Exploring the chemical space of post-translationally modified peptides in Streptomyces with machine learning

Kloosterman, A.M.

Author: Kloosterman, A.M.
Title: Exploring the chemical space of post-translationally modified peptides in Streptomyces with machine learning
Issue Date: 2021-05-12

4

Characterization of a novel lanthipeptide class discovered with a machine-learning genome mining tool

Alexander M. Kloosterman*
Somayah S. Elsayed*
Chao Du
Marnix H. Medema
Gilles P. van Wezel

*These authors contributed equally to this work.

The work described in this chapter is part of the publication:

Kloosterman, et al., *Expansion of RiPP biosynthetic space through integration of pan-genomics and machine learning uncovers a novel class of lantibiotics.* PLOS Biol, 2020. **18**(12): e3002016.

Abstract

Ribosomally synthesized and post-translationally modified peptides (RiPPs) represent a highly diverse and quickly expanding class of natural products that is divided into genetically and chemically distinct subclasses. The identification of novel subclasses is an excellent opportunity to identify chemical scaffolds and expand our knowledge of biosynthetic pathways, but unsuitable for similarity-based genome mining. Here, we report on the characterization of a novel RiPP subclass that has been identified using decRiPPter, a bioinformatic tool for the discovery of novel RiPP subclasses. This RiPP subclass is commonly found among streptomycetes, with one BGC present in every ten genomes. A representative gene cluster from *Streptomyces pristinaespiralis* was selected for characterization. Placing a nearby regulator behind a constitutive promoter resulted in the activation of the BGC, and several masses were detected in the crude chemical extracts of cultures with LCMS. MS/MS fragmentation analysis, chemical labeling and NMR were combined to elucidate the structure one of the most abundant masses, which could be linked to one of the predicted precursor peptides encoded in the gene cluster. Structural analysis showed that this gene cluster specifies lanthipeptides, called the pristinins, despite the fact that the gene cluster did not contain genes for canonical enzymes that create the required lanthionine bridges. The lanthipeptide therefore belongs to a novel class, which we call class V. Through comparisons with previously identified RiPPs, two genes are proposed to encode the enzymes that form the lanthionine bridge in this subclass. These genes are present in a wide variety of genetic contexts, both within *Streptomyces*, but also in other *Actinobacteria* and in *Firmicutes.* This work not only showcases the potential of decRiPPter, but also expands the list of RiPP subclasses and identifies promising enzyme queries that can be used in further genome mining studies.

Introduction

The continuing increase in available sequence data has fueled the identification of many new natural product biosynthetic gene clusters (BGCs) through genome mining [26, 185]. Homology-based genome mining methods expand classes of natural products through the identification of key genetic markers in contexts. Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a highly diverse collection of natural products, which are split into several subclasses that share biosynthetic pathways [42, 48]. Identification of novel RiPP subclasses using bioinformatics alone is a difficult challenge, as these classes by definition lack any known marker. Despite a large pool of available sequence data, this process still mostly depends on traditional high-throughput screening (HTS).

Although many compounds have been found with high-throughput screening, not all BGCs are activated under laboratory conditions and have therefore likely been missed [186]. This means that a large part of the chemical space of natural products is still to be discovered. Especially for streptomycetes, the number of BGCs far exceeds the natural products that have so far been identified. Growing a strain under many different conditions followed by extensive metabolic profiling, the so-called One Strain MAny Compounds (OSMAC) approach, is a good way to find natural products produced by a single strain [187, 188]. Eliciting strategies are complementary to OSMAC, aiming to mimick the ecological growth conditions of the producer strain and hence the activation of cryptic compounds [33, 189, 190]. In contrast to these general methods, BGC-specific methods study the activation of a single BGC of interest, or even force their expression through engineered promoters and heterologous expression [33, 34].

In this study, we characterize a RiPP family previously discovered by decRiPPter (Chapter 2). DecRiPPter identifies RiPP BGCs with a Support Vector Machine (SVM) classifier that identifies RiPP precursors. It is not limited by biosynthetic domains, and could therefore identify new RiPP subclasses. The RiPP family studied here is prevalent in streptomycetes, with one representative BGC present for every ten *Streptomyces* genomes analyzed. A BGC from *S.*

pristinaespiralis from this family is silent under the growth condition tested, but can be activated by placing a nearby regulator behind a strong promoter. The BGC specifies a novel lanthipeptide, called pristinin A3. Since the BGC lacks any homologs of the lanthionine-forming modifying enzymes, a new route must be required for their biosynthesis, meaning pristinin is a class V lanthipeptide. Based on similarities with enzymes encoded in other BGCs, two gene products, called SprPT and SprH3 are proposed as candidates for their biosynthesis. We further show that their encoding genes are found in a wide variety of different contexts, meaning that they could be used as a new handle for RiPP genome mining. Our work not only validates decRiPPter's capabilities to detect novel RiPP subclasses, but also provides new genome mining rules for the expansion of one the best-studied RiPP subclasses.

Results and Discussion

Discovery of a novel family of lanthipeptides

In Chapter 3, we described the applicability of decRiPPter for the mining of *Streptomyces* genomes for RiPP BGCs. To validate the capacity of decRiPPter to find novel RiPP subclasses, we set out to experimentally characterize one of the candidate families (Chapter 3, Figure 4, Other, red marker). Gene clusters belonging to this family shared several genes encoding flavoproteins, methyltransferases, oxidoreductases and occasionally a phosphotransferase. Importantly, the predicted precursor peptides encoded by these putative BGCs showed clear conservation of the N-terminal region, while varying more in the C-terminal region (Text S1). This distinction is typical of RiPP precursors, as the N-terminal leader peptide is used as a recognition site for modifying enzymes, while the C-terminal core peptide can be more variable [43].

One of the gene clusters belonging to this candidate family was identified in *Streptomyces pristinaespiralis* ATCC 25468 (Figure 1A; Table 1). *S. pristinaespiralis* is known for the production of pristinamycin, and was selected for experimental work since the strain was readily available and genetically tractable [191, 192]. The gene cluster was named after its origin (*spr*: *Streptomyces pristinaespiralis* RiPP), and the genes were named after their putative function.

The gene cluster contains four genes encoding putative precursor peptides, although only three of the peptides (SprA1-A3) showed similarity to each other and to the other peptides in the same family (S1 Text). The fourth predicted precursor peptide (encoded by s*prX*) did not align with any of the other peptides and was assumed to be a false positive. The products encoded by *sprA1* and *sprA2* were highly similar to one another compared to the *sprA3* gene product (Fig 1A). Occurrence of two distinct genes for precursors within a single RiPP BGC is typical of two-component lanthipeptides [193].

Most of the modifying enzymes present in the gene cluster had not previously been implicated in RiPP biosynthesis. The predicted *sprF2* gene product, however, shows high similarity to cysteine decarboxylases such as EpiD and CypD. These enzymes decarboxylate C-terminal cysteine residues, which is

Table 1. Annotation of the *Streptomyces pristinaespiralis* RiPP (*spr*) gene cluster.

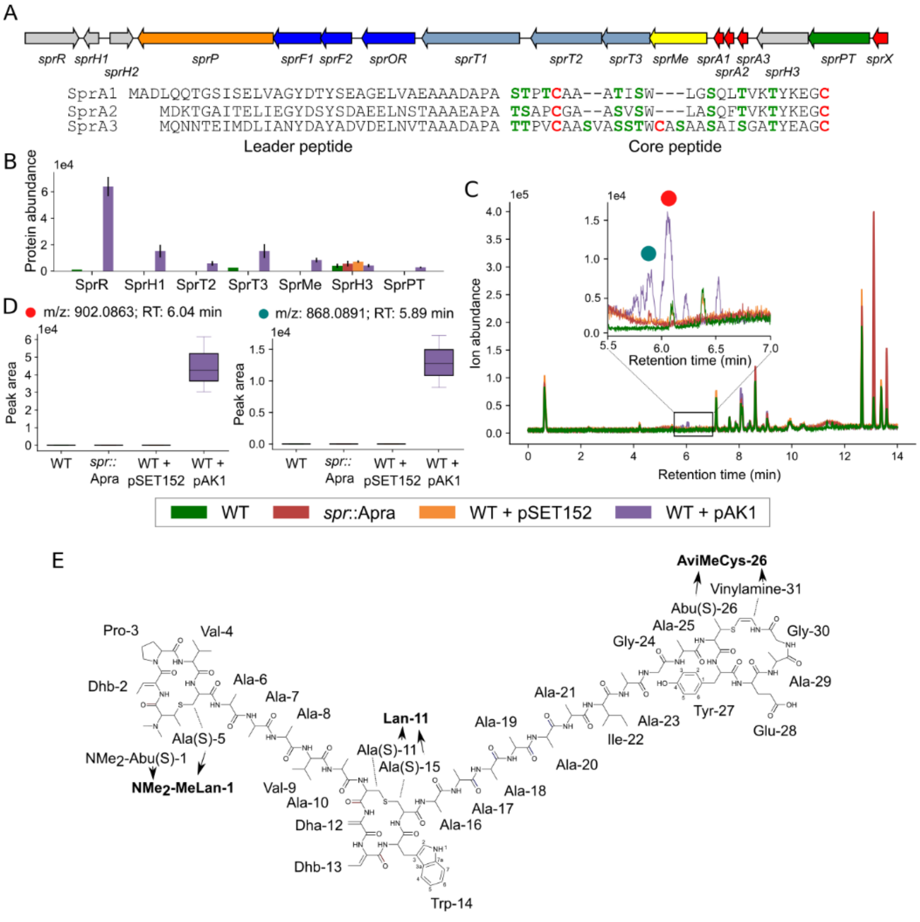
Gene name	NCBI Genbank Accession	NCBI Annotation of gene product	Protein domains	Proposed function
sprR	ALC22061.1	LuxR family transcriptional regulator		Cluster-specific regulator
sprH1	ALC22062.1	hypothetical protein		Unknown
sprH2	ALC22063.1	hypothetical protein		Unknown
sprP	ALC22064.1	Peptidase M16 domain-containing protein	PF00675 PF05193	RiPP maturation protease
sprF1	ALC22065.1	Flavoprotein	PF01636	Cysteine decarboxylation
sprF2	ALC22066.1	Flavoprotein	PF02441	Cysteine decarboxylation
sprOR	ALC22067.1	5,10-methylene tetrahydromethanopterin reductase	PF00291	Reduction of dehydroalanine and dehydrobutyric acid
sprT1	ALC22068.1	ABC transporter ATP-binding protein	PF00005 PF00664	Transport
sprT2	ALC22069.1	ABC transporter	PF12698	Transport
sprT3	ALC22070.1	ABC transporter ATP-binding protein	PF00005 PF13732	Transport
sprMe	ALC22071.1	carminomycin 4-O-methyltransferase	PF00891	N-terminal methylation
sprA1	ALC22072.1	hypothetical protein		RiPP precursor
sprA2	ALC22073.1	hypothetical protein		RiPP precursor
sprA3	ALC22074.1	hypothetical protein		RiPP precursor
sprH3	ALC22075.1	hypothetical protein	PF17914	Dehydration/cyclization
sprPT	ALC22076.1	hypothetical protein	PF01636	Dehydration/cyclization
sprX	ALC22077.1	hypothetical protein		Unknown

the first step in the formation of C-terminal loop structures called *S*-[(*Z*)-2-aminovinyl]-D-cysteine (AviCys) and *S*-[(*Z*)-2-aminovinyl]-(3S)-3-methyl-D-cysteine (AviMeCys) [194]. Several RiPP classes have been reported with this modification, including lanthipeptides, cypemycins and thioviridamides, although they are only consistently present in cypemycins and thioviridamides. This type of modification is less common among lanthipeptides, with only nine out of 120 lanthipeptide gene clusters in MIBiG encoding the required decarboxylase. Genes encoding cysteine-decarboxylating enzymes are present in non-RiPP gene clusters (Table S3) and are also associated with other metabolic pathways [195]. In theory, though, this BGC could have been detected using a bait-based approach using these genes as queries.

A more detailed comparison with the gene clusters in MIBiG [196] showed that two more genes from the thioviridamide gene cluster were homologous to two genes encoding a predicted phosphotransferase (*sprPT*) and a hypothetical protein (*sprH3*), respectively. Taken together with the homologous cysteine decarboxylase, it appeared that our gene cluster was distantly related to the thioviridamide gene cluster [197]. Thioviridamide-like compounds are primarily known for their thioamide residues, for which a TfuA-associated YcaO is thought to be responsible [52, 95]. However, a YcaO homologue was not encoded by the gene cluster, making it unlikely that this gene cluster should produce thioamide-containing RiPPs.

Two strains were created to help determine the natural product specified by the BGC. For the first strain, the entire gene cluster was replaced by an apramycin resistance cassette (aac3(IV)) by homologous recombination with the pWHM3 vector [198]. Both flanking regions were cloned into this vector, creating the vector pAK3. Subsequent homologous recombination resulted in a strain where the gene cluster was replaced by the *aac3(IV)* gene, called *spr*::apra (Materials and Methods). In case the gene cluster was natively expressed, this strain should allow for easy identification of the natural product by comparative metabolomics. In the second approach, we sought to activate the BGC in case it was not natively expressed. To this end, we targeted the cluster-situated *luxR*-family transcriptional regulatory gene *sprR*. The *sprR* gene was expressed from the strong and constitutive *gapdh* promoter from *S. coelicolor* ($\mathrm{p}_{\mathrm{gapdh}}$) on the integrative vector pSET152 [199]. The resulting construct (pAK1) was transformed to *S. pristinaespiralis* by protoplast transformation.

To assess the expression of the gene cluster in the transformants, we analyzed changes in the global expression profiles in 2 days and 7 days old samples of NMMP-grown cultures using quantitative proteomics (Fig 1B). Aside from the regulator itself, six out of the sixteen other proteins were detected in



● **Pristinin A3 (1)**

Figure 1. The pristinin BGC (*spr*) of *S. pristinaespiralis* produces a highly modified RiPP. A) The *spr* gene cluster encodes three putative RiPP precursors, three transporters, a peptidase and an assortment of modifying enzymes (see Table 1). Alignment of the predicted precursor peptides is given below. B) Protein abundance of the products of the *spr* gene cluster in *S. pristinaespiralis* ATCC 25468 and its derivatives. Strains were grown in NMMP and samples were taken after 7 days. Enhanced expression of the regulator (from construct pAK1) resulted in the partial activation of the gene cluster. Proteins that could not be detected are not illustrated. C) Overlay chromatogram of crude extracts from strains grown under the same conditions as under B), samples after 7 days. Several peaks were detected in the extract from the strain with expression construct pAK1 between 7 and 8 minutes. D) Boxplot of two peaks detected only in the strain with pAK1. The two masses could be related to two of the three precursors peptides. E) 2D structure of pristinin A3 (**1**), derived from the SprA3 precursor. The compound has a mass of 2703.235 Da.

the strain containing expression construct pAK1, while only SprPT could be detected in the strain carrying the empty vector pSET152. SprPT was also detected in the proteome of *spr*::apra, however, indicating a false positive. In the wild-type strain, SprT3 and SprR were detected, but only in a single replicate and at a much lower level. Overall, these results suggest that under the chosen growth conditions the gene cluster was expressed at very low amounts in wild-type cells, and was activated when the expression of the likely pathway-specific regulatory gene was enhanced. This makes *spr* a likely silent BGC under the conditions tested.

