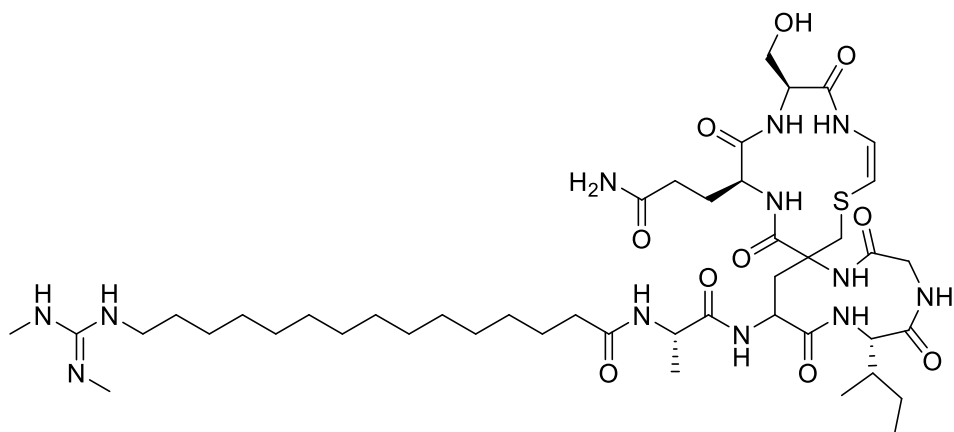
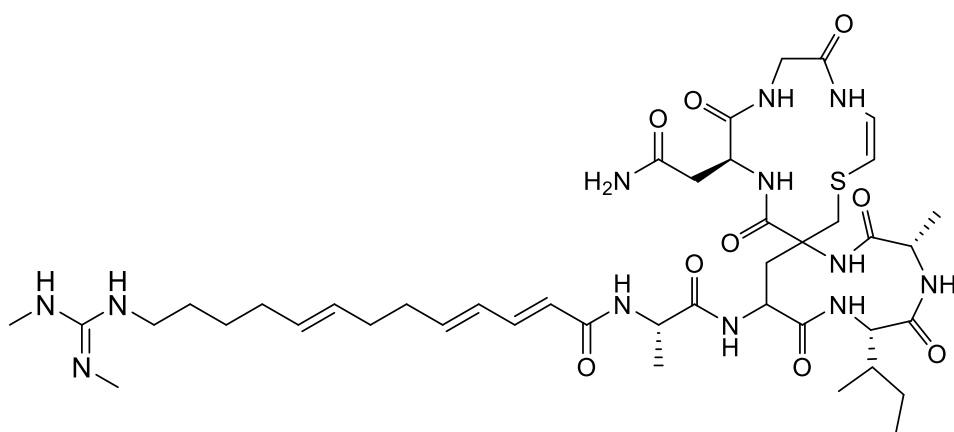
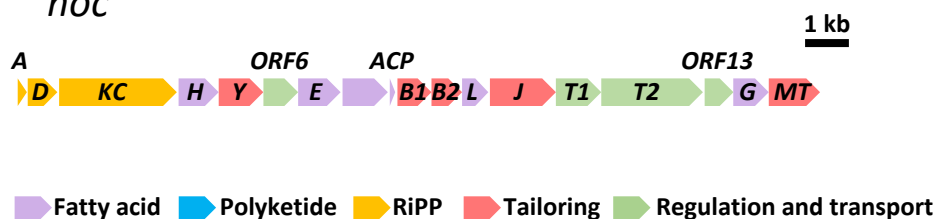


Acyltransferase that catalyses the condensation of polyketide and peptide moieties of goadvionin hybrid lipopeptides

Ryosuke Kozakai, Takuto Ono, Shotaro Hoshino, Hidenori Takahashi, Yohei Katsuyama, Yoshinori Sugai, Taro Ozaki, Kazuya Teramoto, Kanae Teramoto, Koichi Tanaka, Ikuro Abe, Shumpei Asamizu, and Hiroyasu Onaka

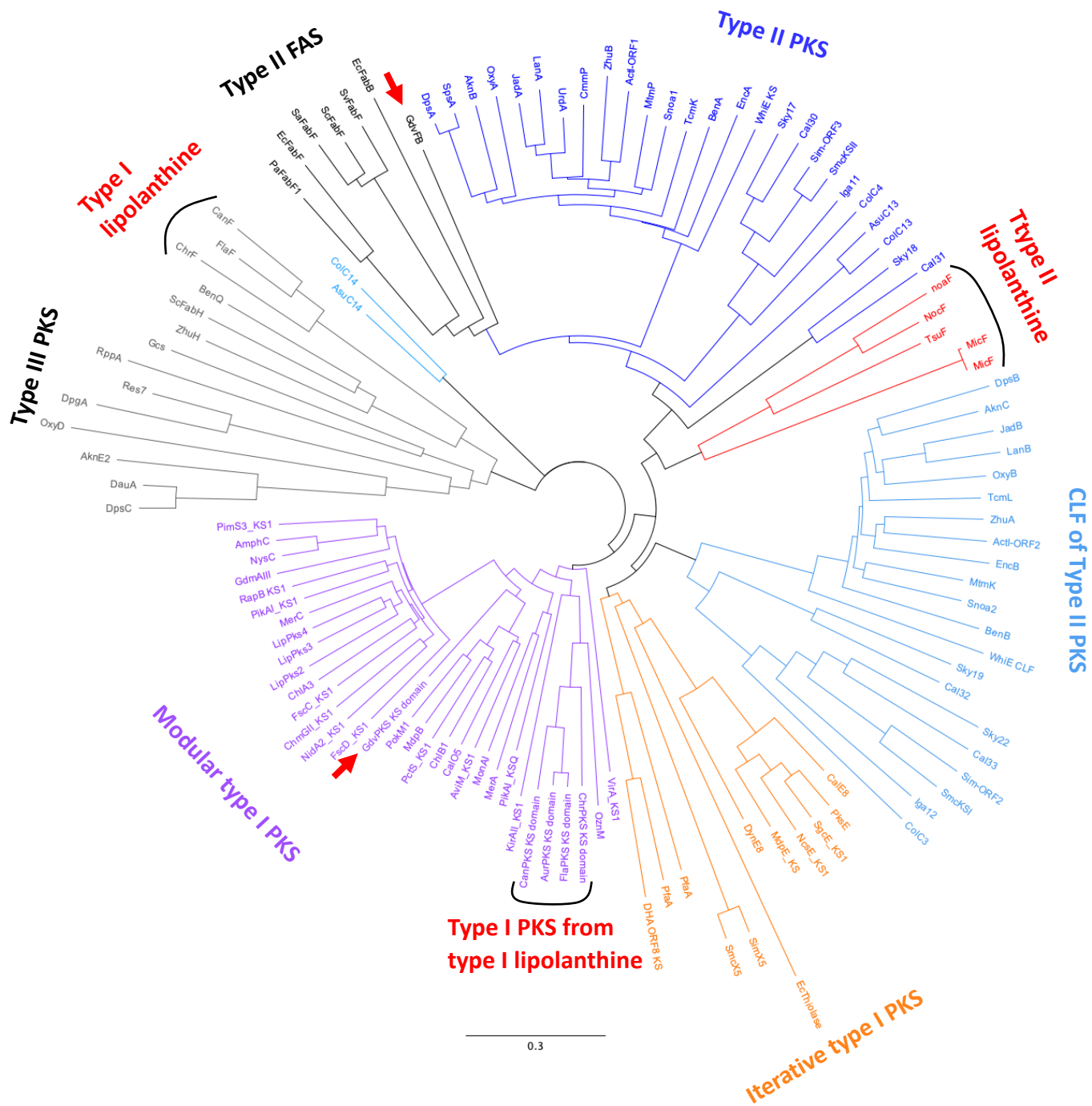
aMicrovionin isolated from *Microbacterium arborescens**mic***b**Nocavionin isolated from *Nocardia terpenica**noc***Supplementary Fig. 1. Chemical structures of lipolanthins.**

(a) Chemical structure of microvionin and gene organization of microvionin biosynthetic gene cluster (*mic*). (b) Chemical structure of nocavionin and nocavionin biosynthetic gene cluster (*noc*).

Supplementary Table 1. Goadvionin biosynthetic gene cluster.

Deduced genes and their proposed functions in *gdv* cluster

Gene	a.a.	Proposed function
<i>gdvC</i>	309	Acyl-ACP desaturase
<i>gdvD</i>	233	Flavoprotein (LanD)
<i>gdvACP</i>	84	Acyl Carrier Protein (ACP)
<i>gdvR</i>	222	LuxR type transcriptional regulator
<i>gdvE</i>	172	Hypothetical
<i>gdvKC</i>	890	Class III lanthipeptide synthetase(LanKC)
<i>gdvA</i>	29	Precursor peptide (goadvionin A)
<i>gdvB</i>	29	Precursor peptide (goadvionin B)
<i>gdvI</i>	535	Fatty acyl-CoA ligase
<i>gdvFB</i>	421	3-Oxoacyl-ACP synthase (KAS I/II, FabB/F)
<i>gdvG</i>	221	N-acetyltransferase
<i>gdvL</i>	595	fatty acyl-AMP ligase (FAAL)
<i>gdvPKS</i>	1902	Type I polyketide synthase (ACP-KS-AT-DH-KR-ACP)
<i>gdvMT</i>	152	Lysine-methyltransferase
<i>gdvFG</i>	234	Ketoreductase (KR, FabG)
<i>gdvFI</i>	252	Enoyl-ACP reductase (ER, FabI)
<i>gdvT1</i>	335	ABC transporter
<i>gdvT2</i>	265	ABC transporter
<i>gdvFZ</i>	149	Acyl-ACP dehydrogenase (DH, FabZ)
<i>gdvH</i>	373	Dioxygenase
<i>gdvJ</i>	561	Acetolactate synthase
<i>gdvY</i>	426	aldehyde dehydrogenase



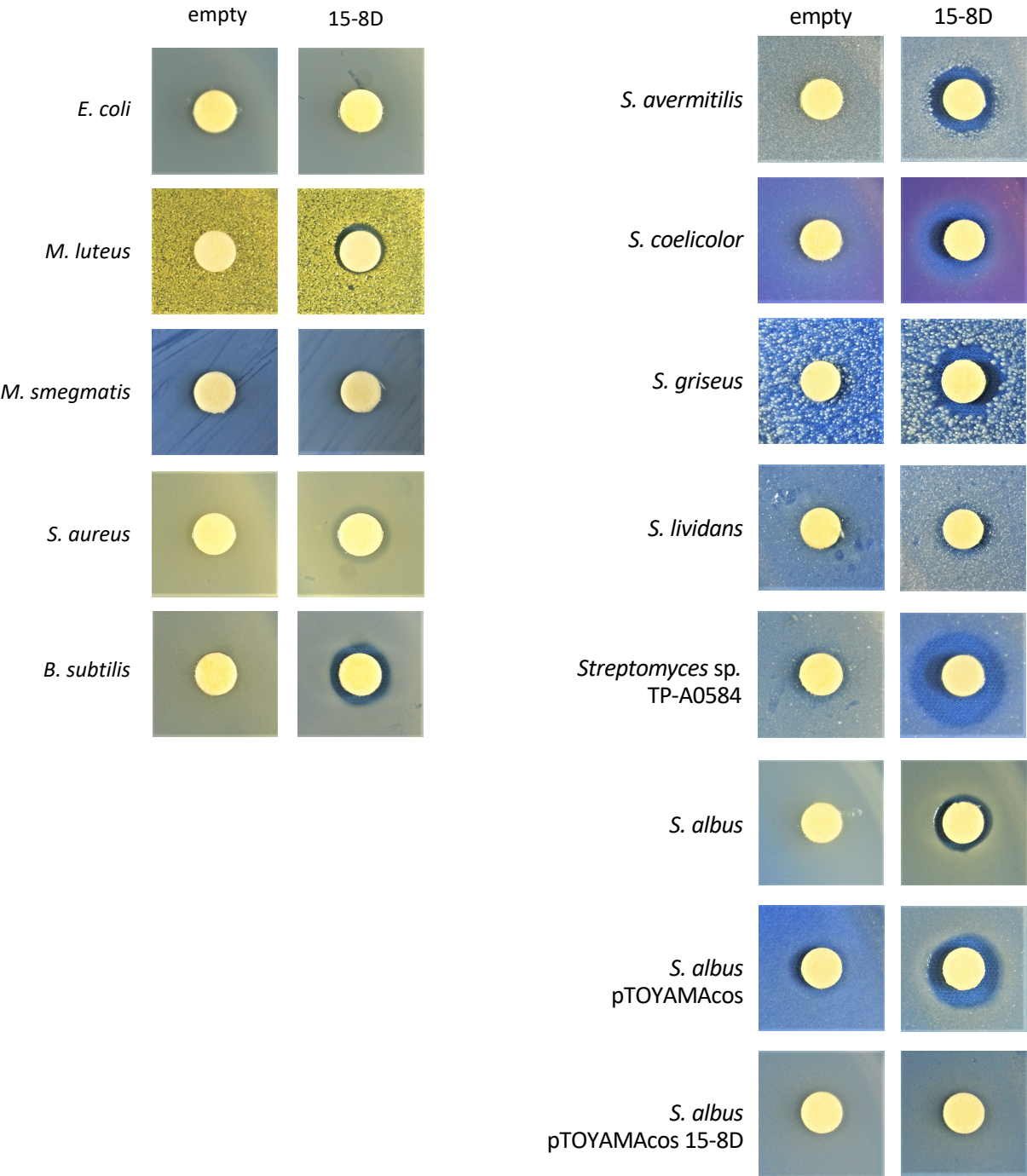
Supplementary Fig. 2. Phylogenetic analysis of various KSs.
The red arrows indicate GdvFB and the KS domain of GdvPKS.

Supplementary Table 2. Calculated and observed mass of compounds (1-10)

	chemical formula	Calculated mass (Da)	Observed mass (Da)
Goadvionin A1 (1)	$C_{63}H_{110}N_{11}O_{17}S^+$	1324.7845	1324.7796
Goadvionin A2 (2)	$C_{63}H_{110}N_{11}O_{16}S^+$	1308.7880	1308.7847
Goadvionin A3 (3)	$C_{63}H_{112}N_{11}O_{17}S^+$	1326.7911	1326.7953
Goadvionin A4 (4)	$C_{63}H_{112}N_{11}O_{16}S^+$	1310.5995	1310.8004
Goadvionin B1 (5)	$C_{65}H_{114}N_{11}O_{17}S^+$	1352.8144	1352.8109
Goadvionin B2 (6)	$C_{65}H_{114}N_{11}O_{16}S^+$	1336.8205	1336.816
Goadvionin B3 (7)	$C_{65}H_{116}N_{11}O_{17}S^+$	1354.8208	1354.8266
Goadvionin B4 (8)	$C_{65}H_{116}N_{11}O_{16}S^+$	1338.8259	1338.8317
Goadpeptin A (9)	$C_{28}H_{43}N_{10}O_9S^+$	695.2946	695.293
Goadpeptin B (10)	$C_{30}H_{47}N_{10}O_9S^+$	723.3261	723.3243

Supplementary table 3. Microorganisms for bioactivity test

Designation	Relevant characteristics
<i>Streptomyces</i> sp. TP-A0584	Gram(+), goadvionin producer
<i>Streptomyces avermitilis</i> MA-4680	Gram(+)
<i>Streptomyces lividans</i> TK23	Gram(+)
<i>Streptomyces coelicolor</i> A3(2)	Gram(+)
<i>Streptomyces albus</i> J1074	Gram(+)
<i>Streptomyces griseus</i> IFO13350	Gram(+)
<i>Bacillus subtilis</i> ATCC6633	Gram(+)
<i>Micrococcus luteus</i>	Gram(+)
<i>Mycobacterium smegmatis</i>	Gram(+)
<i>Staphylococcus aureus</i> 209P JC-1	Gram(+)
<i>Escherichia coli</i> NIHJ JC-2	Gram(-)



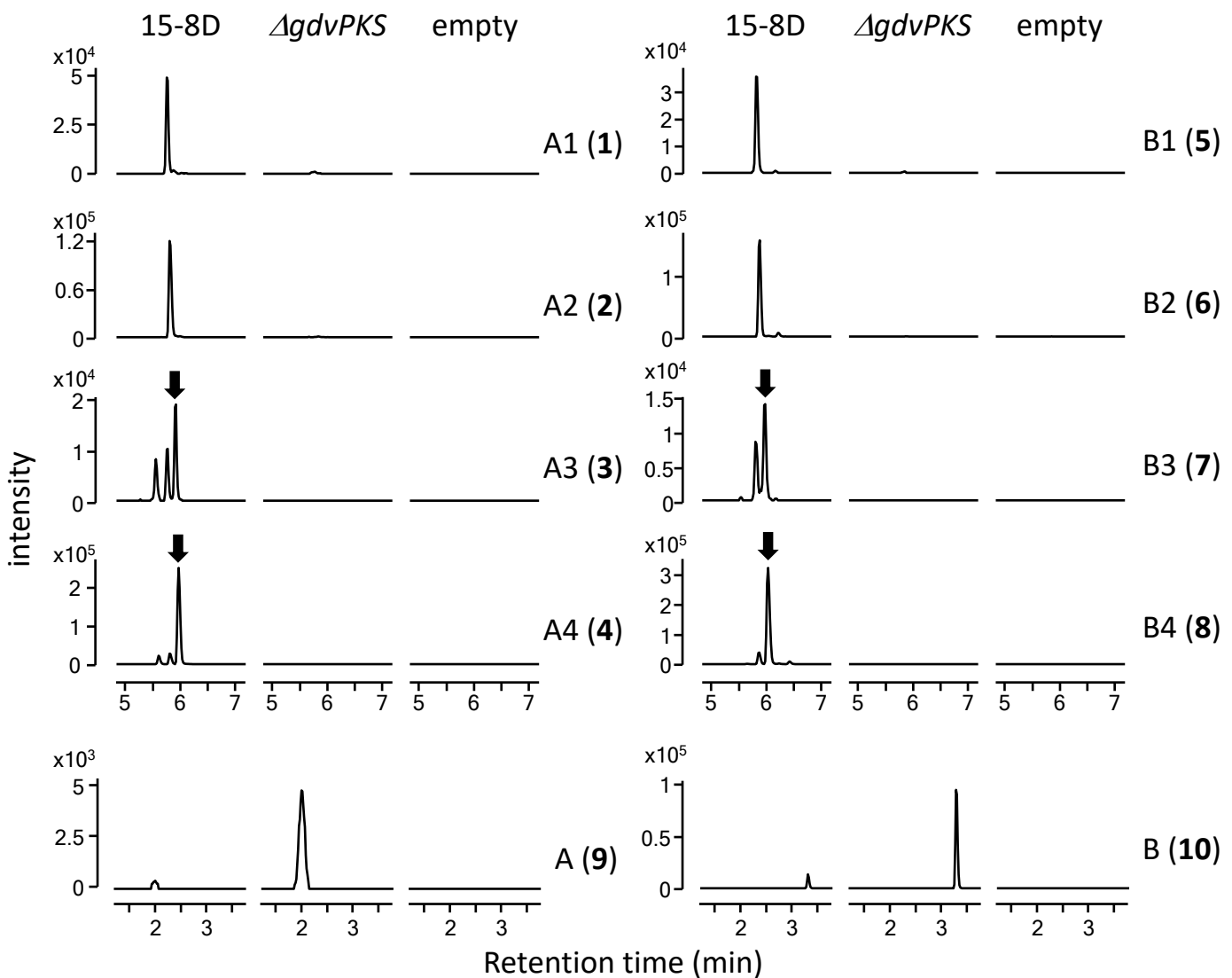
Supplementary Fig. 3. Antimicrobial activity of goadvionins.

Paper disc assay of the extract from *S. albus* pTOYAMAcos 15-8D which is a goadvionin producing strain (15-8D) and *S. albus* pTOYAMAcos which is a goadvionin non-producing strain (empty). The indicator strains for assay is listed in supplementary table 3.

Supplementary table 4. MIC of goadvionin B2.

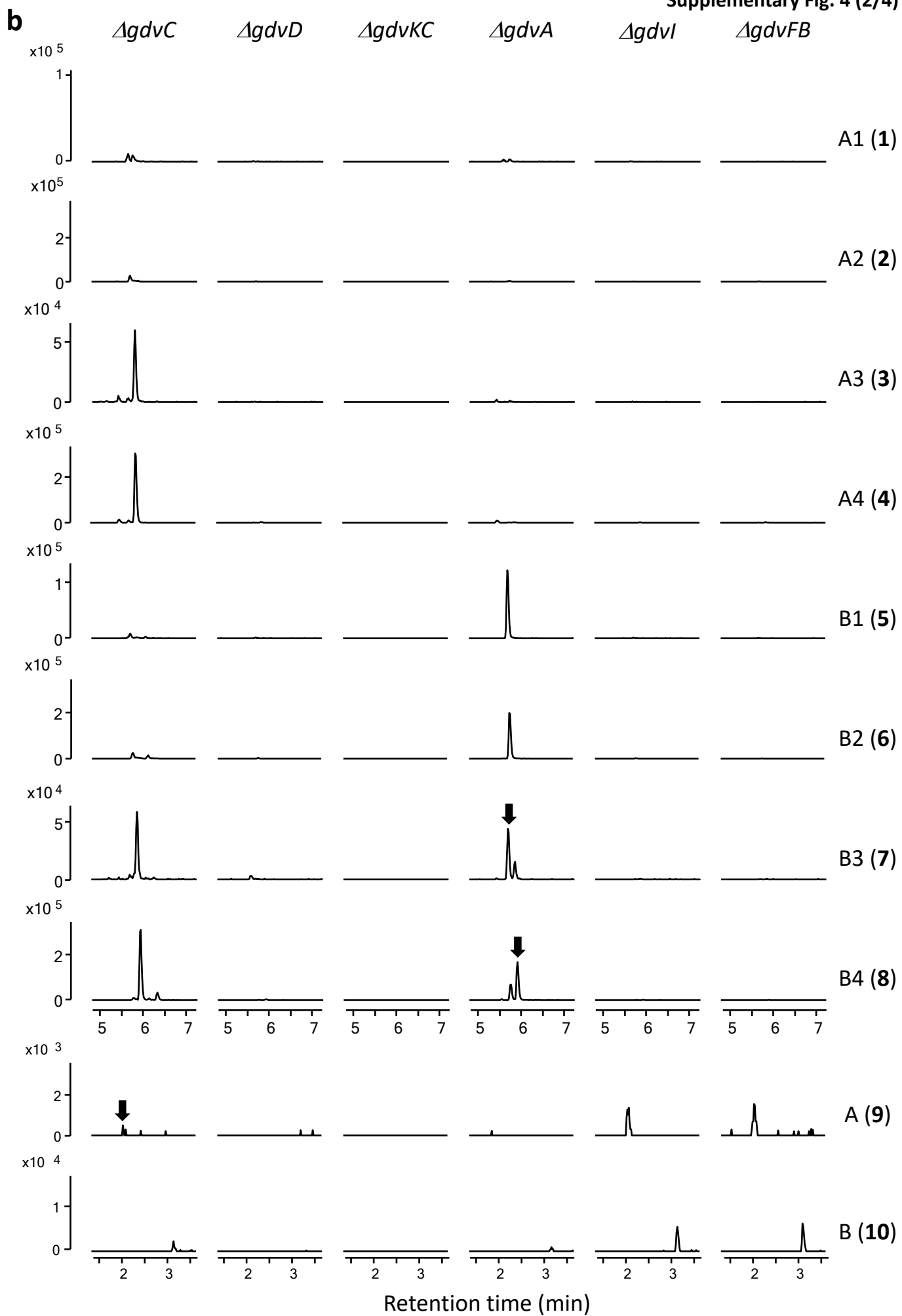
	MIC of goadvionin B2 (µg / mL)
<i>Staphylococcus aureus</i>	6.4
<i>Bacillus subtilis</i>	3.2
<i>Micrococcus luteus</i>	3.2

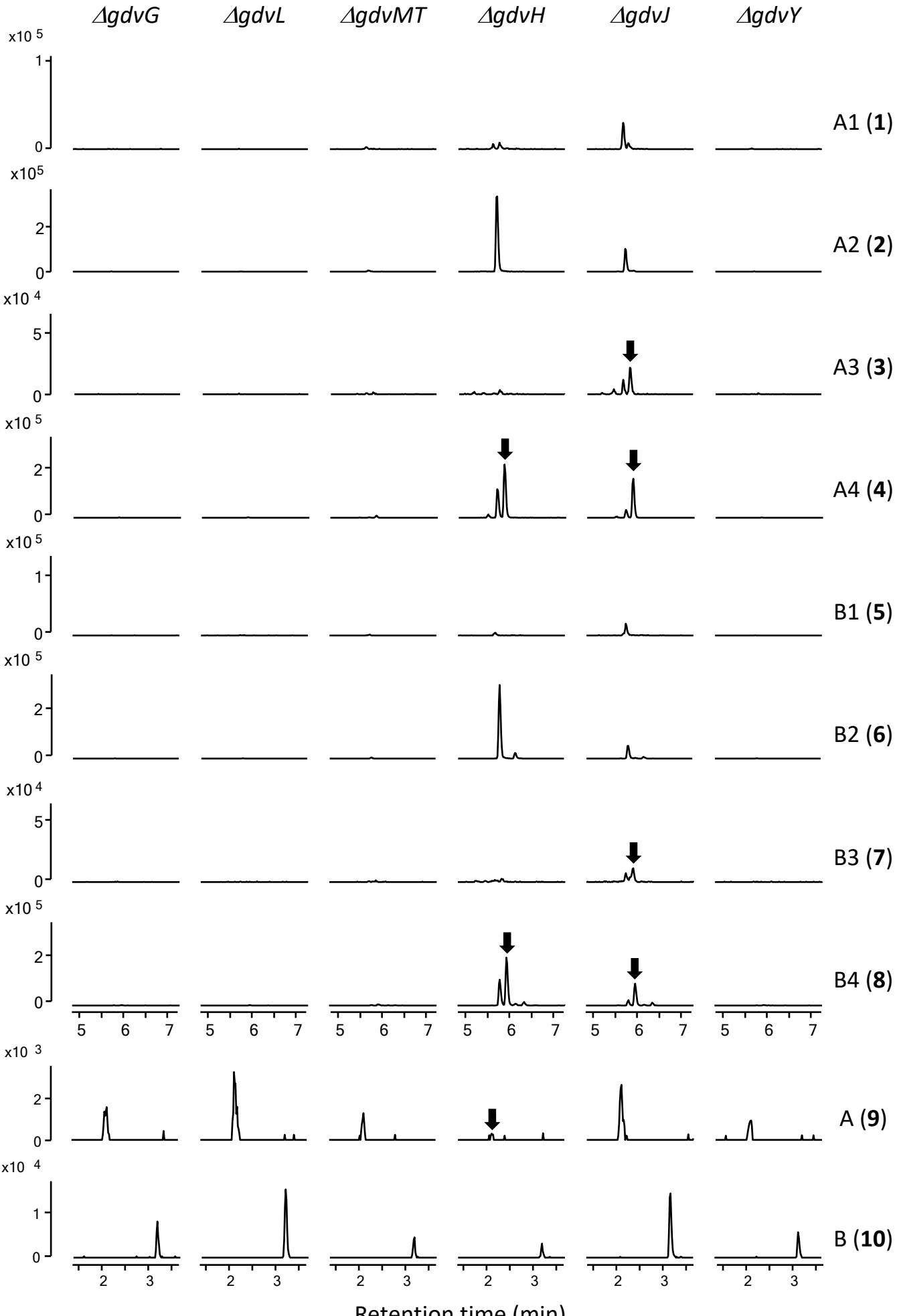
Goadvionin B2 bioactivityies against *Staphylococcus aureus*, *Bacillus subtilis* and *Micrococcus luteus* by 2-fold broth dilution method.

