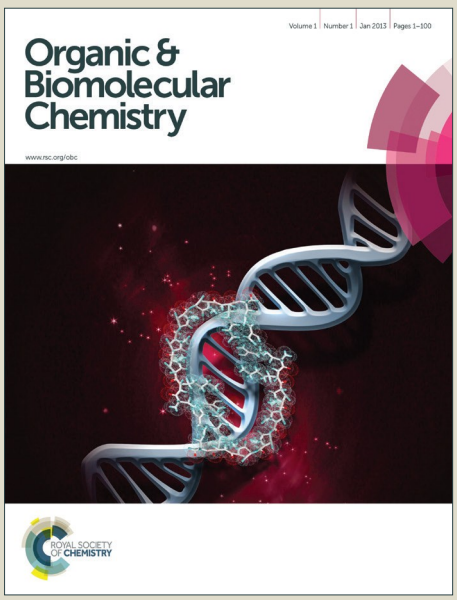
Organic & Biomolecular Chemistry

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: A. Schifrin, Y. Khatri, P. Kirsch, V. Thiel, S. Schulz and R. Bernhardt, *Org. Biomol. Chem.*, 2016, DOI: 10.1039/C6OB00130K.



ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/obc

View Article Online
DOI: 10.1039/C6OB00130K

Organic&Biomolecular Chemistry

PAPER

A Single Terpene Synthase is Responsible for a Wide Variety of Sesquiterpenes in *Sorangium cellulosum* Soce56

Received 00th January 20xx,
Accepted 00th January 20xx

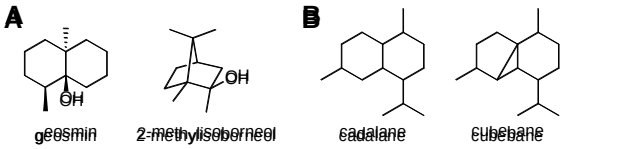
DOI: 10.1039/x0xx00000x
www.rsc.org/

Alexander Schifrin,[a] Yogan Khatri,[a] Philine Kirsch,[a] Verena Thiel,[b] Stefan Schulz,[b] and Rita Bernhardt*[a]

The myxobacterium *Sorangium cellulosum* So ce56 is a prolific producer of volatile sesquiterpenes. The strain harbours one of the largest prokaryotic genomes (13.1 Mbp). However, it codes only for three type I terpene synthases (TSs; sce1440, sce6369, sce8552) and one type II TS (sce4636), responsible for the production of at least 17 sesquiterpenes. We report here the gene expression and biosynthesis of the TS products in *E. coli.* Comparison with the So ce56 volatiles allows the assignment of the terpenes to their synthesizing genes. Both, the geosmin synthase sce1440 and the previously examined (+)-eremophilene synthase sce8552 are highly specific. In contrast, Sce6369, the first characterized 10-*epi*-cubebol synthase, is responsible for the formation of most of the So ce56 sesquiterpenes, mainly cadalanes and cubebanes. Sce4636 does not convert FPP. Having characterized the So ce56 TSs, we screened all the 27 sequenced myxobacterial species from the NCBI and JGI-IMG databases for parent genes to predict the sesquiterpenes produced by them.

Introduction

Terpene synthase (TS)-mediated conversions of farnesyl pyrophosphate (FPP) to sesquiterpenes, the C-15 subgroup of terpenes, belong to the most complex biochemical reactions observed in nature. These reactions can involve up to four C-C bond formations, several consecutive cationic intermediates, hydride shifts and Wagner-Meerwein rearrangements. A single conversion can induce up to six stereogenic centers (and one more in case of water quenching as the final reaction step).[1-3] Over 80 sesquiterpene hydrocarbon skeletons are known to date. Highly specific TSs, which convert FPP to a single product, represent interesting targets for studies of nature's incomparable precision in bioreactions.[4, 5] However, in some cases, less specific TS catalysts produce trace amounts of side products, which bear important structural information. This information can be directly used to draw conclusions about the reaction mechanism.[6-9] The accelerating speed of genomic sequencing reveals a steadily increasing number of bacterial TS genes. Although bacteria are a rich source of structurally diverse terpenoids,[10-12] terpene biosynthesis in the subgroup of myxobacteria has been studied in detail only for the off-flavours geosmin and 2-methylisoborneol (Scheme 1 A).[13-16] The genome sequence data of the strain *Sorangium cellulosum* So ce56 revealed one of the largest bacterial genomes with a size of 13.1 Mbp[17] containing three type I TS genes (sce1440, sce6369, sce8552) and one type II TS gene (sce4636).[18] We recently described the gene product of sce8552 as a specific (+)-eremophilene synthase, clustered with the cytochrome P450 CYP264B1 (sce8551), which oxidizes (+)-eremophilene.[16]



Scheme 1. A) Common myxobacterial terpenoids geosmin and 2-methyl-isoborneol; B) Sesquiterpene hydrocarbon scaffolds of cadalane and cubebane

Despite mutualities of sesquiterpene synthases including the common α-helical fold and the characteristic motifs in the active site (aspartate-rich DDxxD/E motif on helix D and NSE/DTE-triad on helix H),[1] amino acid sequence identities among bacterial sesquiterpene synthases as well as between bacterial and other TSs are rather low.[18] In general, the structure of each sesquiterpene is governed by the shape of the TS binding pocket.[1-3] Furthermore, TSs can undergo large conformational changes observed in an induced-fit reaction mechanism.[19] Hence, the biochemical function of TSs and the structures of their corresponding products can only be assigned by direct experiments.[18] The assignment of bacterial TS functions has been conducted via three main approaches:

1.) *In vitro* conversion of FPP with the purified enzyme and analysis of the products.[16, 20, 21]

2.) Direct analysis of bacterial volatiles in combination with a phylogenetic analysis of the corresponding strain.[22]

3.) Expression of the TS gene in a heterologous host and *in vivo* production of terpenes followed by their analysis.[12, 16, 23, 24]

The first approach is the most rigorous one as it allows an unambiguous functional assignment of the gene and the determination of new terpene structures. Often, purification and handling of the sensitive TSs are tedious and time consuming. In

[a] Dipl. Chem A. Schifrin, Dr. Y. Khatri, B. Sc. P. Kirsch, Prof. Dr. R. Bernhardt
Universität des Saarlandes, Institut für Biochemie, Campus B2.2,
66123 Saarbrücken (Germany)
E-mail: ritabern@mx.uni-saarland.de
[b] Dr. V. Thiel, Prof. Dr. S. Schulz
Technische Universität Braunschweig, Institut für Organische Chemie
Hagenring 30,38106 Braunschweig (Germany)
Fax: (+49) 0531-391-5272
E-mail: stefan.schulz@tu-bs.de

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

View Article Online
DOI: 10.1039/C6OB00130K

addition, the production of new terpenes for structural determination requires a sufficient quantity of pure enzyme. The second approach allows an efficient assignment of TS functions. However, it is only applicable for the analysis of TSs, whose

View Article Online
DOI: 10.1039/C6OB00130K

homologs have been functionally described by an individual experiment. In exceptional cases, new functions of TS genes without any known homologue can be suggested. The function can be deduced by process of elimination, if all the other TSs in the same strain have functional homologs and can thus be identified.[22] Structures of new terpenes and functions of the corresponding genes avoiding the isolation and purification of TSs can be determined via the third approach. However, purification of single terpenes from terpene mixtures and a complex whole-cell matrix remains a challenge. In addition, FPP availability and sufficient expression of the TS in the recombinant host have to be guaranteed.

