study a potential function as diterpene synthases. *E. coli* cultures co-expressing Copu2 or Copu3 with the MEP bottleneck enzymes for 48 h produced a mixture of exclusively 12 and 18 terpene products, respectively. A cultivation temperature of 30 °C and reduced shaking (90 rpm) ensured an adequate bacterial growth rate and the requirements for the production of potentially volatile compounds. Gas chromatography–mass spectrometry (GC–MS) analysis of the liquid extract revealed new product peaks in both cultures with typical mass fragmentation patterns at 161, 207, 222 m/z and 105, 161, 204 m/z (Fig. 2), illustrating the mass patterns for cyclic C_{15} hydrocarbons with and without a single hydroxyl group. Copu3 appeared to be quite selective for the generation of one major STP alcohol (RT: 15.4 min; parent ion at 222 m/z). In contrast, Copu2 appeared to convert FPP into a smaller and slightly different set of cyclization products. The major product of Copu2 fermentations (RT: 14.3 min) was identified as an unhydroxylated STP compound (parent ion at 204 m/z). By contrast, no terpenoid products were detected in *E. coli* cultures expressing Copu1. Neither the co-expression with a GGPP synthase to enable a possible formation of diterpenes nor the evaluation of different fermentation temperatures to avoid eventual evaporation of volatile compounds (e.g., monoterpenes) showed any product formation. Therefore, Copu1 was classified as a nonfunctional TPS sequence.

Terpene purification and structure elucidation

A putative identification of the newly generated STPs was performed by comparing their detailed mass spectra to the National Institute of Standards and Technology (NIST) Database. The comparison of the GC–MS metabolite profiles of *E. coli* cultures expressing Copu2 revealed the putative tricyclic STP β-copaene (RT: 14.3 min; parent ion at 204 m/z; major daughter ions at 105, 161 m/z) as the major product. β-Copaene accounted for about 62% of the total STPs detected by GC-FID. *E. coli* strains expressing Copu3 accumulated the putative tricyclic STP alcohol cubebol (RT: 15.4 min; parent ion at 222 m/z; major daughter ions at 105, 161, 207 m/z) as the major product, representing 75% of the total terpene fraction (Fig. 2). Other minor compounds from both cultures encompassed δ-cadinene, β-elemene, germacrene D and germacrene D-4-ol, whose GC–MS spectra were consistent with NIST database references. Notably, several minor STPs did not match any NIST reference spectra and could therefore not be assigned. In addition to GC–MS analyses, the main products of Copu2 and Copu3 were analyzed by NMR spectroscopy. Organic extraction allowed HPLC-based β-copaene and cubebol purification from 1-L fermentations. The HPLC separation of the Copu2 fermentation products allowed for the extraction of the main product but could not resolve several other STPs, which most likely represent a mixture of different product isoforms. However, another minor product was isolated from the Copu3 fermentation broth.
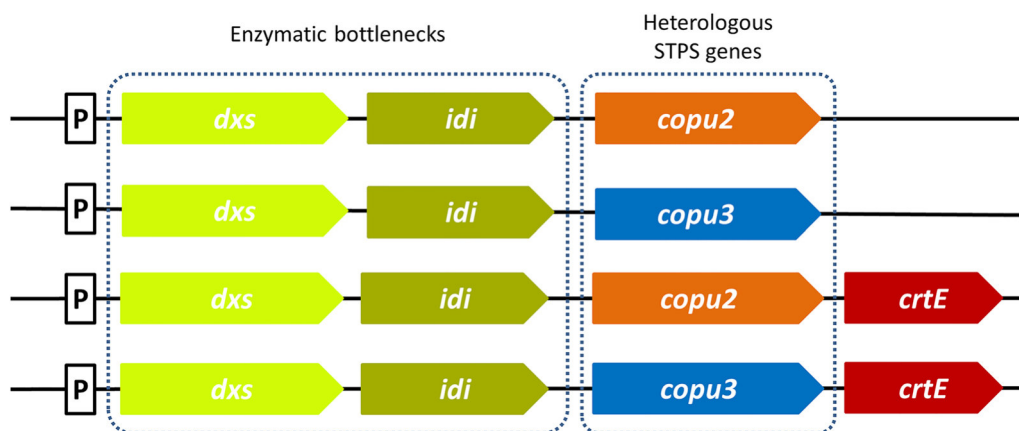
Fig. 1 Design of the synthetic operons for an increased precursor supply and efficient STP production. Enzymatic bottlenecks within the upstream non-mevalonate (MEP) pathway (dxs and idi) as well as the heterologous STPS genes (Copu2 and Copu3) are combined in a single synthetic operon. The integration of an additional geranylgeranyl diphosphate synthase (crtE) allowed to study a potential function as diterpene synthases. A constitutive promoter (P) regulates the pathways

Mischko *et al. Microb Cell Fact* (2018) 17:164

Page 4 of 13
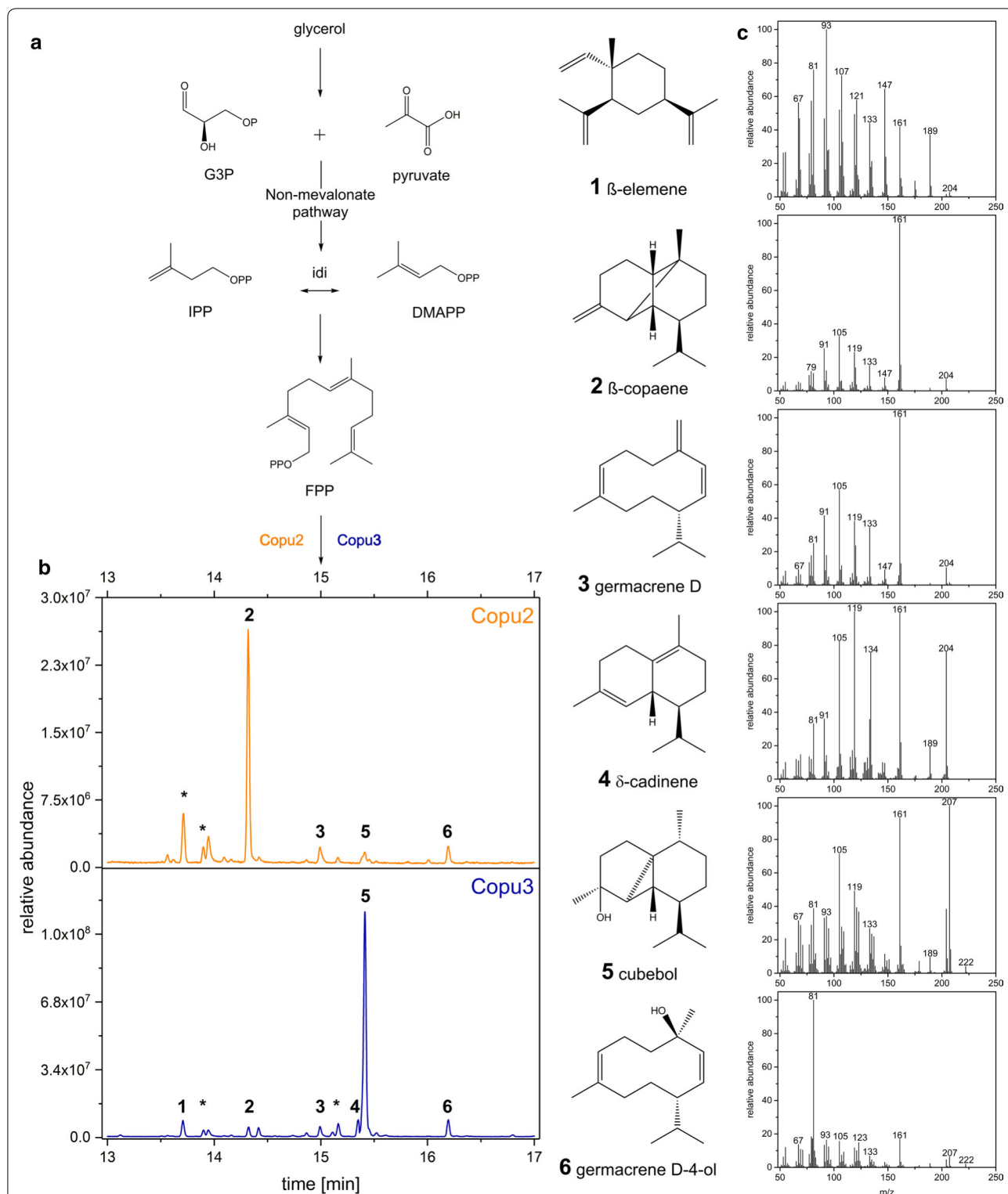
Fig. 2 The biosynthetic generation of β-copaene and cubebol. STP production relies on the precursor supply of the non-mevalonate (MEP) pathway (**a**). The respective GC–MS chromatograms of *E. coli* cultures co-expressing either Copu2 or Copu3 and the corresponding MEP bottleneck enzymes reveal new product peaks and distinct product profiles due to the specific cyclization reaction (**b**). The respective detailed MS spectra allow for a putative compound assignment (**c**). The asterisks refers to terpenes without any MS spectra match in the NIST database

Mischko *et al. Microb Cell Fact* (2018) 17:164

Page 5 of 13

A comparison of NMR spectra to reported references confirmed the presence of β-copaene [26, 27], cubebol [28] and δ-cadinene [29] as products of Copu2 and Copu3, respectively. Based on the NMR data, we designated Copu2 as a new β-copaene synthase. Conversely, Copu3 was designated as a new, highly selective cubebol synthase.

Interestingly, a rearrangement of β-copaene to the better described α-isomer could be observed in this context. The complete conversion took place by storing the pure compound in chloroform for < 12 h (Fig. 3). α-Copaene has been shown to possess important biological properties, including anticarcinogenic as well as antioxidant activity in the field of neurodegenerative diseases [30, 31], or serves as an insect attractant [32].

Phylogenetic synthase evolution

At present, the identification and characterization of terpene cyclases from the group of Basidiomycota is limited. Protein sequence-based phylogenetic analysis of the 29 genetically and biochemically characterized STPSs derived from Basidiomycota revealed four distinct clades (clade I–IV) (Fig. 4). The clustering by sequence conservation suggests that STPSs within one specific clade may catalyze the same or a related cyclization reaction. It also revealed that Copu2 and Copu3 clustered in clade I together with all other Basidiomycota-derived STPSs, generating either β-copaene or cubebol (ACTPS9, Cop4 and Stehi_128017). Additionally, most enzymes that generated cadinene isoforms clustered in clade I. For several candidate enzymes constituting this clade (Cop4, Omp4, Omp5a and 5b), a substrate cyclization mechanism has already been postulated, involving a 1,10-cyclization of (3R) nerolidyl diphosphate (NPP). The conversion of initial FPP involves the formation of a cis-germacradienyl cation, followed by a subsequent 1,6-cyclization. The final result is various STPs derived from a cadinyl cation (Fig. 5) [14, 33, 34]. By contrast, the clade II STPS mechanism involves a 1,10-cyclization of (2E,6E)-FPP to an E,E-germacradienyl cation (Omp1−3, Cop 1−3) [35], generating predominantly α-muurolene and

germacrene A as well as different types of cadinol. Clade III STPSs share a 1,6-cyclization mechanism of (3R)-NPP or (3S)-NPP, leading to a bisabolyl cation [33, 36], forming mainly barbatene or α-cuprenene. Finally, clade IV STPSs share a common 1,11-cyclization mechanism of (2E,6E)-FPP [37]. Except for GME9210, all enzymes in this clade exclusively represent functionally characterized Δ6-protoilludene synthases.
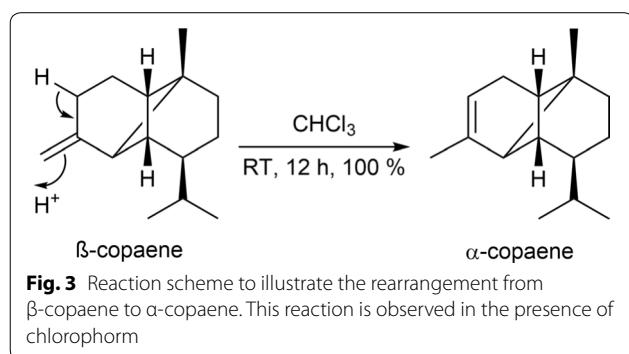
In silico homology modeling [38] of Copu1-3 as well as Cop4 and structural alignments with the selinadiene cyclase (4OKM), a monomeric STPS from *Streptomyces pristinaespiralis*, as a nearest neighbor reference structure were carried out (see Additional file 1). Interestingly, the sequence identity analyses strictly differentiates sequences from distinct organisms while the structural analysis (RMSD calculation) reveals a close structural relation between Copu3 and Cop4. Both enzymes produce δ-cadinene, β-copaene and cubebol.

