DOI: 10.1002/cbic.201300329

Volatile Terpenes from Actinomycetes: A Biosynthetic Study Correlating Chemical Analyses to Genome Data

Patrick Rabe, Christian A. Citron, and Jeroen S. Dickschat*[a]

The volatile terpenes of 24 actinomycetes whose genomes have been sequenced (or are currently being sequenced) were collected by use of a closed-loop stripping apparatus and identified by GC/MS. The analytical data were compared against a phylogenetic analysis of all 192 currently available sequences of bacterial terpene cyclases (excluding geosmin and 2-methylisoborneol synthases). In addition to the several groups of terpenes with known biosynthetic origin, selinadienes were identified as a large group of biosynthetically related sesquiterpenes that are produced by several streptomycetes. The detection of a large number of previously unrecognised side products of known terpene cyclases proved to be particularly important for an in depth understanding of biosynthetic pathways to known terpenes in actinomycetes. Interpretation of the chemical analytical data in the context of the phylogenetic tree of bacterial terpene cyclases pointed to the function of three new enzymes: (*E*)-β-caryophyllene synthase, selina-3,7(11)-diene synthase and aristolochene synthase.

Introduction

Terpenes make up a large class of thousands of natural products. They can have strong biological activities, as exemplified by the cytotoxic diterpenoid taxol, the antimalarial agent artemisinin, the plant growth hormone gibberellic acid and the fragrant limonene. Historically, terpene production was first recognised in plants; reports of terpenes from bacteria have been comparably scarce. More recently, it has been demonstrated that many bacteria,[1] particularly actinomycetes[2–7] but also myxobacteria[8–11] and cyanobacteria,[12,13] produce a large number of structurally diverse terpenes. In the modern era of natural-product chemistry, key to the discovery of new terpenes from bacteria are the fast and low-cost genome sequencing techniques that have revolutionised the field.

Biosynthetically, terpenes are generated from the linear precursors geranyl diphosphate (GPP, monoterpenes), farnesyl diphosphate (FPP, sesquiterpenes) and geranylgeranyl diphosphate (GGPP, diterpenes) by terpene cyclases. Enzymes in this fascinating class are able to convert the linear precursors in cascade reactions via cationic intermediates into (poly)cyclic terpene hydrocarbons, with the simultaneous introduction of several stereocentres in a highly stereocontrolled fashion. Whereas some terpene cyclases are highly specific and yield one dominating product, other enzymes produce a mixture of compounds. A careful analysis of the products of terpene cyclases demonstrates that even highly specific enzymes frequently produce many trace compounds, which give valuable insights into the cyclisation mechanism.[14] Typically, within such mixtures all enzyme products are structurally closely related, because the initial steps of the cyclisation cascade are the same. At the later stages, the reaction cascade can bifurcate into several branches to yield the main and side products. The first cyclisation step in which stereochemical information is introduced usually determines the absolute configurations of all the products.[15]

Some bacterial terpene cyclases have been characterised for their products by purification of the enzyme and in vitro incubation with the native substrate. These include the pentalenene synthase from *Streptomyces exfoliatus*,[16] the *epi*-isozizaene synthase from *Streptomyces coelicolor*,[17] the avermitilol synthase from *Streptomyces avermitilis*,[18] the synthases for δ-cadinene, T-muurolol, 1,8-cineol and linalool/nerolidol from *Streptomyces clavuligerus*,[19–21] the caryolan-1-ol and *epi*-cubenol synthases from *Streptomyces griseus*,[22,23] the germacradien-4-ol and *epi*-α-bisabolol synthases from *Streptomyces citricolor*,[24] the 8-*epi*-α-selinene and germacrene A synthases from *Nostoc* spp.[25] as well as the geosmin synthases[26–28] and 2-methylisoborneol synthases[29–31] from several organisms. Genome sequencing of thousands of bacterial strains has shown that many more terpene cyclases remain to be chemically characterised; the sequence data of genes encoding terpene cyclases are not useful or interesting on their own. Because the data from genome sequencing are accumulating very fast, new approaches that give rapid insights into the products of bacterial terpene cyclases are required. One successful approach is the analysis of a large number of bacterial strains known to encode terpene cyclases for their production of terpenes, and then referencing the analytical data to a phylogenetic tree of terpene cyclases.[4] We have recently demonstrated a second approach in which bacterial terpene cyclases are chemically characterised by heterologous gene expression in *Escherichia coli*.[32] As an extension to our previous analytical work, here we report the analysis of 24 actinomycetes for their

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

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.201300329.

terpene production and the correlation of the obtained analytical data to a phylogenetic analysis of the available terpene cyclase sequences from all bacterial sequenced genomes.

Results and Discussion

Phylogeny of terpene cyclases and bacterial terpene production

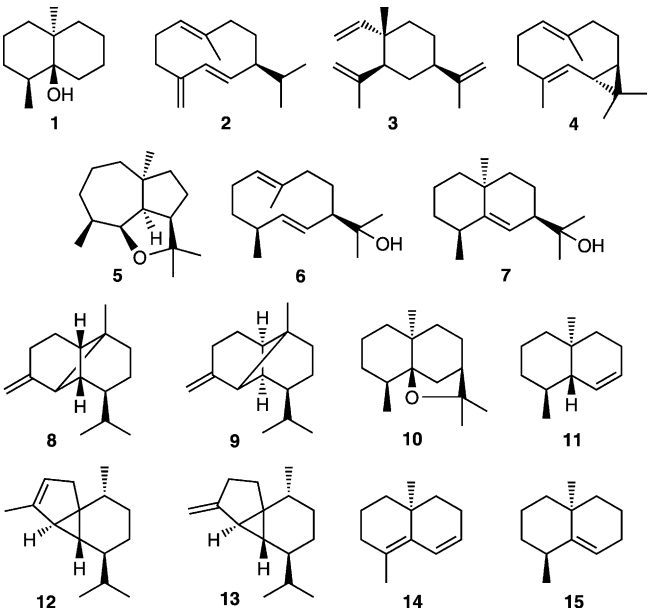
A BLAST search using the amino acid sequence of the pentalenene synthase from *S. exfoliatus* (accession: Q55012) as a probe was performed to identify the terpene cyclases encoded within the genomes of sequenced bacteria. All candidate terpene cyclase sequences were verified by checking for the presence of the highly conserved aspartate-rich motif (DDXXD) and the NSE triad (ND(L,I,V)XSXXXE); only sequences with a minimum of 250 amino acids were included for further analysis. A phylogenetic tree of 192 identified bacterial terpene cyclases (excluding geosmin synthases and 2-methylisoborneol synthases) is shown in Figure S1 of the Supporting Information. Terpene cyclases that have been characterised for the chemical nature of their (main) product are shown in light grey boxes.

During the course of this work 24 actinomycetes of the genera *Streptomyces*, *Saccharomonospora* (here abbreviated *Sm.*, to avoid confusion with *Streptomyces*), *Saccharopolyspora* (*Sp.*) and *Saccharothrix* (*St.*) where analysed for the production of volatile terpenes (Table S1). The volatiles released by agar plate cultures were collected on charcoal filters by using a closed-loop stripping apparatus (CLSA) and analysed by GC/MS. Most strains were grown on two different media (medium 65 and soja flour medium (SFM); for exceptions see Table S1), and each experiment was performed in duplicate to check for reproducibility. The results are summarised in Table S2, and representative chromatograms for each species and medium are shown in Figure S2. As can be seen in the chromatograms for various cases, growth of one organism on different media can give different outcomes of terpene production—a well known phenomenon of bacterial secondary metabolism.[33] The genomes of some of the analysed strains have already been sequenced; others are currently being sequenced. The available information about the terpene cyclases encoded in the genomes of the analysed strains is included in Table S1.