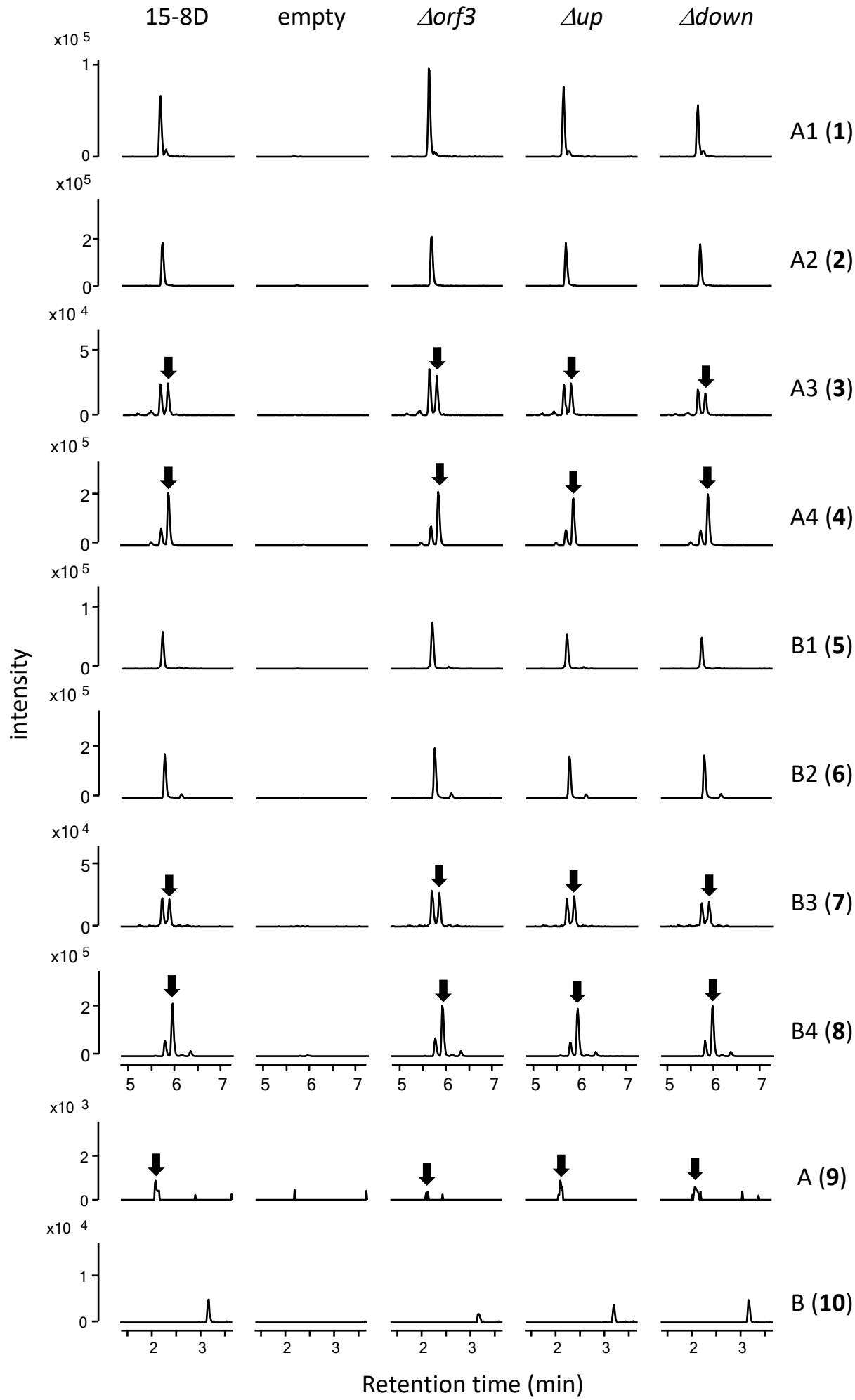
a

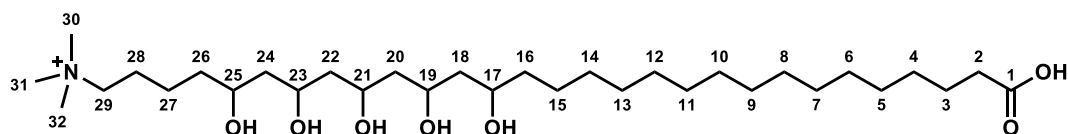
Supplementary Fig. 4. Productivity of goadvionins (1-8) and goadpeptides (9 and 10) in disruptants and heterologous expression strains.

(a) *S. lividans* harboring whole *gdv* cluster without *gdvPKS* ($\Delta gdvPKS$), and pTOYAMAcos 15-8D (15-8D), and pTOYAMAcos (empty). (b) *S. albus* harboring whole *gdv* cluster (pTOYAMAcos 15-8D), whole *gdv* cluster without *gdvC* ($\Delta gdvC$), *gdvD* ($\Delta gdvD$), *gdvKC* ($\Delta gdvKC$), *gdvA* ($\Delta gdvA$), *gdvI* ($\Delta gdvI$), *gdvFB* ($\Delta gdvFB$), *gdvG* ($\Delta gdvG$), *gdvL* ($\Delta gdvL$), *gdvMT* ($\Delta gdvMT$), *gdvH* ($\Delta gdvH$), *gdvJ* ($\Delta gdvJ$), *gdvY* ($\Delta gdvY$), *orf3* ($\Delta orf3$), *orf2* (Δup), and *orf26* ($\Delta down$), and *S. albus* harboring pTOYAMAcos 15-8D (15-8D), and harboring pTOYAMAcos (empty). EIC at 10 compounds, Goadvionin A1 (1324.7796 [M]⁺), Goadvionin A2 (1308.7847 [M]⁺), Goadvionin A3 (1326.7953 [M]⁺), Goadvionin A4 (1310.8004 [M]⁺), Goadpeptide A (695.2930 [M+H]⁺), Goadvionin B1 (1352.8109 [M]⁺), Goadvionin B2 (1336.8160 [M]⁺), Goadvionin B3 (1453.8266 [M]⁺), Goadvionin B4 (1338.8317 [M]⁺) and Goadpeptide B (723.3243 [M+H]⁺), were detected by LC-qTOFMS.

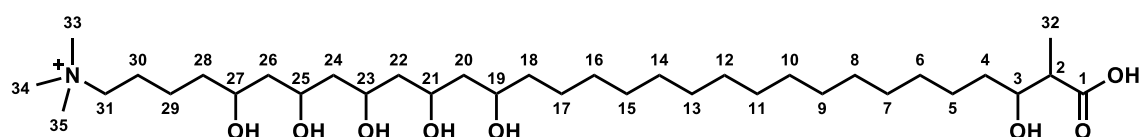




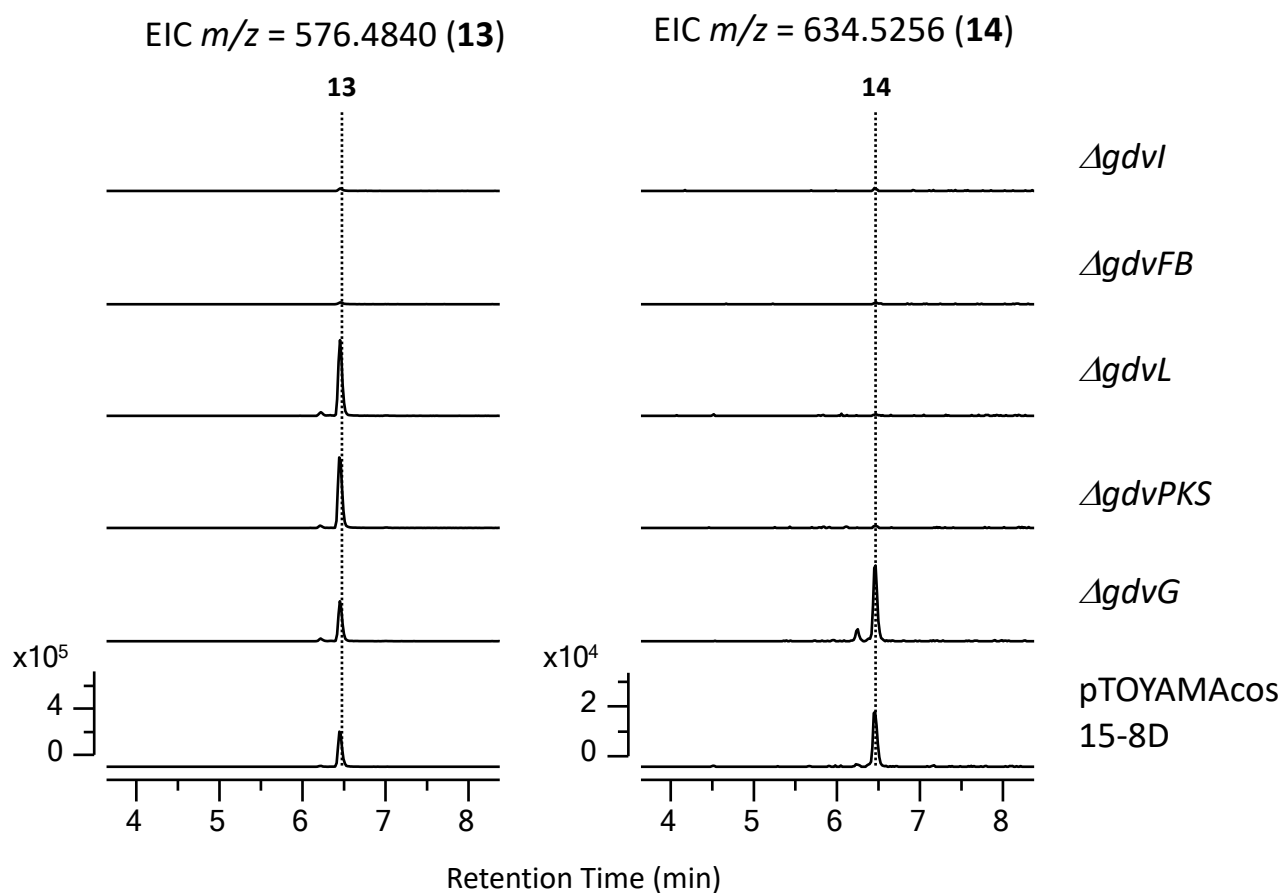


a

Compound **13** ($C_{32}H_{66}NO_7^+$)
Calculated mass: 576.4834 Da

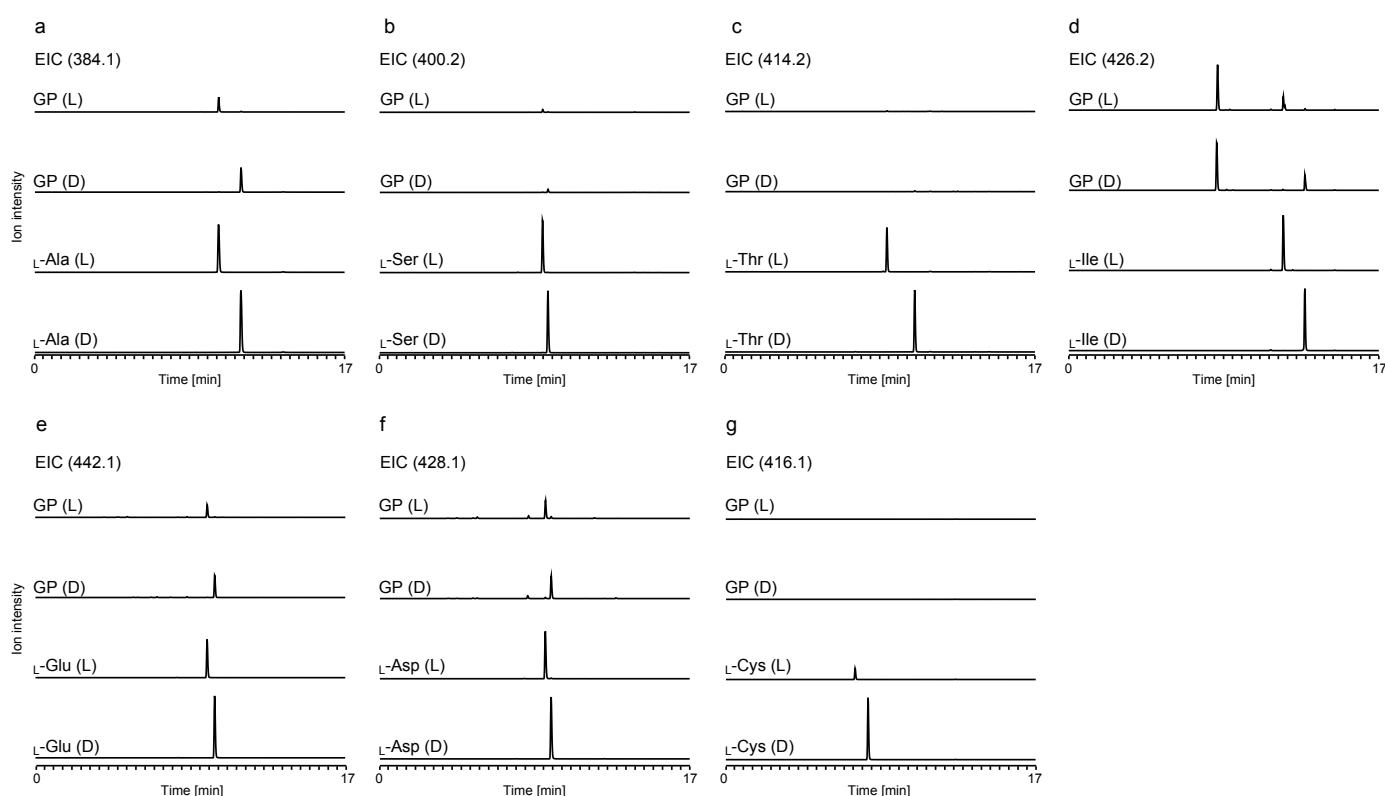


Compound **14** ($C_{35}H_{72}NO_8^+$)
Calculated mass: 634.5252 Da

b

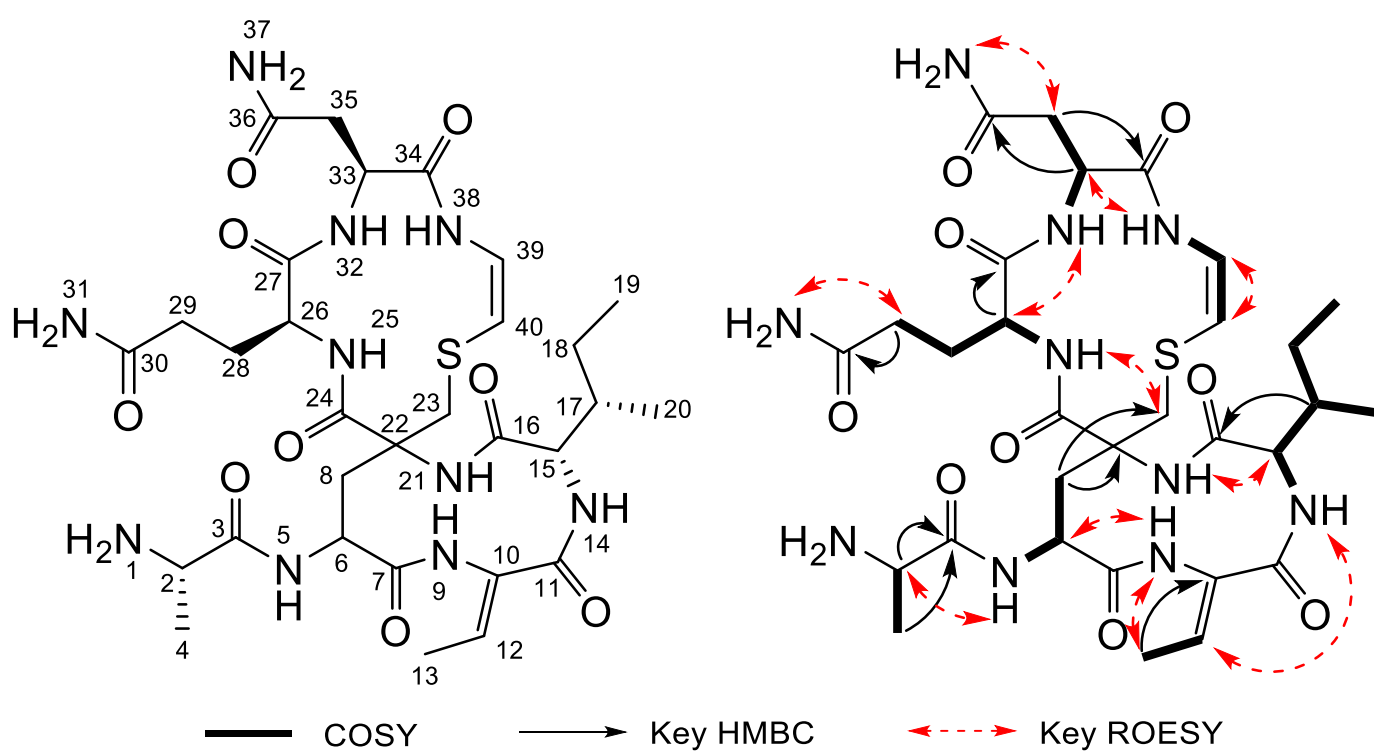
Supplementary Fig.5 Accumulation of acyl chain intermediates.

(a) Proposed structure of fatty chain intermediates. (b) EIC at m/z corresponding to the proposed fatty chain intermediates. The culture extracts of *S. albus* $\Delta gdvI$, $\Delta gdvFB$, $\Delta gdvL$, $\Delta gdvPKS$, $\Delta gdvG$, and pTOYAMAcos15-8D were analyzed.



Supplementary Fig. 6. Amino acid analysis of goadpeptin B.

Goadpeptin B was hydrolyzed under acidic condition, followed by derivatized by L- or D-FDLA. The derivatized samples were analyzed by LC-MS. Figures show comparison of goadpeptin sample and authentic L-Ala (a), L-Ser (b), L-Thr (c), L-Ile (d), L-Glu (e), L-Asp (f), and L-Cys (g). Letters in parenthesis show chiralities of FDLA used for the derivatization.



Supplementary Fig. 7. 2D NMR analyses of goadpeptin B (10)

NMR analysis of goadpeptins (9) and (10)

Consistent with the result of amino acid analysis (Supplementary Fig. 6), 1D and 2D NMR spectra of **10** clearly indicated the presence of four proteogenic amino acid residues (Ala1, Ile4, Gln6, and Asn7), along with the (Z)-dehydrobutyrine unit (Dhb3) derived from Thr residue (Supplementary Fig. 7, Fig. 34-39 and Table 10). The double bond geometry of Dhb moiety was established as Z based on the ROESY correlation between methyl group (H₃-13) and amide proton (NH-9). In addition, ¹H, ¹³C NMR, and HSQC spectra of **10** also displayed the series of signals corresponding to avionin moiety (Avi2/Avi5/AviCys8), including two carbonyls (C-7, C-24), three amide protons (NH-5, NH-21, and NH-38), two vinylic methine (C-39 and C-40), two aliphatic methylenes (C-8 and C-23), one aliphatic methine (C-6), and one quaternary carbon (C-22). The residue of Avi2 and aminovinylthiol (AviCys8) moieties in **10** were constructed based on the COSY correlations of NH-5/H-6/H-8 and NH-38/H-39/H-40, and the *cis*-configuration of AviCys residue was determined based on the coupling constant between H-39 and H-40 (*J* = 6.5 Hz). Furthermore, the complete avionin scaffold (Avi2/Avi5/AviCys8) was confirmed by the HMBC correlations of H-8/C-22 and H-8/C-23 and the molecular formula of **10** (C₃₀H₄₆N₁₀O₉S). Finally, the sequence of each amino acid residue was confirmed based on the ROESY correlations of each amide proton (Supplementary Fig. 7 and 39).

Based on the molecular formula (C₂₈H₄₂N₁₀O₉S) and the amino acid sequence of GdvA, goadpeptin A (**9**) was expected to be an analogue of goadpeptin B, whose Ile4 and Gln6 residues are substituted with Val4 and Asn6, respectively.

NMR analysis of goadvionin A1 (1) and B2 (6)

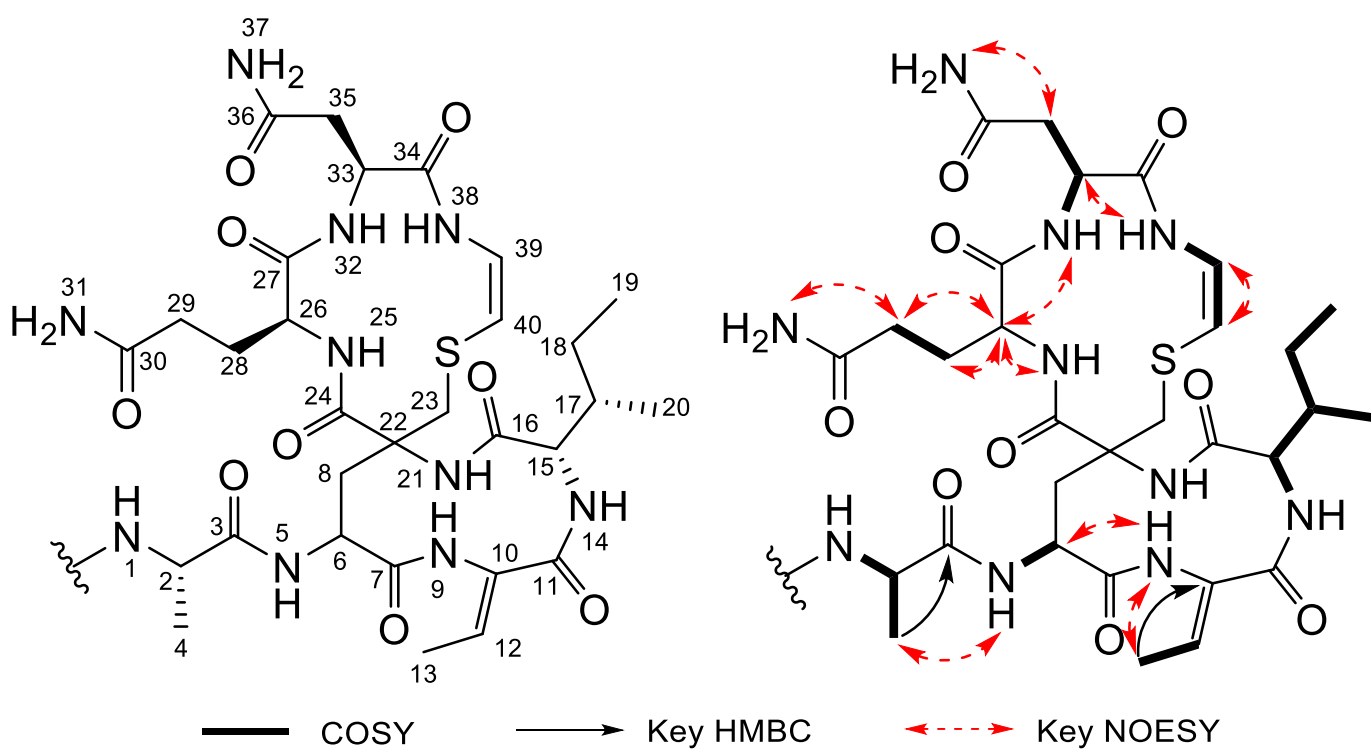
First, we performed the structural analysis of goadvionin B2 (**6**, C₆₅H₁₁₄N₁₁O₁₆S). Notably, **6** consists of an odd number of nitrogen atoms and even number of hydrogen atoms, strongly suggesting that **6** contains a positively-charged factional group such as quaternary ammonium salt. Moreover, LR-ESI-MS/MS spectrum of **6** showed two major fragment ions of *m/z* 652.34 and 685.65 generated from the molecular ion (*m/z* 668.9 [M+H]²⁺, Supplementary Fig. 32), suggesting that **6** consists of a RiPP unit derived from goadpeptin B and long fatty acyl group with positively-charged functional group.

Consistent with the result of ESI-MS/MS analysis, 1D and 2D NMR spectra of **6** also supported the presence of core peptide derived from goadpeptin B (Supplementary Figs. 8, Figs.40-47 and Tables 11 and 12). Furthermore, it was suggested from the downfield shift of H-2 signal (3.45 ppm in **10**/ 4.45 ppm in **6**) along with the appearance of amide proton (NH-1, 7.90 ppm) in **6** that N-terminal amino group of alanine (Ala1) residue is modified with fatty acyl group through amide bond. In addition to the RiPP moiety, ¹H and ¹³C NMR spectra of **6** also displayed the series of signals corresponding to trimethylamino fatty acyl group (Supplementary Table 12), and the detailed 1D and 2D NMR analysis unveiled the presence of four substructures (substructures A-D in Supplementary Fig. 9). Initially, the substructure A (C-1' to C-4' and C-32') was established by HMBC correlation of H₃-32'/C-1' and COSY correlations of H-2'/H-3'/H-4', H-3'/3'-OH, and H-2'/H₃-32'. Furthermore, the amide bond between C-1' and Ala1 NH group was confirmed by the NOESY correlations of NH-1/H-2' and NH-1/H₃-32'. The isolated olefinic bond (substructure B, C-10' to C-13') was inferred from the characteristic ¹H NMR signal at 5.33 ppm (H-11' and 12'), which showed COSY correlation to aliphatic methylene signal at 1.99 ppm (H-10' and H-13'), and HSQC correlation to 130.09 and 130.13 ppm. The geometry of the isolated olefin system was determined to be *Z*, based on the ¹³C NMR chemical shifts of allylic carbons (Supplementary Fig. 10, C-10' and 13' at 27.1 ppm)¹. In addition to the signal of 3'-OH, ¹H NMR also displayed five exchangeable protons (4.46-4.66 ppm) showing clear COSY correlations to methine proton signals from 3.58 to 3.81 ppm, indicating the presence of five additional secondary alcohols (substructure C, C-19', C-21', C-23', C-25' and C-27'). Finally, the terminal part of fatty acid moiety (substructure D, C-29' to C-31' and C-33') was constructed by HMBC correlation of H₃-33'/C-31' along with COSY correlation of H-31'/H-30'/H-29'. The presence of trimethylammonium group was supported by the positively charged nature of **6**, along with characteristic carbon chemical shifts of C-31' (65.9 ppm) and C-33' (52.7 ppm). ¹H and ¹³C NMR spectrum of **6** also displayed large numbers of signals corresponding to aliphatic methylenes derived from fatty acyl moiety, but the severe signal overlaps hampered further NMR analyses of **6** (Supplementary Figs. 41 and 42).

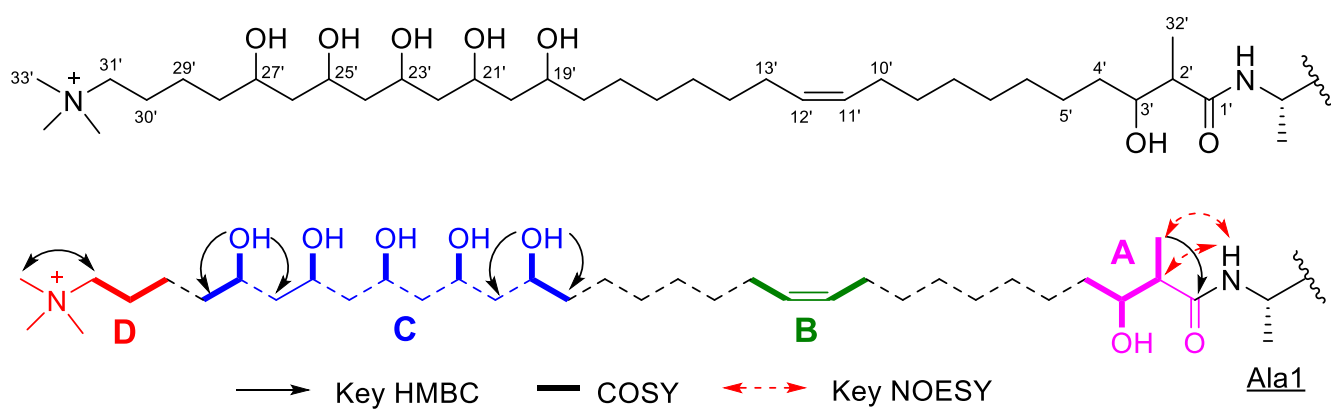
The molecular formula of **1** (C₆₃H₁₁₀N₁₁O₁₇S) strongly suggested that **1** also contains trimethylamino fatty acyl group and becomes positively charged. 1D NMR spectra of **1** were quite similar to those of goadvionin B2 (**6**), and the RiPP moiety of **1** was established by a combination of 1D and 2D NMR analyses (Supplementary Fig. 11 and Supplementary Figs. 48-53 and Supplementary tables 13 and 14). Similar to that of **6**, the RiPP moiety of **1** was derived from goadpeptin A (**9**), and the downfield shift of H-2 (4.45 ppm) along with appearance of an amide proton (NH-1, 7.90 ppm) indicated that the N-terminal alanine residue is modified with the fatty acyl group. Compared with the NMR spectra of **6**, the fatty acyl groups of **1** and **6** were likely to be similar to each other; in fact, detailed NMR analyses revealed that the substructures A-C of the fatty acyl group in **1** were completely identical to those in **6** (Supplementary Fig. 12). On the other hand, the ¹H NMR spectrum of **1** displayed an additional exchangeable proton signal at 5.28 ppm (Supplementary Fig. 13), which showed COSY correlation to the aliphatic methine at 4.02 ppm. Therefore, **1** should contain the seventh hydroxyl group in its fatty acyl group, and we identified its position as C-30' based on the HMBC and COSY spectra (substructure D in Supplementary Fig. 12).

Reference

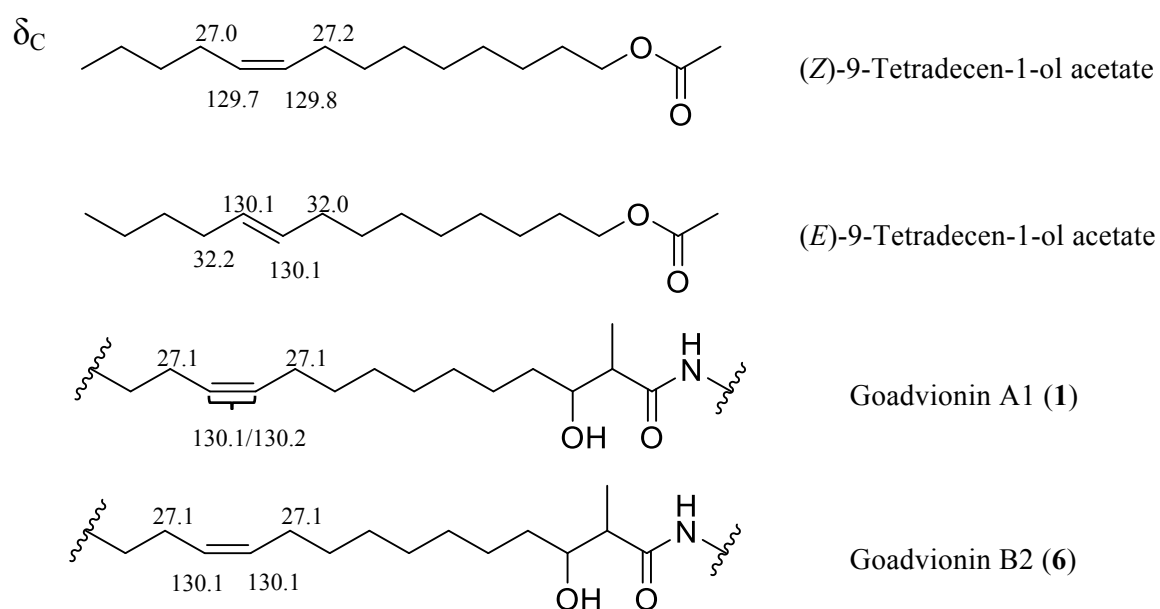
1. Rossi, R., Carpita, A., Quirici, M.G. & Veracini, C.A. Insect Pheromone Components - Use of C-13 Nmr-Spectroscopy for Assigning the Configuration of C=C Double-Bonds of Monoenic or Dienic Pheromone Components and for Quantitative-Determination of *Z/E* Mixtures. *Tetrahedron* **38**, 639-644 (1982).



