Technische Universität München

Fakultät für Chemie

Mechanistic studies on the active site interactions in terpene synthases

Marion Aloisia Ringel

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

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender		Prof. Dr. Tom Nilges
Prüfer der Dissertation	1.	Prof. Dr. Thomas Brück
	2.	apl. Prof. Dr. Wolfgang Eisenreich
	3.	Prof. Dr. Uwe Bornscheuer

Die Dissertation wurde am 10.05.2022 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 25.07.2022 angenommen.

It is not our abilities that show what we truly are. It is our choices.

- Albus Percival Wulfric Brian Dumbledore -

- J. K. Rowling - Harry Potter and the Philosopher's Stone -

I Abstracts

Chapter I

In the first chapter, we present a sustainable, coral-independent, biosynthetic route towards pseudopterosin-bioactives. The *Streptomyces clavuligerus* derived hydropyrene diterpene synthase (HpS) was implemented into an *E. coli* host for effective, heterologous production of its main product hydropyrene, side product Hydropyrenol, and the pseudopterosin precursors isoelisabethatriene A and B. *In-silico* guided, site-directed mutagenesis based on the diterpene synthase CotB2 from *Streptomyces melanosporofaciens* altered the product portfolio of HpS towards isoelisabethatriene A as new main product. Notably, mutagenesis studies revealed the presence of catalytically important methionines which represents the first experimental evidence of the direct involvement of methionines in terpenoid catalysis. Consequently, the production and purification of isoelisabethatriene A and B allowed for lipase B (*Candida antarctica*) mediated, selective oxyfunctionalisation towards the key pseudopterosin intermediate erogorgiaene and 1*R*-epoxy-5,14-elisabethadiene, respectively. The availability of the anti-tubercular bioactive erogorgiaene thus paves the way towards a sustainable and scalable pseudopterosin production, while 1*R*-epoxy-5,14-elisabethadiene might potentially serve as new scaffold for pharmaceutical development.

Chapter II

In the second chapter, we were able to identify two new δ-cadinol synthases from the wood-rotting fungus *Coniophora puteana*. Both enzymes, Copu5 and Copu9 were heterologously expressed in an engineered *E. coli* host, exhibit a high product selectivity and outstanding product titers with 225 mg L^{-1} and 395 mg L^{-1}, respectively. Remarkably, Copu5 and 9 exhibit an identical product spectrum, thus representing the first two class I sesquiterpene synthases from *Basidiomycota* with identical product portfolios. The generation of a high-resolution crystal structure of Copu9 and subsequent model generation of Copu5 led to an *in-silico* guided mutagenesis study targeting identical active site residues in order to shift their respective product profile in a product orientated manner. Consequently, mutagenesis studies of both synthases revealed, that active site residues have an influence on controlling the carbocation reaction trajectory. However, Copu5 and 9 exhibit a difference in structural robustness towards conformational changes suggesting a fundamental importance of tertiary structure elements for terpene synthases' catalysis. Copu5 and Copu9 are thus promising model systems for further mechanistic studies on catalysis in terpene synthases

II Zusammenfassungen

Kapitel I

Das erste Kapitel befasst sich mit der Darstellung eines nachhaltigen, korallen-unabhängigen Biosynthesewegs für bioaktive Pseudopterosinderivate. Mit Hilfe der Hydropyrene Synthase (HpS) aus *Streptomyces clavuligerus* konnte das Hauptprodukt Hydropyrene, das Nebenprodukt Hydropyrenol und die beiden Pseudopterosin-Vorläufer Isoelisabethatriene A und B heterolog in *E. coli* produziert werden. Das Produktspektrum der HpS wurde anschließend durch *in-silico* gestützte, gerichtete Mutagenese basierend auf der Diterpenesynthase CotB2 aus *Streptomyces melanosporofaciens* in Richtung eines neuen Hauptprodukts, Isoelisabethatriene A, verändert. Hervorzuheben ist, dass katalytisch aktive Methionine mit Hilfe der Mutagenese der HpS identifiziert werden konnten. Dies stellt den ersten experimentellen Nachweis von katalytisch aktiven Methioninen in Kontext der Katalyse von Terpensynthasen dar. Die Produktion und Aufreinigung der beiden Pseudopterosin-Vorläufer Isoelisabethatriene A und B ermöglichte zudem die selektive Oxifunktionalisierung durch die Lipase B aus *Candida antarctica*. Dabei konnte zum einen Erogorgiaene und zum anderen 1*R*-Epoxi-5,14-elisabethadiene als Produkte der Lipase Reaktion identifiziert werden. Die Verfügbarkeit des antituberkular wirksamen Erogorgiaenes ebnet somit den Weg zu einer nachhaltigen und skalierbaren Pseudopterosin-Produktion, während 1*R*-Epoxi-5,14-elisabethadiene möglicherweise als neues Target für die pharmazeutische Entwicklung dienen könnte.

Kapitel II

Das zweite Kapitel befasst sich mit der Identifizierung von zwei neuen δ-Cadinol Synthasen aus dem holzzersetzenden Pilz *Coniophora puteana*. Beide Enzyme, Copu5 und Copu9, wurden heterolog in einem gentechnisch veränderten *E. coli*-Wirt exprimiert und weisen eine hohe Produktselektivität und hervorragende Produkttiter (225 mg L^{-1} bzw. 395 mg L^{-1}) auf. Interessanterweise zeigen Copu5 und Copu9 ein identisches Produktspektrum und stellen damit die ersten beiden Sesquiterpene-Synthasen aus einem Basidiomyceten mit identischem Produktportfolio dar. Zur genaueren Untersuchung des molekularen Ursprungs der identischen Produktspezifität der beiden Enzyme wurde eine *in-silico* gestützte, gerichtete Mutagenesestudie durchgeführt. Hierbei wurden, zwischen Copu5 und 9 identische, Aminosäurereste ausgetauscht, mit dem Ziel das Produktportfolio gezielt zu verändern. Dazu wurde jedoch zunächst die Kristallstruktur von Copu9 gelöst und darauf aufbauend ein Homologie Modell von Copu5 erstellt. Die Untersuchungen zeigten, dass die Reste der aktiven Tasche einen Einfluss auf die Lenkung der Umlagerungsreaktionen des Carbokations

während der Katalyse besitzen. Allerdings konnte beobachtet werden, dass sich Copu5 und Copu9 hinsichtlich ihrer strukturellen Robustheit gegenüber Konformationsänderungen innerhalb der Tertiärstruktur stark unterscheiden. Dies weist auf eine grundlegende Bedeutung von Tertiärstrukturelementen für die Katalyse von Terpensynthasen hin. Entsprechend könnten sich Copu5 und Copu9 als vielversprechende Modellsysteme für weiterführende, mechanistische Studien zur Katalyse von Terpensynthasen erweisen.

III Acknowledgements

An erster Stelle möchte ich meinem Doktorvater **Prof. Dr. Thomas Brück** meinen tief empfundenen Dank für die Möglichkeit der Promotion am Werner Siemens-Lehrstuhl für synthetische Biotechnologie aussprechen! Erst seine stete Förderung, sein wissenschaftlicher Rat, die zahlreichen ergebnisoffenen Diskussionen über meine Arbeit, die Freiheit eigene wissenschaftliche Projekte anstoßen zu dürfen aber auch das Vertrauen in mich, im ersten Augenblick herausfordernd wirkende, Aufgaben zu meistern, ermöglichten es mir mein Potential zu entfalten und mich sowohl fachlich als auch persönlich weiterzuentwickeln.

Weiterhin gilt mein Dank **Prof. Dr. Tom Nilges**, **Prof. Dr. Wolfgang Eisenreich** und **Prof. Dr. Uwe Bornscheuer** für die Bewertung und Prüfung meiner Dissertation.

Getreu dem Motto ‚Alleine ist man stark, gemeinsam unschlagbar' wäre das Abenteuer Promotion niemals ganz allein und ohne die Unterstützung hilfsbereiter Kollegen zu bewältigen gewesen.

Ein besonderer Dank gilt daher **Dr. Norbert Mehlmer**, **Dr. Monika Fuchs** und **Dr. Daniel Garbe**. Vielen Dank für all eure guten Ratschläge und die herausragende fachliche Unterstützung, die großen Anteil am Gelingen meiner Experimente haben.

Tausend Dank an **Martina Haack** für ihre unbezahlbare Expertise in jedweder analytischer Fragestellung. Ohne dich hätte so manche Frage keine Antwort gefunden!

Ebenso geht ein großes Dankeschön an meine Kollegen aus anderen Arbeitskreisen, die immer mit Rat und Tat zur Seite standen! Hierbei seien besonders **Dr. Claudia Huber** für ihre Unterstützung in der NMR Strukturaufklärung, sowie **Nicole Dimos** und **Dr. Bernhard Loll** für ihren unermüdlichen Einsatz bei der Auflösung diverser Proteinkristallstrukturen genannt.

An meine Kollegen und Freunde **Zora, Elias, Jan, Felix, Tobi, Sam, Dania, Max, Wolfgang, Michi, Nate, Kevin, Niko, Manfred, Nadim, Philipp, Selina, Linda, Melania, Pariya, Sophia, Mahmoud, Freddy, Petra und Gülnaz** – vielen Dank für euren Rat, eure Unterstützung und den besonderen Teamgeist in unserer Arbeitsgruppe! Ihr habt die letzten Jahre sowohl innerhalb als auch außerhalb des Labors zu etwas besonderem gemacht!

Größter Dank gebührt meiner Familie und meinen Freunden! Insbesondere geht dieser Dank an euch - **Oma**, **Mama**, **Papa**, **Evi** und **Daniel** - ohne eure bedingungslose Unterstützung in jeder Lebenslage und euer Vertrauen in mich wäre ich nicht die, die ich heute bin! Danke schön!

Tom – es gibt keine Wort dafür, wie dankbar ich bin dich in meinem Leben zu haben! Ohne dich geht es nicht!

IV Contents

V List of Abbreviations

°C	degree Celsius
µM	micro molar
ATP	adenosine triphosphate
BPPS	bornyl diphosphate synthase
CDPME	4-diphosphocytidyl-2-C-methyl-D-erythritol
CoA	coenzyme A
CoV	corona virus
DCW	dry cell weight
ddH_2O	double distilled water
DMAPP	dimethylallyl diphosphate
DXP	1-deoxy-D-xylulose-5-phosphate
DXS	DXP synthase
E. coli	*Escherichia coli*
EtOAc	ethyl acetate
EtOH	ethanol
FGGPP	2-fluoro geranylgeranyl diphosphate
FID	flame ionization detector
FPP	farnesyl diphosphate
GC	gas chromatography
GPP	geranyl diphosphate
GGPP	geranylgeranyl diphosphate
HMG-CoA	hydroxy-3-methylglutaryl-CoA
HpS	hydropyrene synthase
HP	hydropyrene
HPol	hydropyrenol
IDI	IPP isomerase
IE	isoelisabethatriene
IES	isoelisabethatriene synthase
IPP	isopentenyl diphosphate

MEP	2-C-methyl-D-erythritol 4-phosphate pathway
mM	millimolar
MS	mass spectrometry
MVA	mevalonate pathway
PCR	polymerase chain reaction
PP	diphosphate moiety
RBS	ribosomal binding site
S. cerevisiae	*Saccharomyces cerevisiae*
SdS	selina-3,7(11)-diene synthase
TAE	tris base, acetic acid and ethylenediaminetetraacetic acid buffer
TDS	trichodiene synthase
TPS	terpene synthase
vvm	volume of medium per min
w/v	weight per volume

1. Introduction

Climate change and infectious diseases – future challenges

The increasing world population and economic growth in uprising industrial nations foster the surge in carbon dioxide emissions thereby supporting the anthropogenic climate change.[1,2] While the effects of climate change on non-human systems[3–5] are widely discussed in the world's leading political committees, less attention has been paid to the interconnection of environmental effects of climate change and their direct impacts on human health.[6] Extreme weather events, environmental degradation and sea-level rise are causing either direct, short-term risks on human health by injury or death from floods, storms and bushfires or show adverse effects on the long run such as desolation of arable land, thereby ultimately leading to shortages in food supply.[6] Furthermore, the environmental effects of climate change foster changes in terrestrial as well as marine ecosystems that lead to an emergence of new infectious diseases.[7]

To that end, the emergence of SARS-CoV-2 originating from South China is the most prominent recent example. For instance, the appearance of (bat-borne) zoonoses like SARS-CoV-2 correlates with the biodiversity of the host organisms which itself is affected by climate conditions.[8] Changing climate conditions can lead to varying geographic distributions of species thus introducing new species with potential pathogens into new areas. As a result the variation of species composition in certain ecological niches facilitates novel host-pathogen interactions thus fostering the evolution of new diseases.[8–10] In particular in biodiversity hot spot areas like South-East-Asia, Central Africa or the Amazonas basin, these effects play an important role in the evolution of new pathogens and facilitate zoonotic spill overs of infectious diseases.[7,8] Moreover, the numbers of vector-borne diseases such as yellow-fever, dengue or malaria previously mainly prevalent in subtropical areas or other diseases like tick-borne encephalitis and Lyme borreliosis are on the rise in regions of moderate climate due to climate change effects.[6,11,12]

In any case, the battle against climate change demands a holistic solution that encompasses all sectors of society. However, the emergence of new infectious diseases as an effect of climate change urges the scientific community to develop new systemic anti-infective, anti-viral and anti-inflammatory drugs as first-line treatment.[13] To this end natural products are a tremendous and to date largely unexploited source of new pharmaceuticals encompassing various biological activities ranging from anti-infective, anti-viral, anti-inflammatory to anti-cancer or insect-repellent and beyond.[14–17]

Terpenoid bioactives

Terpenoids are the largest and structurally most diverse group of natural products. To date this natural product family encompasses over 80 000 characterised compounds including steroids and carotenoids.[18,19] Terpenes are abundant throughout all domains of life as primary or secondary metabolites thereby playing a significant role in biological processes.[20] Some of the numerous functions of terpenoid compounds are essential for prokaryotic organisms such as involvement in cell-wall biosynthesis, membrane function, electron transport or conversion of light into chemical energy.[21] Interestingly, terpenoid compounds seem to be involved in biological processes since the very beginning of life on earth as shown by fossilized terpenes that were found in geological deposits by this means suggesting a fundamental importance for this group of natural products.[22]

While prokaryotic organisms encompass only a small variety of terpenoid secondary metabolites mainly functioning as odour constituents, pigments or antibiotics, eukaryotic systems (e.g., fungi, liverworts or plants) have evolved a huge array of terpenes with different biological roles.[21] In their original host organisms, these secondary metabolites comprise various functions such as phytohormones, pheromones, repellents, protection against abiotic and biotic stress, antifeedants or attractants.[15,21,23] Apart from their function *in vivo*, terpenoids show potent pharmaceutical activities ranging from anti-viral, anti-infective and anti-inflammatory to anti-cancer and beyond.[14–17] These bioactivities render terpenoid natural products as promising candidates for innovative, new drug leads effective against known diseases or as potential cure for challenges along the road. In fact, over the past 20 years more than 50% of the newly developed and registered small-molecule drug leads are either directly originating from natural products or (semi-) synthetic derivatives thereof.[24,25] In addition to the pharmaceutical applications, terpenoids show benefits that can be used in the cosmetic industry (e.g., as insect-repellent or skin care products), the agricultural sector (e.g., as crop protection) or as bulk chemicals and fuel additives.[20]

There are multiple examples of successfully used terpenoid bioactives for pharmaceutical applications. A blockbuster in clinically relevant terpenoids definitely are artemisinin and its derivatives. Artemisinin, which is originally extracted from *Artemisia annua*, is a potent anti-malarial drug lead with the key terpenoid intermediate amorpha-4,11-diene which is subsequently converted to artemisinin.[26,27] To date the main supply route for artemisinin is the biotechnological production of amorpha-4,11-diene in *Escherichia coli* with titers up to 27.4 g L^{-1} [21] and subsequent (semi-) biosynthetic conversion to artemisinin.[28] Another well-known and already established drug is taxol. Originally found in the bark of the pacific yew

tree *Taxus brevifolia*, taxol is a potent anti-cancer drug and has already been issued to clinical trials.[21]

Yet the yield of taxol from its natural source comprises low yields and deforestation of yew tree forests.[21] For instance, also other promising drug candidates such as the novel anti-inflammatory bioactives pseudopterosins from the Caribbean gorgonian coral *Antillogorgia elisabethae* or cyclooctatin from *Streptomyces melanosporofaciens* are hampered from market entry due to low yields and unsustainable extraction processes from their natural sources. The latter, cyclooctatin, exhibits potent anti-inflammatory properties by inhibiting a lysophospholipase during eosinophilic inflammatory reactions.[29] However, pseudopterosins, which belong to the family of diterpene glycosides, comprise potent anti-inflammatory activities exceeding those of their synthetic counterpart indomethacin and thus were issued to clinical trials.[30–33] Unfortunately, clinical trials were stopped in phase II due to insufficient product supply.[34] Nevertheless, pseudopterosins are already commercially used as natural, marine anti-irritants in skincare products and are to date exclusively sourced by unsustainable harvesting and extracting of its natural producer *Antillogorgia elisabethae*.[34,35]

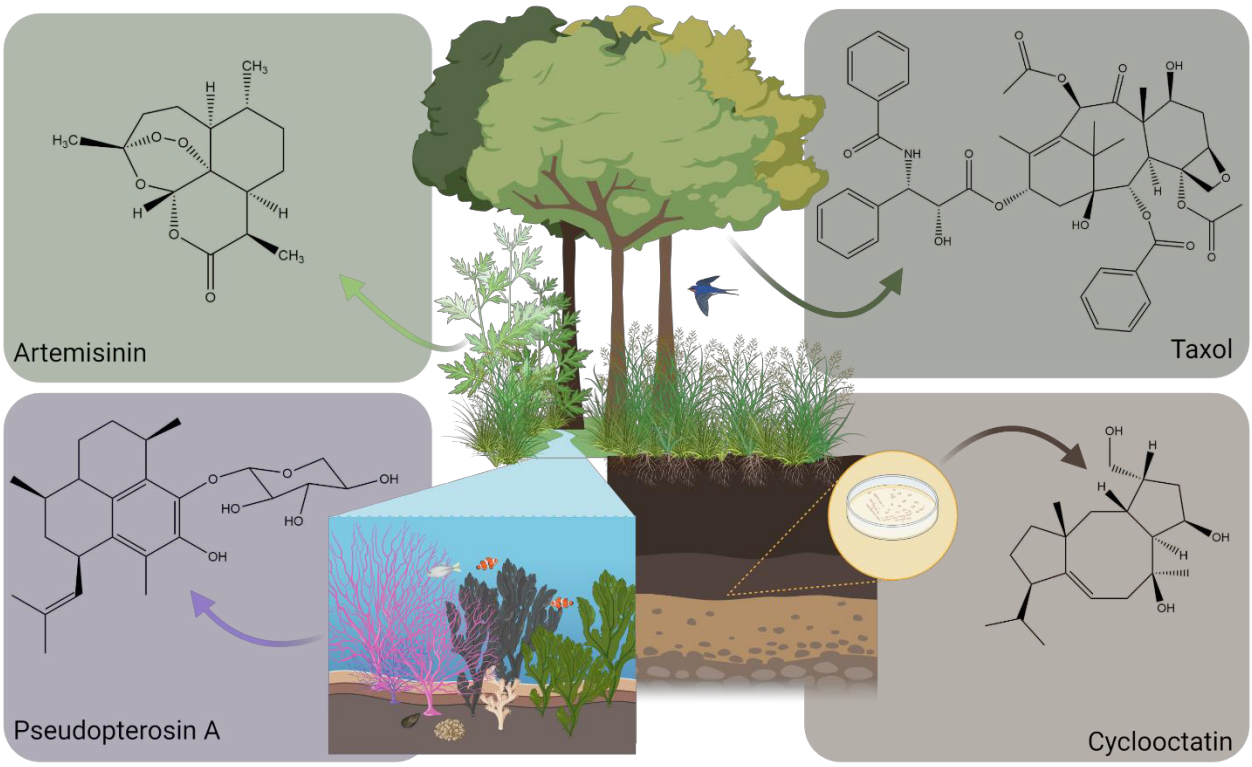


Figure 1: Blockbuster terpenoid bioactives and their natural sources – anti-malarial agent Artemisinin from Artemisia annua[26], cancer therapeutic Taxol from Taxus brevifolia[36], anti-inflammatory bioactive Cyclooctatin from soil bacterium Streptomyces melanosporofaciens[29] and the pseudopterosin group of novel anti-inflammatory agents represented by derivative Pseudopterosin A from the gorgonian coral Antillogorgia elisabethae[30]; Structures drawn with ChemDraw Professional 17.1 and overall figure created with BioRender.

Thus, large scale industrial applications by means of extraction of bioactives from their natural source are neither sustainable nor scalable.[37,38] Alternative chemical synthesis approaches are well established for several terpenoid compounds but usually comprise cost- and workup-intensive metal-organic catalysts and toxic waste streams thus leaving them unsustainable and economically inviable.[20,39–41] Consequently, a sustainable but also economically viable approach towards the production of natural bioactives is of high importance to grant access to sustainably sourced novel drug leads. To this end the biotechnological production of terpenoid bioactives using genetically engineered microbial chassis such as *E. coli* or *Saccharomyces cerevisiae* is a very promising and already well established approach.[21,42–44]

Biosynthesis of terpenes

For development of sustainably sourced, novel drug leads it is crucial to understand the biosynthesis of terpenoid bioactives, within their natural hosts and the opportunities and challenges that might arise during heterologous production in genetically engineered microbial chassis. Generally, all terpenes derive from the universal, five carbon (C_5) building block isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP).[21] Depending on the natural host organism there are two distinct pathways utilising sugar metabolism intermediates for the production of the terpene precursors IPP and DMAPP.[20,45]

First, there is the mevalonate pathway (MVA), which is found in all domains of life, specifically in eucaryotes, archaebacteria, and cytosols of higher plants (c.f. Figure 2).[20,46] Here, the precursor formation starts with the condensation of two acetyl-coenzyme A (CoA) and proceeds via hydroxy-3-methylglutaryl-CoA (HMG-CoA), mevalonate, mevalonate-5-phosphate, mevalonate-5-diphosphate thereby leading to IPP.[20,45,46]

Alternatively, isoprenoid precursors can be synthesised via the 1-deoxy-D-xylulose-5-phosphate pathway (DXP) (also called 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway) which is mainly found in eubacteria, cyanobacteria, green algae, and plastids of higher plants (c.f. Figure 2).[20,46] This pathway commences with pyruvate and D-glyceraldehyde-3-phosphate and leads to IPP via 1-deoxy-D-xylulose-5-phosphate (DXP), 2-C-methyl-D-erythritol-4-phosphate (MEP), 4-diphosphocytidyl-2-C-methyl-D-erythritol (CDPME), 4-diphosphocytidyl-2-C-methyl-D-erythritol 2 phosphate, 2-C-methyl-D-erythritol-2,4-cyclodiphosphate, and 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate.[45–47]

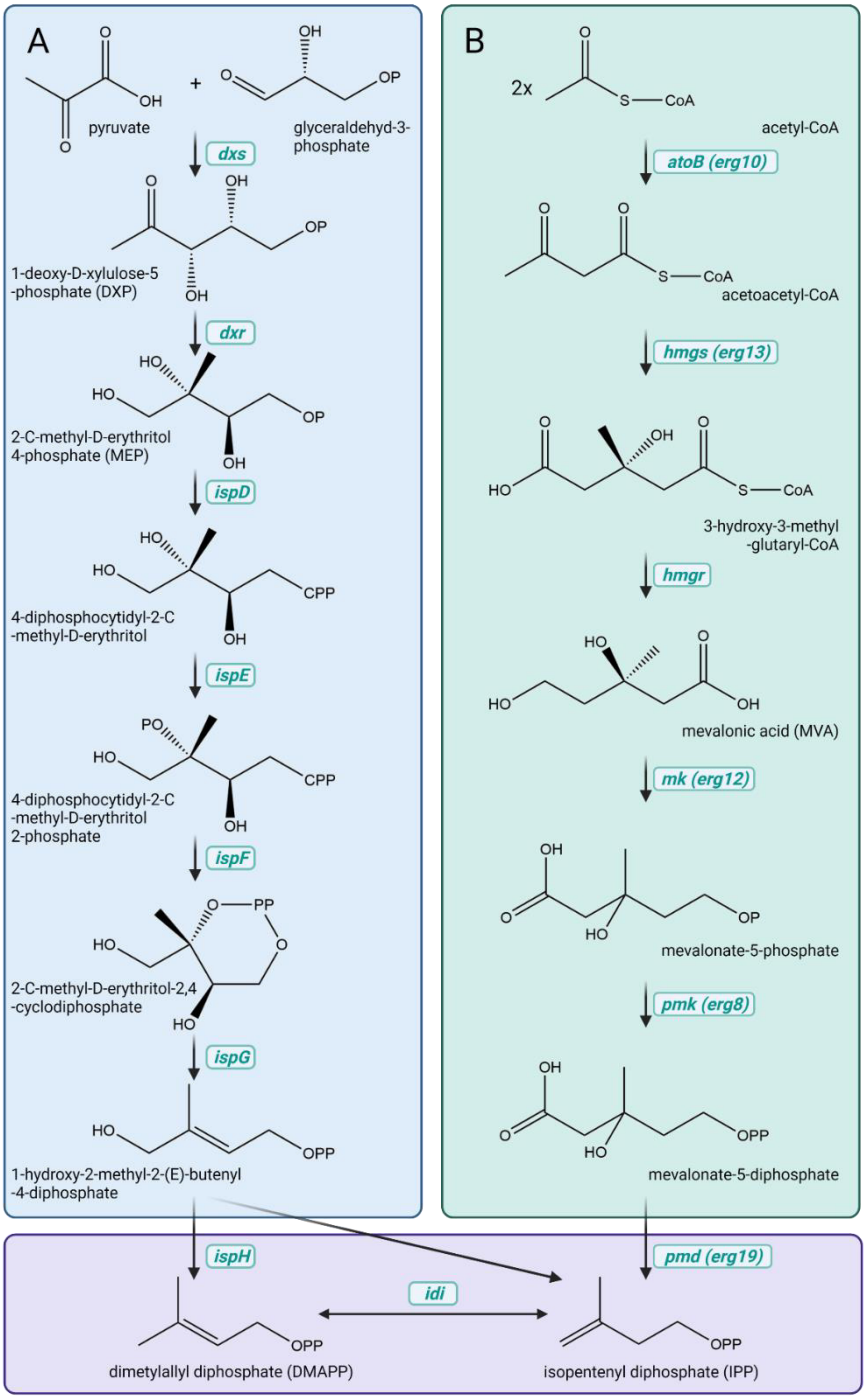


Figure 2: Two biosynthetic pathways for formation of universal terpenoid precursors IPP and DMAPP[45,47]; A: DXP (1-deoxy-D-xylulose-5-phosphate) or also called MEP (2-C-methyl-D-erythritol-4-phosphate) pathway employing seven enzymes for the formation of DMAPP from pyruvate and glyceraldehyd-3-phosphate as initial intermediates; B: MVA (mevalonate) pathway using two molecules of acetyl-CoA as starting point for six enzymes involving pathway yielding IPP; IPP isomerase idi converts IPP and DMAPP to their respective isomer according to the metabolic requirements; Abbreviations: ***dxs*** *– DXP synthase,* ***dxr*** *– DXP reductase,* ***ispD*** *- CDPME synthase,* ***ispE*** *– kinase,* ***ispF*** *- 2-C-methyl-D-erythritol-2,4-diphosphate synthase,* ***ispG*** *- 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate synthase,* ***ispH*** *- 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate reductase,* ***atoB*** *– acetoacetyl-CoA- thiolase,* ***hmgs*** *– hydroxymethylglutaryl-CoA (HMG-CoA) synthase,* ***hmgr*** *– HMG-CoA reductase,* ***mk*** *– mevalonate kinase,* ***pmk*** *– phosphomevalonate kinase,* ***pmd*** *– phosphomevalonate decarboxylase and* ***idi*** *– IPP isomerase[47]; Structures drawn with ChemDraw Professional 17.1 and overall figure created with BioRender.*

Both pathways have previously been successfully implemented in various metabolically engineered microbial hosts such as *E. coli*, *S. cerevisiae* or *Yarrowia lipolytica* thereby allowing for enhanced precursor supply for biotechnological terpene production.[48–52] Studies on the improved production of diterpenes in *E. coli* showed that specifically the DXP synthase DXS and the IPP isomerase IDI are important bottleneck enzymes for improving the metabolic flux towards terpenoid production in an *E. coli* host.[43,53] Furthermore, terpenes are characterized by the number of C_5 isoprene subunits involved in the biosynthesis of their linear terpene precursors thus distinguishing between hemi- (C_5), mono- (C_{10}), sesqui- (C_{15}), di- (C_{20}), tri- (C_{30}), tetra- (C_{40}), and polyterpenes $(C_5)_n$ (c.f. Figure 3).[23] To this end, *all-trans* prenyltransferases catalyse the head to tail 1'-4 condensation of one to three units of IPP and one DMAPP to form the direct, linear precursors geranyl diphosphate (GPP, C_{10}), farnesyl diphosphate (FPP, C_{15}), and geranylgeranyl diphosphate (GGPP, C_{20}).[21,54] Isoprenyl chains exceeding the length of C_{20}, such as $C_{30,}$ or C_{40} linear terpene precursors are subsequently formed by the 1'-1 head-to-head condensation from either two FPP or two GGPP molecules.[21] Interestingly, while terpene precursors are only formed by *all-trans* condensations, so-called regular isoprenoid coupling reactions, also irregular couplings, including *all-cis* condensations occur in nature.[19,54] One prominent example for an *all-cis* terpenoid is natural rubber.[20,54] As in the course of this study only sesquiterpenes and diterpenes originating from *all-trans* condensations are discussed, the complexity of irregular coupling events, which are described thoroughly by Christianson[19], will not be presented in detail.

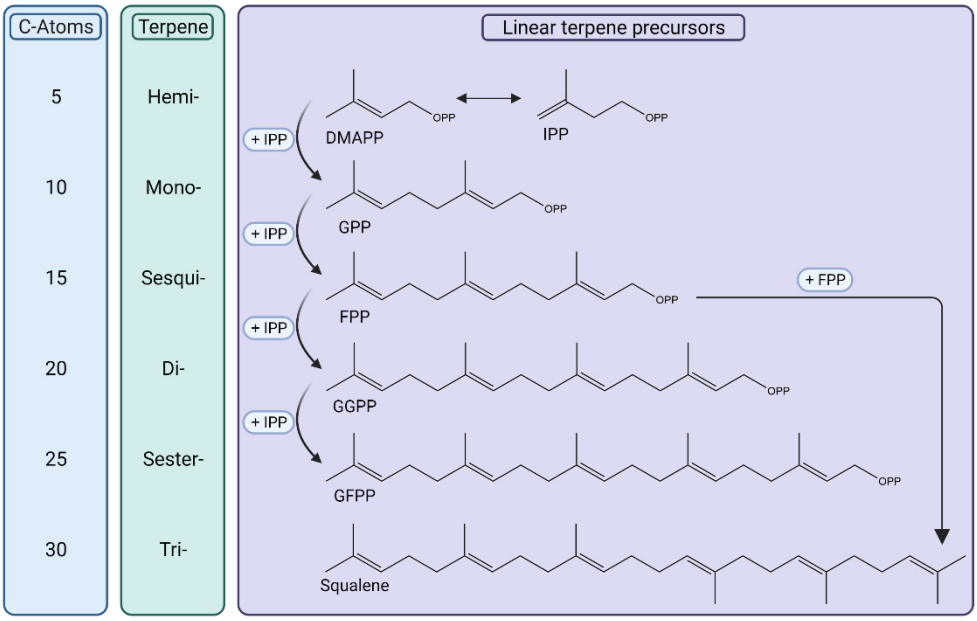


Figure 3: Overview over linear terpene precursors and their nomenclature based on the respective number of C_5 units which are joint by either 1'-4 head to tail condensations (GPP, FPP, GGPP & GFPP) or 1'-1 head to head condensation in case of triterpenoids.[19,20] Structures drawn with ChemDraw Professional 17.1 and overall figure created with BioRender.

Terpene synthases – carbocation chemistry wizards

The vast structural diversity of the terpenoid natural product family, which is solely based on their aliphatic, linear precursors i.e., GPP, FPP, and GGPP, is subsequently created by the terpene synthase (TPS) enzyme family. Generally, terpene synthases can be divided into two distinct classes distinguished from each other by their respective cyclisation initiation reaction. While class I synthases initiate the cyclisation of the aliphatic, linear terpene precursors by abstraction and subsequent ionization yielding the initial carbocation, class II synthases commence catalysis by protonation of the terminal, isoprenyl carbon-carbon double bond.[19] Since only class I TPSs were discussed in the course of this work, the following paragraph will refrain from a detailed assessment of class II TPSs.

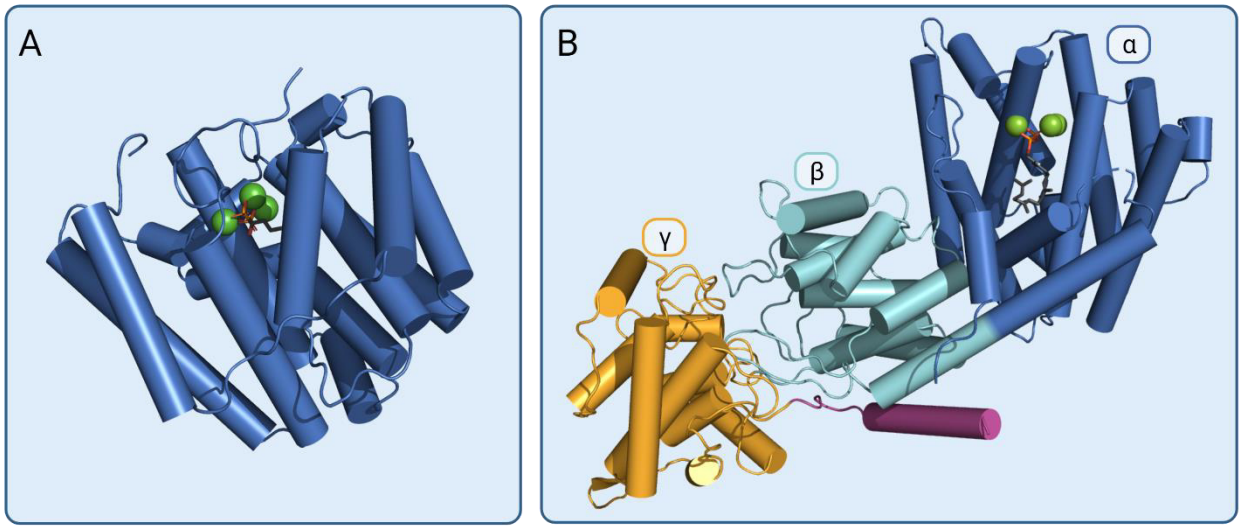


Figure 4: Different class I terpene synthases and respective domain architectures. α-helices are shown as cylinders. A: Diterpene synthase CotB2 (PDB: 6GGI) [55] in complex with 3 Mg^{2+} ions and 2-fluoro-3,7,18-dolabellatriene exhibiting the class I typical alpha-barrel fold. B: Taxadiene synthase (PDB 3P5R) [56] consisting of the three domain architectures α, β and γ (marine: α-domain, turquoise: β-domain, orange: γ-domain, purple: N-terminus of β-domain). Taxadiene synthase is catalytically active in its α-domain, thus crystalized in complex with 3 Mg^{2+} ions and 2-fluorogeranylgeranyl diphosphate. Representation and domain nomenclature of synthases according to Christianson (2017)[19]. Graphical representations were created with PyMOL 2.6.0 (The PyMOL Molecular Graphics System, Schrödinger, LLC.), overall figure created with BioRender.