First, the known products of characterised bacterial terpene cyclases that were identified in the bacterial headspace extracts are briefly discussed; this informed the subsequent discussion on the biosynthesis of additional terpenes of unknown biosynthetic origin.

Geosmin: The smell of earth

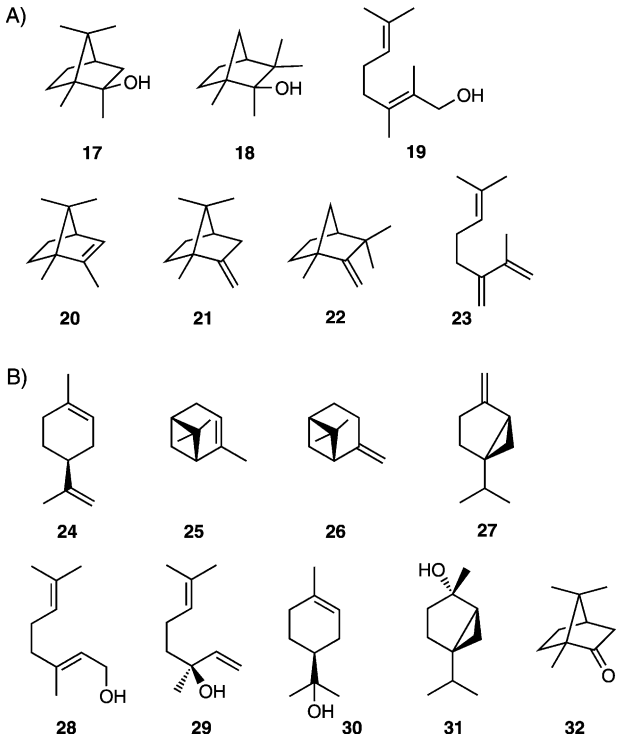
Geosmin (**1**, Scheme 1) was the first identified volatile compound from streptomycetes, and is produced by nearly every member of this genus.[2,4] The compound was also reported in other closely related actinomycetes, in myxobacteria and in cyanobacteria;[1,9,11,12] it is responsible for the typical "earthy" smell, not only of the respective bacterial cultures, but also of soil. A BLAST search revealed the presence of geosmin synthases in more than 100 sequenced bacteria, 13 of which are encoded in the genomes of strains that were included in this study (Table S1). All these strains, apart from *Streptomyces vitaminophilus*, and seven strains whose genome sequences are not yet available were found to emit **1**, frequently observed as the major volatile metabolite (in *Saccharothrix espanaensis*, *Streptomyces antibioticus*, *Streptomyces bottropensis*, *Streptomyces cattleya*, *Streptomyces chartreusis*, *Streptomyces citreus*, *Streptomyces pharetrae*, and *Streptomyces purpureus*). Geosmin biosynthesis has been well investigated,[34–37] and other compounds in Scheme 1 have been assigned as side products of the geosmin biosynthetic pathway.[4,5,8–11,13,26] β-Elemene (**3**) is not directly produced by geosmin synthase, but is known to be formed (according to GC analysis) in a thermal Cope rearrangement from germacrene A.[38] Compound **3** is also found in *Streptomyces filamentosus* and *Streptomyces sudanensis*, two organisms that do not produce geosmin. In these strains germacrene A must be generated by alternative terpene cyclases, as discussed below.



Scheme 1. Geosmin biosynthesis and formation of side products by geosmin synthase. The absolute configurations of all side products were deduced from the known absolute configuration of **1**.

2-Methylisoborneol: A contaminant of drinking water with a musty odour

Like geosmin, the homomonoterpenoid 2-methylisoborneol (2-MIB, **17**, Scheme 2A) was first isolated from a streptomycete,[3] and later found to be produced by various actinomycetes, myxobacteria and cyanobacteria.[1,8,12] The compound has attracted considerable interest as a drinking water contaminant that can, like geosmin, become a serious problem during cyanobacterial blooms in fresh-water reservoirs.[39] Its puzzling biosynthesis was shown to proceed by *S*-adenosylmethionine-dependent methylation of GPP to β-methylgeranyl diphosphate followed by cyclisation to **17**.[8] These two steps are catalysed



Scheme 2. A) 2-Methylisoborneol and homomonoterpenoid side products of the 2-MIB synthase (absolute configurations of side products deduced from the known absolute configuration of **17**). B) Possibly monoterpenoids arising from 2-MIB synthase (relative configurations shown).

by a *C*-methyltransferase and an unusual terpene cyclase, the sequence of which deviates significantly from those of other bacterial terpene cyclases.[29–31,40] As previously described, the genes for *C*-methyltransferase and 2-MIB synthase and a gene for a regulatory protein usually form a three-gene operon.[29–31] The same three genes were also found in four of the sequenced strains in this study (*Saccharopolyspora spinosa*, *Streptomyces acidiscabies*, *Streptomyces collinus*, and *Streptomyces tsukubaensis*); the respective genes are also present, but not clustered, in *St. espanaensis* (Table S1). All five organisms and seven additional strains whose genome sequences are currently not available produced **17** (the main compound in the headspace extracts of *Sp. spinosa*, *S. acidiscabies*, and *Streptomyces rimosus*). Additionally, homomonoterpenoids 2-methyl-β-fenchol (**18**), (*E*)-β-methylgeraniol (**19**), 2-methyl-2-bornene (**20**), 2-methylenebornane (**21**), 1-methylcamphene (**22**) and 2-methylmyrcene (**23**) were occasionally found. These side products were recently identified in other actinomycetes by GC/MS analysis of headspace extracts and synthesis of reference compounds.[14] Monoterpenoids **24**–**32** (Scheme 2B) were only found in species producing exceptionally large quantities of **17** and its side products. These might arise by conversion of GPP instead of β-methylgeranyl diphosphate by the 2-methylisoborneol synthase, as has been discussed previously.[4,14,29]

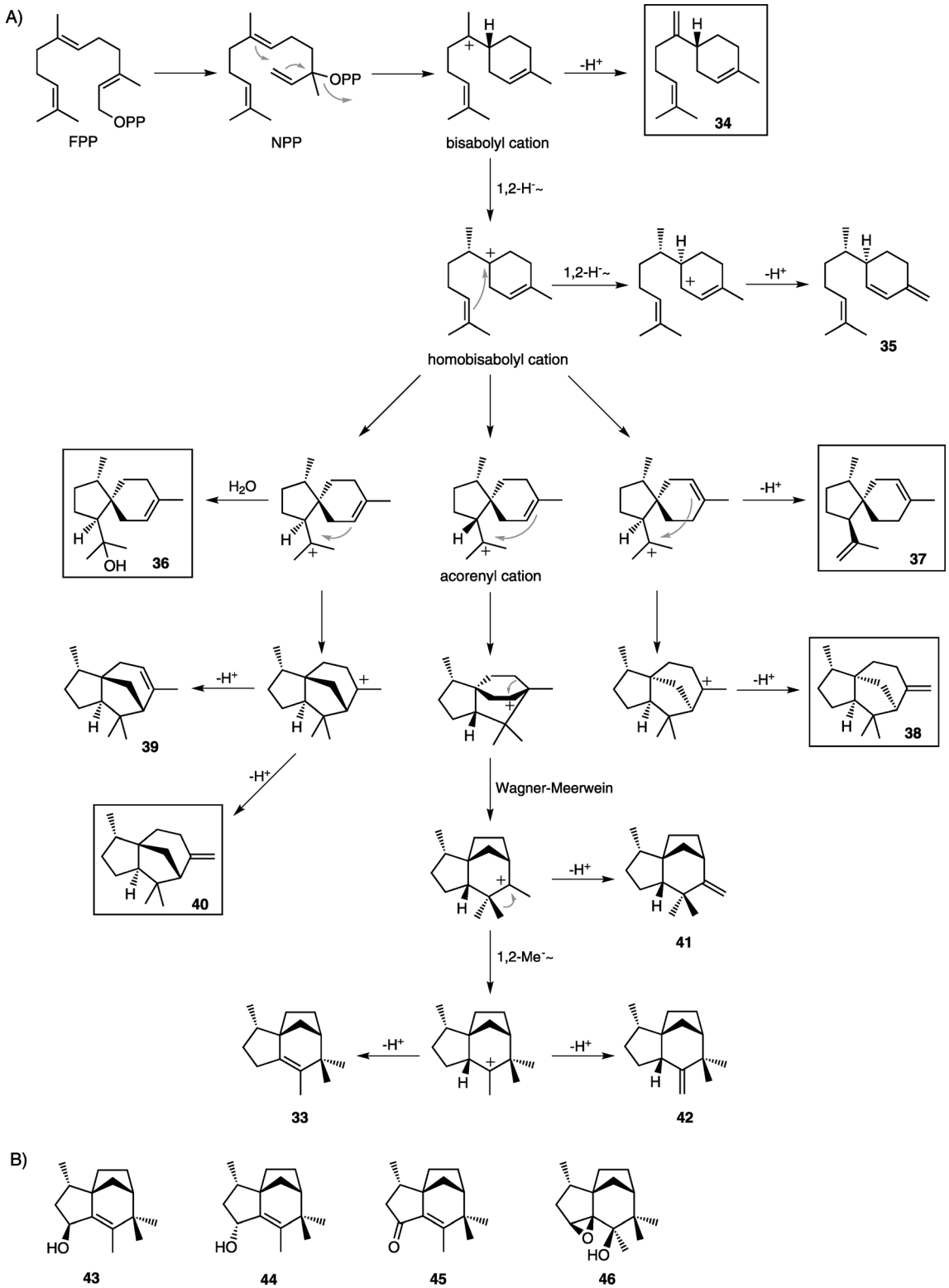
epi-Isozizaene: The precursor to the antibiotic albaflavenone

