TECHNISCHE UNIVERSITÄT MÜNCHEN

Fachgebiet Industrielle Biokatalyse

Validating theoretical approaches for the generation of terpene macrocycles and their functional decoration by targeted protein engineering

Patrick M. Schrepfer

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. rer. nat. Tom Nilges

Prüfer: 1. Univ.-Prof. Dr. rer. nat. Thomas Brück

2. apl. Prof. Dr. rer. nat. Wolfgang Eisenreich

3. Univ.-Prof. Dr. rer. nat. Christian Becker, Universität Wien / Österreich

Die Dissertation wurde am 29.10.2015 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 30.11.2015 angenommen.

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Dissertation selbstständig verfasst, sowie die Ausführungen und Gedanken, welche anderen Schriften sinngemäß oder wörtlich entnommen wurden, sowie weitere Quellen und Hilfsmittel kenntlich gemacht habe. Die vorliegende Arbeit wurde bisher weder in gleicher noch ähnlicher Form einer anderen Prüfungsbehörde vorgelegt oder anderweitig veröffentlicht.

München, den ……..

Acknowledgements

First of all, I would like to thank Prof. Dr. Thomas Brück who gave me the opportunity to carry out this work in the group of Industrial Biocatalysis, for the freedom he offered in terms of following up own scientific ideas and approaches, for his helpful scientific advices and for critical reading on this thesis.

Also I would like to thank Prof. Dr. Wolfgang Eisenreich and Prof. Dr. Christian Becker for accepting co-examination.

I am especially indebted to Christian Görner, Tom Schuffenhauer and Veronika Redai for their help and done work that enabled me to achieve the goals of this thesis.

A big thank you goes to my master and bachelor students, Alexander Büttner, Snighda Majumder, Michael Metterlein, Dora Balogh and Sarah Mannott without their great help and support the goals of the thesis could never have been achieved.

Furthermore I would like to thank the other members of the Brück lab for creating that special working atmosphere.

Another thank you goes to Prof. Dr. Wolfgang Eisenreich, Dr. Martin Haslbeck, Prof. Dr. Ville Kaila und Dr. Ilke Ugur for their help with experimental and computational methodologies and the scientific advices throughout the course of the work.

Many thanks go to all of my friends for the chaperonage throughout the four years of the journey.

Special thanks goes to my family for supporting me throughout the last years and making my studies possible.

Abstract

Sustainable development is an approach that aims to satisfy human needs in an economically viable, environmentally friendly and socially beneficial way. The need of sustainable approaches has priority at a time, where fossil resources are on decline and environment preserving methodologies become more and more important. The concept of integrated biorefineries makes use of sustainable biomass for the production of multiple high-value chemical and pharmaceutical compounds as well as biofuels, thereby minimizing costs and environmental damage. In this respect, the usage of engineered microbial production hosts in combination with optimized pathway enzymes that are able to produce high-value chemical entities from renewable biomass is crucial. Nature has evolved metabolic pathways and high specific enzymes as biocatalysts to produce complex chemical building blocks for the generation of pharmaceutical useful bioactives. Synthetic chemical strategies to copy these natural products are often difficult due to the high complexity of the compounds. Moreover, synthetic chemical routes often suffer from low efficiencies and are characterized by the generation of unwanted by-products.

The sustainable generation of natural, pharmaceutical useful bioactives, *e.g.* terpenoids, alkaloids, flavonoids and polyketides by engineered microbes, is however often limited due to inefficient production hosts, non-specific or non-native biocatalysts and non-optimized metabolic pathways. Therefore, the optimization of known and the elucidation of novel metabolic pathways and enzyme biocatalysts aiming at the efficient and high titer generation of natural bioactives are currently a major focus in biotechnology.

Terpenoids, the class of microbial secondary metabolites with the highest quantity of members, are characterized by a complex architecture and a high degree of functional decoration rendering a synthetic chemical production route inefficient. Examples for high-value bioactive ter-

penoid compounds are Taxol, a potent anticancer drug and cyclooctatin, a potent anti-inflammatory drug. Other important examples are the class of coral-derived verticillene and casbane terpenoid bioactives as well as the class of cembranoid bioactive compounds.

Efficient generation of terpenoid bioactives is based on the availability of their metabolic production pathways, on the knowledge of the enzyme that produces the macrocyclic terpene core (terpene synthase) as well as the enzymes that are responsible for their functional chemical decoration. However, the respective native enzymes are often unknown. One way to overcome this limitation is the redesign of non-native terpene synthases and decorating enzymes to enable the generation of non-accessible or novel terpene macrocylces and their subsequent functional decoration. A prerequisite for these redesign approaches is the detailed knowledge of essential structure-function relationships and underlying catalytic enzyme mechanisms.

Therefore, one aim of the presented work was the elucidation of essential function-relationships of known terpene synthases in order to access enzymatic routes for the generation of novel or today not accessible terpene macrocycles. Another aim was the identification of suitable interaction partners for decorating enzyme complexes that are responsible for the hydroxylation of the generated terpene macrocycles. Their combined integration into a metabolic engineered production system aiming at the *de novo* production of these bioactive terpenoid compounds was the final aim of this work.

Targeted engineering of taxadiene synthase using computational and experimental methodologies.

Taxadiene synthase (TXS) from *Taxus brevifolia*, a diterpene synthase responsible for the first committed key step in Taxol biosynthesis has been under structural and catalytic investigation for decades. However, the discrete catalytic cyclization pathway from geranylgeranylpyrophosphate (GGPP), the linear acyclic precursor of all diterpenoids to taxa-4,5-11,12-diene, a three-cyclic terpene in the active site of the enzyme is unknown.

A crystal structure of the open, inactive complex of TXS is known, however, this complex is not useful to investigate detailed structure-function relationships of this class I terpene synthase model enzyme. Using a combination of computer-assisted modelling of the active closed complex conformation of TXS together with a force field-based modelling approach to locate the proposed carbocationic intermediates, detailed insights into the structure-function relationships of TXS and hence the enzyme family of class I terpene synthases could be obtained. *In silico* redesign of TXS led to the elucidation of a discrete cyclization pathway, evolutionary conserved distinct stabilization motifs for carbocationic intermediates and a bi-functional active site motif that acts as the final cascade terminating base. A detailed mechanism representing the structural basis of the complex cyclization cascade in TXS could be proposed that significantly contributes to the understanding of the underlying cyclization mechanism of class I terpene synthases in general.

The results were experimentally validated using a site-directed mutagenesis approach, where essential *in silico*-predicted residues were mutated to confirm the predicted findings. Eighty five different single mutations were introduced, the mutated enzymes were purified and their ability to generate wild type and intermediate-derived deprotonated carbocationic terpene macrocycles were analysed using GC-MS, 1D- and 2D NMR and CD-spectroscopy. Two major alternate terpene macrocycles, a monocyclic cembranoid and a bi-cyclic verticillene macrocycle, accompanied by minor by-products could be identified by *in silico*-predicted and *in vitro* verified TXS-mutants. Both macrocycles are not accessible by synthetic chemical approaches. Especially for the verticillene macrocycle, a terpene synthase that produces this bi-cyclic hydrocarbon core is unknown, although verticillenes are valuable precursors for a myriad of bioactive terpenoids. The establishment of a bacterial production route for the efficient generation of both alternate macrocycles using the mutated enzymes together with a metabolically engineered natural production pathway (DXP-pathway) led to a fermentation-based efficient production of both macrocycles *in vivo*.

Identification, characterization and adaption of class I P450 redox cascades for the efficient generation of cyclooctatin and sinularcasbane D.

Cyclooctatin, a tri-cyclic and tri-hydroxylated terpenoid from *Streptomyces melanosprofaciens* shows potent anti-inflammatory activity. The biosynthetic gene cluster responsible for the generation of cyclooctatin and its mono- and di-hydroxylated intermediates are known. However, a biosynthetic production route for cyclooctatin is currently unavailable due to the missing of the native redox partner proteins for the cyclooctatin-specific class I P450 hydroxylases CotB3 and CotB4 that hydroxylate the mono- and the diol, respectively. Using a bioinformatic homology search approach covering all available *Streptomyces* genomes combined with *in silico*-based modelling, suitable class I redox partner proteins for both CotB3 and CotB4 could be identified in the genome of *Streptomyces afghaniensis*. The newly identified class I redox system consisted of a ferredoxin (Afx) and a NADH-dependent reductase (AfR). The eleven enzyme-based biosynthetic cyclooctatin pathway was subsequently introduced into a bacterial host. Subsequently, the production of cyclooctatin and its intermediates was screened by GC-MS using silylated derivatives, LC-MS, 1D and 2D NMR as well as CD spectroscopy. Another aim of this project was the structural elucidation of the electron transfer processes and binding interaction between AfR and Afx as well as Afx and CotB3/4. Therefore, the *in silico*-derived structural models and enzyme complexes of AfR and Afx as well as Afx and CotB3/4 were compared with the structurally well-characterized model class I P450 redox system complexes from *Pseudomonas putida*. A structural comparison of the complexes established two residues in Afx (E39 and E106) as the essential ones for electron transfer and binding to CotB3/4. A site-directed mutagenesis approach combined with peptide mass fingerprinting verified the importance of these residues. Additionally, the adaption of the *P. putida* ferredoxin counterpart (Pdx) based on the essential identified residues from Afx led to a significant higher production rate of the di-hydroxylated intermediate. This finding could play an essential role for the adaption of non-native bacterial class I redox systems to any class I biotechnological hydroxylation process where the native redox partner is unknown or the redox system suffers from low

expression rates. The reconstituted bacterial AfR-Afx-CotB3 class I P450 system enabled moreover the elucidation of an uncharacterized substrate promiscuity of CotB3. The wild type CotB3 is shown to hydroxylate (-)-casbene, a monocyclic terpene, in an efficient manner. The obtained and structurally characterized sinularcasbane D has up today only found in a soft coral (*Sinularia sp.)*. This finding shows that the reconstituted AfR-Afx-CotB3-CotB4 system could be a valuable target for the non-native hydroxylation of alternate terpene macrocycles. Most importantly, the production of cyclooctatin using the newly identified redox system in a metabolically engineered production system was the first report of a consecutive *in vivo* hydroxylation of a terpenoid.

Characterization and heterologous production of the cembranoid pathway enzymes from common tobacco

The main volatile leaf-compounds of tobacco plants are mono- and di-hydroxylated monocyclic terpenoids, termed cembranoids. The di-hydroxylated α- and β-cembratriene-4,6-diol epimers (α- and β-CBT-diol) comprise 60% of the mass of the leaf exudate. Their biosynthetic precursors α- and β-CBT-ol, respectively, comprise in contrast only 1.4%. Both di-hydroxylated compounds display a large variety of bioactive activities, *e.g.* anti-proliferative, antimicrobial, antifungal, neuroprotective and antitumor activity. Both have also been shown to inhibit nicotine sensitization in rats, making them a valuable target for biotechnological production in microbial hosts combined with waste stream utilization as carbon source. The mono-hydroxylated epimers show a potent insecticidal activity, rendering a biosynthetic biotechnological production of both forms of hydroxylated compounds biotechnologically desirable. The terpene synthase responsible for the generation of the mono-hydroxylated cembranoid cores is known, but has never been structurally and catalytically characterized. A bacterial production platform for the production of CBT-ols has also never been described. Similarly, the class II P450 monooxygenase responsible for the second hydroxylation step has been described, but never been

experimentally characterized. Production of CBT-diols has moreover never been shown, as the class II membrane-bound NADPH-dependent reductase, the native redox partner protein of the known monooxygenase, remained elusive. Using a combination of *in silico* modelling of the plant CBT-ol producing terpene synthase (CBTS), site-directed mutagenesis and quantum chemical (QM) gas phase calculations, detailed structural and mechanistic insights into the closed active conformation of CBTS could be obtained. A discrete cyclization cascade and the underlying mechanism for the production of valuable by-products could be elucidated. The establishment of a fermentation-based *in vivo* production system enabled an efficient production of the two CBT-ol epimers and the three additional by-products. The CBTS enzyme produces (+)-cembrene, (±)-isocembrene and (-)-casbene, besides the known CBT-ols. These three diterpenoids that are derived from premature quenching products are valuable hydrocarbon precursors for a myriad of cembranoid-type bioactive terpenoids. Furthermore, essential structure-function relationships acting in class I teperne synthase catalysis, identified upon TXS modelling, could also be confirmed for CBTS. Moreover, a mutant could be identified (D495N) that only produces the non-hydroxylated by-products, highlighting an amino-acid-assisted mechanism for the production of hydroxylated compounds in class I terpene synthases.

The second step within this project was the identification of the unknown tobacco-derived class II reductase that allows for the production of CBT-diols by the known class II monooxygenase. Using a bioinformatic homology-based search in the available transcriptome of *N. tabacum* (common tobacco), fragments of a class II reductase could be obtained. These fragments characterized the FMN-, FAD- and NADPH-binding domains of the sought-after enzyme together with the inter-domain comprising sequences, missing the membrane-bound domain. Subsequent membrane engineering of the class II P450 monooxygenase and the combined bi-cistronic integration of the monooxygenase-reductase construct into the CBT-ol production strain enabled the generation of CBT-diols. This strain additionally produces simultaneously the CBT-ols and the three cembranoid-type terpene compounds out of one single metabolic pathway. However, the various bioactivities comprising CBT-diols could only be produced in

low amounts, arguing for a more detailed adaption of the membrane domain of the monooxygenase and the optimization of the bi-cistronic monooxygenase-reductase construct at the molecular level. The production of seven different bioactive compounds out of one metabolic pathway in one microbial host is however a striking result and opens up the way to an economic, efficient and easy-to-handle cembranoid production system.

List of related articles

The following submitted or articles in preparation will contain extracts of this thesis:

I. Schrepfer P., Büttner A., Görner C., Hertel M., Wallrapp F., Eisenreich W., Sieber V., Kourist R. and Brück T. **Carbocations caught in the act: Deciphering details of taxadiene synthase mechanism by *in silico*-guided mutagenesis of its active closed conformation,** PNAS, submitted, review process

II. Görner C., Schrepfer P., Redai V., Wallrapp F., Loll B., Eisenreich W., Haslbeck M. and Brück T. **Identification, characterization and *in silico*-guided adaptation of class I redox systems for the production of functionalized diterpenoids,** ACS Catalysis, submitted, review process

III. Schrepfer P., Ugur I., Metterlein M., Balogh D., Kaila V. and Brück T. **Computer-assisted characterization of the cembranoid pathway enzymes from common tobacco and heterologous production of its functionalized diterpenoid bioactives,** in preparation

Content

Chapter I: General Introduction

1. Biotechnology – a way to sustainable production routes of high value chemical entities

Climate change, a growing world population and limited fossil resources drive manufacturing industries to develop sustainable, biomass-based processes for the production of energy, fuels, chemical building blocks and pharmaceuticals. The challenge for a sustainable economy is to produce material products for society that are based on green and sustainable supply chains.[1,2] Renewable biomass is the only sustainable alternative to fossil fuels as a source of carbon for the chemical and material needs of the economy. The key to enable the utilization of biomass as feedstock for the chemical industry is the so-called biorefinery concept.[3,4] This concept characterizes the integration of biomass conversion as well as the necessary equipment for the production of fuels, power, heat and most importantly value-added chemicals and pharmaceutical precursors from low-value biomass in analogy to the concept of today`s petroleum refinery with its well-known disadvantages. Thereby, a biorefinery is able to simultaneously utilize the various diversified components in biomass and their intermediates for the production of high-value chemicals and low-value, but high-volume biofuels.[5-7] At the same time a parallelized generation of electricity and process heat for own usage and for sale can be accomplished that helps to reduce the energy costs and greenhouse gas emissions (Figure 1). This big advantage of the biorefinery concept maximizes the value obtained from the biomass feedstock. The biorefinery concept profits from recent advances in genetics, biotechnology, process chemistry and engineering as well as bioinformatic approaches.[2,8]

Today, biorefineries, their feedstocks and the produced compounds are classified as first, second or third generation.[3] First generation feedstocks are food crops that compete with food production and land use. Therefore, momentarily the focus is shifting to second and third gen-

eration feedstocks. Second generation feedstocks are defined as mainly lignocellulosic materials as well as biomass from various waste streams (agricultural, municipal and forestry waste). Third generation feedstocks are characterized by the use of microalgae biomass and its biocatalytic conversion to high-value products or *en route* intermediates by various microbial organisms.

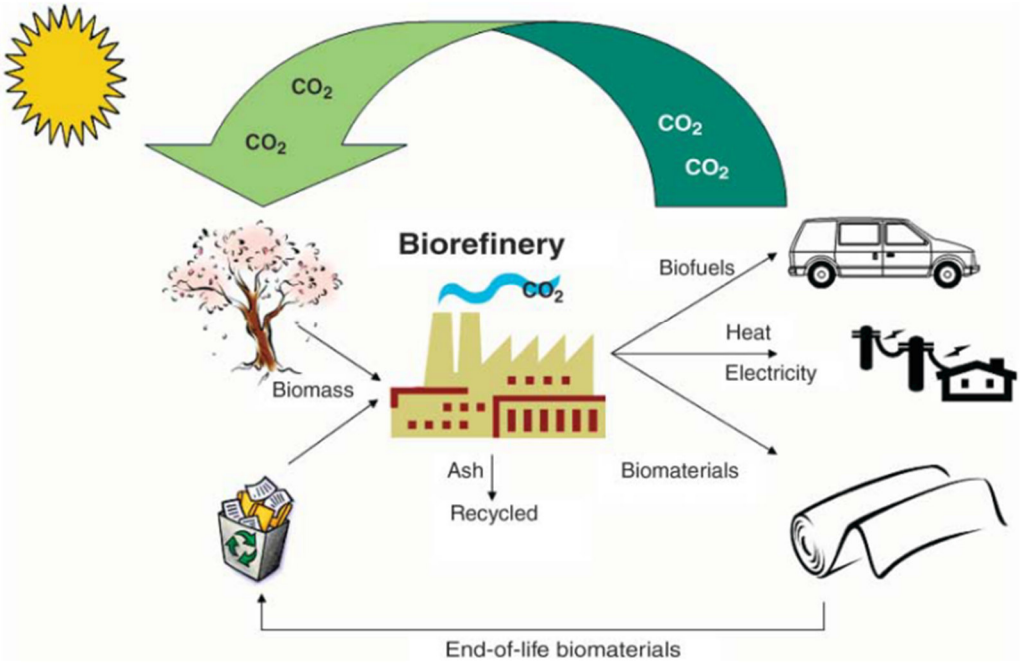


Figure 1: The concept of integrated biorefineries

A fully integrated agro-biofuel-biomaterial-biopower cycle for sustainable technologies. Figure reprinted from (Ragauskas et al., 2006).[2]

One way to process biomass feedstocks into high value useful products is the use of microorganisms and enzymes in biotechnological processes, mostly fermentation processes.[8,9] In these processes, engineered microorganisms harboring optimized biosynthetic metabolic pathways in combination with engineered enzymes are used to *de novo* generate high value products or intermediates thereof from utilizable biomass. An alternative way to use biomass is the so-called thermochemical route that relies on heating the feedstock to high temperatures with little or no oxygen at all.[9] A combined biochemical and thermochemical route is a conceivably alternative in future approaches.

2. Metabolic and enzyme engineering - keys for the production of high-value chemical entities

Genetically engineered microbes in combination with optimized native and/or designed non-native enzymes have been increasingly used to produce high-value chemicals in numerous applications (*e.g.* food, agriculture, chemical and pharmaceutical industries).[8] Important, recent and very successful examples for the use of engineered microbial systems in combination with engineered enzymes are the production of insulin[10], therapeutic antibodies[11] and Fab fragments[12] or the high level production of amorpha-4,11-diene, the important precursor of the anti-malarial sesquiterpenoid from *Artemisia annua*[13], all in the bacterial host *Escherichia coli*. This bacterium combines a fast growth rate, a high genetic accessibility and is able to grow on a variety of culture media, including second and third generation feedstocks with little engineering efforts. Production of high-value chemical entities using this bacterial host does not compete with land use and is not dependent on the use of first generation feedstocks.[8]

In contrast to synthetic chemistry, the use of engineered enzymes, the usage of metabolically engineered microbial systems and a combination of both has significant advantages.[8,14] These advantages are the avoidance of heavy metals, organic solvents, strong acids and bases. Moreover, enzymes display a relatively high substrate specificity resulting in the reduction of non-wanted by-products. A third advantage is that nature has evolutionary developed synthetic pathways for the production of complex structures (*e.g.* alkaloids, terpenoids, flavonoids, vitamins but also amino acids, organic acids and alcohols) (Figure 4), whereas establishing of a chemical *de novo* synthetic pathway is often difficult. Usage of these pathways in heterologous or autologous hosts in combination with metabolic and combinatorial protein engineering allows the creation of improved production titers and the creation of novel bioactive compounds as ultimate objective.[15,16] The elucidation and optimization of established and/or novel metabolic production routes is therefore one of the major objectives in modern biotechnology. The genetic accessibility and viability of *E. coli* and other microbial hosts play an important role in

this respect. The last decades brought up differing methodologies for metabolic engineering of microbial hosts.[17] The plasmid-based approach, where whole metabolic pathways from heterologous hosts were integrated into one or several different plasmids and introduced in a bacterial host. On the other hand the genomic integration approach, where whole heterologous pathways were integrated into the genome of a bacterial host. Recent reports using genomic integration methods showed its superior power compared to the plasmid-based approach.[13,18-20] Integration into the genome bears several important advantages. Plasmids are used for the overexpression of proteins for structural or analytical reasons. However, in metabolic engineering the controlled expression rate of each enzyme of the heterologous or autologous pathway and with it the defined interplay of the pathway enzymes is crucial.[21,22] One copy integration of enzymes into the genome combined with a modular selective steering of their expression rates is observed to be the most promising alternative for the fine-tuning and balancing of the integrated pathway.[17,18,23] Genomic integrated pathways do not rely on antibiotics affecting the genomic stability of the introduced genes in contrast to the plasmid-based approach. Striking examples in *E.coli* are the production of amorpha-4,11-diene and taxa-4,5-11,12-diene (Figure 2), the diterpenoid hydrocarbon precursor of the anti-cancer drug Taxol, where up to eleven pathway enzymes have been genomically integrated.[13,18] In both cases production rates of up to several g/L with simple sugars as sole carbon source have been achieved by genomic integration of the respective pathways into the genome of *E.coli* combined with *in silico* and genomic fine tuning of the metabolic interplay of the respective pathway enzymes (Figure 2).

Besides metabolic engineering and process optimization, the elucidation of structure-function relationships of the respective enzymes and the optimization of the interplay between pathway enzymes are the key-steps to enable an efficient and high titer biotechnological production within an integrated biorefinery concept.[22] The ability to selectively design enzymes leading to optimized stability, higher ligand affinity, higher protein-protein interaction capability or catalytic product and substrate promiscuity based on the detailed understanding of structure-function

relationships have been shown to significantly increase biotechnological production titers.[15,16,24]

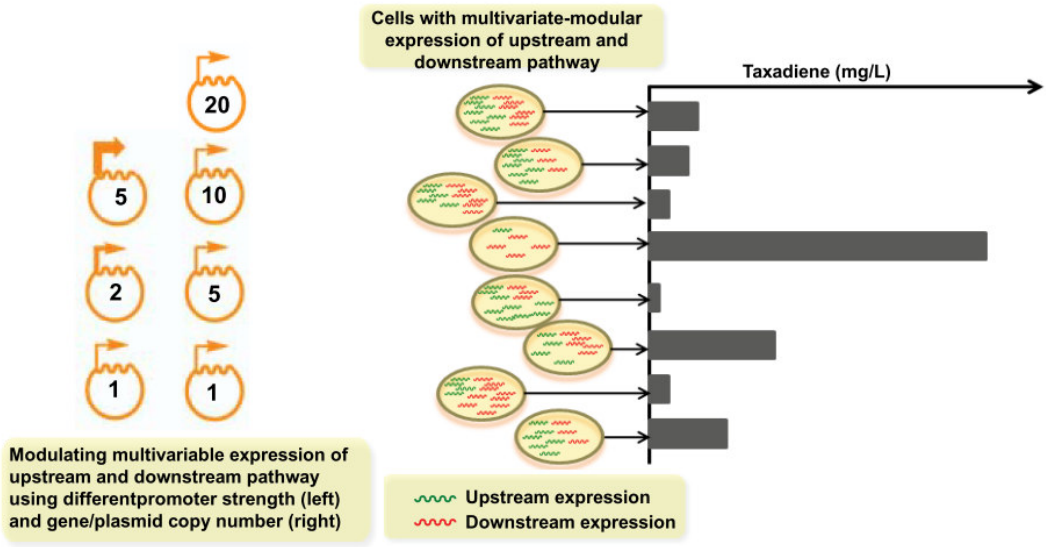


Figure 2: Multivariate-modular approach for isoprenoid pathway optimization

Figure reprinted from (Ajikumar et al., 2010)[18].

There are two distinct approaches to enzyme engineering, directed evolution and rational protein design.[16,24,25] Directed evolution techniques rely on the evolvability of the respective enzyme and create randomly distributed mutations over the whole gene of interest (Figure 3). This methodology does not require the three-dimensional structure of the respective enzyme. However, directed evolution approaches make intense and efficient high-throughput screening efforts necessary during the functional screening step. Sequential directed evolution methods (*e.g.* error-prone PCR) vary single positions within the primary structure, whereas recombinative methods (*e.g.* gene or protein-domain-shuffling) are based on random combinations of larger gene segments or whole protein domains.

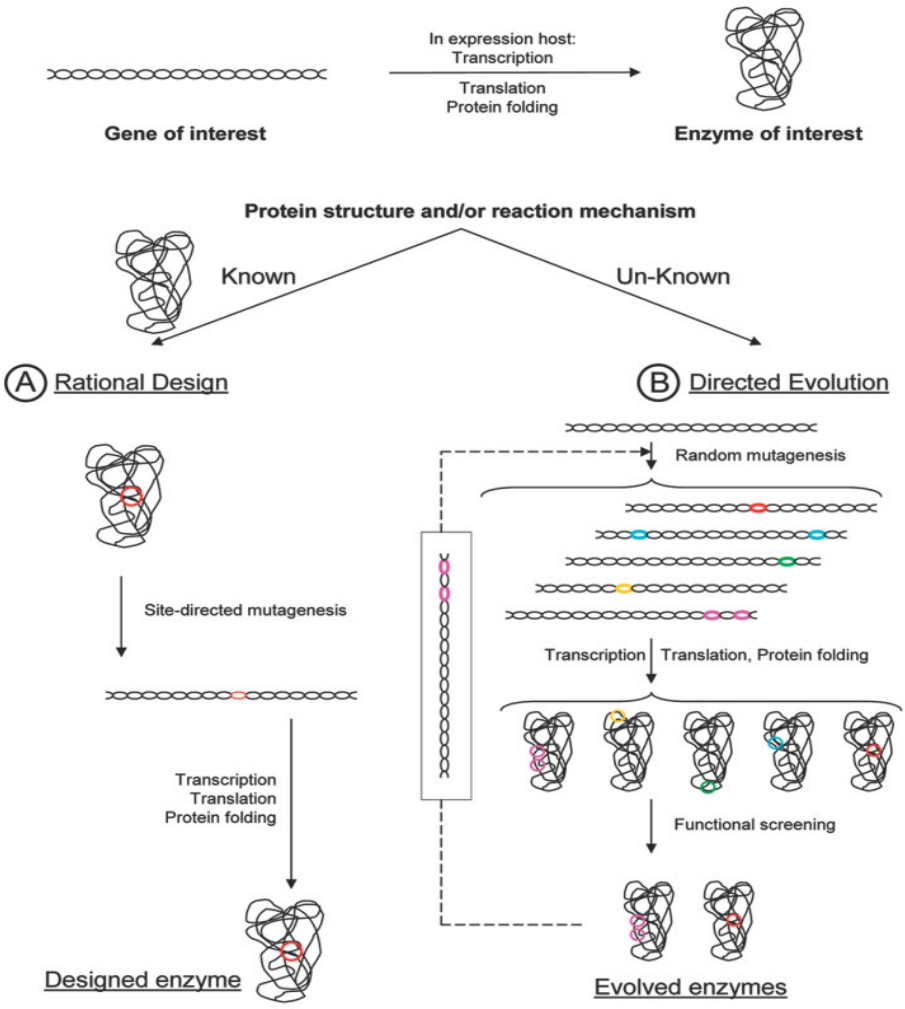


Figure 3: Approaches to designed enzymes

Figure reprinted from (Strohmeier et al., 2011).[25]

In contrast to directed evolution, rational protein design are based on the knowledge of structural, functional and mechanistic information about the enzyme of interest (Figure 3). This methodology mostly rely on computational-based structure analysis (*e.g.* crystal- or NMR-derived protein structures, as well as *ab initio*-based or homology models) and commonly aims at the rational/logical redesign of one or a few residues (so-called “hot-spot”-residues) to alter the substrate/product promiscuity and/or affinity of the respective enzyme. Despite the advances in the *in silico* redesign of enzymes, a subsequent experimental screening for mutants with the desired properties and good expression behaviour is still needed.

The parallel utilization of second and/or third generation feedstocks as carbon source by metabolically and enzymatically engineered microbes to generate high-value known and novel chemical and pharmaceutical entities is therefore the today`s desired goal in red, white, blue, green and brown biotechnology (Figure 4).

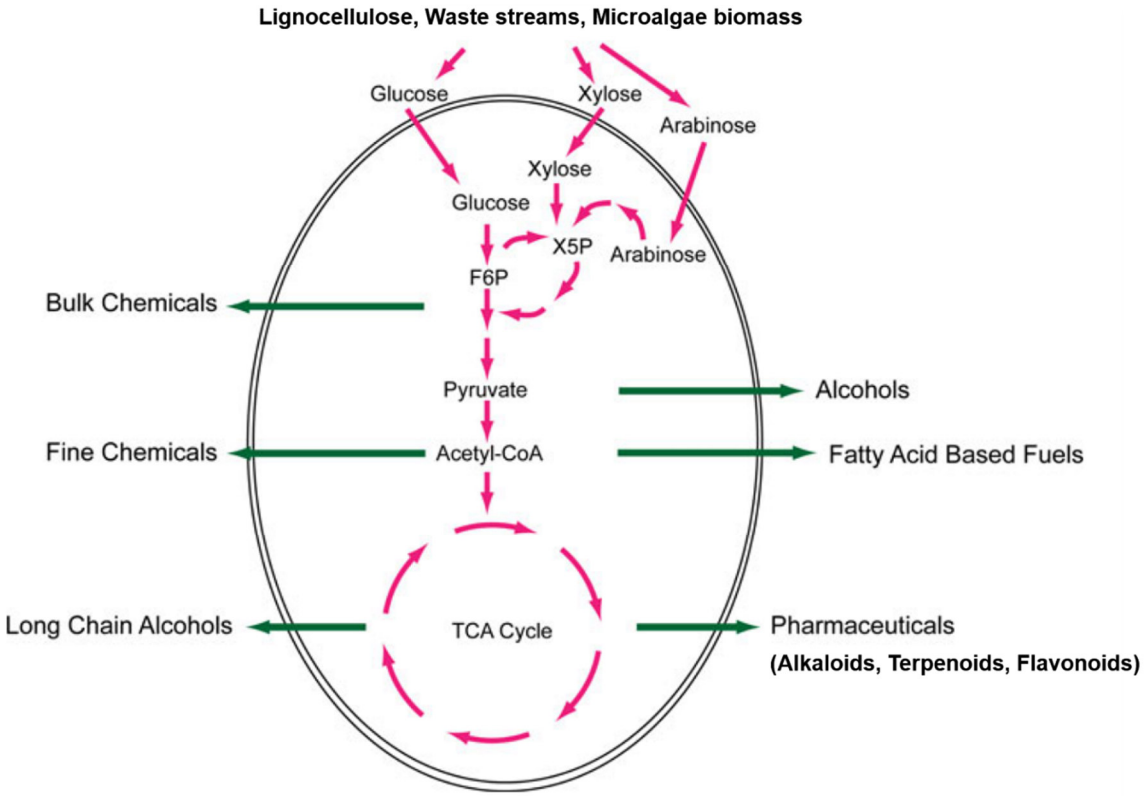


Figure 4: Overview of metabolic flux in engineered microbes

Production of biofuels, chemicals and pharmaceuticals from second and third generation feedstocks using wild type or metabolically and enzymatically engineered microbes. Figure reprinted and adapted from (Du et al., 2011).[8]

3. Terpenoids - a class of valuable natural secondary metabolites

Secondary metabolites from bacteria, fungi and plants that are not associated with growth are referred to as the class of natural products. The class of natural products comprises polyketides, non-ribosomal peptides, terpenoids and alkaloids with > 100.000 compounds. In this class, terpenes and their chemically decorated derivatives, termed terpenoids represent the biggest and structurally most diverse group with 350 family members and > 60.0000 compounds known today.[17,26,27] The hydrocarbon basis of all terpenoids turns out to be a simple universal isoprene unit (2-methylbuta-1, 3-diene). Nevertheless, terpenoids are characterized by a very complex carbon- and stereochemistry that is mostly based on mono- or polycyclic scaffolds, whereas acyclic structures are also possible.[26,28] These scaffolds are classified according to their number of isoprene subunits. Monoterpenes consists of two, sesquiterpenes of three, diterpenes of four and triterpenes of six basic isoprene units. Of the known terpenoids the diterpene group forms the biggest class with > 12.000 known compounds.[29] Terpenes and terpenoids possess an important biotechnological and biopharmaceutical value. They are used as fragrances, flavors, essential oils and cosmetics. Current pharmaceutical blockbuster drugs, *e.g.* paclitaxel® (taxol, an anticancer agent)[30] and artesiminin (antimalarial agent)[31] are naturally occurring terpenoids. A myriad of other terpenoids exhibiting anti-oxidant (i.e. asthaxantin)[32], anti-inflammatory (i.e. cyclooctatin)[33] or insecticidal and cytotoxic (i.e. cembrantriene-ol and -diol)[34] activities have been reported. The biological activities of these compounds arise from their complex molecular architecture and their highly diverse chemical functionalization. Terpenes and terpenoids are typically produced in low amounts in their respective host organisms, providing a limited source of these valuable compounds. Genetic and metabolic engineering strategies using bacterial or fungal microorganisms, however, paved the way to a sustainable production, where the terpene producing and decorating enzymes are known.[22,28,35,36]

4. Upstream and downstream pathways in terpene formation

The generation of the basic hydrocarbon macrocycles common to all terpenes and terpenoids can be classified into upstream and downstream pathways.[18,37] Upstream pathways produce the common precursors DMAPP (dimethylallyl diphosphate) and IPP (isopentenyl diphosphate) (Figure 5). The downstream module synthesizes several different polyprenyldiphosphates that differ in their carbon count by the aid of prenylelongases (Figure 6).[38] These enzymes link DMAPP and IPP to the ten carbon polyprenylphospate geranylpyrophosphate (GPP), one DMAPP and two IPP molecules to farnesylpyrophosphate (FPP) and by adding one molecule IPP to FPP the twenty carbon containing geranylgeranylpyrophosphate (GGPP). Prenylelongases catalyze “head-to-tail” linkage reactions similar to class I terpene synthases, *vide infra*.[38] Two differing upstream pathways that produce the universal isoprenoid precursors are described.[39] The mevalonate pathway (MEV) that is native to most eukaryotes, archaea and some eubacteria. The non-mevalonate pathway, termed MEP or DXP (2C-methyl-d-erythritol-4-phosphate) pathway that is housed by bacteria, plant plastids and some eukaryotic parasites (Figure 5). The eight enzyme containing MEP pathway produces DMAPP and IPP via several intermediate products from pyruvate and glycerinaldehyde-3-phosphate.

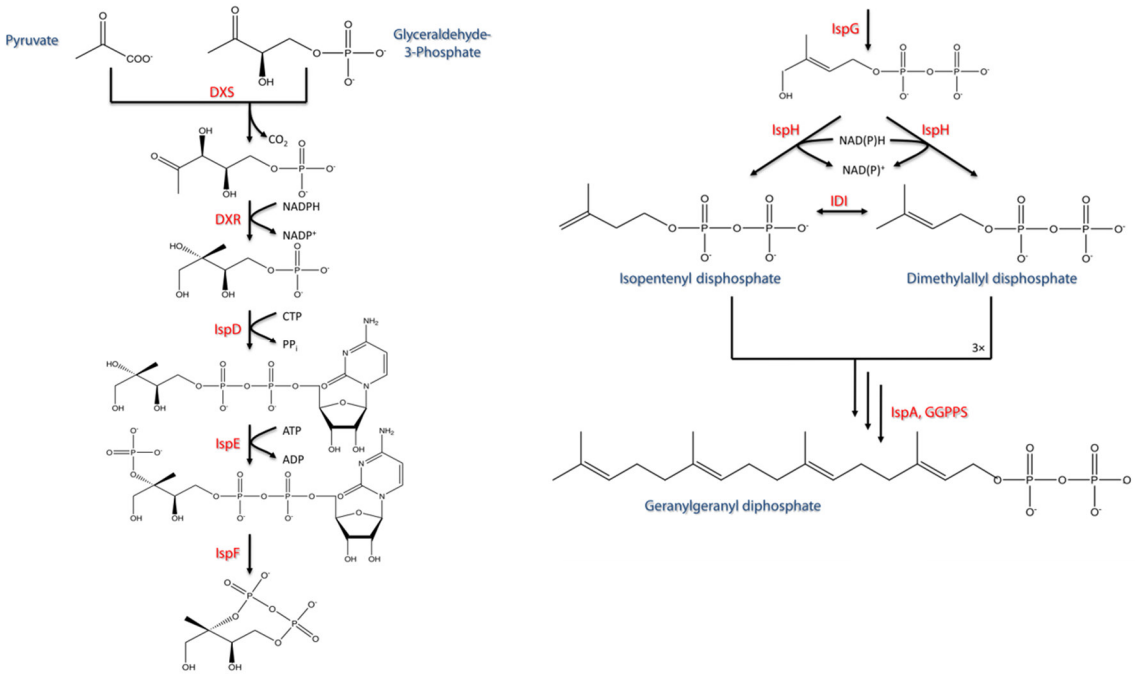


Figure 5: Upstream module of the MEP pathway

See text and list of abbreviations (Appendix) for details. Figure reprinted from (Boghigian et al., 2012).[40]

Diversity in terpenes and terpenoids arise from variable polyisoprene chain elongation (downstream module), subsequent cyclization of the aliphatic, linear hydrocarbon precursors by class I or class II terpene synthases and decorative tailoring reactions of the cyclic terpene products.[41] These decorations predominantly comprise hydroxylation and oxidation by P450 cytochrome c oxidoreductases as well as methylation, esterification, epoxidation and glycosylation by a battery of known and unknown enzymes belonging to the secondary metabolite anabolism.[22]

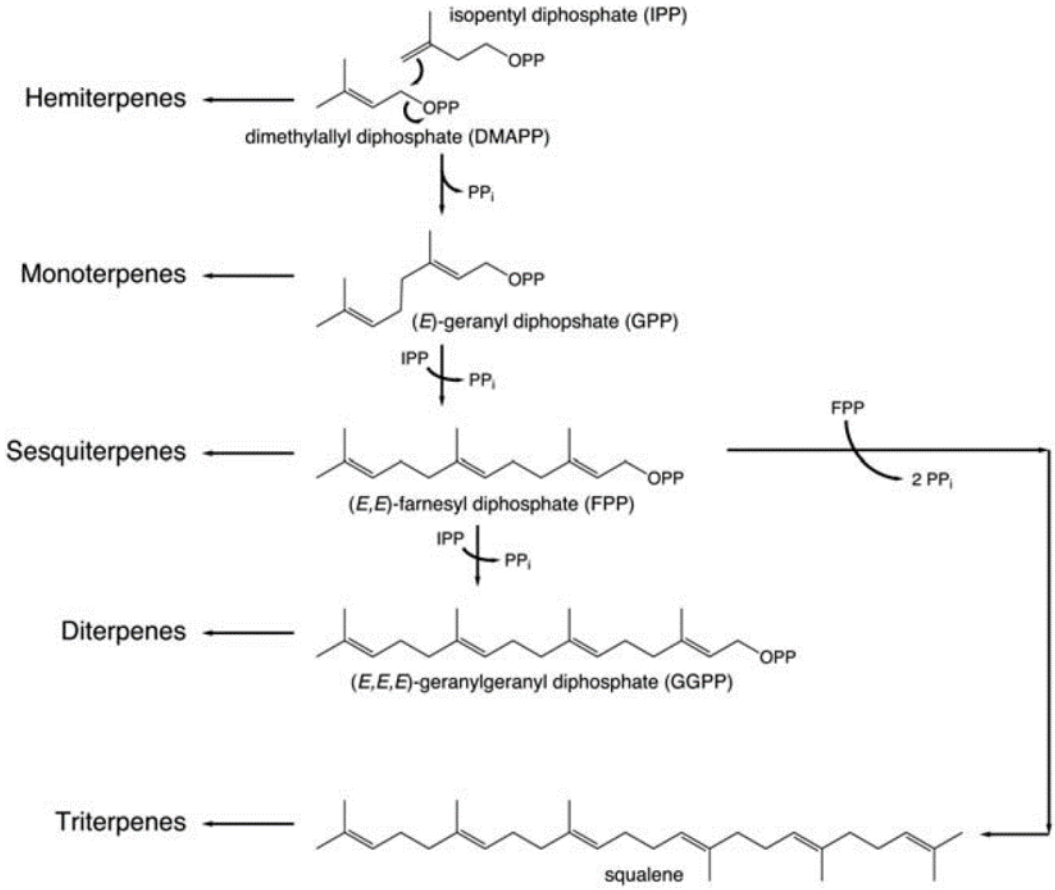


Figure 6: The downstream module of the terpene pathway

See text and list of abbreviations for details. Figure reprinted from (Gao et al., 2012).[42]

5. Terpene synthases

Terpene synthases (TPS) catalyze the cyclization of the acyclic allylic diphosphate precursors that lead to a variety of terpene and terpenoid natural products.[42-44] The cyclization reaction is often characterized by highly regio- and stereoselective formation of multiple chiral centers accompanied by hydride shifts and rare carbon rearrangements. Their products are either a single compound in high fidelity TPS[44,45] or characterized by a diverse product profile in promiscuous TPS.[44,46] The complexity of the cyclization reactions appears to be the main reason for the structural and stereochemical diversity of terpenes. Therefore, terpene synthases rep-

resent the first committed key step in terpenoid biosynthesis.[29,44] These enzymes are found in pro- and eukaryotes. They are structurally characterized by a modular architecture and are mechanistically classified by the initiation of their cyclization reaction. TPS contain one module or a combination of up to six different active or inactive building blocks.[42,43] These building blocks are termed α, β, γ, δ, ε and ζ, whereas α, β and γ comprise α-helical domains (Figure 7). The architecture of the class I terpene synthase superfamily that initiate cyclization by ionization of its respective polyprenyldiphosphate is structurally characterized by a conserved α-helical bundle, the class I associated domain (Figure 7). The initiation of the cyclization reaction of class II terpene synthases depends, in contrast to class I, upon a protonation-dependent mechanism. The class II domain contains a conserved protein fold of double α-barrel domains, the class II associated β- and γ-domains (Figure 7).

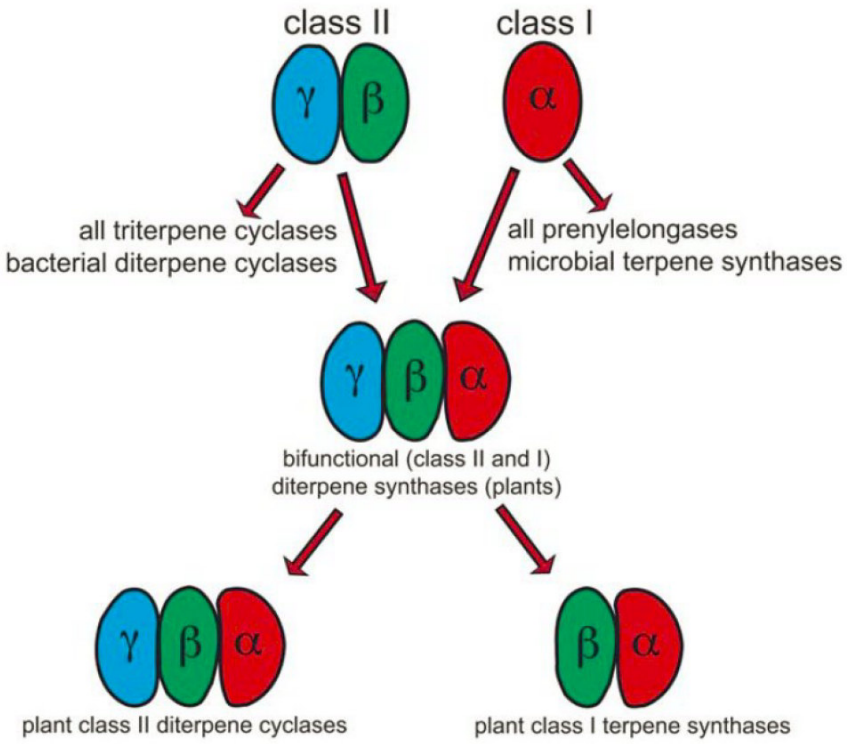


Figure 7: Evolutionary conserved modular architecture of class I and class II terpene synthases

For details, see text. Figure reprinted from (Gao et al., 2012).[42]

6. General catalytic mechanism of class I terpene synthases

Ionization dependent class I terpene synthases (mono-, sesqui- and diterpene synthases) contain a conserved asparagine-rich motif (**D**DXX**D**) and a second typical conserved amino acid signature sequence (**N**,**D**)D(L,I,V)X(**S**,**T**)XXX**E**, the "NSE/DTE" motif (bold residues act in chelating catalytically metal ions) (Figure 8a).[42,43] Typically, these motifs bind three Mg^{2+} ions in class I terpene synthases (Figure 8a).[42,43] The metal ions act in coordination of the respective polyprenyldiphosphate (PPD) upon substrate binding. Additional water molecules in the active site complete the octahedral coordination of the Mg^{2+} ions.[44,47-49] The pyrophosphate moiety (PP_i) of PPD is further thought to be coordinated by a hydrogen bond donor network of active site decorating amino acids, either directly or indirectly via water molecules bound by these amino acids (Figure 8a).[47] Today there are only few examples of class I terpene synthases that have been crystallized.[45,47-54] From these structures, a definite conserved or evolutionary distinct and common mechanism of PPD binding cannot be derived. PPD binding triggers active site closure that brings the active site from an open to a closed state, thereby shielding the initiation of catalysis and the subsequent cyclization steps from premature termination by bulk water solvent. Another important consequence of PPD binding and active site closure is the induction of a conformational change that initiates ionization of PPD that results in a highly reactive unstable allylic carbocation and the separated pyrophosphate moiety (Figure 8b).[53] This conformational change is mediated by a triad consisting of a PP_i-sensor (an arginine residue), a polar linker amino acid and an effector (various amino acids) that provoke abstraction of PP_i by its carbonyl atom directly pointing to the first carbon-carbon double bond of the respective PPD (Figure 8b). Ionization of PPD is the first and rate-limiting step of the carbocation cascade that is common to all class I terpene synthases.[26,55] A concerted addition of the generated primary alkylcarbocation to another π-bond within the same substrate molecule, thereby forming a carbocation at a different carbon center, is generally the second step of the carbocationic cyclization cascade. This step, similar in all class I terpene synthases, leads to a monocyclic carbocation. Formation of this monocyclic cationic structure is thought to be imposed

by the steric decoration of the active site and stabilized by electrostatic interaction with active site amino acids.[42-44] Depending on the terpene synthase this initial monocyclic cation undergoes further nucleophilic intramolecular C=C bond attacks in a sequential and iterative manner often accompanied by key intramolecular proton, hydride and/or methyl shifts, forming multicyclic, often promiscuous structures. For this reason, the orientation of the substrate in the active site is critical with respect to product outcome. Termination of the cyclization cascade generally occurs through quenching of the terminal carbocation either by direct deprotonation (leading to a C=C bond) or via neutralization by a nucleophile.[55,56] Water capture results in a hydroxylated substrate, heterocyclization prior to deprotonation and water capture results in a cyclic ether. The inorganic PP_i moiety can further be used to quench the final cation forming a diphosphate ester.[56,57] As structural elucidation of the closed conformations harboring the native substrate by X-ray crystallography is hampered by the highly dynamic processes intrinsic to terpene synthases, detailed information of structure-function relationships of class I terpene synthases are scarce. Mutagenesis studies and QM/MM simulations of closed complexes, however, have been able to shed a light on the complex cyclization cascades, amino acids that may be involved in catalysis and definitive candidates for the active site base.[53,56-59]

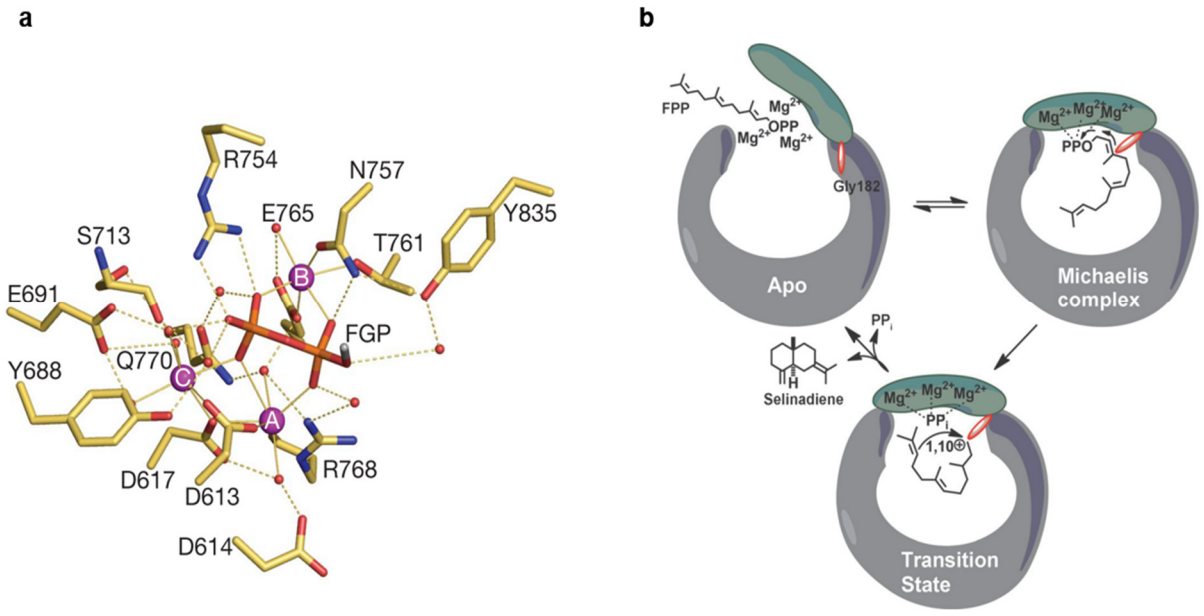


Figure 8: Inner view of taxadiene synthase and induced-fit mechanism upon active site closure

Figure 8**a**: Inner view of the crystal structure of taxadiene synthase highlighting the three catalytic Mg^{2+} ions (A-C) coordinated by water ions (orange), the two catalytic motifs (D613, D614, D617 and N757, T761, E765), further amino acids that coordinate the bulk water solvent and the PP_i group of FGPP (a fluoro-analogue of GGPP). Figure reprinted from (Köksal et al., 2011)[47]. **b**: The induced-fit mechanism of selinadienesynthase upon FPP binding and active site closure. Reprinted from (Baer et al., 2014)[53].

7. Terpenoid-specific macrocycle decoration - cytochrome P450 systems

The decorative functionalization of the hydrocarbon synthons is achieved by an array of additional enzymes that are participating in the secondary metabolite maturation.[17] The well characterized class of cytochrome P450 monooxygenases (CYP450) catalyze the oxidation of non-activated carbon atoms. More than 20.000 genes coding for CYP450s have been found in all domains of life.[60,61] Under mild conditions CYP450s catalyze a variety of chemical reactions, *e.g.* hydroxylation, N-oxidation, oxidative deamination or oxidative dehalogenation. Common to all typical CYP450s is a prosthetic heme B group as cofactor that is attached to the sur-

rounding protein by a single coordinative bond. Specifically, these class of enzymes bind molecular oxygen and catalyze its reduction thereby integrating one O-atom into the respective substrate, the other O-atom is released as water (Figure 9). This reaction requires two electrons supplied by external electron donors, mostly NAD(P)H.[62]

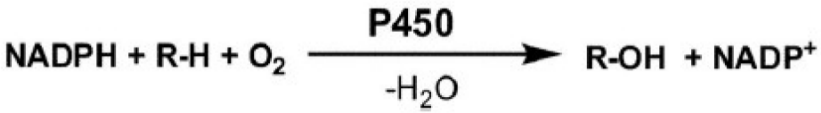


Figure 9: General reaction equation of CYP450 monooxygenases

As a direct electron transfer from NAD(P)H to the CYP450 is uncommon, one or two other proteins are necessary to facilitate this transfer. CYP450s in conjunction with its cognate electron donor proteins are termed cytochrome P450 systems. These systems are classified as depicted in Figure 10.[60-62] Class I CYP450 systems contain the mitochondrial and bacterial system. Upon binding of NAD(P)H by a FAD-containing ferredoxinreductase (FdR) the NAD(P)H electrons are transferred to a soluble iron-sulfur-protein (ferredoxin, Fdx) in both class I systems. Subsequently, the Fdx protein transfers the electrons to the CYP. All proteins of the bacterial class I CYP450 systems exhibit solubility, whereas the CYP and the FdR of the mitochondrial system are membrane-associated or stably integrated. The second class (class II CYP450s) contains the eukaryotic microsomal system. It utilizes a membrane-bound FAD- and FMN-containing cytochrome P450 reductase (CPR) for binding of NAD(P)H and electron transfer to the membrane-bound CYP. There are eight other classes of cytochrome P450 systems, including the well-characterized and widely applicable self-sufficient class VIII CYP450BM3 from *Bacillus megaterium*.[60] These classes are reviewed elsewhere.[60,61]

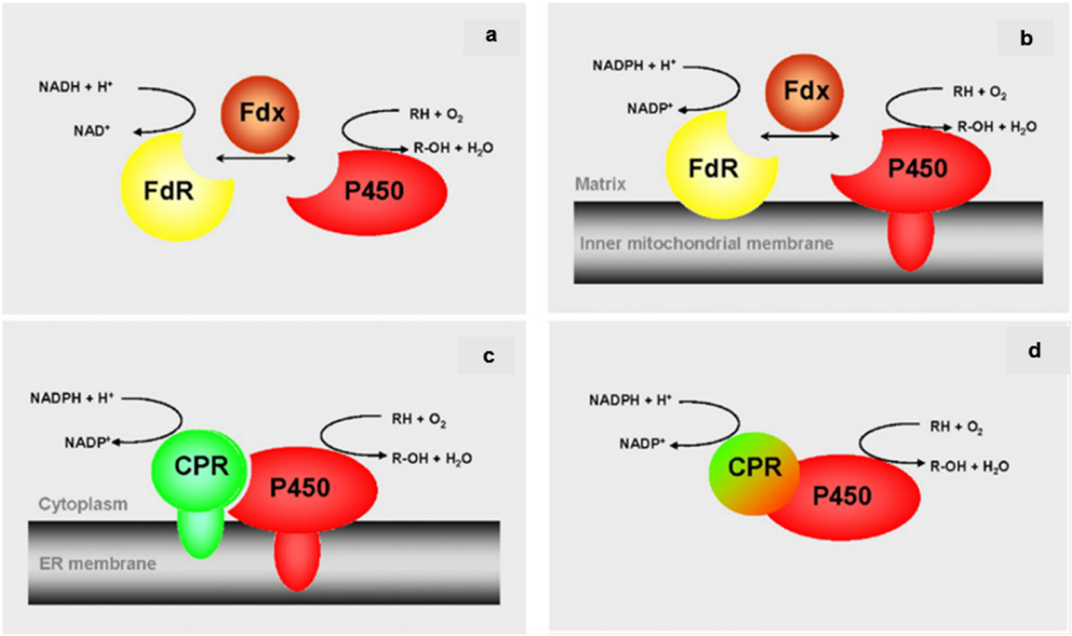


Figure 10: Cytochrome P450 systems

Figure 10**a**: Class I bacterial system, 10**b**: class I mitochondrial system, 10**c**: class II microsomal system and 10**d**: class VIII bacterial CPR-P450 fusion system. For detailed information, see text. Reprinted from (Hannemann et al., 2007).[60]

The catalytic cycle of CYP450 monooxygenases is thought to be characterized by eight different steps.[62] In these steps, the heme B iron cycles between different oxidation states (Fe^{2+}, Fe^{3+} and Fe^{4+}). In the substrate-free state (1, Figure 11) the heme B iron (Fe^{3+}) is hexacoordinated by the four nitrogen atoms of the protoporphyrin ring (equatorial) as well as a distal cysteinate of the enzyme and a proximal water molecule. Binding of the substrate displaces the water molecule and leads to a five-coordinated high spin enzyme-substrate complex (2, Figure 11). The next stage in the cycle is the reduction of the Fe^{3+} ion by an electron transferred from NAD(P)H via an electron transfer chain (class II CPR or class I ferredoxin) (3, Figure 11). After oxygen binding to the Fe^{2+} ion (4, Figure 11), another electron is transferred via the electron transfer chain that results in a peroxo-Fe^{3+}-intermediate substrate-enzyme complex (5, Figure 11). This intermediate complex is subsequently protonated by a proton of the surround-

ing solvent leading to a hydroperoxo-Fe^{3+}-intermediate (6, Figure 11). A second protonation step combined with oxygen cleavage leaves an $(Fe\text{-}O)^{3+}$ complex with the heme B iron in Fe^{4+} oxidation state (7, Figure 11). In a so-called radical "rebound-mechanism"-reaction the Fe-ligated O atom is transferred to the substrate forming the hydroxylated form of the substrate (8, Figure 11). Finally, the product is released and the enzyme returns to the initial state (1, Figure 11). The cycle contains three additional branching reactions that are reviewed elsewhere.[62]

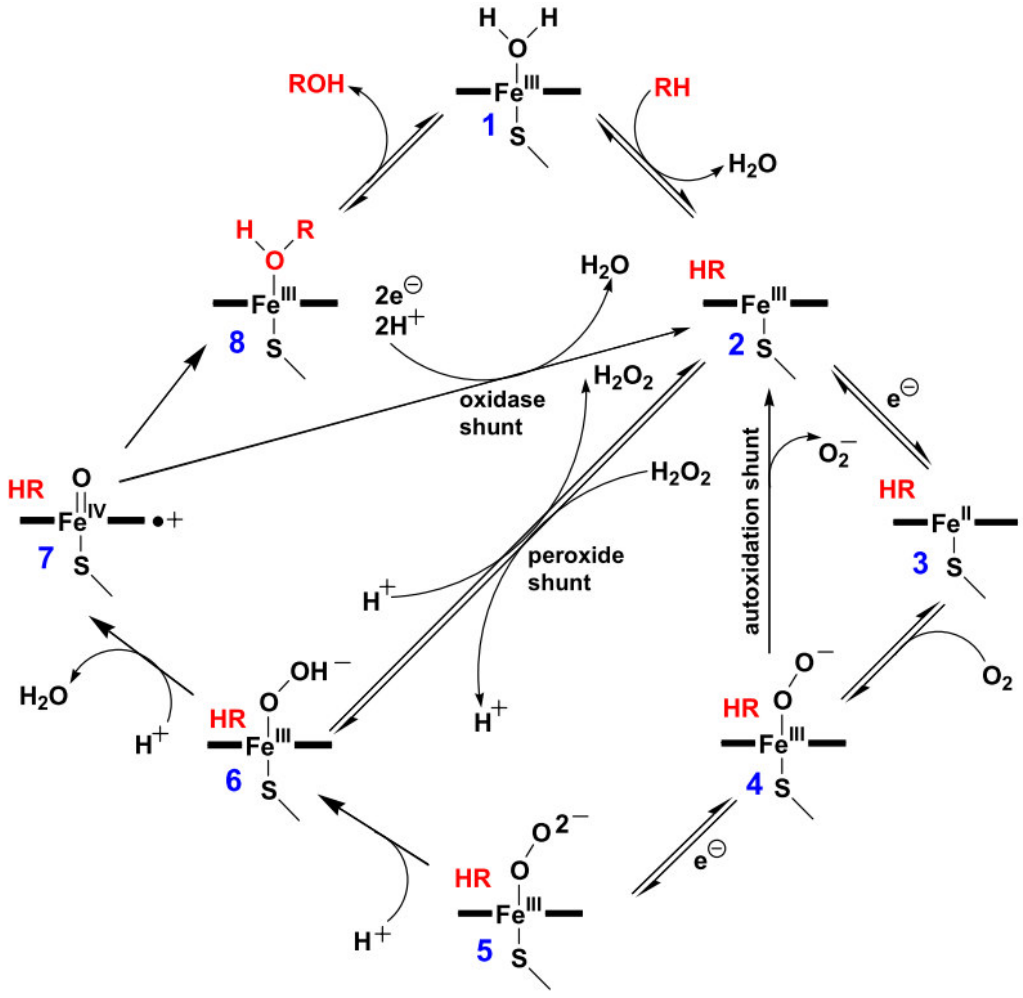


Figure 11: General catalytic cycle of CYP450 monooxygenases

Figure reprinted from (Coon et al., 2005).[62] For detailed information, see text

Chapter II: Carbocations caught in the act: Deciphering details of taxadiene synthase mechanism by *in silico*-guided mutagenesis of its active closed conformation

Patrick Schrepfer [1] , Alexander Büttner [1] , Christian Görner [1] , Michael Hertel [1] , Frank Wallrapp [2] , Wolfgang Eisenreich [3] , Volker Sieber [4] , Robert Kourist [5] & Thomas Brück [1]

[1] Department for Industrial Biocatalysis, Technische Universität München, Lichtenbergstr. 4, 85748 Garching, Germany

[2] Department for Bioinformatics and Computational Biology, Technische Universität München, Boltzmannstraße 3, 85748 Garching, Germany

[3] Chair of Biochemistry, Technische Universität München, Lichtenbergstr. 4, 85748 Garching, Germany

[4] Chair of Chemistry of Biogenic Resources, Technische Universität München, Schulgasse 16, 94315 Straubing, Germany

[5] Junior Research Group for Microbial Biotechnology, Ruhr-Universität Bochum, Universitätsstraße 150, 44780 Bochum, Germany

Author contributions

T.B. and P.S. conceived the study. P.S., A.B., R.K. and T.B. planned and designed the experiments. P.S. evolved the TXS closed complex, subsequent catalytic and variant complexes and docking strategy. P.S., A.B. and M.H. performed mutagenesis, expression and GC-MS-based characterization. P.S., C.G. and W.H. performed NMR, GC-MS and CD-spectroscopic analyses. P.S., A.B., F.W. prepared the manuscript. W.E., R.K., V.S. and T.B. supervised experiments and finalized the manuscript.

1. Introduction

Taxadiene synthase (TXS), a diterpene cyclase from *Taxus brevifolia* catalyzes the cyclization of the twenty-carbon diterpene precursor (E,E,E)- geranylgeranyl diphosphate (**GGPP**) (Figure 12) to taxa-4(5),11(12)-diene, hence named taxadiene as the first commited step in the biosynthesis of Taxol, an established chemotherapy agent for breast and ovarian cancer.[46,63-65] The cyclization of **GGPP** to taxadiene is an exceptional reaction resulting in a three-cyclic macrocycle from the achiral, acyclic substrate comprising formation of three C-C σ bonds, three new stereocenters, and loss of a hydrogen in one reaction cycle (Figure 12).[47,66] The enzyme, a class I terpene synthase uses a trinuclear cluster of the divalent metal ion co-factor Mg^{2+}, coordinated by two conserved binding motifs (**D**DXX**D** and (N,**D**)DXX (**S**,**T**)- XXX**E**), acting in binding the pyrophosphate moiety of **GGPP**, driving active site closure and initiating ionization leading to an enantio- and face-selective carbocation cascade.

Stereochemical labelling experiments and quantum chemical calculations proposed the conformations of the carbocation intermediates **A**-**E** (Figure 12) and the concertedness and synchronicity of the three bond forming events, although the definite catalytic cascade remains elusive (Figure 12).[58,67-71] Mechanistically, the quite exothermic transformation from **GGPP** to **cation C** (Figure 12), accompanied by pyrophosphate loss is thought to be a concerted but asynchronous process and the energy gaining sub-step.[58] Subsequent conformational reorganisation of **cationic states C**-**E** (Figure 12), tentatively existing in equilibrium until site-specific deprotonation occurs, is suggested to be energetically viable by previous experiments and theoretical studies.[58,68] However, elucidation of the mechanism of taxadiene synthase does not include the enzyme`s active site or the pyrophosphate counterion in quantum chemical calculations and insights from labelling experiments have been performed with deuterated or fluorinated **GGPP**-analogues. The enzyme is supposed to largely mediate the complex olefin cation cyclizations including two intramolecular proton transfer steps (**cation C** and **F** (Figure 12)) by conformational control alone without enzymatic intervention in terms of barrierlowering and

without the assistence of an enzyme-bound base, as all class I terpene synthases in general.[44,49] Complementary, recent results assumed also electrostatic effects participating during the carbocation cascade, by pointing out a role for the retained pyrophosphate anion co-product (**PP_i**) as a counter-ion steering carbocation migration towards itself and as the stereospecific base deprotonating the final cation in other (di)terpene synthases.[48,49,55,56,72,73] Literally, substituted polar groups in the active site may lead to stabilization of the different cations leading to premature products, whereas the lack of such stabilization leads to a more complex reaction resulting in final product outcome triggered by the retained pyrophosphate co-product.

A crystal structure of the taxadiene synthase is available, however, showing the open-complex form of the enzyme based on a genomic truncation missing the N-terminus and the so-called J-K loop segment needed for catalysis to initiate.[47] Additionally, the solved structure contains a 2-Fluoro-**GGPP**-analogue as its substrate that is shown not to be in a productive conformation with respect to taxadiene formation.

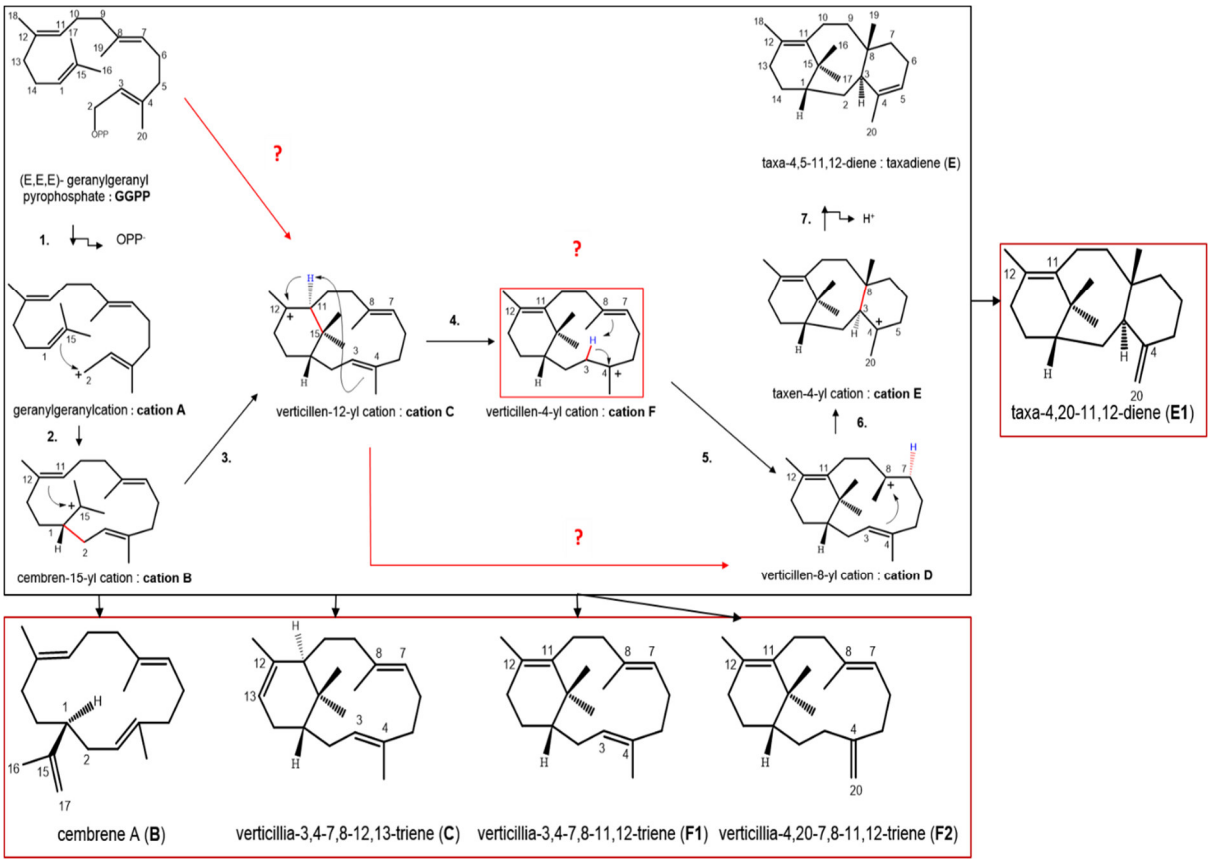


Figure 12: Cascade pathway, carbocation rearrangements and deprotonated intermediates

The black box depicts the postulated carbocationic intermediates during the cascade pathway from **GGPP**->**E** and the numbering of the carbocation cascade steps. The red boxes depict the found deprotonated intermediates, whereby the black arrows indicate the cationic intermediate the deprotonated product date from. The red lines and question marks indicate alternative cyclization steps proposed by QM gas phase calculations.[58] The numbering of cationic and deprotonated intermediates differs from previous reports. [58,69,70] Figure reprinted and adapted from (Hong et al., 2011).[58]

2. Aims of this study

As studying the dynamic processes during class I terpene synthase catalysis is highly challenging by X-ray crystallography or time-resolved NMR spectroscopy we aimed to find an alternative way to get detailed insights into class I terpene synthase mechanisms. We choose TXS as a model class I terpene synthase, as its crystal structure of the open conformation is known and its carbocationic intermediates have been proposed by labelling experiments using substrate analogous and QM calculations. We first modelled the closed active conformation of TXS according to a previous report. Using a molecular mechanical (MM) description of this closed conformation, combined with molecular docking of the known QM gas phase intermediates and an *in silico*-guided mutagenesis strategy, we aimed to decipher the definite catalytic mechanism of TXS as well as structure-function relationships including the long sought active site base.

3. Results

TXS•GGPP complex: Evolutionary conserved motifs initiate the carbocation cascade

Starting point for the modelling was the previously reported X-ray crystal structure of a catalytically not active truncated taxadiene synthase (PDB: 3P5R) (Figure 13, 14a).[47] Active site closure in terpene synthases, after binding of the three Mg^{2+} and the substrate diphosphate group, is thought to involve conformational changes of loops flanking the active site and binding of the N-terminal polypeptide in a groove between these loops important for protecting the reactive carbocation species from premature quenching by bulk solvent.[44,47-49]

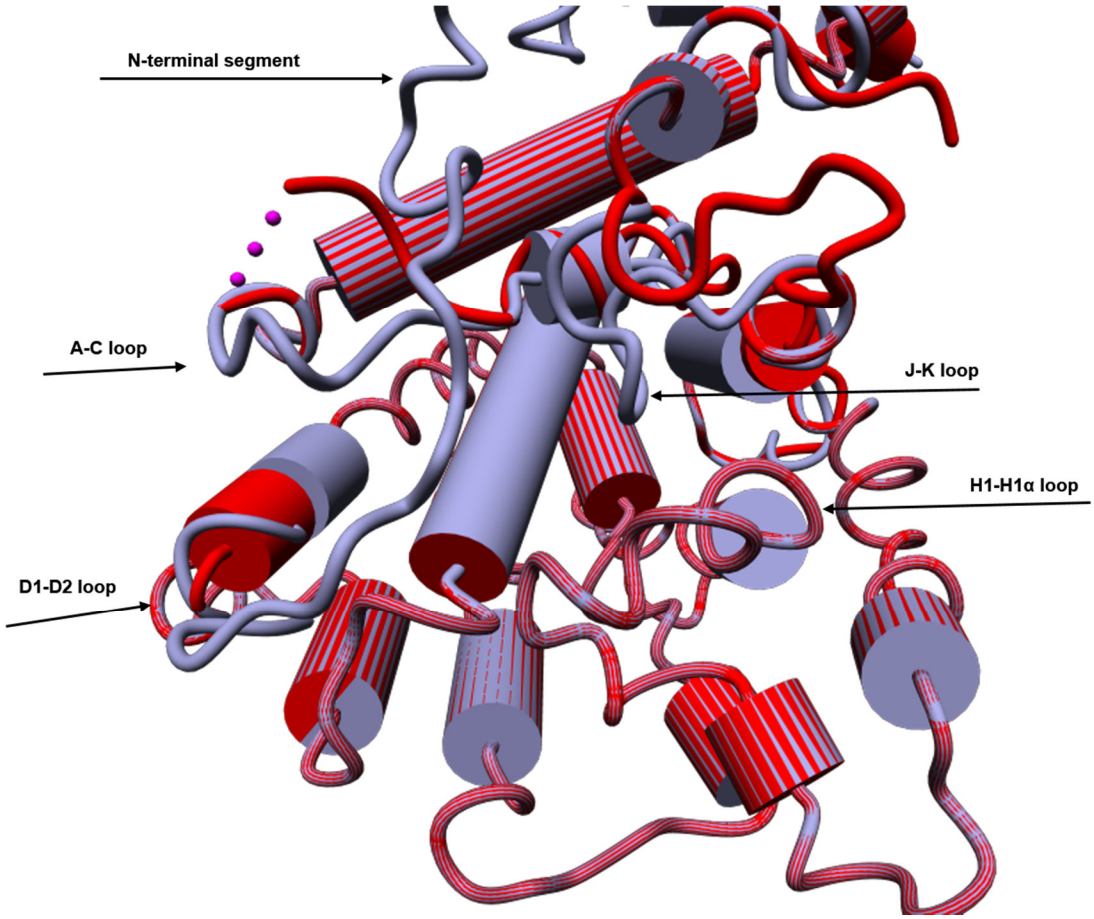


Figure 13: Structural overlay of the open inactive and the active closed conformation of TXS

Figure 14: Structural overlay of the class I domain of the open inactive (red) with the constructed closed conformation of TXS (blue). The closed conformation additionally contains the missing the N-terminus and the so-called J-K and A-C loop segments.

The model of the closed form of taxadiene synthase used in this study, thus contains amino acids 112-862 and additionally the N-terminal segment needed to be catalytically active (amino acids 80-112) (Figure 13, 14b-c). As the crystal structure of TXS additionally contains a 2-F-**GGPP** in an unproductive orientation leading to subsequent carbocations with inverted stereocenters with respect to those shown in Figure 12[58], a geometry optimized structure of **GGPP**, derived from the quantum chemically calculated structure of **cation A**[58], able to produce taxa-

diene with correct relative and absolute configuration was introduced in the modelled active site. The here modelled closed complex illustrates the timepoint imminent to ionization reflecting the start of carbocationic processes. PROCHECK (http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/) analysis of the model (over 90% in the most favoured regions, g-factor of 0.25) and a 10ns MD-simulation confirmed the artificial closed complex to be stable and of good quality. The shape of the J-K and A-C loops and the N-terminus folded into the groove expanded by the A-C and D-D1 loops on one side and the J-K and H-H1α loops on the other, shield in the closed form the initial ionization and carbocationic processes from premature quenching by water (Figure 13). Amino acid-adopted coordination of the three magnesium ions corresponds to that of the open complex and water molecules derived from the crystal structure complete the octahedral coordination (Figure 14c).[47] After a MM-based simulation using the crystal waters of the open complex, we were able to assign the complete water mediated hydrogen bond network of TXS. The closed form impairs several differences in the hydrogen bond network of bound **GGPP** reflecting mechanistically interesting facts concerning ionization and the sequence of carbocation rearrangements. As the A-C and the J-K-loop folded “in” and the N-terminus is bound into the groove between these loops, the pyrophosphate group (**PP$_i$**) accepts now additional water mediated hydrogen bonds from R580 (A-C loop) and Y89 (N-terminus) as shown in Fig .14c.

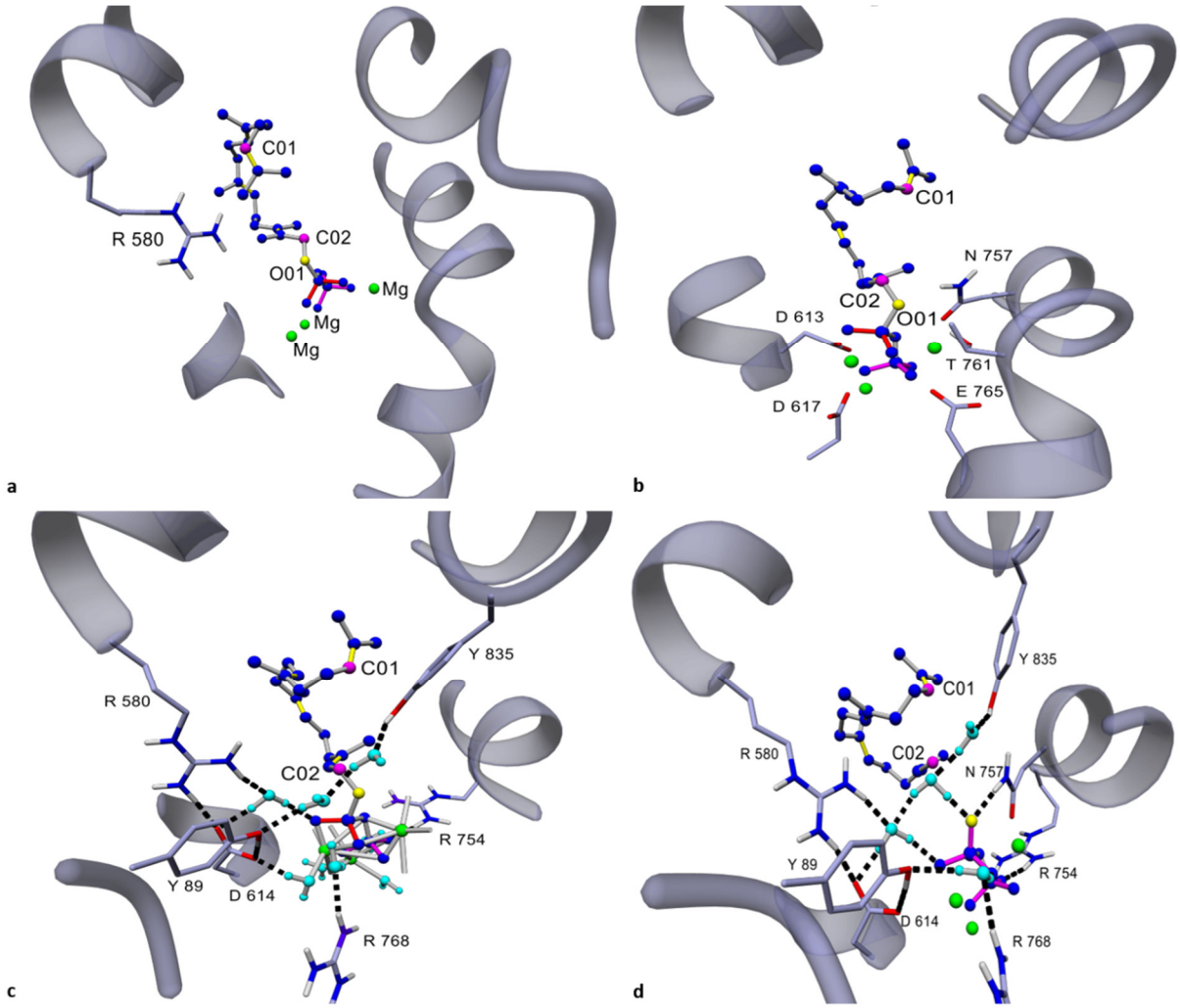


Figure 14: Closed and open conformation TXS model, productive cation A complex

Figure 14**a**: TXS harboring the co-crystallized unproductive 2-F-**GGPP** (PDB: 3RP5) (C01, C02: magenta, O01 of **PP**$_i$: yellow, Mg^{2+}: green, double bonds of 2-F-**GGPP** in yellow). **b**: TXS harboring manually constructed productive **GGPP** (amino acids coordinating Mg^{2+}: D613, D617, N757, T761 and E765). **c**: The closed conformation specific H-bond donor network: Y89, R580, R754, R768 and Y835, coordinated waters: cyan, H-bonds: black, Mg^{2+}-coordinating dative bonds: grey lines. **d**: TXS harboring docked QM-derived productive **cation A**.

Interestingly, amino acid coating of the TXS active site with respect to metal coordination as well as water- and amino acid mediated bonding of productive **GGPP** resembles that of the closed form of BPPS (bornyl diphosphate synthase), a plant class I monoterpene synthase 1:1.[48] This strongly argues for a common ionization and **PP**$_i$ stabilizing mechanism between

class I mono- and diterpene synthases. In both complexes the pyrophosphate group accepts water mediated- and hydrogen bonds from an H-bond donor network composed of Y89/Y60 (TXS/BBPS), R580/R316 (TXS/BPPS), R768/R507, R754/R493, Y835/Y572 (Figure 15).

