

Supplementary Information

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

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

^aMarine Biodiscovery Centre, Department of Chemistry, University of Aberdeen, Aberdeen AB24 3UE Scotland, United Kingdom. Email: h.deng@abdn.ac.uk; m.jaspars@abdn.ac.uk.

^bKey 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.

^cSkaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California, USA.

^dDepartment of Pharmacology and Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California, USA.

^eDepartment of Molecular Microbiology, John Innes Centre, Norwich NR4 7UH, United Kingdom.

^fDepartment of Chemistry, FGO Torto Building, University of Ghana, Legon, Ghana.

^gPharmacognosy Department, School of Pharmacy, Beni-Suef University, Beni-Suef 62514, Egypt.

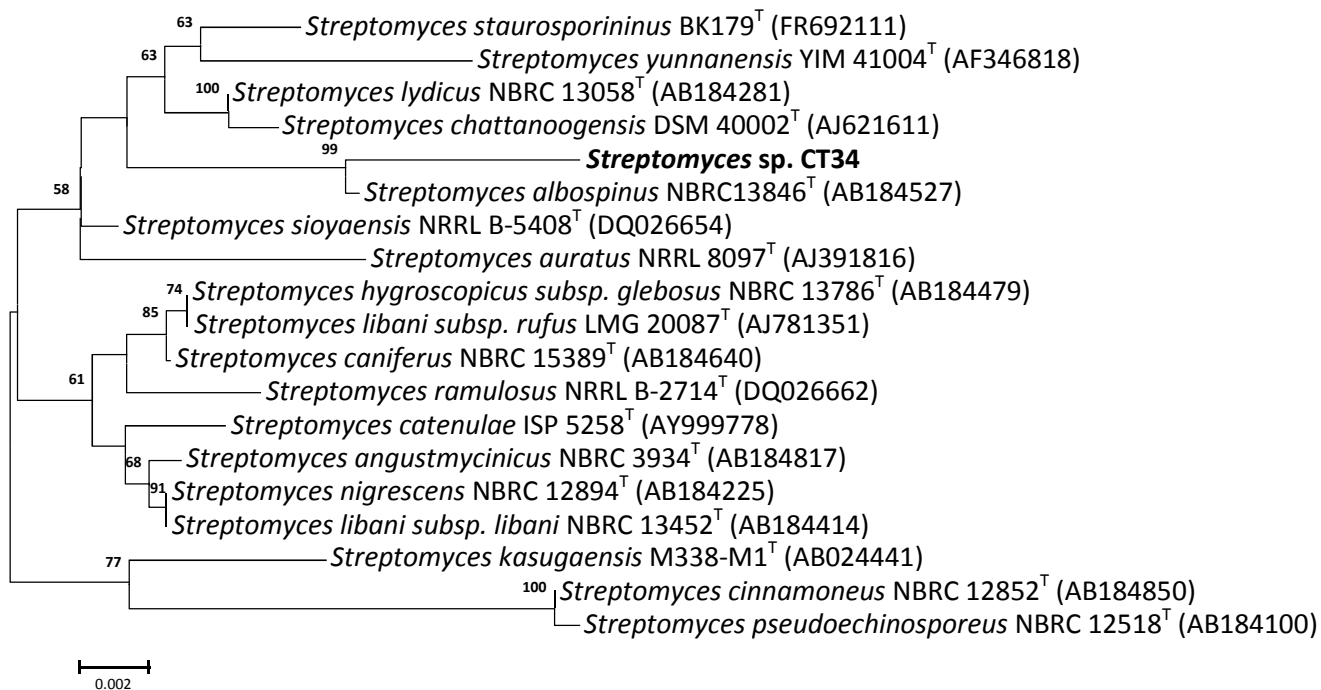


Figure S1: Neighbour-joining phylogenetic tree based on nearly complete 16S rRNA gene sequences. The relationships between strain *Streptomyces* sp. CT34 and the type strains of phylogenetically close species of the genus *Streptomyces* were analyzed. Numbers at nodes are percentage bootstrap values based on 1000 replicates; only values >50 % are shown. Bar, 0.002 substitutions per nucleotide position.

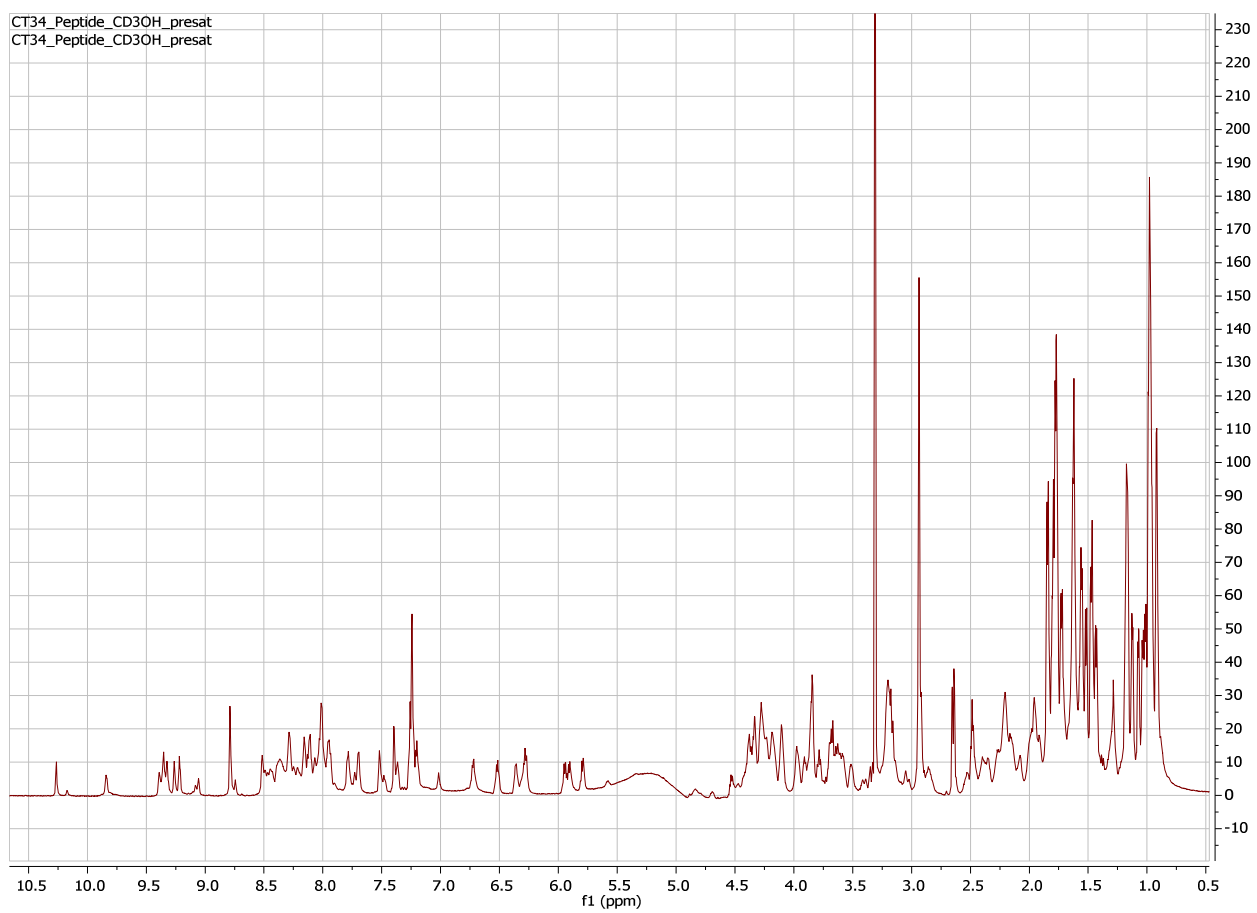


Figure S2. ¹H NMR spectrum of legonaridin 1 in CD₃OH.