Another well studied and exceptionally widespread sesquiterpene cyclase in streptomycetes is the *epi*-isozizaene synthase; it was first identified in *S. coelicolor*[17] and is today known to be encoded in many of the sequenced streptomycetes.[4] Genes for *epi*-isozizaene synthases were also found in the genomes of three of the sequenced strains included in this study (*S. acidiscabies*, *S. chartreusis* and *S. collinus*), whereas production of this terpene hydrocarbon was evident in these three and five additional strains (*S. antibioticus*, *Streptomyces arenae*, *S. citreus*, *S. pharetrae* and *Streptomyces xinghaiensis*). The biosynthesis of *epi*-isozizaene (**33**) suggested by Cane proceeds from FPP via nerolidyl diphosphate (NPP), followed by a sequence of cationic intermediates including the bisabolyl, homobisabolyl and acorenyl cations (Scheme 3A). Subsequent cyclisation to a tricyclic secondary cation, its stabilisation by Wagner–Meerwein rearrangement to a tertiary cation, a 1,2-methyl shift and deprotonation yield **33**.[17] In this study, sesquiterpenes β-bisabolene (**34**), β-sesquiphellandrene (**35**), α-acorenol (**36**), β-acoradiene (**37**), β-funebrene (**38**), α-cedrene (**39**), β-cedrene (**40**), *epi*-prezizaene (**41**) and *epi*-zizaene (**42**) were identified solely in strains capable of *epi*-isozizaene biosynthesis, and their formation can easily be explained by bifurcations from the main pathway towards **33**. The only exception is **34** in *Streptomyces flavidovirens*, which does not release **33**, and, therefore, a different explanation for its formation must apply in this case. The absolute configurations of the side products of *epi*-isozizaene synthase can be deduced from the known absolute configuration of **33**. Compounds **35**, **39**, **41** and **42** were previously reported as side products of *epi*-isozizaene synthase.[4,41] Both enantiomers of albaflavenol (**43** and **44**), albaflavenone (**45**) and 4β,5β-epoxy-2-*epi*-zizaan-6β-ol (**46**) are known oxidation products of **33** (Scheme 3B).[42,43] Compounds **34**, **36**, **37**, **38** and **40** have not been reported as side products of the biosynthetic pathway to **33**.

Selinadienes: A new class of sesquiterpenes from streptomycetes

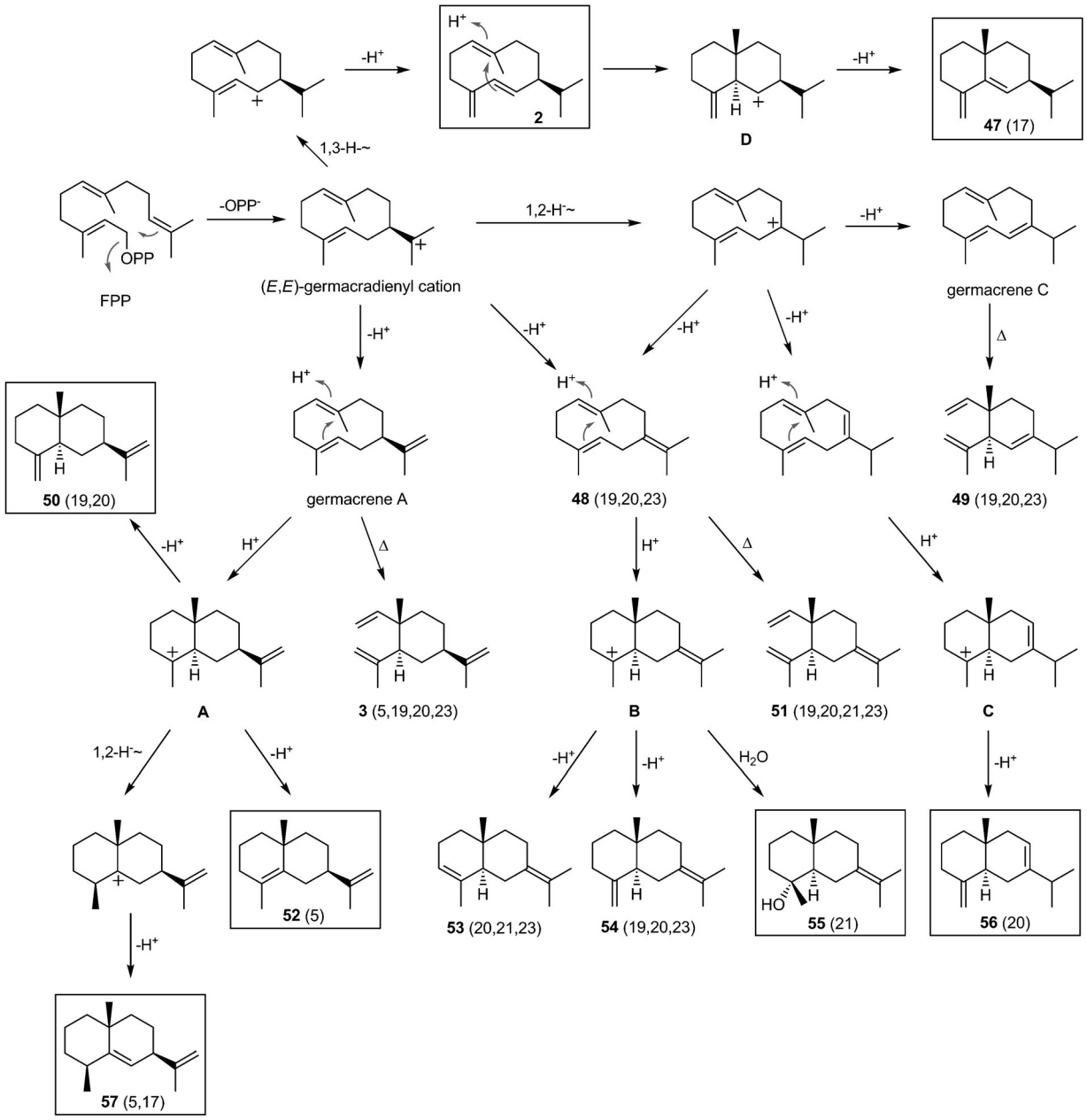
In a recent study from our laboratory, a selina-4(15),7(11)-diene synthase from *Streptomyces pristinaespiralis* was identified by heterologous expression in *E. coli*; this resulted in the production of selina-4(15),7(11)-diene (**54**) and various side products, including germacrene B (**48**), its Cope rearrangement product γ-elemene (**51**), β-elemene (**3**), δ-elemene (**49**, arising from germacrene C), selina-3,7(11)-diene (**53**), δ-selinene and selina-4(15),6-diene.[32] The biosynthesis of these compounds involves an initial cyclisation of FPP to the (*E*,*E*)-germacradienyl cation, followed by conversion to either germacrene A, germacrene B or germacrene C (Scheme 4). Formation of the main product **54** requires additional protonation initiated cyclisation of **48** to cation **B** and subsequent deprotonation.

