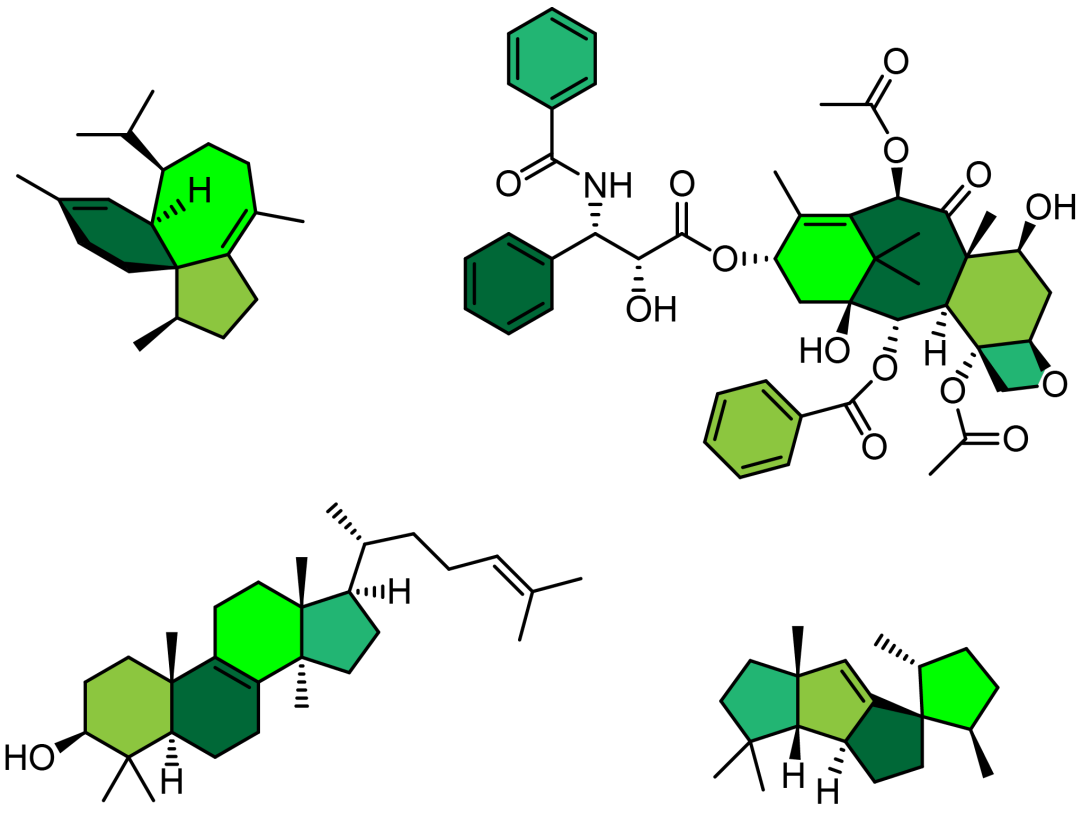
BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Terpenes

Edited by Jeroen S. Dickschat



Imprint

Beilstein Journal of Organic Chemistry
www.bjoc.org
ISSN 1860-5397
Email: journals-support@beilstein-institut.de

The *Beilstein Journal of Organic Chemistry* is published by the Beilstein-Institut zur Förderung der Chemischen Wissenschaften.

Beilstein-Institut zur Förderung der Chemischen Wissenschaften
Trakehner Straße 7–9
60487 Frankfurt am Main
Germany
www.beilstein-institut.de

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Terpenes

Jeroen S. Dickschat

Editorial

Open Access

Address:
Kekulé-Institut für Organische Chemie und Biochemie, Rheinische Friedrich-Wilhelms-Universität Bonn, Gerhard-Domagk-Straße 1, D-53121 Bonn, Germany

Email:
Jeroen S. Dickschat - dickschat@uni-bonn.de

Keywords:
Terpenes

Beilstein J. Org. Chem. **2019,** *15,* 2966–2967.
doi:10.3762/bjoc.15.292

Received: 03 December 2019
Accepted: 11 December 2019
Published: 13 December 2019

This article is part of the thematic issue "Terpenes".

Guest Editor: J. S. Dickschat

With more than 80,000 known compounds terpenes represent the largest class of natural products. Their remarkable structural complexity and diversity is a result of a unique biosynthetic pathway invented by nature that starts with the formation of only a few acyclic precursors termed oligoprenyl diphosphates. These precursors containing multiple olefinic double bonds can be ionised to generate a highly reactive cationic intermediate that can be subject to a cascade reaction through typical carbocation chemistry, including cyclisation reactions, hydride migrations and Wagner–Meerwein rearrangements [1,2]. The cascade is usually terminated by deprotonation or attack of water. The initially formed terpene hydrocarbons can subsequently be modified by enzymatic oxidations that often involve radical chemistry.

The non-functionalised hydrocarbons are volatile and often exhibit interesting odour properties [3]. As a consequence, these compounds may act as chemical signals such as pheromones, but also have a long standing history as fragrances used by humans. The oxidised derivatives may be of high interest because of their bioactivity, with some of the most important drugs used in medical applications being of terpenoid origin. Two of the most fanous examples include the diterpenoid taxol from *Taxus brevifolia* that is used to treat breast and other types of solid cancers, while artemisinin from *Artemisia annua* is used as a cure against malaria [4].

This thematic issue covers all different aspects of terpene chemistry and biochemistry, including compound isolation and structure elucidation, total synthesis, biosynthetic studies, bioactivity and functionalisation for mode of action studies, and other applications of terpenes, e.g. as fragrances. I thank all colleagues who have participated in this issue for their contributions and the whole Beilstein team for the professional support, and I wish the readers of the issue some joyful and stimulating new insights.

Jeroen S. Dickschat

Bonn, December 2019

ORCID® iDs

Jeroen S. Dickschat - https://orcid.org/0000-0002-0102-0631

References

1. Dickschat, J. S. *Angew. Chem., Int. Ed.* **2019,** *58,* 15964–15976. doi:10.1002/anie.201905312

2. Tantillo, D. J. *Angew. Chem., Int. Ed.* **2017,** *56,* 10040–10045. doi:10.1002/anie.201702363
3. Stepanyuk, A.; Kirschning, A. *Beilstein J. Org. Chem.* **2019,** *15,* 2590–2602. doi:10.3762/bjoc.15.252
4. Tu, Y. *Angew. Chem., Int. Ed.* **2016,** *55,* 10210–10226. doi:10.1002/anie.201601967

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

New sesquiterpenoids from the South China Sea soft corals *Clavularia viridis* and *Lemnalia flava*

Qihao Wu[‡1,2], Yuan Gao[‡1], Meng-Meng Zhang[1], Li Sheng[1], Jia Li[1], Xu-Wen Li[1], Hong Wang[*2] and Yue-Wei Guo[*1,2]

Full Research Paper

Open Access

Address:
[1]State Key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zu Chong Zhi Road, Zhangjiang Hi-Tech Park, Shanghai 201203, P. R. China and [2]College of Pharmaceutical Science and Collaborative Innovation Center of Yangtze River Delta Region Green Pharmaceuticals, Zhejiang University of Technology, Hangzhou 310014, P. R. China

Email:
Hong Wang[*] - hongw@zjut.edu.cn; Yue-Wei Guo[*] - ywguo@simm.ac.cn

* Corresponding author ‡ Equal contributors

Keywords:
Clavularia viridis; *Lemnalia flava*; NF-κB; PTP1B; sesquiterpenoid; soft coral, terpenes

Beilstein J. Org. Chem. **2019,** *15,* 695–702.
doi:10.3762/bjoc.15.64

Received: 03 January 2019
Accepted: 01 March 2019
Published: 15 March 2019

This article is part of the thematic issue "Terpenes".

Guest Editor: J. S. Dickschat

Abstract

A detailed chemical investigation of the South China Sea soft corals *Clavularia viridis* and *Lemnalia flava* yielded four new halogenated laurane-type sesquiterpenoids, namely, isobromolaurenisol (**1**), clalaurenol A (**2**), *ent*-laurenisol (**3**), clalaurenol B (**4**), and the new aromadendrane-type sesquiterpenoid claaromadendrene (**6**), together with three known sesquiterpenoids (**5**, **7**, and **8**). Their structures were determined by extensive spectroscopic analysis and by comparison with the previously reported analogues. In a bioassay, compounds **1**, **2** and **4** exhibited interesting inhibitory activities in vitro against PTP1B and NF-κB.

Introduction

Marine soft corals are important sources of biologically active compounds, which made them attractive targets for natural product chemists. Soft corals of the genus *Clavularia* (class Octocorallia, order Alcyonacea, family Clavulariidea), are prolific sources of numerous biologically active compounds [1-4]. A variety of structurally unique sesquiterpenes, including aromadendranes [5], maalianes [5], elemanes [6], and trinorguaianes [7-9], have been isolated since the early 1980s from several species of *Clavularia*. Soft corals of the genus *Lemnalia* are also a rich source of sesquiterpenoids and diterpenoids with various intriguing carbon skeletons, such as nardosinanes, neolemnanes, and ylanganes [10]. Many of these secondary metabolites have attracted a lot of attention for further synthetic and pharmacological studies due to their potent bioactivities ranging from neuroprotective, cytotoxic, to anti-inflammatory properties [10].

In the framework of our ongoing research for the bioactive metabolites from South China Sea soft corals [11,12], we made the collection of the title samples *Clavularia viridis* and *Lemnalia flava* off the Xisha Islands, Hainan Province, China. The chemical investigation of two title animals led to the isolation of four new halogenated laurane-type sesquiterpenoids **1**–**4**, one new aromadendrane-type sesquiterpenoid **6**) together with three related known compounds **5**, **7** and **8** (Figure 1). Herein, the isolation, structure elucidation and bioactivity evaluation of these compounds are presented.

Results and Discussion

The frozen bodies of the two soft corals *C. viridis* and *L. flava* were cut into pieces and exhaustively extracted with acetone. The Et_2O-soluble portion of the acetone extracts were chromatographed repeatedly over silica gel, Sephadex LH-20, and RP-HPLC to yield pure compounds. A total of eight compounds including compounds **1** (1.0 mg), **2** (0.9 mg), **3** (3.4 mg), **4** (1.4 mg), **5** (0.9 mg), **6** (2.8 mg), **7** (7.8 mg), and **8** (6.8 mg) were obtained from the *C. viridis* sample while two compounds **3** (8.6 mg) and **4** (2.3 mg) were obtained from *L. flava*. Among them, the known compounds were readily identified as cupalaurenol (**5**) [13], 1-hydroxyalloaromadendrene (**7**) [14], and humulene epoxide II (**8**) [15] by comparing their NMR spectroscopic data and optical rotation with those reported in the literature.

Isobromolaurenisol (**1**) was obtained as an optically active colorless oil. Its molecular formula, $C_{15}H_{18}OBr_2$, was deduced by HR-ESIMS with ion peaks at *m*/*z* 370.9657, $[M - H]^-$ (calcd for $C_{15}H_{17}OBr_2$, 370.9646), indicating six degrees of unsaturation. The ^{13}C NMR and DEPT spectra contained signals attributable to three methyls, two sp^3 methylenes, one sp^3 methine, one sp^3 quaternary carbon, three sp^2 methines, and five sp^2 quaternary carbons (Table 1). The typical resonances at δ_C 145.6, δ_C 113.0, $\delta_{H/C}$ 7.30/136.8, δ_C 123.4, δ_C 153.0, $\delta_{H/C}$ 6.71/116.8 revealed the presence of a 1,2,4,5-tetrasubstituted benzene ring, and the signals at $\delta_{H/C}$ 6.08/99.1, δ_C 154.2 indicated the existence of a trisubstituted double bond. All the above evidence suggested the laurane nature of this molecule, and literature research revealed that **1** should be an isomer of a known laurane-type terpenoid bromolaurenisol (**1a**) [16,17] due to their extremely similar NMR data and the same molecular weight (Figure 1). In fact, the main difference between **1** and **1a** happened only at the tetrasubstituted benzene ring with the substituents exchange between C-7 and C-10 (Figure 1). The assignment of the planar structure of **1** has been further confirmed by 2D NMR experiments, including $^1H,^1H$ COSY, HSQC, and HMBC, with the key correlations shown in Figure 2. In particular, the hydroxy group (δ_H 4.68, s) was confirmed to be attached at C-10 by the clear HMBC correlation from O*H* to C-10 and C-11.

The relative configuration of **1** was established by a NOESY experiment (Figure 3), in which the correlations of H_3-13 (δ_H 1.29, s) with H-2 (δ_H 3.56, q, J = 7.2 Hz) and H-5β (δ_H 2.34, m) indicated that these protons were on the same side of the molecule and were tentatively assigned to be β-oriented, while the correlation of H-5α (δ_H 1.88, m) and H_3-14 (δ_H 0.74, d, J = 7.3) at C-2 indicated CH_3-14 was α-oriented. Besides, the trisubstituted olefin ($\Delta^{3/15}$) was determined to be in *E* configuration due to the clear NOE correlations of H-15 with H_3-13 and H_3-14. In view of the above evidences, the relative configuration of compound **1** was determined as 1*R**,2*R**, the same as **1a** [16,17].

Compound **2** was isolated as an optically active colorless oil. The molecular formula, $C_{15}H_{19}OBr$, was established by the mo-

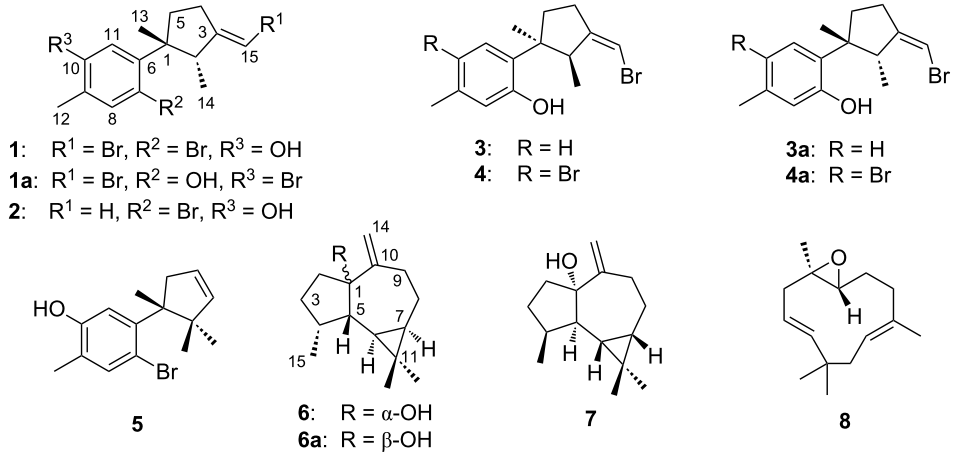


Figure 1: Structures of compounds **1**–**8**.

Table 1: ^{1}H and ^{13}C NMR data of **1–3**[a] recorded in $CDCl_3$.

No.	**1**		**2**		**3**	
	δ_H mult (*J* in Hz)	δ_C	δ_H mult (*J* in Hz)	δ_C	δ_H mult (*J* in Hz)	δ_C
1	–	52.0, qC	–	51.1, qC	–	48.4, qC
2	3.56, q (7.2)	47.5, CH	3.41, q (7.2)	47.1, CH	3.37, q (7.1)	48.1, CH
3	–	154.2, qC	–	157.6, qC	–	153.6, qC
4	2.40, m; 2.51, m	29.0, CH_2	2.48, m	28.2, CH_2	2.49, m	28.0, CH_2
5	1.88, m; 2.34, m	35.1, CH_2	1.79, m; 2.27, m	36.1, CH_2	1.93, m; 2.34, m	34.9, CH_2
6	–	145.6, qC	–	145.6, qC	–	130.4, qC
7	–	113.0, qC	–	113.3, qC	–	153.0, qC
8	7.30, s	136.8, CH	7.31, s	136.7, CH	6.52, s	116.6, CH
9	–	123.4, qC	–	123.1, qC	–	137.2, qC
10	–	153.0, qC	–	152.9, qC	6.71, d (7.7)	121.5, CH
11	6.71, s	116.8, CH	6.73, s	116.8, CH	7.01, d (7.8)	128.2, CH
12	2.18, s	15.0, CH_3	2.18, s	15.0, CH_3	2.28, s	20.7, CH_3
13	1.29, s	25.0, CH_3	1.30, s	25.1, CH_3	1.23, s	25.7, CH_3
14	0.74, d (7.3)	19.2, CH_3	0.73, d (7.2)	19.9, CH_3	0.76, d (7.2)	15.0, CH_3
15	6.08, s	99.1, CH	4.89, s; 4.99, s	106.9, CH_2	5.93, s	97.5, CH
OH	4.68, s	–	4.67, s	–	4.66, s	–

[a]Bruker DRX-500 spectrometer (125 MHz for ^{13}C NMR and 500 MHz for ^{1}H NMR) in $CDCl_3$, chemical shifts (ppm) referred to $CHCl_3$ (δ_C 77.16; δ_H 7.26); assignments were deduced by analysis of 1D and 2D NMR spectra.

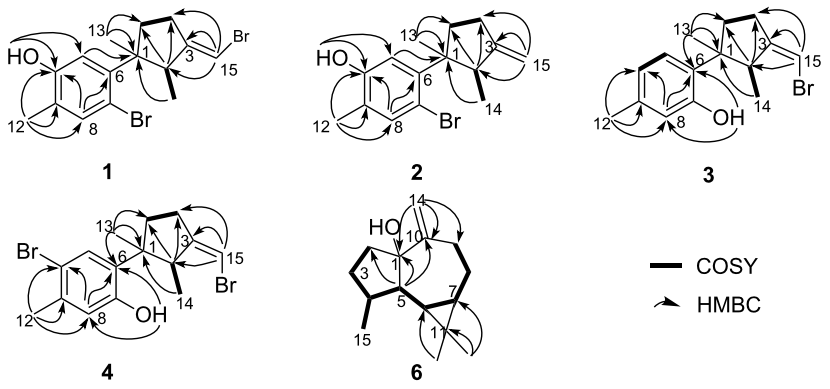


Figure 2: ^{1}H,^{1}H COSY and key HMBC correlations of compounds **1–4** and **6**.

lecular ion peak at *m/z* 293.0548, $[M - H]^-$ (calcd for $C_{15}H_{18}OBr$ 293.0541) in the HR-ESIMS spectrum. The ^{1}H and ^{13}C NMR spectra showed great similarities with those of the co-occurring **1**, which indicated the same laurane skeleton. In fact, compound **2** differed from **1** only by the debromonation at the C-15 position, which was in agree with the lack of 78/80 units in its mass compared to that of **1**. The planar structure of **2** was further confirmed by its 2D NMR data (Figure 2). The relative configurations of the chiral centers on the cyclopentane ring were determined to be the same as **1** by inspection of the proton coupling constants (Table 1) and NOESY experiments (Figure 3). Thus, compound **2** was determined to be the debrominated derivative of **1**, namely, clalaurenol A.

Compound **3** was observed as an optically active colorless oil. The molecular formula, $C_{15}H_{19}OBr$, was deduced by HR-EIMS ion peak at *m/z* 294.0617, $[M]^+$ (calcd for $C_{15}H_{19}OBr$, 294.0619). The ^{1}H and ^{13}C NMR data (Table 1) of **3** were found to be identical to those of laurenisol (**3a**), a halogenated sesquiterpenoid previously isolated from the red alga *Laurencia glandulifera* Kützing [18]. The relative configuration of **3** was established by NOESY correlations (Figure 3) in which the correlations of H_3-13 (δ_H 1.23, s) with H-2 (δ_H 3.37, q, J = 7.1 Hz) and H-5a (δ_H 1.93, m); H_3-14 (δ_H 0.76, d, J = 7.2 Hz) with H-5b (δ_H 2.34, m) indicating that two methyls at C-1 and C-2 were on the opposite side of the molecule. Besides, the NOESY correlation of one olefin proton (H-15,

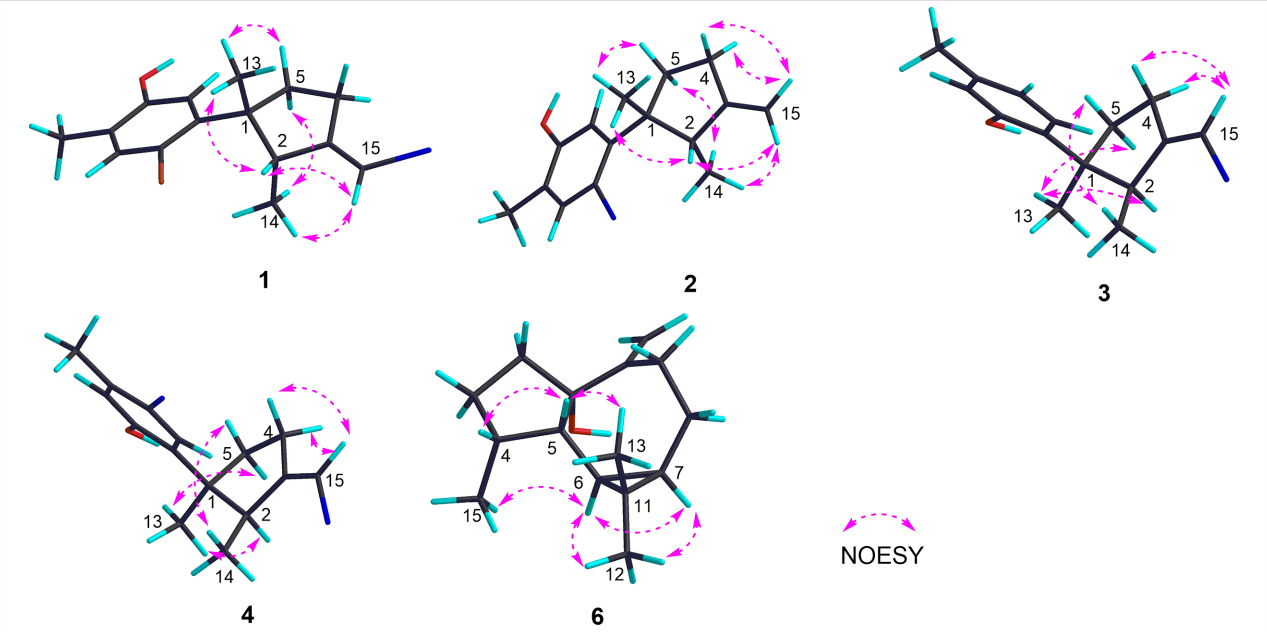


Figure 3: Key NOESY correlations for compounds **1**–**4** and **6**.

δ_H 5.93, s) and two protons at C-4 (δ_H 2.49, m), suggesting the *Z* geometry of the double bond at C-3/C-15. Finally, the sign of the $[\alpha]_D$ values of **3** {$[\alpha]_D^{20}$ −24.2 (*c* 0.06, $CHCl_3$); $[\alpha]_D^{20}$ −16.0 (*c* 0.10, MeOH)} were found to be opposite to that of laurenisol (+85.9) [18]. Thus, compound **3** can be assigned as the enantiomer of **3a**, named *ent*-laurenisol.

Clalaurenol B (**4**) was obtained as an optically active colorless oil. The molecular formula, $C_{15}H_{18}OBr_2$, the same as **4a** [18], was established by HR-ESIMS ion peaks at *m*/*z* 370.9654, $[M - H]^-$ (calcd for $C_{15}H_{17}OBr_2$, 370.9646). The ^1H and ^{13}C NMR data of **4** were identical to those of **4a**, a C-10 bromonated analogue of **3a**. In addition, the NOE correlations between H_3-13 (δ_H 1.21, s) with H-2 (δ_H 3.34, q, *J* = 7.3 Hz) and H-5a (δ_H 1.91, m); H_3-14 (δ_H 0.76, d, *J* = 7.2 Hz) with H-5b (δ_H 2.32, m); a proton at C-15 (H-15, δ_H 5.94, s) and two protons at C-4 (δ_H 2.48, m) suggested the relative configuration (C-1, C-2 and $\Delta^{3,15}$) of **4** is the same as **3**. Moreover, the sign of their $[\alpha]_D$ values {$[\alpha]_D^{20}$ −52.1 (*c* 0.08, $CHCl_3$); $[\alpha]_D^{20}$ −22.5 (*c* 0.10, MeOH)} for **4** and {$[\alpha]_D^{20}$ +74 (*c* 0.58, $CHCl_3$)} for **4a**, indicating that compound **4** should be the enantiomer of **4a** [18].

Compound **6** was isolated as an optically active colorless oil. The molecular formula $C_{15}H_{24}O$, the same as **7** [14] and **6a** [19,20], was established by HR-ESIMS ion peak at *m*/*z* 220.1824 $[M]^+$ (calcd for $C_{15}H_{24}O$, 220.1825). A detailed analysis of 2D NMR experiments (Figure 2), revealed that compound **6** had the same planar structure as **6a** and co-occurring **7** differing only in the stereochemistry. The relative configuration of **6** was established by NOESY correlations (Figure 3) in which the correlations of H-6 (δ_H 0.62, dd, *J* = 11.4, 9.1 Hz) with H-7 (δ_H 0.81, m) and H_3-15 (δ_H 1.01, d, *J* = 7.2 Hz); H_3-12 (δ_H 1.08, s) with H-6 and H-7, indicating that these protons were on the same side of the molecule and were tentatively assigned to be α-oriented, while correlations of H-5 (δ_H 1.60, m) with H-4 (δ_H 2.20, m) and H_3-13 (δ_H 0.99, s) suggesting these protons were on the opposite orientation. In view of the above evidences, the relative configuration of compound **6** was determined as 4*R**,5*S**,6*R**,7*R**. In fact, the only difference between compounds **6** and **6a** was the configuration of the hydroxy group at C-1 with α-orientation for **6** while β-orientation for **6a** [19,20]. Further, due to the influence of the configuration inversions of C-1, the ^{13}C NMR chemical shift of the carbon at C-1 (δ_C 85.5, qC), was apparently upfield shifted ($\Delta\delta$ = −3.0) comparing to compound **6a** (Table 2), giving the further support of the assigned structure for **6** (Figure 1). Thus, compound **6** was determined as a C-1 isomer of *ent*-1-hydroxy-alloaromadendrene (**6a**), namely, claaromadendrene.

In bioassays, all the isolated compounds were tested for protein tyrosine phosphase-1B (PTP1B) and NF-κB inhibitory activity. In the PTP1B inhibitory assay, the inhibitory effects of compounds **1**–**8** were evaluated against PTP1B, and the result showed that compounds **1**, **2** and **4** had a moderate PTP1B inhibitory activity with IC_{50} values of 18.8, 21.8 and 15.6 μM, respectively. The known PTP1B inhibitor oleanolic acid (IC_{50} = 3.0 μM) were used as positive control in this assay. In NF-κB inhibitory assay, compounds **2** and **4** showed the most potent NF-κB signaling pathway inhibition with IC_{50} values of

Table 2: ^{1}H and ^{13}C NMR data of **4**[a] and **6**[a] and ^{13}C NMR data of **6a**[b] and **7**[a] recorded in $CDCl_3$.

No.	**4**		**6**		**6a**[b]	**7**
	δ_H mult (*J* in Hz)	δ_C	δ_H mult (*J* in Hz)	δ_C	δ_H mult (*J* in Hz)	δ_C
1	–	48.4, qC	–	85.5, qC	88.5, qC	88.7, qC
2	3.34, q (7.3)	47.9, CH	1.76, m; 1.96, m	37.1, CH_2	36.5, CH_2	36.6, CH_2
3	–	153.1, qC	1.55, m; 1.90, m	33.3, CH_2	30.6, CH_2	30.7, CH_2
4	2,48, m	27.9, CH_2	2.20, m	34.9, CH	34.3, CH	34.4, CH
5	1.91, m; 2.32, m	34.9, CH_2	1.60, m	46.8, CH	49.0, CH	49.1, CH
6	–	133.2, qC	0.62, dd (11.4, 9.1)	21.8, CH	23.3, CH	23.4, CH
7	–	152.3, qC	0.81, m	27.5, CH	25.4, CH	25.5, CH
8	6.59, s	118.1, CH	0.98, m; 2.02, m	25.2, CH_2	21.2, CH_2	21.4, CH_2
9	–	136.5, qC	2.27, m; 2.47, dd (13.9, 12.6)	34.1, CH_2	32.0, CH_2	32.2, CH_2
10	–	115.5, qC	–	155.1, qC	152.9, qC	153.0, qC
11	7.23, s	131.9, CH	–	19.8, qC	not detected	18.1, qC
12	2.30, s	22.3, CH_3	1.08, s	29.0, CH_3	28.6, CH_3	28.7, CH_3
13	1.21, s	25.4, CH_3	0.99, s	15.7, CH_3	15.9, CH_3	16.0, CH_3
14	0.76, d (7.2)	15.0, CH_3	4.66, t (1.6); 4.80, d (1.7)	108.3, CH_2	111.7, CH_2	111.8, CH_2
15	5.94, s	97.8, CH	1.01 d, (7.2)	18.8, CH_3	16.6, CH_3	16.7, CH_3
OH	4.75, s	–	–	–		

[a]Bruker DRX-500 spectrometer (125 MHz for ^{13}C NMR and 500 MHz for ^{1}H NMR) in $CDCl_3$, chemical shifts (ppm) referred to $CHCl_3$ (δ_C 77.16; δ_H 7.26); assignments were deduced by analysis of 1D and 2D NMR spectra. [b]Data reported in ref. [19] (in $CDCl_3$).

6.8 and 7.3 μM, respectively, while compound **1** showed moderate activity with an IC_{50} value of 19.9 μM (Table 3).

Table 3: PTP1B and NF-κB inhibitory effect of compounds **1**–**8**.

Compounds	IC_{50} (μM)	
	PTP1B	NF-κB
1	18.8	19.9
2	21.8	6.8
3	–	–
4	15.6	7.3
5	–	–
6	–	–
7	–	–
8	–	–
A[#]	3.0	–
B[#]	–	14.0

A[#] and **B**[#], representing oleanolic acid and bortezomib, respectively, were used as the positive controls.

Conclusion

In summary, eight sesquiterpenoids (**1**–**8**), belonging to four different structural types, were isolated from two South China Sea soft corals (*C. viridis* and *L. flava*) for the first time. The discovery of these metabolites extended the structural diversity and complexity of sesquiterpenoids derived from soft corals *C. viridis* and *L. flava*. In fact, to our knowledge, naturally occurring laurane- (**1**–**4**) and cuparane-derived (**5**) sesquiterpenoids, are extremely rare in soft corals. Previously, such sesquiterpenoids have only been isolated from the red algae of the genus *Laurencia* [14,16,17,21] and some sea hares that prey on it [13,22]. In this paper, the chemical investigation of two different soft corals collected off the South China Sea, which belong to two different genera, have resulted in the discovery of two common new halogenated laurane-type sesquiterpenoids (**3** and **4**). Based on these findings, other than prey–predator relationship, the common symbiotic organisms in the algae and the soft corals might be the source of these metabolites. In fact, many investigations have proved that [23] numerous natural products are actually produced by microbes and/or microbial interactions with the "host from whence it was isolated". Further chemical investigation of these soft corals in the South China Sea as well as their associated microorganisms should be conducted to verify the true origin of these metabolites and to further understand the real biological/ecological roles they played in the life cycle of the title animals in the South China Sea.

The promising PTP1B inhibitory activity of laurane-type sesquiterpenoids [24] in a previously report from our group, inspired us to test the PTP1B inhibitory activity of compounds **1**–**4**. Among them, compound **3** was inactive against PTP1B enzyme, whereas compounds **1**, **2** and **4** exhibited considerable PTP1B inhibitory activity with IC_{50} values of 18.8, 21.8, and

15.6 μM, respectively. Compounds **1**, **2** and **4** also showed strong NF-κB inhibitory activity with IC_{50} values of 19.9, 6.8 and 7.3 μM, respectively. With regard to their structure–activity relationship, the bromine atom on the benzene ring may play the key functional role in the inhibitory activity. This study could thus provide a clue for the further biological study and structure modification of marine brominated laurane sesquiterpenoid derivatives towards new effective PTP1B and/or NF-κB inhibitors.

Experimental

General experimental procedures

Optical rotations were measured on a Perkin-Elmer 241MC polarimeter. IR spectra were recorded on a Nicolet-Magna FT-IR 750 spectrometer. EIMS and HR-EIMS spectra were recorded on a Finnigan-MAT-95 mass spectrometer. HR-ESIMS spectra were recorded on a Q-TOF Micro LC–MS–MS mass spectrometer. The NMR spectra were measured on a Bruker DRX-500 spectrometer with the residual $CHCl_3$ (δ_H 7.26 ppm, δ_C 77.2 ppm) as internal standard. Chemical shifts are expressed in δ (ppm) and coupling constants (*J*) in Hz. ^1H and ^{13}C NMR assignments were supported by ^1H,^1H COSY, HSQC, HMBC and NOESY experiments. Commercial silica gel (Qing Dao Hai Yang Chemical Group Co., 300–400 and 500–600 mesh) and Sephadex LH-20 (Amersham Biosciences) were used for column chromatography. Precoated silica gel GF254 plates (Sinopharm Chemical Reagent Co., Shanghai, China) were used for TLC. Reversed-phase (RP) HPLC purification was carried out on an Agilent 1260 series liquid chromatography system equipped with a DAD G1315D detector at 210 and 254 nm and with a semi-preparative ODS-HG-5 column [5 μm, 250 × 9.4 mm]. All solvents used for CC were of analytical grade, and solvents used for HPLC were of HPLC grade.

Collection of biological materials

The soft corals *C. viridis* and *L. flava* were collected by scuba from Xisha Island, Hainan Province, China, in March 23, 2013, at a depth of −15 to −20 m, and identified by Professor Xiu-Bao Li from Hainan University. The voucher samples, both *C. viridis* and *L. flava* are deposited at the Shanghai Institute of Materia Medica, CAS, under registration Nos. 13XS-49 and 13XS-52, respectively.

Extraction and isolation

The lyophilized bodies of *C. viridis* (80 g, dry weight) were minced into pieces and exhaustively extracted with acetone at room temperature (4 × 1 L). The solvent-free actone extract was partitioned between Et_2O and H_2O. The organic phase was evaporated under reduced pressure to give a dark-red residue (1.1 g), which was subjected to a gradient silica gel column chromatography (CC) [Et_2O/petroleum ether (PE), 0–100%] to yield 6 fractions (A–F). Fraction C was subjected to Sephadex LH-20 CC (PE/CH_2Cl_2/MeOH, 2:1:1) to give 4 sub-fractions (C1–C4). Fraction C4 was purified by silica gel CC (500–600 mesh, Et_2O/PE, 4:96) to afford pure **3** (3.4 mg), **4** (1.4 mg) and **8** (6.8 mg). Fraction D eluted with Sephadex LH-20 CC (PE/CH_2Cl_2/MeOH, 2:1:1), followed by CC on silica gel (500–600 mesh, Et_2O/PE, 5:95) to afford pure **1** (1.0 mg), **2** (0.9 mg) and **5** (0.9 mg). Fraction E gave compounds **6** (2.8 mg) and **7** (7.8 mg) after CC on Sephadex LH-20 (PE/CH_2Cl_2/MeOH, 2:1:1) and silica gel (500–600 mesh, Et_2O/PE, 8:92).

The frozen animals *L. flava* (350 g, dry weight) were cut into pieces and extracted exhaustively with acetone at room temperature (6 × 2.0 L). The organic extract was evaporated to give a brown residue, which was then partitioned between H_2O and Et_2O. The upper layer was concentrated under reduced pressure to give a brown residue 8.0 g. The resulted residue was separated into seven fractions (A–G) by gradient silica-gel CC. The resulting fractions were then fractionated into sub-fractions by Sephadex LH-20. The sub-fraction C5 was purified by Semi-preparative HPLC (87% MeOH), yielding compounds **3** (8.6 mg) and **4** (2.3 mg).

Isobromolaurenisol (**1**): Colorless oil; $[\alpha]_D^{20}$ +33.1 (*c* 0.07, $CHCl_3$); ^1H and ^{13}C NMR data, see Table 1; HR-ESIMS *m*/*z*: $[M - H]^-$ 370.9657, 372.9635, 374.9619 (calcd for $C_{15}H_{17}OBr_2$, 370.9646).

Clalaurenol A (**2**): Colorless oil; $[\alpha]_D^{20}$ +69.7 (*c* 0.05, $CHCl_3$); ^1H and ^{13}C NMR data, see Table 1; HR-ESIMS *m*/*z*: $[M - H]^-$ 293.0548, 295.0536 (calcd for $C_{15}H_{18}OBr$, 293.0541).

ent-Laurenisol (**3**): Colorless oil; $[\alpha]_D^{20}$ −24.2 (*c* 0.06, $CHCl_3$); $[\alpha]_D^{20}$ −16.0 (*c* 0.10, MeOH); ^1H and ^{13}C NMR data, see Table 1; HR-EIMS *m*/*z*: $[M]^+$ 294.0617, 296.0594 (calcd for $C_{15}H_{19}OBr$, 294.0619).

Clalaurenol B (**4**): Colorless oil; $[\alpha]_D^{20}$ −52.1 (*c* 0.08, $CHCl_3$); $[\alpha]_D^{20}$ −22.5 (*c* 0.10, MeOH); ^1H and ^{13}C NMR data, see Table 1; HR-ESIMS *m*/*z*: $[M - H]^-$ 370.9654, 372.9637, 374.9618 (calcd for $C_{15}H_{17}OBr_2$, 370.9646).

Claaromadendrene **6**: Colorless oil; $[\alpha]_D^{20}$ −83.0 (*c* 0.20, $CHCl_3$); ^1H and ^{13}C NMR data, see Table 1; HR-EIMS *m*/*z*: $[M]^+$ 220.1824 (calcd for $C_{15}H_{24}O$, 220.1825).

PTP1B inhibitory activity assay

The recombinant PTP1B catalytic domain was expressed and purified according to a previous report [24]. The enzymatic ac-

tivities of the PTP1B catalytic domain were determined at 30 °C by monitoring the hydrolysis of *p*NPP. Dephosphorylation of *p*NPP generated the product *p*NP, which was monitored at an absorbance of 405 nm with an EnVision multilabel plate reader (Perkin–Elmer Life Sciences, Boston, MA). In a typical 100 L assay mixture containing 50 mmol/L 3-morpholinopropanesulfonic acid, pH 6.5, 2 mmol/L *p*NPP, and 30 nmol/L recombinant PTP1B, activities were continuously monitored and the initial rate of hydrolysis was determined by using the early linear region of the enzymatic reaction kinetic curve. The IC_{50} was calculated with Prism 4 software (Graphpad, San Diego, CA) from the nonlinear curve fitting of the percentage of inhibition (% inhibition) vs the inhibitor concentration [*I*] by using the following equation: % inhibition = $1/(1 + [IC_{50}/[I]]^k)$, where *k* is the Hill coefficient; $IC_{50} \geq 50$ μM was considered inactive.

NF-κB signaling pathway inhibitory activity assays

NF-κB signaling pathway inhibitory activity was evaluated according to the previously reported protocol [25]. Stable HEK293/NF-κB cells were plated into 384-well plates at a concentration of approximate 2500 cells per well. After culturing overnight, compounds were added to the medium at a final concentration of 0.1 μg/mL. HEK293/NF-κB cells were seeded into 96-well cell culture plates (Corning, NY, USA) and allowed to grow for 24 h. The cells were then treated with compounds, followed by stimulation with TNF-α. 6 h later, the luciferase substrate was added to each well, and the released luciferin signal was detected using an EnVision microplate reader. The IC_{50} was calculated with Prism 4 software (Graphpad, San Diego, CA) from the nonlinear curve fitting of the percentage of inhibition (% inhibition) versus the inhibitor concentration [*I*] by using the following equation: % inhibition = $100/(1 + [IC_{50}/[I]]^k)$, where *k* is the Hill coefficient. Bortezomib was used as a positive control with an IC_{50} value of 14.0 μM.

Supporting Information

Supporting Information File 1

Spectral data of compounds **1**–**4** and **7**.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-64-S1.pdf]

Acknowledgements

This research work was financially supported by the National Key Research and Development Program of China (No. 2018YFC0310903), the Natural Science Foundation of China (Nos. 21672230, 81603022, 81125023), the SKLDR/SIMM Projects (SIMM 1705ZZ-01). We thank Prof. Xiu-Bao Li from Hainan University for the taxonomic identification of the soft coral material.

ORCID® iDs

Xu-Wen Li - https://orcid.org/0000-0001-7919-9726
Hong Wang - https://orcid.org/0000-0003-0058-060X

References

1. Iguchi, K.; Sawai, H.; Nishimura, H.; Fujita, M.; Yamori, T. *Bull. Chem. Soc. Jpn.* **2002,** *75,* 131–136. doi:10.1246/bcsj.75.131
2. Su, J.; Zhong, Y.; Zeng, L. *J. Nat. Prod.* **1991,** *54,* 380–385. doi:10.1021/np50074a005
3. Duh, C.-Y.; El-Gamal, A. A. H.; Chu, C.-J.; Wang, S.-K.; Dai, C.-F. *J. Nat. Prod.* **2002,** *65,* 1535–1539. doi:10.1021/np0201873
4. Shen, Y.-C.; Cheng, Y.-B.; Lin, Y.-C.; Guh, J.-H.; Teng, C.-M.; Ko, C.-L. *J. Nat. Prod.* **2004,** *67,* 542–546. doi:10.1021/np030435a
5. Iguchi, K.; Fukaya, T.; Yasumoto, A.; Watanabe, K. *J. Nat. Prod.* **2004,** *67,* 577–583. doi:10.1021/np0304013
6. Su, J.-Y.; Chen, S.; Zeng, L.-M.; Wu, H.-M. *Chin. Sci. Bull.* **1996,** *41,* 1877–1880.
7. Kobayashi, M.; Son, B.; Kido, M.; Kyogoku, Y.; Kitagawa, I. *Chem. Pharm. Bull.* **1983,** *31,* 2160–2163. doi:10.1248/cpb.31.2160
8. Kusumi, T.; Hamada, T.; Hara, M.; Ishitsuka, M. O.; Ginda, H.; Kakisawa, H. *Tetrahedron Lett.* **1992,** *33,* 2019–2022. doi:10.1016/0040-4039(92)88129-s
9. Kobayashi, M.; Son, B. W.; Kyogoku, Y.; Kitagawa, I. *Chem. Pharm. Bull.* **1984,** *32,* 1667–1670. doi:10.1248/cpb.32.1667
10. Wu, Q.; Sun, J.; Chen, J.; Zhang, H.; Guo, Y.-W.; Wang, H. *Mar. Drugs* **2018,** *16,* No. 320. doi:10.3390/md16090320
11. Wu, Q.; Li, X.-W.; Li, H.; Yao, L.-G.; Tang, W.; Miao, Z.-H.; Wang, H.; Guo, Y.-W. *Bioorg. Med. Chem. Lett.* **2019,** *29,* 185–188. doi:10.1016/j.bmcl.2018.12.004
12. Ye, F.; Li, J.; Wu, Y.; Zhu, Z.-D.; Mollo, E.; Gavagnin, M.; Gu, Y.-C.; Zhu, W.-L.; Li, X.-W.; Guo, Y.-W. *Org. Lett.* **2018,** *20,* 2637–2640. doi:10.1021/acs.orglett.8b00842
13. Ichiba, T.; Higa, T. *J. Org. Chem.* **1986,** *51,* 3364–3366. doi:10.1021/jo00367a021
14. Wratten, S. J.; Faulkner, D. J. *J. Org. Chem.* **1977,** *42,* 3343–3349. doi:10.1021/jo00441a005
15. Itokawa, H.; Yoshimoto, S.; Morita, H. *Phytochemistry* **1988,** *27,* 435–438. doi:10.1016/0031-9422(88)83115-7
16. Crews, P.; Selover, S. J. *Phytochemistry* **1986,** *25,* 1847–1852. doi:10.1016/s0031-9422(00)81160-7
17. Capon, R. J.; Ghisalberti, E. L.; Mori, T. A.; Jefferies, P. R. *J. Nat. Prod.* **1988,** *51,* 1302–1304. doi:10.1021/np50060a049
18. Suzuki, M.; Kurosawa, E. *Bull. Chem. Soc. Jpn.* **1979,** *52,* 3349–3351. doi:10.1246/bcsj.52.3349
19. Abraham, W.-R.; Kieslich, K.; Stumpf, B.; Ernst, L. *Phytochemistry* **1992,** *31,* 3749–3755. doi:10.1016/s0031-9422(00)97521-6
20. Gijsen, H. J. M.; Wijnberg, J. B. P. A.; van Ravenswaay, C.; de Groot, A. *Tetrahedron* **1994,** *50,* 4733–4744. doi:10.1016/s0040-4020(01)85012-2
21. König, G. M.; Wright, A. D. *J. Nat. Prod.* **1994,** *57,* 477–485. doi:10.1021/np50106a006
22. Appleton, D. R.; Babcock, R. C.; Copp, B. R. *Tetrahedron* **2001,** *57,* 10181–10189. doi:10.1016/s0040-4020(01)01044-4

23. Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2016,** *79,* 629–661. doi:10.1021/acs.jnatprod.5b01055
24. Li, X.-L.; He, W.-F.; Li, J.; Lan, L.-F.; Li, X.-W.; Guo, Y.-W. *J. Asian Nat. Prod. Res.* **2015,** *17,* 1146–1152. doi:10.1080/10286020.2015.1102135
25. Huang, R.-Y.; Chen, W.-T.; Kurtán, T.; Mándi, A.; Ding, J.; Li, J.; Li, X.-W.; Guo, Y.-W. *Future Med. Chem.* **2016,** *8,* 17–27. doi:10.4155/fmc.15.169

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

An efficient synthesis of the guaiane sesquiterpene (−)-isoguaiene by domino metathesis

Yuzhou Wang[1], Ahmed F. Darweesh[2], Patrick Zimdars[1] and Peter Metz[*1]

Full Research Paper

Open Access

Address:
[1]Fakultät Chemie und Lebensmittelchemie, Organische Chemie I, Technische Universität Dresden, Bergstrasse 66, 01069 Dresden, Germany, and [2]Chemistry Department, Faculty of Science, Cairo University, Giza, Egypt

Email:
Peter Metz* - peter.metz@chemie.tu-dresden.de

* Corresponding author

Keywords:
domino reactions; metathesis; Michael addition; organocatalysis; terpenes

Beilstein J. Org. Chem. **2019,** *15,* 858–862.
doi:10.3762/bjoc.15.83

Received: 16 February 2019
Accepted: 29 March 2019
Published: 09 April 2019

This article is part of the thematic issue "Terpenes".

Guest Editor: J. S. Dickschat

Abstract

(−)-Isoguaiene was prepared from (*S*)-citronellal in only 9–10 steps with good overall yields. Either a trienyne or a dienediyne metathesis and highly diastereoselective organocatalytic Michael additions of aldehydes derived from (*S*)-citronellal served as the key transformations.

Introduction

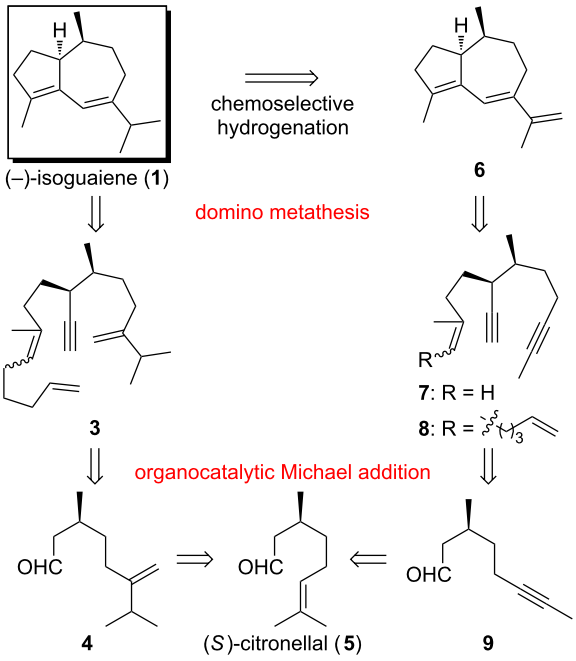
The guaiane sesquiterpene (−)-isoguaiene (**1**) has been isolated from the liverworts *Pellia epiphylla* [1] and *Dumortiera hirsuta* [2] as well as from several *Pimpinella* species [3,4], while the (+)-enantiomer of **1** has been isolated from the roots of *Parthenium hysterophorus* [5]. A recent enantioselective synthesis of (−)-isoguaiene (**1**) from (+)-dihydrocarvone [6] enabled an unambiguous assignment of its absolute configuration as depicted in Figure 1. Due to the structural similarity of **1** and the trisnorsesquiterpene clavukerin A (**2**), we were interested in developing an efficient synthetic access to **1** using a combined organocatalytic/metal-catalyzed strategy related to the one applied to the preparation of **2** [7,8].

(−)-isoguaiene (**1**) (−)-clavukerin A (**2**)

Figure 1: Structures of the sesquiterpene (−)-isoguaiene (**1**) and the trisnorsesquiterpene clavukerin A (**2**).

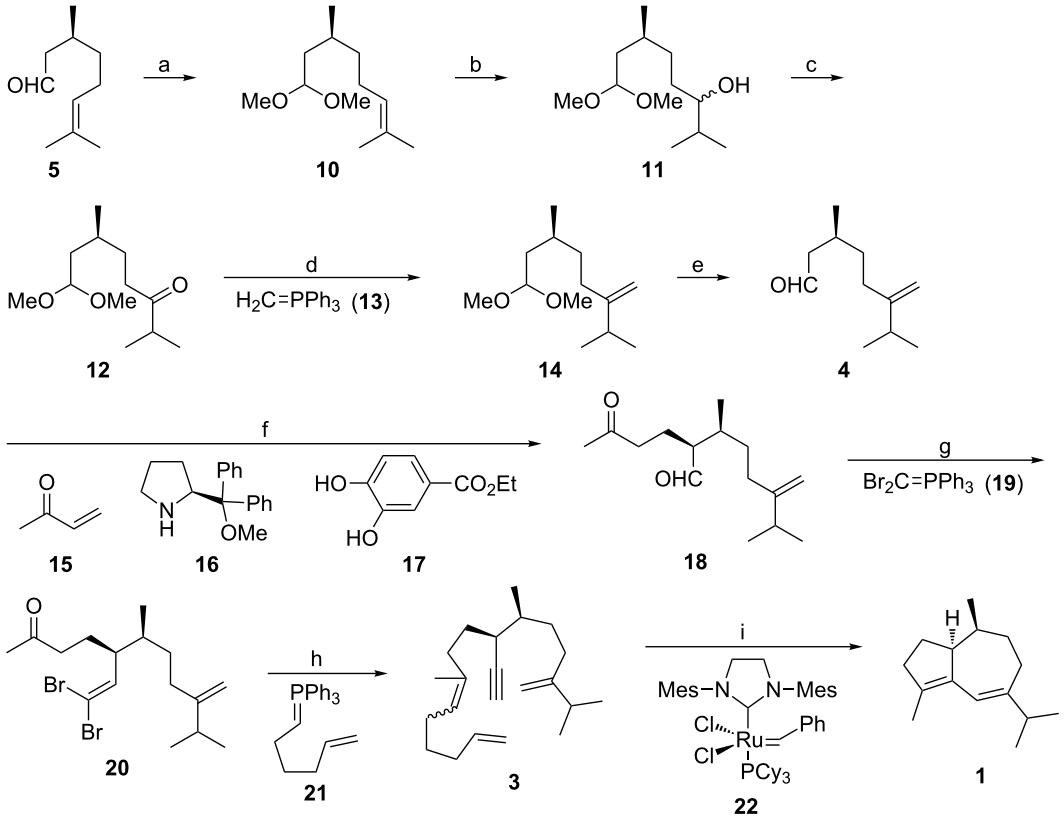
Results and Discussion

As illustrated in Scheme 1, two alternative routes were retrosynthetically devised, both of which feature a domino metathesis event and an organocatalytic Michael addition as the key steps. In closer analogy to our improved synthesis of clavukerin A (**2**) [8], a relay metathesis [9] of trienyne **3** was expected to lead to the hydroazulene **1** selectively. Trienyne **3** was envisioned to result from a stereoselective Michael addition of aldehyde **4** to methyl vinyl ketone [7,8,10] followed by chemoselective elaboration of the two carbonyl functions. Finally, aldehyde **4** was traced back to the commercially available starting material (*S*)-citronellal (**5**). On the other hand, a more rarely used enediyne metathesis [11-14] of compound **7** or its relay surrogate **8** might give rise to the conjugated triene **6**, chemoselective hydrogenation [15] of which would generate the target molecule **1**. Similar to the disconnection of trienyne **3**, the metathesis substrates **7** and **8** can be derived from aldehyde **9**, which is finally also traced back to (*S*)-citronellal (**5**).



Scheme 1: Retrosynthetic analysis for (−)-isoguaiene (**1**).

Scheme 2 illustrates the synthesis of (−)-isoguaiene (**1**) by relay metathesis of trienyne **3**. The unsaturated aldehyde **4** required for the organocatalytic Michael addition was readily prepared in

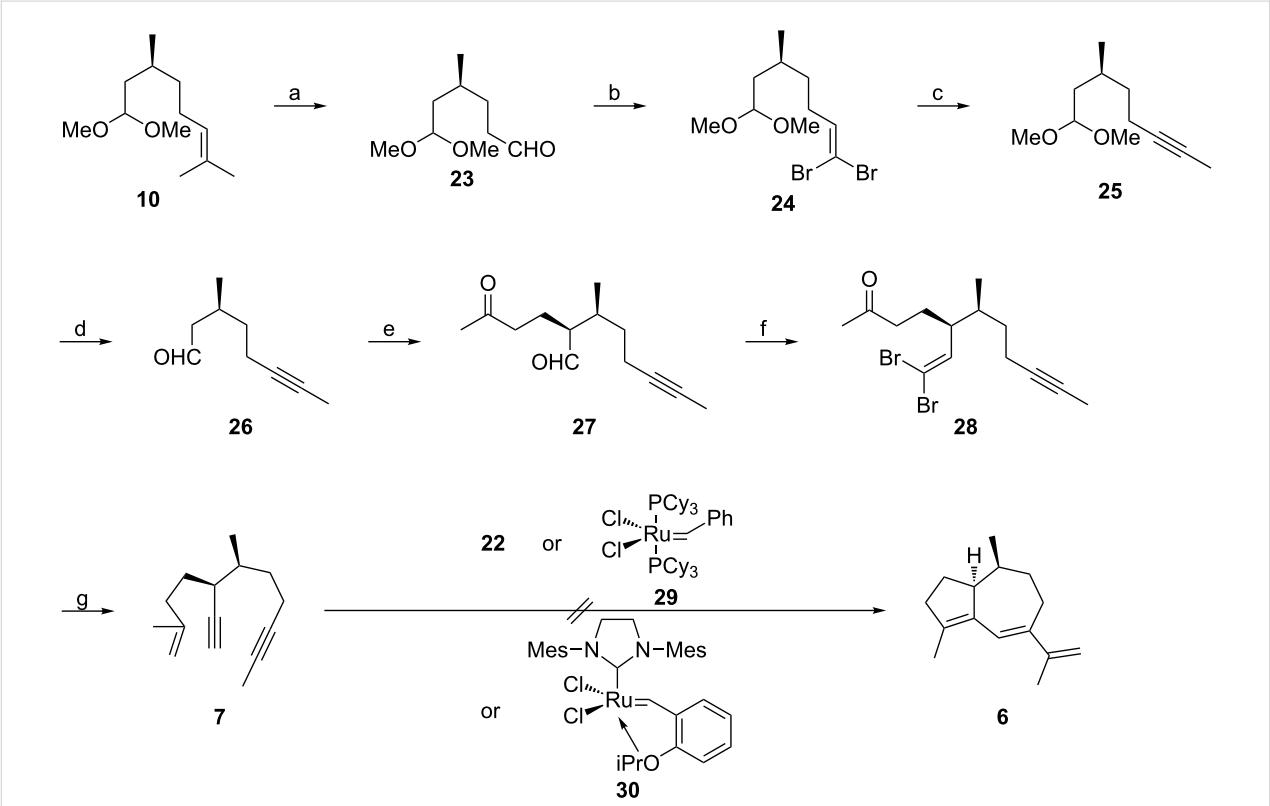


Scheme 2: Synthesis of **1** by relay metathesis of trienyne **3**. a) $HC(OMe)_3$, 4 mol % $LiBF_4$, MeOH, reflux, 80%; b) (i) $BH_3{\cdot}Me_2S$, THF, 0 °C to rt, (ii) 30% H_2O_2, 10% NaOH, 0 °C to rt, 97%; c) 5 mol % TPAP, NMO, CH_2Cl_2, rt, 97%; d) **13**, THF, rt to reflux, 96%; e) 25 mol % TsOH, THF, H_2O, iPrOH, reflux, 92%; f) **15**, 7.5 mol % **16**, 20 mol % **17**, 1 °C, 91%; g) **19**, THF, 0 °C to rt, 84%; h) (i) **21**, THF, −60°C to rt then 50 °C, (ii) BuLi, −78 °C, 76%; i) 30 mol % **22**, benzene, reflux, 51%. TPAP = tetrapropylammonium perruthenate. NMO = *N*-methylmorpholine-*N*-oxide. Ts = *p*-toluenesulfonyl.

five steps commencing with (*S*)-citronellal (**5**). After protection of the aldehyde function as the dimethyl acetal [16-18], hydroboration and oxidative work-up of **10** provided a mixture of epimeric alcohols **11** that was unified by Ley–Griffith oxidation [19] to give ketone **12** [20]. Subsequent Wittig reaction with ylide **13** and acetal cleavage of the resultant olefin **14** delivered aldehyde **4** with considerably higher efficiency compared to the known six-step preparation of *ent*-**4** from (−)-menthone [21]. Asymmetric Michael addition [7,8,10] of aldehyde **4** to methyl vinyl ketone (**15**) proceeded with high catalyst-controlled diastereoselectivity (dr = 23:1) to yield keto aldehyde **18**. Chemoselective dibromoolefination with ylide **19** prepared from dibromomethyltriphenylphosphonium bromide and sodium *tert*-butoxide [22] led to ketone **20** virtually without erosion of the relative configuration (dr = 22:1). After subjecting **20** to carbonyl olefination with unsaturated ylide **21** [8] followed by alkyne generation [23] with butyllithium in a one-pot process, trienyne **3** was obtained as a 1.6:1 mixture of *E* and *Z* olefin isomers. Due to the presence of the isopropyl group at the disubstituted alkene [24] of **3**, 30 mol % of the second generation Grubbs catalyst **22** were required to effect the relay metathesis of **3** to (−)-isoguaiene (**1**) in refluxing benzene in a good yield of 51%.

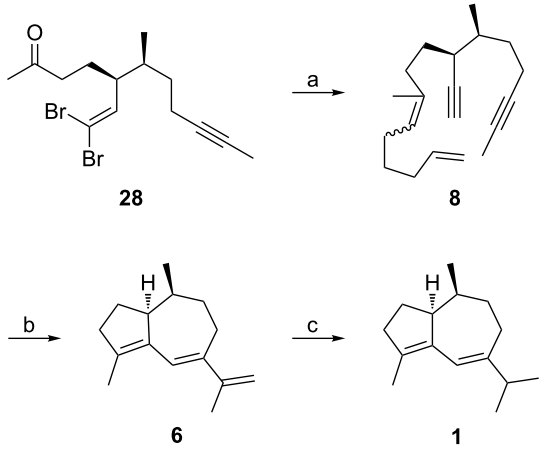
Thus, by application of a domino metathesis strategy featuring trienyne **3**, only 9 steps were needed to secure the guaiane sesquiterpene **1** in 19.7% overall yield starting from (*S*)-citronellal (**5**), which compares favorably with the previous synthesis of **1** from (+)-dihydrocarvone (10 steps, 6.9% overall yield) [6]. Scheme 3 depicts our first attempts to realize an alternative domino metathesis strategy using enediyne **7**.

Ozonolysis of the unsaturated acetal **10** gave aldehyde **23** [17,18] that was subjected to dibromoolefination with ylide **19** as described for the transformation of aldehyde **18**. Use of the preformed ylide **19** led to reproducibly higher yields of **24** in comparison with the application of tetrabromomethane and triphenylphosphine [23]. One-pot alkyne formation and methylation [23] of **24** to furnish **25** and subsequent acetal hydrolysis provided the known aldehyde **26** [25] in very good overall yield. In our hands, the "demethanation" of (*S*)-citronellol to produce the primary alcohol corresponding to aldehyde **26** according to the protocol of Abidi ($NaNO_2$, aqueous AcOH) [26] as a potential shortcut to **26** only proceeded with a maximum yield of 20%. Asymmetric Michael addition of aldehyde **26** to methyl vinyl ketone (**15**) followed immediately by treatment of the resultant unstable keto aldehyde **27** with ylide **19** delivered



Scheme 3: Attempted preparation of **1** by domino metathesis of enediyne **7**. a) (i) O_3, CH_2Cl_2, MeOH, pyridine, −78 °C, (ii) PPh_3, rt, 94%; b) **19**, THF, 0 °C; c) (i) BuLi, −78 °C, (ii) MeI, −78 °C to rt, 92% (2 steps); d) 25 mol % TsOH, THF, H_2O, iPrOH, reflux, 90%; e) **15**, 10 mol % **16**, 20 mol % **17**, 2 °C; f) **19**, THF, 0 °C to rt, 85% (2 steps); g) (i) **13**, THF, rt to reflux, (ii) BuLi, −78 °C, 90%.

dibromo olefin **28** with high diastereocontrol (dr = 19:1). Olefination of ketone **28** with ylide **13** and alkyne formation with butyllithium in a one-pot procedure then gave rise to enediyne **7** in excellent yield. Unfortunately, all attempts to achieve a domino metathesis of **7** to hydroazulene **6** only met with failure. Thus, neither the Grubbs catalysts **22** or **29**, nor the Hoveyda–Blechert catalyst **30** [27,28] in the presence or absence of ethylene effected the desired transformation to triene **6**. As a consequence, we resorted to a relay strategy for enediyne metathesis as well, and the successful execution of this idea is illustrated in Scheme 4.



Scheme 4: Conversion of **28** to **1** by relay metathesis of dienediyne **8**. a) (i) **21**, THF, rt to reflux, (ii) BuLi, −78 °C, 65%; b) 25 mol % **29**, CH_2Cl_2, reflux; c) H_2, 15 mol % $(PPh_3)_3RhCl$, benzene, EtOH, rt, 42% (2 steps).

Similar to the transformation of ketone **20**, the one-pot conversion of ketone **28** by olefination with unsaturated ylide **21** and alkyne formation with butyllithium yielded dienediyne **8** as a 1.3:1 mixture of *E* and *Z* olefin isomers. Gratifyingly, treatment of **8** with 25 mol% of the first generation Grubbs catalyst **29** produced the desired hydroazulene **6** in refluxing dichloromethane. Without purification, the crude sensitive conjugated triene **6** was immediately hydrogenated in the presence of the Wilkinson catalyst [29] to give (−)-isoguaiene (**1**) by chemoselective reduction of only the terminal olefin [15] in satisfactory yield over the 2 steps. Hence, the natural product **1** was available through this domino metathesis strategy featuring dienediyne **8** in 10 steps from (*S*)-citronellal (**5**) in 14.5% overall yield.

Conclusion

In summary, we have accomplished two short and efficient catalytic routes from (*S*)-citronellal (**5**) to the guaiane sesquiterpene (−)-isoguaiene (**1**) using either a trienyne or a dienediyne metathesis and highly diastereoselective organocatalytic Michael additions of aldehydes derived from **5** as the key steps.

Supporting Information

Supporting Information File 1

Experimental procedures and copies of ^1H NMR and ^{13}C NMR spectra of compounds **1**, **3**, **4**, **7**, **8**, **10**, **12**, **14**, **18**, **20**, **23**–**28**.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-83-S1.pdf]

Acknowledgements

This paper is dedicated to Professor Hans-Ulrich Reißig on the occasion of his 70th birthday. Dr. Ahmed Fathy Darweesh thanks the Alexander von Humboldt Foundation for funding his research stays in Dresden, during which a part of this work has been carried out.

ORCID® iDs

Ahmed F. Darweesh - https://orcid.org/0000-0002-7291-6486
Peter Metz - https://orcid.org/0000-0002-0592-9850

References

1. Cullmann, F.; Becker, H. *Phytochemistry* **1998,** *47,* 237–245. doi:10.1016/s0031-9422(97)00414-7
2. Saritas, Y.; Bülow, N.; Fricke, C.; König, W. A.; Muhle, H. *Phytochemistry* **1998,** *48,* 1019–1023. doi:10.1016/s0031-9422(97)00484-6
3. Tabanca, N.; Bedir, E.; Ferreira, D.; Slade, D.; Wedge, D. E.; Jacob, M. R.; Khan, S. I.; Kirimer, N.; Baser, K. H. C.; Khan, I. A. *Chem. Biodiversity* **2005,** *2,* 221–232. doi:10.1002/cbdv.200590005
4. Baser, K. H. C.; Tabanca, N.; Kirimer, N.; Bedir, E.; Khan, I. A.; Wedge, D. E. *Pure Appl. Chem.* **2007,** *79,* 539–556. doi:10.1351/pac200779040539
5. Bohlmann, F.; Zdero, C.; Lonitz, M. *Phytochemistry* **1977,** *16,* 575–577. doi:10.1016/0031-9422(77)80018-6
6. Blay, G.; Garcia, B.; Molina, E.; Pedro, J. R. *Org. Lett.* **2005,** *7,* 3291–3294. doi:10.1021/ol0511023
7. Knüppel, S.; Rogachev, V. O.; Metz, P. *Eur. J. Org. Chem.* **2010,** 6145–6148. doi:10.1002/ejoc.201001087
8. Barthel, A.; Kaden, F.; Jäger, A.; Metz, P. *Org. Lett.* **2016,** *18,* 3298–3301. doi:10.1021/acs.orglett.6b01619
9. Hoye, T. R.; Jeffrey, C. S.; Tennakoon, M. A.; Wang, J.; Zhao, H. *J. Am. Chem. Soc.* **2004,** *126,* 10210–10211. doi:10.1021/ja046385t
10. Chi, Y.; Gellman, S. H. *Org. Lett.* **2005,** *7,* 4253–4256. doi:10.1021/ol0517729
11. Banti, D.; North, M. *Tetrahedron Lett.* **2003,** *44,* 8157–8160. doi:10.1016/j.tetlet.2003.09.035
12. Spiegel, D. A.; Schroeder, F. C.; Duvall, J. R.; Schreiber, S. L. *J. Am. Chem. Soc.* **2006,** *128,* 14766–14767. doi:10.1021/ja065724a
13. Yuan, W.; Wei, Y.; Shi, M. *ChemistryOpen* **2013,** *2,* 63–68. doi:10.1002/open.201300002
14. Zuercher, W. J.; Scholl, M.; Grubbs, R. H. *J. Org. Chem.* **1998,** *63,* 4291–4298. doi:10.1021/jo972279g
15. Zahel, M.; Wang, Y.; Jäger, A.; Metz, P. *Eur. J. Org. Chem.* **2016,** 5881–5886. doi:10.1002/ejoc.201601197

16. Hamada, N.; Kazahaya, K.; Shimizu, H.; Sato, T. *Synlett* **2004,** 1074–1076. doi:10.1055/s-2004-820038
17. Tietze, L. F.; Denzer, H.; Holdgrün, X.; Neumann, M. *Angew. Chem.* **1987,** *99,* 1309–1310. doi:10.1002/ange.19870991225
Angew. Chem., Int. Ed. **1987**, *26*, 1295-1297. doi:10.1002/anie.198712951
18. Naruse, Y.; Yamamoto, H. *Tetrahedron* **1988,** *44,* 6021–6029. doi:10.1016/s0040-4020(01)89790-8
19. Griffith, W. P.; Ley, S. V.; Whitcombe, G. P.; White, A. D. *J. Chem. Soc., Chem. Commun.* **1987,** 1625–1627. doi:10.1039/c39870001625
20. Yarovaya, O. I.; Salomatina, O. V.; Korchagina, D. V.; Polovinka, M. P.; Barkhash, V. A. *Russ. J. Org. Chem.* **2002,** *38,* 1594–1605. doi:10.1023/a:1022501815715
21. Sutherland, A. J.; Sutherland, J. K.; Crowley, P. J. *J. Chem. Soc., Perkin Trans. 1* **1996,** 349–354. doi:10.1039/p19960000349
22. Michel, P.; Gennet, D.; Rassat, A. *Tetrahedron Lett.* **1999,** *40,* 8575–8578. doi:10.1016/s0040-4039(99)01830-4
23. Corey, E. J.; Fuchs, P. L. *Tetrahedron Lett.* **1972,** *13,* 3769–3772. doi:10.1016/s0040-4039(01)94157-7
24. Radtke, L.; Willot, M.; Sun, H.; Ziegler, S.; Sauerland, S.; Strohmann, C.; Fröhlich, R.; Habenberger, P.; Waldmann, H.; Christmann, M. *Angew. Chem.* **2011,** *123,* 4084–4088. doi:10.1002/ange.201007790
Angew. Chem., Int. Ed. **2011,** *50,* 3998–4002. doi:10.1002/anie.201007790
25. Gao, P.; Xu, P.-F.; Zhai, H. *J. Org. Chem.* **2009,** *74,* 2592–2593. doi:10.1021/jo900045k
26. Abidi, S. L. *Tetrahedron Lett.* **1986,** *27,* 267–270. doi:10.1016/s0040-4039(00)83993-3
27. Garber, S. B.; Kingsbury, J. S.; Gray, B. L.; Hoveyda, A. H. *J. Am. Chem. Soc.* **2000,** *122,* 8168–8179. doi:10.1021/ja001179g
28. Gessler, S.; Randl, S.; Blechert, S. *Tetrahedron Lett.* **2000,** *41,* 9973–9976. doi:10.1016/s0040-4039(00)01808-6
29. Osborn, J. A.; Jardine, F. H.; Young, J. F.; Wilkinson, G. *J. Chem. Soc. A* **1966,** 1711–1732. doi:10.1039/j19660001711

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

New terpenoids from the fermentation broth of the edible mushroom *Cyclocybe aegerita*

Frank Surup[1,2], Florian Hennicke[3], Nadine Sella[4], Maria Stroot[1,2], Steffen Bernecker[1,2], Sebastian Pfütze[1,2], Marc Stadler[*1,2] and Martin Rühl[*4,5]

Full Research Paper

Open Access

Address:
[1]Microbial Drugs, Helmholtz Centre for Infection Research GmbH (HZI), Inhoffenstraße 7, 38124 Braunschweig, Germany, [2]German Centre for Infection Research Association (DZIF), Partner Site Hannover-Braunschweig, Inhoffenstraße 7, 38124 Braunschweig, Germany, [3]Junior Research Group Genetics and Genomics of Fungi, Senckenberg Gesellschaft für Naturforschung, Georg-Voigt-Str. 14–16, 60325 Frankfurt am Main, Germany, [4]Institute of Food Chemistry and Food Biotechnology, Justus Liebig University Giessen, Giessen, Germany and [5]Fraunhofer Institute for Molecular Biology and Applied Ecology IME Business Area Bioresources, Heinrich-Buff-Ring 17, 35392 Giessen, Germany

Email:
Marc Stadler[*] - marc.stadler@helmholtz-hzi.de; Martin Rühl[*] - martin.ruehl@uni-giessen.de

* Corresponding author

Keywords:
bioinformatics; gene cluster analysis; natural products; secondary metabolites; structure elucidation; terpenes

Beilstein J. Org. Chem. **2019,** *15,* 1000–1007.
doi:10.3762/bjoc.15.98

Received: 19 February 2019
Accepted: 18 April 2019
Published: 30 April 2019

This article is part of the thematic issue "Terpenes".

Guest Editor: J. S. Dickschat

Abstract

The strophariaceous basidiomycete *Cyclocybe aegerita* (synonyms *Agrocybe aegerita* and *A. cylindracea*) is one of the most praised cultivated edible mushrooms and is being cultivated at large scale for food production. Furthermore, the fungus serves as a model organism to study fruiting body formation and the production of secondary metabolites during the life cycle of Basidiomycota. By studying the secondary metabolite profiles of *C. aegerita*, we found several terpenoids in submerged cultures. Aside from the main metabolite, bovistol (**1**), two new bovistol derivatives B and C (**2**, **3**) and pasteurestin C as a new protoilludane (**4**) were isolated by preparative HPLC. Their structures were elucidated by mass spectrometry and NMR spectroscopy. The relative configurations of **2**–**4** were assigned by ROESY correlations, and $^3J_{H,H}$ coupling constants in the case of **4**. Applying quantitative PCR for gene expression validation, we linked the production of bovistol and its derivatives to the respective biosynthesis gene clusters.

Introduction

The basidiomycete *Agrocybe aegerita* (synonym: *A. cylindracea*) was traditionally accommodated in the genus *Agrocybe* (family Bolbitiaceae) until a recent phylogenetic study based on comparisons of rDNA sequence data has resulted in its placement in the Strophariaceae family and it was accordingly moved into the resurrected genus *Cyclocybe* Velenovsky [1,2].

The original publication by Vizzini et al. [1] was published in a regional journal, rather than in one of the leading peer-reviewed taxonomic journals, and the authors did not follow good scientific practice for typification when proposing these taxonomic changes. However, they subsequently published their entries in Index Fungorum, making them valid according to the current nomenclature rules, and the phylogenetic tree they presented clearly revealed that the species now accommodated in the genus *Cyclocybe* are more closely related to the Strophariaceae and phylogenetically rather distinct from those of the family Bolbitiaceae, including the type species of *Agrocybe* which is *A. praecox*. Therefore, the currently valid scientific name of the fungus is *Cyclocybe aegerita* (V. Brig.) Vizzini.

In fact, *C. aegerita* is a rather important fungal species with regard to practical applications, as i) it belongs to the edible mushrooms that are being cultivated at industrial scale and it is highly esteemed for its excellent aroma and ii) it has been used as a model organism for microbiological and genetic investigations on fruiting body formation for many years. Herzog et al. [3] have recently reported a parental dikaryotic strain *C. aegerita* AAE-3 that completes its life cycle on agar plates in only three weeks, along with a set of sibling monokaryons derived from it. Among these monokaryons, *C. aegerita* AAE-3-32 has been used for histological analysis of monokaryotic fruiting sensu stricto (mushroom formation without previous mating) and, together with *C. aegerita* AAE-3-13, for exploring molecular tools for transformation and gene of interest expression, which has just been published [4]. These strains could serve well for studies exploring the factors regulating monokaryotic fruiting in comparison to dikaryotic mushroom formation. In addition, strains of this fungal species show a reliable growth behaviour in liquid culture and could eventually serve as hosts for heterologous production of secondary metabolites derived from other Basidiomycota that are more difficult or even impossible to culture. With these goals in mind, we have initiated extensive studies of the secondary metabolism of the aforementioned strains, targeting both volatile and non-volatile compounds. The present paper will describe the discovery of one known and three new non-volatile terpenoids (Figure 1) that were isolated from liquid cultures of *C. aegerita* and their physicochemical and preliminary biological characterisation.

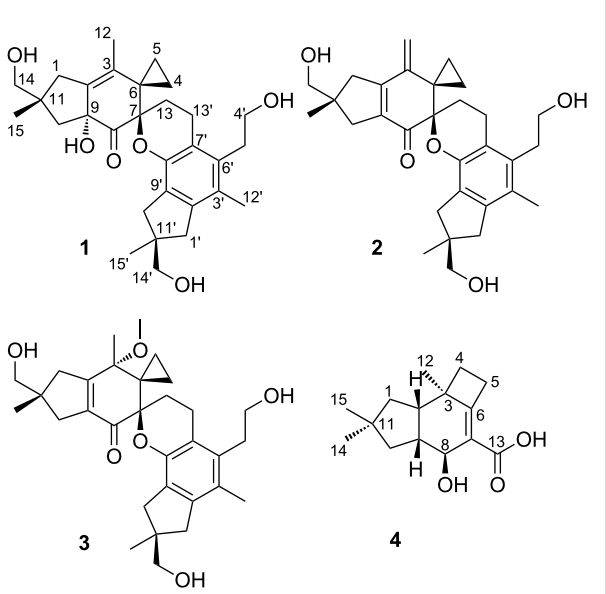


Figure 1: Structures of the isolated metabolites bovistol A (**1**), its new derivatives bovistol B (**2**) and C (**3**), as well as the new protoilludane pasteurestin C (**4**).

Results and Discussion

Both the ethyl acetate extract of the culture filtrate and the acetone extract of the mycelium of *C. aegerita*, grown in ZM/2 medium, contained a major metabolite with a molecular mass of 496 Da, as detected by HPLC–MS analysis. Its molecular formula $C_{30}H_{40}O_6$ was deduced from its $[M + Na]^+$ peak at *m/z* 519.2718 in the HRESIMS spectrum. ^1H and $^1H,^{13}C$-HSQC NMR data (Table 1) indicated the presence of four methyls and twelve methylenes, three of them oxomethylene groups. A database search with this data within the Chapman & Hall Dictionary of Natural Products on DVD suggested its identity as bovistol, which was confirmed by the elucidation of the structure by COSY and HMBC NMR data [5].

In the course of the isolation of **1** the minor metabolites **2** and **3** accrued. Metabolite **2** was analysed for a molecular weight of 478 Da. Its molecular formula $C_{30}H_{38}O_5$, deduced from HRESIMS data, indicated the formal loss of one molecule of water. ^1H and HSQC NMR data (Table 1) of **2** were very similar to those of **1**, with the exception of the replacement of methyl CH_3-12 by an exomethylene group. HMBC correlations from both exo-methylene protons 12-H_2 to C-2, C-3, C-4, in addition to those from both 1-H_2 and 10-H_2 to the olefinic carbons C-2 and C-9, confirmed the structure of **2** (Figure 1). For **3**, HRESIMS data revealed its molecular formula as $C_{31}H_{42}O_6$. The ^1H and ^{13}C NMR data (Table 1) were highly similar to those of **1**, with the key difference being an additional methoxy group (δ_H 3.33/δ_C 50.5). This methoxy was connected to C-3 due to its HMBC correlation to this carbon atom, along to the ones from 1-H_2, 10-H_2 and 12-H_3 to C-3. Compared to **1**, the $\Delta^{2,3}$ double bond is shifted to $\Delta^{2,9}$, explaining the high field shift of the conjugated ketone C-8 (δ_C 197.2). A ROESY correlation between 3-OCH_3 and 13-H_a revealed a 3*S* configuration. The closest known structural relative of **3** is the 3-demethoxy-3-hydroxy derivative of **3**, which has been described as a spontaneous dimerization product of psathyrellon B [5].

Table 1: NMR shifts (^{1}H 700 MHz, ^{13}C 175 MHz) of bovistol A–C (**1**–**3**) in chloroform-*d*.

	1		**2**		**3**	
Atom#	^{13}C, mult.	^{1}H, mult. (*J*, Hz)	^{13}C, mult.	^{1}H, mult. (*J*, Hz)	^{13}C, mult.	^{1}H, mult. (*J*, Hz)
1	38.8, CH_2	2.45, d (15.4)/2.40, d (15.4)	40.0, CH_2	2.32, m/2.70, m	43.3, CH_2	2.75, dt (18.1, 2.0)/2.43, br d (18.1)
2	135.4, C		156.0, C		161.1, C	
3	130.7, C		143.6, C		75.4, C	
4	7.5, CH_2	1.26, m/0.64, ddd (9.8, 5.9, 5.7)	5.2, CH_2	1.05, m/0.87, m	4.0, CH_2	1.11, ddd (9.3, 6.7, 4.8)/0.30, ddd (9.3, 6.6, 4.8)
5	8.6, CH_2	0.89, dt (10.0, 5.7)/0.79, dt (10.0, 5.9)	13.1, CH_2	1.46, m/0.29, m	5.7, CH_2	0.88, m
6	32.4, C		30.9, C		30.0, C	
7	80.9, C		79.3, C		79.7, C	
8	207.6, C		196.7, C		197.2, C	
9	80.8, C		134.2, C		134.2, C	
10	44.9, CH_2	1.94, d (14.2)/1.48, d (14.2)	43.1, CH_2	2.75, m/2.55, m	39.5, CH_2	2.58, dt (16.8, 2.0)/2.24, br d (16.8)
11	41.2, C		42.8, C		42.5, C	
12	15.1, CH_3	1.50, s	112.3, CH_2	5.16, s/5.11, s	18.5, CH_3	1.164, s
13	27.7, CH_2	2.84, m/1.72, td (13.0, 5.8)	29.2, CH_2	2.21, m/1.89, td (13.5, 5.0)	28.5, CH_2	2.78, m/1.72, td (13.5, 5.0)
14	71.6, CH_2	3.26, s	70.3, CH_2	3.49, s	70.4, CH_2	3.48, s
15	26.4, CH_3	1.26, s	24.8, CH_3	1.20, s	24.7, CH_3	1.160, s
1’	43.0, CH_2	2.86, d (16.0)/2.61, d (16.0)	42.8, CH_2	2.86, m/2.60, m	42.8, CH_2	2.84, d (16.0)/2.59, d (16.0)
2’	141.6, C		141.4, C		141.1, C	
3’	124.5, C		124.2, C		124.1, C	
4’	61.7, CH_2	3.69, m	62.0, CH_2	3.70, br t (6.8)	62.1, CH_2	3.71, m
5’	31.9, CH_2	2.83, m	32.2, CH_2	2.86, m	32.1, CH_2	2.88, m
6’	133.0, C		132.4, C		132.2, C	
7’	118.6, C		118.4, C		119.1, C	
8’	148.5, C		149.2, C		149.1, C	
9’	127.2, C		126.8, C		126.6, C	
10’	39.4, CH_2	2.94, d (16.4)/2.55, d (16.4)	39.6, CH_2	2.86, m/2.52, m	39.6, CH_2	2.83, d (16.5)/2.51, d (16.5)
11’	43.8, C		43.9, C		43.9, C	
12’	15.6, CH_3	2.12, s	15.5, CH_3	2.14, s	15.5, CH_3	2.14, s
13’	21.2, CH_2	2.72, m	20.1, CH_2	2.40, m	19.9, CH_2	2.78, m/2.40, m
14’	71.3, CH_2	3.50, s	71.2, CH_2	3.51, s	71.2, CH_2	3.51, s
15’	25.1, CH_3	1.19, s	24.9, CH_3	1.21, s	24.9, CH_3	1.20, s
3OMe					50.5, CH_3	3.33, s

In addition to disesquiterpenoids **1**–**3**, metabolite **4** with a molecular mass of 250 Da was isolated. ^{1}H and ^{1}H,^{13}C-HSQC NMR data (Table 2) revealed the presence of three methyls, four methylenes and three methines, one of them oxygenated. The ^{13}C NMR spectrum indicated furthermore a carboxylic acid in addition to two olefinic and two aliphatic carbons devoid bound protons. A large spin system comprising 1-H_2/2-H/9-H(10-H_2)/8-H was constructed by COSY and TOCSY correlations in addition to the small one of 4-H_2/5-H_2. These spin systems were connected by HMBC correlations to form the protoilludane skeleton, mainly to note the correlations from 14-H_3 and 15-H_3 to 1-H_2/10-H_2, 12-H_3 to C-2/C-3/C-4/C-6, 5-H_2 to C-6/C-7 and 8-H to C-6/C-7/C-13. The strong ROESY correlation between 2-H and 9-H indicated a *cis* configuration between these protons, and the large coupling constant between 8-H and 9-H, observed in the signal of 8-H, a *trans* configuration of 8-H/9-H. Finally, the ROESY correlation between 12-H_3 and 1-H_b indicated methyl C-12 being on the opposite site of the molecular plane as 2-H and 9-H. This assignment was confirmed by the comparison of the ^{13}C NMR shifts to pasteurestin A and B [6,7]. Since the absolute stereochemistry has been demonstrated for pasteurestins A and B by total synthesis, we tentatively conclude a 2*S*,3*R*,8*S*,9*R* absolute configuration for pasteurestin C (**4**). The systematic name

Table 2: NMR data (^{1}H 700 MHz, ^{13}C 175 MHz) of compound **4** in acetone-d_6.

Atom#	C Shift	H Shift	COSY	HMBC
1	41.6, CH_2	1.44, m	1, 2	14, 11, 10, 9
		1.39, m	1, 2	14, 15, 11, 2, 3
2	45.9, CH	2.43, m	1, 1	9, 12, 4, 3, 9, 8
3	47.8, C			
4	36.4, CH_2	1.95, m	5, 5	12, 5, 2, 3, 7, 6
5	29.9, CH_2	3.13, m	4, 5, 8	4, 3, 7, 6
		3.02, m	4, 5	3, 7, 6
6	170.4, C			
7	122.1, C			
8	72.3, CH	4.27, dt (8.0, 2.0)	9, 5	10, 7, 13, 6
9	50.9, CH	2.40, m	10, 10, 8	2, 3, 8
10	47.4, CH_2	1.12, m	10, 9	14, 15, 11, 9, 8
		1.793, br dd (11.2, 7.5)	10, 9	14, 1, 2
11	40.4, C			
12	20.3, CH_3	1.14, m		4, 2, 3, 6
13	167.8, C			
14	27.4, CH_3	0.96, s		15, 11, 1, 10
15	29.8, CH_3	1.09, m		14, 11, 1, 10

for **4** is (4*S*,4a*R*,7a*S*,7b*R*)-4-hydroxy-6,6,7b-trimethyl-2,4,4a,5,6,7,7a,7b-octahydro-1*H*-cyclobuta[*e*]indene-3-carboxylic acid.

Bovistol A (**1**) showed weak cytotoxic effects (IC_{50} for L929 = 15 µg/mL, for KB3.1 = 7 µg/mL), but was inactive against all test organisms in our standard test panel, comprising selected Gram-positive and Gram-negative bacteria as well as fungi [8]. Compound **4** was inactive in all assays of our test panel, and **3** could not be tested due to the insufficient amount isolated.

Our finding of the production of **1**–**4** by *C. aegerita* expands the number of secondary metabolites known from this fungus. From fungal cultures of the genus *Cyclocybe* the production of a broad variety of metabolites is known. This includes polyacetylenes [9,10] as well as sesquiterpenoids with illudine [11], aromadendrane [12], marasmene [13] and fomannosane [14] type skeletons. Although bovistol could formally be supposed to be a triterpene, it is thought to be derived by a hetero-Diels–Alder reaction of two sesquiterpenes to form a dimeric sesquiterpenoid [15].

In the recently published genome of *C. aegerita* [16], two putative sesquiterpene synthase gene clusters have been identified on the basis of the published Δ^6-protoilludene gene cluster of *Omphalotus olearius* [17]. The protein sequences of the genes clustering adjacent to the putative sesquiterpene synthase genes going by the gene IDs AAE3_04120 and AAE3_10454 (http://www.thines-lab.senckenberg.de/agrocybe_genome) reveal the presence of P450 monooxygenases, oxidoreductases as well as one putative Diels-Alderase 1 kb downstream of the putative Δ^6-protoilludene synthase gene going by the gene ID AAE3_04120. To determine the correspondence between both putative sesquiterpene synthases and the analysed secondary metabolites **1**–**4** *C. aegerita* AAE-3 was cultivated in a stirred vessel bioreactor. Mycelial samples were analysed for the presence of the gene transcripts and **1** and **3** (Figure 2). At the beginning of the fermentation, the putative sesquiterpene synthase gene with the gene ID AAE3_04120 was upregulated with a maximum of transcripts at day 7 of cultivation. The second gene with the gene ID AAE3_10454 showed a more slight increase expression peaking at day 11 and day 14. The peak area of **1** in the supernatant increased until day 9 of cultivation and lowered afterwards, whereas in the mycelium the bovistol peak area increased until day 11 and dropped afterwards. This steady increase of **1** until day 9 respectively day 11 resembles the preceding transcriptional upregulation of the expression of the gene with the ID AAE3_04120 which indicates that this gene is presumably involved in the bovistol synthesis pathway. In addition, the deduced protein sequence of the adjacent gene with the ID AAE3_04121 shows similarities to the Diels-Alderase Sol5 from *Alternaria solani*, which is involved in the cycloaddition of prosolanapyrone II into solanapyrone [18]. A similar cycloaddition is needed to form bovistol out of two illudine precursors. Although the respective illudine has not been detected in this study, it is known that

C. aegerita is able to produce several illudanes [11]. As proposed in recent studies on sesquiterpenes in *O. olearius* [17], *Stereum hirsutum* [19] and *Diaporthe* sp. [20] the illudin precursor Δ^6-protoilludene undergoes subsequent reactions catalyzed by enzymes, whose genes cluster with the Δ^6-protoilludene synthase. Nevertheless, evidence is still missing. In *C. aegerita*, Δ^6-protoilludene is seemingly provided by the putative sesquiterpene synthase AAE3_04120. Combining these indications, an enzymatic cascade coded by the AAE3_04120 cluster is transforming Δ^6-protoilludene into illudine(s) and the corresponding bovistol in *C. aegerita* (Figure S2, Supporting Information File 1). Further research has to verify this assumption.

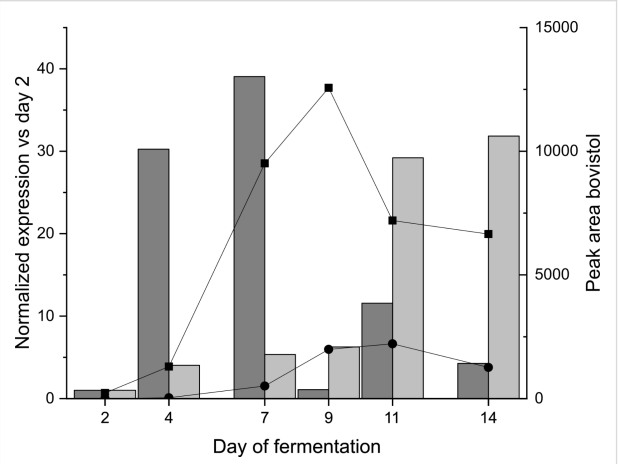


Figure 2: Relative normalized expression of the putative sesquiterpene synthase genes going by the gene IDs AAE3_04120 (dark grey) and AAE3_10454 (light grey) from the *C. aegerita* AAE-3 genome ([16], http://www.thines-ab.senckenberg.de/agrocybe_genome) compared to the peak areas of **1** in the supernatant (black squares) and in the fungal mycelium (black circles).

Conclusion

We identified bovistol A (**1**) as the main metabolite of *C. aegerita* in cultures. In parallel, we isolated its new disesquiterpenoid derivatives bovistol B (**2**) and C (**3**) as well as the new protoilludane pasteurestin C (**4**). By a qPCR approach, we were able to link the production of bovistol to the putative Δ^6-protoilludene synthase AAE3_04120 whose gene is located adjacent to a Diels-Alderase needed for cycloaddition of two illudine monomers. With this information, we made a biosynthesis proposal for these metabolites. Further studies will address this assumption to prove its validity.

Experimental

General

Optical rotations were measured on a Perkin-Elmer 241 spectrometer, the UV spectra on a Shimadzu UV–vis spectrophotometer UV-2450. NMR spectra were recorded with a Bruker Avance III 700 spectrometer, equipped with 5 mm TCI cryoprobe (^{1}H 700 MHz, ^{13}C 175 MHz). Chemical shifts δ were referenced to the solvents chloroform-*d* (^{1}H, δ = 7.27 ppm; ^{13}C, δ = 77.0 ppm), acetone-d_6 (^{1}H, δ = 2.05 ppm; ^{13}C, δ = 29.92 ppm). ESIMS spectra were acquired on an Amazon ion trap mass spectrometer (Bruker Daltonik); HRESIMS spectra were acquired on a Maxis time-of-flight mass spectrometer (Bruker Daltonik), both combined with an Agilent 1200 series HPLC-UV system [column 2.1 × 50 mm, 1.7 µm, C18 Acuity UPLC BEH (Waters), solvent A: H_2O + 0.1% formic acid; solvent B: ACN + 0.1% formic acid, gradient: 5% B for 0.5 min increasing to 100% B in 19.5 min, maintaining 100% B for 5 min, flow rate = 0.6 mL min^{-1}, UV detection at 200–600 nm].

Fermentation in shaking flasks

A culture of the strain *C. aegerita* AAE-3 with a total volume of 4 L was prepared in ZM ½ medium [molasses 47% (Nord Zucker AG Schladen, Germany) 5.00 g/L, oatmeal (Herrnmühle, Harald Feick OHG, Reichelsheim, Germany) 5.00 g/L, D(+)-sucrose (Carl Roth GmbH & Co, Karlsruhe, Germany) 4.00 g/L, D-mannitol (AppliChem GmbH, ITW Company, Darmstadt, Germany) 4.00 g/L, D-glucose monohydrate (Cargill Holding Germany GmbH, Krefeld, Germany) 1.50 g/L, $CaCO_3$ (Carl Roth GmbH & Co KG, Karlsruhe, Germany) 1.50 g/L, lactalbumin hydrolysate (Oxoid LDT, Basingstocke, Hampshire, England) 0.50 g/L, $(NH_4)_2SO_4$ 0.50 g/L]. The pH value of the medium was set to 7.2. To inoculate the medium, 5 mL of a preculture were used. The inoculated shaking flasks were incubated on a lab shaker (NORD Drive systems SK CSX-3H, Getriebebau NORD GmbH & Co KG Drive Systems) at 23 °C. After one week of incubation the glucose concentration in the medium was tested daily. If no more free glucose could be found in the medium, the cultivation was extracted.

Extraction

The mycelium was separated from the fermentation broth with the help of a fluted filter. The filtered mycelium was overlaid with actone (1:1 ratio) and digested in an ultrasonic bath (Sonorex digital 10P, BANDELIN electronic GmbH & Co KG, Berlin, Germany) for 30 min (two times). After the digest the mycelium was separated from the filtrate. The filtrate was narrowed up to the aqueous phase in vacuo. The aqueous phase was shaken out in a 1:1 ratio with ethyl acetate (+ 50 mL deionised H_2O) in a separating funnel (two times) and dried with water-free sodium sulfate (Na_2SO_4, ACROS Organics, New Jersey, USA). The sodium sulfate was filtered off and the extract was narrowed up to the dry with a rotary evaporator. The crude extract was dissolved in 4 mL methanol, evaporated under a nitrogen evaporator (VLM GmbH, Bielefeld, Germany)

and stored in a freezer at −20 °C. The crude extract of the fermentation broth was obtained the same way.

Isolation

The crude extract obtained from the extraction was dissolved in 1 mL of methanol and further separated via RP-LC with deionized water (+ 0.05% TFA, solvent A) and methanol (+ 0.05%, solvent B) by a Gilson RP-HPLC system (Middleton, Wisconsin, USA) equipped with a GX-271 Liquid handler, a diode array detector (DAD) 172 and a 305 and 306 pump. The separation was performed with a VP Nucleodur C18ec (150 × 40 mm, 7 µm; Macherey-Nagel, Düren, Germany) column and a flow rate of 20 mL/min. The gradient was set from 30 to 70% of solvent B in 45 min, with an increase to 100% B in 15 min, followed by isocratic conditions at 100% B for 15 min. All LC fractions were collected according to the UV absorption at 210 nm. Methanol was evaporated in vacuo. The aqueous residues were frozen and then removed by using an Alpha 1-4 LSC freeze dryer (Christ, Osterode, Germany). 12.6 mg of fraction V were obtained as a mixture of compounds **1**, **2** and **3**. Furthermore fraction VI yielded 5.2 mg of pure compound **4**.

Subsequently fraction V was separated via another RP-LC under different conditions. The separation was performed with the same Gilson RP-HPLC system and flow by using a Gemini 10u C18 (250 × 21.20 mm, 10 micron, Phenomenex Inc., Torrance, USA) column. Solvent B was changed to acetonitrile and also no TFA was added. The gradient was set from 50% to 75% of solvent B in 45 min with an increase to 100% B in 15 min, followed by isocratic conditions at 100% B for 15 min. Collection and work-up of the obtained fractions was performed as described above. This separation yielded 4 mg of compound **1**, 2.2 mg of compound **2** and 0.4 mg of compound **3**.

Bovistol A (1): Colourless oil; ^{1}H and ^{13}C NMR data (^{1}H 700 MHz, ^{13}C 175 MHz) in $CDCl_3$: see Table 1; ESIMS (*m/z*): 1015.52 $[2M + Na]^+$, 479.30 $[M + H - H_2O]^+$, 991.64 $[2M - H]^-$, 541.39 $[M + HCOO]^-$; HRESIMS (*m/z*): $[M + Na]^+$ calcd for $C_{30}H_{40}O_6Na$, 519.2717; found, 519.2718.

Bovistol B (2): colourless oil; $[\alpha]_D^{25}$ +3 (*c* 0.1, CH_3OH); ^{1}H and ^{13}C NMR data (^{1}H 700 MHz, ^{13}C 175 MHz) in $CDCl_3$: see Table 1; ESIMS (*m/z*): 979.57 $[2M + Na]^+$, 957.61 $[2M + H]^+$, 479.35 $[M + H]^+$, 523.30 $[M + HCOO]^-$, 477.29 $[M - H]^-$; HRESIMS (*m/z*): $[M + Na]^+$ calcd for $C_{30}H_{38}O_5Na$, 501.2585; found, 501.2606.

Bovistol C (3): Colourless oil; $[\alpha]_D^{25}$ +106 (*c* 0.1, CH_3OH); ^{1}H and ^{13}C NMR data (^{1}H 700 MHz, ^{13}C 175 MHz) in $CDCl_3$: see Table 1; ESIMS (*m/z*): 1043.65 $[2M + Na]^+$, 533.37 $[M + Na]^+$, 493.31 $[M + H - H_2O]^+$, 509.30 $[M - H]^-$, 555.32 $[M + HCOO]^-$; HRESIMS (*m/z*): $[M + Na]^+$ calcd for $C_{31}H_{42}O_6Na$, 533.2874; found, 533.2859; $[M + H]^+$ calcd for $C_{31}H_{43}O_6$, 511.3054; found, 511.3044.

Pasteurestin C (4): Colourless oil; $[\alpha]_D^{25}$ +21 (*c* 0.1, CH_3OH); ^{1}H and ^{13}C NMR data (^{1}H 700 MHz, ^{13}C 175 MHz) in acetone-d_6: see Table 2; ESIMS (*m/z*): 523.32 $[2M + Na]^+$, 233.13 $[M + H - H_2O]^+$, 249.04 $[M - H]^-$; HRESIMS (*m/z*): $[M + Na]^+$ calcd for $C_{15}H_{22}O_3Na$, 273.1461; found, 273.1460; $[M + H]^+$ calcd for $C_{15}H_{23}O_3$, 251.1642; found, 251.1637.

Fermentation in 10 L scale bioreactor

A seed culture of the strain *C. aegerita* AAE-3 with a total volume of 500 mL was prepared in ZM ½ medium. After incubation for 11 days the seed culture was homogenized with an ULTRA-TURRAX under sterile conditions and used for inoculation of a 15 L bioreactor (xCUBIO in-situ bbi biotech) filled with 10 L ZM ½ medium. The pH value was set to 7.2 (unregulated); the DO was also not regulated. The temperature was regulated at 23 °C. Furthermore the submerged aeration rate and stirrer speed was fixed to 0.15 vvm and 200 rpm (rushton turbine). For foam destruction into the process Tego Antifoam D2310 (Evonik Nutrition & Care GmbH) was used. After 29 days of cultivation the fermentation broth was harvested. The biomass and suspended substrates was separated by centrifugation.

RNA extraction, cDNA synthesis and qPCR

During fermentation, mycelial samples were taken at day 2, 4, 7, 9, 11 and 14 and stored in RNAlater (Qiagen, Venlo, Netherlands) until further use. Fungal mycelium was freeze-dried and ground with liquid nitrogen. RNA was extracted from ground mycelium using Ambion TRIzol™ Reagent (life Technologies, Carlsberg, California, USA) according to the manufacturer's instructions with minor changes according to the method of Chomczynski and Sacchi [21]. RNA concentration was determined photometrically by a NanoPhotometer® Pearl (Implen, Munich, Germany). Reverse transcription was performed with the Invitrogen M-MLV Reverse Transcriptase kit (ThermoFisher Scientific, Waltham, Massachusetts, USA) according to the manufacturer's protocol. 10 µL of extracted RNA and 1 µL of 10 µM oligo-$(dT)_{30}$ primer (Eurofins, Waltham, Massachusetts, USA) were used for cDNA synthesis. For removal of RNA in the transcribed cDNA sample, 1 µL of AMRESCO RNase A (VWR International, Radnor, Pennsylvania, USA) was added and the mixture was incubated at 37 °C for 20 min. Primers for qPCR analysis were designed using Geneious 11.0.4. (Biomatters, Auckland, New Zealand). Primer pairs for *C. aegerita* housekeeping genes going

by the gene IDs AAE3_02268 and AAE3_07669 (http://www.thines-lab.senckenberg.de/agrocybe_genome) have been identified and validated by NormFinder and geNorm algorithm to be the best combination for qPCR-based transcription analyses of *C. aegerita* by means of qPCR (data not published). Briefly, KAPA SYBR® FAST qPCR Master Mix (Kapa Biosystems, Wilmington, MA, USA), 900 nM forward primer, 900 nM reverse Primer (Table 3), 10 ng of cDNA and nuclease-free water were mixed. The qPCR reactions were performed in triplicates using the CFX Connect™ RT-PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA). The following conditions were applied: enzyme activation at 94 °C for 20 s followed by 40 cycles of 94 °C for 30 s, 58 °C for 30 s and 72 °C for 10 s.

Supporting Information

Supporting Information File 1

^{1}H and ^{13}C NMR spectra of compound **1** and ^{1}H, ^{13}C, COSY, ROESY, HSQC and HMBC NMR spectra of compounds **2**–**4**.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-98-S1.pdf]

Acknowledgements

We kindly thank C. Kakoschke for recording NMR spectra, A. Gollasch und S. Karwehl for HRESIMS measurements, V. Stiller for technical assistance with cultivation, W. Collisi for conducting bioassays and S. Hüttel and co-workers for the large scale fermentation. N. Sella is financed by the LOEWE program (State-Offensive for the Development of Scientific and Economic Excellence) “AROMAplus”. The authors are thankful for this support.

ORCID® iDs

Frank Surup - https://orcid.org/0000-0001-5234-8525
Marc Stadler - https://orcid.org/0000-0002-7284-8671
Martin Rühl - https://orcid.org/0000-0001-8274-8175

References

1. Vizzini, A. L.; Angelini, C. L.; Ercole, E. N. *Riv. Micol. Rom.* **2014,** *30,* 21–38.
2. Nauta, M. M. Genus *Agrocybe*. In *Flora Agaricina Neerlandica;* Noordeloos, M. E.; Kuyper, T. W.; Velinga, E. C., Eds.; CRC Press: Boca Raton, 2005; Vol. 6, pp 204–221.
3. Herzog, R.; Solovyeva, I.; Rühl, M.; Thines, M.; Hennicke, F. *Mycol. Prog.* **2016,** *15,* 947–957. doi:10.1007/s11557-016-1221-9
4. Herzog, R.; Solovyeva, I.; Bölker, M.; Lugones, L. G.; Hennicke, F. *Mol. Genet. Genomics* **2019**. doi:10.1007/s00438-018-01528-6
5. Rasser, F.; Anke, T.; Sterner, O. *Tetrahedron* **2002,** *58,* 7785–7789. doi:10.1016/s0040-4020(02)00943-2
6. Kögl, M.; Brecker, L.; Warrass, R.; Mulzer, J. *Angew. Chem., Int. Ed.* **2007,** *46,* 9320–9322. doi:10.1002/anie.200703457
7. Kögl, M.; Brecker, L.; Warrass, R.; Mulzer, J. *Eur. J. Org. Chem.* **2008,** 2714–2730. doi:10.1002/ejoc.200800074
8. Surup, F.; Halecker, S.; Nimtz, M.; Rodrigo, S.; Schulz, B.; Steinert, M.; Stadler, M. *Steroids* **2018,** *135,* 92–97. doi:10.1016/j.steroids.2018.03.007
9. Kavanagh, F.; Hervey, A.; Robbins, W. J. *Proc. Natl. Acad. Sci. U. S. A.* **1950,** *36,* 102–106. doi:10.1073/pnas.36.2.102
10. Fushimi, K.; Anzai, K.; Tokuyama, S.; Kiriiwa, Y.; Matsumoto, N.; Sekiya, A.; Hashizume, D.; Nagasawa, K.; Hirai, H.; Kawagishi, H. *Tetrahedron* **2012,** *68,* 1262–1265. doi:10.1016/j.tet.2011.11.049
11. Stránský, K.; Semerdžieva, M.; Otmar, M.; Procházka, Ž.; Buděšínský, M.; Ubik, K.; Kohoutová, J.; Streinz, L. *Collect. Czech. Chem. Commun.* **1992,** *57,* 590–603. doi:10.1135/cccc19920590
12. Zhu, Y.-C.; Wang, G.; Liu, J.-K. *J. Asian Nat. Prod. Res.* **2010,** *12,* 464–469. doi:10.1080/10286020.2010.489822
13. Berg, A.; Dörfelt, H.; Kiet, T. T.; Schlegel, B.; Gräfe, U. *J. Antibiot.* **2002,** *55,* 818–820. doi:10.7164/antibiotics.55.818

Table 3: Primer sequences for qPCR.

Gene ID[a]	Primer sequence (5’ to 3’)	Exon spanning
AAE3_02268	AGATGCGTATTCTGATGGTTGGTC	yes
	CCCACACTGTGAATGAGATGTTC	yes
AAE3_07669	ATTCCTACGATCCTTTTGCCG	yes
	GATCATATTGTTTCGGGAGTCCT	yes
AAE3_04120	GCGAAGACACAGTTTTGACAG	yes
	TGAGCTAACGTCATCAATCCC	yes
AAE3_10454	ATGCGACTTCAATCTTTTGGC	yes
	ATGTCGGCCATGCGTCTT	no

[a]Referring to the gene IDs from the genomic sequence of *C. aegerita* AAE-3 (http://www.thines-lab.senckenberg.de/agrocybe_genome).

14. Liu, L.-Y.; Li, Z.-H.; Dong, Z.-J.; Li, X.-Y.; Su, J.; Li, Y.; Liu, J.-K. *Nat. Prod. Bioprospect.* **2012,** *2,* 130–132. doi:10.1007/s13659-012-0031-2
15. Liao, S.-G.; Yue, J.-M. Dimeric Sesquiterpenoids. In *Progress in the Chemistry of Organic Natural Products;* Kinghorn, A.; Falk, H.; Gibbons, S.; Kobayashi, J., Eds.; Springer International Publishing: Cham, Switzerland, 2016; Vol. 101, pp 1–112. doi:10.1007/978-3-319-22692-7_1
16. Gupta, D. K.; Rühl, M.; Mishra, B.; Kleofas, V.; Hofrichter, M.; Herzog, R.; Pecyna, M. J.; Sharma, R.; Kellner, H.; Hennicke, F.; Thines, M. *BMC Genomics* **2018,** *19,* 48. doi:10.1186/s12864-017-4430-y
17. Wawrzyn, G. T.; Quin, M. B.; Choudhary, S.; López-Gallego, F.; Schmidt-Dannert, C. *Chem. Biol.* **2012,** *19,* 772–783. doi:10.1016/j.chembiol.2012.05.012
18. Kasahara, K.; Miyamoto, T.; Fujimoto, T.; Oguri, H.; Tokiwano, T.; Oikawa, H.; Ebizuka, Y.; Fujii, I. *ChemBioChem* **2010,** *11,* 1245–1252. doi:10.1002/cbic.201000173
19. Quin, M. B.; Michel, S. N.; Schmidt-Dannert, C. *ChemBioChem* **2015,** *16,* 2191–2199. doi:10.1002/cbic.201500308
20. de Sena Filho, J. G.; Quin, M. B.; Spakowicz, D. J.; Shaw, J. J.; Kucera, K.; Dunican, B.; Strobel, S. A.; Schmidt-Dannert, C. *Fungal Biol.* **2016,** *120,* 1050–1063. doi:10.1016/j.funbio.2016.04.001
21. Chomczynski, P.; Sacchi, N. *Anal. Biochem.* **1987,** *162,* 156–159. doi:10.1016/0003-2697(87)90021-2

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Mechanistic investigations on multiproduct β-himachalene synthase from *Cryptosporangium arvum*

Jan Rinkel and Jeroen S. Dickschat*

Full Research Paper

Open Access

Address:
Kekulé-Institute for Organic Chemistry and Biochemistry, University of Bonn, Gerhard-Domagk-Str. 1, 53121 Bonn, Germany

Email:
Jeroen S. Dickschat* - dickschat@uni-bonn.de

* Corresponding author

Keywords:
enzyme mechanisms; isotopes; mass spectrometry; promiscuity; terpenes

Beilstein J. Org. Chem. **2019,** *15,* 1008–1019.
doi:10.3762/bjoc.15.99

Received: 12 March 2019
Accepted: 26 April 2019
Published: 02 May 2019

This article is part of the thematic issue "Terpenes".

Associate Editor: D. Spring

Abstract

A bacterial terpene synthase from *Cryptosporangium arvum* was characterised as a multiproduct β-himachalene synthase. In vitro studies showed not only a high promiscuity with respect to its numerous sesquiterpene products, including the structurally demanding terpenes longicyclene, longifolene and α-longipinene, but also to its substrates, as additional activity was observed with geranyl- and geranylgeranyl diphosphate. In-depth mechanistic investigations using isotopically labelled precursors regarding the stereochemical course of both 1,11-cyclisation and 1,3-hydride shift furnished a detailed catalytic model suggesting the molecular basis of the observed low product selectivity. The enzyme's synthetic potential was also exploited in the preparation of sesquiterpene isotopomers, which provided insights into their EIMS fragmentation mechanisms.

Introduction

The organic chemist usually prefers to work with pure compounds which lead to high requirements for the selectivity of reactions and often to tedious purification procedures, but encountering a pure compound in nature is quite rare. This does not result in reduced requirements for enzyme selectivity. The very opposite is mostly true, because proteins working in a compound mixture need to be precise [1]. However, in some cases, compound mixtures have proven to be superior to the properties of the single compounds by evolution. Examples demonstrating this principle can be found in pheromone chemistry, like the bark beetle aggregation blend of ipsdienol, ipsenol and verbenol, for which synergistic effects were observed compared to the single compounds [2]. Also the sex pheromone of the cranberry white grub *Phyllophaga anxia* was identified as a compound mixture, consisting of L-valine methyl ester and L-isoleucine methyl ester at a 3:1 ratio [3]. Moreover, if there is a single enzyme that can produce a beneficial mixture, the advantage for the producing organism is even higher. Therefore, selectivity is not in every case the highest goal for evolution. An enzyme class, which is highly prone to a regulation of product selectivity for the production of either one or multiple compounds, are terpene synthases (TSs). These enzymes are able to

guide complex cascade reactions from structurally simple oligoprenyl diphosphates to often complex, polycyclic products [4-6] circumventing the low selectivity observed for carbocationic reactions by a defined active-site architecture. Although these enzymes are mostly highlighted for their great product selectivity, TSs producing only one compound are by far not the general case. Mostly, the main product is accompanied by several side products. Prominent examples are the TS identified from the plant *Medicago truncatula* with at least 27 products [7], γ-humulene synthase from *Abies grandis* with 52 products [8], and also the long known trichodiene synthase from *Fusarium sporotrichioides* produces at least 15 sesquiterpenes [9]. Some TSs can even accept multiple chain length substrates [10], a concept which seems to occur frequently in plants [11]. Whether the reduced selectivity of TSs both for substrates and for products can be attributed to imperfect catalysis, or if this function is even beneficial for the producing organism, remains elusive in most cases. Also the structural basis of promiscuous catalysis by TSs is largely unknown [12]. In this study, we present the characterisation of a bacterial TS with a reduced selectivity both for substrates and for products together with the challenging investigation of its cyclisation mechanism by labelling experiments.

Results and Discussion

A bacterial β-himachalene synthase produces numerous side products

Apart from the recently assigned (*Z*)-γ-bisabolene synthase (BbS) [13], the soil-dwelling actinomycete *Cryptosporangium arvum* DSM 44712 also possesses a second TS gene (accession no. WP_035852539). Its encoded amino acid sequence (Figure S1, Supporting Information File 1) shares conserved motifs for TSs, but is phylogenetically distant to BbS and does not possess a close characterised relative among other bacterial TSs (Figure S2, Supporting Information File 1). Therefore, its gene was cloned into the *E. coli* expression vector pYE-Express [14] for functional characterisation (Table S1, Supporting Information File 1). The purified recombinant protein (Figure S3, Supporting Information File 1) was incubated with the common TS substrates geranyl- (GPP, C_{10}), farnesyl- (FPP, C_{15}), geranylgeranyl- (GGPP, C_{20}) and geranylfarnesyl (GFPP, C_{25}) diphosphate. Whereas the latter diphosphate did not lead to any terpene product, the incubation with FPP showed a smooth conversion into several sesquiterpenes (Figure 1A) with compound **1** as the major peak after GC–MS analysis. However, also the incubations with GPP (Figure 1B) and GGPP (Figure 1C) led to several less complex terpene products, demonstrating a broadened substrate range for this enzyme. The annotated peaks were correlated by mass spectral libraries and retention indices (Table 1) to the known natural products **1**–**8** and **10**–**18** (Figure 2).

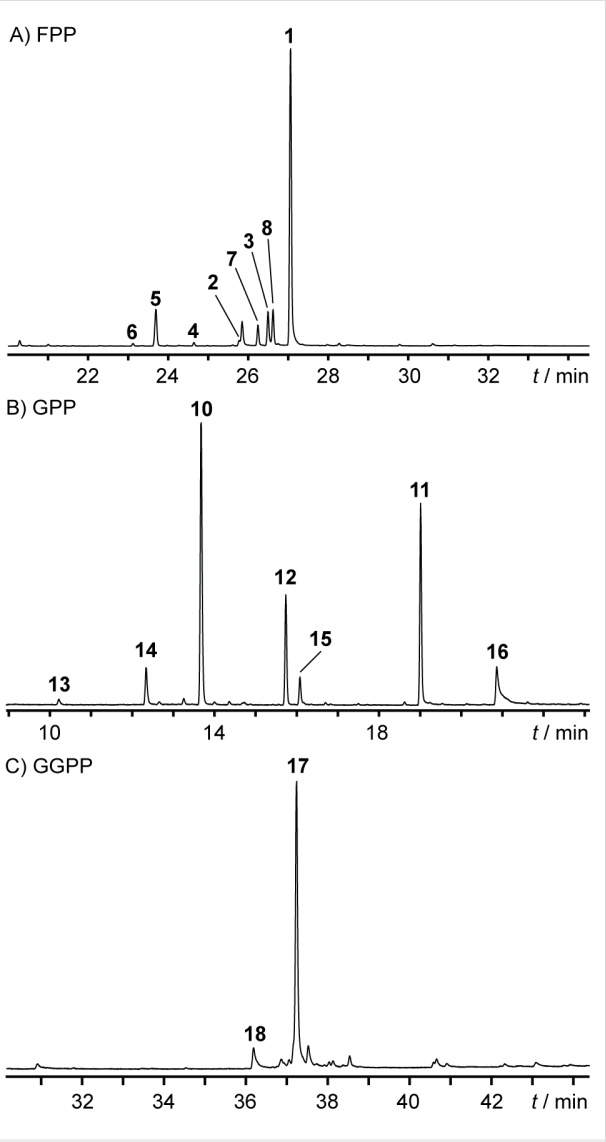


Figure 1: Total ion chromatograms of hexane extracts from the incubations of HcS with A) FPP, B) GPP and C) GGPP. Peak numbering refers to the compounds shown in Figure 2 and listed in Table 1.

In a large scale incubation, β-himachalene (**1**) was isolated, accompanied by smaller amounts of the double oxidation product γ-dehydro-*ar*-himachalene (**9**). Since **9** was only observed after prolonged incubation times, an auto-oxidation mechanism involving oxygen is assumed. Both compounds were analysed by one- and two-dimensional NMR spectroscopy (Tables S2 and S3, Supporting Information File 1). The absolute configuration of **1** was determined as the (+)-enantiomer, unanimously by optical rotary power measurement and an isotopic labelling strategy, which involved conversion of stereoselectively deuterated and at the same position ^{13}C-labelled FPPs by the TS to yield labelled **1** with incorporation of deuterium into diastereotopic hydrogen positions. Together with the relative configuration of the targeted methylene group as deduced by NOESY,

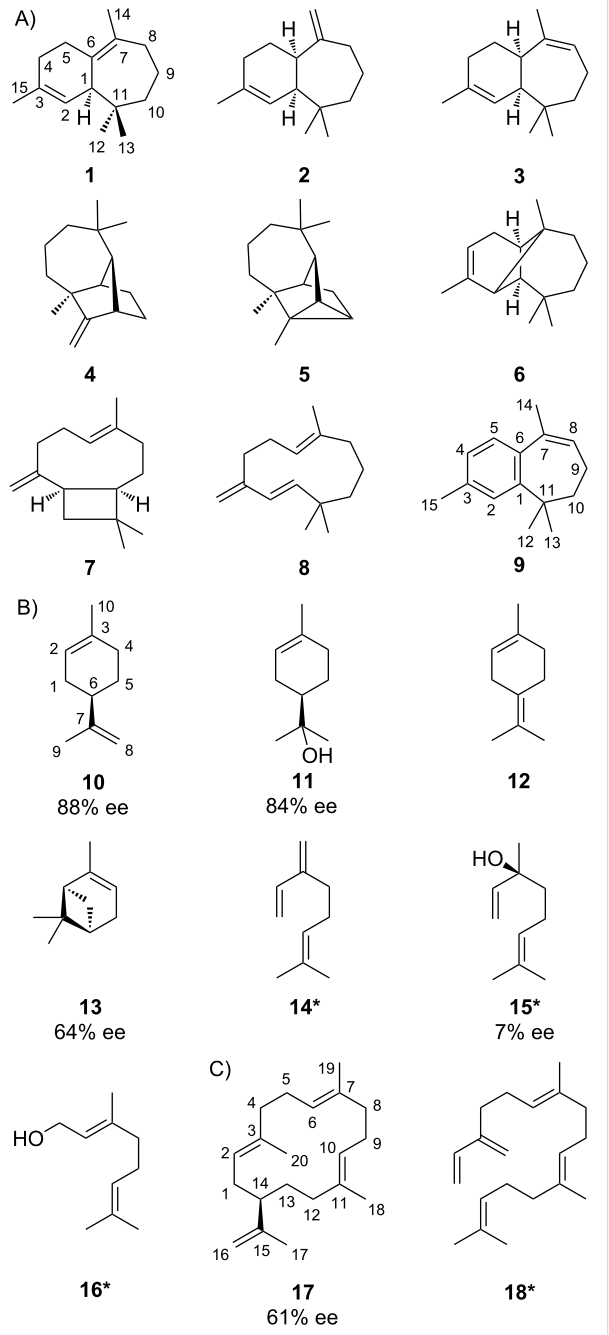


Figure 2: Structures of HcS products arising A) from FPP together with related oxidation product **9**, B) from GPP and C) from GGPP. The carbon numberings of **1** and **9** refer to the carbon positions of FPP as shown in Scheme 3, numberings of **10** and **17** are derived from that of GPP and GGPP, respectively. Compounds known to also originate from non-enzymatic hydrolysis are labelled with an asterisk. The enantiomeric excess values were determined based on GC analysis on a chiral phase.

Table 1: HcS product identification by retention indices.

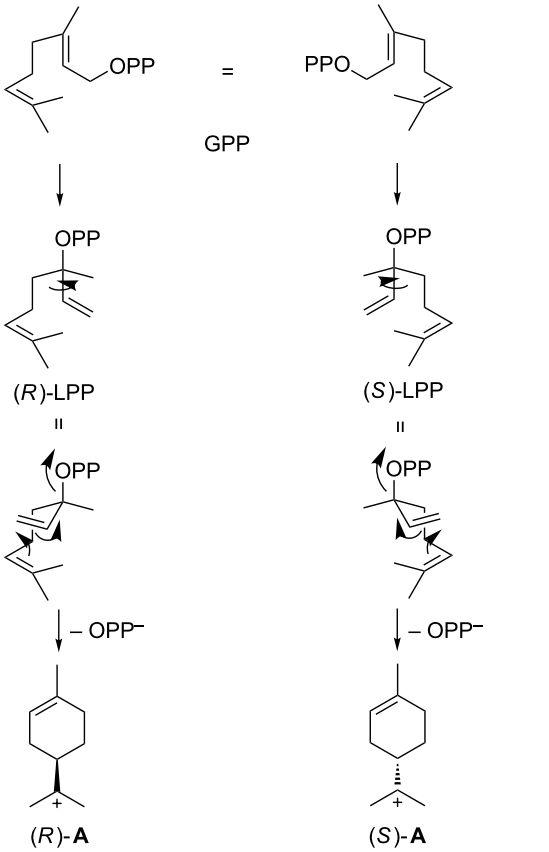
compound	I^a	I (lit.)
from GPP		
α-pinene (**13**)	934	934 [15]
β-myrcene (**14**)	992	992 [16]
limonene (**10**)	1029	1031 [17]
α-terpinolene (**12**)	1089	1088 [18]
linalool (**15**)	1099	1098 [19]
α-terpineol (**11**)	1191	1190 [19]
geraniol (**16**)	1250	1253 [20]
from FPP		
α-longipinene (**6**)	1355	1356 [21]
longicyclene (**5**)	1376	1377 [21]
longifolene (**4**)	1412	1413 [21]
α-himachalene (**2**)	1461	1461 [22]
9-*epi*-β-caryophyllene (**7**)	1475	1471 [23]
γ-humulene (**8**)	1489	1487 [24]
γ-himachalene (**3**)	1490	1489 [22]
β-himachalene (**1**)	1507	1503 [25]
from GGPP		
β-springene (**18**)	1921	1918 [26]
cembrene A (**17**)	1974	1979 [27]

[a]Retention index I on a HP-5MS column.

the stereochemical outcome of these experiments, which can be easily monitored by sensitive HSQC, infers the absolute configuration of **1**. For **1**, C-5 was targeted by (1*R*)- and (1*S*)-(1-^{13}C,1-^{2}H)GPP [28], which were enzymatically elongated with isopentenyl diphosphate (IPP) by farnesyl diphosphate synthase (FPPS) from *Streptomyces coelicolor* [29] with a known stereochemical course [30] (Figure S4, Supporting Information File 1). This principle was also applied to C-4 and C-8 utilising (*Z*)- and (*E*)-(4-^{13}C,4-^{2}H)IPP [31] for the elongation of dimethylallyl diphosphate (DMAPP) catalysed by FPPS in a known course [32]. Since both hydrogens at C-4 possess the same chemical shift, only C-8 could be used to solidify the absolute configuration (Figure S5, Supporting Information File 1).

GC analysis on a homochiral stationary phase was used to assign the absolute configurations of the observed chiral monoterpenes (*R*)-(+)-limonene (**10**), (*R*)-(+)-α-terpineol (**11**), (+)-α-pinene (**13**) and (*S*)-(+)-linalool (**15**) as shown in Figure 2 by comparison with commercially available standards (Figures S6–S9, Supporting Information File 1). The non-enzymatic degradation of GPP as a background reaction to **15** resulted in a substantial loss of stereoinformation for this compound (7% ee). Also the cyclised products **10**, **11** and **13** where not obtained in enantiomerically pure form (ee values were varying between 64% and 88%, as judged by integration), which may point to different possible binding and folding modes within the TS's

active site for GPP involving both enantiomers of linalyl diphosphate (LPP, Scheme 1) and the terpinyl cation (**A**). Other TSs producing an enantiomeric mixture of monoterpenes are also known, e g., from *Pinus taeda* [33]. However, the major enantiomer of each cyclised monoterpene product described herein was found to be derived from (*R*)-**A**.



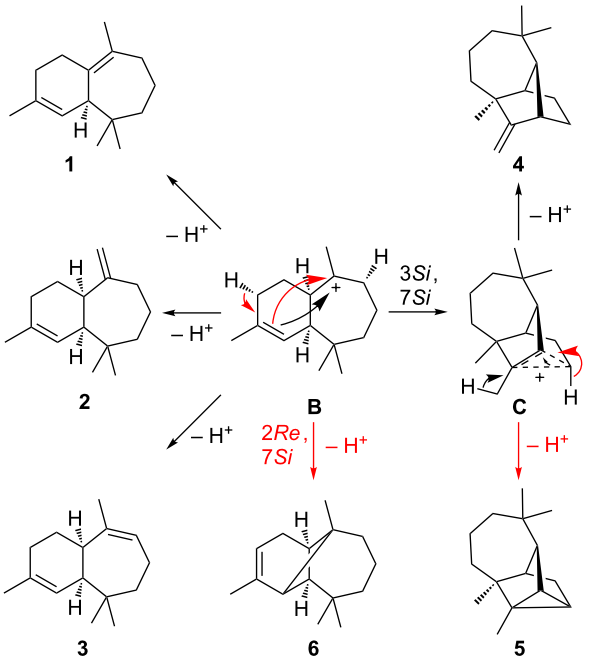
Scheme 1: Initial steps of the cyclisation of GPP towards monoterpene products [34]. Both pathways are likely co-occurring in the TS to explain the formation of both enantiomers of **10**, **11** and **13**, with the major enantiomer in all cases being derived from (*R*)-**A**.

Compound **17** was isolated from a large scale incubation of the TS with GGPP and identified by NMR as cembrene A. Chiral phase GC analysis showed also in this case a mixture of enantiomers with the major one being (–)-cembrene A (61% ee), the enantiomer of the product obtained from a cembrene A synthase (CAS) from *Allokutzneria albata* [27], which was used for comparison (Figure S10, Supporting Information File 1).

Taken together, the overall more sluggish conversion of GPP and GGPP by the TS leading to enantiomeric mixtures, the higher biosynthetic complexity of the obtained sesquiterpenes and the absence of spontaneous hydrolysis products in the incubation with FPP compared to the appearance of **14** and **15** in the incubation with GPP and **18** in the experiment with GGPP, this TS from *C. arvum* is characterised as a multiproduct (+)-β-himachalene synthase (HcS) possessing additional mono- and diterpene cyclase activity.

The structures of its minor products reveal the cyclisation mechanism of HcS

Since **17** is a simple 1,14-cyclisation product, and all cyclised monoterpenes are derived from the extensively studied terpinyl cation [35,36], this work focusses on elucidating the more interesting sesquiterpene cyclase mechanism of HcS. Most sesquiterpene products **1**–**6** of HcS including the main product **1**, but also α-himachalene (**2**), γ-himachalene (**3**), longifolene (**4**), longicyclene (**5**), and α-longipinene (**6**) are traditionally considered to be derived from the himachalyl cation (**B**) [8,37] (Scheme 2).



Scheme 2: Late stage cyclisations of the himachalyl cation **B** to HcS products **1**–**6**. Alternative mechanistic and reaction arrows belonging to branching points are shown in red.

Whereas **1**–**3** are simple deprotonation products of **B**, **4** and **5** require a further 3,7-ring closure, leading to the non-classical cation **C**, which is a derivative of the 2-norbornyl cation [38]. This system either collapses by deprotonation at the methyl group to longifolene (**4**), or by deprotonation at C-4 with formation of a cyclopropane ring to longicyclene (**5**). Starting from **B**, a 2,7-ring closure and deprotonation at the same carbon atom gives α-longipinene (**6**). For the main product **1**, the deprotonation was followed by an incubation of HcS and FPPS with

(2-^{2}H)GPP [39] and IPP, which resulted in unlabelled **1** as observed by GC–MS (Figure 3).

