SUPPORTING INFORMATION: DISCOVERY OF NEW CYCLIC LIPODEPSIPEPTIDE ORFAMIDE N VIA PARTNERSHIP WITH MIDDLE SCHOOL STUDENTS FROM THE BOYS & GIRLS CLUB

Jin Yi Tan[1], Mario Augustinović[1], Ashraf M. Omar[2], Vitor B. Lourenzon[1], Nyssa Krull[1], Xochitl Lopez[1], Manead Khin[1], Gauri Shetye[3], Duc Nguyen[3], Mallique Qader[3], Angela C. Nugent[3], Enock Mpofu[3], Camarria Williams[4], Jonathon Rodriguez[4], Joanna E. Burdette[1], Sanghyun Cho[1,3], Scott G. Franzblau[1,3], Alessandra S. Eustáquio[1], Qibin Zhang[2], Brian T. Murphy[1*]

[1]Department of Pharmaceutical Sciences: Center for Biomolecular Sciences: College of Pharmacy, University of Illinois at Chicago, IL 60612, United States

[2]Center for Translational Biomedical Research, University of North Carolina at Greensboro, NC 27412, United States

[3]Institute for Tuberculosis Research, College of Pharmacy, University of Illinois at Chicago, IL 60612, United States

[4]Boys and Girls Clubs of Chicago, IL, United States

*Corresponding Author: BTM: Phone: (312) 413-9057; E-mail: btmurphy@uic.edu

TABLE OF CONTENTS

Figure S1. Fractionation tree for isolate BCGFaB3.

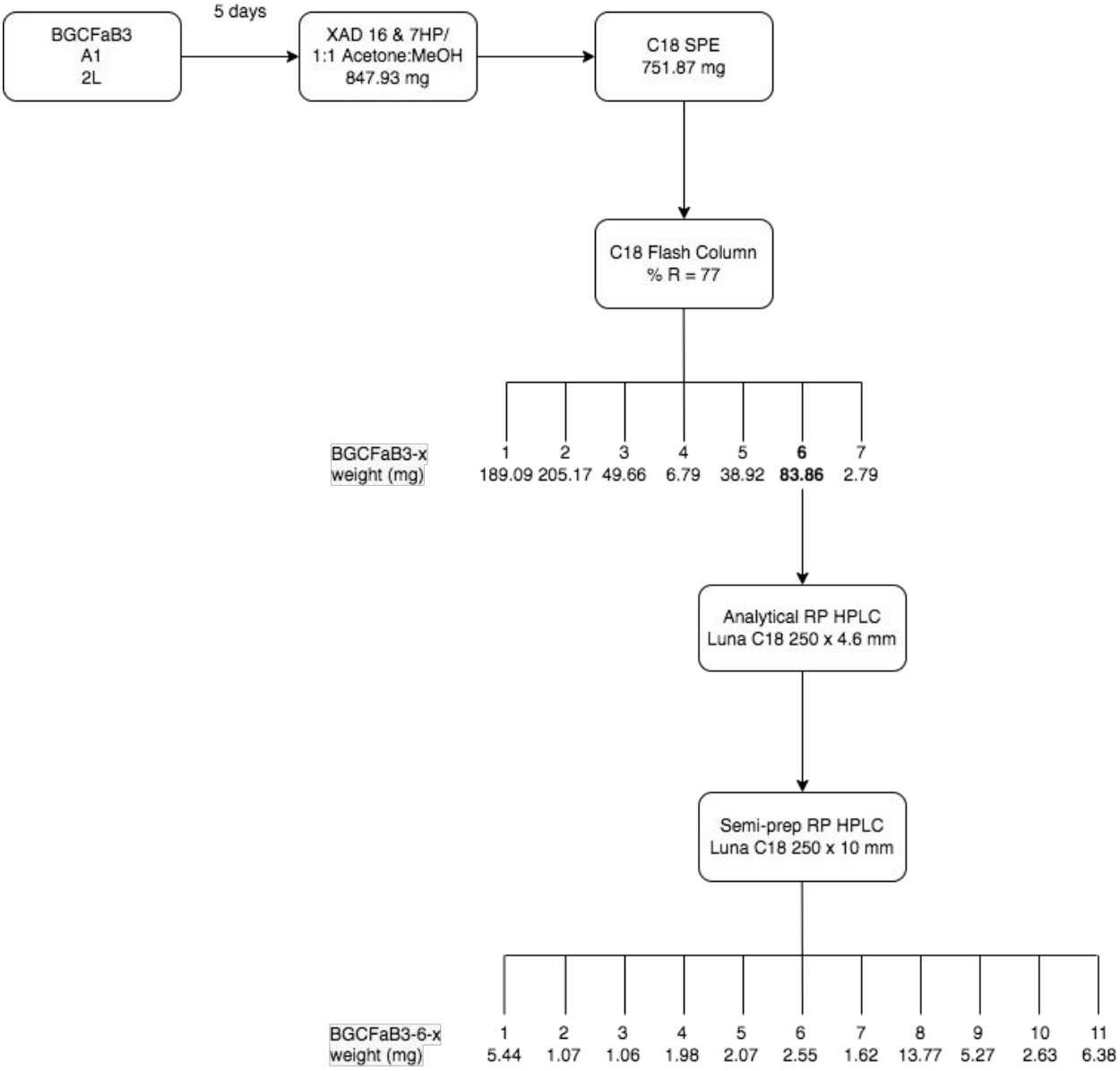


Figure S2. Isolation of orfamide N (**1**), orfamide A (**2**), and orfamide M (**3**) via semi-preparative RP-HPLC.

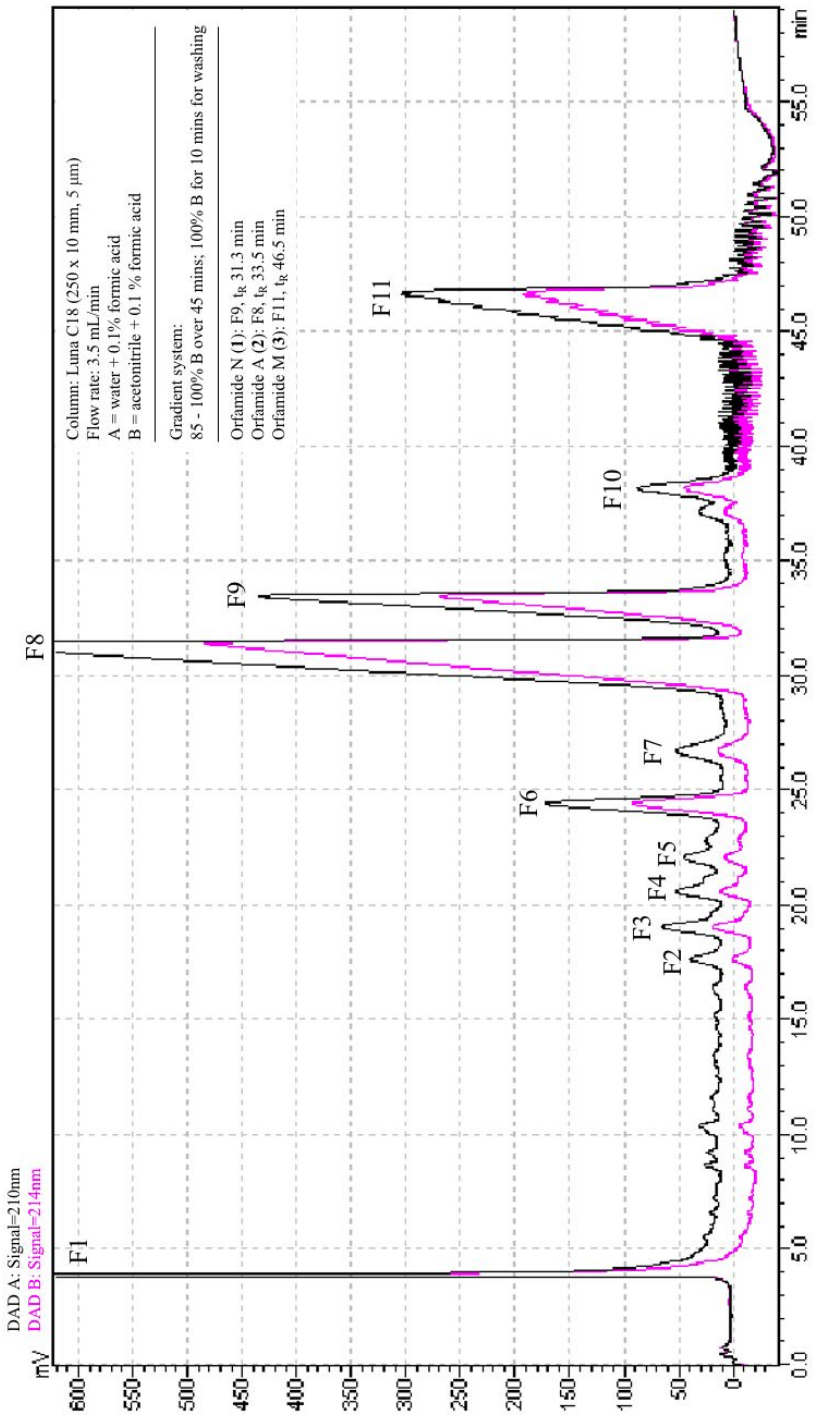


Figure S3. Dereplication of orfamide cyclic lipodepsipeptide analogs via LCMS/MS / GNPS analysis.

Figure S3A. Low-resolution MS/MS spectra of orfamide analogs from subfractions BGCFaB3-6-6 (orfamide B *m/z* 1281.88 $[M+H]^+$), BGCFaB3-6-7 (orfamide F *m/z* 1307.90 $[M+H]^+$), BGCFaB3-6-8 (orfamide A (**2**) *m/z* 1295.90 $[M+H]^+$), BGCFaB3-6-9 (orfamide N (**1**) *m/z* 1321.92 $[M+H]^+$), BGCFaB3-6-10 (orfamide G/H *m/z* 1309.92 $[M+H]^+$), and BGCFaB3-6-11 (orfamide M (**3**) *m/z* 1323.94 $[M+H]^+$).

