Exploring the catalytic cascade of cembranoid biosynthesis by combination of genetic engineering and molecular simulations

Patrick Schrepfer a, Ilke Ugur b,1, Sven Klumpe b,2, Bernhard Loll c,2, Ville R.I. Kaila b,d,1, Thomas Brück a,2

a Werner Siemens-Chair of Synthetic Biotechnology, Department of Chemistry, Technical University of Munich, Lichtenbergstr. 4, 85748 Garching, Germany
b Institute of Chemistry and Biochemistry, Laboratory of Structural Biochemistry, Freie Universität Berlin, Takustr. 6, 14195 Berlin, Germany
c Department of Biochemistry and Biophysics, Stockholm University, Svante Arrhenius väg 16, 10691 Stockholm, Sweden
1 Present address: Department Molecular Structural Biology, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany

A R T I C L E   I N F O

Article history:
Received 13 March 2020
Received in revised form 17 June 2020
Accepted 19 June 2020
Available online 25 June 2020

Keywords:
Terpene cyclization
QM/MM
Oxy-functionalization
Sustainable biocatalysis
Molecular dynamics

A B S T R A C T

While chemical steps involved in bioactive cembranoid biosynthesis have been examined, the corresponding enzymatic mechanisms leading to their formation remain elusive. In the tobacco plant, Nicotiana tabacum, a putative cembratriene-ol synthase (CBTS) initiates the catalytic cascade that leads to the biosynthesis of cembratriene-4,6-diols, which displays antibacterial- and anti-proliferative activities. We report here on structural homology models, functional studies, and mechanistic explorations of this enzyme using a combination of biosynthetic and computational methods. This approach guided us to develop an efficient de novo production of five bioactive non- and monohydroxylated cembranoids. Our homology models in combination with quantum and classical simulations suggested putative principles of the CBTS catalytic cycle, and provided a possible rationale for the formation of premature olefinic side products. Moreover, the functional reconstruction of a N. tabacum-derived class II P450 with a cognate CPR, obtained by transcriptome mining provided for production of bioactive cembratriene-4,6-diols. Our combined findings provide mechanistic insights into cembranoid biosynthesis, and a basis for the sustainable industrial production of highly valuable bioactive cembranoids.

© 2020 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Terpenoids are the largest class of microbial secondary metabolites, characterized by a complex chemical architecture and a high degree of functional decoration that render their total chemical synthesis production challenging. Examples of high-value bioactive terpenoids are taxol, a potent anticancer drug, and artesiminin, an anti-malaria agent [1]. Cembranoids, represent a structurally rather diverse group of monocyclic diterpenoids (C20 terpenoids), which have been reported in soft coral and plants of the genus Pinus and Nicotiana sp. (tobacco) [2]. Most notably, cembranoids represent the major fraction of secondary metabolites detected in the secretions of tobacco plants, which indicates that they may be involved in its natural defense mechanisms. Interestingly, these hydroxylated cembranoids are reported to be of significant biomedical relevance, encompassing anti-bacterial, anti-fungal and anti-proliferative activities [2]. Almost 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). These two compounds alone are reported to display a broad spectrum of antibacterial and antifungal activities. More recently, CBT-diols are flagged as potential therapeutics for neurodegenerative diseases that are increasingly prevalent in industrialized countries due to demographic change [3]. Furthermore, these compounds inhibit nicotine sensitization in rats, rendering them also interesting in treatment of nicotine addiction. The biosynthetic precursors of CBT-diols are (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-epi-isocembrol), respectively. While these CBT-ol precursors are detected in much lower concentrations in planta, they are reported to have potent insecticidal activity specifically against aphids [4].

https://doi.org/10.1016/j.csbj.2020.06.030
2001-0370 © 2020 The Author(s). Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
While the chemical steps in the biosynthesis of cembranoids have been elucidated, the detailed enzymatic mechanisms leading to the formation of this important class of natural products remains elusive [2]. The biosynthesis of CBT-ols takes place in the plastids of the glandular trichome cells of tobacco. In analogy to other plant diterpenes, the biosynthesis of cembranoids is initiated by cyclisation of the aliphatic, achiral (\(E, E, E\))-geranylgeranyl pyrophosphate (GGPP) to form both monocyclic CBT-ol epimers. This process is catalyzed by a putative class I diterpene synthase (dTPS) termed cembratriene-ol synthase (CBTS) [3] (Fig. 1).

A recent tobacco (\(N.\) tabacum) leaf trichome gene expression profiling resulted in annotation of a putative class I dTPS gene, which may be responsible for the cyclization of GGPP to both CBT-ols [5]. However, the definitive function of this putative dTPS gene has not yet been confirmed. Ionization dependent class I mono-, sesqui-, and di-terpene synthases contain a conserved asparagine-rich motif, and a second typical conserved “NSE/DTE” motif, involved in chelating a triade of catalytically important Mg\(^{2+}\) ions [6,7]. This Mg\(^{2+}\) ion triade orients the GGPP pyrophosphate moiety (PPi) within the enzymes active site, followed by active site closure, and GGPP hydrolysis forming the geranylgeranyl carbocation that initiates the catalytic cycle. The structural remodeling of the initial carbocation intermediate towards the final product CBT-ol, may involve sequential hydride shifts, protonation and deprotonation reactions, which are orchestrated and chaperoned by the amino acid lining the enzymes active site [8].

