

Characterization of the Saframycin A Gene Cluster from *Streptomyces lavendulae* NRRL 11002 Revealing a Nonribosomal Peptide Synthetase System for Assembling the Unusual Tetrapeptidyl Skeleton in an Iterative Manner[†]

Lei Li, Wei Deng, Jie Song, Wei Ding, Qun-Fei Zhao, Chao Peng, Wei-Wen Song, Gong-Li Tang,* and Wen Liu*

State Key Laboratory of Bioorganic and Natural Product Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, 354 Fenglin Rd., Shanghai 200032, China

Received 26 May 2007/Accepted 23 October 2007

Saframycin A (SFM-A), produced by *Streptomyces lavendulae* NRRL 11002, belongs to the tetrahydroisoquinoline family of antibiotics, and its core is structurally similar to the core of ecteinascidin 743, which is a highly potent antitumor drug isolated from a marine tunicate. In this study, the biosynthetic gene cluster for SFM-A was cloned and localized to a 62-kb contiguous DNA region. Sequence analysis revealed 30 genes that constitute the SFM-A gene cluster, encoding an unusual nonribosomal peptide synthetase (NRPS) system and tailoring enzymes and regulatory and resistance proteins. The results of substrate prediction and in vitro characterization of the adenylation specificities of this NRPS system support the hypothesis that the last module acts in an iterative manner to form a tetrapeptidyl intermediate and that the colinearity rule does not apply. Although this mechanism is different from those proposed for the SFM-A analogs SFM-Mx1 and safracin B (SAC-B), based on the high similarity of these systems, it is likely they share a common mechanism of biosynthesis as we describe here. Construction of the biosynthetic pathway of SFM-Y3, an aminated SFM-A, was achieved in the SAC-B producer (*Pseudomonas fluorescens*). These findings not only shed new insight on tetrahydroisoquinoline biosynthesis but also demonstrate the feasibility of engineering microorganisms to generate structurally more complex and biologically more active analogs by combinatorial biosynthesis.

Saframycins, belonging to the tetrahydroisoquinoline family of antibiotics, are a group of microbial natural products isolated from *Streptomyces lavendulae* strain 314 (deposited in NRRL with the accession number 11002) (40). Saframycin A (SFM-A), which has antiproliferative activity against a variety of tumor cell lines at low doses, is one of the most potent members of this class of compounds (3, 29). As shown in Fig. 1, SFM-A contains a characteristic bisquinone core with a α -amino nitrile moiety (at the C-21 position), and the departure of the nitrile moiety from C-21 in the presence of reduced cofactors allows the formation of an electrophilic iminium ion that alkylates the guanine residues of double-stranded DNA (17, 18, 23, 34, 35). Although it has frequently been speculated that covalent modification of DNA is essential for the antitumor activity of SFM-A, recently the identification of glyceraldehyde-3-phosphate dehydrogenase (a putative key transcriptional coactivator necessary for entry into the S phase of cell proliferation) as a protein target of SFM-DNA adducts suggested that the action of SFM-A involves a protein-

drug-DNA interaction; thus, a distinct pathway may be involved for SFM-A antiproliferative activity (31, 48).

Ecteinascidin 743 (ET743) (Fig. 1), isolated from marine invertebrates, has a core that is structurally similar to that of SFM-A and is currently in phase II/III clinical trials as an anticancer drug (36, 38, 43). The antiproliferative activity of ET743 is higher than those of many clinically used drugs, such as taxol, by 1 to 3 orders of magnitude. However, the natural scarcity (1 mg of ET743 per 1 kg of tunicate) and the structural complexity may ultimately limit the practical value of preparing the drug either by extraction from the natural source or by total synthesis. Recently, advances in biotechnology have provided a promising alternative to make complex natural products by genetic engineering of the biosynthetic pathways in microorganisms (5, 11, 13). Therefore, an alternative to economically producing ET743 or analogs is by reconstructing the biosynthetic pathway in a recombinant microorganism (8), and the success of this approach critically depends on characterization of the biosynthetic mechanism of ET743-like natural products. The structural similarities between ET743 and SFM-A suggest that they have a common biosynthetic strategy to form a similar intermediate, despite their difference in origin (in fact, tunicates often harbor many symbiotic bacteria that are assumed to be the "real" source of numerous biologically active compounds [12, 20]). Thus, SFM-A, a terrestrial microbial metabolite, serves as a model for this family of compounds to identify a biosynthetic paradigm, and the results of

* Corresponding author. Mailing address: Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, 354 Fenglin Rd., Shanghai, 200032, China. Phone for Wen Liu: 86-21-54925111. Fax: 86-21-64166128. E-mail: wliu@mail.sioc.ac.cn. Phone for Gong-Li Tang: 86-21-54925113. E-mail: gltang@mail.sioc.ac.cn.

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

[‡] Published ahead of print on 2 November 2007.

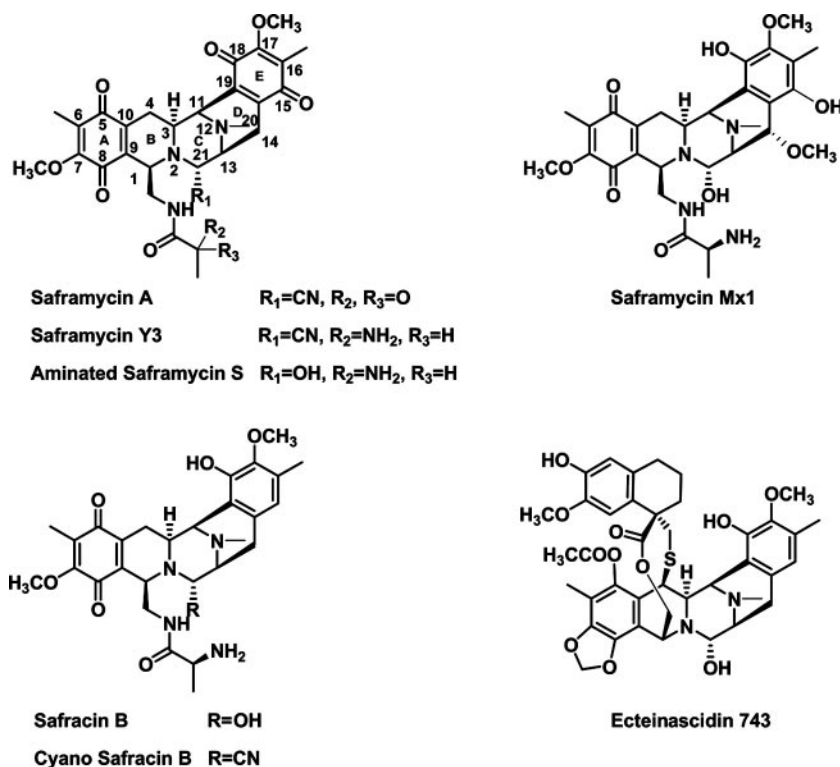


FIG. 1. Structures of saframycins, safracins, and ecteinasidin 743.

studies with SFM-A should provide a genetic and biochemical basis for rationally engineering these complex metabolites and serve as a starting point to access other tetrahydroisoquinoline natural products, such as ET743.

Previous feeding experiments using isotope-labeled substrates showed that the backbone of SFM-A is derived from one alanine (Ala), one glycine (Gly), and two tyrosine (Tyr) residues, suggesting that it is of tetrapeptide origin (28), which is also probably shared among other structurally related analogs, such as saframycin Mx1 (SFM-Mx1) and safracin B (SAC-B) as shown in Fig. 1. The partial biosynthetic gene cluster of SFM-Mx1 (with a hydroquinone form of the E ring, a hydroxy group at the C-21 position, and a reserved α -amino group of Ala in comparison to SFM-A) and the entire biosynthetic gene cluster of SAC-B (one of the structurally simplest members in the SFM family) were cloned from *Myxococcus xanthus* in 1995 (32, 33) and *Pseudomonas fluorescens* in 2005 (46), respectively, indeed revealing a nonribosomal peptide synthetase (NRPS) system for the formation of an identical tetrapeptide intermediate. In both cases, sequential incorporation of Ala, Gly, and Tyr derivatives into the backbone was speculated to be catalyzed by NRPSs in a colinear way according to the substrate specificity of the NRPS modules. This biosynthetic logic was formulated using bioinformatics analysis, but to date, no biochemical studies on SFM-Mx1 and SAC-B have been reported.

We hypothesized that SFM-A is biosynthesized in a manner similar to that of SFM-Mx1 and SAC-B according to their conserved structure. Here we report the cloning and sequencing of the SFM-A biosynthetic gene cluster from *S. lavendulae*

NRRL 11002 and propose biochemical functions for the deduced gene products. Sequence analysis and genetic comparison revealed a common strategy of NRPS-directed tetrapeptide assembly during the biosynthesis of SFM-A, SFM-Mx1, and SAC-B. However, in contrast to speculations from prior reports regarding SFM-Mx1 and SAC-B, we predicted that the same tetrapeptide backbone is catalyzed by NRPSs using the last module in an iterative manner rather than following the typical colinear rule. To confirm this prediction, we heterologously produced and purified proteins containing each adenylation domain of the NRPS modules and determined their substrate specificity using an ATP-PP_i exchange assay. Thus, these findings shed new insight into tetrahydroisoquinoline biosynthesis and afford the opportunity to study iterative events during nonribosomal peptide biosynthesis. Finally, production of SFM-Y3, an aminated analog of SFM-A, was achieved by heterologous expression of the hydroxylase SfmO4 in the SAC-B producer *P. fluorescens* FERM BP-14, demonstrating the feasibility of producing tetrahydroisoquinoline analogs by rationally engineering of an established biosynthetic pathway in microorganisms. The availability of the gene clusters and biosynthetic pathways of SFM-A, SFM-Mx1, and SAC-B has paved the way for future studies regarding the unusual biochemistry found in this pathway and subsequent attempts of applying this knowledge for combinatorial biosynthesis.

MATERIALS AND METHODS

Bacterial strains, plasmids, and reagents. Bacterial strains and plasmids used in this study are summarized in Table 1. Biochemicals, chemicals, media, restriction enzymes, and other molecular biological reagents were from standard commercial sources.

