

Identification of Sesquiterpene Synthases from *Nostoc punctiforme* PCC 73102 and *Nostoc* sp. Strain PCC 7120^{∇†}

Sean A. Agger,¹ Fernando Lopez-Gallego,¹ Thomas R. Hoye,² and Claudia Schmidt-Dannert^{1*}

Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, 1479 Gortner Avenue, St. Paul, Minnesota 55108,¹ and Department of Chemistry, University of Minnesota, 207 Pleasant Street SE, Minneapolis, Minnesota 55455²

Received 28 May 2008/Accepted 11 July 2008

Cyanobacteria are a rich source of natural products and are known to produce terpenoids. These bacteria are the major source of the musty-smelling terpenes geosmin and 2-methylisoborneol, which are found in many natural water supplies; however, no terpene synthases have been characterized from these organisms to date. Here, we describe the characterization of three sesquiterpene synthases identified in *Nostoc* sp. strain PCC 7120 (terpene synthase NS1) and *Nostoc punctiforme* PCC 73102 (terpene synthases NP1 and NP2). The second terpene synthase in *N. punctiforme* (NP2) is homologous to fusion-type sesquiterpene synthases from *Streptomyces* spp. shown to produce geosmin via an intermediate germacradienol. The enzymes were functionally expressed in *Escherichia coli*, and their terpene products were structurally identified as germacrene A (from NS1), the eudesmadiene 8a-*epi*- α -selinene (from NP1), and germacradienol (from NP2). The product of NP1, 8a-*epi*- α -selinene, so far has been isolated only from termites, in which it functions as a defense compound. Terpene synthases NP1 and NS1 are part of an apparent minicluster that includes a P450 and a putative hybrid two-component protein located downstream of the terpene synthases. Coexpression of P450 genes with their adjacent located terpene synthase genes in *E. coli* demonstrates that the P450 from *Nostoc* sp. can be functionally expressed in *E. coli* when coexpressed with a ferredoxin gene and a ferredoxin reductase gene from *Nostoc* and that the enzyme oxygenates the NS1 terpene product germacrene A. This represents to the best of our knowledge the first example of functional expression of a cyanobacterial P450 in *E. coli*.

Terpenoids constitute the largest group of natural products, constituting over 20,000 described compounds (14). These compounds have a wide range of biological functions and are synthesized by plants and microbes as, for example, pigments, hormones and signaling molecules, antibiotics, antifeedants, or pollinator attractants. Many of these compounds are present in minute quantities and have been an important target for synthetic chemists (22, 52) and metabolic engineers (10, 35, 50).

Terpenes are synthesized from linear isoprene diphosphate precursors with various chain lengths, and their enormous structural diversity is the result of the large number of possible enzyme-catalyzed cyclizations and rearrangements that the double-bond-containing isoprene chains can undergo (14, 49). Cyclizations of isoprene diphosphate chains are catalyzed by terpene synthases (also known as terpene cyclases). Depending on the chain length of their isoprene diphosphate substrates, terpene synthases can be classified as mono-, sesqui-, or diterpene synthases, which catalyze the cyclization of geranyl diphosphate (C₁₀), farnesyl diphosphate (FPP; C₁₅), or geranylgeranyl diphosphate, respectively. Well-known examples of cyclic terpenes are the monoterpene 2-methylisoborneol and sesquiterpene geosmin (responsible for the earthy odor of soil

and lake water), the trichothecenes (mycotoxins), the sesquiterpene artemisinin (antimalarial drug), and the diterpene paclitaxel (Taxol; anticancer drug).

Sesquiterpene synthases catalyze the cyclization of FPP into any of 300 known hydrocarbon skeletons. These enzymes are typically either monomeric or homodimeric with molecular masses ranging from 40 to 65 kDa and require Mg²⁺ for catalysis. The Mg²⁺ ions bind to two metal-binding motifs that are conserved among all terpene synthases. These metal ions complex to the substrate pyrophosphate, thereby promoting the ionization of the leaving group of FPP, resulting in the generation of a highly reactive allylic cation. The enzyme then controls carbocation migration through the isoprene chain with concomitant C-C bond formations and alkyl-shifts to produce a terminal carbocation that is finally quenched by a base (14).

Crystal structures of several sesquiterpene synthases (11, 32, 44) have been solved to aid in the investigation of the complex cyclization mechanisms catalyzed by these enzymes (17, 47). All crystallized sesquiterpene synthases share a conserved class I α -helical terpene synthase fold, but their primary sequences are much less conserved, and sequences often share less than 25% identity at the amino acid level. To date, investigations have mostly focused on plant and fungal sesquiterpene synthases, while only a small number of bacterial enzymes have been reported from *Streptomyces*. Germacradienol/geosmin synthases (for convenience, structures of the most relevant sesquiterpenes mentioned throughout this paper are collected in Fig. 1) in both *Streptomyces avermitilis* and *Streptomyces coelicolor* A3(2) have been described (7, 9, 18). This protein

* Corresponding author. Mailing address: Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, 1479 Gortner Avenue, St. Paul, MN 55108. Phone: (612) 625-5782. Fax: (612) 625-5780. E-mail: schmi232@umn.edu.

† Supplemental material for this article may be found at <http://jbb.asm.org/>.

[∇] Published ahead of print on 25 July 2008.

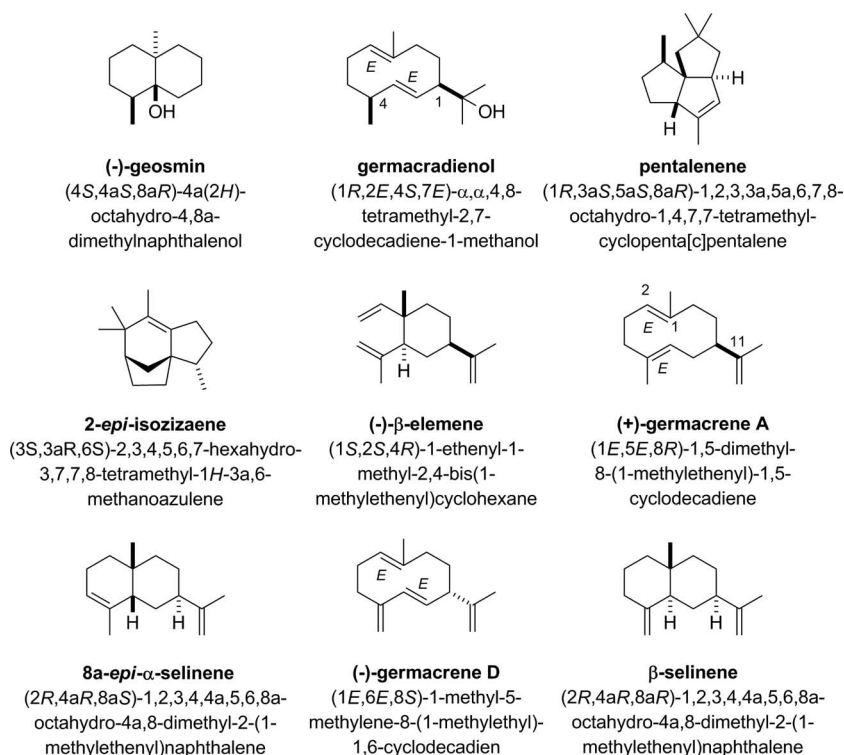


FIG. 1. Structures, common names (in bold), and chemical names in the Chemical Abstracts Service registry for the most relevant sesquiterpenes discussed in the text.

contains two fused complete terpene synthase domains, which result in a bifunctional enzyme. The N-terminal domain of this fusion-type terpene synthase is responsible for catalyzing the formation of germacradienol. Germacradienol is released from the N-terminal active site and becomes the substrate of the C-terminal active site to produce geosmin. Geosmin has an earthy, musty odor and has been implicated in the contamination of water supplies, agricultural products, and wine. Conversion of germacradienol to geosmin under *in vitro* conditions, however, is inefficient, resulting in the formation of only 8 to 15% geosmin among the total terpene products (26). Two other typical single-domain sesquiterpene synthases, pentalenene synthase and *epi*-isozizaene synthase, have also been characterized from *Streptomyces* strains (33, 51). Pentalenene synthase, cloned from both *Streptomyces* sp. strain UC5319 and *S. avermitilis*, has been studied the most extensively and its mechanistic details are well understood. *epi*-Isozizaene synthase has been recently cloned from *S. coelicolor* A3(2) as part of a genome mining effort and shown to make a previously unknown sesquiterpene (33). Very recently, genes involved in the biosynthesis of the monoterpene alcohol 2-methylisobornoneol have been identified in actinomycetes (28).

Actinomycetes like *Streptomyces* are known to produce a range of bioactive natural products. Cyanobacteria are another phylum of bacteria that are typically associated with natural product biosynthesis. These photosynthetic bacteria are found in nearly every habitat imaginable and produce numerous secondary metabolites, including polyketides, non-ribosomal peptides, and terpenes. To date, except for the

biosynthesis of carotenes, little is known about the biosynthesis of other terpenes in cyanobacteria. Cyanobacteria are the major source of the musty smelling and tasting cyclic terpene geosmin and 2-methylisoborneol that are found in many natural water supplies and which are difficult to remove by conventional water treatment methods (31). Two terpene synthase sequences with homology to the *Streptomyces* germacradienol/geosmin synthase sequence have recently been amplified (but not functionally characterized) from a cyanobacterium implicated as the main producer of geosmin in a Saxonian water reservoir (34).

In this study, we have identified three sesquiterpene synthases from *Nostoc punctiforme* PCC 73102 and *Nostoc* sp. strain PCC 7120 (hereafter, *N. punctiforme* and *Nostoc* sp. refer to these two strains, respectively). One of the identified putative terpene synthases (NP2) from *N. punctiforme* shows homology to germacradienol/geosmin synthase from *Streptomyces*, while the other two terpene synthase homologs found in *Nostoc* sp. (NS1) and *Nostoc punctiforme* (NP1) are typical single-domain terpene synthases. Genes encoding the NS1 and NP1 enzymes are located in an apparent gene cluster containing open reading frames (ORFs) encoding a cytochrome P450 and a putative hybrid two-component protein.

MATERIALS AND METHODS

Strains and growth conditions. *Nostoc* sp. strain PCC 7120 (ATCC 27893) was obtained from the ATCC, and *Nostoc punctiforme* PCC 73102 was kindly provided by Jack Meeks at the University of California, Davis, CA. Cyanobacteria were grown at room temperature under fluorescent lights for two months in BG11 media for blue-green algae (cyanobacteria; ATCC medium 616). Cloning

and investigations of in vivo sesquiterpene production were carried out in *Escherichia coli* strain JM109 or BL21(DE3). *E. coli* cultures were grown in Luria-Bertani (LB) medium supplemented with appropriate antibiotics, i.e., ampicillin (100 $\mu\text{g ml}^{-1}$), chloramphenicol (50 $\mu\text{g ml}^{-1}$), kanamycin (30 $\mu\text{g ml}^{-1}$), or streptomycin (30 $\mu\text{g ml}^{-1}$) at 30°C and 250 rpm as indicated.