The analytical work presented here reveals that *S. pristinaespiralis* is not the only streptomycete capable of selinadiene biosynthesis: many other streptomycetes produce closely relat-



Scheme 3. A) *epi*-Isozizaene biosynthesis and formation of side products of *epi*-isozizaene synthase. B) Oxidation products of *epi*-isozizaene. Compounds shown in boxes have not previously been described as side products of the biosynthetic pathway to **33**.

ed sesquiterpenes and encode close relatives of the selina-4(15),7(11)-diene synthase in their genomes (Figure S1). All identified selinadienes arise via germacrenes that are protonated to cyclise to one of the bicyclic cationic intermediates **A**–**D**. In all cases, cyclisation proceeds with the same stereochemical course. For instance, a close homologue of the *S. pristinaespiralis* selina-4(15),7(11)-diene synthase was identified in *S. rimosus* (Figure S1), which also produces **54** as main compound and **3**,

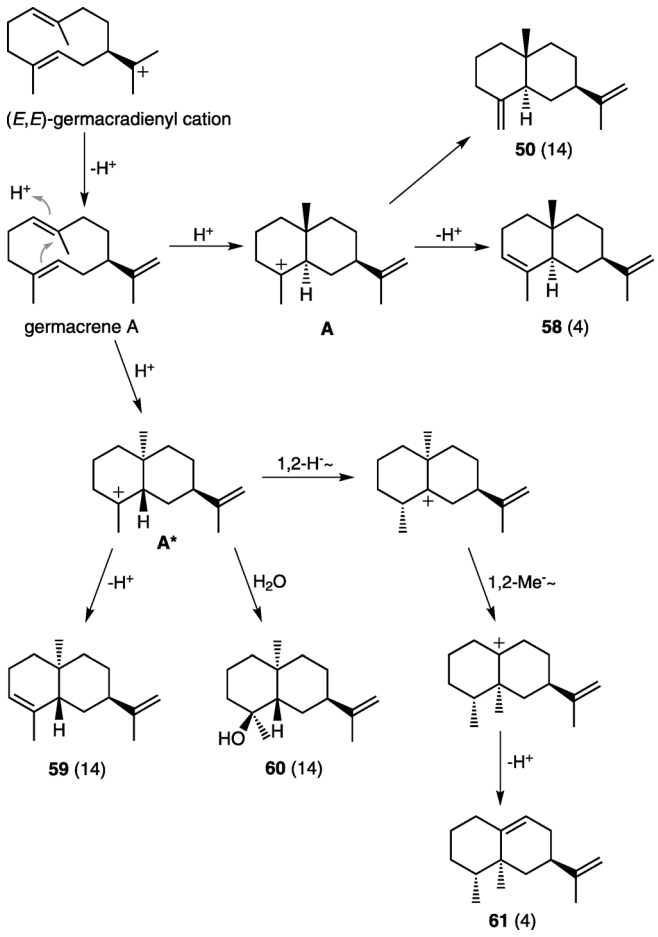


Scheme 4. Biosynthetic pathways to selinadienes (relative configurations shown). Numbers in brackets represent strains producing the particular compound and refer to strain identification numbers in Table S1. Compounds shown in boxes have not been reported for the *S. pristinaespiralis* selina-4(15),7(11)-diene synthase.

48, **49**, **51** and selina-4(15),11-diene (**50**) as side products (compound **50** and all other compounds shown in boxes have not been reported from the selina-4(15),7(11)-diene synthase from *S. pristinaespiralis*). *Streptomyces sudanensis* and *Streptomyces xiamenensis* also released large amounts of **54**, but genetic information about closely related sesquiterpene synthase genes in these species is lacking. Compounds **3**, **48**, **49**, **51** and **53** were observed as side products in both species, whereas **50** and selina-4(15),7-diene (**56**) only occurred in *S. sudanensis*. A slightly more distantly related terpene cyclase is encoded in *S. tsukubaensis*. This organism emits not **54** but **53** as the main product, and juniper camphor (**55**) in minor amounts. These three sesquiterpenes are generated from the same cationic intermediate **B** by alternative deprotonation steps or attack by water. Sesquiterpenes **3**, selina-4,11-diene (**52**) and selina-5,11-diene (**57**) additionally occurred in *S. antibioticus*, and selina-4(15),5-diene (**47**) and **57** were observed in *S. purpureus*. These terpenes can easily be linked to germacrene A or germacrene D (**2**), but neither **53** nor **54** (both formed via germacrene B) were found in these species. Therefore, it remains doubtful as to whether a homologue of the selina-4(15),7(11)-diene synthase or a more distant enzyme is responsible for their biosynthesis.

Aristolochene and β-gurjunene: Sesquiterpenes with a rearranged skeleton

Aristolochene (**61**) was detected in headspace extracts from *S. acidiscabies*. This is the first bacterial aristolochene synthase to be described, although fungal aristolochene synthases from *Penicillium roquefortii* and *Aspergillus terreus* (and many side products of aristolochene biosynthesis in fungi) have been described.[44–49] Its biosynthesis starts similarly to that of the selinadienes, via the (*E*,*E*)-germacradienyl cation and germacrene A (Scheme 5). The subsequent cyclisation to a bicyclic cation **A*** is, however, stereochemically different, and is followed by a cascade of 1,2-hydride migration, rearrangement of the carbon skeleton by a 1,2-methyl shift and deprotonation to yield **61**. α-Selinene (**58**) might be generated in minor amounts by the same terpene cyclase, via cation **A**, but this remains to be elucidated. A candidate enzyme for the biosynthesis of **61** in *S. acidiscabies* (and the only terpene cyclase with unknown function in this organism, at least if sequencing data are complete) is found close to the group of caryolan-1-ol synthases in the phylogenetic tree of bacterial terpene cyclases (Figure S1). In *S. flavidovirens* a sesquiterpenoid intermedeol (**60**) that can also be formed from cation **A*** by attack of water was found. Minor amounts of the deprotonation product 7-*epi*-α-selinene (**59**) and **50** were also present in the headspace extracts.



Scheme 5. Biosynthetic pathway to aristolochene (**61**), β-gurjunene (**62**) and related sesquiterpenes (relative configurations shown). Numbers in brackets represent strains producing the particular compound and refer to strain identification numbers in Table S1.