Despite their low primary sequence identity, TPS share highly conserved tertiary and quaternary structural features, which are dominated by an \(\alpha\)-helical barrel fold. The open, inactive TPS conformation, binds the linear GGPP substrate, which in turn triggers the formation of the closed, active conformation that catalyzes the reaction by stereoelectronic, electrostatic- , and kinetic control [8]. However, structural information of the closed, active conformation of the enzyme is required for understanding its catalytic principles. Recently the closed conformation of a dTPS was modeled and central steps of its catalytic cycle was characterized [8]. A putative mechanism for CBTS catalyzed CBT-ol formation has also been proposed, but the cyclization cascade and formation of promiscuous side-products, remains unclear [9].

The final biocatalytic step in the biosynthesis of bioactive CBT-diols is the hydroxylation of CBT-ols at the C6 atom by a putative class II CYP450 monooxygenase (P450). Class II P450s receive electrons from a class II membrane bound FAD- and FMN- containing NADPH-dependent oxidoreductase (CPR), by a mechanism that has remained unclear. At present, it is challenging to achieve efficient hydroxylation of terpenoid macrocycles by class II P450s in conventional prokaryotic hosts due to their low solubility. A recent study, however, demonstrated that the overall production rate is highly dependent on the choice of the CPR and can be improved by transmembrane engineering in the prokaryotic host [10]. As the native CPR is often unknown, only a few examples for the successful hydroxylation of terpenoid macrocycles have been reported with class II P450s. Catalytic principles of terpene synthases have previously also been subjected to many computational studies that have highlighted key chemical transformation steps [11–22].

We study here catalytic principles leading to \(N.\) tabacum derived CBT-ol and CBT-diol formation by reconstruction of the biosynthetic pathway in an engineered \(E.\) coli host. Initially, we define the activity of a CBTS that is responsible for cyclisation of GGPP to CBT-ol. The catalytic principles are then explored based
on homology models and quantum chemical calculations. Although the homology models do not provide as good starting point as experimentally resolved structures, they nevertheless allowed detailed insights into the cyclization cascade and identification of a catalytically relevant active site. Residues identified in the simulations were probed using experimental protein engineering approaches, which enabled us to develop an efficient de novo production of seven highly bioactive olefin, mono- and di-hydroxylated cembranoids in a metabolically engineered E. coli based microbial production host.

2. Material and methods

2.1. Bacterial strains, genes and vectors, cloning strategies, site-directed mutagenesis

The E. coli strains XL-1 Blue was used for cloning and BL21 (DE3) was used for the diterpene production. All strains and plasmids were obtained from Novagen/ Merk Millipore. Genes were synthesized by Life technologies GmbH containing the appropriate restriction sites and adjusting codon usage for E. coli. For genes, plasmids, and primer sequences used for the cembranoid production see the Supporting information materials (Supporting Tables S1–S3). All cloning procedures were performed according to standard protocols.

2.2. In vivo production of the cembranoid isomers

The in vivo approach was based on a culture volume of 3 × 1 L M9CA (M9 minimal media with 4 g/L casamino acids (Merck, Germany)) was used in 5 L baffled glass flasks. The vectors pColaDuets-1 (dxp, dxs), pCDFDuets-1 (ispD/ispF, idi) and the vector pETDuet-1 (crtE, cbts) were introduced into E. coli BL21 (DE3) by standard transformation procedures. For cultivation in shake flasks, single transformants were grown in 3 × 1 L M9CA medium supplemented with 10 g/L glycerol, 30 µg/mL kanamycin, 50 µg/mL streptomycin and 50 µg/mL carbenicillin. The cultures were inoculated with OD600 of 0.1 from an overnight culture (8 h cultivation at 28°C) and grown at 37°C supplemented with 50 µg/mL kanamycin, 50 µg/mL streptomycin and 50 µg/mL carbenicillin, grown at 37°C and 130 rpm until OD600 of 0.8 and then cooled down to 25°C. At a temperature of 25°C 40 g/L glycerol was added and the cultures were induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside. Cells were grown at 25°C for 3 days.

2.3. Extraction and purification of the CBTS-derived cembranoid isomers

For the cembranoid 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 MgSO4 and evaporated to dryness under continuous N2-stream. The crude extract was solved in 1 mL ethyl acetate and analyzed by GC–MS and GC-FID (Supplemental Information). The crude extract was resolved in 1 mL n-ethyl-acetate 100 µL of that resolved extract were evaporated to dryness under continuous N2-stream, solved in 90 mL of internal standard (1α-humulene, Sigma-Aldrich, final concentration 0.889 mg/L) were 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 ethyl-acetate was added to the reaction solution. After vigorous vortexing (2 × 20 s) phases were separated by centrifugation.

2.4. Batch bioprocess for the production of cembranoid isomers

5 L fermentations were performed in a 10 L bioreactor (LP351, Bioengineering AG, Wald, Switzerland) using M9CA medium 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 NaOH and 5 M H2PO4. 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 × 1 L overnight cultures (8 h cultivation at 37°C supplemented with 50 µg/mL kanamycin, 50 µg/mL streptomycin and 50 µg/mL carbenicillin) at OD600 = 0.25 and run 2 h at 37°C. At a OD600 of 1.95 45 g/L glycerol and 1 mM IPTG 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 OD600. 1 mL was used for determination of glycerol content by HPLC analysis. After the bioprocess 500 mL of culture was used for determination of cembranoid contents using ethyl-acetate extraction of the supernatant. The crude extract was resolved in 1 mL n-ethyl-acetate 100 µL of that resolved extract were evaporated to dryness under continuous N2-stream, solved in 90 mL and 10 mL of the internal standard (±)-isocembrene and (±)-isocembrene and grown at 37°C for further use.