Class I TPSs are metal-, in particular Mg-dependent, enzymes mediating the formation of a multiplicity of terpenes encompassing various ring systems and stereocenters.[19] Monofunctional class I TPSs exhibit a typical α fold with an α, αβ, or αβγ domain architecture with the catalytically active site always located within the α-domain (c.f. Figure 4).[19] Bifunctional TPSs accordingly exhibit two catalytically active subunits with a possibility for combining the functionality of class I and class II synthases thereby showing respective αα (class I – class I) or αβγ (class I – class II) domain architectures.[19]

For the precise chemical rearrangement of the acyclic, aliphatic terpenoid precursors into the immense variety of terpene natural products, class I TPSs are equipped with a unique, bifacial active site. The anterior part of the active site is composed of a highly polar diphosphate (PP) binding region followed by a hydrophobic carbocation binding pocket located deep inside the α-barrel architecture of class I TPSs.[57,58] However, the enzymatic catalysis in class I TPSs, by traditional means of rate enhancement, is limited to the initial C-O bond cleavage step during the abstraction of the diphosphate moiety.[58] This initial step of catalysis in class I TPSs is affected by highly conserved, metal-binding sequence motifs located at the entrance of the active site on helices D and H. Two motifs, namely the **D**DXX**D** dyad and the (**N**,**D**)D(L,I,V)X(**S**,**T**)XXX**E** (NSE) triad are responsible for complexing three Mg^{2+} ions which are essential for catalysis (c.f. Figure 5). By this means, the aspartate-rich **D**DXX**D** motif, placed on helix D, and the NSE triad, located on helix H, are involved in the initial binding and orientation of the synthase's substrate.[19] While the **D**DXX**D** motif usually is highly conserved within class I TPSs, there are some incidences such as CotB2 from *Streptomyces melanosporofaciens*, hydropyrene synthase (HpS) and (+)-T-muurolol synthase from *Streptomyces clavuligerus* or selina-3,7(11)-diene synthase (SdS) from *Streptomyces pristinaespiralis* in which this motif was altered to **D**DX**D** or **D**DXXX**D** respectively, while still maintaining catalytic activity.[59–62]

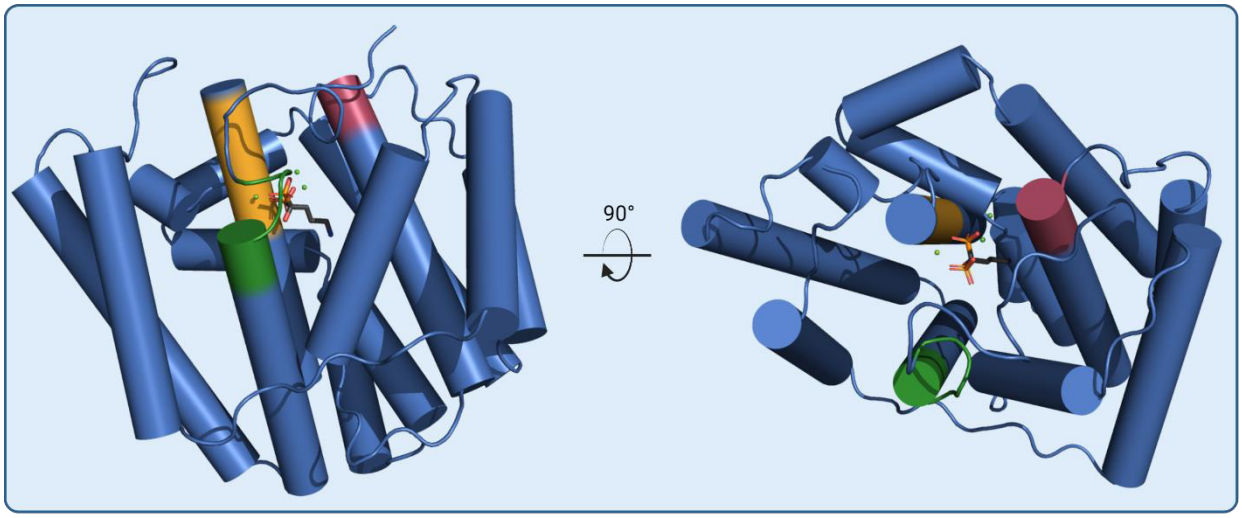


Figure 5: Overall architecture of monomer of diterpene synthase CotB2[55] (side view and rotated by 90° resulting in a view from the top into the active site). α-helices are shown as blue cylinders. The aspartate-rich motif DDXXD is highlighted in red, the NSE motif is coloured in orange and the WxxxxxRY motif in green. CotB2 is depicted in complex with 3 Mg^{2+} ions and 2-fluoro-3,7,18-dolabellatriene. Graphical representations were created with PyMOL 2.6.0 (The PyMOL Molecular Graphics System, Schrödinger, LLC.), overall figure created with BioRender.

Upon substrate binding, class I TPSs are subject to a conformational change from an open (apo) to closed (substrate bound) state including a translation and rotation of several α-helices (B, C, D, F and H) towards the active site thus precisely positioning catalytically important motifs and residues.[55] The structural differences of the open and closed conformation and resulting amino acid interactions were previously shown in detail for the bacterial class I TPS CotB2 in complex with the substrate-analogue 2-fluoro GGPP (FGGPP).[55] The detailed assessment of CotB2's open and closed crystal structure in particular revealed several amino acid interactions directly involved in active side closure (c.f. Table 1). Furthermore, the obtained data in comparison with other bacterial TPS sequences suggests the presence of an additional highly conserved motif, namely WxxxxxRY, which is directly involved in the active site closure by salt bridge formation mediated by the RY dimer (c.f. Figure 5).[55,63] Tryptophan (W) residues in the vicinity of the latter motif are most likely involved in guiding of the product formation as already shown for respective variants of CotB2.[55]

Table 1: CotB2 residues involved in terpene catalysis and their respective function.[55,63]

Residue	Motif	Catalytic role
D110	DDXD	Complexing of Mg^{2+}_A and Mg^{2+}_C and subsequent induction of conformational change of helix D causing active site size reduction and trapping of the substrate
D111	DDXD	Salt bridge formation with R294
-	NSE	Complexing of Mg^{2+}_B
R227	-	Compensation of negative charge of PP
R294	RY-Dimer	Compensation of negative charge of PP
Y295	RY-Dimer	Hydrogen bonds to PP and N220 (NSE)

Subsequently, upon substrate binding and active site closure, the cleavage of the diphosphate moiety affording the initial carbocation is mediated by an effector triad (PP sensor (R), a linker (D) and an effector (G)), which was first shown for the SdS from *Streptomyces pristinaespiralis*.[64] The reaction follows a distinct induced-fit mechanism, in which the effector triad located on the G1/2 helix-break motif is subjected to a major conformational change thus moving the carbonyl oxygen atom of the effector residue towards the bound substrate. In consequence this triggers the abstraction of the PP moiety thereby yielding the initial carbocation.[64]

The resulting carbocation is subsequently subjected to a series of electrophilic cycloadditions and carbocation rearrangements thus maturing the carbocation intermediate along the reaction trajectory to conclusively yield the respective cyclized terpenes.[65–67] The structure of the final product is defined by a precisely decorated, hydrophobic active site which is preshaping the carbocation to chaperone it along its reaction path.[19,58]

In contrast to the first committed step of catalysis, the PP abstraction, the formation of a matured cyclized terpene is a rather fast reaction step due to the reactivity of the carbocation intermediates. In order to control and carefully guide these highly reactive carbocation intermediates along the reaction trajectory the main catalytic challenge rather lies in slowing down the reaction to gain control over the terpenoid intermediate.[58,68] Detailed computational assessment of class I TPSs including the bornyl diphosphate synthase (BPPS), trichodiene synthase (TDS) and CotB2 by Major et al. revealed that the bifacial active site of TPSs is arranged in a layered manner relative to the carbocation binding pocket with alternating negative and positive layers.[58,68] By this means, TPSs create an electrostatic control of the highly reactive carbocation intermediates thus slowing down the reaction in the early stages of catalysis to avoid the formation of numerous side products.[58,69]

To this end it was shown for two fungal class I TPSs, Cop4 and Cop6 from *Coprinus cinereus*, that also the active site size has an influence on the product specificity of TPSs thus rendering a smaller active site more product specific due to a restricted carbocation cyclization route.[70,71] Moreover, TPSs control certain reaction paths by raising the free energy of specific intermediates or by precise positioning of carbocation stabilizing residues such as methionine.[58,69,72] The carbocation is finally matured by either deprotonation or quenching of the carbocation by addition of a water molecule.[73] For instance, as the active site is shielded from entry of bulk solvent by the active site closure, water molecules quenching the carbocation to yield a mono-hydroxylated terpene are also precisely positioned within the active site.[73]

In conclusion, TPSs represent an enigmatic enzyme family encompassing a vast capacity for the catalysis of natural bioactives. To this end, a deep understanding of the mechanistic involved in TPS's carbocation chemistry is crucial for development of industrially viable enzymatic catalysts.

2. Methods

The following paragraph represents an overview of the most important materials, methods and procedures used in this thesis. Further, detailed information is presented in the respective materials and methods parts and the supplementary data sections of the corresponding manuscripts included in this thesis.

Chemicals and reagents

Chemicals used within the presented projects were obtained from standard sources at the highest purity grade available. Components for cultivation media were purchased from Roth chemicals (Karlsruhe, Germany), Sigma-Aldrich (St. Louis, USA), AppliChem GmbH (Darmstadt, Germany), or Merck (Darmstadt, Germany). Ethanol (EtOH), ethyl acetate (EtOAc), n-hexane, acetonitrile (ACN), or methanol (MeOH) were purchased from Roth Chemicals (Darmstadt, Germany) or TH Geyer (Renningen, Germany). $CDCl_3$ and benzene-d_6 used for NMR structure elucidation were obtained from Sigma-Aldrich (St. Louis, USA). Immobilized lipase B from *Candida antarctica* and urea hydrogen peroxide used for *in-vitro* oxyfunctionalisation studies were purchased from Sigma-Aldrich (St. Louis, USA). Enzymes for polymerase chain reaction (PCR), ligation, or restriction digestions were purchased from either Thermo Fisher Scientific (Waltham, USA) or New England Biolabs (NEB) (Ipswich, USA). DNA preparation kits used for plasmid mini preps were obtained from Thermo Fisher Scientific (Waltham, USA), Monarch® PCR & DNA Cleanup Kits and Monarch® DNA Gel Extraction Kits were purchased from New England Biolabs (Ipswich, USA). Terpene standards for quantification were obtained from Biomol (Hamburg, Germany).

Media composition and stock solutions

For all media described below, the pH was adjusted before autoclaving with the appropriate acid or base, respectively. Trace elements, vitamins, sugars, antibiotics, or non-heat-stable media components were either sterile filtrated or autoclaved separately and added to the final media after autoclaving.

Luria-Bertani Broth/ Agar (pH 7)[74]

Yeast extract	5.00 g L^{-1}
NaCl	10.00 g L^{-1}
Tryptone	10.00 g L^{-1}
Agar-Agar (if needed)	15.00 g L^{-1}
ddH_2O	Up to 1 L

SOC medium (pH 7)[75]

Yeast extract	5.00 g L^{-1}
NaCl	0.58 g L^{-1}
KCl	0.18 g L^{-1}
Tryptone	20.00 g L^{-1}
ddH_2O	Up to 980 mL
Added after autoclaving:	
2M $MgSO_4$	10.00 mL L^{-1}
2M Glucose	10.00 mL L^{-1}

Modified R-Media for shake flask experiments (pH 7) (according to Biggs et al.[76])

Batch medium	
K_2HPO_4	5.00 g L^{-1}
KH_2PO_4	10.00 g L^{-1}
$(NH_4)_2HPO_4$	10.00 g L^{-1}
Yeast extract	5.00 g L^{-1}
Citric acid (water free)	1.55 g L^{-1}
1M $MgSO_4$	4.88 mL L^{-1}
0.1M (NH_4) Fe(III) citrate	2.45 mL L^{-1}
Trace element solution (100x)	1.00 mL L^{-1}
50% (w/v) Glycerol (500 g L^{-1})	60.00 mL L^{-1}
1M Thiamin HCl	0.01 mL L^{-1}
Antibiotics (1000x)	1.00 mL L^{-1}
ddH_2O	Up to 1 L

Modified R-Media for fermentation (pH 7)) (according to Biggs et al.[76])

Batch medium		
K_2HPO_4	5.00	g L^{-1}
KH_2PO_4	10.00	g L^{-1}
$(NH_4)_2HPO_4$	10.00	g L^{-1}
Yeast extract	5.00	g L^{-1}
Citric acid (water free)	1.55	g L^{-1}
Glycerol	30.00	g L^{-1}
1M $MgSO_4$	4.88	mL L^{-1}
0.1M (NH_4) Fe(III) citrate	2.45	mL L^{-1}
Trace element solution (100x)	1.00	mL L^{-1}
1M Thiamin HCl	0.01	mL L^{-1}
Antibiotics (1000x)	1.00	mL L^{-1}
ddH_2O	Up to 1 L	

R-Media feeding solution for fermentation

Glycerol	600.00	g L^{-1}
Yeast extract	5.00	g L^{-1}
Collagen hydrolysate	35.00	g L^{-1}
$MgSO_4$	20.00	g L^{-1}
0.1M (NH_4) Fe(III) citrate	2.45	mL L^{-1}
Trace element solution (100x)	20.00	mL L^{-1}
ddH_2O	Up to 1 L	

Trace element solution (100x)

EDTA	5.00	g L^{-1}	13.40	mM
$FeCl_3$*(6 H_2O)	0.83	g L^{-1}	3.10	mM
$ZnCl_2$	84.00	mg L^{-1}	0.62	mM
$CuCl_2$*(2 H_2O)	13.00	mg L^{-1}	76.00	µM
$CoCl_2$*(2 H_2O)	10.00	mg L^{-1}	42.00	µM
H_3BO_3	10.00	mg L^{-1}	162.00	µM
$MnCl_2$*(4 H_2O)	1.60	mg L^{-1}	8.10	µM
ddH_2O	Up to 1 L			

Bacterial strains

All strains used in this thesis were obtained from either Merck Millipore (Darmstadt, Germany), Thermo Fisher Scientific (Waltham, USA), or New England Biolabs (Ipswich, USA). *Escherichia coli* (*E. coli*) strains DH5α or DH10β were employed for cloning and subcloning of all used terpene synthases and their respective mutants.

Heterologous expression of terpene synthases and the 1-deoxy-D-xylulose-5-phosphate (DXP) pathway bottleneck enzymes, all set under constitutive control of a lacI-derived promoter system, was carried out in *E. coli* strains HMS174 (DE3) or ER2566.

Table 2: Employed E. coli strains and their respective genotypes[77]

Strain	Genotype
DH5α	F⁻ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* φ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^- m_K^+$), λ⁻
DH10β	F⁻ Δ(*ara-leu*)*7697*[Δ(*rapA'-cra'*)] Δ(*lac*)*X74*[Δ('*yahH-mhpE*)] duplication(514341-627601)[*nmpC-gltI*] *galK16 galE15* e14⁻(*icd*^{WT} *mcrA*) φ80d*lacZΔM15 recA1 relA1 endA1* Tn*10.10 nupG rpsL150*(Str^R) *rph*⁺ *spoT1* Δ*(mrr-hsdRMS-mcrBC)* λ⁻ Missense(*dnaA glmS glyQ lpxK mreC murA*) Nonsense(*chiA gatZ fhuA*? *yigA ygcG*) Frameshift(*flhC mglA fruB*)
HMS174 (DE3)	F- recA1 hsdR(rK12- mK12+) (DE3) (Rif R)
ER2566	F- λ- fhuA2 [lon] ompT lacZ::T7p07 gal sulA11 Δ(mcrC-mrr)114::IS10 R(mcr-73::miniTn10-TetS)2 R(zgb-210::Tn10)(TetS) endA1 [dcm]

Cloning, plasmid construction and culture conditions

All genes used in this thesis were codon-optimized for the expression in *E. coli* using GeneOptimizer™ software.[78] Genes and oligonucleotides were subsequently synthesised by Eurofins Genomics (Ebersberg, Germany). *E. coli* DH5α cultures were grown in LB broth supplemented with the appropriate antibiotics (ampicillin: 100 µg mL^{-1}; chloramphenicol: 35 µg mL^{-1}; kanamycin: 50 µg mL^{-1}). Ribosomal binding site (RBS) sequences and balanced translational rate were calculated using the SalisLab webtool box including the RBS and operon calculator software package. [79–82]

Standard protocols were employed for polymerase chain reaction (PCR), DNA restriction digestions and ligation. PCR products were checked using 1% (w/v) agarose in TAE buffer and subsequently purified using the Monarch® PCR & DNA Cleanup Kit (New England Biolabs). If needed, gene bands of interest were isolated and purified using Monarch® DNA Gel Extraction Kit (New England Biolabs). Ready cloned plasmids after ligation were transformed into chemically competent cells following the standard heat shock procedure for *E. coli* cells and plated on LB agar supplemented with the appropriate antibiotic.[83]

A colony PCR was conducted for a minimum of 15 clones for each construct. Clones showing a positive colony PCR result were subsequently grown in liquid LB broth supplemented with the appropriate antibiotic for increasing the amount of plasmid DNA for further cloning or expression steps. The GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) was used for plasmid purification. All obtained clones were checked by test restriction digestion and DNA sequencing (Eurofins Genomics). Only clones with a positive sequencing result were used for further experiments. All DNA concentrations in all cloning steps including plasmid preps were checked using a NanoPhotometer® (Implen, Munich, Germany). Wherever applicable, manufacturer's protocols were followed for all procedures.

Gene and protein sequences

In the following section all genes are presented that were used in this thesis. New terpene synthases for functional characterization and heterologous expression were identified by targeted genome mining from *Coniophora puteana (Coniophora puteana RWD-64-598 SS2)*. Based on the amino acid sequence of already characterized Copu3 from *Coniophora puteana* a homology search was performed applying the Basic Local Alignment Search Tool (BLAST).[84,85] The putative, new terpenoid synthases were selected based on the class I sequence specific, conserved motifs DDXXD, the (N,D)D(L,I,V)X(S,T)XXXE (NSE) triad and WxxxxxRY.[19,55]

Terpenoid synthases from *Coniophora puteana*	
Copu5	(XP_007765330)
Copu6	(XP_007773189)
Copu7	(XP_007767204)
Copu9	(XP_007765560)
Copu10	(XP_007766266.1)
Copu11	(XP_007767169.1)

Diterpene synthase from *Streptomyces clavuligerus* (ATCC 27064)[86]	
Hydropyrene synthase (HpS)	(SCLAV_p0765)

MEP pathway bottleneck enzymes[43,53]	
1-deoxy-D-xylulose-5-phosphate synthase from *E. coli* (DXS)	(WP_099145004.1)
isopentenyl-pyrophosphate isomerase from *Haematococcus lacustris* (IDI)	(AAC32208.1)
Geranylgeranyl pyrophosphate synthetase from *Pantoea ananatis* (crtE)	(ADD79325.1)

Mutagenesis of terpene synthases

In order to alter the product profile of the used terpene synthases their respective active site was mutated using a site-directed mutagenesis approach. Standard protocols for cloning were followed as described above. For the identification of target residues for mutagenesis sequence alignments were performed using Clustal Omega.[87,88] If applicable, phylogenetic analysis were conducted with iTOL using the maximum likelihood method.[89,90] In parallel either the web tool iTasser or ROBETTA server [91] were used for generation of homology models of the respective terpene synthases.[92,93] In case of mutagenesis of sesquiterpene synthases, matching amino acid residues as shown by multiple sequence alignments were identified in the active site of the respective synthase. Therefore crystal structure data and homology models were analysed within the environment of UCSF chimera [94,95] and subsequently exchanged for desired amino acids. UCSF chimera was also employed for identification of target residues in all other synthases.

Terpene production in analytical scale

Initial screening of newly identified TPS and other analytical scale terpene production were carried out in either 100 or 500 ml baffled shaking flasks using a New Brunswick Scientific Innova 44R shaking incubator (Eppendorf AG, Hamburg, Germany). Precultures were prepared in either LB or modified R-media (for shaking flask experiments) supplemented with the appropriate antibiotics and incubated at 37 °C, 120 rpm overnight. Main cultures were cultivated in modified R-media (for shaking flask experiments) with a total sugar content of 3% and a starting OD_{600} of 0.1. Depending on the respective terpene product, cultivations were carried out at 22 °C, 120 rpm shaking for 48-72 h (diterpenes) or 30 °C, 100 rpm shaking for 36-48 h (sesquiterpenes), respectively.

Terpene production at technical scale

Fermentation at technical scale for determination of product titers or production of terpenes for further experiments was performed using a DASGIP® 1.3 L parallel rector system (Eppendorf, AG, Germany). Modified R-media (for fermentations) and R-Media feeding solution were used as cultivation broth and feeding solution, respectively. Antifoam 204 (Sigma-Aldrich, USA) was used to control excessive foaming during a fermentation. An overnight culture was used for inoculation of the fermentations (OD_{600}=0.1). The cultivation temperature was kept constant at 23 °C (diterpene production) or 30 °C (sesquiterpene production). Initial stirring velocity and airflow was set to 200 rpm and 0.2 volumes of air per volume of medium per minute (vvm). Dissolved oxygen was kept constant at 30% and maintained by successive increase in stirring velocity (max. 1000 rpm), oxygen content (max. 100%) and airflow (max. 0.8 vvm) during the duration of a fermentation. A pH value of 7.00 was controlled by the addition of 25% aqueous ammonia with a dead band of 0.02. Automated feeding was conducted by a applying a pH-based feeding protocol (feed shot of 40 mL triggered by a rise in pH above 7.05). OD_{600}, dry cell weight (DCW) and terpene content were measured regularly during the course of a fermentation.

Terpene isolation for product screening

In order to functionally characterise and analyse the product profile of terpene synthases in an analytical scale, 20 – 50 ml of engineered *E. coli* cultures in stationary phase (after 48-72 h) were mixed with an equal amount of extraction solution (equal volumes of EtOH, EtOAc and n-hexane). The mixture was shaken vigorously for 2 h at room temperature and subsequently centrifuged (5 min, 8000 rpm) for phase separation. The organic phase was analysed directly via GC-FID/ MS.

Terpene isolation for product purification

For extraction of larger cultivation volumes (e.g.: from fermentations) the whole cultivation broth was mixed with the equal amount of EtOH and shaken (80 rpm) at 20 °C for 12 h. Subsequently the mixture was centrifuged for 20 min at 7000 rpm to separate the supernatant from cell debris. Then EtOAc (50% of the supernatant volume) was added to the mixture and shaken for 5 h (80 rpm, 20 °C). Afterwards the same amount of n-hexane was added to the extraction mixture and shaken for an additional 2 h (80 rpm, 20 °C). Successively, phase separation was conducted using a separation funnel. The upper, organic phase was evaporated using a rotary evaporator. The remaining oily resin was analysed using GC-FID/ MS and subjected to further product purification steps.

Terpene purification

For NMR structure elucidation or further experiments on *in vitro* functionalization, respective terpenes were purified using flash chromatography and HPLC – techniques. All purified compounds were stored in hexane at -20 °C after liquid-liquid extraction from acetonitrile or methanol to hexane.

Separation of terpene fractions from fatty acid residues in crude extracts

First the flash chromatography system PLC 2250 (Gilson, USA) equipped with a Luna 10µm Silica (2) 100A (Phenomenex Inc., Torrance, USA) column, was used for a separation of fatty acid residues and the terpene fraction from the crude *E. coli* extract. n-Hexane and EtOAc were used as mobile phase with a flowrate of 10mL min^{-1} at room temperature. Eluted compounds were analysed by a diode array and an ELSD detector which was flushed with nitrogen gas at 40°C. The oven temperature was set to 40°C.The following gradient was applied: 100% hexane for 15 min, increasing EtOAc within 3 sec to 100%, holding 100% EtOAc for 15 min and then applying 100% hexane for 30 min. Terpene fractions were evaporated under nitrogen flow to approximately 2 mL. The concentration of terpenes of interest were measured using GC-FID. Desired fractions were mixed with Acetonitrile. Subsequently hexane was evaporated until only Acetonitrile containing terpenes of interest remained.

Purification of Diterpenes

In case of products of HpS or IES (Isoelisabethatriene (IE) A/ B, Hydropyrene (HP), Hydropyrenol (HPol)) the further purification was performed by 2d preparative HPLC employing the Ultimate 3000 UHPLC system (Thermo Scientific, USA) containing a binary pump, a diode array detector, an automated fraction collector, and a Jetstream b1.18 column oven. A diode array UV detector at 2,2 mL min^{-1} flow rate using each 25 mg of terpene mixture in Acetonitrile was used. Separation of IEs from Hydropyrenol, Hydropyrene and other terpene derivatives was carried out using a NUCLEODUR® C18 HTec 250/10 mm 5 µm column with guard column NUCLEODUR® C18 HTec 10/8 mm and guard column holder 8 mm (Macherey-Nagel GmbH & Co. KG, Düren, Germany). The separation was performed applying 30% ACN for 5 min, then increasing to 100% within 55 min. This was maintained for 60 min. The oven temperature was set to 30°C.

For separation of IE A and IE B NUCLEODUR® C18 Isis 250/10 mm 5 µm column with guard column NUCLEODUR® C18 Isis 10/8 mm and guard column holder 8 mm (Macherey-Nagel GmbH & Co. KG, Germany) was used applying the following program: 30% MeOH for 5 min, then increasing to 100% within 55 min. This was maintained for 35 min. The oven temperature was set to 30°C.

Purification of sesquiterpenes

In case of sesquiterpenes as the respective cyclisation product, the further purification was carried out using the same instrument setup as described above for the purification of diterpenes. The separation of various sesquiterpenes was carried out on a NUCLEODUR® C18 HTec 250/10 mm and guard column holder 8 mm (Machery-Nagel GmbH & Co. KG, Germany) at 30°C (oven temperature) and a flowrate of 2.2 ml min^{-1} using H_2O and ACN as solvents. The following gradient was applied: 90% ACN for 0.5 min, increased to 100% ACN within 10 min to remain for 12 min, decrease to 90% ACN within 0.1 min to remain for another 10 min.

Analytics

Gas chromatography – FID/ MS analysis

The analysis and quantification of extracted terpenes was performed using a Trace GC-MS Ultra system with DSQII (Thermo Scientific, USA). One microliter (1/10 split ratio) of the respective sample was injected by a TriPlus auto sampler onto a SGE BPX5 column (30 m, I.D. 0.25 mm, film 0.25 µm) with an injector temperature or 280°C. Helium was used as carrier gas with a flow rate of 0.8 ml min^{-1}. Initial oven temperature was set to 50°C for 2 min. The temperature was subsequently increased to 320°C with a rate of 10°C min^{-1} and then held for 3 min. MS data was recorded at 70 eV (EI). Masses were recorded in positive mode in a range between 50 and 650. GC-FID experiments followed the same GC protocol.

Nuclear magnetic resonance (NMR) spectroscopy

Purified compounds for further structural elucidation via NMR experiments were dissolved in either $CDCl_3$ or benzene-d_6. ^{13}C NMR spectra were measured with a Bruker Advance-III 500 MHz spectrometer equipped with a cryo probe head. ^{1}H NMR spectra as well as 2D experiments (HSQC, HMBC, COSY, NOESY) were obtained on an Advance-I 500 MHz system with an inverse probehead (5 mm SEI; $^{1}H/^{13}C$; Z-gradient). The temperature was set to 300 K. Chemical shifts are stated in ppm relative to $CDCl_3$ (δ = 7.26 ppm for ^{1}H and δ = 77.16 ppm for ^{13}C spectra).

3. Research

3.1 Summaries of included publications

Chapter I - Towards a sustainable generation of pseudopterosin-type bioactives

The article "Towards a sustainable generation of pseudopterosin-type bioactives" has been published in Green Chemistry in July 2020 (DOI: 10.1039/d0gc01697g). The author of this thesis, Marion Aloisia Ringel, designed and carried out the experimental work except for identification of residues for site-directed mutagenesis and cloning of the wildtype HpS and respective variants (except for M75L), evaluated the experimental data and wrote the manuscript.

Climate change and increasing population density drive the emergence of new infectious diseases, which demands the development of novel, systemic anti-infective and anti-inflammatory drugs for first-line treatment options.

To this end a prominent example of diterpenoid natural bioactives are the pseudopterosins, which are originally isolated from the Caribbean gorgonian coral, *Antillogorgia elisabethae*. Pseudopterosin-type compounds are known for their potent antibiotic, anti-inflammatory, wound-healing and analgesic activities, with some of them significantly exceeding the efficiency of non-steroidal, synthetic counterparts (i.e., indomethacin), while also causing fewer side effects. The natural extract of *Antillogorgia elisabethae* encompassing pseudopterosin natural products is already applied as anti-irritants in diverse skincare products associated with a multi-billion Euro market value.[34] Importantly, pseudopterosins are of great interest not only because of their superior bioactivities, but also because their biosynthetic precursors feature substantial bioactivities.[35] Particularly, the aromatic, macrocyclic compound erogorgiaene has been demonstrated to have potent activity against multi-drug resistant bacteria such as *Mycobacterium tuberculosis.*[96] However, clinical trials with pseudopterosin-based compounds have not yet progressed past phase II due to an insufficient supply, which is at present exclusively met by extraction from the natural source, a practice that leads to extensive destruction of highly sensitive coral reef ecosystems.[34,37] In light of climate change and global warming, this method is neither scalable nor sustainable.

The presented study shows a sustainable, coral-independent production route for pseudopterosin-type compounds, utilizing the scientific synergies of biotechnology, structural biology, and chemistry. In that context, the enzyme hydropyrene diterpene synthase (HpS) from *Streptomyces clavuligerus*[86] was implemented in an *Escherichia coli* production host.

The main products of the wild-type HpS-catalysed reaction are hydropyrene and hydropyrenol, with isoelisabethatriene A and B (IE A and B) formed in minor quantities. Remarkably, IE A is the direct precursor for erogorgiaene and IE B for novel natural products. The aim of this study was thus to design HpS variants to alter the product spectrum in favour of IE A and B.

Owing to the unavailability of a high-resolution crystal structure of HpS which could facilitate a rational, structure-guided mutagenesis approach, HpS was consequently subjected to an *in silico*-guided mutagenesis, based on the well-characterized bacterial terpene synthase CotB2.[97] Notably, a series of catalytically relevant methionine residues lining the HpS active site were identified. By mutation of these methionine residues, the native HpS product profile was successful altered to an increased IE A and B content, thus providing an efficient route for the conversion of the universal diterpene precursor geranylgeranyl diphosphate (GGPP) to IE A and B, respectively. While the involvement of methionine residues was shown *in silico* for trichodiene synthase[69], the current study represents the first experimental evidence that methionines play an essential role in directing product formation.

The efficient production of the pseudopterosin-specific precursors IE A and B also enabled their selective oxyfunctionalisation, mediated by lipase B (CalB) from *Candida antarctica*. IE A and B showed differential reactivity to form either the well-known pseudopterosin precursor erogorgiaene or the novel natural product 1*R*-epoxy-5,14-elisabethadiene. Consequently, the production of IE A and its subsequent oxidation to erogorgiaene opens a fast-tracked route to a consolidated, sustainable, scalable pseudopterosin production, while 1*R*-epoxy-5,14-elisabethadiene can serve as a new scaffold for pharmaceutical developments.

Chapter II - Biotechnological potential und initial characterization of two novel sesquiterpene synthases from Basidiomycota Coniophora puteana for heterologous production of δ-cadinol

The article "Biotechnological potential und initial characterization of two novel sesquiterpene synthases from Basidiomycota Coniophora puteana for heterologous production of δ-cadinol" has been published in Microbial Cell Factories in April 2022 (DOI: 10.1186/s12934-022-01791-8). The author of this thesis, Marion Aloisia Ringel, and Nicole Dimos contributed equally to the work and writing of this manuscript. Marion Aloisia Ringel identified and functionally characterized the novel terpenoid synthases, carried out site-directed mutagenesis, developed the concept of the publication and carried out experimental work except for protein-crystallography and in-vitro kinetic characterization of the enzymes.

Emerging new diseases and increasing microbial resistance against established drugs highlight the global demand for new, innovative drug leads. Terpenes (and in particular the sesquiterpene subfamily) exhibit antioxidant, anti-inflammatory, antiviral, antimalarial, antibiotic, and antitumor activities.[16,17] Routinely these highly valuable bioactives can only be accessed in minor amounts by extraction from their natural sources. Furthermore, due to the structural complexity of terpenoid drug leads, conventional, chemical synthesis approaches commonly rely on uneconomical, multi-step production routes that are also associated with an adverse ecological profile due to the generation of toxic side products, harsh reaction conditions and the use of metal catalysts.[20] Consequently, the commercialization of various terpenoid drug leads is currently hampered by the lack of sustainable and cost-efficient supply routes.

In order to bypass this limitation, biotechnological production routes for terpenoid bioactives employing genetically engineered microbial hosts and genome mining techniques to identify yet uncharacterized, relevant enzymes from natural sources (i.e., terpene synthases (TPSs)) are on the rise.[84]

This study describes two newly identified class I sesquiterpene synthases from the wood rotting fungus *Coniophora puteana*, Copu5 and Copu9. Both enzymes were demonstrated to be efficient and highly selective (+)-δ-cadinol synthases with product yields of 225 mg L-1 and 395 mg L-1, respectively. (+)-δ-cadinol was previously shown to exhibit cytotoxic activity and may thus emerge as a new, sustainably sourced anti-tumour agent. In addition to their main cyclisation product (+)-δ-cadinol, both Copu5 and Copu9 generate minor side products, including cubebol, a well-known fragrance ingredient originally isolated from plants sources.

