Identification and characterization of andalusicin: N-terminally dimethylated class III lantibiotic from Bacillus thuringiensis sv. andalousiensis

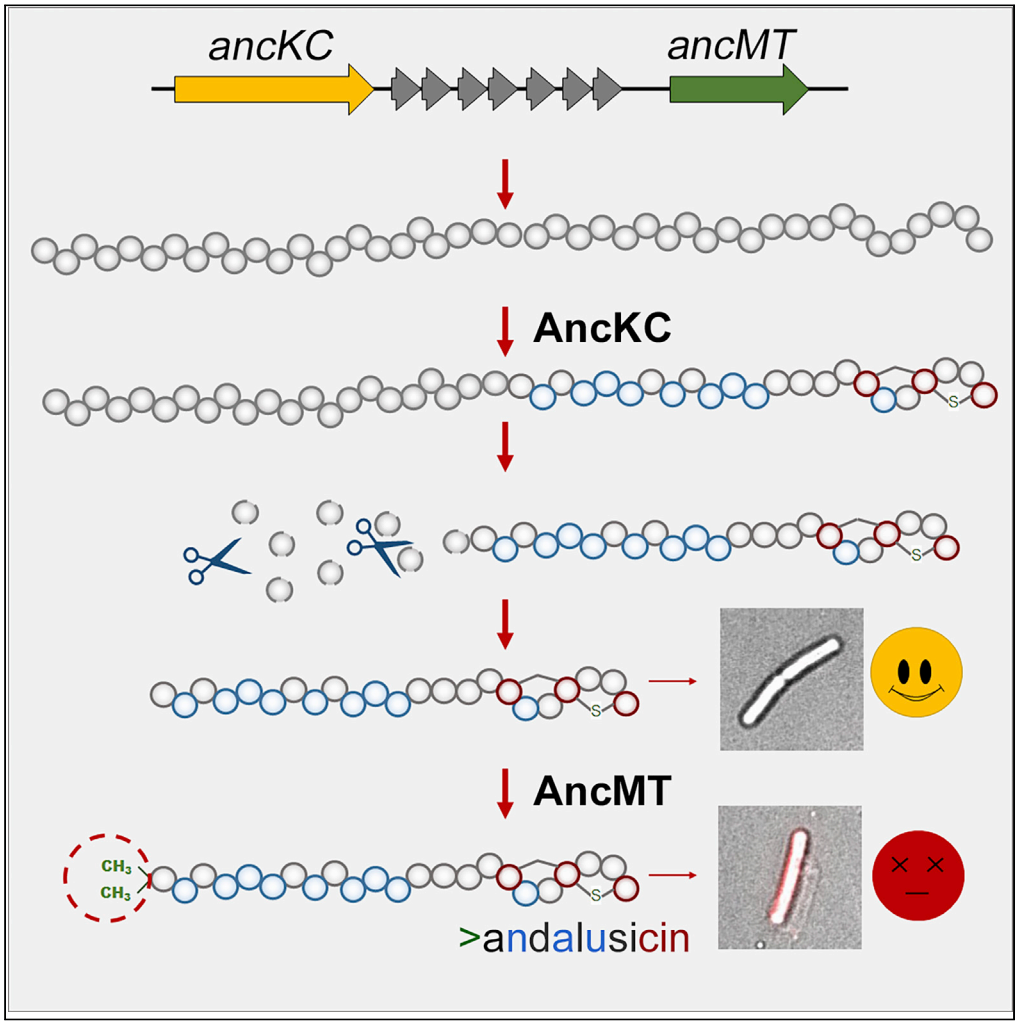
Anastasiia Grigoreva, Julia Andreeva, Dmitry Bikmetov, Anastasiia Rusanova, Marina Serebryakova, Andrea Hernandez Garcia, Darya Slonova, Satish K Nair, Guy Lippens, Konstantin Severinov, et al.

iScience

CellPress
OPEN ACCESS

Article

Identification and characterization of andalusicin: N-terminally dimethylated class III lantibiotic from *Bacillus thuringiensis* sv. *andalousiensis*



Anastasiia Grigoreva, Julia Andreeva, Dmitry Bikmetov, ..., Guy Lippens, Konstantin Severinov, Svetlana Dubiley

severik@waksman.rutgers.edu (K.S.)
svetlana.dubiley@gmail.com (S.D.)

Highlights

Type III lanthipeptide andalusicin A inhibits the growth of Gram-positive bacteria

Andalusicin A has an unusual pastor's crook structure

Genes encoding methyltransferases are frequently found in andalusicin-like BGCs

N-terminal methylation is necessary for andalusicin A biological activity

Grigoreva et al., iScience *24*, 102480
May 21, 2021
https://doi.org/10.1016/j.isci.2021.102480

iScience

CellPress
OPEN ACCESS

Article

Identification and characterization of andalusicin: N-terminally dimethylated class III lantibiotic from *Bacillus thuringiensis* sv. *andalousiensis*

Anastasiia Grigoreva,[1,2] Julia Andreeva,[1,2] Dmitry Bikmetov,[3,4] Anastasiia Rusanova,[2] Marina Serebryakova,[2,5] Andrea Hernandez Garcia,[6] Darya Slonova,[1] Satish K. Nair,[6] Guy Lippens,[7] Konstantin Severinov,[1,3,4,8,9,*] and Svetlana Dubiley[1,2,*]

SUMMARY

Lanthipeptides, ribosomally synthesized and post-translationally modified peptides (RiPPs), can be divided into five classes based on their structures and biosynthetic pathways. Class I and II lanthipeptides have been well characterized, whereas less is known about members of the other three classes. Here, we describe a new family of class III lanthipeptides from *Firmicutes*. Members of the family are distinguished by the presence of a single carboxy-terminal labionin. We identified and characterized andalusicin, a representative of this family. Andalusicin bears two methyl groups at the α-amino terminus, a post-translational modification that has not previously been identified in class III lanthipeptides. Mature andalusicin A shows bioactivity against various Gram-positive bacteria, an activity that is highly dependent on the α-N dimethylation.

INTRODUCTION

Lanthipeptides represent the largest known family of ribosomally synthesized and post-translationally modified peptides (RiPPs). They are synthesized as linear precursor peptides by the ribosome and then undergo post-translational modifications (PTMs) (Arnison et al., 2013; Montalbán-López et al., 2021). The PTMs are localized in the core part of the precursor peptide, whereas the leader sequence recruits the modification enzymes. After the modifications are installed, the leader is cleaved off during (or sometimes after) export from the producing cell (Oman and van der Donk, 2010). Genes required for biosynthesis of bacterial RiPPs usually form compact biosynthetic gene clusters (BGCs). Lanthipeptide BGCs contain gene(s) encoding the precursor peptide(s) and enzymes required for post-translational introduction of the family-defining lanthionine cross-link in the precursor (Figure 1). In addition, some lanthipeptide BGCs contain accessory genes coding for export pumps and/or self-immunity mechanisms (Stein et al., 2003), leader peptidases (Lagedroste et al., 2017), and various tailoring enzymes responsible for secondary PTMs (Acedo et al., 2019; Iorio et al., 2014; Wiebach et al., 2018).

The lanthionine (a β-thioether linkage) is installed in two steps, which are catalyzed by a multifunctional lanthionine synthetase (in class II–IV lanthipeptides), the sequential action of dehydratase and cyclase (in class I lanthipeptides) (Repka et al., 2017), or putative lyase, kinase, and cyclase (in class V) (Xu et al., 2020). Select Ser or Thr residues of the precursor are dehydrated to dehydroalanine (Dha)/dehydrobutyrine (Dhb). Next, a thiol group of a Cys is added to the double bond in a 1,4 conjugate addition reaction to form either a lanthionine (Lan) or a methyl-lanthionine (MeLan), respectively (Knerr and van der Donk, 2012). C-terminal lanthionine and MeLan can undergo decarboxylation resulting in (2-aminovynyl)-cysteine (AviCys) and 2-aminovinyl-3-methyl-cysteine (AviMeCys), respectively (Ortega et al., 2017; Schneider et al., 2000). In some cases lanthipeptides contain more complex labionin (Lab)/methyllabionin (MeLab) structures, which form upon the dehydration and macrocyclization of three residues, Ser/Ser/Cys or Ser/Thr/Cys, respectively (Figure 1) (Meindl et al., 2010).

Class III–V lanthipeptides are the least studied, and only a handful of representatives of either of these classes are known. Class III–IV lanthipeptides are synthetized by a multifunctional lanthionine synthetase composed of three domains (Figure 1): the central kinase domain that activates Ser and Thr residues through phosphorylation; the N-terminal lyase domain, responsible for phosphate elimination; and the

[1]Center of Life Sciences, Skolkovo Institute of Science and Technology, 121205 Moscow, Russia

[2]Institute of Gene Biology, Russian Academy of Sciences, Moscow 119334, Russia

[3]Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Institute of Gene Biology, Russian Academy of Sciences, Moscow 119334, Russia

[4]Institute of Molecular Genetics, Russian Academy of Sciences, Moscow 123182, Russia

[5]A.N. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow 119991, Russia

[6]Department of Biochemistry, University of Illinois at Urbana-Champaign, Champaign, IL 61801 USA

[7]Toulouse Biotechnology Institute (TBI), Université de Toulouse, CNRS, INRA, INSA, Toulouse 31077, France

[8]Waksman Institute for Microbiology, Piscataway, NJ 08854-8020, USA

[9]Lead contact

*Correspondence: severik@waksman.rutgers.edu (K.S.), svetlana.dubiley@gmail.com (S.D.)