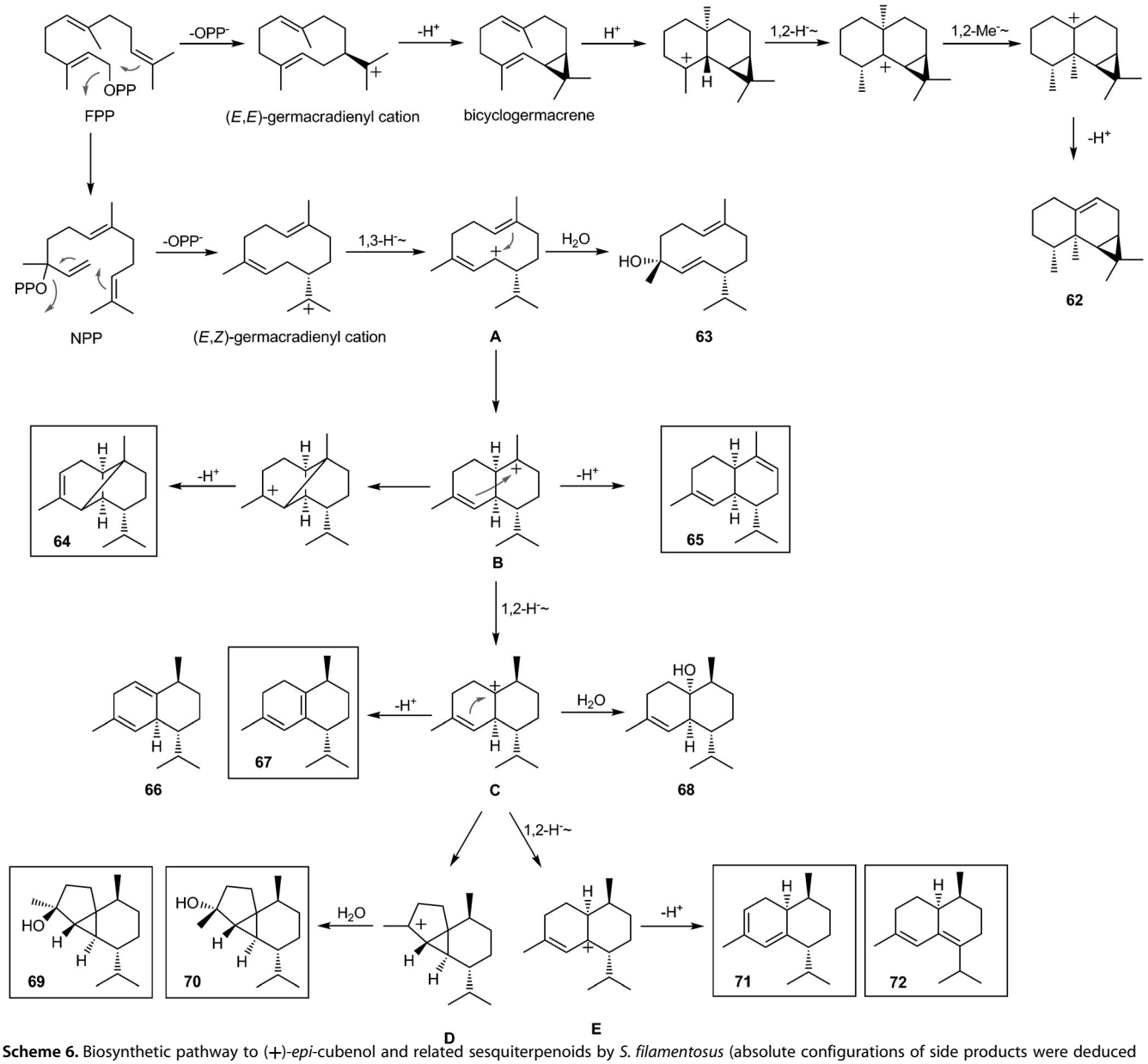
Side products of the (+)-*epi*-cubenol synthase

The *epi*-cubenol synthase from *S. griseus* has recently been characterised,[23] but the precise biosynthetic pathway to *epi*-cubenol (**68**) is the subject of discussion. Cane suggested a pathway via NPP to explain the formation of the *cis*-configured C2=C3 double bond in **68**.[50] A logical alternative was given by Arigoni: germacrene D is a neutral pathway intermediate to cadalane sesquiterpenes.[51] However, germacrene D was not detected in headspace extracts from *S. filamentosus* in this study, although this organism produces large amounts of **68** and encodes a close homologue of the *S. griseus* enzyme in its genome. In addition, a series of sesquiterpenes that can be linked to cationic intermediates en route to *epi*-cubenol was found (Scheme 6 shows Cane's pathway; the formation of all side products can also be explained by Arigoni's pathway). Starting from FPP, the main pathway to **68** proceeds by isomerisation to NPP, cyclisation to the (*E*,*Z*)-germacradienyl cation, a 1,3-hydride shift to cation **A,** ring closure to **B**, 1,2-hydride migration to **C** and final deprotonation.

The formation of germacrene D-4-ol (**63**) involves attack of cation **A** by water; α-copaene (**64**) can arise from **B** by a third cyclisation step and deprotonation. Cation **B** is also the direct precursor of α-muurolene (**65**). Whereas attack by water to **C** yields the main product **68**, different modes of deprotonation give rise to cadina-1,4-diene (**66**) and *trans*-cadina-1(6),4-diene (**67**). Cation **C** can also undergo a third cyclisation to cation **D**, which upon attack by water yields *epi*-cubebol (**69**) and cubebol (**70**), or a 1,2-hydride shift to **E** as the precursor to *trans*-muurola-3,5-diene (**71**) and zonarene (**72**) by deprotonation. Furthermore, *S. filamentosus* was found to release β-gurjunene (**62**). Its biosynthesis proceeds by a very similar sequence of steps as for **61** (Scheme 4), but starting with the formation of bicyclogermacrene instead of germacrene A from the (*E*,*E*)-germacradienyl cation. The side products **62**, **63** and **66** have also been observed in incubation experiments of FPP with the *epi*-cubenol synthase from *S. griseus*.[23] T-muurolol has also been reported, but this compound was not detected in *S. filamentosus* headspace extracts in this work. Compounds **64**, **65**, **67** and **69**–**72** have not previously been reported as side products of *epi*-cubenol biosynthesis (shown in boxes). All these newly identified side products can unambiguously be assigned to *epi*-cubenol biosynthesis, because *S. filamentosus* only encodes an *epi*-cubenol synthase and a caryolan-1-ol synthase. Biosynthesis by the caryolan-1-ol synthase can be excluded, because caryolan-1-ol is structurally unrelated to **68** and its side products.

Side products of (+)-caryolan-1-ol synthase

A second bacterial sesquiterpene cyclase homologue that was recently characterised from *S. griseus* is the caryolan-1-ol synthase.[22] A close homologue is encoded in the genome of *S. fil-*

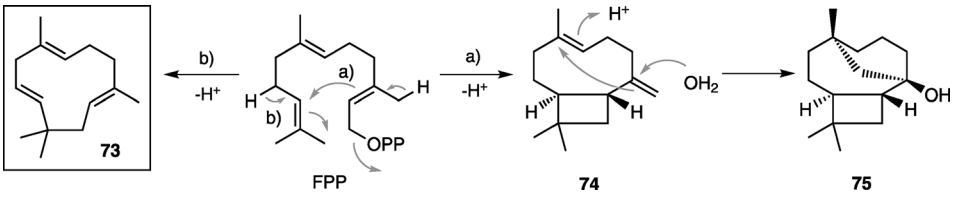


Scheme 6. Biosynthetic pathway to (+)-*epi*-cubenol and related sesquiterpenoids by *S. filamentosus* (absolute configurations of side products were deduced from the known absolute configuration of **68**). Compounds shown in boxes have not previously been reported as side products of *epi*-cubenol biosynthesis.

