DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY
Office for Scholarly Communication

The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis

Citation

Butcher, R. A., F. C. Schroeder, M. A. Fischbach, P. D. Straight, R. Kolter, C. T. Walsh, and J. Clardy. 2007. "The Identification of Bacillaene, the Product of the PksX Megacomplex in Bacillus Subtilis." Proceedings of the National Academy of Sciences 104 (5): 1506–9. https://doi.org/10.1073/pnas.0610503104.

Permanent link

http://nrs.harvard.edu/urn-3:HUL.InstRepos:41483162

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story .

Accessibility

The identification of bacillaene, the product of the PksX megacomplex in *Bacillus subtilis*

Rebecca A. Butcher*, Frank C. Schroeder*, Michael A. Fischbach*, Paul D. Straight†, Roberto Kolter†, Christopher T. Walsh*‡, and Jon Clardy*‡

Departments of *Biological Chemistry and Molecular Pharmacology and †Microbiology and Molecular Genetics, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115

Contributed by Christopher T. Walsh, November 29, 2006 (sent for review November 16, 2006)

The ≈80-kb *pksX* gene cluster in *Bacillus subtilis* encodes an unusual hybrid polyketide/nonribosomal peptide synthase that has been linked to the production of the uncharacterized antibiotic bacillaene. Multiple copies of this synthase, each similar in size to the ribosome, assemble into a single organelle-like complex with a mass of tens to hundreds of megadaltons. The resource requirements of the assembled megacomplex suggest that bacillaene has an important biological role. By coupling a differential NMR spectroscopic technique with genetically manipulated strains of *B. subtilis*, we were able to characterize the structure of this unusual secondary metabolite, which could not be predicted by using bioinformatic analysis. We report that bacillaene is a linear molecule with two amide bonds: the first links an α-hydroxy carboxylic acid to a ω-amino carboxylic acid containing a conjugated hexaene, and the second links the hexaene-containing carboxylic acid to an (ω-1) amino carboxylic acid containing a conjugated triene. Knowledge of bacillaene's structure has enabled us to annotate the *pksX* gene cluster and should facilitate the study of bacillaene's biosynthesis as well as its biological role in *B. subtilis*.

NMR overlay | nonribosomal peptide | polyketide | β-methyl

Two independent lines of inquiry recently converged on bacillaene, a previously obscure antibiotic produced by *Bacillus subtilis* that was first isolated a decade ago (1). The first study, which used bioinformatic analysis of bacterial genomes to identify the unappreciated biosynthetic potential of even well known bacteria, structurally and functionally characterized three giant polyketide synthase (PKS) gene clusters in *Bacillus amyloliquefaciens* FZB 42 (2). One of these giant clusters, an ortholog of the *pksX* gene cluster in *B. subtilis*, was linked to bacillaene production. Bacillaene originally was discovered as a bacteriostatic agent produced by *B. subtilis* that inhibited prokaryotic, but not eukaryotic, protein synthesis by an unknown mechanism (1). The second study, which began with a screen to identify naturally occurring small molecules that modulate interactions between the common soil genera *Bacillus* and *Streptomyces*, also led to the small molecule produced by PksX synthase (3). The ≈80-kb *pksX* gene cluster occupies some 2% of the *B. subtilis* genome, and a single copy of its encoded synthase and auxiliary proteins would form a 2.5-MDa complex, roughly the size of the ribosome. Further investigation revealed that PksX synthase assembled as multiple copies into a single organelle-like enzymatic complex with a mass of tens to hundreds of megadaltons, which is an unprecedented finding in bacterial subcellular organization (3). Although the biological activity of bacillaene is not yet well characterized, the genetic and metabolic resources devoted to its production argue that it is likely to have an important biological role(s).

