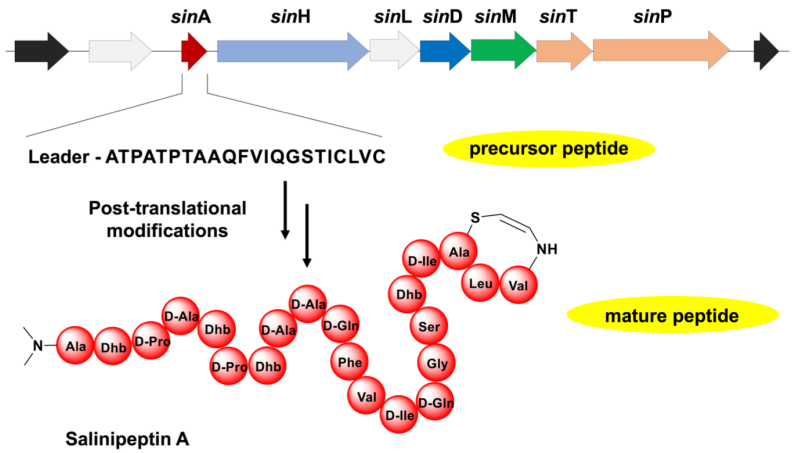
Article

Salinipeptins: Integrated genomic and chemical approaches reveal unusual D-amino acid-containing ribosomally synthesized and post-translationally modified peptides (RiPPs) from a Great Salt Lake Streptomyces sp.

Zhuo Shang, Jaclyn M. Winter, Christopher A. Kauffman, Inho Yang, and William Fenical

ACS Chem. Biol., **Just Accepted Manuscript** • Publication Date (Web): 12 Feb 2019

TOC Graphic:


sinA
sinH
sinL
sinD
sinM
sinT
sinP
Leader - ATPATPTAAQFVIQGSTICLVC
precursor peptide
Post-translational modifications
mature peptide
Salinipeptin A

Salinipeptins: Integrated genomic and chemical approaches reveal unusual D-amino acid-containing ribosomally synthesized and post-translationally modified peptides (RiPPs) from a Great Salt Lake *Streptomyces* sp.

Zhuo Shang,[‡] Jaclyn M. Winter,[*,§] Christopher A. Kauffman,[§] Inho Yang,[‡] William Fenical[*,‡]

[‡] Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093-0204

[§] Department of Medicinal Chemistry, University of Utah, Salt Lake City, UT 84112

KEYWORDS: RiPPs, linaridin, Great Salt Lake actinomycetes, D-amino acids, imidazolidine

ABSTRACT

Analysis of the full genome of an environmentally-unique, halotolerant *Streptomyces* sp. strain GSL-6C, isolated from the Great Salt Lake, revealed a gene cluster encoding the biosynthesis of the salinipeptins, D-amino acid containing members of the rare linaridin subfamily of ribosomally synthesized and post-translationally modified peptides (RiPPs). The sequence organization of the unmodified amino acid residues in salinipeptins A–D (**1**–**4**) were suggested by genome annotation, and subsequently their sequence and post-translational modifications were defined using a range of spectroscopic techniques and chemical derivatization approaches. The salinipeptins are unprecedented linaridins bearing nine D-amino acids, which are uncommon in RiPP natural products and are the first reported in the linaridin subfamily. Whole genome mining of GSL-6C did not reveal any homologs of the reported genes responsible for amino acid epimerization in RiPPs, inferring new epimerases may be involved in the conversion of L- to D-amino acids. In addition, the *N*-oxide and dimethylimidazolidin-4-one moieties in salinipeptins B and C, which are modified from *N*,*N*-dimethylalanine, are unknown in bacterial peptides. The three-dimensional structure of salinipeptin A, possessing four loops generated by significant hydrogen bonding, was established based on observed nuclear Overhauser effect (NOE) correlations. This study demonstrates that integration of genomic information early in chemical analysis significantly facilitates the discovery and structure characterization of novel microbial secondary metabolites.

INTRODUCTION

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a large group of biomedically important natural products of ribosomal origin, which comprises at least 22 subfamilies with a wide variety of structural features, biological activities and ecological functions.[1,2] Among these peptide natural products, the linaridin class is a rare linear RiPP subfamily with extensive enzymatic post-translational modifications (PTMs). At present, this subfamily is composed of only three characterized members, cypemycin,[3] grisemycin[4] and legonaridin (Figure 1).[5] The reported PTMs in the linaridin peptides include dehydration,

methylation, decarboxylation and epimerization, which can produce modified amino acids such as dehydrobutyrine (Dhb), *N*,*N*-dimethyl-alanine/isoleucine, dehydroalanine (Dha), aminovinyl-cysteine (AviCys) and L-*allo*-isoleucine residues. The linaridin class is the only RiPP subfamily subjected to *N*-terminal methylation, which is essential for its antibiotic activity.[6] Besides the uniqueness of their chemical structures, most of the tailoring enzymes and catalytic mechanisms involved in linaridin biosynthesis have been shown to be unprecedented in the RiPP class of natural products. For example, the putative dehydratase CypH is involved in the formation of Dhb in cypemycin biosynthesis and the decarboxylase CypD was shown to catalyze the oxidative decarboxylation of the *C*-terminal Cys to form a thioenol, which can undergo subsequent reactions to produce the AviCys moiety.[6-9] Furthermore, the founding linaridin peptide, cypemycin is a highly selective and potent antibiotic against *Micrococcus luteus* and mouse P388 leukemia cells.[10] A recent genome mining effort revealed more than 50 putative linaridin biosynthetic gene clusters (BGCs) in the sequenced bacteria deposited in GenBank.[7] However, the types of PTMs and the structures and functions of the mature peptides encoded by these putative linaridin BGCs remain undefined. Thus, the huge disparity between the numbers of predicted linaridin BGCs and those that have been characterized significantly impedes our understanding of their chemical diversity and biosynthesis, as well as their ecological and biomedical applications. Herein, we report the identification of a linaridin BGC in a Great Salt Lake-derived *Streptomyces* and the isolation and structural characterization of four new linaridins, the salinipeptins A-D (**1**-**4**) displaying unprecedented PTMs.