2.5. Expression of CBTS, Δ56-CBTS and mutants for in vitro screening

Single transformations of pET28b (+) harboring wild type cbts, Δ56-CBTS or mutations were cultivated for 18 h at 37°C in LB medium (10 g L−1 tryptone, 5 g L−1 yeast extract, and 10 g L−1 NaCl) supplemented with 50 µg mL−1 kanamycin. These preinocula were used to seed fresh 500 mL LB media containing 50 µg mL−1 kanamycin at a starting OD600 of 0.1. After reaching an OD600 of 0.6 the cultures were cooled down to 16°C, induced by 1 mM 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 pH 6.8, 1 mM MgCl2 and 1 mM 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 1 h. The clear supernatant was applied to a pre-equilibrated HisTrap HP 5 mL column (GE Healthcare) at a flow rate of 1 mL min−1 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−1. SDS-PAGE (12%) analysis was used to evaluate protein expression. Selected fractions were combined and dialysed overnight 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 isotropic size exclusion chromatography using a Superdex 26/70 10/300 GL column (GE Healthcare) with buffer A at a flow rate of 0.8 mL min−1. Fractions were combined and concentrated to 2.5 mL. Protein concentration was determined by absorbance at 280 nm and enzyme were stored in 100 µL aliquots at −20°C for further use.

2.6. In vitro assay using purified enzymes

100 µL of enzyme solution (1 mg mL−1) were added to 400 µL of buffer A (see above section). 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 700 mL ethyl-acetate was added to the solution. After vigorous vortexing (2 × 20 s) phases were separated by centrifugation.
for 10 min, 4 °C and 12,000 rpm. Ethyl-acetate extraction was repeated twice; subsequently the organic phases were combined and evaporated to dryness under continuous N₂-stream. The crude extract was resolved in 1 mL ethyl-acetate. 100 µL of that resolved extract were evaporated to dryness under continuous N₂-stream, solved in 90 µL and 10 µL of the internal standard (2-ethyl-humulene, Sigma-Aldrich, final concentration 0.889 mg mL⁻¹) were added. The samples were analyzed by GC–MS.

2.7. In vivo assay for the production of CBT-diols

The four plasmid system was used, comprising pCola-, pCDF-, pET- and pACYC-Duet-1 vector (Supplemental Table S1) with the pACYC-duet-1 vector harboring the codon optimized fusion protein consisting of the membrane-engineered n-CYP450 monoxygenase (Genbank: AF166332.1, n-cyp450, synthesized and codon optimized at Thermo Fischer, GeneArt, Regensburg, Supplemental Fig. S30) and the newly assembled n-CPR (gene: n-cpr, synthesized and codon optimized at Thermo Fischer, GeneArt, Regensburg; Supplemental Fig. S31) combined via a 5'- GSTSGGCAS – 3' linker peptide according to a previous report [10]. The 500 mL shake flask production was performed with M9CA medium instead of LB medium.

2.8. 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 H₂O and centrifuged again (4500 rpm, 25 min). The cells were suspended in 30 mL H₂O and disrupted by sonification (4 × 2 min, 5 × 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 N2 stream. The powder was desorbed to dryness using a rotary evaporator. The samples were analyzed by GC–MS.

2.10. Homology modeling of the open conformation of CBTS and n-CPR

Homology modeling was performed in YASARA Structure [23], based on a PSI-Blast alignment using position specific matrices obtained from the primary alignments. Secondary structure prediction, loop-modeling, and side-chain and hydrogen bond network optimization were performed in presence of explicit solvent molecules. Templates used for the homology modeling were derived from primary and secondary structure alignments from the HHpred server with default settings [25]. (45)-limonene synthase from Mentha spicata (PDB-ID: 2ong chain A; [26]; (+)- bornyl diphosphate synthase from Salvia officinalis (PDB-ID: 1m1b chain A; [27]) 5-epi-aristolochene synthase from N. tabacum (PDB-ID: 3m00, chain A; [28]) abietaadiene synthase from Abies grandis (PDB-ID: 359v, chain A; [29]) taxadiene synthase from T. brevifolia (PDB-ID: 3p3p, chain A; [30]) were employed as templates for open conformation of CBTS. NADPH-cytochrome P450 reductase from Homo sapiens (PDB-ID: 3g2e, chain A; [31]) and nitric-oxide synthase reductase from Rattus norvegicus (PDB-ID: 1tl1, chain A; [32]) were employed as templates for modelling of n-CPR.

2.11. Homology modeling of the closed conformation of CBTS

A molecular model for the truncated CBTS (GenBank accession number: AAS46038.1) lacking the predicted amino-terminal transit sequence (N-terminal amino acids 1–56) was built in YASARA Structure [33] using the same templates as for the wild type open CBTS conformation. The refined model was structurally aligned with the closed conformation of TXS. The structure of the unfolded A-C- and J-K loops, as well as the CBTS N-terminus were replaced by their corresponding secondary structure elements from TXS [30], followed by in silico mutation to obtained the corresponding CBTS residues. The structure was minimized for 10,000 steps, followed by a 10 ns MD simulation using the simulation setup and parameters described above. A second structural alignment between the closed conformation of TXS and the minimized CBTS model were generated, and the productive GGPP and internal water molecules of TXS were integrated into the CBTS complex. A second 10,000 steps energy minimization and 10 ns MD simulation were carried out to compute the final structural model of the closed conformation of CBTS.