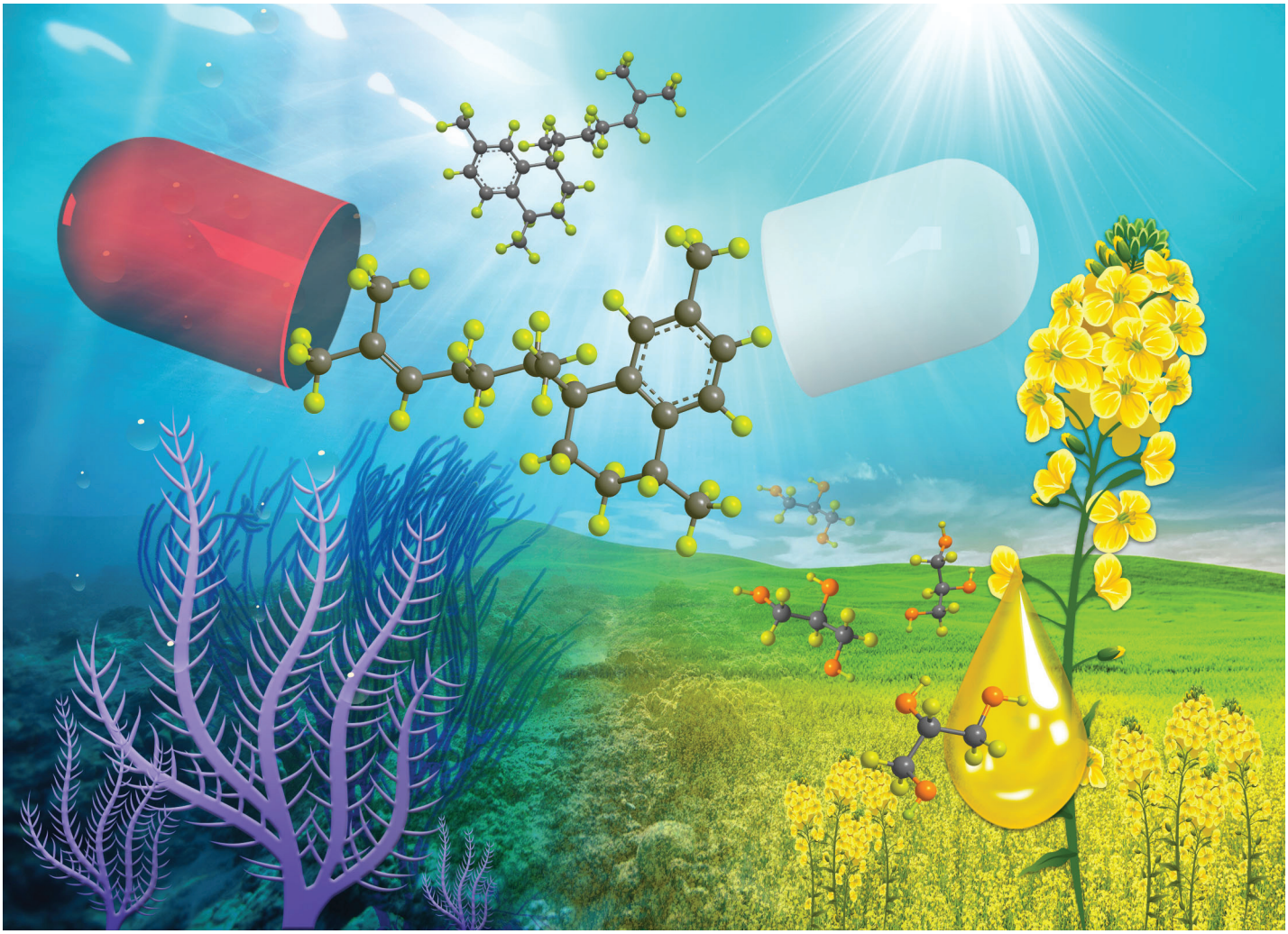
Remarkably, both enzymes share the same product profile, thereby representing the first two terpene synthases from the same organism with identical product profiles. To this end the

generation of the crystal structure of Copu9 (Copu9●Mg3●AHD), which is the first of a class I TPS from *Basidiomycota,* in comparison with a high-quality homology model of Copu5 allowed for detailed, structure sequence related insights into the catalytic mechanism of these two class I TPSs. The detailed inspection of Copu5 and Copu9's active sites revealed several conserved residues potentially involved in catalysis. Subsequently, both class I TPSs were subjected to structurally informed mutagenesis to evaluate whether the product profile of Copu5 and Copu9 could be altered towards that of Copu3, a previously identified cubebol synthase from the same organism, by exchanging the respective amino acids lining the active sites of Copu5 and Copu9 with the corresponding residues of Copu3.

Residues lining the hydrophobic pocket in the active site were shown to have an influence on controlling the carbocation reaction trajectory. However, despite a high degree of conservation in their active sites Copu5 and Copu9 display a difference in structural robustness towards conformational changes. This observation suggests that beyond the previously described catalytically relevant active site motifs addressed by our mutagenesis approach, tertiary structure interactions may play a significant role in shaping the active site and control the carbocation reaction pathway. Copu5 and Copu9 are thus promising model systems for further mechanistic studies of terpenoid catalysis and to design optimized catalysts to produce tailored compounds.

3.2 Full length publications

Towards a sustainable generation of pseudopterosin-type bioactives



Showcasing research from Professor Thomas Brück's laboratory, Werner Siemens Chair of Synthetic Biotechnology, Technical University of Munich, Germany.

Towards a sustainable generation of pseudopterosin-type bioactives

Pseudopterosins (Ps), marine diterpene glycosides derived from the marine octocoral Antillogorgia elisabethae, show potent anti-inflammatory activity. As chemical synthesis is not economical, Ps applications are limited to anti-irritant cosmeceuticals, which are exclusively sourced by coral extraction. The presented two-step biosynthesis approach now opens a sustainable route towards production of the pseudopterosin precursor erogorgiaene. Erogorgiaene itself exhibits potent activity against Mycobacterium tuberculosis. This biosynthetic approach addresses 4 of the 17 UN sustainability goals and showcases the road ahead for developments of a circular economy.

As featured in:

See Thomas Brück *et al.*, *Green Chem.*, 2020, **22**, 6033.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/greenchem

Registered charity number: 207890

Green Chemistry

ROYAL SOCIETY OF CHEMISTRY

PAPER

Cite this: *Green Chem.*, 2020, **22**, 6033

Towards a sustainable generation of pseudopterosin-type bioactives†

Marion Ringel,[a] Markus Reinbold,[a] Max Hirte,[a] Martina Haack,[a] Claudia Huber,[b] Wolfgang Eisenreich,[b] Mahmoud A. Masri,[a] Gerhard Schenk,[c] Luke W. Guddat,[c] Bernhard Loll,[d] Russell Kerr,[e] Daniel Garbe[a] and Thomas Brück*[a]

Pseudopterosins (Ps), marine diterpene glycosides derived from the marine octocoral *Antillogorgia elisabethae*, have potent anti-inflammatory activity demonstrated in phase II clinical trials. As multi-step total chemical synthesis is not economical, Ps applications are limited to anti-irritant cosmeceuticals, which are exclusively sourced by unsustainable coral extraction. While chemical intermediates in Ps biosynthesis have been resolved, the underlying biochemical processes remain elusive. Therefore, a coral independent route to enable sustainable access to Ps and respective bioactive precusors is required. Here, *in silico* guided mutagenesis of the hydropyrene synthase (HpS) from *Streptomyces clavuligerus* reveals five unique, catalytically relevant methionine residues, and affords selective formation of biosynthetic Ps precursors isoelisabethatriene A and B in an *Escherichia coli* host with total terpene yield of HpS M75L of 41.91 ± 1.87 mg L^{-1}. This is the first experimental precedence of methionine residues being involved in terpene synthase catalysis, indicating that HpS may belong to a new subfamily. Further, lipase catalysed chemo-enzymatic oxidation differentially transforms the isomers isoelisabethatriene A and B to the advanced Ps precursor erogorgiaene (yield: 69%) and the new compound 1*R*-epoxy-elisabetha-5,14-diene (EED) (yield: 41%), respectively. As erogorgiaene has significant activity against multi-drug resistant *Mycobacterium tuberculosis*, the process provides a consolidated and scalable access to erogorgiaene, which allows further clinical development of this compound. Moreover, erogorgiaene access also provides a consolidated route for Ps synthesis. Synergistically EED generation affords a new scaffold for Ps-type drug development. These technologies assist in preserving fragile coral reef ecosystem biodiversity and open a fast track for clinical Ps development.

Received 19th May 2020,
Accepted 20th July 2020

DOI: 10.1039/d0gc01697g

rsc.li/greenchem

Introduction

Global population increase paired with unsustainable lifestyles drives climate change and the evolution of new, contractible diseases.[1,2] For the latter, new systemic anti-infective and anti-inflammatory drugs have to be developed as a first-line treatment response.[1] Natural products are a treasure trove for new drug leads, and with over 50 000 characterized compounds terpenes represent the structurally most diverse natural product family. The diterpenoid subfamily encompasses a diverse range of bioactivities including antioxidant, anti-inflammatory, antiviral, antimalarial, antibiotic and antitumor agents, such as the clinically important Taxol.[3] Diterpenes derived from plants, fungi and prokaryotes feature a unique, highly functionalized, structurally complex macrocyclic core. This macrocyclic scaffold is formed by cyclisation of the universal, aliphatic diterpene precursor geranylgeranyl diphosphate (GGPP), a reaction catalysed by the enigmatic family of diterpene synthases.[4,5]

The *Streptomyces*-derived diterpene synthase CotB2, which converts GGPP to cyclooctat-9-en-7-ol, a direct precursor to the anti-inflammatory agent cyclooctatin, belongs to the best characterized diterpene synthases to date.[6,7] Synergistic crystallography, computational and mutagenesis studies allowed for

[a]*Werner Siemens Chair of Synthetic Biotechnology, Dept. of Chemistry, Technical University of Munich (TUM), Lichtenbergstr. 4, 85748 Garching, Germany. E-mail: brueck@tum.de*
[b]*Chair of Biochemistry, Dept. of Chemistry, Technical University of Munich (TUM), Lichtenbergstr. 4, 85478 Garching, Germany*
[c]*School of Chemistry and Molecular Biosciences, The University of Queensland, 68 Cooper Rd, 4702 Brisbane, Australia*
[d]*Institute for Chemistry and Biochemistry, Laboratory of Structural Biochemistry, Freie Universität Berlin, Takustr. 6, 14195 Berlin, Germany*
[e]*Department of Chemistry, and Department of Biomedical Sciences, Atlantic Veterinary College, University of Prince Edward Island, Charlottetown, PE C1A 4P3, Canada*

†Electronic supplementary information (ESI) available. See DOI: 10.1039/d0gc01697g

functional predictions driving rapid diversification of the enzyme's product profile, thereby opening new routes for sustainable drug development.

As diterpene-type natural products predominantly represent secondary metabolites, only minor amounts can routinely be obtained from their respective natural source, often demanding elaborate purification strategies. Moreover, their structural complexity demands uneconomical, multi-step total synthesis approaches.[5] Therefore, commercialization of diterpenoid drug leads is hampered by lack of sustainable and/or cost-efficient supply routes.[5] Prominent examples are the pseudopterosins (Ps), an amphilectane-type diterpene glycoside family, encompassing 31 members, which were originally isolated from the Caribbean gorgonian coral *Antillogorgia elisabethae*.[8] Ps feature potent anti-inflammatory, wound healing and analgesic activities, which significantly exceed those of their non-steroidal, synthetic counterpart indomethacin.[9,10] The superior anti-inflammatory action and reduced side effects are due to a new pharmacological mode of action.[11] Notably, the advanced biosynthetic Ps precursor erogorgiaene exhibits significant antibiotic activity, particularly against *Mycobacterium tuberculosis*, the causative agent of drug-resistant tuberculosis.[12]

Commercially, Ps are applied as natural, marine anti-irritants in diverse skincare products associated with a multi-billion Euro market value.[8,13,14] However, the ever-expanding Ps demand is currently exclusively met by harvesting and extracting its natural source. This practice is neither scalable nor sustainable, as it leads to extensive destruction of sensitive coral reef ecosystems, which are under increasing pressure from climate change effects.[1,15] Moreover, while Ps have progressed to phase II clinical trials as a topical anti-inflammatory agent, further clinical development has been discontinued due to insufficient supply.[13]

While total syntheses for various Ps and their bioactive intermediates (such as erogorgiaene) have been devised,[16–19] none allowed a cost-efficient, sustainable supply of the target compound for further clinical development. Hence, implementation of a sustainable production route *via* an engineered microbial chassis (*e.g. Escherichia coli*) may provide an alternative. In that respect, the most efficient chemical synthesis of the bioactive Ps precursor erogorgiaene and Ps, comprises 8 and 11 steps, respectively, each requiring extensive intermediate work-up procedures.[16,19] All total chemical synthesis are predominantly based on petroleum based educts, require noxious solvents and harsh reaction conditions, which are associated with extensive energy expenditure. Further, multi-step chemical synthesis generate various toxic side streams. This scenario contrasts with biotechnological terpene production processes, which utilize renewable starting materials, are conducted under mild reactions conditions, employ minimal, non-toxic solvents and do not result in toxic side streams.[20,21] Generally, the mass efficiency between chemical and biotechnological routes is dependent on the chemical nature of the target terpene, the cellular production system and bioprocess engineering aspects.[20] However, there are increasing commercial references, such as taxol, arteminisic acid and nootkatone production, where both sustainability, cost – and mass efficiency of biotechnological processes is superior to chemical counterparts.[20,22–24] Therefore, biotechnological generation of chemically complex terpenes is gaining importance for industrial process and product development.

To that end, the chemical steps in Ps biosynthesis have been reported, while the underlying genetics and biochemistry remain unclear.[25–28] Therefore, any biotechnological production scenarios have remained a future prospect. A first step towards this goal has been the isolation and characterisation of elisabethatriene synthase (ELS) from coral tissue, the diterpene synthase responsible for conversion of GGPP to the first biosynthetically committed Ps intermediate, elisabethatriene.[29] However, low yields and complex purification procedures have prohibited the sequence elucidation of this enzyme. Moreover, as marine corals are complex organismic consortia constituting polyps harbouring endosymbiotic microalgae (*Symbiodinium* sp.), as well as various bacteria and fungi,[30,31] the actual Ps production organisms remain elusive.[32,33] Consequently, new, coral-independent approaches for a biotechnological Ps supply route are needed.

The well characterized hydropyrene synthase (HpS) from *Streptomyces clavuligerus* generates hydropyrene (HP) (52%) and hydropyrenol (HPol) (26%) as its main GGPP cyclization products, along with two minor compounds, namely the elisabethatriene isomers isoelisabethatriene (IE) A (12%) and B (9%), respectively.[34] Interestingly, IE A and B only differ in the position of unsaturation within their bicyclic carbon skeleton with reference to the confirmed pseudopterosin precursor, elisabethatriene (Fig. 1).[34,35]

Therefore, these compounds may replace elisabethatriene in a designed biotechnological Ps synthesis cascade (Fig. 2).

This study reports a new, coral independent approach for the biosynthetic production of bioactive intermediates *en-route* to a consolidated, scalable, and sustainable production of Ps-type bioactive compounds. The technologies reported herein

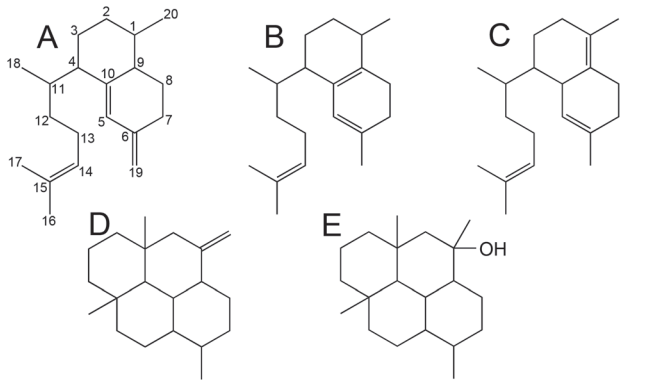


Fig. 1 Structures of elisabethatriene (A), IE A (B), IE B (C), HP (D) and HPol (E); B (12%), C (9%), D (52%) and E (26%) are products of the hydropyrene synthase (HpS) (percentage of product spectrum of wild-type HpS in brackets; carbon numbering as previously described by Kohl and co-workers[27]).

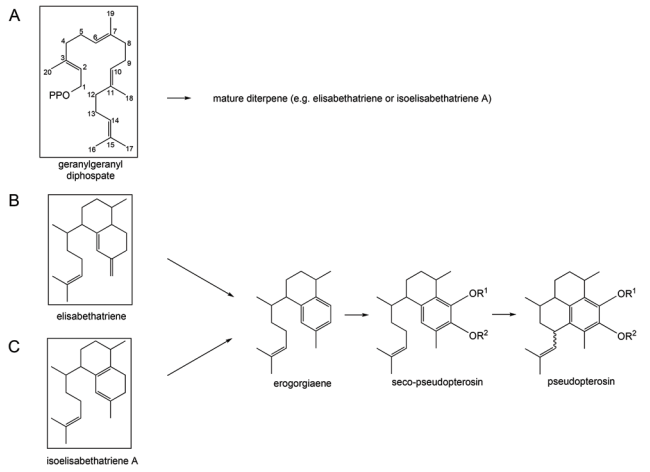


Fig. 2 Early biosynthetic intermediates leading to pseudopterosins; (A) GGPP, the universal diterpene precursor; (B) endogenous coral pathway of pseudopterosin production starting with elisabethatriene *via* erogorgiaene and seco-pseudopterosins; (C) proposed pathway using HpS from *S. clavuligerus* encompassing isoelisabethatriene A; R1,2 = sugar moiety.

will significantly contribute to coral reef protection and provide clinical access to new antibiotic and anti-inflammatory drugs. These can be applied in first line-treatments to control contagion agents and the disease associated excessive inflammatory response during infective epidemics (*i.e.* COVID-19). Moreover, these Ps-type compounds can be applied in chronic inflammatory diseases observed in ageing populations.[36,37]

Results and discussion

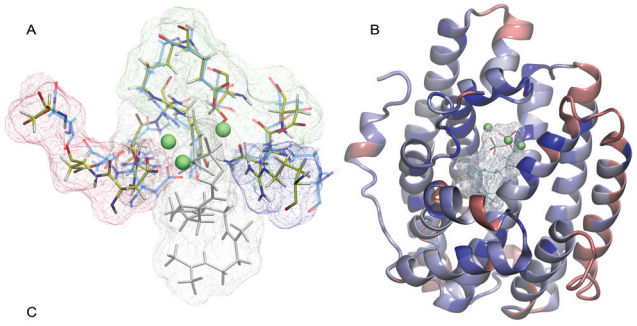
Tailoring *E. coli* for HpS-derived diterpene production

Previous studies reported the formation of early Ps precursors IE A and B by wild-type (wt) HpS, but with low yields.[34] Initial *in vitro* studies with HpS revealed a plausible cyclisation mechanism for GGPP conversion towards the products IE A, IE B, HP and HP-ol (Fig. 1).[35] None of these previous studies specifically aimed for selective production of the Ps precursors IE A and IE B.

In this study, an engineered *E. coli* host harbouring a metabolically balanced two-plasmid terpene production system (previously described by Hirte and co-workers) has been employed for HpS expression.[38] Employing *E. coli* as host system allows for rapid mutagenesis of wt class I HpS and subsequent screening for altered product profiles. Co-transformation of the plasmid carrying the codon optimised HpS gene, together with a separate plasmid harbouring bottleneck enzymes of *E. coli* terpene biosynthesis, led to efficient production of functional HpS. The balanced carbon flux and terpene precursor supply in the tailored *E. coli* host allowed native HpS to efficiently convert GGPP to HP, HP-ol, IE A and IE B (total terpene yield 55.56 ± 2.01 mg L^{-1}, Fig. 2).[34] The HpS-harbouring *E. coli* production system provided rapid and simplified cultivation with high product yields, which were confirmed from purified extracts *via* GC-MS and NMR (Tables S1–S4†).

Delineation of HpS model-based mutagenesis strategy

While class I terpene synthases, such as HpS, share low primary sequence similarity, these enzymes display a significant homology in secondary and tertiary structural features, forming a common α-barrel protein scaffold. Class I terpene synthase catalysis is primed by initial binding and orientation of GGPP *via* its diphosphate (PP) moiety to a conserved Mg^{2+} triade in the active site, characterised by the canonical (DDXX(X)D) motif, and which is located in the centre of the α-barrel (Fig. 3B). Substrate binding initiates active site closure by an induced fit mechanism and subsequent Mg^{2+}-mediated PP hydrolysis, generating a highly reactive, priming carbocation. Solvent water is expelled during active site closure creating a hydrophobic microenvironment that prevents an uncontrolled nucleophilic attack on the carbocation.[6,39] Moreover, specific amino acid residues lining the active site also pre-shape the priming carbocation, thereby significantly influencing the terminal terpene product profile.[40,41] The subsequent intramolecular carbocation rearrangement cascade and terminal cyclization can then commence through C1–C6-, C1–C7-, C1–C10-, C1–C11-, C1–C14- or C1–C15-bond forming reactions (Fig. 2), which are modulated by the relative double bond reactivity of the priming carbocation.[40,42] In addition to the



Motif	Function	CotB2	HpS
DDXX**D**	Metal (Mg^{2+}) binding Diphosphate coordination	DDMD	DDRAID
NSE triad ((**N**,**D**)D(L,I,V)X(**S**, **T**)XXX**E**)	Metal (Mg^{2+}) binding Diphosphate coordination	NDFYSYDRE	NDLXSFARE
WXXXXXRY	active site closure by salt bridge formation	WTTSNKRY	WSARSARY

Fig. 3 (A) Relative position of amino acids of conserved motifs in presence of GGPP and three Mg ions; HpS and CotB2 amino acids are in yellow and blue, respectively; DDX(XX)D motif highlighted in red (HpS: residues ^{82}D, ^{83}D, ^{87}D; CotB2 residues: ^{110}D, ^{111}D, ^{113}D; NDXXSXXRE motif highlighted in green (HpS residues: ^{225}N, ^{226}D, ^{229}S, ^{232}R, ^{233}E, CotB2 residues: ^{220}N, ^{221}D, ^{224}S, ^{227}R, ^{228}E), conserved residue ^{179}R (HpS) and ^{177}R (CotB2), highlighted in dark blue; (B) homology model of HpS with docked substrate GGPP and Mg^{2+} inside the active site with heat map of conserved motifs, dark blue showing highly conserved areas; (C) conserved motifs of class I terpene synthases with their respective catalytic function[7,8,35] (Mg^{2+}-coordinating residues in bold) and the corresponding amino acid sequences for CotB2 and HpS.

inherent carbocation reactivity, the local electrostatic environment created by the substrate-derived PP moiety, as well as transient electronic and ionic interactions with amino acids of the active site, drive and control successive carbocation rearrangements along the reaction trajectory towards an enzyme specific terminal product profile.[41,43] Specifically, terminal cyclisation is induced by amino acid-mediated deprotonation or addition of a water molecule to the final carbocation. These concerted enzyme–substrate interactions facilitate an intense diversity of stereochemically complex diterpene macrocycles, all being derived from the universal precursor GGPP.

Knowledge-based HpS structure–function studies require a structural model to delineate a consolidated mutagenesis strategy. As no HpS crystal structure is available, a homology model[44,45] was constructed, employing the high resolution crystal structure of the taxonomically and secondary structure-related[46] (Fig. S11†) class I terpene synthase CotB2 (PDB-ID 6GGI) as a template.[7,47] The *Streptomyces*-derived diterpene synthase CotB2, which converts GGPP to cyclooctat-9-en-7-ol, a direct precursor to the anti-inflammatory agent cyclooctatin, belongs to one of the best characterized diterpene synthases to date and has been subject to extensive mutagenesis studies.[7,48] So far, CotB2 is the only class I (di)terpene synthase, for which a closed, catalytically relevant structure containing a trapped diterpene reaction intermediate, is available.[7] Computational interrogation of this unique structure in synergy with extensive QM/MD simulations provided detailed insights into the dynamic CotB2 reaction mechanism, highlighting a concerted network of catalytically essential amino acid lining its active site.[7,49] Therefore, CotB2 represents an ideal template for a comprehensive HpS structure–function analysis.

An initial CotB2/HpS comparison indicated that all catalytically relevant class I structural motifs are conserved (Fig. 3). However, the canonical class I DDXXD motif, responsible for initial binding and orientation of the substrate's (GGPP) diphosphate (PP) moiety in the active site, is altered in both CotB2 and HpS to DDXD (^{110}DDMD) and DDXXXD (^{82}DDRAID), respectively. Interestingly, such extensive modifications of the highly conserved DDXXD motif are rare in class I terpene synthases (TPSs).[6,40,42] However, the addition of a single amino acid (X) has also been reported in other TPSs, such as selina-3,7(11)-diene synthase (^{82}DDGYCE) and (+)-T-muurolol synthase (^{83}DDEYCD).[6,50] Other active site motifs are also conserved in HpS and CotB2 (Fig. 3 and Fig. S12†), including the NSE triad[4] and the class I TPS specific WXXXXXRY motif.[7] The HpS-specific amino acid sequences for each catalytically relevant motif are listed in Fig. 3.

Interestingly, a more extensive HpS structural interrogation revealed the distinctive presence of five unique methionine residues (^{71}M, ^{75}M, ^{188}M, ^{300}M and ^{304}M) inside or in the immediate vicinity of the putative HpS active site. A feature that has not been reported or experimentally evaluated for any TPS. The catalytic relevance of these residues is largely unknown, although a computational (QM) study of *Fusarium sporotrichioides* trichodiene synthase (TdS) implicates a methionine residue in interactions with TdS-specific carbocation reaction intermediates.[43] Thus, these methionine residues were included in the mutational strategy to elucidate HpS structure–function relationships to selectively establish the biosynthetic Ps precursors IE A and B as the main GGPP cyclisation products.

Relevant active site residues selected for mutagenesis are listed in Table 1. These include ^{307}W and ^{313}R residues of the conserved ^{307}WXXXXXRY motif. Conservative substitutions of these residues in CotB2 have previously been shown to modulate the product spectrum.[7,47,48,51] In addition to the amino acid changes guided by the CotB2 model, active site residues within a 5 Å sphere of the GGPP binding site were modified. In that regard, target amino acids were either modulated in size or charge to alter the physical properties of the HpS active site.[52] To gain more detailed mechanistic insights into the catalytic function of amino acids surrounding the GGPP binding site, detailed QM/MD simulations are currently in process.

Diterpene-directed product screening of HpS variants

The library of HpS mutants (Fig. 4) was expressed in *E. coli*, and diterpene products were extracted from the cultivation broth.[38] Notably, the total terpene yields of catalytically viable HpS mutants was comparable to that of the wild type enzyme (total terpene yield of HpS M75L: 41.91 ± 1.87 mg l^{-1}; Fig. S2, S4†). Subsequently, all enzyme mutants were evaluated for variations in their product spectrum with respect to wt HpS. A specific focus was given to enhanced IE A and/or B generation.

Mutations Y153A, Y153F, G182K, and W307F did not affect the product spectrum. In contrast, mutations L54A, Y58A, M71R, M71P, M71G, Y78A, A79F, Y153R, R179A, M188G, M188A, M188K, M188Y, M300G, M304D, W307A, W307G and R313A inactivated HpS, indicating that the mutated amino acids are essential for catalysis. Most notably, variants M71Y, M75F, M75L, G182A, G182F, H184A, H184F, M300I and M304C displayed an altered product spectrum with respect to wt HpS (Fig. 5).

Mutation ^{179}R, ^{307}W and ^{313}R, which have analogously been addressed in CotB2 (^{288}W, ^{294}R),[7,47,48] either resulted in a non-

Table 1 Comparison of HpS and CotB2 active site residues used to delineate the HpS mutagenesis strategy. Amino acid residues were chosen due to their potential to alter the product range or stabilize the carbocation intermediate. Residues in bold show identical amino acids in HpS and CotB2

HpS	CotB2	HpS	CotB2
L 54	V 80	G 182	D 180
Y 58	S 84	H 184	G 182
M 71	V 99	M 188	W 186
M 75	N 103	M 300	L 281
Y 78	T 106	M 304	N 292
A 79	F 107	**W 307**	**W 288**
Y 153	F 156	**R 313**	**R 294**
R 179	**R 177**		

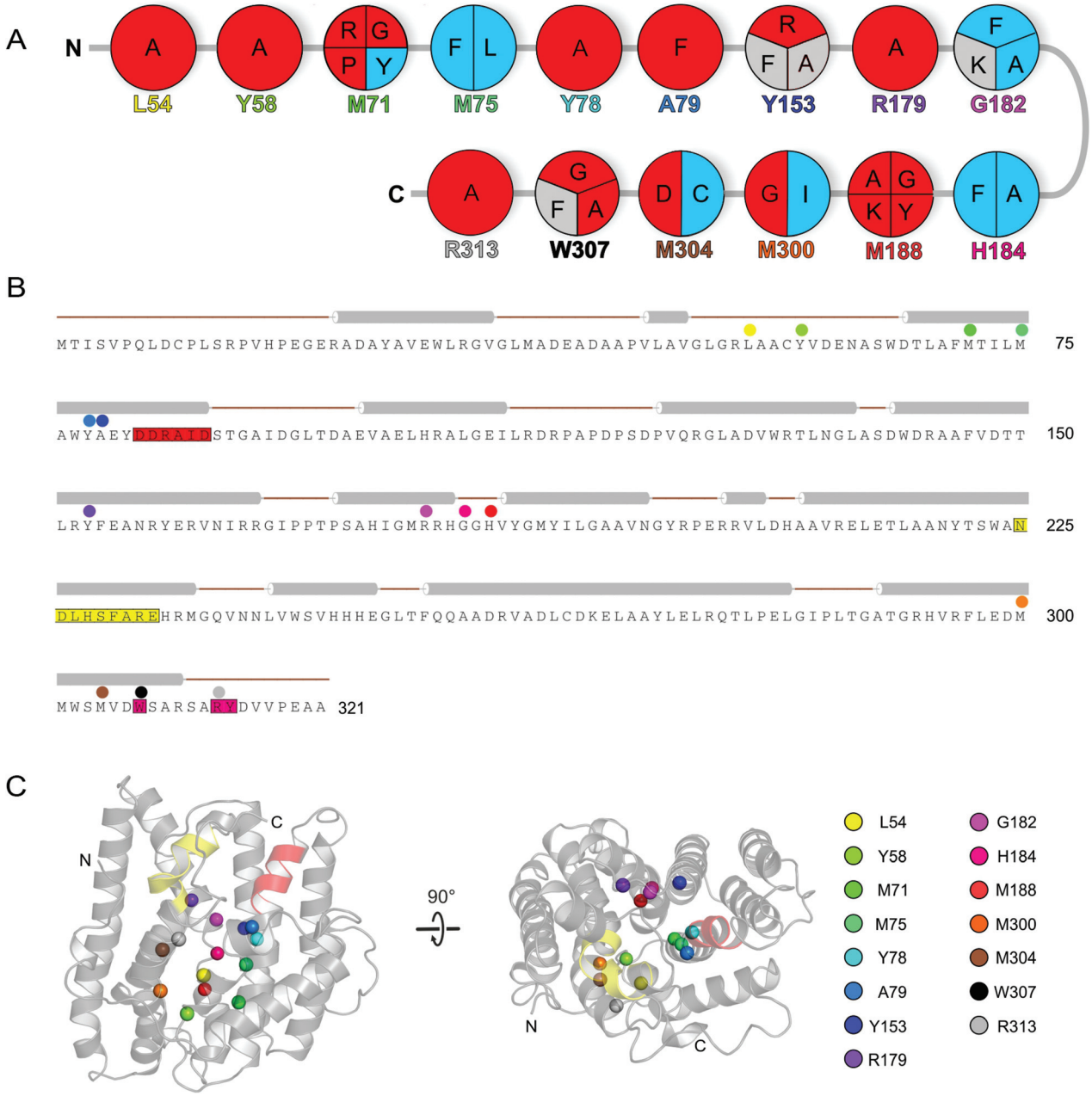
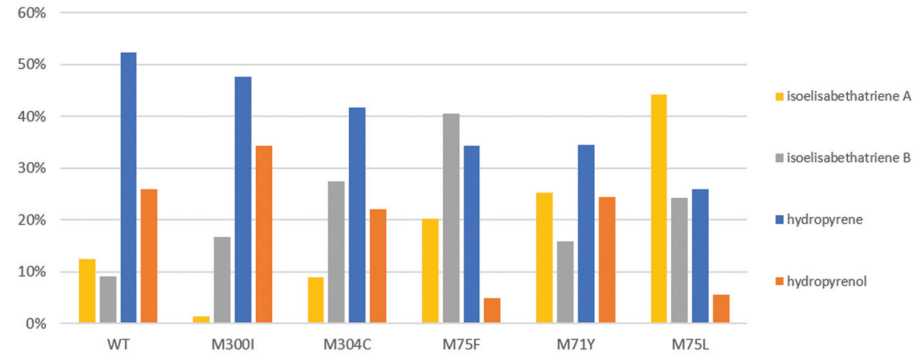


Fig. 4 (A) Variants linked to respective mutation sites (grey: variants with wild type-like product spectrum; red: inactive variants; blue: variants with altered product spectrum), respective amino acids for exchange of wt residues have been chosen according to standard site directed mutagenesis approaches;[52] (B) primary structure and secondary structure elements of HpS (grey cylinders, brown lines and dots indicate alpha helices, beta sheets and mutation sites, respectively; highlighted parts of the primary structure show conserved motifs (red = DDXXD, yellow = NSE; magenta = WXXXXXRY); (C) secondary structure of HpS with highlighted mutation sites.

active variant or displayed the native HpS product spectrum. Therefore, ^{179}R, ^{307}W and ^{313}R are likely to be essential for catalysis, which is consistent with previous reports for CotB2.

Interestingly, mutants M71Y, M75F, M75L, M300I, and M304C targeting HpS-specific methionine residues displayed the most pronounced shifts in the diterpene product profile when compared to wt HpS (Fig. 5). Mutations M300I and M304C lead to a decrease in IE A production and a concomitant increase in IE B. A more significant effect is observed for mutations M71Y, M75F and M75L. Each variant displays a significantly enhanced yield of IE A and B, with a concomitant reduction in HP and HPol production. The most prominent effect is observed for mutations of M75. Notably, M75F showed the highest IE B yield, whereas M75L displayed the highest IE A yield.

Since IE A is a biosynthetic Ps intermediate,[28] its increased yield in the M71Y, M75F and M75L variants is highly encoura-



	WT	M300I	M304C	M75F	M71Y	M75L
isoelisabethatriene A	13%	1%	9%	20%	25%	44%
isoelisabethatriene B	9%	17%	27%	41%	16%	24%
hydropyrene	52%	48%	42%	34%	35%	26%
hydropyrenol	26%	34%	22%	5%	24%	6%

Fig. 5 Relative proportion of the main products of the reaction catalysed by HpS and its variants (displayed as percental ratio of the areas of the respective GC-FID product peaks); HpS variants displayed in order of increasing IE A content.

ging for the ongoing effort to generate a sustainable Ps production platform.

In summary, mutants M75F and M75L were the only variants to shift the product spectrum towards an IE isomer as their major product. M75L is the most promising mutation for Ps production due to its particularly high IE A yield. Therefore, this HpS variant was termed isoelisabethatriene synthase (IES) and used for all downstream efforts to generate advanced Ps intermediates.

In silico driven mechanistic considerations for IE generating mutants

As mutations of ^{71}M and ^{75}M significantly modulated the HpS product spectrum towards IE production it was essential to evaluate the chemical mechanisms that induce these effects. Interestingly, no methionine – carbocation interactions within a distance of ~8 Å have been suggested to be important in the catalytic mechanism of CotB2.[49] Notably, the residue equivalent to ^{75}M in HpS is ^{103}N in CotB2. The latter was proposed to coordinate a water molecule that terminates the CotB2 cyclization cascade[7] or form a dipole-charge interaction during the cyclization reaction.[47,48] Interestingly, the N103A variant CotB2 featured a 3,7,12-dolabellatriene as the major cyclisation product.[47,48] CotB2 has one methionine (^{189}M) that lines its active site but whose replacement by cysteine does not directly interfere with catalysis.[7] The only report that describes the effect of a methionine on terpene synthase catalysis is a computational study of trichodiene synthase, in which ^{73}M is stabilizing selected carbocation conformations *via* a dative sulphur-carbocation bond. Hence, substrate tumbling, and premature deprotonation is prevented.[43] It is thus plausible that methionine residues (especially ^{75}M) in HpS also aid in the stabilisation of carbocation intermediates, providing a route which is crucial for opening a distinct reaction pathway (*e.g. en-route* to HP derivatives; Fig. 6). Therefore, mutagenesis of ^{71}M and ^{75}M in HpS can reroute GGPP cyclisation to either HP or IE derivatives.

Rinkel *et al.* resolved the HpS-specific mechanism for GGPP cyclisation to HP, HP-ol, IE A and IE B, respectively.[35] The initial cyclisation step comprises a 1,10-ring closure, which generates a carbocation at C11. Subsequently, the carbocation (Fig. 7) can be channelled in two distinct reaction pathways.