Gene cloning. Homology searches were performed using the NCBI BLAST software based on the pentalene synthase protein sequence from *Streptomyces* sp. strain UC5319 (accession no. AAA19131). Genomic DNA was isolated from both *Nostoc* sp. strain PCC 7120 and *N. punctiforme* by using standard phenol-chloroform DNA extraction techniques described by Sambrook and Russell (45). Putative terpene synthase genes and P450 genes were amplified by PCR with *Vent* (New England Biolabs, Ipswich, MA) polymerase by using gene-specific primers with added restriction sites. The obtained PCR products were digested with the appropriate restriction enzymes, gel purified, and ligated into pUCmodRBS (for terpene synthase genes) or pET-21b(+) (P450 genes) as described in Table S1 in the supplemental material. Cloned genes were verified by DNA sequencing. Similarly, ferredoxin:NADPH reductase and ferredoxin genes from *Nostoc* sp. strain PCC 7120 and flavodoxin:NADPH reductase and flavodoxin genes from *E. coli* were amplified from genomic DNA and cloned into pUCmodRBS prior to subcloning into pBBR for coexpression of reductase and ferredoxin/flavodoxin pairs. Terpene synthases NS1 and NP1 together with their constitutive *lac* promoter were subcloned from their pUCmodRBS parents into pCDFDuet for coexpression with P450 genes on pET-21b(+) and reductase proteins on pBBR plasmids (see Table S1 in the supplemental material). In addition, NP1 was subcloned in a similar fashion into pACmod, as this plasmid is better suited for high-density fermentation.

Protein expression and purification. For protein purification of terpene synthases, genes were subcloned from pUCmodRBS into pET-21b(+), and a C-terminal six-histidine tag was added in the process. Recombinant *E. coli* BL21(DE3) overnight cultures were used to inoculate 50-ml cultures that were grown at 30°C and 240 rpm until an optical density at 600 nm (OD_{600}) of 0.6, at which point the cultures were cooled on ice and induced with galactose (1 g liter⁻¹) before cultivation was continued overnight at 18°C. The cells were harvested by centrifugation and stored at -20°C until they were used. Cells were resuspended in lysis buffer (10 mM Tris-HCl, 10 mM MgCl₂, 1 mM β -mercaptoethanol buffer, pH 8) and sonicated. Cell debris was cleared by centrifugation (14,000 rpm, 5 min, 4°C). The cleared protein extract was used to purify proteins by metal-affinity chromatography.

Soluble protein was loaded onto a Talon column (Invitrogen, Carlsbad, CA) equilibrated with 10 mM Tris-HCl buffer containing 10 mM MgCl₂ and 1 mM β -mercaptoethanol, pH 8. In the case of NS1, the binding was carried out in the presence of 10 mM imidazole to prevent nonspecific absorption. Following binding, the column was washed three times with 20 mM imidazole. Finally, the protein was eluted with 300 mM imidazole. In the case of NP1, the protein was eluted with 50 mM imidazole. Protein concentrations were determined using Bradford reagent (Bio-Rad, Hercules, CA).

Kinetic parameters. The k_{cat} and K_m for NS1 and NP1 were determined by incubating purified enzyme with various concentrations of 1 to 100 μM FPP. Terpene synthase activity was determined by monitoring the decrease in absorbance at 340 nm as a consequence of the consumption of NADPH coupled to the release of pyrophosphate (PPi). Pyrophosphate was detected using a coupled-enzyme system (containing PPi-dependent fructose-6-phosphate, aldolase, triosephosphate isomerase, and α -glycerophosphate dehydrogenase) supplied as a pyrophosphate reagent by Sigma-Aldrich (product number P7275). This reagent was reconstituted in buffer (16.7 mg in 1 ml 10 mM Tris-HCl, 10 mM MgCl₂, 1 mM β -mercaptoethanol, pH 8). Assay reaction mixtures were prepared in 96-well microplates with 50 μl of pyrophosphate reagent, 90 μl of buffer (10 mM Tris-HCl, 10 mM MgCl₂, 1 mM β -mercaptoethanol, pH 8) and 10 μl of various concentrations of FPP. Similarly, a blank reaction mixture without FPP (instead, 10 μl of buffer was used) was prepared, and both reaction mixtures were preheated at 30°C for 5 min. Next, 5 μl of enzyme (0.2 mg ml⁻¹) was added to both mixtures to start the reactions. The activity was determined as the difference between the decrease in absorbance per minute of the sample and of the blank. Using an extinction coefficient (ϵ_{340}) for NADPH of $6.22 \times 10^3 \text{ M}^{-1} \text{ cm}^{-1}$, one unit of activity was defined as the amount of enzyme needed to release 1 μmol of PPi, inducing the consumption of 2 μmol of NADPH. The K_m and V_{max} values were determined by plotting the reciprocal of the enzymatic reaction velocity versus the reciprocal of the substrate concentration.

Analysis of sesquiterpenes produced by recombinant *E. coli*. Fifty milliliters of LB medium supplemented with ampicillin was inoculated with 1 ml of an overnight culture of *E. coli* transformants harboring the putative sesquiterpene synthases. Culture flasks were sealed with several layers of aluminum foil and grown at 30°C with shaking for 12 h before sampling of the culture headspace for 30 min

by solid-phase microextraction (SPME) by using a 100- μm polydimethylsiloxane fiber (Supelco, Bellefonte, PA). The fiber was then inserted through the tin foil into the flask's headspace (gas phase) and then, after absorption, inserted into the injection port of a gas chromatography (GC) mass spectrometer (MS) for thermal desorption. To detect compounds in the culture medium, a 65- μm polydimethylsiloxane-divinylbenzene fiber (Supelco, Bellefonte, PA) was used to extract the culture supernatant for 4 h.

GC-MS analysis. GC-MS analysis was carried out on a Varian 3800 gas chromatograph coupled to an ion-trap mass spectrometer (Saturn, Palo Alto, CA). Separation was carried out on a DB5 capillary column (length, 30 m; inner diameter, 0.25 mm; film thickness, 1.5 μm) with an injection port temperature of 250°C and helium as a carrier gas. Mass spectra were recorded in an electron impact ionization mode. Compounds bound to SPME fibers were desorbed for 5 min in the injection port. The temperature program started at 100°C and was ramped at 20°C min⁻¹ to a final oven temperature of 300°C. Mass spectra were scanned in the range of 5 to 300 atomic mass units (amu) at 1-s intervals.

High-yield capture of NP1 sesquiterpene product from *E. coli* cultures. A seed culture was started from a single colony of recombinant *E. coli* JM109/pAC-NP1 and grown in 100 ml of LB supplemented with chloramphenicol overnight at 30°C to an OD_{600} of 3.77. The seed culture was used to inoculate a 5-liter (3-liter working volume) bioreactor (Minifors, Boottmtingen, Switzerland). The production media contained 25 g liter⁻¹ of K₂HPO₄, 5 g liter⁻¹ (NH₂)₂SO₄, 2.5 g liter⁻¹ Na₃ citrate, 250 mg liter⁻¹ FeSO₄, 37 mg liter⁻¹ MnSO₄, 2.5 g liter⁻¹ MgSO₄, 62 mg liter⁻¹ CaCl₂, 25 g liter⁻¹ glycerol, and 50 $\mu\text{g ml}^{-1}$ chloramphenicol. The fermentation conditions were as follows: an airflow of 3 liters min⁻¹, vessel pressure of 2.0 lb/in², 37°C temperature, and agitation at 300 rpm. The pH was set at 6.8 and controlled with 30% ammonium hydroxide. Dissolved oxygen was set at 30% and controlled with both agitation and oxygen supplementation. The fermentation was run overnight with a continuous feed of glycerol (500 g glycerol used in total) to reach a final OD_{600} of 123. The off gas of the bioreactor was channeled through a glass column packed with Supelpak-2SV resin (Supelco, Bellefonte, PA). The resin was removed from the column and rinsed with 50 ml of pentane to elute the captured sesquiterpenes. The column captured approximately 1 mg of sesquiterpenes per fermentation run.

Distillation and preparative HPLC. The pentane eluate of captured volatile sesquiterpene from the fermentation was concentrated by distillation prior to preparative high-pressure liquid chromatography (HPLC). Approximately 50 ml of the pentane solution was reduced in volume to ca. 3 ml by distillation through an eight-inch Vigreux column with a pot temperature of 55°C, at which point a microdistillation column was used to further reduce the volume to 800 μl . Terpenes were then resolved on an 1100 HPLC system (Agilent Technologies, Palo Alto, CA) equipped with a photodiode array detector set to 300 nm. Terpene samples (eight 100- μl injections) were loaded onto a reverse-phase Zorbax SB-C18 column (length, 25 cm; diameter, 9.4 mm; Agilent Technologies) and eluted under isocratic conditions with 10% water and 90% methanol at a flow rate of 1 ml min⁻¹. Compounds corresponding to peaks of interest were collected and mixed with an equal volume of water before being loaded onto an octadecyl C₁₈ disposable column (Bakerbond, Phillipsburg, NJ). The column was rinsed with water to remove residual methanol and then dried at an ambient temperature overnight. The resin was removed from the plastic column (in order to prevent leaching of polymers from the plastic into the CDCl₃) and repacked in a glass pasture pipette, and sesquiterpenes were eluted with 100 μl of CDCl₃.

Preparative GC. Sesquiterpenes obtained after preparative HPLC were further purified by preparative GC on a gas chromatograph (6890 series; Agilent, Santa Clara, CA) interfaced with a cold injection system and a preparative fraction collector (PFC) (Gerstel, Baltimore, MD). The gas chromatograph was run in a constant-flow mode with a column flow of 26.5 ml helium per min. Twenty 50- μl samples were injected at 0°C and in a splitless mode (6 min). The cold injection system was then immediately heated to 50°C at 10°C per s, kept at 50°C for 2 min, and heated to 240°C at 12°C per s. Compounds were separated on a DB-1 column (length, 30 m; inner diameter, 0.53 mm; film thickness, 5 μm) from J&W (Folsom, CA) by using a program that ramped the temperature from 40°C (for 1 min) to 250°C at a rate of 25°C per min. The end of the column was split into two lines, one short line (length, 5 cm; inner diameter, 0.05 mm) to the flame ionization detector and one transfer line (length, 30 cm; inner diameter, 0.32 mm) to the PFC. The transfer line to the PFC and the switch device were kept at 250°C. Traps were kept at -60°C. The PFC was set to collect the fraction eluting between 6.6 and 7.5 min. Collected terpenes were eluted from the trap with 0.3 ml of CDCl₃ (400 μg purified terpene) for nuclear magnetic resonance (NMR) analysis.