2.12. Quantum chemical calculations and QM/MM MD simulations

The CBT-ol formation process was studied using in 1) a small QM system of the isolated ligand, with 53 atoms, 2) a cluster model consisting of 280 atoms and 3) a hybrid QM/MM model based on the homology modeled and MD-relaxed structures of CBTS. The small QM system consisted of the geranylgeranyl cation and the conversion of this reactant to cembranyl-4-cation modelled at the B3LYP-D3/def2-TZVP//B3LYP-D3/def2-SVP level of theory in gas phase. Corresponding structures, including the reactant, intermediates, transition states (TS) and the product, were characterized by frequency calculations. For the QM cluster model calculations, residues around 4 Å from the ligand were included. The model comprised the: ligand (geranylgeranyl cation), OPP⁺, 3 Mg²⁺, 9 H₂O, and protein residues R316, R492, R506, D352, D495, D499, E503, S452, A498, LS71, LS66, C494, M491, TS77,
V344, resulting in a system with 281 QM atoms. Amino acids were cut at the C\text{\textgamma}l-positions to provide a computationally feasible sized model system, and fixed during the structure optimization at BBP6-D3 level [34,35], by employing the multipole accelerated resolution of identity approximation (RI-MARIJ) [36], and treating the environmental dielectric with \( \varepsilon = 4 \) using the conductor-like screening model (COSMO) [37]. Single point energies for the protein model are reported at the B3LYP/def2-TZVP level. The isolated ligand was studied at the B3LYP-D3/TZVP/B3LYP/def2-SVP level of theory [38,39]. Previous studies have also employed other functional, such as M06-2X. We however expect the employed theory level provides a balanced and accurate for both the gas phase and protein homology models, which are also subjected to approximation. Entropic and zero-point energy corrections were evaluated from the Hessian matrix. For the QM/MM MD calculations, the CBTS model was first classically relaxed in a water box for 10 ns at \( T = 310 \text{K} \) and 1 bar pressure and using an integration time step of 1 fs, an NPT ensemble, and the CHARMM27 force field [40]. The ligand was treated using electrostatic potential charges calculated at B3LYP/def2-TZVP level and Lennard-Jones parameters derived from the CHARMM27 force field. The ligand was kept fixed during the relaxation. The electrostatics were treated using the particle mesh Ewald (PME) approach with a grid spacing of 1 Å. For further QM/MM simulations, the protein model was trimmed to include a sphere 20 Å around the ligand, with everything beyond 15 Å away from the ligand was kept fixed. The QM/MM model comprised about 5200 atoms, with the ligand described at B3LYP-D3/def2-SVP level of theory. All quantum chemical calculations were performed using TURBOMOLE [41], the QM/MM calculations were performed using the CHARMM/TURBOMOLE implementation [42], and the classical relaxation using NAMD2 [43].

3. Results

3.1. Defining the enzyme activity of a putative \textit{N. tabacum} CBTS by recombinant expression in an engineered \textit{E. coli} host

The \textit{N. tabacum} derived CBTS gene (GenBank: AAS46038.1) was introduced into a previously reported three-plasmid based diterpenoid production system optimized for \textit{E. coli} [8]. Typically, plant class I dTPS contain an N-terminal chloroplast leader sequence. To date, the CBTS protein has not been structurally characterized, hence the complete gene sequence encoding for a 589 amino acid protein was introduced into the diterpenoid producing strain. Three day shake flask cultures provided four unidentified diterpenoid compounds detected via gas chromatography/mass spectrometry (GC/MS) analysis (Fig. 2 and Supplemental Fig. S1). After sequential column purifications, three of the four diterpenoid compounds have subsequently been characterized by nuclear magnetic resonance (NMR) spectroscopy. The two most abundant compounds have subsequently been characterized by nuclear magnetic resonance (NMR) spectroscopy. The two most abundant compounds have subsequently been characterized by nuclear magnetic resonance (NMR) spectroscopy. The two most abundant compounds have subsequently been characterized by nuclear magnetic resonance (NMR) spectroscopy. The two most abundant compounds have subsequently been characterized by nuclear magnetic resonance (NMR) spectroscopy. The two most abundant compounds have subsequently been characterized by nuclear magnetic resonance (NMR) spectroscopy. The two most abundant compounds have subsequently been characterized by nuclear magnetic resonance (NMR) spectroscopy. The two most abundant compounds have subsequently been characterized by nuclear magnetic resonance (NMR) spectroscopy.

The fourth compound that could not be separated from (+)-cembrene upon purification was assigned according to its mass spectrum as (±)-isocembrene (Supplemental Fig. S1b) [47]. The cumulative evidence confirms CBTS to be an active CBT-ol synthase.

To optimize CBT-ol production, the fermentation was scaled to 5 L in a controlled, stirred tank bioreactor. The refined five day batch fermentation (Supplemental Fig. S9) provided final concentrations of 120 mg/L CBT-ol (±)-CBT-ol: 54 mg/L, β-CBT-ol: 66 mg/L, 6.23 mg/L (+)-cembrene, 3.74 mg/L (±)-isocembrene and 1.88 mg/L (−)-casbene referenced against α-humulene as an internal standard (Supplemental Fig. S10).

