AEM
Journals.ASM.org

Unusual Biosynthesis and Structure of Locillomycins from *Bacillus subtilis* 916

Chuping Luo,[a,b] Xuehui Liu,[c] Xian Zhou,[d] Junyao Guo,[a] John Truong,[d] Xiaoyu Wang,[a] Huafei Zhou,[a] Xiangqian Li,[b] Zhiyi Chen[a]

Institute of Plant Protection, Jiangsu Academy of Agricultural Sciences, Nanjing, China[a]; School of Life Science and Chemical Engineering, Huaiyin Institute of Technology, Huaian, China[b]; Institute of Biophysics, Chinese Academy of Sciences, Beijing, China[c]; The National Institute of Complementary Medicine, University of Western Sydney, Sydney, New South Wales, Australia[d]

Three families of *Bacillus* cyclic lipopeptides—surfactins, iturins, and fengycins—have well-recognized potential uses in biotechnology and biopharmaceutical applications. This study outlines the isolation and characterization of locillomycins, a novel family of cyclic lipopeptides produced by *Bacillus subtilis* 916. Elucidation of the locillomycin structure revealed several molecular features not observed in other *Bacillus* lipopeptides, including a unique nonapeptide sequence and macrocyclization. Locillomycins are active against bacteria and viruses. Biochemical analysis and gene deletion studies have supported the assignment of a 38-kb gene cluster as the locillomycin biosynthetic gene cluster. Interestingly, this gene cluster encodes 4 proteins (LocA, LocB, LocC, and LocD) that form a hexamodular nonribosomal peptide synthetase to biosynthesize cyclic nonapeptides. Genome analysis and the chemical structures of the end products indicated that the biosynthetic pathway exhibits two distinct features: (i) a nonlinear hexamodular assembly line, with three modules in the middle utilized twice and the first and last two modules used only once and (ii) several domains that are skipped or optionally selected.

In the competition for nutrients, members of the *Bacillus* genus often produce a vast array of biologically active molecules that potentially inhibit the development of competing organisms. The Gram-positive bacterium *Bacillus subtilis* has an average of 4 to 5% of its genome devoted to antibiotic synthesis and is able to produce more than two dozen antibiotics with an amazing variety of structures (1). Many of these compounds, which have a peptide origin, are synthesized either ribosomally or nonribosomally. Among the nonribosomally generated amphipathic cyclic lipopeptides, surfactins, iturins, and fengycins have well-recognized potential applications in biotechnology and biopharmaceutical products due to their antagonistic activities and surfactant properties (2, 3). Furthermore, the mechanisms behind the observed biocontrol efficacy of different *Bacillus* strains have also been well described (4–6). Lipopeptides are able to induce systemic resistance in plants and to facilitate the multicellular behaviors of the producing strains, such as swarming motility, biofilm formation, and colony morphology (5–7).

Surfactins, iturins, and fengycins are synthesized by nonribosomal peptide synthetases (NRPSs) which exhibit a distinct modular architecture (2, 8–10). A module is typically composed of three core domains, with each domain responsible for a certain biochemical reaction (11). Specifically, the amino acid adenylation domain (A domain) controls the entry of substrates into the peptide structure by recognizing and activating a specific amino acid. The thiolation domain (T domain), also referred to as the peptidyl carrier protein (PCP), contains an invariant serine residue which is essential for the binding of a 4′-phosphopantetheine cofactor. The N-terminal condensation domain (C domain) is required for the coupling of two consecutively bound amino acids (12, 13). These three domains constitute a minimal elongation module, the basic repetitive unit of a multimodular NRPS. Furthermore, modules can be supplemented with domains that catalyze modifications of the activated amino acid, such as N-methylation and epimerization. In some cases, when the first module of an NRPS complex lacks a C domain, the last module contains a termination thioesterase domain (TE domain) to release the end product (14). The order and specificity of the modules within the protein template determine the sequence of the product (for type A, linear NRPSs) (8, 11). Genetic and biochemical analyses have revealed that the modular arrangement of most lipopeptide synthetases is colinear with the amino acid sequences of lipopeptides (1, 2). This assembly line arrangement of the conserved catalytic modules and domains provides the means to construct hybrid NRPSs for use in the synthesis of new lipopeptide compounds (15–18). The prospect of creating numerous bioactive lipopeptides by engineering existing lipopeptide synthetases has stimulated the search for new NRPSs responsible for lipopeptide synthesis (19–24).

To date, only two reported kinds of biosynthetic machinery within the NRPS assembly line do not conform to the rule of colinearity. These include the type B and type C NRPSs, which iteratively use all of their modules and certain domains, respectively, during the assembly of a product (25–28). While our understanding of the nonlinear NRPS biosynthetic mechanism is limited, it is clear that this mechanism has great potential to in-

Received 16 May 2015 **Accepted** 8 July 2015

Accepted manuscript posted online 10 July 2015

Citation Luo C, Liu X, Zhou X, Guo J, Truong J, Wang X, Zhou H, Li X, Chen Z. 2015. Unusual biosynthesis and structure of locillomycins from *Bacillus subtilis* 916. Appl Environ Microbiol 81:6601–6609. doi:10.1128/AEM.01639-15.

Editor: M. A. Elliot

Address correspondence to Xiangqian Li, lixq2002@126.com, or Zhiyi Chen, chzy84390393@163.com.

C. Luo and X. Liu contributed equally to this work.

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.01639-15.

doi:10.1128/AEM.01639-15

TABLE 1 Primers used in this study

Primer	Size (nt)	Sequence (5′-3′)	Note
LocDF	30	TTTGCATGCATGAACTATGATTTATCACAT	Used for Δ*locD* mutant construction
LocDR	29	TTTGAGCTCAGTAAAAATGAGAGGCAATT	Used for Δ*locD* mutant construction
LocCF	30	TTTAAGCTTAATTCAGCTCTTTATGAAGAA	Used for Δ*locC* mutant construction
LocCR	30	TTTGAGCTCATGAAGTTTACTAATGAATAC	Used for Δ*locC* mutant construction
Orf1UF	30	TTTAAGCTTGAATAAATAATTCACGGTAAA	Used for Δ*orf1* mutant construction
Orf1UR	30	TTTTCTAGAATTCAGCTGCTTTATCGTAAG	Used for Δ*orf1* mutant construction
Orf1DF	30	TTTTCTAGACTGTTGCTTTGTCGCATAATG	Used for Δ*orf1* mutant construction
Orf1DR	30	TTTGCATGCAAGAGTGAGTTATCCAGTTGA	Used for Δ*orf1* mutant construction
Orf3F	30	TTTAAGCTTCATTGAACTGAATAAAATGTA	Used for Δ*orf3* mutant construction
Orf3R	30	TTTGAGCTCTTAGTCTCAATCTCAATGTTT	Used for Δ*orf3* mutant construction
Orf7F	30	TTTGCATGCATACGATTAATAAAAGATATG	Used for Δ*orf1* mutant construction
Orf7R	30	TTTGAATTCTTAAATTGTTATATCATCTTT	Used for Δ*orf1* mutant construction
LocAA1F	32	TTTCCATGGTATCTGAAATAGAAATGATTACG	Used for A1 domain expression in *E. coli*
LocAA1R	30	TTTCTCGAGTTCTTTCTGAATAGCTGTTTG	Used for A1 domain expression in *E. coli*
LocBA2F	33	TTTCCATGGTAAAAGATGTAGAAATTATTACAG	Used for A2 domain expression in *E. coli*
LocBA2R	33	TTTCTCGAGCATTTGTTCTTTTTTATTATTTGG	Used for A2 domain expression in *E. coli*
LocBA3F	29	TTTCCATGGTTTCGGAGATTGATATCACG	Used for A3 domain expression in *E. coli*
LocBA3R	29	TTTCTCGAGGGTTTCCTGAACATCATTCT	Used for A3 domain expression in *E. coli*
LocBA4F	31	TTTGGATCCATCTCTTCTATAGATATCATGA	Used for A4 domain expression in *E. coli*
LocBA4R	32	TTTCTCGAGTTATGTTTCCTGCAAAACTGTTT	Used for A4 domain expression in *E. coli*
LocCA5F	28	TTTCCATGGGAGATGTAGGTTTGCTGAC	Used for A5 domain expression in *E. coli*
LocCA5R	33	TTTCTCGAGAACTAATTCTTTTTCAATGTTATT	Used for A5 domain expression in *E. coli*
LocCA6F	32	AAACCATGGTATATCAAATTAACATGATGACT	Used for A6 domain expression in *E. coli*
LocCA6R	30	AAACTCGAGTTTTGTCTCTGTTTCGTTTCT	Used for A6 domain expression in *E. coli*
PEDF	24	CGCAAAGACTGAACCCACTAATTT	Used for quantitative analysis of PEDV
PEDR	24	TTGCCTCTGTTGTTACTTGGAGAT	Used for quantitative analysis of PEDV

