A Bifunctional Leader Peptidase/ABC Transporter Protein Is Involved in the Maturation of the Lasso Peptide Cochonodin I From *Streptococcus Suis*

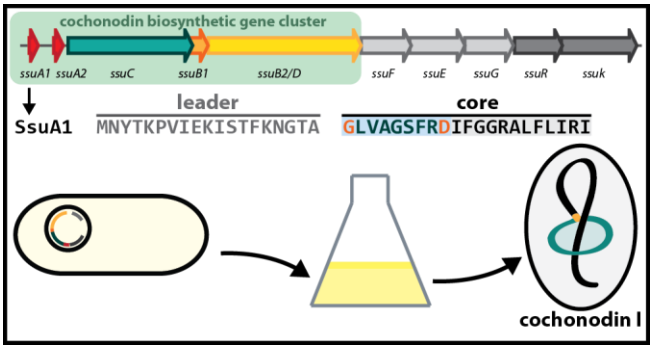
Julian D. Hegemann,[†,*] Kevin Jeanne Dit Fouque,[‡] Miguel Santos-Fernandez,[‡] Francisco Fernandez-Lima[‡]

† Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Saarland University Campus, 66123 Saarbrücken, Germany.

‡ Department of Chemistry and Biochemistry, Florida International University, Miami, Florida 33199, United States.

Lasso peptides are members of the natural product superfamily of ribosomally synthesized and post-translationally modified peptides (RiPPs). Here, we describe the first lasso peptide originating from a biosynthetic gene cluster belonging to a unique lasso peptide subclade defined by the presence of a bifunctional protein harboring both a leader peptidase (B2) and an ABC transporter (D) domain. Bioinformatic analysis revealed that these clusters also encode homologs of the NisR/NisK regulatory system and the NisF/NisE/NisG immunity factors, which are usually associated with the clusters of antimicrobial class I lanthipeptides, such as nisin, another distinct RiPP subfamily. The cluster enabling the heterologous production of the lasso peptide cochonodin I in *E. coli* originated from *Streptococcus suis* LSS65 and the threaded structure of cochonodin I was evidenced through extensive MS/MS analysis and stability assays. It was shown that the ABC transporter domain from SsuB2/D is not essential for lasso peptide maturation. By extensive genome mining dedicated exclusively to other lasso peptide biosynthetic gene clusters featuring bifunctional B2/D proteins, it was furthermore revealed that many bacteria associated with human or animal microbiota hold the biosynthetic potential to produce cochonodin-like lasso peptides, implicating that these natural products might play roles in human and animal health.

Table of Contents/Abstract Graphic



Lasso peptides are fascinating natural products belonging to the superfamily of ribosomally synthesized and post-translationally modified peptides (RiPPs).[1-3] As such, their biosynthesis starts with the ribosomal assembly of a precursor peptide that is then processed by specific biosynthetic enzymes. The precursors themselves are subdivided into a leader region, needed for enzyme recognition, and a core region, into which the processing enzymes incorporate the post-translational modifications. Lasso peptides can exhibit a manifold of different biological properties, e.g., antimicrobial, anticancer, or antiviral activities.[1,3-6]

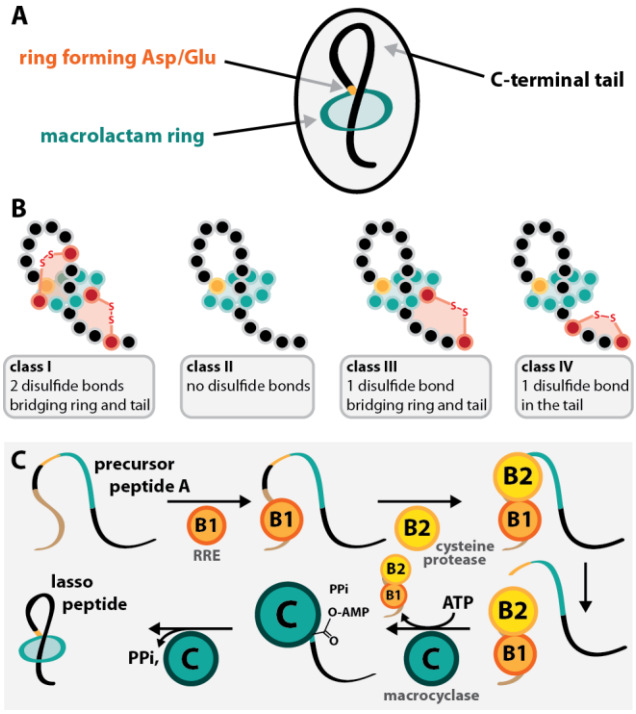


Figure 1. Schematic representation of the (A) lasso fold, and (B) the four currently known lasso peptide classes.[1-3,5,6] (C) Basic principles of lasso peptide biosynthesis.

Members of the lasso peptide RiPP subfamily are defined by a characteristic fold that is reminiscent of a lariat knot (Figure 1A).[1-3] Here, an N-terminal macrolactam ring is threaded by a linear C-terminal tail. These threaded structures are primarily stabilized by steric interactions mediated through the so-called plug residues; amino acids with large, bulky side chains placed

above and below the segment of the tail that threads the ring. Lasso peptides are known to be highly resistant to proteolysis and denaturation through chaotropic reagents and many lasso peptides exhibit a high thermal stability.[1-3] There have also been reports of heat sensitive lasso peptides that unthread into branched-cyclic structures during prolonged incubation at elevated temperatures.[2,7-11] The lasso fold can also be stabilized through the presence of disulfide bonds, which serves as the basis for their further classification (Figure 1B).

The processing of a precursor (A) into a mature lasso peptide requires at least three distinct protein activities[1,3,9,12-15] (Figure 1C): 1.) A RiPP recognition element (RRE) binds to a specific site of the N-terminal leader peptide. 2.) The RRE then recruits a Cys protease that separates leader and core peptide via proteolysis. 3.) A lasso cyclase with homology to asparagine synthetases next activates the carboxylic acid side chain of an Asp or Glu residue at position 7, 8, or 9 of the core peptide in an ATP-dependent manner. Subsequently, the lasso cyclase catalyzes the formation of the macrolactam via condensation of the N-terminal α-amine of the free core peptide with the activated AMP ester.

Besides the lasso cyclase (C), lasso peptide BGCs express either a protein (B) that features two domains fulfilling the functions of the RRE and leader peptidase, respectively, or they encode the RRE (B1) and leader peptidase (B2) as two discrete proteins.[1,3,6,9,11-15] BGCs of antimicrobial lasso peptides usually contain dedicated ABC-transporters for lasso peptide export and self-immunity,[1,3,11,16-19] which are typically annotated as D proteins.

In several previous genome mining campaigns,[6,20,21] a unique lasso peptide subclade was noticed that is predominantly found in Firmicutes. The defining feature of this subclade is that the BGCs encode bifunctional proteins (B2/D), containing both N-terminal leader peptidase and C-terminal ABC transporter domains, that are not present in other lasso peptide operons.

However, so far none of these BGCs have been investigated experimentally and no lasso peptides associated with any of these clusters have been reported.

Here, we describe the novel lasso peptide cochonodin I, which originates from such a lasso peptide BGC found in *Streptococcus suis* LSS65. The mature lasso peptide was heterologously produced in *E. coli* and its lasso fold was confirmed by extensive mass spectrometric (MS) analysis and a combination of heat/carboxypeptidase Y stability assays. It was furthermore shown that cochonodin I biosynthesis is still possible after the ABC transporter domain has been removed from the bifunctional B2/D protein. A targeted genome mining search for members of this specific lasso peptide subclade was performed to give an updated overview of such lasso peptide BGCs in public genomic databases and to allow a closer look at their distribution in other microorganisms. This analysis revealed that these lasso peptides are often found in human and animal microbiota and can potentially be produced by both pathogens as well as commensals, which gives the study of these natural products a particular relevance.

