

# Organic & Biomolecular Chemistry

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

## Legonaridin, a new member of linaridin RiPP from a Ghanaian *Streptomyces* Isolate

Mostafa E. Rateb,<sup>a,g</sup> Yin Zhai,<sup>b</sup> Emmanuelle Ehrner,<sup>a</sup> Christopher M Rath,<sup>c</sup> Xiaoling Wang,<sup>a</sup> Jioji Tabudravu,<sup>a</sup> Rainer Ebel,<sup>a</sup> Mervin Bibb,<sup>e</sup> Kwaku Kyeremeh,<sup>f</sup> Pieter C. Dorrestein,<sup>c,d</sup> Kui Hong,<sup>\*b</sup> Marcel Jaspars,<sup>\*1</sup> Hai Deng<sup>\*a</sup>

<sup>a</sup>Marine Biodiscovery Centre, Department of Chemistry, University of Aberdeen, Aberdeen AB24 3UE Scotland, United Kingdom. Email: [h.deng@abdn.ac.uk](mailto:h.deng@abdn.ac.uk); [m.jaspars@abdn.ac.uk](mailto:m.jaspars@abdn.ac.uk).

<sup>b</sup>Key Laboratory of Combinatory Biosynthesis and Drug Discovery, School of Pharmaceutical Sciences, Wuhan University, 185 East Lake Road, Wuhan 430071, China. Email: [kuihong31@whu.edu.cn](mailto:kuihong31@whu.edu.cn).

<sup>c</sup>Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California, USA.

<sup>d</sup>Department of Pharmacology and Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California, USA.

<sup>e</sup>Department of Molecular Microbiology, John Innes Centre, Norwich NR4 7UH, United Kingdom.

<sup>f</sup>Department of Chemistry, FGO Torto Building, University of Ghana, Legon, Ghana.

<sup>g</sup>Pharmacognosy Department, School of Pharmacy, Beni-Suef University, Beni-Suef 62514, Egypt.

### Abstract

Linaridins are rare linear ribosomally-synthesized and post-translationally modified peptides (RiPPs) and only two, cypemycin and SGR-1832, in this family have been identified so far. Legonaridin **1** has been discovered as a new member of linaridins through chemical isolation, peptidogenomics, comprehensive 1- and 2-D NMR and advanced Marfey's analyses from the soil bacterium *Streptomyces* sp. CT34, an isolate collected from Legon, Ghana. Bioinformatics analysis of the gene cluster suggested that the biosynthesis of legonaridin **1** is different from the ones for cypemycin and SGR-1832. Consistent with bioinformatics and peptidogenomics analyses, **1** has a total of nine post-modifications, 8 dehydrobutyrine residues and *N,N*-dimethylated *N*-terminus with a carboxylic acid at the *C*-terminus. Legonaridin **1** is structurally different from the two known linaridins comprising a new subfamily. This is the first time to use NMR spectroscopy to establish the 2-D structure of a linaridin RiPP.

## Introduction

There is an increasing interest in the discovery of new ribosomally-synthesized and post-translationally modified peptides (RiPPs).<sup>1</sup> RiPPs are a major class of natural products that has only been recently discovered and found in all three domains of life.<sup>2</sup> Although these molecules explore only the 20 proteinogenic amino acids, extensive posttranslational modifications provide these peptide with enormous chemical and biological diversity.<sup>3</sup>

The structural complexity of RiPPs taken together with their often high molecular weight has made their chemical discovery extremely time-consuming.<sup>4</sup> This is exemplified for the discovery of the polytheonamides which were reported in 1994 as cytotoxic constituents of the Japanese sponge *Theonella swinhoei*, which harbours a large diversity of symbiotic bacteria.<sup>5</sup> It took more than 10 years for the same research group to establish the precise structure of these compounds as highly complex 48-mer peptides.<sup>6</sup>

Linaridins are rare RiPPs.<sup>7</sup> Despite increased numbers of RiPPs discovered, only two linaridins have been found so far including cypemycin, the founding member of linaridin family<sup>7</sup> and grisemycin (SGR-1832)<sup>8</sup> (Figure 1). Both linaridins possess common structural features, notably AviCys moiety in C-terminus, *N,N*-dimethylalanine at *N* terminus and dehydrothreonine moieties.<sup>7,8</sup>



**Figure 1.** Cypemycin is the founding member of the linadrinin family from *Streptomyces* sp. OH-4156. Grisemycin (SGR-1832) is the second linadrinin identified from *Streptomyces griseus* IFO 13350. Both compounds contain an *N*-terminal *N,N*-dimethylalanine, a *C*-terminal aminovinyl-cysteine (AviCys) ring and dehydrated threonines (X or dehydrobutyrines). Cypemycin also possesses two *L*-*allo*-isoleucine residues. The stereochemistry of isoleucine residues in SGR-1832 has not been established. The red colour represents the posttranslationally modified amino acid residues.

