## Accepted Article

WILEY-VCH

COMMUNICATION

Two Diterpene Synthases for Spiroalbatene and Cembrene A from *Allokutzneria albata*

Jan Rinkel,[a] Lukas Lauterbach,[a] Patrick Rabe[a] and Jeroen S. Dickschat*[a]

Abstract: Two bacterial diterpene synthases from the actinomycete *Allokutzneria albata* were investigated, resulting in the identification of the structurally unprecedented compound spiroalbatene from the first and cembrene A from the second enzyme. Both enzymes were deeply investigated for their mechanisms by isotopic labelling experiments, site-directed mutagenesis, and variation of metal cofactors and pH. For spiroalbatene synthase the pH- and Mn^{2+}-dependent formation of the byproduct thunbergol was observed that is biosynthetically linked to spiroalbatene.

The astonishing structural complexity of terpenes is reflected by hundreds of known polycyclic carbon skeletons. These are all formed by terpene synthases (TS) in a single enzymatic step from a few linear oligoprenyl diphosphate (OPP) precursors, including geranyl diphosphate (GPP) for monoterpenes (C_{10}), farnesyl PP (FPP) for sesquiterpenes (C_{15}), geranylgeranyl PP (GGPP) for diterpenes (C_{20}), and geranylfarnesyl PP (GFPP) for sesterterpenes (C_{25}).[1] Starting from the pentalenene synthase from *Streptomyces exfoliatus*,[2] previous research has resulted in the discovery of several bacterial mono- (MTSs) and sesquiterpene synthases (STS).[1] In contrast, only a few type I diterpene synthases (DTSs) including those for cyclooct-9-en-7-ol (CotB2), spiroviolene, tsukubadiene, hydropyrene, 18-hydroxydolabella-3,7-diene, and spata-13,17-diene have been characterised,[3] with spata-13,17-diene synthase as the first bacterial enzyme also making sesterterpenes.[3f] Because GGPP has more options to react in comparison to GPP and FPP, the chance to discover new compounds is higher for diterpenes than for mono- and sesquiterpenes. It is not possible to predict directly from the amino acid sequence of a TS which substrate preference it has, but a phylogeny of TSs can unravel candidates with a relationship close enough to a characterised DTSs to suspect DTS activity, and distant enough to ensure that the TS does not make the same product as the known enzyme. Using this approach, we have identified two TSs from the tropical soil isolate *Allokutzneria albata* DSM 44149[4] with close relationship to other bacterial DTSs, as can be seen from a phylogenetic tree constructed from 2728 bacterial TSs (Figure S1).

The genes for both enzymes were cloned into the expression vector pYE-Express[5] by homologous recombination in yeast, followed by gene expression in *Escherichia coli* and protein

purification (Figure S2, accession numbers are given in the SI). The first enzyme did not accept GPP, FPP and GFPP, but GGPP was efficiently converted into a diterpene hydrocarbon (Figure S3). The compound was purified and its structure was elucidated by NMR spectroscopy (Table S2, Figures S4-S10). The most relevant $^1H,^1H$-COSY, HMBC and NOESY correlations are shown in Scheme 1A, resulting in the structure of a spirotetracyclic diterpene with novel skeleton for which we propose the name spiroalbatene (**1**). Thus, the DTS was identified as spiroalbatene synthase (SaS).
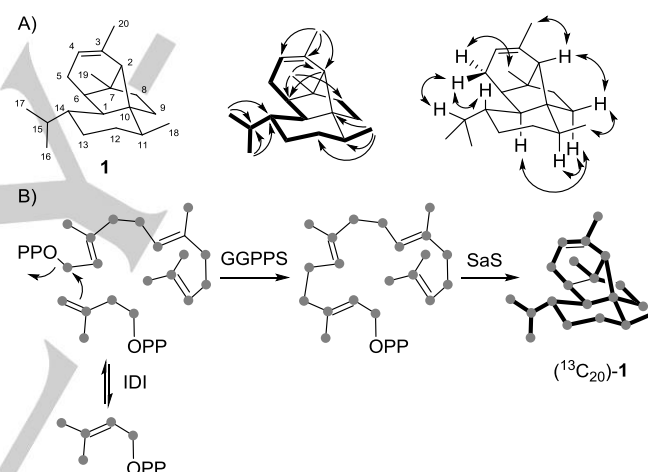
Scheme 1. Spiroalbatene (**1**). A) The carbon numbering for **1** indicates which carbon derives from which carbon of GGPP by same number. Bold lines: $^1H,^1H$-COSY correlations, single-headed arrows: HMBC correlations, double-headed arrows: NOESY correlations. B) Enzymatic synthesis of ($^{13}C_{20}$)-**1**. Grey dots represent ^{13}C atoms and bold lines $^{13}C,^{13}C$-COSY correlations.