Technical scale, fed-batch production of β-copaene and cubebol

To investigate the production performance of Copu2 and Copu3 in an optimized microbial system, fed-batch fermentation experiments were carried out in 1.3-L fermenters. This production scale provided for technically relevant amounts of β-copaene and cubebol. *E. coli* HMS174 (DE3) strains co-expressing either Copu2 or Copu3 and the corresponding MEP bottleneck enzymes were cultured at 30 °C under controlled conditions (Fig. 6). The Copu2-expressing culture reached its stationary phase after 46 h with a final calculated OD_{600} of 130, providing a final β-copaene titer of 215 mg/L. Based on these data, the specific β-copaene production and productivity were 4.4 mg/g dry cell weight (DCW) and 4.7 mg/L/h, respectively. To the authors' knowledge, this is the first report of any quantitative biotechnological production of β-copaene. In comparison, the Copu3-based fermentation generated a cubebol titer of 497 mg/L (calculated OD_{600} of 182), and the specific cubebol production and productivity even reached 7.2 mg/g DCW and 11.2 mg/L/h, respectively. The reported titers in this study exceeded concentrations of alternative approaches obtained by equivalent fermentations based on a plant-derived cubebol synthase (titers of 10 mg/L cubebol) [48] by 50-fold.

Discussion

In this study, three new terpene synthases (Copu1-3) were identified from genome data of the Basidiomycota *Coniophora puteana* and functionally characterized for their product profiles. At present, this filamentous fungus has only been described in the context of cellulytic activity. Therefore, this is the first report of any enzyme

Fig. 3 Reaction scheme to illustrate the rearrangement from β-copaene to α-copaene. This reaction is observed in the presence of chloroform

Mischko *et al. Microb Cell Fact*   (2018) 17:164

Page 6 of 13

Fig. 4 Unrooted neighbor-joining phylogram of all 29 reported and experimentally characterized STPSs from Basidiomycota. The protein sequence based analysis includes STPSs from *Antrodia cinnamomea* [39], *Armillaria gallica* [40], *Boreostereum vibrans* [41], *Coprinopsis cinereus* [35], *Fomitopsis pinicola* [42], *Lignosus rhinocerus* [43] *Stereum hirsutum* [37], *Omphalotus olearius* [42] as well as *Coniophora puteana*. The phylogenetic analysis revealed four distinct clades, divided into different cyclization mechanisms and product profiles. Copu1-3 belong to clade I including all other known β-copaene or cubebol synthases from Basidiomycota. For detailed information about the used sequences and accession numbers, see Additional file 1

system involved in secondary metabolite biosynthesis from this organism. Only a few reports describe the successful functional characterization of TPSs from the phylum of Basidiomycota [35, 37, 39, 42]. In this study, the heterologous expression of Copu2 and Copu3 functionally confirmed their identities as sesquiterpene synthases (STPSs). By contrast, Copu1 expression in either an engineered sesqui- or diterpene *E. coli* production system did not yield any products. The Copu1 sequence may represent a non-functional TPS variant. This could be attributed to the fact that Basidiomycota tends to have very intron-rich genomes, and various functional transcripts

may be generated by alternative splicing events. The in silico identification of the annotated Copu1 coding sequence may only encompass splicing events that result in a catalytically inactive enzyme [14, 49].

The selectivity of terpene synthases varies significantly in this enzyme class. In that respect, highly selective TPSs only provide a single product (e.g., the (+)-d-cadinene synthase from *Gossypium arboreum* [46]), while more promiscuous enzymes can generate in excess of 50 different compounds (e.g., the γ-humulene synthase (*Abies grandis*) [45]). Product promiscuity is prevalent, especially within the group of STPSs [50]. However,
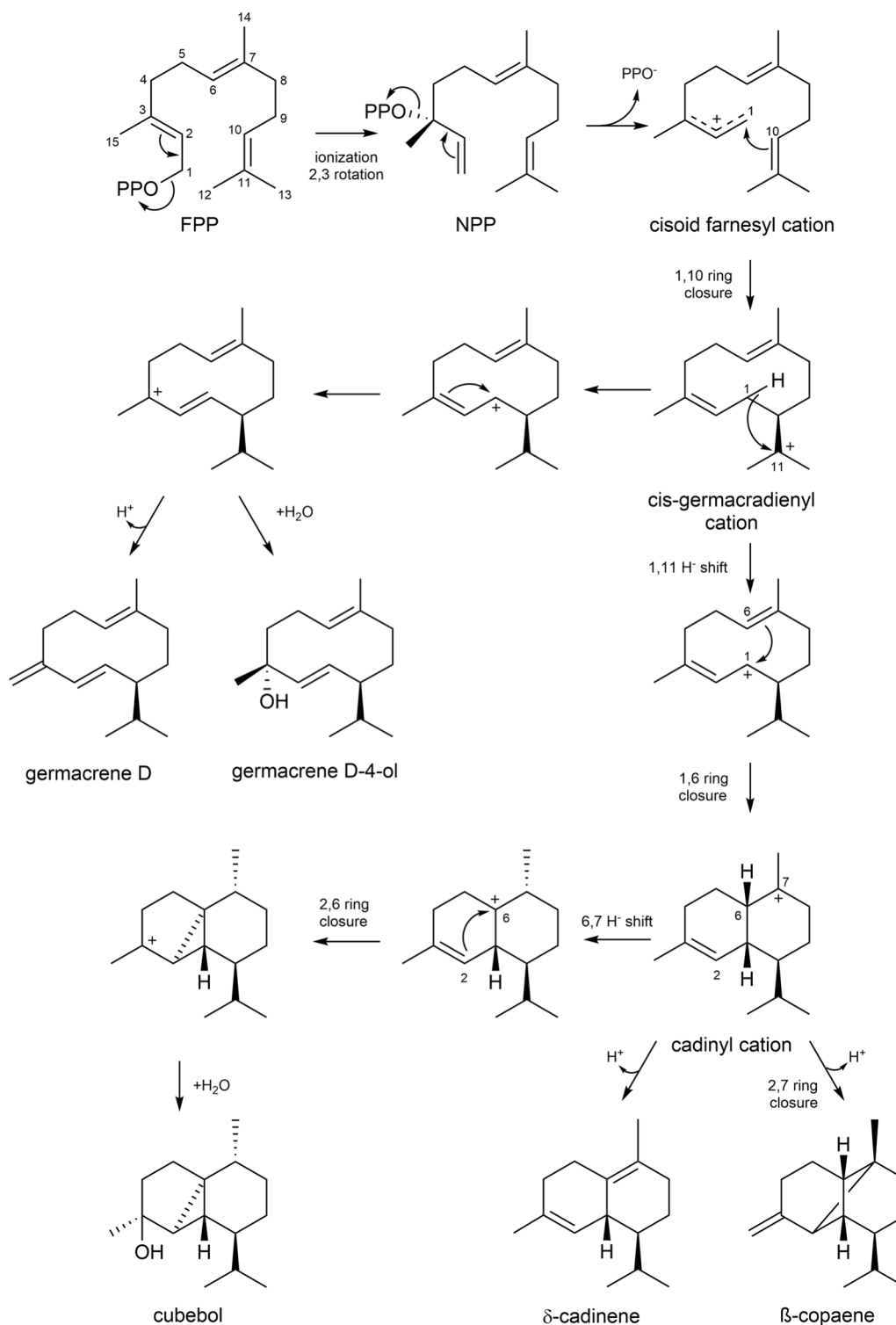
Fig. 5 Proposed cyclization mechanisms for the generation of the main STPs from Copu2 and Copu3. The reaction starts by ionization and isomerization of farnesyl diphosphate (FPP) creating nerolidyl diphosphate (NPP). The first cyclization is achieved by the formation of the C1–C10 bond. Subsequent C6–C1 bond formation yields a cadinyl cation, which represents a general precursor for various cadalane and cubebane-type STPs [14, 44–47]

Mischko *et al. Microb Cell Fact*   (2018) 17:164

Page 8 of 13

Fig. 6 *E. coli* growth curves and time-dependent concentration of β-copaene and cubebol in controlled fed-batch fermentations. Expression of Copu2 (**a**) and Copu3 (**b**) in *E. coli* resulted in a constant concentration increase, which was finally limited by the stationary phase. The error bars represent the mean values ± standard deviation of technical triplicates

heterologously expressed Copu2 and Copu3 display a fairly restricted product profile. As β-copaene accounted for 62% of all terpenoid products in Copu2 fermentations, this enzyme was identified as a β-copaene synthase. By contrast, cubebol was the main product (75%) of Copu3 fermentations, which designated this enzyme as a cubebol synthase. To the best of our knowledge, this is the first report of a selective β-copaene synthase and cubebol synthase. Both enzymes combine high product titers with significant product selectivity. Although we cannot completely rule out an *E. coli*-specific effect on the generated products, we believe that the enzymatic reaction corresponds to the natural conditions in *C. puteana*.

To provide insights into the structural and functional diversity of STPSs from Basidiomycota, we carried out a comprehensive phylogenetic analysis. This analysis encompassed the STPS sequences of *Antrodia cinnamomea* [39], *Armillaria gallica* [40], *Boreostereum vibrans* [41], *Coprinopsis cinereus* [35], *Fomitopsis pinicola* [42], *Lignosus rhinocerus* [43] *Stereum hirsutum* [37] and *Omphalotus olearius* [42], which are all functionally characterized. The phylogenetic analysis suggests that Copu1-3 belong to the same STPS clade, sharing a 1,10-cyclization mechanism with other related enzymes

in this group (Fig. 4). Clade I encompasses all previously reported Basidiomycota STPSs that are capable of generating β-copaene, cubebol or cadinenes, which are all derived from a cadinyl cation intermediate (Fig. 5). However, Copu2 and Copu3 are the only STPSs realizing the generation of significant amounts of β-copaene and cubebol, respectively. Both enzymes form germacrene D and germacrene D-4-ol as minor products, which are derived from an earlier cis-germacradienyl cation intermediate. This carbocation intermediate is present in the biosynthesis of various cadinene type compounds [51]. The phylogenetic analysis based on amino acid sequences clearly illustrates the potential for possibly predicting enzymatic reaction mechanisms and an eventual product profile. Phylogenetic tree generation based on structural information might further improve this functional in silico prediction of an unknown terpene synthase. However, for a more precise prediction of enzymatic functions, more data have to be generated for Basidiomycota since this research area is still at a very early stage and comprises only a few candidate enzymes.