https://doi.org/10.1016/j.isci.2021.102480

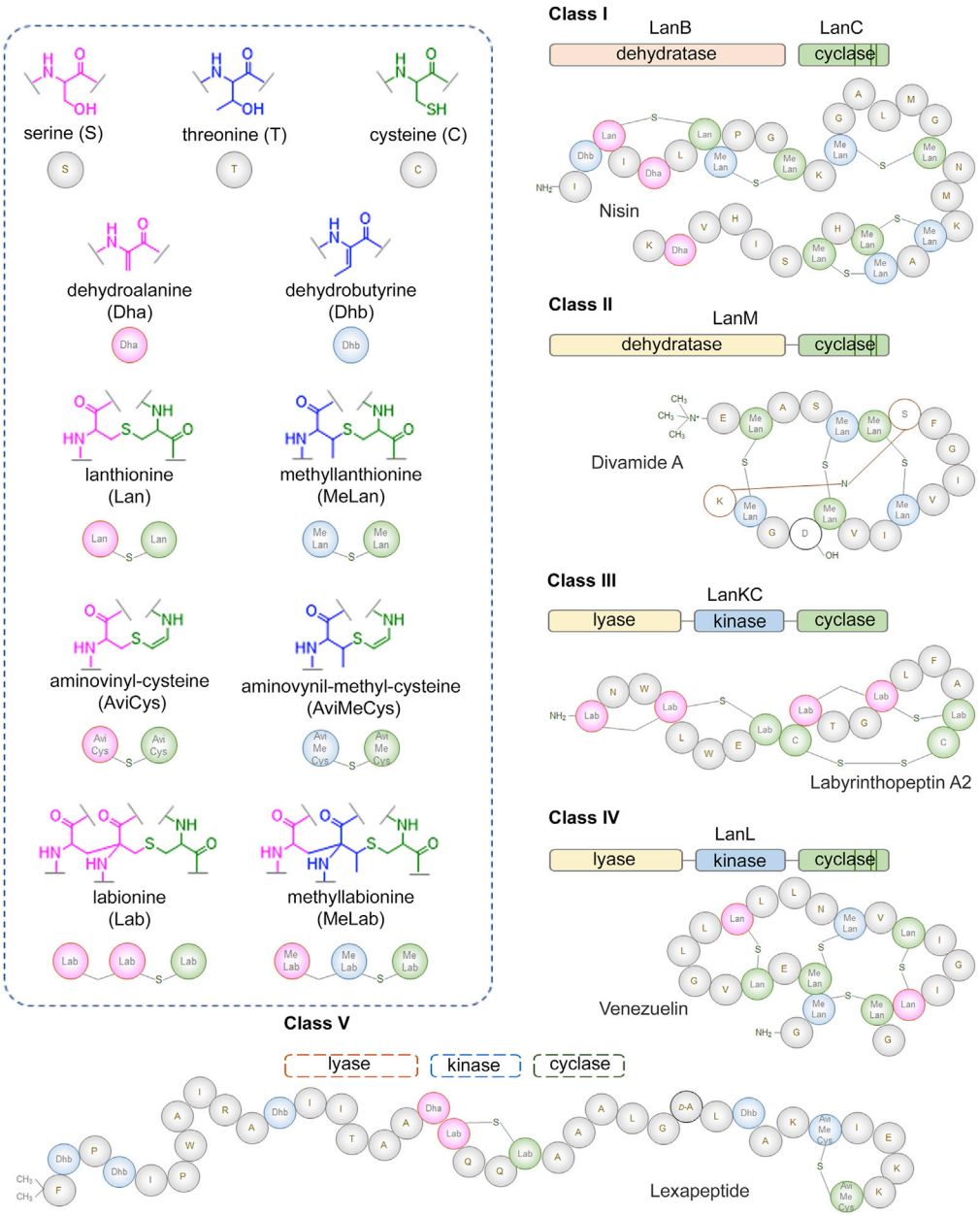


Figure 1. Five classes of lanthipeptides

Schematic domain organizations of lanthionine synthetases from different classes are depicted above representative lanthipeptides from each class. Conserved zinc-binding residues in the cyclase domains are marked by vertical lines. Post-translational modifications common for Ser, Thr, and Cys residues of lanthipeptides are shown in magenta, blue, and green colors, respectively.

C-terminal cyclization domain (LanC) that catalyzes the conjugate addition (Hegemann et al., 2019). The latter domain discriminates class III (LanKC) and IV (LanL) enzymes, as LanKC C-terminal domains do not contain the characteristic zinc-binding residues conserved in other cyclases (Wang and van der Donk, 2012). The unique feature of LanKC enzymes is the ability to install labionin/methyllabionin structures; however, the exact mechanism of this process remains unknown (Hegemann and Süssmuth, 2020).

Here, we report the identification and functional characterization of a novel lanthipeptide biosynthetic gene cluster from *Bacillus thuringiensis* sv. *andalousiensis* NRRL B23139 and structural characterization

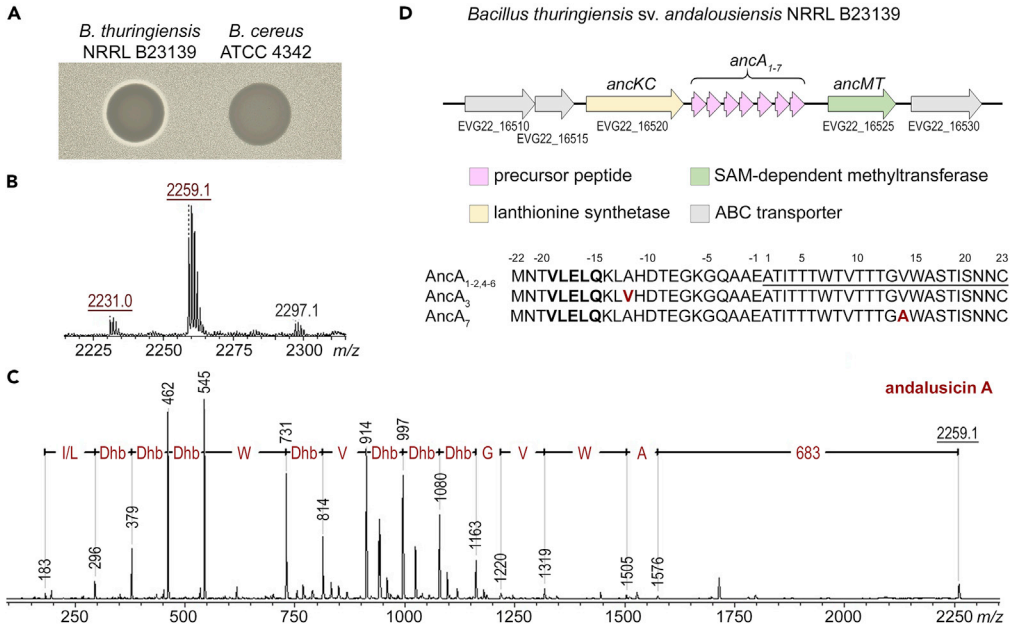


Figure 2. *B. thuringiensis* sv. *andalousiensis* NRRL B23139 produces lanthipeptides

(A) Antimicrobial activity assay of *B. thuringiensis* sv. *andalousiensis* NRRL B23139 against *B. cereus* ATCC 4342. A small clear halo around a patch of *B. thuringiensis* NRRL B23139 cells indicates the production of a secreted antimicrobial compound active against *B. cereus* ATCC 4342 cells forming the surrounding lawn. A patch of *B. cereus* ATCC 4342 cells was used as a negative control (no production of inhibitory compounds expected).

(B) MALDI-TOF-MS analysis of a *B. thuringiensis* cellular extract.

(C) MALDI-TOF-MS/MS spectrum of andalusicin A. Only *b*-ions are labeled for clarity. Amino acids are marked according to standard classification; Dhb stands for dehydrobutyrine. A difference in 683 Da between the precursor ion and the first *b*-ion matches the mass of the C-terminal STISNCC fragment of AncA1/A2 peptides assuming that all Ser and Thr residues are dehydrated.

(D) A scheme of *B. thuringiensis* sv. *andalousiensis* NRRL B23139 lanthipeptide biosynthesis cluster (drawn not to scale) and sequences of AncA precursor peptides. Single amino acid differences are shown in the red-color font. Sequences of core peptides are underlined. Sequences corresponding to known consensus of class III/IV leader peptide (Repka et al., 2017) are shown in bold.

of its labionin-containing product, which we named andalusicin. We identified a methyltransferase responsible for α-*N*-methylation of andalusicin, a tailoring modification previously unknown in class III lanthipeptides. We show that this α-*N*-methylation is required for antibacterial activity of andalusicin.

RESULTS

A novel lantibiotic biosynthetic gene cluster in the *Bacillus thuringiensis* NRRL B23139 genome

During screening of laboratory collection of microorganisms for bioactivity, we observed that *B. thuringiensis* sv. *andalousiensis* NRRL B23139 inhibits growth of *B. cereus* ATCC 4342 (Figure 2A). To identify the compound responsible for antimicrobial activity, the cellular extract of *B. thuringiensis* sv. *andalousiensis* NRRL B23139 was fractionated using reverse-phase high-performance liquid chromatography (HPLC) (Figures S1A and S1B). MALDI-TOF-mass spectrometric (MS) analysis of the active fraction revealed a prominent $[M + H]^+$ ion at *m/z* 2,259.1 (Figure S1B), which was also present in the starting extract (Figure 2B). MALDI-TOF-tandem MS (MS/MS) showed that the corresponding protonated molecule is a heavily modified peptide likely containing a number of dehydrobutyrines (Dhb), arising from dehydration of threonine residues (Figure 2C). Analysis of fragmentation spectra suggested that the bioactive compound may be a RiPP with its precursor a peptide containing an (I/L)TTTWTVTTTGVWA sequence.

We sequenced, assembled, and annotated the genome of *B. thuringiensis* sv. *andalousiensis* NRRL B23139 (GenBank: NZ_CP035727.2). Using the predicted partial sequence of the bioactive peptide as a query for

BlastP search, we identified a compact locus containing seven short open reading frames (ORFs), whose products matched the predicted partial sequence of the bioactive RiPP. Five of the seven genes code for identical 45-amino acid-long peptides (AncA1-2, AncA4-6), whereas two genes encode peptides harboring an alanine to valine substitution in the putative leader (AncA3) and in the core (AncA7) parts, respectively (Figure 2D). A gene encoding a class III lanthionine synthetase is located upstream of the peptide-coding locus. We therefore hypothesized that in addition to dehydrated residues, the identified bioactive compound may contain a lanthionine cross-link. The presence of the LxLQ motif characteristic for class III/IV lanthipeptides (Müller et al., 2011) in the leader part of the AncA peptides supports this hypothesis. Since expected C-terminal ions are absent from the MALDI-TOF-MS/MS spectra in the Figure 2C, a lanthionine could be formed by the C-terminal Cys23 and dehydrated Thr18, Ser17, or Ser20 residues.

Assuming that all Ser and Thr residues in the core parts of AncA peptides are modified and that the 22-amino acid long N-terminal leader is cleaved off, the calculated *m/z* values of protonated mature AncA1-6 should be 2,231.0. However, the observed mass of produced bioactive compound is higher by 28 Da. Careful inspection of mass spectra of the HPLC fractions identified a $[M + H]^+$ ion at *m/z* 2,231.0. No biological activity was evident in fractions containing this compound (Figures S1A and S1C). Although the measured mass of this compound matches the calculated mass of fully modified AncA1-6 peptides, its fragmentation pattern corresponds to AncA7 (Figure S1D). Similar to the major compound, the additional 28 Da were assigned to the N-terminal dipeptide of AncA7-based compound. We named the major compound produced from AncA1-6 peptides andalusicin A, and the minor compound produced from AncA7, andalusicin B.

Andalusicins are dimethylated

The *ancMT* gene encoding a putative S-adenosyl methionine (SAM)-dependent methyltransferase is located immediately downstream of the andalusicin peptide-coding locus. We speculated that *ancMT* is part of the andalusicin BGC and that its product may be responsible for dimethylation of andalusicins. To test this hypothesis, we expressed subsets of *anc* genes in a heterologous host. When a minimal set of two genes, *ancKC-ancA1,* under control of an inducible promoter was introduced in *B. subtillis* 168 strain, a modified peptide (an $[M + H]^+$ ion at *m/z* 2,231.0) accumulated in induced cells. Its molecular mass and fragmentation spectrum matched those of andalusicin A but for the extra 28 Da (Figures 3A, upper panel, and 3B). When the *ancKC-ancA1* genes were introduced together with *ancMT*, a 28-Da shift of the moified peptide appeared (Figure 3A) and the MALDI fragmentation spectrum of the $[M + H]^+$ ion at *m/z* 2,259.1 was identical to that obtained for andalusicin A secreted by the native producer. These results confirm that methyltransferase AncMT is responsible for the addition of two methyl groups to andalusicin. The methylation reaction was also reconstructed *in vitro* using purified desmethyl andalusicin A, recombinant AncMT, and SAM as a methyl donor. As shown in Figure 3C, AncMT successfully converted desmethyl andalusicin A ($[M + H]^+$ ion at *m/z* 2,231.0) to its double-methylated form ($[M + H]^+$ ion at *m/z* 2,259.1). The fragmentation spectrum of this *in vitro* reaction product was identical to that of andalusicin A secreted by the native producer.