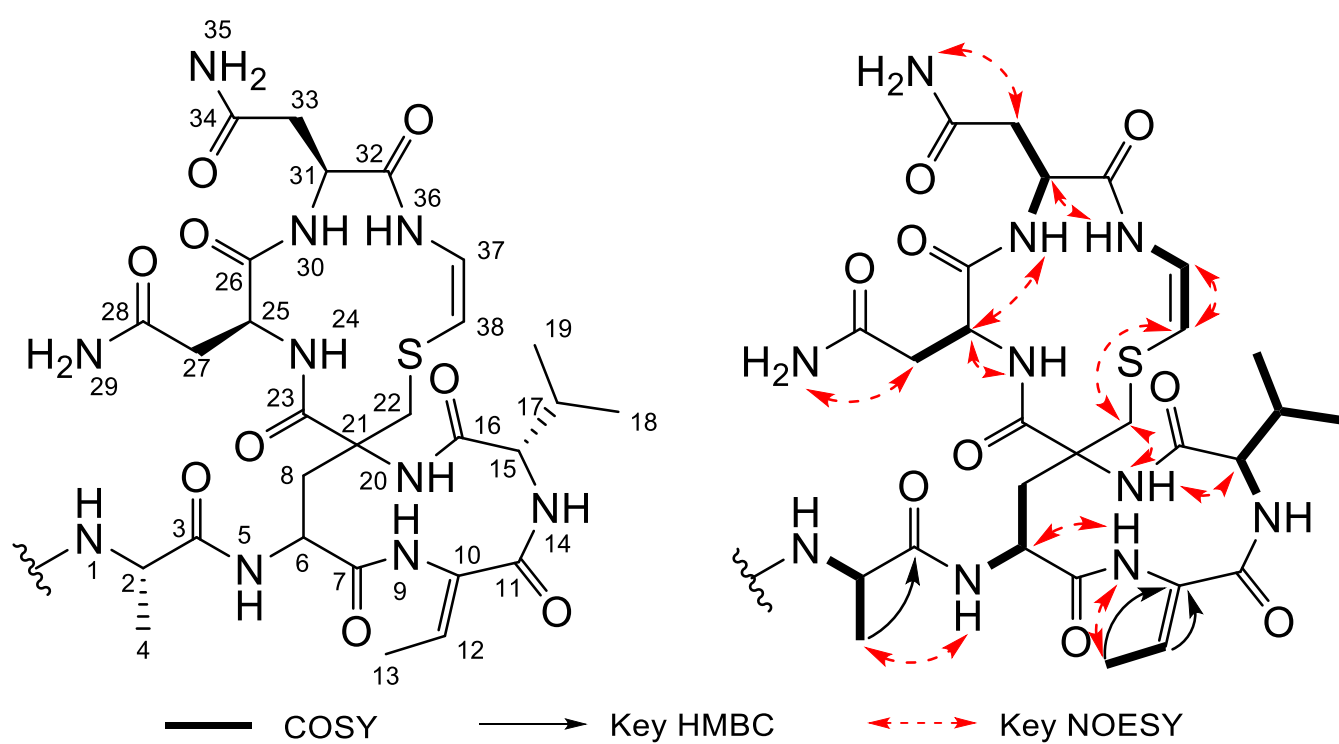
Supplementary Fig. 8. 2D NMR analyses of goadvionin B2 (6, RiPP moiety)



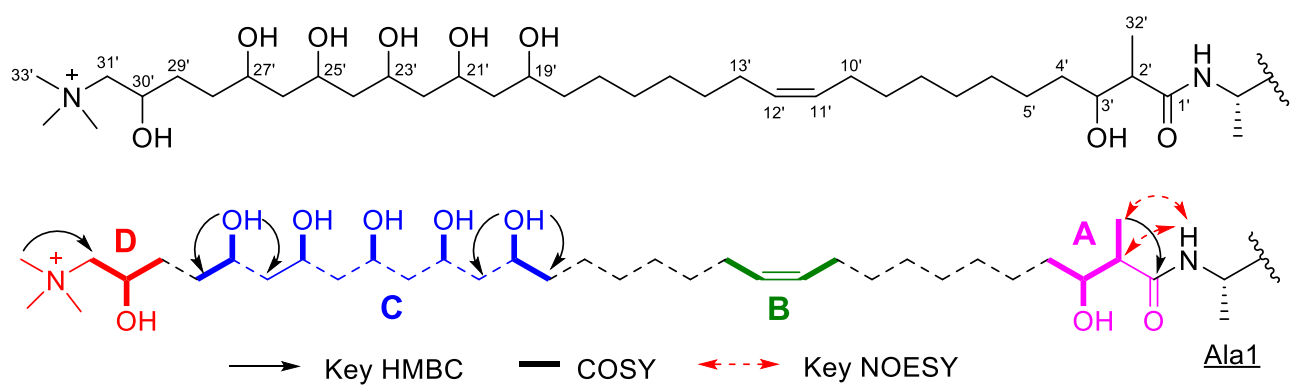
Supplementary Fig. 9. 2D NMR analyses of goadvionin B2 (6, VLCFA moiety)



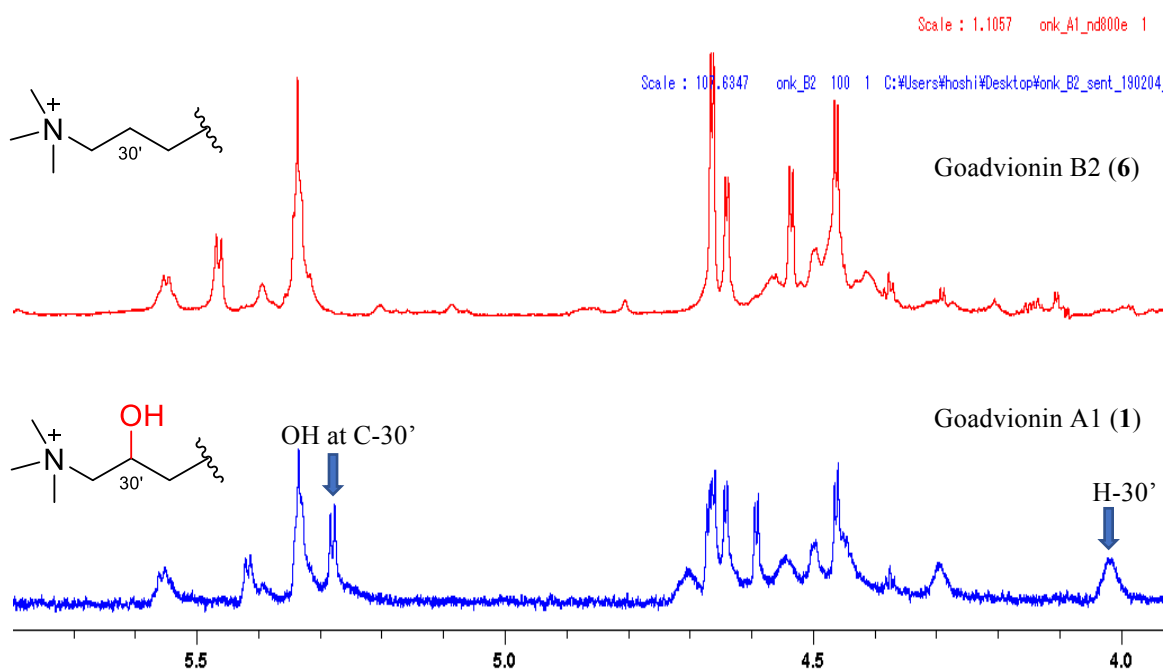
Supplementary Fig. 10. Elucidating the geometries of isolated olefinic bonds in 1 and 6 based on the chemical shifts of allylic carbon.



Supplementary Fig. 11. 2D NMR analyses of goadvionin A1 (1, RiPP moiety)

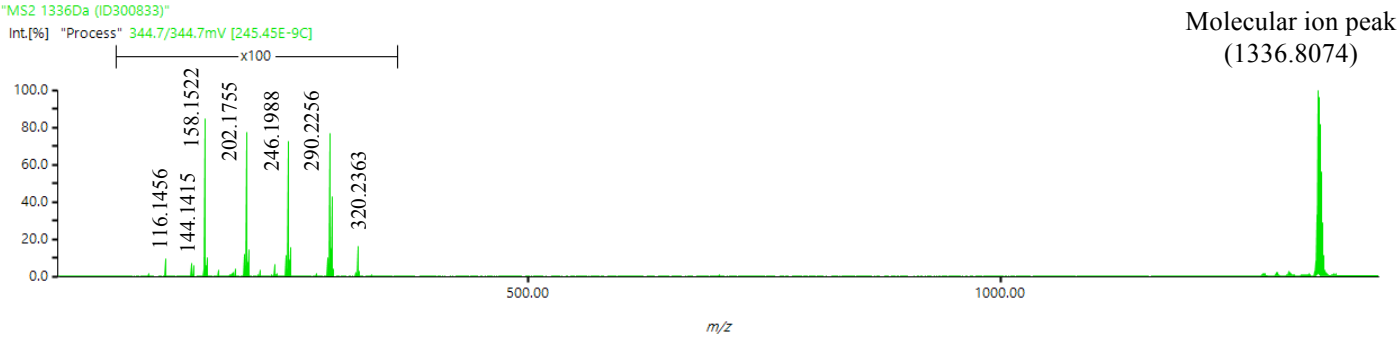
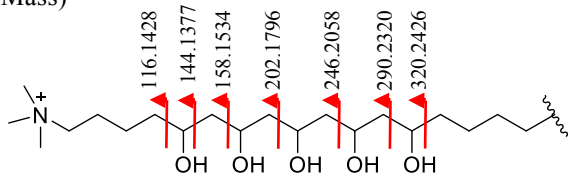


Supplementary Fig. 12. 2D NMR analyses of goadvionin A1 (1, VLCFA moiety)



Supplementary Fig. 13. Comparison of ^1H NMR spectra between 1 and 6 (4.0-5.8 ppm, 800 MHz in $\text{DMSO-}d_6$)

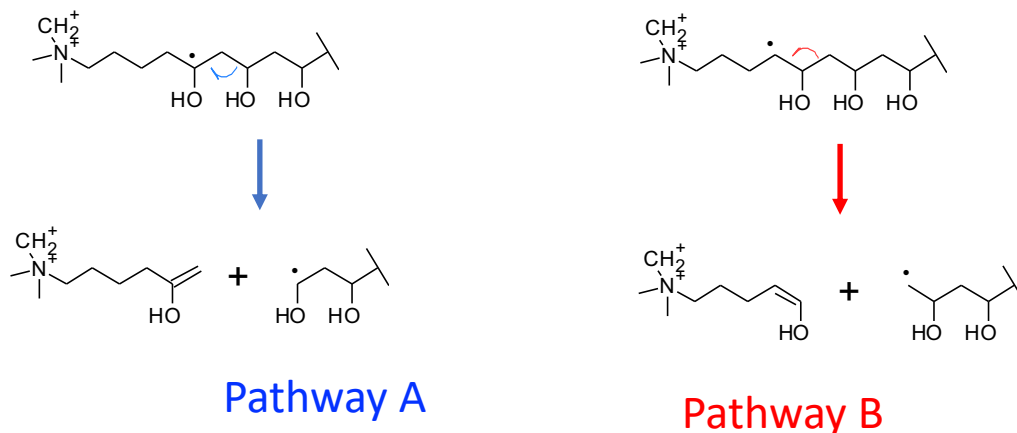
(Calculated Mass)



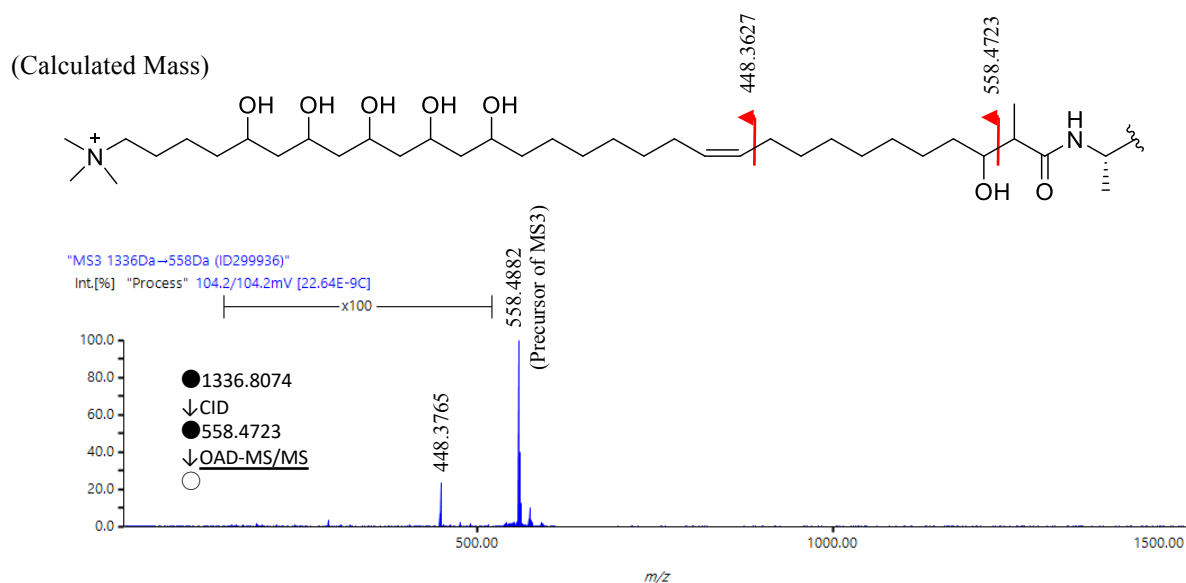
Supplementary Fig. 14. HAD-MS/MS spectrum of goadvionin B2 (6)
Supplemental collisional activation energy = 0 eV.

a

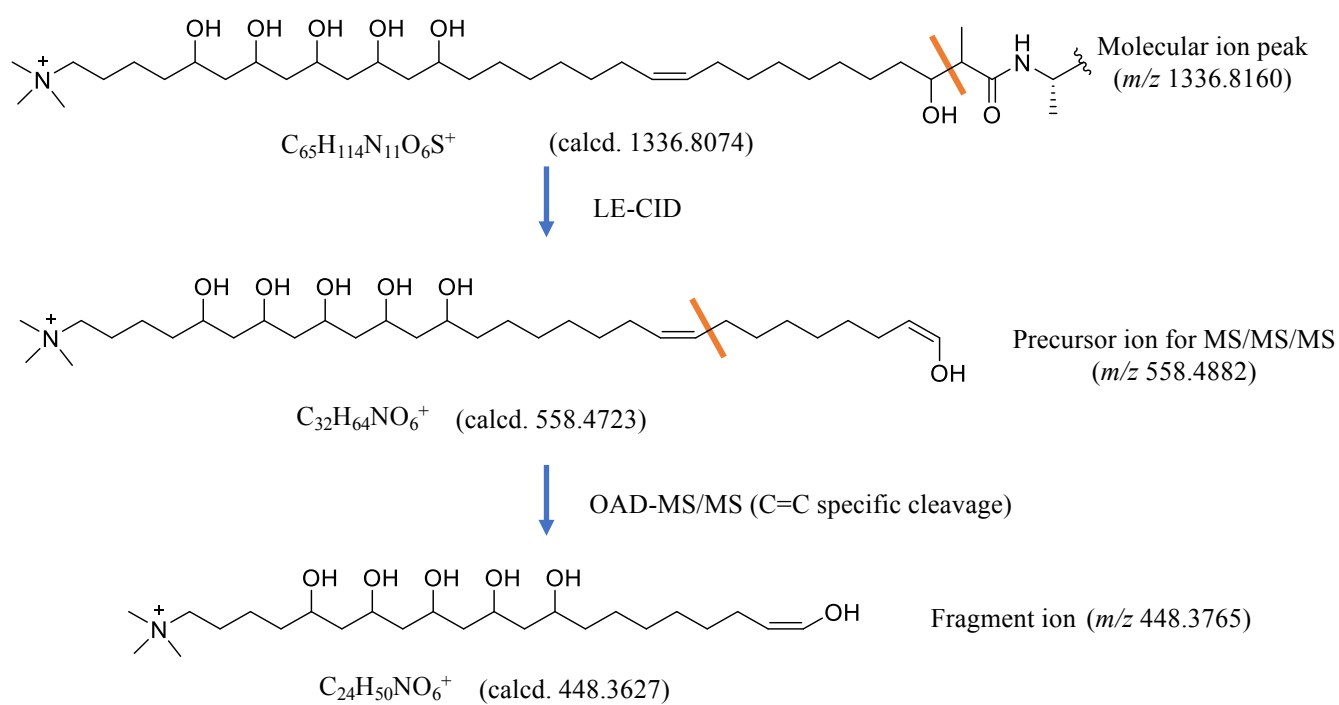
Observed (<i>m/z</i>)	Chemical formula	Predicted structure
116.1456	C ₇ H ₁₈ N ⁺	
144.1415	C ₈ H ₁₈ NO ⁺	
158.1522	C ₉ H ₂₀ NO ⁺	
202.1755	C ₁₁ H ₂₄ NO ₂ ⁺	
246.1988	C ₁₃ H ₂₈ NO ₃ ⁺	
290.2256	C ₁₅ H ₃₂ NO ₄ ⁺	
320.2363	C ₁₆ H ₃₄ NO ₅ ⁺	

b

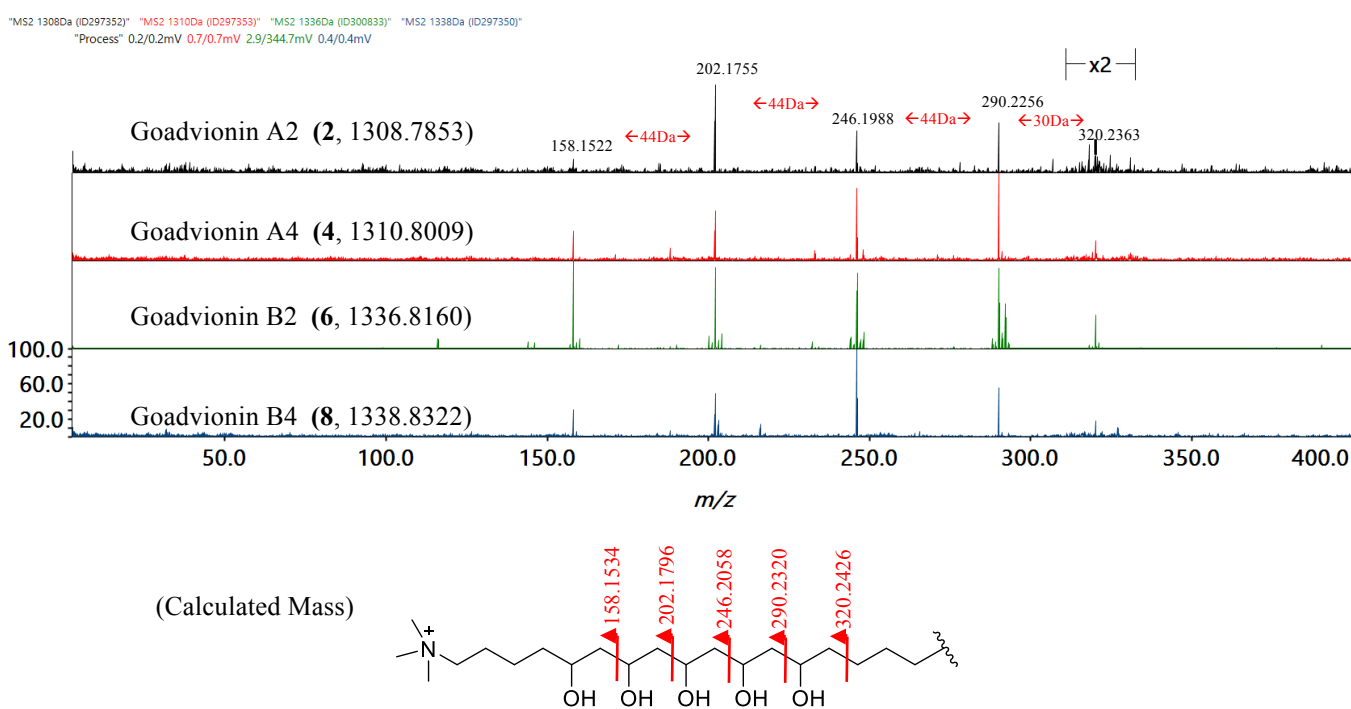
Supplementary Fig. 15. Characterization of HAD-MS/MS fragment ion peaks of goadvionin B2 (a) and proposed mechanism of fragmentation pathway by HAD in goadvionin B2 (6) (b). The major HAD products are likely due to H• abstraction from C-H of the alkyl chain (pathway A and B), not but from hydroxyl groups, because the binding energy of O-H is qualitatively greater than that of C-H. The activation energy required for H• abstraction from hydrogens of the alkyl chain without hydroxyl groups was calculated to be 34 kJ/mol in our previous study [19]. Therefore, it is considered that the presence of hydroxyl group significantly reduces the activation energy required for H• abstraction from the alkyl chain.



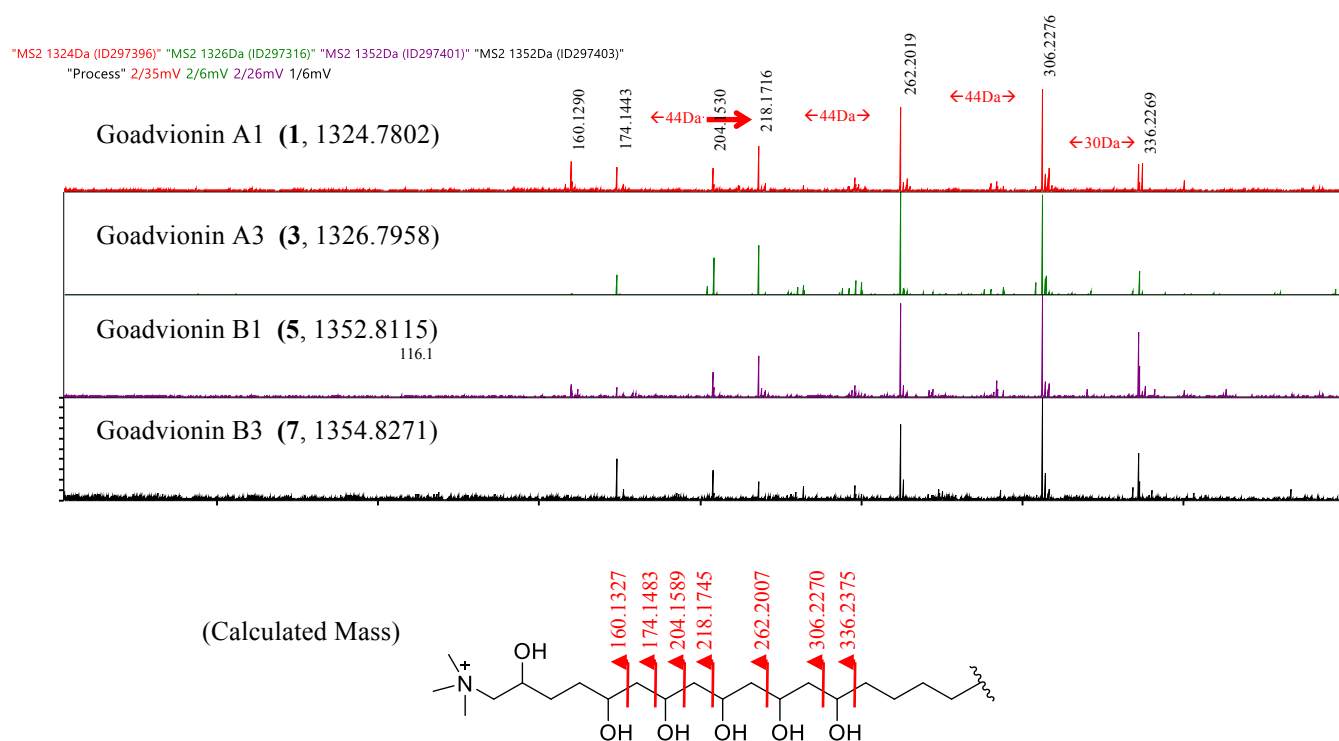
Supplementary Fig. 16. OAD-MS3 spectrum of goadvionin B2 (6)
Supplemental collisional activation energy = 0eV



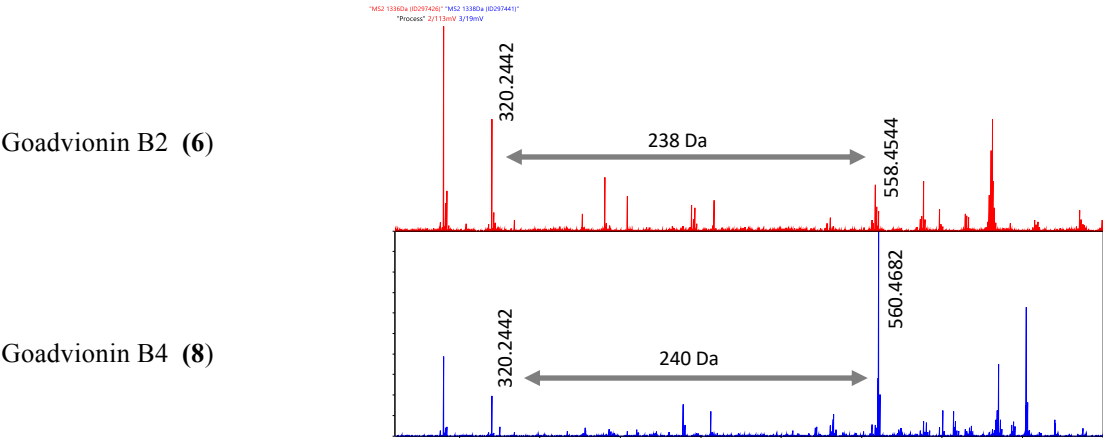
Supplementary Fig. 17. Scheme for OAD-MS3 analysis of goadvionin B2 (6)



Supplementary Fig. 18. HAD-MS/MS spectrum of goadvionins A2, A4, B2, and B4 (m/z 158 to 320)
 Supplemental collisional activation (SA) energy for goadvionin A2, A3, B4 = 0.5eV.
 SA energy for Goadvionin B2 = 0eV.

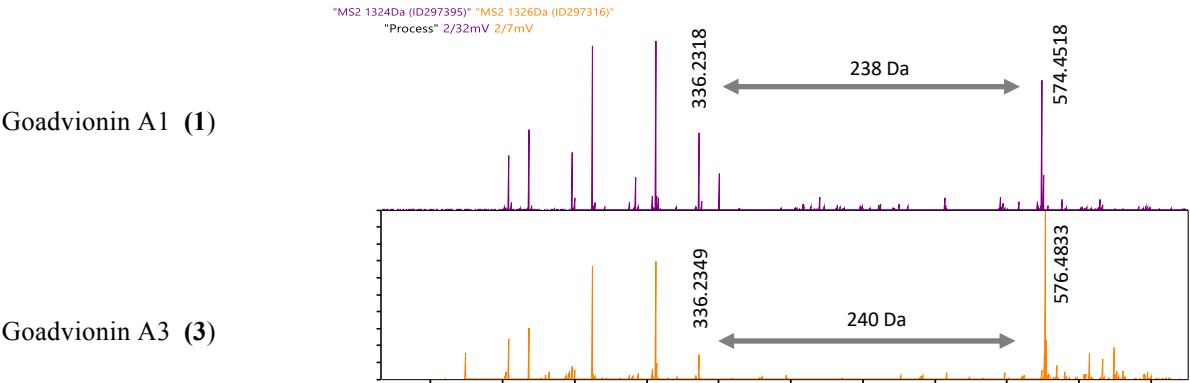


Supplementary Fig. 19. HAD-MS/MS spectrum of goadvionin A1, A3, B1, and B3 (m/z 160 to 336)
Supplemental collisional activation energy = 0.4 eV.



Observed (<i>m/z</i>)	Chemical formula	Calculated mass (<i>m/z</i>)	Predicted structure
320.2442	C ₁₆ H ₃₄ NO ₅ ⁺	320.2426	
558.4544	C ₃₂ H ₆₄ NO ₆ ⁺	558.4723	
560.4682	C ₃₂ H ₆₆ NO ₆ ⁺	560.4879	

Supplementary Fig. 20. HAD-MS/MS spectra of goadvionins B2 (6) and B4 (8)
Supplemental collisional activation energy = 0.3 eV.



