Terpene Biosynthesis

DOI: 10.1002/anie.201209103

Rapid Chemical Characterization of Bacterial Terpene Synthases**

*Patrick Rabe and Jeroen S. Dickschat**

Terpenes make up the largest group of natural products with more than 50000 known compounds. Their skeletons are made by terpene synthases converting linear precursors [geranyl diphosphate (GPP), farnesyl diphosphate (FPP), or geranylgeranyl diphosphate (GGPP)] into (poly)cyclic mono-, sesqui-, or diterpenes.[1] This process is initiated through cation formation by diphosphate abstraction (class I enzymes) or protonation of an olefinic double bond (class II). Successive intramolecular attacks of olefinic double bonds to cationic centers, hydride or proton migrations, Wagner–Meerwein rearrangements, and terminating deprotonations or attack of nucleophiles yield a terpene product. In some cases the initial product may be reionized for a second round of processing.[2]

The active centers of class I enzymes exhibit a highly conserved aspartate-rich motif (DDXXD) and a (N,D)D-(L,I,V)X(S,T)XXXE consensus sequence (NSE/DTE triad) involved in binding of the Mg^{2+} cofactor that complexes the diphosphate for ionization.[3,4] Class II enzymes share a highly conserved DXDD motif for protonation of the substrate.[5,6] In both types of enzymes hydrophobic residues form a cavity in which the oligoprenyl chain is arranged in a product structure predetermining conformation.[4] This pocket also prevents cationic intermediates from the undirected attack of water. Aromatic residues stabilize charged intermediates by cation–π interactions.[7,8]

Although the principal processes of terpene biochemistry are well understood and even a few crystal structures of bacterial terpene synthases are known,[9–11] it is impossible to predict the structure of a terpene from the synthase's amino acid sequence. This results in a dilemma, because the modern sequencing techniques and bioinformatic methods make a large body of information available, showing that several hundreds of terpene synthases are encoded in bacteria, but only a few have been characterized.[12–22] New methods that stay up to date with the accumulating sequencing data are required.

To address this problem we have performed a phylogenetic analysis of bacterial terpene synthases (see Figure 7 in the Supporting Information). Several groups of homologous terpene cyclases were identified, and wherever one enzyme was already chemically characterized, the function of the related enzymes was concluded by analogy.[23] The correctness of this approach was proven by identification of volatile terpenes in headspace extracts. Our analytical data were also used to predict the product of an uncharacterized group of six enzymes (framed box in phylogenetic tree in Figure 7 in the Supporting Information). Three of these enzymes are from bacteria for which the terpenes were analyzed (*Streptomyces flavogriseus* ATCC 33331, *S. griseus* subsp. *griseus* NBRC 13350, and *S. filamentosus* NRRL 15998). All three bacteria produce *epi*-cubenol, thus suggesting that the respective enzymes are *epi*-cubenol synthases (note that no other terpene is commonly produced by these strains and that no other terpene cyclases of these bacteria cluster somewhere else in the tree). This deductive approach fails in case of standalone enzymes in the phylogenetic tree. To circumvent this problem we have developed a method for the rapid and reliable chemical characterization of bacterial terpene synthases based on heterologous expression in *Escherichia coli*. This species was chosen as host, because *E. coli* is able to synthesize FPP and grows fast. *E. coli* can easily be transformed by electroporation, and several vectors and strains optimized for protein expressions are available. Even strains have been designed for efficient readout of codons rarely used by *E. coli*. Altogether, *E. coli* was suggested as ideal system for a rapid analysis of terpene cyclases.

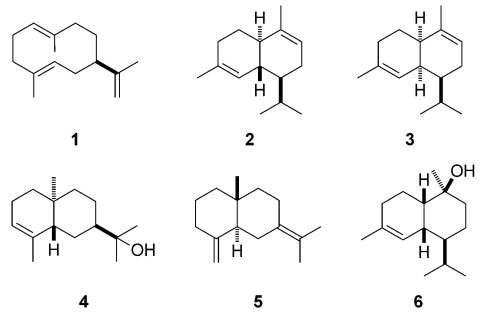
For the identification of the products of six uncharacterized bacterial terpene cyclases the respective genes were amplified by PCR and cloned into the pET28c expression vector, followed by transformation of *E. coli* BL21. After induction of protein expression, the *E. coli* liquid cultures were connected to a closed-loop stripping apparatus (CLSA) for the collection of volatiles on charcoal filters.[23,24] The filters were extracted with CH_2Cl_2 and the obtained extracts were analyzed by GC–MS.