Andalusicin A showed potent antimicrobial activity against *B. cereus* ATCC 4342 in the broth dilution assay with a minimal inhibitory concentration of 16.2 μg/mL, which is comparable to the reported activity of nisin (10 μg/mL) (Pol and Smid, 1999). To test the biological effect of dimethylation, we compared antimicrobial activities of dimethylated and desmethyl andalusicin. Dimethylated andalusicin A was active against all Gram-positive bacteria tested, with closely related *Bacillus* strains being the most sensitive (Table S1). In the absence of methylation, antimicrobial activity against most bacteria tested was abolished. Weak growth inhibition by desmethyl andalusicin A was only observed for *B. cereus* ATCC 14579. We conclude that dimethylation dramatically increases the bioactivity of andalusicin A.

To get an insight into the mechanism of antimicrobial activity of andalusicin, we monitored membrane integrity of *B. cereus* cells treated with andalusicin by staining with propidium iodide (Boulos et al., 1999). As expected, treatment with proton ionophore carbonyl cyanide 3-chlorophenylhydrazone (CCCP) did not result in cell lysis or propidium iodide accumulation inside the cells (Figure 4 and Table S2). In contrast, the addition of nisin caused intracellular accumulation of propidium iodide. Treatment with andalusicin had a similar effect (Figure 4 and Table S2). Nisin binds cell-wall precursor lipid II (Hsu et al., 2004) and induces large pore formation (Wiedemann et al., 2004) and lipid loss in the membrane

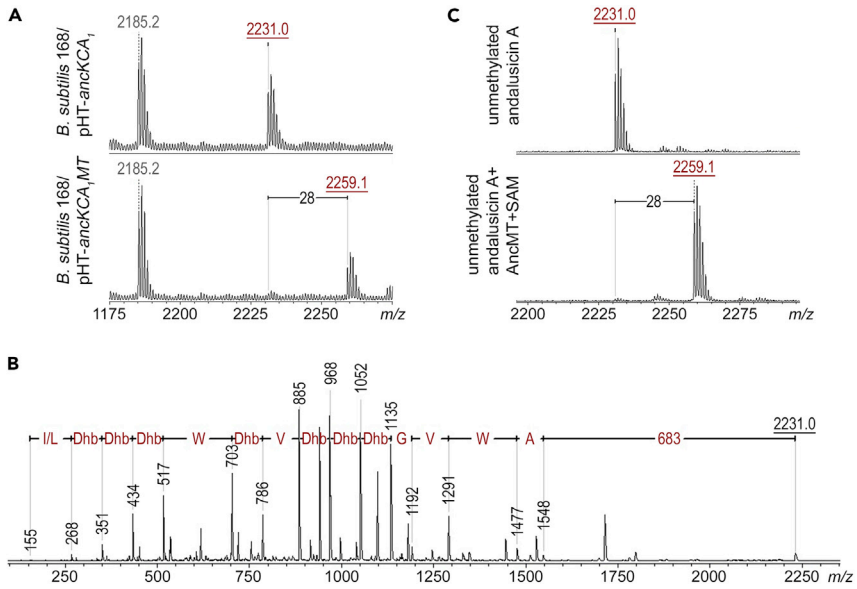


Figure 3. AncMT is required for andalusicin A methylation *in vivo* and *in vitro*

(A) MALDI-TOF-MS analysis of cellular extracts of *B. subtilis* 168 harboring pHT-*ancKCA1* (upper panel) and pHT-*ancKCA1S* (lower panel). An $[M + H]^+$ ion at *m/z* 2,231.0 and $[M + H]^+$ ion at *m/z* 2,259.1 matched desmethyl and dimethylated forms of andalusicin A, respectively. The $[M + H]^+$ ion at *m/z* 2,185.2 corresponds to a *B. subtilis* 168 cellular protein.

(B) MALDI-TOF-MS/MS spectrum of desmethyl andalusicin A. Only *b*-ions are labeled for clarity. Amino acids are marked as in Figure 2C.

(C) The methyltransferase reaction substrate, desmethyl andalusicin A ($[M + H]^+$ ion at *m/z* 2,231.0, upper panel), and the product, dimethylated andalusicin A ($[M + H]^+$ ion at *m/z* 2,259.1, lower panel) are shown. The reaction was carried out in the presence of SAM and the AncMT methyltransferase. The 28 Da difference between the substrate and the product corresponds to the addition of two methyl groups.

of sensitive cells (Prince et al., 2016). Although both andalusicin and nisin induced membrane permeabilization for propidium iodide, their actual mechanism of pore formation might differ.

Structural analysis of andalusicin A

Bioinformatics analysis shows that AncKC is a member of class III lanthionine synthetases. These enzymes are able to install both lanthionine and labionin modifications, the latter being a unique cross-link present in only a handful of lanthipeptides (Hegemann and Süssmuth, 2020). As labionin/lanthionine cross-link formation does not change the net mass of the fully dehydrated peptide, MS alone cannot reveal the nature of the macrocycle. Susceptibility of the peptide to further chemical modifications provides indirect evidence of lanthipeptide cross-link structure. Under mild alkaline conditions, substituted thiols convert the Dha and Dhb residues to S-alkylated cysteine and β-methylcysteine, respectively, while leaving the lanthionine- or labionin-forming serine and threonine residues intact (Lohans and Vederas, 2014). The reaction with β-mercaptoethanol results in a 78 Da increase per Dha/Dhb residue. MS analysis of β-mercaptoethanol-treated andalusicin A revealed a 702-Da shift, which corresponds to alkylation of nine Dha/Dhb residues (Figures S2A–S2C). This suggests that 9 of 11 dehydrated serine and threonine residues in andalusicin A exist as Dha/Dhb, whereas the remaining two Dha/Dhb residues may be involved in a labionin cross-link.

Treatment of lantibiotics with nickel boride leads to peptide linearization through desulfurization of lanthionine cross-link (Martin et al., 2004). Dha and Dhb are converted, respectively, to alanine and α-aminobutyric acid (Abu), whereas lanthionine is converted to two Ala residues that become susceptible to fragmentation during tandem MS. However, the fragmentation pattern of andalusicin A treated with nickel boride lacked product ion peaks in the Ser17–Ser20 region (Figures S2D and S2E). This could be the case if andalusicin contained a labionin in which the thioether bridge between residues 20 and 23 was desulfurized, but the carbocycle-forming residues 17 and 20 remained intact.

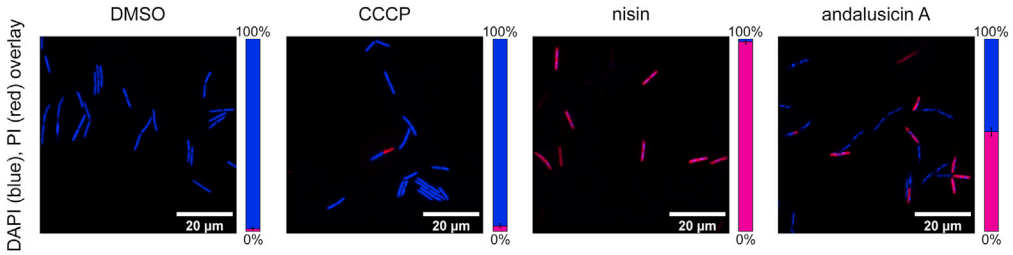


Figure 4. Effect of andalusicin A on *B. cereus* ATCC 4342

Micrographs of DAPI and propidium iodide (PI)-stained *B. cereus* cells treated with DMSO (1%), CCCP (500 µM), nisin (6 µM), and andalusicin A (29 µM). Stacked bar diagrams show the percentage of PI-positive cells with 95% confidence intervals.

For cell counts, see Table S2.

Solubility of andalusicin A in deuterated DMSO proved satisfactory to obtain high-resolution NMR spectra. Although the presence of 8 Dhb residues in the N-terminal part of the peptide hampered complete NMR characterization, we could assign all signals of amino acids with Hα protons based on combined homonuclear Total Correlation Spectroscopy (TOCSY) and Nuclear Overhauser Spectroscopy (NOESY) experiments complemented with ^{1}H-^{13}C Heteronuclear Single Quantum Correlation (HSQC) and Heteronuclear Multiple Bond Correlation (HMBC) spectra. As the only amino acid devoid of an amide proton, we identified the N-terminal Ala as the peak at 3.12/66.8 ppm for its ^{1}Hα/^{13}Cα group and 1.11/16.08 ppm for its methyl moiety (Figure 5A). Importantly, the random coil value for Ala ^{13}Cα is at 50.5 ppm (Wishart et al., 1995), but the 16 ppm downfield shift we observe is compatible with the ^{13}C chemical shift perturbation of 17 ppm observed in Lys side chain ^{13}Cε upon dimethylation (Theillet et al., 2012). Furthermore, exploiting the long-range connectivity in the HMBC spectrum, we observed a distinct cross-peak between the Ala ^{1}Hα/^{13}Cα correlation and an intense peak at 2.226/43.4 ppm, whereby the latter values coincide with the chemical shift values for N-linked methyl groups (Figure 5A). These combined results confirm that the N-terminal Ala in andalusicin carries two methyl groups.

As for the labionin core, we exploited the combined NOESY/TOCSY spectra to identify characteristic NOE peaks. Starting from Gly in position 13, the full NMR assignment could be obtained on the basis of combined TOCSY and NOESY spectra (Table S3). Importantly, we observed a clear NOE contact between the Hα and Hβ protons of the residue in position 16 and the amide proton of residue in position 20. Together with the NOE contact between the α proton of the central residue in position 20 and the amide proton of residue 23 (Figure 5B), this confirms the constrained architecture of the labionin moiety. The physicochemical properties of andalusicin A are summarized in Figure 6.

Mutational analysis of the andalusicin precursor

To test substrate requirements of AncKC, we conducted a mutational analysis of *ancA1* using *B. subtilis* harboring plasmid-borne *ancKC-ancA1* genes. To check if a lanthionine ring can be formed instead of labionin, a mutation resulting in Ser20Ala substitution was introduced into *ancA1*. This substitution did not result in the MS/MS fragmentation pattern change indicating that the lanthionine ring was still formed in the mutant peptide (Figures S3A and S3B). Alanine substitution of Ser17 resulted in a short lanthionine ring, as evidenced by an MS/MS fragmentation pattern that lacked fragmentation peaks between Ser20 and Cys23 (Figures S3A and S3C). Unexpectedly, no mature products of the double S17A/S20A AncA1 mutant were detected, suggesting that Dhb18 or any other preceding Dhb cannot serve as an acceptor for the methyllanthionine cross-link. Taken together, our data indicate that AncKC can catalyze independent nucleophilic attack by AncA1 Cys23 on either one of the dehydrated serine residues, Dha17 or Dha20. To check if lanthionine-containing Ser20Ala mutant can be a substrate for AncMT, *B. subtilis* was transformed with a plasmid harboring the *ancKC-ancA1* (S20A)-*ancMT* operon. As expected, N-terminally dimethylated lanthionine-containing andalusicin was produced (Figure S3D). Unfortunately, the efficiency of production was not sufficient for antibacterial activity analysis.

BGCs encoding class III lanthipeptides in *Firmicutes*

Recent genome mining of prokaryotic genomes revealed that putative BGCs encoding class III lanthipeptide biosynthesis encompass about a quarter of the entire lanthipeptide landscape (Walker et al., 2020).