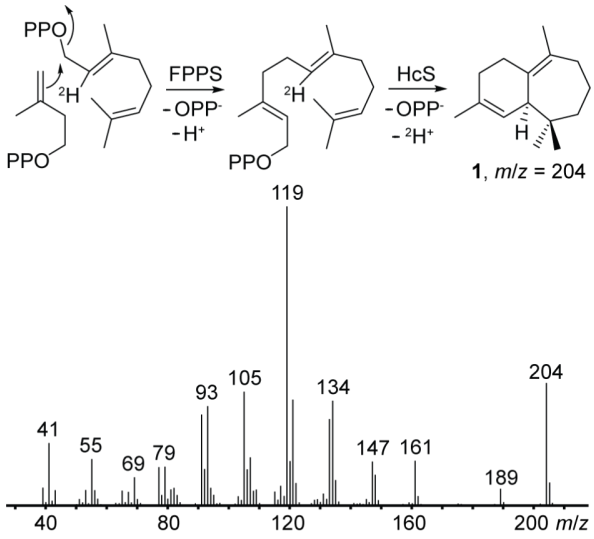
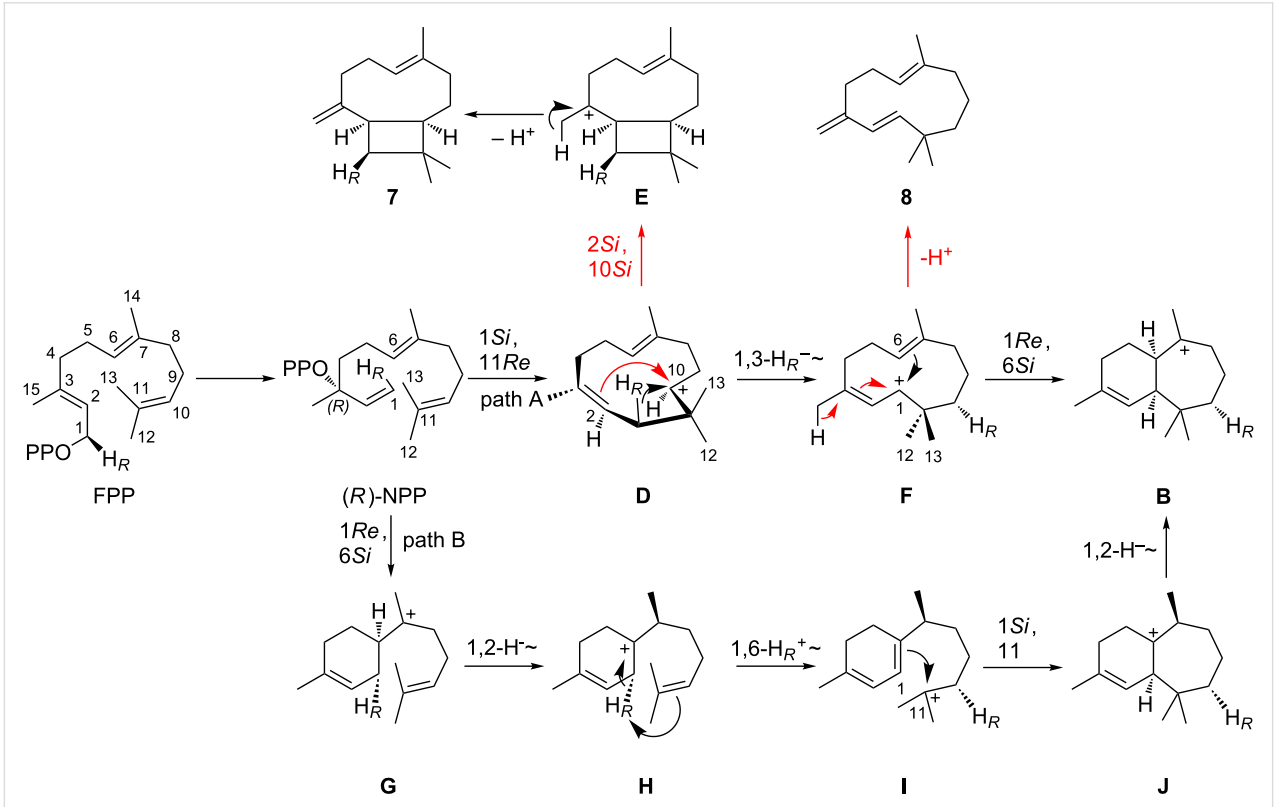


Figure 3: EI mass spectrum of **1** arising from an incubation of (2-^{2}H)GPP and IPP with FPPS and HcS showing a loss of label during deprotonation.

In case of a deprotonation at a methylene group, relevant for the formation of compounds **3**, **5** and **6**, the stereochemical course of these final steps could be followed by stereoselective deuterations. GC–MS analysis of the products obtained from the incubations with HcS, FPPS, DMAPP and (*Z*)- or (*E*)-(4-^{13}C,4-^{2}H)IPP showed a specific loss of H_Z in all cases (Figure 4).

Intriguingly, all deprotonation steps leading to **1**, **3**, **5** and **6** proceed from the same face of **B**. Giving access to most products, cation **B** can be considered as the central branching point within the HcS catalysed cyclisation mechanism. To rationalise the formation of **B** starting from FPP, two different pathways were initially assumed (Scheme 3). Both start with a 1,3-*syn*-allylic rearrangement of OPP to (*R*)-nerolidyl diphosphate (NPP). This step is usually proposed to generate a (*Z*)-configured C-2,C-3 double bond after cyclisation [40]. Following the first mechanism (path A), a 1,11-cyclisation can yield secondary cation **D**, which either stabilises by 2,10-ring closure to give the caryophyllenyl cation **E** that can be deprotonated at the methyl group to yield 9-*epi*-(*E*)-β-caryophyllene (**7**), or **D** undergoes a 1,3-hydride shift to the allylic cation **F**. Deprotonation leads to γ-humulene (**8**), but a 1,6-ring closure gives access to **B**.



Scheme 3: Proposed cyclisation mechanism towards cation **B** via an initial 1,11-cyclisation (path A) and an hypothetical alternative mechanism via an initial 1,6-cyclisation (path B). Alternative mechanistic and reaction arrows belonging to branching points are shown in red.

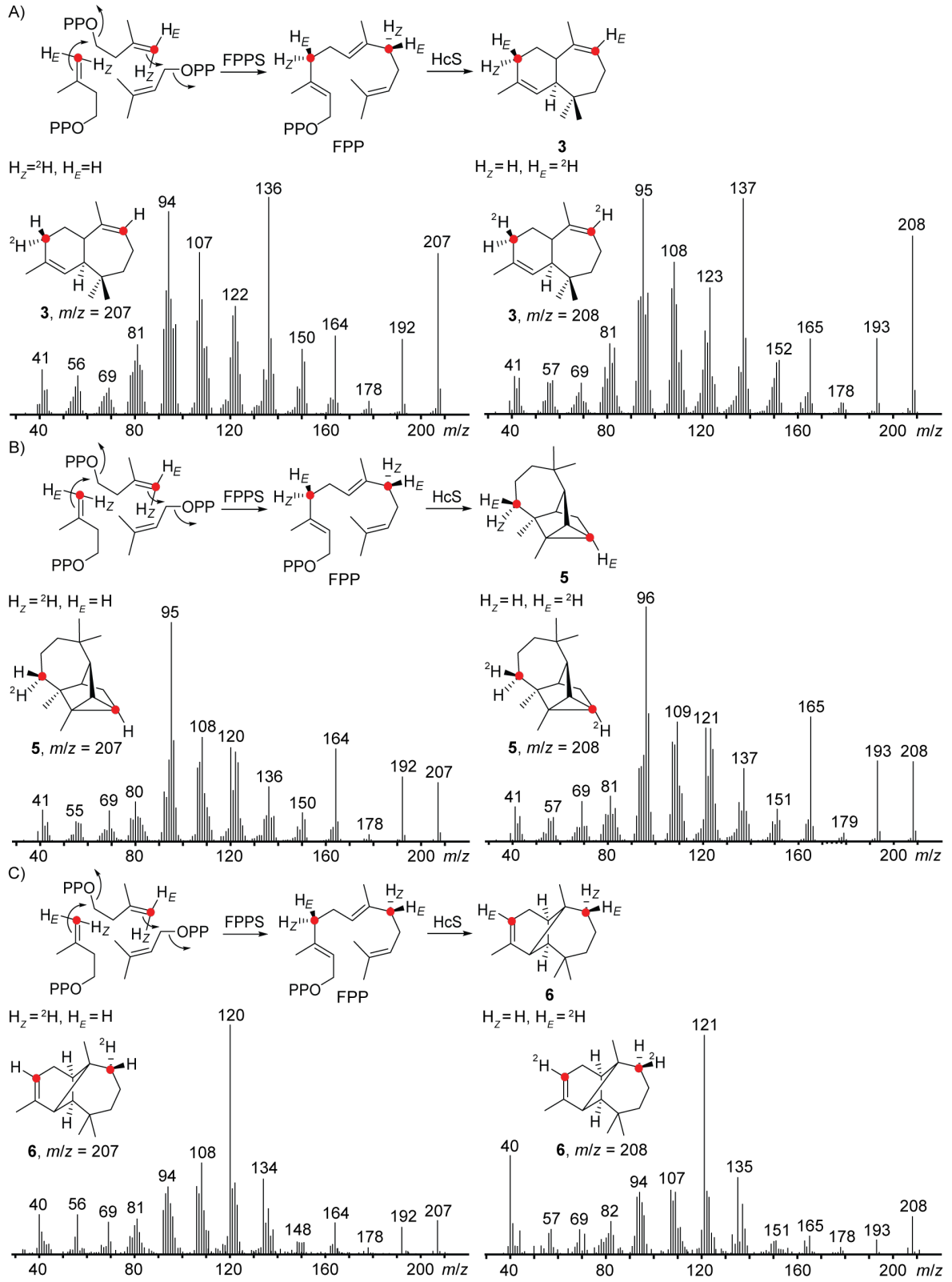


Figure 4: Stereochemical course of the final deprotonation step towards **3**, **5** and **6** investigated by GC–MS. EI mass spectra of labelled A) **3**, B) **5** and C) **6** obtained from the incubation of HcS and FPPS with DMAPP and (*Z*)-(4-^{13}C,4-^{2}H)IPP (left) or (*E*)-(4-^{13}C,4-^{2}H)IPP (right) indicating stereospecific loss of one hydrogen atom. ^{13}C-Labellings are indicated by red dots.

The second shown option, path B, assumes a 1,6-ring closure of (*R*)-NPP to the bisabolyl cation **G**. Proceeding with a 1,2-hydride shift to **H**, the key step is a 1,6-proton shift to give the tertiary cation **I**. This idea is derived from a very similar proton transfer starting from the bisabolyl cation, which occurs in the cyclisation mechanism to trichodiene [41]. A 1,11-cyclisation yields tertiary cation **J**, which can undergo a 1,2-hydride shift to **B**. While path B circumvents the secondary cation intermediate **D**, the HcS products **7** and **8** are hard to explain from path B. Together with the absence of any 1,6-cyclised bisabolene derived molecules in the product mixture their occurrence represent experimental evidence in favour of path A.

Incubation experiments enlighten the stereochemical course of the 1,11-cyclisation and the 1,3-hydride shift

Although the general idea of path A appears to be straightforward at first sight, the details proofed to be challenging as deeper insights for the stereochemistry of each step were obtained by incubation experiments. The question, whether (*R*)-NPP or its enantiomer (*S*)-NPP is involved in the HcS catalysed cyclisation cascade, was addressed by incubation of both enantiomerically pure substrates and (*rac*)-NPP [13] with HcS. The resulting chromatograms (Figure 5) clearly demonstrate (*R*)-NPP as an intermediate, which is a substrate for the production of **1**–**8** in approximately the same ratio as with FPP. In contrast, (*S*)-NPP is selectively converted into (*E*)-β-farnesene (**19**, I = 1460 (HP-5MS), Lit: I = 1459 (HP-5MS) [42]). The same outcome regarding the formation of (*Z*)-γ-bisabolene from (*R*)-NPP and FPP, but of **19** from (*S*)-NPP was recently also observed for BbS [13].

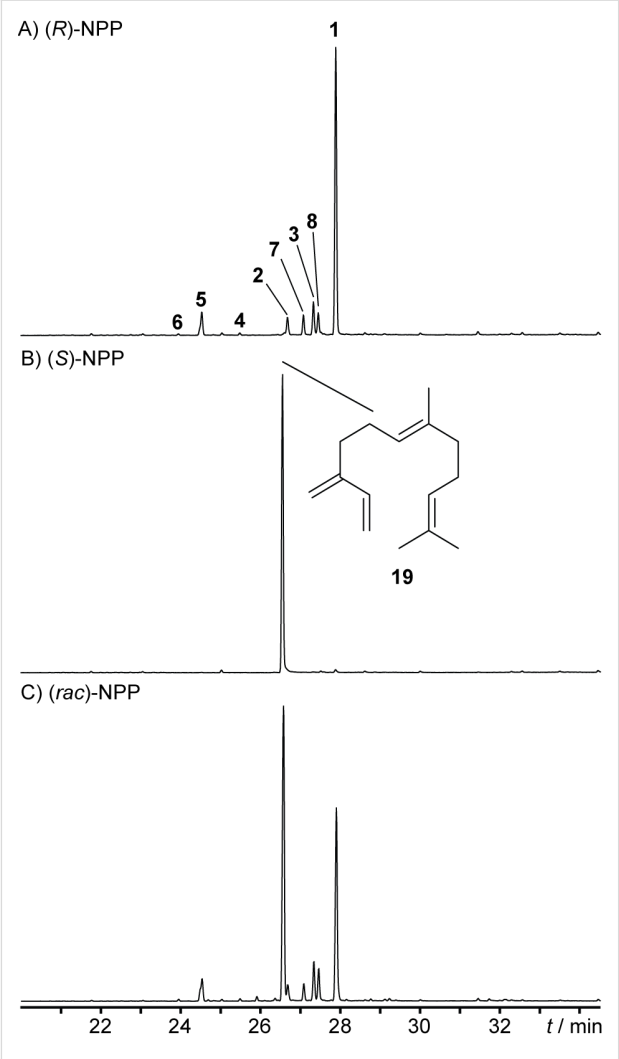


Figure 5: Total ion chromatogram of hexane extracts from HcS incubations with A) (*R*)-NPP, B) (*S*)-NPP and C) (*rac*)-NPP.

Targeting the stereochemical course of the 1,11-cyclisation of (*R*)-NPP to cation **D**, (12-^{13}C)- [43] and enzymatically prepared (13-^{13}C)FPP from (9-^{13}C)GPP [39] and IPP with FPPS were incubated with HcS to follow the fate of the geminal methyl groups for **1** (Figure 6). Combined with the relative orientation of each methyl group deduced by NOESY, these experiments showed an 11*Re* attack preceding the formation of **D**. The observed absolute configurations of the monoterpenes **10**, **11** and **13** and of the diterpene **17** support this finding, because their formation requires involvement of the same face of the terminal isoprenoid double bond (6*Si* from GPP and 14*Si* from GGPP). Therefore, a similar binding conformation for the terminal C_5-unit is reasonable for the three substrates.

To complete the mechanistic picture of the initial 1,11-cyclisation, also the stereochemical course at C-1 was investigated. Unfortunately, this position is disturbed by the follow-up 1,3-hydride shift in **1** and most products. However, in the side product **7** C-1 remains untouched after 1,11-cyclisation, which allows to investigate the stereochemical course of the first cyclisation step for this compound. First, the absolute configuration of **7** was assigned as shown in Figure 2 from the incubation experiments with (*E*)- and (*Z*)-(4-^{13}C,4-^{2}H)IPP, DMAPP, FPPS and HcS targeting the positions C-3 and C-7 (Figure S11, Supporting Information File 1), using published NMR data for **7** [44]. The stereochemical fate for the hydrogens at C-1 was then targeted by the incubation of (1*R*)- and (1*S*)-(1-^{13}C,1-^{2}H)FPP [28] with HcS (Figure 7).

The selective incorporation of deuterium into the diastereotopic positions of **7** is explainable by a 1*Si*,11*Re*-cyclisation of (*R*)-NPP. Given the absolute configuration of NPP and its formation via a 1,3-*syn*-allylic rearrangement from FPP, this ring closure represents an example of a formal *syn*-S_N2' reaction. This is an intriguing observation, since for other TSs a NPP-cyclisation by *anti*-S_N2' is usually described [40,45-47]. This

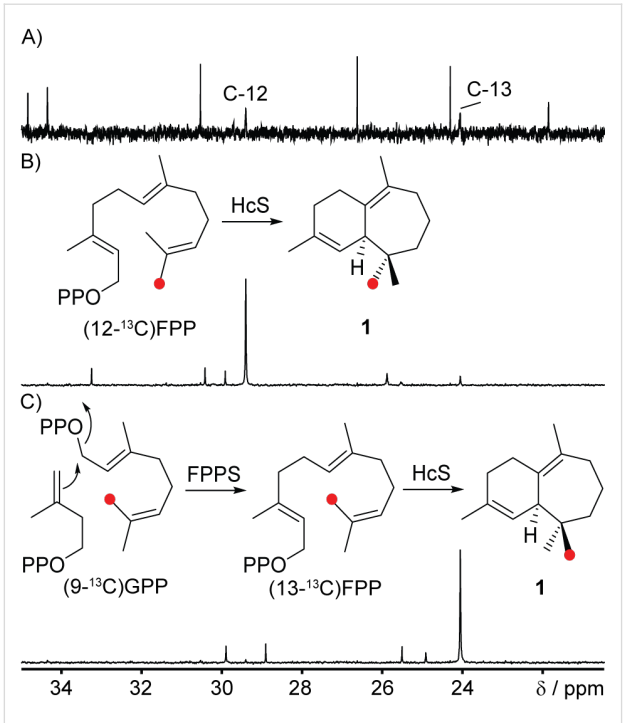


Figure 6: The origin of the two diastereotopic methyl groups in **1**. Partial ^{13}C NMR spectrum of A) unlabelled **1**, B) an extract obtained from the incubation of HcS with (12-^{13}C)FPP and C) an extract obtained from the incubation of HcS with (9-^{13}C)GPP, which is enzymatically converted to (13-^{13}C)FPP. Red dots indicate ^{13}C labelled carbon atoms.

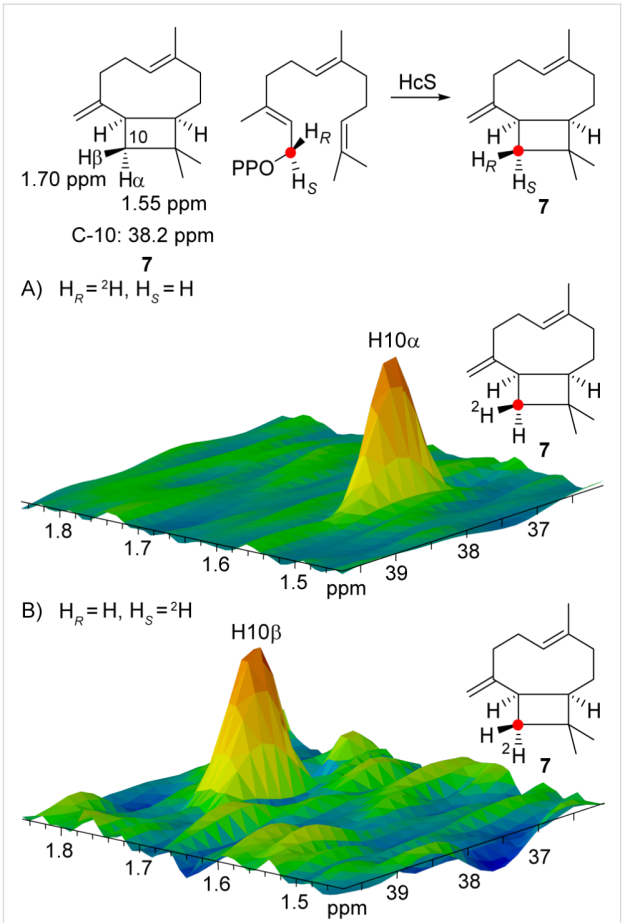


Figure 7: Stereochemical course of the 1,11-cyclisation at C-1 for **7**. Partial HSQC spectra of HcS incubation with A) (1*R*)-(1-^{13}C,1-^{2}H)FPP and B) (1*S*)-(1-^{13}C,1-^{2}H)FPP. The reference chemical shifts for **7** are taken from ref. [44]. Red dots represent ^{13}C-labelled carbon atoms.

cyclisation mechanism is thought to be the predominant case, giving rise to a more energetically favoured transition state, but occasionally also the *syn*-stereochemistry was observed [48]. The rather unexpected stereochemical course of the HcS-catalysed cyclisation of NPP found herein therefore shows, that this step has to be investigated for *anti*- versus *syn*-attack experimentally for every single case, especially for a conformationally flexible situation like a 1,11-cyclisation. Intriguingly, the stereochemical course of the initial cyclisation step can even be substrate dependent. The 1,6-cyclisation towards the monoterpenes **10**, **11** and **13** as investigated by the incubation of (1*S*)- and (1*R*)-(1-^{13}C,1-^{2}H)GPP with HcS and comparison to the NMR data of the commercial available products (Table S4–S8, Supporting Information File 1) clearly obeys the *anti*-S_N2' case (Figures S12–S14, Supporting Information File 1). The observation that **15** was obtained as a nearly racemic mixture contrasts the far more selective incorporation of deuterium into the olefinic positions at C-1 of **15** (Figure S15, Supporting Information File 1). This result supports (*R*)-LPP as an intermediate, formed by a 1,3-*syn*-allylic rearrangement to determine the observed stereochemical course at C-1, while the tertiary diphosphate might then undergo a non-enzymatic degradation to explain the high loss of stereoinformation in **15**. Also for the achiral β-myrcene (**14**), an imbalanced incorporation of deuterium is found at C-1 (Figure S16, Supporting Information File 1). With the opposite stereochemical course than for **15**, **14** is likely derived from the minor enantiomer (*S*)-LPP in analogy to **19** observed from (*S*)-NPP. For the diterpene **17** (Table S9, Supporting Information File 1), similar investigations using (1*S*)- and (1*R*)-(1-^{13}C,1-^{2}H)GGPP [49] with HcS resulted in the expected outcome for a direct 1,14-cyclisation of GGPP (Figure S17, Supporting Information File 1) in line with the results obtained with CAS from *A. albata* for *ent*-**17** [27]. Assuming similar chemical shifts at C-1 for **14** and **18**, the analogous signals for C-1 of **18** gave comparable results with the same stereochemical course as observed for **14**, although with lower preservation of stereoinformation (Figure S18, Supporting Information File 1).

To shed light on the stereochemical course of the 1,3-hydride shift connecting cations **D** and **F**, a series of labelling experiments were conducted to determine the origin of the shifting hydrogen (C-1) and its destination (C-10) for **1** (Figure 8). A

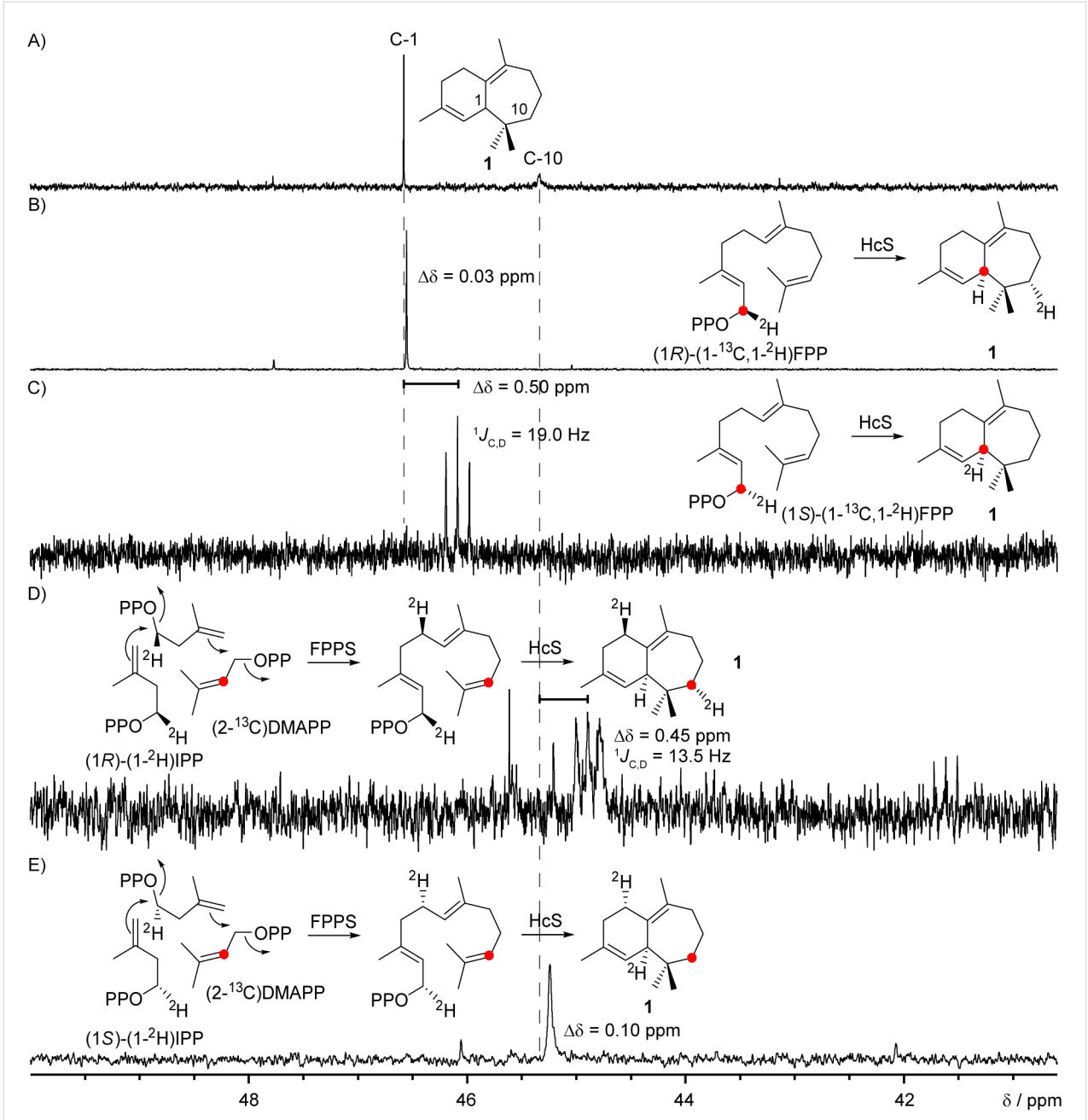


Figure 8: Investigation of the 1,3-hydride shift in the cyclisation towards **1**. Partial ^{13}C NMR spectra of A) unlabelled **1**, and incubations with HcS and B) (1*R*)-(1-^{13}C,1-^{2}H)FPP or C) (1*S*)-(1-^{13}C,1-^{2}H)FPP compared to incubations of HcS with FPPS, (2-^{13}C)DMAPP and D) (1*R*)-(1-^{2}H)IPP or E) (1*S*)-(1-^{2}H)IPP showing a movement of H_R from C-1 to C-10. The singlets for B) and E) are slightly shifted to higher field (designated as Δδ) because of the deuterium location nearby. ^{13}C-Labelled carbon atoms are indicated by red dots. Dashed grey lines show the chemical shifts of the carbon atoms for unlabelled **1**.

comparison of the ^{13}C NMR spectra from the incubations of HcS with (1*R*)- and (1*S*)-(1-^{13}C,1-^{2}H)FPP, resulting in a singlet for the (*R*)-case and a triplet in the (*S*)-case indicating a direct C–D bond, clearly demonstrated the stereospecific migration of H_R from C-1. To complete the observations also for C-10, (2-^{13}C)DMAPP was synthesised from (2-^{13}C)-3-methylbut-2-en-1-ol [43] and incubated with (1*R*)- or (1*S*)-(1-^{2}H)IPP [50], FPPS and HcS resulting in the expected opposite outcome than stated above, namely a triplet in the (*R*)-case and a singlet for the (*S*)-sample.

HSQC analysis of the material obtained from the incubation of (1*R*)-(1-^{2}H)IPP and (2-^{13}C)DMAPP with FPPS and HcS also allowed for the assignment of the newly introduced diastereo-

topic position at C-10 (Figure 9). Together with the assignment of the hydrogens by NOESY in **1**, these data show a stereoselective incorporation of H_R-1 into the H_α-position at C-10 by a vanished crosspeak.

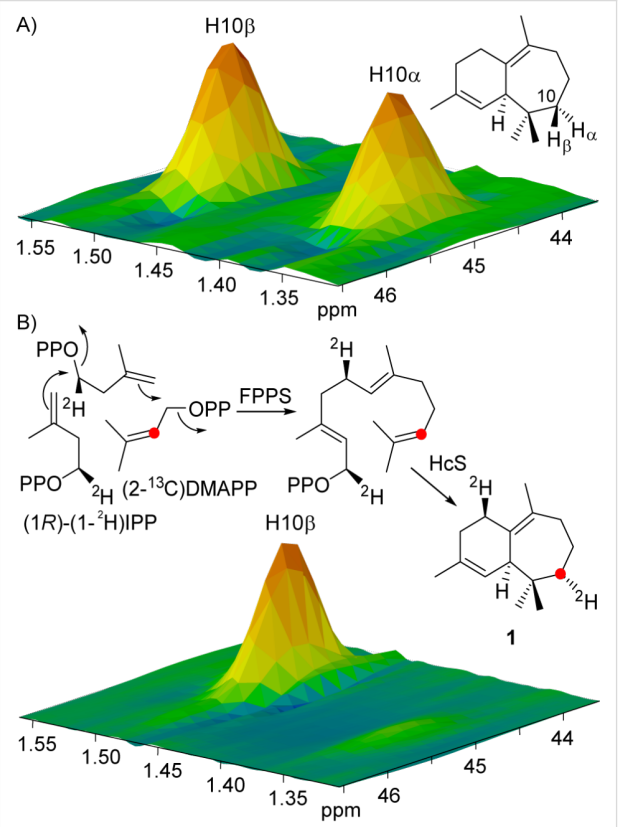


Figure 9: Stereochemical course of the 1,3-hydride shift at C-10 in **1**. Partial HSQC spectra of A) unlabelled **1** and B) labelled **1** arising from the incubation of (1*R*)-(1-^{2}H)IPP and (2-^{13}C)DMAPP with FPPS and HcS. Red dots indicate ^{13}C-labelled carbon atoms.

Combining the information deduced from the extensive incubation experiments stated above, a structural model for the reactive conformation of cation **D** is proposed (Figure S19, Supporting Information File 1). This intermediate, or structurally related transition states for the corresponding concerted reactions to avoid its secondary nature, are of central importance in understanding the initial HcS catalysed cyclisation towards cation **B**. The discussed conformation is imprinted by the structure of **7** with its relative conformation at the four-membered ring system allowing for a 2*Si*,10*Si*-cyclisation to **E** without major rotational changes and also reflects the short distance between H_R and C-10 for the 1,3-hydride shift towards the 10*Si* face leading to **F**. Intriguingly, the unusual *syn*-S_N2' ring closure from (*R*)-NPP leads to the diphosphate moiety (OPP^-) being located close to the "backside" of the cyclising molecule, which may give rise to an explanation of the multiproduct nature of HcS. At this location, OPP^- can easily abstract "backwards" pointing hydrogen atoms from different positions which reflects the observation of the regio- and stereochemistry of the deprotonations.

HcS provides access to labelled sesquiterpenes for EIMS fragmentation studies

Since HcS produces a mixture of structurally interesting sesquiterpenes, its synthetic abilities were also exploited to study EIMS fragmentation mechanisms. Therefore, all fifteen singly-^{13}C labelled FPP isotopomers, either obtained by synthesis or enzymatically [39,43,51], were converted with HcS to result in mixtures of specifically labelled **1**–**8**. The incorporation of label into **1** was checked by ^{13}C NMR (Figure S20, Supporting Information File 1) and all samples were analysed by GC–MS. This allowed for the assignment of carbon positions to specific EI-fragments of the corresponding mass spectrum by observing an increase of +1 Da, if the labelled position is part of the fragment (position specific mass shift analysis, PMA [41,52,53]). Although for many fragments multiple overlaying fragmentation pathways were observed, some of them showed clear position dependent results, which are summarised in Figure 10. The EI mass spectra for each position and molecule laying the basis for the presented three fragments for **1** together with one fragment each for **4**–**8** are depicted in Figures S21–S26 (Supporting Information File 1). Possible EI-fragmentation mechanisms connected to them are discussed in Schemes S1–S3 (Supporting Information File 1).

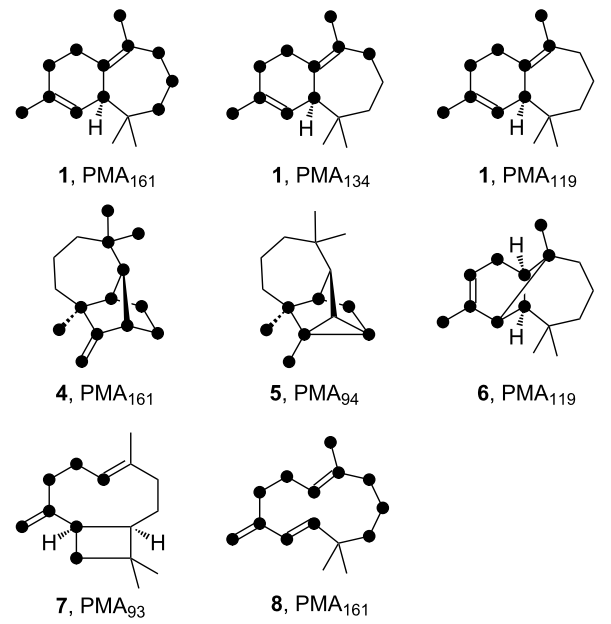


Figure 10: Position specific mass shift analysis for selected EIMS ions of HcS products. Black dots represent an increased mass of the ion (*m*/*z* = +1) in case of a ^{13}C-labelling in this position. Proposed fragmentation mechanisms for these ions are presented in Schemes S1–S3 (Supporting Information File 1).

Conclusion

In summary, a new terpene synthase from *C. arvum* was characterised as a multiproduct (+)-β-himachalene synthase. Accepting GPP, FPP and GGPP, HcS is a promiscuous enzyme, whose catalysis suffers from poor selectivity. Nevertheless, the formation of multiple sesquiterpene products demands for a challenging mechanistic model, which was refined by extensive labelling experiments. Several interesting details were disclosed including the stereochemical course of a 1,3-hydride migration from C-1 to C-10 and the 1,11-cyclisation featuring the unusual *syn*-S_N2' attack. Combining various aspects of the initial cyclisation, the proposed conformer of cation **D** may also rationalise the reduced selectivity of HcS by its positioning of OPP^-. Providing access to labelled isotopomers of its products, including structurally demanding polycyclic terpenes, HcS also served as a platform for investigating selected aspects of their EIMS fragmentation mechanisms. The labelling experiments performed with HcS described in this study therefore represent an encouragement to experimentally explore and elucidate every stereochemical detail of a terpene cyclisation mechanism for a comprehensive picture of the complex reactions, these amazing enzymes are able to catalyse.

Supporting Information

Experimental details for gene cloning, gene expression, protein purification, incubation experiments with isotopically labelled precursors, preparative scale incubation and synthesis of (2-^{13}C)DMAPP. The amino acid sequence of HcS, a phylogenetic tree of bacterial terpene synthases, SDS-PAGE analysis of the recombinant protein, listed NMR data for **1** and **9**, labelling experiments for the determination of the absolute configurations of **1** and **7**, chiral phase GC analysis of **10**, **11**, **13**, **15** and **17**, labelling experiments for the stereochemical course at C-1 of the monoterpenes and diterpenes, a graphical model for cation **D**, NMR spectra for the incubations of singly labelled FPPs with HcS, EIMS data for compounds **1** and **4**-**8** arising from these incubations and discussion of fragmentation mechanisms for selected ions.

Supporting Information File 1

Additional material.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-99-S1.pdf]

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DI1536/7-1).

ORCID® iDs

Jeroen S. Dickschat - https://orcid.org/0000-0002-0102-0631

References

1. Hedstrom, L. *Enzyme Specificity and Selectivity; eLS*; John Wiley & Sons, Ltd.: Chichester, UK, 2010. doi:10.1002/9780470015902.a0000716.pub2
2. Silverstein, R. M.; Rodin, J. O.; Wood, D. L. *Science* **1966,** *154,* 509–510.
3. Zhang, A.; Robbins, P. S.; Leal, W. S.; Linn, C. E., Jr.; Villani, M. G.; Roelofs, W. L. *J. Chem. Ecol.* **1997,** *23,* 231–245. doi:10.1023/b:joec.0000006356.47959.ed
4. Christianson, D. W. *Chem. Rev.* **2017,** *117,* 11570–11648. doi:10.1021/acs.chemrev.7b00287
5. Dickschat, J. S. *Nat. Prod. Rep.* **2016,** *33,* 87–110. doi:10.1039/c5np00102a
6. Tantillo, D. J. *Angew. Chem., Int. Ed.* **2017,** *56,* 10040–10045. doi:10.1002/anie.201702363
7. Garms, S.; Köllner, T. G.; Boland, W. *J. Org. Chem.* **2010,** *75,* 5590–5600. doi:10.1021/jo100917c
8. Steele, C. L.; Crock, J.; Bohlmann, J.; Croteau, R. *J. Biol. Chem.* **1998,** *273,* 2078–2089. doi:10.1074/jbc.273.4.2078
9. Vedula, L. S.; Jiang, J.; Zakharian, T.; Cane, D. E.; Christianson, D. W. *Arch. Biochem. Biophys.* **2008,** *469,* 184–194. doi:10.1016/j.abb.2007.10.015
10. Rinkel, J.; Lauterbach, L.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2017,** *56,* 16385–16389. doi:10.1002/anie.201711142
11. Pazouki, L.; Niinemets, Ü. *Front. Plant Sci.* **2016,** *7,* No. 1019. doi:10.3389/fpls.2016.01019
12. Dixit, M.; Weitman, M.; Gao, J.; Major, D. T. *ACS Catal.* **2017,** *7,* 812–818. doi:10.1021/acscatal.6b02584
13. Rinkel, J.; Dickschat, J. S. *Beilstein J. Org. Chem.* **2019,** *15,* 789–794. doi:10.3762/bjoc.15.75
14. Dickschat, J. S.; Pahirulzaman, K. A. K.; Rabe, P.; Klapschinski, T. A. *ChemBioChem* **2014,** *15,* 810–814. doi:10.1002/cbic.201300763
15. Saroglou, V.; Dorizas, N.; Kypriotakis, Z.; Skaltsa, H. D. *J. Chromatogr. A* **2006,** *1104,* 313–322. doi:10.1016/j.chroma.2005.11.087
16. Hazzit, M.; Baaliouamer, A.; Faleiro, M. L.; Miguel, M. G. *J. Agric. Food Chem.* **2006,** *54,* 6314–6321. doi:10.1021/jf0606104
17. Karioti, A.; Hadjipavlou-Litina, D.; Mensah, M. L. K.; Fleischer, T. C.; Skaltsa, H. *J. Agric. Food Chem.* **2004,** *52,* 8094–8098. doi:10.1021/jf040150j
18. Yáñez, X.; Pinzón, M. L.; Solano, F.; Sánchez, L. R. *Molecules* **2002,** *7,* 712–716. doi:10.3390/70900712
19. Benkaci–Ali, F.; Baaliouamer, A.; Meklati, B. Y.; Chemat, F. *Flavour Fragrance J.* **2007,** *22,* 148–153. doi:10.1002/ffj.1773
20. Yu, Y.; Huang, T.; Yang, B.; Liu, X.; Duan, G. *J. Pharm. Biomed. Anal.* **2007,** *43,* 24–31. doi:10.1016/j.jpba.2006.06.037
21. Zeng, Y.-X.; Zhao, C.-X.; Liang, Y.-Z.; Yang, H.; Fang, H.-Z.; Yi, L.-Z.; Zeng, Z.-D. *Anal. Chim. Acta* **2007,** *595,* 328–339. doi:10.1016/j.aca.2006.12.022
22. Andriamaharavo, N. R., Rentention Data. *NIST Mass Spectrometry Data Center,* 2014.
23. Mesa-Arango, A. C.; Betancur-Galvis, L.; Montiel, J.; Bueno, J. G.; Baena, A.; Durán, D. C.; Martínez, J. R.; Stashenko, E. E. *J. Essent. Oil Res.* **2010,** *22,* 568–574. doi:10.1080/10412905.2010.9700402

24. Bendiabdellah, A.; El Amine Dib, M.; Djabou, N.; Allali, H.; Tabti, B.; Muselli, A.; Costa, J. *Chem. Cent. J.* **2012,** *6,* No. 431. doi:10.1186/1752-153x-6-48
25. Quijano, C. E.; Salamanca, G.; Pino, J. A. *Flavour Fragrance J.* **2007,** *22,* 401–406. doi:10.1002/ffj.1812
26. Heinrich, G.; Pfeifhofer, H. W.; Stabentheiner, E.; Sawidis, T. *Ann. Bot. (Oxford, U. K.)* **2002,** *89,* 459–469. doi:10.1093/aob/mcf062
27. Rinkel, J.; Lauterbach, L.; Rabe, P.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2018,** *57,* 3238–3241. doi:10.1002/anie.201800385
28. Rabe, P.; Rinkel, J.; Dolja, E.; Schmitz, T.; Nubbemeyer, B.; Luu, T. H.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2017,** *56,* 2776–2779. doi:10.1002/anie.201612439
29. Rabe, P.; Rinkel, J.; Nubbemeyer, B.; Köllner, T. G.; Chen, F.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2016,** *55,* 15420–15423. doi:10.1002/anie.201608971
30. Thulasiram, H. V.; Poulter, C. D. *J. Am. Chem. Soc.* **2006,** *128,* 15819–15823. doi:10.1021/ja065573b
31. Lauterbach, L.; Rinkel, J.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2018,** *57,* 8280–8283. doi:10.1002/anie.201803800
32. Cornforth, J. W.; Cornforth, R. H.; Popják, G.; Yengoyan, L. *J. Biol. Chem.* **1966,** *241,* 3970–3987.
33. Phillips, M. A.; Wildung, M. R.; Williams, D. C.; Hyatt, D. C.; Croteau, R. *Arch. Biochem. Biophys.* **2003,** *411,* 267–276. doi:10.1016/s0003-9861(02)00746-4
34. Dickschat, J. S. *Nat. Prod. Rep.* **2011,** *28,* 1917–1936. doi:10.1039/c1np00063b
35. Srividya, N.; Davis, E. M.; Croteau, R. B.; Lange, B. M. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112,* 3332–3337. doi:10.1073/pnas.1501203112
36. Hong, Y. J.; Tantillo, D. J. *Org. Biomol. Chem.* **2010,** *8,* 4589–4600. doi:10.1039/c0ob00167h
37. Isegawa, M.; Maeda, S.; Tantillo, D. J.; Morokuma, K. *Chem. Sci.* **2014,** *5,* 1555–1560. doi:10.1039/c3sc53293c
38. Moss, R. A. *J. Phys. Org. Chem.* **2014,** *27,* 374–379. doi:10.1002/poc.3290
39. Bian, G.; Rinkel, J.; Wang, Z.; Lauterbach, L.; Hou, A.; Yuan, Y.; Deng, Z.; Liu, T.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2018,** *57,* 15887–15890. doi:10.1002/anie.201809954
40. Arigoni, D. *Pure Appl. Chem.* **1975,** *41,* 219–245. doi:10.1351/pac197541010219
41. Hong, Y. J.; Tantillo, D. J. *Org. Lett.* **2006,** *8,* 4601–4604. doi:10.1021/ol061884f
42. Asuming, W. A.; Beauchamp, P. S.; Descalzo, J. T.; Dev, B. C.; Dev, V.; Frost, S.; Ma, C. W. *Biochem. Syst. Ecol.* **2005,** *33,* 17–26. doi:10.1016/j.bse.2004.06.005
43. Rabe, P.; Barra, L.; Rinkel, J.; Riclea, R.; Citron, C. A.; Klapschinski, T. A.; Janusko, A.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2015,** *54,* 13448–13451. doi:10.1002/anie.201507615
44. Hinkley, S. F. R.; Perry, N. B.; Weavers, R. T. *Phytochemistry* **1994,** *35,* 1489–1494. doi:10.1016/s0031-9422(00)86882-x
45. Cane, D. E.; Tandon, M. *J. Am. Chem. Soc.* **1995,** *117,* 5602–5603. doi:10.1021/ja00125a029
46. Hu, Y.; Chou, W. K. W.; Hopson, R.; Cane, D. E. *Chem. Biol.* **2011,** *18,* 32–37. doi:10.1016/j.chembiol.2010.11.008
47. Rinkel, J.; Rabe, P.; Garbeva, P.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2016,** *55,* 13593–13596. doi:10.1002/anie.201608042
48. Benedict, C. R.; Lu, J. L.; Pettigrew, D. W.; Liu, J.; Stipanovic, R. D.; Williams, H. *J. Plant Physiol.* **2001,** *125,* 1754–1765. doi:10.1104/pp.125.4.1754
49. Rinkel, J.; Rabe, P.; Chen, X.; Köllner, T. G.; Chen, F.; Dickschat, J. S. *Chem. – Eur. J.* **2017,** *23,* 10501–10505. doi:10.1002/chem.201702704
50. Rinkel, J.; Lauterbach, L.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2019,** *58,* 452–455. doi:10.1002/anie.201812216
51. Mitsuhashi, T.; Rinkel, J.; Okada, M.; Abe, I.; Dickschat, J. S. *Chem. – Eur. J.* **2017,** *23,* 10053–10057. doi:10.1002/chem.201702766
52. Rabe, P.; Klapschinski, T. A.; Dickschat, J. S. *ChemBioChem* **2016,** *17,* 1333–1337. doi:10.1002/cbic.201600237
53. Rinkel, J.; Rabe, P.; Dickschat, J. S. *Eur. J. Org. Chem.* **2019,** 351–359. doi:10.1002/ejoc.201800217

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Phylogenomic analyses and distribution of terpene synthases among *Streptomyces*

Lara Martín-Sánchez‡[1], Kumar Saurabh Singh‡[2], Mariana Avalos[1,3], Gilles P. van Wezel[1,3], Jeroen S. Dickschat*[1,4] and Paolina Garbeva*[1]

Full Research Paper

Open Access

Address:
[1]Department of Microbial Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Droevendaalsesteeg 10, 6708 PB Wageningen, The Netherlands, [2]College of Life and Environmental Sciences, Biosciences, University of Exeter, Penryn Campus, Penryn, Cornwall TR10 9FE, United Kingdom, [3]Institute of Biology, Leiden University, Sylviusweg 72, 2333 BE Leiden,The Netherlands and [4]University of Bonn, Kekulé-Institute of Organic Chemistry and Biochemistry, Gerhard-Domagk-Straße 1, 53121 Bonn, Germany

Email:
Jeroen S. Dickschat* - dickschat@uni-bonn.de; Paolina Garbeva* - P.Garbeva@nioo.knaw.nl

* Corresponding author ‡ Equal contributors

Keywords:
biosynthesis; evolution; geosmin; *Streptomyces*; terpenes

Beilstein J. Org. Chem. **2019,** *15,* 1181–1193.
doi:10.3762/bjoc.15.115

Received: 14 March 2019
Accepted: 17 May 2019
Published: 29 May 2019

This article is part of the thematic issue "Terpenes".

Associate Editor: K. N. Allen

Abstract

Terpene synthases are widely distributed among microorganisms and have been mainly studied in members of the genus *Streptomyces*. However, little is known about the distribution and evolution of the genes for terpene synthases. Here, we performed whole-genome based phylogenetic analysis of *Streptomyces* species, and compared the distribution of terpene synthase genes among them. Overall, our study revealed that ten major types of terpene synthases are present within the genus *Streptomyces*, namely those for geosmin, 2-methylisoborneol, *epi*-isozizaene, 7-*epi*-α-eudesmol, *epi*-cubenol, caryolan-1-ol, cyclooctat-9-en-7-ol, isoafricanol, pentalenene and α-amorphene. The *Streptomyces* species divide in three phylogenetic groups based on their whole genomes for which the distribution of the ten terpene synthases was analysed. Geosmin synthases were the most widely distributed and were found to be evolutionary positively selected. Other terpene synthases were found to be specific for one of the three clades or a subclade within the genus *Streptomyces*. A phylogenetic analysis of the most widely distributed classes of *Streptomyces* terpene synthases in comparison to the phylogenomic analysis of this genus is discussed.

Introduction

Streptomyces are soil bacteria that belong to the order of actinomycetales and are a rich source of natural products with broad biotechnological interest. Species of this genus have a remarkable genetic potential to produce a large variety of secondary metabolites with different functions including antibiotics, antifungals, pigments or immunosuppressants [1-3]. These

are compounds of diverse chemical nature such as polyketides, peptides, aminoglycosides or terpenoids [4,5].

Terpenoids are the largest and the most diverse class of natural compounds known to date and include the initial products of terpene synthases and all derivatives made from them in tailoring steps. This very diverse class of organic compounds is best known as plant metabolites. However, recent studies revealed that terpenoids can be produced by all kingdoms of life including bacteria, fungi and protists [6-10]. The ability of an organism to produce terpenoids relies on whether the organism contains terpene synthase genes. Biosynthetically, the production of the different types of terpenes depends on the precursors that these synthases can accommodate: geranyl diphosphate (monoterpenes, C10), farnesyl diphosphate (sesquiterpenes, C15) and geranylgeranyl diphosphate (diterpenes, C20). The biological function of terpenes is best studied for plants where they play important roles in aboveground plant–insect, plant–pathogen and plant–plant interactions [11]. However, terpenes might also play important roles in belowground interspecific interactions [12]. Terpene synthases are in fact widely distributed among soil microorganisms, and they have been mainly studied in *Streptomyces* species [9,13]. Some volatile terpenes, such as geosmin and 2-methylisoborneol (2-MIB), responsible for the smell of wet soil after rain, have been known for a long time to be produced by *Streptomyces* species [14,15]. Many terpene synthases from *Streptomyces* have been studied and characterised [13]. However, little is known about the distribution of terpene synthase encoding genes among *Streptomyces*. Are terpene synthase genes specific for certain species or randomly distributed among *Streptomyces*?

To address this question, phylogenomic analyses of *Streptomyces* species were performed, using complete genomes available in the NCBI database and compared the distribution of terpene synthase genes among them. Furthermore, we studied whether phylogenetic trees calculated based on the three most abundant terpene synthases in *Streptomyces* represent the evolution of the *Streptomyces* species based on the whole genome-based phylogenetic analyses.

Results and Discussion

Whole genome-based phylogenetic analyses of *Streptomyces* species

Genome sequences from 93 *Streptomyces* species for which a complete genome was available (represented by a single scaffold and a complete list of annotated protein sequences), were selected to construct a whole genome-based phylogenetic tree (Figure 1). The NCBI database was accessed on September 30th 2018. An orthologues-based approach was adopted to generate a species tree using OrthoFinder. OrthoFinder resulted in a total of 19980 orthologue groups (Table S1, Supporting Information File 1). A total number of 575 single copy orthologues were further selected for the generation of the species tree. Based on these phylogenetic analyses, the *Streptomyces* species clustered in three different clades (indicated in blue, green and red in Figure 1). This separation into three different clades agrees with the study previously reported by McDonald and Currie, 2017 [16]. Based on phylogenetic analyses of 94 housekeeping genes, they showed a separation of *Streptomyces* species in two major clades and a third group of other lineages.

Distribution of terpene synthases in *Streptomyces*

We analysed the distribution of the different types of terpene synthases among *Streptomyces* species with complete genomes. Besides a few rare terpene synthases only occuring in a few or a single species, ten major types of terpene synthases were present among these *Streptomyces* species, including the terpene synthases for geosmin (**1**), 2-methylisoborneol (2-MIB) (**2**), *epi*-isozizaene (**3**), 7-*epi*-α-eudesmol (**4**), *epi*-cubenol (**5**), caryolan-1-ol (**6**), cyclooctat-9-en-7-ol (**7**), isoafricanol (**8**), pentalenene (**9**) and α-amorphene (**10**) (Figure 1 and Figure 2).

The geosmin synthases were the most widely distributed, as they were present in all except one of the *Streptomyces* species (*S. pactum* KLBMP 5084) (Figure 1). This finding suggests that geosmin may have an important ecological function as a chemical signal or as protective specialised metabolite against biotic and abiotic stresses, similarly as the roles played by terpenoids in plants [11]. However, although geosmin was discovered more than 50 years ago [14], its biological or ecological function still remains unclear. *Streptomyces pactum* KLBMP 5084 (the only species included in this study that does not carry geosmin synthases) is an endophytic plant growth-promoting bacterium that provides salt tolerance to the halophytic plant *Limonium sinense* (Plumbaginaceae) [17]. The absence of a geosmin synthase in this bacterium leads us to hypothesise that the role of geosmin may be complemented by the plant host. The only other plant endophyte among the 93 species is *Streptomyces* sp. SAT1 (see Table S9 (Supporting Information File 1) for a list of the isolation sources and habitats of the 93 strains). This strain is an endophyte of the flowering plant *Adenophora trachelioides* from the Campanulaceae family and it does contain a copy of *geoA*, the gene encoding for geosmin synthase. Some species such as *Streptomyces* sp. SirexAA-E harbour a silent geosmin synthase encoding gene in their genomes and do not produce this degraded sesquiterpene under laboratory culture conditions [9]. It will therefore be interesting to investigate whether the geosmin synthase in *Streptomyces* sp. SAT1 is expressed, and to further determine the role of terpenoids in the endophytic life style.

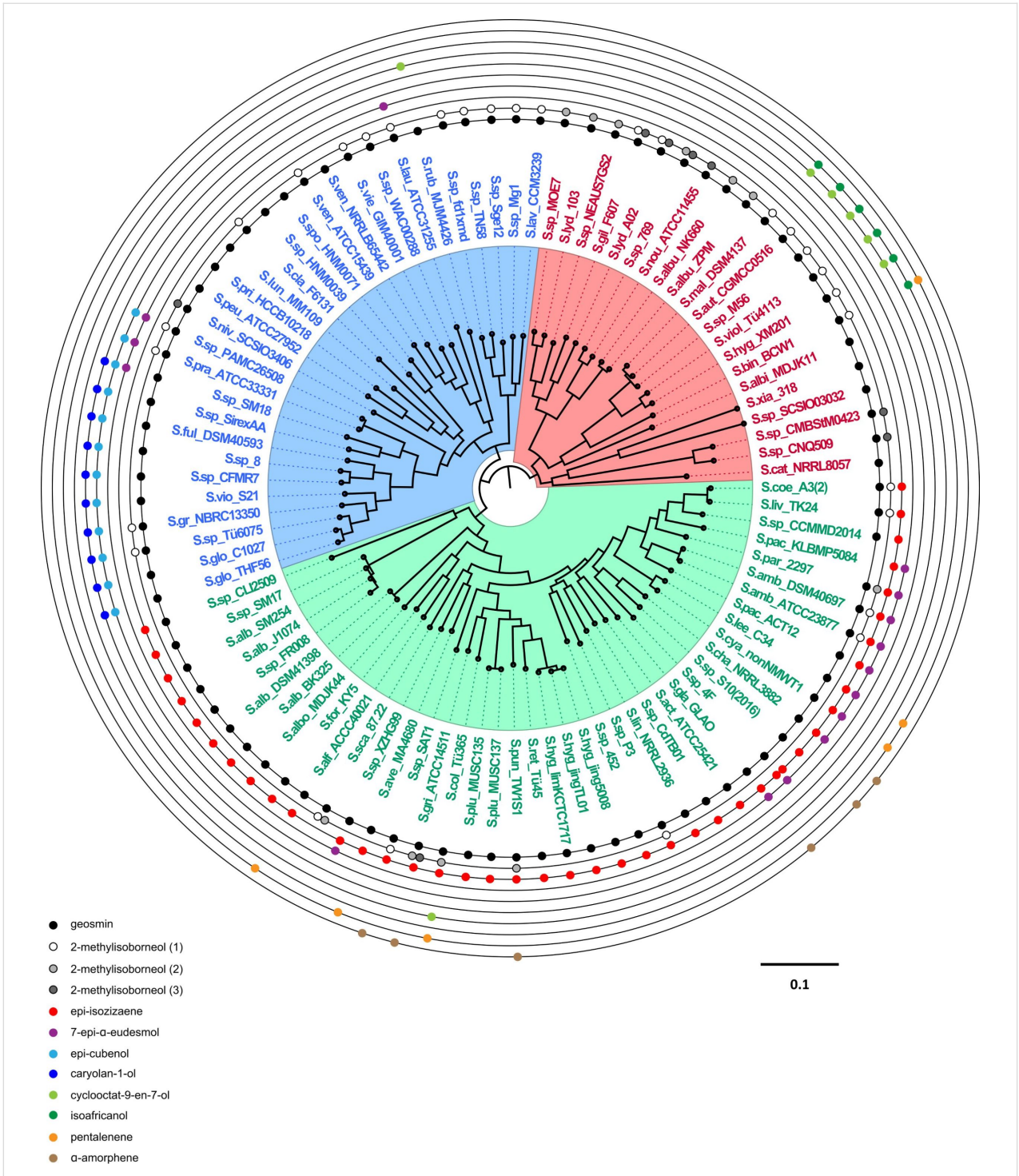


Figure 1: Whole-genome phylogenetic analyses of *Streptomyces* species. Rooted maximum likelihood phylogeny of 93 *Streptomyces* species with fully sequenced genomes based on 575 conserved single copy orthologues. The species separated in three main groups are indicated by different colour-shaded areas. The outer rings show the distribution of different types of terpene synthases in the *Streptomyces* species. Another version of this tree using 5 non-*Streptomyces* species as outgroups can be found in the Figure S1 in Supporting Information File 1. The GenBank accession numbers of the sequences are provided in Table S2 (Supporting Information File 1).

The first geosmin synthase was characterised from *Streptomyces coelicolor* [18]. Geosmin synthases are composed of two domains that both exhibit the typical highly conserved motifs of type I terpene synthases, including the aspartate-rich motif, the NSE triad, the pyrophosphate sensor and the RY pair [19-21]. Both domains have a catalytic activity, the N-terminal domain

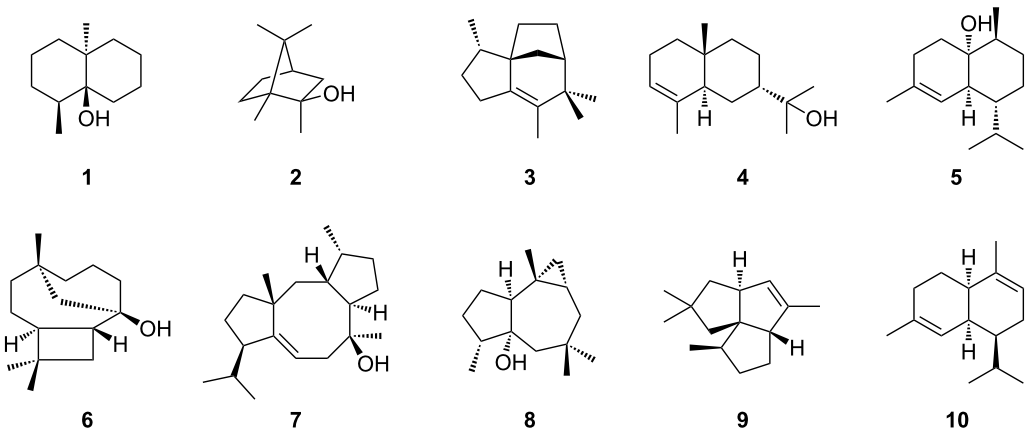
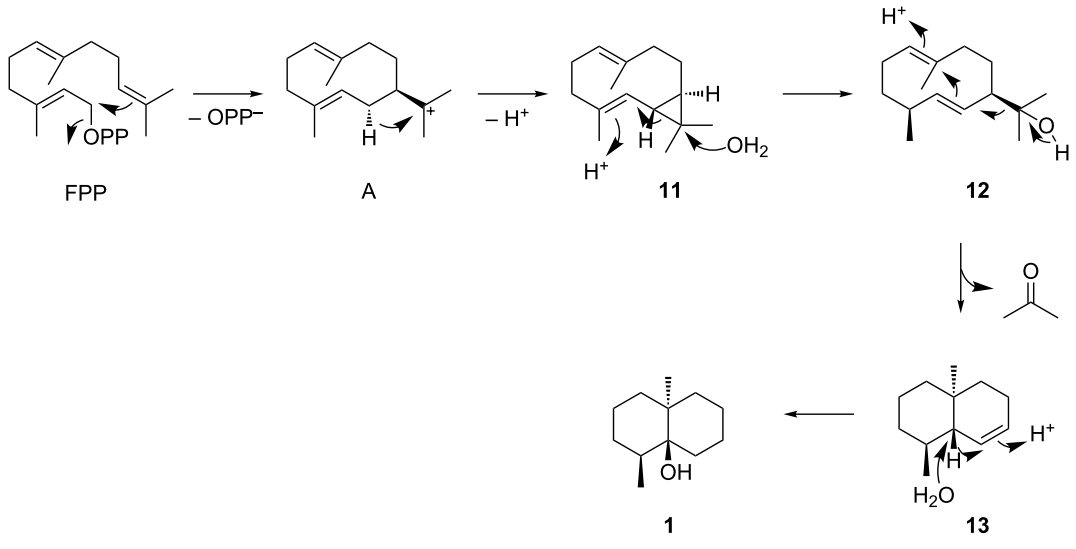


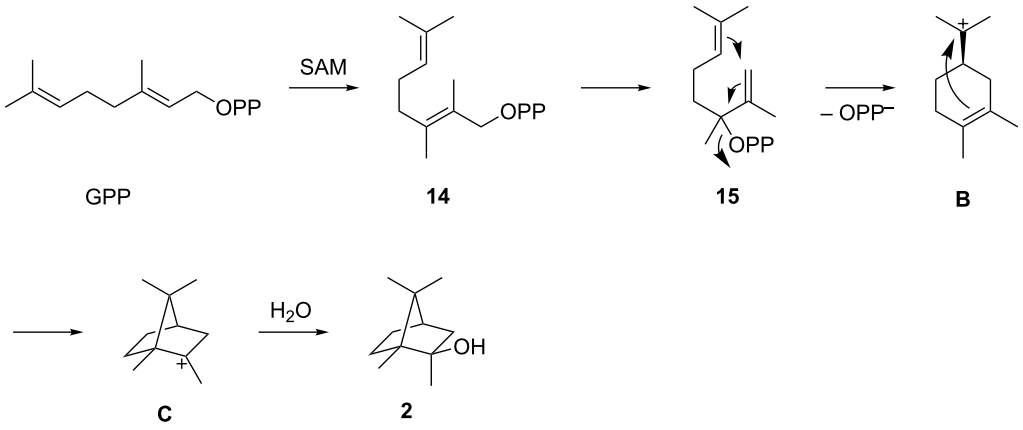
Figure 2: Structures of the products of the ten most abundant terpene synthases in *Streptomyces*.

for the conversion of FPP into the intermediate sesquiterpene alcohol (1(10)*E*,5*E*)-germacradien-11-ol (**12**), and the C-terminal domain for its further conversion into geosmin with cleavage of **12** into acetone and the octalin **13** through a retro-Prins fragmentation (Scheme 1) [22-24]. The proposed neutral intermediate isolepidozene (**11**) has so far only been reported from the S233A enzyme variant of geosmin synthase from *S. coelicolor* [18].

The second most widely distributed terpene synthases are the 2-MIB synthases (Figure 1). As discussed below, the phylogenetic analysis of 2-MIB synthases classifies these enzymes into three different groups. This distribution is also indicated in Figure 1 (white, light gray and dark gray circles). The 2-MIB synthases are present in members of all the three clades from the whole genome phylogenetic tree (Figure 1), but are most abundant in members of the clade depicted in red. These terpene synthases catalyse a unique cyclisation reaction utilizing the modified substrate 2-methyl-GPP to form 2-MIB (**2**) [25,26]. An *S*-adenosylmethionine (SAM) dependent methyl transferase is responsible for the methylation of GPP into 2-methyl-GPP (**14**, Scheme 2). Its isomerisation to **15** allows for a cyclisation via the cationic intermediates **B** and **C** to **2**. Genes encoding for SAM-dependent methyl transferases were found forming a cluster together with the 2-MIB synthase in several *Streptomyces* species [26,27]. Besides the C-terminal domain typical of class I terpene synthases, these enzymes contain an additional proline-rich N-terminal domain that



Scheme 1: Mechanism for the cyclisation of FPP to geosmin.

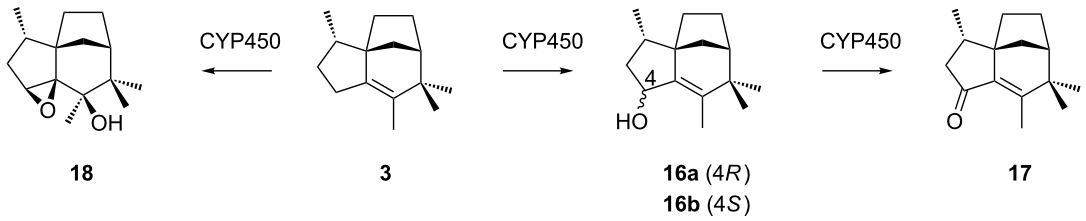


Scheme 2: Biosynthesis of 2-MIB (**2**). First, GPP is methylated to **14** by a SAM-dependent methyltransferase, followed by a terpene synthase catalysed cyclisation through a cationic cascade to **2**.

appears to be disordered in the crystal structure of 2-MIB synthase. The function of this domain is unknown, but it is conserved in most 2-MIB synthases and not present in any other terpene synthase [20,28].

epi-Isozizaene (**3**) is a tricyclic sesquiterpene precursor of the antibiotic albaflavenone (**17**) (Scheme 3) [29]. Furthermore, both enantiomers of the corresponding alcohols (*R*)- and (*S*)-albaflavenol (**16ab**) and the epoxide 4β,5β-epoxy-2-*epi*-zizaan-6β-ol (**18**) are known oxidation products that are all made by a cytochrome P450 monooxygenase [10,29] that is genetically clustered with the *epi*-isozizaene synthase for the cyclisation of FPP to **3** [30]. These enzymes are the most widespread sesquiterpene synthases in bacteria, and their coding genes are present in the genomes of more than 100 of the sequenced *Streptomyces* species [13]. Interestingly, *epi*-isozizaene synthases are only present in members of one clade (indicated as the green clade) in the phylogenetic analyses shown in Figure 1 and occur in almost all species of this clade with one exception (*S. scabiei* 87.22), suggesting an (unknown) ecological function of **3** or one of its oxidation products for streptomycetes of this clade for their adaption to a specific ecological niche.

7-*epi*-α-Eudesmol (**4**) synthases are mostly present in a small group of species within the phylogenomic clade depicted in green in Figure 1, with some exceptions (*S. laurentii* ATCC 31255, *Streptomyces* sp PAMC 26508, *S. pratensis* ATCC 33331, *Streptomyces* sp_SM18 and *Streptomyces* sp. XZHG99, Figure 1). These exceptions may indicate horizontal gene transfer of the genes encoding for these enzymes. The sesquiterpene 7-*epi*-α-eudesmol synthase from *S. viridochromogenes* DSM 40736 has been chemically characterised in vivo by heterologous expression in *E. coli* BL21 and identification of the product in culture headspace extracts by GC–MS [31]. Compound **4** was also isolated from in vitro incubations of FPP with the recombinant enzyme and its optical rotation was shown to be opposite to the material from *Eucalyptus* [32], but the absolute configuration remains unknown. Production of this sesquiterpene by *S. viridochromogenes* DSM 40736 has not been observed [31], but **4** was occassionally reported from other streptomycetes encoding a 7-*epi*-α-eudesmol synthase [33,34].



Scheme 3: Oxidation products derived from **3** by the cytochrome P450 monooxygenase that is genetically clustered with the *epi*-isozizaene synthase in streptomycetes.

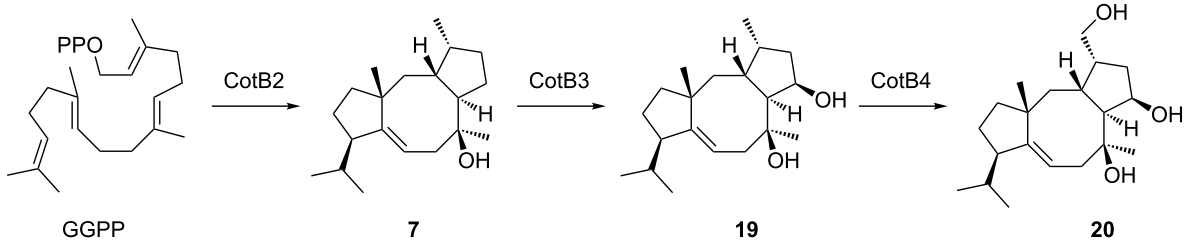
epi-Cubenol (**5**) and caryolan-1-ol (**6**) synthases almost always occur together in one strain. We found only two examples of a strain that has a gene for caryolan-1-ol synthase but not for *epi*-cubenol synthase. These enzymes were found only in a subbranch of closely related *Streptomyces* species from the blue clade and not present in members of any other phylogenomic group (Figure 1). Both enzymes have been identified and characterised in *S. griseus* NBRC 13350 [35,36] and their enzymatic mechanisms for the cyclisation of FPP have been investigated [35,37-39].

Cyclooctat-9-en-7-ol (**7**) and isoafricanol (**8**) synthases are mainly characteristic for a group of very closely related species in the phylogenomic clade depicted in red in Figure 1, with two exceptions, *S. rubrolavendulae* MJM4426 and *S. collinus* Tü 365, members of the other two phylogenomic clades that also present a cyclooctat-9-en-7-ol synthase. Cyclooctat-9-en-7-ol synthase (CotB2) from *S. melanosporofaciens* was the first bacterial type I diterpene cyclase characterised [40] and its crystal structure was the first of a diterpene cyclase of bacterial origin reported [41]. Isoafricanol synthases were first noticed in *S. violaceusniger* and *S. rapamycinicus* based on the presence of **8** in culture headspace extracts as a major sesquiterpene [34,42], followed by the biochemical characterisation of the recombinant enzyme from *Streptomyces malaysiensis* [43]. The diterpene **7** is a precursor to the lysophospholipase inhibitor cyclooctatin (**20**) formed by the action of two genetically clustered cytochrome P450 monooxygenases CotB3 and CotB4 (Scheme 4) [40,44], while no derivatives from **8** are currently known.

Pentalenene (**9**) and α-amorphene (**10**) synthases are the least abundant terpene synthases in *Streptomyces* species, each present in only 6 species (Figure 1). They are mostly present in members of the phylogenomic clade depicted in green in Figure 1, except for one case, *S. bingchenggensis* BCW1, but within the green clade their distribution is scattered and the number of identified genes for these enzymes is too low to draw conclusions on their occurrence in *Streptomyces*. The pentalenene synthase from *S. exfoliatus* was the first characterised bacterial terpene synthase [45,46]. Its crystal structure was also the first reported for a bacterial terpene synthase [47]. Pentalenene synthase catalyses the cyclisation of FPP into pentalenene, which is the first step in the biosynthesis of the antibiotic pentalenolactone. This mechanism has been extensively studied and involves the initial ionisation of the substrate FPP and the formation of a humulyl cation as an intermediate in the biosynthesis of pentalenene [45,46,48,49], while the later steps of the cyclisation cascade were subject to revision based on the findings of quantum chemical calculations [50,51]. The α-amorphene synthase from *S. viridochromogenes* DSM 40736 was characterised by heterologous expression in *E. coli* BL21 [31] and by in vitro experiments with the purified enzyme [32].

Phylogenetic analysis of geosmin synthases

In order to determine if the geosmin synthases co-evolved with the *Streptomyces* species a phylogenetic tree was constructed with the geosmin synthases of all the species present in the full genome tree. As seen in Figure 3, the geosmin synthases separated into different clades. These clades do not fully correspond with specific phylogenomic groups from the genome-based analyses. Most of the geosmin synthases of the green and red phylogenetic clade in the whole genome-based tree of Figure 1 grouped together into one clade. The enzymes from the blue phylogenetic clade in the genome-based tree were the most scattered. All these results may point to the occurrence of horizontal gene transfer within the genus *Streptomyces*. However, if bacteria from other taxonomic groups such as myxobacteria and cyanobacteria and their geosmin synthases are included in a phylogenetic analysis, it can be seen that the geosmin synthase amino acid sequences from distantly related organisms clearly fall into distant clades [33]. Therefore, these results could also be interpreted as evidence for a rapid evolution of secondary metabolite genes to create new natural products with beneficial ecological functions for the producing organism. While many streptomycetes produce geosmin as a major metabolite of their bouquets of volatiles, the number and amounts of geosmin synthase side products associated with it



Scheme 4: Biosynthesis of cyclooctatin (**20**) from **7**.

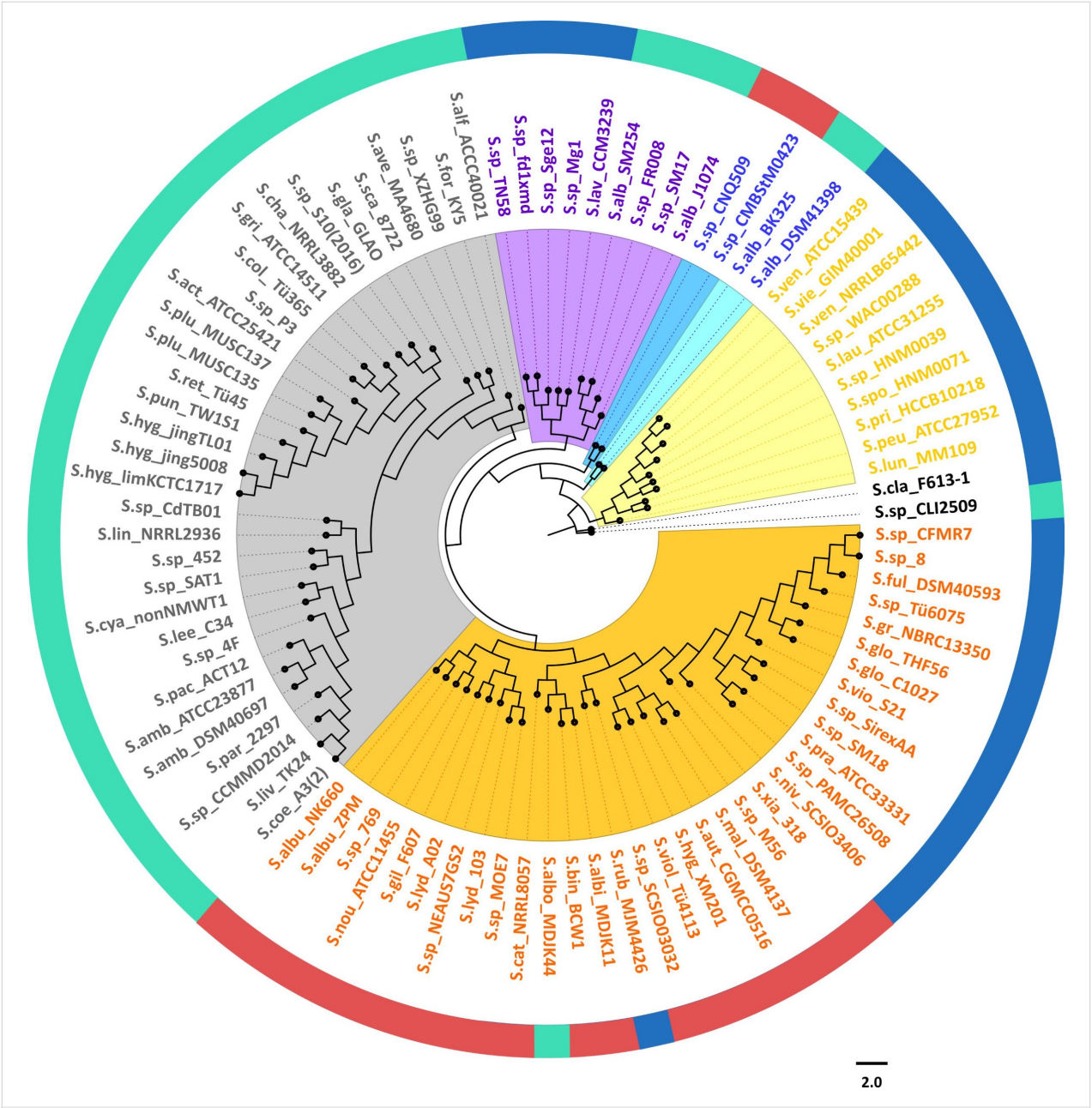


Figure 3: Phylogenetic tree of geosmin synthases. Unrooted maximum likelihood phylogenetic tree of 92 geosmin synthases from the *Streptomyces* species present in the phylogenetic tree in Figure 1. The colours in the outer ring correspond to the colours of the three main phylogenomic groups in the whole-genome species tree and indicate to which phylogenomic group each species belongs. The GenBank accession numbers of the geosmin synthases are listed in Table S3 (Supporting Information File 1).

can vary [33,34], possibly as a result of an evolution of enzyme function.

Phylogenetic analysis of 2-MIB synthases

To gain insights into the evolution of the 2-MIB synthases a phylogenetic analysis of all the enzymes present in the *Streptomyces* species analysed in our study were performed (Figure 4). The phylogenetic tree of the 2-MIB synthases shows a clear separation into three clades (also indicated in Figure 1: group 1, white circles, representing the major clade on the top of Figure 4; group 2, light grey circles, representing the clade on the bottom right; group 3, dark grey circles, representing the clade on the bottom left). Two of them are relatively distant from each other and even more so from the third clade where most species cluster together. This separation does not correspond with the separation observed based on the whole genome phylogenomic analyses. Only some of the enzymes that cluster together belong to species from the same phylogenomic group.

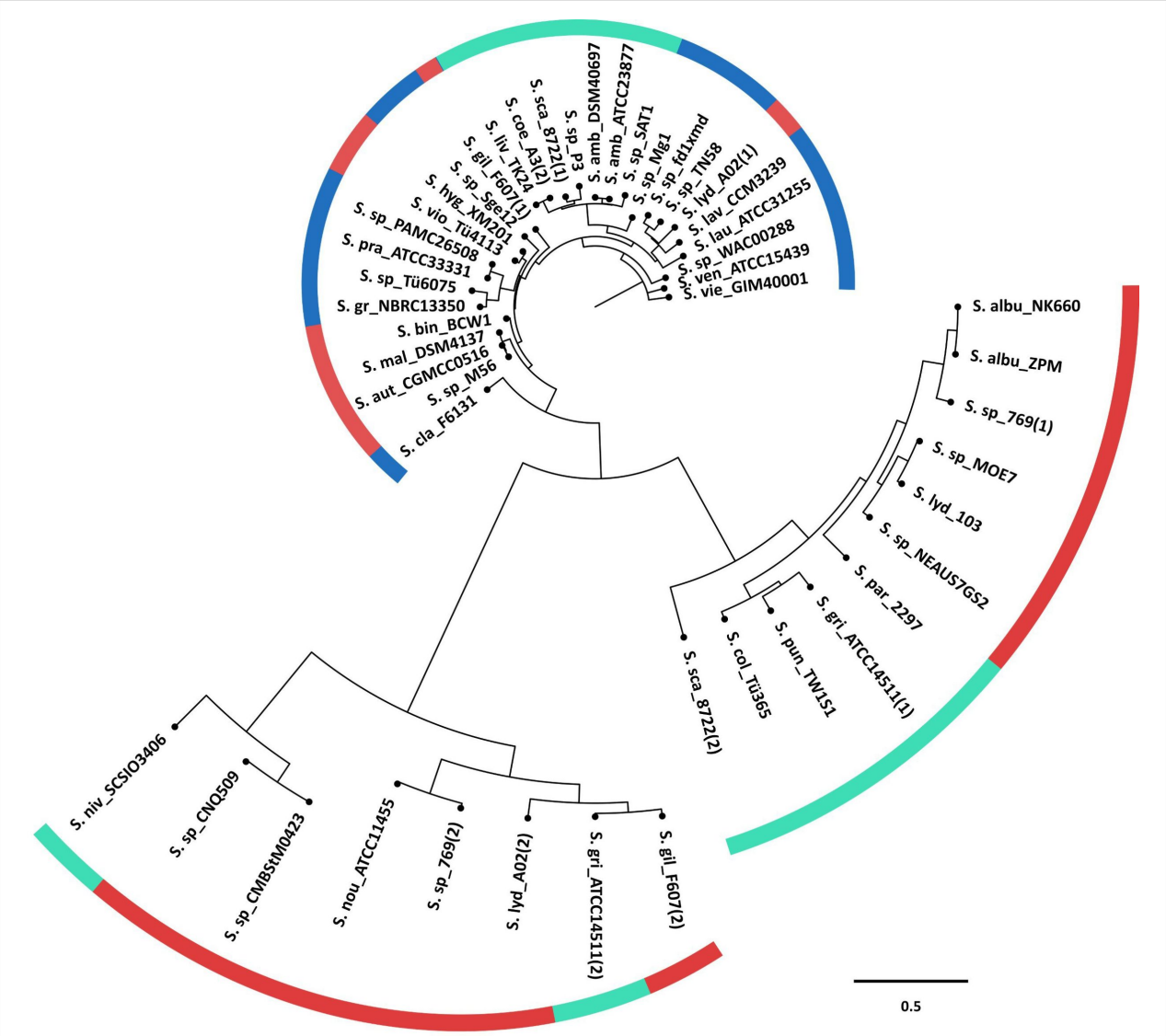


Figure 4: Phylogenetic tree of 2-MIB synthases. Unrooted maximum likelihood phylogenetic tree of 48 2-MIB synthases from the *Streptomyces* species in the phylogenetic tree in Figure 1. The colours of the outer curved lines correspond to the colours of the three main phylogenomic groups in the whole-genome species tree and indicate to which phylogenomic group each species belongs. The GenBank accession numbers of the 2-MIB synthases are listed in Table S4 (Supporting Information File 1).

This indicates that the evolution of these enzymes does not correspond to the evolution of the *Streptomyces* species, and that a different force is driving how these enzymes evolved.

Phylogenetic analysis of *epi*-isozizaene synthases

epi-Isozizaene synthases are terpene synthases belonging only to a specific phylogenomic group of *Streptomyces* species. The phylogenetic analysis presented in Figure 5 shows two clades containing most of *epi*-isozizaene synthases and three other minor clades. Not all the enzymes are clustering in the same way as their containing species based on the whole-genome phylogenetic analyses. For example, *S. pactum* ACT12 *epi*-isozizaene synthase clusters together with that of *Streptomyces* sp. 4F, while these two species were present in different branches in the full-genome-based phylogenomic tree. *Streptomyces* sp. 4F clustered together with *S. qaidamensis* S10(2016) and *S. chartreusis* NRRL 3882 in the phylogenomic tree. However, a second *epi*-isozizaene synthase present in *Streptomyces* sp. 4F clusters together with that of *Streptomyces* sp. SAT1, while these two species were located in separate clades of the phylogenomic tree. The occurrence of two genes for terpene synthases with putatively the same function may more strongly point to horizontal gene transfer events. Other cases include the *epi*-isozizaene synthases from *Streptomyces* sp. 452, *S. glaucescens* GLAO, *S. lincolnensis* NRRL 2936 and *Strepto-*

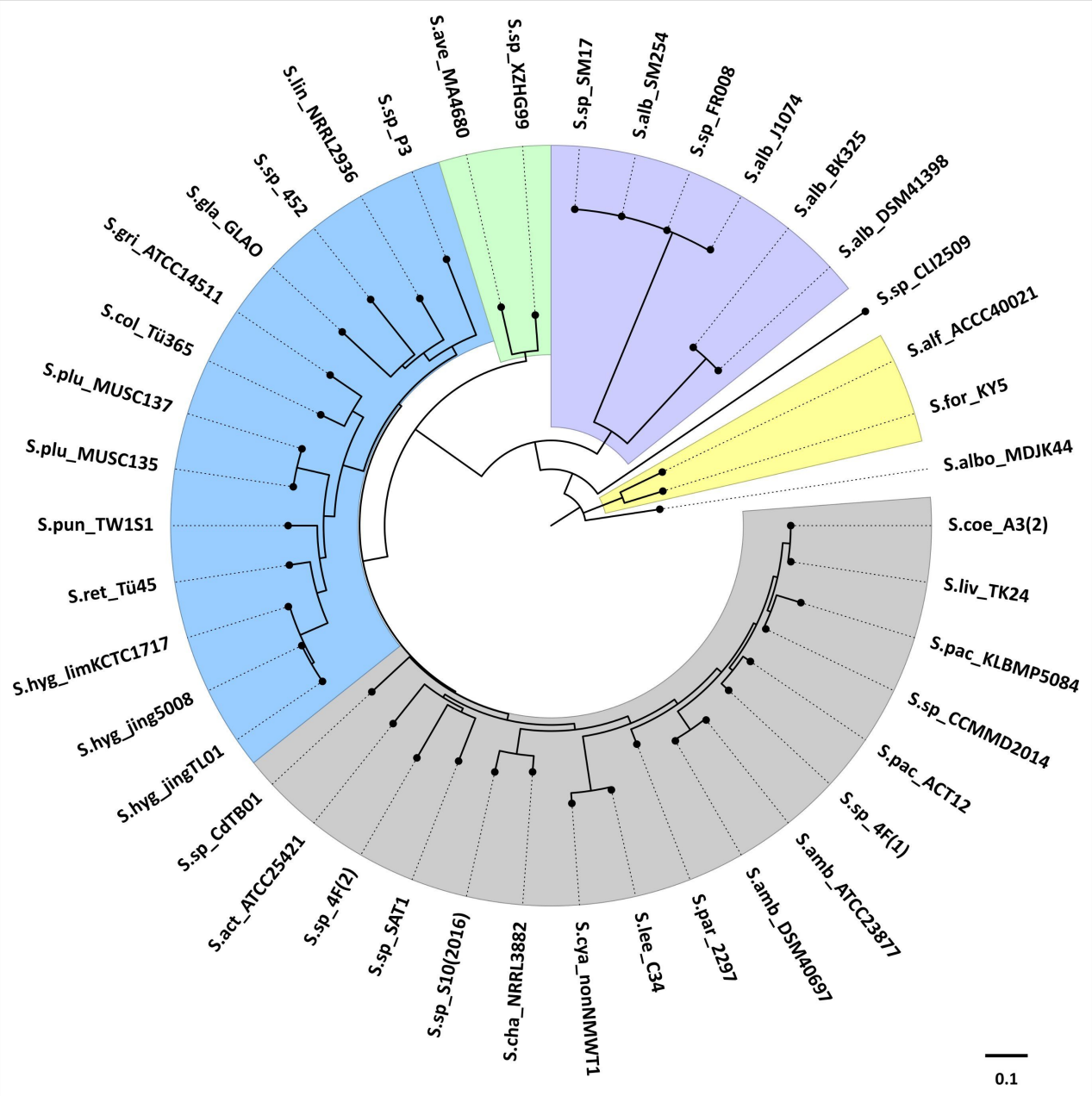


Figure 5: Phylogenetic tree of *epi*-isozizaene synthases. Unrooted maximum likelihood phylogenetic tree of 42 *epi*-isozizaene synthases from the *Streptomyces* species present in the phylogenomic tree in Figure 1. The GenBank accession numbers of the *epi*-isozizaene synthases are listed in Table S5 (Supporting Information File 1).

myces sp. P3 that group together with other enzymes, different to those belonging to species located in their same clade in the whole-genome phylogenomic analyses. This indicates also that some of these terpene synthases have evolved independently of the evolution of the *Streptomyces* species.

Phylogeny of terpene synthases does not correspond to species-level taxonomy

The comparison of the *Streptomyces* species whole genome-based phylogenetic tree and the three terpene synthase trees shows that not all three comparing phylogenies are congruent. All *Streptomyces* strains included in this study carry at least one copy of *geoA*, with one exception. However, the topology of the geosmin synthase tree is not in harmony with the species tree and only some tips of the trees are conserved (Figure S3, Supporting Information File 1). The topological incongruence is even higher for *epi*-isozizaene and 2-MIB synthase trees (Figures S4 and S5, Supporting Information File 1). Tree reconstruction artefacts cannot explain these incongruences because all phylogenies obtained good statistical support. These data

support horizontal gene transfers of terpene synthase genes in *Streptomyces*, but could also point to secondary metabolite genes as being less conserved than housekeeping (primary metabolism) genes. Rapid evolution of secondary metabolism can lead to new natural products with advanced ecological functions in specific ecological niches. If horizontal gene transfer is indeed of high importance, one intriguing question would be why there are almost no *Streptomyces* strains with two or more genes for geosmin synthases and *epi*-isozizaene synthases. This could be explained by the rapid loss of genetic information after uptake of redundant information. It may also reflect the mechanism of integration of the incoming genetic information into the chromosome of the target organism by homologous recombination within identical or highly similar nucleotide sequences. In this study, we searched for the minimal number of events that are required to reconcile the terpene synthase trees with the species tree by performing NOTUNG analyses [52] (for detailed explanations cf. Supporting Information File 1, pp. 33–35). The analyses indicated that the discrepancies between the terpene synthase trees and the species tree, can be explained by horizontal gene transfer of the genes encoding for terpene synthases.

Conclusion

Overall, this study confirmed that *Streptomyces* species divide in three phylogenomic groups, based here on their whole genomes. Analysis of the distribution of the ten most abundant classes of terpene synthases in *Streptomyces* led to the surprising result that some terpene synthases are restricted to one phylogenomic group or even a subgroup which may point to a specific ecological function of the terpene for the respective group of organisms. The phylogenetic analyses of terpene synthases are not congruent with the phylogenomic analyses. Hence, the evolution of these enzymes does not correspond to the evolution of the *Streptomyces* species, possibly pointing to horizontal gene transfers as an important mechanism involved in the distribution of terpene synthase genes.

In this study, we focused on the distribution and evolution of terpenes synthases among *Streptomyces* species. It would be interesting in follow-up studies to assess the distribution and evolution of these genes among other bacteria, fungi, protists and plants. In addition, a deeper knowledge of the ecological function of terpenes in bacteria and in the interaction with their environment is highly desired.

Experimental

Streptomyces genomes selection

Genomes with whole sequences available in the NCBI database (thus not partial sequences) were included. Custom shell scripts (https://github.com/kumarsaurabh20/distribution_of_terpene_synthases) were used to filter and download the nucleotide and protein sequences of all complete genomes including an annotation file in GFF format. The 93 selected sequences and their accession numbers are listed in Table S2 (Supporting Information File 1).

Construction of orthologous gene families

Sequence data of the proteins from the 93 *Streptomyces* species described above were collected. After removing sequences shorter than 50 amino acids, a total of 171,033 sequences were used to construct orthologous gene families using OrthoFinder – v 2.2.6 [53] applying the default setting (BLASTp e-value cut-off = 1e−5; MCL inflation I = 1.5). Using single-copy orthologues a species tree was inferred from unrooted gene trees that were constructed from all single copy genes using the STAG algorithm and the species tree was rooted using the STRIDE algorithm [54]. Both tools are available as core utilities in the OrthoFinder pipeline.

Phylogenetic analyses

Phylogenetic analyses on three different terpene synthases (geosmin synthases, 2-MIB synthases and *epi*-isozizaene synthases) were performed. Protein and nucleotide sequences were extracted from the *Streptomyces* genomes based on their distribution. Phylogenetic trees were generated using the protein dataset. Sequences were aligned with Mafft version 7.313 [55] using default parameters including *--auto* and *--inputorder*. All the alignments were trimmed for gaps and ambiguously aligned regions with BMGE – v 1.12 [56] using default parameters. For phylogenetic analyses, ProtTest – v 3.1.2 [57] was used to evaluate all evolutionary models under a AIC and BIC criterion. Maximum likelihood analyses were performed in RAxML – v 8.2.12 [58] under JTT+I+G (PROTGAMAMALG) model with rapid bootstrapping of 1000 replicates. GenBank accessions for each sequence are shown in Tables S3 to S5 in Supporting Information File 1.

Molecular evolution analysis

The coding DNA sequence (CDS) of the three terpene synthase genes (coding for geosmin synthases, 2-MIB synthases and *epi*-isozizaene synthases) in the 93 *Streptomyces* species were collected and aligned with Mafft version 7.313 using default parameters. Geneious – v 9.1 [59] was used to correct frame shifts and premature stop codons. Scripts published in [60] were used to generate codon-based alignments. We used HyPhy instance [61] to perform molecular evolution analysis. To test if positive selection occurred on a proportion of branches in the terpene synthase trees, the SLAC [62] model was used which is an improved version of the commonly used branch-site model. To test the hypothesis that individual sites have been subjected to episodic, positive or diversifying selection, site-specific model

FUBAR [63] was used. Additionally, aBSREL [64] model was used to infer nonsynonymous (dN) and synonymous (dS) substitution rates on a per-site basis for a given coding sequence alignment and corresponding phylogeny. The treefix-DTL (duplication-transfer-loss) software, version 1.0.2 [64], was applied to fix the topology of each terpene synthase tree under default settings with an alpha value of 0.05 for the paired-site test and the model closest to PROTGAMMALGF available via treefix-DTL (PROTGAMMAJTTF) as RAxML substitution model. To reconstruct the types and numbers of the evolutionary events that explain the discrepancies (if any) between the final topologies, NOTUNG version 2.9 [52] was run under default settings (modified weight parameters edge weight = 0.9; duplication weight = 2.0; transfer weight = 3.0; losses weight = 1.0) except for the permission of horizontal transfers and the use of a DTL cost matrix of 2-3-1, corresponding to default costs used by treefix-DTL.

Supporting Information

Supporting Information File 1

Additional figures and tables.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-115-S1.pdf]

Acknowledgements

This work was supported by NWO ALWOP.178 grant. This is publication 6728 of the NIOO-KNAW.

ORCID® iDs

Lara Martín-Sánchez - https://orcid.org/0000-0001-5514-4712
Kumar Saurabh Singh - https://orcid.org/0000-0001-8352-5897
Gilles P. van Wezel - https://orcid.org/0000-0003-0341-1561
Jeroen S. Dickschat - https://orcid.org/0000-0002-0102-0631

References

1. Barka, E. A.; Vatsa, P.; Sanchez, L.; Gaveau-Vaillant, N.; Jacquard, C.; Klenk, H.-P.; Clément, C.; Ouhdouch, Y.; van Wezel, G. P. *Microbiol. Mol. Biol. Rev.* **2016,** *80,* No. 1. doi:10.1128/mmbr.00019-15
2. Bentley, S. D.; Chater, K. F.; Cerdeño-Tárraga, A.-M.; Challis, G. L.; Thomson, N. R.; James, K. D.; Harris, D. E.; Quail, M. A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C. W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C.-H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabbinowitsch, E.; Rajandream, M.-A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B. G.; Parkhill, J.; Hopwood, D. A. *Nature* **2002,** *417,* 141–147. doi:10.1038/417141a
3. Bérdy, J. *J. Antibiot.* **2005,** *58,* 1–26. doi:10.1038/ja.2005.1
4. Nett, M.; Ikeda, H.; Moore, B. S. *Nat. Prod. Rep.* **2009,** *26,* 1362–1384. doi:10.1039/b817069j
5. Medema, M. H.; Kottmann, R.; Yilmaz, P.; Cummings, M.; Biggins, J. B.; Blin, K.; de Bruijn, I.; Chooi, Y. H.; Claesen, J.; Coates, R. C.; Cruz-Morales, P.; Duddela, S.; Düsterhus, S.; Edwards, D. J.; Fewer, D. P.; Garg, N.; Geiger, C.; Gomez-Escribano, J. P.; Greule, A.; Hadjithomas, M.; Haines, A. S.; Helfrich, E. J. N.; Hillwig, M. L.; Ishida, K.; Jones, A. C.; Jones, C. S.; Jungmann, K.; Kegler, C.; Kim, H. U.; Kötter, P.; Krug, D.; Masschelein, J.; Melnik, A. V.; Mantovani, S. M.; Monroe, E. A.; Moore, M.; Moss, N.; Nützmann, H.-W.; Pan, G.; Pati, A.; Petras, D.; Reen, F. J.; Rosconi, F.; Rui, Z.; Tian, Z.; Tobias, N. J.; Tsunematsu, Y.; Wiemann, P.; Wyckoff, E.; Yan, X.; Yim, G.; Yu, F.; Xie, Y.; Aigle, B.; Apel, A. K.; Balibar, C. J.; Balskus, E. P.; Barona-Gómez, F.; Bechthold, A.; Bode, H. B.; Borriss, R.; Brady, S. F.; Brakhage, A. A.; Caffrey, P.; Cheng, Y.-Q.; Clardy, J.; Cox, R. J.; De Mot, R.; Donadio, S.; Donia, M. S.; van der Donk, W. A.; Dorrestein, P. C.; Doyle, S.; Driessen, A. J. M.; Ehling-Schulz, M.; Entian, K.-D.; Fischbach, M. A.; Gerwick, L.; Gerwick, W. H.; Gross, H.; Gust, B.; Hertweck, C.; Höfte, M.; Jensen, S. E.; Ju, J.; Katz, L.; Kaysser, L.; Klassen, J. L.; Keller, N. P.; Kormanec, J.; Kuipers, O. P.; Kuzuyama, T.; Kyrpides, N. C.; Kwon, H.-J.; Lautru, S.; Lavigne, R.; Lee, C. Y.; Linquan, B.; Liu, X.; Liu, W.; Luzhetskyy, A.; Mahmud, T.; Mast, Y.; Méndez, C.; Metsä-Ketelä, M.; Micklefield, J.; Mitchell, D. A.; Moore, B. S.; Moreira, L. M.; Müller, R.; Neilan, B. A.; Nett, M.; Nielsen, J.; O'Gara, F.; Oikawa, H.; Osbourn, A.; Osburne, M. S.; Ostash, B.; Payne, S. M.; Pernodet, J.-L.; Petricek, M.; Piel, J.; Ploux, O.; Raaijmakers, J. M.; Salas, J. A.; Schmitt, E. K.; Scott, B.; Seipke, R. F.; Shen, B.; Sherman, D. H.; Sivonen, K.; Smanski, M. J.; Sosio, M.; Stegmann, E.; Süssmuth, R. D.; Tahlan, K.; Thomas, C. M.; Tang, Y.; Truman, A. W.; Viaud, M.; Walton, J. D.; Walsh, C. T.; Weber, T.; van Wezel, G. P.; Wilkinson, B.; Willey, J. M.; Wohlleben, W.; Wright, G. D.; Ziemert, N.; Zhang, C.; Zotchev, S. B.; Breitling, R.; Takano, E.; Glöckner, F. O. *Nat. Chem. Biol.* **2015,** *11,* 625–631. doi:10.1038/nchembio.1890
6. Chen, X.; Köllner, T. G.; Jia, Q.; Norris, A.; Santhanam, B.; Rabe, P.; Dickschat, J. S.; Shaulsky, G.; Gershenzon, J.; Chen, F. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113,* 12132–12137. doi:10.1073/pnas.1610379113
7. Song, C.; Mazzola, M.; Cheng, X.; Oetjen, J.; Alexandrov, T.; Dorrestein, P.; Watrous, J.; van der Voort, M.; Raaijmakers, J. M. *Sci. Rep.* **2015,** *5,* 12837. doi:10.1038/srep12837
8. Yamada, Y.; Cane, D. E.; Ikeda, H. Diversity and Analysis of Bacterial Terpene Synthases. In *Natural Product Biosynthesis by Microorganisms and Plant, Part A;* Hopwood, D. A., Ed.; Elsevier: Amsterdam, Netherlands, 2012; Vol. 515, pp 123–162. doi:10.1016/b978-0-12-394290-6.00007-0
9. Yamada, Y.; Kuzuyama, T.; Komatsu, M.; Shin-ya, K.; Omura, S.; Cane, D. E.; Ikeda, H. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112,* 857–862. doi:10.1073/pnas.1422108112
10. Takamatsu, S.; Lin, X.; Nara, A.; Komatsu, M.; Cane, D. E.; Ikeda, H. *Microb. Biotechnol.* **2011,** *4,* 184–191. doi:10.1111/j.1751-7915.2010.00209.x
11. Tholl, D. Biosynthesis and Biological Functions of Terpenoids in Plants. In *Biotechnology of Isoprenoids;* Schrader, J.; Bohlmann, J., Eds.; Advances in Biochemical Engineering/Biotechnology, Vol. 148; Springer: Cham, 2015; pp 63–106. doi:10.1007/10_2014_295
12. Schulz-Bohm, K.; Martín-Sánchez, L.; Garbeva, P. *Front. Microbiol.* **2017,** *8,* No. 2484. doi:10.3389/fmicb.2017.02484
13. Dickschat, J. S. *Nat. Prod. Rep.* **2016,** *33,* 87–110. doi:10.1039/c5np00102a
14. Gerber, N. N.; Lechevalier, H. A. *Appl. Microbiol.* **1965,** *13,* 935–938.

15. Medsker, L. L.; Jenkins, D.; Thomas, J. F.; Koch, C. *Environ. Sci. Technol.* **1969,** *3,* 476–477. doi:10.1021/es60028a008
16. McDonald, B. R.; Currie, C. R. *mBio* **2017,** *8,* e00644-17. doi:10.1128/mbio.00644-17
17. Qin, S.; Feng, W.-W.; Wang, T.-T.; Ding, P.; Xing, K.; Jiang, J.-H. *Plant Soil* **2017,** *416,* 117–132. doi:10.1007/s11104-017-3192-2
18. Jiang, J.; He, X.; Cane, D. E. *Nat. Chem. Biol.* **2007,** *3,* 711–715. doi:10.1038/nchembio.2007.29
19. Baer, P.; Rabe, P.; Fischer, K.; Citron, C. A.; Klapschinski, T. A.; Groll, M.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2014,** *53,* 7652–7656. doi:10.1002/anie.201403648
20. Christianson, D. W. *Chem. Rev.* **2017,** *117,* 11570–11648. doi:10.1021/acs.chemrev.7b00287
21. Seemann, M.; Zhai, G.; de Kraker, J.-W.; Paschall, C. M.; Christianson, D. W.; Cane, D. E. *J. Am. Chem. Soc.* **2002,** *124,* 7681–7689. doi:10.1021/ja026058q
22. Dickschat, J. S.; Bode, H. B.; Mahmud, T.; Müller, R.; Schulz, S. *J. Org. Chem.* **2005,** *70,* 5174–5182. doi:10.1021/jo050449g
23. Jiang, J.; Cane, D. E. *J. Am. Chem. Soc.* **2008,** *130,* 428–429. doi:10.1021/ja077792i
24. Nawrath, T.; Dickschat, J. S.; Müller, R.; Jiang, J.; Cane, D. E.; Schulz, S. *J. Am. Chem. Soc.* **2008,** *130,* 430–431. doi:10.1021/ja077790y
25. Dickschat, J. S.; Nawrath, T.; Thiel, V.; Kunze, B.; Müller, R.; Schulz, S. *Angew. Chem., Int. Ed.* **2007,** *46,* 8287–8290. doi:10.1002/anie.200702496
26. Wang, C.-M.; Cane, D. E. *J. Am. Chem. Soc.* **2008,** *130,* 8908–8909. doi:10.1021/ja803639g
27. Komatsu, M.; Tsuda, M.; Omura, S.; Oikawa, H.; Ikeda, H. *Proc. Natl. Acad. Sci. U. S. A.* **2008,** *105,* 7422–7427. doi:10.1073/pnas.0802312105
28. Köksal, M.; Chou, W. K. W.; Cane, D. E.; Christianson, D. W. *Biochemistry* **2012,** *51,* 3011–3020. doi:10.1021/bi201827a
29. Zhao, B.; Lin, X.; Lei, L.; Lamb, D. C.; Kelly, S. L.; Waterman, M. R.; Cane, D. E. *J. Biol. Chem.* **2008,** *283,* 8183–8189. doi:10.1074/jbc.m710421200
30. Lin, X.; Hopson, R.; Cane, D. E. *J. Am. Chem. Soc.* **2006,** *128,* 6022–6023. doi:10.1021/ja061292s
31. Rabe, P.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2013,** *52,* 1810–1812. doi:10.1002/anie.201209103
32. Rabe, P.; Schmitz, T.; Dickschat, J. S. *Beilstein J. Org. Chem.* **2016,** *12,* 1839–1850. doi:10.3762/bjoc.12.173
33. Citron, C. A.; Gleitzmann, J.; Laurenzano, G.; Pukall, R.; Dickschat, J. S. *ChemBioChem* **2012,** *13,* 202–214. doi:10.1002/cbic.201100641
34. Citron, C. A.; Barra, L.; Wink, J.; Dickschat, J. S. *Org. Biomol. Chem.* **2015,** *13,* 2673–2683. doi:10.1039/c4ob02609h
35. Nakano, C.; Horinouchi, S.; Ohnishi, Y. *J. Biol. Chem.* **2011,** *286,* 27980–27987. doi:10.1074/jbc.m111.265652
36. Nakano, C.; Tezuka, T.; Horinouchi, S.; Ohnishi, Y. *J. Antibiot.* **2012,** *65,* 551–558. doi:10.1038/ja.2012.68
37. Cane, D. E.; Tandon, M. *Tetrahedron Lett.* **1994,** *35,* 5355–5358. doi:10.1016/s0040-4039(00)73498-8
38. Cane, D. E.; Tandon, M. *J. Am. Chem. Soc.* **1995,** *117,* 5602–5603. doi:10.1021/ja00125a029
39. Cane, D. E.; Tandon, M.; Prabhakaran, P. C. *J. Am. Chem. Soc.* **1993,** *115,* 8103–8106. doi:10.1021/ja00071a023
40. Kim, S.-Y.; Zhao, P.; Igarashi, M.; Sawa, R.; Tomita, T.; Nishiyama, M.; Kuzuyama, T. *Chem. Biol.* **2009,** *16,* 736–743. doi:10.1016/j.chembiol.2009.06.007
41. Janke, R.; Görner, C.; Hirte, M.; Brück, T.; Loll, B. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2014,** *70,* 1528–1537. doi:10.1107/s1399004714005513
42. Riclea, R.; Citron, C. A.; Rinkel, J.; Dickschat, J. S. *Chem. Commun.* **2014,** *50,* 4228–4230. doi:10.1039/c4cc00177j
43. Rabe, P.; Samborskyy, M.; Leadlay, P. F.; Dickschat, J. S. *Org. Biomol. Chem.* **2017,** *15,* 2353–2358. doi:10.1039/c7ob00234c
44. Aoyagi, T.; Aoyama, T.; Kojima, F.; Hattori, S.; Honma, Y.; Hamada, M.; Takeuchi, T. *J. Antibiot.* **1992,** *45,* 1587–1591. doi:10.7164/antibiotics.45.1587
45. Cane, D. E.; Abell, C.; Tillman, A. M. *Bioorg. Chem.* **1984,** *12,* 312–328. doi:10.1016/0045-2068(84)90013-0
46. Cane, D. E.; Tillman, A. M. *J. Am. Chem. Soc.* **1983,** *105,* 122–124. doi:10.1021/ja00339a026
47. Lesburg, C. A.; Zhai, G.; Cane, D. E.; Christianson, D. W. *Science* **1997,** *277,* 1820–1824. doi:10.1126/science.277.5333.1820
48. Cane, D. E.; Oliver, J. S.; Harrison, P. H. M.; Abell, C.; Hubbard, B. R.; Kane, C. T.; Lattman, R. *J. Am. Chem. Soc.* **1990,** *112,* 4513–4524. doi:10.1021/ja00167a059
49. Cane, D. E.; Sohng, J.-K.; Lamberson, C. R.; Rudnicki, S. M.; Wu, Z.; Lloyd, M. D.; Oliver, J. S.; Hubbard, B. R. *Biochemistry* **1994,** *33,* 5846–5857. doi:10.1021/bi00185a024
50. Gutta, P.; Tantillo, D. J. *J. Am. Chem. Soc.* **2006,** *128,* 6172–6179. doi:10.1021/ja058031n
51. Zu, L.; Xu, M.; Lodewyk, M. W.; Cane, D. E.; Peters, R. J.; Tantillo, D. J. *J. Am. Chem. Soc.* **2012,** *134,* 11369–11371. doi:10.1021/ja3043245
52. Darby, C. A.; Stolzer, M.; Ropp, P. J.; Barker, D.; Durand, D. *Bioinformatics* **2017,** *33,* 640–649. doi:10.1093/bioinformatics/btw686
53. Emms, D. M.; Kelly, S. *Genome Biol.* **2015,** *16,* 157. doi:10.1186/s13059-015-0721-2
54. Emms, D. M.; Kelly, S. *Mol. Biol. Evol.* **2017,** *34,* 3267–3278. doi:10.1093/molbev/msx259
55. Katoh, K.; Standley, D. M. *Mol. Biol. Evol.* **2013,** *30,* 772–780. doi:10.1093/molbev/mst010
56. Criscuolo, A.; Gribaldo, S. *BMC Evol. Biol.* **2010,** *10,* 210. doi:10.1186/1471-2148-10-210
57. Darriba, D.; Taboada, G. L.; Doallo, R.; Posada, D. *Bioinformatics* **2011,** *27,* 1164–1165. doi:10.1093/bioinformatics/btr088
58. Stamatakis, A. *Bioinformatics* **2014,** *30,* 1312–1313. doi:10.1093/bioinformatics/btu033
59. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; Thierer, T.; Ashton, B.; Meintjes, P.; Drummond, A. *Bioinformatics* **2012,** *28,* 1647–1649. doi:10.1093/bioinformatics/bts199
60. Sydykova, D. K.; Jack, B. R.; Spielman, S. J.; Wilke, C. O. *F1000Research* **2018,** *6,* 1845. doi:10.12688/f1000research.12874.2
61. Pond, S. L. K.; Frost, S. D. W.; Muse, S. V. *Bioinformatics* **2005,** *21,* 676–679. doi:10.1093/bioinformatics/bti079
62. Smith, M. D.; Wertheim, J. O.; Weaver, S.; Murrell, B.; Scheffler, K.; Kosakovsky Pond, S. L. *Mol. Biol. Evol.* **2015,** *32,* 1342–1353. doi:10.1093/molbev/msv022
63. Murrell, B.; Moola, S.; Mabona, A.; Weighill, T.; Sheward, D.; Kosakovsky Pond, S. L.; Scheffler, K. *Mol. Biol. Evol.* **2013,** *30,* 1196–1205. doi:10.1093/molbev/mst030
64. Bansal, M. S.; Wu, Y.-C.; Alm, E. J.; Kellis, M. *Bioinformatics* **2015,** *31,* 1211–1218. doi:10.1093/bioinformatics/btu806

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Molecular basis for the plasticity of aromatic prenyltransferases in hapalindole biosynthesis

Takayoshi Awakawa[*1,2] and Ikuro Abe[*1,2]

Review

Open Access

Address:
[1]Graduate School of Pharmaceutical Sciences, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan and [2]Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-8657, Japan

Email:
Takayoshi Awakawa[*] - awakawa@mol.f.u-tokyo.ac.jp; Ikuro Abe[*] - abei@mol.f.u-tokyo.ac.jp

* Corresponding author

Keywords:
crystal structure; cyanobacteria; Friedel–Crafts reaction; hapalindole; prenyltransferase

Beilstein J. Org. Chem. **2019,** *15,* 1545–1551.
doi:10.3762/bjoc.15.157

Received: 10 May 2019
Accepted: 02 July 2019
Published: 11 July 2019

This article is part of the thematic issue "Terpenes".

Guest Editor: J. S. Dickschat

Abstract

Aromatic prenyltransferases (PTases) are enzymes that catalyze Friedel–Crafts reactions between aromatic compounds and isoprenoid diphosphates. In hapalindole biosynthesis, the aromatic PTases AmbP1 and AmbP3 exhibit surprisingly plastic selectivities. AmbP1 not only transfers the geranyl group on the C-3 of *cis*-indolylvinyl isonitrile, but also on the C-2, which is supressed in the presence of Mg^{2+} ions. AmbP3 transfers the dimethylallyl group on C-2 of hapalindole U in the reverse manner, but on C-2 of its C-10 stereoisomer in the normal manner. This review highlights the molecular bases of the AmbP1 and AmbP3 functions, elucidated through their X-ray crystal structures. The knowledge presented here will contribute to the understanding of aromatic PTase reactions and will enhance their uses as biocatalysts.

Introduction

Aromatic prenyltransferases (PTases) catalyze Friedel–Crafts reactions between aromatic prenyl acceptors and isoprenoid diphosphate prenyl donors to construct C–C, C–O, or C–N bonds, which enrich the structural diversity of aromatic natural products [1,2]. Their reactions are divided into two types, depending on where the cation is generated in the isoprenoid diphosphate: the “normal” prenylation in which the C-1 is attacked and the “reverse” prenylation in which the C-3 is attacked (Figure 1). It is important to study prenylation types for the chemoenzymatic synthesis of bioactive compounds, because the prenylated compounds exhibit better bioactivities due to their improved interactions with biological membranes [3]. The aromatic PTase superfamily involved in the secondary metabolism consists of the ABBA (α-β-β-α barrel)-type [4,5], the dimethylallyltryptophan synthase (DMATS)-type [6,7], and the membrane-bound type PTases [8,9]. Some of them exhibit broad substrate specificities and accept various aromatic compounds as prenyl acceptors. For example, NphB (also called Orf-2), the first reported ABBA-type PTase in naphthterpin biosynthesis, accepts several aromatic compounds, including dihy-

droxynaphthalenes, flavonoids, and resorcinols as prenyl acceptors, and C_{10} geranyl diphosphate (GPP) as a prenyl donor, to generate a variety of O- or C-prenylated aromatic compounds [4,10]. Some PTases accept multiple lengths of isoprenoid diphosphates, as exemplified by the ABBA-type PTase TleC, which accepts C_5 to C_{20} isoprenoid diphosphates in the biosynthesis of teleocidin B [11]. More interestingly, some PTases change their regiospecificity according to the chain length of the isoprenoid diphosphate, as exemplified by the DMATS-type PTase AtaPT [12]. To get knowledge about the molecular bases and their functions, a number of PTases have been subjected to X-ray crystallographical analyses. It is important to compare the multiple X-ray crystal structures with each substrate for the various reactions, to understand their plasticity. Here, we summarize the molecular basis of the two ABBA type PTases, AmbP1 and AmbP3, which catalyze multiple reactions with different sets of substrates [13,14]. Their plasticities in the reactions were revealed by the X-ray crystal structures of their complexes with different substrates.

Review

Hapalindole/ambiguine biosynthesis

Hapalindole/ambiguine alkaloids, isolated from cyanobacteria, are composed of the total C_{15} prenyl moieties derived from dimethylallyl diphosphate (DMAPP) and geranyl diphosphate (GPP), and *cis*-indolylvinyl isonitrile **1** (Figure 2A) [15]. This natural product family includes structurally diverse compounds with beneficial bioactivities, exemplified by 12-*epi*-hapalindole E isonitrile, which exhibits antibacterial, antifungal, and antimycobacterial activities [16], and ambiguine I, which induces apoptosis and cell-cycle arrest through the inhibition of an NF-κB-related regulation pathway [17]. To investigate their biosyntheses, two research groups independently sequenced the genome of a cyanobacterium, *Fischerella ambigua* UTEX

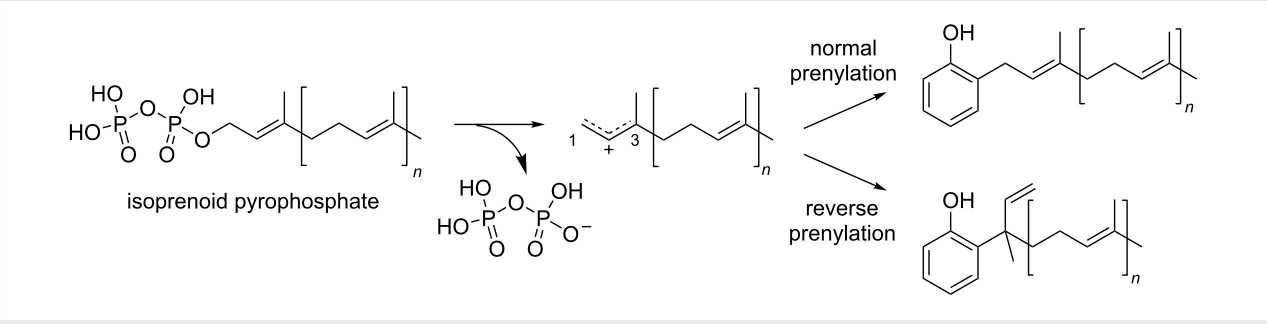


Figure 1: The reactions of aromatic PTases.

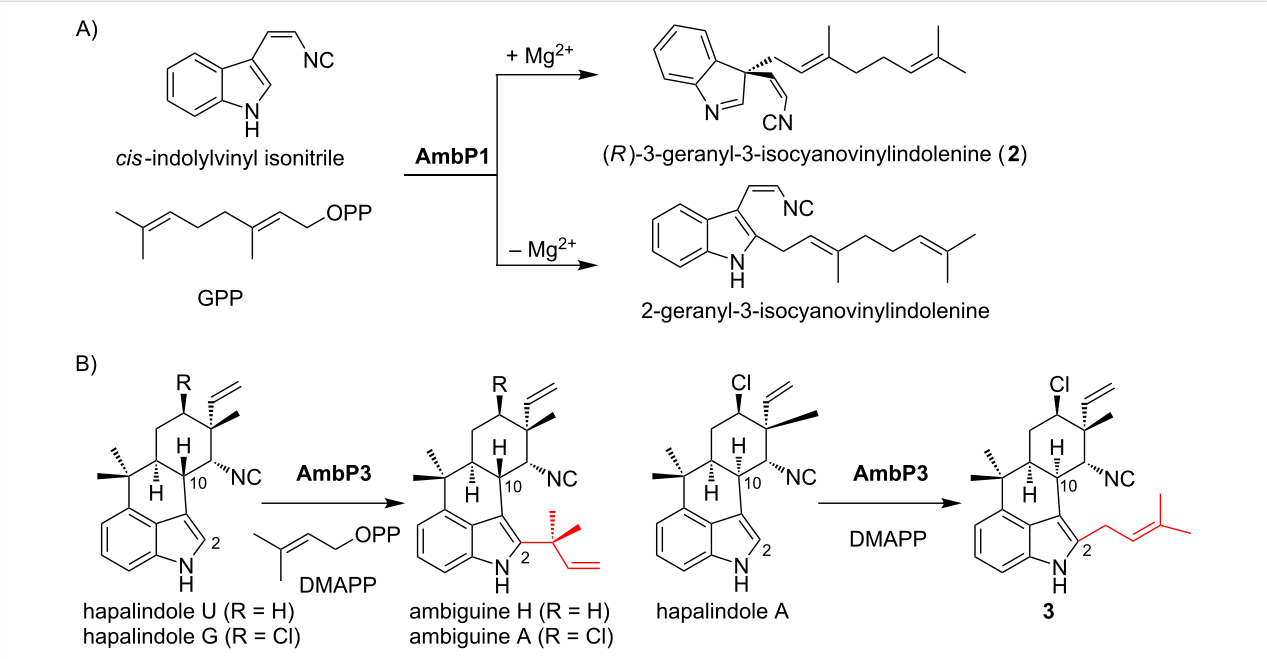


Figure 2: The reactions catalyzed by AmbP1 (A) and AmbP3 (B).

1903, and performed a biosynthetic study [18,19]. Among the identified biosynthetic enzymes, the major contributors of the structural diversity are two prenyltransferases (AmbP1 and AmbP3) [18,20], isomerocyclases [19,21-24], and α-ketoglutarate-dependent oxygenases [25,26]. The prenyltransferase AmbP1 transfers a geranyl group onto C-3 of **1** to yield (*R*)-3-geranyl-3-isocyanovinylindolenine (**2**, Figure 2A) [20]. **2** is cyclized by isomerocyclases to give the hapalindole or fischerindole tetracyclic core structure [19,21-24]. The tetracyclic core is oxidatively halogenated by an α-ketoglutarate-dependent oxygenase [25,26]. Interestingly, AmbP1 also transfers the geranyl group onto the C-2 carbon of **1** to give *cis*-2-geranylindolylvinyl isonitrile, but this undesired side reaction is suppressed in the presence of Mg^{2+} (Figure 2A) [20]. AmbP3 exhibits tolerant substrate specificity with hapalindole substrates [18]. AmbP3 accepts (10*R*)-hapalindole U (HU) and G, and transfers the dimethylallyl group in the reverse prenylation mode to give ambiguine H and A, respectively (Figure 2B). Remarkably, AmbP3 also accepts (10*S*)-hapalindole A (HA), and transfers the dimethylallyl group onto the C-2 carbon of hapalindole A in normal prenylation mode to yield compound **3** (Figure 2B).

X-ray crystal structure analysis of AmbP1

To understand the effect of Mg^{2+} ions on the AmbP1 reaction, an X-ray crystallization analysis was conducted. The *apo* structure of AmbP1 was solved at 2.35 Å, and it adopted an ABBA fold [4,5,13]. Interestingly, the *apo* structure unusually includes the Mg^{2+} ion in a position nearby the β-barrel, stabilized by hydrogen bonding with N41, E63, and D65 (Figure 3A, Mg-1). It might be required for structural integrity, although no mutational study has been performed to support this notion. In order to obtain a structure that is in complex with a substrate, **1** and geranyl *S*-thiodiphosphate (GSPP) were soaked into the crystal of AmbP1 at pH 6.5, which is the same pH as in the reservoir for crystallization. In this structure (Mg^{2+}-free structure), the distance from C-1 of GSPP to C-2 of **1** (Figure 4A, **a**: 3.3 Å) is closer than that to C-3 of **1** (Figure 4A, **b**: 4.6 Å), unexpectedly indicating that this is a model for the C-2 prenylation. Next, the soaking experiment was tested at pH 8. Given that the suppres-

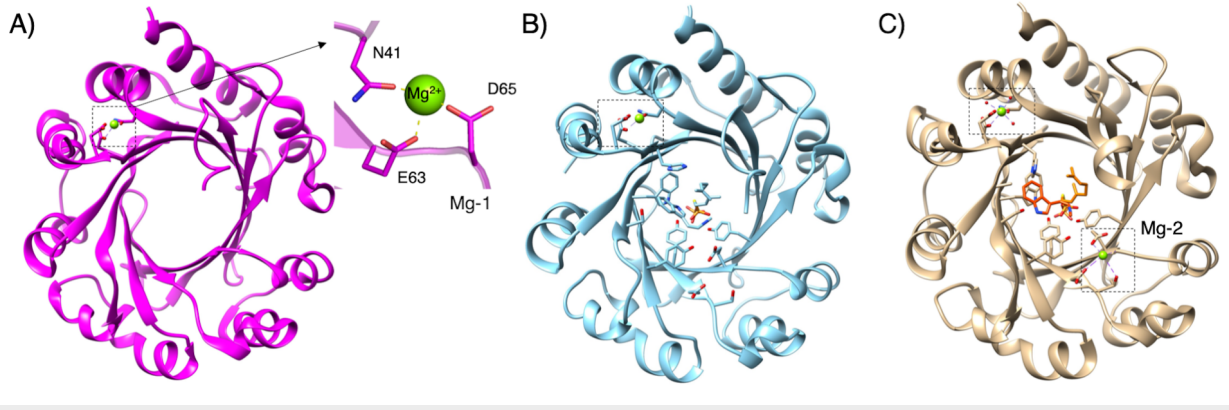


Figure 3: The overall structure of *apo*-AmbP1 (A), the Mg^{2+}-free structure (B), and the Mg^{2+}-bound structure (C).

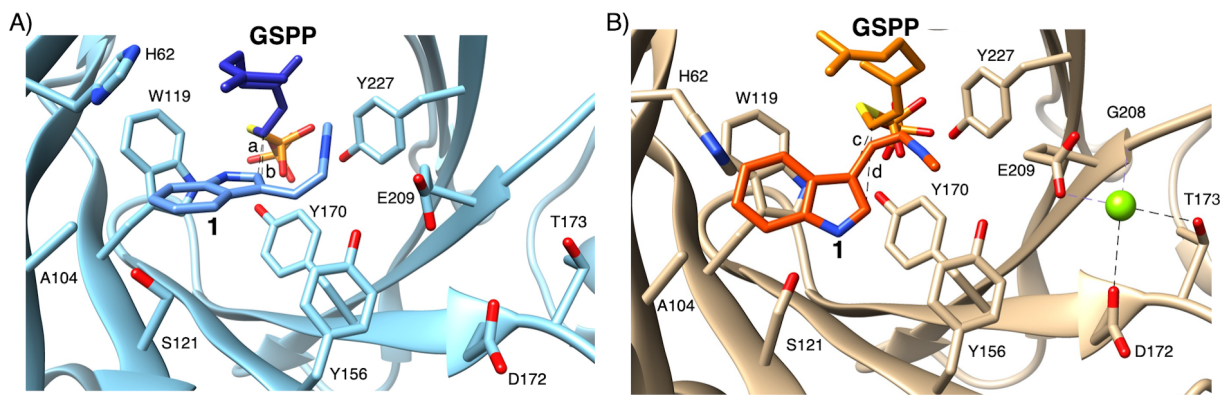


Figure 4: The active site structure of AmbP1. **1** and GSPP were bound in the active site without Mg^{2+} (A, Mg^{2+}-free structure) and with Mg^{2+} (B, Mg^{2+}-bound structure). The green sphere indicates the Mg^{2+} ion.

sion of C-2 prenylation in presence of Mg^{2+} ions is more obvious at pH 8–9 than pH 6 [20], it is expected that this soaking condition will provide the structural model of C-3 prenylation. As expected, the position of **1** dramatically changed, and the distance from C-1 of GSPP to C-3 of **1** (Figure 4B, **c**: 4.6 Å) became closer than that to C-2 of **1** (Figure 4B, **d**: 5.4 Å). Importantly, an additional Mg^{2+} ion (Figure 3C, Mg-2) appeared in the active site close to the isonitrile of **1**, stabilized by the hydrogen bonding with D172, T173, G208, and E209 (Figure 4B). The AmbP1 E209A and E209L mutants completely lost their activities, implying that E209 plays an important role in forming the catalytic cavity as well as binding the Mg^{2+} ion. The active site structure depicted by the surface mode indicated that E209 is important to form the wall of the cavity [13]. More interestingly, the D172A mutation altered the AmbP1 reaction, as it prefers C-3 prenylation even in the presence of Mg^{2+} ions. There are several X-ray crystal structural analyses of PTases that utilize Mg^{2+} as a Lewis acid, such as NphB [4], but this is the first structural analysis of the PTases that utilize a Mg^{2+} ion to reorganise the active site cavity to control the regiospecificity of the prenylation reaction.

X-ray crystal structure analysis of AmbP3

The crystal structures of AmbP3 complexed with DMSPP/hapalindole U (HU structure) and A (HA structure) were each solved at 2.00 Å [14]. The prenyl acceptors, HU and HA, were both surrounded by hydrophobic amino acids, including A44, A102, W117, L119, L259, V284, F288, and M291 [14], and the position of HU was additionally stabilized by hydrogen bonding between the N-1 of HU and E207 (Figure 5A). Remarkably, the terpenoid moieties of HU and HA were located at the same position, but their orientations were completely different. These data indicated that the hydrophobic interaction between the enzyme and the terpenoid moiety is important to support the prenyl acceptor, and the orientation can be altered dependently on their steric structures. The indole of HU, W117, and Y168 formed a cation shield [27,28], which stabilizes the cation intermediate after the removal of the phosphate from DMAPP in the HU structure, and Y225 was substituted with Y168 in the HA structure (Figure 5). The orientation of W117 changed in accordance with the orientation of the indole in HU and HA. As expectedly, W117 was shown to be important for the reaction through a point mutation study, in which W117A and W117F completely lost the catalytic activity. In the HU structure, the distance between C-2 of HU and C-3 of DMSPP (Figure 5A, **a**: 3.6 Å) is shorter than that between C-2 of HU and C-1 of DMSPP (Figure 5A, **b**: 5.4 Å). On the other hand, in the HA structure, the distance between C-2 of HA and C-1 of DMSPP (Figure 5B, **c**: 4.6 Å) is shorter than that between C-2 of HA and C-3 of DMSPP (Figure 5B, **d**: 5.8 Å). These data are consistent with the preference of reverse prenylation on HU and normal prenylation on HA. This is the first X-ray structural model of a PTase that catalyzes both normal and reverse prenylations. The hydrophobic nature of the substrate-binding pocket and the flexibility of the amino acids shielding a cation lead to the plasticity to accept two stereoisomeric hapalindoles and DMAPP, in two different binding poses. This example also illustrates the plasticity of the PTase in the hapalindole biosynthesis. As described above, AtaPT accepts various aromatic compounds as prenyl acceptors, and changes the regiospecificity dependently on the prenyl donor [12]. This plasticity is due to the hydrophobic nature of the prenyl acceptor binding site and the fluctuations of the amino acids forming the cation shield, similarly to AmbP3.