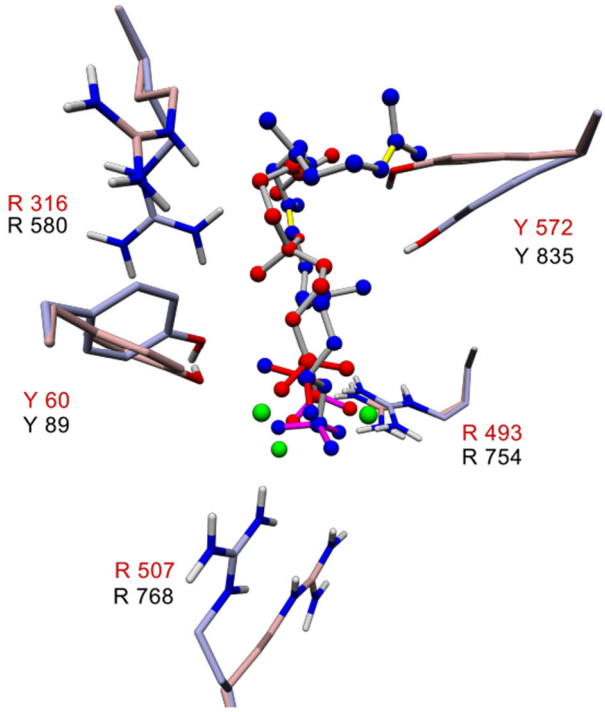


Figure 15: Structural superposition of the H-bond donor amino acid network of BPPS and TXS

Figure 15: Structural superposition of the closed conformation of BPPS (PDB: 1N20) (light red) with TXS (light blue). The amino acids representing the H-bond donor network of BPPS and TXS are shown (BPPS residues in red, the corresponding TXS residues in black).

Structural alignments of TXS, eukaryotic monoterpene synthases BPPS and limonene synthase from *M. spicata* (PDB: 2ONG)[72] as well as eukaryotic diterpene synthases abietadiene synthase from *A. grandis* (PDB: 3S9V)[54], isopimaradiene synthase from *P. abies* (homology model, this work) and casbene synthase from *J. curcas* (homology model, this work) show the same corresponding H-bond donor network with TXS-Arg768 changed to Gln/Glu in abietadiene and isopimaradiene synthase (Figure 16a). The newly solved crystal structures of cyclooct-9-en-7-ol synthase (CotB2), a bacterial diterpene synthase from *S. melanosporofa-*

ciens[45] (PDB: 4OMG; closed conformation model, this work) and selinadiene synthase, a bacterial sesquiterpene synthase (SdS) from *S.pristinaespiralis*[53] (PDB: 4OKZ) crystallized in closed conformation also exhibit this H-bond donor amino acid network (Figure 16b). As this network has also been observed in closed conformations of sesquiterpene synthases from bacteria and fungi[49,52] (Figure 16b), we are now able to generalize that proposed diphosphate-triggered active site closure and charge stabilization mechanism, in that this motive holds **PP$_i$** in place during the complete cascade. Most notably, the R580-**PP$_i$** motif resulting from active site closure is a promising candidate for the active site base, *vide infra*. Importantly, this network is further required to prevent premature quenching by binding of surrounding bulk solvent molecules remaining in the active site after closure. This view is further supported by the fact that the cembratriene-ol synthase from *N. tabacum*[74] (CBTS, closed conformation model, this work*, vide infra*), producing two cembratriene-ol isomers, contains also the three arginine network, albeit misses the tyrosine residues, *vide infra*.

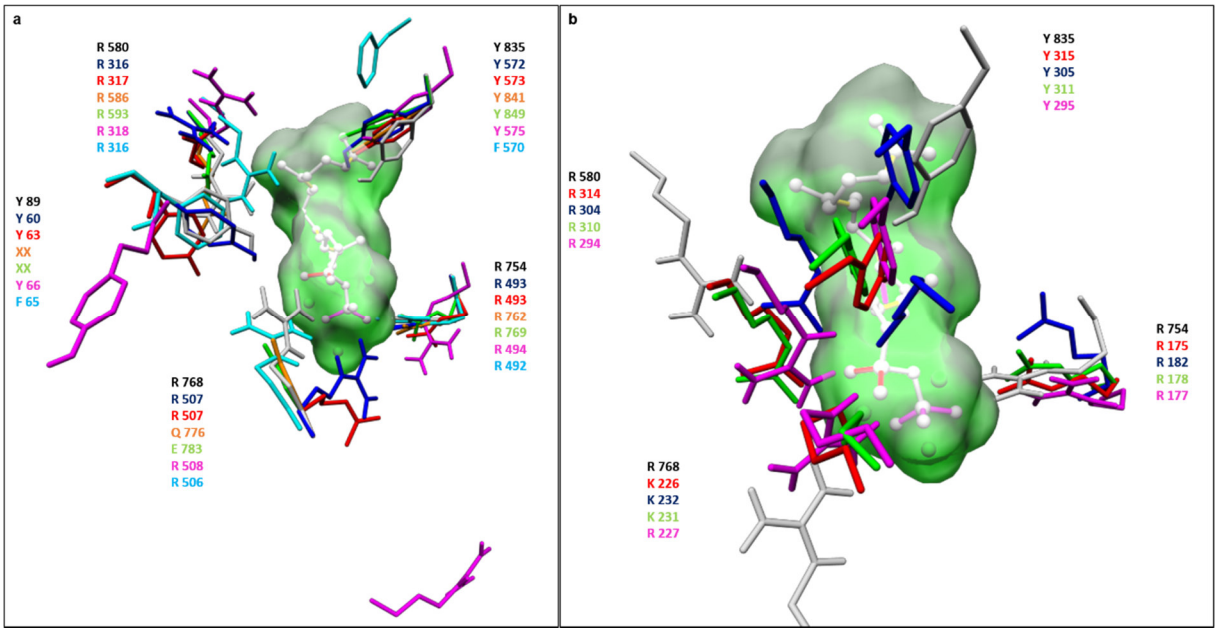


Figure 16: Structural alignments of TXS with eukaryotic and prokaryotic mono-, sesqui- and diterpene synthases

Figure 16**a**: TXS in closed conformation (diterpene synthase: residues in grey, labelling in black, **GGPP** in grey with molecular surface (green)), BPPS in closed conformation (monoterpenesynthase: blue (PDB: 1N20)), limonene synthase from *M. spicata* in closed conformation (monoterpene synthase: red (PDB: 2ONG)), class I domain of abietadiene synthase from *A. grandis* (bifunctional diterpene synthase: orange (PDB: 3S9V)), class I domain of isopimaradiene synthase from *P. abies* (bifunctional diterpene synthase: green (homology model, this work)), casbene synthase from *J. curcas* (diterpene synthase: magenta (open conformation homology model, this work)) and cembratriene-ol synthase (CBTS) from *N. tabacum* (diterpene synthase: light blue (closed conformation homology model, this work)). Note that in the bifunctional diterpene synthases the corresponding second tyrosine (XX) is part of the class II domain and thus could not be taken into account.

Figure 16**b**: TXS in closed conformation (diterpene synthase: residues in grey, labelling in black, **GGPP** in grey with molecular surface), aristolochene synthase from *A. terreus* in closed conformation (fungal sesquiterpene synthase: red (PDB: 2OA6)), trichodiene synthase from *F. sporotrichioides* in closed conformation (fungal sesquiterpene synthase: blue (PDB: 1JFG)), selinadiene synthase form *S.pristinaespiralis* in closed conformation (bacterial sesquiterpene synthase: green (PDB: 4OKZ)) and cyclooctat-5en-7ol (CotB2) from *S. melanosporofaciens* (bacterial diterpene synthase: magenta, closed conformation model, this work: open conformation crystal structure PDB: 4OMG)). Note that bacterial and fungal sesqui- and diterpene synthases comprise a different overall fold with respect to TXS. Nevertheless the class I domain active site cavities correspond to that of their eukaryotic counterparts. The second corresponding tyrosine residue (Y89 in TXS) does not exist in bacterial and fungal sesqui- and diterpene synthases.

At the timepoint imminent to ionization the productive **GGPP** adopts a shape suitable for the *re face* attack of the C1=C15 π-bond on the primary end of the allylic substructure (C2) (Figure 14c). The enzyme seems to reach the proximity of both by protruding the bulky isopropyl group (C15-C17) of linear **GGPP** into a space spanned by W753 on one side and a three almost parallel aromatic ring network, consisting of the benzene rings of F834, Y835 and Y841, on the other (Figure 26a-f). The “so-compressed” productive **GGPP** benefits thus from several hydrophobic interactions from non-polar residues mainly shaping the active site and its π-bonds are stabilized by π-π interactions with aromatic ring systems of W753, F834, Y835 and Y841. The closed complex, hence, seems to prepare ionization of productive **GGPP** during closing the active site by ensuring a close distance between the prospective allylic carbocation and the carbon-carbon double bond to be joined. Proximity of both is thought to aid ionization-initiation of the allylic diphosphate ester bond.[47,58]

Cation A: PP_i stabilization and conformational control imposed by the active site

Cation A is suggested not to be a stable cationic intermediate, but instead reflects exactly the timepoint at that the geranylgeranyl cationic structure captures the productive conformation imminent to cyclization (Figure 12, 14d). Interestingly, **cation A** captures only two distinct conformations in the active site over 999 docking calculations (Figure 17, methods chapter II). One cluster is characterized by a binding energy of 8.472 kcal/mol, the second one by 7.891 kcal/mol (Figure 17, Appendix Table 1). Compared to the localization of unionized **GGPP** in the active site, only the second conformational cluster is derived from the productive **GGPP**, hereafter designated as **cation A** (Figure 17). Clusters of docked cations that are not derived from productive **GGPP** or their former productive cation in respect of the cyclization cascade (Figure 12) are shown in Appendix Figure 1.

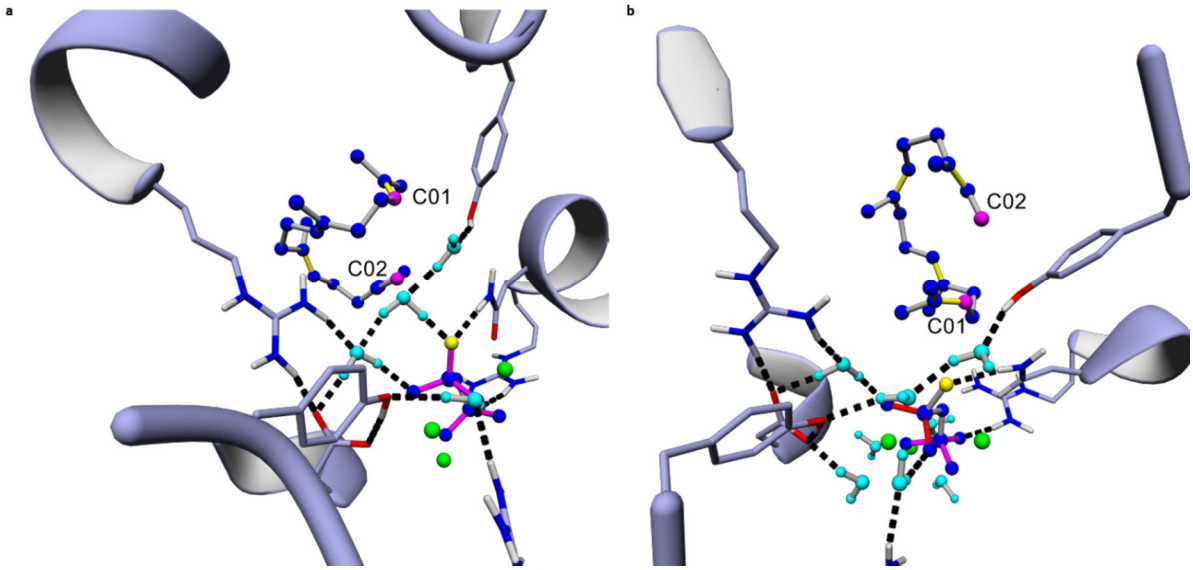


Figure 17: Molecular docking of QM-derived cation A into TXS

Figure 17: Molecular docking of QM-derived **cation A** into the active site of TXS by AutodockVina of YASARA structure. Cluster analysis of 999 docking runs shows two different conformation cluster of **cation A** in the active site. 17**a**: **Conformation A** exhibits a binding energy of 7.206 kcal mol^{-1}, 17**b**: **conformation B** exhibits a binding energy of 7.678 kcal mol^{-1}. Despite a higher binding energy of **conformation B**, the spatial positioning of **conformation B** in the active site does not correspond to the spatial positioning of productive **GGPP** in the modelled

closed complex (Figure 13, 14b-d) and is therefore not in productive sequence **GGPP**->**cation A** during initiation of catalysis. **Conformation A** of **cation A** directly arises from productive **GGPP** after abstraction of $\mathbf{PP_i}$.

Further, using a geranylgeranyl cation (**cation A**) derived from a **GGPP** that reflects the conformation of the 2-F-**GGPP** used during crystallization[47], energy minimization yields also a cationic structure (**A`**) that is productive for cyclization to **B`** (with inverted configuration of C15), but subsequent cyclization from **B`** to **cation C`** yields the unproductive verticillen-12-yl cation with an inverted configuration of H11 that impairs further cyclization to **cation E** (Figure 18). Moreover, **GGPP** in productive conformation without the surrounding enzyme is not able to fold intuitively to **cation A**. This strongly supports the template model hypothesis, *i.e.* terpene synthases mainly dictate reaction outcome by sterically restricting their substrate and discrete reaction intermediates to a subset of possible conformations and thus confirms the accuracy of the modelled closed conformation.[44,55]

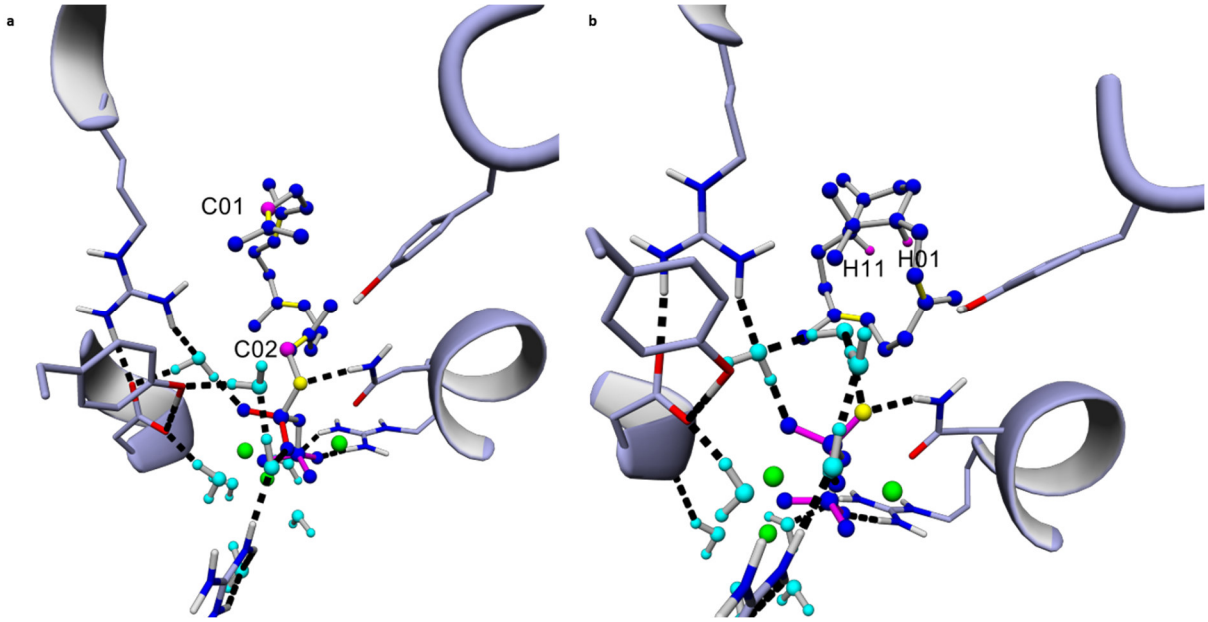


Figure 18: TXS harboring the GGPP conformation derived from the crystal structure

Figure 18**a**: TXS containing a **GGPP** in the unproductive conformation derived from the 2F-**GGPP** conformation of the open complex crystal structure (PDB: 3P5R). **b**: Manual bond formation and energy minimization to **cation C`**. This verticillen-12-yl cation` is characterized by an inverted trans-bridgehead stereochemistry H11 and H1 compared to the verticillen-12-yl cation derived from the productive **GGPP** conformation (Figure 14b), which disturbs the sequence of the hydride shifts. It exhibits moreover an overall inverted positioning in the active and is therefore not productive for the cyclization to **cation F** or -**D**.

Elimination of **PP_i** in TXS is an entropy increasing and exothermic event, as the retained diphosphate anion is stabilized by mesomeric structures, dative binding to the magnesium ions and several water-mediated and amino acid-derived hydrogen bonds.[47,58] The model suggests moreover that hydration of the additional negatively charged O-atom of the retained diphosphate anion, assisted by surrounding amino acids, protruding into the active site during closing of the complex, provides the newly required charge stabilization (Figure 14d), *vide supra*. Transition from open to closed complex brings R580 and N-terminal Y89 into proximity to the bound diphosphate group of **GGPP**. Allocation of charge stabilization *via* protruded amino acid aligned water molecules in the closed conformation strongly argues for an active site-retained **PP_i** counterion during the carbocation cascade. These findings are greatly supported by experimental findings where mutant purified enzymes bearing single substitutions of the H-bond donor network to aliphatic, aromatic or negatively charged amino acids show no or dramatically impaired activities (Appendix Table 2).

Target	Mutation	Activity [a]	E [b]	E1	B	C	F1	F2
TXS		100,0	93,2	4,7	N.D.	0,8	1,1	0,8
V584	V584M[c]	92,3	13,8	0,6	N.D.	83,4	1,4	0,8
V584	V584L	92,1	13,8	0,6	N.D.	83,4	1,4	0,8
S587	S587A[c]	21,8	8,9	N.D.	58,9	32,2	N.D.	N.D.
W753	W753H[c]	51,3	N.D.	N.D.	100,0	N.D.	N.D.	N.D.
Y841	Y841F	41,3	N.D.	N.D.	56,4	43,6	N.D.	N.D.

Table 1: **Characterization of TXS and important mutants**

Table 1**a**: Enzyme activities of mutants (in %) in comparison to TXS (= 100%). The activities were determined by peak integration of the enzyme products in relation to the peak area of the internal standard alpha-humulene (methods chapter II). **b**: Compound numbers **B-E1** refer to compound numbers in Figure 12 and 20. The tabulated values represent the product distributions in % for TXS and each mutant (sum = 100%), (N.D.: Not detectable). c: Mutations and screening have been conducted during and are included in the master thesis of Alexander Büttner.[75]

Cation B: Native existence of a monocyclic intermediate in the carbocation cascade

The next step in carbocation cascade is thought to be conversion of **cation A** to cembren-15-yl cation (**cation B**, Figure 12).[58] However, existence of **cation B** has never been verified using native **GGPP** as a substrate. By manually constructing the C1-C2 bond in docked QM **cation A** followed by energy minimization with stopping before rotation of the C1-C15 isopropyl group occurred, it was possible to detect a structure, which is consistent with the QM-derived **cation B** (Figure 19a).[58] QM calculations suggest existence of **cation B** as distinct energy minimum, although lifetime of **cation B** is predicted to be very small, allowing **cation C** to form without sterical scrambling, as **cation C** is the suggested absolute energy minimum driving the carbocation cascade.[58] Docking of **QM cation B** into its corresponding transient enzyme complex (methods chapter II) further corroborates fitting of **cation B** into the active site of TXS and specifies its conformation as distinctly derived from the docked productive **cation A** complex (Figure 19b, Appendix Table 1).

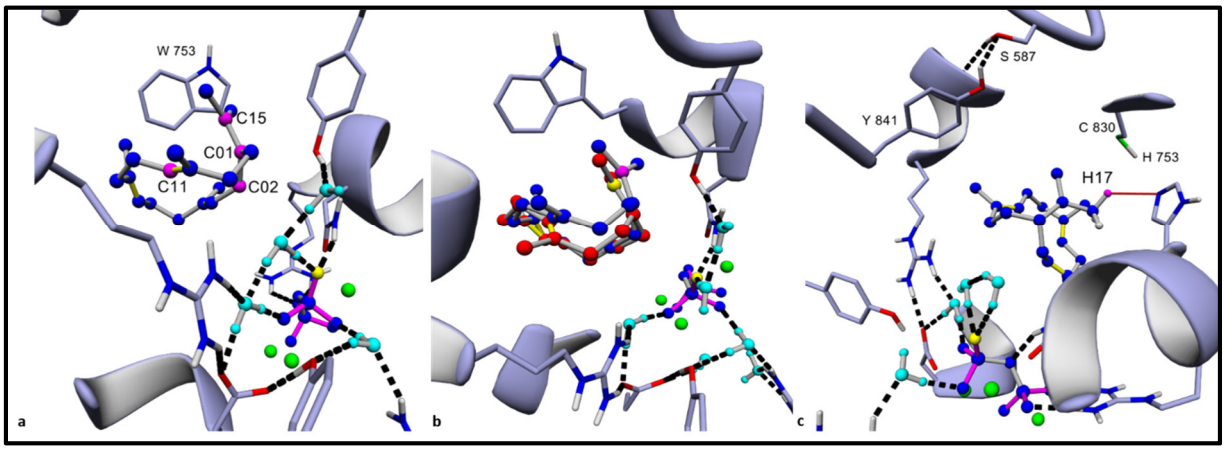


Figure 19: Wild type and mutant cation B complexes

Figure 19**a**: TXS harboring manually constructed **cation B** before rotation of the C1-C15 group occurred (C01, C02, C11 and C15: magenta). **b**: Structural overlay of TXS harboring the manually constructed **cation B** (red, C15: yellow) with docked **QM cation B** (blue, C15: magenta). **c**: TXS-W753H harboring manually constructed **cation B**. Deprotonation of C17 (closest H17: magenta) results in formation of monocyclic **B**. The shielded dipolic network

(S587, Y841) that does not interfere with the cyclization cascade due to hydrogen bonding and spatial separation is further shown.

The **cation B** complex in hand and as it has been speculated that mutant versions of TXS may be expected to form cembranoid isomers from deprotonation of **cation B** as by-products[69], it was consequently attempted to identify the "hot-spot-residues" able to produce monocyclic cembranoid synthons. Although S587, C830, Y835 and Y841 represent polar groups in close proximity to cationic C15 and methyl-hydrogens of C16/C17, none of them act as deprotonating base, as carbocation cascade does not stop at **cation B**, arguing for a protected or a too far removed polar network in the active site that is not interfering with the cyclization processes. Substitutions of Y835 and C830 do not yield any detectable products of *in vitro* incubated mutated TXS enzymes (Appendix Table 2). In contrast, the substitution of W753, operating in stabilization of **cation A** and **-B** to a histidine residue yields a monocyclic structure without any other detectable two- and three-ring structures in an *in vitro* assay using purified enzymes and **GGPP** as substrate (Figure 19c). The W753H product has consequently been assigned as (-)-*R*-cembrene A by MS-spectrometric and CD- as well as NMR spectroscopic analysis (Table 2, Appendix Figure 2-4). Alternatively, H753 may act in deprotonating proximal C830, which leads to a reactive anionic sulphur acting as deprotonating base comparable to the catalytic diad mechanism in cysteine proteases (Figure 19c).[76] The double mutant W753H/C830A yields the same product distribution as W753H. Thus, it seems likely that the histidine residue acts as deprotonating base. Additional substitutions of W753L/V/E/C/A resulted in inactive TXS variants, demonstrating the necessity of an aromatic residue at this position, *vide infra* (Appendix Table 2).

NMR Spectral Data for (-)-R-cembrene A (**B**)			
W753H [a]			R-cembrene A [a]
#	**δc (ppm)**	**δ_H (ppm), J (Hz)**	**δc (ppm)**
1	46.1	2.04 (m)	45.98
2	32.55	1.99 (m)	32.43
3	121.9	5.06 (t, J = 6.1)	121.87
4	134.0	n/a	133.91
5	39.1	2.03 (m)	38.94
6	25.0	2.17 (m)	24.89
7	124.1	5.19 (t, J = 7.2)	124.07
8	134.9	n/a	134.79
9	39.5	2.07 (m)	39.41
10	23.8	2.10 (m)	23.76
11	126	4.98 (t, J = 6.2)	125.90
12	133.5	n/a	133.43
13	34.0	2.13 (m)	33.99
14	28.2	1.96 (m)	28.22
15	149.4	n/a	149.29
16	19.4	1.66 (s)	19.31
17	110.2	4.65 (s), 4.71 (s)	110.10
18	18.1	1.56 (s)	17.99
19	15.4	1.59 (s)	15.25
20	15.6	1.57 (s)	15.48

Table 2: NMR spectral data for (-)-R-cembrene A

Table 2: The chemical shifts of **B** from TXS-W753H are derived from 13C- (Appendix Figure 2) and 1H spectra (Appendix Figure 3). Spectra of **B** have been measured in $CDCl_3$ ([a]). The 13C chemical shifts have been compared with literature.[77]

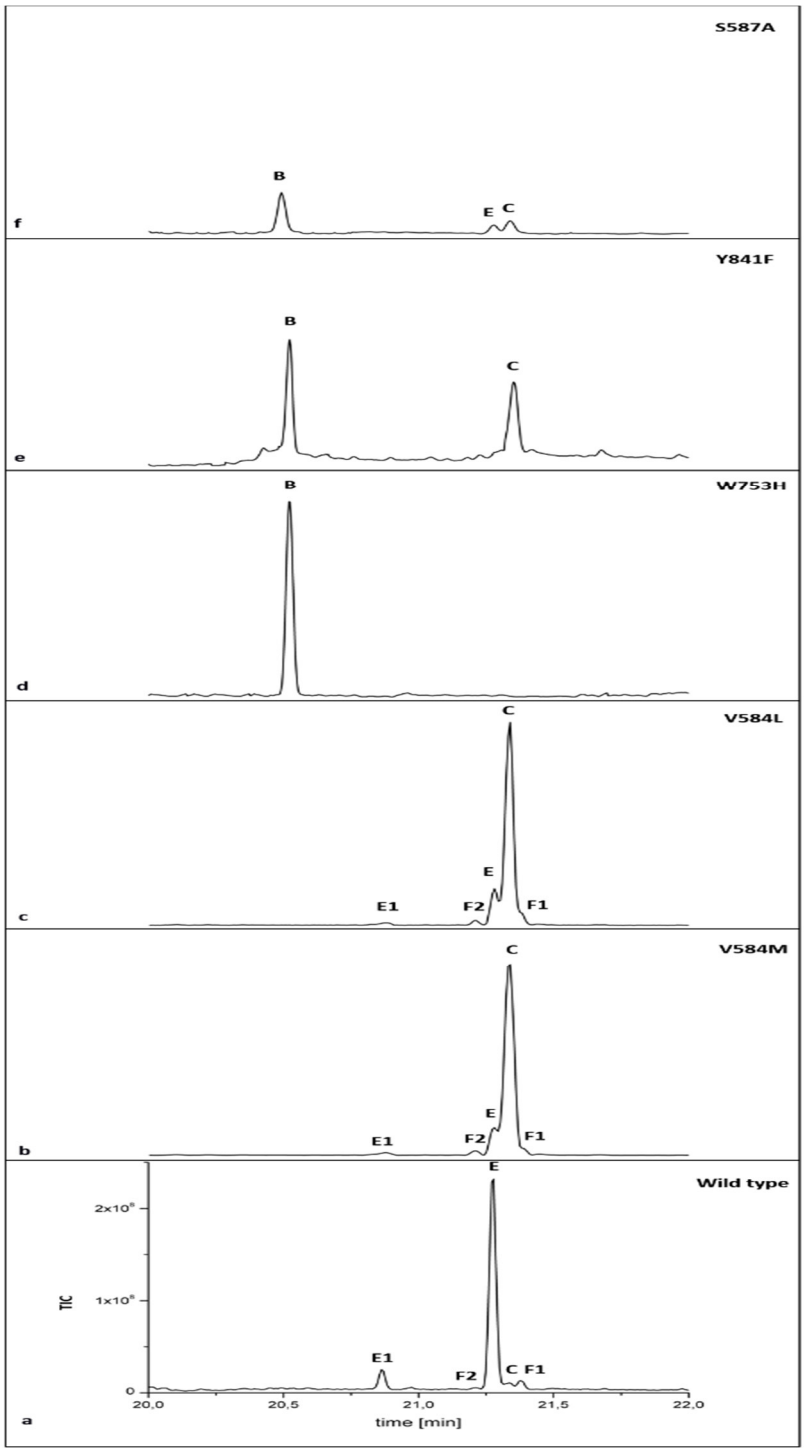


Figure 20: GC-chromatograms of wild type and mutant TXS variants

Figure 20: GC-chromatograms of **a**: wild type TXS, **b**: TXS-V584M, **c**: TXS-V584L, **d**: TXS-W753H, **e**: TXS-Y841F and **f**: TXS-S587A. **E**: taxa-4,5-11,12-diene, **E1**: taxa-4,20-11,12-diene, **B**: cembrene A, **C**: verticilla-3,4-7,8-12,13-triene, **F1**: verticilla-3,4-7,8-11,12-triene, **F2**: verticilla-4,20-7,8-11,12-triene (% activity in comparison to WT corresponding to Table 1, denotation is according to Figure 12).

Cations C-E: Steric and electrostatic control of the carbocation cascade

The verticillen-12-yl cation (**cation C**) is the next cationic intermediate, albeit this intermediate has also never been confirmed using **GGPP** as substrate. However, QM-calculations and incubations of fluoro- and deuterated analogues as well as constitution isomers of **GGPP** lead to verticillene-type products.[58,68,70] In order to get an insight into formation of **cation C**, we again modelled the former by energy minimization of TXS harboring **cation B**, *vide supra*. As expected, a verticillen-12-yl cation was intuitively formed in the active site of TXS with highly similar shape to QM-derived **cation C** (Figure 21a). This cation comprises the known chair like six-membered ring and the trans-bridgehead stereochemistry with H11 axial and H1 equatorial needed for productive cyclization to **cation F** or **-D**.[68] Further, docking analysis confirmed position of **cation C** and establishes it as the distinct conformation derived from productive **cation B** in the active site upon 999 docking runs (Figure 22, Appendix Table 1).

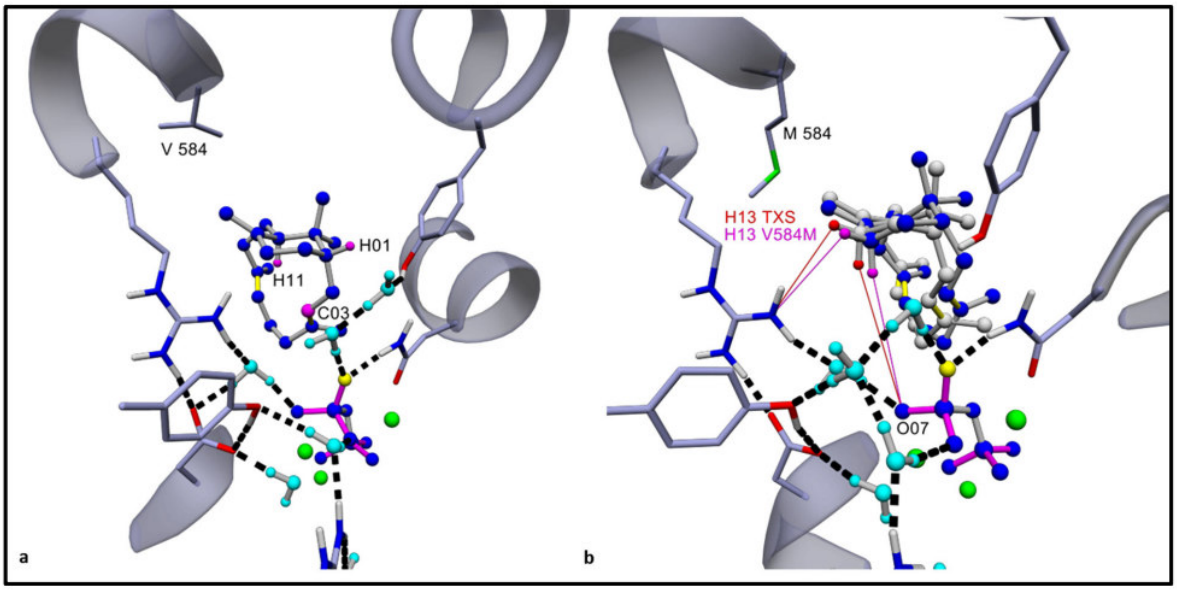


Figure 21: Wild type and mutant cation C complexes

Figure 21**a**: TXS harboring manually constructed **cation C** (H01equatorial, H11axial and C03: magenta). **b**: Super-position of TXS-V584M harboring manually constructed **cation C** (blue) with manually constructed **cation C** of TXS (grey). C13 hydrogens abstracted during R580-**PP**$_i$-assisted deprotonation resulting in formation of **C** are shown (C13 hydrogens of TXS in red, C13 hydrogens of TXS-V584M in magenta). Distance between H13equatorial (H13e) of TXS-V584M and NH2 of R580 is 3.05 Å (magenta line), distance between H13e of TXS and NH2 of R580 is 3.87

Å (red line). Distance between H13axial (H13a) of TXS-V584M and O07 of **PP$_i$** is 4.25 Å, distance between H13a of TXS and O07 of **PP$_i$** is 4.83 Å.

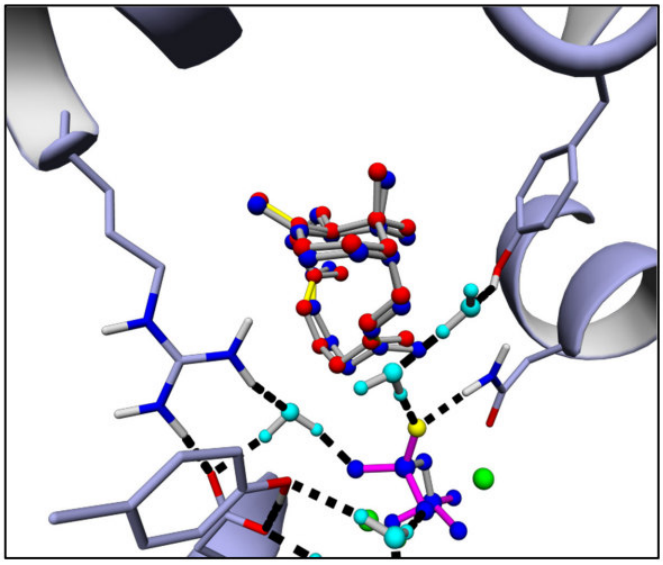


Figure 22: Manually constructed cation C and molecular docking of QM-derived cation C in TXS

Figure 22: Molecular docking of QM-derived **cation C** (red) into TXS and structural superposition with manually constructed **cation C** (blue).

Location of **cation C** in the active site suggests V584 and V610 in close proximity to cationic C 12. Substitutions of V610 to aromatic and polar residues did not result in any detectable diterpenoids during *in vitro* assays (Appendix Table 2). In contrast, single substitutions of V584 to methionine and leucine resulted in taxadiene (14.8%), in a two-cyclic verticillene-type structure (84%) and two minor verticillene-type products (0.8 and 1.4%) in *in vitro* assays as well as in our *in vivo* assay system (Figure 20, Table 1). This product-distribution also could be detected in wild type TXS extracts applying a 10°C/min GC-gradient (Figure 20a). The verticillene-type structure (80%) has subsequently been assigned as (+)-*R*, *R*-verticilla- 3,4-7,8-12,13-diene (**C**), derived from deprotonation of C13 by MS-spectrometric and NMR-

spectroscopic analysis, hereafter designated *endo*-verticillene (Table 3, Appendix Figure 5-11). This premature quenching product has previously been observed, albeit not been characterized.[46] Absolute configuration of this compound has been established by comparison of CD-spectra with the spectra of previously detected 7-fluoro-verticillene products[68] (Appendix Figure 4). Substitutions of V584 to substitutions to N, S, P and R did not show any detectable diterpenoids (Appendix Table 2). However, the V584K substitution results in a 70% *endo*-verticillene and 30% taxadiene distribution. These results greatly confirm the existence of the verticillen-12-yl cation as native intermediate in the carbocation cascade *in vivo* and describes how the enzyme imposes the H1β absolute configuration by the shape of its active site.

NMR Spectral Data for (+)-*R*, *R*-verticilla- 3,4-7,8-12,13-triene (**C**)				
V584M/L				
#	**δc (ppm)**	**δ_H (ppm), *J* (Hz)**	**HMBC**	**NOESY**
1	42.5	1.51	16, 17	
2	34.3	1.84 , 2.64	1	1, 17, 20,
3	124.8	5.29 (br d, J = 11.9)	2, 20	5, 7, 11, 14, 18
4	132.8	n/a	20	
5	41.1	2.02 , 2.17	6, 20	
6	26.8	1.99 , 2.44	5	5
7	130.0	4.83 (br d, J = 12.0)	6, 19	3, 5, 6, 9, 11, 18
8	133.1	n/a	19	
9	39.7	1.96 , 2.16	10, 11	
10	21.6	1.37 , 1.67	9, 16	
11	38.2	2.97	10, 16, 17	18, 3, 7
12	136.0	n/a	10, 14	
13	121.8	5.39 (br s)	14	14, 18
14	30.9	1.87, 2.47	1	2
15	35.8	n/a	1, 16, 17	
16	27.3	0.74 (br s)	17	
17	23.8	0.81 (br s)	16	
18	23.1	1.76 (br s)	10, 11	
19	15.9	1.50 (br s)	9	6, 2
20	15.4	1.58 (br s)	2	6

Table 3: NMR Spectral data of (+)-R,R-verticilla-3,4-7,8-12,13-triene

Table 3: The chemical shifts of **C** from TXS-V584M/L are derived from 13C- (Appendix Figure 6), 1H- (Appendix Figure 7), 1H/13C- HMBC (Appendix Figure 9) and 1H/1H NOESY spectra (Appendix Figure 11). Spectra of **C**, have been measured in $CDCl_3$.

The modelling approach using *in silico* mutated enzymes suggests a mechanism of deprotonation at C13. Note, that carbocation formation with mutated enzymes at the V584 "hot-spot" exhibit respective positions in the active site until **cation C** is reached. Compared to the wild type enzyme, however, **cation C** is localized in closer proximity to the multiple counterion charges of $\mathbf{PP_i}$ due to steric restraints of M and L at position 584 in these mutant enzymes (Figure 21b). As a consequence axial H at C13 is shifted 1 Å towards the R580-$\mathbf{PP_i}$ moiety in comparison to the wild type, providing an indication for premature deprotonation in 84%, most likely by this bi-functional motif. The two minor products in the wild type as well as the V584M/L mutant spectra with retention times 20,77 and 20,93 in the GC chromatogram (Figure 20a-c, Appendix Figure 12a, b) could be characterized according to their mass spectra in comparison with their 7-fluoro-analogues and earlier findings as verticilla- 4,20-7,8-11,12-triene (**F2**) (20.77) and 3,4-7,8-11,12-diene (**F1**) (20.93) (Figure 20a-c). The closed conformation further predicts the existence of these two minor products as premature deprotonated intermediates due to the proximal positioning of **cation F** and the R580-$\mathbf{PP_i}$ motif (Figure 23a, Appendix Figure 12a, b).

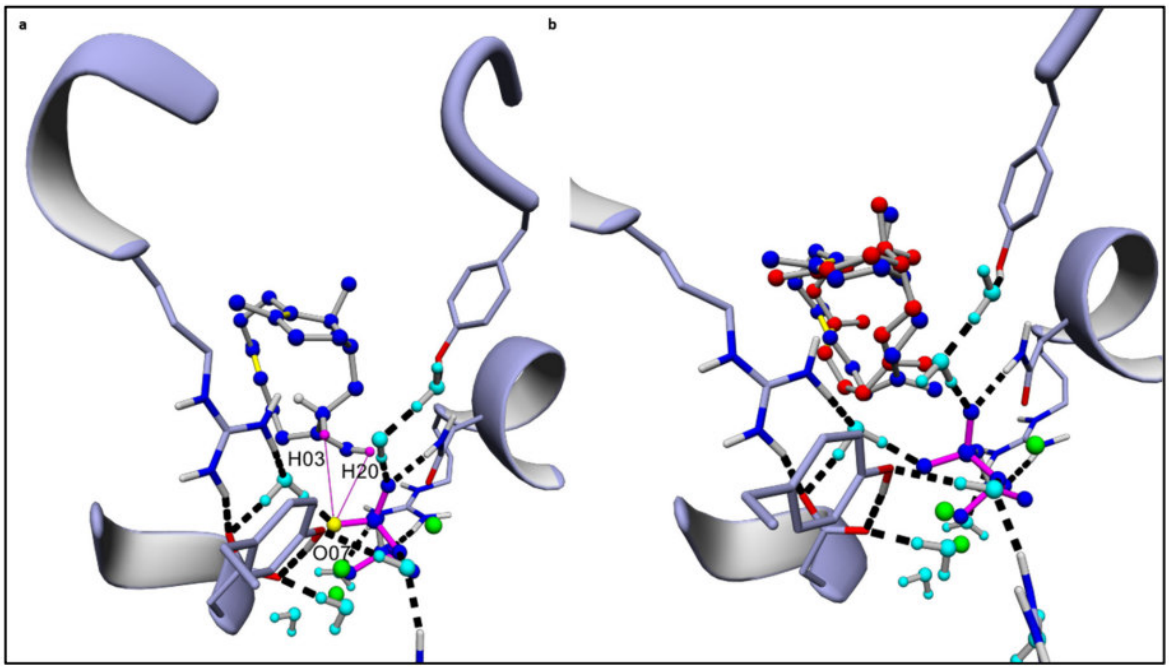


Figure 23: Manually constructed cation F and molecular docking of QM-derived cation F in TXS

Figure 23**a**: TXS containing manually constructed **cation F**. Distance (magenta line) between H03 and O07 of **PP$_i$** is 3.19 Å, distance (magenta line) between H20 (closest) and O07 of **PP$_i$** is 3.69 Å. Abstraction of C03 in this **cation F** by the R580-**PP$_i$** motif results in formation of **F1**, abstraction of one of the C20 hydrogens results in formation of **F2** (Figure 12). The distance between R580-**PP$_i$** and the positively charged C-atom (C03) is 4.06 Å. **b**: Molecular docking of QM-derived **cation F** (red) into TXS and structural superposition with manually constructed **cation F** (blue).

Previous reports further describe single residue switches in eukaryotic diterpene synthases, able to "short-circuit" the cyclization cascade by simple electrostatic stabilization.[35,78,79] Consequently and on the basis of the model we also try to find switches to "short-circuit" the reaction cascade by simple electrostatic stabilization of the pathway intermediates. As mentioned earlier, the dipole network of TXS's active site seems to be protected or is too far removed to interfere with the cyclization cascade (Figure 19c). The model suggests the hydroxyl groups of S587 and Y841, interacting by hydrogen bonding in the wild type, as in

close proximity to **cation C**. Consequently, we generated mutants with disrupted dipole networks. As predicted, single substitutions of S587 to A and Y841 to F were able to "short-circuit" the reaction cascade. The Y841F mutant only produces cembrene A (**B**) and verticilla-3,4-7,8-12,13-diene (**C**), *i.e.* cyclization is fully aborted at **cation C**, whereas S529A mainly produces cembrene A (**B**) and verticilla-3,4-7,8-12,13-diene (**C**) with minor amounts of taxadiene (**E**) (Fig. 20, Table 1).

Carbocation cascade cations A-E: Steering effect of R580-PP$_i$ determines promiscuity

Gas phase formation of **E** proceeds from **cation F** via **cation D** and **-E** (Figure 12), although the direct route **cation C->D->E** cannot be excluded by QM calculations and labeling experiments.[58,68] **Cation D**, which adopts two different **conformations D1** and **D2**, represents the fifth step in the carbocation cascade (Figure 12). MM-based modelling and docking of **cations F** (Figure 23b), **-D1** (Figure 24a, b), **-D2** (Figure 25a, b) and **-E** (Figure 27a) into TXS indicated that each of them adopt a position in TXS that is distinctly derived from the respective former cation (Appendix Table 1).

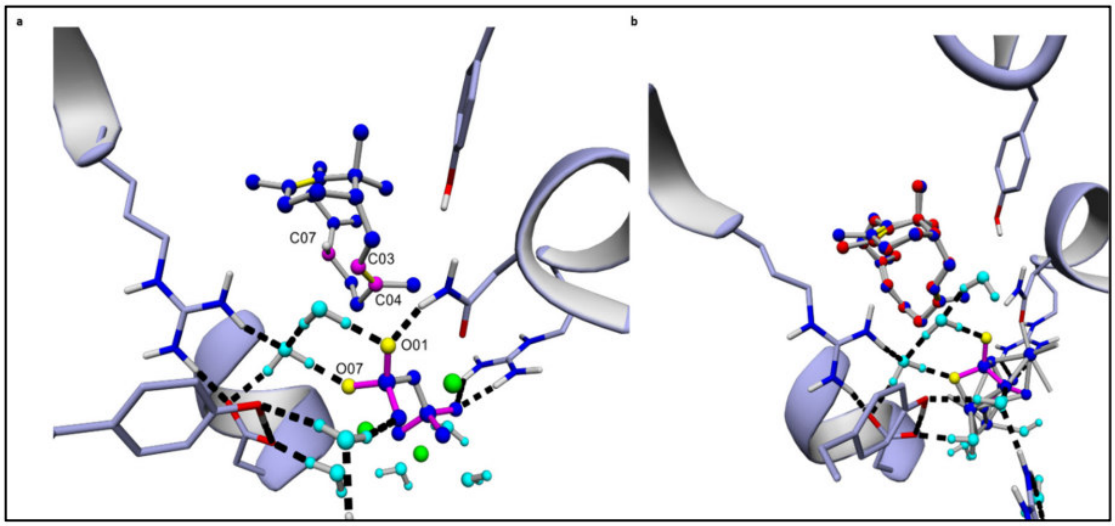


Figure 24: Manually constructed cation D1 and molecular docking of QM-derived cation D1 in TXS

Figure 24**a**: TXS containing manually constructed **cation D1**. **b**: Molecular docking of QM-derived **cation D1** (red) into TXS and structural superposition with manually constructed **cation D1** (blue).

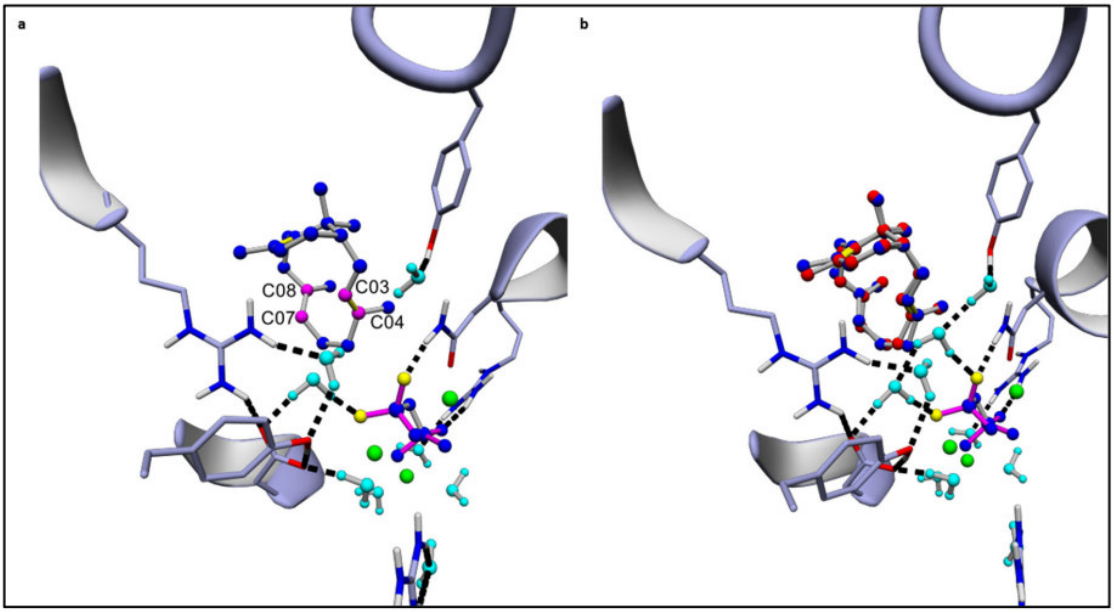


Figure 25: Manually constructed cation D2 and molecular docking of QM-derived cation D2 in

Figure 25**a**: TXS containing manually constructed **cation D2**. **b**: Molecular docking of QM-derived **cation D2** (red) into TXS and structural superposition with manually constructed **cation D2** (blue).

It has been reported that **cation D1** has to undergo a conformational change to **-D2** in order to induce formation of **cation E**.[58] Complex models containing **cations D1** and **-D2** indicate that this change of the **-D1**-C8 methyl group towards an aligned position with the C4 methyl group in **-D2** is induced by an electrostatic repulsion effect of R580-**PP$_i$** (Figure 24, 25). Furthermore, our cumulative data demonstrates a pronounced steering effect of R580-**PP$_i$** that compete with amino acid-assisted electrostatic stabilization during the complete catalytic cascade. The trajectories suggest that after ionization a charge separation between **cation A** and R580-**PP$_i$** occurs that situates the carbocation deep into the hydrophobic pocket, where **cation A** is stabilized by π-π- and cation-π interactions with W753, F834, Y835 and Y841. During transition of **cation A** to **-B** this stabilization persists, however, subsequent **cations C-E** show a reduced π-derived stabilization (Figure 26a-f).

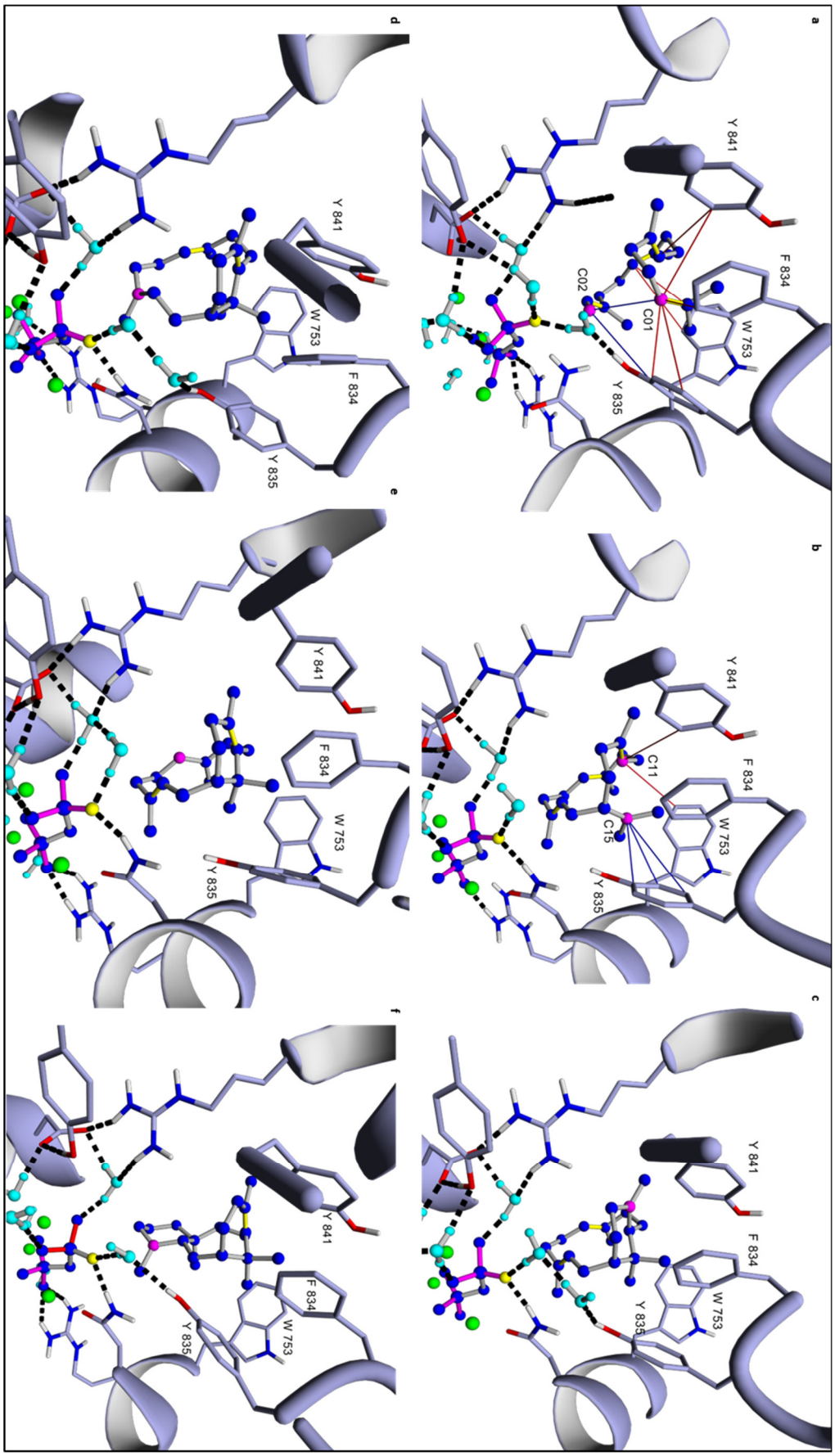


Figure 26: Amino acid-assisted stabilization of cations in TXS

Figure 26**a-f**: Amino acid-assisted π-π (red) and cation-π interactions (blue) acting in stabilization of cations in the active site. **a**: TXS harboring manually constructed **cation A**, **b**: TXS harboring manually constructed **cation B**, **c**: TXS harboring manually constructed **cation C**, **d**: TXS harboring manually constructed **cation F**, **e**: TXS harboring manually constructed **cation D1** (**-D2** is also characterized by no amino acid-assisted π-π and cation-π interactions) and **f**: TXS harboring manually constructed **cation E**.

Therefore, TXS seems to stabilize early (**cation A** and **-B**) but not late stage (**cation C-E**) intermediates. This further suggests that the R580-**PP$_i$** steering and cationic charge-attracting electrostatic effect begins to dominate when the cationic cascade has reached the bi-cyclic stage (**cation C**). This reflects also the experimental observations, as product distributions of TXS as well as V584M/L and S587A mutants show ongoing cyclization to **E** and the existence of all bi- and tricyclic deprotonated intermediates (**C**, **F1**, **F2**, **E** and **E1**) compared to W753H and Y841F mutants (Figure 20, Table 1). This is further corroborated by the F834G mutant that shows all bi- and tricyclic intermediates, but no deprotonated monocyclic structure (**B**) upon *in vitro* incubations with **GGPP** (Appendix Table 2). At **cation C**, this steering electrostatic effect of R580-**PP$_i$** is able to explain the subsequent transition to **cation F**, as the positive charge locally approaches the closest proximity to R580-**PP$_i$** (Figure 23a). This, in turn, suggests that transition from **cation C->F** is an exothermic event in contrast to the gas phase.[58] Despite close proximity of the positive charge at **cation F** with respect to R580-**PP$_i$**, however, other electrostatic effects must counteract those exerted by R580-**PP$_i$** and provide the driving force for progression to **E**. The model indicates that R580-**PP$_i$** and C4 hydrogens of **cation F** are too far removed for terminal deprotonation. Instead, electrostatic attraction of the π-orbital of Δ7,8 in **cation F** directly pointing towards the positive charge seems to initiate the expected hydride shift resulting in transition of **cation F** to **-D1**. This indicates that besides steric and electrostatic control of the cascade, also kinetic effects has to be taken into account. Spatial positioning and dynamic movement of cationic intermediates restricted by the active site do not allow terminal deprotonation in **cation F**. Instead, kinetic control leads to a thermodynamically unfavored transition of **cation F** to **-D1**. However, after **cation D1** has been formed, positive charge migration towards R580-**PP$_i$**, induced by this bi-functional motif and in absence

of π-stabilization seems to be responsible for the transition of **cation D1** to **-D2**. Accordingly, transition of **cation D2** to **-E** is induced by the steering effect of R580-**PP$_i$**. The model of the TXS•**cation E** complex shows in concordance the global distance minimum of the positive charge with respect to R580-**PP$_i$** (Figure 27a). This further suggests that **cation E** is the global energy minimum of the cascade in relation to **cation A**, which is in contrast to QM gas phase calculations.[58] Docking cluster analyses of the modelled trajectories further corroborate these findings (Appendix Table 1).

Furthermore, our model indicates that R580-**PP$_i$** may also be responsible for TXS product promiscuity in addition to its oversized active site cavity. The excess of spatial freedom allows **GGPP** misfolding at initial stages of catalysis and carbocation tumbling at later stages governing alternative product formation. Additionally, imprecise carbocation barrier crossings allow premature deprotonation leading to non-specific quenching product.[44,56] Therefore, tumbling and imprecise barrier crossing due to the lack of π-stabilization at later stages may also be the reason for alternative products observed during the catalytic cascade. Tumbling effects or imprecise barrier crossings can thus explain premature deprotonation in **cation C** and **-F** resulting in formation of **C**, **F1** and **F2** in TXS as well as V584M/L, S587A and F834G mutants (Figure 20, Table 1, and Appendix Table 2). By contrast, multiple product formation derived from **cation E** (**E** (93.2%) and **E1** (4.7%) (Fig. 18)), cannot be explained by tumbling or imprecise barrier crossing alone but may also involve R580-**PP$_i$**-governed electrostatic effects. Interestingly, the models show that the observed product distribution derived from deprotonation of **cation E** is induced by differing relative distances of R580-**PP$_i$** with respect to **cation E**-H5β or -C20 methyl hydrogens (Fig. 27b).

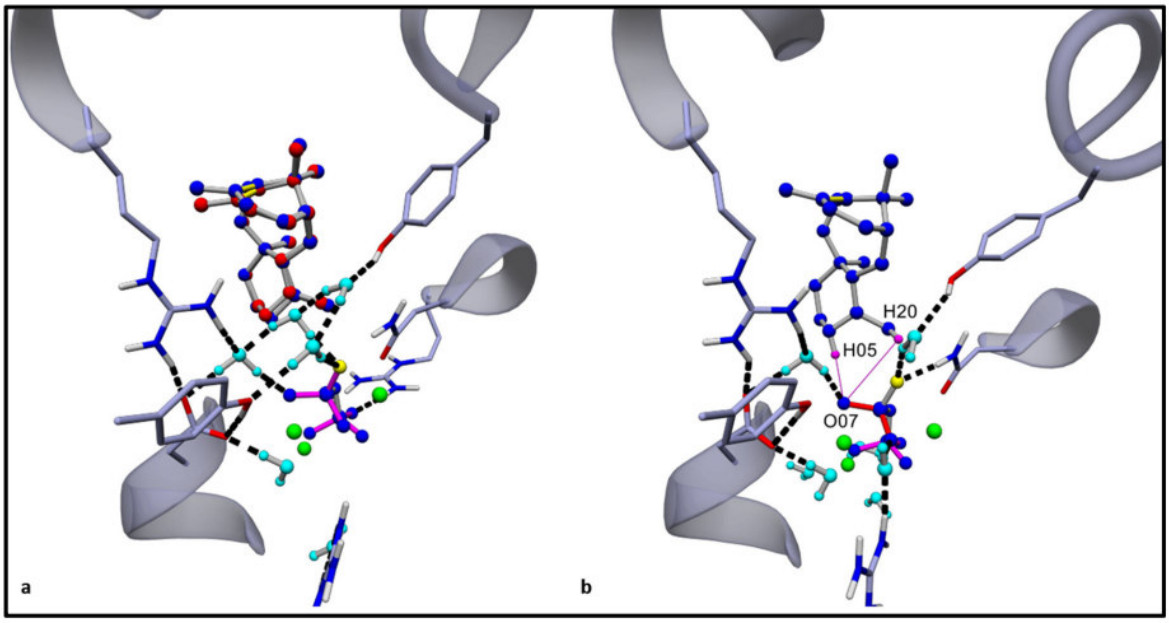


Figure 27: Molecular docking of QM-derived cation E, product promiscuity in cation E

Figure 27**a**: Superposition of TXS harboring manually constructed **cation E** (blue) with the docked QM gas phase-derived **cation E** (red). **b**: TXS harboring **cation E**. Differing distances of H05 equatorial (deprotonation results in formation of **E**) and closest H20 (deprotonation results in formation of **E1**) with respect to O07 of R580-**PP**$_i$ are shown as magenta lines. Distance of equatorial H5 of the newly formed cyclohexane ring (twist-boat conformation) to O07 of R580-**PP**$_i$ is 2.75 Å, distance of closest H20 to O07 is 3.35 Å. The distance between the positively charged C atom with respect to R580-**PP**$_i$ reaches its global minimum of the cascade at formation of **cation E** (3.71 Å).

Biotechnological implications: Production of cembranoid- and verticillene-type synthons

At 30L scale fermentations the three plasmid *in vivo* terpene production system (methods chapter II) was able to produce **B** (W753H) and **C** (V584M) at concentrations of 8 mg/L and 11 mg/L, respectively, with glycerol as the sole carbon source (Figure 28, 29). This is the first time that both compounds were produced in recombinant microbial production systems. While cembranoid synthases have been reported[77,80], this is the first report of a terpene synthase

capable of producing verticillene type skeletons. As minimal functionalization of both macrocycles yields bioactive compounds our methodology provides a sustainable production route for complex natural products, which are not accessible from their natural source.[34,81-83]

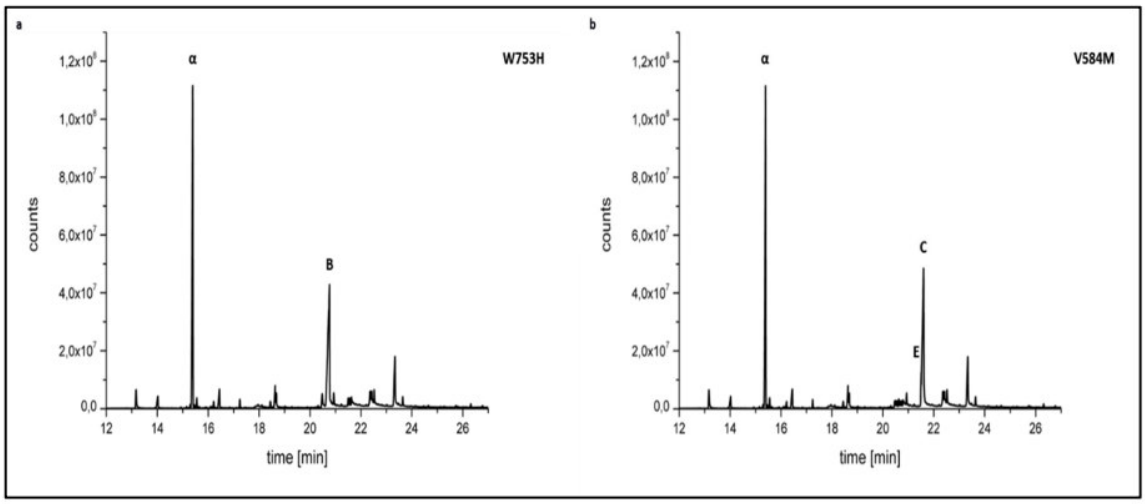


Figure 28: GC-FID Spectra of batch bioprocess supernatants of TXS-W753H and TXS-V584M

Figure 28**a**: GC-FID spectrum of the extracted supernatant (500ml) of the batch bioprocess of TXS-W753H according to method chapter II. (**α**: alpha humulene standard, **B**: cembrene A). **b**: GC-FID spectrum of the extracted supernatant (500ml) of the batch bioprocess of TXS-V584M according to methods chapter II. (**α**: alpha humulene, **C**: verticilla-3,4-7,8-12,13-triene, **E**: taxa-4,5-11,12-diene).

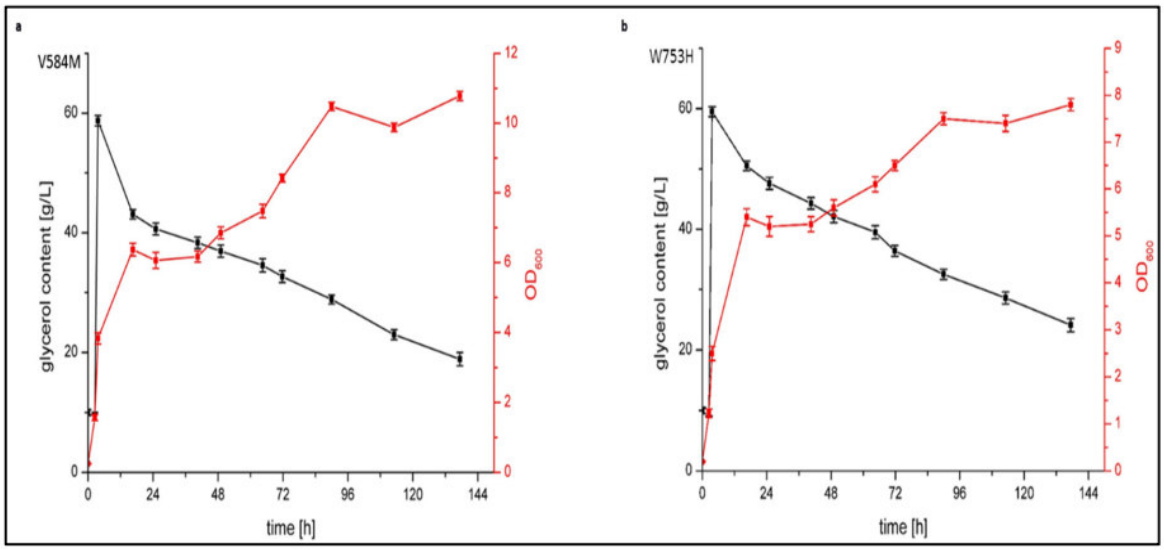


Figure 29: Batch bioprocess characteristic of TXS-W753H and TXS-V584M

Figure 29**a**: Time course of glycerol content (black line) and OD_{600} during the 30L batch fermentation process of TXS-V584M-*in* vivo production system (according to methods chapter II) with glycerol as sole carbon source. **b**: Time course of glycerol content (black line) and OD_{600} during the 30L batch fermentation process of TXS-W753H-*in* vivo production system (according to methods chapter II) with glycerol as sole carbon source.

4. Discussion

Elucidation of the underlying mechanisms of the complex cyclization cascade or snapshots of closed productive complexes of terpene synthases by crystallization or NMR-spectroscopic approaches is highly challenging using native substrates. An alternative approach to get an insight into mechanistic behaviour and to identify “plasticity” residues leading to biotechnologically important quenching products is demonstrated in this study. Construction of MM-based productive closed complexes harboring QM gas phase-calculated or manually-derived cascade intermediates, combined with subsequent *in silico* mutagenesis approaches is a cheap and fast methodology. Especially for the multitude of eukaryotic mono- and diterpene synthases the modelled productive closed conformation of TXS could serve as starting point for the elucidation of cyclization mechanisms and localizing of important residues and switches in

the corresponding active sites bearing significant possibilities *en route* to diversified products with biopharmaceutical implications. Furthermore, the force-field-based MM modelling approach is able to give excellent approximations with respect to localization of cationic intermediates in the active site compared to the docking approach using the proposed QM cascade intermediates. In cases, where intermediates are known and open conformation models exist, this methodology could easily be applied. We verified our *in silico* toolkit for the modelling of the closed conformation of the bacterial diterpene synthase CotB2 (methods chapter II), whose open complex conformation[45] and proposed carbocationic intermediates[84] have recently been established (Appendix Figure 13 and Schrepfer et al., PNAS, submitted).

The identified H-bond donor network greatly argues for a common evolutionary mechanism in terpene synthases throughout the kingdoms of life. Beneath charge stabilization, the model is able to establish an H-bond donor network providing a mechanism for separation of the ionized **PP_i** from the carbocation intermediates to prevent premature recapture. The requirement for charge stabilization and separation also argues for the ionized **PP_i** to be retained in the active site during the whole cyclization cascade, as previously suggested.[48,49,56,72,73] This bears tremendous consequences with regard to electrostatic steering effects and active site base behaviour in class I terpene synthase catalysis.

Further, the fact that the QM-derived **cation A** captures only one distinct productive conformation in the active site at the timepoint imminent to cyclization greatly argues for the view that terpene synthases mainly dictate reaction outcome by sterically restricting their substrate and discrete reaction intermediates to a subset of possible conformations (template model).[44] Moreover this finding confirms the artificial closed complex to be a valuable model for the elucidation of structure-function relationships. The protected TXS dipole network and the observations that disruption of this network via single substitutions lead to premature quenching products in **cation B** and **-C** corroborates the suggested hypothesis. Although **cation B** is a distinct energy minimum in QM-calculations considering only the cationic structures, MM-

based energy minimizations with the surrounding enzyme suggest the direct route from **cation A** to the transition state between **cation B** and **-C** (**TS B-C**), the intermediate state after the rotation of the C1-C15 bond occurred (Figure 30). However, a distinct **cation B** could be detected by manually stopping the energy minimization before the rotation occurs (Figure 19a). Furthermore, the *in vitro*- and *in silico*-confirmed amino acid-assisted deprotonation at C17 in the W753H mutant proofs for the existence of a short lived **cation B** structure during cyclization, albeit not a global energy minimum, as shown by docking cluster analysis (Appendix Table 1). In accordance with the template model, it is not surprising that the cationic structure is preorganized for the direct conversion to **cation C** by active site amino acids as **cation C** displays a far lower energy level. This greatly argues for a concerted but not synchronous cyclization reaction from **GGPP->A->B->C** where distinct **cations A** and **-B** exist as short lived intermediate states.

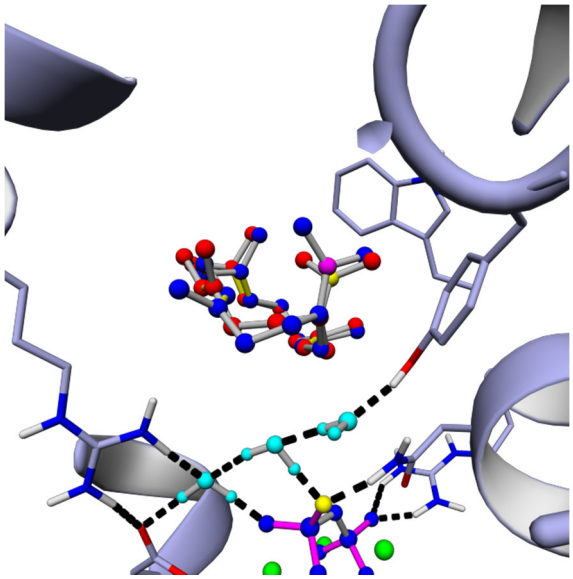


Figure 30: Structural superpositon of manually constructed cation B and TS B-C

Figure 30: Structural superposition of manually constructed **cation B** (red, Figure 19) with manually constructed **TS B-C**. Manual bonding of the C1-C2 bond of **cation A** and subsequent energy minimization yields rather **TS B-C** than **cation B**.

Examination of TXS harboring **cation C**, in turn, combined with experimental mutagenesis experiments enabled the elucidation of plasticity residues that are responsible for diversified diterpene products. Introduction of sterically more demanding residues compared to V584 brings **cation C** into closer proximity of R580-**PP$_i$** leading to an eighty percent premature deprotonation. The existence of a deprotonated *exo*-7-fluoroverticillene (7-fluoroverticilla-3,4-7,8-12,18-triene), found previously upon incubation of TXS with 7-fluoro-**GGPP**[68] (Appendix Figure 15), suggests further a role of R580-**PP$_i$** instead of rather **PP$_i$** alone, as **PP$_i$** is far away from the C12-C18 bond in the closed conformation 7-fluoro-**GGPP**-model (Appendix Figure 14a, b). *In silico* cyclization of 7-fluoro-**GGPP** to the 7-fluoro-**cation C** establishes the location of the 7-fluoro-**cation C** in the active site (Appendix Figure 14a, b). Electrostatic stabilization of cationic center C12 by F7 seems to perturb the fine-tuned kinetics of the cascade while R580-**PP$_i$** provides for a continuous electrostatic attracting force. The other 7-fluoroverticillene-type structures, seem according to the 7-fluoro-models to arise from R580-**PP$_i$**-assisted deprotonation of 7-fluoro-**cation F** leading to *endo*-7-fluoro-verticillene and 7-fluoroverticilla-4,20-7,8-11,12-triene (Appendix Figure 14c). Deprotonation by R580-**PP$_i$** is therefore a more than plausible explanation for the existence of a deprotonated *endo*-verticillene in the wild type TXS spectrum, presumably due to cation tumbling during the cascade.

In addition to this sterical control of product outcome, also electrostatically driven changes in the active site are shown to lead to deprotonated intermediates. Our model suggests that deprotonation in **cationic state B** and **C** in mutants S587A and Y841F is also the result of the disruption of the fine-tuned and electrostatically balanced timeline of cyclization and hydride shifts triggered by the introduction or “freeing” of dipoles in the active site (Appendix Figure 16, 17). In both mutant enzymes, dipole-assisted stabilization of intermediates long enough for direct (**cation C**) and presumably indirect (**cation B**) proton abstraction by R580-**PP$_i$** seems to be the reason for premature deprotonation. As S587 is in closer proximity to **cation B** or **-C** in comparison to Y841, the Y841F substitution that “frees” dipolic S587 is much more efficient in

aborting the cyclization cascade at **cation B** or **-C** (Figure 20). A possible mechanism for deprotonation in Y841F and S529A at **cation B** is the electrostatic stabilization of the proximal cationic C15 by S587 or Y841, accompanied by a hydride-shift of one of the hydrogen atoms of C17 to Δ3, 4. Deprotonation of Δ3, 4 by R580-**PP$_i$** seems then to be similar to deprotonation of **cation F** leading to verticilla-3,4-7,8-11,12-diene (**F1**), *vide supra*. Deprotonation of **cation C** at C13 in Y841F and S587A seems to arise from electrostatic stabilization of cationic C12 by S587 or Y841, accompanied by deprotonation by R580-**PP$_i$**.

The closed conformation model containing **cation C** suggests the route cation **C->F->D** as energetically more favourable as distances between Δ3, 4 and H11 are slightly shorter than between Δ7, 8 and H11 leading to route **C->D**, as previously suggested by gas-phase calculations without enzyme and retained **PP$_i$** (Figure 12).[58] Together with the finding of the two minor verticillene-type products (**F1** and **F2**) in wild type- and V584M/L chromatograms, this represents the first direct experimental evidence for the existence of **cation F**. Deprotonation in the V584M/L and S587A mutants seems not be an absolute event compared to W753H or Y841F mutants as product spectra show an ongoing cyclization to **cation F**, **-D** and **-E** (Figure 20). These results either indicate that **cation C-E** exist in a well-defined equilibrium, where taxadiene formation is just promoted by site-specific deprotonation by R580-**PP$_i$** or alternatively that **cation E** is the absolute energy minimum of the cascade and the cyclization to that point is electrostatically governed by R580-**PP$_i$** and sterically restricted by the coating of the active site.

The results of our modelling approach clearly support the alternative hypothesis as docking cluster analysis and modelling suggest that **cations F** and **-E** comprise a lower energy level in relation to **cation A**, due to their positioning in the active site. Furthermore, the steering and cation-attracting role of retained R580-**PP$_i$** during the whole cascade seems to lead to direct or indirect premature deprotonation in cases where the fine-tuned and electrostatically balanced timeline of cyclization and hydride shifts is disrupted and π-stabilization is reduced.

These facts, in turn, indicate that an R580-**PP$_i$** motif, arguably characterized by an inverse protonation equilibrium, appears to be the prime candidate for the active site base in TXS. A corresponding bi-functional R-**PP$_i$** motif appears thus to be the prime candidate in the other examined terpene synthases that produce terpene synthons. Therefore, extended cyclization to the final cation is just driven by steric coating of the active site and by electrostatic "steering" by the ionized R580-**PP$_i$**-motif.

The mechanistic picture derived in this stuy is entirely consistent with all mechanistic experiments and *in sillico* modelling studies of TXS and other terpene synthases. This mechanism that includes active site amino acid-assisted structural and kinetic chaperonage combined with **PP$_i$**-governed electrostatic steering of the cyclization cascade is generally applicable for any other class I terpene synthase. Additionally, but not less important, the subsequent *in silico*-based mutagenesis methodology provided for biotechnological production of cembranoid- and verticillene-type core structures as basis for known bioactive diterpenoids and may also lead to the development of new semisynthetic or enzyme-based bioactive compounds, if applied to other enzymes of the terpene synthase superfamily.

5. Methods

Quantum chemical calculations (QM)

All calculations were performed with *GAUSSIAN03.*[85] All **cationic intermediates A-E** and **GGPP** in productive conformation were optimized using the B3LYP/6-31+G(d,p) method as described previously.[58]

Model of TXS in closed conformation harboring productive GGPP

Molecular dynamics (MD) simulation and energy minimization steps were performed in a periodic neutralized TIP3P water box at pH 6.8 of YASARA Structure (version 14.12.2, YASARA Biosciences) using the second generation self-parametrizing YASARA2-force field with default settings.[86] Assigned hydrogens, bond orders and charges for **GGPP** and **cations A-E** were obtained using AutoSmiles force field parameter assignment that is part of YASARA Structure.[87,88]

Preparation of closed conformation of TXS

Modeling of the N-terminal residues (80-110) and the J-K loop (residues 834-849) in closed conformation was performed with bornyl diphosphate synthase in complex with three Mg^{2+} ions and 3-azageranyl diphosphate (PDB accession code 1N20) as template[48] and modeled as described previously[47] followed by 10.000 steps of energy minimization and a 10 ns MD simulation in water using YASARA structure. During energy minimization all residues, heavy atoms and 2-F-**GGPP** were unconstrained. Non-bonded cutoff were set to 7.5 Å. PME (Particle Mesh Ewald method) was used for long-range electrostatic forces.[89]

Integration of GGPP in productive conformation

The geometry-optimized **GGPP** in productive conformation obtained by B3LYP/6-31+G(d,p), as described previously[58], was superposed onto the 2-F-**GGPP** in the modeled closed conformation, the 2-F-**GGPP** deleted and the bond between O1 and C1 linked manually. The so

obtained complex was subsequently subjected to 10.000 steps of energy minimization and 10ns MD simulation in water using YASARA Structure.

Molecular Docking of QM derived cation A-E

Molecular Docking of QM-derived **cations A-E** was performed using the AutoDock Vina program environment of YASARA Structure.[90] Cations were docked into a simulation cell (Size: X-size = 14 Å, Y-size = 14 Å, Z-size= 14 Å, angles: alpha = 90°, beta = 90°, gamma = 90°) around the following eight residues Y31, E 583, F602, F612, S752, L827, A764 and Y836. For each cation 999 docking runs were performed while all atoms of the corresponding cations were set as rigid. Cluster analysis was performed in the AutoDock Vina program environment and cluster were characterized by binding energy [kcal mol^{-1}], dissociation constant [pM] and binding energy spread [kcal mol^{-1}] (Appendix Table 1). Two enzyme-ligand complexes belonged to different clusters if the ligand RMSD (root meant square deviation) was larger than 2.0 Å.

Molecular modelling of cations in the active site of TXS

An approach of force-field-based MM modelling combined with molecular docking to derive localization of **cations A-E** was applied. First, using the AutoDockVina program environment of YASARA Structure the **QM-cation A** was docked into the active site of the closed conformation still bearing the retained diphosphate group bound to the tri-nuclear magnesium cluster and hold in place by the established basic amino-acid H-bond donor network. From this initial structure, bonds that reflect cyclization and hydride shifts *en route to* **cation E**, following the path suggested by *in vacuo* geometry optimization[58], were manually constructed. After construction of each closed conformation harboring its corresponding cation ("transient enzyme complex") the whole system was subjected to several rounds of energy minimization and a 1 ns-MD-simulation using the self-parametrizing YASARA2 force-field of YASARA Structure. Subsequently, the corresponding cations derived by quantum chemical calculations were

docked into the active sites of these “transient enzyme complexes” in order to elucidate existence of additional conformations in these “transient enzyme complexes”. Consecutively, localization of docked cations were compared with the manually constructed corresponding cations by structural superposition of Cα-atoms using YASARA structure. Conformational clusters that are not derived from the conformation of productive **GGPP** or the respective former cascade carbocation according to Figure 12 were not considered in cluster analysis. Docking and cluster calculations of QM-derived **cations B-E** and **TS B-C** were used to validate the MM-based manual bond forming method. Simple docking of cationic intermediates into the closed conformation technically harboring unionized **GGPP** would not reflect the dynamic events during the carbocation cascade.

Structural alignments

(Multiple-) structural alignments of different terpene synthases and superposition of Cα-atoms of cations for calculations of RMSDs were performed using the built-in MUSTANG module of YASARA Structure.[91]

Homology modeling of terpene synthases

Homology modelling was performed by the homology modeling module built-in YASARA Structure, according to the manufacturers protocol (http://www.yasara.org/homologymodeling.htm). This approach also includes loop modeling, side-chain rotamer fine-tuning, hydrogen bond network optimization and knowledge-based force field high-resolution refinement with explicit solvent molecules. For homology modelling of Casbene synthase form *J. curcas* the Genbank Accession number AB687998 was used as template. For corresponding homology modelling of isopimaradiene synthase from *P.abies* Genbank Accession number AY473620.2 was used.