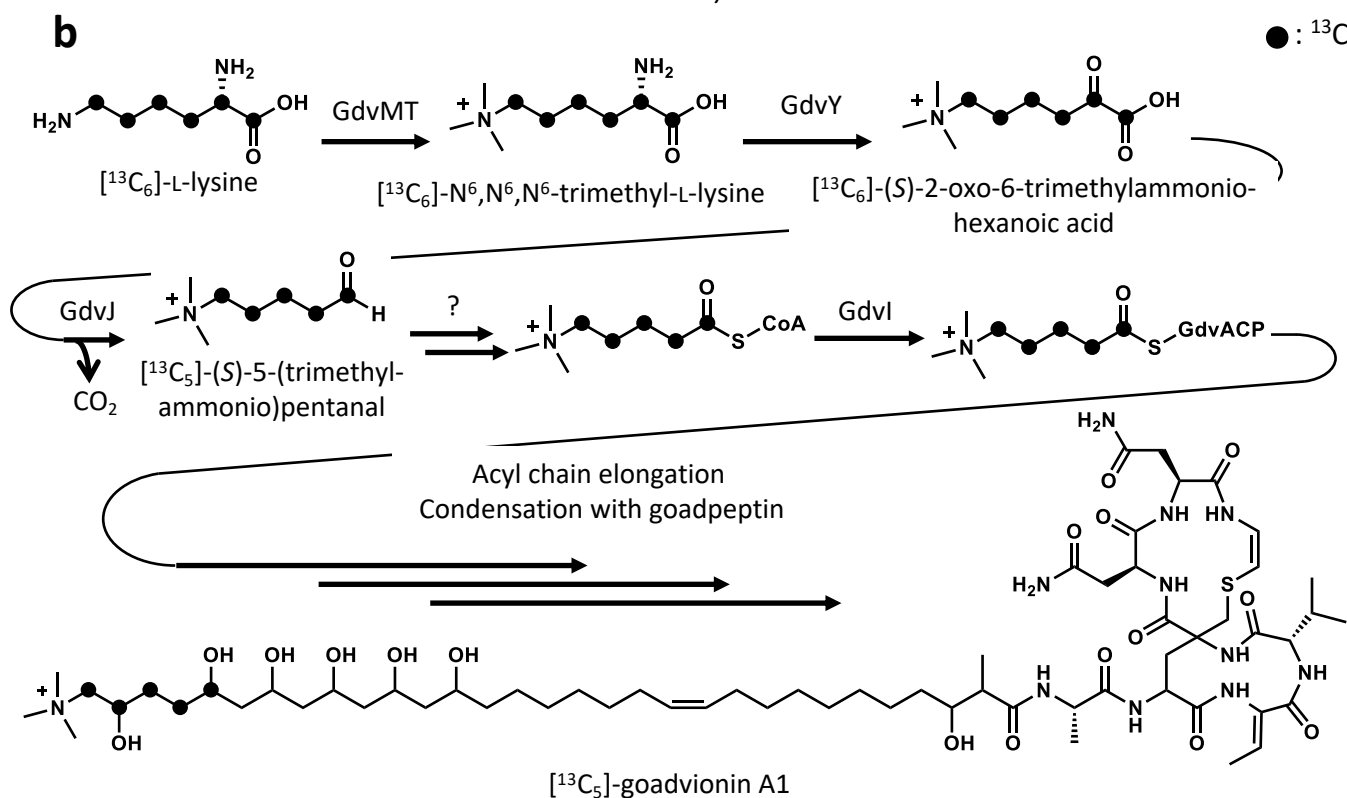
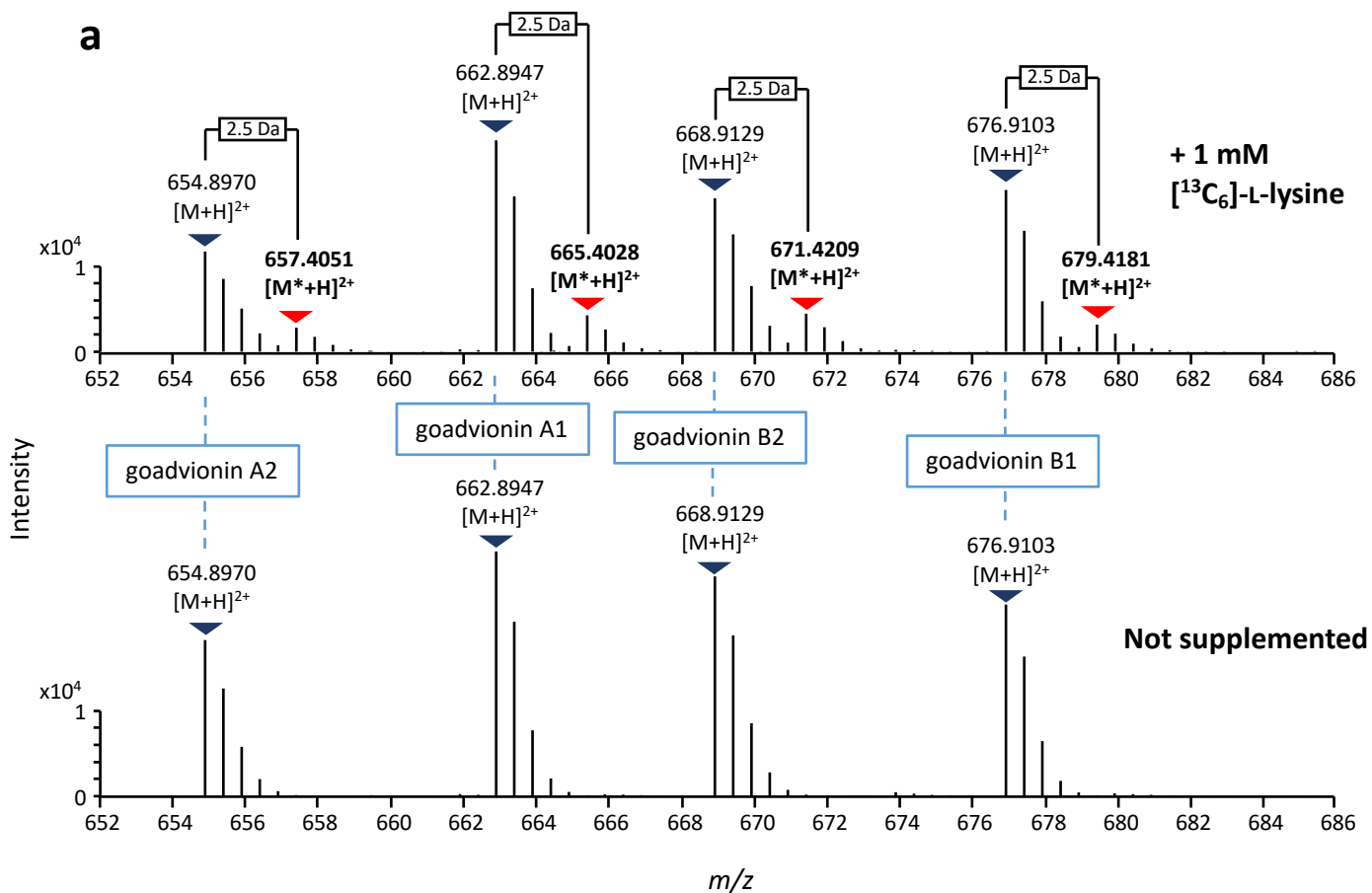
Observed (<i>m/z</i>)	Chemical formula	Calculated mass (<i>m/z</i>)	Predicted structure
336.2318 (336.2349)	C ₁₆ H ₃₄ NO ₆ ⁺	336.2375	
574.4518	C ₃₂ H ₆₄ NO ₇ ⁺	574.4672	
576.4833	C ₃₂ H ₆₆ NO ₇ ⁺	560.4828	

Supplementary Fig. 21. HAD-MS/MS spectra of goadvionin A1 (1) and A3 (3)
Supplemental collisional activation energy = 0.4 eV

RID-MS/MS analysis of goadvionin B2 (**6**)

HAD-MS/MS analysis targeting the molecular ion peak of goadvionin B2 (**6**) revealed seven fragment ion peaks between m/z 116 and 320 (Supplementary Fig. 14, m/z 116, 144, 158, 202, 246, 290, and 320). Based on the accurate mass and chemical formula of each fragment ion (Supplementary Fig. 15), along with the partial structure of **6** established using NMR analysis, all of the unassigned hydroxyl groups were found to be a part of the 1, 3, 5, 7, 9-pentaol substructure and present from C-19' to C-27' positions (Supplementary Fig. 9, substructure C).

For OAD-MS/MS analysis, the molecular ion peak was initially selected as the precursor ion (m/z 1336.8160), but we could not obtain fragment ion peaks with enough intensity. Then, the molecular ion peak was initially subjected to LE-CID and the resulting fragment ion (m/z 558.4882, Supplementary Fig. 16 and 17) was further subjected to OAD-MS3 analysis. Consequently, the fragment ion peak corresponding to double bond-specific cleavage was observed in the OAD-MS3 spectrum (Supplementary Fig. 16), and the position of isolated olefin in the fatty-acyl group of **6** was determined to be C-11' and 12'. Finally, the proposed structure of **6** was further confirmed by the HE-CID-MS/MS fragmentation pattern (Supplementary Fig. 54).

Supplementary Figure 22. $[^{13}\text{C}_6]$ -lysine incorporation to goadvionin.

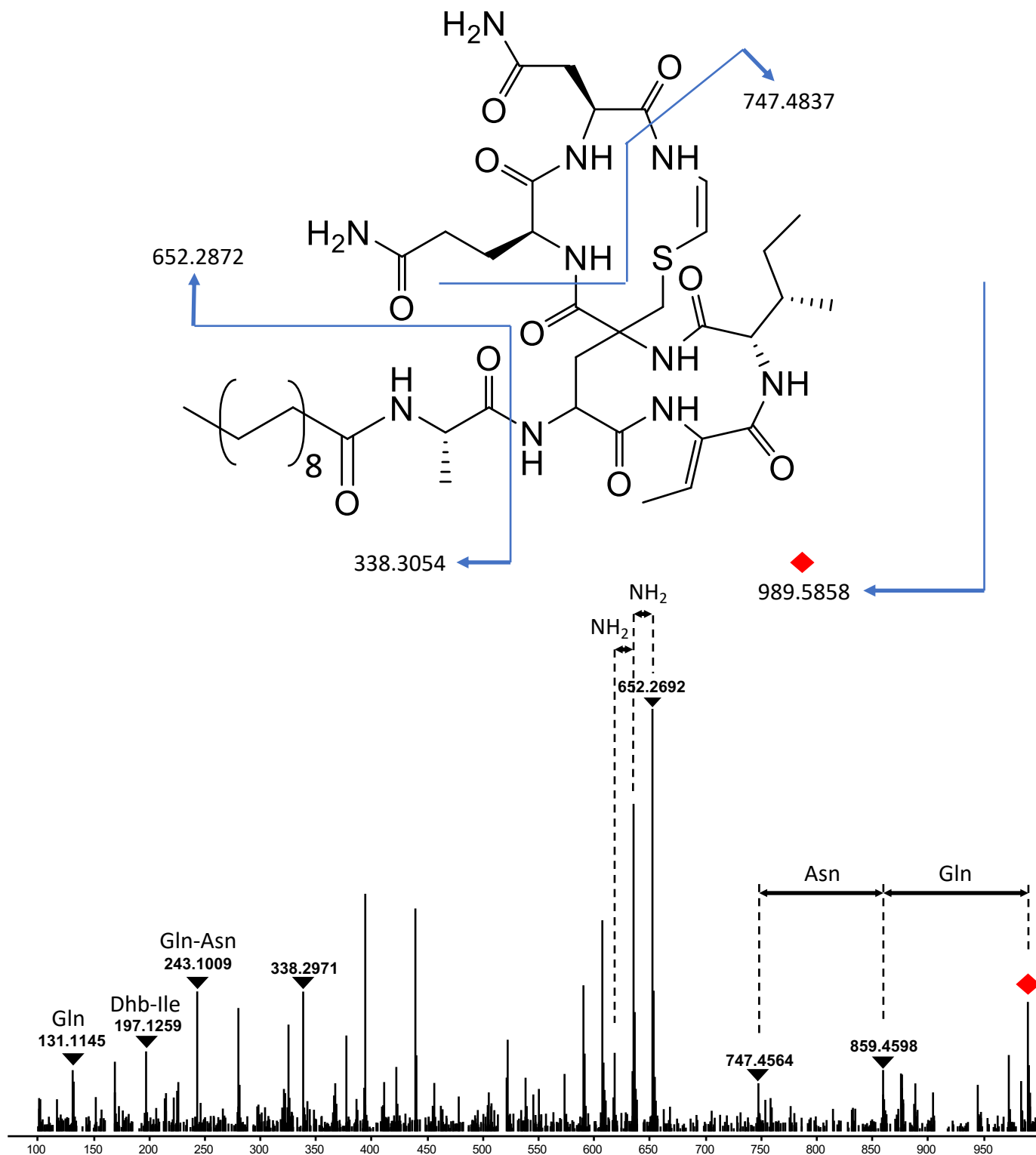
Supplementary Figure 22. [$^{13}\text{C}_6$]-lysine incorporation to goadvionin.

(a) Detected m/z peaks of the goadvionins supplemented with [$^{13}\text{C}_6$]-L-lysine.

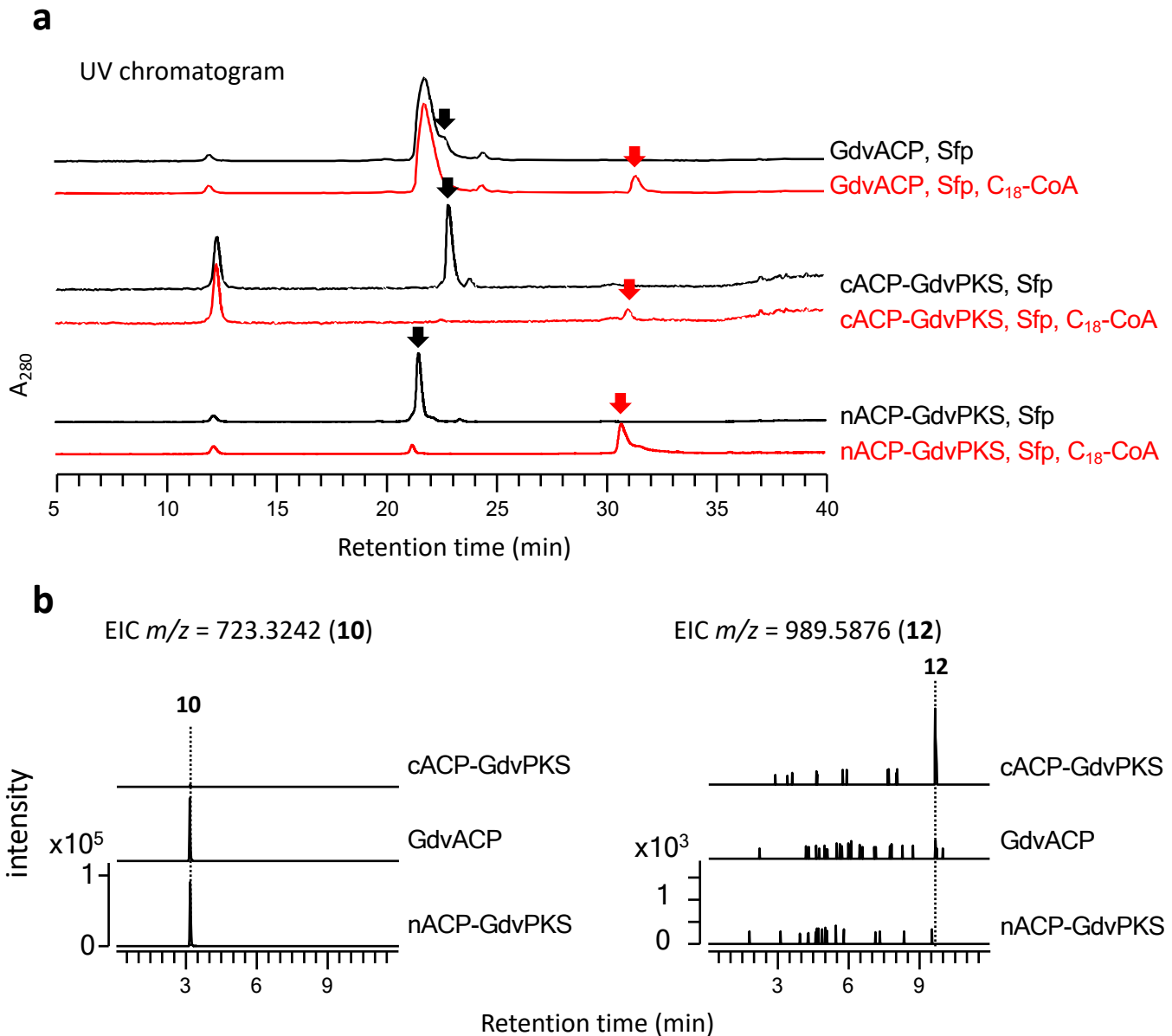
Culture butanol extract was analyzed by LC-MS and peaks found in retention times from 5.6 to 6.0 min are shown. Compared with the goadvionins peaks from negative control (no [$^{13}\text{C}_6$]-L-lysine supplementation culture), m/z 2.508 (for 5.016 Da) shifted peaks were observed, indicating the incorporation of five ^{13}C atoms in the final products. M* refers to the compounds containing heavy atoms.

(b) Proposed goadvionin pathway indicated by [$^{13}\text{C}_6$]-L-lysine supplementation study.

[$^{13}\text{C}_6$]-L-lysine is converted to [$^{13}\text{C}_6$]- N^6,N^6,N^6 -trimethyl-L-lysine by three methylation reaction probably catalyzed by GdvMT. Subsequently, GdvMT product may undergo deamination reaction catalyzed by GdvY to produce [$^{13}\text{C}_6$]-(*S*)-2-oxo-6-trimethylammonio-hexanoic acid. Next, GdvJ may catalyze the decarboxylation reaction to produce [$^{13}\text{C}_5$]-(*S*)-5-trimethylammonio-pentanal. This decarboxylation process reduces the number of ^{13}C atoms from six to five in the compound. Then, [$^{13}\text{C}_5$]-(*S*)-5-trimethylammonio-pentanal may become starting unit for the elongation reaction proceed in the fatty acid biosynthesis system to produce final products, goadvionins.



Supplementary Fig. 23. MS/MS fragment of stearyl goadpeptin B (12)



Supplementary Fig. 24. PK-RiPP condensation reaction with ACPs including *gdv* cluster.

(a) Acylation of ACPs with C_{18} stearoyl-CoA by Sfp. Red traces indicate full set reaction, and black traces indicate the reaction without C_{18} -CoA. Red arrows indicate C_{18} stearoyl ACPs, and black arrows indicate holo-ACPs. (b) *In vitro* condensation reaction were detected by LC-MS. EIC corresponding to both goadpeptin B and stearoyl goadpeptin B were analyzed. Only the reaction with cACP-GdvPKS was proceeded.

N-terminal ACP domain in GdvPKS (nACP-GdvPKS) & GdvACP & C-terminal ACP domain in GdvPKS (cACP-GdvPKS)

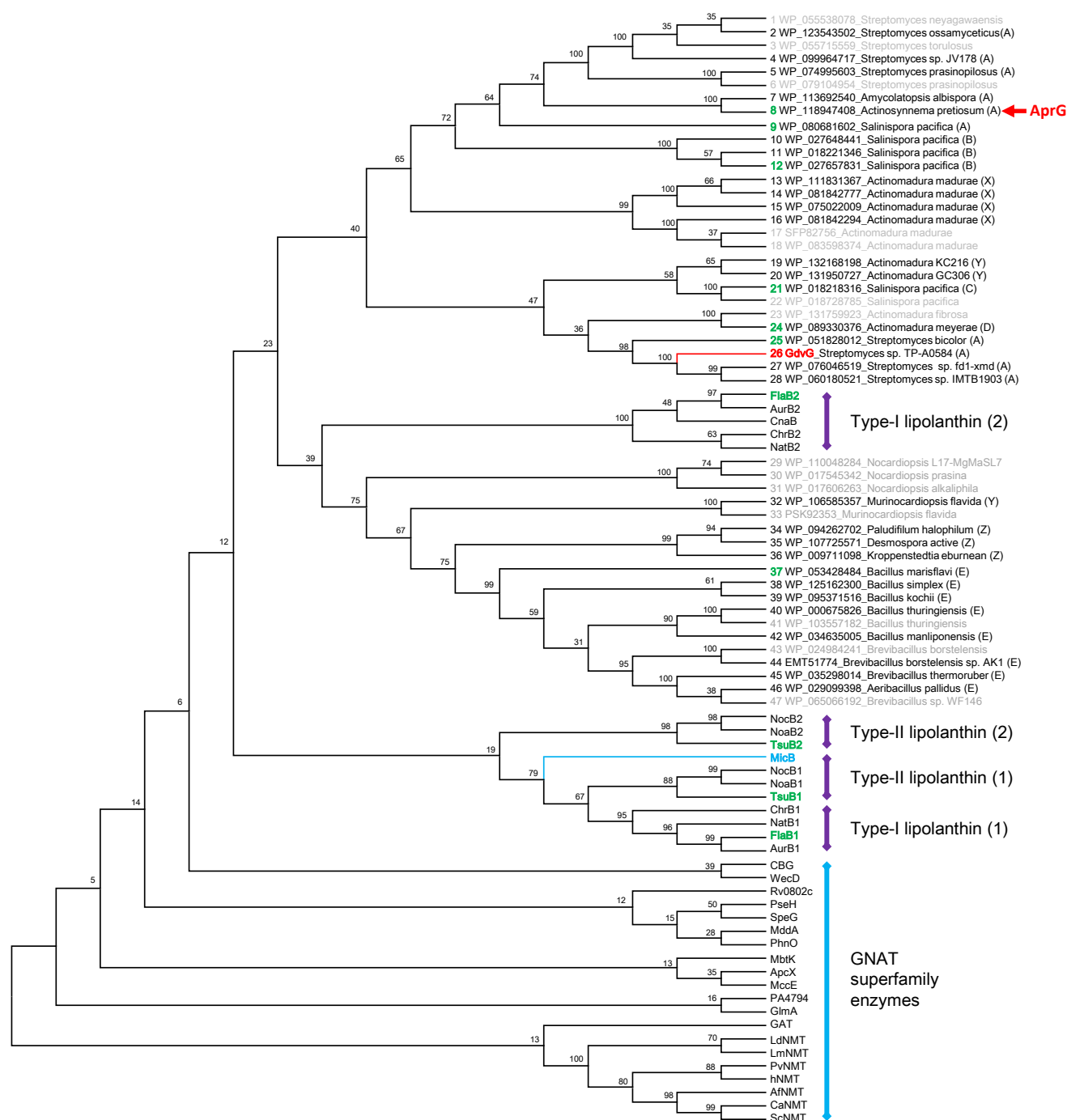
nACP-GdvPKS	-----VTLMDEQALRSWLASWLAAELDMAADD
GdvACP	-----VSTTVTTEKLF EAVTALLEEIAEVPADT
cACP-GdvPKS	PTPLLADLASTVVGRPVSDGTGGAGVAAALARADRAERRELLRRYVREVVAALS L PGPS
	: * : :
nACP-GdvPKS	IDHDQPLG-DLGVSSRMLVELSGALEELLNRPLSPVVAYRAPTVASLARHLTRLDTGVEP
GdvACP	VTRDSHFRDDLDVDSLLMVELSAAIEDQFGVDIPDDEYGAI RTVGDLVD TIHRAQA----
cACP-GdvPKS	ALAPDARFFDLGLDSL LTMTVIRRLKAVVPG-LRPSVLRDNPTVDGLAETLLDLLGELGR
	. ** . . . * : : : : : . : ** . * . :
nACP-GdvPKS	A---
GdvACP	----
cACP-GdvPKS	TTAD

Supplementary Fig. 25. Alignment of cACP-GdvPKS with the other ACP domains in *gdv* cluster.

Comparison with N-terminal ACP domain in GdvPKS (nACP-GdvPKS) & GdvACP & C-terminal ACP domain in GdvPKS (cACP-GdvPKS).

Supplementary table 5. Identity and similarity between proteins in this paper.

Pair of proteins	Identity (%)	Similarity (%)
cACP-GdvPKS & nACP-GdvPKS	20.4 (19/93)	35.5 (33/93)
cACP-GdvPKS & GdvPKS	15.7 (18/115)	30.4 (35/115)
GdvG & AprG	42.0 (87/207)	61.8 (128/207)
GdvG & NocB1	19.9 (41/206)	39.3 (81/206)
GdvG & NocB2	21.4 (45/210)	35.2 (74/210)
GdvG & MicB	18.2 (30/165)	28.5 (47/165)



Supplementary Fig. 26. Phylogenetic tree of GdvG with GNAT superfamily enzymes.

GdvG and MicB are indicated in red and cyan, respectively. Purple bars indicate the GNAT superfamily proteins in type-I and II lipolanthin biosynthetic gene clusters. Cyan bar indicates the conventional GNAT superfamily enzymes and N-myristoyltransferase. Flanking region of protein accession numbers in gray could not be analyzed according to the low sequence quality. The nucleotide accession numbers which were used for analysis are listed in supplementary table 6 and 7. Alphabets in the bracket indicate the type of biosynthetic genes (divided into four categories: LanKC-LanD / type I PKS / FAS / NRPS) found in the flanking region (within around 40 kb) summarized in supplementary table 8. Gene map for the flanking region of number in green are shown in Figure 5.

Supplementary Table 6. Accession number of the nucleotide sequences used for the genome mining.

Nucleotide accession numbers used for the genome mining study are listed.

number in the phylogenetic tree	protein accession number	nucleotide accession number	strain
2	WP_123543502	NZ_RJKY01000001	<i>Streptomyces ossamyceticus</i> SAI-001
4	WP_099964717	NZ_PEKU01000003	<i>Streptomyces</i> sp. JV178
5	WP_074995603	NZ_FMZK01000019	<i>Streptomyces prasinopilosus</i> CGMCC 4.3504
7	WP_113692540	NZ_CP015163	<i>Amycolatopsis albispora</i> WP1
8 <i>apr</i>	WP_118947408	NZ_CP029607	<i>Actinosynnema pretiosum</i> subsp. <i>pretiosum</i> ATCC31280
9	WP_080681602	NZ_AZWA01000042	<i>Salinispora pacifica</i> CNT084
10	WP_027648441	NZ_AZWC01000035	<i>Salinispora pacifica</i> CNT001
11	WP_018221346	NZ_KB904848	<i>Salinispora pacifica</i> CNQ768
12	WP_027657831	NZ_AZWN01000035	<i>Salinispora pacifica</i> CNR894
13	WP_111831367	NZ_UAPP01000005	<i>Actinomadura madurae</i> NCTC11373
14	WP_081842777	NZ_AWOO02000045	<i>Actinomadura madurae</i> LIID-AJ290
15	WP_075022009	NZ_FOVH01000007	<i>Actinomadura madurae</i> DSM 43067
16	WP_081842294	NZ_AWOO02000006	<i>Actinomadura madurae</i> LIID-AJ290
19	WP_132168198	NZ_SMJX01000001	<i>Actinomadura</i> sp. KC216
20	WP_131950727	NZ_SMKM01000022	<i>Actinomadura</i> sp. GC306
21	WP_018218316	NZ_KB900388	<i>Salinispora pacifica</i> DSM 45548 = CNT-148
24	WP_089330376	NZ_FZOR01000050	<i>Actinomadura meyeri</i> DSM 44715
25	WP_051828012	NZ_JOFR01000030	<i>Streptomyces bicolor</i> NRRL B-3897
27	WP_076046519	NZ_CP019798	<i>Streptomyces</i> sp. fd1-xmd
28	WP_060180521	NZ_LQYB01000031	<i>Streptomyces</i> sp. IMTB 1903
	FlaB1/FlaB2	JNZO01000011	<i>Streptomyces flavochromogenes</i>
	TsuB1/TsuB2	CAJY01000026	<i>Tsukamurella</i> sp 1534
32	WP_106585357	NZ_PYGA01000018.1	<i>Murinocardiopsis flavida</i> DSM 45312
34	WP_094262702	NZ_NOWF01000001.1	<i>Paludifilum halophilum</i> DSM 102817
35	WP_107725571	NZ_PZZP01000001.1	<i>Desmospora activa</i> DSM 45169
36	WP_009711098	NZ_FTOD01000015.1	<i>Kroppenstedtia eburnea</i> DSM 45196
37	WP_053428484	NZ_LGUE01000004	<i>Bacillus marisflavi</i> JCM 11544
38	WP_125162300	NZ_RRZF01000020.1	<i>Bacillus simplex</i> BE23
39	WP_095371516	NZ_CP022983.1	<i>Bacillus kochii</i> BDGP4
40	WP_000675826	CP002508.1	<i>Bacillus thuringiensis</i> YBT-020
42	WP_034635005	NZ_JOTN01000001.1	<i>Bacillus manliponensis</i> JCM 15802
44	EMT51774	APBN01000006.1	<i>Brevibacillus borstelensis</i> AK1
45	WP_035298014	NZ_JQMH01000006.1	<i>Brevibacillus thermoruber</i> PM1
46	WP_029099398	NZ_JYCD01000169.1	<i>Aeribacillus pallidus</i> GS3372

Supplemental Table 7: Accession number of the protein sequences used for the phylogenetic analysis. GNAT family enzymes were selected based on reference: Favrot L., et al. Biochemistry. 55(7):989-1002 2016. (doi: 10.1021/acs.biochem.5b01269). *N*-myristoyltransferase were selected based on reference: Salah Ud-Din Al., et al. Int J Mol Sci. 17(7) pii: E1018. 2016. (doi: 10.3390/ijms17071018). GNAT family proteins in type-I and II lipolanthin biosynthetic gene cluster were selected based on reference: Wiebach V., et al. Nat Chem Biol. 14(7):652-654. 2018 (doi: 10.1038/s41589-018-0068-6).

class	name	accession number	amino acids	strain
GNAT family enzyme	PA4794	WP_003111411	160	<i>Pseudomonas aeruginosa</i>
	ApxC	WP_009875272	172	<i>Actinobacillus pleuropneumoniae</i>
	GAT	WP_003180517	146	<i>Bacillus licheniformis</i>
	CBG	WP_003952521	339	<i>Streptomyces clavuligerus</i>
	MccE (GNAT domain)	WP_021528069	312 (278-589)	<i>Escherichia coli</i>
	PseH	WP_120818319	180	<i>Helicobacter pylori</i>
	SpeG	WP_001088091	173	<i>Vibrio cholerae</i>
	WecD	WP_001145166	224	<i>Escherichia coli</i>
	MddA	WP_129406907	171	<i>Salmonella enterica</i>
	MbtK	WP_003406956	210	<i>Mycobacterium tuberculosis</i>
	PhnO	WP_000602316	154	<i>Salmonella enterica</i>
	Rv0802c	WP_003404111	218	<i>Mycobacterium tuberculosis</i>
	GlmA	WP_010963508	320	<i>Clostridium acetobutylicum</i>
N-myristoyltransferase	CaNMT	XP_722713	451	<i>Candida albicans</i>
	LdNMT	XP_001467690	421	<i>Leishmania infantum</i>
	LmNMT	XP_001685320	421	<i>Leishmania major</i>
	ScNMT	NP_013296	455	<i>Saccharomyces cerevisiae</i>
	PvNMT	XP_001616826	410	<i>Plasmodium vivax</i>
	AfNMT	XP_752019	492	<i>Aspergillus fumigatus</i>
	hNMT	NP_004799	498	<i>Homo sapiens</i>
GNAT family proteins in type-II lipolanthin biosynthetic gene cluster	MicB	AVH76812	141	<i>Microbacterium arborescens</i>
	TsuB1	WP_019202646	193	<i>Tsukamurella</i> sp. 1534
	TsuB2	WP_019202647	197	
	NocB1	KZM69139	204	<i>Nocardia terpenica</i>
	NocB2	KZM69140	207	
	NoaB1	WP_069160315	200	<i>Nocardia altamirensis</i>
	NoaB2	WP_069160316	205	
GNAT family proteins in type-I lipolanthin biosynthetic gene cluster	FlaB1	WP_030318150	206	<i>Streptomyces flavochromogenes</i>
	FlaB2	WP_030318155	208	
	ChrB1	WP_017622816	193	<i>Nocardiopsis chromatogenes</i>
	ChrB2	WP_017622812	209	
	NatB1	WP_044362952	202	<i>Streptomyces natalensis</i>
	NatB2	WP_030066285	210	
	CanB	WP_084774387	255	<i>Nonomuraea candida</i>
	AurB1	WP_055599447	206	<i>Streptomyces aureus</i>
	AurB2	WP_055599449	208	

Supplementary table 8. The types of biosynthetic genes surrounding *gdvG* homologs.