110717_CT34_100x_m_FT_1 #1-185 RT: 0.00-2.94 AV: 185 NL: 1.89E6
 F: FTMS + p NSI Full ms [150.00-2000.00]

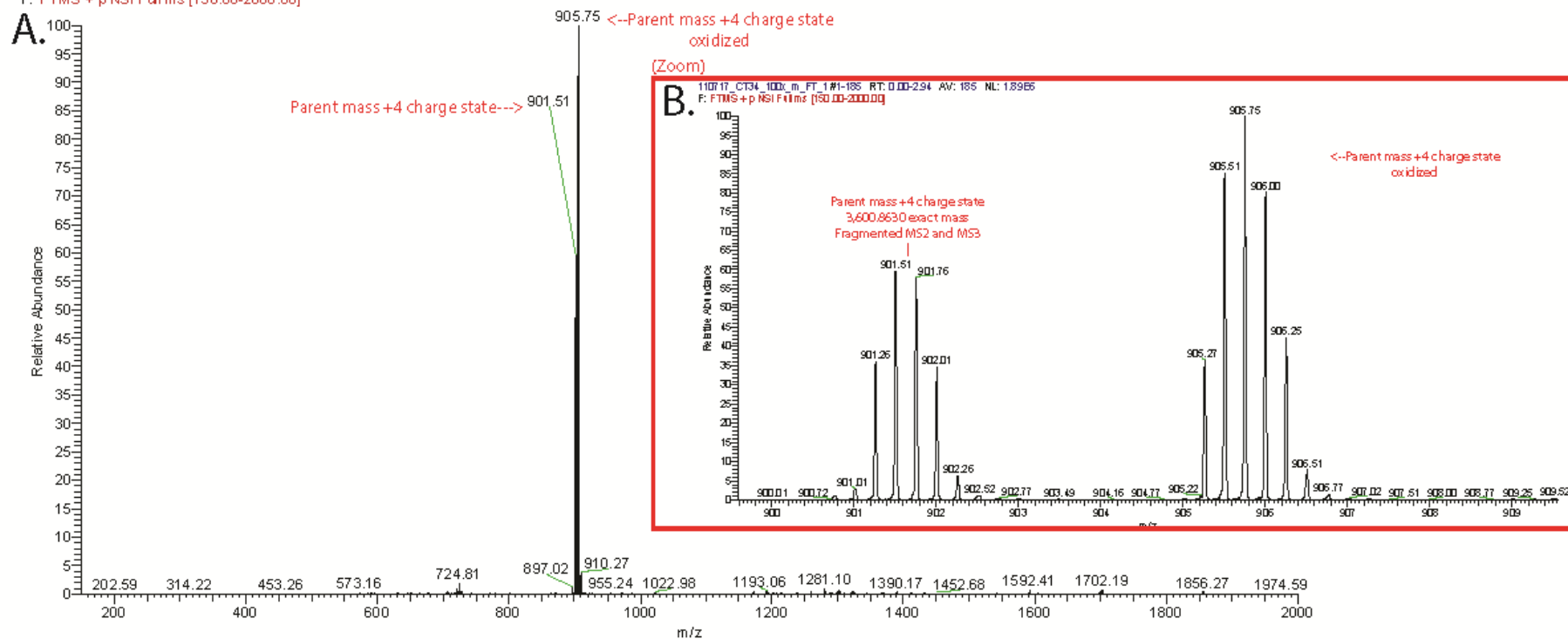


Figure S3 A and B. MS¹ analysis in an LTQ-FT MS on legonaridin **1** to generate 7 reverse-sequence tags. **(A)** MS¹ full scan data; **(B)** MS¹ zoom. All data are presented as m/z versus intensity in arbitrary units. Assigned peaks and mass shifts are noted. The direction of tags cannot be determined from this data. Any residue assigned as I may also be L. MS¹ data are annotated in red, MS² in green, and MS³ in purple.

110717_CT34_100x_m_FT_901_5_256_MS2 #5-248 RT: 0.05-4.09 AV: 244 NL: 5.20E3
 T: FTMS + p NSI Full ms2 901.51@cid30.00 [245.00-2000.00]

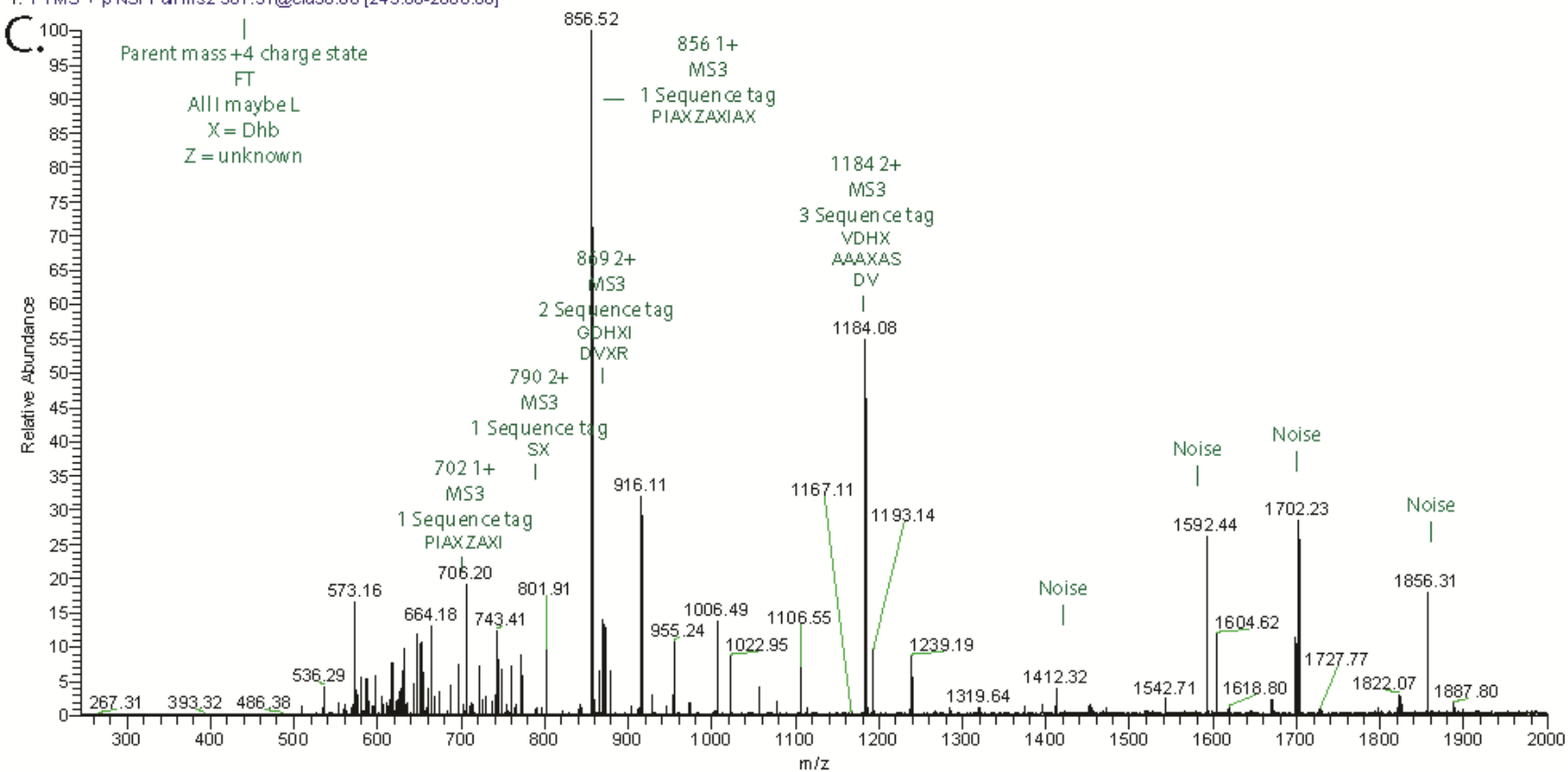


Figure S3C. MSⁿ analysis in an LTQ-FT MS on legonaridin **1** to generate 7 reverse-sequence tags. (C). MS² data. All data are presented as *m/z* versus intensity in arbitrary units. Assigned peaks and mass shifts are noted. The direction of tags cannot be determined from this data. Any residue assigned as I may also be L. MS¹ data are annotated in red, MS² in green, and MS³ in purple.