Here, based on our previous studies with the (+)-eremophilene synthase[16] and in analogy to previous studies with *Actinomycetes*[24] we chose a combinatorial approach of the second and the third methods for the analysis of TSs from the myxobacterium *Sorangium cellulosum* So ce56. The overexpression of TSs in *E. coli* allows an efficient production of terpenes for structural elucidation, and the volatile analysis of So ce56 reconfirms the occurrence of the terpenes *in vivo*.

The present work is based on gas chromatography/mass spectrometry (GC/MS) analyses of the volatile sesquiterpenes emitted by *S. cellulosum* So ce56. For a detailed investigation of their biosynthesis, each of its TS genes has been overexpressed in an *E. coli* whole-cell system. Genes coding for mevalonate pathway enzymes (mevalonate kinase, diphosphomevalonate decarboxylase, phosphomevalonate kinase, isopentenylpyrophosphate isomerase) were simultaneously co-expressed to ensure a sufficient *in vivo* FPP supply. The corresponding terpene products have been analyzed by GC/MS and nuclear magnetic resonance (NMR) spectroscopy and compared with the *S. cellulosum* volatiles to suggest reaction mechanisms. The genomes of all 27 myxobacterial species present in the database of the National Center for Biotechnology Information (NCBI, ncbi.nlm.nih.gov) and the Joint Genome Institute Integrated Microbial Genome Database (JGI-IMG, img.jgi.doe.gov) were investigated for orthologs of the So ce56 TS genes. Thus, the occurrence of certain sesquiterpenes in myxobacterial volatile compositions can be predicted.

Results and discussion

Volatiles from *S. cellulosum* So ce56

The volatiles of *S. cellulosum* have been trapped using a closed loop stripping apparatus (CLSA) from liquid cultures. GC/MS analysis of the volatile extract revealed 17 sesquiterpenes with masses of 202, 204, 222 Da (Figure 1A) and the common myxobacterial sesquiterpene degradation product geosmin (182 Da). Most of the sesquiterpenes were identified via comparison of their mass spectra (MS) and gas chromatographic retention indices (*I*) with authentic standards (Table 1). (+)-Eremophilene (**11**), δ-cadinene (**17**), 10-*epi*-cubebol (**19**), germacra-1(10)*E*,5*E*-dien-11-ol (**26**), and 1,10-di-*epi*-cubenol (**28**) were isolated and their structures proven by NMR spectroscopy. 10-*epi*-α-Cubebene (**1**) and 10-*epi*-β-cubebene (**2**) have not been described before. Their mass spectra were similar to those of α- and β-cubebene, but their retention indices were about 40 units higher. A synthetic sample of **19** contained both **1** and **2** as minor water elimination products, proving the proposed structure.[25] Compound **13** could not be structurally determined. The most prominent peaks were **11**, **17** and **19** (Figure 1 A). Germacra-1(10)*E*,5*E*-dien-11-ol (**26**, 222 Da) is also among the most abundant sesquiterpenes. However, it appears only as a minor peak in Figure 1 due to the low abundance of its 204 Da fragment.

The newly described (+)-eremophilene (**11**) is the single product of the TS gene sce8552. Germacra-1(10)*E*,5*E*-dien-11-ol (**26**) appears in diverse bacterial strains together with geosmin. *In vivo* and *in vitro* conversion experiments with geosmin synthases revealed **26** as the typical sesquiterpene alcohol precursor of geosmin.[13, 26-30] (+)-δ-Cadinene (**17**) has been identified as an intermediate of the phytoalexin gossypol in cotton (*Gossypium arboreum*), but it is also present in other plant species.[31, 32] The cotton (+)-δ-cadinene synthase has been crystallized and its reaction mechanism has been extensively studied. [5] The enantiomer (-)-**17** has been isolated from fungi[8, 33] and bacteria.[34] The termites *Reticulitermes flavipes* and *R. virginicus* both produce **17**, however, its absolute configuration has not been determined.[35, 36] 10-*epi*-Cubebol (**19**) has been extracted from different plant species[37, 38] and a stereospecific synthesis has been performed.[25] However, a bacterial source of **19** has not been described previously.

Interestingly, most of the *S. cellulosum* sesquiterpenes, including **17** and **19**, belong to the structural classes of cadalanes and cubebanes (Scheme 1B). Studies of promiscuous TSs have shown that these sesquiterpene classes are not only structurally closely related, but also share common biosynthetic mechanisms.[22, 39]

Terpene synthases from *S. cellulosum* So ce56

Gene mining in *S. cellulosum* So ce56 revealed three putative type I TSs (sce1440, sce6369, sce8552) and one type II TS (sce4636).[18] In a previous study, we have functionally characterized the two-gene cluster consisting of the highly specific (+)-eremophilene synthase (sce8552) and the (+)-eremophilene oxidase CYP264B1 (sce8551).[16] Here, we additionally characterize the three other TSs regarding FPP conversion. Co-expression of the TS genes with mevalonate pathway genes in an *E. coli* based whole-cell catalyst yielded the corresponding sesquiterpenes after feeding mevalonolactone. In addition to GC/MS analyses of the *E. coli* extracts, the main products of each conversion were analyzed by NMR spectroscopy after purification.

In this study, the type II TS sce4636 did not give any product. Typically, this type of TSs convert the larger GGPP to diterpene derivatives, however rarely FPP.[3]

The amino acid sequence alignment of sce1440 shows high similarities with bifunctional germacra-1(10)*E*,5*E*-dien-11-ol/geosmin synthases having twice the length of usual TSs, and containing two sets of each, the characteristic Mg^{2+}-binding motif (DDxxD) and the NSE/DTE triad.[27-30] Sce1440 yielded exclusively **26**, without geosmin. Sequence analysis of sce1440 does not allow drawing conclusions on why half of the sequence seems inactive. The highly sensitive multi-step conversion of FPP to geosmin involves **26.** However the germacra-1(10)*E*,5*E*-dien-11-ol/geosmin synthase activity and thus the observed intermediates and final product depend on various conditions like the pH and the nature of the divalent metal cation cofactor.[27-30] Therefore, we concluded that sce1440 acts as a germacra-1(10)*E*,5*E*-dien-11-ol/geosmin synthase.

Feeding mevalonolactone to a sce6369 expressing *E. coli* strain resulted in a mixture of 21 sesquiterpene products (Figure 1B). Eight of them – mostly sesquiterpene alcohols – do not match any compound from mass spectral and retention index databases (Table 1, **9, 10, 21-25, 27**). The thirteen structurally identified sesquiterpenes have all cadalane- and cubebane-type scaffolds with the exception of germacrene D (**8**) (Table 1 and Scheme 1B, **1**, **2, 3, 6, 8, 12, 15, 17, 19, 28-31**). The main product was identified as 10-*epi*-cubebol (**19**), which is also one of the major sesquiterpenoids in the *S. cellulosum* extract. Altogether, seven of the substances have been detected in both, the *S. cellulosum* whole cell extract and the sce6369 expressing *E. coli* culture (**1, 2, 3, 6, 8, 17,** and **19**).