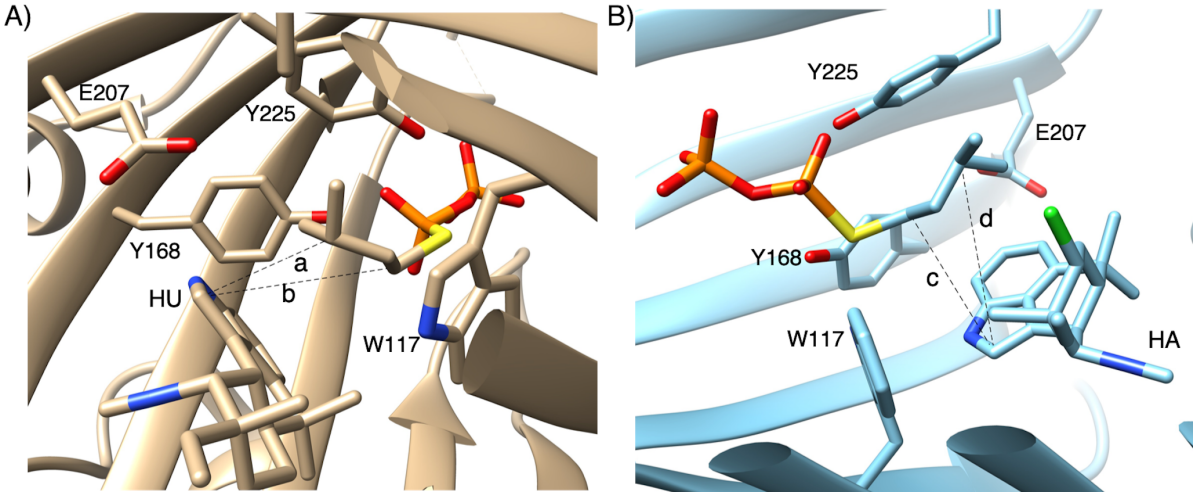


Figure 5: The active site structure of AmbP3 with substrates. The AmbP3 structure in complex with hapalindole U and DMSPP (A), and the AmbP3 structure bound in complex with hapalindole A and DMSPP (B).

Comparison of the AmbP1 and AmbP3 amino acid sequences with other ABBA PTases

The AmbP1 and AmbP3 amino acid sequences were aligned with the other ABBA PTases, including NphB [4], CloQ [28], SCO7190 [10], Fnq26 [29], EpzP [30], NovQ [31], Fur7 [32], SSRG00986 [33], and DzmP [33] (Figure 6). Most of the amino acids involved in pyrophosphate binding, including K117, N168, Y170, R223, Y227, Y276, and K278 in AmbP1, are well conserved among the ABBA family enzymes, but R46 is only

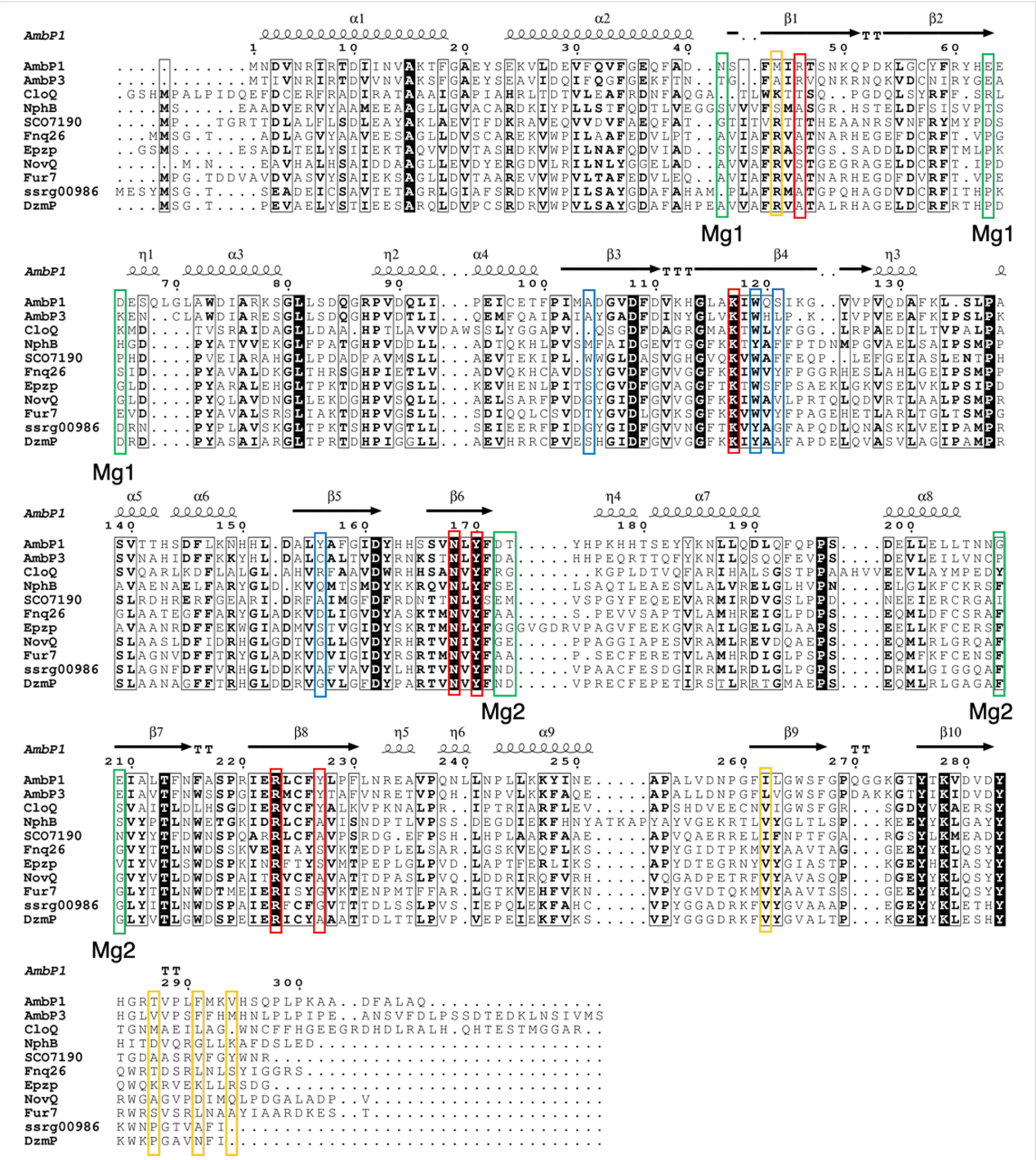


Figure 6: Multiple amino acid sequence alignment of AmbP1, AmbP3, and other ABBA PTases, visualized by ESPript3 [34]. Secondary structure elements: α, α-helices; β, β-strands; η, 3_{10}-helices; TT, strict β-turns. The red frames indicate amino acids that anchor pyrophosphate, the blue frames indicate the residues that support the prenyl acceptor in AmbP1, and the yellow frames indicate the residues that support the prenyl acceptor for AmbP3 that are not conserved in AmbP1, and the green frames indicate the residues that anchor the Mg^{2+} ion.

conserved between AmbP1 and AmbP3, and Y227 is only conserved among AmbP1, AmbP3, and CloQ [28] (Figure 6), indicating that the structures for the pyrophosphate binding pockets in AmbP1 and AmbP3 are slightly different from those of the other enzymes. In fact, the position of the α-phosphate shifts between the Mg^{2+}-free and -bound structures in AmbP1 and between the HU and HA structures in AmbP3, which alters the locations of the substrates in the enzyme. In the AmbP3 structures, Y225 plays an important role to form a cation shield in the HA structure, and the flexibility of Y225 is also an important factor to support both the prenyl acceptor and donor. These observations suggest that the binding mode of the pyrophosphate in AmbP1 and AmbP3 is unique in PTases, and it is likely to be the reason that allows the alternative binding modes of the substrates. The hydrophobic nature of the amino acids that support the prenyl acceptor should be a major factor to allow the alternative substrate binding modes for AmbP1 and AmbP3. Most of the hydrophobic amino acids that support the prenyl acceptor are not conserved among AmbP1, AmbP3, and the other PTases, but W119 is conserved or substituted with tyrosine among all of the aligned PTases (Figure 6). The aromatic amino acids around W119 are also important to support the aromatic substrates and cationic intermediates in the X-ray structural studies of CloQ and EpzP [28,30]. The orientation of W119 significantly changes in accordance with the substrate binding in the HU and HA structures of AmbP3, indicating that the flexible orientation of W119 is also an essential factor for the plasticity. Both two Mg^{2+} binding sites in AmbP1 are not conserved in the other PTases, indicating that they are a unique property of AmbP1. Remarkably, the two Mg^{2+}- binding amino acids are located at the start or end of a β-sheet (Figure 6), which causes the corresponding β-sheet to move through the metal binding. Mg-1 is likely to maintain the overall structure of the enzyme, and Mg-2 defines the shape of the substrate binding site.

Conclusion

The multiple structures of AmbP1 and AmbP3 with different substrate sets provide useful knowledge to understand the molecular basis of the promiscuous PTases. Their promiscuity is mainly caused by the hydrophobic binding pocket for the prenyl acceptor and the flexible positioning of the aromatic residues, which form a cation shield. Although recent progress in chemical synthetic research has established efficient ways to control normal and reverse prenylations with transition metal hydrides [35], it is still important to study the PTases, as they are useful catalysts that control the regiospecificity in an environmentally friendly manner. The information from their X-ray structures will contribute to future engineering of PTases. Furthermore, the structure of AmbP1 can serve as a model to alter the reaction through creating a metal binding site within the PTases, as the natural metalloprotein has been utilized as a model to create an artificial metalloprotein [36,37]. The increasing knowledge obtained from the X-ray structural studies of the PTases will contribute to the development of the enzymology and the chemoenzymatic syntheses of bioactive compounds.

Acknowledgements

This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan (JSPS KAKENHI Grant Number JP16H06443, JP17H04763, and JP18K19139), Japan Science and Technology Agency (JST SICORP Grant Number JPMJSC1701), Suzuken Memorial Foundation, and Kobayashi International Scholarship Foundation. We also thank for a Grant-in-Aid for the Cooperative Research Project from the Institute of Natural Medicine, University of Toyama in 2017.

ORCID® iDs

Takayoshi Awakawa - https://orcid.org/0000-0003-2575-3175
Ikuro Abe - https://orcid.org/0000-0002-3640-888X

References

1. Brandt, W.; Bräuer, L.; Günnewich, N.; Kufka, J.; Rausch, F.; Schulze, D.; Schulze, E.; Weber, R.; Zakharova, S.; Wessjohann, L. *Phytochemistry* **2009,** *70,* 1758–1775. doi:10.1016/j.phytochem.2009.09.001
2. Heide, L. *Curr. Opin. Chem. Biol.* **2009,** *13,* 171–179. doi:10.1016/j.cbpa.2009.02.020
3. Botta, B.; Vitali, A.; Menendez, P.; Misiti, D.; Delle Monache, G. *Curr. Med. Chem.* **2005,** *12,* 713–739. doi:10.2174/0929867053202241
4. Kuzuyama, T.; Noel, J. P.; Richard, S. B. *Nature* **2005,** *435,* 983–987. doi:10.1038/nature03668
5. Saleh, O.; Haagen, Y.; Seeger, K.; Heide, L. *Phytochemistry* **2009,** *70,* 1728–1738. doi:10.1016/j.phytochem.2009.05.009
6. Yu, X.; Li, S.-M. *Methods Enzymol.* **2012,** *516,* 259–278. doi:10.1016/b978-0-12-394291-3.00005-8
7. Rudolf, J. D.; Wang, H.; Poulter, C. D. *J. Am. Chem. Soc.* **2013,** *135,* 1895–1902. doi:10.1021/ja310734n
8. Yazaki, K.; Sasaki, K.; Tsurumaru, Y. *Phytochemistry* **2009,** *70,* 1739–1745. doi:10.1016/j.phytochem.2009.08.023
9. Li, W. *Trends Biochem. Sci.* **2016,** *41,* 356–370. doi:10.1016/j.tibs.2016.01.007
10. Kumano, T.; Richard, S. B.; Noel, J. P.; Nishiyama, M.; Kuzuyama, T. *Bioorg. Med. Chem.* **2008,** *16,* 8117–8126. doi:10.1016/j.bmc.2008.07.052
11. Mori, T.; Zhang, L.; Awakawa, T.; Hoshino, S.; Okada, M.; Morita, H.; Abe, I. *Nat. Commun.* **2016,** *7,* 10849. doi:10.1038/ncomms10849
12. Chen, R.; Gao, B.; Liu, X.; Ruan, F.; Zhang, Y.; Lou, J.; Feng, K.; Wunsch, C.; Li, S.-M.; Dai, J.; Sun, F. *Nat. Chem. Biol.* **2017,** *13,* 226–234. doi:10.1038/nchembio.2263
13. Awakawa, T.; Mori, T.; Nakashima, Y.; Zhai, R.; Wong, C. P.; Hillwig, M. L.; Liu, X.; Abe, I. *Angew. Chem., Int. Ed.* **2018,** *57,* 6810–6813. doi:10.1002/anie.201800855
14. Wong, C. P.; Awakawa, T.; Nakashima, Y.; Mori, T.; Zhu, Q.; Liu, X.; Abe, I. *Angew. Chem., Int. Ed.* **2018,** *57,* 560–563. doi:10.1002/anie.201710682

15. Walton, K.; Berry, J. P. *Mar. Drugs* **2016,** *14,* No. 73. doi:10.3390/md14040073
16. Doan, N. T.; Stewart, P. R.; Smith, G. D. *FEMS Microbiol. Lett.* **2001,** *196,* 135–139. doi:10.1111/j.1574-6968.2001.tb10554.x
17. Acuña, U. M.; Zi, J.; Orjala, J.; Carcache de Blanco, E. J. *Int. J. Cancer Res.* **2015,** *49,* 1655–1662.
18. Hillwig, M. L.; Zhu, Q.; Liu, X. *ACS Chem. Biol.* **2014,** *9,* 372–377. doi:10.1021/cb400681n
19. Li, S.; Lowell, A. N.; Yu, F.; Raveh, A.; Newmister, S. A.; Bair, N.; Schaub, J. M.; Williams, R. M.; Sherman, D. H. *J. Am. Chem. Soc.* **2015,** *137,* 15366–15369. doi:10.1021/jacs.5b10136
20. Liu, X.; Hillwig, M. L.; Koharudin, L. M. I.; Gronenborn, A. M. *Chem. Commun.* **2016,** *52,* 1737–1740. doi:10.1039/c5cc10060g
21. Zhu, Q.; Liu, X. *Angew. Chem., Int. Ed.* **2017,** *56,* 9062–9066. doi:10.1002/anie.201703932
22. Zhu, Q.; Liu, X. *Chem. Commun.* **2017,** *53,* 2826–2829. doi:10.1039/c7cc00782e
23. Li, S.; Lowell, A. N.; Newmister, S. A.; Yu, F.; Williams, R. M.; Sherman, D. H. *Nat. Chem. Biol.* **2017,** *13,* 467–469. doi:10.1038/nchembio.2327
24. Newmister, S. A.; Li, S.; Garcia-Borràs, M.; Sanders, J. N.; Yang, S.; Lowell, A. N.; Yu, F.; Smith, J. L.; Williams, R. M.; Houk, K. N.; Sherman, D. H. *Nat. Chem. Biol.* **2018,** *14,* 345–351. doi:10.1038/s41589-018-0003-x
25. Hillwig, M. L.; Liu, X. *Nat. Chem. Biol.* **2014,** *10,* 921–923. doi:10.1038/nchembio.1625
26. Mitchell, A. J.; Zhu, Q.; Maggiolo, A. O.; Ananth, N. R.; Hillwig, M. L.; Liu, X.; Boal, A. K. *Nat. Chem. Biol.* **2016,** *12,* 636–640. doi:10.1038/nchembio.2112
27. Jost, M.; Zocher, G.; Tarcz, S.; Matuschek, M.; Xie, X.; Li, S.-M.; Stehle, T. *J. Am. Chem. Soc.* **2010,** *132,* 17849–17858. doi:10.1021/ja106817c
28. Metzger, U.; Keller, S.; Stevenson, C. E. M.; Heide, L.; Lawson, D. M. *J. Mol. Biol.* **2010,** *404,* 611–626. doi:10.1016/j.jmb.2010.09.067
29. Haagen, Y.; Unsöld, I.; Westrich, L.; Gust, B.; Richard, S. B.; Noel, J. P.; Heide, L. *FEBS Lett.* **2007,** *581,* 2889–2893. doi:10.1016/j.febslet.2007.05.031
30. Zocher, G.; Saleh, O.; Heim, J. B.; Herbst, D. A.; Heide, L.; Stehle, T. *PLoS One* **2012,** *7,* e48427. doi:10.1371/journal.pone.0048427
31. Ozaki, T.; Mishima, S.; Nishiyama, M.; Kuzuyama, T. *J. Antibiot.* **2009,** *62,* 385–392. doi:10.1038/ja.2009.48
32. Kumano, T.; Tomita, T.; Nishiyama, M.; Kuzuyama, T. *J. Biol. Chem.* **2010,** *285,* 39663–39671. doi:10.1074/jbc.m110.153957
33. Bonitz, T.; Zubeil, F.; Grond, S.; Heide, L. *PLoS One* **2013,** *8,* e85707. doi:10.1371/journal.pone.0085707
34. Robert, X.; Gouet, P. *Nucleic Acids Res.* **2014,** *42,* W320–W324. doi:10.1093/nar/gku316
35. Hu, Y.-C.; Ji, D.-W.; Zhao, C.-Y.; Zheng, H.; Chen, Q.-A. *Angew. Chem., Int. Ed.* **2019,** *58,* 5438–5442. doi:10.1002/anie.201901025
36. Lu, Y.; Berry, S. M.; Pfister, T. D. *Chem. Rev.* **2001,** *101,* 3047–3080. doi:10.1021/cr0000574
37. Schwizer, F.; Okamoto, Y.; Heinisch, T.; Gu, Y.; Pellizzoni, M. M.; Lebrun, V.; Reuter, R.; Köhler, V.; Lewis, J. C.; Ward, T. R. *Chem. Rev.* **2018,** *118,* 142–231. doi:10.1021/acs.chemrev.7b00014

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Archangelolide: A sesquiterpene lactone with immunobiological potential from *Laserpitium archangelica*

Silvie Rimpelová[*1], Michal Jurášek[2], Lucie Peterková[1], Jiří Bejček[1], Vojtěch Spiwok[1], Miloš Majdl[2], Michal Jirásko[3], Miloš Buděšínský[4], Juraj Harmatha[4], Eva Kmoníčková[3,5], Pavel Drašar[*2] and Tomáš Ruml[*1]

Full Research Paper

Open Access

Address:
[1]Department of Biochemistry and Microbiology, University of Chemistry and Technology Prague, Technická 5, 166 28, Prague 6, Czech Republic, [2]Department of Chemistry of Natural Compounds, University of Chemistry and Technology Prague, Technická 5, 166 28, Prague 6, Czech Republic, [3]Charles University in Prague, Faculty of Medicine in Pilsen, 301 66 Pilsen, Czech Republic, [4]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo náměstí 2, 166 10 Prague 6, Czech Republic and [5]Institute of Experimental Medicine, Academy of Sciences of the Czech Republic, v.v.i., 14220 Prague 4, Czech Republic

Email:
Silvie Rimpelová[*] - silvie.rimpelova@vscht.cz; Pavel Drašar[*] - pavel.drasar@vscht.cz; Tomáš Ruml[*] - tomas.ruml@vscht.cz

* Corresponding author

Keywords:
anti-inflammatory properties; archangelolide; dansyl fluorescent conjugate; sarco/endoplasmic reticulum calcium ATPase; sesquiterpene lactone; trilobolide analogue

Beilstein J. Org. Chem. **2019,** *15,* 1933–1944.
doi:10.3762/bjoc.15.189

Received: 31 May 2019
Accepted: 30 July 2019
Published: 13 August 2019

This article is part of the thematic issue "Terpenes".

Guest Editor: J. S. Dickschat

Abstract

Sesquiterpene lactones are secondary plant metabolites with sundry biological effects. In plants, they are synthesized, among others, for pesticidal and antimicrobial effects. Two such compounds, archangelolide and trilobolide of the guaianolide type, are structurally similar to the well-known and clinically tested lactone thapsigargin. While trilobolide has already been studied by us and others, there are only scarce reports on the biological activity of archangelolide. Here we present the preparation of its fluorescent derivative based on a dansyl moiety using azide–alkyne Huisgen cycloaddition having obtained the two sesquiterpene lactones from the seeds of *Laserpitium archangelica* Wulfen using supercritical CO_2 extraction. We show that dansyl-archangelolide localizes in the endoplasmic reticulum of living cells similarly to trilobolide; localization in mitochondria was also detected. This led us to a more detailed study of the anticancer potential of archangelolide. Interestingly, we found that neither archangelolide nor its dansyl conjugate did exhibit cytotoxic effects in contrast to the structurally closely related counterparts trilobolide and thapsigargin. We explain this observation by a molecular dynamics simulation, in which, in contrast to trilobolide, archangelolide did not bind into the sarco/endoplasmic reticular calcium ATPase cavity utilized by thapsigargin. Last, but not least, archangelolide exhibited anti-inflammatory activity, which makes it promising compound for medicinal purposes.

Introduction

Sesquiterpene lactones (SLs) have been attracting interest already for some time due to the plethora of biological effects they elicit. Various SLs show anticancer, antimicrobial, antioxidant, antiprotozoal, antiviral and immunobiological activities (reviewed in [1,2]). Two SLs remarkable for their immunobiological potential are archangelolide (**1**) and trilobolide (**2**), depicted in Figure 1. Compound **2** exhibits strong induction of nitric oxide (NO) in eukaryotic cells, which in turn evokes the synthesis of IL-6, INF-γ and TNF-α (see section "Abbreviations" at the end of the text). Furthermore, compound **2** has a very similar structure to the well-described SL thapsigargin, which is the best-known inhibitor of sarcoplasmic/endoplasmic reticular calcium ATPase (SERCA) with K_i values in the range of nanomoles [3]. The inhibition of SERCA by thapsigargin is stoichiometric and irreversible [4] and results in the depletion of the intracellular calcium storage and an elevated cytosolic calcium concentration, which then can trigger apoptosis in cells. This ability of thapsigargin to trigger programmed cell death prompted the development of mipsagargin/G202 [5], a thapsigargin-derived prodrug that completed phase-I and phase-II clinical trials [6] for several types of solid tumors. The structurally similar compound **2** was found to act by the same mechanism in cells as thapsigargin [7] and was also found to be cytotoxic with localization to the endoplasmic reticulum [8]. The cytotoxicity of compound **2** prompted the preparation and evaluation of compound **2** conjugates with porphyrins [9] and steroids [10] for targeted uptake of compound **2** into cancer cells, and the conjugates showed interesting biological effects including antimycobacterial effects [10]. Nevertheless, cytotoxicity posed a considerable limitation for the utilization of compound **2**. It was, however, soon found that compound **2** may be modified in such a way that its cytotoxicity is reduced while the immunobiological properties are retained [11].

Contrary to compound **2**, compound **1** inhibits NO production and synthesis of IL-1β and INF-γ [12]. Although its structure was already elucidated in the 1970s [13], the information on cytotoxicity of compound **1** is very limited [12] and its intracellular localization and mechanism of action are unknown. Therefore, based on the structural similarity between compound **1** and **2**, we suspected that new interesting and relevant information on this biologically largely undescribed SL could be identified.

In order to study the action of these commercially unavailable compounds, efficient isolation must first be performed in order to gain sufficient amount of pure SLs. Even though the presence of several immuno-active SLs, including compound **1** and **2**, was previously reported in *Laser trilobum* (L.) Borkh and the method of their isolation from petroleum ether extracts of plant roots [12] was described, we aimed to develop a facile method for the isolation of these SLs from the seeds of *Laserpitium archangelica* Wulfen using supercritical CO_2 extraction (SCE; for advantages of this method, see Supporting Information File 1, section 8). In our experience, any part of the plant may be used for extraction of compound **1**, while the roots and seeds give much higher yields than other parts. Utilization of SCE gives improved efficiency of the extraction and shortens the work-up in contrast with classical methods.

By the abovementioned methodology, we obtained sufficient amounts of these compounds for further work including synthetic modifications using azide–alkyne Huisgen cycloaddition. We previously showed the preparation of fluorescent trilobolide conjugates [8] that retained the activity of the parental compounds and proved to be useful for live-cell imaging. In this article, we present a similar approach for compound **1**. Using cancerous, non-transformed, and, also primary cell lines, we evaluated the viability of cells treated with compound **1** and its fluorescent dansyl conjugate, which we then used to study its intracellular localization. Finally, we also determined the anti-inflammatory activity of compound **1** using rat macrophages.

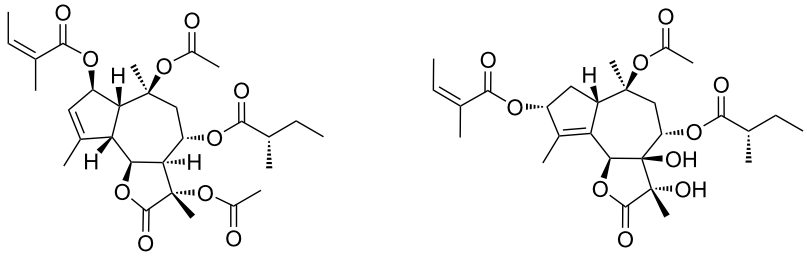


Figure 1: The structure of the sesquiterpene lactones archangelolide (**1**) and trilobolide (**2**).

Results and Discussion

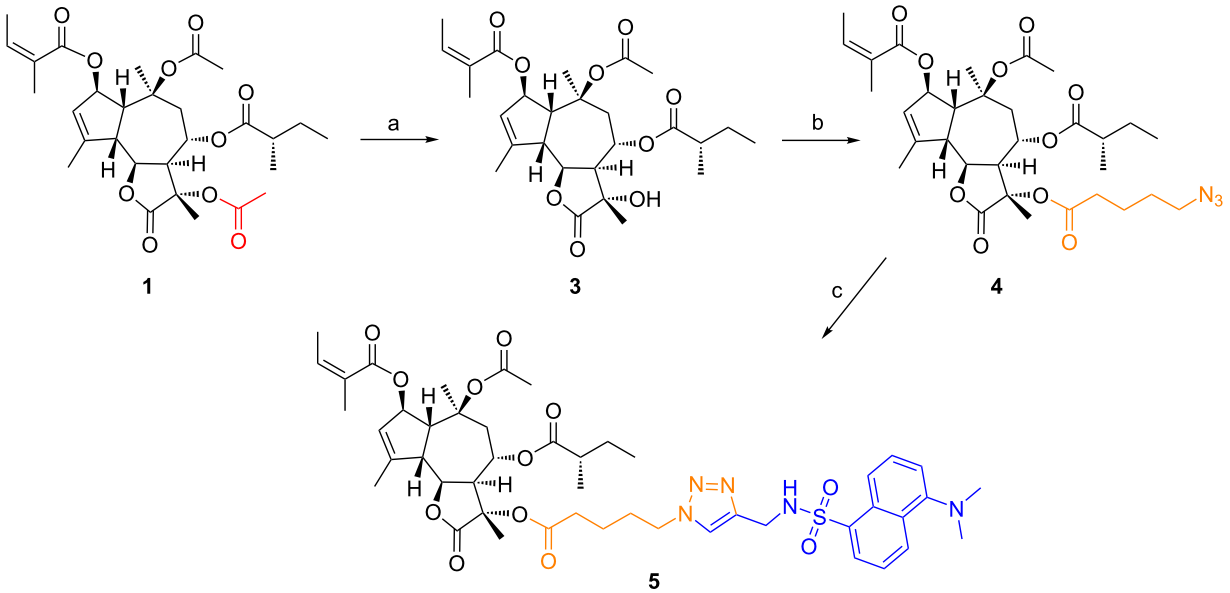
L. archangelica metabolite isolation and identification

For isolation of the major metabolites of *L. archangelica,* we used 100 g of fine ground seeds (Supporting Information File 1, Figure S2) and the method of supercritical CO_2 extraction, which was carried out under pressure of 40 MPa at 40 °C until the extract substances passed from the extracted material (when several subsequent fractions were not adding any addition to the weight to the extract). The obtained yellowish extract (42.5 g) was stored in fridge for a period of 72 h during which the crystals were formed. These were collected by filtration over a frit and washed with hexanes obtaining 7.9 g of the matter that was analyzed by TLC and LRMS analyses. We observed that the matter consisted, in particular, of two terpene-type compounds, compound **1** and **2**, and of β-sitosterol. Using TLC, we found that the mother liquor contained minimal quantities of compound **1** and **2**, but was abundant in β-sitosterol. Therefore, the obtained matter was further re-dissolved in methylene chloride and coated on silica. The purification by column chromatography using a toluene–diethyl ether gradient system (0%→10% of ether) yielded 2.7 and 1.2 g of compound **1** and **2**, respectively. The identity of these two SLs was verified by NMR, HRESIMS and IR analyses (see Supporting Information, section 1). The melting points of the compounds were also in accordance with previously reported data. Compound **1** was crystalized from a mixture of diethyl ether/hexanes (mp 108–111 °C) [14] and compound **2** from a mixture of toluene/Et_2O (mp 190–192 °C) [12].

Derivatization of compound **1** with a dansyl probe

In order to evaluate the intracellular localization and the fate of compound **1** in living cells, we synthesized a blue-emitting derivative employing a dansyl label. This fluorophore is convenient not only for its small size but also for its membrane permeability. Moreover, it was successfully used in other studies of visualizing other natural compounds [15-17].

The synthesis (Scheme 1) of the dansyl-labeled archangelolide **5** started with a mild solvolysis of compound **1** by triethylamine in methanol. Surprisingly, the only product that we obtained after 48 h of treatment was 11-deacetylarchangelolide (**3**), in only 32% yield. The position in the structure of compound **1** where deacylation takes place remarkably differs from that in compound **2** and thapsigargin. This is very likely due to stereochemistry on the lactone ring connection at C7 and the absence of a hydroxy group at the same position. The steric circumstances and the presence of the α-carbonyl group makes the tertiary C11 hydroxy group more acidic and thus a better leaving group in the reaction. Next, the quaternary 11α-hydroxy group was acylated by freshly prepared 5-azidopentanoic anhydride in the presence of 4-DMAP at room temperature. Finally, the azidopentanoate **4** was successively introduced into a click reaction with fluorescent 5-(dimethylamino)-*N*-(prop-2-yn-1-yl)naphthalene-1-sulfonamide (synthesized according to [18]) using CuI and TBTA [19] catalysis. All of the synthesized compounds were thoroughly described by NMR (Supporting Information File 1, Table S1), IR, optical rotation (Support-



Scheme 1: Reagents and conditions: a) MeOH, TEA, 48 h, yield 32%; b) (i) 5-azidopentanoic acid, DCC, DCM, 90 min, rt; (ii) 4-DMAP, DCM, 8 h, rt, 86%; c) dansylated propargylamine, CuI, TBTA, THF, MW, 70 °C, 2 h, yield 64%.

ing Information File 1, section 2) and HRMS (Supporting Information File 1, section 3). All compounds were re-purified on a short silica column prior to testing and afterwards lyophilized from *tert*-butanol. The substances were analyzed by HPLC proving a purity of ≥95% (Supporting Information File 1, section 4). Fluorescent properties of compound **5** were determined in PBS and methanol, the emission maxima corresponded to 484 and 519 nm in these two solvents, respectively (Supporting Information File 1, Figure S17).

Intracellular localization of compound **5**

In this study we aimed to evaluate the biological effects of the poorly described natural compound archangelolide (**1**) and its fluorescent dansyl conjugate **5** in primary, cancer and primary immune cells. First, we analyzed the rate of cell uptake of compound **5** and dansyl amide in U-2 OS (Figure 2 and Supporting Information File 1, Figure S18) and MRC-5 (Supporting Information File 1, Figure S19) cells. In both cell lines, compound **5** localized already after 30 min at 0.5 μM concentration, and the fluorescence intensity increased over a period of 2 h of incubation. In order to identify the intracellular localization of compound **5**, co-localization experiments using endoplasmic reticulum and mitochondrial markers were performed. As it is apparent from Figure 3 and Figure S20 (Supporting Information File 1), compound **5** localized in the endoplasmic reticulum, which is in agreement with the site of localization of another SL, the SERCA inhibitor compound **2** [8]. However, we observed partial co-localization also in mitochondria, in which SERCA is not present. Localization of a control (dansyl fluorophore) was not organelle-specific and the fluorescence in-

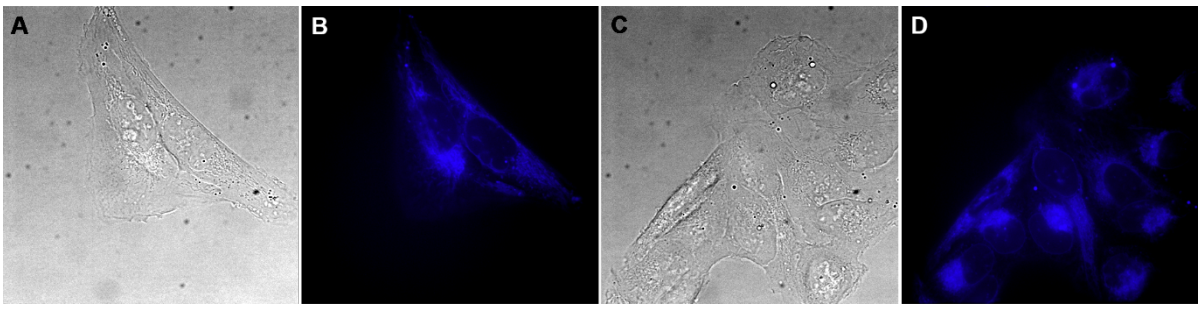


Figure 2: Intracellular localization of archangelolide-dansyl (**5**) in human cells from osteosarcoma (U-2 OS). A, C) Bright-field images; B, D) fluorescence microscopy of living cells treated with 1 μM concentration of compound **5** for 90 min.

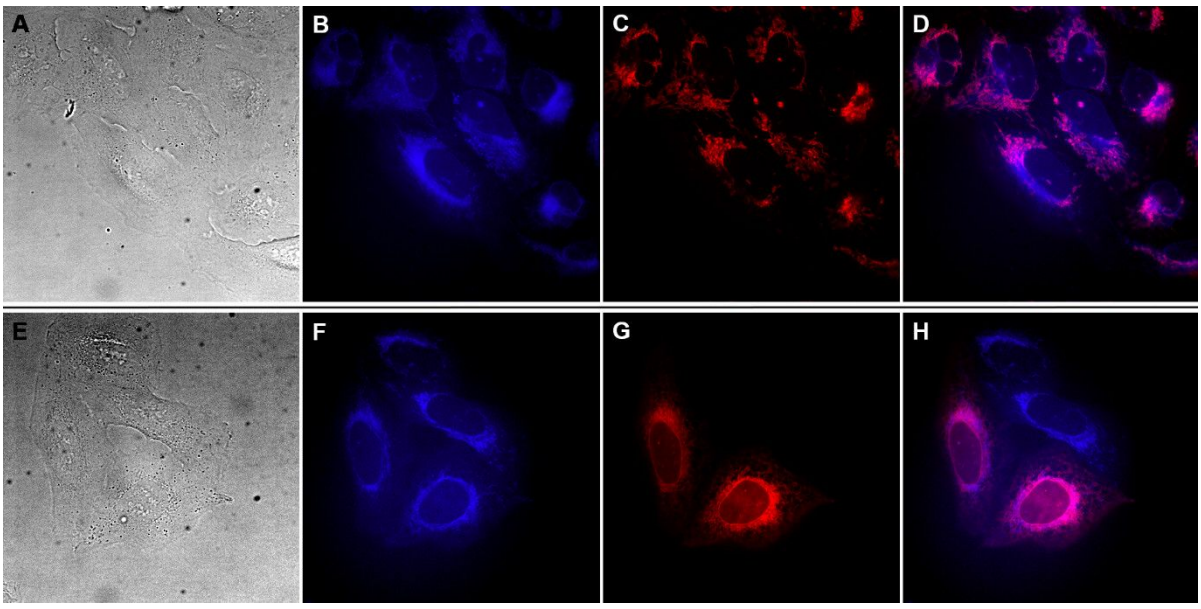


Figure 3: Co-localization of dansylarchangelolide **5** with a marker of endoplasmic reticulum (top row) and with a mitochondrial marker (bottom row) in human cells from osteosarcoma (U-2 OS). A, E) Bright-field images. Fluorescence microscopy of living cells treated with 1 μM concentration of compound **5** (90 min; images B and F) and a mitochondria-specific dye from [20] (10 min; image C) or pDNA coding mCherry-ER (image G). D, H) merged images.

tensity was very weak even at 2 μM concentration (Supporting Information File 1, Figure S18).

Impact of compound 1 on cell viability

In order to reveal whether compound **1** is as potent as thapsigargin or compound **2** in terms of inhibition of cancer cell proliferation, we evaluated its cytotoxicity in a number of cell lines originating from various tumors: prostate, osteosarcoma, breast, colon, pancreas and lung. Cells were treated with compound **1** and its derivatives up to a final concentration of 50 μM for periods of 24, 48 and 72 h. Surprisingly, compound **1** exhibited almost no cytotoxic effect under these conditions (see Supporting Information File 1, Table S2 and Table S3, Figures S21–S23). The IC_{50} values were in the upper micromolar range and up to the tested concentration of 50 μM, 50% cell death was only reached for HEK 293T and LNCaP after 72 h of treatment. Moreover, we also did not detect toxicity of compound **1** in primary and transformed cell lines. Based on these differences from thapsigargin and compound **2**, it seems that this SL does not act as SERCA inhibitor. Therefore, we proceeded to confirm this hypothesis by a molecular dynamics simulation study.

Molecular dynamics simulation of compounds 1 and 2 with SERCA

The binding cavity for thapsigargin and compound **2** in the SERCA protein lies in the transmembrane domain between the helices 3, 5 and 7 [7,21]. Comparing the polarity of these compounds with compound **1** based on the formation of hydrogen bonds, the most polar compound capable of forming the highest number of hydrogen bonds is thapsigargin, the least polar is compound **1** (for details, see Supporting Information File 1, Table S4). Nevertheless, the complex of SERCA with thapsigargin is preferentially stabilized by hydrophobic interactions, [21] and all three mentioned SLs are strongly hydrophobic molecules. We therefore assumed that compounds **1** and **2**, which can form even fewer hydrogen bonds than thapsigargin, may interact with SERCA similarly to thapsigargin. However, this is clearly not the case as represented by our cytotoxicity evaluation using multiple cell lines.

In order to elucidate why compound **1** does not exhibit the same cytotoxicity as its structural counterparts thapsigargin and compound **2**, the potent SERCA inhibitors, manual docking of compound **1** into SERCA followed by MD simulation (1–10 ns) was performed. For this simulation, four different complexes of SERCA and SL were chosen: i) compound **1** positioned correspondingly to the orientation of DTB in the SERCA binding cavity (simulation 1); ii) compound **1** rotated by 180° (simulation 2); iii) compound **2** (for comparison) used as a ligand with the orientation equal to DTB (simulation 3); iv) SERCA positioned in a phospholipid membrane to reflect the fact that it is a transmembrane protein, with compound **2** as a ligand (simulation 4).

The first round of simulations (1–10 ns) of compound **1** docking into SERCA cavity with compound **1** constrained as described (simulation 1) showed the presence of hydrophobic interactions of ligand side chains with Phe256, Val263, Ile829 and the hydrophobic part of Gln259 amino acid residues. Gln259 was the main residue involved in hydrophilic interaction. Interestingly, during the simulations in water environment, compound **1** had the tendency to escape from the SERCA cavity.

In simulation 2 of compound **1** rotated by 180° with constrained positions (10 ns), interactions with Phe256 and hydrophobic parts of Glu255 and Gln259 were detected. When the simulation was performed without constrained positions, compound **1** escaped from the SERCA protein cavity.

After the first simulation of compound **2** (simulation 3) in non-water environment, hydrogen bonds of compound **2** with Glu255 and Gln259 residues, and α-amino group of Ile829 were obvious. The hydrophobic interactions were directed preferentially to Phe256, Val263, Leu260 and Ile829 residues and a hydrophobic part of Lys252. In contrast to simulation 1 with compound **1**, there was no indication of compound **2** escaping from the SERCA cavity (Figure 4 and Figure 5).

During simulations with SERCA positioned in a phospholipid membrane (simulation 4; 1 ns), the free space between individual phospholipids in the cell membrane was filled. SERCA remained stable and tightly embedded in the membrane and did not show any tendency to escape from it. Compound **2** interacted similarly to simulation 3, in which it remained in the thapsigargin SERCA cavity for the duration of the simulation.

When compound **1** was positioned in the same way as compound **2** and DTB during the simulations, it showed a tendency to escape from the SERCA cavity, which was significantly pronounced in the case of compound **1** rotated by 180°. Therefore, we are convinced that especially the orientation and positions of the side chains of compound **1** are very important for its affinity to SERCA or the lack thereof. Winther and co-workers [22] reported that the dissociation constant of thapsigargin derivatives each lacking one of the four thapsigargin side chains (e.g., 2-deoctanoyl-4,5-dihydrothapsigargin or 8-*O*-(dodecanoyl-8-*O*-debutanoyltrilobolide) increases in the following order: *O*-2-deoctanoyl < *O*-8-debutanoyl < *O*-10-deacetyl < *O*-3-deangeloyl. The individual side chains are depicted in Figure 6.

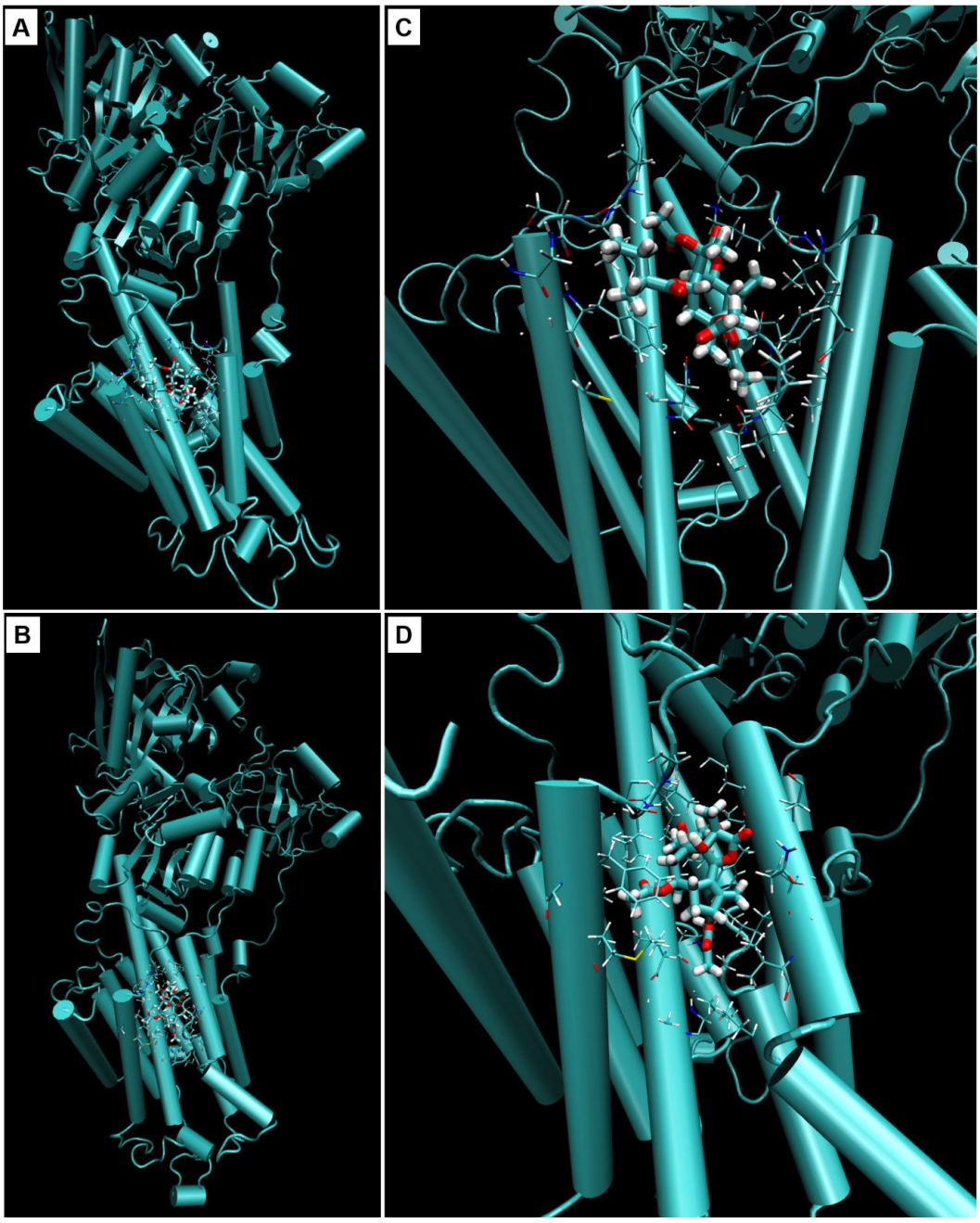


Figure 4: Cartoon representation of sarco/endoplasmic reticulum Ca^{2+} ATPase binding pocket with A, C) archangelolide (**1**) or B, D) trilobolide (**2**) after molecular dynamic simulations. Depicted are also amino acid residues in a range of 5 Å from the respective ligand. The images were created using VMD software, version 1.9.2.

Considering the inhibition potency of SERCA by thapsigargin and compound **2** [7,23] we are not convinced that the missing structural moiety *O*-2-octanoyl (see Figure 6, A) in the case of compound **1** and **2** has a significant influence on the affinity to SERCA. Furthermore, the *O*-8 butanoyl and *O*-10 acetyl moieties (Figure 6, B and C) are not significantly different in compound **1** (presence of a methyl group in the *O*-8 butanoyl side chain) from those of thapsigargin and compound **2**. In contrast, the *O*-3 angeloyl moiety (Figure 6, D) is positioned at C3 in thapsigargin and compound **2**, whereas in compound **1** it is at C2. An *O*-3 deangeloylthapsigargin derivative was previously prepared by Winther and co-workers, who reported that 500–600 times higher concentrations of this derivative than of thapsigargin were needed to achieve 50% inhibition of SERCA, which was the highest of all the derivatives lacking a single side chain [22]. Taking into consideration their result as well as our

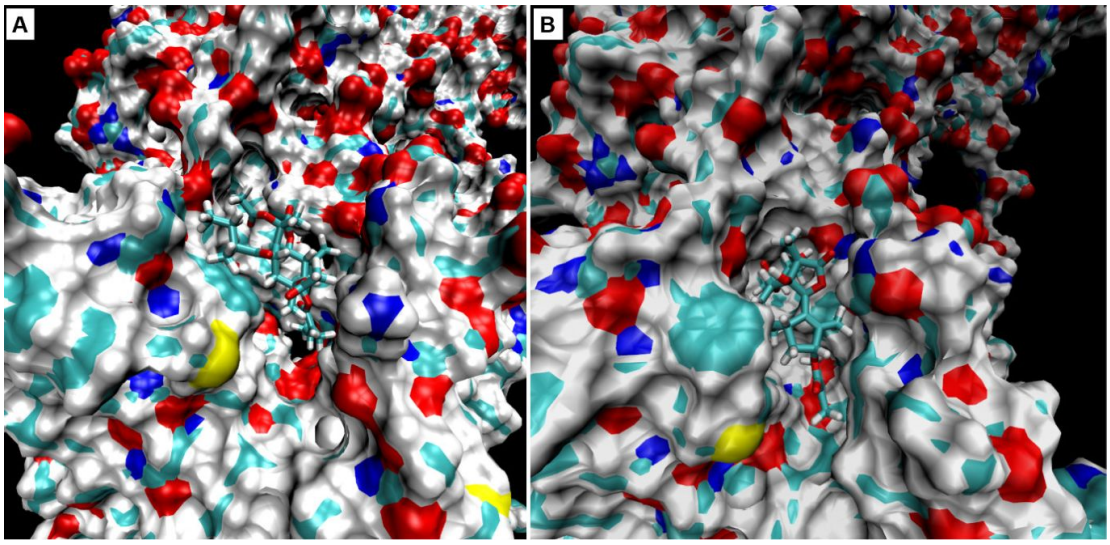


Figure 5: Molecular surface representation of sarco/endoplasmic reticulum Ca^{2+} ATPase binding pocket with A) archangelolide (**1**) and B) trilobolide (**2**) after molecular dynamic simulations. SURF function and probe size 1.4 Å were used. SURF function was written by Amitabh Varshney in University of North Carolina. The images were created using VMD software, version 1.9.2.

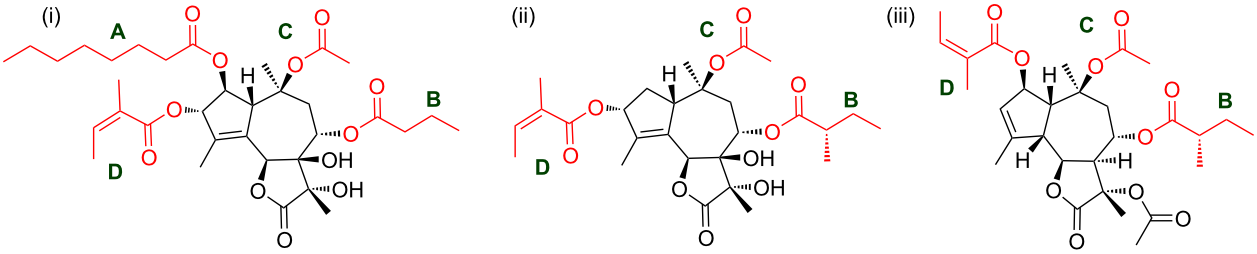


Figure 6: Structural formulae of (i) thapsigargin, (ii) trilobolide (**2**), and (iii) archangelolide (**1**). Red parts show structural moieties of thapsigargin and its derivatives contributing to SERCA binding affinity (according to [22]): A) octanoyl, B) butanoyl or 2-methylbutanoyl, C) acetyl, D) angeloyl.

simulations, we found that compound **1** is unlikely to act as a SERCA inhibitor and, also, it seems that the *O*-3 angeloyl moiety might play a rather significant role in the affinity of thapsigargin/compound **2** to SERCA.

Immunobiological properties of compound 1

SLs, natural compounds predominantly isolated from the species *Asteraceae* and *Apiaceae* represent a rich source of small molecules with potential pharmacological effects. Indeed, some of them have reached clinical applications as antimalarial (artemisinin; [24]) or antitumor (thapsigargin; [25]) agents or have been experimentally investigated with an increasing rate (see PubMed results by year).

In order to assess the immunobiological potential of compound **1**, its cytotoxicity had to be determined first. Therefore, we measured its impact on the viability of rat peritoneal macrophages after 24 h of treatment. No cytotoxicity was detected at tested concentrations of 4 and 40 µM (Figure 7). This is in accordance with results of [12], in which compound **1** and

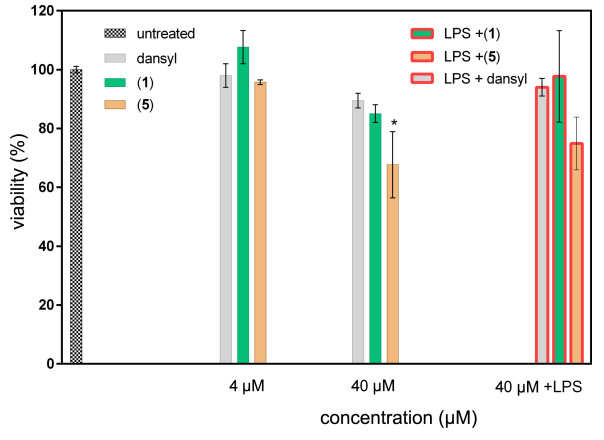


Figure 7: Viability of rat peritoneal cells treated with archangelolide (**1**), dansylarchangelolide **5** and dansyl amide itself. Compounds were applied at 4 µM and 40 µM concentrations and cells were cultured for 24 h. WST-1 assay was used for viability evaluation. The results are expressed as percentage of untreated control ± SEM of n = 6–8 values from two independent experiments. Statistical significance: $^*P < 0.05$, the results of compound **5** are statistically different from those of untreated cells.

2-deangeoylarchangelolide did not exhibit any cytotoxic effect. On the other hand, compound **5** partially decreased the viability of rat peritoneal macrophages to 67% (*$P < 0.05$) at 40 µM concentration. It seems that this phenomenon was not related to the fluorescent dye of compound **5**, since dansyl amide itself did not change the viability of rat macrophages at both concentrations. A decrease in cell viability to 75% (statistically not significant) was also recorded when rat primary macrophages were incubated with 1000 pg·mL^{-1} LPS + 40 µM compound **5**. However, no toxicity was observed for the combination of LPS + dansyl amide at the same concentration.

Immunomodulatory [26] and anti-inflammatory [27] effects of many SLs are well documented. In this study, we investigated the ability of compound **1** and **5** to modulate NO and cytokine TNF-α. We found that when rat primary macrophages were stimulated with 1000 pg·mL^{-1} of LPS, a dose-dependent decrease in NO production was present in cells treated with compound **1** (Figure 8). The decrease in NO was significant (*$P < 0.01$) for 40 µM concentration of compound **1**. The inhibitory effect on NO production was more pronounced in cells treated by compound **5**, with statistics **$P < 0.001$ for both concentrations, 10 and 40 µM. It is possible that the decline in NO at 40 µM concentration of compound **5** was at least partially induced by reduced cell viability. On the other hand, dansyl amide alone did not lower NO production.

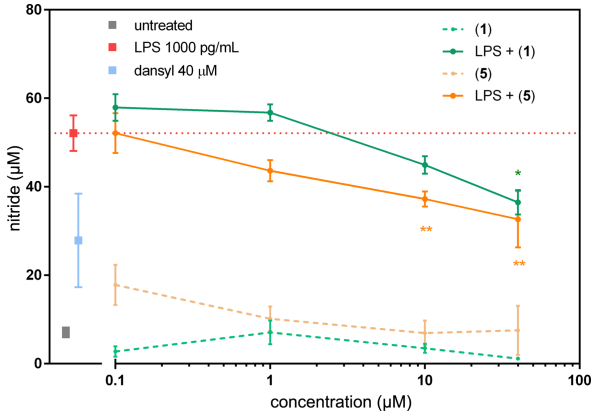


Figure 8: NO production in primary rat macrophages. The cells were treated with archangelolide (**1**) and dansylarchangelolide **5** in the concentration range of 0.1–40 µM for 24 h with or without lipopolysaccharide (LPS, 1000 pg·mL^{-1}) or with solely 40 µM dansyl amide. The results represent the mean ± SEM of three independent experiments, $n = 6$. Statistical significance: * $P < 0.01$, ** $P < 0.001$, the results of the compounds are statistically different from those of the LPS-treated cells.

In contrast to other SLs such as thapsigargin [28] or compound **2** [29] and its derivatives [9], compound **1** did not exhibit immunostimulatory activity, which was confirmed in this study. On the other hand, a slight to moderate effect on the decrease in NO production is clear and corresponds to the result of [12]. Thus, we show the biological activity of a semi-synthetic conjugate of compound **1** as inhibitor of NO production at very low micromolar concentrations.

Furthermore, the anti-inflammatory activity of compounds, which is usually demonstrated by inhibitory activity against TNF-α secretion, has never been explored for compound **1**. Therefore, in this study, we examined the effect of compound **1** and **5** on TNF-α secretion in rat primary macrophages (Figure 9). We found a mild decrease in TNF-α secretion in LPS-treated cells for both compounds. We did not detect any effect of dansyl amide or dansyl amide + LPS (data not shown) on cytokine levels. Previously, weak anti-inflammatory activity was also found for cytokine IL-6 and negligible inhibition of Il-1β was detected for compound **1** and 2-deangeolylarchangelolide [12], whereas a strong anti-inflammatory effect was found for another SL, laserolide [12]. Laserolide (Figure 10) is a germacrane-type SL that has a ten-membered ring adjoining the five-membered lactone ring whereas compound **1** belongs to the guaiane type. In guaianolides, a bond is present between C1 and C5 of the ten-membered ring, creating a seven-membered and a five-membered ring. Nevertheless, the

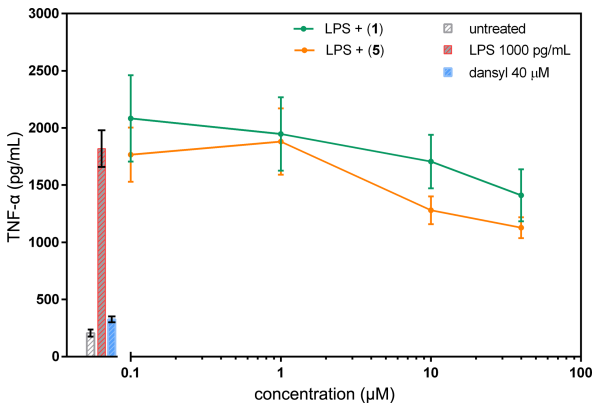


Figure 9: Evaluation of cytokine TNF-α secretion in rat peritoneal cells. Stimulation of primary cells was induced by 1000 pg·mL^{-1} of LPS. Cells were cultured in the presence of archangelolide (**1**) and dansylarchangelolide **5** for 24 h. Cytokine secretion was detected by ELISA. The data are the means ± SEM of two independent experiments, $n = 4$.

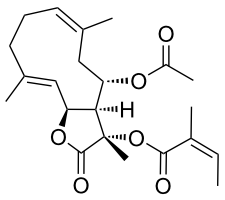


Figure 10: Structure of laserolide.

arrangement of the five-membered lactone ring to the ten-membered part of the molecule with regards to adjoining substituents is similar in both compound **1** and laserolide. This implies that the arrangement of the five-membered ring in a SL molecule towards the rest of the molecule may be related to its anti-inflammatory activity.

Conclusion

Having developed a method for the isolation of two related SLs, archangelolide (**1**) and trilobolide (**2**), from the seeds of *Laserpitium archangelica* Wulfen, we extended the knowledge on the first one, so far almost undescribed. We revealed that despite a high degree of structural similarity with two cytotoxic SLs, compound **2** and thapsigargin, compound **1** exhibited cytotoxicity neither in the tested cancer cell lines nor in primary cells. To explain the reason for different cytotoxicity of compound **1** compared to compound **2** and thapsigargin, we prepared its fluorescent derivative for live-cell imaging studies, which determined its intracellular localization in endoplasmic reticulum and mitochondria of both cancer and normal cells. This led us to examine the binding of compound **1** to SERCA, the target of compound **2** and thapsigargin, by a docking study. Indeed, the results strongly argue that compound **1** is biologically very distinct from the other two SLs, since it does not bind and probably also not inhibit SERCA. Finally, compound **1** exhibited anti-inflammatory activity, which was demonstrated by a decrease in NO production and TNF-α secretion in rat primary macrophages. This makes compound **1** an interesting SL for further study of its immunobiological activity.

Experimental

Source of the natural material

Fresh seeds of *Laserpitium archangelica* Wulfen were provided by the Farmacognostic garden in Poznan (Mazowiecka Street) of the Medical University in Poznan (Poland). The plant was identified by a research specialists of the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic. A voucher specimen has also been deposited at the Department of Natural Compounds at the Faculty of Food and Biochemical Technology, University of Chemistry and Technology, Prague (CZ). Compound **2** was obtained from SciTech spol. s r. o. Prague (CZ).

Synthesis

The synthesis is depicted in Scheme 1.

11-Deacetylarchangelolide (**3**)

Compound **1** (500 mg, 0.91 mmol) was dissolved in freshly distilled MeOH (10 mL) and TEA (2.5 mL) was added dropwise over 1 h period. The mixture was stirred for 48 h at room temperature (ca. 23 °C); then, the solvents were removed under reduced pressure and the product was chromatographed twice (hexanes/AcOEt, gradient 5:1→3:1) to obtain 11-deacetylarchangelolide (148 mg, 0.29 mmol) in 32% yield. Unreacted compound **1** (256 mg) was recovered. The yield based on recovered starting material was 66%. R_f(**1**) = 0.6; R_f(**3**) = 0.4 in hexanes/AcOEt, 3:1. For analytical data, see Supporting Information File 1, sections 2–4.

Synthesis of 11-azidovaleroyl-11-deacetylarchangelolide (**4**)

5-Azidopentanoic acid (80 mg, 0.56 mmol) was dissolved in dry CH_2Cl_2 (3 mL), then DCC (58 mg, 0.28 mmol) was added and the mixture was stirred for 90 min at room temperature (ca. 23 °C) after which the DCU was filtered out. 11α-Deacetylarchangelolide (60 mg, 0.12 mmol) and 4-DMAP (18 mg, 0.14 mmol) were dissolved in CH_2Cl_2 (3 mL) and freshly prepared anhydride of 5-azidopentanoic acid was added to the stirred solution. The mixture was stirred at room temperature for 8 h, after which it was filtered and the solvents were removed under reduced pressure. The residue was chromatographed twice (toluene/Et_2O, 10:1) to obtain the final product (66 mg, 0.1 mmol) in 86% yield with R_f = 0.75 in hexanes/AcOEt, 3:1. For analytical data, see Supporting Information File 1, sections 2–4.

Synthesis of 11-(5-(4-((5-(dimethylamino)naphthalene-1-sulfonamido)methyl)-1*H*-1,2,3-triazol-1-yl)pentanoylarchangelolide (**5**)

To a solution of azidovalerate **4** (40 mg, 0.063 mmol) and 5-(dimethylamino)-*N*-(prop-2-yn-1-yl)naphthalene-1-sulfonamide (20 mg, 0.07 mmol) in THF (3 mL), CuI (50 μL of 1 M water solution) and TBTA (3 mg, 5.6 μmol) were added. The mixture was placed into a microwave reactor and irradiated for 2 h at 70 °C. Then, the solvent was evaporated under reduced pressure and the residue was chromatographed (hexanes/AcOEt, 1:1). The obtained product **5** was re-chromatographed twice to obtain pure product (37 mg, 0.04 mmol) as a slightly yellowish solid in 64% yield with R_f = 0.25 in hexanes/AcOEt, 1:1. For analytical data, see Supporting Information File 1, sections 2–4.

Computational studies

MD Simulations of compound **1** and **2** with sarco/endoplasmic reticular calcium ATPase

For docking and following simulations, a structure of sarco/endoplasmic reticular calcium ATPase (SERCA) protein bound to 8-*O*-(dodecanoyl-8-*O*-debutanoyltrilobolide) [DTB] with 3NAL code was obtained from Protein Data Bank [22]. The structure of compound **1** was drawn using MarvinSketch 6.006 software (ChemAxon Ltd.), the structure of compound **2** was obtained by editing the DTB structure.

First, we computed General Amber Force Field parameters for molecules of compound **1** and **2** in Antechamber software. Charges were calculated by restrained electrostatic potential method (RESP) based on a wave function calculated using quantum chemistry at the HF/6-31G*//HF/6-31G* level. Compound **1** molecule was manually docked into the SERCA binding site for thapsigargin using UCSF CHIMERA 1.10.2 (University of California, San Francisco) software by two approaches. In the first one, the molecule was docked so that the spatial orientation of the central seven-membered ring reached the best agreement with the orientation of DTB. In the other approach, the docked molecule was rotated over by 180°. The following simulations proceeded with the same parameters independently of each other.

The computed complex was transformed into a box with the size of 10.309 × 10.111 × 15.063 nm^3 with periodical boundary conditions and centered. For simulations of energy minimization and molecular dynamics, Gromacs-4.5.5 software was used. The first part of the simulations occurred in vacuum, the other then in water. In order to keep the solution neutral, 24 molecules of water were randomly replaced by sodium ions. The system was minimized and equilibrated by series of restrained simulations (2.02 ns in total).

Simulation of SERCA in a phospholipid bilayer

In order to build a complex of SERCA protein with a phospholipid membrane, we used UCFS Chimera 1.10.2 software (University of California, San Francisco). The membrane composition was as follows: 10% phosphatidylserine, 30% phosphatidylethanolamine and 60% phosphatidylcholine. The membrane itself was created using CHARMM-GUI c40b1 software. Then, phospholipids at any distance greater than 0.35 nm from the enzyme were removed. The system was simulated in a periodic box of the sized of 15.65 × 15.65 × 18.00 nm^3. The system was minimized and equilibrated by series of restrained simulations (3.02 ns in total).

Biological assays

Cell lines and their cultivation

In this study, the following human cancer cell lines were used: LNCaP and PC-3 (prostate carcinoma), U-2 OS (osteosarcoma), MCF-7 (breast carcinoma), HT-29 (colon carcinoma), MIA PaCa-2 (pancreatic carcinoma), A549 (lung carcinoma); one transformed human cell line HEK 293T (embryonic kidney cells); one mouse cell line C2C12 (myoblasts) and a primary cell line MRC-5 (lung fibroblasts). The cell lines were purchased from American Type Culture Collection (ATCC, Manassas, USA) and from Sigma-Aldrich, USA. Unless otherwise specified, cells were cultured in medium (Thermo Fisher, USA) recommended by ATCC with stable glutamine dipeptide and supplemented with 10% fetal bovine serum (FBS; Thermo Fisher, USA). Cells were maintained at exponential phase of growth at 37 °C in humidified atmosphere with 5% CO_2.

Cell viability assay

Viability of cells treated with the tested compounds was determined using WST-1 (Sigma-Aldrich, USA) assay by spectrophotometric detection (450 nm, reference 630 nm) as described in [10]. Briefly, WST-1 assay is based on reduction of a tetrazolium salt on soluble formazan in metabolically active cells. The measured absorbance is directly proportional to the number of metabolically active cells.

For this assay, cells were seeded into individual wells of 96-well plates (5000 cells per well) in 100 µL of cell culture medium supplemented with 10% FBS. The cells were incubated overnight (16 h) under standard cultivation conditions. Then, 100 µL of fresh media was added with the tested compounds (final concentration 0–50 µM) and, the cells were incubated for another 24, 48 and 72 h. Next, the medium was removed and the cells were incubated with 5 µL of WST-1 dissolved in 100 µL of high glucose DMEM without phenol red with 10% FBS for 2 h. Then, the absorbance was measured. Cells incubated with a vehicle (DMSO) in medium were used as control. All experiments were done in quadruplicates. The data were analyzed in Microsoft Excel; the deviations were calculated as standard error of the mean (SEM).

For primary peritoneal cells, a slightly modified procedure was used: the number of $1{\cdot}10^5$ cells per well was cultivated overnight in triplicates, treated with the tested compounds and incubated for another 24 h. The amount of WST-1 was 10 µL per 100 µL of media and the incubation took 3 h. The results are expressed as the percentage of cytotoxicity relative to 100% of dead cells treated with 1% Triton X-100.

Cell uptake study

U-2 OS and MRC-5 cells ($1{\cdot}10^5$ cells per well) were seeded on 35 mm glass bottom dishes for live-cell imaging (MatTek Corporation, USA) and left to adhere for 24 h. Then, the cells were washed with phosphate-buffered saline (PBS) and incubated with compound **5** or dansyl amide fluorophore (0.1–2.0 µM) dissolved in FluoroBrite DMEM medium (Thermo Fisher, USA) for 0.5–2.5 h. After the incubation period, the cells were washed with PBS and fresh FluoroBrite DMEM was added.

Determination of localization in cell organelles

U-2 OS and MRC-5 cells were seeded and influenced by the tested compounds as described in section "Cell uptake study". To assess the intracellular localization of compound **5** in

MRC-5 cells, an endoplasmic reticulum marker ER-Tracker™ Red (120 nM, 30 min; ThermoFisher Scientific, USA) and a mitochondria-specific dye (70 nM, 30 min; UCT Prague, CZ) from [20] were used. In U-2 OS cells, the endoplasmic reticulum was visualized by transfection with a plasmid DNA coding mCherry-ER (0.5 μg) using Fugene HD (Promega, USA; DNA/Fugene HD = 1:3).

Fluorescence microscopy

The intracellular localization of compound **5** was studied by real-time live-cell fluorescence microscopy using an inverse fluorescence microscope Olympus IX-81 operated by xCellence software (Olympus, Japan) and equipped with a high-stability 150 W xenon arc burner and EM-CCD camera C9100-02 (Hamamatsu, Germany). A 60× oil immersion objective (Olympus, Japan) with a numerical aperture of 1.4 was used. All images were deconvolved and background-corrected by xCellence software.

Animals and cells

For isolation of resident peritoneal macrophages, female Wistar rats weighing 175–185 g (VELAZ, Czech Republic) were used. They were kept in plastic cages under standard conditions, i.e., a temperature of 22 ± 2 °C, 12/12 h of light/dark cycle, relative humidity (50 ± 10%) and standard pelleted diet and water were provided ad libitum. Procedures were approved by Institutional protocol MSMT-15894/2013-310. Rats were sacrificed by cervical dislocation. The resident peritoneal macrophages of individual rats were collected by a lavage using sterile saline. The cells were washed, resuspended and seeded into 96-well round-bottom microplates ($2{\cdot}10^5$ cells per well) in complete RPMI-1640 medium containing 10% heat-inactivated FBS, 2 mM L-glutamine, 50 $\mu g{\cdot}mL^{-1}$ gentamicin, and $5{\cdot}10^{-5}$ M 2-mercaptoethanol (all Sigma-Aldrich, St. Louis, MO, USA). The treated or untreated cells (DMSO as a vehicle) were cultured for 24 h and plates were maintained at 37 °C, 5% CO_2, in a humidified Heraeus incubator. Stock solutions of compound **1** and **5** were prepared 100 mM in DMSO. For the cultivation of macrophages and assessment of toxicity and NO level, the majority of chemicals was purchased from Sigma-Aldrich and Merck-Sigma, USA.

Nitrite oxide production in primary rat macrophages

To evaluate immunomodulatory activity of compound **1** and its derivatives, 100 $pg{\cdot}mL^{-1}$ of lipopolysaccharide (LPS) was applied to cells in appropriate wells. After 24 h, the concentration of nitrites was detected in individual supernatants (50 μL) incubated for 10 min at ambient temperature with an aliquot of the Griess reagent (1% sulphanilamide/0.1% naphthylendiamine/2.5% H_3PO_4). The absorbance was recorded at 540 nm using a microplate spectrophotometer (Tecan, Austria). A nitrite calibration curve was used to convert absorbance to concentration (μM) of nitrite.

Cytokine assay

Analogous to NO detection, supernatants of all samples were analyzed for cytokine content after 24 h. The concentration of secreted TNF-α was determined by ELISA assay (R&D Systems, Abingdon, UK) following manufacturer protocol.

Statistical analysis

Analysis of variance (ANOVA) and graphical presentation of data were done using GraphPad Prism 6.05, San Diego, CA.

Abbreviations

4-DMAP, 4-(*N*,*N*-dimethylamino)pyridine; AcOEt, ethyl acetate; DCC, dicyclohexylcarbodiimide; DCU, dicyclohexylurea; Dns, dansyl amide; DMAP, 4-dimethylaminopyridine; DTB, 8-*O*-(dodecanoyl-8-*O*-debutanoyltrilobolid); IL-6, interleukin 6; IL-1β, interleukin 1β; INF-γ, interferon gamma; TBTA, tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine; TEA, triethylamine; THF, tetrahydrofuran; TLC, thin-layer chromatography; TNF-α, tumor necrosis factor alpha; SCE, supercritical CO_2 extraction; SL, sesquiterpene lactones.

Comment on nomenclature

This article uses semitrivial terpene nomenclature and hence the numbering of compounds atoms may differ from these obtained by using IUPAC names.

Supporting Information

Supporting Information File 1

Additional experimental data.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-189-S1.pdf]

Acknowledgements

This work was supported by the Ministry of Health of the CR [15-33018A]; by Martina Roeselová Foundation, Czech Republic; L'Oréal–UNESCO for Women in Science 2019, specific university research [MSMT No 21-SVV/2019], SVV No. 260393, the Czech Science Foundation 14-04329S, and projects [OPPC CZ.2.16/3.1.00/24503]; [LO1601]; [LO1304] and by OP VaVpI project „Improving the quality of laboratory education of students at UCT Prague [KvaLab; 1.05/4.1.00/16.0349]. Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the program "Projects of Large Research, Development, and Innovations Infrastructures.

ORCID® iDs

Silvie Rimpelová - https://orcid.org/0000-0002-3008-1396
Michal Jurášek - https://orcid.org/0000-0002-5069-9716
Michal Jirásko - https://orcid.org/0000-0002-4874-7723
Juraj Harmatha - https://orcid.org/0000-0003-4385-0841
Eva Kmoníčková - https://orcid.org/0000-0002-0940-4605
Pavel Drašar - https://orcid.org/0000-0003-0093-7007

Preprint

A non-peer-reviewed version of this article has been previously published as a preprint doi:10.3762/bxiv.2019.38.v1

References

1. Chadwick, M.; Trewin, H.; Gawthrop, F.; Wagstaff, C. *Int. J. Mol. Sci.* **2013,** *14,* 12780–12805. doi:10.3390/ijms140612780
2. Peterková, L.; Rimpelová, S.; Kmoníčková, E.; Ruml, T. *Chem. Listy* **2019,** *113,* 149–155.
3. Wootton, L. L.; Michelangeli, F. *J. Biol. Chem.* **2006,** *281,* 6970–6976. doi:10.1074/jbc.m510978200
4. Sagara, Y.; Fernandez-Belda, F.; de Meis, L.; Inesi, G. *J. Biol. Chem.* **1992,** *267,* 12606–12613.
5. Quynh Doan, N.; Christensen, S. *Curr. Pharm. Des.* **2015,** *21,* 5501–5517. doi:10.2174/1381612821666151002112824
6. Mahalingam, D.; Peguero, J.; Cen, P.; Arora, S. P.; Sarantopoulos, J.; Rowe, J.; Allgood, V.; Tubb, B.; Campos, L. *Cancers* **2019,** *11,* 833. doi:10.3390/cancers11060833
7. Wictome, M.; Khan, Y. M.; East, J. M.; Lee, A. G. *Biochem. J.* **1995,** *310,* 859–868. doi:10.1042/bj3100859
8. Jurášek, M.; Rimpelová, S.; Kmoníčková, E.; Drašar, P.; Ruml, T. *J. Med. Chem.* **2014,** *57,* 7947–7954. doi:10.1021/jm500690j
9. Tomanová, P.; Rimpelová, S.; Jurášek, M.; Buděšínský, M.; Vejvodová, L.; Ruml, T.; Kmoníčková, E.; Drašar, P. B. *Steroids* **2015,** *97,* 8–12. doi:10.1016/j.steroids.2014.08.024
10. Jurášek, M.; Džubák, P.; Rimpelová, S.; Sedlák, D.; Konečný, P.; Frydrych, I.; Gurská, S.; Hajdúch, M.; Bogdanová, K.; Kolář, M.; Müller, T.; Kmoníčková, E.; Ruml, T.; Harmatha, J.; Drašar, P. B. *Steroids* **2017,** *117,* 97–104. doi:10.1016/j.steroids.2016.08.011
11. Harmatha, J.; Buděšínský, M.; Jurášek, M.; Zimmermann, T.; Drašar, P.; Zídek, Z.; Kmoníčková, E.; Vejvodová, L. *Fitoterapia* **2019,** *134,* 88–95. doi:10.1016/j.fitote.2019.02.002
12. Harmatha, J.; Buděšínský, M.; Vokáč, K.; Kostecká, P.; Kmoníčková, E.; Zídek, Z. *Fitoterapia* **2013,** *89,* 157–166. doi:10.1016/j.fitote.2013.05.025
13. Holub, M.; Samek, Z. *Collect. Czech. Chem. Commun.* **1973,** *38,* 731–738. doi:10.1135/cccc19730731
14. Smítalová, Z.; Buděšínský, M.; Šaman, D.; Holub, M. *Collect. Czech. Chem. Commun.* **1986,** *51,* 1323–1339. doi:10.1135/cccc19861323
15. Wulff, J. E.; Siegrist, R.; Myers, A. G. *J. Am. Chem. Soc.* **2007,** *129,* 14444–14451. doi:10.1021/ja075327f
16. Liu, Y.; Lok, C.-N.; Ko, B. C.-B.; Shum, T. Y.-T.; Wong, M.-K.; Che, C.-M. *Org. Lett.* **2010,** *12,* 1420–1423. doi:10.1021/ol902890j
17. Kim, M.; Kleckley, T. S.; Wiemer, A. J.; Holstein, S. A.; Hohl, R. J.; Wiemer, D. F. *J. Org. Chem.* **2004,** *69,* 8186–8193. doi:10.1021/jo049101w
18. Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, J. C.; Schultz, P. G. *J. Am. Chem. Soc.* **2003,** *125,* 11782–11783. doi:10.1021/ja0370037
19. Hein, J. E.; Tripp, J. C.; Krasnova, L. B.; Sharpless, K. B.; Fokin, V. V. *Angew. Chem., Int. Ed.* **2009,** *48,* 8018–8021. doi:10.1002/anie.200903558
20. Rimpelová, S.; Bříza, T.; Králová, J.; Záruba, K.; Kejík, Z.; Císařová, I.; Martásek, P.; Ruml, T.; Král, V. *Bioconjugate Chem.* **2013,** *24,* 1445–1454. doi:10.1021/bc400291f
21. Paula, S.; Ball, W. J., Jr. *Proteins: Struct., Funct., Bioinf.* **2004,** *56,* 595–606. doi:10.1002/prot.20105
22. Winther, A.-M. L.; Liu, H.; Sonntag, Y.; Olesen, C.; le Maire, M.; Soehoel, H.; Olsen, C.-E.; Christensen, S. B.; Nissen, P.; Møller, J. V. *J. Biol. Chem.* **2010,** *285,* 28883–28892. doi:10.1074/jbc.m110.136242
23. Furuya, Y.; Lundmo, P.; Short, A. D.; Gill, D. L.; Isaacs, J. T. *Cancer Res.* **1994,** *54,* 6167–6175.
24. Muangphrom, P.; Seki, H.; Fukushima, E. O.; Muranaka, T. *J. Nat. Med.* **2016,** *70,* 318–334. doi:10.1007/s11418-016-1008-y
25. Andersen, T.; López, C.; Manczak, T.; Martinez, K.; Simonsen, H. *Molecules* **2015,** *20,* 6113–6127. doi:10.3390/molecules20046113
26. Dirsch, V. M.; Stuppner, H.; Ellmerer-Müller, E. P.; Vollmar, A. M. *Bioorg. Med. Chem.* **2000,** *8,* 2747–2753. doi:10.1016/s0968-0896(00)00202-9
27. Coricello, A.; Adams, J. D.; Lien, E.; Nguyen, C.; Perri, F.; Williams, T. J.; Aiello, F. *Curr. Med. Chem.* **2018,** in press. doi:10.2174/0929867325666180719111123
28. Kmoníčková, E.; Melkusová, P.; Harmatha, J.; Vokáč, K.; Farghali, H.; Zídek, Z. *Eur. J. Pharmacol.* **2008,** *588,* 85–92. doi:10.1016/j.ejphar.2008.03.037
29. Kmoníčková, E.; Harmatha, J.; Vokáč, K.; Kostecká, P.; Farghali, H.; Zídek, Z. *Fitoterapia* **2010,** *81,* 1213–1219. doi:10.1016/j.fitote.2010.08.005

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Analysis of sesquiterpene hydrocarbons in grape berry exocarp (*Vitis vinifera* L.) using in vivo-labeling and comprehensive two-dimensional gas chromatography–mass spectrometry (GC×GC–MS)

Philipp P. Könen and Matthias Wüst*

Full Research Paper

Open Access

Address:
Institute of Nutritional and Food Sciences, Chair of Food Chemistry, University of Bonn, Endenicher Allee 19C, 53115 Bonn, Germany

Email:
Matthias Wüst* - matthias.wuest@uni-bonn.de

* Corresponding author

Keywords:
biosynthesis; deuterium labeling; germacrene; HS-SPME; terpenes; TOF–MS

Beilstein J. Org. Chem. **2019,** *15,* 1945–1961.
doi:10.3762/bjoc.15.190

Received: 28 May 2019
Accepted: 02 August 2019
Published: 14 August 2019

This article is part of the thematic issue "Terpenes".

Guest Editor: J. S. Dickschat

Abstract

Sesquiterpenes are structurally diverse, potent flavoring substances that significantly influence the aroma profile of grapes (*Vitis vinifera* L.) at the time of physiological ripening. To investigate these natural compounds, freshly harvested, ripe berries of the red wine variety Lemberger (*Vitis vinifera* subsp. *vinifera* L.) were analyzed using comprehensive two-dimensional gas chromatography (GC×GC) coupled to a time-of-flight mass spectrometer (TOF–MS) after headspace-solid phase microextraction (HS-SPME). The identification of structurally complex natural compounds, such as sesquiterpenes from fruits and vegetables, is often reported as "tentative", as authentic standards are not commercially available for most of the analytes. For this reason, feeding experiments (in vivo labeling) were carried out using the stable isotope-labeled precursors [5,5-$^2H_2$]-1-deoxy-D-xylulose (d_2-DOX) and [6,6,6-$^2H_3$]-(±)-mevalonolactone (d_3-MVL) to clearly identify the volatiles. Based on the recorded mass spectra of the unlabeled and deuterated compounds, mechanisms for sesquiterpene formation in *V. vinifera* could be proposed and already known pathways could be confirmed or disproved. For example, the HS-SPME–GC×GC–TOF–MS measurements of fed sample material showed that the tricyclic sesquiterpene hydrocarbons α-copaene, β-copaene, α-cubebene, β-cubebene and the bicyclic δ-cadinene were biosynthesized via (*S*)-(−)-germacrene D rather than via (*R*)-(+)-germacrene D as intermediate.

Introduction

The aroma profile of grape berries at the time of physiological ripening is very complex and significantly influenced by potent flavoring substances of isoprenoid origin like mono- and sesquiterpenes [1,2]. Sesquiterpenes form a structurally diverse subgroup of terpenes consisting of three isoprene units [3]. Since the bicyclic sesquiterpene ketone rotundone was

identified for the first time in 2008 as a key aromatic substance for the peppery aroma of the red wine variety Shiraz, sesquiterpenes have increasingly been in the focus of wine research [4].

It is known that these natural products are synthesized in *Vitis vinifera* L. via the mevalonate-dependent biosynthesis pathway (MVA) localized in the cytoplasm as well as via the mevalonate-independent 1-deoxy-D-xylulose 5-phosphate/2-*C*-methyl-D-erythritol 4-phosphate metabolic pathway (DOXP/MEP) localized in plastids [5]. In the MVA pathway, which was first described in yeast and animals, mevalonic acid (MVA) is formed from acetyl-CoA and converted into the products isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), which can further react to farnesyl pyrophosphate (FPP, C_{15}) [6]. It was not until the end of the 1980s that the alternative DOXP/MEP pathway, which also leads to the formation of DMAPP and IPP, was discovered in a study focused on hopanoids [7]. Today it is known that the sesquiterpenes in *Vitis vinifera* are formed from FPP whose biosynthesis relies on both, the DOXP/MEP and the MVA pathway, and that they specifically accumulate in the wax layer of the berry exocarp [8]. The biosynthesis of the diverse sesquiterpene structures in *V. vinifera* is, however, still largely unexplored, especially with respect to the cyclization mechanisms. Studies by Steele et al., Bülow et al., Martin et al. and theoretical studies by Tantillo show the complexity of the various cyclization reactions [9-12].

In order to analyze the biosynthetic pathways of sesquiterpene hydrocarbons in grape berries, a method was developed by us using comprehensive two-dimensional gas chromatography (GC×GC) coupled to a time-of-flight mass spectrometer (TOF–MS) after headspace-solid phase microextraction (HS-SPME).

The comprehensive two-dimensional gas chromatography first demonstrated on an oil sample in 1991 is suitable for the analysis of highly complex samples [13]. Using GC×GC, all analytes of a sample can be separated on two different capillary separation columns of different selectivity, resulting in optimized resolution. Increased resolution makes coelutions less likely and minimizes chromatographic noise. In addition, modulator-controlled focusing effects favor the detection of analytes in the trace range. Furthermore, GC×GC is used to obtain structured chromatograms, i.e., compounds of common structure classes are grouped. The coupling of comprehensive two-dimensional gas chromatography with mass spectrometry as the "third dimension" also provides important information for substance identification. The combination of these advantages makes the GC×GC one of the most powerful separation methods for the chemical analysis of organic compounds in complex matrices [14]. The efficiency of this modern analysis technology has recently been demonstrated using different types of wine in nontarget screenings [15-17].

In our work, we have combined the advantages of multidimensional gas chromatography with those of in vivo labeling. In vivo labeling makes it possible to identify compounds for which authentic standards are not commercially available, which applies to the majority of compounds found in plant foods [18]. We performed feeding experiments on isolated exocarp of freshly harvested grapes (*Vitis vinifera* L.) using the stable isotope-labeled precursors [5,5-$^2H_2$]-1-deoxy-D-xylulose (d_2-DOX) and [6,6,6-$^2H_3$]-(±)-mevalonolactone (d_3-MVL) to unambiguously identify sesquiterpene hydrocarbons. Based on the obtained mass spectra of the genuine and deuterated compounds, multistage cyclisation reactions in the course of sesquiterpene biosynthesis could be substantiated for the first time for several structural classes in *V. vinifera*.

Results and Discussion

The HS-SPME–GC×GC–TOF–MS analysis of the grape variety Lemberger (*Vitis vinifera* subsp. *vinifera*, clone 1Gm, isolated exocarp) showed the presence of several hundred components. Twenty-five of these compounds were unequivocally identified as sesquiterpene hydrocarbons (Figure 1).

Deuterium-labeled compounds eluted earlier than the unlabeled analogues (Figure 3). This is due to the inverse isotope effect described by Matucha et al. which results in the gas chromatographic separation of isotopologic substances [28]. In the following, the fully deuterium-labeled compounds are examined and their possible biosynthesis pathways described. However, it was also possible to detect the peaks of the partially labeled sesquiterpenes, as Supporting Information File 2 shows using compound **10** as an example.

Sesquiterpene biosynthesis

Biosynthesis of sesquiterpene hydrocarbons via (*S*)-(−)-germacrene D

The biosynthetic pathways for the formation of δ-cadinene, α-copaene, β-copaene, α-cubebene and β-cubebene via germacrene D in the legume *Medicago truncatula* have been previously reported by Boland and Garms [29].

The formation of the mentioned sesquiterpenes could a priori take place via farnesyl pyrophosphate (FPP) as well as via (*S*)- and (*R*)-nerolidyl pyrophosphate (NPP) (Scheme 1).

Both enantiomers of NPP are possible as intermediates of terpene biosynthesis, since the absolute configurations of their

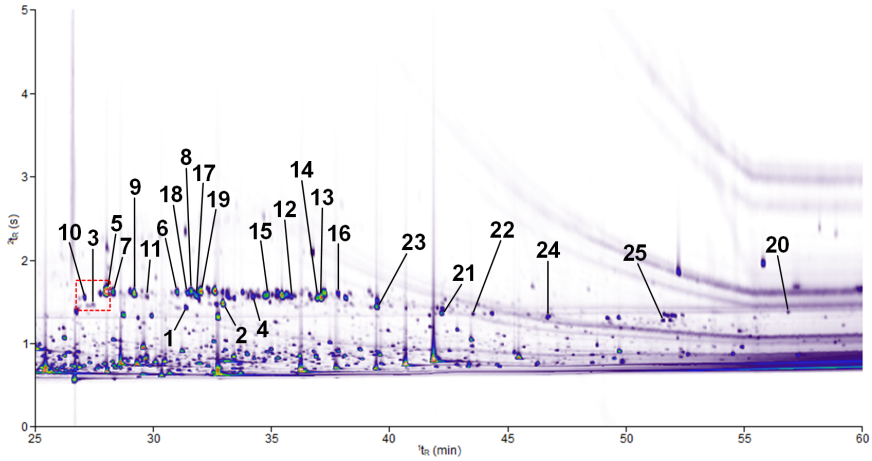


Figure 1: Contour plot of a HS-SPME–GC×GC–TOF–MS chromatogram (TIC) demonstrating the separation of volatile compounds isolated from the headspace of grape berries of the red wine variety Lemberger (*Vitis vinifera* subsp. *vinifera*, clone 1Gm, exocarp). ^1t_R (X-axis) corresponds to the retention time on the primary column and ^2t_R (Y-axis) to the retention time on the secondary column. The color gradient reflects the intensity of the TOF–MS signal on a white background from low (violet) to high (red). Numbers at peaks refer to compound numbers as defined in Figure 2 and Table 1. The area framed in red including the volatile compounds numbered 10, 3 and 5 is enlarged in Figure 3.

Table 1: Sesquiterpene hydrocarbons identified in headspace of Lemberger (*Vitis vinifera* subsp. *vinifera*, clone 1Gm, exocarp).

compound[a]	*I*[b]	*I* (lit.)[c]	identification[d]/verification[e]
α-cubebene (**10**)	1467	1460 [19]	ms (874, 882), ri, d_2-DOX, d_3-MVL
δ-elemene (**3**)	1478	1479 [20]	ms (859, 864), ri, n.d.[f], d_3-MVL
α-ylangene (**5**)	1497	1483 [21]	ms (860, 877), ri, d_2-DOX, d_3-MVL
α-copaene (**7**)	1505	1496 [19]	ms (865, 872), ri, d_2-DOX, d_3-MVL
β-bourbonene (**9**)	1532	1519 [19]	ms (853, 863), ri, d_2-DOX, d_3-MVL
β-cubebene (**11**)	1549	1537 [19]	ms (869, 883), ri, d_2-DOX, d_3-MVL
β-ylangene (**6**)	1588	1568 [19]	ms (852, 856), ri, d_2-DOX, d_3-MVL
β-elemene (**1**)	1598	1592 [22]	ms (903, 906), ri, n.d.[f], d_3-MVL
α-guaiene (**18**)	1600	1591 [23]	ms (871, 892), ri, d_2-DOX, d_3-MVL
β-copaene (**8**)	1605	1598 [24]	ms (847, 856), ri, d_2-DOX, d_3-MVL
(*E*)-β-caryophyllene (**17**)	1612	1604 [25]	ms (928, 939), ri, d_2-DOX, d_3-MVL
guaia-6,9-diene (**19**)	1618	n.a.[g]	ms (824, 833), d_2-DOX, d_3-MVL
γ-elemene (**2**)	1646	1650 [20]	ms (864, 882), ri, d_2-DOX, d_3-MVL
α-humulene (**4**)	1685	1690 [20]	ms (888, 906), ri, d_2-DOX, d_3-MVL
δ-selinene (**15**)	1705	n.a.[g]	ms (841, 887), n.d.[f], d_3-MVL
(+)-valencene (**12**)	1736	1731 [26]	ms (884, 904), ri, d_2-DOX, std, n.d.[h]
δ-cadinene (**14**)	1769	1770 [20]	ms (858, 900), ri, d_2-DOX, d_3-MVL
γ-cadinene (**13**)	1774	1760 [23]	ms (852, 881), ri, d_2-DOX, d_3-MVL
selina-3,7(11)-diene (**16**)	1796	1778 [27]	ms (872, 881), ri, n.d.[f], d_3-MVL
calamenene (isomer) (**23**)	1845	1837 [27]	ms (786, 814), ri, n.d.[f], d_3-MVL
α-calacorene (**21**)	1929	1919 [27]	ms (746, 854), ri, n.d.[f], d_3-MVL
β-calacorene (**22**)	1971	1939 [25]	ms (836, 898), ri, n.d.[f], d_3-MVL
α-corocalene (**24**)	2073	n.a.[g]	ms (864, 882), n.d.[f], d_3-MVL
cadalene (**25**)	2237	2231 [21]	ms (869, 880), ri, n.d.[f], d_3-MVL
guaiazulene (**20**)	2417	n.a.[g]	ms (872, 880), n.d.[f], std, d_3-MVL

[a]Unidentified compounds are not listed. [b]Retention index *I* on a DB-WAX Ultra Inert column. [c]Retention index data from literature. [d]Compound identification is based on matching mass spectrum to a library spectrum (ms, match factor and reverse match factor given in brackets, identical mass spectra would produce a match factor of 1000), identical or closely matching retention index (ri) and comparison to a commercially available standard compound (std). [e]Verification of the found sesquiterpene hydrocarbons was carried out by in vivo labeling with [5,5-$^2H_2$]-1-deoxy-D-xylulose (d_2-DOX) and [6,6,6-$^2H_3$]-(±)-mevalonolactone (d_3-MVL) as stable isotope-labeled precursors. [f]The compound could not be detected in d_2-DOX feeding experiments or the mass spectra could not be evaluated. [g]Retention index data on a WAX column were not available. [h]The compound could not be detected in d_3-MVL feeding experiments or the mass spectra could not be evaluated.

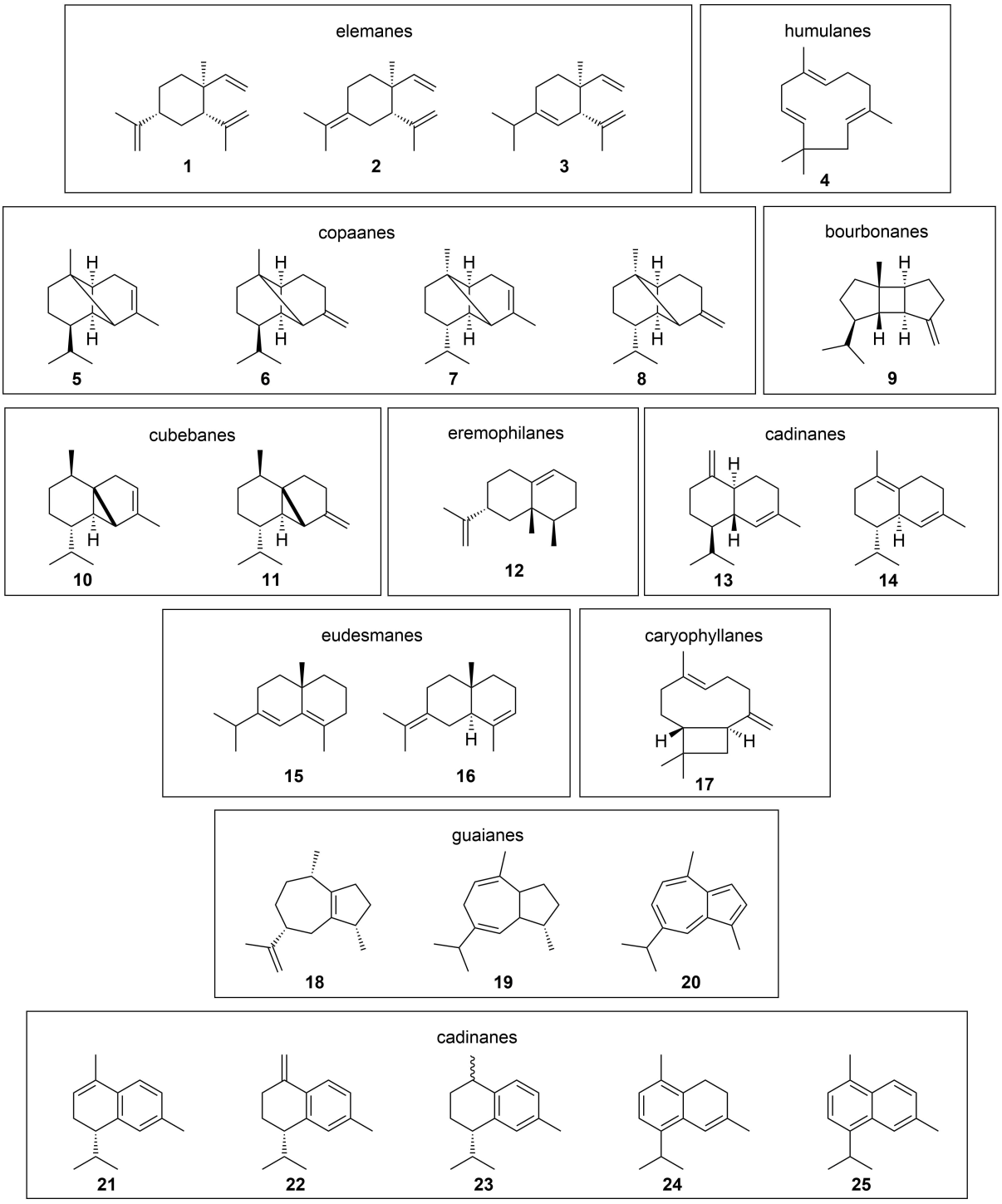


Figure 2: Sesquiterpene hydrocarbons found in the headspace of Lemberger (*Vitis vinifera* subsp. *vinifera*, clone 1Gm, exocarp). **1–4**, monocyclic sesquiterpenes; **5–11**, tricyclic sesquiterpenes; **12–25**, bicyclic sesquiterpenes; **20–25**, aromatic sesquiterpenes. The identified compounds are grouped according to the type of sesquiterpene skeleton.

products from *Vitis vinifera* L. are unknown and the subsequent cyclisation reactions can be explained by the enantiomers of germacrene D [30,31].

In order to investigate whether the formation of δ-cadinene, α-copaene, β-copaene, α-cubebene and β-cubebene occurs via the intermediate farnesyl pyrophosphate (FPP) or via nerolidyl

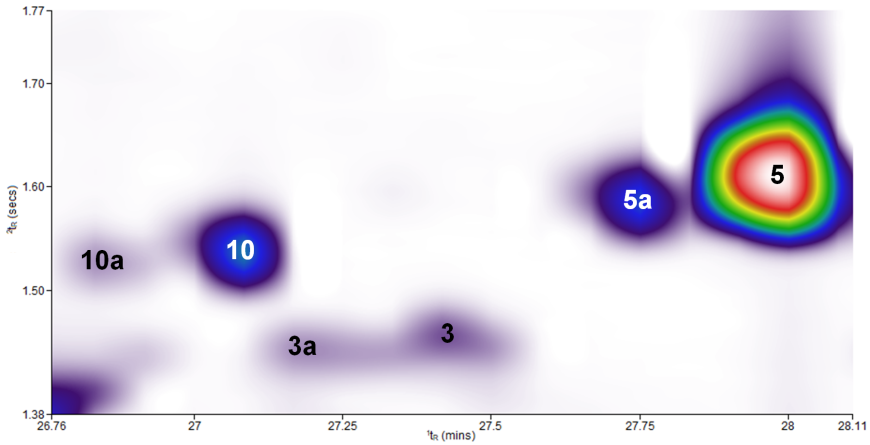
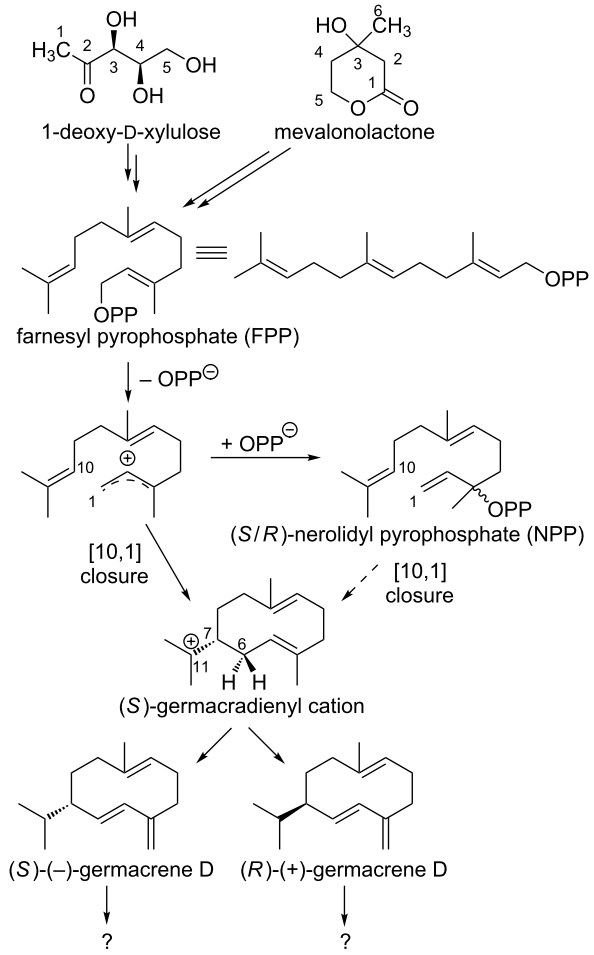


Figure 3: Detailed part of the two-dimensional contour plot (Figure 1) to demonstrate the result of a successful feeding experiment after administration of the stable isotope-labeled precursor [6,6,6-$^2H_3$]-(±)-mevalonolactone (d_3-MVL) to isolated exocarp of grape berries (Lemberger cultivar). The regions labeled with the numbers **10**, **3** and **5** correspond to the genuine sesquiterpene hydrocarbons as defined in Table 1. **10a** (d_8), **3a** (d_9) and **5a** (d_8) are the isotopologues with the highest, possible incorporation of deuterium when d_3-MVL is used as precursor (the maximum possible number of deuterium atoms incorporated is given in brackets). The 3D view of the chromatogram section shown can be found in Supporting Information File 1.



Scheme 1: First steps towards the formation of sesquiterpenes. The (*S*)-germacradienyl cation can be formed from FPP or NPP. The subsequent formation of sesquiterpene hydrocarbons via (*S*)- and (*R*)-germacrene D has not yet been clarified.

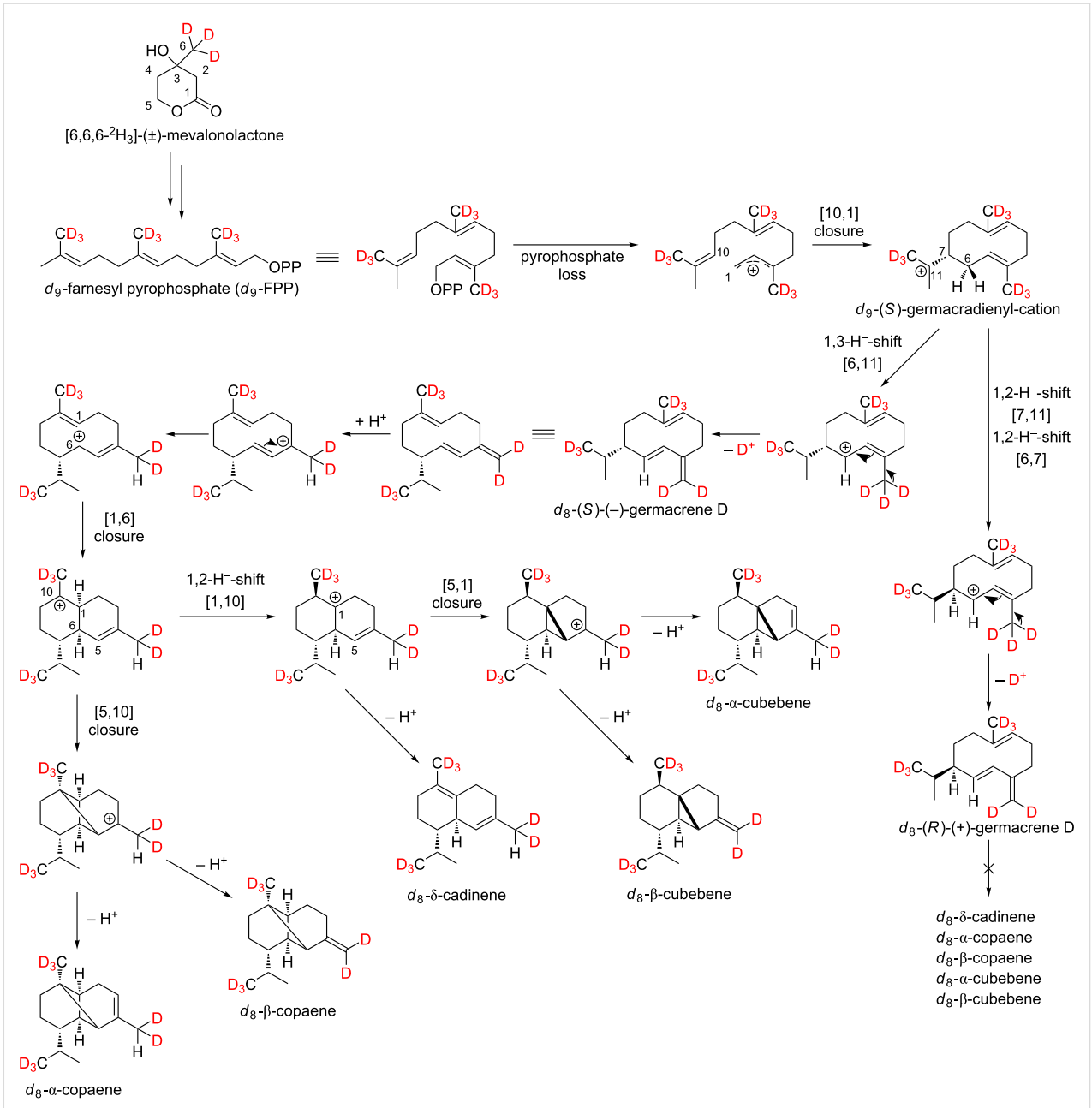
pyrophosphate (NPP), feeding experiments were carried out using the stable isotope-labeled precursor [6,6,6-$^2H_3$]-(±)-mevalonolactone (d_3-MVL, Scheme 2).

As shown in Scheme 2, a maximum incorporation of 8 deuterium atoms was found for the sesquiterpene hydrocarbons, which excludes formation via NPP. Furthermore, these results support the formation of germacrene D, in which the ninth deuterium atom is already missing.

Whether the above-mentioned sesquiterpenes are formed via (*S*)-(−)-germacrene D or (*R*)-(+)-germacrene D was investigated by feeding experiments using the precursor [5,5-$^2H_2$]-1-deoxy-D-xylulose (d_2-DOX, Scheme 3).

Indeed, in Lemberger (*R*)-(+)-germacrene D and (*S*)-(−)-germacrene D are detectable in grape berry exocarp as shown by enantioselective analysis [5,32]. In 1999 Schmidt et al. were able to show that the enantiomers of germacrene D are formed via two different H-transfer pathways in *Solidago canadensis*. (*S*)-(−)-Germacrene D is generated by a cyclization reaction, that includes a 1,3-hydride shift as opposed to the cyclization reaction of (*R*)-(+)-germacrene D, that includes two consecutive 1,2-hydride shifts [33].

In all terpenes in Scheme 3, TOF–EIMS measurements showed that a deuterium atom is located in the isopropyl group. This can be explained by the assumption that in *Vitis vinifera* L. all sesquiterpenes are formed via the (*S*)-germacradienyl cation, which has already been shown for *Solidago canadensis*. The subsequent reactions of the (*S*)-germacradienyl cation to (*S*)-(−)-germacrene D and (*R*)-(+)-germacrene D have already been



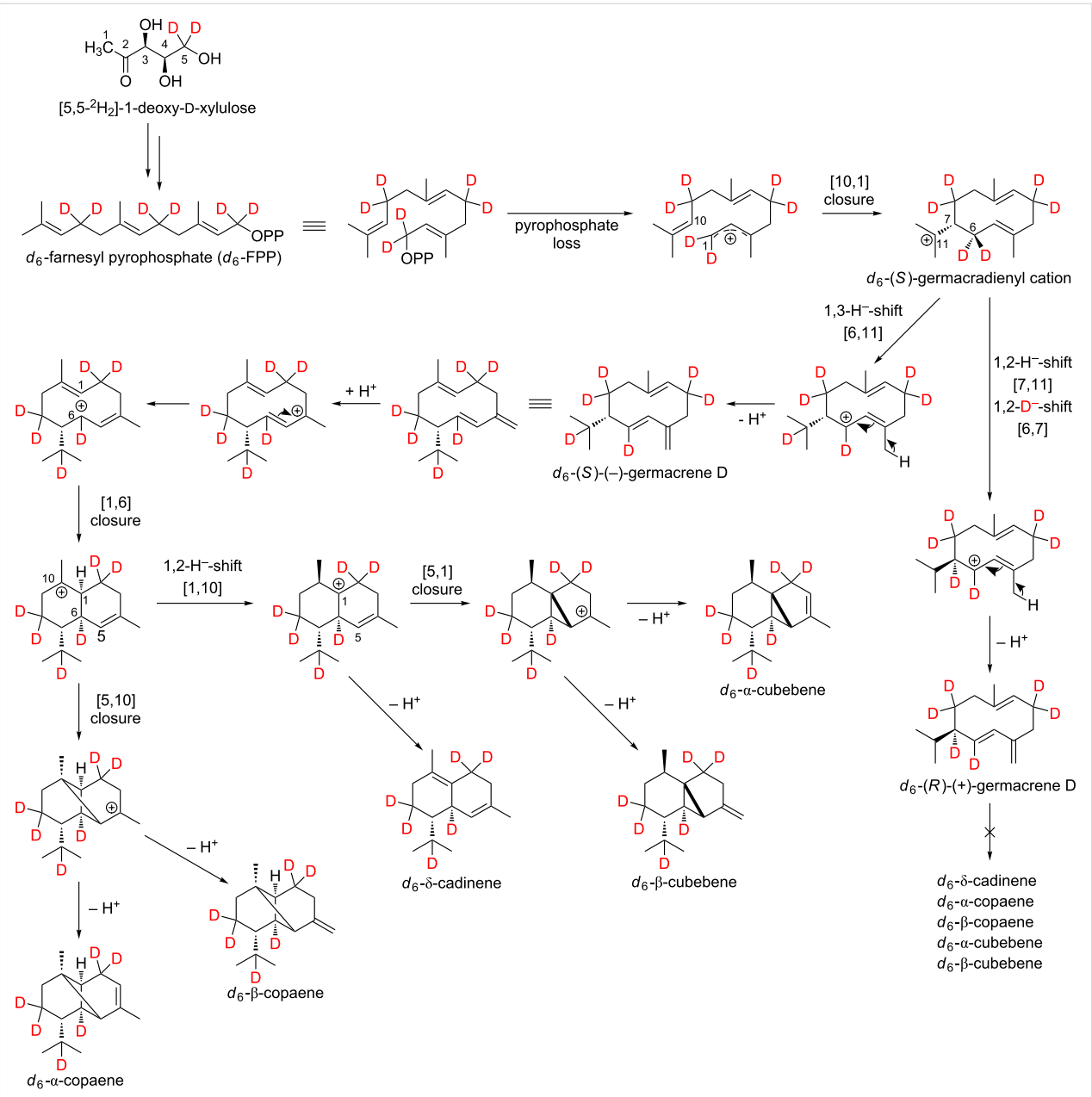
Scheme 2: Possible biosynthetic pathways of the sesquiterpene hydrocarbons d_8-α-copaene, d_8-β-copaene, d_8-α-cubebene, d_8-β-cubebene and d_8-δ-cadinene. As a stable isotope-labeled precursor, d_3-MVL was added to isolated exocarp of ripe grapes of the Lemberger grape variety.

described above. In the case of the formation of δ-cadinene, α-copaene, β-copaene, α-cubebene and β-cubebene, further cyclization takes place only from (*S*)-(−)-germacrene D, from which the absolute configuration of these substances can be derived.

While the detected labeling patterns in our feeding experiments are in full agreement with the 1,3-hydride shift pathway, the cloning and functional characterization of the corresponding synthase as a final proof is yet missing. In the following, the sesquiterpene hydrocarbon α-cubebene is used as an example to illustrate our interpretations in more detail (Figure 4).

Genuine (d_0) α-cubebene has a molecular ion peak $[M]^{+\bullet}$ at m/z = 204 (Figure 4, upper mass spectrum, marked red). The signal with m/z = 161 $[M]^+$ corresponds to the fragment after abstraction of the isopropyl group (marked blue).

Feeding experiments using d_2-DOX showed that according to the biosynthesis pathway demonstrated in Scheme 3, a



Scheme 3: Mechanistic rationale for the generation of the sesquiterpene hydrocarbons δ-cadinene (**14**), α-copaene (**7**), β-copaene (**8**), α-cubebene (**10**) and β-cubebene (**11**) via (*S*)-(−)-germacrene D after feeding isolated exocarp of grapes using the stable isotope-labeled precursor [5,5-$^2H_2$]-1-deoxy-D-xylulose (d_2-DOX).

maximum of 6 deuterium atoms are incorporated into α-cubebene. The presence of d_6-α-cubebene could be confirmed by mass spectrometry as the isotopologue with the highest, possible incorporation of deuterium when d_2-DOX was used as a precursor. The molecular ion peak $[M]^{+\bullet}$ of d_6-α-cubebene therefore has a mass-to-charge ratio of 210. The abstraction of a single deuterated isopropyl group leads to the fragment with m/z = 166. At this point it becomes obvious why it can be assumed that the tricyclic sesquiterpene hydrocarbon α-cubebene is synthesized via (*S*)-(−)-germacrene D and not via (*R*)-(+)-germacrene D. Assuming the biosynthesis of α-cubebene takes place via (*R*)-(+)-germacrene D, the isopropyl group of the molecule would not be deuterated and, therfore, after abstraction of the isopropyl group, the signal with m/z = 167 would dominate, which is not the case.

In the course of the biosynthesis of α-cubebene after feeding with d_3-MVL as precursor, a deuterium atom is abstracted, resulting in d_8-α-cubebene $[M]^{+\bullet}$ with m/z = 212 as highly labeled isotopologue (Figure 4, lower mass spectrum).

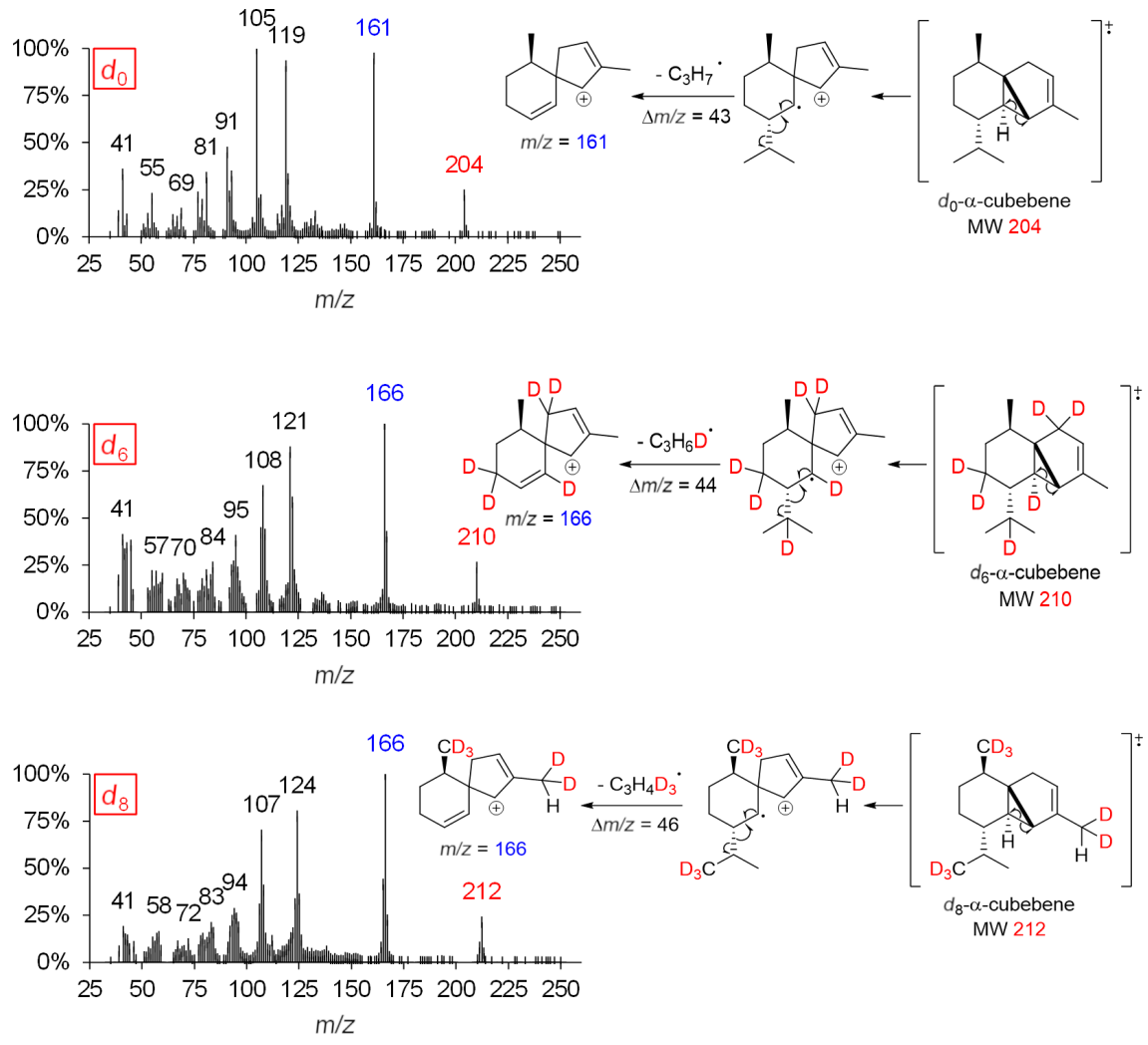


Figure 4: MS spectra of genuine (d_0) and deuterium-labeled (d_6 and d_8) α-cubebene (left panel) after administration of [5,5-$^2H_2$]-1-deoxy-D-xylulose (d_2-DOX) and [6,6,6-$^2H_3$]-(±)-mevalonolactone (d_3-MVL) to isolated exocarp of grape berries (Lemberger cultivar). Expected labeling patterns are depicted for α-cubebene (right panel).

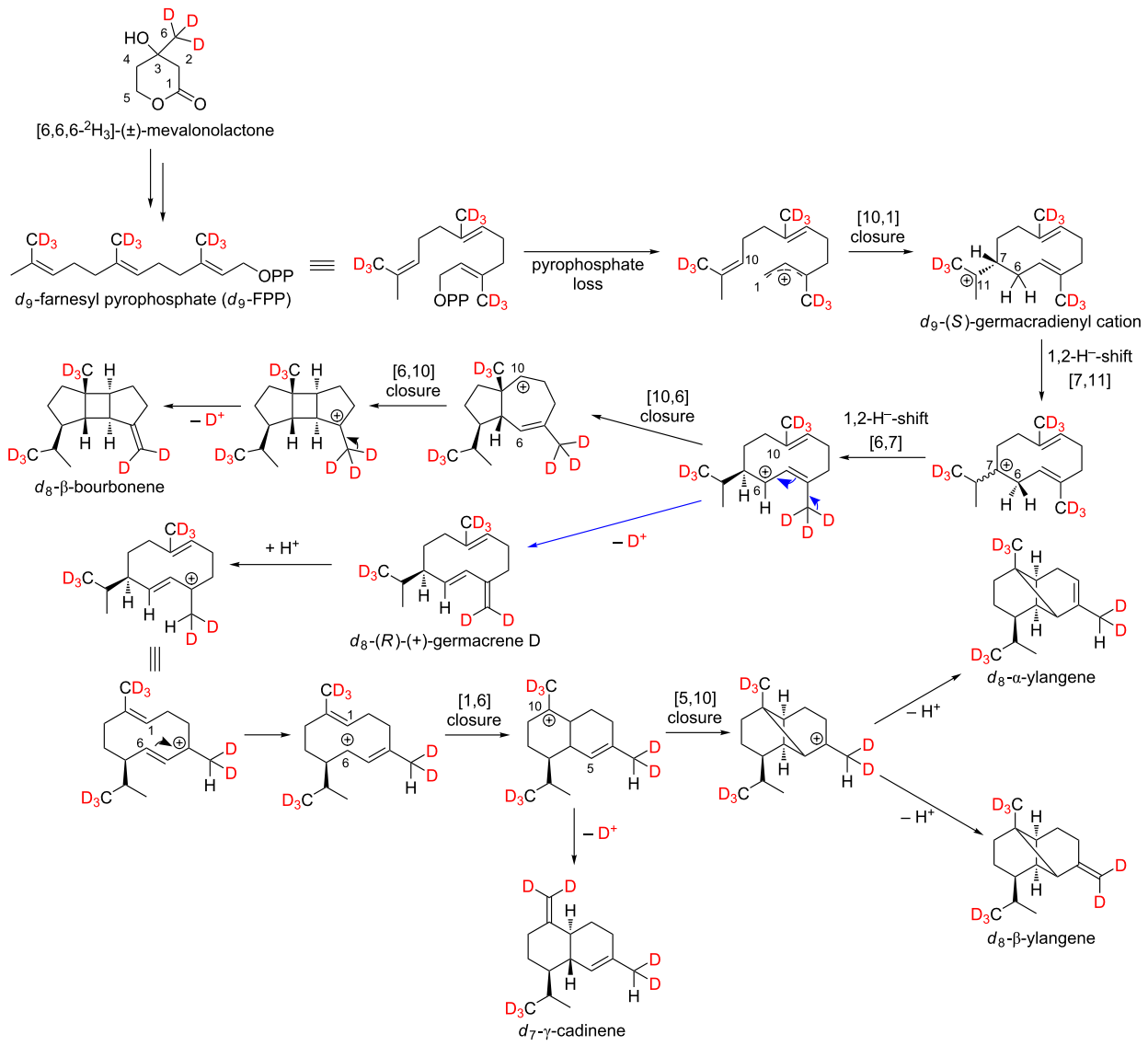
Biosynthesis of sesquiterpene hydrocarbons via (*R*)-(+)-germacrene D

Lodewyk, Gutta and Tantillo postulated an α-ylangene-forming carbocation cascade based on computer calculations [34]. Tantillo and co-workers showed that germacrene D does not have to be involved in the formation of α-ylangene due to carbocation energetics. However, according to our results the formation of this tricyclic sesquiterpene hydrocarbon and its isomer β-ylangene occurs via the intermediate (*R*)-(+)-germacrene D (Scheme 4), which can, indeed, be detected by enantioselective GC in ripening Lemberger berries [5].

The MS spectra and the expected labeling patterns of genuine (A) and deuterium-labeled d_6-α-ylangene (C) after feeding experiments with d_2-DOX are shown in Figure 5.

Figure 5B shows that after feeding experiments with d_3-MVL a maximum of 8 deuterium atoms are incorporated into α-ylangene, so that a molecular ion peak $[M]^{+\bullet}$ with m/z = 212 is detected. After abstraction of the triple deuterated isopropyl group, the fragment ion is formed with m/z = 166. According to the computer simulation of Lodewyk, Gutta and Tantillo, d_9-α-ylangene with m/z = 213 should have been detected after in vivo labeling using [6,6,6-$^2H_3$]-(±)-mevalonolactone (d_3-MVL) as precursor. At this point it should be mentioned that biosynthetic pathways in organisms can differ and computational studies only bear on the intrinsic energetics of possible carbocations.

Furthermore, it is possible that γ-cadinene is formed via (*R*)-(+)-germacrene D. Yoshihara et al. showed that isolated germa-



Scheme 4: Putative formation pathways of the sesquiterpene hydrocarbons α-ylangene (**5**), β-ylangene (**6**), β-bourbonene (**9**) and γ-cadinene (**13**) via (*R*)-(+)-germacrene D after feeding experiments using [6,6,6-$^2H_3$]-(±)-mevalonolactone (d_3-MVL).

crene D reacts to β-bourbonene by irradiation with intense UV light (Hg lamp) [35]. However, the formation of β-bourbonene in plant cells by this photoinduced [2 + 2]-cycloaddition is very unlikely. The biosynthesis of β-bourbonene can be performed by cationic cyclization, as shown in Scheme 4.

A total of 5 bicyclic, aromatic sesquiterpene hydrocarbons (**21–25**) with a cadinane skeleton were found in exocarp of the grape variety Lemberger, whose identity were confirmed by feeding experiments (Figure 6).

The biosynthesis of the five aromatic sesquiterpene hydrocarbons mentioned is still unknown, but we were able to determine the isotopologues with the highest, possible incorporation of deuterium when d_3-MVL is used as precursor.

Figure 7 shows the mass spectrum of genuine (d_0) and deuterium-labeled (d_8) calamenene (isomer) as well as α-calacorene.

Biosynthesis of sesquiterpene hydrocarbons via germacrene A

Faraldos et al. showed that β-elemene is formed by Cope rearrangement when a solution of germacrene A in toluene is heated at reflux [36]. The formation of elemenes requires temperatures >100 °C, so that this reaction cannot take place in the

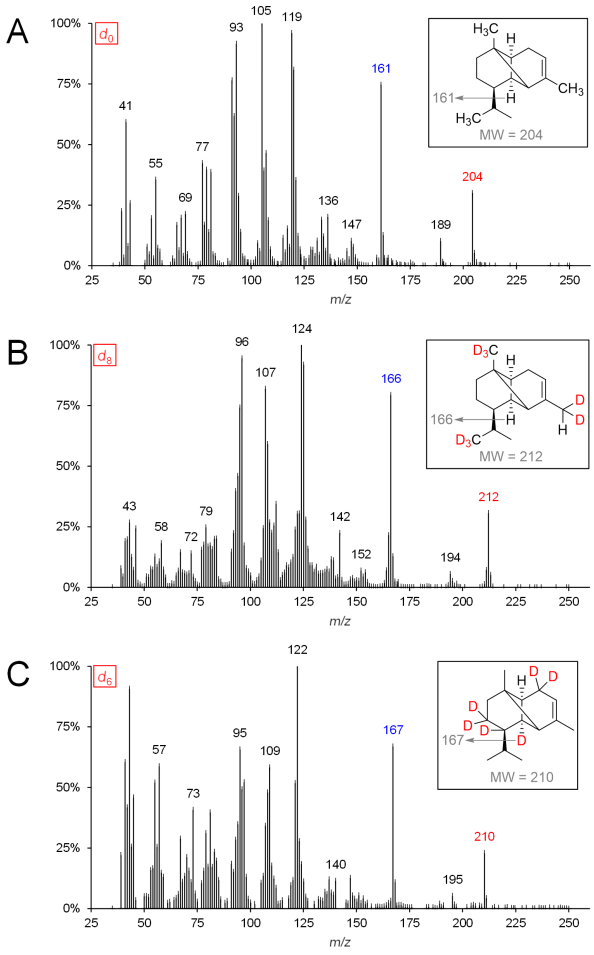


Figure 5: MS spectra and expected labeling patterns of A: d_0-α-ylangene, B: d_8-α-ylangene after administration of d_3-MVL to isolated exocarp of grapes and C: d_6-α-ylangene after feeding experiments using d_2-DOX.

cell, but it must be assumed that these are artifacts of the GC analysis [37]. Nevertheless, elemenes indicate the presence of the corresponding germacrene, as shown by the example of β-elemene and germacrene A (Figure 8 and Scheme 5).

The biosynthesis pathway postulated by Steele et al. for the formation of α-guaiene could also be confirmed [9]. The 2D separation of the sesquiterpene hydrocarbons β-elemene and α-guaiene can be found in Supporting Information File 3. Tantillo suggested the synthesis of (+)-valencene from germacrene A [11]. This biosynthetic pathway could also be confirmed by feeding experiments.

The formation of the three aforementioned sesquiterpene hydrocarbons takes place without deuterium loss, so that with the use of d_3-MVL as precursor (Scheme 5) nine deuterium atoms and with the use of d_2-DOX six deuterium atoms were incorporated at most. This data indicates that the biosynthesis of β-elemene, (+)-valencene and α-guaiene is achieved via germacrene A.