RESULTS AND DISCUSSION

Bioinformatic Analysis of the Lasso Peptide BGC from *Streptococcus suis* LSS65. The lasso peptide BGC from *S. suis* LSS65 (Figure 2A) encodes two similar precursor peptides (SsuA1, 40 aa, and SsuA2, 42 aa; 52.4% identity / 61.9% similarity) and a basic set of lasso peptide biosynthetic enzymes, namely, a lasso cyclase (SsuC, 587 aa), an RRE protein (SsuB1, 83 aa), and a bifunctional leader peptidase/ABC transporter protein (SsuB2/D, 740 aa).

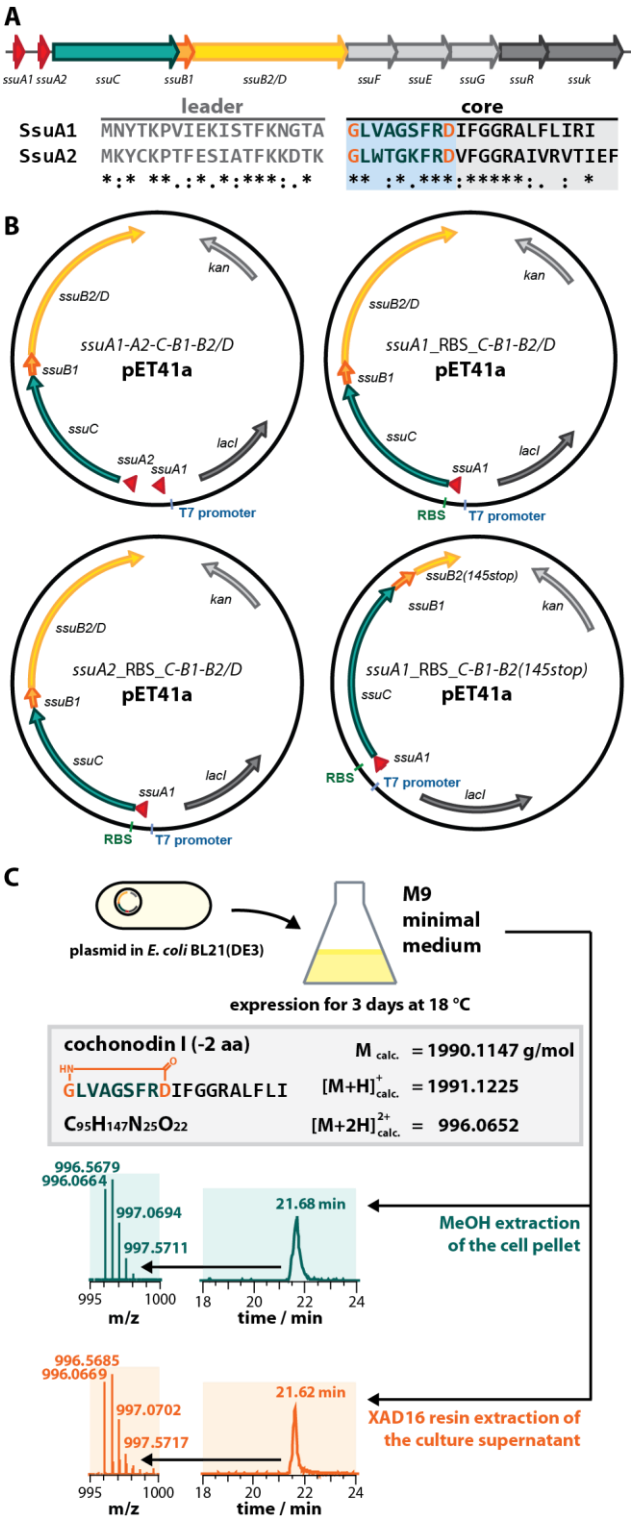


Figure 2. (A) Schematic representation of the cochonodin biosynthetic gene cluster from *S. suis* LSS65 and the sequences of the precursor peptides SsuA1 and SsuA2. (B) Expression plasmids used in this study. (C) Scheme depicting the heterologous production route that yielded cochonodin I (-2 aa). The data shown were obtained by expressing *E. coli* BL21(DE3) carrying the *ssuA1*_RBS_*ssuC-B1-B2/D* pET41a plasmid for 3 days at 18 °C.

The latter is highly unusual and has not been experimentally addressed before, although it is typical for this particular lasso peptide subclade. The A1-A2-C-B1-B2/D organization of the cluster is strikingly different from previously studied lasso peptide BGCs from Firmicutes, which were mostly found in *Bacillus* and *Paenibacillus* species.[6,11,22] For example, the paeninodin BGC from *Paenibacillus dendritiformis* C454 has a C-A-X-B1-B2-D cluster architecture[22] (here, X stands for a lasso peptide specific kinase that phosphorylates the hydroxy group of a C-terminal Ser in the precursor peptide[22,23]). Nonetheless, there is still a reasonable similarity between the corresponding processing enzymes (SI Table S1), which also corroborates that the SsuB2/D protein indeed features an N-terminal leader peptidase domain. Bioinformatic analysis further suggests that the first membrane helix of SsuB2/D starts around residue 178 and shows how the first 145 residues of SsuB2 align well with PadeB2 (SI Figure S1). However, the sequence comparisons also demonstrate that although both lasso peptide subclades are often associated with Firmicutes, they are merely distantly related to each other.

Another important similarity between SsuB2/D and proteins belonging to the ABC-transporter maturation and secretion (AMS)/ peptidase-containing ATP-binding transporter (PCAT) protein family needs to be pointed out. The AMS/PCAT proteins feature an N-terminal peptidase and a C-terminal ABC transporter domain, analogous to SsuB2/D. The peptidase domains of AMS/PCAT exporters belong to the C39 cysteine peptidase family and typically remove an N-terminal leader sequence from their substrates, while the ABC transporter domains mediate the export of the released peptides or proteins.[24-26] However, closer inspection shows that these different types of bifunctional enzymes are less similar on a mechanistic level.

Lasso peptide precursors have a conserved YxxP motif for recognition by their B1 proteins[13-15] and contain a conserved Thr-2 residue that guides the B2 protein to cleave at the correct

position.[13,14] In contrast, the C39 peptidase domains of AMS/PCAT exporters typically recognize their substrates by binding to a LxxxxL motif and cleave after a highly conserved Gly-Gly/Ala/Ser sequence.[24-26]

Indeed, comparisons demonstrate that while the ABC transporter domain of SsuB2/D shows considerable homology with the same domains of the AMS/PCAT proteins LahT and PCAT1 (SI Table S1), there is little conservation observed when aligning the peptidase domains (SI Table S1). In addition, the catalytic Cys nucleophiles of the peptidase domains are located at different sites (Cys78 for SsuB2/D, Cys27 for LahT, Cys21 for PCAT1). Hence, the similarity regarding the attachment of a peptidase to the N-terminus of an ABC transporter seems likely to be a case of convergent evolution instead of a direct relationship between the lasso peptide B2/D proteins and the AMS/PCAT exporters.

This makes sense from a mechanistic point of view. The peptidase domains of AMS/PCAT proteins usually convert inactive precursors into bioactive secondary metabolites that need to be rapidly exported to not harm the producing organism. Conversely, the peptidase domain of SsuB2/D also requires the interaction with SsuB1 for substrate recognition and then merely releases the core peptide for enabling the further processing through SsuC. Hence, it could be hypothesized that a lasso peptide synthetase/export complex is formed from SsuB1, SsuB2/D, and SsuC. This notion is supported by previous studies with the biosynthetic machinery of the lasso peptide microcin J25 showing that these enzymes appear to be associated with the bacterial membrane[27] and suggesting that the enzymes might form a lasso synthetase complex.[12]

The *ssuB2/D* gene is followed by five additional open reading frames (ORFs) and the short distances between each of these genes implies that they might still belong to the lasso peptide operon. Closer inspection of the proteins encoded by these five genes, the analysis of their

conserved domains, and the search for closely related homologs revealed a remarkable similarity of their primary structures to the proteins NisF, NisE, NisG, NisR, and NisK (SI Table S1). These proteins are usually found in context of the nisin BGC from *Lactococcus lactis*.[28-32] Nisin is an antimicrobial lanthipeptide, a RiPP subfamily that shares little commonality with lasso peptides, except following the same biosynthetic logic of a genetically encoded precursor that is turned into a complex natural product by the corresponding processing enzymes.[28,33] Neither the biosynthetic enzymes nor the rings installed into the core peptides of these RiPP subfamilies (a macrolactam for lasso peptides versus β-thioether crosslinks for lanthipeptides) have high similarity.