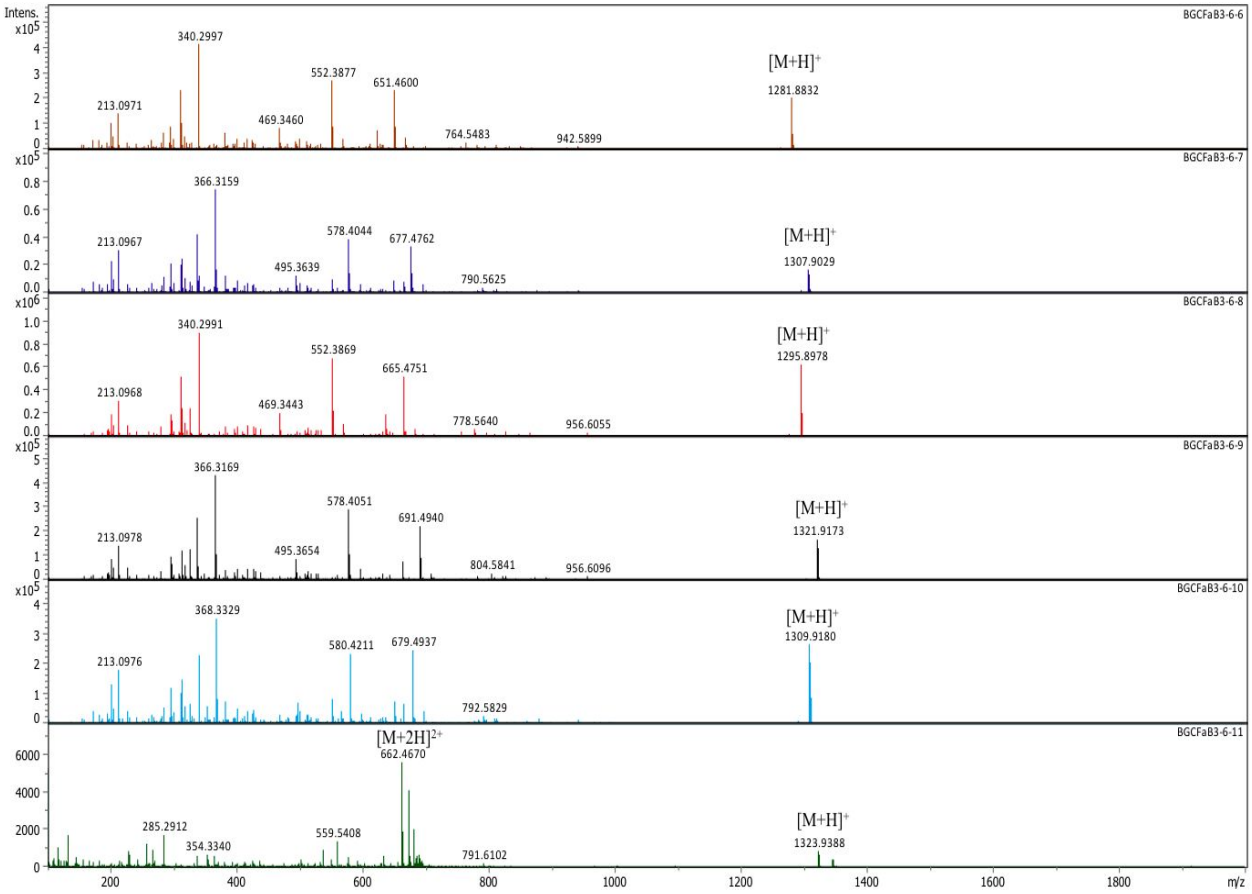
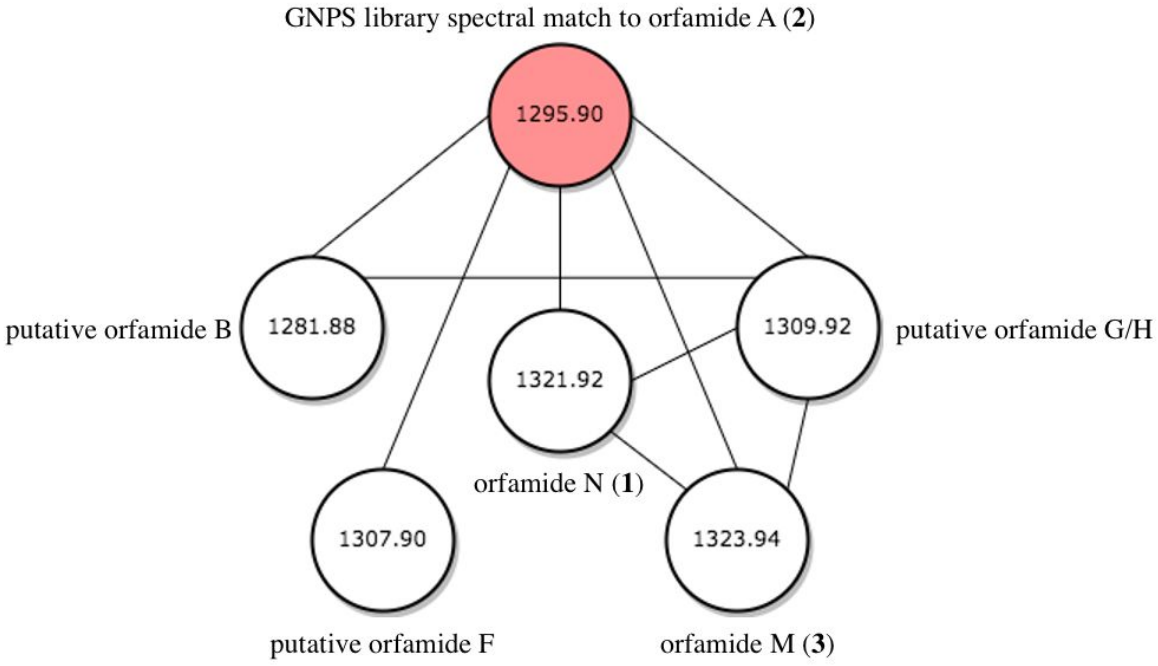


Figure S3B. GNPS analysis and molecular network for orfamide analogs in BGCFaB3-6.



GNPS parameters.

Basic options: Precursor Ion Mass Tolerance, 2 Da; Fragment Ion Mass Tolerance, 0.5 Da

Advanced Network Options: Min Pairs Cos, 0.65; Network TopK, 10; Maximum Connected Component Size, 100; Minimum Matched Fragment Ions, 4; Minimum Cluster Size, 1

Advanced Library Search Options: Library Search Min Matched Peaks, 6; Score Threshold, 0.7; Maximum Analog Search Mass Difference, 100.0

Figure S4. HR-qTOF-MS/MS spectra of **2**.

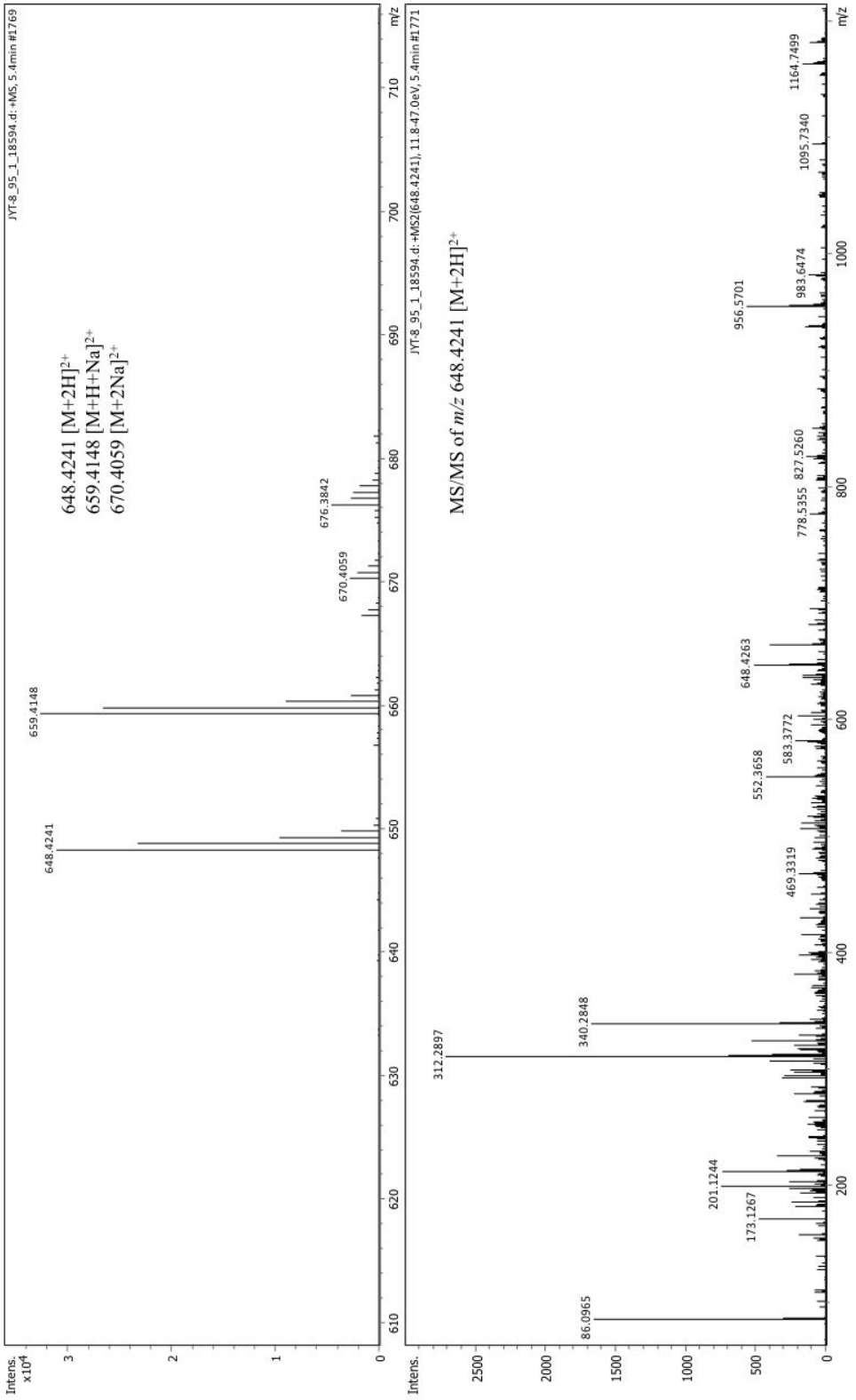


Figure S5. ^{1}H / ^{13}C NMR (600 / 150 MHz, d_3-MeOH) spectra of **2**.

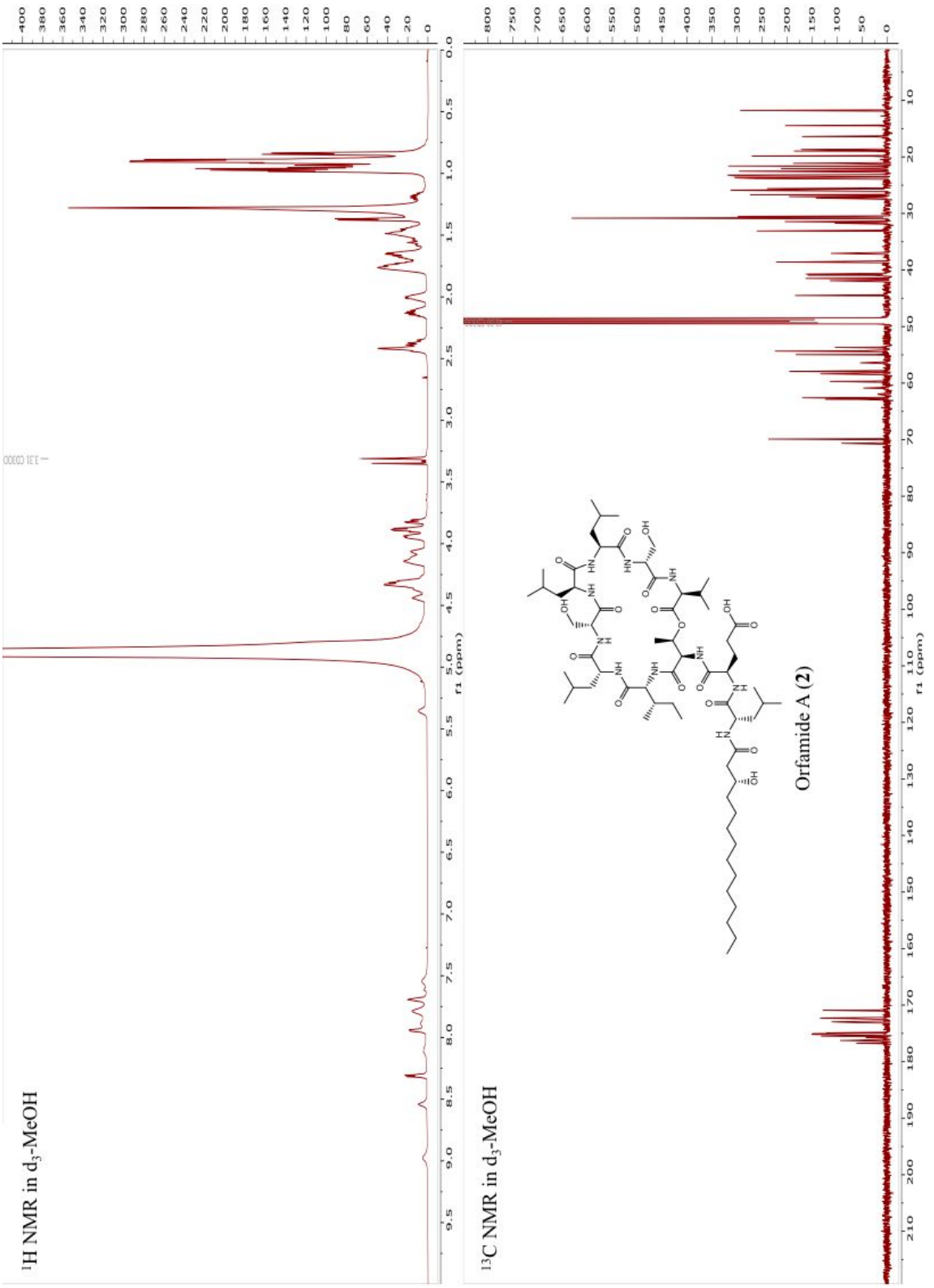


Figure S6. Two-dimensional NMR (600 / 150 MHz, d_3-MeOH) spectra of **2**.

