Supplementary Information

A conserved nonribosomal peptide synthetase in *Xenorhabdus bovienii* produces citrulline-functionalized lipopeptides

Jhe-Hao Li,[1,2] Wooyoung Cho,[1,2] Randy Hamchand,[1,2] Joonseok Oh,[1,2] and Jason M. Crawford[1,2,3,*]

[1] Department of Chemistry, Yale University, New Haven, CT 06511, USA.

[2] Institute of Biomolecular Design & Discovery, Yale University, West Haven, CT 06516, USA.

[3] Department of Microbial Pathogenesis, Yale University School of Medicine, New Haven, CT 06536 USA.

E-mail: jason.crawford@yale.edu

Table of Contents

Supplementary Methods

GPCR PRESTO-Tango Assay

PRESTO-Tango assay was performed according to the protocol reported by Kroeze *et al.* (Kroeze et al., 2015). Briefly, HTLA cells maintained in DMEM containing 10% heat-inactivated fetal bovine serum (HI-FBS) and 1% penicillin/streptomycin were seeded in 96-well tissue culture plates and transfected with 200 ng of C3AR1-Tango plasmid (Addgene plasmid #66231) for 20 h. The medium was replaced with DMEM containing 20 mM HEPES and 1% penicillin/streptomycin 2 h prior to treatment with compounds **5** or **6** (100 μM, 33 μM, 11 μM, 3.7 μM, 1.2 μM, 0.41 μM) in DMSO (final concentration 0.125%), complement protein C3a (10 μM, 3.3 μM, 1.1 μM, 0.37 μM, 0.12 μM, 0.041 μM) in DMSO (final concentration 0.125%), or 0.125% DMSO solvent vehicle. After incubation for 20 h, each well was reacted with 50 μl of Bright-Glo solution (Promega), diluted 20-fold in DPBS with 20 mM HEPES, for 20 min. Luminescence was measured with a Perkin Elmer EnVision 2100 plate reader.

ISG, NF-κB, LDH cytotoxicity assays

THP1-Dual™ (Invivogen) reporter cells contain two inducible promoters that express the secreted luciferase or secreted embryonic alkaline phosphatase (SEAP) reporter genes upon stimulation of the interferon regulatory factor (IRF) or NF-κB pathways, respectively. THP1-Dual™ cells maintained under manufacturer's instructions in RPMI medium (Gibco 11875093) with 10% HI-FBS (Gibco) and 1% penicillin/streptomycin (Gibco) were seeded into 96-well plates and differentiated into a macrophage-like state by incubation with 50 nM phorbol 12-myristate 13-acetate (PMA) (Promega, V1171-5mg) for 72 h. The cells were then treated with compound **5** or **6** (100 μM, 33 μM, 11 μM, 3.7 μM, 1.2 μM, 0.41 μM) in DMSO (final concentration 0.1%), positive control (transfected poly dA:dT for ISG response, 100 ng ml^{-1} lipopolysaccharide (LPS)

for NF-κB response), or solvent vehicle negative control (0.1% DMSO) and incubated for 16-20 h. The supernatants were collected and incubated with the luciferase detection agent QUANTI-Luc™ (Invivogen) to observe ISG responses or with the SEAP detection agent QUANTI-Blue™ (Invivogen) to detect NF-κB responses. The NF-κB assay was also carried out without the differentiation step, as PMA treatment generated higher background readings due to the activation of NF-κB.

Possible cell cytotoxicity of compounds **5** and **6** was measured using CyQUANT™ LDH Cytotoxicity Assay Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The spontaneous LDH release was measured using the supernatant of ~$1x10^5$ cells incubated with 0.1% DMSO and the maximum LDH release was measured by lysing cells grown in the same conditions. Supernatants of the PMA-treated or untreated THP1-Dual™ cells challenged with **5** or **6** (100 μM, 33 μM, 11 μM) in DMSO (final concentration 0.1%) were used for LDH quantification. Absorbance measurements at 490 nm and 680 nm were taken on a Perkin Elmer EnVision 2100 plate reader.