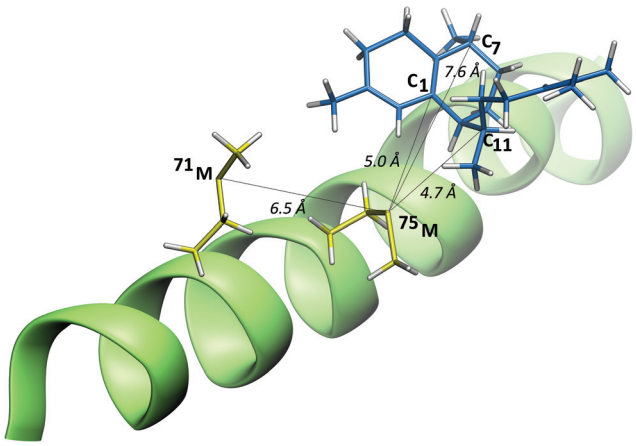


Fig. 6 Distance of ^{75}M to catalytically important carbocations C1 (5.0 Å), C11 (4.7 Å) and C7 (7.6 Å), as well as to ^{71}M (6.5 Å; putatively active as dative bond). The numbering of carbon atoms in the intermediate is based on the numbering of GGPP.

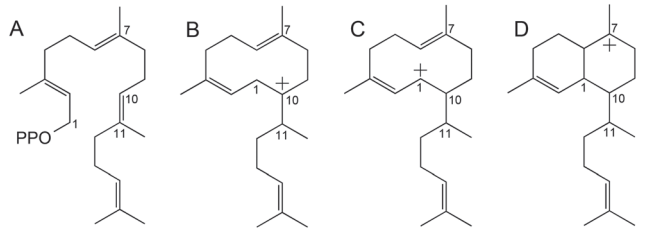


Fig. 7 Important cyclisation intermediates; (A) GGPP; (B) C10 carbocation intermediate; (C) C1 carbocation intermediate (D): C7 carbocation intermediate; numbering of carbon atoms based on numbers of GGPP.

The primary route proceeds *via* the stable *Markovnikov* C11 carbocation followed by a 1,3-hydride shift to form the less stable *anti-Markovnikov* C1 carbocation, which ultimately provides HP or HPol, respectively.[35] In contrast, the formation of IE derivatives requires a 1,3-hydride migration, forming a carbocation at position C7, which leads either to IE A *via* a 1,2-hydrid shift and deprotonation or to IE B *via* simple deprotonation.[35] For effective biotechnological Ps precursor supply, it is crucial to evolve the product spectrum of HpS towards a specific production of IE isomers. As the C11 carbocation is the essential intermediate for changing the preferred pathway to the desired IE A, it is crucial to prevent the 1,3-hydrid shift towards the less stable *anti-Markovnikov* C1 carbocation.[43]

The route towards HP derivative formation requires that the *anti-Markovnikov* C1 carbocation is stabilized within the active site. In wild-type HpS ^{75}M is in close proximity (~5.0 Å) to both C1 and C11 carbons, and it is plausible that in the C1 : C11 carbocation transition the proximal ^{71}M plays a stabilizing role *via* a dative Met–Met interaction (Fig. 6).

Consequently, mutagenesis of ^{71}M and ^{75}M are expected to destabilize the key C1 carbocation intermediate enabling a shift towards IE A and B formation. While this study provides experimental evidence that ^{71}M and ^{75}M are involved in carbocation stabilisation, a plausible catalytic role for the remaining three methionines (^{188}M, ^{300}M, ^{304}M) lining the HpS active site currently cannot be delineated. To that end, solving the crystal structure in synergy with subsequent QM/MD simulation would afford a more insightful dataset.

Identification of hydroxylated IE derivatives

While IE A and B are primary biosynthetic Ps precursors, especially the oxidised IE A variants represent advanced Ps precusors.[25] Culture broth extracts of *E. coli* expressing IES were evaluated for the presence of oxidised IE derivatives using a GC-MS based screening method. Inspection of GC-MS spectra identified a compound with MS spectral similarity to IE but with extended retention time (retention time (Rt) (IE A): 20.46 min; Rt (IE B): 20.87 min; Rt (unknown compound): 22.28 min) and an parent ion mass (*m*/*z*) of 290, indicating the presence of a hydroxyl-moiety (Fig. S17 and S18†). The putative hydroxylated IE derivative can potentially arise by controlled water capture within the HpS active site, which facilitates carbocation quenching along the reaction trajectory. Analogous data have been reported for the class I germacradien-4-ol sesquiterpene synthase.[53]

Moreover, the presence of the aromatized IE derivative, erogorgiaene, a key intermediate in coral-based Ps biosynthesis, was confirmed by comparison with an authentic GC-MS standard isolated from *A. elisabethae* coral tissue.[28,54] However, as erogorgiaene could not be detected, when the *E. coli* extract was analysed directly after the extraction process, it is plausible that oxygen exposure of the analysed extract initiated an oxidative transformation of IE A or B to erogorgiaene. As erogorgiaene is an advanced intermediate in Ps biosynthesis, the current data are consistent with previous reports indicating that hydroxylated elisabethatriene derivatives are direct erogorgiaene precursors in the Ps biosynthetic pathway.[25]

Chemo-enzymatic IE A and B oxidation – a route to advanced Ps precursors

As erogorgiaene formation is a crucial step in Ps biosynthesis, its definitive biosynthetic origin was probed by development of a selective *in vitro* chemo-enzymatic oxidation approach with IE A and B as substrates. Recently, selective functionalization of the macrocyclic diterpene hydrocarbons dolabellatriene and taxadiene *via* lipase-mediated oxidation reactions has been reported.[38,51] Consequently, in a lipase-mediated and chemoenzymatic assay IE A and B were oxidised to establish whether oxyfunctionalization, and therefore activation of the IE hydrocarbon skeleton, is part of the Ps biosynthetic pathway. To ensure future process scalability under economic boundary conditions, we employed the industrially well-established lipase Cal B.[55] The mild lipase-mediated IE oxidation was carried out in ethyl acetate in the presence of urea-hydrogen peroxide with propionic acid, which generates the reactive oxidant.[55–57] The reaction was initiated by *in situ* generation of per-oxo carboxylic acid as the reactive oxidant, which targets olefinic IE bonds either in *re*- or *si*-face conformations to afford a racemic mixture of oxidised products.[58] Reaction progress was monitored by GC-FID analysis, while GC-MS was applied to identify IE A and B specific oxidation products (Fig. 8 and 9).

Identification of the IE B-specific oxidation products

While GC-FID allowed kinetic reaction profiling, parallel GC-MS analysis indicated that the lipase-mediated IE B oxidation resulted in a time dependent formation of IE B mono-

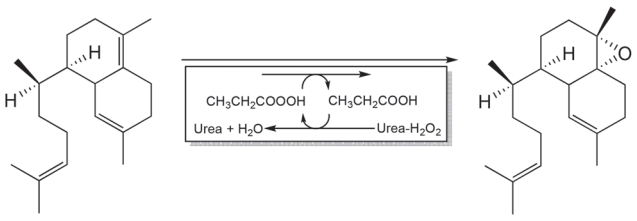


Fig. 8 Lipase-mediated epoxidation of IE B to the new natural product 1*R*-epoxy-5,14-elisabethadiene.

(*m*/*z* 288) and IE B diepoxides (*m*/*z* 304), respectively (Fig. S19–S22†). To enhance product selectivity towards formation of the IE B mono-epoxide, the reaction was terminated after 120 min (yield of 41%). Subsequently, a 2D-HPLC protocol allowed for 1D and 2D NMR spectroscopy-based structure elucidation of the putative IE B-derived mono-epoxide. ^{13}C NMR analysis provided characteristic epoxide-type chemical shifts for C1 and C9 at 62.66 and 64.21 ppm, respectively. Comprehensive NMR signal assignment confirmed the IE B monoepoxide as the new natural product 1*R*-epoxy-5,14-elisabethadiene (EED, Fig. 8, Table S5†).

The epoxidation of the IE B diterpene carbon skeleton enables various downstream biotechnological and chemical functionalization strategies to access a diversified chemical space.[59–61] As most bioactive terpenoids contain at least one functional group, subsequent modification of EED and other IEs is a fundamental step towards the sustainable generation of new pharmaceutical agents.[62,63] Various approaches for hydroxyl group functionalization at the bicyclic Ps carbon skeleton have been applied to generate Ps derivatives and pseudopteroxazoles, which both were active against *M. tuberculosis* and other pathogens.[64]

IE A specific conversion to erogorgiaene

Lipase-mediated oxidation rapidly (90 min) transformed IE A into a single new compound (yield: 69%). Synchronous GC-MS analysis indicated that this compound was the aromatic Ps precursor erogorgiaene (*m*/*z* 270, Fig. 9 and S30–S32†). For structural confirmation, the putative erogorgiaene was purified *via* an optimised 2D-HPLC method and subsequently subjected to 1D and 2D NMR spectroscopy. The resulting NMR signals of the purified compound were in agreement with reported data for (+)-erogorgiaene (Table S6†).[19] While NOESY experiments resolved the relative erogorgiaene stereochemistry, the absolute configuration remained elusive. However, the absolute stereochemistry of the primary HpS cyclisation products was previously resolved using isotopically labelled substrates and CD-spectrophotometric cyclisation product detection.[35] The analysis indicated that HpS converts GGPP to the ((−)-IE A enantiomer, while the *A. elisabethae* coral-derived counterpart constitutes (+)-IE A.[28,35] Similarly, it was deduced that the lipase-based oxidation of HpS derived (−)-IE A leads to the formation of (−)-erogorgiaene, while the coral-derived compound constitutes the (+)-erogorgiaene enantiomer.[19]

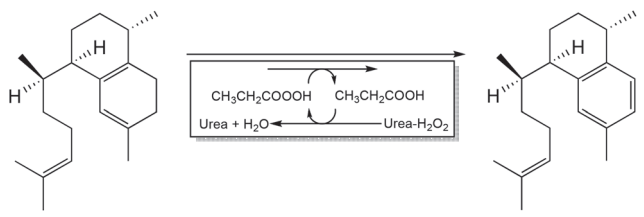


Fig. 9 Lipase-mediated IE A specific conversion of IE A to (−)-erogorgiaene.

The rapid (−)-erogorgiaene formation, precluded observation of any epoxidised IE A intermediates *via* GC-MS. However, mechanistic considerations imply that IE A oxidation proceeds *via* initial epoxidation of the C9–C10 double bond, followed by protonation of the resulting epoxide and a subsequent dehydration, which induces a spontaneous ring system aromatization to afford (−)-erogorgiaene (Fig. S40†).

This mechanistic sequence is supported by detection of elisabethatriene as well as a transient hydroxylated elisabethatriene derivative in crude *A. elisabethae* coral extracts.[25] The spontaneous dehydration of the hydroxylated intermediate to erogorgiaene has been proposed as an essential step in the Ps biosynthesis (Fig. 10).[25] In analogy, the observed chemo-enzymatic conversion of IE A to (−)-erogorgiaene employs the same mechanism. Since erogorgiaene has potent activity against *M. tuberculosis* (with reported MICs as low as 32.25 μg ml^{-1}),[65] the current technology platform can provide a scalable and sustainable access to this interesting natural product.

In light of the accelerated evolution of infectious diseases and the lack of new molecular leads for advanced antibiotic therapy, this platform addresses the essential need for preparedness to fight infection epidemics.[66] The clinical development of new drugs requires bulk access to active pharmaceutical ingredients (API's) under both economical and preferentially ecological constrains.[67,68] Since laboratory scale data for the chemical and biotechnological erogorgiaene synthesis are available, a basic *in silico* scaling and techno-economic analysis (TEA) of both production routes can provide a first glance at process feasibility and provide cues for further technical improvements *en-route* to an industrial process deployment. Therefore, a comparative processes simulation comprising an 8 step chemical synthesis route[19] and the 2 step biotechnological approach described herein was constructed at a scale of 10 kg per batch. The results of each the chemical and biotechnological process simulation provides a basic, initial cost- and productivity overview (Tables S7–18†). The TEA data indicated that initial instrument capital costs for the chemical

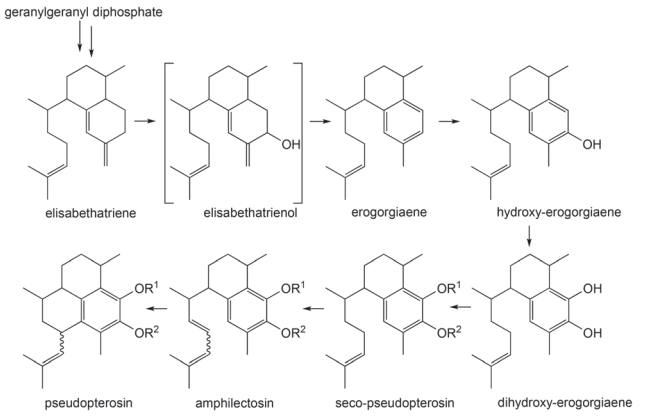


Fig. 10 Postulated biosynthetic pathway of Ps biosynthesis, leading from the geranylgeranyl diphosphate to Ps *via* elisabethatrienol and erogorgiaene (R = -H or -sugar).[25]

synthesis route were 70% lower than for the corresponding biotechnological approach due to different equipment demands in order to realize a 10 kg per batch scale. While the chemical synthesis route at this scale could be conducted in pilot-scale reactors, the biotechnological counterpart would require an industrial-scale fermenter to conduct the first fermentation step. This can be explained by the low fermentative yields (41.91 ± 1.87 mg L^{-1}) currently obtained, which are quite common for terpene production at proof of concept laboratory scale.[20]

However, during industrial process development product yields are commonly significantly (10–1000 fold) enhanced through further genetic engineering and/or change in the microbial production host when combined with process engineering optimizations.[69–71]

Interestingly, operational costs per unit batch (1 kg erogorgiaene) can be considered equivalent. It is however notable, that the raw material costs for biotechnological process are significantly reduced compared to the chemical counterpart, with \$8151 kg^{-1} erogorgiaene and \$20 026 kg^{-1} erogorgiaene, respectively. More specifically, the costs for chemical synthesis involves multiple noxious solvents and (nobel)-metal organic catalysts (*e.g.* Ru-based catalysts),[19] whereas the raw material costs for the biotechnological route can mainly be attributed to high glycerol consumption during the fermentation process. As glycerol is a waste stream of biodiesel production, the raw material stream dominating the biotechnological production route demonstrates favourable economic and ecological performance indicators.[72,73] Further, the time expenditure for one unit production cycle markedly varies between the two approaches requiring 277 h per batch for chemical synthesis and 50 h per batch for biotechnological production, respectively. Summating all process performance indicators, the current erogorgiaene production cost *via* the biotechnological route is significantly lower than for the chemical synthesis route with \$9181 kg^{-1} and \$21 195 kg^{-1}, respectively. Therefore, even at laboratory scale the biotechnological erogorgiaene production route appears to be favourable under economic and ecological constraints. Further yield improvements in the IE A production host combined with an optimized, fermentation process would lead to an industrially viable production process.[74–76]

Conclusions

Ps are potent diterpenoid-type anti-inflammatory compounds, originally isolated from the Caribbean soft coral *A. elisabethae*. Currently, Ps are used as natural anti-irritants in the cosmetics industry, whose supply is entirely sourced from coral extraction. This practice is neither sustainable nor scalable. As efficient total chemical syntheses are not available, this inherent supply issue has also precluded the development of clinically useful compounds from this family of natural products. While the chemical steps in coral-centred Ps biosynthesis are established, the underlying biology remains unclear,[26,30] although some precursors (*e.g.* erogorgiaene) display potent antibiotic activity.[65] This study devised a sustainable, coral-independent route for production of biosynthetic Ps precursors *en-route* to scalable Ps generation. The approach utilised *in silico*-guided mutagenesis of a bacterial HpS to generate the primary Ps precursors IE A and B as the dominant GGPP cyclisation products. Five methionine residues in the HpS active site were identified, that are essential to reroute product formation from HP to IE. The catalytic importance of these residues (especially ^{75}M) is unique among terpene synthases, and current studies aim to establish whether HpS belongs to a new diterpene synthase subclass.

To generate advanced Ps precursors, IE A and B were subjected to a lipase-mediated chemo-enzymatic oxidation. Under identical conditions IE A and B displayed differential reactivity, leading to the formation of the established Ps precursor erogorgiaene and the new natural product 1*R*-epoxy-5,14-elisabethadiene, respectively. Generally, oxyfunctionalisation of diterpenoid backbones provides access to a wide chemical space that enables diversified functionalisation approaches, the basis for the efficient chemo-enzymatic production of various bioactive compounds.[77,78] Given the diversity of functionalised serrulatane diterpenes,[79–81] this development is the basis for the efficient chemoenzymatic production of such bioactive compounds.[82] The concerted application of different biotechnological and chemical functionalisation strategies towards 1*R*-epoxy-5,14-elisabethadiene functionalisation provides a route for development of designed bioactive natural products.

While the bioactivity of IE B oxidation product 1*R*-epoxy-5,14-elisabethadiene has not been evaluated (due to low yields), the IE A oxidation product erogorgiaene has potent antibacterial activity against antibiotic sensitive and multi drug resistant *M. tuberculosis* strains (MICs: 32.25 μg ml^{-1} and 125.00 μg ml^{-1}, respectively).[12,65]

Notably, the chemical synthesis of erogorgiaene requires at least eight steps, utilizing petroleum-based building blocks,[19] the biotechnological approach provides a stereoselective, two-step biosynthetic procedure, solely based on renewable feedstocks.[83] Furthermore, in contrast to chemical synthesis this method does not require metal organic catalysts, nor does it result in any toxic side streams, and is carried out under mild reactions conditions, thereby featuring a superior ecological profile. This consolidated and sustainable production route enables a fast-track pharmaceutical development pipeline for erogorgiaene. As few antibiotic drug leads have been developed to clinical maturity in the past decades, the scalable erogorgiaene supply route addresses an urgent need in the pharmaceutical industry, that underpins new drug development. Such developments are essential to protect an ever-increasing global population from rapidly emerging contagious epidemics (*e.g.* tuberculosis).[2,37,66]

Beyond the generation of new antibiotic entities, erogorgiaene enables consolidated (bio)chemical synthesis to afford Ps-type anti-inflammatory drugs.[9] To that end, sustainably accessed Ps could serve as a new first-line treatment option in

controlling excessive inflammatory symptoms in newly evolving viral epidemics (*e.g.* COVID-19) as well as treating chronic inflammation in aging, industrial populations.[2,37] Ultimately, a sustainable Ps production platform will also replace coral extracts in scalable cosmetic applications, thereby preventing exploitation of fragile reef eco-systems, while protecting marine biodiversity. In summary, the technology platform presented in this study simultaneously complies with all 12 criteria of a green chemistry and addresses four of 17 UN sustainability goals (good health and well-being (goal 3), climate action (13), protecting aquatic (15) and terrestrial (15) life), thereby signalling a path to enhance systemic resilience towards global challenges such as climate change and evolving infectious diseases.

Experimental

Materials

Media components were obtained from Roth chemicals (Karlsruhe, Germany). For terpene extraction technical grade ethanol, ethyl acetate and hexane were purchased from Westfalen AG (Münster, Germany). For all other purposes, highest purity grade chemicals were used. Acetonitrile, ethyl acetate, hexane, methanol, and propionic acid were obtained from Roth chemicals (Karlsruhe, Germany). Immobilized Lipase B from *C. antarctica* (CalB), $CDCl_3$, benzene-*d*6 and urea hydrogen peroxide were purchased from Sigma-Aldrich (St Louis, USA).

Bacterial strains and growth conditions

E. coli strain DH5α was used for plasmid generation and cloning. It was cultivated at 37 °C in Luria–Bertani medium. Terpenes were produced with *E. coli* strain ER2566. During shaking flask experiments *E. coli* ER2566 was grown at 23 °C in either Luria–Bertani or R-Media supplemented with 30 g L^{-1} glucose and 5 g L^{-1} yeast extract. In case of fermentation experiments, *E. coli* ER2566 was cultivated in R-Media supplemented with 30 g L^{-1} glycerol and 5 g L^{-1} yeast extract. Chloramphenicol (30 μg mL^{-1}) and Kanamycin (50 μg mL^{-1}) were added as required.

Gene cloning, plasmid construction

The plasmid for production of diterpene precursors was constructed as described previously.[38] All genes encoding diterpene synthase (Uniprot: SCLAV_p0765) from *S. clavuligerus* (ATCC 27064) were cloned into pACYC-based expression vector system. All genes and primers were synthesized by Eurofins Genomics GmbH (Ebersberg, Germany). Genes were codon optimised for *E. coli* by use of the GeneOptimizer™ software.

Fermentation

Overnight pre-culture was used to inoculate the fermenters of a DASGIP® 1.3 L parallel reactor system (Eppendorf AG, Germany) (OD_{600} = 0.1). Cultivation temperature was kept constant at 23 °C. Stirring velocity, airflow, oxygen content and feeding protocol were set as previously described.[38] Feed solution consisted of 600 g L^{-1} glycerol, 5 g L^{-1} yeast extract, 35 g L^{-1} collagen, 20 g L^{-1} $MgSO_4$, 0.3 g L^{-1} thiamine-HCl, 5 ml L^{-1} 1 M ammonium iron(III) citrate, 20 ml L^{-1} 100× trace elements solution (pH = 7.0) as described previously.[38] To monitor terpene production, samples were taken at different time points.

Protein modelling

Applying the Web tool I-Tasser a homology model of the closed complex of HpS synthase could be generated (https://zhanglab.ccmb.med.umich.edu/I-TASSER/).[44,45] The predicted structure was further analysed and modified within the environment of UCSF Chimera software package including Modeller software package for comparative modelling (http://www.cgl.ucsf.edu/chimera/).[84,85] As previously described by Hirte *et al.*, all substrate docking studies were predicted by AutoDock Vina.[86,87] For subsequent site directed mutagenesis of HpS residues in proximity (3–8 Å) to the docked substrate replacement by more polar or more spacious non-polar residues should allow for quenching of the carbocation intermediate and restrict free folding of the HP skeleton.

For comparative alignment of secondary structure of terpene synthases HHPred applying HMM/HMM comparisons and Ali2D including PSIPRED and MEMSAT software package was used.[46]

Terpene extraction

To extract terpenes from shaking flask experiments and samples taken from fermentation units an equivalent volume of solvent (ethanol : ethyl acetate : hexane = 1 : 1 : 1) was added to the culture broth and mixed for 2 h at room temperature. The solution was centrifuged for 5 min at 8000 rpm to separate the upper organic phase to be analysed by GC-FID and GC-MS. The whole fermentation broth was extracted by addition of the same volume of ethanol. This first process step was carried out on a rotary shaker (80 rpm) at 20 °C for 12 h. Subsequently, the extraction mixture was centrifuged for 20 min at 7000 rpm to separate the supernatant from the cell debris. *Via* addition of ethyl acetate (50% of supernatant volume) a second extraction step on the rotary shaker (80 rpm) was started (20 °C for 5 h). After 5 h the same amount of hexane was added, and the extraction process continued for further 2 h. Finally, the phases were separated by a separation funnel and the organic phase was evaporated.

Terpene purification

Flash chromatography system PLC 2250 (Gilson, USA) allowed for a separation between the fatty acid residues and the terpene fraction. The solvents hexane (A) and ethyl acetate (B) were pumped with a flowrate of 10 mL min^{-1} at room temperature over a Luna 10 μm silica (2) 100 A column. The following gradient was applied: 100% A for 15 min, increasing B in one step to 100%, holding 100% B for 15 min and then applying 100% A for 30 min. Eluted compounds were analysed by a diode array and an ELSD detector which was flushed with

nitrogen gas at 40 °C. Fractions of interest were reduced by nitrogen flow to approximately 2 ml. Fractions containing IEs were mixed with acetonitrile (ACN). Subsequently hexane and ethyl acetate were carefully evaporated until only acetonitrile remained. To further purify the IEs dissolved in ACN, the samples were injected into an Ultimate 3000 UHPLC system (Thermo Scientific, USA) containing a binary pump, a diode array detector, an automated fraction collector, and a Jetstream b1.18 column oven. Separation of isoelisabethatrienes from hydropyrenol, hydropyrene and other terpene derivatives (maximum terpene content of 25 mg) was carried out at 30 °C oven temperature using H_2O (A) and ACN (B) as solvents with a flowrate of 2.2 mL min^{-1} on a NUCLEODUR® C18 HTec 250/10 mm 5 μm column with a guard column NUCLEODUR® C18 HTec 10/8 mm and guard column holder 8 mm (Macherey-Nagel GmbH & Co. KG, Germany). The separation gradient started with 30% B for 5 min, then it increased within 55 min to 100% B. 100% B was maintained for further 60 min.

To separate IE A from B, the same HPLC system was equipped with a NUCLEODUR® C18 Isis 250/10 mm 5 μm column with guard column NUCLEODUR® C18 Isis 10/8 mm and guard column holder 8 mm (Macherey-Nagel GmbH & Co. KG, Germany). The mobile phase consisted of H_2O (A) and MeOH (B). The following program was applied: 30% B for 5 min, then increase to 100% B within 55 min to remain for another 35 min. The oven temperature was set to 30 °C. After liquid–liquid extraction with hexane purified compounds were stored at −20 °C.

Lipase-mediated oxidation of IEs

250 μg mL^{-1} IE A or B was mixed in 5 mL ethyl acetate with 1 μl concentrated propionic acid, 2 mg mL^{-1} immobilised CalB and 2 mg mL^{-1} urea-hydrogen peroxide. Reaction was performed at 22 °C and 1000 rpm in a thermo shaker (Eppendorf AG; Germany). At different time points, samples were taken to monitor the reaction progress by GC-MS analysis. CalB reaction was stopped at appropriate time points by separation of immobilised CalB from reaction mixture by filtration. The remaining solution was diluted with hexane (1 : 4) and filtrated through filter paper. Final product purification occurred in two steps:

In case of IE A, the reaction mixture was first purified by flash chromatography. Hence, the solvents hexane (A) and ethyl acetate (B) were applied at 10 mL min^{-1} to a Luna 10 μm silica (2) 100 A column. After 10 min 100% A, solvent B was increased within 5 min to 100%. Finally, another 30 min the system was operated with 100% A. Subsequently, the fractions were further purified by a preparative HPLC system equipped with a NUCLEODUR® C18 HTec 250/10 mm, 5 μm column with Guard column NUCLEODUR® C18 HTec 10/8 mm and guard column holder 8 mm (Macherey-Nagel GmbH & Co. KG, Germany). The method used an oven temperature of 30 °C and the solvents H_2O (A) and ACN (B) at a flowrate of 2.2 mL min^{-1}. The gradient started with 30% B for 5 min to increase afterwards to 100% B within 55 min. This solvent level was maintained for 60 min.

When purifying products originating from the reaction using IE B, the process also starts with a flash chromatography. This time the gradient was altered to: 1% B for 10 min, increase of B within 41 min to 40%, stay at 40% B for 1 min, further increase to 100% B within the next 3 min and final remain at this level for 10 min. Afterwards the column was cleaned with 100% A for 30 min. Again, the second step consists of a preparative HPLC purification. The solvents remain H_2O (A) and ACN (B), but the following gradient was used: 40% B for 5 min, increase of B to 100% in 30 min and a stay at 100% B for 60 min.

Analytics

Analysis and quantification of terpenes was performed using a Trace GC-MS Ultra system with DSQII (Thermo Scientific, USA). One microliter (1/10 split) of sample was injected by a TriPlus auto sampler onto a SGE BPX5 column (30 m, I.D 0.25 mm, film 0.25 μm) with an injector temperature of 280 °C. Helium was used as carrier gas with a flow rate of 0.8 ml min^{-1}. Initial oven temperature was set to 50 °C for 2 min. The temperature was increased to 320 °C at a rate of 10 °C min^{-1} and then held for 3 min. MS data were recorded at 70 eV (EI). Masses were recorded in positive mode in a range between 50 and 650. GC-FID analysis was performed in the same way.

Compounds for NMR studies were dissolved either in $CDCl_3$ or benzene-d_6. ^{13}C NMR spectra were measured with a Bruker Avance-III 500 MHz spectrometer equipped with a cryo probe head (5 mm CPQNP, $^{1}H/^{13}C/^{31}P/^{19}F/^{29}Si$; Z-gradient). ^{1}H NMR spectra as well as the 2D experiments (HSQC, HMBC, COSY, NOESY) were obtained on an Avance-I 500 MHz system with an inverse probehead (5 mm SEI, $^{1}H/^{13}C$; Z-gradient). The temperature was set to 300 K. Resulting data was processed and analysed by TOPSPIN 3.0 or MestreNova. Chemical shifts were given in ppm relative to $CDCl_3$ (δ = 7.26 ppm for ^{1}H and δ = 77.16 ppm for ^{13}C spectra) or benzene-d_6 (δ = 7.16 ppm for ^{1}H and δ = 128.06 ppm for ^{13}C spectra).

Techno-economic analysis

SuperPro Designer (SPD) V10 (Intelligen, Inc., Scotch Plains, NJ, USA) was used to construct a basic *in silico* cost estimation model for the commercial production of erogorgiaene for each a chemical and a biotechnological production approach.[88,89] The data presented in this study was applied as the base case for the biotechnological process simulation, while the chemical synthesis route was based on results reported by Elford *et al.*, 2011.[19] A production site with a capacity for the production of 10 kg erogorgiaene was the base for this comparison. For determination of the required machine sizing, process cycle timing, and respective purchasing cost (PC), the built-in material and energy balance data in SPD were applied. Raw material scale up and consumption was calculated from the reaction formulas and media composition for each the chemical and biotechnological process approaches. For cost

calculations, recovery of solvents including a 5–10% annual evaporation loss and recovery and recycling (1000×) of catalysts for chemical synthesis steps were assumed. Laboratory scale prices (retrieved from Sigma-Aldrich) for reagents were applied with a discount of 30% in order to delinate bulk procurement pricing. Capital costs including installation, warehousing, site development, additional piping, construction, and contingency were estimated as a percentage of the inside-battery-limit (ISBL) equipment costs based on percentages retrieved from the NREL database.[90] Operational costs were extracted from the build-in model from SPD V10. Details concerning the basic cost estimation model are presented in the ESI.†

Author contribution

MRi, MRe, DG and TB conceived the project. MRi, MRe carried out experiments, supervised by DG and TB. MHa supported the analytical work. CH and WE contributed in structural analysis with NMR expertise. MM carried out the techno-economic analysis. RK supported experiments with authentic material standards and scientific expertise in data analysis. MHi and BL supported protein modelling studies and graphical data representation. GS and LG contributed with scientific expertise in structural biology. MRi drafted the manuscript. All authors revised the manuscript. TB and MRi finalized manuscript preparation.

Conflicts of interest

The authors declare no conflict of interests.

Acknowledgements

TB, MRi and DG gratefully acknowledge funding by the Werner Siemens foundation for establishing the field of Synthetic Biotechnology at TUM. MRe, TB, DG and RK acknowledge funding of the OMCBP project by the German federal ministry of Education and Research (grant no.: 031A276A) under the call initiative BMBF International, which enabled synchronized bilateral German: Canadian projects.

References

1 P. Camill, *Nature Education Knowledge*, 2010, vol. 3.
2 WHO, *Climate change and human health. risks and responses. Summary*, Genève, 2003.
3 D. G. I. Kingston, *Phytochemistry*, 2007, **68**, 1844–1854.
4 D. W. Christianson, *Chem. Rev.*, 2006, **106**, 3412–3442.
5 K. Kemper, M. Hirte, M. Reinbold, M. Fuchs and T. Brück, *Beilstein J. Org. Chem.*, 2017, **13**, 845–854.
6 R. Driller, D. Garbe, N. Mehlmer, M. Fuchs, K. Raz, D. T. Major, T. Brück and B. Loll, *Beilstein J. Org. Chem.*, 2019, **15**, 2355–2368.
7 R. Driller, S. Janke, M. Fuchs, E. Warner, A. R. Mhashal, D. T. Major, M. Christmann, T. Brück and B. Loll, *Nat. Commun.*, 2018, **9**, 3971.
8 F. Berrué, M. W. B. McCulloch and R. G. Kerr, *Bioorg. Med. Chem.*, 2011, **19**, 6702–6719.
9 S. A. Look, W. Fenical, R. S. Jacobs and J. Clardy, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 6238–6240.
10 S. A. Look, W. Fenical, G. K. Matsumoto and J. Clardy, *J. Org. Chem.*, 1986, **51**, 5140–5145.
11 A. M. S. Mayer, P. B. Jacobson, W. Fenical, R. S. Jacobs and K. B. Glaser, *Life Sci.*, 1998, **62**, PL401–PL407.
12 A. D. Rodríguez and C. Ramírez, *J. Nat. Prod.*, 2001, **64**, 100–102.
13 A. Martins, H. Vieira, H. Gaspar and S. Santos, *Mar. Drugs*, 2014, **12**, 1066–1101.
14 A. M. S. Mayer, K. B. Glaser, C. Cuevas, R. S. Jacobs, W. Kem, R. D. Little, J. M. McIntosh, D. J. Newman, B. C. Potts and D. E. Shuster, *Trends Pharmacol. Sci.*, 2010, **31**, 255–265.
15 H. R. Lasker, *PLoS One*, 2013, **8**, e74587.
16 C. G. Newton, S. L. Drew, A. L. Lawrence, A. C. Willis, M. N. Paddon-Row and M. S. Sherburn, *Nat. Chem.*, 2015, **7**, 82–86.
17 D. C. Harrowven and M. J. Tyte, *Tetrahedron Lett.*, 2004, **45**, 2089–2091.
18 H. M. L. Davies and A. M. Walji, *Angew. Chem., Int. Ed.*, 2005, **44**, 1733–1735.
19 T. G. Elford, S. Nave, R. P. Sonawane and V. K. Aggarwal, *J. Am. Chem. Soc.*, 2011, **133**, 16798–16801.
20 Y. Zhang, J. Nielsen and Z. Liu, *FEMS Yeast Res.*, 2017, **17**(8), fox080.
21 P. G. Mountford, in *Green chemistry in the pharmaceutical industry*, ed. P. J. Dunn, Wiley-VCH, Weinheim, 2010, pp. 145–160.
22 E. Brenna and F. Parmeggiani, in *Industrial biotechnology. Products and processes*, ed. C. Wittmann and J. C. Liao, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2017, pp. 271–308.
23 T. Wriessnegger, P. Augustin, M. Engleder, E. Leitner, M. Müller, I. Kaluzna, M. Schürmann, D. Mink, G. Zellnig, H. Schwab and H. Pichler, *Metab. Eng.*, 2014, **24**, 18–29.
24 S. H. Kung, S. Lund, A. Murarka, D. McPhee and C. J. Paddon, *Front. Plant Sci.*, 2018, **9**, 87.
25 C. Duque, M. Puyana, L. Castellanos, A. Arias, H. Correa, O. Osorno, T. Asai, N. Hara and Y. Fujimoto, *Tetrahedron*, 2006, **62**, 4205–4213.
26 L. D. Mydlarz, R. S. Jacobs, J. Boehnlein and R. G. Kerr, *Chem. Biol.*, 2003, **10**, 1051–1056.
27 A. C. Kohl, A. Ata and R. G. Kerr, *J. Ind. Microbiol. Biotechnol.*, 2003, **30**, 495–499.
28 A. C. Kohl and R. G. Kerr, *Mar. Drugs*, 2003, **1**, 54–65.
29 T. B. Brück and R. G. Kerr, *Comp. Biochem. Physiol., Part B: Biochem. Mol. Biol.*, 2006, **143**, 269–278.
30 V. Robertson, B. Haltli, E. P. McCauley, D. P. Overy and R. G. Kerr, *Microorganisms*, 2016, **4**, 23.