Model of closed conformation of CBT-ol synthase from *N. tabacum*

A genomic truncation of CBT-ol synthase (Genbank Accession number: AAS46038.1) lacking the predicted amino-terminal transit sequence (amino acid 1-50), hereafter designated as

CBTS, was subjected to the homology modeling module of YASARA Structure. The refined model was subsequently subjected to a structural alignment with the closed conformation of TXS as template. The structural alignment had a RMSD of 0.31 Å over 498 residues with 26.78% primary sequence identity. The unfolded A-C- and J-K loops as well as the N-terminus of CBTS were then replaced by its corresponding secondary structure elements of TXS followed by "back-mutating" to the corresponding CBTS residues and subsequent 10.000 steps of energy minimization and 10 ns MD simulation in water using YASARA Structure. A second structural alignment between closed conformation of TXS and the "back-mutated" and minimized CBTS were generated and the productive **GGPP** from TXS were integrated into this CBTS complex. Thereafter, water molecules were transferred from TXS to CBTS and dative bonds between $\mathbf{PP_i}$, water and the corresponding amino acids were established. A second energy minimization and MD-simulation procedure yielded the productive closed conformation of CBTS.

Model of closed conformation of Cotb2 from *S. melanosporofaciens*

The modelling approach was equivalent to that of CBTS. The structural alignment of chain D of selinadiene synthase in complex with dihydrofarnesyl pyrophosphate (PDB ID: 4OKZ_D)[53] and the open complex of CotB2 (PDB ID: 4OMG)[45] had a RMSD of 0.18 Å over 287 residues with 13.47% primary sequence identity. After replacement and "backmutating" of the corresponding loops followed by energy minimization and MD-simulation, the productive **GGPP** from TXS was integrated in this CotB2 complex. Establishment of dative bond between $\mathbf{PP_i}$, water and the corresponding amino acids followed by energy minimization and MD-simulation yielded the productive closed conformation of CotB2. Localization of **cations A-F** in the active site of CotB2 were established by manually breaking and linking the corresponding bonds according to the recently published carbocation cascade mechanism followed by energy minimization using the YASARA2 force field of YASARA Structure.

General experimental procedures

All chemicals were obtained from standard sources at the highest purity grade. NMR spectra were recorded in $CDCl_3$ with an Avance-III 500 MHz device (Bruker) at 300 K. ^1H NMR chemical shifts are given in ppm relative to $CHCl_3$ (δ=7.26 ppm) or CD_2HOD (δ=3.31 ppm) and CD_3OH (δ=4.87) (^1H NMR). ^{13}C NMR chemical shifts are given in ppm relative to $CDCl_3$ at δ=77.16 ppm or CD_3OH at δ=49.00 ppm. The 2D experiments (HSQC, HMBC, TOCSY, COSY, NOESY) were performed using standard Bruker pulse sequences and parameters.

GC-MS and GC-FID analysis of diterpene products from n-hexane extractions was done by a Trace GC Ultra with DSQII (Thermo Scientific). One µl sample was applied by TriPlus AS onto a SGE BPX5 column (30 m, I.D 0.25 mm, Film 0.25 µm). The initial column temperature was 50°C (maintained for 2.5 min). A temperature gradient was applied from 50°C – 320°C (10 °C/min), followed by 3 min maintenance at 320°C. MS data were recorded at 70 eV (EI), *m/z* (rel. intensity in %) as TIC, total ion current. The recorded *m/z* range was 50 – 650. Quantification was performed with flame ionization detector (FID) using 1 mg/L α-humulene (Sigma-Aldrich, Germany) as an internal standard.

High-resolution mass spectra of diterpenes were determined with a Thermo Scientific DFS Magnetic Sector GC-HRMS system from n-hexane extractions. One µl sample was applied by TriPlus AS onto a SGE BPX5 column (30 m, I.D 0.25 mm, Film 0.25 µm). The initial column temperature was 50°C (maintained for 2.5 min). A temperature gradient was applied from 50°C – 320°C (10 °C/min), followed by 3 min maintenance at 320°C. The recorded *m/z* range was 50 – 650.

Circular dichroism (CD) spectroscopy was performed using a Chirascan plus spectropolarimeter (Applied Photophysics, United Kingdom). Samples were dissolved in acetonitrile and spectra were recorded in quartz cuvettes with 0.1 cm path length at 20 °C.

Glycerol content of the fed-batch bioprocess was quantified by HPLC, using an Agilent LC 1100 system (Agilent technologies, Waldbronn, Germany), equipped with an autosampler, column oven and a Shodex RI-101 detector (Showa Denko Europe GmbH, Munich). A Rezex ROA-Organic Acid H+ (8%) ion-exclusion column (300 mm, 7.8 mm internal diameter; Phenomenex LTD, Aschaffenburg, Germany) was used for the isocratic separation with 5 mM sulfuric acid at a flow rate of 0.5 ml/minat 70°C.

Bacterial strains, genes and vectors

The *E. coli* strains XL-1 Blue and BL21 (DE3) were used for cloning and diterpene production. All strains and plasmids were obtained from Novagen/Merck Millipore (Germany). Genes were synthesized by GeneArt, Regensburg (Thermo Fisher Scientific) featuring the appropriate restriction sites and adjusting codon usage for *E. coli*.

Genes and vectors used for *in vitro* production of taxa-4,5-11,12-diene

pET28b (+) vector was used for cloning and recombinant expression of a codon optimized version of the M60 truncation of taxadiene synthase from *Taxus brevifolia* (henceforth designated "TXS").[46] The Gene of taxadiene synthase (*txs*) (GenBank: AF326519.1) (Appendix section 6: Genes) was synthesized as M60 truncation including an NdeI restriction site at the 5`- and an XhoI restriction site at the 3`-end Primer used were 5`- AAACATATGGCAATGAGCAGCAGCACC-3` and 5`- AAACTCGAGCTGAATCGGATCGAT-3`. The synthetic gene was introduced into the corresponding restriction site of pET28b (+) vector by standard cloning techniques.

Genes and vectors used for *in vivo* production of taxa-4,5-11,12-diene

The 1-deoxy-D-xylulose 5-phosphate (DXP) pathway was housed by the plasmids pColaDuet-1 and pCDFDuet-1 (Novagen), while the taxadiene biosynthesis genes were carried by the pETDuet-1 plasmid. To overexpress the DXP pathway in *E. coli* BL21(DE3), genes from *E. coli*

of the 1-deoxy-D-xylulose 5-phosphate synthase (*dxs*) (GenBank: YP_001461602.1) (Appendix section 6: Genes), 1-deoxy-D-xylulose 5-phosphate reductoisomerase (*dxr*) (GenBank: NP_414715.1) (Appendix section 6: Genes)), 2-C-methyl-D-erythriol 4-phosphate cytidyltransferase synthase (*ispD*) (GenBank: NP_417227.1) (Appendix section 6: Genes)), 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (*ispF*) (GenBank: NP_289295.1) (Appendix section 6: Genes)) and isopentenyl-diphosphate delta isomerase (*idi*) (GenBank: NP_417365.1) (Appendix section 6: Genes)) were synthesized. *IspD*/*ispF* was created as a bi-cistronic operon (Appendix section 6: Genes)). The synthetic genes were introduced into the appropriate plasmids according to Table 4 by standard cloning techniques

Gene(s)	**Vector**	**Multiple Cloning Site**	**Restriction Sites**
dxr	pColaDuet-1	I	*Nco*I, *EcoR*I
dxs	pColaDuet-1	II	*Nde*I, *Xho*I
ispD/ispF operon	pCDFDuet-1	I	*Nco*I, *EcoR*I
idi	pCDFDuet-1	II	*Nde*I, *Xho*I

Table 4: Plasmids used to construct the overexpressed DXP pathway in* E. coli *BL21 (DE3)

To achieve biosynthesis of taxadiene and its deprotonated intermediates the native geranylgeranyl diphosphate synthase (*crte*) (GenBank: M90698.1) was amplified from *Pantoea agglomerans* (ATCC 27155) using standard protocols. Primers used were 5'-AAA CCA TGG CAA TGG CAA CGG TCT GCG CA-3' and 5'-AAA GAA TTC TTA ACT GAC GGC AGC GAG TTT-

3'. The genes of *crte* and synthetic *txs* were introduced into the appropriate plasmids according to Table 5 by standard cloning techniques.

Gene(s)	**Vector**	**Multiple Cloning Site**	**Restriction Sites**
crtE	pET-Duet-1	I	*Nco*I, *EcoR*I
txsM60	pET-Duet-1	II	*Nde*I, *Xho*I

Table 5: Plasmids used to construct the taxadiene biosynthesis in* E. coli *Bl21 (DE3)

Site-directed mutagenesis

Mutation primers were designed applying the Agilent Technologies QuikChange program (Appendix Table 4). For PCR the following mixture was used: 5 µL Pfu Ultra buffer (10x, Agilent), 1 µL forward/ reverse primer (2 µM stock), 1 µL dNTPs (10 µM stock), 200 ng template DNA (plasmid), 1 µL Pfu Cx Hotstart polymerase (Agilent) and water to a final volume of 50 µL. After the PCR reaction, 2 µL of DpnI (20 u, NEB) were added and the mixture was incubated at 37 °C for at least 2 h. 2 µL of sample were transformed into XL1 blue competent cells (Agilent). After 1 h of incubation at 37 °C in SOB medium, cells were streaked out on LB kanamycin or in case of the *in vivo* approach ampicillin agar plates. Colonies were grown overnight at 37°C. Single colonies were picked and inoculated in LB kanamycin or ampicillin medium overnight. Plasmids were isolated applying a plasmid miniprep kit (Agilent). The mutant TXS genes were verified via sequencing (MWG Eurofins, Ebersberg) using the T7 forward/T7 reverse primer set in case of pET28b (+) or DuetUp2 (Novagen)/T7 reverse primer set in case of pET duet vector.

Expression of TXS and mutant variants for *in vitro* screening

Single transformations of pET28b (+) harboring wildtype *txs* or mutations were cultivated for 18h at 37°C in Luria-Bertani (LB) medium supplemented with 50µg/mL kanamycin. These pre-inocula were used to seed fresh 500 mL LB media containing 50 µg/mL kanamycin at a starting OD_{600} of 0.1. After reaching an OD_{600} of 0.6 the cultures were cooled down to 16°C, induced by 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown at 16°C for 24 h. Cells were harvested by centrifugation at 6.000 g for 15 min and suspended in buffer A (25 mM 3-(N-Morpholino)-propane-sulfonic-acid (MOPS) pH 6.8, 1 mM $MgCl_2$ and 1 mM DTT (Dithiothreitol). Cells were disrupted by sonification on ice four times (60 s on and 90 s off) at medium power. Cell debris was cleared by centrifugation at 30.000 g for 1h. The clear supernatant was applied to a pre-equilibrated HisTrap HP 5 ml column (GE Healthcare) at a flow rate of 1 mL/min with an ÄKTA Purifier system (GE Healthcare). The loaded column was washed one time with 5 column volumes of buffer A plus 20 mM imidazole and after that eluted with a linear gradient of 20-500 mM imidazole at a flow rate of 2.5 mL/min. SDS-PAGE (12%) analysis was used to evaluate expression. Selected fractions were combined and dialysed over night at 4°C against buffer A followed by another 6 h dialysis against fresh buffer A. Subsequently the solution was concentrated to 2 mL and subjected to isocratic size exclusion chromatography using a Superdex S75 10/300GL column (GE Healthcare) with buffer A at a flow rate of 0.8 ml/min. Fractions were combined and concentrated to 2.5 mL. Concentration was determined by absorbance at 280 nm and enzyme were stored in 100 µl aliquots at -20°C for further use.

***In vitro* assay using purified enzymes and GGPP**

100 µl of enzyme solution (1mg/L) were added to 400 µl of buffer A. **GGPP** (Sigma Aldrich) (43.53 µM) was added to the solution. After gentle mixing, reaction solution was incubated for 12 h at 28°C and 400 rpm. For extraction of products 750 µl n-hexane (VWR) was added to the reaction solution. After vigorous vortexing (2 x 20 s) phases were separated by centrifugation for 10 min, 4°C and 12.000 rpm. N-hexane extraction was repeated twice, organic phases

were combined and evaporated to dryness under continuous N_2-stream. The crude extract was resolved in 1 ml n-hexane. 100 µl of that resolved extract were evaporated to dryness under continuous N_2-stream, solved in 90 µl and 10 µl of the internal standard (alpha-humulene, Sigma-Aldrich, final concentration 0.889 mg/L) were added. The samples were analyzed by GC-MS.

In vivo production of deprotonated intermediates using mutated TXS variants

The *in vivo* production system was used to generate sufficient amount for NMR analyses. The *in vivo* approach was based on a culture volume of 3 x 1L LB medium in 5L baffled glass flasks.

The vectors pColaDuet-1 (*dxp, dxs*), pCDFDuet-1 (*ispD/ispF, idi*) and the vector pETDuet-1 (*crte, txs*) were introduced into *E. coli* BL21 (DE3) by standard transformation procedures. For cultivation in shake flasks, single transformants were grown in 3 x 1L medium containing LB-Media (10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl) supplemented with 10 g/L glycerol, 30 µg/mL kanamycin, 50 µg/mL streptomycin and 50 µg/mL carbenicillin. The cultures were inoculated at OD_{600} 0.1 from an overnight culture (8 h cultivation 37 °C supplemented with 30 µg/mL kanamycin 50 µg/mL streptomycin and 50 µg/mL carbenicillin), grown at 37°C and 130 rpm until OD_{600} of 0.8 and then cooled down to 25°C. At 25°C 40 g/L glycerol was added and the cultures were induced by addition of 1 mM IPTG. Cells were grown on 25°C for 3 days.

Extraction and isolation taxa-4,5-11,12-diene and its deprotonated intermediates

For diterpene isolation, cells were sedimented (15 min, 17,500 g, 4 °C). Taxa-4,5-11,12-diene and its deprotonated intermediates were extracted from *E. coli* cells and supernatants, separately. The cell pellets were washed with water and then suspended in 5 mL water. Next, cells were lysed by sonification on a Sonoplus HD2070 device (Bandelin Electronic, Germany) performing 5 repeats on ice (5 min on and 3 min off at 80 % power) and subsequently extracted 3 times with 25 mL n-hexane. Supernatants were extracted 3 times by 200 mL n-hexane.

Organic phases were combined, dried with $MgSO_4$ and evaporated under vacuum to dryness. The crude extract was solved in 1 mL n-hexane and analysed by GC-MS and -FID. Purification of verticillia-3,4-7,8-12,13-triene was carried out by flash chromatography using an isocratic 90/10 n-hexane/ethyl acetate silica step (Silica gel 40, Sigma-Aldrich, Germany). Purification of cembrene A was carried out using the same procedure. The crude extract was resolved in 1 ml n-hexane 100 µl of that resolved extract were evaporated to dryness under continuous N_2-stream, solved in 90 µl and 10 µl of the internal standard (alpha-humulene, Sigma-Aldrich, final concentration 0.889 mg/L) were added. The samples were analyzed by GC-MS.

Batch bioprocess for the production of cembranoid- and verticillene-type macrocycles

30L fermentations were performed in a 75 L bioreactor (LP351, Bioengineering AG, Wald, Switzerland) using LB-Media (10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl) supplemented with 10 g/L glycerol, 30 µg/mL kanamycin, 50 µg/mL streptomycin and 50 µg/mL carbenicillin. The pH was controlled at 6.8 with 4 M NH_4OH and 5 M H_3PO_4. Oxygen saturation was constantly adjusted to 80 %. 1 mg/L Antifoam B (Sigma Aldrich) was automatically added when necessary. The bioprocess was started by inoculation from 4 x 1L overnight cultures (8 h cultivation 37 °C supplemented with 50 µg/mL kanamycin, 50 µg/mL streptomycin and 50 µg/mL carbenicillin) at OD_{600} = 0.25 and run 3.5 h at 37°C. At a OD_{600} = 3.8 40 g/L glycerol and 1 mM ITPG were added aseptically, the culture was cooled down to 25°C and grown for another 5 days. To determine optical densities and glycerol contents 10 mL aliquots were taken as triplicates. 1 mL of each triplicate was used for determination of OD_{600}, 1 mL was used for determination of glycerol content by HPLC analysis. After the bioprocess 500 mL of culture was used for determination of diterpene contents using n-hexane extraction of the supernatant as depicted in section 16. The crude extract was resolved in 1 ml n-hexane 100 µl of that resolved extract were evaporated to dryness under continuous N_2-stream, solved in 90 µl and 10 µl of the internal standard (alpha-humulene, Sigma-Aldrich, final concentration 0.889 mg/L) were added. The samples were analyzed by GC-MS and GC-FID.

Chapter III: Identification, characterization and *in silico*-guided adaptation of class I redox systems for the production of functionalized diterpenoids

Christian Görner [1, +], Patrick Schrepfer [1, +], Veronika Redai [1], Frank Wallrapp [2], Bernhard Loll [3], Wolfgang Eisenreich [4], Martin Haslbeck [5] and Thomas Brück [1]

[1] Department for Industrial Biocatalysis, Technische Universität München, Lichtenbergstr. 4, 85748 Garching, Germany

[2] Department for Bioinformatics and Computational Biology, Technische Universität München, Boltzmannstraße 3, 85748 Garching, Germany

[3] Institute of Chemisty und Biochemistry, Freie Universität Berlin, Takustr. 6, 14195 Berlin, Germany

[4] Chair of Biochemistry, Department of Chemistry, Technische Universität München, Lichtenberg Str. 4, 85748 Garching

[5] Chair of Biotechnology ,Technische Universität München, Lichtenbergstr. 4, 85748 Garching, Germany

[+] These authors contributed equally to this work.

Author contributions:

CG, PS and TB conceived the study, planned and supervised experiments. CG, PS and VR conducted metabolic engineering of *E.coli* and enzyme mutagenesis experiments. CG, PS and MH conducted fermentation and CD measurements. CG, PS and WE conducted NMR data collection and solved structures of fermentation products. PS, FW and BL conducted *in silico* simulations and verified structures. TB, CG and PS finalized the manuscript.

1. Introduction

Fussicoccanes

Fusicoccanes are diterpenoids that are characterized by a 5-8-5 ring motif. They are found in bacteria, fungi and plants and show a broad diversity of biological activities, including bacteriostatic, fungicidal or tumor static.[92] These various effects are based on the diverse functional decoration of the 5-8-5 membered ring system. **Cyclooctatin** (Figure 31), a multifunctionalized tri-hydroxylated member of the fusicoccane family, which was isolated from the soil bacterium *Streptomyces melanosporofaciens* MI614-43F2 shows anti-inflammatory activity. [93,94] Contrary to other anti-inflammatory agents, *e.g.* aspirin® or ibuprofen® that target the cyclooxygenases I or II, **cyclooctatin** inhibits a lyosphospholipase that catalyzes the hydrolysis of the fatty acid ester bonds of lysophospholipids.[45,95]

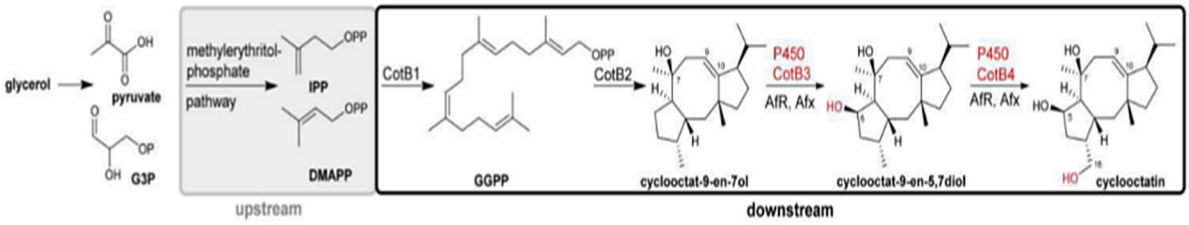


Figure 31: The cyclooctatin gene cluster

Figure 31: The cyclooctatin gene cluster comprises a GGPP synthase (CotB1), a diterpene synthase (CotB2) and two P450 type hydroxylases (CotB3, CotB4).[94]

Casbanes

The group of casbane diterpenoids have been isolated and described mainly from the Euphorbiaceae plant family and additionally have been found in some species of soft corals that belong to the genus *Sinularia*.[96-99] Casbane diterpenoids are characterized by a basic cembrane skeleton and the presence of a dimethyl-cyclopropyl moiety that is fused to the 14-membered cembrane diterpene ring as well as several chemical decorations including hydroxyl-, carbonyl- and epoxide groups. Members of this small group of natural products have been shown to

comprise cytotoxic activity against several tumor cell lines and moreover possess anti-inflammatory activity (Figure 32).[96,99]

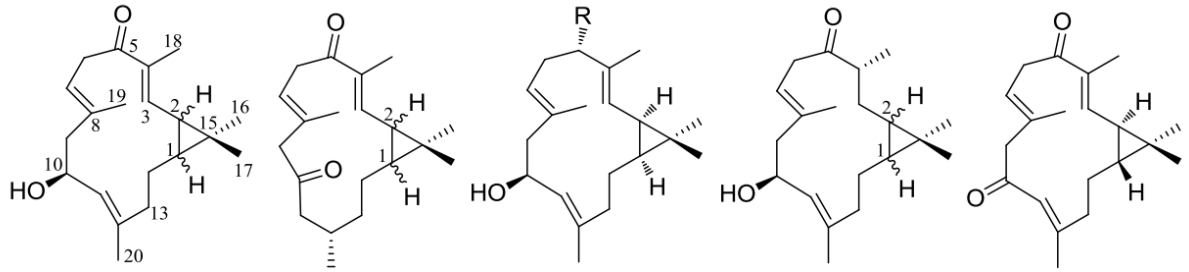


Figure 32: Bioactive casbane diterpenoids

Figure 32: Reprinted from (Yin et al., 2013).[96]

D*e novo* biosynthesis of functionalized diterpenoids, including taxanes, cembranes, casbanes and fusicoccanes is still challenging today. The first step upon functionalization is typically the hydroxylation of one or more nonactivated C-H bonds of the hydrocarbon precursors by pro- or eukaryotic CYP450 systems.[61] Selective oxygenation of nonactivated C-H bonds is one of the most challenging processes in synthetic chemistry. Therefore, CYP450s are attractive biocatalysts in biotechnology.[60,61,100] However, the native CYP450 systems that are responsible for hydroxylation in the respective natural producer are often unknown. In cases, where the CYP450 monoxygenase is known, its redox partner proteins often remained elusive, as they are typically not genetically coded in the same operon in microbial hosts and are widespread over the genome in eukaryotic polyploid organisms.[101] CYP450 monooxygenases display further a high rate of chemo-, regio- and stereoselectivity, which limits their application for functionalization of non-native macrocylic diterpene substrates.[17] Additionally, the usage of eukaryotic-derived membrane affiliated class II CYP450 systems in microbial hosts suffers from minor expression and solubility rates decreasing diterpenoid product titers in contrast to bacterial and mitochondrial class I and VIII CYP450 systems (Figure 10).[18,61] Currently, single hydroxylations of diterpenoids have been achieved using either *de novo* production systems[18] or whole cell catalysis experiments that depend on external substrate addition.[102,103] At present,

de novo production of or whole cell catalysis for multifunctionalized diterpenoids have not been achieved mainly due to the lack of appropriate electron transfer facilitating proteins.

Cyclooctatin biosynthesis

The **cycloocatin** biosynthesis cluster from *S.melanosporofaciens* has recently been shown to comprise four operon-coded enzymes, CotB1 (GGPP synthase), CotB2 (diterpene synthase) and the class I CYP450 monooxygenases CotB3 and CotB4 (Figure 31).[94] The native class I redox partner proteins, however, remained elusive. A subsequent functional reconstitution of the **cyclooctatin** biosynthesis cluster in the homologous host *Streptomyces albus* gave minor product titers whereby the specific redox partners of CotB3 and CotB4 again remained elusive.[94]

2. Aims of this study

The first part of the work was to identify suitable known and also currently unavailable redox systems that are able to activate the *Streptomyces*-derived CYP450 monooxygenases CotB3 and CotB4 in order to achieve production of **cyclooctatin**. Newly identified redox systems should subsequently be structurally characterized *in silico* and compared to the structurally well-characterized and widely applicable redox system from *Pseudomonas putida* (putidaredoxin reductase, PdR and putidaredoxin, Pdx, a 2Fe-2S ferredoxin).[104] A subsequent *in vivo* comparison of their capability to efficiently activate CotB3/4 using the established diterpene production system (see methods chapter II) should be the next part of the work. Finally, if differences between the different redox systems can be found, a methodology should be established to optimize redox systems for the application with non-native bacterial CYP450 systems.

3. Results

Bioinformatic analysis

A search for the native class I redox partners of CotB3/4 from *S. melanosporofaciens* is hampered due to the very limited sequence information. The genome of this soil bacterium is not sequenced up today and neither the transcriptome nor corresponding proteome data is available. Therefore, a multistep computational strategy was applied to all available *Streptomyces* genomes. That included a homology-based primary structure TBLASTN search (https://blast.ncbi.nlm.nih.gov) with the well-characterized class I *P. pudita* redox system[104], covering all available *Streptomyces* sequence databases.

On the primary structure level one putative class I ferredoxin reductase with 37% primary sequence identity and several ferredoxins (with up to 45% primary sequence identity) specifically clustered in the recently shotgun-sequenced genome of *Streptomyces afghaniensis* 772 (Figure 33a, b). This strain is known for the production of a polyketide antibiotic, Taitomycin[105], but not for the production of terpenes or terpenoids. However, as class I CYP450 systems are also used in the functionalization of polyketides, the identified putative class I redox enzymes are likely to operate with bacterial class I CYP450s.[61] The newly identified ferredoxins were subsequently subjected to an HHpred server analysis (http://toolkit.tuebingen.mpg.de/hhpred). The HHpred (Homology detection and structure prediction by HMM-HMM comparison) server tool is able to conduct homology-based secondary structure predictions with all available structures deposited in the current PDB database as template (http://www.rcsb.org/pdb). On the basis of available crystal- or NMR-derived structures the identified ferredoxins were shown to contain several 4Fe-4S ferredoxins, several so-called Rieske proteins and one 2Fe-2S ferredoxin. Rieske proteins are 2Fe-2S ferredoxins that are shown to be components of the mitochondrial cytochrome bc_1 or b_6f complexes operating in oxidative phosphorylation.[106] The found 4Fe-4S ferredoxins belonged according to their structural homology to the HCP_like family of iron-sulfur proteins. This superfamily is thought to play a role in nitrogen metabolism.[107]

a

Streptomyces afghaniensis 772 Seq1701, whole genome shotgun sequence

Sequence ID: gb|AOPY01001701.1| Length: 36012 Number of Matches: 1

Range 1: 26950 to 28173 GenBank Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
258 bits(658)	3e-75	Compositional matrix adjust.	165/416(40%)	240/416(57%)	14/416(3%)	-1

```
Query  4      ANDNVVIVGTGLAGVEVAFGLRASGWEGNIRLVGDATVIPHHLPPLSKAYLAGKATAESL  63
              AND  VIVG GLAG + A  LR  G++G + L+GD    P+  PPLSK YL GK   E +
Sbjct  28173  ANDACVIVGAGLAGAKAAQALREEGFDGPLVLIGDERERPYERPPLSKGYLTGKDAREQI  27994

Query  64     YLRTPDAYAAQNIQLLGGTQVTAINRDRQQVILSDGRALDYDRLVLATGGRPRPLPVASG  123
              Y+  P  YA  N+ +  G  VTA++   +++ L DG  + Y +L+L TG  PR LPV
Sbjct  27993  YVHPPQWYAEHNVDMRLGMAVTAVDPAAREMTLDDGSRVGYGKLLLTTGSAPRRLPVPGA  27814

Query  124    AVGKANNFRYLRTLEDAECIRRQLIADNRLVVIGGGYIGLEVAATAIKANMHVTLLDTAA  183
               + +     YLR +ED++ I+    + +R VVIG G+IGLE  A A  A + VT+L+ A
Sbjct  27813  GLERV---LYLRRVEDSDQIKEAFQSASRAVVIGAGWIGLETTAAARTAGVEVTVLEMAE  27643

Query  184    RVLERVTAPPVSAFYEHLHREAGVDIRTGTQVCGFEMSTDQQKVTAVLCEDGTRLPADLV  243
                L RV    V+  + +LHR+ GVD+R G QV   E++     V  V   DGTR+ AD+V
Sbjct  27642  LPLLRVLGREVAQLFANLHRDHGVDLRFGAQVA--EITGSGGAVDGVRLSDGTRIDADVV  27469

Query  244    IAGIGLIPNCELASAAGLQVDNGIVINEHMQTSDPLIMAVGDCARFHSQLYDRWVRIESV  303
              I G+G+ PN  LA  AGL+VDNGI ++E ++TS P I A GD A     L  + +R+E
Sbjct  27468  IVGVGIKPNIGLAQEAGLEVDNGIRVDERLRTSYPDIYAAGDVAHAFHPLLGKHIRVEHW  27289

Query  304    PNALEQARKIAAILCGKVPRDEAA----PWFWSDQYEIGLKMVGLSE--GYDRIIVRGSL  357
               NAL Q +  A  + G   R++A     P+F++DQY++G++  G  E  GYD+++ RG +
Sbjct  27288  ANALNQPQIAAKAMLG---REDAVYDRIPYFFTDQYDLGMEYAGYVEPGGYDQVVFRGDV  27118

Query  358    AQPDFSVFYLQGDRVLAVDTVNRPVEFNQSKQIITDRLPVEPNLLGDESVPLKEII  413
              A  +F  F+L G+RVLA   VN     +Q + ++     V+  +L D  VPL  ++
Sbjct  27117  AGREFIAFWLAGNRVLAGMNVNIWDVNDQLQTLVRTAQTVDIPMLTDPQVPLGSLL  26950
```

b

Streptomyces afghaniensis 772 Seq1673, whole genome shotgun sequence

Sequence ID: gb|AOPY01001673.1| Length: 42046 Number of Matches: 1

Range 1: 14221 to 14538 GenBank Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
85.1 bits(209)	3e-20	Compositional matrix adjust.	49/106(46%)	64/106(60%)	0/106(0%)	+1

```
Query  1      MSKVVYVSHDGTRRELDVADGVSLMQAAVSNGIYDIVGDCGGSASCATCHVYVNEAFTDK  60
              M KV YVS  G  R +D   G S+MQ AV NG+  I G+CGG  SCATCHV+V+EA  D+
Sbjct  14221  MPKVTYVSDAGEVRVVDGLVGDSVMQTAVRNGVPGITGECGGVLSCATCHVFVDEADLDR  14400

Query  61     VPAANEREIGMLECVTAELKPNSRLCCQIIMTPELDGIVVDVPDRQ  106
              +   +  E  ML+    +  PNSRL CQI ++ EL  + V  P+ Q
Sbjct  14401  LEPVSGLEDEMLDGTVVDRCPNSRLSCQIKLSEELGDLRVTTPEAQ  14538
```

Figure 33: TBLASTN search with PdR and Pdx as template

Figure 33**a-b**: TBLASTN search with **a:** PdR (Query) and **b:** Pdx (Query) as template in the recently shotgun sequenced genome of *Streptomyes afghaniensis 772.* The found homologous contigs (Sbject) are subjected to a secondary structure prediction using the HHpred server tool

The 2Fe-2S ferredoxin was consequently subjected to the MODELLER (https://salilab.org/modeller/) package of the HHpred server tool using the homologues that comprise an e-value lower than 9 x 10^{-26} as template (Figure 34a). The homologues found by the HHpred server tool included several known and well-characterized 2Fe-2S ferredoxins (putidaredoxin[104], the mitochondrial adrenodoxin[108] and terpredoxin[109] from *Pseudomonas sp.*). The constructed homology model was further subjected to optimization, rotamer fine-tuning and energy minimization using the self-parametrizing YASARA02 force field of YASARA structure (methods chapter II and III). The structure refinement ended with a 1 ns MD simulation using a neutralized TIP03 water box. A structural alignment of this refined ferredoxin with the crystal structure of the putidaredoxin[110] (Pdx) resulted in a root mean square deviation (RMSD) of the Cα-atoms of 0.25 Å over 106 aligned amino acids with 27.45% primary sequence identity (Figure 35b), arguing for a shared secondary and tertiary structure and thus similar function.

a

No	Hit	Prob	E-value	P-value	Score	SS	Cols	Query HMM	Template HMM
1	4ltu_A Ferredoxin; 2Fe-2S sing	100.0	9.7E-32	2.7E-36	166.8	11.1	105	1-106	1-106 (107)
2	3lxf_A Ferredoxin; iron, iron-	100.0	1.8E-31	5E-36	164.9	12.0	104	2-106	1-104 (104)
3	2y5c_A Adrenodoxin-like protei	100.0	1.4E-31	4.1E-36	166.4	9.7	105	1-107	5-109 (109)
4	1xlq_A Putidaredoxin, PDX; [2F	100.0	6.8E-31	1.9E-35	162.8	12.1	104	2-106	1-105 (106)
5	2wlb_A ETP1-FD, electron trans	100.0	3.6E-31	1E-35	163.1	10.2	101	1-103	3-103 (103)
6	3hui_A Ferredoxin; cytochrome	100.0	4.1E-31	1.2E-35	168.4	9.8	105	1-106	21-125 (126)
7	1uwm_A Ferredoxin VI, FDVI; el	100.0	1.2E-30	3.5E-35	161.6	10.5	104	2-106	1-105 (106)
8	2mj3_A Iron-sulfur cluster bin	100.0	4.1E-31	1.1E-35	166.1	7.5	105	1-107	1-108 (116)
9	2bt6_A Adrenodoxin 1; rutheniu	100.0	1E-30	2.8E-35	162.4	9.0	101	1-103	6-108 (108)
10	2mjd_A Adrenodoxin homolog, mi	100.0	1.5E-30	4.2E-35	163.3	9.5	104	1-106	3-106 (115)
11	1b9r_A Protein (terpredoxin);	100.0	2.5E-30	7E-35	159.9	6.6	104	2-106	1-105 (105)
12	3n9z_C Adrenodoxin; cytochrome	100.0	3.5E-30	9.8E-35	163.3	1.0	105	1-107	5-111 (123)
13	3ah7_A [2Fe-2S]ferredoxin; [2F	100.0	7.4E-28	2.1E-32	150.5	10.4	98	1-107	1-105 (113)
14	1l5p_A Ferredoxin; [2Fe-2S] cl	99.9	3.2E-27	9.1E-32	143.1	9.3	93	2-102	1-93 (93)
15	1i7h_A Ferredoxin; 2Fe-2S,elec	99.9	1E-26	2.8E-31	145.1	10.1	99	1-107	1-105 (111)
16	3wcq_A Ferredoxin; 2Fe-2S clus	99.9	2.7E-26	7.7E-31	139.9	4.0	94	1-107	1-95 (97)

b

No	Hit	Prob	E-value	P-value	Score	SS	Cols	Query HMM	Template HMM
1	3lxd_A FAD-dependent pyridine	100.0	2.3E-61	6.5E-66	459.3	50.3	408	1-411	6-415 (415)
2	1q1r_A Putidaredoxin reductase	100.0	5.2E-61	1.5E-65	458.6	49.1	407	1-410	1-412 (431)
3	3fg2_P Putative rubredoxin red	100.0	1.4E-60	3.9E-65	452.4	50.8	404	4-411	1-404 (404)
4	2gqw_A Ferredoxin reductase; f	100.0	3.2E-59	9E-64	443.4	46.8	396	1-410	1-403 (408)
5	3ef6_A Toluene 1,2-dioxygenase	100.0	9.2E-59	2.6E-63	440.9	47.3	403	3-411	1-403 (410)
6	3klj_A NAD(FAD)-dependent dehy	100.0	3.6E-54	1E-58	405.1	36.0	377	2-412	7-384 (385)
7	1xhc_A NADH oxidase /nitrite r	100.0	1.1E-51	3E-56	385.8	34.3	355	5-390	9-364 (367)
8	3iwa_A FAD-dependent pyridine	100.0	3.6E-51	1E-55	395.4	39.0	399	1-406	1-439 (472)
9	4fx9_A Coenzyme A disulfide re	100.0	2.6E-49	7.3E-54	380.2	44.2	406	1-415	1-448 (453)
10	2cdu_A NADPH oxidase; flavoenz	100.0	9.1E-50	2.6E-54	383.9	40.2	396	5-414	1-428 (452)
11	3oc4_A Oxidoreductase, pyridin	100.0	3.5E-50	9.8E-55	386.4	36.6	397	3-414	1-423 (452)
12	1nhp_A NADH peroxidase; oxidor	100.0	1.3E-49	3.7E-54	382.2	38.1	397	5-414	1-427 (447)
13	3cgb_A Pyridine nucleotide-dis	100.0	2.3E-49	6.6E-54	383.5	38.8	389	2-396	34-453 (480)
14	1m6i_A Programmed cell death p	100.0	4.4E-50	1.2E-54	389.3	31.9	402	1-408	8-492 (493)
15	3ntd_A FAD-dependent pyridine	100.0	2.5E-48	7.1E-53	384.1	43.1	399	4-406	1-446 (565)
16	2v3a_A Rubredoxin reductase; a	100.0	2.2E-48	6.2E-53	366.6	40.6	375	1-387	1-379 (384)

Figure 34: Primary and secondary structure analysis

Figure 34**a-b**: Primary and secondary structure analysis of **a:** the newly identified 2Fe-2S ferredoxin and **b:** NADH-dependent class I ferredoxin reductase (**b**) from *S. afghaniensis* 772 using the HHpred server

The newly identified NADH-dependent FAD-containing class I ferredoxin reductase was also subjected to the described refinement steps. The model was initially build using homolog structures comprising an e-value lower than 2.2 x 10^{-48} as template and thus included the putidaredoxin reductase[104] (PdR) (Figure 34b). The structural alignment of the refined ferredoxin reductase with the crystal structure of PdR[111] resulted in a RMSD of 0.3 Å over 404 aligned amino acids with 22.05% primary sequence identity (Figure 35a).

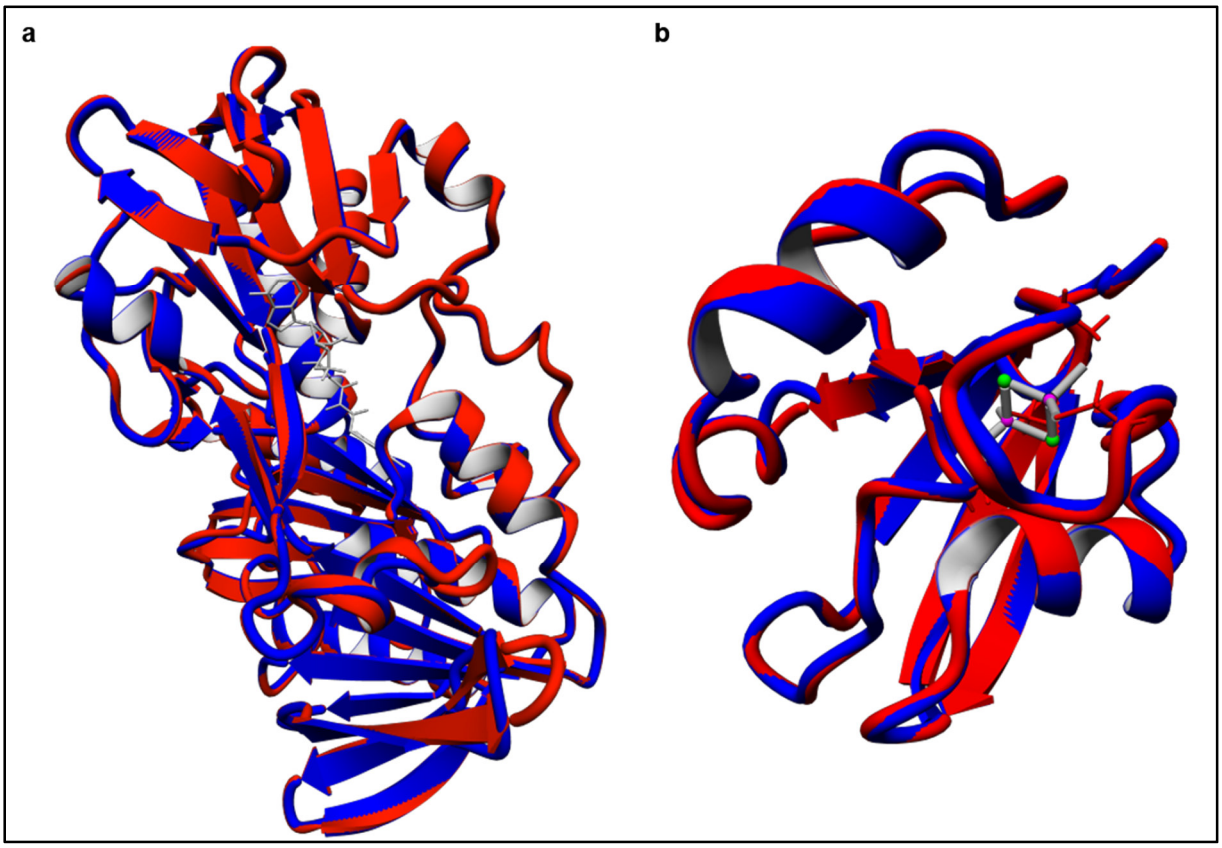


Figure 35: Structural alignment of PdR with AfR and Pdx with Afx

Figure 35**a**: Structural alignment of PdR (blue, FAD cofactor in grey) with AfR (red). **b**: Structural alignment of Pdx (blue, 2Fe-2S cluster in grey) with Afx (red).

The newly identified class I NADH-dependent FAD-containing ferredoxin reductase from *S. afghaniensis* 772 was termed afghanoredoxin reductase (AfR), its corresponding 2Fe-2S ferredoxin was termed afghanoredoxin (Afx). The generated models of AfR and Afx were subjected to a detailed quality assessment to ensure the highest stereochemical quality in the absence of a crystal or NMR structure (Table 6 and methods chapter III, QMEANZ-score quality assessment, see Appendix Figure 32)

	Most favored regions in RAMACHANDRAN Plot [%]	G-factor/ dihedral angles [average score]	G-factor/Main-chain covalent forces [average score]
CotB3	92.5	-0.20	-0.29
CotB4	90.4	-0.27	-0.32
AfR	92.8	-0.01	-0.13
Afx	94.1	-0.02	-0.09

Table 6: PROCHECK results of CotB3/4, AfR and Afx

Table 6*:* PROCHECK results of CotB3/4, AfR and Afx. For detailed information about quality assessments, see methods, chapter III.

Biotechnological production of cyclooctatin and its precursors

The native producer of **cyclooctatin** *S.melanosporofaciens* only produces minor amounts of this tri-hydroxylated compound (0.35 mg/L).[94] A homologous reconstitution of the found **cyclooctatin** biosynthesis cluster in *Streptomyces* albus is further shown to produce only trace amounts.[94]

To test the newly identified redox system (AfR/Afx) and to establish an efficient production of **cyclooctatin**, AfR/Afx and CotB3/4 were introduced as bi-cistronic cassettes into the multiple cloning sites I and II of the pACYC-duet-1 vector, respectively (AfR/Afx strain, methods chapter III). TXS was changed to CotB2 on the pET-duet-1 vector (methods chapter II and III). This genomic diterpenoid production system, consisting of four compatible vectors with 11 enzymes was introduced into *E. coli* BL21 (DE3). To compare this system with the *P. putida* redox system, an equivalent strain harboring PdR/Pdx was engineered (PdR/Pdx strain) by applying the same genetic and metabolic setting.

Three day shake flask cultures (100 mL) of the engineered AfR/Afx and PdR/Pdx strains in triplicates with glycerol as sole carbon source provided several milligrams of **cyclooctatin**,

cyclotat-9-en-7-ol and **cyclooctat-9-en-5,7-diol** (Figure 31, Table 7, Figure 36). Structural analysis by GC/MS- (TMS-derivatives, methods chapter III) and HR-LC/MS spectrometry as well as 1D-/2D-NMR and CD-spectroscopy confirmed their chemical and structural identity in comparison with the published structures[94] (Figure 37, Table 8, 9 and Appendix Figure 18-23). The structural characterization of **cylcooctat-9-en-7-ol** out of this metabolic setting is described elsewhere.[93] This demonstrated that the *in silico*-identified AfR/Afx redox system was able to successfully activate the non-native CYP450s CotB3/4. The application of the *P. putida* redox system in the same setting demonstrated also the successful activation of non-native CotB3/4, however, analysis of the triplicate shake flask cultures showed a drastically altered product distribution. The engineered PdR/Pdx strain provided the tri-hydroxylated **cyclooca-tion** in minor amounts, but compared to the AfR/Afx strain, the chemical precursors of **cy-clooctatin** in higher amounts as shown in Table 7 and Figure 36.

This indicates either a different activation mechanism of Afx vs. Pdx with respect to CotB3/4, despite almost similar secondary and tertiary structure, or differing cytoplasmic availabilities of Afx and Pdx due to different expression rates and/or different solubility characteristics.

	Cyclooctatin [mg/L]	**Cyclooctat-9-en-5,7-diol [mg/L]**	**Cyclooctat-9-en-7-ol [mg/L]**
Cell Pellet AfR/afx strain	8.63 ± 0.78	5.67 ± 1.02	4.33 ± 0.54
Cell Pellet PdR/Pdx strain	1.28 ± 0.26	5.33 ± 0.57	11.39 ± 0.53
Supernatant AfR/Afx strain	7.48 ± 0.41	0.88 ± 0.14	4.47 ± 0.04
Supernatant PdR/Pdx strain	2.43 ± 0.49	1.58 ± 0.06	5.18 ± 0.32

Table 7: Comparison of the Redox-Systems

Table 7: Comparison of the redox systems AfR/Afx from *S. afghaniensis* and PdR/Pdx from *P. putida* in 100 mL shake flask cultivation for 3 days with respect to *de novo* production of **cyclooctatin** and its metabolic dihydroxylated and mono-hydroxylated precursors **cyclooctat-9-en-5,7-diol** and **cyclooctat-9-en-7-ol**, respectively.

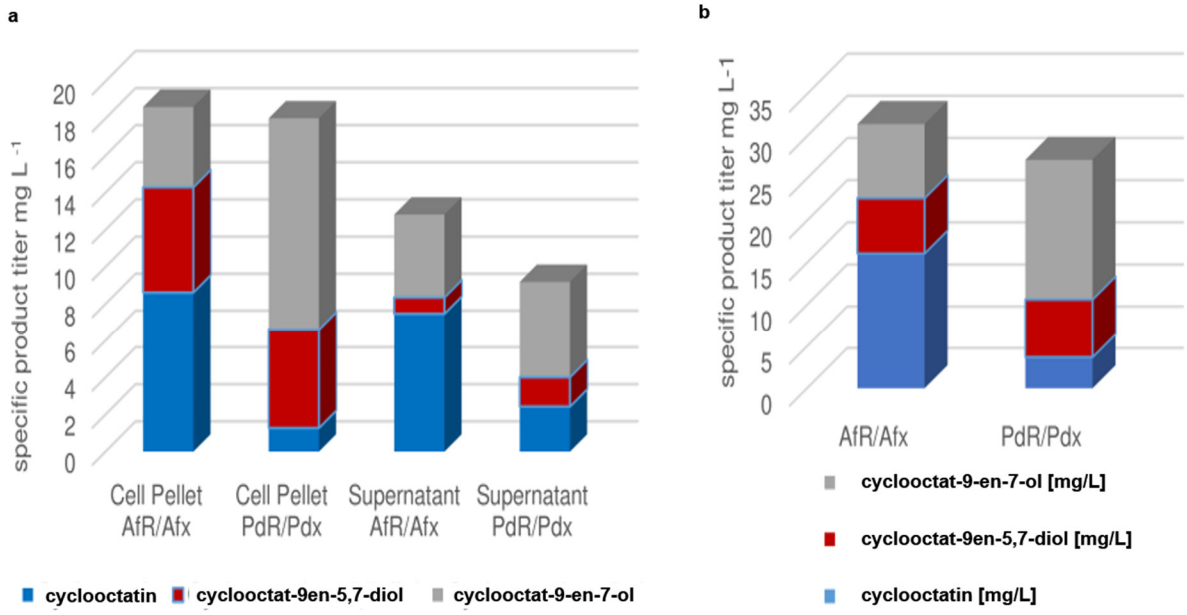


Figure 36: Graphical illustration of Table 7

Figure 36**a**: Graphical illustration of Table 7. **b**: Combined titer out of the AfR/Afx and PdR/Pdx strains.

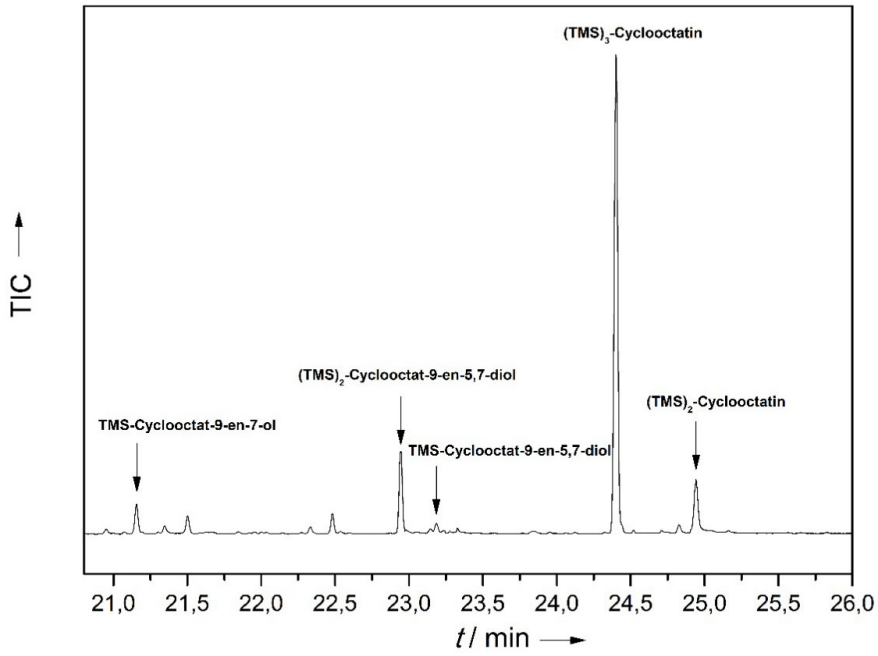


Figure 37: GC-chromatogram of the 100 mL cell extract from the AfR/Afx strain

Figure 37: GC-MS spectra from cell pellet extract of 3 day shake flask cultures (100ml) of TMS-derivatives of the **cyclooctatin** biosynthesis (TMS_2: di-trimethylsilylated, TMS_3: tri-trimethylsilylated).

NMR Spectral Data for **cyclooctat-9-en-5,7-diol** [a]			
			cyclooctat-9-en-5,7-diol [a]
#	**δc (ppm)**	**δ_H (ppm), *J* (Hz)**	**δc (ppm)**
1	44,8	1,56 (br d), 1,05 (t, J = 13,0)	44,7
2	36,3	2,36 (m)	36,4
3	35,5	2,55 (m)	35,6
4	39,9	1,63 (dd), 1,38 dd)	44,0
5	75,5	4,43 (br dd J = 4,8, J = 3,4)	75,5
6	56,9	1,94 (t)	57,0
7	77,2		77,3
8	41,8	2,68 (br t), 1,95 (m)	41,8
9	117,8	5,22 (ddd, J = 10,3, 7,6, 2,0)	117,9
10	153,4		153,3
11	44,8		44,7
12	45,5	1,59 (m), 1,37 (m)	45,6
13	23,2	1,52 (m), 1,35 (m)	23,3
14	53,9	2,27 (m)	53,9
15	29	1,79 (m)	29,0
16	17,3	0,77 (d)	17,4
17	22,1	0,93 (d)	22,2

18	15,3	0,94 (d)	15,4
19	26,6	1,35 (s)	26,6
20	25,0	1,20 (s)	24,9

Table 8: NMR spectral data for cyclooctat-9-en-5,7-diol

Table 8: The chemical shifts of **cyclooctat-9-en-5,7-diol** from extracts of the AfR/Afx strain cultures are derived from 13C- (Appendix Figure 19) and 1H spectra (Appendix Figure 18). Spectra of **cyclooctat-9-en-5,7-diol** have been measured in $CDCl_3$ ([a]). The right column depicts the published 13C chemical shifts.[94]

NMR Spectral Data for **cyclooctatin** [b]			
			cyclooctatin [b]
#	***δc* (ppm)**	***δ*$_H$ (ppm), *J* (Hz)**	***δc* (ppm)**
1	45,5	1,67 (br d), 1,19 (t, J = 12,8)	45,6
2	35,9	2,56 (m)	35,8
3	44,8	2,60 (m)	44,9
4	39,8	1,70 (br dd), 1,38 dt)	39,7
5	75,6	4,43 (br dd J = 4,8, J = 3,4)	75,7
6	58,0	1,97 (t)	58,0
7	78,3		78,4
8	42,1	2,71 (br t), 1,92 (m)	42,2
9	119,1	5,28 (ddd, J = 10,3, 7,4, 2,0)	119,1
10	154,5		154,5
11	45,8		45,9
12	46,5	1,59 (m), 1,41 (m)	46,6
13	24,2	1,55 (m), 1,37 (m)	24,3
14	55,1	2,29 (m)	55,1
15	30,1	1,82 (m)	30,2
16	17,7	0,79 (d)	17,8
17	22,5	0,97 (d)	22,5
18	63,4	3,66 (dd J= 10,8, 7,4), 3,55 (dd, J = 10,8, 6,8)	63,4
19	26,6	1,34 (br s)	26,7
20	25,2	1,25 (s)	25,2

Table 9: NMR spectral data for cyclooctatin

Table 9: The chemical shifts of **cyclooctatin** from extracts of the AfR/Afx strain cultures are derived from 13C- (Appendix Figure 21) and 1H spectra (Appendix Figure 22). Spectra of **cyclooctatin** have been measured in CD_3OD ([b]). The right column depicts the published 13C chemical shifts.[94]

Biotechnological production of sinularcasbane D

As the substrate specificity of the *Streptomyces*-derived CYP450 monooxygenases CotB3/4 have not been elucidated and their efficient *in vivo* activation have been established, their capability to hydroxylate alternative diterpene macrocycles has consequently been tested. Thereby, CotB2 was exchanged with other diterpene synthases and the hydroxylation was assayed using the four plasmid system (methods chapter III). *R*-cembrene A[93] (CotB2 mutant), (-)-casbene (casbene synthase from *Jatropha curcas*[112]), (1R,3E,7E,11S,12S)-3,7,18-dolabellatriene[45] (CotB2 mutant) and taxa-4,5-11,12-diene[46] (TXS) have been tested as alternative hydroxylation substrates. CotB3 was able to hydroxylate 20% of (-)-casbene yielding 1.4 mg/L of a single mono-hydroxylated product, whereas CotB4 do not display any substrate promiscuity. The new compound was identified as **sinularcasbane D** by GC/MS and 1D-/2D-NMR analysis (Figure 38a, b, Table 10 Appendix Figure 24-29). This is the first report of a *de novo*, stereoselective formation of **sinularcasbane D**, which has first been described from the soft coral *Sinularia sp.* For the structural identification of (-)-casbene, see chapter IV.

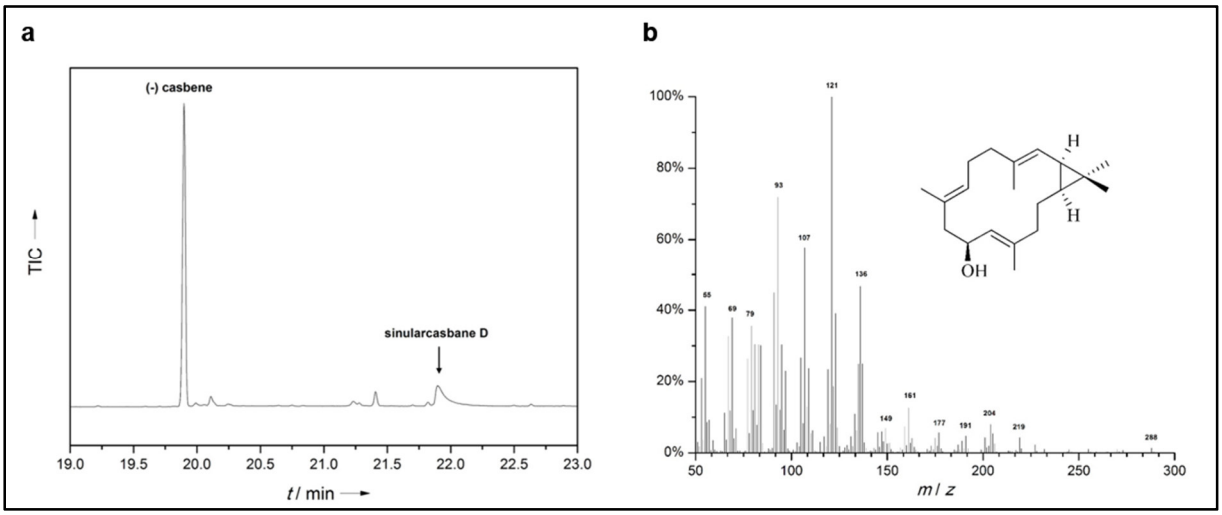


Figure 38: GC chromatogram and EI-MS spectrum of sinularcasbane D

Figure 38: GC chromatogram and EI-MS spectrum of sinularcasbane D from a cell pellet extract of a 72h shake flask culture containing the four plasmid system (with the casbene synthase and CotB3/AfR/Afx, methods chapter III)

NMR Spectral Data for **sinularcasbane D** [a]			
			sinularcasbane D [a]
#	**δc (ppm)**	**δ_H (ppm), *J* (Hz)**	**δc (ppm)**
1	30,8	0,65 (ddd, J = 10,5, 8,7, 1,7)	30,7
2	25,9	1,21 (t = 8,5)	25,9
3	121,4	4,84 (d)	121,2
4	135,9		135,8
5	39,5	2,19 (m), 2,06 (m)	39,4
6	25,1	2,19 (m), 2,05 (m)	25,0
7	127,1	4,87 (ddd, J = 6,7, 4,4, 1,2)	127,0
8	130,9		130,8
9	48,1	2,35 (m), 2,08 (m)	48,0
10	67,0	4,52 (ddd, J = 10,7, 9,1, 3,2)	66,9
11	127	4,98 (dq)	127,0
12	140,2		140,1
13	39,6	2,24 (m), 1,81 (m)	39,5
14	24,0	1,68 (m), 1,13 (m)	23,9
15	20,1		20,0
16	29,0	1,07 (s)	28,9
17	15,7	0,96 (s)	15,6
18	16,0	1,62 (s)	16,0
19	17,2	1,59 (s)	17,1
20	18,7	1,66 (s)	18,6

Table 10: NMR spectral data of sinularcasbane D

Table 10: The chemical shifts of **sinularcasbane D** from extracts of the respective strain cultures are derived from 13C- (Appendix Figure 25) and 1H spectra (Appendix Figure 24). Spectra of **sinularcasbane D** have been measured in $CDCl_3$ ([a]). The right column depicts the published 13C chemical shifts.[96] For additional 2D NMR spectra, see Appendix Figure 26-29.

In silico characterization of the Afx-mediated CYP450 activation

The structural basis for the redox partner recognition of the model class I CYP450 system from *P. putida* was recently elucidated.[110] Crystal structures of a reduced Pdx in complex with its

cognate CYP450 (P450cam) in different conformations as well as earlier NMR and mutagenesis studies established the molecular pattern of binding and electron transfer processes between both redox partners.[113] These results demonstrated that the interacting residues and the electron transfer (ET) processes are different between the reduced, physiologically relevant form of Pdx and the ferric and ferrous dioxygen-bound (oxy-P450cam) forms of P450cam, respectively. However, in both reductive steps Pdx-D38 and -W106 were found to be the key players, albeit with different functions (Figure 39). During ET1 both residues are important for the correct orientation of Pdx with respect to P450cam in order to allow the successful electron transfer, whereas both residues play an underrated role in electron transfer. In ET2, the rate-limiting step of the reaction cycle, Pdx-D38 and -W106 are the key residues that regulate binding and electron transfer between the reduced Pdx and oxy-P450cam. During ET2, Pdx exerts its effector role that makes it highly specific for P450cam. It shifts P450cam towards an open conformation that is important for the proton-coupled electron transfer in ET2, cleavage of the O-O bond and substrate release.

To get an insight into the binding and electron transfer processes between Afx and CotB3/4 as well as Pdx and CotB3/4, *in silico* models of CotB3 and CotB4 were generated (Figure 40a, b, methods chapter III, Appendix Figure 30, 31). For detailed structural quality assessments, see Table 6 and Appendix Figure 32. Subsequently we used both for the modelling of the CotB3/4•Afx (Figure 40a, b) and CotB3/4•Pdx complexes (Figure 41a, b) with the P450cam•Pdx complex (PDB-ID: 4JX1) as a template (methods chapter III). The *P. putida* template complex depicts P450cam in its open state during the second rate-limiting ET step, bound by camphor, its cognate substrate and 5-exo-hydroxycamphor, its cognate product. In this template complex Pdx-D38 ion pairs with P450cam-R112 that also ion pairs with one heme B propionate, Pdx-W106 forms non-bonded interactions with P450cam-A113 and an H-bond with P450cam-N116 (Figure 39). Further, P450cam-R109 ion pairs with the carboxyl group of

Pdx-W106. Formation of this H-bond network is shown to be required to ensure correct orientation of both binding partner in ET2 and also responsible for the second electron transfer step (Figure 39).[104,110]

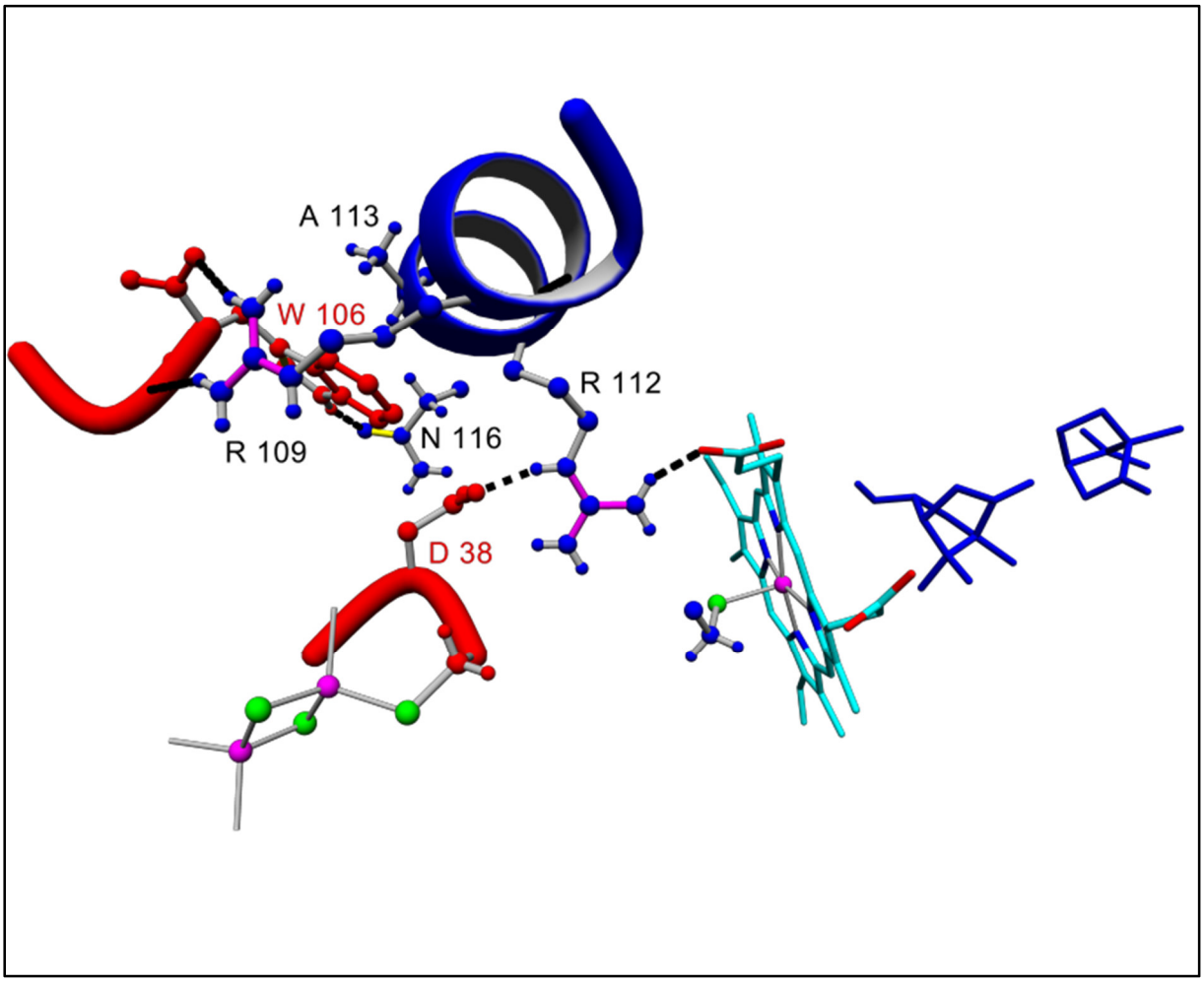


Figure 39: P450cam•Pdx complex

Figure 39: Inner view of the P450cam•Pdx (PDB: 4JX1) complex and interacting important amino acid residues, where P450cam depicts its oxygenated form (blue) and Pdx (red) is reduced. The heme b is depicted in cyan (Fe in magenta, the sulfur of the cysteinate ligand in green), the camphor substrates of the P450cam crystal structure are depicted in blue and the Pdx 2Fe-2S cluster in magenta-green. Coordinative bonding is colored in grey and important hydrogen bonds are colored in broken black lines. For detailed description of important interactions, see text.

Tertiary structural alignments of CotB3/4•Afx with P450cam•Pdx established the structural analogy of Afx-E39 with Pdx-D38 and Afx-E107 with Pdx-W106. The models of the CotB3/4•Afx complexes suggest essential hydrogen bonds between CotB3/4-R110/-R111•Afx-E107, CotB3/4-R113/-R114•Afx-E39 and CotB3/4-R114/-R115•Afx-E107 (Figure 40c, d)

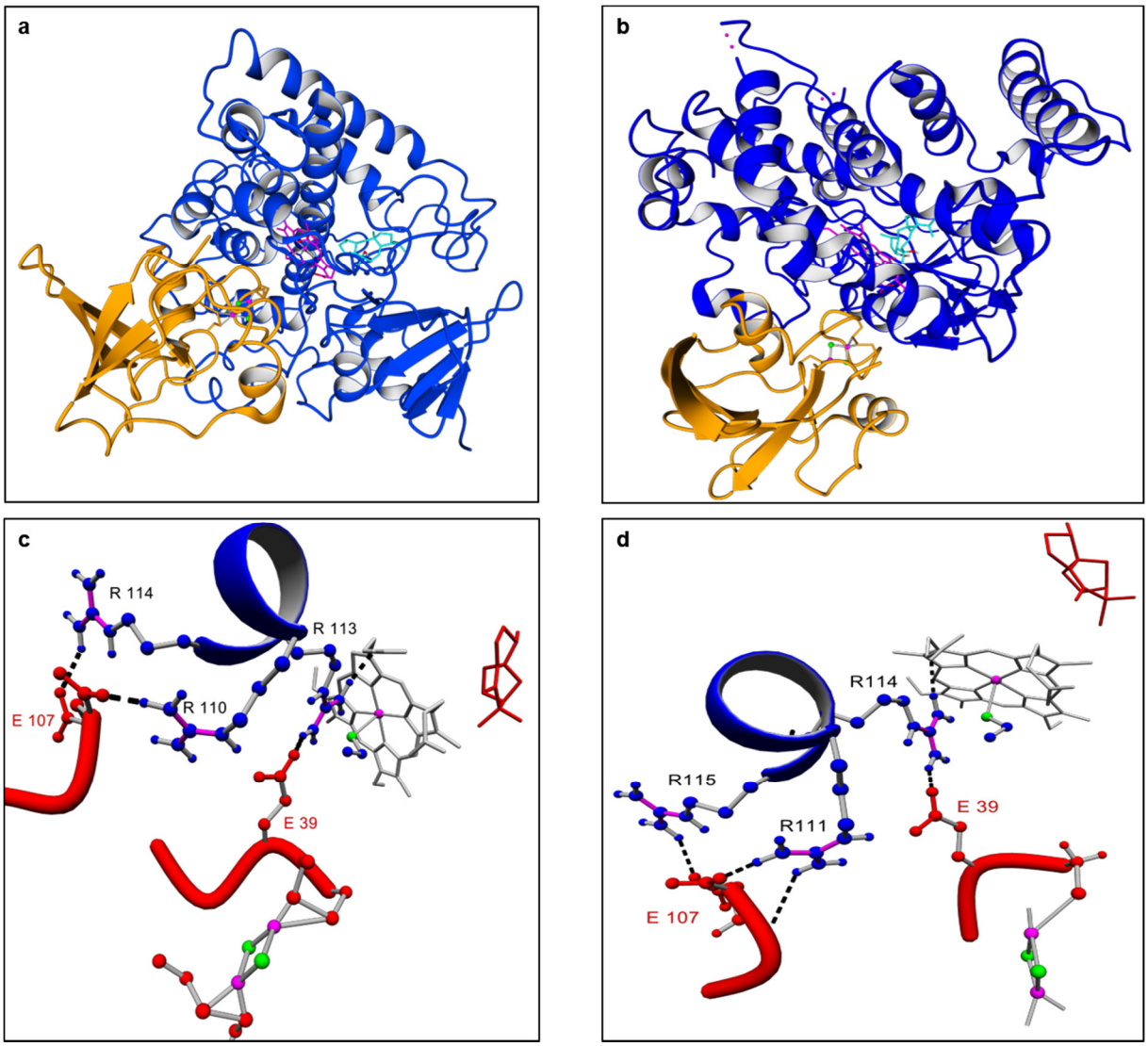


Figure 40: Structural representation of the CotB3•Afx and CotB4•Afx complexes

Figure 40**a**: Structural overview of the CotB3•Afx complex (CotB3 colored in blue, heme B in magenta, docked **cyclooctat-9en-7-ol** in cyan, 2Fe-2S cluster of Afx in magenta-green) , **b**: Structural overview of the CotB4•Afx

complex (CotB4 colored in blue, heme B in magenta, docked **cyclooctat-9en-5,7-diol** in cyan, 2Fe-2S cluster of Afx in magenta-green). **c**: Inner view of the CotB3•Afx complex (CotB3 colored in blue, heme B in grey, Fe in magenta, sulfur of the cysteinate ligand in green, 2Fe-2S cluster of Afx in magenta-green, **cyclooctat-9-en-7-ol** in red), **d**: Inner view of the CotB4•Afx complex (CotB4 colored in blue, heme B in grey, Fe in magenta, sulfur of the cysteinate ligand in green, 2Fe-2S cluster of Afx in magenta-green, **cyclooctat-9-en-5,7-diol** in red). 40**a-b**: Coordinating bonds are colored in grey and important hydrogen bonds between identified essential amino acid residues are depicted as broken black line. For detailed information of the identified interactions, see text.

The interactions of CotB3/4-R110/-R111 with Afx-E107 and CotB3/4-R113/-R114 with Afx-E39 seem to be analogous to the P450cam-R110•Pdx-W106 and P450cam-R113•Pdx-D38 interactions, respectively. However, in contrast to the hydrophobic interaction between P450cam-A113 and Pdx-W106 in the P450cam•Pdx reference, a polar hydrogen bond-based interaction between CotB3/4-R114/-R115 and Afx-E107 could be detected *in silico* in the CotB3/4•Afx complex (Figure 40c, d).

Models of the CotB3/4•Pdx complexes show the ion pair-based interaction between Pdx-D38 and CotB3/4-R113/-R114 that is connected to the heme B propionate via a hydrogen bond (Figure 41a, b). The ion pairing between the carboxyl group of Pdx-W106 and the guanidinium group CotB3/4-R110/-R111 in the CotB3/4•Pdx complexes coincides further with the P450cam•Pdx reference (Figure 39). However, the CotB3/4•Pdx complexes do not show a productive interaction between the guanidinium group of CotB3/4-R114/-R115 and the hydrophobic side chain of Pdx-W106 (Figure 41a, b)

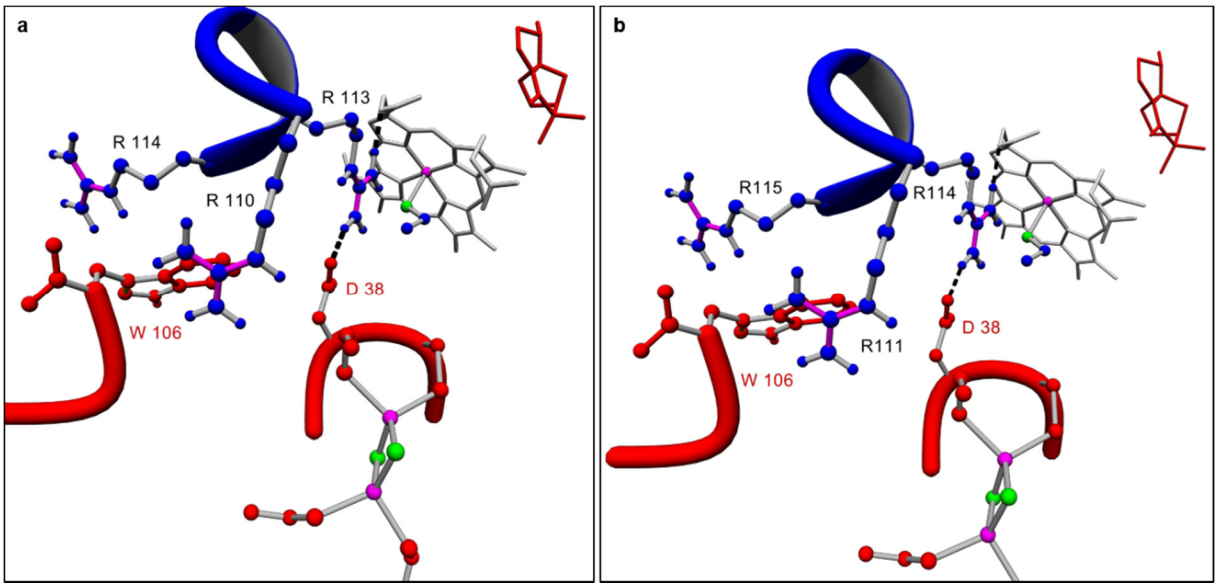


Figure 41: Inner view of the CotB3•Pdx complex, b Inner view of the CotB4•Pdx complex.

Figure 41**a-b**: Coloring scheme is according to Figure 40. For more detailed information see text.

In silico comparisons between CotB3/4•Afx and CotB3/4•Pdx complexes with the P450cam•Pdx reference complex implied the importance of Afx-E39 and Afx-E107 in redox partner recognition, correct orientation of the partners and electron transfer processes in the CotB3/4•Afx complexes. Specifically, these findings argue for the necessity of an acidic polar amino acid at position 39 and 107 in Afx to allow an efficient interaction with CotB3/4.

In vivo characterization of the Afx- and Pdx-mediated activation of CotB3/4

To elucidate the activation efficiency of the redox systems from *S. afghaniensis* and *P. putida* for CotB3 and CotB4, whole cell catalysis experiments were performed. An *in vitro* assay using purified enzymes could not be performed due to significant problems with Afx purification under aerobe conditions and its presumable weak *in vitro* interaction with AfR, which is in line with previous reports for Pdx and PdR.[111] As the four plasmid setting, described above, increases the complexity of an *in vivo* comparison between the two redox systems, *E.coli* Bl21 (DE3)

strains harboring only the pACYC-duet-1 vectors containing either *cotB3* or *cotB4* combined with either *afR-afx* or *pdR-pdx* gene cassettes were generated. Additionally, single and double key residue substitutions of Pdx were created in this system, where Pdx-D38, -W106 and -D38/W106 were substituted with its analogous Afx counterparts (Pdx-D38E, -W106E and –D38E/W106E). These substitutions should be able to verify the *in silico*-derived assumptions with respect to the molecular interactions between the ferredoxins and CotB3/4.

The whole cell catalysis experiments were executed as quintuplicates for 48 hours (methods chapter III). Production of **cyclooctat-9-en-5,7-diol** or **cyclooctatin** was monitored after *in vitro* addition of the corresponding metabolic precursors. In contrast to the AfR•Afx system the PdR•Pdx redox system was not able to produce **cyclooctat-9-en-5,7-diol** with CotB3 (Figure 42a) and only minor amounts of **cyclooctatin** with CotB4 could be detected (Figure 42b). The Pdx-W106E mutant led to a significant increase of cyclooctat-9-en-5,7-diol compared to the AfR•Afx positive control system (Fig. 42a). A slight increase in **cyclooctatin** production could be observed for the Pdx-W106E mutation compared to the PdR•Pdx negative control system (Figure 42b). Substitutions of Pdx-D38E and Pdx-D38E/W106E, however, did not lead to an increase in **cyclooctat-9-en-5,7-diol** and **cyclooctatin** titer compared to the PdR•Pdx negative control system (Figure 42a, b).

These results argue for fundamental differences for the AfR•Afx and PdR•Pdx redox systems with respect to the molecular basis of their interaction with CotB3 and CotB4, as predicted by the *in silico* characterization, *vide supra*.

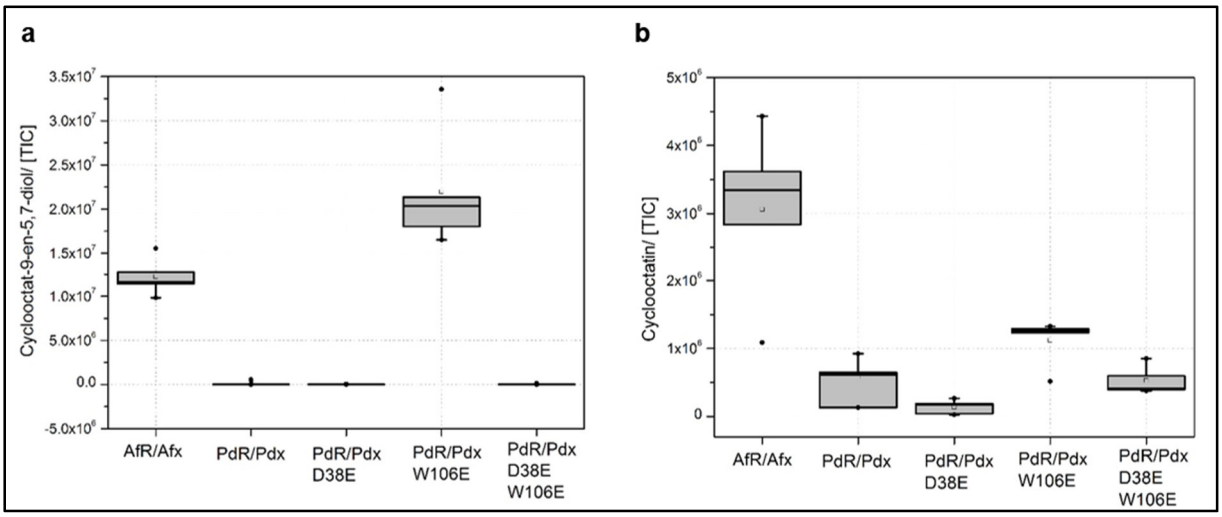


Figure 42: Whole cell hydroxylation experiments

Figure 42**a**: Graphical boxplot representation of the quintuplicate whole cell hydroxylation experiments using *E. coli* harboring *CotB3* combined with either *AfR/Afx* or *PdR/Pdx* or *PdR/Pdx-D38E* or *PdR/Pdx-W106E* or *PdR/Pdx-D38E/W106E*. The diagram y axis represents the relative total ion current of the TMS-derivative of **cycloocat-9-en-5,7-diol** upon GC-MS analysis out of the respective whole cell extracts. GC-FID measurement and representation in mg/L was not possible as in the 10 ml culture used for the whole cell experiments the quantity of produced hydroxylated compounds was too less for an absolute quantity measurement with α-humulene as standard. The boxplot representation was done with the R statistical programming package. Within the box the line represents the median, the boundaries of the box the upper and lower quartiles, the vertical lines outside the box indicate the whiskers (variability outside the upper and lower quartile) and single points indicate extreme outliers. **b**: Graphical boxplot representation of the quintuplicate whole cell hydroxylation experiments using *E.coli* harboring *CotB4* combined with the same metabolic setting as shown in Figure 42**a**. 42**a**: **Cycloocat-9-en-7-ol** or **b**: **cyclooctat-9-en-5,7-diol** have been added *in vitro* (methods chapter III).

However, the increase in product titers of di- or trihydroxylated compounds in the AfR•Afx and the PdR•Pdx-W106E systems also could have been originated from higher expression or slower degradation rates of the respective ferredoxin proteins in the *in vivo* setting. Therefore, their relative expression levels in the whole cell extracts have consequently been tested by comparative SDS-PAGE analysis combined with peptide mass fingerprinting (Figure 43a, b, Table 11, 12). The SDS-PAGE analyses showed significant differences in relative expression levels. Specifically, the PdR•Pdx system showed a high overrepresentation of CotB3/B4 and

Pdx (Figure 43a, b). In contrast to this observation, very low relative expression levels of CotB3/4 and Afx were detected when the AfR•Afx positive control system was applied (Figure 43a, b). In the PdR•Pdx-W106E system the relative expression levels of CotB3 and Pdx-W106E interestingly decreased to the level of CotB3 and Afx observed in the AfR•Afx system (Figure 43a). Contrary, the relative expression levels of CotB3 and Pdx-D38E as well as Pdx-D38E/W106E remained almost constant compared to the levels of CotB3 and Pdx in the PdR•Pdx system (Figure 43a). In the CotB4 setting, CotB4 and Pdx-W106E levels remain at the level PdR•Pdx system when the Pdx-W106 system was applied (Figure 43b). Similarly, CotB4 and the respective mutant Pdx-ferredoxins remain at the level of the PdR•Pdx system when the mutant Pdx-D38E and Pdx-D38E/W106E have been applied in the *in vivo* whole cell setting (Figure 43b).