Reaction Mechanism of Cadalane- and Cubebane-Synthesis

At least eight out of 17 sesquiterpenes from *S. cellulosum* share common cadalane- and cubebane-skeletons. These scaffolds are closely related to each other (Scheme 1B). Indeed, sce6369 is a promiscuous enzyme, which produces 21 sesquiterpenes. The

View Article Online
DOI: 10.1039/C6OB00130K

main products include the cubebanes 10-*epi*-cubebol (Figure 1 B, **19**) and 1,10-di-*epi*-cubebol (**28**) as well as the cadalanes δ-cadinene (**17**), T-cadinol (**29**) and α-cadinol (**31**). However, only seven sesquiterpenes (**1**, **2**, **3**, **6**, **8**, **17**, and **19**) were present in both, the *S. cellulosum* So ce56 extract and the sce6369 expressing *E. coli* culture. Sce6369 is likely to produce different sets of sesquiterpenes when overexpressed in *E. coli* or in its native host *S. cellulosum* So ce56. TS-mediated reactions have shown to be highly sensitive and the selectivities of TSs strongly dependent on the plasticities of their active sites.[40, 41] Furthermore, enzymatic derivatizations of the sesquiterpenes, for instance by one of the 21 cytochromes P450 in So ce56,[16, 42-48] could lead to a change of volatility and, therefore, to a different sesquiterpene pattern in different host organisms.

The early steps of FPP cyclization during the formation of cadalane and cubebane quiterpenes have been intensely investigated by theoretical calculations[49] and experimental data.[4, 20, 50] The promiscuous reactions show the involvement of common cationic intermediates. We thus could propose a comprehensive reaction mechanism, which explains the occurrence of all observed cadalanes and cubebanes in the volatile fraction of *S. cellulosum* So ce56 (Scheme 2, black, red and blue), in the sce6369 expressing in *E. coli* strain (Scheme 2, black and green), and, in addition, in the previously reported sce6369 expressing *S. avermitilis* strain (Scheme 2, pink and red).[12] A reaction mechanism including enantiomers of the terpene structures has been suggested and included as supporting information (Supporting Information, Scheme 1).

FPP isomerizes to NPP to enable the 1,10 cyclization to the germacradienyl cation (Scheme 2, **A**). The following 1,3-hydride shift to yield the allylic cation **B** is energetically favoured. **B** can directly undergo cyclopropanation to bicyclogermacrene (**14**) or terminal proton loss to germacrene D (**8**). An intramolecular nucleophilic attack of the 1,10 double bond in **B** continues the reaction cascade. The C-C bond formation leads either to a cadinane-type (**C**) or a muurolane-type (**D**) tertiary carbocation. The trans-decalin intermediate **C** is a branching point for the formation of the diastereomers T-cadinol (**29**) and α-cadinol (**31**) by water quenching as well as the formation of γ-cadinene (**16**) and α-cadinene (**20**) by hydrogen abstraction. The latter one can isomerize to β-cadinene (Scheme 2, pink) by de- and re-protonation. The cis-decalin intermediate **D** is deprotonated to δ-cadinene (**17**). Hydride shifts of **C** and **D** lead to tertiary allylic cations (**E** and **H**) or tertiary carbocations (**F** and **G**), which are the key intermediates for the synthesis of cubebanes. Cation **E** can be deprotonated to (+)-10-*epi*-zonarene (**12**). Deprotonation of **H** gives bicyclosesquiphellandrene (Scheme 2, pink) or cadina-3,5-diene (**3**). The latter one can aromatize to *trans*-calamenene (**18**). Water quenching of **G** at C-4 gives the fully saturated cubebol (**15**). Water attack at C-1 leads to cubenol (**30**). The 1,2-hydride shift turning carbocation **C** into **F** involves the epimerization of the methyl group at C-10. This epimerization paves the way to the products 10-*epi*-α-cubebene (**1**), 10-*epi*-β-cubebene (**2**), 7βH,10βH-cadina-1(6),4-diene (**5**), *cis*-muurola-4(15)5-diene (**6**), through proton abstraction and to the major products 10-*epi*-cubebol (**19**) and 1,10-di-*epi*-cubenol (**28**) through nucleophilic attack of water. The conversion of **F** to the final products **1**, **2**, **5**, **6**, **19**, and **28**, whether it is direct or over the intermediates **I** or **J**, proceeds always under retention of the stereochemistry at C-7 and C-10.

Screening for Terpene Synthase Orthologs in Myxobacterial Genomes

Since myxobacteria possess a complex physiology and are considered as potential sources of novel natural products,[51, 52] the inherited TS genes could have a multitude of influences on the biosynthesis of natural terpenoids. So far, only few studies on myxobacterial TSs and their products have been performed. Therefore, after the functional characterization of the *S. cellulosum* TSs ce1440, sce6369, and sce8552, we searched for homologue TSs in all 27 sequenced myxobacterial species available in the NCBI and the JGI-IMG genome databases.