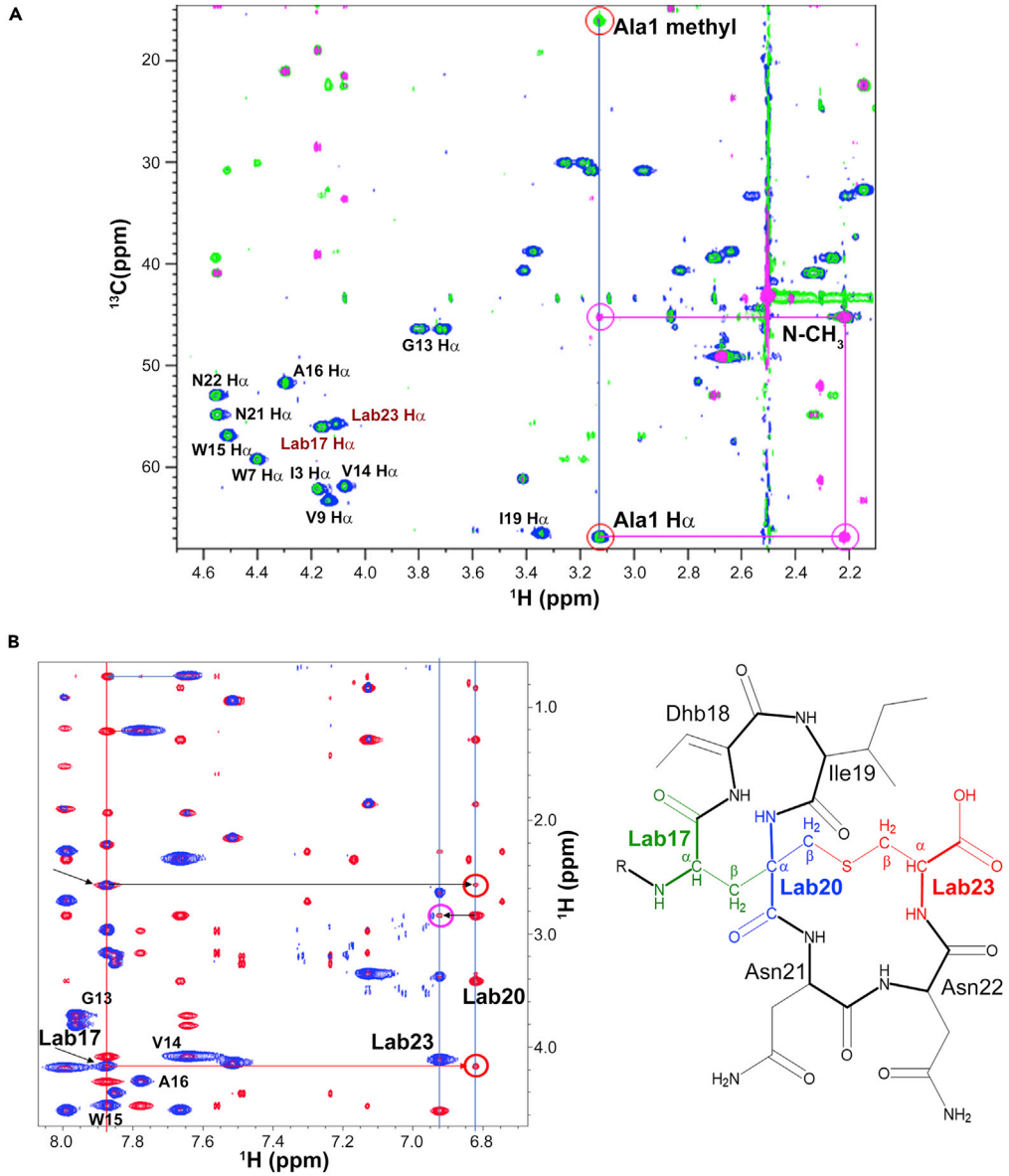


Figure 5. NMR spectra of andalusicin and the structure of C-terminal labionin cross-link

(A) 1H-13C HSQC (blue), HSQC-DIPSI (green), and HMBC (magenta) spectra of the andalusicin peptide in DMSO. The signals of the N-terminal Ala are circled in red, whereas the long-range connectivities with the N-terminal methyl resonance are circled in magenta.

(B) NOESY (red) and TOCSY (blue) spectra of the peptide in the fingerprint region. The distinct nuclear Overhauser effect (NOE) cross peaks for the residue in position 20 to both the residues in position 20 (red circles) and position 23 (magenta circle) confirm its central position in the labionin architecture.

Despite their abundance, class III lanthipeptides remain largely unexplored. As andalusicin is the first characterized labionin-containing lantibiotic from *Firmicutes*, we decided to get a closer look at its homologs. First, we built a sequence similarity network of the LanKC enzymes. As is shown in Figure 7A, at the BLAST score cutoff of 130, five major groups of class III LanKC were identified. Two groups (I and V) comprise solely actinobacterial sequences, groups II and IV contain predominantly firmicutal sequences, while group III has members from different phyla. Andalusicin AncKC belongs to group II. To further evaluate the diversity within this group, we built a phylogenetic tree of lanthionine cyclase domains (LanC), inspected proteins encoded in close proximity to *lanKC* genes, and analyzed ORFs encoding putative precursor peptides (Figure 7B and Table S4).

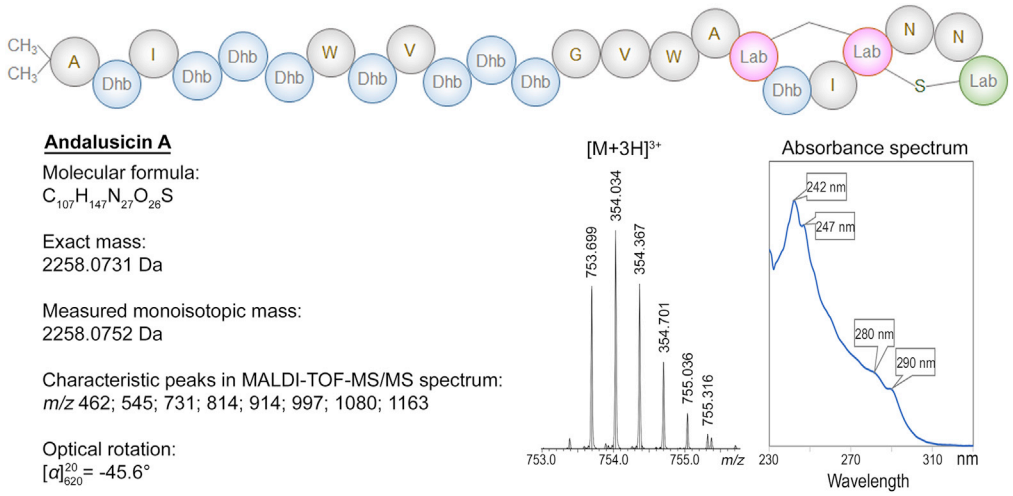


Figure 6. Properties and schematic structure of andalusicin A

The numbers of genes coding for predicted precursor peptides ranged from one to as many as 13, with a mean number of 2.8 and a median of 2. It is worth noting that if multiple genes for precursor peptides are present in a BGC, they encode peptides with identical or almost identical sequences. However, in some cases, predicted precursor peptides differ significantly (e.g., in *Curtobacterium* sp. MCBA15_003 or *Alkalihalobacillus marmarensis* DSM 21297 in Table S4), implying that some class III lanthipeptide BGCs may encode two-component lanthipeptides (McClerren et al., 2006).

We identified conserved motifs in predicted precursor sequences using MEME software (Bailey et al., 2009). As expected, the class III leader peptide motif VLXLQ (motif 1 in Figure 7C) and labionin motifs $S(X)_2S(X)_2C$ or $S(X)_2S(X)_3C$ were found (motifs 3 and 5, Figure 7C). Almost half of the precursors contained an extended tandem labionin motif (motif 2 in Figure 7C). We also identified an additional motif enriched in threonine residues (Figure 7C, motif 4) previously noted by (Walker et al., 2020). Mapping of predicted precursor peptide motif architectures to the LanC phylogenetic tree revealed, remarkably, that precursors with different motif architectures are associated with distinct clades. Precursors presumably containing two labionin macrocycles, either extended tandem motif 2 or various combinations of motifs 3 and 5, are associated with the basal clades of LanKC enzymes. A large group of precursors peptides contain the threonine-rich motif 4 followed by a single C-terminal labionin motif 3 or 5. These precursors are exclusively associated with a relatively recent LanKC clade (Figure 7B). As this clade contains AncKC characterized in this work, we will refer to it as "AncKC-like" and the peptides associated with this clade as "andalusicin-like."

To establish the set of genes involved in firmicutal class III lanthipeptide production, we analyzed genes frequently found in the genomic neighborhood of *lanKC* genes (Table S4). In actinobacterial class III lanthipeptides, the leader peptide removal can be performed either by a bifunctional endo- and aminopeptidase (Chen et al., 2019) or by unspecific housekeeping aminopeptidases (Krawczyk et al., 2012), which sometimes act in concert with the prolyl peptidase encoded in the lanthipeptide BGC (Völler et al., 2013). Genes encoding two types of peptidases, PtrB-like (cl34357) and PqqL-like (cl33975), colocalized, respectively, with group IV *lanKC* genes and the AncKC-like clade in group II (Figure 7B), making them good candidates for leader peptide removal.

Genes encoding ABC transporters responsible for export or self-immunity are often found in lanthipeptide BGCs (Alkhatib et al., 2012). Our analysis revealed that genes coding for MdlB-like ABC transporters colocalize with ca. 70% *lanKC* loci in *Firmicutes*. A pair of genes encoding the ABC-2 type permease and the catalytic CcmA-like ATPase subunit was typical to the AncKC clade (Figure 7B and Table S4). Further studies are required for elucidating their role in export or other functions.

The andalusicin bioactivity depends on the methylation of its N-terminal residue. We analyzed the co-occurrence of genes encoding the AncMT-like family methyltransferases (cl17173) and LanKC enzymes.