troduce structural diversity to secondary metabolites through combinatorial biosynthesis (25, 29, 30). It is necessary to propose nonlinear NRPS biosynthetic models that can further unravel the details of this biosynthetic mechanism. The Gram-positive bacterium *B. subtilis* serves as a model organism and is intensively used in the heterologous expression of commercial metabolites (30–32). However, apart from the present study, no nonlinear NRPS assembly line has been observed in this well-characterized species.

The commercial strain *B. subtilis* 916 was isolated from paddy soils in Jurong County, Jiangsu, China, and has been reported to be effective in the biocontrol of plant diseases (33). While NRPSs for lipopeptide production in *B. subtilis* 916 were reported recently, and the locillomycin also described briefly (34), in the present study we describe in detail the novel structure and unique biosynthesis of the new locillomycin family of nonribosomal lipopeptides and its biological function. In this study, our results strongly suggest that locillomycins have molecular features not observed in the three families of surfactins, iturins, and fengycins, including a unique nonapeptide sequence and macrocyclization. Locillomycins are active against bacteria and viruses, with low hemolytic activity, and can be used in natural health products for therapeutic applications. While the end products were straightforward, the proposed biosynthetic pathway of locillomycins contained several atypical aspects that were not predictable by bioin-

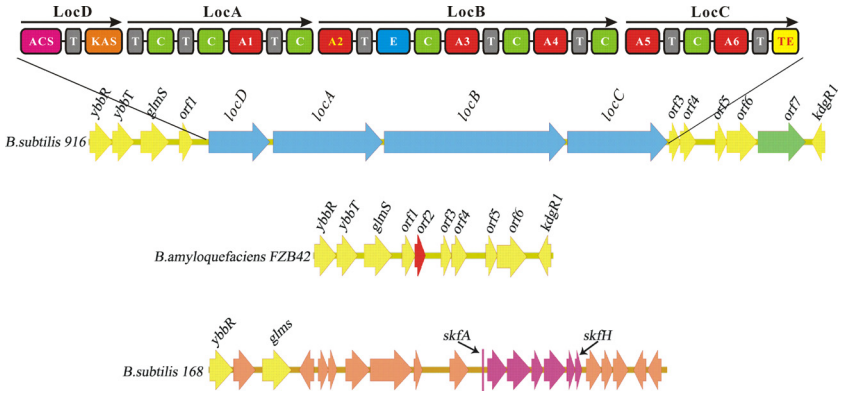

FIG 1 Organization of the locillomycin gene cluster and surrounding genes on the *B. subtilis* 916 chromosome. The same loci of *B. subtilis* 168 and *B. amyloquefaciens* FZB42 are also shown for comparison. Genes with high degrees of homology between the three strains (>95% identity) are shown in yellow. The biosynthetic clusters for locillomycins and sporulation killing factor are shown in blue and purple, respectively. Other, unrelated genes are also shown in green, red, and orange.

formatic analysis. Our study thus adds to the knowledge regarding nonlinear NRPS biosynthesis mechanisms used by *B. subtilis* and further broadens the potential of strain 916 as a model system for nonlinear NRPS studies.

MATERIALS AND METHODS

Cultivation, isolation, and purification. Luria-Bertani (LB) broth was inoculated with *B. subtilis* 916 to a concentration of 5.5×10^5 cells/ml. One hundred cultures of 330 ml each in 2-liter Erlenmeyer flasks were incubated with agitation at 180 rpm at 28°C for 3 days. After removing the biomass by centrifugation, the broth was titrated to pH 2.8 with concentrated hydrochloric acid, and the resulting gray precipitate was extracted with 150 ml methanol (MeOH). The MeOH extract was added to 300 ml H_2O and titrated to pH 7 with 5 M NaOH. The extracted impurities were added to an Agilent amino solid-phase extraction column and sequentially washed with 50% (vol/vol) MeOH-H_2O, 100% MeOH, 1% (vol/vol) formic acid-MeOH, and 2% (vol/vol) formic acid-MeOH. The 2% (vol/vol) formic acid-MeOH eluate was concentrated by nitrogen drying to 10 mg/ml, loaded on an Agilent C_{18} solid-phase extraction column, and sequentially washed with 40, 42, 46, and 48% (vol/vol) MeOH-H_2O. The 44, 46, and 48% (vol/vol) MeOH-H_2O eluates were concentrated to 20 mg/ml by nitrogen drying. These fractions were further processed by high-pressure liquid chromatography (HPLC), using repetitive 50-μl injections, a reversed-phase column (RP-18; 5 μm by 4 mm by 250 mm; Merck), and a 0.5-ml/min flow rate, and were monitored at 230 nm. Fraction I yielded derivative A, which was collected at an average peak retention time of 9 min; fraction II yielded derivative B, with a retention time of 13 min; and fraction III yielded derivative C, with a retention time of 18 min. The fractions were eluted with 50% (vol/vol) CH_3CN, 0.5% (vol/vol) trifluoroacetic acid in H_2O. All collections were concentrated by nitrogen and vacuum drying and resulted in the purified locillomycin A, B, and C derivatives.

NMR structure determination. All nuclear magnetic resonance (NMR) spectra were acquired at 25°C on a Bruker Avance III 600-MHz, Agilent DD2 500-MHz (for ^{13}C-detected spectra), or Agilent DD2 600-MHz spectrometer equipped with a cold probe. Locillomycin A (~8 mg) was dissolved in 500 μl aqueous solvent (20 mM phosphate buffer, pH 6.5, D_2O-H_2O [9:1 {vol/vol}]), while ~8-mg samples of locillomycins B and C were dissolved in 500 μl CD_3OH to make the NMR samples. For each sample, a series of one-dimensional (1D) and 2D spectra, including spectra for ^1H, ^{13}C, distortionless enhancement by polarization transfer (DEPT), ^1H-^1H correlation spectroscopy (COSY), ^1H-^1H total correlation spectroscopy (TOCSY), ^1H-^1H rotating-frame nuclear Overhauser effect spectroscopy (ROESY), ^1H-^{13}C heteronuclear single quantum coherence (HSQC), and ^1H-^{13}C heteronuclear multiple-bond correction (HMBC), were acquired for structure elucidation. The spin-lock time for each TOCSY experiment was 80 ms, and the mixing time for each ROESY experiment was 200 ms. All NMR data were processed and analyzed with MestReNova. The chemical shifts in the aqueous solvent were referenced to the single-deuterium hydrogen oxide (HOD) peak at 4.77 ppm, while those in CD_3OH were referenced to the peak of the methyl group, at 3.30 ppm. Chemical shifts of ^{13}C were referenced indirectly (35) or to the peak of CD_3OH, at 49 ppm. Sequential assignment for the peptide part was done using the protocol developed by Wüthrich (36).