This data indicates that the exchange of Pdx-W106 to an amino acid that is able to facilitate a productive interaction with its corresponding recognition partner R114/R115 in CotB3/4 is the reason for the increased product formation instead of higher expression levels or slower degradation rates of Pdx-W106E compared to wild type Pdx.

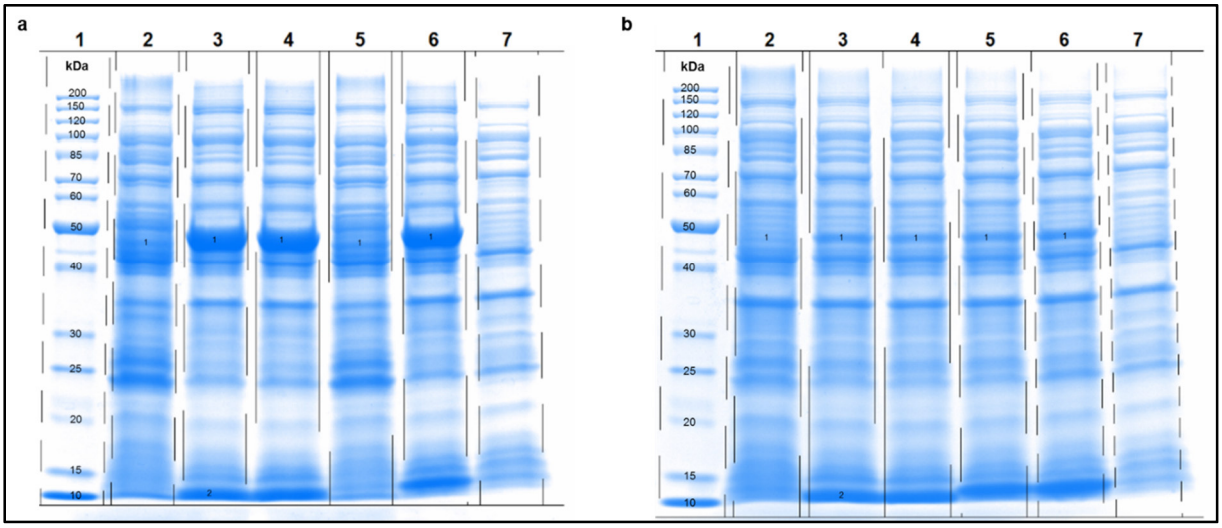


Figure 43: Comparative SDS-PAGE analysis of the whole cell proteomes from the whole cell hydroxylation experiments

Figure 43**a**, b: SDS-Page analysis of the whole cell proteomes from the whole cell catalysis experiments after 24 hours applying the different redox triples as described in the text. Descriptions are shown in Table 11 for **a** and in Table 12 for **b**. For detailed informations, see text.

Lane #	*Lane Description*	*Protein Band #*	*Protein Identified*
1	Marker	1	CotB3
2	pACYC-Duet-1 (*afR*/*afx*, cotB3)	2	Pdx
3	pACYC-Duet-1 (*pdR*/*pdx*, cotB3)	3	
4	pACYC-Duet-1 (*pdR*/*pdx*-D38E, cotB3)	4	
5	pACYC-Duet-1 (*pdR*/*pdx*-W106E, cotB3)	5	
6	pACYC-Duet-1 (pdR/pdx-D38E/W106E, cotB3)	6	
7	pACYC-Duet-1 (empty)		

Table 11: Description of Figure 43a

Lane #	*Lane Description*	*Protein Band #*	*Protein Identified*
1	Marker	1	CotB4
2	pACYC-Duet-1 (*afR*/*afx*, cotB4)	2	Pdx
3	pACYC-Duet-1 (*pdR*/*pdx*, cotB4)	3	
4	pACYC-Duet-1 (*pdR*/*pdx*-D38E, cotB4)	4	
5	pACYC-Duet-1 (*pdR*/*pdx*-W106E, cotB4)	5	
6	pACYC-Duet-1 (*pdR*/*pdx*-D38E/W106E, cotB4)	6	
7	pACYC-Duet-1 (empty)		

Table 12: Description of Figure 43b

4. Discussion

To obtain a high CYP450 activity the employment of the right electron partner proteins is critical. Coupling efficiencies of the electron transfer reactions among the respective CYP450 monooxygenase and the redox partner(s) are too low using ineffective redox partners.[100,101,104] To achieve the maximum CYP450 activity its functional natural redox partners should be used.

However, identification of the native redox partners is still challenging as very few sequence information is available. Further, the genomic loci of the redox partners are typically quite distant from the target CYP450 genes. Homology-based computational-assisted search in the available genomic sequence space using known and well-characterized redox systems as templates is a useful instrument to search for native and/or cognate redox partners for a specific CYP450 monooxygenase. As class I NADH-dependent reductases and 2Fe-2S ferredoxins do not share a high primary sequence identity, secondary structure-based homology-based searches are the methods of choice. However, experimental *in vitro* and/or *in vivo* validation is still necessary to test the functionality and the activation efficiency with respect to a specific CYP450 monooxygenase.

The functional activation of CotB3/4 by the newly identified redox system from *S. afghaniensis* is not that surprising, however it is a significant improve in the production of the bioactive diterpenoid **cyclooctatin**. There are other examples, where *Streptomyces*-derived CYP450 monooxygenases have been successfully activated by non-native eukaryotic and prokaryotic redox systems, however, with modest yield.[101,114-118] The efficiency of the activation by the AfR•Afx system leading to a 50-fold increase in the **cyclooctatin** titer compared to its native producer and the homologous reconstitution of the **cyclooctatin** pathway in *S. albus is* surprising. The results shown in this study are therefore another example that underline the use of at least cognate redox partners and that it is worth to search for suitable redox partners in phylogenetically related organisms. The choice of the redox partner determines the catalytic efficiency of the redox partner recognition in class I CYP450 systems and is thus responsible for the product titer.

S. afghaniensis is known to produce a polyketide-antibiotic Taitomycin, but diterpenoids originating from this organism have not been found. Therefore, the new identified redox system may be affiliated with polyketide hydroxylation in its native host. On the other hand, substrate specificity is solely mediated by the CYP450 monooxygenase and is not originating from the

use of the redox system. As shown for the *de novo* production of **sinularcasbane D**, the catalytically efficient and easy to handle CotB3/4•AfR•Afx system is therefore a valuable protein engineering target system to enable functionalization of other diterpene and polyketide macrocycles.

Structural insights into the partner recognition of bacterial class I redox systems with class I CYP450 monooxygenases come mainly from the *P. putida* system. In this system Pdx is exclusively able to activate its native partner monooxygenase P450cam, as the unique effector role of Pdx during ET2 is the promotion of structural changes in P450cam necessary for O_2 activation, electron transfer and product formation.[104,110] Effective partner recognition that serves as the basis for the correct orientation of both redox partners important for both electron transfer processes as well as mediation of ET2 is shown to be mainly facilitated by Pdx-D38 and -W106.[110] Other structurally characterized bacterial CYP450 monooxygenases like P450cin, acting in hydroxylation of the monoterpenoid cineole shows a somewhat lesser ferredoxin specificity, as it can be activated by Pdx and the mitochondrial ferredoxin adrenodoxin.[104] Recent studies show that P450cin do not rely on the promotion of structural changes upon redox partner binding and electron transfer despite the structural similarity between P450cam and P450cin, especially in the chemical nature at the redox partner binding surface.[119] The studies using the four plasmid setting as well as the whole cell catalysis experiments show that the activation of CotB3/4 is also characterized by a minor distinct ferredoxin specificity. This indicates that unlike P450cam, O_2 activation, ET2 and product release in CotB3/4 do not rely on structural changes derived from ferredoxin binding. However, further structural analyses and especially controlled *in* vitro assays using purified enzymes have to be conducted to ensure this presumption. Moreover, detailed insights into oxidation rates, coupling efficiencies and catalytic turnover rates have to be obtained to fully confirm the situation that arises from the whole cell catalysis experiments shown in this study. The whole cell hydroxylation experiments show a significant effect in the increase of product concentration in

the CotB3/Pdx-W106E experiment. However, the catalytic efficiency have to be determined using purified enzymes to corroborate these experiments. The weak expression rate of Afx and low solubility, even with stabilizing protein tags, during expression of Afx renders this approach hardly feasible at the moment (personal communication with Dr. Loll).

Another well-characterized bacterial class I CYP450 monooxygenase•reductase•ferredoxin system is the CYP1199A2•PuR•Pux system from *Rhodopseudomonas palustris*.[120] While PuR (reductase) can be efficiently replaced by PdR, Pux (ferredoxin) cannot be replaced by Pdx despite a high degree of structural similarity. Structural comparisons between Pdx and Pux show a substitution of D to E at position 38 in Pux and a W to R substitution at position 106 in Pux. The change of an aromatic amino acid to an arginine at the important position W106 is thought to eliminate the ability of Pux to serve as a redox partner for P450cam.[104]

This is in agreement with and further supports the results of this study. The change of W106 in Pdx to its analogous residue from the newly identified Afx seems to enable Pdx-W106E to efficiently activate the non-native CYP450 monooxygenases CotB3/4. SDS-PAGE analyses and peptide mass fingerprinting indicate that the increase in product titers originates from a favorable molecular redox partner recognition, as predicted *in silico,* instead of higher expression or slower degradation levels of Pdx-W106E in the whole cell catalysis experiments. The interesting fact was the observation that product titer increases although the relative CotB3 and mutant Pdx expression levels in the PdR/Pdx-W106E system decreases to a similar less relative expression level that is observed when the CotB3/AfR/Afx system is applied. After several repeats using the same metabolic setting and resequencing of the constructs, this observation persists. A definite explanation of this observation remains elusive, however it seems that the expression levels adjust upon a productive and effective interaction of both redox partners that results in increased product titers in comparison to the less productive CotB3/PdR/Pdx and CotB3/PdR/Pdx-D38E and -D38E/W106E whole cell catalysis experiments. A more quantitative approach using mass spectrometry or purified enzyme assays are

again the way to definitively answer this question. However, the application of whole cell catalysis experiments using a simple and controlled setting was able to corroborate the *in silico* prediction and to experimentally observe a significant increase in the production rate of **cyclooctat-9-en-5,7-diol** and a more modest increase in the production rate of **cyclooctatin** compared to the negative and positive control systems.

Pdx-W106E was further able to functionally activate CotB4, albeit in significantly minor rates. This result was rather unexpected as CotB3 and CotB4 show a high structural similarity and both were predicted *in silico* to display similar productive interactions with Afx at the molecular level. The modelling of the CotB3•Afx and CotB4•Afx complexes displays the structural situation during the second rate-limiting electron transfer step, where Afx-E107 and Afx-E39 are supposed to play the key roles in partner recognition and electron transfer in analogy to its *P. putida* counterpart.[110] The structural situation during the first electron transfer step could not be elucidated due to the lack of template structures. It is known for Pdx that Pdx-D38 and Pdx-W106 play important roles during this event[104], however, a lot of other amino acid residues in both redox partners participate in the first electron transfer step.[121] Therefore, the predictions for CotB4 derived from *in silico* modeling may be accurate and similar to that of CotB3, however, partner recognition as basis for the first electron transfer step may be impaired due to the high degree of divergence in primary sequence of Pdx-W106E compared to Afx.

A similar situation arose from the ineffectiveness of the Pdx-D38E and –D38E/W106 mutants in the activation of CotB3 and CotB4. A detailed comparative modelling approach of PdR and AfR assume that the reductase•ferredoxin electron transfer displays significant differences (Figure 44). In the PdR•Pdx complex the partner recognition is shown to incorporate hydrogen bridge interactions between Pdx-D38 and PdR-R310, whereas an efficient electron transfer requires the interplay between a tryptophan residue (PdR-W330) and NADH.[111] The model of the AfR•Afx complex also shows the requirement of a tryptophan residue (AfR-W297) for the interaction with NADH, the ferredoxin•reductase partner recognition, however, seems to be

mechanistically different from the PdR•Pdx template (Figure 44). The glutamic acid residue at position AfR-39 seems to be involved in partner recognition due to the overall similar fold between PdR and AfR. The model suggest a glutamine (AfRQ 305) as its corresponding interaction partner, but this is rather speculative. The change of Pdx-D38 to E in analogy to the respective residue in Afx may thus disturb the partner recognition of PdR with Pdx-D38E and eliminate the electron transport mechanisms in line with a previous report.[122] This suggests that the interaction of Pdx-D38E and -D38E/W106E with CotB3 and CotB4 may be productive, as predicted by *in silico* modelling. However, as the interaction between both mutants and its native reductase is impaired no titer increasing effect can be detected.

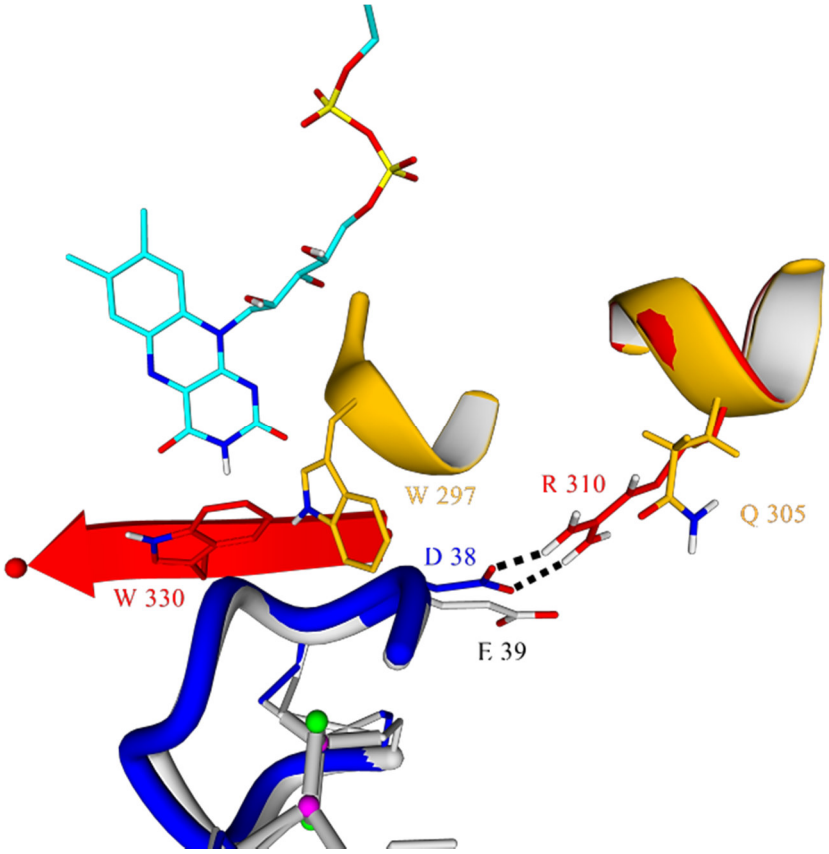


Figure 44: Structural superposition of PdR•Pdx with the modelled AfR•Afx complex

Figure 44: Structural superposition of the PdR•Pdx (red·blue) with the modelled AfR•Afx complex (ocher•grey). Prosthetic FAD group of PdR and modelled AfR are shown in light blue. 2Fe-2S cluster of Pdx and modelled Afx are shown in magenta-green. The important hydrogen bridge (broken black line) between Pdx_{D38} and PdR_{R310}

needed for binding as well as PdR_{W330} facilitating electron transport are shown.[111] The differing positioning of PdR_{W330} that corresponds to W297 in AfR, presumably acting in electron transfer from FAD to Pdx is shown in ocher. Analogous interactions facilitating binding of AfR to Afx could not be deciphered. Substitution of Pdx_{D38} to E38 is suggested to impair electron transport by an alternate binding mode of Pdx_{D38E} to native PdR.

In conclusion, *in silico* complex modelling of class I redox partner complexes with well-characterized structures as templates is a way to analyze and optimize partner recognition between CYP450 monooxygenases and non-native ferredoxins. Adaption of the ferredoxin C-terminal residue in order to ensure a chemically productive interaction with the respective CYP450 monooxygenase is a promising result that could help to improve class I redox-dependent biotechnological processes. A recent study shows the opposite way of improving the product titers in class I redox-coupled processes.[123] Instead of optimization of the CYP450 monooxygenase binding site in the ferredoxin, the ferredoxin binding site in the CYP450 monooxygenase was optimized. However, for the general application of such approaches, more structures of redox complexes have to be analyzed and generally more possible non-native redox partners have to be found.

5. Methods

General experimental procedure

All chemicals were obtained from standard sources at the highest purity grade. NMR spectra were recorded in $CDCl_3$ or CD_3OD with an Avance-III 500 MHz (Bruker) at 300 K. ^1H NMR chemical shifts are given in ppm relative to $CHCl_3$ (δ=7.26 ppm) or CD_2HOD (δ=3.31 ppm) and CD_3OH (δ=4.87) (^1H NMR). ^{13}C NMR chemical shifts are given in ppm relative to $CDCl_3$ at δ=77.16 ppm or CD_3OH at δ=49.00 ppm. The 2D experiments (HSQC, COSY, NOESY) were performed using standard Bruker pulse sequences and parameters.

GC-MS analysis of diterpene products from ethyl acetate extractions was done by a Trace GC Ultra with DSQII (Thermo Scientific). Probes were silylated (1 h at 70°C) prior to measurement by 37.5 % (v/v) N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) and 25 % (v/v) chlorotrimethylsilan (TMS-Cl) in pyridine. For the GC-MS analysis of dodecane probes, 50 µL of the dodecane probe was mixed with 45 µL pyridine, 45 uL N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) and 10 µL% (v/v) chlorotrimethylsilan (TMS-Cl). Silylation was done for 4 h at 70 °C. One µl sample was applied by TriPlus AS onto a SGE BPX5 column (30 m, I.D 0.25 mm, Film 0.25 µm). The initial column temperature was 50°C (maintained for 2.5 min). A temperature gradient was applied from 50°C – 320°C (10 °C/min), followed by 3 min maintenance at 320°C. MS data were recorded at 70 eV (EI), *m/z* (rel. intensity in %) as TIC, total ion current. The recorded *m/z* range was 50 – 650. Quantification was performed with flame ionization detector (FID) using 1 mg/mL α-humulene (Sigma-Aldrich, Germany) as an internal standard. To apply the FID signal of α-humulene to the silylated diterpene products, correlation factors were calculated according to the literature. From the determined mass concentrations of the silylated diterpene products the mass concentrations of the non-derivate products were calculated.

High-resolution mass spectra of diterpenes were determined with Thermo Fisher Acela HPLC system linked to a Thermo Fisher Scientific LTQ Orbitrap XL mass. Electrospray ionization was done in positive ion mode.

MALDI-MS analysis using an Ultraflex I (Bruker Daltronics) was applied to identify the recombinant expressed proteins in the observed bands on SDS-PAGE. Bands were excised and digested with trypsin as described previously. Prior to MS-analysis, peptides were concentrated and purified using C18 ZipTips (Merck Millipore) following the manufacturer's protocol. Data analysis was performed using the MASCOT software program (Matrix Science, London, UK) along with the National Center for Biotechnology Information database.

Circular dichroism (CD) spectroscopy was performed using a Chirascan plus spectropolarimeter (Applied Photophysics). Samples were dissolved in acetonitrile and spectra were recorded in quartz cuvettes with 0.1 cm path length at 20 °C.

Bacterial strains, genes and vectors

The *E. coli* strains XL-1 Blue and Bl21 (DE3) were used for cloning and diterpene production. All strains and plasmids were obtained from Novagen/ Merk Millipore. Genes were synthesized by Life technologies GmbH featuring the appropriate restriction sites and adjusting codon usage for *E. coli*.

Plasmids for the cyclooctatin production

Plasmids used to construct the overexpressed DXP pathway in *E. coli* Bl21 (DE3) are described in Table 4.

To achieve biosynthesis of cyclooctatin the native geranylgeranyl diphosphate synthase (*crte*) (GenBank: M90698.1) sequence was amplified from *Pantoea agglomerans* (ATCC 27155) using standard protocols. Primers used were 5'-AAA CCA TGG CAA TGG CAA CGG TCT GCG CA-3' and 5'-AAA GAA TTC TTA ACT GAC GGC AGC GAG TTT-3'. Genes of *cotB2* (GenBank: BAI44338.1) (Appendix section 6), *cotB3* (GenBank: BAI44339.1) (Appendix section 6), *cotB4* (GenBank: BAI44340.1) (Appendix section 6) from *Streptomyces melanosporofaciens* MI614-43F2 as well as the genes to express the reductase (*afR*) (GenBank: WP_020277402) (Appendix section 6) and ferrodoxin (*afx*) (GenBank: WP_020276845) (Appendix section 6)

from *Streptomyces afghaniensis* were synthesized. *CotB3*/*B4* (Appendix section 6), *afR*/*afx* (Appendix section 6) and *pdR*/*pdx* (Appendix section 6) from *Pseudomonas putida* were expressed from bi-cistronic operons, introduced, respectively, into the appropriate plasmids according to Table 13 by standard cloning techniques.

Gene(s)	**Vector**	**Multiple Cloning Site**	**Restriction Sites**
crtE	pET-Duet-1	I	*Nco*I, *EcoR*I
cotB2 or txs *or* cs (casbene synthase)	pET-Duet-1	II	*Nde*I, *Xho*I
afR/afx operon or *pdR/pdx* operon	pACYCDuet-1	I	*Nco*I, *Not*I
cotB3/cotB4 operon	pACYCDuet-1	II	*Nde*I, *Xho*I

Table 13: Plasmids used to construct the cyclooctatin biosynthesis in* E. coli *Bl21 (DE3)

Plasmids used to evaluate hydroxylation of diterpenes by CotB3/4

To evaluate if the P450 hydroxylases CotB3/4 are capable to functionalize other diterpene skeletons, the diterpene synthase gene *cotb2* in the vector pET-Duet-1 (Table 13) was exchanged by a series of other genes shown on Table 14.

Gene(s)	**Description**	**Diterpene produced**
cs (Appendix section 6)	casbene synthase	(-)-casbene
txs (Appendix section 6)	taxadiene synthase	taxa-4,5-11,12-diene
$cotb2_{W288G}$[45]	cyclooctatin synthase	(1R,3E,7E,11S,12S)-3,7,18-dolabellatriene
$cotb2_{F107A}$[93]	cyclooctatin synthase	(R)-cembrene A

Table 14: Plasmids used to evaluate hydroxylation of diterpenes by CotB3/4

Plasmids to evaluate different redox-system variants

The catalytic activities of single CotB3 and CotB4 proteins were evaluated using different redox-system variants. The redox system from *S. afghaniensis* (*afR*/*afx*), *P. putida* (*pdR*/*pdx*) and mutant variants of *P. putida* (*pdR*/*pdx*) were tested. Consequently, the vector pACYCDuet-1 either carried *cotB3* or *cotB4* expressing genes in multiple cloning site II and the bi-cistronic operon of reductase/ ferredoxin in multiple cloning site I (Table 13). Different variants of the pACYCDuet-1 vector are shown on Table 15.

P450 Proteins	**Redox system**	**Mutations**
CotB3	AfR/Afx	-
CotB3	PdR/Pdx	-
CotB3	PdR/Pdx	D38E
CotB3	PdR/Pdx	W106E
CotB3	PdR/Pdx	D38E, W106E
CotB4	AfR/Afx	-
CotB4	PdR/Pdx	-
CotB4	PdR/Pdx	D38E
CotB4	PdR/Pdx	W106E
CotB4	PdR/Pdx	D38E, W106E

Table 15: Plasmids used to compare the activity of CotB3 and CotB4 hydroxylases with different redox system variants

The *cotB3* expressing gene was amplified from the bi-cistronic operon (*cotB3/cotB4*) using the primers: 5'-ATT ACA TAT GCG TGA ACG TGG T-3' and 5'-ATA TCT CGA GTT AAC GCG GTT CAC AAA CCA T-3'. *CotB4* was amplified from the bi-cistronic operon *(cotB3/cotB4*) using the primers: 5'-ATT ACA TAT GAA AGA TTT TTT TCG TAT GCG CAC-3' and 5'-ATA TCT CGA GTT AAC GAG GTT CCG-3'. The bi-cistronic operons expressing (*afR*/*afx*) / (*pdR*/*pdx*) and *cotB3/B4* were introduced into the pACYCDuet-1 according by standard cloning techniques. Point mutations in Pdx protein from the redox system were introduced by the QuikChange site-directed mutagenesis protocol. Primers for pdx_{D38E} were: 5'-GAT ATT GTT GGT GAA TGT GGT GGT AGC G-3' and 5'-CGC TAC CAC CAC ATT CAC CAA CAA TAT C-3'. Primers pdx_{W106E} were: 5'-GAT GTT CCG GAT CGT CAG GAA TAA GCG GCC GCA TAA TG-3' and 5'-CAT TAT GCG GCC GCT TAT TCC TGA CGA TCC GGA ACA TC-3'. The different pACYCDuet-1 (*afR*/*afx* or *pdR*/*pdx*, *cotB3* or *cotB4*) vectors were transformed into Bl21 (DE3) cells.

Production of cyclooctatin

The vectors pColaDuet-1 (*dxr*, *dxs*), pCDFDuet-1 (*ispD*/*ispF*, *idi*), pETDuet-1 (*crte*, *cotB2*) and the vector pACYCDuet-1 (*afR*/*afx*, cotB3/cotB4) were introduced into *E. coli* Bl21 (DE3) by standard transformation procedures. Cultivation conditions were based on Boghigian *et al.*[36]

Shake-flask production culture of cyclooctatin

For cultivation in shake flasks, cells were grown in 100 mL medium containing LB-Media (10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl) supplemented with 100 mM HEPES (pH 7.6), 40 g/L glycerol, 30 µg/mL kanamycin, 50 µg/mL streptomycin, 50 µg/mL carbenicillin, 34 µg/mL chloramphenicol, 1 mM δ-aminolevulinic acid (ALA), 1 mM $FeSO_4$ x $7H_2O$, 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 40 µL 30 % Antifoam A) in 5 L baffled glass flasks. The culture was inoculated at OD_{600} 0.1 from an overnight culture (8 h cultivation 37 °C supplemented with 30 µg/mL kanamycin 50 µg/mL streptomycin, 50 µg/mL carbenicillin, 34 µg/mL chloramphenicol) and cultivated at 25 °C for 4 days.

Production of sinularcasbene D

The vectors pColaDuet-1 (*dxr*, *dxs*), pCDFDuet-1 (*ispD*/*ispF*, *idi*), pETDuet-1 (*crte*, *cs*) and the vector pACYCDuet-1 (*afR*/*afx*, *cotB3*) were introduced into *E. coli* Bl21 (DE3) by standard transformation procedures. Cultivation conditions were based on Boghigian *et al.*[36]

Batch bioprocess of sinularcasbene D

For cultivation in shake flasks, cells were grown in 400 mL medium containing LB-Media (10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl), 40 g/L glycerol, 30 µg/mL kanamycin, 50 µg/mL streptomycin, 50 µg/mL carbenicillin, 34 µg/mL chloramphenicol, 1 mM δ-aminolevulinic acid (ALA), 1 mM $FeSO_4$ x $7H_2O$, 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 40 µL 30 % Antifoam A) in 5 L baffled glass flasks. The culture was inoculated at OD_{600} 0.1 from an overnight culture (8 h cultivation 37 °C supplemented with 30 µg/mL kanamycin 50 µg/mL streptomycin, 50 µg/mL carbenicillin, 34 µg/mL chloramphenicol) and cultivated at 25 °C for 3 days.

Extraction and isolation of cyclooctatin and its biosynthetic precursors

For the diterpene isolation, cells were pelleted (15 min, 17,500 g, 4 °C). Cyclooctatin and its biosynthetic precursors were extracted from *E. coli* cells and supernatant, separately. The cell pellet was washed with water and then suspended in 5 mL water. Next, cells were lysed by sonification using a Sonoplus HD2070 (Bandelin Electronic) performing 5 repeats on ice (5 min on and 3 min off at 80 % power) and subsequently extracted 3 times with 25 mL ethyl acetate. The supernatant was extracted 3 times by 200 mL ethyl acetate. Organic phases were combined, dried with $MgSO_4$ and evaporated under vacuum to dryness. The crude extract was solved in 1 mL ethyl acetate and analysed by GC-MS and -FID. Purification of cyclooctat-9-en-7-ol was carried out by flash chromatography. An isocratic 70/30 hexane/ethyl acetate silica step (Silica gel 40, Sigma-Aldrich,) was followed by an isocratic 30/70 water/acetonitrile reversed phase chromatography step on a Polygoprep 60-50, C_{18} column (Macherey-Nagel).

For purification of cyclooctat-9-en-5,7-diol by flash chromatography, the hexane/ethyl acetate solvent was changed to 50/50 and the water/acetonitrile solvent to 10/90. Cyclooctatin was purified by flash chromatography using hexane/ethyl acetate 30/70 as solvent followed by a 10/90 water acetonitrile reversed phase.

Extraction and isolation of (-)- casbene and sinularcasbene D

For the (-)- casbene and sinularcasbene D isolation, cells were pelleted (15 min, 17,500 g, 4 °C). Sinularcasbene D and its biosynthetic precursor (-)- casbene were extracted from *E. coli* cells and supernatant, separately. The cell pellet was washed with water and then suspended in 5 mL water. Next, cells were lysed by sonification on a Sonoplus HD2070 (Bandelin Electronic) performing 5 repeats on ice (5 min on and 3 min off at 80 % power) and subsequently extracted 3 times with 25 mL ethyl acetate. The supernatant was extracted 3 times by 200 mL ethyl acetate. Organic phases were combined, dried with $MgSO_4$ and evaporated under vacuum to dryness. The crude extract was solved in 1 mL ethyl acetate and analysed by GC-MS and -FID. Purification of (-)- casbene and sinularcasbene D was carried out by flash chromatography. An isocratic 90/10 hexane/ethyl acetate chromatography on silica (Silica gel 40, Sigma-Aldrich) was applied.

Evaluating CotB3/4 for activity on different diterpene skeletons

The vectors pColaDuet-1 (*dxr*, *dxs*), pCDFDuet-1 (*ispD*/*ispF*, *idi*), pETDuet-1 (*crte*, *cs*)/ pETDuet-1 (*crte*, *txs*)/ pETDuet-1 (*crte*, $cotb2_{W288G}$) / pETDuet-1 (crte, $cotb2_{F107A}$) and the vector pACYCDuet-1 (afR/afx, cotB3)/ pACYCDuet-1 (afR/afx, cotB4) were introduced into *E. coli* Bl21 (DE3) by standard transformation procedures.

For cultivation in shake flasks, cells were grown in 400 mL medium containing LB-Media (10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl), 40 g/L glycerol, 30 µg/mL kanamycin, 50 µg/mL streptomycin, 50 µg/mL carbenicillin, 34 µg/mL chloramphenicol, 1 mM δ-aminolevulinic

acid (ALA), 1 mM $FeSO_4$ x $7H_2O$, 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 40 µL 30 % Antifoam A) in 5 L baffled glass flasks. The culture was inoculated at OD_{600} 0.1 from an overnight culture (8 h cultivation 37 °C supplemented with 30 µg/mL kanamycin 50 µg/mL streptomycin, 50 µg/mL carbenicillin, 34 µg/mL chloramphenicol) and cultivated at 25 °C for 3 days. The cells were pelleted (15 min, 17,500 g, 4 °C). Terpenes were extracted from *E. coli* cells and supernatant, separately. The cell pellet was washed with water and then suspended in 5 mL water. Next, cells were lysed by sonification using a Sonoplus HD2070 (Bandelin Electronic, Germany) performing 5 repeats on ice (5 min on and 3 min off at 80 % power) and subsequently extracted 3 times with 25 mL ethyl acetate. The supernatant was extracted 3 times by 200 mL ethyl acetate. Organic phases were combined, dried with $MgSO_4$ and evaporated under vacuum to dryness. The crude extract was solved in 1 mL ethyl acetate and analysed by GC-MS and FID.

Whole cell hydroxylation **assay for comparison of redox-systems**

The electron transfer of different redox system variants was evaluated to promote catalytic efficiency of CotB3 and CotB4. Therefore, Bl21 (DE3) cells harboring different pACYCDuet-1 (*afR*/*afx* or *pdR*/*pdx*, *cotB3* or *cotB4*) vectors were used in an in-vivo assay. In this assay the P450 and a corresponding redox system were expressed in the absence of the DXP pathway. The metabolic precursor cyclooctat-9-en-7-ol (for CotB3) or cyclooct-9-en-5,7-diol (for CotB4) was added to the media and the amount of catalytic products cyclooct-9-en-5,7-diol (for CotB3) or cyclooctatin (for CotB4) was measured.

From an overnight culture (LB media, 34 µg/mL chloramphenicol) Bl21(DE3) cells harboring different pACYCDuet-1 encoding (*afR*/*afx* or *pdR*/*pdx*, *cotB3* or *cotB4*) vectors were grown in 100 mL baffled shake flask cultures at 37°C using 50 mL LB media supplemented with 34 µg/mL chloramphenicol. Bl21 (DE3) harboring the empty pACYCDuet-1 vector was used as negative control. All strains were cultivated five times. After the cultures reached the OD_{600} 0.8,

1 mM δ-aminolevulinic acid (ALA), 1 mM $FeSO_4$ x $7H_2O$ and 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added and the cultures were incubated at 30 °C for 24 hours. For the in vivo assay, the cell density of the cultures were adjusted to OD_{600} 6 by centrifugation (3000 g for 5 minutes) and re-suspension in the appropriate volume. 10 mL of each culture (OD_{600} = 6) were transferred to 50 mL baffled shake flasks and supplemented with 100 µM of cyclooctat-9-en-7-ol (CotB3) or cyclooct-9-en-5,7-diol (CotB4). Terpenes were added in acetonitrile to a final concentration of 0.1 % (*v/v*). The cultures were sealed by membranes (Greiner bio-one breathseal 676051) to avoid evaporation and incubated at 30°C for 2 days. To analyze the whole cell proteins 100 µL samples were taken every 6 hours. The culture samples were centrifuged (10000 g for 5 minutes), the supernatant was discarded and the cell pellets were stored at -20°C. The samples were analyzed by a 12 % SDS-Page gel according to Lämmli. After cultivation the cultures were directly lysed by sonification on a Sonoplus HD2070 (Bandelin Electronic, Germany) performing 5 repeats on ice (5 min on and 3 min off at 80 % power) and subsequently extracted 3 times with 25 mL ethyl acetate. Organic phases were combined, dried with $MgSO_4$ and evaporated under vacuum to dryness. The crude extract was solved in 1 mL ethyl acetate and analyzed by GC-MS. The GC-MS spectra were normalized to the total icon count of α-humulene (Sigma-Aldrich, Germany), which was used as an internal standard. Quantification was done by determining the total ion count of cyclooctatin and cyclooct-9-en-5,7-diol, respectively.

Modelling of CotB3 and CotB4

Homology models were built with Schrödinger's Prime version 3.4 (Schrödinger, LLC, New York) from alignments derived from PROMALS3D.[124] For CotB3, we applied a chimera model using the crystal structure of *Streptomyces coelicolor* cytochrome P450 (PDB 3EL3, 39% sequence identity) as main template and crystal structure of human cytochrome P450 (PDB 3LD6, 23% sequence identity) as template for missing loops at positions 194 to 209 and 250 to 258. We applied the same approach for CotB4, with sequence identities of 39% and 20% for templates 3EL3 and 3LD6, respectively. Resulting models were subject to protonation state

assignment at pH = 7.0 using PROPKA (http://nbcr-222.ucsd.edu/pdb2pqr_2.0.0/) and protein relaxation using restrained minimization at a convergence factor of RMSD = 0.3 Å for the heavy atoms. The prosthetic heme group was kept frozen during this process and the formal charge of its iron atom set to +3 (high-spin sexted Fe (III)), as we simulated the enzyme-substrate complex right after the resting state where the substrate has replaced the distal water.

Docking of CotB3, CotB4 with its substrates

The ligand structures of cyclooct-9-en-7-ol and cyclooct-9-en-5,7-diol were built with Schrödinger's Maestro version 0.6 (Schrödinger, LLC, New York). Their conformations and partial charges were taken from QM minimization with a dielectric constant of 4.8, simulating protein environment. Ligand docking was done with Schrödinger's induced fit docking protocol. First, the receptor residues within 5.0 Å of the ligand are trimmed, then the ligand flexibly docked into the cavity and finally the side chains of residues within 5.0 Å vicinity are sampled and minimized together with the ligand. The prosthetic heme b group was kept frozen throughout this process and resulting docking poses were sorted by the IFDScore (Schrödinger manual).

Modelling of complexes

Secondary structure analysis and modeling of the tertiary structures of putative reductases and ferredoxins from *S. afghaniensis* whole shotgun-sequences was done using the HHpred server tool (see methods chapter II). Structural alignments have been performed using the MUSTANG approach of YASARA structure. The model of the AfR•Afx complex has been derived by subsequent alignments of both AfR and Afx against the PdR•Pdx (PDB-ID: 3LB8) complex. Models of CotB3•/CotB4•Afx and CotB3•/CotB4•Pdx complexes have been derived by the same approach using the P450cam•Pdx complex (PDB-ID: 4JX1) as template.

Assigned complexes have subsequently been subjected to 10.000 steps of energy minimization in a TIP3P water box at pH 7.0 using the YAMBER03 force field of YASARA structure. The force field parameters of cyclooct-9-en-7-ol and cyclooct-9-en-5,7-diol, the prosthetic heme group and the Fe2-S2-cluster was obtained using the AutoSmiles force field parameter

assignment implemented in YASARA structure (see methods chapter II). During energy minimization all residues, heavy atoms, prosthetic groups and both substrates were unconstrained, non-bonded cutoff were set to 7.5 Å and PME method (Particle-Mesh-Ewald) was used for long-range electrostatic forces (see methods chapter II).

Molecular Docking of (-)-casbene and sinularcasbene D in the cotB3 active site

Molecular Docking was performed using the AutoDock Vina program environment of YASARA structure. Sinularcasbene D was docked into a simulation cell (Size: X-size = 16Å, Y-size = 16Å, Z-size= 16Å, angles: alpha = 90°, beta = 90°, gamma = 90°) around the following four residues Cys87, Arg185, Ala282 and Leu361. 999 docking runs were performed while all atoms of (-)-casbene and sinularcasbene D were set as rigid. Cluster analysis were performed in the AutoDock Vina program environment and characterized by binding energy[kcal/mol], dissocation constant [pM] and contacting receptor residues (see methods chapter II).

Quality assessments

15.1 Stereochemical quality assessments: PROCHECK software package (Ramachandran plot, g-factor)

"The PROCHECK analyses provide an idea of the stereochemical quality of all protein chains in a given PDB structure. They highlight regions of the proteins which appear to have unusual geometry and provide an overall assessment of the structure as a whole."

http://www.ebi.ac.uk/pdbsum/

a. Interpretation of PROCHECK derived Ramachandran plot statistics

*"Based on an analysis of **118** structures of resolution of at least **2.0** Angstroms and R-factor no greater than **20.0** a good quality model would be expected to have over **90%** in the most favored regions"*

http://www.ebi.ac.uk/pdbsum/

b. Bond-length and bond-angles G-factors

"For the main-chain bond lengths and bond angles, the G-factors are computed using the Engh & Huber (1991) small-molecule means and standard deviations. G-factors provide a measure of how unusual, or out-of-the-ordinary, a property is.

Values below -0.5* - unusual

Values below -1.0** - highly unusual"

http://www.ebi.ac.uk/pdbsum/

15.2 Assessment of absolute structural model qualities

*"The **QMEAN Z-score** provides an estimate of the absolute quality of a model by relating it to reference structures solved by X-ray crystallography] The QMEAN Z-score is an estimate of the "degree of nativeness" of the structural features observed in a model by describing the likelihood that a model is of comparable quality to high-resolution experimental structure.*

The model's QMEAN score is compared to the scores obtained for experimental structures of similar size (model size +/- 10%) and a Z-score is calculated. A Z-score (or standard score) is a score which is normalised to mean 0 and standard deviation 1. Thus the QMEAN Z-score directly indicates how many standard deviations the model's QMEAN score differs from expected values for experimental structures.

QMEAN, which stands for Qualitative Model Energy Analysis, is a composite scoring function describing the major geometrical aspects of protein structures. Five different structural descriptors are used. The local geometry is analyzed by a new kind of torsion angle potential over three consecutive amino acids. A secondary structure-specific distance-dependent pairwise residue-level potential is used to assess long-range interactions. A solvation po-

tential describes the burial status of the residues. Two simple terms describing the agreement of predicted and calculated secondary structure and solvent accessibility, respectively, are also included."

http://swissmodel.expasy.org/qmean/cgi/index.cgi

Chapter IV: Computer-assisted characterization of the cembranoid pathway enzymes from common tobacco and heterologous production of its functionalized diterpenoid bioactives

Patrick Schrepfer [1], **Ugur Ilke** [2], **Metterlein Michael** [1], **Balogh Dóra** [1], **Kaila Ville** [2] **and Thomas Brück** [1]

[1] **Department of Industrial Biocatalysis, Technische Universität München, Lichtenbergstr. 4, 85748 Garching, Germany**

[2] **Department of Computational Biocatalysis, Technische Universität München, Lichtenbergstr. 4, 85748 Garching, Germany**

Author contributions:

T.B. and P.S. conceived the study, planned and supervised experiments. P.S., M.M. and D.B. conducted metabolic engineering of *E.coli* and enzyme mutagenesis experiments. P.S. conducted fermentation experiments. P.S. conducted NMR data collection and verified produced structures. P.S. conducted *in silico* modelling and design of CBTS. I.U. and V.K. conducted QM gas phase modelling. P.S., V.K. and T.B. will finalize the manuscript.

1. Introduction

Cembranoids are a subgroup of monocyclic diterpenoids. The cembrane ring is a 14-carbon ring with an isopropyl group at C1 and methyl groups at C4, C8 and C12 (Figure 45). Cembranoids are characterized by a varying number of oxygen atoms. More than 300 natural cembranoids are known and most of them are isolated from marine invertebrates, plants and insects.[34,98,102,125-128] The surface of tobacco plants (*Nicotiana spp.)* contains high amounts of non-volatile cembranoids. 60% of the dry weight of the resinous substance covering the tobacco leaves contains (1S,2E,4S,6R,7E,11E)-2,7,11-cembratriene-4,6-diol (**α-CBT-diol**) and its epimer, (1S,2E,4R,6R,7E,11E)-2,7,11-cembratriene-4,6-diol (**β-CBT-diol**) (Figure 45).[128] Additionally, this exudate contains their biosynthetic precursors, (1S,2E,4S,7E,11E)-2,7,11-cembratriene-4-ol (**α-CBT-ol**, thunbergol) and (1S,2E,4R,7E,11E)-2,7,11-cembratriene-4-ol (**β-CBT-ol**, 4-epiisocembrol) (Figure 45), respectively, however in a much lower quantity (~1.4% of dry exudate weight).[128]

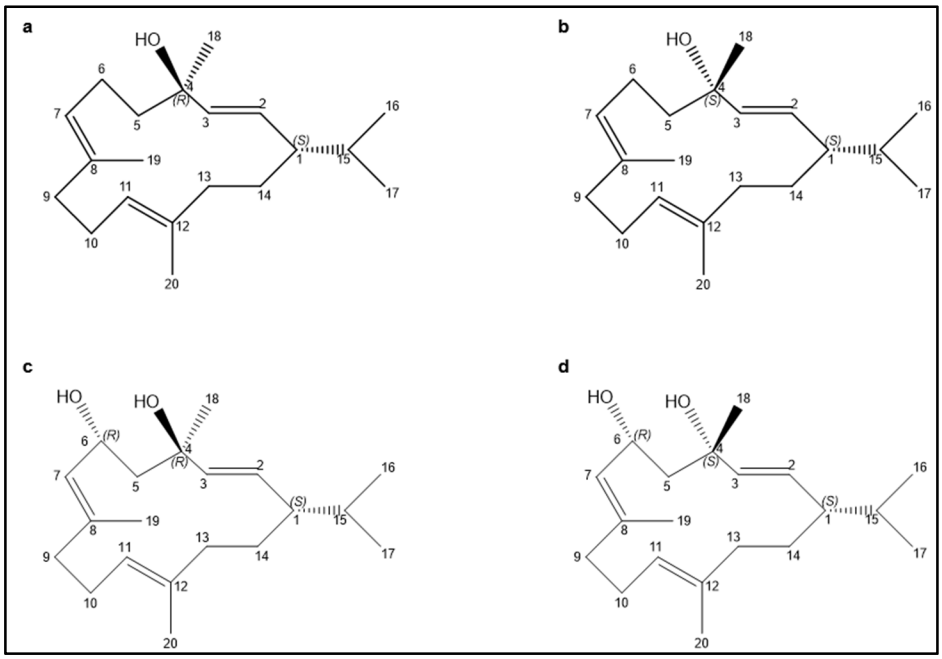


Figure 45: Stereostructures of the major tobacco cembranoids

Figure 45**a**: (1S,2E,4S,7E,11E)-2,7,11-cembratriene-4-ol (**α-CBT-ol**, thunbergol), **b**: (1S,2E,4R,7E,11E)-2,7,11-cembratriene-4-ol (**β-CBT-ol**, 4-epiisocembrol), **c**: (1S,2E,4S,6R,7E,11E)-2,7,11-cembratriene-4,6-diol (**α-CBT-diol**) and **d**: (1S,2E,4R,6R,7E,11E)-2,7,11-cembratriene-4,6-diol (**β-CBT-diol**).

CBT-diols show a high degree of anti-proliferative activity against different types of cancer.[34,129] Both were reported to significantly decrease the process of tumor formation by inhibiting the induction of tumor promoters (such as Epstein-Barr virus early antigen, ornithin decarboxy-lase). **α-CBT-diol** depicts a broad spectrum of antibacterial and antifungal activity against drug resistant strains (*Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Candida albicans*), while **β-CBT-diol** also shows promising antibacterial and -fungal activity against specific bacteria and fungi.[34] Furthermore, **CBT-diols** are inhibitors of muscle-type and neuronal nicotinic acetylcholine receptors, and therefore are suggested to be potential therapeutics against neurodegenerative diseases.[130,131] In addition, they also show neuroprotective effects. Their neuroprotective function is probably enhanced by their anti-inflammatory properties, as both inhibit prostaglandin synthesis even at low concentrations.[132] **CBT-diols** are also shown to inhibit nicotine sensitization in rats, rendering them interesting in treatment of nicotine addiction.[133] **CBT-ols**, in turn, are shown to exhibit a potent insecticidal activity against aphids.[128]

The biosynthesis of **CBT-ols** and **CBT-diols** takes place in the plastids of the glandular trichome cells of tobacco.[128] As common to all diterpenoids, the biosynthesis of **CBT-diol** starts with **GGPP**, which is converted into both **CBT-ol** epimers by the cembratrienol synthase (CBTS), a class I diterpene synthase. Subsequently, both **CBT-ols** are hydroxylated at C6 by a class II CYP450 monooxygenase that receives electrons from an unknown class II membrane bound FAD- and FMN- containing NADPH-dependent oxidoreductase (CPR).[128] Guo and Wagner proposed a basic mechanism of **CTB-ol** formation in 1995 (Figure 46).[134] However, the detailed reaction pathway of **CBT-ol** synthesis starting with **GGPP** and the native carbocation intermediates during the cyclization reaction in the tobacco CBTS are currently unknown.

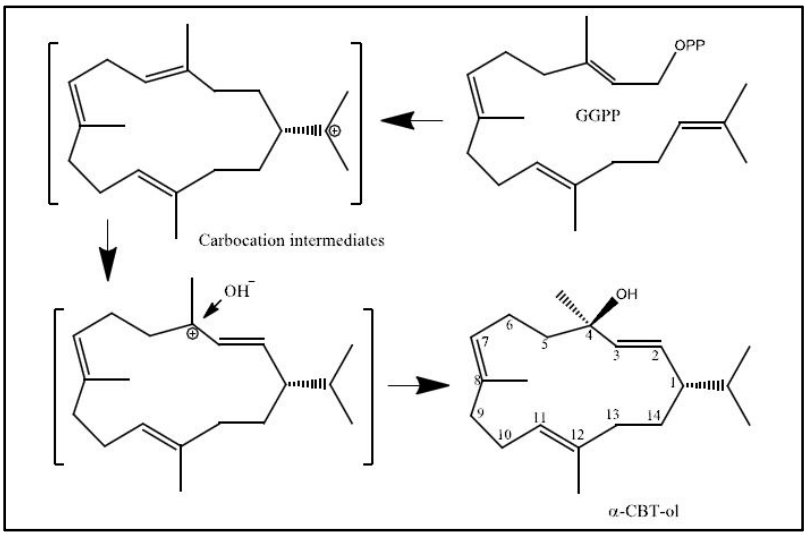


Figure 46: Proposed carbocation cyclization mechanism of N.tabacum CBTS

The gene of a putative diterpene synthase from common tobacco, *N.* tabacum that is proposed to be responsible for the cyclization of **GGPP** to both **CBT-ols** was recently found upon gene expression profile analyses of tobacco leaf trichomes.[135] Initial homology-based sequence searches using the BLAST heuristic algorithms suggested a high similarity to other tobacco-derived terpene synthases.[135] However, the functionality of the translated protein from this gene sequence has not been experimentally confirmed. A CBTS that could possibly represent the translated corresponding gene has also been partially purified from leafs of *N. tabacum* during the work of Guo and Wagner in 1995.[134] However, no sequence data nor structural or mechanistic data for their partially purified protein is available.

On the other hand, the membrane bound class II CYP450 monooxygenase responsible for hydroxylation of **CBT-ols** at C6 is known. This enzyme (CYP71D16) was identified upon preparation of *N. tabacum* gland-specific cDNA libraries in conjunction with BLAST homology searches.[128] The functionality of the monooxygenase was established by co-suppression and antisense strategies *in planta*.[128] The suppressed plants were shown to produce an increased amount of the **CBT-ol** precursors and, thus, to display a significantly higher aphicidal activity compared to the control plants.[128]

A biotechnological *de novo* production route for tobacco **CBT-ols** has been established in *N. benthamiana* using the known homologous CBTS from the wood-tobacco, *N. sylvestris.* There, sufficient amounts have been produced for subsequent NMR-spectroscopic analysis.[136] For the various bioactivities displaying **CBT-diols** no report for a biotechnological production route has been described.

2. Aims of this study

The first part of the work was to test the functionality of the previously identified gene for the putative CBTS from *N. tabacum*. Subsequently, an efficient recombinant heterologous biotechnological production route for the insecticidal **CBT-ols** are to be established. The second part was the structural *in silico* characterization of the CBTS via MM modelling techniques combined with targeted protein engineering as well as the deciphering of the discrete cyclization mechanism of this CBTS by QM gas phase calculations. The last part dealt with the computer-based search for a suitable eukaryotic CPR in order to establish **CBT-diol** production *in vivo.*

3. Results

Biotechnological production of non- and mono-hydroxylated cembranoids in *E.coli*

The gene of the putative CBTS from *N.tabacum* (GenBank: AAS46038.1) was introduced into the three plasmid diterpene production system on the pET-duet-1 vector (methods chapter II). Typically, plant class I diterpene synthases contain an N-terminal chloroplast leader sequence.[42] However, as the protein has not been structurally characterized, the complete gene sequence coding for a 589 amino acids protein was introduced into the diterpenoid producing strain and TXS was replaced by CBTS (methods chapter II and III). Five day shake flask cultures in M9CA minimal medium with glycerol as the sole carbon source yielded several diterpenoid compounds upon GC/MS analysis (Figure 47, 48). Subsequently, several flash chromatography and preparative TLC analytic steps of the organic extracts separated three of the four diterpenoid compounds, enough for structural characterization of the compounds for NMR spectroscopy (methods chapter IV).

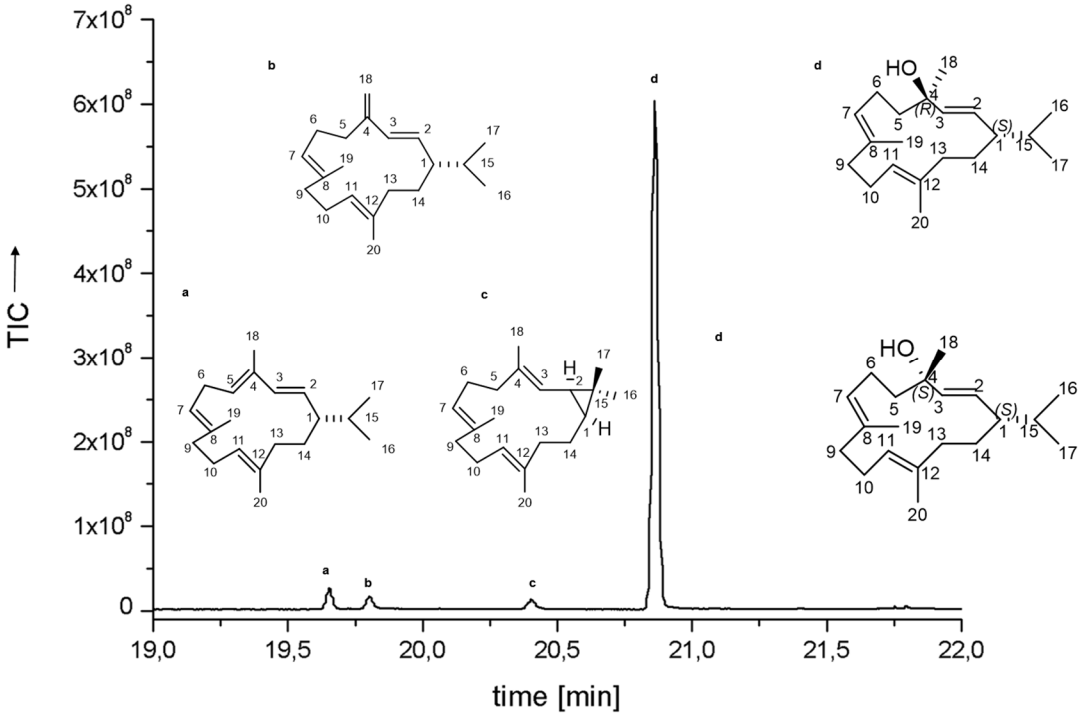


Figure 47: GC chromatogram of the shake flask experiments

Figure 47: GC chromatogram of the ethyl acetate extract of the five day shake flask experiments. **a**: **(+)-cembrene, b**: **(±)-isocembrene**, **c**: **(-)-casbene** and **d**: **α-** and **β-CBT-ol**.

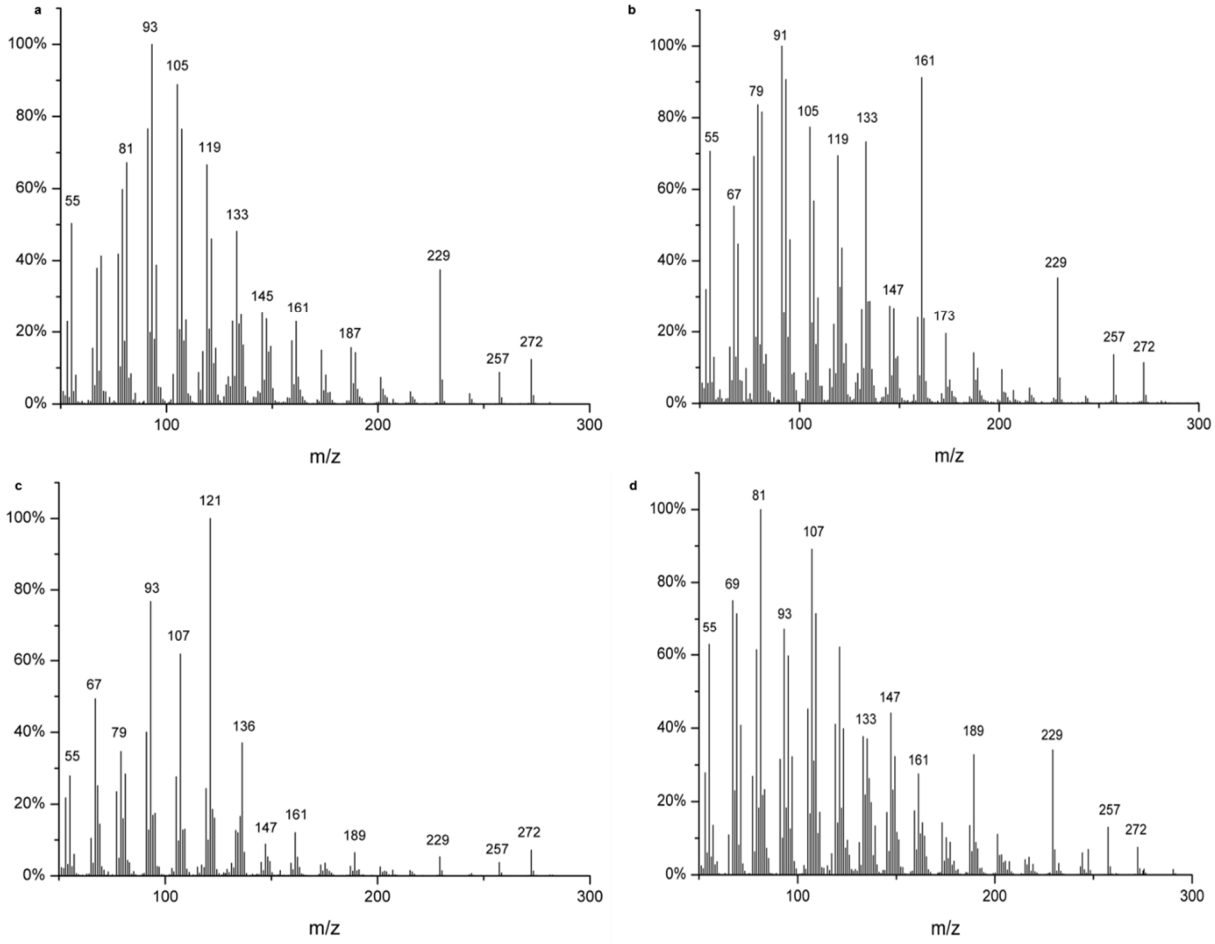


Figure 48: EI-MS spectra of the compounds from Figure 47

Figure 48a: EI-MS spectrum of **a**: **(+)-cembrene**, **b**: **(±)-isocembrene**, **c**: **(-)-casbene** and **d**: combined spectrum of **α-** and **β-CBT-ol**. For more information, see text.

NMR Spectral Data for **(+)-cembrene,** (**a**) [a]			
			(+)-cembrene [a]
#	**δc (ppm)**	**δ_H (ppm), *J* (Hz)**	**δc (ppm)**
1	48,1		48,1
2	131	5,20 (dd, J = 15,5, 9,8)	131,1
3	130,2	6,07 (d, J =15,5)	130,4
4	135,1		135,2
5	125,3	5,53 (t, J = 7,6)	125,4
6	26,1		26,1
7	126,4	5,14 (d, J = 10,6)	126,5
8	131,2		131,4
9	38,8		38,7
10	23,3		23,3
11	125,6	4,88 (d, J = 6,8)	125,7
12	132,5		132,6
13	36,4		36,4
14	27,6		27,6
15	32,7		32,6
16	20,7		20,8
17	19,9		19,9
18	19,8		19,9
19	14,3		14,4
20	14,3		14,4

Table 16: NMR spectral data for (+)-cembrene

Table 16: The chemical shifts of **(+)-cembrene** (**a**) are derived from 13C- (Appendix Figure 34) and 1H spectra (Appendix Figure 33). Spectra of **(+)-cembrene** have been measured in $CDCl_3$ ([a]). The right column depicts the published 13C chemical shifts.[137]

NMR Spectral Data for **(-)-casbene, (c)** [a]			
			(-)-casbene [a]
#	***δc* (ppm)**	***δ*$_H$ (ppm), *J* (Hz)**	***δc* (ppm)**
1	30,76		30,6
2	25,8		25,7
3	121,3	4,88 (d, J = 8,1)	121,2
4	135,6		135,3
5	39,6		39,5
6	25,09		24,9
7	123,5	4,90 (ddd, J = 10,7, 9,1, 3,2)	123,4
8	125,6		125,5
9	39,5		39,3
10	24,09		24,0
11	133,4	4,97 (t, J = 6,7)	133,2
12	136,2		136,1
13	40,5		40,3
14	24,09		24,0
15	19,95		19,8
16	29,01		28,8
17	15,88		15,7
18	15,92		15,7
19	16,3		16,3
20	16,6		16,4

Table 17: NMR spectral data for (-)-casbene

Table 17: The chemical shifts of **(-)-casbene** are derived from 13C- (Appendix Figure 36) and 1H spectra (Appendix Figure 35). Spectra of **(-)-casbene** have been measured in $CDCl_3$ ([a]). The right column depicts the published 13C chemical shifts.[138]

In comparison with NMR spectra from previous reports two of the diterpenoid compounds of the *E.coli*-CBTS extract were characterized as **(+)-cembrene** (**a**) and **(-)-casbene** (**c**) (Table 16, 17). Upon purification and GC/MS analysis both **CBT-ol** epimers could not be separated. Therefore, derivatization using TMS as silylating agent was used to separate both for GC/MS analyses (methods chapter III and IV). A comparison with a TMS-silylated thunbergol standard was able to identify the **α**- and **β-CBT-ol**-TMS derivatives in the GC-chromatogram. (Figure 49, 50). NMR-characterization of the purified combined **CBT-ol** extract confirmed the presence

of both epimers in comparison with the thunbergol authentic standard spectral data (Table 18, 19 and Appendix Figure 37- 40).

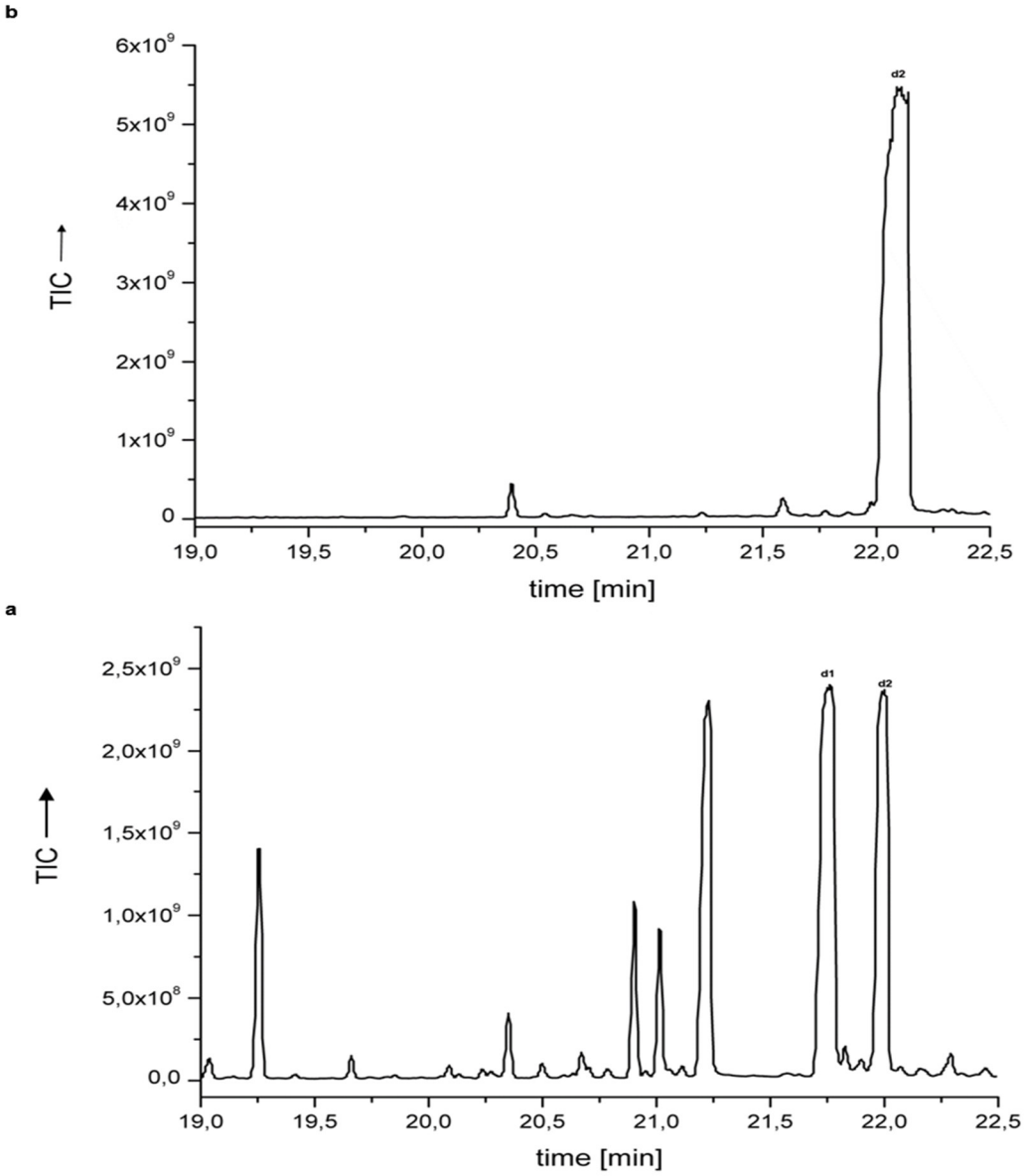


Figure 49: GC-chromatogram of the TMS-derivatives of α- and β-CBT-ol

Figure 49**a**: GC-chromatogram of the TMS-derivatives of the ethyl acetate extract of the five day shake flask experiments, **d1**: **β-CBT-ol**, **d2**: **α-CBT-ol** (thunbergol). **b**: GC-chromatogram of the TMS-derivative of the authentic thunbergol standard (1mg/L in ethyl acetate), **d2**: **α-CBT-ol** (thunbergol).

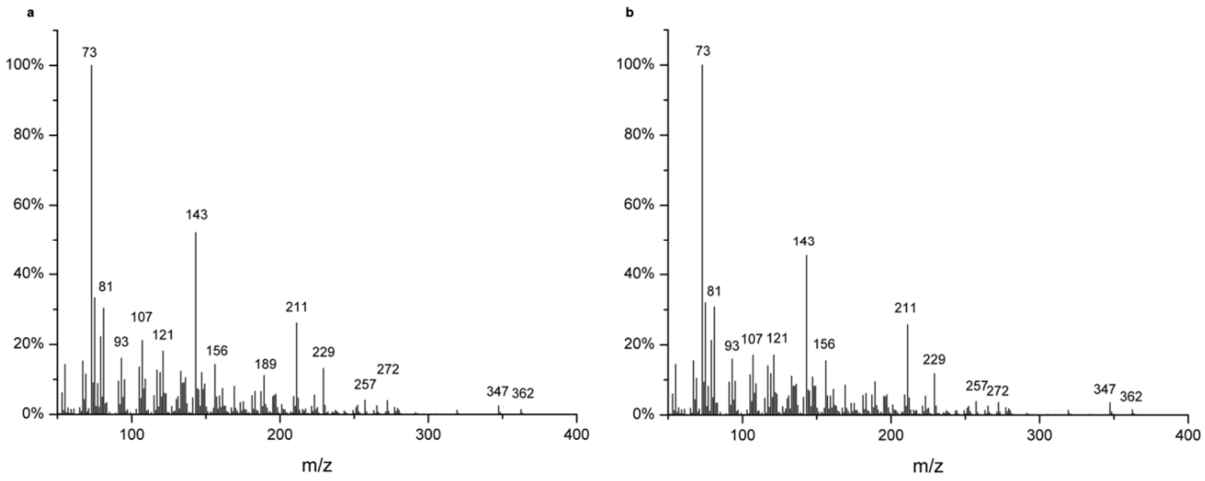


Figure 50: EI-MS spectra of TMS-α- and TMS-β -CBT-ol

Figure 50**a**: EI-MS spectrum of TMS-**α-CBT-ol** from the ethyl acetate extract and the authentic standard, **b**: EI-MS spectrum of the TMS-**β-CBT-ol** from the ethyl acetate extract.

NMR Spectral Data for 4-epiisocembrol (**β-CBT-ol**), **d1** [a]			
			4epi-isocembrol [a]
#	***δc* (ppm)**	***δ*$_H$ (ppm), *J* (Hz)**	***δc* (ppm)**
1	46		46,2
2	126,8	5,26 (dd, J = 15,4, 9,2)	126,9
3	138,7	5,56 (d, J = 15,4)	138,9
4	73,8		73,8
5	43,9		44,1
6	23,4		23,5
7	127,7	5,21 (m)	127,9
8	132,5		132,4
9	36,8		36,9
10	23,6		23,7
11	124,7	5,0-5,05 (m)	124,9
12	132,8		132,7
13	38,9		39,1
14	27,8		28,0
15	32,9		33,1
16	20,6	0,83 (d, J = 6,8)	20,7
17	19,3	0,79 (d, J = 6,8)	19,4
18	29,2	1,28 (s)	29,3
19	14,7	1,58 (s)	14,8
20	15,0	1,51 (s)	15,1

Table 18: NMR spectral data for β-CBT-ol (4-epiisocembrol)

Table 18: The chemical shifts of **β-CBT-ol** (**d1**) are derived from 13C- (Appendix Figure 38) and 1H spectra (Appendix Figure 37). Spectra of **β-CBT-ol** have been measured in $CDCl_3$ ([a]). The right column depicts the published 13C chemical shifts.[139]

NMR Spectral Data for thunbergol (α-CBT-ol), **d2** [a] and the authentic standard			
			thunbergol [a]
#	**δ_C (ppm)**	**δ_H (ppm), *J* (Hz)**	**δ_C (ppm)**
1	45,8		45,9
2	129,1	5,19 (dd, J = 15,4, 9,2)	129,0
3	137,9	5,70 (d, J = 15,4)	138,3
4	72,6		72,5
5	42,9		43,1
6	22,4		22,6
7	128,4	5,19 (m)	128,6
8	132,3		132,2
9	36,7		36,8
10	23,6		23,8
11	125	4,98-5,03 (m)	125,2
12	132,4		132,3
13	39,1		39,2
14	27,4		27,6
15	32,8		33,0
16	20,4	0,84 (d, J = 6,8)	20,4
17	19,4	0,81 (d, J = 6,8)	19,5
18	28,0	1,34 (s)	28,1
19	14,6	1,60 (s)	14,7
20	15,0	1,51 (s)	15,0

Table 19: NMR spectral data for α-CBT-ol (thunbergol) and the authentic thunbergol standard

Table 19: The chemical shifts of **α-CBT-ol** (**d2**) and the authentic standard are derived from 13C- (Appendix Figure 38 and 40) and 1H spectra (Appendix Figure 37 and 39). Spectra of **α-CBT-ol** and the authentic standard have been measured in $CDCl_3$ ([a]). The right column depicts the published 13C chemical shifts.[139]

The relative abundance of **α**- and **β-CBT-ol** produced by the CBTS in the five day shake flask experiments were determined using 1H-NMR spectroscopy (Figure 51, Appendix Figure 37). The analysis revealed a relative ratio of 1:0,3 **β-**:**α-CBT-ol** (Figure 51).

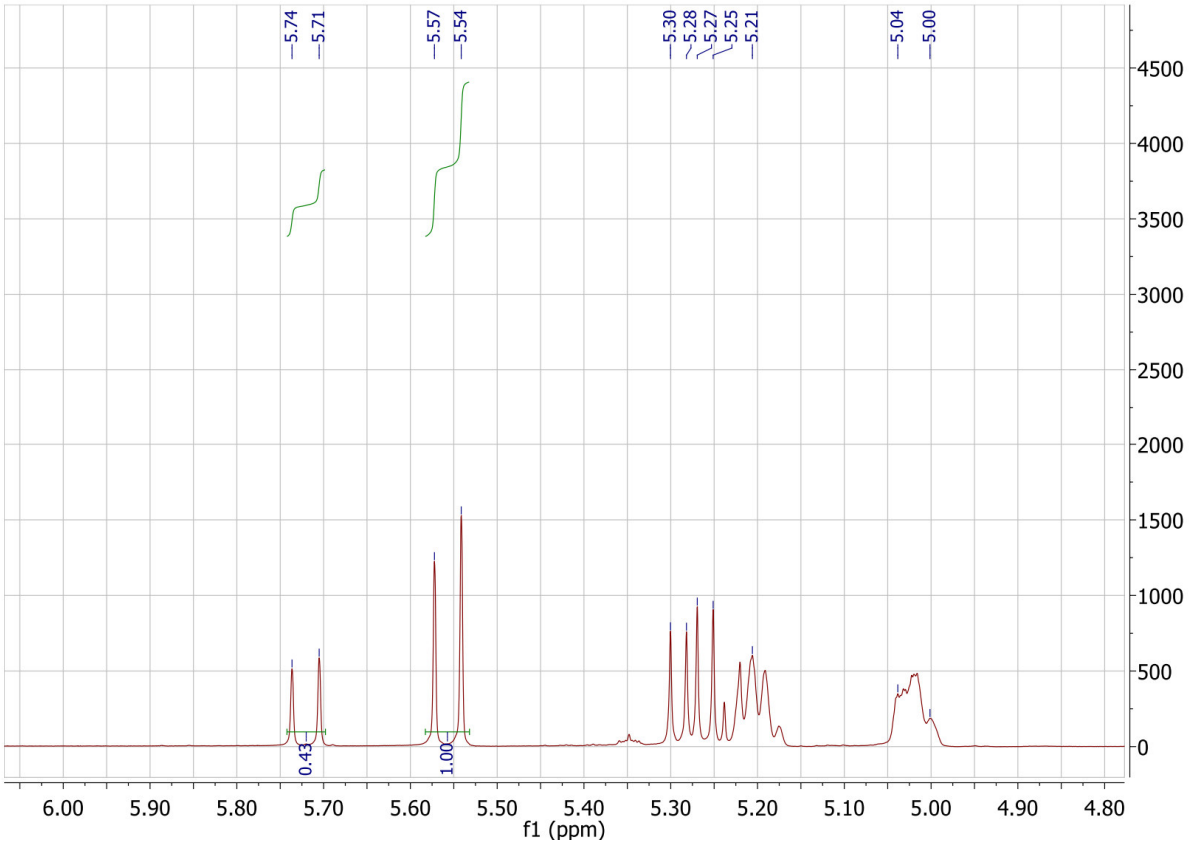


Figure 51: 1H spectrum (mid-field) of the combined purified CBT-ol extract

Figure 51: 1H-spectrum (mid-field) of the purified **CBT-ol** extract from the ethyl acetate extract shake flask experiment. The duplet at δ 5.71 corresponds to the olefinic H3 of **α-CBT-ol**, the duplet at δ 5.54 corresponds to the olefinic H3 of **β-CBT-ol**.

α- and **β-CBT-ol** has been shown to result in a mixture of **(+)-cembrene** (**a** in Figure 47, 48) and **(±)-isocembrene** (**b** in Figure 47, 48) upon dehydratation. Therefore, the compound that could not be separated from **(+)-cembrene** upon purification steps was assigned according to it mass spectrum as **(±)-isocembrene** which is in line with previous reports (Figure 48).[139,140]

The next step was the fermentative upscaling of the **CBT-ol** production using the controlled environment of a 5L reactor (Figure 52). A five day batch bioprocess in M9CA medium with glycerol as sole carbon source provided a final **CBT-ol** concentration of 120 mg/L (**α-CBT-ol**: 54 mg/L (**d2**), **β-CBT-ol**: 66 mg/L (**d1**)). By contrast, **(+)-cembrene** (**a**) was produced in a concentration of 6,23 mg/L, **(±)-isocembrene** (**b**) in a concentration of 3,74 mg/L and **(-)-**

casbene (**c**) in a final concentration of 1,88 mg/L, with α-humulene (**α**) as internal standard (Figure 53).

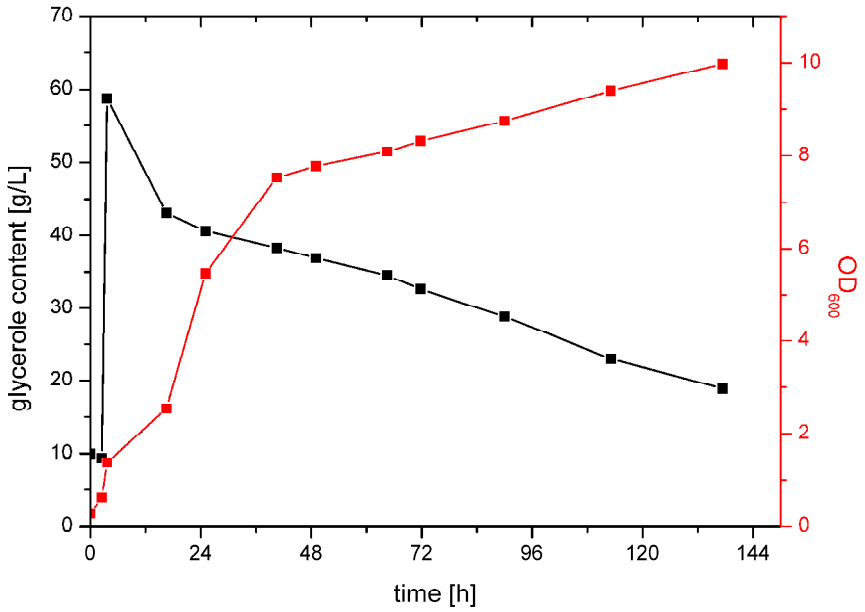


Figure 52: Fermentation characteristic of a 5L bioreactor experiment

Figure 52: Fermentation characteristic of a five day batch fermentation in a 5L bioreactor with glycerol as sole carbon source. The strain used was the same strain used for the experiment shown in Figure 47. X-axis depicts the time scale of the fermentation in min, the y1-axis depicts the glycerol content at the beginning and its decrease during the fermentation measured by HPLC, the y2-axis depicts the OD_{600} growth curve of the used *E.coli* BL21 (DE3) strain. For more information, see methods chapter IV.

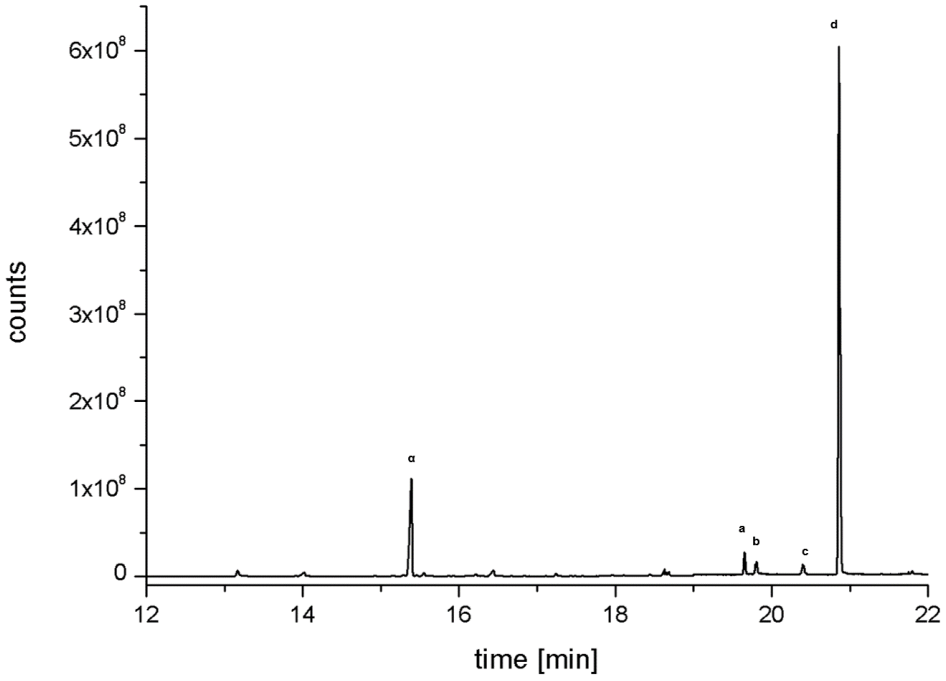


Figure 53: GC-FID analysis of the 5L bioreactor experiment

Figure 53: GC-FID analysis of the ethyl acetate extracts of the 5L bioreactor experiment. **a**: **(+)-cembrene**, **b**: **(±)-isocembrene**, **c**: **(-)-casbene**, **d**: combined **α-** and **β-CBT-ol**, **α**: internal alpha-humulene standard (1mg/L).

Structural and mechanistic characterization of the CBTS from *N.tabacum*

No structural information of a CBTS is currently available and only two plant diterpene synthases, taxadiene synthase and abietadiene synthase have been structurally characterized.[47,54] Therefore, a homology model of CBTS was constructed using the HHpred server tool (methods chapter III). Templates during homology search, secondary structure prediction and model building were chosen to be all structurally characterized pro- and eukaryotic mono-, sesqui- and diterpene synthases showing an e-value of 1.0 x10 $^{-120}$ during homology search and secondary prediction. The model of the CBTS from *N. tabacum* showed the typical fold of mono-functional plant class I terpene synthases according to Fig. 7 (Figure 54).[42] However, the *in silico*-derived model displayed the open conformation together with an unfolded N-terminal peptide (Figure 54). Bioinformatic analysis via the SignalIP server

(http://www.cbs.dtu.dk/services/SignalP/), able to predict cellular signal-peptides and the TMpred server (http://www.ch.embnet.org/software/TMPRED_form.html), able to search for transmembrane regions in given sequences do not show, in contrast to TXS, the existence of a chloroplast leader sequence or a transmembrane region in the protein sequence of CBTS. Recombinant expression of the full length CBTS in the pET28b-vector and subsequent Ni-NTA- and desalting chromatography (methods chapter IV) is in full accordance with the absence of transmembrane domains and leader peptides, as it showed a good expression rate for the full length synthase (Figure 55a).