Biosynthesis of sesquiterpene hydrocarbons via (*E*,*E*)-germacrene B

The mechanisms proposed by Steele et al. for the formation of selina-3,7(11)-diene and γ-elemene via (*E*,*E*)-germacrene B are consistent with our results (Scheme 6) [9]. In Figure 9 the recorded mass spectra of d_0-γ-elemene and d_9-γ-elemene are shown as examples.

Biosynthesis of guaiazulene, δ-elemene, guaia-6,9-diene and δ-selinene

The formation of the aromatic sesquiterpene guaiazulene is still unknown. Our feeding experiments with the precursor d_3-MVL

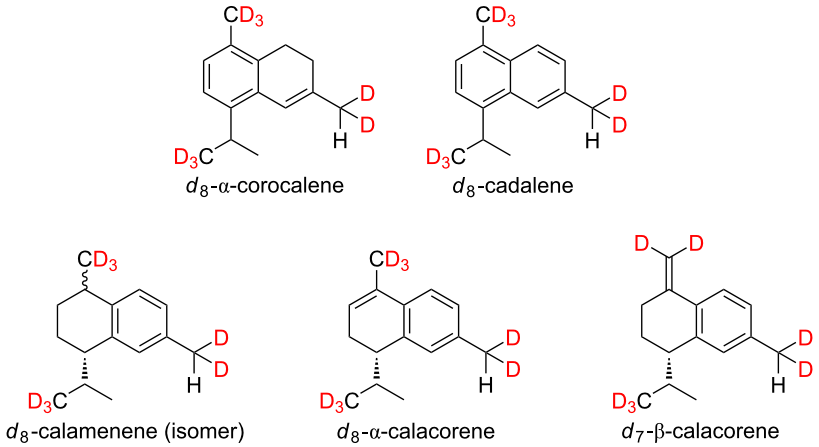


Figure 6: Expected labeling patterns of deuterium-labeled, aromatic sesquiterpenes after administration of [6,6,6-$^2H_3$]-(±)-mevalonolactone (d_3-MVL) to isolated exocarp of grape berries (Lemberger cultivar).

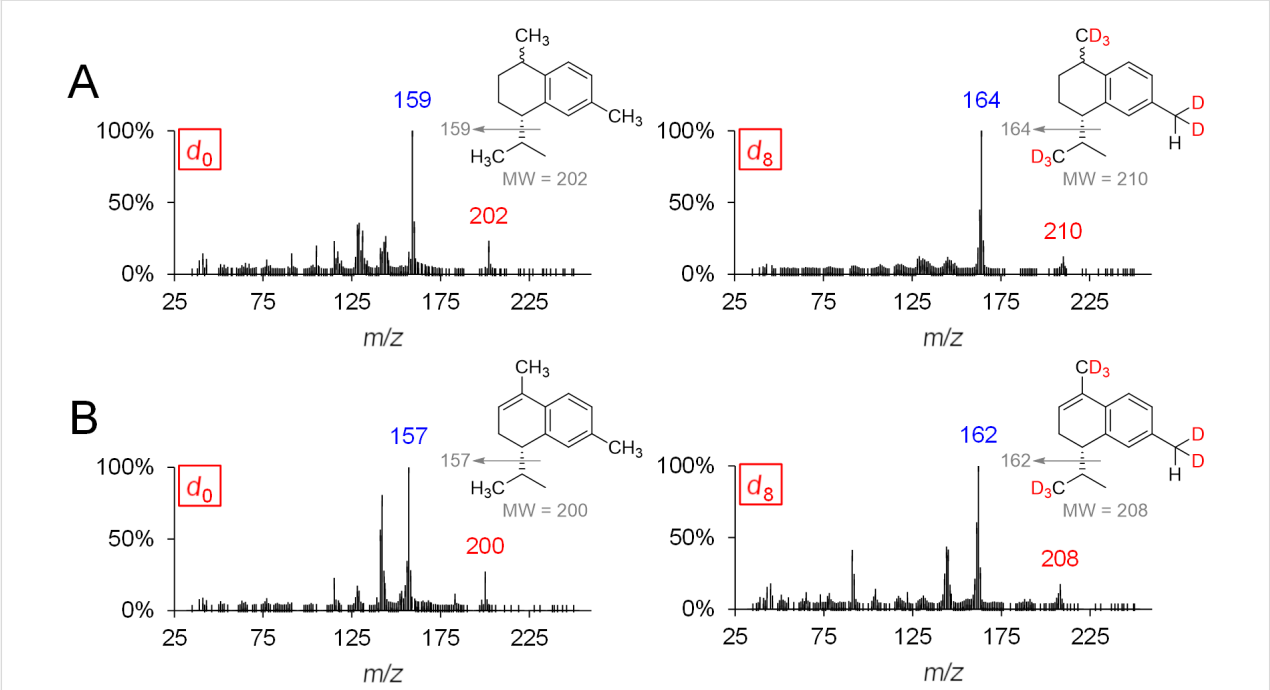


Figure 7: MS spectra and expected labeling patterns of genuine and deuterium-labeled A: calamenene (isomer) and B: α-calacorene after feeding experiments using d_3-MVL.

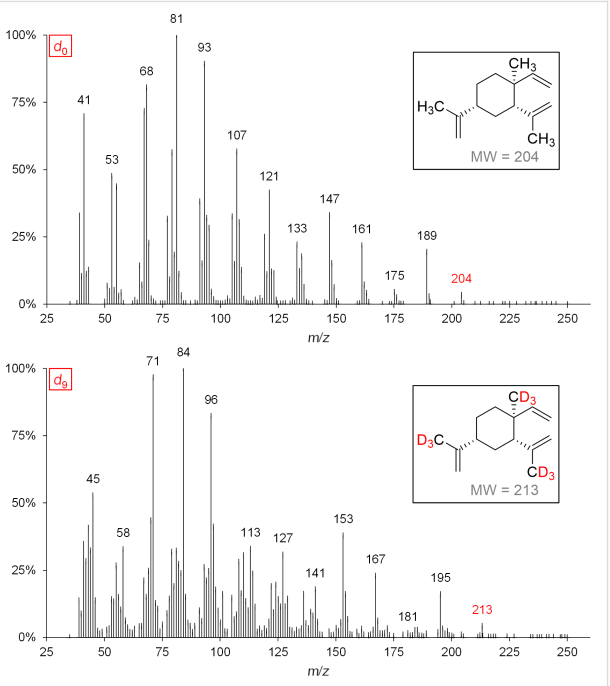


Figure 8: MS spectra and expected labeling patterns of genuine (d_0) and deuterium-labeled (d_9) β-elemene after feeding experiments of grape berry exocarp using d_3-MVL.

show that no deuterium loss is to be expected during the biosynthesis of guaiazulene in grapes, and that a maximum of 9 deuterium atoms are incorporated into guaiazulene (Figure 10).

The biosynthesis of guaiazulene could be performed via germacrene C (Scheme 7).

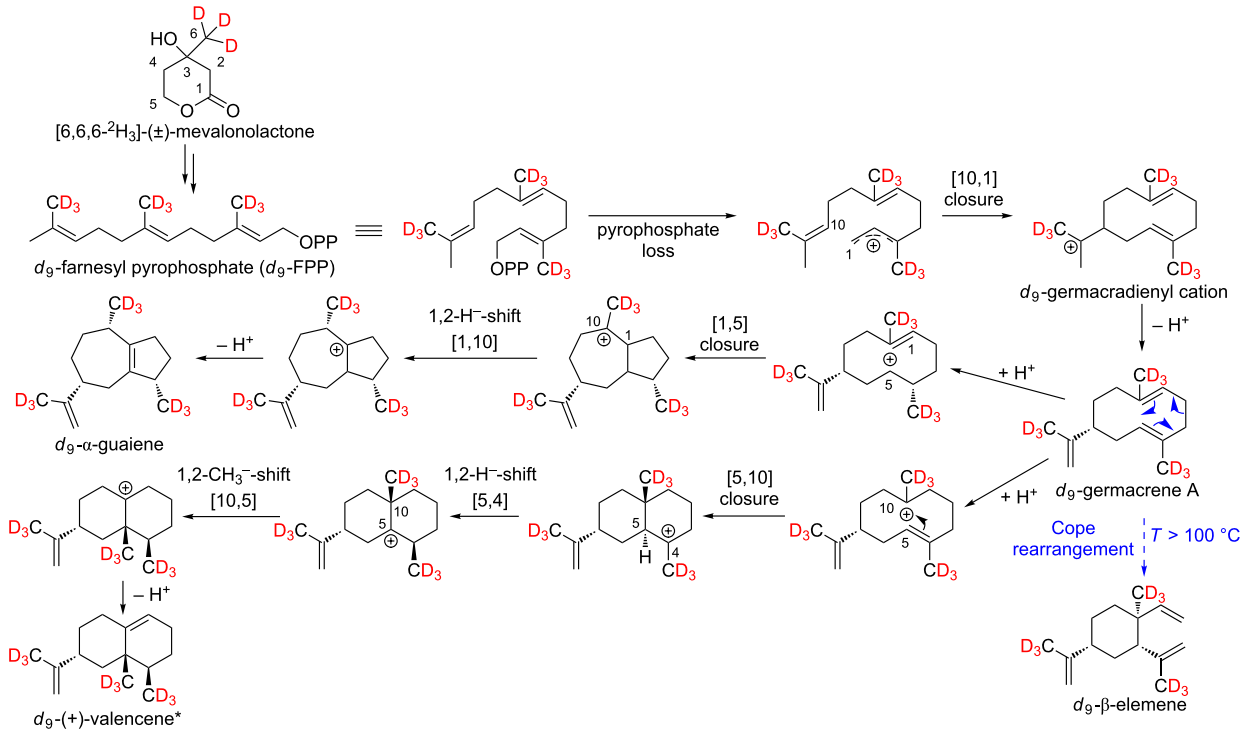
It is conceivable that double bonds are introduced by desaturases in multistage reactions to form the aromatic compound guaiazulene. The participation of desaturases in the formation of aromatic terpenes was already described in 1978 by Poulose et al. explaining the conversion of γ-terpinene to *p*-cymene [38].

May described the formation of δ-elemene from germacrene C via Cope rearrangement [32]. Considering that high temperatures are required for this reaction, our results also indicate that δ-elemene may be formed from germacrene C. The mechanisms described by Steele et al. for the formation of guaia-6,9-diene and δ-selinene via germacrene C could also be confirmed by feeding experiments (Scheme 7) [9].

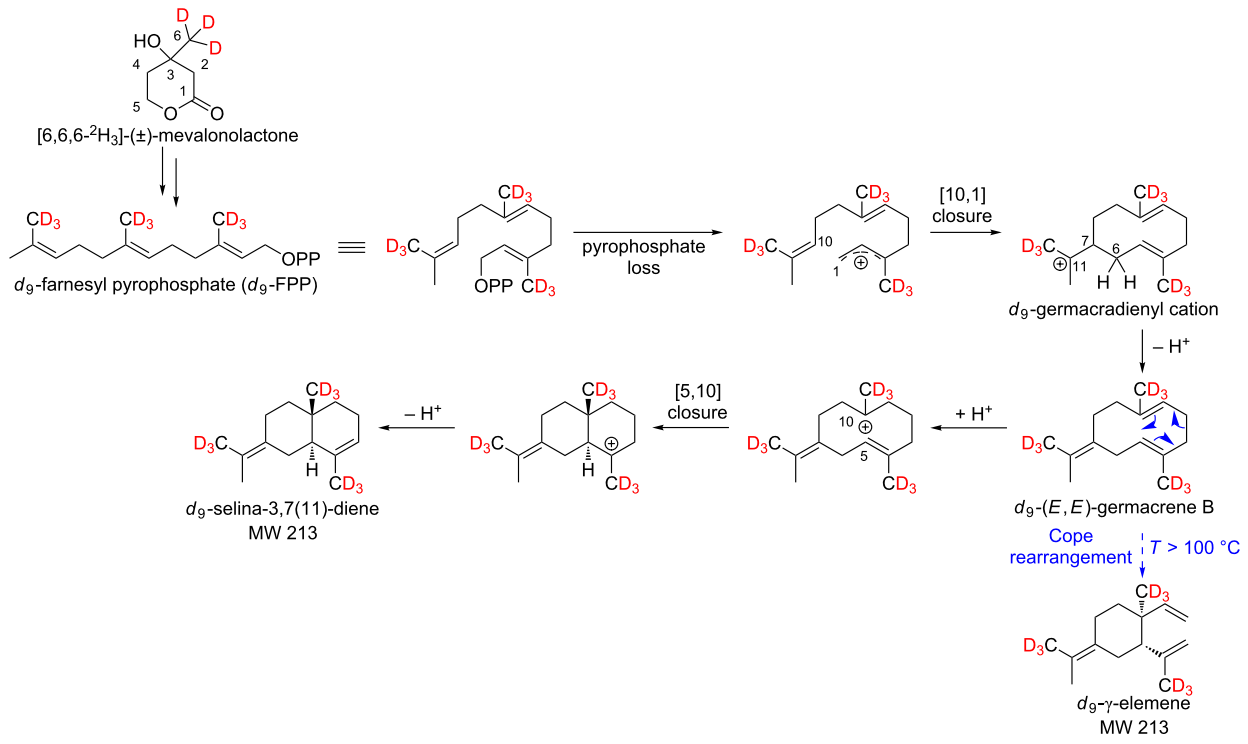
Biosynthesis of (*E*)-β-caryophyllene and α-humulene

The biosynthetic pathway postulated by Boland and Garms for the formation of (*E*)-β-caryophyllene from farnesyl pyrophosphate (FPP) is consistent with our feeding experiments (Scheme 8) [29].

The incorporation of deuterium atoms into (*E*)-β-caryophyllene during biosynthesis is shown by the following mass spectra (Figure 11).



Scheme 5: Possible biosynthesis of d_9-β-elemene, d_9-(+)-valencene and d_9-α-guaiene via germacrene A. *An incorporation of deuterium atoms into (+)-valencene could be detected, but due to coeluting substances no characteristic mass spectrum of d_9-(+)-valencene could be obtained.



Scheme 6: Mechanistic rationale for the generation of the sesquiterpene hydrocarbons γ-elemene and selina-3,7(11)-diene via (*E*,*E*)-germacrene B. The shown deuterium incorporation results from feeding experiments of exocarp (Lemberger cultivar) using the precursor d_3-MVL.

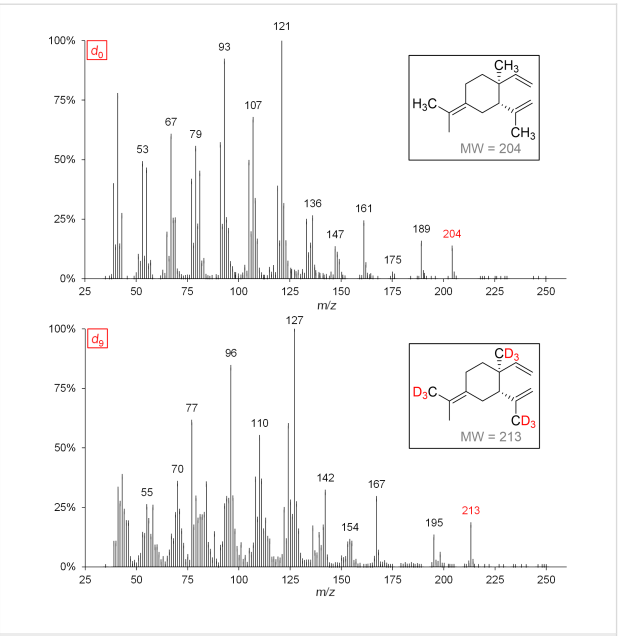


Figure 9: Mass spectra and associated structural formulas of d_0-γ-elemene and d_9-γ-elemene after administration of [6,6,6-$^2H_3$]-(±)-mevalonolactone (d_3-MVL) to isolated exocarp of grape berries (Lemberger cultivar).

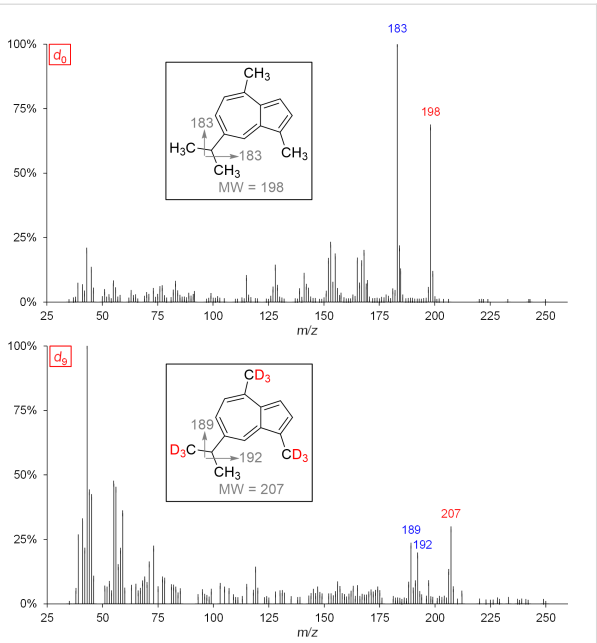
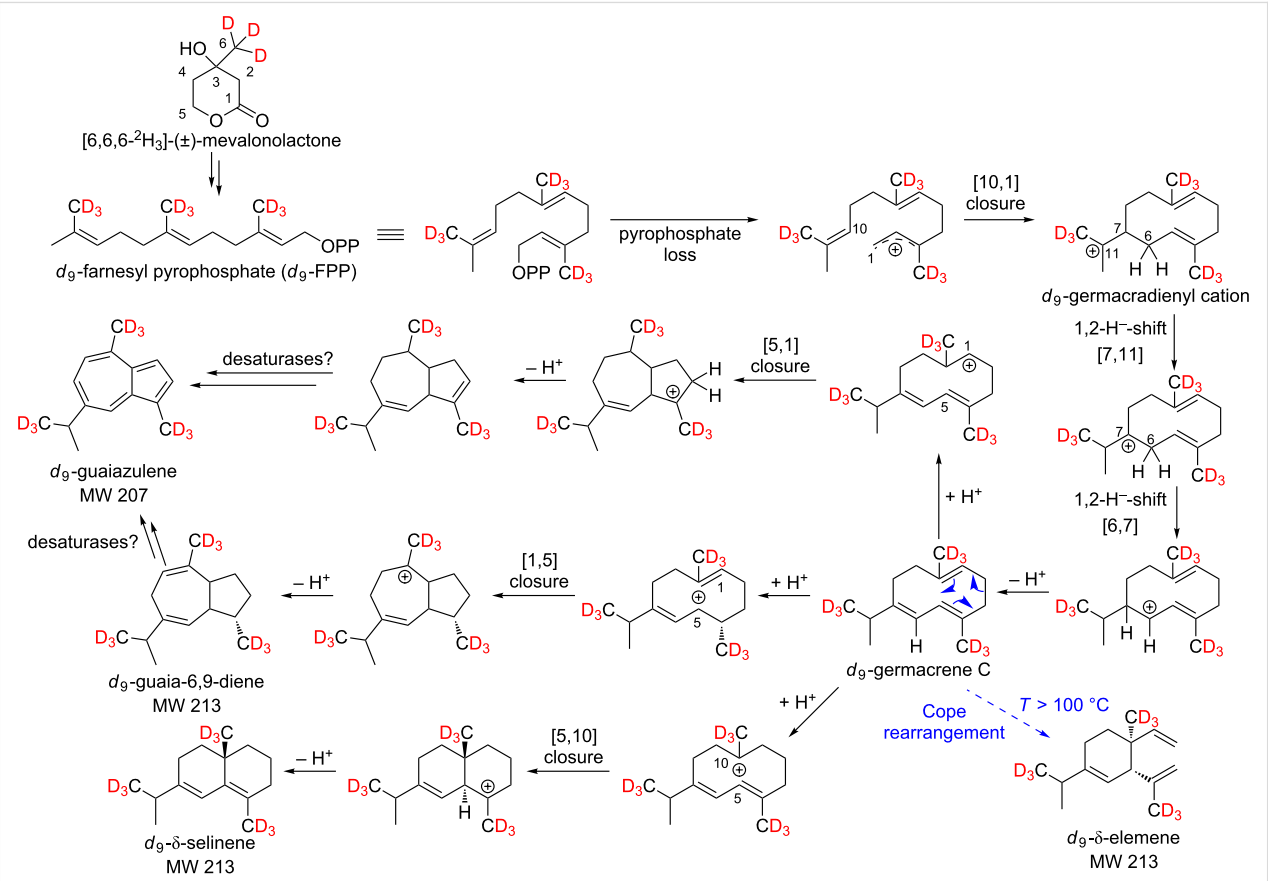
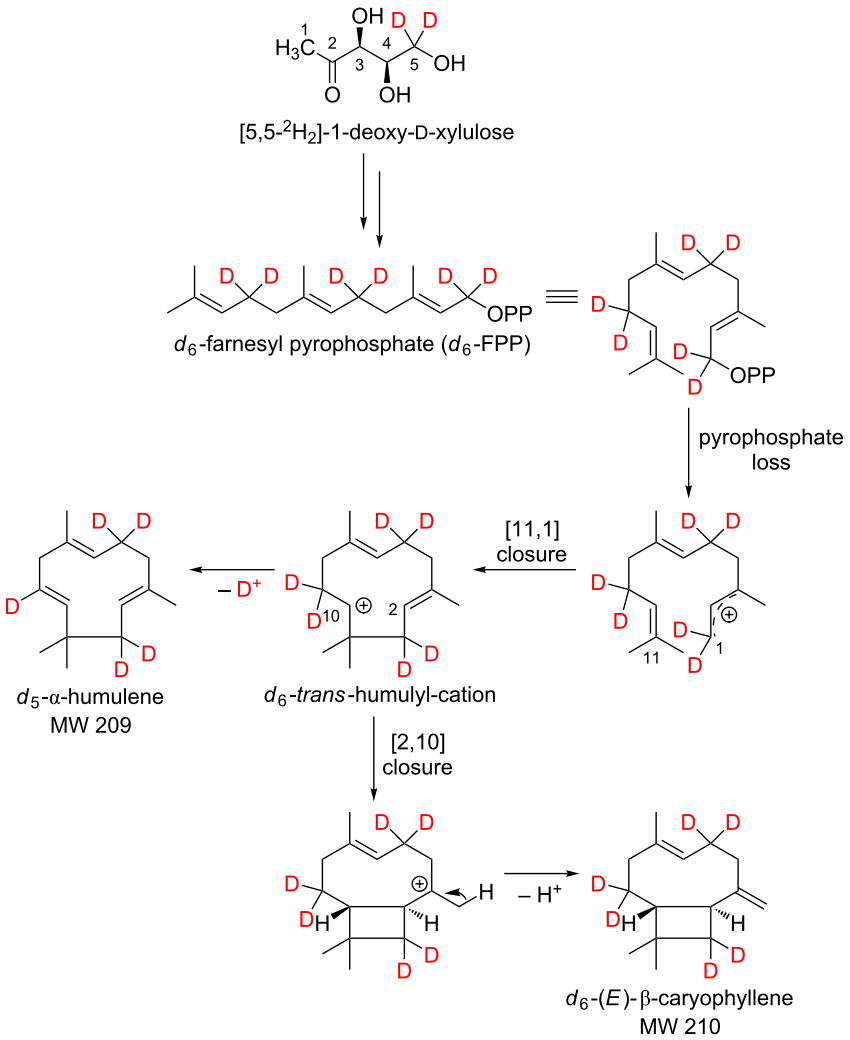


Figure 10: MS spectra and expected labeling patterns of genuine (d_0) and deuterium-labeled (d_9) guaiazulene after feeding experiments of grape berry exocarp using [6,6,6-$^2H_3$]-(±)-mevalonolactone.



Scheme 7: Possible synthesis of d_9-guaiazulene, d_9-δ-elemene, d_9-guaia-6,9-diene and d_9-δ-selinene via germacrene C after administration of [6,6,6-$^2H_3$]-(±)-mevalonolactone (d_3-MVL) to isolated exocarp of grape berries (Lemberger cultivar).



Scheme 8: Possible biosynthesis of d_6-(*E*)-β-caryophyllene and d_5-α-humulene starting from farnesyl pyrophosphate (FPP) after administration of [5,5-$^2H_2$]-1-deoxy-D-xylulose (d_2-DOX) to isolated exocarp of grape berries (Lemberger cultivar).

The assumed biosynthetic pathway for the formation of (*E*)-β-caryophyllene after administration of the precursor [6,6,6-$^2H_3$]-(±)-mevalonolactone (d_3-MVL) can be found in Supporting Information File 4.

Steele et al. described the formation of α-humulene via a *trans*-humulyl cation [9]. This metabolic pathway could also be confirmed by feeding experiments.

Conclusion

In summary, 25 sesquiterpene hydrocarbons were found in the exocarp of ripe grapes (*Vitis vinifera* L.) of the red wine variety Lemberger. The unequivocal identification of these structurally complex natural compounds was performed by combining multidimensional gas chromatography with in vivo labeling using the stable isotope-labeled precursors [5,5-$^2H_2$]-1-deoxy-D-xylulose (d_2-DOX) and [6,6,6-$^2H_3$]-(±)-mevalonolactone (d_3-MVL). Based on the recorded mass spectra of genuine sesquiterpene hydrocarbons and their deuterated analogs, the respective labeling patterns were deduced. Subsequently, it was examined whether the multistage cyclization reactions for the formation of sesquiterpene hydrocarbons postulated so far are consistent with our data. Contrary to previous assumptions, it could be shown that the tricyclic compounds α-copaene and β-copaene may be biosynthesized via (*S*)-(−)-germacrene D, whereas their stereoisomers α-ylangene and β-ylangene are formed via (*R*)-(+)-germacrene D. Furthermore, our results showed that the isomeric compounds α-cubebene and β-cubebene are also formed via (*S*)-(−)-germacrene D. In addition, 6 aromatic sesquiterpene hydrocarbons from the head-

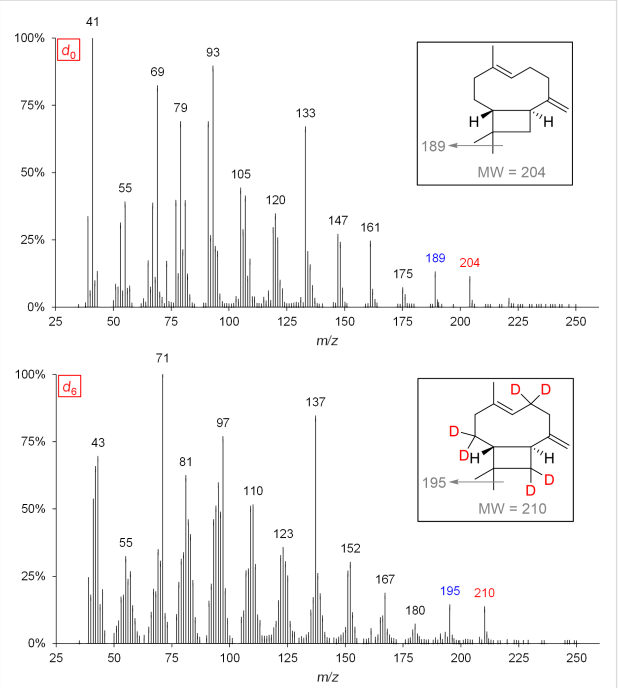


Figure 11: MS spectra and expected labeling patterns of d_0-(*E*)-β-caryophyllene and d_6-(*E*)-β-caryophyllene after successful administration of d_2-DOX to isolated exocarp of grapes of the red wine variety Lemberger.

space of grape berry exocarp of the red wine variety Lemberger were confirmed by in vivo labeling.

Experimental

Plant material

Ripe grape berries of the red wine variety Lemberger also called Blauer Limberger (*Vitis vinifera* subsp. *vinifera*, clone 1Gm) from the Hochschule Geisenheim University (Geisenheim, HE, Germany), harvested at the beginning of September 2018, were analyzed. The grapes were transported directly after the harvest to the University of Bonn (Bonn, NW, Germany) for sample preparation.

Chemicals

The sesquiterpene biosynthesis precursor [5,5-$^2H_2$]-1-deoxy-D-xylulose (d_2-DOX) was synthesized according to Meyer et al. [39], while [6,6,6-$^2H_3$]-(±)-mevalonolactone (d_3-MVL) was purchased from C/D/N Isotopes Inc. (Pointe-Claire, Quebec, Canada). Retention indices were calculated using a C_7–C_{30} saturated alkanes standard solution obtained from Sigma-Aldrich (St. Louis, MO, U.S.A.).

The analytical standard (+)-valencene (purity: 80.0%) was also acquired by Sigma-Aldrich (St. Louis, MO, U.S.A.). Guaiazulene (purity: 99.8%) was purchased from Alfa Aesar (Ward Hill, MA, U.S.A.). Water for dilutions was purified using a Milli-Q water purification system (Merck Millipore, Billerica, MA, U.S.A.). Parafilm® M was purchased from Sigma-Aldrich (St. Louis, MO, U.S.A.) to seal the sample vials to prevent the possible evaporation of volatile analytes.

SPME-Fiber

A divinylbenzene/carboxen/polydimethylsiloxane (DVB/CAR/PDMS, 50/30 µm) SPME fiber from Supelco (Bellefonte, PA, U.S.A.) was used.

Sample preparation

Sample preparation was based on a slightly modified method described by May, Lange and Wüst [5]. The exocarp of freshly harvested grapes was isolated with a scalpel. One hundred µL of a 0.1% d_3-MVL solution were pipetted into a 10 mL headspace vial and the isolated exocarp was inserted into the solution using tweezers. Similarly, the isolated exocarps of wineberries were placed in 100 µL of a 0.1% d_2-DOX solution. The filled 10 mL headspace vials were closed with screw cap, septum and parafilm. The fed sample material was incubated at room temperature for exactly 48 hours. Subsequently, the samples were stored at −20 °C until measurement. The SPME extraction conditions were based on a slightly modified method described by Welke et al. [17]. Prior to SPME extraction, the samples were held at an incubation temperature of 45 °C for 10 min without stirring. The extraction was carried out at 45 °C for 30 min. After sampling the headspace using a mixed fiber, the volatile and semi-volatile compounds were desorbed in the GC inlet at 250 °C for 5 min. In order to avoid carryover, the fiber was reconditioned for 15 min at 250 °C prior to each analysis. No sample carryover was observed.

GC×GC–TOF–MS analysis

All chromatographic separations were performed by using an Agilent 7890B gas chromatograph (Agilent Technologies, Palo Alto, CA, U.S.A.) equipped with a liquid nitrogen cryogenic modulator (Consumable-Free ZX2 thermal modulator) coupled to a time-of-flight mass spectrometer (Markes International Ltd, Llantrisant, RCT, UK). Liquid nitrogen was used for cryofocusing of the analytes eluting from the first dimension (1D) column, whereas heated air was used for releasing and reinjecting these compounds onto the second dimension (2D) column. A modulation period of 5 s and a hot jet duration of 0.35 s were used. The column set consisted of a 30 m × 0.25 mm i.d. × 0.25 µm d_f DB-WAX Ultra Inert (Agilent Technologies, Bellefonte, PA, U.S.A.) as a high polar primary column with poly(ethylene glycol) as stationary phase coupled to a 1.7 m × 0.10 mm i.d. × 0.10 µm d_f MEGA-17 MS FAST (MEGA s.n.c., Legnano, MI, IT) moderately polar second dimension with poly(50%-phenyl-50%-methylsiloxane) as stationary phase. The separation was performed using the

following temperature program: 1D oven ramp: 35 °C (5 min), 5 °C min^{-1} to 120 °C (0 min), 3 °C min^{-1} to 220 °C (5 min); 2D oven ramp: 60 °C (5 min), 5 °C min^{-1} to 145 °C (0 min), 3 °C min^{-1} to 245 °C (5 min). The injector was operated at 250 °C in the splitless mode. Helium was used as carrier gas at a constant flow of 1.0 mL min^{-1} and an initial inlet pressure of 166 kPa. The transfer line was maintained at 250 °C and the analytes were ionized by operating in electron impact (EI) mode at −70 eV. The ion source temperature was 250 °C and the detector voltage was set to −2236.8 V. Ions in the mass range 35–250 amu were acquired at a rate of 100 spectra s^{-1}. All samples were analyzed in triplicate.

Data analysis

Data processing was performed using the software ChromSpace (Markes International Ltd, Llantrisant, RCT, UK; version 1.2). First, a dynamic background compensation (dbc) was performed with a peak width of 2 s. The genuine sesquiterpene hydrocarbons were identified using the NIST Mass Spectral Search Program (NIST, Gaithersburg, MD, U.S.A.; version 2.2) by comparing the recorded mass spectra with those of the NIST database (when available), taking into account the match factors and reverse match factors. A series of *n*-alkanes (C_7–C_{30}) were analyzed. The retention times of the measured *n*-alkanes on the first GC column were used to calculate the retention indices *I* according to the method of Van Den Dool and Kratz for temperature-programmed gas chromatography [40]. Subsequently, the calculated retention indices of the identified sesquiterpenes were compared with literature values *I* (lit.). The confirmation of the found sesquiterpenes was done via the corresponding deuterium-labeled compounds starting from the precursors d_3-MVL and d_2-DOX after incubation.

Supporting Information

Supporting Information File 1

3D view of signals from genuine and deuterated α-cubebene, α-ylangene and δ-elemene after feeding experiments using d_3-MVL.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-190-S1.pdf]

Supporting Information File 2

Extracted ion chromatograms (EIC) of genuine (d_0), partially labeled (d_4, d_6) and fully deuterium-labeled (d_8) α-cubebene after administration of [6,6,6-$^2H_3$]-(±)-mevalonolactone (d_3-MVL) to isolated exocarp of grape berries (Lemberger cultivar).
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-190-S2.pdf]

Supporting Information File 3

Enlarged 2D chromatogram showing the separation of β-elemene, α-guaiene and β-ylangene.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-190-S3.pdf]

Supporting Information File 4

Putative formation pathway of d_8-(*E*)-β-caryophyllene and d_9-α-humulene starting from d_9-farnesyl pyrophosphate.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-190-S4.pdf]

Acknowledgements

The sample material was collected at the Department of Grapevine Breeding, Hochschule Geisenheim University. We would like to particularly thank E.-H. Rühl and B. May for their support. We also thank the DFG for financial support (INST 217/783-1 FUGG).

ORCID® iDs

Philipp P. Könen - https://orcid.org/0000-0002-4847-7807
Matthias Wüst - https://orcid.org/0000-0001-6808-5555

References

1. Coelho, E.; Rocha, S. M.; Delgadillo, I.; Coimbra, M. A. *Anal. Chim. Acta* **2006,** *563,* 204–214. doi:10.1016/j.aca.2005.11.018
2. Schreier, P.; Drawert, F.; Junker, A. *Z. Lebensm.-Unters. Forsch.* **1976,** *160,* 271–274. doi:10.1007/bf01132291
3. Breitmaier, E. *Terpenes. Flavors, fragrances, pharmaca, pheromones;* Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2006. doi:10.1002/9783527609949
4. Wood, C.; Siebert, T. E.; Parker, M.; Capone, D. L.; Elsey, G. M.; Pollnitz, A. P.; Eggers, M.; Meier, M.; Vössing, T.; Widder, S.; Krammer, G.; Sefton, M. A.; Herderich, M. J. *J. Agric. Food Chem.* **2008,** *56,* 3738–3744. doi:10.1021/jf800183k
5. May, B.; Lange, B. M.; Wüst, M. *Phytochemistry* **2013,** *95,* 135–144. doi:10.1016/j.phytochem.2013.07.021
6. Bloch, K. *Science* **1965,** *150,* 19–28. doi:10.1126/science.150.3692.19
7. Flesch, G.; Rohmer, M. *Eur. J. Biochem.* **1988,** *175,* 405–411. doi:10.1111/j.1432-1033.1988.tb14210.x
8. May, B.; Wüst, M. Sesquiterpene profiles of different grape varieties. In *Advances and Challenges in Flavor Chemistry & Biology;* Hofmann, T.; Meyerhof, W.; Schieberle, P., Eds.; Deutsche Forschungsanstalt für Lebensmittelchemie: Freising, Germany, 2011; pp 323–327.
9. Steele, C. L.; Crock, J.; Bohlmann, J.; Croteau, R. *J. Biol. Chem.* **1998,** *273,* 2078–2089. doi:10.1074/jbc.273.4.2078
10. Martin, D. M.; Aubourg, S.; Schouwey, M. B.; Daviet, L.; Schalk, M.; Toub, O.; Lund, S. T.; Bohlmann, J. *BMC Plant Biol.* **2010,** *10,* 226. doi:10.1186/1471-2229-10-226
11. Tantillo, D. J. *Nat. Prod. Rep.* **2011,** *28,* 1035–1053. doi:10.1039/c1np00006c
12. Bülow, N.; König, W. A. *Phytochemistry* **2000,** *55,* 141–168. doi:10.1016/s0031-9422(00)00266-1

13. Liu, Z.; Phillips, J. B. *J. Chromatogr. Sci.* **1991,** *29,* 227–231. doi:10.1093/chromsci/29.6.227
14. Meinert, C.; Meierhenrich, U. J. *Angew. Chem.* **2012,** *124,* 10610–10621. doi:10.1002/ange.201200842
15. Robinson, A. L.; Boss, P. K.; Heymann, H.; Solomon, P. S.; Trengove, R. D. *J. Chromatogr. A* **2011,** *1218,* 504–517. doi:10.1016/j.chroma.2010.11.008
16. Weldegergis, B. T.; Crouch, A. M.; Górecki, T.; de Villiers, A. *Anal. Chim. Acta* **2011,** *701,* 98–111. doi:10.1016/j.aca.2011.06.006
17. Welke, J. E.; Manfroi, V.; Zanus, M.; Lazarotto, M.; Alcaraz Zini, C. *J. Chromatogr. A* **2012,** *1226,* 124–139. doi:10.1016/j.chroma.2012.01.002
18. Birkemeyer, C.; Luedemann, A.; Wagner, C.; Erban, A.; Kopka, J. *Trends Biotechnol.* **2005,** *23,* 28–33. doi:10.1016/j.tibtech.2004.12.001
19. Paolini, J.; Costa, J.; Bernardini, A.-F. *J. Chromatogr. A* **2005,** *1076,* 170–178. doi:10.1016/j.chroma.2005.03.131
20. Gränicher, F.; Christen, P.; Kapetanidis, I. *Phytochemistry* **1995,** *40,* 1421–1424. doi:10.1016/0031-9422(95)00492-p
21. Blanc, M.-C.; Bradesi, P.; Gonçalves, M. J.; Salgueiro, L.; Casanova, J. *Flavour Fragrance J.* **2006,** *21,* 324–332. doi:10.1002/ffj.1605
22. Duquesnoy, E.; Dinh, N. H.; Castola, V.; Casanova, J. *Flavour Fragrance J.* **2006,** *21,* 453–457. doi:10.1002/ffj.1676
23. Cavalli, J.-F.; Tomi, F.; Bernardini, A.-F.; Casanova, J. *Flavour Fragrance J.* **2006,** *21,* 111–114. doi:10.1002/ffj.1531
24. Shimizu, Y.; Imayoshi, Y.; Kato, M.; Maeda, K.; Iwabuchi, H.; Shimomura, K. *Flavour Fragrance J.* **2009,** *24,* 251–258. doi:10.1002/ffj.1938
25. Cavalli, J.-F.; Tomi, F.; Bernardini, A.-F.; Casanova, J. *Flavour Fragrance J.* **2003,** *18,* 532–538. doi:10.1002/ffj.1263
26. Brat, P.; Rega, B.; Alter, P.; Reynes, M.; Brillouet, J.-M. *J. Agric. Food Chem.* **2003,** *51,* 3442–3447. doi:10.1021/jf026226y
27. Gonny, M.; Cavaleiro, C.; Salgueiro, L.; Casanova, J. *Flavour Fragrance J.* **2006,** *21,* 99–106. doi:10.1002/ffj.1527
28. Matucha, M.; Jockisch, W.; Verner, P.; Anders, G. *J. Chromatogr.* **1991,** *588,* 251–258. doi:10.1016/0021-9673(91)85030-j
29. Boland, W.; Garms, S. *Flavour Fragrance J.* **2010,** *25,* 114–116. doi:10.1002/ffj.1979
30. Rinkel, J.; Rabe, P.; Garbeva, P.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2016,** *55,* 13593–13596. doi:10.1002/anie.201608042
31. Arigoni, D. *Pure Appl. Chem.* **1975,** *41,* 219–245. doi:10.1351/pac197541010219
32. May, B. Biosynthese und Analytik von Sesquiterpenen in Weinbeeren (Vitis vinifera). Ph.D. Thesis, University of Bonn, Bonn, Germany, 2015. http://hss.ulb.uni-bonn.de/2016/4356/4356.pdf
33. Schmidt, C. O.; Bouwmeester, H. J.; Franke, S.; König, W. A. *Chirality* **1999,** *11,* 353–362. doi:10.1002/(sici)1520-636x(1999)11:5/6<353::aid-chir2>3.0.co;2-l
34. Lodewyk, M. W.; Gutta, P.; Tantillo, D. J. *J. Org. Chem.* **2008,** *73,* 6570–6579. doi:10.1021/jo800868r
35. Yoshihara, K.; Ohta, Y.; Sakai, T.; Hirose, Y. *Tetrahedron Lett.* **1969,** *10,* 2263–2264. doi:10.1016/s0040-4039(01)88136-3
36. Faraldos, J. A.; Wu, S.; Chappell, J.; Coates, R. M. *Tetrahedron* **2007,** *63,* 7733–7742. doi:10.1016/j.tet.2007.04.037
37. Dickschat, J. S.; Celik, E.; Brock, N. L. *Beilstein J. Org. Chem.* **2018,** *14,* 900–910. doi:10.3762/bjoc.14.77
38. Poulose, A. J.; Croteau, R. *Arch. Biochem. Biophys.* **1978,** *187,* 307–314. doi:10.1016/0003-9861(78)90039-5
39. Meyer, O.; Hoeffler, J.-F.; Grosdemange-Billiard, C.; Rohmer, M. *Tetrahedron* **2004,** *60,* 12153–12162. doi:10.1016/j.tet.2004.10.016
40. van Den Dool, H.; Dec. Kratz, P. *J. Chromatogr.* **1963,** *11,* 463–471. doi:10.1016/s0021-9673(01)80947-x

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Isolation and characterisation of irinans, androstane-type withanolides from *Physalis peruviana* L.

Annika Stein[‡1], Dave Compera[‡1], Bianka Karge[2], Mark Brönstrup[1,2] and Jakob Franke[*1]

Full Research Paper

Open Access

Address:
[1]Centre of Biomolecular Drug Research, Leibniz University Hannover, Schneiderberg 38, 30167 Hannover, Germany and [2]Helmholtz Centre for Infection Research, Inhoffenstrasse 7, 38124 Braunschweig, Germany

Email:
Jakob Franke[*] - jakob.franke@oci.uni-hannover.de

* Corresponding author ‡ Equal contributors

Keywords:
androstanes; *Physalis peruviana*; steroids; structure elucidation; withanolides

Beilstein J. Org. Chem. **2019,** *15,* 2003–2012.
doi:10.3762/bjoc.15.196

Received: 31 May 2019
Accepted: 07 August 2019
Published: 23 August 2019

This article is part of the thematic issue "Terpenes".

Guest Editor: J. S. Dickschat

Abstract

Withanolides are steroidal lactones widespread in Nightshade plants with often potent antiproliferative activities. Additionally, the structural diversity of this compound class holds much potential for the discovery of novel biological activity. Here, we report two newly characterised withanolides, named irinans, from *Physalis peruviana* with highly unusual truncated backbones that resemble mammalian androstane sex hormones. Based on biomimetic chemical reactions, we propose a model that links these compounds to withanolide biosynthesis. Irinans have potent antiproliferative activities, that are however lower than those of 4ß-hydroxywithanolide E. Our work establishes androwithanolides as a new subclass of withanolides.

Introduction

Traditional medicine has long been a source of inspiration for modern drug research. An important example is *Withania somnifera,* also known as ashwaghanda or Indian ginseng, which has been used in Ayurvedic medicine to treat a large variety of ailments [1]. Extensive studies revealed withanolides, a class of steroidal lactones, to be primarily responsible for the medicinal effects [1,2]. A large range of pharmacological properties has been assigned to withanolides, with antiproliferative activities being the most potent ones [1]. Withanolides have been also discovered in numerous genera other than *Withania*, for example *Datura*, *Dunalis*, *Iochroma*, *Jaborosa*, *Lycium* and *Physalis* [3], resulting in more than 300 known representatives [3]. *Physalis peruviana* is a withanolide producer of particular relevance as it is widely cultivated for its edible berries [4]. So far, several withanolides have been reported from *P. peruviana* and other *Physalis* species, most prominently physalins, perulactones and 4ß-hydroxywithanolide E (**1**) [5-16]. As part of our ongoing programme focussed on the biochemistry of withanolides, our aim was to gain further insights into the withanolide profile of *P. peruviana*. Here we report irinans A (**2**) and B (**3**), two unusual truncated withanolides with androstane backbones. We show that oxidative, but not acidic or basic conditions

enable conversion of the putative precursor 4ß-hydroxywithanolide E (**1**) to irinan A (**2**). Based on this intrinsic reactivity we propose a biosynthetic model that will serve as further guidance for elucidating the enzymatic basis of androstane formation in plants in the future.

Results and Discussion

To isolate withanolides from *P. peruviana*, we used a purification strategy based on previous reports [17-19]. Nine weeks old whole *P. peruviana* plants (140 g) were extracted with H_2O/MeOH (3:1) and divided into fractions soluble in petroleum ether, chloroform, and *n*-butanol, respectively. The chloroform fraction was further separated by flash chromatography on a C_{18} stationary phase, resulting in three major subfractions F1–F3. Final purification by preparative HPLC followed by NMR analysis revealed 4β-hydroxywithanolide E (**1**) as the major compound (50 mg) as well as the known metabolites withanolide E (**4**), withanolide F (**5**) and perulactone H (**6**) by comparison to literature data (Figure 1) [9,20].

Two additional compounds attracted our attention based on their unusual ^{1}H NMR spectra (Table 1). Both showed two multiplets in the olefinic region, which are highly characteristic for withanolides with A-ring Michael acceptors. However, compared to other withanolides, several signals were missing. Typically, withanolides show five singlets for methyl groups in the aliphatic region, as well as the H-22 oxymethine proton of the lactone moiety. Surprisingly, both compounds showed only two putative methyl signals, and no signal which might correspond to H-22. Thus, we reasoned that both unknown compounds might be truncated withanolide-like compounds.

HRESIMS suggested a sum formula of $C_{19}H_{24}O_5$ for the first compound, which was supported by the ^{13}C spectrum (Table 1). By comparing the spectrum to NMR data of other withanolides, we quickly identified the Michael system in ring A based on two olefinic protons (δ_H 6.94 and 6.22 ppm), a secondary alcohol at C-4 (δ_H 3.79 ppm), a 5,6-epoxide (δ_H 3.37 (H-6)), and a tertiary alcohol at C-14 (δ_C 80.9 ppm). COSY correlations supported by HMBC analysis (Figure 2A) revealed an intact ABCD ring system with a substitution pattern identical to 4β-hydroxywithanolide E (**1**). Only a single, striking difference was noted: C-17 was shifted from 87.8 to 218.0 ppm, strongly suggesting the presence of a ketone instead of an alcohol. In agreement with the predicted sum formula and the absence of all side chain carbons, this completed the structure of the first unknown compound, which we named irinan A (**2**, Figure 1).

The second unknown compound had a sum formula of $C_{19}H_{24}O_3$ based on HRESIMS and ^{13}C NMR (Table 1). In contrast to the first compound, no epoxide and no secondary alcohol at C-4 was present, in agreement with the different elemental composition. Instead, ^{13}C NMR indicated a double bond at C5–C6 (δ_C 135.7 and 124.3 ppm). Otherwise, all spin systems and correlations indicated a typical withanolide ABCD

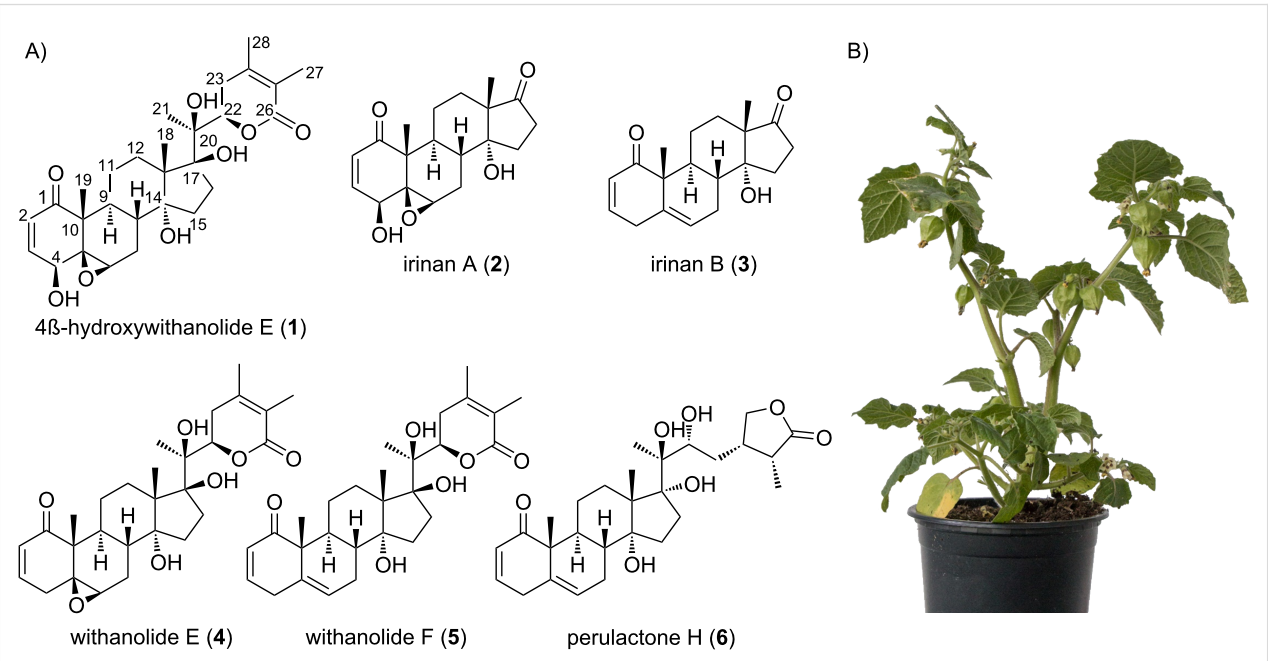


Figure 1: Withanolides from *Physalis peruviana*. A) Structures of the newly characterised truncated withanolides irinan A (**2**) and B (**3**) with an androstane backbone, together with the known withanolides **1** and **4**–**6** isolated from *P. peruviana*. B) *P. peruviana* plant.

Table 1: ^{13}C and ^{1}H NMR data ($CDCl_3$, 500 MHz, 298 K) of irinans A (**2**) and B (**3**) in comparison to the known compound 4β-hydroxywithanolide E (**1**, $CDCl_3$, 400 MHz, 298 K, δ in ppm, *J* in Hz). For carbon numbering see Figure 1 and Figure 2.

	^{13}C			^{1}H		
position	**1**	**2**	**3**	**1**	**2**	**3**
1	201.9	202.0	203.8	–	–	–
2	133.2	132.3	128.1	6.22 (1H, d, 9.9)	6.22 (1H, d, 10.0)	5.91 (1H, ddd, 10.0, 3.1, 1.2)
3	141.4	142.1	145.3	6.92 (1H, dd, 9.9, 6.1)	6.94 (1H, dd, 10.0, 5.8)	6.79 (1H, ddd, 10.0, 5.0, 2.6)
4	70.4	69.9	33.6	3.74 (1H, d, 6.1)	3.79 (1H, dd, 5.8, 2.4)	3.31 (1H, dddddd, 21.3, 2.8, 2.8, 2.8, 2.8, 2.8)[a] 2.88 (1H, dd, 21.2, 4.9)
5	64.2	63.9	135.7	–	–	–
6	63.1	63.1	124.3	3.28 (1H, br s)	3.37 (1H, m)	5.64 (1H, dt, 5.7, 2.0)
7	26.0	24.9	24.1	2.03 (2H, m)	2.11 (1H, dt, 14.2, 3.1) 1.84 (1H, ddd, 14.1, 11.7, 1.4)	2.08 (1H, m) 1.95 (1H, m)
8	34.3	32.6	35.5	1.83 (1H, m)	1.90 (1H, m)	1.88 (1H, m)
9	36.7	38.1	37.1	1.69 (1H, m)	1.51 (1H, m)	2.10 (1H, m)
10	47.9	47.8	50.9	–	–	–
11	21.5	20.4	21.6	1.72 (1H, m) 1.56 (1H, m)	1.91 (1H, m) 1.46 (1H, m)	2.34 (1H, m) 1.52 (1H, m)
12	29.8	24.3	25.0	2.25 (1H, m) 1.28 (1H, m)	1.66 (1H, d, 13.2) 1.55 (1H, m)	1.86 (1H, m) 1.63 (1H, m)
13	54.6	52.6	52.5	–	–	–
14	81.9	80.9	81.0	–	–	–
15	32.5	30.0	29.9	1.66 (1H, m) 1.59 (1H, m)	1.92 (2H, m)	1.96–1.85 (2H, m)
16	38.0	33.1	33.1	2.72 (1H, m) 1.45 (1H, m)	2.44 (1H, ddd, 18.9, 7.6, 4.1) 2.33 (1H, dt, 18.8, 8.8)	2.35–2.46 (2H, m)
17	87.8	218.0	218.5	–	–	–
18	20.4	17.9	18.1	1.07 (3H, s)	1.01 (3H, s)	1.05 (3H, s)
19	16.9	17.8	19.2	1.42 (3H, s)	1.45 (3H, s)	1.27 (3H, s)
20	79.2	–	–	–	–	–
21	19.8	–	–	1.42 (3H, s)	–	–
22	79.7	–	–	4.88 (1H, dd, 11.8, 5.3)	–	–
23	34.4	–	–	2.51 (2H, m)	–	–
24	150.8	–	–	–	–	–
25	121.6	–	–	–	–	–
26	166.0	–	–	–	–	–
27	12.5	–	–	1.88 (3H, s)	–	–
28	20.8	–	–	1.94 (3H, s)	–	–
14-OH				n.d.	1.41 (1H, br s)	1.41 (1H, br s)
4-OH				n.d.	2.57 (1H, d, 2.50)	-

[a]Apparent dsext. See Figure S19 (Supporting Information File 1) for details. n.d. not detected.

ring system. Again, a carbon with a distinct downfield shift of 218.4 ppm was found, demonstrating the presence of a ketone at C-17. Highly unusually, H-4β appeared as a doublet of sextets (1:5:10:10:5:1) (dsext) in the ^{1}H NMR spectrum. This multiplet was explained as "dddddd" by a total of six COSY correlations (Figure S19, Supporting Information File 1). The resulting compound was named irinan B (**3**, Figure 1). A compound of putatively identical structure was isolated from *P. peruviana* before, but only fragmentary physicochemical data has been reported so far [14].

To elucidate the relative stereochemistry of irinans A (**2**) and B (**3**), we analysed NOESY data (Figure 2). In the case of irinan A (**2**), the β configuration of OH-4 was deduced by the NOESY correlation OH-4/CH_3-19. OH-14 was assigned as α based on the correlations OH-14/H-12α and H-9/H-12α. The 5,6-epoxide was determined as β by a correlation from H-6 to H-3. These assignments are in complete agreement with the relative stereochemistry of 4ß-hydroxywithanolide E (**1**). In irinan B, the configuration of OH-14 could not be unambiguously inferred from NOE data due to the signal overlap of H-15 with H-12 and

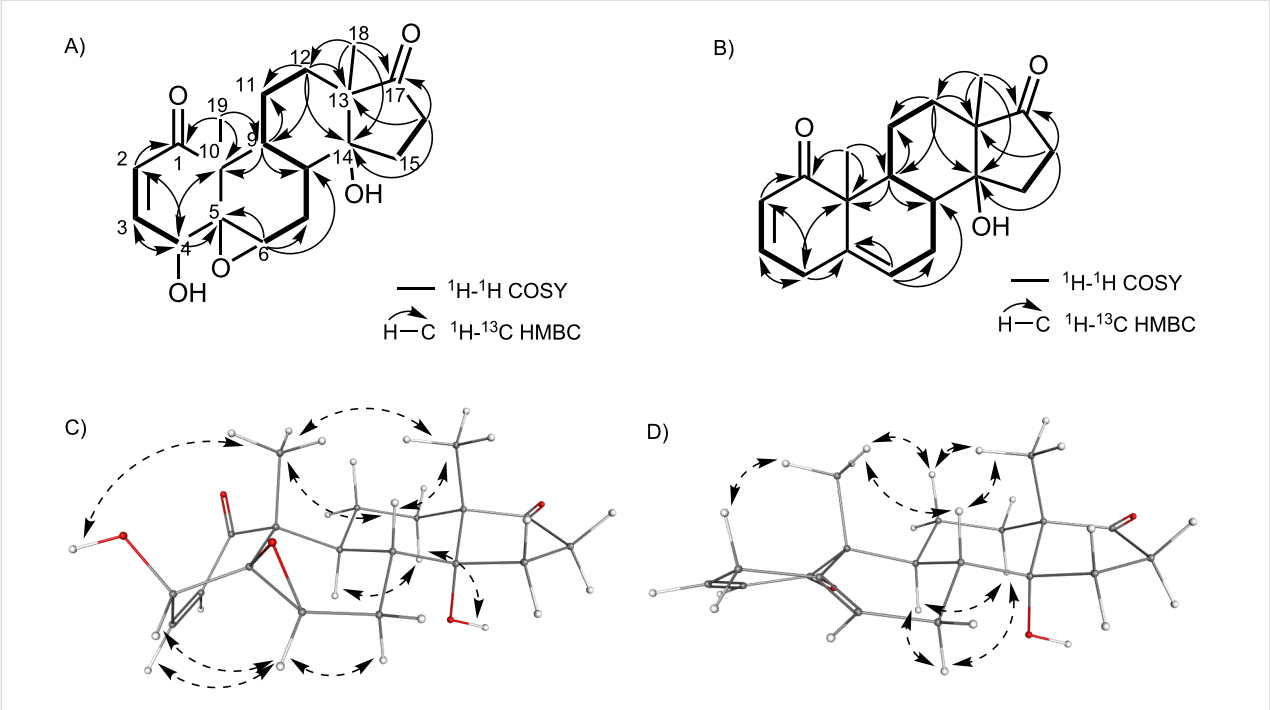


Figure 2: Key NMR correlations. (A) COSY and HMBC correlations for irinan A (**2**). (B) COSY and HMBC correlations for irinan B (**3**). (C) Key NOESY correlations for irinan A (**2**). (D) Key NOESY correlations for irinan B (**3**).

other protons. As an alternative, OH-14 α configuration was deduced from the chemical shifts of C-12 and C-9, which experience a strong shielding γ-gauche effect for OH-14α configurations [21]. These data indicate a relative stereochemistry of irinan B matching withanolide F (**5**).

Irinans represent highly unusual withanolide derivatives, as they lack the side-chain lactone ring that is a common structural feature of virtually all known withanolides [3], but possess an androstane backbone instead. While androstanes such as androsterone (**7**) are well-known human sex hormones (Figure 3A) [22], their occurrence in plants is rare [23-26]. Only a single withanolide androstane has been fully characterised before, cinedione (**8**), isolated from *Physalis cinerascens* (Figure 3A) [23]. We propose the name androwithanolides for this withanolide subclass, which so far appears to be characteristic of *Physalis* species.

The biosynthesis of androstanes in mammals requires three enzymatic steps starting from cholesterol (**9**, Figure 3B) [27]. Cholesterol (**9**) is converted to pregnenolone (**10**) by the cytochrome P450 cholesterol side-chain cleavage enzyme (P450scc), which cleaves the C20–C22 bond [27]. Then, the bifunctional P450c17 acts as a 17α-hydroxylase and 17,20-lyase to give rise to androstanes [27]. Related enzymes have not been reported from plants. We searched transcriptome data of *P. peruviana* for putative homologues of these enzymes [28]. The best hits only had amino acid sequence identities of 22–28%, indicating that no P450 enzymes of these clans exist in *P. peruviana*. Although enzymes with similar catalytic activity might have evolved convergently in plants, the different substitution pattern in the side chain suggests that a side-chain cleavage mechanism distinct from mammals is involved. While the order of oxidative steps in withanolide biosynthesis is still completely elusive [29], we propose that this fragmentation occurs at a late stage, when most typical withanolide functionalisations have already been introduced. Indeed, irinan A (**2**), irinan B (**3**) and cinedione (**8**) can be directly linked to the known withanolides 4ß-hydroxywithanolide E (**1**), withanolide F (**5**) and withanolide S [23], respectively (Figure S20, Supporting Information File 1). If the fragmentation occurred early in the biosynthesis, this would imply that several biosynthetic enzymes have to tolerate substrates without the lactone side chain. We therefore propose that the side-chain cleavage enzyme in withanolide biosynthesis acts at a late stage, using common pathway end products such as 4ß-hydroxywithanolide E (**1**) as its substrates. Two mechanisms are conceivable for this transformation (Figure 3C): A non-oxidative Grob fragmentation could make use of a push–pull mechanism between C-17 and C-22, building on acid–base catalysis. Alternatively, an enzyme could cleave the C17–C20 diol oxidatively. Several P450 enzymes have been reported to be capable of cleaving diols, presumably via a ferric peroxo intermediate (Figure 3C) [30,31].

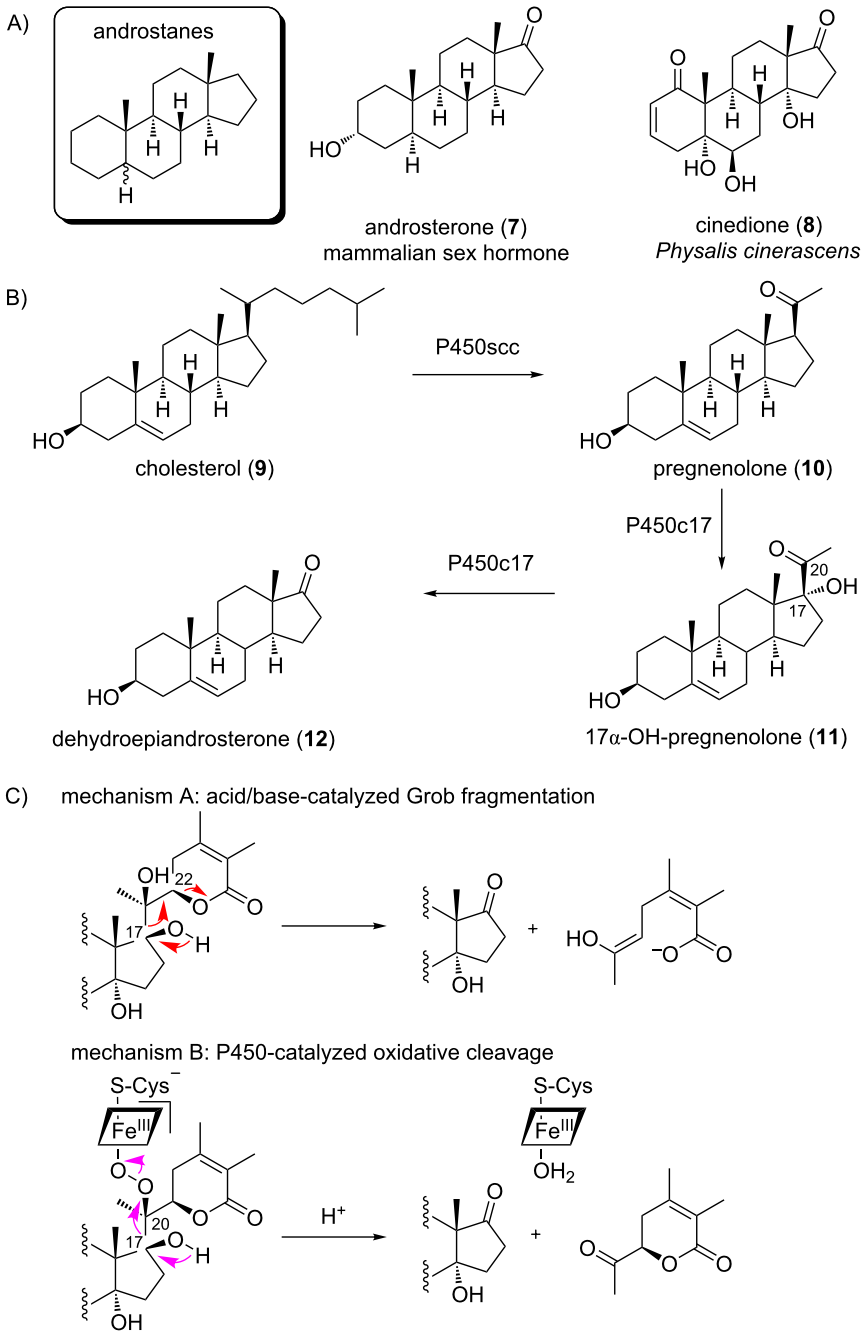


Figure 3: Structures and biosynthesis of androstanes. (A) Androstane backbone and androsterone (**7**) as a typical mammalian sex hormone. Cinedione (**8**) is the only other fully characterised androwithanolide known. (B) Biosynthesis of androstanes in mammals. (C) Possible cleavage mechanisms involved in androwithanolide biosynthesis in plants.

To gain further insights into the biosynthetic route and to exclude that androwithanolides are isolation artefacts [32], we exposed 4ß-hydroxywithanolide E (**1**) as the likely precursor to irinan A (**2**) to various chemical conditions (Figure 4). In general, **1** was stable in all solvents tested, namely chloroform, methanol, DMSO and acetonitrile (data not shown). Treatment with acid at pH 3 caused no reaction at all when heating up to 70 °C (Figure 4A). At pH 0, several unidentified compounds appeared, but not irinan A (**2**). In basic conditions, only a single unidentified product was formed at pH 11 and 70 °C. Next, we tested whether **1** could be oxidatively cleaved [33]. Incubation of **1** with $NaIO_4$ at room temperature did not result in any reaction (data not shown). However, although it has been reported that periodates are not capable of cleaving ditertiary glycols

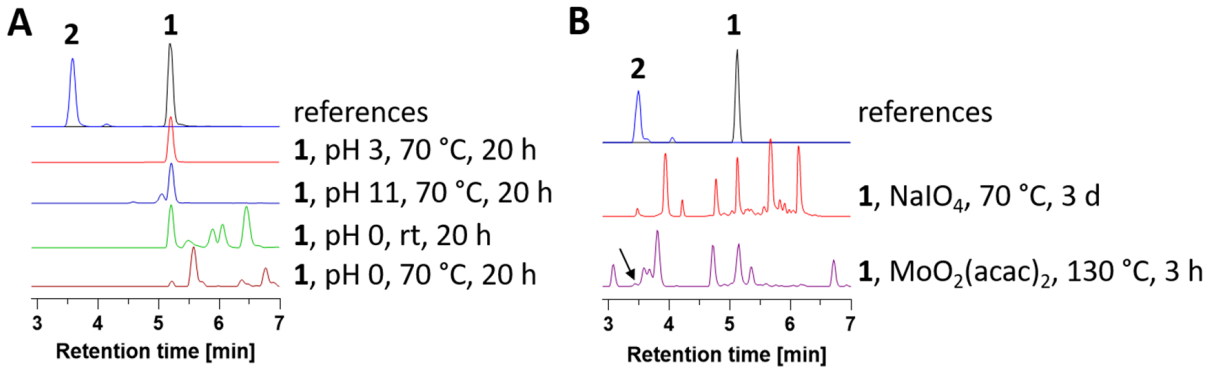


Figure 4: Intrinsic reactivity of 4ß-hydroxywithanolide E (**1**) under acidic/basic and oxidative conditions, respectively. (A) LC–MS chromatograms (ELS detection) of **1** incubated at different pH values. (B) LC–MS chromatograms (UV detection at 200–400 nm) of **1** treated with different oxidative reagents. The formation of **2** in the $NaIO_4$ reaction was confirmed by NMR analysis. See also Figure S21 in Supporting Information File 1 for extracted ion chromatograms (EICs).

[33,34], we noted formation of small quantities of irinan A (**2**) when performing the reaction at 70 °C (Figure 4B and Figure S21 in Supporting Information File 1). The identity of irinan A (**2**) was verified by isolation of the corresponding compound by preparative HPLC (4% yield) followed by NMR analysis. This result confirms our NMR-based stereochemical assignment and unambiguously links irinan A (**2**) to 4ß-hydroxywithanolide E (**1**). We also performed an oxidative cleavage reaction with catalytic amounts of $MoO_2(acac)_2$ in DMSO as described by García et al. [34], which also led to the formation of trace amounts of irinan A (**2**). Our experiments suggest that irinan A (**2**) and most likely all androwithanolides are not isolation artefacts but true natural products, which require an oxidative enzyme to facilitate the C–C bond cleavage. Future studies will shed light on the enzymatic basis of androwithanolide formation.

Considering the potent bioactivities of androstanes as well as withanolides, we wondered whether the loss of the side-chain lactone would negatively impact the antiproliferative activity. Irinan A (**2**) and B (**3**) together with 4ß-hydroxywithanolide E (**1**) as a positive control were evaluated against a panel of four cell lines (Table 2). In our assays we observed decreasing activities during the third and fourth replicates, resulting in large standard deviations and potentially indicating limited stability of these compounds. Nonetheless, EC_{50} values of 4ß-hydroxywithanolide E (**1**) were in good agreement with previously published values [14,35,36]. Irinans A (**2**) and B (**3**) were 1.3 to 10-fold less active than 4ß-hydroxywithanolide E (**1**), with the exception of irinan B (**3**) in A549 cells, which was equipotent. However, irinan A (**2**) and B (**3**) samples had a purity of 90% and 80%, respectively. We therefore cannot exclude that unidentified impurities, which could not be removed by repeated preparative HPLC, obscure the true EC_{50} values of irinans. We conclude that irinans possess potent antiproliferative activity, that is however reduced compared to 4ß-hydroxywithanolide E (**1**). Our results demonstrate the importance of the lactone side chain for bioactivity.

Conclusion

We have discovered and characterised irinans A and B, two new withanolides from *P. peruviana* with truncated backbones. They resemble mammalian sex hormones of the androstane class. The relative stereochemistry was elucidated based on NOESY analysis. Chemical studies support a model that these compounds are formed by an oxidative process. We

Table 2: Antiproliferative activities in different cell lines. Data indicate EC_{50} values ± SD in µM. A549 = human lung carcinoma; L929 = mouse fibroblast; KB-3-1 = human cervix carcinoma; MCF-7 = human breast cancer cell line.

Compound	A549	L929	KB-3-1	MCF-7
4ß-hydroxywithanolide E (**1**)	3.74 ± 0.50	0.27 ± 0.30	1.11 ± 0.98	10.65 ± 6.18
irinan A (**2**)[a]	5.01 ± 5.27	2.29 ± 0.88	4.62 ± 5.76	17.88 ± 7.27
irinan B (**3**)[b]	3.45 ± 1.91	1.68 ± 1.78	2.40 ± 2.32	13.56 ± 9.18
staurosporine (positive control)	1.19 ± 0.99	<0.003	0.04 ± 0.01	0.16 ± 0.02
auranofin (positive control)	>7.03	2.35 ± 0.83	1.59 ± 0.37	2.06 ± 0.60

[a]Estimated 90% purity based on ^{1}H NMR. [b]Estimated 80% purity based on ^{1}H NMR.

propose the name androwithanolides for this withanolide subclass.

Experimental

General experimental procedures

Seeds of *Physalis peruviana* were obtained from FloraSelf, Sperli and Quedlinburger Saatgut. Plants were initially grown in seed starter soil (Kölle's Beste Anzuchterde) and later transferred to potting soil (Kölle's Beste Pflanzerde). Plants were grown under LED illumination (SANlight S2W) at 350 µmol s^{-1} m^{-2} PPFD with a 12 h photoperiod and at 18–25 °C without temperature and humidity control. Plants were watered twice per week with tap water as needed.

NMR spectra were recorded using Bruker Ascend™ 400 or DRX 500 MHz spectrometers operating at 400 and 500 MHz for ^{1}H NMR and at 100 and 125 MHz for ^{13}C NMR where $CDCl_3$ was used as solvent. Chemical shifts were referenced relative to the residual solvent signal of $CDCl_3$ (δ_H = 7.26 ppm, δ_C = 77.16 ppm) and expressed in δ values (ppm), with coupling constants reported in Hz. Analysis was conducted with TopSpin (Version 4.0.6, Bruker). ATR-IR analysis was performed for the range of 400–4000 cm^{-1} using a Shimadzu IRAffinity 1S spectrometer with samples dissolved in chloroform. Optical rotations were measured with a Perkin Elmer 341 polarimeter. Using methanol as solvent, the wavelength for maximum absorption was determined on a Jasco V-630 spectrophotometer. Flash purification was performed on a Biotage Isolera One using columns described below. HRMS measurements were carried out on a Waters Alliance 2695 HPLC coupled to a Micromass LCT Premier mass spectrometer.

For analytical and preparative LC–MS a Waters instrument was used consisting of a Waters 2767 autosampler, Waters 2545 pump system, Waters 2998 diode array detector, Waters 2424 ELS detector, and a Waters SQ Detector 2 for mass spectrometry in ESI^+ and ESI^- modes between *m/z* 150 and 1000. In analytical mode, a Phenomenex Kinetex column (2.6 µm, C_{18}, 100 Å, 4.6 × 100 mm) was used with a gradient of [solvent A: H_2O + 0.05% formic acid; solvent B: acetonitrile + 0.045% formic acid; gradient: 10% to 90% B over 10 min, 1 mL/min]. Samples were dissolved to a concentration of 10 mg/mL in MeOH and 20 µL injected. In preparative mode, a Phenomenex Kinetex Axia column (5 µm, C_{18}, 100 Å, 21.2 × 250 mm) equipped with a Phenomenex Security Guard precolumn (Luna, C_5, 300 Å) was used in combination with the separation gradient described below.

Extraction and isolation of withanolides

140 g of 9 weeks old, whole *Physalis peruviana* plants were frozen in liquid nitrogen and ground to a fine powder. The powder was extracted with 500 mL H_2O/MeOH (3:1) at room temperature for 3 h. After filtration and evaporation of the solvent under reduced pressure, the crude extract was resuspended in 300 mL H_2O and defatted with 300 mL petroleum ether. The remaining aqueous layer was further extracted with 2 × 300 mL $CHCl_3$ followed by 2 × 300 mL *n*-BuOH. This resulted in a 660 mg petroleum ether fraction, 386 mg $CHCl_3$ fraction and 1174 mg *n*-BuOH fraction.

The $CHCl_3$ fraction was separated via reversed-phase flash chromatography (Biotage SNAP KP-C18-HS 30 g column) with a H_2O/MeOH gradient. Samples were adsorbed onto Celite under reduced pressure for dry loading. A gradient from 30% to 95% MeOH was used. Fractions were pooled guided by UV maximum absorbance to form main fraction F1 (subfractions 1–21, 102 mg), F2 (subfractions 22–31, 12 mg) and F3 (subfractions 32–45, 48 mg). No withanolides were detected in F2 based on LC–MS analysis and therefore discarded.

Fraction F1 was further separated by preparative LC–MS. The sample was dissolved in MeOH to a concentration of 15 mg/mL. 100 µL was injected per run. A separation gradient was used [solvent A: H_2O + 0.05% formic acid; solvent B: acetonitrile + 0.045% formic acid; gradient: 10% to 90% B over 10 min, 20 mL/min]. The post-column flow was split (100:1) and the minority flow made up to 1 mL/min with MeOH + 0.045% formic acid for in-line analysis by UV, ELSD and MS. The majority flow was collected. The following peaks were collected and identified by NMR: t_R = 5.8–6.0 min (irinan A (**2**), 6 mg); 7.2–8.0 min (4β-hydroxywithanolide E (**1**), 49 mg); 9.0–9.2 min (irinan B (**3**), 1 mg). The collected fractions were evaporated under reduced pressure using a Christ RVC 2-25 CDplus rotational vacuum concentrator.

Main fraction F3 was also separated by preparative LC–MS as described above, yielding the known compounds withanolide E (**4**) (t_R = 7.3–7.6 min, 6 mg), perulactone H (**6**) (t_R = 7.6–7.8 min, 9 mg) and withanolide F (**5**) (t_R = 7.8–8.5 min, 9 mg) which were identified by NMR [20].

Analytical data

4β-Hydroxywithanolide E (**1**) was isolated as a white crystalline powder. NMR data of **1** is listed in Table 1. All spectroscopic properties matched literature data [20].

Irinan A (**2**): white crystalline powder; $[\alpha]_D^{20}$ +10.48 (β = 0.62; MeOH); UV (MeOH) λ_{max} (log ε) 239 nm (3.93); IR (ATR, $CHCl_3$) ν_{max}: 3460, 2967, 2930, 1734, 1674, 1454, 1373, 1092, 1036, 986, 922, 754 cm^{-1}; for ^{1}H and ^{13}C data see Table 1; HRESIMS *m/z*: $[M + Na]^+$ calcd for $C_{19}H_{24}O_5Na^+$, 355.1516; found, 355.1519.

Irinan B (**3**): white crystalline powder; $[\alpha]_D^{20}$ –10.00 (β = 0.06; MeOH); UV (MeOH) λ_{max} (log ε) 251 nm (3.94); IR (ATR, $CHCl_3$) ν_{max}: 3402, 2955, 2930, 1682, 1383, 1259, 1215, 1136, 1088, 1016, 966, 806, 748 cm^{-1}; for ^1H and ^{13}C data see Table 1; HRESIMS *m/z*: $[M + Na]^+$ calcd for $C_{19}H_{24}O_3Na^+$, 323.1618; found, 323.1626

BLAST search of known androstane biosynthesis enzymes

The known androstane biosynthesis enzymes *Homo sapiens* P450scc (UniProtKB accession P05108) and *Homo sapiens* P450c17 (P05093) were used to search reported *Physalis peruviana* transcriptome data [28] via the tBLASTn algorithm. Both enzymes yielded several full-length hits with amino acid sequence identities of 22–28%.

Oxidative cleavage of 4β-hydroxywithanolide E (**1**) to irinan A (**2**) by $NaIO_4$

57.8 mg of $NaIO_4$ (270.2 µmol, 7.0 equiv) in 400 µL hot H_2O was added to 19.4 mg 4ß-hydroxywithanolide E (**1**, 38.6 µmol, 1.0 equiv) in 1 mL MeOH. The reaction was incubated at 70 °C for 72 h in a heat block with shaking at 1000 rpm. After that time a peak with *m/z* 315 corresponding to $[M + H - H_2O]^+$ with a retention time of 3.5 min was observed by LC–MS, co-eluting with authentic irinan A (**2**). The reaction mixture was separated by preparative LC–MS as described above to give a white crystalline powder (0.5 mg, 4%), which was confirmed to be irinan A (**2**) by ^1H NMR spectroscopy.

Oxidative cleavage of 4β-hydroxywithanolide E (**1**) to irinan A (**2**) by $MoO_2(acac)_2$

2 µL of a $MoO_2(acac)_2$ stock solution in DMSO (100 µg/µL, 0.6 µmol, 0.02 equiv) was added to 14.6 mg hydroxywithanolide E (**1**, 29.0 µmol, 1.0 equiv) in 100 µL DMSO. The reaction was incubated at 130 °C for 3 h in an oil bath with stirring at 400 rpm. After that time a peak with *m/z* 315 corresponding to $[M + H - H_2O]^+$ with a retention time of 3.5 min was observed, co-eluting with authentic irinan A (**2**). H_2O (5 mL) and $CHCl_3$ (5 mL) were added to the reaction mixture. The layers were separated and the aqueous phase was extracted with chloroform (3 × 5 mL). Combined organic layers were washed with water (5 mL), dried over $MgSO_4$, filtered and the solvent was evaporated under reduced pressure. The resulting crude reaction product was then analysed by analytical HPLC as described above.

Antiproliferative assays

The effect of compounds on cell viability was probed with a WST-1 test using the procedure of Ishiyama et al. [37] as modified by Sasse et al. [38]. The following cell lines were used: mouse fibroblast cell line L929 (DSM ACC 2), human cervix carcinoma cell line KB-3-1 (DSM ACC 158), the human lung carcinoma cell line A549 (DSMZ ACC 107) and human breast cancer cell line MCF-7 (DSM ACC 115). The subconfluent cells were briefly washed with Earle's Balanced Salt Solution (Gibco) without Ca and Mg, trypsinized and re-suspended in Dulbecco's modified eagle's medium that contained 5% fetal bovine serum (FBS; L929, KB-3-1, A549) or Roswell Park Memorial Institute medium that contained 5% FBS, 0.5% Minimum Essential Medium Non-Essential Amino Acids, Gibco (MEM NEAA), 0.5% GlutaMAX (Gibco) and insulin at 5 µg/mL (MCF-7). 25 µL of serial dilutions of the test compounds (64–0.06 µg/mL, that were made with a pipetting robot (epMotion, Eppendorf, Hamburg, Germany), were added to 25 µL aliquots of a cell suspension (1500 cells for KB-3-1, L929 and A549, 3000 cells for MCF-7) in 384 well microtiter plates. Blank and solvent controls were incubated under identical conditions. After an incubation period of 5 days, 3 µL WST-1 (ready to use solution by Roche) was added. The incubation time of the plates at 37 °C varied between the cell lines from 20 min for KB-3-1 and A549, L929 for 30 min, and 2 h for MCF-7 before measuring absorbance at 450 nm (reference 600 nm) with an Infinite 200 PRO plate reader (Tecan, Männedorf, Switzerland). As positive control compounds, Auranofin and Staurosporin were applied. The absorbance of the solvent control was set to 100%. The EC_{50} values were determined with Sigma Plot. All data are average values from four biological replicates.

Supporting Information

Supporting Information File 1

NMR, MS, UV and IR spectra of irinan A (**2**) and irinan B (**3**). NMR data of withanolide E (**4**), withanolide F (**5**) and perulactone H (**6**).
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-196-S1.pdf]

Acknowledgements

This work has been carried out within the framework of the SMART BIOTECS alliance between the Technische Universität Braunschweig and the Leibniz Universität Hannover. This initiative is supported by the Ministry of Science and Culture (MWK) of Lower Saxony, Germany. LC–MS (INST 187/621) and NMR (INST 187/686-1) instruments funded by the Deutsche Forschungsgemeinschaft (DFG) were used. We thank Prof. Russell Cox for his support and helpful discussions, and Marcel Arndt for preliminary work. We thank Dr. Jörg Fohrer and colleagues for support with NMR measurements and Katja Körner and colleagues for excellent lab support.

ORCID® iDs

Mark Brönstrup - https://orcid.org/0000-0002-8971-7045
Jakob Franke - https://orcid.org/0000-0002-7603-6232

References

1. Rai, M.; Jogee, P. S.; Agarkar, G.; Santos, C. A. d. *Pharm. Biol. (Abingdon, U. K.)* **2016,** *54,* 189–197. doi:10.3109/13880209.2015.1027778
2. Maurya, R. *J. Pharm. Pharmacol.* **2010,** *62,* 153–160. doi:10.1211/jpp.62.02.0001
3. Chen, L.-X.; He, H.; Qiu, F. *Nat. Prod. Rep.* **2011,** *28,* 705–740. doi:10.1039/c0np00045k
4. Fischer, G.; Herrera, A.; Almanza, P. J. Cape Gooseberry (Physalis Peruviana L.). In *Postharvest Biology and Technology of Tropical and Subtropical Fruits;* Yahia, E. M., Ed.; Woodhead Publishing Series in Food Science, Technology and Nutrition; Elsevier: Amsterdam, Netherlands, 2011; pp 374–397. doi:10.1533/9780857092762.374
5. Sang-ngern, M.; Youn, U. J.; Park, E.-J.; Kondratyuk, T. P.; Simmons, C. J.; Wall, M. M.; Ruf, M.; Lorch, S. E.; Leong, E.; Pezzuto, J. M.; Chang, L. C. *Bioorg. Med. Chem. Lett.* **2016,** *26,* 2755–2759. doi:10.1016/j.bmcl.2016.04.077
6. Park, E.-J.; Sang-Ngern, M.; Chang, L. C.; Pezzuto, J. M. *J. Nat. Prod.* **2019,** *82,* 492–499. doi:10.1021/acs.jnatprod.8b00861
7. Kirson, I.; Abraham, A.; Sethi, P. D.; Subramanian, S. S.; Glotter, E. *Phytochemistry* **1976,** *15,* 340–342. doi:10.1016/s0031-9422(00)89029-9
8. Frolow, F.; Ray, A. B.; Sahai, M.; Glotter, E.; Gottlieb, H. E.; Kirson, I. *J. Chem. Soc., Perkin Trans. 1* **1981,** 1029–1032. doi:10.1039/p19810001029
9. Sakurai, K.; Ishii, H.; Kobayashi, S.; Iwao, T. *Chem. Pharm. Bull.* **1976,** *24,* 1403–1405. doi:10.1248/cpb.24.1403
10. Neogi, P.; Sahai, M.; Ray, A. B. *Phytochemistry* **1986,** *26,* 243–247. doi:10.1016/s0031-9422(00)81520-4
11. Dinan, L. N.; Sarker, S. D.; Šik, V. *Phytochemistry* **1997,** *44,* 509–512. doi:10.1016/s0031-9422(96)00553-5
12. Fang, S.-T.; Liu, J.-K.; Li, B. *Steroids* **2012,** *77,* 36–44. doi:10.1016/j.steroids.2011.09.011
13. Ahmad, S.; Malik, A.; Yasmin, R.; Ullah, N.; Gul, W.; Khan, P. M.; Nawaz, H. R.; Afza, N. *Phytochemistry* **1999,** *50,* 647–651. doi:10.1016/s0031-9422(98)00567-6
14. Lan, Y.-H.; Chang, F.-R.; Pan, M.-J.; Wu, C.-C.; Wu, S.-J.; Chen, S.-L.; Wang, S.-S.; Wu, M.-J.; Wu, Y.-C. *Food Chem.* **2009,** *116,* 462–469. doi:10.1016/j.foodchem.2009.02.061
15. Fang, S.-T.; Li, B.; Liu, J.-K. *Helv. Chim. Acta* **2009,** *92,* 1304–1308. doi:10.1002/hlca.200900005
16. Zhang, W.-N.; Tong, W.-Y. *Chem. Biodiversity* **2016,** *13,* 48–65. doi:10.1002/cbdv.201400435
17. Xia, G.; Li, Y.; Sun, J.; Wang, L.; Tang, X.; Lin, B.; Kang, N.; Huang, J.; Chen, L.; Qiu, F. *Steroids* **2016,** *115,* 136–146. doi:10.1016/j.steroids.2016.09.002
18. Xu, Y.-m.; Bunting, D. P.; Liu, M. X.; Bandaranayake, H. A.; Gunatilaka, A. A. L. *J. Nat. Prod.* **2016,** *79,* 821–830. doi:10.1021/acs.jnatprod.5b00911
19. Chen, L.-X.; Xia, G.-Y.; He, H.; Huang, J.; Qiu, F.; Zi, X.-L. *RSC Adv.* **2016,** *6,* 52925–52936. doi:10.1039/c6ra07031k
20. Ozawa, M.; Morita, M.; Hirai, G.; Tamura, S.; Kawai, M.; Tsuchiya, A.; Oonuma, K.; Maruoka, K.; Sodeoka, M. *ACS Med. Chem. Lett.* **2013,** *4,* 730–735. doi:10.1021/ml400144e
21. Zhang, H.; Timmermann, B. N. *J. Nat. Prod.* **2016,** *79,* 732–742. doi:10.1021/acs.jnatprod.5b00648
22. Kicman, A. T. *Br. J. Pharmacol.* **2008,** *154,* 502–521. doi:10.1038/bjp.2008.165
23. Maldonado, E.; Alvarado, V. E.; Torres, F. R.; Martínez, M.; Pérez-Castorena, A. L. *Planta Med.* **2005,** *71,* 548–553. doi:10.1055/s-2005-864157
24. Siddiqui, B. S.; Usmani, S. B.; Begum, S.; Siddiqui, S. *Phytochemistry* **1993,** *33,* 925–928. doi:10.1016/0031-9422(93)85306-c
25. Sanogo, R.; Germano, M. P.; de Tommasi, N.; Pizza, C.; Aquino, R. *Phytochemistry* **1998,** *47,* 73–78. doi:10.1016/s0031-9422(97)00477-9
26. Pupo, M. T.; Vieira, P. C.; Fernandes, J. B.; das G.F. da Silva, M. F.; Fo, E. R. *Phytochemistry* **1997,** *45,* 1495–1500. doi:10.1016/s0031-9422(97)00167-2
27. Miller, W. L.; Auchus, R. J. *Endocr. Rev.* **2011,** *32,* 81–151. doi:10.1210/er.2010-0013
28. Fukushima, A.; Nakamura, M.; Suzuki, H.; Yamazaki, M.; Knoch, E.; Mori, T.; Umemoto, N.; Morita, M.; Hirai, G.; Sodeoka, M.; Saito, K. *Front. Plant Sci.* **2016,** *7,* 1883. doi:10.3389/fpls.2016.01883
29. Dhar, N.; Razdan, S.; Rana, S.; Bhat, W. W.; Vishwakarma, R.; Lattoo, S. K. *Front. Plant Sci.* **2015,** *6,* 1031. doi:10.3389/fpls.2015.01031
30. Strushkevich, N.; MacKenzie, F.; Cherkesova, T.; Grabovec, I.; Usanov, S.; Park, H.-W. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108,* 10139–10143. doi:10.1073/pnas.1019441108
31. Ortiz de Montellano, P. R. Substrate Oxidation by Cytochrome P450 Enzymes. In *Cytochrome P450: Structure, Mechanism, and Biochemistry;* Ortiz de Montellano, P. E., Ed.; Springer International Publishing: Cham, Switzerland, 2015; pp 111–176. doi:10.1007/978-3-319-12108-6_4
32. Capon, R. J. *Nat. Prod. Rep.* **2019**. doi:10.1039/c9np00013e
33. Zviely, M.; Goldman, A.; Kirson, I.; Glotter, E. *J. Chem. Soc., Perkin Trans. 1* **1986,** 229–231. doi:10.1039/p19860000229
34. García, N.; Rubio-Presa, R.; García-García, P.; Fernández-Rodríguez, M. A.; Pedrosa, M. R.; Arnáiz, F. J.; Sanz, R. *Green Chem.* **2016,** *18,* 2335–2340. doi:10.1039/c5gc02862k
35. Yen, C.-Y.; Chiu, C.-C.; Chang, F.-R.; Chen, J. Y.-F.; Hwang, C.-C.; Hseu, Y.-C.; Yang, H.-L.; Lee, A. Y.-L.; Tsai, M.-T.; Guo, Z.-L.; Cheng, Y.-S.; Liu, Y.-C.; Lan, Y.-H.; Chang, Y.-C.; Ko, Y.-C.; Chang, H.-W.; Wu, Y.-C. *BMC Cancer* **2010,** *10,* 46. doi:10.1186/1471-2407-10-46
36. Xu, Y.-M.; Wijeratne, E. M. K.; Babyak, A. L.; Marks, H. R.; Brooks, A. D.; Tewary, P.; Xuan, L.-J.; Wang, W.-Q.; Sayers, T. J.; Gunatilaka, A. A. L. *J. Nat. Prod.* **2017,** *80,* 1981–1991. doi:10.1021/acs.jnatprod.6b01129
37. Ishiyama, M.; Tominaga, H.; Shiga, M.; Sasamoto, K.; Ohkura, Y.; Ueno, K. *Biol. Pharm. Bull.* **1996,** *19,* 1518–1520. doi:10.1248/bpb.19.1518
38. Sasse, F.; Steinmetz, H.; Schupp, T.; Petersen, F.; Memmert, K.; Hofmann, H.; Heusser, C.; Brinkmann, V.; Matt, P. V.; Höfle, G.; Reichenbach, H. *J. Antibiot.* **2002,** *55,* 543–551. doi:10.7164/antibiotics.55.543

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Bipolenins K–N: New sesquiterpenoids from the fungal plant pathogen *Bipolaris sorokiniana*

Chin-Soon Phan[1], Hang Li[1], Simon Kessler[1], Peter S. Solomon[2], Andrew M. Piggott[*3] and Yit-Heng Chooi[*1]

Full Research Paper

Open Access

Address:
[1]School of Molecular Sciences, The University of Western Australia, Perth, WA 6009, Australia, [2]Research School of Biology, Australian National University, Canberra, ACT 2601, Australia and [3]Department of Molecular Sciences, Macquarie University, Sydney, NSW 2109, Australia

Email:
Andrew M. Piggott[*] - andrew.piggott@mq.edu.au; Yit-Heng Chooi[*] - yitheng.chooi@uwa.edu.au

* Corresponding author

Keywords:
Bipolaris sorokiniana; phytotoxicity; sesquiterpenes; terpenes

Beilstein J. Org. Chem. **2019,** *15,* 2020–2028.
doi:10.3762/bjoc.15.198

Received: 04 July 2019
Accepted: 16 August 2019
Published: 26 August 2019

This article is part of the thematic issue "Terpenes".

Guest Editor: J. S. Dickschat

Abstract

Chemical investigation of the barley and wheat fungal pathogen *Bipolaris sorokiniana* BRIP10943 yielded four new sativene-type sesquiterpenoid natural products, bipolenins K–N (**1**–**4**), together with seven related known analogues (**5**–**11**), and a sesterterpenoid (**12**). Their structures were determined by detailed analysis of spectroscopic data, supported by TDDFT calculations and comparison with previously reported analogues. These compounds were evaluated for their phytotoxic activity against wheat seedlings and wheat seed germination. The putative biosynthetic relationships between the isolated sesquiterpenoids were also explored.

Introduction

Fungi belonging to the genus *Bipolaris* (teleomorph: *Cochliobolus*) have been reported to produce a diverse array of secondary metabolites, including sesquiterpenes [1-7], sesquiterpene-xanthones [8], diterpenes [9], sesterterpenes [10], cochlioquinones and peptides [11]. Moreover, several of these secondary metabolites are known to play important roles in mediating the virulence of these fungi against plant hosts [12]. Well-known examples include the host-specific toxins victorin and T-toxin and other non-host-specific toxins such as the ophiobolins [11]. *Bipolaris sorokiniana* (syn. *Cochliobolus sativus*) has been identified as the causative agent of multiple diseases on wheat and barley and is a major threat to yield improvement and food security in Central Asia [13]. Recent genome sequencing of 35 Australian strains of *B. sorokiniana* identified a known proteinaceous necrotrophic effector, *ToxA*, which confers host-specific virulence proteins and is proposed to be acquired through horizontal gene transfer [14]. To date, only three studies have explored phytotoxins from *B. sorokiniana* [2,7,10]. Therefore, in the framework of furthering our understanding of the roles of *B. sorokiniana* sec-

ondary metabolites in crop disease, we investigated the compounds produced by the *ToxA*-containing strain BRIP10943 (CS10) [14] and their phytotoxicity. This led to the isolation of four new sativene-type sesquiterpenoid natural products along with seven related known analogues and one sesterterpenoid. Herein, the isolation, structure elucidation and phytotoxic activities of these compounds are presented.

Results and Discussion

B. sorokiniana was cultivated for 22 days in Fries medium supplemented with rolled oats. The resulting broth and mycelia were extracted with methanol and the extracts were partitioned against EtOAc/MeOH/acetic acid (89.9:10:0.1 ratio). The combined organic layer was chromatographed repeatedly with silica gel and RP-HPLC to afford four new sativene-type sesquiterpenoids, bipolenins K–N (**1**–**4**), along with eight previously reported compounds (**5**–**12**), which were identified as sativene-type sesquiterpenoids prehelminthosporol lactone (**5**) [1], helminthosporic acid (**6**) [1], helminthosporol (**7**) [15], bipolenin A (**8**) [3], secolongifolene diol (**9**) [15], dihydroprehelminthosporol (**10**) [1] and sorokinianin (**11**) [2], and the cytotoxic sesterterpenoid, terpestacin (**12**) [16,17] (Figure 1).

Bipolenin K (**1**) was isolated as a colourless oil. Its molecular formula was determined as $C_{15}H_{22}O_3$ from the HRESIMS $[M + H]^+$ ion at *m/z* 251.1646 (calcd for $C_{15}H_{23}O_3^+$, 251.1642), corresponding to five degrees of unsaturation. The IR absorption bands at 3445 and 1729 cm^{-1} revealed the presence of hydroxy and ester moieties, respectively. The ^{13}C NMR spectrum (Figure S2, Supporting Information File 1) showed 15 distinct carbon signals, while ^{13}C and ^1H NMR data (Table 1) indicated an isopropyl unit (δ_C 28.3, 29.2 and 73.1; δ_H 1.20 and 1.24), one tertiary methyl (δ_C 20.1; δ_H 1.25), a disubstituted olefin (δ_C 105.7 and 155.2; δ_H 5.12 and 4.89) and four methines (δ_C 39.6, 51.1, 51.6 and 54.0; δ_H 1.73, 1.88, 2.65 and 3.77), which are typical resonances for sativene-type sesquiterpenoids. The NMR data for **1** were very similar to those for prehelminthosporol lactone (**5**) except for the replacement of a methine group (δ_C 32.1; δ_H 1.42) at C-9 in **5** with a hydroxylated quaternary carbon (δ_C 73.1) in **1**. This suggested that **1** was the 9-hydroxy analogue of **5**, which was further confirmed by detailed analysis of key 2D NMR correlations (Figure 2). Compound **1** was previously reported in 1970 as a semi-synthetic analogue of 9-hydroxyprehelminthosporol [18], but has not been previously isolated and characterised from a natural source.

The relative configuration of **1** was established based on NOESY correlations (Figure 3) of H-1 to H_2-5 and H_3-10; and H_2-14 to H-7, H_3-8 and H-13. Due to the constrained bicyclo[3.2.1]octane ring system, these NOESY correlations indicated that H-1 was β-oriented, while H-6, H-7, H_3-8 and H-13

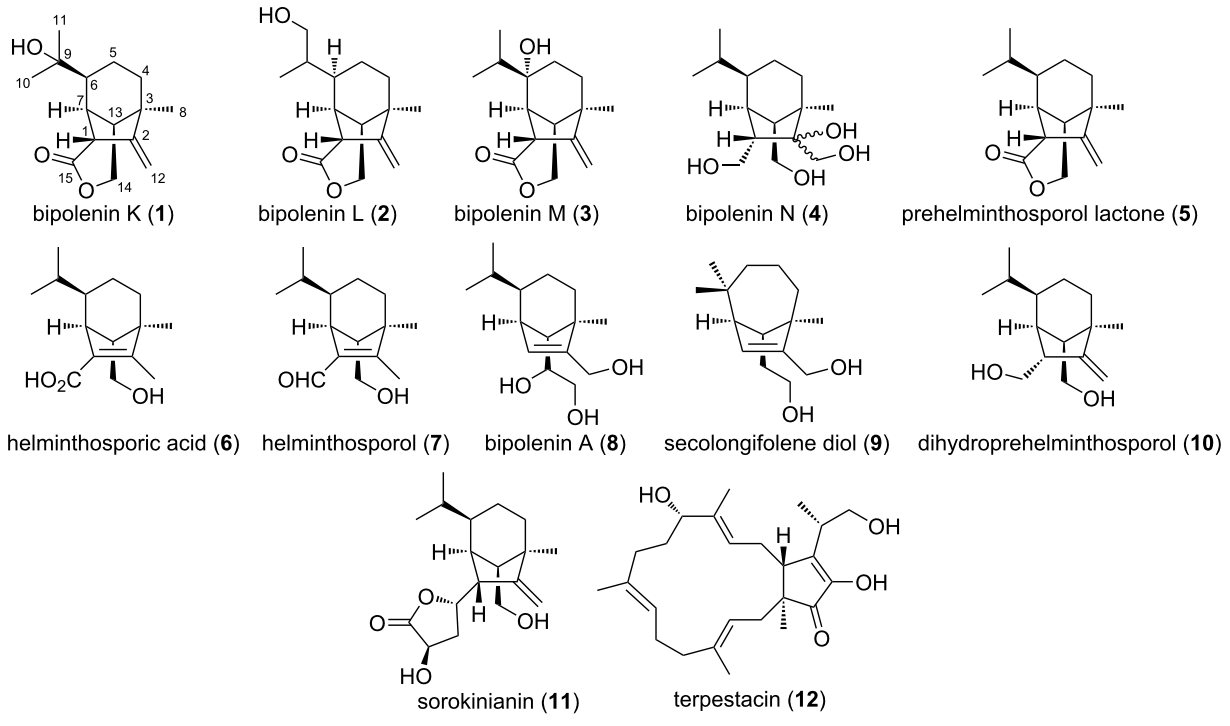


Figure 1: Structures of compounds **1**–**12** isolated from *B. sorokiniana*.

Table 1: ^{1}H and ^{13}C NMR data for bipolenins K–N (**1**–**4**).

No.	Bipolenin K (**1**)[a]		Bipolenin L (**2**)[b]		Bipolenin M (**3**)[b]		Bipolenin N (**4**)[c]	
	δ_C	δ_H, mult (*J* in Hz)	δ_C	δ_H, mult (*J* in Hz)	δ_C	δ_H, mult (*J* in Hz)	δ_C	δ_H, mult (*J* in Hz)
1	54.0	3.77, br s	52.4	3.36, br s	53.1	3.31, m	47.4	2.19, dt (10.4, 3.1)
2	155.2		155.1		154.3		88.1	
3	46.3		46.0		45.6		50.0	
4	42.4	1.64, m	42.4	1.64, dd (12.8, 4.7)	39.3	1.85, td (14.8, 5.2)	30.6	1.60, m
		1.53, m		1.48, ddd (12.8, 5.6, 2.0)		1.34, m		
5	22.3	1.72, m	26.3	1.80, m	32.5	1.69, m	24.6	1.68, m
		1.52, m		1.27, m		1.34, m		1.31, m
6	51.1	1.73, m	41.9	1.70, m	74.8		45.8	1.18, m
7	39.6	2.65, br s	40.7	2.47, br s	46.3	2.44, br s	43.5	1.53, br s
8	20.1	1.25, s	20.3	1.25, s	20.2	1.25, s	19.4	0.98, s
9	73.1		39.7	1.44, m	35.8	1.67, septet (6.9)	30.4	1.34, m
10	29.2	1.24, s	15.8	1.02, d (6.9)	16.5	0.91, d (6.9)	21.4	0.91, d (6.9)
11	28.3	1.20, s	65.8	3.58, dd (10.8, 3.8)	16.5	0.91, d (6.9)	20.3	0.84, d (6.9)
				3.44, dd (10.8, 6.0)				
12	105.7	5.12, s	106.3	5.15, s	107.4	5.19, s	62.2	3.79, d (11.8)
		4.89, s		4.90, s		4.96, s		3.65, d (11.8)
13	51.6	1.88, dd (4.5, 1.8)	51.0	1.92, d (4.5)	46.0	2.38, d (4.4)	52.2	1.50, d (3.5)
14	71.3	4.46, d (11.7)	71.6	4.48, d (11.7)	72.1	4.53, d (11.7)	69.8	3.90, dd (7.8, 3.5)
		4.30, dd (11.7, 4.5)		4.30, dd (11.7, 4.5)		4.23, dd (11.7, 4.4)		3.36, d (7.8)
15	174.1		174.4		174.0		62.9	3.77, t (10.4)
								3.47, dd (10.4, 3.1)

[a]Recorded at 500/125 MHz for ^{1}H/^{13}C in CD_3OD; [b]Recorded at 600/150 MHz for ^{1}H/^{13}C in CD_3OD; [c]Recorded at 600/150 MHz for ^{1}H/^{13}C in $CDCl_3$.

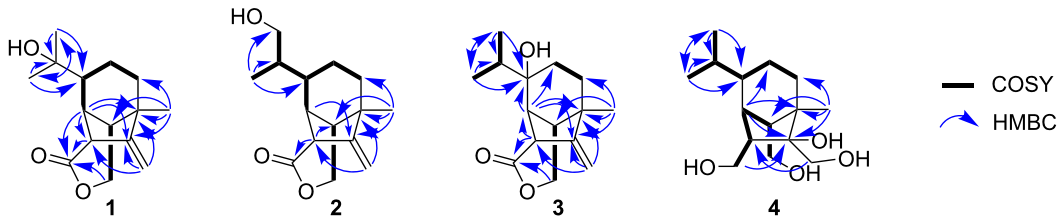


Figure 2: Key 2D NMR correlations of bipolenins K–N (**1**–**4**).

were α-oriented. The absolute configuration of **1** was determined to be 1*R*,3*R*,6*S*,7*R*,13*S* by comparison of the experimental electronic circular dichroism (ECD) spectrum with time-dependent density functional theory (TDDFT)-calculated ECD spectra of the two possible enantiomers of **1** (Figure 4).

Bipolenin L (**2**) was isolated as a colourless oil. The HRESIMS $[M + H]^+$ ion at *m*/*z* 251.1649 corresponded to a molecular formula $C_{15}H_{22}O_3$ (calcd for $C_{15}H_{23}O_3^+$, 251.1642), which is isomeric with **1**. The ^{1}H and ^{13}C NMR data for **2** (Table 1) were also very similar to those for **5**, with the only significant difference being the presence of a hydroxymethylene group (δ_C 65.8; δ_H 3.58 and 3.44) in place of the methyl group at C-11. Thus, the structure of **2** was assigned as the 11-hydroxy analogue of **5**. The absolute configurations of the chiral centres in **2** were established to be the same as for **1** after investigation of the proton coupling constants (Table 1), NOESY correlations (Figure 3) and ECD spectra (Figure 4). The configuration at C-9

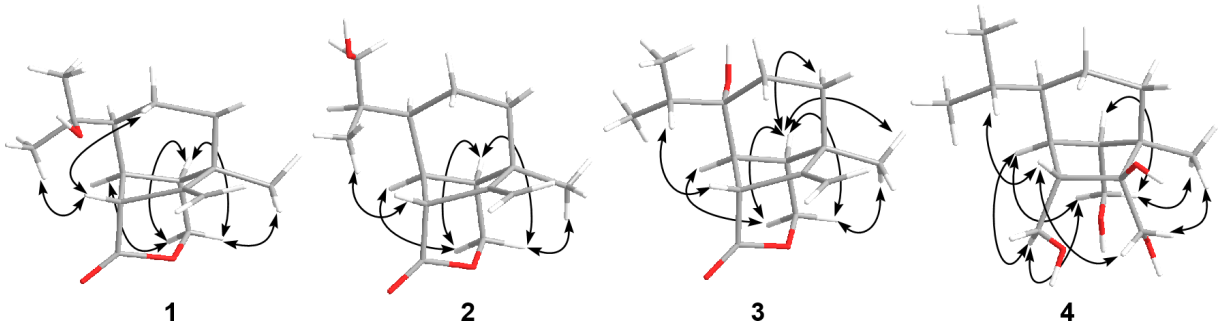


Figure 3: Key NOESY correlations of bipolenins K–N (**1**–**4**).

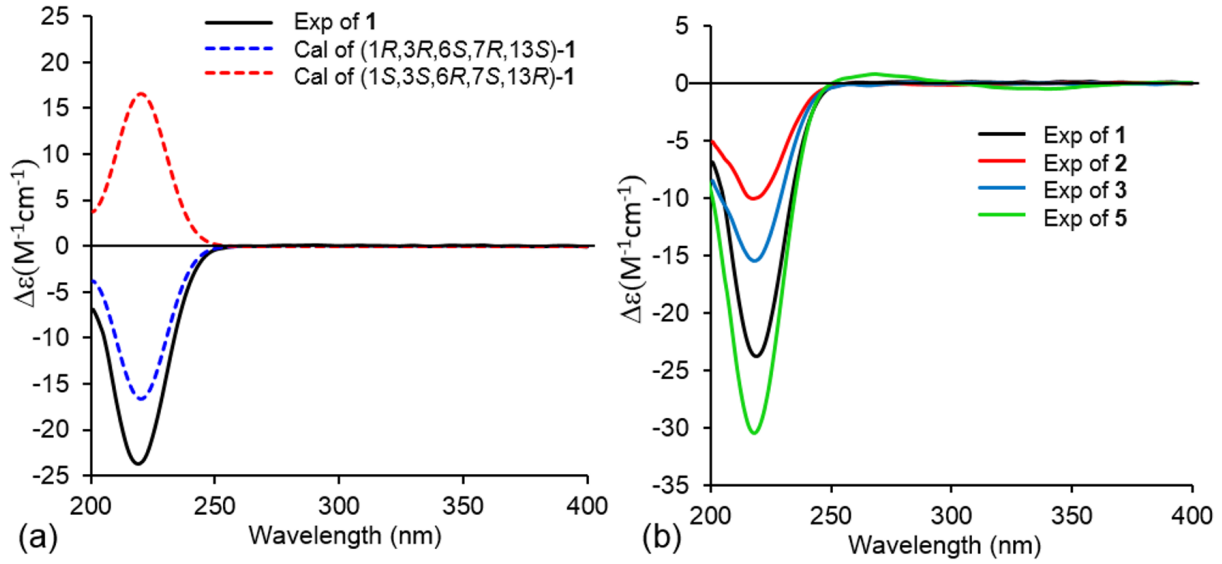


Figure 4: (a) Experimental ECD spectrum of **1** (MeOH) compared to TDDFT-calculated spectra (B3LYP-D3/def2-TZVPP) for the two possible enantiomers of **1**, which were blue-shifted by 9 nm. (b) Comparison of experimental ECD spectra of **1**–**3** and **5** (MeOH).

was not determined. Hence, structure **2** was determined as shown in Figure 1.

Bipolenin M (**3**) was purified as a colourless oil. The molecular formula $C_{15}H_{22}O_3$ was based on a HRESIMS $[M + H]^+$ ion at *m/z* 251.1647 (calcd for $C_{15}H_{23}O_3^+$, 251.1642), and is isomeric with **1** and **2**. The ^1H and ^{13}C NMR data for **3** (Table 1) were very similar to those for **5**, with the only significant difference being the replacement of the methine group (δ_C 47.1; δ_H 1.40) at C-6 in **5** with a hydroxylated quaternary carbon (δ_C 74.8) in **3**. This suggested that **3** was the 6-hydroxy analogue of **5**, which was further confirmed by detailed analysis of key 2D NMR correlations (Figure 2). The absolute configurations of the stereocentres in **3** were established to be identical to **1** and **2** based on the analysis of proton coupling constants (Table 1), NOESY correlations (Figure 3) and ECD spectra (Figure 4).

Bipolenin N (**4**) was acquired as a colourless oil. Its molecular formula was determined to be $C_{15}H_{28}O_4$ from the HRESIMS $[M + H - 2H_2O]^+$ ion at *m/z* 237.1859 (calcd for $C_{15}H_{25}O_2^+$, 237.1849). The UV–vis spectra of **1**–**3** (Figure S37, Supporting Information File 1) and previously reported congener **5** were almost identical, while **4** displayed no significant UV–vis absorptions, suggesting the absence of both the ester and alkene moieties. This was confirmed by the analysis of the ^1H and ^{13}C NMR data for **4** (Table 1), which revealed the absence of ester and alkene resonances and the presence of three hydroxylated methylenes at C-12 (δ_C 62.2; δ_H 3.79 and 3.65), C-14 (δ_C 69.8; δ_H 3.90 and 3.36) and C-15 (δ_C 62.9; δ_H 3.77 and

3.47), and one hydroxylated quaternary carbon at C-2 (δ_C 88.1). This suggested **4** was related to **5**, but with reduction of the lactone ring to the dialcohol and dihydroxylation of the $\Delta^{2,12}$ double bond. Detailed analysis of the 2D NMR data for **4** (Figure 2) confirmed the seco-sativene-type scaffold. The relative configurations at C-1, C-3, C-6, C-7 and C-13 were determined to be the same as those of **1**–**3** and other reported analogues based on NOESY correlations (Figure 3), while the configuration at C-2 was not determined. The ECD spectrum of **4** (Figure S38, Supporting Information File 1) was measured, but no significant Cotton effect was observed. Therefore, the structure of **4** was determined as shown in Figure 1.

Equipped with the compounds, we tested **1**–**12** for phytotoxic activity against wheat seedlings. The compounds all showed negligible activities at 200 ppm, although **6** and **10** showed signs of necrosis at 500 ppm (Figure S40, Supporting Information File 1). In addition, the activities of **1**, **6**–**10** and **12** against wheat seed germination were also tested, with **7** inhibiting germination at 100 ppm (Figure S41, Supporting Information File 1). This corresponds to a previous report of the inhibitory effects of **7** on lettuce seed germination [19]. This activity could be due to the presence of an aldehyde moiety in **7**. Interestingly, an earlier study showed that **7** promoted the elongation of the shoots of rice seedlings [20]. Compound **12** was reported to have a broad spectrum of biological activities, including phytotoxicity on juvenile plant *Bromus tectorum* [21], syncytium formation inhibitory effects on cells infected with respiratory syncytial virus [22,23], induction of aerial mycelium formation in *Fusarium culmorum* [24], and as an inhibitor of ubiquinol-cytochrome c reductase binding protein, blocking mitochondrial ROS-mediated vascular endothelial growth factor receptor type 2 signalling pathways in endothelial cells [25]. However, **12** showed no activity against wheat seedlings or wheat seed germination in this study.

The sativene-type sesquiterpenoids contain a bicyclo[3.2.1]octane backbone and are related to seco-sativene and isosativene scaffolds [15] (Figure 5a). They were also proposed to be related to the bicyclo[4.2.1]nonane-containing longifolene and seco-longifolene sesquiterpenoids, as they were often co-isolated [3,4,6,15,26,27]. A closer examination of the biosynthetic relationship between sativene and longifolene scaffolds suggests that the two pathways branched early at the nerolidyl cation (Figure 5b) [28-31]. The biosynthesis of **1**–**8** and **10**–**11** are likely to be derived from sativene with a key oxidation at C-15 followed by a Baeyer–Villiger oxidation to break the C-14–C-15 bond (Figure 5c). Based on an isotope labelling study, the γ-butyrolactone moiety on **11** has been proposed to be derived from oxaloacetic acid or similar TCA-cycle intermediates [32]. Compound **9**, which contains the seco-longifolene scaffold, is likely to be derived from longifolene via a similar Baeyer–Villiger mechanism proposed above for **1**–**8** and **10** and **11**.

Several sativene-type sesquiterpenoids have been previously reported from fungi, including from *Bipolaris* sp. [1], *B. sorokiniana* [2], *B. eleusines* [3-6,8], *Cochliobolus* sp. [26], *Cochliobolus sativus* [18], *Helminthosporium sativum* [20,33,34], *Drechslera* sp. [27], *Drechslera dematioidea* [15], and *Veronaea* sp. [35], most of which are Dothideomycetes. Significantly, this is the first report pertaining to sativene-type sesquiterpenoids from *B. sorokiniana* in 25 years, since the first and only literature account was published in 1994 [2]. Furthermore, structure **4** has a seco-sativene type scaffold without an olefin unit at C-1/C-2 or C-2/C-12. In contrast, all the previously known seco-sativene-type sesquiterpenoids possessed a double bond either at C-1/C-2 or C-2/C-12 [1,2,15,18,20,33], except drechslerine C, which contains a decarboxylated seco-sativene-type scaffold [15]. To the best of our knowledge, the previously reported sativene-type sesquiterpenoids **5** (isolated from *B. eleusines* and *Cochliobolus* sp., and from semi-synthetic analogue of prehelminthosporol with pyridinium chlorochromate or chromic acid) [1,4,18,26], **8** (isolated from *B. eleusines*) [3,6], and **10** (isolated from *Bipolaris* sp. and *Cochliobolus* sp.) [1,26], were reported in *B. sorokiniana* for the first time, while, known metabolites **6** and **7** [20], **9** [34], **11** [2], and **12** [16] were previously reported from *B. sorokiniana* (syn. *C. sativus* and *H. sativum*).

The terpene synthase responsible for the biosynthesis of the sativene/longifolene backbone of **1**–**11** remains unknown. Given that the genome of *B. sorokiniana* BRIP10943 has been sequenced [14], we surveyed the genome for potential terpene synthases that may be responsible for the biosynthesis of these compounds. Four putative sesquiterpene synthases were found, corresponding to the genes COCSADRAFT_31812, COCSADRAFT_346586, COCSADRAFT_83129 and COCSADRAFT_26102 annotated in the published genome *B. sorokiniana* ND90Pr in GenBank. However, it is difficult to determine which sesquiterpene synthase is responsible for biosynthesis of the sativene-type sesquiterpene backbone at this stage.

The biosynthetic gene cluster (*tpc*) for terpestacin (**12**) has been recently identified from *Bipolaris maydis* [36]. A didomain sesterterpene synthase (*tpcA*) with a terpene cyclase domain and polyprenyltransferase domain was demonstrated to be responsible for the production of the sesterterpene backbone of **12**. A BLASTp search using *tpcA* as query against the genome of *B. sorokiniana* ND90Pr and BRIP10943 identified COCSADRAFT_342920 in ND90Pr (and its homolog in

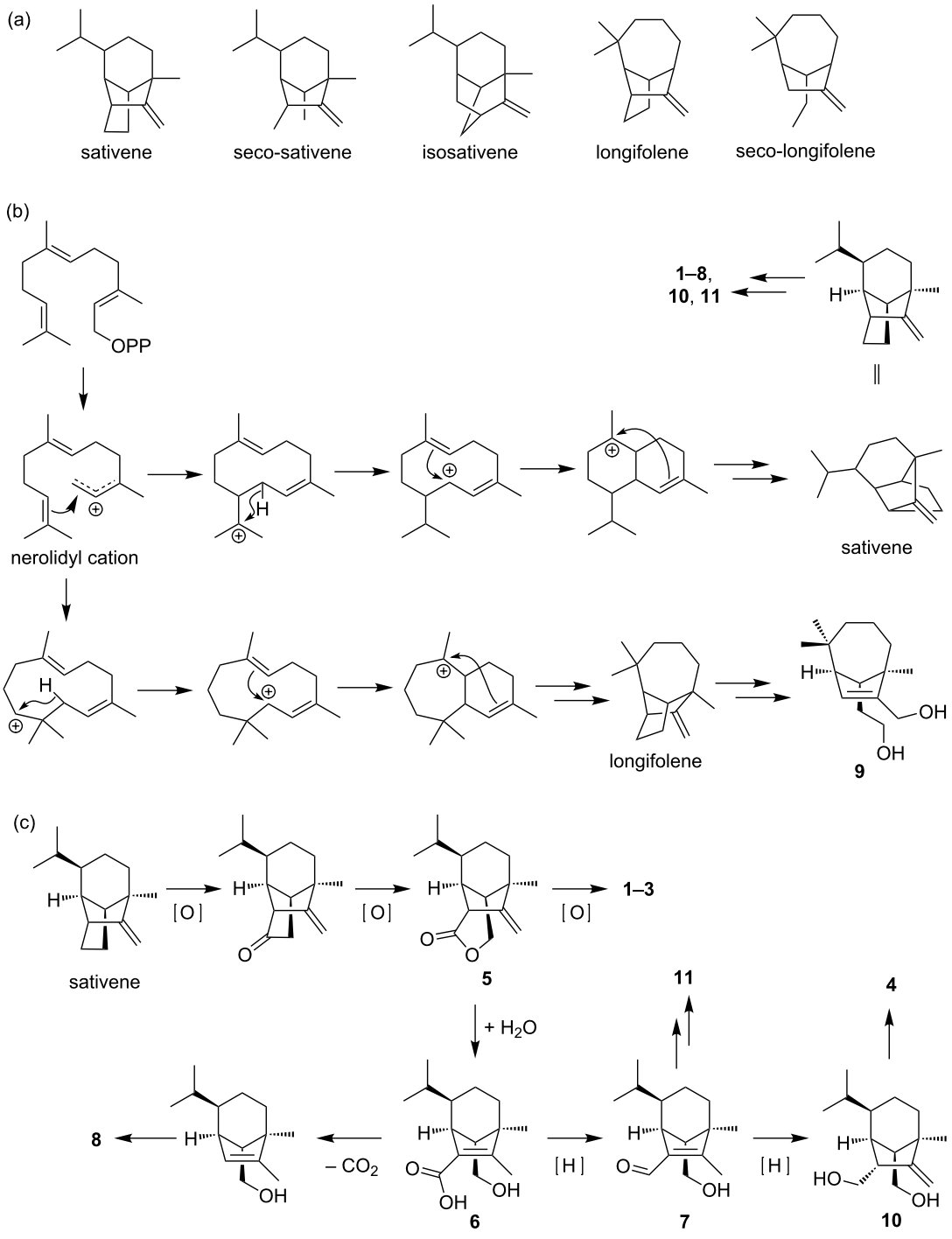


Figure 5: Relationship of sesquiterpenoids isolated in this study. A) Different groups of sativene/longifolene-type sesquiterpenoid scaffolds; B) The branched pathways to sativene- and longifolene-type sesquiterpenoids. C) Detailed proposed pathways to the sativene-derived sesquiterpenoids from this study.

BRIP10943), which shares 96% identity to *tpcA*. In the vicinity of the sesterterpene synthase gene, we also identified homologs for the two P450 oxygenases (*tpcB* and *tpcC*) encoded in the *tpc* cluster as COCSADRAFT_342924 (92% identity) and COCSADRAFT_146541 (98% identity), respectively, while the *tpcD* homolog in *B. sorokiniana* corresponded to COCSADRAFT_94398 (91% identity). This suggests that the homologous gene cluster in *B. sorokiniana* is likely responsible

for production of **12**. We are currently investigating the genetic basis for the biosynthesis of the sativene-type terpenoid compounds identified from *B. sorokiniana*.

Conclusion

Following the first and only reported isolation of a sativene-type sesquiterpenoid, sorokinianin (**11**), from *B. sorokiniana* in 1994 [2], we have expanded the number of reported analogues to eleven. These include the new sesquiterpenoid natural products, bipolenins K–N (**1**–**4**), as well as the previously reported sesquiterpenoids prehelminthosporol lactone (**5**), helminthosporic acid (**6**), helminthosporol (**7**), bipolenin A (**8**), secolongifolene diol (**9**), dihydroprehelminthosporol (**10**) and sorokinianin (**11**), together with a sesterterpenoid, terpestacin (**12**). We demonstrated that **6** and **10** have weak necrotic activity against wheat leaves, while **7** inhibited wheat seed germination at 100 ppm. These compounds served as markers for identifying putative sesquiterpene synthase genes in the genome of *B. sorokiniana* BRIP10943, allowing the molecular genetic basis for their biosynthesis and their roles in mediating the virulence of *B. sorokiniana* against wheat to be explored.

Experimental

General experimental procedures

Optical rotations were measured on an A. Krüss Optronic P8000 polarimeter. The IR spectra were collected on a Perkin Elmer Spectrum One FTIR spectrometer. The HR-ESIMS spectra were recorded on a Waters LCT Premier XE mass spectrometer. The ESIMS spectra were recorded on an Agilent 1260 LC system equipped with a DAD detector and coupled to an Agilent 6130 Quadrupole MS with an ESI source. The NMR spectra were recorded on Bruker Avance III HD 500 or AV600 spectrometers. The ECD spectra were recorded on a Jasco J-810 spectropolarimeter with MeOH as solvent. Flash cartridge (Reveleris, HP-silica, 12 g, 20 µm), Kinetex C18 (Phenomenex, 2.6 µm, 2.1 × 100 mm), and semi-preparative C18 (Grace, 5 µm, 10 × 250 mm) were used. All solvents used for extraction were analytical grade, and solvents for HPLC were HPLC grade.

Biological material

The fungal strain *B. sorokiniana* BRIP10943 was obtained from Queensland Plant Pathology Herbarium (BRIP). It was isolated from a wheat field at Hermitage, QLD, Australia. The fungus was maintained on potato dextrose agar (PDA).

Extraction and isolation

B. sorokiniana was cultured on 16 plates of V8PDA at 25 °C for 14 days, then inoculated in 4 L shake-flask culture (25 °C, 180 rpm for 22 days) in Fries medium supplemented with oat. The Fries medium was filtered and extracted by partition with EtOAc/MeOH/acetic acid at 89.9:10:0.1 ratio. The cells were extracted with MeOH and partition with EtOAc/MeOH/acetic acid at 89.7:10:0.3 ratio. Both organic partitioned layers were combined to obtain a light-yellow crude extract (205 mg), which was fractionated on a Reveleris flash chromatography (Grace) using gradient mode of H_2O/MeOH equipped with the flash cartridge, UV and evaporative light scattering detector. The resulting fractions were further purified by RP-HPLC on gradient mode of H_2O/MeCN equipped with the C_{18} column, and DAD detector to yield **1** (1.5 mg), **2** (0.4 mg), **3** (0.4 mg), **4** (0.4 mg), **5** (0.5 mg), **6** (1.0 mg), **7** (4.0 mg), **8** (1.8 mg), **9** (2.8 mg), **10** (3.0 mg), **11** (0.4 mg), and **12** (1.3 mg). The Kinetex C18 on RP-HPLC (Phenomenex, 2.6 µm, 2.1 × 100 mm, 0.75 mL/min, DAD detection 200–800 nm, gradient: 0–10 min 5–95% MeCN with 0.1% formic acid, 10–15 min 95% MeCN with 0.1% formic acid) eluted **1** (t_R 5.10 min), **2** (t_R 4.79 min), **3** (t_R 5.28 min), **4** (t_R 5.96 min), **5** (t_R 7.64 min), **6** (t_R 6.00 min), **7** (t_R 6.45 min), **8** (t_R 4.85 min), **9** (t_R 5.69 min), **10** (t_R 6.37 min), **11** (t_R 6.48 min) and **12** (t_R 6.88 min).

Bipolenin K (**1**): Colourless oil; $[\alpha]_D^{20}$ −59 (*c* 0.15, MeOH); IR (KBr) λ_{max} 3445, 2926 and 1729 cm^{-1}; ^1H and ^{13}C NMR data, see Table 1; HRESIMS *m*/*z*: $[M + H]^+$ calcd for $C_{15}H_{23}O_3^+$, 251.1642; found, 251.1646, and $[M + H - H_2O]^+$ calcd for $C_{15}H_{21}O_2^+$, 233.1536; found, 233.1531.

Bipolenin L (**2**): Colourless oil; $[\alpha]_D^{20}$ −63 (*c* 0.04, MeOH); IR (KBr) λ_{max} 3419, 2920 and 1730 cm^{-1}; ^1H and ^{13}C NMR data, see Table 1; HRESIMS *m*/*z*: $[M + H]^+$ calcd for $C_{15}H_{23}O_3^+$, 251.1642; found, 251.1649, and $[M + H - H_2O]^+$ calcd for $C_{15}H_{21}O_2^+$, 233.1536; found, 233.1535.

Bipolenin M (**3**): Colourless oil; $[\alpha]_D^{20}$ −57 (*c* 0.04, MeOH); IR (KBr) λ_{max} 3418, 2927 and 1720 cm^{-1}; ^1H and ^{13}C NMR data, see Table 1; HRESIMS *m*/*z*: $[M + H]^+$ calcd for $C_{15}H_{23}O_3^+$, 251.1642; found, 251.1647, and $[M + H - H_2O]^+$ calcd for $C_{15}H_{21}O_2^+$, 233.1536; found, 233.1545.

Bipolenin N (**4**): Colourless oil; $[\alpha]_D^{20}$ +38 (*c* 0.04, MeOH); IR (KBr) λ_{max} 3334, 2927 and 1045 cm^{-1}; ^1H and ^{13}C NMR data, see Table 1; HRESIMS *m*/*z*: $[M + H - 2H_2O]^+$ calcd for $C_{15}H_{25}O_2^+$, 237.1849; found, 237.1859.