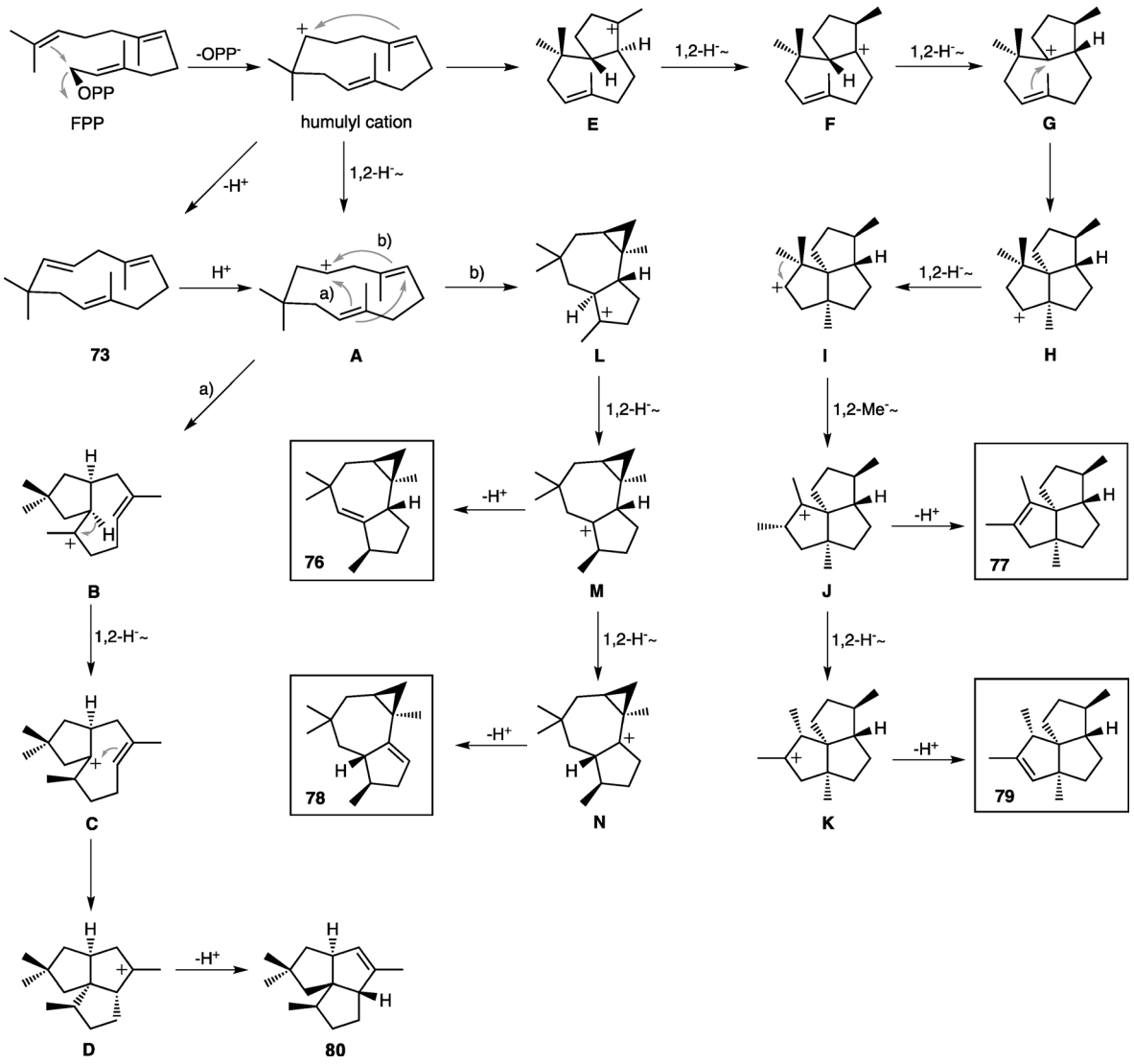
amentosus. One of the major compounds in the headspace extracts of this strain was identified as caryolan-1-ol (**75**), in addition to the biosynthetic intermediate (*E*)-β-caryophyllene (**74**) and the side product α-humulene (**73**; Scheme 7). Compound **74** (but not **73**) was also detected after incubation of FPP with the *S. griseus* enzyme.[22]

Pentalenene and side products of pentalenene synthase

The pentalenene synthase from *S. exfoliatus* is one of the best-studied bacterial terpene synthases.[16,52–54] The homologue from *S. avermitilis* has also been characterised,[55] and genome sequencing has revealed closely related enzymes in *S. arenae*, *Streptomyces bingchenggensis* and *S. collinus*. A more distantly related enzyme is found in *St. espanaensis* (Figure S1). Various reaction cascades from FPP to pentalenene (**80**) have been discussed,[56,57] and one plausible route is shown in Scheme 8. The



Scheme 7. Biosynthetic pathway to (+)-caryolan-1-ol and related sesquiterpenoids by *S. filamentosus* (absolute configuration of **74** and **75** as in ref. [22]). Compound **73** (shown in box) has not previously been reported as side product of caryolan-1-ol biosynthesis.



Scheme 8. A) Biosynthetic pathway to pentalenene (**80**) and side products of pentalenene synthase. B) Biosynthetic pathway to (*E*)-β-caryophyllene (**74**) in *St. espanaensis*. Compounds shown in boxes have not previously been reported as side products of pentalenene biosynthesis.

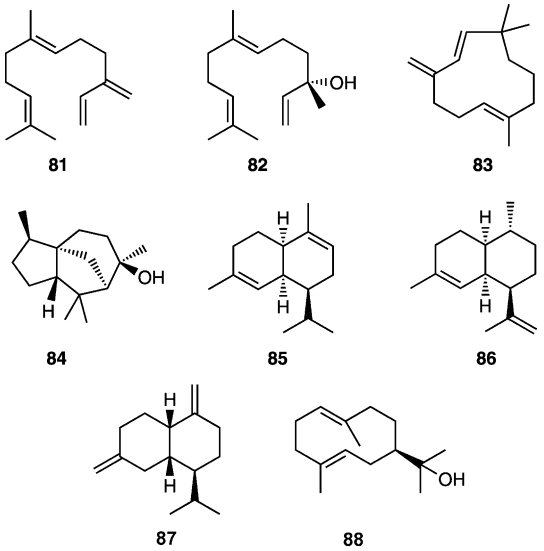
pathway proceeds by cyclisation of FPP to the humulyl cation; this can undergo 1,2-hydride migration to cation **A**, followed by cyclisation to **B** (pathway a), a second 1,2-hydride shift to **C** and cyclisation to **D**; a final deprotonation step yields **80**. In this study *S. arenae* and *S. collinus* were shown to produce **80**, whereas α-humulene (**73**), which can arise from the humulyl cation by deprotonation, was detected in *S. collinus*. Its reprotonation might provide alternative access to **A**. In *S. arenae* headspace extracts, two additional sesquiterpene hydrocarbons with a 5-5-5 tricyclic motif were found: silphiperfol-6-ene (**77**) and silphiperfol-5-ene (**79**). Compound **77** was also present in *S. collinus*, and both compounds were previously observed in *S. avermitilis*.[4] The structural similarity to **80** and frequent production of these compounds by organisms encoding pentalenene synthases suggest their formation by this enzyme. Our suggestion for their biosynthesis starts from the humulyl cation by cyclisation to cation **E**, two consecutive 1,2-hydride shifts via **F** to **G**, and a third cyclisation to tricyclic **H**. Another 1,2-hydride migration to **I** followed by a 1,2-methyl shift to **J** and deprotonation gives rise to **77**. Cation **J** can undergo one more 1,2-hydride shift to **K** and deprotonation to **79**. Furthermore, african-1-ene (**76**) and african-5-ene (**78**) were present in headspace extracts of *S. collinus*. Neither the *epi*-isozizaene pathway nor the geosmin or 2-methylisoborneol pathways of this species deliver a satisfactory explanation for their formation, yet the compounds are accessible by a few steps from cation **A** of the pentalenene pathway. Alternatively, double cyclisation of **A** (pathway b) yields cation **L**, from which 1,2-hydride migration to **M** and deprotonation results in **76**. Another 1,2-hydride shift from **M** gives access to **N** as the precursor of **78** by deprotonation. Compounds **76**–**79** have not been reported as side products of pentalenene synthase before (boxed in Scheme 8).