Antimicrobial assays

Compounds **5** and **6** were dissolved in DMSO to a concentration of 50 mM. Each compound was diluted in the corresponding medium to yield final concentrations of 500, 250, 125, 62.5, 31.25, 15.625, 0 μM (final DMSO concentration 1%) in triplicate wells of a 96-well microplate. *Escherichia coli* DH5α was grown in LB broth overnight in a shaking incubator at 250 RPM and 37^{o}C. *Bacillus subtilis* and *Saccharomyces cerevisiae* were grown overnight in LB and YPD media, respectively, in a shaking incubator at 250 RPM and 30^{o}C. The bacterial/yeast cultures were diluted to (OD_{600}) of 0.5, then diluted further to yield 1e6 cells ml^{-1}. 50 μl of the diluted bacterial cultures was added to each well containing 50 μl of compound or control. The microplate was

incubated overnight in a stationary 37°C (*E. coli*) or 30°C (*B. subtilis*, *S. cerevisiae*) incubator and the OD_{600} of each well was measured using a Perkin Elmer EnVision 2100 plate reader.

Table S1 *XBJ1_2367* homologs in *X. bovienii* strains. Genes are shown with magnifying genomes (MAGE) accession codes.

***X. bovienii* strain name**	**XBJ1_2367 orthologue**	**Amino acid identity (%)**
puntauvense	XBP1v2_400023	92.53
felitiae France	XBXBFFR1v2_630020	92.47
Jollieti	XBJ2v2_1260017	100
oregonense	XBO1v2_1300059	92.11
kraussei Quebec	XBKQ1v2_850017	92.09
Intermedium	XBI1v2_1260106	93.04
CS03	XBW1v5_2161	92.47
	XBW1v5_2162 (terminal PCP domain)	
felitiae Florida	XBFFL1v2_1440021	92.47
kraussei Becker Underwood	XBKB1v2_2960007	92.23
felitiae Moldova	XBFM1v2_1510028	92.44

Table S2 *XBJ1_2367* homologs in *Xenorhabdus* species and their adenylation domain specificity codes.

Species	NCBI Accession	Amino Acid Identity	A1 domain	A2 domain	A3 domain
X. beddingii	WP_086111596.1	68.0%	D A W L L G A V	D V W H F S L I	D I S N I G G V
X. bovienii SS-2004	CBJ81493.1	100.0%	D A W L I G A V	D L Y N N A L -	D I S N I G A I
X. budapestensis	WP_099134826.1	68.5%	D A W I I G V I	D V W H L S L I	D I S N I G A V
X. doucetiae	WP_071827274.1	73.4%	D G W L I G A V	D L Y N N A L T	D I S N I G A V
X. eapokensis	WP_074023869.1	68.3%	D A W L L G A V	D V W F L S L I	D I S N I G A I
X. ehlersii	WP_099131307.1	69.3%	D A W L L G A V	D V W Y L S L I	D I S N I G A I
X. griffiniae [a]	WP_053014213.1	68.4%	D A W L M G A V	D V W Y L S L I	D I S N I G A I
X. hominickii	WP_069316711.1	69.4%	D V Q F I A H -	D A W X V A A I	D I S N I G A V
X. innexi	WP_086956251.1	65.0%	D A W L L G A V	D V W H L S L I	D I S N I G A I
X. ishibashii	WP_099116303.1	68.6%	D A W L L G A V	D V W Y L S L I	D I S N I G A I
X. japonica	WP_092519020.1	70.8%	D A W L L G A V	D V W H L S L I	D I S N I G A V
X. khoisanae SF87	WP_084728559.1	69.7%	D A W L M G A V	D V W Y L S L V	D I S N I G A I
X. kozodoii	WP_099141660.1	68.3%	D A W L L G A I	D V W Y L S L I	D I S N I G A I
X. mauleonii	WP_092507495.1	68.1%	D V W F I - G I	D L Y N N A L T	D I S N I G A V
X. miraniensis	WP_099114446.1	70.2%	D A W L I G A V	D V W Y L S L V	D I S N I G A I
X. nematophila ATCC19061	WP_050986630.1	70.2%	D A W L L G A V	D L Y N N A L T	D I S N I G A V
X. poinarii	WP_052708256.1	73.7%	D A W I L G A I	D L Y N N A M -	D I S N I G A V
X. stockiae	WP_099125076.1	64.9%	D A W L L G A I	D V W H L S L I	D I S N I G A I
X. szentirmaii DSM16338	PHM32961.1	68.6%	D V W F I - G V	D L Y X N A L T	D I S N I G A V
X. thuongxuanensis	WP_074019651.1	68.7%	D A W L L G A V	D V W Y L S L I	D I S N I G A I
X. vietnamensis	WP_086108473.1	70.7%	D A W L L G A I	D V W H L S L I	D I S N I G A V
Xenorhabdus sp. NBAII XenSa04	WP_052189507.1	67.4%	D A W I I G A I	D V W H F S L I	D I S N I G A V
Xenorhabdus sp. KK7.4	WP_099120489.1	64.9%	D S W L L G A I	D V W H L S L I	D I S N I G A I
Xenorhabdus sp. KJ12.1	WP_099109723.1	65.0%	D A W L L G A I	D V W H L S L I	D I S N I G A I