Gene disruption of *B. subtilis* 916 and mutant analysis. *In vivo* generation of targeted mutations in *B. subtilis* 916 was achieved by a modified protocol originally developed for *B. subtilis* 168 (9). The knockout plasmids were constructed from vector pUC19 by inserting homologous fragments from *B. subtilis* 916 genomic DNA and a neomycin resistance cassette from pBEST501. An example for disruption of LocD (Δ*locD*) is detailed below. A 2.5-kb fragment was amplified by a PCR using the primers LocDF and LocDR and cloned into vector pUC19, generating pLocD1. The latter was digested with XbaI, which cuts in the middle of the PCR fragment. Simultaneously, pBEST501 was cut with the same enzyme to obtain the neomycin resistance cassette (~1.3 kb), which was then ligated

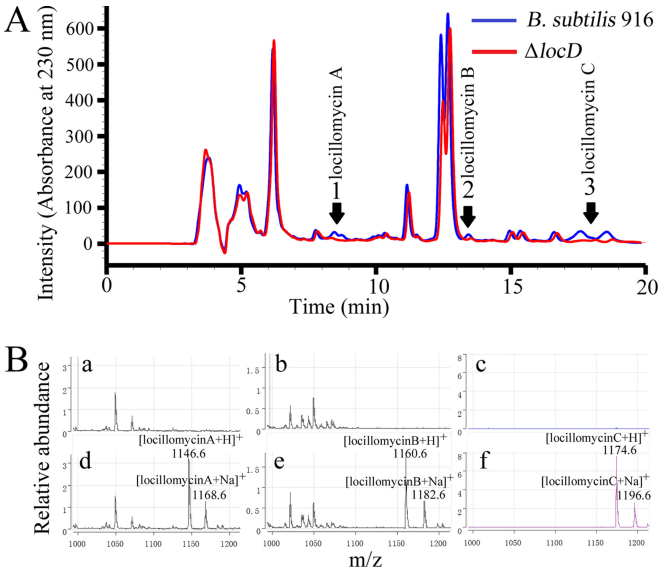

FIG 2 HPLC-MS chromatograms for locillomycins produced by wild-type *B. subtilis* 916 and its Δ*locD* mutant. (A) Locillomycins A to C were detected in *B. subtilis* 916 broth cultures but were below the detectable level in Δ*locD* cultures. (B) MS spectra for extracts of the Δ*locD* mutant (a to c) and wild-type *B. subtilis* 916 (d to f). Molecular weights of locillomycins A, B, and C were determined for *B. subtilis* 916 broth cultures, but these compounds were not observed in the Δ*locD* mutant culture broth.

to pLocD1, resulting in pLocD2. This was subsequently transformed into the naturally competent strain *B. subtilis* 916, in which it was introduced into the genome via double-crossover homologous recombination. The disruption of the *locD* gene in neomycin-resistant colonies was confirmed by PCR with appropriate primers and by Southern hybridization. The locillomycins were extracted from the mutants and further analyzed by HPLC-mass spectrometry (HPLC-MS) as described above.

Heterologous expression and purification of internal adenylation domains. The pET28a expression system was used to clone PCR products in which NcoI and XhoI sites were introduced by use of 5′-modified primers (Table 1). PCR amplification from the *B. subtilis* 916 chromosome was carried out in a 50-μl reaction volume containing 2.5 U Ex*Taq* DNA polymerase, 5 μl 10× Mg-free reaction buffer, a 200 μM concentration of each deoxynucleoside triphosphate (dNTP), 2.5 mM $MgCl_2$, and 1 μM (each) primers. Amplification conditions were as follows: 30 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 3 min, followed by a final extension at 72°C for 10 min. PCR products were digested with NcoI and XhoI and cloned into the digested pET28a vector. PCR products cloned into the XhoI site of pET28a resulted in appendage of a poly-His tag at the carboxyl ends of recombinant proteins. Transformants of *Escherichia coli* BL21(DE3) were selected by growth at 28°C in LB medium containing 100 μg/ml ampicillin. Cells were induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) at an optical density at 600 nm (OD_{600}) of 0.6 and allowed to grow for an additional 4 h before being harvested. Purification of the His_6-tagged proteins was carried out by Ni^{2+}-affinity chromatography. The proteins were separated by 10% (wt/vol) SDS-PAGE, stained with Coomassie blue, and analyzed by density scanning using an image analysis system (Bio-Rad).

Pyrophosphate exchange assay. Amino acid-dependent ATP-sodium pyrophosphate assays were performed by using the spectrophotometric assay furnished by an EnzChek pyrophosphate assay kit (Molecular Probes). Each 200-μl reaction mixture contained 75 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 5 mM dithiothreitol (DTT), 5 mM ATP, 400 mM MesG, 0.2 U purine nucleoside phosphorylase, 0.2 U inorganic pyrophosphatase, and 2 mM amino acids. A reaction mixture without amino acids