The products of two terpene cyclases from *Chitinophaga pinensis* DSM 2588 were identified. Expression of the first enzyme (accession number YP_003121494) resulted in the production of β-elemene as main compound (Table 1 and Figure 1 in the Supporting Information). β-Elemene is the product of a Cope rearrangement of germacrene A (**1**, Scheme 1) and is formed in a thermal reaction in the GC injection port.[25] Minor amounts of **1** were also detected. Therefore, this terpene cyclase is a germacrene A synthase. Further products were (*E*)-nerolidol, γ-elemene and δ-elemene (the Cope rearrangement products of germacrenes B and C, respectively), (*E*)-β-farnesene, germacrene B, α-selinene, β-selinene, hedycaryol, and germacradien-4-ol. The production of all these terpenes by one enzyme is rationalized via common cationic intermediates (Scheme 1 in the Support-

[*] P. Rabe, Dr. J. S. Dickschat
Institut für Organische Chemie
Technische Universität Braunschweig
Hagenring 30, 38106 Braunschweig (Germany)
E-mail: j.dickschat@tu-bs.de

[**] Funding by the Deutsche Forschungsgemeinschaft DFG with an Emmy Noether fellowship (to J.S.D., DI1536/1-1) and a grant "Duftstoffe aus Actinomyceten" (DI1536/2-1), and by the Beilstein Institut zur Förderung der Chemischen Wissenschaften with a scholarship (to P.R.) is gratefully acknowledged. We thank Stefan Schulz for support of our group.

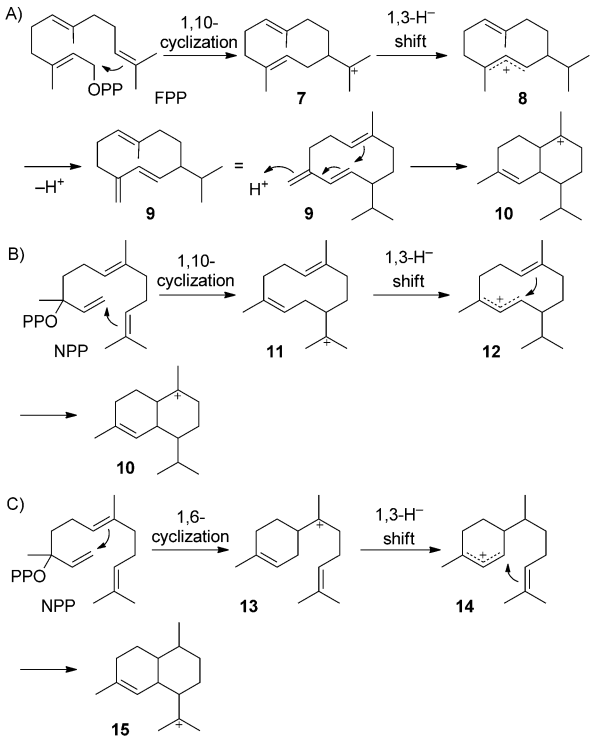
Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201209103.



Scheme 1. Main products of bacterial terpene cyclases.

ing Information). None of the terpenes was detected in headspace extracts of *C. pinensis*, thus suggesting that the germacrene A synthase was not expressed under standard culture conditions.

Expression of a second terpene cyclase from *C. pinensis* (YP_003124367) resulted in γ-cadinene (**2**) along with α- and δ-cadinene, α-muurolene, *cis*-muurola-4(15),5-diene, 10-*epi*-zonarene, α-copaene, (*E*)-β-farnesene, β- and γ-elemene, and germacrene D-4-ol (Table 2 and Figure 2A in the Supporting Information). The biosynthesis of **2** is explainable via the (*E*,*E*)-germacradienyl cation (**7**) and germacrene D (**9**), as suggested for the biosynthesis of cadalanes in *Medicago truncatula* (mechanism A, Scheme 2).[26,27] However, **9** was not identified in *E. coli* headspace extracts. An alternative mechanism is the 1,10-cyclization of nerolidyl diphosphate (NPP) to the (*E*,*Z*)-germacradienyl cation (**11**) followed by a 1,3-hydride shift and cyclization to **10** (mechanism B, Scheme 2). A biosynthetic scheme to all products of the γ-cadinene synthase based on mechanism A is presented in Scheme 2 in the Supporting Information. The main product **2** was also detected in *C. pinensis* (Figure 2B in the Supporting Information).



Scheme 2. Three alternative mechanisms for the biosynthesis of cadalane sesquiterpenes as discussed in the literature.[26–28] OPP = $OP_2O_6^{3-}$.