[a] Gene assigned to *X. griffiniae* based on Magnifying Genome (MaGe) platform.

Table S3 Primers used in this study

Primer Name	Sequence
ACYC_PacI	GTAAAAAATTAATTAATCTCCTTCTTATACTTAACTAATATA CTAAGATGG
ACYC_HindIII	GTAAAAAATAAGCTTAAAAAACTCGAGTAACCTAGGCTGCTG
2367_I_PacI	GTAAAAAATTTAATTAAATGTCACACAGCCAAAATCG
2367_I_HindIII	GTAAAAAATAAGCTTGTTGTACCCAGGTG
2367_II_HindIII	GTAAAAAATAAGCTTTCCGCCCTGAA
2367_II_XhoI	GTAAAAAATCTCGAGTTATTTAGAATAAAAATTCTTTGCTAATT TTGT

Table S4 *XBJ1_2367* dependent ions in *E. coli* BAP1 and *X. bovienii*

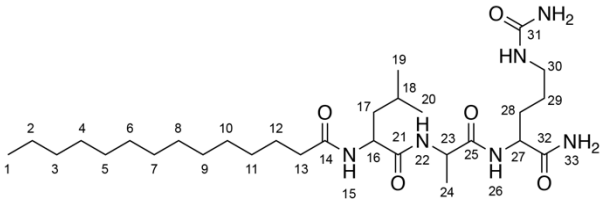
m/z	Retention time (min)[a]	AUC [b] in *E. coli* BAP1	AUC in *X. bovienii* SS-2004
567.4221	22.1	1.16E+06	1.01E+07
568.4077	22.7	1.26E+06	3.66E+06
569.4403	24.1	3.55E+06	4.56E+07
570.4243	24.7	6.62E+06	3.28E+07
585.4332	21.5	1.37E+06	2.76E+06
586.4173	22.1	6.48E+05	8.13E+05
595.4544	25.0	1.82E+06	4.23E+06
596.4384	25.5	4.90E+06	5.08E+06
597.4714	27.5	1.30E+06	5.23E+06
598.4537	27.5	4.29E+06	0.00E+00
598.4537	28.0	2.65E+06	6.03E+06
624.4712 [d]	28.2	2.35E+06	0.00E+00
624.4712 [d]	28.6	1.48E+06	1.56E+06
626.4864	30.8	1.43E+06	0.00E+00
652.4999	31.3	6.11E+05	0.00E+00
541.4059	21.1	5.55E+05	3.09E+06
542.3914	21.6	2.61E+05	5.66E+05
584.4384	26.3	2.44E+06	3.57E+06
612.4377	22.7	2.23E+04	0.00E+00
612.4339	23.0	1.19E+05	0.00E+00
614.4481	24.5	7.06E+05	0.00E+00
642.4923	27.5	5.03E+05	0.00E+00
640.4648	24.5	9.81E+04	0.00E+00
557.4021	18.8	8.16E+04	5.83E+05
583.4171	20.0	9.35E+04	1.90E+05
584.4021	20.5	1.65E+04	3.66E+04
598.4648	27.6	4.10E+05	3.61E+05
583.4550	25.8	1.36E+06	4.10E+06