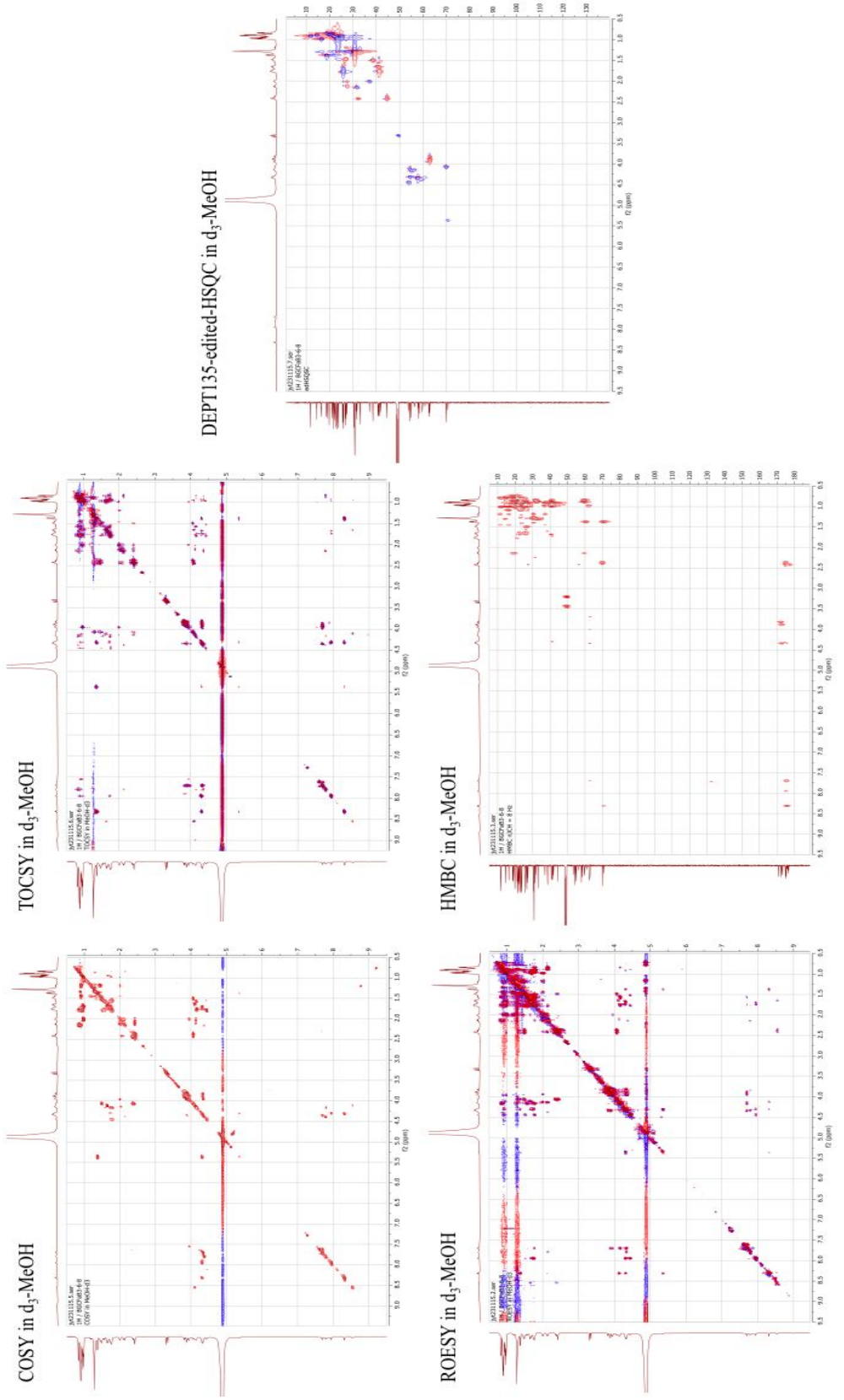


Figure S7. HR-qTOF-MS/MS spectra of **3**.

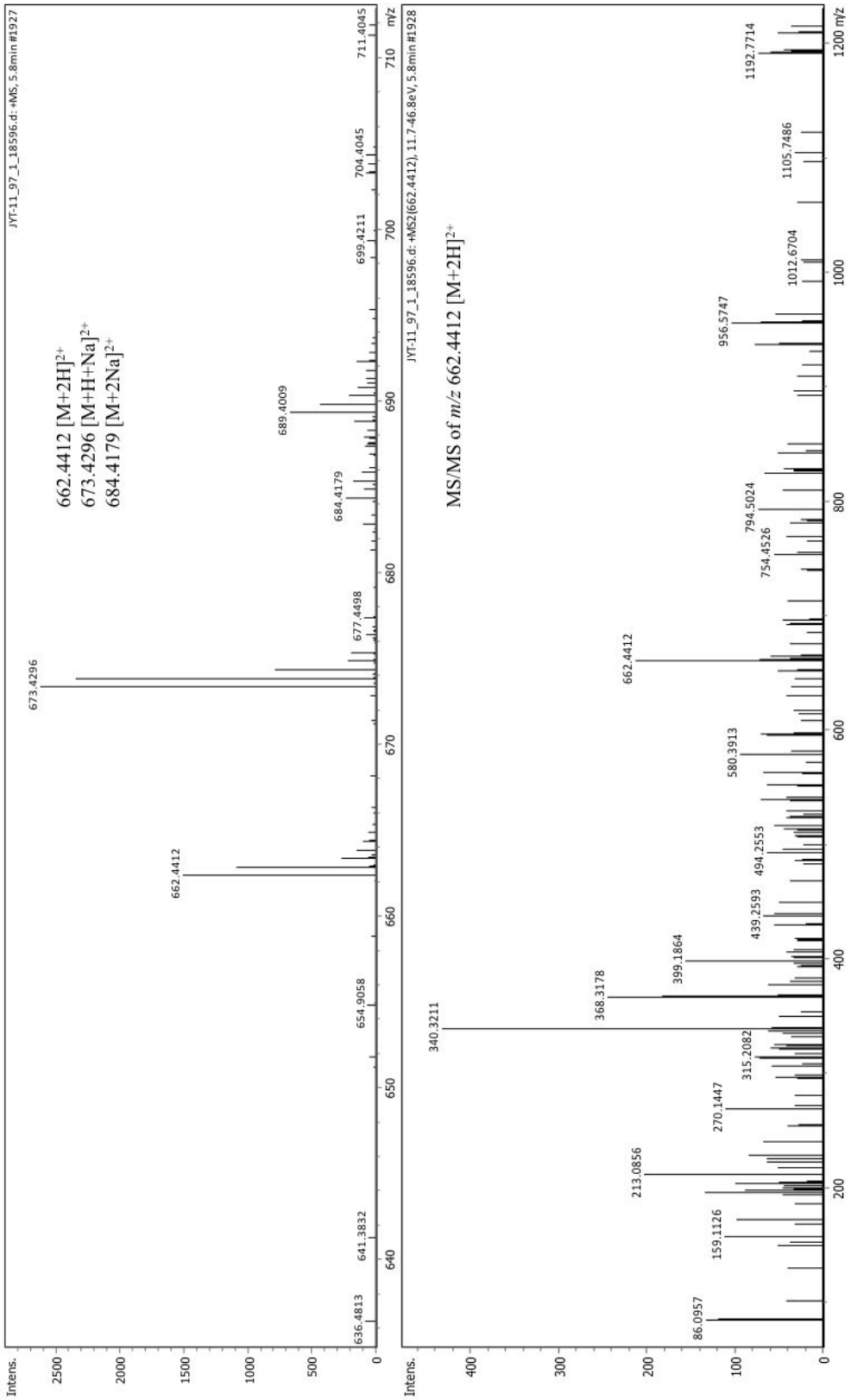


Figure S8. ^{1}H / ^{13}C NMR (600 / 150 MHz, d_3-MeOH) spectra of **3**.

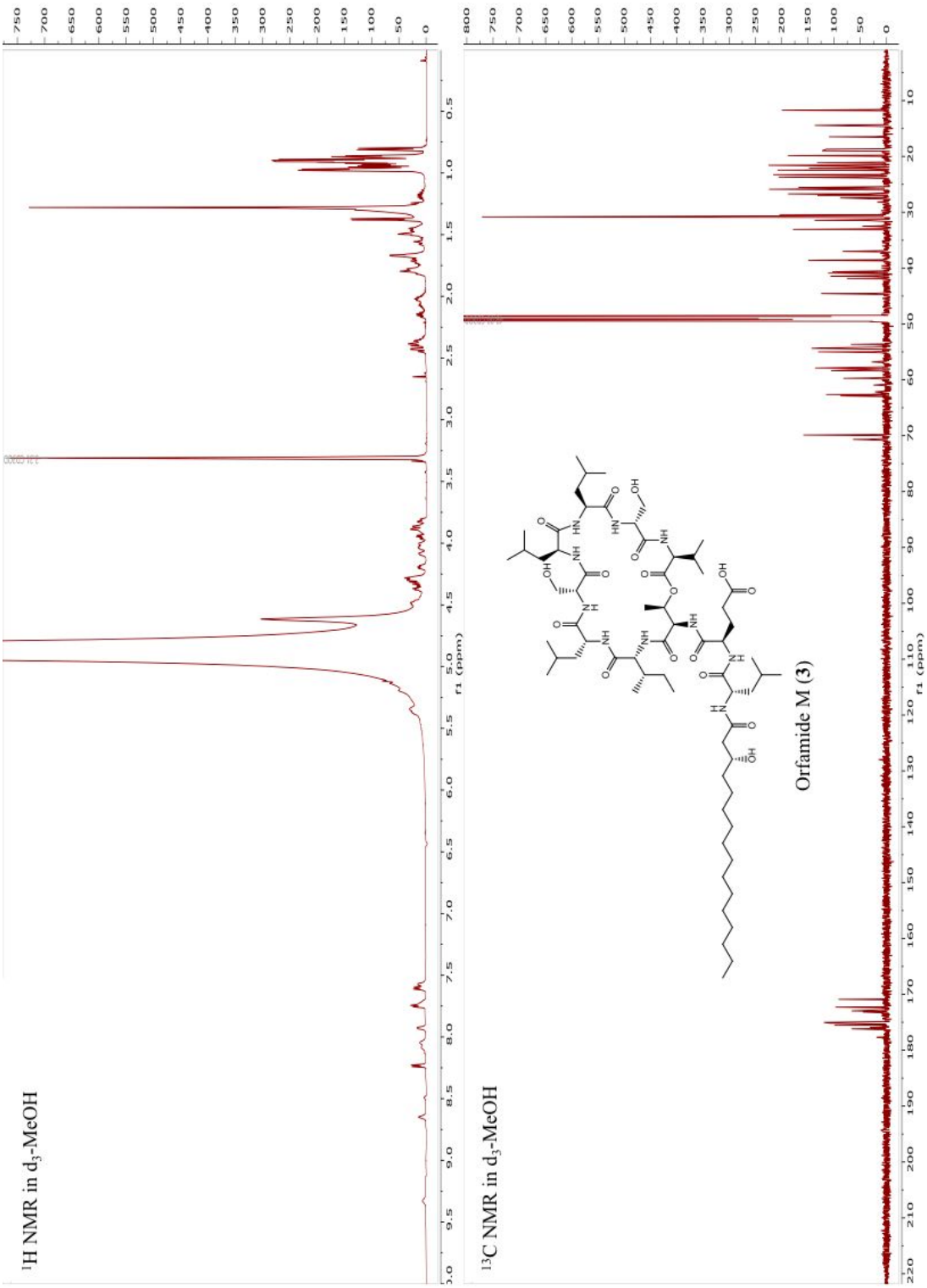


Figure S9. HR-qTOF-MS/MS spectra of **1**.