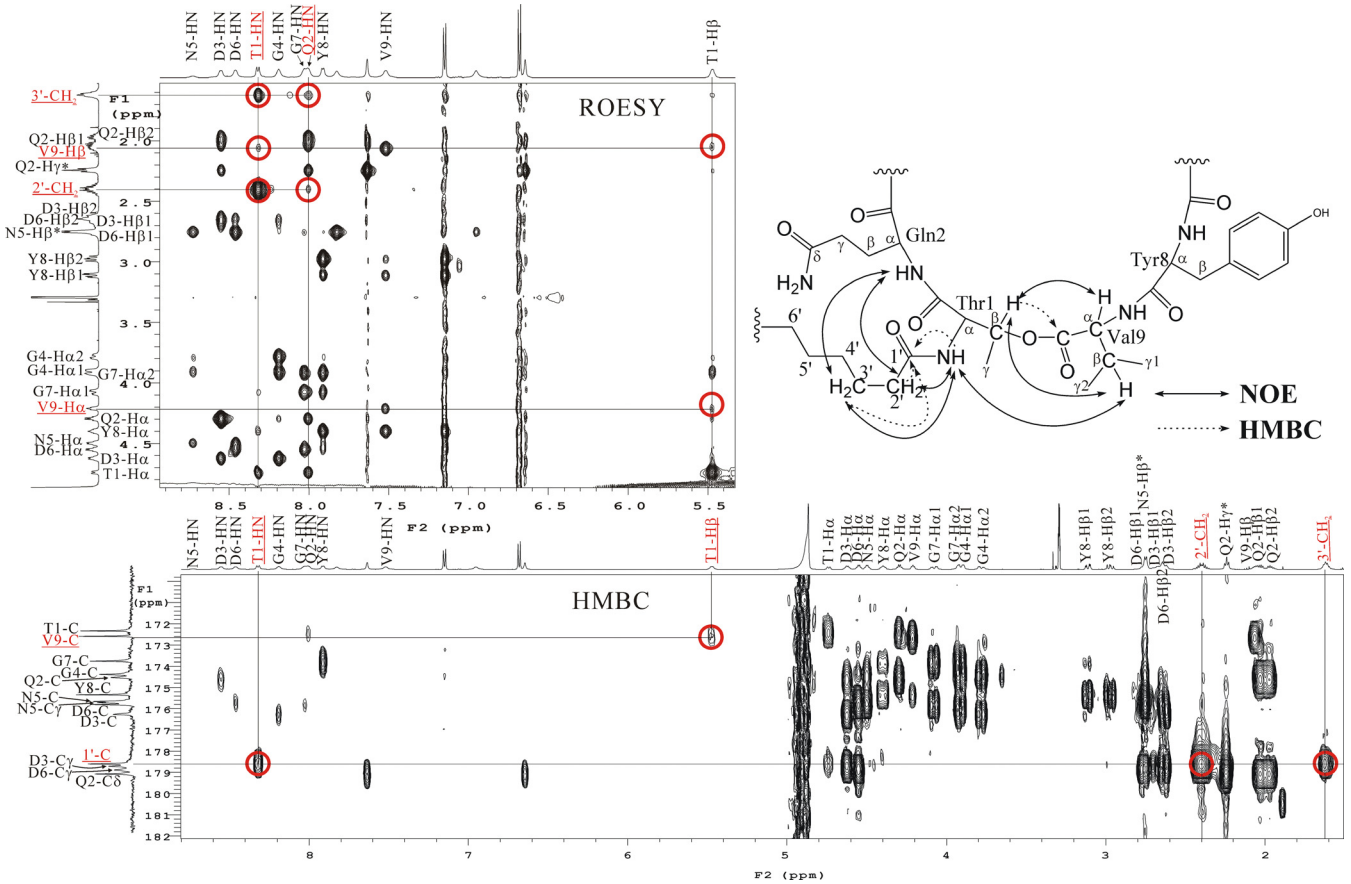

FIG 3 Important nuclear Overhauser effect (NOE) and HMBC interactions for determination of the attachment site for the acyl group and the cyclic structure of the peptide moiety are illustrated in the structure (top right) and the spectra (top left and bottom). The ROESY (top left) and HMBC (bottom) spectra are also labeled with assigned signals along the sides. The assigned nucleic names of the key residues are underlined and displayed in red, and the red circles show relevant NOE and HMBC interactions.

was used as the control. The reaction was initiated by the addition of enzyme, and the mixture was incubated at 30°C for 30 min. The absorbance at 360 nm was measured in a Perkin-Elmer Lambda 6-vis spectrophotometer. All reactions were performed in triplicate.

Nucleotide sequence accession numbers. The complete genome sequence of *B. subtilis* 916 and the sequence of the gene cluster for biosynthesis of locillomycins described in this work have been submitted to GenBank under accession numbers CP009611 and KF866134, respectively.

RESULTS AND DISCUSSION

Identification and *in vivo* gene disruption of *loc* gene cluster. Identification and bioinformatic analyses of the *loc* gene cluster were employed as described in our previous work (34). The draft genome and complete genome sequences of *B. subtilis* 916 were recently analyzed (GenBank accession no. AFSU00000000.1 and CP009611), and four NRPS gene clusters were identified (see Fig. S1 in the supplemental material) (37). In addition to the three conventional gene clusters, *srf* (for surfactins), *bmy* (for bacillomycin Ls), and *fen* (for fengycins), *B. subtilis* 916 also contains a fourth NRPS gene cluster, *loc* (GenBank accession no. KF866134) (34). Further analysis of the *loc* open reading frames led to the conclusion that the *loc* gene cluster was a potential candidate for the biosynthesis of an unknown family of lipopeptides called locillomycins. Interestingly, this particular locus in *B. subtilis* 916 shares a high degree of homology (>95% identity) with the same locus found in *B. subtilis* 168 and *Bacillus amyloquefaciens* FZB42 (Fig. 1).

Four NRPS genes, designated *locD*, *-A*, *-B*, and *-C*, are contiguous in the *loc* gene cluster and correspond to six peptide extension modules (Fig. 1). The first gene, *locD*, encodes a 145.8-kDa protein containing the polyketide synthase module and including domains with homology to the proteins involved in the synthesis of fatty acids and polyketides: a fatty acid acyl-coenzyme A synthetase (ACS) domain, an acyl carrier protein/thiolation domain (T domain), and a β-ketoacyl synthetase (KAS) domain. The second gene, *locA*, encodes a 267.4-kDa protein with only one extension module, containing the core elongation domains: the C, A, and T domains. Two C domains exist independently upstream of the A domain. The third gene, *locB*, encodes a 443.1-kDa protein containing three extension modules. The first module contains a predicted epimerization (E) domain, suggesting that the corresponding amino acid appears in the D-configuration. The fourth gene, *locC*, encodes a 243.9-kDa protein containing two extension modules and a termination module, which contains a C-terminal TE domain. A similarity analysis showed that the nucleotide sequence of the *loc* gene cluster has low similarity (<50%) to other lipopeptide NRPS gene clusters (see Table S1 in the supplemental mate-

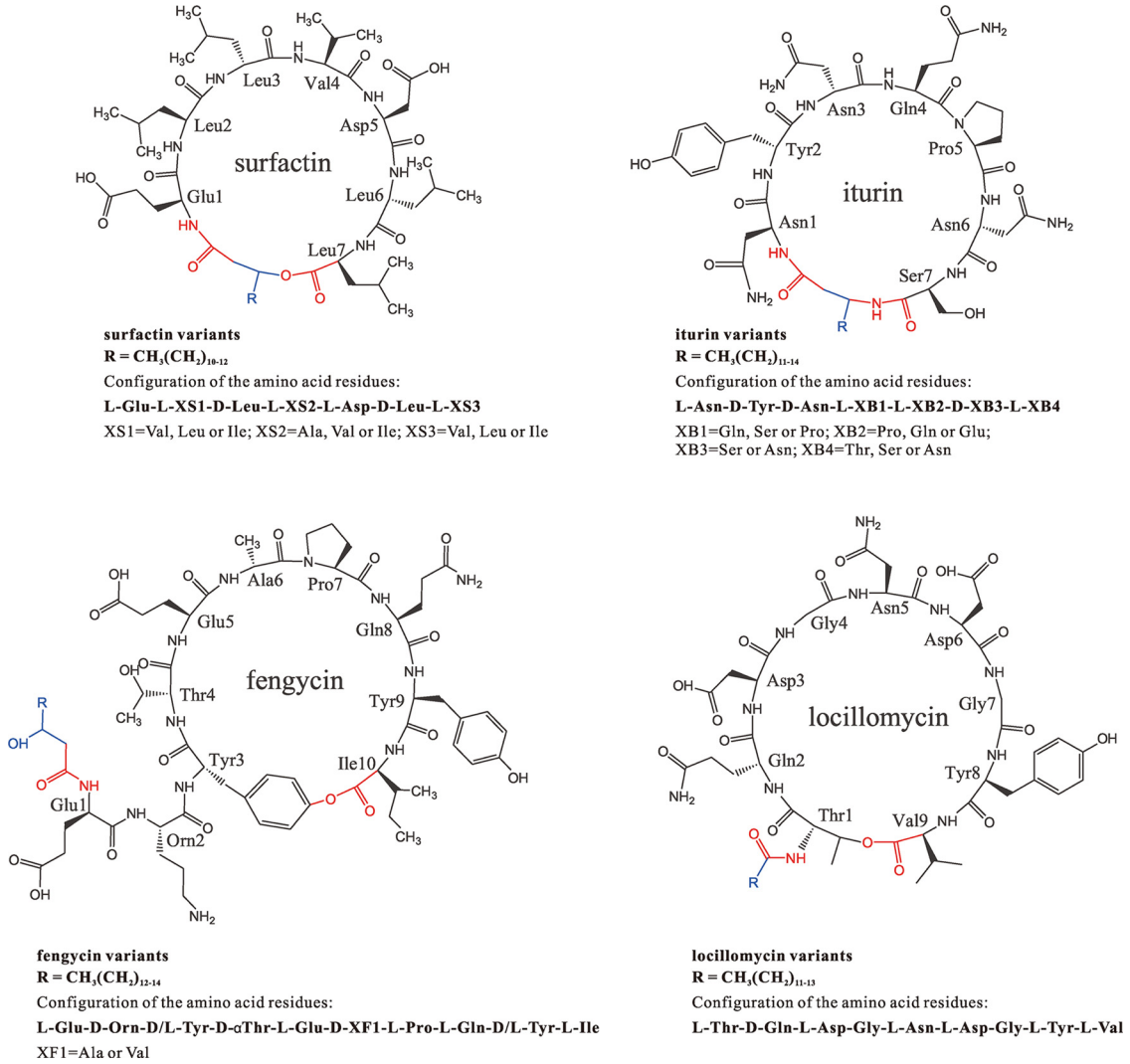

FIG 4 Structural comparison of locillomycins, surfactins, iturins, and fengycins. Some of the key bonds are colored red and blue for clarity. The residue numbers are labeled next to the α-carbons. Statistical sequences and the configuration for each residue are listed below the formulas.

rial). In addition, the combination and organization of the enzymatic components of *loc* are uniquely structured (Fig. 1).