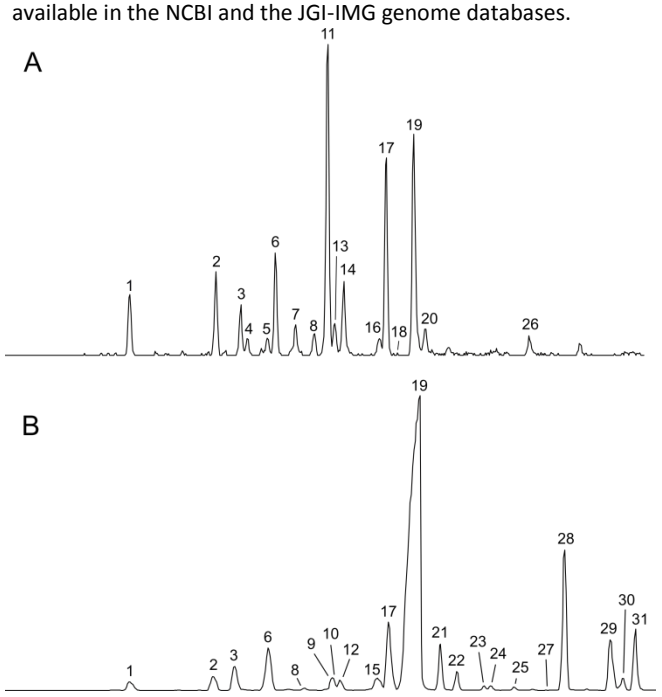


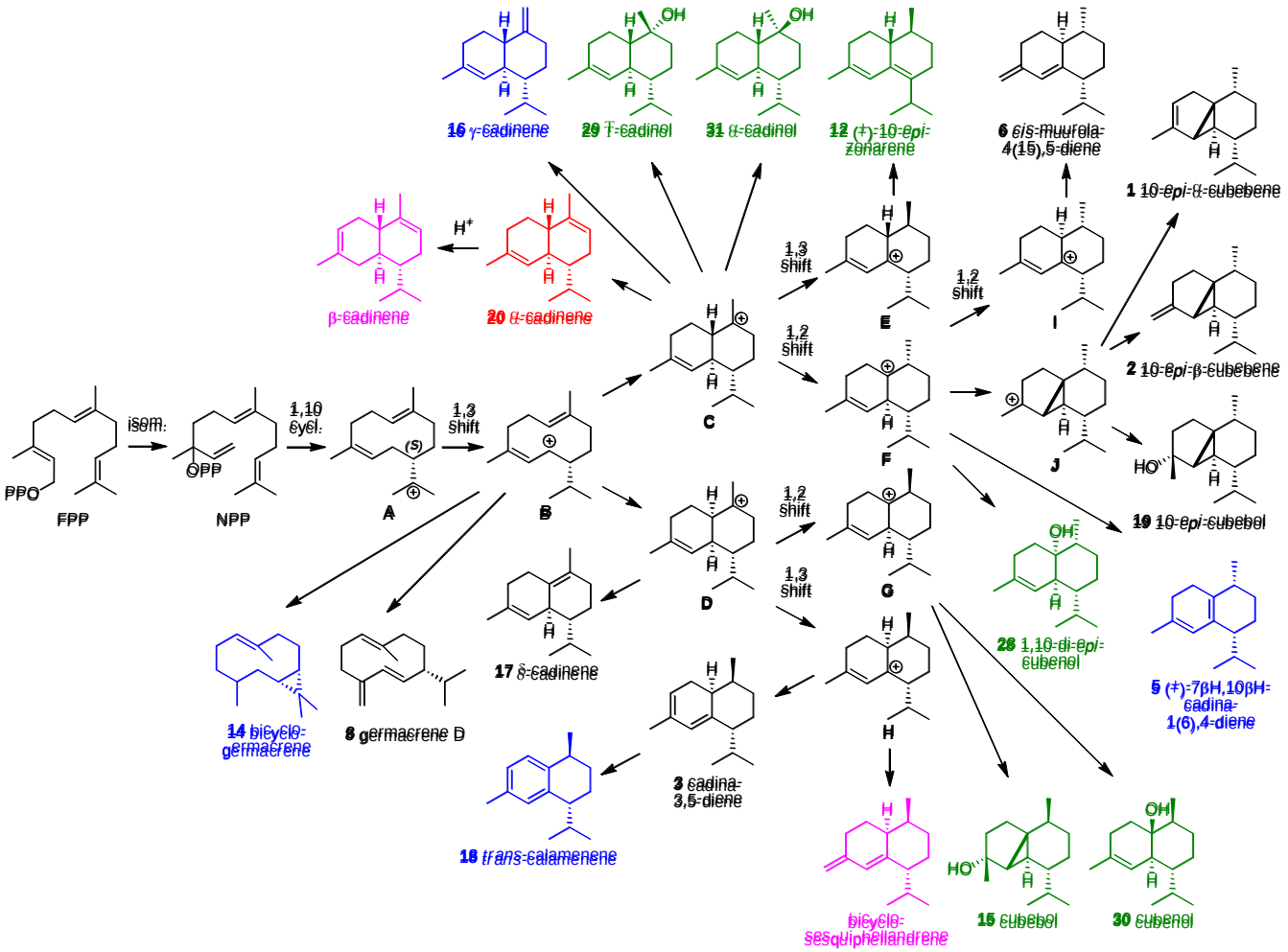
Figure 1. Selected ion chromatogram (SIM) of sesquiterpenes with *m/z* 204.A) The sesquiterpenes in the extract of volatiles of *S. cellulosum* So ce56; B) The sesquiterpenes produced by sce6369 in *E. coli* whole-cell cultures.

We observed that the germacra-1(10)*E*,5*E*-dien-11-ol/geosmin synthases (sce1440) are widespread among myxobacteria and thus, orthologs with over 50% sequence identitiy and over 65% similarity have been identified in 15 out of the 27 myxobacterial species (Supporting Information, Table 1 and Figure 1). Eight gene sequences related to the type II TS sce4636 were found in only four myxobacteria (Supporting Information, Table 2 and Figure 2). Since sce4636 does not convert FPP, its homologs are not likely to contribute to the production of sesquiterpenes in myxobacteria. The 10-*epi*-cubebol synthase (sce6369) has seven homologous genes with sequence identities between 24% and 34% (similarities from 38% to 47%) within myxobacterial genomes (Supporting Information, Table 3 and Figure 3). The (+)-eremophilene synthase (sce8552) shares 68.9% identity (81.2% similarity) with a gene from the related *S. cellulosum* strain So0157-2. Sequence identities with the other twelve related genes range in between 22% and 30% (similarities between 37% and 47%) (Supporting Information, Table 4 and Figure 4).

Functional orthologs of the 10-*epi*-cubebol synthase and (+)-eremophilene synthase from other strains have not been described yet. However, various TSs producing cubebane-and cadalane-type compounds have been characterized in actinomycetes.[53] The genes include *epi*-cubebol synthase (*Streptosporangium roseum* DSM 43021)[54] *epi*-cubenol synthase (*Streptomyces griseus*),[4, 55] δ-cadinene synthase (*Streptomyces clavuligerus*),[34] γ-cadinene synthase (*Chitinophaga pinensis* DSM 2588),[24] T-muurolol synthases (*S. clavuligerus*, *Roseiflexus castenholzii* DSM 13941).[12, 24, 34] In addition, studies of the myxobacterium *Chondromyces crocatus* showed the presence of (+)-eremophilene (**11**), cubenol (**30**) and other cadalanes (cadina-1,4-diene, α-muurolene, zonarene) among the volatiles.[56] Its genes related to 10-*epi*-cubebol synthase (WP_050430829) and (+)-eremophilene synthase (WP_012241161) are considered to be highly promising candidates for the synthesis of the corresponding terpenes. Taken together, this data provides the availability of

View Article Online
DOI: 10.1039/C6OB00130K

sets of different TSs which can be used for biotechnological purposes. However, experimental analyses of these TSs and their products are necessary to assign the gene functions.



Scheme 2. Proposed mechanism for the biosynthesis of cadalane- and cubebane-type sesquiterpenes by sce6369; sesquiterpenes observed in *S. cellulosum* and the sce6369 expressing *E. coli* strain are represented in black; sesquiterpenes observed only in *S. cellulosum* are represented in blue; sesquiterpenes observed only in the sce6369 expressing *E. coli* strain are represented in green; sesquiterpenes observed in *S. cellulosum* and the sce6369 expressing *Streptomyces avermitilis* SUKA22[12] are represented in red; sesquiterpenes observed only in the sce6369 expressing *S. avermitilis* SUKA22[12] are represented in pink; cadina-3,5-diene (**3**) was observed in *S. cellulosum* as well as both *E. coli* and *S. avermitilis* SUKA22 expressing sce6369[12]

Table 1. Sesquiterpenes identified in an extract of volatiles of *S. cellulosum* So ce56 and the corresponding terpene synthase (TS) genes. ST: sesquiterpene, *I*: retention index, MS: mass spectrum, NMR: nmr spectrum, n.d.: no corresponding TS gene determined

	Sesquiterpenoid	*I*	Mass, Empirical formula	Identification	So ce56 Extract	TS gene
1	10-*epi*-α-cubebene	1395	204, $C_{15}H_{24}$	MS, *I*	x	sce6369
2	10-*epi*-β-cubebene	1440	204, $C_{15}H_{24}$	MS, *I*	x	sce6369
3	cadina-3,5-diene	1453	204, $C_{15}H_{24}$	MS, *I*	x	sce6369[a]
4	(*E*)-β-farnesene	1456	204, $C_{15}H_{24}$	MS, *I*	x	n. d.
5	7βH,10βH-cadina-1(6),4-diene	1460	204, $C_{15}H_{24}$	MS, *I*	x	sce6369[b]
6	*cis*-muurola-4(15)5-diene	1469	204, $C_{15}H_{24}$	MS, *I*	x	sce6369
7	γ-gurjunene	1480	204, $C_{15}H_{24}$	MS, *I*	x	n. d.