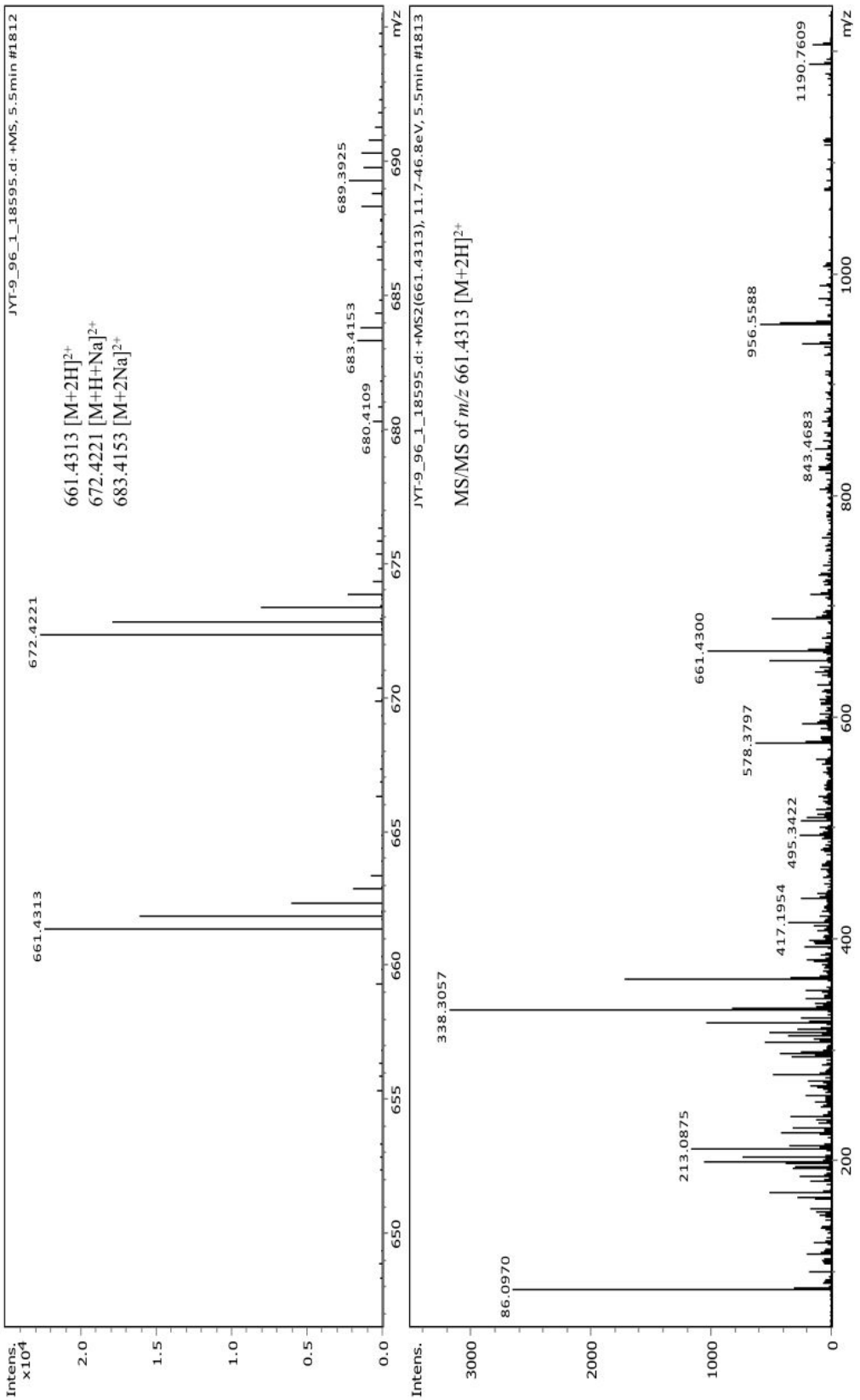


Figure S10. ^{1}H NMR (600 MHz, d_3-MeOH) spectrum of **1**.

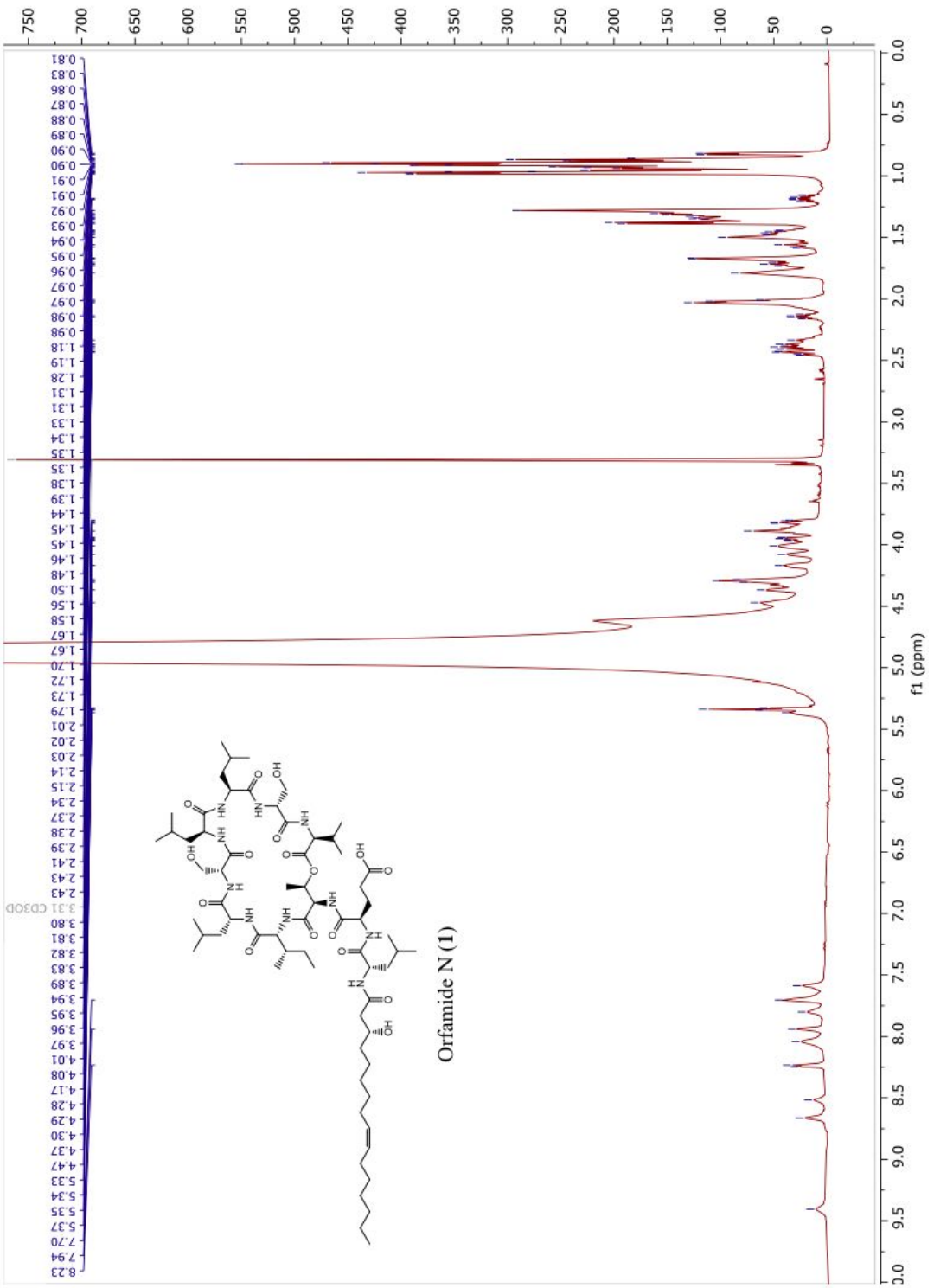


Figure S11. ^{13}C NMR (150 MHz, d_3-MeOH) spectrum of **1**.

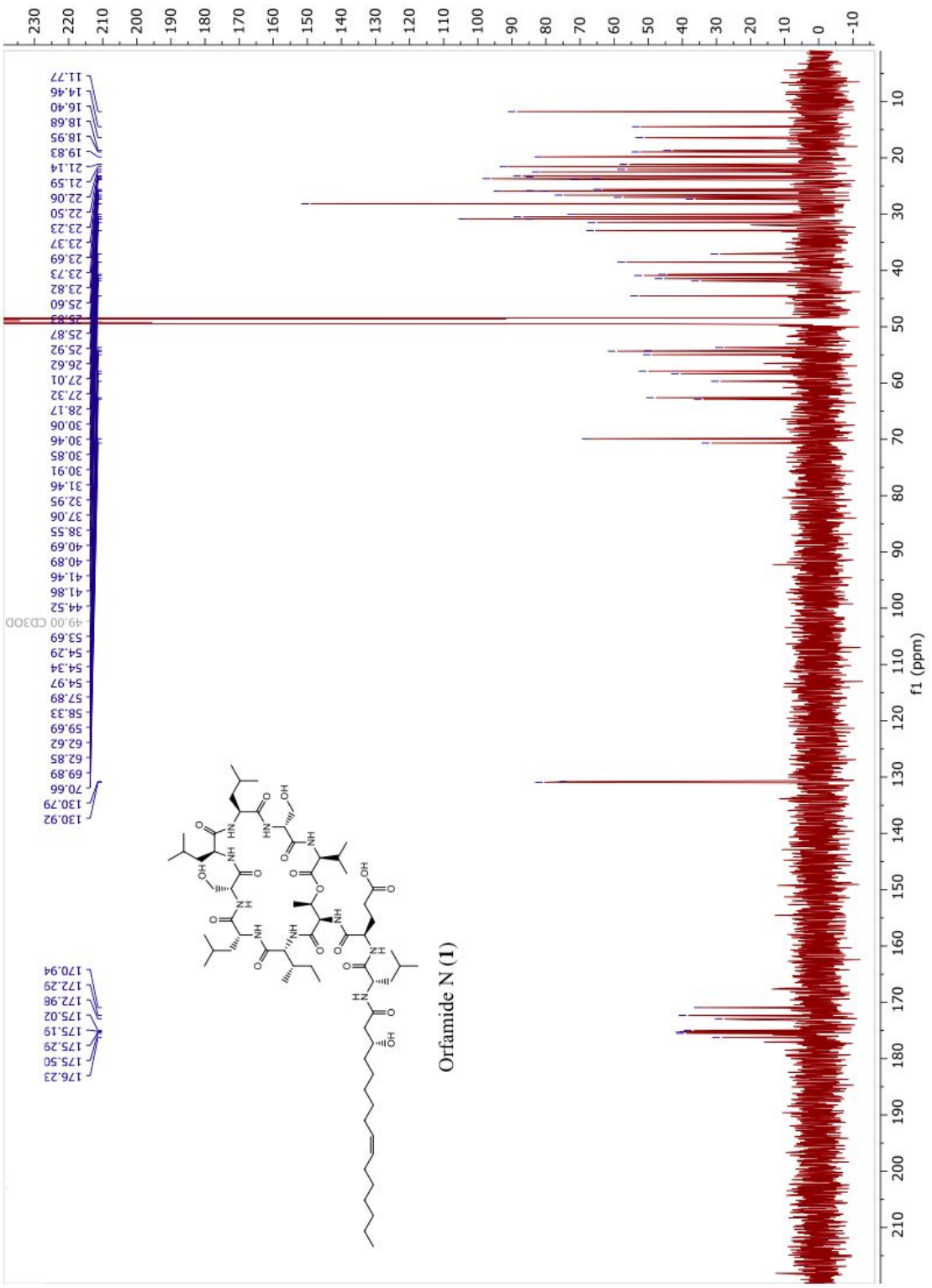


Figure S12. COSY (600 MHz, d_3-MeOH) spectrum of **1**.

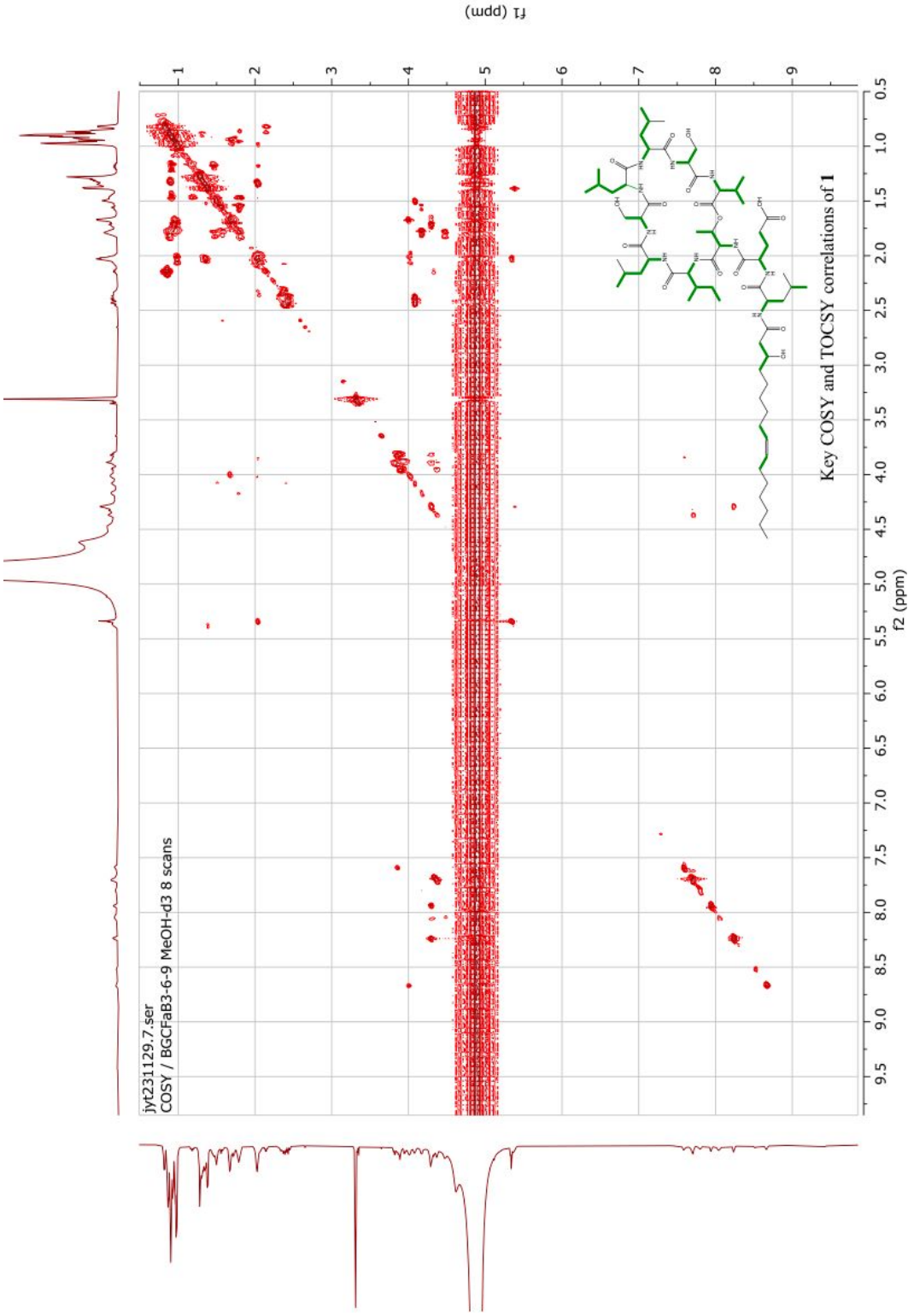


Figure S13. TOCSY (600 MHz, d_3-MeOH) spectrum of **1**.