Phytotoxicity assays

The phytotoxicity assays on wheat leaves and seeds were carried out as previously reported [37]. Briefly, the leaves of 17-days-old wheat seedlings in 10 cm planting pots were grown at 20 °C under a 16 h/8 h light/dark cycle regime. Compounds were dissolved in 0.2% MeOH/H_2O and 30 µL of dissolved solution was infiltrated on the adaxial face of leaves at concen-

trations of 100, 200 and 500 ppm (serial dilution) using a 1 mL syringe. The leaves were examined for the presence of necrosis or chlorosis after 24 h and 48 h. The control consisted of 30 µL of 0.2% MeOH/H_2O without dissolved compound. Two wheat seeds (sterilised by 10% EtOH) were placed on top of the agar (1.5% agar in 1 mL of tap water) containing 100 ppm of compound. The control was agar containing 30 µL of MeOH. The seeds were monitored for the progress of germination on day 5 and day 7.

Calculation of ECD spectra

Structures were initially subjected to a LowModeMD conformational search using the Molecular Operating Environment 2019.0101 package. The lowest energy geometry for each molecule was further optimised by DFT at the B3LYP-D3/def2-TZVPP level of theory using Turbomole 7.1 [38] and ECD spectra were calculated in Turbomole using TDDFT (B3LYP-D3/def2-TZVPP).

Supporting Information

Supporting Information File 1

NMR, IR and MS spectra of compounds **1**–**4**.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-198-S1.pdf]

Acknowledgements

This study was funded, in part, by the Australian Research Council (FT130100142 and FT160100233) and the Cooperative Research Centres Projects scheme (CRCPFIVE000119). The research was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI), which is supported by the Australian Government. Authors acknowledge the facilities of Microscopy Australia at the Centre for Microscopy, Characterisation & Analysis, The University of Western Australia, a facility funded by the University, State and Commonwealth Governments.

ORCID® iDs

Peter S. Solomon - https://orcid.org/0000-0002-5130-7307
Andrew M. Piggott - https://orcid.org/0000-0002-5308-5314
Yit-Heng Chooi - https://orcid.org/0000-0001-7719-7524

References

1. Pena-Rodriguez, L. M.; Armingeon, N. A.; Chilton, W. S. *J. Nat. Prod.* **1988,** *51,* 821–828. doi:10.1021/np50059a001
2. Nakajima, H.; Isomi, K.; Hamasaki, T.; Ichinoe, M. *Tetrahedron Lett.* **1994,** *35,* 9597–9600. doi:10.1016/0040-4039(94)88520-6
3. Ai, H.-L.; Yang, M.-S.; Zi, S.-H.; Guo, H.-C. *J. Asian Nat. Prod. Res.* **2015,** *17,* 982–987. doi:10.1080/10286020.2015.1041929
4. Yang, M.-S.; Cai, X.-Y.; He, Y.-Y.; Lu, M.-Y.; Liu, S.; Wang, W.-X.; Li, Z.-H.; Ai, H.-L.; Feng, T. *Nat. Prod. Bioprospect.* **2017,** *7,* 147–150. doi:10.1007/s13659-016-0116-4
5. Li, Z.-H.; Ai, H.-L.; Yang, M.-S.; He, J.; Feng, T. *Phytochem. Lett.* **2018,** *27,* 87–89. doi:10.1016/j.phytol.2018.07.007
6. He, J.; Li, Z.-H.; Ai, H.-L.; Feng, T.; Liu, J.-K. *Nat. Prod. Res.* **2018**. doi:10.1080/14786419.2018.1486313
7. Jahani, M.; Aggarwal, R.; Gupta, S.; Sharma, S.; Dureja, P. *Cereal Res. Commun.* **2014,** *42,* 252–261. doi:10.1556/crc.2013.0053
8. He, J.; Yang, M.-S.; Wang, W.-X.; Li, Z.-H.; Elkhateeb, W. A. M.; Wen, T.-C.; Ai, H.-L.; Feng, T. *RSC Adv.* **2019,** *9,* 128–131. doi:10.1039/c8ra09861a
9. Wang, Q.-x.; Qi, Q.-y.; Wang, K.; Li, L.; Bao, L.; Han, J.-j.; Liu, M.-m.; Zhang, L.-x.; Cai, L.; Liu, H.-w. *Org. Lett.* **2013,** *15,* 3982–3985. doi:10.1021/ol401736z
10. Nihashi, Y.; Lim, C. H.; Tanaka, C.; Miyagawa, H.; Ueno, T. *Biosci., Biotechnol., Biochem.* **2002,** *66,* 685–688. doi:10.1271/bbb.66.685
11. Muria-Gonzalez, M. J.; Chooi, Y.-H.; Breen, S.; Solomon, P. S. *Mol. Plant Pathol.* **2015,** *16,* 92–107. doi:10.1111/mpp.12162
12. Chooi, Y.-H.; Solomon, P. S. *Front. Microbiol.* **2014,** *5,* No. 640. doi:10.3389/fmicb.2014.00640
13. Kumar, J.; Schäfer, P.; Hückelhoven, R.; Langen, G.; Baltruschat, H.; Stein, E.; Nagarajan, S.; Kogel, K.-H. *Mol. Plant Pathol.* **2002,** *3,* 185–195. doi:10.1046/j.1364-3703.2002.00120.x
14. McDonald, M. C.; Ahren, D.; Simpfendorfer, S.; Milgate, A.; Solomon, P. S. *Mol. Plant Pathol.* **2018,** *19,* 432–439. doi:10.1111/mpp.12535
15. Osterhage, C.; König, G. M.; Höller, U.; Wright, A. D. *J. Nat. Prod.* **2002,** *65,* 306–313. doi:10.1021/np010092l
16. Lim, C. H.; Miyagawa, H.; Ueno, T.; Takenaka, H.; Sung, N. D. *Agric. Chem. Biotechnol. (Engl. Ed.)* **1996,** *39,* 241–244.
17. Chan, J.; Jamison, T. F. *J. Am. Chem. Soc.* **2004,** *126,* 10682–10691. doi:10.1021/ja0470968
18. Aldridge, D. C.; Turner, W. B. *J. Chem. Soc. C* **1970,** 686–688. doi:10.1039/j39700000686
19. Qader, M. M.; Kumar, N. S.; Jayasinghe, L.; Araya, H.; Fujimoto, Y. *Mycology* **2017,** *8,* 17–20. doi:10.1080/21501203.2016.1269844
20. Tamura, S.; Sakurai, A.; Kainuma, K.; Takai, M. *Agric. Biol. Chem.* **1965,** *29,* 216–221. doi:10.1080/00021369.1965.10858370
21. Masi, M.; Meyer, S.; Górecki, M.; Pescitelli, G.; Clement, S.; Cimmino, A.; Evidente, A. *Molecules* **2018,** *23,* 1734. doi:10.3390/molecules23071734
22. Oka, M.; Iimura, S.; Narita, Y.; Furumai, T.; Konishi, M.; Oki, T.; Gao, Q.; Kakisawa, H. *J. Org. Chem.* **1993,** *58,* 1875–1881. doi:10.1021/jo00059a045
23. Oka, M.; Iimura, S.; Tenmyo, O.; Sawada, Y.; Sugawara, M.; Okhusa, N.; Yamamoto, H.; Kawano, K.; Hu, S. L.; Fukagawa, Y.; Oki, T. *J. Antibiot.* **1993,** *46,* 367–373. doi:10.7164/antibiotics.46.367
24. Schlegel, B.; Schmidtke, M.; Dörfelt, H.; Kleinwächter, P.; Gräfe, U. *J. Basic Microbiol.* **2001,** *41,* 179–183. doi:10.1002/1521-4028(200107)41:3/4<179::aid-jobm179>3.0.co;2-h
25. Jung, H. J.; Kim, Y.; Chang, J.; Kang, S. W.; Kim, J. H.; Kwon, H. J. *J. Mol. Med. (Heidelberg, Ger.)* **2013,** *91,* 1117–1128. doi:10.1007/s00109-013-1049-6
26. Zhang, G.-F.; Guo, Z.-K.; Wang, W.; Cui, J.-T.; Tan, R.-X.; Ge, H.-M. *J. Asian Nat. Prod. Res.* **2011,** *13,* 761–764. doi:10.1080/10286020.2011.585608

27. Abdel-Lateff, A.; Okino, T.; Alarif, W. M.; Al-Lihaibi, S. S. *J. Saudi Chem. Soc.* **2013,** *17,* 161–165. doi:10.1016/j.jscs.2011.03.002
28. Lodewyk, M. W.; Gutta, P.; Tantillo, D. J. *J. Org. Chem.* **2008,** *73,* 6570–6579. doi:10.1021/jo800868r
29. Little, D. B.; Croteau, R. B. *Arch. Biochem. Biophys.* **2002,** *402,* 120–135. doi:10.1016/s0003-9861(02)00068-1
30. Steele, C. L.; Crock, J.; Bohlmann, J.; Croteau, R. *J. Biol. Chem.* **1998,** *273,* 2078–2089. doi:10.1074/jbc.273.4.2078
31. Van Vranken, D.; Weiss, G. A. *Introduction to Bioorganic Chemistry and Chemical Biology,* 1st ed.; Garland Science: New York, U.S.A., 2013.
32. Nakajima, H.; Toratsu, Y.; Fujii, Y.; Ichinoe, M.; Hamasaki, T. *Tetrahedron Lett.* **1998,** *39,* 1013–1016. doi:10.1016/s0040-4039(97)10803-6
33. de Mayo, P.; Williams, R. E. *J. Am. Chem. Soc.* **1965,** *87,* 3275. doi:10.1021/ja01092a066
34. Dorn, F.; Arigoni, D. *Experientia* **1974,** *30,* 851–852. doi:10.1007/bf01938319
35. Zhou, L.; Zheng, X.; Wan, C.-P.; Yu, Z.-F.; Zhang, K.-Q.; Li, G.-H. *Chem. Nat. Compd.* **2015,** *51,* 270–272. doi:10.1007/s10600-015-1259-y
36. Narita, K.; Minami, A.; Ozaki, T.; Liu, C.; Kodama, M.; Oikawa, H. *J. Org. Chem.* **2018,** *83,* 7042–7048. doi:10.1021/acs.joc.7b03220
37. Li, H.; Hu, J.; Wei, H.; Solomon, P. S.; Vuong, D.; Lacey, E.; Stubbs, K. A.; Piggott, A. M.; Chooi, Y.-H. *Org. Lett.* **2018,** *20,* 6148–6152. doi:10.1021/acs.orglett.8b02617
38. TURBOMOLE V7.1; a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Genome mining in *Trichoderma viride* J1-030: discovery and identification of novel sesquiterpene synthase and its products

Xiang Sun, You-Sheng Cai, Yujie Yuan, Guangkai Bian, Ziling Ye, Zixin Deng and Tiangang Liu*

Full Research Paper

Open Access

Address:
Key Laboratory of Combinatorial Biosynthesis and Drug Discovery, Ministry of Education and School of Pharmaceutical Sciences, Wuhan University, Wuhan, 430071, P. R. China

Email:
Tiangang Liu* - liutg@whu.edu.cn

* Corresponding author

Keywords:
genome mining; metabolic engineering; natural products; sesquiterpene synthase; terpenes; *Trichoderma viride* J1-030

Beilstein J. Org. Chem. **2019,** *15,* 2052–2058.
doi:10.3762/bjoc.15.202

Received: 25 May 2019
Accepted: 12 August 2019
Published: 28 August 2019

This article is part of the thematic issue "Terpenes".

Associate Editor: A. Kirschning

Abstract

Sesquiterpene synthases in *Trichoderma viride* have been seldom studied, despite the efficiency of filamentous fungi for terpenoid production. Using the farnesyl diphosphate-overexpressing *Saccharomyces cerevisiae* platform to produce diverse terpenoids, we herein identified an unknown sesquiterpene synthase from *T. viride* by genome mining and determined the structure of its corresponding products. One new 5/6 bicyclic sesquiterpene and its esterified derivative were characterised by GC–MS and 1D and 2D NMR spectroscopy. To the best of our knowledge, this is the first well-identified sesquiterpene synthase from *T. viride* to date.

Introduction

Terpenoids represent the most diverse group of natural products, with a wide distribution in microorganisms, plants, insects and various marine invertebrates [1,2]. More than 80,000 terpenoids have been identified and characterised [3-5]. These diverse and complex natural products are mostly derived from carbocation cyclisation with linear C5 isoprene precursors, which are catalysed by terpene synthases (TPSs) [6]. TPSs can be classified into three types based on their amino acid sequence. Type I TPSs are metal-dependent enzymes that initiate cyclisation by the elimination of diphosphate groups from precursors and carbocation formation, and type II TPSs initiate the catalytic process by the protonation of an olefinic double bond [7]. The recently reported type III TPSs, UbiA-related TPSs, also catalyse cascade reactions by diphosphate elimination [8]. In addition, each type of TPS is characterised by a unique aspartate-rich motif; most type I TPSs have a DDXXD/E motif and an NSE/DTE motif, whereas type II TPSs have the DXDD motif [9,10].

The C_{15} sesquiterpenoids constitute a large class of terpenoids with a wide range of industrial and commercial applications, including uses in flavours and perfumes, as bioactive molecules

in the pharmaceutical industry, and in health care products [11]. Sesquiterpenoids are biosynthesised from the universal linear precursor farnesyl pyrophosphate (FPP) and assembled by FPP synthases, using dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP) as substrates. The subsequent elimination of diphosphate from FPP is catalysed by sesquiterpene synthases, with further cyclisation steps to form structurally diverse (poly)cyclic core skeletons [3,12]. A set of post-modification enzymes can transform core sesquiterpene skeletons into different kinds of sesquiterpenoids with potential anticancer, cytotoxic and antibiotic functions [13]. More than 121 skeleton structures derived from the sesquiterpene precursor FPP via sesquiterpene synthase have been described. Nearly 75% of these structures have at least one six-membered ring; 69% of these contain five-membered rings, occupying a large portion. Three- and seven-membered ring structures account for just 21% and 24% of these structures, respectively. Four (10%) and eight (7%) membered ring structures (e.g. asteriscanolide) are seldom found [14]. With the lower costs of gene sequencing, recent developments in genome mining by sequencing and annotation have led to the discovery of a large number of functionally unknown terpene synthases [15-17], generating diverse complex structures and several bioactive products (e.g., 6α,9α,15-trihydroxycadinan-4-en-3-one, (+)-3,11,12-trihydroxycalamenene, and (−)-3,10,11,12-tetrahydroxycalamenene) [18].

Filamentous fungi are powerful producers of terpenoid products [19]. Many terpenoids produced by these fungi have recently been characterised; these terpenoids exhibit diverse complex structures and uncommon catalytic mechanisms [20]. However, a limited number of sesquiterpenes have been characterised from a few fungal taxa (e.g., trefolane A and sterhirsutins) [21]. *Trichoderma viride* is a filamentous fungus that has received considerable attention as an effective biocontrol agent against two fungal pathogens, *Fusarium oxysporum* f. sp. *adzuki* and *Pythium arrhenomanes*, infecting soybean. This fungus is a competent mycoparasite and strong producer of secondary metabolites [22,23]. However, *T. viride* terpenoids have rarely been studied and the low concentrations of products under natural conditions have limited the pace of research in this field. Metabolic engineering makes the overproduction of different terpenoids from *T. viride* possible [21,24]. To increase the discovery efficiency of terpenoid products, heterologous expression of various sources of terpene synthases in *Escherichia coli* and *Saccharomyces cerevisiae* is a feasible approach [25,26].

In this study, a combination of genome mining and metabolic engineering was used for sesquiterpenoid discovery, utilizing farnesyl diphosphate-overexpressing *S. cerevisiae* as a platform (Figure 1). By the heterologous expression of predicted terpene synthases from the genome of *T. viride*, an unknown sesquiterpene synthase was identified and characterised. Furthermore, a new compound produced by this enzyme and its esterified product were detected and characterised by GC–MS and 1D and 2D NMR, revealing a 5/6 bicyclic sesquiterpene and its C-11 esterified structure. Based on a literature search, to our knowledge, this is the first report of the characterisation of a sesquiterpene synthase in *T. viride*. In addition, this study demonstrates the effectiveness of the combination of genome mining and heterologous expression of predicted terpene synthases for detecting unknown terpenoids from rarely studied fungi.

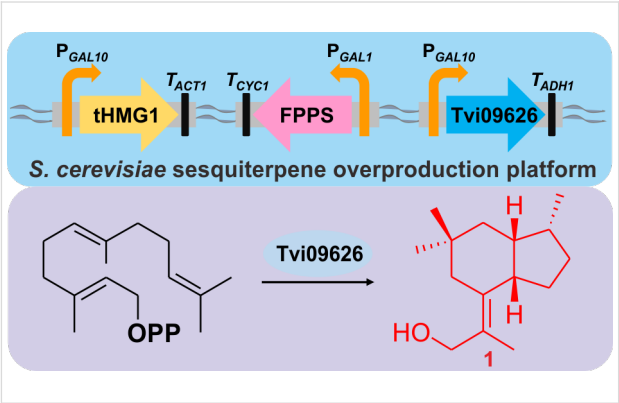


Figure 1: Schematic diagram of the *S. cerevisiae* sesquiterpene overproduction platform and the products of Tvi09626.

Results and Discussion

Prediction and analysis of terpene synthase genes in *T. viride* J1-030

Through genome sequencing of *T. viride* J1-030 and prediction of the potential terpene synthases in J1-030 genome, gene *Tvi09626* was selected and the following bioinformatics analysis of the function of this unidentified terpene synthase was performed. A protein blast search against the NCBI database was performed with Tvi09626, revealing sequence identities of 89.66% and 85.23% with the enzymes from the strain *T. virens* Gv29-8 [27] and *T. reesei* QM6a [28], respectively, with only predicted functions. Thereafter, an amino acid sequence alignment with several known terpene synthases showed that Tvi09626 had the typical highly conserved ^{128}**DDxxD/E** aspartate-rich motif, ^{276}**NSE/DTE** triad, ^{366}**RY** dimer and ^{230}**R** monomer (Figure S1, Supporting Information File 1) [29-31]. Furthermore, in a phylogenetic analysis (Figure 2), Tvi09626 belonged to Clade V of Class I terpene synthases. In previous studies, terpene synthases have been studied in the genus *Trichoderma*, such as trichodiene synthase homologous gene isolation and characterisation in *T. harzianum* [32] and functional identification of terpene synthase vir4 in *T. virens* [33]. However, owing to the general lack of previous studies of

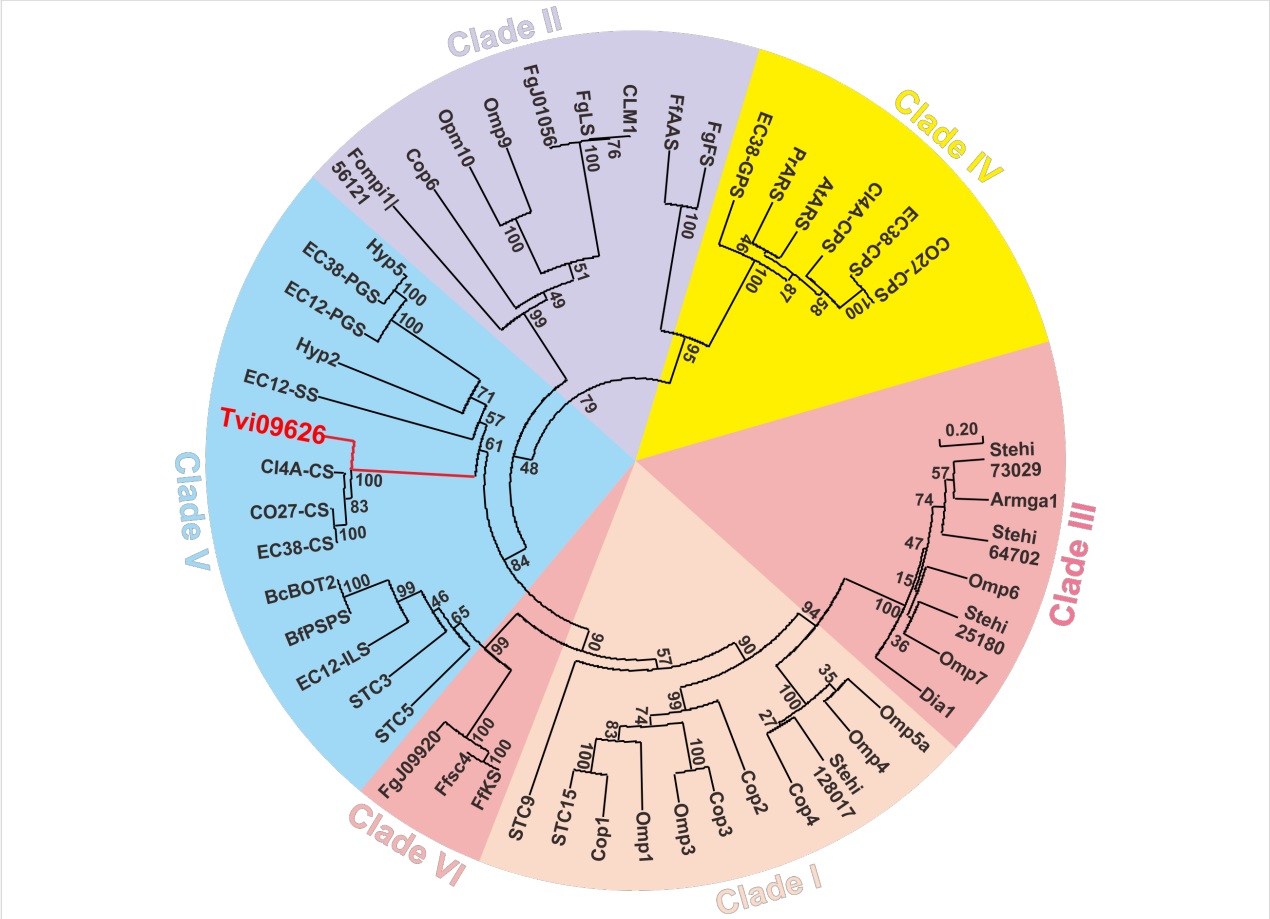


Figure 2: Phylogenetic analysis of Tvi09626 with other characterised terpene synthases. Six clades are marked with different colours and Tvi09626 is labelled in red in Clade V. Percentages indicate branch support based on 1,000 bootstrap replicates.

terpene synthases in *T. viride*, Tvi09626 is the first terpene synthase well-identified with products characterised in *T. viride*.

In vitro analysis of Tvi09626 function

To confirm the function of the candidate enzyme, the DNA sequence of *Tvi09626* was amplified by touchdown PCR from the *T. viride* genome. The gene fragment was cloned into a pET28a (+) vector to construct the plasmid pXS222. Next, pXS222 was transformed into BL21 to overexpress and purify Tvi09626 (Figure S2, Supporting Information File 1). The substrates GPP, FPP and GGPP were incubated with the protein individually and the products were detected and analysed by gas chromatography/mass spectrometry (GC–MS) [30,34]. In vitro assays clearly showed that Tvi09626 could use FPP as its only substrate to produce compound **1** (Figure 3 and Figure S3, Supporting Information File 1).

Heterologous expression of Tvi09626 in *S. cerevisiae*

To further verify the function of the putative terpene synthase, a metabolic engineering strategy was used to reconstruct an FPP

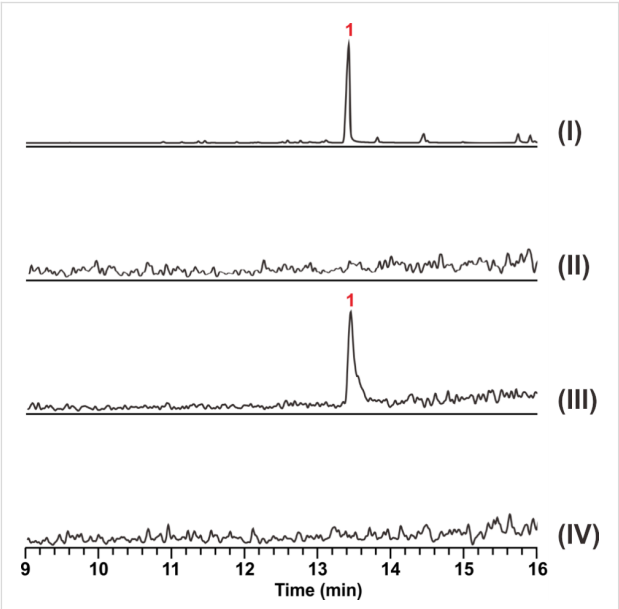


Figure 3: GC–MS chromatogram of products in vivo (I), in yeast YZL141 (II), in vitro Tvi09626 with FPP (III), and boiled Tvi09626 with FPP (IV).

overproduction platform in *S. cerevisiae* in order to obtain sufficient quantities of the products of Tvi09626 for chemical structural characterisation. *S. cerevisiae* YZL141, engineered previously [21], was used owing to its ability to provide enough IPP, DMAPP, and FPP for the production of terpenoids (Figure 1). After 72 h of shaken-flask fermentation, the strain was extracted with hexane/ethyl acetate (4:1), pre-separated by silica gel column chromatography, and detected by GC–MS. Similar to the in vitro assay results, compound **1** was the final product (Figure 4). Interestingly, during the extraction process, compound **2** was detected at a retention time of 14.53 min and identified as esterified compound **1** (Figure 4 and Table S4, Supporting Information File 1). Using the metabolic engineering strategy, the product of Tvi09626 was efficiently enriched via an abundant FPP supply.

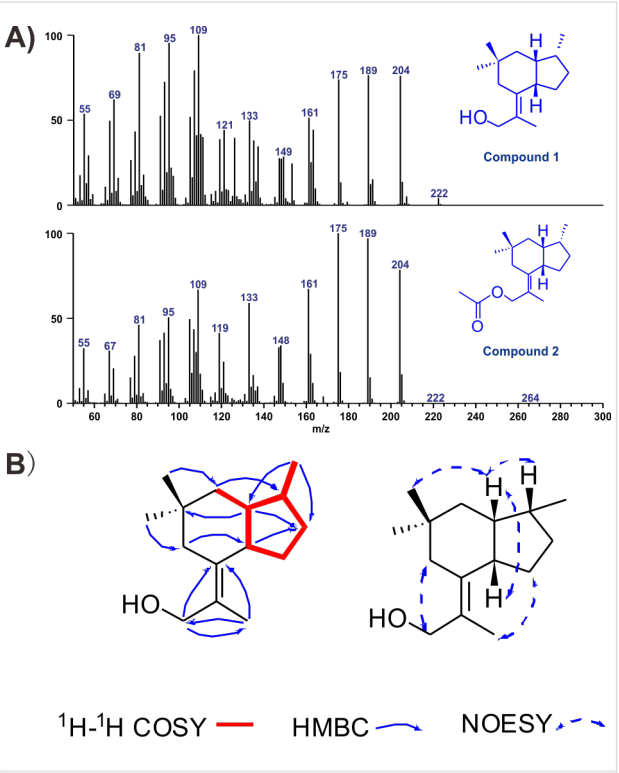


Figure 4: Characterisation of Tvi09626 products. (A) Mass spectra of compound **1** at t_R = 13.46 min with *m/z* 222 and compound **2** at t_R = 14.53 min with *m/z* 264. (B) COSY, HMBC and NOESY correlations for compound **1**.

Detection and characterisation of Tvi09626 products

By semi-preparative high-performance liquid chromatography (HPLC), we purified compound **1** and compound **2** (23.1 mg and 13.2 mg, respectively). The structures of the two new compounds were characterised by 1D and 2D NMR spectroscopy (Table 1, Table S4, and Figures S4–S15, Supporting Information File 1).

Table 1: ^{1}H NMR (400 MHz, $CDCl_3$) and ^{13}C NMR (100 MHz) data for compound **1** in $CDCl_3$.

Position	δ_C	δ_H
1	45.2	1.68 (dd, 8.2, 4.3 Hz, 1H)
2	40.6	1.38 (ddd, *J* = 12.8, 3.9, 1.6 Hz, 1H), 1.16 (t, *J* = 12.9 Hz,1H)
3	33	—
4	43.8	2.35 (dd, *J* = 13.8, 1.4 Hz, 1H), 1.60 (d, *J* = 13.8 Hz, 1H)
5	137.4	—
6	47.6	1.96 (dd, *J* = 12.3, 6.1 Hz, 1H)
7	30.9	2.10 (m, 1H), 1.60 (m, 1H)
8	33.3	2.07 (m, 1H), 1.07 (m, 1H)
9	31.6	1.99 (m, 1H)
10	126.5	—
11	65.2	4.18 (d, *J* = 11.6 Hz, 1H),3.97 (d, *J* = 9.8 Hz, 1H)
12	17.85	1.87 (t, *J* = 1.2 Hz, 3H)
13	32	0.96 (s, 3H)
14	26.1	0.83 (s, 3H)
15	17.78	0.80 (d, *J* = 7.0 Hz, 3H)
16	—	3.47 (s, 1H)

Compound **1** was a new compound with a known skeleton [35], isolated as a white powder. ^{1}H and ^{13}C NMR data showed four methyl groups at δ_H 1.87 (t, *J* = 1.2 Hz, 3H,), δ_H 0.96 (s, 3H), δ_H 0.83 (s, 3H), and δ_H 0.80 (d, *J* = 7.0 Hz, 3H). Five methylenes were detected, including an oxygenated one at δ_H 4.18 (d, *J* = 11.6 Hz, 1H), 3.97 (d, *J* = 9.8 Hz, 1H), as well as three methines and three quaternary carbons including a double bond at δ_C 137.4 (C-5), 126.5 (C-10). The 2D NMR data indicated that compound **1** is a 5/6 bicyclic sesquiterpene with the molecular formula $C_{15}H_{26}O$ (Figure 1). Interestingly, compound **1** contained a quaternary carbon with two methyl groups, which is uncommon for the cyclization mechanism of sesquiterpenoids and needs further investigation.

Compound **2** was purified as a white powder. ^{1}H and ^{13}C NMR data showed chemical shifts of five methyl groups at δ_H 1.83 (t, *J* = 1.2 Hz, 3H), δ_H 0.96 (s, 3H), δ_H 0.82 (s, 3H), δ_H 0.80 (d, *J* = 7.0 Hz, 3H), and δ_H 2.06 (s, 3H). Five methylenes were identified, including an esterified group at C-11 with a resonance of δ_H 4.64 (d, *J* = 11.6 Hz 1H), 4.46 (d, *J* = 11.6 Hz, 1H), three methines, and three quaternary carbons including a double bond at δ_C 140.37 (C-5), 122.04 (C-10). Compared with 2D NMR information of compound **1**, compound **2** was a C-11 esterified **1** with the molecular formula $C_{17}H_{28}O_2$ (Table S4, Supporting Information File 1), which may represent the esterification reaction during the extraction process.

The 5/6 bicyclic sesquiterpene identified and characterised in this study was a brasilane-type sesquiterpenoid; this sesquiter-

penoid type is typically isolated from cultures of the basidiomycete *Coltricia sideroides* in combination with its two new alkane derivatives colisiderin A and (7*E*,9*E*)-undeca-7,9-diene-2,4,5-triol [35] and from the organic extract of the red alga *Laurencia obtusa* [36]. However, to the best of our knowledge, ours is the first report of these brasilane-type sesquiterpenes obtained via biosynthetic genes.

Metal ion dependency of Tvi09626 and its kinetics

As reported previously, most terpene synthases are active in the presence of Mg^{2+} ions [8,37]. To test the Mg^{2+} dependency of Tvi09626, an in vitro assay was performed. The GC–MS analysis showed that in the presence of Mg^{2+}, compound **1** can be obtained, whereas without Mg^{2+} or added EDTA (2.5 mM), compound **1** cannot be detected (Figure 5). This assay demonstrated that Tvi09626 was a Mg^{2+}-dependent sesquiterpene synthase. In a kinetics analysis, the turnover rate (k_{cat}) of the enzyme with FPP was $(15 \pm 0.3) \times 10^{-2}$, which is similar to those of omp6 and omp7. Its substrate affinity (K_m) was $(0.44 \pm 0.11) \times 10^{-6}$, one-tenth that of omp6 and nearly a quarter that of omp7. The catalytic efficiency (k_{cat}/K_m) of Tvi09626 was $(35.32 \pm 0.57) \times 10^3$, higher than that of omp6 and lower than that of omp7 [21,38].

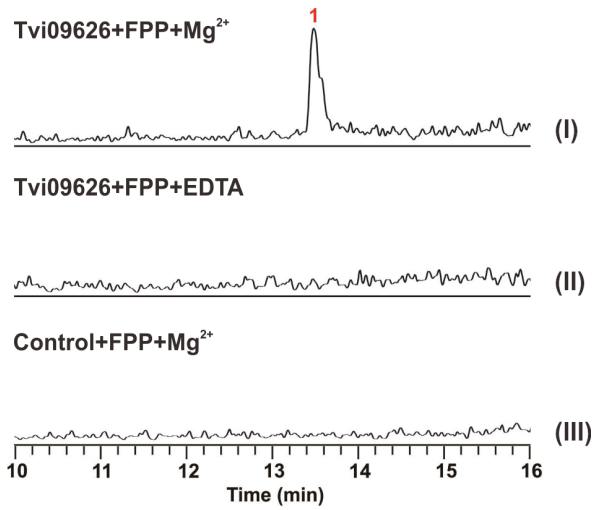


Figure 5: GC–MS chromatogram for the metal ion dependency assay.

Conclusion

In conclusion, we identified a novel sesquiterpene synthase, Tvi09626, from *T. viride* using a strong sesquiterpene overproduction platform with *S. cerevisiae* YZL141; it is the first biochemically identified and characterised sesquiterpene synthase from this filamentous fungus. In an analysis of its relative structure, the product of this enzyme was characterised as a 5/6 bicyclic sesquiterpene compound **1** oxygenated at C-11. Interestingly, esterified compound **1** was isolated during the product extraction process, suggesting an esterification reaction. It contained a quaternary carbon with two methyl groups, which is uncommon of the cyclization mechanism of sesquiterpenoids and needs to be studied in the future. To the best of our knowledge, this study reports the first use of a biosynthetic gene to obtain a brasilane-type sesquiterpene.

Supporting Information

Supporting Information File 1

Experimental part and supplementary figures and tables.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-202-S1.pdf]

Acknowledgements

We thank Dr. You-Sheng Cai (Wuhan University) for suggestions regarding structural characterisation. This work was financially supported by funding from the National Key R&D Program of China (2018YFA0900400), the Medical Science Advancement Program (Clinical Medicine) of Wuhan University, the state Key Laboratory of Microbial Metabolism, Shanghai Jiao Tong University (grant MMLKF 18-12) and the National Natural Science Foundation of China (31800032).

ORCID® iDs

Xiang Sun - https://orcid.org/0000-0001-6703-0413
Tiangang Liu - https://orcid.org/0000-0001-8087-0345

References

1. Bian, G. K.; Ma, T.; Liu, T. G. *Methods Enzymol.* **2018,** *608,* 97–129. doi:10.1016/bs.mie.2018.04.025
2. Huber, T.; Weisheit, L.; Magauer, T. *Beilstein J. Org. Chem.* **2015,** *11,* 2521–2539. doi:10.3762/bjoc.11.273
3. Christianson, D. W. *Chem. Rev.* **2017,** *117,* 11570–11648. doi:10.1021/acs.chemrev.7b00287
4. Pemberton, T. A.; Chen, M.; Harris, G. G.; Chou, W. K. W.; Duan, L.; Köksal, M.; Genshaft, A. S.; Cane, D. E.; Christianson, D. W. *Biochemistry* **2017,** *56,* 2010–2023. doi:10.1021/acs.biochem.7b00137
5. Lauterbach, L.; Rinkel, J.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2018,** *57,* 8280–8283. doi:10.1002/anie.201803800
6. Huang, A. C.; Hong, Y. J.; Bond, A. D.; Tantillo, D. J.; Osbourn, A. *Angew. Chem., Int. Ed.* **2018,** *57,* 1291–1295. doi:10.1002/anie.201711444
7. Bian, G.; Rinkel, J.; Wang, Z.; Lauterbach, L.; Hou, A.; Yuan, Y.; Deng, Z.; Liu, T.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2018,** *57,* 15887–15890. doi:10.1002/anie.201809954
8. Yang, Y.-l.; Zhang, S.; Ma, K.; Xu, Y.; Tao, Q.; Chen, Y.; Chen, J.; Guo, S.; Ren, J.; Wang, W.; Tao, Y.; Yin, W.-B.; Liu, H. *Angew. Chem., Int. Ed.* **2017,** *56,* 4749–4752. doi:10.1002/anie.201700565

9. Köksal, M.; Jin, Y.; Coates, R. M.; Croteau, R.; Christianson, D. W. *Nature* **2011,** *469,* 116–120. doi:10.1038/nature09628
10. Rabe, P.; Rinkel, J.; Nubbemeyer, B.; Köllner, T. G.; Chen, F.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2016,** *55,* 15420–15423. doi:10.1002/anie.201608971
11. Peng, B.; Plan, M. R.; Chrysanthopoulos, P.; Hodson, M. P.; Nielsen, L. K.; Vickers, C. E. *Metab. Eng.* **2017,** *39,* 209–219. doi:10.1016/j.ymben.2016.12.003
12. Matsuda, Y.; Mitsuhashi, T.; Quan, Z.; Abe, I. *Org. Lett.* **2015,** *17,* 4644–4647. doi:10.1021/acs.orglett.5b02404
13. Kim, D.; Lee, E.; Lee, J.; Leutou, A.; Shin, Y.-H.; Choi, B.; Hwang, J.; Hahn, D.; Choi, H.; Chin, J.; Cho, S.; Hong, Y.; Ko, J.; Seong, C.; Maloney, K.; Oh, D.-C.; Yang, I.; Hwang, H.; Nam, S.-J. *Mar. Drugs* **2018,** *16,* 130. doi:10.3390/md16040130
14. Klapschinski, T. A.; Rabe, P.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2016,** *55,* 10141–10144. doi:10.1002/anie.201605425
15. Ye, Y.; Minami, A.; Mandi, A.; Liu, C.; Taniguchi, T.; Kuzuyama, T.; Monde, K.; Gomi, K.; Oikawa, H. *J. Am. Chem. Soc.* **2015,** *137,* 11846–11853. doi:10.1021/jacs.5b08319
16. Nakano, C.; Kudo, F.; Eguchi, T.; Ohnishi, Y. *ChemBioChem* **2011,** *12,* 2271–2275. doi:10.1002/cbic.201100418
17. Hu, Y.; Chou, W. K. W.; Hopson, R.; Cane, D. E. *Chem. Biol.* **2011,** *18,* 32–37. doi:10.1016/j.chembiol.2010.11.008
18. Cao, L.; Shehla, N.; Tasneem, S.; Cao, M.; Sheng, W.; Jian, Y.; Li, B.; Peng, C.; Choudhary, M. I.; Atta-ur-Rahman; Liao, D.-f.; Wang, W. *Molecules* **2019,** *24,* 1664. doi:10.3390/molecules24091664
19. Yan, Y.; Liu, Q.; Zang, X.; Yuan, S.; Bat-Erdene, U.; Nguyen, C.; Gan, J.; Zhou, J.; Jacobsen, S. E.; Tang, Y. *Nature* **2018,** *559,* 415–418. doi:10.1038/s41586-018-0319-4
20. Minami, A.; Ozaki, T.; Liu, C.; Oikawa, H. *Nat. Prod. Rep.* **2018,** *35,* 1330–1346. doi:10.1039/c8np00026c
21. Bian, G.; Hou, A.; Yuan, Y.; Hu, B.; Cheng, S.; Ye, Z.; Di, Y.; Deng, Z.; Liu, T. *Org. Lett.* **2018,** *20,* 1626–1629. doi:10.1021/acs.orglett.8b00366
22. Mannina, L.; Segre, A. L.; Ritieni, A.; Fogliano, V.; Vinale, F.; Randazzo, G.; Maddau, L.; Bottalico, A. *Tetrahedron* **1997,** *53,* 3135–3144. doi:10.1016/s0040-4020(97)00024-0
23. John, R. P.; Tyagi, R. D.; Prévost, D.; Brar, S. K.; Pouleur, S.; Surampalli, R. Y. *Crop Prot.* **2010,** *29,* 1452–1459. doi:10.1016/j.cropro.2010.08.004
24. Bian, G.; Deng, Z.; Liu, T. *Curr. Opin. Biotechnol.* **2017,** *48,* 234–241. doi:10.1016/j.copbio.2017.07.002
25. Peng, B.; Nielsen, L. K.; Kampranis, S. C.; Vickers, C. E. *Metab. Eng.* **2018,** *47,* 83–93. doi:10.1016/j.ymben.2018.02.005
26. Yap, H.-Y. Y.; Muria-Gonzalez, M. J.; Kong, B.-H.; Stubbs, K. A.; Tan, C.-S.; Ng, S.-T.; Tan, N.-H.; Solomon, P. S.; Fung, S.-Y.; Chooi, Y.-H. *Microb. Cell Fact.* **2017,** *16,* 103. doi:10.1186/s12934-017-0713-x
27. Kubicek, C. P.; Herrera-Estrella, A.; Seidl-Seiboth, V.; Martinez, D. A.; Druzhinina, I. S.; Thon, M.; Zeilinger, S.; Casas-Flores, S.; Horwitz, B. A.; Mukherjee, P. K.; Mukherjee, M.; Kredics, L.; Alcaraz, L. D.; Aerts, A.; Antal, Z.; Atanasova, L.; Cervantes-Badillo, M. G.; Challacombe, J.; Chertkov, O.; McCluskey, K.; Coulpier, F.; Deshpande, N.; von Döhren, H.; Ebbole, D. J.; Esquivel-Naranjo, E. U.; Fekete, E.; Flipphi, M.; Glaser, F.; Gómez-Rodríguez, E. Y.; Gruber, S.; Han, C.; Henrissat, B.; Hermosa, R.; Hernández-Oñate, M.; Karaffa, L.; Kosti, I.; Le Crom, S.; Lindquist, E.; Lucas, S.; Lübeck, M.; Lübeck, P. S.; Margeot, A.; Metz, B.; Misra, M.; Nevalainen, H.; Omann, M.; Packer, N.; Perrone, G.; Uresti-Rivera, E. E.; Salamov, A.; Schmoll, M.; Seiboth, B.; Shapiro, H.; Sukno, S.; Tamayo-Ramos, J. A.; Tisch, D.; Wiest, A.; Wilkinson, H. H.; Zhang, M.; Coutinho, P. M.; Kenerley, C. M.; Monte, E.; Baker, S. E.; Grigoriev, I. V. *Genome Biol.* **2011,** *12,* R40. doi:10.1186/gb-2011-12-4-r40
28. Martinez, D.; Berka, R. M.; Henrissat, B.; Saloheimo, M.; Arvas, M.; Baker, S. E.; Chapman, J.; Chertkov, O.; Coutinho, P. M.; Cullen, D.; Danchin, E. G. J.; Grigoriev, I. V.; Harris, P.; Jackson, M.; Kubicek, C. P.; Han, C. S.; Ho, I.; Larrondo, L. F.; de Leon, A. L.; Magnuson, J. K.; Merino, S.; Misra, M.; Nelson, B.; Putnam, N.; Robbertse, B.; Salamov, A. A.; Schmoll, M.; Terry, A.; Thayer, N.; Westerholm-Parvinen, A.; Schoch, C. L.; Yao, J.; Barabote, R.; Nelson, M. A.; Detter, C.; Bruce, D.; Kuske, C. R.; Xie, G.; Richardson, P.; Rokhsar, D. S.; Lucas, S. M.; Rubin, E. M.; Dunn-Coleman, N.; Ward, M.; Brettin, T. S. *Nat. Biotechnol.* **2008,** *26,* 553–560. doi:10.1038/nbt1403
29. Yuan, Y.; Litzenburger, M.; Cheng, S.; Bian, G.; Hu, B.; Yan, P.; Cai, Y.; Deng, Z.; Bernhardt, R.; Liu, T. *ChemBioChem* **2019,** *20,* 677–682. doi:10.1002/cbic.201800670
30. Burkhardt, I.; Siemon, T.; Henrot, M.; Studt, L.; Rösler, S.; Tudzynski, B.; Christmann, M.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2016,** *55,* 8748–8751. doi:10.1002/anie.201603782
31. Burkhardt, I.; Kreuzenbeck, N. B.; Beemelmanns, C.; Dickschat, J. S. *Org. Biomol. Chem.* **2019,** *17,* 3348–3355. doi:10.1039/c8ob02744g
32. Gallo, A.; Mulè, G.; Favilla, M.; Altomare, C. *Physiol. Mol. Plant Pathol.* **2004,** *65,* 11–20. doi:10.1016/j.pmpp.2004.11.005
33. Crutcher, F. K.; Parich, A.; Schuhmacher, R.; Mukherjee, P. K.; Zeilinger, S.; Kenerley, C. M. *Fungal Genet. Biol.* **2013,** *56,* 67–77. doi:10.1016/j.fgb.2013.05.003
34. Zhu, F.; Zhong, X.; Hu, M.; Lu, L.; Deng, Z.; Liu, T. *Biotechnol. Bioeng.* **2014,** *111,* 1396–1405. doi:10.1002/bit.25198
35. Hu, D.-B.; Zhang, S.; He, J.-B.; Dong, Z.-J.; Li, Z.-H.; Feng, T.; Liu, J.-K. *Fitoterapia* **2015,** *104,* 50–54. doi:10.1016/j.fitote.2015.05.009
36. Iliopoulou, D.; Vagias, C.; Galanakis, D.; Argyropoulos, D.; Roussis, V. *Org. Lett.* **2002,** *4,* 3263–3266. doi:10.1021/ol026506z
37. Takino, J.; Kozaki, T.; Sato, Y.; Liu, C.; Ozaki, T.; Minami, A.; Oikawa, H. *J. Am. Chem. Soc.* **2018,** *140,* 12392–12395. doi:10.1021/jacs.8b08925
38. Wawrzyn, G. T.; Quin, M. B.; Choudhary, S.; López-Gallego, F.; Schmidt-Dannert, C. *Chem. Biol.* **2012,** *19,* 772–783. doi:10.1016/j.chembiol.2012.05.012

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Harnessing enzyme plasticity for the synthesis of oxygenated sesquiterpenoids

Melodi Demiray, David J. Miller and Rudolf K. Allemann*

Full Research Paper

Open Access

Address:
School of Chemistry, Cardiff University, Main Building, Park Place, Cardiff, CF10 3AT. United Kingdom

Email:
Rudolf K. Allemann* - AllemannRK@cardiff.ac.uk

* Corresponding author

Keywords:
artemisinin; amorphadiene synthase; oxygenated terpenoids; sesquiterpenoids; substrate engineering; terpenes

Beilstein J. Org. Chem. **2019,** *15,* 2184–2190.
doi:10.3762/bjoc.15.215

Received: 09 June 2019
Accepted: 26 August 2019
Published: 17 September 2019

This article is part of the thematic issue "Terpenes".

Guest Editor: J. S. Dickschat

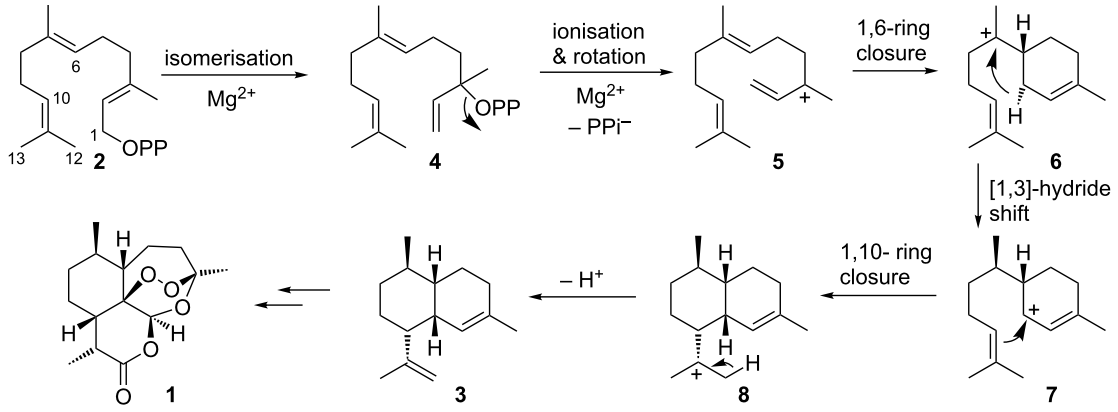
Abstract

8-Methoxy-γ-humulene, (*E*)-8-methoxy-β-farnesene, 12-methoxy-β-sesquiphellandrene and 12-methoxyzingiberene can be synthesised in amorphadiene synthase-catalysed reactions from 8- and 12-methoxyfarnesyl diphosphates due to the highly plastic yet tightly controlled carbocationic chemistry of this sesquiterpene cyclase.

Introduction

Amorphadiene synthase (ADS) from *Artemisia annua* is a key enzyme involved in the biosynthesis of the antimalarial sesquiterpene drug artemisinin (**1**) [1-4]. ADS catalyses the Mg^{2+}-dependent conversion of farnesyl diphosphate (FDP, **2**) to amorpha-4,11-diene (**3**) with high regio- and stereochemical control (Scheme 1) [5-7]. The carbocationic reaction mechanism of ADS involves an isomerisation to nerolidyl diphosphate (NDP, **4**) followed by breakage of the carbon–oxygen bond to generate the allylic cation **5**. This allows rotation around the C2–C3 bond and 1,6-ring closure to form the bisabolyl cation (**6**). A [1,3]-hydride shift to form carbocation **7** and 1,10-ring closure yield the amorphyl cation (**8**). Finally, deprotonation generates amorpha-4,11-diene (**3**) [8,9].

Several sesquiterpene synthases including ADS accept FDP analogues containing a variety of heteroatoms and functional groups to generate unnatural sesquiterpenoids that are not easily accessible by conventional organic synthesis [10-19]. Creating novel sesquiterpenoids, not normally found in nature, is of great interest due to the important applications of terpenoids in healthcare and agriculture as well as the potential to tailor their properties to specific needs. For example, fluorinated derivatives of (*E*)-β-farnesene, a potent alarm pheromone for aphids [20-23], are more effective as pheromones than the parent compound, and finding high potency derivatives of (*E*)-β-farnesene may be of significant benefit in agriculture [24]. While (*S*)-germacrene D is a highly volatile but unstable olfactory signal that repels invertebrate arthropod pests (insects, ticks, mites) that affect humans and livestock as well as arable crops, (*S*)-14,15-dimethylgermacrene D acts as an attractant of aphids [10]. β-Sesquiphellandrene and α-curcumene are both found in turmeric (*Curcuma longa*) and have been shown to have anti-



Scheme 1: Mechanism of the ADS-catalysed conversion of FDP (**2**) to amorpha-4,11-diene (**3**), a biosynthetic precursor of artemisinin (**1**).

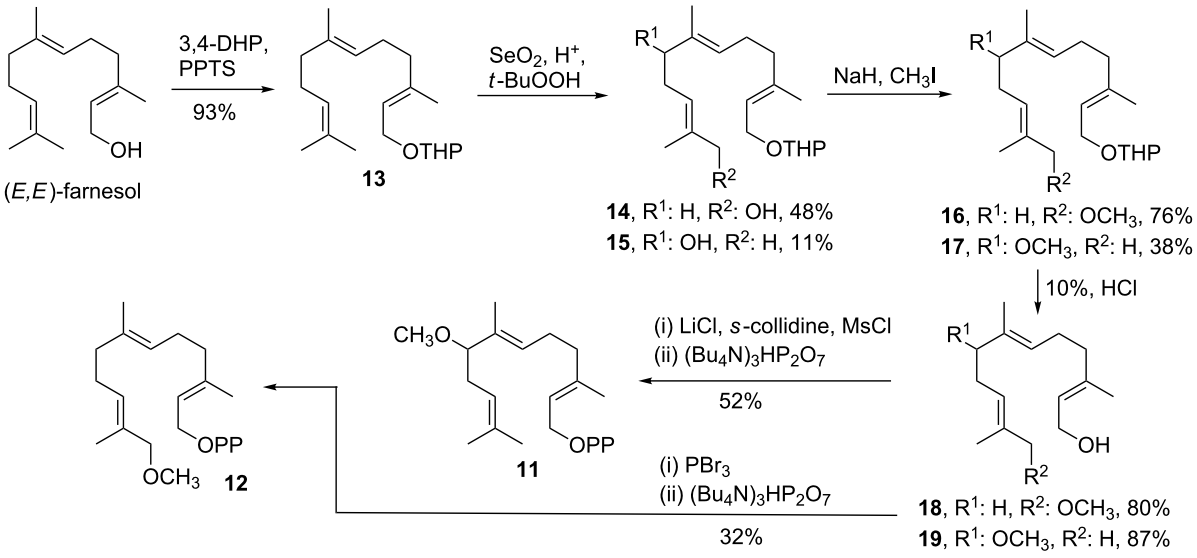
cancer activity [25,26]. The oxygenated α-curcumene and β-sesquiphellandrene derivatives α- and β-turmerone are reported to possess anticonvulsant properties and are used to treat epilepsy [27,28]. This array of important compounds shows the potential of generating novel sesquiterpenoids with desirable bio-properties.

ADS is a high fidelity sesquiterpene synthase that produces almost exclusively a single product. Its active site plasticity nevertheless allows the conversion of 12-hydroxy-FDP (**9**) to dihydroartemisinic aldehyde (**10**), a biosynthetic intermediate and valuable precursor in the synthesis of artemisinin [29].

Results and Discussion

Here we report that ADS accepts the bulkier FDP analogues 8-methoxy-FDP (**11**) and 12-methoxy FDP (**12**) as substrates, thereby opening up direct and efficient synthetic routes to oxidised sesquiterpenoids. This represents in essence a reversal of the biosynthetic pathways to oxidised sesquiterpenoids since cyclisation occurs after FDP 'oxidation'. 8-Methoxy-FDP (**11**) and 12-methoxy-FDP (**12**) were both prepared in the same manner, beginning with a tetrahydropyranyl (THP) protection of (*E*,*E*)-farnesol to form **13** [29,30]. This was followed by a selenium dioxide oxidation that produced 12-hydroxyfarnesol (**14**) as the major product (48%, already published in [29]) in addition to 8-hydroxyfarnesol (**15**, 11%) [31]. Both of these alcohols were methylated with methyl iodide to yield **16** and **17** [32]. To produce the final FDP analogues, THP was removed to generate alcohols **18** and **19**, and subsequently diphosphorylated via halogenated intermediates (Scheme 2 and Supporting Information File 1) [33-35].

GC–MS analysis (Figure 1) of the organic soluble products formed from 8-methoxy FDP (**11**) through ADS catalysis



Scheme 2: Synthesis of 8-methoxy-FDP (**11**) and 12-methoxy-FDP (**12**) (for full synthesis details see Supporting Information File 1).

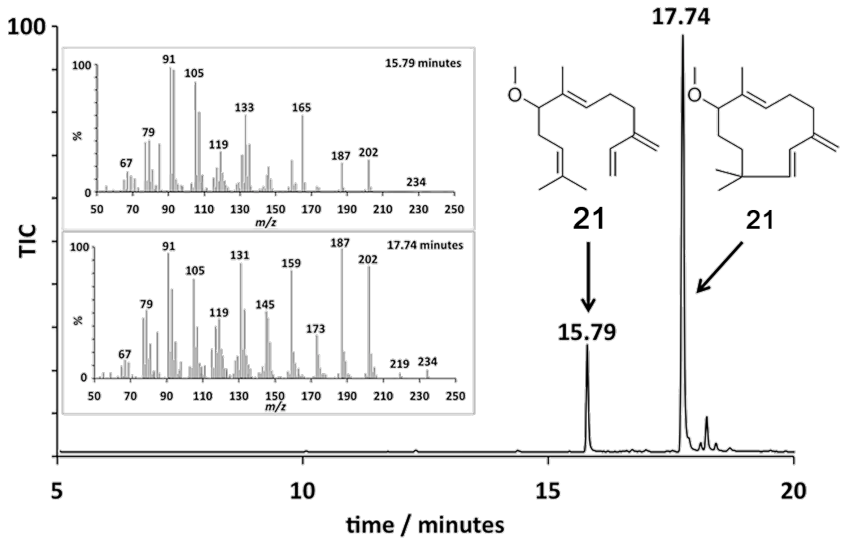


Figure 1: Total-ion chromatogram of the pentane extractable products formed in an incubation of ADS with 8-methoxy-FDP (**11**). Inset: EI$^+$ Mass spectra of the eluted compounds **20** and **21**.

revealed the formation of a major sesquiterpenoid of molecular mass 234 (85%). No organic soluble products were detected when ADS was omitted from the incubation mixture. This product was identified as 8-methoxy-γ-humulene (**20**) by NMR spectroscopy and comparison of its ^{1}H NMR spectrum (Figure 2) with that of 8-oxo-γ-humulene, a natural sesquiterpenoid isolated from the plant *Cineraria fruticulorum* [36]. The signals for H5 (δ_H = 5.44 ppm, d, $J_{H,H}$ = 16.0 Hz), H6 (δ_H = 6.06, d, $J_{H,H}$ = 16.0 Hz), H15 (δ_H = 5.00, bs) and H15' (δ_H = 5.02, bs) of 8-oxo-γ-humulene [36] are in agreement with their corresponding equivalents in **20** (H2 (δ_H = 5.48, d, $J_{H,H}$ = 16.0 Hz), H1 (δ_H = 5.88, d, $J_{H,H}$ = 16.0 Hz), H15 (δ_H = 4.81, d, $J_{H,H}$ = 2.5 Hz) and H15' (δ_H = 4.89, d, $J_{H,H}$ = 2.5 Hz)). The resonances for the methyl groups at C12 and C13 are also analogous, appearing as singlets at 1.03 ppm in 8-oxo-γ-humulene and 0.96 and 0.97 ppm in **20**. The identity of the minor product (15%) was confirmed as (*E*)-8-methoxy-β-farnesene (**21**) by GC co-elution and comparison of its mass spectrum with an authentic sample prepared by exposing (*E*)-β-farnesene synthase to diphosphate **11** (Supporting Information

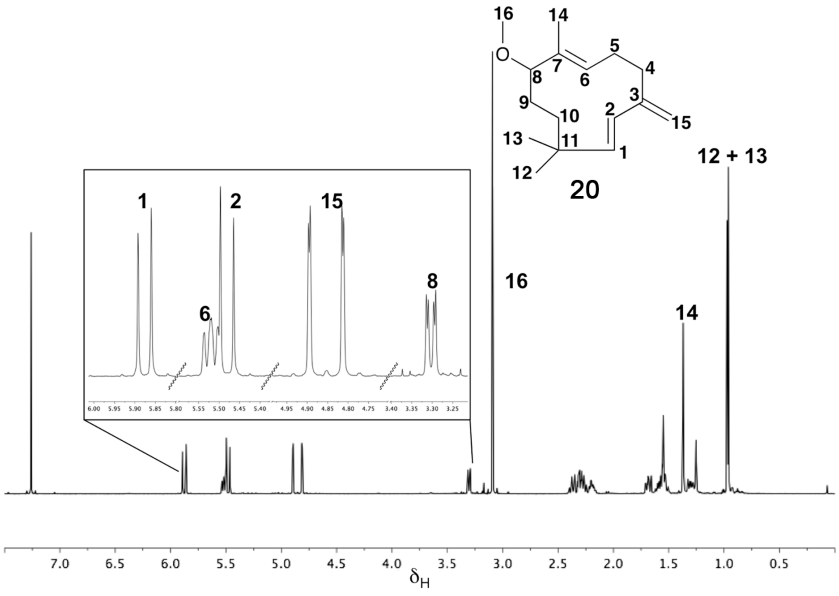


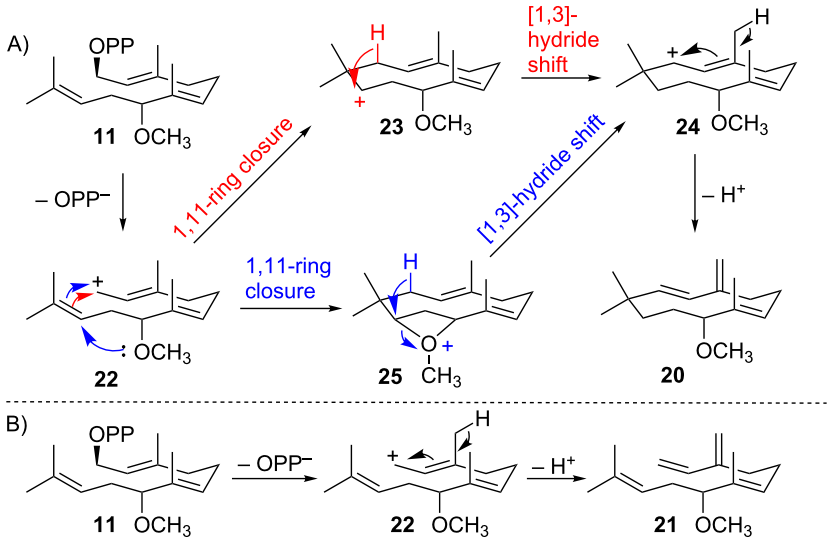
Figure 2: ^{1}H NMR spectrum (500 MHz, $CDCl_3$) of the 8-methoxy-γ-humulene (**20**) generated by ADS from 8-methoxy-FDP (**11**).

File 1). Further support for the structure of the minor product came from the excellent agreement of the diagnostic ^{1}H NMR signals of **21** (H1 (δ_H = 5.25, d, $J_{H,H}$ = 17.5 Hz), H1' (δ_H = 5.06, d, $J_{H,H}$ = 11.0 Hz), H2 (δ_H = 6.38, dd, $J_{H,H}$ = 17.5 and 11.0 Hz), H15 (δ_H = 5.02, s), H15' (δ_H = 5.00, s)) with those reported for the parent sesquiterpene (*E*)-β-farnesene (Supporting Information File 1) [37]. The proton on C6 (δ_H = 5.34, br, t) resonates further downfield for **21** compared to the corresponding proton of (*E*)-β-farnesene (δ_H = 5.17, t, $J_{H,H}$ = 7.0 Hz) due to the presence of the methoxy group two carbon atoms away.

The ADS-catalysed 1,11-cyclisation of diphosphate **11** suggests that the 8-methoxy group prevents the formation of a conformation conducive to isomerisation to NDP (**4**, Scheme 1) and hence 1,6-cyclisation to generate a bisabolyl cation and the amorphane skeleton. Rather the active site conformations of **11** and cation **22** appear to enable a 1,11-cyclisation to **23**.

A subsequent [1,3]-hydride shift to cation **24** and deprotonation from C15 lead to 8-methoxy-γ-humulene (**20**, Scheme 3A). Alternatively, the nucleophilic 8-methoxy group could react at C10 to induce a fast 1,11-cyclisation, forming cation **25**, which effectively competes with the isomerization of **11** to 8-methoxy-NDP. A subsequent [1,3]-hydride shift leads to **24** (Scheme 3A). Direct deprotonation of **22** at C15 forms the minor reaction product (*E*)-8-methoxy-β-farnesene (**21**) (Scheme 3B). Due to the racemic nature of the starting diphosphate **11b** it is, however, also possible that each enzymatic product arises independently from the individual enantiomers.

GC–MS analysis of the organic soluble products generated from an incubation of ADS with 12-methoxy-FDP (**12**) revealed the formation of a 1:2.4 mixture of two sesquiterpenoids of mass *m/z* 234 (Figure 3). Again, no organic soluble products were detected when ADS was omitted from the incubation mixture. The ^{1}H NMR spectrum of this product mixture (Figure 4) and comparison with the ^{1}H NMR spectra of the bisabolyl-derived sesquiterpenes β-sesquiphellandrene [38-40] and zingiberene [40], a hydrocarbon with antifertility, antiviral and anticancer activity [41], suggested that the major compound was 12-methoxy-β-sesquiphellandrene (**26**), while the minor product was identified as 12-methoxyzingiberene (**27**). The two doublets observed at 0.85 and 0.87 ppm correspond to the C14H$_3$ groups for both compounds. The ^{13}C-DEPT-135 spectrum (see Supporting Information File 1) showed an inverted peak at ≈109 ppm, implying the presence of an olefinic CH_2 that couples to the two overlapped doublets at 4.74 ppm. The integration of these doublets as 2 protons suggested that the exocyclic alkene was present in only one of the products. In the literature, this exocyclic alkene in β-sesquiphellandrene is observed at 4.72 ppm as a multiplet [39]. A triplet at 5.39 ppm that integrates for 2 protons can be assigned to H10 in both **26** and **27**, which resonates further downfield than in β-sesquiphellandrene and zingiberene due to the methoxy group positioned two carbons away. The olefinic protons H1 (δ_H = 5.67, d, $J_{H,H}$ = 10.0 Hz) and H2 (δ_H = 6.14, d, $J_{H,H}$ = 10.0 Hz) of **26** are in agreement with the equivalent protons H1 (δ_H = 5.66, dd, $J_{H,H}$ = 10.0 and 2.5 Hz) and H2 (δ_H = 6.13, d, $J_{H,H}$ = 10.0 Hz) in β-sesquiphellandrene. Similarly, the signals for the olefinic protons H1 (δ_H = 5.63), H2 (δ_H = 5.77) and H4 (δ_H = 5.45) in



Scheme 3: Potential mechanisms for the ADS-catalysed conversion of 8-methoxy-FDP (**11**) to 8-methoxy-γ-humulene (**20**) and (*E*)-8-methoxy-β-farnesene (**21**).

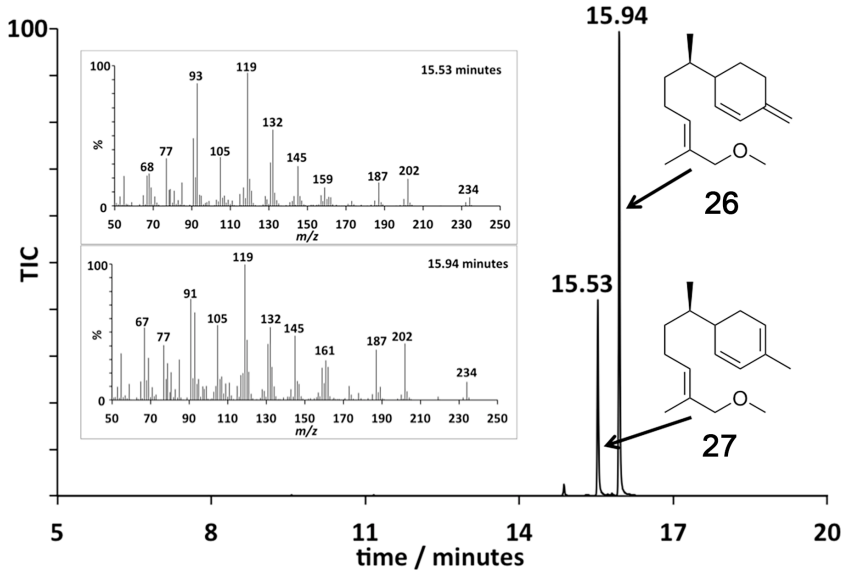


Figure 3: Total-ion chromatogram of the pentane extractable products formed in an incubation of ADS with 12-methoxy-FDP (**12**). Inset: Mass spectra of **26** and **27**.

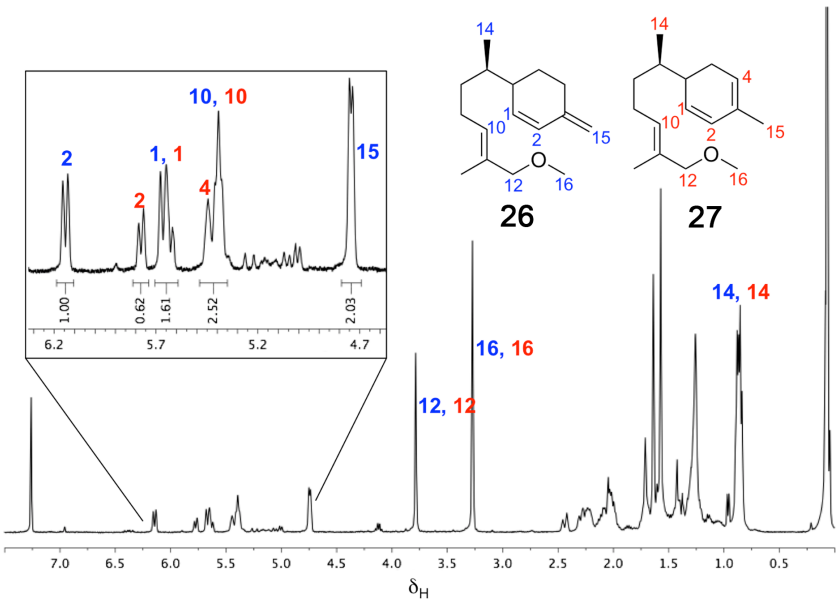
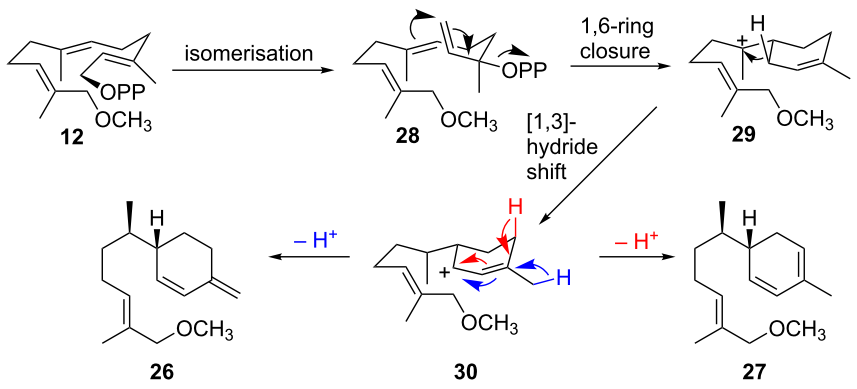


Figure 4: ^{1}H NMR spectrum (400 MHz, $CDCl_3$) of 12-methoxy-β-sesquiphellandrene (**26**) and 12-methoxyzingiberene (**27**) produced by ADS from 12-methoxy-FDP (**12**).

27 correspond to the equivalent protons (δ_H = 5.61), H2 (δ_H = 5.57) and H4 (δ_H = 5.42) in zingiberene [40].

Mechanistically, isomerisation of **12** to the methoxy-NDP analogue **28** allows for an ADS-catalysed 1,6-cyclisation to the 12-methoxy-bisabolyl cation (**29**) followed by a [1,3]-hydride shift, which relocates the positive charge on C1 in **30**. With FDP and 12-hydroxy-FDP, a subsequent 1,10-ring closure has been proposed previously [29]. However, the presence of monocyclic products indicates that this second ring closure does not occur with 12-methoxy-FDP (**12**).

It is suggested that the 12-methoxy group enforces an orientation of the distal 10,11-double bond that is not conducive to the



Scheme 4: Possible mechanisms for the ADS-catalysed conversion of 12-methoxy-FDP (**12**) to 12-methoxy-β-sesquiphellandrene (**26**) and 12-methoxyzingiberene (**27**).

second ring closure. Subsequent deprotonation from C15 and C4 from intermediate **30** affords 12-methoxy-β-sesquiphellandrene (**26**) and 12-methoxyzingiberene (**27**), respectively (Scheme 4).

Conclusion

In conclusion, the class I sesquiterpene cyclase amorphadiene synthase facilitates the efficient conversion of readily accessible synthetic methoxy-FDPs to sesquiterpenoids that may have applications in healthcare and agriculture. These results inform us of both the utility and limitations that non-natural functional groups have upon terpene cyclase-catalysed reaction cascades supporting the design of future biocatalytic syntheses. In particular, the presence of an ethereal oxygen atom containing π-acid functionality alongside its inductive withdrawal effect has a profound effect on the carbocationic reactivity of the intermediates. Of course a fully comprehensive interpretation of these results, regarding potentially interesting aspects such as anchimeric assistance is hampered by unknowns such as the conformation of binding to the enzyme and what effect the extra bulk of the substituents has upon the results observed, but nevertheless such empirical results will accumulate to inform future investigations. This reversal of the biosynthetic reaction order is expandable to other terpene synthases to generate libraries of unnatural sesquiterpenoids with a wide range of potential uses and applications across many areas of human activity.

Supporting Information

Supporting Information File 1

Experimental part.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-215-S1.pdf]

Acknowledgements

We thank Dr. Veronica Gonzalez for helpful discussions. The help of Dr. Juan A. Faraldos with early parts of this work and of Dr. Rob Jenkins and Mr Thomas Williams with mass spectrometry and NMR spectroscopy is gratefully acknowledged. This work forms part of the Ph.D. project carried out by Dr. Melodi Demiray and is contained within her Ph.D. Thesis (Cardiff University, 2017) and was supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC) through grants (BB/H01683X/1, BB/M022463/1, BB/N012526/1) and the Engineering and Physical Sciences Research Council (EPSRC) through grant (EP/K502819/1).

ORCID® iDs

Melodi Demiray - https://orcid.org/0000-0002-0317-4124
David J. Miller - https://orcid.org/0000-0003-4956-2656
Rudolf K. Allemann - https://orcid.org/0000-0002-1323-8830

References

1. Klayman, D. *Science* **1985,** *228,* 1049–1055. doi:10.1126/science.3887571
2. White, N. J. *J. Clin. Invest.* **2004,** *113,* 1084–1092. doi:10.1172/jci21682
3. Krishna, S.; Bustamante, L.; Haynes, R. K.; Staines, H. M. *Trends Pharmacol. Sci.* **2008,** *29,* 520–527. doi:10.1016/j.tips.2008.07.004
4. Kremsner, P. G.; Krishna, S. *Lancet* **2004,** *364,* 285–294. doi:10.1016/s0140-6736(04)16680-4
5. Bouwmeester, H. J.; Wallaart, T. E.; Janssen, M. H. A.; van Loo, B.; Jansen, B. J. M.; Posthumus, M. A.; Schmidt, C. O.; De Kraker, J.-W.; König, W. A.; Franssen, M. C. R. *Phytochemistry* **1999,** *52,* 843–854. doi:10.1016/s0031-9422(99)00206-x
6. Mercke, P.; Bengtsson, M.; Bouwmeester, H. J.; Posthumus, M. A.; Brodelius, P. E. *Arch. Biochem. Biophys.* **2000,** *381,* 173–180. doi:10.1006/abbi.2000.1962

7. Chang, Y.-J.; Song, S.-H.; Park, S.-H.; Kim, S.-U. *Arch. Biochem. Biophys.* **2000,** *383,* 178–184. doi:10.1006/abbi.2000.2061
8. Picaud, S.; Mercke, P.; He, X.; Sterner, O.; Brodelius, M.; Cane, D. E.; Brodelius, P. E. *Arch. Biochem. Biophys.* **2006,** *448,* 150–155. doi:10.1016/j.abb.2005.07.015
9. Kim, S.-H.; Heo, K.; Chang, Y.-J.; Park, S.-H.; Rhee, S.-K.; Kim, S.-U. *J. Nat. Prod.* **2006,** *69,* 758–762. doi:10.1021/np050356u
10. Touchet, S.; Chamberlain, K.; Woodcock, C. M.; Miller, D. J.; Birkett, M. A.; Pickett, J. A.; Allemann, R. K. *Chem. Commun.* **2015,** *51,* 7550–7553. doi:10.1039/c5cc01814e
11. Cascón, O.; Touchet, S.; Miller, D. J.; Gonzalez, V.; Faraldos, J. A.; Allemann, R. K. *Chem. Commun.* **2012,** *48,* 9702–9704. doi:10.1039/c2cc35542f
12. Faraldos, J. A.; Miller, D. J.; González, V.; Yoosuf-Aly, Z.; Cascón, O.; Li, A.; Allemann, R. K. *J. Am. Chem. Soc.* **2012,** *134,* 5900–5908. doi:10.1021/ja211820p
13. Miller, D. J.; Yu, F.; Allemann, R. K. *ChemBioChem* **2007,** *8,* 1819–1825. doi:10.1002/cbic.200700219
14. Faraldos, J. A.; Zhao, Y.; O'Maille, P. E.; Noel, J. P.; Coates, R. M. *ChemBioChem* **2007,** *8,* 1826–1833. doi:10.1002/cbic.200700398
15. Huynh, F.; Grundy, D. J.; Jenkins, R. L.; Miller, D. J.; Allemann, R. K. *ChemBioChem* **2018,** *19,* 1834–1838. doi:10.1002/cbic.201800218
16. Huynh, F.; Miller, D. J.; Allemann, R. K. *Methods Enzymol.* **2018,** *608,* 83–95. doi:10.1016/bs.mie.2018.05.004
17. Oberhauser, C.; Harms, V.; Seidel, K.; Schröder, B.; Ekramzadeh, K.; Beutel, S.; Winkler, S.; Lauterbach, L.; Dickschat, J. S.; Kirschning, A. *Angew. Chem., Int. Ed.* **2018,** *57,* 11802–11806. doi:10.1002/anie.201805526
18. Rising, K. A.; Crenshaw, C. M.; Koo, H. J.; Subramanian, T.; Chehade, K. A. H.; Starks, C.; Allen, K. D.; Andres, D. A.; Spielmann, H. P.; Noel, J. P.; Chappell, J. *ACS Chem. Biol.* **2015,** *10,* 1729–1736. doi:10.1021/acschembio.5b00145
19. Lauchli, R.; Rabe, K. S.; Kalbarczyk, K. Z.; Tata, A.; Heel, T.; Kitto, R. Z.; Arnold, F. H. *Angew. Chem., Int. Ed.* **2013,** *52,* 5571–5574. doi:10.1002/anie.201301362
20. Beale, M. H.; Birkett, M. A.; Bruce, T. J. A.; Chamberlain, K.; Field, L. M.; Huttly, A. K.; Martin, J. L.; Parker, R.; Phillips, A. L.; Pickett, J. A.; Prosser, I. M.; Shewry, P. R.; Smart, L. E.; Wadhams, L. J.; Woodcock, C. M.; Zhang, Y. *Proc. Natl. Acad. Sci. U. S. A.* **2006,** *103,* 10509–10513. doi:10.1073/pnas.0603998103
21. Yu, X.-D.; Pickett, J.; Ma, Y.-Z.; Bruce, T.; Napier, J.; Jones, H. D.; Xia, L.-Q. *J. Integr. Plant Biol.* **2012,** *54,* 282–299. doi:10.1111/j.1744-7909.2012.01107.x
22. Yu, X.; Zhang, Y.; Ma, Y.; Xu, Z.; Wang, G.; Xia, L. *The Crop J.* **2013,** *1,* 50–60. doi:10.1016/j.cj.2013.07.005
23. Bhatia, V.; Maisnam, J.; Jain, A.; Sharma, K. K.; Bhattacharya, R. *Ann. Bot. (Oxford, U. K.)* **2015,** *115,* 581–591. doi:10.1093/aob/mcu250
24. Briggs, G. G.; Cayley, G. R.; Dawson, G. W.; Griffiths, D. C.; Macaulay, E. D. M.; Pickett, J. A.; Pile, M. M.; Wadhams, L. J.; Woodcock, C. M. *Pestic. Sci.* **1986,** *17,* 441–448. doi:10.1002/ps.2780170415
25. Tyagi, A. K.; Prasad, S.; Yuan, W.; Li, S.; Aggarwal, B. B. *Invest. New Drugs* **2015,** *33,* 1175–1186. doi:10.1007/s10637-015-0296-5
26. Shin, Y.; Lee, Y. *Toxicol. Res. (Seoul, Repub. Korea)* **2013,** *29,* 257–261. doi:10.5487/tr.2013.29.4.257
27. Orellana-Paucar, A. M.; Serruys, A.-S. K.; Afrikanova, T.; Maes, J.; De Borggraeve, W.; Alen, J.; León-Tamariz, F.; Wilches-Arizábala, I. M.; Crawford, A. D.; de Witte, P. A. M.; Esguerra, C. V. *Epilepsy Behav.* **2012,** *24,* 14–22. doi:10.1016/j.yebeh.2012.02.020
28. Orellana-Paucar, A. M.; Afrikanova, T.; Thomas, J.; Aibuldinov, Y. K.; Dehaen, W.; de Witte, P. A. M.; Esguerra, C. V. *PLoS One* **2013,** *8,* e81634. doi:10.1371/journal.pone.0081634
29. Demiray, M.; Tang, X.; Wirth, T.; Faraldos, J. A.; Allemann, R. K. *Angew. Chem., Int. Ed.* **2017,** *56,* 4347–4350. doi:10.1002/anie.201609557
30. Wuts, P. G. M.; Greene, T. W. *Greene's Protective Groups in Organic Synthesis;* John Wiley & Sons, Inc.: Hoboken, NJ, U.S.A., 2006. doi:10.1002/0470053488
31. Rose, M. W.; Rose, N. D.; Boggs, J.; Lenevich, S.; Xu, J.; Barany, G.; Distefano, M. D. *J. Pept. Res.* **2005,** *65,* 529–537. doi:10.1111/j.1399-3011.2005.00261.x
32. Marco, J. A.; García-Pla, J.; Carda, M.; Murga, J.; Falomir, E.; Trigili, C.; Notararigo, S.; Díaz, J. F.; Barasoain, I. *Eur. J. Med. Chem.* **2011,** *46,* 1630–1637. doi:10.1016/j.ejmech.2011.02.011
33. Davisson, V. J.; Woodside, A. B.; Neal, T. R.; Stremler, K. E.; Muehlbacher, M.; Poulter, C. D. *J. Org. Chem.* **1986,** *51,* 4768–4779. doi:10.1021/jo00375a005
34. Huang, Z.; Poulter, C. D. *Org. Synth.* **1988,** *66,* 211–215. doi:10.15227/orgsyn.066.0211
35. Keller, R. K.; Thompson, R. *J. Chromatogr.* **1993,** *645,* 161–167. doi:10.1016/0021-9673(93)80630-q
36. Bohlmann, F.; Singh, P.; Jakupovic, J. *Phytochemistry* **1982,** *21,* 2531–2535. doi:10.1016/0031-9422(82)85251-5
37. Arkoudis, E.; Stratakis, M. *J. Org. Chem.* **2008,** *73,* 4484–4490. doi:10.1021/jo800355y
38. Tori, M.; Aoki, M.; Asakawa, Y. *Phytochemistry* **1994,** *36,* 73–76. doi:10.1016/s0031-9422(00)97015-8
39. Kreiser, W.; Körner, F. *Helv. Chim. Acta* **1999,** *82,* 1427–1433. doi:10.1002/(sici)1522-2675(19990908)82:9<1427::aid-hlca1427>3.0.co;2-s
40. Sy, L.-K.; Brown, G. D. *Magn. Reson. Chem.* **1997,** *35,* 424–425. doi:10.1002/(sici)1097-458x(199706)35:6<424::aid-omr84>3.0.co;2-c
41. Togar, B.; Turkez, H.; Tatar, A.; Hacimuftuoglu, A.; Geyikoglu, F. *Cytotechnology* **2015,** *67,* 939–946. doi:10.1007/s10616-014-9729-9

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Isolation of fungi using the diffusion chamber device FIND technology

Benjamin Libor, Henrik Harms, Stefan Kehraus, Ekaterina Egereva, Max Crüsemann and Gabriele M. König*

Full Research Paper

Open Access

Address:
Institute for Pharmaceutical Biology, University of Bonn, Nussallee 6, 53115 Bonn, Germany

Email:
Gabriele M. König* - g.koenig@uni-bonn.de

* Corresponding author

Keywords:
FIND; fungal one-step isolation device; *Heydenia* cf. *alpina*; natural products; terpenes

Beilstein J. Org. Chem. **2019,** *15,* 2191–2203.
doi:10.3762/bjoc.15.216

Received: 28 June 2019
Accepted: 13 August 2019
Published: 19 September 2019

This article is part of the thematic issue "Terpenes".

Guest Editor: J. S. Dickschat

Abstract

Fungi are an important source of bioactive metabolites. The Fungal one-step IsolatioN Device (FIND) technology allows the isolation of rare fungi from terrestrial and marine samples. The FIND comprises a multi-chambered micro agar plate, where initially only one fungal part (e.g., hyphal cell, mycelial fragment or spore) is located in each chamber. After inoculation the device is placed back into the original natural environment of sample collection, to ensure favourable growth conditions. Experiments were carried out with terrestrial soil and marine sediment, as well as sea water samples to validate this method. This yielded axenic cultures of 12 different filamentous fungi, one of them being the marine-adapted fungal strain *Heydenia* cf. *alpina*. The latter produced two new terpenoids, which are the first secondary metabolites from this genus.

Introduction

Natural products play a prominent role as lead structures in drug discovery [1]. Apart from bacteria, fungi are an impressive group of microorganisms in this respect, due to the high structural diversity of their rich secondary metabolism. Fungi are known as producers of many therapeutically important drugs, such as antibiotics (e.g., cephalosporins), immunosuppressants (e.g., mycophenolic acid) and immunomodulators (e.g., fingolimod), as well as cholesterol lowering agents (statins). Generally, fungal metabolites were shown to have antiviral, cytotoxic, antineoplastic, cardiovascular, anti-inflammatory, immune-stimulating, and anticancer activities [2]. Since the World Health Organization (WHO) declared antibiotic resistance a global threat and urged the search for new antibiotics, the interest in fungal metabolites increased recently.

In order to find new structural types, novel source organisms have to be targeted. It is however, most difficult to isolate unusual, to date not yet researched microbes in axenic form and to cultivate them. Overall it was estimated that only 1% of all microbes present in our environments are cultivated today

("great plate count anomaly") [3]. Recently the new bacterial species *Eleftheria terrae*, alongside other new microbial species, was isolated from an environmental sample with a device known as iChip. This technology is based on the incubation of a single bacterial cell in its natural environment. After this incubation, strains can be transferred as axenic cultures to the laboratory [4]. *E. terrae* produced a novel antibiotic called teixobactin with no detectable resistance, as it interacts with key functions of cell wall synthesis that do not undergo fast evolutionary changes [5].

Inspired by this approach we developed the Fungal one-step IsolatioN Device (FIND), in order to obtain axenic cultures of rare and less examined filamentous fungi, as several tries before with the iChip were not successful. Apart from terrestrial soil probes we focused on the marine habitat, probing sediment and sea water samples for the presence of fungi. With 150 000 species named and classified, it was suggested that 95% of fungi to date remain unknown [6]. The so-called dark matter fungi (DMF) are a promising source for novel compounds [7-10] and their isolation and culture is one of the keys to find new pharmacological agents.

Results

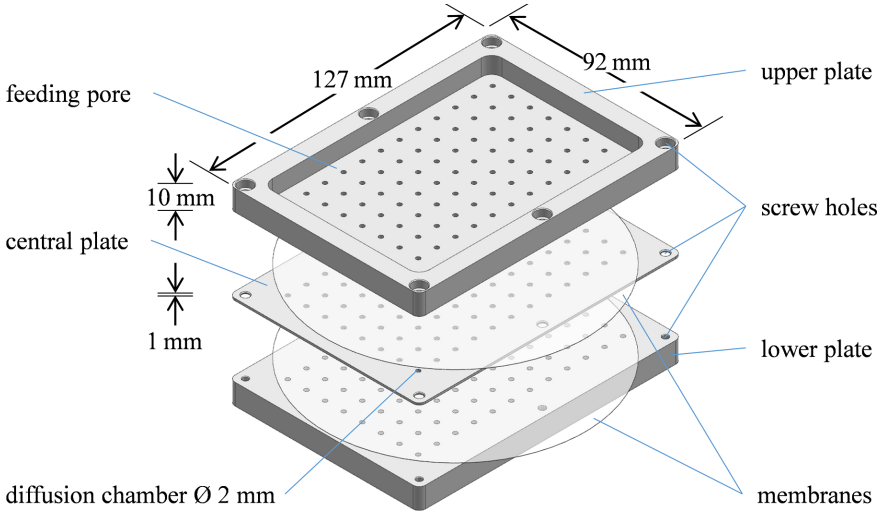
Isolation of axenic fungal strains using FIND

For the isolation of fungal strains from terrestrial and marine environments a device was constructed, i.e., FIND that is similar to the one described by Epstein et al. for "unculturable" bacteria [4]. For the FIND we adjusted the isolation procedure and technical features, e.g., the dimensions of the through holes utilized as growing chambers, to fungi. The FIND protocol now suits fungal physiology and the prevalence of fungi in different habitats.

In general, the FIND technology is a multichambered micro agar plate. After inoculation, the device is placed back into the original natural environment of sample collection, to ensure favourable growth conditions for specific fungi. Inoculation is performed in only one step and thus suitable for high throughput approaches. Scheme 1 illustrates the FIND, consisting of merely three plastic parts crafted from polyoxymethylene, which are solid and durable enough for use in most natural environments. The central plate is a thin plastic board with 96 through holes that function as compartments, capable of holding inoculated agar plugs. Semipermeable membranes are attached to both sides of this plate, this way covering the through holes and preventing contamination by other microbes. These membranes, at the same time, permit diffusion of nutrients and growth factors from the environment through the feeding pores of the upper and lower plates. This way slow growing fungi or to date uncultured fungi can be isolated.

To test the FIND technology, we used samples from terrestrial and marine environments, i.e., soil from Wuppertal, Germany (experiment 1), marine sediments from the Aegean Sea, Greece (experiment 2), brackish soil samples from the Schlei region around Kappeln, Germany (experiment 3), and sea water samples from Katwijk aan Zee, Netherlands (experiment 4).

Adjustment of the sample concentration (i.e., number of fungal parts per volume), which is crucial for the success of a FIND experiment, has to take the abundance of fungal parts, i.e., hyphae, mycelial fragments or spores in the collected environmental samples into account. Ideally, for incubation one fungal part per through hole should be present. Thus, for the here presented FIND with 96 through holes each with a volume of 3×10^{-3} mL, a concentration of 320 000 fungal parts/L must be

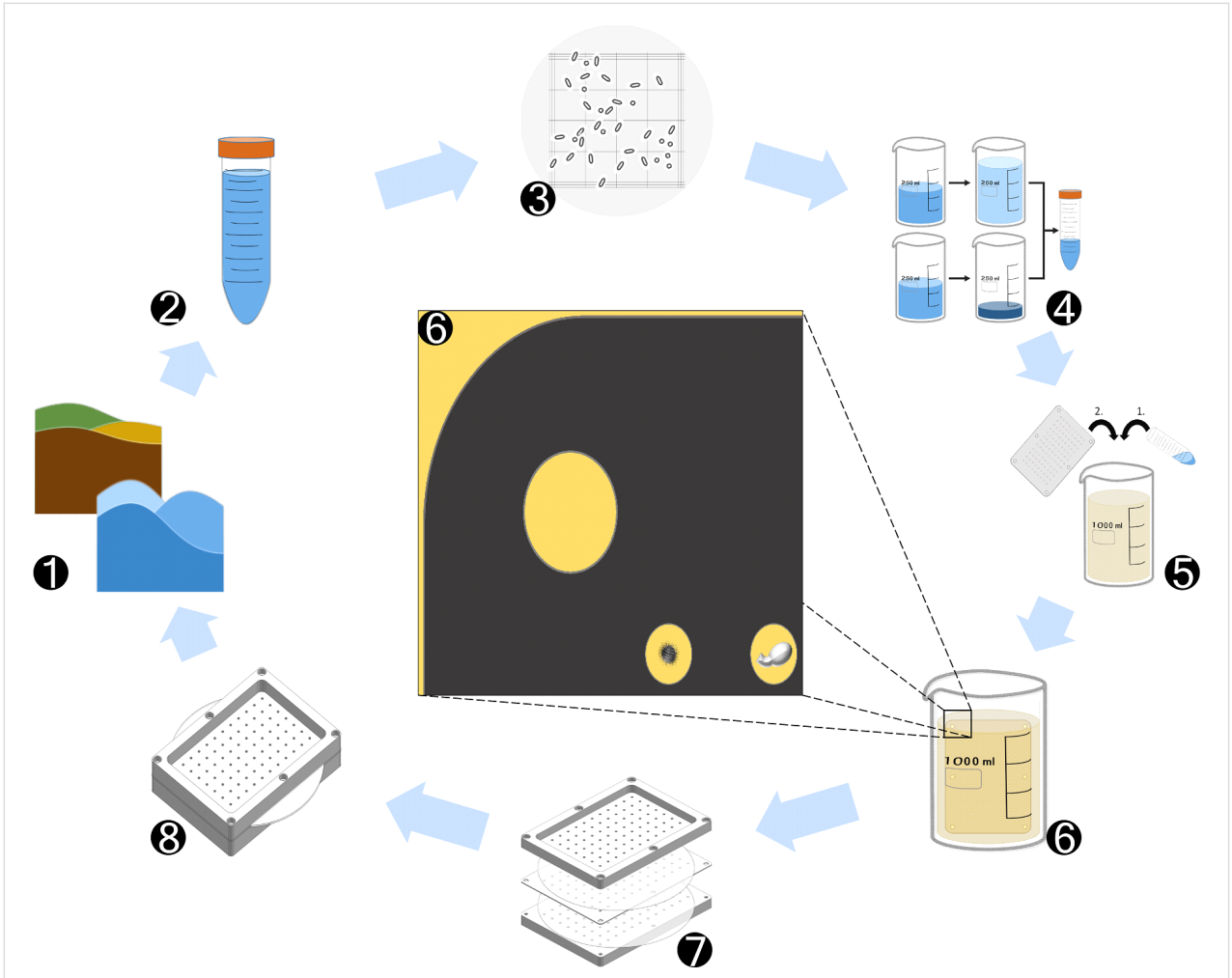


Scheme 1: Design and functional parts of the FIND technology.

reached in the agar solution, with which the central plate is loaded. The number of fungi in a given volume of a sample is experimentally hard to determine. We thus counted bacterial cells, and applied the ratio of bacteria-to-fungi, which is described in the literature for most environments, to estimate suitable dilution or concentration steps in our experiments. Scheme 2 illustrates each individual step of a FIND experiment.

For experiment 1 (terrestrial soil sample) a small volume of soil was treated with a sterile-filtered isotonic NaCl solution [11-13]. After centrifugation cell counting was performed on the supernatant using a Neubauer Improved or Thoma chamber [14] to determine the abundance of bacterial cells. For our soil sample $2 \pm 0.25 \times 10^{13}$ bacteria/L were determined. For experiment 2 (marine sediment) and 3 (brackish water sediment), samples were treated alike, but instead of a NaCl solution, sea water from the respective sampling site was used. For experiment 4 (sea water), sea water was analysed directly. The extant abundances of bacterial cells/L in experiments 2, 3 and 4 were found to be $7 \pm 0.7 \times 10^9$, $13 \pm 5 \times 10^9$ and $8 \pm 0.8 \times 10^9$ bacterial cells/L, respectively.

The samples were then diluted, concentrated or used undiluted according to the bacteria-to-fungi ratio, in order to ultimately reach an approximate target concentration of 320 000 fungal parts/L. Literature documented a bacteria-to-fungi ratio as 10^3 to 1 for terrestrial soil [15]. Thus, the sample from Wuppertal (experiment 1) was diluted sequentially to a concentration of 2×10^{10} bacterial cells/L, which should be equivalent to 2×10^7 fungal parts/L, and 16 mL of this solution were necessary to be added to one litre of agar solution to obtain the required target concentration.



Scheme 2: Isolation of fungal strains with the FIND technology. 1. Collection of terrestrial or marine sample. 2. Sample preparations. 3. Determination of the number of microbial cells. 4/5. Dilution or concentration of sample and merging with molten agar. 6. Adjustment of the concentration of fungal parts in such a way that in average one fungal part is finally placed in one growing chamber. 7. Attachment of semipermeable membranes to the central plate and sealing with upper and lower plates. 8. Placement of the loaded and assembled device into the original environment for incubation.

For marine sediment samples (experiments 2 and 3), in general, the bacteria-to-fungi ratio was hardly known. Thus, we applied the ratio described for terrestrial soil samples, i.e., 10^3 to 1. With abundances of 7×10^9 and 13×10^9 bacterial cells/L concentrations of 7×10^6 and 13×10^6 fungal parts/L, respectively, were assumed. The samples were thus used undiluted and 42 and 24 mL, respectively, were added to one litre of agar solution to obtain the required target concentration. Using this ratio, in our experiments, however, we had hardly any success (apart from one fungal isolate, i.e., *Cladosporium allicinum* and *Heydenia* cf. *alpina* from samples 2 and 3, respectively). Obviously in these experiments the concentration of fungal parts was too low, and thus we addressed this question in experiment 4 more carefully.

Abundances of around 1000 fungi/L sea water were documented for samples from the North Sea [16,17]. Considering this information, we performed experiment 4 on the assumption of an abundance of 1000 fungi/L, which required fungal cells of 8 L of sea water in a volume of 25 mL agar solution in step 5 of the FIND procedure. Since our bacterial count on this unprocessed sea water sample had revealed an abundance of 8×10^9 cells per L, we ourselves calculated a bacteria-to-fungi ratio of around 10^7 to 1 for subsurface samples from littoral areas of the North Sea. This is in accordance to literature reports for surface slick and subsurface samples from littoral areas of Florida, where ratios of 10^5 to 1 and higher were determined [18].

Vacuum filtration of the marine water sample (10 L) from Katwijk aan Zee/North Sea through paper disks with a pore size of approx. 2 µm, i.e., wide enough to let pass bacteria, but small enough to concentrate fungal spores and mycelia in the filter cake, was performed. Around 5 mL of filter cake was resuspended in another 5 mL of sterile sea water and mixed with 15 mL agar solution.

In summary, the diluted sample from experiment 1, the undiluted solutions from experiments 2 and 3, and the concentrated, resuspended sea water solution from experiment 4 were each mixed with a warm agar preparation, containing a minimal nutrient medium and antibiotics (experiment 1: amino acids/peptone, antibiotics; experiments 2–4: ditto, but in sea water) to yield the desired concentration of roughly 320 000 fungal parts/L. Then, the central plate of FIND (Scheme 1) was lowered into these preparations to fill each chamber with an agar solution, containing in average one fungal part. As the agar solidified, the fungal spore or mycelium was trapped in the chamber and separated from other ones, e.g., competing organisms. Semipermeable membranes were attached to both sides of the central plate and the device was sealed with the upper and lower parts and fixed with screws. To guarantee that the device was sealed properly, commercial aquarium safe silicone without antifungal additive was applied to the edges. All materials and tools were sterilized by autoclaving or surface disinfection and the procedure was performed under the laminar airflow cabinet.

The sample-loaded and sealed device was then inserted into an appropriate container with 40 L of the original soil for experiment 1, sea water and sediment from the collection sites for experiments 2 and 3, whereas for experiment 4 (sea water) 20 L of sea water from the sample site were used to mimic environmental conditions of the corresponding sample sites.

At this stage the trapped organisms are separated from each other, but connected to their original environment via semipermeable membranes. This way the fungal seeds are provided with growth factors exclusively found in their natural habitat. The device was incubated for four weeks. Deionized water was applied to the soil of experiment 1 every 2 days. An aquarium water pump provided water current simulation and evaporated tank water was replaced by autoclaved deionized water in experiments 2–4.

After incubation the device was removed from the tanks, rinsed with autoclaved deionized water to remove sediment and soil and opened under a laminar air flow cabinet. Each agar plug was transferred to a corresponding agar plate for cultivation. The agar plates contained the same minimal nutrient medium as the agar solution into which the central plate was dipped earlier. Alternatively, a stamp can be used to punch all agar plugs simultaneously into the wells of a 96 well plate. Fermentation on agar plates was performed in an incubator at 25 °C for another four weeks. Morphology of the axenic colonies was then analysed microscopically, and their taxonomy was additionally identified by DNA sequencing.

In total, experiments 1–4 led to 76 fungal isolates (41 fungi in experiment 1, 1 fungus in experiment 2, 1 fungus in experiment 3 and 33 fungi in experiment 4). All fungal cultures were obtained in single isolation steps directly via the FIND procedure and found to be axenic. From experiment 1, 39 isolates were microscopically identified as belonging to genera common in soil, e.g., *Fusarium*, *Penicillium* or *Trichoderma* and thus were not further processed. The addition of antibiotics prevented bacterial contamination in all marine samples. However, a few antibiotic resistant strains were isolated from soil but also not further processed. One isolate from experiment 4 was found to belong to the genus *Cladosporium*, and therefore was also not considered further. The remaining 36 isolates were identified at least to the genus level (Table 1). The pooled results from com-

Table 1: Fungal taxa obtained by four FIND experiments and identified by sequence comparison with the best BLASTn match within the NCBI GenBank database.

Species/genus [GenBank accession number]	ID%	Number of bp analysed	Taxonomic class
Clonostachys rosea (Link:Fries) Schroers et al.[1] [KT323182]	99[a]	506	Sordariomycetes
Ilyonectria europaea A. Cabral, Rego & Crous[1] [JF735294]	100[b]	531	Sordariomycetes
Cladosporium allicinum (Fries) Bensch, U. Braun & Crous[2] [MH118272; MH567104]	100[c]; 100[d]	511; 241	Dothideomycetes
Heydenia cf. *alpina*[3] [KF574887]	100[a]	549	incertae sedis
Alternaria armoraciae E.G. Simmons & G.F. Hill[4] [KC584638]	100[d]	236	Dothideomycetes
Auxarthron cfr. *umbrinum* (Boudin) G.F. Orr & Plunkett[4] [MH857026]	99[b]	509	Eurotiomycetes
Cadophora sp.[4] [FJ820724; KF636777]	99[b]; 99[e]	552; 999	Leotiomycetes
Chaetomium globosum aggr.[4] [MH644079; MG890100]	100[f]; 100[g]	380; 585	Sordariomycetes
Chrysosporium sp.[4] [MF939600]	99[h]	538	Eurotiomycetes
Leucothecium sp.[4] [LT608439]	100[i]	200	Sordariomycetes
Metarhizium carneum (Duché & R. Heim) Kepler, S.A. Rehner & Humber[4] [MH864783]	100[b]	624	Sordariomycetes
Scopulariopsis brevicaulis Bainier[4] [LM652465]	100[b]	609	Sordariomycetes

[1]Eskesberg, Wuppertal (DE), soil sample, [2]Aegean Sea (GR), marine sediment, [3]Kappeln (DE), brackish water sediment, [4]Katwijk aan Zee (NL), sea water
[a]18S rRNA gene, ITS1, 5.8S rRNA gene, ITS2, [b]18S rRNA gene, ITS1, 5.8S rRNA gene, ITS2, 28S rRNA gene, [c]ITS1, 5.8S rRNA gene, ITS2, [d]tef1α, [e]28S rRNA, [f]β-tubulin, [g]RPBII, [h]ITS1, 5.8S rRNA gene, ITS2 and 28S rRNA gene, [i]ITS2

parison (BLAST) of DNA sequences (ITS1, ITS2, 18S, 5.8S, 28S, RPBII, tef1α and β-tubulin) with reference sequences deposited in public databases (GenBank) and microscopical analysis of morphological characteristics, identified 12 ascomycetes from 12 different genera. Except for *Scopulariopsis brevicaulis* (25 isolates), all strains were encountered only once, due to the ideally adjusted microbial concentration and homogenisation of the sample material. Nine of those 12 different isolates could be identified to species level, whereas for the remaining three (*Cadophora* sp., *Chrysosporium* sp., *Leucothecium* sp.) no species could be assigned. In the case of fungi of the genus *Cadophora*, genetic information for all 25 hitherto described species (MycoBank) are deposited in GenBank. As the alignments of our sequences with the best matching ones of a cultured species, e.g., *C. orchidicola* reached only 97% identity in BLAST, it can be assumed that our isolate is an undescribed species, as it shows 99% identity with sequences of an undescribed species (Table 1). For the 116 legitimate species of *Chrysosporium*, BLAST of our sequences revealed 99% identity with reference sequences of *Ch. submersum*. Unfortunately, sequences for only 33 *Chrysosporium* species are available as reference sequences, and thus no identification on the species level could be achieved. Interestingly, the sequence of our isolated *Leucothecium* species is 100% identical to the reference sequence of an uncultured fungus isolated from cotton in 2014 and seems to be closely related to the type species *L. emdenii,* which is one of only two known species among *Leucothecium*. Most of the isolates obtained in our experiments belong to the Sordariomycetes, the by far biggest class of ascomycetes, and to the Dothideomycetes. Some of our cultures, however, are from less examined classes, e.g., Eurotiomycetes and Leotiomycetes. Only *Chaetomium globosum* aggr. comes into consideration as a common air mould.

In a next series of experiments, we set out to shed light on the halotolerance of our isolates from experiments 2 and 3 by testing their salt tolerance or dependency. Thus, we analysed the growth behaviour of *Cladosporium allicinum* and *Heydenia* cf. *alpina* on agar plates containing growth medium of increasing salinity from 0–35‰ (sea water has a salinity of approx. 35‰). As outcome the diameter of radial growth was measured after 14 days. All experiments were performed in triplicate on 130 mm agar plates containing bio malt agar medium. For the isolate from the Mediterranean Sea (experiment 2), i.e., *C. allicinum*, our results showed optimal growth conditions at 35‰ salinity, i.e., a diameter of 107 ± 2 mm was reached. Only poor growth of 54 ± 1 to 89 ± 1 mm was observed at levels of salinity of 14‰ and below. For the isolate from brackish water (experiment 3), i.e., *Heydenia* cf. *alpina* levels of salinity above 7‰ were necessary to achieve an adequate growth of 124 ± 1 to 130 ± 0 mm. Below 7‰ salinity the radial growth rates of latter cultures dropped dramatically to only 67 ± 1 mm in diameter. In both cases growth rates were nearly halved when the growth medium was lacking appropriate concentrations of sea salts. Our results show that these fungi require high salt concentrations for optimal growth, and this suggests an adaption of the respective strains to marine conditions (see Supporting Information File 1, Tables S1 and S2 for full experimental data).

With the FIND technology we managed to isolate rare fungi. To evaluate the potential for the biosynthesis of bioactive secondary metabolites we performed screenings on antimicrobial activities against Gram-negative and Gram-positive bacteria, an ascomycete, a basidiomycete and a zygomycete, i.e., *Escherichia coli*, *Bacillus megaterium, Eurotium rubrum*, *Microbotryum violaceum* and *Mycotypha microspora*. Ethyl acetate extracts of *C. rosea*, *I. europaea* grown on BM medium and of the other 10 isolates grown on BMS medium were tested in disk diffusion assays using 50 µg of extract per disk. None of the extracts was active against the Gram-negative bacterium *E. coli*. Extracts of *A. armoraciae, Auxarthron umbrinum, C. globosum, Chrysosporium* sp. 831*, C. allicinum, C. rosea, H.* cf. *alpina* and *I. europaea* inhibited the growth of *B. megaterium*, i.e., inhibition zones of 1, 2, 1.5, 1, 4, 5, 5 and 3 mm were reached, respectively (streptomycin, equally concentrated, was used as positive control with an inhibition zone of 10 mm). Additionally, the extract of *I. europaea* inhibited the growth of *M. violaceum* with an inhibition zone of 3 mm in diameter (miconazole, half concentrated, was used as positive control with an inhibition zone of 10 mm; see Supporting Information File 1, Table S3 for full experimental data).

Detailed chemical investigation of *H.* cf. *alpina* including structure elucidation of the new metabolites **1** and **2**

On the basis of its preferred growth on saline media, the *H.* cf. *alpina* (strain No. 824), isolated during this study, was selected for further chemical investigations. To the best of our knowledge this is the first chemical investigation of a member of the *Heydenia* genus.

From the solid, sea-salt containing cultures of *H.* cf. *alpina* the ethyl acetate soluble organic compounds were analysed. A first fractionation was achieved by vacuum liquid chromatography (VLC) on reversed-phase material yielding 4 fractions. ^{1}H NMR analysis of these indicated the presence of chemically diverse secondary metabolites in the major VLC fraction 3. Detailed HRESIMS investigation was thus performed with VLC fraction 3, which showed prominent *m/z* values for metabolites with molecular weights of 248 and 250 Da.

Subsequent repeated fractionation of VLC fraction 3 by RP-HPLC resulted in the isolation of two, structurally closely related, new secondary metabolites, **1** and **2** (Figure 1).

Compound **1** was obtained as a yellow oil with a molecular formula of $C_{15}H_{20}O_3$, implying 6 degrees of unsaturation (DOU), as deduced by HRESIMS (*m/z* 249.1481 $[M + H]^+$; calcd. for $C_{15}H_{21}O_3$, 249.1485). The IR spectrum displayed an absorption band at 1670 cm^{-1} indicating the presence of an α,β-unsaturated carbonyl moiety [19].

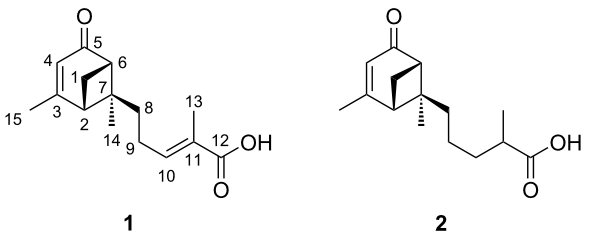


Figure 1: Secondary metabolites isolated from *H.* cf. *alpina*.

The ^{1}H NMR spectrum (Table S4, Supporting Information File 1) exhibited characteristic singlets for three methyl groups, as well as two olefinic protons, together with the MS data pointing towards a terpenoid like structure. The ^{13}C NMR and DEPT-135 spectra of **1** (Table S4, Supporting Information File 1) displayed 15 carbon signals for three methyl, three methylene, two sp^3 methine and two sp^3 quaternary carbons. Four characteristic shifts in the ^{13}C NMR spectrum at δ_C 174.1 (C-3), 122.1 (C-4), 143.0 (C-10) and 129.5 (C-11) evidenced two carbon–carbon double bonds. Together with a carbonyl carbon at δ_C 206.6 and a carboxylic carbon at δ_C 177.8, the two remaining DOU were accounted to a bicyclic ring system.

After assigning the proton resonances to the ^{13}C resonances of directly attached carbon atoms with an HSQC experiment, analysis of the ^{1}H,^{1}H-COSY data led to the identification of two major ^{1}H–^{1}H spin systems, one corresponding to the H-2–H-1$_b$–H-6 moiety (spin system A) and the second to the H$_2$-8–H$_3$-13 subunit (spin system B; Figure 2).

Spin system B was extended by a carboxylic group as indicated by ^{1}H,^{13}C-HMBC correlations from both H-10 (δ_H 6.86) and H$_3$-13 (δ_H 1.89) to carbon C-12 (δ_C 177.8), and identified the partial structure 2-methylpentenoic acid. HMBC correlations between the resonances of H-6 (δ_H 2.73) to both C-4 (δ_C 122.1) and C-5 (δ_C 206.6), between H$_3$-15 (δ_H 2.10) and C-2 (δ_C 49.8), C-3 (δ_C 174.1) and C-4 (δ_C 122.1), and between H$_2$-1 (δ_H 2.12; 2.93) and both C-3 (δ_C 174.1) and C-5 (δ_C 206.6) placed the methyl group CH$_3$-15 at carbon C-3 (δ_C 174.1), and proved the positions of the carbonyl group C-5 (δ_C 206.6) and the double bond $\Delta^{3,4}$ in the molecule, as shown in Figure 2. The second partial structure was thus a cyclohexenone ring. The proton resonances of H-2 (δ_H 2.65) and H-6 (δ_H 2.73) appear as a characteristic triplet of doublets caused by a vicinal coupling with proton H-1$_b$ (δ_H 2.93; J = 6.0 Hz), an unusual large "W" coupling to each other (J = 6.0 Hz) as well as an allylic coupling to H-4 (δ_H 5.78; J = 1.3 Hz). These couplings can only be ex-

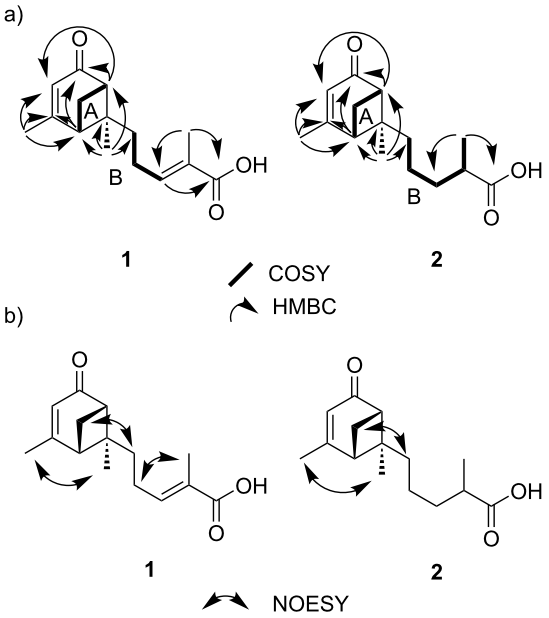


Figure 2: a) Significant $^{1}H,^{1}H$-COSY and $^{1}H,^{13}C$-HMBC correlations for compounds **1** and **2**. b) Key NOESY correlations for compounds **1** and **2**.

plained with the rigid structure of **1** supporting the presence of a bicyclo[3.1.1] ring system resembling that of verbenone [20].

HMBC correlations of H_3-14 (δ_H 1.07) to C-2 (δ_C 49.8), C-6 (δ_C 57.2), C-7 (δ_C 58.6) and C-8 (δ_C 38.2) confirmed the presence of a verbenone moiety and connected the methylpentenoic acid moiety to the quaternary carbon C-7, completing the planar structure of **1**. Compound **1** is thus a bergamotene type sesquiterpene, containing a bicyclo[3.1.1] ring system as well as a 2-methylpent-2-enoic acid moiety, accounting for all ring double bond equivalents (Figure 2).

The relative configuration of **1** at C-2, C-6 and C-7 was established by a NOESY spectrum. NOESY correlations between H-1_b (δ_H 2.93) and H_2-8 (δ_H 2.07) indicated that these protons were on the same side of the molecule and together with the NOESY correlation between H_3-14 and H_3-15 the relative configurations of C-2, C-6 and C-7 could thus be delineated as 2*S**, 6*S** and 7*S**, respectively.

The absolute configurations of C-2, C-6 and C-7 were determined by comparison of our experimental with simulated circular dichroism (CD) spectra. The latter ones were performed by Wu et al. using (2*S*,6*S*)-verbenone and (2*R*,6*R*)-verbenone and can be applied here, because of the assumed minor effect of the side chain on the CD spectrum [21]. Experimental data for **1** showed a negative Cotton effect in the range of 320–350 nm (321 nm), which suggested according to the calculated CD spectrum of Wu et al. (321 nm) that the absolute configurations of C-2, C-6 and C-7 in **1** are 2*S*, 6*S*, and 7*S*, respectively (see Supporting Information File 1, Figure S2 for the CD spectrum of **1**).

The configuration of the double bond $\Delta^{10,11}$ could be determined as *E* by NOESY correlations between proton signals H_2-9 (δ_H 2.32) and H_3-13 (δ_H 1.89). This was confirmed by the strong shielding in the ^{13}C NMR spectrum of carbon C-13 typically occurring in double bonds with *E* configuration [22]. Therefore, the structure was established as (*E*)-2-methyl-5-((1*S*,5*S*,6*S*)-2,6-dimethyl-4-oxobicyclo[3.1.1]hept-2-en-6-yl)pent-2-enoic acid, for which the trivial name heydenoic acid A is suggested. Notable is the occurrence of peaks containing *m/z* values attributable to isomers of the isolated metabolite.

Compound **2** has a molecular formula of $C_{15}H_{22}O_3$, (HRESIMS *m/z*: 251.1641 $[M + H]^+$; calcd. for $C_{15}H_{23}O_3$, 251.1642) implying only 5 DOU as compared to **1**.

^{1}H and ^{13}C NMR data (Table S5, Supporting Information File 1) of **2** were similar to those of **1**, except for the chemical shifts of CH_2-9 (δ_H 1.37; δ_C 23.4), CH_2-10 (δ_H 1.49, 1.73; δ_C 35.5), and CH_3-13 (δ_H 1.21; δ_C 17.8). Additional proton signals for H-11 and H_2-10 (δ_H 2.50; H-11, δ_H 1.49 and δ_H 1.73; H_2-10) proved that **2** is a derivative of **1**, in which $\Delta^{10,11}$ is saturated. COSY and HMBC data of **2** fully supported this assumption. NOESY correlation between H_b-1 (δ_H 2.89) and H_2-8 (δ_H 1.94) indicated the relative configurations of C-2, C-6 and C-7 to be the same as in **1**. The CD spectrum of **2** showed also a negative Cotton effect as **1**, indicating the absolute configurations of C-2, C-6 and C-7 to be all *S* (see Supporting Information File 1, Figure S3 for the CD spectrum of **2**). The configuration at C-11 could not be determined. Thus, the structure was established as 2-methyl-5-((1*S*,5*S*,6*S*)-2,6-dimethyl-4-oxobicyclo[3.1.1]hept-2-en-6-yl)pentanoic acid, named as heydenoic acid B.

Compounds **1** and **2** were also tested in antimicrobial assays and showed no effect.

Discussion

Our experiments demonstrated the potential of the FIND technology for the isolation of fungi from various habitats including terrestrial, littoral and marine areas. We developed an isolation method that is straight forward to use, rapidly performed and provides interesting isolates in a single experiment. In contrast to other isolation methods, e.g., direct plating or dilution plating, where the former can be experimentally difficult and the latter needs several days for colony counts, the FIND procedure is less labour intensive and the count can be per-

formed in a few minutes. A strain of *A. umbrinum* was isolated by Alvi et al. using dilution plating on a soil sample [23]. The sample was therefore diluted and spread over the plates and incubated for colony growth. To obtain plates for colony picking several dilutions have to be prepared and inoculated plates have to be incubated only to use a fraction of them for further processing. More challenging is the contamination by fast growing fungi that additionally suppress growth of other species on the same plate. The advantage of FIND lies in the adjustment of the correct dilution before incubation on the one hand and the spatial separation which allows slow growing fungi to form axenic colonies in a single step on the other hand. In times of 3D printers, the device may also be constructed in such a manner. In conventional methods we often experience a mix of species in one colony, which is hard or impossible to separate. In FIND, however, all collected colonies were found to be axenic and ready to be cultivated for further purposes like taxonomic identification or natural product isolation. Isolation methods like baiting may lead to bait-specific fungi only. A strain of *A. umbrinum* was isolated by Deshmukh et al. using a hair baiting technique. The majority of those isolates showed keratinolytic activity which proves the bias towards bait-specific fungi [24]. In FIND, fungi regardless of nutritional preferences can be obtained, as they are cultured within the respective environment. In this respect, the most striking feature of the FIND technology is the possibility to reconnect the isolates with their environment, in order to exchange nutrients and growth factors and therefore to ensure species-specific parameters.

An important feature to know before inoculating the FIND is the concentration of fungal parts in the collected environmental sample. Our experiments and literature data showed that the abundances of fungi or fungal parts, i.e., hyphal cells, mycelial fragments or spores, in terrestrial soil samples (2×10^9 fungal parts/L) is huge compared to those in marine sediments (5.000–25.000 fungal parts/L). Only extremely small numbers of fungal parts can be found in sea water (0.3–1000 fungal parts/L) [15-17,25]. These findings together with the high bacteria-to-fungi ratios described in the literature [18] explain the low recovery of fungal strains of experiments 2 and 3. To improve the outcome of experiments performed on sea sediment samples we propose using tenfold of the reported sediment and corresponding sea water in step 2 of the FIND procedure. Additionally, it is crucial to apply the herein described concentration protocol (step 4) and to perform the experiment in a volume as small as possible, i.e., 25 mL to achieve the target concentration (step 5).

Using FIND we isolated and identified 12 ascomycetes belonging to 4 different taxonomic classes (Table 1) from 3 different habitats. For half of these species, e.g., *A. armoraciae, C. allicinum, H.* cf. *alpina, I. europaea* and *M. carneum* to date no secondary metabolites are described in the literature. *A. armoraciae and C. allicinum* are special within their genus, as both belong to chemically underexamined clades, e.g., *Chalastospora* clade and *Cladosporium herbarum* complex, respectively [26,27]. For *C. allicinum* our experiments point towards a salt dependency of growth, i.e., an adaption to the marine environment. Halotolerance was to date exclusively reported for members of the *Cladosporium sphaerospermum* complex [28].

Analysis of genes involved in secondary metabolite production suggested *Cadophora malorum* as a producer of bioactive compounds [29], as later confirmed by the results of Almeida et al. [30], which highlights FINDs potential in the isolation of bioactive compound producing strains. Using FIND enabled the isolation of rare species of the genus *Chrysosporium*, which are scarcely found in soil [31] and of which the world register of marine species only lists four entries, e.g., *C. corda*, *C. pannorum*, *C. magnasporum* and *C. oceanitesii* [32]. With *Leucothecium* sp. a rare as well as fastidious fungus was isolated in this study, where cultivation was reported to be difficult for at least one of the two species belonging to this genus. *L. coprophilum* showed poor growth when cultured under laboratory conditions as it was lacking interactions to certain bacteria coexisting in its natural habitat [33]. In contrast to this, there were no difficulties regarding isolation or cultivation using the FIND technology, thus proving the benefits of the device, i.e., isolation of rare fungal strains that are challenging to cultivate in the laboratory as crucial growth factors are missing in artificial medium.

For only four species among the 12 isolates (*A. umbrinum, C. globosum, C. rosea, S. brevicaulis)* secondary metabolites were described in the literature. In the case of *C. globosum* these findings were limited to mycotoxins [34] where in the other cases the isolated structures showed remarkable activities in the field of anticancer and antimicrobial agents [35-40].

H. alpina is one of only six species known from the less examined genus *Heydenia,* and was mistaken for a member of the genus *Onygena* by Fischer because of its morphological similarities [41,42]. Molecular data now allow to distinguish both genera from each other [43]. Strains of *H. alpina* were isolated from areas with very harsh conditions, e.g., Antarctic lakes and naked silican rock [43,44]. On the basis of its extraordinary growth behaviour, the *H.* cf. *alpina* (strain 824) isolated during this study was selected for further chemical investigations. This led to the isolation of two novel compounds (**1**, **2**). Compounds **1** and **2** belong to the bergamotene type of sesquiterpenoids, for which manifold ecological functions were reported [45]. The bi-

ological role of the closely related (+)-(*E*)-*endo*-β-bergamoten-12-oic acid in *Lycopersicon hirsutum* was well investigated. Present in the trichomes of the leaves, it serves as an oviposition stimulant to the moth *Heliothis zea* in LA 1777 [46]. On the other hand, (+)-(*E*)-*endo*-β-bergamoten-12-oic acid showed repellent activity against arthropod mites. In fact, *L. hirsutum* is highly resistant to arthropod herbivores [47]. The bergamotene derivatives clavigerins, first isolated from the liverwort *Lepidolaena clavigera* also showed insect antifeedant activities [48]. Further research is necessary to shed light on the ecological role of **1** and **2**.

Conclusion

In conclusion the FIND adds a new isolation technique for filamentous fungi to the existing methods. FIND allows high throughput approaches and overcomes limitations in nutrition and growth conditions compared to artificial broths and laboratory cultivation. It enables to obtain axenic fungal cultures in a single isolation experiment and at the same time restrains fast growing fungi in favour of rare and less examined species. FIND ultimately supports the isolation of novel natural products by providing promising specimen from an environment of interest.

Experimental

FIND

The device was crafted from polyoxymethylene by the technical staff of the Helmholtz-Institute for Radiation- and Nuclear Physics, Nussallee 14–16, 53115 Bonn. The central plate (127 mm by 92 mm by 1 mm) and the two symmetrical upper and lower plates, the latter with screw threads, contain 96 through-holes 2 mm in diameter arranged in an 8 by 12 grid in a way that it perfectly aligns with the wells of a 96-well plate. The array of through-holes can be covered by one 142 mm polycarbonate membrane filter with a pore size of 0.03 micron (Sterlitech corp., Kent, WA). The device is sealed by six hex bolts size M3.

General experimental procedures

Optical rotations were measured with a Jasco DIP 140 polarimeter. ECD spectra were taken on a Jasco J-810 CD spectropolarimeter. UV and IR spectra were obtained using Perkin-Elmer Lambda 40 and Perkin-Elmer Spectrum BX FTIR instruments, respectively. All NMR spectra were recorded in MeOH-d_4 using a Bruker Avance 300 DPX spectrometer. Spectra were referenced to residual solvent signals with resonances at $\delta_{H/C}$ 3.35/49.0. HRESIMS were recorded on a LTQ Orbitrap mass spectrometer.

HPLC was performed on a Waters HPLC system equipped with a 1525µ binary pump, a 2998 PDA detector, Breeze 2 software and a Rheodyne 7725i injection system. A Macherey-Nagel Nucleoshell C_{18} column (250 mm × 4.6 mm; 5 µm), Nucleodur PolarTec column (250 mm × 4.6 mm; 5 µm), Pyramid C_{18} column (250 mm × 4.6 mm; 5 µm) and Phenomenex Kinetex C_{18} column (250 mm × 4.6 mm; 5 µm) as well as a Phenomenex Aqua C_{18} column (250 mm × 10 mm; 5 µm) were used.

Taxonomic identification of strains

Sequencing of the obtained isolates from the FIND experiments was performed by P. Massart and C. Decock at the Belgium coordinated collections of microorganisms (BCCM/IHEM, Brussels) by extracting DNA using their usual protocols and genomic DNA extraction kits and generation of partial sequences of DNA loci using standard primers of ITS1, ITS2, ITS4, 5.8S, RPBII, tef1α and β-tubulin. Digitalized and formatted DNA sequences were compared to reference sequences using BLASTn. ID% of the ITS regions was used to identify each strain. If multiple strains came into question the ID% of other gene sequences were taken into consideration to find the best match. Additionally, macromorphology was examined by colony observation using a Leica MZ6 stereomicroscope (Leica, Wetzlar, Germany) and micromorphology was examined with an Olympus BX51 polarizing microscope (Olympus, Shinjuku, Japan) by analysis of conidia and mycelium to support the bioinformatic suggestions.

Fermentation, extraction and isolation

For cultivation of the marine isolates from the diffusion chamber device experiments inside and outside of the device a minimal nutrient medium – FIND agar medium (FAM: 0.05% casamino acids, 0.025% fish peptone, 1.5% agar in sea water from the given sample) – was compiled to oppress growth of fast-growing fungi like common air moulds. For soil samples fish peptone was exchanged with meat peptone and deionized water was added instead of sea water. To the cooled autoclaved medium 250 mg penicillin G potassium salt (Roth, CN29.1) and streptomycin sulfate (Sigma-Aldrich, S6501-26G) were added via sterile filtration obtaining a concentration of 250 mg/L each to prevent bacterial growth.

For scale-up fermentation, *H.* cf. *alpina* was cultivated in 42 Fernbach flasks containing 150 mL of BMS agar medium (20 g biomalt in 1 L artificial seawater) each at 25 °C for 28 days. The fermented material of every two Fernbach flasks was extracted three times with 300 mL ethyl acetate (EtOAc) and the organic solvent was evaporated under vacuum to yield 2 g of crude extract. This crude extract was dissolved in 100 mL 80% (v/v) aqueous methanol (MeOH) and successively shaken three times with 100 mL petroleum ether in a separation funnel. MeOH solubles (250 mg) were fractionated by vacuum liquid

chromatography (VLC) over Polygoprep 60-50 C_{18} stationary phase (Macherey-Nagel) using gradient elution from 0:100 (MeOH/H_2O) to 100 % MeOH (25% steps) to yield 4 fractions. 100 mL of the mobile phase was used for each fraction. Fraction 3 was again fractionated by VLC under the before mentioned conditions to yield 4 subfractions: 3.1–3.4. Fraction 3.2 (44 mg) was analysed with the NMR system. Complex spectra obtained from VLC fraction 3.2 by NMR analysis indicated the presence of a wide range of secondary metabolites. This VLC fraction was thus subsequently subjected to further chromatographic separation.

RP-HPLC separation of VLC fraction 3.2 (MeOH/H_2O 55:45 Phenomenex Luna column (250 × 10 mm, 5 µm, 2 mL/min) obtained 10 fractions (3.2H1–3.2H9). Further HPLC of 3.2H7 and 3.2H9 resulted in the isolation of **1** (3 mg) and **2** (4 mg).