To see if a RiPP was produced, the same cultures used for proteomics were separated into mycelial biomass and supernatant. The biomass was extracted with methanol, while HP20 beads were added to the supernatants to adsorb secreted natural products. Analysis of the crude methanol extracts and the HP20 eluents with HPLC-MS revealed several peaks eluting between 5.5 and 7 minutes in the methanol extracts (Figure 1C), which were not found in extracts from wild-type strain or the strain containing the empty vector. Feature detection with MZmine followed by statistical analysis with MetaboAnalyst revealed seven unique peaks, with *m/z* between 707.3534 and 918.0807 (Figure S1). The isotope patterns of these peaks showed that six of the identified ions were triply charged. Careful analysis of adduct ions and looking for mass increases consistent with Na- or K-addition, led to the conclusion that these peaks corresponded to the $[M+3H]^+$ adduct, suggesting monoisotopic masses in the range of 2,604.273 and 2,754.242 Da. The highest signal came from the compound with a monoisotopic mass of 2,703.245. Four of the other masses seemed to be related to this mass, as they were different in increments of 4, 14, or 16 Da (Table S4). We therefore reasoned that this mass was the product of one of the precursor peptides, while others were incompletely processed peptides. Another mass of 2,601.2433 could not be directly linked to the mass of 2,703.245. This mass was nevertheless only detected in extracts of the strain harboring pAK1 (Figure 1D), suggesting it is the product of another precursor peptide, although whether or not it is the final product remains unclear.

To further verify that the identified masses indeed belonged to the RiPP precursors in our gene cluster, we first removed the apramycin resistance cassette from *spr*::apra using the pUWLCRE vector [200], creating strain Δ*spr* (Materials and Methods). The expression construct pAK1 and an empty pSET152 vector were transformed to the *spr* null mutant. When these transformants were grown under the same conditions, the aforementioned peaks were not detected, further suggesting that they were products of this gene cluster (Figure 2A).

Most masses were detected in only low amounts. In order to resolve this, we created a similar construct as pAK1, but this time using the low-copy shuttle vector pHJL401 as the vector [201]. The plasmid pAK2 was introduced into *S. pristinaespiralis* and the transformants grown in NMMP for 7 days. Extraction of the mycelial biomass with methanol resulted in a higher abundance of the masses previously detected (Figure 2B). Consistent with the MS profiles of pAK1 transformants, also pAK2 transformants produced an abundant peak corresponding to a monoisotopic mass of 2,703.245 Da, as well as a second peak corresponding to a monoisotopic mass of 2,553.260 Da. Many more masses were detected, most of which could be related to one of these two masses, suggesting these are the final products, related to two distinct precursors (Figure 2CD, Table S4 and Table S5).

We then performed MS/MS analysis of the extracts of the pAK2 transformants to identify the metabolites. Building on the hypothesis that the abundant peaks corresponded to the modified SprA1-A3 peptides, we used their peptide sequences to map the fragments. The fragmentation pattern of the peak with a mass of 2,703.245 Da could indeed be assigned to the sprA3 precursor sequence, but only when several mass adjustments of -16 Da, -18 Da, +28 Da and -46 Da were applied (Figure S2A, Table S6). Similarly, fragments for the mass of 2,553.260 could be matched to the SprA2 precursor sequence considering the same mass adjustments (Figure S2B, Table S7). The compounds were named the pristinins, and the individual compounds were named after their precursors (pristinin A3 and pristinin A2, respectively).

All of the -18 and -16 Da adjustments were predicted on serine and threonine fragments. These mass differences are typical of dehydration (-18 Da) of the residues to dehydroalanine (Dha) and dehydrobutyrine (Dhb). Reduction of these dehydrated amino acids (+2 Da) would then give rise to alanine and

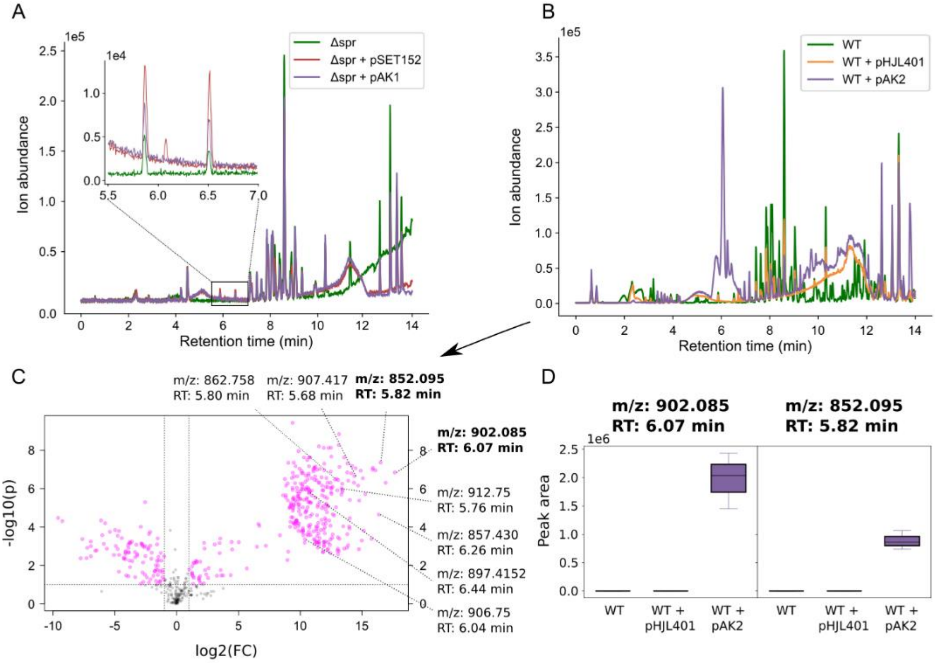


Figure 2. Chromatograms comparing the extracted compounds in knockout strains and highly producing strains. A) Strains lacking the *spr* gene cluster are unable to produce the extracted products, even when transformed with pAK1. B) Chromatogram of methanol extracts made from *S. pristinaespiralis* harboring no vector (WT) an empty pHJL401 vector (pHJL401), or pAK2 (pHJL401 with *sprR* behind p_{gap}). A large peak can be seen in the extracts of strains harboring pAK2, not seen in extracts of the other strains. C) Volcano plot comparing extracts of the strain containing pAK2 with the strain containing pHJL401. Peaks in pink had p-value ≥ 0.1 and a fold-change of ≥ 2. A large collection of peaks can be identified with log2(fold-change) ≥ 10. The two largest peaks (bold) corresponding to different monoisotopic masses could be related to the SprA2 and SprA3 precursors by MS/MS (S11 Fig). Many of the other masses eluted at comparable times, and had masses that were close to the two major peaks, suggesting they were derived from them. Clear mass differences could be identified for some of the identified masses (Table S5). Whether the largest peaks indeed correspond to the final product remains to be determined. D) Extracted ion chromatograms of the two major peaks identified from the volcano plot. The two masses were only detected in the strain harboring pAK2.

butyric acid residues, a modification that has been reported for lanthipeptides [202]. A modification of +28 Da suggests a dual methylation among the five N-terminal residues, which is consistent with the methyltransferase SprMe that is encoded by the *spr* gene cluster. The loss of -46 Da could be attributed to the C-terminal cysteine. This mass difference correlates to oxidative decarboxylation, which is consistent with the cysteine decarboxylase SprF2 that is encoded by the cluster. The loss of -18 Da in a threonine residue close to the modified cysteine suggests the presence of an AviMeCys group at the C-terminal end of the peptide. The lack of fragments for the residues $T^{-18}YEAGC^{-46}$ in the fragments pristinin A3 further supports the presence of an AviMeCys-containing C-terminal ring.

Surprisingly, no fragments were found of the residues $S^{-18}S^{-18}T^{-18}WC$ in the center of pristinin A3, or for the N-terminal $[T^{-18}T^{-18}PVC]^{+28}$ region. Considering the other modifications typical of lanthipeptides, and the likely presence of a thioether crosslink in the AviMeCys group, we hypothesized the presence of thioether crosslinks between the Dhbs and cysteines. To find further support for this hypothesis, we treated the purified product of SprA3 with iodoacetamide (IAA). Iodoacetamide alkylates free cysteines, while cysteines in thioether bridges remain unmodified [203]. In agreement with our hypothesis, treatment with iodoacetamide did not affect the observed masses, despite the presence of three cysteines in the peptide (Figure S7).

NMR confirms the presence of lanthionine bridges in the pristinins

To further ascertain the presence of the proposed modifications, we purified the peak corresponding to pristinin A3. Since the products were not detected when cultures were grown in 500 mL cultures, we grew 100 × 20mL cultures (2L total) of a transformant harboring the expression plasmid pAK2. The culture was then extracted and the extract was subjected to a series of chromatographic fractionations, which resulted in the purification of pristinin A3 (**1**) (Materials and Methods). The purified compound was dissolved in deuterated DMSO (DMSO-d_6) for NMR analysis. Extensive purification allowed us to purify 1.1 mg of the compound. While the amount of material meant that the NMR signal was low, we could derive many key features of the peptide in the ^{1}H NMR spectrum (Figure S3, Figure S4A). The NH signals in the ^{1}H NMR spectrum were very broad using DMSO-d_6 as solvent. We therefore changed to $CD_3CN:H_2O$ 9:1 as the solvent, which showed very good NH signals for the recently identified similar peptide cacaoidin [204]. Indeed, sharper peaks and better HMBC correlations could be observed (Figure S5 and S6). Re-analysis of pristinin A3 (**1**) using LC-MS showed that the compound was partially oxidized, i.e. a mixture of compounds was analyzed in the NMR run using CD_3CN:H_2O as a solvent (Table S11). MS/MS fragmentation suggested that the oxidation occurred consistently in the center and N-terminal ring structures (Table S12).

Combined analysis of the 2D COSY, TOCSY, HSQC, HMBC and NOESY NMR spectra obtained in DMSO-*d₆* (Figure S4, Table S9) supported the proposed structure of pristinin A3 (**1**) (Figure 1E). In the 2D spectra several spin systems were identified, which were consistent with the amino acid sequence of SprA3 and the MS/MS fragmentation data (Figure S3). These amino acid residues were 2 Val, 2 Gly, 1 Pro, 1 Trp, 1 Ile, 1 Tyr, 1 Glu and multiple mostly overlapping Ala. Additionally, we identified spin systems consistent with the proposed modified amino acid residues. These were 2 Dhbs, 2 β-thioalanines (Ala(S)), 1 Dha, 1 β-thioaminobutyric acid (Abu(S)), and 1 aminovinyl group. Due to weak signals, we could not use the HSQC-TOCSY spectra to further support the identified residues. There was no clear evidence in the NMR spectra of the presence of Thr or Ser amino acid residues, which corroborated the hypothesis that all the Thr and Ser residues identified in SprA3 had been modified.

We next sought evidence for the connectivity of the identified amino acids. The connectivity of the amino acid residues through NMR could be readily established through the Hα-NH (i, i+1), Hβ-NH (i, i+1), and NH-NH (i, i+1) NOESY correlations. Based on this, the AviMeCys-containing C-terminal ring and its extension up to Ala-21 could be unambiguously established to be in accordance with the proposed structure through the MS/MS data (Figure S3). Importantly, the same structural fragment could be clearly observed in the sample analyzed in CD_3CN:H_2O 9:1, supporting the observation from the MS/MS data that the oxidation of pristinin A3 (**1**) was in the rings closer to the N-terminus. The NMR data in CD_3CN:H_2O confirmed the sequence of Ala-25 up to Glu-28, because some of the Hα and NH signals for these residues, that were overlapping in DMSO-*d₆*, were well separated in CD_3CN:H_2O (Figure S4, S5 and S6, Table S10). Additionally, HMBC correlations could be observed to the carboxyl group of Glu in CD_3CN:H_2O. The NOESY correlations in DMSO-*d₆* further unambiguously confirmed the peptide sequence observed in MS/MS for Dhb-2 to Ala-10, Dha-12 to Dhb-13, and Trp14 to Ala-16 (Table S9). The sequence of Ala-17 to Ala-20 had overlapping Hα and NH signals. However, the correlation pattern observed

Table 2. Summary of the different methods used to identify the amino acid residues of pristinin A3. Symbols indicate whether residues and their connectivity were confirmed (+), partly confirmed (±), or not confirmed (-).

Amino acid residues	Gene seq.	HRMS/MS	NMR	Acid hydrolysis[a]	Amino acid residues	Gene seq.	HRMS/MS	NMR	Acid hydrolysis[a]
NMe_2-MeLan-1*	-	+	±[b]	+	Ala-17	-	+	+	+
Dhb-2	-	+	+	-	Ala-18	+	+	+	+
Pro-3	+	+	+	+	Ala-19	+	+	+	+
Val-4	+	+	+	+	Ala-20	-	+	+	+
Ala-6	+	+	+	+	Ala-21	+	+	+	+
Ala-7	+	+	+	+	Ile-22	+	+	+	+
Ala-8	-	+	+	+	Ala-23	-	+	+	+
Val-9	+	+	+	+	Gly-24	+	+	+	-
Ala-10	+	+	+	+	Ala-25	+	+	+	+
Lan-11*	-	+	±[c]	+	AviMeCys-26*	-	+	±	-
Dha-12	-	+	+	-	Tyr-27	+	+	+	+
Dhb-13	-	+	+	-	Glu-28	+	+	+	+
Trp-14	+	+	±[d]	-	Ala-29	+	+	+	+
Ala-16	+	+	+	+	Gly-30	+	+	+	-

[a] Acid hydrolysis only confirms the amino residues, but not their connectivity.

[b] Only Ala(S)-5 could be observed in NMR;

[c] Only Ala(S)-15 could be observed in NMR

[d] Trp-14 and its connectivity to Ala(S)-15 could be confirmed by NMR, but its connectivity to Dhb-13 could not be confirmed.

* NMe2-MeLan-1 = NMe_2-Abu(S)-1 + Ala(S)-5; Lan-11 = Ala(S)-11 + Ala(S)-15; AviMeCys-26 = Abu(S)-26 + Vinylamine-31

and the peak integration support a series of alanine residues to be the connection between Ala-16 and Ala-21, as was also indicated by the MS/MS data.