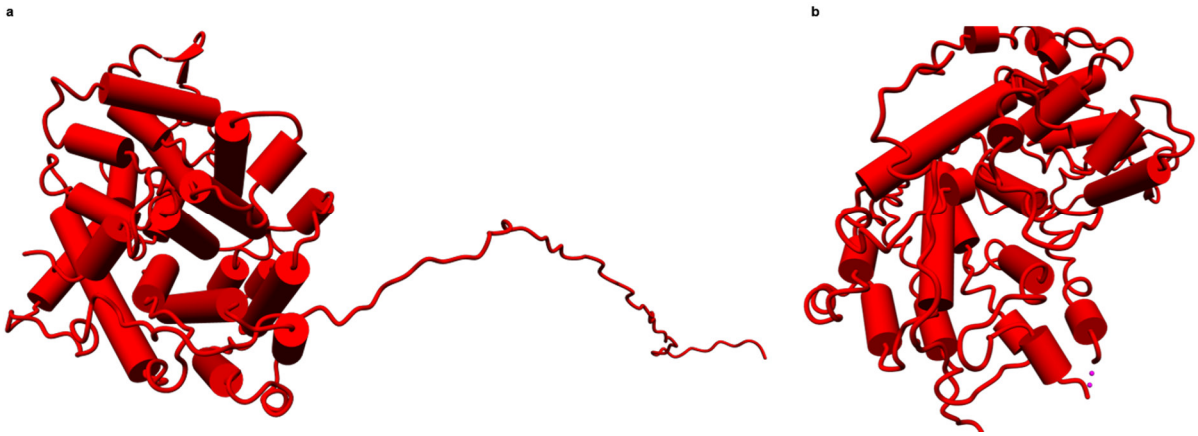


Figure 54: Modelled CBTS with leader peptide and the open conformation form

Figure 54**a**: Homology model of the full-length CBTS from *N. tabacum* containing the unfolded N-terminal peptide, **b**: the model from Figure 54**a** without the 50 amino acids N-terminal peptide used for further *in silico* modelling approaches.

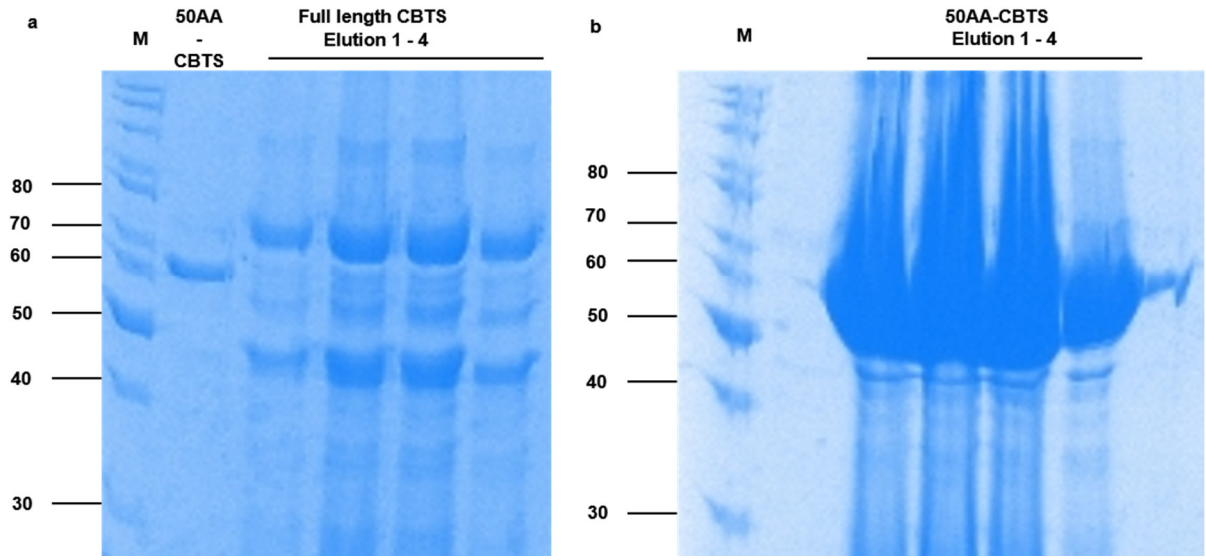


Figure 55: SDS-PAGE analysis of the desalted His-full-length CBTS and the His-tagged 50AA-truncated version

Figure 55**a**: SDS-PAGE analysis of the desalted his-tagged full length CBTS (lane 3-6, MW = 70.5 kilodalton) with the desalted his-tagged 50AA-CBTS truncation in comparison (lane 2, MW = 62.75kilodalton). **b**: SDS-PAGE analysis of the desalted his-tagged 50AA-CBTS truncation (lane 3-6).

The unfolded N-terminal region in CBTS contained the first 50 amino acids, whereas TXS exhibit an N-terminal unfolded peptide that consists of 79 residues. TXS is shown not to rely on these 79 residues, as a genomic truncation of the first 79 amino acids is still catalytically active. Therefore, during modelling of the closed complex of CBTS the first 50 residues were excluded (methods chapter II). However, to ensure that the 50 amino acids (50AA) truncated version is also soluble and active, the truncated version has also been recombinantly expressed using the pET28b-vector setting and has been purified by Ni-NTA affinity and desalting chromatography (Figure 55a, b). A structural alignment of the closed complex of TXS with that of CBTS exhibits an RMSD of 0.51Å over 321 aligned residues with 25.23% primary sequence identity, arguing for an identical fold and thus similar structure-function relationships (Figure 56).

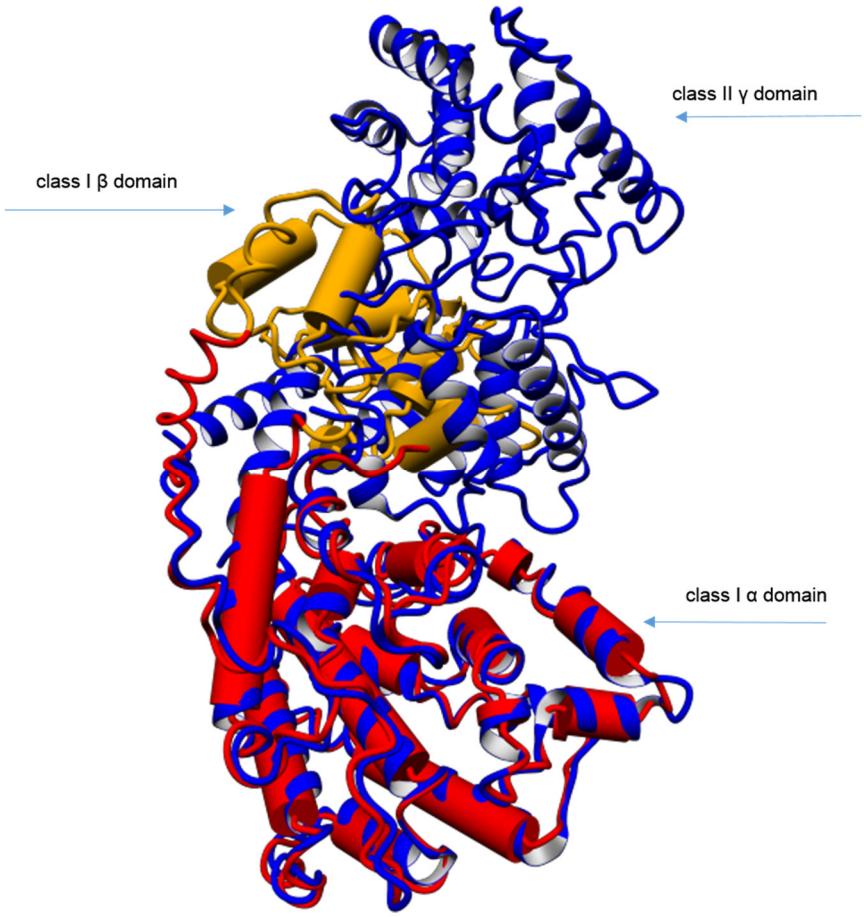


Figure 56: Structural alignment of the closed form of TXS with the closed form of CBTS

Figure 56: Structural alignment of the closed form of TXS (blue) with the modelled closed form of CBTS (50AA truncation) (red-ochre). The closed CBTS conformation depicts a typical fold for plant class I terpene synthases containing the catalytically active α class I domain (red) and the β class I domain (ochre).[42] TXS exhibits, besides both domains, another catalytically inactive γ domain harboring a class II terpene synthase domain.

The TXS:CBTS alignment moreover showed the structural analogy of the two magnesium binding motifs (Figure 57, 58a). In CBTS both motifs are housed on corresponding helices in analogy to TXS. According to the modelling of the closed complex of TXS, the productive TXS-derived **GGPP**, coordinated by the three Mg^{2+} ions and additional water molecules derived from the TXS crystal structure (methods chapter II) were modelled into the closed complex of CBTS. Subsequently the water molecules were integrated and dative bonds with the metal

binding motifs were constructed (methods chapter II). After several rounds of energy minimization and MD-simulation using the self-parametrizing YASARA2 force field, the **GGPP**-containing productive closed complex of CBTS was obtained, hereafter designated CBTS.

```
CLUSTAL

CBT_ol_Synthase_M50/1-549 ------------------------------------------------------------
TXS_M60/1-783             DDIPRLSANYHGDLWHHNVIQTLETPFRESSTYQERADELVVKIKDMFNALGDGDISPSA

CBT_ol_Synthase_M50/1-549 ------------------------------------------------------------
TXS_M60/1-783             YDTAWVARLATISSDGSEKPRFPQALNWVFNNQLQDGSWGIESHFSLCDRLLNTTNSVIA

CBT_ol_Synthase_M50/1-549 ------------------------------------------------------------
TXS_M60/1-783             LSVWKTGHSQVQQGAEFIAENLRLLNEEDELSPDFQIIFPALLQKAKALGINLPYDLPFI

CBT_ol_Synthase_M50/1-549 -----NRAMDLSSSSRHLADFPST-----------IWGD---------HFLSYNSEIT-
TXS_M60/1-783             KYLSTTREARLTDVSAAADNIPANMLNALEGLEEVIDWNKIMRFQSKDGSFLSSPASTAC

CBT_ol_Synthase_M50/1-549 -EITTQEKNEHEMLKEIVRKML----VETPDNSTQKLVLIDTIQRLGLAYHFNDEIENSI
TXS_M60/1-783             VLMNTGDEKCFTFLNNLLDKFGGCVPCMYSIDLLERLSLVDNIEHLGIGRHFKQEIKGAL

CBT_ol_Synthase_M50/1-549 QNIFNLSQNS-----EDDDEHNLYVAALRFRLARQQGYYMSSDVFKQFTNHDGKFKENHT
TXS_M60/1-783             DYVYRHWSERGIGWGRDSLVPDLNTTALGLRTLRMHGYNVSSDVLNNFKDENGRFFSSAG

CBT_ol_Synthase_M50/1-549 N---DVQGLLSLYEAAHMRVHDEEILEEALIFTTTHLESVIPN-LS--NSLKVQVTEALS
TXS_M60/1-783             QTHVELRSVVNLFRASDLAFPDERAMDDARKFAEPYLREALATKISTNTKLFKEIEYVVE

CBT_ol_Synthase_M50/1-549 HPIRKAIPRVGARKYIHIYENI-----------GTHNDLLLKFAKLDFNMLQKLHRKEL
TXS_M60/1-783             YPWHMSIPRLEARSYIDSYDDNYVWQRKTLYRMPSLSNSKCLELAKLDFNIVQSLHQEEL

CBT_ol_Synthase_M50/1-549 NELTSWWKDLDRANKFPYAKDRLVEAYFWTVGIYFEPQYSRSRSLVTKVVKMNSIIDDTY
TXS_M60/1-783             KLLTRWWKESGMA-DINFTRHRVAEVYFS--SATFEPEYSATRIAFTKIGCLQVLFDDMA

CBT_ol_Synthase_M50/1-549 DAYATFDELVLFTDAIQRWDEGAMDLLPTYLRPIYQGLLDVFNEMEEVLAKEGKADHIYY
TXS_M60/1-783             DIFATLDELKSFTEGVKRWDTSLLHEIPECMQTCFKVWFKLMEEVNNDVVKVQGRDMLAH

CBT_ol_Synthase_M50/1-549 AKKEMKKVAEVYFKEAEWLNANYIPKCEEYMKNGLVSSTGPMYGIISLVVMEEIITKEAF
TXS_M60/1-783             IRKPWELYFNCYVQEREWLEAGYIPTFEEYLKTYAISVGLGPCTLQPILLMGELVKDDVV

CBT_ol_Synthase_M50/1-549 EWLTNEPLILRAASTICRLMDDMADHEVEQQRGHVASFVECYMKEYGVSKQETYV-EMRK
TXS_M60/1-783             EKVHYPSNMFELVSLSWRLTNDTKTYQAEKARGQQASGIACYMKDNPGATEEDAIKHICR

CBT_ol_Synthase_M50/1-549 KITNAWKDINKELLRPT-AVPMFILERSLNFSRLADTFLKDDDGYTNPKSKVKDLIASLF
TXS_M60/1-783             VVDRALKEASFEYFKPSNDIPMGCKSFIFNLRLCVQIFYKFIDGYGIANEEIKDYIRKVY

CBT_ol_Synthase_M50/1-549 VESVDI
TXS_M60/1-783             IDPIQV
```

Figure 57: Primary sequence alignment of TXS with CBTS

Figure 57: Primary sequence alignment of the closed form of TXS with the closed form of CBTS. The black lines indicate the common magnesium binding motifs. The alignment was generated using CLUSTAL omega.

This complex exhibits a similar H-bond donor network, characterized by three arginine residues R316, R492 and R506 in analogy to TXS and other eu- and prokaryotic terpene synthases (Figure 14c, 58b). However, the two tyrosine residues, important for binding of remaining bulk water upon active site closure in TXS are missing. In contrast to TXS, casbene and isopimaradiene synthase that all depend on the tyrosine-assisted protection of their active site from premature water quenching, CBTS harbors two phenylalanine residues (F65, F570) on the analogous positions (Fig. 58b, c). These phenylalanine residues are chemically not able to

coordinate bulk water solvent in order to shield the active site. Furthermore, in TXS, N757 coordinates the terminal O-atom of **PP$_i$** via a hydrogen bridge of its amide group during the whole cyclization cascade (Figure 14). In CBTS the analogous position is filled by an aspartate residue D495. Hence, this aspartate is not able to form this interaction. The investigation of the closed conformation of CBTS indicates that D495 coordinates a water molecule that is derived from the remaining bulk solvent upon active site closure (Figure 58a-c). There are other examples of class I terpene synthases, where a coordinated water is bound in the active site, that is important for diphosphate-assisted deprotonation or is directly engaged in water capture by the carbocationic intermediates.[48]

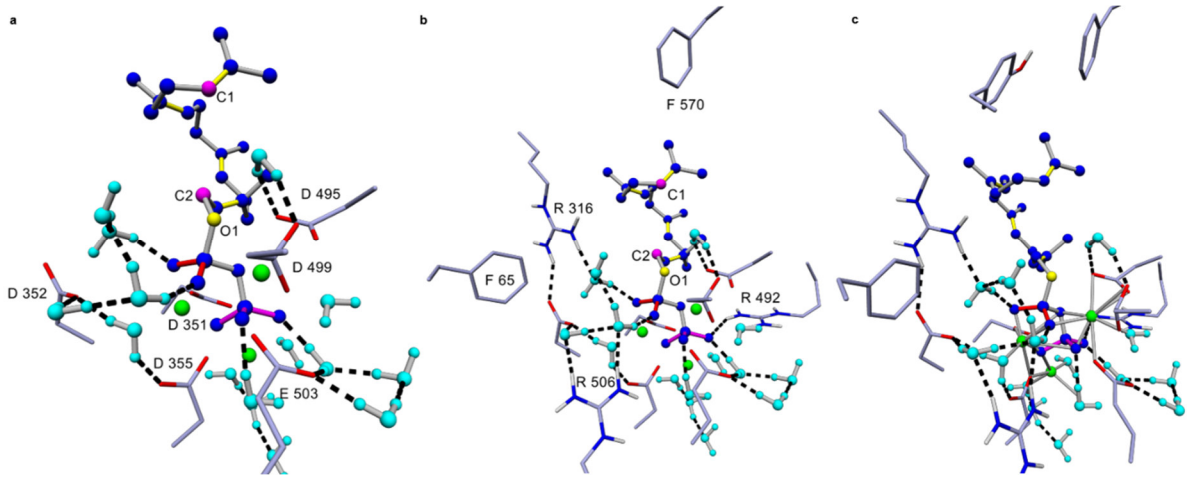


Figure 58: H-bond donor network and Mg^{2+} coordinating amino acids of CBTS

Figure 58**a**: Inner view of the modelled productive closed form of CBTS. The Mg^{2+} coordinating motifs (D351, D352 and D355) and (D495, D499 and E503) are shown. **GGPP** is colored in blue (C1, C2 in magenta, O01 of **PP$_i$** in yellow), Mg^{2+} ions are colored in green and bulk water is colored in cyan. **b**: The H-bond donor network of CBTS is shown, containing R316, R492 and R506 in analogy to TXS. CBTS harbors two phenylalanine residues (F65 and F570) instead of the two tyrosine residues Y89 and Y835 in TXS that protect the active site of premature water quenching). **c**: The dative bonds of the two Mg^{2+} coordinating motifs are shown as grey lines.

To consequently test the hypothesis that the *in silico* identified H-bond donor network displays a similar function compared to that of TXS, the network constituting amino acids were substituted with aliphatic, aromatic and other polar residues. Additionally, the proposed amino acid-

assisted mechanism controlling water-assisted quenching products were tested by substitutions of D495 and S452 that are supposedly important for water coordination in the active site. Further, as shown for TXS, π-π and cation-π stabilization of the carbocationic intermediates are important for product outcome. To test, whether CBTS depends on similar stabilization mechanism, despite a much simpler reaction mechanism, all aromatic residues spanning the active site were substituted by aliphatic, polar and other aromatic residues (Figure 59) and tested in an *in vitro* assay using the purified 50AA-truncated CBTS and **GGPP**, according to TXS (method chapter II).

Single substitutions of the three arginine H-bond donor network showed no activity upon incubation with **GGPP** during *in vitro* assays using purified enzymes (Appendix Table 3). Substitutions of aromatic residues, W323, F570 and Y577 to aliphatic and non-aromatic polar residues showed also no activity (Appendix Table 3). However, the substitution W323Y, F570Y and Y577F displayed a similar activity compared to the wild type CBTS (Appendix Table 3). As substitutions of single amino acids that are shown to be proximal to cationic substrate C atoms result in premature quenching products in TXS, it was consequently been tested, whether analogous substitutions in the active site of CBTS also lead to premature quenching products. The MM-based *in silico* model of CBTS harboring **GGPP** suggests that C491, L566 and V344 are in close proximity to **GGPP** and carbocationic intermediates in the active site (Figure 59). However, substitutions of L566H, W, A and S, V344K, M, L and S as well as C491A, S and Y displayed no activity in *in vitro* incubations with **GGPP**, although expression rates of these mutants were similar to the wild type protein. Only the C491L mutant showed a similar product profile as the wild type CBTS (Appendix Table 3). Substitutions of S452A, T and L as well as D495T, A and E were inactive. Interestingly, the D495N substitution showed a significantly altered product profile. This substitution did not show any hydroxylated products, however, the produced amounts of the hydrocarbon mono- and bi-cyclic compounds remained similar to their production rate in wild type CBTS (Appendix Table 3, Fig. 60b). This strongly argues for

the direct involvement of D495 in the control of water-assisted quenching products, as predicted *in silico*.

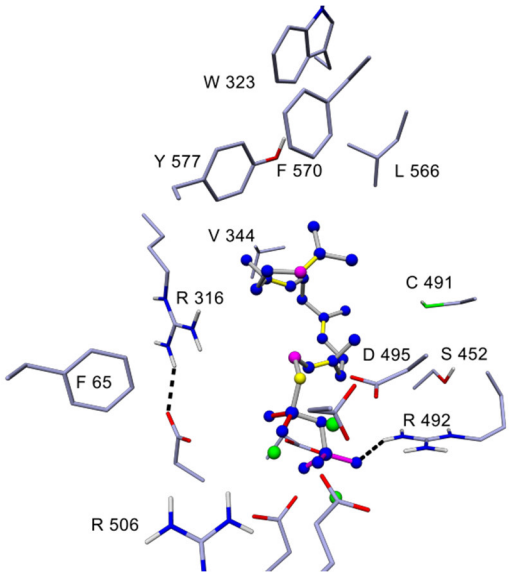


Figure 59: Structural representation of the mutated residues in CBTS

Figure 59: Graphical representation is according to Figure 58.

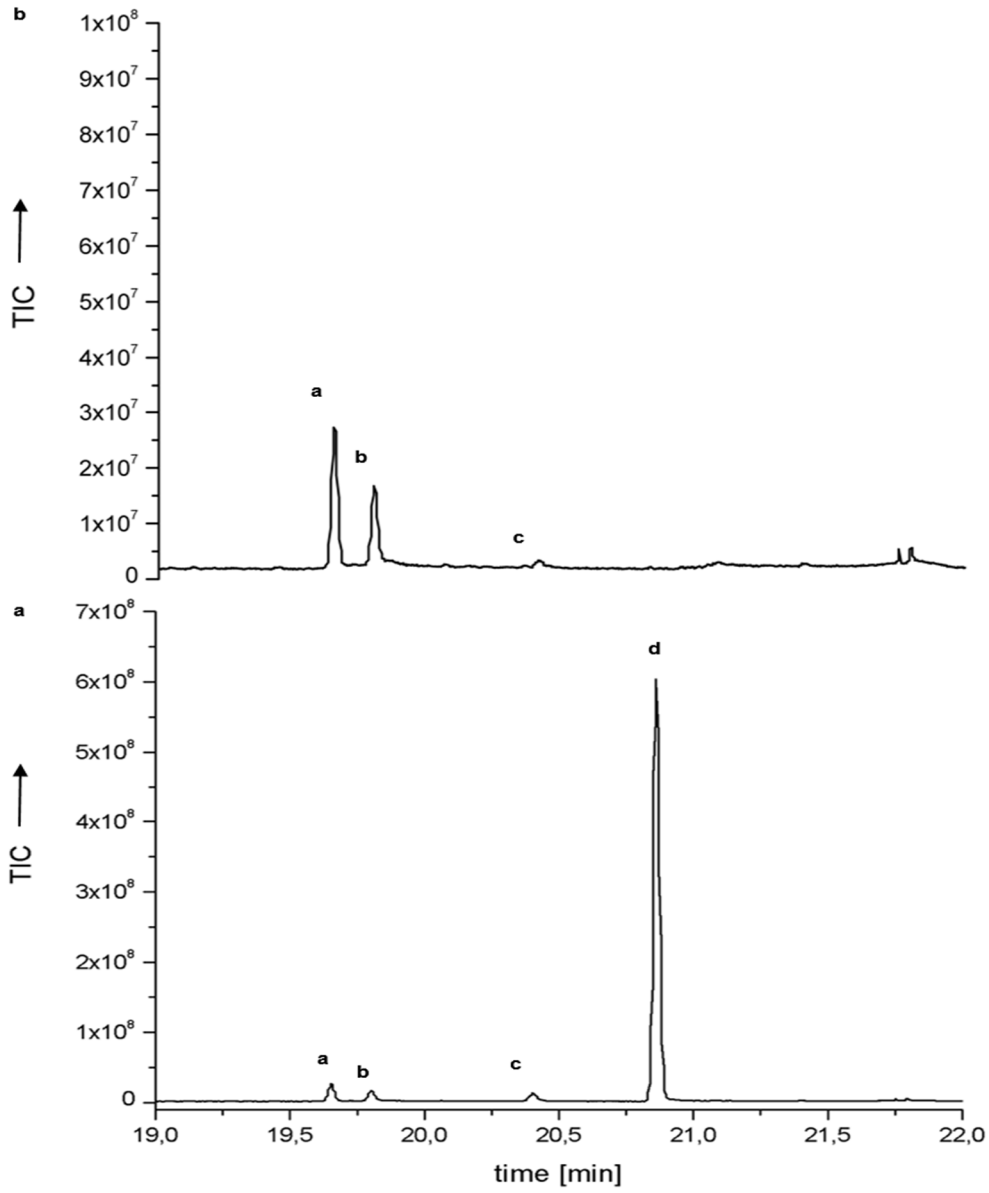


Figure 60: GC-chromatograms of wild type CBTS and CBTS-D495N

Figure 60**a**: GC-chromatogram of an ethyl acetate extract from the *in vitro* assay using purified 50AA-CBTS and **GGPP**. **b**: GC-chromatogram of an ethyl acetate extract from the *in vitro* assay using purified CBTS-D495N and **GGPP**.

Elucidation of the discrete cyclization pathway and the molecular basis of the amino acid-assisted water capture

A detailed cyclization mechanism of **CBT-ol** formation in CBTS including hydride shifts and the existence of carbocationic intermediates is not available, however a basic cyclization reaction was proposed (Figure 46).[134] As the *in vivo* experiments using the CBTS from *N. tabacum* demonstrated the existence of several catalytic non-hydroxylated by-products, the discrete catalytic carbocation pathway during formation of the monocyclic cembrane ring in conjunction with subsequent hydroxylation was elucidated with QM-gas phase modelling using various QM-DFT methodologies. The work was performed together with the group of Computational Biocatalysis at the TUM. The data demonstrates the detailed cyclization mechanism, the energetics of cyclization *in vacuo* and the hydrogen shifts leading to the formation of the cembra-4-yl cation (**cation C**) (Figure 61). QM gas phase calculations indicates an exothermic formation of **cation C** and additionally small energetic barriers between the different cations and transition states.

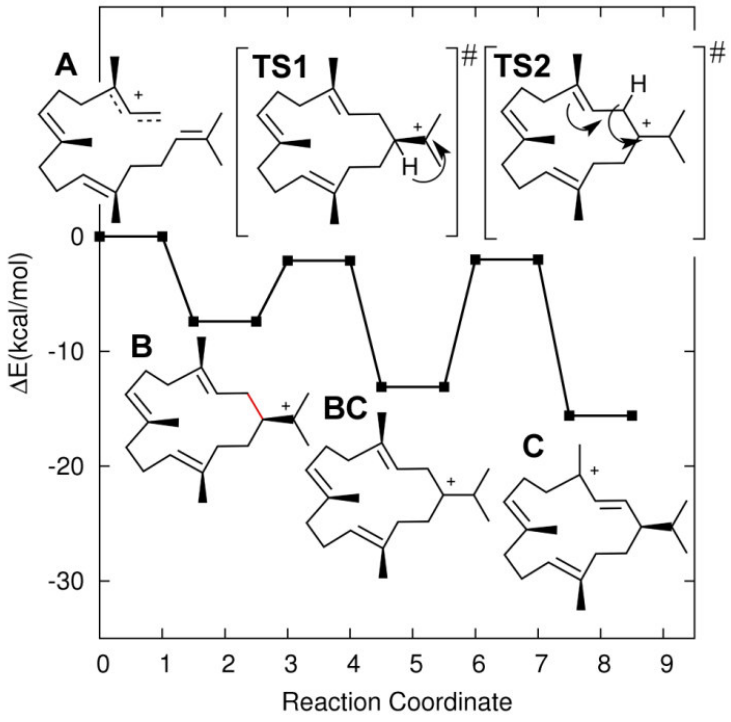


Figure 61: QM-gas phase calculation for the discrete cyclization pathway of CBTS

Figure 61: The figure shows the energetics of the consecutive stepwise hydrogen shifts from the geranylgeranyl cation (**cation A**, according to Figure 12) to the cembranyl-4-cation (**cation C**). The energetics propose that the cyclization of **GGPP** to **α**- and **β-CBT-ol**, derived from **cation C**, consists of four distinct intermediates (**cation A**, **B**, **BC** and **C**) as well as two transition state intermediates (**TS1** and **TS2**). Two productive hydrogen shifts during transition of **cation B** to **-BC** and **cation BC** to **-C** are needed to reach **cation C**.

The next step was the confirmation of the **CBT-ol** product distribution obtained in the *in vivo* experiments (Figure 51). Therefore the closed productive CBTS•**GGPP** complex obtained after MM-modelling was subjected to a DFT-QM calculation (Figure 62a). This approach indicates an almost equal distribution in the formation of **α-CBT-ol** (PR) and **β-CBT-ol** (PS) *in vacuo* (Figure 62b). In contrast, calculations of the formation rate of both compounds in the protein environment of the closed complex CBTS-QM model system indicate a significantly higher formation rate of **β-CBT-ol** (Figure 62b). This is in good accordance with the experimentally observed product formation, as shown in Figure 51.

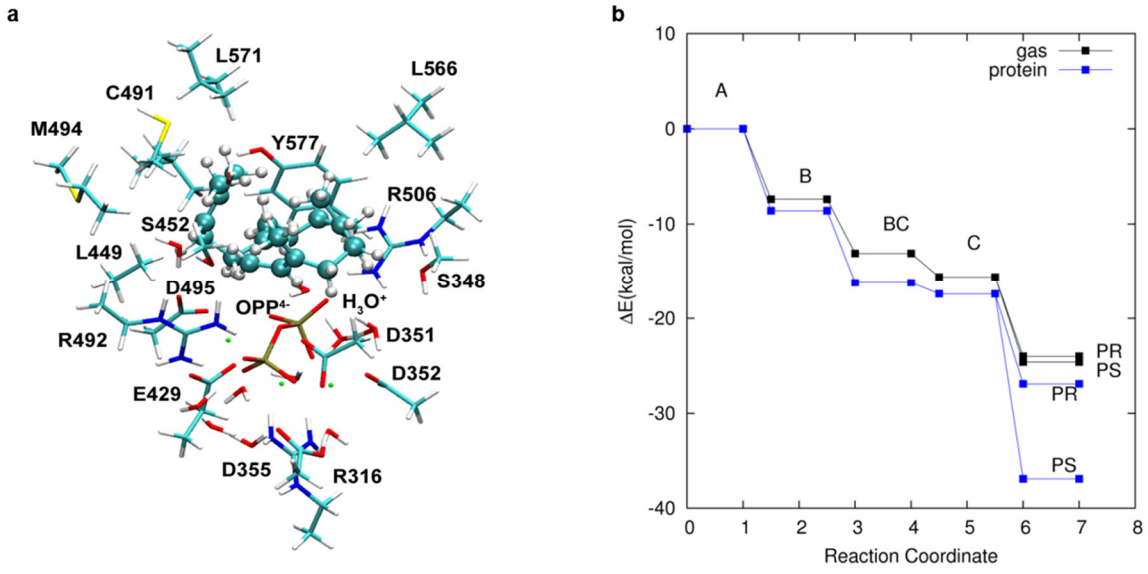


Figure 62: The structural model of QM system of CBTS and the calculation of product distribution

Figure 62**a**: The structural model of the QM system (281 atoms of the active site and **GGPP**) that is derived from the MM-based model of the closed productive conformation of CBTS (Figure 58). **b**: Calculation of the product

distribution of the **CBT-ols** in the gas phase (black line, PR: **α-CBT-ol**, PS: **β-CBT-ol**) and within the protein environment of the closed conformation of CBTS (blue line) described as the 281 atoms containing QM system shown in Figure 62a. The reaction coordinate is described according to Figure 61 following the exothermic path **cation A**-> **cation C**->**CBT-ols**.

The QM calculations within the protein environment suggest further a direct involvement of D495 and S452 in the proposed amino acid-assisted control mechanism leading to water-assisted quenching products (Figure 63). The calculations suggest that D495 is connected to S452 via a water and that D495 is a possible proton acceptor for the remaining water hydrogen upon hydroxylation at C4 leading to **α-CBT-ol** and **β-CBT-ol** formation. S452 seems further to be involved in the hydrogen bond network stabilizing both of the products (Figure 63). As **β-CBT-ol** stabilization by the hydrogen bond network is much more effective due to its availability to build more hydrogen bonds, based on the steric positioning of its hydroxyl group in contrast to **α-CBT-ol** (Figure 63), this is a proper explanation for the product distribution observed in *in vivo* experiments and *in silico* calculations.

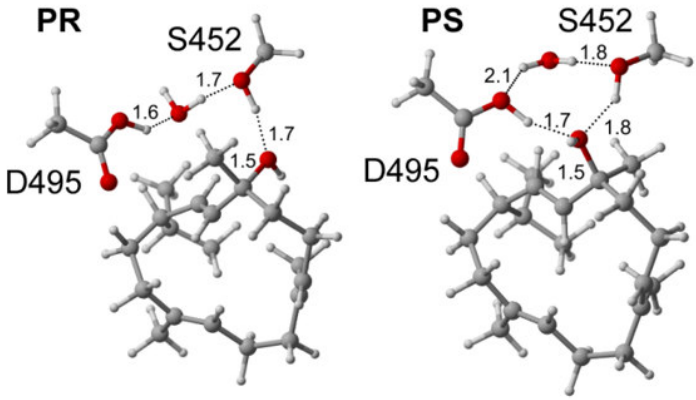


Figure 63: Proposed explanation for the observed CBT-ol product distribution

Figure 63: Relevant atoms of the optimized structures of **α-** and **β-CBT-ol** (PS and PR) and D495 and S452 of the QM system of CBTS are shown. Carbon, oxygen, and hydrogens are colored as grey, red and white, respectively. For detailed information, see text.

Biotechnological production of bioactive CBT-diols: *In silico* identification and experimental characterization of the native class II reductase from *N. tabacum*

Hydroxylation of non- and functionalized diterpenoids have been achieved using plant class II CYP450 monooxygenases.[18,141,142] However, the use of class II CYP450 systems is challenging as the monoxygenase and the reductase are membrane-bound.[100] Expression of these systems in bacterial hosts suffer from low product yields due to a significantly decreased solubility compared to class I and class VIII systems (Figure 10).[18] Engineering of the transmembrane regions of both partners is shown to increase their solubility in bacterial hosts.[37,143,144] However, the main problem using highly specific class II CYP450 systems for hydroxylation of terpenoids in general is the lack of native class II NADPH-dependent FAD- and FMN containing oxidoreductases (CPR). Recently, taxadiene hydroxylation at C5 has been achieved using protein engineering of the transmembrane region of the taxadiene 5-alpha-ol hydroxylase (t-CYP450) and its innate CPR (t-CPR) from *Taxus cuspidata*.[18] The native class II CPR of the known **CBT-diols** forming CYP450 monooxygenase (Genbank: AAD47832.1, n-CYP450)[128] is not available. Therefore, bioinformatics was used to find a suitable CPR in the genome of *N. tabacum* with the t-CPR as template. Sequences covering the whole genome of *Nicotiana tabacum* are not available. Concordantly, homology-based searches covering the limited genomic sequences was not successful. In contrast, the transcriptomic sequence information in form of expressed sequence tags, ESTs, is less limited for *N. tabacum*. Homology-based searches in the EST database (http://www.ncbi.nlm.nih.gov/nucest/) with t-CPR as template provided several hits (Figure 64). However, the identified ESTs do not cover the whole sequence of t-CPR. Instead, only parts of t-CPR could be aligned with differing ESTs from the *N. tabacum* transcriptome (Figure 64). These parts resemble the domains that are responsible for FAD, FMN and NADPH binding (Figure 64), according to a previous report.[145] The *N.tabacum* CPR (n-CPR) was therefore assembled from these differing EST hits, as all four used

TBLASTN hits constitute overlapping sequences (Figure 64, 65a). A structural homology model suggested further a proper folding of the assembled n-CPR (Figure 66).

The t-CPR constitutes a 74 N-terminal transmembrane helix that was truncated during expression in *E.coli* upon taxa-5-alpha-ol (single hydroxylation of taxa-4,5-11,12-diene at C5) production.[18] The assembled n-CPR do not contain a transmembrane helix, as the expressed sequence tags do not show any transmembrane consensus motif. Therefore, no transmembrane engineering was needed for soluble expression in *E.coli*.

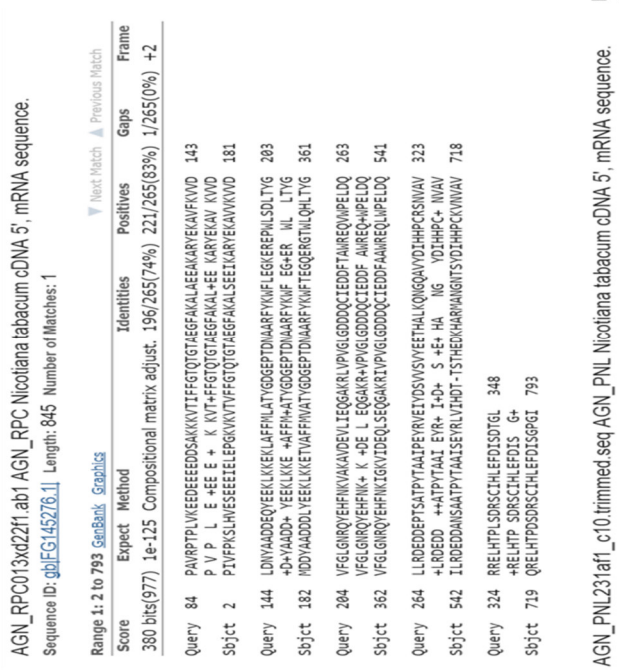


Figure 64: Primary sequence alignments of tobacco ESTs with the t-CPR template

Figure 64: Primary sequence alignment of the four overlapping tobacco ESTs with the t-CPR as template upon TBLASTN homology search. (Query: protein sequence of t-CPR, Sbjct: aligning sequences from the tobacco EST sequence database).

a

```
Global Protein alignment.    Mol 1 Taxus Red full length (1 to 717)    Mol 2  N.tabacum CPR (1 to 634)
Scoring matrix:  BLOSUM 62

  Sequence View:  Similarity Format, Color areas of high matches at same base position

Taxus Red full l       1 mqansntvegasqgkslldisrldhifalllngkggdlgamtgsaliltensqnlmilttalavlvacvf
N.tabacum CPR            ------------------------------------------------------------------------

Taxus Red full l      71 ffvwrrggsdtqkpavrptplvkeedeeeeddsakkkvtiffgtqtgtaegfakalaeeakaryekavfk
N.tabacum CPR          1 ----------makpivfpkslhvepeeeielepgkvkvtvffgtqtgtaegfakalseeikaryekavvk

Taxus Red full l     141 vvdldnyaaddeqyeeklkkeklaffmlatygdgeptdnaarfykwflegkerepwlsdltygvfglgnr
N.tabacum CPR         61 vvdmddyaadddlyeeklkketvaffmvatygdgeptdnaarfykwftegqergtwlqhltygvfglgnr

Taxus Red full l     211 qyehfnkvakavdevlieqgakrlvpvglgdddqcieddftawreqvwpeldqllrdeddeptsatpyta
N.tabacum CPR        131 qyehfnkigkvideqlseqgakrivpvglgdddqcieddfaawretlwpeldqllrdeddeptsatpyta

Taxus Red full l     281 aipeyrveiydsvvsvyeethalkqngqavydihhpcrsnvavrrelhtplsdrscihlefdisdtgliy
N.tabacum CPR        201 aipeyrlvihdtaas-hddkhasmangntsydihhpcqvnvavqrelhtpesdrscihlefdicgtgisy

Taxus Red full l     351 etgdhvgvhtensietveeaakllgyqldtifsvhgdkedgtplggsslpppfpgpctlrtalaryadll
N.tabacum CPR        270 etgdhvgvyaenpeeiveeaarllgqsldlifsihtdkedgtalgg-slpppfpgpctlraalarfadll

Taxus Red full l     421 npprkaaflalaahasdpaeaerlkflsspagkdeysqwvtasqrslleimaefpsakpplgvffaaiap
N.tabacum CPR        339 npprkatlvalaahaaepseaeklkflaspqgkdeysawvvasqrsllevmaefpsakpplgvffaavsp

Taxus Red full l     491 rlqpryysissprfapsrihvtcalvygpsptgrihkgvcsnwmknslpseethdcswapvfvrqsnfk
N.tabacum CPR        409 rlqpryysissprfaparihvtcalvygptptgrihkgvcstwmknavpleksqncssapifirpsnfk

Taxus Red full l     561 lpadsttpivmvgpgtgfapfrgflqeraklqeageklgpavlffgcrnrqmdyiyedelkgyvekgilt
N.tabacum CPR        479 lpadnkvpiimigpgtglapfrgflqerlalkkegaelgpavlffgcrnrqmdyiyqeeldnfleagals

Taxus Red full l     631 nlivafsregatkeyvqhkmlekasdtwsliaqgggylyvcgdakgmardvhrtlhtivqeqesvdsskae
N.tabacum CPR        549 elvvafsregpnkeyvqhkmsekaadiwnmisqgggyvyvcgdakgmardvhrtlhtiaqdqgsldsskae

Taxus Red full l     701 flvkklqmdgrylrdiw
N.tabacum CPR        619 glvknlqttgrylrdv-
```

b

MAKPIVFPKSLHVEPEEEIELEPGKVKVTVFFGTQTGTAEGFAKALSEEIKARYEKAVVKVVDMDDYAADDDLYEEKL
KKETVAFFMVATYGDGEPTDNAARFYKWFTEGQERGTWLQHLTYGVFGLGNRQYEHFNKIGKVIDEQLSEQGAK
RIVPVGLGDDDQCIEDDFAAWRETLWPELDQLLRDEDDEPTSATPYTAAIPEYRLVIHDTAASHDDKHASMANGN
TSYDIHHPCQVNVAVQRELHTPESDRSCIHLEFDICGTGISYETGDHVGVYAENPEEIVEEAARLLGQSLDLIFSIHTDK
EDGTALGGSLPPPFPGPCTLRAALARFADLLNPPRKATLVALAAHAAEPSEAEKLKFLASPQGKDEYSAWVVASQRS
LLEVMAEFPSAKPPLGVFFAAVSPRLQPRYYSISSSPRFAPARIHVTCALVYGPTPTGRIHKGVCSTWMKNAVPLEKS
QNCSSAPIFIRPSNFKLPADNKVPIIMIGPGTGLAPFRGFLQERLALKKEGAELGPAVLFFGCRNRQMDYIYQEELDN
FLEAGALSELVVAFSREGPNKEYVQHKMSEKAADIWNMISQGGYVYVCGDAKGMARDVHRTLHTIAQDQGSLDS
SKAEGLVKNLQTTGRYLRDV

Figure 65: Primary sequence alignment of the assembled n-CPR with t-CPR and domain representation of the n-CPR

Figure 65**a**: Primary sequence alignment of the n-CPR assembled from the four overlapping EST contigs from Figure 64 (*N. tabacum* CPR) with the template t-CPR (Taxus Red full) using CLUSTAL omega (http://toolkit.tuebingen.mpg.de/clustalw). Identical residues are colored in orange. **b**: Primary sequence representation of the assembled FAD- (blue), FMN- (red) and NADPH-binding (purple) domains from the newly identified class II NADPH-dependent FAD- and FMN containing diflavin oxidoreductase from *N. tabacum* (n-CPR).

Figure 66: Homology model of the n-CPR

Figure 66: Homology model of the n-CPR after homology modelling using YASARA structure (methods chapter II). The prosthetic FAD is shown in blue, prosthetic FMN in purple and the bound NADPH cofactor is shown in grey.

To establish **CBT-diol** production, the n-CYP450 had to be solubly expressed in *E.coli*. The *Taxus*-CYP450 (t-CYP450) was shown to comprise an N-terminal 42 residue transmembrane region (TM).[18] However, to establish taxa-5-alpha-ol production, only a 24 residue truncation of the TM instead of a truncation covering the whole 42 residues, and the additional insertion of a bovine leader peptide at the N-terminal end of the CYP450, shown to increase solubility of class II CYP450 monooxygenases in *E.coli*[143], leaded to the efficient production of this hydroxylated taxane derivate.[18] In contrast, the n-CPR comprised only a 19 N-terminal residue transmembrane sequence upon TMpred analysis. A subsequent primary sequence alignment between t-CYP450 and n-CYP450 established the homologous truncation region in n-CYP450

(Figure 67). Therefore, in n-CYP450 only the first 4 amino acids were substituted by the MALLLAVF bovine leader peptide without further truncations in the TM region, the truncated CYP450 hereafter designated as nt-CYP450. Using the established setting of the four plasmid system (methods chapter II and III), the engineered nt-CYP450 was introduced in MCS I of the pACYC-duet-1 vector. To establish optimal expression conditions in *E.coli* the assembled CPR (n-CPR) was genomically fused to the nt-CYP450 via an N-terminal n-CPR GSTSSGSGAS-linker peptide as a similar genomic construct was used for a soluble expression of the *T. cus-pidata* class II CYP450 system.[18]

Three day, 500ml shake flask production in M9CA medium with glycerol as the sole carbon source demonstrated the production of **CBT-diols** and thus the functional reconstitution of the n-CYP450 with the assembled n-CPR (Figure 68a). Subsequent purification steps using flash column chromatography 1D- und 2D-TLC allowed the separation of the two produced major **CBT-diols** from the ethyl acetate extract of the shake flask culture (Figure 68b, c). Separate GC-MS analysis of authentic **α**- and **βCBT-diol** standards identified the GC peaks at 22.37 min and 22.33 min as **α**- and **ß-CBT-diol** in the shake flask culture extracts (Figure 69a, b and Appendix Figure 41a-d). For NMR-spectroscopic identification of the produced **CBT-diols** the yields of the shake flask cultures have been too low. However, **α-CBT-diol** seems to be the preferred substrate of the CYP450 monooxygenase of *N. tabacum* (n-CYP450), as the relative titer of **α-CBT-diol** (**e**) is increased compared to the **β-CBT-diol** titer (**f**), which is in line with a previous report (Figure 68a).[128]

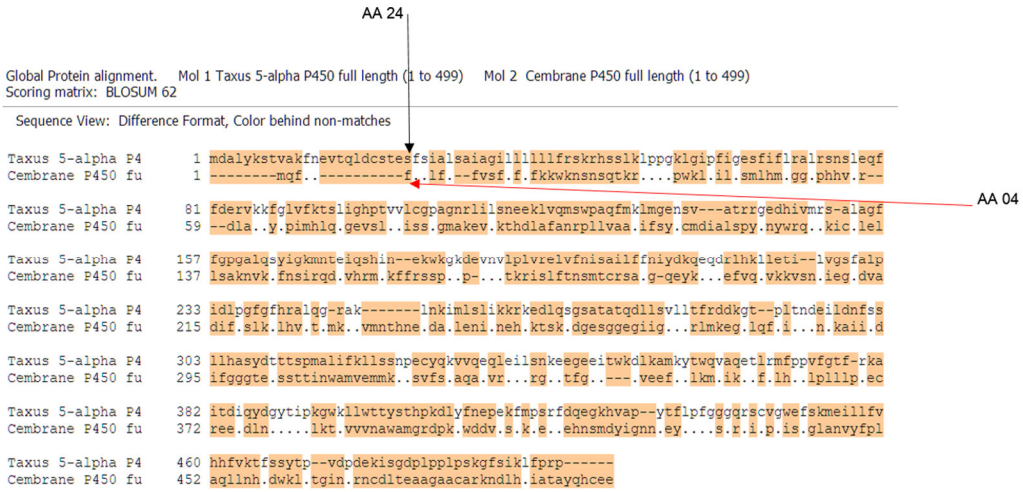


Figure 67: Primary sequence alignment of t-CYP450 with n-CYP450

Figure 67: Primary sequence alignment of the t-CYP450 (Taxus 5-alpha) with the n-CPR (Cembrane P450) using CLUSTAL omega (http://toolkit.tuebingen.mpg.de/clustalw). Identical or similar residues are colored in orange. The black arrow depicts the 24th amino acid of the t-CYP450, the red arrow the corresponding 4th amino acid of the n-CYP450, at which position the bovine leader sequence was introduced. For detailed information, see text.

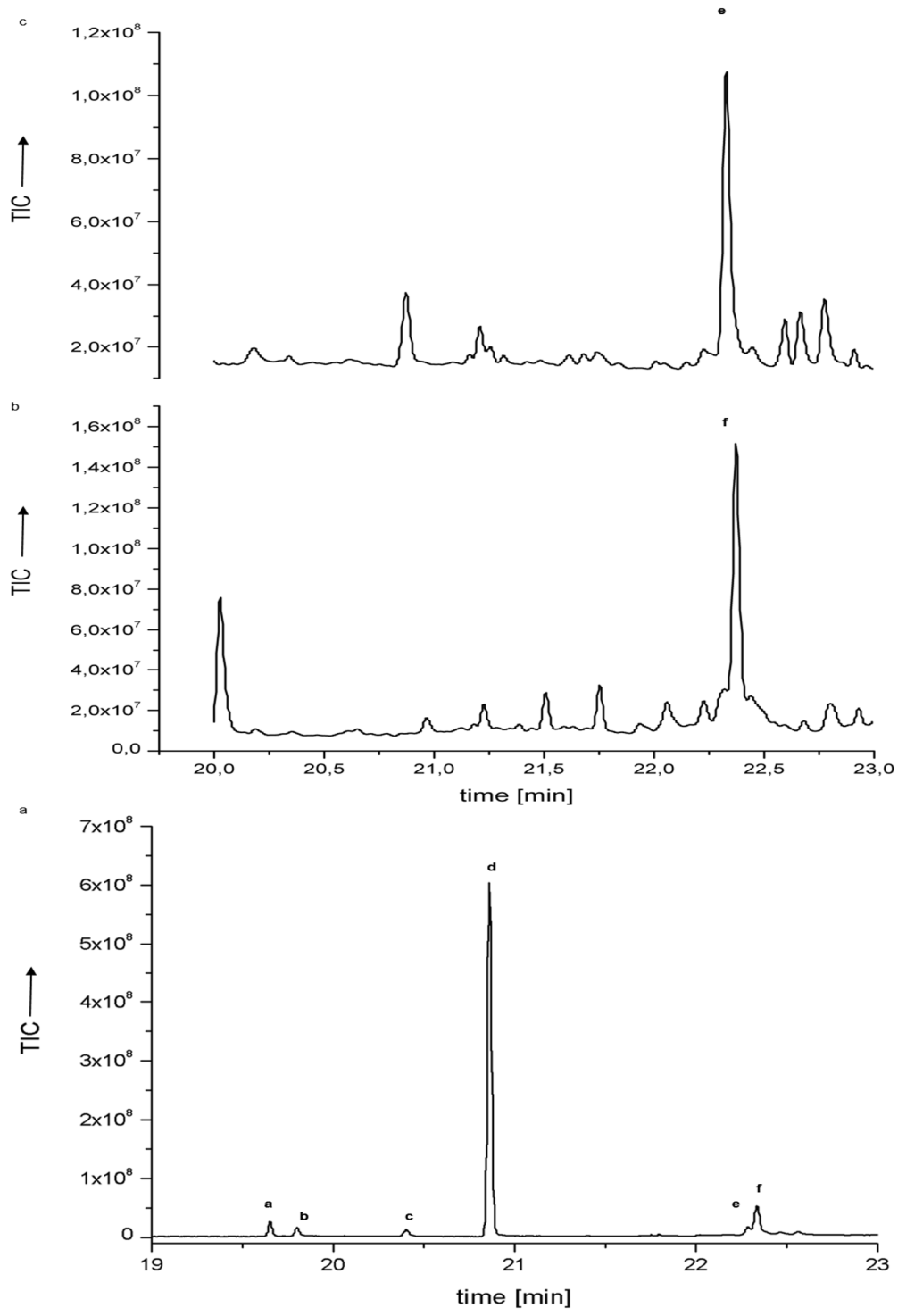


Figure 68: GC-chromatograms of the CBT-diol extract and purified CBT-diols

Figure 68**a**: GC-chromatogram of the ethyl acetate extract of the four plasmid system harboring the DXP pathway on the pCola- and pCDF-duet-1 vectors, the **CBT-ol** upstream pathway on the pET-duet-1 vector and the nt-CYP450-n-CPR fusion construct on the pACYC-duet-1 vector (**a**: **(+)-cembrene**, **b**: **(±)-isocembrene**, **c**: **(-)-**

casbene, **d**: **α**- and **β-CBT-ol**, **e**: **α-CBT-diol** and **f**: **β-CBT-diol**.) **b**: GC-chromatogram of the purified **α-CBT-diol** (**e**), RT: 22.37 min. **c**: GC-chromatogram of the purified **β-CBT-diol** (**f**), RT: 22.33 min.

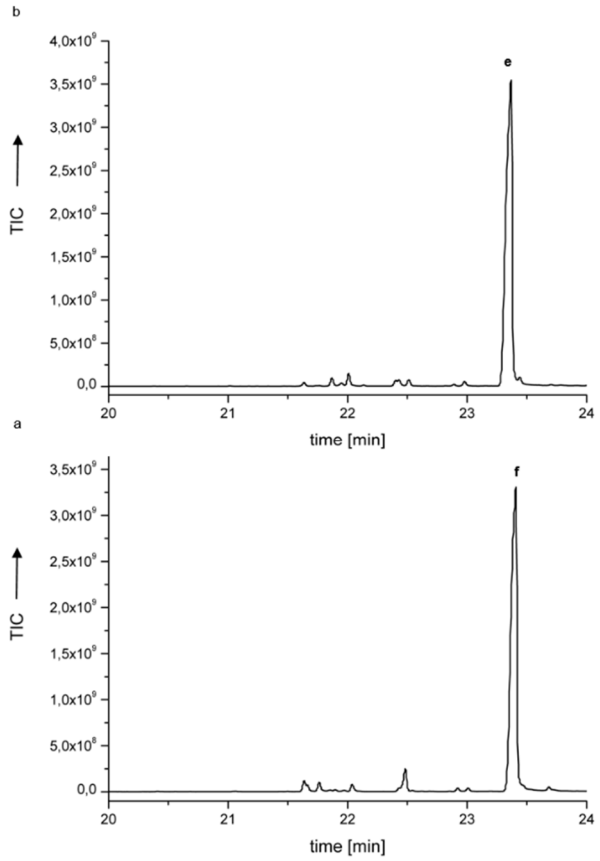


Figure 69: GC-chromatograms of the authentic α- and β-CBT-diol standards

Figure 69**a**: GC-chromatogram of an authentic **α-CBT-diol** standard, RT: 23.41 min. **b**: GC-chromatogram of an authentic **β-CBT-diol** standard, RT: 23.37 min. Note that for the **CBT-diol** standards a new DB-5 GC column was used. For all other spectra an already used shortened DB-5 GC-column was used.

4. Discussion

Tobacco cembranoids display a broad range of biological activities. The non-functionalized olefins ((**(+)-cembrene**, **(±)-isocembrene** and **(-)-casbene**) were shown to exhibit cytotoxic activity against various human cancer cell lines by interaction with tubulin at the so-called colchicine binding site.[146] The mono-functionalized **CBT-ols** were shown to significantly enhance the aphicidic resistance of tobacco plants.[128] The bi-functionalized **CBT-diols** display a high variety of biological activities including cytotoxicity, neuroprotectivity, antimicrobial and antifungal as well as insecticidal activity.[34,130] Efficient *de novo* production of these monocyclic compounds that are not efficiently accessible by classical chemical synthesis have not been shown in bacterial hosts.

Here, an efficient biotechnological production route for tobacco cembranoids in *E. coli* was established. Using a simple three plasmid system non- and mono-functionalized cembranoid bioactives have been produced in good yields. Especially, high titer production of **CBT-ols** with up to 120mg/L yield using glycerol as the sole carbon source is a good sustainable alternative compared to *in planta* production.[136] The most striking result of this study is the fact, that one diterpene synthase, CBTS, is able to produce five different cembranoid bioactives as major or by-products. *In silico* characterization of the *N. tabacum* CBTS and QM gas phase calculations of the discrete cyclization pathway from **GGPP** to **α**- and **β-CBT**-ol further paved the way to the understanding of the structural basis for the existence of the three major non-functionalized by-products.

Additionally, the structural basis for the observed distribution of **α**- and **β-CBT-ol** produced by CBTS could be established by QM gas phase modelling and verified by comparable NMR spectroscopy. These results further corroborate that *in silico* modelling of a closed complex with a homology model as template and subsequent QM calculations are able to get deep insights into the highly dynamic processes of a diterpenoid synthase in absence of a crystal structure.

The *in silico* data combined with targeted mutagenesis demonstrated further the existence of an H-bond donor network that acts in stabilization of **PP_i** similar to its function in TXS, also predicted from *in silico* modelling. Targeted *in silico*-guided protein engineering, however, was not able to enhance production of one or more of the by-products. Homologous substitutions of the respective amino acids resulted in a similar product profile, whereas non-homologous substituted enzymes showed no **GGPP** cyclization activity. This indicates that the native decorative coating of the active site of CBTS is exclusively designed to produce monocyclic diterpenoid structures. The data indicates further that **(+)-cembrene**, **(±)-isocembrene** and **(-)-casbene** are derived from initial **GGPP** misfolding, imprecise barrier crossing and substrate tumbling during the cyclization cascade in CBTS, in line with structural investigations of TXS, *vide supra*. The closed complex models of CBTS harboring **cations B** and **-C** further suggest that, in accordance with TXS, the bi-functional CBTS-motif depicted by R316-**PP_i** is the prime candidate for the premature deprotonation of the non-hydroxylated cembranoids. This argues again for a common **PP_i** stabilization and deprotonation mechanism in plant diterpene synthases. Premature deprotonation by this bi-functional motif at **cation B** stage would lead to the formation of **(-)-casbene** (Figure 70 a, b). Cation tumbling during the cascade and/or initial **GGPP** misfolding would result in the premature deprotonation at C4 or C20 of **cation C** by R316-**PP_i** that lead to the formation of **(+)-cembrene** or **(±)-isocembrene** instead of formation of the **CBT-ols** (Figure 70c).

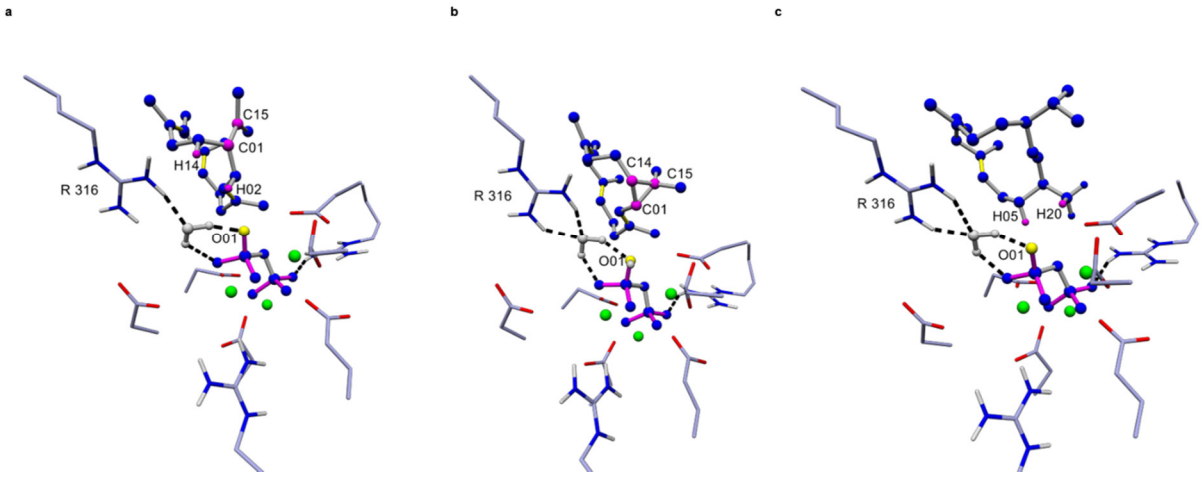


Figure 70: Graphical representation of CBTS harboring cation B, (-)-casbene and cation C

Figure 70**a**: Closed conformation of CBTS harboring **cation B**, according to Figure 61. Coloring is according to Figure 50. **b**: Closed conformation of CBTS harboring **(-)-casbene** that could be derived from **cation B** of Figure 70**a**, if premature deprotonation at H02 by the R316-**PP$_i$** bi-functional motif occurs, due to **GGPP** misfolding or cation tumbling. A subsequent H14 -> C02 hydride shift accompanied with nucleophilic attack of cationic C15 at the then cationic C14 would enable the formation of the dimethyl-cyclopropyl moiety of **(-)-casbene**. **c**: Closed conformation of CBTS harboring **cation C**. Premature deprotonation of H05 upon cation tumbling by the R316-**PP$_i$** bi-functional motif would lead to the formation of **(+)-cembrene**, premature deprotonation of H20 upon cation tumbling during the cascade by this motif would lead to the formation of **(±)-isocembrene**.

The D495N substitution was shown to only produce non-hydroxylated cembranoids in a similar distribution compared to the wild type CBTS. The observed outcome of the D495N mutant strongly corroborates the *in* silico-derived assumption that this residue is directly engaged in binding a water molecule. The closed complex model of the D495N mutant shows that the amide group of N495 is directly engaged in ion pairing with **PP$_i$** due to its close proximity and is thus presumably not able to act in the coordination of a remaining bulk water molecule. Combined with QM calculations of the closed complex this data shows an amino acid-assisted quenching control mechanism for functionalized cembranoids in CBTS.

In addition to the biotechnological production of non- and mono-functionalized cembranoids,

the four plasmid system containing the n-CYP450-n-CPR redox couple was able to produce di-hydroxylated cembranoids. This is the first report of a biotechnological production route for these highly bioactive compounds. Application of transmembrane signal engineering and bio-informatics were crucial to produce **α**- and **β-CBT-diols**. However, both compounds were only produced in minor amounts. A few examples for diterpenoid hydroxylations using class II CYP450 systems have been achieved.[18,141,142] In all cases production rates are very low compared to the production rate of the non-functionalized precursors. The main limitation for pathway reconstruction in bacterial hosts is the expression levels of the CYP450 monooxygenase and its CPR redox partner protein. Natively, both redox partners are integrated into the ER or thylakoid membrane in their respective organism. In addition, the metabolic pathway that produces the non-functionalized cyclized precursors is often located in a different cellular compartment, thereby reducing the metabolic cellular stress. Transmembrane signal engineering is a helpful methodology, however, expression in eukaryotic cells or plant tissues would provide better alternatives. Imitation of class VIII bacterial CYP450 systems by combining both, the CYP450 monooxygenase with its CPR counterpart in one protein, aiming at faster catalytic turnover rates, is another helpful instrument to achieve efficient electron transfer processes, as applied study.[61] However, these class II CYP450 systems containing CYP450 one-peptide complexes are often characterized by a molecular weight of over 120kDa that dramatically limits its soluble bacterial expression. Concordantly, biotechnological production of **CBT-diols** using a bi-cistronic construct with the n-CYP450 monooxygenase and the identified n-CPR in different multiple cloning sites harbored by the pACYC-duet-1 vector is shown to yield slightly higher **CBT-diol** titers (Master thesis Dòra Balogh).[147]

The main limitation for the efficient production of multi-functionalized diterpenoids in bacterial hosts is, however, the simple fact that CYP450-mediated decorated compounds are usually more toxic than its respective olefinic precursors.[61] Therefore, the choice of a eukaryotic host in combination with a more detailed transmembrane engineering and the balancing of the metabolic flux through the up- and downstream pathway of cembranoid biosynthesis may result in

an efficient biotechnological production of the two **CBT-diol** epimers. Alternatively, a class I P450 system could be used that is characterized by a better solubility in bacterial hosts, *vide supra*. A justified try to establish an efficient production of both **CBT-diol** epimers would be the application of the aforementioned AfR-Afx-CotB3/4 system (chapter III). Additionally, a site-directed mutagenesis approach could be applied to the active site of CotB3/4, based on *in silico* protein design and/or molecular docking studies of both **α**- and **β**-CTB-ol epimers into the active site of CotB3/4.

5. Methods

General experimental procedure

See methods chapter II. The authentic thunbergol standard was obtained from Apin Chemicals, UK, in a concentration of 50 mg/L. **α**- and **β-CBT-diol** standards were obtained from Biomol, Germany, both in a concentration of 10 mg/L.

Expression and purification of CBTS

See methods chapter II. The pET28b-vector (Novagen) was used for expression of the full length *cbts* (restriction sites NdeI, XhoI, Appendix section 6). Primer used were 5`-AAACATATGAGCCAGAGCATTAGTCCGCTGATTTG-3` and 5`-AAACTCGAGATCAACGCTTTCAACAAACAG-3`.

Site directed mutagenesis

See methods chapter II. Used primer are depicted in Appendix Table 5.

In vitro assay of CBTS with GGPP

See methods chapter II.

In vivo assay of CBTS

For the *in vivo* production of tobacco-derived cembranoids the three plasmid system (methods, chapter II and III, pCola-, pCDF- and pET-duet-1 vector) was used. Instead of TXS, this plasmid system contained the full length *cbts* gene (Appendix section 6) on the pET-duet-1 vector in MCS II (NdeI, XhoI). Instead of LB medium, M9CA (M9 minimal media with 4g/L casamino acids (Merck, Germany) was used. All other conditions are based on methods in chapter II.

Batch bioprocess of CBTS in the three plasmid system

A 5L-biorector (Braun, Melsungen) and M9CA medium with 40g/L glycerol was used. All other

conditions are based on methods in chapter II.

Extraction and Purification of the CBTS-derived diterpenoids

For the diterpene isolation, cells were pelleted (15 min, 17,500 g, 4 °C). The supernatant was extracted 3 times by 200 mL ethyl acetate. Organic phases were combined, dried with $MgSO_4$ and evaporated under vacuum to dryness. The crude extract was solved in 1 mL ethyl acetate and analysed by GC-MS and -FID. Purification of **(+)-cembrene**, **(±)-isocembrene** and **(-)-casbene** were carried out by flash chromatography. An isocratic 90/10 hexane/ethyl acetate silica step (Silica gel 40, Sigma-Aldrich,) was followed by a 1D-TLC with 90/10 hexane/ethyl acetate as mobile phase. For the purification of both **CBT-ol** epimers by flash chromatography, the hexane/ethyl acetate solvent during the 1D-TLC run was changed to 50/50 followed by a 70/30 hexane/ethyl acetate mobile phase.

Modelling of the closed complex of CBTS

See methods chapter II. Modelling of the carbocationic intermediates in the active site was performed according to the TXS-derived carbocationic intermediates shown in methods, chapter II, based on the proposed CBTS carbocation mechanism.

QM gas phase calculations and QM calculation in the protein environment

For gas phase QM calculations of the cyclization cascade of CBTS the B3LYP-D3/def2-TZVP and B3LYP-D3/def2-SVP DFT-hybrid functionals with their according basis sets were used. For calculations within the 281 atom containing QM system of CBTS the B3LYP-D3/def2-TZVP and BP68-D3/def2-SVP hybrid functionals with their according basis sets were used.

Cloning of the nt-CYP450-n-CPR fusion construct

The first step of the cloning procedure was the introduction of the synthesized nt-CYP450 from *N. tabacum* (Appendix section 6) containing the MALLLAVF bovine leader sequence at 5´-end and the GSTSSGSGAS linker at the 3`end that includes a Nhe1 restriction site into the pET-28b vector (NcoI, NheI, Appendix section 6). Primer for this integration were 5`-AAAAC-CATGGCACTGCTGCTGGCAGTTTT-3` and 5`-AAAAGCTAGCACCGC-TACTGAGCTGGTGCT-3`. Cloning, transformation, plasmid preparation and sequencing was conducted using standard cloning procedures. The second step was the genomic assembling of the n-CPR gene (Appendix section 6) onto the nt-CYP450 gene in the pET28b vector using the NheI restriction site at 5`end and a HindIII restriction site combined with a stop codon at the 3`-end. Primer for this approach were 5`-AAAAGCTAGCATGGCCAAAC-CGATTGTTTTTC-3`and 5`-AAAAAGCTTGCTCCAAACGTCACGCAGATAA-3`. After the confirmation of the productive fusion of nt-CYP450 and n-CPR (Appendix Section 6) by a standard sequencing procedure (see General experimental procedures, chapter II and III) the fusion construct was integrated into MCS I of the pACYC-duet vector using the NcoI restriction site at the 5` end and the HindIII restriction site at the 3`end. The primer used were 5`-AAAAC-CATGGCACTGCTGCTGGCAGTTTT-3` and 5`-AAAAAGCTTGCTCCAAACGTCACGCAGA-TAA-3`.

***In vivo* assay for the production of CBT-diols**

The four plasmid system was used according to methods, chapter II (pCola-, pCDF-, pET- and pACYC-duet-1 vector) with the pACYC-duet-1 vector harboring the codon optimized fusion protein consisting of the membrane-engineered n-CYP450 monooxygenase (Genbank: AF166332.1, *n-cyp450,* synthesized and codon optimized at Thermo Fischer, GeneArt, Regensburg, Appendix section 6) and the newly assembled n-CPR (gene: *n-cpr*, synthesized and codon optimized at Thermo Fischer, GeneArt, Regensburg Appendix section 6*)* combined via

a 5`- GSTSSGSGAS -3` linker peptide according to a previous report (Appendix section 6).[18] The 500 mL shake flask production was performed according methods, chapter III with M9CA medium instead of LB medium.

Extraction and purification of CBT-diols

The cultures were centrifuged (5000 rpm, 15 min). Supernatants were discarded and the cell pellets were washed in 100 ml H_2O and centrifuged again (4500 rpm, 25 min). The cells were suspended in 30 ml H_2O and disrupted by sonification (4 x 2 min, 5 x 10 % intensity, 50 % power). The lysates were extracted twice with 20 ml ethyl acetate and centrifuged at 15000 rpm for 15 min. Organic phases were combined and evaporated to dryness using a rotary evaporator. Remaining ethyl acetate was removed under a continuous N_2 stream. The powder was dissolved in 1 ml ethyl acetate and subjected to a silica phase flash column chromatography with a 95/5 hexane/ethyl acetate isocratic elution to remove olefinic and single hydroxylated compounds. The mobile phase was subsequently changed to 70/30 hexane/ethyl acetate and both major **CBT-diol** epimers eluted as a combined fraction. This fraction was the further purified by a 2D-TLC. The sample was evaporated to dryness and solved in 100 % ethyl acetate. 10 % of this sample was used for the stained control plate. 90 % of the sample was filled on the edge of a preparative TLC plate. The plates were first subjected to an isocratic gradient of 90/10 hexane/ethyl acetate and after the run was finished horizontally subjected to an isocratic 70/30 hexane/ethyl acetate run. Both **CBT-diol** spots were finally analyzed by GC/MS analysis.

Chapter V: Concluding remarks and outlook

Terpenoids are the largest class of microbial and plant secondary metabolites. The range of bioactive activities of these family of complex compounds is overwhelming, rendering them valuable targets for the pharmaceutical industry. However, the successful sustainable production of functionalized terpenoids with their large variety of terpene hydrocarbon core cycles and diversified functional chemical decoration is limited. There are a few important exceptions like the biotechnological production of artemisinin, whose production demanded detailed massive scientific efforts over decades as prerequisite.[13,17,148,149] On the other hand, enormous efforts have been undertaken and will go on to elucidate novel terpene synthases by next-generation sequencing approaches, homology-based searches in a variety of bacterial and plant genomes and high-throughput experimental methodologies. Similar massive efforts have been undertaken and will go on in the field of metabolic engineering of microbial hosts to optimize the efficient production of known and novel complex terpene hydrocarbon macrocycles.[22,23] As the downstream pathway for terpenoid production remains the same for all types of terpenoids, the upstream pathway, in contrast, varies for each terpene macrocycle, the big challenge is and will be to optimize the interplay between both.

Another challenge is the efficient biocatalytic functional decoration of the produced terpene macrocycles. Limited information with respect to decorating enzyme complexes is available and therefore only a few terpene macrocycles have been functionally *de novo* decorated today.[18,100,148] The described disadvantages of class II P450 systems, the large molecular weight of class VIII P450 system imitations and the missing of native and/or at least effective non-native redox partner proteins for class I and class II P450 systems makes the efficient functional decoration of various terpene macrocycles a tremendous challenge for the biotechnological future.

Therefore, the methodologies described in this study could help to simplify the search for novel terpene macrocycles by *in silico* predictions and design of essential catalytic active site amino

acids. The modelled closed complex of taxadiene synthase and the observed mechanistic picture driving the complex cyclization cascade can generally be applied for a better understanding of the cyclization mechanisms of any class I terpene synthase. This leads to a significant increase within the design horizon of class I terpene synthases usable for the generation of novel unidentified or still inaccessible terpene macrocycles. The method of closed conformation complex modelling in combination with MM-based carbocation modelling can be applied to any eukaryotic class I terpene synthase open conformation structure, closed conformation structure harboring substrate analogues or high-class homology model that comprises a functional class I associated alpha domain. The closed conformation model of TXS can in this respect serve as a well-characterized model enzyme. During the course of the experimental work, the modelling of the closed conformation of the bacterial class I diterpene synthase, cyclooctat-9-en-7-ol synthase (CotB2) has been additionally achieved using the open conformation crystal structure as template.[45] The proposed carbocationic intermediates of CotB2 have been subsequently modelled according to the established methodology for TXS.[84] The closed conformation model of CotB2 can thus further serve as a template model for any microbial terpene synthase that comprises a functional class I associated alpha domain.

The identified and *de novo* produced terpene and terpenoid structures, *endo*-verticillene, cembrene A, (+)-cembrene, (±)-isocembrene, (-)-casbene as well as the produced hydroxylated compounds, cyclooctatin and its intermediates as well as the CBT-ols and –diols, could moreover serve as core structures for the generation of further functionalized, bioactive terpenoids. Especially, the generated bi-cyclic verticillene-type hydrocarbon macrocycles possess a high degree of bioactivity when decorated with one or two hydroxyl groups.[82,83] Biocatalytic conversion of this core structure by the cembranoid class II P450 system or the *Streptomyces*-derived class I bi-functional CYP450 system or rational designed mutants thereof seems promising. As most of the produced compounds are intrinsically characterized by various bioactivities[146] and are shown to be produced efficiently and in good amounts, detailed investigations should be undertaken to decipher their spectrum of activity.

The complex modelling of the CotB3/4•Afx and CotB3/4•Pdx/Pdx-mutants with the P450cam•Pdx complex as template and the subsequent adaption of Pdx to superiorly activate CotB3 opens up an interesting possibility to adapt non-native class I redox systems for the application in any class I CYP450 system-based biotechnological process, a missing link in today`s biocatalytic methodologies. The background of this *in silico* strategy is an easy-to-use and fast approach to define important residues participating in electron transfer and redox partner binding. The newly identified, efficiently working class I redox system from *Streptomyces afghaniensis* could moreover serve as a promising class I redox system for the activation of any *Streptomyces*-derived class I CYP450 monooxygenases. The phylum *Streptomyces* possess an overwhelming variety of bioactives-producing strains, rendering the application of the identified redox system an interesting future perspective.[29,92,150,151]

Another future perspective based on the findings in this study could be the introduction of more than one terpene synthase into one bacterial host using the modular duet-vector-based metabolic production system. However, recent advances opening up efficient strategies for the genomic integration of large metabolic pathways into any bacterial host renders these approaches more promising and effective for the integration of more than one terpene synthase.[13,18,23] An ultimate goal would be to combine the efficient production of terpene macrocycles and their efficient functional decoration in a metabolically optimized and large-scale adapted microbial host that is able to use sustainable second or third generation feedstocks as sole carbon source.

References

1 Clark, D., Farmer. The Integration of green chemistry into future biorefineries. *Biofuels, Bioprod. Bioref.* **3**, 72-90 (2009).

2 Ragauskas, A. J. *et al.* The path forward for biofuels and biomaterials. *Science* **311**, 484-489, doi:10.1126/science.1114736 (2006).

3 Azapagic, A. Sustainability considerations for integrated biorefineries. *Trends in biotechnology* **32**, 1-4, doi:10.1016/j.tibtech.2013.10.009 (2014).

4 Demirbas. Biorefineries for biofuel upgrading: A critical review. *Applied Energy* **86**, 151-161 (2009).

5 da Silva, T. L., Gouveia, L. & Reis, A. Integrated microbial processes for biofuels and high value-added products: the way to improve the cost effectiveness of biofuel production. *Applied microbiology and biotechnology* **98**, 1043-1053, doi:10.1007/s00253-013-5389-5 (2014).

6 FitzPatrick, M., Champagne, P., Cunningham, M. F. & Whitney, R. A. A biorefinery processing perspective: treatment of lignocellulosic materials for the production of value-added products. *Bioresource technology* **101**, 8915-8922, doi:10.1016/j.biortech.2010.06.125 (2010).

7 Kokossis, Y. On the use of system technologies and a systematic approach for the synthesis and the design of future biorefineries. *Computers and Chemical Engineering* **34**, 1397-1405 (2010).

8 Du, J., Shao, Z. & Zhao, H. Engineering microbial factories for synthesis of value-added products. *Journal of industrial microbiology & biotechnology* **38**, 873-890, doi:10.1007/s10295-011-0970-3 (2011).

9 Wright, B. Comparative economics of biorefineries based on the biochemical and thermochemical platforms. *Biofuels, Bioprod. Bioref.* **1** (2007).

10 Baeshen, N. A. *et al.* Cell factories for insulin production. *Microbial cell factories* **13**, 141, doi:10.1186/s12934-014-0141-0 (2014).

11 Chames, P., Van Regenmortel, M., Weiss, E. & Baty, D. Therapeutic antibodies: successes, limitations and hopes for the future. *British journal of pharmacology* **157**, 220-233, doi:10.1111/j.1476-5381.2009.00190.x (2009).

12 Spadiut, O., Capone, S., Krainer, F., Glieder, A. & Herwig, C. Microbials for the production of monoclonal antibodies and antibody fragments. *Trends in biotechnology* **32**, 54-60, doi:10.1016/j.tibtech.2013.10.002 (2014).

13 Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nature biotechnology* **21**, 796-802, doi:10.1038/nbt833 (2003).

14 Chen, Y. & Nielsen, J. Advances in metabolic pathway and strain engineering paving the way for sustainable production of chemical building blocks. *Current opinion in biotechnology* **24**, 965-972, doi:10.1016/j.copbio.2013.03.008 (2013).

15 Yoshikuni, Y., Ferrin, T. E. & Keasling, J. D. Designed divergent evolution of enzyme function. *Nature* **440**, 1078-1082, doi:10.1038/nature04607 (2006).

16 Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485**, 185-194, doi:10.1038/nature11117 (2012).

17 Brück, K., Loll. Production of Macrocyclic Sesqui- and Diterpenes in Heterologous Microbial Hosts: A Systems Approach to Harness Nature`s Molecular Diversity. *ChemCatChem* **6**, 1142-1165 (2014).

18 Ajikumar, P. K. *et al.* Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science* **330**, 70-74, doi:10.1126/science.1191652 (2010).

19 Klein-Marcuschamer, D., Ajikumar, P. K. & Stephanopoulos, G. Engineering microbial cell factories for biosynthesis of isoprenoid molecules: beyond lycopene. *Trends in biotechnology* **25**, 417-424, doi:10.1016/j.tibtech.2007.07.006 (2007).

20 Wang, Y. & Pfeifer, B. A. 6-deoxyerythronolide B production through chromosomal localization of the deoxyerythronolide B synthase genes in E. coli. *Metabolic engineering* **10**, 33-38, doi:10.1016/j.ymben.2007.09.002 (2008).

21 Yuan, L. Z., Rouviere, P. E., Larossa, R. A. & Suh, W. Chromosomal promoter replacement of the isoprenoid pathway for enhancing carotenoid production in E. coli. *Metabolic engineering* **8**, 79-90, doi:10.1016/j.ymben.2005.08.005 (2006).

22 Yadav, V. G., De Mey, M., Lim, C. G., Ajikumar, P. K. & Stephanopoulos, G. The future of metabolic engineering and synthetic biology: towards a systematic practice. *Metabolic engineering* **14**, 233-241 (2012).

23 Keasling, J. D. Manufacturing molecules through metabolic engineering. *Science* **330**, 1355-1358, doi:10.1126/science.1193990 (2010).

24 Bornscheuer, U. T. & Pohl, M. Improved biocatalysts by directed evolution and rational protein design. *Current opinion in chemical biology* **5**, 137-143 (2001).

25 Strohmeier, G. A., Pichler, H., May, O. & Gruber-Khadjawi, M. Application of designed enzymes in organic synthesis. *Chemical reviews* **111**, 4141-4164, doi:10.1021/cr100386u (2011).

26 Ueberbacher, B. T., Hall, M. & Faber, K. Electrophilic and nucleophilic enzymatic cascade reactions in biosynthesis. *Natural product reports* **29**, 337-350, doi:10.1039/c2np00078d (2012).

27 Oldfield, E. Targeting isoprenoid biosynthesis for drug discovery: bench to bedside. *Accounts of chemical research* **43**, 1216-1226, doi:10.1021/ar100026v (2010).

28 Ajikumar, P. K. *et al.* Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Molecular pharmaceutics* **5**, 167-190, doi:10.1021/mp700151b (2008).

29 Smanski, M. J., Peterson, R. M., Huang, S. X. & Shen, B. Bacterial diterpene synthases: new opportunities for mechanistic enzymology and engineered biosynthesis. *Current opinion in chemical biology* **16**, 132-141, doi:10.1016/j.cbpa.2012.03.002 (2012).

30 Horwitz, S. B. Taxol (paclitaxel): mechanisms of action. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* **5 Suppl 6**, S3-6 (1994).

31 Meshnick, S. R. Artemisinin: mechanisms of action, resistance and toxicity. *International journal for parasitology* **32**, 1655-1660 (2002).

32 Riccioni, G. Marine carotenoids and oxidative stress. *Marine drugs* **10**, 116-118, doi:10.3390/md10010116 (2012).

33 Aoyama, T., Naganawa, H., Muraoka, Y., Aoyagi, T. & Takeuchi, T. The structure of cyclooctatin, a new inhibitor of lysophospholipase. *The Journal of antibiotics* **45**, 1703-1704 (1992).