Cubebol is of interest to the cosmetic and flavor industry as it is already a registered product with a pronounced cooling effect [21] and is therefore formulated into dietary supplements and flavoring agents [9, 20]. Also the performed conversion of β-copaene to the more extensively investigated α-isoform opens up access to a potential pharmaceutical market, based on the proposed anticarcinogenic as well as antioxidant activity in the field of neurodegenerative diseases [30, 31]. At present, both compounds are commercially extracted from plant material. Specifically, β-copaene is a component of the essential oil (15%) from *Piper nemorense* leaves [52]. Similarly, cubebol is a constituent in the berry oil (5–10%) of *Piper cubeba* [53, 54] or the volatile leaf oil (3.3%) extracted from *Juniperus convallium* [55]. The quality and quantity of the target compounds from plant extracts vary seasonally and with global production location [53]. Consequently, purification from these natural extracts is technically challenging and cost-intensive [11, 12] because of various isoforms and structurally related impurities in complex mixtures. In contrast to the plant extracts, the biotechnological production of both β-copaene and cubebol would provide an efficient and sustainable alternative with simplified purification procedures. This could ensure a constant product quality, which is of high interest to the flavor industry [12].

The biotechnological formation of cubebol or β-copaene could only be provided by very few characterized STPSs. The δ-cadinene synthase VvPNCuCad (HM807407.1) from grapevine (*Vitis vinifera*) encoding a multi-product STPS shows 20.5% cubebol selectivity when expressed in *E. coli* [56]. The fungal δ-cadinene

Mischko *et al. Microb Cell Fact*   (2018) 17:164

Page 9 of 13

synthase Cop4 (A8NU13.1) from *C. cinereus* is also designated as a cubebol and β-copaene synthase. The product profile of recombinant Cop4 is reported to generate 30% β-copaene and 10% cubebol with respect to the total terpene production titer [35]. Even under optimized in vitro conditions, Cop4 does not generate in excess of 34.2% cubebol [47]. Other fungal STPSs with related enzymatic activity have been cloned and characterized from *A. cinnamomea* (AcTPS9) [39] and *S. hirsutum* (Stehi_128017) [37] but with significantly lower production rates. Even though several other plant enzymes are listed for the more common α-copaene isoform [57, 58], Cop4 represented the only known enzyme with a relevant formation of β-copaene to date. Therefore, Copu2 and 3 isolated from *C. puteana* are the only enzymes capable of the highly selective generation of their main products β-copaene (62% product selectivity) and cubebol (75% product selectivity) and represent an excellent basis for biotechnological production.

The only described biotechnological production approach reported for the quantitative generation of cubebol utilized a patented plant enzyme (CQ813505.1 from grapefruit *Citrus x paradisi*; 28% cubebol selectivity) but only provided titers of 10 mg/L [48], which is well below the titers we report with Copu3 fermentations. In our *E. coli*-based approach, we were able to demonstrate cubebol titers of 497 mg/L even without significant optimization of the fermentation procedures. The production of 215 mg/L β-copaene represents the first biotechnological process with technically relevant titers of this compound. Our data expand the set of functionally characterized STPSs, which can be used for biotechnological processes. With respect to literature data and current state-of-the art technology of essential oil extraction [59, 60], our production system provided, within < 2 days, cubebol concentrations (w/w) that were similar to the natural producer (ripe berries of *piper cubeba*) [53] over a whole season. At the same time, it represents a 50-fold increase compared to the only reported target-oriented biotechnological approach in yeast [48].

Conclusion
Fungi have an enormous capacity for biosynthesis of versatile natural terpenoids and therefore represent an outstanding resource of new metabolic pathways for biotechnological production. Regarding the increasing availability of complete fungal genomes, the quantity of genetic information is expanding continuously. However, the complex biochemical characterization of the in silico-annotated enzyme activities is lagging far behind. While most enzyme-focused studies involve Ascomycota-derived sequences, the biosynthetic diversity of

secondary metabolites, particularly related to the terpenome of Basidiomycota, is largely unexplored.

In this study, we focused on the discovery and investigation of sesquiterpene synthases (STPSs) from the Basidiomycota *Coniophora puteana*. We identified the STPSs Copu1-3, which were subsequently expressed in an engineered *E. coli* host capable of either sesqui- or diterpene production. While the expression of Copu1 did not show any terpene accumulation, the sesquiterpene β-copaene was the main product of Copu2. Hence, Copu2 was designated as the first exclusive β-copaene synthase (62% product selectivity). Similarly, Copu3 was identified as the most efficient cubebol synthase (75% product selectivity) to date. The metabolic optimization of a microbial production host, including the introduction of MEP pathway enzymes and the fungal enzymes Copu2 and Copu3, created microbial cell factories for the de novo production of 215 mg/L β-copaene and 497 mg/L cubebol. Although further work is needed to optimize the product titers, the current whole-cell systems could serve as a promising basis for the development of large-scale biotechnological production of these compounds. Patent filed.

Methods
Gene cloning, plasmid construction and culture condition
Escherichia coli strain DH5α was used for cloning and *E. coli* strain HMS174 (DE3) for terpene production. Genes encoding for the sesquiterpene synthases Copu1 (327 AA; XP_007772164.1); Copu2 (340 AA; XP_007771895.1) and Copu3 (332 AA; XP_007765978.1) from *Coniophora puteana* were ordered codon-optimized (Eurofins Genomics) for improved efficiency in *E. coli* and cloned into a pACYC-based expression vector system. The final production construct contained a single operon with selected bottleneck enzymes of the MEP pathway (DXS; 1-deoxy-D-xylulose-5-phosphate synthase from *E. coli*; WP_099145004.1) (idi; isopentenyl pyrophosphate: dimethylallyl pyrophosphate isomerase from *Haematococcus lacustris*; AAC32208.1) and the corresponding cyclase under the control of a lac-I-derived constitutive promoter [61]. Cultures were grown in modified R-media [62] (13.3 g/L KH_2PO_4, 4.0 g/L $(NH_4)_2HPO_4$, 1.7 g/L citric acid, 5.0 g/L yeast extract, 35 g/L glycerol, 4.9 mL/L 1 M $MgSO_4$, 2.45 mL/L 0.1 M Fe(III) citrate, 10 mL/L $100 \times$ trace elements solution (5 g/L EDTA, 0.83 g/L $FeCl_3$–$6H_2O$, 84 mg/L $ZnCl_2$, 13 mg/L $CuCl_2$–$2H_2O$, 10 mg/L $CoCl_2$–$2H_2O$, 10 mg/L H_3BO_3, 1.6 mg/L $MnCl_2$–$4H_2O$), 1 mg/L Thiamin) supplemented with the appropriate antibiotics, ampicillin (100 μg/mL) or chloramphenicol (34 μg/mL), at 30 °C and 90 rpm shaking.

Mischko *et al. Microb Cell Fact*   (2018) 17:164

Page 10 of 13

Terpene isolation

For analytical terpene isolation within the screening process, 35 mL of the selected cultures was mixed with 15 mL of an extraction solution (ethyl acetate, hexane and ethanol; 1:1:1). The suspension was then strongly shaken for 15 min followed by a 60 s centrifugation step at 8000g for phase separation. A defined volume of the upper organic phase was then sampled and analyzed by GC–MS.

Fermentation and preparative extraction

All fermentations were performed in a DASGIP® 1.3 L parallel reactor system (Eppendorf AG) using modified R-media [62] supplemented with 35 g/L glycerol. An overnight preculture was used for inoculation (OD = 0.1). The cultivation temperature was kept constant at 30 °C. The initial stirrer velocity and airflow were set to 200 rpm and 0.5 volumes of air per volume of medium per minute (vvm), respectively. The dissolved oxygen (DO) was kept at 30% and controlled by stirrer velocity (up to 1000 rpm), oxygen content (up to 100%) and airflow. A pH value of 7.00 was maintained by adding 25% sodium hydroxide solution as needed. A pH-based feed was activated by pH values exceeding 7.05, which triggered a feed shot of 40 mL. The feed solution consisted of 600 g/L glycerol, 20 g/L $MgSO_4 \cdot 7H_2O$, 15 mL/L $100 \times$ trace elements solution (5 g/L EDTA, 0.83 g/L $FeCl_3 – 6H_2O$, 84 mg/L $ZnCl_2$, 13 mg/L $CuCl_2 – 2H_2O$, 10 mg/L $CoCl_2 – 2H_2O$, 10 mg/L H_3BO_3, 1.6 mg/L $MnCl_2 – 4H_2O$), 70 g/L collagen hydrolysate and 7.5 g/L yeast extract (pH = 7.00). Samples were taken regularly to measure the OD_{600} and the respective terpene concentration.

The whole cultivation broth was extracted by adding an equal volume of ethyl acetate and ethanol (1:1) to the cell culture. The suspension was shaken for 12 h at 22 °C followed by a centrifugation step for 15 min at 7000g. After separating the supernatant from the pellet, an additional 1/4 volume of hexane was added. The mixture was shaken for 2 h before the upper phase was isolated using a separation funnel. The organic layer was concentrated using a rotary evaporator. The remaining crude oil containing terpenoids was dissolved in an ACN and H_2O (9:1) solution for further HPLC purification.

Terpene purification, identification and quantification

The purification was performed by preparative HPLC using a NUCLEODUR® C18 HTec 250/10 mm 5 µm column (MACHEREY–NAGEL GmbH & Co. KG) and a diode array UV detector at 2.2 mL/min flow rate. The injection volume was 2 mL at a concentration of 25 mg/mL of crude extract in ACN and H_2O (9:1). The separation was performed applying an ACN gradient starting at 90% and increasing to 100% within 10 min. This was maintained for 30 min. The oven temperature was set to 30 °C. The terpene peaks were detected at 210 nm wavelength. Fractions containing the pure product, determined by GC–MS analysis, were pooled and concentrated using a rotary evaporator.

Escherichia coli whole-cell conversion extracts from *C. puteana* Copu1-3 were analyzed by a Trace GC Ultra with DSQ II (Thermo Scientific). The sample was loaded by TriPlus AS onto an SGE BPX5 column (30 m, I.D. 0.25 mm, film 0.25 µm). The initial oven temperature was set at 50 °C for 2 min, increased to 320 °C at a rate of 10 °C/min, and held for 3 min. MS data were recorded at 70 eV (EI) and m/z (rel. intensity in %) as total ion current (TIC). Data were collected in full scan mode (m/z 50–650). Structural determination of terpenes was conducted by comparison to mass spectra data of the NIST Standard Reference Database. Concentrations were quantified by correlating the FID peak area to a defined α-humulene and cubebol standard of known quantity (see Additional file 1 for calibration curves).

The NMR spectra of the products were recorded in $CDCl_3$ (cubebol and δ-cadinene) or C_6D_6 (β-copaene) with a Bruker Ascend™ 400 MHz NMR spectrometer. All chemical shifts are relative to $CDCl_3$ at δ = 7.26 (1H-NMR) and $CDCl_3$ at δ = 77.16 (13C-NMR) or C_6D_6 at δ = 7.16 (1H-NMR) and C_6D_6 at δ = 128.06.16 (13C-NMR) using the standard δ notation in parts per million.