It was not possible to establish the connection between Dhb-13 and Trp-14 using NMR. At the same time, a Dha–Dhb sequence could be clearly established using NMR. The fact that Dha and Dhb are the products of modified Ser and Thr residues, respectively, and the fact that the only Ser–Thr sequence in the SprA3 precursor lies before Trp, inevitably means that the observed Dha–Dhb structural fragment is connected to Trp-14 and positioned as Dha12 and

Dhb-13. Finally, the thioether crosslinks of the proposed N-terminal and center ring structures could not be completely resolved based on NMR data alone. This is because the ^{1}H NMR resonance for a CH/CH_2 group attached to a sulfur atom should be around δ_H 3 ppm, which is close to the area where the water signal in DMSO-*d₆* (δ_H 3.3 ppm) is suppressed in the NMR experiments. Water suppression greatly affected the smaller signals around this area. Nevertheless, we managed to establish and position Ala(S)-5 and Ala(S)-15, both of which have to be part of a thioether bond as proven through the IAA labelling experiment discussed earlier. This left only one residue in each of the two additional rings observed in MS/MS, which was not accounted for by NMR (Figure S3). Based on this, an NMe_2-Abu(S)-1 and Ala(S)-11 could be proposed to form thioether bridges with Ala(S)-5 and Ala(S)-15, respectively, resulting in the formation of *N*,*N*-dimethyl-β-methyllanthionine (NMe_2-MeLan) and lanthionine (Lan) residues, respectively. As a further evidence, we hydrolyzed the purified peptide with 6M HCl at 110°C for 24h. Under these conditions, the amide bond should be hydrolyzed, while the thioether bond should be unaffected [205]. The resulting mixture of amino acids was analyzed using LC-HRMS and was indeed found to contain peaks with exact masses corresponding to NMe_2-MeLan and Lan (Table S8). Thus, the primary sequence of the peptide, the MS/MS fragmentation data, the NMR data, acid hydrolysis and labelling experiments (Table 2) allowed us to elucidate the 2D structure of pristinin A3 (**1**; Figure 1E).

The RiPPs characterized here contain a number of modifications that have previously been identified in other RiPPs. A recent study, which appeared around the time of submission for this paper, describes a RiPP found by activity-based screening, called cacaoidin, that has many of the same modifications [204], and is additionally glycosylated. The serines converted to alanines in cacaoidin were all D-alanines. It therefore seems probable that the converted serines in pristinin A3 (**1**) were also converted to D-alanines, which could be determined by further chemical analyses. BLAST analysis shows that the genes of the cacaoidin BGC show low similarity to those in the *spr* BGC, and the precursor genes do not seem directly related. However, the same Pfam domains are found in both BGCs, indicating that both BGCs belong to the same RiPP class. The authors describing cacaoidin remark that these modifications were found previously in linaridins and lanthipeptides, and therefore named this class the lanthidins. While some enzymes encoded by the BGCs of this RiPP class indeed show low similarity to enzymes involved in the biosynthesis of characterized RiPPs, the combination of modifications makes it a novel RiPP subclass that was not previously detected by other RiPP genome mining tools. Overall, these findings further support the potential of decRiPPter to identify novel RiPP BGCs.

SprH3/SprPT are candidates for the enzymes that install lanthionines

Taken together, we have shown that pristinin A3 contained a number of posttranslational modifications that are typical of lanthipeptides. The conversion of serine/threonine to alanine/butyric acid via reduction, the creation of an AviCys moiety and the crosslinks to form thioether bridges are all found in lanthipeptides, and are dependent on dehydration of serine and threonine residues. No homologs of known lanthionine-forming enzymes were found to be encoded by the gene cluster studied. However, *sprH3* and *sprPT* showed homology to two uncharacterized genes of the thioviridamide BGC. Thioviridamide contains an AviCys moiety, the formation of which requires a dehydrated serine residue. The enzymes responsible for dehydration and subsequent cyclization have not been identified yet [94, 206]. Another RiPP subclass with an AviCys moiety is the linaridin subclass. Dehydration of the required serine is thought to be catalyzed by LinH or LinL, neither of which show similarity to the proteins encoded by the thioviridamide BGC or the *spr* BGC. Of note, the cacaoidin BGC also encoded two proteins with the same domains as SprH3 and SprPT (i.e. PF01636 and PF17914). Since the thioviridamide, cacaodin and *spr* gene clusters share a common modification for which the enzyme is unknown, we hypothesize that SprH3 and SprPT carry out the dehydration and cyclization reactions and are therefore likely involved in the maturation of many different RiPPs, with dehydrated residues, AviCys moieties, or thioether bridges. In the latter group, these enzymes candidate as core modifying enzymes of a new lanthipeptide subclass, which we designated lanthipeptide class V.

To find experimental support for the hypothesis that SprH3 and SprPT are the sought-after modifying enzymes, we replaced the gene pair *sprH3*/*PT* with an apramycin resistance cassette (aac3(IV); Materials and Methods). To this end, the flanking regions were amplified with PCR, and placed into the shuttle vector pWHM3. An apramycin resistance was placed between the flanks through restriction and ligation. The resulting vector pAK8 was transformed to *S. pristinaespiralis* and the exchange of genes was confirmed by PCR. The strain was named *sprH3PT*::Ap. This strain was then transformed with the pAK2 vector in order to activate the *spr* BGC, or with an empty pHJL401 vector as a control. The three resulting strains were grown and extracted using the same conditions as described above. Despite this, no masses were detected with HPLC that relate 6to the products of the *spr* BGC (Figure S8A). The genetic modifications made to the BGC therefore disrupt the biosynthesis of the pristinin A2 and A3 precursors.

The extracts obtained from the *sprH3PT*::Ap strain and derivatives suggest that the two products of removed gene pair are indeed involved in the biosynthesis of pristinin A2 and A3. An alternative explanation is that the genetic modifications themselves disrupt the transcription and/or translation of the *spr* BGC, which appears to be in an operon-like structure. We aimed to rule out the latter explanation by providing an additional copy of the removed genes to the *sprH3PT*::Ap strain. The gene pair was amplified with PCR, either with or without the native promoter. When no native promoter was amplified, the amplified gene pair was placed behind the upstream region of XNR_3799, a strong promoter for streptomycetes amplified from *S. lividans* (Zhang, L., personal communication). The two different gene pair regions were then placed either in pHJL401, to create the control constructs pAK4 (native promoter) and pAK6 (XNR promoter), or the pAK2 vector, creating the constructs pAK5 (native promoter) and pAK7 (XNR promoter). A t0 terminator was placed between the DNA fragment harboring the *sprH3/PT* gene pair and the fragment harboring the *sprR* gene in the pAK5 and pAK7 vectors to prevent transcriptional read-through.

The resulting four vectors were transformed to the *sprH3PT*::Ap strain, cultured and extracted as described above. Under these conditions, none of the masses related to the *spr* BGC were detected in the extracts (Figure S8B). This meant that the complementation vectors either do not express the gene pair, or disruption of the *spr* BGC extends beyond the targeted gene pair to also affect the rest of the BGC. The removed fragment is evidently important for the production of the mature RiPP product. However, whether this is due the presence of promoter regions within that fragment that regulate the expression of the BGC, or due to the relevance of the encoded products could not be determined from these experiments. Further experiments, such as *in vitro* enzymology experiments are still required to confirm the function of the SprH3/PT proteins.

The *sprH3/sprPT* gene pair is present in a wide variety of predicted RiPP BGCs

Lanthipeptide core modifying enzymes catalyze the most prominent reaction in lanthipeptide maturation, and as such, are present in many different genetic contexts [68, 73]. To find support for the proposed role of the gene products we studied the distribution of the *SprH3*/*PT* gene pair across *Streptomyces* genomes analyzed by decRiPPter. Using CORASON [185] with the s*prPT* gene as a query yielded 195 homologs in various gene clusters (Figure 3, Materials and Methods). The *sprPT/sprH3* gene pair was completely conserved across all gene clusters for which an uninterrupted contig of DNA was available, strongly supporting their functional interaction and joint involvement. Using the *sprH3* gene as a query yielded similar results. A total of 391 orthologs of the gene pair were found outside *Streptomyces*, particularly in Actinobacteria (219) and Firmicutes (161; Figure S9). Distantly similar homologs of the gene pair were also identified in Cyanobacteria, Planctomycetes and Proteobacteria.

Among the 195 identified gene clusters in *Streptomyces*, the majority (131) overlapped with a gene cluster detected by decRiPPter, indicating that the gene pair was within short intergenetic distance from predicted precursor gene in the same strand orientation. A large fraction (80) also passed the strictest filtering (Table 1), showing that among these gene clusters were many encoding biosynthetic machinery, peptidases and regulators. In contrast, only nine of the gene clusters overlapped with a BGC identified by antiSMASH [39]. Four of these showed the gene pair in apparent operative linkage with a bacteriocin gene cluster, marked as such by the presence of a DUF692 domain. This domain is often associated with small prepeptides, such as the precursor peptides of methanobactin [207]. Another four gene clusters detected by decRiPPter were only overlapping due to the gene pair being on the edge of a neighboring gene cluster.

The genetic context of the gene pairs showed a wide variation (Figure 3, right side). While some gene clusters were mostly homologous to the *spr* gene cluster (Figure 3, group g-h), others shared only a few genes (groups a and d), and some only shared the gene pair itself (Figure 3, b, c and e; Table 3). Many other predicted enzyme families were found to be encoded inside these gene clusters, including YcaO-like proteins, glycosyltransferases, sulfotransferases and aminotransferases. The large variation in genetic contexts combined with the clear association with a predicted precursor indicates that this gene pair likely plays a role in many different RiPP-associated genetic contexts, supporting their proposed role as a core gene pair. We would like to emphasize, however, that not all of these BGCs necessarily specify lanthipeptides. Assuming that the proposed role for the products of *sprH3/PT* in dehydration of serine and threonine residues is correct, these modifications could also lead to AviCys moieties, such as in thioviridamide-like products, or simply remain dehydrated residues. Further genetic and biochemical elucidation of the role of these enzymes is necessary to completely determine the scope of their reactions.

Furthermore, we searched for genes encoding enzymes whose functions are dependent on a lanthipeptide dehydration in their substrate, to find if they were associated with the *sprPT/sprH3* gene pair. Both within and outside *Streptomyces*, homologs of *sprF1* and *sprF2* were often found associated with the gene pair (*sprF1*: 251/586; 40.1%; *sprF2*: 281/586; 48.0%; Table S13).

Table 3. Co-occurrence of genes found in the pristinin gene cluster (*spr*) with homologs of *sprPT* in the analyzed 1,295 *Streptomyces* strains.

Gene name	Co-occurrence with s*prPT* (percentage)	Gene name	Co-occurrence with s*prPT* (percentage)
sprH3	99.49	*sprP*	38.5
sprMe	20	*sprH1*	9.0
sprT1	35.38	*sprH2*	2.0
sprT2	12.31	*sprR*	28.5
sprT3	12.82	*sprA1*	1.03
sprOR	64.62	*sprA2*	1.03
sprF1	39.5	*sprA3*	16.92
sprF2	68.72		

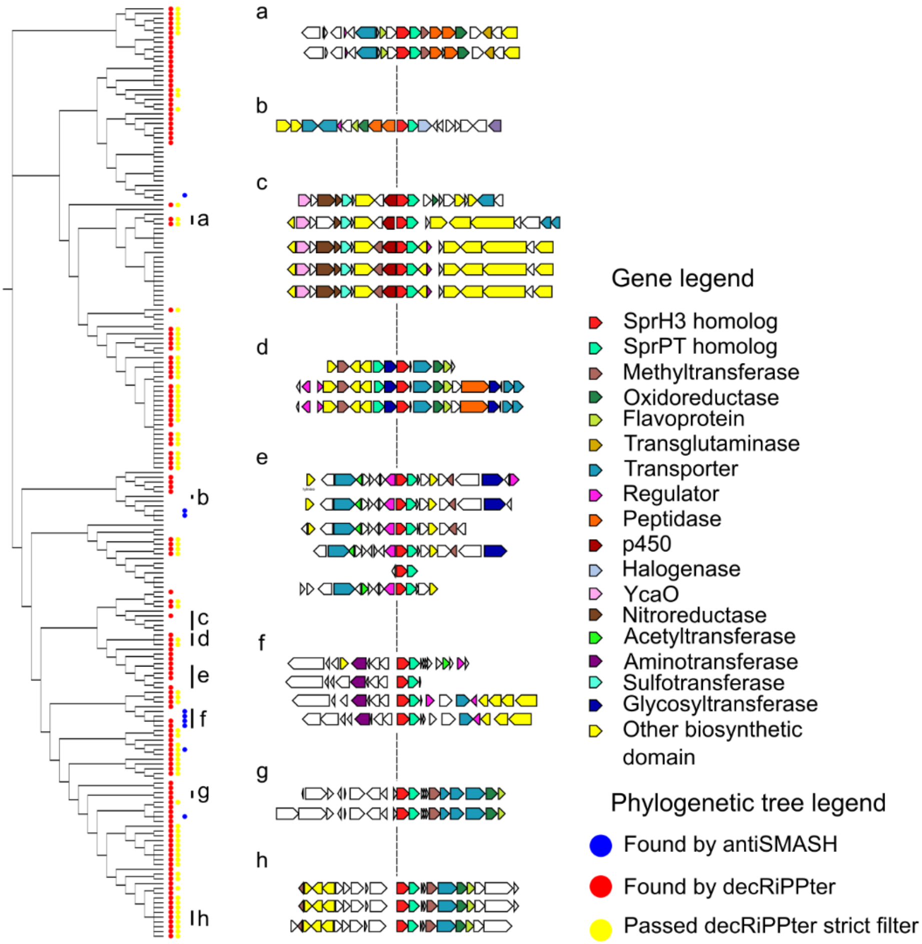


Figure 3. Orthologs of *sprPT* and *sprH3* cooccur in a wide variety of genetic contexts. (Left side) Phylogenetic tree of gene clusters containing homologs of *sprPT* and *sprH3*, visualized by CORASON. A red dot indicates that the genes were present in a gene cluster found by decRiPPter, a yellow dot that it passed the strict filter (Chapter 2). A blue dot indicates overlap with a BGC identified by antiSMASH. (Right side) Several gene clusters with varying genetic contexts are displayed. Group (g) represents the query gene cluster. The genetic context varies, while the gene pair itself is conserved. Color indicates predicted enzymatic activity of the gene products as described in the legend.

Another modification dependent on the presence of dehydrated serine and threonine residues is the conversion of these to alanine and butyric acid, respectively. This conversion is catalyzed either by a zinc-dependent dehydrogenase (LanJ$_A$, also known as as LtnJ) or an NAD(P)H-dependent FMN reductase family enzyme (LanJ$_B$, also known as CrnJ) in lanthipeptides [202]. Outside *Streptomyces*, the genomic surroundings of the *sprPT/sprH3* gene pair occasionally contained homologs of the *lanj$_A$* gene (40/391; 10.1%). An example of such a BGC is that of pediocin A, a known antimicrobial compound of which the structure has yet to be resolved [208]. These gene associations further imply that the SprH3/SprPT gene products apply the canonical dehydration reactions.