34 Aqil, F. *et al.* Antimicrobial, antioxidant, and antimutagenic activities of selected marine natural products and tobacco cembranoids. *Drug and chemical toxicology* **34**, 167-179, doi:10.3109/01480545.2010.494669 (2011).

35 Morrone, D., Xu, M., Fulton, D. B., Determan, M. K. & Peters, R. J. Increasing complexity of a diterpene synthase reaction with a single residue switch. *Journal of the American Chemical Society* **130**, 5400-5401, doi:10.1021/ja710524w (2008).

36 Boghigian, B. A., Myint, M., Wu, J. & Pfeifer, B. A. Simultaneous production and partitioning of heterologous polyketide and isoprenoid natural products in an Escherichia coli two-phase bioprocess. *Journal of industrial microbiology & biotechnology* **38**, 1809-1820, doi:10.1007/s10295-011-0969-9 (2011).

37 Boghigian, B. A., Salas, D., Ajikumar, P. K., Stephanopoulos, G. & Pfeifer, B. A. Analysis of heterologous taxadiene production in K- and B-derived Escherichia coli. *Applied microbiology and biotechnology* **93**, 1651-1661, doi:10.1007/s00253-011-3528-4 (2012).

38 Oldfield, E. & Lin, F. Y. Terpene biosynthesis: modularity rules. *Angewandte Chemie* **51**, 1124-1137, doi:10.1002/anie.201103110 (2012).

39 Kuzuyama, T. Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. *Bioscience, biotechnology, and biochemistry* **66**, 1619-1627 (2002).

40 Boghigian, B. A., Armando, J., Salas, D. & Pfeifer, B. A. Computational identification of gene over-expression targets for metabolic engineering of taxadiene production. *Applied microbiology and biotechnology* **93**, 2063-2073, doi:10.1007/s00253-011-3725-1 (2012).

41 Leonard, E. *et al.* Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proc Natl Acad Sci U S A* **107**, 13654-13659, doi:10.1073/pnas.1006138107 (2010).

42 Gao, Y., Honzatko, R. B. & Peters, R. J. Terpenoid synthase structures: a so far incomplete view of complex catalysis. *Natural product reports* **29**, 1153-1175, doi:10.1039/c2np20059g (2012).

43 Cao, R. *et al.* Diterpene cyclases and the nature of the isoprene fold. *Proteins* **78**, 2417-2432, doi:10.1002/prot.22751 (2010).

44 Christianson, D. W. Unearthing the roots of the terpenome. *Current opinion in chemical biology* **12**, 141-150, doi:10.1016/j.cbpa.2007.12.008 (2008).

45 Janke, R., Gorner, C., Hirte, M., Bruck, T. & Loll, B. The first structure of a bacterial diterpene cyclase: CotB2. *Acta crystallographica. Section D, Biological crystallography* **70**, 1528-1537, doi:10.1107/S1399004714005513 (2014).

46 Williams, D. C. *et al.* Heterologous expression and characterization of a "Pseudomature" form of taxadiene synthase involved in paclitaxel (Taxol) biosynthesis and evaluation of a potential intermediate and inhibitors of the multistep diterpene cyclization reaction. *Archives of biochemistry and biophysics* **379**, 137-146, doi:10.1006/abbi.2000.1865 (2000).

47 Koksal, M., Jin, Y., Coates, R. M., Croteau, R. & Christianson, D. W. Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. *Nature* **469**, 116-120, doi:10.1038/nature09628 (2011).

48 Whittington, D. A. *et al.* Bornyl diphosphate synthase: structure and strategy for carbocation manipulation by a terpenoid cyclase. *Proc Natl Acad Sci U S A* **99**, 15375-15380, doi:10.1073/pnas.232591099 (2002).

49 Shishova, E. Y. *et al.* X-ray crystallographic studies of substrate binding to aristolochene synthase suggest a metal ion binding sequence for catalysis. *The Journal of biological chemistry* **283**, 15431-15439, doi:10.1074/jbc.M800659200 (2008).

50 Starks, C. M., Back, K., Chappell, J. & Noel, J. P. Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science* **277**, 1815-1820 (1997).

51 Caruthers, J. M., Kang, I., Rynkiewicz, M. J., Cane, D. E. & Christianson, D. W. Crystal structure determination of aristolochene synthase from the blue cheese mold, Penicillium roqueforti. *The Journal of biological chemistry* **275**, 25533-25539, doi:10.1074/jbc.M000433200 (2000).

52 Rynkiewicz, M. J., Cane, D. E. & Christianson, D. W. Structure of trichodiene synthase from Fusarium sporotrichioides provides mechanistic inferences on the terpene cyclization cascade. *Proc Natl Acad Sci U S A* **98**, 13543-13548, doi:10.1073/pnas.231313098 (2001).

53 Baer, P. *et al.* Induced-fit mechanism in class I terpene cyclases. *Angewandte Chemie* **53**, 7652-7656, doi:10.1002/anie.201403648 (2014).

54 Zhou, K. *et al.* Insights into diterpene cyclization from structure of bifunctional abietadiene synthase from Abies grandis. *The Journal of biological chemistry* **287**, 6840-6850, doi:10.1074/jbc.M111.337592 (2012).

55 Zhou, K. & Peters, R. J. Electrostatic effects on (di)terpene synthase product outcome. *Chemical communications* **47**, 4074-4080, doi:10.1039/c0cc02960b (2011).

56 Major, D. T. & Weitman, M. Electrostatically guided dynamics--the root of fidelity in a promiscuous terpene synthase? *Journal of the American Chemical Society* **134**, 19454-19462, doi:10.1021/ja308295p (2012).

57 Hong, Y. J. & Tantillo, D. J. Quantum chemical dissection of the classic terpinyl/pinyl/bornyl/camphyl cation conundrum-the role of pyrophosphate in manipulating pathways to monoterpenes. *Organic & biomolecular chemistry* **8**, 4589-4600, doi:10.1039/c0ob00167h (2010).

58 Hong, Y. J. & Tantillo, D. J. The taxadiene-forming carbocation cascade. *Journal of the American Chemical Society* **133**, 18249-18256, doi:10.1021/ja2055929 (2011).

59 Srividya, N., Davis, E. M., Croteau, R. B. & Lange, B. M. Functional analysis of (4S)-limonene synthase mutants reveals determinants of catalytic outcome in a model monoterpene synthase. *Proc Natl Acad Sci U S A* **112**, 3332-3337, doi:10.1073/pnas.1501203112 (2015).

60 Hannemann, F., Bichet, A., Ewen, K. M. & Bernhardt, R. Cytochrome P450 systems--biological variations of electron transport chains. *Biochimica et biophysica acta* **1770**, 330-344, doi:10.1016/j.bbagen.2006.07.017 (2007).

61 Urlacher, V. B. & Girhard, M. Cytochrome P450 monooxygenases: an update on perspectives for synthetic application. *Trends in biotechnology* **30**, 26-36, doi:10.1016/j.tibtech.2011.06.012 (2012).

62 Coon, M. J. Cytochrome P450: nature's most versatile biological catalyst. *Annual review of pharmacology and toxicology* **45**, 1-25, doi:10.1146/annurev.pharmtox.45.120403.100030 (2005).

63 Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P. & McPhail, A. T. Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from Taxus brevifolia. *Journal of the American Chemical Society* **93**, 2325-2327 (1971).

64 Koepp, A. E. *et al.* Cyclization of geranylgeranyl diphosphate to taxa-4(5),11(12)-diene is the committed step of taxol biosynthesis in Pacific yew. *The Journal of biological chemistry* **270**, 8686-8690 (1995).

65 Lin, X., Hezari, M., Koepp, A. E., Floss, H. G. & Croteau, R. Mechanism of taxadiene synthase, a diterpene cyclase that catalyzes the first step of taxol biosynthesis in Pacific yew. *Biochemistry* **35**, 2968-2977, doi:10.1021/bi9526239 (1996).

66 Williams, D. C. *et al.* Intramolecular proton transfer in the cyclization of geranylgeranyl diphosphate to the taxadiene precursor of taxol catalyzed by recombinant taxadiene synthase. *Chemistry & biology* **7**, 969-977 (2000).

67 Jin, Q., Williams, D. C., Hezari, M., Croteau, R. & Coates, R. M. Stereochemistry of the macrocyclization and elimination steps in taxadiene biosynthesis through deuterium labeling. *The Journal of organic chemistry* **70**, 4667-4675, doi:10.1021/jo0502091 (2005).

68 Jin, Y., Williams, D. C., Croteau, R. & Coates, R. M. Taxadiene synthase-catalyzed cyclization of 6-fluorogeranylgeranyl diphosphate to 7-fluoroverticillenes. *Journal of the American Chemical Society* **127**, 7834-7842, doi:10.1021/ja050592r (2005).

69 Chow, S. Y., Williams, H. J., Huang, Q., Nanda, S. & Scott, A. I. Studies on taxadiene synthase: interception of the cyclization cascade at the isocembrene stage with GGPP analogues. *The Journal of organic chemistry* **70**, 9997-10003, doi:10.1021/jo0517489 (2005).

70 Siew Yin Chow, H. J. W., James D. Pennington, Samik Nanda‡, Joseph H. Reibenspies, A. Ian Scott. Studies on taxadiene synthase: interception of the cyclization cascade at the verticillene stage and rearrangement to phomactatriene. *Tetrahedron* **63**, 6204-6209 (2007).

71 Gutta, P. & Tantillo, D. J. A promiscuous proton in taxadiene biosynthesis? *Organic letters* **9**, 1069-1071, doi:10.1021/ol070007m (2007).

72 Hyatt, D. C. *et al.* Structure of limonene synthase, a simple model for terpenoid cyclase catalysis. *Proc Natl Acad Sci U S A* **104**, 5360-5365, doi:10.1073/pnas.0700915104 (2007).

73 Hong, Y. J. & Tantillo, D. J. A potential energy surface bifurcation in terpene biosynthesis. *Nature chemistry* **1**, 384-389, doi:10.1038/nchem.287 (2009).

74 Hong Cui, S.-T. Z., Hui-Juan Yang, Hao Ji and Xiu-Jie Wang. Gene expression profile analysis of tobacco leaf trichomes. *BMC Plant Biology* **11:76** (2011).

75 Büttner, A. Mechanistic Studies of Taxadiene Synthases. *Master Thesis TUM* (2012).

76 Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M. & Wolthers, B. G. Structure of papain. *Nature* **218**, 929-932 (1968).

77 Meguro, A., Tomita, T., Nishiyama, M. & Kuzuyama, T. Identification and characterization of bacterial diterpene cyclases that synthesize the cembrane skeleton. *Chembiochem : a European journal of chemical biology* **14**, 316-321, doi:10.1002/cbic.201200651 (2013).

78 Xu, M., Wilderman, P. R. & Peters, R. J. Following evolution's lead to a single residue switch for diterpene synthase product outcome. *Proc Natl Acad Sci U S A* **104**, 7397-7401, doi:10.1073/pnas.0611454104 (2007).

79 Wilderman, P. R. & Peters, R. J. A single residue switch converts abietadiene synthase into a pimaradiene specific cyclase. *Journal of the American Chemical Society* **129**, 15736-15737, doi:10.1021/ja074977g (2007).

80 Kirby, J. *et al.* Cloning of casbene and neocembrene synthases from Euphorbiaceae plants and expression in Saccharomyces cerevisiae. *Phytochemistry* **71**, 1466-1473, doi:10.1016/j.phytochem.2010.06.001 (2010).

81 Sugano, M. *et al.* Phomactin E, F, and G: new phomactin-group PAF antagonists from a marine fungus Phoma sp. *The Journal of antibiotics* **48**, 1188-1190 (1995).

82 Ya-Ching Shen, Y.-B. C., Jun'ichi Kobayashi, Takaaki Kubota, Yohei Takahashi, Yuzuru Mikami , Junji Ito and Yun-Sheng Lin. Nitrogen-Containing Verticillene Diterpenoids from the Taiwanese Soft Coral Cespitularia taeniata. *J. Nat. Prod.* **70** 1961-1965 (2007).

83 Shen, Y. C. *et al.* Cespihypotins Q-V, verticillene diterpenoids from Cespitularia hypotentaculata. *Journal of natural products* **71**, 1993-1997, doi:10.1021/np8005327 (2008).

84 Meguro, A. *et al.* An Unusual Terpene Cyclization Mechanism Involving a Carbon-Carbon Bond Rearrangement. *Angewandte Chemie*, doi:10.1002/anie.201411923 (2015).

85 M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision. (2004).

86 Krieger, E. *et al.* Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. *Proteins* **77 Suppl 9**, 114-122, doi:10.1002/prot.22570 (2009).

87 Jakalian, A., Jack, D. B. & Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of computational chemistry* **23**, 1623-1641, doi:10.1002/jcc.10128 (2002).

88 Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *Journal of computational chemistry* **25**, 1157-1174, doi:10.1002/jcc.20035 (2004).

89 Krieger, E., Nielsen, J. E., Spronk, C. A. & Vriend, G. Fast empirical pKa prediction by Ewald summation. *Journal of molecular graphics & modelling* **25**, 481-486, doi:10.1016/j.jmgm.2006.02.009 (2006).

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

91 Jung, J. & Lee, B. Protein structure alignment using environmental profiles. *Protein engineering* **13**, 535-543 (2000).

92 de Boer, A. H. & de Vries-van Leeuwen, I. J. Fusicoccanes: diterpenes with surprising biological functions. *Trends in plant science* **17**, 360-368, doi:10.1016/j.tplants.2012.02.007 (2012).

93 Görner, C. H., I. Schrepfer, P. Eisenreich, W. Brück, T. Targeted Engineering of Cyclooctat-9-en-7-ol Synthase: AStereospecific Access to Two New Non-naturalFusicoccane-Type Diterpenes. *ChemCatChem* **5**, 3289-3298, doi:10.1002/c ctc.201300285 (2013).

94 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* **16**, 736-743, doi:10.1016/j.chembiol.2009.06.007 (2009).

95 Aoyagi, T. *et al.* Cyclooctatin, a new inhibitor of lysophospholipase, produced by Streptomyces melanosporofaciens MI614-43F2. Taxonomy, production, isolation, physico-chemical properties and biological activities. *The Journal of antibiotics* **45**, 1587-1591 (1992).

96 Yin, J. *et al.* New casbane diterpenoids from a South China Sea soft coral, Sinularia sp. *Marine drugs* **11**, 455-465, doi:10.3390/md11020455 (2013).

97 Shao, F. G. *et al.* Two new casbane diterpenoids from the roots of Euphorbia pekinensis. *Journal of Asian natural products research* **13**, 805-810, doi:10.1080/10286020.2011.596828 (2011).

98 Yang, B. *et al.* New cembrane diterpenoids from a Hainan soft coral Sinularia sp. *Marine drugs* **10**, 2023-2032, doi:10.3390/md10092023 (2012).

99 Vasas, A. & Hohmann, J. Euphorbia diterpenes: isolation, structure, biological activity, and synthesis (2008-2012). *Chemical reviews* **114**, 8579-8612, doi:10.1021/cr400541j (2014).

100 Bernhardt, R. Cytochromes P450 as versatile biocatalysts. *Journal of biotechnology* **124**, 128-145, doi:10.1016/j.jbiotec.2006.01.026 (2006).

101 Pandey, B. P. *et al.* Identification of the specific electron transfer proteins, ferredoxin, and ferredoxin reductase, for CYP105D7 in Streptomyces avermitilis MA4680. *Applied microbiology and biotechnology* **98**, 5009-5017, doi:10.1007/s00253-014-5525-x (2014).

102 Lee-Huu, H., Claasen, Laschat, Urlacher. Chemo - , Regio - , and Stereoselective Oxidation of the Monocyclic Diterpenoid β- Cembrenediol by P450 BM3. *ACS Catalysis* **5**, 1772-1780 (2015).

103 Bleif, S. *et al.* Identification of CYP106A2 as a regioselective allylic bacterial diterpene hydroxylase. *Chembiochem : a European journal of chemical biology* **12**, 576-582, doi:10.1002/cbic.201000404 (2011).

104 Sevrioukova, I. F. & Poulos, T. L. Structural biology of redox partner interactions in P450cam monooxygenase: a fresh look at an old system. *Archives of biochemistry and biophysics* **507**, 66-74, doi:10.1016/j.abb.2010.08.022 (2011).

105 Komatsu, N., Nakazawa, S., Hamada, M., Shimo, M. & Tomosugi, T. Studies on the taitomycin, a new antibiotic produced by Streptomyces afghaniensis. IV. Effect of taitomycin on experimental psittacosis. *The Journal of antibiotics* **12**, 173-176 (1959).

106 Link, T. A. The role of the 'Rieske' iron sulfur protein in the hydroquinone oxidation (Q(P)) site of the cytochrome bc1 complex. The 'proton-gated affinity change' mechanism. *FEBS letters* **412**, 257-264 (1997).

107 van den Berg, W. A., Hagen, W. R. & van Dongen, W. M. The hybrid-cluster protein ('prismane protein') from Escherichia coli. Characterization of the hybrid-cluster protein, redox properties of the [2Fe-2S] and [4Fe-2S-2O] clusters and identification of an associated NADH oxidoreductase containing FAD and [2Fe-2S]. *European journal of biochemistry / FEBS* **267**, 666-676 (2000).

108 Grinberg, A. V. *et al.* Adrenodoxin: structure, stability, and electron transfer properties. *Proteins* **40**, 590-612 (2000).

109 Mo, H., Pochapsky, S. S. & Pochapsky, T. C. A model for the solution structure of oxidized terpredoxin, a Fe2S2 ferredoxin from Pseudomonas. *Biochemistry* **38**, 5666-5675, doi:10.1021/bi983063r (1999).

110 Tripathi, S., Li, H. & Poulos, T. L. Structural basis for effector control and redox partner recognition in cytochrome P450. *Science* **340**, 1227-1230, doi:10.1126/science.1235797 (2013).

111 Sevrioukova, I. F., Poulos, T. L. & Churbanova, I. Y. Crystal structure of the putidaredoxin reductase x putidaredoxin electron transfer complex. *The Journal of biological chemistry* **285**, 13616-13620, doi:10.1074/jbc.M110.104968 (2010).

112 Nakano, Y., Ohtani,M., Polsri,W., Usami,T., Sambongi,K. and & Demura, T. Characterization of the casbene synthase homolog from Jatropha. *Plant Biotechnol.* **29**, 185-189 (2012).

113 Hiruma, Y. *et al.* The structure of the cytochrome p450cam-putidaredoxin complex determined by paramagnetic NMR spectroscopy and crystallography. *Journal of molecular biology* **425**, 4353-4365, doi:10.1016/j.jmb.2013.07.006 (2013).

114 Pandey, B. P. *et al.* Regioselective hydroxylation of daidzein using P450 (CYP105D7) from Streptomyces avermitilis MA4680. *Biotechnology and bioengineering* **105**, 697-704, doi:10.1002/bit.22582 (2010).

115 Pandey, B. P. *et al.* Screening of bacterial cytochrome P450s responsible for regiospecific hydroxylation of (iso)flavonoids. *Enzyme and microbial technology* **48**, 386-392, doi:10.1016/j.enzmictec.2011.01.001 (2011).

116 Yang, W. *et al.* Molecular characterization of a class I P450 electron transfer system from Novosphingobium aromaticivorans DSM12444. *The Journal of biological chemistry* **285**, 27372-27384, doi:10.1074/jbc.M110.118349 (2010).

117 Makino, T., Katsuyama, Y., Otomatsu, T., Misawa, N. & Ohnishi, Y. Regio- and stereospecific hydroxylation of various steroids at the 16alpha position of the D ring by the Streptomyces

griseus cytochrome P450 CYP154C3. *Applied and environmental microbiology* **80**, 1371-1379, doi:10.1128/AEM.03504-13 (2014).

118 Hussain, H. A. & Ward, J. M. Enhanced heterologous expression of two Streptomyces griseolus cytochrome P450s and Streptomyces coelicolor ferredoxin reductase as potentially efficient hydroxylation catalysts. *Applied and environmental microbiology* **69**, 373-382 (2003).

119 Madrona, Y. *et al.* Crystal structure of cindoxin, the P450cin redox partner. *Biochemistry* **53**, 1435-1446, doi:10.1021/bi500010m (2014).

120 Bell, S. G. *et al.* Cytochrome P450 enzymes from the metabolically diverse bacterium Rhodopseudomonas palustris. *Biochemical and biophysical research communications* **342**, 191-196, doi:10.1016/j.bbrc.2006.01.133 (2006).

121 Hiruma, Y. *et al.* Hot-spot residues in the cytochrome P450cam-putidaredoxin binding interface. *Chembiochem : a European journal of chemical biology* **15**, 80-86, doi:10.1002/cbic.201300582 (2014).

122 Kuznetsov, V. Y., Poulos, T. L. & Sevrioukova, I. F. Putidaredoxin-to-cytochrome P450cam electron transfer: differences between the two reductive steps required for catalysis. *Biochemistry* **45**, 11934-11944, doi:10.1021/bi0611154 (2006).

123 Yasutake, Y., Nishioka, T., Imoto, N. & Tamura, T. A single mutation at the ferredoxin binding site of P450 Vdh enables efficient biocatalytic production of 25-hydroxyvitamin D(3). *Chembiochem : a European journal of chemical biology* **14**, 2284-2291, doi:10.1002/cbic.201300386 (2013).

124 Pei, J. & Grishin, N. V. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods in molecular biology* **1079**, 263-271, doi:10.1007/978-1-62703-646-7_17 (2014).

125 Liu, X. *et al.* Bioactive cembranoids from the South China Sea soft coral Sarcophyton elegans. *Molecules* **20**, 13324-13335, doi:10.3390/molecules200713324 (2015).

126 Kobayashi, M., Kobayashi, K., Nomura, M. & Munakata, H. Conformational study of the cembranoid sarcophytol A, a potent anti-tumor-promoter. *Chemical & pharmaceutical bulletin* **38**, 815-817 (1990).

127 Xi, Z. *et al.* Sarcophyolides B-E, new cembranoids from the soft coral Sarcophyton elegans. *Marine drugs* **11**, 3186-3196, doi:10.3390/md11093186 (2013).

128 Wang, E. *et al.* Suppression of a P450 hydroxylase gene in plant trichome glands enhances natural-product-based aphid resistance. *Nature biotechnology* **19**, 371-374, doi:10.1038/86770 (2001).

129 El Sayed, K. A. & Sylvester, P. W. Biocatalytic and semisynthetic studies of the anticancer tobacco cembranoids. *Expert opinion on investigational drugs* **16**, 877-887, doi:10.1517/13543784.16.6.877 (2007).

130 Ferchmin, P. A. *et al.* Actions of octocoral and tobacco cembranoids on nicotinic receptors. *Toxicon : official journal of the International Society on Toxinology* **54**, 1174-1182, doi:10.1016/j.toxicon.2009.02.033 (2009).

131 Ferchmin, P. A. *et al.* 4R-cembranoid protects against diisopropylfluorophosphate-mediated neurodegeneration. *Neurotoxicology* **44**, 80-90, doi:10.1016/j.neuro.2014.06.001 (2014).

132 Martins, A. H. *et al.* Neuroprotective activity of (1S,2E,4R,6R,-7E,11E)-2,7,11-cembratriene-4,6-diol (4R) in vitro and in vivo in rodent models of brain ischemia. *Neuroscience* **291**, 250-259, doi:10.1016/j.neuroscience.2015.02.001 (2015).

133 Ferchmin, P. A. *et al.* Tobacco cembranoids block behavioral sensitization to nicotine and inhibit neuronal acetylcholine receptor function. *Journal of neuroscience research* **64**, 18-25 (2001).

134 Guo, W. Biosynthesis of cembratrienols in cell-free extracts from trichomes of Nicotiana tabacum. *Plant Science* **110**, 1-10, doi:10.1016/0168-9452(95)04174-S (1995).

135 Cui, H., Zhang, S. T., Yang, H. J., Ji, H. & Wang, X. J. Gene expression profile analysis of tobacco leaf trichomes. *BMC Plant Biol* **11**, 76, doi:10.1186/1471-2229-11-76 (2011).

136 Bruckner, K. & Tissier, A. High-level diterpene production by transient expression in Nicotiana benthamiana. *Plant methods* **9**, 46, doi:10.1186/1746-4811-9-46 (2013).

137 Bai, S. & Jain, M. 1H and 13C assignments of five cembrenes from guggul. *Magnetic resonance in chemistry : MRC* **46**, 791-793, doi:10.1002/mrc.2252 (2008).

138 Crombie, L., Kneen, G., Pattenden, G. & Whybrow, D. Total synthesis of the macrocyclic diterpene (-)-casbene, the putative biogenetic precursor of lathyrane, tigliane, ingenane, and related terpenoid structures. *Journal of the Chemical Society, Perkin Transactions 1*, 1711-1717, doi:10.1039/P19800001711 (1980).

139 Wahlberg, W., Narbonne, Nishida, Enzell. Note on the Stereostructures of Thunbergol (Isocembrol) and 4-Epiisocemb. *Acta Chemica Scandinavica* **35**, 65-68, doi:10.3891/acta.chem.scand.35b-0065 (197).

140 Pattenden, G. & Smithies, A. J. Concise syntheses of cembrenes based on radical-mediated vinylcyclopropane ring-opening reactions in casbene. *Journal of the Chemical Society, Perkin Transactions 1*, 57-61, doi:10.1039/P19960000057 (1996).

141 Swaminathan, S., Morrone, D., Wang, Q., Fulton, D. B. & Peters, R. J. CYP76M7 is an ent-cassadiene C11alpha-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice. *The Plant cell* **21**, 3315-3325, doi:10.1105/tpc.108.063677 (2009).

142 Wang, Q., Hillwig, M. L., Wu, Y. & Peters, R. J. CYP701A8: a rice ent-kaurene oxidase paralog diverted to more specialized diterpenoid metabolism. *Plant physiology* **158**, 1418-1425, doi:10.1104/pp.111.187518 (2012).

143 Leonard, E. & Koffas, M. A. Engineering of artificial plant cytochrome P450 enzymes for synthesis of isoflavones by Escherichia coli. *Applied and environmental microbiology* **73**, 7246-7251, doi:10.1128/AEM.01411-07 (2007).

144 Chang, M. C., Eachus, R. A., Trieu, W., Ro, D. K. & Keasling, J. D. Engineering Escherichia coli for production of functionalized terpenoids using plant P450s. *Nature chemical biology* **3**, 274-277, doi:10.1038/nchembio875 (2007).

145 Yamada, T., Imaishi, H., Oka, A. & Ohkawa, H. Molecular cloning and expression in Saccharomyces cerevisiae of tobacco NADPH-cytochrome P450 oxidoreductase cDNA. *Bioscience, biotechnology, and biochemistry* **62**, 1403-1411, doi:10.1271/bbb.62.1403 (1998).

146 Villanueva, S. Cembrene Diterpenoids: Conformational Studies and Molecular Docking to Tubulin *Records of Natural Products* **4**, 115 (2010).

147 Balogh, D. Engineering a recombinant biosynthetic pathway for the production of 2,7,11-cembratriene-4,6-diol in *Escherichia coli*. *Master Thesis TUM* (2014).

148 Ro, D. K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940-943, doi:10.1038/nature04640 (2006).

149 Renault, H., Bassard, J. E., Hamberger, B. & Werck-Reichhart, D. Cytochrome P450-mediated metabolic engineering: current progress and future challenges. *Current opinion in plant biology* **19**, 27-34, doi:10.1016/j.pbi.2014.03.004 (2014).

150 Cane, D. E. & Ikeda, H. Exploration and mining of the bacterial terpenome. *Accounts of chemical research* **45**, 463-472, doi:10.1021/ar200198d (2012).

151 Citron, C. A., Barra, L., Wink, J. & Dickschat, J. S. Volatiles from nineteen recently genome sequenced actinomycetes. *Organic & biomolecular chemistry* **13**, 2673-2683, doi:10.1039/c4ob02609h (2015).

152 Metterlein, M. Mechanistic Studies of the Terpenoid Cyclase from Common Tobacco. *Master Thesis TUM* (2013).

Chapter VI: Appendix

1. Abbreviations

Amino acids

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Trypthophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Used enzymes and pathways

TXS	taxa-4,5-11,12-diene synthase, class I diterpene synthase
DXS	1-deoxy-D-xylulose 5-phosphate synthase
DXR	1-deoxy-D-xylulose 5-phosphate reductoisomerase
IspD	2-C-methyl-D-erythriol 4-phosphate cytidyltransferase synthase
IspE	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase
IspF	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
IspG	(E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate synthase
IspH	(E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate reductase
IDI	isopentenyl-diphosphate delta isomerase
ispA	geranylgeranylpyrophosphate synthase
GGPPS	geranylgeranylpyrophosphate synthase
Crte	geranylgeranylpyrophosphate synthase from *P. agglomerans*
CotB1	geranylgeranylpyrophosphate synthase
CotB2	class I diterpene synthase
CotB3	class I cyctochrome c P450 oxidoreductase

CotB4	class I cyctochrome c P450 oxidoreductase
CYP450	cyctochrome c P450 oxidoreductase
CPR	class II NADPH-cyctochrome c-P450 reductase
FdR	class I ferredoxin reductase
Fdx	class I ferredoxin
AfR	class I ferredoxin reductase from *S. afghaniensis*
Afx	class I ferredoxin from S. afghaniensis
PdR	class I ferredoxin reductase from *P. putida*
Pdx	class I ferredoxin from P. putida
CS	(-)-casbene synthase, class I diterpene synthase
CBTS	2,7,11-cembratriene-4-ol synthase, class I diterpene synthase
DXP	2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate pathway
MEP	2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate pathway
MEV	mevalonate pathway

Used abbreviations in the text

%	percent	ALA	5-aminolevulinic DNA
% v/v	percent volume per volume	BLAST	Basic Local Alignment Search Tool
% v/w	percent volume per weight	CD	circular dichroism
&	and	COSY	correlation spectroscopy
°C	degree	DTT	dithiothreitol
µ	micro ($1 * 10^{-6}$)	*E. coli*	*Escherichia coli*
µl	microliter	EI	electron ionization
1D	one-dimensional	ESI	electrospray ionization
2D	two-dimensional	ET	electron transfer
Å	Angstrom	FAD	flavin adenine dinucleotide
C	carbon	FID	flame ionization detector
CA	casamino acids	FMN	flavin mononucleotide
CBT	2,7,11-cembratriene-4-	FPLC	fast-protein-liquid-chromatography
$CDCl_3$	deuterated chloroform	Fw	forward
$CDOD_3$	deuterated ethanole	GC	gaschromatograph
cm	centimeter	HCP	hybride-cluster-like
dd	double distilled	HMBC	heteronuclear multiple bond correlation
DMAPP	dimethylallyl diphosphate	HPLC	high-performance-liquid-chromatography
Fe	iron	HR	high resolution
FPP	farnesylpyrophosphate	HSQC	heteronuclear single quantum coherence
g	gram	IPTG	isopropyl-ß-1-thiogalactopyranoside
GGPP	geranylgeranylpyrophosphate	LB	Luria-Bertani-broth
GPP	geranylpyrophosphate	LC	liquid-chromatography
H	hydrogen	M9	M9 minimal media
IPP	isopentenyl diphosphate	MCS	multiple cloning site
K	Kelvin	MD	molecular dynamic
kcal	kilocalorie	MM	molecular mechanic
L	Liter	MOPS	3-(N-morpholino)propanesulfonic acid

m	milli ($1 * 10^{-3}$)	MS	mass spectrometry
M	molar	N.D.	not detectable
m/z	mass-to-charge-ratio	NAD(P)H	nicotinamide adenine dinucleotide phosphate
mg	milligram	NADH	nicotinamide adenine dinucleotide
Mg	magnesium	NEB	New England Biolabs
$MgCl_2$	magnesiumchloride	Ni-NTA	nickel-nitriloacetic acid
$MgSO_4$	magnesiumsulfate	NMR	nuclear resonance spectroscopy
Mhz	megahertz	NOESY	nuclear Overhauser effect spectroscopy
min	minute	OD_{600}	optical density at 600 nM
mL	milliliter	PAGE	polyacrylamide gel electrophoresis
mM	millimolar	PCR	polymerase chain reaction
mm	millimeter	PDB	protein data base
n	nano ($1 * 10^{-9}$)	Pfu	*Pyrococcus furiosus polymerase*
N	nitrogen	QM	quantum mechanic
NaCl	natrium chloride	RMSD	root mean square deviation
Ni	nickel	RT	retention time
nM	nanomolar	Rv	reverse
ns	nanosecond	SDS	sodium dodecyl sulfate
O	oxygen	SOB	super optimal broth
p	pico ($1 * 10^{-12}$)	TBLASTN	search translated nucleotides using a proteins
pH	pH-value	TIC	total ion current
pM	picomolar	TIP3P	water model in the CHARMM force field
PPD	polyprenyl diphosphate	TLC	thin layer chromatography
PP_i	pyrophosphate	TM	transmembrane
ppm	parts per milion	TMS	trimethylsilyl
rpm	rounds per minute	TMS-CL	chlorotrimethylsilyl
s	second	TPS	terpene synthase
S	sulfur	TS	transition state
δ	chemical shift in ppm	UV	ultra violett
Δ	double bond	VIS	visible
π	pi bond	WT	wild type
σ	sigma bond		

2. List of Figures

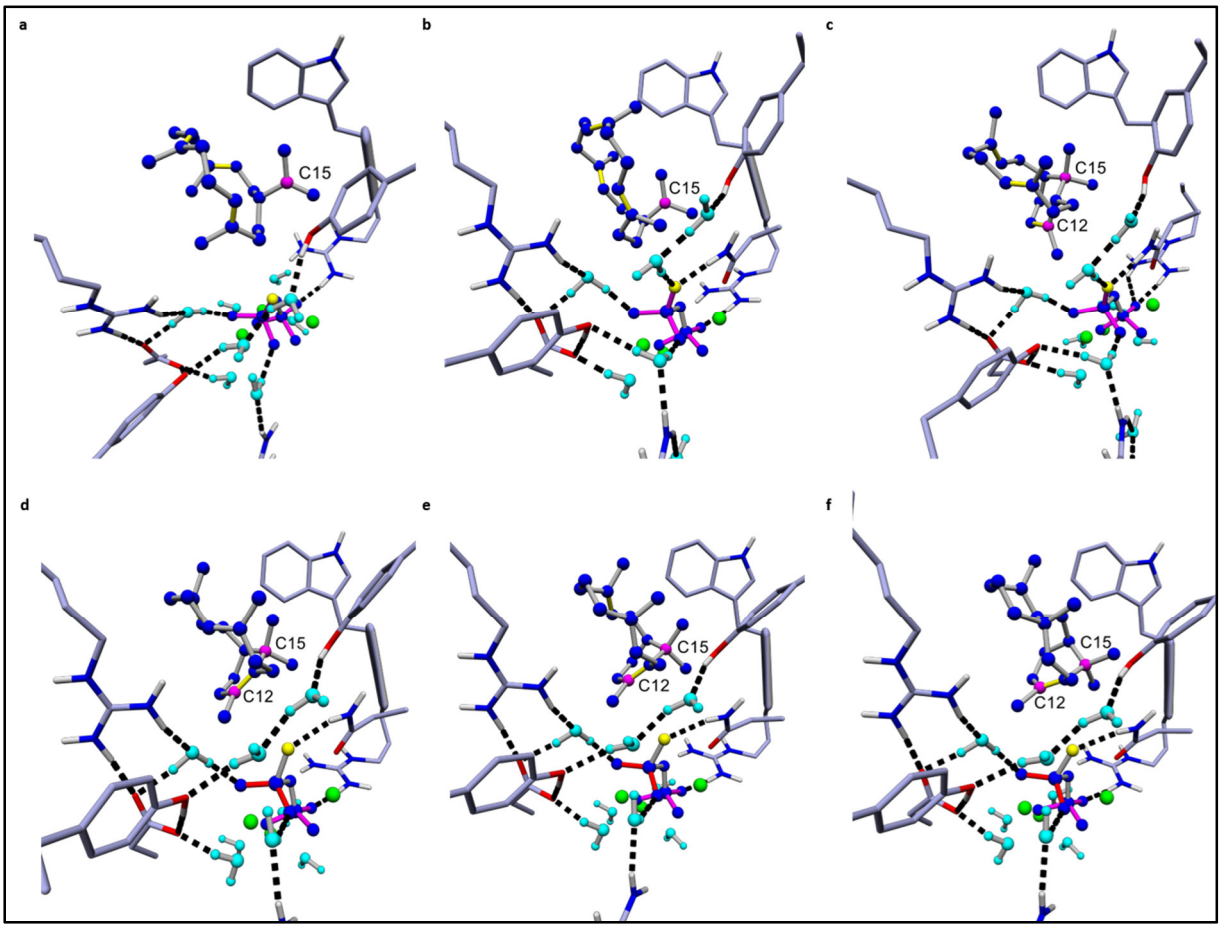
3. List of Tables

4. Appendix Figures and Tables

	Predicted free energy of binding for the cationic ligand with lowest dissociation constant in the respective cluster [kcal mol^{-1}]	Dissociation constant for the cationic ligand with highest predicted free energy of binding in the respective cluster [pM]	Free energy of binding spread of cationic ligands in the respective clusters [kcal mol^{-1}] (average and standard deviation)
Cation A			
Cluster			
1	8.4720	616512.1875	8.472 +- 0.0
2	**7.8910**	**1643691.5000**	**7.891 +- 0.0**
Cation B			
Cluster			
1	10.2850	28906.4688	9.3118 +- 0.5918
2	**8.6750**	**437663.8750**	**8.1593 +- 0.4584**
TS B-C			
Cluster			
1	10.0050	46370.0703	6.9122 +- 2.4141
2	**7.9690**	**1440938.8750**	**6.7142 +- 1.6325**
Cation C			
Cluster			
1	9.3070	150618.5781	7.5646 +- 1.2115
2	**8.4770**	**611331.2500**	**7.3673 +- 0.8611**
Cation F			
Cluster			
1	**10.0840**	**40581.6758**	**7.1243 +- 2.0671**
Cation D1			
Cluster			
1	9.1340	201692.2500	7.4407 +- 1.1357
2	**8.1720**	**1022926.8750**	**6.8544 +- 1.5005**
Cation D2			
Cluster			
1	9.1340	201692.2500	7.0101 +- 1.6486
2	**9.0110**	**248227.2344**	**6.7997 +- 1.7352**
Cation E			
Cluster			
1	**11.3550**	**4749.7925**	**6.7744 +- 2.4352**
2	7.9450	1500506.1250	6.7014 +- 1.0455

Appendix Table 1: **Cluster analyses of docked QM-carbocations into the transient enzyme intermediate complexes**

Cluster analyses of **QM-cations A-E** and **-TS B-C** docked into their corresponding transient enzyme intermediate complexes according to methods chapter II. **Cations A**, **-B**, **-C**, **-D1**, **-D2**, **-E** and **TS B-C** exhibit two distinct conformational clusters in their corresponding transient enzyme intermediate complexes. **Cation F** exhibits only one distinct conformation. Two conformations belonged to different clusters, if the ligand RMSD was larger than 2 Å. Only cluster that are distinctly derived from the conformation of unionized **GGPP** (cation A) or the respective former cascade cation (**cation B-E** and **TS B-C**), depicted in red were considered for validation and structural superposition of the respective manually constructed cations. These cationic structures are shown in Figure 27 (**cation E**), Figure 14 and 17 (**cation A**), 19 and 30 (**cation B** and **TS B-C**), 22 (**cation C**), 23 (**cation F**), 24 (**cation D1**) and 25 (**cation D2**). The conformational cluster that is not structurally derived from unionized **GGPP** or the respective former cascade cation, depicted in black are shown in Figure 17 (**cation A**) and Appendix Figure 1 (**cation B-E** and **TS B-C**). Note that single conformations of **cations F** and **-E** in their respective considered cluster exhibit the overall highest predicted free energies of binding (second column), indicating the most favorable interaction with their respective receptor. This in line with the assumption that **cations C-E** are not in equilibrium state as suggested by QM gas phase calculations.[58] Instead, it indicates that **cation C->F** transition is an exothermic event and **cation E** is indeed the global energy minimum of the reaction cascade in relation to **cation A**. Moreover, cluster analyses demonstrate that according to the predicted spread in free energies of binding (fourth column) the average predicted free energies of binding of the cationic ligands decrease from cation **B->E** and that **cation A** and -**B** exhibit the highest average free energies of binding as well as the lowest standard deviations. This indicates that a high percentage of conformations in the cluster of **cation A** and **-B** exhibit a more favorable interaction with the respective receptor than **cations C-E**. This, in turn, is in line with our observation that **cations C-E** suffer from a reduced binding in TXS, resulting in cation tumbling and imprecise barrier crossings that lead to the observed side products derived from these cationic intermediates.



Appendix Figure 1: Representation of unproductive docking cluster

Representation of unproductive docking cluster in accordance to Appendix Table 1. **a**: Representation of the conformational cluster of **cation B**, that is, due to its localization it the active site, structurally not derived from productive **cation A** (Figure 14 and 17), obtained upon docking of QM-derived **cation B** into its corresponding transient enzyme complex. **b**: Representation of the conformational cluster of **TS B-C**, that is, due to its localization it the active site, structurally not derived from productive **cation A**, obtained upon docking of QM-derived **TS B-C** into its corresponding transient enzyme complex. **c**: Representation of the conformational cluster of **cation C**, that is, due to its localization in the active site, structurally not derived from productive **TS B-C**, obtained upon docking of QM-derived **cation C** into its corresponding transient enzyme complex. **d**: Representation of the conformational cluster of **cation D1**, that is, due to its localization it the active site, structurally not derived from productive **cation C**, obtained upon docking of QM-derived **cation D1** into its corresponding transient enzyme complex. **e**: Representation of the conformational cluster of **cation D2**, that is, due to its localization in the active site, structurally not derived from productive **cation D1**, obtained upon docking of QM-derived **cation D2** into its corresponding transient enzyme complex. **f**: Representation of the conformational cluster of **cation E**, that is, due to its localization in the active site,

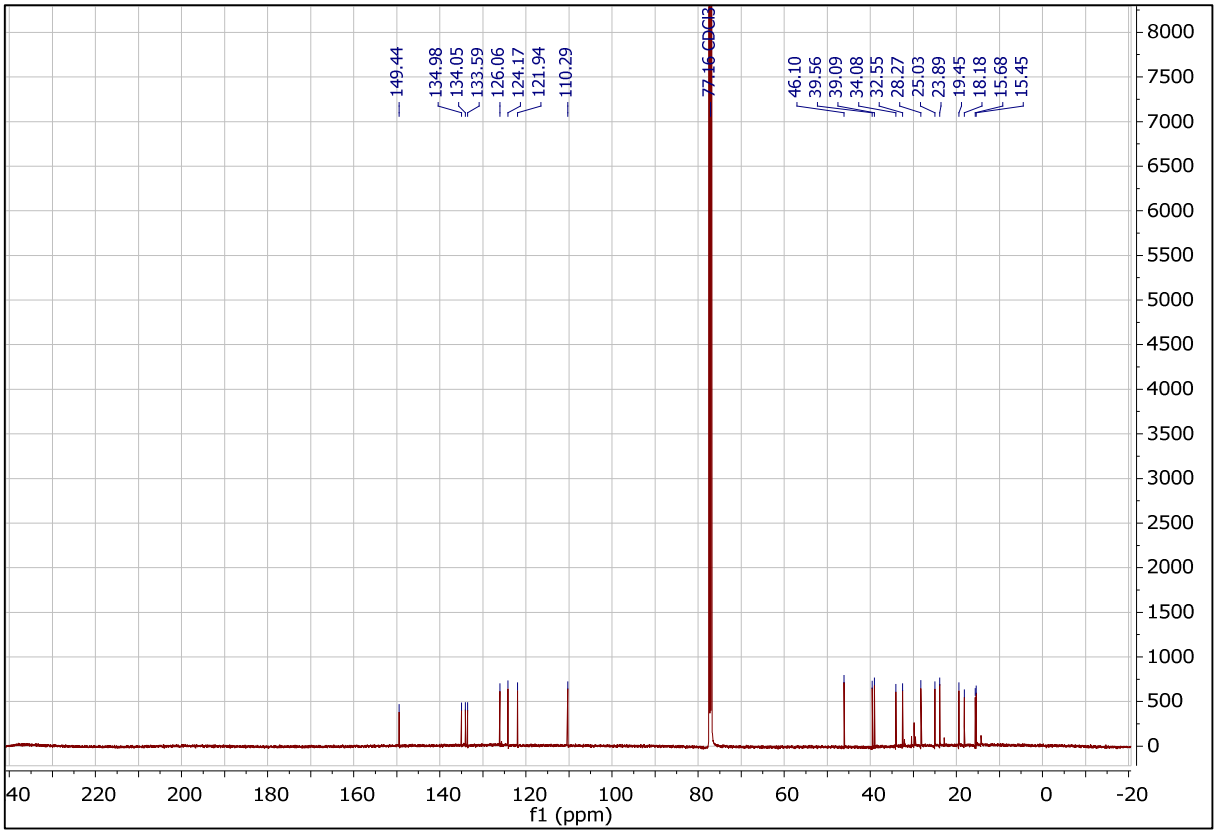
structurally not derived from productive **cation D2**, obtained upon docking of QM-derived **cation E** into its corresponding transient enzyme complex.

Target	Mutation	Activity [a]	E [b]	E1	B	C	F2	F1
TXS		100,0	93,2	4,7	N.D.	0,8	0,2	1,1
Y89	Y89F	0,0						
	Y89A	0,0						
	Y89E	2,8	100,0	N.D.	N.D.	N.D.	N.D.	N.D.
R580	R580H[c]	0,0						
	R580A[c]	0,0						
	R580E[c]	0,0						
R754	R754H	0,0						
	R754A	0,0						
	R754E	0,0						
R768	R768H	0,0						
	R768A	0,0						
	R768E	0,0						
V584	V584M[c]	92,3	13,8	0,6	N.D.	83,4	0,8	1,4
	V584K[c]	89,1	29,8	0,6	N.D.	67,4	0,8	1,4
	V584L	92,1	13,8	0,6	N.D.	83,4	0,8	1,4
	V584N[c]	0,0						
	V584S[c]	0,0						
	V584P[c]	0,0						
	V584R	0,0						
S587	S587D	0,0						
	S587Y	0,0						
	S587K	0,0						
	S587L	0,0						
	S587A[c]	21,8	8,9	N.D.	58,9	32,2	N.D.	N.D.
	S587G[c]	0,0						
F602	F602W[c]	9,0	100,0	N.D.	N.D.	N.D.	N.D	N.D.
	F602I[c]	22,0	93,4	4,9	N.D.	N.D.	N.D.	1,7
	F602A[c]	1,0	100,0	N.D.	N.D.	N.D.	N.D.	N.D.
V610	V610H	0,0						
	V610S	0,0						
	V610F	0,0						
	V610A	0,0						
S713	S713T	97,4	92,7	5,1	N.D.	1,0	0,4	0,8
	S713A	0,0						
	S713L	0,0						
V714	V714T	1,2	100,0	N.D.	N.D.	N.D.	N.D.	N.D.
	V714A	8,7	94,3	5,7	N.D.	N.D.	N.D.	N.D.
	V714I	10,4	94,1	5,9	N.D.	N.D.	N.D.	N.D.

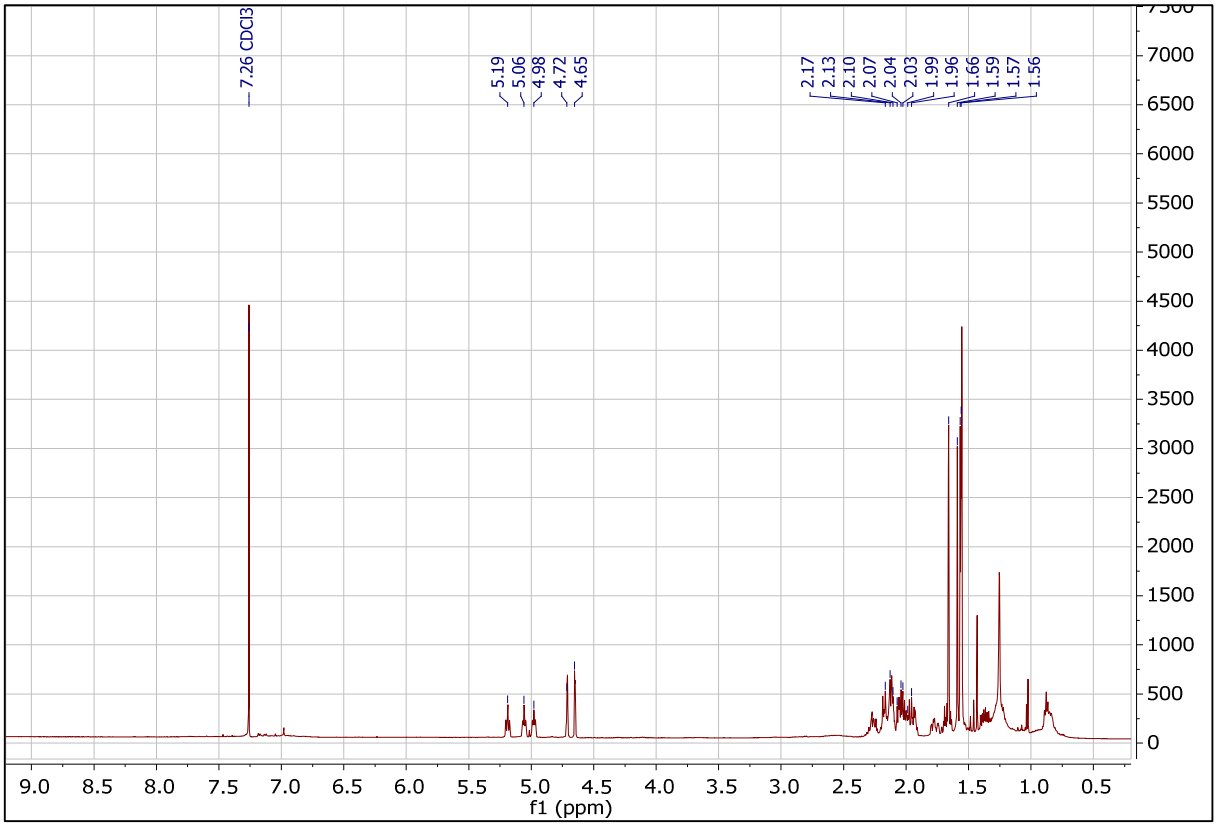
	V714G	0,0						
	V714P	0,0						
G715	G715A[c]	0,0						
	G715S[c]	0,0						
S713	S713T	97,4	N.D.	N.D.	100,0	N.D.	N.D.	N.D.
	S713A	0,0						
	S713L	0,0						
W753	W753H[c]	51,3	N.D.	N.D.	100,0	N.D.	N.D.	N.D.
	W753L[c]	0,0						
	W753V[c]	0,0						
	W753E[c]	0,0						
	W753C[c]	0,0						
	W753H/C830A	48,5	N.D.	N.D.	100,0	N.D.	N.D.	N.D.
	W753A[c]	0,0	93,5	4,6	N.D.	0,7	0,2	1,0
C830	C830A	88,5	93,1	4,8	N.D.	0,8	0,2	1,1
	C830S	92,4	79,3	7,3	N.D.	1,1	2,1	10,2
F834	F834A	25,6	38,3	6,5	N.D.	8,8	21,9	24,5
	F834G	28,3	87,4	6,1	N.D.	0,8	2,5	3,2
	F834Y	93,1						
	F834H	0,0						
Y835	Y835F[c]	3,5	100,0	N.D.	N.D.	N.D.	N.D.	N.D.
	Y835W[c]	0,0						
	Y835A[c]	0,0						
Y841	Y841A	0,0						
	Y841F	41,3	N.D.	N.D.	56,4	43,6	N.D.	N.D.
	Y841T	0,0						

Appendix Table 2: *Characterization of TXS and mutants*

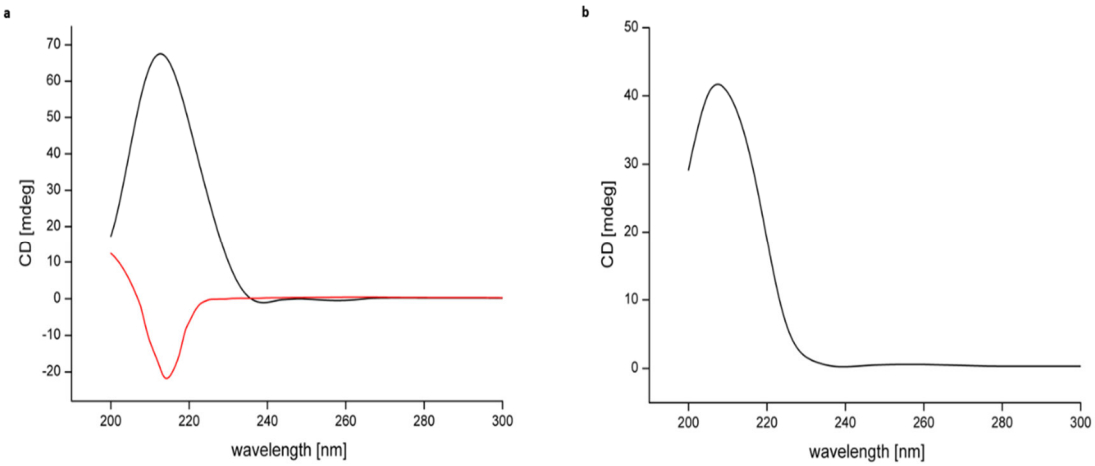
Appendix Table 2**a**: Enzyme activities of mutants (in %) in comparison to TXS (= 100%). The activities were determined by peak integration of the enzyme products in relation to the peak area of the internal standard alpha-humulene (methods chapter II). **b**: Compound numbers **B-E1** refer to compound numbers in Figure 12 and Table 1. The tabulated values represent the product distributions in % for TXS and each mutant (sum = 100%). (N.D.: Not detectable). c: The mutations have been conducted during and included in the master thesis of Alexander Büttner.[75]



Appendix Figure 2: 13C NMR Spectrum of *B*

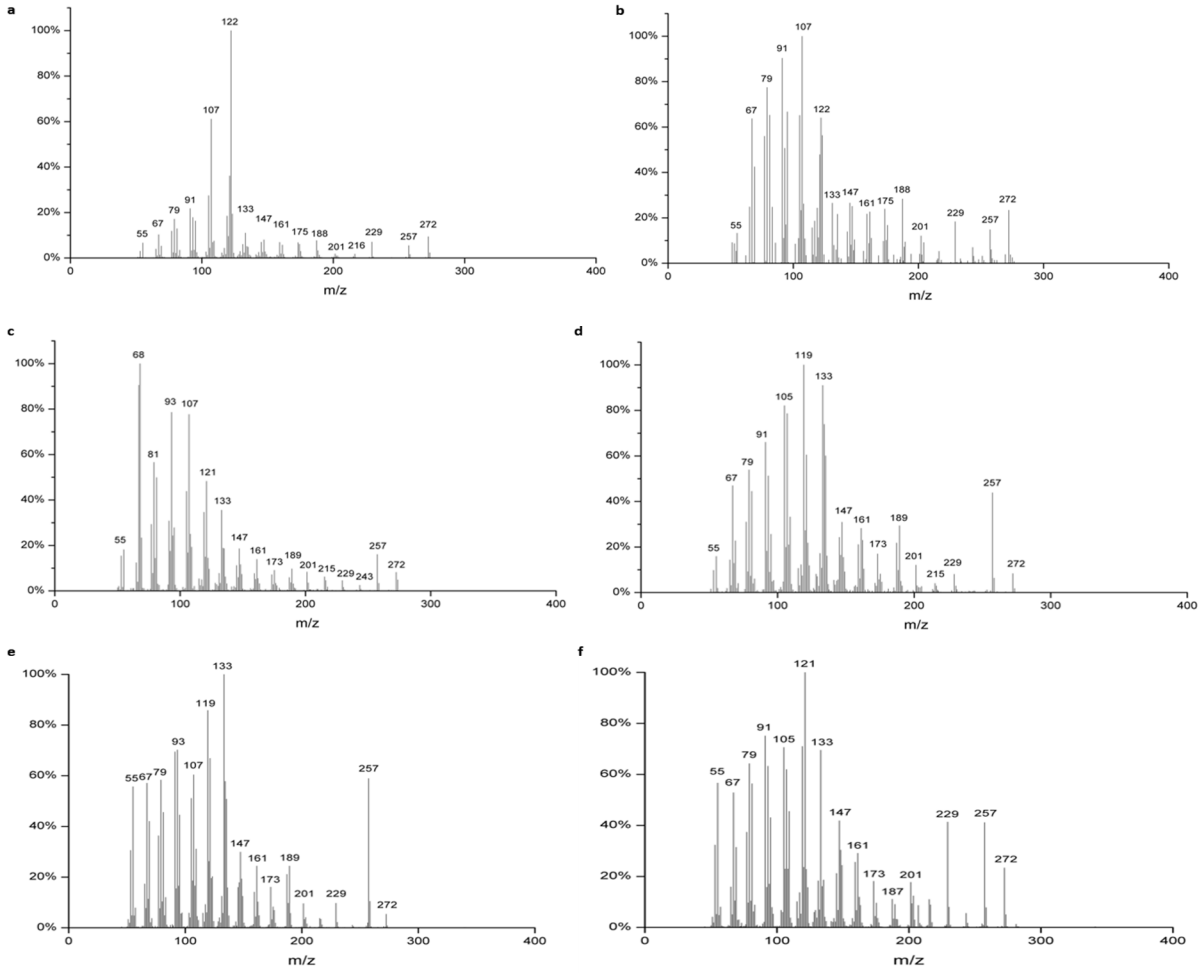


Appendix Figure 3: 1H NMR Spectrum of B



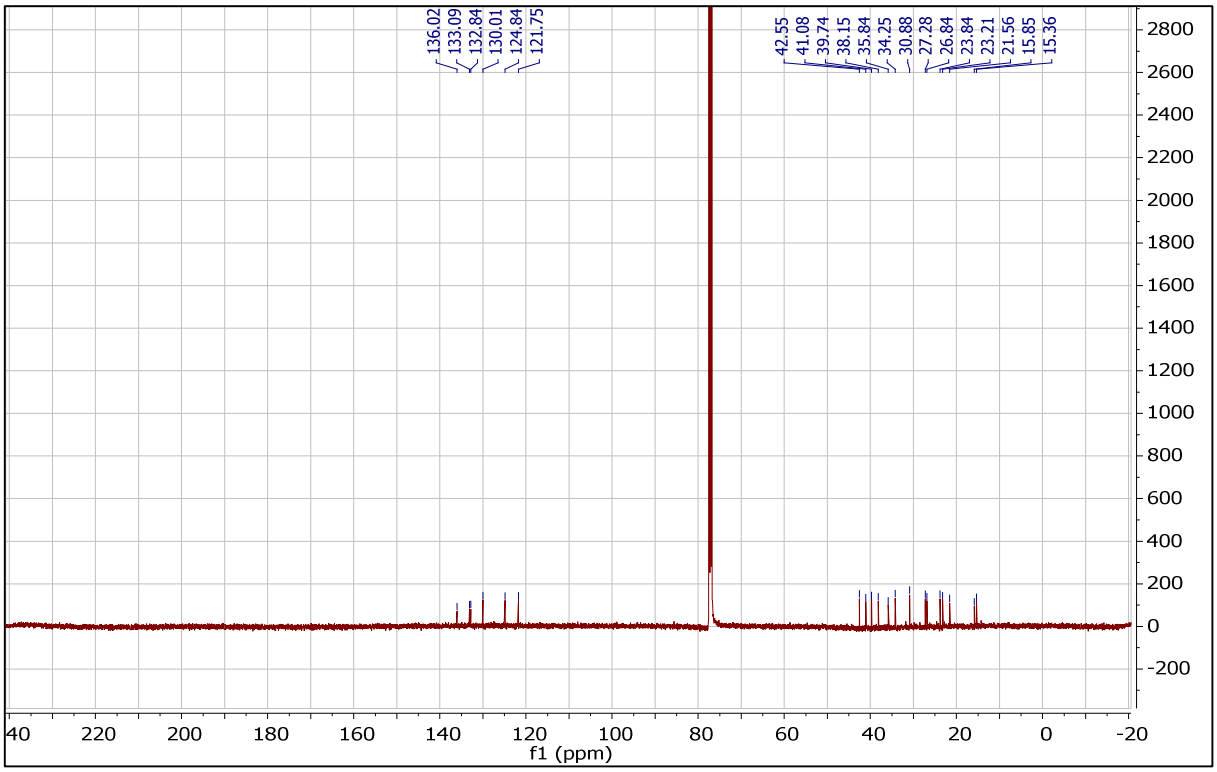
Appendix Figure 4: CD spectra of B, (+)-cembrene and C

Appendix Figure 4: CD-spectra of **a**: **(+)-cembrene** (black line) and **B** (red line), **b**: CD-spectrum of **C** (black line).

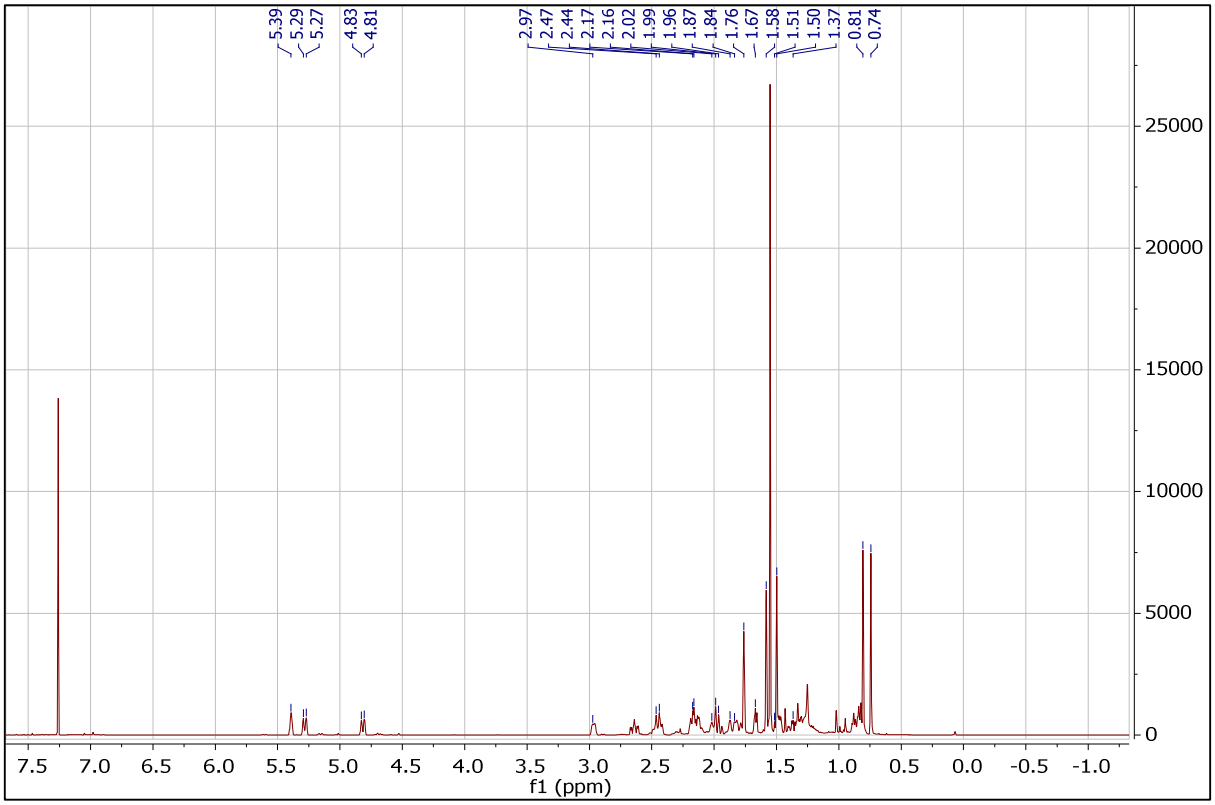


Appendix Figure 5: MS-Spectra of deprotonated intermediates B, C, F1, F2, E and E1

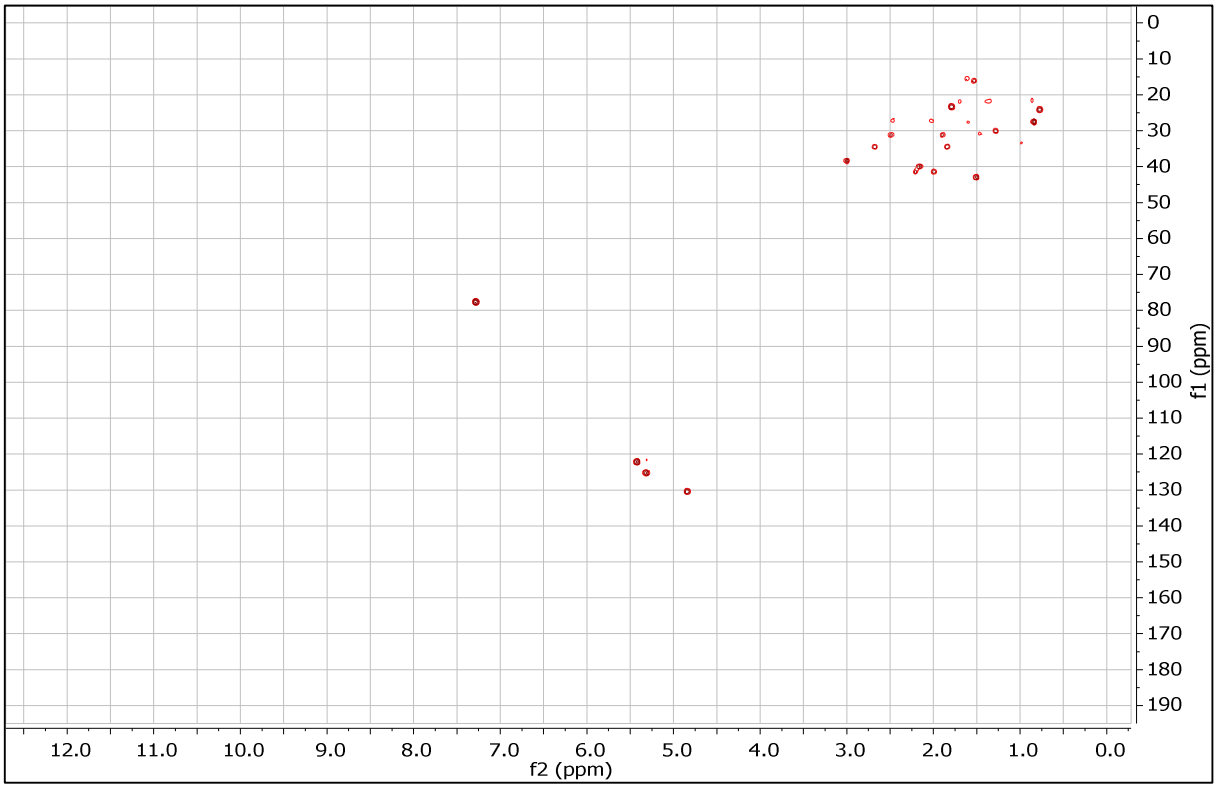
Appendix Figure 5: Mass spectra of deprotonated intermediates according to Figure 12. (**a**) **E**: taxa-4,5-11,12-diene, (**b**) **E1**: taxa-4,20-11,12-diene, (**c**) **B**: cembrene A, (**d**) **C**: verticilla-3,4-7,8-12,13-triene, (**e**) **F2**: verticilla-4,20-7,8-11,12-triene, (**f**) **F1**: verticilla-3,4-7,8-11,12-triene.



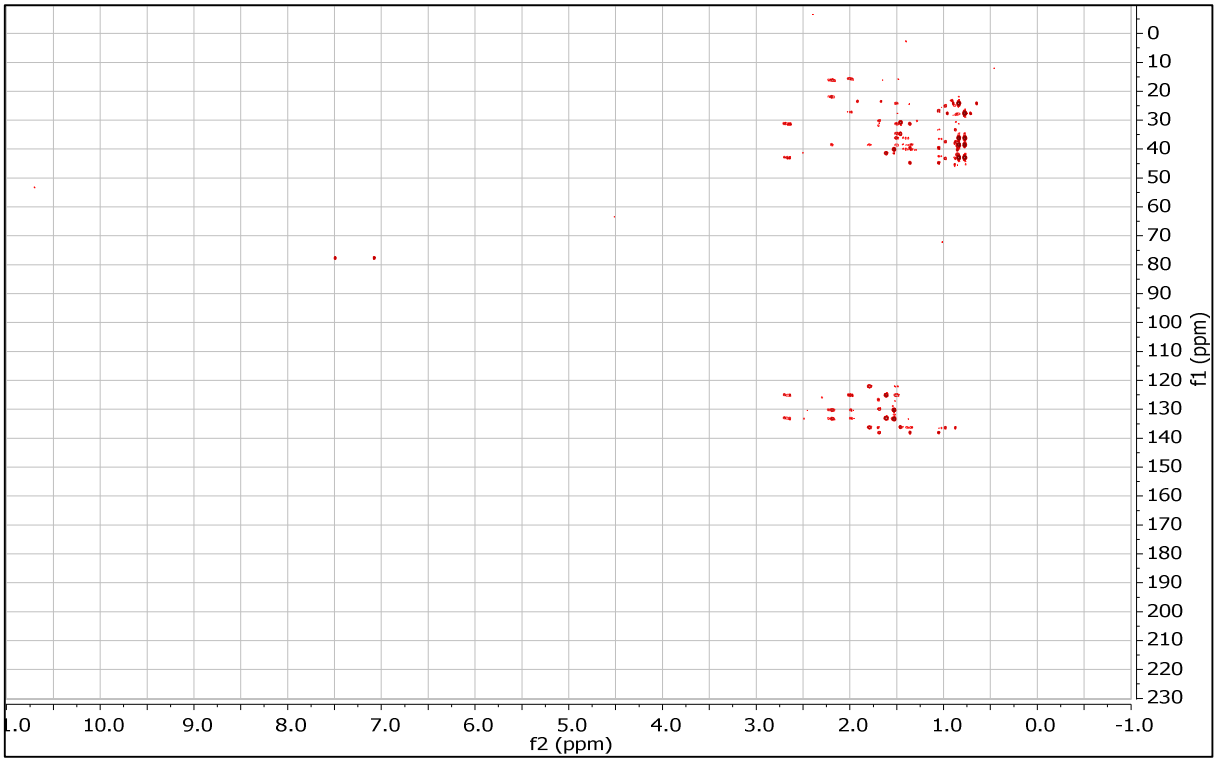
Appendix Figure 6: 13-C Spectrum of C



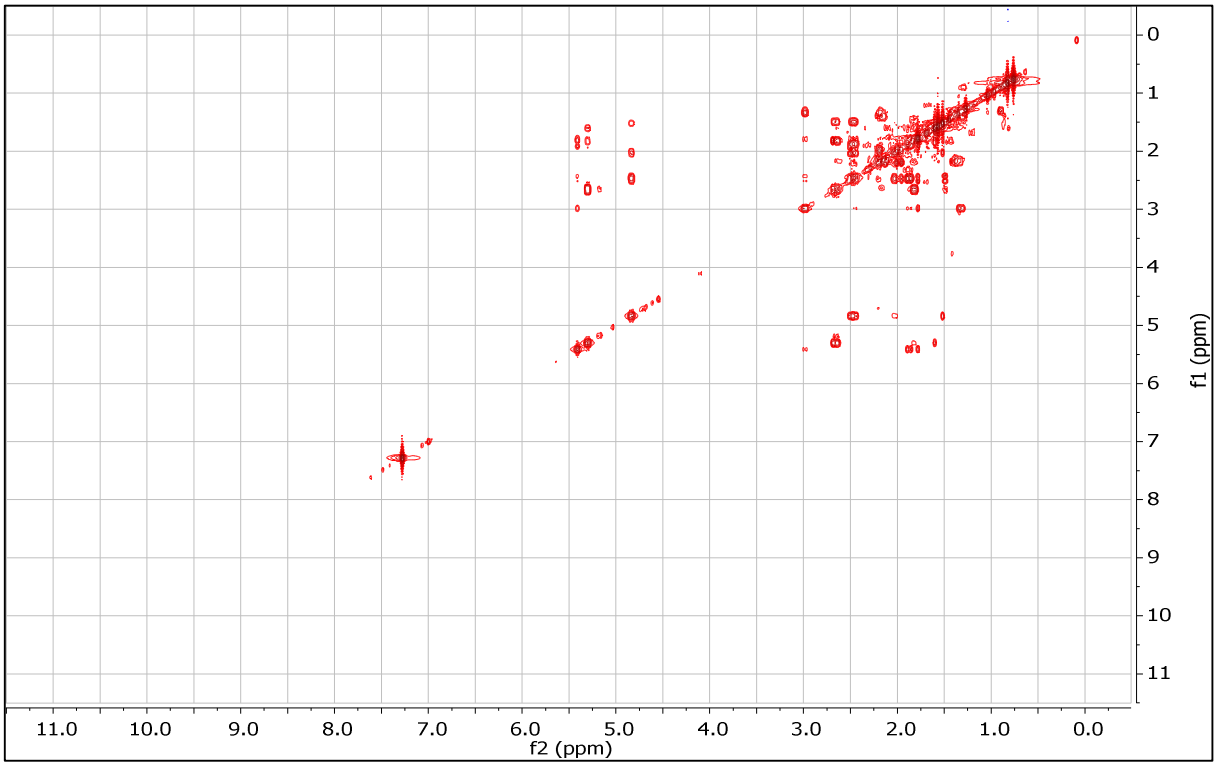
Appendix Figure 7: 1-H Spectrum of C



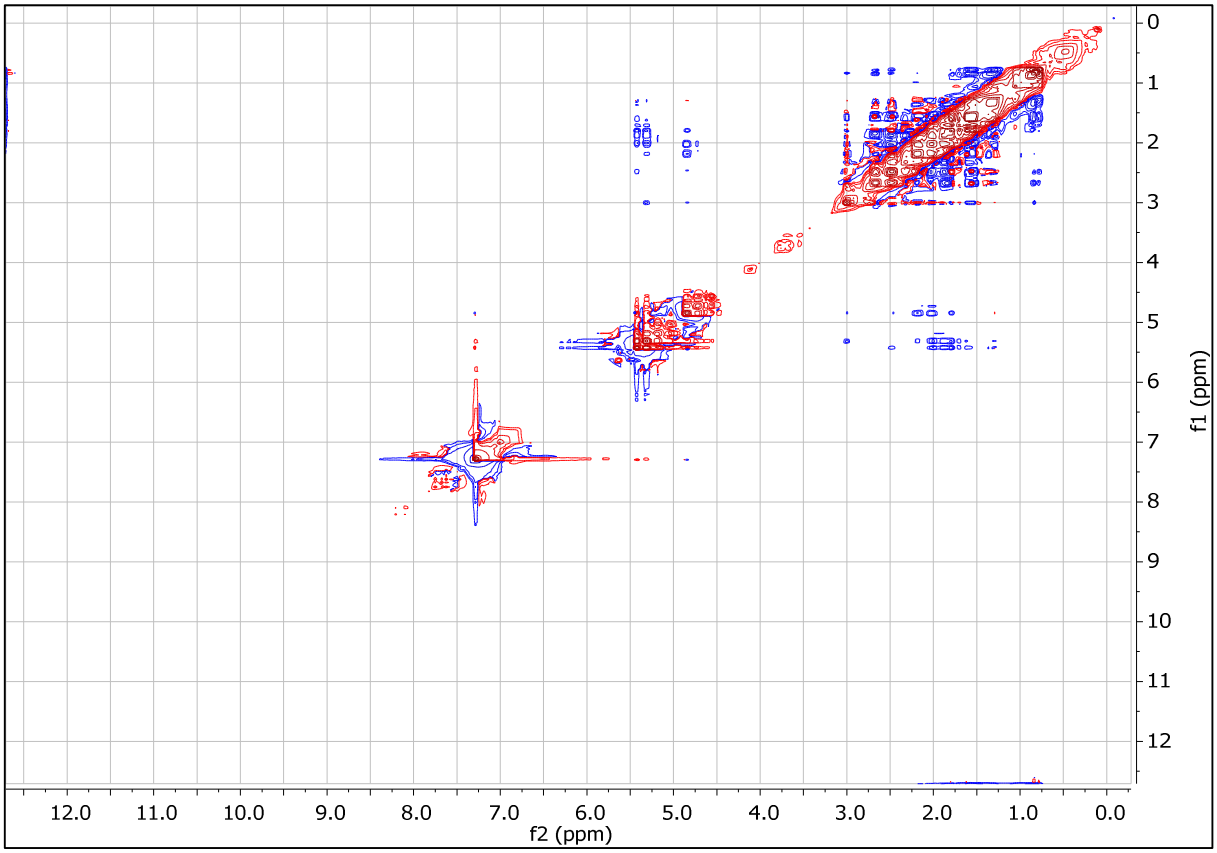
Appendix Figure 8: HSQC Spectrum of C



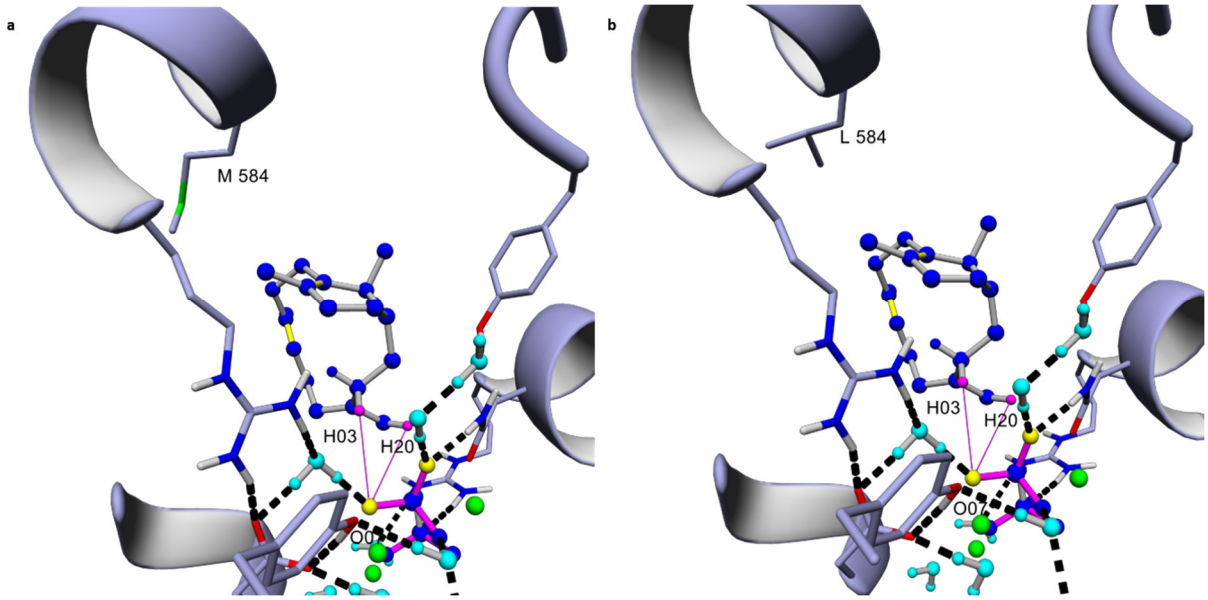
Appendix Figure 9: HMBC Spectrum of C



Appendix Figure 10: COSY Spectrum of C

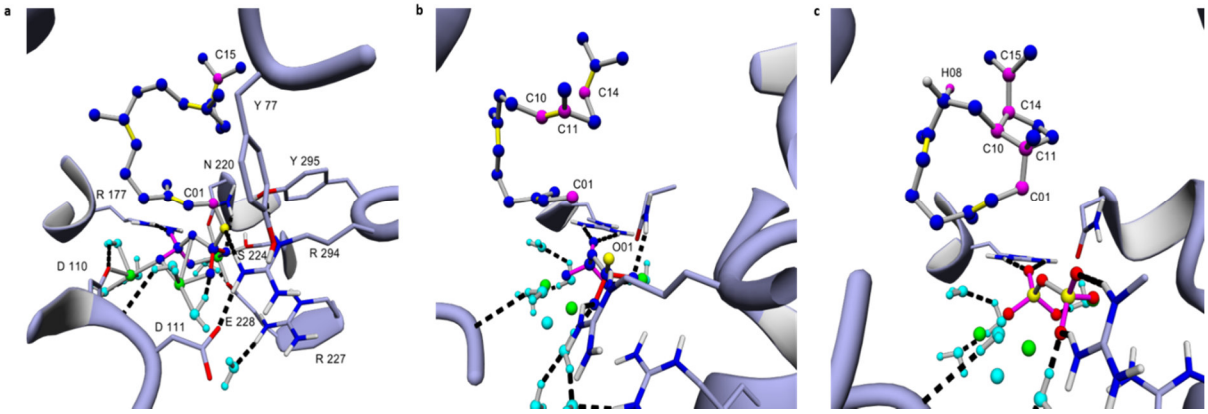


Appendix Figure 11: NOESY Spectrum of C



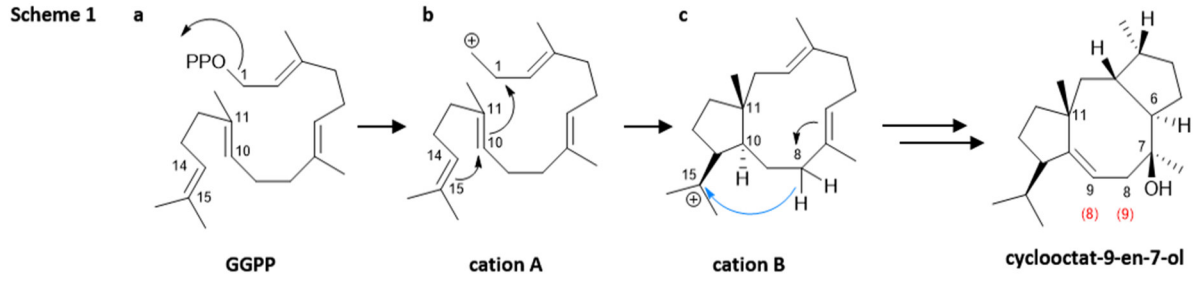
Appendix Figure 12: TXS-V584M/L harboring cation F

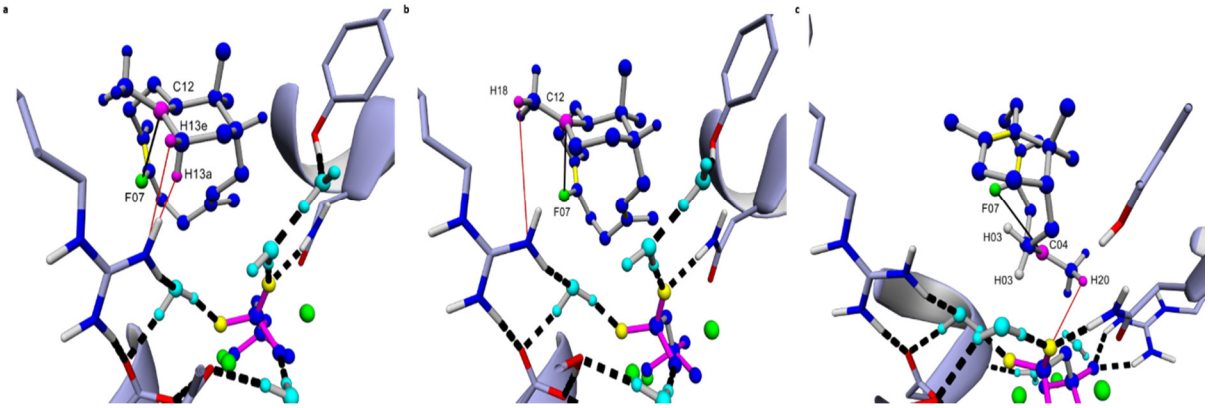
Appendix Figure 12**a**: TXS-V584M harboring manually constructed **cation F**. b: TXS-V584L harboring manually constructed **cation F**. Magenta lines indicate the distance of H03 and H20 (closest) to the R580-**PP$_i$** bi-functional motif.