NMR spectroscopy. ¹H NMR measurements were carried out on a Varian 600-MHz NMR spectrometer with a triple-resonance probe. The software was VNMRj 1.1D. One- and two-dimensional experiments (correlation spectros-

copy), heteronuclear single-quantum coherence (HSQC), heteronuclear multi-quantum coherence (HMOC), heteronuclear multibond correlation (HMBC) spectroscopy, and total correlated spectroscopy were performed at an ambient temperature using a Shigemi tube.

Analysis of oxygenated sesquiterpenes produced by recombinant *E. coli*. *E. coli* transformants coexpressing terpene synthase (with pCDFDuet-NP1 or pCDFDuet-NS1), P450 (with pET-P450NP1 or pET-P450NS1), reductase proteins (with pBBR-FER-RED), and rare tRNAs (with pRARE) were grown in 50-ml LB cultures at 30°C with shaking (250 rpm) to an OD₆₀₀ of 0.6, at which time galactose (1 g liter⁻¹ final concentration) was added to induce P450 gene expression. Following induction, the cultivation was continued for 12 h at 18°C. Sesquiterpenes in the culture medium and headspace were captured by SPME and analyzed by GC-MS as described above.

RESULTS

Cyanobacterial sesquiterpene synthase homologs. A BLAST analysis of published prokaryotic sequences from NCBI and Joint Genome Institute with the experimentally characterized pentalenene synthase from *Streptomyces* sp. strain UC5319 (accession no. Q55012) (8) identified putative sesquiterpene synthases in a number of bacterial sequences (Fig. 2). Sesquiterpene synthase gene homologs can be found in the genomes of several of actinobacteria and proteobacteria (delta, beta, and gamma subgroups). Two of the four *Chloroflexii* genera (*Roseiflexus* and *Herpetosiphon*) for which genome sequences have been published also contain sesquiterpene synthase homologs. However, the genome sequences of only two genera of cyanobacteria (*Nostoc* and *Anabaena*) out of more than a dozen genera with published genome sequences contain sesquiterpene synthase homologs (Fig. 2). Several of the terpene synthase homologs share homology with the bifunctional germacadienol/geosmin synthases from *Streptomyces* (7, 26), and a number of bacterial species contain several paralogs of either the typical single-domain or bifunctional fusion-type sesquiterpene synthase (Fig. 2; see Fig. S1 in the supplemental material). Interestingly, *Herpetosiphon aurantiacus* has three sesquiterpene synthase gene paralogs, of which two (corresponding to ZP_01425931 and ZP_01425930) are located next to one another, suggesting a gene duplication event similar to the duplication event that after transcriptional fusion may have led to the creation of the bifunctional germacadienol/geosmin synthases.

The closely related free-living freshwater cyanobacteria *Nostoc* sp. (previously *Anabaena* sp.) and *Anabaena variabilis* ATCC 29413 each have one putative single-domain terpene synthase, while two putative terpene synthases (a single-domain synthase and a fusion-type terpene synthase) are present in the plant and fungal symbiont *N. punctiforme* PCC 73102 (48). The single-domain terpene synthase gene homologs in *Nostoc* and *Anabaena* are located together with a cytochrome P450 gene and a large putative hybrid two-component protein gene in an apparent gene cluster not found in other sequenced bacterial genomes (Fig. 3). ORFs flanking the putative terpene synthase mini-gene cluster differ in the three cyanobacterial genomes, suggesting functional significance of this cluster of genes. The fusion-type terpene synthase homolog in *N. punctiforme* shares a greater than 50% sequence identity (see Fig. S1 in the supplemental material) with the germacadienol/geosmin synthases from *Streptomyces*, suggesting that it may be involved in geosmin biosynthesis in *N. punctiforme*. Analysis of *N. punctiforme* and *Nostoc* sp. cultures that were grown for up

to two months with illumination revealed, as expected, that geosmin was produced only by *N. punctiforme* and not by *Nostoc* sp. cultures (see Fig. S2 in the supplemental material).

For further analysis of cyanobacterial sesquiterpene biosynthesis, terpene synthase homologs identified in the free-living *Nostoc* sp. (single-domain enzyme named NS1) and the symbiont *N. punctiforme* (single-domain enzyme named NP1 and fusion-type terpene synthase named NP2) were selected for cloning and characterization. In addition, cytochrome P450 homologs located adjacent to terpene synthase homologs NS1 (P450NS) and NP1 (P450NP) were targeted for cloning and characterization.

Cloning and expression of single-domain sesquiterpene synthases NS1 and NP1 in *E. coli*. To determine the terpene products of NS1 and NP1, the terpene synthase genes were amplified from genomic DNA and cloned into the expression vector pUCmodRBS. The resulting plasmids (pUC-NS1 and pUC-NP1) were transformed into *E. coli* JM109, and the headspaces of 50-ml cultures in log phase of growth were analyzed by SPME and GC-MS for the presence of terpene products.

The culture headspace of *E. coli* cells expressing the *Nostoc* sp. strain PCC 7120 terpene synthase NS1 contained one major new product along with a minor compound. The major volatile compound had a mass fragmentation pattern typical for a sesquiterpene, with a parent ion with an *m/z* of 204 and characteristic daughter ions with *m/z* values of 189, 147, 107, 93, and 81 (Fig. 4A). Comparison of this spectrum with those of published reference terpene spectra in the terpene library of MassFinder (software version 3) (29) and in the National Institute of Standards and Technology (NIST) MS database allowed us to identify this compound as β -elemene (94% match of the spectrum with the reference spectrum in terpene library). The minor compound was identified as germacrene A (90% match of the spectrum with the reference spectrum in the terpene library). Germacrene A is a thermally unstable compound, which under the high temperature of the GC injection port is known to undergo a Cope rearrangement, forming β -elemene (1, 16). As expected, when the injection port temperature was set at 250°C or higher, 98% of the final product was β -elemene, and conversely, reduction of the injection port temperature increased the germacrene A peak relative to the β -elemene peak (see Fig. S3 in the supplemental material).

Expression of the *N. punctiforme* terpene synthase NP1 in *E. coli* resulted in the accumulation of one major product along with several minor products in the culture headspace (Fig. 4B). The parent ion with an *m/z* of 204 and the mass fragmentation pattern of this major product confirmed it to be a sesquiterpene hydrocarbon. Moreover, its precise mass of 204.1876 amu was fully supportive of a molecular formula of C₁₅H₂₄ (theoretically, 204.1873 amu). However, comparison of retention time and mass fragments of this compound with reference data in the MassFinder and NIST database did not result in any match that would allow a structural prediction of this terpene with a high confidence.

Structural identification of the NP1 sesquiterpene product. The unidentified sesquiterpene synthesized by NP1 was then characterized by NMR spectroscopy. To obtain enough sesquiterpene for NMR analysis, large-scale in vitro preparations were first attempted with isolated recombinant terpene syn-

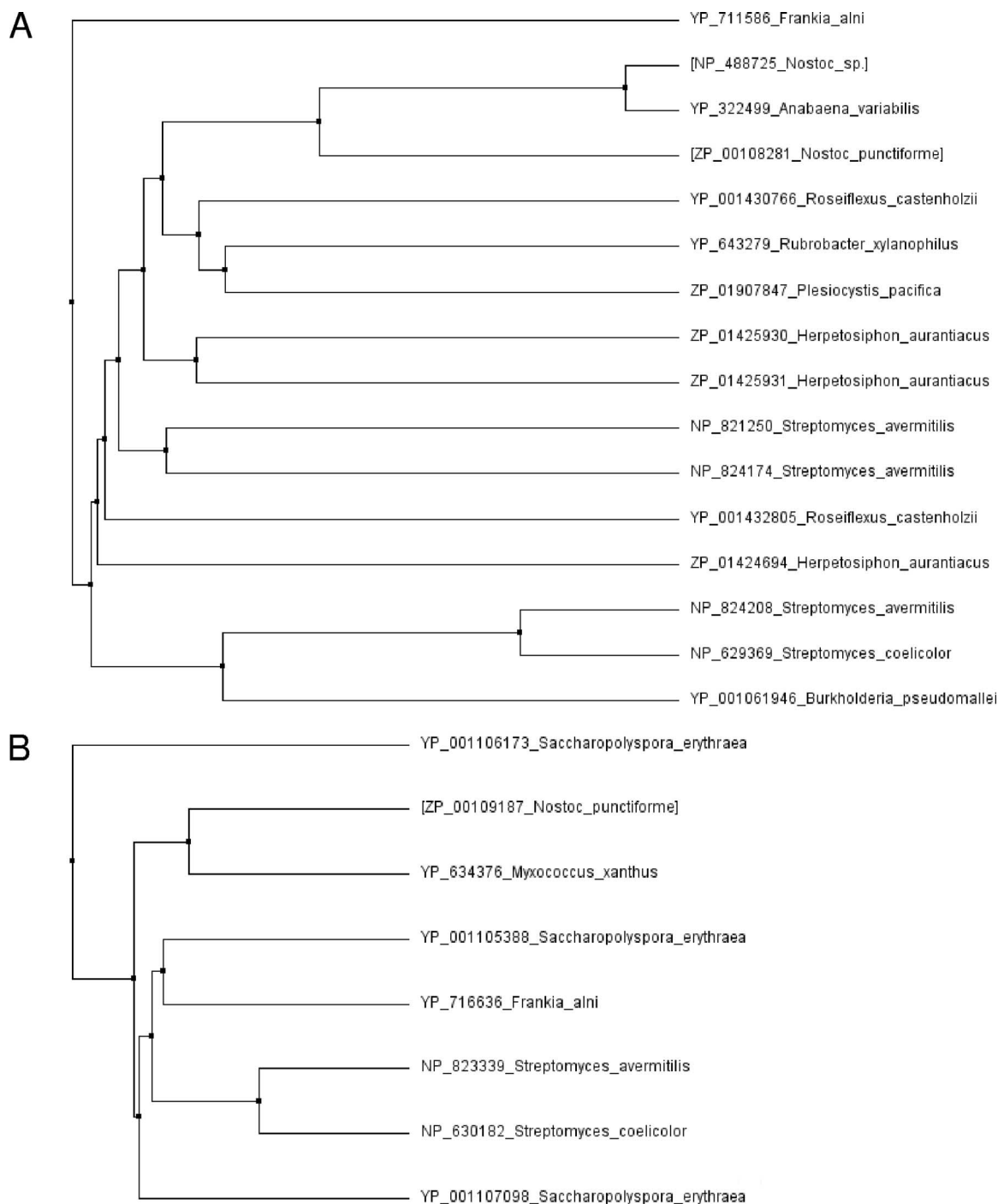


FIG. 2. Unrooted phylogenetic trees of sesquiterpene synthase homologs identified in completed bacterial sequences available in the NCBI database. (A) Unrooted phylogenetic tree of single-domain sesquiterpene synthase homologs. (B) Unrooted phylogenetic tree of germacradienol/geosmin synthase-like homologs composed of two fused terpene synthase domains. Sequences were identified by BLAST using known bacterial sesquiterpene synthases. In the case in which sequences of several strains of a bacterial species were available, the sesquiterpene synthase sequence(s) of only one representative strain was selected for sequence analysis. Alignments of sequences used for tree building are shown in Fig. S1 in the supplemental material. Phylogenetic tree branches are labeled with accession numbers followed by species names for each sesquiterpene synthase sequence. *Nostoc* terpene synthase described in this study are indicated in brackets. Note that some species have several terpene synthase paralogs.