3.2. Structural and mechanistic characterization of the CBTS from \textit{N. tabacum}

Initially, we intended to gain structural information of CBTS by crystallization and X-ray crystallography. Since we failed in obtaining crystals of CBTS, we constructed a CBTS homology model from \textit{N. tabacum} based on its reported primary sequence and the five X-ray structures of class I TPSs in order to structurally and mechanistically characterize CBTS. The resulting model displays the typical α-barrel fold of mono-functional plant class I terpene synthases, comprising both the signature α class I and the β class I domains (Supplemental Fig. S11). This structure represents the open, catalytically inactive conformation together with an unfolded N-terminal peptide. Comparison of the CBTS structure against taxadiene synthase (TXS), a structurally well characterized class I plant terpene synthases model, indicated that CBTS lacks the presence of a chloroplast leader sequence and a transmembrane region. This result is consistent with recombinant expression of the full length CBTS in the pET28b-vector, showing a good expression rate for CBTS in \textit{E. coli} (Supplemental Fig. S12). The unfolded N-terminal region in CBTS comprises the first 56 amino acids, whereas TXS exhibits an N-terminal unfolded peptide that consists of 79 residues. However, TXS does not seem to rely on these 79 residues, as a genetic truncation retains catalytic activity [30].

Hence, for modeling of the closed CBTS complex, we excluded the first 56 N-terminal residues. To ensure that the generated Δ56-CBTS model still represents a soluble and active form of the enzyme, we expressed the truncated protein in \textit{E. coli} and probed its activity by in vitro addition of GGPP. Both native and Δ56-CBTS proteins show identical expression levels (Supplemental Fig. S12), in vitro activities and product profiles.
Subsequently, we constructed the CBTS closed complex model as described previously [8]. The structural alignment of the α class I domain of the closed complex of A79-TXS with that of Δ56-CBTS exhibits a root mean square deviation (RMSD) of 0.31 Å over 321 residues with 27% primary sequence identity (Supplemental Fig. S11). The TXS/CBTS superposition suggests, that the two CBTS magnesium binding motifs are spatially located on α-helical secondary structure elements, that correspond to those reported for TXS [30]. Using the closed TXS complex containing GGPP in its productive conformation as a template [8], we constructed a putative closed complex of CBTS with GGPP coordinated by the three Mg2+ ions and corresponding water molecules. Water molecules and amino acid mediated bonding of productive GGPP resembles that of the productive closed conformation of TXS-GGPP template. Interestingly, the closed CBTS-GGPP complex exhibits a similar hydrogen-bonding donor network as observed in the TXS-GGPP template and other pro- and eukaryotic class I terpene synthases [6]. Specifically, this hydrogen bonding network is characterized by three arginine residues R316, R492 and R506. In TXS, this conserved triade, stabilizes PPi during the catalytic cascade. More importantly, the CBTS closed complex lacks the tyrosine residues, Y89 and Y835, which in closed TXS-GGPP model are important for binding of a water molecule adjacent to the PPi moiety upon active site closure. Instead, CBTS harbors two non-polar residues (F65, L571) in the analogous positions. These residues are chemically not able to coordinate water molecules, and may thus shield the active site. Furthermore, in contrast to TXS, where NT57 coordinates the terminal oxygen-atom of PPi, via a hydrogen bond, the analogous position in CBTS is occupied by an aspartate residue (D495) that does not form a similar coordinating bond. Instead, our obtained CBTS closed complex structure suggests that D495 might coordinate a water molecule, which originates from the bulk solvent upon active site closure (Fig. 3B). Interestingly, this water molecule is in much closer contact to the carbocation intermediate in comparison to the “protected” TXS-active site. Therefore, the water molecule in CBTS may experience enhanced mobility, as the coordination with Y835 in TXS is not present. The enhanced water mobility in the CBTS active site and its close proximity to the carbocation intermediate may facilitate premature carbocation quenching that could facilitate the formation of monocyclic cembranes as the main product. This hypothesis was examined using in silico guided CBTS protein engineering.

Initially, we examined whether the identified CBTS hydrogen-bonding network displays a similar function as in TXS. Therefore, we substituted the amino acid network surrounding R316, R492 and R506 with aliphatic, aromatic, and other polar residues (Supplemental Table S3). Additionally, we substituted D495 and S452, which may be important for water coordination within the active site, to other small aliphatic and polar residues. Previously, we reported that in the TXS closed complex carbocation intermediates are stabilized by π-π and cation-π interaction [8], which coordinates the reaction cascade and determines the specific product outcome. In analogy, we replaced all residues with aromatic character, which decorate the CBTS active site, with aliphatic, polar and other aromatic residues and tested their reactivity in vitro. As expected, single substitutions of the three essential arginine residues (R316, R492 and R506) resulted in inactive CBTS variants (Supplemental Table S10), consistent with the results for TXS [8]. Moreover, substitution of the aromatic CBTS residues W323, F570 and V577 by aliphatic and non-aromatic polar residues resulted in inactive variants (Supplemental Table S10). By contrast, the amino acid substitutions W323Y, F570V and V577F resulted in CBTS variants with similar activity and product profiles as the wild type enzyme. The CBTS-GGPP closed complex model suggests that residues C491, L566 and V344 are in close proximity to the substrate and the corresponding carbocation intermediates generated during catalysis (Supplemental Fig S13). Although expression rates of CBTS variants L566H/W/A/S, V344K/M/L/S, C491A/S/Y, S452A/T/D and D495T/A/E were similar to that of the wild type enzyme, these amino acid substitutions resulted in inactive enzyme variants. Notably, only the CBTS C491L mutant showed a similar product profile as the wild type enzyme, while the D495N substitution does result in a significantly altered product profile were hydroxylated cembranes are absent (Supplemental Table S10). Nonetheless, the generated concentrations of mono- and bi-cyclic hydrocarbons remains comparable to wild type CBTS. This cumulative data strongly suggests that D495 is involved in the control of water-assisted quenching of carbocation intermediates during the catalytic cycle. The in silico model of the D495N mutant shows that the amide group of the asparagine coordinates the terminal γ-oxygen of PPi, via a hydrogen bridge instead of bulk water coordination (Supplemental Fig. S14), which leads to a displacement of the water molecule from the carbocation intermediate, and might then explain the lack of hydroxylated hydrocarbon products observed in this mutant.