It was previously shown that the amino acid specificities of the individual modules can be predicted by comparing the active site residues of known NRPS A domains (see Table S2 in the supplemental material) (38). According to the colinearity rule, the locillomycins biosynthesized by the *loc* gene cluster may be composed of a long fatty acid chain of 13 to 18 carbons linked to a hexapeptide moiety, with molecular masses ranging from 800 to 1,000 Da.

A set of gene disruption experiments were carried out as previously described to verify the identified NRPS *loc* gene cluster and to test the necessity for various genes in the biosynthesis of locillomycins. The knockout targets included *locD* (encoding a polyketide synthase module), *locC* (including the coding sequence for a termination module), *orf1* (encoding an upstream ABC transporter protein), *orf3* (encoding a putative regulatory protein containing the sensor histidine kinase), and *orf7* (encoding a putative multidrug transporter protein). The primers used in this study are listed in Table 1. The production of locillomycins was completely abolished in the Δ*locD* and Δ*locC* strains, which demonstrated that these two genes were essential for the biosynthesis of locillomycins. The disruption of *orf3* also significantly reduced the yield of locillomycins compared to that of the wild type (reduced to <5%), with only trace amounts of locillomycins as indicated by mass ion extraction. This result suggested that *orf3* plays an important role in the regulation of locillomycin biosynthesis. Contrary to our expectations, the disruption of *orf1* and *orf7* had no apparent impact on the production of locillomycins (see Fig. S3 in the supplemental material). These results suggest that a variety of transporter proteins may be involved in the transportation of locillomycins. In general, the *in vivo* gene disruption experiments suggested that the identified *loc* gene cluster isolated from *B. subtilis* 916 is responsible for the biosynthesis of locillomycins.

Given the obvious differences in molecular masses of locillomycins detected by HPLC-MS (1,145.6, 1,159.6, and 1,173.6 Da) (Fig. 2) and the molecular masses predicted by the bioinformatic analysis (800 to 1,000 Da), the results strongly suggested that the biosynthesis of locillomycins in *B. subtilis* 916 does not obey the

colinearity rule. To further elucidate the mechanism of biosynthesis of locillomycins, it was necessary to characterize the structures of locillomycins and to detail the six internal adenylation domains encoded by the *loc* gene cluster.

Isolation and structural characterization of locillomycins. The procedures for cultivation, isolation, and purification of locillomycins are described in detail in Materials and Methods. Briefly, the purification method consisted of organic solvent extraction, amino solid-phase chromatography, C_{18} solid-phase chromatography, and repetitive HPLC purification (see Fig. S4 in the supplemental material). Detailed NMR and MS approaches were pursued simultaneously with the gene disruption analysis work to determine the chemical structures of the locillomycins. Locillomycin C was used as a representative example to explain the structural characterization of all locillomycins (Fig. 3). The amino acid residue types and connectivity were obtained by analyzing the ROESY and TOCSY spectra (36). We concluded that there are nine residues in locillomycin C and that these residues are connected in the sequence Thr1-Gln2-Asp3-Gly4-Asn5-Asp6-Gly7-Tyr8-Val9. These observations were consistent with the data from the tandem MS analysis (see Table S5 and Fig. S8). Several Thr1-Val9 interactions observed in the ROESY spectrum indicate that these two residues are close in space, implying the possibility of a cyclic structure connected end to end. The existence of a long alkyl group is evident from the large peaks around 1.2 to 1.4 ppm in the ^{1}H spectrum and corresponding peaks in the ^{13}C spectrum. Protonated carbons were assigned after the assignment of most protons, with the aid of the ^{13}C-edited (CH_2 negative, CH_3, and CH positive) HSQC spectrum. All known fragments of the compound accounted for 13 carbonyl signals, while there were 14 carbonyl signals in the ^{13}C spectrum. This indicated that there must be a carbonyl group in the long alkyl chain of this compound. By analyzing both the HMBC and ROESY spectra, the attachment site between the carbonyl alkyl chain (actually a long-chain acyl group) and the peptide moiety was determined. Moreover, both spectra confirmed that the peptide chain was indeed cyclical in nature, through the end-to-end connection of Thr1 and Val9 (Fig. 3). Finally, the number of CH_2 groups was determined by integrating the multi-CH_2 area, and the result is in accordance with the results of mass spectrum analysis. No fragment from the alkyl group was identified by MS, and neither the DEPT nor proton spectrum indicated that there was branching in this alkyl group. Stereochemistry analysis of the amino acids was performed by HPLC after complete degradation of the compounds (see Fig. S9). Only one of the nine amino acids was found to be in the D-configuration (D-Gln). The resulting chemical structure is shown in Fig. 4. As for locillomycins A and B, analysis of NMR and MS spectra showed that they were very similar to locillomycin C, and there was a difference in molecular mass of only 14 Da between the different locillomycins (C > B > A) (see Tables S3 to S5 and Fig. S6 to S8). This difference is derived from the different lengths of the long-chain acyl group shown in Fig. S10, and the method of structural determination is not detailed here.

Unlike the surfactin and iturin lipopeptide families, in which the peptide residues are interlinked with a β-hydroxy or β-amino fatty acid to form a cyclic lactone or lactam, the locillomycins form an internal lactone ring within the peptidic moiety. This construction is more similar to that of the fengycin family, which also forms a lactone without the participation of a β-hydroxyl fatty acid. Importantly, however, the number of residues in locillomycins is 9 rather than the 10 observed in the fengycin family (8 of 10 residues form the lactone ring) (Fig. 4). These nine-membered cyclic lipopeptides are also different from the iturin and surfactin families, which form seven-membered rings. Furthermore, only one D-amino acid (D-Gln2) is incorporated into the members of the locillomycin family, while all the members of the other three families possess more than two D-amino acid residues. The lengths of the fatty acid chains are similar to those of other *Bacillus* lipopeptides.

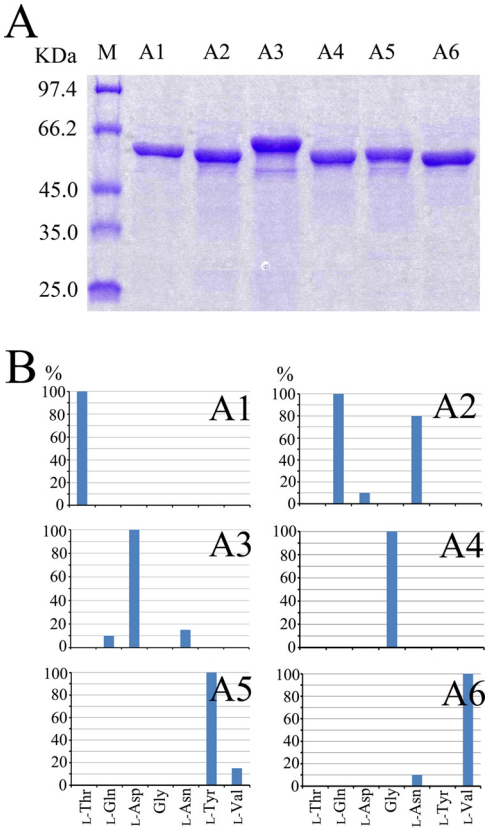

FIG 5 Expression, purification, and relative substrate specificities of internal adenylation domains (A domains). (A) Expression and purification of the A domains were analyzed in 10% SDS-polyacrylamide slab gels. (B) A domains were investigated in terms of activity in the ATP-PP_i exchange reaction, using the amino acids of locillomycin and a control without amino acids. The highest activities were set at 100%. The background was below 5%. The specificities of the different domains coincide with the primary structures of the locillomycins.