View Article Online
DOI: 10.1039/C6OB00130K

8	germacrene D	1489	204, $C_{15}H_{24}$	MS, *I*	x	sce6369
9	ST	1493	204, $C_{15}H_{24}$	MS		sce6369
10	ST alcohol	1495	222, $C_{15}H_{26}O$	MS		sce6369
11	(+)-eremophilene	1496	204, $C_{15}H_{24}$	MS, *I*, NMR[16],polarimetry	x	sce8552[c]
12	10-*epi*-zonarene	1499	204, $C_{15}H_{24}$	MS, *I*		sce6369
13	ST	1499	204, $C_{15}H_{24}$	MS	x	n. d.
14	bicyclogermacrene	1503	204, $C_{15}H_{24}$	MS, *I*	x	sce6369[b]
15	cubebol	1517	222, $C_{15}H_{26}O$	MS, *I*		sce6369
16	γ-cadinene	1522	204, $C_{15}H_{24}$	MS, *I*	x	sce6369[b]
17	δ-cadinene	1524	204, $C_{15}H_{24}$	MS, *I*, NMR	x	sce6369
18	*trans*-calamenene	1530	202, $C_{15}H_{24}$	MS, *I*	x	sce6369[b]
19	10-*epi*-cubebol	1537	222, $C_{15}H_{26}O$	MS, *I*, NMR	x	sce6369
20	α-cadinene	1543	204, $C_{15}H_{24}$	MS, *I*	x	sce6369[d]
21	ST alcohol	1549	222, $C_{15}H_{26}O$	MS		sce6369
22	ST alcohol	1558	222, $C_{15}H_{26}O$	MS		sce6369
23	ST alcohol	1573	222, $C_{15}H_{26}O$	MS		sce6369
24	ST alcohol	1576	222, $C_{15}H_{26}O$	MS		sce6369
25	ST alcohol	1590	222, $C_{15}H_{26}O$	MS		sce6369
26	germacra-1(10)*E*,5*E*-dien-11-ol	1594	222, $C_{15}H_{26}O$	MS, *I*, NMR	x	sce1440
27	ST alcohol	1609	222, $C_{15}H_{26}O$	MS		sce6369
28	1,10-di-*epi*-cubenol	1616	222, $C_{15}H_{26}O$	MS, *I*, NMR		sce6369
29	T-cadinol	1642	222, $C_{15}H_{26}O$	MS, *I*		sce6369
30	cubenol	1649	222, $C_{15}H_{26}O$	MS, *I*		sce6369
31	α-cadinol	1655	222, $C_{15}H_{26}O$	MS, *I*		sce6369

[a] reported here and in Ref. [12]
[b] these cadinanes and cubebanes have not been found in the sce6369 expressing *E. Coli* extract, however, their occurrence can be reasonably deduced from the reaction mechanism
[c] reported in Ref. [16] and in Ref. [12]
[d] reported only in Ref. [12]

ROYAL SOCIETY OF CHEMISTRY View Article Online
DOI: 10.1039/C6OB00130K

Organic&Biomolecular Chemistry

PAPER

Conclusions

We used a combined approach to assign the sesquiterpenes present in *Sorangium cellulosum* So ce56 to their corresponding genes: Besides the direct analysis of the myxobacterial volatiles, each of the four TSs (sce1440, sce4636, sce6369, and sce8552) was individually overexpressed in *E. coli*. Simultaneous co-expression of the mevalonate pathway genes provided FPP as a TS substrate *in vivo*.

Out of the 17 sesquiterpenes identified among the volatiles of the myxobacterium (Figure 1A and Table 1), (+)-eremophilene (**11**) and germacradienol (**26**) could be unambiguously assigned as products of their corresponding specific TSs sce8552[16] and sce1440, respectively. Among the remaining sesquiterpenes, mainly of the cadalane- and cubebane-type, seven (**1**, **2**, **3**, **6**, **8**, **17**, and **19**) were observed in the extract of the sce6369 overexpressing *E. coli* strain identifying this TS as the most versatile one of So ce56. 10-*Epi*-cubebol (**19**), the main product of sce6369 and also one of the most abundant terpenes in the cell extract of So ce56, designates the enzyme function. In contrast, α-cadinene (**20**) was previously detected in a sce6369 overexpressing *S. avermitilis* strain along with cadina-3,5-diene (**3**), β-cadinene, and bicyclosesquiphellandrene.[12] This difference in the product spectrum of sce6369 might be due to conformational changes of the enzyme in different environments. As shown before, TS selectivities are strongly dependent on the plasticities of the enzyme's active sites. Another possibility for the observed differences in the product patterns could be subsequent reactions catalyzed by autologous enzymes of the various expression strains. To the best of our knowledge, differences in the product spectrum of the same TS, overexpressed in different hosts have not been reported yet and, thus, should be investigated in more detail.

Furthermore, a comprehensive reaction mechanism could be established, which explains the formation of all observed cadalanes including **5**, **16**, and **18**, as well as germacranes **8** and **14** by sce6369 (Scheme 2). The formation of (*E*)-β-farnesene (**4**), γ-gurjunene (**7**)[57] and the unidentified sesquiterpene **13** could not be related to any of the examined TS.

Besides the previously described specific (+)-eremophilene synthase (sce8552),[16] sce1440 was characterized as a germacra-1(10)*E*,5*E*-dien-11-ol/geosmin synthase, sce4363 does not convert FPP. Thus, sce6369 is the first described, however, highly promiscuous 10-*epi*-cubebol (**19**) synthase. Volatile analysis has shown, that *C. crocatus* is the only investigated myxobacterium besides *S. cellulosum* that produces (+)-eremophilene (**11**) and cadalanes.[56] As shown by our bioinformatics analyses, orthologs of sce6369 and sce8552 might be involved in their synthesis. Further analyses will show if the sesquiterpene synthases in *Sorangium cellulosum* So ce56 are unique throughout the bacterial kingdom or if other strains are able to shape FPP to the same compounds.