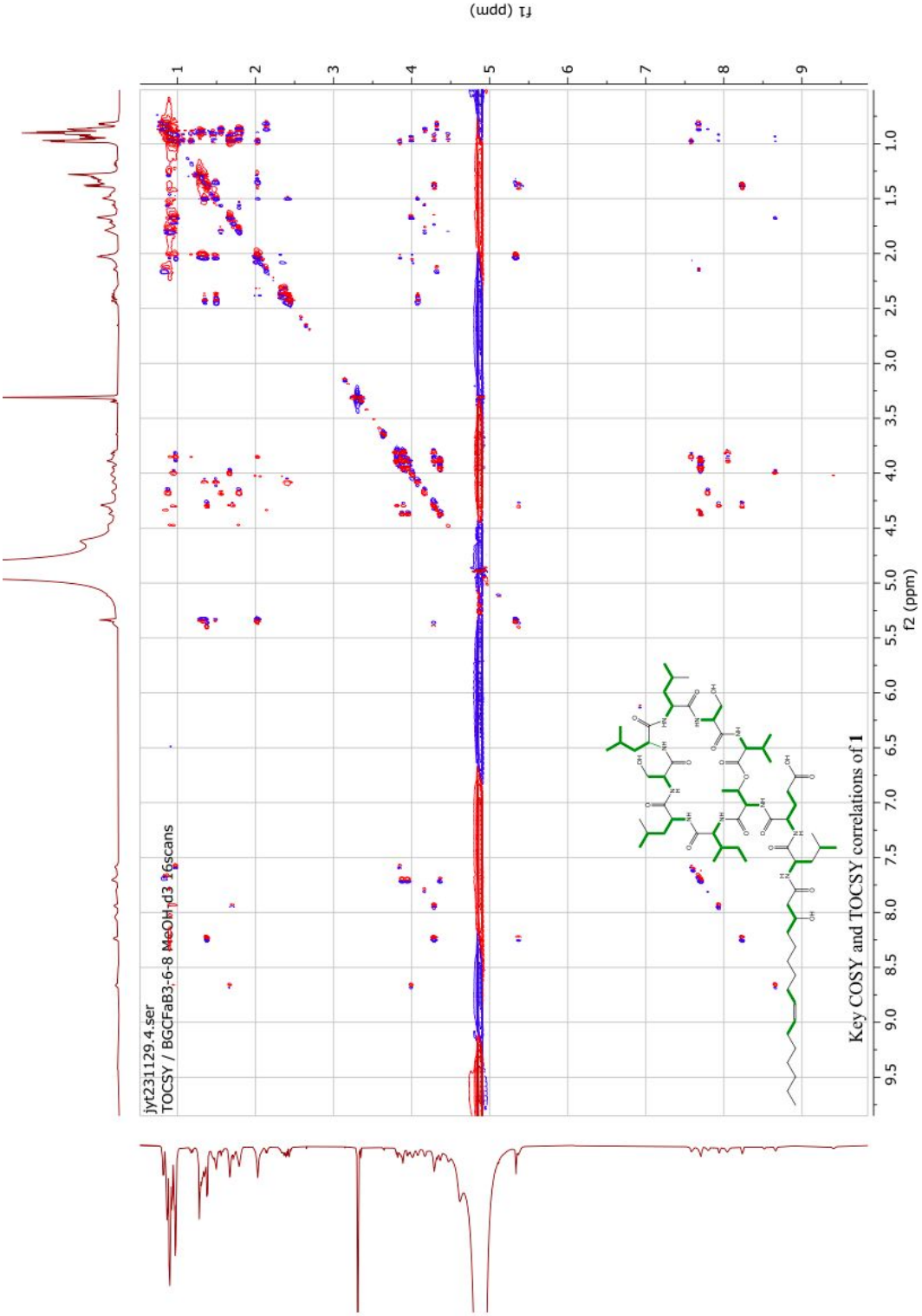


Figure S14. HSQC (600 MHz, d_3-MeOH) spectrum of **1**.

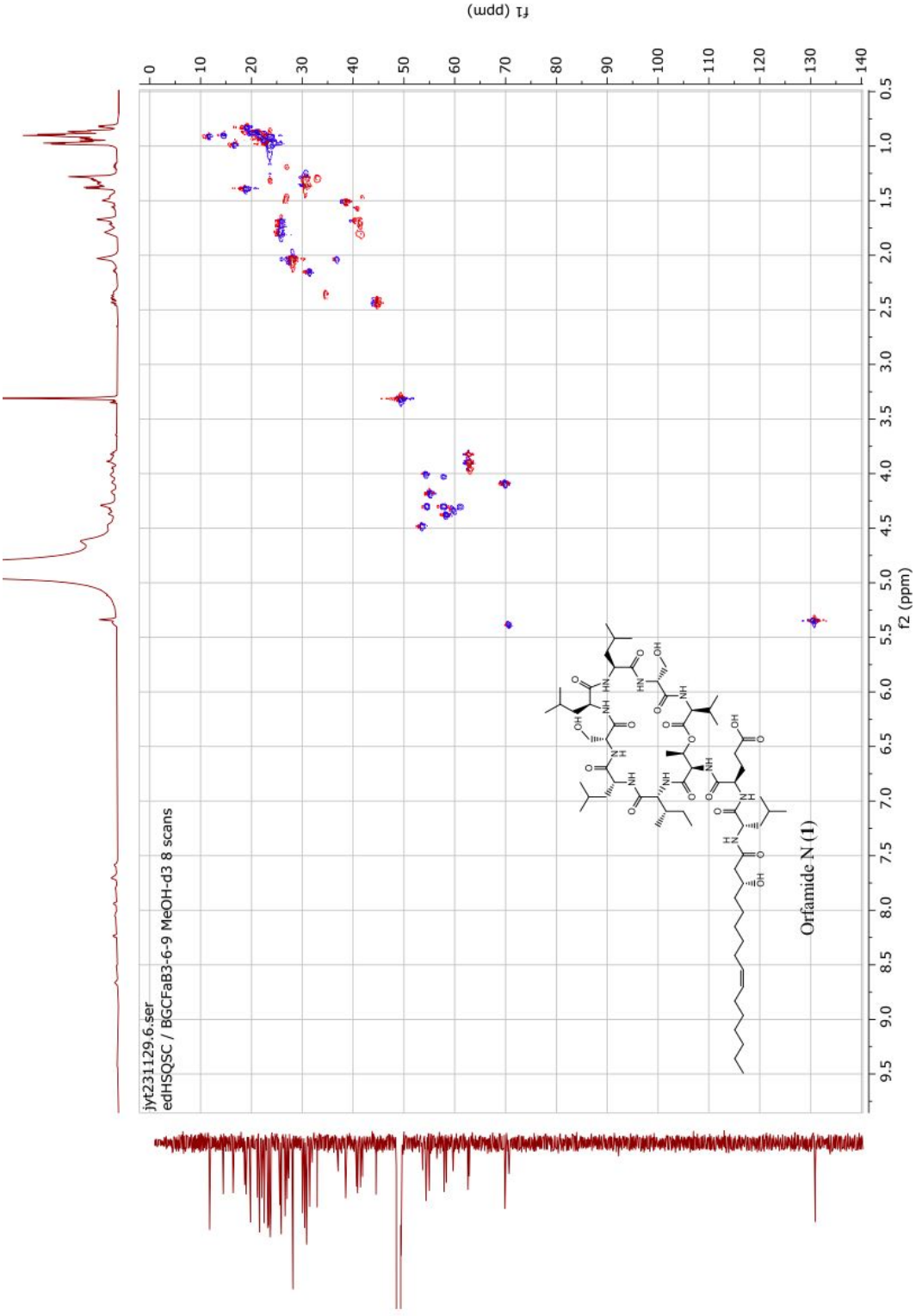


Figure S15. HMBC (600 MHz, d_3-MeOH) spectrum of **1**.

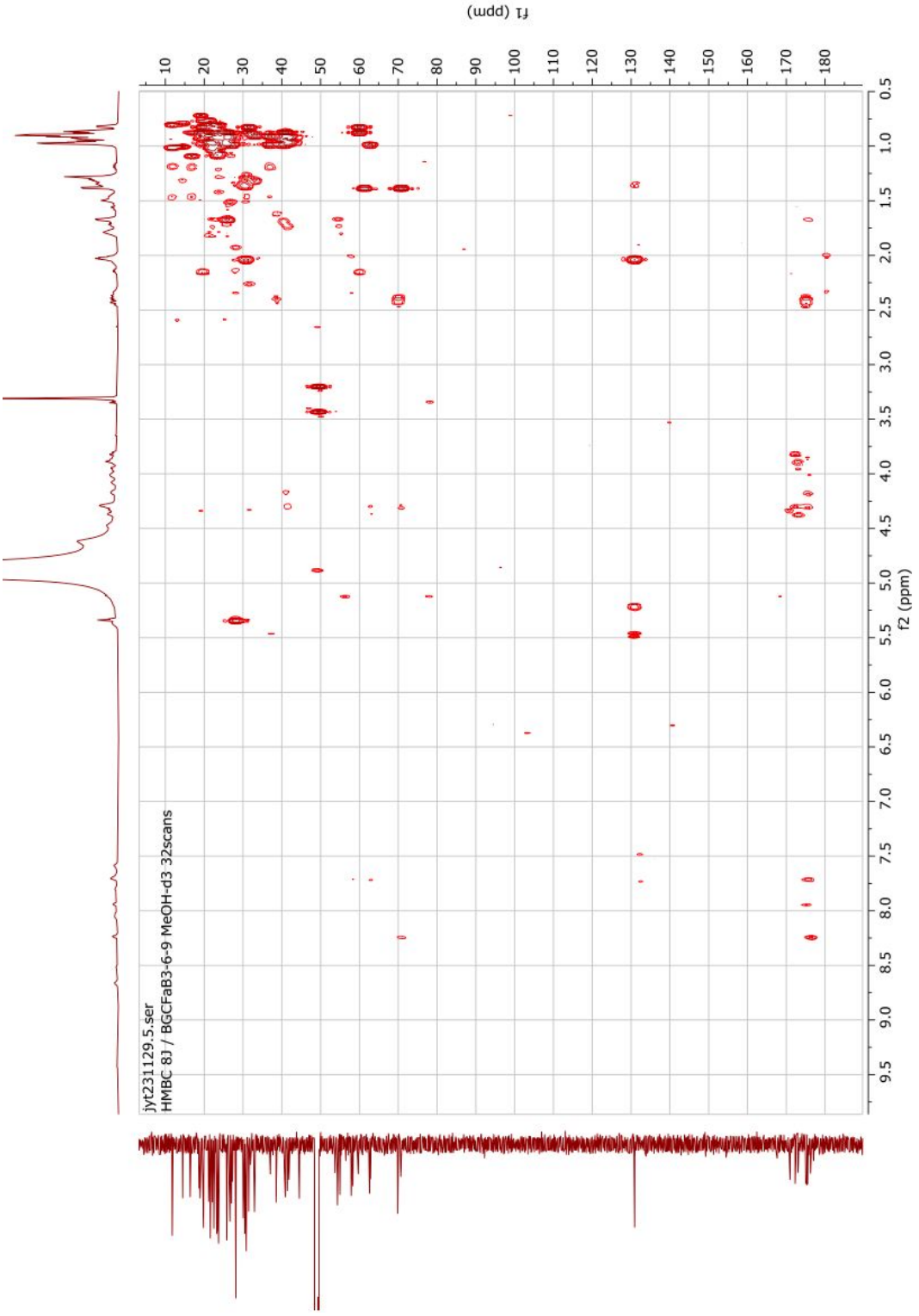


Figure S16. ROESY (600 MHz, d_3-MeOH) spectrum of **1**.