The *pksX* gene cluster encodes a hybrid PKS/nonribosomal peptide synthetase (NRPS). PKSs and NRPSs are large, modular enzymes that iteratively couple malonyl derivatives and amino acids, respectively (4). These systems use varied building blocks and tailoring enzymes to create a diverse array of secondary metabolites, many of which have therapeutic uses. Because NRPS and PKS enzymes are modular and the genes that encode them are often colinear with the order of molecular assembly, bioinformatic prediction of product structure (or part structure) sometimes is possible (5, 6). However, a number of unusual features present in the PksX synthase have frustrated efforts to use such analysis to link the biosynthetic genes to a small molecule structure for bacillaene. These features include unconventional module organization, a number of novel domains, and multiple enzymes that act in trans rather than in the standard assembly-line mode that is characteristic of polyketide and nonribosomal peptide biosynthesis.

Isolation of bacillaene using traditional activity-based fractionation is challenging given the molecule's chemical instability. To avoid decomposition while identifying the structure of bacillaene, we adopted a purification strategy based on minimal chromatography. We then used an NMR overlay technique to identify compounds present in extracts from strains with the *pksX* gene cluster (*pksX*$^+$) and absent in extracts from strains in which the cluster was deleted (*pksX*$^-$) (7). The structure of bacillaene is remarkable given that it is linear, has a high degree of conjugation with several cis double bonds, and incorporates a β-branch, a feature seen in only a few other polyketides (8–12). Based on bacillaene's structure, we were able to annotate further the *pksX* gene cluster, which contains two unusual split modules—modules in which the enzymatic domains of a single polyketide module are split between two separate proteins. Additionally, we were able to propose a mechanism by which the synthase incorporates the β-branch methyl group into the hexaene portion of the molecule.

Results and Discussion

To identify the structure of bacillaene, we focused our efforts on two known *B. subtilis* producers: 3610, which was used by Straight *et al.* (3), and 55422, which was used by Patel *et al.* (1). Early in our work, we observed that *B. subtilis* 3610 produces an antibiotic activity against *Streptomyces avermitilis* that requires the *pksX* gene cluster. This antibiotic activity is simpler to assay after removing the background activity of subtilosin (product of the *sbo–alb* gene cluster) by genetic deletion of the *albA* gene (3). Thus, we pursued bacillaene produced by the 3610 *albA*$^-$ *pksX*$^+$ strain by using growth inhibition of *S. avermitilis* as an assay. In addition, a previous report had suggested that the *B. subtilis*

Author contributions: R.A.B., F.C.S., P.D.S., and J.C. designed research; R.A.B. and P.D.S. performed research; R.A.B., F.C.S., and M.A.F. analyzed data; and R.A.B., F.C.S., M.A.F., P.D.S., R.K., C.T.W., and J.C. wrote the paper.

The authors declare no conflict of interest.

Abbreviations: PKS, polyketide synthase; NRPS, nonribosomal peptide synthetase; dqfCOSY, double-quantum-filtered correlation spectroscopy; α-HIC, α-hydroxy-isocaproic acid; KR, ketoreductase; AT, acyltransferase; T, thiolation.

‡To whom correspondence may be addressed. E-mail: christopher_walsh@hms.harvard.edu or jon_clardy@hms.harvard.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0610503104/DC1.

© 2007 by The National Academy of Sciences of the USA

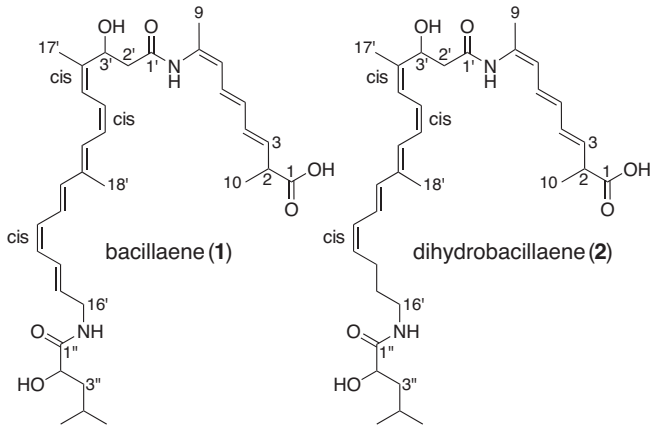

Fig. 1. Structures of bacillaene stereoisomer **1** (*Left*) and dihydrobacillaene stereoisomer **2** (*Right*). Other prominent isomers include the C6′–C7′ trans double-bond isomer of **1** and the C12′–C13′ trans double-bond isomer of **2**. Additional stereoisomers, differing in the geometries of the various double bonds in the hexaene/pentaene fragment, were present as minor components.