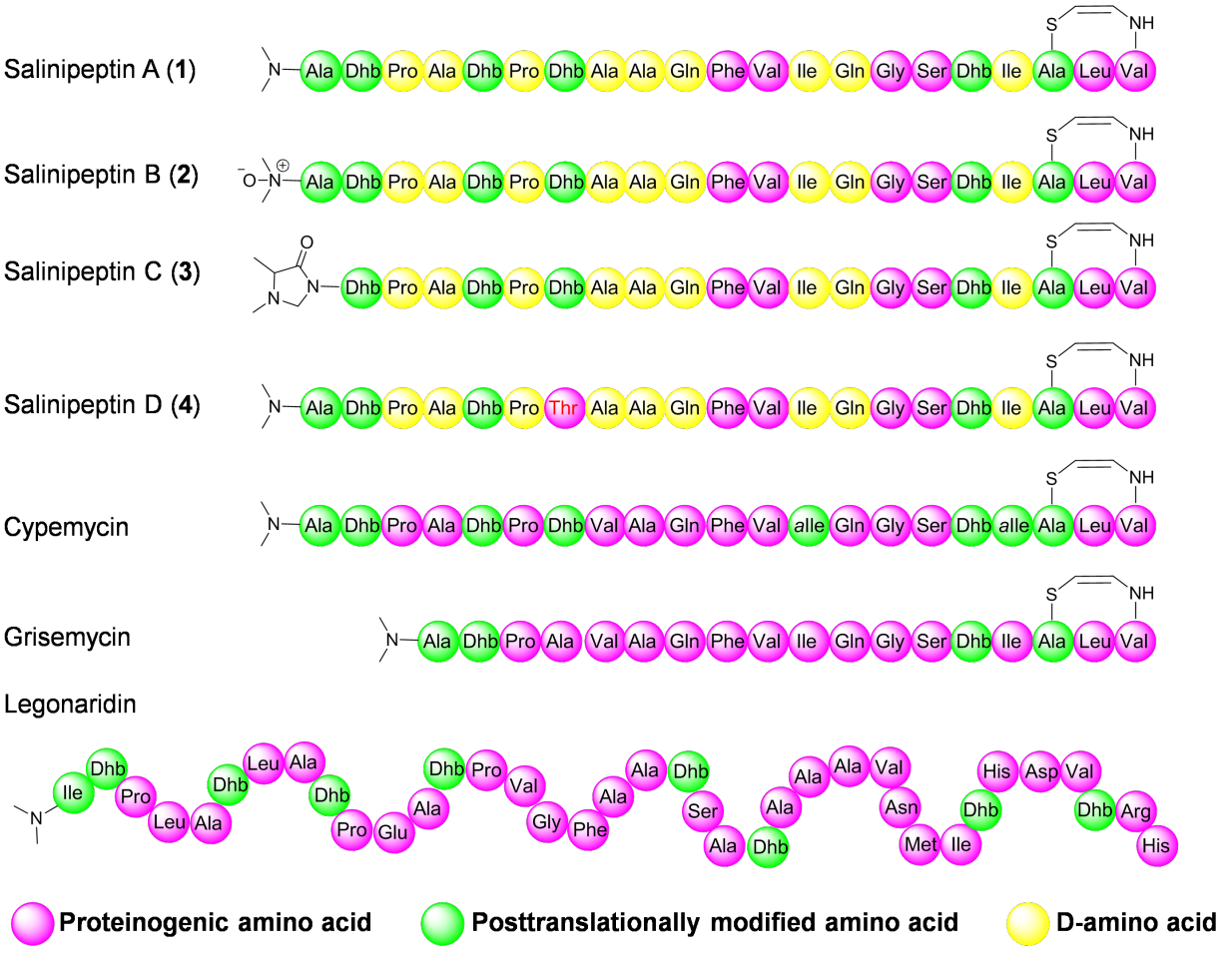


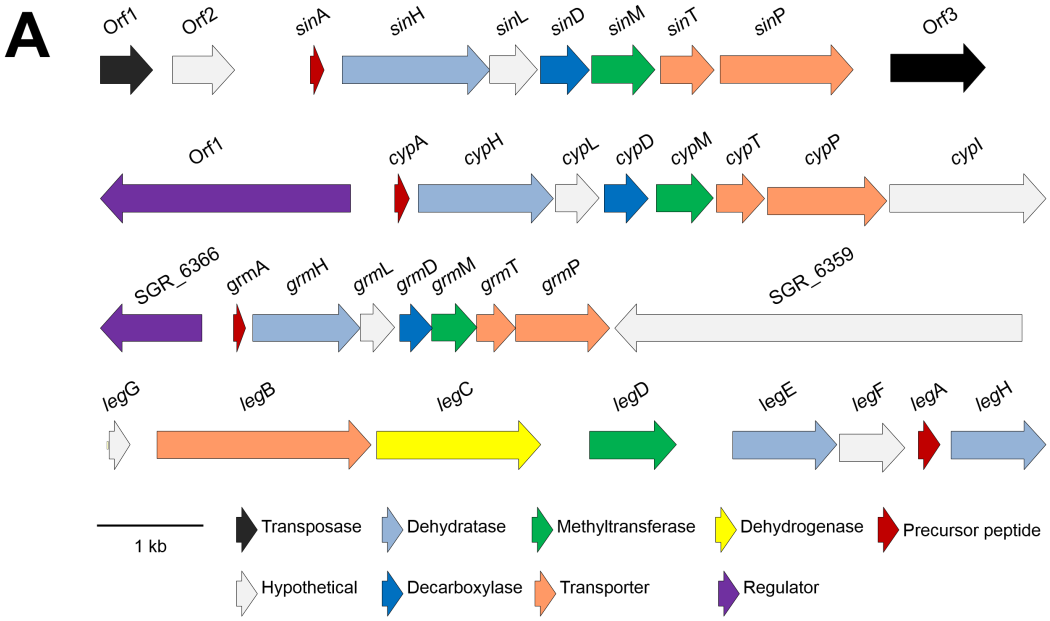
Figure 1. Structures of the salinipeptins A-D (**1**-**4**) and other known RiPPs belonging to the linaridin subfamily. The absolute configurations for **1**-**4** and legonaridin have been determined by advanced Marfey's analysis. However, the absolute configuration of the *N*-

terminal Ala residue in **1**-**4** could not been determined by advanced Marfey's method. It has been reported that cypemycin contains all L-amino acids and the structure is shown to reflect this. Due to insufficient quantities during its isolation, the absolute configuration of amino acids in grisemycin is still unknown.

RESULTS AND DISCUSSION

Salinipeptin Gene Cluster Analysis

We recently initiated a natural product drug discovery campaign aimed at the isolation and examination of halophilic bacteria isolated from the uniquely hypersaline environment of the Great Salt Lake (GSL). In this study ten bacterial strains belonging to the order Actinomycetales were isolated from shallow sediment samples collected from Farmington Bay. 16S rRNA gene sequencing was used to initially identify all strains. Curiously, by their 16S sequences, these bacteria were all closely related to bacteria isolated from marine environments world-wide (Table S1). Concurrent with the initial chemical analysis, which is discussed below, the genome of *Streptomyces* sp. GSL-6C was sequenced and assembled. By 1D and 2D NMR experiments, the structural fragment Gly-Ser-Dhb-Ile-AviCys was elucidated. As this substructure was also reported in cypemycin, we speculated that the peptide was a cypemycin congener and that the amino acid residues Gly-Ser-Thr-Ile-Cys, prior to post-translationally modifications, should exist in a precursor peptide. Genome mining using the structural fragment Gly-Ser-Thr-Ile-Cys led to the identification of a 10,208 bp BGC containing seven open reading frames, hereby named the *sin* cluster (GenBank accession MG788286), that is similarly organized to the BGCs responsible for cypemycin, grisemycin and legonaridin biosynthesis (Figure 2A, Figure S3 and Table S2). *sin*A codes for the precursor peptide and its core sequence shows high sequence similarity with *cyp*A, with the only difference being the swapping of Val-8 in cypemycin with Ala-8 (Figure 2B and Table S3), suggesting the salinipeptins are new members of the linaridin RiPP subfamily. Additional genes associated with post-translational modifications, including a putative dehydratase *sin*H and methyltransferase *sin*M, are encoded in the *sin* cluster and are predicted to modify the precursor peptide to yield the non-proteinogenic amino acids Dhb and *N*,*N*-dimethylalanine (Me_2Ala), respectively. Additionally, the *sin* cluster contains a putative decarboxylase gene *sin*D that is homologous to *cyp*D. In cypemycin biosynthesis, AviCys is formed between two cysteine residues and *Cyp*D was recently shown to catalyze the oxidative decarboxylation of the *C*-terminal cysteine residue,[8,9] which is subsequently processed to form AviCys. To date, the AviCys residue is a terminal functionality only reported in lanthipeptide and linaridin RiPP subfamilies.[11]



B

```
SinA  VRSELTSTRPDSSATHPESLAAQEFANTALGAATPGFHADCETPAMATP----ATPTAAQ--FVIQGSTICLVC---------
CypA  VRSEMTLTSTNSA----EALAAQDFANTVLSAAAPGFHADCETPAMATP----ATPTVAQ--FVIQGSTICLVC---------
GrmA  MRLD-SIATQETATALPESMATQDFANSVLAGAVPGFHSDAETPAM-------ATPAVAQ--FVIQGSTICLVC---------
LegA  ------------------MSVLAEFANTELVDVEPGRLGSEATPTMITPLATLATPEATPVGFAATSATAAAVNMITHDVTRH
```

Figure 2. Organization of linaridin BGCs. **A**) Organization of the 10,208 bp *sin* gene cluster and comparison of its gene organization to other linaridin BGCs. *cyp*, cypemycin identified from *S.* sp. OH-4156; *grm*, grisemycin identified from *S. griseus* IFO 13350; *leg*, legonaridin identified in *S.* sp. CT34. **B**) Sequences of the precursor peptides for salinipeptins, cypemycin, grisemycin and legonaridin. The red arrow indicates where the leader sequence is cleaved to yield the mature peptide.

Isolation and Structure Elucidation of Salinipeptins A–D (1–4)

HPLC-DAD-HRMS/MS analysis of the EtOAc extracts of the ten GSL bacterial strains revealed that three strains (i.e., GSL-6C, GSL-9, GSL-20) produce similar peptide-derived natural products with observed molecular weights ranging from 2,000–2,100 Da (Figure S1). This conclusion is based on accurate mass values and isotopic distributions of pseudo-molecular parent ions (always larger than 1,000 Da with a cluster of multiple isotopic peaks) as well as MS/MS fragmentation patterns (neutral losses for amino acid residues). Natural product databases (e.g., MarinLit, AntiBase, and Dictionary of Natural Products) were searched and failed to show any known peptide-derived metabolites with the observed molecular weights. To isolate and identify the peptides, *Streptomyces* sp. GSL-6C was cultivated in 60 x 1 L scale in a saline nutrient medium for seven days, followed by EtOAc extraction and chromatographic fractionation (Figure S4). Four unusual D-amino acid-containing RiPPs, salinipeptins A-D (**1-4**), of the rare linaridin subfamily were obtained and fully structurally characterized using integrated genomic and chemical approaches (Figure 1).

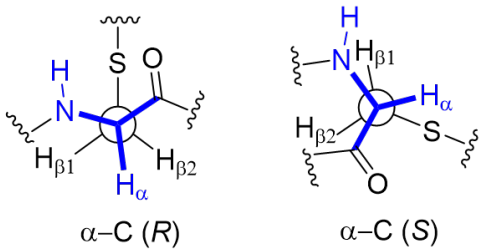
The core peptide sequence derived from the gene cluster was effectively used as the starting point for the structure elucidation of the modified peptides. The major peptide, salinipeptin A (**1**), was isolated as white amorphous powder. HRESI (+) MS analysis of **1** showed a doubly protonated pseudo-molecular ion at *m/z* 1034.5577 $[M+2H]^{2+}$, indicative of the molecular formula $C_{97}H_{150}N_{24}O_{24}S$ (Δppm +1.63). The ^{1}H NMR spectral data for **1** (Table S4, Figure S24) exhibited diagnostic signals for multiple exchangeable protons (δ_H 6.5–10.0), numerous α-protons of amino acids (δ_H 3.0–4.5), two chemically equivalent protons (δ_H 2.36) assigned to a *N*,*N*-dimethyl group, and six olefinic protons (δ_H 5.4–7.2) that are absent in proteinogenic amino acids, thus

illustrating that salinipeptin A (**1**) had been highly modified. This speculation was further supported by the observation of resonances for 21 amide carbonyl carbons (δ_C 166–178) and six olefinic carbons (δ_C 95–135) in the ^{13}C NMR spectrum ($CDCl_3$) (Figure S25). Despite the highly overlapping signals in the 1D proton NMR spectrum, the spin system of each amino acid residue was successfully resolved by detailed interpretation of 2D NMR data (HSQC, HMBC, COSY, TOCSY and HSQC-TOCSY) (Figures S26–S33), which permitted the assignment of 15 proteinogenic amino acids (Ala × 3, Val × 2, Pro × 2, Gln × 2, Ile/*allo*-Ile × 2, Leu × 1, Phe × 1, Ser × 1 and Gly × 1) and 6 non-proteinogenic amino acids including four Dhbs, one Me_2Ala and one AviCys residue in the structure of **1** (Figure 1). As the highly modified AviCys residue in the linaridins is formed between two cysteine residues via dehydration and oxidative decarboxylation-catalyzed cyclization at the *C*-terminus,[11] we deduced that AviCys-19 is cyclized with Val-21 in **1**, a conclusion confirmed by a key HMBC NMR correlation from -NH (δ_H 8.78) in the aminovinyl moiety to the C=O (δ_C 170.5) in Val-21 (Figure S5).