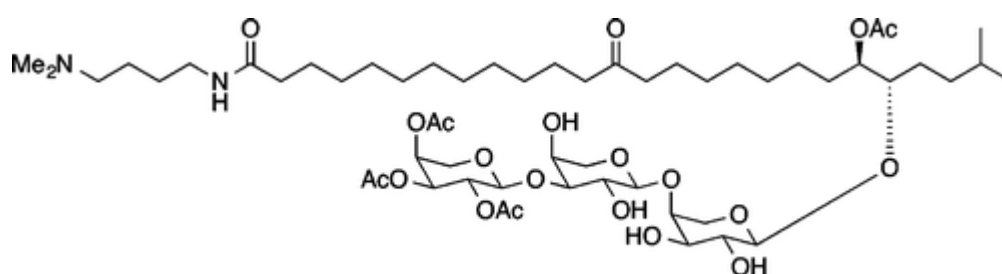
		peptide synthesis		acyl chain synthesis	
cluster type	GNAT superfamily	LanKC- LanD	NRPS	Type 1 PKS	FAS (Type II PKS)
A	○	○		○	○
B	○	○	○	○	
C	○	○			○
D	○		○	○	○
E	○		○		○
X	○			○	
Y	○			○	○
Z	○		○		

The biosynthetic genes were divided into four categories: LanKC-LanD / type I PKS / FAS / NRPS, and classified to 8 types by the combination of the genes existence.

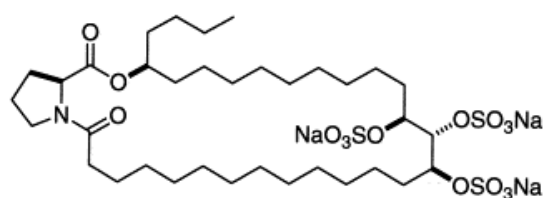
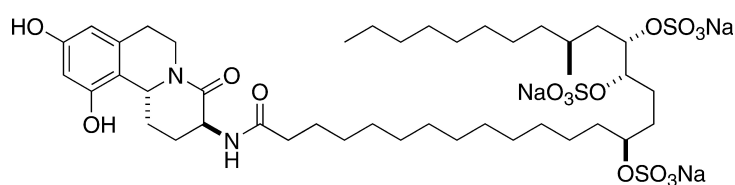
GdvG	- - -MRGDFVSLQ R CTDADFER I ADWSMSAASTYSSGAPQLLTGEAI R QAARQSGMSFLM
AtdA	MPSAHSV K L R P L E R ED L RYVHQLDNNASVM R YWF E EPY E AFV E LS D LY D KH I HDQS E RRF
SpeG	- - MNSQLTL R AL E R G DL R FIHNLNNNN N IMSYWF E EPY E S F DE L EELYN K HIHDNA E RRF
PaiA	- - -MSV K M K K C S R ED L QTLQQLS - - - - -I E TFND T F K EQNS P EN M KAY L ESAFN
GdvG	VVTH E GE - R IGAVNWGPLAYHGSYTI G SVIGAAGLWTQGYGV E ANVLLLNHLFHALN A RR
AtdA	VV E CD G E - K AGLV E LV E INHVR R A E FQ I IIS P EYQ G KGLAT R AA K LAM D YGFTVLNLY K
SpeG	VV E DAQ K N L IGLV E L I EINYIHR S A E FQ I IAP E HQ G KGFAT L IN R AL D YSFTILNL H K
PaiA	T E Q L E K ELSNMSSQFFFIY F D H E I AGYV K VNI D DAQS E EMGA E S L E I ER I Y I KNSFQ K HG
GdvG	VQMTLGTHN R HMLQIVVNAGFV V E G IL R DYFFAD G Q Y FDAVTC S LL R EEFYALAE E AE L G
AtdA	LYLIV D K E NE K A I HIY R K L GF S V E G E LMH E FFINGQ Y R N A I RM C IFQH Q YLA E H K TPG Q T
SpeG	IY L HVAV E NP K AVHLY E ECGF V E E G H LV E EFFING R Y Q D V K R MYILQ S KYLN R S E - - - -
PaiA	L G KHLLN - K A I E I AL E R N K K NIWL G V E K N ENAI A F Y K KMGFVQ T GAHSFYMG D E E Q T D L
GdvG	LAAD T VPAD D D K AAAR R MLTAHL D H R SAAY L R N LF E R
AtdA	LL K PTAQ - - - - -
SpeG	- - - - -
PaiA	IM A K T LI - - - - -

Supplementary Fig. 27. Alignment of GdvG, SpeG, AtdA, and PaiA.

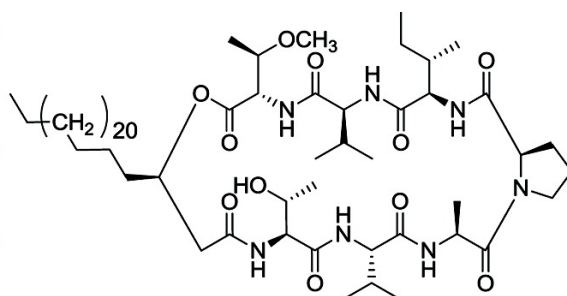
GdvG is aligned with AtdA from Spermidine *N*¹- acetyltransferase of *Escherichia coli* K-12 (UniProt ID: P0A951), SpeG from Spermidine/spermine *N*¹-acetyltransferase of *Vibrio cholerae* serotype O1 (PDB ID: 5UG4) and PaiA from Spermidine/spermine *N*¹-acetyltransferase of *Bacillus subtilis* strain 168 (UniProt ID: P21340) by using the ClustalW. Acidic amino acid residues are colored by red and basic amino acid residues are colored by blue. Tyrosine which is considered as an important residue is shown in bold.



Stellettoside A1 (Glycolipid)

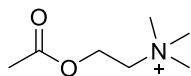
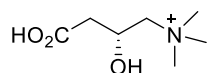
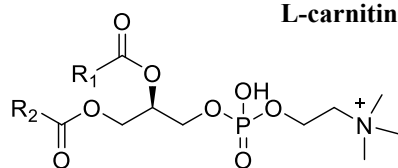
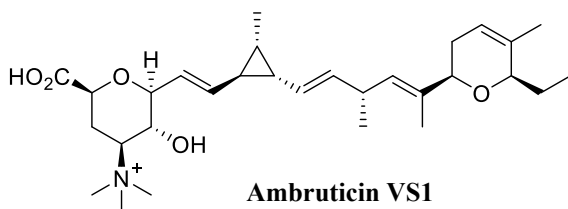
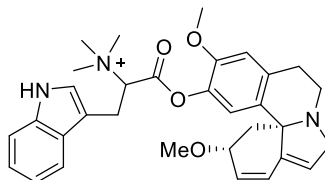
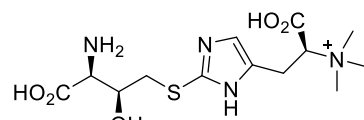
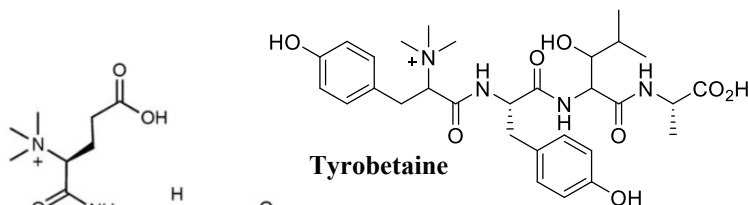
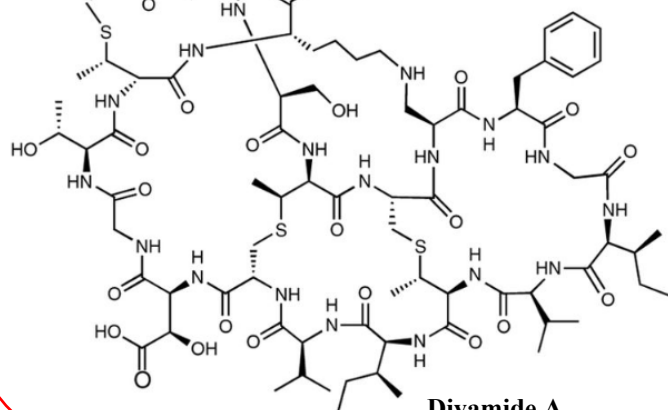
Penarolide sulfate A₁ (Marine alkaloid)

Schulzeine A (Marine alkaloid)



Peptidolipin C (Lipopeptide)

Supplementary Fig. 28. Examples of VLCFA-containing Natural product

Primary metabolites**Acetylcholine****L-carnitine****phosphatidylcholine****Polyketide****Ambruticin VS1****Secondary metabolites derived from amino acid betaine****Eryosphorine****Clithioneine****Tyrobetaine****Divamide A****Supplementary Fig. 29. Examples of natural products containing trimethyl ammonium group**

GdvG & AprG

GdvG MRGDFVSLQ-RCTDADFERIADWSMSAASTYSSGAPQLLTGEAIRQAARQSGMSFLMVVT
 AprG LQGELVSLHPASTDEDYELLAAWASSTTSVYSSGQVRPVTPAQLREAGELHGMRYLVVRD
 :::::***: .** *: * : * : * : * .***** : : * : * : * . ** : * : *

GdvG HEGERIGAVNWGPLAYHGSYTIGSVIGAAGLWTQGYGVEANVLLLNHLFHALNARRVQMT
 AprG REGEFVGGVNWRPVTYEGHFEVGNVAGPPQLWDLGYGVESVLLLLLEELFHQRNAHRVHLM
 :*** : * .*** : * : * . * : : * . : * . ** ***** : : * : * .*** ** : * : :

GdvG LGTHNRHMLQIVVNAGFVVEGILRDYFFADGQYFDAVTCSSLREEFYALAEAEELGLAAD
 AprG TAAFNKRMQIFTSGLIHVEGVLRDYFYLDGAYHHAVIGSILRDEYYRHAGLDAVRR-MD
 . . . * : * : * . * . . . : * : * : * : * : * * . * . * * : * : * : * * : * : *

GdvG TVPADDKAAARRMLTAHLDHRSAAAYLR-----
 AprG LVPRRDKEEAVELLTAHLAAHPVVAGRGTSATPAQTGPPREGVAWIDPLRTDPPPTAADP
 * * * * . : * : * : * : * : * . . . *

GdvG --NLFER-----
 AprG ALAAFSRRTNGEGA
 * . *

GdvG & NocB1

GdvG MRGDFVSLQRCTDADFERIADWSMS-AASTYSSGAPQLLTGEAIRQAARQSGMSFLMVVT
 NocB1 VLKFFALEPVLQTQEDTARAARLVSAAGPVSAVSSAGPQWVREREATALLDRSAVAFVVAGG
 : . . * : * * * : . . * : * . * : . : * : * : * .

GdvG HEGERIGAVNWGPLAYHGSYTIGSVIGAAGLWTQGYGVEANVLLLNHLFHALNARRVQMT
 NocB1 EAG--IGVLRAERLQTPGVFELALHSADDALETIGVAAGFDEIVRYLTSTTEVHRIELW
 . * * . : . * * : : : * : : * : * :

GdvG LGTHNRHMLQIVVN-AGFVVEGILRDYFFADGQYFDAVTCSSLREEFYALAEAEELGLAA
 NocB1 FGVYNRLLLDHCLESEHFRCCEGMVRDRYFAAGRYWDGLVWSITGPRLRAIREADPAAAES
 : * : * : * : : * * : * : * : * : * : * : * : * : * : *

GdvG DTVPADDKAAARRMLTAHLDHRSAAAYLRNLFER
 NocB1 ERAYARTSARLRRRAIRDRLADPEPAS-----
 : . * . * * : : * . . . *

GdvG & NocB2

GdvG MRGDFVSLQRCTDADFERIADWSMSAASTYSSGAPQLLTGEAIRQAARQ-SGMSFLMVVT
 NocB2 ---VTIHGLWTRLEESRGADPSRMLARVFESAAQRLFAGSAVAPVDDRGLGALADASDR
 . . * : . * * * * . . . * : * : * : * : . : *

GdvG HEGERIGAVNWGP-----LAYHGSYTIGSVIGAAGLWTQGYGVEANVLLLNHL
 NocB2 MLGYIYRRDDGGPGAPRPIGIYDCRLAEPARMHVGVLVLAGDAAGDVGMDAMLALLRYL
 * : * * * . : * * : * : * : * : * : * : *

GdvG FHALNARRVQMTLGTHNRHMLQIVVNAGFVVEGILRDYFFADGQYFDAVTCSSLREEFYA
 NocB2 FHFDGRRAFEMTAAVHNPHVTVRIMVHRAVRPEALLADVADGSRLVELAVGTLTEPEFAE
 * * . * : * * . * * : * : * : . . * : * * . . . : * . *

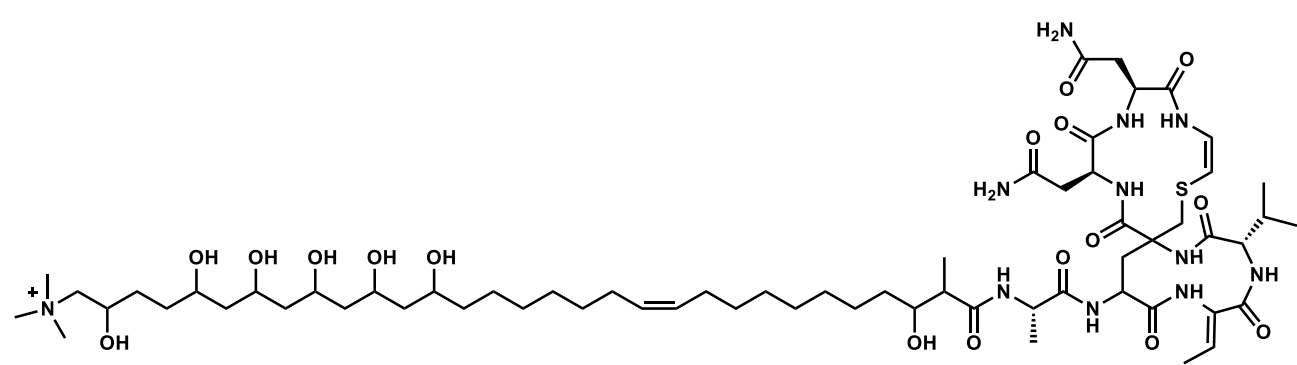
GdvG LAEEAEELGLAADTVPADDKAAARRMLTAHLDHRSAAAYLRNLFER
 NocB2 VGYRVPG--GSEFTDTRWREVRERVLAQIGAQS-----
 : * . : * : * : *

GdvG & MicB

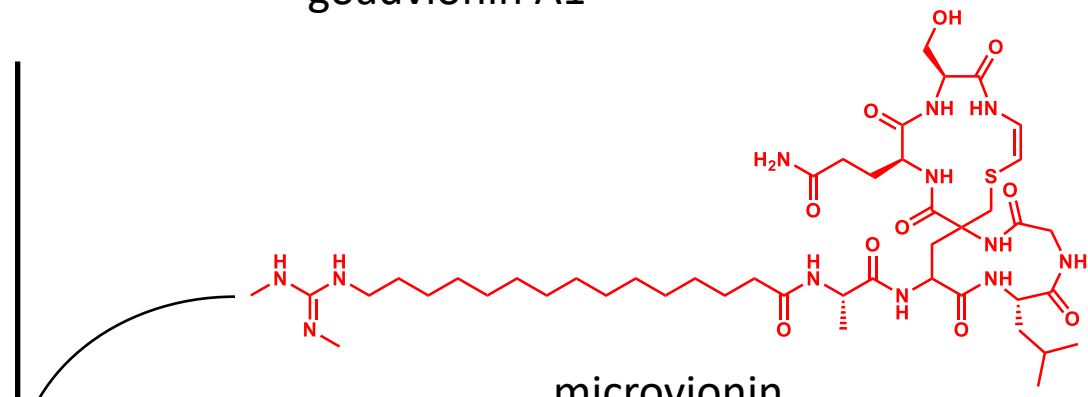
GdvG	MRGDFVSLQRCTDADFERIADWSMSAASTYSSGAPQLLTGEAIRQAARQSGMSFLMVVTH
MicB	VSALTRMLRIPNTAASIGVPAWFDDAAVDEFVGAG-----PVTV
	: . * : . * : . * . ** ** **
GdvG	EGERIGAVNWGPLAYHGSYTIGSVIGAAGLWTQGYGVEANVLLLNHLFHALNARRVQMTL
MicB	TGDRACHVGFSLAPEHPTATCFVAIPDG--HGTPDWVAAAVTEAASAAGARGAKRFEAWV
	* : * * . : . . * : * . * . * * * * . * : * . : :
GdvG	GTHNRHMLQIVVNAGFVVEGILRDYFFADGQYFDAVTCSLLREEFYALAEELGLAADT
MicB	PAWN-DAIARATERSMTREGLLEKAVFAYGERWDLTLYSVQDGGR-----
	: * . : . : . : ** : * . . ** * : : * . * :
GdvG	VPADDKAAARRMLTAHLDHRSAAAYLRNLFER
MicB	-----

Supplementary Fig. 30. Alignment of GdvG with the other GNAT superfamily enzymes.

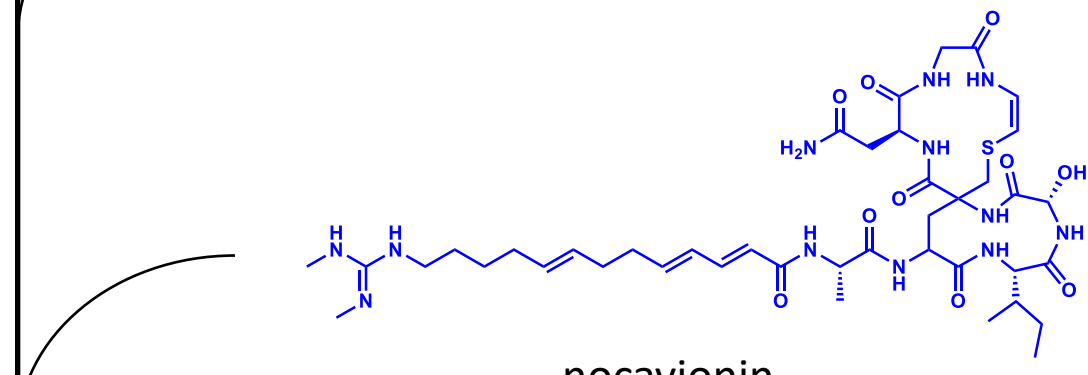
The alignment was created with clustalW. The catalytic tyrosine residues predicted in the alignment with SpeG prepared in Supplementary Fig. 27 are indicated in yellow.



goadvionin A1

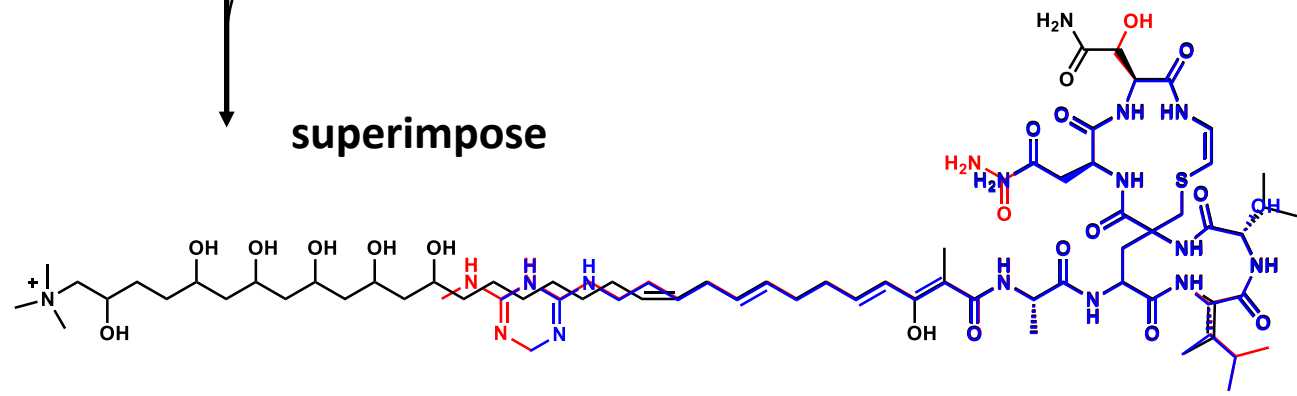


microvionin



nocavionin

superimpose

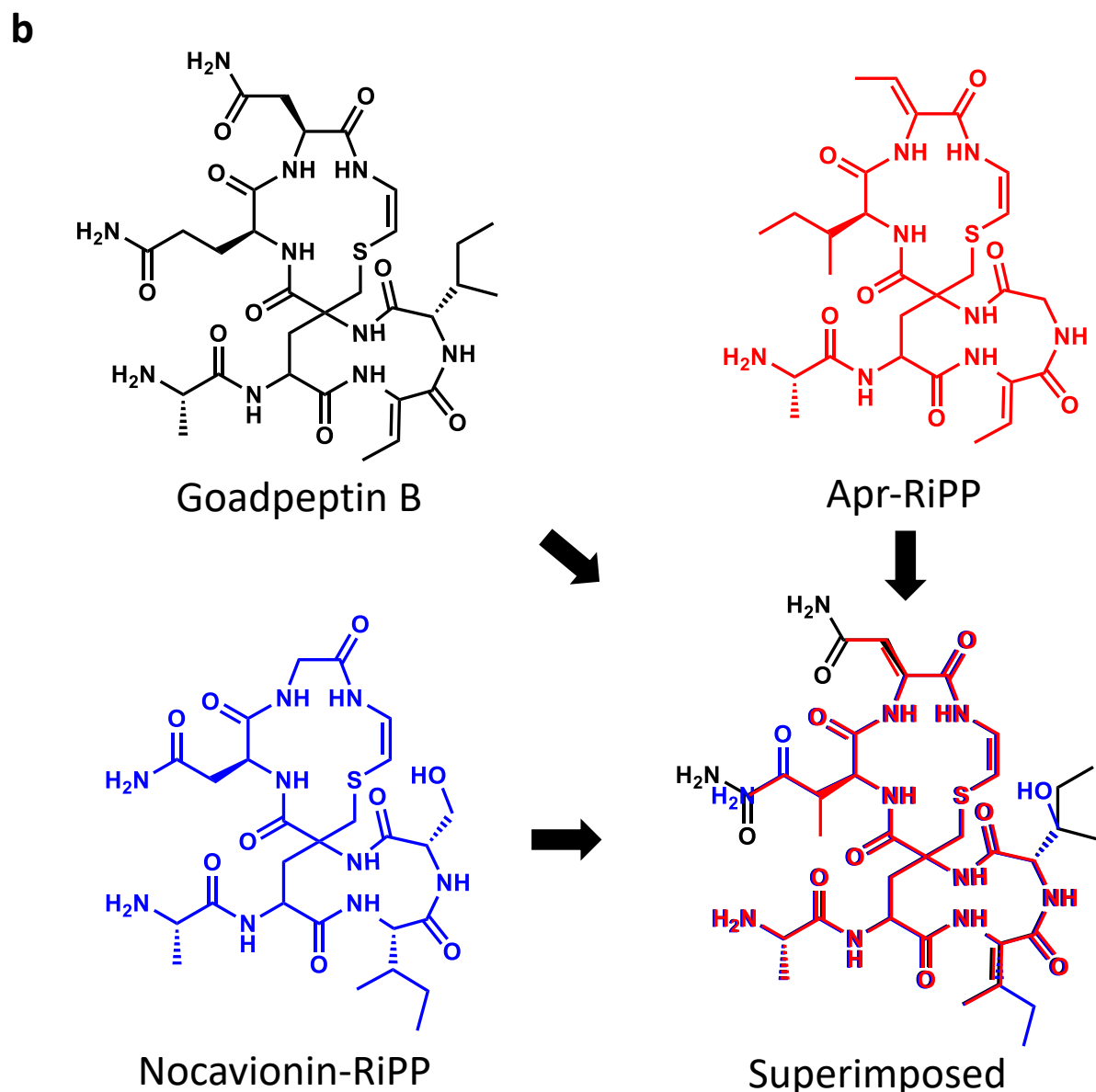


Supplementary Fig. 31. Comparison of chemical structure between goadvionin and lipolanthins.

Cleavage site

a

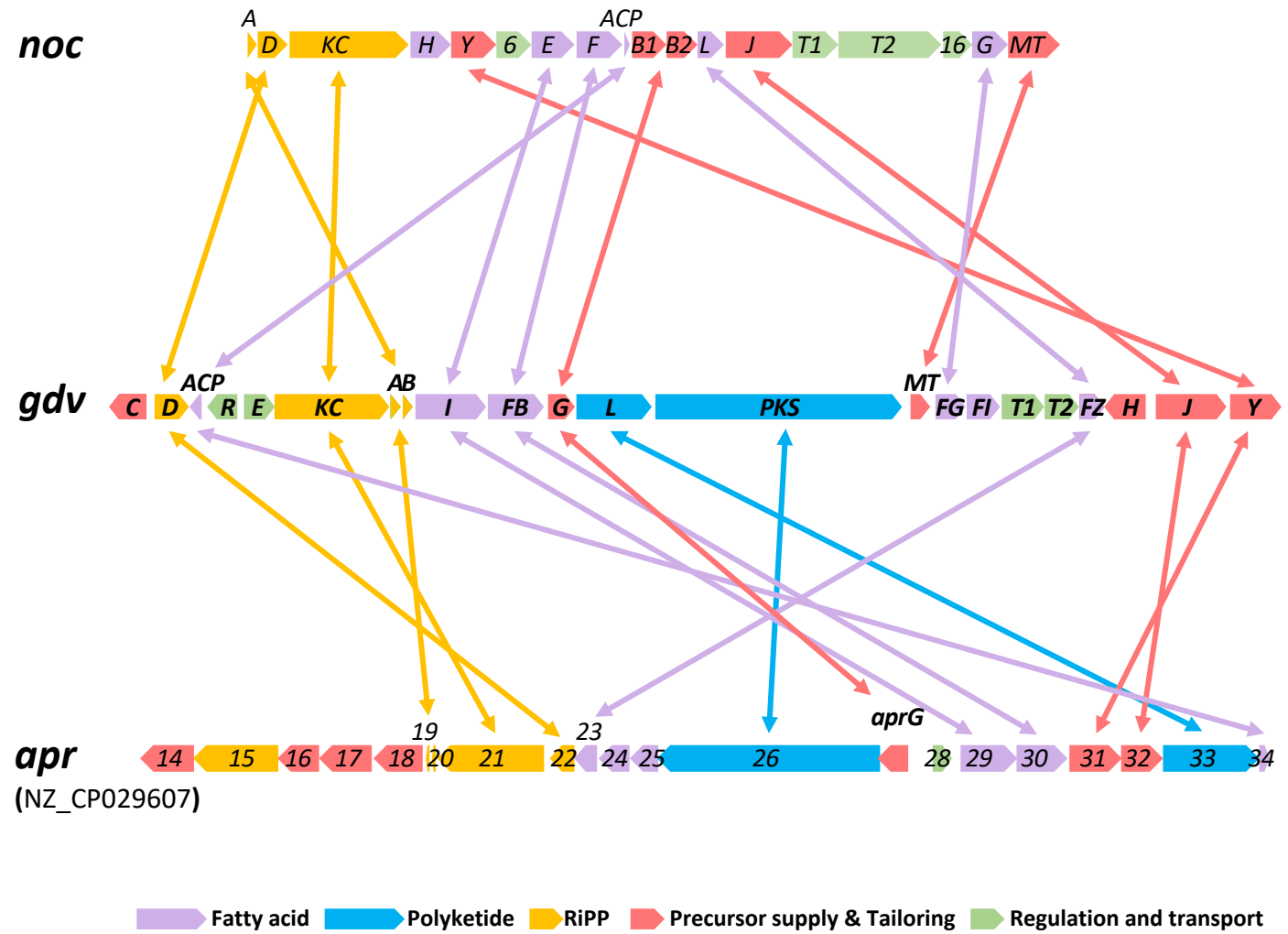
GdvA	----MESILDLQELETs-EEESALMA		ASTVSNNC
GdvB	----MESVLDLQELETs-EEESALMA		ASTISQNC
MicA	-----MSLEQLEALDAS-SEA-AEMA		ASLGSQSC
NocA	-MIDVTNIAELHELDSTSTSASAELV		ASISSNGC
AprA	--VTHMDIRELQSLVID-VEE-GDLP		ASTGSITC
AprB	MGVTHMDIRELQSLVID-VEE-GDLP		ASTGSITC



Supplementary Fig. 32. Comparison of RiPP moieties

(a) Amino acid comparison between the precursor sequences of RiPP moieties in *gdv*, *mic*, *noc* cluster and putative PK/RiPP hybrid compound biosynthetic gene cluster containing *aprG* in *Actinosynnema pretiosum* ATCC31280 (*apr* cluster). (b) Comparison of chemical structure of RiPP moieties.