55422 strain produced bacillaene in significant amounts (>0.8 mg/liter) (1). Because it was unclear whether the materials produced by the two strains were identical or similar, we pursued the activity in both.

Bacillaene's rapid decomposition when exposed to light or room temperature, which had frustrated earlier efforts to characterize it, dictated an identification strategy based on minimal chromatography coupled with detailed NMR spectroscopic analysis of mixtures. Strains were grown in rich media for 24 h, extracted with ethyl acetate or C18 resin, and separated into fractions by silica gel chromatography. Active fractions were characterized by using double-quantum-filtered (dqf) COSY NMR spectra, which showed both the 3610 *pksX*$^+$ *albA*$^-$ and 55422 active fractions were complex mixtures of polyunsaturated compounds. To pinpoint compounds derived from the *pksX* gene cluster, active fractions from the 3610 *pksX*$^+$ *albA*$^-$ strain and the 55422 strain and crude extracts from the knockout strain 3610 *pksX*$^-$ *albA*$^-$ were compared by using a recently developed dqfCOSY overlay technique (7). These comparisons clearly defined distinct subsets of signals present in 3610 *pksX*$^+$ *albA*$^-$ and 55422 but missing in 3610 *pksX*$^-$ *albA*$^-$ [see supporting information (SI) Fig. 3]. Additional heteronuclear multiple-quantum correlation, heteronuclear multiple-bond correlation, and ROESY NMR spectra showed these signals represented α-hydroxy-isocaproic acid (α-HIC), a conjugated triene, and a conjugated hexaene moiety of the bacillaenes, and as each occurred multiple times, the presence of multiple stereoisomers seemed likely (Fig. 1 and SI Table 1 and SI Figs. 4–13). In addition, we detected a related series of compounds representing stereoisomers of a pentaene analog of the hexaene unit, which represent the dihydrobacillaenes in agreement with Chen *et al.* (2). The NMR spectroscopic data indicated that the α-HIC and triene units are linked to the central hexaene/pentaene core via two amide linkages (Fig. 1). The heteronuclear multiple-bond correlation between the C-9 methyl protons and carbonyl C-1′ link the pentaene/hexaene fragment and triene fragment, and the correlation between the C-16′ methylene protons and the carbonyl C-1″ link the pentaene/hexaene and α-HIC fragments. Multiple ^{1}H-^{13}C heteronuclear multiple-bond correlations between the two amide N–H protons and nearby carbons were observed, and multiple ^{1}H-^{15}N heteronuclear multiple-bond correlations between the two amide nitrogens and nearby protons were observed (SI Figs. 14 and 15). High-resolution positive-ion electrospray MS supported this interpretation with strong signals at 581.3583 for bacillaene plus H ($C_{34}H_{49}N_2O_6$) and 583.3749 for dihydrobacillaene plus H ($C_{34}H_{51}N_2O_6$) for the active 3610 pksX$^+$ *albA*$^-$ and 55422 fractions. Earlier research had obtained a different mass for bacillaene, 603.3332 (M plus Na), and assigned the empirical formula $C_{35}H_{48}O_7$ to the molecule (1).

The spectroscopic data indicate that the bacillaenes and dihydrobacillaenes are mixtures of stereoisomers, differing in the geometries of the double bonds in the hexaene/pentaene fragment. Coupling constants, as well as through-space interactions obtained from ROESY/NOESY experiments, allowed double-bond geometries to be assigned (SI Table 1 and SI Fig. 5). Under our experimental conditions, **1** and **2** were the predominant stereoisomers isolated. Other prominent isomers identified include the C6′–C7′ trans double-bond isomer of **1** and the C12′–C13′ trans double-bond isomer of **2** (SI Table 2). The 3610 *pksX*$^+$ *albA*$^-$ strain produced mostly bacillaenes, and the 55422 strain produced mostly dihydrobacillaenes, although both strains made both types of molecules. The complexity of the mixtures, coupled with the lability of these conjugated polyenes with elimination-prone hydroxy and amide functionality, raised the possibility that isomerization had occurred during the silica gel chromatography step. Differential analysis of dqfCOSY spectra acquired with crude extracts and chromatographed material from 3610 *pksX*$^+$ *albA*$^-$ clearly showed that the bacillaenes are present in approximately the same stereoisomeric ratios before and after chromatography (data not shown).