Two terpene cyclases from *Streptomyces viridochromogenes* DSM 40736 have been investigated. Expression of the first enzyme (ZP_07302078) resulted in α-amorphene (**3**) and traces of γ-amorphene, α- and β-ylangene, α-himachalene, α-cuprenene, β-bisabolene, β-sesquiphellandrene, and (*E*)-β-farnesene (Table 3 and Figure 3A in the Supporting Information). The oxidation products α- and β-calacorene were also found. The production of **3** may proceed by an initial 1,10-ring closure via **9**, but also for the α-amorphene synthase no production of **9** was observed. The formation of monocyclic sesquiterpenes such as β-bisabolene requires isomerization of FPP to NPP, and therefore a biosynthesis through mechanism B may be more likely. A third alternative is the initial 1,6-ring closure of NPP to the bisabolyl cation (**13**) followed by 1,3-hydride shift and cyclization to **15** (mechanism C, Scheme 2), as suggested for the biosynthesis of δ-cadinene in *Gossypium hirsutum*.[28] This would also allow for the formation of compounds such as β-bisabolene, however, the cationic center in **15** is not in the correct position for the formation of **3**. A complex series of hydride migrations from **15** to **10** would be required, but no deprotonation products of such cations were observed. A detailed pathway through mechanism B to all products of the α-amorphene synthase is shown in Scheme 3 in the Supporting Information. α-Amorphene, γ-amorphene, α- and β-ylangene were also found in *S. viridochromogenes* (Figure 3B in the Supporting Information).

The second enzyme from *S. viridochromogenes* (ZP_07308339) was highly selective for 7-*epi*-α-eudesmol (**4**; Table 4 and Figure 4 in the Supporting Information). Minor amounts of 10-*epi*-γ-eudesmol, 5-*epi*-7-*epi*-α-eudesmol, hedycaryol, valerianol, β-dihydroagarofuran, α-selinene, 7-*epi*-α-selinene, (*E*)-β-caryophyllene, (*E*)-β-farnesene, germacrene A, and β-elemene were also detected. All products are made via a 1,10-ring closure to **7**, with the exception of (*E*)-β-caryophyllene, which requires a 1,11-cyclization (Scheme 4 in the Supporting Information). The biosynthesis of β-dihydroagarofuran involves intramolecular attack of a hydroxy function at a tertiary cationic center to establish the tetrahydrofuran ring, while valerianol requires a 1,2-methyl migration. Production of **4** or any side product was not observed in *S. viridochromogenes*.

One sesquiterpene cyclase from *Streptomyces pristinaespiralis* ATCC 25486 (ZP_06911744) revealed a high selectivity for selina-4(15),7(11)-diene (**5**) besides selina-3,7(11)-diene, selina-4(15),6-diene, δ-selinene, germacrene B, β-, γ-, and δ-elemene (Table 5 and Figure 5A in the Supporting Information). The formation of **5** is explained by cyclization of FPP to **7**, deprotonation to germacrene B, reprotonation for a second cyclization step, and final loss of a proton (Scheme 5 in the Supporting Information). The formation of some side products can likewise be explained via germacrene A and C. All sesquiterpenes with exception of selina-

4(15),6-diene were also detected in *S. pristinaespiralis* (Figure 5B in the Supporting Information). Closely related enzymes of the selina-4(15),7(11)-diene synthase are encoded in the genomes of *Streptomyces somaliensis* DSM 40738 (ZP_11214367) and *S. tsukubaensis* NRRL 18488 (ZP_10070539, Figure 7 in the Supporting Information). Investigation of headspace extracts from *S. tsukubaensis* revealed the production of mainly selina-3,7(11)-diene and minor amounts of **5**. Therefore, the enzyme from *S. tsukubaensis* is a selina-3,7(11)-diene synthase. The biosyntheses of **5** and selina-3,7(11)-diene deviate only in the last deprotonation.

The product spectrum of a sesquiterpene cyclase from the chloroflexus *Roseiflexus castenholzii* DSM 13941 (YP_001430766) was relatively simple and only composed of T-muurolol (**6**) along with δ-cadinene, α- and γ-muurolene (Table 6 and Figure 6 in the Supporting Information). All four compounds derive from one cationic species by attack of water or alternative modes of deprotonation (Scheme 6 in the Supporting Information). Extracts of *R. castenholzii* did not contain **6** or the side products.

Expressions of bacterial sesquiterpene cyclases in *E. coli* and direct analysis of the headspace extracts is a valuable and rapid method for the characterization of enzymes with unknown products. The method also works for terpene cyclases that are not active in their native hosts. Laborious and time-consuming purification of the heterologously expressed terpene cyclase and reaction products obtained from incubations are avoided. For unambiguous compound identifications by GC–MS high-quality MS data bases including information about retention indices are required and sufficient. Such libraries are available[29,30] for our method that is suitable to keep pace with the fast developments of DNA sequencing.