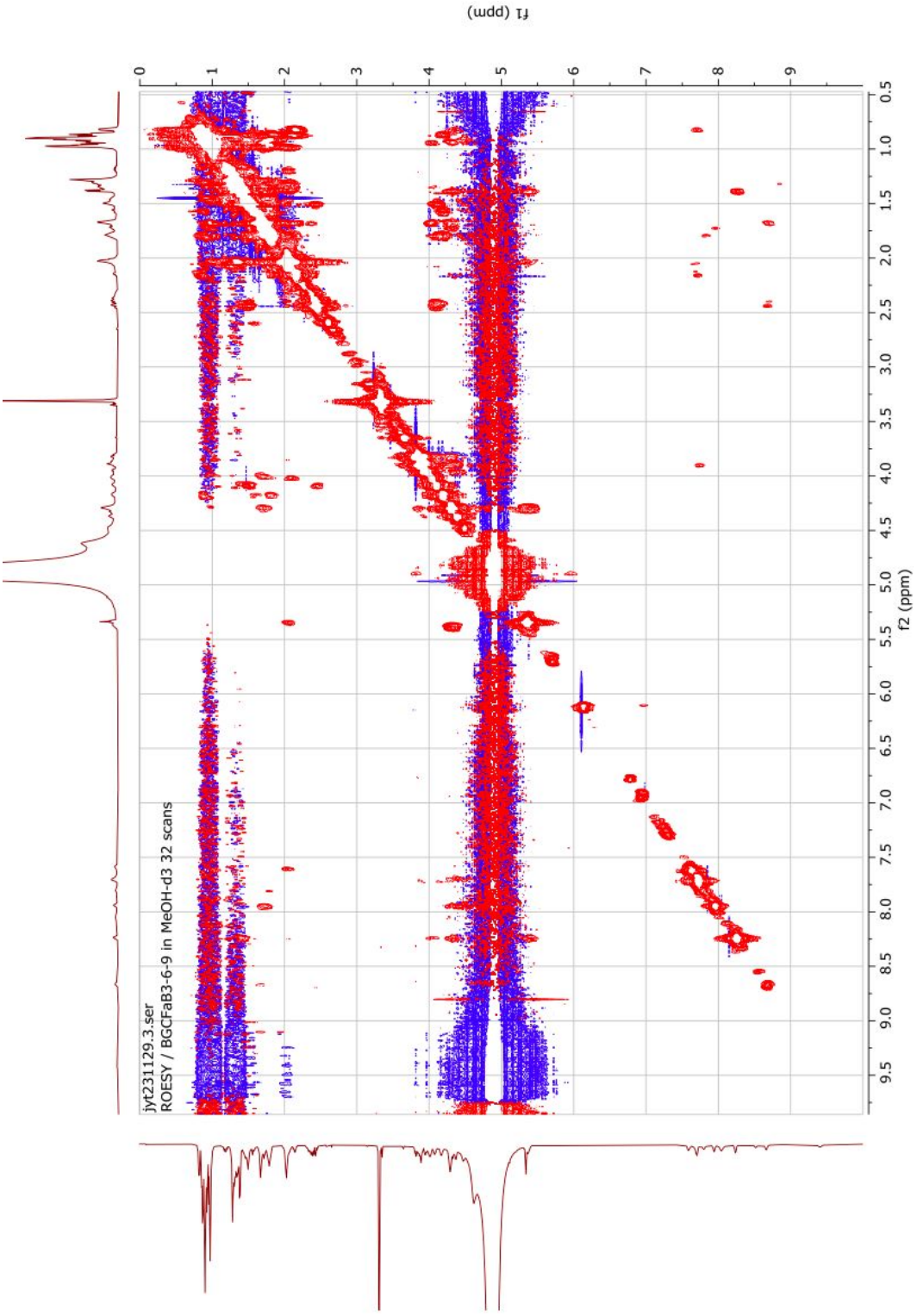


Figure S17. Whole genome phylogenetic analysis of isolate BGCFaB3. The whole genome phylogenetic tree was built with Type (Strain) Genome Server (TYGS) to identify the isolate BGCFAB3. *P. idahonensis* 1D357 is the closest strain to the isolate BGCFAB3 (highlighted in orange).

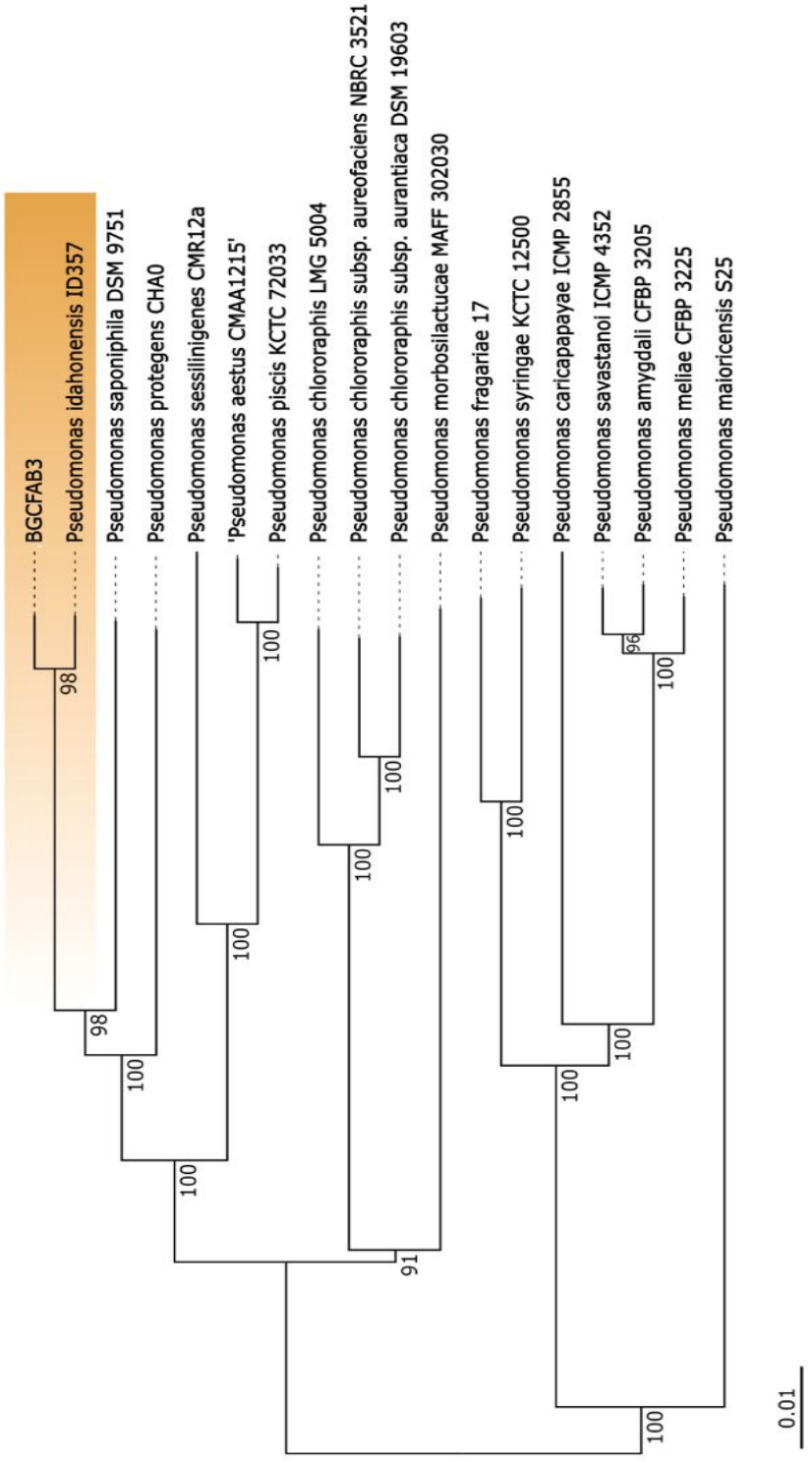


Figure S18. A phylogenetic analysis of the C-domains in the orfamide N biosynthetic gene cluster.

Figure S18A. A phylogenetic tree of the condensation (C) domains (amino acid level) of the *orf* biosynthetic gene clusters (BGC) in *P. idahonensis* BGCFaB3 (orfamide N-producing strain) and a comparison with that of *P. protegens* Pf-5 (orfamide A-producing strain). The AntiSMASH prediction of the condensation domains showed that the C-domain on module 1 (C1) encodes for a starting condensation domain, frequently capable of incorporating non-amino structures such as polyketides or lipid chains. Domains C8 and C9 were predicted as $^{L}C_{L}$, condensation domains forming the bond between two L-amino acids. The remaining C domains (C2, C3, C4, C5, C6, C7, and C10) were predicted as C_{DUAL}, which are dual epimerization/condensation domains capable of catalyzing epimerization of the amino acid incorporated in the previous module into a D-amino acid. These three different types of C-domains fall into distinct clades and are clustered together.

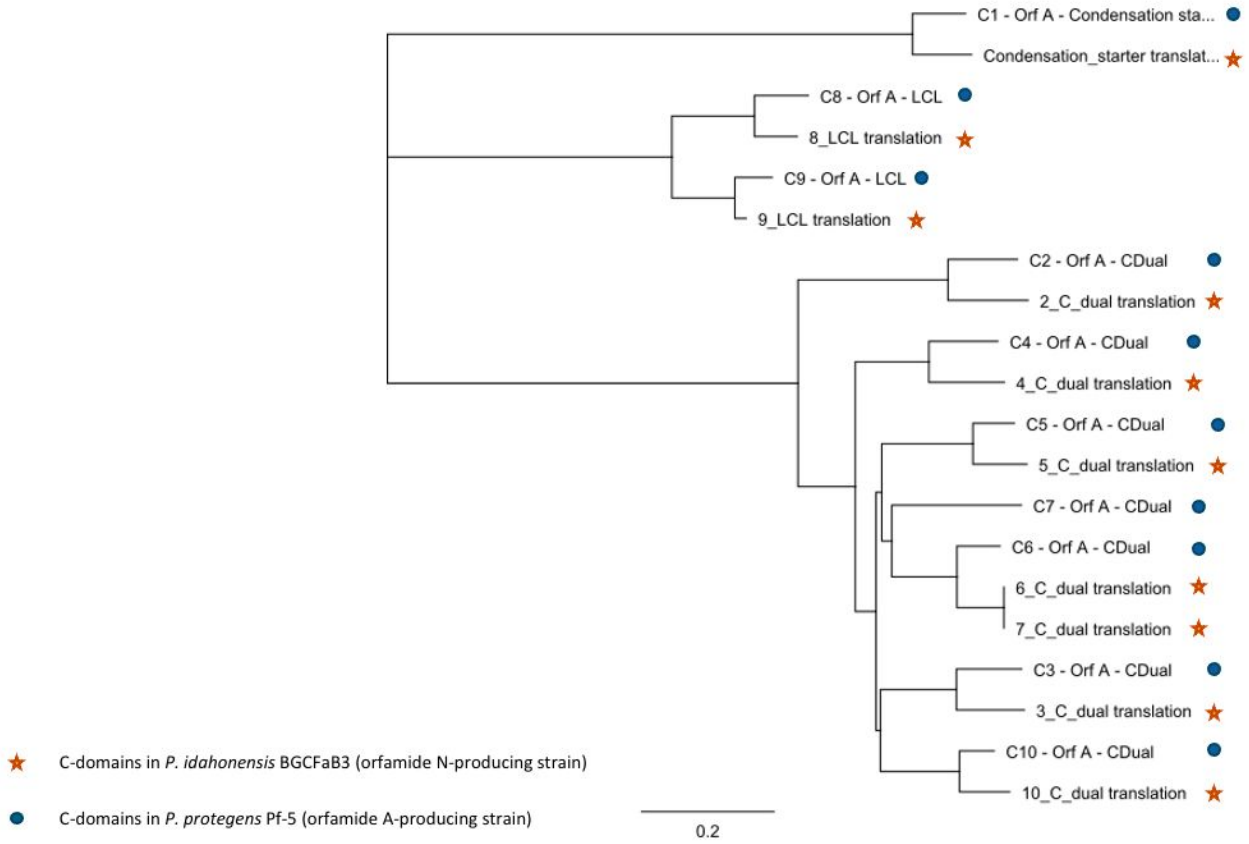
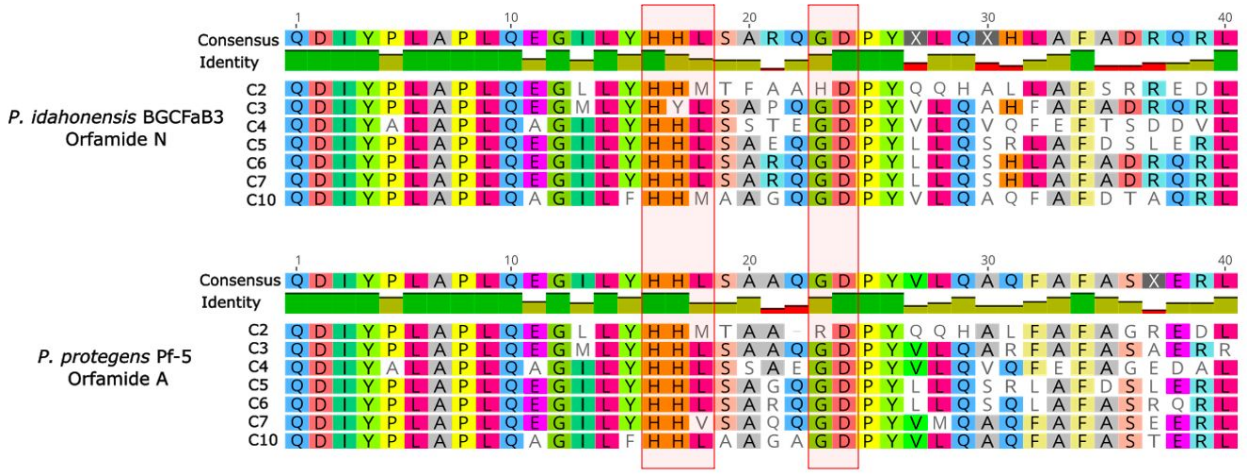