The structures of these unusual molecules lead to a proposed scheme for their biosynthesis by the PksX proteins (Fig. 2). The bacillaene synthase is composed of 13 PKS and 3 NRPS modules. Because there are no acyltransferase (AT) domains within the large, multimodular PksX synthase, it becomes another member of the growing class of PKSs in which monomers are introduced in trans by free-standing AT domains (13). The bacillaene synthase is unique even among trans-AT synthases because it encodes three free-standing AT domains (PksCDE), one of which is fused to a predicted flavin mononucleotide-dependent oxidoreductase (PksE). PksC has been characterized as a malonyl-CoA-specific AT that accepts AcpK as a thiolation (T) domain substrate (8), but it is unknown whether PksC services other T domains in the PksX synthase, and PksD and PksE have not yet been characterized.

The first module of PksJ initiates the chain with α-HIC, and α-hydroxy acids are rarely found in nonribosomal peptides or polyketides. The one characterized example is formed by the ketoreductase (KR)-catalyzed reduction of a thioester-linked α-keto-acid (14), but no such KR is found in the initiation module. The second NRPS module inserts glycine to complete the dipeptide starter unit, and the third NRPS module inserts alanine as part of an *N*-(β-OH-propionyl)-Ala spacer between the hexaene and the triene. The substrate specificities of PksJ–A2 (glycine) and PksN–A1 (alanine) agree with our previous report (15).

The bacillaene synthase contains two unusual split modules at the interfaces of PksJ–PksL and PksL–PksM (colored blue in Fig. 2*B*). In standard PKS modules that process the β-keto group to an α,β-enoyl moiety, the domain organization is KS-(AT)-DH-KR-T, and the KR and DH act sequentially after KS-catalyzed chain elongation (4). In contrast, the PksJ–PksL and PksL–PksM split modules appear to divide reduction and dehydration into discrete steps. After the reduction of the β-ketoacyl-*S*-T to a β-hydroxyacyl-*S*-T, the second KS likely serves as a transthiolation catalyst in an intramodule, interprotein chain transfer to move the β-hydroxyacyl-*S*-T intermediate from the first T domain to the second. The DH domain then dehydrates the β-hydroxyacyl-*S*-T to an α,β-enoyl-*S*-T, the substrate for further elongation. Interestingly, both of these split modules may introduce an unusual cis olefin into the growing chain.