Further support for the structure of **1** was obtained by $^{13}C,^{13}C$-COSY NMR. For this purpose, we prepared completely isotopically substituted ($^{13}C_{20}$)GGPP from ($^{13}C_{15}$)FPP.[6] During our previous synthesis of this compound also ($^{13}C_5$)-3,3-dimethylallyl alcohol was made available that was converted into ($^{13}C_5$)DMAPP. Incubation of both compounds with the isopentenyl diphosphate isomerase (IDI) from *Serratia plymuthica*, the GGPP synthase (GGPPS) from *Streptomyces cyaneofuscatus*,[3c] and SaS resulted in the efficient production of ($^{13}C_{20}$)-**1** (Scheme 1B). Analysis of the compound by $^{13}C,^{13}C$-COSY NMR (Figure S11) resulted in the observation of crosspeaks for all C-C-connectivities apart from one.

A biosynthetic model for **1** is shown in Scheme 2. Starting from GGPP, a 1,14-cyclisation to cation **A** is followed by a 1,2-hydride migration to **B**. Another 1,2-hydride shift to **C** and ring closure results in **D** that upon a third 1,2-hydride shift and cyclisation

[a]   Prof. Dr. Jeroen S. Dickschat, Jan Rinkel, Lukas Lauterbach, Dr.
        Patrick Rabe
        Kekulé-Institute of Organic Chemistry and Biochemistry
        University of Bonn
        Gerhard-Domagk-Straße 1, 53121 Bonn, Germany
        E-mail: dickschat@uni-bonn.de

        Supporting information for this article is given via a link at the end of
   the document.

COMMUNICATION

yields the cyclopropane intermediate **E**. Opening of the three-membered ring with concomitant ring closure produces **F**. If the **E**-to-**F** conversion is concerted, a secondary cation intermediate can be avoided. A last ring closure and deprotonation result in **1**. This biosynthetic hypothesis was tested by conversion of all twenty isotopomers of (^{13}C)GGPP that were prepared by synthesis or enzymatically from corresponding ^{13}C-labelled FPP and IPP isotopomers.[3c,6] For all substrates the label ended up in a position as expected for the model (Figure S12).