Protein modeling

The webservice RaptorX was applied for homology modeling studies (http://raptorx.uchicago.edu) [63]. Subsequently, the predicted structures (Cop4, Copu 1-3) were analyzed by Visual Molecular Dynamics (VMD) (http://www.ks.uiuc.edu/Research/vmd/) [64]. Protein structures were aligned and compared using the MultiSeq [65], and scenes rendered by Tachyon [66], both, implemented in the VMD software package. Phylogenetic trees of these five proteins were calculated using sequence identity or the rmsd value as tree generation criteria.

Structure function analyses of Copu3 were performed as previously described [38].

Phylogenetic analysis

A multiple sequence alignment was generated with Clustal Omega (https://www.ebi.ac.uk/Tools), using seeded guide trees and HMM profile techniques. Evolutionary analyses were conducted with iTOL (https://itol.embl.de/). The phylogenetic tree was inferred using the maximum likelihood method.

Mischko *et al. Microb Cell Fact*   (2018) 17:164

Page 11 of 13

Abbreviations

AA: amino acid; DCW: dry cell weight; DMAPP: dimethylallyl pyrophosphate; FPP: farnesyl pyrophosphate; GC–MS: gas chromatography–mass spectrometry; IPP: isopentenyl pyrophosphate; MEP: non-mevalonate pathway; NIST: National Institute of Standards and Technology; NPP: nerolidyl diphosphate; STP: sesquiterpenoid; STPS: sesquiterpene synthase; TPS: terpene synthase.

Authors' contributions

Conceived the project: WM and MH. Designed and performed the experiments: WM. Analyzed the data: WM and MH. Prepared the manuscript: WM and TB. Supervised the whole work: TB, MF, and NM. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its Additional file. Additional data required is available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Zhong J-J, Xiao J-H. Secondary metabolites from higher fungi: discovery, bioactivity, and bioproduction. Adv Biochem Eng Biotechnol. 2009;113:79–150. https://doi.org/10.1007/10_2008_26.
2. Pateraki I, Heskes AM, Hamberger B. Cytochromes P450 for terpene functionalisation and metabolic engineering. Adv Biochem Eng Biotechnol. 2015;148:107–39. https://doi.org/10.1007/10_2014_301.
3. Christianson DW. Structural and chemical biology of terpenoid cyclases. Chem Rev. 2017;117:11570–648. https://doi.org/10.1021/acs.chemrev.7b00287.
4. Kornienko A, Evidente A, Vurro M, Mathieu V, Cimmino A, Evidente M, et al. Toward a cancer drug of fungal origin. Med Res Rev. 2015;35:937–67. https://doi.org/10.1002/med.21348.
5. Hassan SB, Gali-Muhtasib H, Göransson H, Larsson R. Alpha terpineol: a potential anticancer agent which acts through suppressing NF-kappaB signalling. Anticancer Res. 2010;30:1911–9.
6. Fernandes ES, Passos GF, Medeiros R, da Cunha FM, Ferreira J, Campos MM, et al. Anti-inflammatory effects of compounds alpha-humulene and (−)-trans-caryophyllene isolated from the essential oil of *Cordia verbenacea*. Eur J Pharmacol. 2007;569:228–36. https://doi.org/10.1016/j.ejphar.2007.04.059.
7. Rogerio AP, Andrade EL, Leite DFP, Figueiredo CP, Calixto JB. Preventive and therapeutic anti-inflammatory properties of the sesquiterpene alpha-humulene in experimental airways allergic inflammation. Br J Pharmacol. 2009;158:1074–87. https://doi.org/10.1111/j.1476-5381.2009.00177.x.
8. Zhao B, Lin X, Lei L, Lamb DC, Kelly SL, Waterman MR, Cane DE. Biosynthesis of the sesquiterpene antibiotic albaflavenone in *Streptomyces coelicolor* A3(2). J Biol Chem. 2008;283:8183–9. https://doi.org/10.1074/jbc.M710421200.
9. Velazco MI, Wuensche L, Deladoey P. Use of cubebol as a flavoring ingredient 2001. Google Patents.
10. Jones CG, Moniodis J, Zulak KG, Scaffidi A, Plummer JA, Ghisalberti EL, et al. Sandalwood fragrance biosynthesis involves sesquiterpene synthases of both the terpene synthase (TPS)-a and TPS-b subfamilies, including santalene synthases. J Biol Chem. 2011;286:17445–54. https://doi.org/10.1074/jbc.M111.231787.
11. Chemat F, Vian MA, Cravotto G. Green extraction of natural products: concept and principles. Int J Mol Sci. 2012;13:8615–27. https://doi.org/10.3390/ijms13078615.
12. Liu X, Ding W, Jiang H. Engineering microbial cell factories for the production of plant natural products: from design principles to industrial-scale production. Microb Cell Fact. 2017;16:125. https://doi.org/10.1186/s12934-017-0732-7.
13. Bian G, Deng Z, Liu T. Strategies for terpenoid overproduction and new terpenoid discovery. Curr Opin Biotechnol. 2017;48:234–41. https://doi.org/10.1016/j.copbio.2017.07.002.
14. Quin MB, Flynn CM, Schmidt-Dannert C. Traversing the fungal terpenome. Nat Prod Rep. 2014;31:1449–73. https://doi.org/10.1039/c4np00075g.
15. Schmidt-Dannert C. Biosynthesis of terpenoid natural products in fungi. Adv Biochem Eng Biotechnol. 2015;148:19–61. https://doi.org/10.1007/10_2014_283.
16. Wriessnegger T, Augustin P, Engleder M, Leitner E, Müller M, Kaluzna I, et al. Production of the sesquiterpenoid (+)-nootkatone by metabolic engineering of *Pichia pastoris*. Metab Eng. 2014;24:18–29. https://doi.org/10.1016/j.ymben.2014.04.001.
17. Viitanen HA. Modelling the time factor in the development of brown rot decay in pine and spruce sapwood—the effect of critical humidity and temperature conditions. Holzforschung. 1997;51:99–106. https://doi.org/10.1515/hfsg.1997.51.2.99.
18. Schmidhalter DR, Canevascini G. Isolation and characterization of the cellobiose dehydrogenase from the brown-rot fungus *Coniophora puteana* (Schum ex Fr.) Karst. Arch Biochem Biophys. 1993;300:559–63. https://doi.org/10.1006/abbi.1993.1077.
19. Schmidhalter DR, Canevascini G. Purification and characterization of two exo-cellobiohydrolases from the brown-rot fungus *Coniophora puteana* (Schum ex Fr.) Karst. Arch Biochem Biophys. 1993;300:551–8. https://doi.org/10.1006/abbi.1993.1076.
20. Deladoey P, Velazco MI, Wunsche L. Verwendung von Cubebol als Aromastoff 2005. Google Patents.
21. Bharate SS, Bharate SB. Modulation of thermoreceptor TRPM8 by cooling compounds. ACS Chem Neurosci. 2012;3:248–67. https://doi.org/10.1021/cn300006u.
22. Leeper FJ, Vederas JC. Biosynthesis: aromatic polyketides, isoprenoids, alkaloids. Berlin: Springer; 2000.
23. Baer P, Rabe P, Fischer K, Citron CA, Klapschinski TA, Groll M, Dickschat JS. Induced-fit mechanism in class I terpene cyclases. Angew Chem Int Ed Engl. 2014;53:7652–6. https://doi.org/10.1002/anie.201403648.
24. Rabe P, Schmitz T, Dickschat JS. Mechanistic investigations on six bacterial terpene cyclases. Beilstein J Org Chem. 2016;12:1839–50. https://doi.org/10.3762/bjoc.12.173.

Mischko *et al. Microb Cell Fact*   (2018) 17:164

Page 12 of 13

25. Dickschat JS. Bacterial terpene cyclases. Nat Prod Rep. 2016;33:87–110. https://doi.org/10.1039/c5np00102a.

26. Wenkert E, Bookser BC, Arrhenius TS. Total syntheses of (.+−.)-.alpha.- and (.+−.)-.beta.-copaene and formal total syntheses of (.+−.)-sativene, (.+−.)-cis-sativenediol, and (.+−.)-helminthosporal. J Am Chem Soc. 1992;114:644–54. https://doi.org/10.1021/ja00028a034.

27. Kulkarni YS, Niwa M, Ron E, Snider BB. Synthesis of terpenes containing the bicyclo[3.1.1]heptane ring system by the intramolecular [2+2] cycloaddition reaction of vinylketenes with alkenes. Preparation of chrysanthenone,beta.-pinene,beta.-cis-bergamotene,beta.-trans-bergamotene,beta.-copaene, and.beta.-ylangene and.beta.-ylangenol. J Org Chem. 1987;52:1568–76. https://doi.org/10.1021/jo00384a035.

28. Orjala J, Wright AD, Rali T, Sticher O. Aduncamide, a cytotoxic and antibacterial b-phenylethylamine-derived amide from *Piper aduncum*. Nat Prod Lett. 2006;2:231–6. https://doi.org/10.1080/10575639308043814.

29. Schifrin A, Khatri Y, Kirsch P, Thiel V, Schulz S, Bernhardt R. A single terpene synthase is responsible for a wide variety of sesquiterpenes in *Sorangium cellulosum* Soce56. Org Biomol Chem. 2016;14:3385–93. https://doi.org/10.1039/c6ob00130k.

30. Türkez H, Celik K, Toğar B. Effects of copaene, a tricyclic sesquiterpene, on human lymphocytes cells in vitro. Cytotechnology. 2014;66:597–603. https://doi.org/10.1007/s10616-013-9611-1.

31. Turkez H, Togar B, Tatar A. Tricyclic sesquiterpene copaene prevents H_2O_2-induced neurotoxicity. J Intercult Ethnopharmacol. 2014;3:21. https://doi.org/10.5455/jice.20131229104710.

32. Kendra PE, Owens D, Montgomery WS, Narvaez TI, Bauchan GR, Schnell EQ, et al. α-Copaene is an attractant, synergistic with quercivorol, for improved detection of *Euwallacea nr. fornicatus* (Coleoptera: Curculionidae: Scolytinae). PLoS ONE. 2017;12:e0179416. https://doi.org/10.1371/journal.pone.0179416.

33. Faraldos JA, Miller DJ, González V, Yoosuf-Aly Z, Cascón O, Li A, Allemann RK. A 1,6-ring closure mechanism for (+)-δ-cadinene synthase? J Am Chem Soc. 2012;134:5900–8. https://doi.org/10.1021/ja211820p.

34. Gennadios HA, Gonzalez V, Di Costanzo L, Li A, Yu F, Miller DJ, et al. Crystal structure of (+)-delta-cadinene synthase from Gossypium arboreum and evolutionary divergence of metal binding motifs for catalysis. Biochemistry. 2009;48:6175–83. https://doi.org/10.1021/bi900483b.

35. Agger S, Lopez-Gallego F, Schmidt-Dannert C. Diversity of sesquiterpene synthases in the basidiomycete *Coprinus cinereus*. Mol Microbiol. 2009;72:1181–95. https://doi.org/10.1111/j.1365-2958.2009.06717.x.

36. Faraldos JA, O'Maille PE, Dellas N, Noel JP, Coates RM. Bisabolyl-derived sesquiterpenes from tobacco 5-epi-aristolochene synthase-catalyzed cyclization of (2Z,6E)-farnesyl diphosphate. J Am Chem Soc. 2010;132:4281–9. https://doi.org/10.1021/ja909886q.