1 kb



Supplementary Fig. 33. Biosynthetic gene clusters containing GdvG, NocB1 and AprG.

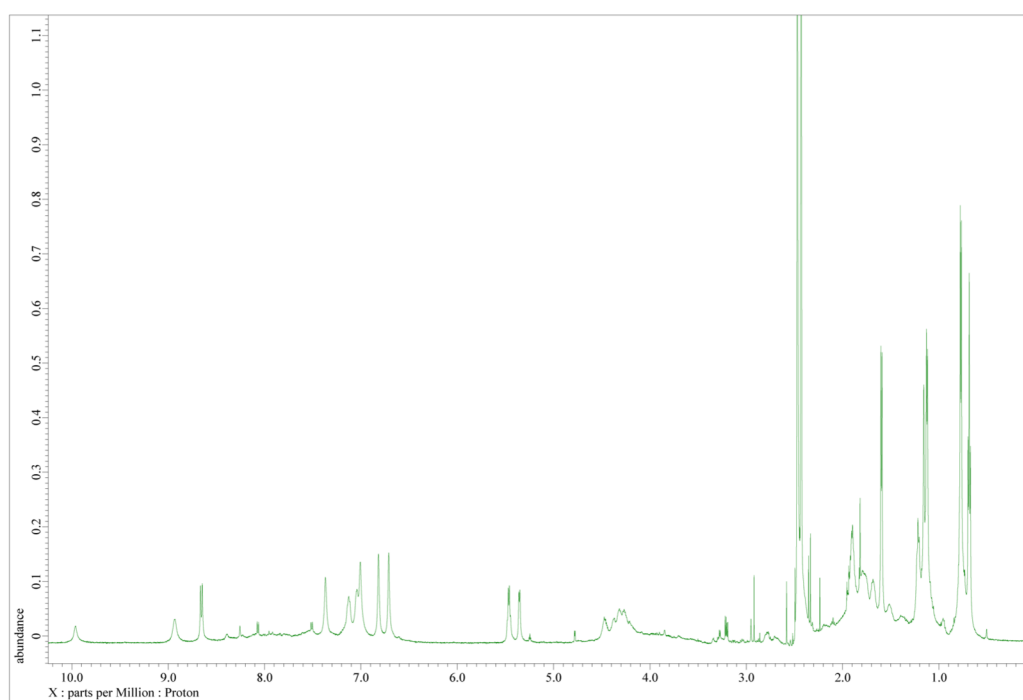
Double headed arrows indicate conserved genes. *apr* cluster was organized from the retrieved nucleotide sequence (NZ_CP029607)

Supplementary Table 9. Putative PK-RiPP hybrid compound biosynthetic gene cluster containing *aprG* in *Actinosynnema pretiosum* subsp. *pretiosum* ATCC31280

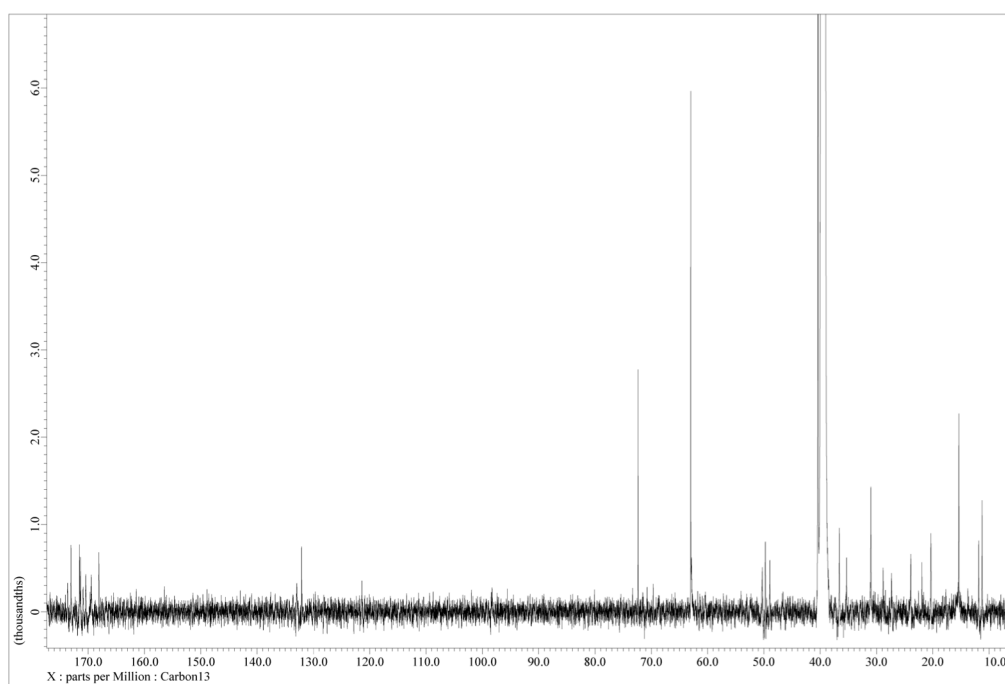
ORF	putative function
1	ricin-type beta-trefoil lectin domain protein
2	serine/threonine protein kinase
3	serine/threonine protein kinase
4	hypothetical protein
5	hypothetical protein
6	DUF1697 domain-containing protein
7	cytochrome P450
8	hypothetical protein
9	polyprenol monophosphomannose synthase
10	TetR family transcriptional regulator
11	NAD(P)H-binding protein
12	hypothetical protein
13	MarR family transcriptional regulator
14	Glycosyltransferase*
15	serine-type endopeptidase*
16	NAD-dependent epimerase/dehydratase*
17	Glycosyltransferase*
18	cytochrome P450*
19	precursor peptide
20	precursor peptide
21	LanKC-like serine/threonine protein kinase
22	LanD-like flavoprotein
23	MaoC family dehydratase
24	dehydrogenase/reductase SDR*
25	dehydrogenase/reductase SDR*
26	Type I polyketide synthase (ACP-KS-AT-DH-KR-ACP)
27(<i>aprG</i>)	GdvG-like N-acetyltransferase
28	Transposase*
29	Long chain fatty acid CoA ligase
30	Beta-ketoacyl synthase (Type II fatty acid synthase)
31	aldehyde dehydrogenase
32	pyruvate oxidase/decarboxylase
33	AMP-dependent synthetase and ligase
34	acyl carrier protein
35	ABC transporter ATP-binding protein
36	TetR family transcriptional regulator
37	MFS transporter
38	MFS transporter
39	hypothetical protein
40	NAD(P)/FAD-dependent oxidoreductase

Fatty acid synthase gene
polyketide synthase gene
RiPP biosynthetic gene
Tailoring gene
Other

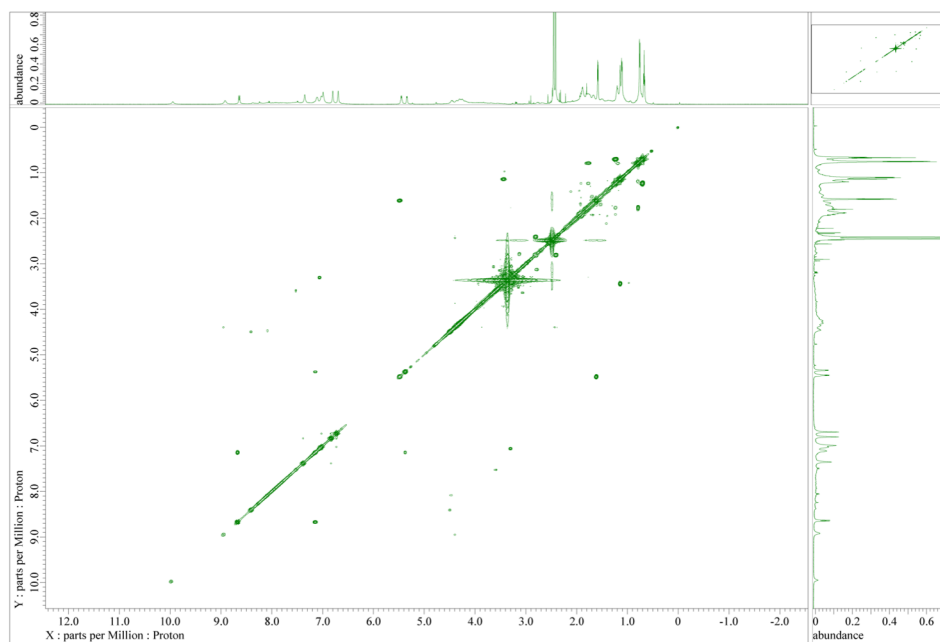
Asterisk indicates genes not present in *gdv* cluster.
The nucleotide accession number is NZ_CP029607



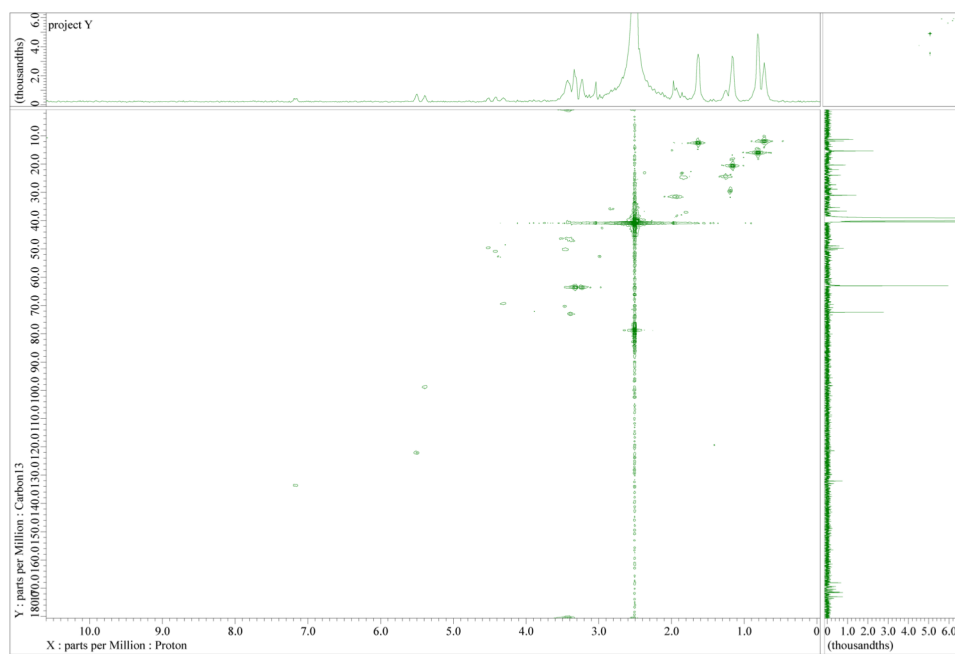
Supplementary Fig. 34. ^1H NMR spectrum of goadpeptin B (10) (600 MHz, in $\text{DMSO}-d_6$)



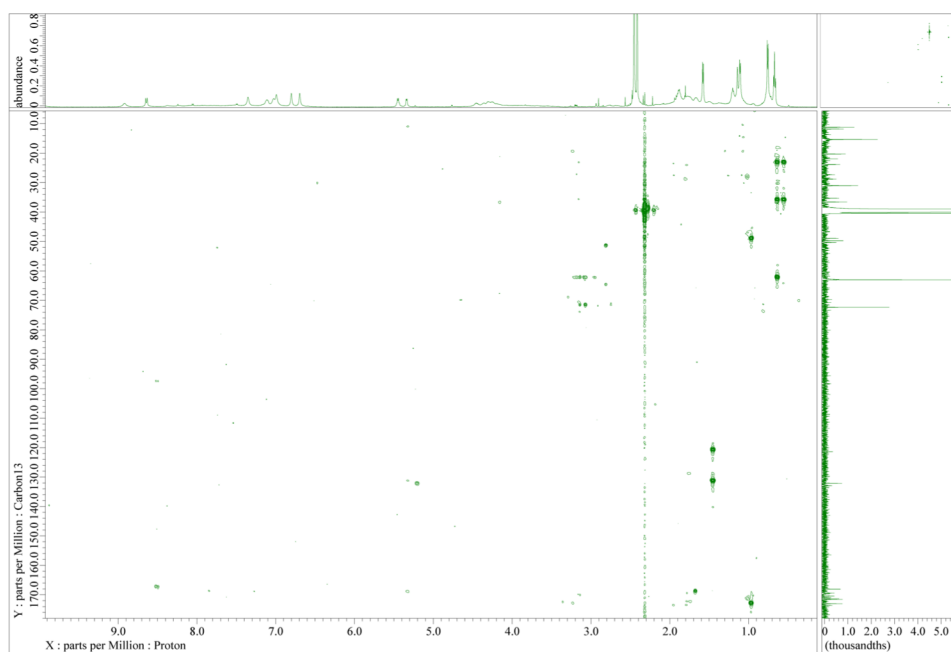
Supplementary Fig. 35. ^{13}C NMR spectrum of goadpeptin B (10) (150 MHz, in $\text{DMSO}-d_6$)



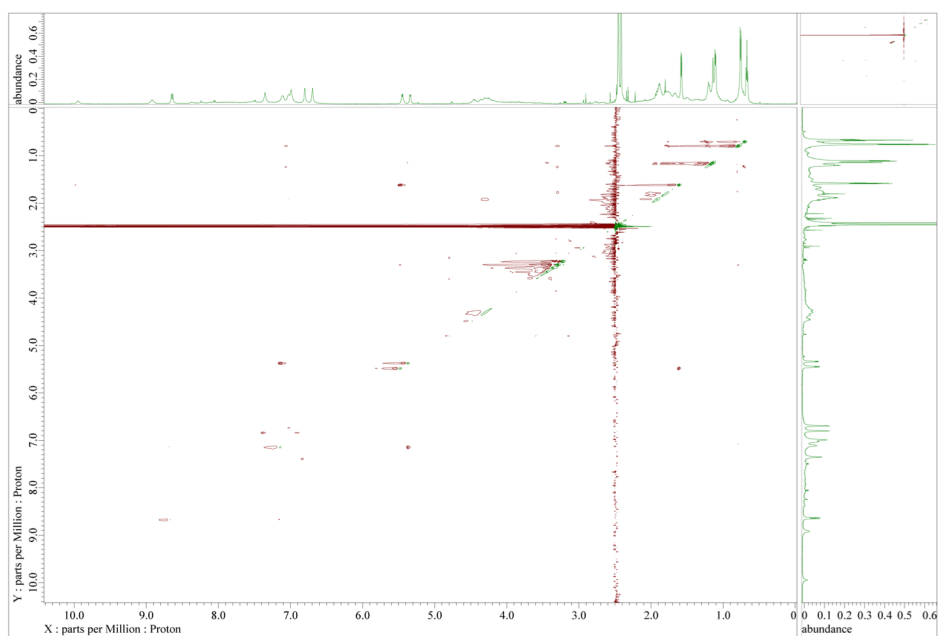
Supplementary Fig. 36. COSY spectrum of goadpeptin B (10) (600 MHz, in DMSO-*d*₆)



Supplementary Fig. 37. HMQC spectrum of goadpeptin B (10) (600 MHz, in DMSO- d_6)



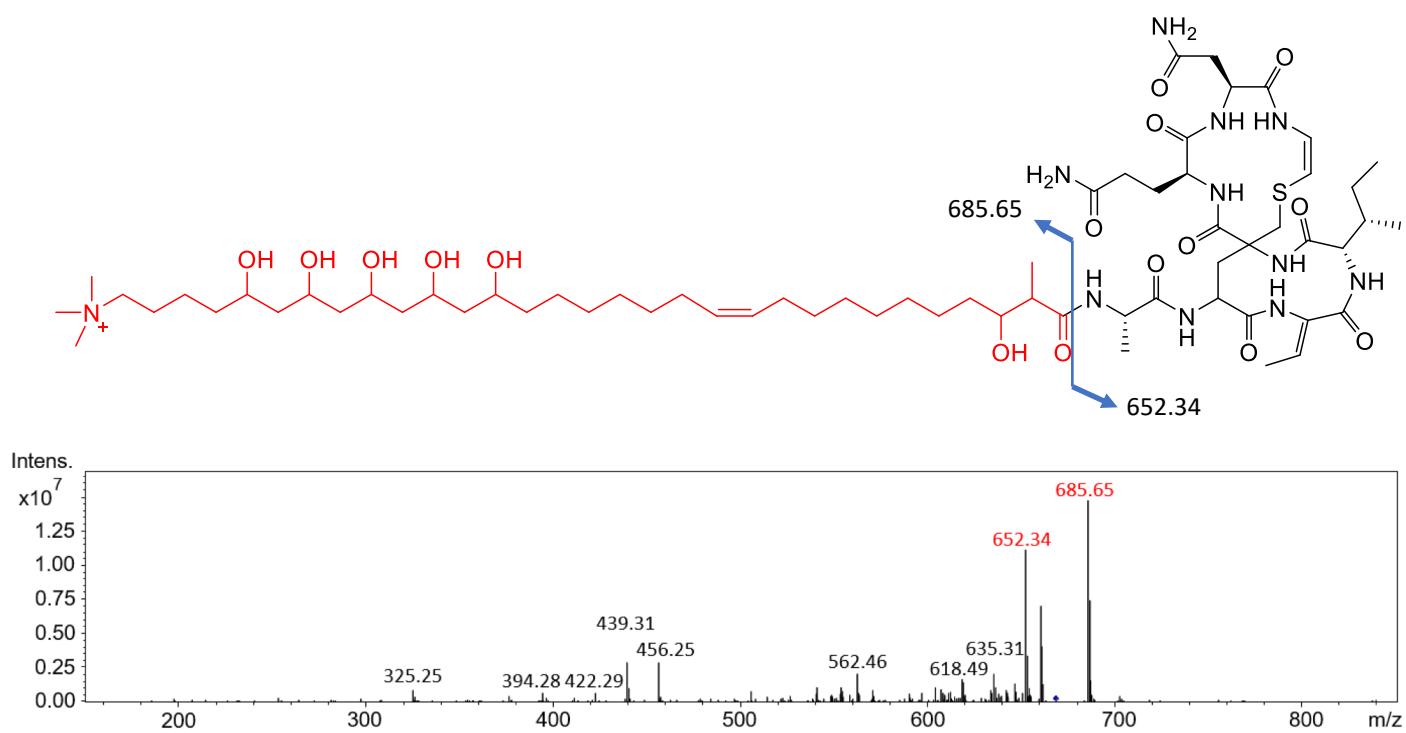
Supplementary Fig. 38. HMBC spectrum of goadpeptin B (10) (600 MHz, in DMSO-*d*₆)



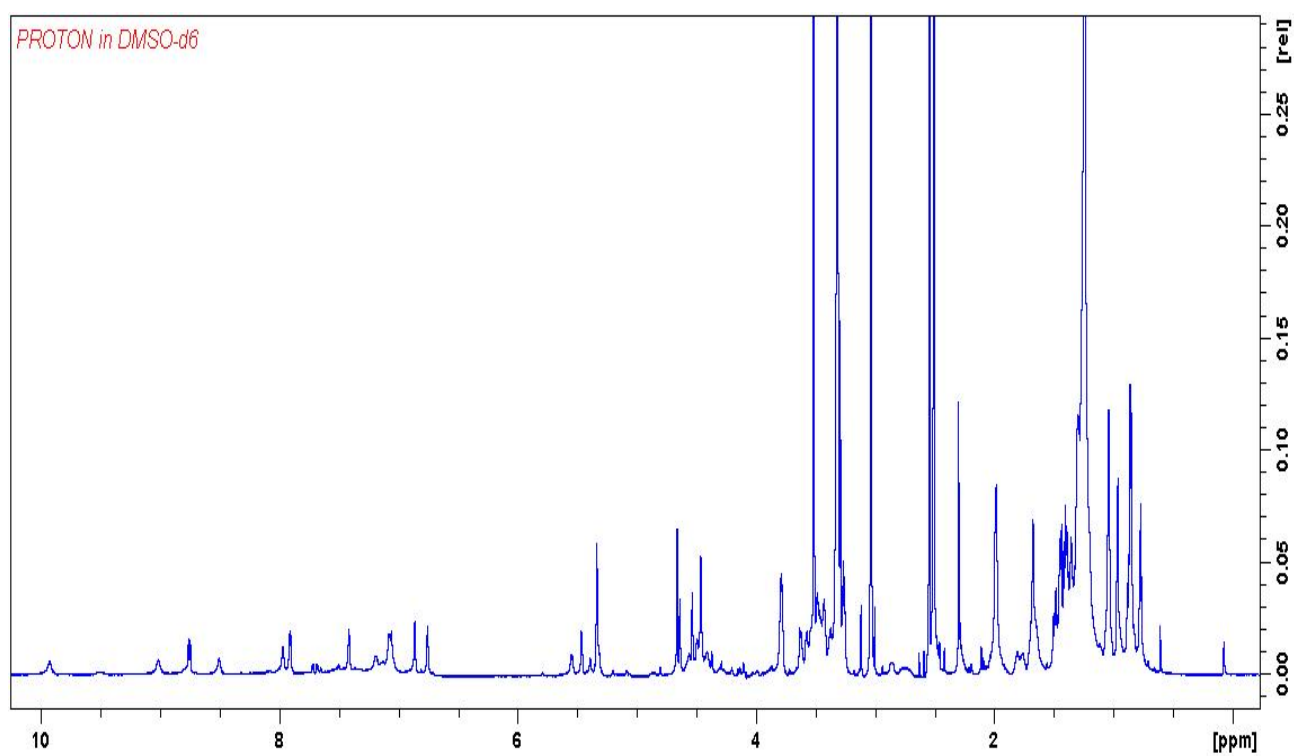
Supplementary Fig. 39. ROESY spectrum of goadpeptin B (10) (600 MHz, in DMSO-*d*₆)

Supplementary Table 10. NMR data of goadpeptin B (10)

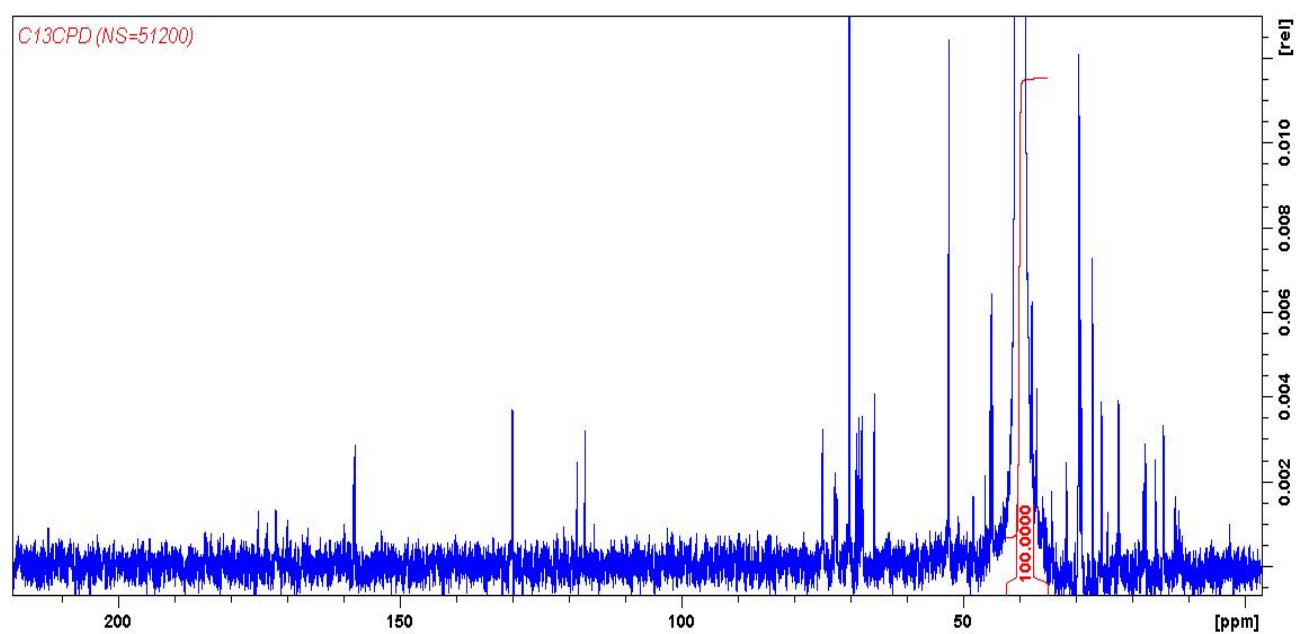
		δ_H , mult (J in Hz)	δ_C , type of carbon
Ala1	1 (NH ₂)	-	-
	2	3.45, m	49.7, CH
	3	-	172.8, C
	4	1.15, d (6.5)	20.3, CH ₃
Dha2	5 (NH)	8.41, brs	-
	6	4.49, m	48.9, CH
	7	-	168.8, C ^a
	8	3.15, m 2.04, m	31.2, CH ₂
Dhb3	9 (NH)	9.98, brs	-
	10	-	133.0, C
	11	-	168.9, C
	12	5.49, m	121.4, CH
	13	1.62, d (7.0)	11.8, CH ₃
Ile4	14 (NH)	7.08, brs	-
	15	3.24, m	63.0, CH
	16	-	168.7, C
	17	1.79, m	36.6, CH
	18	1.24, m	23.9, CH ₂
	19	0.71, t (6.5)	11.2, CH ₃
	20	0.79, d (6.5)	15.4, CH ₃
Dha5	21 (NH)	7.41, brs	-
	22	-	64.1, C
	23	-	171.6, C ^a
	24	3.25, m 2.82, m	35.8, CH ₂
	25 (NH)	7.55, m	-
Gln6	26	4.29, m	50.3, CH
	27	-	169.7, C
	28	1.70, m 1.56, m	21.9, CH ₂
	29	1.94, m	31.0, CH ₂
	30	-	173.6, C
	31 (NH ₂)	7.03, s 6.74, s	-
	32 (NH)	8.96, brs	-
Asn7	33	4.37, m	52.6, CH
	34	-	170.6, C
	35	2.83, m 2.42, m	35.3, CH ₂
	36	-	169.7, C
	37 (NH ₂)	7.38, s 6.85, s	-
	38 (NH)	8.71, d (11.0)	-
AviCys8	39	7.15, m	132.1, CH
	40	5.58, d (6.5)	98.3, CH
^a interchangeable			



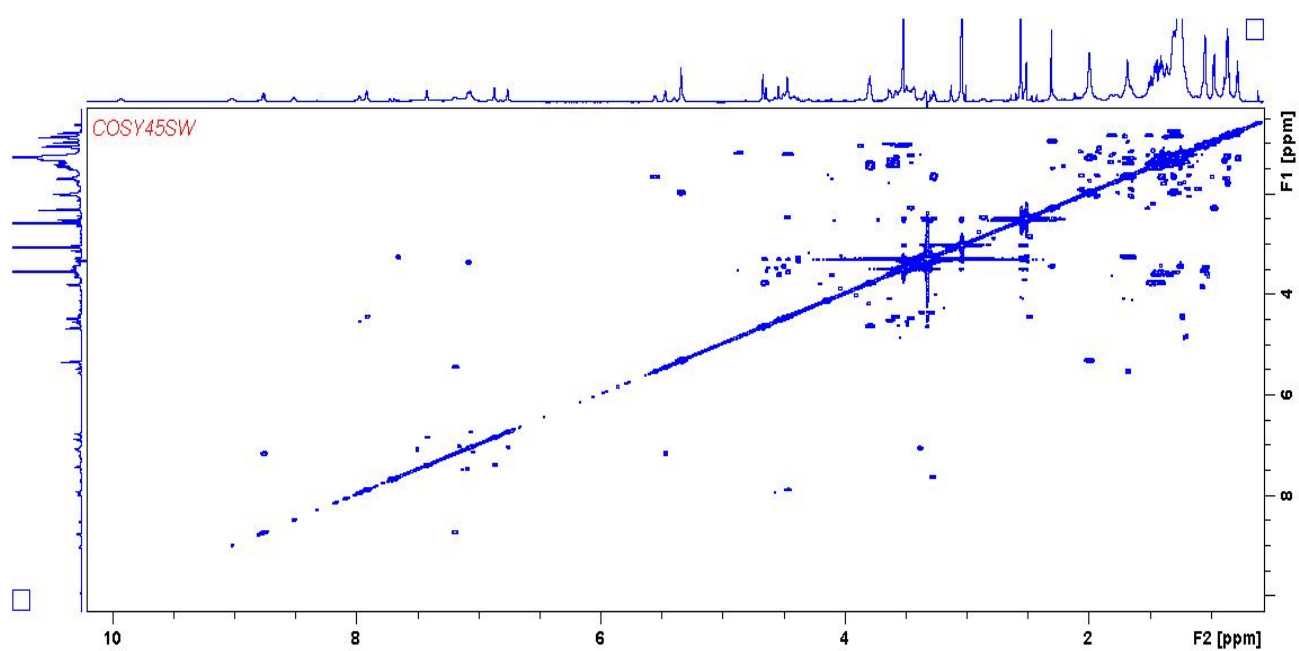
Supplementary Fig. 40. LR-ESI-MS/MS spectrum of 6 generated from the molecular ion peak (m/z 668.9 $[M+H]^{2+}$)