TABLE 1. Bacterial strains and plasmids used in this study

Strain or plasmid	Characteristic(s)	Source or reference
<i>E. coli</i> strains		
DH5 α	Host for general cloning	Invitrogen
CC118 (λ pir)	Host for general cloning	9
XL1-Blue MRF'	Host for constructing the genomic library	Stratagene
S17-1	Donor strain for conjugation between <i>E. coli</i> and <i>Streptomyces</i>	26
S17-1 (λ pir)	Donor strain for conjugation between <i>E. coli</i> and <i>Pseudomonas</i>	9
BL21(DE3)	Donor strain for protein overexpression	Novagen
<i>S. lavendulae</i> strains		
NRRL 11002	Wild-type strain, SFM-A producing	NRRL
TL2001	NRPS allele mutant of P1, SFM-A producing	This study
TL2002	NRPS allele mutant of P2, SFM-A nonproducing	This study
TL2003	Δ <i>sfmB</i> gene replacement mutant, SFM-A nonproducing	This study
TL2004	TL2003 derivative containing pTL2007, expression of <i>sfmB</i> under the control of <i>Perme*</i> promoter, SFM-A producing	This study
TL2005	<i>orf</i> (-1) gene disruption mutant, SFM-A producing	This study
TL2006	<i>orf</i> (+1) gene disruption mutant, SFM-A producing	This study
<i>P. fluorescens</i> strains		
FERM BP-14	Wild-type strain, SAC-B producing	FERM ^a
TL2101	FERM BP-14 containing pTL2011, expression of <i>sfmO</i> to <i>sfmK</i> under the control of the <i>Ptac</i> promoter	This study
TL2102	FERM BP-14 containing pTL2014, expression of <i>sfmO4</i> under the control of the <i>Ptac</i> promoter, aminated SFM-S producing	This study
Plasmids		
pGEM-T Easy	<i>E. coli</i> subcloning vector	Promega
pGEM-7zf	<i>E. coli</i> subcloning vector	Promega
pSP72	<i>E. coli</i> subcloning vector	Promega
pANT841	<i>E. coli</i> subcloning vector	GenBank (accession no. AF438749)
pVLT33	Heterologous expression vector in <i>P. fluorescens</i> A2-2	9
pTL2001	1.2-kb PCR product of the NRPS gene (P1) in pGEM-T Easy	This study
pTL2002	1.2-kb PCR product of the NRPS gene (P2) in pGEM-T Easy	This study
pTL2003	0.8-kb PCR product of the NRPS RE gene (P3) in pGEM-T Easy	This study
pTL2004	1.2-kb EcoRI/SpeI internal fragment of pTL2001 in pOJ260	This study
pTL2005	1.2-kb EcoRI/HindIII internal fragment of pTL2002 in pKC1139	This study
pTL2006	<i>sfmB</i> replacement construct in which <i>sfmB</i> was inactivated by insertion of <i>ermE</i>	This study
pTL2007	5.2-kb fragment containing <i>sfmB</i> under the control of <i>Perme*</i> in pTGV-4	This study
pTL2008	1.4-kb PCR fragment containing <i>sfmO3</i> to -K in pANT841	This study
pTL2009	1.4-kb NdeI/HindIII fragment containing <i>sfmO3</i> to -K in pET28a	This study
pTL2010	1.4-kb XbaI/HindIII fragment containing <i>sfmO3</i> to -K in pVLT33	This study
pTL2011	1.4-kb PCR fragment containing <i>sfmO4</i> in pANT841	This study
pTL2012	1.4-kb NdeI/HindIII fragment containing <i>sfmO4</i> in pET28a	This study
pTL2013	1.4-kb XbaI/HindIII fragment containing <i>sfmO4</i> in pVLT33	This study
pTL2014	2.0-kb EcoRI/HindIII fragment that encodes the N terminus of SfmA in pSP72	This study
pTL2015	5.5-kb fragment that encodes SfmA in pET28a	This study
pTL2016	3.4-kb fragment that encodes the truncated SfmA (C1-A1-PCP1) in pET28a	This study
pTL2017	0.6-kb EcoRI/PstI fragment that encodes the N terminus of SfmB in pSP72	This study
pTL2018	0.5-kb EcoRI/HindIII fragment that encodes the C terminus of SfmB in pSP72	This study
pTL2019	3.3-kb fragment that encodes SfmB in pET28a	This study
pTL2020	1.7-kb EcoRI/SphI fragment that encodes the N terminus of SfmC in pSP72	This study
pTL2021	0.2-kb XbaI/HindIII fragment that encodes the C terminus of SfmC in pGEM-7zf	This study
pTL2022	4.5-kb fragment that encodes SfmC in pET28a	This study
pTL2023	0.3-kb PCR product encoding the internal fragment of <i>orf</i> (-1) in pSP72	This study
pTL2024	0.4-kb PCR product encoding the internal fragment <i>orf</i> (+1) in pSP72	This study
pTL2025	0.3-kb EcoRI/XbaI internal fragment of pTL2023 in pKC1139	This study
pTL2026	0.4-kb EcoRI/XbaI internal fragment of pTL2024 in pKC1139	This study
pTL2101	<i>S. lavendulae</i> NRRL 11002 genomic library cosmid	This study
pTL2102	<i>S. lavendulae</i> NRRL 11002 genomic library cosmid	This study
pTL2103	<i>S. lavendulae</i> NRRL 11002 genomic library cosmid	This study

^a FERM, Fermentation Research Institute, Agency of Industrial Science and Technology, Japan.

DNA isolation, manipulation, and sequencing. DNA isolation and manipulation in *Escherichia coli* and *Streptomyces* were carried out according to standard methods (19, 37). PCR amplifications were carried out on an Eppendorf authorized thermal cycler (Eppendorf AG, Hamburg, Germany) using either *Taq* DNA polymerase or PfuUltra high-fidelity DNA polymerase. Primer synthesis

and DNA sequencing were performed at the Shanghai GeneCore Biotechnology, Inc., and Chinese National Human Genome Center.

Genomic library construction and screening. A genomic library of *S. lavendulae* NRRL 11002 was constructed in Super-Cos1 according to a previously published protocol (22). *E. coli* VCS257 and Gigapack III XL packaging extract

(Stratagene, La Jolla, CA) were used for library construction according to the manufacturer's instructions. The NRPS gene probes for library screening were obtained by PCR amplification and confirmed by sequencing. For PCR products P1 and P2, a 1.2-kb fragment was obtained by using the primers 5'-TACACGT CCGGCACSAACGCAARCCNAARGG-3' and 5'-AWCGAGKSGCCSGGG SMGAAGAA-3'. For PCR product P3, a 0.8-kb fragment was amplified by using the primers 5'-GACAACTTCTTCGAGCTGGGSGSSAYTC-3' and 5'-GCG GACCAACTTCTCCGSRCCAYTTRCT-3'. The genomic library (6.0×10^3 colonies) was screened by colony hybridization with P2 as a probe, and resultant positive clones were further confirmed by Southern hybridization with P2 and P3 as probes.

Sequence analysis. The open reading frames (ORFs) were deduced from the sequence by performing FramePlot 3.0beta program (<http://watson.nih.gov/jp/~jun/cgi-bin/frameplot-3.0b.pl>). The corresponding deduced proteins were compared with other known proteins in the databases by available BLAST methods (<http://www.ncbi.nlm.nih.gov/BLAST/>). Amino acid sequence alignments were performed by the CLUSTALW method, and the DRAWTREE and DRAWGRAM methods, respectively, from BiologyWorkBench 3.2 software (<http://workbench.sdsc.edu>). Prediction of amino acid specificity of individual NRPS A domains was performed by using the BLAST server provided at the website <http://bix.umbi.umd.edu/Projects/nrps/> (see Table S2 in the supplemental material).

Production, isolation, and analysis of SFM-A in *S. lavendulae*. *S. lavendulae* wild-type and recombinant strains were grown on ISP-2 (0.4% glucose, 0.4% yeast extract, and 1% malt extract [pH 7.2]) agar plates (with the appropriate antibiotic for recombinant strains) at 30°C for sporulation. For fermentation, 200 μ l of spore suspension (1.0×10^6 to 1.0×10^7 /ml) of each *S. lavendulae* strain was inoculated onto a YSA (0.1% yeast, 0.5% soluble starch, 1.5% agar, pH 7.5) plate and incubated at 27°C for 7 days. A piece of YSA with spores was then transferred into a 500-ml flask containing 50 ml of fermentation medium (0.1% glucose, 1% soluble starch, 0.5% NaCl, 0.1% K₂HPO₄, 0.5% casein acid hydrolysate, 0.5% meat extract, pH 7.0) and incubated at 27°C and 250 rpm for 30 to 36 h.

For SFM-A isolation, each 50 ml of the culture broth was filtered and adjusted to a pH of 6.8. After treatment with 1 mM KCN at 35°C for 30 min, the filtered broth was extracted thrice with 30 ml of ethyl acetate, and the combined extract was finally concentrated to 100 μ l in a vacuum.

High-performance liquid chromatography (HPLC) analysis was carried out on a Microsorb-MV 100-5 C₁₈ column (4.6 by 250 mm) (catalog no. SN 281505; Varian). The column was equilibrated with 50% solvent A (H₂O, 0.05% trifluoroacetic acid) and solvent B (CH₃CN, 0.05% trifluoroacetic acid) and developed with the following program: 0 to 5 min, 90% solvent A and 10% solvent B; 5 to 25 min, a linear gradient from 90% solvent A and 10% solvent B to 15% solvent A and 85% solvent B; 25 to 27 min, constant 15% solvent A and 85% solvent B; 27 to 29 min, a linear gradient from 15% solvent A and 85% solvent B to 90% solvent A and 10% solvent B; and 29 to 30 min, constant 90% solvent A and 10% solvent B. This was carried out at a flow rate of 1 ml/min and UV detection at 270 nm using an Agilent 1100 HPLC system (Agilent Technologies, Palo Alto, CA). The identity of a compound was confirmed by coinjection with standard SFM-A and liquid chromatography-mass spectrometry (LC-MS) analysis performed on an LCMS-2010 A liquid chromatograph mass spectrometer (Shimadzu, Japan) under the same conditions. SFM-A showed [M+H]⁺ ion at *m/z* 563.0, consistent with the molecular formula C₂₉H₃₀N₄O₈.

Production, isolation, and analysis of SACs and SFMs in *P. fluorescens*. For fermentation, 50 μ l of frozen vegetative stock of each *P. fluorescens* strain was transferred into a 250-ml flask containing 50 ml of YMP3 medium (1% glucose, 0.25% beef extract, 0.5% Bacto peptone, and 0.8% CaCO₃, pH 6.5). After incubation at 27°C for 30 h, 2% of this seed medium was then transferred into a 500-ml flask containing 100 ml of M-16B medium [15.2% D-mannitol, 3.5% dried brewer's yeast, 1.4% (NH₄)₂SO₄, 0.001% FeCl₃, and 2.6% CaCO₃, pH 6.5]. Strains were incubated at 27°C for 40 h, and the expression of *sfmO3* to *sfmK* or expression of *sfmO4* was induced by the addition of isopropyl- β -D-thiogalactopyranoside (IPTG) to a final concentration of 0.2 mM, and cultures were incubated at 24°C for an additional 72 to 96 h. The same fermentation procedure was applied to the wild-type strain as a compatible control.

For the production of cyano-substituted analogs, the supernatants of fermentation cultures were treated with 1 mM KCN at 35°C for 30 min.