these NP1 as reported for the structural identification of other terpene synthase products (23). Terpene synthase NP1 with an added C-terminal six-histidine tag was therefore cloned into the vector pET-21b(+) to allow for overexpression from the

strong T7 promoter in *E. coli*. However, despite various expression conditions and coexpression of helper plasmids encoding chaperones and/or rare tRNAs, NP1 could not be overexpressed at a level that would provide sufficient protein yields

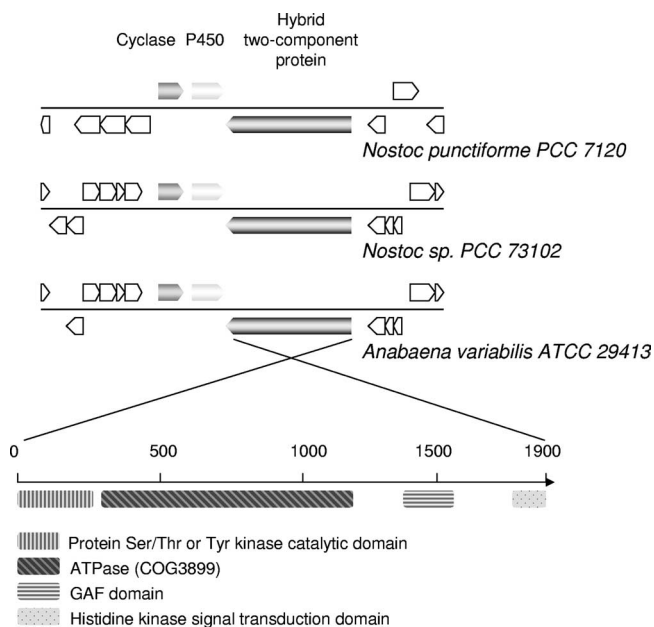


FIG. 3. Clustering of cyanobacterial sesquiterpene synthase ORFs from *Nostoc* and *Anabaena* with ORFs encoding a cytochrome P450 and a putative hybrid two-component protein. ORFs flanking the terpene synthase mini-gene cluster differ in the three cyanobacterial genomes, suggesting functional significance of the gene cluster. The lower panel shows the different domains identified in the putative hybrid two-component protein by searching the NCBI conserved domain database. The GAF domain acronym derives from the names of the first three protein classes identified to contain this domain: mammalian cGMP-stimulated phosphodiesterases, *Anabaena* adenyl cyclases, and *E. coli* transcription factor FhlA.

for preparative-scale *in vitro* sesquiterpene synthesis from FPP. Consequently, *in vivo* production of sufficient quantities of sesquiterpene was attempted by growing recombinant *E. coli* expressing NP1 under high-density fermentation conditions.

NP1 was subcloned from the high-copy-number plasmid pUCmod into the medium-copy-number plasmid pACmod, which also contains a selectable marker for chloramphenicol. This plasmid's lower copy number and selectable marker make it more stable under high-density fermentation conditions. *E. coli* pACmod-NP1 was grown to high cell densities in a 3-liter bioreactor. Volatiles, including the yet-unidentified NP1 sesquiterpene, were continuously collected from the off gas during the fermentation by using a resin-packed trap. Following elution from the resin and subsequent distillation of the eluate, about 1 mg of pure NP1 sesquiterpene for NMR analysis was obtained following final purification using a combination of preparative HPLC and GC.

The molecular formula of $C_{15}H_{24}$ indicated four degrees of unsaturation for the NP1 hydrocarbon. The one-dimensional proton NMR spectrum showed three 3H resonances at δ 1.00 (s), 1.55 (ddd, $J = 2.5, 1.5, 1.5$ Hz), and 1.73 (dd, $J = 1.5, 1$ Hz), ascribable to one quaternary alkyl group and two allylic methyl groups, and alkene resonances at δ 5.43 (br m), 4.70 (dq, $J = 2, 1$ Hz), and 4.69 (dq, $J = 2, 1.5$ Hz). These data pointed us to the structure of the rare sesquiterpene 8*a*-*epi*- α -selinene (structure 1) (Fig. 1 and 5) isolated from several genera of termite soldier defense secretions (41). Various two-dimen-

sional proton and carbon (correlation spectroscopy, HMQC, and HMBC) NMR spectra were also collected. Their iterative interpretation led to the full assignment of skeletal connectivity and of the chemical shifts for all proton and carbon resonances (Table 1). 8*a*-*epi*- α -Selinene (1) has been synthesized, and the ^{13}C NMR chemical shifts for that sample (2) match very well those we deduced from analysis of the HMQC and HMBC spectra. Nearly all 1H - 1H coupling constant (J) values were also determined, and they are further supportive of the stereostructure of structure 1. Some of the particularly diagnostic J values, associated with H_{1ax} and H_{1eq} , which solidified the relative configuration assignment among the three stereogenic centers in structure 1, are specifically shown in Fig. 5. We have no basis upon which to assign the absolute configuration of the sample produced by NP1, so the enantiomer drawn in structure 1 throughout is arbitrarily chosen.

Kinetic characterization of NS1 and NP1. For *in vitro* assays, NS1 and NP1 with an added C-terminal histidine tag were cloned into the expression vector pET-21b(+). In contrast to NP1, NS1 was overexpressed at high levels in *E. coli* BL21 (approximately 40 to 50% of total protein, compared to <5% for NP1). The recombinant terpene synthases were purified by metal-affinity chromatography resulting in enzyme preparations with >90% purity for NS1 and approximately 50 to 60% purity for NP1 (see Fig. S1 in the supplemental material).

GC-MS analysis of reactions with purified terpene synthases and FPP confirmed *in vitro* the formation of 8*a*-*epi*- α -selinene by NP1 and germacrene A by NS1 as observed in the recombinant *E. coli* cultures. Enzyme kinetics were measured using a coupled enzyme assay in which pyrophosphate released from FPP is detected. The calculated values for k_{cat} , K_m , and catalytic efficiency of the two enzymes were similar, although the catalytic efficiency for NS1 was slightly higher than that for NP1 (Table 2). The kinetic parameters of NS1 and NP1 agree well with those reported for other sesquiterpene synthases (8, 12, 39).

Characterization of cytochrome P450 located downstream of NP1 and NS1. The genes encoding NP1 and NS1 are both flanked by an ORF encoding a putative cytochrome P450 monooxygenase (P450NP and P450NS) (Fig. 3). The P450 enzymes may modify the sesquiterpene hydrocarbon products of NP1 and NS1 to generate oxidized terpenes. To analyze the activities of the two P450s, both genes were cloned into the expression vector pET-21b(+). Cell extracts from *E. coli* transformants expressing either P450NS or P450NP were then added to *in vitro* assay reactions containing FPP, purified sesquiterpene synthases NP1 or NS1 (depending on the P450 studied) and commercially available spinach ferredoxin and ferredoxin:NADPH reductase to supply reducing equivalents to the P450 enzyme (36). Although the spinach ferredoxin reductase system has been used successfully for an electron donor for other P450s (54), no oxygenated terpene products were detected in these experiments.

Because the bacterial P450s may not be compatible with the plant reductase system, *in vitro* reactions were also conducted with recombinant reductase systems from *E. coli* and *Nostoc* sp. *E. coli* NADPH:flavodoxin reductase (encoded by *fpr*) and flavodoxin (encoded by *fld*) were cloned and overexpressed in *E. coli*, and cell lysates were added to *in vitro* reactions with terpene synthases and P450s. The *E. coli* reductase system has

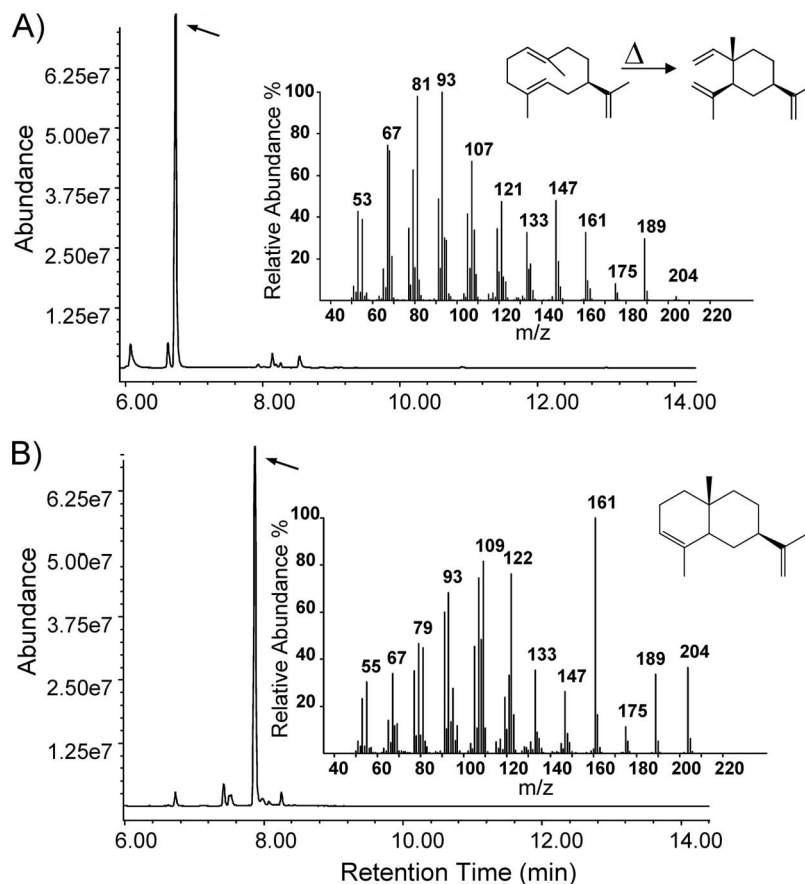


FIG. 4. GC-MS analysis of volatile organic compounds produced by *E. coli* transformants expressing single-domain sesquiterpene synthases from *Nostoc*. (A) *E. coli* cells expressing terpene synthase NS1 from *Nostoc* sp. produced germacrene A which undergoes a Cope rearrangement to form β -elemene under the high temperatures of the injection port. (B) *E. coli* cells expressing terpene synthase NP1 from *N. punctiforme* produced a germacrene-like compound that was structurally identified by NMR spectroscopy as 8a-epi- α -selinene.