In the nisin system, NisFEG are membrane proteins that together form an ABC transporter complex that is an additional immunity factor for self-protection of the *L. lactis* producer strain against nisin.[30,31] The antimicrobial activity of nisin is derived from nisin binding to the peptidoglycan precursor lipid II in the membrane of Gram-positive bacteria and then triggering the formation of membrane pores.[28-30] NisRK are needed for the regulation of the nisin biosynthesis.[29-32] NisK is a sensor histidine kinase, while NisR is a response regulator. In response to the detection of nisin in the extracellular space, NisK first autophosphorylates one of its own His residues and then transfers the phosphate group onto NisR. Phosphorylated NisR in turn acts as a transcriptional regulator that upregulates both the expression of the nisin precursor (NisA) and the corresponding nisin processing enzymes, as well as the NisFEG immunity cassette.[29-31] Hence, the presence of genes encoding homologs to both NisFEG and NisRK suggests that the natural products produced by this lasso peptide BGC could also exhibit antimicrobial activities and could be regulated in a similar manner to the nisin system.

Heterologous Production of Cochonodin I. To avoid working with BSL-2 classified *S. suis* strains, a heterologous production route for the isolation of the lasso peptides from the *S. suis* LSS65 BGC was established using *E. coli* BL21(DE3) as an expression host (Figure 2B and 2C). Towards this goal, the *ssuA1-A2-C-B1-B2/D* genes were cloned into the pET41a expression vector (*ssuA1-A2-C-B1-B2/D* pET41a), which harbors an IPTG-inducible T7 promotor. As it was previously shown that co-expression of two precursors at once is detrimental to the yields of the individual lasso peptides,[34] two additional expression plasmids containing only a single precursor gene were generated (*ssuA1*_RBS_*ssuC-B1-B2/D* pET41a and *ssuA2*_RBS_*ssuC-B1-B2/D* pET41a). Furthermore, the intergenic region between *ssuA1*/*ssuA2* and *ssuC* was replaced with an *E. coli* optimized ribosomal binding site (RBS) as this approach was previously shown to be an effective strategy to increase processing enzyme expression and in turn the overall yield of the mature lasso peptide.[34-36]

Only a weak MS signal was detected for the lasso peptide derived from SsuA1 when expressing *ssuA1-A2-C-B1-B2/D* pET41a, while a significantly higher MS signal was obtained when expressing *ssuA1*_RBS_*ssuC-B1-B2/D* pET41a (Figure 2C and SI Figure S2). In contrast, no MS signals related to SsuA2-derived lasso peptides were detected after expression of either *ssuA1-A2-C-B1-B2/D* pET41a or *ssuA2*_RBS_*ssuC-B1-B2/D* pET41a. Hence, it was decided to focus on the characterization of the SsuA1-derived lasso peptide, which was named cochonodin I (the prefix *cochon-* is the French word for pig and was chosen as not only *S. suis* is a known pathogen of pigs but also because *suis* is a latin adjective referring to pigs. The *-nodin* suffix was established previously and is derived from the latin word *nodus* which translates into “knot”). As observed for other heterologously produced lasso peptides,[7,8,34,35] only a truncation of cochonodin I lacking the last two C-terminal amino acids (cochonodin I (-2 aa), $C_{95}H_{147}N_{25}O_{22}$,

calculated M = 1990.1147 g/mol, calculated $[M+2H]^{2+}$ = 996.0652) was extracted from *E. coli* production cultures. Notably, the MS signal relating to cochonodin I (-2 aa) was detected in extracts from both the culture supernatant (SN) and the cell pellet (Figure 2C and SI Figure S2).

The ABC-transporter domain of SsuB2/D is Not Essential for Lasso Peptide Biosynthesis. To investigate if the attachment of the cysteine protease to the ABC transporter domain is required for lasso peptide maturation, it was attempted to delete the ABC transporter domain and thus only co-express SsuA1 with the known basic biosynthetic components needed for lasso peptide maturation. To identify the right position to separate both domains, the location of the transmembrane helices was predicted bioinformatically (SI Figure S1A). This prediction was complemented by an alignment of the N-terminal region of SsuB2/D with other discrete B2 proteins and by the prediction of the secondary structures of these proteins (SI Figure S1B). The proteins chosen for this alignment were taken from well-known lasso peptide BGCs and represent B2 proteins from a range of different organisms. Namely, an alignment was performed with CptB2 (from the chaxapeptin BGC found in *Streptomyces leeuwenhoekii* str. C58)[36] LarB2 (from the lariatin BGC found in *Rhodococcus jostii* K01-B0171),[37] and PadeB2 (from the paeniodin BGC found in *Paenibacillus dendritiformis* C454).[22] Based on these alignments and on the prediction of the transmembrane helices, it was decided to delete the complete ABC transporter domain after position 145 of SsuB2/D and thereby generating the plasmid *ssuA1*_RBS_*ssuC-B1-B2(145Stop)* pET41a.

Expression of this plasmid still led to the production of cochonodin I (-2 aa) (SI Figure S2C), demonstrating that, as expected, the transporter domain of SsuB2/D is not essential for precursor processing and that an artificially generated discrete SsuB2 protein retains its proteolytic activity.

Interestingly, cochonodin I (-2 aa) could still be detected in the culture SN in the absence of the ABC transporter domain, albeit with a significantly reduced MS signal intensity (SI Figure 2C) compared to the SN extract from the co-expression with intact SsuB2/D (SI Figure 2B). A possible explanation for this observation is that *E. coli* BL21(DE3) expresses another transport protein able to export cochonodin I (-2 aa), although it might act at a reduced efflux rate compared to SsuB2/D. While this assumption remains speculative, such a notion is not unheard of in the context of lasso peptides. For example, it was previously reported that the *E. coli* ABC transporter YojI can mediate the export of microcin J25 and that YojI overexpression increases the resistance of *E. coli* against this antimicrobial lasso peptide.[17,38]

Characterization of the Lasso Fold of Cochonodin I (-2 aa) via tandem MS/MS. As the lasso peptide yields were too low to allow cochonodin I (-2 aa) isolation from either the cell pellet or the culture SN, the elucidation of the three-dimensional structure of cochonodin I (-2 aa) via NMR spectroscopy or crystallography remained out of the scope of this study. The characterization of the lasso fold of cochonodin I (-2 aa) and the validation of the primary structure derived from the SsuA1 amino acid sequence was instead accomplished by employing an extensive tandem MS/MS-based toolkit previously established for lasso peptide analysis.[1-3,22,34,36,39-43]

The ESI-MS analysis of cochonodin I (-2 aa) resulted in the observation of a single charge state species ($[M+2H]^{2+}$) under denaturing conditions. The tandem MS/MS using collision-induced dissociation (CID) of the $[M+2H]^{2+}$ ion of cochonodin I (-2 aa) (*m/z* 996.1) exhibited a fragmentation pattern of b_i/y_j ions, corresponding to fragmentations in the C-terminal tail of cochonodin I (-2 aa), as well as internal fragment ions, denoted as $(b_iy_j)_n$, and mechanically interlocked product ions, denoted as $[(b_i)^*(y_j)]$, with associated b_i and y_j fragments (Figure 3A).