Appendix Figure 13: Closed conformation of CotB2 with productive GGPP, cation A and cation B

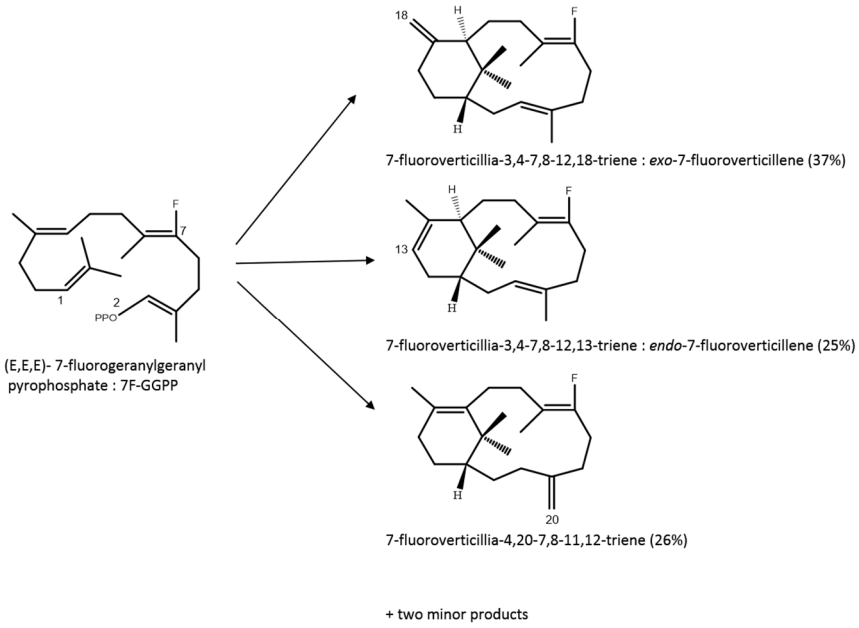
Appendix Figure 13**a**: Modelled closed conformation of CotB2 harboring productive **GGPP** (C01 and C15 of **GGPP** colored in magenta, Mg^{2+} ions in green, O01 of **GGPP** in yellow, the magnesium and bulk water coordinating amino acids are shown. For more information, see (Schrepfer et al., PNAS, submitted). **b**: CotB2 harboring **cation A** (scheme 1). **c**: CotB2 harboring **cation B** (scheme 1).





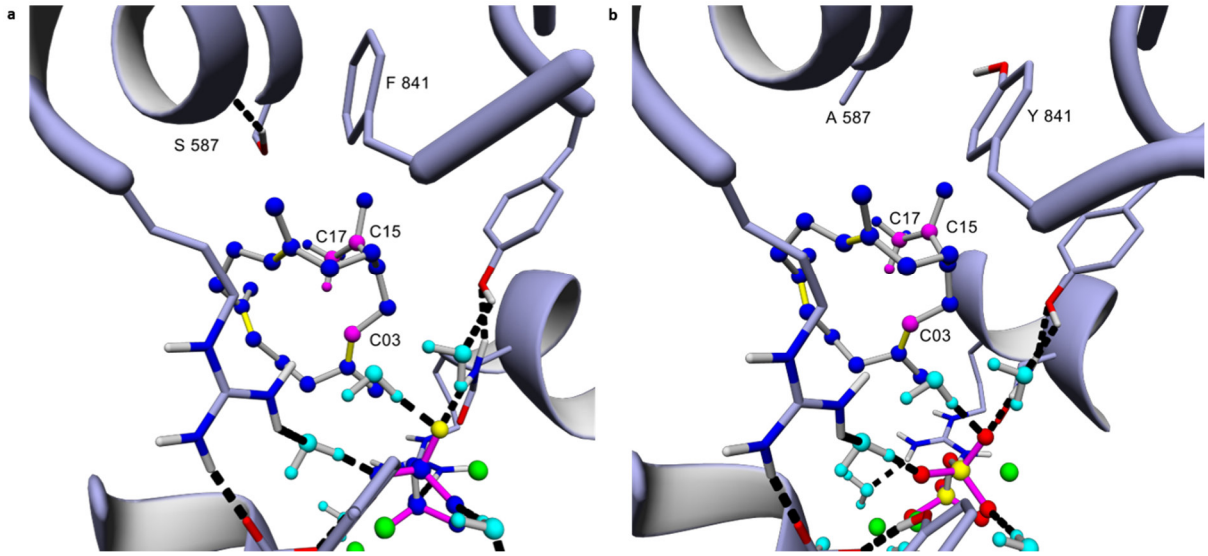
Appendix Figure 14: TXS•cation C and -F complexes harboring 7-fluoro-GGPP-derived cations C and F

Appendix Figure 14**a-b**: TXS harboring the manually constructed **7-fluoro-GGPP-derived cation C**. **a**: The 7-fluoro-ligand is shown in green, positive charged C12 and H13axial (H13a) as well as H13equatorial (H13e) presumably abstracted by R580-PP_i, resulting in formation of *endo*-7-fluoroverticillene are shown in magenta. Distance between C12 and F07 is 3.012 Å, distance between H13a and R580-PP_i is 3.158 Å, distance between H13e and R580-PP_i is 3.290 Å. **b**: The 7-fluoro-ligand is shown in green, positive charged C12 and closest H18 presumably abstracted by R580-**PP_i**, resulting in formation of *exo*-7-fluoroverticillene are shown in magenta. Distance between C12 and F07 is 3.012 Å, distance between closest H18 and R580-**PP_i** is 4.306 Å. **c**: TXS harboring manually constructed **7-fluoro-cation F**. The 7-fluoro-ligand is shown in green, positive charged C04 and closest H20 presumably abstracted by R580-**PP_i**, resulting in formation of 7-fluoroverticillia-4,20-7,8-11,12-triene are shown in magenta. H03 atoms are shown in grey. Distance between C04 and F07 is 3.334 Å, distance between closest H20 and R580-**PP_i** is 3.411 Å.



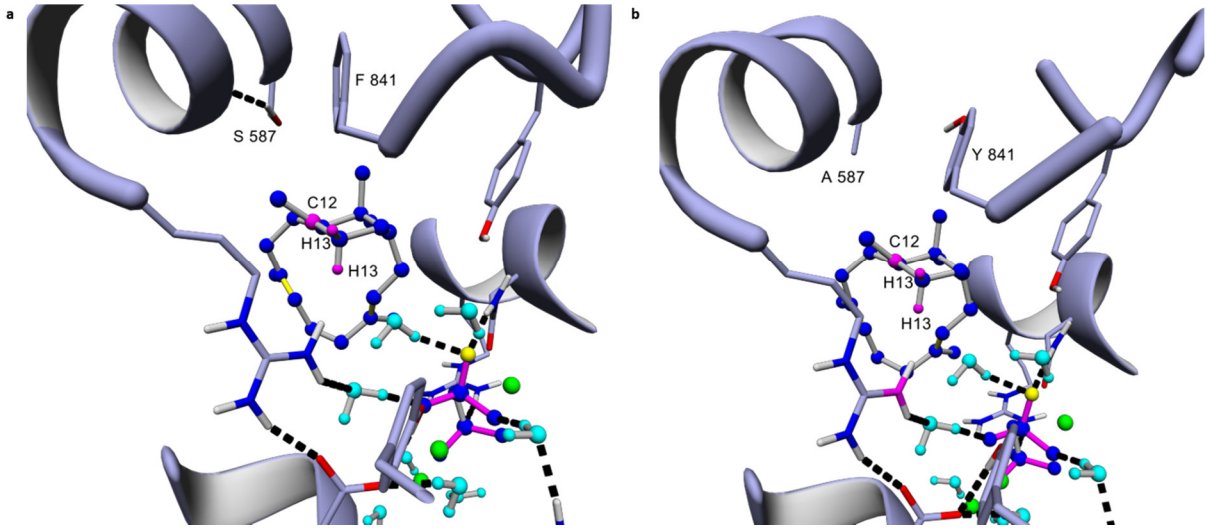
Appendix Figure 15: Observed product distribution of TXS incubated with 7-fluoro-GGPP

Appendix Figure 15: The TXS**•cation C** and **-F** complexes harboring the manually constructed **7-fluoro-GGPP-derived cations C and -F** (Appendix Figure 14) are able to give an indication for the product distributions observed during incubation with 7-fluoro-**GGPP** (Appendix Figure 15).[68] These experiments conducted by Jin and coworkers, show a 37% outcome of *exo*-7-fluoroverticillene. Our models (Appendix Figure 14a, b) show closer distances between H13a/e, abstracted in formation of *endo*-7-fluoroverticillene, and R580-**PP$_i$**, compared to the distance between closest H18 (to R580-**PP$_i$**), abstracted in formation of *exo*-7-fluoroverticillene and R580-**PP$_i$**. This would suggest a favored *endo*-7-fluoroverticillene formation. However, the models further indicate that the 7-fluoro ligand is able to counteract H13a/e abstraction by a transient electrostatic attraction of H13a/e. This attraction is reduced in respect to closest H18, resulting in the observed product distribution by Jin and coworkers, derived from deprotonation of 7**-fluoro-cation C**. The distance between F07 and H13a is 3.247 Å and the distance between F07 and H13e is 4.358 Å. In contrast, distance between closest H18 and F07 is 4.859 Å. The electrostatic attraction of F07 with H03 protons in **7-fluoro-cation F** (Appendix Figure 14c) is moreover able to explain the absence of a 7-fluoroverticillia-3,4-7,8-11,12-triene as another major product in their observed product distribution, despite close proximity of H03 protons to R580-**PP$_i$**, corresponding to the observed formation of **F1** in Figure 23a.



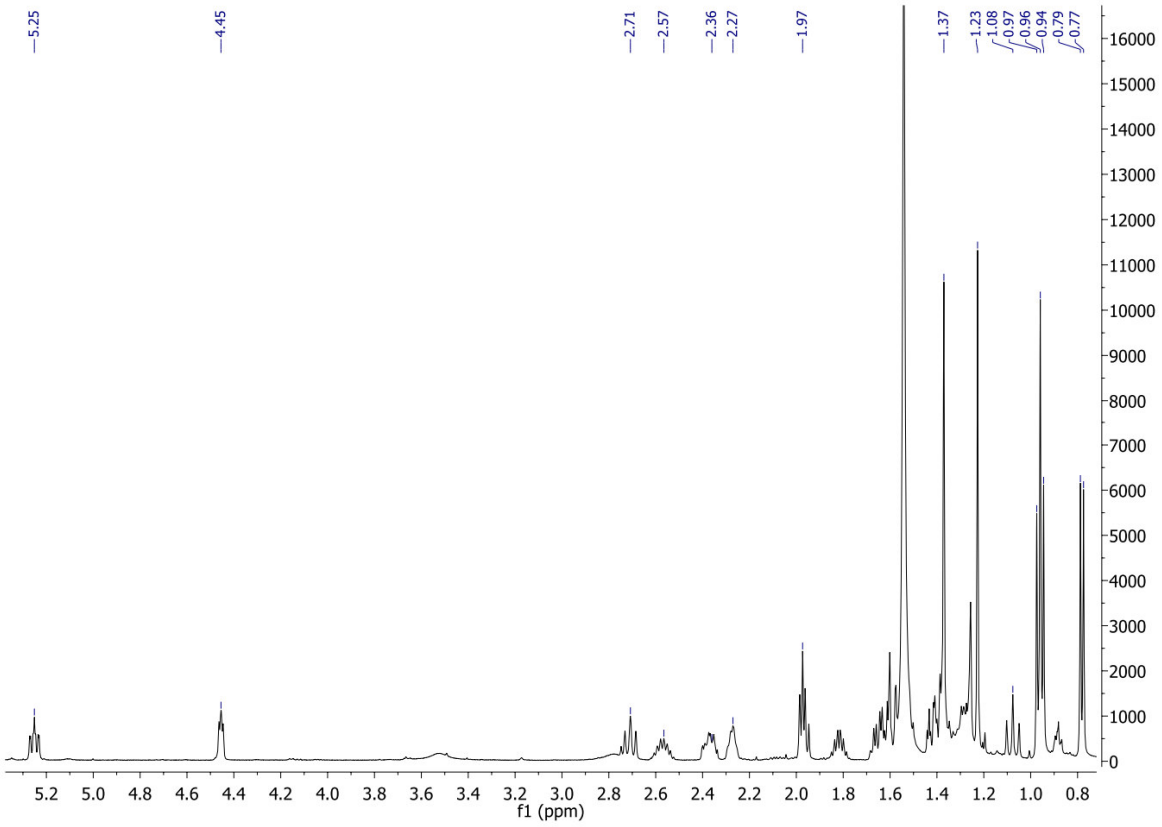
Appendix Figure 16: TXS-Y841F and TXS-S587A harboring cation B

Appendix Figure 16**a**: TXS-Y841F harboring manually constructed **cation B**. Distance between "liberated" and thus cyclization interfering S587-OH and cationic C15 is 4.162 Å. **b**: TXS-S587A harboring manually constructed **cation B**. Distance between "liberated" and thus cyclization interfering Y841-OH and cationic C15 is 4.862 Å. Formation of **B** in Y841F and S587A mutants can be explained by electrostatic stabilization of the proximal positive charge at C15 by S587 or Y841. An interfered transition of **cation B** to **-C** would enable, governed by R580-**PP$_i$**, a not native hydride-shift of a single C17 hydrogen to Δ3, 4 in **cation B**. This, in turn, brings the positive charge into close proximity to R580-**PP$_i$** with consequent deprotonation of C3.

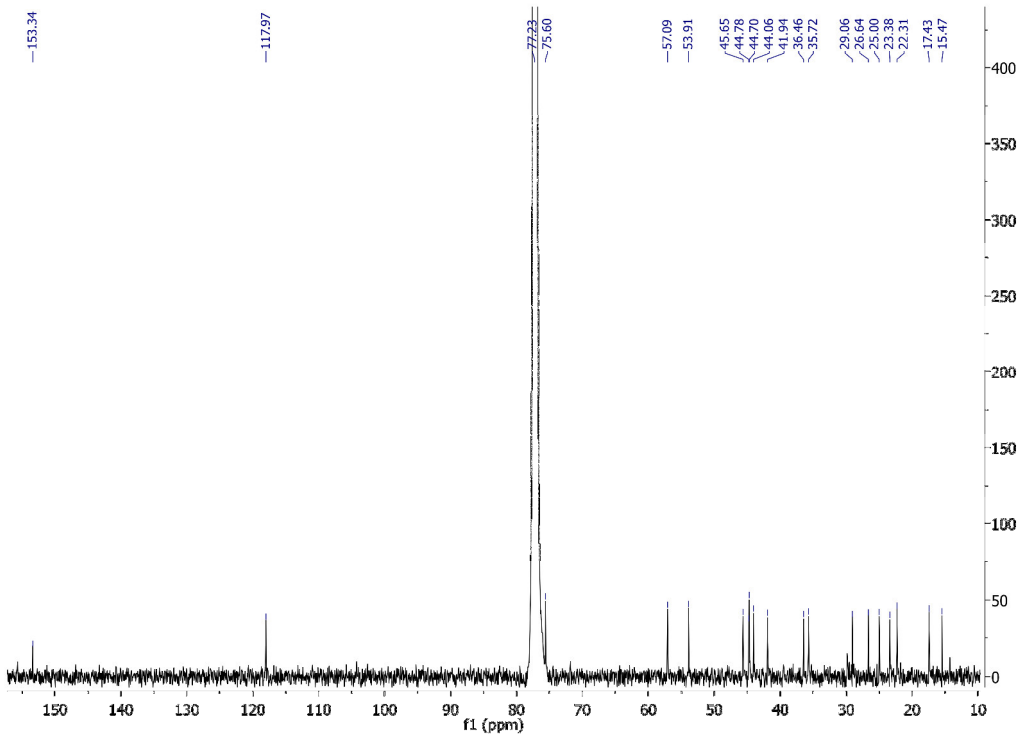


Appendix Figure 17: TXS-Y841F and TXS-S587A harboring cation C

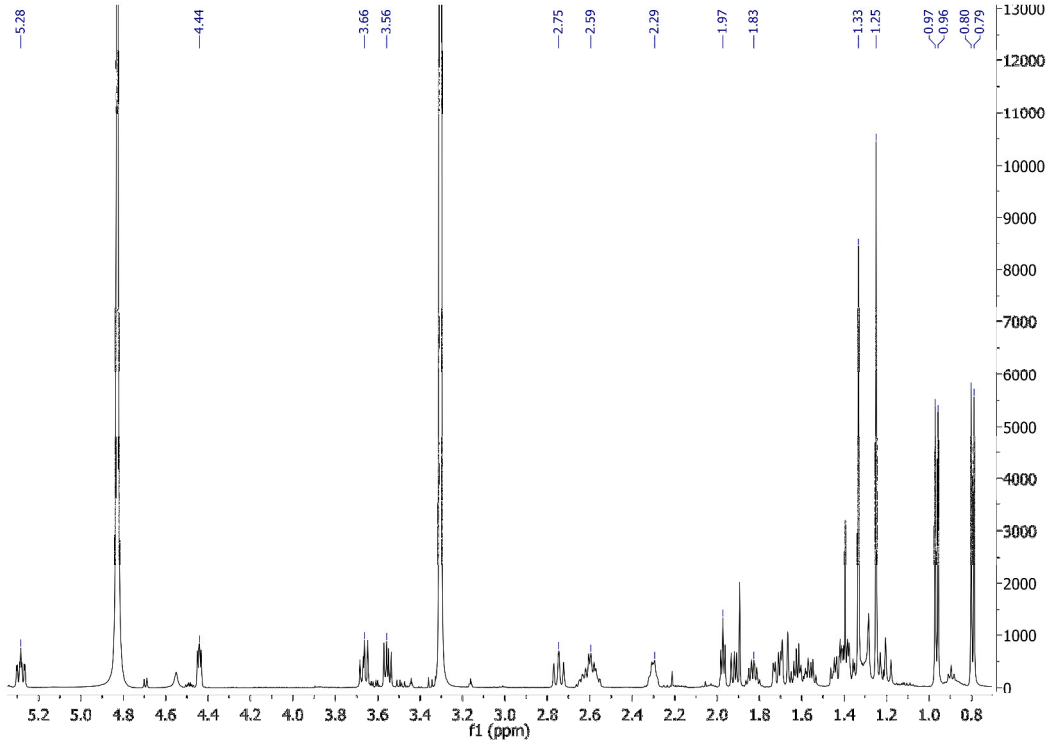
Appendix Figure 17**a**: TXS-Y841F harboring manually constructed **cation C**. Distance between "liberated" and thus cyclization interfering S587-OH and cationic C12 is 3.859 Å. **b**: TXS-S587A harboring manually constructed **cation C**. Distance between "liberated" and thus cyclization interfering Y841-OH and cationic C12 is 5.805 Å. Deprotonation of C13 resulting in formation of **C** seems to be facilitated by electrostatic stabilization of cationic C12 by "liberated" S587 or Y841, respectively. This stabilization leads consequently to an interfered and slowed transition to **cation F**, accompanied by simultaneous cation steering by R580-**PP$_i$** towards itself. The bi-functional motif seems then accordingly to be the active site base.



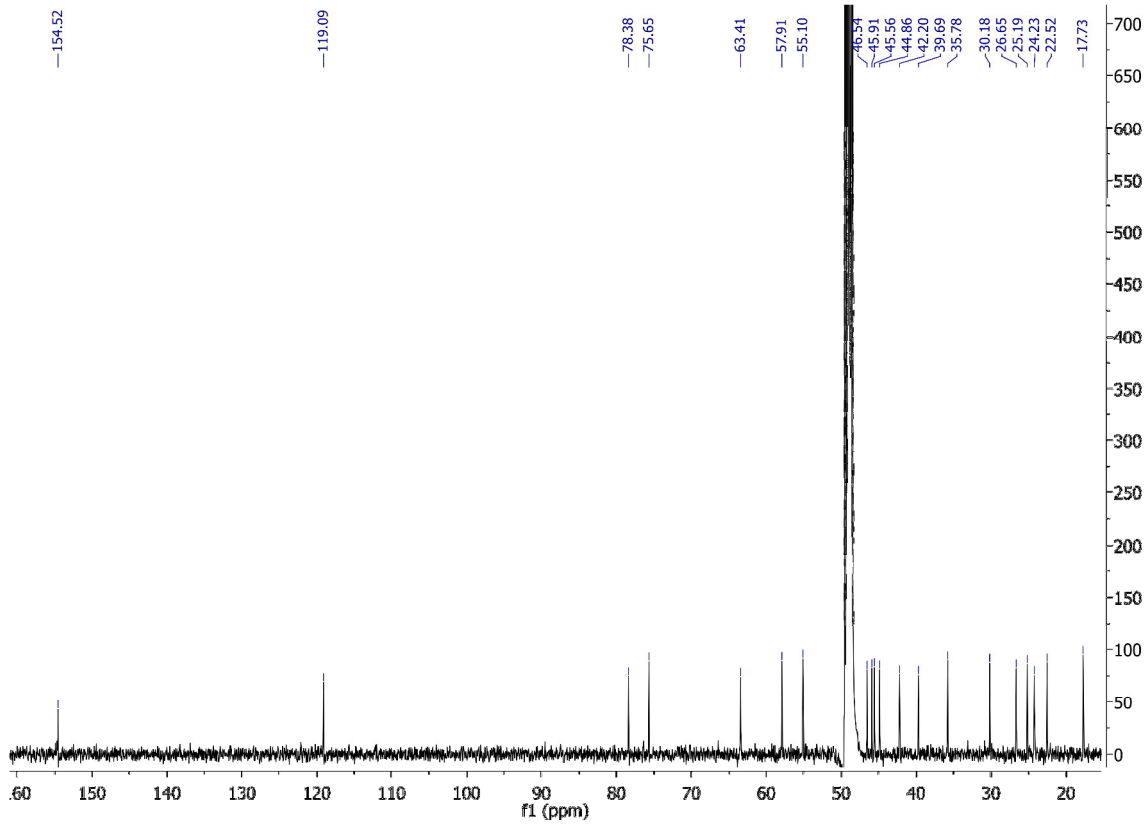
Appendix Figure 18: 1H spectrum of cyclooctat-9-en-5,7-diol



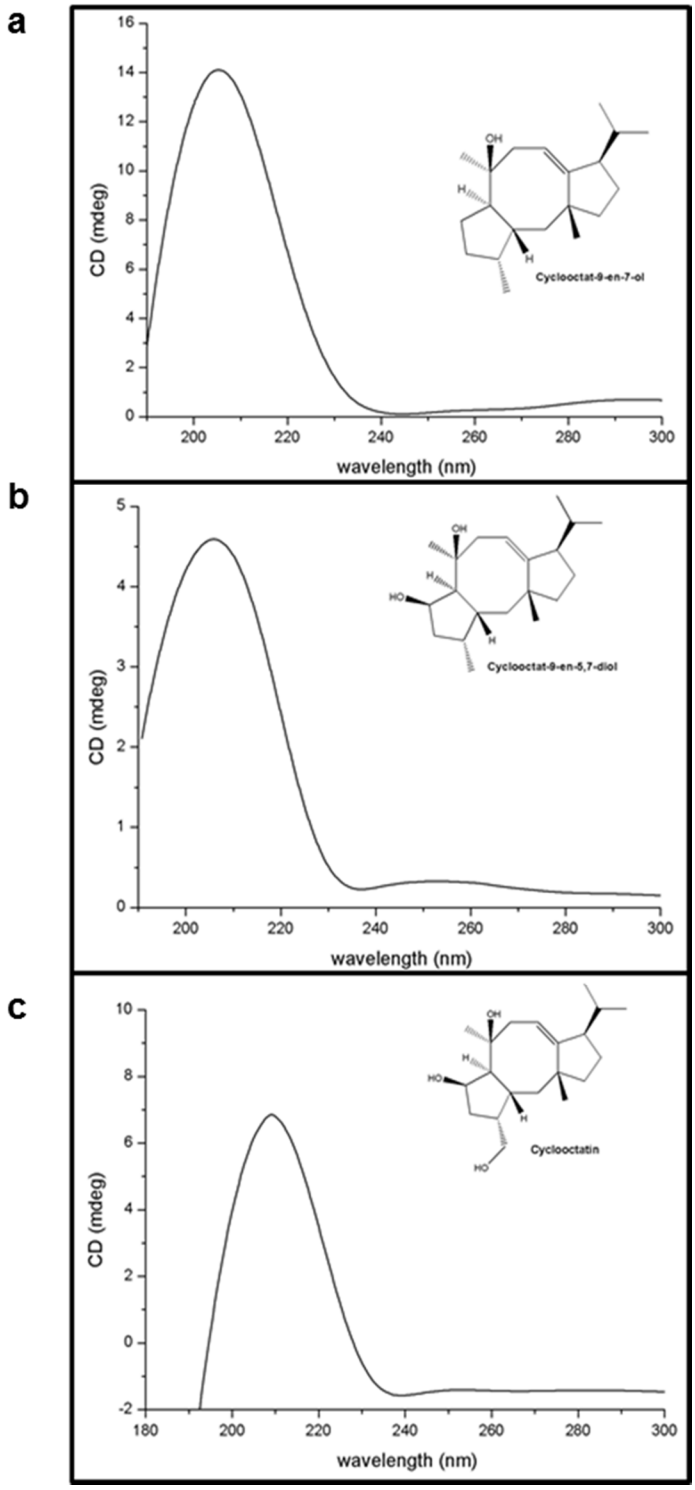
Appendix Figure 19: 13C spectrum of cyclooctat-9-en-5,7-diol



Appendix Figure 20: 1H spectrum of cyclooctatin

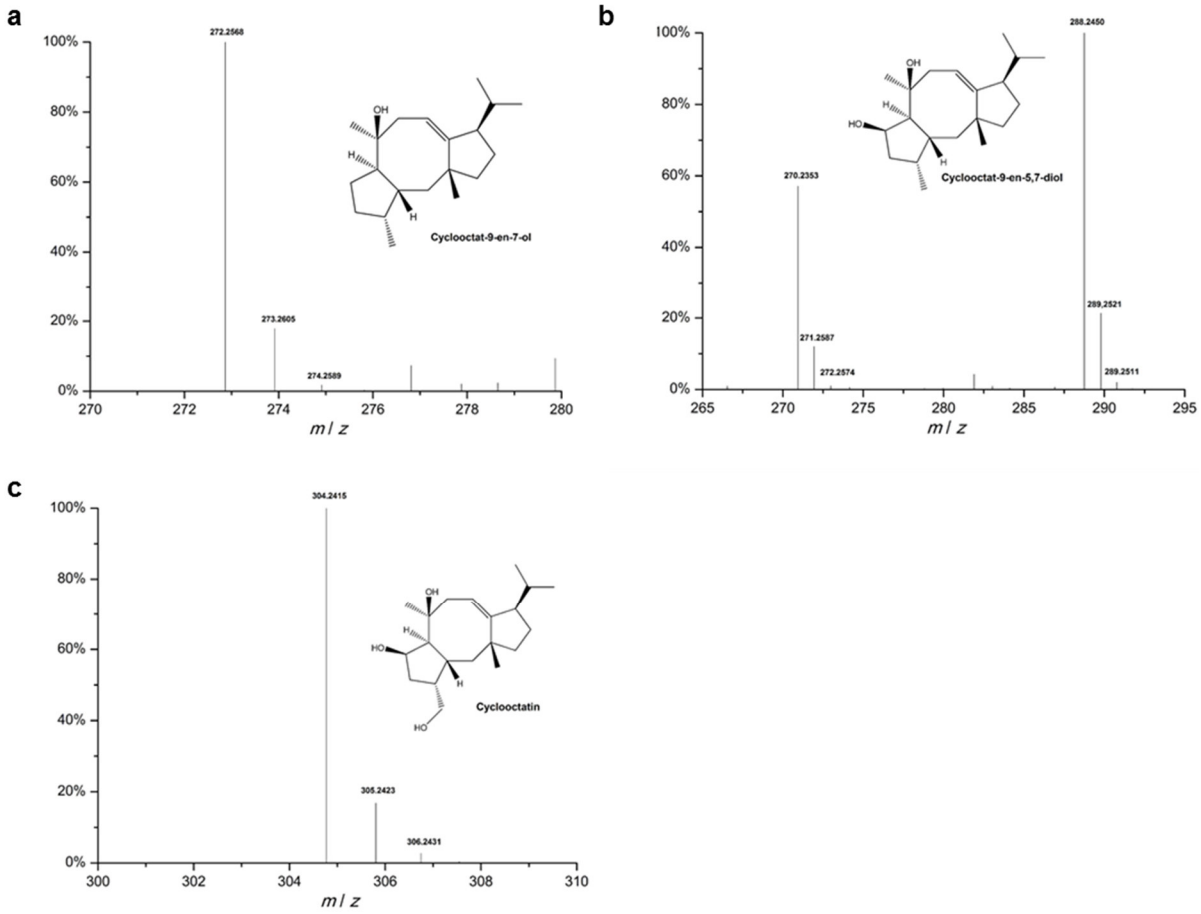


Appendix Figure 21: 13C spectrum of cyclooctatin



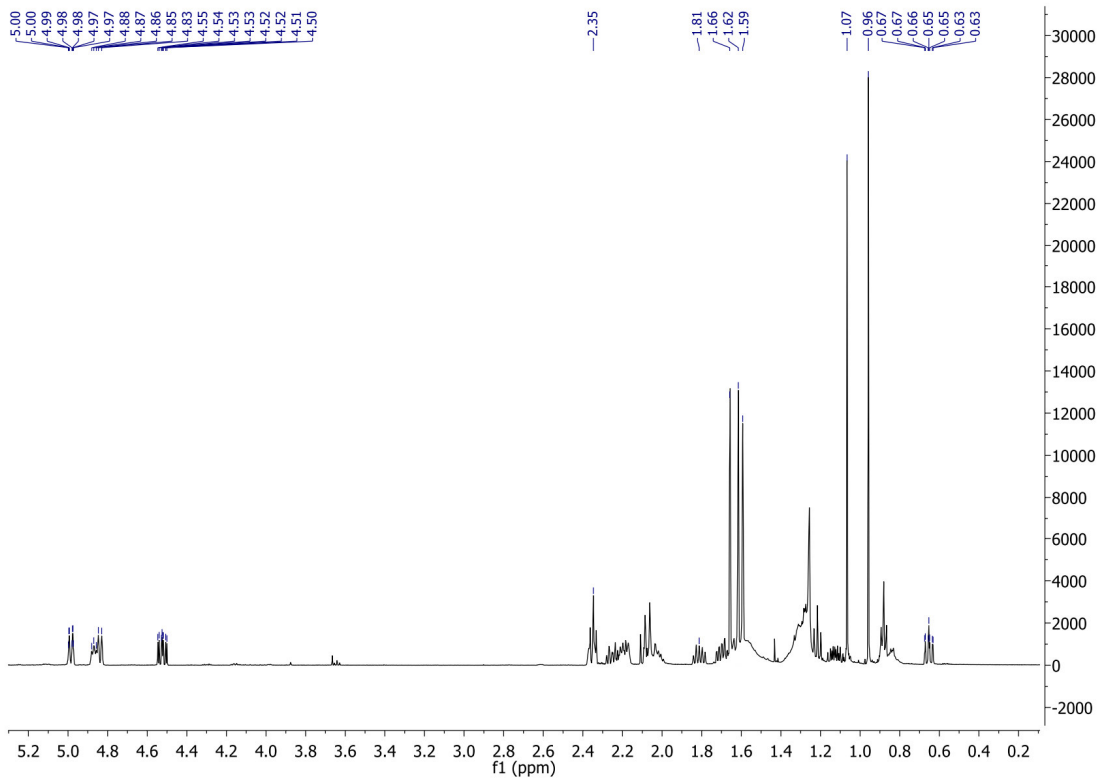
Appendix Figure 22: CD spectra of cyclooctatin and its metabolic precursor

Appendix Figure 22**a**: CD spectrum in HCN of **cyclooctat-9-en-7-ol**, **b**: **cyclooctat-9-en-5,7-diol** and **c**: **cyclooctatin**.

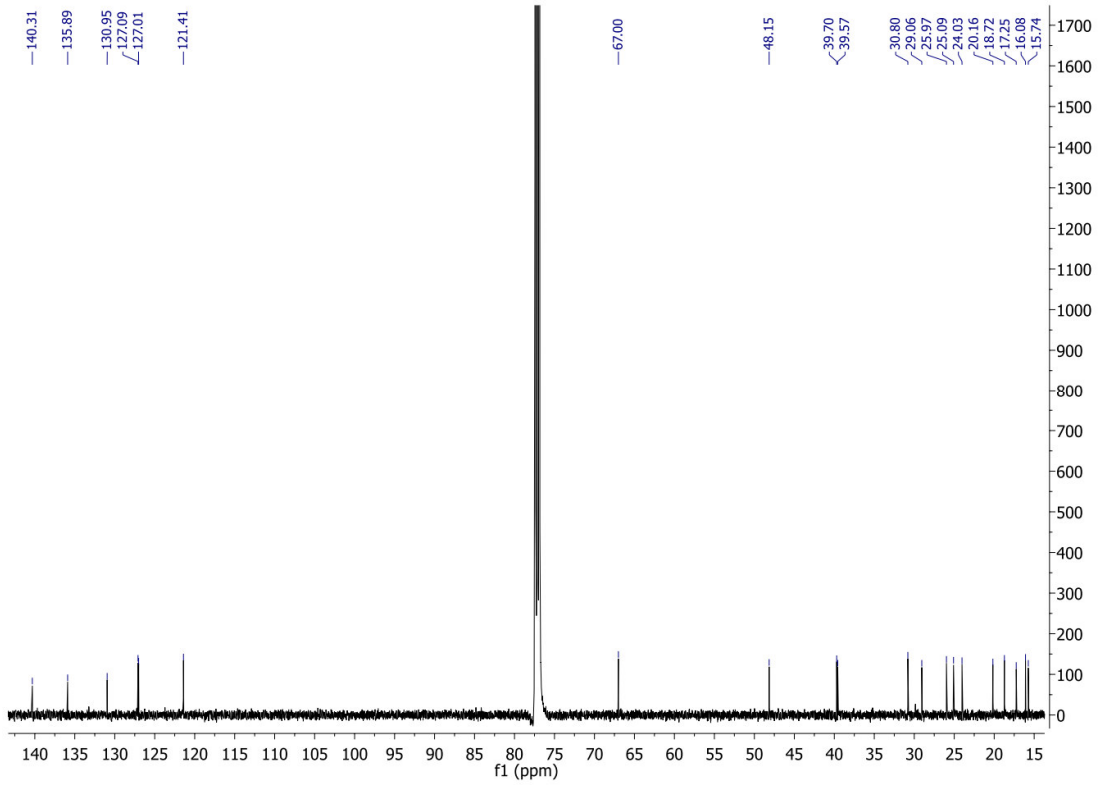


Appendix Figure 23: High resolution LC-MS spectra for cyclooctatin and ist metabolic precursor

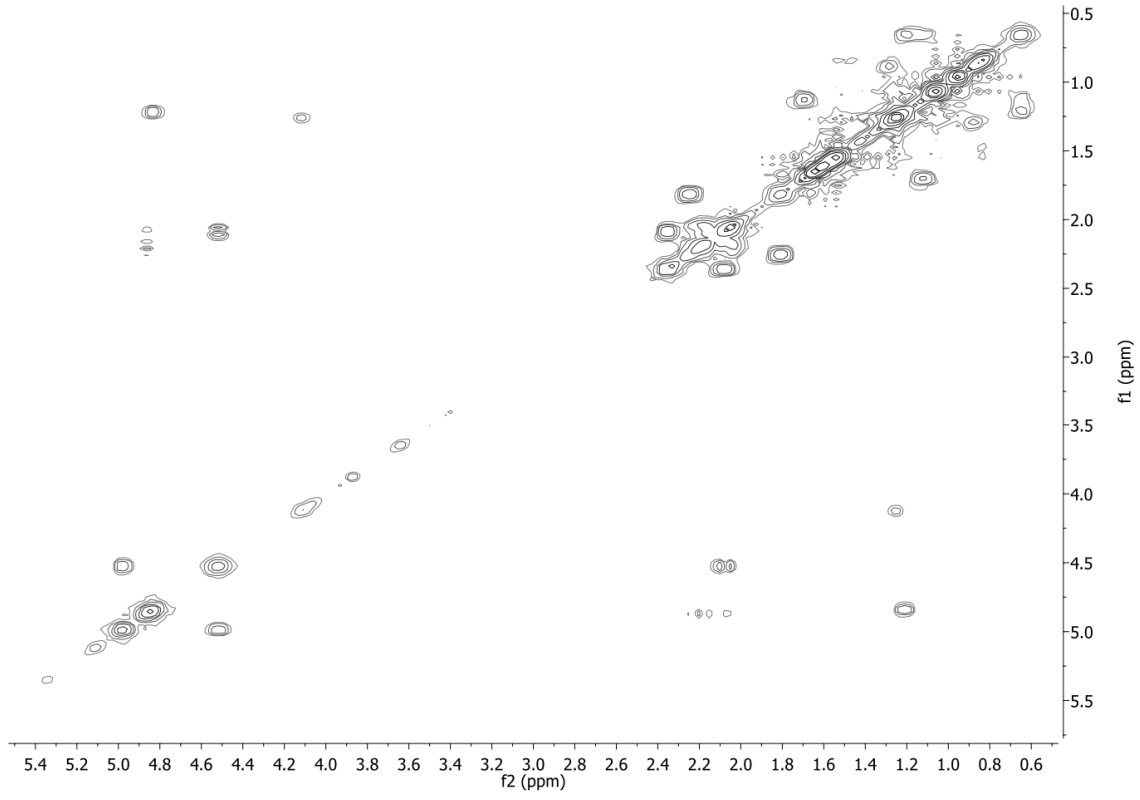
Appendix Figure 23: High resolution electron spray ionization of **a**: **cyclooctat-9-en-7-ol**, *m/z* ($[M–H_2O]^+$ 272.2568, calculated $C_{20}H_{32}$ 272.2504, **b**: **cyclooctat-9-en-5-7-diol**, *m/z* ($[M–H_2O]^+$ 272.2450, calculated $C_{20}H_{32}O$ 288.2453 and **c**: **cyclooctatin**, *m/z* ($[M–H_2O]^+$ 304.2415, calculated $C_{20}H_{32}O_2$ 304.2402.



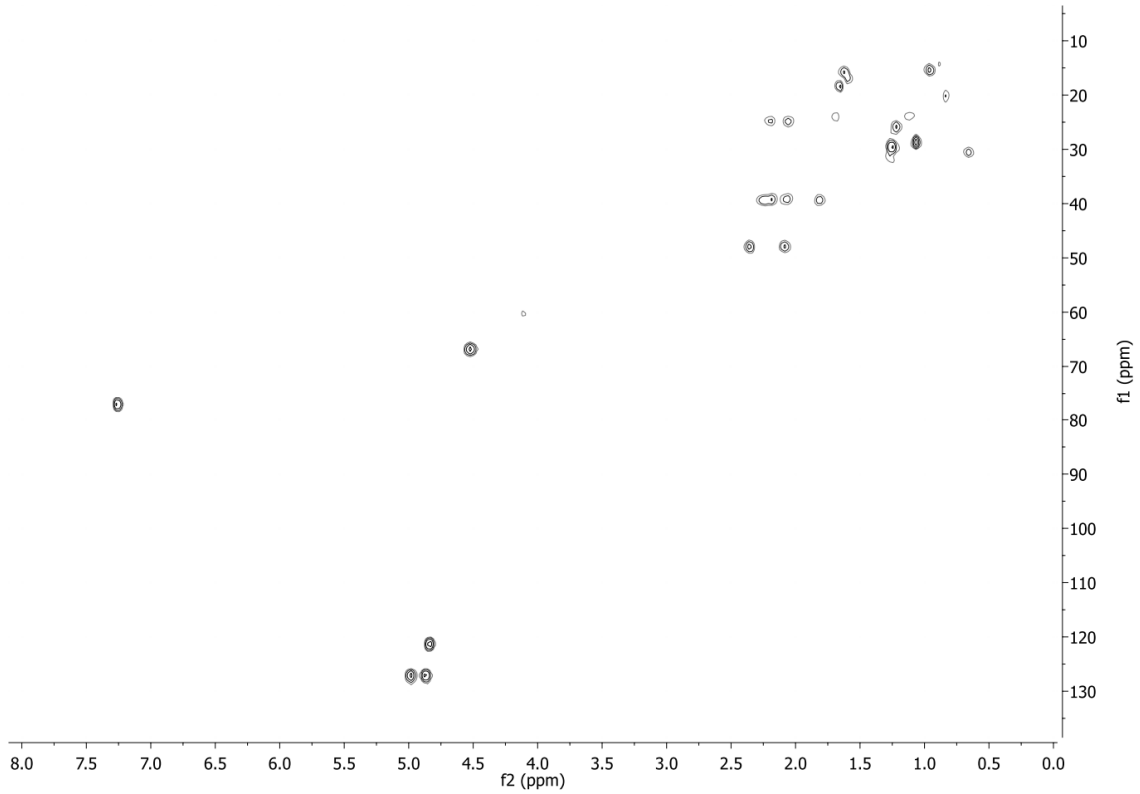
Appendix Figure 24: 1H spectrum of sinularcasbane D



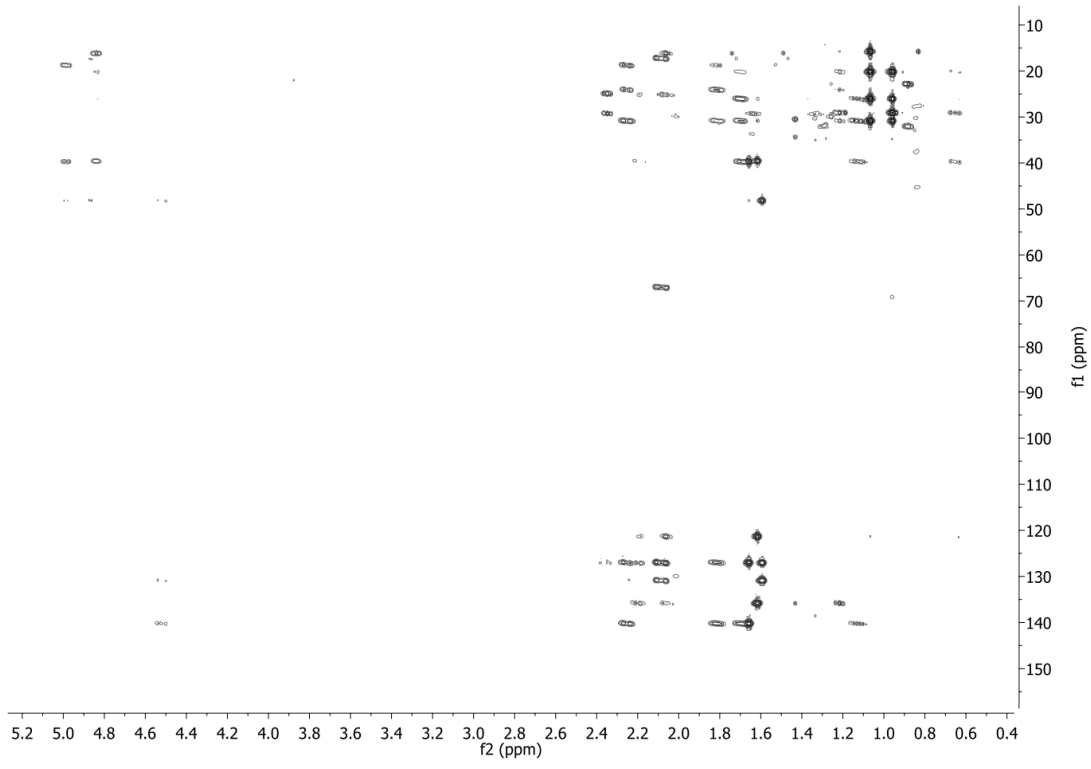
Appendix Figure 25: 13C spectrum of sinularcasbane D



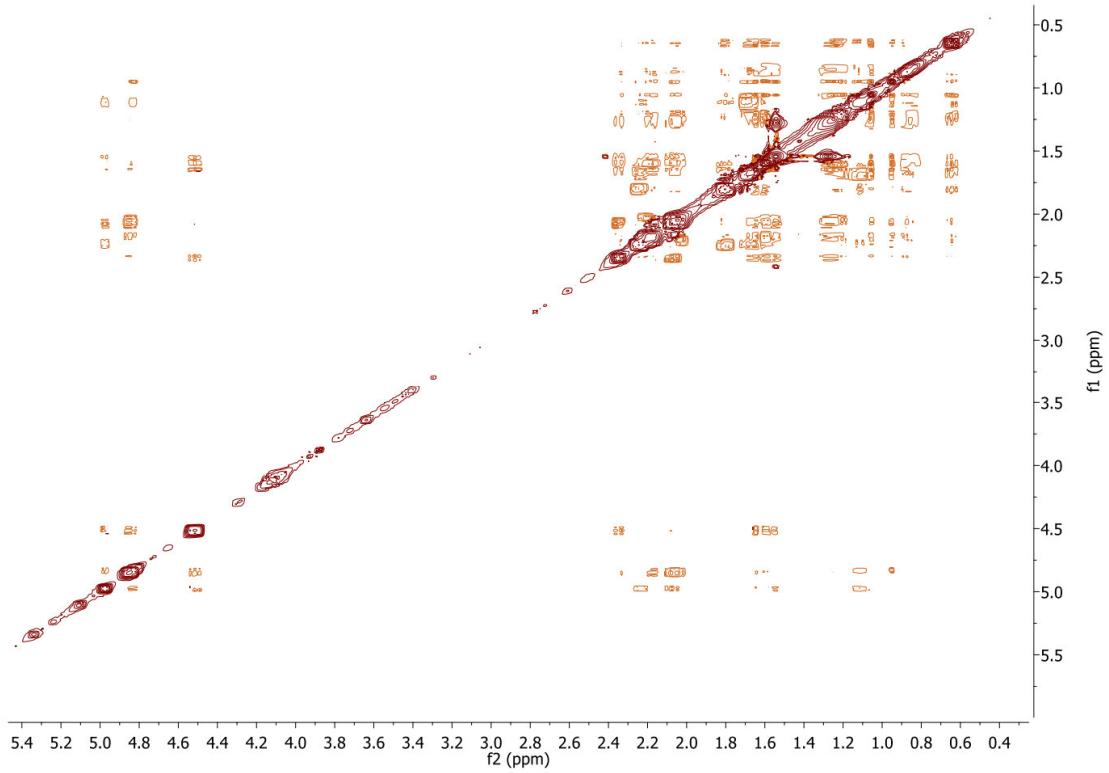
Appendix Figure 26: COSY spectrum of sinularcasbane D



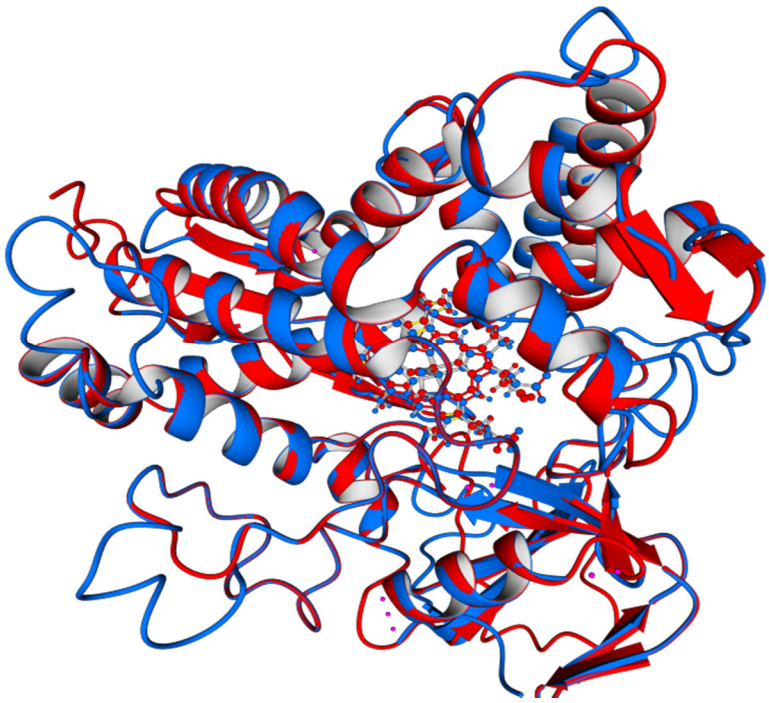
Appendix Figure 27: HSQC spectrum of sinularcasbane D



Appendix Figure 28: HMBC spectrum of sinularcasbane D

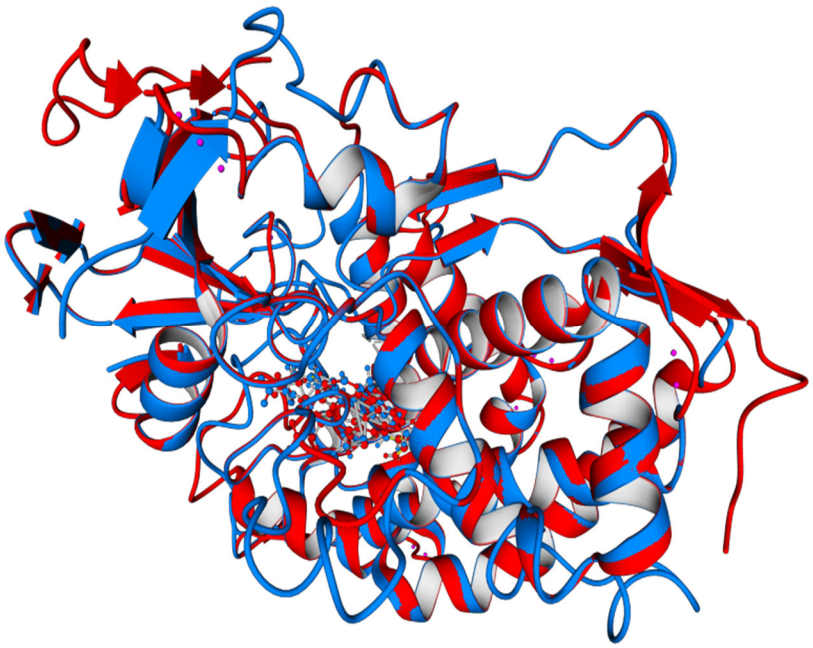


Appendix Figure 29: NOESY spectrum of sinularcasbane D



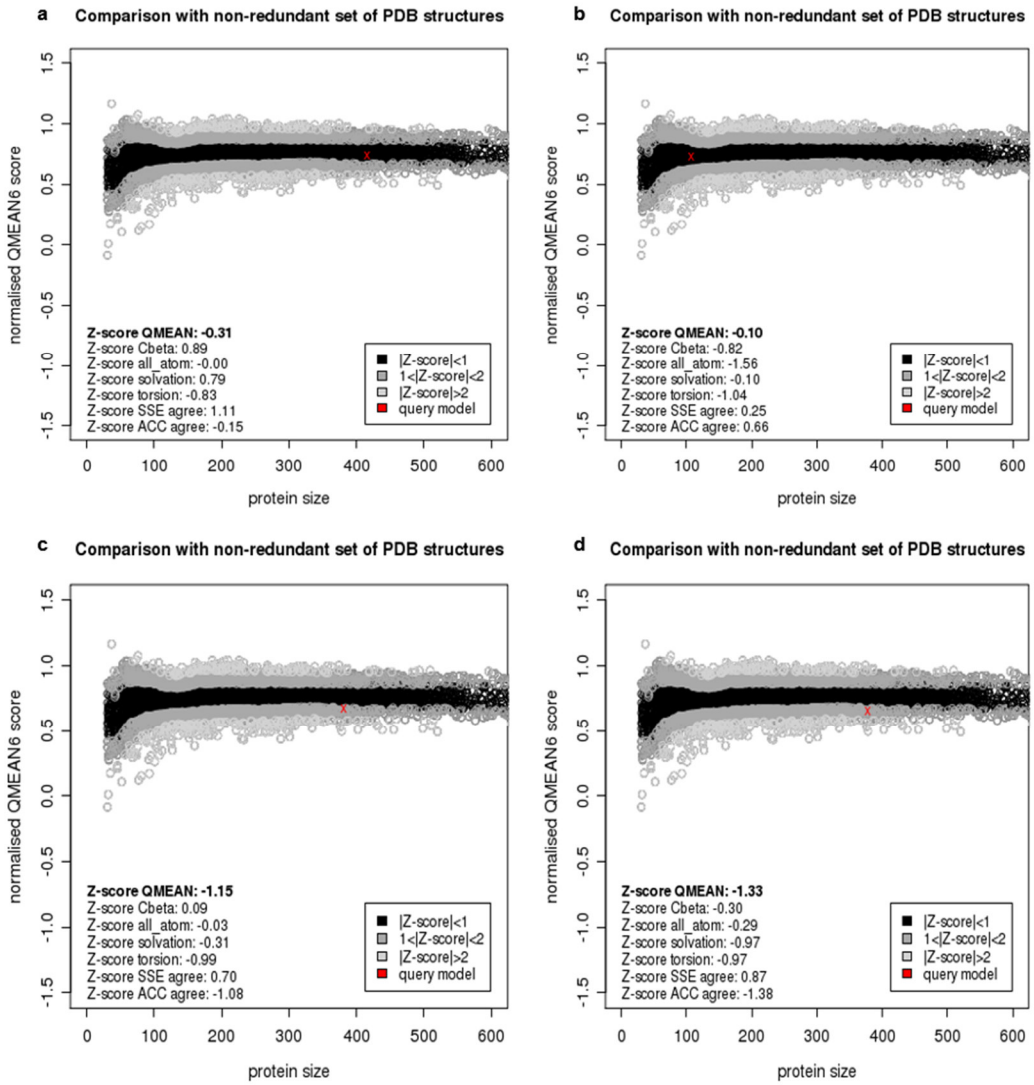
Appendix Figure 30: Structural alignment of CotB3 with P450cam

Appendix Figure 30: Structural alignment of CotB3 (cyan) and P450cam (PDB-ID: 4JX1, red). The alignment has an RMSD of 0.142Å over 313 aligned residues with 19% sequence identity. Note that the RMSD value considers only atoms that are present in both structures. Not matching loops in CotB3 are not incorporated into the alignment. The interaction region of CotB3 with Afx/Pdx is part of the matched helical structures.



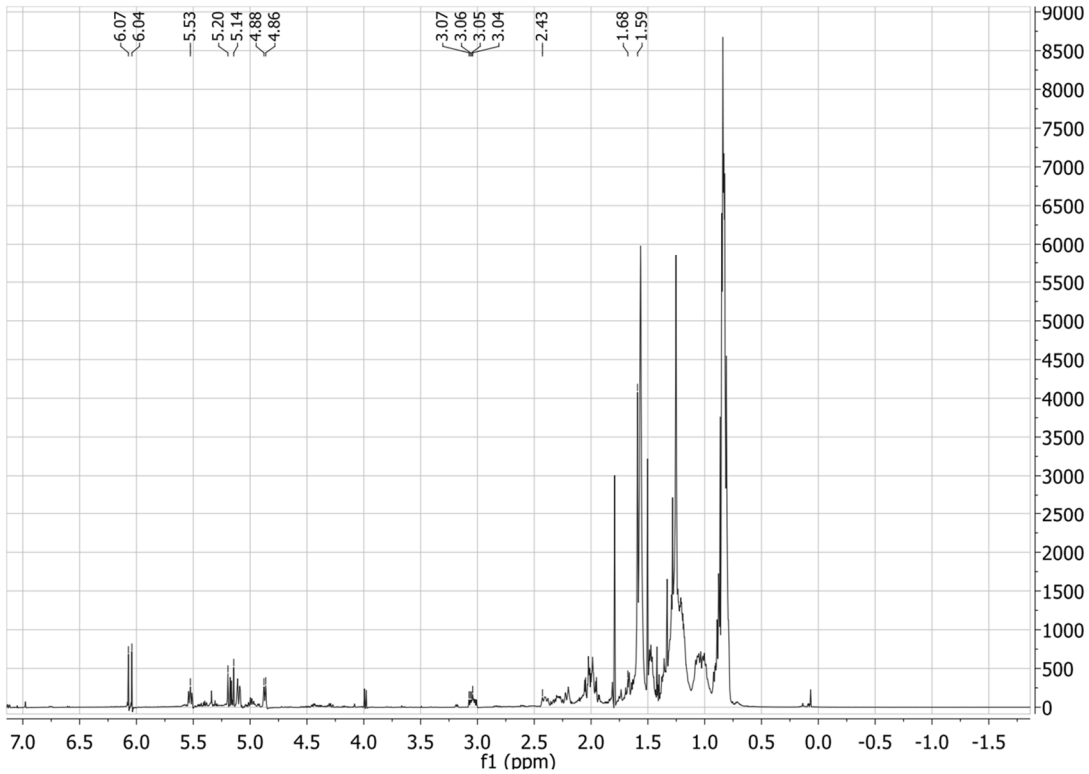
Appendix Figure 31: Structural alignment of CotB4 with P450cam

Appendix Figure 31: Structural alignment of CotB4 (cyan) and P450cam (PDB-ID: 4JX1, red). The alignment has an RMSD of 0.144Å over 339 aligned residues with 18% sequence identity. Note that the RMSD value considers only atoms that are present in both structures. Not matching loops in CotB4 are not incorporated into the alignment. The interaction region of CotB4 with Afx/Pdx is part of the matched helical structures.

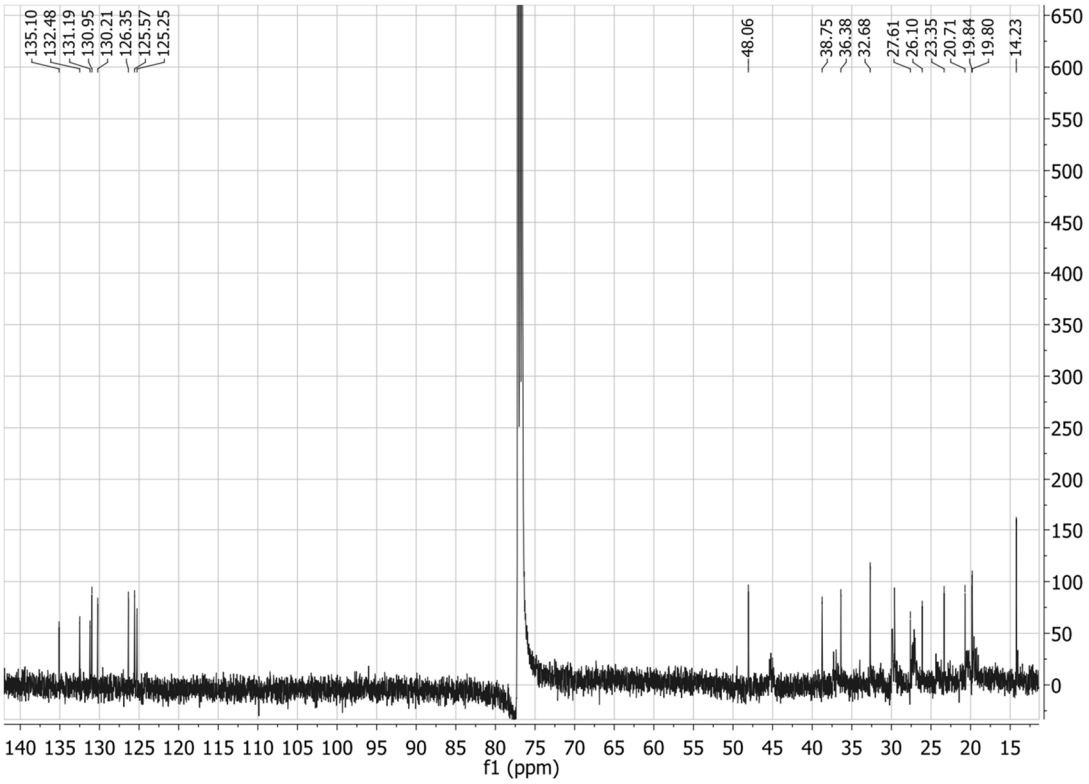


Appendix Figure 32: QMEANZ-score of the AfR, Afx, CotB3 and CotB4 models

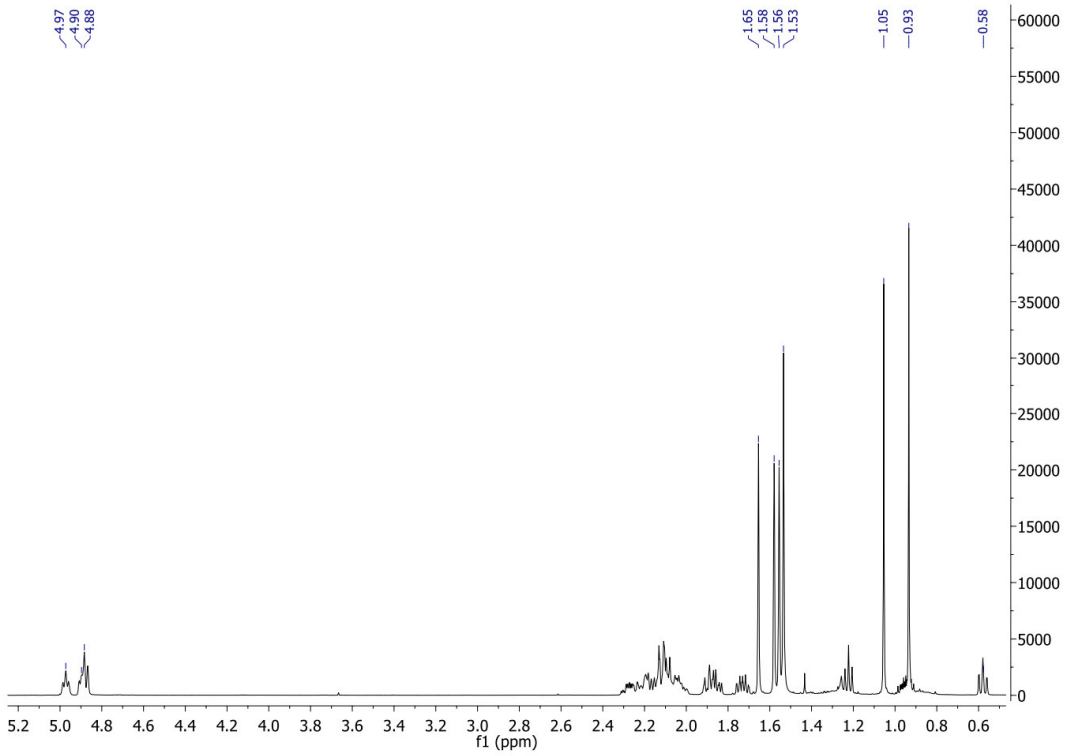
Figure 32**a**: QMEANZ-score of the AfR model (red cross) (-0-31), **b**: QMEANZ-score of the Afx model (-0.10), **c**: QMEANZ-score of the CotB3 model (-1.15) and **d**: QMEANZ-score of the CotB4 model (-1.33). For detailed information to QMEANZ-score quality assessment, see methods chapter III. The QMEAN Z-score of the aforementioned modeled proteins suggest that these models are of comparable quality to high resolution experimental structures. The reference structures are a non-redundant subset of the PDB sharing less than 30 % sequence identity among each other and are solved at a resolution below than 2 Å.



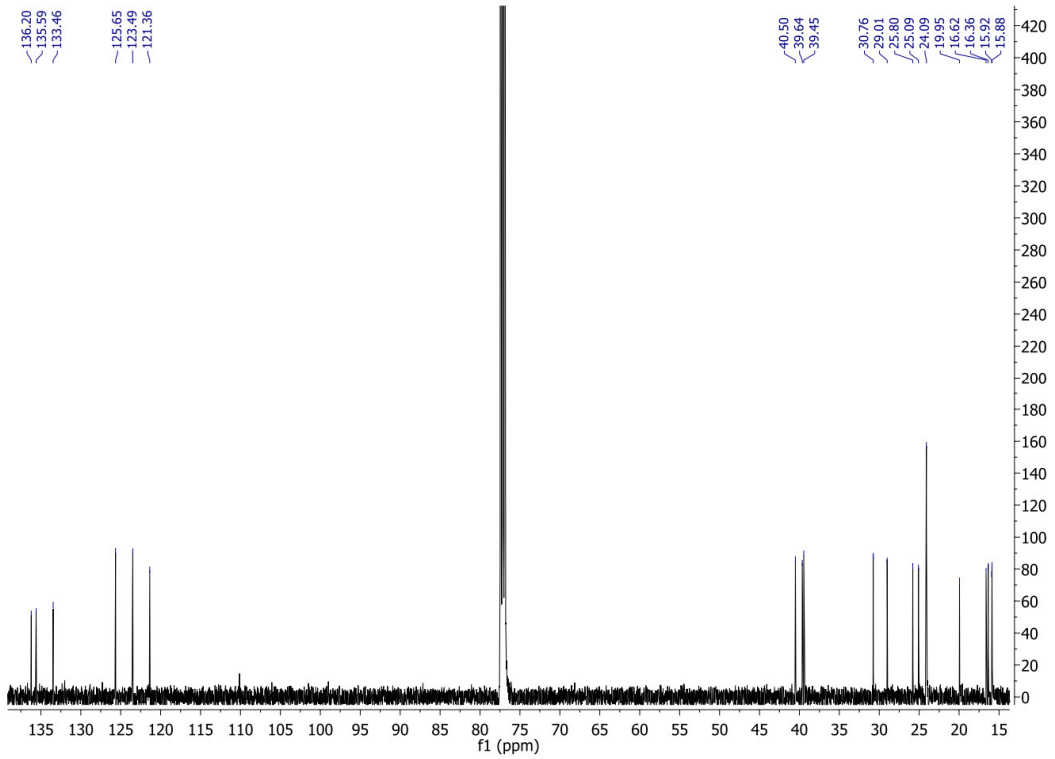
Appendix Figure 33: 1H spectrum of (+)-cembrene



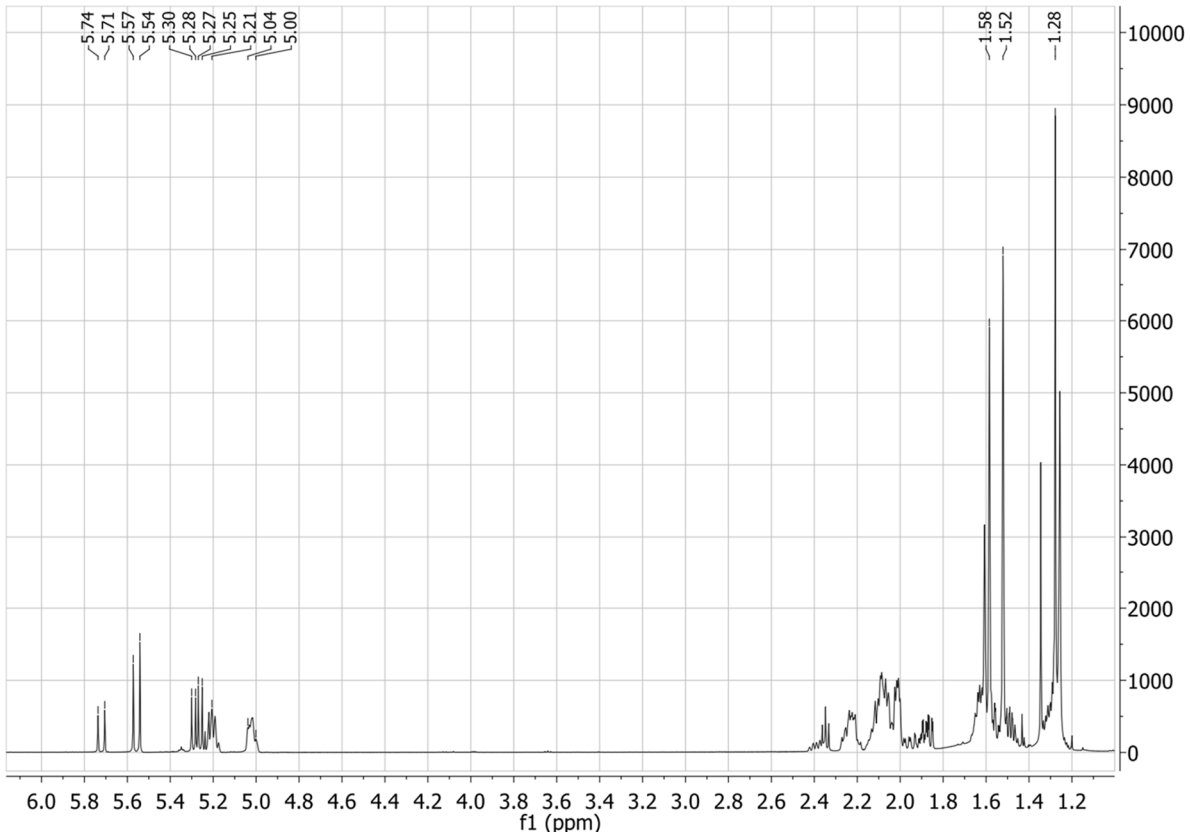
Appendix Figure 34: 13C spectrum of (+)-cembrene



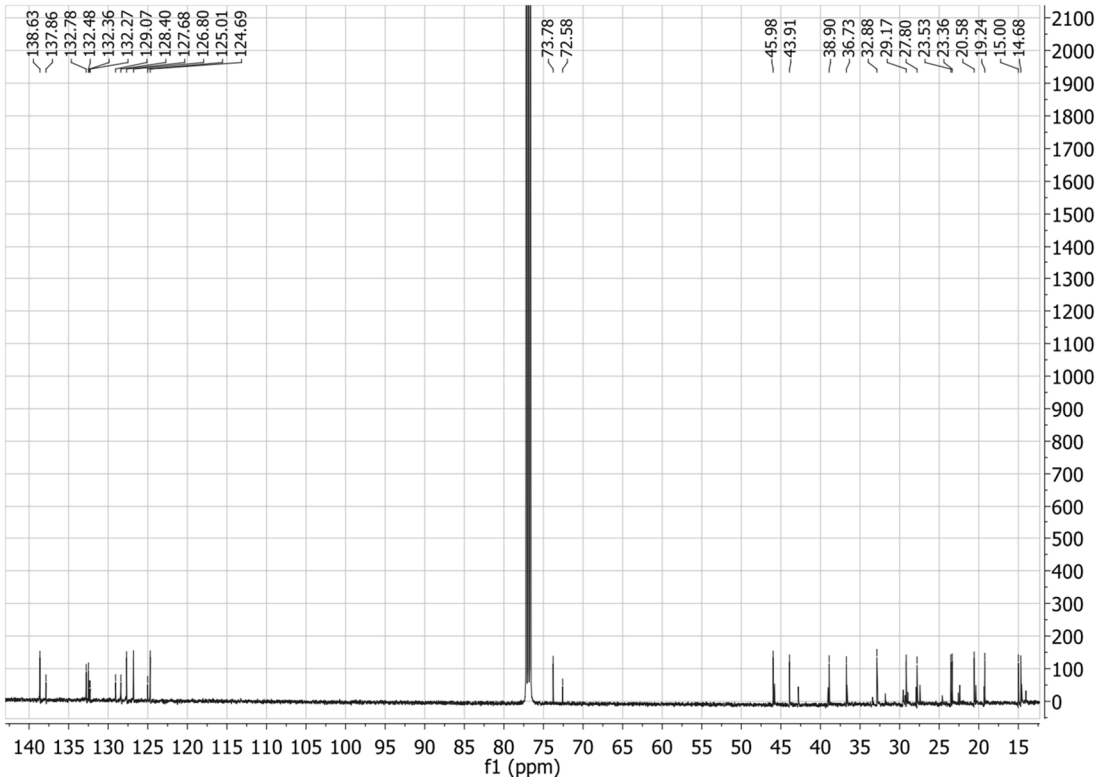
Appendix Figure 35: 1H spectrum of (-)-casbene



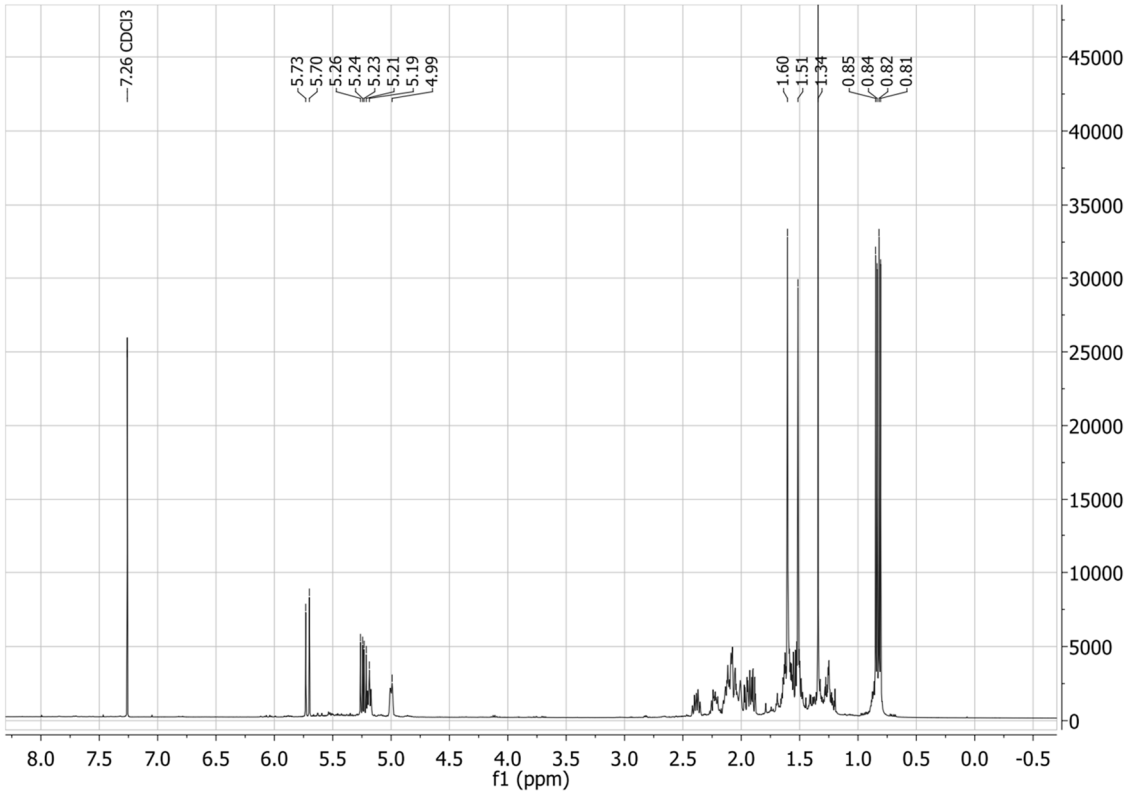
Appendix Figure 36: 13C spectrum of (-)-casbene



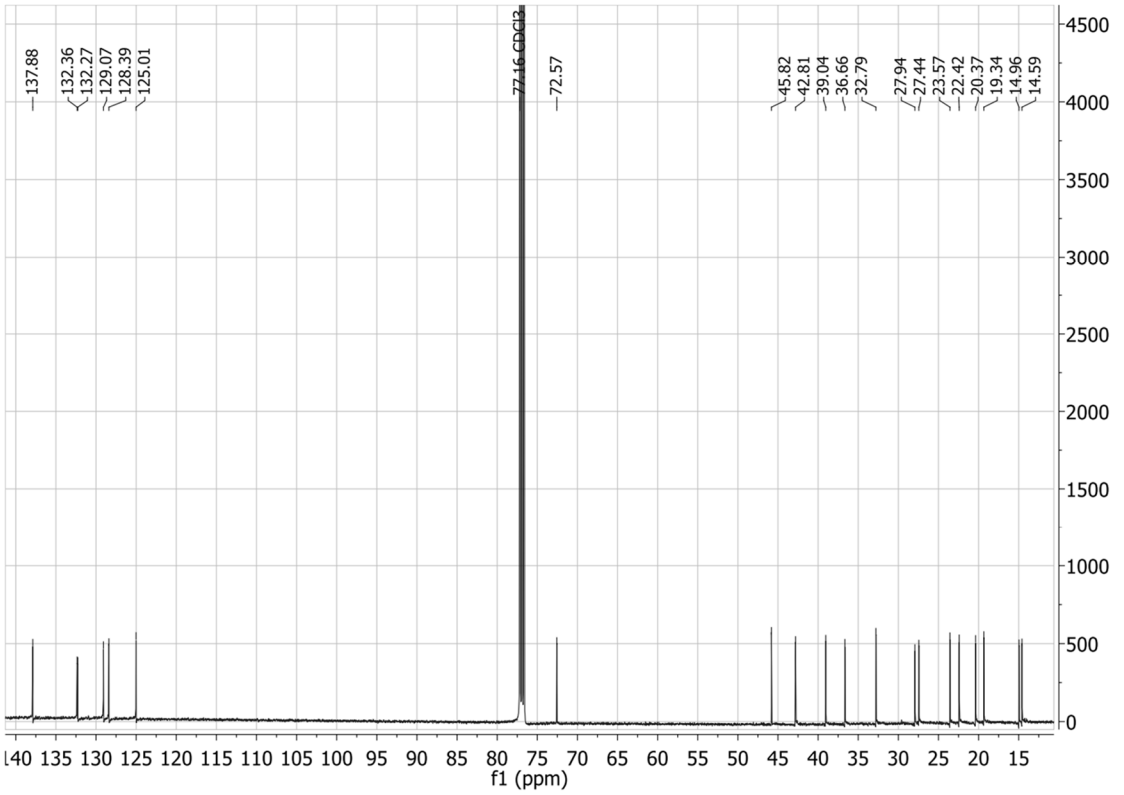
Appendix Figure 37: 1H spectrum of the combined purified α- and β-CBT-ol extract



Appendix Figure 38: **13C spectrum of the combined purified α- and β-CBT-ol extract**



Appendix Figure 39: 1H spectrum of the authentic α-CBT-ol standard

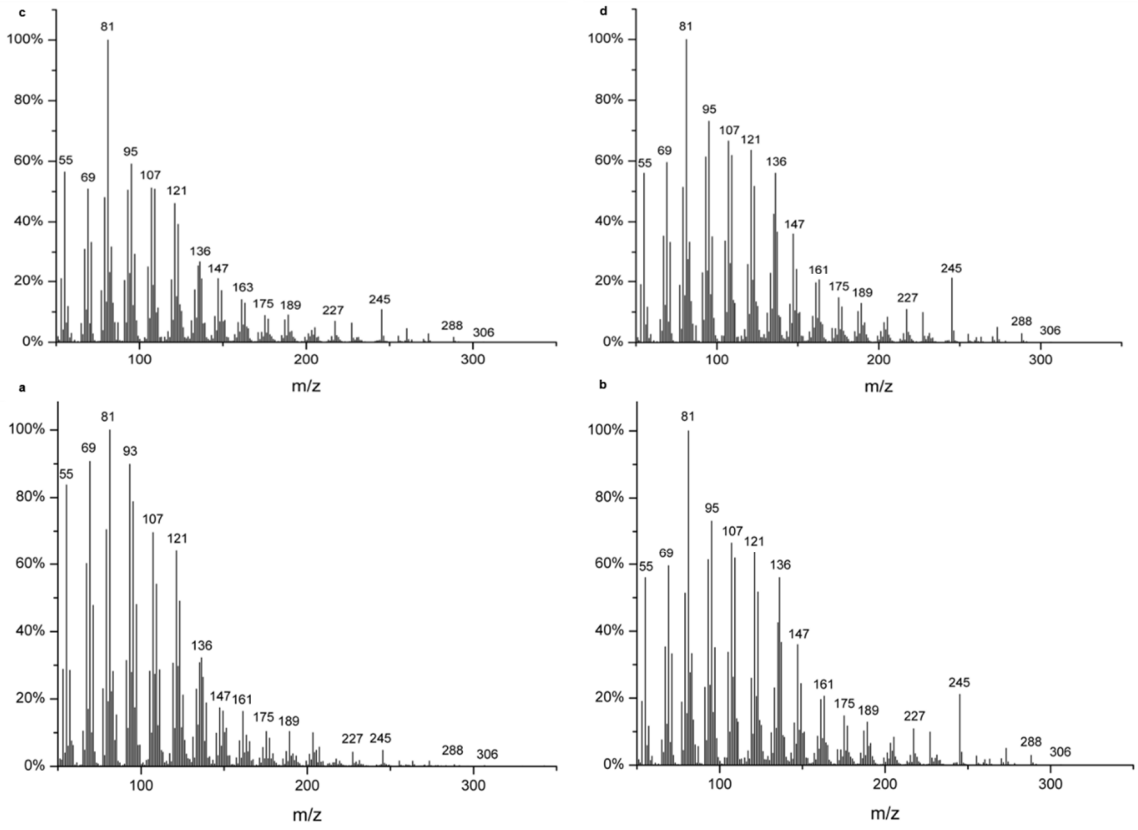


Appendix Figure 40: **13C spectrum of the authentic α-CBT-ol standard**

Target	Mutation	Activity [a]	a [b]	b	c	d
CBTS		100,0	4,79	2,86	0,67	91,70
F65	F65Y	0,0				
	F65A	0,0				
	F65E	0,0				
R316	R316L	0,0				
	R316H	0,0				
	R316Y	0,0				
W323	W323H[c]	0,0				
	W323L[c]	0,0				
	W323Y[c]	92,4	4,79	2,86	0,67	91,70
V344	V344M	0,0				
	V344K	0,0				
	V344L	0,0				
S452	S452A[c]	0,0				
	S452T	0,0				
C491	C491A[c]	5,2	4,50	2,45	N.D.	93,05
	C491L	91,6	4,75	2,56	0,61	92,08
R492	R492L	0,0				
	R492H	0,0				
	R492E	0,0				
D495	D495A	0,0				
	D495T	0,0				
	D495N	89,9	57,59	34,36	8,05	
R506	R506L	0,0				
	R506H	0,0				
	R506E	0,0				
L566	L566H	0,0				
	L566W	0,0				
F570	F570H[c]	0,0				
	F570D[c]	0,0				
	F570Y[c]	95,3	4,61	2,76	0,51	92,12
Y577	Y577A	0,0				
	Y577F	96,1	4,81	2,89	0,70	91,60

Appendix Table 3: Characterization of CBTS mutants

Appendix Table 3**a**: Enzyme activities of mutants (in %) in comparison to wild type CBTS (= 100%). The activities were determined by peak integration of the enzyme products in relation to the peak area of the internal standard alpha-humulene (methods section chapter II and IV). **b**: Compound numbers **a-d** refer to compound numbers in Figure 47 and 48. The tabulated values represent the product distributions in % for CBTS and each mutant (sum = 100%). (N.D.: Not detectable). **c**: The mutation and screening have been conducted during and included in the master thesis of Michael Metterlein.[152]



Appendix Figure 41: EI-MS spectra of CBT-diols

Appendix Figure 41**a**: EI-MS spectrum of **α-CBT-diol** purified from cell extract, **b**: spectrum of **β-CBT-diol** purified from cell extract, **c**: spectrum of **α-CBT-diol** (authentic standard) and **d**: spectrum of **β-CBT-diol** (authentic standard).