110717_CT34_100x_m_FT_901_5_128_MS2_30 #1-18 RT: 0.00-0.07 AV: 18 NL: 1.65E4
 F: ITMS + p NSI Full ms2 901.51@cid30.00 [245.00-2000.00]

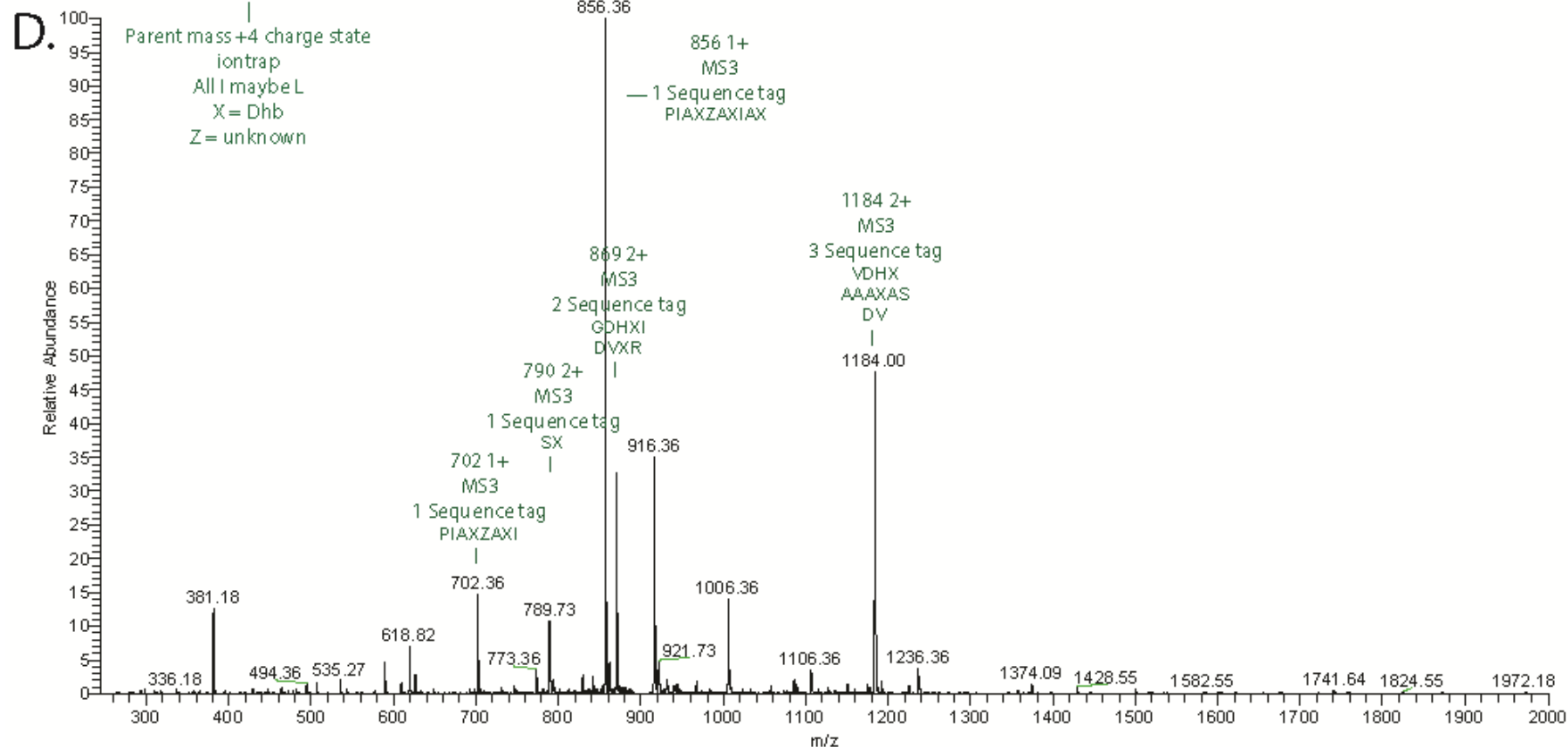


Figure S3D. MSⁿ analysis in an LTQ-FT MS on legonaridin **1** to generate 7 reverse-sequence tags. (D) MS² data. All data are presented as *m/z* versus intensity in arbitrary units. Assigned peaks and mass shifts are noted. The direction of tags cannot be determined from this data. Any residue assigned as I may also be L. MS¹ data are annotated in red, MS² in green, and MS³ in purple.

110717 CT34 100x m IT 901 702 MS3 256 #1-255 RT: 0.00-1.42 AV: 255 NL: 2.51E2
 T: ITMS + p NSI Full ms3 901.00@cid30.00 702.00@cid25.00 [190.00-2000.00]

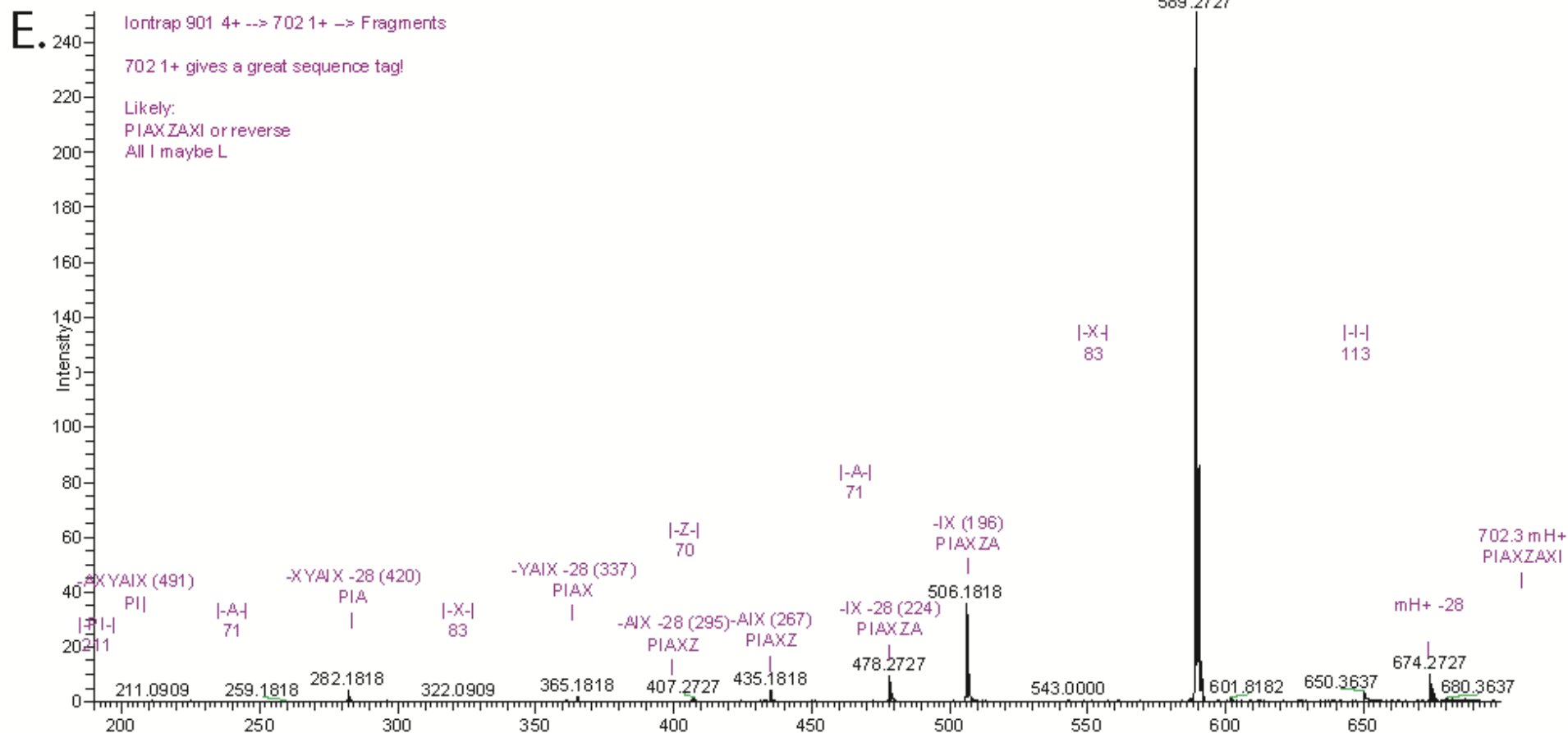


Figure S3E. MS⁰ analysis in an LTQ-FT MS on legonaridin **1** to generate 7 reverse-sequence tags. (E) MS³ is illustrated with annotations. All data are presented as *m/z* versus intensity in arbitrary units. Assigned peaks and mass shifts are noted. The direction of tags cannot be determined from this data. Any residue assigned as I may also be L. MS¹ data are annotated in red, MS² in green, and MS³ in purple.