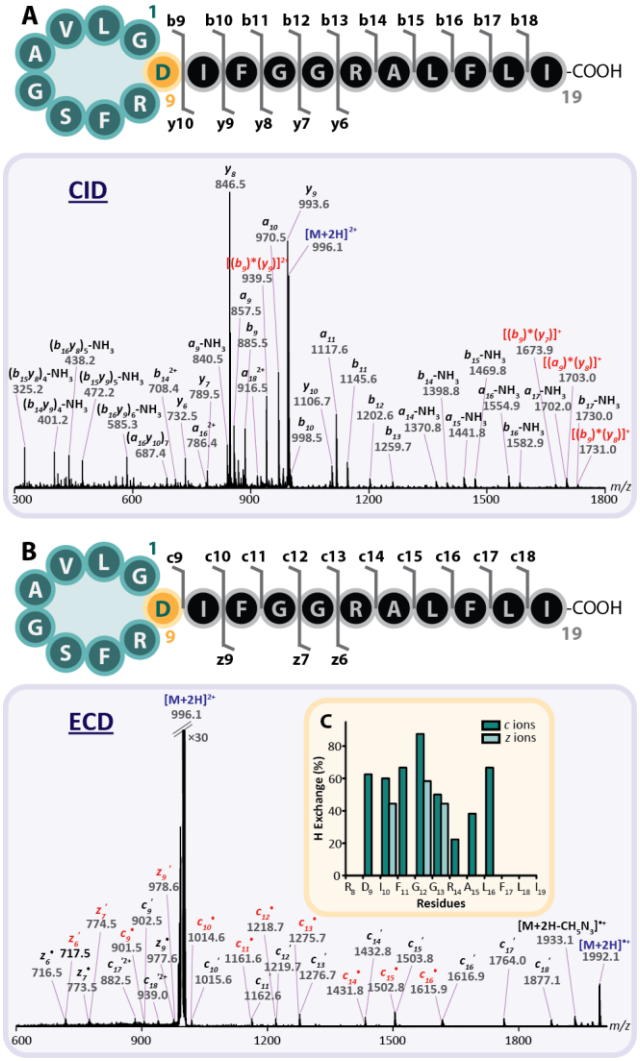


Figure 3. (A) Tandem MS/MS CID spectrum of the $[M+2H]^{2+}$ species of cochonodin I (-2 aa). (B) Tandem MS/MS ECD spectrum of the $[M+2H]^{2+}$ species of cochonodin I (-2 aa). (C) Relative H• transfer as observed in the ECD spectrum. The bar plots were calculated by correcting the contributions of the theoretical isotope patterns to the experimentally observed isotopic distributions. Schematic depictions of the primary structure of cochonodin I (-2 aa) are shown with assigned *b*/*y*- or *c*/*z*-fragments as observed in the CID or ECD spectra. Instead of the threaded lasso structure, a linear arrangement of residues is depicted in these schematics to improve the clarity of the signal assignments.

Interlocked fragment ions result from two bond cleavages in the loop region, for which the b_i and y_j product ions remain associated through steric interactions of the plug residues (in the y_j fragment) with the macrolactam ring (in the b_i fragment). These interlocked fragment ions are typical for the [1]rotaxane structures found in lasso peptides as previously reported.[1-3,39-41]

Therefore, the presence of such mechanically interlocked product ions (highlighted in red in Figure 3A) are clear indicators of the threaded lasso topology of cochonodin I (-2 aa).

Cochonodin I and capistruin both feature a threaded Gly1-Asp9 macrolactam ring.[2,44] Based on reports that either Arg or Phe residues can stabilize the lasso fold in capistruin[2,44] and our tandem MS/MS data, we reasoned that either the bulky Arg14 or Phe17 residues could be likely candidates to fulfill the role as lower plug in the threaded lasso structure. However, the Arg14 residue can be easily excluded as the lower plug based on the presence of mechanically interlocked fragments, as it was previously demonstrated how the loop has to be strictly longer than four residues for interlocked fragments to occur, while with Arg14 as the lower plug the loop would comprise exactly four residues.[41] Thus, we propose that the Phe17 residue acts as the lower plug in cochonodin I.

Tandem MS/MS using electron capture dissociation (ECD) was also performed on the $[M+2H]^{2+}$ ion (*m/z* 996.1) of cochonodin I (-2 aa) (Figure 3B). The tandem MS/MS ECD spectrum displayed the charge-reduced $[M+2H]^{\bullet+}$ ion (*m/z* 1992.1), as signature ion of the ECD events, and a classic c_i'/$z_j^{\bullet}$ fragment series resulting from cleavages within the C-terminal tail of cochonodin I (-2 aa). In addition, extensive hydrogen migration events (H• transfer) were observed, for which the c_9' to c_{16}' fragment ions were partially shifted by 1 Da from a loss of H• ($c_9^{\bullet}$ to $c_{16}^{\bullet}$, highlighted in red in Figure 3B), while the $z_6^{\bullet}$, $z_7^{\bullet}$ and $z_9^{\bullet}$ product ions were partially shifted by 1 Da from a capture of H• (z_6', z_7' and z_9', highlighted in red in Figure 3B). Extensive hydrogen migration was previously demonstrated as another typical feature for a large variety of lasso peptides[39-41] and supports the prediction of the threaded lasso structure of cochonodin I (-2 aa).

Such hydrogen migration events were shown to largely occur in the loop region of a threaded lasso peptide, where H• transfer is facilitated through the close proximity of the loop residues to each other, while residues located outside of the macrolactam ring are positioned in a way that does not allow H• transfer to occur as frequently. The extensive H• transfer (Figure 3B and 3C) observed for the Ile10-Leu16 region of cochonodin I (-2 aa) therefore delineates these residues as the likely loop of cochonodin I, which, in agreement with the observations from the tandem MS/MS CID data, provides further evidence for Phe17 fulfilling the function of the lower plug in cochonodin I.

Extended Thermostability and Carboxypeptidase Y Assays of Cochonodin I (-2 aa). A combined thermal and carboxypeptidase Y stability assay, previously established for this purpose,[1-3,5,8,9,22,34] was performed to further corroborate the lasso fold (SI Figure S3).

Whereas heating of cochonodin I (-2 aa) for 4 h at 95 °C did not cause a significant shift in retention time, carboxypeptidase Y treatment confirmed the heat-induced unthreading of the lasso peptide. When cochonodin I (-2 aa) was directly treated with carboxypeptidase Y, only one additional C-terminal amino acid was lost, meaning that the full conversion to cochonodin I (-3 aa) took place. This observation is also in favor of the notion that the Phe17 residue acts as the lower plug as suggested by the tandem MS/MS data. In contrast, if a sample of cochonodin I (-2 aa) was first incubated for 4 h at 95 °C and then treated with carboxypeptidase Y, the peptide was degraded to a mixture of -11 aa and -12 aa species, which correspond to just the macrolactam ring (cochonodin I (-12 aa)) or the ring with only one additional amino acid attached to it (cochonodin I (-11 aa)). This behavior is characteristic for the carboxypeptidase Y treatment of a branched-cyclic peptide with an N-terminal macrolactam ring that arises from the thermally induced unthreading of a lasso peptide.[1-3,6-8,11,22,34]

The high resistance to carboxypeptidase Y degradation of the compound prior to heating is on the other hand typical for a lasso fold, where the threaded segment of the linear C-terminal tail is shielded from carboxypeptidase Y access.[1-3,6-8,11,22,34] Taken together, these experiments demonstrate that cochonodin I (-2 aa) is a heat-sensitive lasso peptide.