The analysis on each filtered culture was carried out by using the same set of HPLC and LC-MS conditions as those for SFM-A detection described above. The column was equilibrated with 50% solvent A (10 mM NH₄Ac, 1% diethanolamine, pH 4.0) and solvent B (CH₃CN), and was developed with the following program: 0 to 5 min, 93% solvent A and 7% solvent B; 5 to 25 min, a linear gradient from 93% solvent A and 7% solvent B to 15% solvent A and 85%

solvent B; 25 to 27 min, constant 15% solvent A and 85% solvent B; 27 to 29 min, a linear gradient from 15% solvent A and 85% solvent B to 93% solvent A and 7% solvent B; and 29 to 30 min, constant 93% solvent A and 7% solvent B. This was carried out at a flow rate of 1 ml/min and UV detection at 268 nm. The identity of compound was confirmed by LC-MS analysis under the same conditions. SAC-B showed [M+H]⁺ ion at *m/z* 541.2, consistent with the molecular formula C₂₈H₃₆N₄O₇. Cyano-substituted SAC-B showed [M+H]⁺ ion at *m/z* 550.2, consistent with the molecular formula C₂₉H₃₅N₅O₆. Aminated SFM-S showed [M+H]⁺ ion at *m/z* 555.2, consistent with the molecular formula C₂₈H₃₄N₄O₈. SFM-Y3 showed [M+H]⁺ ion at *m/z* 564.0, consistent with the molecular formula C₂₉H₃₃N₅O₇.

Determination of substrate specificities of SfmA, SfmB, and SfmC. For amino acid-dependent ATP-PP_i exchange assay, a typical reaction (100 μ l) was carried out in 75 mM Tris-HCl (pH 8.0) buffer, containing 50 to 100 nM NRPS protein, 5 mM ATP, 1.0 mM PP_i, with 0.5 μ Ci of ³²P-ATP (40.02 Ci/mmol; NEN Life Science Products, Boston, MA), 10 mM MgCl₂, 5.0 mM dithiothreitol, and a 1.0 mM concentration of each of various amino acids. After incubation at 30°C for 30 min, each assay was stopped by the addition of 0.5 ml of 1% (wt/vol) activated charcoal in 4.5% (wt/vol) tetrasodium pyrophosphate and 3.5% (vol/vol) perchloric acid. The precipitate was collected on a glass-fiber filter (2.4 cm) (G-4; Fisher, Pittsburgh, PA), washed successively with 10 ml of 40 mM sodium pyrophosphate plus 1.4% perchloric acid, 10 ml of water, and 5 ml of 95% ethanol and briefly dried in air. The filter was mixed with 5 ml of scintillation fluid (ScintiSafe gel; Fisher) and counted on a Beckman LS-6800 scintillation counter to determine the radioactivity.

Supplemental material. See the supplemental material for supporting data including deduced function of open reading frames beyond the SFM-A biosynthetic gene cluster boundary, prediction of amino acid recognitions of NRPS A domains, structures of SFMs, inactivation and complementation in *S. lavendulae*, biotransformation in *P. fluorescens*, overexpression and purification of SFM NRPSs, and chemical synthesis of Tyr derivatives in this study.

Nucleotide sequence accession number. The sequence reported in this paper has been deposited into GenBank under the accession number DQ838002.

RESULTS

Cloning and sequencing of the SFM-A gene cluster from *S. lavendulae* NRRL 11002. NRPSs catalyze the assembly of non-ribosomal peptides from proteinogenic and nonproteinogenic amino acids and usually possess a multimodular structure (42). Each module consists minimally of an adenylation domain (A domain) responsible for amino acid activation; a peptidyl carrier protein (PCP) domain, which usually resides adjacent to the A domain, for thioesterification of the activated amino acid; and a condensation domain (C domain) for transpeptidation between the upstream and downstream peptidyl and amino acyl thioesters to elongate the growing peptide chain. Based on the high sequence similarity among various A and PCP domains in the database, two conserved motifs (YTSGT TGKPKG in A domains and FFXLGGXSX in PCP domains) were used to design a pair of degenerate primers to clone the putative SFM-A NRPS genes by PCR. With the genomic DNA of *S. lavendulae* as the template, a distinct product with the expected size of 1.2 kb was readily amplified. Sequencing and analyzing of the selected clones revealed two gene sequences (P1 and P2), both of which are highly similar to those of known NRPS genes. To determine their roles in SFM-A biosynthesis, we set out to inactivate the target alleles in *S. lavendulae*. While the P1 allele mutant strain TL2001 retained the ability to produce SFM-A and even produced SFM-A to a level higher than that of the wild-type strain, the P2 allele mutant strain TL2002 completely lost the ability to produce SFM-A, confirming that the P2-contained NRPS gene is involved in SFM-A biosynthesis (data not shown). As a result, after screening approximately 6×10^3 clones of the *S. lavendulae* genomic library with the 1.2-kb P2 fragment as a probe, we

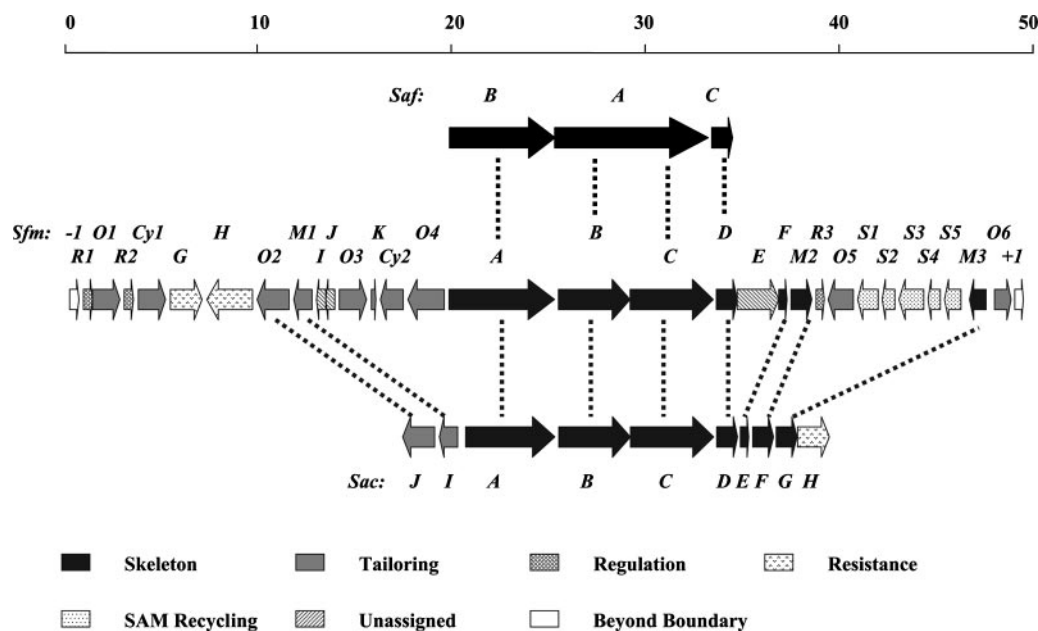


FIG. 2. Genetic organization and comparison of the *saf* (SFM-Mx1), *sfm* (SFM-A), and *sac* (SAC-B) biosynthetic gene clusters. The proposed functions of individual ORFs are shown and summarized in Table 1.

isolated 20 overlapping cosmids that span a 60- to 65-kb DNA region, as exemplified by pTL2101, pTL2102, and pTL2103 (see Fig. S2 in the supplemental material).

Previous studies on SFM-Mx1 and SAC-B biosynthesis revealed a relatively rare reductase (RE) domain that contains a NAD(P)H binding site at the C-terminal ends of SafA and SacC (33, 46), both of which are NRPSs involved in the tetrapeptidyl backbone formation. These RE domains may act on the PCP-tethered polypeptidyl intermediate and reductively release it from the PCP as a linear aldehyde (21), instead of the thioesterase (TE) functionality at the C-terminal ends of typical NRPSs. Since the RE domain is conserved and always resides next to the last PCP, we adopted an alternative strategy to specifically clone the putative RE gene fragment by PCR according to two motifs [DNFFEL(G/D)GHS in the PCP domain and RVLKEAVWKS in the RE domain]. A single product with the expected size of 0.8 kb was readily amplified and cloned from the genomic DNA of *S. lavendulae*. Sequence analysis of the randomly selected clones confirmed that 80% of them contain an identical product, P3, and the sequence of P3 exhibits significant similarity to the sequences of the RE domains of the *safA* and *sacC* genes. To identify the locus on the chromosome, Southern analysis of the previously identified cosmids obtained by P2 screening was performed using the 0.8-kb P3 fragment as a probe. Intriguingly, most cosmids showed a positive signal, and a single 7.4-kb BamHI fragment that harbors both P2 and P3 was detected. All together, these results provided strong support that we have cloned the SFM-A gene cluster in *S. lavendulae*.

The DNA region represented by cosmid pTL2101 (partial), pTL2102 (entire), and pTL2103 (partial) was selected for sequencing, yielding a 62,804-bp contiguous sequence with 71.86% of the overall GC content that is characteristic for *Streptomyces* DNA. Bioinformatic analysis of the sequenced

region revealed 47 ORFs, and 30 of the ORFs from *sfmR1* to *sfmO6* were proposed to constitute the SFM-A gene cluster according to functional assignment of their deduced products and genetic comparison with the gene cluster for SAC-B biosynthesis (Fig. 2 and Table 2; see also Table S1 in the supplemental material). Mutant strains with inactivated *orf(-1)* and *orf(+1)* retained the ability to produce SFM-A as deduced from HPLC analysis (see Fig. S3 in the supplemental material), supporting the idea that they are outside the *sfm* gene cluster. Consistent with the structure of SFM-A, the ORFs within the *sfm* cluster presumably include eight genes encoding enzymes involved in the biosynthesis of the SFM-A tetrapeptidyl backbone, nine genes encoding the tailoring enzymes, as well as three regulatory genes, two resistance genes, five genes involved in *S*-adenosylmethionine (SAM) recycling, and three additional genes whose functions could not be predicted or assigned for SFM-A production.