been shown previously to support P450 activity in vitro (24, 25), but it did not facilitate sesquiterpene oxidation in assays with the two *Nostoc* P450s. Under the assumption that an NADPH:ferredoxin reductase and ferredoxin pair from *Nostoc* may reconstitute a native reductase system for P450NS and P450NP, both proteins were cloned from *Nostoc* sp. This cyanobacterium has two NADPH:ferredoxin reductase and ferredoxin pairs: one pair has been shown to be involved in photosystem I electron transfer (37), while the other pair likely is

TABLE 1. ^{13}C and ^1H NMR one-dimensional spectral data for 8a-epi- α -selinene (structure 1) (CDCl_3 , 600 MHz)

Atom no.	Carbon δC	Proton resonance		
		δ_H	Multiplicity	J (Hz)
1ax	29.3	1.60	ddd	13.0, 11.5, 4.4
1eq	29.3	1.78	dddd	13.5, 2.8, 2.8, 2.8
2	40.3	1.71	dddd	11, 11, 3.3, 3.3
3ax	26.9	1.47	dddd	12.7, 12.7, 11.0, 3.8
3eq	26.9	1.52	dddd	12.9, 4, 4, 4, 2
4ax	31.2	1.61	ddd	12.9, 12.9, 4.3
4eq	31.2	1.08	ddd	13.0, 3.5, 3.5
4a	31.5			
5ax	36.1	1.36	ddd	13.2, 10.9, 5.9
5eq	36.1	1.39	ddd	13.1, 6.8, 2.5
6ax	22.6	2.07	br m	Σ Js ca. 47
6eq	22.6	1.92	br d	Σ Js ca. 34
7	122.7	5.43	br m	Σ Js ca. 11
8	135.2			
8a	44.4	1.94	br m	Σ Js ca. 13
9	150.5			
10E/Z	108.1	4.70	dq	2, 1
10Z/E	108.1	4.69	dq	2, 1.5
11	21.3	1.73	dd	1.5, 1
12	27.4	1.00	s	
13	22.1	1.66	ddd	2.5, 1.4, 1.4

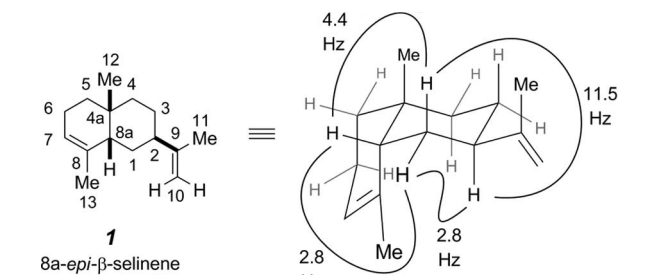


FIG. 5. Two structural representations (with atom numbering) of 8a-epi- α -selinene (structure 1). Coupling constant (J) values on the right support the all-*cis* orientation of the dominant substituents at C-2, C-4a, and C-8a (i.e., their relative configuration). Me, methyl.

TABLE 2. Kinetic parameters of *Nostoc* sesquiterpene synthases

Enzyme	K_m (μM)	k_{cat} (s^{-1})	k_{cat}/K_m ($\text{s}^{-1} \text{M}^{-1}$)
NS1	3.4 ± 0.40	$(4.4 \pm 1.8) \times 10^{-2}$	1.3×10^4
NP1	2.5 ± 0.46	$(6.3 \pm 3.1) \times 10^{-2}$	2.4×10^4

involved in nonphotosynthetic electron transfer processes. The latter pair was therefore chosen for use in the in vitro assays. However, as before, no new sesquiterpene products were obtained.

Since the in vitro studies with the two P450s were unsuccessful (partly because P450s are notorious for having low levels of activity in vitro and for being difficult to isolate as active proteins), P450 activities were investigated in vivo in *E. coli*. Plasmids expressing sesquiterpene synthase (NS1 and NP1 cloned into pDUET) and corresponding P450s [P450NS and P450NP cloned into pET-21b(+)] were transformed into *E. coli* BL21 containing plasmids encoding the *Nostoc* sp. NADPH:ferredoxin reductase and ferredoxin (pBBR-RED-FER) and supplemental tRNAs (pRARE). Media and culture headspace of the recombinant *E. coli* cultures were analyzed by SPME and GC-MS for the presence of sesquiterpenes.

One major new peak and two minor new peaks with a parent ion of an m/z of 220 were detected in the medium of *E. coli* cultures expressing the sesquiterpene synthase and P450 from *Nostoc* sp. (Fig. 6). The mass (220 amu) of each of the three new products suggests the addition of one oxygen atom to the sesquiterpene product of NS1. However, the mass spectra of these compounds did not match any of the reference terpenoid spectra in the MassFinder (29) and NIST MS databases.

Control cultures expressing only the sesquiterpene synthase accumulated β -elemene in the medium, and as expected, β -elemene was no longer detected when P450 and terpene synthase were coexpressed. However, both the control and P450-coexpressing culture accumulated considerable quantities of volatile β -elemene in their headspace, indicating that in *E. coli*, the activity of P450 is much lower than that of the terpene synthase. As a result, most of the volatile hydrocarbon sesquiterpene leaves the cell before it can be oxygenated, and only small quantities of the oxygenated terpenes accumulate in the culture medium, which prevented their further structural analysis by NMR spectroscopy.

Coexpression of the sesquiterpene synthase and P450 from *N. punctiforme* did not produce any new oxygenated terpene products. Sodium dodecyl sulfate-polyacrylamide gel electro-

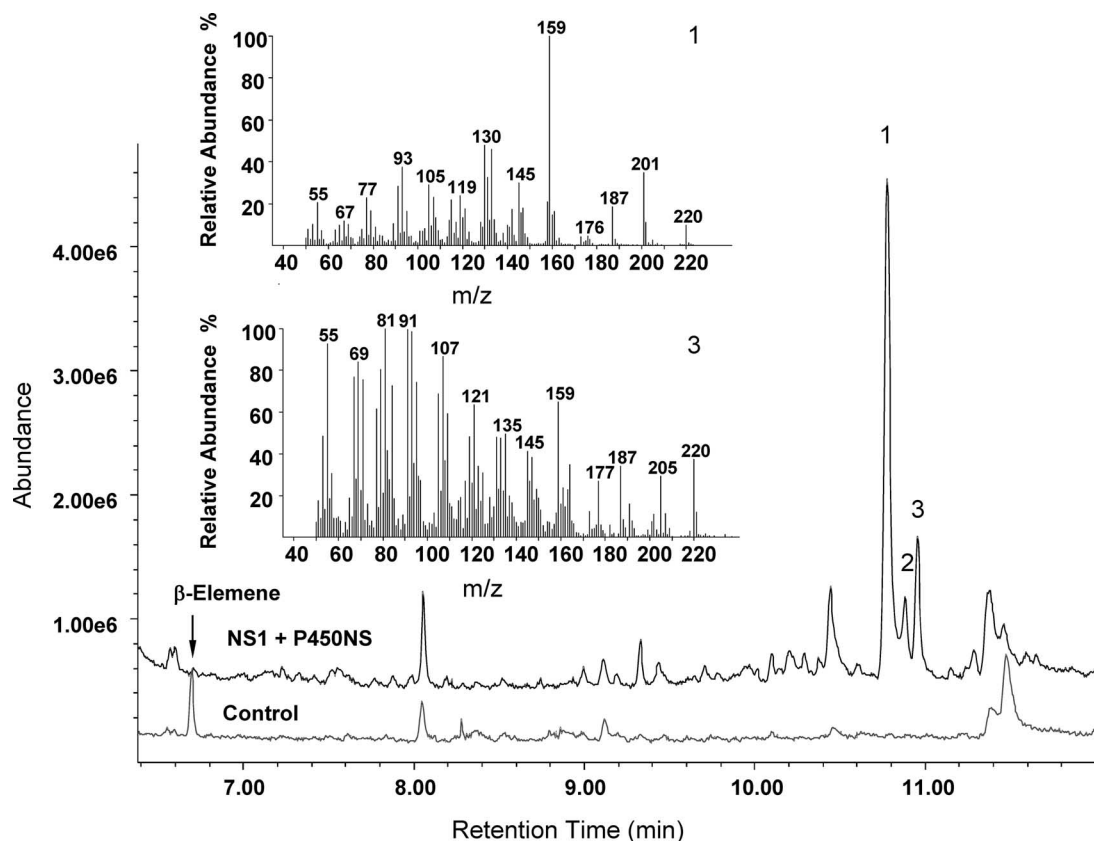


FIG. 6. GC-MS analysis of oxidized sesquiterpenes produced from germacrene A by *Nostoc* sp. P450. Culture supernatant of *E. coli* cells coexpressing single-domain sesquiterpene synthase NS1 and P450NS from *Nostoc* sp. along with the ferredoxin/ferredoxin:NADPH reductase system from *Nostoc* sp. were analyzed for the accumulation of new oxygenated sesquiterpenes. Three new peaks (labeled 1 to 3; top trace), each with a molecular ion of an m/z of 220, appeared in the chromatogram compared to control cells (bottom trace) that coexpress only terpene synthase NS1 and reductase proteins. β -Elemene (the rearrangement product of germacrene A) present in the supernatant of the control culture is not present in the culture expressing terpene synthase and P450. Mass spectra for the two larger peaks are shown and suggest the addition of one hydroxyl group to germacrene A.