Genome Mining Reveals the Distribution of Cochonodin-like BGCs in Sequenced Microorganisms. To get a better overview of the distribution and abundance of cochonodin-like lasso peptides, genome mining for closely related BGCs was performed. Towards this goal the unique nature of the bifunctional B2/D proteins was leveraged by using the SsuB2/D sequence from the cochonodin BGC as query for the *Basic Local Alignment Search Tool* (BLAST).[45] Since the larger ABC transporter domain shares a general homology to other ABC transporters, this search yields many false positives in form of ABC transporters outside of cochonodin-like gene clusters. However, the absence of an N-terminal Cys protease domain, which is reflected by query coverages below a certain threshold, enables the easy identification and elimination of such unrelated ABC transporters. Moreover, this approach has the advantage that lasso peptide BGCs not encoding a B2/D protein are excluded from the hits by the search algorithm since the comparatively small size of discrete B2 proteins gives too low scores when aligned with the much bigger bifunctional B2/D proteins.

Through manual curation of the hits identified from this search, a variety of clusters encoding almost 100 precursors were identified (SI Figure S4, SI Table S2, and SI Table S3). Most of these were found in *Streptococci* and *Enterococci*, albeit the large number of sequenced genomes of *E. faecalis*, *S. suis*, and *S. pneumoniae* spp. that are present in public databases likely provides a certain bias towards these particular strains.

Clusters found in *S. suis* genomes tend to contain two precursor genes as described for the cochonodin BGC from *S. suis* LSS65. Conversely, BGCs found in *S. pneumoniae* and *E. faecalis* genome sequences always encode only one precursor. Furthermore, four BGCs were discovered that encoded more than two precursors. The BGC from *Peptidiphaga gingivicola* NCTC370 encodes four, the clusters from *Actinobaculum* sp. oral taxon 183 str. F0552 and *Actinomyces* sp. oral taxon 848 str. F0332 encode five, and the BGC from *Ruminococcus flavefaciens* ATCC 19208 even encodes six different precursor peptides.

Remarkably, either NisR/NisK pairs, NisF/NisE/NisG cassettes, or most commonly both are encoded in proximity of all identified lasso peptide BGCs (SI Table S2). This observation demonstrates that not only the B2/D proteins but also the association with these potential regulatory elements and immunity factors are benchmarks of this kind of lasso peptide BGC.

To give a better overview of the results of our genome mining efforts, a structure similarity network (SSN) analysis[46] was performed based on the core peptide sequences associated with each BGC. The resulting SSN (SI Figure 4A) shows that there are two larger and one smaller clade based on the primary structures of the predicted lasso peptides as well as a number of singleton sequences.

Clade I comprises of predicted lasso peptides similar to the hypothetical cochonodin II. The putative lasso peptides derived from almost all BGCs encoding only a single precursor peptide belong to clade I; the only exceptions stem from two out of the three identified clusters originating from a *Paenibacillus* strain. Clade II contains lasso peptides with high homology to cochonodin I. These lasso peptides are exclusively found in BGCs with two precursor peptides like the *ssu* operon. In these BGCs, one of the two precursors always bears high resemblance to cochonodin I, whereas the second precursor typically represents a member of clade I.

To visualize the sequence space covered by cochonodin-like lasso peptides, conserved motifs were predicted for the leader and core peptide regions of the putative lasso peptide precursors. The observed leader peptide motif (SI Figure 4B) contains the conserved Thr at the penultimate position of the leader that is typical for lasso peptide precursors[14,22,36] and allows the B2 protein to recognize the site of leader cleavage.[14] In addition, the presence of the YxxP motif, typically found in precursors originating from clusters encoding discrete B1 and B2 proteins, was also observed. The conserved YxxP sequence was previously demonstrated to mediate the recognition by the lasso peptide RRE, which initiates the precursor maturation process.[6,9,13-15,36] An additional Glu residue in proximity to the YxxP motif (YxxPxxE) is furthermore fully conserved amongst all cochonodin-type precursors. This was not observed in any previously investigated lasso peptide precursors and hence it is unclear if this Glu is important for precursor maturation. The same is true for other residues that show at least a rather high, albeit not complete, conservation between the identified precursors.

The comparison of the identified core peptide motifs shows that lasso peptide BGCs featuring B2/D proteins exclusively produce lasso peptides with Gly1-Asp9 macrolactams (SI Figure 4C). Whereas some positions in the ring and close to the C-terminus show some sequence variety, other regions of the core peptides seem to be fully or at least strongly conserved. Particularly striking is the FGGRA sequence following shortly after the ring forming Asp9 residue. The proximity of these residues to the isopeptide bond demands that they must be, at least in part, situated in the lasso peptide loop, which is usually one of the most rigid regions of a lasso peptide scaffold. Hence, the conservation at these positions implies that this sequence might be of some importance for the natural function of cochonodin-like lasso peptides, although what this function is remains elusive.

When comparing only the core peptides relating to a distinct clade (SI Figure 4D), it becomes apparent that the structural variety of lasso peptides belonging to clade I and clade III is significantly higher when compared to the nigh identical sequences of clade II lasso peptides, where only slight differences at positions 4, 6, and 20 are observed. This points towards a close relationship of the corresponding BGCs encoding two precursors that were found in *S. vestibularis* F0396, *S.* sp. 19428wD3 AN2, *Lactococcus garvieae* Tac2, and a number of *S. suis* and *S. salivarius* strains.

Another interesting aspect is that the genome mining data provide an overview of the distribution of cochonodin-like lasso peptides in microorganisms whose genomes have already been sequenced (SI Figure 4A, SI Table S2, and SI Table S3). Intriguingly, such lasso peptide BGCs seem to be commonly associated with human and animal microbiota, including both potential pathogens as well as commensal bacteria. Thus, it would be worthwhile to direct future efforts towards either improving heterologous cochonodin I production or finding a native producer that yields high amounts of such lasso peptides under laboratory conditions so that their isolation and investigation of their biological activities becomes possible. These specialized metabolites may be important virulence factors or could even represent new antimicrobials, as may be inferred from the presence of the dedicated ABC transporter domain in SsuB2/D, as well as a NisFEG-like immunity cassette.

EXPERIMENTAL SECTION

General Experimental Procedures. Oligonucleotide primers were obtained from Millipore Sigma. Phusion DNA polymerase, restriction enzymes, and T4 DNA ligase were purchased from New England Biolabs. XAD16 resin was bought from Millipore Sigma. DNA sequences of

cloned or mutated plasmids were confirmed by dideoxy sequencing carried out by GATC Biotech.

High-resolution MS analysis of the extracts and the combined heat and carboxypeptidase Y stability assays was carried out on an Thermo Fisher Scientific LTQ Orbitrap XL instrument that was coupled with an Agilent 1200 HPLC system using a Grom-Sil-120-ODS-4-HE column (50 mm in length, 2 mm in diameter) for separation.

Tandem MS/MS CID/ECD experiments were carried out on a custom built nanoESI-TIMS-q-EMS-CC-ToF MS/MS (Bruker Daltonics Inc., Billerica, MA) instrument.[47] The nanoESI emitters were pulled from quartz capillaries (O.D. = 1.0 mm and I.D. = 0.70 mm) using a Sutter Instrument Co. P2000 laser puller (Sutter Instruments, Novato, CA). Peptide sample solutions were loaded in a pulled-tip capillary, housed in a mounted custom built XYZ stage in front of the MS inlet, and sprayed at 850 V via a tungsten wire inserted inside the nanoESI emitters.

A custom built electromagnetostatic cell[48] (EMS, e-MSion Inc., Corvallis, OR), capable of performing ECD without the need for long reaction times or ultrahigh vacuum, was mounted at the quadrupole exit before a shortened collision cell. The 19 mm long EMS cell was composed of seven cylindrical electrostatic lenses (L1-L7), two ring magnets and a heated rhenium filament (Scientific Instrument Services, Ringoes, NJ) housed in L4, where electrons were generated at the center of the cell. Electrons were confined along the ion longitudinal axis. The filament was operated at a current of 2.5 A. The electrostatic lenses applied to the EMS cell were tuned to get optimal ECD fragmentation events. The collision cell was operated using high purity argon (oxygen free) to enhance the cooling of the ions.