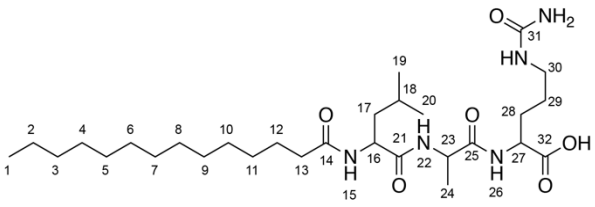
[a] Retention time was determined with LC-MS coupled to a Kinetex 5µ C18 100 Å column (250 × 4.6 mm) with a water:acetonitrile gradient from 5 to 100% in 30 minutes.
[b] Area under curve.
[c] 5-mL cultures of *E. coli* and *X. bovienii* SS-2004 were extracted to compare the relative abundance.
[d] These two features are distinct.

Table S5 ^{1}H and ^{13}C NMR assignments of bovienimide A (**5**) and bovienimide B (**6**)

	5[a]		**6**[b]	
position	δ_H, mult. (*J* in Hz)	δ_C	δ_H, mult. (*J* in Hz)	δ_C
1	0.82, t (6.9)	14.38, CH_3	0.90, t (7.0)	13.00, CH_3
2-10[c]	1.20-1.33, m	29.39, CH_2	1.26-1.34, m	22.3, CH
11	1.19, m	29.29, CH_2	1.59, m	25.7, CH_2
12	1.44, m	29.67, CH_2	1.32, m	28.8, CH_2
13	2.05, m 2.09, m	35.48, CH_2	2.21, m	35.10, CH_2
14		173.02, C		175.12, C
15	7.96, d (7.4)		8.09, d (6.4)	
16	4.19, m	51.81, CH	4.26, m	52.36, CH
17	1.39, m	40.88, CH_2	1.58, m 1.54, m	39.95, CH_2
18	1.54, m	24.64, CH	1.66, m	24.54, CH
19	0.80, d 6.6)	23.35, CH_3	0.94, d (6.5)	20.63, CH_3
20	0.85, d (6.6)	22.01, CH_3	0.98, d (6.5)	21.90, CH_3
21		172.77, C		174.04, C
22	8.10, d (7.4)		8.37, d (7.2)	
23	4.21, quin (7.4)	49.31, CH	4.35, quin (7.2)	49.03, CH
24	1.16, d (7.4)	18.53, CH_3	1.36, d (7.2)	16.57, CH_3
25		172.37, C		173.39, C
26	7.86, d (8.4)		7.98, d (8.1)	
27	4.11, td (4.82, 8.4)	52.55, CH	4.38, m	51.87, CH
28	1.44, m 1.64, m	29.64, CH_2	1.75, m 1.91, m	28.30, CH_2
29	1.29, m	27.04, CH_2	1.49, m	26.12, CH_2
30	2.90, m	29.13, CH_2	3.08, td (2.64, 6.83)	38.58, CH_2
31		159.17, C		160.65, C
32		174.03, C		175.53, C
33	7.00, d (2.17) 7.24, d (2.28)			

[a]DMSO-d_6, [b]methanol-d_4, [c]methylenes comprising C2-10 are indistinguishable