31 J. A. J. M. van de Water, D. Allemand and C. Ferrier-Pagès, *Microbiome*, 2018, **6**, 64.
32 L. R. Hunt, S. M. Smith, K. R. Downum and L. D. Mydlarz, *Mar. Drugs*, 2012, **10**, 1225–1243.
33 H. Choi and D.-C. Oh, *Arch. Pharmacal Res.*, 2015, **38**, 1591–1605.
34 Y. Yamada, S. Arima, T. Nagamitsu, K. Johmoto, H. Uekusa, T. Eguchi, K. Shin-ya, D. E. Cane and H. Ikeda, *J. Antibiot.*, 2015, **68**, 385–394.
35 J. Rinkel, P. Rabe, X. Chen, T. G. Köllner, F. Chen and J. S. Dickschat, *Chemistry*, 2017, **23**, 10501–10505.
36 K. E. Park, Y. Qin and A. A. Bavry, *Aging Health*, 2012, **8**, 167–177.
37 J. Stebbing, A. Phelan, I. Griffin, C. Tucker, O. Oechsle, D. Smith and P. Richardson, *Lancet Infect. Dis.*, 2020, **20**, 400–402.
38 M. Hirte, W. Mischko, K. Kemper, S. Röhrer, C. Huber, M. Fuchs, W. Eisenreich, M. Minceva and T. B. Brück, *Green Chem.*, 2018, **20**, 5374–5384.
39 P. Baer, P. Rabe, K. Fischer, C. A. Citron, T. A. Klapschinski, M. Groll and J. S. Dickschat, *Angew. Chem., Int. Ed.*, 2014, **53**, 7652–7656.
40 D. W. Christianson, *Chem. Rev.*, 2017, **117**, 11570–11648.
41 D. T. Major, *ACS Catal.*, 2017, **7**, 5461–5465.
42 J. S. Dickschat, *Nat. Prod. Rep.*, 2016, **33**, 87–110.
43 M. Dixit, M. Weitman, J. Gao and D. T. Major, *ACS Catal.*, 2017, **7**, 812–818.
44 A. Roy, A. Kucukural and Y. Zhang, *Nat. Protoc.*, 2010, **5**, 725–738.
45 J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson and Y. Zhang, *Nat. Methods*, 2015, **12**, 7–8.
46 L. Zimmermann, A. Stephens, S.-Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. N. Lupas and V. Alva, *J. Mol. Biol.*, 2018, **430**, 2237–2243.
47 R. Janke, C. Görner, M. Hirte, T. Brück and B. Loll, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2014, **70**, 1528–1537.
48 T. Tomita, S.-Y. Kim, K. Teramoto, A. Meguro, T. Ozaki, A. Yoshida, Y. Motoyoshi, N. Mori, K. Ishigami, H. Watanabe, M. Nishiyama and T. Kuzuyama, *ACS Chem. Biol.*, 2017, **12**, 1621–1628.
49 K. Raz, R. Driller, T. Brück, B. Loll and D. T. Major, *Beilstein J. Org. Chem.*, 2020, **16**, 50–59.
50 Y. Hu, W. K. W. Chou, R. Hopson and D. E. Cane, *Chem. Biol.*, 2011, **18**, 32–37.
51 C. Görner, M. Hirte, S. Huber, P. Schrepfer and T. Brück, *Front. Microbiol.*, 2015, **6**, 1115.
52 J. Yazmin and G. Saab-Rinco, in *Site-Directed Mutagenesis in the Research of Protein Kinases - The Case of Protein Kinase CK2*, ed. D. Figurski, INTECH Open Access Publisher, 2013.
53 D. J. Grundy, M. Chen, V. González, S. Leoni, D. J. Miller, D. W. Christianson and R. K. Allemann, *Biochemistry*, 2016, **55**, 2112–2121.
54 T. A. Ferns and R. G. Kerr, *J. Org. Chem.*, 2005, **70**, 6152–6157.
55 E. G. Ankudey, H. F. Olivo and T. L. Peeples, *Green Chem.*, 2006, **8**, 923–926.
56 J. M. R. da Silva and M. d. G. Nascimento, *Process Biochem.*, 2012, **47**, 517–522.
57 S. Ranganathan, S. Zeitlhofer and V. Sieber, *Green Chem.*, 2017, **19**, 2576–2586.
58 F. Björkling, S. E. Godtfredsen and O. Kirk, *J. Chem. Soc., Chem. Commun.*, 1990, **3**, 1301–1303.
59 C. Wang, L. Luo and H. Yamamoto, *Acc. Chem. Res.*, 2016, **49**, 193–204.
60 Z. Wang, Y.-T. Cui, Z.-B. Xu and J. Qu, *J. Org. Chem.*, 2008, **73**, 2270–2274.
61 O. Hauenstein, M. Reiter, S. Agarwal, B. Rieger and A. Greiner, *Green Chem.*, 2016, **18**, 760–770.
62 *Biotechnology of isoprenoids*, ed. J. Schrader, J. Bohlmann and J. Alonso-Gutierrez, Springer, Cham, 2015, vol. 148.
63 C. Khosla and J. D. Keasling, *Nat. Rev. Drug Discovery*, 2003, **2**, 1019–1025.
64 M. W. B. McCulloch, B. Haltli, D. H. Marchbank and R. G. Kerr, *Mar. Drugs*, 2012, **10**, 1711–1728.
65 C. A. Incerti-Pradillos, M. A. Kabeshov, P. S. O'Hora, S. A. Shipilovskikh, A. E. Rubtsov, V. A. Drobkova, S. Y. Balandina and A. V. Malkov, *Chemistry*, 2016, **22**, 14390–14396.
66 K. Lewis, *Cell*, 2020, **181**, 29–45.
67 J. L. Tucker, *Org. Process Res. Dev.*, 2006, **10**, 315–319.
68 P. Nightingale, *Ind. Corp. Change*, 2000, **9**, 315–359.
69 Z.-Q. Li, Y. Liu, B.-Y. Liu, H. Wang, H.-C. Ye and G.-F. Li, *J. Integr. Plant Biol.*, 2006, **48**, 1486–1492.
70 P. J. Westfall, D. J. Pitera, J. R. Lenihan, D. Eng, F. X. Woolard, R. Regentin, T. Horning, H. Tsuruta, D. J. Melis, A. Owens, S. Fickes, D. Diola, K. R. Benjamin, J. D. Keasling, M. D. Leavell, D. J. McPhee, N. S. Renninger, J. D. Newman and C. J. Paddon, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, E111–E118.
71 J. V. Pham, M. A. Yilma, A. Feliz, M. T. Majid, N. Maffetone, J. R. Walker, E. Kim, H. J. Cho, J. M. Reynolds, M. C. Song, S. R. Park and Y. J. Yoon, *Front. Microbiol.*, 2019, **10**, 1404.
72 V. K. Garlapati, U. Shankar and A. Budhiraja, *Biotechnol. Rep.*, 2016, **9**, 9–14.
73 M. R. Monteiro, C. L. Kugelmeier, R. S. Pinheiro, M. O. Batalha and A. da Silva César, *Renewable Sustainable Energy Rev.*, 2018, **88**, 109–122.
74 B. E. Min, H. G. Hwang, H. G. Lim and G. Y. Jung, *J. Ind. Microbiol. Biotechnol.*, 2017, **44**, 89–98.
75 S. Zhang, J. Chu and Y. Zhuang, in *Biomanufacturing*, ed. J.-J. Zhong, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 97–150.
76 Y. Wang, J. Chu, Y. Zhuang, Y. Wang, J. Xia and S. Zhang, *Biotechnol. Adv.*, 2009, **27**, 989–995.
77 K. Härtl, K. McGraphery, J. Rüdiger and W. Schwab, in *Biotechnology of Natural Products*, Springer, 2018, pp. 219–263.
78 W. Schwab, B. M. Lange and M. Wüst, *Biotechnology of Natural Products*, Springer International Publishing, Cham, 2018.
79 H. J. Barnes, M. P. Arlotto and M. R. Waterman, *Proc. Natl. Acad. Sci. U. S. A.*, 1991, 5597–5601.

80 C. P. Ndi, S. J. Semple, H. J. Griesser, S. M. Pyke and M. D. Barton, *Phytochemistry*, 2007, **68**, 2684–2690.
81 R. Escarcena, J. Perez-Meseguer, E. Del Olmo, B. Alanis-Garza, E. Garza-González, R. Salazar-Aranda and N. Waksman de Torres, *Molecules*, 2015, **20**, 7245–7262.
82 F. Cao, C.-L. Shao, Y.-F. Liu, H.-J. Zhu and C.-Y. Wang, *Sci. Rep.*, 2017, **7**, 12548.
83 W. Mischko, M. Hirte, S. Roehrer, H. Engelhardt, N. Mehlmer, M. Minceva and T. Brück, *Green Chem.*, 2018, **68**, 1844.
84 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605–1612.
85 N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M.-Y. Shen, U. Pieper and A. Sali, *Current protocols in bioinformatics*, 2006, ch. 5, unit-5.6.
86 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455–461.
87 M. Hirte, N. Meese, M. Mertz, M. Fuchs and T. B. Brück, *Front. Chem.*, 2018, **6**, 101.
88 D. Petrides, *Biosep. Sci. Eng.*, 2000, 1–83.
89 A. Toumi, C. Jurgens, C. Jungo, B. Maier, V. Papavasileiou and D. Petrides, *Pharm. Eng.*, 2010, **30**, 1–9.
90 D. Humbird, R. Davis, L. Tao, C. Kinchin, D. Hsu, A. Aden, P. Schoen, J. Lukas, B. Olthof and M. Worley, *Process design and economics for biochemical conversion of lignocellulosic biomass to ethanol: dilute-acid pretreatment and enzymatic hydrolysis of corn stover*, National Renewable Energy Lab. (NREL), Golden, CO, United States, 2011.

Biotechnological potential and initial characterization of two novel sesquiterpene synthases from Basidiomycota *Coniophora puteana* for heterologous production of δ-cadinol

Ringel *et al. Microbial Cell Factories* (2022) 21:64
https://doi.org/10.1186/s12934-022-01791-8

Microbial Cell Factories

RESEARCH Open Access

Biotechnological potential and initial characterization of two novel sesquiterpene synthases from Basidiomycota *Coniophora puteana* for heterologous production of δ-cadinol

Marion Ringel[1†], Nicole Dimos[2†], Stephanie Himpich[2], Martina Haack[1], Claudia Huber[3], Wolfgang Eisenreich[3], Gerhard Schenk[4], Bernhard Loll[2*] and Thomas Brück[1*]

Abstract

Background: Terpene synthases are versatile catalysts in all domains of life, catalyzing the formation of an enormous variety of different terpenoid secondary metabolites. Due to their diverse bioactive properties, terpenoids are of great interest as innovative ingredients in pharmaceutical and cosmetic applications. Recent advances in genome sequencing have led to the discovery of numerous terpene synthases, in particular in Basidiomycota like the wood rotting fungus *Coniophora puteana*, which further enhances the scope for the manufacture of terpenes for industrial purposes.

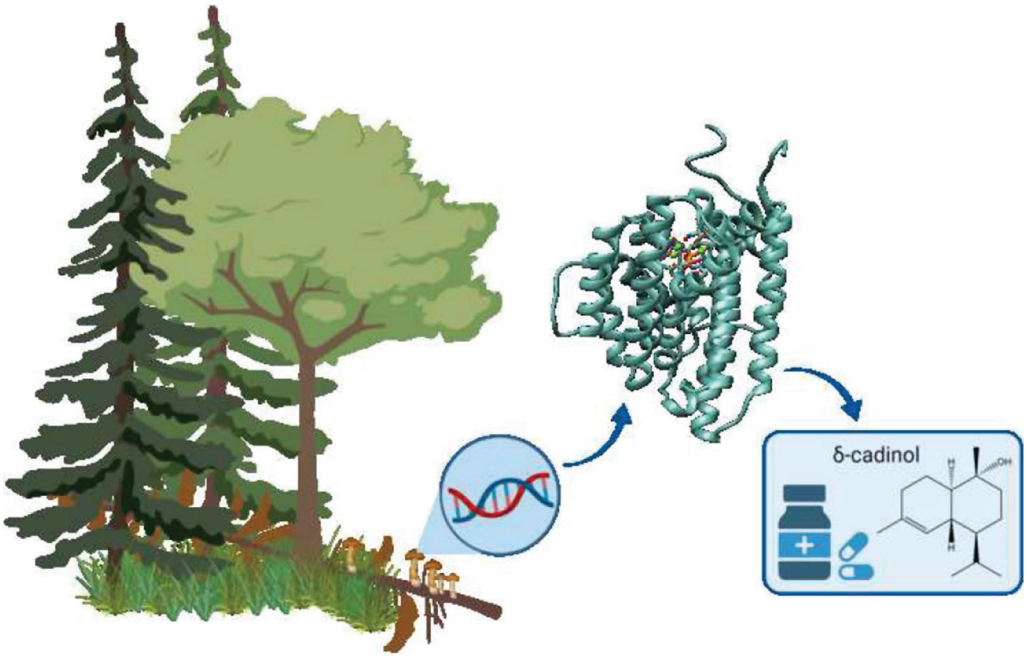
Results: In this study we describe the identification of two novel (+)-δ-cadinol synthases from *C. puteana,* Copu5 and Copu9. The sesquiterpene (+)-δ-cadinol was previously shown to exhibit cytotoxic activity therefore having an application as possible, new, and sustainably sourced anti-tumor agent. In an *Escherichia coli* strain, optimized for sesquiterpene production, titers of 225 mg l^{-1} and 395 mg l^{-1}, respectively, could be achieved. Remarkably, both enzymes share the same product profile thereby representing the first two terpene synthases from Basidiomycota with identical product profiles. We solved the crystal structure of Copu9 in its closed conformation, for the first time providing molecular details of sesquiterpene synthase from Basidiomycota. Based on the Copu9 structure, we conducted structure-based mutagenesis of amino acid residues lining the active site, thereby altering the product profile. Interestingly, the mutagenesis study also revealed that despite the conserved product profiles of Copu5 and Copu9 different conformational changes may accompany the catalytic cycle of the two enzymes. This observation suggests that the involvement of tertiary structure elements in the reaction mechanism(s) employed by terpene synthases may be more complex than commonly expected.

*Correspondence: loll@chemie.fu-berlin.de; brueck@tum.de
†Marion Ringel and Nicole Dimos contributed equally to this manuscript
[1] Werner Siemens Chair of Synthetic Biotechnology, Department of Chemistry, Technical University of Munich, Lichtenbergstr. 4, 85748 Garching, Germany
[2] Institute for Chemistry and Biochemistry, Structural Biochemistry Laboratory, Freie Universität Berlin, Takustr. 6, 14195 Berlin, Germany
Full list of author information is available at the end of the article

Conclusion: The presented product selectivity and titers of Copu5 and Copu9 may pave the way towards a sustainable, biotechnological production of the potentially new bioactive (+)-δ-cadinol. Furthermore, Copu5 and Copu9 may serve as model systems for further mechanistic studies of terpenoid catalysis.

Keywords: δ-cadinol, Sesquiterpene, Basidiomycota, Terpene synthases, Active site architecture, Mutagenesis

Graphical Abstract



Introduction

The rapid emergence of new diseases (e.g., Covid-19) and the excess use of common drugs such as antibiotics continuously forces the scientific community towards the development of new, innovative drug leads. Over the past decades filamentous fungi have mainly been subject to research focused on the biosynthesis of major antibiotic agents, whereas the identification of bioactive terpenoids was largely based on the analysis of secondary metabolites of medicinal plants [1]. Little attention has been paid to filamentous fungi as potential source of new terpenoid-based bioactives. To date terpenoids represent the largest and structurally most diverse group of natural products encompassing over 80,000 characterized compounds [2]. Terpenoids (and in particular the sesquiterpenoid subfamily) are widely used in medicine and health care for their anti-insect, anti-inflammatory, anti-viral, anti-malarial, anti-microbial and anti-tumor activities [3, 4]. A prominent example for a clinically relevant sesquiterpenoid bioactive is artemisinin, which is a first line treatment against malaria [5]. All sesquiterpenoids feature a complex C_{15} carbohydrate skeleton which is formed by the cyclization of the universal, aliphatic precursor farnesyl diphosphate (FPP), a reaction catalyzed by enzymes from the sesquiterpene synthase family. Sesquiterpene synthases typically belong to the class I terpene synthase (TPS) family, commonly exhibiting a α, αβ or αβγ domain architecture for monofunctional enzymes, with the catalytic site located in their respective α-domains. Bifunctional class I TPS also exist, exhibiting additional catalytic functions in either a second α- or the γ-domain (displaying the class I—class I αα or class I—class II αβγ domain architecture) [2]. All class I TPSs share highly conserved sequence motifs in their respective α-domains, such as the aspartate-rich **D**DXX**D** (DD) dyad and the (**N**,**D**)D(L,I,V)X(**S**,**T**)XXX**E** (NSE) triad, with both being involved in complexing of three Mg^{2+} ions that are essential for catalysis [2]. Furthermore, class I TPSs share a WxxxxxRY sequence motif that facilitates the closure of the active site via salt bridge formation

upon substrate binding [6, 7]. In this regard, the *Streptomyces*-derived class I di-TPS CotB2 [8] and the trichodiene synthase from *Fusarium sporotrichioides* [9], yielding the sesquiterpene trichodiene, belong to the best studied TPSs to date [6, 10]. Detailed computational studies on the mechanism employed by trichodiene synthase have highlighted the relevance of a bifacial active site architecture consisting of a highly polar region to promote binding of diphosphate (PP region) and a hydrophobic pocket lined with aromatic amino side chains to guide the propagation of carbocations [10, 11]. Upon the initial substrate binding within the PP region, which also involves the complexation of the diphosphate moiety by the tri-Mg^{2+}-cluster, the active site is closed (induced fit) [7, 10]. In the first step of the chemical reaction, C–O bond cleavage and the abstraction of the PP moiety result in the formation of a farnesyl carbocation intermediate. Subsequently, the carbocation relocates to the hydrophobic pocket; the rate of this rearrangement, and therefore the maturing of the carbocation to the respective cyclized (sesqui-) terpene, is controlled by electrostatic interactions within the active site [11]. The structure of the mature terpene is ultimately defined by the amino acid residues lining the hydrophobic binding pocket [2, 11]. Therefore, detailed knowledge about catalytically essential residues within the active site of TPSs may guide the engineering of such enzymes towards targeted products (e.g., new natural bioactives).

This study reports the identification of two new sesqui-TPSs, Copu5 and Copu9, from the wood rotting fungus *Coniophora puteana* via a genome mining approach. Both enzymes can be classified as class I terpene synthases and catalyze the formation of (+)-δ-cadinol as main product from its direct precursor FPP. (+)-δ-cadinol, which is also known as "torreyol" or "pilgerol" [12], has been subject to numerous studies over the past decades including bioactivity tests [13, 14]; plant extracts containing cadinene-type sesquiterpenes (e.g., δ-cadinol) were shown to have anti-microbial, anti-fungal and anti-inflammatory properties [15, 16]. Furthermore, purified (+)-δ-cadinol exhibited cytotoxic activity against MCF7 cells with an IC_{50} of 3.5 ± 0.58 µg ml^{-1} [17]. To date only two other terpene synthases predominantly forming δ-cadinol have been identified, BvCS from *Boreostereum vibrans* [18] and GME3638 from *Lignosus rhinocerotis* [17] with sequence similarities of 41.4% and 58.3% compared to Copu9, respectively (Additional file 1: Fig. S3). Interestingly, exactly like Coniophora puteana, both organisms belong to the division of Basidiomycota. Analyses of the amino acid sequences and corresponding tertiary structure elements of Copu5 and Copu9, employing protein crystallization and homology modelling techniques, revealed that both enzymes share an almost identical active site architecture. Structure-based mutagenesis was employed to probe the role of catalytically important residues, with a view to provide a platform for product-targeted engineering of TPSs.

Results and discussion

Identification of potential sesquiterpene synthases

A previous study reported the identification and functional characterization of several putative TPSs within the genome of the wood rotting fungi *C. puteana*. Two of these candidates were indeed shown to be efficient and highly selective sesqui-TPSs producing cubebol and β-copaene, respectively [19]. However, the majority of these putative TPSs still await characterization. To gain further insight into the terpenom of *C. puteana*, the genome of *C. puteana* was probed for the presence of additional TPS-like sequences using a Basic Local Alignment Search Tool (BLAST) with the amino acid sequence of the recently identified cubebol synthase Copu3 [19] as reference. Six candidates were identified (Copu5: XP_007765330, Copu6: XP_007773189, Copu7: XP_007767204, Copu9: XP_007765560, Copu10: XP_007766266.1 and Copu11: XP_007767169.1), all of which contain the aspartate-rich DDXXD motif and the NSE triad, which provide ligands for the three essential Mg^{2+} ions in the active site (Additional file 1: Fig. S1 and S2) [2, 7]. The class I TPS-specific WxxxxxRY motif, a component of the induced fit mechanism, is also conserved [6]. Beyond these motifs the level of sequence conservation is considerably lower, with the six newly identified putative TPS sequences sharing only 24–37% sequence similarity with Copu3.

Heterologous expression in *Escherichia coli* and characterization of produced sesquiterpenes

In order to biochemically and functionally characterize the novel TPSs from *C. puteana*, their respective open reading frames (ORFs) were codon-optimized for heterologous expression in *E. coli* and cloned into a single operon expression system as described previously[19]. The employed expression system, also includes the native *E. coli* non-mevalonate pathway (MEP) bottleneck enzymes 1-deoxy-D-xylulose-5-phosphate synthase (DXS; WP_099145004.1) and isopentenyl-pyrophosphate isomerase (IDI; AAC32208.1) to enhance sesquiterpene production as previously described [19, 20]. In addition, the ORFs of the TPS candidates were also cloned into a two-plasmid diterpene production system, which includes a geranylgeranyl diphosphate synthase (crtE

from *Pantoea ananatis*; ADD79325.1) [21], to test for possible catalytic activity towards diterpene production. In preliminary experiments Copu6, Copu7, Copu10 and Copu11 did not show any catalytic activity when tested for sesquiterpene production using the single operon expression vector nor for diterpene production applying the two-plasmid system. Therefore, these four putative TPSs are considered as non-viable protein sequences. Basidiomycota like *C. puteana* are prone to alternative splicing events which especially occur in organisms under stress conditions [22, 23]. Therefore, it is likely that the in silico annotated TPS coding sequences encompass variations of the enzymes, that result in inactive enzyme variants. Active variants of the annotated genes might occur in vivo during splicing events caused by exogenous stress conditions. In contrast, GC–MS analysis of Copu5- and Copu9-expressing *E. coli* extracts showed catalytic activity towards the production of five different sesquiterpenes, when co-expressed with the MEP bottleneck enzymes (Fig. 1). Both cell extracts show typical fragmentation patterns at 105, 119, 161 and 204 m/*z* indicative of non-functionalized, cyclic sesquiterpenes. Moreover, *m/z* values at 105, 119, 161, 204 and 222 suggest that the generated sesquiterpenes are decorated with a single hydroxyl group. Interestingly, Copu5 and Copu9 appear to be very selective showing only one prominent GC–MS signal with a typical mass pattern representative of a mono-hydroxylated sesquiterpene (parent ion 222 m/*z*; RT: 17.09 min). Remarkably, the cyclisation products generated by both Copu5- and Copu9 display the same

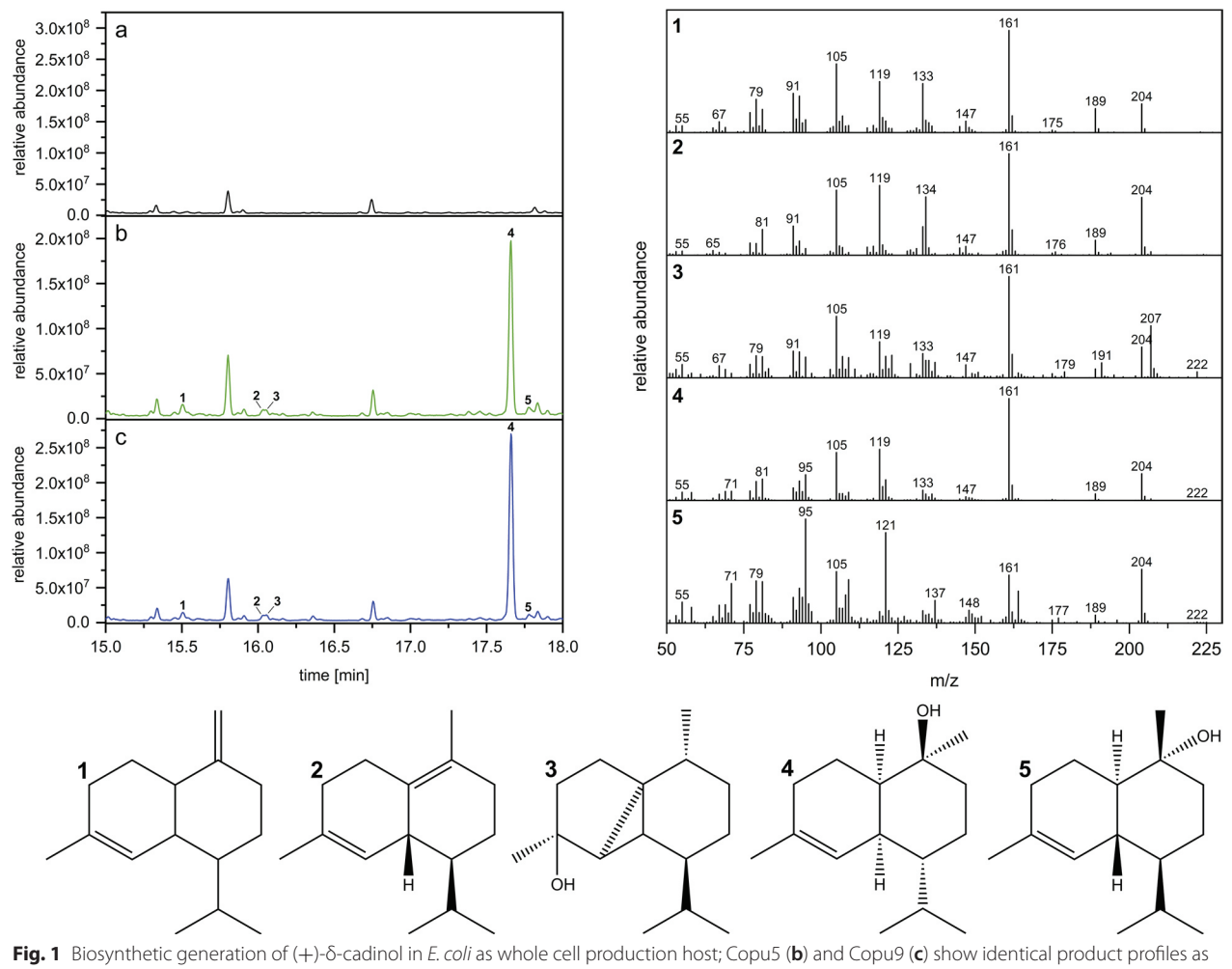


Fig. 1 Biosynthetic generation of (+)-δ-cadinol in *E. coli* as whole cell production host; Copu5 (**b**) and Copu9 (**c**) show identical product profiles as shown in the GC chromatograms when compared to a E. coli extract without the biocatalyst as control (**a**). Detailed GC–MS analysis based on NIST database comparisons allows for compound assignment of (1) tau-muurolene (RT: 15.51 min), (2) delta-cadinene (RT: 16.03 min), (3) cubebol (RT: 16.05 min), (4) (+)-δ-cadinol (RT: 17.66 min) and (5) α-cadinol (RT: 17.78 min)

fragmentation pattern, indicative of formation of identical sesquiterpenes (Fig. 1).

Structure elucidation of generated sesquiterpenes

In order to structurally identify the sesquiterpenes generated by Copu5 and Copu9 a detailed comparison of their mass spectra with the National Institute of Standards and Technology (NIST) database was performed [24]. This evaluation of the *E. coli* extracts expressing Copu5 and Copu9 revealed (+)-δ-cadinol (RT: 17:66 min; parent ion mass 222 m/*z*, major daughter ions at 119, 161 and 204 m/*z*) as the main product (Fig. 1). The identity of this product was confirmed via NMR experiments (Standardized on solvent ($CDCl_3$) peak: $^1H = 7.26$ ppm, $^{13}C = 77.2$ ppm ^1H NMR (500 MHz, $CDCl_3$) δ 5.51 (dq, J = 5.3, 1.6 Hz, 1H), 1.98 (m, J = 13.9, 10.4, 6.2 Hz, 4H), 1.92–1.85 (m, 1H), 1.66 (s, 3H); 1.63–1.45 (m, 5H), 1.35–1.25 (m, 1H), 1.29 (s, 3H), 1.09 (qd, J = 13.2, 4.2 Hz, 1H), 0.88 (d, J = 6.9 Hz, 3H), 0.81 (d, J = 6.9 Hz, 3H); ^{13}C NMR (126 MHz, $CDCl_3$) δ 134.36, 124.61, 72.55, 45.55, 44.08, 36.77, 35.31, 31.14, 27.97, 26.41, 23.66, 21.70, 21.52, 18.51, 15.31.) in conjunction with a comparison to reported NMR data (Additional file 1: Figs. S4–S11) [17]. In addition to the main cyclisation product (+)-δ-cadinol, the sesquiterpenes tau-muurolene (RT: 15.51 min), delta-cadinene (RT: 16.03 min), cubebol (RT: 16.05 min) and α-cadinol (RT: 17.78 min) were putatively assigned as minor products in both extracts as indicated by comparison of GC–MS fingerprint spectra with NIST database references (Fig. 1). Based on these product profiles both Copu5 and Copu9 can be designated as new, highly selective (+)-δ-cadinol synthases. The product selectivity of (sesqui-) TPSs varies significantly within this versatile enzyme family ranging from single product formation (e.g., (+)-δ-cadinene synthase from *Gossypium arboreum*) to a product portfolio of over 50 different compounds (e.g., γ-humulene synthase from *Abies grandis*) [19, 25, 26]. In contrast to the common function of an enzyme as an accelerator of a reaction rate, the catalytic challenge for TPSs rather lies in the control of the highly reactive carbocation intermediates alongside their reaction trajectory [27]. The product distribution in TPSs is guided by several factors such as: (i) the activation of the C–O bond by the pyrophosphate-Mg^{2+}-cluster in the active site, (ii) electrostatic interactions that lead to the sequestration of the active site, (iii) the specific positioning of water molecules or acidic/basic residues, that facilitate site-specific hydroxylations or (de)protonations, and (iv) a specific active site architecture that pre-shapes the carbocation intermediate [11, 27]. For instance, for two fungal sesquiterpene synthases, Cop4 and Cop6 from *Coprinus cinereus*, it was demonstrated that a smaller carbocation binding pocket lead to a more specific product profile as the carbocation intermediate is more restricted along its potential cyclization routes [28, 29]. At present, all functionally characterized sesquiterpene synthases from *C. puteana* show a highly specific product distribution, indicating that their active site architectures may be very effective in restricting the carbocation intermediates, thereby preventing undesired side reactions.

Technical scale production of (+)-δ-cadinol

Mischko and co-workers demonstrated that the TPSs from *C. puteana* have both a high product selectivity as well as high product titers [19]. In order to investigate the performance of the newly identified (+)-δ-cadinol synthases Copu5 and Copu9 in an optimized *E. coli* production host, technical scale, fed-batch fermentation experiments were carried out using a 1.3 L parallel fermentation system as described previously [19]. *Escherichia coli* cultures co-expressing Copu5 and the respective MEP bottleneck enzymes reached stationary phase after 48 h with a final OD_{600} of 88 and a (+)-δ-cadinol titer of 225 mg l^{-1} (Fig. 2). Based on this data a Copu5-specific productivity of 4.7 mg l^{-1} h^{-1} was calculated. In contrast, Copu9-expressing cultures reached a final OD_{600} of 126 and a (+)-δ-cadinol titer of 395 mg l^{-1}, entering

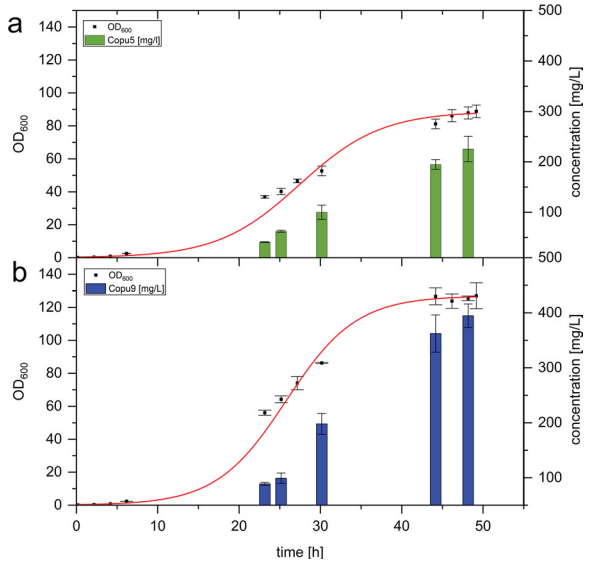


Fig. 2 Growth curves of technical scale for the fed-batch fermentation of **a** Copu5 and **b** Copu9 in respective E. coli production hosts and time-dependent (+)-δ-cadinol titers. Error bars represent the mean values ± standard deviation over triplicates

stationary phase after 48 h (Fig. 2). The calculated productivity of Copu9 was 8.2 mg l^{-1} h^{-1}, respectively. Since equal fermentation parameters were maintained in the fermentation of Copu5- and Copu9-expressing strains, the only varying factor was the used TPS itself. The resulting different biomass (OD_{600}) and δ-cadinol accumulation is thus likely to be a result of a difference in metabolic burden. To date only a few attempts for the biotechnological production of δ-cadinol by microbial hosts have been reported, all of them resulting only in minor yields (no larger than 1 mg l^{-1}) [17, 18], and hence, despite its promising pharmaceutical properties, δ-cadinol is mainly referred to as constituent of various plant extracts used in traditional medicine [15, 30, 31]. Therefore, our titers not only significantly exceed those of previous studies, but also, provide an opportunity to sustainably and scalably generate this highly valuable natural bioactive.

Structural comparison of Copu5 and Copu9

Copu5 and Copu9 exhibit the same product profile (Fig. 1) but differ in their productivities (Fig. 2). The two enzymes share 52.7% sequence identity and a homology of 65.2% (Additional file 1: Figs. S2 and S3). In order to gain insight into residues, that promote the high selectivity of these enzymes, but also the enhanced productivity of Copu9, crystallization trials were carried out. Copu9 could not be crystallized in its open, resting state conformation, but co-crystallization with the non-hydrolysable FPP substrate mimic (4-amino-1-hydroxybutylidene)bisphosphonic acid (alendronate, AHD) and $MgCl_2$ resulted in crystals suitable for the collection of X-ray diffraction data. The obtained structure thus represents the closed, catalytically active Copu9 (Copu9·Mg_3·AHD) conformation. No crystals were obtained for Copu5. In order to further understand the differences of Copu9 and Copu5 a thermal shift assay was performed revealing a significantly lower melting temperature for Copu5 in its apo state as well as bound to AHD [27.6 ± 1.0/33.5 ± 0.8 °C compared to 37.2 ± 0.6/44.3 ± 0.6 °C of Copu9 in pyrophosphate containing Copu5 buffer (Additional file 1 Fig. S15)]. This might be an explanation for the difficulties in purification of Copu5.