37. Quin MB, Flynn CM, Wawrzyn GT, Choudhary S, Schmidt-Dannert C. Mushroom hunting by using bioinformatics: application of a predictive framework facilitates the selective identification of sesquiterpene synthases in basidiomycota. ChemBioChem. 2013;14:2480–91. https://doi.org/10.1002/cbic.201300349.

38. Hirte M, Meese N, Mertz M, Fuchs M, Brück TB. Insights into the bifunctional aphidicolin-16-β-ol synthase through rapid biomolecular modeling approaches. Front Chem. 2018;6:389. https://doi.org/10.3389/fchem.2018.00101.

39. Lin Y-L, Ma L-T, Lee Y-R, Shaw J-F, Wang S-Y, Chu F-H. Differential gene expression network in terpenoid synthesis of *Antrodia cinnamomea* in mycelia and fruiting bodies. J Agric Food Chem. 2017;65:1874–86. https://doi.org/10.1021/acs.jafc.6b05386.

40. Engels B, Heinig U, Grothe T, Stadler M, Jennewein S. Cloning and characterization of an *Armillaria gallica* cDNA encoding protoilludene synthase, which catalyzes the first committed step in the synthesis of antimicrobial melleolides. J Biol Chem. 2011;286:6871–8. https://doi.org/10.1074/jbc.M110.165845.

41. Zhou H, Yang Y-L, Zeng J, Zhang L, Ding Z-H, Zeng Y. Identification and characterization of a δ-cadinol synthase potentially involved in the formation of Boreovibrins in *Boreostereum vibrans* of Basidiomycota. Nat Prod Bioprospect. 2016;6:167–71. https://doi.org/10.1007/s13659-016-0096-4.

42. Wawrzyn GT, Quin MB, Choudhary S, López-Gallego F, Schmidt-Dannert C. Draft genome of *Omphalotus olearius* provides a predictive framework for sesquiterpenoid natural product biosynthesis in Basidiomycota. Chem Biol. 2012;19:772–83. https://doi.org/10.1016/j.chembiol.2012.05.012.

43. Yap H-YY, Muria-Gonzalez MJ, Kong B-H, Stubbs KA, Tan CS, Ng S-T, et al. Heterologous expression of cytotoxic sesquiterpenoids from the medicinal mushroom *Lignosus rhinocerotis* in yeast. Microb Cell Fact. 2017;16:103. https://doi.org/10.1186/s12934-017-0713-x.

44. Benedict C. The enzymatic formation of δ-cadinene from farnesyl diphosphate in extracts of cotton. Phytochemistry. 1995;39:327–31. https://doi.org/10.1016/0031-9422(95)00066-G.

45. Steele CL, Crock J, Bohlmann J, Croteau R. Sesquiterpene synthases from grand fir (*Abies grandis*). Comparison of constitutive and wound-induced activities, and cDNA isolation, characterization, and bacterial expression of delta-selinene synthase and gamma-humulene synthase. J Biol Chem. 1998;273:2078–89. https://doi.org/10.1074/jbc.273.4.2078.

46. Yoshikuni Y, Martin VJJ, Ferrin TE, Keasling JD. Engineering cotton (+)-delta-cadinene synthase to an altered function: germacrene D-4-ol synthase. Chem Biol. 2006;13:91–8. https://doi.org/10.1016/j.chembiol.2005.10.016.

47. Lopez-Gallego F, Agger SA, Abate-Pella D, Distefano MD, Schmidt-Dannert C. Sesquiterpene synthases Cop4 and Cop6 from *Coprinus cinereus*: catalytic promiscuity and cyclization of farnesyl pyrophosphate geometric isomers. ChemBioChem. 2010;11:1093–106. https://doi.org/10.1002/cbic.200900671.

48. Asadollahi MA, Maury J, Schalk M, Clark A, Nielsen J. Enhancement of farnesyl diphosphate pool as direct precursor of sesquiterpenes through metabolic engineering of the mevalonate pathway in *Saccharomyces cerevisiae*. Biotechnol Bioeng. 2010;106:86–96. https://doi.org/10.1002/bit.22668.

49. Da Lage J-L, Binder M, Hua-Van A, Janeček S, Casane D. Gene make-up: rapid and massive intron gains after horizontal transfer of a bacterial α-amylase gene to Basidiomycetes. BMC Evol Biol. 2013;13:40. https://doi.org/10.1186/1471-2148-13-40.

50. Khersonsky O, Tawfik DS. Enzyme promiscuity—evolutionary and mechanistic aspects. In: Mander LN, Liu H-W, editors. Comprehensive natural products II chemistry and biology. 1st ed. Amsterdam: Elsevier; 2010. p. 47–88. https://doi.org/10.1016/b978-008045382-8.00155-6.

51. Bülow N, König WA. The role of germacrene D as a precursor in sesquiterpene biosynthesis: investigations of acid catalyzed, photochemically and thermally induced rearrangements. Phytochemistry. 2000;55:141–68. https://doi.org/10.1016/S0031-9422(00)00266-1.

52. da Silva JK, da Trindade R, Alves NS, Figueiredo PL, Maia JGS, Setzer WN. Essential oils from neotropical piper species and their biological activities. Int J Mol Sci. 2017. https://doi.org/10.3390/ijms18122571.

53. Bos R, Woerdenbag HJ, Kayser O, Quax WJ, Ruslan K, Elfami. Essential oil constituents of *Piper cubeba* L. fils. from Indonesia. J Essent Oil Res. 2007;19:14–7. https://doi.org/10.1080/10412905.2007.9699217.

54. Vonášek F, Herout V, Šorm F. On terpenes. CVII. The composition of essential oil from false cubebs and the structure of cubeb camphor. Collect Czech Chem Commun. 1960;25:919–26. https://doi.org/10.1135/cccc19600919.

55. Adams RP, Shao-Zhen Z, Ge-lin C. The volatile leaf oil of *Juniperus convallium* Rehd. & Wils. from Gansu, China. J Essent Oil Res. 1993;5:571–3. https://doi.org/10.1080/10412905.1993.9698281.

56. Martin DM, Aubourg S, Schouwey MB, Daviet L, Schalk M, Toub O, et al. Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. BMC Plant Biol. 2010;10:226. https://doi.org/10.1186/1471-2229-10-226.

57. Göpfert JC, Macnevin G, Ro D-K, Spring O. Identification, functional characterization and developmental regulation of sesquiterpene synthases from sunflower capitate glandular trichomes. BMC Plant Biol. 2009;9:86. https://doi.org/10.1186/1471-2229-9-86.

58. Xie X, Kirby J, Keasling JD. Functional characterization of four sesquiterpene synthases from *Ricinus communis* (castor bean). Phytochemistry. 2012;78:20–8. https://doi.org/10.1016/j.phytochem.2012.02.022.

59. Moncada J, Tamayo JA, Cardona CA. Techno-economic and environmental assessment of essential oil extraction from Oregano (*Origanum vulgare*) and Rosemary (*Rosmarinus officinalis*) in Colombia. J Clean Prod. 2016;112:172–81. https://doi.org/10.1016/j.jclepro.2015.09.067.

60. Gong HY, Liu WH, Lv GY, Zhou X. Analysis of essential oils of *Origanum vulgare* from six production areas of China and Pakistan. Revista Brasileira de Farmacognosia. 2014;24:25–32. https://doi.org/10.1590/0102-695X2014241434.

Mischko *et al. Microb Cell Fact*   (2018) 17:164

Page 13 of 13

61. Müller J, Oehler S, Müller-Hill B. Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. J Mol Biol. 1996;257:21–9. https://doi.org/10.1006/jmbi.1996.0143.

62. Riesenberg D, Schulz V, Knorre WA, Pohl H-D, Korz D, Sanders EA, et al. High cell density cultivation of *Escherichia coli* at controlled specific growth rate. J Biotechnol. 1991;20:17–27. https://doi.org/10.1016/0168-1656(91)90032-Q.

63. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, Xu J. Template-based protein structure modeling using the RaptorX web server. Nat Protoc. 2012;7:1511–22. https://doi.org/10.1038/nprot.2012.085.

64. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph. 1996;14:33–8. https://doi.org/10.1016/0263-7855(96)00018-5.

65. Roberts E, Eargle J, Wright D, Luthey-Schulten Z. MultiSeq: unifying sequence and structure data for evolutionary analysis. BMC Bioinform. 2006;7:382. https://doi.org/10.1186/1471-2105-7-382.

66. Stone JE, University of Missouri-Rolla. Graduate School. An efficient library for parallel ray tracing and animation: a thesis presented to the Faculty of the Graduate School of the University of Missouri-Rolla in partial fulfillment of requirements for the Degree of Master of Science in Computer Science: University of Missouri-Rolla; 1998.

Outlook

Strain Optimization – High through put screening

To further raise the efficiency of the terpene production platform, the terpenoid metabolic flux inside the cell needs to be further optimized and increased. Novel approaches to engineer high product yielding hosts take mathematic model calculations into account[74, 115]. However, the predicted model, btw. optimal combinatorics of the enzyme cascade in optimal expression strength, still requires elaborate computational effort and subsequent experimental evaluation. Moreover, it still remains difficult to predict the behavior of a genetically modified production host. Accordingly, iterative optimization approaches to raise terpene yields are normally applied[82, 116, 117]. However, such methodology requires man and lab power or an efficient and automated screening system.

The generation of a constitutive production platform for lycopene formation is an optimal basis for the development of an automated screening system for strain optimization[9]. The constitutive expression provides for a direct correlation between cell growth and lycopene formation. Accordingly, the production platform can be randomly varied and quickly evaluated. For example, the impact of expression strength modulation or the addition of other proteins to the expression system can investigated easily.

One strategy to optimize the production platform would be the random substitution of bases within the polycistronic gene sequence by error prone PCR or other mutation inducing experiments. In this context, hundreds of mutants generated require a specific and precise screening. To identify a high performing clone from the rest, automated cell sorting based on the lycopene content per cell should be considered.

Design of experiment – Process optimization

Fermentation process optimization is a major task in industrial process development[10]. In this context, design of experiments (DoE) minimizes experimental effort by mathematic-based prediction of the optimal parameter's values. The system is characterized in different dimensions and each dimension is delineated by two defined settings. Accordingly, a system can only be operated in the dimension defined. The experimental characterization of the dimension's corners and their outcome is used to mathematically identify correlations and to predict the optimum value of the screened dimensions. DoE is used in biotechnology to identify optimal operation settings for growth temperature or oxygen uptake rates, but also for media optimization[118].

In the context of this study, process optimization was carried out step by step and hence, only optimized in one dimension at a time. No interacting correlations were investigated. This should be addressed by future researchers, which potentially will result in a terpene titer increase.

Another impact identified during this research was the composition of the media applied[9]. For example, using tryptone/peptone as nutrient supplement in the media resulted in extensive production of indole that lowers the terpene yield substantially[82]. Moreover, substitution of major parts of yeast extract by corn steep liquor lead to an increase in TD concentration per cell dry weight[9]. Hence, variation of the components within the culture media and feed will affectively influence the terpene yield outcome. A factorial DoE should be applied to investigate the impact of media components in order to identify correlations between terpene production and media components. The constitutive lycopene expression platform should be used for the reasons mentioned above.