110717 CT34 100x m IT 901 790 MS3 256 #1-256 RT: 0.00-1.42 AV: 256 NL: 4.36E1
T: ITMS + p NSI Full ms3 901.00@cid30.00 789.70@cid25.00 [215.00-2000.00]

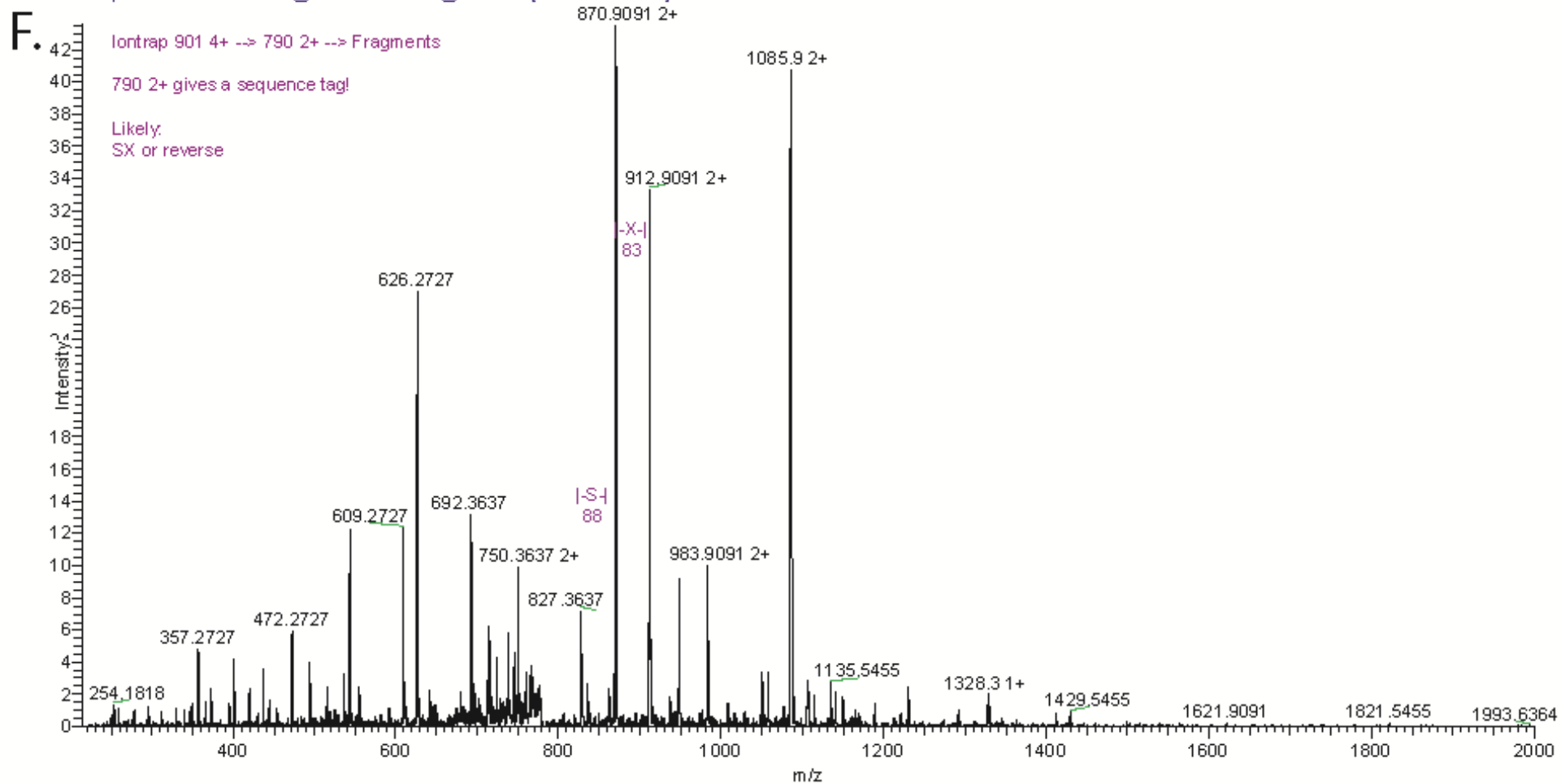


Figure S3F. MSⁿ analysis in an LTQ-FT MS on legonaridin **1** to generate 7 reverse-sequence tags. (F) is illustrated with annotations. All data are presented as m/z versus intensity in arbitrary units. Assigned peaks and mass shifts are noted. The direction of tags cannot be determined from this data. Any residue assigned as I may also be L. MS¹ data are annotated in red, MS² in green, and MS³ in purple.

110717 CT34 100x m IT 901 856 MS3 256 #1-256 RT: 0.00-1.41 AV: 256 NL: 2.41E3
 T: ITMS + p NSI Full ms3 901.00@cid30.00 856.00@cid30.00 [235.00-2000.00]

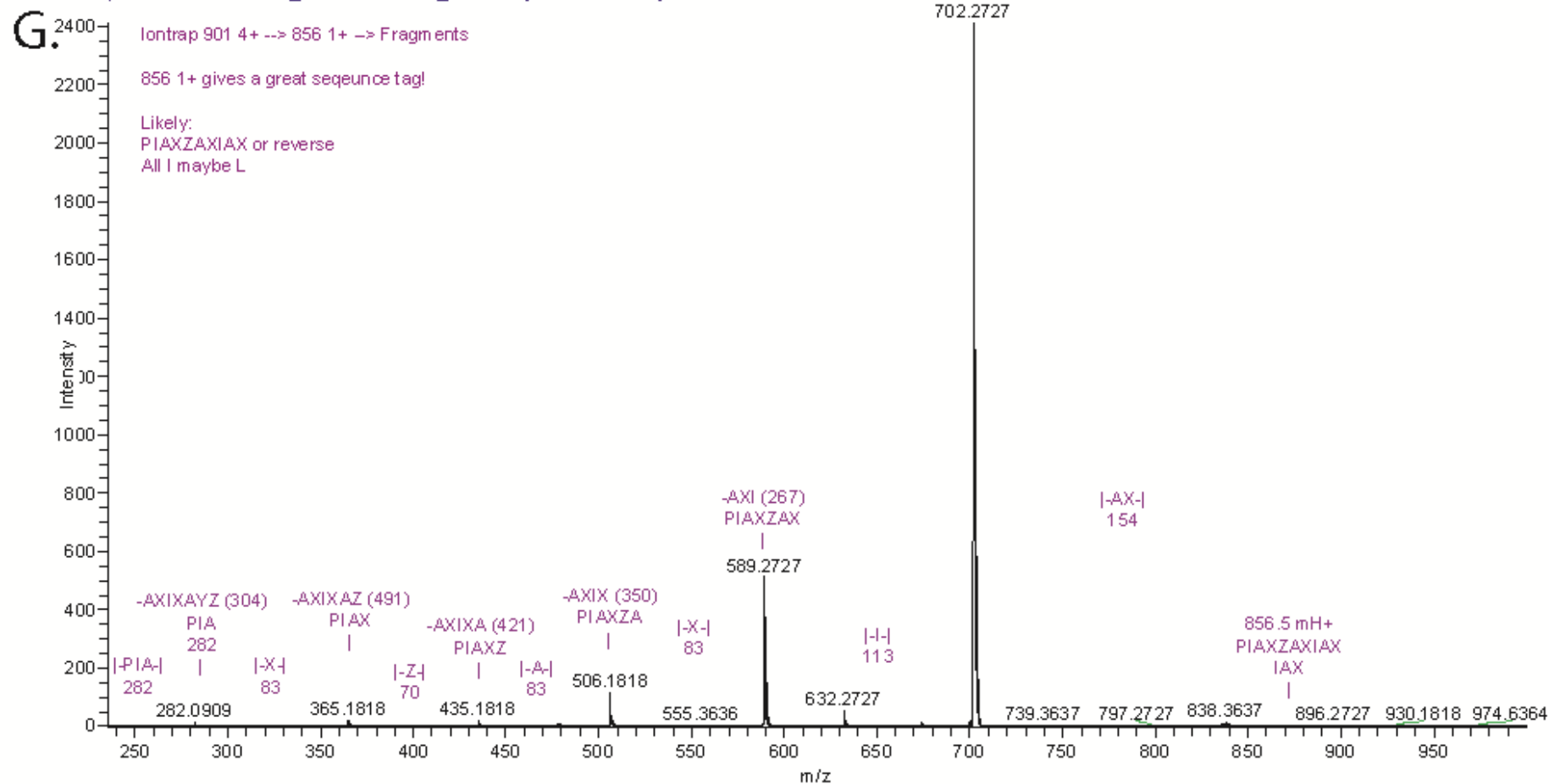


Figure S3G. MSⁿ analysis in an LTQ-FT MS on legonaridin **1** to generate 7 reverse-sequence tags. (G) MS³ is illustrated with annotations. All data are presented as *m/z* versus intensity in arbitrary units. Assigned peaks and mass shifts are noted. The direction of tags cannot be determined from this data. Any residue assigned as I may also be L. MS¹ data are annotated in red, MS² in green, and MS³ in purple.