The elementary steps involving hydrogen migrations or deprotonations were investigated by use of (stereo)selectively deuterated probes. In several cases an additional ^{13}C substitution was introduced to enhance the carbon's signal in ^{13}C-NMR analyses and to make resulting ^{13}C-^{2}H bonds observable by triplet signals. The first 1,2-hydride shift from **A** to **B** was followed by conversion of (14-^{2}H)GGPP with SaS, prepared from IPP and (2-^{2}H)DMAPP[3e] with *Streptomyces coelicolor* FPP synthase (FPPS)[7] and GGPPS.[3c] GC/MS analysis of the product showed incorporation of one deuterium by an increased molecular ion, while the fragment ion of **1** at $m/z = 229$ representing the loss of the *i*Pr group did not change (Figure S13A). Therefore, deuterium incorporation was located in the *i*Pr group as expected. A 1,3-hydride migration from C1 into the *i*Pr group as an alternative to the two sequential 1,2-hydride migrations from **A** to **C** was ruled out by incubation of (*S*)- and (*R*)-(1-^{2}H)GGPP[3c] with SaS that yielded products with a fragment ion at $m/z = 230$, meaning that deuterium stayed within the tetracyclic core structure of **1** in both cases (Figure S13B). The ^{13}C-labelling from (16-^{13}C)- and (17-^{13}C)GGPP showed a scrambling (Figure S12) that was previously observed for other terpene cyclisations with 1,2-hydride migrations into the *i*Pr group[3e,8] and reflects the similar distances for the migrating hydrogen to the two diastereotopic faces of the cation in **A**. In contrast, for 1,3-hydride migrations always a strict stereochemical course is observed.[9] The stereochemical course for the second 1,2-hydride shift from **B** to **C** was investigated by conversion of (*S*)- and (*R*)-(1-^{13}C,1-^{2}H)GGPP[3e] with SaS. While the *S*-enantiomer yielded a product with a singlet for C1 in the ^{13}C-NMR, the *R*-enantiomer resulted in a product with a triplet, establishing the specific migration of the *pro-S*-hydrogen from C1 of GGPP (Figure S14A). The 1,2-hydride transfer from **D** to **E** was investigated with (3-^{13}C,2-^{2}H)GPP, synthesised according to a published route (Scheme S1),[3c] that was elongated with IPP and GGPPS to (11-^{13}C,10-^{2}H)GGPP followed by cyclisation with SaS. The product showed the expected triplet in the ^{13}C-NMR spectrum (Figure S14B). The terminal deprotonation step to **1** proceeds from C4 of GGPP. Which of the two enantiotopic protons is lost can be addressed by stereoselective deuteration. For this purpose, (*E*)- and (*Z*)-(4-^{2}H)IPP were synthesised (Scheme S2) and used for the GGPPS-catalysed elongation of FPP to obtain (*S*)- and (*R*)-(4-^{2}H)GGPP (Figure S15). The stereochemical course for this reaction includes attack of IPP at C4 from the *Si* face.[10] Further conversion with SaS revealed specific loss of deuterium from (*E*)-(4-^{2}H)IPP, but retainment from (*Z*)-(4-^{2}H)IPP, revealing a deprotonation of **G** as shown in Scheme 2.

We have recently developed a method that uses the stereoselectively deuterated precursors (*S*)- and (*R*)-(1-^{13}C,1-^{2}H)FPP and (*S*)- and (*R*)-(1-^{13}C,1-^{2}H)GPP for determination of the absolute configurations of terpenes.[3c] The elongation of OPPs with IPP by OPP synthases proceeds with inversion of

configuration at C1.[11] After conversion of the enantioselectively deuterated substrates with GGPPS and SaS the defined stereochemical anchors at the deuterated carbons were used to solve the absolute configuration of **1** by determining the relative configurations of the stereospecifically deuterated products. The additional ^{13}C substitution was introduced for highly sensitive product analysis by HSQC (Figures S16 and S17). The results with all four probes consistently pointed to the absolute configuration of **1** as shown in Scheme 1.
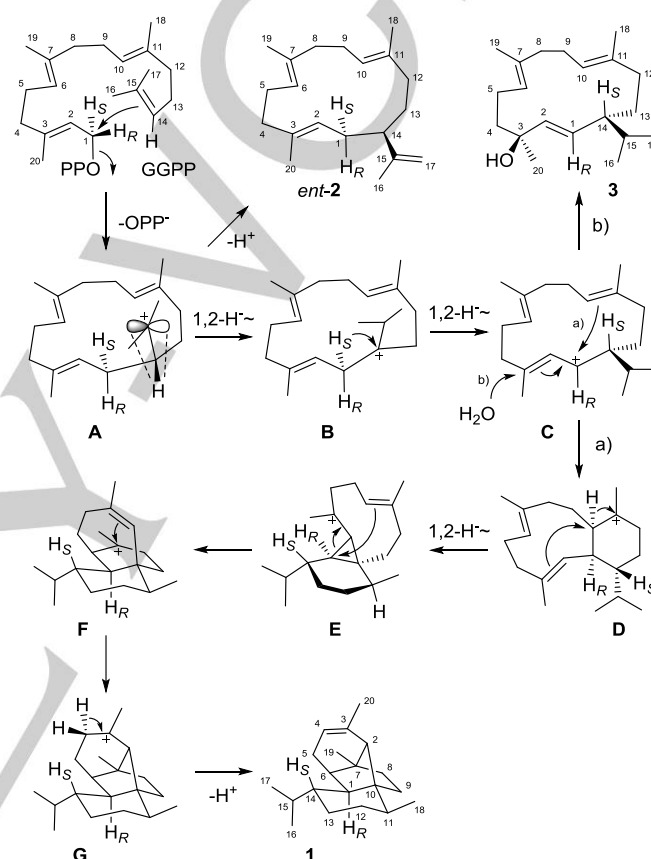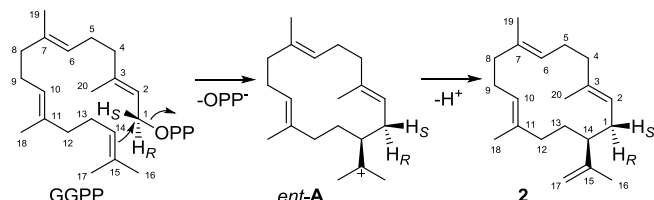