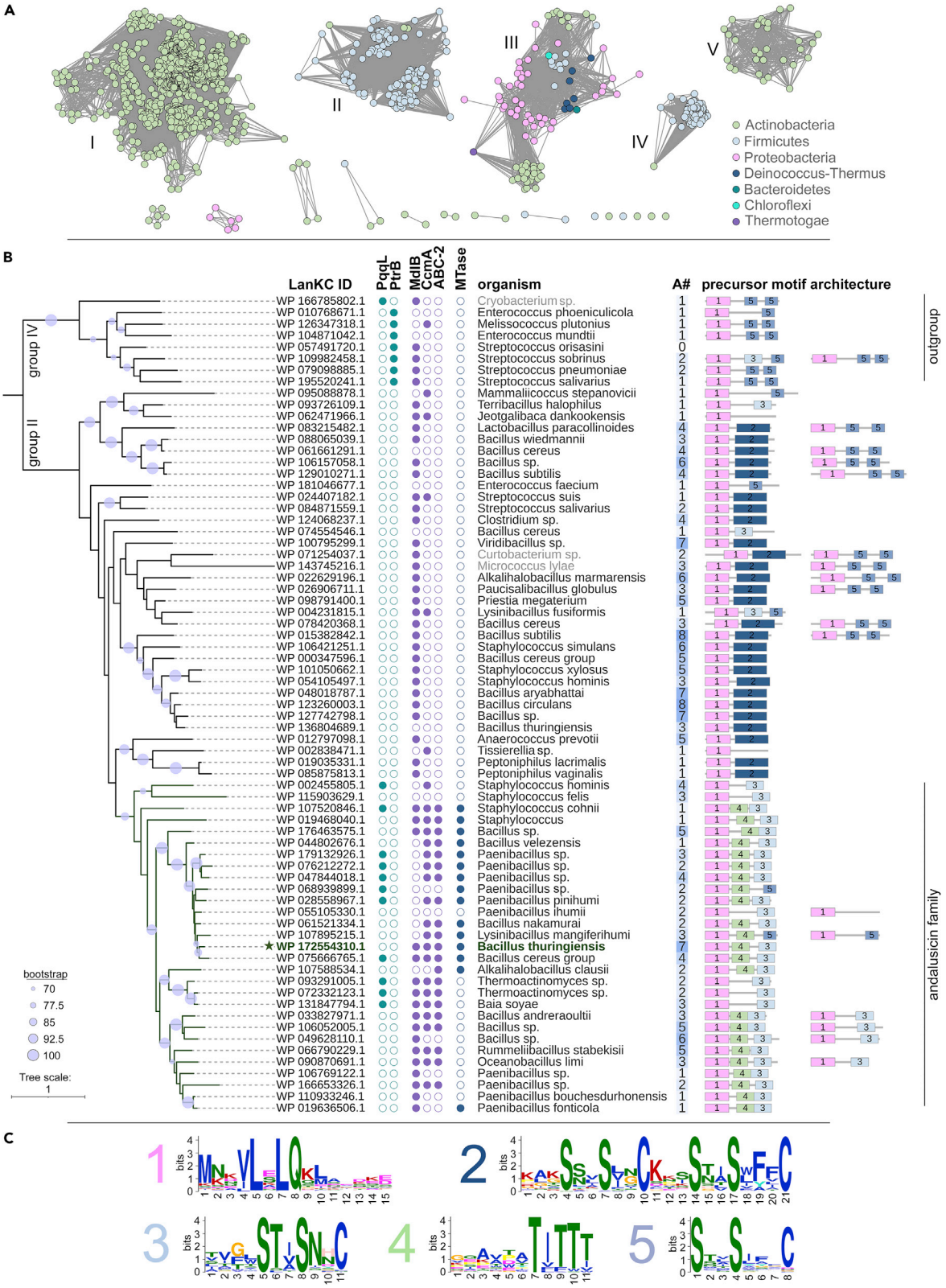

A
I
II
III
IV
V
Actinobacteria
Firmicutes
Proteobacteria
Deinococcus-Thermus
Bacteroidetes
Chloroflexi
Thermotogae
B
LanKC ID
PqqL
PtrB
MdlB
CcmA
ABC-2
MTase
organism
A#
precursor motif architecture
group IV
group II
outgroup
andalusicin family
bootstrap
70
77.5
85
92.5
100
Tree scale:
1
C
bits

Figure 7. Phylogenetic analysis of class III lanthionine synthetases from *Firmicutes*

(A) Sequence similarity network of class III LanKC enzymes generated using EFI-EST web tool with a strict similarity cutoff (alignment score 130 equivalent to an E-value of 10^{-130}) visualized in Cytoscape. Nodes are colored according to phyla. Groups with more than 10 nodes are numbered.

(B) A maximum likelihood phylogenetic tree of LanC domains from groups II and IV. Sequences of LanC domains from group IV were added to serve as an outgroup for tree rooting. Bootstrap values are shown with light blue circles on corresponding branches. The LanKC clade associated with andalusicin-like peptides is marked by a dark green color. Protein ID for andalusicin AncKC is marked with a star. Colocalization of *lanKC* with genes coding for specified peptidases, transporters, and methyltransferase is indicated by closed circles. Genus names of LanKCs from the *Actinomyces* are shown in gray. The sequence motif architecture variants of predicted precursor peptides are schematically represented. Motifs are numbered as in (C). The number of predicted precursor peptide genes found in corresponding BGCs is shown in column A#.

(C) Motifs found in precursor peptides associated with *lanKC* genes from groups II and IV from (A) are represented as sequence logos. Motifs are ordered according to the E-value.

As is shown in Figure 7B, AncMT-like MTases are associated solely with the AncKC clade, which allows us to speculate that at least some of the lanthipeptides encoded in these BGCs possess methylation-dependent antimicrobial activity.

DISCUSSION

Recent systematic genome mining identified numerous class III lanthipeptide BGCs in *Firmicutes* (Walker et al., 2020), although none of these compounds have been isolated to date. In this work, we characterized andalusicin, a novel class III lanthipeptide from *B. thuringiensis* sv. *andalousiensis* NRRL B23139. Andalusicin has a pastor's crook-like structure: it contains a long linear N-terminal part with a stretch of dehydrated threonine residues and a single C-terminal macrocycle. It makes andalusicin structurally more similar to certain linaridins (Minami et al., 1994; Shang et al., 2019) than to other known class III lanthipeptides (Iorio et al., 2014; Jungmann et al., 2016). The andalusicin precursor peptides have a characteristic $SX_2SX_{2\text{-}3}C$ motif typical for labionin-containing lanthipeptides (Hegemann and Süssmuth, 2020). However, the presence of such a motif is not sufficient for labionin formation as some LanKC enzymes predominantly generate lanthionine crosslink with a dehydrated central serine residue (Krawczyk et al., 2012). Mutational analysis confirms that AncKC can install both large and small lanthionine crosslinks. However, the major product of the *anc* BGC, andalusicin A, contains the labionin macrocycle. Yet another unique feature of andalusicin A is dimethylation of its N-terminal residue. Similar to the linear azoline-containing peptide plantazolicin (Molohon et al., 2011) and the linaridin cypemicin (Claesen and Bibb, 2010), N-terminal dimethylation is required for andalusicin bioactivity. Andalusicin A without this tailoring modification showed weak or no antimicrobial activity, while the dimethylated form inhibited the growth of various Gram-positive bacteria, including *S. aureus*. Our preliminary data indicate that andalusicin A causes severe damage to cellular membrane of the sensitive *B. cereus* cells, however, the exact mechanism of its antimicrobial activity remains to be determined.

We identified the gene cluster responsible for biosynthesis of andalusicin and determined the minimal set of genes required for its production. The andalusicin BGC comprises genes coding for a LanKC-like class III lanthionine synthetase AncKC, a methyltransferase AncMT, and seven genes encoding three nearly identical precursor peptides. Expression of three genes, *ancKC*, *ancMT*, and *ancA1* is sufficient for andalusicin A synthesis in a heterologous host. Two genes encoding a permease (EVG22_16,520) and an ATP-binding protein (EVG22_16,515) that may form a putative two-component ABC transporter, are located upstream of *ancKC*. Another gene, encoding a putative MdlB-like ABC transporter (EVG22_16,530), is located downstream of the *ancMT* gene. Phylogenetic analysis shows that genes coding for both kinds of transporters are frequently associated with andalusicin-like BGCs (Figure 7B). Analogously to nisin BGC (Siegers and Entian, 1995), one of the transporters may be responsible for lanthipeptide secretion, while another can serve as an immunity determinant. Unfortunately, we were unable to establish whether EVG22_16,515, EVG22_16,520, and EVG22_16,530 genes are part of the *anc* cluster.

Similar to most other studied class III lanthipeptides BGCs, a gene required for leader peptide processing is absent from andalusicin BGC. Since heterologous production of andalusicin in *B. subtilis* resulted in ineffective (albeit correct) processing of the leader peptide, we speculate that the leader can be removed by a dedicated peptidase(s) encoded elsewhere in the genome of the producer strain (Chen et al., 2019). Our bioinformatics analysis revealed that genes encoding PqqL-like peptidases are frequently co-localized with andalusicin-like clusters (Figure 7B). As a number of peptidases belonging to this family are encoded in the genome of *B. thuringiensis* sv. *andalousiensis* NRRL B23139, at least one of them may be responsible for the cleavage of the leader part of andalusicin.

Phylogenetic analysis shows that methyltransferases are not common in class III lanthipeptide BGCs from *Firmicutes*. Indeed, the acquisition of a methyltransferase gene appeared to be a single event that happened relatively early in the evolution of andalusicin-like lanthipeptides. It is worth noting that the methyltransferase gene was lost in two daughter clades of andalusicin-like LanKCs. Further studies will show if products of clusters from these clades are functional and whether N-terminal methylation can impact on the biological activity of these compounds.

Taken together, andalusicin is the first representative of a large group of firmicutal class III lanthipeptides that possess a unique structure and a secondary modification rarely found in lanthipeptides. The N-terminal methylation required for andalusicin bioactivity, if installed on other lantibiotics, may stimulate their activity and thus can find practical application.

Limitations of the study

We characterized andalusicin, the first labionin-containing class III lantibiotic from *Firmicutes* and identified its biosynthetic gene cluster. Andalusicin contains two methyl groups on its N terminus, a modification previously unknown for this class of lanthipeptides. Under standard laboratory cultivation conditions, the production level of andalusicin A was relatively low. This precluded further physicochemical analyses of this compound. Importantly, at this stage, we were unable to conduct crystallization experiments to establish the absolute conformation of the compound. Phylogenetic analysis of LanKC encoded in prokaryotic genomes showed that andalusicin is a representative of a large group of structurally similar lanthipeptides. Genes likely responsible for transport, leader peptide removal, and immunity were predicted, but these predictions remain to be experimentally verified. Andalusicin is bioactive against *B. cereus* and closely related species. Further investigations will be required to uncover the exact mechanism of andalusicin antibacterial activity and the role of N-terminal dimethylation in lanthipeptide stability, affinity to the target, or cellular permeability.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Bacterial cultures
- METHOD DETAILS
  - Plasmid construction
  - Peptide products expression and purification
  - AncMT activity
  - Andalusicin antimicrobial activity
  - Microscopy
  - Chemical modifications
  - Physicochemical methods
  - Whole-genome DNA sequencing, genome assembly, and annotation
  - Bioinformatic analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102480.

ACKNOWLEDGMENTS

We are grateful to Dmitry Travin (Skoltech, Russia) and Dmitry Ghilarov (Jagiellonian University, Poland) for the critical reading of the manuscript. We thank Tatyana Smirnova and the Center for Precision Genome Editing and Genetic Technologies for Biomedicine, IGB RAS (Moscow, Russia), for microscopy. We thank Dr. Viktor Zgoda and "Human Proteome" Core Facilities of the Institute of Biomedical Chemistry (Moscow,

Russia) for high-resolution mass spectrometry. We are grateful to Prof. V. Minaev and the Shared Use Equipment Center for high-precision measuring in photonics (ckp.vniiofi.ru, VNIIOFI, Moscow, Russia) for help with optical rotation measurement. This work was supported by the Russian Foundation of Basic Research grant RFBR 19-04-01163 to J.A., NIH RO1 grant AI117270 (S.K.N. and K.S.), and GM079038 (S.K.N.). Mutational analysis of andalusicin was supported by Russian Science Foundation grant RSF 19-14-00266 to S.D. MALDI-TOF-MS facility became available to us in the framework of the Moscow State University Development Program PNG 5.13. The NMR facility is funded through MetaToul (Toulouse metabolomics & fluxomics facilities, www.metatoul.fr) as part of the French National Infrastructure for Metabolomics and Fluxomics MetaboHUB-AR-11-INBS-0010 (www.metabohub.fr). It receives support from the Région Midi-Pyrénées, the ERDF, the SICOVAL, and the French Minister of Education & Research, who are all gratefully acknowledged.