110717 CT34 100x m IT 901 869 MS3 256 #1-256 RT: 0.00-1.42 AV: 256 NL: 2.72E2
 T: ITMS + p NSI Full ms3 901.00@cid30.00 869.00@cid25.00 [235.00-2000.00]

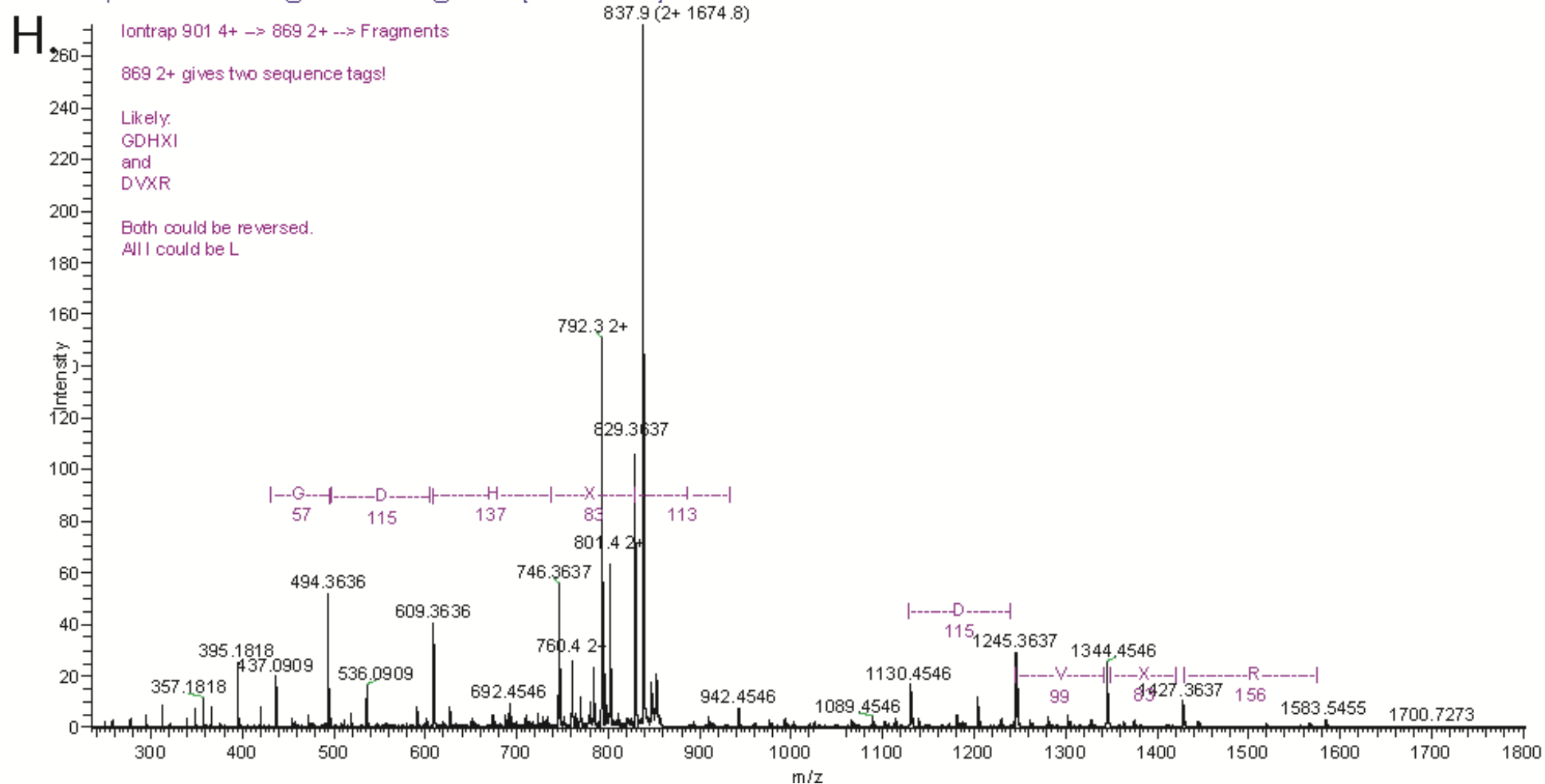


Figure S3H. MSⁿ analysis in an LTQ-FT MS on legonaridin **1** to generate 7 reverse-sequence tags. (H) MS³ is illustrated with annotations. All data are presented as *m/z* versus intensity in arbitrary units. Assigned peaks and mass shifts are noted. The direction of tags cannot be determined from this data. Any residue assigned as I may also be L. MS¹ data are annotated in red, MS² in green, and MS³ in purple.

110717 CT34 100x m IT 901 1184 MS3 1028 #1-365 RT: 0.00-1.94 AV: 365 NL: 3.25E2
 ITMS + p NSI Full ms3 901.00@cid30.00 1184.00@cid25.00 [325.00-2000.00]

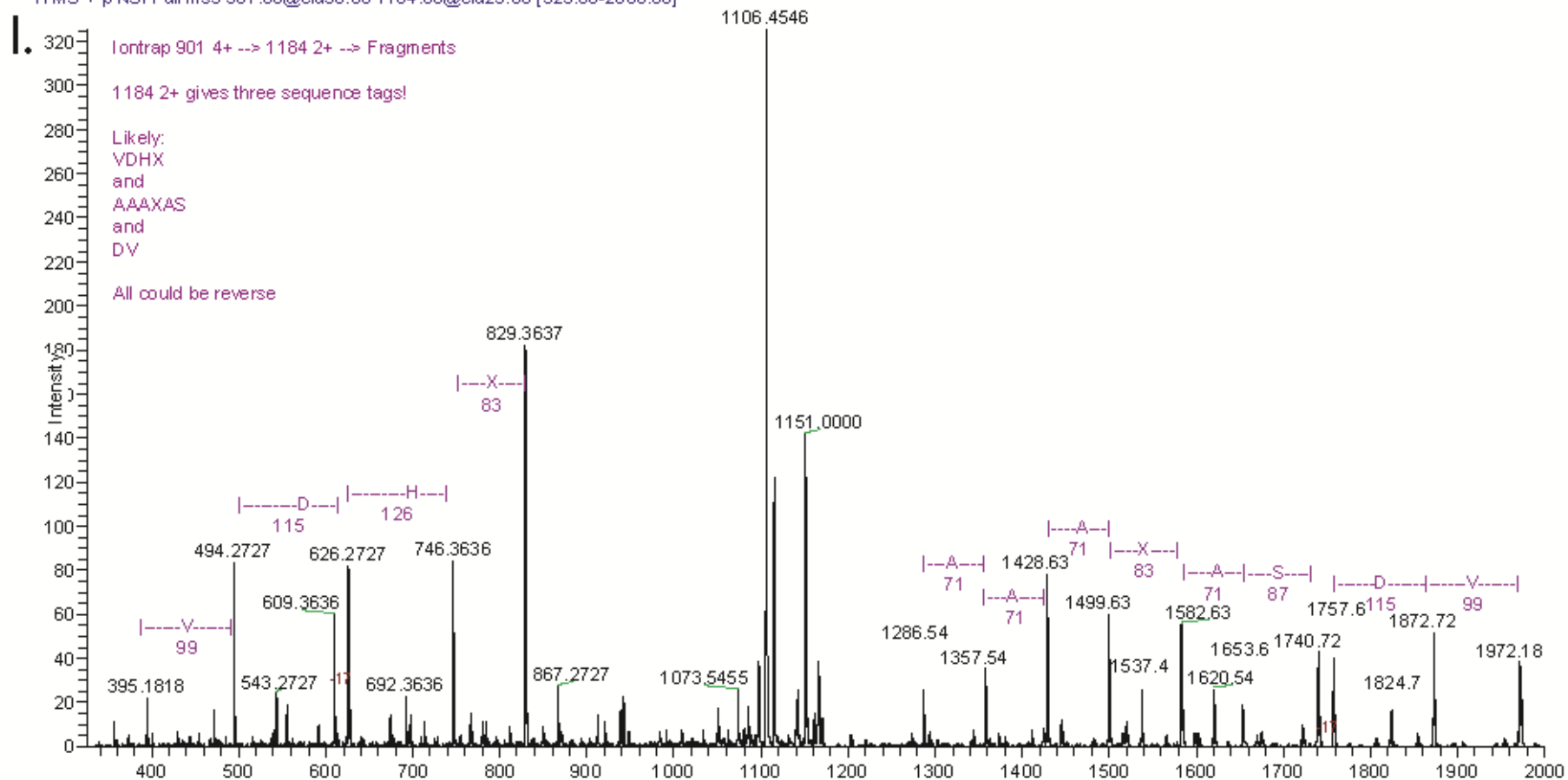


Figure S3I. MSⁿ analysis in an LTQ-FT MS on legonaridin **1** to generate 7 reverse-sequence tags. (I). MS³ is illustrated with annotations. All data are presented as *m/z* versus intensity in arbitrary units. Assigned peaks and mass shifts are noted. The direction of tags cannot be determined from this data. Any residue assigned as I may also be L. MS¹ data are annotated in red, MS² in green, and MS³ in purple.