Experimental

GC/MS analysis of the *S. cellulosum*So ce56 volatiles and the *E. coli* whole-cell conversion extracts

The strain So ce 56 was grown in liquid MD 1 medium as described before.[14] A closed-loop-stripping-apparatus (CLSA) was used to collect the volatiles during 24 h. The volatile fraction was then analyzed by GC/MS, according to the method of Fuhlendorff et al.[56] Terpene production in *E. coli* C43(DE3) cells transformed with pAC-Mv_sce1440, pAC-Mv_sce4636, or pAC-Mv_sce6369 and sampling was conducted as described previously.[16] Online GC/MS analyses of the *E. coli* extracts to observe the occurrence of terpenes was carried out on a GC/MS (Focus GC/DSQ II, Thermo Scientific) equipped with a HP-5ms capillary column (25 m x 0,32 mm i.d., 0,52 μm film, Agilent) using the previously described conditions.[16] GC/MS analyses for structural determination of the terpenes were conducted according to the method of Fuhlendorff et al.[56]

Bioinformatic analysis and molecular cloning of the TS genes

The coding sequences for the TSs sce1440, sce4636, sce6369, and sce8552were identified previously.[18] The sequences for sce1440, sce4636, and sce6369 were amplified by polymerase chain reaction (PCR) from the genomic DNA of *S. cellulosum* So ce56 using primers:
sce1440 forward primer (5'-cca at*catatg*t cat ctgatcgaacta gcg-3') and reverse primer (5'-ccaat*aagctt*tcag tgatggtgatggtgatgtccagcgcgcacgatca-3').
sce4636 forward primer (5'-ccaat*catatg*tcatctgatcgaactagcg-3') and reverse primer (5'-ccaat*aagctt*tcagtgatggtgatggtgatgtccagcgcgcacgatca-3').
sce6369 forward primer (5'-ccaat*catatg*tcatctgatcgaactagcg-3') and reverse primer (5'-ccaat*aagctt*tcag tgatggtgatggtgatgtccagcgcgcacgatca-3').
The purified PCR products were inserted into vector pCR™-Blunt II-TOPO® using the Zero Blunt® TOPO® PCR Cloning Kit (Invitrogen). The *E. coli* expression vectors pAC-Mv_sce1440, pAC-Mv_sce4636, and pAC-Mv_sce6369 were constructed using the *Nde*I and *Hind*III restriction sites in analogy to pAC-MvGo (containing sce8552) from pAC-Mv as described by Schifrin et al.[16]

Product purification and NMR characterization

Products for purification were obtained from *in vivo* conversions and a total reaction volume of 1.9 L. δ-cadinene (**17**),10-*epi*-cubebol (**19**) and germacra-1(10)*E*,5*E*-dien-11-ol (**26**) were purified on silica gel with isocratic eluent mixtures of n-hexane and ethyl acetate (9:1). Fractions containing the products were identified via TLC in a solvent tank containing the same eluent and stained with anis aldehyde.

1,10-di-*epi*-cubenol (**28**) was purified using a setup produced by Waters (Eschborn, Germany) containing a 2767 Sample Manager, a 2545 binary gradient pump, a 2998 PDA detector and a 3100 electron spray mass spectrometer. After separation the solvent flow has been split using a flow splitter and a 515 HPLC pump for makeup flow. Water containing 0.1% formic acid and acetonitrile containing 0.1% formic acid were used as solvents for the analysis and separation. A Waters X-Bridge column (C18, 150 x 19 mm, 5 μM) has been used with a flow of 20 mL/min for the separation. A gradient of 60% to 95% acid containing acetonitrile in eight minutes has been applied.

The obtained products were used for NMR characterization. The NMR spectra of theproducts were recorded in $CDCl_3$ with a Bruker DRX 500 NMR spectrometer. All chemical shifts are relative to $CHCl_3$ at δ=7.26 (^{1}H-NMR) or $CDCl_3$ at δ=77.00 (^{13}C-NMR) using the standard δ notation in parts per million.

Genome mining of Myxobacteria for terpene synthases

The homologues of the terpene synthase encoding open reading frames (ORFs) of respective synthase genes from *Sorangium cellulosum* So ce56 were obtained by comparing the published genomic sequences of the myxobacteria *Anaeromyxobacter* (BAZG00000000), *Myxococcus xanthus* DK 1622 (NC_008095), *Sorangiumcellulosum*So0157-2 (NC_010162.1), *Myxococcus fulvus* HW-1 (NC_015711.1), *Anaeromyxobacter dehalogenans* 2CP-C (NC_007760.1), *Haliangium ochraceum* (NC_013440.1), *Plesiocystis pacifica* (NZ_ABCS00000000.1), *Stigmatella aurantiaca* DW4/3-1 (NC_014623.1), *Byssovorax cruenta* (BBEO00000000.1), *Cystobacter violaceus* Cb vi76 (NZ_JPMI00000000.1), *Corallococcus* (CBAX000000000.1), *Myxococcus* (NZ_CP012109.1), *Myxococcus stipitatus* DSM 14675 (NC_020126.1), *Cystobacter fuscus* DSM 2262 (NZ_ANAH00000000.2), *Chondromyces apiculatus* DSM 436 (NZ_ASRX00000000.1), *Corallococcus coralloides* DSM 2259 (NC_017030.1), *Myxococcales bacterium* SG8_38 (LJTN00000000.1), *Myxococcales bacterium* SG8_38_1 (LJTO00000000.1), *Labilithrix luteola* (CP012333.1), *Vulgatibacter incomptus* (NZ_CP012332.1), *Chondromyces crocatus* (NZ_CP012159.1), *Archangium gephyra* (NZ_CP011509.1), *Sandaracinus amylolyticus* (NZ_CP011125.1), *Hyalangium minutum* (NZ_JMCB00000000.1), and *Enhygromyxa salina* (JMCC00000000.2).
The homologues of the genes were searched using the global NCBI database (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) or the DOE Joint Genome Institute database (https://img.jgi.doe.gov/cgi-bin/m/main.cgi). Thus obtained homologous protein sequences were aligned using CLUSTAL W[58] and the radial view of an unrooted phylogenetic tree was obtained by MEGA4 (version 4.0)[59] analysis for the determination of relatedness of the TSs from *Sorangium cellulosum* So ce56 with the TSs from other myxobacterial species.

δ-cadinene (17)

^{1}H-NMR (500MHz, $CDCl_3$): δ0.77, d, J = 6.9 Hz, 3H and 0.94, d, J = 6.9 Hz, 3H (H-12 and H-13); 1.03, m (H-7); 1.15, m (H-8a); 1.59, m (H-8b); 1.63, s, 3H (H-14); 1.65, s, 3H (H-15); 2.06-1.84, m, 6H (H-11, H-9a, H-9b, H-3a, H-3b, H-2a); 2.50, br d, J = 9.5 Hz (H-6); 2.69, ddd, J = 12.8, 4.7 and 2.5 Hz (H-2b); 5.43, m (H-5).^{13}C-NMR (125MHz, $CDCl_3$): δ15.7and 21.8 (C-12 and C-13);18.5 (C-14); 21.2 (C-8); 23.6, (C-15); 26.7 and 26.8 (C-2 and C-11); 32.0 and 32.3 (C-3 and C-9); 39.5 (C-6); 45.4 (C-7); 124.5 (C-10); 124.7 (C-5); 129.9 (C-1); 134.2 (C-4)

10-*epi*-cubebol (19)