Collision-induced dissociation (CID) experiments were performed in the collision cell located after the EMS cell (operated in transmission mode). The tandem MS/MS CID spectra of the

$[M+2H]^{2+}$ ions were obtained using argon as the collision gas with the collision cell operated at a collision voltage of ~50 V.

Bacterial Strains. Cloning and mutagenesis steps were accomplished using *E. coli* TOP10 cells, while heterologous expression experiments were carried out with *E. coli* BL21(DE3).

Generation of the Cochonodin Production Plasmids. The plasmid *ssuA1-A2-C-B1-B2/D* pET41a was generated via restriction enzyme-based cloning. The cochonodin gene cluster was amplified with Phusion DNA polymerase using the primers listed in SI Table S4, which introduced the NcoI and XhoI cut sites used for cloning. Amplified DNA was purified via agarose gel electrophoresis and isolated with a QIAGEN gel extraction kit using the manufacturer's protocol. Next, the purified DNA and empty pET41a vector were treated with NcoI and XhoI, again purified via agarose gel electrophoresis, and then extracted from the respective gel slices. The *ssuA1-A2-C-B1-B2/D* pET41a plasmid was then assembled via T4 DNA ligase treatment and the ligation reaction was transformed into competent *E. coli* TOP10 cells. Colonies were selected on LB agar plates containing 50 μg/mL kanamycin and the sequences of isolated plasmids were confirmed via dideoxy DNA sequencing.

Mutagenesis of *ssuA1-A2-C-B1-B2/D* pET41a was carried out using 5'-phosphorylated primers (SI Table S5) and blunt end ligation. Here, the primers annealed to the site of the plasmid where the RBS sequence was supposed to be incorporated and/or adjacent to the region of the plasmid that was supposed to be deleted. Amplified 5'-phosphorylated DNA was purified via agarose gel electrophoresis, extracted from the gel slice, and then the blunt end ligation was accomplished via treatment with T4 DNA ligase. After transformation of the ligation reactions into *E. coli* TOP10 cells and selection on kanamycin-containing LB agar, plasmids were isolated

from single colony overnight cultures and their sequences were confirmed via dideoxy sequencing.

The *ssuA1-A2-C-B1-B2/D* pET41a plasmid was converted into *ssuA1*_RBS_*C-B1-B2/D* pET41a by use of the primers FP_RBS-ssuC and RP_ssuA1-RBS and into *ssuA1-A2*_RBS_*C-B1-B2/D* pET41a by use of the primers FP_RBS-ssuC and RP_ssuA1/A2-RBS. The primers FP_ΔssuA1 and RP_ΔssuA1 were then employed to convert *ssuA1-A2*_RBS_*C-B1-B2/D* pET41a into *ssuA2*_RBS_*C-B1-B2/D* pET41a, whereas the primers FP_ssuB2(145Stop) and RP_ssuB2(145Stop) enabled the conversion of *ssuA1*_RBS_*C-B1-B2/D* pET41a into *ssuA1*_RBS_*C-B1-B2(145Stop)* pET41a.

Heterologous Production of Cochonodin I in *E. coli* BL21(DE3). Production tests for each construct were carried out in 1 L aliquots of M9 minimal medium (8.9 g/L $Na_2HPO_4 \cdot 2\ H_2O$, 3.0 g/L KH_2PO_4, 0.5 g/L NaCl, 1.0 g/L NH_4Cl, 1.0 mL/L $MgSO_4$ (2 M in dH_2O), 0.2 mL/L $CaCl_2$ (0.5 M in dH_2O), pH 7.0; after autoclaving, 10 mL/L sterile 40% (m/V in dH_2O) glucose solution, 2.0 mL/L sterile M9 vitamin mix (see SI Table S6), and 1 mL/L sterile kanamycin stock solution (50 mg/mL in H_2O) were added).

For the lasso peptide production, the medium was prewarmed for 1 h at 37 °C, then inoculated 1:100 with an LB overnight culture, and subsequently grown at 37 °C until an optical density at 600 nm (OD_{600}) of ~0.4-0.5 was reached. The medium was then moved into another shaker that was already cooled to 18 °C and incubated for another hour. At this point, the culture had reached an OD_{600} ~0.5-0.7 and the expression was then induced by addition of 0.2 mL of a 0.5 M sterile isopropyl β-D-1-thiogalactopyranoside (IPTG) solution. Cultures were subsequently shaken for 3 days at 18 °C and then harvested by centrifugation.

The cell pellet from 1 L of culture was resuspended in 50 mL of MeOH and extracted by shaking overnight at 4 °C. On the next day, the pellet extract was clarified by centrifugation. Next, the cleared pellet extract was completely dried under reduced pressure and then resuspended in 1 mL of a 50% (V/V) dH_2O:MeOH mixture. Before further analysis, the concentrated extract was again clarified by centrifugation.

The 1 L of culture SN was extracted by addition of 10 mL of a suspension of XAD16 resin (50 g of XAD16 resin mixed with 200 mL of dH_2O), then stirring slowly for 1 h at room temperature (RT), and finally collecting the resin with the bound peptide by filtration. After washing the resin thrice with 5 mL of dH_2O, the SN extract was eluted from the resin through stepwise addition of a total of 50 mL of MeOH. Next, the SN extract was completely dried under reduced pressure, resuspended in 1 mL of a 50% (V/V) dH_2O:MeOH mixture, and clarified by centrifugation before subsequent use.

Heat and Carboxypeptidase Y Stability Assays of Cochonodin I. For these assays, 200 μL of SN extract from an *ssuA1*_RBS_*C-B1-B2/D* pET41a *E. coli* BL21(DE3) culture were diluted with 1.3 mL of dH_2O, frozen in liquid nitrogen, and then freeze-dried in a lyophilizer. The dried sample was next resuspended in 200 μL of MES buffer (50 mM MES, 1 mM $CaCl_2$, pH = 6.75) and divided into four 50 μL aliquots.

One aliquot was kept untreated for reference. One aliquot was directly treated with carboxypeptidase Y by addition of 10 μL of a carboxypeptidase Y stock solution (1 mg/mL in MES buffer) and, after mixing, incubation overnight at RT. The remaining two aliquots were incubated for 4 h at 95 °C in a thermocycler with a heated lid. One of these aliquots was subsequently also treated with carboxypeptidase Y as described.

For quenching of the peptidase reactions, 5 µL of pure MeOH were added to a sample at the next morning and, after mixing, the precipitated protein was removed by centrifugation. For comparability, 10 µL of plain MES buffer and 5 µL of pure MeOH were also added to the samples that were not treated with carboxypeptidase Y. For LC-MS analysis, 50 µL of each sample were used.

High-Resolution LC-MS Analysis of Extracts and Combined Heat and Carboxypeptidase Y Stability Assays. For high-resolution LC-MS analysis, the following gradient of water/0.1% formic acid (solvent A) and MeCN/0.1% formic acid (solvent B) at a flow rate of 0.2 mL/min was employed: holding 2% B for 2 min; linear increase from 2% to 30% B in 18 min; linear increase from 30% to 95% B in 15 min; holding 95% B for 2 min. For the initial measurements of the pellet and SN extracts as well as for the combined heat and carboxypeptidase Y stability assays, 50 µL of sample were injected. In this way, the production of cochonodin 1 (-2 aa) ($C_{95}H_{147}N_{25}O_{22}$, calculated M = 1990.1147 g/mol, calculated $[M+2H]^{2+}$ = 996.0652) was verified by detection of the $[M+2H]^{2+}$ product ion.