We then established the sequence of the 21 amino acid residues in **1**, fully aware of the unmodified core peptide sequence, first by inspection of sequential NOE correlations between the amide groups of two adjacent amino acid residues. Next we confirmed the sequence by inspection of NOE correlations from the α-H and β-H to the next amide group (Figure S6), and coupled these data with HMBC correlations from –NH to C=O within an amide group. Due to the presence of two proline residues whose tertiary amides obstruct the sequential HMBC correlations, as well as eight overlapping –NH signals (δ_H 7.2–7.7), only the partial sequences Me_2Ala-Dhb, Ala-Dhb and Gly-Ser-Dhb-Ile-AviCys of **1** could be unambiguously established by NMR methods.

To further corroborate the genomic and NMR-based sequence of **1**, tandem mass spectrometric analysis (MS^n) was applied to generate peptide fragment ions (Figure S13). The doubly protonated ion at *m*/*z* 1035.00 $[M+2H]^{2+}$ was selected as the precursor ion and subjected to MS/MS fragmentation (Figures S14 and S65). The product ions at *m*/*z* 1131.60 (b_{12} ion) and *m*/*z* 938.50 (y_9 ion) were chosen for successive MS/MS/MS analysis (Figures S66-68). Annotation and analysis of the fragment ions (mainly b and y ions) allowed for unambiguous validation of the amino acid sequence of **1** (Figure S13).

After the amino acid composition and the sequence of **1** had been confirmed, the absolute configurations of the amino acids were next evaluated using the advanced Marfey's method.[12] LC-MS analysis of L- and D-FDAA derivatives of the acid hydrolysate of **1,** and the comparison with FDAA derivatives of standard amino acids, indicated that all Ala, Pro, Gln and Ile/*allo*-Ile in **1** are D-amino acids, while the remaining proteinogenic amino acid residues are in L configurations (Table S8 and Figure S21). Although the FDAA derivatives of Ile and *allo*-Ile could not be differentiated after many attempts, including the use of different reverse phase HPLC columns and elution conditions, derivatization of the acid hydrolysate of **1** with 2,3,4,6-tetra-*O*-acetyl-*β*-D-glucopyranosyl isothiocyanate (GITC) allowed for the separation of Ile from *allo*-Ile (Table S9 and Figure S22).[13] Comparing GITC derivatives of L- and D-amino acid standards, we concluded that **1** contains D-Ile, consistent with the observation of the absence of the CypI homolog gene in the BGC of **1**, which was proposed to epimerize Ile to *allo*-Ile in cypemycin biosynthesis.[6] With regard to the stereochemistry of the AviCys moiety, a *Z* configuration of the double bond was determined based on the small coupling constant (^3J 7.0 Hz), while the absolute configuration of the α-H in AviCys was tentatively assigned as *R* by inspection of the energy-minimized structure of **1** that gives the best fit with the coupling constant of the α-H (^3J 9.7, 4.5 Hz) in the AviCys residue (Figure 3). Also, the double bonds in all four Dhb residues were assigned *Z* configurations on the basis of strong NOE correlations between the methyl group and the amide proton within each Dhb moiety (Figure S6). Thus, the initial sequence of **1** was proposed by genomic analysis and then the post-translational modifications assigned by comprehensive 1D and 2D NMR analyses.



	α-C (*R*) prediction	α-C (*S*) prediction	NMR
$^3J_{Hα,NH}$	8-10 Hz	8-10 Hz	9.7 Hz
$^3J_{Hα,Hβ1}$	0-2 Hz	0-2 Hz	not observed
$^3J_{Hα,Hβ2}$	2-6 Hz	12-14 Hz	4.5 Hz

Figure 3. Newman projections from the α-C to the β-C in **1** after ChemDraw 3D energy-minimization showing the *R*- versus *S*-AviCys residue with proposed and experimental coupling constants. Direct measurements strongly favor the α-C = *R* configuration.

In addition to **1**, three minor metabolites, salinipeptins B–D (**2**–**4**), were isolated in limited amounts (ca. 0.5 mg each). The structures of these analogs were fully characterized using 1.7 mm cryoprobe high-resolution NMR techniques coupled with high-resolution tandem mass spectrometry. HRESI (+) MS analysis indicated that the molecular weights of **2** and **4** are 16 Da and 18 Da higher than those of **1**, respectively, while the molecular weight of **3** is 2 Da lower. The changes in molecular weight suggest oxidation, hydration and dehydrogenation of these minor peptides compared to **1**. To quickly probe structural differences between **1** and **2**–**4**, differential analysis of MS/MS and HSQC NMR spectra as peptide fingerprints were performed.

MS/MS analysis of salinipeptin B (**2**) suggested that an additional oxygen is installed in one of the amino acid residues from Me_2Ala-1 to Dhb-5 in **1** to yield **2** (Figures S15, S16 and S69). Detailed comparison of HSQC NMR data for **2** with those for **1** (Figures S36–S39) illustrated that Me_2Ala in **1** is oxidized to *N*,*N*-dimethylalanine-*N*-oxide (oxiMe_2Ala) in **2**. This was also indicated by the significantly downfield-shifted α-H ($\Delta\delta_H$ +0.83, $\Delta\delta_C$ +10.0), as well as the two chemically nonequivalent methyl groups ($\Delta\delta_H$ +1.05, $\Delta\delta_C$ +17.3; $\Delta\delta_H$ +0.83, $\Delta\delta_C$ +12.6) in the oxiMe_2Ala residue (Table S5, Figures S7, S8, and S40-S44). The chemical shifts of the oxiMe_2Ala residue in **2** are in close agreement with those from a plant-derived cyclic peptide alkaloid *N*-oxide-mauritine A,[14] a published natural product containing the unusual oxiMe_2Ala residue. The presence of the amine *N*-oxide in **2** was further demonstrated by quantitative conversion of **1** to **2** using *meta*-chloroperoxybenzoic acid (*m*CPBA) (Figure S23).

Comparative MS^n analysis of salinipeptin C (**3**) with **1** (Figures 4A, S17, S18 and S70-S73) revealed that 2 Da of mass had to be subtracted from residues Me_2Ala-1 to Dhb-5 in **1** in order to satisfy the lower molecular weight. Differential analysis of HSQC NMR data for **1** and **3** (Figure 4B, Table S6, Figures S46-S49) showed the absence of one *N*-methyl group and the presence of a methylene (δ_H 4.14/4.58, δ_C 70.9) in **3**, along with the loss of an amide proton in the Dhb-2 residue. This suggested cyclization between one *N*-methyl group in Me_2Ala-1 and the amide nitrogen in Dhb-2 to yield an unusual 1,5-dimethylimidazolidin-4-one moiety (diMeIDL) at the *N*-terminus in **3**. This suggestion was confirmed by HMBC correlations from the methylene (δ_H 4.14/4.58) to *N*-Me (δ_C 38.3), α-C (δ_C 61.0) and amide carbonyl (δ_C 174.6) as well as the NOE correlation between the methyl group (δ_H 1.85) in Dhb-2 and the methylene (δ_H 4.14/4.58) in diMeIDL (Figures S9, S10, S50-S54). The uncommon diMeIDL moiety was previously reported only in the structure of apetaline B,[14] a plant-derived cyclic peptide from the same source as for *N*-oxide-mauritine A.