^{1}H-NMR (500MHz, $CDCl_3$): δ0.77, t, J = 2.9 Hz (H-6); 0.92, m (H-5); 0.92, d, J = 6.7 Hz, 3H and 0.95, d, J = 6.7 Hz, 3H (H-12 and H-13); 0.97, d, J = 6.8 Hz, 3H (H-14); 1.03 – 1.09, m (H-8a); 1.13 – 1.16, m (H-8b); 1.21 – 1.26, m (H-7); 1.28, d, J = 0.7 Hz, 3H (H-15); 1.30 – 1.35, m, 2H (H-2); 1.33, m (H-3a); 1.52, m (H-9a); 1.53, m (H-3b); 1.63, m (H-11); 1.73, m (H-9b); 1.92, m (H-10). ^{13}C-NMR (125MHz, $CDCl_3$): δ17.4 (C-14); 19.2 and 19.9 (C-12 and C-13); 19.6 (C-8); 20.8 (C-6); 28.0, (C-15); 29.3 (C-2 and C-10); 30.8 (C-9); 32.9 (C-1); 33.7 (C-11); 36.4 (C-3); 41.1 (C-5); 42.4 (C-7); 80.3 (C-4)

germacra-1(10)*E*,5*E*-dien-11-ol (26)

^{1}H-NMR (500MHz, $CDCl_3$): δ1.22, d, J = 2.0 Hz, 3H and 1.32, d, J = 2.0 Hz, 3H (H-12 and H-13); 1.25, d, J = 6.9 Hz, 3H (H-15); 1.33 – 1.42, m, (H-8a); 1.61 – 1.66, m, (H-8b); 1.69, br s, 4H (H-14 and H-3a); 1.81 – 1.88, m, (H-8b); 2.04, br d, J = 14.8 Hz (H-2a); 2.33 – 2.43, m, 3H (H-7 and H-9); 2.54, m, (H-2b); 2.57, m, (H-4); 5.13, m, (H-6); 5.17, d, J = 11.6 Hz (H-1); 5.80, d, J = 16.0 Hz (H-5). ^{13}C-NMR (125MHz, $CDCl_3$): δ14.7 (C-15); 16.7 (C-14); 22.0 (C-2), 23.7 (C-8); 26.2 and 26.8 (C-12 and C-13); 32.7 (C-3); 33.8 (C-4); 41.2 (C-9); 58.9 (C-7); 71.9 (C-11); 123.6 (C-6); 130.6 (C-1); 131.1 (C-10); 143.3 (C-5)

1,10-di-*epi*-cubenol (28)

^{1}H-NMR (500MHz, $CDCl_3$): δ0.84, d, J = 6.9 Hz, 3H and 0.90, d, J = 6.9 Hz, 3H (H-12 and H-13); 1.01, d, J = 7.1 Hz, 3H (H-14); 1.13, sept, J = 4.9 (H-7); 1.30 – 1.32, m, 2H (H-2); 1.35 – 1.41, m (H-8a); 1.48 - 1.53, m, 2H (H-9); 1.68, s, 3H (H-15); 1.78, m (H-10); 1.96 – 2.00, m (H-3a and H-11); 2.05, m (H-6); 2.09 – 2.18, m, 2H (H-3b and H-8b); 5.36, m, (H-1). ^{13}C-NMR (125MHz, $CDCl_3$): δ13.9 (C-14); 16.6 and 21.8 (C-12 and C-13), 18.9 (C-2); 23.4 (C-15); 27.2(C-11); 27.8 (C-3); 28.5 (C-9); 31.2 (C-8); 37.1 (C-10); 41.8 (C-6); 49.1 (C-7); 72.3 (C-1); 123.3 (C-5); 133.7 (C-4)

Acknowledgements

This work was supported by a grant from the Deutsche Forschungsgemeinschaft to R. B. (Be1343/23). The authors want to express their gratitude to Reiner Wintringer (Universität des Saarlandes, Bioanalytik) for help with the GC/MS measurements, to Dr. Carsten Börger (PharmBioTec GmbH) for purification of 1,10-di-epi-cubenol, and to Dr. Josef Zapp (Universität des Saarlandes, Pharmazie) for NMR measurements. We thank N. Misawa for providing the plasmid pAC-Mv and D. M. Hodgson for an authentic sample of 10-epi-cubebol.

References

1. D. W. Christianson, *Chemical Reviews*, 2006, **106**, 3412-3442.
2. D. J. Miller and R. K. Allemann, *Nat. Prod. Rep.*, 2012, **29**, 60-71.
3. Y. Gao, R. B. Honzatko and R. J. Peters, *Nat. Prod. Rep.*, 2012, **29**, 1153-1175.
4. D. E. Cane, M. Tandon and P. C. Prabhakaran, *Journal of the American Chemical Society*, 1993, **115**, 8103-8106.
5. X.-Y. Chen, Y. Chen, P. Heinstein and V. J. Davisson, *Archives of Biochemistry and Biophysics*, 1995, **324**, 255-266.
6. C. L. Steele, J. Crock, J. Bohlmann and R. Croteau, *Journal of Biological Chemistry*, 1998, **273**, 2078-2089.
7. B. Felicetti and D. E. Cane, *Journal of the American Chemical Society*, 2004, **126**, 7212-7221.
8. S. Agger, F. Lopez-Gallego and C. Schmidt-Dannert, *Molecular Microbiology*, 2009, **72**, 1181-1195.
9. C. A. Citron, L. Barra, J. Wink and J. S. Dickschat, *Organic & Biomolecular Chemistry*, 2015, **13**, 2673-2683.
10. D. E. Cane and H. Ikeda, *Acc. Chem. Res.*, 2012, **45**, 463-472.
11. C. A. Citron and J. S. Dickschat, *Synlett*, 2014, **25**, 766-782.
12. Y. Yamada, T. Kuzuyama, M. Komatsu, K. Shin-ya, S. Omura, D. E. Cane and H. Ikeda, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 857-862.
13. J. S. Dickschat, H. B. Bode, T. Mahmud, R. Müller and S. Schulz, *The Journal of Organic Chemistry*, 2005, **70**, 5174-5182.
14. J. S. Dickschat, T. Nawrath, V. Thiel, B. Kunze, R. Müller and S. Schulz, *Angewandte Chemie International Edition*, 2007, **46**, 8287-8290.
15. N. L. Brock, S. R. Ravella, S. Schulz and J. S. Dickschat, *Angewandte Chemie International Edition*, 2013, **52**, 2100-2104.
16. A. Schifrin, T. T. B. Ly, N. Günnewich, J. Zapp, V. Thiel, S. Schulz, F. Hannemann, Y. Khatri and R. Bernhardt, *ChemBioChem*, 2015, **16**, 337-344.
17. S. Schneiker, O. Perlova, O. Kaiser, K. Gerth, A. Alici, M. O. Altmeyer, D. Bartels, T. Bekel, S. Beyer, E. Bode, H. B. Bode, C.

J. Bolten, J. V. Choudhuri, S. Doss, Y. A. Elnakady, B. Frank, L. Gaigalat, A. Goesmann, C. Groeger, F. Gross, L. Jelsbak, J. Kalinowski, C. Kegler, T. Knauber, S. Konietzny, M. Kopp, L. Krause, D. Krug, B. Linke, T. Mahmud, R. Martinez-Arias, A. C. McHardy, M. Merai, F. Meyer, S. Mormann, J. Munoz-Dorado, J. Perez, S. Pradella, S. Rachid, G. Raddatz, F. Rosenau, C. Ruckert, F. Sasse, M. Scharfe, S. C. Schuster, G. Suen, A. Treuner-Lange, G. J. Velicer, F. J. Vorholter, K. J. Weissman, R. D. Welch, S. C. Wenzel, D. E. Whitworth, S. Wilhelm, C. Wittmann, H. Blöcker, A. Pühler and R. Müller, *Nat. Biotechnol.*, 2007, **25**, 1281-1289.