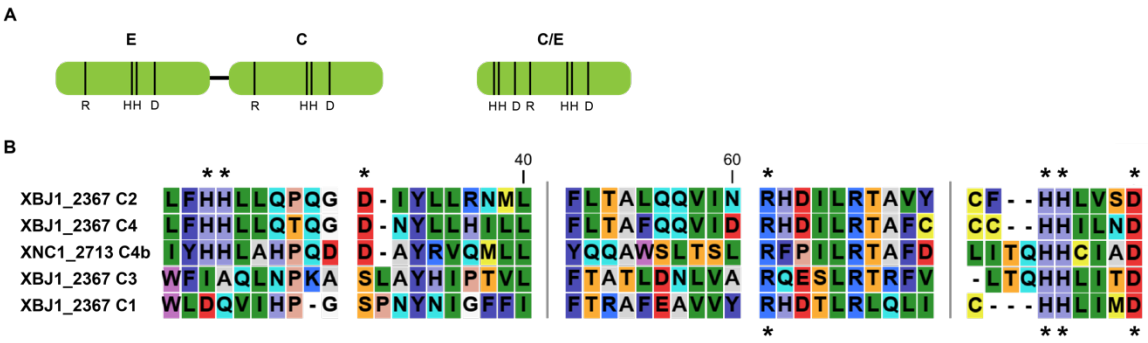


Figure S1 Sequence alignment of C/E domain key residues. (A) Schematic representation of key amino acids in epimerization-condensation didomain proteins and condensation/epimerization dual functional domain. (B) Partial alignment of condensation domains in *XBJ1_2367* and representative C/E domain in xenematide biosynthesis cluster (*XNC1_2713* C4b). Asterisks stand for conserved catalytic residues.

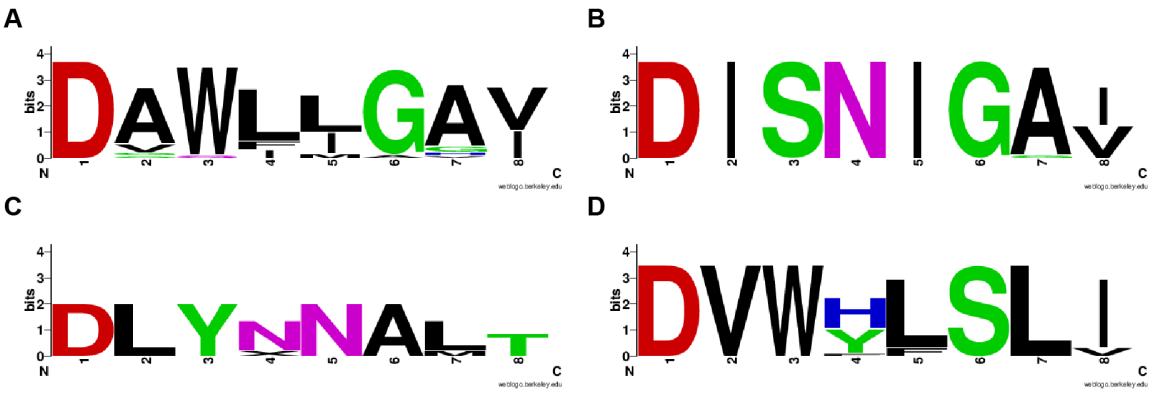


Figure S2 The sequence logo showing the consensus sequences of specificity codes in the A domains. **A,** A1-domain. **B,** A3-domain. **C,** Group I A2-domain. **D,** Group II A2-domain.