Overlapping MS Fragments

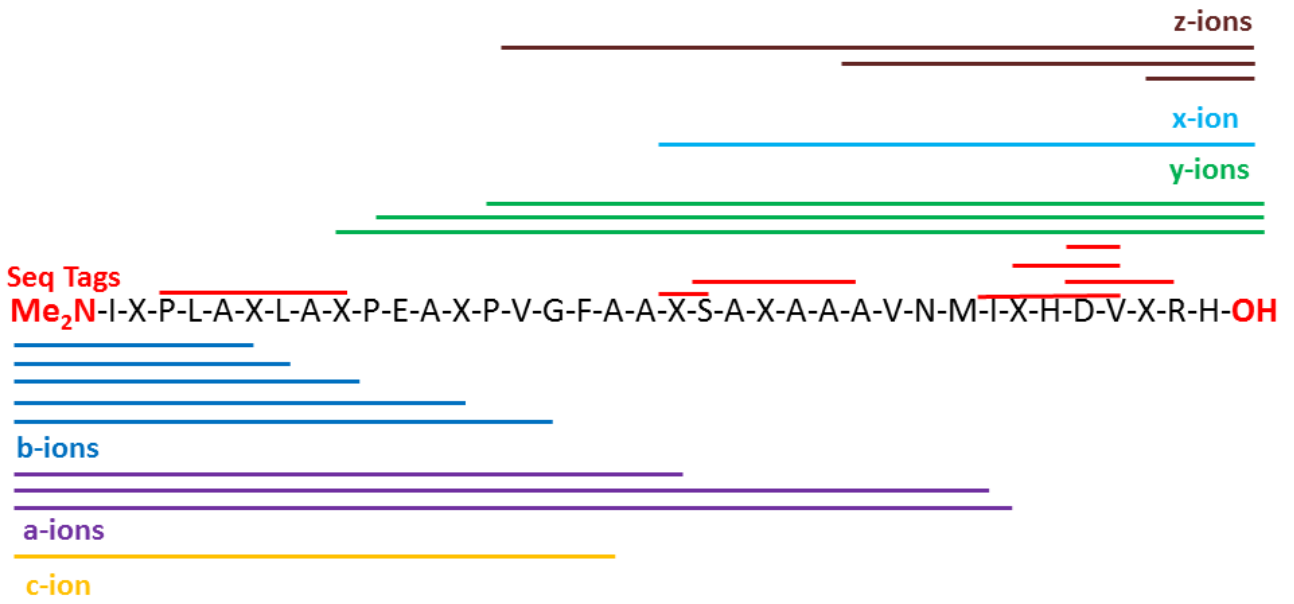


Figure S4 overlapping MS analysis fragments according to the predicted sequence of the mature peptide legonaridin.

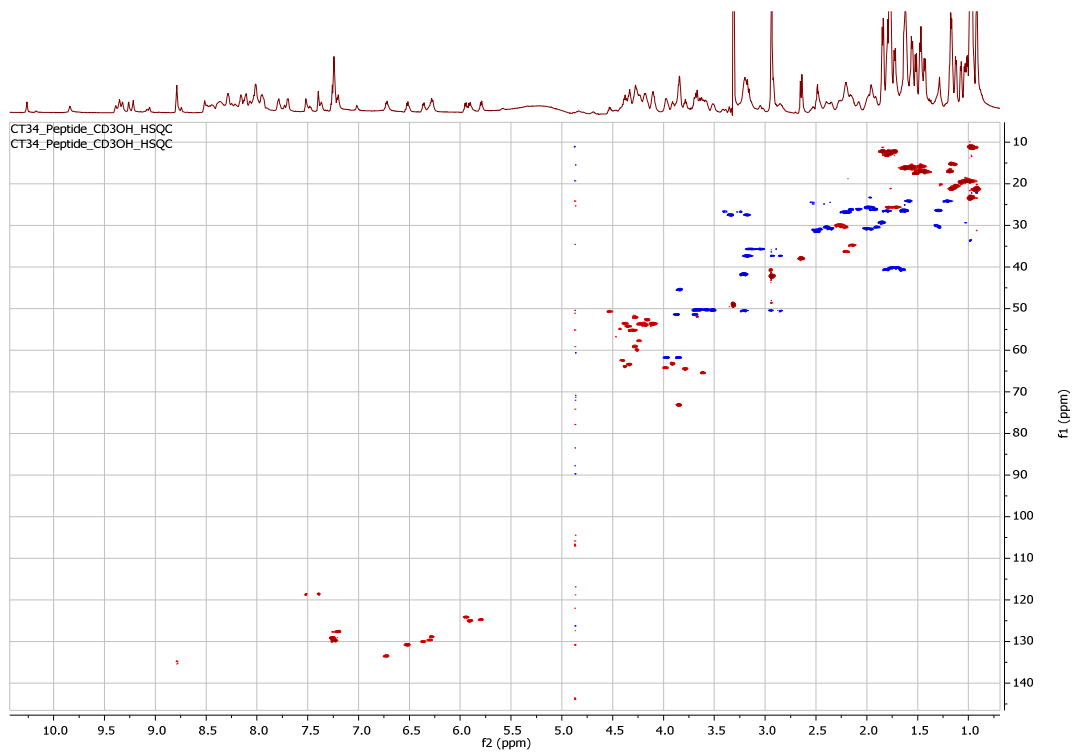


Figure S5. HSQC spectrum of legonaridin **1** in CD₃OH.

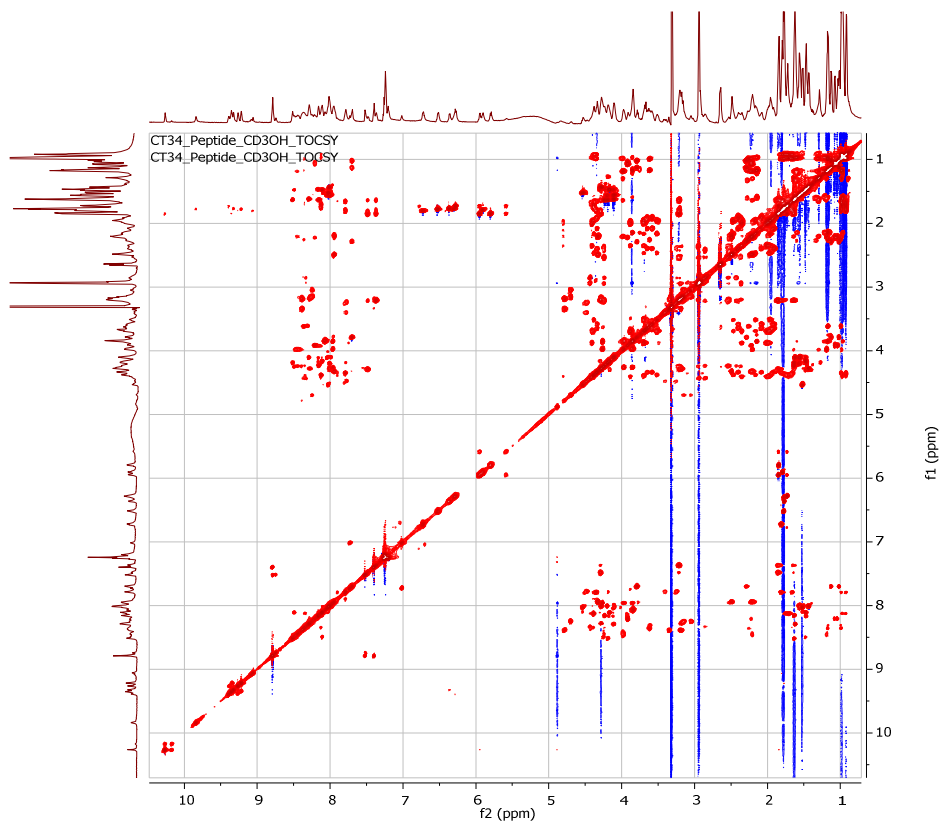


Figure S6. TOCSY spectrum of legonaridin **1** in CD₃OH.