During the course of our investigation on biodiversity in Ghana, we were able to isolate several new soil actinomycetes including *Streptomyces* sp. MA37<sup>9</sup> and two new mangrove strains, *Micromonospora* sp. K310 that produces butremycin, a new derivative of ikarugamycin,<sup>10</sup> and *Verrucosipora* sp. K51G, the producer of a new pyrazinone butrepyrazinone.<sup>11</sup>

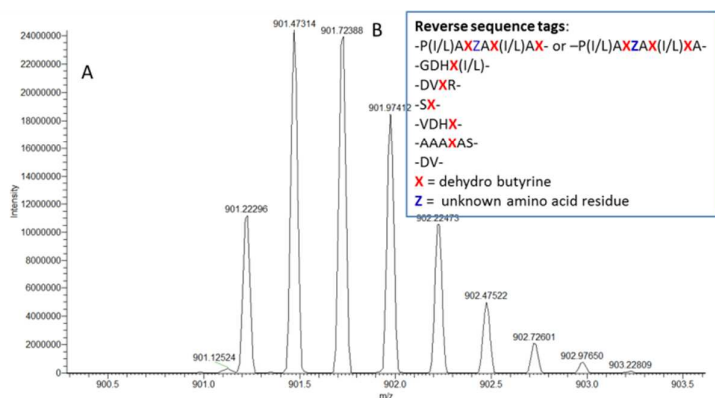
Here we report the discovery and characterization of a new linaridin, legonaridin **1**, using a combination of chemical isolation, peptidogenomics, and comprehensive NMR spectroscopic analysis, from *Streptomyces* sp. CT34 (Figure 2). The identity of the gene cluster directing the biosynthesis of **1** was confirmed through gene inactivation. The absolute structure of **1** was established through advanced Marfey's analysis, suggesting that **1** is structurally different to cypemycin and grisemycin, and belongs to a new subgroup of linaridin RiPP family, consistent with our bioinformatics analysis.

## Results and discussion

### Identification of peptide from *Streptomyces* sp. CT34

*Streptomyces* sp. CT34 is a new *Streptomyces* strain that was isolated from a rhizosphere sample from Legon, Ghana (Figure S1). Metabolite profiling of the strain revealed the presence of a high molecular-weight metabolite (3600.8 Da) only under the acidic culturing condition (pH 5.5). Large scale fermentation, extraction and solvent partitioning using modified Kupchan method,<sup>12</sup> followed by RP-HPLC of the CH<sub>2</sub>Cl<sub>2</sub> fraction allowed the isolation of pure compound **1** (9 mg).

Analysis of high resolution electrospray ionization mass spectrometry (HR-ESIMS) gave monoisotopic ion  $[M + 4H]^{4+}$  at  $m/z$  of 901.2230 (Figure 2A). Inspection of <sup>1</sup>H NMR indicated the presence of a group of  $\alpha$ -H signals around 4.5 ppm, suggesting that **1** is a peptidic natural product (Figure S2). The rest of proton signals were, however, heavily overlapped, which would have prevented full structural characterization. For this reason a different approach was needed to fully establish the structure of legonaridin **1**.



**Figure 2.** (A). Mass spectrometric identification of the 3600 Da legonaridin **1** from the organic extract of *Streptomyces* sp. CT34 (ion  $[M+4H]^{4+}$ ); (B) Generation of 7 reverse sequence tags from the purified legonaridin **1** by FTICR-mass spectrometric detection. X = dehydrobutyryne and Z = unidentified amino acid residue.

We then applied the peptidogenomics methodology<sup>8</sup> developed by one of our authors (PCD) and generated 7 identifiable peptide fragments (Figure 2B). This was obtained by using tandem MS ( $MS^n$ ) fragments to generate sequence tags for this peptide which was then subjected to collision-induced dissociation (CID) fragmentation

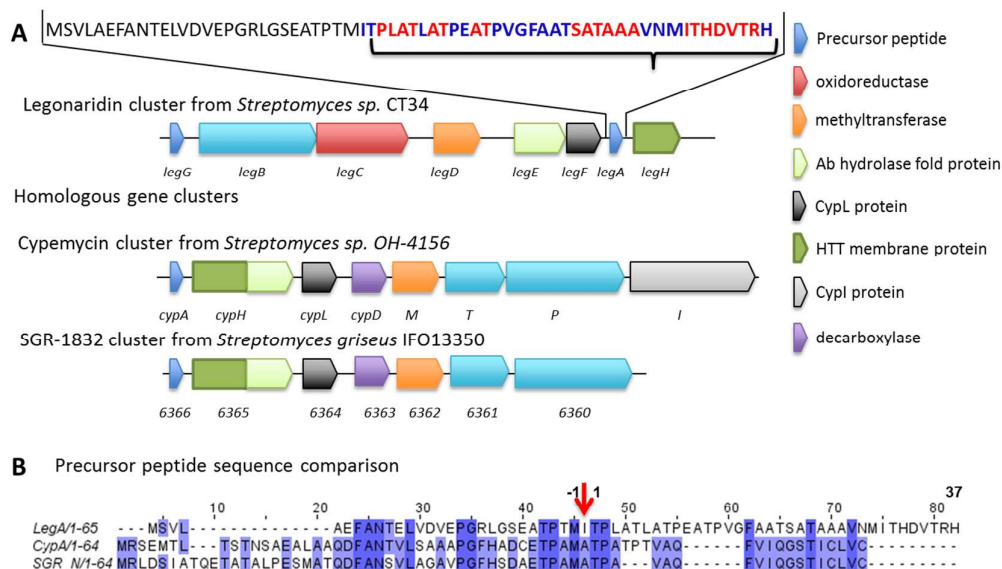
and the sequential fragment ions were assigned through *de novo* analysis of MS<sup>n</sup> spectra (Figure S3-4). Almost all of the mass shifts generated can be substituted with proteinogenic amino acids except mass shifts of 83 Da which were assigned to the non-proteinogenic amino acid dehydrobutyrine (X or Dhb). Interestingly, serine residue(s) in the sequence tags remained intact. A similar feature was also observed in the RiPP natural products such as lantibiotics and other linaridins in which serine residues are less favoured than threonine residues for dehydration.<sup>13</sup> This suggested that this compound may be a ribosomally-derived peptide.