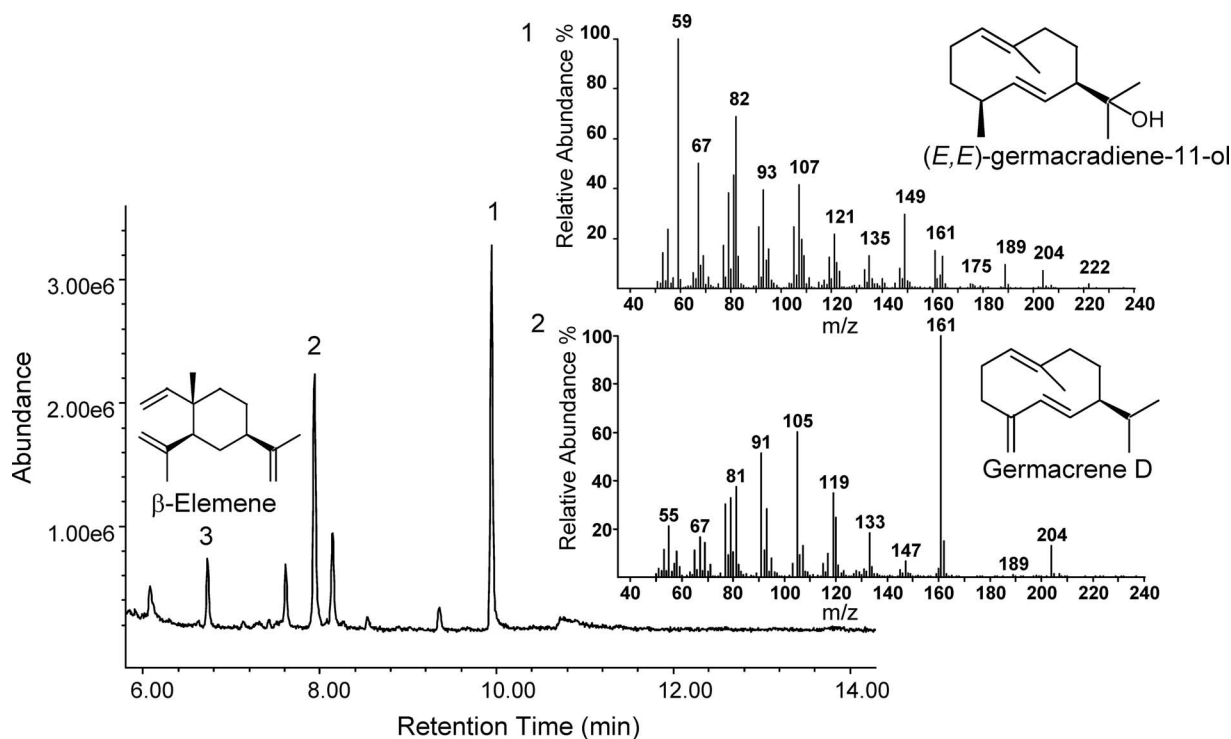


FIG. 7. GC-MS analysis of sesquiterpenoids produced by fusion-type terpene synthase NP2 from *N. punctiforme*. Culture supernatant of *E. coli* cells expressing fusion-type sesquiterpene synthase NP2 for *N. punctiforme* was analyzed for the accumulation of sesquiterpene products. One major peak (peak 1) with an m/z of 222 and a typical sesquiterpene alcohol fragmentation pattern was identified as (*E,E*)-germacradienol. Two other peaks were identified as germacrene D (peak 2) and β -elemene (peak 3; Cope rearrangement product of germacrene A). Mass spectra for peaks 1 and 2 are shown.

phoresis analysis of recombinant *E. coli* cells showed that P450NP is not at all or else only very poorly expressed compared to the homologous enzyme from *Nostoc* sp. (data not shown). Coexpression of sesquiterpene synthases from one *Nostoc* strain with the P450 from the other *Nostoc* strain did not result in the synthesis of new oxygenated sesquiterpenes, indicating that functional identified P450 from *Nostoc* sp. does not accept the *N. punctiforme* sesquiterpene as a substrate.

Cloning and characterization of fusion-type terpene synthase NP2. The second sesquiterpene synthase gene homolog identified in the genome sequence of *N. punctiforme* was amplified by PCR from genomic DNA based on the annotation in the NCBI database and cloned into the constitutive expression vector pUCmodRBS. The encoded protein shared 50% sequence identity with the germacradienol/geosmin synthases from *Streptomyces* (7, 9, 18). However, expression of this protein in *E. coli* did not result in the synthesis of sesquiterpenes. Inspection of the annotated ORF (1,893 bp) encoding this protein, along with surrounding sequences, revealed a probable sequencing error in the published gene sequence which would cause a truncated ORF. Immediately downstream of the putative terpene synthase lies an ORF that encodes a protein sequence that shows significant homology to *Streptomyces* germacradienol/geosmin synthase. The entire region containing both ORFs was therefore amplified and cloned into pUCmodRBS to give plasmid pUC-NP2. Sequencing of four independent clones showed that the published sequence contained an additional T at position 1,848 of the annotated NP2 ORF

(along with additional point mutations), causing a sequence frameshift. The cloned sequence contained an ORF (2,262 bp) with a size similar to those of the characterized germacradienol/geosmin synthase genes from *Streptomyces* (7, 9, 18) (see Fig. S4 in the supplemental material).

Plasmid pUC-NP2 was then transformed into *E. coli* JM109, and culture headspace and media were analyzed by GC-MS for the production of sesquiterpenes. A major peak with a parent ion of an m/z of 222 and a typical sesquiterpene alcohol fragmentation pattern was detected in the culture medium (Fig. 7). This compound was identified as (*E,E*)-germacradienol (the product of the N-terminal domain of germacradienol/geosmin synthases from *Streptomyces*) (7, 9, 18) by comparison with reference spectra in the MassFinder database (93% match with reference spectrum in the MassFinder terpene library). Two other sesquiterpene products, germacrene D and β -elemene (rearrangement product of germacrene A; see above) were also identified (>90% match with reference spectra in library) (Fig. 7).

To verify (*E,E*)-germacradienol synthesis *in vitro* with purified terpene synthase NP2 and to attempt to detect the C-terminal terpene synthase domain product geosmin reported to be produced in small quantities by the *Streptomyces* enzyme *in vitro* (26), NP2 was subcloned with a C-terminal histidine tag into pET-21b(+) for overexpression in *E. coli* BL21. However, overexpression of this protein under a variety of different conditions, including different cultivation temperatures and inducing conditions, always resulted in insoluble protein. NP2

could be purified from inclusion bodies, but the refolded protein did not show activity under in vitro conditions.

DISCUSSION

To date, the majority of terpene synthases have been characterized from plants and fungi (14). Only three sesquiterpene synthases (7, 8, 26, 27, 33), one diterpene synthase (13, 21), and very recently a new group of monoterpene synthases (28) have been biochemically characterized from bacteria. Of the characterized bacterial terpene synthases, all are from the genus *Streptomyces*. Sesquiterpene synthases from *Streptomyces* have been shown to produce pentalenene (8), *epi*-isozizaene (33), and geosmin (9, 26). Pentalenene is a precursor to the pentalenolactone family of antimicrobials (43). *epi*-Isozizaene has very recently been shown to be converted to the antibiotic albaflavenone by two sequential hydroxylation steps catalyzed by a cytochrome P450 monooxygenase that is clustered with the terpene synthase (55). Geosmin is produced by a new type of sesquiterpene synthase that contains two terpene synthase domains that together cyclize FPP via germacradienol (catalyzed by the N-terminal domain) into geosmin (catalyzed by the C-terminal domain) (7, 9, 26). Several members of a new group of terpene synthases identified in actinomycetes were shown to synthesize the volatile monoterpene alcohol 2-methylisoborneol from methylated geranyldiphosphate (28). These enzymes contain the two conserved metal-binding motifs characteristic of terpene synthases, but their overall sequence similarities to other known bacterial terpene synthases are low, resulting in their classification as a new group of bacterial terpene synthases.

N. punctiforme has both types of sesquiterpene synthases, while *Nostoc* sp. has one single-domain terpene synthase. The fact that cyanobacteria are major producers of the water contaminant geosmin suggested that the fusion-type terpene synthase NP2 present in *N. punctiforme* is involved in geosmin biosynthesis. The products of recombinant NP2 in *E. coli* are germacradienol (50%), germacrene D (40%), and β -elemene (10%); no geosmin was detected (Fig. 7). The amino acid sequence of NP2 cloned in this study is similar both in length and sequence to those of the two characterized enzymes from *Streptomyces* (9, 26) (see Fig. S4 in the supplemental material). The C-terminal domain, which catalyzes the conversion of germacradienol to geosmin (9, 26), therefore should be functional in the cloned NP2 enzyme. However, conversion of germacradienol to geosmin reported for the recombinant germacradienol/geosmin synthase from *S. coelicolor* A3(2) is relatively inefficient resulting in a terpene product mixture that contained 15% geosmin, 68% geosmin precursor germacradienol, and 13% dead-end product germacrene D (along with 4% octalin) (26). It appears that the C-terminal domain and/or substrate channeling of germacradienol/geosmin synthase does not optimally function under nonnative conditions. NP2, however, must be functionally expressed in *N. punctiforme*, since geosmin was detected in the culture headspace of this cyanobacterium (see Fig. S2 in the supplemental material). Moreover, neither germacradienol nor any of the other derailment products produced by recombinant NP2 were detected in the culture supernatant or headspace of *N. punctiforme*, suggesting

that the native enzyme efficiently catalyzes cyclization of FPP to geosmin.

The genes encoding single-domain sesquiterpene synthases present in the genomes of *N. punctiforme* (NP1) and *Nostoc* sp. (NS1) are part of an apparent minicuster that, in addition to the terpene synthase gene, contains genes encoding a P450 monooxygenase and a large hybrid two-component protein (Fig. 2). The two terpene synthases share greater than 50% amino acid sequence identity and were expected to synthesize similar sesquiterpene products. Indeed, as shown in Fig. 8, cyclization of FPP (structure 2) likely proceeds with both synthases via germacren-11-yl cation 3. Deprotonation from C-12 in this cation produces germacrene A (structure 4), the final product of the enzyme NS1. NP1, which produces 8a-*epi*- α -selinene (structure 1) instead, could convert germacrene A (structure 4) further into the eudesman-8-yl cation 6 by reprotonation of structure 4 at C-2 to generate the germacren-1-yl cation 5 (cf. *epi*-aristolochene) from aristolochene synthase (6). Proton loss from C-7 in structure 6 then would yield 8a-*epi*- α -selinene (structure 1), the product of NP1. However, the fact that germacrene A (structure 4) is known to undergo acid-induced rearrangement to give, principally, β -selinene (cf. Fig. 1) rather than α -selinene (1) (in fact, we observed the same [partial] rearrangement when our sample of isolated structure 4 from NS1 was heated in CDCl_3) suggests that neutral structure 4 may not be an intermediate on the path from structure 2 to structure 1. A mechanistic variant that avoids neutralization of cation 3 by proton loss to an external (active site) base, is the intramolecular proton transfer represented in the structure 3-to-structure 7 conversion shown at the bottom of Fig. 8. This would produce the germacren-8-yl tertiary carbenium ion 7, which could, in turn, undergo a second, now-transannular proton transfer to C-2 to provide an alternate entry to cation 5. In this regard, it is intriguing to consider the possible role of proton sandwich structure 8, in light of the interesting observations about such species from the Tantillo laboratories (19, 20, 40). Finally, it is even possible to consider an analogous delocalized intermediate, namely, structure 9, for the initial 1,6-proton transfer taking structure 3 to structure 7.