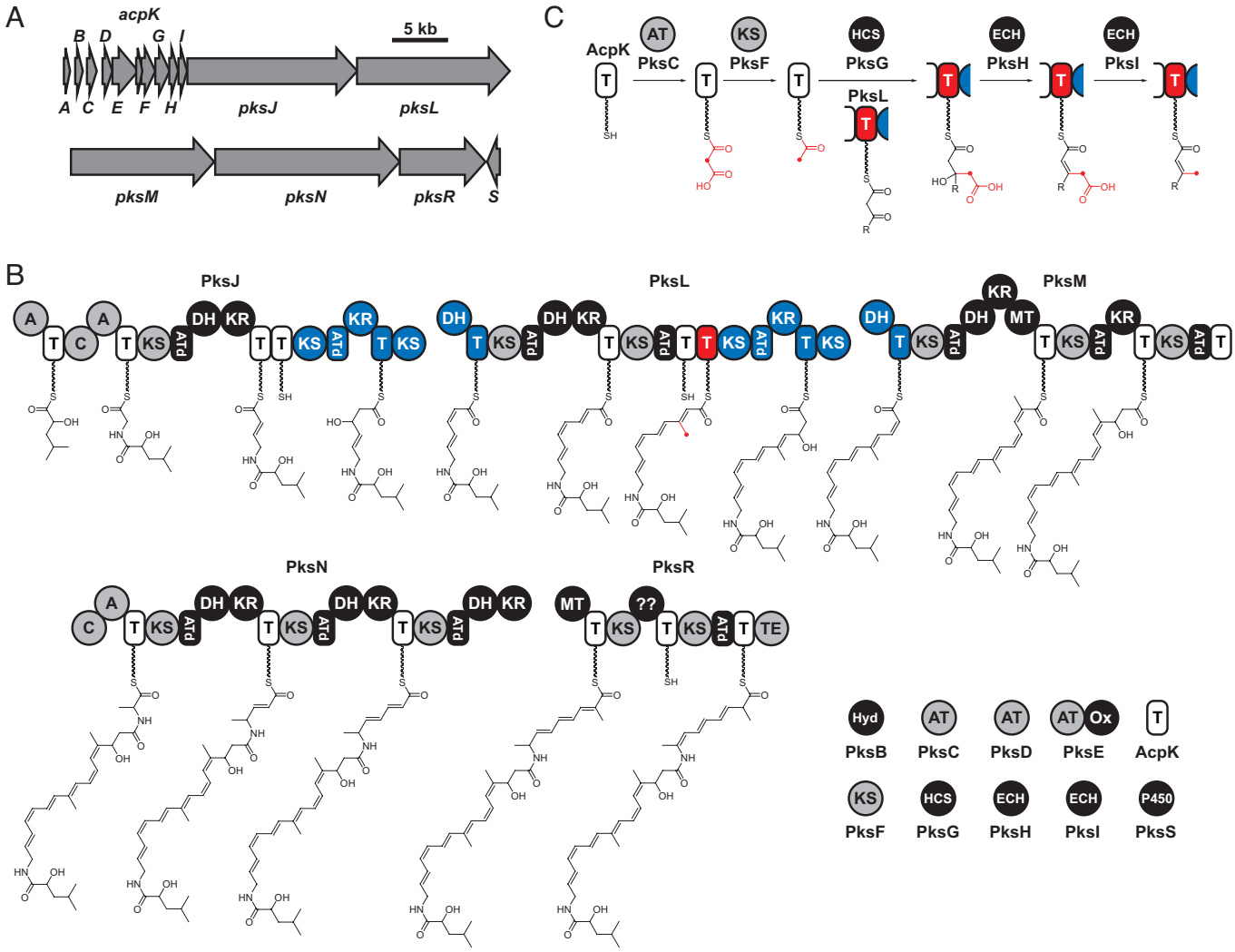

Fig. 2. The bacillaene PksX synthase. (*A*) The *pksX* gene cluster. (*B*) Proposed model for bacillaene biosynthesis. (*C*) Insertion of a β-methyl branch on PksL by the PksCFGHI/AcpK subcluster (8). KS, ketosynthase; AT, acyltransferase; T, thiolation; DH, dehydratase; KR, ketoreductase; MT, methyltransferase; A, adenylation; C, condensation; ATd, AT-docking; Hyd, Zn-dependent hydrolase; Ox, flavin mononucleotide-dependent oxidase; HCS, HMG-CoA synthase; ECH, enoyl-CoA hydratase/isomerase; TE, thioesterase.

Earlier work on the stereochemistry of the KR reaction in polyketide biosynthesis linked the presence/absence of a conserved aspartate residue in the KR domain with the stereochemical outcome of the subsequent dehydration reaction (16–18). Indeed, the three modules in the PksX synthase that generate a cis double bond lack this conserved aspartate in the KR domain (i.e., PksJ–KR2, PksL–KR2, and PksM–KR1). However, we do not rule out the possibility that other structural differences in the KR domains or other factors specific to the split modules may contribute to cis double-bond formation.

PksCFGHI and AcpK comprise a subcluster (Fig. 2*C*) of proteins that use a variation of standard isoprene biosynthesis to construct a Δ^2-isoprenyl-*S*-T branch on the T_2 motif of PksL from a β-ketoacyl-*S*-T intermediate (8). The resulting methyl branch lies at a β-position in the middle of the hexaene moiety. In contrast to the more common strategy of α-methylation by *S*-adenosyl methionine-dependent methyltransferase domains (such as those in PksM and PksR), β-alkylation requires the action of five enzymes and one external T domain (8, 19).