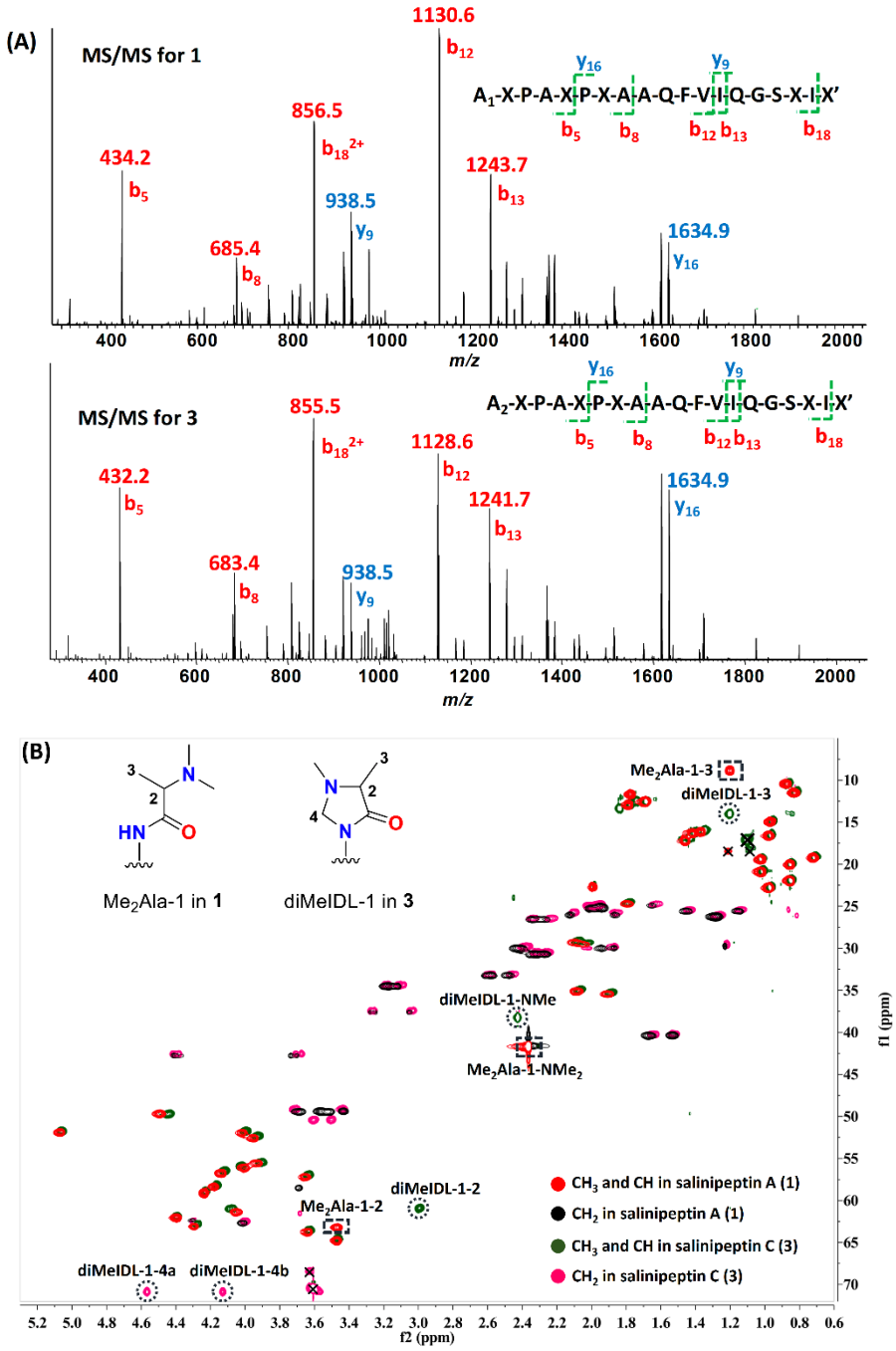


Figure 4. Differential spectroscopic analysis of **1** and **3**. **A**) Comparative analysis of MS/MS spectra for **1** (precursor ion *m/z* 1035.00 $[M+2H]^{2+}$) and **3** (precursor ion *m/z* 1034.00 $[M+2H]^{2+}$) narrowed down the location of PTMs in **3**. b ions and y ions were labelled in red and blue, respectively. In the peptide sequence, A_1 = Me_2Ala, A_2 = diMeIDL, X = Dhb, X' = cyclized tripeptides at *C*-terminus. **B**) Comparative analysis of ge-2D multiplicity-edited HSQC spectra for **1** and **3** targeting the cyclized dimethylalanine in **3**. Dash circles and dash rectangles indicate the non-overlapped cross peaks for diMeIDL moiety in **3** and Me_2Ala moiety in **1**, respectively.

MS^n analysis of salinipeptin D (**4**) suggested that Pro-6 or Dhb-7 in **1** undergoes hydration (Figures S19, S20, and S74-76). Having the knowledge that Thr is the precursor of Dhb in the biosynthesis of RiPPs, we hypothesized that Dhb-7 in **1** is replaced by Thr-7 in **4**. Consistent with this hypothesis, comparison of HSQC NMR data of **4** with those of **1** exhibited a spin

system for Thr ($\delta_{H\alpha}$ 3.86, $\delta_{H\beta}$ 4.04, $\delta_{H\gamma}$ 1.07) in **4** instead of those for Dhb ($\delta_{H\beta}$ 5.98, $\delta_{H\gamma}$ 1.68) (Table S7 and Figures S11, S12, S55-S64). Thus, the Thr residue in **4** does not undergo dehydration.

Identifying salinipeptin analogs in Great Salt Lake bacteria using molecular networking

To evaluate the chemical diversity of salinipeptins produced by *Streptomyces* sp. GSL-6C, we analyzed MS/MS data on the Global Natural Products Social Molecular Networking (GNPS) web platform. The GNPS is a recently developed method that automatically detects, correlates and visualizes the sets of spectra from related molecules based on their MS/MS fragmentation patterns.[15] The resulting MS/MS network of GSL-6C displays 25 clusters containing at least three nodes (Figure S2A). Among them, only two clusters (Clusters 1 and 2) are composed of the nodes with parent mass ranging from 2,000 to 2,100 Da. The nodes for **1**–**3** were readily identified from Cluster 1, while **4** was found separately in Cluster 2. Careful analysis of the remaining unknown nodes, together with MS ion extraction, identified three minor salinipeptins associated with **1**–**4**, suggesting that GSL-6C is rich in minor salinipeptins. We chose not to pursue these unidentified salinipeptins due to their trace amounts produced at less than 0.005 mg/L. Additionally, using the *sin*D and *sin*H sequences, the genomes of our other Great Salt Lake-derived actinomycetes were scanned for putative salinipeptin-like biosynthetic gene clusters. Two strains, *Streptomyces* spp. GSL-9 and GSL-20 were shown to contain *sin*-like clusters and HPLC-DAD-HRMS/MS analysis of their EtOAc extracts revealed that the strains produced similar peptide-derived natural products with observed molecular weights ranging from 2,000–2,100 Da (Figure S2B).

Relatedness of the salinipeptins to other linaridin RiPPs

Although RiPPs are a large family of natural products known to possess diverse structures with a wide range of biological activities, the linaridin subfamily is rare with only three characterized members being reported so far. The founding member, cypemycin, was identified from *Streptomyces* sp. OH-4156 in 1994, as a potent and selective antibiotic against *Micrococcus luteus* and a potent inhibitor of murine leukemia cells *in vitro*.[3,10] Remarkably different from those for lanthipeptide RiPPs, the BGC of cypemycin (*cyp*) identified in 2010, revealed distinct sets of enzymes and catalytic mechanisms that modify the precursor peptide.[6] First, the dehydratase enzymes CypH and CypL that are proposed to dehydrate Thr to Dhb in cypemycin are not homologous to those found in lanthipeptide biosynthesis. Secondly, the decarboxylase CypD, responsible for the oxidative decarboxylation of the *C*-terminal Cys in cypemycin, has little homology to those in lanthipeptide biosynthesis. Third, the AviCys moiety in cypemycin is derived from two Cys resides rather than one Ser and one Cys as observed in AviCys-containing lanthipeptides. Lastly, the formation of Me_2Ala catalyzed by the methyltransferase CypM in cypemycin has never been observed in lanthipeptide BGCs. With the precedence of the cypemycin BGC, recent genome mining of other *Streptomyces* isolates resulted in the isolation of grisemycin from *S. griseus* IFO 13350 and legonaridin from *S.* sp. CT34, the second and third members of linaridin RiPPs (Figure 1).[4,5] Of note, legonaridin is a linaridin variant containing 37 amino acid residues with a free carboxylic acid group at the *C*-terminus.[5] Herein, our discovery of salinipeptins A–D (**1**–**4**) significantly expands the subfamily of linaridin RiPPs. Furthermore, **2** and **3** contain unusual dimethylalanine amine oxide and dimethylimidazolidin-4-one moieties at their *N*-termini. To the best of our knowledge, these two moieties were only observed in a plant-derived cyclic peptide alkaloid in 2011 and have never been reported in any RiPPs.[14]

Of major significance, salinipeptin A was found to contain nine D-amino acids (Ala × 3, Pro × 2, Gln × 2 and Ile × 2) as demonstrated by Marfey's analysis and GITC derivatization. Advanced Marfey's analysis was used to confirm that legonaridin contains only L-amino acids, and amino acid analysis of hydrolyzed cypemycin showed that its structure contained two *allo*-Ile residues. However, the absolute stereochemistry of all amino acid residues in grisemycin and cypemycin remains unclear. To the best of our knowledge, the salinipeptins are the first reported members in the linaridin subfamily containing D-amino acids. Particularly, D-proline residues have not been reported previously in RiPPs. The isomerization of L-amino acids into the D-forms

occurs during post-translational modification of the precursor peptide and only a limited number of bacterial RiPPs containing D-amino acids such as lactocin S,[16] lacticin 3147,[17] siamycin/MS-271,[18,19] specialicin,[20] carnolysin,[21] bicereucin,[22] and polytheonamides[23] have been reported. The majority of these peptides belong to the lanthipeptide subfamily. The proposed mechanisms of post-translational isomerization for bacterial RiPPs can be categorized as follows: (1) Dehydration of L-Ser or L-Thr residues to generate dehydro-amino acids such as Dha or Dhb, followed by hydrogenation to give D-Ala or D-α-aminobutyric acid. Examples would include the enzymes LtnM/LtnJ involved in the biosynthesis of lacticin 3147,[24] CrnM/CrnJ in carnolysin,[21] and BsjM/BsjJ$_b$ in bicereucin;[22] (2) Formation of resonance-stabilized radicals catalyzed via abstraction of the α-H in amino acid residues, followed by hydrogen transfer or switch to yield D-amino acids. PoyD, a proteusin radical *S*-adenosyl methionine epimerase discovered from an uncultured bacterium, is the first and only enzyme reported thus far capable of introducing multiple D-amino acids in RiPPs.[23,25] However, mining the genome of GSL-6C with the sequences of these characterized isomerases did not return any homologs, inferring that new enzymes or catalytic mechanisms may be involved in the generation of D-amino acids in the salinipeptins. Further investigation into this mechanism is presently underway.

Biosynthetically, the salinipeptin precursor peptide assembled in the ribosome is subjected to post-translational modifications initiated by a variety of tailoring enzymes (Figures 1 and 2), with the most important modifications being formation of the AviCys end group, dehydration of Thr residues to Dhb, and epimerization of multiple L-amino acid residues to their D-isomers. Except for the partially dehydrated **4** retaining one Thr residue, the major peptide constituent, salinipeptin A (**1**), further undergoes *N*-terminal oxidation to yield **2**, followed by proposed oxidation, dehydration or elimination and cyclization to afford **3** (Figure 5). The identification and biochemical characterization of the enzymes involved in modifying the *N*-terminus in these peptides are in progress.