To probe possible water-quenching mechanism and the energetics of the discrete cyclization pathway, we performed quantum chemical density functional theory (DFT) and hybrid QM/MM-MD calculations of the isolated substrates and on the closed CBTS model. As a starting model for these calculations, we employed the closed conformation of CBTS harboring the docked geranylgeranyl cation A (Supplemental Table S9). We expect that the ca. 30% overall sequence identity of the homology model provides a sufficiently accurate starting point for exploration of the chemical principles.

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

We first studied the energetics in the gas-phase at the density functional theory level (B3LYP-D3/def2-TZVP//B3LYP-D3/def2-SVP), using the isolated ligand, with 53 atoms, as a model system. We find that the conversion of the geranylgeranyl cation A (Fig. 4A and B) into the first cyclic intermediate B is exothermic by −10.8 kcal mol−1, consistent with the expected stability of a tertiary carbocation, but the overall free energy for this step is 0.3 kcal mol−1, since the ring formation has a negative entropic contribution. Hong et al. [24] previously found a reaction energy of −2.9 kcal mol−1 (by including electronic energies and zero point corrections at the MPWB1K/6–31 + G(d,p)//B3LYP/6–31 + G(d,p) level), and a barrierless cyclization step, consistent with our QM/MM MD simulations (see below). Our calculations further suggest that formation of the cembra-4-yl cation (C) from cation B could occur via a stepwise mechanism involving the formation of the intermediate BC, via two low-energy transition states, TS1 and TS2, with barriers of 1.5 and 3.5 kcal mol−1, respectively, whereas the concerted route is energetically highly unfavorable by ca. 40 kcal mol−1, and is thus unlikely to occur. Our DFT calculations further suggest that formation of the cation C is exergonic by −6.5 kcal mol−1 in the gas phase (Fig. 4A).

Next we examined the hydride shift reaction, as well as formation of α-CBT-ol (PR) and β-CBT-ol (PS) within the active site of the CBTS. Our model consists of first and second sphere protein surroundings of the active site (about 280 atoms) that were constructed based on the homology models (see Methods, Supplemental Fig. S15A). We optimized these cluster models to obtain the reaction mechanism elements we found in our gas phase calculations. The formation of the cation C is exothermic by −16.2 kcal mol−1, similar to the gas-phase exothermicity of −18.3 kcal mol−1. Moreover, formation of the β-CBT-ol (PS) is about 10 kcal mol−1 more favorable than the α-CBT-ol (PR), in agreement with the experimentally observed product (Fig. 2).
The QM calculations suggest that D495 and S452 could stabilize the reaction product by formation of a hydrogen-bond with a nearby water molecule (Fig. 4D). D495 may thus indeed function as a primary proton acceptor for the water molecule, providing a precursor for the hydroxylation at C4 (see Fig. 1 for the numbering) that precedes the formation of \( \alpha \)-CBT-ol and \( \beta \)-CBT-ol. In addition to this hydrogen-bonded network, \( \beta \)-CBT-ol is also in close contact with the OPP\(^{4-} \), suggesting that formation of the \( \beta \)-isomer is energetically preferred. Although the stabilization of the CBT-ol product is strongly stabilized by the hydrophilic amino acids and the OPP\(^{4-} \) network, the formation of the cyclic carbocation intermediates could be controlled by the hydrophobic cage within the active site (Fig. 3). We therefore probed how the protein environment may control the formation of the first cycloadduct B. To this end, we modeled all possible monocyclic products, which could form via cation-\( \pi \) interactions involved in cation A using the isolated ligand model (Supplemental Table S9). We suggest that formation of a covalent bond between C7 and C2 is somewhat more favorable than the natural product (B) in CBTS (Supplemental Table S9). This alternative product involves a 6-membered ring, which is energetically more stable than a 19-membered cycle (B), but this structure might be sterically hindered from forming in the protein environment (Supplemental Fig. S15B). In order to probe the formation of possible cyclization products in the protein active site, we performed hybrid QM/MM MD simulations starting from compound A. We performed two independent 5 ps, simulations, with the D495 in its deprotonated (D495, Fig. 4C) and protonated (D495\(^+ \); Fig. 4C) states. When D495 is modeled in its protonated state (D495\(^+ \)), we observe a spontaneous formation of B around in about 3 ps, as indicated by the decrease in the C1-C2 distance from 3 Å to...
1.6 Å (Fig. 4C and E). This rapid cyclization process is consistent with the previously suggested barrierless ring formation in gas-phase [24]. We further find that the 6-membered ring or the other possible species are not formed in the protein due to the unique orientation of A within the active site in the homology model (Supplemental Table S9). Conformational changes that are difficult to model could, however, in principles also allow for such side pathways. The active site of the CBTS contains several charged and polar residues, as well as the OPP$^4\text{C}_0$ and its Mg$^{2+}$ counter ion that may favor the deprotonation of water. Considering the proximity and hydration energetics (Fig. 4), D495 may serve as proton acceptor, and providing a site for the further attack of the hydroxyl with the carbocation to form CBT-ol.