Heydenoic acid A (**1**): yellow oil; $[\alpha]_D^{25}$ −50.0 (*c* 0.24, MeOH); UV (MeOH) λ_{max} (log ε): 203 (4.0), 261 (3.5) nm; CD (*c* 4.0 × 10^{-3} M, MeOH) λ_{max} (Δ ε) 203 (1.33), 207 (1.21), 213 (−0.57), 321 (−0.75); IR ν_{max}: 3648, 2930, 1706, 1683, 1558, 1540, 1507, 1435, 1376, 1237, 634 cm^{-1}; HRESI-TOF-MS *m/z*: $[M + H]^+$ calcd for $C_{15}H_{21}O_3$ 249.1481; found, 249.1485. For ^1H and ^{13}C NMR data see Supporting Information File 1, Table S4.

Heydenoic acid B (**2**): yellow oil; $[\alpha]_D^{25}$ −34.9 (*c* 0.33, MeOH); UV (MeOH) λ_{max} (log ε): 203 (3.8), 258 (3.5) nm; CD (*c* 4.0 × 10^{-3} M, MeOH) λ_{max} (Δ ε) 208 (0.57), 218 (0.45), 321 (−0.38); IR ν_{max}: 3346, 2927, 1670, 1558, 1540, 1507, 1436, 1376, 1237, 1037, 634 cm^{-1}; HRESI-TOF-MS *m/z*: $[M + H]^+$ calcd for $C_{15}H_{23}O_3$ 251.1641; found, 251.1642. For ^1H and ^{13}C NMR data see Supporting Information File 1, Table S5.

Agar diffusion assay

Culture plates (5% sheep blood Columbia agar, BD) were overlaid with 3 mL tryptic soy soft agar, inoculated with TSB (Tryptic soy broth, Oxoid) growth suspension of the bacteria to be tested. Compounds were diluted to a concentration of 1 mg/mL (syringomycin 0.5 mg/mL) with DMSO and 3 µL of this dilution were placed on the surface of the agar. Compounds diffuse into the agar and the size of the inhibition zone was measured after 24 hours of incubation at 37 °C.

Disk diffusion assay

Antimicrobial tests of isolated compounds were performed by Ekaterina Egereva (Institute for Pharmaceutical Biology, University of Bonn) following the method described by Schulz et al. [49]. The bacteria *Bacillus megaterium* and *Escherichia coli* were used as representatives for Gram-positive and Gram-negative bacteria. *Microbotryum violaceum* (Basidiomycete), *Eurotium rubrum* (Ascomycete), and *Mycotypha microspora* (Zygomycete) were used as fungal test organisms.

Raw extracts were dissolved in MeOH to give a concentration of 1 mg/mL per test sample. 50 µL (equivalent to 50 µg) of each solution were pipetted onto sterile filter disks (diameter: 9 mm, Carl Roth GmbH & Co. KG, KA08.1) which was then placed onto the appropriate agar medium and sprayed with a suspension of the test organism. Growth media, preparation of spraying suspensions and conditions of incubation were carried out according to Schulz et al. [49]. For tested samples, a growth inhibition zone ≥3 mm and/or a complete inhibition ≥1 mm, measured from the edge of the filter disk, were regarded as a positive result. Growth inhibition was defined as follows: growth of the appropriate test organism was significantly inhibited compared to a negative control; total inhibition: no growth at all in the appropriate zone. Benzylpenicillin (1 mg/mL MeOH), streptomycin (1 mg/mL MeOH) and miconazole (0.5 mg/mL DCM) were used as positive controls.

Cytotoxicity assay

HEK293 cells were seeded into a PDL-coated 96-well plate (25.000/well) and allowed to settle for 2 hours prior to addition of the compounds (solubilized in water). After incubating for 22 hours, CellTiter Blue reagent (Promega) has been added. After another 2 hours of incubation the fluorescence has been detected using FlexStation 3 (Molecular Devices).

FIND procedure

For experiment 1, 5 mL of soil (Falcon tube) was diluted in 45 mL of sterile-filtered NaCl 0.85% solution (Ringer solution). This solution underwent manual agitation for 10 minutes and ultrasonication using a Bandelin Sonorex at 100% amplitude (35 kHz) two times for 5 seconds to detach cells from sample surfaces. After centrifugation using an Thermo Scientific Heraeus Multifuge X1R at 130 gravities for 5 minutes, cell counting was performed on the supernatant using a Neubauer Improved or Thoma chamber on an Olympus BX51 polarizing microscope at 100× magnification, and the abundance of bacteria was calculated by counting bacterial cells in all 25 medium squares of the central large square or all 16 medium squares of one of the large square, respectively. With Neubauer Improved chambers the supernatant was exchanged three times, for Thoma chambers three different large squares were counted. All the cells within each medium square (totally) together with those that are over the top and right sides of the square (even partially) were considered inside of a square. Mean cell counts were calculated and converted into concentrations by multiplying cell counts with the reciprocal value of the volume of a small square in L, considering possible dilution steps. For experiments 2 and 3, samples were treated alike, but instead of a

Ringers solution, sea water from the respective sampling site was used. For experiment 4 only sea water was used.

With the calculated abundances 2×10^{13} as well as 7×10^9, 13×10^9 and 8×10^9 bacterial cells/L for experiments 1–4, respectively, volumes between 10 and 50 mL of the solutions were taken according to the corresponding assumed bacteria-to-fungi ratios of the given sample type. These volumes, e.g., 16, 24 and 42 mL, respectively, were then added to a volume of FAM solution to obtain a concentration of 320.000 fungal parts per litre.

For experiment 4, 500 mL portions of a 10 L sample were poured through a sieve with a mesh size of 35 (0.5 mm) onto Whatman filter paper disks 602H with a pore size of 2 µm attached to a Büchner funnel. The funnel was put onto a 2 L suction bottle which was connected to a KNF Neuberger N811 KN.18 vacuum pump that provided negative pressure in the bottle and therefore ensured filtration of the sample portions. The residues on the filter paper disks (5 mL) were resuspended in 5 mL autoclaved sea sample and also mixed with 15 mL warm FAM solution resulting in a concentration of approximately 320 000 fungal parts/L. The prepared FAM solutions containing the corresponding samples were homogenized by stirring and the central plate was dipped into it. The desired target concentration is necessary to allow only one fungal part per through hole in the central plate. As the agar solidifies, the specimen is trapped in the chamber. A semipermeable polycarbonate membrane filter was attached to both sides of the central plate. The upper and lower parts of the FIND were connected to each other with six M3 hex bolt screws, sandwiching the central plate with the attached membranes between them. To ensure the devices seal commercial aquarium safe and antifungal-less silicone (Soudal) was applied to the edges of the device. The silicone was given time to dry and the device was inserted into a tank containing 40 L of soil for experiment 1, 2 L of sea water and sediment for experiment 2, 20 L of sea water and sediment for experiment 3 and sea water for experiment 4. An aquarium powerhead pump provided water current for simulation, oxygenation and distribution purposes for experiments 2–4. Evaporated tank water was replaced by autoclaved deionized water for experiments 2–4 and the soil in experiment 1 was watered with autoclaved deionized water every two days.

After four weeks of incubation inside of the tank the device was removed, rinsed with autoclaved deionized water to remove sediment and soil and opened under a laminar air flow cabinet to transfer each agar plug to a corresponding agar plate for cultivation. For transfer into 96 well plates a Boekel replica plater was used to punch all agar plugs from the central plate into the corresponding wells containing liquid FAM medium. For transfer onto agar plates, autoclaved wooden toothpicks were used to pick each agar plug from the central plate manually and streak them onto agar plates. The plates were stored in a Memmert BE500 incubator at 25 °C for four weeks.

Salt dependency experiments

Artificial sea water was used for the preparations of the agar plates. The stock solution containing 23.48 g NaCl, 10.61 g $MgCl_2 \cdot 6H_2O$, 3.92 g Na_2SO_4, 1.47 g $CaCl_2 \cdot 2H_2O$, 0.66 g KCl, 0.19 g $NaHCO_3$, 0.1 g KBr, 0.04 g $SrCl_2$, 0.03 g H_3BO_3 per litre deionized water equals 35‰ salinity. Dilutions with 7, 14, 21 and 28‰ water were achieved by adding deionized water to portions containing 480, 360, 240, 120 mL of the stock solutions, respectively, to obtain 600 mL each. For the 0‰ medium 600 mL deionized water was used. 20 mg biomalt and 15 mg agar per litre were added to the solutions and pH was adjusted at 7 by adding NaOH/HCl. After autoclaving for 20 minutes at 121 °C using an H+P Varioklav autoclave 100 mL of cooled agar solution were poured into petri dishes 130 mm in diameter. Agar plugs (1 cm^2) cut from pure Petri dish cultures of *Cladosporium allicinum* and *Heydenia* cf. *alpina* were transferred onto the centre of the prepared agar plates. The experiment was performed in triplicate for each salinity level.

Supporting Information

Supporting Information File 1

Genomic sequences of isolated fungi, data on bioactivity and halotolerance, spectroscopic data of compounds **1** and **2** from *Heydenia* cf. *alpina* strain 824.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-216-S1.pdf]

Acknowledgements

We are grateful to Tobias Meinert and Max Schirp Schönen (RWTH Aachen, Aachen, Germany) for 3D modeling of the FIND and to Tobias Benkel (University of Bonn, Bonn, Germany) for performing of the toxicity assays.

ORCID® iDs

Henrik Harms - https://orcid.org/0000-0003-3888-0536
Max Crüsemann - https://orcid.org/0000-0001-6660-2715
Gabriele M. König - https://orcid.org/0000-0003-0003-4916

References

1. Koehn, F. E. *Med. Chem. Commun.* **2012,** *3,* 854–865. doi:10.1039/c2md00316c
2. Dias, D. A.; Urban, S.; Roessner, U. *Metabolites* **2012,** *2,* 303–336. doi:10.3390/metabo2020303

3. Staley, J. T.; Konopka, A. *Annu. Rev. Microbiol.* **1985,** *39,* 321–346. doi:10.1146/annurev.mi.39.100185.001541
4. Nichols, D.; Cahoon, N.; Trakhtenberg, E. M.; Pham, L.; Mehta, A.; Belanger, A.; Kanigan, T.; Lewis, K.; Epstein, S. S. *Appl. Environ. Microbiol.* **2010,** *76,* 2445–2450. doi:10.1128/aem.01754-09
5. Ling, L. L.; Schneider, T.; Peoples, A. J.; Spoering, A. L.; Engels, I.; Conlon, B. P.; Mueller, A.; Schäberle, T. F.; Hughes, D. E.; Epstein, S.; Jones, M.; Lazarides, L.; Steadman, V. A.; Cohen, D. R.; Felix, C. R.; Fetterman, K. A.; Millett, W. P.; Nitti, A. G.; Zullo, A. M.; Chen, C.; Lewis, K. *Nature* **2015,** *517,* 455–459. doi:10.1038/nature14098
6. Willis, K. J. Definition and Diversity. *State of the World's Fungi 2018;* 2018; pp 9 ff.
7. Lok, C. *Nature* **2015,** *522,* 270–273. doi:10.1038/522270a
8. Ahrendt, S. R.; Quandt, C. A.; Ciobanu, D.; Clum, A.; Salamov, A.; Andreopoulos, B.; Cheng, J.-F.; Woyke, T.; Pelin, A.; Henrissat, B.; Reynolds, N. K.; Benny, G. L.; Smith, M. E.; James, T. Y.; Grigoriev, I. V. *Nat. Microbiol.* **2018,** *3,* 1417–1428. doi:10.1038/s41564-018-0261-0
9. Grossart, H.-P.; Wurzbacher, C.; James, T. Y.; Kagami, M. *Fungal Ecol.* **2016,** *19,* 28–38. doi:10.1016/j.funeco.2015.06.004
10. Hawksworth, D. L.; Rossman, A. Y. *Phytopathology* **1997,** *87,* 888–891. doi:10.1094/phyto.1997.87.9.888
11. Meliani, A.; Bensoltane, A.; Mederbel, K. *Am. J. Plant Nutr. Fert. Technol.* **2012,** *2,* 10–18. doi:10.3923/ajpnft.2012.10.18
12. Zhang, Z.; Qu, Y.; Li, S.; Feng, K.; Wang, S.; Cai, W.; Liang, Y.; Li, H.; Xu, M.; Yin, H.; Deng, Y. *Sci. Rep.* **2017,** *7,* No. 4837. doi:10.1038/s41598-017-05260-w
13. Bressan, M.; Trinsoutrot Gattin, I.; Desaire, S.; Castel, L.; Gangneux, C.; Laval, K. *Appl. Soil Ecol.* **2015,** *88,* 60–68. doi:10.1016/j.apsoil.2014.12.007
14. Camacho-Fernández, C.; Hervás, D.; Rivas-Sendra, A.; Marín, M. P.; Seguí-Simarro, J. M. *Plant Methods* **2018,** *14,* No. 30. doi:10.1186/s13007-018-0297-4
15. Jedicke, E. *Böden - Entstehung, Ökologie, Schutz;* O. Maier: Ravensburg, Germany, 1989.
16. Gaertner, A. *Plant Biol.* **1969,** *82,* 287–306.
17. Gaertner, A. *Veroeff. Inst. Meeresforsch. Bremerhaven* **1968,** *3,* 75–92.
18. Crow, S. A.; Ahearn, D. G.; Cook, W. L.; Bourquin, A. W. *Limnol. Oceanogr.* **1975,** *20,* 644–646. doi:10.4319/lo.1975.20.4.0644
19. Pretsch, E.; Bühlmann, P.; Badertscher, M. *Spektroskopische Daten zur Strukturaufklärung organischer Verbindungen,* 5th ed.; Springer: Berlin, Germany, 2010; p 289. doi:10.1007/978-3-540-76866-1
20. Jacobsen, N. E. Chirality and Stereochemistry: Natural Products. *NMR Data Interpretation Explained: Understanding 1D and 2D NMR Spectra of Organic Compounds and Natural Products;* John Wiley and Sons, Inc.: Hoboken, NJ, U.S.A., 2017; pp 294–297.
21. Wu, Z.-Y.; Wu, Y.; Chen, G.-D.; Hu, D.; Li, X.-X.; Sun, X.; Guo, L.-D.; Li, Y.; Yao, X.-S.; Gao, H. *RSC Adv.* **2014,** *4,* 54144–54148. doi:10.1039/c4ra10365c
22. Tanaka, Y.; Sato, H.; Kageyu, A.; Tomita, T. *J. Chromatogr.* **1985,** *347,* 275–283. doi:10.1016/s0021-9673(01)95493-7
23. Khisal, A. A.; Rabenstein, J. *J. Ind. Microbiol. Biotechnol.* **2004,** *31,* 11–15. doi:10.1007/s10295-003-0106-5
24. Deshmukh, S. K.; Verekar, S. A. *J. Mycol. Med.* **2014,** *24,* 319–327. doi:10.1016/j.mycmed.2014.08.004
25. Gonçalves, V. N.; Vitoreli, G. A.; de Menezes, G. C. A.; Mendes, C. R. B.; Secchi, E. R.; Rosa, C. A.; Rosa, L. H. *Extremophiles* **2017,** *21,* 1005–1015. doi:10.1007/s00792-017-0959-6
26. Segers, F. J. J.; Meijer, M.; Houbraken, J.; Samson, R. A.; Wösten, H. A. B.; Dijksterhuis, J. *PLoS One* **2015,** *10,* e0145415. doi:10.1371/journal.pone.0145415
27. Woudenberg, J. H. C.; Groenewald, J. Z.; Binder, M.; Crous, P. W. *Stud. Mycol.* **2013,** *75,* 171–212. doi:10.3114/sim0015
28. Zalar, P.; de Hoog, G. S.; Schroers, H.-J.; Crous, P. W.; Groenewald, J. Z.; Gunde-Cimerman, N. *Stud. Mycol.* **2007,** *58,* 157–183. doi:10.3114/sim.2007.58.06
29. Rédau, V.; Kumar, A.; Hainaut, M.; Henrissat, B.; Record, E.; Barbier, G.; Burgaud, G. *Genome Announce.* **2016,** *4,* e00467-16. doi:10.1128/genomea.00467-16
30. Almeida, C.; Eguereva, E.; Kehraus, S.; Siering, C.; König, G. M. *J. Nat. Prod.* **2010,** *73,* 476–478. doi:10.1021/np900608d
31. Schönborn, C. *Mykosen* **1968,** *11,* 847–864. doi:10.1111/j.1439-0507.1968.tb04450.x
32. Crous, P. W.; Wingfield, M. J.; Guarro, J.; Cheewangkoon, R.; van der Bank, M.; Swart, W. J.; Stchigel, A. M.; Cano-Lira, J. F.; Roux, J.; Madrid, H.; Damm, U.; Wood, A. R.; Shuttleworth, L. A.; Hodges, C. S.; Munster, M.; de Jesús Yáñez-Morales, M.; Zúñiga-Estrada, L.; Cruywagen, E. M.; De Hoog, G. S.; Silvera, C.; Najafzadeh, J.; Davison, E. M.; Davison, P. J. N.; Barrett, M. D.; Barrett, R. L.; Manamgoda, D. S.; Minnis, A. M.; Kleczewski, N. M.; Flory, S. L.; Castlebury, L. A.; Clay, K.; Hyde, K. D.; Maússe-Sitoe, S. N. D.; Chen, S.; Lechat, C.; Hairaud, M.; Lesage-Meessen, L.; Pawłowska, J.; Wilk, M.; Śliwińska-Wyrzychowska, A.; Mętrak, M.; Wrzosek, M.; Pavlic-Zupanc, D.; Maleme, H. M.; Slippers, B.; Mac Cormack, W. P.; Archuby, D. I.; Grünwald, N. J.; Tellería, M. T.; Dueñas, M.; Martín, M. P.; Marincowitz, S.; de Beer, Z. W.; Perez, C. A.; Gené, J.; Marin-Felix, Y.; Groenewald, J. Z. *Persoonia - Mol. Phylog. Evol. Fungi* **2013,** *31,* 188–296. doi:10.3767/003158513x675925
33. Valldosera, M.; Guarro, J.; Figueras, M. J. *Mycol. Res.* **1991,** *95,* 243–246. doi:10.1016/s0953-7562(09)81020-4
34. Fogle, M.; Douglas, D.; Jumper, C.; Straus, D. *Int. J. Mol. Sci.* **2008,** *9,* 2357–2365. doi:10.3390/ijms9122357
35. Yamagishi, Y.; Matsuoka, M.; Odagawa, A.; Kato, S.; Shindo, K.; Mochizuki, J. *J. Antibiot.* **1993,** *46,* 884–887. doi:10.7164/antibiotics.46.884
36. Yamagishi, Y.; Kato, S.; Kawai, H. *J. Antibiot.* **1993,** *46,* 888–891. doi:10.7164/antibiotics.46.888
37. Clark, B. R.; Capon, R. J.; Lacey, E.; Tennant, S.; Gill, J. H. *Org. Lett.* **2006,** *8,* 701–704. doi:10.1021/ol052880y
38. Dias, A.; Ruiz, N.; Couzinet-Mossion, A.; Bertrand, S.; Duflos, M.; Pouchus, Y.-F.; Barnathan, G.; Nazih, H.; Wielgosz-Collin, G. *Mar. Drugs* **2015,** *13,* 4934–4948. doi:10.3390/md13084934
39. Zhai, M.-M.; Qi, F.-M.; Li, J.; Jiang, C.-X.; Hou, Y.; Shi, Y.-P.; Di, D.-L.; Zhang, J.-W.; Wu, Q.-X. *J. Agric. Food Chem.* **2016,** *64,* 2298–2306. doi:10.1021/acs.jafc.6b00556
40. Yu, Z.; Lang, G.; Kajahn, I.; Schmaljohann, R.; Imhoff, J. F. *J. Nat. Prod.* **2008,** *71,* 1052–1054. doi:10.1021/np070580e
41. Fischer, E. *Mitt. Naturforsch. Ges. Bern* **1920,** *99,* No. 19.
42. Iwanicki, N. S.; Pereira, A. A.; Botelho, A. B. R. Z.; Rezende, J. M.; Moral, R. d. A.; Zucchi, M. I.; Delalibera Júnior, I. *Sci. Rep.* **2019,** *9,* No. 4443. doi:10.1038/s41598-019-38594-8
43. Leuchtmann, A.; Clémençon, H. *Mycol. Prog.* **2012,** *11,* 699–710. doi:10.1007/s11557-011-0779-5
44. Gonçalves, V. N.; Vaz, A. B. M.; Rosa, C. A.; Rosa, L. H. *FEMS Microbiol. Ecol.* **2012,** *82,* 459–471. doi:10.1111/j.1574-6941.2012.01424.x

45. Palmer-Young, E. C.; Veit, D.; Gershenzon, J.; Schuman, M. C. *PLoS One* **2015,** *10,* e0127296. doi:10.1371/journal.pone.0127296
46. Coates, R. M.; Denissen, J. F.; Juvik, J. A.; Babka, B. A. *J. Org. Chem.* **1988,** *53,* 2186–2192. doi:10.1021/jo00245a012
47. Snyder, J. C.; Guo, Z.; Thacker, R.; Goodman, J. P.; Pyrek, J. S. *J. Chem. Ecol.* **1993,** *19,* 2981–2997. doi:10.1007/bf00980597
48. Perry, N. B.; Burgess, E. J.; Foster, L. M.; Gerard, P. J.; Toyota, M.; Asakawa, Y. *J. Nat. Prod.* **2008,** *71,* 258–261. doi:10.1021/np070644l
49. Schulz, B.; Boyle, C.; Draeger, S.; Römmert, A.-K.; Krohn, K. *Mycol. Res.* **2002,** *106,* 996–1004. doi:10.1017/s0953756202006342

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Current understanding and biotechnological application of the bacterial diterpene synthase CotB2

Ronja Driller[1,2,3], Daniel Garbe[4], Norbert Mehlmer[4], Monika Fuchs[4], Keren Raz[5], Dan Thomas Major[*5], Thomas Brück[*4] and Bernhard Loll[*1]

Review

Open Access

Address:
[1]Institute of Chemistry and Biochemistry, Laboratory of Structural Biochemistry, Freie Universität Berlin, Takustr. 6, 14195 Berlin, Germany, [2]present address: Department of Molecular Biology and Genetics, Aarhus University, Gustav Wieds Vej 10, 8000 Aarhus C, Denmark, [3]present address: Danish Research Institute of Translational Neuroscience - DANDRITE, Nordic-EMBL Partnership for Molecular Medicine, Aarhus C, Denmark, [4]Werner Siemens Chair of Synthetic Biotechnology, Dept. of Chemistry, Technical University of Munich (TUM), Lichtenbergstr. 4, 85748 Garching, Germany and [5]Department of Chemistry, Bar-Ilan University, Ramat-Gan 52900, Israel

Email:
Dan Thomas Major[*] - majort@biu.ac.il; Thomas Brück[*] - brueck@tum.de; Bernhard Loll[*] - loll@chemie.fu-berlin.de

* Corresponding author

Keywords:
biotechnology; CotB2; crystal structure; cyclooctatin; diterpene; reaction mechanism; terpene synthase

Beilstein J. Org. Chem. **2019,** *15,* 2355–2368.
doi:10.3762/bjoc.15.228

Received: 10 July 2019
Accepted: 12 September 2019
Published: 02 October 2019

This article is part of the thematic issue "Terpenes".

Guest Editor: J. S. Dickschat

Abstract

CotB2 catalyzes the first committed step in cyclooctatin biosynthesis of the soil bacterium *Streptomyces melanosporofaciens.* To date, CotB2 represents the best studied bacterial diterpene synthase. Its reaction mechanism has been addressed by isoptope labeling, targeted mutagenesis and theoretical computations in the gas phase, as well as full enzyme molecular dynamic simulations. By X-ray crystallography different snapshots of CotB2 from the open, inactive, to the closed, active conformation have been obtained in great detail, allowing us to draw detailed conclusions regarding the catalytic mechanism at the molecular level. Moreover, numerous alternative geranylgeranyl diphosphate cyclization products obtained by CotB2 mutagenesis have exciting applications for the sustainable production of high value bioactive substances.

Introduction

Terpenes represent one of the most diverse groups of natural biomolecules [1-3]. Sesqui- and diterpenes are a diverse class of secondary metabolites derived predominantly from plants, marine invertebrates, fungi and some prokaryotes [4-8]. Properties of these natural products include antitumor, anti-oxidant, anti-inflammatory, antiviral, antimalarial, antibiotic, neuroprotective and even insecticidal activities, which makes these compounds high-value commercial targets for the chemical and

pharmaceutical industry [9,10]. Structural diversity of diterpenes is created by the terpene synthase (TPS) enzyme family, which use acyclic isoprenoid precursors to generate a vast number of (poly)cyclic hydrocarbon scaffolds. Remarkably, this complex chemical reaction, comprising changes in bonding, hybridization as well as the introduction of specific stereochemistry, is performed in a single reaction cascade without consumption of a cofactor [11]. In this review we will focus particularly on bacterial diterpene synthases, in context with other sesqui- and ditperpene synthases of bacterial, fungal and plant origin.

The initial step in diterpene biosynthesis (Figure 1) is the incremental condensation of dimethylallyl diphosphate (**1**) and isopentylen diphosphate (**2**) [12] to the acyclic terpene synthase substrate geranylgeranyl diphosphate **3** (GGDP) [1,13-16]. Following initial substrate binding and folding in a product-like conformation, the cyclization reaction can be subdivided into three steps: (1) generation of a reactive allyl carbocation as a result of heterolytic cleavage of the pyrophosphate–hydrocarbon bond or protonation of a double bond, (2) propagation of the carbocation along the forming terpene skeleton as a result of ring formations, hydride and/or methyl shifts, de- and reprotonation of intermediates, the creation of a terminal carbocation (3) and finally the quenching of the carbocation by a base or water [16,17].

TPSs can be divided into two distinct classes, which are distinguished by their substrate activation mechanism. Whereas ionization of an isoprenoid diphosphate is caused by hydrolysis of the pyrophosphate by a trinuclear metal cluster in class I terpenoid cyclases, class II reactions proceed via a protonation of the terminal carbon–carbon double bond of an isoprenoid substrate [3]. In addition to the differences in the activation mechanism, the two classes of TPSs have an unrelated overall fold. Class I TPSs establish an α-helical bundle fold [18] and are structurally related to the isoprenyl diphosphate synthases such as the farnesyl (C15) or geranylgeranyl (C20) diphosphate synthase, which generate the substrates of sesqui- and diterpene synthases, respectively. In bacteria, diterpene synthases almost exclusively belong to class I TPSs. A few exceptions exist, for example the terpentedienyl-diphosphate synthase from *Kitasatospora griseola* [19], which belongs to class II, and PtmT1 from *Streptomyces platensis* in the platensimycin gene cluster [20], which is neither characteristic for class I nor class II. In fungi and plants, they can be both class I or class II TPSs and even mixed class I/II [21]. Class I TPSs harbor two distinct catalytic motifs, which are crucial for binding and correctly positioning the substrate for catalysis, the aspartate-rich motif (DDxxD) and the NSE/DTE motif, that bind to three Mg^{2+} ions [1,22]. Notably, the PtmT1 diterpene synthase from *S. platensis* lacks both catalytic motifs [20,23]. The NSE/DTE motif illustrates the distribution of DTE motifs, which are predominantly found in plant TPSs, and NSE motifs, which are more common in bacterial and fungal TPSs. For the description of the bacterial TPS CotB2, we will therefore refer to the NSE motif only.

Using host microorganisms, such as bacteria or baker's yeast for the heterologous synthesis of terpenes increases the sustainability of bioactive terpene production by saving resources, as the production host can be fed with residual organic waste streams, such as milling or forestry waste. Additionally, the heterologous terpene production minimizes waste generation as the targeted production of a single compound reduces extraction and purification steps [24,25]. Additionally, heterologous production enables protein engineering to optimize product ratios or to alter the native product portfolio of the enzyme [9,26,27]. Furthermore, production by engineered microorganisms considerably reduces the cost compared to total chemical synthesis or extraction from natural sources, since the target compounds are produced from inexpensive carbon sources. Prominent examples of optimized terpene production pathways in *E. coli* are taxadiene, a precursor of the anticancer drug taxol

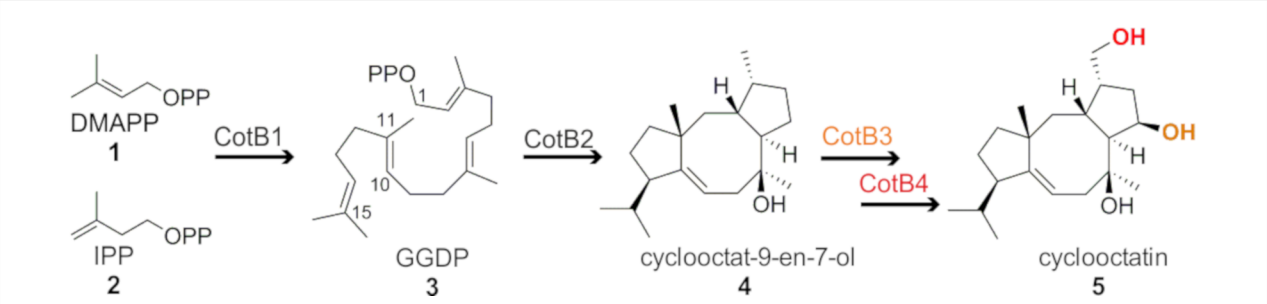


Figure 1: CotB1 synthesizes geranylgeranyl diphosphate (GGDP) **3** from the substrates dimethylallyl diphosphate (DMAPP) **1** and isopentenyl diphosphate (IPP) **2**. The acyclic substrate GGDP (**3**) is stereospecifically cyclized by CotB2 to cyclooctat-9-en-7-ol (**4**), with a fusicoccane 5–8–5 fused ring system. Two cytochrome P450 enzymes, CotB3 and CotB4, subsequently functionalize cyclooctat-9-en-7-ol (**4**) to the bioactive compound cyclooctatin (**5**).

[28], amorpha-4,11-diene, an antimalarial drug precursor [29], and cyclooctatin [30].

The scope of this review encompasses a detailed consideration of the cyclooctatin biosynthetic gene cluster in particular the TPS CotB2 from the soil bacterium *Streptomyces melanosporofaciens* MI614-43F2 (Figure 1) [31]. Cyclooctatin **5**, with its distinct 5–8–5 ring motif, belongs to the fusicoccane diterpenoids that encompass a wide range of bioactivities, such as bacteriostatic, fungicidal and tumorstatic effects [32]. A key player in the biosynthesis of cyclooctatin **5** is the bacterial diterpene synthase CotB2.

Different research teams have investigated CotB2 by means of biochemical [30,31,38], biophysical [33-35], structural biology [36-39] and computational modeling approaches [35,37,39,40], revealing a rather unusual reaction mechanism. Given the diverse bioactivities of cyclooctatin **5**, CotB2 is of biotechnological relevance as well. Cyclooctatin **5** can be manufactured biotechnologically in *E. coli* in milligram scales [26,30], which is important for an industrial application. Within the subsequent paragraphs, we would like to summarize the current understanding of the unusual reaction mechanism and the biotechnological applications.

Review

The cyclooctatin biosynthetic gene cluster

The cyclooctatin gene cluster from the soil bacterium *S. melanosporofaciens* MI614-43F2 comprises four enzymes: GGDP synthase (CotB1), a diterpene cyclase (CotB2) and two P450 cytochromes (Figure 1) [31]. The cyclization of GGDP **3** to cyclooctat-9-en-7-ol (**4**) is performed by the TPS CotB2. Compound **4** is further decorated with two hydroxy functions introduced by two P450 cytochromes, CotB3 and CotB4, to cyclooctatin (**5**). Depending on the functionalization of the ring system, the compounds demonstrate a broad diversity of biological activities, among others fungicidal or tumor static [32,43-45]. Cyclooctatin (**5**) inhibits a lysophospholipase, which catalyzes the hydrolysis of the fatty acid ester bonds of lysophospholipids [31,41]. Moreover, cyclooctatin (**5**) was shown to be effective against *Plasmodium falciparum* with an IC_{50} of 7.14 µg/mL along with very low cytotoxicity [42].

CotB2 belongs to the class of cyclopentane-forming diterpene synthases [46], a class of enzymes that is widely distributed among plants, fungi and bacteria. CotB2 has evolved to convert the acyclic, achiral substrate GGDP to the 5–8–5 ring motif of cyclooctat-9-en-7-ol that contains six chiral stereocenters. Hence, CotB2 has been fine tuned to perform a highly specific regio- and stereochemical reaction. The cyclization mechanism of CotB2 has been investigated extensively in recent years. By isotope labeling and NMR spectroscopic investigations [35], it has been shown that CotB2 catalyzes the complex regio- and stereospecific cyclization reaction with an unusual carbon–carbon bond rearrangement. Hong and Tantillo performed the first theoretical study of the reaction mechanism in CotB2 via gas phase density functional theory (DFT) calculations [33]. Simultaneously, Sato et al. also studied the CotB2 reaction mechanism in the gas phase using DFT, combined with experimental deuterium labeling [34]. Another important step towards a deeper understanding of the cyclization mechanism was the availability of crystal structures (Table 1) revealing structural snapshots along the reaction trajectory commencing with the open, inactive conformation completing with the closed, active conformation of CotB2. [36-38].

Overall structure of CotB2^{wt} in the open, inactive conformation

The structure of CotB2^{wt} (PDB-ID 4OMG [38] and PDB-ID 5GUC [36]) is complete, except for the 15 N-terminal and 12 C-terminal residues. CotB2 consists of ten core α-helices (A to J) that are connected by short loop segments and additional five short α-helices (α1 to α5; (Figure 2A and B)). CotB2 resembles the classical α-helical fold of TPSs [18] with significant differences in its primary sequence compared to other TPSs. The core α-helices surround a large, deep cleft, which

Table 1: Overview of available crystal structures of CotB2 and its variants.

protein/variant	PDB-ID	conformation	ligands	reference
CotB2^{wt}	4OMG	open	–	[38]
CotB2^{wt}	5GUC	open	–	[36]
CotB2^{F149L}	4OMH	open	–	[38]
CotB2$^{\Delta C}$	6GGK	open	–	[37,39]
CotB2^{F107A}·Mg^{2+}_{B}	6GGL	intermediate	1 Mg^{2+}	[37,39]
CotB2^{wt}·Mg^{2+}_{B}·GGSDP	5GUE	intermediate	1 Mg^{2+}, geranyl geranyl thiodiphosphate (GGSDP)	[36]
CotB2^{wt}·Mg^{2+}_{3}·F-Dola	6GGI	closed	3 Mg^{2+}, diphosphate, 2-fluoro-3,7,18-dolabellatriene	[37,39]
CotB2^{wt}·Mg^{2+}_{3}·AHD	6GGJ	closed	3 Mg^{2+}, 4-amino-1-hydroxy-1-phosphonobutyl phosphonic acid	[37,39]

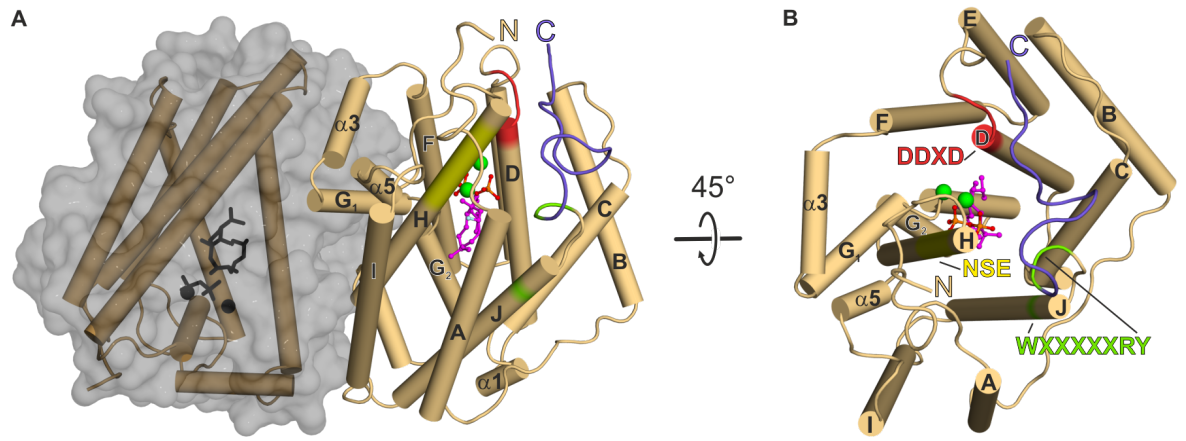


Figure 2: The bacterial diterpene synthase CotB2^{wt}·Mg^{2+}_3·F-Dola in the closed, active conformation (PDB-ID 6GGI; [37]). (A) The two monomers of CotB2 are shown in cartoon representation with the α-helices drawn as cylinders and colored in light brown. One monomer of CotB2 is shown in gray surface representation. The location of the aspartate-rich motif is indicated in red and the NSE-motif is marked in yellow. The WXXXXXRY motif is indicated in light-green. The last 12 C-terminal residues of the lid are drawn in violet. The three Mg^{2+} ions are represented as green spheres and the bound intermediate F-Dola is shown in magenta. The cleaved diphosphate is shown in orange. (B) View of panel A rotated by 45°. For clarity, one monomer has been removed. The view is into the active site of CotB2.

forms the active site (Figure 2A and B). CotB2 is arranged as a homo-dimer in crystallo [36,38] and it was demonstrated that CotB2 exists as a homo-dimer in solution as well [38]. The two active sites of CotB2 are arranged in an antiparallel fashion, resembling the arrangement initially observed for the monoterpene (+)-bornyl diphosphate synthase [47] and the sesquiterpene trichodiene synthase [48], but is in contrast to the parallel dimer as described for farnesyl diphosphate synthase [49]. Structurally, CotB2 is most closely related to the fungal monoterpene synthase aristolochene (PDB-ID 2OA6; [50]), and epi-isozizaene synthase (PDB-ID 3LGK; [51]), but not to the known plant structures of diterpene TPSs, such as the ent-copalyl diphosphate synthase (PDB-ID 3PYA; [52]) or taxadiene synthase (PDB-ID 3P5R; [53]).

Two crystal structures of CotB2 resembling two distinct precatalytic states

The structure of CotB2^{F107A} was determined with one single Mg^{2+} (Mg^{2+}_B) bound to the NSE motif (CotB2^{F107A}·Mg^{2+}_B; PDB-ID 6GGL [37]). The overall structure of CotB2 remains unchanged compared to the open state. This suggests that binding of Mg^{2+}_B is the initial step to prearrange CotB2 for substrate binding.

Tomita et al. obtained a crystal structure of CotB2^{wt} with bound inert substrate analogue geranylgeranyl thiodiphosphate (GGSDP; CotB2^{wt}·Mg^{2+}_B·GGSDP; PDB-ID 5GUE; [36]), resembling the next precatalytic state of CotB2. Surprisingly, there is only Mg^{2+}_B bound to the enzyme and not the full set of three Mg^{2+}-ions needed for catalysis. The position of Mg^{2+}_B is identical to the structure of CotB2^{F107A}·Mg^{2+}_B, but two coordinating water molecules have now been exchanged with two oxygen atoms of the diphosphate function of GGSDP. Upon binding of GGSDP, a bidentate salt bridge between D111 of the DDXD motif and R294 of the WXXXXXRY motif is formed, designating the transition from the open to the closed conformation. However as for the open, inactive conformation, the remaining C-terminal residues have not been traceable in the electron density (Figure 3), and hence CotB2 is not yet ready for catalysis. The position and tilting of the core α-helices in the structure of CotB2^{wt}·Mg^{2+}_B·GGSDP is practically indistinguishable from the open conformation of CotB2 (Figure 3). Since in CotB2^{wt}·Mg^{2+}_B·GGSDP two Mg^{2+} ions are absent the diphosphate moiety is not properly coordinated, resulting in significant substrate flexibility within the catalytically active site. Moreover, the two missing Mg^{2+} ions prevent helix D, accommodating the Asp-rich motif, from shifting towards the active site (Figure 3). Consequently, the active site remains partially open and the C-terminus cannot fold over the active site, making a proper substrate positioning and subsequent cyclization unlikely.

Crystal structure of CotB2^{wt}·Mg^{2+}_3·F-Dola resembles the closed, active conformation

The structure of CotB2^{wt}·Mg^{2+}_3·F-Dola has been obtained by co-crystallization of CotB2^{wt} and FGGDP (PDB-ID 6GGI; [37]), representing the closed, active conformation. Fluorinated substrates, such as FGGDP, have been used for the crystallization of other TPSs and have been shown to stick to the active site without undergoing cyclization, trapping the enzyme in a

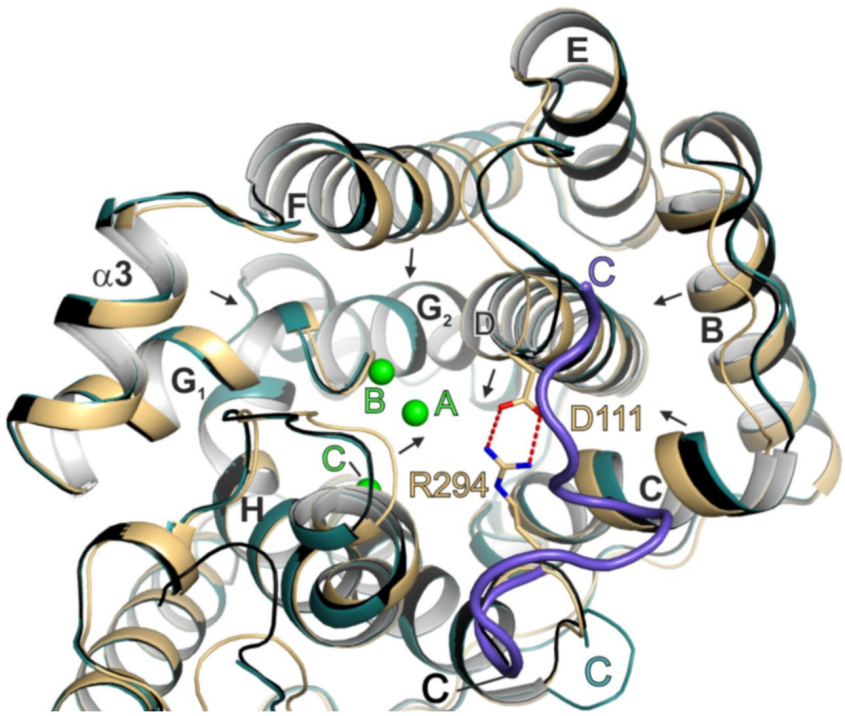


Figure 3: Conformational changes of CotB2 upon ligand binding. Superposition of CotB2's open (teal), pre-catalytic (black, $CotB2^{wt}{\cdot}Mg^{2+}{}_{B}{\cdot}GGSDP$), and fully closed (light-brown, $CotB2^{wt}{\cdot}Mg^{2+}{}_{3}{\cdot}F\text{-}Dola$) conformation. The overall fold of $CotB2^{wt}{\cdot}Mg^{2+}{}_{B}{\cdot}GGSDP$ (PDB-ID 5GUE; [36]) is more similar to $CotB2^{wt}$ (PDB-ID 4OMG; [38]) than to $CotB2^{wt}{\cdot}Mg^{2+}{}_{3}{\cdot}F\text{-}Dola$ (PDB-ID 6GGI; [37]). The salt bridge between D111 and R294, in stick representation, is shown by red, dashed lines. Mg^{2+} ions are colored in green. Black arrows indicate movement of secondary structure elements from the open to the closed conformation of CotB2. The different C-termini are labeled in all three structures. The C-terminus in the structure of $CotB2^{wt}{\cdot}Mg^{2+}{}_{3}{\cdot}F\text{-}Dola$ is colored in purple.

closed state [52,53]. Comparing the overall structure of $CotB2^{wt}{\cdot}Mg^{2+}{}_{3}{\cdot}F\text{-}Dola$ to the open conformation of CotB2 reveals major differences. The binding of FGGDP and three Mg^{2+} ions induces a translation and rotation of α-helices B, C, D, F and H towards the active site to accurately position the Asp-rich motif and to subsequently transfer the active site into a product-shaped conformation (Figure 3). Now, a significant change in the C-terminal region of the enzyme is observed as the former unstructured C-terminus becomes folded, thereby acting as a lid to shield the active site from bulk solvent (Figure 2). To exclude that the folded C-terminus is due to crystal contacts and hence a crystallographic artifact, CotB2 was co-crystallized with AHD, a compound that mimics the diphosphate group of GGDP and acts as a suicide inhibitor. The crystal structure of $CotB2^{wt}{\cdot}Mg^{2+}{}_{3}{\cdot}AHD$ (PDB-ID 6GGJ; [37]) confirmed the active site architecture as observed in the structure of $CotB2^{wt}{\cdot}Mg^{2+}{}_{3}{\cdot}F\text{-}Dola$ and more importantly, that the folding of the C-terminus is not a crystallographic artifact. To demonstrate the significance of the C-terminus, a CotB2 truncation ($CotB2^{\Delta C}$) was designed terminating at R294, thereby missing the 12 C-terminal residues, that correspond to the lid [37]. In activity experiments, no substrate conversion by $CotB2^{\Delta C}$ was observed, stressing the fact, that the salt bridge D111-R294 is insufficient for catalysis. Instead, the full C-terminus is required for catalysis.

In $CotB2^{wt}{\cdot}Mg^{2+}{}_{3}{\cdot}F\text{-}Dola$, there is no electron density observed for the intact FGGDP molecule. Instead two distinct and clearly separated electron densities are visible that could be readily interpreted as the reaction intermediate F-Dola, as well as a single diphosphate moiety. The FGGDP molecule – functionalized with a fluorine atom at the C2 position of its isoprenoid unit – can interfere with the propagation of the generated carbocation(s). Such cyclization reactions of fluorinated substrates have been reported previously [54,55]. However, the structure of $CotB2^{wt}{\cdot}Mg^{2+}{}_{3}{\cdot}F\text{-}Dola$ represents the first reported structure of a TPS with an in crystallo formed and bound reaction product, together with the cleaved-off diphosphate. This is in contrast to the structure of taxadiene synthase (PDB-ID 3P5R; [53]), co-crystallized with FGGDP, where the substrate is bound but the cyclization reaction has not been initialized and consequently the substrate is not yet cleaved. Identical observations have been made for the crystal structure of aristolochene synthase co-crystallized with 2-fluoro diphos-

phate (PDB-ID 3BNY; [56]). A structural comparison of CotB2^{wt}·Mg$^{2+}{}_3$·F-Dola and CotB2^{wt}·Mg$^{2+}{}_B$·GGSDP reveals that the hydrophobic tail of GGSDP adopts a position similar to the cyclized intermediate in our CotB2^{wt}·Mg$^{2+}{}_3$·F-Dola structure (Figure 4), likely reflecting the already product-shaped architecture of the active site.

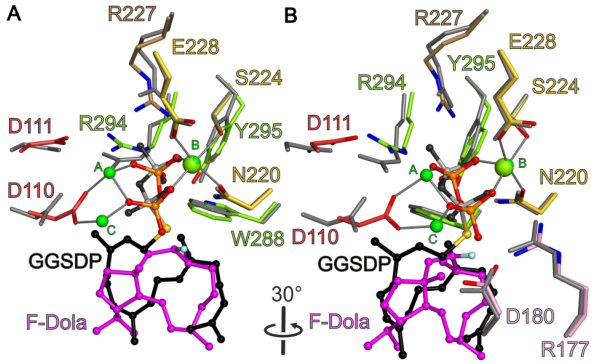


Figure 4: View into the active site of CotB2^{wt}·Mg$^{2+}{}_3$·F-Dola [37] superimposed with CotB2^{wt}·Mg$^{2+}{}_B$·GGSDP [36]. (A) The bound F-Dola reaction intermediate is shown in magenta with the fluorine atom colored in light blue, and Mg^{2+}ions are shown in green. Residues of the DDXD motif are shown in red and residues of the NSE motif in yellow. The conserved residues of the WXXXXXRY motif are drawn in light green and the R227 that interacts with diphosphate moiety in light brown. The cleaved diphosphate is shown in orange. Gray lines represent the coordination sphere of the Mg^{2+} ions. Identical amino acid residues located in the structure of CotB2^{wt}·Mg$^{2+}{}_B$·GGSDP are shown in gray. The GGSDP molecule is shown in black with the sulfur atom colored in yellow. The position of Mg$^{2+}{}_B$ is identical in both structures. For clarity water molecules have been omitted. (B) View in panel A rotated by 30°. In addition to the motifs shown in panel A, the pyrophosphate sensor motif is depicted in pink.

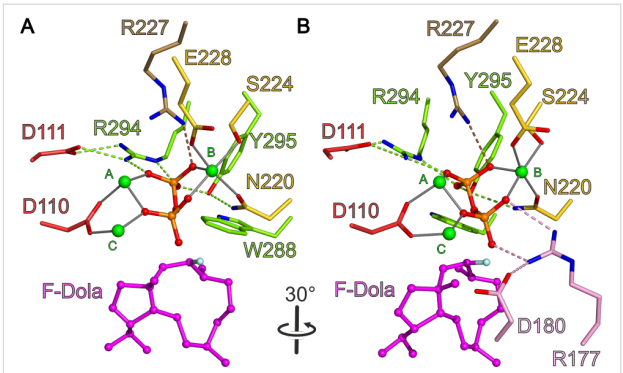


Figure 5: View into the active site of CotB2^{wt}·Mg$^{2+}{}_3$·F-Dola [37]. Identical view as in Figure 4. (A) The bound F-Dola reaction intermediate is shown in magenta with the fluorine atom colored in light blue, and Mg^{2+}-ions are shown in green. Residues of the DDXD motif are shown in red and residues of the NSE motif in yellow. The conserved residues of the WXXXXXRY motif are drawn in light green and the R227 that interacts with diphosphate moiety in light brown. The cleaved diphosphate is shown in orange. Hydrogen bonds and salt bridges are indicated by dashed lines in the same color-coding as the involved motifs. Gray lines represent the coordination sphere of the Mg^{2+}-ions. For clarity water molecules have been omitted. (B) View in panel A rotated by 30°. In addition to the motifs shown in panel A, the pyrophosphate sensor motif is depicted in pink.

Catalytic motifs of CotB2

Mg^{2+} coordination by the aspartate-rich and NSE motif

Instead of the classical DDxxD motif for class I TPS, CotB2 contains a modified ^{110}DDxD^{113} motif, which is neither characteristic for class I nor class II terpene synthases. In the literature, other TPSs with modified aspartate-rich motifs have been reported, encompassing selina-3,7(11)-diene synthase: ^{82}DDGYCE^{87} [57] and (+)-T-muurolol synthase: ^{83}DDEYCD^{88} [58]. It is commonly observed that the aspartate-rich motif resides at the lower part of α-helix D in the structure of class I TPSs. Interestingly, in CotB2, α-helix D is rather short as a proline residue, adjacent to the third aspartate of the aspartate-rich motif, introduces a kink. In the closed conformation D110 is directly involved in the coordination of Mg$^{2+}{}_A$ and Mg$^{2+}{}_C$, whereas D111 forms a salt bridge with R294 (Figure 5). A third Mg^{2+} (Mg$^{2+}{}_B$) is bound by the residues N220, S224 and E228 of the NSE motif [59], that is located on α-helix E of CotB2 (Figure 5). D113 is surface exposed and points away from the active site and hence is likely not be involved in catalysis. Latter structural findings are supported by site-directed mutagenesis studies [36,38], which showed that, whereas D110 and D111 are crucial for catalysis, the solvent exposed D113 is not. The here described structural situation is different compared to TPSs with a canonical DDXXD motif. For instance in the structure of epi-isozizaene synthase in complex with diphosphate, 3 Mg^{2+} ions and *N*-benzyl-*N*,*N*-diethylethanaminium [51], the α-helix is longer. Compared to CotB2, the first aspartate is as well involved in the coordination of Mg$^{2+}{}_A$ and Mg$^{2+}{}_C$ and the second aspartate forms a salt bridge with the RY-pair, but the third aspartate is involved in coordination of water molecules in the water network around the catalytic Mg^{2+} ions.

The pyrophosphate sensor motif

In a recent review, the presence of a "pyrophosphate sensor", an arginine 46 amino acid residues upstream of the NSE motif, was discussed as a universal feature of bacterial TPSs [11]. Structural information on selina-4(15),7(11)-diene synthase (SdS), revealed that upon substrate binding by an induced fit mechanism, R178 changes its side chain conformation thereby approaching and interacting with the diphosphate function of the substrate analogue and forming a salt bridge to the neighboring D181 [60]. Moreover, a conformational change of the kink of the α-helix G1 moves the catalytically important G182 (effector) towards the active site. Initiated by this induced fit mechanism, the active site is being closed and the Michaelis complex is formed. In CotB2, R177 represents the pyrophosphate sensor (Figure 5B) establishing identical interactions to

the diphosphate moiety of the substrate and to D180. In contrast to SdS, in CotB2 an additional residue is inserted between D180 and the effector G182. A comparison of the open and closed conformation of SdS revealed a large conformation change of the G1/G2 helix kink, which is not observed in the structures of CotB2 [36-38]. This might be explained by the observation that the pyrophosphate sensor of CotB2 R177 and D180 is already involved in a salt bridge in the open conformation.

The WXXXXXRY motif

Analyzing protein sequences of other bacterial TPSs for the presence of the RY pair, also called "basic pair" [61,62] and their flanking regions, led to the identification of a conserved tryptophan six amino acids upstream of the RY pair (Figure 6) [37]. In CotB2, residues of the motif are located at the end of the C-terminal core α-helix J. The RY pair is a conserved feature among most bacterial and several types of fungal TPSs [62]. It is not only important for active site closure by establishing a salt bridge, but moreover for interaction with the substrate. In the closed conformation of CotB2, R294 is involved in a bidentate salt bridge with D111 of the aspartate-rich motif, and two salt bridges to the diphosphate moiety derived from FGGDP (Figure 3 and Figure 5). Y295 establishes hydrogen bonds to the diphosphate moiety as well as to N220 of the NSE motif (Figure 5). In the closed conformation of the labdane-related diterpene synthase [63] the WXXXXXRY motif is present as well and adopts a nearly identical conformation as observed in the closed conformation of CotB2^{wt}·Mg^{2+}_3·F-Dola (PDB-ID 6GGI; [37]). Hence, the WXXXXXRY motif seems to be conserved not only sequence-based (Figure 6) but also structurally in other diterpene synthases. Hence, the discovery of this novel motif might help in the identification and functional assignment of novel TPSs.

The importance of the tryptophan residue in the WXXXXXRY motif of CotB2 was proven by mutation to glycine (W288G), where the product was changed to 3,7,18-dolabellatriene (**12**) [38]. If a less drastic mutation was introduced (W288F), preserving the aromatic character of the side chain, the product was not changed but the activity of CotB2 was reduced [37]. In other TPSs, a similar observation was made. Mutation of the corresponding residue W308F in the bacterial sesquiterpene synthase pentalenene synthase leads to a product mixture [64], whereas mutation of W273 in the plant-derived epi-aristolochene synthase resulted in total loss of enzymatic function [65]. Consequently, the aromatic character of the side chain in the WXXXXXRY motif is important for the propagation of the carbocation.

Mechanistic aspects of the cyclization reaction and point mutations

The floor of the active site is decorated by hydrophobic residues V80, F107, W109, F149, V150, I181, F185, M189, W186, L281, and W288 conferring the overall shape to the cavity that serves as a template for the binding of GGDP and the cyclization reaction. Based on ^{2}H- as well as ^{13}C-isotope labeling experiments a surprising reaction mechanism has been derived (Scheme 2) [35]. Strong support for the proposed reaction mechanism has been predominantly provided by site-directed mutagenesis of amino acids with an aromatic function [30,36-38] (Table 2 and Scheme 1), which significantly affect the product profile.

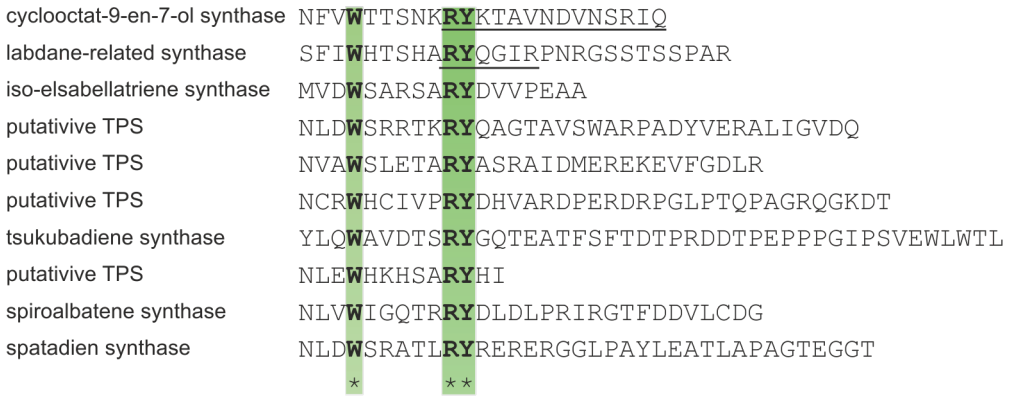


Figure 6: The WXXXXXRY motif in protein sequences of diterpene TPS from different bacteria. Highlighted is the WXXXXXRY motif in green. Underlined sequences refer to crystal structures of the respective diterpene TPS in the closed conformation, with a structured C-terminus. Cyclooctat-9-en-7-ol synthase (WP_093468823), *S. melanosporofaciens;* labdane-related diterpene synthase (WP_019525557), *Streptomyces* K155; iso-elsabellatriene synthase (WP_003963279), *Streptomyces clavuligerus*; terpene synthase (WP_046708564.1), *Streptomyces europaeiscabiei*; terpene synthase (WP_012394883), *Myobacterium marinum*; diterpene synthase (BAP82229), *Streptomyces* sp. ND90; tsukubadiene synthase (EIF90392), *Streptomyces tsukubaensis* NRRL 18488; terpene synthase (ZP_00085244), *Pseudomonas fluorescens* PfO-1; spiroalbatene synthase (WP_030426588.1), *Allokutzneria albata*; spatadien synthase (WP_095757924.1) *Streptomyces xinghaiensis*.

Table 2: CotB2 and its variants as well as the altered products.

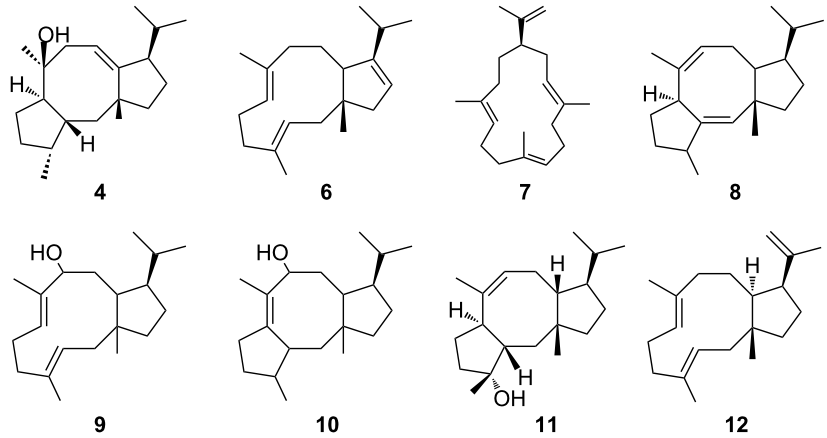
variant	product	reference
wild-type	cyclooctat-9-en-7-ol (**4**)	[31]
N103A	3,7,12-dolabellatriene (**6**)	[36]
F107A	*R*-cembrene A (**7**)	[30]
F107Y	cyclooctat-1,7-diene (**8**)	[30]
F107L	cyclooctat-9-en-7-ol (**4**) 3,7-dolabelladiene-9-ol (**9**) cyclooctat-6-en-8-ol (**10**)	[30,36]
F149L	cyclooctat-7-en-3-ol (**11**)	[30]
F185A	cyclooctat-9-en-7-ol (**4**) cyclooctat-6-en-8-ol (**10**)	[36]
W186L	cyclooctat-9-en-7-ol (**4**) cembrane A (**7**) 3,7,18-dolabellatriene (**12**)	[36]
W186F	cyclooctat-9-en-7-ol (**4**) cyclooctat-7-en-3-ol (**11**) 3,7-dolabelladiene-9-ol (**9**) cyclooctat-6-en-8-ol (**10**)	[36]
W186H	cyclooctat-7-en-3-ol (**11**) 3,7,18-dolabellatriene (**12**)	[36,38]
W288G	3,7,18-dolabellatriene (**12**)	[38]

Further support for the reaction mechanism was generated by gas-phase calculations [33,34] as well as in silico multiscale modeling [37], which suggest an active role of the enzyme during catalysis.

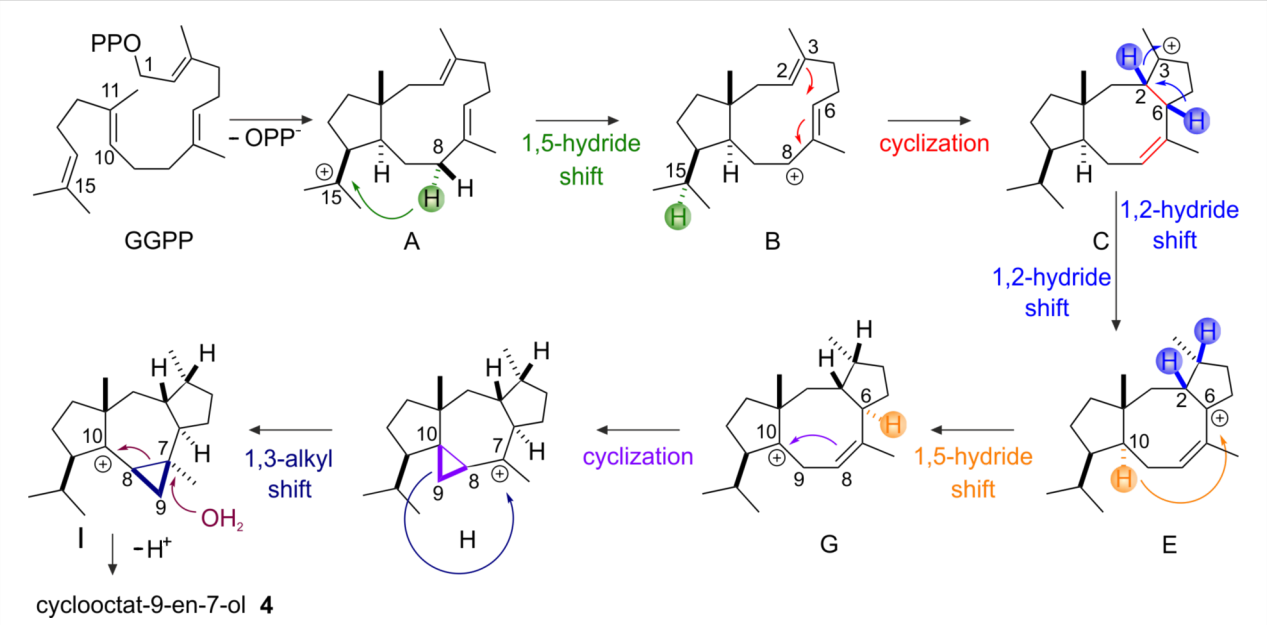
Cyclization is initiated by cleavage of the GGDP diphosphate moiety. After two consecutive cyclization reactions, a dolabellatrienyl cation (**A**) is generated, stabilized by π-cation interactions with W186 (Scheme 2 and Figure 7). Whereas mutation of this residue to amino acids with aromatic character mainly lead to different migration of the double bound and different hydroxylation pattern (Table 2 and Scheme 1). An exchange to leucine drastically changes the product to cembrane A (**7**) and 3,7,18-dolabellatriene **12** (Table 2 and Scheme 1) [36]. The cation migrates via a 1,5 hydride shift, as shown by deuterium labeling [33-35], to form the carbocation located at position C8 of dolabellatrienyl (**B**). By ring closure and formation of a novel C–C bond the tricyclic 5–8–5 ring system (**C**) is established. F107, F149, and the pyrophosphate moiety stabilize the carbocation at position C3 (Scheme 2 and Figure 7). F107 has been targeted by site-directed mutagenesis as well (Table 2 and Scheme 1) [30], leading to compounds *R*-cembrene A (**7**) and cyclooctat-1,7-diene (**8**). Now the cationic intermediate has two possibilities to react to **G**, either by a 1,3- and 1,5-hydride shift or by two 1,2-hydride shifts followed by a 1,5-hydride shift. The sequential 1,2-hydride shift (**C** to **E**) route was initially suggested by theoretical calculations, which supported the overall mechanism [33,34], and this was verified via isotope labeling [34]. Such series of two 1,2-hydride shifts have previously been demonstrated experimentally for tsukubadiene synthase [66]. Compound **E** is stabilized by F107 (Scheme 2 and Figure 7) and the backbone carbonyl of I181. Compound **G** is obtained by another 1,5-hydride shift as proven by isotope labeling [35]. The carbocation located at C10 is stabilized by N103 and T106. A mutation of N103 to alanine, bearing the possibility to stabilize the carbocation, results in 3,7,12-dolabellatriene **6** (Table 2 and Scheme 1). Finally, the cyclopropyl ring is opened by a 1,3-alkyl shift and a subsequent nucleophilic attack by a water molecule at C7 yields the product cycloocat-9-en-7-ol (**4**). Notably, no water has been observed in the structural investigations that could resume the nucleophilic attack.

Biotechnological applications to exploit the chemistry of CotB2

The mutants described in Table 2 are not only important for resolving structure–function relationships, they also convey the



Scheme 1: Overview of the altered product portfolio as a result of introduced point mutations in the active site of CotB2. For an overview of the point mutations see Table 2.



Scheme 2: Catalytic mechanism of CotB2, derived from isotope labeling experiments [34,35], density functional theory calculations [33] as well as QM/MM simulations [37].

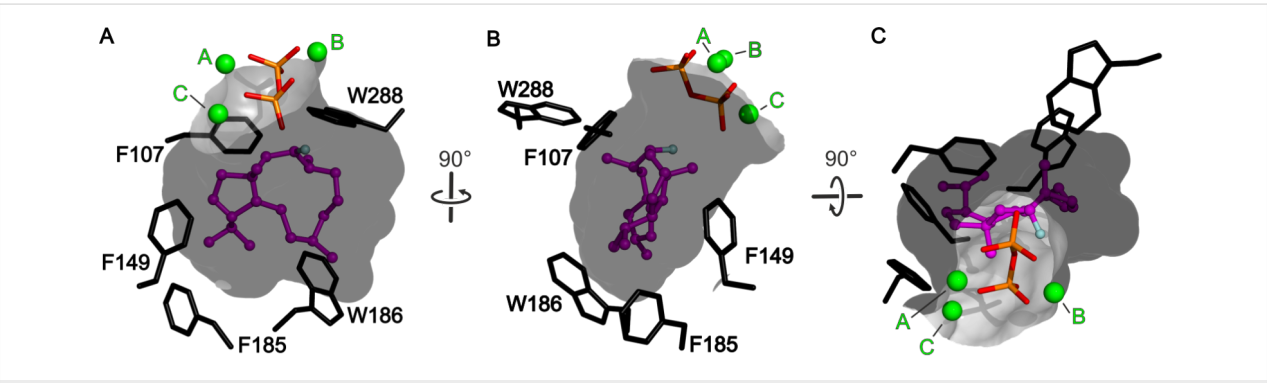
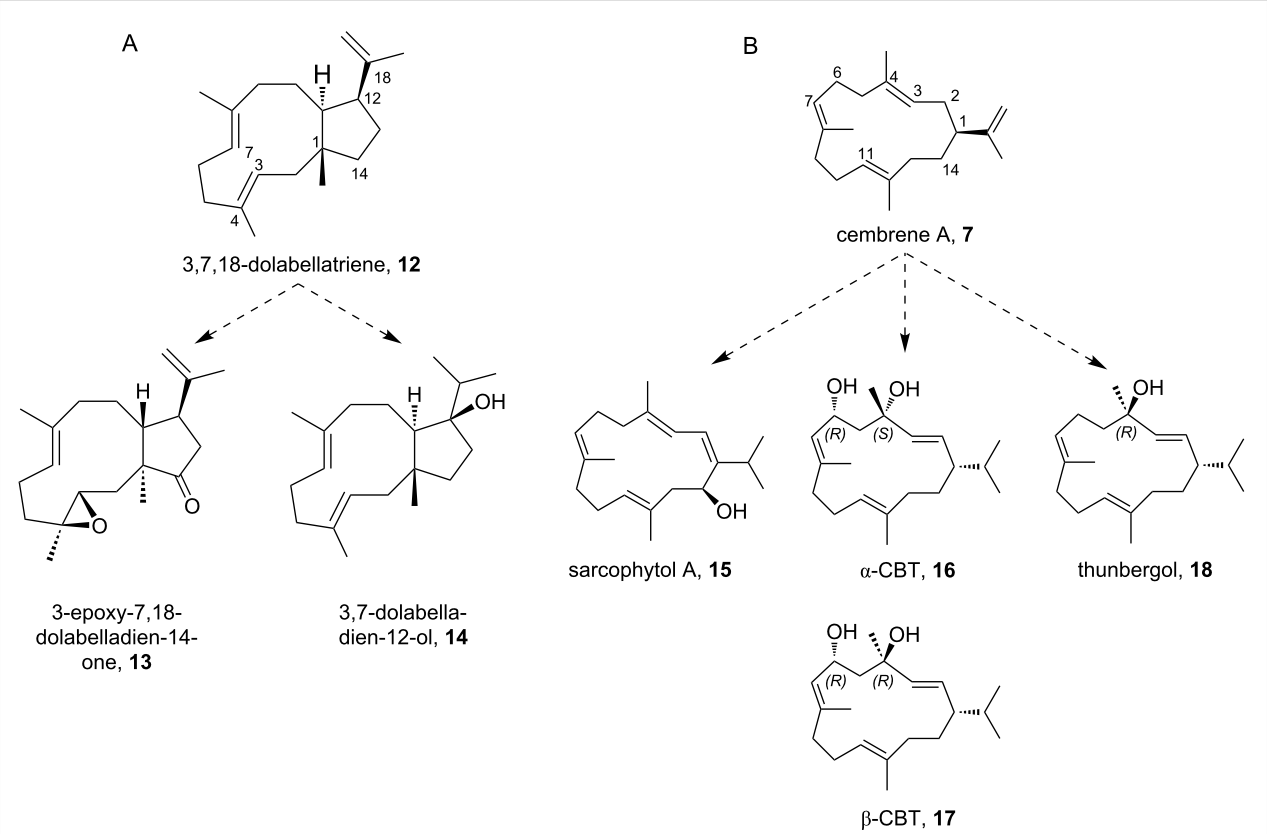


Figure 7: (A) The inner surface of the active site is shown in gray. The bound F-Dola reaction intermediate is shown in magenta with the fluorine atom colored in light blue. Aromatic residues are shown in black stick representation (B) View of panel A rotated by 90°. (C) View from the top into the active site of CotB2. View of panel A rotated by 90°.

conversion to carbon skeletons of other bioactive natural product classes. In light of this fact, the exchange of only one amino acid changes the product portfolio from cyclooctatin (**5**) belonging to the fusicoccane diterpene family to compounds of the dolabellane [67] or the cembranoid class [68].

The dolabellane diterpenes synthesized by the respective synthase mutants are missing the bond between C2 and C6 compared to cyclooctatin (**5**). Hence, they are comprising a 5,11-fused bicyclic skeleton. In contrast, cembranoids are monocyclic compounds consisting of a 14 membered ring structure. Common to all compounds available via the CotB2 mutants is the necessity of further decoration with functional groups, like hydroxylations or epoxidations, to induce the desired bioactivity (Scheme 3). Modification descriptions in the next chapters are composed according to the atom numbering in (Scheme 3). Lead structure for the dolabellane derivatives is 3,7,18-dolabellatriene (**12**) and cembrene A (**7**) for the cembranoid family, respectively (Scheme 3).

Dolabellanes comprise a diverse group of bioactive diterpenes, with the first compounds initially isolated from the mollusc *Dolabella californica* in 1977 [69]. Later, the product family was widened by molecules isolated from other marine organisms, like sponges, sea whips and the brown algae of this genus Dictyota [70]. Two modifications would transform 3,7,18-dolabellatriene (**12**, Scheme 3) into a potent cytotoxic molecule with activity towards murine leukemia cells or human non-small cell



Scheme 3: Variants of CotB2 open the route to a novel product portfolio with altered cyclic carbon skeletons, which can be converted into bioactive compounds by chemo-enzymatic methodologies. Modification descriptions are composed according to the atom numbering. (A) Lead structure for the dolabellane derivatives is 3,7,18-dolabellatriene (**12**) and (B) cembrene A (**7**) for the cembranoid family.

lung cancer with ED_{50} values of 6.5 and 16.7 μg mL^{-1}, respectively [71]. To obtain this interesting compound isolated from *Dictyota dichotoma*, C14 has to be modified with a keto function and the double bond between C3 and C4 has to be converted into an epoxide. Additionally, 3,7,18-dolabellatriene (**12**) possesses a mainly weak antibiotic activity against different types of *Staphylococcus aureus* strains (two epidemic MRSA strains, a macrolide-resistant variant and two multidrug-resistant efluxing strains) [72]. By contrast, 3,7,18-dolabellatriene (**12**, Scheme 3), only altered at C14 with a hydroxy group, showed a significant increase in potency (compound **14**). Here the MIC values were as low as 4 μg mL^{-1} for a macrolide-resistant and one multidrug-resistant effluxing variant, as well as 2 μg mL^{-1} for two epidemic MRSA and another multidrug-resistant effluxing strain.

The first bioactive molecule of the cembranoid family sarcophytol A (**15**), isolated in 1979 from the soft coral *Sarcophyton glaucum* [73], is structurally closely related to *R*-cembrene A (**7**, Scheme 3). In order to obtain sarcophytol A (**15**), two reactions have to be applied to compound **7**: First a base induced double bond shift to gain cembrene C and second a hydroxylation at C14. Sarcophytol A (**15**), is a promising target as it possesses inhibitory activity against potent tumor promotors, like teleocidin [74]. Interestingly, sarcophytol A (**15**) was already inhibitory at equimolar amounts in contrast to other natural antitumor promotors, like flavonoids or dihydroxycembranoids from tobacco leaves, which has to be applied at a factor of thousand or more compared to the promotor quantity to be reasonably active. Moreover, sarcophytol A (**15**) inhibits the expression of the tumor necrosis factor-α (TNF-α) mRNA and the release of TNF-α by BALB/3T3 cells [75]. This cancer prevention activity can be initiated at an IC_{50} of only 2.5 μM.

Other bioactive target molecules structurally accessible from compound **7** could be both epimers of (6*R*)-2,7,11-cembratriene-4,6-diol either 4S: α-CBT **16** or 4R: β-CBT **17** (Scheme 3). Both were isolated in 1985 from cigarette smoke condensate and identified as anticancer agents [76]. With quite similar potency (α-CBT: 25.2 μM and β-CBT: 21.9 μM [77]), they showed inhibition of the induction of Epstein–Barr virus early antigen by lymphoblastoid cancer cells. In the tobacco plant itself, both substances, in addition to being key flavor

ingredients [78], play a major role in the defence against insects, pathogenic microbes and herbivores [79]. Furthermore, β-CBT **17** showed additional antibiotic activity against multidrug resistant *S. aureus* strains, [80], as well as neuroprotective activity [81].

Interestingly, the β-CBT derivative missing the hydroxy group at C6 called thunbergol (**18**, Scheme 3), already inhibited the growth of two parasites *Trypanosoma brucei rhodesiense*, causing African sleeping sickness, as well as *Plasmodium falciparum* causing Malaria tropica, while having only minor human cytotoxicity [82]. Further, thunbergol (**18**) is reported to repel aphids within 48 h by 70%, if wheat seedlings are topically treated with a 0.25% (w/w) solution in ethyl acetate, compared to untreated plants [25]. During application in agriculture a negative impact on useful insects, like bees, should be avoided. A hint in this direction could be that in an in vitro experiment with *Spodoptera frugiperda* insect cells, a high resilience could be detected with an IC_{50} of 68 μM. Furthermore, Gram-positive bacteria are significantly growth inhibited through thunbergol (**18**) exposure with an IC_{50} for *Bacillus subtilis* of 9 μM and 10 μM for *Micrococcus luteus,* respectively [25].

For the sustainable, high yield production of bioactive diterpenoids various aspects have to be considered. One key issue in yielding high recombinant terpene production titers, are metabolic bottlenecks in the precursor supply, which have to be circumvented [83]. In that regard, there are different ways of enhancing *E. coli* host productivity. First, the general production pathway is usually genetically established. Choosing a plasmid-based production pathway has the advantage, that single genes can be exchanged or added quite rapidly [84]. Furthermore, decreasing the metabolic burden of the plasmid construct on the native host metabolism [85] can be achieved by using polycistronic operons to reduce the amount of plasmid in a cell. Additionally, computer aided fine-tuning [86] of transcription rates by promotor [24] and RBS [87] variations will further enhance the production rate [25]. Alternatively, permanent integration of the heterologous genes into the host genome is an alternative strategy to circumvent metabolic stress by antibiotics, which are required to maintain a plasmid in the production host [28]. Nevertheless, genomic integration procedures into the *E. coli* genome are time-consuming and have the disadvantage of having only a single copy in the host genome. Thus, expression and subsequently production rates of a previously optimized plasmid based system are not transferable to a genomically integrated system. In that respect, each operon has to be optimized de novo, whereby also the loci of integration have a severe impact on the expression rates of genome integrated heterologous gene material [88-90].

Perspectives

CotB2 is an exciting example of how nature evolved an enzyme to perform very sophisticated chemistry. Mechanistically, CotB2 is very well understood, based on data from various different disciplines [30,31,33,35-38]. TPSs play an active role in all steps from the initialization of the diphosphate cleavage, and end with the nucleophilic attack of a water molecule. Yet, there is a dogmatic dispute in the TPS community about the respective roles of the protein acting as a scaffold and the "inherent" carbocation reactivity [91] in orchestrating an enzyme specific reaction cascade. With regard to the latter, the "inherent reactivity" of carbocations is no doubt an important ingredient in terpene biosynthesis. The enzyme could be understood as passive catalyst, essentially chaperoning the intermediates during the reaction cascade. It is clear that much of terpene biosynthesis can be understood by this concept. "Inherent reactivity" largely relies on theoretical calculations performed in the gas phase [92,93]. The latter approach per se disregards the role of the protein scaffold during catalysis. However, there are many examples of mono-, sesqui-, and diterpene synthases, where the contribution of the diphosphate and selected amino acid residues, which decorate the active site, on carbocation stability has been demonstrated [94-98]. These studies reveal a significant effect of individual active site moieties on TPS reactions in respect of product formation or alteration of the product profile. Additionally, numerous theoretical studies have emphasized the role of the enzyme environment in guiding the reaction cascade [94,99-102]. In case of CotB2, mutagenesis studies of plasticity residues of CotB2 [30,36-38] have been demonstrated to drastically change the propagation of the carbocations and consequently alter the product portfolio (Table 2 and Scheme 1).

Notably, in the crystal structures of CotB2 (Table 1) [36-38], we perceive that the amino acid side chains with an aromatic character in the active site (Figure 7) frequently adopt energetically unfavorable side chain conformations. Moreover, the crystal structures of the $CotB2^{F107A}$ [37] and $CotB2^{F149L}$ [38] variants revealed that other aromatic amino acid side chains adopt different side chain conformations compared to the structure of $CotB2^{wt}$. This hints to long-range effects of the introduced mutation on other amino acid side chains in the active cavity. We are convinced that this plasticity of side chains in the active site of CotB2 plays not only an important role in stabilization and hence propagation of carbocations.

In future experiments, it would be interesting to investigate the influence of double or triple mutations within the active site. Moreover, it would be interesting to prepare a CotB2 variant with a canonical aspartate-rich motif and to study the influence on catalysis, whether the product portfolio is affected or not.

Since until now no water molecule could be located in the electron density maps, in terms of catalysis the question remains which entity performs the nucleophilic attack at position C7 to give the final product cyclocotat-9-en-7-ol (**12**).

Another direction, to expand the chemical space, is the exploitation of TPSs to generate novel compounds that could be further functionalized by classical organic chemistry. Examples of this include epoxidation resulting in 3,4-epoxy-7,18-dolabelladien-14-one (**13**) or hydroboration of 3,7,18-dolabellatriene (**12**, Scheme 1) that has been previously biotechnologically manufactured using CotB2^{W288G} [103]. Another successful example is the oxidative transformation of cattleyene and phomopsen [104]. Yet another approach is the use of heteroatom-modified farnesyl diphosphates that could be still cyclized by TPSs yielding unnatural terpenoids [105].

Acknowledgements

R. Driller was supported by Elsa-Neumann and Nüsslein-Volhard stipends. D. T. Major acknowledges support from the Israel Science Foundation (Grant #2146/15 and 1683/18). T. Brück gratefully acknowledges funding by the Werner Siemens foundation for establishing the field of Synthetic Biotechnology at the Technical University of Munich (TUM). Additionally, the German Federal Ministry of Education and Research supported TB and DG through the project OMCBP (Grant #031A276A), while TB and MF were supported by the project SysBioTerp (Grant #031A305A). Furthermore, TB and NM acknowledge funding by the Bavarian Ministry of Economic Affairs, Regional Development and Energy for the bioinsecticide project (Grant #1340/68351/13/2013). We are grateful to M. Wahl for continuous encouragement and support.

ORCID® iDs

Ronja Driller - https://orcid.org/0000-0001-8834-9087
Daniel Garbe - https://orcid.org/0000-0001-7203-8751
Monika Fuchs - https://orcid.org/0000-0002-6954-2639
Dan Thomas Major - https://orcid.org/0000-0002-9231-0676
Thomas Brück - https://orcid.org/0000-0002-2113-6957
Bernhard Loll - https://orcid.org/0000-0001-7928-4488

References

1. Christianson, D. W. *Chem. Rev.* **2006,** *106,* 3412–3442. doi:10.1021/cr050286w
2. Christianson, D. W. *Science* **2007,** *316,* 60–61. doi:10.1126/science.1141630
3. Christianson, D. W. *Chem. Rev.* **2017,** *117,* 11570–11648. doi:10.1021/acs.chemrev.7b00287
4. Mayer, A. M. S.; Gustafson, K. R. *Int. J. Cancer* **2003,** *105,* 291–299. doi:10.1002/ijc.11080
5. Ebel, R. *Mar. Drugs* **2010,** *8,* 2340–2368. doi:10.3390/md8082340
6. Elissawy, A.; El-Shazly, M.; Ebada, S.; Singab, A.; Proksch, P. *Mar. Drugs* **2015,** *13,* 1966–1992. doi:10.3390/md13041966
7. Shirai, M.; Okuda, M.; Motohashi, K.; Imoto, M.; Furihata, K.; Matsuo, Y.; Katsuta, A.; Shizuri, Y.; Seto, H. *J. Antibiot.* **2010,** *63,* 245–250. doi:10.1038/ja.2010.30
8. Dickschat, J. S. *Angew. Chem., Int. Ed.* **2019,** in press. doi:10.1002/anie.201905312
9. Brück, T.; Kourist, R.; Loll, B. *ChemCatChem* **2014,** *6,* 1142–1165. doi:10.1002/cctc.201300733
10. Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2016,** *79,* 629–661. doi:10.1021/acs.jnatprod.5b01055
11. Dickschat, J. S. *Nat. Prod. Rep.* **2016,** *33,* 87–110. doi:10.1039/c5np00102a
12. Poulter, C. D.; Argyle, J. C.; Mash, E. A. *J. Biol. Chem.* **1978,** *253,* 7227–7233.
13. Maimone, T. J.; Baran, P. S. *Nat. Chem. Biol.* **2007,** *3,* 396–407. doi:10.1038/nchembio.2007.1
14. Christianson, D. W. *Curr. Opin. Chem. Biol.* **2008,** *12,* 141–150. doi:10.1016/j.cbpa.2007.12.008
15. Sacchettini, J. C.; Poulter, C. D. *Science* **1997,** *277,* 1788–1789. doi:10.1126/science.277.5333.1788
16. Cane, D. E. *Chem. Rev.* **1990,** *90,* 1089–1103. doi:10.1021/cr00105a002
17. Lesburg, C. A.; Caruthers, J. M.; Paschall, C. M.; Christianson, D. W. *Curr. Opin. Struct. Biol.* **1998,** *8,* 695–703. doi:10.1016/s0959-440x(98)80088-2
18. Wendt, K. U.; Schulz, G. E. *Structure* **1998,** *6,* 127–133. doi:10.1016/s0969-2126(98)00015-x
19. Hamano, Y.; Kuzuyama, T.; Itoh, N.; Furihata, K.; Seto, H.; Dairi, T. *J. Biol. Chem.* **2002,** *277,* 37098–37104. doi:10.1074/jbc.m206382200
20. Smanski, M. J.; Yu, Z.; Casper, J.; Lin, S.; Peterson, R. M.; Chen, Y.; Wendt-Pienkowski, E.; Rajski, S. R.; Shen, B. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108,* 13498–13503. doi:10.1073/pnas.1106919108
21. Chen, M.; Harris, G. G.; Pemberton, T. A.; Christianson, D. W. *Curr. Opin. Struct. Biol.* **2016,** *41,* 27–37. doi:10.1016/j.sbi.2016.05.010
22. Aaron, J. A.; Christianson, D. W. *Pure Appl. Chem.* **2010,** *82,* 1585–1597. doi:10.1351/pac-con-09-09-37
23. Smanski, M. J.; Peterson, R. M.; Huang, S.-X.; Shen, B. *Curr. Opin. Chem. Biol.* **2012,** *16,* 132–141. doi:10.1016/j.cbpa.2012.03.002
24. Hirte, M.; Mischko, W.; Kemper, K.; Röhrer, S.; Huber, C.; Fuchs, M.; Eisenreich, W.; Minceva, M.; Brück, T. B. *Green Chem.* **2018,** *20,* 5374–5384. doi:10.1039/c8gc03126f
25. Mischko, W.; Hirte, M.; Roehrer, S.; Engelhardt, H.; Mehlmer, N.; Minceva, M.; Brück, T. *Green Chem.* **2018,** *20,* 2637–2650. doi:10.1039/c8gc00434j
26. Kemper, K.; Hirte, M.; Reinbold, M.; Fuchs, M.; Brück, T. *Beilstein J. Org. Chem.* **2017,** *13,* 845–854. doi:10.3762/bjoc.13.85
27. Hirte, M.; Meese, N.; Mertz, M.; Fuchs, M.; Brück, T. B. *Front. Chem. (Lausanne, Switz.)* **2018,** *6,* 101. doi:10.3389/fchem.2018.00101
28. Ajikumar, P. K.; Xiao, W.-H.; Tyo, K. E. J.; Wang, Y.; Simeon, F.; Leonard, E.; Mucha, O.; Phon, T. H.; Pfeifer, B.; Stephanopoulos, G. *Science* **2010,** *330,* 70–74. doi:10.1126/science.1191652
29. Anthony, J. R.; Anthony, L. C.; Nowroozi, F.; Kwon, G.; Newman, J. D.; Keasling, J. D. *Metab. Eng.* **2009,** *11,* 13–19. doi:10.1016/j.ymben.2008.07.007

30. Görner, C.; Häuslein, I.; Schrepfer, P.; Eisenreich, W.; Brück, T. *ChemCatChem* **2013,** *5,* 3289–3298. doi:10.1002/cctc.201300285
31. Kim, S.-Y.; Zhao, P.; Igarashi, M.; Sawa, R.; Tomita, T.; Nishiyama, M.; Kuzuyama, T. *Chem. Biol.* **2009,** *16,* 736–743. doi:10.1016/j.chembiol.2009.06.007
32. de Boer, A. H.; de Vries-van Leeuwen, I. J. *Trends Plant Sci.* **2012,** *17,* 360–368. doi:10.1016/j.tplants.2012.02.007
33. Hong, Y. J.; Tantillo, D. J. *Org. Biomol. Chem.* **2015,** *13,* 10273–10278. doi:10.1039/c5ob01785h
34. Sato, H.; Teramoto, K.; Masumoto, Y.; Tezuka, N.; Sakai, K.; Ueda, S.; Totsuka, Y.; Shinada, T.; Nishiyama, M.; Wang, C.; Kuzuyama, T.; Uchiyama, M. *Sci. Rep.* **2016,** *5,* 18471. doi:10.1038/srep18471
35. Meguro, A.; Motoyoshi, Y.; Teramoto, K.; Ueda, S.; Totsuka, Y.; Ando, Y.; Tomita, T.; Kim, S.-Y.; Kimura, T.; Igarashi, M.; Sawa, R.; Shinada, T.; Nishiyama, M.; Kuzuyama, T. *Angew. Chem., Int. Ed.* **2015,** *54,* 4353–4356. doi:10.1002/anie.201411923
36. Tomita, T.; Kim, S.-Y.; Teramoto, K.; Meguro, A.; Ozaki, T.; Yoshida, A.; Motoyoshi, Y.; Mori, N.; Ishigami, K.; Watanabe, H.; Nishiyama, M.; Kuzuyama, T. *ACS Chem. Biol.* **2017,** *12,* 1621–1628. doi:10.1021/acschembio.7b00154
37. Driller, R.; Janke, S.; Fuchs, M.; Warner, E.; Mhashal, A. R.; Major, D. T.; Christmann, M.; Bruck, T.; Loll, B. *Nat. Commun.* **2018,** *9,* 3971. doi:10.1038/s41467-018-06325-8
38. Janke, R.; Görner, C.; Hirte, M.; Brück, T.; Loll, B. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2014,** *70,* 1528–1537. doi:10.1107/s1399004714005513
39. Driller, R. Structural studies of biotechnologically relevant enzymes. Ph.D. Thesis, Freie Universität Berlin, Berlin, Germany, 2018.
40. Das, S.; Shimshi, M.; Raz, K.; Nitoker Eliaz, N.; Mhashal, A. R.; Ansbacher, T.; Major, D. T. *J. Chem. Theory Comput.* **2019,** *15,* 5116–5134. doi:10.1021/acs.jctc.9b00366
41. Aoyagi, T.; Aoyama, T.; Kojima, F.; Hattori, S.; Honma, Y.; Hamada, M.; Takeuchi, T. *J. Antibiot.* **1992,** *45,* 1587–1591. doi:10.7164/antibiotics.45.1587
42. Supong, K.; Sripreechasak, P.; Tanasupawat, S.; Danwisetkanjana, K.; Rachtawee, P.; Pittayakhajonwut, P. *Appl. Microbiol. Biotechnol.* **2017,** *101,* 533–543. doi:10.1007/s00253-016-7804-1
43. Toyomasu, T.; Tsukahara, M.; Kaneko, A.; Niida, R.; Mitsuhashi, W.; Dairi, T.; Kato, N.; Sassa, T. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104,* 3084–3088. doi:10.1073/pnas.0608426104
44. Subbarao, G. V.; Nakahara, K.; Hurtado, M. P.; Ono, H.; Moreta, D. E.; Salcedo, A. F.; Yoshihashi, A. T.; Ishikawa, T.; Ishitani, M.; Ohnishi-Kameyama, M.; Yoshida, M.; Rondon, M.; Rao, I. M.; Lascano, C. E.; Berry, W. L.; Ito, O. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106,* 17302–17307. doi:10.1073/pnas.0903694106
45. Rasoamiaranjanahary, L.; Marston, A.; Guilet, D.; Schenk, K.; Randimbivololona, F.; Hostettmann, K. *Phytochemistry* **2003,** *62,* 333–337. doi:10.1016/s0031-9422(02)00551-4
46. Minami, A.; Ozaki, T.; Liu, C.; Oikawa, H. *Nat. Prod. Rep.* **2018,** *35,* 1330–1346. doi:10.1039/c8np00026c
47. Whittington, D. A.; Wise, M. L.; Urbansky, M.; Coates, R. M.; Croteau, R. B.; Christianson, D. W. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99,* 15375–15380. doi:10.1073/pnas.232591099
48. Rynkiewicz, M. J.; Cane, D. E.; Christianson, D. W. *Proc. Natl. Acad. Sci. U. S. A.* **2001,** *98,* 13543–13548. doi:10.1073/pnas.231313098
49. Tarshis, L. C.; Yan, M.; Poulter, C. D.; Sacchettini, J. C. *Biochemistry* **1994,** *33,* 10871–10877. doi:10.1021/bi00202a004
50. Shishova, E. Y.; Di Costanzo, L.; Cane, D. E.; Christianson, D. W. *Biochemistry* **2007,** *46,* 1941–1951. doi:10.1021/bi0622524
51. Aaron, J. A.; Lin, X.; Cane, D. E.; Christianson, D. W. *Biochemistry* **2010,** *49,* 1787–1797. doi:10.1021/bi902088z
52. Köksal, M.; Hu, H.; Coates, R. M.; Peters, R. J.; Christianson, D. W. *Nat. Chem. Biol.* **2011,** *7,* 431–433. doi:10.1038/nchembio.578
53. Köksal, M.; Jin, Y.; Coates, R. M.; Croteau, R.; Christianson, D. W. *Nature* **2011,** *469,* 116–120. doi:10.1038/nature09628
54. Miller, D. J.; Yu, F.; Allemann, R. K. *ChemBioChem* **2007,** *8,* 1819–1825. doi:10.1002/cbic.200700219
55. Jin, Y.; Williams, D. C.; Croteau, R.; Coates, R. M. *J. Am. Chem. Soc.* **2005,** *127,* 7834–7842. doi:10.1021/ja050592r
56. Shishova, E. Y.; Yu, F.; Miller, D. J.; Faraldos, J. A.; Zhao, Y.; Coates, R. M.; Allemann, R. K.; Cane, D. E.; Christianson, D. W. *J. Biol. Chem.* **2008,** *283,* 15431–15439. doi:10.1074/jbc.m800659200
57. Arigoni, D. *Pure Appl. Chem.* **1975,** *41,* 219–245. doi:10.1351/pac197541010219
58. Hu, Y.; Chou, W. K. W.; Hopson, R.; Cane, D. E. *Chem. Biol.* **2011,** *18,* 32–37. doi:10.1016/j.chembiol.2010.11.008
59. Cane, D. E.; Kang, I. *Arch. Biochem. Biophys.* **2000,** *376,* 354–364. doi:10.1006/abbi.2000.1734
60. Baer, P.; Rabe, P.; Fischer, K.; Citron, C. A.; Klapschinski, T. A.; Groll, M.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2014,** *53,* 7652–7656. doi:10.1002/anie.201403648
61. Rabe, P.; Schmitz, T.; Dickschat, J. S. *Beilstein J. Org. Chem.* **2016,** *12,* 1839–1850. doi:10.3762/bjoc.12.173
62. Vedula, L. S.; Rynkiewicz, M. J.; Pyun, H. J.; Coates, R. M.; Cane, D. E.; Christianson, D. W. *Biochemistry* **2005,** *44,* 6153–6163. doi:10.1021/bi050059o
63. Centeno-Leija, S.; Tapia-Cabrera, S.; Guzmán-Trampe, S.; Esquivel, B.; Esturau-Escofet, N.; Tierrafría, V. H.; Rodríguez-Sanoja, R.; Zárate-Romero, A.; Stojanoff, V.; Rudiño-Piñera, E.; Sánchez, S.; Serrano-Posada, H. *J. Struct. Biol.* **2019,** *207,* 29–39. doi:10.1016/j.jsb.2019.04.010
64. Seemann, M.; Zhai, G.; de Kraker, J.-W.; Paschall, C. M.; Christianson, D. W.; Cane, D. E. *J. Am. Chem. Soc.* **2002,** *124,* 7681–7689. doi:10.1021/ja026058q
65. Starks, C. M.; Back, K.; Chappell, J.; Noel, J. P. *Science* **1997,** *277,* 1815–1820. doi:10.1126/science.277.5333.1815
66. Rabe, P.; Rinkel, J.; Dolja, E.; Schmitz, T.; Nubbemeyer, B.; Luu, T. H.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2017,** *56,* 2776–2779. doi:10.1002/anie.201612439
67. Chen, J.; Li, H.; Zhao, Z.; Xia, X.; Li, B.; Zhang, J.; Yan, X. *Mar. Drugs* **2018,** *16,* 159. doi:10.3390/md16050159
68. Yang, B.; Liu, J.; Wang, J.; Liao, S.; Liu, Y. Cytotoxic Cembrane Diterpenoids. In *Handbook of Anticancer Drugs from Marine Origin;* Kim, S.-K., Ed.; Springer International Publishing: Cham, Switzerland, 2015; pp 649–672. doi:10.1007/978-3-319-07145-9_30
69. Ireland, C.; Faulkner, D. J. *J. Org. Chem.* **1977,** *42,* 3157–3162. doi:10.1021/jo00439a010
70. Cai, X.-H.; Wang, Y.-Y.; Zhao, P.-J.; Li, Y.; Luo, X.-D. *Phytochemistry* **2010,** *71,* 1020–1024. doi:10.1016/j.phytochem.2010.03.005
71. Bouaïcha, N.; Tringali, C.; Pesando, D.; Malléa, M.; Roussakis, C.; Verbist, J. *Planta Med.* **1993,** *59,* 256–258. doi:10.1055/s-2006-959663
72. Ioannou, E.; Quesada, A.; Rahman, M. M.; Gibbons, S.; Vagias, C.; Roussis, V. *J. Nat. Prod.* **2011,** *74,* 213–222. doi:10.1021/np1006586

73. Kobayashi, M.; Nakagawa, T.; Mitsuhashi, H. *Chem. Pharm. Bull.* **1979,** *27,* 2382–2387. doi:10.1248/cpb.27.2382
74. Fujiki, H.; Suganuma, M.; Suguri, H.; Yoshizawa, S.; Takagi, K.; Kobayashi, M. *J. Cancer Res. Clin. Oncol.* **1989,** *115,* 25–28. doi:10.1007/bf00391595
75. Suganuma, M.; Okabe, S.; Sueoka, E.; Iida, N.; Komori, A.; Kim, S. J.; Fujiki, H. *Cancer Res.* **1996,** *56,* 3711–3715.
76. Saito, Y.; Takizawa, H.; Konishi, S.; Yoshida, D.; Mizusaki, S. *Carcinogenesis* **1985,** *6,* 1189–1194. doi:10.1093/carcin/6.8.1189
77. Saito, Y.; Nishino, H.; Yoshida, D.; Mizusaki, S.; Ohnishi, A. *Oncology* **1988,** *45,* 122–126. doi:10.1159/000226545
78. Wahlberg, I.; Eklund, A. M. *Dev. Food Sci.* **1994,** *35,* 449–462.
79. Wagner, G. J. *Plant Physiol.* **1991,** *96,* 675–679. doi:10.1104/pp.96.3.675
80. Aqil, F.; Zahin, M.; El Sayed, K. A.; Ahmad, I.; Orabi, K. Y.; Arif, J. M. *Drug Chem. Toxicol.* **2011,** *34,* 167–179. doi:10.3109/01480545.2010.494669
81. Martins, A. H.; Hu, J.; Xu, Z.; Mu, C.; Alvarez, P.; Ford, B. D.; El Sayed, K.; Eterovic, V. A.; Ferchmin, P. A.; Hao, J. *Neuroscience* **2015,** *291,* 250–259. doi:10.1016/j.neuroscience.2015.02.001
82. Thao, N.; Luyen, B.; Brun, R.; Kaiser, M.; Van Kiem, P.; Van Minh, C.; Schmidt, T.; Kang, J.; Kim, Y. *Molecules* **2015,** *20,* 12459–12468. doi:10.3390/molecules200712459
83. Ajikumar, P. K.; Tyo, K.; Carlsen, S.; Mucha, O.; Phon, T. H.; Stephanopoulos, G. *Mol. Pharmaceutics* **2008,** *5,* 167–190. doi:10.1021/mp700151b
84. Shetty, R. P.; Endy, D.; Knight, T. F., Jr. *J. Biol. Eng.* **2008,** *2,* 5. doi:10.1186/1754-1611-2-5
85. Borkowski, O.; Ceroni, F.; Stan, G.-B.; Ellis, T. *Curr. Opin. Microbiol.* **2016,** *33,* 123–130. doi:10.1016/j.mib.2016.07.009
86. Salis, H. M. *Methods Enzymol.* **2011,** *498,* 19–42. doi:10.1016/b978-0-12-385120-8.00002-4
87. Zelcbuch, L.; Antonovsky, N.; Bar-Even, A.; Levin-Karp, A.; Barenholz, U.; Dayagi, M.; Liebermeister, W.; Flamholz, A.; Noor, E.; Amram, S.; Brandis, A.; Bareia, T.; Yofe, I.; Jubran, H.; Milo, R. *Nucleic Acids Res.* **2013,** *41,* e98. doi:10.1093/nar/gkt151
88. Englaender, J. A.; Jones, J. A.; Cress, B. F.; Kuhlman, T. E.; Linhardt, R. J.; Koffas, M. A. G. *ACS Synth. Biol.* **2017,** *6,* 710–720. doi:10.1021/acssynbio.6b00350
89. Juhas, M.; Ajioka, J. W. *Microb. Biotechnol.* **2015,** *8,* 726–738. doi:10.1111/1751-7915.12296
90. Juhas, M.; Ajioka, J. W. *PLoS One* **2015,** *10,* e0123007. doi:10.1371/journal.pone.0123007
91. Tantillo, D. J. *Angew. Chem., Int. Ed.* **2017,** *56,* 10040–10045. doi:10.1002/anie.201702363
92. Hare, S. R.; Tantillo, D. J. *Beilstein J. Org. Chem.* **2016,** *12,* 377–390. doi:10.3762/bjoc.12.41
93. Tantillo, D. J. *Nat. Prod. Rep.* **2011,** *28,* 1035–1053. doi:10.1039/c1np00006c
94. Dixit, M.; Weitman, M.; Gao, J.; Major, D. T. *ACS Catal.* **2017,** *7,* 812–818. doi:10.1021/acscatal.6b02584
95. Weitman, M.; Major, D. T. *J. Am. Chem. Soc.* **2010,** *132,* 6349–6360. doi:10.1021/ja910134x
96. Xu, M.; Wilderman, P. R.; Peters, R. J. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104,* 7397–7401. doi:10.1073/pnas.0611454104
97. Leferink, N. G. H.; Ranaghan, K. E.; Karuppiah, V.; Currin, A.; van der Kamp, M. W.; Mulholland, A. J.; Scrutton, N. S. *ACS Catal.* **2018,** *8,* 3780–3791. doi:10.1021/acscatal.8b00692
98. Miller, D. J.; Allemann, R. K. *Nat. Prod. Rep.* **2012,** *29,* 60–71. doi:10.1039/c1np00060h
99. Ansbacher, T.; Freud, Y.; Major, D. T. *Biochemistry* **2018,** *57,* 3773–3779. doi:10.1021/acs.biochem.8b00452
100. Freud, Y.; Ansbacher, T.; Major, D. T. *ACS Catal.* **2017,** *7,* 7653–7657. doi:10.1021/acscatal.7b02824
101. Major, D. T.; Freud, Y.; Weitman, M. *Curr. Opin. Chem. Biol.* **2014,** *21,* 25–33. doi:10.1016/j.cbpa.2014.03.010
102. Major, D. T. *ACS Catal.* **2017,** *7,* 5461–5465. doi:10.1021/acscatal.7b01328
103. Görner, C.; Hirte, M.; Huber, S.; Schrepfer, P.; Brück, T. *Front. Microbiol.* **2015,** *6,* 1115. doi:10.3389/fmicb.2015.01115
104. Rinkel, J.; Steiner, S. T.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2019,** *58,* 9230–9233. doi:10.1002/anie.201902950
105. Oberhauser, C.; Harms, V.; Seidel, K.; Schröder, B.; Ekramzadeh, K.; Beutel, S.; Winkler, S.; Lauterbach, L.; Dickschat, J. S.; Kirschning, A. *Angew. Chem., Int. Ed.* **2018,** *57,* 11802–11806. doi:10.1002/anie.201805526

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Synthetic terpenoids in the world of fragrances: Iso E Super® is the showcase

Alexey Stepanyuk and Andreas Kirschning*

Review

Open Access

Address:
Institute of Organic Chemistry and Center of Biomolecular Drug Research (BMWZ), Leibniz Universität Hannover, Schneiderberg 1b, 30167 Hannover, Germany

Email:
Andreas Kirschning* - andreas.kirschning@oci.uni-hannover.de

* Corresponding author

Keywords:
asymmetric synthesis; fragrances; odorants; sandalwood; scents; terpenes; terpenoids

Beilstein J. Org. Chem. **2019,** *15,* 2590–2602.
doi:10.3762/bjoc.15.252

Received: 22 July 2019
Accepted: 26 September 2019
Published: 31 October 2019

This article is part of the thematic issue "Terpenes".

Guest Editor: J. S. Dickschat

Abstract

The history of fragrances is closely associated with the chemistry of terpenes and terpenoids. For thousands of years mankind mainly used plant extracts to collect ingredients for the creation of perfumes. Many of these extracts contain complex mixtures of terpenes, that show distinct olfactoric properties as pure compounds. When organic synthesis appeared on the scene, the portfolio of new scents increased either in order to substitute natural fragrances without change of olfactoric properties or to broaden the scope of scents. This short review describes the story of the most successful synthetic fragrance ever which is called Iso E Super® as it is an ingredient in a large number of perfumes with varying percentages and is the first example being used as a pure fragrance. Structurally, it is related to natural terpenes like many other synthetic fragrances. And indeed, the story began with a classic in the field of fragrances, the natural product ionone.

Review

"Iso E Super® is to perfume what Tango Nuevo is to Tango Argentino" [1]

Introduction – classical terpenes in perfumes

Perfumes (Latin "per fumus", which means "through smoke") have accompanied mankind for thousands of years dating back well before biblical times [2,3]. Plants and resins served as source for perfumes after alcoholic extraction. These extracts were not only used as fragrances but also as medicine (aqua mirabilis), aphrodisiac and elixir of life (aquavitae).

In 1882 *'Fougere Royale'* was created, a composition of coumarin, oak moss, geranium and bergamot, commercially launched by Houbigant [4]. The major constituents of geranium oil include myrcene (**1**), menthone (**2**), α-pinene (**3**), geraniol (**4a**), geranyl acetate (**4b**), geranyl butyrate (**4c**), citronellol (**5**), limonene (**6**) and linalool (**9a**). As for the bergamot oil monoterpenes limonene (**6**, 37%), γ-terpinene (**7**, 7%), β-pinene (**8**, 6%), linalool (**9a**, 9%) and linalyl acetate (**9b**, 30%) are key ingredients (Figure 1). The ratio of (*R*)-linalool and (*R*)-linalyl acetate (commonly >99.3% ee) is one of the

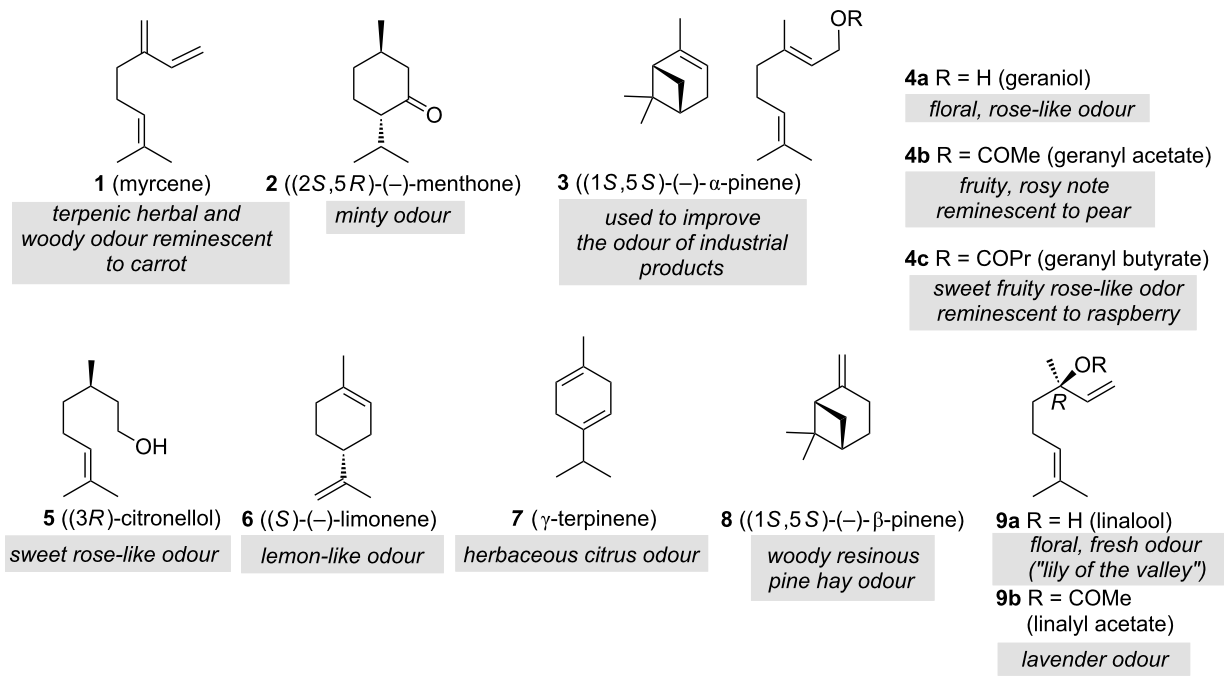


Figure 1: Terpene constituents **1**–**9** found in geranium and bergamot oils and specified odours of individual components. Perfumes are composed of compounds that are perceived immediately (top notes), that form the principal bouquet (heart notes) and compounds that are mainly perceived in the dry-out (bottom notes) [6]. Classification of 1. top notes: a) citrus, b) aldehydic, c) fruity, d) green and e) herbaceous/herbal, 2. heart notes: a) floral light, b) floral green, c) floral fresh, d), floral fruity, e) floral heavy and f) floral woody and 3. bottom notes: a) aromatic, b) balsamic, c) moss/leather/animalic, d) musk, e) amber and f) wood.

quality indices as it affects the aroma of the essence of bergamot [5].

Nowadays, these historically important oils, rich in monoterpenes, are complemented by other essential oils from flowers, roots, fruit, wood, and moss [6], e.g., lavender and petitgrain oils are rich in linalyl acetate (**9b**) and lemon oil in γ-terpinene (**7**) and β-pinene (**8**). Commonly, essential oils are obtained by distillation with water or steam, and separation from the aqueous phase upon cooling. Many of these essential oils contain substantial fractions of mono- and sesquiterpenes, the most prominent examples and their olfactory properties being shown in Figure 2. α-Terpinene (**10**) is found in cardamom and marjoram oils, while isomeric terpinolene (**11**) is present in pine oils. α-Phellandrene (**12**) is a constituent of elemi oil, whereas red/pink pepper oils are rich both in α- phellandrene (**12**) as well as β-phellandrene (**13**). Rosemary and eucalyptus oils contain the monoterpene ether 1,8-cineole (**14**) and camphene (**21**) can be isolated from the Siberian fir needle oil. Cymbopogon oils provide among other components borneol (**16b**), geranyl acetate (**4b**) and citronellol (**5**). Besides limonene (**6**), (−)-carvone (**17**) is one of the main constituents in caraway oil and dill seed oil yields (+)-carvone (**17'**). Not surprisingly, camphor oil is rich in camphor (**16a**). Cypress oil yields 3-carene (**20**) as one major constituent. Mint oils serve as one possible source for menthone (**2**), menthol (**15**) and (−)-carvone (**17**). Essential oils collected from eucalyptus are rich in 1,8-cineole (**14**) and from fennel oil in fenchone (**18**). Farnesol (**23**) is present in many essential oils such as citronella, neroli, cyclamen and lemon grass. Nerolidol (**24**) is present in neroli, ginger, jasmine, lavender, tea tree and other essential oils. Finally, vetiver oil contains the sesquiterpene khusimol (**25**) from which the acetate (**26**) can be prepared by semisynthesis [6].

Obviously, nature served as starting point and guideline for creating scents and these lists reveal, that terpenes, particularly mono- and sesquiterpenes, have played a rather dominant role in the fragrance industries [6]. Over the last decades, the demand for fragrances has grown dramatically, so that plantations serve to provide the raw materials. In parallel, synthetic efforts also dramatically expanded to fulfil the huge demand of the consumer markets, new olfactory experiences included. Indeed, synthetic compounds were not introduced until the dawn of 19th century and first and foremost coumarin played a key role, first synthesised by Perkin in 1868 [7,8]. Following this breakthrough, many other perfumes were created based on synthetic molecules born from the newly established discipline

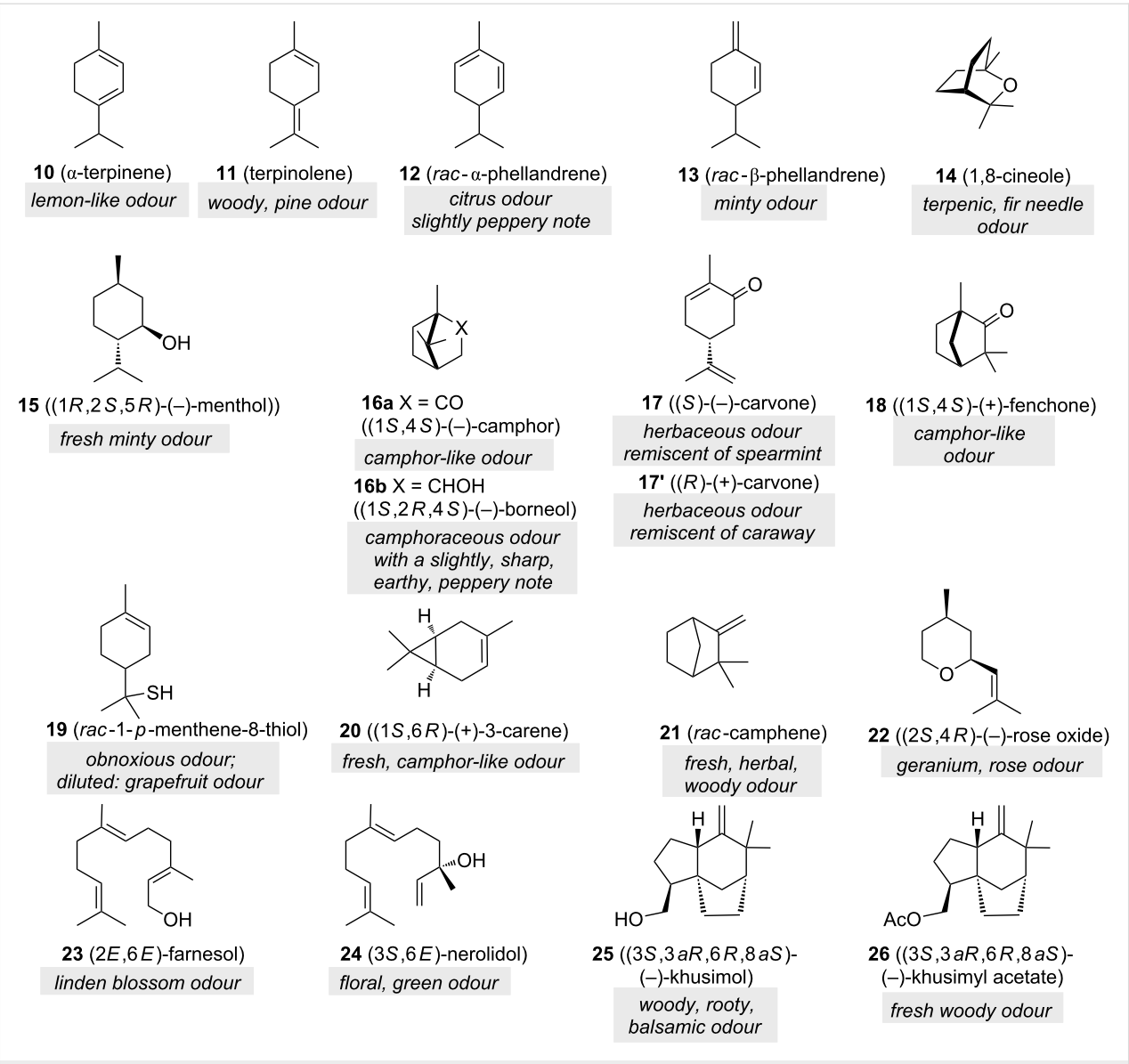


Figure 2: Other selected mono- and sesquiterpenes (**10**–**26**) as fragrance materials [6].

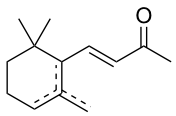
synthetic organic chemistry. This made odorants available for the broad masses and perfumes to be worn according to one's daily mood [2,3].

Key enabling milestones were the musk ketone accidentally discovered in 1894, being an important compound not derived directly from nature. Other musky compounds are (−)-(3*R*)-muscone, isolated in small-yields from glandular secretion of the musk deer, and 15-pentadecanolide were utilised too [9].

The discovery and modern applications of Iso E Super®

It has to be stressed that musky odours were not the only scents of interest but also the spectrum of fragrances from violet flower oils. In fact, these were the most expensive of all available essential oils. Exorbitant quantities of flower petals were extracted to collect the oil, used directly in cosmetic formulations or spread on laundry to generate a characteristic smell. As for musk fragrances, there was a quest in the perfume industries to find a synthetic solution to create scents that mimic violet flower oils. First, a similarly smelling but more affordable orris root oil (*Iris pallida* Lam., fam. Iridaceae) was chosen for structural analysis. Thiemann and Krüger isolated irone (**27**, Figure 3), whose molecular formula was first falsely assigned as $C_{13}H_{20}O$ [10].

In an attempt to recreate this compound by condensation of acetone with citral (**28**) a compound with "a strange but not very

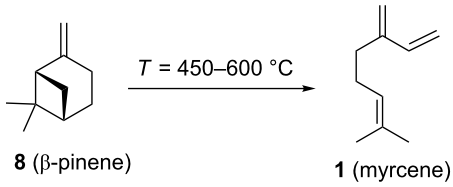


27

Figure 3: Main constituents of natural iris oil: irone (**27**).

characteristic odour" was formed, later named pseudoionone (**29**, Scheme 1). It turned out not to be suited for further investigations. However, after cleaning the glassware with sulfuric acid, a distinctive scent of violets was noted which later was linked to ionone (**30**) being created in the acidic medium. Thus, the category of synthetic ionone (**30**) and woody smelling compounds was born in 1893 and investigated further in the following years [10-13].

Following this invention, many derivatives were produced to find new viable targets. These studies mostly focused on Diels–Alder cycloadditions to create structures that resemble terpenoids readily available from easily accessible and affordable starting materials like myrcene (**1**). One of the newly found products was Ambrelux (**32**, Scheme 2) that was further cyclised in a similar fashion previously mentioned for ionone compounds. This process yielded Isocyclemone E® (**33**), later rebranded to the famous name Iso E Super® (**33**) that is valid until today [9,14,15]. Indeed, myrcene (**1**) is one of the most versatile monoterpenes to be used as starting material for generating products in various industries. These include polymers, insect repellents, vitamins, flavours and fragrances [16]. Commercially, it is obtained from turpentine, a side product in paper manufacturing. Its main constituents are α-pinene (**3**) and β-pinene (**8**), 3-carene (**20**), limonene (**6**) and camphene (**21**). Since no large-scale source for myrcene (**1**) was available, a short route from readily available monoterpenes was established. Under pyrolytic conditions β-pinene (**8**) as constituent of turpentine undergoes a rearrangement to myrcene (**1**) (Scheme 3) [17-21].

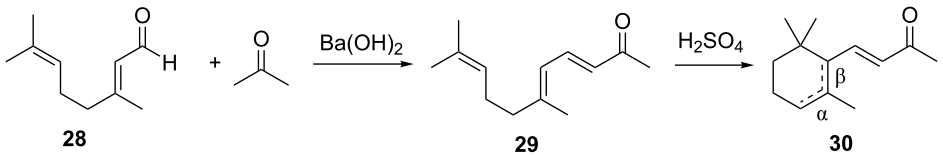


8 (β-pinene) **1** (myrcene)

Scheme 3: Industrial synthesis of myrcene (**1**) by pyrolysis of β-pinene (**8**).

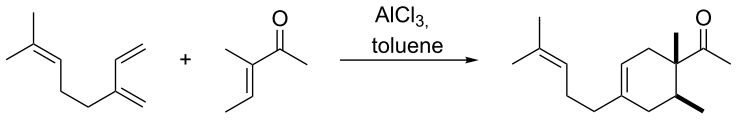
To produce Ambrelux (**32**), myrcene (**1**) is reacted with dienophile (**31**) in a Diels–Alder cycloaddition promoted under Lewis-acidic conditions. In order to obtain Iso E Super® (**33**), Brønstedt acid-mediated cyclisation, similar to the one utilised for the first synthesis of ionone (**30**), proved feasible on large scale. As it turned out, not only the one depicted, but several other cyclisation products formed. The main constituent was Iso E Super® (**33**). A minor byproduct is now referred to as Iso E Super Plus® (**34**, Scheme 4). Small modifications of the reaction conditions yielded other geometric isomers. In 2007, a thorough study was published by Fráter et al. disclosed of how such variations of parameters affect product formation and composition [22].

Interestingly, Iso E Super® (**33**) itself shows a comparably high odour threshold of 500 ng L^{-1} as was reported in the original patent [15]. An impurity of ca. 5%, now called Iso E Super Plus® (**34**), was made responsible for the characteristic smell having an odour threshold as low as 5 ng L^{-1} [23]. Naturally,



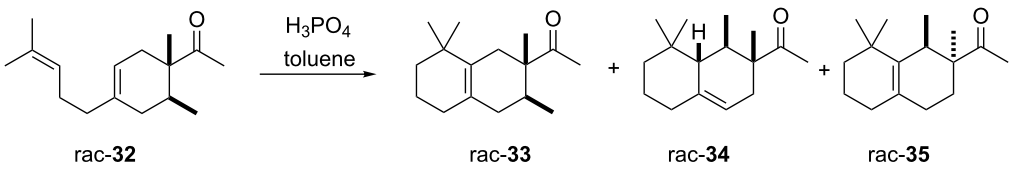
28 **29** **30**

Scheme 1: First synthesis of ionone (**30**) [11].



1 **31** rac-**32**

Scheme 2: First synthesis of Ambrelux (**32**) [14].



Scheme 4: First synthesis of Iso E Super® (**33**), Iso E Super Plus® (**34**) and Georgywood® (**35**) as a mixture of isomers [15].

this impurity was thoroughly analysed in the laboratories of Givaudan SA and finally secured in a patent as Iso E Super Plus® (**34**). Later, also the second impurity Georgywood® (**35**) with a higher odour threshold of 15 to 30 ng L^{-1} but better odour characteristics was patented [17-20]. Further details on the individual components of this complex mixture are listed in Table 1.

It must be noted that the conditions for the synthesis of all Iso E Super® related compounds vary slightly. The main difference lies in a prolonged isomerisation process of the Diels–Alder product **32** before and after the second cyclisation step. Georgywood® (**35**) named after Georg Fráter is industrially produced with, e.g., methanol as additive to enforce isomerisation and suppress premature cyclisation [24-26].

Today, Iso E Super® (**33**) and its isomers are widely used in a variety of perfumery products. From Haliston *Woman* that only contains a very small portion of this component and Christian Dior´s *Fahrenheit* consisting of 25% Iso E Super® (**33**), LancÔme *Trésor* (18%) and Shiseido´s *Feminite du Bois* (43%), *Bois de Violette*, *Bois et Fruits*, *Bois et Musc*, *Un Bois*

Table 1: Individual components of the complex Iso E Super® mixture.

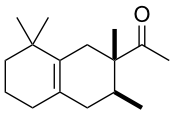
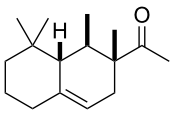
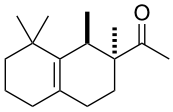
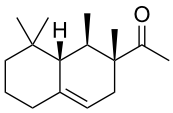
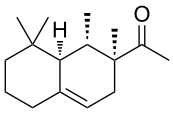
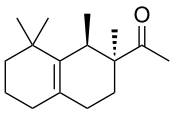
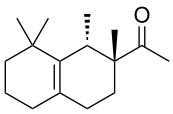
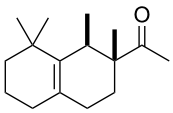
Component	Commercial name	CAS number	Odour (threshhold)
 *rac*-**33**	Iso E Super® (**33**)	59056-94-9	woody, floral, ambergris, violet, old-wood, lemony (500 ng L^{-1})
 *rac*-**34**	Iso E Super Plus® (**34**)	140194-26-9	woody–ambery, very strong odour (5 pg L^{-1})
 *rac*-**35**	Georgywood® (**35**)	185429-83-8	(15 pg L^{-1}–30 pg L^{-1})
 (+)-**34**	(+)-E Super Plus® (**34**) (Corey also referred to it as “arborone”)	356088-93-2	intense woody odor, clean and pleasant (5 pg L^{-1})
 (−)-**34**	(−)-E Super Plus® (**34**)	356088-90-9	faint odor
 (−)-**35**	(−)-Georgywood® (**35**)	828933-31-9	woody-ambery. bottom note: fresh, minty, green, sweet (20 pg L^{-1})

Table 1: Individual components of the complex Iso E Super® mixture. (continued)

 (+)-**35**	(+)-Georgywood® (**35**)	828933-41-1	weakly woody bottom note: unpleasant, acrid, musty (3.5 ng L^{-1})
 *rac*-**53**	**53**	260792-30-1	n.a.

Sepia, *Un Bois Vanille*, Christian Dior´s *Dolce Vita*, just to mention a few (Table 2). Only recently, it was probed, whether it is possible to further increase the amount of Iso E Super® (**33**) in a commercial perfume [25,26].

This trend culminated in Schön´s creation of *Molecule 01* in the year 2005, a perfume that contains nothing else than Iso E Super® (**33**). *Orb_ital* from Nomenclature (75% Iso E Super®) followed in the year 2015. This fragrance collection has set itself the task of using a range of synthetic fragrances as "overdoses" in perfumes.

The name *Orb_ital* derives from Orbitone, a brand name from the olfactory active (2*R*, 3*R*)-Iso E Super® (**33**) [27]. It has to be stressed, that all compounds related to Iso E Super® are not handled as single isomers but rather as varying mixtures because none of the industrial syntheses is very stereo- and regioselective as shown by GC analysis in Figure 4. So far efforts in industrial production have been directed towards product mixtures that are dominated by one isomer with favourable olfactory properties.

What seems to be counterintuitive for purely synthetically oriented or medicinal chemists, can be rationalised, when briefly considering the biochemical mechanism of the smell and the operation of scents. The odour impression is created by olfactory receptor neurons inside the nose. Since olfaction is a very complicated and broad field, it is hard to predict how molecules and mixtures of different molecules affect the perception. This is especially complex since odour impressions may change when concentrations are altered. On the lowest level, compounds of interest interact with so-called G-protein receptors consisting of seven intermembrane domains [28]. The quaternary structure including the membrane set up the active site.

Table 2: Top fragrances with regard to their volume percentage (listed down to about 20%; the large number of perfumes with lower percentages are not listed) of Iso E Super® in perfume oil [1].