### Identification of the peptide sequence

Illumina-based genome scanning of *Streptomyces* sp. CT34 was carried out, generating 8.05 Mbp of sequence with 310 contigs, and the draft genome was deposited in NCBI database.<sup>14</sup> The sequence was annotated using the AntiSMASH software.<sup>15</sup> Through peptidogenomics-based matching, we found one putative peptide, LegA, that was annotated as a homolog of the precursor peptide CypA of cypemycin. In comparison with the mature peptides of cypemycin and grisemycin, we proposed that the mature peptide of **1** started with isoleucine (Ile) and have 37 amino acids in total (Figure 3A and B), indicating that **1** is a new member of the emerging linaridin RiPP family, which we named legonaridin after its association with Legon, Ghana, the location of the University of Ghana.

Analysis of the surrounding genetic environment of *legA* allowed identification of a candidate gene cluster (*leg*) (Figure 3A and Table S1). Gene inactivation of the entire cluster completely abolished the production of **1**, indicating that the *leg* cluster indeed directs the biosynthesis of **1**. The *leg* cluster possesses 8 ORFs, 2 of which, LegD and F, share moderate sequence identities with CypM (30%) and CypL (37%), respectively, in the cypemycin pathway (Table S1). Biochemical assay demonstrated that CypM is responsible for the *N,N* dimethylation of the *N*-terminal Ala in the biosynthesis of cypemycin,<sup>7</sup> an unusual peptide modification that was also observed in SGR-1832 linaridin peptide.<sup>8</sup> Therefore, it is proposed that LegD may generate *N*-methylation on the *N*-terminus Ile of **1**. The two genes, *legE* and *legH*, encode Alpha/beta(Ab) hydrolase fold protein and HTT membrane protein, in the close proximity of *legA*, respectively (Figure 3A). Interestingly, LegE has moderate sequence identity (31% identity) with the hydrolase domain of CypH whereas LegH has moderate sequence identity (40% identity) with the HTT domain of CypH (Table S1). It appears that the *cypH* homolog in the *leg* cluster is split into two individual genes. The presence of both encoded CypL and CypH are required for the production of cypemycin. CypL and/or CypH are believed to catalyze the dehydration of threonine residues into the mature peptide.<sup>7</sup> The presence of genes *legE*, F, and H in the *leg* cluster suggests that these gene products may be required for the production of **1**, possibly playing the same roles as CypL and H in the maturation of cypemycin. However, one cannot exclude the possibility that LegC, a putative FAD-dependent oxidoreductase, may catalyse the dehydration of all of Thr residues. The exact *in vivo* roles of these four genes remain to be determined. In the cypemycin biosynthesis, CypD is responsible for the decarboxylation of the cysteine residue in the *C*-terminus, the essential biosynthetic step to construct the AviCys motif.<sup>7</sup> The *cypD* homolog gene was also observed in the SGR-1832 cluster.<sup>8</sup> However, there is no *cypD* homolog gene in the *leg* cluster in the CT34 strain, suggesting that legonaridin is a completely linear peptide with a free carboxylic acid group at the *C*-terminus. Interestingly, the presence of *cypD* homolog genes is not

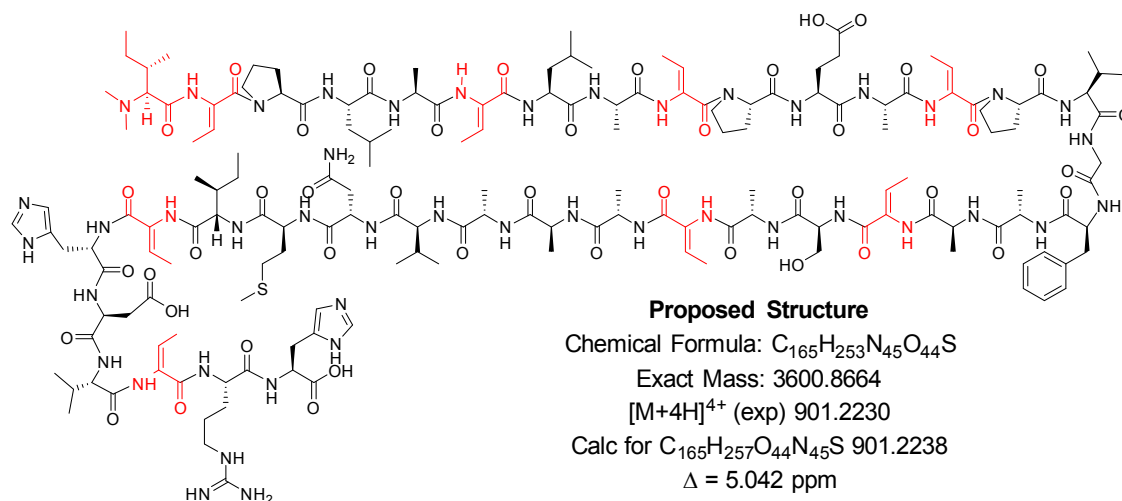
common in the putative linaridin clusters and was only observed in the cypemycin and SGR-1832 clusters.<sup>7,8</sup> Moreover, no CypI homolog gene was observed in the close proximity of leg cluster. CypI was proposed to epimerize L-isoleucine residue to L-*allo*-isoleucine.<sup>7</sup> Therefore, it is likely that the L-configuration of the two isoleucine residues in the legonaridin **1** has been retained.