Genes encoding NRPSs and associated enzymes for the biosynthesis of the core. Three NRPS genes, *sfmA*, *sfmB* and *sfmC*, were identified within the *sfm* cluster (Fig. 2). As shown in Fig. 3B (also see Fig. 6A), their deduced products constitute an NRPS system that contains a set of characteristic domains arranged as follows: acyl coenzyme A ligase (AL)-PCP0-C1-A1-PCP1-C2-A2-PCP2-C3-A3-PCP3-RE. The products are similar from head to tail in both domain organization and amino acid sequence to those for SFM-Mx1 and SAC-B biosynthesis (see Fig. 6B and C [the SAC-B NRPS system lacks the first module AL-PCP0]). The *sfmB* mutant strain TL2003 completely lost its ability to produce SFM-A (Fig. 4A, panel III), which was restored by expressing *sfmB* in *trans* (Fig. 4A, panel IV), confirming the essential role of this NRPS system for SFM-A biosynthesis. The first module of SfmA resembles a family of NRPS N-terminal modules, which consist of a domain with high similarity to ALs and an PCP-like domain,

TABLE 2. Deduced functions of ORFs in the SFM-A biosynthetic gene cluster

Gene	Size ^a	Protein homolog ^b
<i>orf</i> (-1 to -10)		Beyond the <i>sfm</i> cluster boundary
<i>sfmR1</i>	195	StropDRAFT_1652 (ZP_01431206; 57/71), TetR type regulatory protein from <i>Salinispora tropica</i> CNB-440
<i>sfmO1</i>	304	SAMR0789 (CAJ88498; 62/75), putative oxidoreductase from <i>Streptomyces ambofaciens</i> ATCC 23877
<i>sfmR2</i>	169	MitQ (AAD28455; 63/78), putative regulatory protein in mitomycin biosynthesis
<i>sfmCy1</i>	512	MitR (AAD28454; 67/78), similar to mitomycin C oxidase McrA
<i>sfmG</i>	479	DQ915964.1 (ABL09967; 37/56), AraJ-like transmembrane efflux protein from <i>Streptomyces echinatus</i>
<i>sfmH</i>	819	CmrX (CAE17542; 57/72), UV repair protein in chromomycin biosynthesis
<i>sfmO2</i>	521	SacJ (AAL33754; 43/58), putative monooxygenase/hydroxylase in SAC-B biosynthesis
<i>sfmM1</i>	199	SacI (AAL33755; 43/56), SAM-dependent methyltransferase in SAC-B biosynthesis
<i>sfmI</i>	160	MflvDRAFT_0798 (ZP_01194885; 35/46), unknown protein from <i>Mycobacterium flavescens</i> PYR-GCK
<i>sfmJ</i>	165	MflvDRAFT_0799 (ZP_01194886; 41/58), unknown protein from <i>M. flavescens</i> PYR-GCK
<i>sfmO3</i>	395	Orf3 (AAD28449; 50/67), cytochrome P450 hydroxylase from <i>S. lavendulae</i>
<i>sfmK</i>	61	Fas2 (P46374, 44/57), ferredoxin-like protein from <i>Rhodococcus fascians</i>
<i>sfmCy2</i>	504	MitR (AAD28454; 45/60), similar to mitomycin C oxidase McrA
<i>sfmO4</i>	475	HctH (AA42400; 25/45), cytochrome P450 monooxygenase from <i>Lynghya majuscula</i>
<i>sfmA</i>	1,836	SafB (AAC44128; 38/51), NRPS in SFM-Mx1 biosynthesis
<i>sfmB</i>	1,082	SacB (AAL33757; 42/58), NRPS in SAC-B biosynthesis
<i>sfmC</i>	1,485	SacC (AAL33758; 47/60), NRPS in SAC-B biosynthesis
<i>sfmD</i>	365	SacD (AAL33759; 38/51), hydroxylase in SAC-B biosynthesis
<i>sfmE</i>	789	Orf (-15) (AAN85499; 32/43), putative peptidase from <i>Streptomyces atroolivaceus</i>
<i>sfmF</i>	73	SacE (AAL33760; 60/68), protein containing MbtH-like domain in SAC-B biosynthesis
<i>sfmM2</i>	366	SacF (AAL33761, 63/76), SAM-dependent methyltransferase in SAC-B biosynthesis
<i>sfmR3</i>	178	MitQ (AAD28455; 51/69), regulatory protein in mitomycin biosynthesis
<i>sfmO5</i>	389	ComPD (AAK81837; 42/52), prephenate dehydrogenase in complestatin biosynthesis
<i>sfmS1</i>	484	Fnq16 (CAL34094; 91/95), putative adenosylhomocysteinase from <i>Streptomyces cinnamomensis</i>
<i>sfmS2</i>	313	Fnq15 (CAL34093, 76/85), putative 5,10-methylene-tetrahydrofolate reductase from <i>S. cinnamomensis</i>
<i>sfmS3</i>	1,160	Fnq14 (CAL34092; 86/92), putative methionine synthase from <i>S. cinnamomensis</i>
<i>sfmS4</i>	327	Fnq13 (CAL34091; 67/76), putative adenosine kinase from <i>S. cinnamomensis</i>
<i>sfmS5</i>	401	Fnq12 (CAL34090, 85/91), putative S-adenosylmethionine synthase from <i>S. cinnamomensis</i>
<i>sfmM3</i>	334	SacG (AAL33762; 43/62), SAM-dependent methyltransferase in SAC-B biosynthesis
<i>sfmO6</i>	376	MtmOII (CAK50777; 37/46), FAD-dependent oxygenase in mithramycin biosynthesis
<i>orf</i> (+1 to +10)		Beyond the <i>sfm</i> cluster boundary

^a The size of each protein is shown as the number of amino acids.

^b NCBI accession numbers and percent identity/percent similarity are given in parentheses.

such as BlmVI NRPS-5 (40% identity) that is presumably involved in the biosynthesis of the β -aminoalaniamide moiety of bleomycin as a starter module (10). Although its function on initiation of the peptidyl intermediate biosynthesis remains unclear, sequence alignment revealed that the SfmA AL-like domain lacks most of the conserved motifs of A domains, supporting that the first module of SfmA does not function as a typical NRPS module to incorporate amino acid residues into the tetrapeptidyl skeleton. To predict the substrate of the SFM-A NRPS system, the eight specificity-conferring codes for each A domain were identified as follows by sequence alignment with the A domain of PheA (gramicidin synthase A) (6, 44) (see Table S2 in the supplemental material): DLFNNALT for SfmA-A1 (100% identity to those for SafB-A1 and SacA-A1), DILQLGLI for SfmB-A2 (87.5% identity to those for SafA-A2 and 100% identity to those for SacB-A2), and DPWGLGLI for SfmC-A3 (100% identity to those for SafA-A3 and SacC-A3). Among the SFM-A, SFM-Mx1, and SAC-B NRPS systems, the identity of these eight codes for each group of A domains (with the exception that the Val at residue position 330 in SafA-A2 is replaced by Ile in SfmB-A2 or SacB-A2) indicates that they recognize and activate a same amino acid substrate. Together with the same order of modules for substrate incorporation (see Fig. 6), a common strategy of assembling the tetrapeptidyl intermediate is proposed to occur in SFM-A, SFM-Mx1, and SAC-B biosynthesis. Subsequently,

SfmA-A1, SfmB-A2, and SfmC-A3 were predicted to activate Ala or Gly, Gly or 3-hydroxy-5-methy-O-methyltyrosine (3h5mOmTyr), and 3h5mOmTyr, respectively, by an analysis program provided at the website <http://bix.umbi.umd.edu/Projects/nrps/>. Further biochemical determination of substrate specificities of these NRPS A domains in vitro and mechanistic analysis of this NRPS system for the tetrapeptidyl intermediate assembly are described below. The RE domain of SfmC exhibits high similarity to a few NRPS C-terminal reductase domains that release peptidyl intermediates from NRPSs as reductive products including aldehydes, alcohols, and macrocyclic imines. Very recently, the terminal reductase domain of NcpB, an NRPS for nostocyclopeptide biosynthesis, was biochemically characterized to catalyze the reductive release of the matured peptide chain as an aldehyde and then trigger the spontaneous formation of the imino head-to-tail linkage (21), instead of the more commonly found TE domains for hydrolysis, lactamization, or lactonization. In a mechanistic analogy, SfmC-RE may catalyze the reductive release of the resulting tetrapeptidyl intermediate tethered to SfmC-PCP3 as an aldehyde and then trigger the spontaneous intramolecular cyclization to close the C ring (Fig. 3B).

Four genes, *sfmD*, *sfmF*, *sfmM2*, and *sfmM3*, are proposed to encode the enzymes involved in biosynthesis of the nonproteinogenic amino acid 3h5mOmTyr (Fig. 3A), the pathway of which was established in SAC-B biosynthesis by heterologous

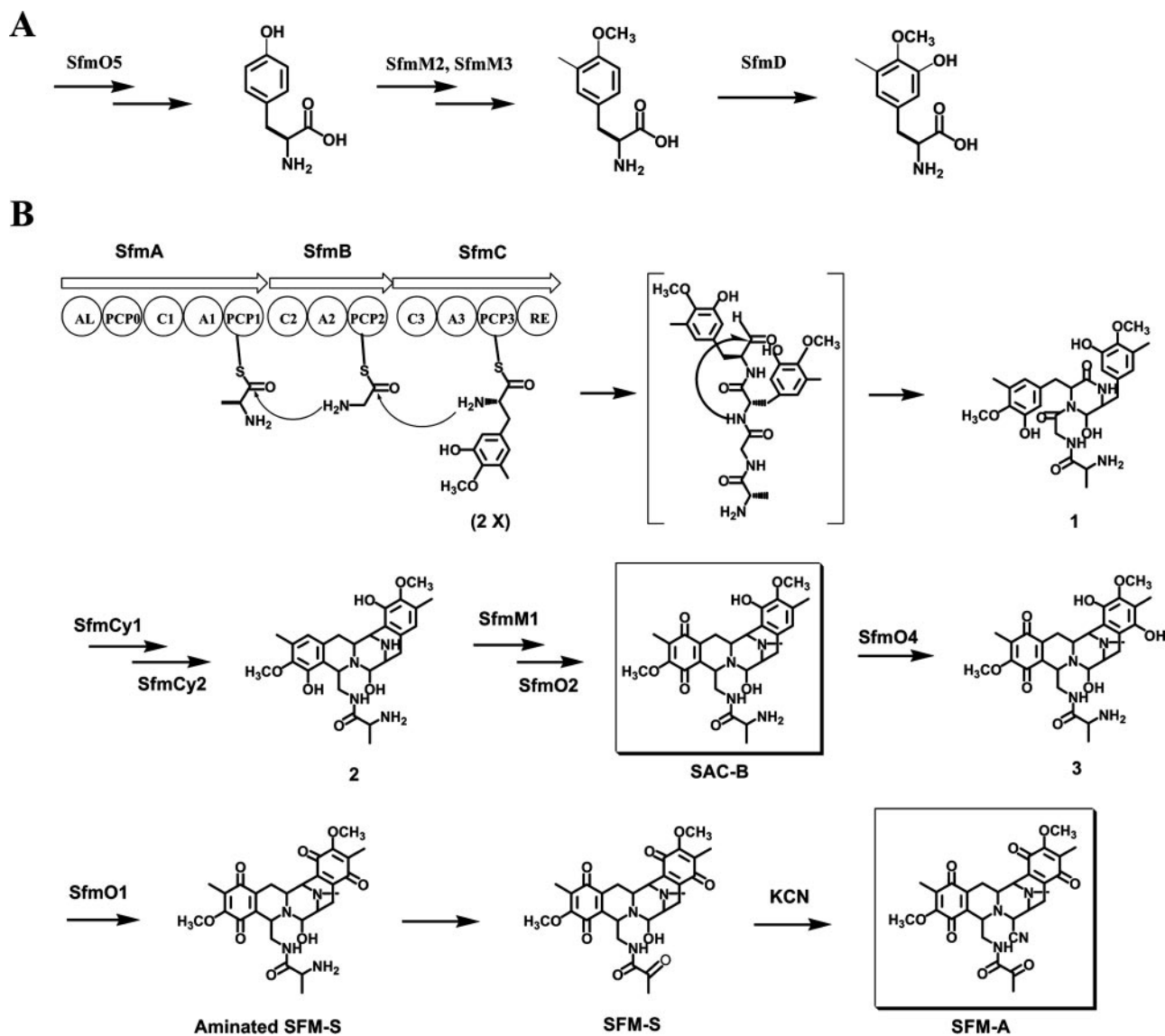


FIG. 3. Proposed biosynthetic pathways for 3-hydroxy-5-methoxytyrosine (A) and saframycin A (B).