Scheme 2. Cyclisation mechanism from GGPP to **1** by SaS, hypothetical formation of *ent*-**2** from **A**, and formation of **3** from **C**.

The second enzyme from *A. albata* also did not accept GPP, FPP and GFPP, but efficiently converted GGPP into one main and a few side products (Figure S18). The main product was purified and showed identical NMR data to those of cembrene A (Table S3),[12] while the optical rotation was $[\alpha]_D^{20} = +12.0$ (*c* 0.13, CHCl$_3$). With the published optical rotation for the synthetic *S* ($[\alpha]_D = +19.5$)[13] and *R* enantiomers ($[\alpha]_D^{20} = -12$, *c* 0.85, CHCl$_3$),[12] these data pointed to the structure of (*S*)-(+)-cembrene A (**2**, Scheme 3) and identified the enzyme from *A. albata* as cembrene A synthase (CAS). The cyclisation mechanism from GGPP to **2** requires a 1,14-cyclisation to the cembranyl cation (*ent*-**A**) and deprotonation from one of the geminal Me groups. SaS could potentially yield *ent*-**2** from intermediate **A**, reflecting the phylogenetic distance between the two enzymes, but *ent*-**A** is not observed from SaS. Incubation of (16-^{13}C)- and (17-^{13}C)GGPP

COMMUNICATION

revealed a distribution of labelling and therefore a relaxed stereochemical course for the deprotonation step (Figure S19). Incubation of (S)- and (R)-(1-^{13}C,1-^{2}H)GGPP with CAS and product analysis by HSQC showed inversion of configuration at C1 for the 1,14-cyclisation of GGPP (Figure S20).

Scheme 3. Cyclisation mechanism from GGPP to **2**.

Type I TSs generally exhibit the same α-helical fold as first recognised for avian FPPS.[14] Bacterial enzymes, apart from geosmin synthase with two α domains,[15] adopt a single α domain architecture with several highly conserved motifs involved in binding of the metal cofactor and substrate recognition.[16] This includes the aspartate-rich motif **DD**XX(X)(D,E), the NSE triad **N**D(L,I,V)X**S**XX(R,K)(E,D), the diphosphate (PP) sensor (**R** near the helix G break), and the R**Y** dimer near the C-terminus. For SaS the Mg^{2+} cofactor can be substituted by Mn^{2+} resulting in a reduced activity of 27±17%, while for CAS a higher activity (172±12%) is observed with Mn^{2+} (Figure S21). No activity was observed with GGPP and SaS, if no metal cofactor was added, demonstrating that not residual Mg^{2+} in buffers or enzyme preparations accounts for activity in the experiments with Mn^{2+} (and vice versa).