**Figure 3.** (A) Schematic representation of the biosynthetic gene clusters of linaridin RiPPs. The sequence tags generated by MSn from legonaridin **1** are highlighted in Red. (B) The sequence alignment of legonaridin precursor peptide sequence and cypemycin and grisemycin (SGR-1832) precursor peptide sequences. The red arrow indicates the possible first amino acid in the matured legonaridin.

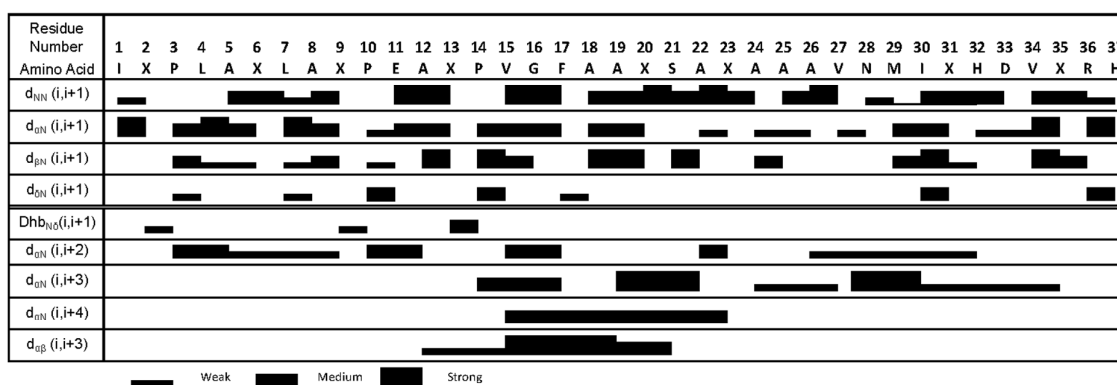
### NMR analysis confirms that **1** is a new linaridin RiPP

*In silico* analysis allow us to propose the structure of **1** as shown in Figure 4, which was corroborated by MS<sup>n</sup> fragments, providing strong supportive evidence that the predicted structure is correct. The molecular formula of **1** was established as C<sub>165</sub>H<sub>254</sub>N<sub>45</sub>O<sub>44</sub>S based on the HR-ESIMS analysis (observed [M + 4H]<sup>4+</sup> = 901.2230, calculated [M + 4H]<sup>4+</sup> = 901.2238, Δ = 5.042 ppm. We then revisited the NMR experiments including <sup>1</sup>H, COSY, TOCSY, HSQC, HMBC and NOESY (Table 1 and Figures S2, 5-9). Despite of overlapping peaks, all protons were successfully assigned through a strategy based on definition of the amino acid spin systems<sup>16-18</sup> by inspection of COSY and TOCSY NMR data, followed by confirmation of the overall amino acid sequence using NOESY data (Figure 5). The most predominant NOE correlations were those of N<sub>(i)</sub>-N<sub>(i+1)</sub> and H<sub>α(i)</sub>-N<sub>(i+1)</sub> and were strong to medium between all neighbouring residues with the exception of X2-P3, X9-P10, X13-P14 and F17-A18 (Figure 5). The latter of the four pairs proved difficult to determine as the amide protons of F17 and A18 shared a similar <sup>1</sup>H shift (8.28 ppm and 8.29 ppm respectively) (Figure 5). A weak NOE correlation was however observed between 7.23 ppm (F17<sub>δ</sub>) to 8.29 ppm and 1.62 ppm (A18<sub>β</sub>), suggesting the F17-A18 sequence (Figure 5).



**Figure 4.** Structural elucidation demonstrated that **1** is a new class of linaridin family that the C-terminal is unmodified. The post-translational modifications of amino acids residues has been highlighted in red.

NOE correlations between H<sub>δ</sub> in proline and the neighbouring H in *N*-atom finally confirmed the sequences of X2-P3, X9-P10 and X13-P14 linkages, as illustrated in Figure 4. Many of the <sup>13</sup>C could also be assigned, and in some cases inter residue connectivity could be corroborated using HMBC correlations. Using the advanced Marfey's analysis,<sup>19</sup> we confirmed that the absolute configuration of the unmodified amino acid residues is *L*-configuration, consistent with the above bioinformatics analysis. When tested for various antibacterial and cytotoxic activities, **1** displayed weak activity. The exact biological role of this peptide remains undefined.