Biochemical investigation of internal adenylation domains. To further elucidate the biosynthetic process for locillomycins, DNA fragments encoding the adenylation domains of modules LocA1 (A1), LocB1 (A2), LocB2 (A3), LocB3 (A4), LocC1 (A5), and LocC2 (A6) (see Fig. 6) were amplified from chromosomal *B. subtilis* 916 DNA and cloned into an IPTG-inducible expression vector as described in Materials and Methods. The constructs were confirmed by sequencing of the fusion sites. The enzyme fragments were overexpressed in *E. coli* as His_6-tagged proteins and were purified by Ni^{2+}-affinity chromatography. All proteins were found within the soluble fraction after French press lysis of *E. coli* cells, as confirmed in Coomassie blue-stained SDS-polyacrylamide gels (Fig. 5A), with calculated molecular masses of 60.0 to

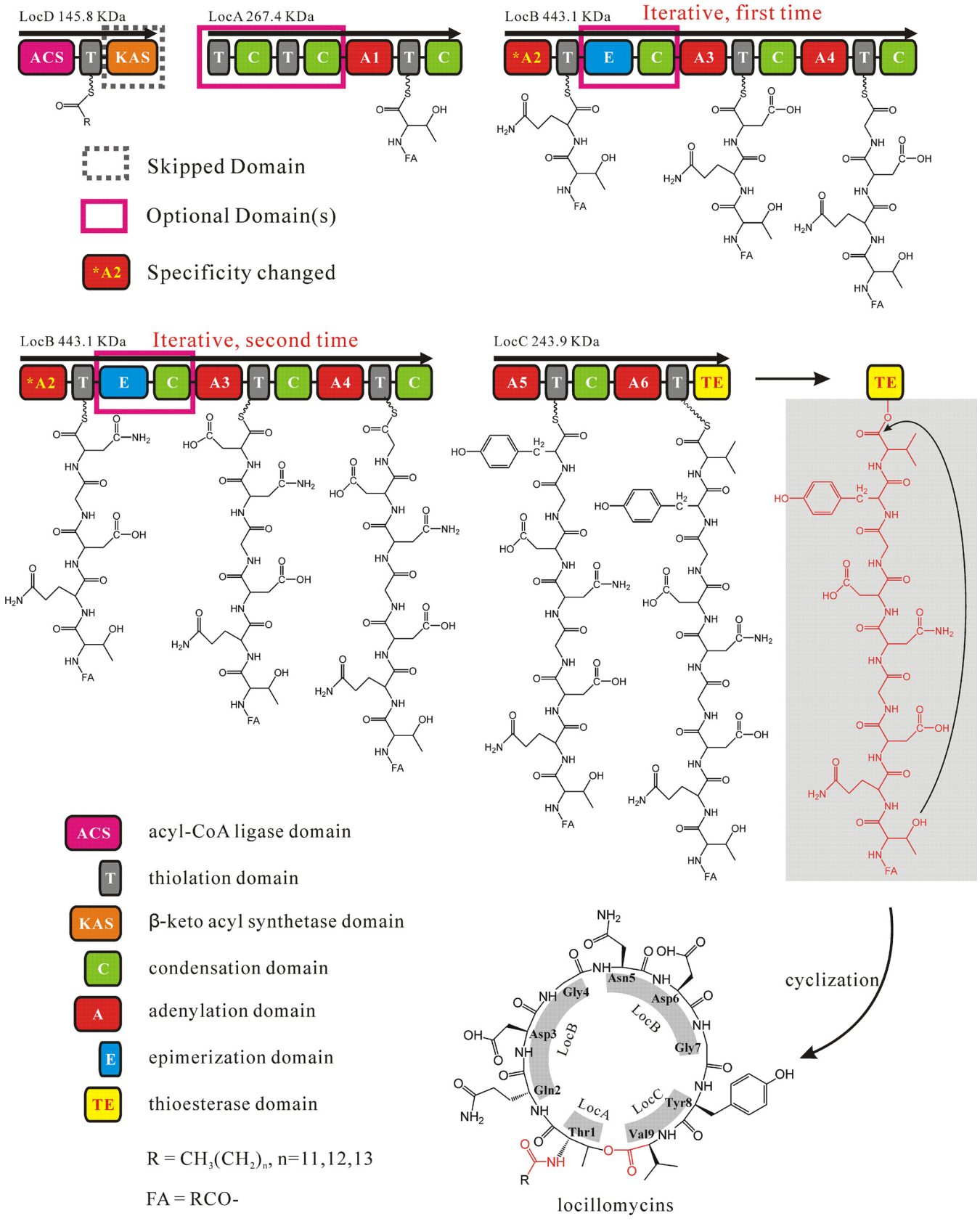

FIG 6 Proposed biosynthetic pathway for locillomycins. Nonribosomal peptide synthetase genes (*locA*, *locB*, and *locC*) are represented by arrows, and domains encoded by the respective genes are shown underneath; the domains encoded by *locB* are proposed to act iteratively. The fatty acid chain (FA) is coupled to the peptide by formation of an amide bond with the amino group of the threonine residue, and a nonapeptide is cyclized by formation of a macrolactone between the threonine hydroxyl and the valine carboxylate. The KAS domain is skipped during the synthesis and is labeled with a dashed box, while domains used optionally are labeled with purple boxes. *locB* appears twice because it is used iteratively and the A2 domain exhibits different specificities in the two rounds. The E domain following the A2 domain also loses its function in the second round.

65.0 kDa. The purified adenylation domains were biochemically investigated with respect to their activity and specificity in the ATP-PP_i exchange reaction. In order to determine amino acid specificity, each protein was incubated with all constituent amino acids of locillomycins (the control was incubated without amino acids). The six adenylation domains were found to activate the amino acids corresponding to the constituent amino acids of locillomycins. Specifically, A1 exclusively activated L-Thr; A2 activated L-Gln (100%), L-Asn (80%), and L-Asp (10%); A3 activated L-Asp (100%), L-Asn (15%), and L-Gln (10%); A4 exclusively activated Gly; A5 activated L-Tyr (100%) and L-Val (15%); and A6 activated L-Val (100%) and L-Asn (10%) (the highest activity was set at 100%) (Fig. 5B). A1, the only adenylation domain of LocA, exclusively activated L-Thr and is likely to be involved in position 1 amino acid biosynthesis. A2, A3, and A4, the three adenylation domains of LocB, are likely to act iteratively to incorporate the 2nd to 7th amino acids. In particular, A2 was able to efficiently activate both L-Gln and L-Asn. A5 and A6, the last two adenylation domains of LocC, optimally activated L-Tyr and L-Val, respectively, which is in agreement with the last two amino acids of locillomycins.