expression of homologous genes and cocultivation among mutant strains (46). SfmO5, closely related to the prephenate dehydrogenases (e.g., ComPD in complestatin biosynthesis [7]; 42% identity) that catalyze the *p*-hydroxyphenylpyruvate formation, might enhance the biosynthesis of Tyr that serves as the precursor of 3h5mOmTyr. SfmF, with high similarity to SacE (60% identity), belongs to a family of MbtH-like proteins that contain three fully conserved tryptophan (Trp) residues. Most members of this family are found in known antibiotic gene clusters, such as VioN (60% identity) in viomycin biosynthesis (45); however, their roles remain to be established. SfmM2 exhibits high sequence similarity to SacF (63% identity) and likely functions as a *C*-methyltransferase to introduce a methyl group at the C-3 position of Tyr. SfmM3 exhibits high sequence homology to SacG (43% identity) and various *O*-methyltransferases (e.g., CalO6 in calicheamicin biosynthesis [1]; 36% identity), supporting its role for *O* methylation at the

C-4 position. SfmD, with no other homologous proteins found in the database, exhibits relatively high sequence similarity to SacD (38% identity), the function of which was deduced to be responsible for the hydroxyl group substitution at C-5 to convert 3-methyl-*O*-methyltyrosine (3mOmTyr) into 3h5mOmTyr.

Genes encoding tailoring enzymes. Postmodifications on the tetrapeptidyl intermediate compound 1, including cyclization, methylation, oxidoreduction, and nitrile moiety substitution, are postulated to proceed with a set of tailoring enzymes in the SFM-A biosynthetic pathway as outlined in Fig. 3B. The gene products of *sfmCy1* and *sfmCy2* exhibit high sequence similarity to MitR (67% and 45% identity, respectively), which might be responsible for the C8a-C9 bond formation in mitomycin biosynthesis (24). In a mechanistic analogy, SfmCy1 and SfmCy2 may act on compound 1 as cyclases to close the B and D rings at C9-C1 and C19-C11, although their regiospecificities need to be determined. Noticeably, such homologous genes

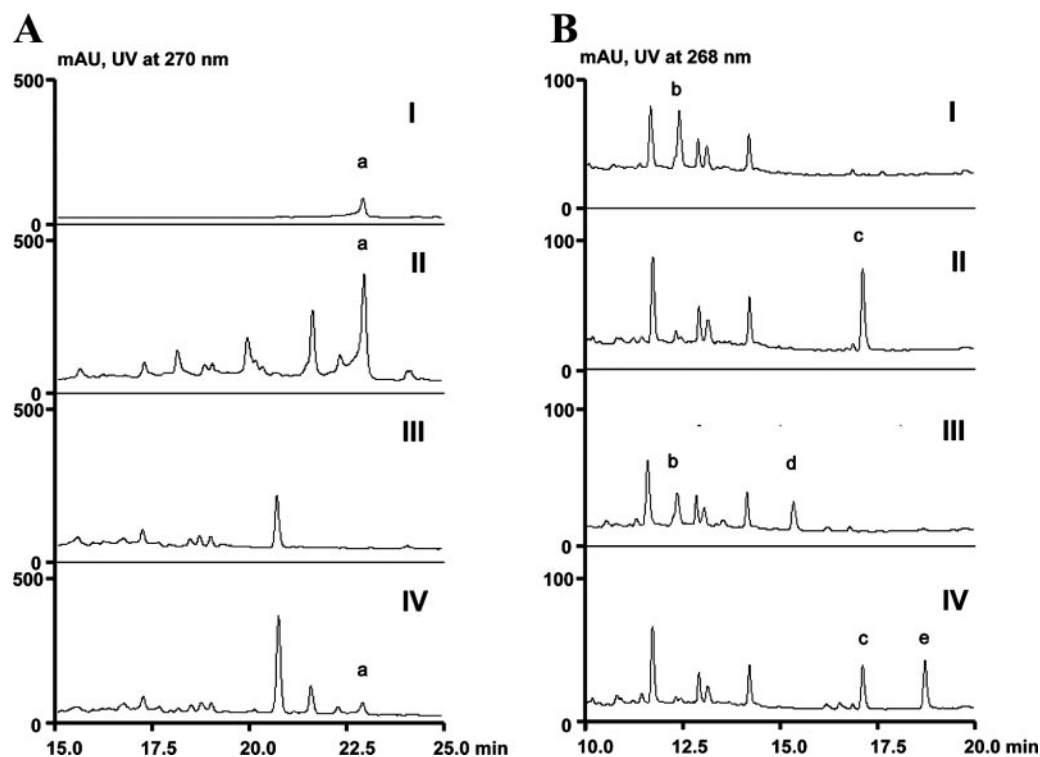


FIG. 4. HPLC analysis of saframycin and safracins. (A) Saframycin A (a) isolated from an authentic standard (I), *S. lavendulae* wild-type strain (II), strain TL2003 ($\Delta sfmB$) (III), and strain TL2004 (*PerME*::sfmB*) (IV). (B) Safracin B (b) isolated from a *P. fluorescens* wild-type strain (I), cyano-substituted safracin B (c) isolated from a *P. fluorescens* wild-type strain treated with 1 mM KCN (II), aminated saframycin S (d) isolated from strain TL2102 (*Ptac::sfmO4*) (III), and saframycin Y3 (e) isolated from strain TL2102 (*Ptac::sfmO4*) treated with 1 mM KCN (IV). Absorbance of UV at 270 nm (A) or 268 nm (B) in milliabsorbance units (mAU) is shown on the y axes, and time (in minutes) is shown on the x axes.

have not been identified within the SAC-B biosynthetic gene cluster in *P. fluorescens*. It would be interesting to further determine whether these homologs can be identified in the *P. fluorescens* chromosome or whether they have any function regarding the formation of the B and E rings.

sfmM1 and *sfmO2* encode proteins that have high sequence similarity to SacI (43% identity) and SacJ (43% identity), respectively, both of which have been functionally assigned to catalyze the last two steps for SAC-B biosynthesis on the basis of identifying shunt metabolites resulted from *sacI* or *sacJ* inactivation (46). Although the catalytic order remains to be established, it is likely that SfmM1 acts as a *N*-methyltransferase to introduce a methyl group at the N-12 position, and SfmO2 serves as a monooxygenase responsible for hydroxylation at the C-5 position on the A ring, which then undergoes a dehydrogenation to form the quinone ring of SAC-B. Based on this hypothesis, interestingly, SAC-B might be a key intermediate in the SFM-A biosynthetic pathway (Fig. 3B).

SFM-A structurally differs from SAC-B with a heavily oxidized E ring. Heterologous expression of SfmO4 in the SAC-B producer resulted in aminated SFM-S production (described below), supporting the hypothesis that SfmO4 acts on SAC-B at the C-15 position for a hydroxyl substitution. *sfmO1*, encoding a putative NAD(P)⁺-dependent oxidoreductase, might catalyze oxidation of the resultant hydroxyl on the E ring in vivo, producing the characteristic bisquinone core scaffold (Fig. 3B). The further desamination of Ala and substitution of a nitrile moiety at the C-21 position are proposed to yield SFM-A.

Alternatively, it could not be excluded that the oxidative desamination step, which is predicted to be catalyzed by a putative FAD-dependent monooxygenase SfmO6, occurs at an earlier stage during the tailoring process. Since previous studies showed that treatment of SFM-S (an SFM-A precursor with a hydroxyl group instead of a nitrile moiety) with sodium cyanide led to the formation of SFM-A (2), the substitution of a nitrile moiety may be spontaneous, consisting with no obvious gene candidate identified within the *sfmA* cluster.

The *S. lavendulae* wild-type strain also produces a series of SFM derivatives with additional oxidation or O methylation at the C-14 position (as shown as a hydroxyl, methoxy, or keto group in Fig. 1) (40), suggesting that the branched biosynthetic pathways may start with the intermediate compound 3 and aminated SFM-S (or their desamino derivatives). The putative cytochrome P450 enzyme SfmO3 (coupled with the ferredoxin-like protein SfmK) presumably initiates the oxidative bioconversion at this position.

Genes encoding regulation, resistance, and other functions. Three genes (Fig. 2B), *sfmR1*, *sfmR2*, and *sfmR3*, are presumed to encode pathway-specific regulatory proteins. While SfmR1 resembles the TetR family of transcriptional regulators widely found in many microorganisms, SfmR2 and SfmR3, with high similarity to MitQ (63% and 51% identity, respectively) in the mitomycin biosynthetic pathway (24), belong to the OmpR family of DNA binding regulators in the two-component system.

Two resistance genes (Fig. 2B), *sfmG* and *sfmH*, were found

in the *sfm* cluster. In contrast to SfmG that belongs to a family of transmembrane efflux permeases that usually exhibits multiple drug resistance, such as AraJ (37% identity) in the aranciamycin biosynthetic pathway (41), SfmH shows high sequence similarity to a family of UV repair proteins, such as CmrX (57% identity) in the chromomycin biosynthetic pathway (27), representing a more specific resistance protein in agreement with the mechanism of action of SFM-A as a DNA alkylation agent.