A similar modification was observed for pristinin A2 and A3, despite that no homologs of the genes encoding LanJ$_A$ or LanJ$_B$ were identified within the *spr* gene cluster. However, *sprOR* encodes a putative oxidoreductase, and thus is a candidate for this modification. Supporting this, orthologs of *sprOR* were found frequently associated with either canonical lanthipeptide BGCs or the *sprPT/sprH3* gene pair (lanthipeptide: 124/462; *sprPT/sprH3*: 137/462; Table S13). One of these lanthipeptide BGCs showed high homology to the lacticin 3147 BGCs from *Lactococcus lactis*. Lacticin 3147 contains several D-alanine residues as a result of conversion of dehydrated serine residues [209]. While all the genes, including the precursors, were well conserved between the two gene clusters, the *ltnJ* gene had been replaced by an *sprOR* homolog, suggesting that their gene products catalyze similar functions (Figure S10). A recent paper describes a BGC with a gene also encoding a luciferase-like monooxygenase. The product of this BGC contains serine residues that are converted to alanine residues [73], further suggesting that this enzyme applies this modification.

Interestingly, many *sprOR*, *sprF1* and *sprF2* homologs were found not present in either a lanthipeptide BGC or close to the *sprPT/H3* gene pair. These three genes products all require a dehydrated serine or threonine residue to carry out their reaction. The presence of these homologs therefore provides a promising lead for core-dependent genome mining. Assuming the products of the homologs still carry out the same reaction, investigation of these homologs could lead to the discovery of even more lanthipeptide core modifying enzymes.

Conclusions and final perspectives

Most RiPP genome mining strategies expand previously characterized RiPP subclasses. These efforts can lead to novel natural products when new RiPP precursors are identified in conjunction with previously characterized modification machinery. However, the detection of completely novel RiPP subclasses remains a more challenging ordeal, and currently used genome mining tools can only identify these if there are similarities between the known and the novel RiPP subclass.

In this work, we have characterized a candidate novel RiPP subclass, whose BGC was identified with decRiPPter. The product of one of the gene clusters associated with this candidate class was characterized as the first member of a new class of lanthipeptides (termed ‘class V’). BGCs of this class were not detected by any other RiPP genome mining tool. Variants of this gene cluster are widespread across *Streptomyces* species, further expanding one of the best-studied RiPP subclasses. The fact that no less than five different sets of lanthionine-forming enzymes have been reported highlights the importance of this crosslink. Furthermore, this subclass is one of the few RiPP subclasses that has been prioritized purely through the use of bioinformatics, showcasing the potential of these methods for natural product genome mining when properly applied. Since no fewer than 42 different candidate families were discovered in *Streptomyces* alone, the potential of decRiPPter to further expand the list of RiPP subclasses is an exciting prospect.

In addition, two core genes were proposed based on their similarity to genes associated with other RiPP subclasses, which share a common modification. These genes were used to expand the family by finding additional homologs in Actinobacteria and Firmicutes. These homologs could be present in many different genetic contexts, suggesting that a wide variety of new RiPPs and RiPP modifications could be identified among these BGCs. Taken together, this work shows that known RiPP families only cover part of the complete genomic landscape, and that many more RiPP families likely remain to be discovered, especially when expanding the search space to the broader bacterial tree of life.

Materials and Methods

Experimental procedures

Bacterial strain and growth conditions

Streptomyces pristinaespiralis ATCC 25468 was purchased from DSMZ (DSM number 40338). Media components were purchased from Thermo Fisher Scientific, Sigm-Aldrich or Duchefa Biochemie. For strain cultivation on solid media, *Streptomyces* spores were spread on mannitol soya flour agar (SFM; 20 g/L Agar, 20 g/L mannitol, 20 g/L soya flour, supplemented with tap water) prepared as described previously [210], and incubated at 30°C. Spores were harvested after 4-7 days of growth when the strain started to produce a grey pigment, by adding water directly to the plate and releasing the spores with a cotton swab. Spores were centrifuged and stored in 20% glycerol.

For cultivation in liquid media, 20-50 µL of a dense spore stock was inoculated into 100 mL shake flasks with coiled coils containing 20 mL of the medium of interest. For extractions, NMMP was used (0.60 mg/L $MgSO_4$, 5 mg/L NH_4SO_4, 5 g/L Bacto casaminoacids, 1 mL trace elements (1 g/L $ZnSO_4.7H_2O$, 1 g/L $FeSO_4.7H_2O$, 1 g/L $MnCl_2.4H_2O$, 1 g/L $CaCl_2$, anhydrous)), while for genomic DNA isolation, a 1:1 mixture of TSBS: YEME with 0.5% glycine and 5 mM $MgCl_2$ was used (TSBS: 30 g/L Bacto Tryptic Soy Broth, 100 g/L sucrose; YEME: Bacto Yeast Extract: 3 g/L, Bacto Peptone 5 g/L, Bacto Malt Extract 3 g/L, glucose 10 g/L, sucrose 340 g/L).

E. coli strains JM109 and ET8 were used for general cloning purposes and demethylation, respectively. Strains were cultivated in liquid LB and on LB-agar plates at 37°C.

Molecular biology

All materials and primers were purchased from Sigma-Aldrich or Thermo Fisher Scientific unless stated otherwise. Restriction enzymes and T4 ligase were purchased from NEB. Restriction and ligation protocols were followed as per manufacturer's description. For amplification of DNA fragments with PCR, Pfu polymerase was used. Primers were designed with T_m of the annealing region roughly equal to 60°C. Standard PCR protocols consisted of 30 cycles (45 second DNA melting @ 95 °C, 45 second primer annealing @55°C-65°C, 60s-180s primer elongation @ 72°C), but PCR protocols were optimized where necessary.

Deletion mutants were created by replacing the gene cluster or targeted genes with an *aac(3)IV* apramycin resistance cassette via homologous recombination, as described [211]. For the deletion of the entire gene cluster, the -1507/-39 and +135/+1641 regions upstream and downstream of the cluster were amplified by PCR with the spr_LF_F/spr_LF_R and spr_RF_F/spr_RF_R primer pairs (table S1) respectively, and inserted into the pWHM3-oriT vector (Table S2) into the EcoRI/HindIII sites. The *aac(3)IV* apramycin resistance cassette was inserted into the XbaI site, creating the vector pAK3. pAK3 was transformed to *E.coli ET8* for DNA demethylation, purified, and transformed to *S. pristinaespiralis* by protoplast transformation. Transformation mixtures were plated out on R5, prepared as described earlier [210]. After 14-18 hours, the plates were overlaid with 1.2 mL H_2O containing 10 µg thiostrepton and 25 µg apramycin. Three colonies were picked after 4 days of growth and spread onto SFM plates without added antibiotic to allow for homologous recombination. Colonies containing the correct phenotype (apramycin-resistant, thiostrepton-sensitive) were picked and the homologous recombination was confirmed by PCR, using the spr_del_check_F/spr_del_check_R primer pair.

For the deletion of the gene pair *sprH3/sprPT*, the primers sprH3PT_LF_F/ sprH3PT_LF_R and sprH3PT_RF_F/ sprH3PT_RF_R were used to amplify the -1430/-54 and +2/+1483 flanking regions. These resulting fragments were used to create the pAK8 vector, and the mutants were created as above. Confirmation of the mutants was done with PCR using the primer pair sprH3PT_check_F/sprH3PT_check_R.

Removal of the apramycin cassette was done by transforming the pUWLCRE shuttle vector to the mutant strain as described above. Three colonies were picked and grown on SFM without antibiotics. The antibiotic resistance phenotype was monitored by growing the spores on plates with the relevant antibiotics. Strains that were apramycin-resistant, thiostrepton-sensitive were picked as candidate full deletion mutants, which was confirmed by PCR using the same checking primers as used for the apramycin resistance mutants.

Constructs for the overexpression of the *sprR* regulator were constructed as follows: the *sprR* gene was amplified from the genomic DNA of *S. pristinaespiralis* using the sprR_F/sprR_R primer pair, and placed into the EcoRI/XbaI site of the pSET152 vector. The -0/-457 upstream region of glyceraldehyde 3-phosphate dehydrogenase amplified from the genome of *S. coelicolor*, was obtained from previous studies [212, 213] and inserted into the EcoRI site and the engineered NdeI site, placing it directly upstream of the *sprR* gene. To create vector pAK2, the entire region between the EcoRI and XbaI sites was excised and inserted into the pHJL401 vector.

To make the complementation constructs for the *sprH3/PT* deletion strain, we aimed at placing a DNA fragment containing both genes (-0/+0), preceded by a promoter, in the XbaI/HindIII site behind the regulator in the pAK2 vector. An additional terminator sequence was placed between the the *sprR* gene and the amplified fragment, to prevent transcriptional read-through. To this end, a single fragment containing both genes with the preceding 519 bp was amplified with the sprH3PT_compl_F_t0_prom/sprH3PT_R primer pair, and placed in either pHJL401 (creating pAK4) or in the pAK2 vector (creating pAK5). The SprH3PT_compl_F_t0_prom primer contains a t0 terminator sequence. Overexpression constructs with the XNR_3799 promoter were created by amplifying the -695/+3 upstream region of XNR_3799 of *S. lividans* with a preceding t0 terminator sequence from an in-house plasmid on which these two sequences were adjacent, using the XNR_t0_F and XNR_t0_R primers. Using the XbaI/NdeI restriction site, this fragment was placed behind the *sprH3/PT* genes, amplified without their native promoter using the sprH3PT_compl_F/ sprH3PT_compl_R primer pair. The resulting fragment containing the t0 terminator, the XNR3170 promoter and the *sprH3*/*PT* gene pair was placed on pHJL401 (creating pAK6) and on pAK2 (creating pAK7).

Extractions

Strains were cultured in 100 mL shake flasks containing 20 mL NMMP, with coiled coils at 30°C for 7 days. 20 µg/mL thiostrepton was added to cultivate strains containing pHJL401. Mycelium was collected by centrifugation, washed twice with sterile MiliQ water and extracted with 5 mL methanol by shaking overnight at 4°C. The methanol was collected and centrifuged at 4°C to clear it of cellular debris and precipitates. The crude extracts were dried and weighed, and dissolved in methanol at a concentration of 1 mg/mL for further analysis.

Peptide purification

For large-scale extraction, 2L NMMP prepared as above was inoculated with 2.5 mL of a dense spore stock *S. pristinaespiralis* with pAK3, and split over one hundred 100 mL shake flasks. The cultures were grown for 14 days, pooled together and extracted with an equivalent volume of butanol. The butanol extracted was then evaporated *in vacuo* to yield 1.7g of crude extract. The resulting crude extract was adsorbed on silica gel 60 (40–60 µm, Sigma Aldrich), and dry loaded on a VLC column (3 × 30 cm) packed with the same material. The column was eluted with 200 mL fractions of a gradient comprised of (v/v): hexane, hexane–EtAc (1:1), EtAc, EtAc-MeOH (3:1), EtAc-MeOH (1:1), EtAc–MeOH (1:3), and finally MeOH. The fractions containing the compound of interest were pooled, concentrated and further purified using Waters preparative HPLC system comprised of 1525 pump, 2707 autosampler, and 2998 PDA detector. The pooled fraction (112.9 mg) was injected into a SunFire C_{18} column (10 µm, 100 Å, 19 × 150 mm). The column was run at a flow rate of 12.0 mL/min, using solvent A (0.1% FA in H_2O) and solvent B (0.1% FA in ACN), and a gradient of 30–60% B over 20 min. HPLC purification was monitored at 254 nm, and eventually resulted in compound **1** (1.1 mg).

LCMS analysis

LC-MS/MS acquisition was performed using Shimadzu Nexera X2 UHPLC system, with attached PDA, coupled to Shimadzu 9030 QTOF mass spectrometer, equipped with a standard ESI source unit, in which a calibrant delivery system (CDS) is installed. The dry extracts were dissolved in MeOH to a final concentration of 1 mg/mL, and 2 µL were injected into a Waters Acquity Peptide BEH C_{18} column (1.7 µm, 300 Å, 2.1 × 100 mm). The column was maintained at 40 °C, and run at a flow rate of 0.5 mL/min, using 0.1% formic acid in H_2O as solvent A, and 0.1% formic acid in acetonitrile as solvent B. A gradient was employed for chromatographic separation starting at 5% B for 1 min, then 5 – 85% B for 9 min, 85 – 100% B for 1 min, and finally held at 100% B for 4 min. The column was re-equilibrated to 5% B for 3 min before the next run was started. The LC flow was switched to the waste the first 0.5 min, then to the MS for 13.5 min, then back to the waste to the end of the run. The PDA acquisition was performed in the range 200 – 400 nm, at 4.2 Hz, with 1.2 nm slit width. The flow cell was maintained at 40 °C.

The MS system was tuned using standard NaI solution (Shimadzu). The same solution was used to calibrate the system before starting. System suitability was checked by including a standard sample made of 5 µg/mL thiostrepton; which was analyzed regularly in between the batch of samples. All the samples were analyzed in positive polarity, using data dependent acquisition mode. In this regard, full scan MS spectra (*m/z* 400 – 4000, scan rate 20 Hz) were followed by three data dependent MS/MS spectra (*m/z* 400 – 4000, scan rate 20 Hz) for the three most intense ions per scan. The ions were selected when they reach an intensity threshold of 1000, isolated at the tuning file Q1 resolution, fragmented using collision induced dissociation (CID) with collision energy ramp (CE 10 – 40 eV), and excluded for 0.05 s (one MS scan) before being re-selected for fragmentation. The parameters used for the ESI source were: interface voltage 4 kV, interface temperature 300 °C, nebulizing gas flow 3 L/min, and drying gas flow 10 L/min.

LC-MS based comparative metabolomics

All raw data obtained from LC-MS analysis were converted to mzXML centroid files using Shimadzu LabSolutions Postrun Analysis. The converted files were imported and processed MZmine 2.5.3 [214]. Throughout the analysis, *m/z* tolerance was set to 0.002 *m/z* or 10.0 ppm, RT tolerance was set to 0.05 min, noise level was set to 2.0E2 and minimum absolute intensity was set to 5.0E2 unless specified otherwise. Features were detected (polarity: positive, mass detector: centroid) and their chromatograms were built using the ADAP chromatogram builder [215] (minimum group size in number of scans: 10; group intensity threshold: 2.0E2). The detected peaks were smoothed (filter width: 9), and the chromatograms were deconvoluted (algorithm: local minimum search; Chromatographic threshold: 90%; search minimum in RT range: 0.05; minimum relative height: 1%; minimum ratio of peak top/edge: 2; peak duration 0.03 – 3.00 min). The detected peaks were deisotoped (maximum charge: 5; representative isotope: lowest *m/z*). Peak lists from different extracts were aligned (weight for RT = weight for *m/z*; compare isotopic pattern with a minimum score of 50%). Missing peaks detected in at least one of the sample were filled with the gap filling algorithm (RT tolerance: 0.1 min). Among the peaks, we identified fragments (maximum fragment peak height: 50%), adducts ($[M+Na]^+$, $[M+K]^+$, $[M+NH_4]$, maximum relative adduct peak height: 3000%) and complexes (Ionization method: $[M+H]^+$, maximum complex height: 50%). Duplicate peaks were filtered. Artifacts caused by detector ringing were removed (*m/z* tolerance: 1.0 *m/z* or 1000.0 ppm) and the results were filtered down to the retention time of interest. The aligned peaks were exported to a MetaboAnalyst file. From here, peaks were additionally filtered to keep only peaks present in all three replicates, using in-house scripts. The resulting peak list was uploaded to MetaboAnalyst [216], log transformed and normalized with Pareto scaling without prior filtering. Missing values were filled with half of the minimum positive value in the original data. Heatmaps and volcano plots were generated using default parameters.