Proposed locillomycin biosynthesis pathway. The gene cluster for locillomycin biosynthesis identified in the present study sets the stage for further delineating the intricate chemical assembly of this family of lipopeptide antibiotics. Based on the chemical structures of locillomycins, the analysis of *in vivo* gene disruptions, internal adenylation domain experiments, the amino acid sequence homology of *loc*-encoded proteins to enzymes, and previous knowledge of lipopeptide biosynthetic pathways, we propose a model for the locillomycin biosynthetic pathway (Fig. 6). In brief, it is a pathway where a cyclic liponanopeptide is assembled by multifunctional enzymes of a hexamodular NRPS. In the initial step, the ACS domain couples coenzyme A to a long-chain fatty acid, presumably a tridecanoic to pentadecanoic acid, in an ATP-dependent reaction. The activated fatty acid is then transferred to the 4-phosphopantetheine cofactor of the T domain. In the second step, the fatty acid is coupled to the activated threonine directly, catalyzed by one of the condensation domains preceding the A1 domain of the peptide synthetases, with the KAS domain, two T domains, and one C domain skipped (Fig. 6). In subsequent condensation reactions, the donor part, containing the fatty acid and thiolated threonine, is connected to the acceptor part, containing the A2 domain-activated glutamine, which is epimerized by the E domain following A2 before the next extension step (Fig. 6). The process continues in a canonical fashion until the A4 domain-activated glycine is connected. When this occurs, the whole LocB peptide is used a second time for further extension (Fig. 6). This time A2 activates asparagine rather than glutamine, which is connected to the donor without epimerization. Extension then proceeds through LocB and LocC all the way down to the TE domain, which catalyzes the cyclization of the synthesized linear peptide by connecting the carbonyl group of Val9 and the hydroxyl group of the Thr1 side chain before releasing the mature cyclic lipopeptide of locillomycins (Fig. 6) (39). One should notice in particular that the first A domain (A1) and the last two A domains (A5 and A6) are used only once, whereas the three A domains in the middle (A2, A3, and A4) are used iteratively during the biosynthesis (Fig. 6). Moreover, the A2 domain in LocB somehow exhibits a different specificity in the iteration. Also, the epimerization domain following the A2 domain, which is responsible for conversion of L-Gln to D-Gln the first time, is skipped the second time (Fig. 6). The different selectivities of these catalytic domains within the same NRPS assembly line are quite unusual compared to other NRPS assembly lines. Thus, compared to the iterative uses in type B and C NRPSs, the *loc*-encoded NRPS was designated a type D NRPS (Fig. 6) (25–28).

A typical lipopeptide biosynthesis pathway involves the KAS domain catalyzing the formation of a β-ketoacyl thioester, which is subsequently converted into a β-substituted fatty acid by a transamination or transhydroxylation reaction. This is followed by transfer of the fatty acid to a T domain and is coupled to the first activated amino acid. In contrast, with regard to the synthesis of locillomycin, only the ACS domain, not the KAS domain, is responsible for the activation of the fatty acid. The skipping of the KAS domain may be due to the lack of any functional domains, such as an aminotransferase or hydroxyl transferase, downstream of the KAS domain. The functions of the redundant T and C domains before LocA are unclear, since one condensation domain preceding the first A domain of the peptide synthetase is enough to couple the fatty acid to the first activated amino acid (2). While a few of these atypical biochemical features of single-module or multiple-module iteration have been observed previously, the locillomycins represent a rare, nonlinear NRPS biosynthetic model with a combination of iteration, skipping, alternation, and specificity changes (Fig. 6).

Biological functions of locillomycins. Biological activity assays revealed that locillomycins have moderate antibacterial activities (see Fig. S11 in the supplemental material). The antibacterial MICs of locillomycins A, B, and C were 24.3, 18.8, and 17.6 μg/ml, respectively, for methicillin-resistant *Staphylococcus aureus* (MRSA) and were 6.3, 5.8, and 5.4 μg/ml, respectively, for *Xanthomonas oryzae* pv. oryzae. Antiviral results for different locillomycin concentrations revealed that porcine epidemic diarrhea virus (PEDV) infection could be inhibited effectively by these compounds. In particular, at a concentration of 10 μg/ml, locillomycins reduced the number of virus copies 300-fold and were about 10 times as efficient as surfactins (see Fig. S12). Locillomycins also have lower hemolytic activities than surfactins and are presumed to reduce toxicity against eukaryotic cells, which could enhance their potential for use in therapeutic applications (see Fig. S12) (3).

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (grant 31570061), the National High-Tech R&D Program of China (grant 2011AA10A201), and the Jiangsu Independent Innovation Fund [grant CX(12)5001].

We are grateful to L. Du (Department of Chemistry, University of Nebraska-Lincoln) for a critical review and revision.

We declare no competing financial interests. The manuscript was written through contributions of all authors. All authors have given approval for the final version of the manuscript.

REFERENCES

1. **Stein T.** 2005. *Bacillus subtilis* antibiotics: structures, syntheses and specific functions. Mol Microbiol **56:**845–857. http://dx.doi.org/10.1111/j.1365-2958.2005.04587.x.
2. **Ongena M, Jacques P.** 2008. *Bacillus* lipopeptides: versatile weapons for plant disease biocontrol. Trends Microbiol **16:**115–125. http://dx.doi.org/10.1016/j.tim.2007.12.009.
3. **Rodrigues L, Banat IM, Teixeira J, Oliveira R.** 2006. Biosurfactants:

potential applications in medicine. J Antimicrob Chemother **57:**609–618. http://dx.doi.org/10.1093/jac/dkl024.