3.4. Functional reconstitution of class II P450 hydroxylase through in-silico indentification of a cognate N. tabacum reductase aimed at the biotechnological production of bioactive CBT-diols

Taxa-4,5-diene hydroxylation at C5 has been achieved in E. coli using protein engineering of the transmembrane region of the native taxadiene 5-α-ol hydroxylase (t-P450) and a cognate reductase (t-CPR) from Taxus cuspidate [10]. By contrast, the native class...
II CPR of CBT-diols-forming P450 monooxygenase (GenBank: AAD47832.1, n-P450) is, however, not available. We therefore employed several bioinformatic approaches to search for a suitable CPR in the genome of *N. tabacum* using the t-CPR as a template. This task was complicated as the whole genome sequence of *N. tabacum* is not available. In contrast, *N. tabacum* transcriptome sequence information is in form of expressed sequence tags (ESTs) available. Homology-based searches in the EST database with t-CPR as template provided several hits (Supplementary Fig. S16). However, the identified ESTs did not cover the entire t-CPR sequence space. Instead, only parts of t-CPR could be aligned with differing *N. tabacum* ESTs (Supplementary Fig. S17). According to a previous report, the identified parts resemble the domains that are responsible for FAD, FMN and NAPDH binding [48]. Based on this information, the *N. tabacum* CPR (n-CPR) was assembled from these differing EST hits, since all of four TBLASTN hits constituted overlapping sequences. As expressed sequence tags did not display any transmembrane consensus motifs, the assembled n-CPR does not contain a transmembrane spanning α-helix, which renders this reductase naturally soluble in our *E. coli* production system. However, to establish CBT-diol production in our *E. coli* system, also the n-P450 had to be solubly expressed. This was accomplished by transmembrane engineering of the n-P450, in analogy to the transmembrane engineering of the 5α-ol hydroxylase from *T. cuspidate* (t-P450), which resulted in production of taxadiene-5α-ol [10].

To establish optimal expression conditions in *E. coli*, the assembled CPR (n-CPR) was genetically fused to the n-P450 via a N-terminal n-CPR GSTSSGSGAS-linker peptide, analogous to the genomic construct used for a soluble expression of the *E. coli* CPR (Fig. 1), display a high variety of biological activities including cytotoxicity, neuro-protectivity, antimicrobial and antifungal as well as insecticidal activity. The latter reconstitution of the P450-CRP system required in *silico* identification and assembly of a cognate CRP extracted from fragmented transcriptomic data of *N. tabacum*.

Using a simple *E. coli* production system, we produced non-functionalized and mono-functionalized cembranoid bioactives. Moreover, our study demonstrated that the high titer production of CBT-ols with up to 120 mg L\(^{-1}\) yield using glycerol as the sole carbon source is a good sustainable alternative compared to *in planta* production. Remarkably, we could also demonstrate that CBTS is able to produce five different cembranoid bioactives as major products or as side-products. *In silico* characterization of the *N. tabacum* CBTS by homology modeling and quantum chemical calculations further allowed us to understand the structural basis for three major non-functionalized by-products and a possible basis for the quenching process from GGPP to α- and β-CBT-ol.

Our quantum chemical calculations suggested that although the hydride shift and cyclization reaction of the geranylgeranyl cation have low barriers and are highly favorable in gas-phase, CBTS provides and important surroundings that favor ring closure at the correct C1-C2 position. We could observe spontaneous ring closure in the QM/MM MD simulations on picosecond timescales, particularly after protonation of D495. Moreover, our DFT cluster models further suggested that hydroxylation of cation C leading to formation of β-CBT-ol (−26.9 kcal/mol [10], Fig. 4B, PS) is energetically favored over α-CBT-ol formation (−17.4 kcal/mol [10], Fig. 4B, PS), due to the hydrogen-bonding network formed between D495 and S452, consistent with our NMR experiments and site-directed mutagenesis experiments showing that both residues are crucial for the catalytic activity of CBTS. These results further corroborate, that in absence of a crystal structure CBTS, our *in silico* modeled structures provide important insights into the function of the enzyme. Based on the computational data, we could further design targeted mutagenesis studies to demonstrate that the hydrogen-bonded donor network stabilizes PP\(_i\), similar to its function in TXS. We were not, however, able to enhance production of one or more of the side-products in our targeted protein engineering experiments. Homologous substitutions of the respective amino acids resulted in a similar product profile, whereas non-homologous substituted enzymes showed no GGPP cyclization activity. This, together with the results obtained from our QM/MM MD simulations indicate that the native decorative coating of the active site of CBTS is exclusively designed to produce monocyclic cembranoid structures.

Our combined data indicates that (+)-cembrene, (±)-isocembrene and (−)-casbene could derive from initial GGPP misfolding, imprecise barrier crossing and substrate tumbling during the cyclization cascade in CBTS, in line with structural investigations of TXS [8]. Our closed complex model of CBTS harboring cations B and C suggest that, in analogy to TXS, the bi-functional CBTS-motif depicted by R316-PP, is the prime candidate for the premature deprotonation of the non-hydroxylated cembranoids (Fig. 5). Premature deprotonation by this bi-functional motif at cation B stage is expected to lead to the formation of (−)-casbene (Fig. 1). 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, that lead to the formation of (+)-cembrene or (±)-isocembrene instead of formation of the CBT-ols (Fig. 1).