18. Y. Yamada, D. E. Cane and H. Ikeda, in *Methods in Enzymology*, ed. A. H. David, Academic Press, 2012, vol. Volume 515, pp. 123-162.
19. P. Baer, P. Rabe, K. Fischer, C. A. Citron, T. A. Klapschinski, M. Groll and J. S. Dickschat, *Angewandte Chemie International Edition*, 2014, **53**, 7652-7656.
20. D. E. Cane and M. Tandon, *Journal of the American Chemical Society*, 1995, **117**, 5602-5603.
21. P. Rabe, L. Barra, J. Rinkel, R. Riclea, C. A. Citron, T. A. Klapschinski, A. Janusko and J. S. Dickschat, *Angewandte Chemie International Edition*, 2015, DOI: 10.1002/anie.201507615, n/a-n/a.
22. P. Rabe, C. A. Citron and J. S. Dickschat, *ChemBioChem*, 2013, **14**, 2345-2354.
23. H. Harada, F. Yu, S. Okamoto, T. Kuzuyama, R. Utsumi and N. Misawa, *Appl. Microbiol. Biotechnol.*, 2009, **81**, 915-925.
24. P. Rabe and J. S. Dickschat, *Angewandte Chemie International Edition*, 2013, **52**, 1810-1812.
25. D. M. Hodgson, S. Salik and D. J. Fox, *The Journal of Organic Chemistry*, 2010, **75**, 2157-2168.
26. D. Spiteller, A. Jux, J. Piel and W. Boland, *Phytochemistry*, 2002, **61**, 827-834.
27. D. E. Cane and R. M. Watt, *Proceedings of the National Academy of Sciences*, 2003, **100**, 1547-1551.
28. D. E. Cane, X. He, S. Kobayashi, S. Omura and H. Ikeda, *J Antibiot*, 2006, **59**, 471-479.
29. J. Jiang, X. He and D. E. Cane, *Journal of the American Chemical Society*, 2006, **128**, 8128-8129.
30. J. Jiang, X. He and D. E. Cane, *Nat Chem Biol*, 2007, **3**, 711-715.
31. X. Xie, J. Kirby and J. D. Keasling, *Phytochemistry*, 2012, **78**, 20-28.
32. P. Schreier, F. Drawert and A. Junker, *Z Lebensm Unters Forsch.*, 1976, **160**, 271-274.
33. D. E. Cane, B. J. Rawlings and C. C. Yang, *J Antibiot (Tokyo).* 1987, **40**, 1331-1334.
34. Y. Hu, W. K. W. Chou, R. Hopson and D. E. Cane, *Chemistry & Biology*, 2011, **18**, 32-37.
35. L. Zalkow, R. Howard, L. Gelbaum, M. Gordon, H. Deutsch and M. Blum, *J Chem Ecol*, 1981, **7**, 717-731.
36. R. Bartelt, A. Cossé, B. Zilkowski, D. Weisleder and F. Momany, *J Chem Ecol*, 2001, **27**, 2397-2423.
37. J.-J. Filippi, E. Belhassen, N. Baldovini, H. Brevard and U. J. Meierhenrich, *Journal of Chromatography A*, 2013, **1288**, 127-148.
38. T. Hieda, M. Tazaki, Y. Morishita, T. Aoki and S. Nagahama, *Phytochemistry*, 1996, **42**, 159-162.
39. F. Lopez-Gallego, S. A. Agger, D. Abate-Pella, M. D. Distefano and C. Schmidt-Dannert, *ChemBioChem*, 2010, **11**, 1093-1106.
40. J. Degenhardt, T. G. Köllner and J. Gershenzon, *Phytochemistry*, 2009, **70**, 1621-1637.
41. K. Zhou and R. J. Peters, *Chemical Communications*, 2011, **47**, 4074-4080.
42. K. M. Ewen, F. Hannemann, Y. Khatri, O. Perlova, R. Kappl, D. Krug, J. Hüttermann, R. Müller and R. Bernhardt, *J. Biol. Chem.*, 2009, **284**, 28590-28598.
43. Y. Khatri, F. Hannemann, K. M. Ewen, D. Pistorius, O. Perlova, N. Kagawa, A. O. Brachmann, R. Müller and R. Bernhardt, *Chem. Biol.*, 2010, **17**, 1295-1305.
44. A. Schifrin, M. Litzenburger, M. Ringle, T. Ly Thi Bich and R. Bernhardt, *ChemBioChem*, 2015, DOI: 10.1002/cbic.201500417, n/a-n/a.
45. T. B. Ly, Y. Khatri, J. Zapp, M. Hutter and R. Bernhardt, *Appl. Microbiol. Biotechnol.*, 2012, **95**, 123-133.
46. M. Litzenburger, F. Kern, Y. Khatri and R. Bernhardt, *Drug Metab. Dispos.*, 2014, DOI: 10.1124/dmd.114.061937.
47. F. Kern, T. K. F. Dier, Y. Khatri, K. M. Ewen, J.-P. Jacquot, D. A. Volmer and R. Bernhardt, *Scientific Reports*, 2015, **5**, 14881.
48. M. Litzenburger and R. Bernhardt, *Appl. Biochem. Biotechnol.*, 2016, **in press**.
49. M. W. Lodewyk, P. Gutta and D. J. Tantillo, *The Journal of Organic Chemistry*, 2008, **73**, 6570-6579.
50. D. Arigoni, *Pure Appl. Chem.*, 1975, **41**, 219.
51. Y. Khatri, F. Hannemann, O. Perlova, R. Müller and R. Bernhardt, *FEBS Letters*, 2011, **585**, 1506-1513.
52. S. C. Wenzel and R. Müller, *Nat. Prod. Rep.*, 2009, **26**, 1385-1407.
53. J. S. Dickschat, *Natural Product Reports*, 2016, DOI: 10.1039/c5np00102a.
54. J. S. Dickschat, K. A. K. Pahirulzaman, P. Rabe and T. A. Klapschinski, *ChemBioChem*, 2014, **15**, 810-814.
55. C. Nakano, T. Tezuka, S. Horinouchi and Y. Ohnishi, *J Antibiot*, 2012, **65**, 551-558.
56. S. Schulz, J. Fuhlendorff and H. Reichenbach, *Tetrahedron*, 2004, **60**, 3863-3872.
57. C. O. Schmidt, H. J. Bouwmeester, N. Bülow and W. A. König, *Archives of Biochemistry and Biophysics*, 1999, **364**, 167-177.
58. J. D. Thompson, D. G. Higgins and T. J. Gibson, *Nucleic Acids Research*, 1994, **22**, 4673-4680.
59. K. Tamura, J. Dudley, M. Nei and S. Kumar, *Mol. Biol. Evol.*, 2007, **24**, 1596-1599.