Crystals of Copu9·Mg_3·AHD diffracted to a resolution of 1.83 Å (Additional file 1: Table S1). The enzyme forms a homodimer both *in crystallo* and in solution as observed in size exclusion chromatography. Inspection of the electron density clearly revealed bound AHD and the presence of three Mg^{2+} cations (Additional file 1: Fig. S13) in a closed conformation. Both poly-peptide chains are practically identical with a root mean square deviation (rmsd) of 0.32 Å for 330 pairs of Cα atoms. The structure of Copu9 is complete except for its 13 N-terminal residues. While TPSs generally are helical bundle proteins lacking any β-strands, two short β-strands are present in Copu9, one at the N- and one at the C-terminal ends (Fig. 3a, b and Additional file 1: Fig. S12). These two β-strands (T17-L21 and R335-L339) form an antiparallel

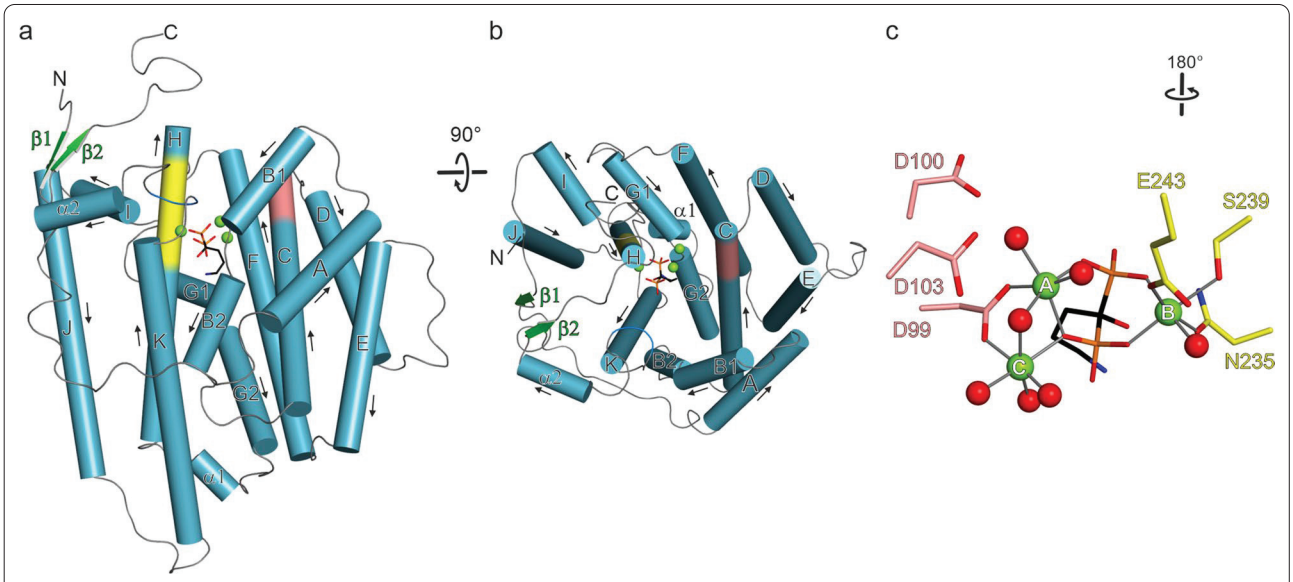


Fig. 3 Overall architecture of one monomer of Copu9; **a** α-helices are drawn as blue cylinders and β-sheets as green arrows. The Asp-rich motif is colored in salmon and the NSE motif in yellow. The three Mg^{2+} ions are shown as green spheres and the AHD in stick representation; **b** View of panel a rotated by 90°, resulting in a view from the top into the active site; **c** detailed view of the Asp-rich and the NSE motif. The three Mg^{2+} ions are octahedrally coordinated by side chains of the catalytic motifs, water molecules and the phosphate functions of AHD

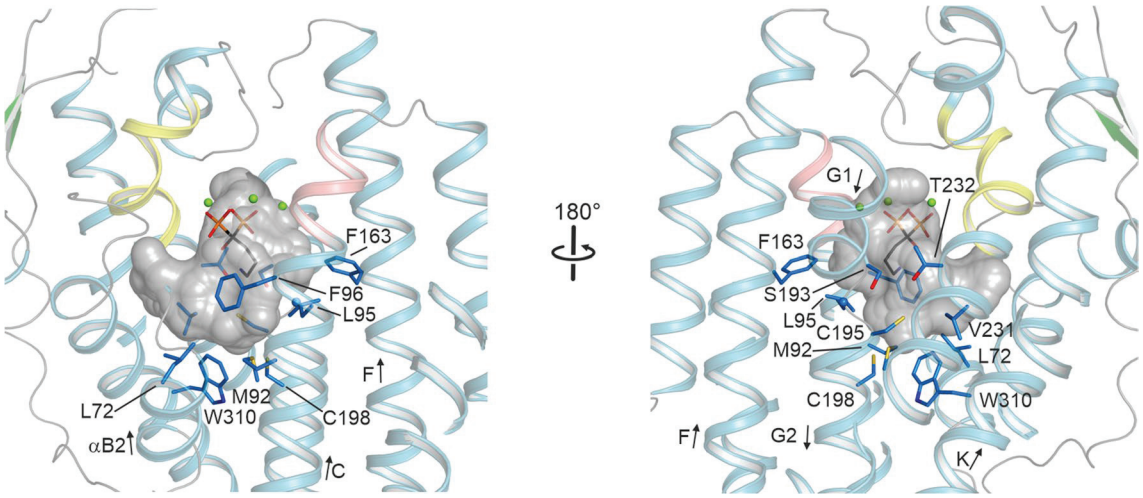


Fig. 4 Overall view on residues lining the active site of Copu9. Same color coding as in Fig. 3. The three Mg^{2+} ions are shown as green spheres and the AHD in stick representation. The cavity of the active site pocket is indicated as an inverted surface displayed in gray

β-sheet, which might further stabilize the closed conformation. Copu9 shows the classical ^{99}DDWLD^{103} (located on α-helix C) and ^{235}NSE^{243} (located on the opposing α-helix H) motifs. The C-terminal ^{317}WxxxxxRY^{324} motif adopts a random coil conformation and folds onto the active site, reflecting the closed conformation of Copu9. Latter conformation of the ^{317}WxxxxxRY^{324} segment is identical as previously observed in CotB2 [6]. Therefore, both side chains of R223 and Y324 point towards the active site. R223 established a bidentate salt-bridge to Asp99 of the Asp-rich motif. R324 forms a hydrogen bond to one phosphate function of AHD. The pyrophosphate sensor [32] R188 is located on α-helix G1 and establishes a bidentate salt-bridge to the second phosphate group of AHD.

The active site is mainly lined by hydrophobic residues: L72, M92, L95, F96, F163, S193, G194, C195, C198, V231, T232 and W310 (Fig. 4). Based on a DALI search [33], the closest structural homologue to Copu9 is Selinadiene synthase (SdS; PDB-ID 4OKM [7]) (Additional file 1: Table S2). The two structures superimpose with a rmsd of 1.45 Å for 296 pairs of Cα atoms.

Since we could not obtain an experimental structure of Copu5, we predicted the structure by the ROBETTA server [34]. To validate the prediction, we initially predicted the structure of Copu9. The obtained model, in its open, inactive conformation superimposes very well, with the experimentally obtained structure of Copu9·Mg2 + 3·AHD (Additional file 1: Fig. S16 and Table S3). The largest differences in the protein backbone are observed in the N- and C-terminal extensions of the protein (Additional file 1: Fig. S16). As anticipated, the active site of the modeled structure is wider, due to the absence of the Mg2+ ions as well as alendronate, since both α-helices harboring the metal binding motifs are tilted away from the active side. A similar observation is made in the model of Copu5 that largely resembles the fold of Copu9 (Additional file 1: Fig. S16b, c and Table S3). The amino acid sequences in their hydrophilic PP binding pockets are highly conserved (DDXXD: Copu5: ^{92}DD**W**SD^{96}, Copu9: ^{99}DD**W**LD^{103}; NSE: Copu5: ^{227}NDV**F**S**YNK**E^{235}, Copu9: ^{235}NDI**F**S**YNK**E^{243}; WxxxxxRY: Copu5: ^{309}W**SF**ETERY^{316}, Copu9: ^{317}W**SF**DSHRY^{324}) (Additional file 1: Fig. S2). Furthermore, the 12 residues involved in either pre-shaping the geometry or the propagation of the carbocations in the hydrophobic pocket of the active site are identical in the two enzymes (Additional file 1: Fig. S12 and Table S4) which is likely to be the cause of their identical catalytic activity. To the best of our knowledge Copu5 and Copu9 are the first reported TPSs from the same organism with the same product profile and an almost equivalent active site decoration. To this end, there are two residues located in the second shell of residues lining the active site and thus merely surrounding first shell residues, which are different between Copu9/Copu5: namely (F91/Y84), and (C198, V190), respectively. By contrast, both Copu5 and Copu9 only share four of the 12 relevant residues in the hydrophobic pocket with the previously reported cubebol synthase Copu3 [19] (Additional file 1: Fig. S12 and Table S4). The remaining eight amino acid side chains are thus likely to play an important role in guiding the product profiles of Copu3 and Copu5/Copu9.

Structure-based mutagenesis targeting active site residues

In order to evaluate their roles, each of them was iteratively changed in Copu5 and Copu9 to their counterpart present in Copu3. As the main product of Copu3 is cubebol [19], which is also produced in minor amounts by Copu5 and Copu9 (Fig. 1), the mutations introduced into Copu5 and Copu9 were anticipated to shift their product profile towards that of Copu3, i.e., generation of cubebol as the main cyclization product (Fig. 5).

The Copu5 variants T66C, C69V and T184N and the Copu9 variants T73C, C76V, T192N, G194A, C195V and P197C did not affect their product spectrum. In contrast, Copu5 variants G186A, C187V and S306N and Copu9 variants N228A and S314N showed increased synthesis of minor products (Fig. 1, Additional file 1: Fig. S17 and S18), while the P189C variant of Copu5 exhibited lower product formation indicating, that this residue is either essential for catalysis or interferes with the catalytically active, closed conformation. Notably, Copu5 variants C187V, N220A and S306N as well as Copu9 variants N228A and S314N showed the formation of an additional side product, germacrene D-4-ol (Additional file 1: Figs. S17–S19; RT: 16.85 min; identified by a comparison to the NIST database [24]). Interestingly, germacrene D-4-ol is also a side product of Copu3.[12].

In order to further evaluate the influence of the conducted point mutations on the synthases' catalytic properties in vitro kinetic experiments were performed. All kinetic parameters obtained from Copu9 and its variants reaction with FPP are listed in Table 1 (Fig. 6, Additional file 1: Fig. S21). Copu9 WT and all variants show comparable binding affinity (K_m) towards FPP. However, Copu9 variants C76V, N228A and S314N show a minor decrease in catalytic turnover (k_{cat}) compared to Copu9 WT, while variants T73C, T192N, G194A, C195V and P197C either retain WT k_{cat} or show a slightly increased catalytic turnover compared to Copu9 WT (Table 1, Fig. 6). Copu9 WT and all variants also show comparable catalytic efficiency (k_{cat}/K_m). However, variant C76V, which showed a minor decrease in catalytic turnover, exhibits slightly higher substrate specificity. Interestingly, the variants N228A and S314N, show a tendency towards decreased substrate specificity which possibly reflects the minor changes in their product spectrum. All kinetic parameters observed for Copu9 WT and its variants are within range of the respective kinetic constants of the kinetically well characterized fungal sesquiterpene synthases Cop4 and Cop6 from *Coprinus cinereus* [29]. In contrast to Copu9, it was only possible to purify Copu5 using pyrophosphate containing buffers due to its significant in-vitro stabilizing effect. Hence, we were not able to determine the in-vitro kinetics of Copu5 and its variants.

However, none of the variants neither significantly changed the product profile towards another major product, such as cubebol, nor showed a drastic change in its respective catalytic properties. Copu5 appears to be more receptive to single amino acid changes whereas Copu9 largely compensates mutations and retains its

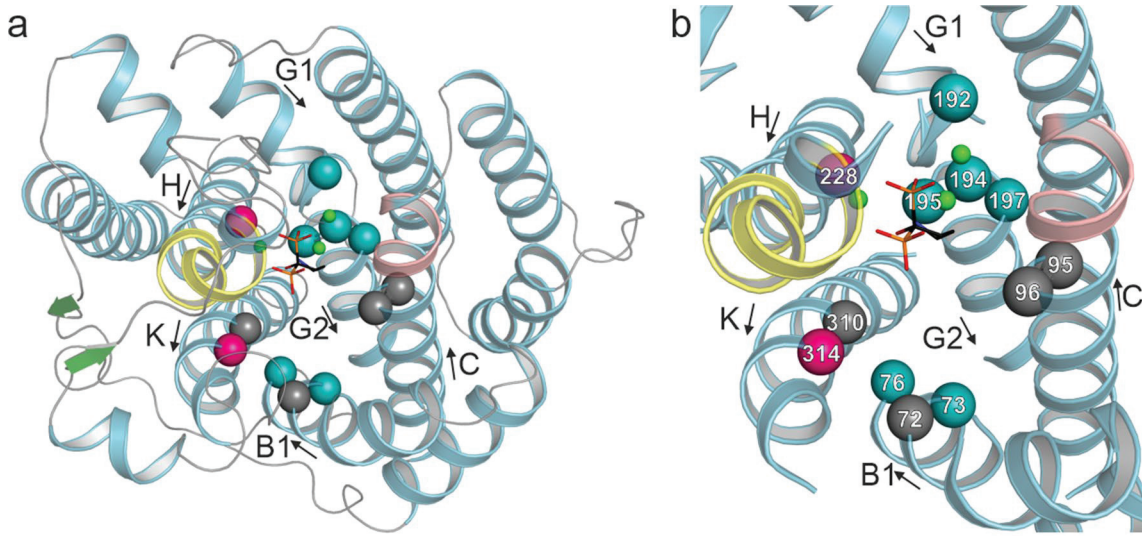


Fig. 5 Location of residues subjected to mutagenesis. **a** The same perspective and color-coding as in Fig. 3b were used. α-Helices are drawn as spirals. Identical amino acids between Copu3 and Copu9 are indicated as grey spheres. Residues that have been subjected to mutagenesis without altering the product are shown as teal-colored spheres. Single amino acid exchanges with an effect on the product profile are drawn as magenta-colored spheres. **b** Magnified view into the active site (same perspective and color coding as in panel a), with residues numbered

Table 1 Steady-state kinetic parameters of Copu9 and its variants calculated from the EnzChek™ pyrophosphate assay

Copu9	k_{cat} [s^{-1}]	K_m [μM]	k_{cat}/K_m ($\times 10^3$) [$s^{-1}\,M^{-1}$]
WT	$(3.24 \pm 0.03) \times 10^{-2}$	4.59 ± 0.17	7.067
T73C	$(5.02 \pm 0.26) \times 10^{-2}$	6.28 ± 1.37	7.995
C76V	$(2.91 \pm 0.07) \times 10^{-2}$	3.14 ± 0.45	9.252
T192N	$(3.96 \pm 0.08) \times 10^{-2}$	8.28 ± 0.60	4.782
G194A	$(3.60 \pm 0.05) \times 10^{-2}$	7.00 ± 0.37	5.135
C195V	$(3.98 \pm 0.07) \times 10^{-2}$	8.48 ± 0.51	4.692
P197C	$(3.81 \pm 0.10) \times 10^{-2}$	6.82 ± 0.77	5.584
N228A	$(2.62 \pm 0.07) \times 10^{-2}$	4.42 ± 0.59	5.940
S314N	$(1.92 \pm 0.05) \times 10^{-2}$	5.10 ± 0.54	3.755

wild-type product profile. Recently, it was demonstrated that mutations in loop regions distinct from the active site may result in drastic catalytic differences in related synthases [28]. Considering the proposed cyclization mechanism required to form cadinene-type sesquiterpenes (Additional file 1: Fig. S20) every mutation carried out for Copu5 and Copu9 would be anticipated: (i) to have an effect (stabilizing or destabilizing) on the reaction of the carbocation intermediates as well as, (ii) to allow or restrict water molecule(s) to enter the active site [19, 26, 35]. In particular, residues C187, N220 and S306 in Copu5 and corresponding residues in Copu9 (N228 and S314) appear to have an impact on controlling the carbocation reaction trajectory of the germacryl cation, which is the common intermediate of the cadinene and germacrene cyclization trajectories (Additional file 1:Fig. S20) [19, 26, 29]. However, the impact of single amino acid substitutions in Copu5 and Copu9 is relatively small when compared to other TPSs (e.g., CotB2) [36–39]. In that context, residues lining the entrance of the active site (e.g., residues G186 and C187 in Copu5) affect product specificity.

The directed modulation of a product profile towards a defined product was previously demonstrated for (+)-δ-cadinene synthase from *G. arboreum* [26]. The main product of that synthase was altered towards the production of germacrene D-4-ol by specific mutations within the G1 and G2 α-helices in the hydrophobic pocket of the active site of this enzyme. These helices and the small linker region connecting them, were shown to be important for product specificity in class I TPSs [7, 26, 40]. Upon substrate binding this segment is subject to a conformational change triggered by a conserved effector triad, which was first identified in the C-terminal end of the G1 α-helix of SdS (residues Arg178, Asp181 and Gly182 [7]). This induced-fit mechanism triggers substrate ionization and therefore represents the starting point of the subsequent carbocation cyclization trajectory towards germacrene D-4-ol. By a superimposition of the SdS and Copu9 structures in conjunction with an inspection of a sequence alignment that also includes Copu5, this effector triad could also be identified in the C-terminal ends of the G1 α-helices of Copu5 (R182, D183 and S185) and Copu9 (R190, D191 and S193), respectively. However, instead of the commonly found glycine at the third position of this triad, both Copu5 and Copu9 employ a serine residue (S185 and S193, respectively). Single mutations in the vicinity of the effector triad of Copu9 (i.e., T192, G194, C195 and P197) have no significant effect on the catalytic properties nor product profile of that enzyme (vide supra). However, mutations of the corresponding glycine (G186) and cysteine (C187) residues in Copu5, both located in the linker region between the α-helices G1 and G2, alter the product profile, with the G186A variant also promoting the formation of germacrene D-4-ol (Additional file 1: Figs. S15–S17). This glycine and cysteine residues are the closest residues within the hydrophobic pocket to the phosphate moiety of the bound substrate (see G194 and C195 in Fig. 5).

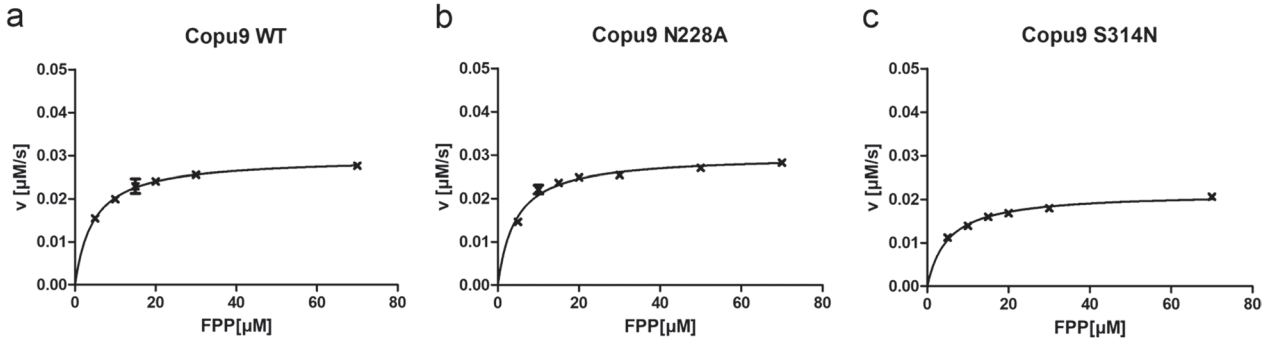


Fig. 6 In vitro Michaelis–Menten kinetics of Copu9 WT (**a**) and the variant N228A (**b**) and S314N (**c**) using the EnzChek pyrophosphate Assay. A non-linear regression analysis was performed on the data collected from the time-resolved steady- state kinetic assay

Therefore, the results presented herein point out: (i) that the serine residue in the effector triad may be important in promoting the production of (+)-δ-cadinol, but also (ii) that there are likely additional tertiary structural elements, that influence the product profile of sesqui-TPSs, possibly by affecting the interface between the hydrophilic and hydrophobic regions in the active site, and/or by interaction with the linker region between α-helices G1 and G2. This is in agreement with studies exploring the influence of domain-domain interactions on the catalytic function of TPSs [41, 42]. It was shown that different bifunctional TPSs exhibit varying degrees of domain-domain interdependence regarding their catalytic activity. Upon separation of the α and βγ domain in Ent-kaurene synthase from *Phaeosphaeria* sp. [42], the single domains still showed catalytic activity but catalytic activity was decreased by 30-fold compared to the wild-type full-length protein. Taxadiene synthase from *Taxus brevifolia* also showed severely compromised catalytic activity in its α-domain when being separated from the βγ-domain [43]. In contrast, the separation of the same domains in abietadiene synthase from *Abies grandis* [41] led to a complete loss of function of the separate domains. Further studies on domain-domain chimeras of fusicoccadiene synthase from *Phomopsis amygdala* as well as ophiobolin F synthase from *Aspergillus clavatus* showed significantly altered cyclization fidelity and catalytic activity compared to the wild-type enzymes [43]. The architecture of multi-domain TPS might contribute to the overall stability of the proteins. Moreover, tertiary structure interactions play a significant role in shaping the active site for precise chemical control along the carbocation reaction trajectory. Especially for residues involved in conformational changes during catalysis or flexible tertiary structure elements this might play a bigger role than hitherto expected.

Conclusion

The genomes of terrestrial and marine organisms bear an enormous and at present widely uncharacterized capacity of TPSs mediating the formation of natural bioactives [44, 45]. Similarly, although fungi are known for their broad variety of bioactive terpenoids, fungal TPSs have not yet been explored in depth for their potential for biotechnology applications [35]. In this study six novel putative TPSs from the Basidiomycota *C. puteana* were characterized. Only two of them, Copu5 and Copu9, were functional and identified as efficient and highly product selective (+)-δ-cadinol synthases (Fig. 1). Notably, these two synthases are the first TPSs that originate from the same organism and yet have virtually identical product profiles. The observed product selectivity of both synthases is likely to be a results of highly conserved sequence motifs and a carefully and equally decorated active site. Both display excellent production and productivity for cadinol, exceeding currently available biosynthesis systems by far (Fig. 2).

The crystal structure of Copu9 in complex with three catalytically essential Mg^{2+} ions and the substrate mimic AHD was solved to a resolution of 1.83 Å (Fig. 3). This structure is the first of a class I TPS from Basidiomycota. Structure-informed mutations in the hydrophobic pockets of the active sites provided proof-of-concept, that single amino acid changes can alter the product profile of Copu5 and Copu9 from (+)-δ-cadinol to germacrene D-4-ol. However, despite having identical residues in the hydrophobic substrate binding pocket, Copu5 appears to be more flexible towards changing its product profile than Copu9. For Copu5 mutations targeting the linker between the α-helices G1 and G2 affected the product spectrum (Additional file 1: Fig. S12). This structural segment has previously been shown to be important for the product specificity in class I TPSs [7, 26, 40]. To this end, the herein presented results in conjunction with the results previously shown for a (+)-δ-cadinene synthase [26] suggest that the helix-turn-helix motif may be a hotspot for product shifting mutations and thus serve as a guiding significance for the improvement of related class I terpene synthases. The question remains why Copu9, unlike other members of this class of enzymes, appears to be less affected by mutations in this region. This observation may suggest a structural robustness of Copu9 and may also be connected to its high product yield and specificity when expressed in *E. coli*. It is likely, that additional structural elements and associated conformational changes play an important role in modulating the catalytic performance and product profile of class I TPSs. In this respect, Copu5 and Copu9 present ideal model systems to enhance mechanistic insight into this important class of enzymes for applications in biotechnology and synthetic biology.

Experimental section

General

All media components and HPLC grade chemicals were purchased from Roth chemicals (Karlsruhe, Germany). Technical grade solvents were obtained from Westfalen AG (Münster, Germany). $CDCl_3$ and Benzene-d_6 were purchased from Sigma-Aldrich (St. Louis, USA).

Gene cloning, plasmid construction and culture condition

Escherichia coli strain DH5α was used for all cloning steps and *E. coli* strain HMS174 (DE3) for terpene production. Genes encoding putative TPSs Copu5 (XP_007765330), Copu6 (XP_007773189), Copu7 (XP_007767204), Copu9

(XP_007765560), Copu10 (XP_007766266.1) and Copu11 (XP_007767169.1) from *C. puteana* were codon-optimized for *E. coli* using the GeneOptimizer™ software and subsequently synthesized by Eurofins Genomics GmbH (Ebersberg, Germany). For sesquiterpene production all genes were cloned into a pACYC-based vector system containing a single operon with selected bottleneck enzymes of the MEP pathway as previously described [19] and the respective TPS all set under the control of a *lac*-I-derived constitutive promoter [21]. For determination of potential catalytic activity towards diterpene production all genes were cloned into a pACYC-based expression vector system and co-transformed with a plasmid containing essential bottleneck enzymes for diterpene production as previously described [21]. All cloning experiments were performed according to standard protocols.

Cultures were grown in modified R-Media [46] (13.3 g l^{-1} KH_2PO_4, 4.0 g l^{-1} $(NH_4)_2HPO_4$, 1.7 g l^{-1} citric acid, 5.0 g l^{-1} yeast extract, 4.88 ml l^{-1} 1 M $MgSO_4$, 2.45 ml l^{-1} 0.1 M Fe(III) citrate, 1.00 ml l^{-1} 100 × Trace Element Solution (5.0 g l^{-1} EDTA, 84 mg l^{-1} $ZnCl_2$, 13 mg l^{-1} $CuCl_2$*2 H_2O, 10 mg l^{-1} $CoCl_2$*2 H_2O, 10 mg l^{-1} H_3BO_3, 1.6 mg l^{-1} $MnCl_2$*4 H_2O) and 30 g l^{-1} glycerol at 30 °C and 100 rpm shaking. The appropriate antibiotics kanamycin (50 µg ml^{-1}), chloramphenicol (35 µg ml^{-1}) or ampicillin (100 µg ml^{-1}) were added as needed.

Fermentation

Fermentation on a technical scale was performed using a DASGIP® 1.3 L parallel reactor system (Eppendorf AG, Germany) with a modified R-media as described above. An overnight preculture was used for the inoculation of the fermenters (OD_{600} = 0.1). Cultivation temperature was kept constant at 30 °C. Initial stirring velocity and air flow were set to 200 rpm and 0.2 volumes of air per volumes of medium per minute (vvm), respectively. Dissolved oxygen was kept constant at 30% and maintained by a gradual increase in stirring velocity (max. 1000 rpm), oxygen content (max. 100%) and airflow (max. 0.8 vvm) during fermentation. A pH value of 7.0 was controlled by the addition of 25% aqueous ammonia. A pH-based feeding protocol was set as previously described [19, 21]. The feeding solution consisted of 600 g l^{-1} glycerol, 5 g l^{-1} yeast extract, 35 g l^{-1} collagen, 20 g l^{-1} MgSO4, 0.3 g l^{-1} thiamine-HCl, 5 ml l^{-1} 1 M ammonium iron(III) citrate, 20 ml l^{-1} 100 × trace element solution (pH = 7.0) [19, 47].

Terpene extraction

To extract terpenes during the screening process 20 ml of the *E. coli* culture broth were mixed with 20 ml of an extraction solution (ethanol, ethyl acetate and hexane; 1:1:1). The mixture was shaken for 4 h at room temperature and subsequently centrifuged down for 5 min at 8000 rpm to separate the organic phase. A sample from the organic phase was then analyzed via GC–MS.

The cultivation broth from either large-scale shaking flask experiments (1 l) or fermentation using a DASGIP® 1.3 l parallel reactor system (Eppendorf AG, Germany) was extracted by adding the same volume of ethanol. This mixture was shaken on a rotary shaker (80 rpm) at 20 °C for 12 h. Subsequently ½ volume of ethyl acetate was added and shaken for 3 h (20 °C, 80 rpm) followed by a centrifugation step for 15 min at 7000*g* to separate the supernatant from the cell debris. Afterwards the ½ volume of hexane was added to the supernatant and the extraction was carried out for 3 h (20 °C, 80 rpm). The organic phase was separated using a separation funnel and subsequently concentrated using a rotary evaporator.

Terpene purification

The crude extract was evaporated until only the oily resin remained. The resin was dissolved in 10 ml of hexane. Subsequently, flash chromatography was carried out to separate the terpene fraction from fatty acid residues using the flash chromatography system PLC 2250 (Gilson, USA) equipped with a Luna 10 µm silica (2) 100 A column at a flow rate of 10 ml min^{-1}. Peaks were detected using an Evaporative Light Scattering Detector (ELSD) flushed with nitrogen gas and a diode array detector at 40 °C. The following gradient was applied: 100% hexane (solvent A) for 15 min, followed by a rapid change (within 3 s) to 100% EtOAc (solvent B), a 15 min wash with solvent B, a return to 100% solvent A within 3 s and a final wash with that solvent for further 30 min. Fractions of interest were reduced to approximately 2 ml using a rotary evaporator and mixed with acetonitrile. Subsequently, the residual hexane was evaporated until only acetonitrile (ACN) remained.

For further purification of the products, the samples were injected into an Ultimate 3000 UHPLC system (Thermo Scientific, USA) containing a binary pump, a diode array detector, an automated fraction collector, and a Jetstream b1.18 column oven. The purification of the respective sesquiterpenes was carried out on a NUCLEODUR® C18 HTec 250/10 mm and guard column holder 8 mm (Machery-Nagel GmbH & Co. KG, Germany) at 30 °C and a flowrate of 2.2 ml min^{-1} using H_2O and ACN as solvents. The following gradient was applied: 90% ACN for 0.5 min, increased to 100% ACN within 10 min to remain for 12 min, decrease to 90% ACN within 0.1 min to remain for another 10 min. Fractions containing the sesquiterpene of interest were evaporated under low nitrogen flow to dryness and subsequently dissolved

in the solvent of interest (hexane for GC–MS analysis or $CDCl_3$ for NMR analysis).

Analytics

Analysis and quantification of terpenes was performed using a Trace GC–MS Ultra system with DSQII (Thermo Scientific, USA). The sample (1 µl, 1/10 split) was injected by a TriPlus auto sampler onto a SGE BPX5 column (30 m, I.D 0.25 mm, film 0.25 µm) with an injector temperature of 280 °C. Helium was used as carrier gas with a flow rate of 0.8 ml min^{-1}. Initial oven temperature was set to 50 °C for 2 min. The temperature was increased to 320 °C at a rate of 10 °C/min^{-1} and then held for 3 min. MS data were recorded at 70 eV (EI) in positive mode in a range between 50 and 650. GC-FID analysis was carried out accordingly. Quantification of sesquiterpenes was carried out by correlation of the FID peak area to a defined α-humulene standard of known quantity as previously described [19].

Purified compounds for further NMR analysis were dissolved in $CDCl_3$. ^{13}C NMR spectra were measured with a Bruker Avance-III 500 MHz spectrometer equipped with a cryo probe head (5 mm CPQNP, $^{1}H/^{13}C/^{31}P/^{19}F/^{29}Si$; Z-gradient). ^{1}H NMR spectra as well as 2D experiments (HSQC, HMBC, COSY, NOESY) were obtained on an Avance-I 500 MHz system with an inverse probehead (5 mm SEI; $^{1}H/^{13}C$; Z-gradient). The temperature was set to 300 K. Resulting data were processed and analyzed by TOPSPIN 3.2 or MestreNova 11.0. Chemical shifts were given in ppm relative to $CDCl_3$ ($\delta = 7.26$ ppm for ^{1}H and $\delta = 77.16$ ppm for ^{13}C spectra).

Protein expression and purification for crystallization experiments

The codon optimized genes encoding Copu5 and Copu9 were fused to an N-terminal hexa-histidine-tag in a pET-M11 vector and transformed into *E. coli* BL21 RIL DE3. Overexpression was performed using auto-induction medium at 37 °C until an OD ~0.7 was reached and subsequently cooled down to 18 °C [48]. Cells grew 48 h and were harvested by centrifugation (10 min, 6000 rpm at 4 °C). For resuspension of the cell pellets, buffer A was used (Copu9: 20 mM Tris/HCl pH 7.5, 500 mM NaCl, 5 mM $MgCl_2$; Copu5: 100 mM Tris/HCl pH 7.5, 500 mM NaCl, 5 mM $MgCl_2$, 10% (w/v) glycerol, 10 mM sodium pyrophosphate, 1 mM DTT). Cells were lysed by homogenization at 4 °C and the lysate was cleared by centrifugation (1 h, 21,000 rpm at 4 °C). Ni^{2+}-NTA beads (cv ~1 ml; GE Healthcare) were equilibrated with buffer A. Copu9 was loaded on the column and washed with 10 cv of buffer A containing additional 30 mM imidazole. Copu9 was eluted with buffer A containing 250 mM imidazole. Size exclusion chromatography was performed with a HighLoad Superdex S200 16/60 column (GE Healthcare), equilibrated with buffer B [(Copu9: 20 mM Tris/HCl pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$; Copu5: 100 mM Tris/HCl pH 7.5, 250 mM NaCl, 5 mM $MgCl_2$, 10% (w/v) glycerol, 10 mM sodium pyrophosphate, 1 mM DTT)]. Pooled protein fractions were concentrated with an Amicon-Ultra 30,000 cell. Calibration runs were performed with the high molecular weight standard (GE Healthcare).

Thermal shift assay

Melting temperatures of Copu9 variants and Copu5 were measured with the Mx3005P qPCR system (Agilent) in a 96-well plate format with and without alendronate. Each well contained 8 µl SEC buffer (either of Copu9 or Copu5, as stated in the manuscript), 10 µl protein (0.15 µg $µl^{-1}$) with 1× SYPRO Orange dye (Invitrogen) end concentration and either 2 µl water or 2 µl alendronate (0.6 mg/ml) dissolved in water. The program consisted of three steps: step 1 was a pre-incubation for 1 min at 20 °C, and steps 2 and 3 were cycles comprising the temperature increase of 1 °C within 20 s. The temperature gradient proceeded from 25 to 95 °C at 1 °C per minute. Samples were measured in triplicates. The data was acquired with MxPro QPCR software (Agilent, Germany) and analyzed with DSF Analysis v3.0.1 tool (ftp://ftp.sgc.ox.ac.uk/pub/biophysics) and Graphpad Prism 5.0.0.228 (Graph Pad Software Inc.). A t-test was performed with Graphpad Prism to validate the significance of the results.

In-vitro kinetics

Time resolved kinetics were measured using the EnzChek™ pyrophosphate assay kit (Invitrogen™, ThermoFisher Scientific) in a 96-well-plate with a reduced volume of 200 µl. The assay was performed as stated in the manual using 0.8–1.2 µM of the Copu9 variants and FPP (Sigma-Aldrich) concentrations ranging from 0 to 70 µM dissolved in methanol. Everything except the substrate was mixed and preincubated at 25 °C for 10 min. The substrate was added shortly before measurement and mixed for 5 s prior the first measurement. The enzymatic reaction was performed at 25 °C and the absorption was measured at 360 nm in 30 s increments using a CM Spark plate reader (Tecan, Germany). The data was analyzed using Graphpad Prism 5.0.0.228 (Graph Pad Software Inc.).