We could show that the D495N substitution only produced non-hydroxylated cembranoids in a similar distribution compared to the wild type CBTS (Supplementary Fig. S21), supporting the putative role that this residue is catalysis. The closed complex model of the D495N mutant suggested 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 coordina-
tion of a remaining bulk water molecule. Combined with results from our QM calculations, this suggests that the residue is involved in an amino acid-assisted quenching reaction for functionalized cembranoids in CBTS.

In addition to the biotechnological production of non- and mono-functionalized cembranoids, the four plasmid system containing the n-P450-n-CPR redox couple, we were able to produce di-hydroxylated cembranoids, and provides a first report of a biotechnological production route for these bioactive compounds. Application of trans-membrane signal engineering and bioinformatics were crucial to produce α- and β-CBT-diols. Nevertheless, these compounds were produced only in minor amounts. A few examples for diterpenoid hydroxylations using class II P450 systems have been achieved. In all cases, production rates are low compared to the production rate of the non-functionalized precursors. Therefore, the choice of a eukaryotic host in combination with a more detailed trans-membrane 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. Our integrated experimental and computational approach could provide a basis for the rational design of engineered cembranoid enzymes.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was in part supported by the German BMBF Grant SysBioTerp (support code 031A305A) to T.B. and the German BMBF Grant ELP (support code 1340/68351/3/11) to T.B. and P.S. We acknowledge support by the Open Access Publication Fund of the Freie Universität Berlin. We thank Ronja Driller and Max Hirte for support in preparing figures.

Author contributions

T.B. and P.S. designed the study, planned and supervised experiments. P.S. conducted metabolic engineering of E. coli and enzyme mutagenesis experiments. P.S. conducted fermentation experiments, NMR data collection and verified produced structures, and the homology modelling of CBTS. I.U., S.K. and V.R.I.K. performed the quantum chemical calculations. P.S., V.R.I.K., B.L., and T.B. wrote the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2020.06.030.

References

[1] Brück T, Kourist R, Loll B. Production of macrocyclic sesqui- and diterpenes in heterologous microbial hosts: a systems approach to harness nature’s molecular diversity. ChemCatChem 2014;6:1142–65.
[2] Yan N, Du Y, Liu X, Zhang H, Liu Y, Zhang P, et al. Chemical structures, biosynthesis, bioactivities, biocatalysis and semisynthesis of tobacco cembranoids: An overview. Ind Crops Prod 2016;83:66–80.
[3] El Sayed KA, Sylvester PW. Biocatalytic and semisynthetic studies of the anticancer tobacco cembranoids. Expert Opin Invest Drugs 2007;16:877–87.
[4] Mischko W, Hirte M, Fuchs M, Mehmer N, Bruck TB. Identification of sesquiterpene synthases from the Basidimycota Coniophora puteana for the efficient and highly selective beta-copaene and cubebol production in E. coli. Microb Cell Fact 2018;17:164.
[5] Ennadjoua H, Vachon G, Giacalone C, Besse I, Sallaud C, Herzog M, et al. Trichome specific expression of the tobacco (Nicotiana sylvestris) cembratrienoid synthase gene is controlled by both activating and repressing cis-regions. Plant Mol Biol 2010;73:673–85.
[6] Christianson DW. Structural and chemical biology of terpenoid cyclases. Chem Rev 2017;117:11570–648.
[7] Aaron JA, Christianson DW. Trinuclear metal clusters in catalysis by terpenoid synthases. Pure Appl Chem 2010;82:1585–97.
[8] Schreper P, Buettner A, Goerner C, Hertel M, van Rijn J, Wallrpap F, et al. Identification of amino acid networks governing catalysis in the closed complex of class I terpene synthases. Proc Natl Acad Sci USA 2016;113: E958–67.
[9] Guo W. Biosynthesis of cembratrienols in cell-free extracts from trichomes of Nicotiana tabacum. Plant Sci 1995;110:1–10.
[10] Ajikumar PK, Xiao WH, Tyo KE, Wang Y, Simeon F, Leonard E, et al. Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. Science 2010;330:70–3.
[11] Major DT, Freund Y, Weitman M. Catalytic control in terpene cyclases: multiscale modeling of thermodynamic, kinetic, and dynamic effects. Curt Opin Chem Biol 2014;21:25–33.
[12] Freund Y, Anscher T, Major DT. Catalytic control in the facile proton transfer in taxadiene synthase. ACS Catal 2017;7:7853–7.
[13] O’Brien TE, Bertolani SJ, Zhang Y, Siegel JB, Tantillo DJ. Predicting productive binding modes for substrates and carbocation intermediates in terpene synthases-borneyl diphosphate synthase as a representative case. ACS Catal 2018;8:3322–30.
[14] Major DT. Electrostatic control of chemistry in terpene cyclases. ACS Catal 2017;7:5461–5.
[15] Shomaker SR, Zhang Y, O’Brien T, Tantillo D, Siegel JB. Hybrid quantum mechanics/molecular modeling mechanistic study of a terpene synthase. Biophys J 2020;118:304a-a.
[16] van Rijn JPM, Escorcia AM, Thiel W. QM/MM study of the taxadiene synthase mechanism. J Comput Chem 2019;40:1902–10.