The last module of PksM appears to be skipped during assembly of bacillaene, in analogy to other recently described examples of module skipping in PKSs (13). This predicted skipping is consistent with this module's absence in BaeM, the analogous component of the otherwise highly homologous *B. amyloliquefaciens* FZB 42 bacillaene synthase (2). The penultimate step of bacillaene biosynthesis is likely to be isomerization of the terminal triene out of conjugation with the adjacent thioester while the chain is tethered to one of the two full modules of PksR, neither of which incorporates an extender unit into the chain. Lastly, the thioesterase domain catalyzes the attack of a nucleophilic water molecule on the acyl-*O*-thioesterase intermediate, releasing bacillaene as a linear-free carboxylate.

Small molecules like bacillaene, which link genotype (the *pksX* gene cluster) with phenotype (antibiotic and likely other activities), are central to chemical biology, and the approach used here should be generally useful for their characterization. Knowledge of the structure of bacillaene facilitates the study of its unusual biosynthetic steps, and *B. subtilis* is an ideal organism in which to address these questions because its genetic manipulability makes genetic knockout, complementation, and reconfiguration possible. Production of such unusual secondary metabolites by *B. subtilis* also demonstrates that this genetically characterized and manipulable host has an unappreciated met-

abolic potential for heterologous production of complex natural products. Additionally, it will be interesting to explore the biological role of the bacillaenes and their structural analogs, which now can be approached through either genetic modifications to the biosynthetic pathway or chemical synthesis.

Materials and Methods

Strains and Media. *B. subtilis* 3610 *albA*Δ::*kanR* and 3610 *albA*Δ::*kanR pksX*Δ::*specR* were constructed as described (3). The *B. subtilis* 55422 strain was obtained from ATCC (Manassas, VA). YEME7 media consisted of 4 g/liter yeast extract, 10 g/liter malt extract, 4 g/liter glucose, and 0.5 ml of 1 M sodium hydroxide per liter. After autoclaving, filter-sterilized 10× Mops (209.3 g/liter Mops, 3.4 g/liter potassium phosphate monobasic, and 4.36 g/liter potassium phosphate dibasic; add sodium hydroxide until pH 7.0 is reached) was added to 1×. YEME7 plates included 15 g of agar per liter.

Bioassay. A lawn of $\approx 1 \times 10^6$ *S. avermitilis* spores was spread on a YEME7 agar plate, and plates were placed in a hood until dry ($\approx$15 min). Fractions were spotted in 5 μl of dimethyl sulfoxide in the center of the plates, which were incubated for 2 days at 30°C, and zones of growth inhibition were observed.

Characterization of Dihydrobacillaenes from 55422. For the 55422 strain, 8 liters of bacteria were grown in YEME7 media for 24 h in a darkened shaker at 37°C, and then the culture was extracted with ethyl acetate. The layers were separated by centrifugation at 5,000 × *g* for 15 min, and the ethyl acetate layer was evaporated by rotary evaporation at room temperature. The residue then was fractionated by silica gel chromatography with a 1:20 to 1:5 methanol:methylene chloride gradient. Care was taken to avoid exposure to light. Fractions were assayed by using the *S. avermitilis* assay. NMR spectroscopic experiments were performed in CD_3OD or CD_3OH with a symmetrical NMR microtube susceptibility-matched with methanol (Shigemi, Inc., Allison Park, PA) on a Varian (Palo Alto, CA) INOVA 600 NMR. High-resolution phase-sensitive nongradient dqfCOSY spectra were acquired by using the following parameters: number of transients (nt) = 8, 16, or 32; acquisition time (at) = 0.5 or 0.6 s; and number of increments (ni) = 128 per 2 ppm of sweep width.