Crystallization

For co-crystallization experiments, Copu9 was concentrated to 28 mg ml^{-1} as measured by the absorbance at 280 nm and incubated with a 10 -fold molar excess of AHD for 30 min on ice. Initial crystals were obtained by the sitting-drop vapor-diffusion method at 18 °C with a

reservoir solution composed of 15.0% (w/v) polyethylene glycol 3350, 100 mM Tris/HCl at pH 8.5 and 100 mM Mg formate. Initial, inter-grown crystals were used to prepare a seed stock. With a cat whisker, seeds were transferred to a freshly prepared crystallization drop and crystallization plates were subsequently stored at 4 °C. Prior to flash-freezing in liquid nitrogen, crystals were transferred to a cryo-protectant solution composed of the reservoir solution supplemented with 25% (v/v) ethylene glycol.

Structure determination and refinement

Synchrotron diffraction data were collected at the beamline 14.1 of the MX Joint Berlin laboratory at BESSY (Berlin, Germany). Diffraction data were processed with XDS [49] (Additional file 1: Table S1). The structure was determined by molecular replacement using the coordinates of SVS_A2 (PDB-ID: 6TJZ [50]) as search model using the PHASER software [51]. The structure was refined by maximum-likelihood restrained refinement in PHENIX Model building [52, 53] and water picking was performed with COOT [54]. Model quality was evaluated with MolProbity [55] and the JCSG validation server (JCSG Quality Control Check v3.1). Secondary structure elements were assigned with DSSP [56] and ALSCRIPT [57] was used for secondary structure based sequence alignments. Figures were prepared using PyMOL (Schrödinger, Inc.). The DALI server [33] was used to identify structures closely related to Copu9.

Protein modelling and phylogenetic analysis

Sequence alignments were performed using Clustal Omega [58] by employing seeded guide trees and HMM profile techniques as previously described [19]. To predict the tertiary structure of Copu5 and Copu9 the ROBETTA server was used [34]. The primary sequence of both proteins was submitted, and the server was run with standard settings. All calculated models have been analyzed and verified by SAVES 6.0 (https://saves.mbi.ucla.edu/) which were subsequently analyzed within the UCSF Chimera environment [59, 60]. For further comparison of the sesquiterpene synthases sequences AliView [61] was used.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12934-022-01791-8.

Additional file 1. Additional figures and tables.

Author contributions

MR conceived the project, designed, and performed the experiments, analyzed the data, and prepared the manuscript. ND and BL performed protein purification, crystallization and kinetic experiments, performed diffraction experiments, refined the structure, and contributed to manuscript preparation. SH performed protein purification, TSA measurements and kinetic experiments. MH supported the analytical work. CH and WE contributed to structural analysis with NMR expertise. GS contributed with scientific expertise in data analysis. TB supervised the work and contributed to manuscript preparation and finalization. MR and ND contributed equally to this manuscript. All authors read and approved the final manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL. MR, MH, and TB gratefully acknowledge funding by the Werner Siemens foundation for establishing the field of Synthetic Biotechnology at TUM. ND and BL acknowledge access to beamlines of the BESSY II storage ring (Berlin, Germany) via the Joint Berlin MX-Laboratory sponsored by the Helmholtz Zentrum Berlin für Materialien und Energie, the Freie Universität Berlin, the Humboldt-Universität zu Berlin, the Max-Delbrück-Centrum, the Leibniz-Institut für Molekulare Pharmakologie and Charité—Universitätsmedizin Berlin. CH and WE would like to acknowledge the support of the German ministry for Education and Research (BMBF; grant number 031A305A and 031B0823D). We are grateful to Gunther Stier for sharing the pET-M11 vector. We are grateful to Elvenstar's Pukipon for the kind donation of cat whiskers. The graphical abstract was created with BioRender.com. We acknowledge C. Freund for access to the plate reader.

Availability of data and materials

All data generated of analyzed during this study are included in this article and the supporting information. Coordinates and structure factors have been deposited in the PDB (7OFL). Diffraction images have been deposited at proteindiffraction.org (https://doi.org/10.18430/m37OFL).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Werner Siemens Chair of Synthetic Biotechnology, Department of Chemistry, Technical University of Munich, Lichtenbergstr. 4, 85748 Garching, Germany. [2]Institute for Chemistry and Biochemistry, Structural Biochemistry Laboratory, Freie Universität Berlin, Takustr. 6, 14195 Berlin, Germany. [3]Bavarian NMR Center - Structural Membrane Biochemistry, Department of Chemistry, Technical University of Munich, 85748 Garching, Germany. [4]School of Chemistry and Molecular Biosciences, The University of Queensland, 68 Cooper Rd, Brisbane 4702, Australia.

Received: 14 September 2021 Accepted: 7 April 2022
Published online: 19 April 2022

References

1. Demain AL, Martens E. Production of valuable compounds by molds and yeasts. J Antibiot. 2017;70:347–60. https://doi.org/10.1038/ja.2016.121.
2. Christianson DW. Structural and chemical biology of terpenoid cyclases. Chem Rev. 2017;117:11570–648. https://doi.org/10.1021/acs.chemrev.7b00287.
3. Jaeger R, Cuny E. Terpenoids with special pharmacological significance: a review. Nat Prod Commun. 2016;11:1934578X1601100946.
4. Joshee N, Dhekney SA, Parajuli P. Medicinal plants: from farm to pharmacy. Cham: Springer; 2019.
5. Meshnick SR. Artemisinin: mechanisms of action, resistance and toxicity. Int J Parasitol. 2002;32:1655–60. https://doi.org/10.1016/S0020-7519(02)00194-7.

6. Driller R, Janke S, Fuchs M, Warner E, Mhashal AR, Major DT, et al. Towards a comprehensive understanding of the structural dynamics of a bacterial diterpene synthase during catalysis. Nat Commun. 2018;9:3971. https://doi.org/10.1038/s41467-018-06325-8.
7. Baer P, Rabe P, Fischer K, Citron CA, Klapschinski TA, Groll M, Dickschat JS. Induced-fit mechanism in class I terpene cyclases. Angew Chem Int Ed Engl. 2014;53:7652–6. https://doi.org/10.1002/anie.201403648.
8. Janke R, Görner C, Hirte M, Brück T, Loll B. The first structure of a bacterial diterpene cyclase: CotB2. Acta Crystallogr D Biol Crystallogr. 2014;70:1528–37. https://doi.org/10.1107/S1399004714005513.
9. Rynkiewicz MJ, Cane DE, Christianson DW. Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade. Proc Natl Acad Sci. 2001;98:13543–8. https://doi.org/10.1073/pnas.231313098.
10. Dixit M, Weitman M, Gao J, Major DT. Chemical control in the battle against fidelity in promiscuous natural product biosynthesis: the case of trichodiene synthase. ACS Catal. 2017;7:812–8. https://doi.org/10.1021/acscatal.6b02584.
11. Major DT. Electrostatic control of chemistry in terpene cyclases. ACS Catal. 2017;7:5461–5. https://doi.org/10.1021/acscatal.7b01328.
12. Borg-Karlson A-K, Norin T, Talvitie A. Configurations and conformations of torreyol (δ-cadinol), α-cadinol, T-muurolol and T-cadinol. Tetrahedron. 1981;37:425–30. https://doi.org/10.1016/S0040-4020(01)92031-9.
13. Kundu A, Saha S, Walia S, Shakil NA, Kumar J, Annapurna K. Cadinene sesquiterpenes from *Eupatorium adenophorum* and their antifungal activity. J Environ Sci Health B. 2013;48:516–22. https://doi.org/10.1080/03601234.2013.761921.
14. Bande-Borujeni S, Zandi-Sohani N, Ramezani L. Chemical composition and bioactivity of essential oil from *Eucalyptus occidentalis* leaves against two stored product pests. Int J Trop Insect Sci. 2018;38:216–23. https://doi.org/10.1017/S1742758418000085.
15. Mulyaningsih S, Youns M, El-Readi MZ, Ashour ML, Nibret E, Sporer F, et al. Biological activity of the essential oil of *Kadsura longipedunculata* (Schisandraceae) and its major components. J Pharm Pharmacol. 2010;62:1037–44. https://doi.org/10.1111/j.2042-7158.2010.01119.x.
16. Chang H-T, Cheng Y-H, Wu C-L, Chang S-T, Chang T-T, Su Y-C. Antifungal activity of essential oil and its constituents from *Calocedrus macrolepis* var. formosana florin leaf against plant pathogenic fungi. Bioresour Technol. 2008;99:6266–70. https://doi.org/10.1016/j.biortech.2007.12.005.
17. Yap H-YY, Muria-Gonzalez MJ, Kong B-H, Stubbs KA, Tan C-S, Ng S-T, et al. Heterologous expression of cytotoxic sesquiterpenoids from the medicinal mushroom *Lignosus rhinocerotis* in yeast. Microb Cell Fact. 2017;16:103. https://doi.org/10.1186/s12934-017-0713-x.
18. Zhou H, Yang Y-L, Zeng J, Zhang L, Ding Z-H, Zeng Y. Identification and characterization of a δ-cadinol synthase potentially involved in the formation of boreovibrins in *Boreostereum vibrans* of Basidiomycota. Nat Prod Bioprospect. 2016;6:167–71. https://doi.org/10.1007/s13659-016-0096-4.
19. Mischko W, Hirte M, Fuchs M, Mehlmer N, Brück TB. Identification of sesquiterpene synthases from the Basidiomycota *Coniophora puteana* for the efficient and highly selective β-copaene and cubebol production in *E. coli*. Microb Cell Fact. 2018;17:164. https://doi.org/10.1186/s12934-018-1010-z.
20. Kemper K, Hirte M, Reinbold M, Fuchs M, Brück T. Opportunities and challenges for the sustainable production of structurally complex diterpenoids in recombinant microbial systems. Beilstein J Org Chem. 2017;13:845–54. https://doi.org/10.3762/bjoc.13.85.
21. Hirte M, Mischko W, Kemper K, Röhrer S, Huber C, Fuchs M, et al. From microbial upcycling to biology-oriented synthesis: combining whole-cell production and chemo-enzymatic functionalization for sustainable taxanoid delivery. Green Chem. 2018;20:5374–84. https://doi.org/10.1039/c8gc03126f.
22. Sieber P, Voigt K, Kämmer P, Brunke S, Schuster S, Linde J. Comparative study on alternative splicing in human fungal pathogens suggests its involvement during host invasion. Front Microbiol. 2018;9:2313. https://doi.org/10.3389/fmicb.2018.02313.
23. Da Lage J-L, Binder M, Hua-Van A, Janeček S, Casane D. Gene make-up: rapid and massive intron gains after horizontal transfer of a bacterial α-amylase gene to basidiomycetes. BMC Evol Biol. 2013;13:40. https://doi.org/10.1186/1471-2148-13-40.
24. Linstrom P. NIST chemistry WebBook, NIST standard reference database 69. Gaithersburg: National Institute of Standards and Technology; 1997.
25. Steele CL, Crock J, Bohlmann J, Croteau R. Sesquiterpene synthases from grand fir (*Abies grandis*). Comparison of constitutive and wound-induced activities, and cDNA isolation, characterization, and bacterial expression of delta-selinene synthase and gamma-humulene synthase. J Biol Chem. 1998;273:2078–89. https://doi.org/10.1074/jbc.273.4.2078.
26. Yoshikuni Y, Martin VJJ, Ferrin TE, Keasling JD. Engineering cotton (+)-delta-cadinene synthase to an altered function: germacrene D-4-ol synthase. Chem Biol. 2006;13:91–8. https://doi.org/10.1016/j.chembiol.2005.10.016.
27. Raz K, Levi S, Gupta PK, Major DT. Enzymatic control of product distribution in terpene synthases: insights from multiscale simulations. Curr Opin Biotechnol. 2020;65:248–58. https://doi.org/10.1016/j.copbio.2020.06.002.
28. López-Gallego F, Wawrzyn GT, Schmidt-Dannert C. Selectivity of fungal sesquiterpene synthases: role of the active site's H-1 alpha loop in catalysis. Appl Environ Microbiol. 2010;76:7723–33. https://doi.org/10.1128/AEM.01811-10.
29. Lopez-Gallego F, Agger SA, Abate-Pella D, Distefano MD, Schmidt-Dannert C. Sesquiterpene synthases Cop4 and Cop6 from *Coprinus cinereus*: catalytic promiscuity and cyclization of farnesyl pyrophosphate geometric isomers. ChemBioChem. 2010;11:1093–106. https://doi.org/10.1002/cbic.200900671.
30. Ali NAA, Wursterb M, Denkert A, Arnold N, Fadail I, Al-Didamony G, et al. Chemical composition, antimicrobial, antioxidant and cytotoxic activity of essential oils of *Plectranthus cylindraceus* and *Meriandra benghalensis* from Yemen. Nat Prod Commun. 2012;7:1099–102.
31. Moiteiro C, Esteves T, Ramalho L, Rojas R, Alvarez S, Zacchino S, Bragança H. Essential oil characterization of two Azorean *Cryptomeria japonica* populations and their biological evaluations. Nat Prod Commun. 2013;8:1785–90.
32. Dickschat JS. Bacterial terpene cyclases. Nat Prod Rep. 2016;33:87–110. https://doi.org/10.1039/c5np00102a.
33. Holm L. DALI and the persistence of protein shape. Protein Sci. 2020;29:128–40. https://doi.org/10.1002/pro.3749.
34. Hiranuma N, Park H, Baek M, Anishchenko I, Dauparas J, Baker D. Improved protein structure refinement guided by deep learning based accuracy estimation. Nat Commun. 2021;12:1340. https://doi.org/10.1038/s41467-021-21511-x.
35. Quin MB, Flynn CM, Schmidt-Dannert C. Traversing the fungal terpenome. Nat Prod Rep. 2014;31:1449–73. https://doi.org/10.1039/c4np00075g.
36. Driller R, Garbe D, Mehlmer N, Fuchs M, Raz K, Major DT, et al. Current understanding and biotechnological application of the bacterial diterpene synthase CotB2. Beilstein J Org Chem. 2019;15:2355–68. https://doi.org/10.3762/bjoc.15.228.
37. Görner C, Häuslein I, Schrepfer P, Eisenreich W, Brück T. Targeted engineering of cyclooctat-9-en-7-ol synthase: a stereospecific access to two new non-natural fusicoccane-type diterpenes. ChemCatChem. 2013;5:3289–98. https://doi.org/10.1002/cctc.201300285.
38. Raz K, Driller R, Brück T, Loll B, Major DT. Understanding the role of active site residues in CotB2 catalysis using a cluster model. Beilstein J Org Chem. 2020;16:50–9. https://doi.org/10.3762/bjoc.16.7.
39. Schrepfer P, Buettner A, Goerner C, Hertel M, van Rijn J, Wallrapp 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. https://doi.org/10.1073/pnas.1519680113.
40. Karunanithi PS, Zerbe P. Terpene synthases as metabolic gatekeepers in the evolution of plant terpenoid chemical diversity. Front Plant Sci. 2019;10:1166. https://doi.org/10.3389/fpls.2019.01166.
41. Peters RJ, Carter OA, Zhang Y, Matthews BW, Croteau RB. Bifunctional abietadiene synthase: mutual structural dependence of the active sites for protonation-initiated and ionization-initiated cyclizations. Biochemistry. 2003;42:2700–7. https://doi.org/10.1021/bi020492n.
42. Kawaide H, Sassa T, Kamiya Y. Functional analysis of the two interacting cyclase domains in ent-kaurene synthase from the fungus *Phaeosphaeria* sp. L487 and a comparison with cyclases from higher plants. J Biol Chem. 2000;275:2276–80. https://doi.org/10.1074/jbc.275.4.2276.
43. Pemberton TA, Chen M, Harris GG, Chou WKW, Duan L, Köksal M, et al. Exploring the influence of domain architecture on the catalytic function

of diterpene synthases. Biochemistry. 2017;56:2010–23. https://doi.org/10.1021/acs.biochem.7b00137.
44. Jia Q, Li G, Köllner TG, Fu J, Chen X, Xiong W, et al. Microbial-type terpene synthase genes occur widely in nonseed land plants, but not in seed plants. Proc Natl Acad Sci USA. 2016;113:12328–33. https://doi.org/10.1073/pnas.1607973113.
45. Gross H, König GM. Terpenoids from marine organisms: unique structures and their pharmacological potential. Phytochem Rev. 2006;5:115–41. https://doi.org/10.1007/s11101-005-5464-3.
46. Biggs BW, Lim CG, Sagliani K, Shankar S, Stephanopoulos G, de Mey M, Ajikumar PK. Overcoming heterologous protein interdependency to optimize P450-mediated taxol precursor synthesis in *Escherichia coli*. Proc Natl Acad Sci USA. 2016;113:3209–14. https://doi.org/10.1073/pnas.1515826113.
47. Ringel M, Reinbold M, Hirte M, Haack M, Huber C, Eisenreich W, et al. Towards a sustainable generation of pseudopterosin-type bioactives. Green Chem. 2020;22:6033–46. https://doi.org/10.1039/d0gc01697g.
48. Studier FW. Protein production by auto-induction in high density shaking cultures. Protein Expr Purif. 2005;41:207–34. https://doi.org/10.1016/j.pep.2005.01.016.
49. Kabsch W. XDS. Acta Crystallogr D Biol Crystallogr. 2010;66:125–32. https://doi.org/10.1107/S0907444909047337.
50. Schriever K, Saenz-Mendez P, Rudraraju RS, Hendrikse NM, Hudson EP, Biundo A, et al. Engineering of ancestors as a tool to elucidate structure, mechanism, and specificity of extant terpene cyclase. J Am Chem Soc. 2021;143:3794–807. https://doi.org/10.1021/jacs.0c10214.
51. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. J Appl Crystallogr. 2007;40:658–74. https://doi.org/10.1107/S0021889807021206.
52. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr. 2010;66:213–21. https://doi.org/10.1107/S0907444909052925.
53. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, et al. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr. 2012;68:352–67. https://doi.org/10.1107/S0907444912001308.
54. Casañal A, Lohkamp B, Emsley P. Current developments in coot for macromolecular model building of electron cryo-microscopy and crystallographic data. Protein Sci. 2020;29:1069–78. https://doi.org/10.1002/pro.3791.
55. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci. 2018;27:293–315. https://doi.org/10.1002/pro.3330.
56. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers. 1983;22:2577–637. https://doi.org/10.1002/bip.360221211.
57. Barton GJ. ALSCRIPT: a tool to format multiple sequence alignments. Protein Eng. 1993;6:37–40. https://doi.org/10.1093/protein/6.1.37.
58. McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, et al. Analysis tool web services from the EMBL-EBI. Nucleic Acids Res. 2013;41:W597-600. https://doi.org/10.1093/nar/gkt376.
59. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem. 2004;25:1605–12. https://doi.org/10.1002/jcc.20084.
60. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M-Y, et al. Comparative protein structure modeling using modeller. Curr Protoc Bioinformatics. 2006. https://doi.org/10.1002/0471250953.bi0506s15.
61. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics. 2014;30:3276–8. https://doi.org/10.1093/bioinformatics/btu531.

Publisher's Note

4. Discussion and Outlook

Advancing biocatalysts for drug development

Advances in computational drug design and natural product synthesis have pioneered and set the pace in modern drug development over the past years.[98–100] Although chemical synthesis has elaborated largely over the past decades thus enabling the synthesis of a large number of natural products, this technique still has to obey to certain limits that ultimately render it inefficient (e.g. toxicity issues, low yields and production of racemic mixtures).[101] In this context biocatalytic strategies are gaining more and more importance for pharmaceutical industries by providing new opportunities to which existing synthetic techniques only offer insufficient solutions and by opening a yet untouched chemical space (e.g. C-H activation).[101] A prominent example for biocatalytic synthesis strategies successfully used in an industrial scale is the chemo-enzymatic production of statins.[102,103] Moreover, considering the emerging climate change effects, biocatalytic approaches combined with biotechnological techniques offer a sustainable route towards the production of natural drug leads. Hence natural bioactives (e.g.: terpenoids) can be produced enzymatically while completely avoiding the reliability on fossil fuel based fine chemicals and the production of toxic waste streams. This renders enzymatic catalysis truly suitable for development of sustainable processes for the pharmaceutical industry as enzymes themselves are i.e. produced from inexpensive raw material and are biodegradable, thus comply to central principles of green chemistry.[104,105] Further, enzymatic catalysis offers the opportunity to fine tune the desired catalytic activity by means of product specificity, substrate scope and even regio-, stereo-, and enantioselectivity via targeted protein engineering.[106] In contrast to rather universally applicable chemical catalysts this approach gives rise to tailor-made, product-targeted and thus highly specific biocatalysts.[106,107]

To that end, the herein presented study by Ringel et al. entitled ‘Towards a sustainable generation of pseudopterosin-type bioactives’[62] clearly shows the advantages of enzymatic catalysis compared to conventional extraction methods or total chemical synthesis. In this setup we were able to present a sustainable route towards the key pseudopterosin intermediate, erogorgiaene, mediated by just two enzymatic steps. First, we performed *in silico* guided mutagenesis of the hydropyrene synthase from *Streptomyces clavuligerus* to efficiently produce isoelisabethatriene A as main product in a whole cell bio catalysis approach. Secondly, we were able to effectively transform isoelisabethatriene A to the key intermediate erogorgiaene using a chemo-enzymatic oxidation mediated by lipase CalB with a yield of 69%. Pseudopterosins, which represent the final product target, as well as the intermediate erogorgiaene exhibit promising pharmaceutical properties such as anti-inflammatory and anti-

tubercular, respectively.[32,108] The extraction from natural sources for both compounds is highly unsustainable as it involves harvesting of coral reefs thus destroying one of the most important marine ecosystems and global carbon dioxide sinks.[37,109,110] Further, total chemical synthesis of pseudopterosins and erogorgiaene are to date only established in laboratory scale involving numerous cost-intensive and toxic metal-organic catalysts.[39,111,112] Based on these laboratory scale data we thus were able to proof that the biocatalytic approach for erogorgiaene production is not only more sustainable but also more favourable under economic constraints compared to published chemical synthesis routes thereby showcasing the power of targeted protein engineering combined with biotechnological production systems.[62]

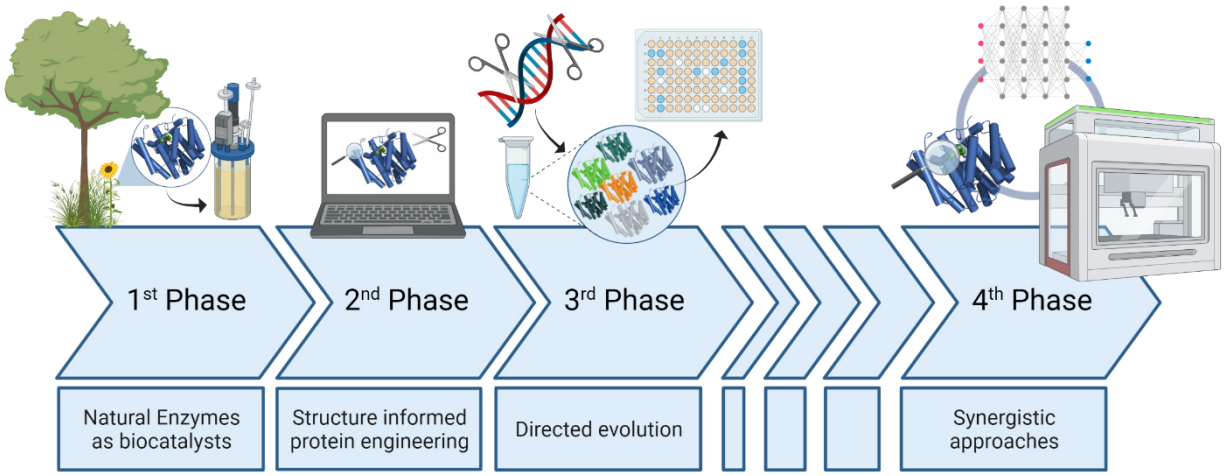


Figure 6: Evolution of protein engineering from a 1st phase of protein engineering employing natural enzymes as biocatalysis, a 2nd employing structure informed protein engineering approaches to alter the enzymes' functions, a 3rd phase using directed evolution approaches with enhanced screening techniques towards the future of protein engineering by entering a 4th phase which utilizes the synergistic effects of structure informed protein engineering, deep learning algorithms and modern high-end screening techniques.[106,107,113] Graphical representations were created with PyMOL 2.6.0 (The PyMOL Molecular Graphics System, Schrödinger, LLC.), Protein structure of diterpene synthase CotB2 (PDB ID 6GGI) was retrieved from rcsb.org[114], overall figure created with BioRender.

Over the past decades protein engineering approaches have already elaborated immensely starting from a so-called first phase of protein engineering in which naturally occurring enzymes were used to mediate desired reactions. In a second phase, structurally informed engineering approaches were used to alter the substrate scope of biocatalysts while directed evolution approaches accelerated the pace in protein engineering in a third phase.[107,115] In this regard directed evolution has advanced to the most important protein engineering strategy by gradually integrating favourable mutations through iterative rounds of random mutagenesis (techniques used are i.e. error-prone PCR or DNA shuffling).[101,106,116] Although protein engineering has thereby already evolved to an extremely powerful tool giving rise to enzymes with new-to-nature activities, to date the limiting factor for evolution of tailor-made enzymes

definitely is the screening time needed for identification of the latter.[106] The optimisation of proteins by directed evolution depends on creation and screening of vast variant libraries thus leaving it impossible to exhaustively screen the available sequence space of the enzyme of interest using currently available high-throughput screening and computational methods.[106,116] Moreover, the functional sequence space that can be addressed by mutagenesis while retaining enzymatic activity is considerably small compared to the multiplicity of mutations that leave the enzyme either non-functional, non-soluble or in the best case simply non-optimised (optimisation plateaus).[106,116,117] To this end, even single point mutations can either change the product scope or productivity completely as well as disrupt the whole functional, secondary structure of an enzyme.[116,118] This leads to the inevitable question how to efficiently traverse an enzyme's functional sequence space to reach the desired optimisation maximum and avoid a plethora of non-sense variants that would only waste screening time. To find an elegant solution to this issue, one must look deep into all different aspects of protein engineering to find a synergistic route taking advantage of all currently available state of the art techniques thereby entering a fourth phase of enzyme engineering.[115] Clearly, the different aspects of protein engineering are revolving around the two major elements of a conventional directed evolution workflow, first the variant design, and second the screening techniques.[101,106]

Rational directed evolution for design of 'small but smart' [106] variant libraries

In contrast to conventional directed evolution workflows generating a vast number of enzyme variants, approaches that rather rely on a more rational, structure-function informed protein engineering strategy provide the opportunity to create "small but smart" variant libraries that are more likely to yield optimised enzyme variants.[106,113,119] Hence, an in-depth knowledge of structure-function relations of the enzyme of interest including features like active-site location, co-factor binding motifs or secondary structure movements upon substrate binding is immensely important.[120] In order to obtain insights into the structure-function relations, high resolution crystal structures of the enzymes of interest, preferably in open (apo) and closed (substrate-bound) state, are mandatory.[97,121] In addition, multiple sequence alignments provide a detailed insight into conserved sequence motifs of the specific enzyme families and thus help in the analysis of structure-function relations.[122] Although rational protein engineering strategies based on the above mentioned knowledge are already very effective, it is rare to yield completely altered product portfolios and even harder to predict the specific effects of certain mutations.[120]

Therefore, combining rational, structure-informed, and random approaches consequently allows for effective targeting of specific enzyme domains to yield optimisation results. Several rational or semi-rational directed evolution approaches utilising the structure-function based

knowledge can be conducted based on the degree and nature of the available structural and functional information.[119] In this context, mutations targeting the active site of an enzyme will most probably result in an altered substrate scope or product portfolio while mutations of residues not directly involved in catalysis (these might be in the vicinity of the active site or far away) rather effect the stability or activity of the enzyme. [119,123,124] If structural information is generally available for identification of target residues, an option combining rational protein design and directed evolution would be to choose specific residues that should be subjected to mutagenesis but subsequently randomly alter these positions.[119] In this case also simultaneous randomisations at various positions may have a synergistic effect.[119] In case there is only limited access to crystal structure or high fidelity model data[125] of a desired enzyme, key residues can be identified by iterative rounds of random mutagenesis and further modified by site-saturation of these specific positions. Moreover, if specific residues are known but also benefits of unpredictable mutation sites should be exploited, simultaneous conduction of random and site-saturation (of defined residues) mutagenesis can be a valid way forward.[119]

State of the art insights into the engineering of terpene synthases

In this regard the engineering efforts of terpene synthases can build on solid data regarding the availability of crystal structures and terpenoid related protein sequences for multiple sequence alignments. Thus, workflows for genome mining for new terpene synthases, rational engineering and building of computational models based on existing structures can benefit from this large knowledge base. To this end in the herein presented study 'Biotechnological potential und initial characterization of two novel sesquiterpene synthases from Basidiomycota *Coniophora puteana* for heterologous production of δ-cadinol'[126] we could demonstrate the identification and characterisation of several terpene synthases which were carefully selected from the genome of wood rotting fungus *Coniophora puteana* using the in-depth, structure-function relation based knowledge of catalytically important conserved motifs (DDXXD, NSE, WxxxxxRY) of class I terpene synthases.[19,55] We could thus prove, that detailed knowledge about catalytically essential residues can indeed minimise the screening effort for identification of functionally active terpenoid biocatalysts (in this case two active synthases out of seven selected enzymes).[126]

To date mostly rational engineering approaches (e.g., site-directed mutagenesis) based on structural information or sequence comparison have been used for mutagenesis studies revolving around terpene synthases thereby elucidating their characteristic catalytic mechanism.[127] In fact, extensive mutagenesis studies based on structural as well as computational models were performed i.e. for the class I diterpene synthase CotB2 from *Streptomyces melanosporofaciens* resulting in significant changes of the product portfolio of

CotB2.[55,97,128,129] Specifically residues lining the active site were shown to have an influence on the product formation. In this regard, CotB2 was demonstrated to be very receptive towards the alteration of its product portfolio by mutagenesis of single residues. In addition, the residue W288, which is part of the conserved motif WxxxxxRY, was proven to have an influence on the product formation.[55,129] When comparing with literature references, a general importance in guiding of the product formation was suggested for this specific residue (CotB2 W288) as mutagenesis throughout different synthases including the hydropyrene synthase in our studies yielded either a change in product profile, a decrease in activity or complete loss of function.[62,130,131] To this end, mutagenesis of hydropyrene synthase conducted in the course of our studies revealed several methionines within the active site with particular catalytic relevance by stabilising carbocation intermediates.[62] So far, the catalytic involvement of methionines was only suggested via computational methods for trichodiene synthase, thus our studies represent the first experimental evidence for a catalytic relevance of the latter.[69]

Although the wide array of mutagenesis studies on different terpene synthases already gave insights into the catalytic mode of action, the specific task of each residue within the active site still remains unclear. To unravel the specific residues which are important for product formation within the active site of a terpene synthase we investigated the precise active site architecture of the two (+)-δ-cadinol synthases identified in the herein presented study entitled 'Biotechnological potential und initial characterization of two novel sesquiterpene synthases from Basidiomycota *Coniophora puteana* for heterologous production of δ-cadinol'.[126] Interestingly both synthases show highly conserved motifs and an equally decorated hydrophobic pocket for the first shell of residues.[126] Thus, we concluded that these identical residues are important for guiding the product formation. However, subsequent product-targeted mutagenesis studies of the latter residues aiming for a shift towards the product profile of the previously characterized cubebol synthase Copu3 [84] only showed minor changes in the product profile of Copu5 and Copu9, respectively.[126] Generally, these results suggest that both synthases are considerably unreactive towards single amino acid changes compared to other class I terpene synthases e.g., CotB2. [55,97,126,128,129] Nevertheless, Copu5 showed a higher receptiveness towards point mutations than Copu9. Interestingly, the addressed residues which were proven to have a higher impact on Copu5 are located in the close vicinity of the effector triade [64] involved in the initial carbocation formation located at the helix-turn-helix motif of the alpha helices G1 and G2. This specific motif is subjected to a conformational change upon substrate binding thereby closing the active site.[64] Since, the effect of mutations in this specific region showed contrasting results for Copu5 and Copu9, respectively, we concluded that the classical view of catalysis in terpene synthases involving a rather restricted focus on the active site architecture possibly needs to consider an essential importance of

tertiary structure elements specifically located near the G1/G2 helix-turn-helix motif.[126] To this end, our results are in conjunction with a large mutagenesis study of the (+)-δ-cadinene synthase from *Gossypium arboreum* employing a combination of rational design and random mutagenesis to alter the product profile of the latter enzyme. [132]

However, the prediction of effects of mutating a specific residue remains highly erratic especially as different enzymes might show diverging results when subjected to the same mutations. Therefore, the combination of rational design and random mutagenesis provides a valid way towards tailor-made design of biocatalysts.

Advances in high throughput protein engineering

For instance, little research efforts have been conducted in the field of directed evolution or random mutagenesis of terpene synthases aiming for altered product portfolios or optimised enzymatic properties so far. Yoshikuni et al. successfully showed the combination of rational design and random mutagenesis to alter the production portfolio of the (+)-δ-cadinene synthase from *Gossypium arboretum* towards the production of germacrene D4-ol.[132] However, the main obstacle to tackle is in fact the lack of a sophisticated and convenient high-throughput screening method for TPS product formation and/ or distribution. To this end, Lauchli and co-workers have worked on directed evolution of the sesquiterpene synthase BcBOT2 in order to obtain a more heat stable enzyme.[133] In the due course they were able to develop a high-throughput screening method for the BcBOT2 specific enzymatic reaction using a non-natural substrate.[133] Moreover, Agger et al. established a time-resolved, photometric assay for determination of TPS kinetic parameters using a coupled enzyme assay for detection of released PP moiety during TPS catalysis[71,134] as well as Vardakou et al. presented a PP depended method relying on malachite green as colouring agent for high-throughput characterisation of TPS.[135] To this end, in the herein presented study 'Biotechnological potential und initial characterization of two novel sesquiterpene synthases from Basidiomycota *Coniophora puteana* for heterologous production of δ-cadinol'[126] we were able to adapt the commercially available EnzCheck™ pyrophosphate assay kit (Invitrogen™, ThermoFisher Scientific) for fast, time-resolved determination of TPS kinetic parameters thus providing an additional method for PP depended TPS kinetics determination. However, although all mentioned methods provide valuable insights into biochemical characteristics of TPS none of them is able to give further information on substrate specificity or product distribution. In this context, Furubayashi and co-workers have invented a substrate-independent screening method for the cellular activity of terpene synthases which is relying on the substrate usage in carotenoid pathway engineered cells.[127] The described method allows for discrimination between the production of mono-/ sesqui- and diterpenes,

respectively, thus making a fast screening of altered substrate specificities introduced by e.g., directed evolution approaches available.[127]

However, to date none of the described methods enables direct screening of product-targeted mutagenesis approaches as the screening methods only provide a “yes” (active enzyme) or “no” (inactive enzyme) answer, detailed insights into the kinetic parameters of variant libraries and potentially a first glance on the product size when employing the screening method by Furubayashi et al.[127]. Hence, to gain insights into the specific product distribution of active enzymes a coupled spectrometric (e.g., GC-MS) readout is needed i.a., in the testing step of a design-make-test cycle employing rational directed evolution.[115] In this regard, there are several options to engage in a detailed compound analysis for high-throughput screenings of TPS. Certainly, one possibility would be to couple the spectrometric analysis right to the screening e.g., by using a microfluidics approach with a coupled mass spectrometry for analysis of metabolites in single droplets. For instance, Fidalgo et al. developed a technique for online MS analysis of individual microdroplets via on-chip extraction of the latter.[136] However, the feasibility of spectrometric analysis is directly depending on the resolution limits of the desired analytical device. To this end, the latter method only allows for analysis as low as 500µM [136] which to date is non-achievable in the context of single-cell terpene synthases screening systems. Thus, with respect to current analytical limitations the only valid way forward is a conventional, non-miniaturized GC-MS readout for product portfolio analysis.