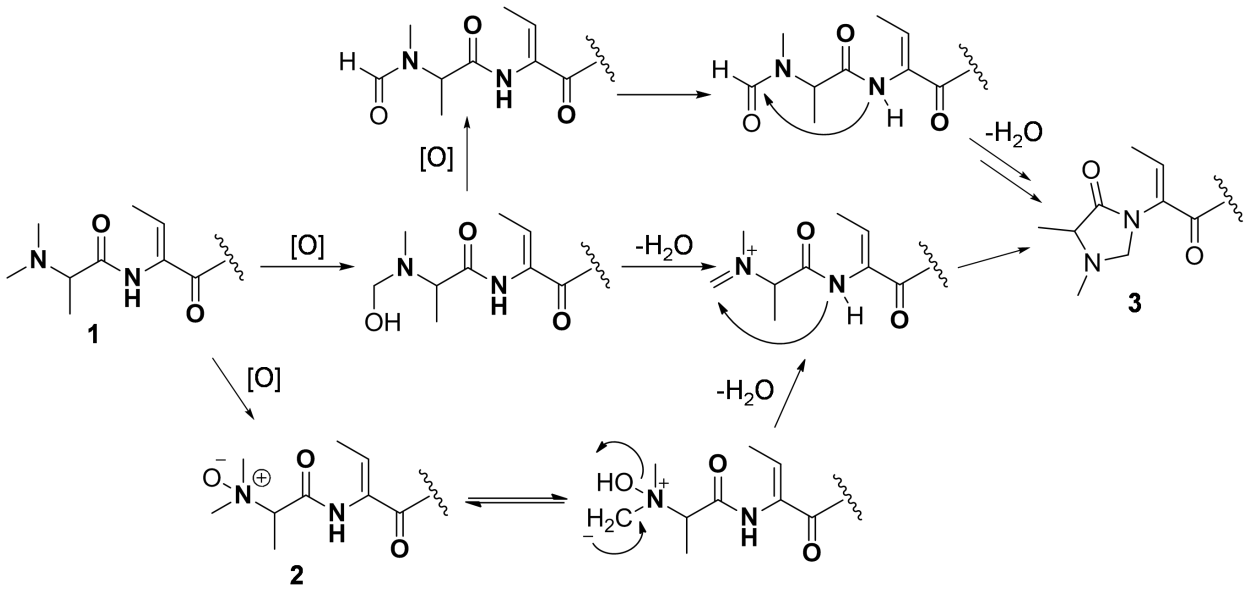


Figure 5. Possible mechanisms for the *N*-terminus modification of the mature peptide linking salinipeptins A–C (**1**–**3**).

Peptides and proteins exert their physiological and biological functions by their three-dimensional tertiary structures, which are specifically recognized by and bound to other micro- or macro-molecules. To achieve the lowest energy, peptides and proteins bend and twist driven largely by stabilizing forces, such as hydrophobic and ionic interactions of amino acid side chains, the

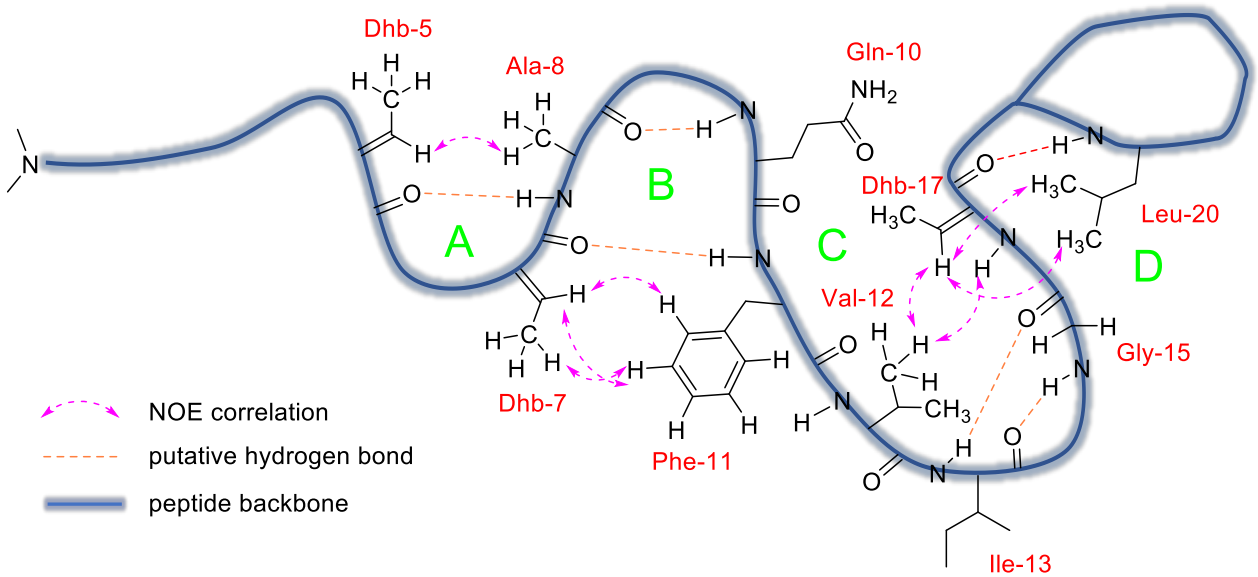


Figure 6. Proposed three-dimensional structure of salinipeptin A (**1**) showing key experimental NOE correlations and putative hydrogen bonds.

formation of disulfide bridges and hydrogen bonds formed between carbonyl and amine groups in different amino acid residues within proximity. Several computational programs have been developed to predict the three-dimensional structures of polypeptides and proteins. However, these programs are not applicable to the salinipeptins due to the presence of highly modified amino acids and D-amino acids in their structures. Therefore, we set out to apply experimental NOE measurements and molecular models to propose a three-dimensional structure for **1**. The NOESY NMR spectrum of **1** suggested that each amino acid pair Dhb-5 and Ala-8, Dhb-7 and Phe-11,Val-12 and Dhb-17, and Dhb-17 and Leu-20 positions closely in space respectively, as evidenced from key long-range NOE correlations between Dhb-5 (δ_H 5.95, H_β) and Ala-8 (δ_H 1.36, H_β), Dhb-7 (δ_H 5.98, H_β) and Phe-11 (δ_H 7.08 and 7.18, H_{Ar}), Val-12 (δ_H 0.71, H_γ) and Dhb-17 (δ_H 9.91, H_{NH}), Val-12 (δ_H 0.71, H_γ) and Dhb-17 (δ_H 6.71, H_β), and Dhb-17 (δ_H 6.71, H_β) and Leu-20 (δ_H 1.51/1.64, H_β; δ_H 0.83/0.96, H_δ) (Figure S34). Analysis of the predominant NOE correlations showed the conformation of **1** exhibited four major loops, which were stabilized through putative hydrogen bonds as shown in Figure 6. Specifically, loop A was proposed to be fixed by a hydrogen bond between Dhb-5 carbonyl (C=O) and Ala-8 amine (NH) groups, leading to the proximity of Dhb-5 and Ala-8. The formation of loop B is apparently driven by two hydrogen bonds (Dhb-7 C=O and Phe-11 NH; Ala-8 C=O and Gln-10 NH) and the hydrophobic effect between the olefinic side chain in Dhb-7 and benzene ring in Phe-11. With regard to loop C, we suggest that two hydrogen bonds (Ile-13 NH and Gly-15 C=O; Ile-13 C=O and Gly-15 NH) along with hydrophobic interactions between the side chains of Val-12 and Dhb-17 contribute to the formation of loop C. Finally, the hydrogen bond between Dhb-17 C=O and Leu-20 NH may be attributed to the generation of the small loop D next to the C-terminus tripeptide ring. Hence, the proposed three-dimensional structure of salinipeptin A (**1**) in $CDCl_3$ is suggested to be in a 4-loop configuration as illustrated in Figure 6.

It has been reported while that cypemycin exhibits both cytocidal activity against P388 leukemia cells and antibiotic activity against *Micrococcus luteus*,[10] its congener grisemycin lacked any activity against *M. luteus*.[4] As salinipeptin A is more similar to cypemycin than grisemycin in terms of number of amino acid residues and post-translational modifications, we set forth to assess its antibiotic and cytotoxic activities. Salinipeptin A showed antibacterial activity against Group A *Streptococcus pyogenes* M1T1 with an LD_{50} of 12.5 μg/mL, but was essentially inactive (>100 μg/mL) against vancomycin-resistant *Enterococcus faecalis*, *Acinetobacter baumannii* 5705, *Klebsiella pneumoniae* BB11, *Staphylococcus aureus*, *Escherichia coli* K1RS and *Micrococcus luteus*. Growth inhibition testing of salinipeptin A (**1**) against cancer cell lines showed modest activity against U87 glioblastoma and HCT-116 colon carcinoma (LD_{50} of 15 μg/mL), but was inactive toward MDA-MB-231 breast cancer. Interestingly, cypemycin was initially reported to have *in vitro* cytotoxicity against P388 leukemia cells at an IC_{50} value of 1.3 μg/mL and antimicrobial activity against *M. luteus* with an MIC of 0.2 μg/mL, but showed no activity against other Gram-positive or Gram-negative bacteria.[10] With the exception of a Val to Ala swap at position eight, the major structural differences between cypemycin and salinipeptin reside in the stereochemistry of Ile, Pro, Ala, and Gln, and the differences in *in vitro* activity can be attributed to these modifications. While the absolute configurations of amino acid residues in cypemycin have not yet been confirmed, it is highly likely that they are indeed L.

CONCLUSIONS

In summary, herein we described the full structures of salinipeptins A–D (**1**–**4**), four novel members of the rare linaridin RiPP subfamily, using integrated genomics, metabolomics and spectroscopic approaches. Salinipeptins are highly post-translationally modified RiPPs composed of multiple D-amino acids, particularly D-proline, *N*,*N*-dimethylalanine-*N*-oxide and dimethylimidazolidin-4-one subcomponents that have never been reported from other bacterial RiPPs. These observations suggest that unique enzymes and catalytic mechanisms are involved in the biosynthesis of the salinipeptins. This study has significantly expanded the chemical diversity of the linaridin subfamily and illustrated the existence of a potentially new epimerase creating the D-amino acids in these peptides. Biotesting of salinipeptin A against pathogenic bacteria and three cancer cell lines showed modest levels of activity. Further biochemical studies will shed light on the unprecedented modifications observed in the salinipeptins.