Sequence analysis within the *sfm* cluster revealed five genes (Fig. 2B), *sfmS1* to *sfmS5* as a complete set for the recycling of SAM from *S*-adenosylhomocysteine (SAH, a by-product in the SAM-dependent methylation reaction). SfmS1, a putative *S*-adenosyl-L-homocysteine hydrolase, may cleave SAH to adenosine and homocysteine. The latter could be methylated and converted into methionine by SfmS2, a putative methionine synthase, with *N*⁵-methyl tetrahydrofolate (*N*⁵-methyl THF) as the cosubstrate. SfmS5, closely related to a family of SAM synthetases, might be responsible for the generation of SAM from methionine and ATP. *N*⁵-methyl THF as a methyl donor originates from *N*⁵,*N*¹⁰-methylene THF, requiring SfmS2 as a putative *N*⁵,*N*¹⁰-methylene THF reductase. SfmS4 shows high sequence similarity to a family of adenosine kinases, presumably in charge of ATP regeneration by converting adenosine to AMP. The pathway for recycling SAH to SAM has been well established in primary metabolism and recently was identified to be involved in a few biosynthetic pathways for secondary metabolites, such as the polyketide-isoprenoid compound furanonephthoquinone I (16). Since multiple SAM-dependent methylations at C, O, and N positions occur in the SFM-A biosynthetic process, the advantage of involvement of this complete pathway might facilitate enhancement of the supply of the SAM precursor.

Finally, three genes within the *sfm* cluster could not be functionally assigned on the basis of sequence analysis alone (Fig. 2B). *sfmE* encodes a protein that resembles proteins in the peptidase M28 family. The deduced products of two coupled genes, *sfmI* and *sfmJ*, exhibit high sequence similarity to MfvDRAFT_0798 (35% identity) and MfvDRAFT_0799 (41% identity), respectively, and the genes encoding MfvDRAFT_0798 and MfvDRAFT_0799 are clustered within the genome of *Mycobacterium flavescens* PYR-GCK (under the NCBI accession number NZ_AAPA01000017). Although SfmJ contains a putative pyridoxamine 5'-phosphate oxidase domain, their roles in SFM-A biosynthesis could not be speculated.

Determination of substrate specificities of SfmA, SfmB, and SfmC by utilizing amino acid-dependent ATP-PP_i exchange assay. Initial attempts to directly determine substrate specificities of individual A domains (i.e., SfmA-A1, SfmB-A2, and SfmC-A3) or intact SfmA from the SFM-A NRPS system were hampered by either poor solubility of the resultant proteins in *E. coli* or low enzymatic activities (data not shown). Thus, truncated SfmA (C1-A1-PCP1) and the intact SfmB (C2-A2-PCP2) and SfmC (C3-A3-PCP3-RE) were heterologously expressed in *E. coli* by using pET28a yielding soluble N-terminal His-tagged proteins with a yield around 5 to 10 mg/liter. Using nickel affinity chromatography, gel filtration, or anion-exchange chromatography in tandem, all proteins were purified to near homogeneity as shown in Fig. 5A, and sodium dodecyl

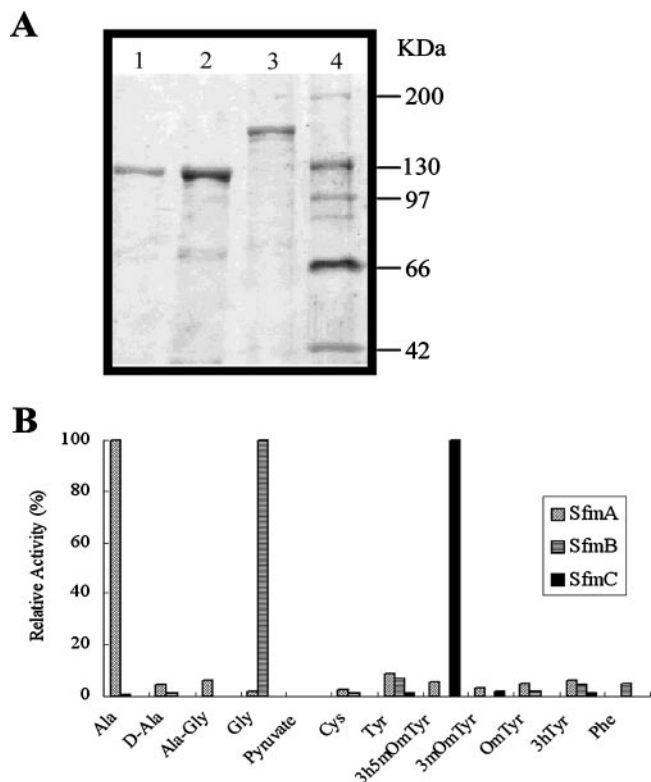


FIG. 5. (A) Purified SfmA-C1-A1-PCP1 (lane 1), SfmB-C2-A2-PCP2 (lane 2), and SfmC-C3-A3-PCP3-RE (lane 3) as analyzed by electrophoresis on a 7.5% sodium dodecyl sulfate-polyacrylamide gel. The positions of molecular mass markers (lane 4) (in kilodaltons) are shown to the right of the gel. (B) Substrate specificities as determined by the ATP-PP_i exchange reaction with the amino acids predicted to be incorporated into SFM-A (100% relative activity corresponds to 51,540 cpm for SfmA-C1-A1-PCP1, 44,300 cpm for SfmB-C2-A2-PCP2, and 51,500 cpm for SfmC-C3-A3-PCP3-RE).

sulfate-polyacrylamide gel electrophoresis revealed a dominant band consistent with their deduced molecular masses (124 kDa, 119 kDa, and 163 kDa, respectively). Among the substrates used in the ATP-PP_i exchange assay (L-Ala, D-Ala, L-Ala-L-Gly, L-Gly, pyruvate, L-Cys, L-Tyr, L-3h5mOmTyr, L-3mOmTyr, L-OmTyr, L-3hTyr, and L-Phe), as shown in Fig. 5B, truncated SfmA (C1-A1-PCP1), SfmB (C2-A2-PCP2), and SfmC (C3-A3-PCP3-RE) specifically recognized and activated L-Ala, L-Gly, and L-3h5mOmTyr, respectively.

Construction of the SFM-Y3 biosynthetic pathway in the SAC-B producer. According to the proposed SFM-A biosynthetic pathway (Fig. 3B), SAC-B, originally produced by *P. fluorescens*, may serve as a key intermediate. Two putative cytochrome P450 hydroxylase genes, *sfmO3* and *sfmO4*, were identified within the *sfm* cluster, and one of these genes encodes a protein that may be involved in further oxidation of the E ring. To verify this hypothesis, the constructs that carry *sfmO3* to *sfmK* (SfmK is a ferredoxin-like protein as a putative electron donor) and *sfmO4* were individually introduced into the SAC-B-producing strain, yielding the recombinant strains TL2101 and TL2102, respectively. Using the *P. fluorescens* wild-type strain as a control, strains TL2101 and TL2102 were cultured, and the resulting compounds were detected by HPLC

analysis. While the culture of strain TL2101 exhibited an HPLC profile similar to that of the wild-type strain (Fig. 4B, panel I), the production of SAC-B in TL2102 decreased. Accordingly, it was partially transformed into a new compound (Fig. 4B, panel III). LC-MS analysis revealed this compound with an $[M+H]^+$ ion at m/z 555.2, consistent with the molecular formula $C_{28}H_{34}N_4O_8$. By comparison with SAC-B and SFM-A, it was deduced to be an aminated SFM-S that contains a heavily oxidized quinone E ring. Thus, SfmO4 may act at C-15 position and yield the hydroquinone derivative compound 3, which may not be stable and is rapidly converted into aminated SFM-S during the purification process. To further confirm this oxidation step, the cultures of wild-type and mutant strains were treated with potassium cyanide at an earlier stage of purification. Upon HPLC (Fig. 4B, panels II and IV) and LC-MS analyses, as we expected, SAC-B from the wild-type strain and aminated SFM-S from strain TL2013 were correspondingly transformed into cyano-SAC-B and SFM-Y3 (the structure of the latter was further supported by tandem MS spectrometry analysis. See the supplemental material for detailed information), respectively. SFM-Y3 that is distinct from SFM-A only by a preserved amino group on Ala was previously found in the *S. lavendulae* strain culture supplemented with additional Ala and Gly or Ala-Gly dipeptide as substrates (Fig. 1) (4). These results not only clearly confirmed the function of SfmO4 but also showed that the SAC-B biosynthetic machinery in *P. fluorescens* is amenable to be rationally engineered for the production of structurally more complex analogs by introducing additional tailoring genes from the SFM-A biosynthetic pathway.

DISCUSSION

SFM-A is a bisquinone alkaloid that has significant antiproliferative activity. Previous feeding experiments indicated that the skeleton of SFM-A originates from a highly modified tetrapeptide (28), and studies with the structurally related compounds, SFM-Mx1 (33) and SAC-B (46), revealed that the biosynthesis of these compounds is mediated by an NRPS system (Fig. 1). Based on the assumption that SFM-A is biosynthesized in a similar manner, we attempted to clone the *sfm* cluster by amplifying putative NRPS gene fragments by PCR from *S. lavendulae* NRRL 11002. Using the PCR-amplified fragment P2 (obtained with the general NRPS primer set designed according to the conserved motifs of A and PCP domains) and P3 (obtained with the specific NRPS primer set designed according to the conserved motifs of PCP and RE domains) as probes, we screened the genomic library of *S. lavendulae* and identified the NRPS gene locus. Genetic analysis of a sequenced 62,804-bp DNA region revealed 47 ORFs, 30 of which are proposed to constitute the *sfm* gene cluster that contains a set of unusual NRPS genes with numerous novel genes for the subsequent tailoring steps to produce SFM-A and its analogs (Fig. 2 and 3B). The inactivation of NRPS gene *sfmB* completely abolished the ability to produce SFM-A, and subsequent complementation of this mutation by expressing *sfmB* in *trans* restored the production of SFM-A, unambiguously confirming that the cloned gene cluster is essential for SFM-A biosynthesis (Fig. 4A).

SfmA, SfmB, and SfmC constitute an NRPS system that

exhibits similarities in domain organization and amino sequence from head to tail to those for SFM-Mx1 and SAC-B biosynthesis (Fig. 6). With the exception of the first module that is not found in the SAC-B NRPS system, the eight specificity-conferring amino acids for each A domain in the remaining three modules are almost identical, suggesting a common logic for the assembly of the tetrapeptide intermediate during SFM-A, SFM-Mx1, and SAC-B biosynthesis. Based on the colinearity rule (25), wherein the NRPS module organization parallels the order of the amino acid residues in the resultant polypeptide, sequential incorporation of Ala, Gly, and two Tyr derivatives into the tetrapeptide in SAM-Mx1 biosynthesis was previously speculated to be directed by four successive modules (i.e., SafB-AL-PCP0, SafB-C1-A1-PCP1, SafA-C2-A2-PCP2, and SafA-C3-A3-PCP3-RE) (Fig. 6B) (33). Since SacA in SAC-B biosynthesis lacks the first module AL-PCP0, a bifunctional adenylation activation by SacA or direct incorporation of an Ala-Gly dipeptide into the tetrapeptide by SacA was hypothesized (Fig. 6C) (46). In both systems, the last two modules (SafA-C2-A2-PCP2 and SafA-C3-A3-PCP3-RE or SacB-C2-A2-PCP2 and SacC-C3-A3-PCP3-RE) were suggested to be responsible for activation and incorporation of each Tyr derivative, 3h5mOmTyr.