**Figure 5.** NOE correlations of legonaridin **1** confirming the sequence of the amino acid residues. X = Dhb residue.

**Table 1.** NMR spectroscopic data (600 MHz, CD<sub>3</sub>OH, 298K) for legonaridin **1**

Residue	C	H <sub>N</sub>	C <sub>α</sub>	H <sub>α</sub>	C <sub>β</sub>	H <sub>β</sub>	C <sub>γ</sub>	H <sub>γ</sub>	Others
---------	---	----------------	----------------	----------------	----------------	----------------	----------------	----------------	--------

<b>I1</b>	175.9	N/A	73.1	3.85	34.7	2.14	24.1	1.21, 1.59	Cδ/ Hδ 11.3/0.96, NMe <sub>2</sub> 42.1/2.94
							15.2	1.17	
<b>X2</b>	168.4	10.27	130.7	N/A	124.0	5.95	12.2	1.85	
<b>P3</b>	<i>nk</i>	N/A	62.4	4.41	30.7	2.01, 2.36	25.6	1.99	Cδ/ Hδ 51.4/3.69, 3.88
<b>L4</b>	168.7	8.13	53.5	4.38	40.1	1.71, 1.75	25.6	1.71	Cδ/ Hδ 23.2/0.98 Cδ/ Hδ 21.1/0.92
<b>A5</b>	178.0	8.11	51.6	4.17	16.9	1.48	N/A	N/A	
<b>X6</b>	167.2	9.22	130.7	N/A	130.7	6.52	12.3	1.78	
<b>L7</b>	<i>nk</i>	7.80	54.2	4.35	40.6	1.65, 1.79	25.6	1.77	Cδ/ Hδ 23.5/0.99 Cδ/ Hδ 21.4/0.93
<b>A8</b>	174.4	8.02	50.6	4.53	17.4	1.52	N/A	N/A	
<b>X9</b>	168.7	9.26	131.9	N/A	124.9	5.91	12.2	1.79	
<b>P10</b>	<i>nk</i>	N/A	63.4	4.34	30.4	1.92, 2.40	26.1	2.08, 1.92	Cδ/ Hδ 50.2/3.58, 3.65
<b>E11</b>	175.1	7.95	63.4	4.33	26.7	2.21	31.4	2.49	Cδ 175.9
<b>A12</b>	175.5	8.01	52.1	4.28	17.1	1.43	N/A	N/A	
<b>X13</b>	169.3	9.36	131.9	N/A	124.7	5.80	12.2	1.84	
<b>P14</b>	<i>nk</i>	N/A	63.9	4.38	30.9	1.97, 2.47	26.1	2.17, 1.95	Cδ/ Hδ 50.3/3.52, 3.70
<b>V15</b>	175.6	7.70	64.4	3.79	30.0	2.29	20.6	1.13	
							19.3	1.02	
<b>G16</b>	166.3	8.07	45.4	3.84	N/A	N/A	N/A	N/A	
<b>F17</b>	176.1	8.28	59.0	4.28	37.2	3.20, 3.16	N/A	N/A	C1 - 137.4 C2,6 - 129.7 / 7.23 C3,5 - 127.6 / 7.20 C4 - 129.1 / 7.26
<b>A18</b>	177.5	8.29	53.6	4.11	16.2	1.62	N/A	N/A	
<b>A19</b>	177.1	8.52	53.6	4.19	16.2	1.63	N/A	N/A	
<b>X20</b>	167.7	9.33	130.2	N/A	129.9	6.37	12.6	1.76	
<b>S21</b>	174.2	7.96	59.9	4.27	61.7	3.97, 3.85	N/A	N/A	
<b>A22</b>	178.0	8.17	53.9	4.11	15.8	1.47	N/A	N/A	
<b>X23</b>	167.7	9.39	130.8	N/A	128.8	6.28	12.2	1.72	
<b>A24</b>	177.8	8.04	53.6	4.10	16.0	1.56	N/A	N/A	
<b>A25</b>	177.8	8.01	53.5	4.20	16.0	1.57	N/A	N/A	
<b>A26</b>	177.3	8.15	53.6	4.22	16.2	1.63	N/A	N/A	
<b>V27</b>	177.4	8.36	65.4	3.63	30.0	2.25	17.0	1.18	
							19.4	0.99	
<b>N28</b>	173.5/176.6	8.34	55.0	4.44	37.3	2.85, 2.92	173.5/176.6	N/A	γNH <sub>2</sub> - 7.02, 7.21
<b>M29</b>	174.4	8.16	57.7	4.24	31.1	2.54	50.4	3.21	εCH <sub>3</sub> - 38.0 / 2.65
<b>I30</b>	175.9	8.45	64.3	3.98	36.3	2.20	26.4	1.29, 1.64	Cδ/ Hδ 11.1/0.98
							16.9	1.18	
<b>X31</b>	168.2	9.84	131.3	N/A	129.6	6.29	12.5	1.77	
<b>H32</b>	<i>nk</i>	7.78	56.6	4.47	26.6	3.24, 3.40	N/A	N/A	C1 131.07 C2 - / 7.51 C3 135.0 / 8.76