Figure S18B. Amino acid sequence alignment of C_{DUAL} domains in *P. idahonensis* BGCFaB3 (orfamide N-producing strain) and *P. protegens* Pf-5 (orfamide A-producing strain), highlighting the conserved HHLxxxxGD motif at the *N*-terminus which is characteristic of C_{DUAL} domains[1]. To resolve the position of the single D-leucine in orfamide N, the amino acid sequence of the condensation domains for all C_{DUAL} were aligned. The C-domain from module 2 presents two mutations in the catalytic regions essential for C_{DUAL} activity, suggesting the domain is incapable of performing the epimerization of the leucine incorporated by module 1. On the other hand, this region was fully conserved for C-6 and consistent with published analogs, suggesting the epimerization of L-Leu to D-Leu at module 5.



Reference:

(1) C.J. Balibar, F.H. Vaillancourt, C.T. Walsh, Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chem Biol*. **12**, 1189-200 (2005).

Figure S19. Chiral HPLC analysis of isoleucine residues.

Figure S19A. Chiral HPLC analysis of the underivatized hydrolysate of **1** was performed using a Phenomenex Chirex 3126 D-penicillamine column (4.6 x 250 mm) with an isocratic flow of 1 mM copper (II) sulfate in water/isopropanol (95:5) at 1 mL/min. The retention time of the Ile residue in orfamide N (**1**) was compared to that of D- and D-*allo*-Ile standards. Retention times were measured by HPLC at 254 nm. Identifications were confirmed by co-injection with standards (Fig. S19B).

	Retention time (min)		**Measured**	**Assignment**
	D	**D-*allo***		
Ile	56.4	45.9	46.2	D-*allo*

Figure S19B. Chromatogram for (a) D- and D-*allo*-Ile standards (b) the underivatized hydrosylate of **1 (**c) co-injection with D-*allo*-Ile standard.

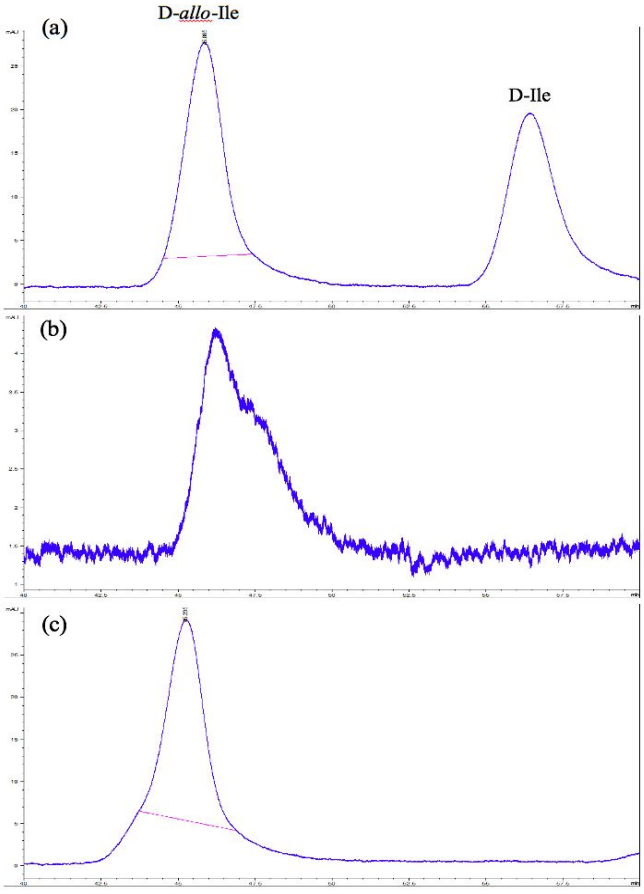


Figure S20. Advanced Marfey's analysis of orfamide N (1).

Figure S20A. The retention times of derivatized amino acids in orfamide N were compared to that of derivatized amino acid standards. Retention times were measured by UPLC-DAD-ESIMS extracted ion chromatograms.

Amino acid	**Retention time (min)**			**Assignment**
	L	**D**	**Measured**	
Ser	0.68	0.72	0.72	D
Glu	0.86	1.03	1.03	D
Leu	4.15	5.81	4.16, 5.81	3 L and 1 D[a]
Val	2.68	4.2	2.67	L
Thr	0.78	1.18	0.94	D-*allo*-Thr
allo-Thr	0.78	0.93		
Ile	3.91	5.64	5.66	D-*allo*-Ile[b]
allo-Ile	3.93	5.65		

[a]The ratio of L-leucines to D-leucines was determined by the area under the curve of 2317.187 to 749.299 (Fig. S20C).

[b]Marfey's analysis results were unable to differentiate between D- and D-*allo*-Ile. To resolve the configuration of the remaining Ile residue, the underivatized hydrolysate of **1** was subjected to chiral HPLC analysis, which showed it to be of D-*allo* configuration (Fig. S19).

Figure S20B. Extracted ion chromatogram (EIC *m/z* 370.13) for (a) orfamide N derivatized hydrosylate (b) FDAA-L-valine (c) FDAA-D-valine.

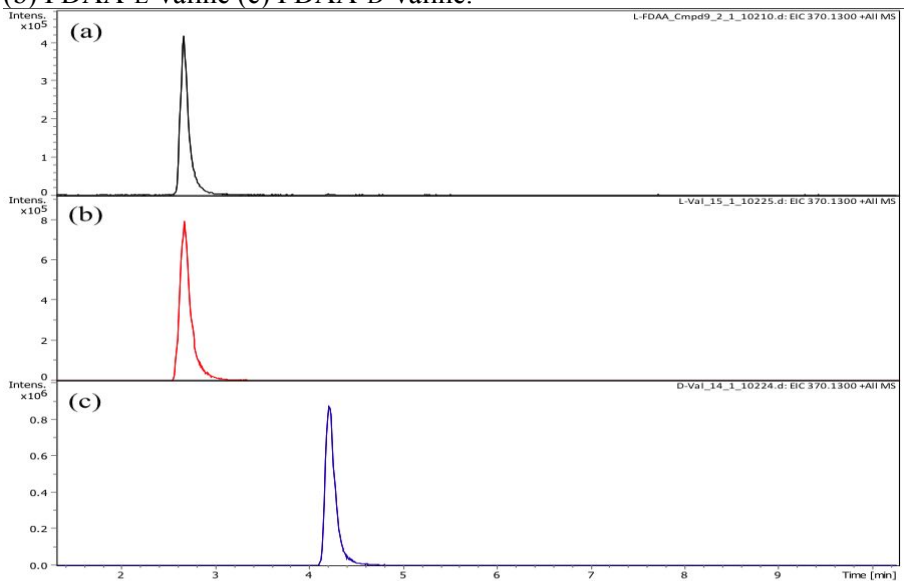


Figure S20C. Extracted ion chromatogram (EIC *m/z* 384.15) for (a) orfamide N derivatized hydrosylate (b) FDAA-L-leucine (c) FDAA-D-leucine (d) FDAA-L-isoleucine (e) FDAA-L-*allo*-isoleucine (f) FDAA-D-isoleucine (g) FDAA-D-*allo*-isoleucine.

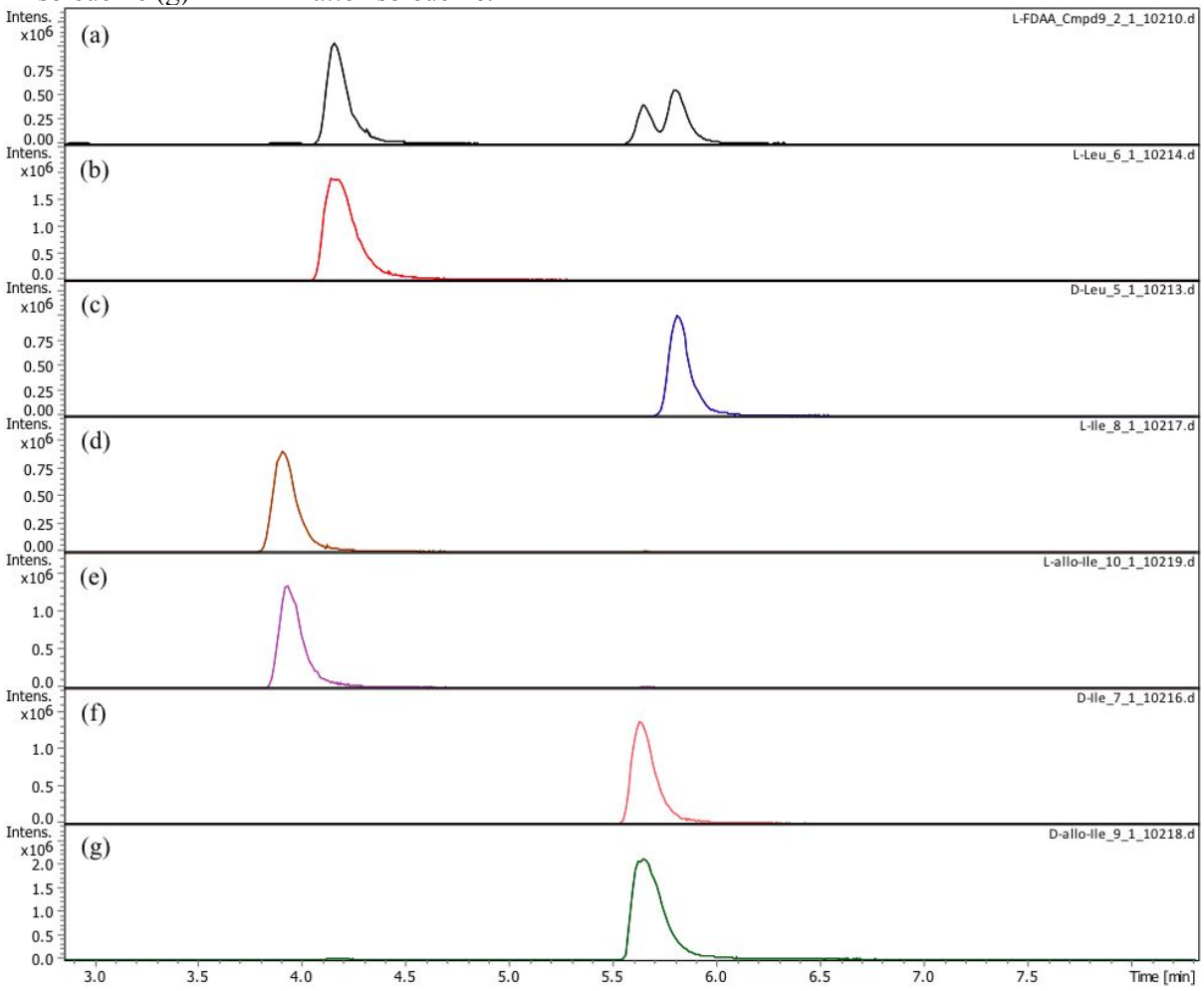


Figure S20D. Extracted ion chromatogram (EIC *m/z* 358.10) for (a) orfamide N derivatized hydrosylate (b) FDAA-L-serine (c) FDAA-D-serine.