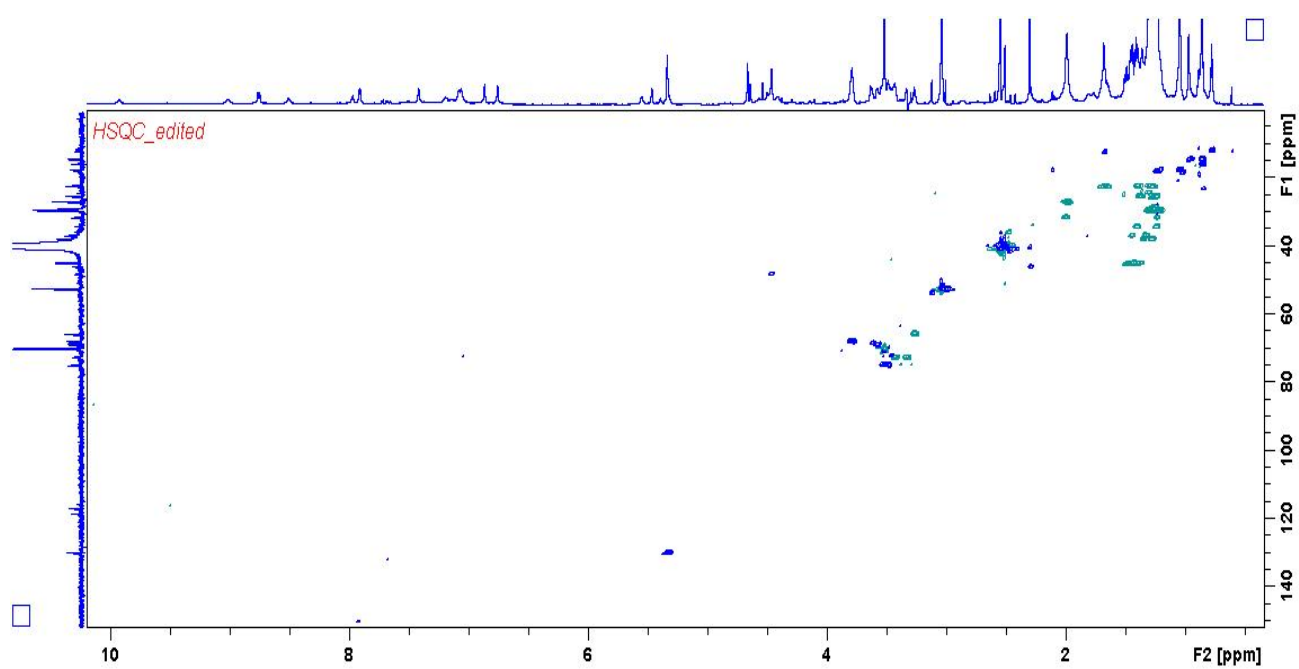
Supplementary Fig. 41. ^1H NMR spectrum of goadvionin B2 (6) (800 MHz, in DMSO- d_6)



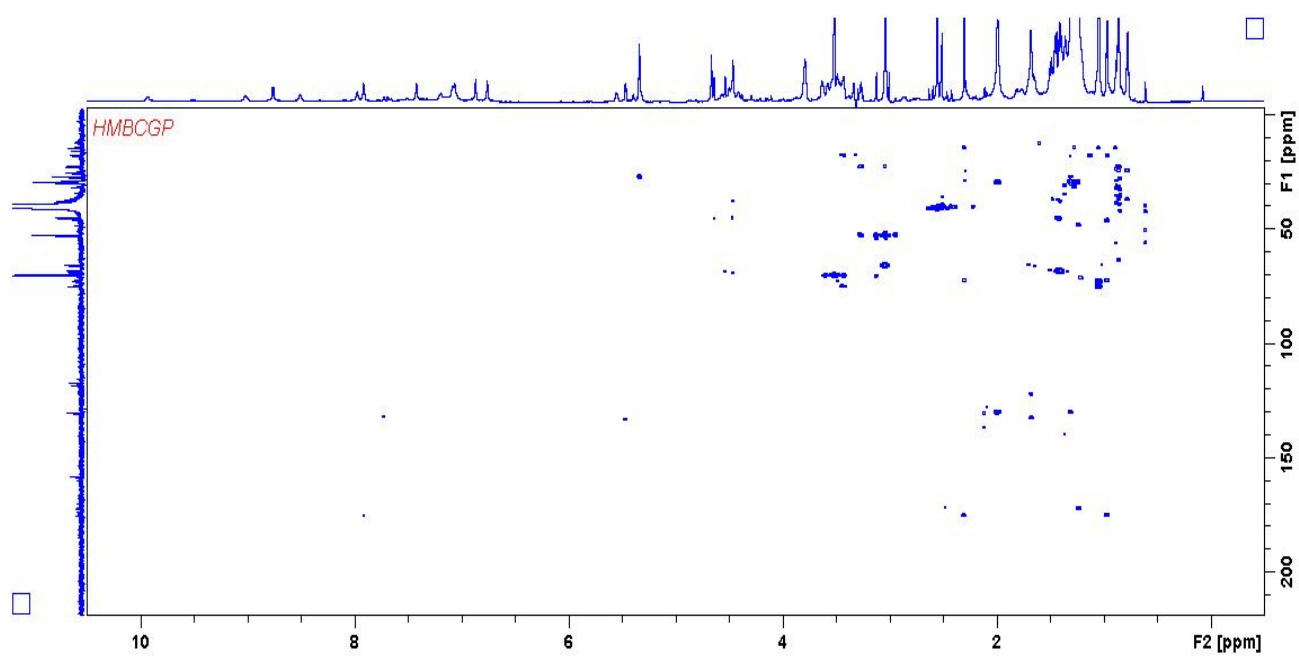
Supplementary Fig. 42. ^{13}C NMR spectrum of goadvionin B2 (6) (200 MHz, in $\text{DMSO-}d_6$)



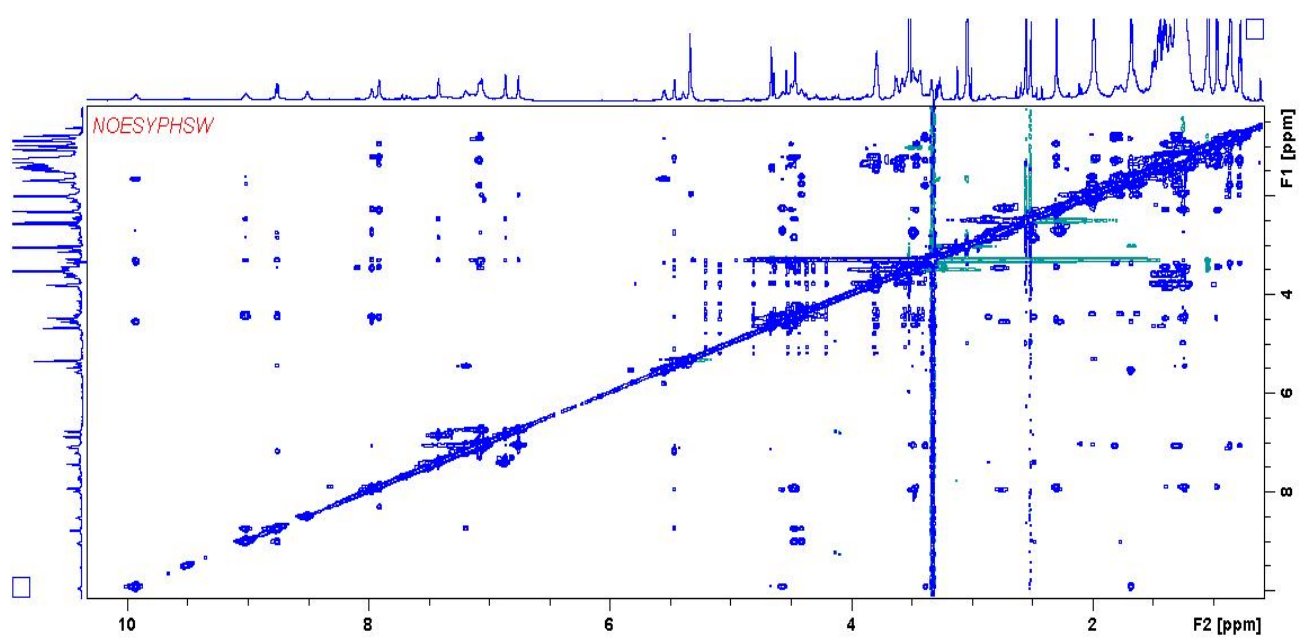
Supplementary Fig. 43. COSY spectrum of goadvionin B2 (6) (800 MHz, in DMSO-*d*₆)



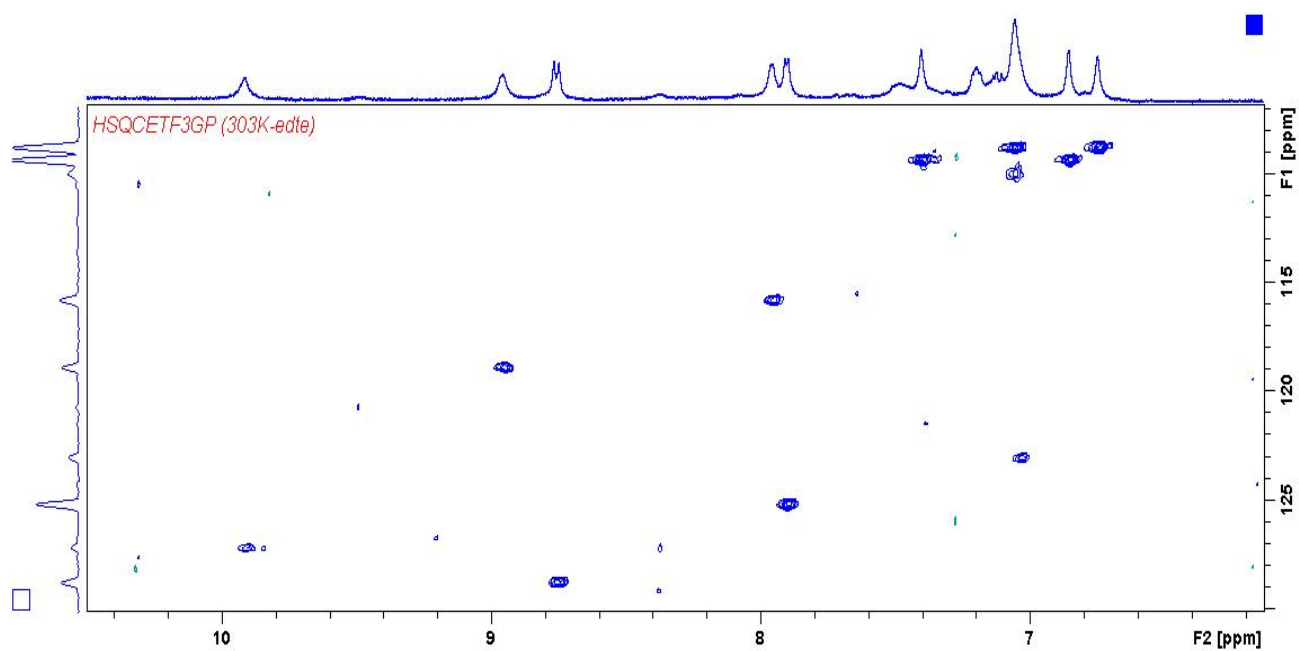
Supplementary Fig. 44. Edited-HSQC spectrum of goadvionin B2 (6) (800 MHz, in DMSO-*d*₆)



Supplementary Fig. 45. HMBC spectrum of goadvionin B2 (6) (800 MHz, in DMSO-*d*₆)



Supplementary Fig. 46. NOESY spectrum of goadvionin B2 (6) (800 MHz, in DMSO-*d*₆)



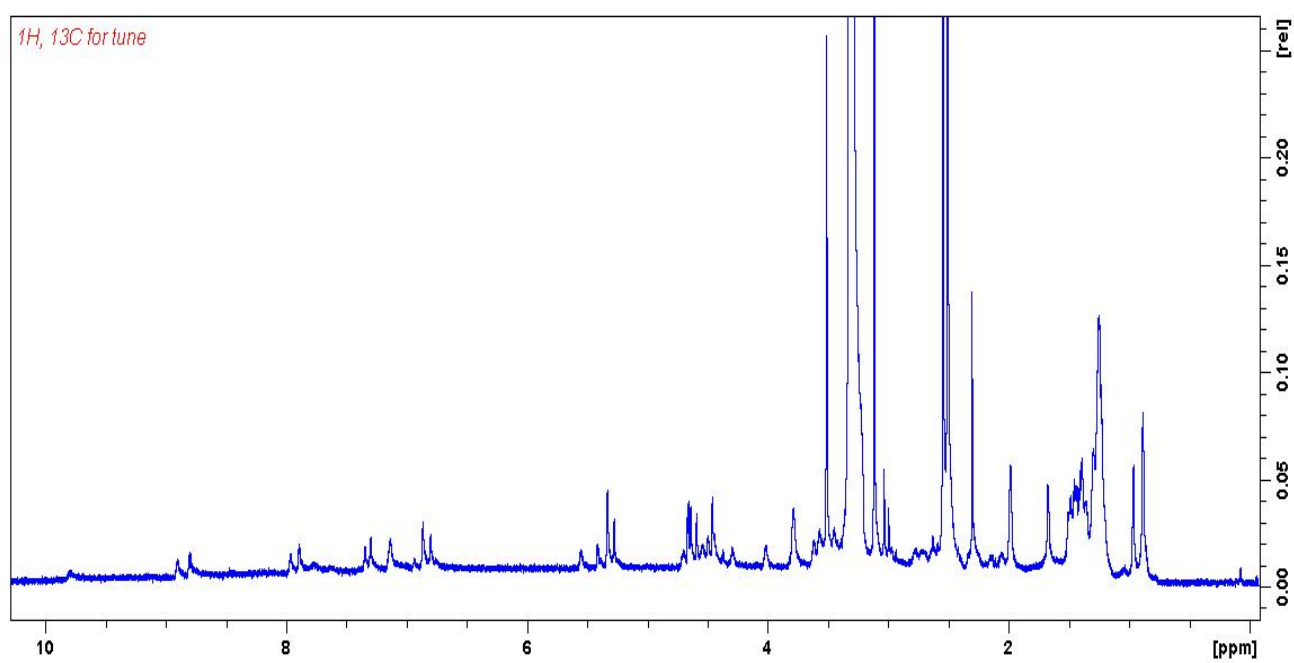
Supplementary Fig. 47. ^1H - ^{15}N HSQC spectrum of goadvionin B2 (6) (600 MHz, in $\text{DMSO}-d_6$)

Supplementary Table 11. NMR data of goadvionin B2 (6, RiPP moiety)

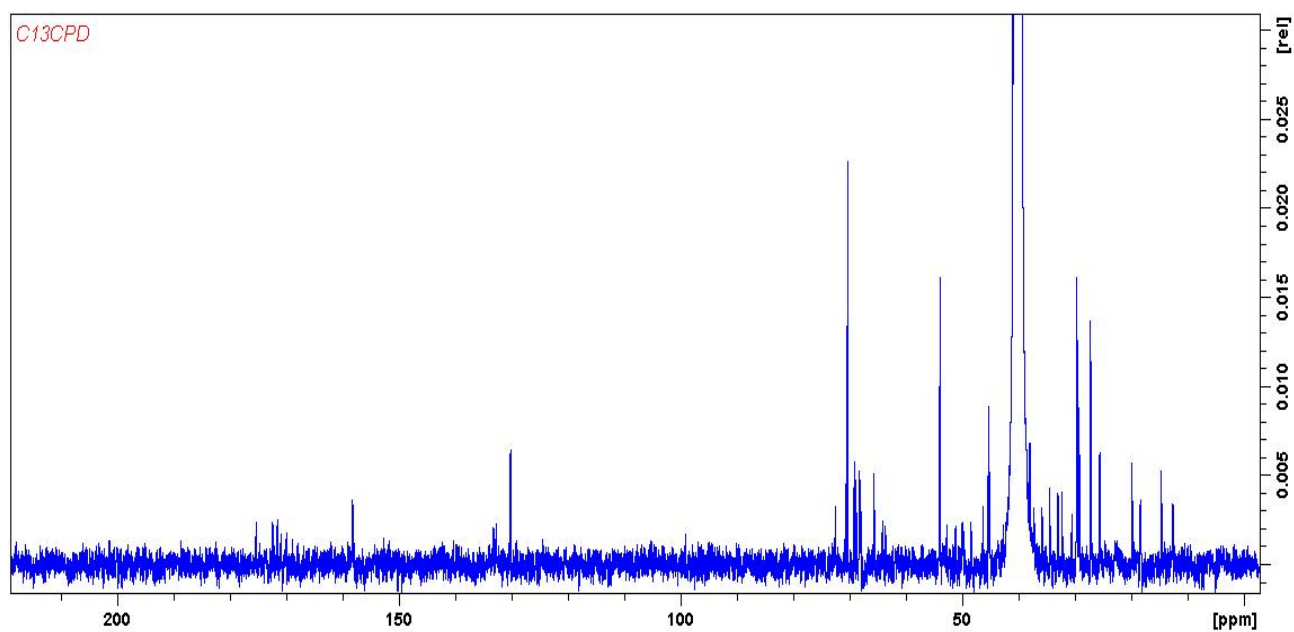
		δ_H , mult (J in Hz)	δ_C , type of carbon	δ_N
Ala1	1 (NH)	7.91, d (8.5)	-	125.1
	2	4.46, m	48.3, CH	-
	3	-	172.2, C	-
	4	1.23, m	18.1, CH ₃	-
Dha2	5 (NH)	7.97, brs	-	115.8
	6	4.57, m	49.8, CH	-
	7	-	-	-
	8	2.76, m 2.28, m	-	-
Dhb3	9 (NH)	9.93, brs	-	127.1
	10	-	132.6, C	-
	11	-	-	-
	12	5.55, m	122.2, CH	-
	13	1.68, d (7.0)	12.5, CH ₃	-
Ile4	14 (NH)	7.08, d (6.0)	-	111.0
	15	3.38, m	63.4, CH	-
	16	-	-	-
	17	1.81, m	37.1, CH	-
	18	1.31, m	24.4, CH ₂	-
	19	0.77, t (7.0)	11.8, CH ₃	-
	20	0.86, d (7.0)	14.4, CH ₃	-
Dha5	21 (NH)	7.03, m	-	122.9
	22	-	-	-
	23	-	-	-
	24	-	-	-
Gln6	25 (NH)	7.53, m	-	118.0
	26	4.41, m	53.1, CH	-
	27	-	-	-
	28	1.77, m 1.63, m	27.9, CH ₂	-
	29	2.00, m	27.1, CH ₂	-
	30	-	-	-
	31 (NH ₂)	7.06, s 6.76, s	-	108.7
Asn7	32 (NH)	9.02, brs	-	118.8
	33	4.46, m	50.9, CH	-
	34	-	-	-
	35	2.87, m 2.48, m	35.9, CH ₂	-
	36	-	-	-
	37 (NH ₂)	7.42, s 6.86, s	-	109.3
AviCys 8	38 (NH)	8.76, d (11.5)	-	128.8
	39	7.20, m	133.3, CH	-
	40	5.46, d (7.0)	99.4, CH	-

Supplementary Table 12. NMR data of goadvionin B2 (6, RiPP moiety)

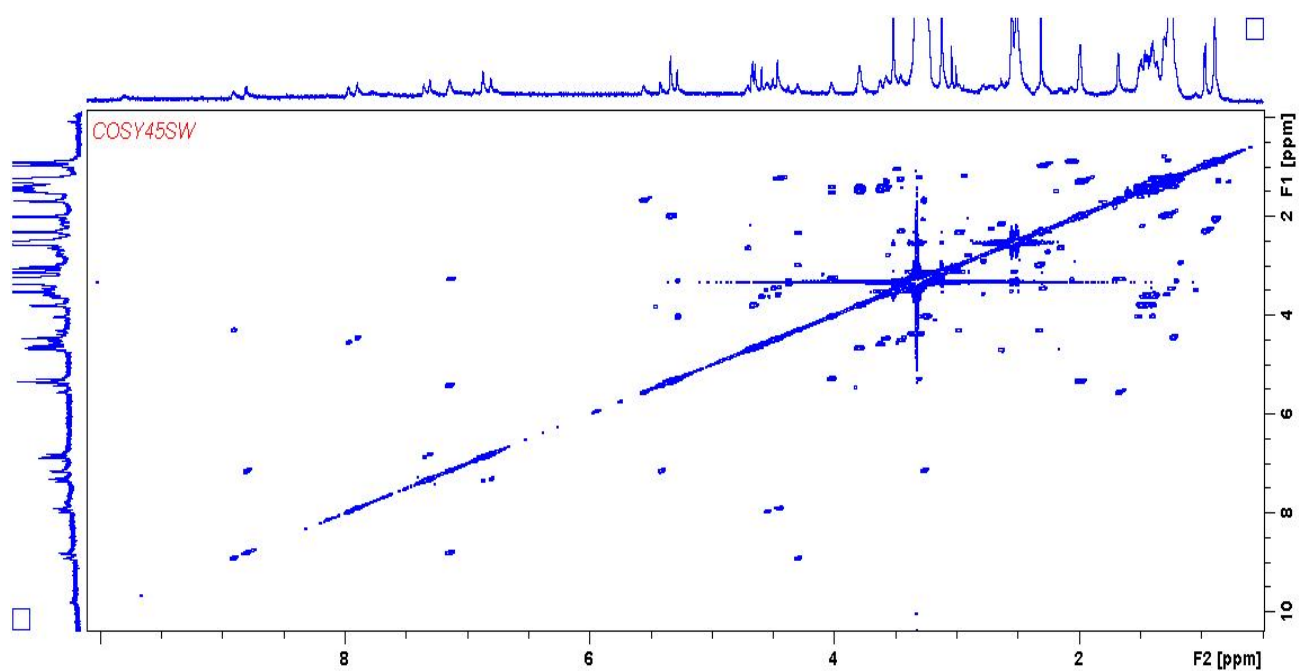
	δ_H , mult (J in Hz)	δ_C , type of carbon
1'	-	175.2, C
2'	2.30, m	46.2, CH
3'	3.46, m	72.4, CH
4'	1.40, m 1.25, m	34.4, CH ₂
5'	1.20-1.33, m	29.1-29.7, CH ₂
6'	1.20-1.33, m	29.1-29.7, CH ₂
7'	1.20-1.33, m	29.1-29.7, CH ₂
8'	1.20-1.33, m	29.1-29.7, CH ₂
9'	1.20-1.33, m	29.1-29.7, CH ₂
10'	1.99, m	27.1, CH ₂
11'	5.33, m	130.1, CH
12'	5.33, m	130.1, CH
13'	1.99, m	27.1, CH ₂
14'	1.20-1.33, m	29.1-29.7, CH ₂
15'	1.20-1.33, m	29.1-29.7, CH ₂
16'	1.20-1.33, m	29.1-29.7, CH ₂
17'	1.20-1.33, m	29.1-29.7, CH ₂
18'	1.36, m 1.28, m	37.8, CH ₂
19'	3.58, m	69.1, CH
20'	1.40-1.50, m	45.0, CH ₂
21'	3.77-3.81, m	67.9, CH ^b
22'	1.40-1.50, m	45.2, CH ₂
23'	3.77-3.81, m	68.0, CH ^b
24'	1.40-1.50, m	45.2, CH ₂
25'	3.77-3.81, m	68.2, CH ^b
26'	1.40-1.50, m	45.0, CH ₂
27'	3.63, m	68.6, CH
28'	1.45, m 1.33, m	37.0, CH ₂
29'	1.40, m 1.30, m	22.4, CH ₂
30'	1.70, m 1.65, m	22.6, CH ₂
31'	3.27, m	65.9, CH ₂
32'	0.97, d (7.0)	14.6, CH ₃
33'	3.04, s	52.7, CH ₃
3'-OH	4.50, m	-
19'-OH	4.46, d (4.5)	-
21'-OH	4.64, d (4.0) ^a	-
23'-OH	4.66, d (4.0) ^a	-
25'-OH	4.66, d (4.0) ^a	-
27'-OH	4.54, d (5.0)	-
^{a,b} interchangeable		



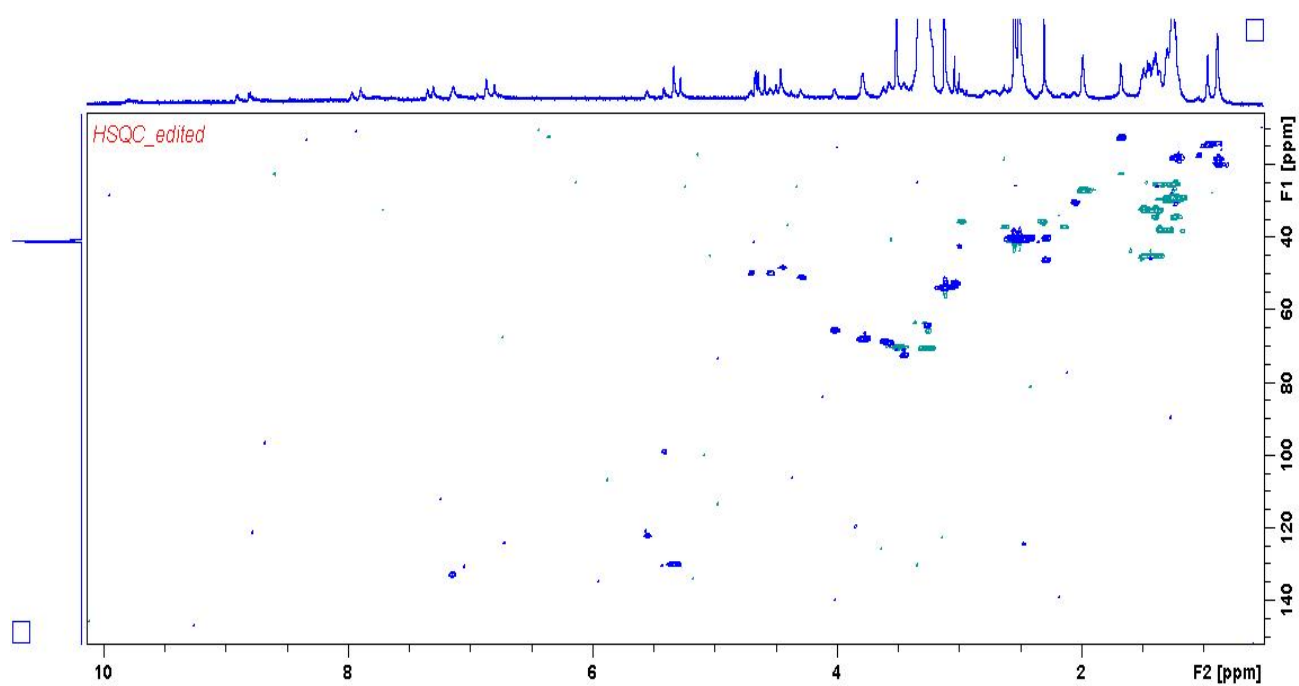
Supplementary Fig. 48. ¹H NMR spectrum of goadvionin A1 (1) (800 MHz, in DMSO-*d*₆)



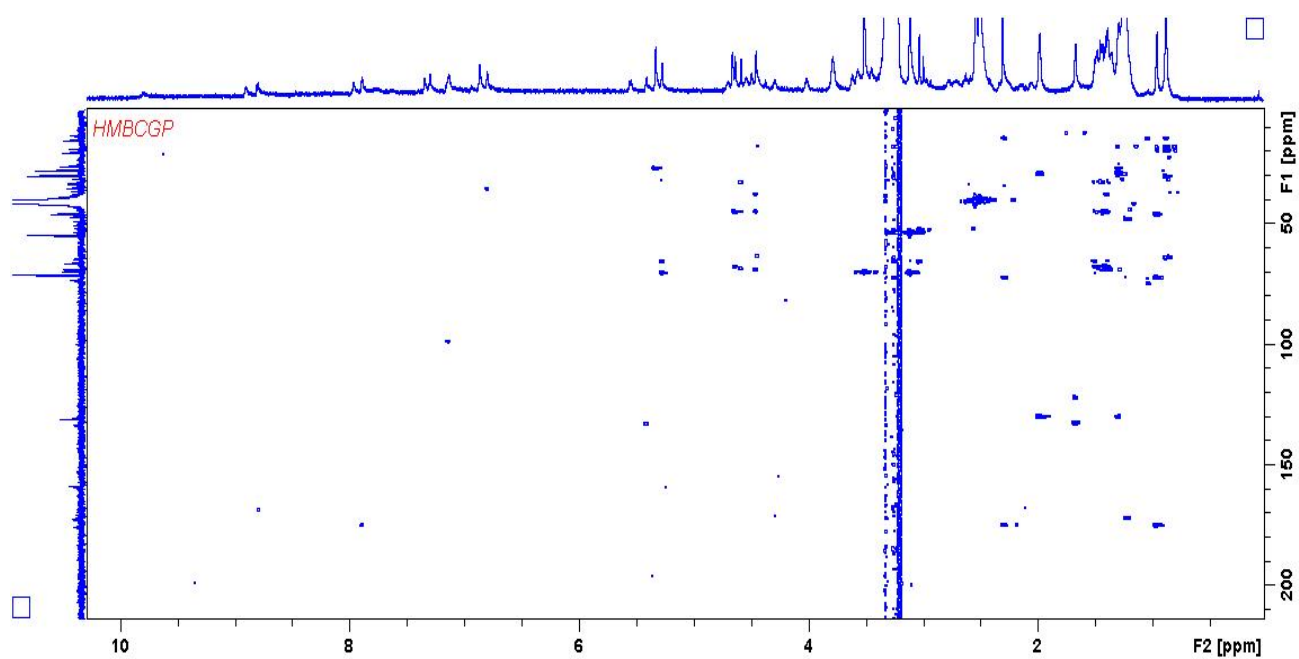
Supplementary Fig. 49. ^{13}C NMR spectrum of goadvionin A1 (1) (200 MHz, in $\text{DMSO}-d_6$)