Bioinformatic Analysis of the SsuB2/D Protein. The putative positions of transmembrane helices in SsuB2/D were analyzed with the DeepTMHMM webtool (SI Figure S1A).[49] The first transmembrane helix was thereby predicted to start at position 178, which implies that the leader peptidase domain must reside within the first 177 residues of SsuB2/D. To further identify a suitable site to separate the peptidase and ABC transporter domains, the first 177 residues of SsuB2/D were aligned with other known B2 proteins (SI Figure S1B) and the secondary structure motifs of these B2 proteins and the leader peptidase domain of SsuB2/D were predicted using the PsiPred webtool (SI Figure S1B).[50] The B2 protein sequences employed were taken from the chaxapeptin (CptB2),[36] lariatin (LarB2),[37] and paeninodin (PadeB2) BGCs.[22] Based on

these data, it was decided to remove the ABC transporter domain from SsuB2/D by deletion of every residue following position 145. The bioinformatic data on which this decision was based on are summarized in SI Figure S1.

Genome Mining for Cochonodin Homologs. For genome mining, the SsuB2/D sequence was used as input for two rounds of a position-specific iterated protein BLAST (PSI-BLAST)[45] search against public genomic databases. For removal of false positive hits that were other ABC transporters lacking the N-terminal leader peptidase domain, any hit with a query coverage <80% was immediately discarded. To facilitate the identification of lasso peptide precursors and the other processing enzymes in proximity of the remaining hits, their accession numbers were used as input for the lasso peptide scoring module of the RODEO webtool.[6] The RODEO output was further manually curated to identify the correct precursor sequences, to double check the correct annotation of the leader and core peptide regions in the identified precursors, and to check for the presence of NisF, NisE, NisG, NisR, and NisK homologs. To ensure that no B2/D protein was overlooked, the search was repeated with a B2/D protein sequence from a non-*Streptococcus* species as BLAST input, namely with the *R. albus* 8 B2/D protein, which exhibits 50.3% identity and 73.1% similarity with SsuB2/D. However, no additional B2/D homologs in comparison to the SsuB2/D-based search were found. SI Table S2 contains a summary of all lasso peptide BGCs identified by this approach.

Structure Similarity Network Analysis. For a better comparison of the lasso peptides arising from the identified SsuA1 homologs, a *sequence similarity network* (SSN) analysis was performed via the *Enzyme Function Initiative - Enzyme Similarity Tool* (ESI-EST)[46] using only the core peptide regions of the identified precursor peptides and setting the alignment score threshold to 7.

Conserved Motif Prediction and Protein Sequence Comparison. For prediction of conserved motifs in the leader and core peptide regions of the identified precursors, the *Multiple Expectation Maximizations for Motif Elicitation* (MEME) algorithm and the corresponding MEME webtool were used.[51] General motif logos were based on 95 out of 95 leader sequences and 85 out of 95 core peptide sequences. The clade-specific core peptide motifs were based on 56 out of 58 clade 1, 18 out of 18 clade 2, and 5 out of 5 clade 3 core peptide sequences.

For direct comparison of protein sequences, the EMBOSS Needle webtool was employed.[52]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: .

sequence alignments, table with sequence alignment scores, secondary structure predictions, additional LC-MS spectra of tested production plasmids, heat and carboxypeptidase Y stability assays, tables and figures with information about the cochonodin-like BGCs identified through genome mining, figures with conserved sequence motifs based on the genome mining results, tables with oligonucleotide primer sequences, recipe for the M9 vitamin mix

AUTHOR INFORMATION

Corresponding Author

* E-mail (J. D. Hegemann): jdhegemann@googlemail.com.

ORCID

Julian D. Hegemann: 0000-0002-3859-8744

Kevin Jeanne Dit Fouque: 0000-0002-2974-1350

Miguel Santos-Fernandez: 0000-0003-4233-0203

Francisco Fernandez-Lima: 0000-0002-1283-4390

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Prof. Duncan Maskel (formerly at the University of Cambridge (UK), currently at the University of Melbourne (Australia)) for kindly providing *S. suis* LSS65 gDNA to us. The authors at FIU further acknowledge the financial support from the National Science Foundation Division of Chemistry, under CAREER award CHE-1654274, with co-funding from the Division of Molecular and Cellular Biosciences to FFL and funding from National Institutes of General Medicine (R01GM134247).

REFERENCES

(1) Hegemann, J. D.; Zimmermann, M.; Xie, X.; Marahiel, M. A. *Acc. Chem. Res.* **2015**, *48*, 1909-1919.

(2) Hegemann, J. D. *ChemBioChem* **2020**, *21*, 7-18.

(3) Hegemann, J. D.; Dit-Foque, K. J.; Xie, X. In *Comprehensive Natural Products III*; Liu, H.-W., Begley, T. P., Eds.; Elsevier: Oxford, **2020**, p 206-228.

(4) Pati, N. B.; Doijad, S. P.; Schultze, T.; Mannala, G. K.; Yao, Y.; Jaiswal, S.; Ryan, D.; Suar, M.; Gwozdzinski, K.; Bunk, B.; Mraheil, M. A.; Marahiel, M. A.; Hegemann, J. D.; Sproer, C.; Goesmann, A.; Falgenhauer, L.; Hain, T.; Imirzalioglu, C.; Mshana, S. E.; Overmann, J.; Chakraborty, T. *Sci. Rep.* **2018**, *8*, 5392.

(5) Guerrero-Garzón, J. F.; Madland, E.; Zehl, M.; Singh, M.; Rezaei, S.; Aachmann, F. L.; Courtade, G.; Urban, E.; Rückert, C.; Busche, T.; Kalinowski, J.; Cao, Y. R.; Jiang, Y.; Jiang, C. L.; Selivanova, G.; Zotchev, S. B. *iScience* **2020**, *23*, 101785.

(6) Tietz, J. I.; Schwalen, C. J.; Patel, P. S.; Maxson, T.; Blair, P. M.; Tai, H. C.; Zakai, U. I.; Mitchell, D. A. *Nat. Chem. Biol.* **2017**, *13*, 470-478.

(7) Fage, C. D.; Hegemann, J. D.; Nebel, A. J.; Steinbach, R. M.; Zhu, S.; Linne, U.; Harms, K.; Bange, G.; Marahiel, M. A. *Angew. Chem. Int. Ed.* **2016**, *55*, 12717-12721.

(8) Zong, C.; Wu, M. J.; Qin, J. Z.; Link, A. J. *J. Am. Chem. Soc.* **2017**, *139*, 10403-10409.

(9) DiCaprio, A. J.; Firouzbakht, A.; Hudson, G. A.; Mitchell, D. A. *J. Am. Chem. Soc.* **2019**, *141*, 290-297.

(10) Long, T.; Liu, L.; Tao, Y.; Zhang, W.; Quan, J.; Zheng, J.; Hegemann, J.; Uesugi, M.; Yao, W.; Tian, H.; Wang, H. *Angew. Chem. Int. Ed.* **2021**, *60*, 13414-13422.

(11) Zyubko, T.; Serebryakova, M.; Andreeva, J.; Metelev, M.; Lippens, G.; Dubiley, S.; Severinov, K. *Chem. Sci.* **2019**, *10*, 9699-9707.

(12) Yan, K. P.; Li, Y.; Zirah, S.; Goulard, C.; Knappe, T. A.; Marahiel, M. A.; Rebuffat, S. *ChemBioChem* **2012**, *13*, 1046-1052.

(13) Zhu, S.; Fage, C. D.; Hegemann, J. D.; Mielcarek, A.; Yan, D.; Linne, U.; Marahiel, M. A. *Sci. Rep.* **2016**, *6*, 35604.

(14) Hegemann, J. D.; Schwalen, C. J.; Mitchell, D. A.; van der Donk, W. A. *Chem. Commun.* **2018**, *54*, 9007-9010.

(15) Sumida, T.; Dubiley, S.; Wilcox, B.; Severinov, K.; Tagami, S. *ACS Chem. Biol.* **2019**, *14*, 1619-1627.