In addition to a geosmin synthase and 2-methylisoborneol synthase (thus yielding several products of the two pathways as described above), *St. espanaensis* encodes a third terpene

cyclase, of which the closest relatives are the pentalenene synthases. This terpene cyclase might be responsible for the production of (*E*)-β-caryophyllene (**74**), as the initial cyclisation step to this compound is conversion of FPP to the humulyl cation (as for pentalenene). Further proof should be obtained by in vitro incubation of the *St. espanaensis* enzyme with FPP, or by heterologous expression.

Other sesquiterpenes from actinomycetes

A few other sesquiterpenes and sesquiterpene alcohols were occasionally found in headspace extracts of the strains investigated in this study (Scheme 9), but these compounds could



Scheme 9. Further sesquiterpenes and sesquiterpene alcohols identified in the headspace extracts.

not be logically or unambiguously linked to any of the pathways discussed above. These compounds include (*E*)-β-farnesene (**81**) and nerolidol (**82**) in *Streptomyces vitaminophilus*, probably made by the same (but unidentified) terpene cyclase. γ-Humulene (**83**), α-cedrol (**84**), α-amorphene (**85**), amorpha-4,11-diene (**86**), ε-muurolene (**87**) and hedycaryol (**88**) were found in different strains. Completion of genome sequencing efforts might give further insight into their biosynthetic origins.

Conclusions

Terpene cyclases are particularly widespread in actinomycetes. Modern sequencing techniques have revealed 192 bacterial terpene cyclases, but only a few of these have been functionally characterised. Phylogenetic analysis revealed groups of highly homologous terpene cyclases, but also showed that many uncharacterised bacterial terpene cyclases do not have a closely related homologue of known function. Such terpene cyclases will likely catalyse the formation of products other than those of the characterised enzymes, thus making them particularly interesting for further study. Because the function of a terpene cyclase cannot be deduced from its amino acid sequence, characterisation by in vitro incubation experiments with the respective linear precursors (GPP, FPP or GGPP) or by heterologous expression is required. In this study we have shown that in many actinomycetes, groups of structurally closely related terpenes can be found; these can be linked to cationic intermediates along the pathways to the main compounds, thereby resulting in the identification of new side products of the biosynthetic pathways to *epi*-isozizaene, selina-4(15),7(11)-diene, *epi*-cubenol, caryolan-1-ol and pentalenene. Furthermore, by connecting the chemical analytical data to the phylogenetic analysis of bacterial terpene cyclases, three bacterial enzymes are suggested to function as an (*E*)-β-caryophyllene synthase (YP_007039597, *St. espanaensis* DSM 44229), as a selina-3,7(11)-diene synthase (ZP_10070539, *S. tsukubaensis* NRRL 18488), and as an aristolochene synthase (ZP_10457327, *S. acidiscabies* 84-104). The assignment of side products to a known terpene synthase was based on structural similarities between the main and side products, and by the easily explained common formation from the same cationic intermediates along the cyclisation cascade. As this method does not give direct proof for the biosynthetic origin of each terpene (although the assignments presented here give the best explanations and are quite plausible), final proof must be obtained in future studies by using direct methods, such as heterologous expressions and in vitro incubation. Major aims of this work were to identify interesting candidate enzymes for such studies, to lay the groundwork for future biochemical investigations and to give further insights into the complexity of mixtures of terpenes produced by actinomycetes.

Experimental Section

Strains and culture conditions: The bacterial strains used in this study (Table S1) were obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ, Braunschweig, Germany) and the US Department of Agriculture Culture Collection (NRRL collection, Peoria, IL). Liquid precultures were grown for a few days at 28 °C with shaking (160 rpm) in the media listed in Table S1. (Medium 84: rolled oats (20.0 g; boiled in 1 L of distilled water for 20 min, than oats were filtered off) with trace elements (1.0 mL; $FeSO_4 \cdot 7H_2O$ (1 g L^{-1}), $MnCl_4 \cdot 4H_2O$ (1 g L^{-1}), $ZnSO_4 \cdot 7H_2O$ (1 g L^{-1})); other media as in ref. [4].) Agar plates containing the same media were inoculated by plating of liquid cultures (1 mL). After incubation (~2 days, 28 °C), volatiles were collected with a closed-loop stripping apparatus over 18–24 h as described previously.[4] The obtained headspace extract (30 μL) was immediately analysed by GC-MS (1 μL sample) or stored at −80 °C.

GC/MS analysis of headspace extracts: GC-MS analyses were performed with a HP7890A GC system connected to a HP5975C Mass Selective Detector fitted with a HP-5 MS fused silica capillary column (30 m × 0.22 mm, 0.25 μm, Agilent): inlet pressure, 67 kPa; He, 23.3 mL min^{-1}; injection volume, 1 μL; injector, 250 °C; transfer line, 300 °C; electron energy, 70 eV. The GC was programmed as follows: 50 °C (5 min isothermic) increasing (5 °C min^{-1}) to 320 °C, and operated in splitless mode (60 s valve time); carrier gas (He), 1.2 mL min^{-1}. Retention indices were calculated from a homologous series of *n*-alkanes (C_8–C_{32}). Identification of compounds was based

on comparison of mass spectra with database spectra of commercially available libraries and comparison of the measured retention indices with tabulated data (Table S2).

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft with an Emmy Noether grant (to J.S.D., DI1536/1-3), a Heisenberg grant (DI1536/4-1), and the grant "Duftstoffe aus Actinomyceten" (DI1536/2-1), and by the Beilstein Institut zur Förderung der Chemischen Wissenschaften with a scholarship (to P.R.). We thank the US Department of Agriculture for strains Streptomyces cattleya *NRRL 8057,* Streptomyces filamentosus *NRRL 11379,* Streptomyces mobaraensis *NBRC 13819 (=NRRL B-3729),* Streptomyces pharetrae *CZA 14 (=NRRL B-24333),* Streptomyces purpureus *KA281 (=NRRL B-5737),* Streptomyces polyantibioticus *SPR (=NRRL B-24448),* Streptomyces tsukubaensis *NRRL 18488 and* Streptomyces xinghaiensis *S187 (=NRRL B-24674). We also thank Tilmann Weber (Tübingen) for sequences of terpene cyclases from Streptomyces collinus Tü 365, and Stefan Schulz (Braunschweig) for his support.*

Keywords: actinomycetes · biosynthesis · gas chomatography · terpene cyclases · volatiles