Consequently, saving time during the screening efforts in terpene synthases engineering cannot build on analytical speed but rather implies a limitation of the number of clones to be screened and thus the creation of ‘small but smart’ libraries as mentioned above.[106] Computational biology has made significant progress in the last few years regarding computational enzyme engineering hence providing a very valuable addition to conventional directed evolution workflows by efficiently limiting the number of enzyme designs that have to be checked experimentally.[137] The toolbox of computational enzyme engineering involves several different strategies such as statistical, computational or biophysical protein design.[137] Statistical approaches are relying on existing protein structure datasets, often combined with ancestral sequence information to gain insights into structure-function relations of protein structures.[137,138] To this end, especially assessment of ancestral sequence information can help to improve e.g., the thermo stability or activity of the enzyme of interest.[137,139,140] Computational and biophysical design approaches include detailed computational assessment of protein/ligand or sidechain/backbone conformations based on high-resolution crystal structures by calculating geometries and energy terms of the latter.[137] The created knowledge can subsequently be used in the context of redesigning enzyme functions and validation of the new variants to a certain extent. In this regard, ‘EnzyDock’[141] is a recent

example for a multistate docking program which is capable of identifying and predicting possible binding modes of a substrate along a reaction coordinate (e.g., folding of a carbocation intermediate to the mature terpene). Needless to say, that especially a detailed knowledge about different substrate binding states and the involved residues is immensely important for predicting target residues in the context of tailor-made terpenoid biocatalysts. Hence, the extensive combination of computational biology and conventional protein engineering techniques, such as rational mutagenesis and directed evolution, provides a huge chance for enzyme design within the current fourth phase of protein engineering. Particularly in the light of current advances in the fields of quantum computing, artificial intelligence, machine learning and deep-learning neural networks which will provoke a quantum leap in prediction of enzyme designs, the road ahead looks indeed very promising for protein engineering.[137,142,143]

Concluding remarks

Protein engineering has evolved immensely over the past decades and will certainly be a hot topic in synthetic biology research in the future given to recent advances in genomics, metabolomics, proteomics, and especially computational biology. Thus, the vision of tailor-made biocatalysts capable of mediating every desired reaction in a sheer endless chemical space seems to be coming within reach. Nevertheless, the road ahead, although promising, is still long.

The herein presented work is providing further insights into the catalytical mode of action of terpene synthases thereby offering a guiding significance for further terpene synthase engineering efforts. In summary, we were able to prove the catalytic involvement of methionines as well as provide a first hint that tertiary structure elements might play a bigger role that expected. Furthermore, the presented synthases Copu5 and Copu9 bear the potential for biotechnological production of the potential new drug candidate δ-cadinol with excellent titers. In addition, we could present an alternative and sustainable route towards the pharmaceutically highly interesting pseudopterosins via isoelisabethatriene A and erogorgiaene.

In conclusion, the work presented herein still offers overwhelming potential for further experiments to unravel the catalytic magic of terpene synthases but also provides the first steps towards new sustainable production processes of new bioactives.

5. List of Publications

Towards a sustainable generation of pseudopterosin-type bioactives

Marion Ringel, Markus Reinbold, Max Hirte, Martina Haack, Claudia Huber, Wolfgang Eisenreich, Mahmoud A. Masri, Gerhard Schenk, Luke W. Guddat, Bernhard Loll, Russell Kerr, Daniel Garbe and Thomas Brück

Green Chem., 2020,22, 6033-6046
https://doi.org/10.1039/D0GC01697G

Biotechnological potential und initial characterization of two novel sesquiterpene synthases from Basidiomycota Coniophora puteana for heterologous production of δ-cadinol

Marion Ringel, Nicole Dimos, Stephanie Himpich, Martina Haack, Claudia Huber, Wolfgang Eisenreich, Gerhard Schenk, Bernhard Loll and Thomas Brück

Microbial Cell Factories (2022) 21:64
https://doi.org/10.1186/s12934-022-01791-8

The Impression of a Nonexisting Catalytic Effect: The Role of CotB2 in Guiding the Complex Biosynthesis of Cyclooctat-9-en-7-ol

Keren Raz, Ronja Driller, Nicole Dimos, **Marion Ringel**, Thomas Brück, Bernhard Loll and Dan Thomas Major

J. Am. Chem. Soc. 2020, 142, 51, 21562–21574
https://doi.org/10.1021/jacs.0c11348

Patent: Modified terpene synthases and their use for production of pseudopterosin intermediates and/or pseudopterosins

Marion Ringel, Thomas Brück, Markus Reinbold and Daniel Garbe

Patent Publication No. EP 3933038; publication date: 05.01.2022

6. Reprint Permission

Towards a sustainable generation of pseudopterosin-type bioactives

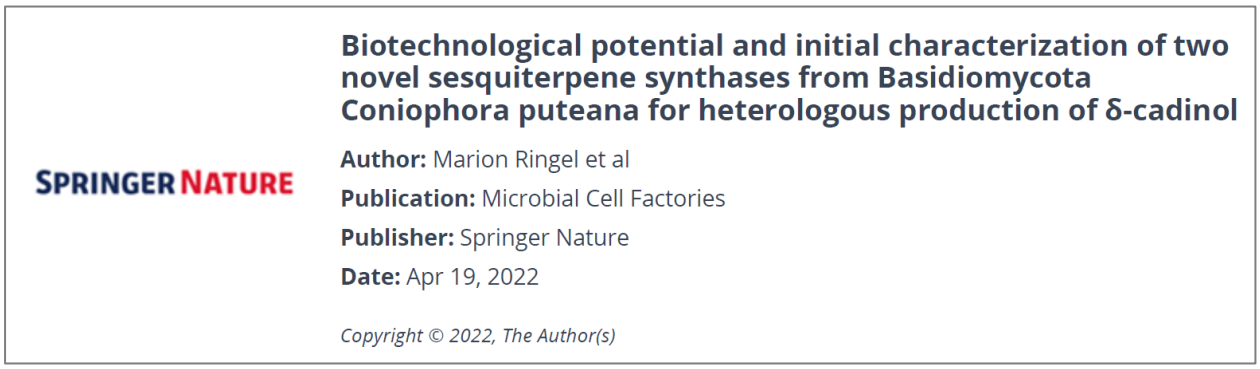
Published in Green Chemistry (Green Chem., 2020,22, 6033-6046)

Reproduced from Green Chem., 2020,22, 6033-6046 (https://doi.org/10.1039/D0GC01697G) with permission of Royal Society of Chemistry.

Biotechnological potential und initial characterization of two novel sesquiterpene synthases from Basidiomycota Coniophora puteana for heterologous production of δ-cadinol

Published in Microbial Cell Factories (Microbial Cell Factories (2022) 21:64)

The article is published and available under the Creative Commons Attribution License (CC BY).



7. Figures & Tables

8. References

1. Malhi, Y. *et al.* Climate change and ecosystems: threats, opportunities and solutions. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **375,** 20190104; 10.1098/rstb.2019.0104 (2020).

2. Camill, P. Global change. *Nature Education Knowledge* **3** (2010).

3. Pereira, H. M., Navarro, L. M. & Martins, I. S. Global Biodiversity Change: The Bad, the Good, and the Unknown. *Annu. Rev. Environ. Resour.* **37,** 25–50; 10.1146/annurev-environ-042911-093511 (2012).

4. Parmesan, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421,** 37–42; 10.1038/nature01286 (2003).

5. Keong, C. Y. The Ocean Carbon Sink and Climate Change: A Scientific and Ethical Assessment. *IJESD* **10,** 246–251; 10.18178/ijesd.2019.10.8.1181 (2019).

6. McMichael, A. J., Woodruff, R. E. & Hales, S. Climate change and human health: present and future risks. *The Lancet* **367,** 859–869; 10.1016/S0140-6736(06)68079-3 (2006).

7. Rodó, X., San-José, A., Kirchgatter, K. & López, L. Changing climate and the COVID-19 pandemic: more than just heads or tails. *Nature medicine* **27,** 576–579; 10.1038/s41591-021-01303-y (2021).

8. Beyer, R. M., Manica, A. & Mora, C. Shifts in global bat diversity suggest a possible role of climate change in the emergence of SARS-CoV-1 and SARS-CoV-2. *The Science of the total environment* **767,** 145413; 10.1016/j.scitotenv.2021.145413 (2021).

9. Carlson, C. J. *et al. Climate change will drive novel cross-species viral transmission* (2020).

10. Retel, C., Märkle, H., Becks, L. & Feulner, P. G. D. Ecological and Evolutionary Processes Shaping Viral Genetic Diversity. *Viruses* **11**; 10.3390/v11030220 (2019).

11. Patz, J. A. *et al.* Climate change and infectious diseases. *Climate change and human health: risks and responses* **6,** 103–137 (2003).

12. Shope, R. Global climate change and infectious diseases. *Environmental health perspectives* **96,** 171–174 (1991).

13. Adamson, C. S. *et al.* Antiviral drug discovery: preparing for the next pandemic. *Chemical Society reviews* **50,** 3647–3655; 10.1039/d0cs01118e (2021).

14. Ludwiczuk, A., Skalicka-Woźniak, K. & Georgiev, M. I. Terpenoids. In *Pharmacognosy* (Elsevier2017), pp. 233–266.

15. Gershenzon, J. & Dudareva, N. The function of terpene natural products in the natural world. *Nature chemical biology* **3,** 408–414; 10.1038/nchembio.2007.5 (2007).

16. Jaeger, R. & Cuny, E. Terpenoids with special pharmacological significance: a review. *Natural product communications* **11,** 1934578X1601100946 (2016).

17. Joshee, N., Dhekney, S. A. & Parajuli, P. (eds.). *Medicinal plants. From farm to pharmacy* (Springer International Publishing, Cham, 2019).

18. Cox-Georgian, D., Ramadoss, N., Dona, C. & Basu, C. Therapeutic and Medicinal Uses of Terpenes. In *Medicinal plants. From farm to pharmacy,* edited by N. Joshee, S. A. Dhekney & P. Parajuli (Springer International Publishing, Cham, 2019), pp. 333–359.

19. Christianson, D. W. Structural and Chemical Biology of Terpenoid Cyclases. *Chemical reviews* **117,** 11570–11648; 10.1021/acs.chemrev.7b00287 (2017).

20. Kemper, K., Hirte, M., Reinbold, M., Fuchs, M. & Brück, T. Opportunities and challenges for the sustainable production of structurally complex diterpenoids in recombinant microbial systems. *Beilstein journal of organic chemistry* **13,** 845–854; 10.3762/bjoc.13.85 (2017).

21. Schrader, J., Bohlmann, J. & Alonso-Gutierrez, J. (eds.). *Biotechnology of isoprenoids* (Springer, Cham, 2015).

22. Hillier, S. G. & Lathe, R. Terpenes, hormones and life: isoprene rule revisited. *The Journal of endocrinology* **242,** R9-R22; 10.1530/JOE-19-0084 (2019).

23. Breitmaier, E. *Terpenes. Flavors, fragrances, pharmaca, pheromones* (Wiley-VCH, Weinheim, 2006).

24. Newman, D. J. & Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **79,** 629–661; 10.1021/acs.jnatprod.5b01055 (2016).

25. Li, F., Wang, Y., Li, D., Chen, Y. & Dou, Q. P. Are we seeing a resurgence in the use of natural products for new drug discovery? *Expert opinion on drug discovery* **14,** 417–420; 10.1080/17460441.2019.1582639 (2019).

26. Meshnick, S. R. Artemisinin: mechanisms of action, resistance and toxicity. *International Journal for Parasitology* **32,** 1655–1660; 10.1016/S0020-7519(02)00194-7 (2002).

27. Westfall, P. J. *et al.* Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proceedings of the National Academy of Sciences of the United States of America* **109,** E111-8; 10.1073/pnas.1110740109 (2012).

28. Dietrich, J. A. *et al.* A novel semi-biosynthetic route for artemisinin production using engineered substrate-promiscuous P450(BM3). *ACS chemical biology* **4,** 261–267; 10.1021/cb900006h (2009).

29. 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; 10.7164/antibiotics.45.1587 (1992).

30. Look, S. A., Fenical, W., Jacobs, R. S. & Clardy, J. The pseudopterosins: anti-inflammatory and analgesic natural products from the sea whip Pseudopterogorgia elisabethae. *Proceedings of the National Academy of Sciences* **83,** 6238–6240 (1986).

31. Look, S. A., Fenical, W., Matsumoto, G. K. & Clardy, J. The pseudopterosins: a new class of antiinflammatory and analgesic diterpene pentosides from the marine sea whip Pseudopterogorgia elisabethae (Octocorallia). *J. Org. Chem.* **51,** 5140–5145; 10.1021/jo00376a016 (1986).

32. Mayer, A. M., Jacobson, P. B., Fenical, W., Jacobs, R. S. & Glaser, K. B. Pharmacological characterization of the pseudopterosins: Novel anti-inflammatory natural products isolated from the Caribbean soft coral, Pseudopterogorgia elisabethae. *Life Sciences* **62,** PL401-PL407; 10.1016/S0024-3205(98)00229-X (1998).

33. Berrué, F., McCulloch, M. W. B. & Kerr, R. G. Marine diterpene glycosides. *Bioorganic & medicinal chemistry* **19,** 6702–6719; 10.1016/j.bmc.2011.06.083 (2011).

34. Martins, A., Vieira, H., Gaspar, H. & Santos, S. Marketed marine natural products in the pharmaceutical and cosmeceutical industries: tips for success. *Marine Drugs* **12,** 1066–1101; 10.3390/md12021066 (2014).

35. Mayer, A. M. S. *et al.* The odyssey of marine pharmaceuticals: a current pipeline perspective. *Trends in pharmacological sciences* **31,** 255–265; 10.1016/j.tips.2010.02.005 (2010).

36. Wani, M. C. & Horwitz, S. B. Nature as a remarkable chemist: a personal story of the discovery and development of Taxol. *Anti-Cancer Drugs* **25,** 482–487; 10.1097/CAD.0000000000000063 (2014).

37. Lasker, H. R. Recruitment and resilience of a harvested Caribbean octocoral. *PloS one* **8,** e74587; 10.1371/journal.pone.0074587 (2013).

38. Liu, W. C., Gong, T. & Zhu, P. Advances in exploring alternative Taxol sources. *RSC Adv.* **6,** 48800–48809; 10.1039/C6RA06640B (2016).

39. Elford, T. G., Nave, S., Sonawane, R. P. & Aggarwal, V. K. Total synthesis of (+)-erogorgiaene using lithiation-borylation methodology, and stereoselective synthesis of each of its diastereoisomers. *Journal of the American Chemical Society* **133,** 16798–16801; 10.1021/ja207869f (2011).

40. Kanda, Y. *et al.* Two-Phase Synthesis of Taxol. *J. Am. Chem. Soc.* **142,** 10526–10533; 10.1021/jacs.0c03592 (2020).

41. Movahhed, S. *et al.* Total Synthesis of (+)-Erogorgiaene and the Pseudopterosin A-F Aglycone via Enantioselective Cobalt-Catalyzed Hydrovinylation. *Chemistry (Weinheim an der Bergstrasse, Germany)*; 10.1002/chem.202101863 (2021).

42. Ajikumar, P. K. *et al.* Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science (New York, N.Y.)* **330,** 70–74; 10.1126/science.1191652 (2010).

43. Hirte, M. *et al.* From microbial upcycling to biology-oriented synthesis. Combining whole-cell production and chemo-enzymatic functionalization for sustainable taxanoid delivery. *Green Chem* **20,** 5374–5384; 10.1039/c8gc03126f (2018).

44. Paramasivan, K. & Mutturi, S. Progress in terpene synthesis strategies through engineering of Saccharomyces cerevisiae. *Critical reviews in biotechnology* **37,** 974–989; 10.1080/07388551.2017.1299679 (2017).

45. Lange, B. M., Rujan, T., Martin, W. & Croteau, R. Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proceedings of the National Academy of Sciences* **97,** 13172–13177; 10.1073/pnas.240454797 (2000).

46. Kuzuyama, T. Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. *Bioscience, biotechnology, and biochemistry* **66,** 1619–1627; 10.1271/bbb.66.1619 (2002).

47. Zhang, C. Biosynthesis of Carotenoids and Apocarotenoids by Microorganisms and Their Industrial Potential. In *Progress in Carotenoid Research,* edited by L. Q. Zepka, E. Jacob-Lopes & V. V. de Rosso (InTech2018).

48. Ye, V. M. & Bhatia, S. K. Pathway engineering strategies for production of beneficial carotenoids in microbial hosts. *Biotechnology Letters* **34,** 1405–1414; 10.1007/s10529-012-0921-8 (2012).

49. Kim, S.-W. & Keasling, J. D. Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway inEscherichia coli enhances lycopene production.

Biotechnology and bioengineering **72,** 408–415; 10.1002/1097-0290(20000220)72:4<408::AID-BIT1003>3.0.CO;2-H (2001).

50. Khosla, C. & Keasling, J. D. Metabolic engineering for drug discovery and development. *Nature reviews. Drug discovery* **2,** 1019–1025; 10.1038/nrd1256 (2003).

51. Lu, Y., Yang, Q., Lin, Z. & Yang, X. A modular pathway engineering strategy for the high-level production of β-ionone in Yarrowia lipolytica. *Microb Cell Fact* **19,** 49; 10.1186/s12934-020-01309-0 (2020).

52. Zhang, F., Rodriguez, S. & Keasling, J. D. Metabolic engineering of microbial pathways for advanced biofuels production. *Current opinion in biotechnology* **22,** 775–783; 10.1016/j.copbio.2011.04.024 (2011).

53. Mischko, W. *et al.* Modular biomanufacturing for a sustainable production of terpenoid-based insect deterrents. *Green Chem* **68,** 1844; 10.1039/c8gc00434j (2018).

54. Wang, K. Isoprenyl diphosphate synthases. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1529,** 33–48; 10.1016/S1388-1981(00)00136-0 (2000).

55. Driller, R. *et al.* Towards a comprehensive understanding of the structural dynamics of a bacterial diterpene synthase during catalysis. *Nature communications* **9,** 3971; 10.1038/s41467-018-06325-8 (2018).

56. Koksal, M. & Christianson, D. W. *Crystal Structure of Taxadiene Synthase from Pacific Yew (Taxus brevifolia) in complex with Mg2+ and 2-fluorogeranylgeranyl diphosphate* (2010).

57. Christianson, D. W. Structural biology and chemistry of the terpenoid cyclases. *Chemical reviews* **106,** 3412–3442; 10.1021/cr050286w (2006).

58. Major, D. T. Electrostatic Control of Chemistry in Terpene Cyclases. *ACS Catal.* **7,** 5461–5465; 10.1021/acscatal.7b01328 (2017).

59. Driller, R. *et al.* Current understanding and biotechnological application of the bacterial diterpene synthase CotB2. *Beilstein journal of organic chemistry* **15,** 2355–2368; 10.3762/bjoc.15.228 (2019).

60. Dickschat, J. S. Bacterial terpene cyclases. *Natural product reports* **33,** 87–110; 10.1039/c5np00102a (2016).

61. Hu, Y., Chou, W. K. W., Hopson, R. & Cane, D. E. Genome mining in Streptomyces clavuligerus: expression and biochemical characterization of two new cryptic

sesquiterpene synthases. *Chemistry & Biology* **18,** 32–37; 10.1016/j.chembiol.2010.11.008 (2011).

62. Ringel, M. *et al.* Towards a sustainable generation of pseudopterosin-type bioactives. *Green Chem* **22,** 6033–6046; 10.1039/d0gc01697g (2020).

63. Rabe, P., Schmitz, T. & Dickschat, J. S. Mechanistic investigations on six bacterial terpene cyclases. *Beilstein journal of organic chemistry* **12,** 1839–1850; 10.3762/bjoc.12.173 (2016).

64. Baer, P. *et al.* Induced-fit mechanism in class I terpene cyclases. *Angewandte Chemie (International ed. in English)* **53,** 7652–7656; 10.1002/anie.201403648 (2014).

65. Weitman, M. & Major, D. T. Challenges posed to bornyl diphosphate synthase: diverging reaction mechanisms in monoterpenes. *J. Am. Chem. Soc.* **132,** 6349–6360; 10.1021/ja910134x (2010).

66. 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; 10.1039/c0ob00167h (2010).

67. Tantillo, D. J. Biosynthesis via carbocations: theoretical studies on terpene formation. *Natural product reports* **28,** 1035–1053; 10.1039/C1NP00006C (2011).

68. Raz, K. *et al.* The Impression of a Nonexisting Catalytic Effect: The Role of CotB2 in Guiding the Complex Biosynthesis of Cyclooctat-9-en-7-ol. *J. Am. Chem. Soc.* **142,** 21562–21574; 10.1021/jacs.0c11348 (2020).

69. Dixit, M., Weitman, M., Gao, J. & Major, D. T. Chemical Control in the Battle against Fidelity in Promiscuous Natural Product Biosynthesis: The Case of Trichodiene Synthase. *ACS catalysis* **7,** 812–818; 10.1021/acscatal.6b02584 (2017).

70. López-Gallego, F., Wawrzyn, G. T. & Schmidt-Dannert, C. Selectivity of fungal sesquiterpene synthases: role of the active site's H-1 alpha loop in catalysis. *Applied and environmental microbiology* **76,** 7723–7733; 10.1128/AEM.01811-10 (2010).

71. Lopez-Gallego, F., Agger, S. A., Abate-Pella, D., Distefano, M. D. & Schmidt-Dannert, C. Sesquiterpene synthases Cop4 and Cop6 from Coprinus cinereus: catalytic promiscuity and cyclization of farnesyl pyrophosphate geometric isomers. *Chembiochem : a European journal of chemical biology* **11,** 1093–1106; 10.1002/cbic.200900671 (2010).

72. Raz, K., Levi, S., Gupta, P. K. & Major, D. T. Enzymatic control of product distribution in terpene synthases: insights from multiscale simulations. *Current opinion in biotechnology* **65,** 248–258; 10.1016/j.copbio.2020.06.002 (2020).

73. Grundy, D. J. *et al.* Mechanism of Germacradien-4-ol Synthase-Controlled Water Capture. *Biochemistry* **55,** 2112–2121; 10.1021/acs.biochem.6b00115 (2016).

74. LB (Luria-Bertani) liquid medium. *Cold Spring Harb Protoc* **2006,** pdb.rec8141; 10.1101/pdb.rec8141 (2006).

75. SOC medium for HT. *Cold Spring Harb Protoc* **2008,** pdb.rec11457; 10.1101/pdb.rec11457 (2008).

76. Biggs, B. W. *et al.* Overcoming heterologous protein interdependency to optimize P450-mediated Taxol precursor synthesis in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **113,** 3209–3214; 10.1073/pnas.1515826113 (2016).

77. OpenWetWare. *E. coli genotypes -- OpenWetWare*. Available at https://openwetware.org/mediawiki/index.php?title=E._coli_genotypes&oldid=1098511 (2021).

78. Raab, D., Graf, M., Notka, F., Schödl, T. & Wagner, R. The GeneOptimizer Algorithm: using a sliding window approach to cope with the vast sequence space in multiparameter DNA sequence optimization. *Systems and synthetic biology* **4,** 215–225; 10.1007/s11693-010-9062-3 (2010).

79. Espah Borujeni, A., Channarasappa, A. S. & Salis, H. M. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic acids research* **42,** 2646–2659; 10.1093/nar/gkt1139 (2014).

80. Farasat, I. *et al.* Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Molecular systems biology* **10,** 731; 10.15252/msb.20134955 (2014).

81. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Bio/Technology* **27,** 946–950; 10.1038/nbt.1568 (2009).

82. Ng, C. Y., Farasat, I., Maranas, C. D. & Salis, H. M. Rational design of a synthetic Entner-Doudoroff pathway for improved and controllable NADPH regeneration. *Metabolic engineering* **29,** 86–96; 10.1016/j.ymben.2015.03.001 (2015).

83. Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of Escherichia coli with plasmids. *Gene* **96,** 23–28 (1990).

84. Mischko, W., Hirte, M., Fuchs, M., Mehlmer, N. & Brück, T. B. Identification of sesquiterpene synthases from the Basidiomycota Coniophora puteana for the efficient and highly selective β-copaene and cubebol production in E. coli. *Microbial cell factories* **17,** 164; 10.1186/s12934-018-1010-z (2018).

85. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of molecular biology* **215,** 403–410; 10.1016/S0022-2836(05)80360-2 (1990).

86. Yamada, Y. *et al.* Novel terpenes generated by heterologous expression of bacterial terpene synthase genes in an engineered Streptomyces host. *The Journal of antibiotics* **68,** 385–394; 10.1038/ja.2014.171 (2015).

87. Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Science* **27,** 135–145; 10.1002/pro.3290 (2018).

88. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology* **7,** 539; 10.1038/msb.2011.75 (2011).

89. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic acids research*; 10.1093/nar/gkab301 (2021).

90. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics (Oxford, England)* **23,** 127–128; 10.1093/bioinformatics/btl529 (2007).

91. Hiranuma, N. *et al.* Improved protein structure refinement guided by deep learning based accuracy estimation. *Nature communications* **12,** 1340; 10.1038/s41467-021-21511-x (2021).

92. Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols* **5,** 725–738; 10.1038/nprot.2010.5 (2010).

93. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nature methods* **12,** 7–8; 10.1038/nmeth.3213 (2015).

94. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25,** 1605–1612; 10.1002/jcc.20084 (2004).

95. Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics* **Chapter 5,** Unit-5.6; 10.1002/0471250953.bi0506s15 (2006).

96. Yadav, J. S., Thirupathaiah, B. & Al Khazim Al Ghamdi, A. Protecting Group Free Formal Total Synthesis of the Antitubercular Agent Erogorgiaene. *Eur. J. Org. Chem.* **2012,** 2072–2076; 10.1002/ejoc.201101269 (2012).

97. Janke, R., Görner, C., Hirte, M., Brück, T. & Loll, B. The first structure of a bacterial diterpene cyclase: CotB2. *Acta crystallographica. Section D, Biological crystallography* **70,** 1528–1537; 10.1107/S1399004714005513 (2014).

98. Schaduangrat, N. *et al.* Towards reproducible computational drug discovery. *J Cheminform* **12,** 9; 10.1186/s13321-020-0408-x (2020).

99. Truax, N. J. & Romo, D. Bridging the gap between natural product synthesis and drug discovery. *Natural product reports* **37,** 1436–1453; 10.1039/D0NP00048E (2020).

100. Tomar, V., Mazumder, M., Chandra, R., Yang, J. & Sakharkar, M. K. Small Molecule Drug Design. In *Encyclopedia of Bioinformatics and Computational Biology* (Elsevier2019), pp. 741–760.

101. Devine, P. N. *et al.* Extending the application of biocatalysis to meet the challenges of drug development. *Nat Rev Chem* **2,** 409–421; 10.1038/s41570-018-0055-1 (2018).

102. Müller, M. Chemoenzymatic synthesis of building blocks for statin side chains. *Angewandte Chemie (International ed. in English)* **44,** 362–365; 10.1002/anie.200460852 (2005).

103. Wolberg, M. *et al.* Chemoenzymatic synthesis of the chiral side-chain of statins: application of an alcohol dehydrogenase catalysed ketone reduction on a large scale. *Bioprocess Biosyst Eng* **31,** 183–191; 10.1007/s00449-008-0205-9 (2008).

104. Tang, S. L. Y., Smith, R. L. & Poliakoff, M. Principles of green chemistry: PRODUCTIVELY. *Green Chem* **7,** 761; 10.1039/b513020b (2005).

105. Macquarrie, D. J. & Clark, J. H. *Handbook of green chemistry and technology* (Blackwell Science, Oxford England, Malden, MA, 2002).

106. Badenhorst, C. P. S. & Bornscheuer, U. T. Getting Momentum: From Biocatalysis to Advanced Synthetic Biology. *Trends in Biochemical Sciences* **43,** 180–198; 10.1016/j.tibs.2018.01.003 (2018).

107. Bornscheuer, U. T. *et al.* Engineering the third wave of biocatalysis. *Nature* **485,** 185–194; 10.1038/nature11117 (2012).

108. Incerti-Pradillos, C. A. *et al.* Asymmetric Total Synthesis of (-)-Erogorgiaene and Its C-11 Epimer and Investigation of Their Antimycobacterial Activity. *Chemistry (Weinheim an der Bergstrasse, Germany)* **22,** 14390–14396; 10.1002/chem.201602440 (2016).

109. Kinsey, D. W. & Hopley, D. The significance of coral reefs as global carbon sinks—response to Greenhouse. *Palaeogeography, Palaeoclimatology, Palaeoecology* **89,** 363–377; 10.1016/0031-0182(91)90172-N (1991).

110. Roberts, C. M. *et al.* Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science* **295,** 1280–1284; 10.1126/science.1067728 (2002).

111. Newton, C. G. & Sherburn, M. S. Total synthesis of the pseudopterosin aglycones. *Natural product reports* **32,** 865–876; 10.1039/c5np00008d (2015).

112. Newton, C. G. *et al.* Pseudopterosin synthesis from a chiral cross-conjugated hydrocarbon through a series of cycloadditions. *Nature chemistry* **7,** 82–86; 10.1038/NCHEM.2112 (2015).

113. Bornscheuer, U. T. & Höhne, M. (eds.). *Protein engineering. Methods and protocols* (Humana Press, New York, NY, 2018).

114. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic acids research* **28,** 235–242; 10.1093/nar/28.1.235 (2000).

115. Truppo, M. D. Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Med. Chem. Lett.* **8,** 476–480; 10.1021/acsmedchemlett.7b00114 (2017).

116. Currin, A., Swainston, N., Day, P. J. & Kell, D. B. Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. *Chemical Society reviews* **44,** 1172–1239; 10.1039/C4CS00351A (2015).

117. Keefe, A. D. & Szostak, J. W. Functional proteins from a random-sequence library. *Nature* **410,** 715–718; 10.1038/35070613 (2001).

118. Kuhlman, B. & Baker, D. Native protein sequences are close to optimal for their structures. *Proceedings of the National Academy of Sciences* **97,** 10383–10388; 10.1073/pnas.97.19.10383 (2000).

119. Chica, R. A., Doucet, N. & Pelletier, J. N. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Current opinion in biotechnology* **16,** 378–384; 10.1016/j.copbio.2005.06.004 (2005).

120. Bornscheuer, U. T. & Pohl, M. Improved biocatalysts by directed evolution and rational protein design. *Current opinion in chemical biology* **5,** 137–143; 10.1016/S1367-5931(00)00182-4 (2001).

121. Spiliopoulou, M. *et al.* Applications of X-ray Powder Diffraction in Protein Crystallography and Drug Screening. *Crystals* **10,** 54; 10.3390/cryst10020054 (2020).

122. Thompson, J. D., Linard, B., Lecompte, O. & Poch, O. A comprehensive benchmark study of multiple sequence alignment methods: current challenges and future perspectives. *PloS one* **6,** e18093; 10.1371/journal.pone.0018093 (2011).

123. Morley, K. L. & Kazlauskas, R. J. Improving enzyme properties: when are closer mutations better? *Trends in Biotechnology* **23,** 231–237; 10.1016/j.tibtech.2005.03.005 (2005).

124. Park, S. *et al.* Focusing mutations into the P. fluorescens esterase binding site increases enantioselectivity more effectively than distant mutations. *Chemistry & Biology* **12,** 45–54; 10.1016/j.chembiol.2004.10.012 (2005).

125. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589; 10.1038/s41586-021-03819-2 (2021).

126. Ringel, M. *et al.* Biotechnological potential and initial characterization of two novel sesquiterpene synthases from Basidiomycota Coniophora puteana for heterologous production of δ-cadinol. *Microb Cell Fact* **21**; 10.1186/s12934-022-01791-8 (2022).

127. Furubayashi, M. *et al.* A high-throughput colorimetric screening assay for terpene synthase activity based on substrate consumption. *PloS one* **9,** e93317; 10.1371/journal.pone.0093317 (2014).

128. Tomita, T. *et al.* Structural Insights into the CotB2-Catalyzed Cyclization of Geranylgeranyl Diphosphate to the Diterpene Cyclooctat-9-en-7-ol. *ACS chemical biology* **12,** 1621–1628; 10.1021/acschembio.7b00154 (2017).

129. Görner, C., Häuslein, I., Schrepfer, P., Eisenreich, W. & Brück, T. Targeted Engineering of Cyclooctat-9-en-7-ol Synthase: A Stereospecific Access to Two New Non-natural Fusicoccane-Type Diterpenes. *ChemCatChem* **5,** 3289–3298; 10.1002/cctc.201300285 (2013).

130. 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; 10.1126/science.277.5333.1815 (1997).

131. Seemann, M. *et al.* Pentalenene synthase. Analysis of active site residues by site-directed mutagenesis. *J. Am. Chem. Soc.* **124,** 7681–7689; 10.1021/ja026058q (2002).

132. Yoshikuni, Y., Martin, V. J. J., Ferrin, T. E. & Keasling, J. D. Engineering cotton (+)-delta-cadinene synthase to an altered function: germacrene D-4-ol synthase. *Chemistry & Biology* **13,** 91–98; 10.1016/j.chembiol.2005.10.016 (2006).

133. Lauchli, R. *et al.* High-throughput screening for terpene-synthase-cyclization activity and directed evolution of a terpene synthase. *Angewandte Chemie International Edition* **52,** 5571–5574; 10.1002/anie.201301362 (2013).

134. Agger, S. A., Lopez-Gallego, F., Hoye, T. R. & Schmidt-Dannert, C. Identification of sesquiterpene synthases from Nostoc punctiforme PCC 73102 and Nostoc sp. strain PCC 7120. *Journal of bacteriology* **190,** 6084–6096; 10.1128/JB.00759-08 (2008).

135. Vardakou, M., Salmon, M., Faraldos, J. A. & O'Maille, P. E. Comparative analysis and validation of the malachite green assay for the high throughput biochemical characterization of terpene synthases. *MethodsX* **1,** 187–196; 10.1016/j.mex.2014.08.007 (2014).

136. Fidalgo, L. M. *et al.* Coupling Microdroplet Microreactors with Mass Spectrometry: Reading the Contents of Single Droplets Online. *Angew. Chem.* **121,** 3719–3722; 10.1002/ange.200806103 (2009).

137. Chowdhury, R. & Maranas, C. D. From directed evolution to computational enzyme engineering—A review. *AIChE J* **66**; 10.1002/aic.16847 (2020).

138. Dong, Q., Wang, X. & Lin, L. Novel knowledge-based mean force potential at the profile level. *BMC bioinformatics* **7,** 324; 10.1186/1471-2105-7-324. (2006).

139. Arenas, M. & Bastolla, U. ProtASR2: Ancestral reconstruction of protein sequences accounting for folding stability. *Methods Ecol Evol* **11,** 248–257; 10.1111/2041-210X.13341 (2020).

140. Spence, M. A., Kaczmarski, J. A., Saunders, J. W. & Jackson, C. J. Ancestral sequence reconstruction for protein engineers. *Current opinion in structural biology* **69,** 131–141; 10.1016/j.sbi.2021.04.001 (2021).

141. Das, S. *et al.* EnzyDock: Protein-Ligand Docking of Multiple Reactive States along a Reaction Coordinate in Enzymes. *J. Chem. Theory Comput.* **15,** 5116–5134; 10.1021/acs.jctc.9b00366 (2019).

142. Feehan, R., Montezano, D. & Slusky, J. S. G. Machine learning for enzyme engineering, selection and design. *Protein engineering, design & selection : PEDS* **34**; 10.1093/protein/gzab019 (2021).

143. Narayanan, H. *et al.* Machine Learning for Biologics: Opportunities for Protein Engineering, Developability, and Formulation. *Trends in pharmacological sciences* **42,** 151–165; 10.1016/j.tips.2020.12.004 (2021).