<b>D33</b>	<i>nk</i>	8.25	<i>nk</i>	4.70	35.6	3.05, 3.15	<i>nk</i>	N/A	
<b>V34</b>	175.9	8.21	63.2	3.91	30.4	2.22	19.5	1.08	
							19.4	1.04	
<b>X35</b>	166.3	9.06	130.3	N/A	133.4	6.72	13.0	1.8	
<b>R36</b>	<i>nk</i>	7.48	55.2	4.29	29.3	1.85	26.4	1.65	Cδ/ Hδ 50.5/3.20, Cε 158.4 Nε 7.37
<b>H37</b>	172.8	8.39	52.4	4.79	27.5	3.18, 3.35	N/A	N/A	C1 130.7 C2 118.6/ 7.40 C3 134.7 / 8.79

n/k = not known, not found n/a = not applicable

### Database mining suggests that **1** belongs to a new subgroup of linaridin RiPPs

A homolog search of the precursor peptide LegA in the public domain indicated that LegA also share high sequence identity (60%-70% AA identity) with five hypothetic proteins in other *Streptomyces* strains (Figure S10), which offers six gene clusters to compare and contrast their immediate environment. These five putative gene clusters possess striking similarity to the *leg* cluster. So far the genes encoding methyltransferases and CypL homolog proteins are highly conserved in the clusters of cypemycin, grisemycin, legonaridin and these five putative linaridins. ABC transporter, FAD-dependent oxidoreductase, Alpha/beta(Ab) hydrolase fold protein and HTT membrane protein are unique to the ones of legonaridin and the five putative linaridins. No *cypI* and *cypD* homologs can be identified in these five new putative clusters and the *leg* cluster. Phylogenetic analysis suggested that these five new putative precursor peptides and LegA form a distinct branch with the ones of cypemycin and grisemycin (Figure S11). Taken together, we proposed that legonaridin **1** should be a new subgroup of linaridin RiPPs.

### Conclusions

In conclusion, legonaridin **1**, a new member of linaridin RiPP was discovered from the soil bacterium *Streptomyces* sp. CT34 through metabolite profiling and chemical isolation. MS<sup>n</sup>-guided genome mining approach allowed identification of the precursor peptide of legonaridin and the putative *leg* gene cluster responsible for the biosynthesis of legonaridin. The absolute structure of legonaridin **1** was established through peptidogenomics, genomic-driven analysis, extensive spectroscopic analysis, and advanced Marfey's analysis revealed that **1** is a 37-mer linear peptide with some unusual chemical modification, including an *N*-terminal *N,N*-dimethylalanine and dehydrothreonine moieties. Despite of the same chemical features in **1** as the ones observed in cypemycin and grisemycin, legonaridin **1** has a structurally different *C*-terminus with a free carboxylic acid, suggesting that **1** forms a new subgroup of the rare linaridin RiPP family, in line with our bioinformatics analysis.

### Experimental Section

### General information

NMR data were acquired on a Varian NMR spectrometer operating at 600 MHz for proton and 150 MHz for carbon. High resolution mass spectrometric data were obtained using a Thermo Instruments MS system (LTQ XL/LTQ Orbitrap Discovery) coupled to a Thermo Instruments HPLC system (Accela PDA detector, Accela PDA autosampler and Accela pump). The following conditions were used: capillary voltage 45V, capillary temperature 260 °C, auxiliary gas flow rate 10-20 arbitrary units, sheath gas flow rate 40-50 arbitrary units, spray voltage 4.5 kV, mass range 100-2000 amu (maximum resolution 30,000). HPLC separations were carried out using a Phenomenex Luna reversed-phase (C18 250 × 10mm, L × i.d.) column connected to a Waters 1525 Binary HPLC pump Chromatograph with a 2998 PDA detector, column heater and in-line degasser. Detection was achieved on-line through a scan of wavelengths from 200-400 nm. The optical rotation was measured using Bellingham+Stanley 37-410 polarimeter, UK. The UV spectrum was acquired using Agilent Cary 60 UV-Vis spectrometer, UK. The IR spectrum was measured using Perkin Elmer FT-IR spectrometer Ultra Two, UK. Diaion HP-20 was obtained from Resindion S.R.L., a subsidiary of Mitsubishi Chemical Co., Binasco, Italy. All solvents used throughout were HPLC-grade and purchased from Sigma-Aldrich (Germany) through a Ghana based agent (Huge Limited). Sephadex LH-20 (25-100µm) was purchased from GE Healthcare (Little Chalfont, UK).

Plasmids and bacterial strains are summarized in Table S2. Restriction enzymes were purchased from New England Biolabs and chemicals from Sigma-Aldrich. Oligonucleotide primers (Table S2) were purchased from Sangon Biotech.

### Isolation of *Streptomyces* sp. CT34

The rhizosphere sample was obtained near a *Caesalpinioideae* tree (*Tamarindus Indica*, Africa) in the Botanic Garden in the University of Ghana, Legon (5°39'32.72" N, 0°11'55.26" W). *Streptomyces* sp. CT34 was isolated using the procedure outlined as follows: the soil sample was suspended in sterile water. The suspension was transferred to modified ISP2 agar plates (oxoid agar, 15g; glucose, 4g; yeast extract, 4g; malt extract, 10g; pH 5.5) supplemented with nalidixic acid (25 µg/mL) and nystatin (25 µg/mL) and incubated at 28 °C for 7 days. The pure strain was obtained after repeatedly transferred and streaked on different fresh ISP2 agar plates.

### Cultivation of *Streptomyces* sp. CT34

The pure strain was maintained on ISP2 agar plates at 28 °C. For seed culture, *Streptomyces* sp. CT34 spores were inoculated into modified ISP2 liquid medium (50 mL, pH 5.5) in an Erlenmeyer flask (250 mL) and cultivated in a rotary incubator (180 rpm, Barnstead Lab-line MaxQ 5000) at 30 °C for 5 days. For the large scale fermentation, the seed culture was transformed to modified ISP2 (pH 5.5) (4 litres), containing glucose (4g/L), yeast extract (4g/L) and malt extract (10g/L). Diaion HP-20 (50g/L) was also added during the fermentation.