Fragrance Name	Company	Launch year	Iso E Super® (**33**) and Iso E Super Plus® (**34**)
Molecule 01	Escentric Molecules	2005	100%
Perles	Lalique	2007	80%
Orb_ital	Nomenclature	2015	75%
Poivre Samarcande	Hermès	2004	71%
Escentric 01	Escentric Molecules	2005	65%
Terre d'Hermès	Hermès	2006	55%
Incense Kyoto	Comme des Garçons	2002	55%
Incense Jaisalmer	Comme des Garçons	2002	51%
Fierce for Men	Abercrombie & Fitch	2002	48%
Kenzo Air	Kenzo	2003	48%
Encre noire	Lalique	2006	45%
Feminite du Bois	Shiseido	1992	43%
Fahrenheit	Christian Dior	1988	25%
Tresor	Lancome	1990	18%
Aventus	Creed	2010	18%

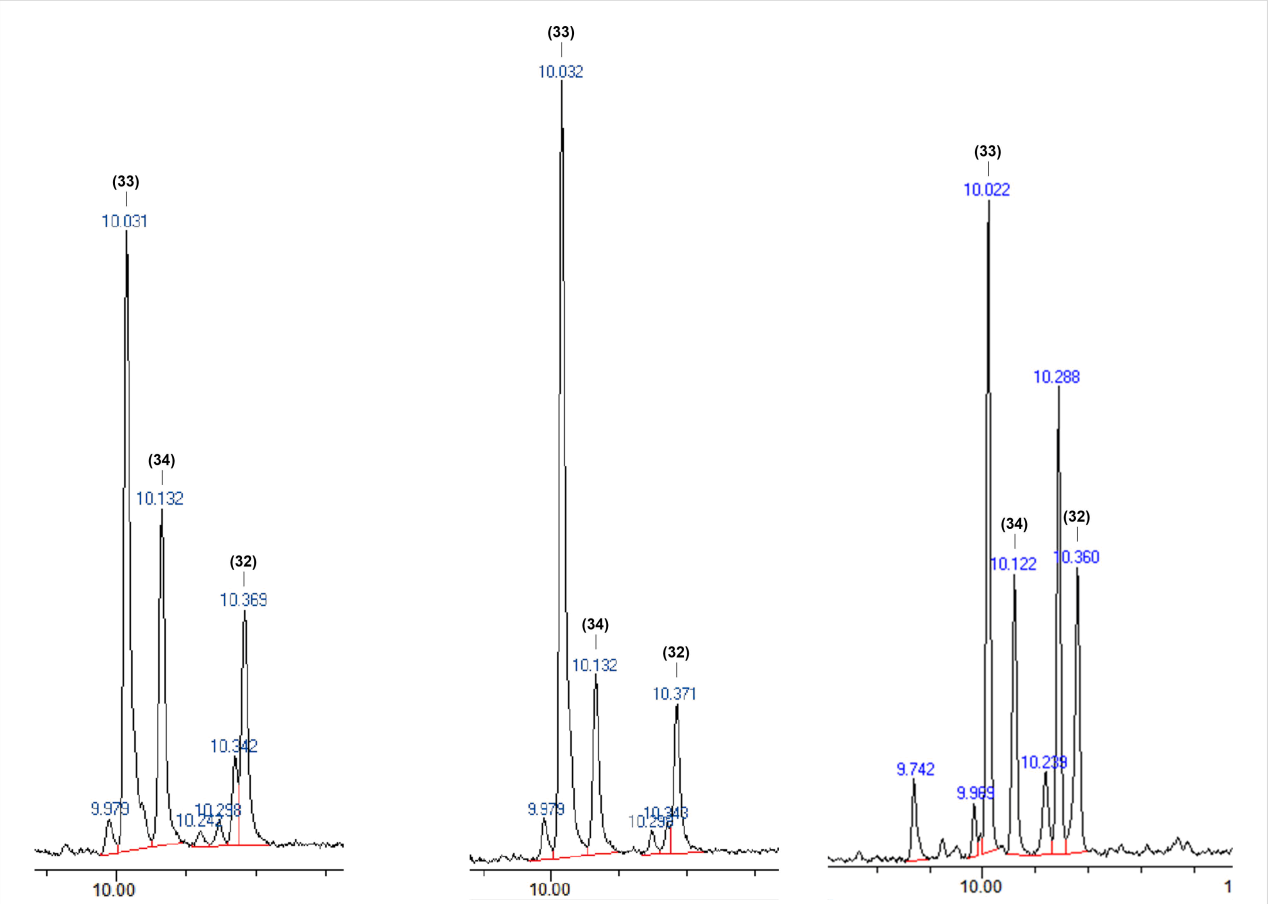


Figure 4: Iso E Super® region of GC spectra of *Molecule 01* (left, 75 €–100 € per 100 mL; march 2019), a low-priced counterpart from TOP2BASE (middle, 20 € per 100 mL) and product produced in our laboratories under flow conditions. Note, although both samples provide a similar smell, the olfactory effects slightly vary because of different proportions of products and isomers produced in very small amounts that impact the nuances of a scent significantly (GC analyses were conducted in the author´s laboratories on a Zebron™ phenomenex ZB-1MS F&F (20 m, 0.18 mm, 0.18 µm) column).

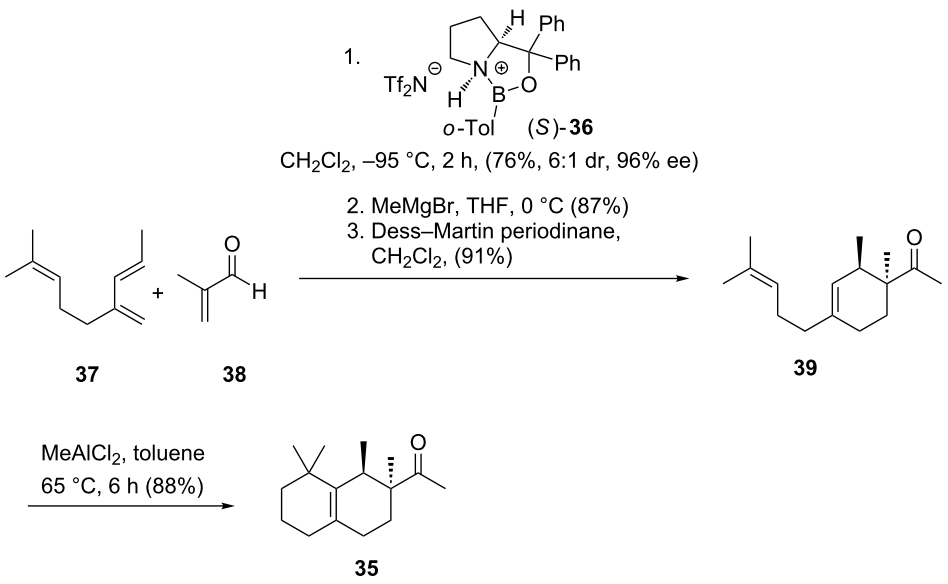
Approximately 370 different G-type proteins are known, that are linked with the odour perception. Because molecules can bind to an array of olfactory receptors generating a complex odour impression, an exact determination which proteins are linked to which smells or molecules is a very ambitious task. Hence, studies towards understanding interactions led to a Nobel Award in 2002 [29-31]. Even today correct modelling and protein crystallisation are immense challenges to be solved. Hydrophilic and hydrophobic interactions with the unpolar lipid layer make the tendency to yield suitable crystals even more difficult. Nevertheless, Palczewski and co-workers were able to crystallise the first GPCR (G-protein-coupled receptor) in 2000 confirming the previously described structure [28,32].

Synthetic aspects of individual Iso E Super® components

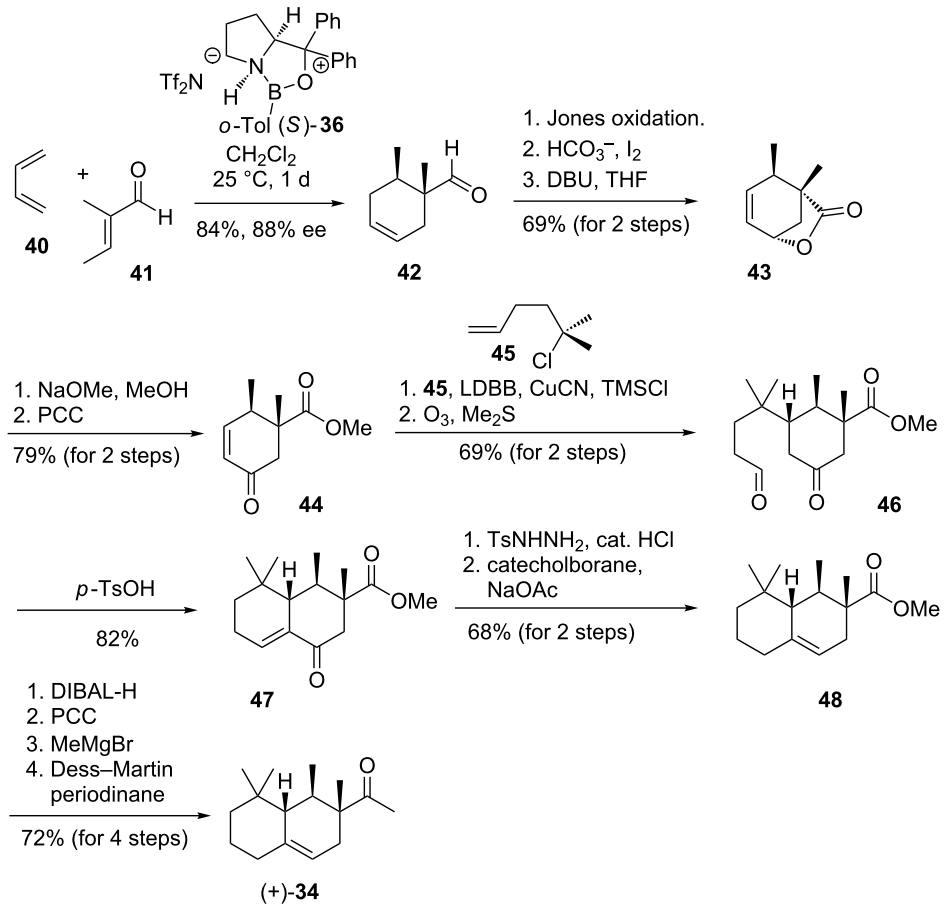
The first target-specific synthesis of (−)-Georgywood® (**35**) utilised the (*S*)-Corey–Bakshi–Shibata catalyst (**36**) for the enantioselective Diels–Alder cycloaddition (Scheme 5). The corresponding enantiomer (+)-Georgywood® (**35**) was also prepared using the corresponding (*R*)-CBS catalyst (**36**).

In contrast, the enantiomer (+)-Georgywood® (**35**) was found to possess a relatively weak odour which was described as distinctly unpleasant and acrid-musty by several members of the Corey group [33]. The same approach led to the discovery of (+)-Iso E Super Plus® (**34**) as a highly active component (Scheme 6). Fráter et al. confirmed these experiences after isolation of active olfactory compounds of Iso E Super Plus® (**34**) and Georgywood® (**35**). Racemic resolution provided a crystalline material that served to obtain an X-ray structure of the oxime derivative of (−)-(1*R*,2*S*)-Georgywood® ((−)-**35**) [33,34].

Corey´s asymmetric synthesis of Iso E Super Plus® ((+)-**34**) is initiated by a stereoselective Diels–Alder cycloaddition utilizing the CBS catalyst (**36**) to yield the cyclohexene derivative **42** with good facial selectivity [33]. Oxidation, iodocar-



Scheme 5: First synthetic route to (−)-Georgywood® (**35**) by Corey and Hong [33].

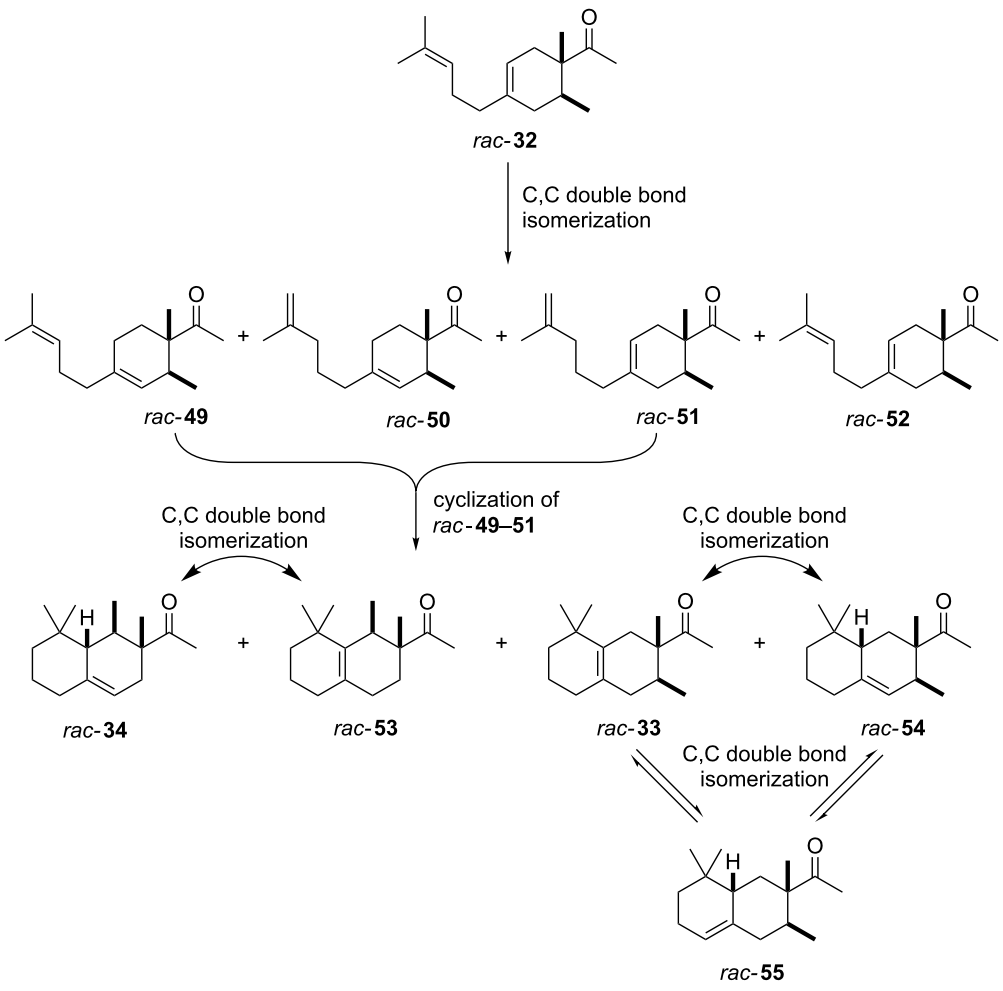


Scheme 6: First synthetic route to the odour-active (+)-enantiomer of Iso E Super Plus® (+)-**34** [33].

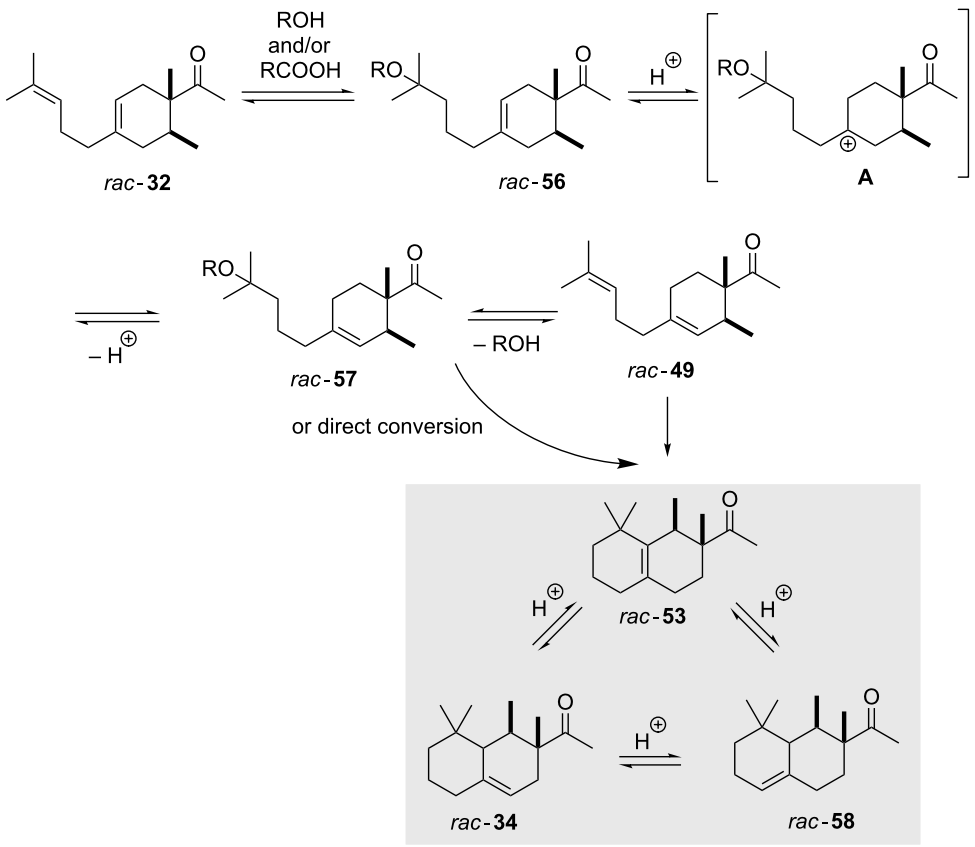
boxylation and elimination yielded the lactone **43**. A series of functional group manipulations provided enone **44**, which underwent a cuprate-mediated Michael addition and liberation of the aldehyde **46** upon ozonolysis. After intramolecular aldol condensation the resulting enone **47** was transformed into cyclohexene **48** with shifted olefinic group by means of a reductive variant of the Wolff–Kishner deoxygenation. A straightforward four-step sequence finally yielded Iso E Super Plus® ((+)-**34**).

Industrially pursued syntheses do not involve a specific stereoinducing step. In fact, it is mentioned in the patents that the standard industrial process of Iso E Super® (**33**) utilises technical grade chemicals for both synthetic steps. The mixture of resulting isomers is then used in perfumes, when the smell meets standard criteria by quality control [14]. As encountered earlier, the second step of production is the most important one for product formation and composition. Therefore, several patents exist describing the isomerisation and cyclisation steps involved. In the first step, both olefinic double bonds of the primary Diels–Alder product **32** can isomerise, thereby creating several precursors **49**–**52** that, accept for **52**, are suited to undergo a second cyclisation as depicted in Scheme 7. After the following cyclisation step, the double bond of the racemic products obtained isomerises between α, β and γ [25].

Furthermore, Erman and co-workers from Millenium Speciality Chemicals Inc. described a process, which involves methanol and other alcohols or alternatively organic acids as nucleophilic additives that can reversibly be introduced and removed again (Scheme 8). Typically, methanol, ethanol, isopropanol and 2-methoxyethanol served as suitable alcohols. According to patent information di- or polyols can also serve as "dummy" additives. Alternatively, also acetic acid was suggested. Using this method, the desired Iso E Super Plus® (**34**) concentration ranged from 5% to 7% as judged by GC analysis [26].



Scheme 7: Analysis of the isomerisation process and formation of products. Most importantly, Iso E Super® (**33**), Iso E Super Plus® (**34**) and Iso Gamma (**55**) are formed [1]. All compounds are obtained as racemates [25].



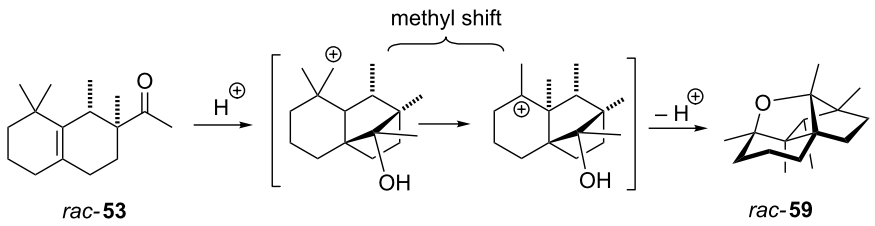
Scheme 8: Isomerisation using additives such as alcohols or carboxylic acids. The product with the γ-positioned double bond is the desired Iso E Super Plus® (**34**). Products **58** (α double bond) and product **53** (β double bond) are not desired [26].

Fráter and Schröder discovered that Iso E Super Plus® (**34**) can undergo an additional cyclisation through compound *rac*-**53** (Scheme 9). This is initiated by the acid employed in the second step of the synthesis. Thus, the ketone is protonated and the highly electrophilic carbon atom reacts with the alkene moiety. The resulting tertiary carbocation undergoes a 1,2-methyl shift to yield a new cation, which in turn is nucleophilically trapped by the carbinol moiety. The resulting tetrahydrofuran **59** is chemically stable and this observation was used as rationale for the erosion of the isomeric ratio observed during prolonged reaction times.

In the same piece of work Fráter et al. investigated the influence of Brønstedt and Lewis acids on the formation of Georgywood® (**35**). It was found that Lewis acids such as $AlCl_3$ shift the equilibrium towards Georgywood® (**35**) type products especially when employed in over-stoichiometric amounts. Using different Brønstedt acids, the ratios between the products obtained can change drastically [22].

Conclusion and Outlook

Here, we presented a short story on Iso E Super® and derivatives formed during synthesis, a group of molecules that has



Scheme 9: Iso E Super Plus® (**34**) can undergo a third cyclisation to tetrahydrofuran **59** through compound *rac*-**53** [22].

changed the perfume industry, but has its roots in the terpenoid ingredients of classical essential oils geranium and bergamot (Figure 5). Starting from ionone (**30**) an "evolutionary process" towards synthetic products with similar olfactory properties led to Iso E Super® (**34**), Iso E Super Plus® (**35**) and Georgywood® (**35**), a development that took almost hundred years and saw koavone and timberole as intermediates. An analysis of today´s fine fragrances reveals that almost all of them combine synthetic scent molecules with traditional essential oils, despite the fact, that the ongoing consumer trend is towards natural ingredients. Avoiding synthetics like Iso E Super® (**33**) would rule out many favourite scents. In fact, about 100 natural fragrance ingredients are known, but perfumers have more than 3,000 synthetic molecules at hand of which several examples **60**–**66** with terpene-like structures are listed in Figure 6. Noteworthy, the fragrance properties of synthetically-derived unnatural compounds commonly mimic those of natural products.

Biotechnology is another way to harness fragrance components be it enzymatic or microbial. Nowadays engineered microbes are at hand that, e.g., produce scents, such as patchouli, by fermenting sugar. Patchouli is a complex mixture of sesquiterpenes ((−)-patchoulol, (+)-norpatchoulenol, (+)-α-bulnesene, (−)-α-guajene, (−)-β-patchoulene and (−)-seychellene) with a slightly camphoraceous, woody balsamic odour [6].

Enzymatic derivatisation of terpenes by means of biocatalysis is another opportunity to create new fragrance molecules or to

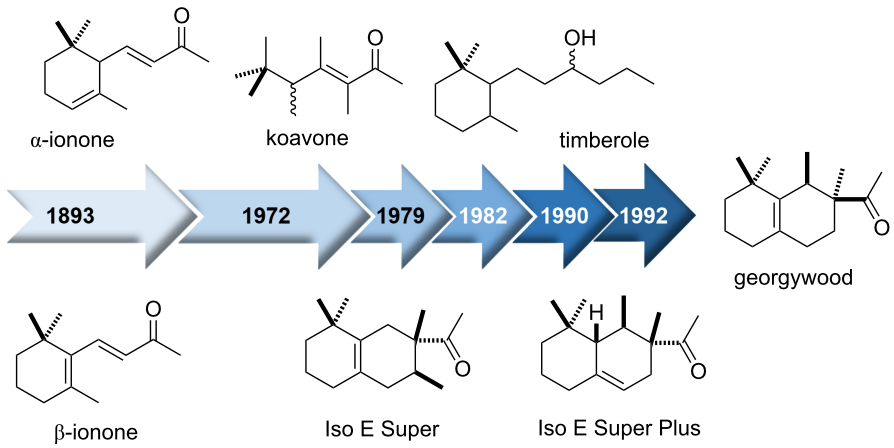


Figure 5: (Adapted from ref. [8]) Ionone (**30**, 1893, odour threshold: 0.8 ng L^{-1}), koavone (1982, odour threshold: 75 ng L^{-1}), Iso E Super® (**33**) (1972, odour threshold: 500 ng L^{-1}), timberole (1982, odour threshold: 26 ng L^{-1}) Iso E Super Plus® (**34**) (1990, odour threshold: 5 pg L^{-1}) and Georgywood® (**35**, 1996, odour threshold: 15 pg L^{-1}–30 pg L^{-1}) [9,15,33,35-38].

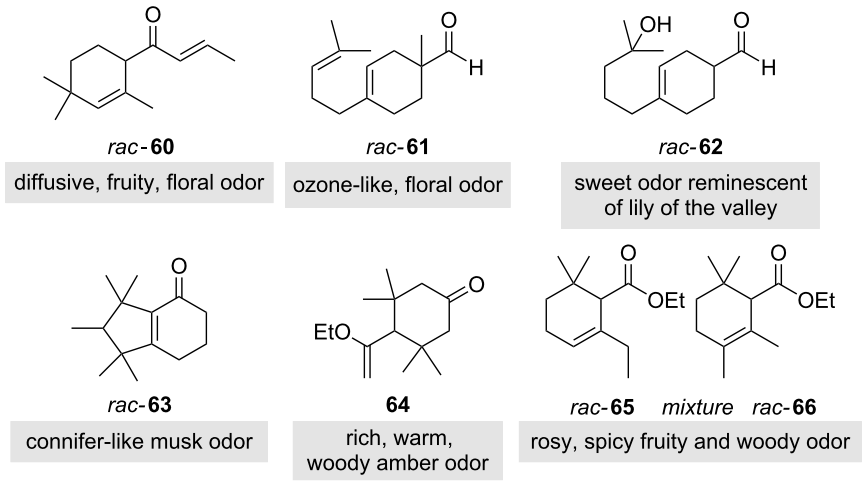
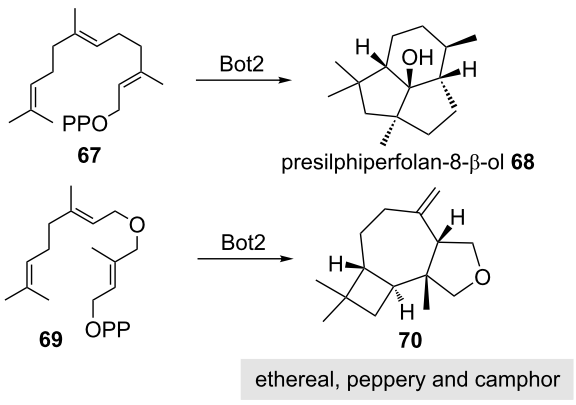


Figure 6: Branched, terpene-like cyclohexene derivatives, that are synthetic fragrance components: **60**: Iso damascone, **61**: Precyclemone B, **62**: Lyral®, **63**: Cashmeran®, **64**: Kephalis, and **65**, **66**: Givescone®.

achieve chiral resolution of racemates. The former process is commonly associated with oxidation reactions, while the latter process is often based on the action of lipases. Very recently, a new concept was disclosed that probed sesquiterpene cyclases to accept unnatural farnesyl pyrophosphates and generate unnatural cyclisation products with unusual backbones. Thus, in the presence of presilphiperfolan-8-β-ol synthase (Bot2) a novel tricyclic product **70** was obtained from unnatural farnesyldiphosphate ether **69**. The olfactory analysis revealed an ethereal, peppery and camphoric scent (Scheme 10) [39].



Scheme 10: New unnatural terpenoid **70** from unnatural farnesyl pyrophosphate derivative **69** and comparison with natural biotransformation (**67**→**68**) and olfactory property of tetrahydrofuran **70** [39].

Future prospects of the fragrance industry will be linked with a bouquet of methods to broaden the platform of molecules with favourable olfactory properties. These include chemical synthesis, microbiology and molecular biology associated with biotechnology and combinations based on these methods. Hence also the most recent developments in synthetic biology will appear on the stage of the world of fragrances [40,41].

Acknowledgements

We thank Dr. Johannes Panten (Symrise AG, Holzminden) for inspirational discussions on all aspects of the world of fragrances.

ORCID® iDs

Alexey Stepanyuk - https://orcid.org/0000-0001-5944-8924
Andreas Kirschning - https://orcid.org/0000-0001-5431-6930

References

1. Schön, G. *Chem. Biodiversity* **2008,** *5,* 1154–1158. doi:10.1002/cbdv.200890092
2. Morris, E. T. *Scents of Time: Perfume from Ancient Egypt to the 21st Century;* Prestel, 1999.
3. Kraft, P.; Bajgrowicz, J. A.; Denis, C.; Fráter, G. *Angew. Chem., Int. Ed.* **2000,** *39,* 2980–3010. doi:10.1002/1521-3773(20000901)39:17<2980::aid-anie2980>3.0.co;2-#
4. Fougere Royale - Houbigant Parfum Paris - Men's Fragrances. https://www.houbigant-parfum.com/eu_en/fougere-royale.html/ (accessed July 8, 2019).
5. Sawamura, M.; Onishi, Y.; Ikemoto, J.; Tu, N. T. M.; Phi, N. T. L. *Flavour Fragrance J.* **2006,** *21,* 609–615. doi:10.1002/ffj.1604
6. Surburg, H.; Panten, J., Eds. *Common Fragrance and Flavor Materials,* 6th ed.; Wiley-VCH Verlag GmbH: Weinheim, Germany, 2016. doi:10.1002/9783527693153
7. Perkin, W. H. *Justus Liebigs Ann. Chem.* **1868,** *148,* 203–207. doi:10.1002/jlac.18681480210
8. Groom, N. *The Perfume Handbook;* Springer Science+Business Media S.A., 1992. doi:10.1007/978-94-011-2296-2
9. Gautschi, M.; Bajgrowicz, J. A.; Kraft, P. *Chimia* **2001,** *55,* 379–387.
10. Schinz, H. Die Veilchenriechstoffe. *Fortschritte der Chemie Organischer Naturstoffe / Progress in the Chemistry of Organic Natural Products / Progrès Dans la Chimie des Substances Organiques Naturelles;* Springer Vienna: Vienna, Austria, 1951; pp 146–206. doi:10.1007/978-3-7091-7172-1_5
11. Tiemann, F.; Krüger, P. *Ber. Dtsch. Chem. Ges.* **1893,** *26,* 2675–2708. doi:10.1002/cber.18930260373
12. Ohloff, G.; Mignat, S. *Justus Liebigs Ann. Chem.* **1962,** *652,* 115–126. doi:10.1002/jlac.19626520116
13. Nussbaumer, C.; Frater, G. *J. Org. Chem.* **1987,** *52,* 2096–2098. doi:10.1021/jo00386a038
14. Ohloff, G. Method of producing derivatives of the 1,1-dimethyl-octahydro- naphthalene series. G.B. Patent 896,039A, July 18, 1959.
15. Hall, J. B.; Sanders, J. M. Perfume compositions and perfurme articles containing one isomer of an octahydrotetramethyl acetonaphthone. U.S. Patent 3,929,677A, Dec 30, 1975.
16. Behr, A.; Johnen, L. *ChemSusChem* **2009,** *2,* 1072–1095. doi:10.1002/cssc.200900186
17. Goldblatt, L. A.; Palkin, S. Process for converting nopinene to myrcene. U.S. Patent 2,420,131, May 6, 1947.
18. Savich, T. R.; Goldblatt, L. A. Process for producing myrcene from beta-pinene. U.S. Patent 2,507,546, May 16, 1950.
19. Rives Resiniques, T. E. R. P. Myrcene prodn. by isomerisation - using reduced press. French Patent 2,274,584, June 17, 1976.
20. Chibiryaev, A. M.; Yermakova, A.; Kozhevnikov, I. V.; Sal'nikova, O. I.; Anikeev, V. I. *Russ. Chem. Bull.* **2007,** *56,* 1234–1238. doi:10.1007/s11172-007-0186-x
21. Yermakova, A.; Chibiryaev, A. M.; Kozhevnikov, I. V.; Anikeev, V. I. *J. Supercrit. Fluids* **2008,** *45,* 74–79. doi:10.1016/j.supflu.2007.12.005
22. Fráter, G.; Schröder, F. *J. Org. Chem.* **2007,** *72,* 1112–1120. doi:10.1021/jo061668k
23. Nussbaumer, C.; Fráter, G.; Kraft, P. *Helv. Chim. Acta* **1999,** *82,* 1016–1024. doi:10.1002/(sici)1522-2675(19990707)82:7<1016::aid-hlca1016>3.0.co;2-y
24. Bajgrowicz, J. A.; Bringhen, A.; Frater, G.; Müller, U. Novel perfume. Eur. Patent 0,743,297B1, Nov 20, 1996.
25. Erman, M. B.; Williams, M. J.; Cárdenas, C. G. Process for obtaining fragrant ketone. U.S. Patent 6,423,874B1, July 23, 2000.
26. Erman, M. B.; Cardenas, C. G.; Hoffmann, H. M. Process for obtaining mixtures of acyloctahydronaphthalenes. U.S. Patent 6.160.182, Dec 12, 2000.

27. Yudov, M. Die Geschichte des Iso E Super® in der Parfümerie ~ Duftnoten. https://www.fragrantica.de/Neuigkeiten/Die-Geschichte-des-Iso-E-Super-in-der-Parfuemerie-3436.html (accessed April 21, 2019).
28. Palczewski, K.; Kumasaka, T.; Hori, T.; Behnke, C. A.; Motoshima, H.; Fox, B. A.; Le Trong, I.; Teller, D. C.; Okada, T.; Stenkamp, R. E.; Yamamoto, M.; Miyano, M. *Science* **2000,** *289,* 739–745. doi:10.1126/science.289.5480.739
29. Buck, L.; Axel, R. *Cell* **1991,** *65,* 175–187. doi:10.1016/0092-8674(91)90418-x
30. Buck, L. B. *Nutr. Rev.* **2004,** *62,* S184–S188. doi:10.1111/j.1753-4887.2004.tb00097.x
31. Malnic, B.; Gonzalez-Kristeller, D.; Gutiyama, L. *Odorant Receptors;* Frontiers in Neuroscience; CRC Press: Boca Raton, FL, U.S.A., 2009.
32. Lagerström, M. C.; Schiöth, H. B. *Nat. Rev. Drug Discovery* **2008,** *7,* 339–357. doi:10.1038/nrd2518
33. Hong, S.; Corey, E. J. *J. Am. Chem. Soc.* **2006,** *128,* 1346–1352. doi:10.1021/ja057483x
34. Fráter, G.; Müller, U.; Schröder, F. *Tetrahedron: Asymmetry* **2004,** *15,* 3967–3972. doi:10.1016/j.tetasy.2004.11.003
35. Etzweiler, F. Acetyl-tri- and tetramethyl-octahydronaphthalenes and fragrance compositions containing same. U.S. Patent 5,180,709, Jan 19, 1991.
36. Etzweiler, F.; Helmlinger, D.; Nussbaumer, C.; Pesaro, M. Acetyl-tri- and tetramethyl-octahydronaphthalenes and fragrance compositions containing the same. CA Patent 2,042,673C, Oct 16, 1990.
37. Bajgrowicz, J. A.; Bringhen, A.; Fráter, G.; Müller, U. Odorant compounds and compositions. U.S. Patent 005,707,961A, Jan 13, 1996.
38. Kraft, P. *Chem. Biodiversity* **2008,** *5,* 970–999. doi:10.1002/cbdv.200890108
39. Oberhauser, C.; Harms, V.; Seidel, K.; Schröder, B.; Ekramzadeh, K.; Beutel, S.; Winkler, S.; Lauterbach, L.; Dickschat, J. S.; Kirschning, A. *Angew. Chem., Int. Ed.* **2018,** *57,* 11802–11806. doi:10.1002/anie.201805526
40. Braga, A.; Guerreiro, C.; Belo, I. *Food Bioprocess Technol.* **2018,** *11,* 2217–2228. doi:10.1007/s11947-018-2180-8
41. Brenna, E.; Fuganti, C.; Gatti, F. G.; Serra, S. *Chem. Rev.* **2011,** *111,* 4036–4072. doi:10.1021/cr100289r

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Nanangenines: drimane sesquiterpenoids as the dominant metabolite cohort of a novel Australian fungus, *Aspergillus nanangensis*

Heather J. Lacey[1,2], Cameron L. M. Gilchrist[3], Andrew Crombie[1], John A. Kalaitzis[1,4], Daniel Vuong[1], Peter J. Rutledge[2], Peter Turner[2], John I. Pitt[1], Ernest Lacey[1,4], Yit-Heng Chooi[*3] and Andrew M. Piggott[*4]

Full Research Paper

Open Access

Address:
[1]Microbial Screening Technologies, Smithfield, NSW 2164, Australia, [2]School of Chemistry, The University of Sydney, NSW 2006, Australia, [3]School of Molecular Sciences, The University of Western Australia, WA 6009, Australia and [4]Department of Molecular Sciences, Macquarie University, NSW 2109, Australia

Email:
Yit-Heng Chooi[*] - yitheng.chooi@uwa.edu.au; Andrew M. Piggott[*] - andrew.piggott@mq.edu.au

* Corresponding author

Keywords:
Aspergillus; biosynthesis; drimane; secondary metabolites; sesquiterpenoid; terpenes

Beilstein J. Org. Chem. **2019,** *15,* 2631–2643.
doi:10.3762/bjoc.15.256

Received: 22 July 2019
Accepted: 14 October 2019
Published: 05 November 2019

This article is part of the thematic issue "Terpenes".

Guest Editor: J. S. Dickschat

Abstract

Chemical investigation of an undescribed Australian fungus, *Aspergillus nanangensis*, led to the identification of the nanangenines – a family of seven new and three previously reported drimane sesquiterpenoids. The structures of the nanangenines were elucidated by detailed spectroscopic analysis supported by single crystal X-ray diffraction studies. The compounds were assayed for in vitro activity against bacteria, fungi, mammalian cells and plants. Bioinformatics analysis, including comparative analysis with other acyl drimenol-producing Aspergilli, led to the identification of a putative nanangenine biosynthetic gene cluster that corresponds to the proposed biosynthetic pathway for nanangenines.

Introduction

The fungal genus *Aspergillus* is well recognised as a source of structurally diverse terpenoids comprising monoterpenoids [1], sesquiterpenoids [2-5], diterpenoids [6], sesterterpenoids [7-9], triterpenoids [10] and prenylated polyketide meroterpenoids [11-15] isolated from soil, endophytes and marine strains. Of this genus, *A. ustus* [16], *A. calidoustus* [17], *A. insuetus* [17], *A. insulicola* [18], *A. bridgeri* [18], *A. sclerotiorum* [19], *A. variecolor* [19], *A. parasiticus* [20], *A. oryzae* [21], *A. ochraceus* [22], *A. pseudodeflectus* [17], *A. carneus* [23] and *Aspergillus* sp. strain IBWF002-96 [4,5] are biosynthetic sources of the drimane sesquiterpenoids. Drimane sesquiterpenoids, which are derived from a parent C_{15} pentamethyl-

trans-decalin skeleton, are known to occur in plants, sponges, molluscs and other fungi as well and possess a wide range of bioactivities [24]. The drimane sesquiterpenoids isolated from *Aspergillus* spp. have exhibited in vitro anti-inflammatory [5] and antiviral [22] activities as well as cytotoxicity against several mammalian cell lines [4,16,22].

Continuing our chemotaxonomic exploration of unusual Australian species of the genus *Aspergillus* [25-28], a soil survey was completed in the Kingaroy District of the South Burnett region of South East Queensland. One fungal strain, isolated from soil collected near the town of Nanango, Queensland, showed atypical growth patterns with distinct macro- and micro-morphological differences to other Aspergilli. This strain was considered to be a new species, *Aspergillus nanangensis*, belonging to the subgenus *Circumdati*, section *Jani*. The detailed morphological, genomic and chemotaxonomic characterisation of *A. nanangensis* will be reported elsewhere in due course. Herein, we report the isolation, structure elucidation and bioassay of a family of drimane sesquiterpenoids from *A. nanangensis*, which we named the nanangenines. Notably, *A. nanangensis* distinguishes itself within the genus by the production of terpenoids as the dominant biosynthetic class of secondary metabolites.

Results and Discussion

Purification and identification

The metabolite profile of *A. nanangensis* was examined on a limited range of solid and liquid media suitable for fungal metabolite production. The metabolite profile remained consistent despite variations to the carbon and nitrogen sources across agars and liquid media, but the productivity was superior on grains, notably rice and barley. A search of the UV–vis profiles against our in-house library of type species (>25,000 spectra from 2,000 species, including 205 *Aspergillus* type species) and unidentified but metabolically talented fungi (>60,000 spectra from 3,000 species) returned no similar metabolite cohorts, suggesting an unknown species. Individual retention time/UV–vis searches of the dominant 15 secondary metabolites against our in-house pure metabolite library (>7,100 standards) also failed to provide a single known secondary metabolite, further suggesting the strain was a hitherto unaccounted species of *Aspergillus*.

A. nanangensis was cultivated separately on jasmine rice and pearl barley for 21 days, which resulted in confluent and thick mycelial coverage of the grains. Extraction of the grains with acetone, followed by partitioning of the aqueous residue with EtOAc and defatting with hexane, provided an enriched extract of non-polar secondary metabolites. Fractionation by reversed-phase preparative HPLC (Figures S1 and S2 in Supporting Information File 1) yielded ten drimane metabolites shown in Figure 1: one trihydroxylated drimane sesquiterpenoid lactone, nanangenine A (**1**), four drimane lactones bearing C_6/C_8 acyl chains, nanangenines B, C, D and E (**2**, **4**, **5** and **6**), two acylated drimanes bearing isomeric lactones, isonanangenines B and D (**3** and **7**), and three acylated drimanes, nanangenines F–H (**8**–**10**), which are putative biosynthetic intermediates. The structures of **1**–**10** were elucidated by detailed spectroscopic analysis, while absolute configurations were determined by single crystal X-ray diffraction analysis of selected analogues.

High-resolution positive electrospray ionisation mass spectrometry (HRESI(+)MS) analysis of nanangenine A (**1**) revealed an adduct ion ($[M + Na]^+$ *m/z* 305.1363) indicative of a molecular formula $C_{15}H_{22}O_5$ requiring five double bond equivalents (DBE). No distinguishing absorption maxima were observed in the UV–vis spectrum, while absorptions at 3354 and 1738 cm^{-1} in the IR spectrum were indicative of hydroxy and carbonyl groups, respectively. The ^{13}C NMR data for **1** (Table 1) indicated the presence of one carbonyl carbon (δ_C 179.3, C-11) and two olefinic carbons (δ_C 128.2, C-7 and δ_C 131.1, C-8), accounting for two DBE and thus requiring **1** to be tricyclic. The ^1H and ^{13}C NMR data also revealed the presence of one hydroxylated quaternary carbon (C-9), two aliphatic quaternary carbons (C-4, C-10), two oxymethines (C-1, C-6), one aliphatic methine (C-5), one oxymethylene (C-12), two aliphatic methylenes (C-2, C-3), three methyl groups (C-13, C-14, C-15) and three hydroxy groups (1-OH, 6-OH, 9-OH). Detailed analysis of the 2D NMR data for **1** (Table S3 in Supporting Information File 1) confirmed the presence of a drimane sesquiterpenoid lactone scaffold. A search of the literature revealed **1** to be almost identical to strobilactone B, previously reported from *A. ustus* [29], with the only difference being hydroxylation at C-1 in **1**, instead of at C-2. Therefore, the structure of **1** was assigned as shown in Figure 1. The absolute configuration of **1** was confirmed to be 1*R*,5*S*,6*R*,9*R*,13*R* by single crystal X-ray diffraction analysis (Table S2 and Figure S3 in Supporting Information File 1).

HRESI(+)MS analysis of nanangenine B (**2**) revealed an adduct ion ($[M + Na]^+$ *m/z* 403.2096) indicative of a molecular formula $C_{21}H_{32}O_6$. The NMR data for **2** (Table S4 in Supporting Information File 1) were very similar to those for **1**, with the only significant differences being the absence of the 6-OH proton, the presence of additional signals for a C_6 acyl chain, and significant deshielding of H-6 from δ_H 4.30 to 5.47 ppm. Therefore, the structure of **2** was assigned as the 6-*O*-hexanoyl analogue of **1**, as shown in Figure 1. Compound **2** was previously reported in 2014 as an unnamed metabolite from *Aspergillus* sp. IBWF002-96 [5], and we have assigned the trivial name nanangenine B for consistency. The absolute configuration of **2** was

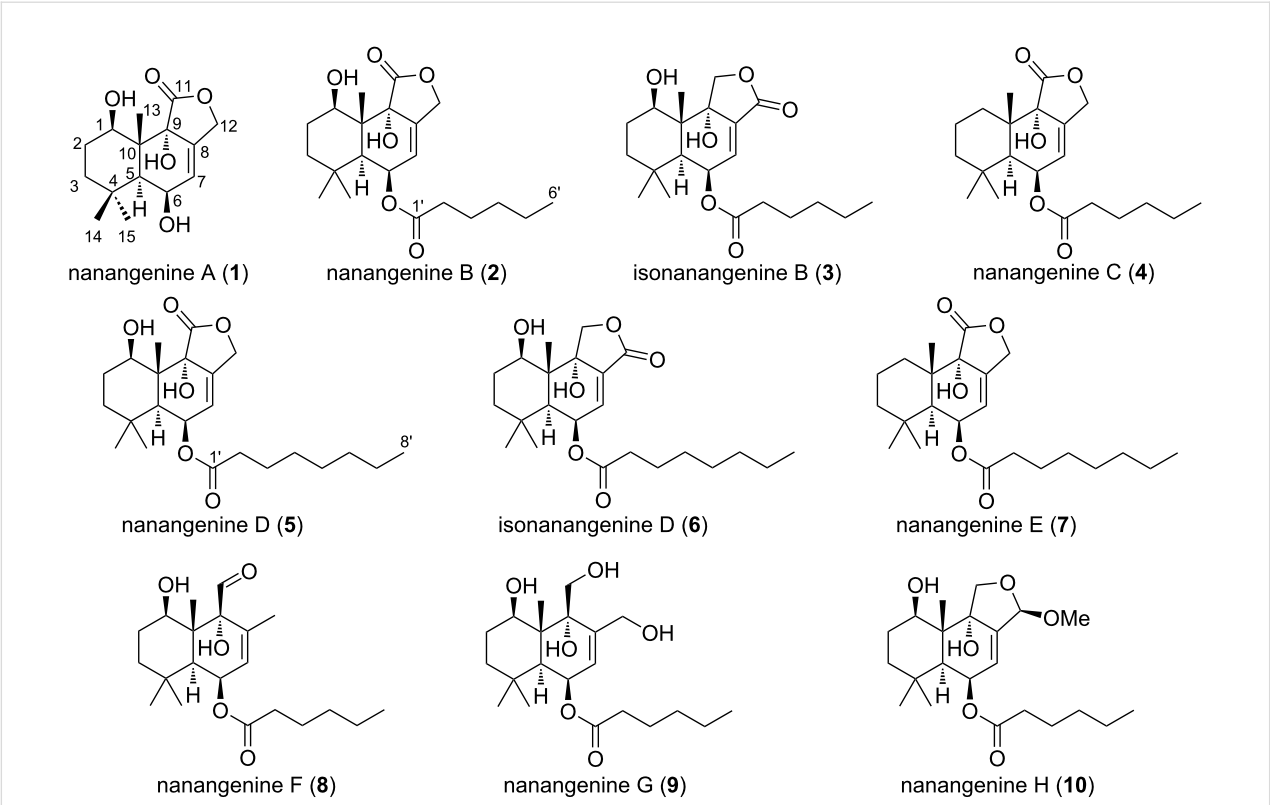


Figure 1: Structures of nanangenines **1**–**10** isolated from *A. nanangensis*.

Table 1: NMR data for nanangenine A (**1**) in DMSO-d_6.

Position	δ_C[a]	δ_H, mult. (*J* in Hz)[b]
1	69.0	4.12, dd (8.4, 7.8)
2	26.2	1.54, m
3	41.5	1.22, m
4	33.6	
5	45.0	1.52, d (5.0)
6	63.3	4.30, m
7	128.2	5.92, m
8	131.1	
9	76.1	
10	42.4	
11	179.3	
12a	70.6	5.00, ddd (12.2, 2.5, 2.4)
12b		4.91, ddd (12.2, 1.3, 1.2)
13	12.6	0.92, s
14	24.3	1.23, s
15	31.9	1.02, s
1-OH		4.76, s
6-OH		4.81, br d (5.8)
9-OH		6.50, s

[a]Acquired at 125 MHz; [b]acquired at 500 MHz.

assigned based on single crystal X-ray diffraction analysis of a 9-*O*-(4-bromobenzoyl) derivative (**2b**; Table S2 and Figure S4 in Supporting Information File 1).

HRESI(+)MS analysis of isonanangenine B (**3**) revealed an adduct ion ($[M + Na]^+$ *m/z* 403.2094) indicative of a molecular formula $C_{21}H_{32}O_6$, which is isomeric with **2**. Comparison of the NMR data for **3** (Table S5 in Supporting Information File 1) with those for **2** revealed the only difference to be the position of the lactone carbonyl group, which was determined to be at C-12 instead of C-11 based on key HBMC correlations from H-7 to C-12 and 9-OH to C-11. Therefore, the structure of **3** was assigned as shown in Figure 1. Compound **3** was previously reported in 2013 as an unnamed metabolite (code number SF002-96-1) from *Aspergillus* sp. IBWF002-96 [4], and we have assigned the trivial name isonanangenine B for consistency. The absolute configuration of **3** was assigned based on single crystal X-ray diffraction analysis of a 1-*O*-(4-bromobenzoyl) derivative (**3b**; Table S2 and Figure S5 in Supporting Information File 1).

HRESI(+)MS analysis of nanangenine C (**4**) revealed an adduct ion ($[M + Na]^+$ *m/z* 387.2147) indicative of a molecular formula $C_{21}H_{32}O_5$, which has one fewer oxygen atom than **2** and **3**.

Comparison of the NMR data for **4** (Table S6 in Supporting Information File 1) with those for **2** revealed the only significant differences to be the absence of the H-1 and 1-OH protons and the presence of an additional methylene group (H_2-1). Therefore, the structure of **4** was assigned to be the 1-deoxy analogue of **2**, as shown in Figure 1. Compound **4** was previously reported in 2014 as an unnamed metabolite from *Aspergillus* sp. IBWF002-96 [5], and we have assigned the trivial name nanangenine C for consistency. The absolute configuration of **4** was assigned based on single crystal X-ray diffraction analysis (Table S2 and Figure S6 in Supporting Information File 1).

Nanangenine D (**5**), isonanangenine D (**6**) and nanangenine E (**7**) had virtually identical spectroscopic data to **2**, **3** and **4**, respectively, with the only significant differences being the presence of two additional resonances in the aliphatic region of the ^{13}C NMR spectra (Table 2) and an increase in area of four protons in the methylene envelopes of the ^{1}H NMR spectra. Therefore, the structures of **5**–**7** were assigned to be the C_8 homologs of **2**–**4**, respectively, as shown in Figure 1. The absolute configurations of **5**–**7** were assigned to be the same as **2**–**4** based on their similar NMR data and optical rotations.

HRESI(+)MS analysis of nanangenine F (**8**) revealed a protonated molecule ($[M + H]^+$ *m*/*z* 367.2483) indicative of a molecular formula $C_{21}H_{34}O_5$, requiring one fewer DBE than **2**. Indeed, the NMR data for **8** (Table 3) were very similar to those for **2**, with the main differences being the absence of signals for the lactone carbonyl and oxymethylene group, and the presence

Table 2: NMR data for nanangenine D (**5**), isonanangenine D (**6**) and nanangenine E (**7**) in DMSO-d_6.

Position	Nanangenine D (**5**)		Isonanangenine D (**6**)		Nanangenine E (**7**)	
	δ_C[a]	δ_H, mult. (*J* in Hz)[b]	δ_C[a]	δ_H, mult. (*J* in Hz)[b]	δ_C[a]	δ_H, mult. (*J* in Hz)[b]
1a	68.7	4.17, ddd (10.9, 5.9, 0.9)	68.6	3.95, ddd (12.0, 5.9, 3.9)	29.4	1.95, ddd (15.0, 13.5, 4.3)
1b						1.81, dm (13.5)
2a	25.9	1.57, m	27.2	1.53, m	17.4	1.59, m
2b						1.47, m
3a	41.3	1.28, m	41.7	1.31, ddd (13.3, 3.5, 3.5)	44.3	1.33, dm (13.0)
3b				1.23, m		1.19, m
4	33.1		32.9		33.3	
5	43.8	1.87, d (4.9)	44.1	1.99, d (5.2)	44.0	1.97, d (4.9)
6	66.0	5.46, m	66.0	5.59, dd (5.1, 3.9)	66.0	5.47, m
7	121.9	5.88, m	131.6	6.45, d (3.9)	121.3	5.77, m
8	135.5		133.2		136.5	
9	75.6		75.0		73.1	
10	42.4		43.2		37.2	
11a	178.4		76.4	4.41, dd (10.2, 0.5)	174.3	
11b				4.19, br d (10.2)		
12a	70.3	5.03, ddd (12.8, 4.8, 2.5)	168.6		68.2	4.87, ddd (12.8, 2.5, 2.5)
12b		4.94, ddd (12.8, 1.7, 1.2)				4.74, ddd (12.8, 1.2, 1.2)
13	12.2	0.94, s	12.3	0.94, s	18.1	0.99, s
14	24.0	1.06, s	24.2	1.08, s	24.2	1.08, s
15	31.5	0.90, s	31.9	0.91, s	32.1	0.91, s
1'	172.2		172.2		172.2	
2'a	34.0	2.33, m	33.9	2.36, m	34.0	2.32, m
2'b		2.27, m		2.27, m		2.25, m
3'	24.2	1.53, m	24.2	1.53, m	24.2	1.52, m
4'	28.2	1.24, m	28.2	1.23, m	28.3	1.24, m
5'	28.3	1.24, m	28.2	1.23, m	28.2	1.24, m
6'	31.0	1.22, m	31.0	1.21, m	31.0	1.21, m
7'	21.9	1.24, m	21.9	1.23, m	21.9	1.24, m
8'	13.9	0.84, t (7.9)	13.9	0.83, t (7.2)	13.9	0.84, t (7.2)
1-OH		4.71, d (0.9)		4.62, d (5.1)		–
9-OH		6.77 s		5.61 br s		6.24 s

[a]Acquired at 150 MHz; [b]acquired at 600 MHz.

Table 3: NMR data for nanangenines F-H (**8**–**10**) in DMSO-d_6.

Position	Nanangenine F (**8**)		Nanangenine G (**9**)		Nanangenine H (**10**)	
	δ_C[a]	δ_H, mult. (*J* in Hz)[b]	δ_C[c]	δ_H, mult. (*J* in Hz)[d]	δ_C[a]	δ_H, mult. (*J* in Hz)[b]
1	68.5	3.88, ddd (11.9, 5.8, 5.8)	69.1	3.93 dd (11.2, 4.3)	68.4	3.93, ddd (11.8, 5.4, 5.0)
2a	27.3	1.51, m	28.3	1.56, m	27.5	1.53, m
2b		1.47, m				1.48, m
3a	41.6	1.26, m	41.6	1.23, m	41.9	1.27, m
3b		1.21, m				1.22, m
4	32.8		33.5		32.8	
5	43.5	1.84, d (4.7)	44.9	1.83, d (4.4)	44.3	1.95, d (4.0)
6	66.5	5.37, m	67.0	5.41, m	67.1	5.46, ddd (5.0, 3.9, 1.4)
7	124.0	5.53, dq (5.1, 1.5)	121.4	5.78, br d (5.3)	120.3	5.57, dd (3.9,1.4)
8	137.2		144.1		145.0	
9	80.1		75.6		78.3	
10	47.0		45.9		43.5	
11a	203.3	9.49, br s	62.2	3.73, m	74.5	3.84, d (9.7)
11b				3.67, m		3.76, d (9.7)
12a	19.5	1.51, dd, (1.3, 1.3)	61.0	4.05, m	101.9	5.31, t (1.4)
12b				4.02, m		
13	11.5	1.06, s	12.5	1.06, s	11.7	0.92, s
14	24.3	1.05, s	24.4	1.03, s	24.3	1.08, s
15	31.8	0.89, s	32.4	0.87, br s	32.1	0.89, s
1'	172.3		172.7		172.3	
2'a	34.1	2.30, m	34.5	2.24, m	34.1	2.32, m
2'b		2.23, m		2.21, m		2.23, m
3'	23.9	1.53, m	24.2	1.53, m	24.0	1.52, m
4'	30.6	1.25, m	30.9	1.25, m	30.5	1.24, m
5'	21.7	1.26, m	22.0	1.26, m	21.7	1.25, m
6'	13.7	0.84, t (7.0)	13.9	0.83, t (7.1)	13.9	0.84, t (7.1)
12-OMe	–	–	–	–	54.1	3.26, s
1-OH		4.74, d, (5.7)		5.07, br s		4.37, d (5.4)
9-OH		5.46, s		4.41, br s		4.93, s
11-OH		–		5.27, br s		–
12-OH		–		4.88, br s		–

[a]Acquired at 150 MHz; [b]acquired at 600 MHz; [c]acquired at 125 MHz; [d]acquired at 500 MHz.

of signals for putative aldehyde (δ_C 203.3; δ_H 9.49, s) and olefinic methyl (δ_C 19.5; δ_H 1.51, dd) groups. Key HMBC correlations (Table S10 in Supporting Information File 1) from 9-OH to C-11 and from H-11 to C-9 positioned the aldehyde on C-9, while HMBC correlations from H-7 to C-12 and from H_3-12 to C-9 positioned the methyl on C-8. Therefore, the structure of **8** was assigned as the seco analogue of **2**, as shown in Figure 1. The absolute configuration of **8** was assigned to be the same as **2** based on their similar NMR data and optical rotations, supported by key ROESY correlations between H_3-13 and H-11, and between H-5 and 9-OH (Table S10 in Supporting Information File 1).

HRESI(+)MS analysis of nanangenine G (**9**) revealed an adduct ion ($[M + Na]^+$ *m/z* 407.2406) indicative of a molecular formula $C_{21}H_{36}O_6$. The NMR data for **9** (Table 3) were very similar to those for **2**, with the main differences being the absence of a signal for the lactone carbonyl group and the presence of additional signals for an additional oxymethylene (δ_C 62.2; δ_H 3.73/3.67, m) and two additional hydroxy groups (δ_H 5.27, br s and 4.88, br s). The ^{1}H NMR signals for oxymethylene H_2-12 were also shifted upfield by ≈1 ppm compared to those in **2**, suggesting the pendant oxygen atom was no longer attached to a carbonyl carbon. A detailed analysis of the 2D NMR data for **9** (Table S11 in Supporting Information File 1) confirmed the structure to be the seco analogue of **2**, as shown in Figure 1. The absolute configuration of **9** was then confirmed to be the same as **2** by single crystal X-ray diffraction analysis (Table S2 and Figure S7 in Supporting Information File 1).

HRESI(+)MS analysis of nanangenine H (**10**) revealed a protonated dehydration product ($[M - H_2O + H]^+$ *m/z* 379.2473) indicative of a molecular formula $C_{22}H_{36}O_6$. The NMR data for **10** (Table 3) were very similar to those for **3**, with the main differences being the absence of a signal for the lactone carbonyl group and the presence of additional signals for acetal (δ_C 101.9; δ_H 5.31, t) and methoxy (δ_C 54.1; δ_H 3.26, s) groups. Key HMBC correlations (Table S12 in Supporting Information File 1) from H-7 to C-12 and from 12-OMe to C-12 positioned the methoxy group and acetal proton on C-12. Thus, the structure of **10** was assigned as shown in Figure 1. The configuration of **10** at C-12 was determined to be 12*R* based on a key ROESY correlation between H-12 and 9-OH (Table S12, Supporting Information File 1), with the absolute configuration of the remainder of the molecule assumed to be the same as **3** based on their similar NMR data and optical rotations.

Bioassays

The nanangenines were assayed for in vitro activity against four mammalian cell lines, two bacteria, one fungus and one plant (Table 4). Compound **1** was inactive up to 100 μg mL^{-1} in all of the assays performed, suggesting acylation at 6-OH is important for biological activity. Compounds **4**, **5** and **7** showed moderate antibacterial activity against *B. subtilis*, with weaker activity observed for **2**, **3**, **6** and **8**. Compounds **2**, **5**, **8**, **9** and **10** exhibited low levels of cytotoxicity against the four mammalian cell lines tested, while **3**, **6** and **7** were considerably more cytotoxic. Notably, **4** showed strong activity against the mouse myeloma NS-1 cell line, but no activity against the three human cell lines. None of the compounds tested showed any activity up to 100 μg mL^{-1} against the Gram-negative bacterium *Escherichia coli* (ATCC 25922), the fungus *Candida albicans* (ATCC 10231) or the plant *Eragrostis tef* (teff).

Proposed biosynthesis and gene cluster

The biosynthesis of drimane-type sesquiterpenoids from farnesyl diphosphate is proposed to proceed via the protonation-initiated mechanism (class II terpene synthases) [24], which is distinct from the ionisation-initiated mechanism (class I) terpene synthases, where a carbocation is generated by the release of a diphosphate group [30]. Therefore, the drimane synthase is likely to be different from the commonly observed sesquiterpene synthase, which belongs to the class I terpene synthases. Recently, the drimane synthase AstC involved in biosynthesis of the astellolides was identified and shown to be a novel member of the terpene synthase family, showing similarity to haloacid dehalogenase (HAD)-like hydrolases [21]. Thus, we suspected that a related enzyme may be involved in the biosynthesis of the nanangenines, and used the amino acid sequence of AstC to probe the *A. nanangensis* genome. We also hypothesised that the acyl side chains present in the (iso)nanangenines could be derived either from a fatty acid synthase (FAS) or polyketide synthase (PKS). For example, in aflatoxin biosynthesis, the hexanoyl started unit is supplied by a FAS [31], while in the meroterpenoid fumagillin biosynthesis, the unsaturated acyl chain is synthesised by a PKS [32].

The *A. nanangensis* genome was sequenced and a draft assembly of the genome was generated. A local BLASTp search of the *A. nanangensis* genome using the drimane synthase AstC as query returned a hit on scaffold 3, FE257_006542, which was immediately flanked by a highly-reducing PKS (FE257_006541) and a FAD-binding oxidoreductase (FE257_006543). AstC, though annotated as a HAD-like hydrolase and having low sequence homology to other characterised terpene synthases, contains sequence motifs conserved across both class I (DDxxD/E) and class II (DxDD,

Table 4: In vitro bioassay results for nanangenines **1**–**10**.

Compound	IC_{50} (μg mL^{-1})[a]				
	Bs[b]	NS-1[c]	DU-145[d]	MCF-7[e]	Nff[f]
1	>100	>100	>100	>100	>100
2	62 ± 2	38 ± 1.5	>100	>100	>100
3	56 ± 6	0.16 ± 0.01	0.9 ± 0.2	0.19 ± 0.05	0.9 ± 0.1
4	5.7 ± 0.1	4.1 ± 0.1	>100	>100	>100
5	9.4 ± 0.4	19 ± 2	37 ± 2	22 ± 1	37 ± 2
6	4%[g]	1.0 ± 0.1	32 ± 3	3.6 ± 0.4	2.1 ± 0.3
7	2.3 ± 0.3	2.4 ± 0.1	15 ± 1	9.4 ± 0.7	6.8 ± 0.6
8	78 ± 2	49 ± 2	95 ± 3	49 ± 1	84 ± 2
9	>100	47 ± 2	>100	>100	>100
10	>100	10 ± 2	7%[g]	62 ± 2	60 ± 1
control	0.4[h]	0.02[i]	18[i]	2[i]	0.2[i]

[a]Experiments were conducted in triplicate. IC_{50} values are mean ± standard error; [b]*Bacillus subtilis* (ATCC 6633); [c]mouse myeloma NS-1 cell line (ATCC TIB-18); [d]human prostate cancer DU-145 cell line (ATCC HTB-81); [e]Human breast adenocarcinoma MCF-7 cell line (ATCC HTB-22); [f]Human fibroblast NFF cell line (ATCC PCS-201); [g]incomplete dose response, % inhibition reached at 100 μg mL^{-1}; [h]tetracycline; [i]staurosporine.

QW) terpene synthases [33-35]. Specifically, AstC contains a DxDTT motif [21], which is a variant of the DxDD motif known to be involved in class II-type protonation-initiated terpene cyclisation [36]. Importantly, the DDxxD motif in AstC contains a substitution of the second Asp for Asn, leading to a loss of ionisation-activated cyclisation activity. Two other HAD-like hydrolases, AstI and AstK, with the class I (DDxxD/E) motif, are therefore required to catalyse other dephosphorylation steps. Despite overall low sequence homology, alignment with the amino acid sequences of AstC, AstI, AstK and related homologs confirmed the presence of both conserved class I and class II motifs in FE257_006542 (Figure S50 and Table S15 in Supporting Information File 1).

Drimane sesquiterpenoids are produced by several species across the *Aspergillus* genus, with compounds possessing acyl side chains observed specifically in section *Usti*, such as *A. ustus* [37,38], *A. insuetus* [39], *A. pseudodeflectus* [40] and other unnamed species [17]. *A. calidoustus*, another member of section *Usti*, also produces drimane sesquiterpenoids, though no compounds with acyl side chains have yet been reported [41]. Notably, the acyl chains that these compounds possess are unsaturated polyenes, while the acyl chains of the nanangenines are fully saturated. As noted earlier, **2**–**4** were previously reported as unnamed metabolites from *Aspergillus* sp. IBWF002-96 [4,5], which has an internal transcribed spacer (ITS) sequence that shares 100% identity with an uncultured soil sample (NCBI accession GQ921753.1) and 97% identity with *A. janus* (NCBI accession EU021598). The ITS sequence of *A. nanangensis* (deposited on NCBI, accession MK979278) also shares 100% sequence identity with the uncultured soil sample, and 97% identity with *A. janus* (Figure S51 in Supporting Information File 1). No public sequence could be found for IBWF002-96. Interestingly, the uncultured soil sample was also isolated from soil near the town of Nanango in Queensland, Australia, as part of a survey of soil fungal communities in monoculture *Araucaria cunninghamii* plantations, though it remained unidentified [42]. While we cannot conclusively determine each strain to be *A. nanangensis*, the ITS sequences of all three strains are identical and they clearly clade in section *Jani* (Figure S48 in Supporting Information File 1).

A comparative genomics survey of all publicly available genomes on NCBI and the Joint Genome Institute (JGI) genome portal of known drimane sesquiterpenoid producers was undertaken (Figure S49 in Supporting Information File 1). Homologous gene clusters were observed in all drimane sesquiterpenoid-producing members of section *Usti*: *A. ustus* CBS 3.3904, *A. calidoustus* CBS 12160 and *A. insuetus* CBS 107.25 (Figure 2). A homologous cluster was also found in *A. pseudodeflectus* CBS 756.74, though it is located on a truncated assembly scaffold. In total, six genes are conserved across these species: the HR-PKS (FE257_006541), HAD-like terpene synthase (FE257_006542) and FAD-binding oxidoreductase (FE257_006543), as well as a cytochrome P450 (FE257_006544), an alpha/beta hydrolase (FE257_006545) and a short-chain dehydrogenase (FE257_006546). Notably, no homologous clusters were identified in other members of section *Jani* (*A. brevijanus* and *A. janus*) that have not been reported to produce such drimane sesquiterpenoids. A phylogenetic tree of all AstC homologs identified in the genomes of all known drimane sesquiterpenoid producers was constructed, in which

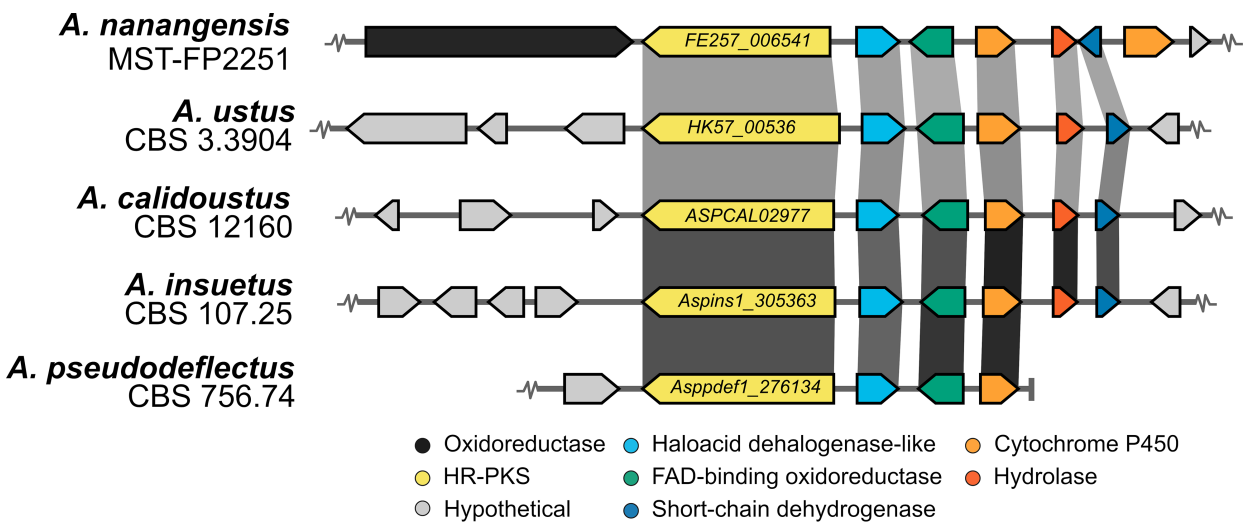


Figure 2: Putative nanangenine biosynthetic gene cluster in *A. nanangensis* MST-FP2251 and homologs identified in other drimane sesquiterpenoid-producing species of *Aspergillus* section *Usti*. Gene models are drawn to scale; shaded boxes that link gene models represent amino acid identity (0% transparent; 100% black).

the putative HAD-like terpene synthases (drimane synthases) from the section *Usti* form a distinct clade, separate from the AstC type (Figure S52 in Supporting Information File 1). The analysis above, and the discovery of other homologous gene clusters in *Aspergillus* spp. that are known to produce acyl drimane sesquiterpenoids, support the identification of this as the putative biosynthetic gene cluster for nanangenines.

Based on the analyses above, a biosynthetic pathway to the nanangenines was proposed (Figure 3). Unlike the *ast* cluster, where there are multiple HAD-like enzymes encoded (one terpene synthase and two phosphatases), the putative nanangenine cluster only encodes one such enzyme, FE257_006542. However, given that FE257_006542 contains both class I (DDxxD/E) and class II (DxDD, QW) terpene synthases, we propose that FE257_006542 is responsible for both the cyclisation into drimanyl diphosphate and the hydrolysis of the diphosphate into drim-8-ene-11-ol. We propose the next step in the pathway is hydroxylation at C-6 or C-9, as hydroxylation at both sites is common to all of the (iso)nanangenines. The 9-hydroxylation also results in migration of the double bond on the decalin to $\Delta^{7,8}$. The two hydroxylations could be catalysed by the FAD-dependent oxidoreductase or one of the cytochrome P450 oxygenases. From this point, the pathway branches out to the different (iso)nananagenines via combinations of hydroxylation at C-1 (or not) and one or multiple oxidations at C-11 and C-12. The oxidations of methyl to alcohol and alcohol to aldehyde could be catalysed by one of the two P450 oxygenases, or the short-chain dehydrogenase could oxidise the alcohol (at C-11 or C-12) to an aldehyde. Depending on whether C-11 or C-12 formed an aldehyde, further oxidation of the aldehyde to a carboxylic acid and condensation with the γ-OH group would afford the butyrolactone ring in nanagenines A–E and isonanagenines B/D, respectively. The C-6 and C-8 lipid chain is likely produced by the HR-PKS encoded by gene FE257_006541, while the acylation could be attributed to the enzyme encoded by FE257_006545, which contains the conserved domain of the alpha/beta-hydrolase fold superfamily (which includes thioesterases and acyltransferases, e.g., LovD [43]). Therefore, the identified putative gene cluster encodes all the genes predicted to be required for biosynthesis of the (iso)nanangenines. We are currently in the process of verifying

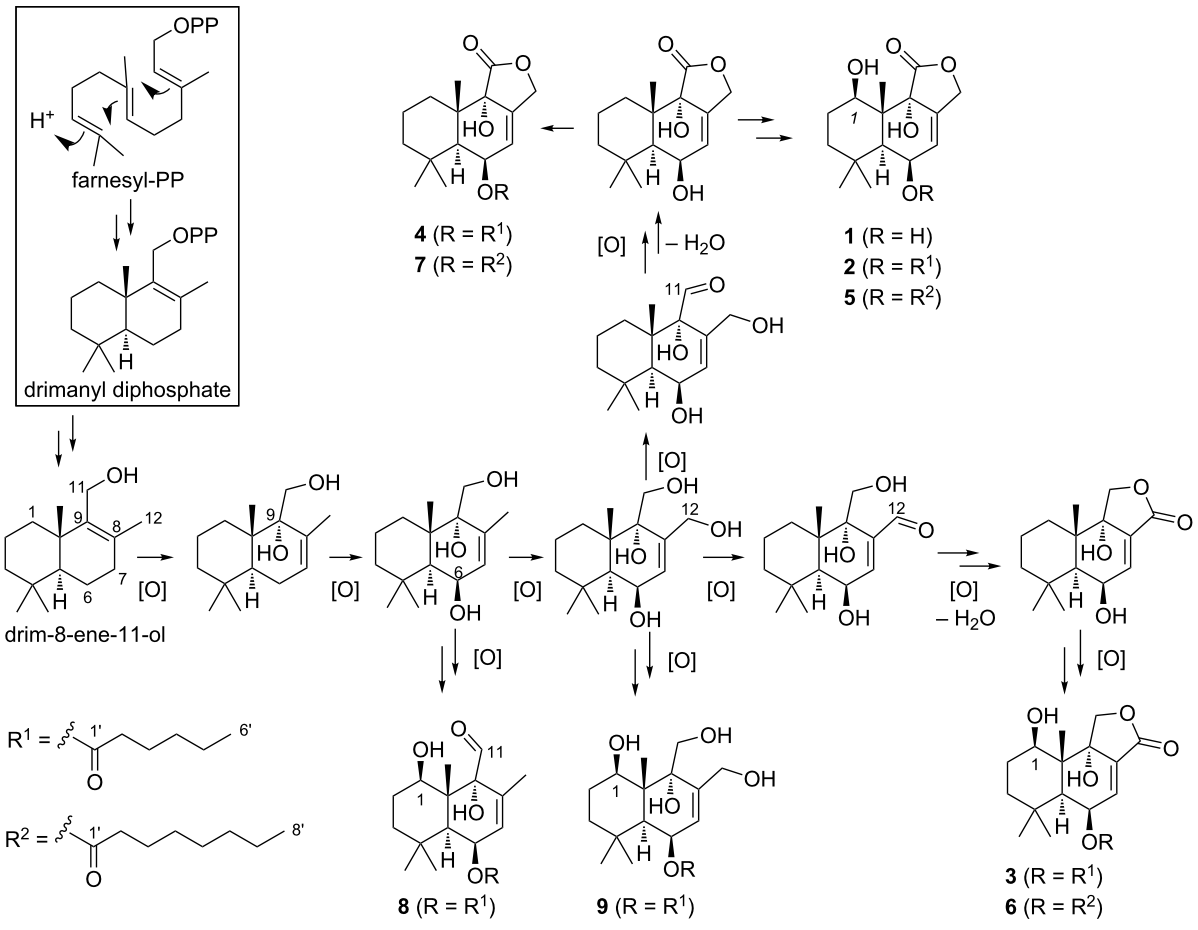


Figure 3: Putative biosynthetic pathway to the nanangenines.

the putative biosynthetic gene cluster by genetic deletion and heterologous expression.

Conclusion

Our investigations into unique Australian microbial biodiversity have led to discovery of a family of drimane sesquiterpenoids, the nanangenines, which serve as key chemotaxonomic markers for the novel fungal species, *A. nanangensis*. Although related drimane sesquiterpenoid lactones have been found in several Aspergilli, drimane sesquiterpenoids with saturated fatty acyl chains have only been reported in one *Aspergillus* strain (IBWF002-96), which appears to share an identical ITS sequence with *A. nanangensis* [4,5]. Interestingly, drimane sesquiterpenoids are not produced uniformly within the *Aspergillus* genus, instead clustering within the sections *Usti*, *Circumdati* and *Flavi* (Figure S49 in Supporting Information File 1). The drimane sesquiterpenoids are not among the dominant metabolite classes commonly used for *Aspergillus* chemotaxonomy, but have been used for chemotyping of Aspergilli in the section *Usti* [17]. Here, we showed for the first time that an *Aspergillus* species from the section *Jani* produces such drimane sesquiterpenoids and has a secondary metabolite repertoire that is distinct from close members in the section *Jani*. Comparative analysis identified a group of putative homologous acyl drimane sesquiterpenoid biosynthetic gene clusters that are shared among *A. nanangensis* and Aspergilli in the section *Usti*. Interestingly, the putative nanangenine biosynthetic gene cluster, and those homologs in *Aspergillus* section *Usti*, only shared low homology to the astellolide biosynthetic gene cluster identified in *A. oryzae*. Further investigation by genetic knockout or heterologous expression is required to confirm the identity of the putative nanangenine biosynthetic gene cluster. Over the range of biological assays performed, isonanangenines **3** and **6**, in particular, exhibited strong activity against human and murine tumour cell lines. This suggests that the regioisomeric lactone moiety could play an important role in the observed cytotoxicity. Interestingly, **3** (SF002-96-1 [4]) was previously shown to inhibit survivin, which is a member of the inhibitor of apoptosis (IAP) family and a potential cancer target. Further structure–activity relationship studies are required to fully characterise this potential anticancer pharmacophore. This study again demonstrates how taxonomy-guided exploration of fungi is a highly effective strategy for identifying novel bioactive metabolites.

Experimental

General experimental details

Optical rotations were acquired in MeOH on a Perkin-Elmer Model 341 polarimeter in a 50 × 5 mm cell or on a Jasco P-1010 polarimeter in a 100 × 3.5 mm cell. UV–vis spectra were acquired in MeCN on a Varian Cary 4000 spectrophotometer or a Jasco V-760 spectrophotometer in a 10 × 10 mm quartz cuvette. Analytical HPLC was performed on a gradient Shimadzu VP HPLC system equipped with a Shimadzu SPD-M10A VP diode array detector and an LC-10AT VP gradient chromatograph. Preparative HPLC was performed on a gradient Shimadzu HPLC system comprising two LC-8A preparative liquid pumps with static mixer, SPD-M10AVP diode array detector and SCL-10AVP system controller with standard Rheodyne injection port. The columns used in the purification of the metabolites were selected from either a Hypersil C_{18} column (50 × 100 mm, 5 µm; Grace Discovery), a Vydac C_{18} column (50 × 100 mm, 5 µm; Grace Discovery), a Zorbax SB-C_{18} column (50 × 150 mm, 5 µm; Agilent) or an Alltima C_{18} column (22 × 250 mm, 5 µm, Grace Discovery), eluted isocratically with acetonitrile/water mixtures with or without 0.01% TFA modifier, as described for each separation. LCMS was performed on an Agilent 1260 Infinity series HPLC equipped with an Agilent 6130 Infinity series single quadrupole mass detector in both positive and negative ion modes. High-resolution electrospray ionisation mass spectra (HRESIMS) were obtained on a Bruker Apex Qe 7T Fourier Transform Ion Cyclotron Resonance mass spectrometer equipped with an Apollo II ESI/MALDI Dual source or a Q Exactive Plus hybrid quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) by direct infusion. NMR data were recorded in DMSO-d_6 on either a Bruker Avance III 500 or a Bruker Avance II DRX-600K spectrometer. All NMR spectra were recorded at 25 °C, processed using Bruker Topspin 3.5 software and referenced to residual non-deuterated solvent signals (DMSO-d_6: δ_H 2.49, δ_C 39.5 ppm).

Collection and cultivation

A. nanangensis MST-FP2251 was isolated by serial dilution of an aqueous suspension of soil collected in Nanango, Queensland, Australia in 2004. Cultures were freeze-dried and accessioned into the CSIRO Agriculture and Food (FRR) culture collection, North Ryde, NSW, Australia, with the code FRR6048. The growth of *A. nanangensis* was optimised on four different agar and liquid media and four types of grains. Agar and liquid versions of glycerol casein (AC), Czapek-Dox (CZ), malt extract (MA) and yeast extract (YS) media, and the grains, pearl barley (PB), cracked wheat (BL), jasmine rice (JR) and basmati rice (BS) were prepared in accordance with recipes provided in Table S1 in Supporting Information File 1. The culture was recovered from storage at −80 °C onto three malt agar plates. At day 7 the recovery plates showed good confluent growth and no signs of contamination. Agar plates were sliced and deposited into three Erlenmeyer flasks containing sterilised Milli-Q water (100 mL). A spore suspension was created by shaking for 1 h prior to inoculation. The agar plates were inoculated with spore suspension (500 µL), liquid media (50 mL)

were inoculated with spore suspension (2 mL), and grains were inoculated with spore suspension (10 mL). The grain, agar and liquid media were placed in a dark room at 24 °C, with the liquid media flasks on a shaker pad rotating at 90 rpm. The grains and agar cultivars were sub-sampled on days 7, 14 and 21, while the liquids were sub-sampled on day 7. The sub-samples were extracted in MeOH for 2 h prior to being analysed by analytical HPLC (C_{18}) to determine the secondary metabolite profile.

Preparative cultivation, extraction and isolation

Two preparative-scale cultivations were carried out to produce sufficient quantities of the nanangenine family for characterisation and bioassay. Cultivation conditions were identical with the exception of growth medium. Culture A was prepared with jasmine rice (1.5 kg) and Culture B with pearl barley (1.5 kg). Each grain was hydrated during sterilisation (1 L water; 120 °C for 40 min). A spore suspension of a 7-day-old Petri plate of *A. nanangensis* was used to inoculate 42 × 250 mL Erlenmeyer flasks each containing 75 g of sterile medium. The flasks were incubated at 24 °C for 21 days, by which time the cultures had grown extensively throughout the grain and reached maximal metabolite productivity. Cultures were extracted separately using the same process (see Figures S1 and S2 in Supporting Information File 1). Grains were pooled from individual flasks, and extracted with acetone (2 × 3 L). The combined extracts were evaporated under vacuum to produce an aqueous slurry (500 mL). The slurry was partitioned against ethyl acetate (2 × 2 L) and the ethyl acetate was reduced in vacuo to an oily residue. The residue was dissolved in methanol (300 mL) and defatted using hexanes (2 × 500 mL) to give the crude extract.

Initial fractionation of culture A (13.5 g) was achieved using Sephadex LH-20 size-exclusion chromatography in methanol (25 × 600 mm). The crude extract was loaded onto the column and eluted with methanol at a flow rate of 4 mL min^{-1}. The fractions (100 mL) were sub-sampled and analysed by C_{18} analytical HPLC. Fractions 3 and 4 were pooled and evaporated to give a solid residue (3.7 g), as were fractions 5–7 (7.8 g). Fraction 3–4 was dissolved in MeOH (8 mL) and insoluble material was pelleted by centrifugation (4000 rpm for 7 min). The supernatant was fractionated by preparative HPLC (Hypersil C_{18}, isocratic 70% MeCN/H_2O, 60 mL min^{-1}) to give an enriched fraction (302.4 mg), which was further purified by preparative HPLC (Hypersil C_{18}, isocratic 55% MeCN/H_2O, 60 mL min^{-1}) to yield nanangenine B (**2**; t_R = 41.0 min; 92.1 mg). Fraction 5–7 was dissolved in MeOH (9 mL) and insoluble material was pelleted by centrifugation (4000 rpm for 7 min). The supernatant was fractionated by preparative HPLC (Hypersil C_{18}, isocratic 45% MeCN/H_2O containing 0.01% TFA, 60 mL min^{-1}) which yielded four subfractions (1, 3, 4 and 7) regarded as candidates for further purification. Subfraction 1 (337.8 mg) was purified by preparative HPLC (Hypersil C_{18}, isocratic 27.5% MeCN/H_2O containing 0.01% TFA, 60 mL min^{-1}) to give an enriched fraction (53.2 mg), which was further fractionated by preparative HPLC (Alltima C_{18}, isocratic 25% MeCN/H_2O containing 0.01% TFA, 60 mL min^{-1}) to yield nanangenine A (**1**; t_R = 12.6 min; 25.9 mg). Subfraction 3 (942.1 mg) was purified by preparative HPLC (Hypersil C_{18}, isocratic 55% MeCN/H_2O, 60 mL min^{-1}) to give an enriched fraction, which was further fractionated by preparative HPLC (Alltima C_{18}, isocratic 50% MeCN/H_2O, 60 mL min^{-1}) to yield nanangenine G (**9**; t_R = 16.5 min; 211.3 mg). Subfraction 4 (632.0 mg) was purified by preparative HPLC (Hypersil C_{18}, isocratic 60% MeCN/H_2O, 60 mL min^{-1}) to yield isonanangenine B (**3**; t_R = 19.1 min; 106.3 mg). Subfraction 7 (490.1 mg) was purified by preparative HPLC (Hypersil C_{18}, isocratic 80% MeCN/H_2O containing 0.01% TFA, 60 mL min^{-1}) to yield nanangenine C (**4**; t_R = 11.5 min; 117.4 mg).

Initial fractionation of culture B (13.5 g) was achieved using 100 g silica gel (Davisil, Grace Discovery 50 μm) with 20 g of silica used to dry load the sample atop the column bed (55 × 250 mm). The column was eluted with $CHCl_3$ (500 mL) containing increasing concentrations of MeOH. The fractions containing terpenoids (0-2% MeOH, 3.0 g) were pooled and fractionated using isocratic preparative HPLC (Zorbax C_{18}, 65% MeCN/H_2O, 60 mL min^{-1}) to give five fractions containing terpenoids. Fraction 1 (785.3 mg) was purified by isocratic preparative HPLC (Zorbax C_{18}, isocratic 45% MeCN/H_2O, 20 mL min^{-1}) to yield nanangenine H (**12**; t_R = 15.6 min; 41.7 mg). Fraction 2 (93.7 mg) was purified by isocratic preparative HPLC (Zorbax C_{18}, isocratic 70% MeCN/H_2O, 20 mL min^{-1}) to yield nanangenine F (**8**; t_R = 12.9 min; 7.2 mg). Fraction 3 (30.6 mg) was purified by isocratic preparative HPLC (Zorbax C_{18}, isocratic 70% MeCN/H_2O, 20 mL min^{-1}) to yield isonanangenine D (**6**; t_R = 15.8 min; 10.7 mg). Fraction 4 (48.8 mg) was purified by isocratic preparative HPLC (Zorbax C_{18}, isocratic 70% MeCN/H_2O, 20 mL min^{-1}) to yield nanangenine D (**5**; t_R = 21.8 min; 29.9 mg). Fraction 5 (24.4 mg) was purified by isocratic preparative HPLC (Zorbax C_{18}, isocratic 75% MeCN/H_2O, 20 mL min^{-1}) to yield nanangenine E (**7**; t_R = 12.8 min; 8.7 mg).

Characterisation of compounds

Nanangenine A (1): white powder; $[\alpha]_D^{20}$ −205 (*c* 0.03, MeOH); UV (MeCN) λ_{max} (log ε) 200 (4.19); 212 (3.71) nm; HRMS–ESI (+, *m/z*): $[M + Na]^+$ calcd. for $C_{15}H_{22}NaO_5^+$, 305.1359; found, 305.1363.

Nanangenine B (2): white powder; $[\alpha]_D^{20}$ −226 (*c* 0.06, MeOH); UV (MeCN) λ_{max} (log ε) 200 (4.08); 208 (3.76) nm; HRMS–ESI (+, *m*/*z*): $[M + Na]^+$ calcd. for $C_{21}H_{32}NaO_6^+$, 403.2091; found, 403.2096.

Isonanangenine B (3): white powder; $[\alpha]_D^{20}$ −180 (*c* 0.04, MeOH); UV (MeCN) λ_{max} (log ε) 200 (4.42); 207 (4.21) nm; HRMS–ESI (+, *m*/*z*): $[M + Na]^+$ calcd. for $C_{21}H_{32}NaO_6^+$, 403.2091; found, 403.2094.

Nanangenine C (4): white powder; $[\alpha]_D^{20}$ −289 (*c* 0.2, MeOH); UV (MeCN) λ_{max} (log ε) 200 (4.09); 208 (3.91) nm. HRMS–ESI (+, *m*/*z*): $[M + Na]^+$ calcd. for $C_{21}H_{32}NaO_5^+$, 387.2142; found, 387.2147.

Nanangenine D (5): white powder; $[\alpha]_D^{24}$ −246 (*c* 0.28, MeOH); UV (MeCN) λ_{max} (log ε) 200 (3.90); 208 (3.81) nm; HRMS–ESI (+, *m*/*z*): $[M + H]^+$ calcd. for $C_{23}H_{37}O_6^+$, 409.2585; found, 409.2579.

Isonanangenine D (6): white powder; $[\alpha]_D^{24}$ −234 (*c* 0.13, MeOH); UV (MeCN) λ_{max} (log ε) 200 (4.47); 206.5 (4.42) nm; HRMS–ESI (+, *m*/*z*): $[M + H]^+$ calcd. for $C_{23}H_{37}O_6^+$, 409.2585; found, 409.2581.

Nanangenine E (7): white powder; $[\alpha]_D^{24}$ −285 (*c* 0.18, MeOH); UV (MeCN) λ_{max} (log ε) 200 (4.17); 206.5 (4.08) nm; HRMS–ESI (+, *m*/*z*): $[M + H]^+$ calcd. for $C_{23}H_{37}O_5^+$, 393.2636; found, 393.2628.

Nanangenine F (8): white powder; $[\alpha]_D^{24}$ −268 (*c* 0.24, MeOH); UV (MeCN) λ_{max} (log ε) 200 (3.96); 203 (3.82) nm; HRMS–ESI (+, *m*/*z*): $[M + H]^+$ calcd. for $C_{21}H_{35}O_5^+$ $[M + H]^+$, 367.2479; found, 367.2483.

Nanangenine G (9): white powder; $[\alpha]_D^{20}$ −231 (*c* 0.05, MeOH); UV (MeCN) λ_{max} (log ε) 200 (4.17) nm; HRMS–ESI (+, *m*/*z*): $[M + Na]^+$ calcd. for $C_{21}H_{36}NaO_6^+$ $[M + Na]^+$, 407.2404; found, 407.2406.

Nanangenine H (10): white powder; $[\alpha]_D^{24}$ −270 (*c* 0.13, MeOH); UV (MeCN) λ_{max} (log ε) 200 (3.98) nm; HRMS–ESI (+, *m*/*z*): $[M - H_2O + H]^+$ calcd. for $C_{22}H_{35}O_5^+$ $[M - H_2O + H]^+$, 379.2479; found, 379.2473.

Crystallography

Crystals suitable for single crystal X-ray diffraction analysis were obtained for **1**, **4** and **9**, providing confirmation of the spectroscopic structural elucidation. Indirect confirmation of the structures of **2** and **3** was obtained following their conversion to 4-bromobenzoyl ester derivatives, with substitution at the C-9 and C-1 hydroxy groups respectively, and the growth of crystals of these derivatives (**2b** and **3b**). The X-ray diffraction data obtained from single crystals of **4** and **9** did not allow a definitive determination of their absolute structures, however their anomalous dispersion statistics indicate that the assignments are very likely to be correct. The crystal structure of **1** has four crystallographically-independent molecules, while the structures of **2b** and **3b** have two crystallographically-independent molecules, and the structures of **4** and **9** have one. The cyclohexane rings of all molecules have chair conformations, the cyclohexene rings of all molecules have half-chair conformations, and the furan rings of all molecules have flattened envelope conformations. See Supporting Information File 1 for detailed methods.

Genomics and bioinformatic analysis

Genomic DNA was extracted from *A. nanangensis* (see Supporting Information File 1) and sent to the Australian Genome Research Facilities for de novo genome sequencing using Illumina HiSeq 2000. A draft genome assembly of *A. nanangensis* was obtained using SPAdes v3.13.0 via the AAFTF pipeline (https://github.com/stajichlab/AAFTF). To facilitate the identification of a putative biosynthetic gene cluster, a local BLAST database was created from the resulting assembly [44]. Though Shinohara et al. list the revised annotation of AstC as AORIB40_05408, its amino acid sequence of was retrieved from AspGD [45] under the locus tag AO090026000582. AstC homologs were identified in *A. nanangensis* via the tBLASTn [44] functionality in Geneious 10.2.6 [46]. The genome sequences for drimane sesquiterpenoid producing Aspergilli were obtained from NCBI where available. This included *A. calidoustus* SF006504 (GCA_001511075.1), *A. ochraceus* fc-1 (GCA_004849945.1), *A. parasiticus* SU-1 (GCA_000956085.1), *A. sclerotiorum* HBR18 (GCA_000530345.1) and *A. ustus* 3.3904 (GCA_000812125.1). Homologous gene clusters were identified in these genomes using a custom Python script, named clusterblaster (https://github.com/gamcil/clusterblaster). A DIAMOND [47] database was built from protein sequences extracted from these genomes, and clusterblaster was used to rapidly identify biosynthetic gene cluster homologs in *A. ustus* and *A. calidoustus*. Genomes for *A. insuetus* CBS 107.25 and *A. pseudodeflectus* CBS 756.74 were obtained from the Joint Genome Institute (JGI) under genome portals Aspins1 and Asppdef1, respectively. Manual tBLASTn searches using the putative nanangenine biosynthetic gene cluster sequences were performed to identify homologs in these genomes. Homology between these clusters was visualised using another custom Python script named crosslinker (https://github.com/gamcil/crosslinker). Further details regarding clusterblaster and crosslinker are given in Supporting Information File 1.

Supporting Information

Supporting Information File 1

Details of cultivation media, fractionation schemes, NMR spectra and tabulated 2D NMR data for all compounds, detailed X-ray crystallographic details and CCDC deposition numbers, bioassay procedures and genomic data.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-256-S1.pdf]

Supporting Information File 2

Crystal structure information files for nanangenines A–C and G.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-256-S2.zip]

Acknowledgements

We thank Dr Matthew McKay (MQ) and Dr Nick Proschogo (USyd) for acquisition of HRMS data, Dr Ian Luck (USyd) for acquisition of preliminary NMR data and Si Shu (UWA) for preparation of genomic DNA for genome sequencing. This work was funded, in part, by the Australian Research Council (FT130100142, FT160100233) and the Cooperative Research Centres Projects scheme (CRCPFIVE000119).

ORCID® iDs

Peter J. Rutledge - https://orcid.org/0000-0002-0767-5196
Ernest Lacey - https://orcid.org/0000-0002-5387-2721
Yit-Heng Chooi - https://orcid.org/0000-0001-7719-7524
Andrew M. Piggott - https://orcid.org/0000-0002-5308-5314

References

1. Heddergott, C.; Calvo, A. M.; Latgé, J. P. *Eukaryotic Cell* **2014,** *13,* 1014–1025. doi:10.1128/ec.00074-14
2. Huang, J.-H.; Lv, J.-M.; Wang, Q.-Z.; Zou, J.; Lu, Y.-J.; Wang, Q.-L.; Chen, D.-N.; Yao, X.-S.; Gao, H.; Hu, D. *Org. Biomol. Chem.* **2019,** *17,* 248–251. doi:10.1039/c8ob02832j
3. Liu, X.-H.; Miao, F.-P.; Qiao, M.-F.; Cichewicz, R. H.; Ji, N.-Y. *RSC Adv.* **2013,** *3,* 588–595. doi:10.1039/c2ra22701k
4. Felix, S.; Sandjo, L. P.; Opatz, T.; Erkel, G. *Beilstein J. Org. Chem.* **2013,** *9,* 2866–2876. doi:10.3762/bjoc.9.323
5. Felix, S.; Sandjo, L. P.; Opatz, T.; Erkel, G. *Bioorg. Med. Chem.* **2014,** *22,* 2912–2918. doi:10.1016/j.bmc.2014.04.015
6. de Souza, J. J.; Vieira, I. J. C.; Rodrigues-Filho, E.; Braz-Filho, R. *Molecules* **2011,** *16,* 10604–10618. doi:10.3390/molecules161210604
7. Cueto, M.; Jensen, P. R.; Fenical, W. *Org. Lett.* **2002,** *4,* 1583–1585. doi:10.1021/ol0258076
8. Hensens, O. D.; Zink, D.; Williamson, J. M.; Lotti, V. J.; Chang, R. S. L.; Goetz, M. A. *J. Org. Chem.* **1991,** *56,* 3399–3403. doi:10.1021/jo00010a040
9. Zhang, D.; Fukuzawa, S.; Satake, M.; Li, X.; Kuranaga, T.; Niitsu, A.; Yoshizawa, K.; Tachibana, K. *Nat. Prod. Commun.* **2012,** *7,* 1411–1414. doi:10.1177/1934578x1200701102
10. Afiyatullov, S.; Zhuravleva, O. I.; Antonov, A. S.; Kalinovsky, A. I.; Pivkin, M. V.; Menchinskaya, E. S.; Aminin, D. L. *Nat. Prod. Commun.* **2012,** *7,* 497–500. doi:10.1177/1934578x1200700421
11. Bai, Z.-Q.; Lin, X.; Wang, J.; Zhou, X.; Liu, J.; Yang, B.; Yang, X.; Liao, S.; Wang, L.; Liu, Y. *Mar. Drugs* **2015,** *13,* 237–248. doi:10.3390/md13010237
12. Liaw, C.-C.; Yang, Y.-L.; Lin, C.-K.; Lee, J.-C.; Liao, W.-Y.; Shen, C.-N.; Sheu, J.-H.; Wu, S.-H. *Org. Lett.* **2015,** *17,* 2330–2333. doi:10.1021/acs.orglett.5b00739
13. Liu, Z.; Liu, H.; Chen, Y.; She, Z. *Nat. Prod. Res.* **2018,** *32,* 2652–2656. doi:10.1080/14786419.2017.1375924
14. Qi, C.; Bao, J.; Wang, J.; Zhu, H.; Xue, Y.; Wang, X.; Li, H.; Sun, W.; Gao, W.; Lai, Y.; Chen, J.-G.; Zhang, Y. *Chem. Sci.* **2016,** *7,* 6563–6572. doi:10.1039/c6sc02464e
15. Shaaban, M.; El-Metwally, M. M.; Abdel-Razek, A. A.; Laatsch, H. *Nat. Prod. Res.* **2018,** *32,* 2437–2446. doi:10.1080/14786419.2017.1419230
16. Liu, H.; Edrada-Ebel, R.; Ebel, R.; Wang, Y.; Schulz, B.; Draeger, S.; Müller, W. E. G.; Wray, V.; Lin, W.; Proksch, P. *J. Nat. Prod.* **2009,** *72,* 1585–1588. doi:10.1021/np900220r
17. Kozlovskii, A. G.; Antipova, T. V.; Zhelifonova, V. P.; Baskunov, B. P.; Ivanushkina, N. E.; Kochkina, G. A.; Ozerskaya, S. M. *Microbiology (Moscow, Russ. Fed.)* **2017,** *86,* 176–182. doi:10.1134/s0026261717020114
18. Rahbæk, L.; Christophersen, C.; Frisvad, J.; Bengaard, H. S.; Larsen, S.; Rassing, B. R. *J. Nat. Prod.* **1997,** *60,* 811–813. doi:10.1021/np970142f
19. Gould, R. O.; Simpson, T. J.; Walkinshaw, M. D. *Tetrahedron Lett.* **1981,** *22,* 1047–1050. doi:10.1016/s0040-4039(01)82862-8
20. Hamasaki, T.; Kuwano, H.; Isono, K.; Hatsuda, Y.; Fukuyama, K.; Tsukihara, T.; Katsube, Y. *Agric. Biol. Chem.* **1975,** *39,* 749–751. doi:10.1080/00021369.1975.10861681
21. Shinohara, Y.; Takahashi, S.; Osada, H.; Koyama, Y. *Sci. Rep.* **2016,** *6,* 32865. doi:10.1038/srep32865
22. Fang, W.; Lin, X.; Zhou, X.; Wan, J.; Lu, X.; Yang, B.; Ai, W.; Lin, J.; Zhang, T.; Tu, Z.; Liu, Y. *Med. Chem. Commun.* **2014,** *5,* 701–705. doi:10.1039/c3md00371j
23. Zhuravleva, O. I.; Afiyatullov, S. S.; Denisenko, V. A.; Ermakova, S. P.; Slinkina, N. N.; Dmitrenok, P. S.; Kim, N. Y. *Phytochemistry* **2012,** *80,* 123–131. doi:10.1016/j.phytochem.2012.05.008
24. Jansen, B. J. M.; de Groot, A. *Nat. Prod. Rep.* **2004,** *21,* 449–477. doi:10.1039/b311170a
25. Chaudhary, N. K.; Pitt, J. I.; Lacey, E.; Crombie, A.; Vuong, D.; Piggott, A. M.; Karuso, P. *J. Nat. Prod.* **2018,** *81,* 1517–1526. doi:10.1021/acs.jnatprod.7b00816
26. Lacey, H. J.; Vuong, D.; Pitt, J. I.; Lacey, E.; Piggott, A. M. *Aust. J. Chem.* **2016,** *69,* 152–160. doi:10.1071/ch15488
27. Li, H.; Gilchrist, C. L. M.; Lacey, H. J.; Crombie, A.; Vuong, D.; Pitt, J. I.; Lacey, E.; Chooi, Y.-H.; Piggott, A. M. *Org. Lett.* **2019,** *21,* 1287–1291. doi:10.1021/acs.orglett.8b04042
28. Pitt, J. I.; Lange, L.; Lacey, A. E.; Vuong, D.; Midgley, D. J.; Greenfield, P.; Bradbury, M. I.; Lacey, E.; Busk, P. K.; Pilgaard, B.; Chooi, Y.-H.; Piggott, A. M. *PLoS One* **2017,** *12,* e0170254. doi:10.1371/journal.pone.0170254
29. Shiono, Y.; Hiramatsu, F.; Murayama, T.; Koseki, T.; Funakoshi, T.; Ueda, K.; Yasuda, H. Z. *Z. Naturforsch., B: J. Chem. Sci.* **2007,** *62,* 1585–1589. doi:10.1515/znb-2007-1218
30. Gao, Y.; Honzatko, R. B.; Peters, R. J. *Nat. Prod. Rep.* **2012,** *29,* 1153–1175. doi:10.1039/c2np20059g

31. Hitchman, T. S.; Schmidt, E. W.; Trail, F.; Rarick, M. D.; Linz, J. E.; Townsend, C. A. *Bioorg. Chem.* **2001,** *29,* 293–307. doi:10.1006/bioo.2001.1216
32. Lin, H.-C.; Chooi, Y.-H.; Dhingra, S.; Xu, W.; Calvo, A. M.; Tang, Y. *J. Am. Chem. Soc.* **2013,** *135,* 4616–4619. doi:10.1021/ja312503y
33. Christianson, D. W. *Chem. Rev.* **2006,** *106,* 3412–3442. doi:10.1021/cr050286w
34. Nakano, C.; Okamura, T.; Sato, T.; Dairi, T.; Hoshino, T. *Chem. Commun.* **2005,** 1016–1018. doi:10.1039/b415346d
35. Cao, R.; Zhang, Y.; Mann, F. M.; Huang, C.; Mukkamala, D.; Hudock, M. P.; Mead, M. E.; Prisic, S.; Wang, K.; Lin, F.-Y.; Chang, T.-K.; Peters, R. J.; Oldfield, E. *Proteins: Struct., Funct., Bioinf.* **2010,** *78,* 2417–2432. doi:10.1002/prot.22751
36. Nakano, C.; Hoshino, T. *ChemBioChem* **2009,** *10,* 2060–2071. doi:10.1002/cbic.200900248
37. Lu, Z.; Wang, Y.; Miao, C.; Liu, P.; Hong, K.; Zhu, W. *J. Nat. Prod.* **2009,** *72,* 1761–1767. doi:10.1021/np900268z
38. Zhou, H.; Zhu, T.; Cai, S.; Gu, Q.; Li, D. *Chem. Pharm. Bull.* **2011,** *59,* 762–766. doi:10.1248/cpb.59.762
39. Cohen, E.; Koch, L.; Thu, K. M.; Rahamim, Y.; Aluma, Y.; Ilan, M.; Yarden, O.; Carmeli, S. *Bioorg. Med. Chem.* **2011,** *19,* 6587–6593. doi:10.1016/j.bmc.2011.05.045
40. Hayes, M. A.; Wrigley, S. K.; Chetland, I. A. N.; Reynolds, E. E.; Ainsworth, A. M.; Renno, D. V.; Latif, M. A.; Cheng, X.-M.; Hupe, D. J.; Charlton, P.; DoHerty, A. M. *J. Antibiot.* **1996,** *49,* 505–512. doi:10.7164/antibiotics.49.505
41. Slack, G. J.; Puniani, E.; Frisvad, J. C.; Samson, R. A.; Miller, J. D. *Mycol. Res.* **2009,** *113,* 480–490. doi:10.1016/j.mycres.2008.12.002
42. Curlevski, N. J. A.; Xu, Z.; Anderson, I. C.; Cairney, J. W. G. *J. Soils Sediments* **2010,** *10,* 1278–1288. doi:10.1007/s11368-010-0239-x
43. Gao, X.; Xie, X.; Pashkov, I.; Sawaya, M. R.; Laidman, J.; Zhang, W.; Cacho, R.; Yeates, T. O.; Tang, Y. *Chem. Biol.* **2009,** *16,* 1064–1074. doi:10.1016/j.chembiol.2009.09.017
44. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T. L. *BMC Bioinf.* **2009,** *10,* 421. doi:10.1186/1471-2105-10-421
45. Cerqueira, G. C.; Arnaud, M. B.; Inglis, D. O.; Skrzypek, M. S.; Binkley, G.; Simison, M.; Miyasato, S. R.; Binkley, J.; Orvis, J.; Shah, P.; Wymore, F.; Sherlock, G.; Wortman, J. R. *Nucleic Acids Res.* **2014,** *42,* D705–710. doi:10.1093/nar/gkt1029
46. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; Thierer, T.; Ashton, B.; Meintjes, P.; Drummond, A. *Bioinformatics* **2012,** *28,* 1647–1649. doi:10.1093/bioinformatics/bts199
47. Buchfink, B.; Xie, C.; Huson, D. H. *Nat. Methods* **2015,** *12,* 59–60. doi:10.1038/nmeth.3176

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Emission and biosynthesis of volatile terpenoids from the plasmodial slime mold *Physarum polycephalum*

Xinlu Chen[1], Tobias G. Köllner[2], Wangdan Xiong[1], Guo Wei[1] and Feng Chen[*1]

Full Research Paper

Open Access

Address:
[1]Department of Plant Sciences, University of Tennessee, Knoxville, TN 37996, USA and [2]Department of Biochemistry, Max Planck Institute for Chemical Ecology, Hans-Knöll-Strasse 8, D-07745 Jena, Germany

Email:
Feng Chen[*] - fengc@utk.edu

* Corresponding author

Keywords:
amoebae; evolution; terpene synthases; volatiles

Beilstein J. Org. Chem. **2019,** *15,* 2872–2880.
doi:10.3762/bjoc.15.281

Received: 01 October 2019
Accepted: 15 November 2019
Published: 28 November 2019

This article is part of the thematic issue "Terpenes".

Guest Editor: J. S. Dickschat

Abstract

Terpene synthases (TPSs) are pivotal enzymes for the production of diverse terpenes, including monoterpenes, sesquiterpenes, and diterpenes. In our recent studies, dictyostelid social amoebae, also known as cellular slime molds, were found to contain *TPS* genes for making volatile terpenes. For comparison, here we investigated *Physarum polycephalum*, a plasmodial slime mold also known as acellular amoeba. Plasmodia of *P. polycephalum* grown on agar plates were found to release a mixture of volatile terpenoids consisting of four major sesquiterpenes (α-muurolene, (*E*)-β-caryophyllene, two unidentified sesquiterpenoids) and the monoterpene linalool. There were no qualitative differences in terpenoid composition at two stages of young plasmodia. To understand terpene biosynthesis, we analyzed the transcriptome and genome sequences of *P. polycephalum* and identified four *TPS* genes designated *PpolyTPS1–PpolyTPS4*. They share 28–73% of sequence identities. Full-length cDNAs for the four *TPS* genes were cloned and expressed in *Escherichia coli* to produce recombinant proteins, which were tested for sesquiterpene synthase and monoterpene synthase activities. While neither PpolyTPS2 nor PpolyTPS3 was active, PpolyTPS1 and PpolyTPS4 were able to produce sesquiterpenes and monoterpenes from the respective substrates farnesyl diphosphate and geranyl diphosphate. By comparing the volatile profile of *P. polycephalum* plasmodia and the in vitro products of PpolyTPS1 and PpolyTPS4, it was concluded that most sesquiterpenoids emitted from *P. polycephalum* were attributed to PpolyTPS4. Phylogenetic analysis placed the four *PpolyTPSs* genes into two groups: *PpolyTPS1* and *PpolyTPS4* being one group that was clustered with the TPSs from the dictyostelid social amoeba and *PpolyTPS2* and *PpolyTPS3* being the other group that showed closer relatedness to bacterial TPSs. The biological role of the volatile terpenoids produced by the plasmodia of *P. polycephalum* is discussed.

Introduction

Volatile organic compounds (VOCs) are used by many living organisms as chemical languages for communication [1,2]. Rapid progress has been made in our understanding of the VOC world of microbes, especially bacteria [3,4] and fungi [5,6]. Not only the chemical diversity of microbial VOCs is continuingly to be discovered, our understanding of their biosynthesis is also

growing rapidly [7,8]. Among the diverse VOCs, terpenoids are the largest group. Terpenoids are biosynthesized from two C5 diphosphate compounds isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP), which are produced by either the mevalonate (MVA) pathway or the methylerythritol phosphate (MEP) pathway [9,10]. The MVA pathway is found in eukaryotes, archaea, and a few bacteria, and the MEP pathway is present in several photosynthetic eukaryotes and bacteria [11]. Isoprenyl diphosphate synthases (IDSs) catalyze the formation of prenyl diphosphates of various chain length [12]. After that, terpene synthases (TPSs) catalyze the conversion of prenyl diphosphates to diverse terpenes [13]. Because all living organisms produce prenyl diphosphates, whether an organism has the ability to produce terpenes depends on whether it contains *TPS* genes.

Recently we could show that dictyostelid social amoebae contain *TPS* genes. *TPS* genes were found in 6 species of sequenced amoeba, including *Dictyostelium discoideum*, *D. purpureum, Cavenderia fasciculata* (formerly *D. fasciculatum*), *Tieghemostelium lacteum* (formerly *D. lacteum*), *Heterostelium album* (formerly *Polysphondylium pallidum*), and *Actyostelium subglobosum* [14]. The number of *TPS* genes ranges from 1 to 21 in these species. Some of the *TPS* genes among these species have conserved catalytic functions. For example, TPSs of one orthologous group that include DdTPS6, DpTPS1, AsTPS1, DiTPS1, DfTPS1, and PpTPS18 all catalyze the formation of the sesquiterpene protoillud-7-ene [15,16]. Among paralogs, there is dramatic functional divergence. For instance, *D. discoideum* contains 9 *TPS* genes with diverse catalytic activities [14]. In *D. discoideum*, most *TPS* genes showed expression during multicellular development [14,15]. Consistent with the catalytic activities and gene expression patterns, the products of most DdTPSs were released as volatiles from *D. discoideum* at the multicellular developmental stage [14,15].

TPS genes previously were known to exist only in bacteria, fungi, and plants [13,17,18]. The identification of *TPS* genes in dictyostelid social amoeba now indicates a broader distribution of *TPS* genes. To understand whether *TPS* genes occur in other groups of amoebae, in this study, we investigated *Physarum polycephalum*. *P. polycephalum* belongs to the class of Myxogastria whereas social amoeba belongs to the class of Dictyostelia, but they both belong to the same infraphylum Mycetozoa. *P. polycephalum* is called plasmodial amoeba because of the plasmodium formed during the vegetative phase. Plasmodium is a single cell containing millions of nuclei, which also gives the name of acellular amoeba. *P. polycephalum* has been a popular model organism for studying a diversity of topics [19-21], ranging from cytoplasmic streaming to primitive intelligence [22]. In a manuscript deposited at arXiv [23], it was described that complex mixtures of volatiles including some terpenoids were detected from the headspace of *P. polycephalum* using two extraction temperatures. In our study, we aimed I) to determine whether *P. polycephalum* releases volatile terpenoids under normal growing conditions and II) to identify and characterize the genes for terpene biosynthesis in *P. polycephalum*. Our results will enable us to compare terpene chemistry and their underpinning biosynthetic genes in the two lineages of amoeba.

Results

Plasmodia of *P. polycephalum* release a mixture of volatile terpenoids

To determine whether *P. polycephalum* releases volatile terpenoids like dictyostelid social amoebae, plasmodia of *P. polycephalum* were cultured on agar plates with oak flakes as nutrient source and subjected to volatile profiling at two time points: 8 days and 18 days after the transfer of plasmodia to a fresh agar plate. Volatiles were collected from the headspace using a solid phase-microextraction fiber and analyzed using gas chromatography–mass spectrometry (GC–MS). At the 8th day after the transfer, a total of five volatiles were detected, including three known compounds and two unidentified compounds (Figure 1A). The three known compounds are all terpenoids, including one monoterpene linalool and two sesquiterpenes (*E*)-β-caryophyllene and α-muurolene (Figure 1B). The two unidentified are putative sesquiterpenoids. Compound **3** is a putative hydrocarbon sesquiterpene with a molecular mass of 204 (Figure 1C). In contrast, compound **4** has a molecular formula of $C_{15}H_{22}O$ and a molecular mass of 218 (Figure 1C). It was predicated to be a sesquiterpene aldehyde. At the 18th day after the transfer, essentially the same profile of volatiles was detected (Figure 1A).

Four terpene synthase genes were identified in *P. polycephalum*

With the identification of terpenes from the headspace of *P. polycephalum* (Figure 1), the next question was how they are synthesized. The genome of *P. polycephalum* has been sequenced [24] and there are multiple transcriptome datasets available for this species (http://www.physarum-blast.ovgu.de). Because the genome sequence was not annotated, we searched the transcriptomes for *TPS* genes. A total of four full-length putative *TPS* genes were identified from the transcriptomes. They were designated as *PpolyTPS1*, *PpolyTPS2*, *PpolyTPS3,* and *PpolyTPS4*. The length of the proteins encoded by *PpolyTPS1*, *PpolyTPS2*, *PpolyTPS3,* and *PpolyTPS4s* is 334, 347, 353, and 337 amino acids, respectively. Among the four proteins, the highest sequence similarities occurred between PpolyTPS1 and PpolyTPS4 (72%) and between PpolyTPS2 and PpolyTPS3 (64%). PpolyTPS1/4 and PpolyTPS2/3, however,

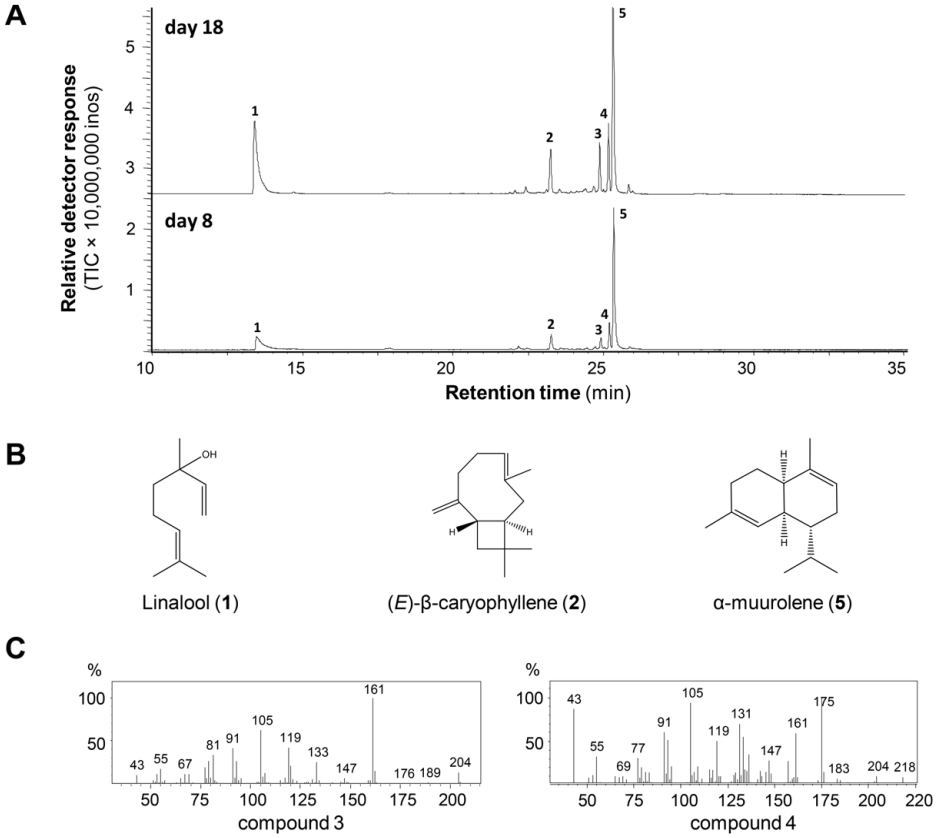


Figure 1: Plasmodia of *P. polycephalum* emit a mixture of volatiles predominated by terpenoids. A) GC chromatograms of volatiles collected from plasmodia of *P. polycephalum* grown at 8 and 18 days after transferring to a fresh agar plate. This experiment was repeated three times with similar results. **1**, linalool; **2**, (*E*)-β-caryophyllene; **3**, unidentified sesquiterpene hydrocarbon (compound **3**); **4**, unidentified putative sesquiterpene aldehyde (compound **4**); **5**, α-muurolene. B) Structures of known terpenoids. The structure of compound **2** shows the (−)-enantiomer of (*E*)-β-caryophyllene. However, since we have not performed chiral analysis, the absolute configuration of (*E*)-β-caryophyllene emitted from *P. polycephalum* plasmodia is still unknown. C) Mass spectra of unidentified terpenoids.

showed only ≈30% sequence similarity to each other. Terpene synthases can be classified into class I and class II, based on the reaction mechanisms they catalyze. These two types of terpene synthases are associated with conserved motifs: class I TPSs contain a 'DDxxD/E' and a 'NSD/DTE' motif while class II TPSs contain a 'DxDD' motif. While PpolyTPS1 and PpolyTPS4 contain the 'DDxxD' motif, PpolyTPS2 and PployTPS3 contain a 'DDxxE' motif. PpolyTPS1, PpolyTPS4, and PpolyTPS2 contain the 'NDxxSxxxE' motif. This motif is changed to "NDxxLxxxE" in PpolyTPS3. Based on their motifs, all four PpolyTPSs can be predicted to be class I TPSs. Also observed in all four PpolyTPSs is the "WxxxxxRY" motif (Figure 2A), which is frequently found in TPSs [25]. In our analysis of the genome sequence of *P. polycephalum*, the coding sequences of all four *PpolyTPS* genes were identified. *PpolyTPS* genes contain six to nine introns (Figure 2B). This is in contrast to the *TPS* genes from dictyostelid social amoeba, which consist zero to three introns [14].

Biochemical activities of PpolyTPSs

To determine whether *PpolyTPS* genes encode functional terpene synthases, full-length cDNAs were amplified by RT-PCR and cloned into the protein expression vector pEXP-5-CT/TOPO.

Recombinant PpolyTPSs were heterologously expressed in *Escherichia coli* and then tested for terpene synthase activities using geranyl diphosphate (GPP) and (*E,E*)-farnesyl diphosphate (FPP) as substrates. PpolyTPS2 and PpolyTPS3 did not show detectable terpene products with either GPP or FPP. In contrast, PpolyTPS1 could convert GPP into a mixture of cyclic and acyclic monoterpenes, including myrcene and linalool (**1**, Figure 3). PpolyTPS4 showed only trace activity with GPP and produced very small amounts of myrcene. When using FPP as substrate, PpolyTPS1 produced a mixture of sesquiterpenes with γ-muurolene as the most abundant compound and (*E*)-β-caryophyllene (**2**), α-muurolene (**5**), and four unidentified

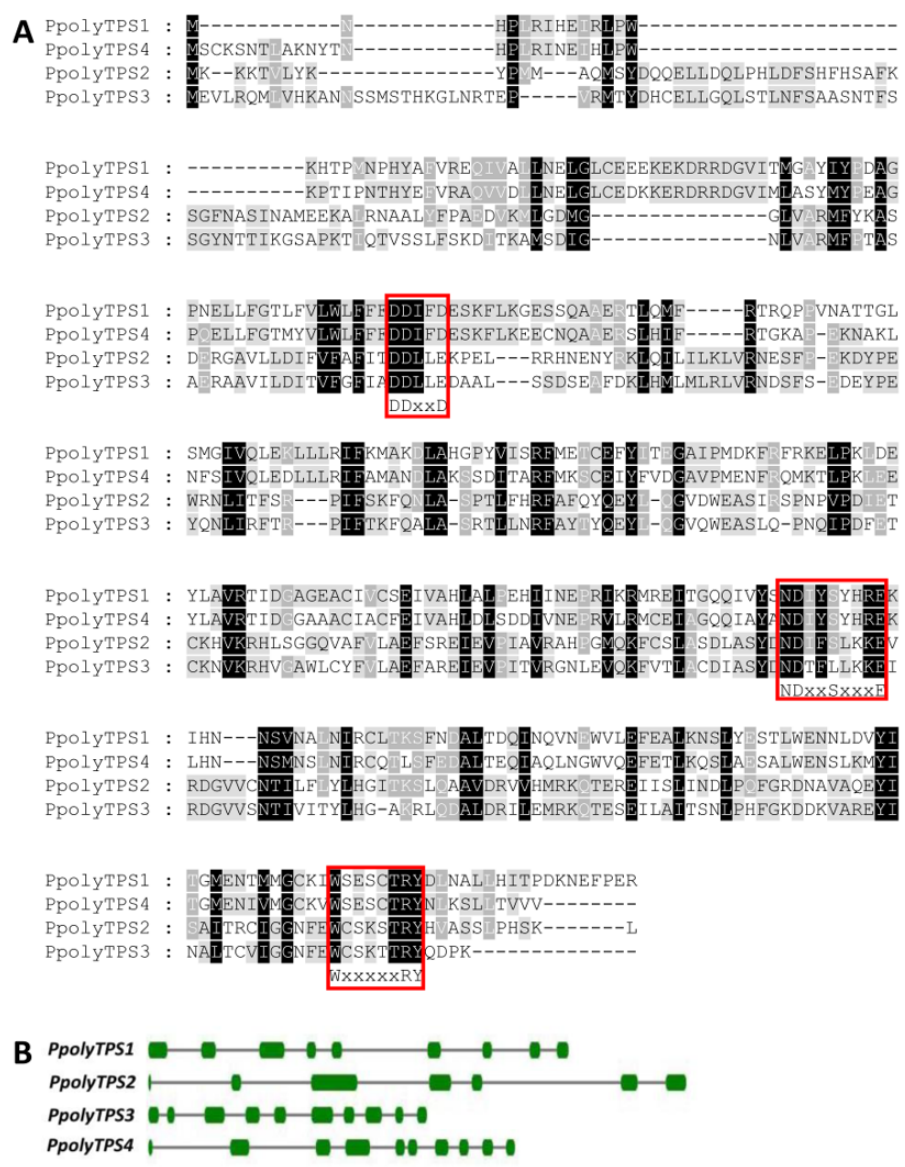


Figure 2: *P. polycephalum* contains four terpene synthase genes. A) Multiple sequence alignment of the protein sequences of the four PpolyTPSs. The sequences for PpolyTPS1-PpolyTPS4 reported in this paper have been deposited in the GenBank database (accession numbers. MN523652–MN523655). Three signature motifs for terpene synthases were boxed. Shadings in black and gray indicate identical and similar residues, respectively. B) Intron/exon organization of the four *PpolyTPS* genes. Boxes and lines indicate exons and introns, respectively.

sesquiterpenes (Supporting Information File 1, Figure S1) as minor components. PpolyTPS4 converted FPP into α-muurolene, (*E*)-β-caryophyllene, and one unidentified sesquiterpene (Figure 3). As negative controls, neither denatured PpolyTPS1 nor denatured PpolyTPS4 produced any terpene products using FPP (Supporting Information File 1, Figure S2).

Relatedness of PpolyTPSs to the TPSs from dictyostelid social amoebae, fungi, and bacteria

When individual *PpolyTPS* genes were used as query to search against the nonredundant protein database at NCBI, the top hits for PpolyTPS2 and PpolyTPS3 were all from bacteria. In contrast, the top hits for both PpolyTPS1 and PpolyTPS4 were from eukaryotes (Supporting Information File 1, Table S1). To further understand the evolutionary relatedness of PpolyTPSs to other TPSs, we performed a phylogenetic analysis of PpolyTPSs with TPSs from dictyostelid social amoeba, another amoeba *Naegleria gruberi*, fungi, and bacteria. The four PpolyTPSs genes were divided into two groups (Figure 4). *PpolyTPS1* and *PpolyTPS4* formed one group that was clustered with the TPSs from the dictyostelid social amoeba and the amoeba *N. gruberi*, which together were more related to fungal TPSs. In contrast, *PpolyTPS2* and *PpolyTPS3*

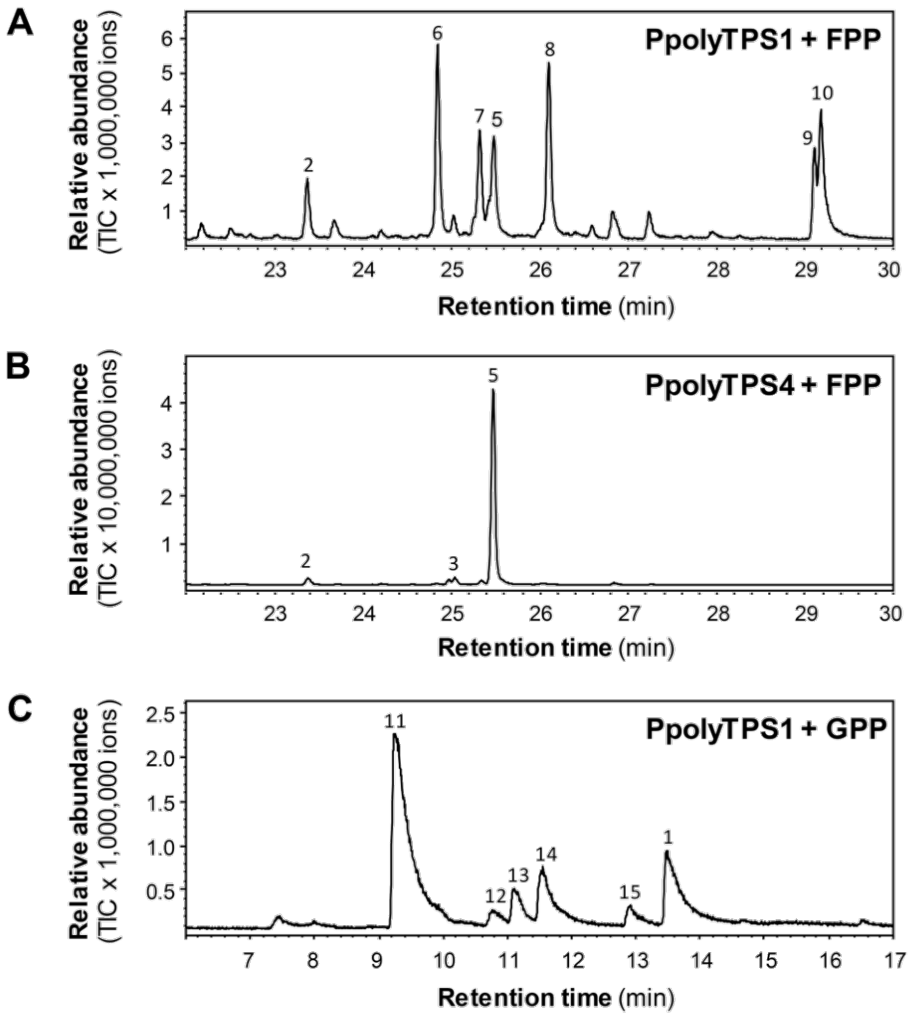


Figure 3: PpolyTPS1 and PpolyTPS4 have terpene synthase activities. A) GC chromatogram of sesquiterpenes produced by recombinant PpolyTPS1 incubated with FPP. **2**, (*E*)-β-caryophyllene; **6**, γ-muurolene; **7**, unidentified sesquiterpene 1; **5**, α-muurolene; **8**, unidentified sesquiterpene 2; **9**, unidentified sesquiterpene 4; **10**, unidentified sesquiterpene 5. B) GC chromatogram of sesquiterpenes produced by recombinant PpolyTPS4 incubated with FPP. **2**, (*E*)-β-caryophyllene; **3**, unidentified sesquiterpene 3; **5**, α-muurolene. C) GC chromatogram monoterpenes produced by recombinant PoplyTPS1 incubated with GPP. **1**, linalool; **11**, myrcene; **12**, limonene; **13**, (*Z*)-β-ocimene; **14**, (*E*)-β-ocimene.

formed the other group that showed closer relatedness to bacterial TPSs.