Our sequence analysis revealed that the N-terminal domain of SfmA in SFM-A biosynthesis, as well as that of SafB in SFM-Mx1 biosynthesis, lacks the expected conserved motifs of A domains and closely resembles the AL family, suggesting that SfmA-AL-PCP0 or SafB-AL-PCP0 might not be involved in amino acid incorporation like typical NRPS modules. Furthermore, the lack of this complete module in SAC-B NRPS system suggests that it may not be essential for tetrapeptide biosynthesis. Comparisons of the eight specificity-conferring amino acids of the A2 domain (DILQLGLI) and the A3 domain (DPWGLGLI) show they are significantly different, and thus, it is unlikely that the third module (SfmB-C2-A2-PCP2, SafA-C2-A2-PCP2, or SacB-C2-A2-PCP2) activates and incorporates the same 3h5mOmTyr residue as the fourth module does (SfmC-C3-A3-PCP3-RE, SafA-C3-A3-PCP3-RE, or SacC-C3-A3-PCP3-RE). In fact, question on the original assignments of substrate specificities for the SFM-Mx1 NRPS system was raised in bleomycin biosynthesis (10), in which reexamination of the SafAB sequence suggested that SafB-A1 serves as a candidate for Ala activation, while SafA-A2 recognizes and activates Gly. Consequently, these NRPS systems most likely catalyze the formation of a tetrapeptide intermediate using the last module in an iterative manner rather than following a typical colinearity rule, as shown in Fig. 6A. To confirm this hypothesis, we performed an amino acid-dependent ATP-PP_i exchange assay to determine the substrate specificities of SfmA, SfmB, and SfmC. As we anticipated, truncated SfmA (C1-A1-PCP1), SfmB (C2-A2-PCP2), and SfmC (C3-A3-PCP3-RE) showed exclusive activities with L-Ala, L-Gly, and L-3h5mOmTyr, respectively, strongly supporting the hypothesis that SfmC acts twice to incorporate two L-3h5mOmTyr residues into the tetrapeptide (Fig. 6A). This finding, as well as the recent emergence of many unusual NRPS systems (30, 39, 47), indicates a rich variety of biochemistry and architecture of NRPSs beyond those previously appreciated.

NRPS systems that are distinct from the current paradigm

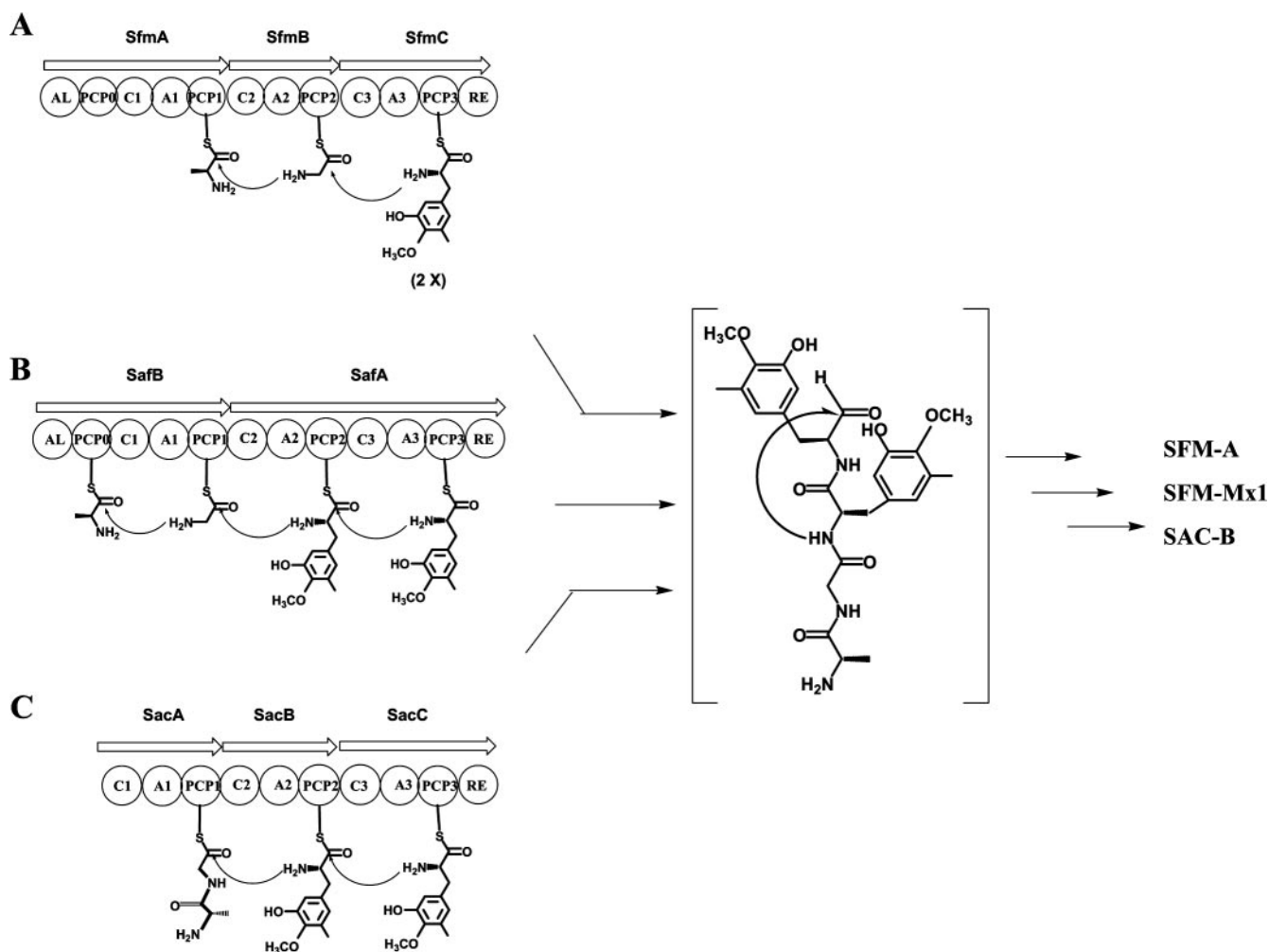


FIG. 6. Domain organizations and proposed enzymatic mechanisms of NRPS systems for SFM-A (A), SFM-Mx1 (B) (33), and SAC-B (C) (46) biosynthesis.

fall into one of two categories: they either contain at least one module with an atypical arrangement of the core domains, exemplified by the syringomycin biosynthesis (15), or use all their modules iteratively for product assembly with the TE domain channeling the multimeric intermediates, exemplified by enterobactin biosynthesis (14). While the SFM-A NRPS system shares the classical A-PCP-(C-A-PCP)_n domain organization found in linear NRPSs, the last module (SfmC), in contrast to the SfmA and SfmB modules, acts in an iterative fashion to produce a tetrapeptide intermediate. Furthermore, the TE domain that is typically found as the C-terminal domain of NRPS is substituted with a terminal RE domain. Since the C domain catalyzes transpeptidation between amino acyl thioesters without covalently binding the intermediate and since NRPSs are usually recognized to function as monomers (39), SfmC might channel the tripeptidyl or tetrapeptidyl intermediate in an unusual iterative manner that is distinct from any known NRPS system and mechanistically different from iterative events found in type I polyketide synthases (47). Knowledge of the structure of SfmC will be necessary to understand the selectivity and interaction of the domains involved and to

eventually engineer novel NRPS enzymes, like SfmC, for combinatorial biosynthesis.

As shown in Fig. 3B, a RE domain that resides at the C terminus of SfmC may reductively release the tetrapeptide intermediate from the PCP3 as a linear aldehyde. A series of intramolecular cyclizations lead to the formation of the B, C, and D rings, characteristic of the tetrahydroisoquinoline family. Subsequently, regiospecific methylation, oxidation, desamination, and substitution of a cyano group successively occur to produce SFM-A. Genetic comparison revealed that the *sfm* gene cluster contains all the structural genes for SAC-B biosynthesis (Fig. 2), supporting our hypothesis that SAC-B, which was originally isolated from a *Pseudomonas* strain, might serve as a key intermediate for SFM-A biosynthesis. It is not surprising that the *sfm* cluster harbors additional genes, since their functions are required for further modifications of the shared intermediate, such as multiple oxidations of the E ring. Heterologous expression of SfmO4 (a hydroxylase responsible for the introduction of a hydroxyl group at C-15 position) in the SAC-B producer resulted in a bisquinone derivative, aminated SFM-S, which was then converted into an aminated

SFM-A analog, SFM-Y3, by treatment of the fermentation culture with potassium cyanide (Fig. 4B). Thus, our results are consistent with parallel biosynthetic pathways for SFMA-A and SAC-B production.

In conclusion, the availability of the *sfm* biosynthetic gene cluster described here provides an excellent opportunity to access the unusual enzymatic mechanism of SFM-A biosynthesis. Sequence analysis and genetic comparison revealed a common strategy for tetrapeptidyl assembly among SFM-A and its analogs (SFM-Mx1 and SAC-B), and biochemical determination of substrate specificities for the A domains supported that the backbone formation is catalyzed by a multimodular NRPS system in a semi-iterative manner rather than following a previously proposed colinear rule. On the basis of the SAC-B biosynthetic machinery, we reconstructed an aminated SFM-A pathway by heterologously expressing a regiospecific hydroxylase SfmO4 in the SAC-B-producing *Pseudomonas* strain, thereby demonstrating a unified mechanism of biosynthesis among the tetrahydroisoquinoline compounds and the feasibility of engineering bacterial strains to generate new or otherwise scarce bisquinone alkaloid analogs.

ACKNOWLEDGMENTS

We thank Steven G. Van Lanen, School of Pharmacy, University of Wisconsin—Madison, for reading the manuscript and providing helpful comments; Andrew G. Myers, Department of Chemistry and Chemical Biology, Harvard University, for providing the authentic SFM-A standard; Victor De Lorenzo, Centro de Astrobiología (Instituto Nacional de Técnica Aeroespacial-CSIC), Spain, for providing the expression system in a *Pseudomonas* strain; and Linquan Bai, Shanghai Jiaotong University, China, for helpful discussions.

This work was supported in part by grants from the National Natural Science Foundation of China (20621062, 20402021, 30425003, and 30525001), the Ministry of Science and Technology of China (2006AA02Z185), the Chinese Academy of Science (KJCX2-YW-H08), and the Science and Technology Commission of Shanghai Municipality (04DZ14901 and 05QMX1466).