Germacrene A is the product of several plant sesquiterpene synthases and many plant natural products are derived from this sesquiterpenoid hydrocarbon scaffold by subsequent oxidation steps (4, 5, 15, 16, 42). 8a-*epi*- α -Selinene is a compound that has been identified previously only from cephalic secretions of soldier termites of the genus *Noditermes wasambaricus*. In termites, this sesquiterpenoid is thought to act as an oily antihealant when added to wounds inflicted by the elongated mandibles of the soldier termites (38).

The putative P450 ORFs located adjacent to the germacrene A synthase gene in *Nostoc* sp. and the 8a-*epi*- α -selinene synthase gene in *N. punctiforme* are expected to produce enzymes that modify these sesquiterpenoid hydrocarbons into various oxygenated forms. After several unsuccessful attempts to reconstitute P450 activities in vitro by using different reductase systems and conditions, we resorted to an in vivo strategy by coexpressing P450, NADPH:ferredoxin reductase, and ferredoxin from *Nostoc* sp. and terpene synthase in *E. coli*. This strategy was not successful for the P450 enzyme associated with 8a-*epi*- α -selinene synthase (NP1) in *N. punctiforme*, most likely due to poor protein expression of the P450 enzyme (data not

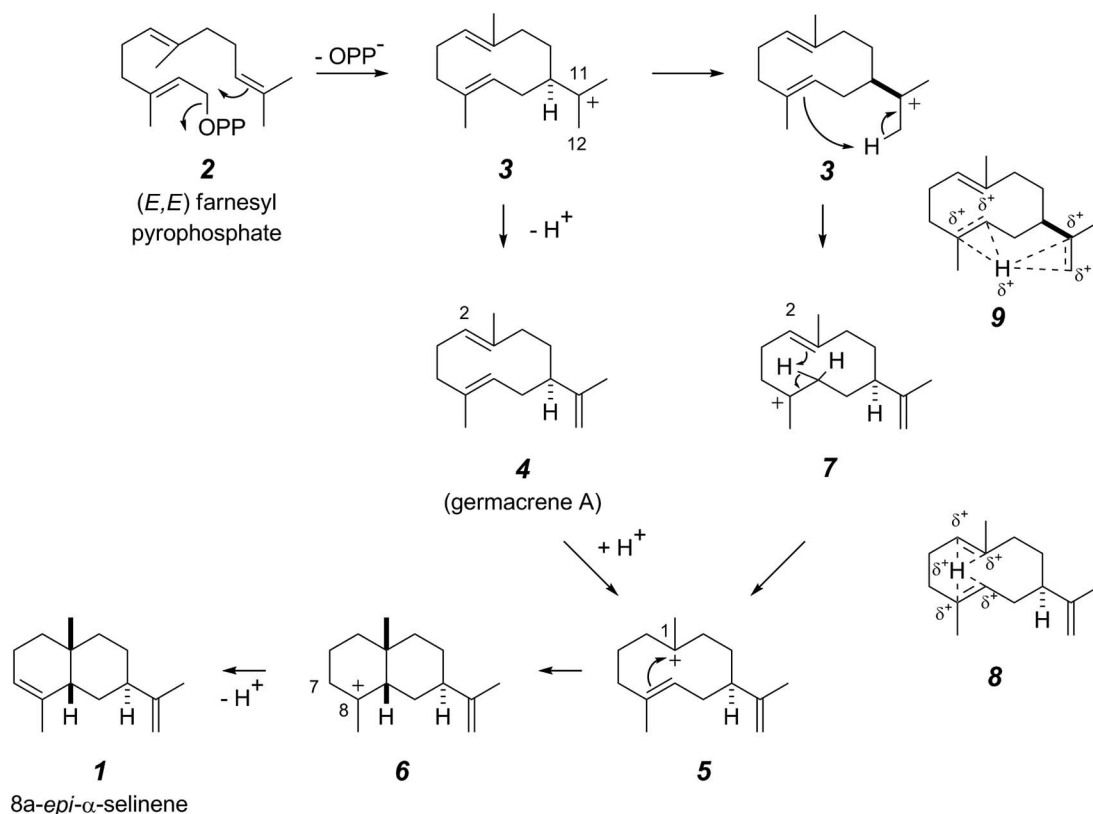


FIG. 8. Proposed mechanism for the formation of 8a-*epi*- α -selinene. Shown are alternative routes and possible intermediate species 8 and 9.

shown). However, a new major compound with a mass of 220 amu, suggesting the addition of one atom of oxygen (likely via hydroxylation), was observed when germacrene A synthase (NS1) and its corresponding P450 were coexpressed in *E. coli* (Fig. 7). A hydroxylase oxidizing the isopropenyl side chain of germacrene A to form germacratrien-12-ol (220 amu) in chicory root extract, in which it is involved in the biosynthesis of bitter-tasting sesquiterpenoid lactones in this plant, has been described (16). The quantities of the oxygenated germacrene A produced in *E. coli*, however, were too small for further NMR analysis, which would have enabled comparison of its spectra with those of known monooxygenated germacrenes. Interestingly, 8a-*epi*- α -selinene, which is structurally fairly similar to germacrene A, was not a substrate of this P450 in recombinant *E. coli*, suggesting that this enzyme must either be specific for the overall structure of germacrene A or hydroxylate a part of the structure in germacrene A that is different in 8a-*epi*- α -selinene.

Germacrene A, 8a-*epi*- α -selinene, or oxygenated derivatives of these sesquiterpenoids were never detected in the culture supernatant or headspace of the two *Nostoc* strains. Unlike that for geosmin in *N. punctiforme*, the biosynthetic genes for these terpenoids may not be expressed under standard laboratory conditions. Terpene synthase and P450 in *Nostoc* appear to be clustered with an unusual hybrid two-component protein (Fig. 3). Such a minicluster is not found (except in the closely related *Anabaena variabilis*) with any of the sesquiterpene syn-

these genes identified in published genome sequences. The presence of this complex two-component protein may indicate that terpene synthases and their associated P450 may be responsible for the synthesis of a signaling molecule that is recognized by the two-component protein similar to quorum-sensing systems (53) or that their expression is under complex regulation mediated by the two-component protein. In the latter case, the produced terpenoids may constitute some type of defense mechanism. Genome sequences of cyanobacteria have shown that they possess an extraordinary repertoire of two-component proteins with different organizations of sensor, receiver, transmitter, and response domains (3) that regulate their cellular activities in response to the environment. Terpene biosynthesis by *Nostoc* strains may be a part of this regulatory network. Additional studies are required to investigate whether the expression of terpene cyclase, P450, and hybrid two-component genes are coregulated and under which conditions they are expressed.

Investigations on the biosynthesis of terpenoid natural products in bacteria are only in their infancy. Characterization of sesquiterpene synthases identified from a variety of bacterial species may lead to new bioactive compounds. Further studies on the significance of the two-component protein in the putative terpene minicluster in *Nostoc* should provide new insights about the biological function of these sesquiterpenoids as signaling defense molecules.

ACKNOWLEDGMENTS

This research was supported by the David and Lucile Packard Foundation (grant 2001-18996) and the Academic Health Center of the University of Minnesota (grant 2005-12).