### Extraction and Isolation

After 7-day fermentation, HP-20 resin was filtered under the reduced pressure. HP-20 resin was washed with distilled water, and then extracted with methanol (3x500 mL). The successive MeOH extracts were combined and concentrated under reduced pressure, followed solvent partition using the modified Kupchan method. The dichloromethane (CH<sub>2</sub>Cl<sub>2</sub>) fraction was subjected to semi-preparative reversed-phase HPLC (C18 Sunfire™ 250x10 mm Waters; A: water, B: Methanol; gradient 0-40 min 60-100% B; flow rate 1.5 mL/min) to afford 9 mg of pure **1**.

Compound **1**: white powder;  $[\alpha]_D^{25}$  -1.1 (c 0.06 in MeOH); UV (MeOH)  $\lambda_{max}$  (log  $\epsilon$ ) 281 (4.20) 213 (4.20) nm; IR (film)  $\nu_{max}$  3416, 2978, 1647, 1453, 1383, 1087 cm<sup>-1</sup>; <sup>1</sup>H, HSQC, COSY, TOCSY, HMBC and ROESY NMR data see Table 1; HRESIMS  $m/z$  901.223 [M + 4H]<sup>4+</sup>  $\Delta$  5.042 ppm calculated for as C<sub>165</sub>H<sub>254</sub>N<sub>45</sub>O<sub>44</sub>S.

#### De novo MS/MS for identification of sequence tags.

Data were acquired on a Thermo LTQ-FT 7T instrument using an Advion Nanomate nanoelectrospray infusion robot. The sample was diluted in 50/50 acetonitrile water with 0.1% formic acid. Samples were infused at approximately 200 nL/min with an applied voltage of 1.8 kV. Gas settings were all set to 0 with an inlet temperature of 250 °C. The instrument was manually controlled during acquisition and MS parent ion spectra were collected from 150-2,000  $m/z$  in profile mode as analyzed in the FTMS cell. MS<sup>2</sup> and MS<sup>3</sup> spectra were collected at varying collision energies, isolation windows, and acquisition windows optimized for optimal signal and isolation efficiency. Acquired spectra were manually interpreted. High-resolution, but lower signal to noise FT spectra were compared against low-resolution but higher signal to noise iontrap spectra. FT peaks without isotopes and absent from the iontrap spectra were identified as electronic noise. FT transients of were collected at 128-256k. All spectra analyzed were averaged (18-356).

#### Genome sequence

The genome sequence of CT34 was deposited in GenBank under the accession no: JSFP00000000.<sup>14</sup> The legonaridin biosynthetic gene cluster was identified using seven sequence tags generated by MS<sup>n</sup> fragmentation as the sequence query.

The open reading frames (ORFs) were deduced from the sequence by performing FramePlot 4.0 beta program (<http://nocardia.nih.gov/fp4/>). The corresponding deduced proteins were compared with other known proteins in the databases using the Basic Local Alignment Search Tool (BLAST) (<http://www.ncbi.nlm.nih.gov/blast/>). Amino acid sequence alignments were performed with the CLUSTALW algorithm from BIOLOGYWORKBENCH 3.2 software (<http://workbench.sdsc.edu>).

#### Advanced Marfey's method

A 0.5 mg quantity of Legonaridin **1** was hydrolyzed in 0.3 mL of 6 N HCl at 115 °C for 6 h; the reaction was then cooled rapidly by placing the reaction vial into ice water for 3 min. The reaction solvent was evaporated *in vacuo*, and residual HCl was completely removed by adding 0.5 mL of H<sub>2</sub>O and removing the solvent three times. The dried hydrolysate was lyophilized for 24 h. The hydrolysate containing the free amino acids was

divided into two portions, and each portion was transferred into a 10 mL vial. The hydrolysate was then dissolved in 50  $\mu$ L of 1 N NaHCO<sub>3</sub>. To each of the two vials containing the dissolved free amino acids, 50  $\mu$ L of either 10 mg/mL L-FDLA (1-fluoro-2,4-dinitrophenyl-5-L-leucine amide) or D-FDLA (1-fluoro-2,4-dinitrophenyl-5-D-leucine amide) in acetone was added. The reaction mixtures were incubated at 50 °C for 30 min. A 50  $\mu$ L aliquot of 2 N HCl was added to neutralize the reaction, followed by the addition of 300  $\mu$ L of aqueous 50% CH<sub>3</sub>CN. A 10  $\mu$ L aliquot of each reaction mixture was analysed by LC-MS with a gradient solvent system (20% to 100% CH<sub>3</sub>CN containing 0.1% formic acid over 40 min, C18 reversed-phase column 100  $\times$  4.6 mm). L-FDAA derivatives eluted before D-FDAA derivatives for all amino acid residues in the hydrolysate. Therefore, the absolute configurations of the unmodified amino acid residues in **1** were determined as L.