AUTHOR CONTRIBUTIONS

Conceptualization, S.D.; methodology, S.K.N., M.S., and S.D.; investigation, A.G., A.R., J.A., D.B., M.S., D.S., A.H.G., and G.L.; writing – original draft, S.D.; writing – review & editing, S.K.N. and K.S.; resources, M.S. and G.L.; supervision, K.S. and S.D.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 24, 2020
Revised: March 21, 2021
Accepted: April 23, 2021
Published: May 21, 2021

REFERENCES

Acedo, J.Z., Bothwell, I.R., An, L., Trouth, A., Frazier, C., and van der Donk, W.A. (2019). O-Methyltransferase-Mediated incorporation of a β-amino acid in lanthipeptides. J. Am. Chem. Soc. *141*, 16790–16801.

Alkhatib, Z., Abts, A., Mavaro, A., Schmitt, L., and Smits, S.H.J. (2012). Lantibiotics: how do producers become self-protected? J. Biotechnol. *159*, 145–154.

Arnison, P.G., Bibb, M.J., Bierbaum, G., Bowers, A.A., Bugni, T.S., Bulaj, G., Camarero, J.A., Campopiano, D.J., Challis, G.L., Clardy, J., et al. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat. Prod. Rep. *30*, 108–160.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). Meme SUITE: tools for motif discovery and searching. Nucleic Acids Res. *37*, W202–W208.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. *19*, 455–477.

Boulos, L., Prévost, M., Barbeau, B., Coallier, J., and Desjardins, R. (1999). LIVE/DEAD BacLight : application of a new rapid staining method for direct enumeration of viable and total bacteria in drinking water. J. Microbiol. Methods *37*, 77–86.

Chen, S., Xu, B., Chen, E., Wang, J., Lu, J., Donadio, S., Ge, H., and Wang, H. (2019). Zn-dependent bifunctional proteases are responsible for leader peptide processing of class III lanthipeptides. Proc. Natl. Acad. Sci. U S A *116*, 2533–2538.

Claesen, J., and Bibb, M. (2010). Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of posttranslationally modified peptides. Proc. Natl. Acad. Sci. U S A *107*, 16297–16302.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

Gerlt, J.A., Bouvier, J.T., Davidson, D.B., Imker, H.J., Sadkhin, B., Slater, D.R., and Whalen, K.L. (2015). Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): a web tool for generating protein sequence similarity networks. Biochim. Biophys. Acta *1854*, 1019–1037.

Hegemann, J.D., Shi, L., Gross, M.L., and van der Donk, W.A. (2019). Mechanistic studies of the kinase domains of class IV lanthipeptide synthetases. ACS Chem. Biol. *14*, 1583–1592.

Hegemann, J.D., and Süssmuth, R.D. (2020). Matters of class: coming of age of class III and IV lanthipeptides. RSC Chem. Biol. *1*, 110–127.

Hsu, S.-T.D., Breukink, E., Tischenko, E., Lutters, M.A.G., de Kruijff, B., Kaptein, R., Bonvin, A.M.J.J., and van Nuland, N.A.J. (2004). The nisin-lipid II complex reveals a pyrophosphate cage that provides a blueprint for novel antibiotics. Nat. Struct. Mol. Biol. *11*, 963–967.

Iorio, M., Sasso, O., Maffioli, S.I., Bertorelli, R., Monciardini, P., Sosio, M., Bonezzi, F., Summa, M., Brunati, C., Bordoni, R., et al. (2014). A glycosylated, labionin-containing lanthipeptide with marked antinociceptive activity. ACS Chem. Biol. *9*, 398–404.

Jungmann, N.A., van Herwerden, E.F., Hügelland, M., and Süssmuth, R.D. (2016). The supersized class III lanthipeptide stackepeptin displays motif multiplication in the core peptide. ACS Chem. Biol. *11*, 69–76.

Knerr, P.J., and van der Donk, W.A. (2012). Discovery, biosynthesis, and engineering of lantipeptides. Annu. Rev. Biochem. *81*, 479–505.

Krawczyk, B., Völler, G.H., Völler, J., Ensle, P., and Süssmuth, R.D. (2012). Curvopeptin: a new lanthionine-containing class III lantibiotic and its Co-substrate promiscuous synthetase. ChemBioChem *13*, 2065–2071.

Lagedroste, M., Smits, S.H.J., and Schmitt, L. (2017). Substrate specificity of the secreted nisin leader peptidase NisP. Biochemistry *56*, 4005–4014.

Letunic, I., and Bork, P. (2019). Interactive Tree of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. *47*, W256–W259.

Li, Q., Xu, L.Z., Zou, T., Ai, P., Huang, G.H., Li, P., and Zheng, A.P. (2015). Complete genome sequence of Bacillus thuringiensis strain HD521. Stand. Genomic Sci. *10*, 62.

Lohans, C.T., and Vederas, J.C. (2014). Structural characterization of thioether-bridged bacteriocins. J. Antibiot. (Tokyo). *67*, 23–30.

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M., Hurwitz,

D.I., Marchler, G.H., Song, J.S., et al. (2020). CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res. *48*, D265–D268.

Martin, N.I., Sprules, T., Carpenter, M.R., Cotter, P.D., Hill, C., Ross, R.P., and Vederas, J.C. (2004). Structural characterization of lacticin 3147, a two-peptide lantibiotic with synergistic activity. Biochemistry *43*, 3049–3056.

McClerren, A.L., Cooper, L.E., Quan, C., Thomas, P.M., Kelleher, N.L., and van der Donk, W.A. (2006). Discovery and in vitro biosynthesis of haloduracin, a two-component lantibiotic. Proc. Natl. Acad. Sci. U S A *103*, 17243–17248.

Meindl, K., Schmiederer, T., Schneider, K., Reicke, A., Butz, D., Keller, S., Gühring, H., Vértesy, L., Wink, J., Hoffmann, H., et al. (2010). Labyrinthopeptins: a new class of carbacyclic lantibiotics. Angew. Chem. Int. Ed. *49*, 1151–1154.

Meyer, H.E., Heber, M., Eisermann, B., Korte, H., Metzger, J.W., and Jung, G. (1994). Sequence analysis of lantibiotics: chemical derivatization procedures allow a fast access to complete edman degradation. Anal. Biochem. *223*, 185–190.

Minami, Y., Yoshida, K., Azuma, R., Urakawa, A., Kawauchi, T., Otani, T., Komiyama, K., and Ōmura, S. (1994). Structure of cypemycin, a new peptide antibiotic. Tetrahedron Lett. *35*, 8001–8004.

Molohon, K.J., Melby, J.O., Lee, J., Evans, B.S., Dunbar, K.L., Bumpus, S.B., Kelleher, N.L., and Mitchell, D.A. (2011). Structure determination and interception of biosynthetic intermediates for the plantazolicin class of highly discriminating antibiotics. ACS Chem. Biol. *6*, 1307–1313.

Montalbán-López, M., Scott, T.A., Ramesh, S., Rahman, I.R., van Heel, A.J., Viel, J.H., Bandarian, V., Dittmann, E., Genilloud, O., Goto, Y., et al. (2021). New developments in RiPP discovery, enzymology and engineering. Nat. Prod. Rep. *38*, 130–239.

Müller, W.M., Ensle, P., Krawczyk, B., Süssmuth, R.D., Müller, W.M., Ensle, P., Krawczyk, B., and Süssmuth, R.D. (2011). Leader peptide-directed processing of labyrinthopeptin A2 precursor peptide by the modifying enzyme LabKC. Biochemistry *50*, 8362–8373.

Oman, T.J., and van der Donk, W.A. (2010). Follow the leader: the use of leader peptides to guide natural product biosynthesis. Nat. Chem. Biol. *6*, 9–18.

Ortega, M.A., Cogan, D.P., Mukherjee, S., Garg, N., Li, B., Thibodeaux, G.N., Maffioli, S.I., Donadio, S., Sosio, M., Escano, J., et al. (2017). Two flavoenzymes catalyze the post-translational generation of 5-chlorotryptophan and 2-aminovinyl-cysteine during NAI-107 biosynthesis. ACS Chem. Biol. *12*, 548–557.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., et al. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res. *42*, D206–D214.

Pol, I.E., and Smid, E.J. (1999). Combined action of nisin and carvacrol on Bacillus cereus and Listeria monocytogenes. Lett. Appl. Microbiol. *29*, 166–170.

Prince, A., Sandhu, P., Ror, P., Dash, E., Sharma, S., Arakha, M., Jha, S., Akhter, Y., and Saleem, M. (2016). Lipid-II independent antimicrobial mechanism of nisin depends on its crowding and degree of oligomerization. Sci. Rep. *6*, 37908.

Repka, L.M., Chekan, J.R., Nair, S.K., and van der Donk, W.A. (2017). Mechanistic understanding of lanthipeptide biosynthetic enzymes. Chem. Rev. *117*, 5457–5520.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Schneider, T.R., Kärcher, J., Pohl, E., Lubini, P., and Sheldrick, G.M. (2000). Ab initio structure determination of the lantibiotic mersacidin. Acta Crystallogr. D. Biol. Crystallogr. *56*, 705–713.

Seabold, S., and Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with python, 9th Python Sci. Conf. https://www.statsmodels.org/stable/index.html.

Shang, Z., Winter, J.M., Kauffman, C.A., Yang, I., and Fenical, W. (2019). Salinipeptins: integrated genomic and chemical approaches reveal unusual d-amino acid-containing ribosomally synthesized and post-translationally modified peptides (RiPPs) from a great salt lake streptomyces sp. ACS Chem. Biol. *14*, 415–425.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Siegers, K., and Entian, K.D. (1995). Genes involved in immunity to the lantibiotic nisin produced by Lactococcus lactis 6F3. Appl. Environ. Microbiol. *61*, 1082–1089.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

Steenwyk, J.L., Buida, T.J., Li, Y., Shen, X.-X., and Rokas, A. (2020). ClipKIT: a multiple sequence alignment trimming software for accurate phylogenomic inference. PLoS Biol. *18*, e3001007.

Stein, T., Heinzmann, S., Solovieva, I., and Entian, K.-D. (2003). Function of Lactococcus lactis nisin immunity genes nisI and nisFEG after coordinated expression in the surrogate host Bacillus subtilis. J. Biol. Chem. *278*, 89–94.

Steinegger, M., and Söding, J. (2017). MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat. Biotechnol. *35*, 1026–1028.

Theillet, F.-X., Smet-Nocca, C., Liokatis, S., Thongwichian, R., Kosten, J., Yoon, M.-K., Kriwacki, R.W., Landrieu, I., Lippens, G., and Selenko, P. (2012). Cell signaling, post-translational protein modifications and NMR spectroscopy. J. Biomol. NMR *54*, 217–236.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272.

Völler, G.H., Krawczyk, B., Ensle, P., and Süssmuth, R.D. (2013). Involvement and unusual substrate specificity of a prolyl oligopeptidase in class III lanthipeptide maturation. J. Am. Chem. Soc. *135*, 7426–7429.

Walker, M.C., Eslami, S.M., Hetrick, K.J., Ackenhusen, S.E., Mitchell, D.A., and van der Donk, W.A. (2020). Precursor peptide-targeted mining of more than one hundred thousand genomes expands the lanthipeptide natural product family. BMC Genomics *21*, 387.