Mass spectrometry-based quantitative proteomics

20 uL of dense spore stocks were inoculated in NMMP and grown for 7 days as described above. 1 mL samples were taken after 2 and 7 days. Mycelium was gathered by centrifugation and washed with disruption buffer (100 mM Tris-HCl, pH 7.6, 0.1 M dithiothreitol). The samples were sonicated for 5 minutes (in cycles off 5s on, 5s off) to disrupt the cell wall, and centrifuged at max speed for 10 minutes to collect the proteins. Proteins were then precipitated using chloroform-methanol [217]. The dried proteins were dissolved in 0.1% RapiGest SF surfactant (Waters) at 95°C. Protein digestion steps were done according to van Rooden et al [218]. After digestion, formic acid was added for complete degradation and removal of RapiGest SF. Peptide solution containing 8 µg peptide was then cleaned and desalted using the STAGETipping technique [219]. Final peptide concentration was adjusted to 40 ng/µl with 3% acetonitrile, 0.5% formic acid solution. 200 ng of digested peptide was injected and analysed by reverse-phase liquid chromatography on a nanoAcquity UPLC system (Waters) equipped with HSS-T3 C18 1.8 µm, 75 µm X 250 mm column (Waters). A gradient from 1% to 40% acetonitrile in 110 min was applied, [Glu^1]-fibrinopeptide B was used as lock mass compound and sampled every 30 s. Online MS/MS analysis was done using Synapt G2-Si HDMS mass spectrometer (Waters) with an UDMS^E method set up as described [218].

Mass spectrum data were generated using ProteinLynx Global SERVER (PLGS, version 3.0.3), with MS^E processing parameters with charge 2 lock mass 785.8426 Da. Reference protein database was downloaded from GenBank with the accession number GCA_001278075.1. The resulting data were imported to ISOQuant [220] for label-free quantification. TOP3 quantification result from ISOQuant was used when further investigating the data.

Iodoacetamide treatment

Reaction mixtures were prepared based on earlier reported studies [203]. 20 µL reaction mixtures containing 0.25 mg/mL purified peptide, 13 mM TCEP, 25 mM IAA and 250 mM HEPES (pH = 8.0) in H_2O were left at room temperature for 1 hour in the dark. Reaction mixtures were cleaned using the STAGETipping technique [219].

Protein hydrolysis

0.2 mg of purified peptide was dissolved in 3 mL 6M HCl and sealed inside a glass ampule, based on earlier studies[221]. The mixture was heated to 110°C for 24 hours. The HCl was removed by repeated drying and dissolving of the peptide with H_2O. The peptide was afterwards dissolved in 50 µL H_2O and analyzed with LCMS as described above.

NMR

NMR data were recorded on Bruker Ascend 850 NMR spectrometer (Bruker BioSpin GmbH), equipped with a 5 mm cryoprobe. The sample was measured in a 3 mm NMR tube through the use of an adapter. All NMR experiments were performed with suppression of the water peak in the solvent.

Data analysis

Genomic context analysis

CORASON [185] was used with the number of flanking genes set to 15, on the *Streptomyces* genomes analyzed with the query of interest. Results were parsed using in-house scripts and compared to decRiPPter output. NCBI BLAST was used to find additional homologs of genes of interest within the clusters, with a cutoff of 30 percent ID similarity.

Supplementary information for Chapter 4

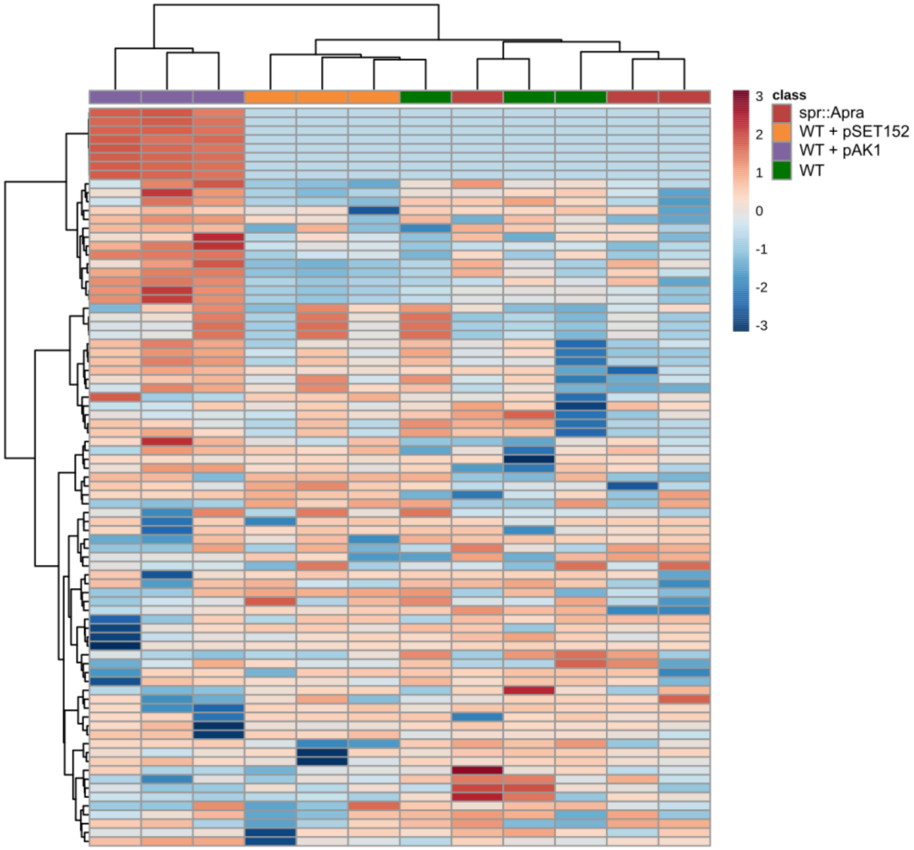


Figure S1. Heatmap of extracted peaks reveals seven peaks that are uniquely observed in strains containing the expression construct pAK1. Each row represents a single mass feature and each column represents a single extract, while the colour scale indicates the $\log_{10}$-scaled intensity of the mass features for each extract.

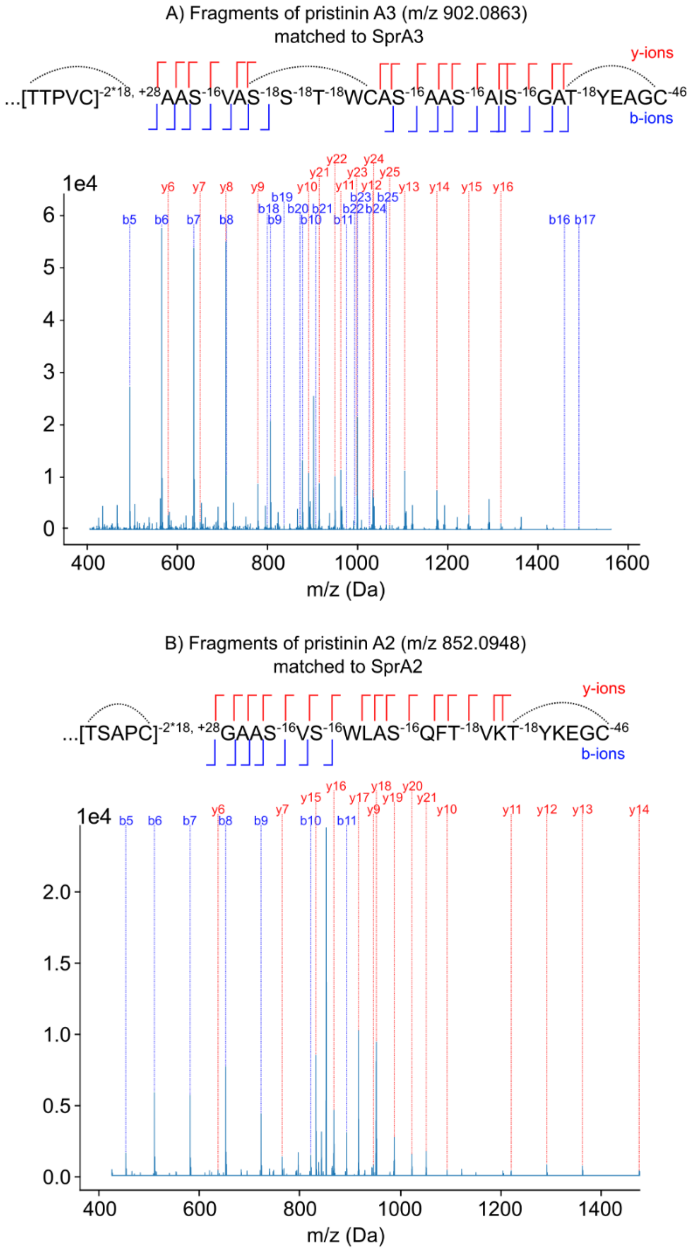


Figure S2. Fragmentation patterns of two highly extracted peaks can be matched to the SprA2 and SprA3 precursors. A full list of the modifications applied can be found in Table S6 and Table S7.

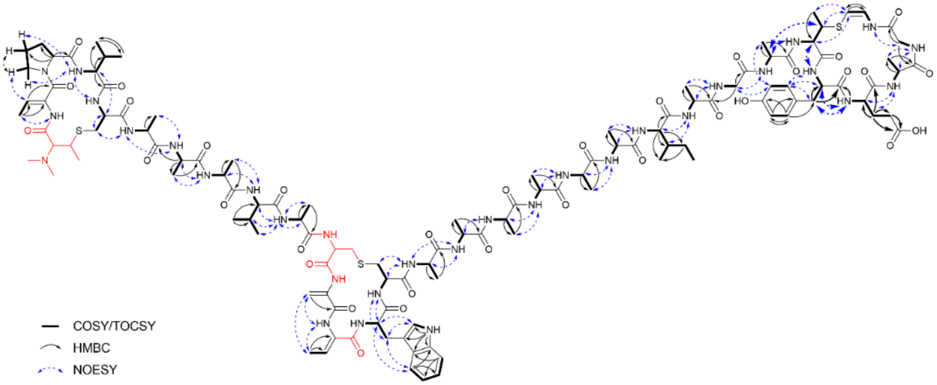
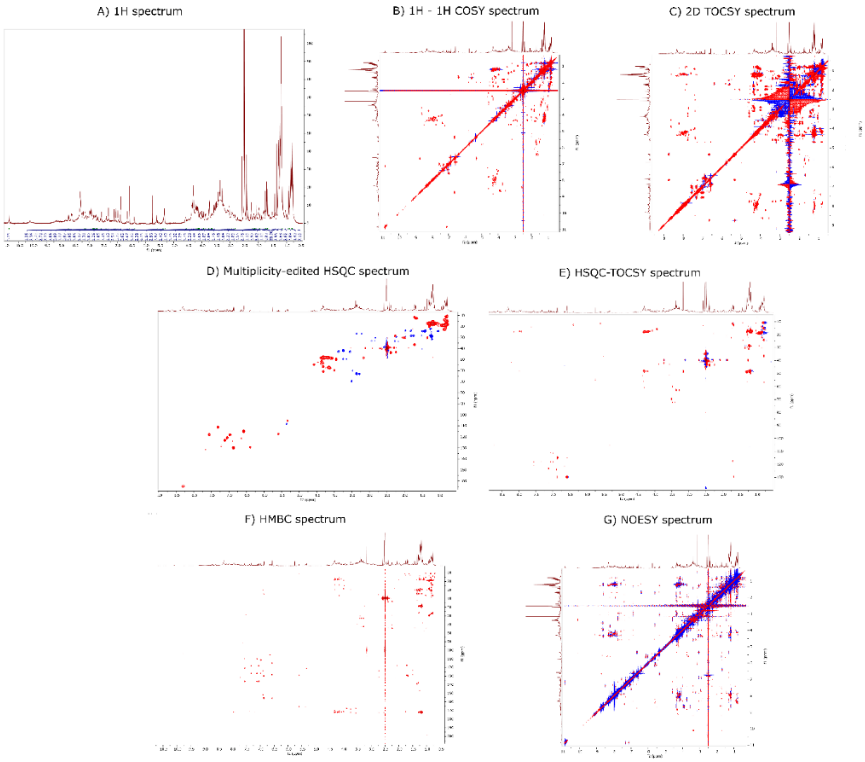


Figure S3. Key 2D NMR correlations observed for pristinin A3 (1). No clear correlations could be observed for the red parts of the structure, which were confirmed through other techniques. Bold arrows are for correlations which were better observed in CD_3CN:H_2O 9:1.


A) 1H spectrum

B) 1H - 1H COSY spectrum

C) 2D TOCSY spectrum

D) Multiplicity-edited HSQC spectrum

E) HSQC-TOCSY spectrum

F) HMBC spectrum

G) NOESY spectrum

Figure S4. NMR spectra of pristinin A3 (850 MHz, in DMSO-*d6*, 298K). A) ^1H NMR spectrum with water suppression. The peak at 3.17 ppm is due to traces of methanol in the sample. B) ^1H–^1H COSY spectrum. C) 2D TOCSY spectrum. D) Multiplicity-edited HSQC spectrum. E) HSQC-TOCSY spectrum. F) HMBC spectrum. G) NOESY spectrum. Full data available on request.

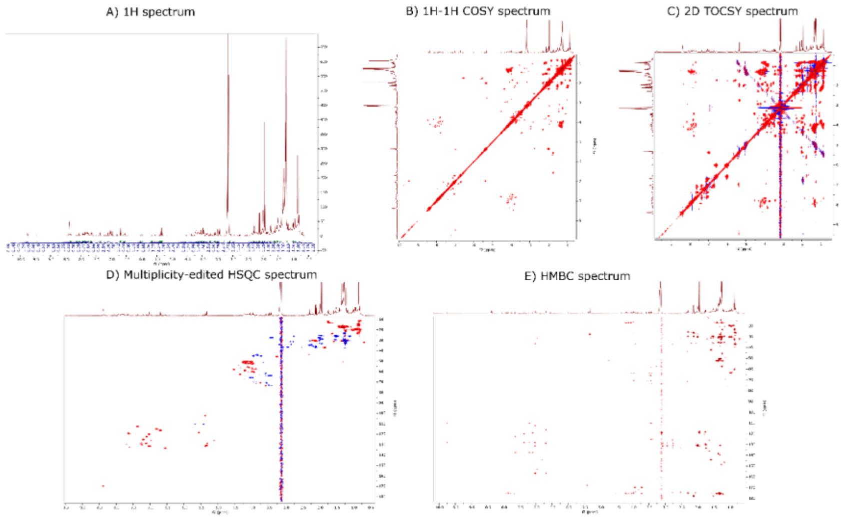


Figure S5. NMR spectra of pristinin A3 (850 MHz, in CD_3CN:H_2O 9:1, 297 K, first run). A) ^{1}H NMR spectrum with water suppression. B) ^{1}H–^{1}H COSY spectrum. C) 2D TOCSY spectrum. D) Multiplicity-edited HSQC spectrum. E) HMBC spectrum. Full data available on request.