The importance of several conserved residues in TSs for catalytic activity has been demonstrated by site-directed mutagenesis (SDM, critical residues identified in previous mutational studies are shown in bold above).[3f,15,17] SaS exhibits all the conserved motifs as usual, with the exception of a Gly instead of the Ser in the NSE triad (Figure S22 and S23). Among all >50 characterised bacterial TSs and their nearly 2500 close homologs from sequenced bacteria (Figure S1), in only four cases a substitution of this Ser is observed (Table S4). Installation of Ser by SDM (G229S) gave no soluble enzyme, suggesting that Gly with its conformational flexibility is of structural importance for enzyme folding of wildtype SaS. The crystal structure of selina-4(15),7(11)-diene synthase (SdS)[17d] shows hydrogen bridges between the PP sensor Arg178, a highly conserved Tyr174 found in >90% of bacterial TSs, and the first Asp225 of the NSE triad (Figure S23). In SaS and in ca. 5% of bacterial TSs a Phe175 is found instead of Tyr. The F175Y variant gave a good enzyme yield, but the activity dropped to 28±6% with Mg^{2+}, and a similar loss of activity was observed with Mn^{2+}.

For spata-13,17-diene synthase (SpS) from *Streptomyces xinghaiensis* a critical role of the highly conserved Pro and Leu located 21 and 14 positions upstream of the Asp-rich motif was recently demonstrated.[3f] The corresponding L72A variant of SaS gave no soluble protein, while P65A yielded wildtype levels of enzyme, but the activity dropped significantly with Mg^{2+} and Mn^{2+}. As can be seen in the crystal structure of SdS,[17d] these two residues are placed at the link between helices C and D (Figure S24) and are likely of structural relevance. In CAS the corresponding position to Pro61 is naturally occupied by Ala. In this case the A64P variant has lower activity than the wildtype, but only with Mg^{2+}, while no effect was observed with Mn^{2+}.

A highly conserved Glu residue seen in >93% of bacterial TSs is occupied by Gln160 in SaS. In SdS this residue is involved in metal cofactor binding (Figure S25). While gene expression of the SaS Q160E variant worked well, the enzyme showed no activity regardless of using Mg^{2+} or Mn^{2+}. For CAS the opposite exchange (E179Q) gave a strongly reduced activity with both metal cations, and the E179K variant was inactive. An interesting structural feature of SdS is a salt bridge between a highly conserved Arg144 and a less conserved Glu192. We have recently shown for SpS that the D217E mutation within the salt bridge (Arg169 and Asp217) results in an inactive enzyme variant in combination with Mn^{2+}, while an increased activity was observed with Mg^{2+}.[3f] The corresponding residues for SaS (Arg145 and Glu193) were mutated with shortening of the salt bridge by the R145K or E193D exchanges, resulting in a reduced (with Mg^{2+}) or complete loss of activity (with Mn^{2+}). For the R145M variant an only slightly decreased activity was observed for both metal cofactors. Of particular interest is the observation that both SpS[3f] and CAS have a naturally short salt bridge with participation of an Asp instead of the more regular Glu, and both enzymes are more efficient with Mn^{2+} than with Mg^{2+}. For CAS the D212E mutation resulted in a strongly decreased activity with Mg^{2+} and Mn^{2+}. These data suggest an influence of the length of this salt bridge likely by alteration of the angle of slope between helices G and F on the active site architecture, with consequences for metal binding. A short salt bridge with participation of Asp seems to favour catalysis with Mn^{2+}, while a long salt bridge with Glu fosters Mg^{2+} binding, but further experiments with other TSs are required to investigate whether this finding is generalisable.

A highly conserved Pro was identified 22 residues downstream of the PP sensor. In SdS this residue is located at the bottom of helix H and seems to have an important structural function (Figure S26). In a few TSs the corresponding position is occupied by Thr or Ser, but the P201T variant of SaS showed a significantly reduced activity with both metal cofactors.

SaS was also investigated for an effect of changing the pH from pH = 7.4 used in the experiments above to pH = 8.2. While no significant change in the product profile was observed with the wildtype enzyme in combination with Mg^{2+}, incubations with Mn^{2+} resulted in a new byproduct (Figure S3). The compound was isolated and identified by NMR (Table S5 and Figures S27-S33) as thunbergol (**3**). Its absolute configuration of (1R,4S)-**3** was evident from the optical rotation ([α]$_D^{20}$ = –48, c 0.85, CHCl$_3$) by comparison to (1S,4R)-**3** from *Pseudotsuga menziesii* ([α]$_D^{20}$ = +75.2, c 1.2, CHCl$_3$),[18] establishing another example for a bacterial TS producing the opposite enantiomer as in plants.[1,19] Compound **3** can be formed via the same intermediate **C** as **1** (Scheme 2) which is corroborated by the same configuration of the stereocentre at C14 for both products. The formation of **1** and **3** via **C** as a common intermediate is also mechanistically supported: Incubation of (2-^{2}H)DMAPP with IPP, GGPPS and SaS followed by GC/MS analysis of the product **3** revealed the same 1,2-hydride shift from **A** to **B** as determined for **1** (Figure S34), again with a distribution of labelling from C16 and C17