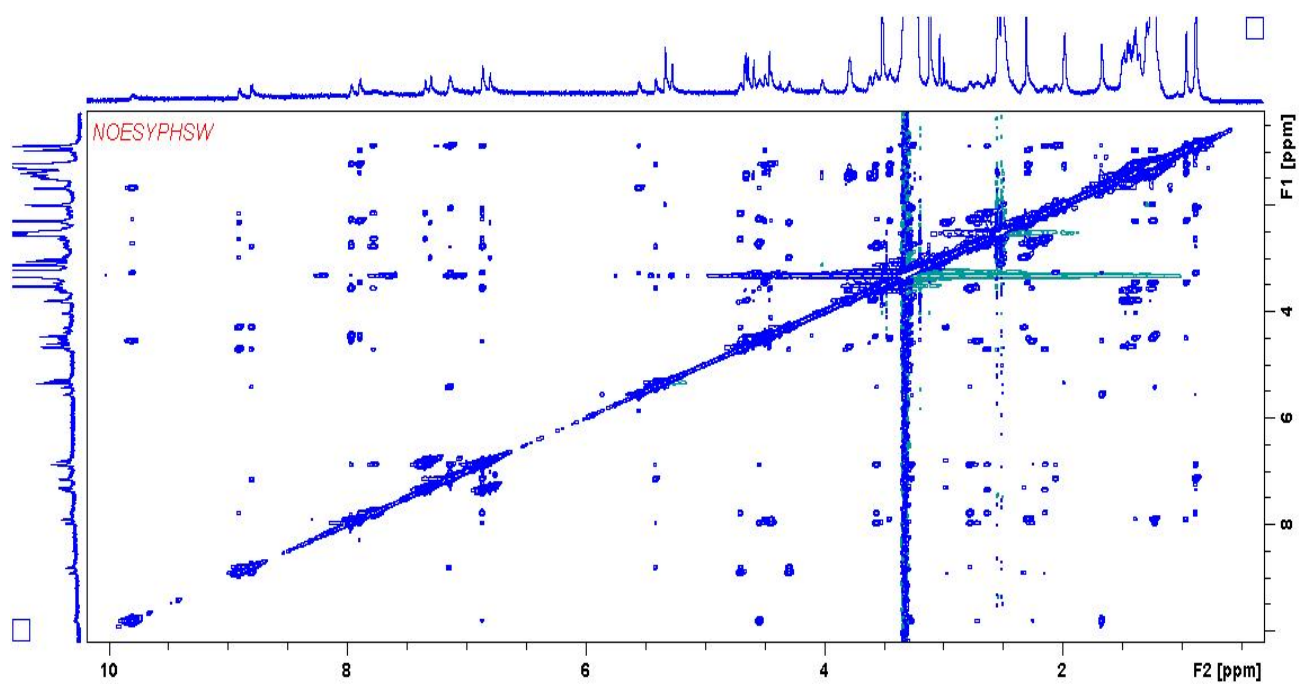
Supplementary Fig. 50. COSY spectrum of goadvionin A1 (1) (800 MHz, in DMSO- d_6)



Supplementary Fig. 51. Edited-HSQC spectrum of goadvionin A1 (1) (800 MHz, in DMSO- d_6)



Supplementary Fig. 52. HMBC spectrum of goadvionin A1 (1) (800 MHz, in DMSO-*d*₆)



Supplementary Fig. 53. NOESY spectrum of goadvionin A1 (1) (800 MHz, in DMSO-*d*₆)

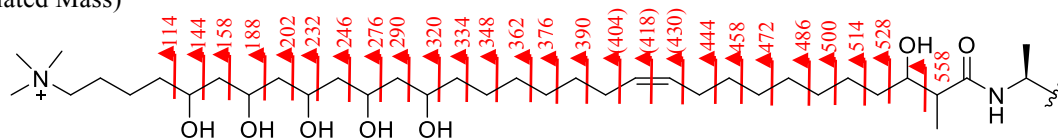
Supplementary Table 13. NMR data of goadvionin A1 (1, RiPP moiety)

		δ_H , mult (J in Hz)	δ_C , type of carbon
Ala1	1 (NH)	7.90, brs	-
	2	4.45, m	48.3, CH
	3	-	172.2, C
	4	1.22, d (7.0)	18.1, CH ₃
Dha2	5 (NH)	7.97, brs	-
	6	4.55, m	49.8, CH
	7	-	169.0-173.0, C
	8	2.72, m 2.27, m	n.d.
Dhb3	9 (NH)	9.80, brs	-
	10	-	132.7, C
	11	-	169.0-173.0, C
	12	5.56, m	122.2, CH
	13	1.67, d (7.0)	12.5, CH ₃
Val4	14 (NH)	7.13, m	-
	15	3.27, m	64.1, CH
	16	-	169.0-173.0, C
	17	2.06, m	30.4, CH
	18	0.88, m	19.8, CH ₃
	19	0.89, m	18.3, CH ₃
Dha5	20 (NH)	6.87, s	-
	21	-	n.d.
	22	-	169.0-173.0, C
	23	3.56, m 2.78, m	40.5, CH ₂
Asn6	24 (NH)	7.78, d (7.0)	-
	25	4.70, m	49.7, CH
	26	-	169.0-173.0, C
	27	2.63, m 2.15, m	37.2, CH ₂
	28	-	169.0-173.0, C
	29 (NH ₂)	7.35, s 6.87, s	-
Asn7	30 (NH)	8.91, brs	-
	31	4.29, m	51.1, CH
	32	-	-
	33	2.98, m 2.32, m	35.7, CH ₂
	34	-	169.0-173.0, C
	35 (NH ₂)	7.30, s 6.80, s	-
AviCys8	36 (NH)	8.81, d (11.5)	-
	37	7.15, m	133.1, CH
	38	5.42, d (7.0)	99.1, CH

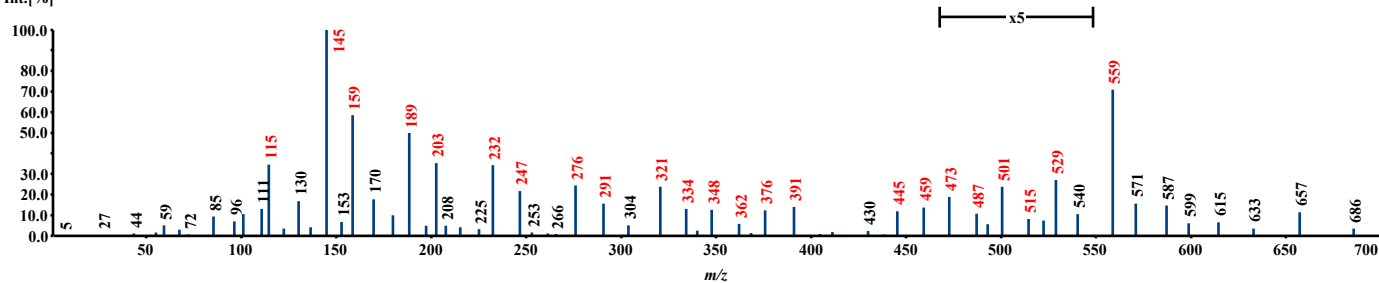
Supplementary Table 14. NMR data of goadvionin A1 (1, RiPP moiety)

	δ_H , mult (J in Hz)	δ_C , type of carbon
1'	-	175.3, C
2'	2.29, m	46.2, CH
3'	3.46, m	72.4, CH
4'	1.39, m 1.25, m	34.3, CH ₂
5'	1.20-1.33, m	29.2-29.7, CH ₂
6'	1.20-1.33, m	29.2-29.7, CH ₂
7'	1.20-1.33, m	29.2-29.7, CH ₂
8'	1.20-1.33, m	29.2-29.7, CH ₂
9'	1.20-1.33, m	29.2-29.7, CH ₂
10'	1.99, m	27.1, CH ₂
11'	5.33, m	130.1, CH ^b
12'	5.33, m	130.2, CH ^b
13'	1.99, m	27.1, CH ₂
14'	1.20-1.33, m	29.2-29.7, CH ₂
15'	1.20-1.33, m	29.2-29.7, CH ₂
16'	1.20-1.33, m	29.2-29.7, CH ₂
17'	1.20-1.33, m	29.2-29.7, CH ₂
18'	1.36, m 1.29, m	37.9, CH ₂
19'	3.58, m	69.1, CH
20'	1.40-1.50, m	45.0, CH ₂
21'	3.77-3.80, m	67.9, CH ^c
22'	1.40-1.50, m	45.2, CH ₂
23'	3.77-3.80, m	68.1, CH ^c
24'	1.40-1.50, m	45.2, CH ₂
25'	3.77-3.80, m	68.2, CH ^c
26'	1.40-1.50, m	45.0, CH ₂
27'	3.62, m	68.7, CH
28'	1.48, m 1.39, m	32.9, CH ₂
29'	1.51, m 1.39, m	32.2, CH ₂
30'	4.02, m	65.6, CH
31'	3.31, m 3.25, m	70.5, CH ₂
32'	0.96, d (7.0)	14.6, CH ₃
33'	3.12, s	53.9, CH ₃
3'-OH	4.50, m	-
19'-OH	4.46, d (4.5)	-
21'-OH	4.64, d (4.0) ^a	-
23'-OH	4.66, d (4.0) ^a	-
25'-OH	4.67, d (4.0) ^a	-
27'-OH	4.59, d (4.5)	-
30'-OH	5.28, d (5.5)	-
^{a-c} interchangeable		

(Calculated Mass)



"MS1 (ID297388)"
Int. [%]



Supplementary Fig. 54. HE-CID-MS/MS spectrum of goadvionin B2 (m/z 404, 418, and 430 were not observed)

Name	Sequence	Purpose
GSlas-LanKC-F	ACTTCTCCAACGGAGGCGGTGTCT	Screening for <i>gdv</i> cluster from cosmid library.
GSlas-LanKC-R	CACAGAACAGCAGCGGATGGTACG	
Dup_F	ACGACGGTGGCATTGCCAGCGCCCCGCCCGCGATTCCGC GGCGAGTTTATAAAGCAAAAGGGGATGATAAGT	Disruption of UP region
Dup_R	CTCATAACTCAGGCACAGGGGCTCCGCCACGGCGAGCCCCT GTTTTTTATTATAATCAGCCAATCGACTGGCGAG	
Ddown_F	TCCCCCGCCGCCGCTGCTGCCGTCCAGCGCGGCCAGCCGCC CCGTCGAATGCATAGCAAAAGGGGATGATAAGT	Disruption of DOWN region
Ddown_R	TAACCGCGGCGCGACGGTCGCGAGGTGGTCGTCCCCGCACGT GCGGGGGTATGCATTAGCCAATCGACTGGCGAG	
Dorf14_F	CGCTTAAAAAACAGGGGCTCCGCCGTGGCGAGCCCCTGTGC CTGAGTTAACTAGTAGCAAAAGGGGATGATAAGT	Disruption of <i>orf14</i> gene
Dorf14_R	TCATCGTCACCGCCCCGCTGTGGGCCGTGTCGCGGCGGCCGCA CTGGATGACTAGTTCAGCCAATCGACTGGCGAG	
DgdvC_F	TGCTGGCGGCGCGTCTGGGCACCCGGCCCTGAGCGCGCCAGG GACGGTCAACTAGTAGCAAAAGGGGATGATAAGT	Disruption of <i>gdvC</i> gene
DgdvC_R	AGCCTGCAAATCGTCGAAAACCGACCCGCGCAGGAAAGAGCG GCTTCATGACTAGTTCAGCCAATCGACTGGCGAG	
DgdvD_F	TCCGATGGCCTTTGTTCTTTGTTGCGGCGACCTATTAGAAAGGT CTCATGCAGTGAGTTCGAGCGACTCGAAGT	Disruption of <i>gdvD</i> gene
DgdvD_R	CGGCGGCATGTGTTGGAGCACTCGGGCACGTGCTGTGGGACG AGCCCTCACTGGTACCGAGCGAACGCGTT	
DgdvKC_F	GCCGTCTGTCCCCCTATTGCGCATCCTGTTGGAATTAGGTACG AGGATGCAGTGAGTTCGAGCGACTCGAAGT	Disruption of <i>gdvKC</i> gene
DgdvKC_R	CCTTTCTCGTCTGTCGGTCGACCGACCGGTTGGTGGGGTCA CATTTCACTGGTACCGAGCGAACGCGTT	
DgdvA_F	CCAACCGGTGCGGTGACCGACGAGCGAGGAAAGGAGGTGA TTTCGTATGACTAGTAGCAAAAGGGGATGATAAGT	Disruption of <i>gdvA</i> gene
DgdvA_R	GCGCCGCAAGTCCCATCGGCGGCTCGGCGAGGCCGCCGGTGG CCCGTTCAACTAGTTCAGCCAATCGACTGGCGAG	
DgdvI_F	CTTCCCGACCAACCACGACCAGATGCCCGAGGAGCAGCGAT CCCCATGACTAGTAGCAAAAGGGGATGATAAGT	Disruption of <i>gdvI</i> gene
DgdvI_R	GCATGCGTGGGGGCGGCGGCGGGCGGTGCGCCGCGCCGC GACGGTTCAACTAGTTCAGCCAATCGACTGGCGAG	
DgdvFB_F	GCATGCCCCGACAGGGCGGCGCCTCACCAGCACCGAAGGAC TTCCCATGACTAGTAGCAAAAGGGGATGATAAGT	Disruption of <i>gdvFB</i> gene
DgdvFB_R	CGAGACGAAGTCTCCGCGCATCAGCCCTCCAGCGGCGGACG GCCAGCACACTAGTTCAGCCAATCGACTGGCGAG	
DgdvG_F	TTCGGCGGCCACAACGCAGTGCTGGCCGTCCGCCGCTGGGAG GGCTGATGACTAGTAGCAAAAGGGGATGATAAGT	Disruption of <i>gdvG</i> gene
DgdvG_R	TTCCGGTCCTGAGAGTGGCATAGCTGGATCTTTCTCGCGCGAG CGGCTCAACTAGTTCAGCCAATCGACTGGCGAG	
DgdvL_F	CGCAATCTGTTGAGCGCTGAGCCGCTCGCGGAGAAAGATCC AGCTATGACTAGTAGCAAAAGGGGATGATAAGT	Disruption of <i>gdvL</i> gene
DgdvL_R	GTCCATGAGGGTCACCATTCCAAGTCGTACGGTGCAAAGTGC CTTCGGCACTAGTTCAGCCAATCGACTGGCGAG	
DgdvMT_F	CGCACCGAAGGGCCGGGAGGCCGAGGACGCAACGTGAGGA ATACCTATGCAGTGAGTTCGAGCGACTCGAAGT	Disruption of <i>gdvMT</i> gene
DgdvMT_R	CCAGACAGGAGCGGCTCATGTGTCATGTCCCTTCTCGGTCGCGGC GCGCTCACTGGTACCGAGCGAACGCGTT	
DgdvH_F	TCACCGTCGTCTAGGGCCGGATCCGTACGCGTAGTTCACCGC TGCGCAGACTAGTAGCAAAAGGGGATGATAAGT	Disruption of <i>gdvH</i> gene
DgdvH_R	ACGCGGCGGTCGTCCGGCAGTCGGATGCGCGGCGCGGAACCG GCGTGATGACTAGTTCAGCCAATCGACTGGCGAG	

Name	Sequence	Purpose
DgdvJ_F	TGTCCCTGCCATCCCAGCCGGCCGCGCTGCCGACGCCCCGAC	Disruption of <i>gdvJ</i> gene
DgdvJ_R	GACCATGACTAGTAGCAAAAGGGGATGATAAGT	
DgdvY_F	CGGGCCCGCCGGGTGCGGGGCGACGCCGGCGAGCGGACGGT	Disruption of <i>gdvY</i> gene
DgdvY_R	CGGTGATCAACTAGTTCAGCCAATCGACTGGCGAG	
pColdI_cACP-GdvPKS_F	TCGACCCGCTCCCGACCGATCTCCGCCCGGCCGCGCCGCACGC	Heterologous expression of cACP-GdvPKS.
pColdI_cACP-GdvPKS_R	GCCCCTGACTAGTAGCAAAAGGGGATGATAAGT	
pET16b_GdvG_F	GCCGCGGTTACGCGCCCCGCGTCGGCGGCGCCACTAATTCGC	Heterologous expression of GdvG.
pET16b_GdvG_R	TCCGTCAACTAGTTCAGCCAATCGACTGGCGAG	
pET16b_nACP-GdvPKS_F	CATCATATCGAAGGTAGGCATATGCCGACGCCGCTCCTGG	Heterologous expression of nACP-GdvPKS.
pET16b_nACP-GdvPKS_R	GTCGACAAGCTTGAATTCGGATCCTCAGTCGGCGGTGGTG	
pColdI_GdvACP_F	TTCATATGCGCGGAGACTTCGTCTCGCTC	Heterologous expression of GdvACP.
pColdI_GdvACP_R	TTGGATCCTCAGCGCTCGAACAGATTGCGCAG	
pET26b_NocB1_F	CCATATCGAAGGTCGTCATATGGTGACCTCATGGACGAG	Heterologous expression of NocB1.
pET26b_NocB1_R	GGGCTTTGTTAGCAGCCGGATCCTCACGCCGTTCCACCC	
pET26b_NocB2_F	ATCATATCGAAGGTAGGCATATGTTGAAGTACCATGAGAAAAA	Heterologous expression of NocB2.
pET26b_NocB2_R	AAGGGG	

Supplementary table 16. Optimized expression gene sequences used for this study.

Name	Sequence
cACP-AprPKS	CGTACCCCACTGCTGGCGGAACTGGCCGGTGAAGGTGCCCCAACCCGTGCCGCCACAGTCCAGATCCAGTTCA GCATCGCGAACTGCTGCGCGAACACGTTTCGACCGTTATCGCCGGTGCGATGGGCATCACCCAGAAGCGCTG AATCCGCGCGCCAATCTGTTTTCGCTGGGTGTGGATAGTCTGATGAGTCTGAGCATCATTCGTACGCTGGCCGC CACCACGGGTATCATATCCCGCCGACCGTTCTGCGTGAACGTCCAACGGCGCTGGCCCTCGCGGATCATCTGG CCGAAGCCGAACAGACCACCGAAGGCGAACAGCGCTAA
AprG	CTCCAAGGCGAACTGGTGAGTCTGCACCCGGCGAGCACGGATGAGGATTATGAACTGCTGGCGGCGTGGGCG AGTAGTACCACCAGCGTTTACAGTAGCGGCCAAGTTCGTCCAGTTACCCAGCGCAACTGCGTGAAGCCGGCG AACTGCATGGCATGCGCTATCTGGTGGTTCGTGACCGCGAAGGCGAACCAGTTGGCGGCGTTAACTGGCGCCC GGTGACCTATGAGGGTCACTTCAAGTTGGCAACGCCGTTGGTCCGCCACAGCTGTGGGATCTGGGCTATGGC GTTGAGAGCGTGCTGCTGCTGCTGGAAGAACTGTTTCATCAGCGCAACGCCATCGCGTGCATCTGATGACGGC GGCCTTTAACAAGCGCATGGTGAGATCTTCACCAGCGGTCTGATCCACGTGGAAGGCGTGCTGCGCGACTACT TTTATCTGGATGGCGCCTACCATCACGCGTTATCGGTAGCATTCTGCGCGACGAATACTACCGTCACGCGGGT CTGGATGCCGTTCCGCGCATGGATCTGGTGCCGCGTCGCGATAAAGAGGAAGCGGTTGAACTGCTGACCGCCC ATCTCGCGGCCCATCCAGTGTTGCCGTCGTGGTACCAGTGCGACCCAGCCCAAACGGGTCCACCACGCGA AGGTGTTGCGTGATCGATCCACTCCGTACCGATCCACCGCCAACGGCGGCCGATCCAGCCCTCGCGGCGTTCA GTCGTCGTACCAATGGTGAGGGCGCTAA
NocB1	CTGAAATTCGTGGCGCTGGAACCGGTGCTGACCCAAGAAGATACCGCCCGTGCGGCGCGTCTGGTTAGTGCGG GTCCAGTTAGCGCCGTTAGCAGTGCGGGCCCGCAATGGGTTCTGTAACGTGAAGCGACCGCGCTGCTGGATCG TAGTGCCGTTGCCTTTGTTGTTGCCGGCGGTGAAGCGGGCATCGGTGTTCTGCGTGCGGAACGCCTCCAGACGC CGGGCGTTTTCGAACTGGCGCTGCATAGTGCCGACGATGCCGCGCTGGAACCATTTGGCGTTGCCGCGGGCTT CGATGAAATTGTGCGCTATCTGACCAGCACGACCGAGGTTTCATCGCATCGAACTGTGGTTTGGCGTGACAATC GTCTGCTGCTGGATCACTGCCTCGAAAGCGAGCATTTCGCTGCGAAGGCATGGTGCGCGATCGCTATTTTCGCG GCCGGTCGTTATTGGGATGGCCTCGTTTGGAGCATCACCGGTCCACGTCTGCGTGCCATTCTGTAAGCCGATCC AGCCGCCGCGGAAAGCGAACGTGCCTATGCGCGTACCAGTGCCCGTCTGCGTCGCGCCATCCGTGATCGTCTG GCGGACCCAGAACCGGCGAGCTAA
NocB2	ACGATCCATGGTCTGTGGACGCGTCTGGAAGAAAGCCGTGGTGCCGATCCGAGTCGCATGCTGGCGCGTGTGT TTGAAAGTGCCGCCAACGTCTCTTTGCCGGTAGCGCGGTTGCGCCGTTGATGATCGTGGTCTGGGTGCGCTG GCGGATGCGAGTGATCGCATGCTGGGCTATATCTATCGCCGCGATGACGGTGGTCCGGGCGCCCCACGTCCGA TTGGCATTACGATTGCCGTCTCGCCGAGCCGGCCCGTATGCATGTGGGTGTTGTTCTGGCGGGTGATGCGGCG TGGGGTGATGGTGTGGCATGGATGCGATGCTGGCGCTGCTGCGTTATCTGTTTCACTTTGATGGCCGTGCGCG CTTTGAAATGACCGCGGCGGTTCAACCCGCATACGGTTCGATTATGGTTACCGTGCGGTTCTGCCAGAAG CGCTGCTGGCGGATGTTGTTGCGGATGGTAGCCGTCTGGTTGAACTGGCGGTTGGCACGCTGACCGAACCGA GTTTCGCGGAAGTTGGTTATCGCGTGCCGGGCGGTAGCGAGTTACGGATACCCGCTGGCGCGAAGTTCTGAA CGGTTCTGGCCAGATTGGCGCCAGAGCTAA
NocACP	ACCGCCGATGCCAATACCACCATCGAAGACCGCTTTCTGGAAGTTGTTGCCGCCGAAGTGGAACGCATCAACGA TCTGGATGACACCCCAACCGTTACGGTGGATGATGCCATCATCGATGATCTGGAAGTGGATAGTCTGGGTGTTT TGGAAGTGGTTCTGCGTCTGCAGAGCGTTTTTAGCGTGCGCTGAGCGAAGAAGAAGTGGTTGCGGCGACCAC CGTTGGCGATCTGCTGAAACTGGTGAATCCGGTGCGCCCGTGCTAA
MicB	AGCGCGCTGACCCGCATGCTGCGTATCCGAATACCGCGGCCAGCATCGGTGTTCCAGCGTGTTTCGATGATGC GGCCGTTGATGAATTGTTGGCGCCGGTCCGGTTACCGTTACCGGTGATCGCGCGTGCCACGTTGGTTTCAGTC TGGCCCCAGAACATCCGACCGCCACGTGCTTCGTTGCGATCCCGGACGGCCATGGTACCCAGATTGGGTTGCC GCGGCCGTTACCGAAGCCGCCAGTGCCGCCGGTGCGCGTGGTGCGAAACGTTTCGAAGCGTGGGTTCCGGCGT GGAATGATGCGATTGCGCGTGCCACCGAACGCAGCATGACGCGTGAAGGTCTGCTGGAAAAAGCGGTGTTTGC CTACGGTGAACGCTGGGATCTGACGCTGTACAGCGTGCAAGATGGCGGTGCTAA
MicACP	ACCGATAGCGAACTGTACGATTTCTGTCGCTGGAAGCGGGCCGCATCAATGAAGAAAGCTACGCCGACGTGG TGCTGAGCGATGATACCGTTCTGGCGGATGATCTGGACATCGATAGTCTGGCCATGCTGGAAGTGTGCATCCGC GTTGAAGAAGTGTTCCGGCTTGCGGTTGCCGATGAAGTGGCCGCCGAAATCAAACCGTTGGCGATCTGCGCC GCTTTCTGGATAGCGCCGAAACCGTTTCGCGCTAA

Supplementary table16. Optimized expression gene sequences used for this study.

Name	Sequence
FlaB1	GATGGTCGCGCGGGCACGAGTCTGGAATTTATCCCGGTGCTGACCGAGAACGAGGAAAACTGCTGGATGGC GTGCTGGCCGCGCCATTAGCTACGTTAGCACGGAAGGTCCAGTTTGGCGTACCGCCGCCATGCGTCGTCGCAT CGTTGAAGATGATCACCGCAGCCATCTGCTGGTTCCCGCGATGAACGTGTTATTGGCGCCGTTAGCTGGAGCG CCGGTCAAACCCCGGGTTACCATCGTCTGGTTCTGACGAGCTGCAGCGATGGCGATTGGACCCCGGGCACGGC CGATGAAGCGATTCTGCGCGCTGAGTGCGCTGCTGCGTGCCGCGGAAGTGAAACGCGTGGAAGTCTGCTGGTT GCGACCTACAATGAGGTGCTGCTGGAATACCTACCGAATACAGCGGCTTCATCGTTGAAGGCGTTCTGCGCGA TCGTTCTTCATCGACGGCGAGTTTTGGCCGGGCATCATCTGCCGCGCCGACGTTACCGAATTCGCGCAGCGTG GTGAACGCCTCCAGAAAAGTCGTGCGGCGCTGCTGACCAGTCTGCGCAAGAAAAACGCTGGAGGATCTGAAAGC GGCCCATGATGGCACCAGTTGGAGCGCTAA
FlaB2	AAAGGTCTGCTGTGCGAACTGGTTCCGCTGACCCCGGATAGTCTGGATAGTGTTGCCCGTTGGCTGAGCGGCAC GGATTGCGAAACGCTGCTGCGTGGTACCAAGTAGTCGTCTGGTTCGCCGCAGCCGCTACGAAACGCTGGAAGGC GATGATCGCCGTGCGGCCCTCATCACCGATCGTCAAGGCACCCAGTTGGCGTTGCCATTATTGGCGTCTGGG TGAAGGCGTGTATGAAGTTGGCGGTGCGACCGGCGATGAACGTCTGTGGCATAACGGGCATTGGCATTGAAGCC GCCATGCTGCTCGTGGAATTATCTGTTCAAGTGCTGAACTGCCGCCGCGTTGAATTCACCACCGGTCTGCACAAC AACACCACGCTGCAGCTGGTGCTGACCGACGGTCTGAGTCTGGAAGCGGTGTGCGAGGATTACTTCCAGACCC CAAATGGCACGCTGCCGGCCGTGAAAAGCAGTCTGACCCGTGAGGAGTATCGCCGCCCGTACCCACATTATCA GCCACGCGTGGGTGCGCGCTTGATCGTGATCGCGTGCGCACGCTGACGGAAGGTCTGGTTGGTCGTGCCGCG GCCGATCCGGGCAGTAGCGAGGAAAAACAAAGCGTAA
cACP-FlaPKS	GATGCCCCGGCGCTGCGCGGTCTGCTGGCCGCCATGCCGGGCGCGGGTCTGTTGTTCAAGCGAAAGCGGTGGT TTTCGTGCGGCCTTTGATGCGGCGAGCAGCGGTGTTCTGGCGAACTGCTGCTGGCCGCGTTCGTGAACAAGT TGCGCGCGTTCTGGGCGAAACCGATCCGGATGCCGGCGAAACCGATGATCTGGATGTGTTTGCCCGTGGTCTG GATAGTCTGACCGCCATCGAACTGCGCCGCAATCTGCAGCGTGCGCTGGGTTGCCGTCTGGGTGAGAGTGCCG TTTACGATCATCCAACCGTTGCCGCGCTGGCCGGTCATCTGGGTGAACTGATCGAAGATGACGGTAGCGGCACC CGCTAA