EXPERIMENTAL PROCEDURES

General experimental procedures

Specific optical rotations ($[\alpha]_D$) were measured on a JASCO P-2000 polarimeter (JASCO Analytical Instruments, Easton, MD, USA) in a 10 x 3.5 mm cell at 25 °C. UV-visible spectra were obtained on a Beckman Coulter DU 800 spectrophotometer (Beckman Coulter Life Sciences, Indianapolis, IN, USA) with 1 cm quartz cells. Nuclear magnetic resonance (NMR) spectra were acquired on a Bruker 600 MHz NMR spectrometer with a 1.7 mm inverse detection triple resonance (H-C/N/D) cryoprobe with z-gradient, a JEOL ECZ 500 NMR spectrometer equipped with a 3 mm inverse detection probe with z-gradient, and Varian VX 500 MHz NMR spectrometer equipped with Varian XSens 2 channel ($^1H/^{13}C$) NMR cold probe optimized for direct observance of ^{13}C NMR spectra. In all cases spectra were acquired at 25 °C (unless otherwise specified) in solvents as specified in the text, with referencing to residual ^1H or ^{13}C signals in the deuterated solvents. High-resolution mass spectrometry (HR-MS) and tandem mass spectrometry (MS^n) experiments were carried out on a Thermo Scientific LTQ Orbitrap XL Hybrid Ion Trap-Orbitrap Mass Spectrometer with the Thermo IonMax ESI interface. Liquid chromatography-mass spectrometry (LC-MS) data were collected on an Agilent 1260 series separation module coupled with a diode array multiple wavelength detector and an Agilent 6530 Accurate Mass Q-TOF mass spectrometer. The drying N_2 gas was at a flow rate of 11.0 L/min and a source temperature of 350 °C was set for the Q-TOF mass spectrometer. Semi-preparative HPLC was performed using

a Shimadzu HPLC system with a LC-10AT pump, a SCL-10A vp system controller, a SPD-M10A vp UV/vis diode array detector and a FRC-10A fraction collector.

Isolation, identification and genome mining of *Streptomyces* sp. GSL-6C

Streptomyces sp. GSL-6C was isolated from sediment collected from Farmington Bay in the Great Salt Lake, UT, USA. Sediment samples were collected at a depth of 0.6 m and dried for 48 hours in a biological safety cabinet. Dried sediment, 0.5 g, was added to agar plates prepared with yeast-peptone-mannitol (YPM) agar (4 g/L mannitol, 2 g/L yeast extract, 2 g/L peptone, 18 g/L agar and 42 g/L Instant Ocean Aquarium Sea Salt Mixture, Spectrum Brands, USA). The plates were incubated at 30 °C for up to 90 days and bacterial colonies were subcultured on YPM medium until pure isolates were obtained. 16S rRNA gene sequencing was used to identify GSL-6C as a *Streptomyces* sp. (GenBank accession number for the partial 16S rRNA gene sequence is MG517285). For gDNA isolation for sequencing, a 1 cm x 1 cm agar plug was used to inoculate 50 mL of GYEC media (15 g/L glucose, 3 g/L yeast extract, 5 g/L peptone, 25 g/L Instant Ocean Aquarium Sea Salt Mixture) in a 250 mL Erlenmeyer flask. The culture was shaken at 30 °C for five days and the cells were harvested by centrifugation (4,000 x g at 4 °C for 15 minutes). The cell pellet was ground to a fine powder under liquid nitrogen, suspended in lysis buffer (50 mM Tris, 50 mM EDTA, 2 % SDS, pH 8.0) and incubated in a 65 °C water bath for 30 minutes, followed by one hour's incubation on ice. The mixture was centrifuged at 4,000 x g for 15 minutes at 4 °C and the supernatant was transferred and extracted with phenol/chloroform/isoamyl alcohol at pH 8.0. gDNA was isolated from the aqueous layer using ethanol precipitation.

Library construction and sequencing was performed at the High-Throughput Genomics Center in the Huntsman Cancer Institute at the University of Utah. A 180 bp PCR-free DNA library was constructed and sequenced using Illumina HiSeq (125 cycle paired-end). Genome assembly was performed on the Future Systems server using the IDBA-UD (k-mer = iterative) software package.[26] AntiSMASH and the NCBI Blast+ program were used to mine the genome of *Streptomyces.* sp. GSL-6C and identify the salinipeptin biosynthetic gene cluster (GenBank accession MG788286).[27]

Small scale cultivation and chemical profiling of *Streptomyces* sp. GSL-6C

Streptomyces sp. GSL-6C, inoculated from 10 mL of seed culture, was cultivated in a 2.5 L Ultra Yield (Thomson Scientific, Oceanside, CA) shake flask containing 1 L of A1 broth (10 g soluble starch, 4 g yeast extract, 2 g peptone, 750 mL natural seawater and 250 mL deionized water) at 180 rpm at 27 °C for 7 days. After cultivation, the broth was extracted with 1 L of EtOAc and the organic layer was dried *in vacuo* to yield the organic extract (52.6 mg). After removal of non-polar fractions by hexane partition, the remaining extract was dissolved in MeOH (5 mg/mL) and subjected to HPLC-DAD-HRMS/MS analysis (Phenomenex Luna RP-C_{18} column, 100 × 4.6 mm, 5 μm, 0.7 mL/min gradient elution from 90% H_2O/MeCN to 100% MeCN over 20 min without or with 0.1% formic acid). The QTOF-MS/MS parameters were set as follows: mass range *m/z* 200–1700, positive ion mode, MS scan rate 1/s, MS/MS scan rate 3/s, fixed collision energy 20 keV, drying N_2 gas flow 11.0 L/min, source temperature 350 °C, nebulizer 45 psig. HPLC and MS data were analyzed on Agilent OpenLAB CDS ChemStation A.01.04. and Agilent MassHunter Qualitative Analysis B.05.00., respectively.

Molecular networking analysis of the *Streptomyces* sp. GSL-6C EtOAc extract

MS/MS data from the GSL-6C extract, in Agilent MassHunter file format (.d), were converted into mzXML files using ProteoWizard MSConvert.[28] After uploading the mzXML MS/MS files onto the GNPS server (gnps.ucsd.edu), a molecular network was created using the online workflow at GNPS.[29] The data were clustered with MS-Cluster with a parent mass tolerance of 0.05 Da and a MS/MS fragment ion tolerance of 0.05 Da to create consensus spectra. Further, consensus spectra that contained less

than two spectra were discarded. A network was then created where edges were filtered to have a Cosine score above 0.7 and more than 10 matched peaks. Further edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. To visualize the network of nodes and edges, the results were displayed and analyzed using Cytoscape 3.6.0.[30] For compound dereplication, the resulting spectra in the network were searched against GNPS spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks. For analogue search, the nodes in each cluster and the corresponding MS/MS spectra were compared and analyzed manually.

Large-scale cultivation and fractionation of *Streptomyces* sp. GSL-6C extract

Streptomyces sp. GSL-6C was inoculated into sixty 2.5 L Ultra Yield flasks (Thomson Scientific, Oceanside, CA) with each containing 1 L of A1 broth and shaken at 180 rpm at 27 °C for 7 days. The resulting broth was extracted with EtOAc (1 L × 60) and the combined organic phase was concentrated *in vacuo* to yield the organic extract (3.74 g), which was sequentially partitioned in hexane to remove non-polar substances. The hexane insoluble material (1.41 g) was re-dissolved in MeOH and subjected to Sephadex LH-20 fractionation eluting with MeOH to afford 15 fractions. HPLC-DAD-MS analysis revealed compounds **1**–**4** were eluted in the first fraction (51.2 mg), which was further fractionated by HPLC (Phenomenex Luna RP-C_{18}, 250 × 10 mm, 5 μm, 100 Å, 2.0 mL/min gradient elution from 75% to 35% H_2O/MeCN over 60 min, with a constant 0.1% formic acid modifier) to yield pure salinipeptin A (**1**) (t_R = 30.6 min; 15.4 mg, 0.4% of EtOAc extract), salinipeptin B (**2**) (t_R = 32.2 min; 1.4 mg, 0.04% of EtOAc extract) and salinipeptin C (**3**) (t_R = 41.9 min; 0.8 mg, 0.02% of EtOAc extract) as well as an impure form of salinipeptin D (**4**). The latter was subsequently purified by HPLC (Phenomenex Luna RP-C_{18}, 250 × 10 mm, 5 μm, 100 Å, 2.0 mL/min gradient elution from 75% to 55% H_2O/MeCN over 60 min, with a constant 0.1% formic acid modifier) to give the pure salinipeptin D (**4**) (t_R = 44.3 min; 1.2 mg, 0.03% of EtOAc extract).

Physicochemical properties of salinipeptins A–D (1–4)

Salinipeptin A (1). White amorphous powder; $[\alpha]_D^{25}$–55.8 (*c* 0.33, MeOH); UV-vis (MeOH)λ_{max} (log ε) 215–235 (sh) nm; NMR (500 MHz, $CDCl_3$) see Table S2; ESI(+)MS *m/z* 1045.55 $[M+H+Na]^{2+}$, 1034.56 $[M+2H]^{2+}$, 690.05 $[M+3H]^{3+}$; ESI(–)MS *m/z* 2066.09 $[M-H]^-$, 1032.65 $[M-2H]^{2-}$; HRMS (ESI-Orbitrap) *m/z*: $[M+2H]^{2+}$calcd for $C_{97}H_{152}N_{24}O_{24}S$ 1034.5560; Found 1034.5577.

Salinipeptin B (2). White amorphous powder; $[\alpha]_D^{25}$–34.5 (*c* 0.27, MeOH); UV-vis (MeOH) λ_{max}(log ε) 215–235 (sh) nm; NMR (600 MHz, $CDCl_3$) see Table S3; ESI(+)MS *m/z* 1053.57 $[M+H+Na]^{2+}$, 1042.54 $[M+2H]^{2+}$, 695.35 $[M+H]^{3+}$; ESI(–)MS *m/z* 2082.09 $[M-H]^-$, 1040.56 $[M-2H]^{2-}$; HRMS (ESI-Orbitrap) *m/z*: $[M+2H]^{2+}$calcd for $C_{97}H_{152}N_{24}O_{25}S$ 1042.5535; Found 1042.5509.