COMMUNICATION

(Figure S35). And as observed for **1**, the enzymatic conversion of (*R*)- and (*S*)-(1-^{13}C,1-^{2}H)GGPP with SaS gave evidence for a stereoselective migration of the 1-*pro-S* hydrogen to **C** also for **3** (Figure S36).

None of the diterpenes identified from SaS or CAS were observed in headspace extracts from *A. albata*, while spiroviolene was reported in our previous study,[3c] suggesting that both enzymes are not expressed in laboratory cultures. Alternatively, **1** may be oxidised to an unknown product by a genetically clustered cytochrome P450 (Figure S37 and Table S6).

In summary, we have identified the products of two DTSs from *A. albata*, one of them exhibiting the unprecedented structure of spiroalbatene. While the general principles of terpene biosynthesis are well understood, the recently accumulating data from isotopic labelling experiments, enzyme crystallisations, enzyme mutagenesis and variation of incubation conditions lead to more detailed insights that may allow for predictive approaches how to get access to new terpenes in the future. The phylogenetic tree of bacterial TSs (Figure S1) shows that the products of many enzymes are known today, but there are also many dark areas for continuing research on this interesting enzyme class.

Acknowledgements

[1] J. S. Dickschat, *Nat. Prod. Rep.* **2016**, *33*, 87.

[2] D. E. Cane, C. Pargellis, *Arch. Biochem. Biophys.* **1987**, *254*, 421.

[3] a) A. Meguro, Y. Motoyoshi, K. Teramoto, S. Ueda, Y. Totsuka, Y. Ando, T. Tomita, S.-Y. Kim, T. Kimura, M. Igarashi, R. Sawa, T. Shinada, M. Nishiyama, T. Kuzuyama, *Angew. Chem. Int. Ed.* **2015**, *54*, 4353; b) Y. Yamada, T. Kuzuyama, M. Komatsu, K. Shin-ya, S. Omura, D. E. Cane, H. Ikeda, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 857; c) P. Rabe, J. Rinkel, E. Dolja, T. Schmitz, B. Nubbemeyer, T. H. Luu, J. S. Dickschat, *Angew. Chem. Int. Ed.* **2017**, *56*, 2776; d) J. S. Dickschat, J. Rinkel, P. Rabe, A. Beyraghdar Kashkooli, H. J. Bouwmeester, *Beilstein J. Org. Chem.* **2017**, *13*, 1770; e) J. Rinkel, P. Rabe, X. Chen, T. G. Köllner, F. Chen, J. S. Dickschat, *Chem. Eur. J.* **2017**, *23*, 10501; f) J. Rinkel, L. Lauterbach, J. S. Dickschat, *Angew. Chem. Int. Ed.* **2017**, *56*, 16385.

[4] K. Tomita, Y. Hoshino, T. Miyaki, *Int. J. Syst. Bacteriol.* **1993**, *43*, 297.

[5] J. S. Dickschat, K. A. K. Pahirulzaman, P. Rabe, T. A. Klapschinski, *ChemBioChem* **2014**, *15*, 810.

[6] P. Rabe, L. Barra, J. Rinkel, R. Riclea, C. A. Citron, T. A. Klapschinski, A. Janusko, J. S. Dickschat, *Angew. Chem. Int. Ed.* **2015**, *54*, 13448.

[7] P. Rabe, J. Rinkel, B. Nubbemeyer, T. G. Köllner, F. Chen, J. S. Dickschat, *Angew. Chem. Int. Ed.* **2016**, *55*, 15420.