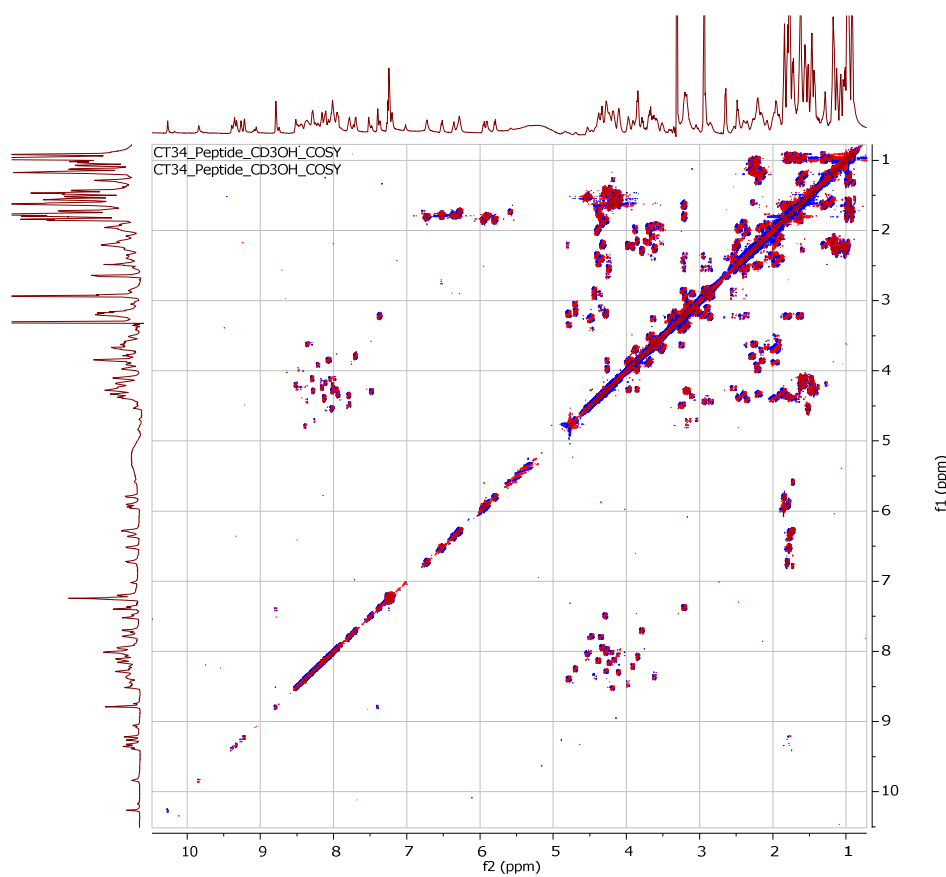


Figure S7. COSY spectrum of legonaridin **1** in CD₃OH.



Figure S8. HMBC spectrum of legonaridin **1** in CD₃OH.

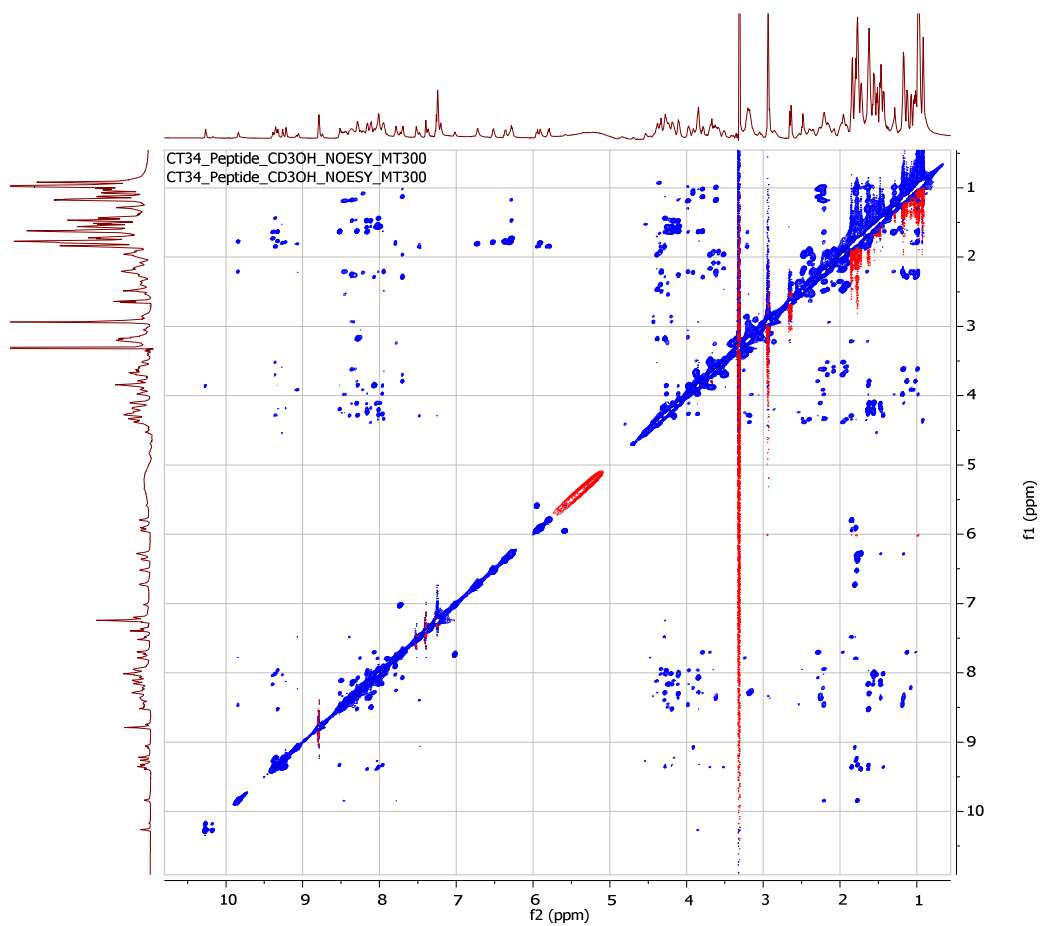
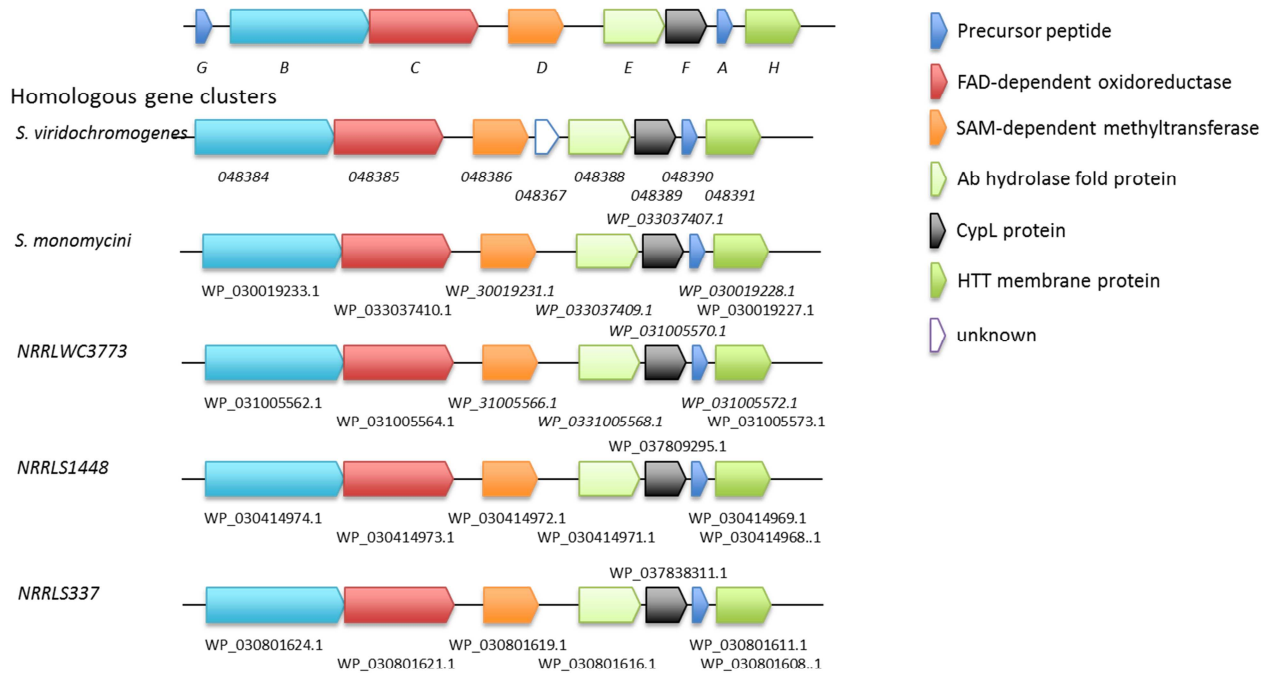


Figure S9. NOESY spectrum of legonaridin **1** in CD₃OH.