Received: November 13, 2012
Published online: January 10, 2013

Keywords: gas chromatography · heterologous expression · mass spectrometry · terpenoids · volatiles

[1] J. S. Dickschat, *Nat. Prod. Rep.* **2011**, *28*, 1917 – 1936.
[2] D. E. Cane, P. C. Prabhakaran, J. S. Oliver, D. B. J. McIlwaine, *J. Am. Chem. Soc.* **1990**, *112*, 3209 – 3210.
[3] M. N. Ashby, P. A. Edwards, *J. Biol. Chem.* **1990**, *265*, 13157 – 13164.
[4] D. W. Christianson, *Chem. Rev.* **2006**, *106*, 3412 – 3442.
[5] K. U. Wendt, K. Poralla, G. E. Schulz, *Science* **1997**, *277*, 1811 – 1815.
[6] C. Feil, R. Süßmuth, G. Jung, K. Poralla, *Eur. J. Biochem.* **1996**, *242*, 51 – 55.
[7] D. A. Dougherty, *Science* **1996**, *271*, 163 – 168.
[8] C. A. Lesburg, J. M. Caruthers, C. M. Paschall, D. W. Christianson, *Curr. Opin. Struct. Biol.* **1998**, *8*, 695 – 703.
[9] C. A. Lesburg, G. Zhai, D. E. Cane, D. W. Christianson, *Science* **1997**, *277*, 1820 – 1824.
[10] J. A. Aaron, X. Lin, D. E. Cane, D. W. Christianson, *Biochemistry* **2010**, *49*, 1787 – 1797.
[11] M. Köksal, W. K. W. Chou, D. E. Cane, D. W. Christianson, *Biochemistry* **2012**, *51*, 3011 – 3020.
[12] D. E. Cane, J. K. Sohng, C. R. Lamberson, S. M. Rudnicki, Z. Wu, M. D. Lloyd, J. S. Oliver, B. R. Hubbard, *Biochemistry* **1994**, *33*, 5846 – 5857.
[13] W. K. W. Chou, I. Fanizza, T. Uchiyama, M. Komatsu, H. Ikeda, D. E. Cane, *J. Am. Chem. Soc.* **2010**, *132*, 8850 – 8851.
[14] Y. Hu, W. K. W. Chou, R. Hopson, D. E. Cane, *Chem. Biol.* **2011**, *18*, 32 – 37.
[15] C. Nakano, S. Horinouchi, Y. Ohnishi, *J. Biol. Chem.* **2011**, *286*, 27980 – 27987.
[16] C. Nakano, F. Kudo, T. Eguchi, Y. Ohnishi, *ChemBioChem* **2011**, *12*, 2271 – 2275.
[17] C. Nakano, H.-K. Kim, Y. Ohnishi, *ChemBioChem* **2011**, *12*, 1988 – 1991.
[18] C. Nakano, M. H.-K. Kim, Y. Ohnishi, *ChemBioChem* **2011**, *12*, 2403 – 2407.
[19] S. A. Agger, F. Lopez-Gallego, T. R. Hoye, C. Schmidt-Dannert, *J. Bacteriol.* **2008**, *190*, 6084 – 6096.
[20] J. Jiang, X. He, D. E. Cane, *Nat. Chem. Biol.* **2007**, *3*, 711 – 715.
[21] a) C.-M. Wang, D. E. Cane, *J. Am. Chem. Soc.* **2008**, *130*, 8908 – 8909; b) M. Komatsu, M. Tsuda, S. Omura, H. Oikawa, H. Ikeda, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 7422 – 7427.
[22] X. Lin, R. Hopson, D. E. Cane, *J. Am. Chem. Soc.* **2006**, *128*, 6022 – 6023.
[23] C. A. Citron, J. Gleitzmann, G. Laurenzano, R. Pukall, J. S. Dickschat, *ChemBioChem* **2012**, *13*, 202 – 214.
[24] K. Grob, F. Zürcher, *J. Chromatogr. A* **1976**, *117*, 285 – 294.
[25] P. J. Teisseire, *Chemistry of Fragrant Substances*, Wiley-VCH, New York, **1994**.
[26] D. Arigoni, *Pure Appl. Chem.* **1975**, *41*, 219 – 245.
[27] S. Garms, T. G. Köllner, W. Boland, *J. Org. Chem.* **2010**, *75*, 5590 – 5600.
[28] J. A. Faraldos, D. J. Miller, V. Gonzalez, Z. Yoosuf-Ali, O. Cascon, A. Li, R. K. Allemann, *J. Am. Chem. Soc.* **2012**, *134*, 5900 – 5908.
[29] D. Joulain, W. A. König, *The Atlas of Spectral Data of Sesquiterpene Hydrocarbons*, E.B.-Verlag, Hamburg, **1998**.
[30] R. P. Adams, *Identification of Essential Oil Components by Gas Chromatography/Mass Spectrometry*, Allured Pub Corp, Carol Stream, **2009**.