[8] a) I. Burkhardt, T. Siemon, M. Henrot, L. Studt, S. Rösler, B. Tudzynski, M. Christmann, J. S. Dickschat, *Angew. Chem. Int. Ed.* **2016**, *55*, 8748.

[9] a) J. Rinkel, P. Rabe, P. Garbeva, J. S. Dickschat, *Angew. Chem. Int. Ed.* **2016**, *55*, 13593; b) P. Rabe, T. Schmitz, J. S. Dickschat, *Beilstein J. Org. Chem.* **2016**, *12*, 1839.

[10] a) J. W. Cornforth, R. H. Cornforth, G. Popjak, L. Yengoyan, *J. Biol. Chem.* **1966**, *241*, 3970; b) H. V. Thulasiram, C. D. Poulter, *J. Am. Chem. Soc.* **2006**, *128*, 15819.

[11] J. W. Cornforth, R. H. Cornforth, C. Donninger, G. Popjak, *Proc. R. Soc. London, Ser. B*, **1966**, *163*, 492.

[12] R. Schwabe, I. Farkas, H. Pfander, *Helv. Chim. Acta* **1988**, *71*, 292.

[13] T. Kato, M. Suzuki, T. Kobayashi, B. P. Moore, *J. Org. Chem.* **1980**, *45*, 1126.

[14] L. C. Tarshis, M. Yan, C. D. Poulter, J. C. Sacchettini, *Biochemistry* **1994**, *33*, 10871.

[15] J. Jiang, X. He, D. E. Cane, *Nat. Chem. Biol.* **2007**, *3*, 711.

[16] D. W. Christianson, *Chem. Rev.* **2017**, *117*, 11570.

[17] a) M. Seemann, G. Z. Zhai, J. W. de Kraker, C. M. Paschall, D. W. Christianson, D. E. Cane, *J. Am. Chem. Soc.* **2002**, *124*, 7681; b) M. Seemann, G. Zhai, K. Umezawa, D. Cane, *J. Am. Chem. Soc.* **1999**, *121*, 591; c) P. Baer, P. Rabe, C. A. Citron, C. C. de Oliveira Mann, N. Kaufmann, M. Groll, J. S. Dickschat, *ChemBioChem* **2014**, *15*, 213; d) P. Baer, P. Rabe, K. Fischer, C. A. Citron, T. A. Klapschinski, M. Groll, J. S. Dickschat, *Angew. Chem. Int. Ed.* **2014**, *53*, 7652.

[18] I. Wahlberg, I. Wallin, C. Narbonne, T. Nishida, C. R. Enzell, *Acta. Chem. Scand.* **1981**, *35b*, 65.

[19] a) P. Rabe, T. Schmitz, J. S. Dickschat, *Beilstein J. Org. Chem.* **2016**, *12*, 1839; b) L. Ding, H. Goerls, K. Dornblut, W. Lin, A. Maier, H.-H. Fiebig, C. Hertweck, *J. Nat. Prod.* **2015**, *78*, 2963.

WILEY-VCH

COMMUNICATION

COMMUNICATION

Two diterpene synthases from *Allokutzneria albata* were investigated for their products, resulting in the identification of the structurally complex compound spiroalbatene and of cembrene A. The enzyme mechanisms were studied in isotopic labelling experiments, by mutations and variations of incubation conditions. For spiroalbatene synthase changed incubation conditions resulted in the formation of thunbergol.
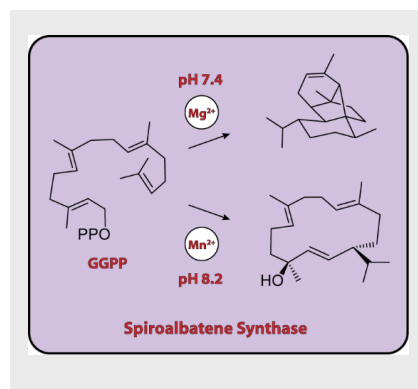
Jan Rinkel, Lukas Lauterbach, Patrick Rabe and Jeroen S. Dickschat*

Page No. – Page No.

Two Diterpene Synthases for Spiroalbatene and Cembrene A from *Allokutzneria albata*