Salinipeptin C (3). White amorphous powder; $[\alpha]_D^{25}$–20.0 (*c* 0.25, MeOH); UV-vis (MeOH) λ_{max} (log ε) 215–235 (sh) nm; NMR (600 MHz, $CDCl_3$) see Table S4; ESI(+)MS *m/z* 1044.53 $[M+H+Na]^{2+}$, 1033.55 $[M+2H]^{2+}$, 689.37 $[M+3H]^{3+}$; ESI(–)MS *m/z* 2064.06 $[M-H]^-$, 1031.55 $[M-2H]^{2-}$; HRMS (ESI-Orbitrap) *m/z*: $[M+2H]^{2+}$calcd for $C_{97}H_{150}N_{24}O_{24}S$ 1033.5482; Found 1033.5484.

Salinipeptin D (4). White amorphous powder; $[\alpha]_D^{25}$–30.7 (*c* 0.30, MeOH); UV-vis (MeOH) λ_{max} (log ε) 215–235 (sh) nm; NMR (600 MHz, $CDCl_3$) see Table S5; ESI(+)MS *m/z* 1054.54 $[M+H+Na]^{2+}$, 1043.56 $[M+2H]^{2+}$, 696.05 $[M+3H]^{3+}$; ESI(–)MS

m/z 2084.10 $[M-H]^-$, 1041.54 $[M-2H]^{2-}$; HRMS (ESI-Orbitrap) *m/z*: $[M+2H]^{2+}$calcd for $C_{97}H_{154}N_{24}O_{25}S$ 1043.5613; Found 1043.5596.

Tandem mass spectrometric (MS^n) analysis of salinipeptins A–D (1–4)

Compounds **1–4** (0.2 mg/mL in MeOH) were diluted 3-fold to 50-fold (depending on signals) in 50% MeOH/H_2O containing 0.1% acetic acid. Each sample was infused into the Thermo Scientific LTQ Orbitrap Mass Spectrometer at the flow rate of 5 μL/min. The ESI source parameters were set as follows: capillary temperature 275 °C, sheath gas flow 8 units, positive polarity, source voltage 5.0 kV, capillary voltage 47 V, and tube lens 140 V. The Fourier transform MS (Orbitrap) parameters were: FTMS AGC 1e6, FTMS microscans averaged 2, FTMS full scan maximum ion time 500 ms, and mass range from *m/z* 100 to 2000. The resolution parameter was 7,500 (peak *m/z* divided by peak width given as full width at half maximum, at 400 *m/z*). For the MS-MS CID spectra, a normalized collision energy of 35% was used at first, and then varying collision energies were tested for optimal signals. The acquired MS^n data were analyzed using Thermo Xcalibur 2.0.7 software and mass peaks were annotated manually.

Amino acid analysis of salinipeptin A (1) by hydrolysis and chiral derivatization (Marfey's and GITC derivatives)

Salinipeptin A (**1**) (300 μg in 600 μL of 6 M HCl) was heated to 100 °C for 24 h with stirring in a sealed thick-walled reaction vessel, after which the hydrolysate was concentrated to dryness under N_2. The resulting hydrolysate was divided into three portions (3 × 100 μg) for chemical derivatization with 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide (L-FDAA), 1-fluoro-2, 4-dinitrophenyl-5-D-alanine amide (D-FDAA), and 2,3,4,6-tetra-O-acetyl-β-D-glucopyranosyl isothiocyanate (GITC).

(i) ***L- and D-FDAA derivatization of amino acids.*** Two aliquots (2 × 100 μg) of the hydrolysate of **1** were treated with 1 M $NaHCO_3$ (100 μL) and L- or D-FDAA (1% solution in acetone, 100 μL) at 40 °C for 1 h, respectively. Samples were neutralized with 1 M HCl (100 μL), diluted with MeCN (100 μL) and filtered (0.45 μm PTFE) prior to HPLC-DAD-MS analysis (Phenomenex Prodigy ODS-2 column, 250 x 4.6 mm, 5 μm, 0.7 mL/min gradient elution from 90% H_2O/MeCN to 100% MeCN with 0.1% formic acid over 45 min, 340 nm detector, positive and negative ion modes). The L- and D-FDAA derivatives of the peptide hydrolysate was detected by UV absorption (340 nm) and extracted ion chromatograms (EICs). The configurations of the amino acid residues in **1** were determined by comparison of their retention times and elution orders with those for FDAA derivatives of amino acid standards (see Table S8). For the preparation of FDAA-amino acid standard derivatives, 50 mM of L-amino acid (L-Ala, L-Val, L-Ser, L-Leu, L-Ile, L-*allo*-Ile, L-Phe, L-Pro, L-Gln, L-Glu, L-Thr, L-*allo*-Thr and L-Cys) dissolved in H_2O (50 μL) was treated with 1 M $NaHCO_3$ (20 μL) and 1% L- or D-FDAA (100 μL) at 40 °C for 1 h, respectively. After reaction, the solution was quenched with 1 M HCl (20 μL) and diluted with MeCN (810 μL) for HPLC-DAD-MS analysis using the same column and elution condition as above.

(ii) ***GITC derivatization of amino acids.*** An aliquot (100 μg) of the hydrolysate of **1** was incubated with 6% triethylamine (50 μL) and GITC (1% solution in acetone, 50 μL) at room temperature for 20 min, and then quenched with 5% acetic acid (50 μL). The reaction mixture was dried under N_2 and re-dissolved in 50 μL MeOH for HPLC-DAD-MS analysis (Phenomenex Prodigy ODS-2 column, 250 x 4.6 mm, 5 μm, 0.7 mL/min gradient elution from 90% H_2O/MeCN to 30% H_2O/MeCN with 0.1% formic acid over 45 min, 254 nm detector, positive and negative ion modes). The GITC derivatives of the salinipeptin A hydrolysate were detected by UV absorption (254 nm) and extracted ion chromatograms (EICs). The configurations of amino acid residues in **1** were determined by comparison of their retention times and elution orders with those for GITC derivatives of L- and D-amino acid standards (see Table S9). Regarding the GITC derivatives of amino acid standards, 50 mM of each L- and D-amino acid (L/D-Ala, L/D-Leu, L/D-Ile, L/D-*allo*-Ile, L/D-Pro and L/D-Glu) dissolved in H_2O (50 μL) was treated with

6% triethylamine (10 μL) and 1% GITC (20 μL) at room temperature for 20 min, followed by quenching with 5% acetic acid (10 μL) for HPLC-DAD-MS analysis using the same column and elution condition as above. To achieve a satisfactory resolution for Leu, Ile and *allo*-Ile, an optimized HPLC condition (0.7 mL/min gradient elution from 70% H_2O/MeCN to 60% H_2O/MeCN with 0.1% formic acid over 80 min) was applied to reanalyze the reaction solution and amino acid standards.

Oxidative conversion of salinipeptin A (1) to salinipeptin B (2)

An aliquot (2 mg) of **1** in CH_2Cl_2 (1 mL) was cooled to –78 °C and treated with *meta*-chloroperoxybenzoic acid (*m*CPBA, 77% purity) (20 μL, 13.6 mg/mL in CH_2Cl_2, 1.25 mol equiv.). The reaction was kept at –20 °C for 20 min followed by quenching with K_2CO_3 solution (20 μL, 10.8 mg/mL in H_2O, 1 mol equiv.). The product mixture was analyzed by HPLC-DAD-MS (Phenomenex Luna RP-C_{18} column, 100 × 4.6 mm, 5 μm, 0.7 mL/min gradient elution from 90% H_2O/MeCN to 100% MeCN over 20 min with 0.1% formic acid modifier) and further fractionated on semi-preparative HPLC (Phenomenex Luna RP- C_{18} column, 250 × 10 mm, 5 μm, 100 Å, 2.0 mL/min gradient elution from 75% to 35% H_2O/MeCN over 60 min, with a constant 0.1% formic acid modifier) to afford **2** (0.9 mg, 45% yield).

Bioactivity testing

Using the standard broth dilution antibiotic testing method,[31] salinipeptin A (**1**) was screened for antimicrobial activity against Group A *Streptococcus pyogenes* M1T1, vancomycin-resistant *Enterococcus faecalis*, *Acinetobacter baumannii* 5705, *Klebsiella pneumoniae* BB11, *Staphylococcus aureus*, *Escherichia coli* K1RS and *Micrococcus luteus*. No antimicrobial activity against *S. aureus* or *M. luteus* was observed at 128 μg/mL. Cancer cell cytotoxicity against U87 Glioblastoma, HCT-116 Carcinoma, and MDA-MB-231 breast cancer cell lines was measured using the standard Alamar Blue method.[32] Briefly, cancer cells are cultured in 96 well plates using Dulbecco's Modified Eagle's Culture Medium (DMEM) and maintained at 37 degrees in a humidified atmosphere. Cancer cells are initially added to the wells at the following concentrations: HCT-116: 3,000 cells/well, MDA-MB-231: 3,000 cells/well, U87: 1,000 cells/well. In order to establish IC_{50} curves, cells are plated on Day 1, treated on Day 2, and analyzed for cell viability on Day 5 using Alamar Blue fluoresence as a colormetric assay. Treatment consisted of cells in quadruplicate with a 6 step dilution of the compound from 10 μM to 0.0001 μM. At Day 5, the metabolic activity of viable cells to reduce the blue dye to a fluorescent pink form is measured colorimetrically at 590 nm providing a relative measure of cell death.

ASSOCIATED CONTENT

Supporting Information. Gene annotation and comparison, full spectral data for salinipeptins A-D, LC-MS chromatograms, NMR and tandem mass spectra of **1**–**4**, and molecular network analysis of GSL-6C. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*wfenical@ucsd.edu

*jaclyn.winter@utah.edu

ORCID

Zhuo Shang: 0000-0002-5755-2629

Jaclyn M. Winter: 0000-0001-6273-5377

Inho Yang: 0000-0003-0990-1465

William Fenical: 0000-0002-8955-1735

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the University of Utah, Department of Medicinal Chemistry Start-Up Funds and the ALSAM Foundation (No. 10046217) to J.M.W., and by partial support from the NIH under grant R37 CA044848 to W.F. The authors are grateful for the assistance from L. Gross and Y. Su (UCSD) for tandem mass spectrometry advice. The authors also thank B. M. Duggan and A. A. Mrse (UCSD) for their help with NMR methods, S. A. Bell (U. Utah) for assistance with genome assembly of *S.* sp. GSL-6C. J. Li and Q. Zhang (UCSD) for valuable discussions, and reviewer 1 for suggesting an alternative pathway for diMeIDL formation. We thank S. Dahesh and V.Nizet (UCSD) and M. Langer (U. Utah), for providing antibacterial testing and J. Rodvold (UCSD) for cancer cell bioassay data.