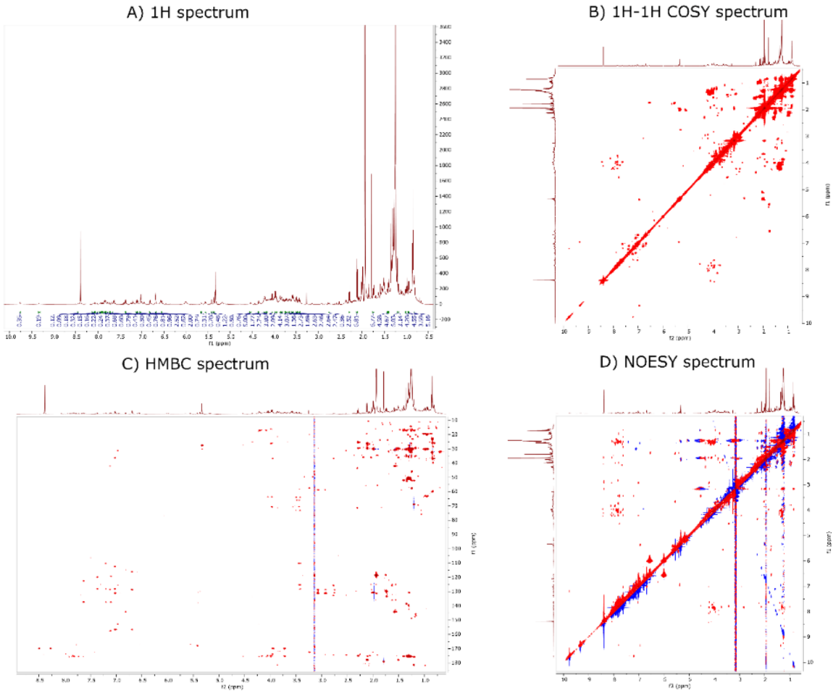


Figure S6. NMR spectra of pristinin A3 (850 MHz, in CD_3CN:H_2O 9:1, 297 K, second run). A) ^{1}H NMR spectrum. B) ^{1}H–^{1}H COSY spectrum. C) NOESY spectrum. D) HMBC spectrum. Full data available on request.

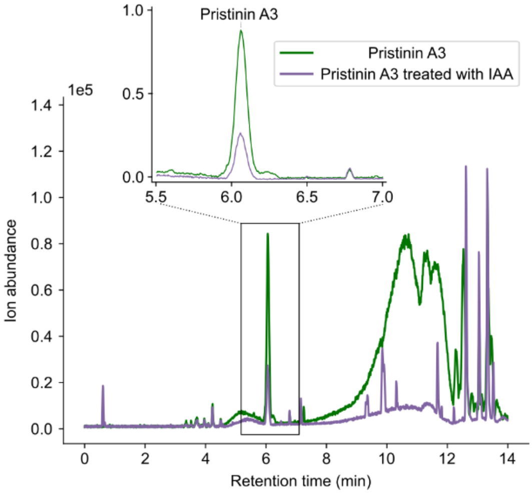


Figure S7. Labeling experiments with iodoacetamide (IAA) provide further support for the proposed structure of pristinin A3. (Purple) IAA covalently attaches to free sulfur groups of cysteines. However, the SprA3 peak was unaltered by IAA treatment, despite the presence of three cysteines in the peptide, strongly suggesting that these cysteines are not free.

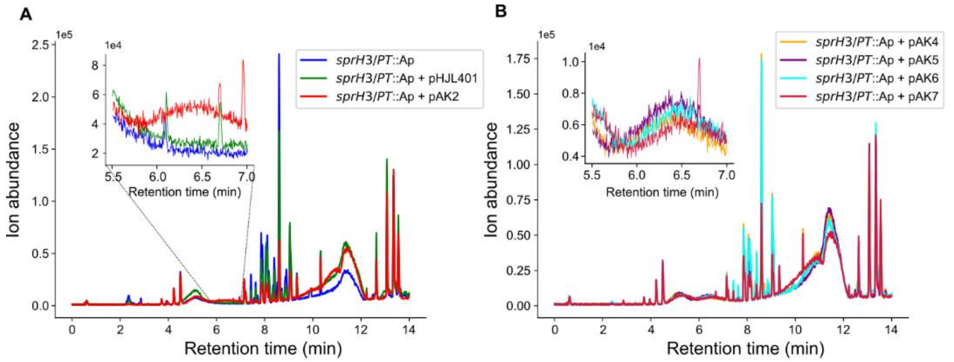


Figure S8. A mutant strain in which *sprPT* and *sprH3* are deleted no longer produces the *spr* RiPPs, but the production is not restored by complementation. A) LCMS analysis of crude extracts made of *sprH3PT*::Ap, with pHJL401, pAK2 or no vector. Lacking the *sprH3/PT* gene pair, the strain no longer produces the previously identified RiPPs. B) Complementation of the *sprH3PT* genes does not restore RiPP production. Whether its own native promoter was used (pAK4, pAK6) or the strong XNR_3799 promoter (pAK5, pAK7), even in combination with the *sprR* gene behind the gap promoter (pAK6, pAK7), the masses corresponding to the RiPPs were no longer detected.

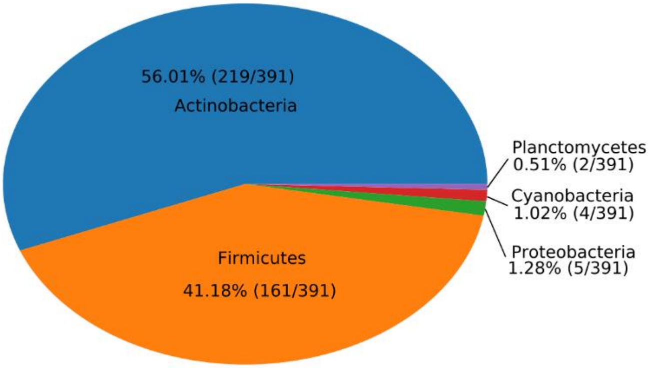


Figure S9. Homologs of the sprPT and sprH3 gene pair are present outside Streptomyces. Most homologs were found in Actinobacteria and Firmicutes, although a few additional candidates were found in Proteobacteria, Cyanobacteria and Planctomycetes.

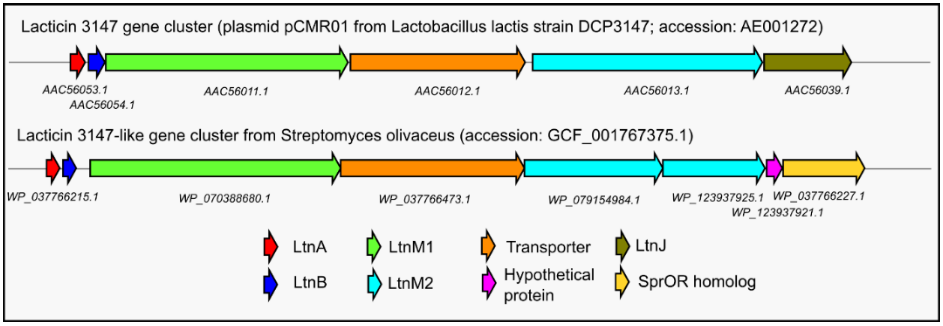


Figure S10. Comparison of the lacticin 3147-like gene cluster from Lactococcus lactis with a homologous cluster from Streptomyces olivaceus. Genes encoding both precursors, both LanM-like modifying enzymes and the transporter are well conserved between the clusters. The gene encoding LtnJ, however, responsible for the reduction in the conversion to alanine and butyric acid, was not conserved. Instead, a homolog to sprOR was found, suggesting it may carry out a similar function.

```
prod_559746     1 ---MHTM-TETDLLSGYTAYTTAEELDQFDGKAAPAATTPVLAPILI-----RASIIAARSSQQCA----AGIAAAGGGIWRTIRKVC
prod_4312120    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGATAGPAFSPTV-PWAI-----QATVISARSSQACA----AALGSLAA---KTVEKKC
prod_4312121    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGATAGPAFSPTV-PWAI-----QATVISARSSQACA----AALGSLAA---KTVEKKC
prod_9638834    1 ---MQNV-TEQDLFDGYTAYTSAEELGLHDGKDAAPAFSPTI-PWAI-----RATIISARSSQQCA----AALGSLAA---KTVENKC
prod_1888002    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_1892473    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_1898975    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_2702012    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_4125916    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_4204099    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_5620390    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_5701534    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_5937191    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_6249001    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_6819619    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_710895     1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_7443641    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_7703323    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_8242019    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_8466597    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_8698113    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_8721923    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_8902069    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_9724047    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGQEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_1317692    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGKEAAPAFSPTI-PWAI-----RATIISARSSQQCA----AALGSLAA---KTVENKC
prod_3048582    1 ---MQNV-TEKDLFDGYTAYTSAEELGLHDGKEAAPAFSPTI-PWAI-----RATIITARSSQQCA----AALGSLAA---KTVENKC
prod_398364     1 ---MQSTQNEKDLFEGYTAYTSAEELGLYDGKDAAPAFSPTI-PWAI-----RATIITARSSQQCA----AAIGSLTA---KTIENKC
prod_7467458    1 ---MQSTQNEKDLFEGYTAYTSAEELGLYDGKDAAPAFSPTI-PWAI-----RATIITARSSQQCA----AAIGSLTA---KTIENKC
prod_5042396    1 ---MQNV-NEKDLFDGYTAYTSAEELGLYDGKDAAPAFSPTI-PWAI-----RAGLITARSSQQCA----AAIGSFTA---RTIESKC
prod_1644796    1 ---MNAS---AHLIAGYTAYTTAAEFDA-SITADAPAVTPAT-P--------SIALSIAESSYACG----AGVGASIG---ITFTKGC
prod_7595003    1 ---MNAS---AHLIAGYTAYTTAAEFDA-SITADAPAVTPAT-P--------SIALSIAESSYACG----AGVGASIG---ITFTKGC
prod_9224211    1 ---VNTT---ENLIAGYTAYTSAQEIEA-THAEEAPGATPSV---------LSFIATSGWA---CG----AGIGTSIG---VTAAKGC
prod_4694754    1 ---VNTT---DTLLAGYAAYTSADEIAA-AQDGGAPEISPVS----L-----SIAVSIAESSYACS----AGLSMSVG---VTVGKGC
prod_4694755    1 ---VNTT---DTLLAGYAAYTSADEIAA-AQDGGAPEISPVS----L-----SIAVSIAESSYACS----AGLSMSVG---VTVGKGC
prod_7200544    1 ---MNTS---DNLMAGYATYTSADEIAA-TLDGGAPEISPVS----L-----SIAVSITESSYACG----AGISLSVG---WTVGKGC
prod_9224208    1 ---MNTA---DQLMAGYAVYTTSDEIGA-GAAADAPAISPVS---IF-----SAASSVECAIFSAG----VVTSASAG---GTVAGNC
prod_4694758    1 ---MNTA---DQLIAGYTAYTDSAEIAA-DATAEAPAISPTT----------TITIVSVESVLASI----GASASFSAG--YTVSSGC
prod_7200547    1 ---MNNT---DQLIAGYTAYTDSAEIAA-DASAEAPAITPTT----------TITIVSVESAVFSI----GGAASFSAG--YTISSGC
prod_326225     1 --MSHDQNTLEELVTGYESYADADEIEV-DAVTGAPATTPFCGA--------AASFMLSYV---------------------TTNGPG
prod_326226     1 --MTNDQSTLEDLVTGYESYADADEIEV-DAVTGAPATTPFCGA--------VASFALSYV---------------------TTNGPG
prod_6174086    1 ---VKTQ----DLIAGYAAYVDVAELNV-SAASEAPATSPVCFAAATSSAACLAATSSGWCVAGAG----AGVGGGIA---QSVKHGC
prod_6174087    1 ---M----ELDEMISGYDTYVDVAELDV-SAQSEAPATSPTC-----------FIASVGLS--------------------YQITKDL
prod_9167739    1 ---MQNDIEIMELVGGFEAYTEAAELNM-EASVEAPAATPTA------------TIVYTKFS----------VASVT----LTAKKGC
prod_2743547    1 VQKNDTV-DIMELVGGFEAYAEAAELNF-EASADAPAITPTL-----------TTIAYTKVS----------VASVSA----SIKVGC
prod_5868070    1 VQNIENV-EIMELVGGFEAYAQAAELNF-EASADAPAITPTL-----------TTIAYTKVTVAGT----A----------ASIKWTC
prod_1221493    1 ---MDTH----ELIEGFDAYVEAEELNE-DAMVDAPATTVPC-----------------------------TVASF-----ATGYFSC
prod_3289459    1 ---MDTH----ELIEGFDAYVEAEELNE-DAMVDAPATTVPC-----------------------------TVASF-----ATGYFSC
prod_467494     1 ---MDTH----ELIEGFDAYVEAEELNE-DAMVDAPATTVPC-----------------------------TVASF-----ATGYFSC
prod_1100745    1 ---MEKATSIVELLSGYEAYSSVEEINL-SAASDAPATTWGCAA---------VSASISWM-----------SGQVVS---KTVDDGC
prod_2725163    1 ---MEKATSIVELLSGYEAYSSVEEINL-SAASDAPATTWGCAA---------VSASISWM-----------SGQVVS---KTVDDGC
prod_3616888    1 ---MEKATSIVELLSGYEAYSSVEEINL-SAASDAPATTWGCAA---------VSASISWM-----------SGQVVS---KTVDDGC
prod_5244387    1 ---MEKATSIVELLSGYEAYSSVEEINL-SAASDAPATTWGCAA---------VSASISWM-----------SGQVVS---KTVDDGC
prod_6023772    1 ---MEKATSIVELLSGYEAYSSVEEINL-SAASDAPATTWGCAA---------VSASISWM-----------SGQVVS---KTVDDGC
prod_6473304    1 ---MEKATSIVELLSGYEAYSSVEEINL-SAASDAPATTWGCAA---------VSASISWM-----------SGQVVS---KTVDDGC
prod_6857183    1 ---MEKATSIVELLSGYEAYSSVEEINL-SAASDAPATTWGCAA---------VSASISWM-----------SGQVVS---KTVDDGC
prod_8409183    1 ---MEKATSIVELLSGYEAYSSVEEINL-SAASDAPATTWGCAA---------VSASISWM-----------SGQVVS---KTVDDGC
prod_9246364    1 ---MEKATSIVELLSGYEAYSSVEEINL-SAASDAPATTWGCAA---------VSASISWM-----------SGQVVS---KTVDDGC
prod_9674514    1 ---MEKATSIVELLSGYEAYSSVEEINL-SAASDAPATTWGCAA---------VSASISWM-----------SGQVVS---KTVDDGC
prod_1949441    1 ---MEKATSIVELLSGYEAYSSAEEINL-SAATDAPATTWGCAA---------VSASVSWM-----------SGQVVS---KTVDDGC
prod_5478099    1 ---MEKATSIVELLSGYEAYSSAEEINL-SAATDAPATTWGCAA---------VSASVSWM-----------SGQVVS---KTVDDGC
prod_326224     1 ---MDNA-SMMDLVAGYNTYAEASELGI-QAVADAPATTPVCAATIA-----ASAVSSGWC---AS----AAASAAGG---ATYKLGC
sprA3           1 ---MQNNTEIMDLIANYDAYADVDELNV-TAAADAPATTPVCAA---------SVASSTW----CA----SAASAISG---ATYEAGC
prod_8036387    1 --MSNKSTVIADLVAGYDAYTEVDELNV-SAAAGAPATTWVC----------VSVVASRASSVKCGAWASAGASAVSG---ATYEITC
prod_2805062    1 --MDNKSTVITDLVAGYSTYTEAGELNV-SAAAGAPATTYIC----------ASVAISRVSSPRCA----ASASAVSG---ATYEWTC
prod_297319     1 --MDNKSTVITDLVAGYSTYTEAGELNV-SAAAGAPATTYIC----------ASVAISRVSSPRCA----ASASAVSG---ATYEWTC
prod_5421071    1 --MDNKSTVITDLVAGYSTYTEAGELNV-SAAAGAPATTYIC----------ASVAISRVSSPRCA----ASASAVSG---ATYEWTC
prod_5527162    1 --MDNKSTVITDLVAGYSTYTEAGELNV-SAAAGAPATTYIC----------ASVAISRVSSPRCA----ASASAVSG---ATYEWTC
prod_6582107    1 --MDNKSTVITDLVAGYSTYTEAGELNV-SAAAGAPATTYIC----------ASVAISRVSSPRCA----ASASAVSG---ATYEWTC
prod_8403914    1 --MDNKSTVITDLVAGYSTYTEAGELNV-SAAAGAPATTYIC----------ASVAISRVSSPRCA----ASASAVSG---ATYEWTC
prod_9151868    1 --MDNKSAVITDLVAGYSTYTEAGELNV-SAAAGAPATTYIC----------ASVAISRTSSVKCS----AAASAISG---ATYEWTC
prod_9381790    1 --MDNKSAVITDLVAGYSTYTEAGELNV-SAAAGAPATTYIC----------ASVAISRTSSVKCS----AAASAISG---ATYEWTC
prod_8036386    1 ---MKTT-TIMELAAGYDAYTGAEELEV-GATAEAPASTPLCAA--------AASAGVSWM-----------ASQFSA---RTISGGC
prod_9151867    1 ---MKTT-AIMELVAGYEVYADSAELQV-DATVNAPASTPAC-----------GAATVSWI-----------VSQFSA---KTVKDGC
prod_9381789    1 ---MKTT-AIMELVAGYEVYADSAELQV-DATVNAPASTPAC-----------GAATVSWI-----------VSQFSA---KTVKDGC
prod_2805061    1 ---MKTT-AIMELVAGYEVYADSAELQV-DAAVDAPASTPAC-----------AAATLSWI-----------VSQFSG---KTVKDGC
prod_297320     1 ---MKTT-AIMELVAGYEVYADSAELQV-DAAVDAPASTPAC-----------AAATLSWI-----------VSQFSG---KTVKDGC
prod_5421070    1 ---MKTT-AIMELVAGYEVYADSAELQV-DAAVDAPASTPAC-----------AAATLSWI-----------VSQFSG---KTVKDGC
prod_5527161    1 ---MKTT-AIMELVAGYEVYADSAELQV-DAAVDAPASTPAC-----------AAATLSWI-----------VSQFSG---KTVKDGC
prod_6582106    1 ---MKTT-AIMELVAGYEVYADSAELQV-DAAVDAPASTPAC-----------AAATLSWI-----------VSQFSG---KTVKDGC
prod_8403913    1 ---MKTT-AIMELVAGYEVYADSAELQV-DATVDAPASTPAC-----------AAATLSWI-----------VSQFSG---KTVKDGC
sprA1           1 MADLQQTGSISELVAGYDTYSEAGELVA-EAAADAPASTPTC-----------AAATISWL-----------GSQLTV---KTYKEGC
sprA2           1 ---MDKTGAITELIEGYDSYSDAEELNS-TAAAEAPATSAPC-----------GAASVSWL-----------ASQFTV---KTYKEGC
```