Characterization of Bacillaenes from 3610. For the 3610 strain, 24 liters of bacteria were grown in YEME7 media for 24 h in a darkened shaker at 37°C. Because isolation of the bacillaenes from 3610 by ethyl acetate extraction was inefficient and inconsistent, the culture supernatant instead was passed over 10-g C18 Sep-pak Cartridges (Waters, Milford, MA), washed with water, and eluted with 1:1 ethyl acetate:hexaene at room temperature. The elution fractions were evaporated by rotary evaporation. The residue then was fractionated by silica gel chromatography with a 1:20 to 1:5 methanol:chloroform gradient. Care was taken to avoid exposure to light. Fractions were assayed by using the *S. avermitilis* assay. NMR spectroscopic experiments were performed as described above.

MS. High-resolution MS was performed on a Micromass Q-TOF-2 spectrometer equipped with an Alliance 2690/2695 HPLC system (Waters). The exact mass obtained for [bacillaene plus H] was 581.3583, and the exact mass obtained for [dihydrobacillaene plus H] was 583.3749.

This work was supported by National Institutes of Health Grant CA24487 (to J.C.). R.A.B. was supported by a Kirschstein National Service Research Award from the National Institutes of Health, and M.A.F. was supported by a graduate fellowship from the Hertz Foundation.

1. Patel PS, Huang S, Fisher S, Pirnik D, Aklonis C, Dean L, Meyers E, Fernandes P, Mayerl F (1995) *J Antibiot (Tokyo)* 48:997–1003.
2. Chen XH, Vater J, Piel J, Franke P, Scholz R, Schneider K, Koumoutsi A, Hitzeroth G, Grammel N, Strittmatter AW, *et al.* (2006) *J Bacteriol* 188:4024–4036.
3. Straight PD, Fischbach MA, Walsh CT, Rudner DZ, Kolter R (2007) *Proc Natl Acad Sci USA* 104:305–310.
4. Fischbach MA, Walsh CT (2006) *Chem Rev* 106:3468–3496.
5. Lautru S, Deeth RJ, Bailey LM, Challis GL (2005) *Nat Chem Biol* 1:265–269.
6. McAlpine JB, Bachmann BO, Piraee M, Tremblay S, Alarco AM, Zazopoulos E, Farnet CM (2005) *J Nat Prod* 68:493–496.
7. Schroeder FC, Gibson DM, Churchill ACL, Sojikul P, Wursthorn EJ, Krasnoff SB, Clardy J (December 20, 2006) *Angew Chem Int Ed*, 10.1002/anie.200603821.
8. Calderone CT, Kowtoniuk WE, Kelleher NL, Walsh CT, Dorrestein PC (2006) *Proc Natl Acad Sci USA* 103:8977–8982.
9. Fortman JL, Sherman DH (2005) *ChemBioChem* 6:960–978.
10. Edwards DJ, Marquez BL, Nogle LM, McPhail K, Goeger DE, Roberts MA, Gerwick WH (2004) *Chem Biol* 11:817–833.
11. Chang Z, Sitachitta N, Rossi JV, Roberts MA, Flatt PM, Jia J, Sherman DH, Gerwick WH (2004) *J Nat Prod* 67:1356–1367.
12. Gu L, Jia J, Liu H, Hakansson K, Gerwick WH, Sherman DH (2006) *J Am Chem Soc* 128:9014–9015.
13. Wenzel SC, Muller R (2005) *Curr Opin Chem Biol* 9:447–458.
14. Magarvey NA, Ehling-Schulz M, Walsh CT (2006) *J Am Chem Soc* 128:10698–10699.
15. Dorrestein PC, Blackhall J, Straight PD, Fischbach MA, Garneau-Tsodikova S, Edwards DJ, McLaughlin S, Lin M, Gerwick WH, Kolter R, *et al.* (2006) *Biochemistry* 45:1537–1546.
16. Carvalho R, Reid R, Viswanathan N, Gramajo H, Julien B (2005) *Gene* 359:91–98.
17. Reid R, Piagentini M, Rodriguez E, Ashley G, Viswanathan N, Carney J, Santi DV, Hutchinson CR, McDaniel R (2003) *Biochemistry* 42:72–79.
18. Caffrey P (2003) *ChemBioChem* 4:654–657.
19. Walsh CT, Chen H, Keating TA, Hubbard BK, Losey HC, Luo L, Marshall CG, Miller DA, Patel HM (2001) *Curr Opin Chem Biol* 5:525–534.