### Generation of Knockout Constructs

To delete the *leg* gene cluster, a 1880 bp upstream fragment and a 1990 bp downstream fragment were amplified from genomic DNA of *Streptomyces* sp. CT34 by PCR using the primers U1/U2 and D1/D2, respectively (Table S2). The PCR products were purified using the Omega PCR purification kit and were cloned in pUC-T simple vector to generate constructs, pWHU2261 and pWHU2262 (Table S2). After sequencing verification, the fragments were digested by NdeI/XhoI and XhoI/HindIII from pWHU2261 and pWHU2262 respectively. The obtained fragments were cloned into the NdeI and HindIII site of pYH7 to fulfil the in-frame deletion construct pWHU2263 which was then transferred into CT34 strain via *E. coli-Streptomyces* conjugation.<sup>20</sup> A PCR-based screening allowed identification of the in-frame deletion mutant M-ZY using two PCR primer pairs of L1/L2 and R1/R2 (table S2).

### Acknowledgements

KK, MJ, RE and HD are grateful for financial support through the Leverhulme Trust-Royal Society Africa award (AA090088). MJ, RE, HD, JT and KH thank EU FP7 for financial support (contract no. 312184). HD thanks School of Natural and Computing Sciences, University of Aberdeen, for a PhD scholarship to XLW. PCD gratefully acknowledges grants from National Institute Health (GM097509 and GMS10RR029121). We thank the Bruker Therapeutic Discovery Mass Spectrometry Center for recording the MS<sup>n</sup> spectra.

### Notes and references

- 1 P. J. Knerr, W. A. van der Donk, *Annu. Rev. Biochem.* 2012, **81**, 479–505.
- 2 P. G. Arnison, *et al. Nat. Prod. Rep.* 2013, **30**, 108–160.
- 3 J. McIntosh, M. S. Donia, E. W. Schmidt, *Nat. Prod. Rep.* 2009, **26**, 537–559.
- 4 H. Itoh, M. Inoue, *Acc. Chem. Res.* 2013, **46**, 1567–1578.
- 5 T. Sugawara, S. Matsunaga, *Tetrahed.* 1994, **35**, 719–720.
- 6 T. Hamada, S. Matsunaga, G. Yano, N. Fusetani, *J. Am. Chem. Soc.* 2005, **127**, 110–118.
- 7 J. Claesen, M. Bibb, *Proc. Natl. Acad. Sci. USA* 2010, **107**, 16297–16302.
- 8 R. D. Kersten, Y.-L. Yang, Y. Xu, P. Cimermancic, S.-J. Nam, W. Fenical, M. Fischbach, B. S. Moore, P. C. Dorrestein, *Nat. Chem. Biol.* 2011, **7**, 794–802.
- 9 a) H. Deng, L. Ma, N. Bandaranayaka, Z. Qin, G. Mann, K. Kyeremeh, Y. Yu, T. Shepherd, J. H. Naismith, D. O'Hagan, *ChemBioChem* 2014, **15**, 364–368; b) L. Ma, A. Bartholome, M. H. Tong, Z. Qin, Y. Yu, T. Shepherd, K. Kyeremeh, H. Deng, D. O'Hagan, *Chem. Sci* 2015, **6**, 1414–19; c) S. Huang, J. Tabudravu, S. S. Elsayed, J. Travert, D. Peace, M. Tong H., K. Kyeremeh, L. Trembleau, R. Ebel, M. Jaspars, Y. Yu, H. Deng, *Angew Chemie Int. Ed.* 2015, Online, DOI:10.1002/anie.201502902.

- 10 K. Kyeremeh, K. S. Acquah, A. Sazak, W. Houssen, J. Tabudravu, H. Deng, M. Jaspars, *Mar. Drugs* 2014, **12**, 999–1012.
- 11 K. Kyeremeh, K. S. Acquah, M. Çamaş, J. Tabudravu, W. Houssen, H. Deng, M. Jaspars, *Mar. Drugs* 2014, **12**, 5197–5208.
- 12 L. A. Morris, J. Jantina, K. V. D. Bosch, K. Versluis, S. Thompson, M. Jaspars, *Tetrahed* 2000, **56**, 8345–8353.
- 13 R. Rink, A. Kuipers, E. de Boef, K. J. Leenhouts, A. J. M. Driessen, G. N. Moll, O. P. Kuipers, *Biochem* 2005, **44**, 8873–8882.
- 14 Y. Zhai, B. Cheng, J. Hu, K. Kyeremeh, X. Wang, M. Jaspars, H. Deng, Z-X. Deng, K. Hong, *Genome Announc.* 2015, **3**, e01508–14.
- 15 K. Blin, H. M. Medema, D. Kazempour, M. A. Fischbach, R. Breitling, E. Takano, T. Weber, *Nucleic Acids Res.* 2013, **41**, W204–W212.
- 16 V. Prabhu, *et al. J. Plant Physiol.* 1996, **149**, 246–250.
- 17 D. S. Wishart, *et al., Biochem.* 1992, **31** 1647–1651.
- 18 P. Crews, *et al., Organic Structure Analysis*, Oxford University Press, New York, 2<sup>nd</sup> Ed. 2010.
- 19 K. Fujii, Y. Ikai, T. Mayumi, H. Oka, M. Suzuki, K. Harada, *Anal. Chem.* 1997, **69**, 3346–3352.
- 20 T. Kieser, M. J. Bibb, M. J. Buttner, K. F. Chater, D. A. Hopwood *Practical Streptomyces Genetics*, **2000**. Norfolk, United Kingdom. The John Innes Foundation.