ALegonaridin cluster from *Streptomyces* sp. CT34**B**

Precursor peptide sequence comparison



Figure S10. (A) Schematic representation of five putative biosynthetic gene clusters of linaridin RiPPs with the one for **1**. (B) The sequence alignment of legonaridin **1** precursor peptide sequence and five putative precursor peptide sequences. The red arrow indicated the possible first amino acid the mature legonaridin RiPPs.

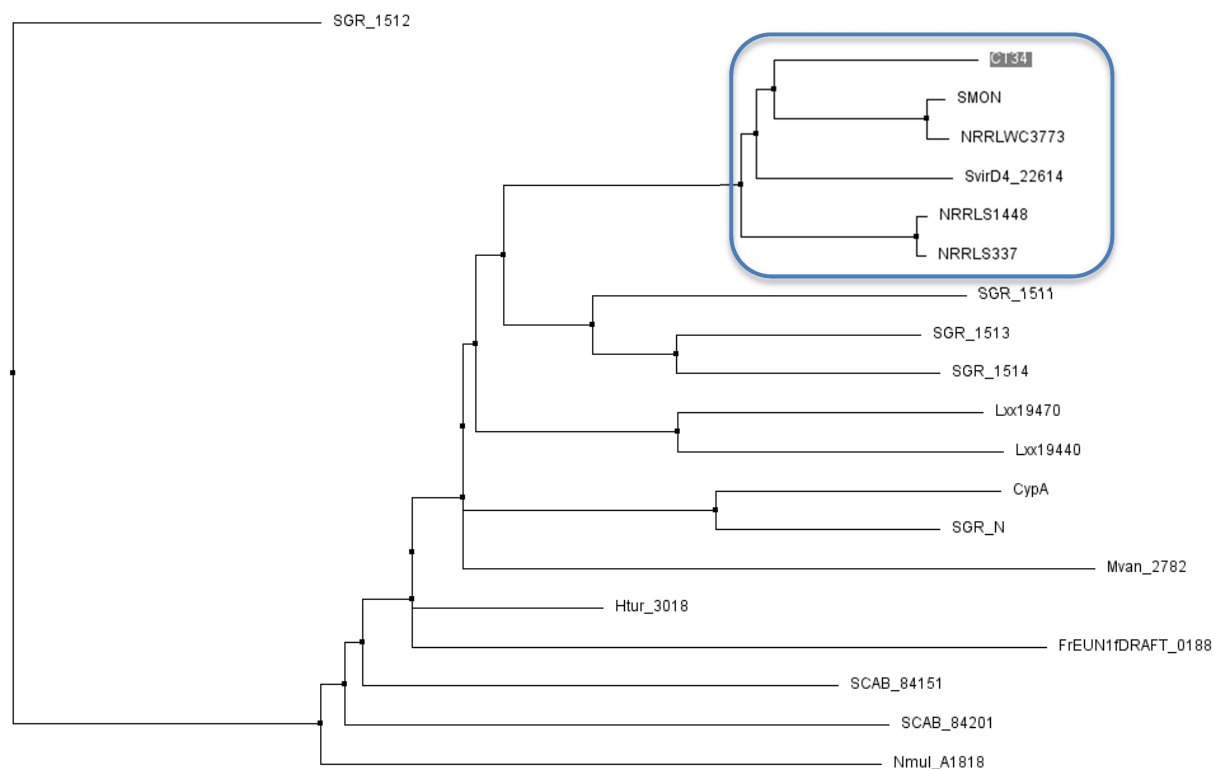


Figure S11. Phylogenetic analysis of prepeptide sequences from linaridin RiPP family, suggesting that legonaridin is a subgroup of the linaridin RiPP family. CT34 is the prepeptide of legonaridin; SvirD4_22614 is a hypothetical protein from *Streptomyces viridochromogenes* (WP_003991980.1); SMON is a hypothetical protein from *Streptomyces monomycini* (WP_030019228.1); NRRLwc3773 is from *Streptomyces sp.* NRRL WC-3773 (WP_031005572.1), NRRLS1448 is from *Streptomyces sp.* NNRL S-1448 (WP_030414969.1); NRRLS337 is from *Streptomyces sp.* S-337 (WP_030801611.1); CypA is the prepeptide of cypemycin; SGR_N is the prepeptide of SGR-1832 RiPP.

Table S1. Proposed functions of the encoded proteins of the identified *leg* gene cluster directing the biosynthesis of legonaridin **1** in *Streptomyces* sp. CT34.

ORF	proposed function	AA	Sequence identity	homologues	NCBI accession no
G	hypothetical protein	63	31%	LegA, no homolog found in NCBI database	WP_043265479.1
B	ABC transporter	617	77%	ABC transporter ATPase [<i>S. viridochromogenes</i>] (WP_003991974.1)	WP_043270371.1
C	FAD dependent oxidoreductase	495	89%	dehydrogenase [<i>Streptomyces</i> sp. TAA486]	WP_043270372.1
D	methyltransferase	252	30%	CypM	WP_043270380.1
E	Ab hydrolase fragment	278	31%	Alpha/beta(Ab) hydrolase fold domain of CypH	WP_043270373.1
F	hypothetical protein	195	37%	CypL	WP_043270374.1
A	Precursor peptide	65	46%	CypA	WP_043265481.1
H	HTT membrane protein fragment	259	40%	HTT membrane domain of CypH	WP_043265482.1

Table S2. Plasmids, strains, and primers used in this study

Plasmid/strains/primers	Description	Ref. or source
pYH7	shuttle vector used for gene knockout of <i>Streptomyces</i> and <i>E.coli</i>	Ref. 1,2
pUC-T simple	vector used for general cloning	
Gene-deletion constructs		
pWHU2261	pUC-T based construct with 1880 bp of upstream of the <i>leg</i> cluster	This work
pWHU2262	pUC-T based construct with 1990 bp of downstream of the <i>leg</i> cluster	This work
pWHU2263	pYH7 based construct	This work
<i>Escherichia coli</i>		
DH5 α	strain used for general cloning and plasmid maintenance	Stratagene
ET12567	dam, dcm, hsdM, hsdS, hsdR, cat, tet	
<i>Streptomyces</i> sp. CT34		
WT	Wild type strain; used to construct the strain CT34 mutants	This work
M-ZY	CT34 mutants	This work
Primer	Sequence	Site
Primers for pWHU2263 constructs		
upstream homologous arm 1 (U1)	<u>CATATG</u> CCCGCCAGGTAGTTGACGC	NdeI
upstream homologous arm 2 (U2)	<u>CTCGAG</u> CGCCCTTCCTCGGTTTCG	XhoI
downstream homologous arm 1 (D1)	<u>CTCGAG</u> AACTGGCCTCGGTCGACAGT	XhoI

downstream homologous arm 2
(D2)

AAGCCTCGTCATCGAGCCGAGGTACG

HindIII

Primers for verify pYH7

P 1 AGAGGTTCCACTCCAGCGAG

P 2 CCACGTAGACGTTGCGTGAG

Primers for confirming

L 1 CCGAAGTGGTGGACGGAGTG

L 2 GCCCGATCTCGTCGTAGTGG

R 1 TGCAGCAATGCGATTCACGG

R 2 CGAACTCAGCCAGCACAGAC

References:

- (1). Sun, Y.; Hong, H.; Samborskyy, M.; Mironenko, T.; Leadlay, P.F.; Haydock, S.F. *Microbiology*, **2006**, *152* 3507-3515.
- (2). Sun, Y.; Zhou, X.; Liu, J.; Bao, K.; Zhang, G.; Tu, G.; Kieser, T.; Deng, Z. *Microbiology*, **2002**, *148*: 361-371.
- (3). Zhai, Y.; Cheng, B.; Hu, J.; Kyremeh, K.; Wang, X.; Jaspars, M.; Deng, H.; Deng, Z-X.; Hong, K. *GenomeA*. **2015**, *3*, e01508-14.