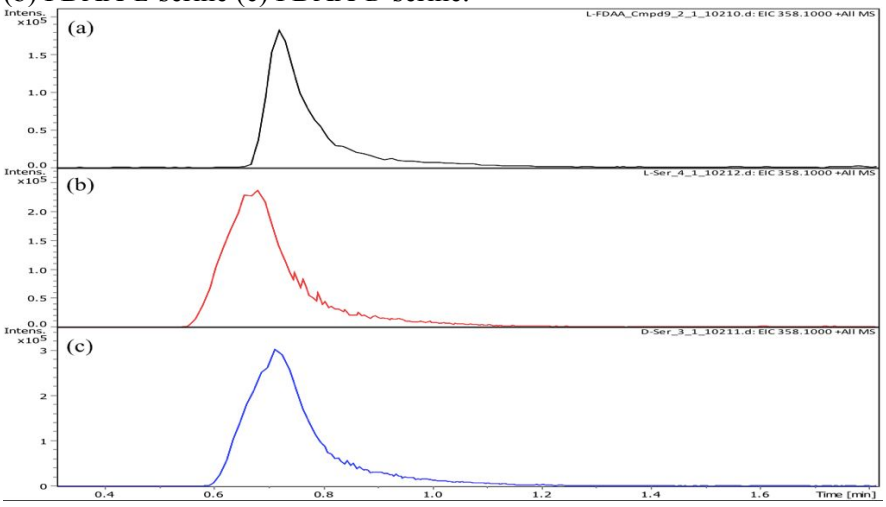


Figure S20E. Extracted ion chromatogram (EIC *m/z* 400.11) for (a) orfamide N derivatized hydrosylate (b) FDAA-L-glutamic acid (c) FDAA-D-glutamic acid.

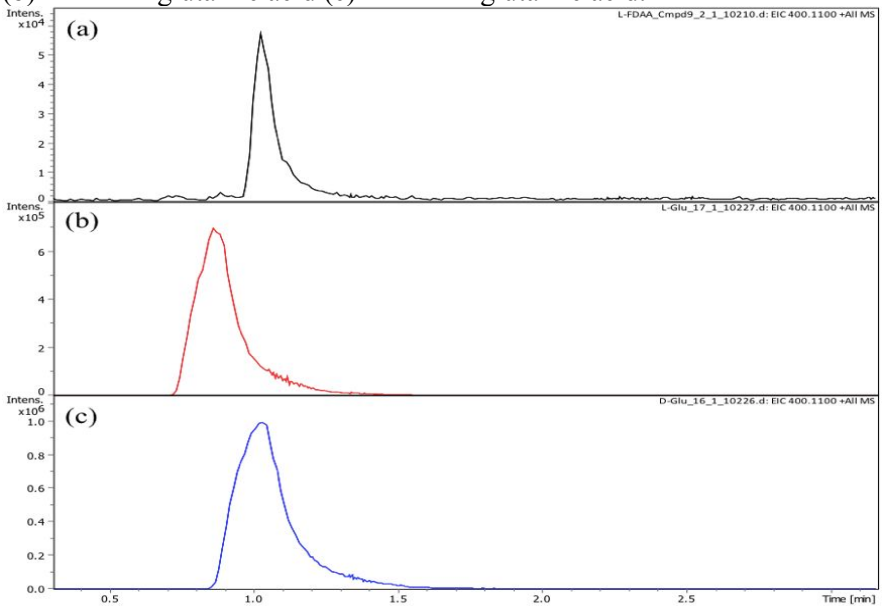


Figure S20F. Extracted ion chromatogram (EIC *m/z* 372.11) for (a) orfamide N derivatized hydrosylate (b) FDAA-L-threonine (c) FDAA-L-*allo*-threonine (d) FDAA-D-threonine (e) FDAA-D-*allo*-threonine.

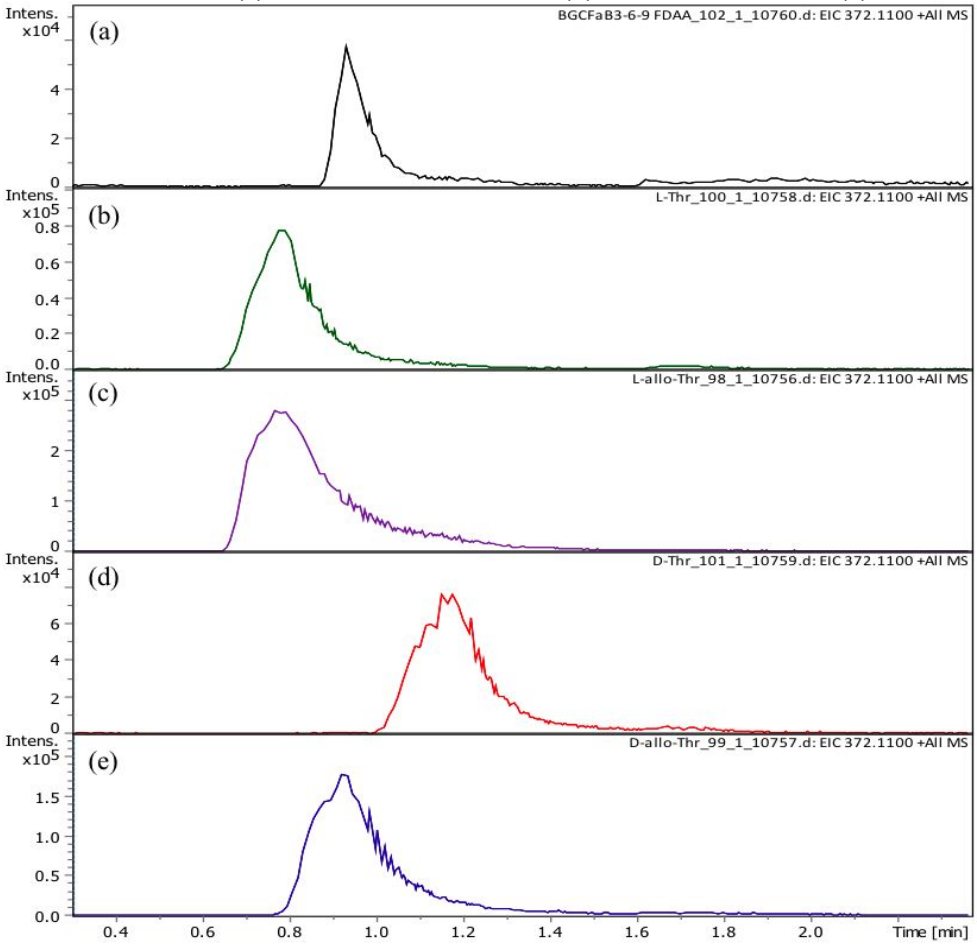


Figure S21. Chiral LC-MS analysis to determine the configuration of the β-hydroxy acid in 1.

Figure S21A. The retention time of 3-OH-hexadecanoic acid synthesized from the free 3-OH-hexadec-9-enoic acid isolated from orfamide N (**1**) was compared to a 3*R/S*-OH-hexadecanoic acid standard. Retention times were measured by UPLC-DAD-ESIMS extracted ion chromatograms at *m/z* 271.2544.

	Retention time (min)		**Measured**	**Assignment**
	S	*R*		
3-OH-hexadecanoic acid	17.8	18.4	18.3	*R*

Figure S21B. UPLC-DAD-ESIMS extracted ion chromatogram at *m/z* 271.544 for (a) 3*R/S*-OH-hexadecanoic acid standard (b) 3*R*-OH-hexadecanoic acid synthesized from the free 3*R*-OH-hexadec-9-enoic acid isolated from orfamide N (**1**).

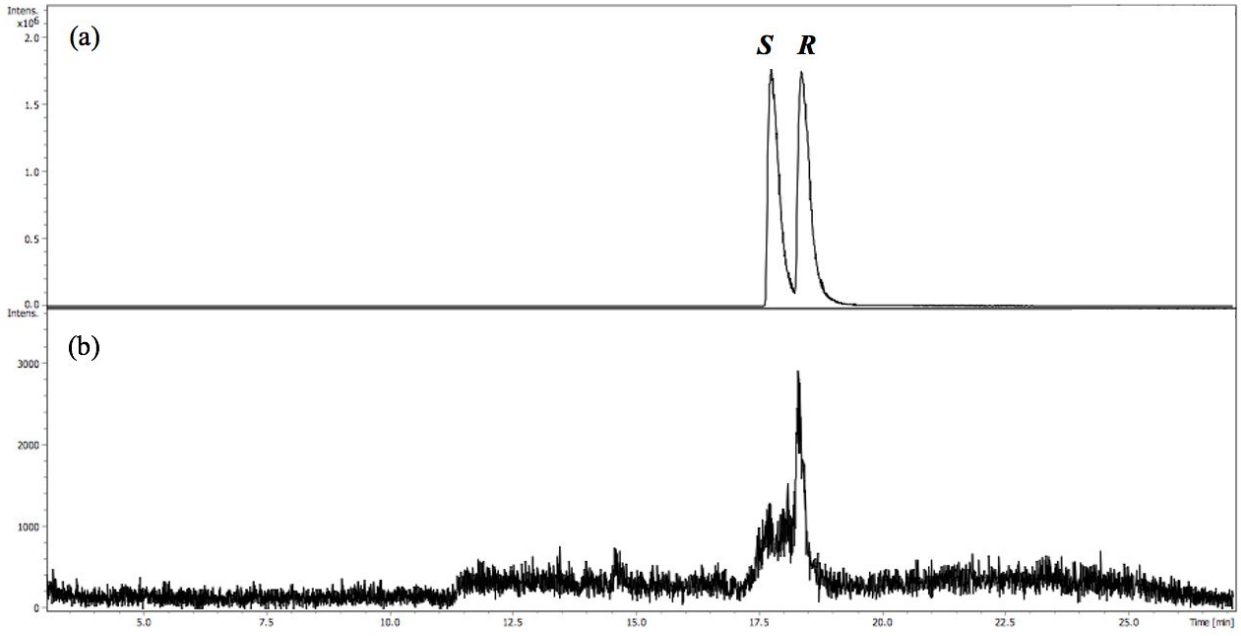


Figure S22. Results of antibacterial assay. Known CFU/mL of bacteria [*Enterobacter aerogenes* (ATCC 13048), E*nterococcus faecium* (ATCC 35667), *Staphylococcus aureus* (ATCC 29213), and *Escherichia coli* (ATCC 25922)] and yeast [*Candida albicans* (ATCC 18804)] were diluted in Mueller Hinton II Broth (Cation- Adjusted) (CAMH) and Sabouraud Dextrose Broth (SDB) respectively to achieve final inoculum density of 2 × 105 CFU/mL. 100 microliters of diluted cell cultures were added per well containing equal volumes of the same media with the test compound and were incubated overnight at 37 °C. Optical density was measured at 570 nm. Moxifloxacin and doxycycline were used as bacterial positive controls, while amphotericin B was used for *C. albicans*. MICs were determined as described by the European Committee on Antimicrobial Susceptibility Testing (EU-CAST).

	MIC (mg/mL)				
Compound ID	***Staphylococcus aureus* (ATCC 29213)**	***Enterobacter aerogenes* (ATCC 13048)**	***Enterococcus faecium* (ATCC 35667)**	***Escherichia coli* (ATCC 25922)**	***Candida albicans* (ATCC 18804)**
Orfamide N (1)	>100	>100	>100	>100	>100
Standard drugs	**MIC (mg/mL)**				
Moxifloxacin	<0.195	<0.195	1.56	<0.195	
Doxycycline	<0.195	2.84	<0.195	0.48	
Amphotericin B					<0.195

Reference:

(1) European Committee on Antimicrobial Susceptibility Testing (EUCAST), 2022, https://www.eucast.org/ast_of_bacteria/mic_determination.