Wang, H., and van der Donk, W.A. (2012). Biosynthesis of the class III lantipeptide catenulipeptin. ACS Chem. Biol. *7*, 1529–1535.

Wiebach, V., Mainz, A., Siegert, M.-A.J., Jungmann, N.A., Lesquame, G., Tirat, S., Dreux-Zigha, A., Aszodi, J., Le Beller, D., and Süssmuth, R.D. (2018). The anti-staphylococcal lipolanthines are ribosomally synthesized lipopeptides. Nat. Chem. Biol. *14*, 652–654.

Wiedemann, I., Benz, R., and Sahl, H.-G. (2004). Lipid II-mediated pore formation by the peptide antibiotic nisin: a black lipid membrane study. J. Bacteriol. *186*, 3259–3261.

Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S., and Sykes, B.D. (1995). 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. J. Biomol. NMR *5*, 67–81.

Xu, M., Zhang, F., Cheng, Z., Bashiri, G., Wang, J., Hong, J., Wang, Y., Xu, L., Chen, X., Huang, S.-X., et al. (2020). Functional genome mining reveals a class V lanthipeptide containing a d-amino acid introduced by an F420 H2 -dependent reductase. Angew. Chem. Int. Ed. *59*, 18029–18035.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Bacterial strains are listed in Table S5		N/A
Deposited data		
Whole genome sequence of *Bacillus thuringiensis* sv. *andalousiensis* NRRL B23193	NCBI	GenBank: NZ_CP035727.2
Annotated andalusicin biosynthetic gene cluster	MIBiG	MIBiG: BGC0002111
Experimental models: organisms/strains		
Bacillus thuringiensis sv. *andalousiensis* NRRL B23193	NRRL	NRRL B-23139
Oligonucleotides		
For oligonucleotides, see Table S6		N/A
Recombinant DNA		
pHT01 vector	MoBiTech, GmbH	cat# PBS001
pET_His6_TEV_LIC	Scott Gradia, Addgene	RRID:Addgene_29653
pHT-ancKCA1	This study	N/A
pHT-ancKCA1(S17A)	This study	N/A
pHT-ancKCA1(S20A)	This study	N/A
pHT-ancKCA1(S17A-S20A)	This study	N/A
pHT-ancKCA1MT	This study	N/A
pHT-ancKCA1(S20A)MT	This study	N/A
pET_His6_ancMT	This study	N/A
Software and algorithms		
ImageJ	(Schneider et al., 2012)	https://imagej.nih.gov/ij/
SPAdes v. 3.9.1	(Bankevich et al., 2012)	https://cab.spbu.ru/software/spades/
RAST	(Overbeek et al., 2014)	https://rast.nmpdr.org/rast.cgi
MEME Suite v. 5.0.2	(Bailey et al., 2009)	https://meme-suite.org/meme/tools/meme
EFI-EST	(Gerlt et al., 2015)	https://efi.igb.illinois.edu/efi-est/
Cytoscape	(Shannon et al., 2003)	https://cytoscape.org/
MMseqs2 version 12.113e3	(Steinegger and Söding, 2017)	https://github.com/soedinglab/MMseqs2
MUSCLE v3.8.31	(Edgar, 2004)	http://www.drive5.com/muscle/
ClipKIT v1.0.7	(Steenwyk et al., 2020)	https://github.com/JLSteenwyk/ClipKIT
RaxML	(Stamatakis, 2014)	https://cme.h-its.org/exelixis/web/software/raxml/
Batch CD-Search	(Lu et al., 2020)	https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi
iTOL	(Letunic and Bork, 2019)	https://itol.embl.de/
SciPy v.1.3.1	(Virtanen et al., 2020)	https://www.scipy.org/
statsmodels v. 0.10.1	(Seabold and Perktold, 2010)	https://www.statsmodels.org/stable/index.html
custom scripts for bioinformatic analysis	This study	https://github.com/bikdm12/andalusicin

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to the lead contact, Konstantin Severinov (severik@waksman.rutgers.edu).

Materials availability

Plasmids generated in this study will be made available by the lead contact upon reasonable request.

Data and code availability

The accession number for the assembled genome sequence of *B. thuringiensis* sv. *andalousiensis* B23139 reported in this paper is GenBank: NZ_CP035727. The accession number for the annotated andalusicin biosynthetic gene cluster is MIBiG: https://mibig.secondarymetabolites.org BGC0002111.

Table S4 includes BGCs, CDD IDs, and manually predicted precursor peptides analyzed in this study. Data used for the sequence similarity network construction was retrieved from Walker et al., 2020). Custom Python scripts used for bioinformatic and statistical analyses are available at https://github.com/bikdm12/andalusicin.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bacterial cultures

The bacterial strains used in this study are listed in the Table S5.

B. thuringiensis sv. *andalousiensis* NRRL B23139 was grown aerobically in 2xYT medium (1.6% tryptone, 1% yeast extract, 0.5% NaCl, pH 7.0) at 30°C for 16 h. Aliquots (150 μl) were spread onto the 2xYT agar plates and incubated at 37°C for 3 days. Cells were collected and used for andalusicin extraction. *B. subtilis* 168 was used as a heterologous host for the production of andalusicin and its mutant forms. *B. subtilis* cells were transformed with pHT01 vector and its derivatives and grown on the 2xYT agar plates supplemented with 10 μg/mL chloramphenicol at 30°C.

METHOD DETAILS

Plasmid construction

The sequences of oligonucleotides used in this study are listed in the Table S6.

The *B. thuringiensis* sv. *andalousiensis* NRRL B23139 genomic DNA was used as a template for PCR. Phusion DNA polymerase, T4 DNA ligase, and restriction endonucleases were obtained from ThermoFisher Scientific, USA. Custom oligonucleotide synthesis was performed by Evrogen, Russia.

A sequence containing EagI-PstI restriction sites was introduced into pHT01 vector using the whole-plasmid PCR with pHT_FPst and pHT_REagPst primers, followed by digestion with PstI endonuclease and self-ligation of the resulting PCR product. For the heterologous expression of *ancKC-ancA1* genes, the corresponding DNA fragment was amplified with ancKC_BamHI_F and ancA1_EagI_R primers. For introduction of mutations in *ancA1*, PCR was performed with the same forward primer and one of ancA1_S17A_EagI_R (S17A substitution), ancA1_S20A_EagI_R (S20A substitution), or ancA1_S17A-S20A_EagI_R (for double S17A and S20A substitution) reverse primers. The PCR products were digested with BamHI and EagI restriction endonucleases and inserted into the pHT01-NotI vector linearized with the same endonucleases. To complement the minimal cluster with the methyltransferase gene *ancMT*, its coding sequence was amplified from genomic DNA with Pre_ancMT_EagI_F and ancMT_PstI_R primers; the PCR product was digested with EagI and PstI restriction endonucleases. The ancMT PCR was combined with the BamHI and EagI digested *ancKC-ancA1* or *ancKC-ancA1*_S20A PCR product, the pHT01-NotI vector linearized with BamHI and PstI, and ligated and transformed into *E. coli* DH5α cells.

For the recombinant AncMT methyltransferase production, its coding sequence was amplified from the genomic DNA with ancMT_His6_LIC_F and ancMT_His6_LIC_R primers; pET_His6_TEV_LIC vector was linearized with SspI restriction enzyme. Digested PCR product and pET_His6_TEV_LIC vector were assembled with Gibson Assembly Master Mix (NEB) to create a construction encoding N-terminal fusion protein with a hexahistidine (His6) tag. Ligation mixtures were transformed into *E. coli* DH5α cells.

Plasmids were purified using GeneJET Plasmid Miniprep Kit (ThermoFisher Scientific, USA) and transformed into appropriate bacterial strains for expression. Vectors used in the study are listed in the Table S7.

Peptide products expression and purification

B. thuringiensis cells (~7 g wet weight) were harvested from the agar plates, resuspended in 50 mL of 70% (v/v) methanol, and incubated at 37°C for 3 hours with occasional shaking. Cells were removed by centrifugation (3,200 x *g*, 4°C for 20 min), the supernatant was collected and evaporated in a centrifugal vacuum concentrator (Genevac EZ-2, SP Scientific, USA). The dried cellular extract was dissolved in 30 mL of buffer A (0.1% trifluoroacetic acid (TFA)/30% acetonitrile), briefly centrifuged to remove the insoluble material, and applied to a C18 gravity column (Mega Bound Elut-C18 10 g, Agilent Technologies, USA). The column was washed with 50 mL of buffer A, followed by elution with 25 mL of buffer B (0.1% TFA in 50% acetonitrile). Eluted fractions were dried and dissolved in 2 mL of 40% acetonitrile and subjected to further purification using HPLC.

For andalusicin production in the heterologous host, the overnight culture of *B. subtilis* 168 harboring the relevant plasmid was diluted 1:100 in 500 mL of 2xYT medium supplemented with 10 μg/mL chloramphenicol. Cells were grown at 37°C, 180 rpm until the optical density at 600 nm reached 0.7 – 0.8. Expression was induced by adding isopropyl-β-d-1-thiogalactopyranoside (IPTG) to the final concentration 0.5 mM; incubation continued at 18°C for 16 h. Cells were pelleted by centrifugation at 3,200 x *g* for 20 min. The supernatant was collected, combined with 200 mL of 0.1% TFA in acetonitrile, and applied to the C18 gravity column (Mega Bound Elut-C18 1 g, Agilent Technologies, USA). The column was washed with 30 mL of buffer A, eluted with 10 mL of buffer B. Eluted andalusicin was further purified using HPLC.

HPLC purification of peptide compounds was performed using a Zorbax Eclipse Plus C18 column (5-μm, 4.6 x 250 mm, Agilent Technologies, USA), connected to 1220 Infinity II LC system (Agilent Technologies, USA). A linear gradient of 45 to 60% acetonitrile in 0.1% TFA was applied for 30 min. Chromatographic peaks were detected by the absorption at 280 nm. Fractions were collected and tested for biological activity in the agar diffusion test as described below. Fractions containing andalusicin or its forms were additionally purified using the same column and linear gradient of acetonitrile (40-60%) in 0.1 M triethylamine acetate buffer (pH 6.0) for 30 min. Purified compounds were dried and stored at -20°C.

AncMT activity

Recombinant AncA was produced in *E. coli* BL21(DE3) transformed with the pET_His6_TEV_LIC plasmid. The cells were grown in 200 mL of 2xYT medium supplemented with 100 μg/mL ampicillin to OD600 = 0.5. Protein expression was induced with 0.2 mM IPTG, and incubation was continued at 18°C, 180 rpm for 16 h. The cells were harvested by centrifugation at 8,000 x *g*, resuspended in an ice-cold wash buffer (20 mM Tris-HCl, 500 mM NaCl, 10 mM $MgCl_2$, 5 mM imidazole, pH 8.0), supplemented with 1 mM phenylmethylsulfonyl fluoride (PMSF), and disrupted by sonication. The lysate was cleared by centrifugation at 20,000 x *g* at 4°C for 15 min and applied to a pre-equilibrated Talon CellThru Co^{2+} resin (Clontech-Takara Biotechnology, USA). The resin was washed with 10-bed volumes of the wash buffer, followed by elution with 5-bed volumes of the elution buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM $MgCl_2$, 0.5 M imidazole, 10% glycerol). Eluted protein fractions were flash-frozen in liquid nitrogen and stored at -80°C. The purity of AncMT was determined by SDS-PAGE; the identity of the protein was confirmed by tryptic peptide fingerprint using MALDI-TOF-MS.