Process optimization

The efficiency of the terpene production process can be further optimized. One strategy is the development of a continuous fed-batch fermentation process[119]. The constitutive terpene production platform provides for a continuous production of terpene as long as the cells keep on growing. By mathematic calculations the feed rate and concentration can be determined for an optimal process flow. The extended production period will result in a higher time-space yield, which might be relevant for industrial scale production.

One particular processing problem during downstream operations was the high hydrophobicity of the terpenoid molecules. Extraction and purification were affected by emulsion formation and solubility issues. Therefore, we devised a new purification process of microbial derived terpenoids by liquid/liquid chromatography[8, 9]. Distribution coefficient of terpenoids in biphasic mixtures based on classical thermodynamics were the most interesting fact, I completely neglected in the beginning of my PhD study. Biphasic systems should be screened and identified that can remove major contaminants of the microbial derived extract to reduce the complexity of the extract. This simple and fast pre-purification can lead to an already concentrated terpene rich extract that can be directly applied for HPLC purification without preparation issues.

Exploring the fungal genetic code

Time and cost needed to solve and annotate genomes has drastically decreased in the last decades[120]. Accordingly, lots of genetic information are freely available (e.g. National Center for Biotechnology Information). Most of this information are solely annotated enzyme function and only few of the enzymes have been characterized in experimental settings so far. The characterization of the putative terpene synthases of *C. puteana* were very successful and expression in *E. coli* was possible without limitations[85]. Moreover, the phylogenetic classification revealed significant correlation between sesquiterpene product and sequence similarity[85]. This strategy should be evaluated and further improved.

Clustering of biosynthetic routes in fungi also allow for the specific selection of CYPs potentially modifying the macrocycle core[121, 122].

Functionalization of terpenes by CYPs

Hydroxylation reactions of terpenes are still challenging[40, 42]. In the context of this study, a new strategy was developed that facilitates activity screening of CYPs and reductases on diverse terpene scaffolds by an in-vitro biotransformation approach[107, 108]. This system needs to be further improved, initially by raising the promoter expression strength and stationary phase induction. During our experiments we also identified a strong influence of the N-terminal sequence. This protein sequence should be optimized to guarantee soluble und functional protein expression.

The transferability of reductase systems should be evaluated by other CYPs. 5HTD is a special CYP because it performs an epoxidation reaction of TD[40]. It might be possible that only for CYP epoxidation reactions replacement of a flavin dependent reductase remains functional. A negative control (expressing CotB3, not expressing a heterologous reductase) has to be conducted for the hydroxylation of cycloocat-9-en-7-ol in order to prove that the electron transfer can also be conducted by a CPR.

CotB3 is a promising CYP for new-to-nature combinations due to its activity on diverse diterpene macrocylces[107]. Virtual protein modelling should be applied to specifically alter the enzymes active site[26]. Enzyme mutants could have improved activity on non-natural substrates.

References

1. Markard, J., R. Raven, and B. Truffer, *Sustainability transitions: An emerging field of research and its prospects.* Research Policy, 2012. **41**(6): p. 955-967.
2. Kircher, M., *The transition to a bio-economy: national perspectives.* Biofuels, Bioproducts and Biorefining, 2012. **6**(3): p. 240-245.
3. (BMBF), F.M.o.E.a.R., *National Research Strategy – Bioeconomy 2030.* [Online]. Available at: https://www.pflanzenforschung.de/files/4514/7886/1937/German_bioeconomy_Strategy_2030.pdf (February 20, 2019).
4. Bugge, M.M., T. Hansen, and A. Klitkou, *What Is the Bioeconomy? A Review of the Literature.* Sustainability, 2016. **8**(7): p. 691.
5. Pfau, S.F., et al., *Visions of Sustainability in Bioeconomy Research.* Sustainability, 2014. **6**(3): p. 1222.
6. OECD, *Future prospects for industrial biotechnology.* [Online]. OECD publishing: 13-15 (2011). Available online at: https://read.oecd-ilibrary.org/science-and-technology/future-prospects-for-industrial-biotechnology_9789264126633-en#page1 (February 20, 2019).
7. Lokko, Y., et al., *Biotechnology and the bioeconomy—Towards inclusive and sustainable industrial development.* New Biotechnology, 2018. **40**: p. 5-10.
8. Mischko, W., et al., *Modular biomanufacturing for a sustainable production of terpenoid-based insect deterrents.* Green Chemistry, 2018. **20**(11): p. 2637-2650.
9. Hirte, M., et al., *From microbial upcycling to biology-oriented synthesis: combining whole-cell production and chemo-enzymatic functionalization for sustainable taxanoid delivery.* Green Chemistry, 2018. **20**(23): p. 5374-5384.
10. Wenda, S., et al., *Industrial biotechnology—the future of green chemistry?* Green Chemistry, 2011. **13**(11): p. 3007-3047.
11. Research, F.M.o.E.a., *E:Bio Innovation Competition Systems Biology.* 2010, [online] Available at: https://www.bmbf.de/foerderungen/bekanntmachung.php?B=608 (February 20, 2019).
12. Pateraki, I., A.M. Heskes, and B. Hamberger, *Cytochromes P450 for Terpene Functionalisation and Metabolic Engineering*, in *Biotechnology of Isoprenoids*, J. Schrader and J. Bohlmann, Editors. 2015, Springer International Publishing: Cham. p. 107-139.
13. Hanson, J.R., *Diterpenoids of terrestrial origin.* Natural product reports, 2015. **32**(1): p. 76-87.
14. Barros de Alencar, M.V.O., et al., *Diterpenes as lead molecules against neglected tropical diseases.* Phytotherapy Research, 2017. **31**(2): p. 175-201.
15. Islam, M.T., et al., *Diterpenes: Advances in Neurobiological Drug Research.* Phytotherapy Research, 2016. **30**(6): p. 915-928.
16. Jansen, D.J. and R.A. Shenvi, *Synthesis of medicinally relevant terpenes: reducing the cost and time of drug discovery.* Future medicinal chemistry, 2014. **6**(10): p. 1127-1148.
17. Jones, B., *Diterpenoids: Types, Functions and Research.* 2017: Nova Science Publishers, Incorporated.
18. Kemper, K., et al., *Opportunities and challenges for the sustainable production of structurally complex diterpenoids in recombinant microbial systems.* Beilstein Journal of Organic Chemistry, 2017. **13**: p. 845-854.
19. Lorenzo, C. and A. Eugenio, *Use of Terpenoids as Natural Flavouring Compounds in Food Industry.* Recent Patents on Food, Nutrition & Agriculture, 2011. **3**(1): p. 9-16.
20. Christianson, D.W., *Structural and Chemical Biology of Terpenoid Cyclases.* Chemical Reviews, 2017. **117**(17): p. 11570-11648.
21. Thomson, R.H., *The Chemistry of Natural Products.* 1993: Springer Netherlands.
22. Breitmaier, E., *Terpenes: Flavors, Fragrances, Pharmaca, Pheromones.* 2006: Wiley.
23. Schrepfer, P., et al., *Identification of amino acid networks governing catalysis in the closed complex of class I terpene synthases.* Proceedings of the National Academy of Sciences, 2016. **113**(8): p. E958-E967.

24. Gao, Y., R.B. Honzatko, and R.J. Peters, *Terpenoid synthase structures: a so far incomplete view of complex catalysis.* Natural Product Reports, 2012. **29**(10): p. 1153-1175.

25. Zhou, K., et al., *Insights into Diterpene Cyclization from Structure of Bifunctional Abietadiene Synthase from Abies grandis.* Journal of Biological Chemistry, 2012. **287**(9): p. 6840-6850.

26. Hirte, M., et al., *Insights Into the Bifunctional Aphidicolan-16-ß-ol Synthase Through Rapid Biomolecular Modeling Approaches.* Frontiers in Chemistry, 2018. **6**(101).

27. Görner, C., et al., *Identification, characterization and molecular adaptation of class I redox systems for the production of hydroxylated diterpenoids.* Microbial Cell Factories, 2016. **15**: p. 86.

28. Mafu, S., et al., *Probing the promiscuity of ent-kaurene oxidases via combinatorial biosynthesis.* Proceedings of the National Academy of Sciences, 2016. **113**(9): p. 2526-2531.

29. Maier, M.E., *Design and synthesis of analogues of natural products.* Organic & Biomolecular Chemistry, 2015. **13**(19): p. 5302-5343.

30. Wani, M.C., et al., *Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from Taxus brevifolia.* Journal of the American Chemical Society, 1971. **93**(9): p. 2325-2327.

31. Yeung, T.K., et al., *The Mode of Action of Taxol: Apoptosis at Low Concentration and Necrosis at High Concentration.* Biochemical and Biophysical Research Communications, 1999. **263**(2): p. 398-404.

32. Schiff, P.B., J. Fant, and S.B. Horwitz, *Promotion of microtubule assembly in vitro by taxol.* Nature, 1979. **277**(5698): p. 665-667.

33. Dunn, P.J., A. Wells, and M.T. Williams, *Green Chemistry in the Pharmaceutical Industry.* 2010: Wiley.

34. Nicolaou, K.C., et al., *Total synthesis of taxol.* Nature, 1994. **367**(6464): p. 630-634.

35. Leone, L.M. and S.C. Roberts, *Accessing Anti-cancer Natural Products by Plant Cell Culture*, in *Natural Products and Cancer Drug Discovery*, F.E. Koehn, Editor. 2013, Springer New York: New York, NY. p. 193-211.

36. Jennewein, S. and R. Croteau, *Taxol: biosynthesis, molecular genetics, and biotechnological applications.* Appl Microbiol Biotechnol, 2001. **57**(1-2): p. 13-9.

37. Kingston, D.G.I., A.A. Molinero, and J.M. Rimoldi, *The Taxane Diterpenoids*, in *Fortschritte der Chemie organischer Naturstoffe / Progress in the Chemistry of Organic Natural Products*, W. Herz, et al., Editors. 1993, Springer Vienna: Vienna. p. 1-165.

38. Köksal, M., et al., *Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis.* Nature, 2010. **469**: p. 116.

39. Barton, N.A., et al., *Accessing low-oxidation state taxanes: is taxadiene-4(5)-epoxide on the taxol biosynthetic pathway?* Chemical Science, 2016. **7**(5): p. 3102-3107.

40. Edgar, S., et al., *Mechanistic Insights into Taxadiene Epoxidation by Taxadiene-5a-Hydroxylase.* ACS Chemical Biology, 2016. **11**(2): p. 460-469.

41. Biggs, B.W., et al., *Overcoming heterologous protein interdependency to optimize P450-mediated Taxol precursor synthesis in Escherichia coli.* Proceedings of the National Academy of Sciences, 2016. **113**(12): p. 3209-3214.

42. Biggs, B.W., et al., *Orthogonal Assays Clarify the Oxidative Biochemistry of Taxol P450 CYP725A4.* ACS Chemical Biology, 2016. **11**(5): p. 1445-1451.

43. Dalziel, W., et al., *The structure and absolute configuration of the antibiotic aphidicolin: a tetracyclic diterpenoid containing a new ring system.* Journal of the Chemical Society, Perkin Transactions 1, 1973(0): p. 2841-2851.

44. Starratt, A.N. and S.R. Loschiavo, *The production of aphidicolin by Nigrospora sphaerica.* Canadian Journal of Microbiology, 1974. **20**(3): p. 416-417.

45. Toyomasu, T., et al., *Cloning of a Gene Cluster Responsible for the Biosynthesis of Diterpene Aphidicolin, a Specific Inhibitor of DNA Polymerase a.* Bioscience, Biotechnology, and Biochemistry, 2004. **68**(1): p. 146-152.