REFERENCES

- Adio, A. M., C. Paul, H. Tesso, P. Kolth, and W. A. Koenig. 2004. Absolute configuration of helminthogermacrene. *Tetrahedron Asymmetry* **15**:1631–1635.
- Ando, M., K. Kikuchi, K. Isogai, T. Ishiwatari, N. Hirata, and H. Yamazaki. 1994. Synthetic studies of sesquiterpenes with a cis-fused decalin system. *J. Nat. Prod.* **57**:1889–1899.
- Ashby, M. K., and J. Houmard. 2006. Cyanobacterial two-component proteins: structure, diversity, distribution, and evolution. *Microbiol. Mol. Biol. Rev.* **70**:472–509.
- Berteau, C. M., A. Voster, F. W. Verstappen, M. Maffei, J. Beekwilder, and H. J. Bouwmeester. 2006. Isoprenoid biosynthesis in *Artemisia annua*: cloning and heterologous expression of a germacrene A synthase from a glandular trichome cDNA library. *Arch. Biochem. Biophys.* **448**:3–12.
- Bouwmeester, H. J., J. Kodde, F. W. A. Verstappen, I. G. Altug, J. W. de Kraker, and T. E. Wallaart. 2002. Isolation and characterization of two germacrene A synthase cDNA clones from chicory. *Plant Physiol.* **129**:134–144.
- Calvert, M. J., P. R. Ashton, and R. K. Allemann. 2002. Germacrene A is a product of the aristolochene synthase-mediated conversion of farnesylpyrophosphate to aristolochene. *J. Am. Chem. Soc.* **124**:11636–11641.
- Cane, D. E., X. He, S. Kobayashi, S. Omura, and H. Ikeda. 2006. Geosmin biosynthesis in *Streptomyces avermitilis*. Molecular cloning, expression, and mechanistic study of the germacradienol/geosmin synthase. *J. Antibiot. (Tokyo)* **59**:471–479.
- Cane, D. E., J. K. Sohng, C. R. Lamberson, S. M. Rudnicki, Z. Wu, M. D. Lloyd, J. S. Oliver, and B. R. Hubbard. 1994. Pentalenene synthase. Purification, molecular cloning, sequencing, and high-level expression in *Escherichia coli* of a terpenoid cyclase from *Streptomyces* UC5319. *Biochemistry* **33**:5846–5857.
- Cane, D. E., and R. M. Watt. 2003. Expression and mechanistic analysis of a germacradienol synthase from *Streptomyces coelicolor* implicated in geosmin biosynthesis. *Proc. Natl. Acad. Sci. USA* **100**:1547–1551.
- Carter, O. A., R. J. Peters, and R. Croteau. 2003. Monoterpene biosynthesis pathway construction in *Escherichia coli*. *Phytochemistry* **64**:425–433.
- Caruthers, J. M., I. Kang, M. J. Rynkiewicz, D. E. Cane, and D. W. Christianson. 2000. Crystal structure determination of aristolochene synthase from the blue cheese mold, *Penicillium roqueforti*. *J. Biol. Chem.* **275**:25533–25539.
- Croteau, R. B., J. J. Shaskus, B. Renstrom, N. M. Felton, D. E. Cane, A. Saito, and C. Chang. 1985. Mechanism of the pyrophosphate migration in the enzymatic cyclization of geranyl and linalyl pyrophosphates to (+)- and (–)-bornyl pyrophosphates. *Biochemistry* **24**:7077–7085.
- Dairi, T., Y. Hamano, T. Kuzuyama, N. Itoh, K. Furihata, and H. Seto. 2001. Eubacterial diterpene cyclase genes essential for production of the isoprenoid antibiotic terpentecin. *J. Bacteriol.* **183**:6085–6094.
- Davis, E. M., and R. Croteau. 2000. Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes. *Biosynthesis Aromatic Polyketides Isoprenoids Alkaloids* **209**:53–95.
- de Kraker, J. W., H. J. Bouwmeester, M. C. Franssen, and A. de Groot. 1999. (+)-Germacrene A synthesis in chicory (*Cichorium intybus* L.): the first step in sesquiterpene lactone biosynthesis. *Acta Bot. Gallica* **146**:111–115.
- de Kraker, J. W., M. C. Franssen, M. C. Dalm, A. de Groot, and H. J. Bouwmeester. 2001. Biosynthesis of germacrene A carboxylic acid in chicory roots. Demonstration of a cytochrome P450 (+)-germacrene a hydroxylase and NADP⁺-dependent sesquiterpenoid dehydrogenase(s) involved in sesquiterpene lactone biosynthesis. *Plant Physiol.* **125**:1930–1940.
- Faraldos, J. A., Y. Zhao, P. E. O'Maille, J. P. Noel, and R. M. Coates. 2007. Interception of the enzymatic conversion of farnesyl diphosphate to 5-epi-aristolochene by using a fluoro substrate analogue: 1-fluorogermacrene A from (2E,6Z)-6-fluorofarnesyl diphosphate. *ChemBiochem* **8**:1826–1833.
- Gust, B., G. L. Challis, K. Fowler, T. Kieser, and K. F. Chater. 2003. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. USA* **100**:1541–1546.
- Gutta, P., and D. J. Tantillo. 2005. Proton sandwiches: nonclassical carbocations with tetracoordinate protons. *Angew Chem. Int. Ed. Engl.* **44**:2719–2723.
- Gutta, P., and D. J. Tantillo. 2006. Theoretical studies on farnesyl cation cyclization: pathways to pentalenene. *J. Am. Chem. Soc.* **128**:6172–6179.
- Hamano, Y., T. Kuzuyama, N. Itoh, K. Furihata, H. Seto, and T. Dairi. 2002. Functional analysis of eubacterial diterpene cyclases responsible for biosynthesis of a diterpene antibiotic, terpentecin. *J. Biol. Chem.* **277**:37098–37104.
- He, W., J. Huang, X. Sun, and A. J. Frontier. 2008. Total synthesis of (+/–)-merrillactone A. *J. Am. Chem. Soc.* **130**:300–308.
- He, X., and D. E. Cane. 2004. Mechanism and stereochemistry of the germacradienol/germacrene D synthase of *Streptomyces coelicolor* A3(2). *J. Am. Chem. Soc.* **126**:2678–2679.
- Jenkins, C. M., I. Pikuleva, and M. R. Waterman. 1998. Expression of eukaryotic cytochromes P450 in *E. coli*. *Methods Mol. Biol.* **107**:181–193.
- Jenkins, C. M., and M. R. Waterman. 1998. NADPH-flavodoxin reductase and flavodoxin from *Escherichia coli*: characteristics as a soluble microsomal P450 reductase. *Biochemistry* **37**:6106–6113.
- Jiang, J., X. He, and D. E. Cane. 2007. Biosynthesis of the earthy odorant geosmin by a bifunctional *Streptomyces coelicolor* enzyme. *Nat. Chem. Biol.* **3**:711–715.
- Jiang, J., X. He, and D. E. Cane. 2006. Geosmin biosynthesis. *Streptomyces coelicolor* germacradienol/germacrene D synthase converts farnesyl diphosphate to geosmin. *J. Am. Chem. Soc.* **128**:8128–8129.
- Komatsu, M., M. Tsuda, S. Omura, H. Oikawa, and H. Ikeda. 2008. Identification and functional analysis of genes controlling biosynthesis of 2-methylisoborneol. *Proc. Natl. Acad. Sci. USA* **105**:7422–7427.
- Konig, W. A., N. Bulow, and Y. Saritas. 1999. Identification of sesquiterpene hydrocarbons by gas phase analytical methods. *Flavour Fragrance J.* **14**:367–378.
- Reference deleted.
- Lawton, L. A., and G. A. Codd. 1991. Cyanobacterial (blue-green algal) toxins and their significance in UK and European waters. *J. Inst. Water Eng. Manag.* **5**:460–465.
- Lesburg, C. A., G. Z. Zhai, D. E. Cane, and D. W. Christianson. 1997. Crystal structure of pentalenene synthase: mechanistic insights on terpenoid cyclization reactions in biology. *Science* **277**:1820–1824.
- Lin, X., R. Hopson, and D. E. Cane. 2006. Genome mining in *Streptomyces coelicolor*: molecular cloning and characterization of a new sesquiterpene synthase. *J. Am. Chem. Soc.* **128**:6022–6023.
- Ludwig, F., A. Medger, H. Bornick, M. Opitz, K. Lang, M. Gottfert, and I. Roske. 2007. Identification and expression analyses of putative sesquiterpene synthase genes in *Phormidium* sp. and prevalence of *geoA*-like genes in a drinking water reservoir. *Appl. Environ. Microbiol.* **73**:6988–6993.
- Martin, V. J., D. J. Pitera, S. T. Withers, J. D. Newman, and J. D. Keasling. 2003. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.* **21**:796–802.
- Munro, A. W., H. M. Girvan, and K. J. McLean. 2007. Cytochrome P450-reductase partner fusion enzymes. *Biochim. Biophys. Acta* **1770**:345–359.
- Navarro, J. A., M. Hervas, C. G. Genzor, G. Cheddar, M. F. Fillat, M. A. de la Rosa, C. Gomez-Moreno, H. Cheng, B. Xia, Y. K. Chae, H. Yan, B. Wong, N. A. Straus, J. L. Markley, J. K. Hurley, and G. Tollin. 1995. Site-specific mutagenesis demonstrates that the structural requirements for efficient electron transfer in *Anabaena* ferredoxin and flavodoxin are highly dependent on the reaction partner: kinetic studies with photosystem I, ferredoxin:NADP⁺ reductase, and cytochrome c. *Arch. Biochem. Biophys.* **321**:229–238.
- Naya, Y., G. D. Prestwich, and S. G. Spanton. 1982. Sesquiterpenes from termite soldiers: structure of amiteol, a new 5-beta, 7-beta, 10-beta-eudesmane from *Amiteles excellens*. *Tetrahedron Lett.* **23**:3047–3050.
- Picaud, S., M. E. Olsson, M. Brodelius, and P. E. Brodelius. 2006. Cloning, expression, purification and characterization of recombinant (+)-germacrene D synthase from *Zingiber officinale*. *Arch. Biochem. Biophys.* **452**:17–28.
- Ponec, R., P. Bultinck, P. Gutta, and D. J. Tantillo. 2006. Multicenter bonding in carbocations with tetracoordinate protons. *J. Phys. Chem. A* **110**:3785–3789.
- Prestwich, G. D., I. Abe, Y. F. Zheng, B. J. Robustell, and T. Y. Dang. 1999. Enzymatic cyclization of squalene analogs. *Pure Appl. Chem.* **71**:1127–1131.
- Prosser, I., A. L. Phillips, S. Gittings, M. J. Lewis, A. M. Hooper, J. A. Pickett, and M. H. Beale. 2002. (+)-(10R)-Germacrene A synthase from goldenrod, *Solidago canadensis*; cDNA isolation, bacterial expression and functional analysis. *Phytochemistry* **60**:691–702.
- Quaderer, R., S. Omura, H. Ikeda, and D. E. Cane. 2006. Pentalenolactone biosynthesis. Molecular cloning and assignment of biochemical function to PtlI, a cytochrome P450 of *Streptomyces avermitilis*. *J. Am. Chem. Soc.* **128**:13036–13037.
- Rynkiewicz, M. J., D. E. Cane, and D. W. Christianson. 2001. Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade. *Proc. Natl. Acad. Sci. USA* **98**:13543–13548.
- Sambrook, J., and D. Russell. 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Reference deleted.
- Shishova, E. Y., L. Di Costanzo, D. E. Cane, and D. W. Christianson. 2007. X-ray crystal structure of aristolochene synthase from *Aspergillus terreus* and evolution of templates for the cyclization of farnesyl diphosphate. *Biochemistry* **46**:1941–1951.
- Svenning, M. M., T. Eriksson, and U. Rasmussen. 2005. Phylogeny of symbiotic cyanobacteria within the genus *Nostoc* based on 16S rDNA sequence analyses. *Arch. Microbiol.* **183**:19–26.
- Takahashi, S., and T. Koyama. 2006. Structure and function of cis-prenyl chain elongating enzymes. *Chem. Rec.* **6**:194–205.
- Takahashi, S., Y. Yeo, B. T. Greenhagen, T. McMullin, L. Song, J.

- Maurina-Brunker, R. Rosson, J. P. Noel, and J. Chappell.** 2007. Metabolic engineering of sesquiterpene metabolism in yeast. *Biotechnol. Bioeng.* **97**:170–181.
51. **Tetzlaff, C. N., Z. You, D. E. Cane, S. Takamatsu, S. Omura, and H. Ikeda.** 2006. A gene cluster for biosynthesis of the sesquiterpenoid antibiotic pentalenolactone in *Streptomyces avermitilis*. *Biochemistry* **45**:6179–6186.
52. **White, D. E., I. C. Stewart, R. H. Grubbs, and B. M. Stoltz.** 2008. The catalytic asymmetric total synthesis of elatol. *J. Am. Chem. Soc.* **130**:810–811.
53. **Williams, P., K. Winzer, W. C. Chan, and M. Camara.** 2007. Look who's talking: communication and quorum sensing in the bacterial world. *Philos. Trans. R. Soc. Lond. B* **362**:1119–1134.
54. **Woithe, K., N. Geib, K. Zerbe, D. B. Li, M. Heck, S. Fournier-Rousset, O. Meyer, F. Vitali, N. Matoba, K. Abou-Hadeed, and J. A. Robinson.** 2007. Oxidative phenol coupling reactions catalyzed by OxyB: a cytochrome P450 from the vancomycin producing organism. Implications for vancomycin biosynthesis. *J. Am. Chem. Soc.* **129**:6887–6895.
55. **Zhao, B., X. Lin, L. Lei, D. C. Lamb, S. L. Kelly, M. R. Waterman, and D. E. Cane.** 2008. Biosynthesis of the sesquiterpene antibiotic albaflavenone in *Streptomyces coelicolor* A3(2). *J. Biol. Chem.* **283**:8183–8189.