In vitro methylation assay was conducted in a reaction buffer (50 mM Tris-HCl, pH 7.0) containing 3 mM S-adenosylmethionine (SAM), 5 μM AncMT, and 5 μM desmethyl andalusicin A at 37°C for 3 h. Prior to mass spectrometry analysis, reaction mixtures were desalted using P10 ZipTip C18 (Millipore, Ireland) according to the manufacturer's protocol.

Andalusicin antimicrobial activity

The minimal inhibitory concentration of andalusicin A against *B. cereus* ATCC 4342 was measured in a broth dilution test. Andalusicin was serially diluted in DMSO and added to the wells containing 200 μl of LB media inoculated with 10^5 cells of *B. cereus* ATCC 4342. The final concentration of DMSO in the wells was 3%. Plates were incubated at 37°C with moderate shaking (30 rpm) for 16 hours.

The antimicrobial activity of andalusicin on a panel of bacterial strains was determined in agar diffusion assay. Soft LB agar plates were prepared by seeding of 20 mL of molten 0.5% LB agar with 50 μl of overnight bacterial culture; 3 μl-drops of 2.5 mg/mL solutions of andalusicin A and its desmethyl form were deposited on the surface of the solidified seeded soft agar. Size of growth inhibition zone was evaluated after 24 h of

incubations at 30°C, ranged and converted into "+", "++", "+++", where "+" corresponds to the smallest zone, low inhibitory activity and "+++" corresponds to the largest zone and strongest inhibitory activity.

All measurements were performed in triplicates.

Microscopy

The overnight culture of *B. cereus* ATCC 4342 was diluted 1:50 in fresh LB medium and allowed to grow at 28°C with vigorous shaking to OD600 = 0.4. 100-μL aliquots of the cells were combined with the corresponding fluorescent dyes and 1.5 μL of either inhibitor (1.34 μg/mL nisin in 100% DMSO; 4.36 μg/mL andalusicin A in 100% DMSO; 33 mM carbonyl cyanide 3-chlorophenylhydrazone (CCCP) in 100% DMSO) or 100% DMSO as a negative control. The fluorescent dyes were used at the final concentration of 1 μg/ml for 4′,6-diamidino-2-phenylindole (DAPI) and 30 μM for Propidium Iodide (PI). After 30 min of incubation at 28°C, 1-μL aliquots of the treated cells were applied to agarose pads (1.2% agarose in PBS) and imaged using Nikon Eclipse Ti microscope equipped with the Nikon Plan Apo VC 100X1.40 oil objective and Nikon DS-Qi2 digital monochrome camera. Images were processed using ImageJ software (Schneider et al., 2012).

Chemical modifications

Chemical modification of the Dha/Dhb residues in andalusicin with β-mercaptoethanol was performed according to the procedure described in (Meyer et al., 1994). Lanthipeptide (0.2 mg) was dissolved in the reaction mixture consisting of 280 μl ethanol, 200 μl water, 65 μl 5 M NaOH, and 60 μl β-mercaptoethanol and incubated at 50°C for 2 h. After the incubation completion, the reaction mixture was concentrated by rotary evaporation to 1/10 of the initial volume, desalted using P10 ZipTip C18 (Millipore, Ireland) according to manufacturer's protocol, and subjected to MALDI-TOF-MS analysis.

Andalusicin treatment with Ni_2B was carried out as described in (Martin et al., 2004). Andalusicin A (0.2 mg) and $NiCl_2$ (1 mg) were dissolved in 2 mL of 50% (v/v) methanol and combined with NaBH4 (1.0 mg). The mixture was incubated at 50°C for 1 h with occasional vortexing and then centrifuged to remove the Ni_2B precipitate. The supernatant was transferred into a fresh tube; the residual peptide was additionally extracted from the Ni_2B precipitate by adding 50% (v/v) methanol. The supernatants were combined, dried in a centrifugal vacuum concentrator (Genevac EZ-2, SP Scientific, USA). For MALDI-TOF-MS analysis, the lanthipeptide was dissolved in a 50% (v/v) acetonitrile.

Physicochemical methods

UV absorbance spectrum of a 32-micromolar solution of andalusicin A in 50% acetonitrile, pH 5.0 was recorded using NanoDrop 2000 UV-Vis spectrophotometer (ThermoFisher Scientific, USA). The optical rotation value was measured on a Russian State primary standard of unit of polarization plane rotation angle polarimeter (ckp.vniiofi.ru, VNIIOFI). Measurement was performed in a 10-cm microcell at 20°C, He-Ne laser (632 nm), compound concentration 0.9 mg/mL in 50% acetonitrile.

For MALDI-TOF-MS analysis, sample aliquots were mixed (2:1) with the matrix solution (20 mg/mL 2,5-dihydroxybenzoic acid (Sigma-Aldrich), 0.5% TFA in 30% acetonitrile) on a steel target. *M/z* values of molecular ions were measured in reflector mode on UltrafleXtreme II MALDI-TOF-TOF (Bruker Daltonics) equipped with Nd laser. The accuracy of the obtained masses was within 0.1 Da. Spectra of fragmentation were obtained in LIFT mode; the accuracy of product ions measurement was within 1 Da range. Mass-spectra were processed using FlexAnalysis 3.2 software (Bruker Daltonics, Germany) and analyzed manually. For high-resolution mass spectrometry andalusicin A was diluted to 0.5 μg/mL in 0.1% formic acid/ 25% acetonitrile. The exact mass was measured on FTICR MS Apex Ultra (Bruker Daltonics).

The NMR sample was prepared by dissolving 3.5 mg of andalusicin in 200 μl DMSO-d6 (Sigma-Aldrich, US), and after adding 1 mM of TMS for spectral referencing, it was introduced in a 3mm NMR tube. All spectra were recorded at 30°C on an 800 MHz AvanceIII NMR spectrometer equipped with a cryogenic QCP probe head (Bruker, Billerica, USA). Homonuclear NOESY and TOCSY spectra were acquired as a matrix of 16k x 512 complex points, with a 1 s relaxation delay. The TOCSY spin lock time was set to 60 ms, and 4 scans per increment were recorded. For the NOESY spectrum, the mixing time was set to 200 ms, and 32 scans per increment were recorded. 1H-13C spectra were acquired with the standard Bruker pulse sequences (hsqcetgpsisp2.2, hsqcdietgpsisp.2, and hmbcetgpl3nd) as matrices of 4k x 512 complex points. In the

13C dimension, the carrier was set to 40 ppm and the spectral window to 80 ppm. Spectra were acquired with 32 scans for the HSQC sequences and 64 scans for the HMBC.

All spectra were transformed with the Bruker topspin3.5 program, and the assignment was manually performed with the same software.

Whole-genome DNA sequencing, genome assembly, and annotation

Genomic DNA of *B. thuringiensis* sv. *andalousiensis* NRRL B23139 was extracted with GeneJET Genomic DNA Purification Kit (ThermoFisher Scientific, USA) according to the manufacturer's protocol. The library was prepared using NEBNext Ultra II kit (New England Biolabs, USA) and sequenced using HiSeq platform (Illumina, USA). Genome assembly was performed using SPAdes v. 3.9.1 (Bankevich et al., 2012) with *B. thuringiensis* HD521 genome sequence (GenBank: CP010106.1) (Li et al., 2015) as a reference. Genome annotation was performed using RAST software (Overbeek et al., 2014). Genes encoding andalusicin precursor peptides were identified using BLASTp against products of all ORFs with a partial sequence of the peptide predicted by MALDI-TOF-MS/MS used as a query.

Bioinformatic analysis

A dataset of LanKC sequences and corresponding automatically predicted precursor peptides was retrieved from work by (Walker et al., 2020). 1,646 class III LanKC sequences were acquired from the NCBI Protein database (https://www.ncbi.nlm.nih.gov/protein/). The sequence of andalusicin LanKC (AncKC) was added, and the resulting dataset was used for a sequence similarity network construction using the EFI-EST web tool (Gerlt et al., 2015) with a strict alignment score cutoff (130, equivalent to an e-value of 10^{-130}). For visualization of the network, the Cytoscape program was used (Shannon et al., 2003). The cluster (connected component) containing AncKC was selected for further analysis (Group II, Figure 7A). Another cluster (Group IV, Figure 7A) consisting mainly of LanKC from Firmicutes was also included in phylogenetic reconstruction to serve as an outgroup for tree rooting. Protein sequences were clustered using MMseqs2 version 12.113e3 (Steinegger and Söding, 2017) with 90% identity and 90% coverage cutoffs to remove redundancy. Obtained representative sequences were trimmed to C-terminal 400 amino acids representing LanC domains and aligned using MUSCLE v3.8.31 (Edgar, 2004) with default settings. Alignment was then manually corrected, and highly variable C- and N- terminal portions were removed. Columns containing more than 90% of gaps were removed using ClipKIT v1.0.7 (Steenwyk et al., 2020). Phylogenetic tree was reconstructed using the maximum likelihood method as implemented in RaxML (Stamatakis, 2014). Reconstruction was performed with the LG amino acid substitution model, gamma-distributed evolutionary rates, and a rapid bootstrapping that converged after 500 iterations according to autoMRE criterion. All proteins encoded within 10 kbp upstream or downstream to the *lanKC* present on the tree were obtained from RefSeq database on February 2021 (https://www.ncbi.nlm.nih.gov/refseq/) and functionally annotated using conserved domain database (CDD) database via Batch CD-Search web server (Lu et al., 2020).

Dataset for *de novo* motif discovery consisted of the precursor peptides associated with all LanKCs from the network clusters selected for the phylogenetic analysis. For LanKCs shown on the tree, the prediction of ORFs encoding potential precursor peptides was refined manually (Table S4). An ORF was considered as encoding a precursor peptide if it was preceded by a strong ribosome binding site, was located in the proximity to *lanKC,* and its product harbored characteristic leader motif LxLQ described previously (Müller et al., 2011). Prior to motif discovery the duplicates were removed from the dataset via clustering with MMseqs2 (Steinegger and Söding, 2017) with 95% identity and 95% coverage cutoffs. Motifs in the representative precursor peptide sequences were identified using MEME v5.0.2 (Bailey et al., 2009) with the following parameters: any number of sites per sequence, minimum motif width 6, maximum width 60, and 30 minimum sites per motif. Manually predicted sequences of precursor peptides associated with LanKCs present on the phylogenetic tree were then scanned for the presence of motifs with MAST (Bailey et al., 2009) using per sequence composition in significance calculation and omitting motifs with e-value greater than 0.05. The phylogenetic tree, along with gene co-occurrence and precursor peptide motif architectures was visualized using iTOL (Letunic and Bork, 2019). Custom python scripts used in this study are available at https://github.com/bikdm12/andalusicin.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed in Python 3.7 using SciPy 1.3.1 (Virtanen et al., 2020) and statsmodels 0.10.1 (Seabold and Perktold, 2010) packages. The proportion of cells with the compromised membrane was calculated by manual counting the PI-positive and DAPI-positive cells. At least 3 independent measurements were performed for each of the inhibitors (Table S2). First, we compared the replicates using Fisher's exact test. For all inhibitors, we found no difference between replicates at the significance level 0.05 (with Bonferroni multiplicity correction). Thus, we pooled the replicates and conducted another series of Fisher's exact tests with Bonferroni adjustment and the significance level 0.05. No significant difference was observed only between negative control (DMSO) and CCCP samples. The 95% confidence intervals for the proportions of PI-stained cells were calculated using Wilson method.