Text S1. Alignment of precursors belonging to the characterized family of type V lanthipeptides. Precursors were aligned with MUSCLE [153] and visualized with BoxShade.

Table S1. Primers used in this study.

Primer name	Primer sequence
sprR_F	gatc GAATTC CAT ATGACCGTCAACGACCTGTCC
sprR_R	gatc TCTAGA CGCGGCCCACGGATCAGACC
spr_LF_F	gatc GAATTC CTCGCGGCCCTCGGCATTCTGG
spr_LF_R	gcta TCTAGA GTGGCGTGCGCGGCGTTGG
spr_RF_F	gcta TCTAGA CGCCCGGAAACAGGCATGAAGG
spr_RF_R	gcta AAGCTT ATGTCGCGGTGGACGACACCC
spr_del_check_F	GGGCTACATGCCTACTTTGC
spr_del_check_R	GTGCCCTCTGATTCCTTTCC
sprH3PT_LF_F	gcta GAATTC CCTCCTTGCGGAAGGCAGC
sprH3PT_LF_R	gcta TCTAGA CCTTACGACGGCTGAGGCGG
sprH3PT_RF_F	gcta TCTAGA CGTCCCGAAGCGCTCTGAC
sprH3PT_RF_R	gcta AAGCTT GCTTTCTTCCTCGTCATCGGCGG
sprH3PT_check_F	gcta TCTAGA TTCCTCGTTCGCGCTTTCTCCG
sprH3PT_check_R	gcta TCTAGA CCTTACGACGGCTGAGGCGG
sprH3PT_compl_F_t0_prom	gatc TCTAGA TTGTTCAGAACGCTCGGTCTTGCACACCGGGCG TTTTTTCTTTGTGAGTCCA GGGTGCCCTCTGATTCCTTTCCG
sprH3PT_compl_F	gatc TCTAGA CAT ATGACAGTGATGCTGGAGGCCACG
sprH3PT_compl_R	gatc AAGCTT TCAGCGGCGAGGCAGATTCC
XNR_t0_F	gcta TCTAGA TTGTTCAGAACGCTCGGTCTTGC
XNR_t0_R	gcta AAGCTT gatc CATATGCCGACCTCCCCCTTCG

Table S2. Plasmids used in this study.

Plasmid	Description	Reference
pSET152	Integrative *E. coli* / *Streptomyces* shuttle vector.	Bierman *et al.* [199]
pHJL401	*E. coli/Streptomyces* shuttle vector with intermediate copy number.	Larson *et al*. [201]
pWHM3	Unstable *E. coli/Streptomyces* shuttle vector with high copy number; used for homologous recombination.	Vara *et al*. [198]
pUWLCRE	Unstable *E. coli/Streptomyces* shuttle vector containing the Cre recombinase enzyme, behind a constitutive promoter.	Fedoryshyn *et al*. [200]
pAK1	pSET152 containing *sprR* behind GAPDH promoter from *S. coelicolor* (SCO1947).	This work.
pAK2	pHJL401 containing *sprR* behind GAPDH promoter from *S. coelicolor* (SCO1947).	This work.
pAK3	pWHM3 containing regions flanking the *spr* gene cluster.	This work.
pAK4	pHJL401 containing the *sprH3/PT* gene pair with native promoter.	This work.
pAK5	pHJL401 containing *sprR* behind GAP promoter from *S. coelicolor*, a t0 terminator, the *sprH3/PT* gene pair with their native promoter.	This work.
pAK6	pHJL401 containing the *sprH3/PT* gene pair behind the XNR_3170 promoter.	This work.
pAK7	pHJL401 containing *sprR* behind GAP promoter from *S. coelicolor*, a t0 terminator, the XNR_3170 promoter and the *sprH3/PT* gene pair.	This work.
pAK8	pWHM3 containing regions flanking the *sprH3/PT* gene pair.	This work.

Table S3. Proteins containing a flavoprotein domain (PF02441) are present in both RiPP and non-RiPP BGCs. While proteins with this domain are known in RiPP biosynthesis for the decarboxylation of C-terminal cysteines, their presence is not restricted to RiPP BGCs.

MIBiG BGC ID	BGC class	RiPP class (if applicable)	Protein accession
BGC0000157	Polyketide		ABI94381.1
BGC0000158	Polyketide		ABV91288.1
BGC0000171	Polyketide		CCC21124.1
BGC0000203	Polyketide		ADI71473.1
BGC0000203	Polyketide		ADI71437.1
BGC0000373	NRP		EFG10345.1
BGC0000807	Saccharide		ADD45285.1
BGC0000889	Other		BAM73626.1
BGC0000932	Other		AFO93363.1
BGC0001115	NRP/Polyketide		CBK62752.1
BGC0001193	NRP		AJI44175.1
BGC0001362	Other		AFO93363.1
BGC0001592	Other		AVI10267.1
BGC0000508	RiPP	Lanthipeptide	CAA44255.1
BGC0000514	RiPP	Lanthipeptide	ABC94905.1
BGC0000527	RiPP	Lanthipeptide	CAB60260.1
BGC0000529	RiPP	Lanthipeptide	ADK32557.1
BGC0000530	RiPP	Lanthipeptide	EMC15126.1
BGC0000531	RiPP	Lanthipeptide	AAG48568.1
BGC0000533	RiPP	Lanthipeptide	AAD56146.1
BGC0001618	RiPP	Lanthipeptide	ARD24448.1
BGC0001669	RiPP	Lanthipeptide	AVH76813.1
BGC0000582	RiPP	Linaridin	ADR72965.1
BGC0000583	RiPP	Linaridin	YP_001827875.1
BGC0000625	RiPP	Thioamide-containing peptide	BAN83921.1
BGC0001802	RiPP	Thioamide-containing peptide	ATJ00796.1
BGC0001803	RiPP	Thioamide-containing peptide	BAN83921.1
BGC0001696	RiPP	Thioamide-containing peptide	BBC15202.1

Table S4. Peaks unique to strains containing pAK1 appear to be mostly derived from a single mass. Charges were predicted from isotope patterns, and monoisotopic masses were calculated on assuming M+H ions.

Peak *m/z*	Predicted charge	Monoisotopic mass	Description
707.3534	1	706.3454	Fragment of 2703.2349
868.0891	3	2601.2433	
902.0863	3	2703.2349	
903.4186	3	2707.2318	2703.2349 + 4 Da (2*H_2)
907.4167	3	2719.2261	2703.2349 + 16 Da (O)
908.7487	3	2723.2221	2703.2349 + 20 Da (O + 2*H_2)
914.9003	NA	NA	
918.0807	3	2751.2181	2703.2349 + 48 Da (3*O)

Table S5. Many detected masses from strains containing the expression construct pAK2 appear to be derived from two masses. The two base masses were also the most abundant, making it likely these form final products, while the other masses may be incompletely processed products.

Description	Calculated *m/z* M+3H (Da)	Observed *m/z* pAK2	Δppm
Most abundant mass #1	902.088	902.085	3.3
+ oxygen	907.42	907.417	2.8
	907.42	907.417	2.4
+2 oxygen	912.751	912.748	3.6
	912.751	912.75	1.3
+3 oxygen	918.083	918.08	2.6
	918.083	918.081	1.6
	918.083	918.082	1.2
+4 oxygen	923.414	923.412	3.1
+ methyl	906.76	906.757	3.5
- methyl	897.416	897.415	0.8
- 2 methyl	892.744		
Most abundant mass #2	852.097	852.095	2.3
+ oxygen	852.376	857.426	0.9
+2 oxygen	862.758	862.758	0
+3 oxygen	868.09	868.09	0.6
	868.09	868.089	0.7
- 2 methyl	842.751	842.747	4.2

Table S6. Observed masses for fragments of a mass of 2703.235 Da can be matched to the SprA3 precursor. See also Figure S2A.

Ion	Calculated *m/z*	Calculated *m/z* (z = 2)	Observed *m/z*	Δppm	Ion	Calculated *m/z*	Calculated *m/z* (z = 2)	Observed *m/z*	Δppm
b1	112.0762				y1	76.0221			
b2	195.1134				y2	133.0436			
b3	292.1661				y3	204.0807			
b4	391.2345				y4	333.1233			
b5	494.2437		494.2437	0.05	y5	496.1866			
b6	565.2808		565.2808	0.06	y6	579.2237		579.224	0.49
b7	636.3179		636.3185	0.87	y7	650.2608		650.2596	1.88
b8	707.3551		707.3555	0.61	y8	707.2823		707.2825	0.30
b9	806.4235		806.4248	1.64	y9	778.3194		778.322	3.33
b10	877.4606		877.4608	0.25	y10	891.4035		891.4061	2.96
b11	946.4821		946.4810	1.11	y11	962.4406		962.4424	1.90
b12	1015.5035				y12	1033.4777		1033.4777	0.00
b13	1098.5406	549.7742			y13	1104.5148		1104.5161	1.17
b14	1284.6200	642.8139			y14	1175.5519		1175.5535	1.35
b15	1387.6291	694.3185			y15	1246.5890	623.7984	1246.5909	1.50
b16	1458.6663	729.8370	1458.6639	1.61	y16	1317.6261	659.3170	1317.6302	3.09
b17	1529.7034	765.3556	1529.7031	0.17	y17	1420.6353	710.8216		
b18	1600.7405	800.8742	800.8757	1.94	y18	1606.7146	803.8612		
b19	1671.7776	836.3927	836.3925	0.25	y19	1689.7518	845.3798		
b20	1742.8147	871.9113	871.9155	4.86	y20	1758.7732	879.8905		
b21	1813.8518	907.4298	907.4316	1.96	y21	1827.7947	914.4013	914.4006	0.72
b22	1926.9359	963.9719	963.9677	4.31	y22	1898.8318	949.9198	949.9199	0.09
b23	1997.9730	999.4904	999.4840	6.41	y23	1997.9002	999.4540	999.4547	0.68
b24	2054.9945	1028.0011	1028.0005	0.62	y24	2068.9373	1034.9726	1034.9719	0.66
b25	2126.0316	1063.5197	1063.5239	3.95	y25	2139.9744	1070.4911	1070.4915	0.34
b26	2209.0687	1105.0383			y26	2211.0116	1106.0097		
b27	2372.1320	1186.5699			y27	2314.0207	1157.5143		
b28	2501.1746	1251.0912			y28	2413.0892	1207.0485		
b29+1	2572.2117	1286.6098			y29	2510.1419	1255.5749		
b30+1	2629.2332	1315.1205			y30	2593.1790	1297.0934		
b31+1	2686.2369	1343.6224			y31	2704.2475	1352.6276		

Table S7. Observed masses for fragments of a peak corresponding to a monoisotopic mass of 2553.260 Da can be matched to the SprA2 precursor. See also Figure S2B.