46. Baranovskiy, A.G., et al., *Structural basis for inhibition of DNA replication by aphidicolin.* Nucleic Acids Research, 2014. **42**(22): p. 14013-14021.

47. Ikegami, S., et al., *Aphidicolin prevents mitotic cell division by interfering with the activity of DNA polymerase-a.* Nature, 1978. **275**: p. 458.

48. Pedrali-Noy, G., et al., *Synchronization of HeLa cell cultures by inhibition of DNA polymerase alpha with aphidicolin.* Nucleic Acids Research, 1980. **8**(2): p. 377-387.

49. Edwards, T.G., et al., *Human Papillomavirus Episome Stability Is Reduced by Aphidicolin and Controlled by DNA Damage Response Pathways.* Journal of Virology, 2013. **87**(7): p. 3979-3989.

50. Starczewska, E., et al., *Targeting DNA repair with aphidicolin sensitizes primary chronic lymphocytic leukemia cells to purine analogs.* Oncotarget, 2016. **7**(25): p. 38367-38379.

51. Fujii, R., et al., *Total Biosynthesis of Diterpene Aphidicolin, a Specific Inhibitor of DNA Polymerase a: Heterologous Expression of Four Biosynthetic Genes in Aspergillus oryzae.* Bioscience, Biotechnology, and Biochemistry, 2011. **75**(9): p. 1813-1817.

52. Oikawa, H., et al., *Proposed Mechanism for the Reaction Catalyzed by a Diterpene Cyclase, Aphidicolan-16β-ol Synthase: Experimental Results on Biomimetic Cyclization and Examination of the Cyclization Pathway by ab Initio Calculations.* Journal of the American Chemical Society, 2002. **124**(31): p. 9145-9153.

53. TAKAYUKI AOYAMA, H.N., YASUHIKO MURAOKA, TAKAAKI AOYAGI, TOMIO TAKEUCHI, *THE STRUCTURE OF CYCLOOCTATIN, A NEW INHIBITOR OF LYSOPHOSPHOLIPASE.* J-Stage, 1992. **45**(10): p. 1703-1704.

54. Janke, R., et al., *The first structure of a bacterial diterpene cyclase: CotB2.* Acta Crystallogr D Biol Crystallogr, 2014. **70**(Pt 6): p. 1528-37.

55. Görner, C., et al., *Targeted Engineering of Cyclooctat-9-en-7-ol Synthase: A Stereospecific Access to Two New Non-natural Fusicoccane-Type Diterpenes.* ChemCatChem, 2013. **5**(11): p. 3289-3298.

56. Kim, S.-Y., et al., *Cloning and Heterologous Expression of the Cyclooctatin Biosynthetic Gene Cluster Afford a Diterpene Cyclase and Two P450 Hydroxylases.* Chemistry & Biology, 2009. **16**(7): p. 736-743.

57. Görner, C., et al., *Identification, characterization and molecular adaptation of class I redox systems for the production of hydroxylated diterpenoids.* Microbial Cell Factories, 2016. **15**(1): p. 86.

58. Llc, B., *Carotenoids: Carotene, Lycopene, Astaxanthin, Carotenoid, Beta-Carotene, Lutein, Tretinoin, Xanthophyll, Phytofluene, Canthaxanthin, Phytoene.* 2010: General Books LLC.

59. Landrum, J.T., *Carotenoids: Physical, Chemical, and Biological Functions and Properties.* 2009: CRC Press.

60. Krinsky, N.I., *Carotenoid protection against oxidation*, in *Pure and Applied Chemistry*. 1979. p. 649.

61. Krinsky, N.I., S.T. Mayne, and H. Sies, *Carotenoids in Health and Disease.* 2004: CRC Press.

62. Vandamme, E.J. and J.L. Revuelta, *Industrial Biotechnology of Vitamins, Biopigments, and Antioxidants.* 2016: Wiley.

63. Alper, H., et al., *Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli.* Metabolic Engineering, 2005. **7**(3): p. 155-164.

64. Chen, Y., et al., *Lycopene overproduction in Saccharomyces cerevisiae through combining pathway engineering with host engineering.* Microbial Cell Factories, 2016. **15**(1): p. 113.

65. Leavell, M.D., D.J. McPhee, and C.J. Paddon, *Developing fermentative terpenoid production for commercial usage.* Current Opinion in Biotechnology, 2016. **37**: p. 114-119.

66. Moser, S. and H. Pichler, *Identifying and engineering the ideal microbial terpenoid production host.* Applied Microbiology and Biotechnology, 2019. **103**(14): p. 5501-5516.

67. Paddon, C.J., et al., *High-level semi-synthetic production of the potent antimalarial artemisinin.* Nature, 2013. **496**(7446): p. 528-532.

68. Tsuruta, H., et al., *High-Level Production of Amorpha-4,11-Diene, a Precursor of the Antimalarial Agent Artemisinin, in Escherichia coli.* PLOS ONE, 2009. **4**(2): p. e4489.

69. Meadows, A.L., et al., *Rewriting yeast central carbon metabolism for industrial isoprenoid production.* Nature, 2016. **537**: p. 694.

70. Larroude, M., et al., *A synthetic biology approach to transform Yarrowia lipolytica into a competitive biotechnological producer of β-carotene.* Biotechnology and Bioengineering, 2018. **115**(2): p. 464-472.

71. Beuttler, H., et al., *Biosynthesis of zeaxanthin in recombinant Pseudomonas putida.* Applied Microbiology and Biotechnology, 2011. **89**(4): p. 1137-1147.

72. Zhou, K., et al., *Optimization of amorphadiene synthesis in bacillus subtilis via transcriptional, translational, and media modulation.* Biotechnology and Bioengineering, 2013. **110**(9): p. 2556-2561.

73. Engels, B., P. Dahm, and S. Jennewein, *Metabolic engineering of taxadiene biosynthesis in yeast as a first step towards Taxol (Paclitaxel) production.* Metabolic Engineering, 2008. **10**(3): p. 201-206.

74. Boghigian, B.A., et al., *Analysis of heterologous taxadiene production in K- and B-derived Escherichia coli.* Applied Microbiology and Biotechnology, 2012. **93**(4): p. 1651-1661.

75. Wang, W., et al., *Bacteriophage T7 transcription system: an enabling tool in synthetic biology.* Biotechnology Advances, 2018. **36**(8): p. 2129-2137.

76. Jones, J.A., et al., *ePathOptimize: A Combinatorial Approach for Transcriptional Balancing of Metabolic Pathways.* Scientific Reports, 2015. **5**: p. 11301.

77. Moore, J.T., et al., *Overcoming Inclusion Body Formation in a High-Level Expression System.* Protein Expression and Purification, 1993. **4**(2): p. 160-163.

78. Nikel, P.I. and V. de Lorenzo, *Pseudomonas putida as a functional chassis for industrial biocatalysis: From native biochemistry to trans-metabolism.* Metabolic Engineering, 2018. **50**: p. 142-155.

79. Wagner, S.G., et al., *pTRA – A reporter system for monitoring the intracellular dynamics of gene expression.* PLOS ONE, 2018. **13**(5): p. e0197420.

80. Bernhard, B., *The lac operon*, in *Biological Chemistry*. 1997. p. I.

81. Arditti, R.R., J.G. Scaife, and J.R. Beckwith, *The nature of mutants in the lac promoter region.* Journal of Molecular Biology, 1968. **38**(3): p. 421-426.

82. Ajikumar, P.K., et al., *Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in Escherichia coli.* Science, 2010. **330**(6000): p. 70-74.

83. Shetty, R.P., D. Endy, and T.F. Knight, *Engineering BioBrick vectors from BioBrick parts.* Journal of Biological Engineering, 2008. **2**(1): p. 5.

84. Zelcbuch, L., et al., *Spanning high-dimensional expression space using ribosome-binding site combinatorics.* Nucleic Acids Research, 2013. **41**(9): p. e98.

85. Mischko, W., et al., *Identification of sesquiterpene synthases from the Basidiomycota Coniophora puteana for the efficient and highly selective β-copaene and cubebol production in E. coli.* Microbial Cell Factories, 2018. **17**(1): p. 164.

86. Snyder, L.R., J.J. Kirkland, and J.L. Glajch, *Practical HPLC Method Development.* 2012: Wiley.

87. Marston, A. and K. Hostettmann, *Counter-current chromatography as a preparative tool —applications and perspectives.* Journal of Chromatography A, 1994. **658**(2): p. 315-341.

88. Klamt, A. and F. Eckert, *COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids.* Fluid Phase Equilibria, 2000. **172**(1): p. 43-72.

89. Nelson, D.R., et al., *P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature.* Pharmacogenetics, 1996. **6**(1): p. 1-42.

90. de Montellano, P.R.O., *Cytochrome P450: Structure, Mechanism, and Biochemistry.* 2007: Springer US.

91. Jung, C., *The mystery of cytochrome P450 Compound I: A mini-review dedicated to Klaus Ruckpaul.* Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics, 2011. **1814**(1): p. 46-57.

92. Giulivi, C. and E. Cadenas, *Heme Protein Radicals: Formation, Fate, and Biological Consequences.* Free Radical Biology and Medicine, 1998. **24**(2): p. 269-279.

93. Hannemann, F., et al., *Cytochrome P450 systems—biological variations of electron transport chains.* Biochimica et Biophysica Acta (BBA) - General Subjects, 2007. **1770**(3): p. 330-344.

94. Schiffler, B. and R. Bernhardt, *Bacterial (CYP101) and mitochondrial P450 systems—how comparable are they?* Biochemical and Biophysical Research Communications, 2003. **312**(1): p. 223-228.

95.	Kumar, S., *Engineering cytochrome P450 biocatalysts for biotechnology, medicine and bioremediation.* Expert Opinion on Drug Metabolism & Toxicology, 2010. **6**(2): p. 115-131.

96.	Munro, A.W., et al., *P450 BM3: the very model of a modern flavocytochrome.* Trends in Biochemical Sciences, 2002. **27**(5): p. 250-257.

97.	McLean, K.J., H.M. Girvan, and A.W. Munro, *Cytochrome P450/redox partner fusion enzymes: biotechnological and toxicological prospects.* Expert Opinion on Drug Metabolism & Toxicology, 2007. **3**(6): p. 847-863.

98.	Lipinski, C.A., et al., *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.* Advanced Drug Delivery Reviews, 1997. **23**(1): p. 3-25.

99.	Lipinski, C.A., *Lead- and drug-like compounds: the rule-of-five revolution.* Drug Discovery Today: Technologies, 2004. **1**(4): p. 337-341.

100.	Görner, C., et al., *Stereoselective chemo-enzymatic oxidation routes for (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene.* Frontiers in Microbiology, 2015. **6**: p. 1115.

101.	Le-Huu, P., et al., *Chemo-, Regio-, and Stereoselective Oxidation of the Monocyclic Diterpenoid β-Cembrenediol by P450 BM3.* ACS Catalysis, 2015. **5**(3): p. 1772-1780.

102.	Oberleitner, N., et al., *From waste to value – direct utilization of limonene from orange peel in a biocatalytic cascade reaction towards chiral carvolactone.* Green Chemistry, 2017. **19**(2): p. 367-371.

103.	Sigel, A., H. Sigel, and R.K.O. Sigel, *The Ubiquitous Roles of Cytochrome P450 Proteins.* 2007: Wiley.

104.	Lovering, F., J. Bikker, and C. Humblet, *Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success.* Journal of Medicinal Chemistry, 2009. **52**(21): p. 6752-6756.