Discussion

With our previous work, now we have shown that both the dictyostelid social amoebae [14,15] and *Physarum polycephalum* (Figure 1A) release mixtures of volatiles predominated by terpenoids. Despite the fact that both are amoeba, dictyostelids and *P. polycephalum* have a long evolutionary distance and exhibit strikingly different life styles. In social amoeba, volatile terpenoids are exclusively released at the multicellular stage and were not detected from the unicellular vegetative stage [15]. In contrast, volatile terpenoids from *P. polycephalum* were released from plasmodia, its vegetative stage. Interestingly, the volatile terpenoids from both types of amoeba are predominated by sesquiterpenoids. However, α-muurolene, the most abundant sesquiterpene released from *P. polycephalum*, was not found in *D. discoideum* and *D. purpureum*. There were differences between the volatiles we detected in this study and the ones detected from *P. polycephalum* in a previous study [23]. In the Kateb and Costello study [23], a total of 87 compounds and 79 compounds were identified at the incubation temperature of 75 °C and 30 °C, respectively, from the plasmodia of *P. polycephalum.* Besides the differences in temperature (75 °C is not biologically relevant), the plasmodia were extracted from the agar plate before headspace collection in the Kateb and Costello study [23], which would disrupt the culture. Therefore, the much larger number of

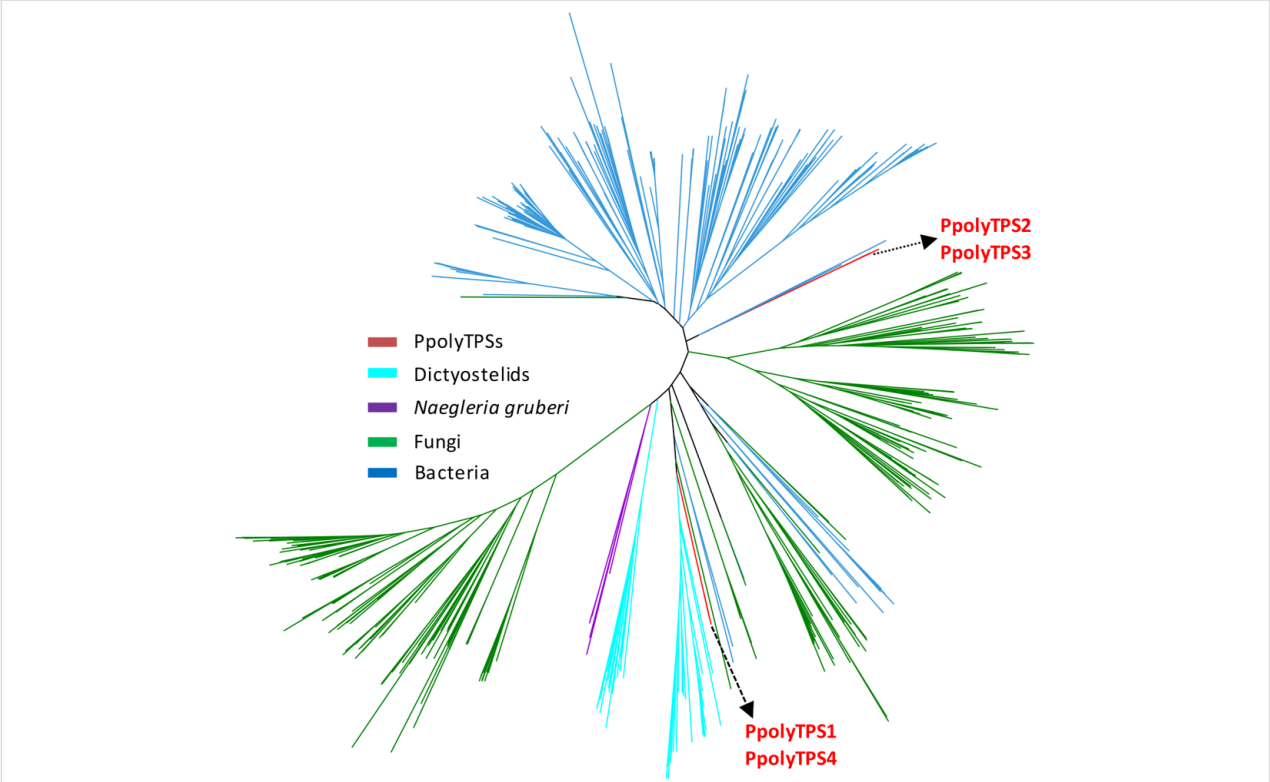


Figure 4: Phylogenetic analysis of PpolyTPSs with TPSs from dictyostelid social amoebae (Dictyostelids), the amoebae *Naegleria gruberi*, fungi, and bacteria. TPSs from dictyostelid social amoebae were obtained from the study [15]. TPSs from another amoeba *N. gruberi*, fungi, and bacteria were obtained from the study [14]. TPSs from different lineages were color-coded.

compounds detected in that study may be partly due to the destructive nature of their method. It is notable that α-muurolene (**5**) was detected from both studies.

Like in social amoeba, the production of most volatile terpenoids in *P. polycephalum* can be attributed to specific terpene synthases. PpolyTPS4 produced α-muurolene (**5**) as a major product, and (*E*)-β-caryophyllene (**2**) and an unidentified sesquiterpene hydrocarbon as minor products. These three terpenes could also be found in the headspace of *P. polycephalum* with α-muurolene as the most abundant constituent (Figure 1A). It is thus conceivable that PpolyTPS4 is responsible for the formation of these three compounds in vivo. PpolyTPS1 produced also α-muurolene, however, the in vitro product profile was dominated by other sesquiterpenes that were not detected in the headspace of *P. polycephalum*. Nonetheless, PpolyTPS1 possessed monoterpene synthase activity in vitro and produced a mixture of monoterpenes. One of the PpolyTPS1 monoterpene products, linalool (**1**), could be detected in the headspace of *P. polycephalum*, but the other monoterpene products including myrcene, limonene, (*Z*)-β-ocimene, and (*E*)-β-ocimene were not detected from the headspace of *P. polycephalum*. It is thus unclear whether and how PpolyTPS1 contributes to volatile terpene formation in the vegetative stage of *P. polycephalum*. It is possible that *P. polycephalum* contains other *TPS* genes, which were not identified in this study due to incomplete transcriptome and genome information, but contribute to the in vivo biosynthesis of linalool. It is interesting to note that only PpolyTPS1 and PpolyTPS4 are closely related to TPSs of dictyostelid social amoebae whereas PpolyTPS2 and PpolyTPS3 are closely related to bacteria TPSs (Figure 4), suggesting two evolutionary origins of *PpolyTPS* genes. It is tempting to speculate that *PpolyTPS2* and *PpolyTPS3* may be derived from bacteria through horizontal gene transfer, which was recently demonstrated to have occurred from bacteria to fungi for the evolution of *TPS* genes [26]. It is also interesting to note that under our standard assay conditions, neither PpolyTPS2 nor PpolyTPS3 showed activity with either GPP or FPP. Because catalytic motifs are present in both TPSs (Figure 2), the inactivity is somehow puzzling. Some efforts are needed to discern whether they are instable enzymes, active with different prenyl diphosphate substrates, or require different conditions for catalysis. It is also possible that they become inactive genes in the process of pseudogenization.

What are the biological functions of volatile terpenoids emitted from *P. polycephalum*? In plants and other organisms, volatile terpenoids have many biological/ecological functions. They

may serve as chemical defense [27,28] or as signals for attracting beneficial organisms. For social amoeba, multiple functions have been proposed, including chemical defense, attracting spore dispersers, and regulating development. In *D. discoideum*, a mutant strain with a disrupted *DdTPS8* gene showed slower progression in development [29], adding genetic evidence on the role of *TPS* genes and their terpene products in development. The volatiles emitted from *P. polycephalum* may have similar functions as those in dictyostelid social amoeba or have lineage-specific functions. *P. polycephalum* has a unique biology. It sends information in the form of nutrient concentrations through the tubular network using the streaming cytoplasm [30]. It will be interesting to test whether volatile terpenoids function in internal communication of *P. polycephalum*. When foraging, *P. polycephalum* marks the territory that have been explored. Some chemicals have been proposed to function in this process [31]. It is well known that some insects such as fire ants use terpenes as trace pheromone [32]. It will be interesting to determine whether some terpenoids produced by *P. polycephalum* have similar functions. When studying chemotaxis of *P. polycephalum*, a number of exogenously applied volatiles were tested for attraction and repellence. One sesquiterpene farnesene was found to be a strong chemoattractant of *P. polycephalum* [31]. Thus, it will also be intriguing to ask whether volatile terpenoids emitted from *P. polycephalum* are involved in chemotaxis.

Conclusion

In this study, we have successfully identified and characterized terpene synthase (*TPS*) genes that are involved in making volatile terpenoids from a plasmodial slime mold *Physarum polycephalum*. The volatiles emitted from the plasmodium of *P. polycephalum* were mainly sesquiterpenes, but also included monoterpenes (Figure 1). These volatile terpenoids are probably constitutively produced, because the plasmodia at two developmental stages of *P. polycephalum* did not display qualitative differences in terpene profiles (Figure 1). Four *TPS* genes, designated *PpolyTPS1–PpolyTPS4*, were identified from *P. polycephalum*. When *E. coli*-expressed PpolyTPS proteins were tested for terpene synthase activities, PpolyTPS1 was demonstrated to be a sesquiterpene synthase and PpolyTPS4 a monoterpene synthase (Figure 3). Their in vitro terpene products were also detected from the headspace of *P. polycephalum* plasmodium. When PpolyTPSs were compared with those from dictyostelid social amoebae, only *PpolyTPS1* and *PpolyTPS4* were shown to be clustered with the TPSs from the dictyostelid social amoeba (Figure 4). Despite this difference as well as the difference in life style of *P. polycephalum* and dictyostelid social amoeba, it appears that both of these two types of organisms use volatile terpenoids for certain biological and ecological processes. This study provides novel information on the occurrence of terpenoids and their biosynthetic genes among eukaryotes. Because *P. polycephalum* is a popular model system for investigating many aspects of biology and other disciplines such as engineering and physics [19,22,33-36], our results most likely will stimulate new research directions using *P. polycephalum* as a model system centered upon volatile terpenoids.

Experimental

P. polycephalum culture

P. polycephalum was purchased from Carolina Biological Supply Company (https://www.carolina.com). The plasmodium of *P. polycephalum* was cultured on agar plates with oat flakes as a nutrient source under continuous darkness.

Headspace collection and GC–MS analysis

On the 8th day and 18th day after the transfer of plasmodia to a fresh agar plate, volatiles were collected from the headspace of plasmodia on the agar plate using solid phase microextraction (SPME) (https://www.sigmaaldrich.com). After collection for one hour, the SPME fiber was withdrawn and then inserted into the injector port of a Shimadzu 17A gas chromatograph coupled to a Shimadzu QP5050A quadrupole mass selective detector for volatile identification and identification. Separation was performed on a Restek Rxi-5Sil MS column (30 m × 0.25 mm i.d. × 0.25 μm thickness; Restek) with helium as the carrier gas and a temperature program from 60 °C to 300 °C at 5 °C per minute rate. The experiment was performed with three biological replicates.

Sequence retrieval and analysis

Transcriptome datasets for *P. polycephalum* were downloaded from the *Physarum polycephalum* Genome Resources database (http://www.physarum-blast.ovgu.de). Protein sequences were predicated using TransDecoder (5.0.2) [37]. Putative terpene genes were searched against SmMTPSLs HMM profile [38] using HMMER 3.0 hmmsearch [39] with an E-value of $1e^{-5}$. The coding region of each *PpolyTPS* gene was identified from the genome sequence of *P. polycephalum*. For phylogenetic reconstruction, a multiple sequence alignment was first performed using MAFFT (v7.450) [40] in accurate strategy (L-INS-i) with 1000 iteration of improvement. Then phylogenetic trees were built by Fasttree [41] and visualized using interactive tree of life (ITOL) (https://itol.embl.de/).

Cloning of full-length cDNA of terpene synthase genes in *P. polycephalum*

Plasmodia were harvested from agar plates, placed in 2 mL centrifuge tubes and disrupted using Qiagen TissueLyser II according to manufacturer's manual (https://www.qiagen.com). Total RNA was isolated using Plant RNA Purification Reagent

(https://www.thermofisher.com). cDNA was prepared using 1st strand cDNA synthesis kit (https://www.gelifesciences.com). Full length cDNAs of individual *PpolyTPS* genes were amplified using gene specific primers (Table S2, Supporting Information File 1), cloned into pEXP5 CT/TOPO vector (https://www.thermofisher.com) and fully sequenced.

Terpene synthase enzyme assays

pEXP5 CT/TOPO vector containing individual *PpolyTPS* genes was transformed into *E. coli* strain BL21 Codon Plus (DE3). Heterologous expression of individual *PpolyTPS* genes in *E. coli* and recombinant protein preparation and terpene synthase enzyme assays were performed as previously described [42]. Each PoplyTPS recombinant protein was tested with both geranyl diphosphate and farnesyl diphosphate as substrates and terpene products were analyzed using GC–MS as described for volatile profiling.

Supporting Information

Supporting Information File 1

Additional figures and tables.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-281-S1.pdf]

Acknowledgements

Wangdan Xiong was supported by a scholarship from the China Scholarship Council. Kevin Chen from Farragut High School, Knoxville, is acknowledged for his assistance with chemical profiling and TPS enzyme assays for this project.

ORCID® iDs

Tobias G. Köllner - https://orcid.org/0000-0002-7037-904X
Guo Wei - https://orcid.org/0000-0001-5977-0681
Feng Chen - https://orcid.org/0000-0002-3267-4646

References

1. Chen, F.; Ludwiczuk, A.; Wei, G.; Chen, X.; Crandall-Stotler, B.; Bowman, J. L. *Crit. Rev. Plant Sci.* **2018,** *37,* 210–231. doi:10.1080/07352689.2018.1482397
2. Schulz, S.; Dickschat, J. S. *Nat. Prod. Rep.* **2007,** *24,* 814–842. doi:10.1039/b507392h
3. Audrain, B.; Farag, M. A.; Ryu, C.-M.; Ghigo, J.-M. *FEMS Microbiol. Rev.* **2015,** *39,* 222–233. doi:10.1093/femsre/fuu013
4. Schulz-Bohm, K.; Geisen, S.; Wubs, E. R. J.; Song, C.; de Boer, W.; Garbeva, P. *ISME J.* **2017,** *11,* 817–820. doi:10.1038/ismej.2016.144
5. Hung, R.; Lee, S.; Bennett, J. W. *Appl. Microbiol. Biotechnol.* **2015,** *99,* 3395–3405. doi:10.1007/s00253-015-6494-4
6. Schmidt, R.; Etalo, D. W.; de Jager, V.; Gerards, S.; Zweers, H.; de Boer, W.; Garbeva, P. *Front. Microbiol.* **2016,** *6,* No. 1495. doi:10.3389/fmicb.2015.01495
7. McGenity, T. J.; Crombie, A. T.; Murrell, J. C. *ISME J.* **2018,** *12,* 931–941. doi:10.1038/s41396-018-0072-6
8. Wisecaver, J. H.; Slot, J. C.; Rokas, A. *PLoS Genet.* **2014,** *10,* e1004816. doi:10.1371/journal.pgen.1004816
9. Lichtenthaler, H. K. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1999,** *50,* 47–65. doi:10.1146/annurev.arplant.50.1.47
10. Sapir-Mir, M.; Mett, A.; Belausov, E.; Tal-Meshulam, S.; Frydman, A.; Gidoni, D.; Eyal, Y. *Plant Physiol.* **2008,** *148,* 1219–1228. doi:10.1104/pp.108.127951
11. Lombard, J.; Moreira, D. *Mol. Biol. Evol.* **2011,** *28,* 87–99. doi:10.1093/molbev/msq177
12. Nagel, R.; Schmidt, A.; Peters, R. J. *Planta* **2019,** *249,* 9–20. doi:10.1007/s00425-018-3052-1
13. Chen, F.; Tholl, D.; Bohlmann, J.; Pichersky, E. *Plant J.* **2011,** *66,* 212–229. doi:10.1111/j.1365-313x.2011.04520.x
14. Chen, X.; Köllner, T. G.; Jia, Q.; Norris, A.; Santhanam, B.; Rabe, P.; Dickschat, J. S.; Shaulsky, G.; Gershenzon, J.; Chen, F. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113,* 12132–12137. doi:10.1073/pnas.1610379113
15. Chen, X.; Köllner, T. G.; Shaulsky, G.; Jia, Q.; Dickschat, J. S.; Gershenzon, J.; Chen, F. *Sci. Rep.* **2018,** *8,* 14361. doi:10.1038/s41598-018-32639-0
16. Rabe, P.; Rinkel, J.; Nubbemeyer, B.; Köllner, T. G.; Chen, F.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2016,** *55,* 15420–15423. doi:10.1002/anie.201608971
17. Yamada, Y.; Kuzuyama, T.; Komatsu, M.; Shin-ya, K.; Omura, S.; Cane, D. E.; Ikeda, H. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112,* 857–862. doi:10.1073/pnas.1422108112
18. Schmidt-Dannert, C. Biosynthesis of Terpenoid Natural Products in Fungi. In *Biotechnology of Isoprenoids;* Schrader, J.; Bohlmann, J., Eds.; Advances in Biochemical Engineering/Biotechnology, Vol. 148; Springer: Cham, Switzerland, 2014; pp 19–61. doi:10.1007/10_2014_283
19. Alim, K.; Andrew, N.; Pringle, A.; Brenner, M. P. *Proc. Natl. Acad. Sci. U. S. A.* **2017,** *114,* 5136–5141. doi:10.1073/pnas.1618114114
20. Taylor, B.; Adamatzky, A.; Greenman, J.; Ieropoulos, I. *BioSystems* **2015,** *127,* 42–46. doi:10.1016/j.biosystems.2014.10.005
21. Bonifaci, V.; Mehlhorn, K.; Varma, G. *J. Theor. Biol.* **2012,** *309,* 121–133. doi:10.1016/j.jtbi.2012.06.017
22. Nakagaki, T.; Yamada, H.; Tóth, Á. *Nature* **2000,** *407,* 470. doi:10.1038/35035159
23. Kateb, H. a.; Costello, B. d. L. *arXiv* **2013,** 1307.
24. Schaap, P.; Barrantes, I.; Minx, P.; Sasaki, N.; Anderson, R. W.; Bénard, M.; Biggar, K. K.; Buchler, N. E.; Bundschuh, R.; Chen, X.; Fronick, C.; Fulton, L.; Golderer, G.; Jahn, N.; Knoop, V.; Landweber, L. F.; Maric, C.; Miller, D.; Noegel, A. A.; Peace, R.; Pierron, G.; Sasaki, T.; Schallenberg-Rüdinger, M.; Schleicher, M.; Singh, R.; Spaller, T.; Storey, K. B.; Suzuki, T.; Tomlinson, C.; Tyson, J. J.; Warren, W. C.; Werner, E. R.; Werner-Felmayer, G.; Wilson, R. K.; Winckler, T.; Gott, J. M.; Glöckner, G.; Marwan, W. *Genome Biol. Evol.* **2016,** *8,* 109–125. doi:10.1093/gbe/evv237
25. Driller, R.; Janke, S.; Fuchs, M.; Warner, E.; Mhashal, A. R.; Major, D. T.; Christmann, M.; Brück, T.; Loll, B. *Nat. Commun.* **2018,** *9,* No. 3971. doi:10.1038/s41467-018-06325-8
26. Jia, Q.; Chen, X.; Köllner, T. G.; Rinkel, J.; Fu, J.; Labbé, J.; Xiong, W.; Dickschat, J. S.; Gershenzon, J.; Chen, F. *Sci. Rep.* **2019,** *9,* 9223. doi:10.1038/s41598-019-45532-1

27. Rasmann, S.; Köllner, T. G.; Degenhardt, J.; Hiltpold, I.; Toepfer, S.; Kuhlmann, U.; Gershenzon, J.; Turlings, T. C. J. *Nature* **2005,** *434,* 732–737. doi:10.1038/nature03451
28. Köllner, T. G.; Held, M.; Lenk, C.; Hiltpold, I.; Turlings, T. C. J.; Gershenzon, J.; Degenhardt, J. *Plant Cell* **2008,** *20,* 482–494. doi:10.1105/tpc.107.051672
29. Chen, X.; Luck, K.; Rabe, P.; Dinh, C. Q.; Shaulsky, G.; Nelson, D. R.; Gershenzon, J.; Dickschat, J. S.; Köllner, T. G.; Chen, F. *eLife* **2019,** *8,* e44352. doi:10.7554/elife.44352
30. Dussutour, A.; Latty, T.; Beekman, M.; Simpson, S. J. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107,* 4607–4611. doi:10.1073/pnas.0912198107
31. de Lacy Costello, B. P.; Adamatzky, A. I. *Commun. Integr. Biol.* **2013,** *6,* e25030.
32. Suckling, D. M.; Stringer, L. D.; Bunn, B.; El-Sayed, A. M.; Vander Meer, R. K. *J. Chem. Ecol.* **2010,** *36,* 744–750. doi:10.1007/s10886-010-9810-6
33. Adamatzky, A. *Parallel Process. Lett.* **2009,** *19,* 105–127. doi:10.1142/s0129626409000109
34. Romeo, A.; Dimonte, A.; Tarabella, G.; D'Angelo, P.; Erokhin, V.; Iannotta, S. *APL Mater.* **2015,** *3,* 014909. doi:10.1063/1.4902817
35. Tero, A.; Kobayashi, R.; Nakagaki, T. *J. Theor. Biol.* **2007,** *244,* 553–564. doi:10.1016/j.jtbi.2006.07.015
36. Tsuda, S.; Aono, M.; Gunji, Y.-P. *BioSystems* **2004,** *73,* 45–55. doi:10.1016/j.biosystems.2003.08.001
37. Haas, B. J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P. D.; Bowden, J.; Couger, M. B.; Eccles, D.; Li, B.; Lieber, M.; MacManes, M. D.; Ott, M.; Orvis, J.; Pochet, N.; Strozzi, F.; Weeks, N.; Westerman, R.; William, T.; Dewey, C. N.; Henschel, R.; LeDuc, R. D.; Friedman, N.; Regev, A. *Nat. Protoc.* **2013,** *8,* 1494–1512. doi:10.1038/nprot.2013.084
38. Jia, Q.; Li, G.; Köllner, T. G.; Fu, J.; Chen, X.; Xiong, W.; Crandall-Stotler, B. J.; Bowman, J. L.; Weston, D. J.; Zhang, Y.; Chen, L.; Xie, Y.; Li, F.-W.; Rothfels, C. J.; Larsson, A.; Graham, S. W.; Stevenson, D. W.; Wong, G. K.-S.; Gershenzon, J.; Chen, F. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113,* 12328–12333. doi:10.1073/pnas.1607973113
39. Finn, R. D.; Clements, J.; Eddy, S. R. *Nucleic Acids Res.* **2011,** *39,* W29–W37. doi:10.1093/nar/gkr367
40. Katoh, K.; Standley, D. M. *Mol. Biol. Evol.* **2013,** *30,* 772–780. doi:10.1093/molbev/mst010
41. Price, M. N.; Dehal, P. S.; Arkin, A. P. *Mol. Biol. Evol.* **2009,** *26,* 1641–1650. doi:10.1093/molbev/msp077
42. Li, G.; Köllner, T. G.; Yin, Y.; Jiang, Y.; Chen, H.; Xu, Y.; Gershenzon, J.; Pichersky, E.; Chen, F. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109,* 14711–14715. doi:10.1073/pnas.1204300109

License and Terms

BEILSTEIN JOURNAL OF ORGANIC CHEMISTRY

Bacterial terpene biosynthesis: challenges and opportunities for pathway engineering

Eric J. N. Helfrich[‡1], Geng-Min Lin[‡2], Christopher A. Voigt[2] and Jon Clardy[*1]

Review

Open Access

Address:
[1]Harvard Medical School, Department of Biological Chemistry and Molecular Pharmacology, Boston, United States and [2]Massachusetts Institute of Technology, Department of Biological Engineering, Cambridge, United States

Email:
Jon Clardy[*] - jon_clardy@hms.harvard.edu

* Corresponding author ‡ Equal contributors

Keywords:
bacterial sesquiterpenes and diterpenes; cytochrome P450; pathway engineering; synthetic biology; terpene biosynthesis; terpene cyclase

Beilstein J. Org. Chem. **2019,** *15,* 2889–2906.
doi:10.3762/bjoc.15.283

Received: 27 August 2019
Accepted: 01 November 2019
Published: 29 November 2019

This article is part of the thematic issue "Terpenes".

Guest Editor: J. S. Dickschat

Abstract

Terpenoids are the largest and structurally most diverse class of natural products. They possess potent and specific biological activity in multiple assays and against diseases, including cancer and malaria as notable examples. Although the number of characterized terpenoid molecules is huge, our knowledge of how they are biosynthesized is limited, particularly when compared to the well-studied thiotemplate assembly lines. Bacteria have only recently been recognized as having the genetic potential to biosynthesize a large number of complex terpenoids, but our current ability to associate genetic potential with molecular structure is severely restricted. The canonical terpene biosynthetic pathway uses a single enzyme to form a cyclized hydrocarbon backbone followed by modifications with a suite of tailoring enzymes that can generate dozens of different products from a single backbone. This functional promiscuity of terpene biosynthetic pathways renders terpene biosynthesis susceptible to rational pathway engineering using the latest developments in the field of synthetic biology. These engineered pathways will not only facilitate the rational creation of both known and novel terpenoids, their development will deepen our understanding of a significant branch of biosynthesis. The biosynthetic insights gained will likely empower a greater degree of engineering proficiency for non-natural terpene biosynthetic pathways and pave the way towards the biotechnological production of high value terpenoids.

Introduction

Evolutionary diversification of terpene biosynthetic pathways has resulted in the largest and most structurally diverse class of specialized metabolites on the planet. To date more than 70,000 terpenoids (dictionary of natural products) have been characterized and grouped into more than 400 structural families – the vast majority of which have been isolated from plants and fungi [1]. Their structural diversity reflects the breadth of their functional roles, which range from widely distributed metabolites like cholesterol, to those with more restricted distribution like vitamins A and D, carotenoids, and steroid hormones, to some

with highly restricted distribution like pheromones, fragrances, and defense metabolites [2,3]. Many of the more restricted members possess significant biological activities, like the anticancer agent taxol (**1**) [4] or the antimalarial agent artemisinin (**2**, Figure 1) [5,6]. The structural diversity and functional utility of this class of specialized metabolites have combined to encourage efforts to apply the tools of synthetic biology to engineer pathways that will expand molecular diversity, especially around scaffolds associated with high-value compounds.

The biosynthetic logic of terpene formation differs significantly from the logic employed by other classes of secondary metabolite biosynthetic pathways. Bacterial thiotemplated assembly lines, such as type I polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS), are modular, with each module contributing a distinct fragment to the final product's core structure – a short-chain carboxylic acid (PKS) or an amino acid (NRPS). The modularly defined template can be further modified by tailoring enzymes, but the core structure can be inferred from the organization of the biosynthetic genes and the modular architecture of the associated proteins [7,8].

Terpene biosynthesis has a very different logic. Five-carbon units called isoprenes are joined to create a linear polyene with branching methyl groups that form the core hydrocarbon structure in a single enzyme-catalyzed step [9]. The enzyme, which is called terpene cyclase, holds the linear methyl-branched polyene in a defined conformation that initiates a series of carbocation-driven cyclizations and rearrangements, creating the basic hydrocarbon skeleton of a terpene [10,11]. This basic hydrocarbon skeleton is then modified to generate a large number of terpenoid structures, which can be further modified by addition of other building blocks, like sugars, amino acids, or fatty acids [12]. Terpenes are named for the number of five-carbon units that form their hydrocarbon skeletons. Our review focuses on sesqui- (C_{15}) and diterpenes (C_{20}) because these subgroups, after undergoing extensive oxidative modification, have molecular characteristics that are most similar to those of known drugs. Emphasis will be placed on terpene pathways from bacteria, as their biosynthetic pathways usually have the genes encoding the terpene cyclase and modifying enzyme in close proximity, which simplifies both analysis and pathway engineering. The review will begin with a brief description of terpene families with a special focus on a few illustrative examples with pharmaceutical significance that highlight general hallmarks of terpene biosynthesis, then describe the natural biosynthetic pathways in more detail before moving on to describing the progress, promise, and obstacles to engineering terpene biosynthetic pathways.

Review

Remaining challenges in terpene (bio)synthesis

While the large number of characterized terpenes and their biological and medical significance would suggest that there are good tools for terpene synthesis along with a good understanding of terpene biosynthesis, significant gaps in both remain. Taxol (**1**), a plant-derived terpenoid, provides an illustrative example. Taxol was structurally characterized in 1971 [4] and approved by the FDA as an anticancer agent in 1992 [13]. Today, almost 50 years after its initial report, despite its blockbuster status in cancer therapy [14] and multiple research efforts, there is no sustainable synthetic or biosynthetic approach to this molecule. More than half a dozen total syntheses,

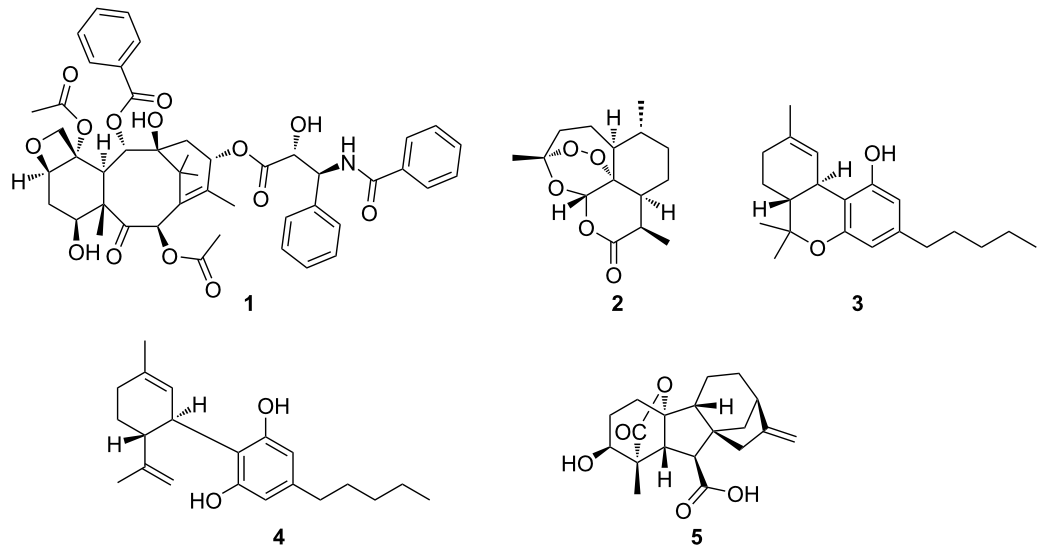


Figure 1: Examples of bioactive terpenoids.

each requiring 37 steps [15,16] or more [14], have been reported, yielding at best 32 mg of the drug in toto, while the yearly pharmaceutical requirements are greater than 10^9 mg [17]. The cyclase that forms the hydrocarbon framework has been known for more than 20 years [18,19], but the precise biosynthetic modifications leading to the final molecule are still not fully understood [20]. These shortcomings were successfully addressed in a practical sense through semi-synthesis from biotechnologically produced taxol precursors [21], but a completely engineered pathway could provide a more efficient solution to taxol and specific analogs.

There are additional puzzles about terpene biosynthesis that engineered pathways could address. As many other terpenes, **1** is produced along with a large suite of related molecules, the taxanes [17]. This pathway promiscuity can be illustrated with another popular example of a terpene family: the cannabinoids, with more than 100 members used for both medical and recreational purposes. While its most prominent member (−)-*trans*-Δ^9-tetrahydrocannabinol, better known as THC (**3**, Figure 1), is widely known as the principal psychoactive constituent of cannabis, cannabidiol (CBD, **4**, Figure 1) for example, shows strongly diminished psychoactive properties but has promising anti-inflammatory [22], antischizophrenic [23], and antiepileptic bioactivities [24,25]. Modifying the cannabinoid profile through engineering would therefore be of significant interest.

The biosynthetic promiscuity, the hallmark of terpene biosynthesis, sets terpenes apart from other natural product classes and is a product of their distinctive biosynthetic logic. Biosynthetic core enzymes of well-characterized classes of natural products, such as modular thiotemplate assembly lines (NRPSs, PKSs), are usually highly specific and produce only a few closely related natural product analogs. Adenylation domains in NRPSs [26-28], acyltransferase (AT) domains in *cis*-AT PKSs [29], and ketosynthase domains in *trans*-AT PKSs [30,31] are highly substrate-specific and function as gatekeepers to ensure that a single natural product is produced with high efficiency. Erroneous or stalled intermediates are removed from the assembly lines [32,33]. Novelty in these assembly lines is achieved through the recombination of module series (*trans*-AT PKSs) [34], the exchange of large subdomains (NRPSs) [35], and the duplication followed by diversification of entire modules (NRPSs and *cis*-AT PKSs) [36]. As a result, dysfunctionality is a constant threat for a metabolic pathway during evolution.

In contrast, the underlying mechanistic logic of terpene biosynthesis is based on repetitive electrophilic and nucleophilic functionalities in each oligomeric substrate, similar to nonmodular type II PKSs, coupled with conformational flexibility for enzyme-mediated juxtaposition of complementary pairs of these functionalities (Figure 2) [9]. It is this intrinsic and repetitive reactivity that can be easily tuned by natural selection [9]. As a result, a single terpene cyclase (TC) can produce dozens of hydrocarbon scaffolds that can differ significantly from each other [9,37]. This biosynthetic promiscuity should be regarded as a valuable feature rather than a bug in the system. According to the “screening hypothesis" [38,39] (more recently and appropriately also referred to as the “diversity-based hypothesis" [40]), potent biological activity is a rare property of any molecule in nature [39]. Therefore, evolution would likely favor organisms that can generate and retain chemical diversity at a low cost [38]. As a result, producing and “screening” a large number of specialized metabolites for potent TCs that can generate multiple products from simple building blocks is a huge evolutionary advantage; even more so when combined with tailoring enzymes that have a broad substrate tolerance and catalyze multiple, sequential tailoring reactions [40]. This diversification process resembles the strategy embedded in the construction of combinatorial libraries by organic chemists to generate chemical diversity. The biosynthetic promiscuity is not a result of intrinsic "sloppiness" of terpenoid biosynthetic pathways, as terpenes that are classified as primary metabolites, such as cholesterol, are produced with high fidelity [40]. In contrast, more promiscuous terpene pathways might only be a few mutations away from novel highly bioactive natural products that might meet new selective needs [40]. All it takes is the ability to generate a small amount of terpene side product that nature can utilize as a starting point to reinforce and refine the pathway [9,40]. As such, terpene pathways allow for quantita-

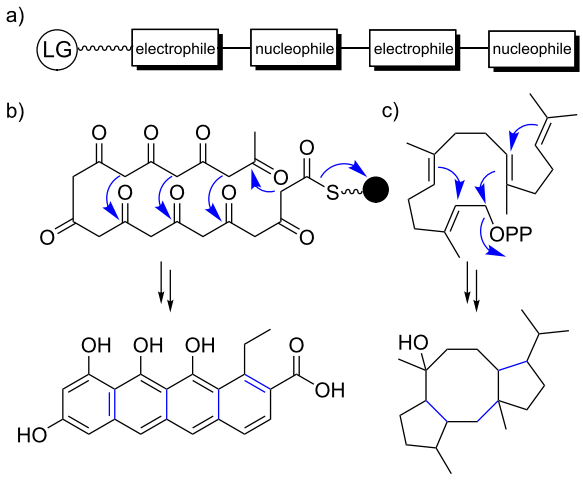


Figure 2: Repetitive electrophilic and nucleophilic functionalities in terpene and type II PKS-derived polyketide biosynthesis. a) Schematic representation. b) Type II PKS-derived polyketide biosynthesis. c) Terpene biosynthesis. LG: leaving group, black ball: acyl carrier protein.

tive control of product outcomes, while PKS and NRPS pathways feature qualitative control [40]. The gibberellins (e.g., gibberellin A4 (**5**), Figure 1) are an extreme example illustrating the promiscuity of terpene biosynthetic pathways with more than 130 different family members reported [40]. In fact, different pathways have evolved in plants, fungi, and bacteria for this fascinating compound family in an extreme case of convergent evolution [41,42]. While the plant and fungal biosynthetic pathways are well studied [42], the bacterial pathway was studied to a lesser degree until recently [41] – something that can be largely attributed to the historic perception that bacteria are not capable of producing complex terpenoids.

The historic neglect of bacterial terpenes becomes most apparent when the literature is mined for characterized bacterial terpene biosynthetic gene clusters (BGCs). While the antiSMASH database (online repository of BGCs predicted by the genome mining platform antiSMASH) lists more than 4,000 bacterial terpene BGCs [43], only 127 have been characterized and deposited in the MIBiG database (repository of characterized BGCs) to date (Figure 3) [44].

The comparatively small number of characterized bacterial terpenes can likely be attributed to three factors: 1) the early misconception that bacteria are not capable of producing complex terpenoids, 2) the lack of genome mining platforms for terpene biosynthetic gene clusters and the inability to perform rational structure predictions based on genome sequence information, and finally 3) low production titers of bacterial terpenoids under standard laboratory conditions. These factors, in combination with general terpene properties such as lack of UV-absorbing functional groups, poor ionization properties, and ubiquitous odiferous terpenes that overshadow characteristic terpene signals (branching methyl groups) in NMR experiments, render the targeted isolation of terpenes highly challenging. Therefore, heterologous expression in modified host organisms could be the method of choice in most studies.

Understanding the mechanistic logic of terpene biosynthesis

The entry points of terpene biosynthesis are isopentenyl pyrophosphate (IPP, **6**) and dimethylallyl pyrophosphate (DMAPP, **7**), which are assembled through the 2-*C*-methyl-D-erythritol 4-phosphate (MEP) or the mevalonic acid (MVA) pathway. Each pathway uses different precursors and enzymes, and different organisms utilize either or both pathways. A typical textbook description then divides terpene biosynthesis into two phases (Figure 4): 1) hydrocarbon backbone assembly and cyclization catalyzed by oligoprenyl synthetases and terpene cyclases (terpene synthases, TCs), respectively, to

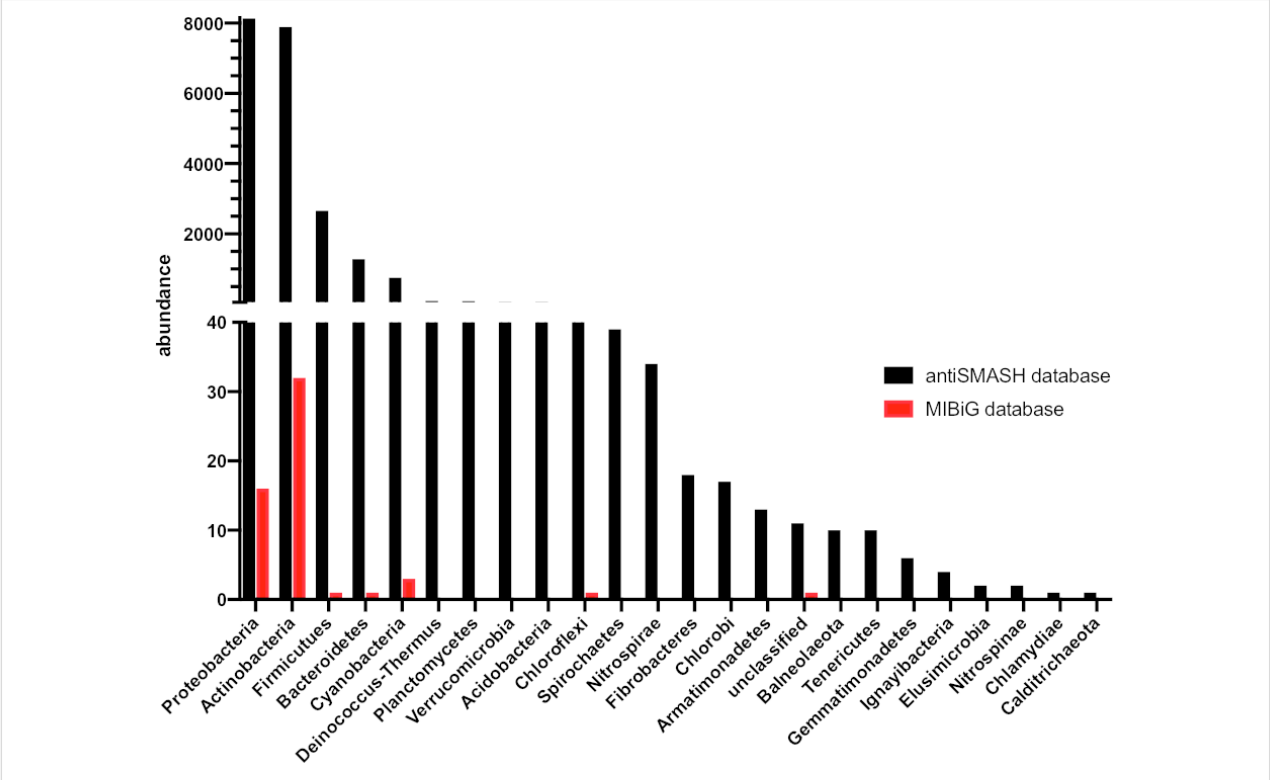


Figure 3: Abundance and distribution of bacterial terpene biosynthetic gene clusters as determined by genome mining (black) compared to experimentally characterized bacterial terpene BGCs (red).

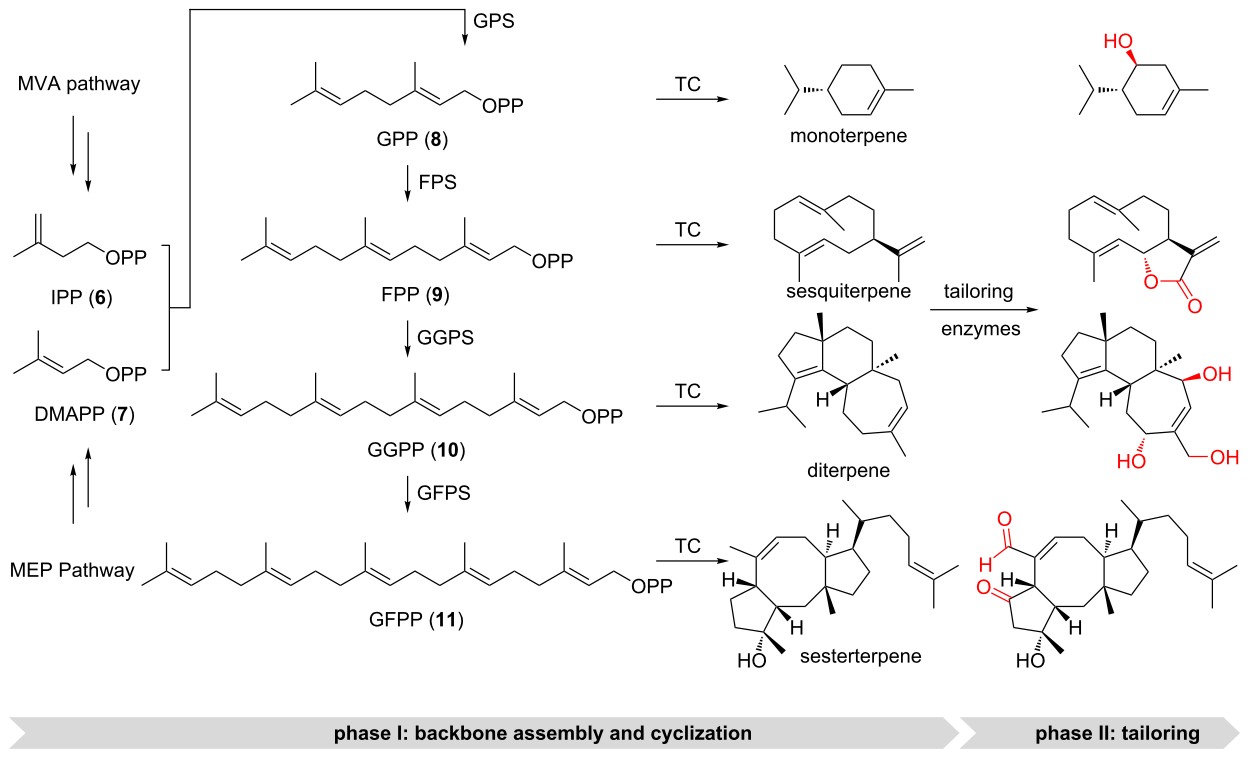


Figure 4: Terpenoid biosynthesis. Terpenoid biosynthesis is divided into two phases, 1) terpene scaffold generation and 2) terpene scaffold functionalization. Representative examples are shown with the modifications installed in the tailoring step highlighted in red. MEP: 2-*C*-methyl-D-erythritol 4-phosphate pathway, MVA: mevalonic acid pathway, IPP: isopentenyl pyrophosphate, DMAPP: dimethylallyl pyrophosphate, GPS: geranyl pyrophosphate synthase, GPP: geranyl pyrophosphate, FPS: farnesyl pyrophosphate synthase, FPP: farnesyl pyrophosphate, GGPS: geranylgeranyl pyrophosphate synthase, GGPP: geranylgeranyl pyrophosphate, GFPS: geranylfarnesyl pyrophosphate synthase, and GFPP geranylfarnesyl pyrophosphate.

generate terpene scaffolds and 2) the decoration phase, during which the terpene scaffold can get heavily modified by tailoring enzymes [45-47].

Phase 1) terpene scaffold generation

Unlike core biosynthetic enzymes or domains of other natural product classes (PKSs and NRPSs), bacterial TCs show only little overall sequence similarity [1]. The lack of conservation in primary sequence has slowed down our understanding of terpene cyclization and, as a result, hampered the development of efficient genome mining platforms for the detection of bacterial terpene BGCs. antiSMASH [48] is the only open source web application that annotates bacterial terpene BGCs. As a result of the low conserved sequence similarity between TCs, terpene BGCs are missed more frequently by antiSMASH than other BGC classes [49].

In contrast to the well-characterized biosynthetic rules of thiotemplate biosynthetic pathways that enable relatively precise predictions of natural product core structures (e.g., the colinearity rule in NRPSs and *cis*-acyltransferase polyketide synthases [8]), no such rules exist for the predictions of the cyclic hydrocarbon backbone produced by TCs [1]. This is likely a result of the inherent differences between these biosynthetic machineries. In modular assembly lines, each individual domain is responsible for one defined biosynthetic transformation, while TCs rather act as chaperones [10] to guide oligoprenyl pyrophosphate precursors through a cascade of cyclization reactions.

Textbook TCs are categorized into two classes and differ in their mechanisms of substrate activation as well as their protein folds [37,50-52]. Type I TCs trigger the formation of highly reactive allylic cations by heterolytic cleavage of the terminal pyrophosphate of farnesyl diphosphate (FPP, **9**) or geranylgeranyl diphosphate (GGPP, **10**, Figure 5a) [11,53]. Upon binding of the precursor, the TC undergoes conformational changes to seal the hydrophobic binding pocket [54]. This induced-fit mechanism locks the acyclic precursor into a defined, preorganized conformation that positions the leaving pyrophosphate group and the nucleophilic alkenes in proximity to initiate the C–C-bond forming, carbocation-mediated cascade reactions [10]. The hydrophobic binding pocket stabilizes the reaction intermediates and tames the propagation of carbocations

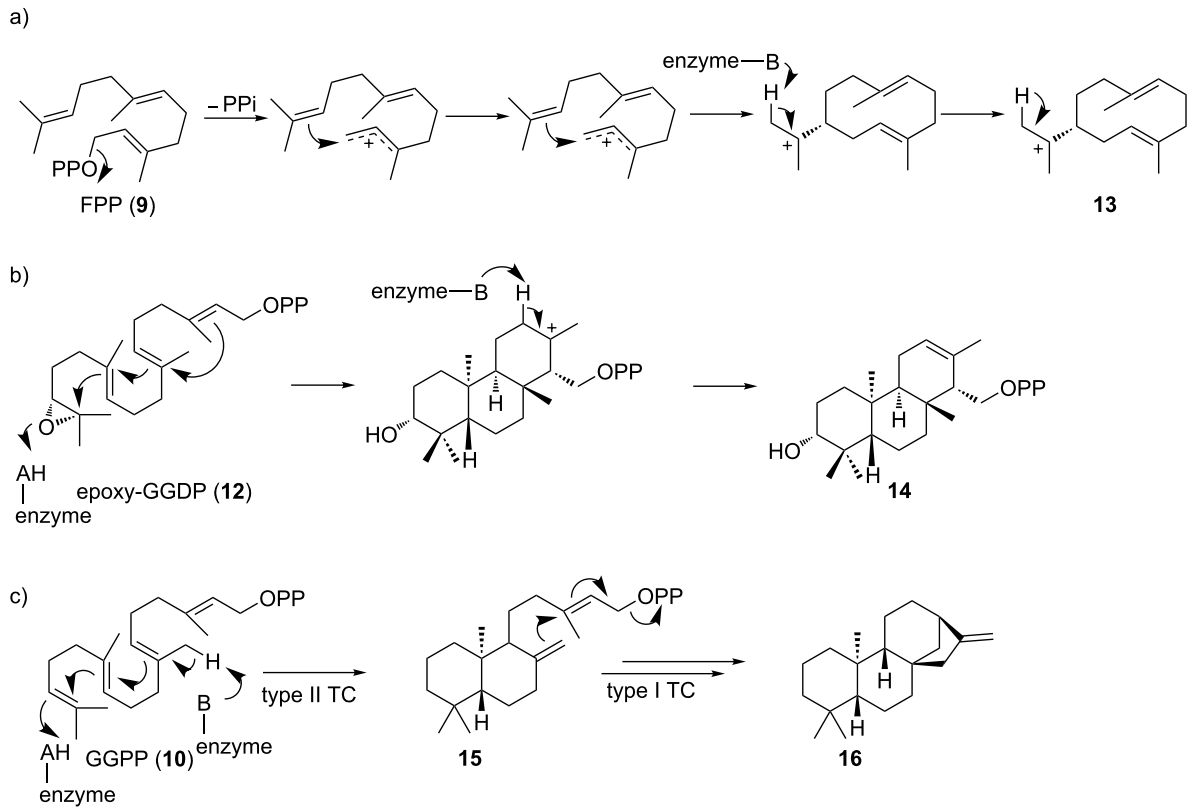


Figure 5: Mechanisms for type I, type II, and type II/type I tandem terpene cyclases. a) Tail-to-head class I germacrene A (**13**) cyclase. b) Head-to-tail type II brasilicardin (**14**) cyclase. c) Type II ent-copalyl diphosphate (**15**) synthase and type I ent-kaurene (**16**) cyclase.

through cation–π and other electrostatic interactions [54]. Moreover, TCs also assist intramolecular atom transfer and rearrangements including hydride or proton transfer and carbon shifts [10]. Eventually, the carbocation is quenched by deprotonation (E1-like) or nucleophilic attack (S_N1-like) of water [45]. In contrast to sesquiterpenes (type I TCs), diterpenes can either be generated by type I TCs or type II TCs [11,55]. Type II TCs initiate carbocation formation by Brønsted acid catalysis to protonate a terminal isoprene double bond or an epoxide ring (Figure 5b) [56]. Thus, cyclization mediated by type II TCs leads to an inverted direction of charge propagation along the oligoisoprene chain in the cyclization cascade [53]. The resulting cyclized product retains the pyrophosphate moiety and can serve as the substrate for a second cyclization catalyzed by type I TCs to generate even more complicated diterpene backbones (Figure 5c).

Although the majority of terpene backbones results from the direct cyclization of FPP (**9**) and GGPP (**10**) catalyzed by either type I or mixed type II/I cyclases, there are cases known to divert from these canonical paths. The two homoterpene pathways, responsible for 2-methlyisoborenol [57] and sodorifen [58] biosynthesis for example, require the action of a methyltransferase preceding cyclization. Surprisingly, more and more terpene BGCs are characterized that do not harbor classical TCs, but instead use a variety of different enzyme classes for the cyclization reaction. These atypical terpene cyclization reactions have recently been extensively reviewed [53]. Many of these alternative cyclization routes have been shown to resemble classical type I and II cyclization mechanisms initiated by the formation of highly reactive short-lived carbocation intermediates (e.g., prenyltransferases [59,60], vanadium-dependent bromoperoxidases [61,62], or methyltransferases [63]).

In addition, enzymes typically classified as tailoring enzymes, such as flavin-dependent monooxygenases [64] and cytochrome P450s (CYPs) [65], were reported to be involved in noncanonical terpene cyclization. Furthermore, both oxidative [66] and reductive cyclizations [67] have been described. The reductive cyclization reaction is particularly noteworthy, as the reduction and cyclization step are catalyzed by two distinct enzymes [67]. These examples show that the classical division of terpene biosynthesis into a cyclization and decoration phase needs to be modified, as it is now known that enzyme families, traditionally regarded as tailoring enzymes, can potentially be involved

in the cyclization reaction. The ever-increasing number of alternative terpene cyclization mechanisms suggests that nature has likely evolved additional, but so far undiscovered, modes of cyclization.

In vitro enzyme assays have been the method of choice for studying terpene cyclization (Supporting Information File 1, Table S1). The purified enzymes are particularly suitable for mechanistic investigations using strategically labeled substrates, as elegantly demonstrated in the studies on pentalenene synthase [68], geosmin synthase [69-71], and *epi*-isozizaene synthase [72]. Recently, all possible ^{13}C-labeled FPP (**9**) [73] and GGPP (**10**) [74] precursors were synthesized and used to track the movement of carbon atoms and test mechanistic hypotheses. Using enantiotopically doubly ^{13}C-^{2}H-labeled substrates, it is possible to determine the stereochemistry of a cyclization product by locating the ^{2}H atom and its relative position to other stereocenters [75]. In vitro terpene biosynthesis, however, might not always result in the production of the terpene skeleton produced natively or in the heterologous hosts (Figure 6a) [76]. In fact, it is occasionally observed that the same terpene cyclase generates different terpene skeletons dependent on the heterologous host in which it was expressed. While the exact reasons for the observed product variations are not fully understood, subtle differences in folding and reaction environment might be an explanation. In order to obtain larger quantities of terpenes, increasing the metabolic flux into isoprenoid pathways is often required. Overexpression of the bottleneck enzymes in the endogenous MEP pathway (*dxs*, *idi*, *ispD*, *ispF*, *fps/ggps*) or the entire exogenous MVA pathway in *E. coli* have been demonstrated to raise terpene titers dramatically [77] and have been applied for the production of several plant-derived terpene backbones, including amorpha-4,11-diene (**21**) [78,79] and taxadiene (**22**, Figure 6b) [80].

Despite efforts to characterize individual terpene cyclases and their modes of cyclization, no biosynthetic rules have so far been deduced that can be applied to a broad range of unrelated terpene cyclases.

Phase 2) terpene scaffold functionalization

Tailoring enzymes are important elements that further increase the structural complexity of terpenoids by adding various functional groups. The size of bacterial terpene BGCs can vary significantly, which can be largely attributed to the presence or absence of tailoring genes. Some BGCs can be as large as modular thiotemplate BGCs (e.g., phenalinolactone (42 kb) [81] and platensimycin/platencin (47/41 kb) [82]) while most are less than 10 kb in size and harbor only one or two putative tailoring enzymes. Among them, oxidoreductases, especially CYPs, are the most abundant tailoring enzymes involved in the diversification process (Supporting Information File 1, Table S2).

CYPs are heme-dependent iron proteins that catalyze a wide range of reactions [83,84]. The reactions typically involve substrate radical generation by the activated iron species and subsequent hydroxylation. Terpenes are mainly composed of nonactivated hydrocarbons that are mostly chemically indistinguishable, while many P450s are known to selectively act on specif-

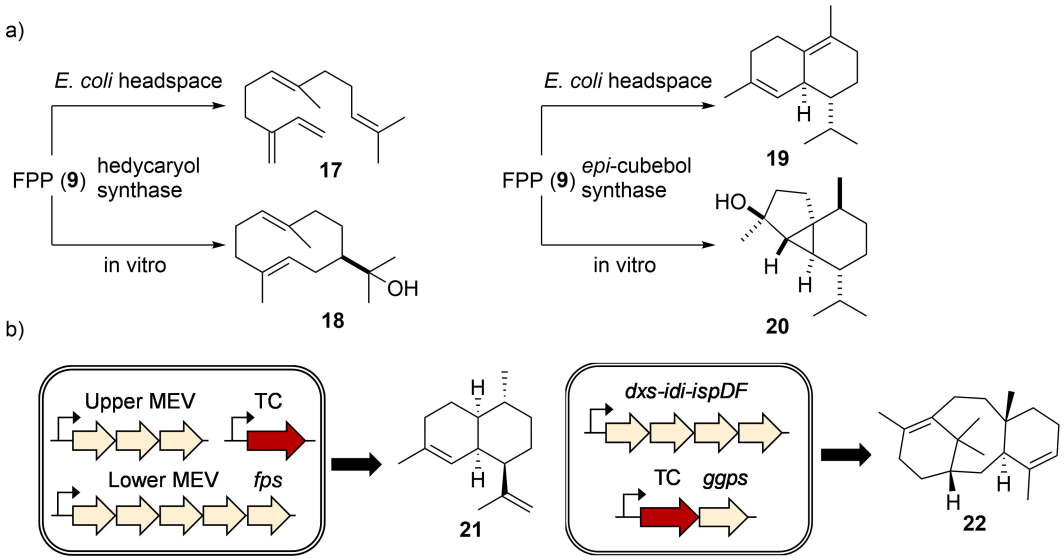


Figure 6: Functional TC characterization. a) Different terpenes were produced when hedycaryol (**18**) synthase and *epi*-cubebol (**20**) synthase were expressed in *E. coli* vs incubated with FPP (**9**) in vitro. b) Engineering the isoprenoid flux increases the titer of terpene production.

ic carbon atom(s). A continuous electron input is required to keep CYPs functional [83]. In general, soluble bacterial CYPs utilize class I redox systems, where the electron from NAD(P)H is delivered to CYPs through a ferredoxin reductase (FdR) and ferredoxin (FdX) [85]. Most bacteria have more than one FdR and FdX pair encoded in the genome [86]. It remains challenging to determine a priori which combination can reconstitute an active CYP, especially as the proximity of CYPs to these redox enzymes in the genome does not guarantee their partnership [87].

Studies of CYPs in bacterial terpenoid biosynthesis lags behind those of cyclases, and only a handful of examples are found in literature (Supporting Information File 1, Table S2). CYPs mostly catalyze prototypical hydroxylations, sometimes followed by further oxidation to carbonyls or carboxylates. Other reactions, such as epoxidation, ether bond formation, and structural rearrangement have also been reported (Figure 7). CYP114 in gibberellin (**5**) biosynthesis, for example, catalyzes the unique oxidation/six-membered to five-membered ring contraction of *ent*-kaurenoic acid (**27** and **28**, respectively, Figure 7b) [41].

Characterization of CYPs is typically achieved by in vitro or in vivo studies. *E. coli* is the most popular host for obtaining proteins for in vitro studies, and proper selection of a redox system is usually the obstacle to reconstitute CYP activity. Substrates for in vivo studies are either added to the culture or supplied by coexpression of upstream enzymes in the heterologous hosts. If *E. coli* is selected as the host, coexpression of a redox system is required, as *E. coli* does not harbor any CYPs [88].

In comparison to TCs, tailoring genes, such as CYPs, can be annotated with high confidence, yet likewise no predictions about the function of the corresponding enzyme's functions and stereo- or regiospecificity can be made. This lack of biosynthetic understanding is largely due to the relatively small number of characterized, bacteria-derived terpene-modifying enzymes. In addition, the fact that several CYPs have been shown to have relaxed substrate specificity, act on several intermediates, or catalyze multiple reactions, further complicates the in silico prediction.

(Bio)synthetic production of complex terpenoids

Heterologous expression is the most widely used method to study complex terpenoid biosyntheses. Since many bacterial terpenoid BGCs are actinomycete-derived, terpene BGCs are often expressed in model *Streptomyces* hosts, such as *Streptomyces albus*, *Streptomyces avermitilis*, *Streptomyces coelicolor*, and *Streptomyces lividans*, under the control of exogenous promoters [92-94]. To minimize the cellular resource competi-

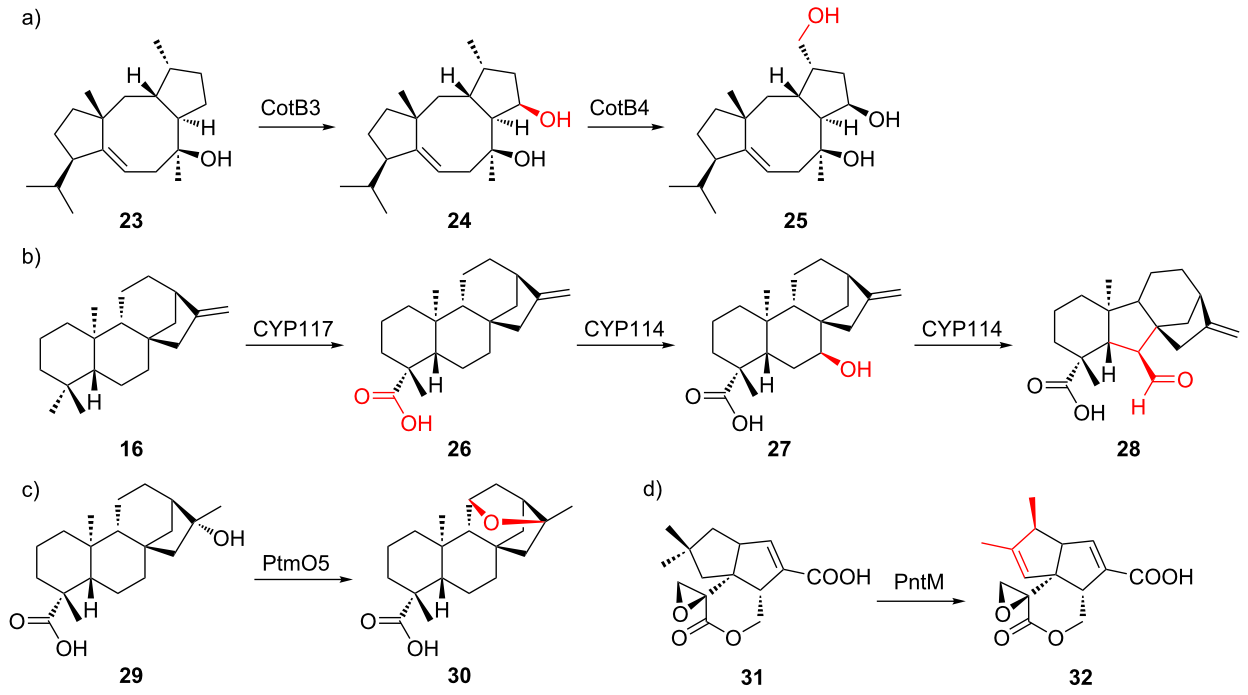


Figure 7: Selected examples of terpene modification by bacterial CYPs. a) Hydroxylation [89]. b) Carboxylation, hydroxylation, and ring contraction [41]. c) Ether formation [90]. d) Rearrangement [91].

tion and facilitate cleaner analysis, many of these hosts have been engineered to remove native secondary-metabolite BGCs and for optimized terpene precursor supply [1,95-97]. Like *Streptomyces* hosts, many *E. coli* hosts have been engineered to increase precursor supply, provide redox partners for CYP enzymes, and overproduce oligoprenyl diphosphate. As is the case for other natural product classes, off-target effects of heterologous host enzymes have been reported to alter terpenoid structures significantly (Figure 8) [98].

Despite difficulties in engineering terpene biosynthesis, total synthesis of complex terpenoids does not appear to be a viable alternative. Traditionally, complex terpenes are synthesized from small and cheap chiral terpenes as starting material in (linear) multistep syntheses [99]. This way, pre-existing oxidized functionalities need to be maintained and propagated with complex protecting group strategies en route to the desired target [100]. Most importantly, however, these total syntheses can easily require 50 or more steps, with poor overall yields, and are hence far away from nature's efficiency and pathway-encoded diversity [47,101]. The insights gained into the biosynthesis of complex terpenes have resulted in alternative approaches towards complex terpenoids – biomimetic syntheses. Cyclization reactions have been successfully mimicked using a wide variety of conditions. In analogy to nature's biosynthetic strategies outlined above, cationic, Diels–Alder, oxidative, and radical cyclization strategies have been successfully applied in total syntheses of terpenoids [53]. In addition, chemists are currently creatively exploring means to mimic classical terpene cyclases and their chaperone-like properties [102-104]. Unlike the generation of hydrocarbon backbones in nature, even the biomimetic synthetic construction of terpene scaffolds is usually a multistep process [17]. The second phase of a biomimetic synthesis, in analogy to terpene biosynthesis, involves the challenging introduction of functional groups (mostly hydroxy groups) into nonactivated C–H bonds [46,105-112]. The absence of directing functional groups, however, renders these regioselective oxidations of nonactivated carbon atoms highly challenging, as the chemical functionalization of nonactivated carbon atoms is one of the remaining challenges in organic synthesis [100,112,113]. Despite the progress made in total synthesis of terpenoids and the introduction of the two-phase mimicry of terpene biosynthesis during biomimetic syntheses, it has become obvious that total synthesis is currently not an efficient alternative to either the biotechnological production of terpenoids or the biosynthetic expansion of terpene chemical space [100,113].

Engineering terpene biosynthetic features to produce novel, non-natural terpenoids

Terpene cyclases

Most bacterial TCs produce one major product, while few of them are known to produce multiple products [1,50,76,114-117]. Among them, 10-*epi*-cubebol synthase from *Sorangium cellulosum* Soce56 is the most prolific bacterial TC known. It produces at least 25 different sesquiterpene skeletons in addi-

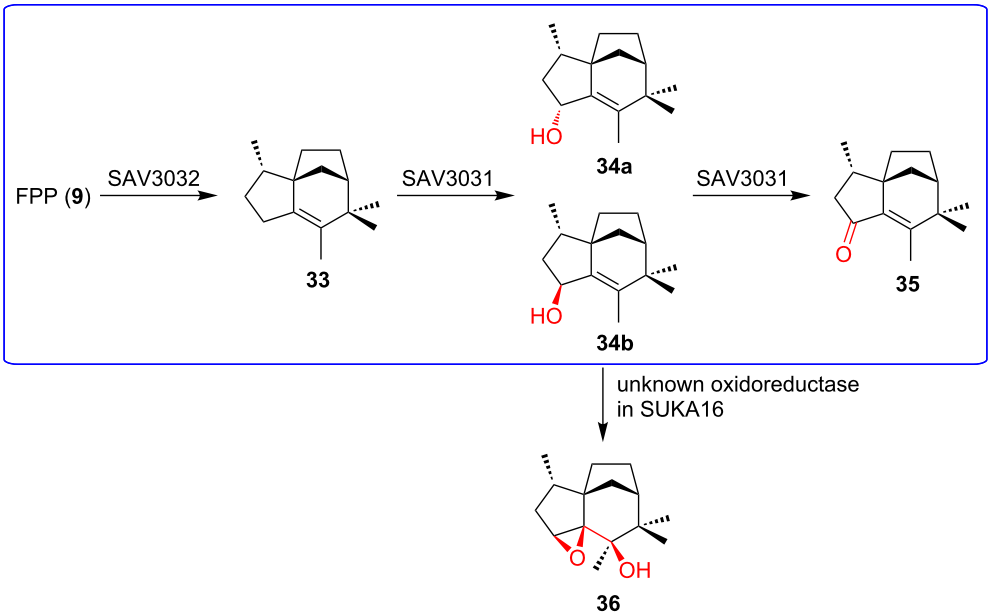


Figure 8: Off-target effects observed during heterologous expression of terpenoid BGCs. Unexpected oxidation of **34b** by an oxidoreductase in the heterologous host *S. avermitilis* SUKA16. The native pathway is highlighted in blue.

tion to 10-*epi*-cubebol [116]. In some cases, bacterial TCs are able to accept oligoprenyl diphosphates with different lengths. Spata-13,17-diene (**39**) synthase is an extreme example that can convert FPP (**9**), GGPP (**10**), and geranylfarnesyl diphosphate (GFPP, **11**) into sesquiterpenes, diterpenes, and sesterterpenes, respectively (**37**–**40**), though only diterpenes were observed in the culture headspace of the wild type producer (Figure 9a) [118]. These examples suggest a remarkable degree of plasticity of TC active sites in order to direct the intermediate to different trajectories and accommodate substrates with variable lengths. Indeed, besides the influence on kinetic properties by changing the conserved motifs, many bacterial TCs are able to produce novel skeletons through mutations of other active-site residues. This could result in either the arrest of catalytic intermediates or the creation of new trajectories to quench the cationic species. For instance, remodeling the hydrophobic pocket of the active site by single-point mutation, *epi*-isozizaene (**33**) synthase was engineered to produce various linear,

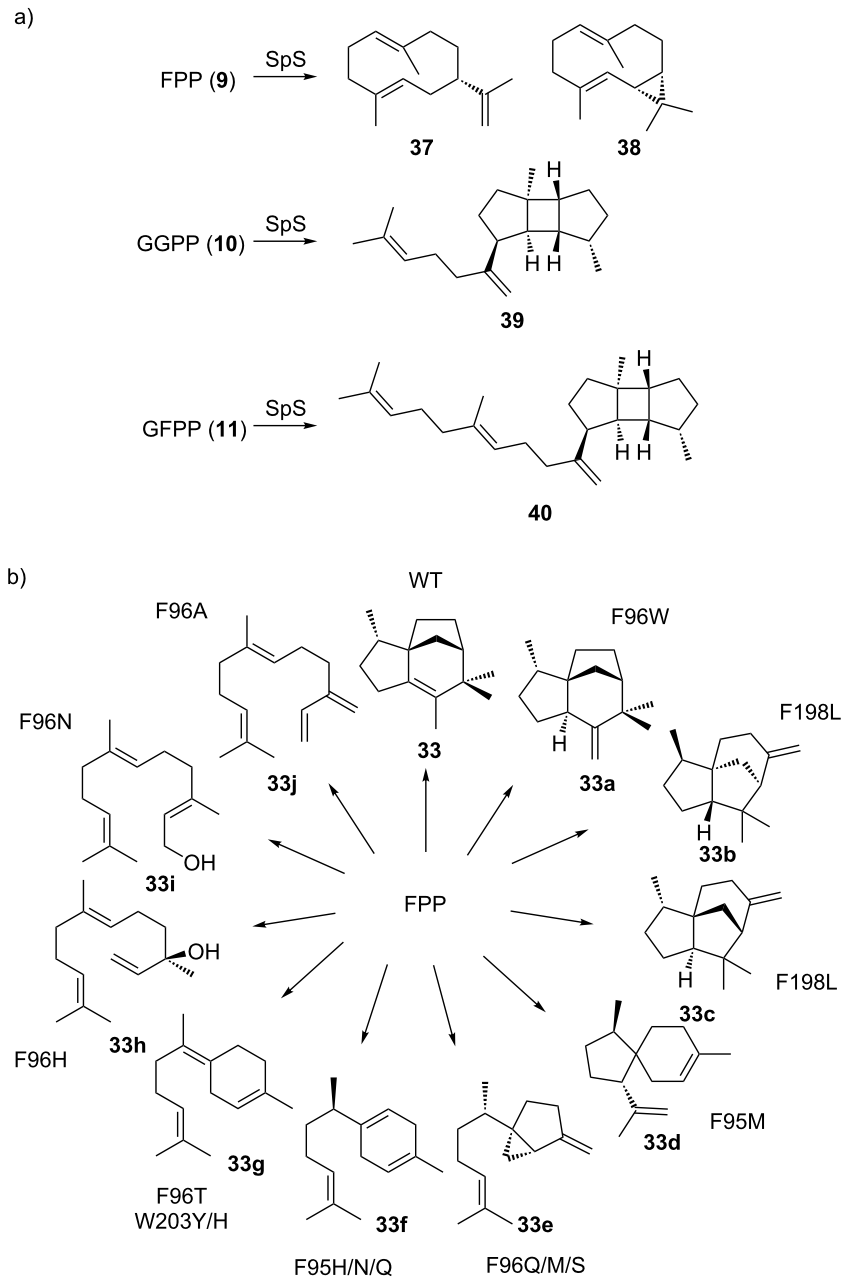


Figure 9: TC promiscuity and engineering. a) Spata-13,17-diene (**39**) synthase (SpS) can take C_{15} and C_{25} oligoprenyl diphosphate as substrates in addition to its C_{20} natural substrate. b) Selected examples of *epi*-isozizaene (**33**) synthase mutants that produce different sesquiterpene skeletons.

monocyclic, bicyclic, and tricyclic terpene skeletons (Figure 9b) [119-121]. Another prominent example is cyclooctat-9-en-7-ol (**23**, Figure 7a) synthase, for which the active-site residues responsible for cationic intermediate stabilization were identified through analysis of the crystal structure [122] and structural modeling [123]. Mutations of the identified residues were shown to alter product profiles and yielded several new terpene skeletons. With the growing mechanistic and structural knowledge of bacterial TCs, and the fact that many TCs naturally produce multiple products, there is great potential for the discovery of novel skeletons or enhancing the production of desired backbones among other side products, simply by rationally mutating the active-site residues.

The majority of bacterial type II diterpene TCs produce bicyclic labdane, halimadane, or clerodane skeletons with different stereochemistry, levels of unsaturation, and hydroxylation patterns [124], which undergo further conversion with their cognate type I diterpene TCs. It was therefore hypothesized that promiscuous type I TCs might be able to process different type II products to generate non-natural structural diversity. Indeed, pairing of type II TCs with type I TCs generated 13, 41, and 49 novel terpene skeletons, respectively, in three separate studies [125-127]. Though mostly plant TCs were employed, the few bacterial cyclases tested were also shown to be productive, suggesting this promiscuity is not limited to plant TCs. Type II/I diterpene TC tandems, however, are less prevalent in bacteria [11,55]. With the ever-increasing number of bacterial genome sequences available and the development of advanced bioinformatics tools, more type II/I diterpene TC tandems are likely to be discovered. By applying the same combinatorial concept, it is expected that the diterpene chemical space can be significantly expanded.

Cytochrome P450s

CYPs have the ability to install functional groups at nonactivated C–H bonds with exceptional selectivity. In contrast to the common perception of enzymes being highly substrate-specific, many CYPs are able to modify non-natural substrates but retain their high stereo- and regiochemical selectivity observed from the transformation of their natural substrate. For the exploration of substrate promiscuity of plant-derived *ent*-kaurene oxidases, twenty combinations of type II/I diterpene TCs were coexpressed with one of the two *ent*-kaurene oxidases (CYPs) and their native CPRs from different plant sources, which led to the production of 12 novel oxidized labdane-related diterpenes [128]. Recently, commercially available monoterpenes were incubated with cell-free extracts from *E. coli* expressing well-studied bacterial CYPs, resulting in 27 previously unreported terpenoids (selected examples are shown in Figure 10a) [129].

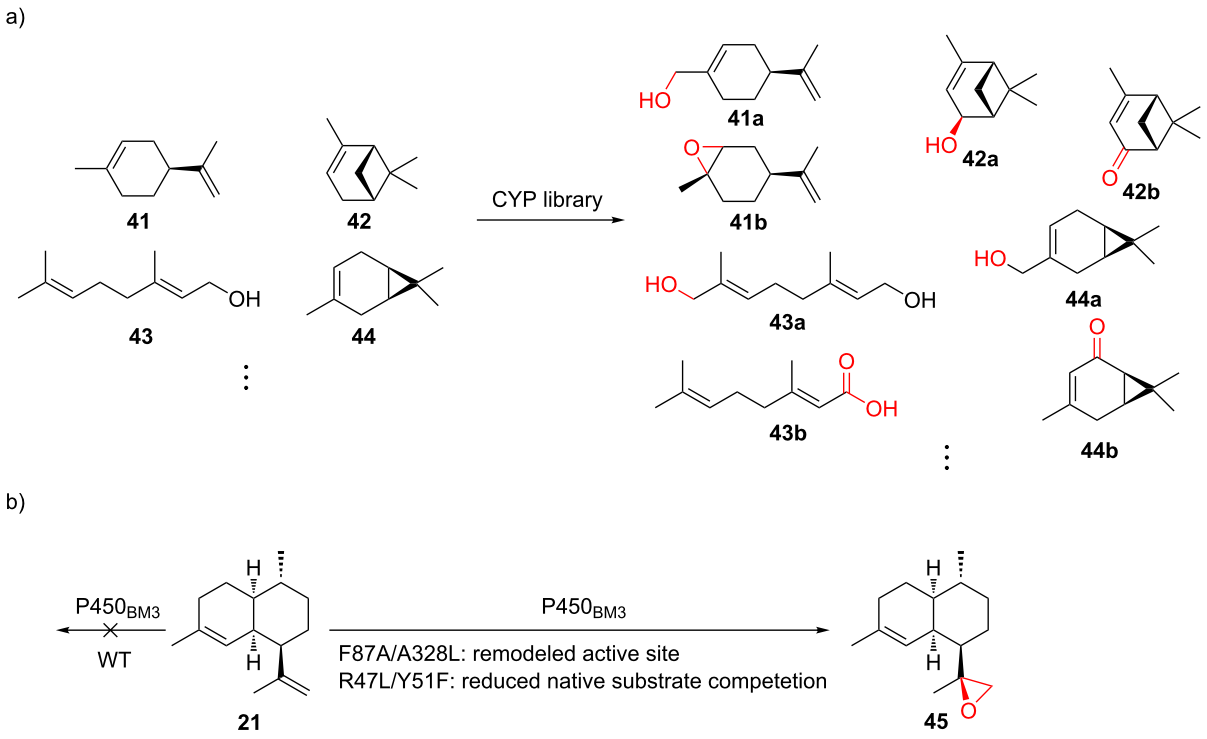


Figure 10: Substrate promiscuity and engineering of CYPs. a) Selected examples from using a CYP library to oxidize various monoterpenes. b) Rational engineering of $P450_{BM3}$ for epoxidation of amorphadiene (**21**). F87A/A328L and R47L/Y51F: mutations.

While only a limited number of bacterial CYPs for terpene modification has been characterized, many of them were reported to exhibit substrate promiscuity (Supporting Information File 1, Table S2). These promiscuous CYPs do not necessarily co-localize with TC-encoding genes in the genomes, suggesting their potential to diversify terpenes out of their biosynthetic contexts.

Directed evolution of CYPs and other heme-containing proteins has a long history [130] and could potentially be applied to terpenoid diversification. Rational design, on the other hand, relies on crystal structures, homology models, or mechanistic information to select the residues to be mutated. For example, $P450_{BM3}$ (CYP102A1) was rationally evolved to catalyze the epoxidation of amorphadiene (**21**) by expanding the active site, minimizing competing reactions, and facilitating substrate access (Figure 10b) [131]. Similarly, by mutating two residues of $P450_{BM3}$ that control the shape of the substrate binding cavity, the enzyme was engineered to accept different linear, monocyclic, and bicyclic terpenes [132]. The alteration of the native sequences uses either method or the combination of the two, serving as an indispensable tool to introduce another dimension for terpenoid diversification.

Expanding terpene chemical space through pathway engineering

Engineering of terpenoid biosynthetic pathways has focused on enhancing the metabolic flux to supply isoprenoid precursors (Figure 11a) [77,133]. Common strategies include expressing an entire exogenous isoprenoid pathway (e.g., the MVA pathway in *E. coli*) and overexpressing bottleneck enzymes (e.g., *dxs*, *idi*, *ispD*, *ispF* in the MEP pathway; *mk*, *hmgr* in the MVA pathway) [78-80,134-136]. Attempts are also being made to search for and engineer more active pathway enzyme variants [136-138]. As the balanced pathway expression is often the key to high terpenoid titers, these pathway genes have been grouped into "modules", using a multivariate-modular approach [80], in vitro enzyme activity assays [139], proteomics-based principal component analysis [140], qPCR/proteomics [141], and iterative grid search [142] to troubleshoot and guide the tuning of the promoter strengths and copy numbers of each module/enzyme (Figure 11b).

Toxicity of biosynthetic intermediates, endogenous regulation, and stability of genetic constructs are the main concern during these engineering efforts. IPP (**6**), FPP (**9**), and HMG-CoA (3-hydroxy-3-methylglutaryl-CoA) are known to be toxic

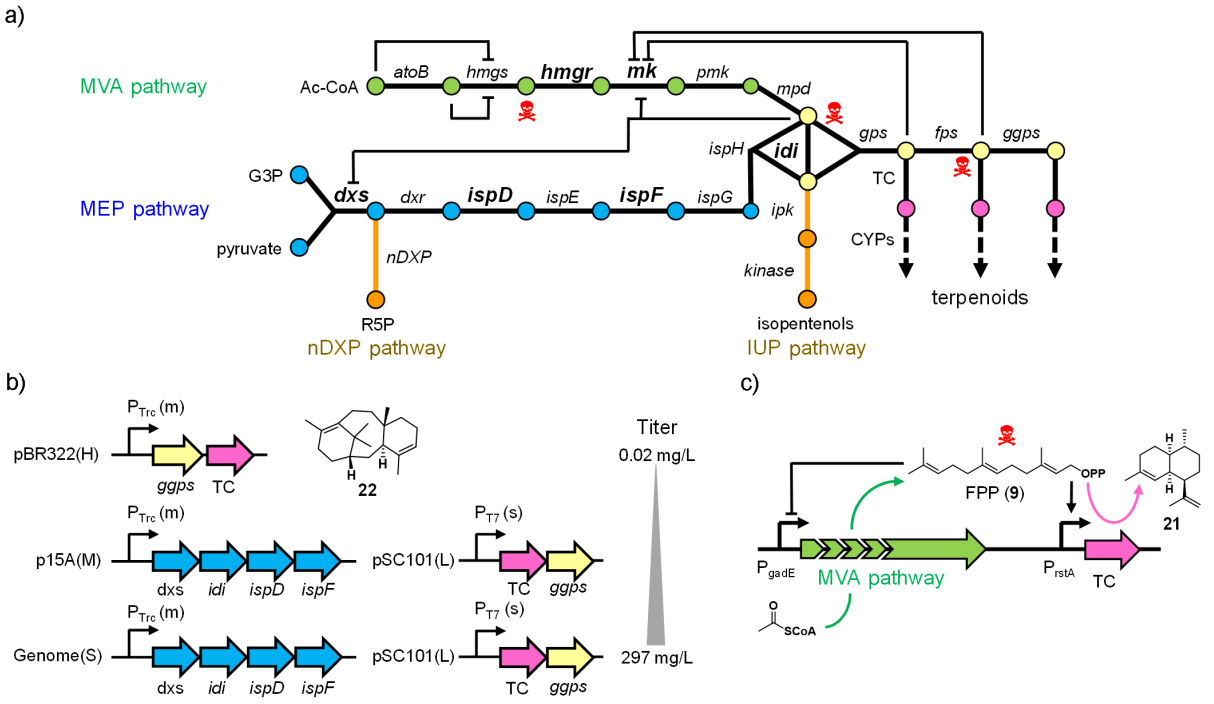


Figure 11: Engineering of terpenoid pathways. a) Metabolic network of terpenoid biosynthesis. Toxic intermediates are labeled with skull signs and known enzyme inhibition by intermediates is indicated. Bottleneck enzymes that have been subjected to optimization/engineering are highlighted in bold. Two novel pathways, nDXP and IUP, provide alternative entry points to terpenoid biosynthesis. b) Balancing gene expression for taxadiene (**22**) production in *E. coli*. The copy number of the constructs are labeled with upper case letters: H: high, M: medium, L: low, and S: single copy. The relative strength of the promoter is labeled with lower case letters: s: strong and m: medium. This tuning led to 15,000-fold increase in taxatidene titer. c) Two FPP (**9**)-responsive promoters were implemented to dynamically control the expression of MVA pathway genes and the terpene cyclase gene in order to prevent the accumulation of toxic FPP (**9**).

[78,143], and FPP-dependent stress-responsive promoters have recently been identified through microarray experiments to dynamically regulate pathways and minimize the accumulation of FPP (Figure 11c) [143]. Alternatively, in vitro pathway reconstitution led to the successful production of several monoterpenes, circumnavigating toxicity to the heterologous hosts [144]. Several intermediates are also known to inhibit enzyme activities in the pathway. Novel pathways, such as the novel 1-deoxy-D-xylulose 5-phosphate pathway (nDXP) [145] and the isopentenol utilization pathway (IUP) [146], that bypass (part of) the natural MVA or MEP pathway were designed to circumvent this complex endogenous interaction and regulation (Figure 11a). Lastly, integration of pathways into the *E. coli* genome and the subsequent tuning via recombineering or CRISPR/Cas9 tools has been pursued to solve the instability of plasmid-based systems [147,148].

The comparably small size of terpenoid BGCs, the promiscuity of the enzymes involved, and the fact that the core is usually built up in one step and then modified by tailoring reactions make terpenoid biosynthesis particularly amenable for generating non-natural variants through engineering. It is easy to imagine that the combinatorial pairing of promiscuous TCs and CYPs harbors great potential to achieve this goal, and the integration of machine learning and retrobiosynthetic algorithms could facilitate the design of constructs for specific terpenoid variants [149]. While it is now relatively straightforward to direct the flux to produce terpene skeletons, less is known about how to effectively support function of CYPs beyond natural/surrogate redox partners or fusing CYPs with redox partners [80,136,141,150,151]. One common observation is that CYP activity is not able to match the high flux of isoprenoids, leaving a significant portion of terpenes unmodified. Solutions to this problem, such as in silico redox partner prediction, remain to be explored [152].

One major bottleneck in these studies is the lack of high-throughput screening tools for terpenoid production and characterization. Most terpenoids do not have any chromogenic groups and their detection still relies on liquid chromatography–mass spectrometry setups, limiting the number of screens that can be done. Though genetic construction, culturing bacteria, terpenoid extraction, and chromatographic analysis can be carried out (semi-)automatedly, terpenoid structure elucidation constitutes another major bottleneck of these endeavors. At this end, the development of powerful chemoinformatic platforms is likely to assist in detecting putative terpenoids.

In the future, synthetic biology will likely allow the expansion of terpene chemical space by efficiently producing novel terpenes in genome-optimized host strains and the generation of non-natural terpenes through the fine control of terpene cyclases and tailoring enzymes. The reduced cost of DNA synthesis [153] and advanced cloning techniques [154], such as Gibson Assembly [155] and Golden Gate cloning [156], has made pathway construction more economically efficient, and the genetic parts can also be retrieved, recombined, and reused easily for different applications. Insights into terpene biosynthesis, along with advancements in synthetic biology, will pave the way towards the sustainable and selective biotechnological production of high-value terpenes, such as taxol (**1**). One recent study describes the successful reconstitution of THC (**3**) biosynthesis in yeast by upregulating MVA pathway genes, mutating *erg20* (*fps*) in favor of GPP (**8**) over FPP (**9**) production, using genes from multiple different species to supply hexonyl-CoA, and identifying key coupling enzyme through bioinformatic analysis [157]. THC variants could also be obtained via feeding the engineered strain with substrate analogs. This integration of multiple strategies highlights the great potential of synthetic biology for the production of valuable complex terpenoids [157].

Conclusion

Despite the ubiquitous distribution of terpene biosynthetic pathways in bacteria, only a few terpenes of bacterial origin have been characterized, and their biosynthesis is for the most part poorly understood. The lack of bio-/chemoinformatics platforms to predict terpene core structures coupled with their physicochemical properties have rendered the targeted isolation of complex terpenoids from their native producers highly challenging. Since terpene biosynthetic pathways are comparably small, heterologous expression of entire pathways in suitable host strains is the method of choice to retrieve the biosynthetic treasures hidden in bacterial genomes. Moreover, the biosynthetic logic of terpene biosynthetic pathways with a single enzyme for hydrocarbon backbone cyclization and multiple tailoring enzymes is ideally suited for synthetic biology approaches to engineer new pathways. These new pathways would enable the combinatorial expansion of terpene chemical space and the creation of non-natural biosynthetic pathways for the production of high-value terpenoids, especially pharmaceutical agents. The development of highly manipulable host strains with multiple orthogonally inducible promoters allows for the tight control, timing, and analysis of these combinatorial endeavors. In combination with directed mutagenesis to relax, alter, and expand cyclization, tailoring patterns, and substrate scope, synthetic biology is ideally suited to generate non-natural terpenes with promising drug-like properties and biological activities. Results obtained from these studies will also assist in refining our understanding of bacterial terpenoid biosynthesis.

Supporting Information

Supporting Information File 1

Additional figures and tables.
[https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-15-283-S1.pdf]

ORCID® iDs

Geng-Min Lin - https://orcid.org/0000-0001-8123-0921
Jon Clardy - https://orcid.org/0000-0003-0213-8356

References

1. Yamada, Y.; Kuzuyama, T.; Komatsu, M.; Shin-ya, K.; Omura, S.; Cane, D. E.; Ikeda, H. *Proc. Natl. Acad. Sci. U. S. A.* **2015,** *112,* 857–862. doi:10.1073/pnas.1422108112
2. Chappell, J. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1995,** *46,* 521–547. doi:10.1146/annurev.pp.46.060195.002513
3. Gershenzon, J.; Dudareva, N. *Nat. Chem. Biol.* **2007,** *3,* 408–414. doi:10.1038/nchembio.2007.5
4. Wani, M. C.; Taylor, H. L.; Wall, M. E.; Coggon, P.; McPhail, A. T. *J. Am. Chem. Soc.* **1971,** *93,* 2325–2327. doi:10.1021/ja00738a045
5. Tu, Y. *Angew. Chem., Int. Ed.* **2016,** *55,* 10210–10226. doi:10.1002/anie.201601967
6. Wang, G. T. W.; Bidigare, R. R. Terpenoids As Therapeutic Drugs and Pharmaceutical Agents. In *Natural Products - Drug Discovery and Therapeutic Medicine;* Zhang, L.; Demain, A. L., Eds.; Humana Press: Totowa, NJ, 2005; pp 197–227. doi:10.1007/978-1-59259-976-9_9
7. Fischbach, M. A.; Walsh, C. T. *Chem. Rev.* **2006,** *106,* 3468–3496. doi:10.1021/cr0503097
8. Hertweck, C. *Angew. Chem., Int. Ed.* **2009,** *48,* 4688–4716. doi:10.1002/anie.200806121
9. Austin, M. B.; O'Maille, P. E.; Noel, J. P. *Nat. Chem. Biol.* **2008,** *4,* 217–222. doi:10.1038/nchembio0408-217
10. Driller, R.; Janke, S.; Fuchs, M.; Warner, E.; Mhashal, A. R.; Major, D. T.; Christmann, M.; Bruck, T.; Loll, B. *Nat. Commun.* **2018,** *9,* No. 3971. doi:10.1038/s41467-018-06325-8
11. Dickschat, J. S. *Nat. Prod. Rep.* **2016,** *33,* 87–110. doi:10.1039/c5np00102a
12. Janocha, S.; Schmitz, D.; Bernhardt, R. Terpene Hydroxylation with Microbial Cytochrome P450 Monooxygenases. In *Biotechnology of Isoprenoids;* Schrader, J.; Bohlmann, J., Eds.; Springer: Cham, 2015; pp 215–250. doi:10.1007/10_2014_296
13. Donehower, R. C. *Oncologist* **1996,** *1,* 240–243. doi:10.1177/174498719600100317
14. Kingston, D. G. I. *Chem. Commun.* **2001,** 867–880. doi:10.1039/b100070p
15. Wender, P. A.; Badham, N. F.; Conway, S. P.; Floreancig, P. E.; Glass, T. E.; Houze, J. B.; Krauss, N. E.; Lee, D.; Marquess, D. G.; McGrane, P. L.; Meng, W.; Natchus, M. G.; Shuker, A. J.; Sutton, J. C.; Taylor, R. E. *J. Am. Chem. Soc.* **1997,** *119,* 2757–2758. doi:10.1021/ja963539z
16. Wender, P. A.; Badham, N. F.; Conway, S. P.; Floreancig, P. E.; Glass, T. E.; Gränicher, C.; Houze, J. B.; Jänichen, J.; Lee, D.; Marquess, D. G.; McGrane, P. L.; Meng, W.; Mucciaro, T. P.; Mühlebach, M.; Natchus, M. G.; Paulsen, H.; Rawlins, D. B.; Satkofsky, J.; Shuker, A. J.; Sutton, J. C.; Taylor, R. E.; Tomooka, K. *J. Am. Chem. Soc.* **1997,** *119,* 2755–2756. doi:10.1021/ja9635387
17. Mendoza, A.; Ishihara, Y.; Baran, P. S. *Nat. Chem.* **2012,** *4,* 21–25. doi:10.1038/nchem.1196
18. Lin, X.; Hezari, M.; Koepp, A. E.; Floss, H. G.; Croteau, R. *Biochemistry* **1996,** *35,* 2968–2977. doi:10.1021/bi9526239
19. Wildung, M. R.; Croteau, R. *J. Biol. Chem.* **1996,** *271,* 9201–9204. doi:10.1074/jbc.271.16.9201
20. Jennewein, S.; Croteau, R. *Appl. Microbiol. Biotechnol.* **2001,** *57,* 13–19. doi:10.1007/s002530100757
21. Wuts, P. G. *Curr. Opin. Drug Discovery Dev.* **1998,** *1,* 329–337.
22. Lodzki, M.; Godin, B.; Rakou, L.; Mechoulam, R.; Gallily, R.; Touitou, E. *J. Controlled Release* **2003,** *93,* 377–387. doi:10.1016/j.jconrel.2003.09.001
23. Leweke, F. M.; Piomelli, D.; Pahlisch, F.; Muhl, D.; Gerth, C. W.; Hoyer, C.; Klosterkötter, J.; Hellmich, M.; Koethe, D. *Transl. Psychiatry* **2012,** *2,* No. e94. doi:10.1038/tp.2012.15
24. Devinsky, O.; Cilio, M. R.; Cross, H.; Fernandez-Ruiz, J.; French, J.; Hill, C.; Katz, R.; Di Marzo, V.; Jutras-Aswad, D.; Notcutt, W. G.; Martinez-Orgado, J.; Robson, P. J.; Rohrback, B. G.; Thiele, E.; Whalley, B.; Friedman, D. *Epilepsia* **2014,** *55,* 791–802. doi:10.1111/epi.12631
25. Mechoulam, R.; Peters, M.; Murillo-Rodriguez, E.; Hanuš, L. O. *Chem. Biodiversity* **2007,** *4,* 1678–1692. doi:10.1002/cbdv.200790147
26. Challis, G. L.; Ravel, J.; Townsend, C. A. *Chem. Biol.* **2000,** *7,* 211–224. doi:10.1016/s1074-5521(00)00091-0
27. Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. *Chem. Biol.* **1999,** *6,* 493–505. doi:10.1016/s1074-5521(99)80082-9
28. Phelan, V. V.; Du, Y.; McLean, J. A.; Bachmann, B. O. *Chem. Biol.* **2009,** *16,* 473–478. doi:10.1016/j.chembiol.2009.04.007
29. Reeves, C. D.; Murli, S.; Ashley, G. W.; Piagentini, M.; Hutchinson, C. R.; McDaniel, R. *Biochemistry* **2001,** *40,* 15464–15470. doi:10.1021/bi015864r
30. Jenner, M.; Afonso, J. P.; Bailey, H. R.; Frank, S.; Kampa, A.; Piel, J.; Oldham, N. J. *Angew. Chem., Int. Ed.* **2015,** *54,* 1817–1821. doi:10.1002/anie.201410219
31. Helfrich, E. J. N.; Ueoka, R.; Dolev, A.; Rust, M.; Meoded, R. A.; Bhushan, A.; Califano, G.; Costa, R.; Gugger, M.; Steinbeck, C.; Moreno, P.; Piel, J. *Nat. Chem. Biol.* **2019,** *15,* 813–821. doi:10.1038/s41589-019-0313-7
32. Jensen, K.; Niederkrüger, H.; Zimmermann, K.; Vagstad, A. L.; Moldenhauer, J.; Brendel, N.; Frank, S.; Pöplau, P.; Kohlhaas, C.; Townsend, C. A.; Oldiges, M.; Hertweck, C.; Piel, J. *Chem. Biol.* **2012,** *19,* 329–339. doi:10.1016/j.chembiol.2012.01.005
33. Yeh, E.; Kohli, R. M.; Bruner, S. D.; Walsh, C. T. *ChemBioChem* **2004,** *5,* 1290–1293. doi:10.1002/cbic.200400077
34. Ueoka, R.; Uria, A. R.; Reiter, S.; Mori, T.; Karbaum, P.; Peters, E. E.; Helfrich, E. J. N.; Morinaka, B. I.; Gugger, M.; Takeyama, H.; Matsunaga, S.; Piel, J. *Nat. Chem. Biol.* **2015,** *11,* 705–712. doi:10.1038/nchembio.1870
35. Crüsemann, M.; Kohlhaas, C.; Piel, J. *Chem. Sci.* **2013,** *4,* 1041–1045. doi:10.1039/c2sc21722h
36. Jenke-Kodama, H.; Dittmann, E. *Phytochemistry* **2009,** *70,* 1858–1866. doi:10.1016/j.phytochem.2009.05.021
37. Gao, Y.; Honzatko, R. B.; Peters, R. J. *Nat. Prod. Rep.* **2012,** *29,* 1153–1175. doi:10.1039/c2np20059g

38. Firn, R. D.; Jones, C. G. *Nat. Prod. Rep.* **2003,** *20,* 382–391. doi:10.1039/b208815k
39. Jones, C. G.; Firn, R. D. *Philos. Trans. R. Soc., B* **1991,** *333,* 273–280. doi:10.1098/rstb.1991.0077
40. Fischbach, M. A.; Clardy, J. *Nat. Chem. Biol.* **2007,** *3,* 353–355. doi:10.1038/nchembio0707-353
41. Nett, R. S.; Montanares, M.; Marcassa, A.; Lu, X.; Nagel, R.; Charles, T. C.; Hedden, P.; Rojas, M. C.; Peters, R. J. *Nat. Chem. Biol.* **2017,** *13,* 69–74. doi:10.1038/nchembio.2232
42. Hedden, P.; Phillips, A. L.; Rojas, M. C.; Carrera, E.; Tudzynski, B. *J. Plant Growth Regul.* **2001,** *20,* 319–331. doi:10.1007/s003440010037
43. Blin, K.; Pascal Andreu, V.; de los Santos, E. L. C.; Del Carratore, F.; Lee, S. Y.; Medema, M. H.; Weber, T. *Nucleic Acids Res.* **2019,** *47,* D625–D630. doi:10.1093/nar/gky1060
44. Medema, M. H.; Kottmann, R.; Yilmaz, P.; Cummings, M.; Biggins, J. B.; Blin, K.; de Bruijn, I.; Chooi, Y. H.; Claesen, J.; Coates, R. C.; Cruz-Morales, P.; Duddela, S.; Düsterhus, S.; Edwards, D. J.; Fewer, D. P.; Garg, N.; Geiger, C.; Gomez-Escribano, J. P.; Greule, A.; Hadjithomas, M.; Haines, A. S.; Helfrich, E. J. N.; Hillwig, M. L.; Ishida, K.; Jones, A. C.; Jones, C. S.; Jungmann, K.; Kegler, C.; Kim, H. U.; Kötter, P.; Krug, D.; Masschelein, J.; Melnik, A. V.; Mantovani, S. M.; Monroe, E. A.; Moore, M.; Moss, N.; Nützmann, H.-W.; Pan, G.; Pati, A.; Petras, D.; Reen, F. J.; Rosconi, F.; Rui, Z.; Tian, Z.; Tobias, N. J.; Tsunematsu, Y.; Wiemann, P.; Wyckoff, E.; Yan, X.; Yim, G.; Yu, F.; Xie, Y.; Aigle, B.; Apel, A. K.; Balibar, C. J.; Balskus, E. P.; Barona-Gómez, F.; Bechthold, A.; Bode, H. B.; Borriss, R.; Brady, S. F.; Brakhage, A. A.; Caffrey, P.; Cheng, Y.-Q.; Clardy, J.; Cox, R. J.; De Mot, R.; Donadio, S.; Donia, M. S.; van der Donk, W. A.; Dorrestein, P. C.; Doyle, S.; Driessen, A. J. M.; Ehling-Schulz, M.; Entian, K.-D.; Fischbach, M. A.; Gerwick, L.; Gerwick, W. H.; Gross, H.; Gust, B.; Hertweck, C.; Höfte, M.; Jensen, S. E.; Ju, J.; Katz, L.; Kaysser, L.; Klassen, J. L.; Keller, N. P.; Kormanec, J.; Kuipers, O. P.; Kuzuyama, T.; Kyrpides, N. C.; Kwon, H.-J.; Lautru, S.; Lavigne, R.; Lee, C. Y.; Linquan, B.; Liu, X.; Liu, W.; Luzhetskyy, A.; Mahmud, T.; Mast, Y.; Méndez, C.; Metsä-Ketelä, M.; Micklefield, J.; Mitchell, D. A.; Moore, B. S.; Moreira, L. M.; Müller, R.; Neilan, B. A.; Nett, M.; Nielsen, J.; O'Gara, F.; Oikawa, H.; Osbourn, A.; Osburne, M. S.; Ostash, B.; Payne, S. M.; Pernodet, J.-L.; Petricek, M.; Piel, J.; Ploux, O.; Raaijmakers, J. M.; Salas, J. A.; Schmitt, E. K.; Scott, B.; Seipke, R. F.; Shen, B.; Sherman, D. H.; Sivonen, K.; Smanski, M. J.; Sosio, M.; Stegmann, E.; Süssmuth, R. D.; Tahlan, K.; Thomas, C. M.; Tang, Y.; Truman, A. W.; Viaud, M.; Walton, J. D.; Walsh, C. T.; Weber, T.; van Wezel, G. P.; Wilkinson, B.; Willey, J. M.; Wohlleben, W.; Wright, G. D.; Ziemert, N.; Zhang, C.; Zotchev, S. B.; Breitling, R.; Takano, E.; Glöckner, F. O. *Nat. Chem. Biol.* **2015,** *11,* 625–631. doi:10.1038/nchembio.1890
45. Davis, E. M.; Croteau, R. Cyclization Enzymes in the Biosynthesis of Monoterpenes, Sesquiterpenes, and Diterpenes. In *Biosynthesis - Aromatic Polyketides, Isoprenoids, Alkaloids;* Leeper, F. J.; Vederas, J. C., Eds.; Springer: Berlin, 2000; pp 53–95. doi:10.1007/3-540-48146-x_2
46. Chen, K.; Baran, P. S. *Nature* **2009,** *459,* 824–828. doi:10.1038/nature08043
47. Maimone, T. J.; Baran, P. S. *Nat. Chem. Biol.* **2007,** *3,* 396–407. doi:10.1038/nchembio.2007.1
48. Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. *Nucleic Acids Res.* **2019,** *47,* W81–W87. doi:10.1093/nar/gkz310
49. Medema, M. H.; Blin, K.; Cimermancic, P.; de Jager, V.; Zakrzewski, P.; Fischbach, M. A.; Weber, T.; Takano, E.; Breitling, R. *Nucleic Acids Res.* **2011,** *39,* W339–W346. doi:10.1093/nar/gkr466
50. Christianson, D. W. *Chem. Rev.* **2006,** *106,* 3412–3442. doi:10.1021/cr050286w
51. Christianson, D. W. *Curr. Opin. Chem. Biol.* **2008,** *12,* 141–150. doi:10.1016/j.cbpa.2007.12.008
52. Wendt, K. U.; Schulz, G. E. *Structure* **1998,** *6,* 127–133. doi:10.1016/s0969-2126(98)00015-x
53. Baunach, M.; Franke, J.; Hertweck, C. *Angew. Chem., Int. Ed.* **2015,** *54,* 2604–2626. doi:10.1002/anie.201407883
54. Baer, P.; Rabe, P.; Fischer, K.; Citron, C. A.; Klapschinski, T. A.; Groll, M.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2014,** *53,* 7652–7656. doi:10.1002/anie.201403648
55. Dickschat, J. S. *Angew. Chem., Int. Ed.* **2019,** *58,* 15964–15976. doi:10.1002/anie.201905312
56. Smanski, M. J.; Peterson, R. M.; Huang, S.-X.; Shen, B. *Curr. Opin. Chem. Biol.* **2012,** *16,* 132–141. doi:10.1016/j.cbpa.2012.03.002
57. Wang, C.-M.; Cane, D. E. *J. Am. Chem. Soc.* **2008,** *130,* 8908–8909. doi:10.1021/ja803639g
58. von Reuss, S.; Domik, D.; Lemfack, M. C.; Magnus, N.; Kai, M.; Weise, T.; Piechulla, B. *J. Am. Chem. Soc.* **2018,** *140,* 11855–11862. doi:10.1021/jacs.8b08510
59. Hillwig, M. L.; Zhu, Q.; Liu, X. *ACS Chem. Biol.* **2014,** *9,* 372–377. doi:10.1021/cb400681n
60. Erickson, H. K.; Poulter, C. D. *J. Am. Chem. Soc.* **2003,** *125,* 6886–6888. doi:10.1021/ja034520g
61. Carter-Franklin, J. N.; Parrish, J. D.; Tschirret-Guth, R. A.; Little, R. D.; Butler, A. *J. Am. Chem. Soc.* **2003,** *125,* 3688–3689. doi:10.1021/ja029271v
62. Carter-Franklin, J. N.; Butler, A. *J. Am. Chem. Soc.* **2004,** *126,* 15060–15066. doi:10.1021/ja047925p
63. Awakawa, T.; Zhang, L.; Wakimoto, T.; Hoshino, S.; Mori, T.; Ito, T.; Ishikawa, J.; Tanner, M. E.; Abe, I. *J. Am. Chem. Soc.* **2014,** *136,* 9910–9913. doi:10.1021/ja505224r
64. Xu, Z.; Baunach, M.; Ding, L.; Hertweck, C. *Angew. Chem., Int. Ed.* **2012,** *51,* 10293–10297. doi:10.1002/anie.201204087
65. Chooi, Y.-H.; Hong, Y. J.; Cacho, R. A.; Tantillo, D. J.; Tang, Y. *J. Am. Chem. Soc.* **2013,** *135,* 16805–16808. doi:10.1021/ja408966t
66. Taura, F.; Morimoto, S.; Shoyama, Y.; Mechoulam, R. *J. Am. Chem. Soc.* **1995,** *117,* 9766–9767. doi:10.1021/ja00143a024
67. Lichman, B. R.; Kamileen, M. O.; Titchiner, G. R.; Saalbach, G.; Stevenson, C. E. M.; Lawson, D. M.; O'Connor, S. E. *Nat. Chem. Biol.* **2019,** *15,* 71–79. doi:10.1038/s41589-018-0185-2
68. Cane, D. E.; Weiner, S. W. *Can. J. Chem.* **1994,** *72,* 118–127. doi:10.1139/v94-019
69. He, X.; Cane, D. E. *J. Am. Chem. Soc.* **2004,** *126,* 2678–2679. doi:10.1021/ja039929k
70. Jiang, J.; He, X.; Cane, D. E. *J. Am. Chem. Soc.* **2006,** *128,* 8128–8129. doi:10.1021/ja062669x
71. Jiang, J.; Cane, D. E. *J. Am. Chem. Soc.* **2008,** *130,* 428–429. doi:10.1021/ja077792i
72. Lin, X.; Cane, D. E. *J. Am. Chem. Soc.* **2009,** *131,* 6332–6333. doi:10.1021/ja901313v

73. Rabe, P.; Barra, L.; Rinkel, J.; Riclea, R.; Citron, C. A.; Klapschinski, T. A.; Janusko, A.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2015,** *54,* 13448–13451. doi:10.1002/anie.201507615
74. Rinkel, J.; Lauterbach, L.; Rabe, P.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2018,** *57,* 3238–3241. doi:10.1002/anie.201800385
75. Rabe, P.; Rinkel, J.; Dolja, E.; Schmitz, T.; Nubbemeyer, B.; Luu, T. H.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2017,** *56,* 2776–2779. doi:10.1002/anie.201612439
76. Dickschat, J. S.; Pahirulzaman, K. A. K.; Rabe, P.; Klapschinski, T. A. *ChemBioChem* **2014,** *15,* 810–814. doi:10.1002/cbic.201300763
77. Wong, J.; Rios-Solis, L.; Keasling, J. D. Microbial Production of Isoprenoids. In *Consequences of Microbial Interactions with Hydrocarbons, Oils, and Lipids: Production of Fuels and Chemicals;* Lee, S. Y., Ed.; Springer: Cham, 2017; pp 1–24. doi:10.1007/978-3-319-31421-1_219-1
78. Martin, V. J. J.; Pitera, D. J.; Withers, S. T.; Newman, J. D.; Keasling, J. D. *Nat. Biotechnol.* **2003,** *21,* 796–802. doi:10.1038/nbt833
79. Anthony, J. R.; Anthony, L. C.; Nowroozi, F.; Kwon, G.; Newman, J. D.; Keasling, J. D. *Metab. Eng.* **2009,** *11,* 13–19. doi:10.1016/j.ymben.2008.07.007
80. Ajikumar, P. K.; Xiao, W.-H.; Tyo, K. E. J.; Wang, Y.; Simeon, F.; Leonard, E.; Mucha, O.; Phon, T. H.; Pfeifer, B.; Stephanopoulos, G. *Science* **2010,** *330,* 70–74. doi:10.1126/science.1191652
81. Dürr, C.; Schnell, H.-J.; Luzhetskyy, A.; Murillo, R.; Weber, M.; Welzel, K.; Vente, A.; Bechthold, A. *Chem. Biol.* **2006,** *13,* 365–377. doi:10.1016/j.chembiol.2006.01.011
82. Smanski, M. J.; Peterson, R. M.; Rajski, S. R.; Shen, B. *Antimicrob. Agents Chemother.* **2009,** *53,* 1299–1304. doi:10.1128/aac.01358-08
83. Ortiz de Montellano, P. R. Substrate Oxidation by Cytochrome P450 Enzymes. In *Cytochrome P450 - Structure, Mechanism, and Biochemistry,* 4th ed.; Ortiz de Montellano, P. R., Ed.; Springer: Cham, 2015; Vol. 1, pp 111–176. doi:10.1007/978-3-319-12108-6_4
84. Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I. *Chem. Rev.* **2005,** *105,* 2253–2278. doi:10.1021/cr0307143
85. Hannemann, F.; Bichet, A.; Ewen, K. M.; Bernhardt, R. *Biochim. Biophys. Acta, Gen. Subj.* **2007,** *1770,* 330–344. doi:10.1016/j.bbagen.2006.07.017
86. Atkinson, J. T.; Campbell, I.; Bennett, G. N.; Silberg, J. J. *Biochemistry* **2016,** *55,* 7047–7064. doi:10.1021/acs.biochem.6b00831
87. Ortega Ugalde, S.; de Koning, C. P.; Wallraven, K.; Bruyneel, B.; Vermeulen, N. P. E.; Grossmann, T. N.; Bitter, W.; Commandeur, J. N. M.; Vos, J. C. *Appl. Microbiol. Biotechnol.* **2018,** *102,* 9231–9242. doi:10.1007/s00253-018-9299-4
88. Rudolf, J. D.; Chang, C.-Y.; Ma, M.; Shen, B. *Nat. Prod. Rep.* **2017,** *34,* 1141–1172. doi:10.1039/c7np00034k
89. Kim, S.-Y.; Zhao, P.; Igarashi, M.; Sawa, R.; Tomita, T.; Nishiyama, M.; Kuzuyama, T. *Chem. Biol.* **2009,** *16,* 736–743. doi:10.1016/j.chembiol.2009.06.007
90. Rudolf, J. D.; Dong, L.-B.; Manoogian, K.; Shen, B. *J. Am. Chem. Soc.* **2016,** *138,* 16711–16721. doi:10.1021/jacs.6b09818
91. Duan, L.; Jogl, G.; Cane, D. E. *J. Am. Chem. Soc.* **2016,** *138,* 12678–12689. doi:10.1021/jacs.6b08610
92. Huo, L.; Hug, J. J.; Fu, C.; Bian, X.; Zhang, Y.; Müller, R. *Nat. Prod. Rep.* **2019,** *36,* 1412–1436. doi:10.1039/c8np00091c
93. Rutledge, P. J.; Challis, G. L. *Nat. Rev. Microbiol.* **2015,** *13,* 509–523. doi:10.1038/nrmicro3496
94. Baltz, R. H. *J. Ind. Microbiol. Biotechnol.* **2010,** *37,* 759–772. doi:10.1007/s10295-010-0730-9
95. Gomez-Escribano, J. P.; Bibb, M. J. *Microb. Biotechnol.* **2011,** *4,* 207–215. doi:10.1111/j.1751-7915.2010.00219.x
96. Khalid, A.; Takagi, H.; Panthee, S.; Muroi, M.; Chappell, J.; Osada, H.; Takahashi, S. *ACS Synth. Biol.* **2017,** *6,* 2339–2349. doi:10.1021/acssynbio.7b00249
97. Myronovskyi, M.; Rosenkränzer, B.; Nadmid, S.; Pujic, P.; Normand, P.; Luzhetskyy, A. *Metab. Eng.* **2018,** *49,* 316–324. doi:10.1016/j.ymben.2018.09.004
98. Takamatsu, S.; Lin, X.; Nara, A.; Komatsu, M.; Cane, D. E.; Ikeda, H. *Microb. Biotechnol.* **2011,** *4,* 184–191. doi:10.1111/j.1751-7915.2010.00209.x
99. Brill, Z. G.; Condakes, M. L.; Ting, C. P.; Maimone, T. J. *Chem. Rev.* **2017,** *117,* 11753–11795. doi:10.1021/acs.chemrev.6b00834
100. Razzak, M.; De Brabander, J. K. *Nat. Chem. Biol.* **2011,** *7,* 865–875. doi:10.1038/nchembio.709
101. Davies, H. M. L. *Nature* **2009,** *459,* 786–787. doi:10.1038/459786a
102. Zhou, Q.; Chen, X.; Ma, D. *Angew. Chem., Int. Ed.* **2010,** *49,* 3513–3516. doi:10.1002/anie.201000888
103. Molawi, K.; Delpont, N.; Echavarren, A. M. *Angew. Chem., Int. Ed.* **2010,** *49,* 3517–3519. doi:10.1002/anie.201000890
104. Major, D. T. *Nat. Catal.* **2018,** *1,* 567–568. doi:10.1038/s41929-018-0130-5
105. Davies, H. M. L.; Manning, J. R. *Nature* **2008,** *451,* 417–424. doi:10.1038/nature06485
106. Mello, R.; Fiorentino, M.; Fusco, C.; Curci, R. *J. Am. Chem. Soc.* **1989,** *111,* 6749–6757. doi:10.1021/ja00199a039
107. Costas, M.; Mehn, M. P.; Jensen, M. P.; Que, L. *Chem. Rev.* **2004,** *104,* 939–986. doi:10.1021/cr020628n
108. Groves, J. T.; Bonchio, M.; Carofiglio, T.; Shalyaev, K. *J. Am. Chem. Soc.* **1996,** *118,* 8961–8962. doi:10.1021/ja9542092
109. Brodsky, B. H.; Du Bois, J. *J. Am. Chem. Soc.* **2005,** *127,* 15391–15393. doi:10.1021/ja055549i
110. Yang, J.; Gabriele, B.; Belvedere, S.; Huang, Y.; Breslow, R. *J. Org. Chem.* **2002,** *67,* 5057–5067. doi:10.1021/jo020174u
111. Chen, M. S.; White, M. C. *Science* **2007,** *318,* 783–787. doi:10.1126/science.1148597
112. Ishihara, Y.; Baran, P. S. *Synlett* **2010,** 1733–1745. doi:10.1055/s-0030-1258123
113. Willot, M.; Christmann, M. *Nat. Chem.* **2010,** *2,* 519–520. doi:10.1038/nchem.715
114. Meguro, A.; Tomita, T.; Nishiyama, M.; Kuzuyama, T. *ChemBioChem* **2013,** *14,* 316–321. doi:10.1002/cbic.201200651
115. Rabe, P.; Pahirulzaman, K. A. K.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2015,** *54,* 6041–6045. doi:10.1002/anie.201501119
116. Schifrin, A.; Khatri, Y.; Kirsch, P.; Thiel, V.; Schulz, S.; Bernhardt, R. *Org. Biomol. Chem.* **2016,** *14,* 3385–3393. doi:10.1039/c6ob00130k
117. Lauterbach, L.; Rinkel, J.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2018,** *57,* 8280–8283. doi:10.1002/anie.201803800
118. Rinkel, J.; Lauterbach, L.; Dickschat, J. S. *Angew. Chem., Int. Ed.* **2017,** *56,* 16385–16389. doi:10.1002/anie.201711142
119. Aaron, J. A.; Lin, X.; Cane, D. E.; Christianson, D. W. *Biochemistry* **2010,** *49,* 1787–1797. doi:10.1021/bi902088z
120. Li, R.; Chou, W. K. W.; Himmelberger, J. A.; Litwin, K. M.; Harris, G. G.; Cane, D. E.; Christianson, D. W. *Biochemistry* **2014,** *53,* 1155–1168. doi:10.1021/bi401643u

121. Blank, P. N.; Barrow, G. H.; Chou, W. K. W.; Duan, L.; Cane, D. E.; Christianson, D. W. *Biochemistry* **2017,** *56,* 5798–5811. doi:10.1021/acs.biochem.7b00895
122. Tomita, T.; Kim, S.-Y.; Teramoto, K.; Meguro, A.; Ozaki, T.; Yoshida, A.; Motoyoshi, Y.; Mori, N.; Ishigami, K.; Watanabe, H.; Nishiyama, M.; Kuzuyama, T. *ACS Chem. Biol.* **2017,** *12,* 1621–1628. doi:10.1021/acschembio.7b00154
123. Görner, C.; Häuslein, I.; Schrepfer, P.; Eisenreich, W.; Brück, T. *ChemCatChem* **2013,** *5,* 3289–3298. doi:10.1002/cctc.201300285
124. Peters, R. J. *Nat. Prod. Rep.* **2010,** *27,* 1521–1530. doi:10.1039/c0np00019a
125. Jia, M.; Potter, K. C.; Peters, R. J. *Metab. Eng.* **2016,** *37,* 24–34. doi:10.1016/j.ymben.2016.04.001
126. Andersen-Ranberg, J.; Kongstad, K. T.; Nielsen, M. T.; Jensen, N. B.; Pateraki, I.; Bach, S. S.; Hamberger, B.; Zerbe, P.; Staerk, D.; Bohlmann, J.; Møller, B. L.; Hamberger, B. *Angew. Chem., Int. Ed.* **2016,** *55,* 2142–2146. doi:10.1002/anie.201510650
127. Jia, M.; Mishra, S. K.; Tufts, S.; Jernigan, R. L.; Peters, R. J. *Metab. Eng.* **2019,** *55,* 44–58. doi:10.1016/j.ymben.2019.06.008
128. Mafu, S.; Jia, M.; Zi, J.; Morrone, D.; Wu, Y.; Xu, M.; Hillwig, M. L.; Peters, R. J. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113,* 2526–2531. doi:10.1073/pnas.1512096113
129. Hernandez-Ortega, A.; Vinaixa, M.; Zebec, Z.; Takano, E.; Scrutton, N. S. *Sci. Rep.* **2018,** *8,* No. 14396. doi:10.1038/s41598-018-32816-1
130. Jung, S. T.; Lauchli, R.; Arnold, F. H. *Curr. Opin. Biotechnol.* **2011,** *22,* 809–817. doi:10.1016/j.copbio.2011.02.008
131. Dietrich, J. A.; Yoshikuni, Y.; Fisher, K. J.; Woolard, F. X.; Ockey, D.; McPhee, D. J.; Renninger, N. S.; Chang, M. C. Y.; Baker, D.; Keasling, J. D. *ACS Chem. Biol.* **2009,** *4,* 261–267. doi:10.1021/cb900006h
132. Seifert, A.; Vomund, S.; Grohmann, K.; Kriening, S.; Urlacher, V. B.; Laschat, S.; Pleiss, J. *ChemBioChem* **2009,** *10,* 853–861. doi:10.1002/cbic.200800799
133. Paramasivan, K.; Mutturi, S. *Crit. Rev. Biotechnol.* **2017,** *37,* 974–989. doi:10.1080/07388551.2017.1299679
134. Harada, H.; Yu, F.; Okamoto, S.; Kuzuyama, T.; Utsumi, R.; Misawa, N. *Appl. Microbiol. Biotechnol.* **2009,** *81,* 915–925. doi:10.1007/s00253-008-1724-7
135. Westfall, P. J.; Pitera, D. J.; Lenihan, J. R.; Eng, D.; Woolard, F. X.; Regentin, R.; Horning, T.; Tsuruta, H.; Melis, D. J.; Owens, A.; Fickes, S.; Diola, D.; Benjamin, K. R.; Keasling, J. D.; Leavell, M. D.; McPhee, D. J.; Renninger, N. S.; Newman, J. D.; Paddon, C. J. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109,* E111–E118. doi:10.1073/pnas.1110740109
136. Alonso-Gutierrez, J.; Chan, R.; Batth, T. S.; Adams, P. D.; Keasling, J. D.; Petzold, C. J.; Lee, T. S. *Metab. Eng.* **2013,** *19,* 33–41. doi:10.1016/j.ymben.2013.05.004
137. Leonard, E.; Ajikumar, P. K.; Thayer, K.; Xiao, W.-H.; Mo, J. D.; Tidor, B.; Stephanopoulos, G.; Prather, K. L. J. *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107,* 13654–13659. doi:10.1073/pnas.1006138107
138. Tashiro, M.; Kiyota, H.; Kawai-Noma, S.; Saito, K.; Ikeuchi, M.; Iijima, Y.; Umeno, D. *ACS Synth. Biol.* **2016,** *5,* 1011–1020. doi:10.1021/acssynbio.6b00140
139. Zhu, F.; Zhong, X.; Hu, M.; Lu, L.; Deng, Z.; Liu, T. *Biotechnol. Bioeng.* **2014,** *111,* 1396–1405. doi:10.1002/bit.25198
140. Alonso-Gutierrez, J.; Kim, E.-M.; Batth, T. S.; Cho, N.; Hu, Q.; Chan, L. J. G.; Petzold, C. J.; Hillson, N. J.; Adams, P. D.; Keasling, J. D.; Garcia Martin, H.; Lee, T. S. *Metab. Eng.* **2015,** *28,* 123–133. doi:10.1016/j.ymben.2014.11.011
141. Biggs, B. W.; Lim, C. G.; Sagliani, K.; Shankar, S.; Stephanopoulos, G.; De Mey, M.; Ajikumar, P. K. *Proc. Natl. Acad. Sci. U. S. A.* **2016,** *113,* 3209–3214. doi:10.1073/pnas.1515826113
142. Meyer, A. J.; Segall-Shapiro, T. H.; Glassey, E.; Zhang, J.; Voigt, C. A. *Nat. Chem. Biol.* **2019,** *15,* 196–204. doi:10.1038/s41589-018-0168-3
143. Dahl, R. H.; Zhang, F.; Alonso-Gutierrez, J.; Baidoo, E.; Batth, T. S.; Redding-Johanson, A. M.; Petzold, C. J.; Mukhopadhyay, A.; Lee, T. S.; Adams, P. D.; Keasling, J. D. *Nat. Biotechnol.* **2013,** *31,* 1039–1046. doi:10.1038/nbt.2689
144. Korman, T. P.; Opgenorth, P. H.; Bowie, J. U. *Nat. Commun.* **2017,** *8,* No. 15526. doi:10.1038/ncomms15526
145. Kirby, J.; Nishimoto, M.; Chow, R. W. N.; Baidoo, E. E. K.; Wang, G.; Martin, J.; Schackwitz, W.; Chan, R.; Fortman, J. L.; Keasling, J. D. *Appl. Environ. Microbiol.* **2015,** *81,* 130–138. doi:10.1128/aem.02920-14
146. Chatzivasileiou, A. O.; Ward, V.; Edgar, S. M.; Stephanopoulos, G. *Proc. Natl. Acad. Sci. U. S. A.* **2019,** *116,* 506–511. doi:10.1073/pnas.1812935116
147. Ye, L.; Zhang, C.; Bi, C.; Li, Q.; Zhang, X. *Microb. Cell Fact.* **2016,** *15,* No. 202. doi:10.1186/s12934-016-0607-3
148. Alonso-Gutierrez, J.; Koma, D.; Hu, Q.; Yang, Y.; Chan, L. J. G.; Petzold, C. J.; Adams, P. D.; Vickers, C. E.; Nielsen, L. K.; Keasling, J. D.; Lee, T. S. *Biotechnol. Bioeng.* **2018,** *115,* 1000–1013. doi:10.1002/bit.26530
149. Lin, G.-M.; Warden-Rothman, R.; Voigt, C. A. *Curr. Opin. Syst. Biol.* **2019,** *14,* 82–107. doi:10.1016/j.coisb.2019.04.004
150. Pateraki, I.; Heskes, A. M.; Hamberger, B. Cytochromes P450 for Terpene Functionalisation and Metabolic Engineering. In *Biotechnology of Isoprenoids;* Schrader, J.; Bohlmann, J., Eds.; Springer: Cham, 2015; pp 107–139. doi:10.1007/10_2014_301
151. Chang, M. C. Y.; Eachus, R. A.; Trieu, W.; Ro, D.-K.; Keasling, J. D. *Nat. Chem. Biol.* **2007,** *3,* 274–277. doi:10.1038/nchembio875
152. Gricman, Ł.; Weissenborn, M. J.; Hoffmann, S. M.; Borlinghaus, N.; Hauer, B.; Pleiss, J. *ChemistrySelect* **2016,** *1,* 1243–1251. doi:10.1002/slct.201600369
153. Kosuri, S.; Church, G. M. *Nat. Methods* **2014,** *11,* 499–507. doi:10.1038/nmeth.2918
154. Casini, A.; Storch, M.; Baldwin, G. S.; Ellis, T. *Nat. Rev. Mol. Cell Biol.* **2015,** *16,* 568–576. doi:10.1038/nrm4014
155. Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A., III; Smith, H. O. *Nat. Methods* **2009,** *6,* 343–345. doi:10.1038/nmeth.1318
156. Engler, C.; Gruetzner, R.; Kandzia, R.; Marillonnet, S. *PLoS One* **2009,** *4,* No. e5553. doi:10.1371/journal.pone.0005553
157. Luo, X.; Reiter, M. A.; d'Espaux, L.; Wong, J.; Denby, C. M.; Lechner, A.; Zhang, Y.; Grzybowski, A. T.; Harth, S.; Lin, W.; Lee, H.; Yu, C.; Shin, J.; Deng, K.; Benites, V. T.; Wang, G.; Baidoo, E. E. K.; Chen, Y.; Dev, I.; Petzold, C. J.; Keasling, J. D. *Nature* **2019,** *567,* 123–126. doi:10.1038/s41586-019-0978-9