Ion	Calculated *m/z* (*z*=1)	Observed *m/z*	Δppm	Ion	Calculated *m/z* (*z*=1)	Calculated *m/z* (*z*=2)	Observed *m/z*	Δppm
b1	112.0771			y1	76.02193			
b2	181.0986			y2	133.0434			
b3	252.1357			y3	262.086			
b4	349.1885			y4	390.1809			
b5	452.1977	452.1976	0.1	y5	553.2443			
b6	509.2191	509.219	0.2	y6	636.2814		636.2841	4.3
b7	580.2562	580.2552	1.8	y7	764.3764		764.3755	1.1
b8	651.2933	651.2933	0.0	y8	863.4448			
b9	722.3304	722.3316	1.6	y9	946.4819		946.4829	1.1
b10	821.3989	821.3968	2.5	y10	1093.55		1093.554	3.5
b11	892.436	892.4338	2.4	y11	1221.609		1221.606	2.1
b12	1078.515			y12	1292.646		1292.651	3.9
b13	1191.599			y13	1363.683		1363.672	8.1
b14	1262.636			y14	1476.767		1476.769	1.0
b15	1333.674			y15	1662.846	831.9272	831.9276	0.4
b16	1461.732			y16	1733.884	867.4458	867.4469	1.3
b17	1608.801			y17	1832.952	916.98	916.9802	0.2
b18	1691.838			y18	1903.989	952.4986	952.4948	3.9
b19	1790.906			y19	1975.026	988.0171	988.0172	0.1
b20	1919.001			y20	2046.063	1023.536	1023.54	4.0
b21	2002.038			y21	2103.085	1052.046	1052.043	3.6
b22	2165.101			y22	2206.094			
b23	2293.196			y23	2303.147			
b24	2422.239			y24	2374.184			
b25	2479.26			y25	2443.205			

Table S8. Cysteines linked to serine and threonine residues are detected after acidic hydrolysis of pristinin A3. Most of the predicted masses of the amino acids can be detected by HPLC-MS, including the cysteines linked to dehydrated serine and threonine residues.

Amino acid	Calculated *m/z* ($M+H^+$)	Observed *m/z*	Δppm
Glycine	76.04	NA	NA
Serine $^{-18}$	88.04	NA	NA
Alanine/Serine^{-16}	90.056	90.055	13
Threonine $^{-18}$	102.056	NA	NA
Proline	116.072	116.071	8.9
Valine	118.087	118.086	9.3
Isoleucine	132.103	132.102	7.9
Glutamate	148.062	148.06	7.1
Decarboxylated cysteine – threonine	177.07	NA	NA
Tyrosine	182.082	182.081	5.5
Tryptophan	205.098	NA	NA
Cysteine – Serine	209.06	209.059	2
Cysteine – Threonine (twice methylated)	251.108	251.106	5.5

Table S9. ^{1}H and ^{13}C NMR data for pristinin A3 (DMSO-d_6, 850 MHz, 298 K).

Residue	Position	δ_C, type	δ_H, mult.[a] (J in Hz)	Residue	Position	δ_C, type	δ_H, mult.[a] (J in Hz)
Dhb-2	α	129.1, C		**Dha-12**	α	ND	
	β	117.4, CH	5.59, q (7.2)		β	108.2, CH_2	5.36, d (16.1)
	γ	11.5, CH_3	1.71, d (7.2)		CO	166.0, C	
	CO	165.5, C			NH		ND
	NH		9.20	**Dhb-13**	α	129.9, C	
Pro-3	α	60.7, CH	4.31		β	129.4, CH	6.39, q (6.9)
	β	29.4, CH_2	2.23		γ	12.5, CH_3	1.68, d (6.9)
	γ	23.8, CH_2	a: 1.97		CO	ND	
			b: 1.82				
	δ	49.5, CH_2	a: 3.75		NH		10.1
			b: 3.65				
	CO	172.2, C		**Trp-14**	α	54.4, CH	4.58
Val-4	α	59.7, CH	4.00, t (8.2)		β	26.4, CH_2	a: 3.37
							b: 3.22
	β	28.2, CH	2.26		1 (indole)		10.84, br s
	γ	18.9, CH_3	0.94, d (6.9)		2 (indole)	123.1, CH	7.12, br s
	γ′	19.1, CH_3	0.90, d (6.9)		3 (indole)	109.9, C	
	CO	171.0, C			3a (indole)	126.9, C	
	NH		7.40		4 (indole)	118.0, CH	7.56, d (8.0)
Ala(S)-5	α	53.8, CH	4.36		5 (indole)	118.2, CH	6.98, dd (8.0, 7.6)
	β	33.1[b], CH_2	a: 3.08		6 (indole)	120.7, CH	7.05, dd (8.3, 7.6)
			b: 2.84				
	CO	ND			7 (indole)	111.2, CH	7.32, d (8.3)
	NH		7.89		7a (indole)	136.0, C	
Ala-6	α	48.1, CH	4.4.29		CO	ND	
	β	17.5	1.19		NH		7.76
	CO	171.6		**Ala(S)-15**	α	48.1, CH	4.30
	NH		8.13		β	33.1[b], CH_2	2.94
Ala-7	α	48.3, CH	4.17		CO	ND	
	β	17.6, CH_3	1.22		NH		7.70
	CO	171.8		**Ala-16**	α	48.8, CH	4.07
	NH		7.78		β	17.4, CH_3	1.14, d (7.2)
Ala-8	α	48.1, CH	4.27		CO	171.8, C	
	β	17.6, CH_3	1.21		NH		7.60
	CO	172.1		**Ala-17**	α	48.3, CH	4.15
	NH		7.84		β	17.4, CH_3	1.20
Val-9	α	57.6, CH	4.12		CO	172.9, C	
	β	30.3, CH	1.99		NH		7.92
	γ	17.8, CH_3	0.83, d (6.8)	**Ala-18**	α	48.2, CH	4.22
	γ′	19.0, CH_3	0.83, d (6.8)		β	18.1, CH_3	1.19
	CO	ND			CO	172.9, C	
	NH		7.90		NH		7.93
Ala-10	α	48.1, CH	4.32	**Ala-19**	α	48.2, CH	4.18
	β	17.9, CH_3	1.28		β	17.4, CH_3	1.19
	CO	172.1			CO	171.8, C	
	NH		8.23		NH		7.97

Table S9 (continued).

Ala-20	α	48.2, CH	4.22
	β	18.1, CH_3	1.19
	CO	172.9, C	
	NH		7.93
Ala-21	α	48.1, CH	4.29
	β	17.4, CH_3	1.17
	CO	171.8, C	
	NH		7.98
Ile-22	α	56.9, CH	4.15
	β	36.4, CH	1.70
	γ	24.1, CH_2	a: 1.41
			b: 1.04
	δ	10.8, CH_3	0.78
	β-CH_3	15.0, CH_3	0.78
	CO	170.4, C	
	NH		7.82
Ala-23	α	48.1, CH	4.32
	β	17.6, CH_3	1.18
	CO	172.2, C	
	NH		8.14
Gly-24	α	41.7, CH_2	3.73
	CO	168.4, C	
	NH		8.22
Ala-25	α	48.1, CH	4.32
	β	17.4, CH_3	1.29
	CO	172.1, C	
	NH		8.23
Abu(S)-26	α	56.2, CH	4.31
	β	39.7^b, CH	3.04
	γ	17.9, CH_3	1.29
	CO	ND	
	NH		8.81
Tyr-27	α	48.1, CH	4.32
	β	34.5^b, CH_2	a: 3.04
			b: 2.92
	1 (phenol)	128.1, C	
	2/6 (phenol)	129.8, CH	6.87, d (8.0)
	3/5 (phenol)	114.8, CH	6.57, d (8.0)
	4 (phenol)	155.5, C	
	CO	ND	
	NH		8.31
Glu-28	α	51.4, CH	4.35
	β	28.4,CH_2	a: 1.81
			b: 1.74
	γ	33.5, CH_2	2.16
	δ	ND	
	CO	ND	
	NH		8.29
Ala-29	α	49.8, CH	3.94
	β	15.7, CH_3	1.19
	CO	173.0, C	
	NH		8.61
Gly-30	α	42.7, CH_2	a: 3.88, dd (17.0, 6.7)
			b: 3.55, dd (17.0, 4.7)
	CO	167.7, C	
	NH		8.63
Vinyl-amine-31	α	121.3, CH	6.64, dd (9.9, 8.5)
	β	105.3, CH	5.33, d (8.5)
	NH		8.73

ND: Not clearly detected

a Multiplicities not given to overlapping or broad signals

b Very weak 13C NMR signal in the HSQC

Table S10. ^{1}H **and** ^{13}C **NMR data for the G24, A25 and the C-terminal ring of pristinin A3 (CD_3CN:H_2O 9:1, 850 MHz, 297 K)**

Residue	Position	δ_C, type	δ_H, mult.[a] (*J* in Hz)	Residue	Position	δ_C, type	δ_H, mult.[a] (*J* in Hz)
Gly-24	α	44.1, CH_2	3.83, dd (11.9, 5.6)	**Glu-28**	α	55.7, CH	4.09 [4.11]
	CO	171.9			β	28.7, CH_2	1.94 [1.92]
	NH		8.15 [8.12]		γ	34.3, CH_2	2.24 [2.18]
Ala-25	α	50.5, CH	4.23		δ	180.7, C	
	β	17.2, CH_3	1.27		CO	174.4, C	
	CO	175.2			NH		8.35 [8.41]
	NH		8.06 [7.96]	**Ala-29**	α	50.6, CH	4.22
Abu(S)-26	α	57.6, CH	3.87 [3.95]		β	17.3, CH_3	1.28
	β	45.6, CH	3.26		CO	175.3	
	γ	20.2, CH_3	1.24		NH		8.06
	CO	172.5, C		**Gly-30**	α	44.9, CH_2	a: 3.95 b: 3.73
	NH		8.46 [8.44]		CO	169.6, C	
Tyr-27	α	58.8, CH	4.20 [4.22]		NH		7.81 [7.82]
	β	36.4, CH_2	a: 3.07 [3.06] b: 2.94 [2.92]	**Vinylamine-31**	α	125.8, CH	6.98
	1 (phenol)	128.7, C			β	102.2, CH	5.37, d (7.5)
	2/6 (phenol)	131.3, CH	7.02		NH		9.46 [9.35]
	3/5 (phenol)	116.3, CH	6.69, d (8.0)				
	4 (phenol)	156.8, C					
	CO	174.0					
	NH		8.12 [8.08]				

[*] Following the long time acquisition of the COSY, TOCSY, HSQC and HMBC spectra; the solvent evaporated from the NMR tube. The sample was then re-prepared for additional NOESY and longer HMBC experiments. It was noted from the ^{1}H NMR spectrum of the second run that some ^{1}H signals, especially those of the NH, have slightly shifted. Accordingly, a COSY experiment was repeated in the second run, to relate the ^{1}H NMR resonances of the two sets of data. The ^{1}H NMR resonance in the second run is given in square brackets, if it is different from the first one.

[a] Multiplicities not given to overlapping or broad signals

Table S11. Ratio of oxidized product in samples analyzed by NMR. The relative ares indicate the integrated areas divided over the integrated areas of the unmodified base peak.

Extract	NMR solvent	Base peak +16 Da relative area	Base peak +48 Da relative area
Crude extract	NA	0.152	0.066
Pristinin A3	DMSO-d_6	0.103	ND
Pristinin A3	CD3CN:H2O 9:1	0.287	3

NA: Not applicable; ND: Not determined

Table S12. Fragmentation data of oxidized products. X's indicate that a mass was observed within 10 ppm. A mixture of oxidized and non-oxidized fragments can be observed when the fragments do not contain the center ring structure. When the fragments do contain the center ring structure, they are always oxidized, suggesting the center ring contains the oxidation.

Ion	Obs. *m/z*	Obs. *m/z* (+16 Da)	Obs. *m/z* (+32 Da)	Obs. *m/z* (+48 Da)	Ion	Obs. *m/z*	Obs. *m/z* (+16 Da)	Obs. *m/z* (+32 Da)	Obs. *m/z* (+48 Da)
b1					y1				
b2					y2				
b3					y3				
b4					y4				
b5	x				y5				
b6	x	x			y6	x			
b7	x	x			y7	x			
b8	x	x			y8	x	x		
b9	x	x			y9	x	x		
b10	x	x			y10	x	x		
b11	x				y11	x	x		
b12					y12	x			
b13					y13	x			
b14					y14	x			
b15					y15	x			
b16					y16				
b17				x	y17				
b18			x	x	y18				
b19				x	y19				
b20				x	y20				
b21				x	y21				x
b22				x	y22				x
b23					y23				x
b24					y24				x
b25				x	y25				x
b26					y26				
b27					y27				
b28					y28				
b29					y29				
b30					y30				
b31					y31				

Table S13. Homologs of the genes *lanJ$_A$*, *sprF1, sprF2 and sprOR* are found associated with both known lanthipeptide BGCs and close to the *sprPT/sprH3* gene pair. Homology was determined at a cutoff of 30% amino acid identity of the gene products. Within *Streptomyces* genomes, all homologs were found within the analyzed 1,295 genomes. It was then checked whether these homologs overlapped with an antiSMASH-detected lanthipeptide BGC, or were within 15 genes of the *sprPT/sprH3* gene pair. *sprOR* homologs were found within canonical lanthipeptide BGCs as well as associated with the *spriPT/sprH3* gene pair, suggesting its association with lanthipeptide BGCs. For non-*Streptomyces* genomes, the *sprPT/sprH3* gene pair was first detected, and homologs of the given queries were found within the 15 surrounding genes. Homologs of *lanJ$_A$* and *sprF1* are often found associated with *sprPT/sprH3* gene pair, suggesting they are involved in lanthipeptide biosynthesis.

Streptomyces genomes

Query	**Overlap with lanthipeptide BGC**	***sprPT/sprH3* gene pair**	**Overlap with both**	**Overlap with neither**
lanjA	0	0	0	5
sprOR	124	137	2	199
sprF1	0	124	2	16
sprF2	13	135	2	348

Non-*Streptomyces* gene clusters

Query	**Overlap with lanthipeptide BGC**	***sprPT/sprH3* gene pair**	**Overlap with both**	**Overlap with neither**
lanjA	0	40	0	0
sprOR	0	108	0	0
sprF1	0	111	0	0
sprF2	0	146	0	0