105.	Lovering, F., *Escape from Flatland 2: complexity and promiscuity.* MedChemComm, 2013. **4**(3): p. 515-519.

106.	Ajikumar, P.K., et al., *Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli.* Science, 2010. **330**.

107.	Mertz, M., *Herstellung vielfältiger Diterpenoide durch Arten übergreifende Verknüpfung von Diterpenzyklasen mit Cytochrom P450 Monooxygenasen*, in *Werner Siemens Chair of synthetic biotechnology.* 2018, TU Munich: Munich.

108.	Chandarasekaran, S., *BIOTRANSFORMATION OF DITERPENOIDS in whole and resting cells*, in *Werner Siemens Chair of synthetic biotechnology.* 2018, TU Munich: Munich.

109.	Karplus, M. and J.A. McCammon, *Molecular dynamics simulations of biomolecules.* Nature Structural Biology, 2002. **9**(9): p. 646-652.

110.	Eswar, N., et al., *Comparative Protein Structure Modeling Using Modeller.* Current Protocols in Bioinformatics, 2006. **15**(1): p. 5.6.1-5.6.30.

111.	Pettersen, E.F., et al., *UCSF Chimera—A visualization system for exploratory research and analysis.* Journal of Computational Chemistry, 2004. **25**(13): p. 1605-1612.

112.	Trott, O. and A.J. Olson, *AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.* Journal of Computational Chemistry, 2010. **31**(2): p. 455-461.

113.	Hanwell, M.D., et al., *Avogadro: an advanced semantic chemical editor, visualization, and analysis platform.* Journal of Cheminformatics, 2012. **4**(1): p. 17.

114.	Vanommeslaeghe, K., et al., *CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields.* Journal of Computational Chemistry, 2010. **31**(4): p. 671-690.

115.	Alper, H., K. Miyaoku, and G. Stephanopoulos, *Characterization of lycopene-overproducing E. coli strains in high cell density fermentations.* Applied Microbiology and Biotechnology, 2006. **72**(5): p. 968-974.

116.	Morrone, D., et al., *Increasing diterpene yield with a modular metabolic engineering system in E. coli: comparison of MEV and MEP isoprenoid precursor pathway engineering.* Applied Microbiology and Biotechnology, 2010. **85**(6): p. 1893-1906.

117.	Chang, M.C.Y. and J.D. Keasling, *Production of isoprenoid pharmaceuticals by engineered microbes.* Nat Chem Biol, 2006. **2**(12): p. 674-681.

118. Lee, K.-M. and D.F. Gilmore, *Statistical experimental design for bioprocess modeling and optimization analysis.* Applied Biochemistry and Biotechnology, 2006. **135**(2): p. 101-115.

119. Malek, I., *Continuous Cultivation of Microorganisms: Proceedings of the Second Symposium held in Prague June 18-23, 1962.* 2012: Elsevier Science.

120. Davies, K., *The $1,000 Genome: The Revolution in DNA Sequencing and the New Era of Personalized Medicine.* 2010: Free Press.

121. Oikawa, H., et al., *Cloning and Functional Expression of cDNA Encoding Aphidicolan-16β-ol Synthase: A Key Enzyme Responsible for Formation of an Unusual Diterpene Skeleton in Biosynthesis of Aphidicolin.* Journal of the American Chemical Society, 2001. **123**(21): p. 5154-5155.

122. Keller, N.P., G. Turner, and J.W. Bennett, *Fungal secondary metabolism — from biochemistry to genomics.* Nature Reviews Microbiology, 2005. **3**: p. 937.

Part III

Appendix

1
Abbreviations

5HTD	5-α-hydoxytaxadiene
AC	Aphidicolin
API	Active pharmaceutical ingredient
ATP	Adenosine triphosphate
CO	Cyclooctatin
CPR	Cytochrome P450 reductase
CTP	Cytosine triphosphate
CYP	Cytochrome P450 monooxygenase
DMADP	Dimethylallyl diphosphate
Fd	Ferredoxin
FDP	Farnesyl diphosphate
GDP	Geranyl diphosphate
GGDP	Geranylgeranyl diphosphate
IDP	Isopentenyl diphosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
TD	Taxadiene

2
List of figures

3

Protocols

Split_stream.pl


```perl
use strict;
use File::Basename;


my $rtf="";
my $prm="";
my $junk="";

my $r=\$junk;
my $warn_rtf=0;
my $warn_prm=0;

my($basename, $directories, $suffix) = fileparse($ARGV[0]);
$basename=~s/.str$//;
$basename=~s/^toppar_//;
my $ortf="top_$basename.rtf";
my $oprm="par_$basename.prm";


while(my $l=<>) {
        if($l=~/^read rtf card/i) {
                $r=\$rtf;
                if($$r) {
                        $$r.="! WARNING -- ANOTHER rtf SECTION FOUND\n";
                        $warn_rtf++;
                }
        }
        elsif($l=~/^read param? card/i) {
                $r=\$prm;
                if($$r) {
                        $$r.="! WARNING -- ANOTHER para SECTION FOUND\n";
                        $warn_prm++;
                }
        }
        elsif($l=~/^end/i) { $r=\$junk; }
        else {
                $$r .= $l;
        }
}

if($warn_rtf || $warn_prm) {
        print STDERR "Warning - duplicate sections found. You need to revise the outputs marked
with [*]\n";
}

my $w;
$w="*"x$warn_rtf;
```

```perl
print STDERR "Creating $ortf $w\n";
open R,">$ortf";
print R $rtf;
print R "END\n";
close R;

$w="*"x$warn_prm;
print STDERR "Creating $oprm $w\n";
open P,">$oprm";
print P $prm;
print P "END\n";
close P;
```

Generate pdb and psf files by VMD (TK-console command)

For the ligand (requires: ligand.pdb; top_ligand.rtf):

```
package require psfgen
topology top_all36_cgenff.rtf
topology top_ligand.rtf
segment Z {pdb ligand.pdb}
coordpdb ligand.pdb Z
guesscoord
writepdb ligand-VMD.pdb
writepsf ligand-VMD.psf
```

For the protein (requires receptor.pdb):

```
package require psfgen
topology top_all27_prot_lipid.inp
pdbalias residue HIS HSE
pdbalias atom ILE CD1 CD
segment A {pdb receptor.pdb}
coordpdb receptor.pdb A
guesscoord
writepdb receptor-VMD.pdb
writepsf receptor-VMD.psf
```

Add water sphere around the protein in VMD (TK-console command)

```
set molname complex2

mol new ${molname}.psf
mol addfile ${molname}.pdb

### Determine the center of mass of the molecule and store the coordinates
set cen [measure center [atomselect top all] weight mass]
set x1 [lindex $cen 0]
```

```
set y1 [lindex $cen 1]
set z1 [lindex $cen 2]
set max 0

### Determine the distance of the farthest atom from the center of mass
foreach atom [[atomselect top all] get index] {
  set pos [lindex [[atomselect top "index $atom"] get {x y z}] 0]
  set x2 [lindex $pos 0]
  set y2 [lindex $pos 1]
  set z2 [lindex $pos 2]
  set dist [expr pow(($x2-$x1)*($x2-$x1) + ($y2-$y1)*($y2-$y1) + ($z2-$z1)*($z2-$z1),0.5)]
  if {$dist > $max} {set max $dist}
  }

mol delete top

### Solvate the molecule in a water box with enough padding (15 A).
### One could alternatively align the molecule such that the vector
### from the center of mass to the farthest atom is aligned with an axis,
### and then use no padding
package require solvate
solvate ${molname}.psf ${molname}.pdb -t 15 -o del_water

resetpsf
package require psfgen
mol new del_water.psf
mol addfile del_water.pdb
readpsf del_water.psf
coordpdb del_water.pdb

### Determine which water molecules need to be deleted and use a for loop
### to delete them
set wat [atomselect top "same residue as {water and ((x-$x1)*(x-$x1) + (y-$y1)*(y-$y1) + (z-$z1)*(z-$z1))<($max*$max)}"]
set del [atomselect top "water and not same residue as {water and ((x-$x1)*(x-$x1) + (y-$y1)*(y-$y1) + (z-$z1)*(z-$z1))<($max*$max)}"]
set seg [$del get segid]
set res [$del get resid]
set name [$del get name]
for {set i 0} {$i < [llength $seg]} {incr i} {
  delatom [lindex $seg $i] [lindex $res $i] [lindex $name $i]
  }
writepsf ${molname}_ws.psf
writepdb ${molname}_ws.pdb

mol delete top

mol new ${molname}_ws.psf
mol addfile ${molname}_ws.pdb
puts "CENTER OF MASS OF SPHERE IS: [measure center [atomselect top all] weight mass]"
puts "RADIUS OF SPHERE IS: $max"
```

mol delete top

Simulation.conf (run by namd2.exe)

```
###################################################################
###
## JOB DESCRIPTION                              ##
###################################################################
###

# Minimization and Equilibration of
# a protein with a ligand in a water box


###################################################################
###
## ADJUSTABLE PARAMETERS                        ##
###################################################################
###


structure          ../complex_ws.psf
coordinates        ../complex_ws.pdb

set temperature    300
set outputname     complex-MD

firsttimestep      0



###################################################################
###
## SIMULATION PARAMETERS                        ##
###################################################################
###

# Input
paraTypeCharmm          on
parameters        ../par_all27_prot_lipid_na.inp
parameters                ../par_all36_cgenff.prm
parameters                ../par_ligand.prm
mergeCrossterms     yes
temperature         $temperature


# Force-Field Parameters
exclude           scaled1-4
1-4scaling          1.0
cutoff             12.0
switching           on
switchdist          10.0
pairlistdist        14.0
```

```
# Integrator Parameters
timestep          1.0  ;# 2fs/step
rigidBonds        all  ;# needed for 2fs steps
nonbondedFreq     1
fullElectFrequency  2
stepspercycle     10


# Constant Temperature Control
langevin          on    ;# do langevin dynamics
langevinDamping   5     ;# damping coefficient (gamma) of 1/ps
langevinTemp      $temperature
langevinHydrogen  off   ;# don't couple langevin bath to hydrogens


# Periodic Boundary Conditions
cellBasisVector1   150.0   0.   0.0
cellBasisVector2   0.0    150.0  0.0
cellBasisVector3   0.0     0.0   150.0
cellOrigin         6.054964065551758 -0.5199195742607117 -19.352827072143555

wrapAll           on


# PME (for full-system periodic electrostatics)
PME               yes
PMEGridSpacing    1.1

#manual grid definition
#PMEGridSizeX      128
#PMEGridSizeY      128
#PMEGridSizeZ      128


# Constant Pressure Control (variable volume)
useGroupPressure      yes ;# needed for rigidBonds
useFlexibleCell       yes
useConstantArea       no

langevinPiston        on
langevinPistonTarget  1.01325 ;#  in bar -> 1 atm
langevinPistonPeriod  100.0
langevinPistonDecay   50.0
langevinPistonTemp    $temperature


# Output
outputName        $outputname
```

```
restartfreq        500     ;# 500steps = every 1ps
dcdfreq            250
xstFreq            250
outputEnergies     100
outputPressure     100


#################################################################
###
## EXTRA PARAMETERS                           ##
#################################################################
###


#################################################################
###
## EXECUTION SCRIPT                           ##
#################################################################
###

# Minimization
minimize           100
reinitvels         $temperature

run 50000 ;# 100ps
```