REFERENCES

- Ahlert, J., E. Shepard, N. Lomovskaya, E. Zazopoulos, A. Staffa, B. O. Bachmann, K. Huang, L. Fonstein, A. Czisny, R. E. Whitwam, C. M. Farnet, and J. S. Thorson. 2002. The calicheamicin gene cluster and its iterative type I enediyne PKS. *Science* **297**:1173–1176.
- Arai, T., K. Takahashi, K. Ishiguro, and K. Yazawa. 1980. Increased production of saframycin A and isolation of saframycin S. *J. Antibiot.* **33**:951–960.
- Arai, T., K. Takahashi, K. Ishiguro, and Y. Mikami. 1980. Some chemotherapeutic properties of two new antitumor antibiotics, saframycins A and C. *Gann* **71**:790–796.
- Arai, T., K. Yazawa, K. Takahashi, A. Maeda, and Y. Mikami. 1985. Directed biosynthesis of new saframycin derivatives with resting cells of *Streptomyces lavendulae*. *Antimicrob. Agents Chemother.* **28**:5–11.
- Baltz, R. H. 2006. Molecular engineering approaches to peptide, polyketide and other antibiotics. *Nat. Biotechnol.* **24**:1533–1540.
- Challis, G. L., J. Ravel, and C. A. Townsend. 2000. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **7**:211–224.
- Chiu, H. T., B. K. Hubbard, A. N. Shah, J. Eide, R. A. Fredenburg, C. T. Walsh, and C. Khosla. 2001. Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. *Proc. Natl. Acad. Sci. USA* **98**:8548–8553.
- Cuevas, C., M. Perez, M. J. Martin, J. L. Chicharro, C. Fernandez-Rivas, M. Flores, A. Francesch, P. Gallego, M. Zarzuelo, F. De La Calle, J. Garcia, C. Polanco, I. Rodriguez, and I. Manzanares. 2000. Synthesis of ecteinascidin ET-743 and phthalascidin Pt-650 from cyanosafrafrin B. *Org. Lett.* **2**:2545–2548.
- De Lorenzo, V., L. Eltis, B. Kessler, and K. N. Timmis. 1993. Analysis of *Pseudomonas* gene products using *lacI*^q/*P**trp-lac* plasmids and transposons that confer conditional phenotypes. *Gene* **123**:17–24.
- Du, L., C. Sanchez, M. Chen, D. J. Edwards, and B. Shen. 2000. The biosynthetic gene cluster for the antitumor drug bleomycin from *Streptomyces verticillus* ATCC 15003 supporting functional interactions between nonribosomal peptide synthetases and a polyketide synthase. *Chem. Biol.* **7**:623–642.
- Fischbach, M. A., and C. T. Walsh. 2006. Directing biosynthesis. *Science* **314**:603–605.
- Fortman, J. L., and D. H. Sherman. 2005. Utilizing the power of microbial genetics to bridge the gap between the promise and the application of marine natural products. *ChemBiochem* **6**:960–978.
- Galm, U., and B. Shen. 2006. Expression of biosynthetic gene clusters in heterologous hosts for natural product production and combinatorial biosynthesis. *Expert Opin. Drug Discov.* **1**:409–437.
- Gehring, A. M., I. Mori, and C. T. Walsh. 1998. Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry* **37**:2648–2659.
- Guenzi, E., G. Galli, I. Grgurina, D. C. Gross, and G. Grandi. 1998. Characterization of the syringomycin synthetase gene cluster. A link between prokaryotic and eukaryotic peptide synthetases. *J. Biol. Chem.* **273**:32857–32863.
- Haagen, Y., K. Gluck, K. Fay, B. Kammerer, B. Gust, and L. Heide. 2006. A gene cluster for prenylated naphthoquinone and prenylated phenazine biosynthesis in *Streptomyces cinnamonensis* DSM 1042. *ChemBiochem* **7**:2016–2027.
- Ishiguro, K., K. Takahashi, K. Yazawa, S. Sakiyama, and T. Arai. 1981. Binding of saframycin A, a heterocyclic quinone anti-tumor antibiotic to DNA as revealed by the use of the antibiotic labeled with [¹⁴C]tyrosine or [¹⁴C]cyanide. *J. Biol. Chem.* **256**:2162–2167.
- Ishiguro, K., S. Sakiyama, K. Takahashi, and T. Arai. 1978. Mode of action of saframycin A, a novel heterocyclic quinone antibiotic. Inhibition of RNA synthesis in vivo and in vitro. *Biochemistry* **17**:2545–2550.
- Kieser, T., M. Bibb, M. Butter, K. F. Chater, and D. A. Hopwood. 2000. Practical *Streptomyces* genetics. The John Innes Foundation, Norwich, United Kingdom.
- Konig, G. M., S. Kehraus, S. F. Seibert, A. Abdel-Lateff, and D. Muller. 2006. Natural products from marine organisms and their associated microbes. *ChemBiochem* **7**:229–238.
- Kopp, F., C. Mahlert, J. Grunewald, and M. A. Marahiel. 2006. Peptide macrocyclization: the reductase of the nostocyclopeptide synthetase triggers the self-assembly of a macrocyclic imine. *J. Am. Chem. Soc.* **128**:16478–16479.
- Liu, W., and B. Shen. 2000. Genes for production of the enediyne antitumor antibiotic C-1027 in *Streptomyces globisporus* are clustered with the *cagA* gene that encodes the C-1027 apoprotein. *Antimicrob. Agents Chemother.* **44**:382–392.
- Lown, J. W., A. V. Joshua, and J. S. Lee. 1982. Molecular mechanisms of binding and single-strand scission of deoxyribonucleic acid by the antitumor antibiotics saframycins A and C. *Biochemistry* **21**:419–428.
- Mao, Y., M. Varoglu, and D. H. Sherman. 1999. Molecular characterization and analysis of the biosynthetic gene cluster for the antitumor antibiotic mitomycin C from *Streptomyces lavendulae* NRRL 2564. *Chem. Biol.* **6**:251–263.
- Marahiel, M. A., T. Stachelhaus, and H. D. Mootz. 1997. Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* **97**:2651–2674.
- Mazodier, H., R. Petter, and C. Thompson. 1989. Intergeneric conjugation between *Escherichia* and *Streptomyces* species. *J. Bacteriol.* **171**:3583–3585.
- Menendez, N., M. Nur-e-Alam, A. F. Brana, J. Rohr, J. A. Salas, and C. Mendez. 2004. Biosynthesis of the antitumor chromomycin A3 in *Streptomyces griseus*: analysis of the gene cluster and rational design of novel chromomycin analogs. *Chem. Biol.* **11**:21–32.
- Mikami, Y., K. Takahashi, K. Yazawa, T. Arai, M. Namikoshi, S. Iwasaki, and S. Okuda. 1985. Biosynthetic studies on saframycin A, a quinone antitumor antibiotic produced by *Streptomyces lavendulae*. *J. Biol. Chem.* **260**:344–348.
- Mikami, Y., K. Yokoyama, H. Tabeta, K. Nakagaki, and T. Arai. 1981. Saframycin S, a new saframycin group antibiotic. *J. Pharmacobiol. Dyn.* **4**:282–286.
- Mootz, H. D., D. Schwarzer, and M. A. Marahiel. 2002. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBiochem* **3**:490–504.
- Plowright, A. T., S. E. Schaus, and A. G. Myers. 2002. Transcriptional response pathways in a yeast strain sensitive to saframycin A and a more potent analog: evidence for a common basis of activity. *Chem. Biol.* **9**:607–618.
- Pospiech, A., B. Cluzel, J. Bietenhader, and T. Schupp. 1995. A new *Mycococcus xanthus* gene cluster for the biosynthesis of the antibiotic saframycin Mx1 encoding a peptide synthetase. *Microbiology* **141**:1793–1803.
- Pospiech, A., J. Bietenhader, and T. Schupp. 1996. Two multifunctional peptide synthetases and an O-methyltransferase are involved in the biosynthesis of the DNA-binding antibiotic and antitumor agent saframycin Mx1 from *Mycococcus xanthus*. *Microbiology* **142**:741–746.
- Rao, K. E., and J. W. Lown. 1990. Mode of action of saframycin antitumor

- antibiotics: sequence selectivities in the covalent binding of saframycins A and S to deoxyribonucleic acid. *Chem. Res. Toxicol.* **3**:262–267.
35. Rao, K. E., and J. W. Lown. 1992. DNA sequence selectivities in the covalent bonding of antibiotic saframycins Mx1, Mx3, A, and S deduced from MPE. Fe(II) footprinting and exonuclease III stop assays. *Biochemistry* **31**:12076–12082.
36. Rinehart, K. L. 2000. Antitumor compounds from tunicates. *Med. Res. Rev.* **20**:1–27.
37. Sambrook, J., and D. W. Russell. 2001. *Molecular cloning: a laboratory manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
38. Schwartzmann, G., A. Brondani da Rocha, R. G. Berlinck, and J. Jimeno. 2001. Marine organisms as a source of new anticancer agents. *Lancet Oncol.* **2**:221–225.
39. Schwarzer, D., R. Finking, and M. A. Marahiel. 2003. Nonribosomal peptides: from genes to products. *Nat. Prod. Rep.* **20**:275–287.
40. Scott, J. D., and R. M. Williams. 2002. Chemistry and biology of the tetrahydroisoquinoline antitumor antibiotics. *Chem. Rev.* **102**:1669–1730.
41. Sianidis, G., S. E. Wohler, C. Pozidis, S. Karamanou, A. Luzhetskyy, A. Vente, and A. Economou. 2006. Cloning, purification and characterization of a functional anthracycline glycosyltransferase. *J. Biotechnol.* **125**:425–433.
42. Sieber, S. A., and M. A. Marahiel. 2005. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem. Rev.* **105**:715–738.
43. Simmons, T. L., E. Andrianansolo, K. McPhail, P. Flatt, and W. H. Gerwick. 2005. Marine natural products as anticancer drugs. *Mol. Cancer Ther.* **4**:333–342.
44. Stachelhaus, T., H. D. Mootz, and M. A. Marahiel. 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* **6**:493–505.
45. Thomas, M. G., Y. A. Chan, and S. G. Ozanick. 2003. Deciphering tuberactinomycin biosynthesis: isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster. *Antimicrob. Agents Chemother.* **47**:2823–2830.
46. Velasco, A., P. Acebo, A. Gomez, C. Schleissner, P. Rodriguez, T. Aparicio, S. Conde, R. Munoz, F. De La Calle, J. L. Garcia, and J. M. Sanchez-Puelles. 2005. Molecular characterization of the safracin biosynthetic pathway from *Pseudomonas fluorescens* A2-2: designing new cytotoxic compounds. *Mol. Microbiol.* **56**:144–154.
47. Wenzel, S. C., and R. Muller. 2005. Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from textbook biosynthetic logic. *Curr. Opin. Chem. Biol.* **9**:447–458.
48. Xing, C., J. R. Lacob, J. K. Barbay, and A. G. Myers. 2004. Identification of GAPDH as a protein target of the saframycin antiproliferative agents. *Proc. Natl. Acad. Sci. USA* **101**:5862–5866.