ABBREVIATIONS

RiPPs, ribosomally synthesized and post-translationally modified peptides; PTMs, post-translational modifications; Dhb, dehydrobutyrine; AviCys, aminovinyl-cysteine; BGCs, biosynthetic gene clusters; GSL, Great Salt Lake; Me_2Ala, *N*,*N*-dimethylalanine; oxiMe_2Ala, *N*,*N*-dimethylalanine-*N*-oxide; diMeIDL, 1,5-dimethylimidazolidin-4-one; FDAA, 1-fluoro-2-4-dinitrophenyl-5-alanine amide; GITC, 2,3,4,6-tetra-*O*-acetyl-β-D-glucopyranosyl isothiocyanate; *m*CPBA, *meta*-chloroperoxybenzoic acid; GNPS, Global Natural Products Social Molecular Networking; EIC, extracted ion chromatogram.

REFERENCES

(1) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K-D.; Fischbach, M. A.; Garavelli, J. S.; Göransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Müller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J. T.; Rebuffat, S.; Ross, R. P.; Sahl, H.-G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Süssmuth, R. D.; Tagg, J. R.; Tang, G-L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. A. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep. 30*, 108−160.

(2) Ortega, M. A.; van der Donk, W. A. (2016) New insights into the biosynthetic logic of ribosomally synthesized and post-translationally modified peptide natural products. *Cell Chem. Biol. 23*, 31−44.

(3) Minami, Y.; Yoshida, K. I.; Azuma, R.; Urakawa, A.; Kawauchi, T.; Otani, T.; Komiyama, K.; Omura, S. (1994) Structure of cypemycin, a new peptide antibiotic. *Tetrahedron Lett. 35*, 8001−8004.

(4) Claesen, J.; Bibb, M. J. (2011) Biosynthesis and regulation of grisemycin, a new member of the linaridin family of ribosomally synthesized peptides produced by *Streptomyces griseus* IFO 13350. *J. Bacteriol. 193*, 2510–2516.

(5) Rateb, M. E.; Zhai, Y.; Ehrner, E.; Rath, C. M.; Wang, X.; Tabudravu, J.; Ebel, R.; Bibb, M.; Kyeremeh, K.; Dorrestein, P. C.; Hong, K.; Jaspars, M.; Deng, H. (2015) Legonaridin, a new member of linaridin RiPP from a Ghanaian *Streptomyces* isolate. *Org. Biomol. Chem. 13*, 9585–9592.

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

(7) Mo, T.; Liu, W. Q.; Ji, W.; Zhao, J.; Chen, T.; Ding, W.; Yu, S.; Zhang, Q. (2017) Biosynthetic insights into linaridin natural products from genome mining and precursor peptide mutagenesis. *ACS Chem, Biol. 12*, 1484–1488.

(8) Ding, W.; Yuan, N.; Mandalapu, D.; Mo, T.; Dong, S.; Zhang, Q. (2018) Cypemycin decarboxylase CypD is not responsible for aminovinyl-cysteine (AviCys) ring formation. *Org. Lett. 20*, 7670–7673.

(9) Ding, W.; Mo, T.; Mandalapu, D.; Zhang, Q. (2018) Substrate specificity of the cypemycin decarboxylase CypD. *Synth. Syst. Biotechnol. 3*, 159–162.

(10) Komiyama, K.; Otoguro, K.; Segawa, T.; Shiomi, K.; Yang, H.; Takahashi, Y.; Hayashi, M.; Otani, T.; Omura, S. (1993) A new antibiotic, cypemycin. Taxonomy, fermentation, isolation and biological characteristics. *J. Antibiot. 46*, 1666–1671.

(11) Sit, C. S.; Yoganathan, S.; Vederas, J. C. (2011) Biosynthesis of aminovinyl-cysteine-containing peptides and its application in the production of potential drug candidates. *Acc. Chem. Res. 44*, 261–268.

(12) Fujii, K.; Ikai, Y.; Mayumi, T.; Oka, H.; Suzuki, M.; Harada, K. (1997) A nonempirical method using LC/MS for determination of the absolute configuration of constituent amino acids in a peptide: elucidation of limitations of Marfey's method and of its separation mechanism. *Anal. Chem. 69*, 3346–3352.

(13) Hess, S.; Gustafson, K. R.; Milanowski, D. J.; Alvira, E.; Lipton, M. A.; Pannell, L. K. (2004) Chirality determination of unusual amino acids using precolumn derivatization and liquid chromatography-electrospray ionization mass spectrometry. *J. Chromatogr. A, 1035*, 211–219.

(14) Han, J.; Ji, C. J.; He, W. J.; Shen, Y.; Leng, Y.; Xu, W. Y.; Fan, J. T.; Zeng, G. Z.; Kong, L. D.; Tan, N. H. (2011) Cyclopeptide alkaloids from *Ziziphus apetala*. *J. Nat. Prod. 74*, 2571–2575.

(15) Wang, M.; et al. (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol. 34*, 828–837.

(16) Skaugen, M.; Nissen-Meyer, J.; Jung, G.; Stevanovic, S.; Sletten, K.; Inger, C.; Abildgaard, M.; Nes, I. F. (1994) *In vivo* conversion of L-serine to D-alanine in a ribosomally synthesized polypeptide. *J. Biol. Chem. 269*, 27183–27185.

(17) Ryan, M. P.; Jack, R. W.; Josten, M.; Sahl, H. G.; Jung, G.; Ross, R. P.; Hill, C. (1999) Extensive post-translational modification, including serine to D-alanine conversion, in the two-component lantibiotic, lacticin 3147. *J. Biol. Chem. 274*, 37544–37550.

(18) Feng, Z.; Ogasawara, Y.; Nomura, S.; Dairi, T. (2018) Biosynthetic gene cluster of a D-tryptophan-containing lasso peptide, MS-271. *Chembiochem, 19*, 2045–2048.

(19) Yano, K.; Toki, S.; Nakanishi, S.; Ochiai, K.; Ando, K.; Yoshida, M.; Matsuda, Y.; Yamasaki, M. (1996) MS-271, a novel inhibitor of calmodulin-activated myosin light chain kinase from *Streptomyces* sp.--I. Isolation, structural determination and biological properties of MS-271. *Bioorg. Med. Chem. 4*, 115–120.

(20) Kaweewan, I.; Hemmi, H.; Komaki, H.; Harada, S.; Kodani, S. (2018) Isolation and structure determination of a new lasso peptide specialicin based on genome mining. *Bioorg. Med. Chem. 26*, 6050–6055.

(21) Lohans C. T.; Li, J. L.; Vederas, J. C. (2014) Structure and biosynthesis of carnolysin, a homologue of enterococcal cytolysin with D-amino acids. *J. Am. Chem. Soc. 136*, 13150–13153.

(22) Huo, L.; van der Donk, W. A. (2016) Discovery and characterization of bicereucin, an unusual D-amino acid-containing mixed two-component lantibiotic. *J. Am. Chem. Soc. 138*, 5254–5257.

(23) Freeman, M. F.; Gurgui, C.; Helf, M. J.; Morinaka, B. I.; Uria, A. R.; Oldham, N. J.; Sahl, H. G.; Matsunaga, S.; Piel, J. (2012) Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. *Science*, *338*, 387–390.

(24) Cotter, P. D.; O'Connor, P. M.; Draper, L. A.; Lawton, E. M.; Deegan, L. H.; Hill, C.; Ross, R. P. (2005) Posttranslational conversion of L-serines to D-alanines is vital for optimal production and activity of the lantibiotic lacticin 3147. *Proc. Natl. Acad. Sci. U S A*, *102*, 18584–18589.

(25) Morinaka, B. I.; Vagstad, A. L.; Helf, M. J.; Gugger, M.; Kegler, C.; Freeman, M. F.; Bode, H. B.; Piel, J. (2014) Radical *S*-adenosyl methionine epimerases: regioselective introduction of diverse D-amino acid patterns into peptide natural products. *Angew. Chem. Int. Ed. Engl. 53*, 8503–8507.

(26) Peng, Y.; Leung, H. C.; Yiu, S. M.; Chin, F. Y. (2012) IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics 28*, 1420–1428.

(27) Weber, T.; Blin, K.; Duddela, S.; Krug, D.; Kim, H. U.; Bruccoleri, R.; Lee, S. Y.; Fischbach, M. A.; Müller, R.; Wohlleben, W.; Breitling, R.; Takano, E.; Medema, M. H. (2015) antiSMASH 3.0-a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res. 43*, W237–243.

(28) Adusumilli, R.; Mallick, P. (2017) Data conversion with ProteoWizard msConvert. *Methods Mol. Biol. 1550*, 339–368.

(29) UCSD Computational Mass Spectrometer Website; http://gnps.ucsd.edu/ (accessed 2/22/2018)

(30) Cytoscape Software; http://www.cytoscape.org/ (accessed 2/23/2018)

(31) Stalons, D.R.; Thornsberry, C. Broth-dilution method for determining the antibiotic susceptibility of anaerobic bacteria. *Antimicrob. Agents & Chemother.* (1975) *7*, 15-21.

(32) O'Brien, J.; Wilson, I.; Orton, T.; Pognan, F. (2000) Investigation of the Alamar Blue (resazurin) fluorescent dye for the assessment of mammalian cell cytotoxicity. *Eur. J. Biochem. 267*, 5421-5426.