[1] S. Schulz, J. S. Dickschat, *Nat. Prod. Rep.* **2007**, *24*, 814 – 842.
[2] N. N. Gerber, H. A. Lechevalier, *Appl. Microbiol.* **1965**, *13*, 935 – 938.
[3] L. L. Medsker, D. Jenkins, J. F. Thomas, C. Koch, *Environ. Sci. Technol.* **1969**, *3*, 476 – 477.
[4] C. A. Citron, J. Gleitzmann, G. Laurenzano, R. Pukall, J. S. Dickschat, *ChemBioChem* **2012**, *13*, 202 – 214.
[5] J. S. Dickschat, T. Martens, T. Brinkhoff, M. Simon, S. Schulz, *Chem. Biodiversity* **2005**, *2*, 837 – 865.
[6] C. E. G. Schöller, H. Gürtler, R. Pedersen, S. Molin, K. Wilkins, *J. Agric. Food Chem.* **2002**, *50*, 2615 – 2621.
[7] K. Wilkins, C. Schöller, *Actinomycetologica* **2009**, *23*, 27 – 33.
[8] J. S. Dickschat, T. Nawrath, V. Thiel, B. Kunze, R. Müller, S. Schulz, *Angew. Chem.* **2007**, *119*, 8436 – 8439; *Angew. Chem. Int. Ed.* **2007**, *46*, 8287 – 8290.
[9] J. S. Dickschat, H. B. Bode, S. C. Wenzel, R. Müller, S. Schulz, *ChemBioChem* **2005**, *6*, 2023 – 2033.
[10] J. S. Dickschat, H. Reichenbach, I. Wagner-Döbler, S. Schulz, *Eur. J. Org. Chem.* **2005**, 4141 – 4153.
[11] J. S. Dickschat, S. C. Wenzel, H. B. Bode, R. Müller, S. Schulz, *ChemBioChem* **2004**, *5*, 778 – 787.
[12] J. A. L. Tabachek, M. Yurkowski, *J. Fish. Res. Board Can.* **1976**, *33*, 25 – 35.
[13] C. Höckelmann, P. G. Becher, S. H. von Reuß, F. Jüttner, *Z. Naturforsch.* **2009**, *64c*, 49 – 55.
[14] N. L. Brock, S. R. Ravella, S. Schulz, J. S. Dickschat, *Angew. Chem.* **2013**, *125*, 2154 – 2158; *Angew. Chem. Int. Ed.* **2013**, *52*, 2100 – 2104.
[15] J. S. Dickschat, *Nat. Prod. Rep.* **2011**, *28*, 1917 – 1936.
[16] 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.
[17] X. Lin, R. Hopson, D. E. Cane, *J. Am. Chem. Soc.* **2006**, *128*, 6022 – 6023.
[18] W. K. W. Chou, I. Fanizza, T. Uchiyama, M. Komatsu, H. Ikeda, D. E. Cane, *J. Am. Chem. Soc.* **2010**, *132*, 8850 – 8851.
[19] Y. Hu, W. K. W. Chou, R. Hopson, D. E. Cane, *Chem. Biol.* **2011**, *18*, 32 – 37.
[20] C. Nakano, M. H.-K. Kim, Y. Ohnishi, *ChemBioChem* **2011**, *12*, 1988 – 1991.
[21] C. Nakano, M. H.-K. Kim, Y. Ohnishi, *ChemBioChem* **2011**, *12*, 2403 – 2407.
[22] C. Nakano, S. Horinouchi, Y. Ohnishi, *J. Biol. Chem.* **2011**, *286*, 27980 – 27987.
[23] C. Nakano, T. Tezuka, S. Horinouchi, Y. Ohnishi, *J. Antibiot.* **2012**, *65*, 551 – 558.
[24] C. Nakano, F. Kudo, T. Eguchi, Y. Ohnishi, *ChemBioChem* **2011**, *12*, 2271 – 2275.
[25] S. A. Agger, F. Lopez-Gallego, T. R. Hoye, C. Schmidt-Dannert, *J. Bacteriol.* **2008**, *190*, 6084 – 6096.
[26] D. E. Cane, R. M. Watt, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 1547 – 1551.
[27] D. E. Cane, X. He, Y. Kobayashi, S. Omura, H. Ikeda, *J. Antibiot.* **2006**, *59*, 471 – 479.
[28] S. Giglio, J. Jiang, C. P. Saint, D. E. Cane, P. T. Monis, *Environ. Sci. Technol.* **2008**, *42*, 8027 – 8032.
[29] C.-M. Wang, D. E. Cane, *J. Am. Chem. Soc.* **2008**, *130*, 8908 – 8909.
[30] M. Komatsu, M. Tsuda, S. Omura, H. Oikawa, H. Ikeda, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 7422 – 7427.
[31] S. Giglio, W. K. W. Chou, H. Ikeda, D. E. Cane, P. T. Monis, *Environ. Sci. Technol.* **2011**, *45*, 992 – 998.
[32] P. Rabe, J. S. Dickschat, *Angew. Chem.* ***2013****, 125,* 1855 – 1857; *Angew. Chem. Int. Ed.* ***2013****, 52,* 1810 – 1812.
[33] H. B. Bode, B. Bethe, R. Höfs, A. Zeeck, *ChemBioChem* **2002**, *3*, 619 – 627.
[34] J. S. Dickschat, H. B. Bode, T. Mahmud, R. Müller, S. Schulz, *J. Org. Chem.* **2005**, *70*, 5174 – 5182.
[35] J. Jiang, X. He, D. E. Cane, *J. Am. Chem. Soc.* **2006**, *128*, 8128 – 8129.
[36] J. Jiang, D. E. Cane, *J. Am. Chem. Soc.* **2008**, *130*, 428 – 429.
[37] T. Nawrath, J. S. Dickschat, R. Müller, J. Jiang, D. E. Cane, S. Schulz, *J. Am. Chem. Soc.* **2008**, *130*, 430 – 431.
[38] P. J. Teisseire, *Chemistry of Fragrant Substances*, Wiley-VCH, New York, **1994**.
[39] G. Izaguirre, C. J. Hwang, S. W. Krasner, M. J. McGuire, *Appl. Environ. Microbiol.* **1982**, *43*, 708 – 714.
[40] W. K. W. Chou, H. Ikeda, D. E. Cane, *Tetrahedron* **2011**, *67*, 6627 – 6632.
[41] X. Lin, D. E. Cane, *J. Am. Chem. Soc.* **2009**, *131*, 6332 – 6333.
[42] B. Zhao, X. Lin, L. Lei, D. C. Lamb, S. L. Kelly, M. R. Waterman, D. E. Cane, *J. Biol. Chem.* **2008**, *283*, 8183 – 8189.
[43] S. Takamatsu, X. Lin, A. Nara, M. Komatsu, D. E. Cane, H. Ikeda, *Microb. Biotechnol.* **2011**, *4*, 184 – 191.
[44] T. M. Hohn, R. D. Plattner, *Arch. Biochem. Biophys.* **1989**, *272*, 137 – 143.
[45] R. H. Proctor, T. M. Hohn, *J. Biol. Chem.* **1993**, *268*, 4543 – 4548.
[46] E. Y. Shishova, F. Yu, D. J. Miller, J. A. Faraldos, Y. Zhao, R. M. Coates, R. K. Allemann, D. E. Cane, D. W. Christianson, *J. Biol. Chem.* **2008**, *283*, 15431 – 15439.
[47] M. J. Calvert, P. R. Ashton, R. K. Allemann, *J. Am. Chem. Soc.* **2002**, *124*, 11636 – 11641.
[48] H. H. Jeleń, *J. Agric. Food Chem.* **2002**, *50*, 6569 – 6574.
[49] N. L. Brock, J. S. Dickschat, *ChemBioChem* **2013**, *14*, 1189 – 1193.
[50] D. E. Cane, M. Tandon, P. C. Prabhakaran, *J. Am. Chem. Soc.* **1993**, *115*, 8103 – 8106.
[51] D. Arigoni, *Pure Appl. Chem.* **1975**, *41*, 219 – 245.
[52] M. Seemann, G. Zhai, K. Umezawa, D. Cane, *J. Am. Chem. Soc.* **1999**, *121*, 591 – 592.
[53] M. Seemann, G. Zhai, J.-W. de Kraker, C. M. Paschall, D. W. Christianson, D. E. Cane, *J. Am. Chem. Soc.* **2002**, *124*, 7681 – 7689.
[54] C. A. Lesburg, G. Zhai, D. E. Cane, D. W. Christianson, *Science* **1997**, *277*, 1820 – 1824.
[55] C. N. Tetzlaff, Z. You, D. E. Cane, S. Takamatsu, S. Omura, H. Ikeda, *Biochemistry* **2006**, *45*, 6179 – 6186.
[56] P. Gutta, D. J. Tantillo, *J. Am. Chem. Soc.* **2006**, *128*, 6172 – 6179.
[57] L. Zu, M. Xu, M. W. Lodewyk, D. E. Cane, R. J. Peters, D. J. Tantillo, *J. Am. Chem. Soc.* **2012**, *134*, 11369 – 11371.

Received: May 22, 2013
Published online on October 8, 2013