4. **Romero D, de Vicente A, Rakotoaly RH, Dufour SE, Veening JW, Arrebola E, Cazorla FM, Kuipers OP, Paquot M, Pérez-García A.** 2007. The iturin and fengycin families of lipopeptides are key factors in antagonism of *Bacillus subtilis* toward *Podosphaera fusca*. Mol Plant Microbe Interact **20:**430–440. http://dx.doi.org/10.1094/MPMI-20-4-0430.
5. **Ongena M, Jourdan E, Adam A, Paquot M, Brans A, Joris B, Arpigny JL, Thonart P.** 2007. Surfactin and fengycin lipopeptides of *Bacillus subtilis* as elicitors of induced systemic resistance in plants. Environ Microbiol **9:**1084–1090. http://dx.doi.org/10.1111/j.1462-2920.2006.01202.x.
6. **Zeriouh H, de Vicente A, Pérez-García A, Romero D.** 2014. Surfactin triggers biofilm formation of *Bacillus subtilis* in melon phylloplane and contributes to the biocontrol activity. Environ Microbiol **16:**2196–2211. http://dx.doi.org/10.1111/1462-2920.12271.
7. **Han Q, Wu F, Wang X, Qi H, Shi L, Ren A, Liu Q, Zhao M, Tang C.** 2015. The bacterial lipopeptide iturins induce *Verticillium dahliae* cell death by affecting fungal signalling pathways and mediate plant defence responses involved in pathogen-associated molecular pattern-triggered immunity. Environ Microbiol **17:**1166–1188. http://dx.doi.org/10.1111/1462-2920.12538.
8. **Duitman EH, Hamoen LW, Rembold M, Venema G, Seitz H, Saenger W, Bernhard F, Reinhardt R, Schmidt M, Ullrich C, Stein T, Leenders F, Vater J.** 1999. The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. Proc Natl Acad Sci U S A **96:**13294–13299. http://dx.doi.org/10.1073/pnas.96.23.13294.
9. **Tsuge K, Akiyama T, Shoda M.** 2001. Cloning, sequencing, and characterization of the iturin A operon. J Bacteriol **183:**6265–6273. http://dx.doi.org/10.1128/JB.183.21.6265-6273.2001.
10. **Tosato V, Albertinij AM, Zotti M, Sondal S, Bruschi CV.** 1997. Sequence completion, identification and definition of the fengycin operon in *Bacillus subtilis* 168. Microbiology **143:**3443–3450. http://dx.doi.org/10.1099/00221287-143-11-3443.
11. **Kessler N, Schuhmann H, Morneweg S, Linne U, Marahiel MA.** 2004. The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains. J Biol Chem **279:**7413–7419. http://dx.doi.org/10.1074/jbc.M309658200.
12. **Stachelhaus T, Mootz HD, Bergendahl V, Marahiel MA.** 1998. Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. J Biol Chem **273:**22773–22781.
13. **Belshaw PJ, Walsh CT, Stachelhaus T.** 1999. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. Science **284:**486–489. http://dx.doi.org/10.1126/science.284.5413.486.
14. **Sieber SA, Marahiel MA.** 2003. Learning from nature's drug factories: nonribosomal synthesis of macrocyclic peptides. J Bacteriol **185:**7036–7043. http://dx.doi.org/10.1128/JB.185.24.7036-7043.2003.
15. **Mootz HD, Schwarzer D, Marahiel MA.** 2000. Construction of hybrid peptide synthetases by module and domain fusions. Proc Natl Acad Sci U S A **97:**5848–5853. http://dx.doi.org/10.1073/pnas.100075897.
16. **Hoefler BC, Gorzelnik KV, Yang JY, Hendricks N, Dorrestein PC.** 2012. Enzymatic resistance to the lipopeptide surfactin as identified through imaging mass spectrometry of bacterial competition. Proc Natl Acad Sci U S A **109:**13082–13087. http://dx.doi.org/10.1073/pnas.1205586109.
17. **Zaman M, Toth I.** 2013. Immunostimulation by synthetic lipopeptide-based vaccine candidates: structure-activity relationships. Front Immunol **4:**318. http://dx.doi.org/10.3389/fimmu.2013.00318.
18. **Sieber SA, Marahiel MA.** 2005. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. Chem Rev **105:**715–738. http://dx.doi.org/10.1021/cr0301191.
19. **Robbel L, Marahiel MA.** 2010. Daptomycin, a bacterial lipopeptide synthesized by a nonribosomal machinery. J Biol Chem **285:**27501–27508. http://dx.doi.org/10.1074/jbc.R110.128181.
20. **Magarvey NA, Haltli B, He M, Greenstein M, Hucul JA.** 2006. Biosynthetic pathway for mannopeptimycins, lipoglycopeptide antibiotics active against drug-resistant gram-positive pathogens. Antimicrob Agents Chemother **50:**2167–2177. http://dx.doi.org/10.1128/AAC.01545-05.
21. **Mascio CTM, Mortin LI, Howland KT, Van Praagh ADG, Zhang S, Arya A, Chuong CL, Kang C, Li T, Silverman JA.** 2012. *In vitro* and *in vivo* characterization of CB-183,315, a novel lipopeptide antibiotic for treatment of *Clostridium difficile*. Antimicrob Agents Chemother **56:**5023–5030. http://dx.doi.org/10.1128/AAC.00057-12.
22. **Miao V, Coëffet-Legal M-F, Brian P, Brost R, Penn J, Whiting A, Martin S, Ford R, Parr I, Bouchard M, Silva CJ, Wrigley SK, Baltz RH.** 2005. Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry. Microbiology **151:**1507–1523. http://dx.doi.org/10.1099/mic.0.27757-0.
23. **Owen JG, Reddy BVB, Ternei MA, Charlop-Powers Z, Calle PY, Kim JH, Brady SF.** 2013. Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. Proc Natl Acad Sci U S A **110:**11797–11802. http://dx.doi.org/10.1073/pnas.1222159110.
24. **Rosconi F, Davyt D, Martínez V, Martínez M, Abin-Carriquiry JA, Zane H, Butler A, de Souza EM, Fabiano E.** 2013. Identification and structural characterization of serobactins, a suite of lipopeptide siderophores produced by the grass endophyte *Herbaspirillum seropedicae*. Environ Microbiol **15:**916–927. http://dx.doi.org/10.1111/1462-2920.12075.
25. **Felnagle EA, Jackson EE, Chan YA, Podevels AM, Berti AD, McMahon MD, Thomas MG.** 2008. Nonribosomal peptide synthetases involved in the production of medically relevant natural products. Mol Pharm **5:**191–211. http://dx.doi.org/10.1021/mp700137g.
26. **Tang MC, Fu CY, Tang GL.** 2012. Characterization of SfmD as a heme peroxidase that catalyzes the regioselective hydroxylation of 3-methyltyrosine to 3-hydroxy-5-methyltyrosine in saframycin A biosynthesis. J Biol Chem **287:**5112–5121. http://dx.doi.org/10.1074/jbc.M111.306316.
27. **Dimise EJ, Widboom PF, Bruner SD.** 2008. Structure elucidation and biosynthesis of fuscachelins, peptide siderophores from the moderate thermophile Thermobifida fusca. Proc Natl Acad Sci U S A **105:**15311–15316. http://dx.doi.org/10.1073/pnas.0805451105.
28. **Peng C, Pu J, Song L, Jian X, Tang M, Tang G.** 2012. Hijacking a hydroxyethyl unit from a central metabolic ketose into a nonribosomal peptide assembly line. Proc Natl Acad Sci U S A **109:**8540–8545. http://dx.doi.org/10.1073/pnas.1204232109.
29. **Moss SJ, Martin CJ, Wilkinson B.** 2004. Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. Nat Prod Rep **21:**575–593. http://dx.doi.org/10.1039/b315020h.
30. **Gruenewald S, Mootz HD, Stehmeier P, Stachelhaus T.** 2004. *In vivo* production of artificial nonribosomal peptide products in the heterologous host *Escherichia coli*. Appl Environ Microbiol **70:**3282–3291. http://dx.doi.org/10.1128/AEM.70.6.3282-3291.2004.
31. **Tsuge K, Inoue S, Ano T, Itaya M, Shoda M.** 2005. Horizontal transfer of iturin A operon, itu, to *Bacillus subtilis* 168 and conversion into an iturin A producer. Antimicrob Agents Chemother **49:**4641–4648. http://dx.doi.org/10.1128/AAC.49.11.4641-4648.2005.
32. **Park SY, Choi SK, Kim J, Oh TK, Park SH.** 2012. Efficient production of polymyxin in the surrogate host *Bacillus subtilis* by introducing a foreign ectB gene and disrupting the abrB gene. Appl Environ Microbiol **78:**4194–4199. http://dx.doi.org/10.1128/AEM.07912-11.
33. **Luo C, Zhou H, Zou J, Wang X, Zhang R, Xiang Y.** 2015. Bacillomycin L and surfactin contribute synergistically to the phenotypic features of *Bacillus subtilis* 916 and the biocontrol of rice sheath blight induced by *Rhizoctonia solani*. Appl Microbiol Biotechnol **99:**1897–1910. http://dx.doi.org/10.1007/s00253-014-6195-4.
34. **Luo C, Liu X, Zhou H, Wang X, Chen Z.** 2015. Nonribosomal peptide synthase gene clusters for lipopeptide biosynthesis in *Bacillus subtilis* 916 and their phenotypic functions. Appl Environ Microbiol **81:**422–431. http://dx.doi.org/10.1128/AEM.02921-14.
35. **Markley J, Bax A, Arata Y, Hibers C, Kaptein R, Sykes B, Wright PE, Wuthrich K.** 1998. Recommendations for the presentation of NMR structures of proteins and nucleic acids—IUPAC-IUBMB-IUPAB Inter-Union Task Group on the standardization of data bases of protein and nucleic acid structures determined by NMR spectroscopy. Eur J Biochem **256:**1–15. http://dx.doi.org/10.1046/j.1432-1327.1998.2560001.x.
36. **Wüthrich K.** 1986. NMR of proteins and nucleic acids. Wiley, New York, NY.
37. **Wang X, Luo C, Chen Z.** 2012. Genome sequence of the plant growth-promoting rhizobacterium *Bacillus* sp. strain 916. J Bacteriol **194:**5467–5468. http://dx.doi.org/10.1128/JB.01266-12.
38. **Lautru S, Challis GL.** 2004. Substrate recognition by nonribosomal peptide synthetase multi-enzymes. Microbiology **150:**1629–1636. http://dx.doi.org/10.1099/mic.0.26837-0.
39. **Du L, Lou L.** 2010. PKS and NRPS release mechanisms. Nat Prod Rep **27:**255–278. http://dx.doi.org/10.1039/B912037H.