(16) Choudhury, H. G.; Tong, Z.; Mathavan, I.; Li, Y.; Iwata, S.; Zirah, S.; Rebuffat, S.; van Veen, H. W.; Beis, K. *Proc. Natl. Acad. Sci.* **2014**, *111*, 9145-9150.

(17) Romano, M.; Fusco, G.; Choudhury, H. G.; Mehmood, S.; Robinson, C. V.; Zirah, S.; Hegemann, J. D.; Lescop, E.; Marahiel, M. A.; Rebuffat, S.; De Simone, A.; Beis, K. *ACS Chem. Biol.* **2018**, *13*, 1598-1609.

(18) Smits, S. H. J.; Schmitt, L.; Beis, K. *FEBS Lett.* **2020**, *594*, 3920-3942.

(19) Beis, K.; Rebuffat, S. *Res. Microbiol.* **2019**, *170*, 399-406.

(20) Maksimov, M. O.; Pelczer, I.; Link, A. J. *Proc. Natl. Acad. Sci.* **2012**, *109*, 15223-15228.

(21) Sabino, Y. N. V.; de Araújo, K. C.; de Assis, F.; Moreira, S. M.; Lopes, T. D. S.; Mendes, T. A. O.; Huws, S. A.; Mantovani, H. C. *Front. Microbiol.* **2020**, *11*, 576738.

(22) Zhu, S.; Hegemann, J. D.; Fage, C. D.; Zimmermann, M.; Xie, X.; Linne, U.; Marahiel, M. A. *J. Biol. Chem.* **2016**, *291*, 13662-13678.

(23) Zhu, S.; Fage, C. D.; Hegemann, J. D.; Yan, D.; Marahiel, M. A. *FEBS Lett.* **2016**, *590*, 3323-3334.
(24) Bobeica, S. C.; Dong, S. H.; Huo, L.; Mazo, N.; McLaughlin, M. I.; Jimenez-Oses, G.; Nair, S. K.; van der Donk, W. A. *Elife* **2019**, *8*, e42305.
(25) Kieuvongngam, V.; Olinares, P. D. B.; Palillo, A.; Oldham, M. L.; Chait, B. T.; Chen, J. *Elife* **2020**, *9*, e51492.
(26) Rahman, S.; McHaourab, H. S. *J. Biol. Chem.* **2020**, *295*, 14678-14685.
(27) Clarke, D. J.; Campopiano, D. J. *Org. Biomol. Chem.* **2007**, *5*, 2564-2566.
(28) Repka, L. M.; Chekan, J. R.; Nair, S. K.; van der Donk, W. A. *Chem. Rev.* **2017**, *117*, 5457-5520.
(29) Chen, J.; van Heel, A. J.; Kuipers, O. P. *mBio* **2020**, *11*, e02825-02820.
(30) Lubelski, J.; Rink, R.; Khusainov, R.; Moll, G. N.; Kuipers, O. P. *Cell Mol. Life Sci.* **2008**, *65*, 455-476.
(31) de Ruyter, P. G.; Kuipers, O. P.; Beerthuyzen, M. M.; van Alen-Boerrigter, I.; de Vos, W. M. *J. Bacteriol.* **1996**, *178*, 3434-3439.
(32) Kuipers, O. P.; Beerthuyzen, M. M.; de Ruyter, P. G.; Luesink, E. J.; de Vos, W. M. *J. Biol. Chem.* **1995**, *270*, 27299-27304.
(33) Hegemann, J. D.; Süssmuth, R. D. *RSC Chem. Biol.* **2020**, *1*, 110-127.
(34) Hegemann, J. D.; Zimmermann, M.; Zhu, S.; Klug, D.; Marahiel, M. A. *Biopol. Pept. Sci.* **2013**, *100*, 527-542.
(35) Bratovanov, E. V.; Ishida, K.; Heinze, B.; Pidot, S. A.-O.; Stinear, T. P.; Hegemann, J. A.-O.; Marahiel, M. A.; Hertweck, C. A.-O. X. *ACS Chem. Biol.* **2020**, *15*, 1169-1111.
(36) Martin-Gomez, H.; Linne, U.; Albericio, F.; Tulla-Puche, J.; Hegemann, J. D. *J. Nat. Prod.* **2018**, *81*, 2050-2056.
(37) Inokoshi, J.; Matsuhama, M.; Miyake, M.; Ikeda, H.; Tomoda, H. *Appl. Microbiol. Biotechnol.* **2012**, *95*, 451-460.
(38) Delgado, M. A.; Vincent, P. A.; Farías, R. N.; Salomón, R. A. *J. Bacteriol.* **2005**, *187*, 3465-3470.
(39) Dit Fouque, K. J.; Scutelnic, V.; Hegemann, J. D.; Rebuffat, S.; Maître, P.; Rizzo, T. R.; Fernandez-Lima, F. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1096-1104.
(40) Jeanne Dit Fouque, K.; Bisram, V.; Hegemann, J. D.; Zirah, S.; Rebuffat, S.; Fernandez-Lima, F. *Anal. Bioanal. Chem.* **2019**, *411*, 6287-6296.
(41) Jeanne Dit Fouque, K.; Lavanant, H.; Zirah, S.; Hegemann, J. D.; Fage, C. D.; Marahiel, M. A.; Rebuffat, S.; Afonso, C. *Analyst* **2018**, *143*, 1157-1170.
(42) Dit Fouque, K. J.; Moreno, J.; Hegemann, J. D.; Zirah, S.; Rebuffat, S.; Fernandez-Lima, F. *Anal. Chem.* **2018**, *90*, 5139-5146.
(43) Fouque, K. J. D.; Lavanant, H.; Zirah, S.; Hegemann, J. D.; Zimmermann, M.; Marahiel, M. A.; Rebuffat, S.; Afonso, C. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 315-322.
(44) Knappe, T. A.; Linne, U.; Robbel, L.; Marahiel, M. A. *Chem. Biol.* **2009**, *16*, 1290-1298.
(45) Boratyn, G. M.; Camacho, C.; Cooper, P. S.; Coulouris, G.; Fong, A.; Ma, N.; Madden, T. L.; Matten, W. T.; McGinnis, S. D.; Merezhuk, Y.; Raytselis, Y.; Sayers, E. W.; Tao, T.; Ye, J.; Zaretskaya, I. *Nucleic Acids Res.* **2013**, *41*, W29-33.
(46) Gerlt, J. A.; Bouvier, J. T.; Davidson, D. B.; Imker, H. J.; Sadkhin, B.; Slater, D. R.; Whalen, K. L. *Biochim. Biophys. Acta* **2015**, *1854*, 1019-1037.
(47) Jeanne Dit Fouque, K.; Kaplan, D.; Voinov, V. G.; Holck, F. H. V.; Jensen, O. N.; Fernandez-Lima, F. *Anal. Chem.* **2021**, *27*, 9575-9582.

(48) Voinov, V. G.; Beckman, J. S.; Deinzer, M. L.; Barofsky, D. F. *Rapid. Commun. Mass Spectrom.* **2009**, *23*, 3028-3030.
(49) Hallgren, J.; Tsirigos, K. D.; Almagro Armenteros, J. J.; Marcatili, P.; Nielsen, H.; Krogh, A.; Winther, O. *DeepTMHMM* **2021**, https://biolib.com/DTU/DeepTMHMM.
(50) Buchan, D. W. A.; Jones, D. T. *Nucleic Acids Res.* **2019**, *47*, W402-W407.
(51) Bailey, T. L.; Boden, M.; Buske, F. A.; Frith, M.; Grant, C. E.; Clementi, L.; Ren, J.; Li, W. W.; Noble, W. S. *Nucleic Acids Res.* **2009**, *37*, W202-208.
(52) Madeira, F.; Park, Y. M.; Lee, J.; Buso, N.; Gur, T.; Madhusoodanan, N.; Basutkar, P.; Tivey, A. R. N.; Potter, S. C.; Finn, R. D.; Lopez, R. *Nucleic Acids Res.* **2019**, *47*, W636-W641.