5. Primer used for the mutation of TXS and CBTS

TXS Mutant	Sequence 5`-> 3`
Fw_R580H	CTTTACCCGTCATCATGTTGCCGAAGTG
Rv_R580H	CACTTCGGCAACATGATGACGGGTAAAG
Fw_R580A	GATATTAACTTTACCCGTCATGCCGTTGCCGAAGTGTATTTTAG
Rv_R580A	CTAAAATACACTTCGGCAACGGCATGACGGGTAAAGTTAATATC
Fw_R580E	GATATTAACTTTACCCGTCATGAGGTTGCCGAAGTGTATTTTAGC
Rv_R580E	GCTAAAATACACTTCGGCAACCTCATGACGGGTAAAGTTAATATC
Fw_R754H	GTAAGCCTGAGCTGGCATCTGACCAATGATAC
Rv_R754H	GTATCATTGGTCAGATGCCAGCTCAGGCTTAC
Fw_R754A	GGTAAGCCTGAGCTGGGCGCTGACCAATGATACC
Rv_R754A	GGTATCATTGGTCAGCGCCCAGCTCAGGCTTACC
Fw_R754E	GAGCTTGTATCCTTGAGCTGGGAACTAACAAACGACACCAAAAC
Rv_R754E	GTTTTGGTGTCGTTTGTTAGTTCCCAGCTCAAGGATACAAGCTC
Fw_R768H	CAGGCTGAAAAGGCTCATGGACAACAAGCCTCAG
Rv_R768H	CTGAGGCTTGTTGTCCATGAGCCTTTTCAGCCTG
Fw_R768A	CAGGCTGAAAAGGCTGCGGGACAACAAGCCTC
Rv_R768A	GAGGCTTGTTGTCCCGCAGCCTTTTCAGCCTG
Fw_R768E	CAGGCTGAAAAGGCTGAAGGACAACAAGCCTC
Rv_R768E	GAGGCTTGTTGTCCTTCAGCCTTTTCAGCCTG
Fw_V584M	GACACCGAGTGGCGGAGATGTATTTTTCATCAGCTAC
Rv_V584M	GTAGCTGATGAAAAATACATCTCCGCCACTCGGTGTC
Fw_V584K	CGACACCGAGTGGCGGAGAAATATTTTTCATCAGCTAC
Rv_V584K	GTAGCTGATGAAAAATATTTCTCCGCCACTCGGTGTCG
Fw_V584L	CATCGTGTTGCCGAACTGTATTTTAGCAGTG
Rv_V584L	CACTGCTAAAATACAGTTCGGCAACACGATG
Fw_V584N	GACACCGAGTGGCGGAGAATTATTTTTCATCAGCTAC
Rv_V584N	GTAGCTGATGAAAAATAATTCTCCGCCACTCGGTGTC
Fw_V584S	GACACCGAGTGGCGGAGAGCTATTTTTCATCAGCTAC
Rv_V584S	GTAGCTGATGAAAAATAGCTCTCCGCCACTCGGTGTC
Fw_V584P	GACACCGAGTGGCGGAGCCGTATTTTTCATCAGCTAC
Rv_V584P	GTAGCTGATGAAAAATACGGCTCCGCCACTCGGTGTC
Fw_V584H	CGTCATCGTGTTGCCGAACACTATTTTAGCAGTGCAAC
Rv_V584H	GTTGCACTGCTAAAATAGTGTTCGGCAACACGATGACG
Fw_V584R	GTCATCGTGTTGCCGAAAGGTATTTTAGCAGTGCAAC
Rv_V584R	GTTGCACTGCTAAAATACCTTTCGGCAACACGATGAC
Fw_S587D	GTGGCGGAGGTTTATTTTGATTCAGCTACATTTGAACCCG
Rv_S587D	CGGGTTCAAATGTAGCTGAATCAAAATAAACCTCCGCCAC
Fw_S587Y	GGCGGAGGTTTATTTTTATTCAGCTACATTTGAACCC
Rv_S587Y	GGGTTCAAATGTAGCTGAATAAAAATAAACCTCCGCC
Fw_S587K	GTGGCGGAGGTTTATTTTAAATCAGCTACATTTGAACCC

Rv_S587K	GGGTTCAAATGTAGCTGATTTAAAATAAACCTCCGCCAC
Fw_S587L	GTGGCGGAGGTTTATTTTCTGTCAGCTACATTTGAACCC
Rv_S587L	GGGTTCAAATGTAGCTGACAGAAAATAAACCTCCGCCAC
Fw_S587A	GTTGCCGAAGTGTATTTTGCCAGTGCAACCTTTGAACCG
Rv_S587A	CGGTTCAAAGGTTGCACTGGCAAAATACACTTCGGCAAC
Fw_S587G	GTTGCCGAAGTGTATTTTGGCAGTGCAACCTTTGAAC
Rv_S587G	GTTCAAAGGTTGCACTGCCAAAATACACTTCGGCAAC
Fw_F602W	CTGCAGGTCCTGTGGGATGATATGGCC
Rv_F602W	GGCCATATCATCCCACAGGACCTGCAG
Fw_F602I	GTCTGCAGGTCCTGATCGATGATATGGCC
Rv_F602I	GGCCATATCATCGATCAGGACCTGCAGAC
Fw_F602A	GTCTGCAGGTCCTGGCAGATGATATGGCCG
Rv_F602A	CGGCCATATCATCTGCCAGGACCTGCAGAC
Fw_V610H	CAAAATTGGTTGTCTGCAGCACCTGTTCGATGATATGGC
Rv_V610H	GCCATATCATCGAACAGGTGCTGCAGACAACCAATTTTG
Fw_V610S	CAAAATTGGTTGTCTGCAGAGCCTGTTCGATGATATGGC
Rv_V610S	GCCATATCATCGAACAGGCTCTGCAGACAACCAATTTTG
Fw_V610F	CAAAATTGGTTGTCTGCAGTTCCTGTTCGATGATATGG
Rv_V610F	CCATATCATCGAACAGGAACTGCAGACAACCAATTTTG
Fw_V610A	CAAAATTGGTTGTCTGCAGGCCCTGTTCGATGATATGGCC
Rv_V610A	GGCCATATCATATCATCGAACAGGGCCAGCAGACAACCAATTTTG
Fw_V714A	CTTATGCTATATCAGCGGGCCTTGGACCGTG
Rv_V714A	CACGGTCCAAGGCCCGCTGATATAGCATAAG
Fw_V714I	GACTTATGCTATATCAATTGGCCTTGGACCGTGTAC
Rv_V714I	GTACACGGTCCAAGGCCAATTGATATAGCATAAGTC
Fw_V714T	GAAAACCTATGCAATTAGCACCGGTCTGGGTCCGTGTACC
Rv_V714T	GGTACACGGACCCAGACCGGTGCTAATTGCATAGGTTTTC
Fw_V714G	GAAAACCTATGCAATTAGCGGCGGTCTGGGTCCGTGTACCC
Rv_V714G	GGGTACACGGACCCAGACCGCCGCTAATTGCATAGGTTTTC
Fw_V714P	GACTTATGCTATATCACCGGGCCTTGGACCGTGTAC
Rv_V714P	GTACACGGTCCAAGGCCCGGTGATATAGCATAAGTC
Fw_G715A	CTTATGCTATATCAGTAGCGCTTGGACCGTGTACCCTAC
Rv_G715A	GTAGGGTACACGGTCCAAGCGCTACTGATATAGCATAAG
Fw_G715S	CTTATGCTATATCAGTAAGCCTTGGACCGTGTAC
Rv_G715S	GTACACGGTCCAAGGCTTACTGATATAGCATAAG
Fw_S713T	CTTAAAGACTTATGCTATAACAGTAGGCCTTGGACCGTG
Rv_S713T	CACGGTCCAAGGCCTACTGTTATAGCATAAGTCTTTAAG
Fw_S713A	CTTAAAGACTTATGCTATAGCGGTAGGCCTTGGACCGTGTAC
Rv_S713A	GTACACGGTCCAAGGCCTACCGCTATAGCATAAGTCTTTAAG
Fw_S713L	CTTAAAGACTTATGCTATACTGGTAGGCCTTGGACCGTGTAC
Rv_S713L	GTACACGGTCCAAGGCCTACCAGTATAGCATAAGTCTTTAAG
Fw_W753H	GAACTGGTAAGCCTGAGCCACCGTCTGACCAATGATACC
Rv_W753H	GGTATCATTGGTCAGACGGTGGCTCAGGCTTACCAGTTC
Fw_W753L	GGTAAGCCTGAGCTTGCGTCTGACCAATG
Rv_W753L	CATTGGTCAGACGCAAGCTCAGGCTTACC
Fw_W753V	CTGGTAAGCCTGAGCGTGCGTCTGACCAATG

Rv_W753V	CATTGGTCAGACGCACGCTCAGGCTTACCAG
Fw_W753E	GAGCTTGTATCCTTGAGCGAACGACTAACAAACGACAC
Rv_W753E	GTGTCGTTTGTTAGTCGTTCGCTCAAGGATACAAGCTC
Fw_W753C	GTATCCTTGAGCTGCCGACTAACAAACGAC
Rv_W753C	GTCGTTTGTTAGTCGGCAGCTCAAGGATAC
Fw_W753A	GAACTGGTAAGCCTGAGCGCACGTCTGACCAATGATAC
Rv_W753A	GTATCATTGGTCAGACGTGCGCTCAGGCTTACCAGTTC
Fw_C830A	CCTTTATCTTTAATCTGCGTCTGGCCGTGCAGATCTTCTATAAATTC
Rv_C830A	GAATTTATAGAAGATCTGCACGGCCAGACGCAGATTAAAGATAAAGG
Fw_C830S	CTTTATCTTTAATCTGCGTCTGAGCGTGCAGATCTTCTATAAATTC
Rv_C830S	GAATTTATAGAAGATCTGCACGCTCAGACGCAGATTAAAGATAAAG
Fw_F834A	CTTAGATTGTGTGTCCAAATCGCCTACAAGTTTATAGATGGGTAC
Rv_F834A	GTACCCATCTATAAACTTGTAGGCGATTTGGACACACAATCTAAG
Fw_F834G	CTTAGATTGTGTGTCCAAATCGGCTACAAGTTTATAGATGGGTAC
Rv_F834G	GTACCCATCTATAAACTTGTAGCCGATTTGGACACACAATCTAAG
Fw_F834Y	GATTGTGTGTCCAAATCTACTACAAGTTTATAGATGG
Rv_F834Y	CCATCTATAAACTTGTAGTAGATTTGGACACACAATC
Fw_F834H	CGTCTGTGCGTGCAGATCCACTATAAATTCATTGATG
Rv_F834H	CATCAATGAATTTATAGTGGATCTGCACGCACAGACG
Fw_Y835F	GTGTGTCCAAATCTTTTTTAAGTTTATAGATGGGTAC
Rv_Y835F	GTACCCATCTATAAACTTAAAAAAGATTTGGACACAC
Fw_Y835W	GTGTGTCCAAATCTTTTGGAAGTTTATAGATGGGTAC
Rv_Y835W	GTACCCATCTATAAACTTCCAAAAGATTTGGACACAC
Fw_Y835A	GATTGTGTGTCCAAATCTTTGCGAAGTTTATAGATGGGTACG
Rv_Y835A	CGTACCCATCTATAAACTTCGCAAAGATTTGGACACACAATC
Fw_Y841A	CTATAAATTCATTGATGGTGCCGGCATTGCCAACGAAGAG
Rv_Y841A	CTCTTCGTTGGCAATGCCGGCACCATCAATTTATAG
Fw_Y841F	CTATAAATTCATTGATGGTTTCGGCATTGCCAACGAAGAG
Rv_Y841F	CTCTTCGTTGGCAATGCCGAAACCATCAATGAATTTATAG
Fw_Y841T	CTATAAATTCATTGATGGTACCGGCATTGCCAACGAAGAG
Rv_Y841T	CTCTTCGTTGGCAATGCCGGTACCATCAATGAATTTATAG
Fw_Y89F	GTCTGAGCGCAAATTTTCATGGTGATCTGTG
Rv_Y89F	CACAGATCACCATGAAAATTTGCGCTCAGAC
Fw_Y89A	ACTCTCCGCCAATGCGCATGGCGATCTGTG
Rv_Y89A	CACAGATCGCCATGCGCATTGGCGGAGAGT
Fw_Y89E	CGACTCTCCGCCAATGAGCATGGCGATCTGTG
Rv_Y89E	(CACAGATCGCCATGCTCATTGGCGGAGAGTCG

Appendix Table 4: Primer used for mutations of TXS

Appendix Table 4: Primer used for mutations of TXS (Fw: Forward primer, Rv: Reverse primer).

CBTS Mutant	Sequence 5`-> 3`
Fw_F65Y	GTCATCTGGCAGATTATCCGAGCACCATTTG
Rv_65Y	CAAATGGTGCTCGGATAATCTGCCAGATGAC
Fw_F65A	CGTCATCTGGCAGATGCGCCGAGCACCATTTG
Rv_F65A	CAAATGGTGCTCGGCGCATCTGCCAGATGACG
Fw_F65E	CCGTCATCTGGCAGATGAACCGAGCACCATTTGG
Rv_F65E	CCAAATGGTGCTCGGTTCATCTGCCAGATGACGG
Fw_R316L	CGTATGCAAAAGATCTGCTGGTGGAAGCATA
Rv_R316L	TATGCTTCCACCAGCAGATCTTTTGCATACG
Fw_R316H	CCGTATGCAAAAGATCATCTGGTGGAAGCATA
Rv_R316H	TATGCTTCCACCAGATGATCTTTTGCATACGG
Fw_R316Y	CCGTATGCAAAAGATTATCTGGTGGAAGCAT
Rv_R316Y	ATGCTTCCACCAGATAATCTTTTGCATACGG
Fw_W323H	CTGGTGGAAGCATATTTTCATACCGTTGGTATCTATTT
Rv_W323H	AAATAGATACCAACGGTATGAAAATATGCTTCCACCAG
Fw_W323L	GGAAGCATATTTTTTGACCGTTGGTATCT
Rv_W323L	AGATACCAACGGTCAAAAAATATGCTTCC
Fw_W323Y	GGTGGAAGCATATTTTTATACCGTTGGTATCTATT
Rv_W323Y	AATAGATACCAACGGTATAAAAATATGCTTCCACC
Fw_V344M	CTGGTTACCAAAGTTATGAAAATGAACAGC
Rv_V344M	GCTGTTCATTTTCATAACTTTGGTAACCAG
Fw_V344K	GTAGTCTGGTTACCAAAGTTAAAAAAATGAACAGCATTATCG
Rv_V344K	CGATAATGCTGTTCATTTTTTTAACTTTGGTAACCAGACTAC
Fw_V344L	GTAGTCTGGTTACCAAAGTTCTGAAAATGAACAGCATTAT
Rv_V344L	ATAATGCTGTTCATTTTCAGAACTTTGGTAACCAGACTAC
Fw_S452A	AATGGTCTGGTTAGCGCGACCGGTCCGATGTATG
Rv_S452A	CATACATCGGACCGGTCGCGCTAACCAGACCATT
Fw_S452T	TGGTCTGGTTAGCACCACCGGTCCGATGT
Rv_S452T	ACATCGGACCGGTGGTGCTAACCAGACCA
Fw_C491A	GTGCAGCAAGTACCATTGCGCGTCTGATGGATGATATG
Rv_C491A	CATATCATCCATCAGACGCGCAATGGTACTTGCTGCAC
Fw_C491L	GTGCAGCAAGTACCATTCTGCGTCTGATGGATGATATG
Rv_C491L	CATATCATCCATCAGACGCAGAATGGTACTTGCTGCAC
Fw_R492L	GCAAGTACCATTTGTCTGCTGATGGATGATATGGC
Rv_R492L	GCCATATCATCCATCAGCAGACAAATGGTACTTGC
Fw_R492H	GCAAGTACCATTTGTCATCTGATGGATGATATGG
Rv_R492H	CCATATCATCCATCAGATGACAAATGGTACTTGC
Fw_R492E	GCAGCAAGTACCATTTGTGAACTGATGGATGATATGGCAG
Rv_R492E	CTGCCATATCATCCATCAGTTCACAAATGGTACTTGCTGC
Fw_D495A	CATTTGTCGTCTGATGGCGGATATGGCAGATCATG
Rv_D495A	CATGATCTGCCATATCCGCCATCAGACGACAAATG
Fw_D495T	GTACCATTTGTCGTCTGATGACCGATATGGCAGATCATGAAG

Rv_D495T	CTTCATGATCTGCCATATCGGTCATCAGACGACAAATGGTAC
Fw_D495N	CCATTTGTCGTCTGATGAATGATATGGCAGATCATG
Rv_D495N	CATGATCTGCCATATCATTCATCAGACGACAAATGG
Fw_R506L	GAAGTTGAACAGCAGCTGGGTCATGTTGCGAGC
Rv_R506L	GCTCGCAACATGACCCAGCTGCTGTTCAACTTC
Fw_R506H	GAAGTTGAACAGCAGCATGGTCATGTTGCGAG
Rv_R506H	CTCGCAACATGACCATGCTGCTGTTCAACTTC
Fw_R506E	CATGAAGTTGAACAGCAGGAAGGTCATGTTGCGAGCTTTG
Rv_R506E	CAAAGCTCGCAACATGACCTTCCTGCTGTTCAACTTCATG
Fw_L566H	GTCTGAATTTTAGCCGTCATGCCGATACCTTTCTGAAAG
Rv_L566H	CTTTCAGAAAGGTATCGGCATGACGGCTAAAATTCAGAC
Fw_L566W	GTCTGAATTTTAGCCGTTGGGCCGATACCTTTCTG
RV_L566W	CAGAAAGGTATCGGCCCAACGGCTAAAATTCAGAC
Fw_F570H	CGTCTGGCCGATACCCATCTGAAAGATGACG
Rv_F570H	CGTCATCTTTCAGATGGGTATCGGCCAGACG
Fw_F570D	CGTCTGGCCGATACCGATCTGAAAGATGACG
Rv_F570D	CGTCATCTTTCAGATCGGTATCGGCCAGACG
Fw_F570Y	GTCTGGCCGATACCTATCTGAAAGATGACG
Rv_F570Y	CGTCATCTTTCAGATAGGTATCGGCCAGAC
Fw_Y577A	CTGAAAGATGACGATGGTGCGACCAACCCGAAAAGCAAAG
Rv_Y577A	CTTTGCTTTTCGGGTTGGTCGCACCATCGTCATCTTTCAG
Fw_Y577F	GAAAGATGACGATGGTTTTACCAACCCGAAAAGC
Rv_Y577F	GCTTTTCGGGTTGGTAAAACCATCGTCATCTTTC

Appendix Table 5: Primer used for mutation of CBTS

Appendix Table 5: Primer used for mutations of CBTS (Fw: Forward primer, Rv: Reverse primer).

6. Genes

5`-
CATATGGCAATGAGCAGCAGCACCGGCACCAGCAAAGTTGTTAGCGAAACCAGCAGTACCATTG
TTGATGATATTCCGCGTCTGAGCGCAAATTATCATGGTGATCTGTGGCATCATAATGTGATTCAGA
CCCTGGAAACCCCGTTTCGTGAAAGCAGCACCTATCAAGAACGTGCAGATGAACTGGTTGTGAAA
ATCAAAGATATGTTTAACGCACTGGGTGATGGTGATATTAGCCCGAGCGCCTATGATACCGCATG
GGTTGCACGTCTGGCAACCATTAGCAGTGATGGTAGCGAAAAACCGCGTTTTCCGCAGGCACTG
AATTGGGTTTTTAACAATCAGCTGCAGGATGGTAGTTGGGGTATTGAAAGCCATTTTAGCCTGTGT
GATCGTCTGCTGAATACCACCAATAGCGTTATTGCACTGAGCGTTTGGAAAACCGGTCATAGCCA
GGTTCAGCAGGGTGCAGAATTTATTGCAGAAAATCTGCGCCTGCTGAATGAAGAAGATGAGCTGA
GTCCGGATTTTCAGATTATCTTTCCGGCACTGCTGCAGAAAGCAAAAGCACTGGGTATTAATCTG
CCGTATGATCTGCCGTTTATCAAATATCTGAGCACCACCCGTGAAGCACGTCTGACCGATGTTAG
CGCAGCAGCAGATAATATTCCGGCAAATATGCTGAATGCACTGGAAGGTCTGGAAGAAGTTATTG
ACTGGAACAAAATTATGCGCTTCCAGAGCAAAGATGGTAGCTTTCTGAGTAGTCCGGCAAGCACC
GCATGTGTTCTGATGAATACCGGTGATGAAAAATGCTTTACCTTCCTGAATAACCTGCTGGATAAA
TTTGGTGGTTGTGTTCCGTGTATGTATAGCATTGATCTGCTGGAACGTCTGAGCCTGGTTGATAAT
ATTGAACATCTGGGTATTGGTCGCCACTTCAAACAAGAAATTAAAGGTGCACTGGATTACGTGTAT
CGTCATTGGAGCGAACGTGGTATTGGTTGGGGTCGTGATAGCCTGGTTCCGGATCTGAATACAA
CCGCACTGGGCCTGCGTACCCTGCGTATGCATGGTTATAATGTTAGCTCAGATGTGCTGAACAAC
TTTAAAGATGAAAACGGTCGCTTTTTTAGCAGCGCAGGTCAGACCCATGTTGAACTGCGTAGCGT
TGTTAACCTGTTTCGTGCAAGCGATCTGGCATTTCCGGATGAACGTGCAATGGATGATGCACGTA
AATTTGCAGAACCGTATCTGCGTGAAGCCCTGGCCACCAAAATTAGCACCAATACAAAACTGTTTA
AAGAAATCGAATATGTGGTCGAGTATCCGTGGCACATGAGCATTCCTCGTCTGGAAGCACGTAGC
TATATTGATAGCTATGATGATAACTATGTGTGGCAGCGTAAAACCCTGTATCGTATGCCGAGCCTG
AGCAATAGCAAATGTCTGGAACTGGCAAAACTGGATTTTAACATTGTTCAGAGCCTGCACCAAGA
AGAACTGAAACTGCTGACCCGTTGGTGGAAAGAAAGCGGTATGGCAGATATTAACTTTACCCGTC
ATCGTGTTGCCGAAGTGTATTTTAGCAGTGCAACCTTTGAACCGGAATATAGCGCAACCCGTATT
GCCTTTACCAAAATTGGTTGTCTGCAGGTCCTGTTCGATGATATGGCCGATATTTTTGCAACCCTG
GATGAACTGAAAAGTTTTACCGAAGGTGTTAAACGTTGGGATACCAGTCTGCTGCATGAAATCCC
GGAATGTATGCAGACCTGTTTTAAAGTGTGGTTTAAACTGATGGAAGAGGTGAATAACGATGTGG
TTAAAGTTCAGGGTCGCGATATGCTGGCCCATATTCGTAAACCGTGGGAACTGTATTTCAACTGC
TATGTTCAAGAACGCGAATGGCTGGAAGCCGGTTATATTCCGACCTTTGAAGAATATCTGAAAAC
CTATGCAATTAGCGTTGGTCTGGGTCCGTGTACCCTGCAGCCGATTCTGCTGATGGGTGAACTG
GTGAAAGATGATGTTGTTGAGAAAGTTCATTACCCGAGCAACATGTTTGAACTGGTAAGCCTGAG
CTGGCGTCTGACCAATGATACCAAAACCTATCAGGCAGAAAAAGCACGTGGTCAGCAGGCAAGC
GGTATTGCATGTTATATGAAAGACAATCCGGGTGCAACCGAAGAGGATGCAATCAAACATATTTG
TCGTGTTGTTGATCGTGCACTGAAAGAAGCCAGCTTTGAATATTTCAAACCGAGCAACGATATTCC
GATGGGCTGTAAATCCTTTATCTTTAATCTGCGTCTGTGCGTGCAGATCTTCTATAAATTCATTGAT
GGTTACGGCATTGCCAACGAAGAGATCAAAGATTATATCCGCAAAGTGTATATCGATCCGATTCA
GGTTTAACTCGAG-3`

Appendix Figure 42: Gene of the codon optimized M60 truncation polypeptide of TXS from *Taxus brevifolia*

Appendix Figure 42: The red underscore depicts the used restriction sites for integration into the pET-duet-1 vector in MCS II (multiple cloning site) and the pET28b vector, according to methods chapter II. The red capitals depict the sequence of primer used for integration. This scheme of coloring and description will range over all following genes.

5`-
CCATGGATGACGGTCTGCGCAAAAAAACACGTTCACCCCACTCGCAGTGCTGCTGAGCAATTACT
GGCTGATATTGATCGGCGGCTTGAGCAGTTATTGCCTGTCGAGGGAGAACGGGATTTTGTAGGT
GCCGCGATGCGTGAGGGCGCGCTGGCACCGGGTAAACGTATTCGCCCGATGCTGTTGTTGCTC
ACAGCACGCGATCTGGGCTGCGCTGTCAGCCACGAGGGATTGCTGGATATCGCCTGCGCCGTG
GAAATGGTGCACGCGGCATCGTTGATTCTTGATGATATGCCCTGTATGGATGATGCGCAGATCGT
CGGCGGCCGCCCTACCGTTCACTGTCAGTACGGCGAGCATGTGGCAATACTGGCGGCGGTCGC
CTTACTGAGCAAAGCCTTTGGGGTCATTGCTGATGCAGATGGCCTGACCGCGCTGGCAAAAAAC
CGCGCAGTGTCAGAATTATCAAACGCCATTGGTATGCAAGGTTTGGTACAGGGTCAGTTTAAGGA
TCTGTCAGAAGGCGATAAGCGCCGCAGCGCTGAAGCCATTTTGATGACGAACCACTTTAAAACCA
GCACCCTGTTTTGCGCGTCTATGCAGATGGCCTCAATTGTGGCGAATGCCTCCAGCGAACGACG
CGATTACCTACACCGTTTTTCGCTTGATCTCGGCCAAGCGTTTCAACTGCTGGACGACTTAACGG
ATGGCATGGCTGACACGGGTAAGGACATCAATCAGGATGAAGGCAAATCGACCCTGGTCAATTTA
TTAGGCTCCAGAGCAGTTGAAGAGCGTCTGCGGCACAGTCTGCATCTCGCCAGTGAGCACCTCT
CAGCCGCGTGTCAAAACGGTCATTCTACTCAACAATTTATCCAGGCCTGGTTTGACAAAAAACTC
GCTGCCGTCAGTTAAGGATTC-3`

Appendix Figure 43: Gene of the codon optimized Crte polypeptide from *Pantoea agglomerans*

The red underscore depicts the used restriction sites for integration into the pET-duet-1 vector in MCS I, according to methods chapter II.

5`-
CCATGGATGAGTTTTGATATTGCCAAATACCCGACCCTGGCACTGGTCGACTCCACCCAGGAGTT
ACGACTGTTGCCGAAAGAGAGTTTACCGAAACTCTGCGACGAACTGCGCCGCTATTTACTCGACA
GCGTGAGCCGTTCCAGCGGGCACTTCGCCTCCGGGCTGGGCACGGTCGAACTGACCGTGGCGC
TGCACTATGTCTACAACACCCCGTTTGACCAATTGATTTGGGATGTGGGGCATCAGGCTTATCCG
CATAAAATTTTGACCGGACGCCGCGACAAAATCGGCACCATCCGTCAGAAAGGCGGCCTGCACC
CGTTCCCGTGGCGCGGCGAAAGCGAATATGACGTATTAAGCGTCGGGCATTCATCAACCTCCAT
CAGTGCCGGAATTGGTATTGCGGTTGCTGCCGAGAAAGAAGGCAAAAATCGCCGCACCGTCTGT
GTCATTGGCGATGGCGCGATTACCGCTGGCATGGCGTTTGAAGCGATGAATCACGCGGGCGATA

TCCGTCCTGATATGCTGGTGGTCCTCAACGACAATGAAATGTCGATTTCCGAAAATGTCGGCGCG
CTCAATAACCATCTGGCACAGCTGCTTTCCGGTAAGCTTTACTCTTCGCTGCGCGAAGGCGGGAA
AAAAGTTTTCTCTGGCGTTCCGCCAATTAAAGAGCTGCTCAAACGTACCGAAGAACATATTAAAGG
CATGGTAGTGCCTGGCACGTTGTTTGAAGAGCTGGGCTTTAACTACATCGGCCCGGTTGACGGT
CACGATGTGCTGGGGCTTATCACCACGCTGAAGAACATGCGCGACCTGAAAGGCCCGCAGTTCC
TGCATATCATGACCAAAAAAGGTCGTGGTTATGAACCGGCAGAAAAAGACCCCATCACTTTCCAC
GCCGTGCCTAAATTTGATCCCTCCAGCGGTTGTTTGCCGAAAAGTAGCGGCGGTTTGCCGAGCT
ATTCAAAAATCTTTGGCGACTGGTTGTGCGAAACGGCAGCGAAAGACAACAAGCTGATGGCGATT
ACTCCGGCGATGCGTGAAGGTTCCGGCATGGTCGAGTTTTCACGTAAATTCCCGGATCGTTACTT
CGACGTGGCAATCGCCGAGCAACACGCGGTGACCTTTGCCGCCGGTCTGGCGATTGGTGGGTA
CAAACCCATTGTCGCGATTTACTCCACTTTCCTGCAACGCGCCTATGATCAGGTGCTGCATGACG
TGGCGATTCAAAAGCTCCCGGTCCTGTTCGCCATCGACCGCGCGGGCATTGTTGGTGCTGACGG
TCAAACCCATCAGGGCGCTTTTGACCTCTCTTACCTGCGCTGTATACCGGAAATGGTCATTATGA
CCCCGAGCGATGAAAACGAATGTCGCCAGATGCTCTATACCGGCTATCACTATAACGACGGCCC
GTCCGCGGTGCGCTACCCGCGCGGTAACGCGGTTGGCGTGGAACTGACGCCGCTGGAAAAACT
GCCAATTGGCAAAGGCATTGTGAAGCGTCGTGGCGAGAAACTGGCGATCCTTAACTTTGGTACG
CTGATGCCAGACGCGGCGAAAGTCGCTGAATCGCTGAACGCTACGCTGGTCGATATGCGTTTTG
TGAAACCGCTTGATGAAGCGTTAATTCTGGAAATGGCCGCCAGCCATGAAGCGCTGGTCACCGT
AGAAGAAAACGCCATTATGGGCGGCGCAGGCAGCGGCGTGAACGAAGTGCTAATGGCCCATCG
TAAACCAGTACCCGTGCTGAACATTGGCCTGCCTGACTTCTTTATTCCACAAGGAACTCAGGAAG
AAATGCGCGCCGAACTCGGCCTCGATGCCGCCGGTATGGAAGCCAAAATCAAGGCCTGGCTGG
CATAAGAATTC-3`

Appendix Figure 44: Gene of the Dxs polypeptide from *E.coli*

The red underscore depicts the used restriction sites for integration into the pCola-duet-1 vector in MCS I, according to methods chapter II.

5`-
CATATGAAGCAACTCACCATTCTGGGCTCGACCGGCTCGATTGGTTGCAGCACGCTGGACGTGG
TGCGCCATAATCCCGAACACTTCCGCGTAGTTGCGCTGGTGGCAGGCAAAAATGTCACTCGCAT
GGTAGAACAGTGCCTGGAATTCTCTCCCCGCTATGCCGTAATGGACGATGAAGCGAGTGCGAAA
CTTCTTAAAACGATGCTACAGCAACAGGGTAGCCGCACCGAAGTCTTAAGTGGGCAACAAGCCG
CTTGCGATATGGCAGCGCTTGAGGATGTTGATCAGGTGATGGCAGCCATTGTTGGCGCTGCTGG
GCTGTTACCTACGCTTGCTGCGATCCGCGCGGGTAAAACCATTTTGCTGGCCAATAAAGAATCAC
TGGTTACCTGCGGACGTCTGTTTATGGACGCCGTAAAGCAGAGCAAAGCGCAATTGTTACCGGT
CGATAGCGAACATAACGCCATTTTTCAGAGTTTACCGCAACCTATCCAGCATAATCTGGGATACG
CTGACCTTGAGCAAAATGGCGTGGTGTCCATTTTACTTACCGGGTCTGGTGGCCCTTTCCGTGAG
ACGCCATTGCGCGATTTGGCAACAATGACGCCGGATCAAGCCTGCCGTCATCCGAACTGGTCGA
TGGGGCGTAAAATTTCTGTCGATTCGGCTACCATGATGAACAAAGGTCTGGAATACATTGAAGCG
CGTTGGCTGTTTAACGCCAGCGCCAGCCAGATGGAAGTGCTGATTCACCCGCAGTCAGTGATTC

ACTCAATGGTGCGCTATCAGGACGGCAGTGTTCTGGCGCAGCTGGGGGAACCGGATATGCGTAC
GCCAATTGCCCACACCATGGCATGGCCGAATCGCGTGAACTCTGGCGTGAAGCCGCTCGATTTT
TGCAAACTAAGTGCGTTGACATTTGCCGCACCGGATTATGATCGTTATCCATGCCTGAAACTGGC
GATGGAGGCGTTCGAACAAGGCCAGGCAGCGACGACAGCATTGAATGCCGCAAACGAAATCACC
GTTGCTGCTTTTCTTGCGCAACAAATCCGCTTTACGGATATCGCTGCGTTGAATTTATCCGTACTG
GAAAAAATGGATATGCGCGAACCACAATGTGTGGACGATGTGTTATCTGTTGATGCGAACGCGCG
TGAAGTCGCCAGAAAAGAGGTGATGCGTCTCGCAAGCTGACTCGAG-3`

Appendix Figure 45: Gene of the Dxr polypeptide from *E.coli*

The red underscore depicts the used restriction sites for integration into the pCola-duet-1 vector in MCS II, according to methods chapter II.

5`-
CCATGGATGGCAACCACTCATTTGGATGTTTGCGCCGTGGTTCCGGCGGCCGGATTTGGCCGTC
GAATGCAAACGGAATGTCCTAAGCAATATCTCTCAATCGGTAATCAAACCATTCTTGAACACTCGG
TGCATGCGCTGCTGGCGCATCCCCGGGTGAAACGTGTCGTCATTGCCATAAGTCCTGGCGATAG
CCGTTTTGCACAACTTCCTCTGGCGAATCATCCGCAAATCACCGTTGTAGATGGCGGTGATGAGC
GTGCCGATTCCGTGCTGGCAGGTTTGAAAGCCGCTGGCGACGCGCAGTGGGTATTGGTGCATG
ACGCCGCTCGTCCTTGTCTGCATCAGGATGACCTCGCGCGATTGTTGGCGTTGAGCGAAACCAG
CCGCACGGGAGGGATCCTAGCCGCACCAGTGCGCGATACGATGAAACGTGCCGAACCGGGCAA
AAATGCCATTGCTCATACCGTTGATCGCAACGGCTTATGGCACGCGCTGACGCCGCAATTTTTCC
CTCGTGAGCTGTTACATGACTGTCTGACGCGCGCTCTAAATGAAGGCGCGACTATTACCGACGAA
GCCTCGGCGCTGGAATATTGCGGATTCCATCCTCAGTTGGTCGAAGGCCGTGCGGATAACATTA
AAGTCACGCGCCCGGAAGATTTGGCACTGGCCGAGTTTTACCTCACCCGAACCATCCATCAGGA
GAATACATAAGCAGGAGCAGGAGCAGAAGGAGGAGCAGGAATGCGAATTGGACACGGTTTTGAC
GTACATGCCTTTGGCGGTGAAGGCCCAATTATCATTGGTGGCGTACGCATTCCTTACGAAAAAGG
ATTGCTGGCGCATTCTGATGGCGACGTGGCGCTCCATGCGTTGACCGATGCATTGCTTGGCGCG
GCGGCGCTGGGGGATATCGGCAAGCTGTTCCCGGATACCGATCCGGCATTTAAAGGTGCCGAC
AGCCGCGAGCTGCTACGCGAAGCCTGGCGTCGTATTCAGGCGAAGGGTTATACCCTGGGCAAC
GTCGATGTCACTATCATCGCTCAGGCACCGAAGATGTTGCCGCACATTCCACAAATGCGCGTATT
TATTGCCGAAGATCTCGGCTGCCATATGGATGATGTTAACGTGAAAGCCACTACTACGGAAAAAC
TTGGATTTACCGGACGTGGGGAAGGGATTGCCTGTGAAGCGGTGGCGCTACTCATTAAGGCAAC
AAAATGAGAATTC-3`

Appendix Figure 46: Gene of the IspD/F polypeptide from *E.coli*

The red underscore depicts the used restriction sites for integration into the pCDF-duet-1 vector in MCS I, according to methods chapter II.

5`-
CATATGCAAACGGAACACGTCATTTTATTGAATGCACAGGGAGTTCCCACGGGTACGCTGGAAAA
GTATGCCGCACACACGGCAGACACCCGCTTACATCTCGCGTTCTCCAGTTGGCTGTTTAATGCCA
AAGGACAATTATTAGTTACCCGCCGCGCACTGAGCAAAAAAGCATGGCCTGGCGTGTGGACTAA
CTCGGTTTGTGGGCACCCACAACTGGGAGAAAGCAACGAAGACGCAGTGATCCGCCGTTGCCGT
TATGAGCTTGGCGTGGAAATTACGCCTCCTGAATCTATCTATCCTGACTTTCGCTACCGCGCCAC
CGATCCGAGTGGCATTGTGGAAAATGAAGTGTGTCCGGTATTTGCCGCACGCACCACTAGTGCG
TTACAGATCAATGATGATGAAGTGATGGATTATCAATGGTGTGATTTAGCAGATGTATTACACGGT
ATTGATGCCACGCCGTGGGCGTTCAGTCCGTGGATGGTGATGCAGGCGACAAATCGCGAAGCCA
GAAAACGATTATCTGCATTTACCCAGCTTAAATAACTCGAG-3`

Appendix Figure 47: Gene of the Idi polypeptide from *E.coli*

The red underscore depicts the used restriction sites for integration into the pCDF-duet-1 vector in MCS II, according to methods chapter II.

5`-
CATATGACCACCGGTCTGAGCACCGCAGGCGCACAGGATATTGGTCGTAGCAGCGTTCGT
CCGTATCTGGAAGAATGTACCCGTCGTTTTCAAGAAATGTTTGATCGTCATGTTGTTACC
CGTCCGACCAAAGTTGAACTGACCGATGCAGAACTGCGTGAAGTTATTGATGATTGTAAT
GCAGCAGTTGCACCGCTGGGTAAAACCGTTAGTGATGAACGTTGGATTAGCTATGTTGGT
GTTGTTCTGTGGTCACAGAGTCCGCGTCATATTAAAGATATGGAAGCATTTAAAGCCGTG
TGCGTTCTGAATTGTGTTACCTTTGTTTGGGATGATATGGACCCTGCACTGCATGATTTT
GGTCTGTTTCTGCCTCAGCTGCGTAAAATTTGCGAAAAATACTATGGTCCGGAAGATGCC
GAAGTTGCCTATGAAGCAGCACGTGCATTTGTTACCAGCGATCACATGTTTCGTGATAGC
CCGATTAAAGCAGCACTGTGTACCACCAGTCCGGAACAGTATTTTCGTTTTCGTGTTACC
GATATTGGCGTGGATTTTTGGATGAAAATGAGCTATCCGATTTATCGCCATCCGGAATTT
ACCGAACATGCAAAAACCAGCCTGGCAGCACGTATGACCACCCGTGGTCTGACCATTGTT
AATGATTTCTATAGCTATGATCGCGAAGTTAGCCTGGGTCAGATTACCAATTGTTTTCGT
CTGTGTGATGTGAGTGATGAAACCGCCTTTAAAGAATTTTTTCAGGCACGTCTGGATGAC
ATGATCGAAGATATTGAATGCATCAAAGCGTTTGATCAGCTGACACAGGATGTTTTTCTG
GATCTGATTTATGGCAATTTTGTGTGGACCACCTCCAACAAACGTTATAAAACCGCAGTG
AATGATGTGAACAGCCGTATTCAATAACTCGAG-3`

Appendix Figure 48: Gene of the codon optimized CotB2 polypeptide from *Streptomyces melanosporofaciens*

The red underscore depicts the used restriction sites for integration into the pET-duet-1 vector in MCS II, according to methods chapter III.

5`-
CATATGCGTGAACGTGGTCCGGTTACACCGGCAAAAAGCAGCGCACCGCCTGAACGTCCG
TGGACCACCGGTACAGCACCGGGTAGCGTTCCGCTGCTGGGTCATACAATGGCACTGTGG
CGTCGTCCGCTGCAATTTCTGGCAAGCCTGCCTGCACATGGTGATCTGGTTGAAGTTCGT
CTGGGTCCGAGCCGTGCATATCTGGCATGTCATCCGGAACTGGTTCGTCAGGTTCTGCTG
AATCCGCGTGTTTTTGATAAAGGTGGCGTTTTCGATAAAGCACGTCAGCTGCTGGGCAAT
AGCCTGAGCGTTAGCCGTGGTGAAGATCATCGTTATCAGCGTCGTATGATTCAGCCTGCA
TTTCATACCCCGAAAATTGCCGCATATACCGCAGCAGTTGCAGATGATACCCGTGCAGCA
ATTGGTAGCTGGGAACCGGGTCGTACCCTGGATATTAGCGATACCATGCATGCACTGCTG
ATGCGTGTTGCAGCACGTACCCTGTTTAGCACCGGTATTGATGAAGCAACCATCGATGAA
GCCCGTCATTGTCTGCGTATTGTTAGTGATGGTATCTATAAACGTACGATGGCACCGCTG
GGTATTATGGAAAAACTGCCGACACCGGGTAATCGTCGTTATGATCGTGCAAATGCACGT
CTGCGTCAGATTGTTGATGAAATGATTCGTGAACGTCGTCGTAGCGGTGCAGATCATGGC
GATCTGCTGAGCACCCTGCTGCGTGCAGAACATCCGGAAACCGGTAAAGGTCTGGATGAT
GGTGAAGTTCTGGATCAGGTTGTTACCTTTCTGGTTGCAGGTAGCGAAACCACCGCAAGC
ACCCTGGCATTTGTTTTTCATCTGCTGGGTGCCCATCCGGAAGTTGAAAAACGTGTTCAT
GCAGAAATCGATGAAATTCTGGAAGGTCGTAGCCCGACCTTTGAAGATCTGCCGAGCCTG
GAATATACCCGTGGTGTTATTACCGAAAGCCTGCGTCTGTATCCGCCTAGCTGGATGGCA
ATGCGTGTTACCGCAGCCGAAACCGAACTGGGTGGTCGTACCGTTCCGGCAGGCACCATG
ATTCTGTATAGCGCACAGGCACTGCATCATAACCCGGAACTGTTTCCAGATCCGGAACGT
TTTGATCCTGAACGTTGGCTGGGTGATCGTGCCAAAGAAGTTGAACGCGGAGCACTGCTG
CCGTTTGGTGCCGGTAGCCATAAATGTATTGGTGATGTGCTGGCACTGACCGAAACAGCA
CTGATTGTTGCAACCATTGCAAGCCGTTGGCGTCTGCGTCCGGTTCCGGGTACAACCCTG
CGTCCTGAACCGAAAGCAACCCTGGAACCTGGTCCGCTGCCGATGGTTTGTGAACCGCGT
TAACTAGTAGGAGGAAAACATCATGAAAGATTTTTTTCGTATGCGCACCGCACAGCAGCC
TGCCACACGTCATTGGCGTCATACCGTTGCACCGGGTGGTCTGCCGCTGGCAGGTCACGC
TCTGCTGATGGCACGTAAACCTCTGCAATTCCTGGCCTCACTGCCAGCCCACGGCGATCT
GGTGGAACTGCGCCTGGGACCGCGTCCGGTGTATCTGCCGTGCCACCCTGAACTGGTGCA
GCAGGTACTGGTTAATGCACGTGTTTATGATACAGGCGGTCCGGTGAAAGAAAAAGCAAA
ACCGATTCTGGGTAATGGTCTGATTACCAGCGATTGGGCTGATCATCGTCGCCAGCGTCG
CCTGGTTCAGCCAGCCTTTCACACCGCACGTATTGCAAAGTATGCCGAAGTTATGGAACG
TGAATGTGAAGCAGAAAGCACCGCATGGACCGCACGTCGTCCGATTGATGTTAGCCATGA
AATGCTGGCCCTGACAGCACGCGTTACCGCACGTGCACTGTTTTCAACCGATATGGCTCC
GCATGCCGTTGCAGAAATTCAGCATTGCCTGCCGATTGTTGTTGAAGGTGCATATCGTCA
GGCAATTGATCCGACCGGTCTGCTGGCCAAACTGCCTCTGGCAGCAAATCGTCGCTTTGA
TGATGCACTGGCACGTCTGAACCAGCTGATTGATCGCATGATTGATGATTACAAAGCCAG
TGATGATGGCGATCGTGGTGATGTTCTGAGCGCACTGTTTGCAGCACAGGATGATGAAAC
CGGTGGTACAATGAGCGATCAAGAAATTCATGATCAGGTTATGACACTGCTGCTGGCTGG
TATTGAAACCACAGCCAGCGCACTGACCTGGGCATGGTTTCTGCTGGGACGTAATCCGGG
TGCGGAAGCAGCACTGCATGCGGAAGTGGATGAAGTGCTGGGTGGCCGTGCACCGCGTTA
TGCAGATGTTCCGCGTCTGGCATATACACAGCGTGTGTTTAGCGAAGCCCTGCGCCTGTT

TCCTCCGGCATGGCTGTTTACCCGTACCACCACCGAAACCACGGAACTGGGAGGCCGTCG
TCTGCCTCCGGCTTCAGATGTGCTGATTAGCCCGTATGTGCTGCATCGTGATCCAGCACT
GTTTCCGCGTCCTGATAGCTTTGACCCGGATCGTTGGCTGCCAGAACGCGCAAAAGAAGT
AACACGCGGTAGCTATCTGCCTTTTGGTGGTGGTTCACGTAAATGCATTGGCGACGTTTT
TGGTATGACCGAAGCAACACTGGCACTGGCAGCGATTGCCGGTCGTTGGCGGATGCGTCC
TATTCCTGGCACCAAAATTCGTCCGCGTCCGCAGATGAGCCTGACCGCAGGTCCTCTGCG
CATGATTCCGGAACCTCGTTAACTCGAG-3`

Appendix Figure 49: Gene of the CotB3/CotB4 polypeptide from *S. melanosporofaciens*

The red underscore depicts the used restriction sites for integration into the pACYC-duet-1 vector in MCS II, according to methods chapter III.

5`-
CCATGGGCCCTGCAAATGATGCCTGTGTTATTGTTGGTGCAGGTCTGGCAGGCGCAAAAG
CAGCACAGGCACTGCGTGAAGAGGGTTTTGATGGTCCGCTGGTTCTGATTGGTGATGAAC
GTGAACGTCCGTATGAACGTCCTCCGCTGAGCAAAGGTTATCTGACCGGTAAAGATGCAC
GTGAGCAGATTTATGTTCATCCGCCTCAGTGGTATGCAGAACATAATGTTGATATGCGTC
TGGGTATGGCAGTTACCGCAGTTGATCCGGCAGCACGTGAAATGACCCTGGATGATGGTA
GCCGTGTTGGTTATGGTAAACTGCTGCTGACCACCGGTAGCGCACCGCGTCGTCTGCCGG
TTCCTGGTGCTGGCCTGGAACGTGTTCTGTATCTGCGTCGTGTTGAAGATAGCGATCAGA
TTAAAGAAGCATTTCAGAGCGCAAGCCGTGCAGTTGTTATTGGTGCCGGTTGGATTGGTC
TGGAAACCACCGCAGCAGCACGTACCGCAGGCGTTGAAGTTACCGTTCTGGAAATGGCAG
AACTGCCGCTGCTGCGTGTTCTGGGTCGTGAAGTTGCACAGCTGTTTGCAAATCTGCATC
GTGATCATGGTGTTGATCTGCGTTTTGGTGCACAGGTTGCAGAAATTACCGGTAGTGGTG
GTGCAGTTGATGGTGTTCGTCTGAGTGATGGCACCCGTATTGATGCAGATGTTGTGATTG
TGGGTGTTGGCATTAAACCGAATATTGGCCTGGCACAAGAAGCCGGTCTGGAAGTTGATA
ATGGTATTCGTGTGGATGAACGTCTGCGTACCAGCTATCCTGATATTTATGCAGCCGGTG
ATGTTGCACATGCATTTCATCCGCTGCTGGGTAAACATATTCGCGTTGAACATTGGGCAA
ATGCACTGAATCAGCCGCAGATTGCAGCAAAAGCAATGCTGGGTCGCGAAGATGCCGTTT
ATGATCGTATTCCGTATTTTTTCACCGATCAGTATGATCTGGGCATGGAATATGCAGGTT
ATGTTGAACCGGGTGGTTATGATCAGGTTGTTTTTCGTGGTGATGTGGCAGGTCGTGAAT
TTATTGCATTTTGGCTGGCAGGCAATCGTGTGCTGGCAGGTATGAATGTGAATATTTGGG
ATGTTAATGATCAGCTGCAGACCCTGGTTCGTACCGCACAGACCGTTGATATTCCGATGC
TGACCGATCCGCAGGTTCCGCTGGGTAGCCTGCTGCCTGATCCTCAGCATCGTTAATATA
TTAGTTAAGTATAAGAAGGAGATATAATGCCGAAAGTTACCTATGTTAGTGATGCCGGTG
AAGTTCGTGTTGTTGATGGTCTGGTTGGTGATAGCGTTATGCAGACCGCAGTTCGTAATG
GTGTTCCGGGTATTACCGGTGAATGTGGTGGTGTTCTGAGCTGTGCAACCTGTCATGTTT
TTGTTGATGAAGCAGATCTGGATCGTCTGGAACCGGTTAGCGGTCTGGAAGATGAAATGC

TGGATGGCACCGTTGTTGATCGTTGTCCGAATAGCCGTCTGAGCTGCCAGATTAAACTGA
GCGAAGAACTGGGTGATCTGCGTGTTACCACACCGGAAGCACAAGAATAAGCGGCCGC-3`

Appendix Figure 50: Gene of the AfR/Afx polypeptide from *Streptomyces afghaniensis*

The red underscore depicts the used restriction sites for integration into the pACYC-duet-1 vector in MCS I, according to methods chapter III.

5`-
CCATGGGCAACGCCAATGATAATGTTGTTATTGTTGGCACCGGTCTGGCAGGCGTTGAAGTTGCA
TTTGGTCTGCGTGCAAGCGGTTGGGAAGGTAATATTCGTCTGGTTGGTGATGCAACCGTTATTCC
GCATCATCTGCCTCCGCTGAGCAAAGCATATCTGGCAGGTAAAGCAACCGCAGAAAGCCTGTAT
CTGCGTACACCGGATGCCTATGCAGCACAGAATATTCAGCTGCTGGGTGGTACACAGGTTACCG
CAATTAATCGTGATCGTCAGCAGGTTATTCTGAGTGATGGTCGTGCACTGGATTATGATCGTCTG
GTGCTGGCAACCGGTGGTCGTCCGCGTCCGCTGCCGGTTGCAAGTGGTGCAGTTGGTAAAGCC
AATAACTTTCGTTATCTGCGCACCCTGGAAGATGCAGAATGTATTCGTCGTCAGCTGATTGCAGAT
AATCGCCTGGTTGTGATTGGTGGTGGTTATATTGGTCTGGAAGTTGCAGCAACCGCCATTAAAGC
AAATATGCATGTTACCCTGCTGGATACCGCAGCACGTGTTCTGGAACGTGTTACCGCACCGCCTG
TTAGCGCCTTTTATGAACATCTGCATCGTGAAGCCGGTGTTGATATTCGTACCGGCACCCAGGTT
TGTGGTTTTGAAATGAGCACCGATCAGCAGAAAGTTACCGCAGTTCTGTGTGAAGATGGCACCCG
TCTGCCTGCAGATCTGGTTATTGCAGGTATTGGCCTGATTCCGAATTGTGAACTGGCAAGCGCAG
CAGGTCTGCAGGTTGATAATGGTATTGTTATTAACGAACACATGCAGACCAGCGATCCGCTGATT
ATGGCAGTTGGTGATTGTGCACGTTTTCATAGCCAGCTGTATGATCGTTGGGTTCGTATTGAAAG
CGTTCCGAATGCACTGGAACAGGCACGTAAAATTGCAGCAATTCTGTGTGGTAAAGTTCCGCGTG
ATGAAGCAGCACCGTGGTTTTGGAGCGATCAGTATGAAATTGGTCTGAAAATGGTTGGTCTGAGC
GAAGGTTATGATCGCATTATTGTTCGTGGTAGCCTGGCACAGCCGGATTTTTCAGTTTTTTATCTG
CAGGGTGATCGTGTGCTGGCAGTTGATACCGTTAATCGTCCGGTTGAATTTAACCAGAGCAAACA
AATTATCACCGATCGTCTGCCGGTGGAACCGAATCTGCTGGGAGATGAAAGCGTGCCGCTGAAA
GAAATTATTGCAGCAGCAAAAGCAGAACTGAGCAGCGCATAATATATTAGTTAAGTATAAGAAGGA
GATATAATGAGCAAAGTTGTGTATGTTAGCCATGATGGCACCCGTCGTGAACTGGATGTTGCAGA
TGGTGTGAGCCTGATGCAGGCAGCAGTTAGCAATGGTATTTATGATATTGTTGGTGATTGTGGTG
GTAGCGCAAGCTGTGCAACCTGTCATGTTTATGTTAATGAAGCCTTCACCGATAAAGTTCCGGCA
GCAAATGAACGTGAAATTGGTATGCTGGAATGTGTTACCGCAGAACTGAAACCGAATAGCCGTCT
GTGTTGTCAGATTATTATGACACCGGAACTGGACGGTATTGTTGTTGATGTTCCGGATCGTCAGT
GGTAAGCGGCCGC-3`

Appendix Figure 51: Gene of the PdR/Pdx polypeptide from *Pseudomonas putida*

The red underscore depicts the used restriction sites for integration into the pACYC-duet-1 vector in MCS I,

according to methods chapter III.

5`-
CATATGGCAAGCACCAAAAGCGAAACCGAAGCACGTCCGCTGGCATATTTTCCGCCTACCGTTTG
GGGTGATCGTCTGGCAAGCCTGACCTTTAATCAGCCTGCATTTGAACTGCTGAGTAAACAGGTTG
AGCTGCTGAACGAGAAAATCAAAAAAGAAATGCTGAATGTGAGCACCAGCGATCTGGCAGAAAAA
ATCATTCTGATTGATAGCCTGTGTCGTCTGGGTGTTAGCTATCATTTTGAAGAAGAAATCCAAGAA
AACCTGACCCGCATCTTTAATACCCAGCCGAATTTTCTGAACGAAAAAGATTATGATTTATTCACC
GTGGCCGTGATCTTTCGTGTTTTTCGTCAGCATGGCTTTAAAATCAGCAGCGACGTTTTTAACAAA
TTCAAAGATAGCGACGGGAAATTCAAAGAAAGCCTGCTGAATGATATTAAAGGCATCCTGAGCCT
GTTTGAAGCAACCCATGTTAGCATGCCGAATGAACCGATTCTGGATGAAGCACTGGCATTTACCA
AAGCATTTCTGGAAAGCAGCGCAGTTAAATCCTTTCCGAATTTTGCCAAACATATTAGCAGCGCAC
TGGAACAGCCGGTTCATAAAGGTATTCCGCGTCTGGAAGCACGCAAATATATCGATCTGTATGAA
GTTGATGAAAGCCGCAATGAAACCGTGCTGGAACTGGCGAAACTGGATTTTAATCGTGTTCAGCT
GCTGCATCAAGAAGAACTGAGCCAGTTTAGCAAATGGTGGAAAAGCCTGAATATTAGTGCCGAAG
TTCCGTATGCACGTAATCGTATGGCAGAAATCTTTTTTTGGGCAGTGAGCATGTATTTTGAACCGC
AGTATGCAAAAGCCCGTATGATTGTTAGCAAAGTGGTTCTGCTGATTAGCCTGATTGATGATACCA
TTGATGCCTATGCCACCATCGATGAAATCCATCGTGTTGCAGATGCAATTGAACGTTGGGATATG
CGTCTGGTTGATCAGCTGCCGAATTATATGAAAGTGATTTATCGCCTGATCATCAACACCTTTGAT
GAGTTTGAAAAAGATCTGGAAGCCGAGGGTAAAAGCTATAGCGTTAAATATGGTCGTGAAGCCTA
TCAAGAACTGGTGCGTGGTTATTACCTGGAAGCAATTTGGAAAGCGGATGGTAAAGTTCCGAGCT
TCGATGAGTATATCTATAATGGTGGTGTTACCACCGGTCTGCCGCTGGTTGCCACCGTTAGCTTT
ATGGGTGTTAAAGAAATCAAAGGCACCAAAGCCTTTCAGTGGCTGAAAACCTATCCGAAACTGAA
TCAGGCAGGCGGTGAATTTATTCGTCTGGTTAATGATGTTATGAGCCATGAAACCGAACAGGATC
GTGGTCATGTTGCAAGCTGTATTGATTGCTATATGAAACAGTATGGCGTGAGCAAAGAAGAAGCC
GTTGAAGAGATCCAGAAAATGGCAACCAATGAGTGGAAAAAACTGAACGAACAGCTGATTGTTCG
TAGCACCGAAGTTGTTCCGGTTAATCTGCTGATGCGTATTGTTAATCTGGTTCGTCTGACCGATGT
GAGCTATAAATACGGTGATGGTTATACCGATAGCTCCCAGCTGAAAGAATATGTGAAAGGCCTGT
TTATTGAACCGATCGCAACCGGTACCTCTTAACTCGAG-3`

Appendix Figure 52: Gene of the Casbene Synthase (CS) polypeptide from *Jatropha curcas*

The red underscore depicts the used restriction sites for integration into the pET-duet-1 vector in MCS II, according to methods chapter III.

5`-
CATATGAGCCAGAGCATTAGTCCGCTGATTTGTAGCCATTTTGCCAAATTTCAGAGCAATATTTGG
CGTTGTAATACCAGCCAGCTGCGTGTTATTCATAGCAGCTATGCCAGCTTTGGTGGTCGTCGTAA
AGAACGTGTTCGTCGTATGAATCGTGCAATGGATCTGAGCAGCAGCAGCCGTCATCTGGCAGATT
TTCCGAGCACCATTTGGGGTGATCATTTTCTGAGCTATAATAGCGAAATTACCGAGATCACCACC
CAAGAAAAAAACGAACATGAAATGCTGAAAGAAATCGTGCGCAAAATGCTGGTTGAAACACCGGA
TAATAGCACCCAGAAACTGGTTCTGATTGATACCATTCAGCGTCTGGGTCTGGCATATCATTTTAA
CGATGAAATCGAAAACAGCATCCAGAACATCTTTAATCTGTCCCAGAATAGCGAAGATGATGATG
AACATAATCTGTATGTTGCAGCACTGCGTTTTCGTCTGGCACGTCAGCAGGGCTATTATATGAGC
AGTGATGTCTTTAAACAGTTCACCAACCATGATGGCAAATTCAAAGAAAACCATACCAATGATGTT
CAGGGTCTGCTGAGCCTGTATGAAGCAGCACACATGCGTGTTCATGATGAAGAAATTCTGGAAGA
AGCCCTGATTTTTACCACCACCCATCTGGAAAGCGTTATTCCGAATCTGAGCAATAGCCTGAAAG
TTCAGGTTACCGAAGCACTGAGCCATCCGATTCGTAAAGCAATTCCGCGTGTTGGTGCACGCAAA
TATATCCATATTTATGAAAACATCGGCACCCACAATGATCTGCTGCTGAAATTCGCAAAACTGGAT
TTTAACATGCTGCAGAAACTGCACCGCAAAGAACTGAATGAACTGACCAGTTGGTGGAAAGACCT
GGATCGTGCAAACAAATTTCCGTATGCAAAAGATCGTCTGGTGGAAGCATATTTTTGGACCGTTG
GTATCTATTTTGAACCGCAGTATAGCCGTAGCCGTAGTCTGGTTACCAAAGTTGTGAAAATGAACA
GCATTATCGATGATACCTATGATGCCTATGCGACCTTTGATGAACTGGTGCTGTTTACCGATGCAA
TCCAGCGTTGGGATGAAGGCGCTATGGACCTGCTGCCGACCTATCTGCGTCCGATTTATCAGGG
CCTGCTGGATGTTTTTAATGAAATGGAAGAAGTGCTGGCCAAAGAAGGTAAAGCCGATCATATCT
ATTACGCCAAAAAAGAAATGAAAAAAGTGGCCGAAGTGTATTTCAAAGAAGCCGAATGGCTGAAC
GCAAACTATATTCCGAAATGCGAAGAGTACATGAAAAATGGTCTGGTTAGCAGCACCGGTCCGAT
GTATGGTATTATTAGCCTGGTTGTTATGGAAGAAATCATCACCAAAGAAGCGTTTGAGTGGCTGA
CCAATGAACCGCTGATTCTGCGTGCAGCAAGTACCATTTGTCGTCTGATGGATGATATGGCAGAT
CATGAAGTTGAACAGCAGCGTGGTCATGTTGCGAGCTTTGTTGAATGTTACATGAAAGAATATGG
CGTGAGCAAACAAGAAACCTATGTGGAAATGCGCAAAAAAATCACCAATGCCTGGAAAGATATTA
ACAAAGAGCTGCTGCGTCCTACCGCAGTTCCGATGTTTATTCTGGAACGTAGTCTGAATTTTAGC
CGTCTGGCCGATACCTTTCTGAAAGATGACGATGGTTATACCAACCCGAAAAGCAAAGTTAAAGA
CCTGATTGCAAGCCTGTTTGTTGAAAGCGTTGATCTCGAG-3`

Appendix Figure 53: Gene of the CBTS polypetide from *Nicotiana tabacum*

The red underscore depicts the used restriction sites for integration into the pET-duet-1 vector in MCS II and the pET28b vector according to methods chapter IV.

5`-
ATGCAGTTCTTCAACTTCTTTAGCCTGTTTCTGTTTGTTAGCTTTCTGTTCCTGTTCAAAAAATGGA
AAAACAGCAACAGCCAGACCAAACGTCTGCCTCCGGGTCCGTGGAAACTGCCGATTCTGGGTAG
CATGCTGCACATGCTGGGTGGTCTGCCGCATCATGTTCTGCGTGATCTGGCCAAAAAATACGGTC
CGATTATGCATCTGCAGCTGGGTGAAGTTAGCCTGGTTGTTATTAGCAGTCCGGGTATGGCAAAA
GAAGTTCTGAAAACCCACGATCTGGCATTTGCCAATCGTCCGCTGCTGGTTGCAGCAAAAATCTT
TAGCTATAACTGCATGGATATTGCCCTGAGCCCGTATGGTAATTATTGGCGTCAGATGCGTAAAAT
TTGTCTGCTGGAACTGCTGAGCGCCAAAAATGTGAAATCCTTTAATAGCATTCGCCAGGATGAAG
TGCATCGGATGATTAAATTCTTTCGTAGCTCACCGGGTAAACCGGTTAATGTTACCAAACGTATTT
CCCTGTTTACCAATAGCATGACCTGTCGTAGCGCATTTGGTCAAGAATATAAAGAACAGGATGAG
TTTGTCCAGCTGGTGAAAAAAGTTAGCAATCTGATTGAAGGTTTCGATGTGGCAGATATTTTTCCG
AGCCTGAAATTTCTGCATGTTCTGACCGGTATGAAAGCCAAAGTTATGAATACCCATAATGAGCTG
GATGCCATTCTGGAAAACATTATCAACGAGCACAAAAAAACCAGCAAAAGTGATGGTGAAAGCGG
TGGTGAAGGTATTATTGGTGTTCTGCTGCGTCTGATGAAAGAAGGTGGCCTGCAGTTTCCGATTA
CCAACGATAACATTAAAGCCATCATCAGCGACATCTTTGGTGGTGGCACCGAAACCAGCAGTACC
ACCATTAATTGGGCAATGGTTGAGATGATGAAAAACCCGAGCGTTTTTAGCAAAGCACAGGCCGA
AGTTCGTGAAATTCTGCGTGGTAAAGAAACCTTTGGTGAAATCGATGTGGAAGAGTTCAAATATCT
GAAAATGGTGATCAAAGAAACGTTCCGTCTGCATCCGCCTCTGCCGCTGCTGCTGCCTCGTGAAT
GTCGTGAAGAAATTGATCTGAATGGTTATACCATCCCGCTGAAAACCAAAGTGGTTGTTAATGCAT
GGGCTATGGGTCGTGATCCGAAATATTGGGATGATGTTGAATCGTTTAAACCGGAACGCTTTGAA
CACAACAGCATGGATTATATCGGCAACAACTATGAGTATCTGCCGTTTGGTAGCGGTCGTCGTAT
TTGTCCTGGTATTAGCTTTGGCCTGGCCAATGTTTATTTTCCGCTGGCCCAGCTGCTGAATCATTT
TGATTGGAAACTGCCTACCGGTATTAATCCGCGTAATTGTGATCTGACCGAAGCAGCCGGTGCAG
CCTGTGCACGTAAAAATGATCTGCATCTGATTGCAACCGCATATCAGCATTGTGAAGAATGA-3`

Appendix Figure 54: Gene of the wildtype n-CYP450 polypeptide from *Nicotiana tabacum*

5`-
ATGGCCAAACCGATTGTTTTTCCGAAAAGCCTGCATGTTGAACCGGAAGAAGAAATTGAACTGGA
ACCGGGTAAAGTTAAAGTGACCGTTTTTTTTGGCACCCAGACCGGCACCGCAGAAGGTTTTGCCA
AAGCACTGAGCGAAGAAATTAAAGCCCGTTATGAAAAAGCCGTTGTGAAAGTTGTTGATATGGAT
GATTATGCAGCCGATGATGATCTGTATGAAGAAAAACTGAAAAAAGAAACCGTGGCCTTTTTTATG
GTTGCCACCTATGGTGATGGTGAACCGACCGATAATGCAGCACGTTTCTATAAATGGTTTACCGA
AGGTCAAGAACGTGGCACCTGGCTGCAGCATCTGACCTATGGCGTTTTTGGTCTGGGTAATCGT
CAGTATGAACACTTTAACAAAATCGGCAAAGTGATCGATGAACAGCTGAGCGAACAGGGTGCAAA
ACGTATTGTTCCGGTTGGCCTGGGTGATGATGACCAGTGTATTGAAGATGATTTTGCCGCATGGC
GTGAAACCCTGTGGCCTGAACTGGATCAGCTGCTGCGTGATGAGGATGATGAACCTACCAGCGC
AACCCCGTATACCGCAGCAATCCCGGAATATCGTCTGGTTATTCATGATACAGCAGCAAGCCATG

ATGATAAACATGCAAGCATGGCAAATGGCAATACCAGCTATGATATTCATCACCCGTGTCAGGTTA
ATGTTGCAGTGCAGCGTGAACTGCATACACCGGAAAGCGATCGTAGCTGTATTCATCTGGAATTT
GATATTTGTGGCACCGGCATTAGCTATGAAACCGGTGATCATGTTGGTGTGTATGCCGAAAATCC
GGAAGAAATCGTTGAAGAAGCAGCACGTCTGCTGGGTCAGAGCCTGGATCTGATTTTTAGCATTC
ATACCGATAAAGAAGATGGCACCGCACTGGGTGGTAGCCTGCCTCCGCCTTTTCCGGGTCCGTG
TACCCTGCGTGCAGCACTGGCACGTTTTGCCGATCTGCTGAATCCGCCTCGTAAAGCAACCCTG
GTTGCACTGGCAGCACATGCAGCAGAACCGAGCGAAGCAGAGAAACTGAAATTTCTGGCAAGTC
CGCAGGGTAAAGATGAATATAGCGCATGGGTTGTTGCAAGCCAGCGTAGCCTGCTGGAAGTTAT
GGCAGAATTTCCGAGCGCAAAACCGCCTCTGGGTGTGTTTTTTGCAGCAGTTAGTCCGCGTCTG
CAGCCTCGTTATTATAGCATTAGCAGCAGTCCGCGTTTTGCTCCGGCACGTATTCATGTTACCTGT
GCACTGGTTTATGGTCCGACCCCGACAGGTCGTATTCATAAAGGTGTTTGTAGCACCTGGATGAA
AAATGCGGTTCCGCTGGAAAAAAGCCAGAATTGTAGCAGCGCACCGATTTTTATCCGTCCGAGCA
ACTTTAAACTGCCTGCCGATAATAAAGTGCCGATTATTATGATTGGTCCGGGTACAGGTCTGGCA
CCGTTTCGTGGTTTTCTGCAAGAACGTCTGGCCCTGAAAAAAGAGGGTGCAGAACTGGGTCCGG
CAGTTCTGTTTTTTGGTTGTCGTAATCGCCAGATGGATTACATCTATCAAGAAGAACTGGACAACT
TTCTGGAAGCCGGTGCACTGAGTGAACTGGTTGTTGCCTTTAGCCGTGAAGGTCCGAATAAAGAA
TATGTGCAGCATAAAATGAGCGAGAAAGCAGCCGATATTTGGAATATGATTAGCCAGGGTGGTTA
TGTTTATGTTTGTGGTGATGCAAAAGGTATGGCACGTGATGTTCATCGTACCCTGCATACCATTGC
ACAGGATCAGGGTAGCCTGGATAGCAGCAAAGCAGAAGGTCTGGTTAAAAATCTGCAGACCACC
GGTCGTTATCTGCGTGACGTTTGGAGCTGA-3`

Appendix Figure 55: Gene of the assembled n-CPR polypeptide from *Nicotiana tabacum*

5`-
ACCATGGCACTGCTGCTGGCAGTTTTTAGCCTGTTTCTGTTTGTTAGCTTTCTGTTCCTGTTCAAA
AAATGGAAAAACAGCAACAGCCAGACCAAACGTCTGCCTCCGGGTCCGTGGAAACTGCCGATTC
TGGGTAGCATGCTGCACATGCTGGGTGGTCTGCCGCATCATGTTCTGCGTGATCTGGCCAAAAA
ATACGGTCCGATTATGCATCTGCAGCTGGGTGAAGTTAGCCTGGTTGTTATTAGCAGTCCGGGTA
TGGCAAAAGAAGTTCTGAAAACCCACGATCTGGCATTTGCCAATCGTCCGCTGCTGGTTGCAGCA
AAAATCTTTAGCTATAACTGCATGGATATTGCCCTGAGCCCGTATGGTAATTATTGGCGTCAGATG
CGTAAAATTTGTCTGCTGGAACTGCTGAGCGCCAAAAATGTGAAATCCTTTAATAGCATTCGCCAG
GATGAAGTGCATCGGATGATTAAATTCTTTCGTAGCTCACCGGGTAAACCGGTTAATGTTACCAAA
CGTATTTCCCTGTTTACCAATAGCATGACCTGTCGTAGCGCATTTGGTCAAGAATATAAAGAACAG
GATGAGTTTGTCCAGCTGGTGAAAAAAGTTAGCAATCTGATTGAAGGTTTCGATGTGGCAGATATT
TTTCCGAGCCTGAAATTTCTGCATGTTCTGACCGGTATGAAAGCCAAAGTTATGAATACCCATAAT
GAGCTGGATGCCATTCTGGAAAACATTATCAACGAGCACAAAAAAACCAGCAAAAGTGATGGTGA
AAGCGGTGGTGAAGGTATTATTGGTGTTCTGCTGCGTCTGATGAAAGAAGGTGGCCTGCAGTTTC
CGATTACCAACGATAACATTAAAGCCATCATCAGCGACATCTTTGGTGGTGGCACCGAAACCAGC

AGTACCACCATTAATTGGGCAATGGTTGAGATGATGAAAAACCCGAGCGTTTTTAGCAAAGCACA
GGCCGAAGTTCGTGAAATTCTGCGTGGTAAAGAAACCTTTGGTGAAATCGATGTGGAAGAGTTCA
AATATCTGAAAATGGTGATCAAAGAAACGTTCCGTCTGCATCCGCCTCTGCCGCTGCTGCTGCCT
CGTGAATGTCGTGAAGAAATTGATCTGAATGGTTATACCATCCCGCTGAAAACCAAAGTGGTTGTT
AATGCATGGGCTATGGGTCGTGATCCGAAATATTGGGATGATGTTGAATCGTTTAAACCGGAACG
CTTTGAACACAACAGCATGGATTATATCGGCAACAACTATGAGTATCTGCCGTTTGGTAGCGGTC
GTCGTATTTGTCCTGGTATTAGCTTTGGCCTGGCCAATGTTTATTTTCCGCTGGCCCAGCTGCTG
AATCATTTTGATTGGAAACTGCCTACCGGTATTAATCCGCGTAATTGTGATCTGACCGAAGCAGCC
GGTGCAGCCTGTGCACGTAAAAATGATCTGCATCTGATTGCAACCGCATATCAGCATTGTGAAGA
AGGTAGCACCAGCTCAGGTAGCGGTGCTAGCATGGCCAAACCGATTGTTTTTCCGAAAAGCCTG
CATGTTGAACCGGAAGAAGAAATTGAACTGGAACCGGGTAAAGTTAAAGTGACCGTTTTTTTTGG
CACCCAGACCGGCACCGCAGAAGGTTTTGCCAAAGCACTGAGCGAAGAAATTAAAGCCCGTTAT
GAAAAAGCCGTTGTGAAAGTTGTTGATATGGATGATTATGCAGCCGATGATGATCTGTATGAAGA
AAAACTGAAAAAAGAAACCGTGGCCTTTTTTATGGTTGCCACCTATGGTGATGGTGAACCGACCG
ATAATGCAGCACGTTTCTATAAATGGTTTACCGAAGGTCAAGAACGTGGCACCTGGCTGCAGCAT
CTGACCTATGGCGTTTTTGGTCTGGGTAATCGTCAGTATGAACACTTTAACAAAATCGGCAAAGTG
ATCGATGAACAGCTGAGCGAACAGGGTGCAAAACGTATTGTTCCGGTTGGCCTGGGTGATGATG
ACCAGTGTATTGAAGATGATTTTGCCGCATGGCGTGAAACCCTGTGGCCTGAACTGGATCAGCTG
CTGCGTGATGAGGATGATGAACCTACCAGCGCAACCCCGTATACCGCAGCAATCCCGGAATATC
GTCTGGTTATTCATGATACAGCAGCAAGCCATGATGATAAACATGCAAGCATGGCAAATGGCAAT
ACCAGCTATGATATTCATCACCCGTGTCAGGTTAATGTTGCAGTGCAGCGTGAACTGCATACACC
GGAAAGCGATCGTAGCTGTATTCATCTGGAATTTGATATTTGTGGCACCGGCATTAGCTATGAAA
CCGGTGATCATGTTGGTGTGTATGCCGAAAATCCGGAAGAAATCGTTGAAGAAGCAGCACGTCT
GCTGGGTCAGAGCCTGGATCTGATTTTTAGCATTCATACCGATAAAGAAGATGGCACCGCACTGG
GTGGTAGCCTGCCTCCGCCTTTTCCGGGTCCGTGTACCCTGCGTGCAGCACTGGCACGTTTTGC
CGATCTGCTGAATCCGCCTCGTAAAGCAACCCTGGTTGCACTGGCAGCACATGCAGCAGAACCG
AGCGAAGCAGAGAAACTGAAATTTCTGGCAAGTCCGCAGGGTAAAGATGAATATAGCGCATGGG
TTGTTGCAAGCCAGCGTAGCCTGCTGGAAGTTATGGCAGAATTTCCGAGCGCAAAACCGCCTCT
GGGTGTGTTTTTTGCAGCAGTTAGTCCGCGTCTGCAGCCTCGTTATTATAGCATTAGCAGCAGTC
CGCGTTTTGCTCCGGCACGTATTCATGTTACCTGTGCACTGGTTTATGGTCCGACCCCGACAGGT
CGTATTCATAAAGGTGTTTGTAGCACCTGGATGAAAAATGCGGTTCCGCTGGAAAAAAGCCAGAA
TTGTAGCAGCGCACCGATTTTTATCCGTCCGAGCAACTTTAAACTGCCTGCCGATAATAAAGTGC
CGATTATTATGATTGGTCCGGGTACAGGTCTGGCACCGTTTCGTGGTTTTCTGCAAGAACGTCTG
GCCCTGAAAAAAGAGGGTGCAGAACTGGGTCCGGCAGTTCTGTTTTTTGGTTGTCGTAATCGCCA
GATGGATTACATCTATCAAGAAGAACTGGACAACTTTCTGGAAGCCGGTGCACTGAGTGAACTGG
TTGTTGCCTTTAGCCGTGAAGGTCCGAATAAAGAATATGTGCAGCATAAAATGAGCGAGAAAGCA
GCCGATATTTGGAATATGATTAGCCAGGGTGGTTATGTTTATGTTTGTGGTGATGCAAAAGGTATG
GCACGTGATGTTCATCGTACCCTGCATACCATTGCACAGGATCAGGGTAGCCTGGATAGCAGCA
AAGCAGAAGGTCTGGTTAAAAATCTGCAGACCACCGGTCGTTATCTGCGTGACGTTTGGAGCTGA
AAGCTT-3`

Appendix Figure 56: Gene of the assembled nt-CYP450-n-CPR polypeptide from *Nicotiana tabacum*

The red underscore depicts the used restriction sites for the integration into the pET28b-vector during generation of the fusion construct and for the subsequent integration into the pACYC-duet-1 vector in MCS I, according to methods chapter IV.