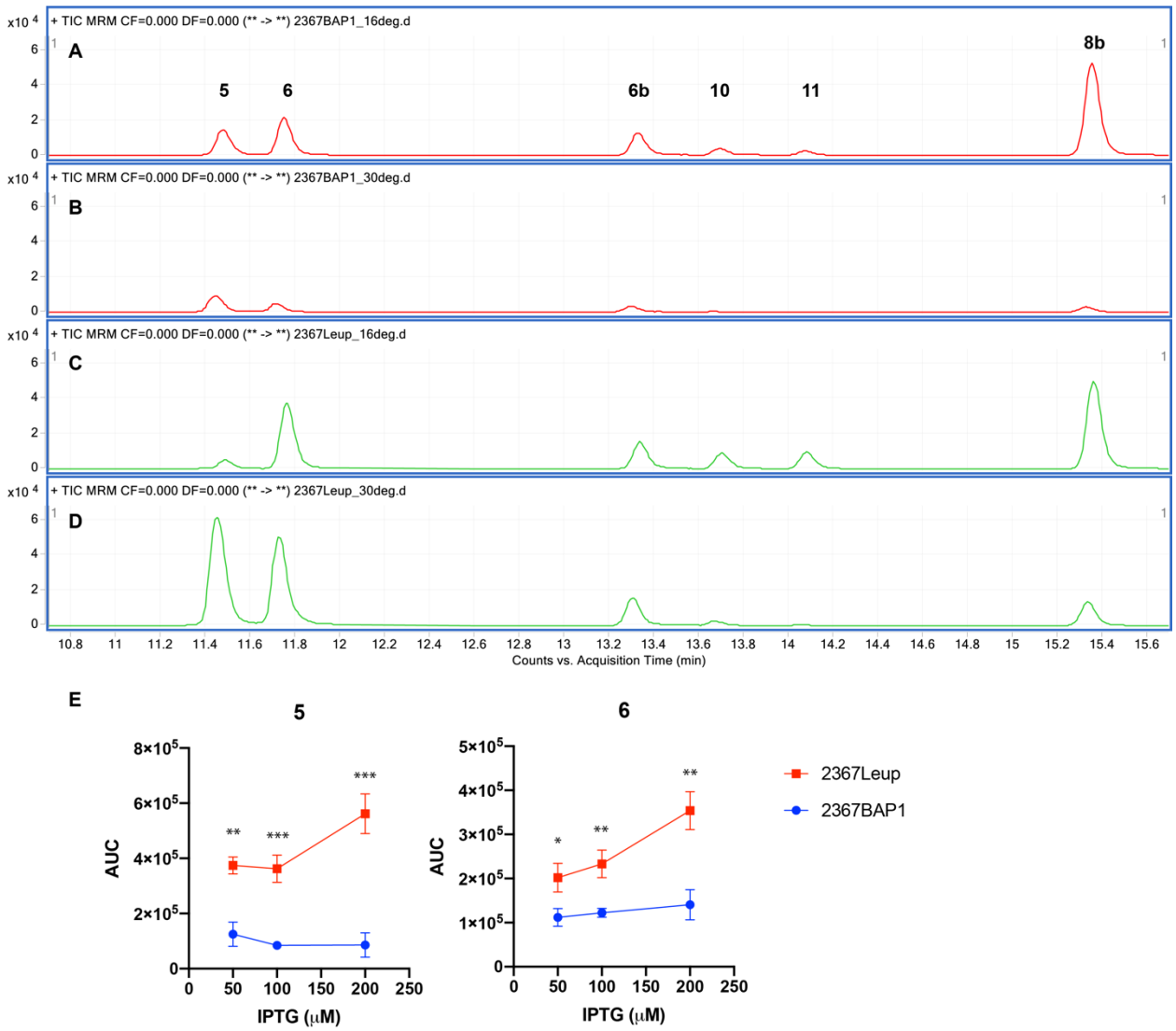


Figure S3 Co-expression of pACYC-*XBJ1_2367* and pCDF-*Leup* resulted in higher production of **5** and **6** at 30°C. 5-mL LB culture extracted with ethyl acetate was analyzed by triple quadrupole with dynamic MRM mode with a linear gradient using water and acetonitrile solvent systems containing 0.1% formic acid at 0.3 mL/min, 0-15 min, 20 to 80% acetonitrile. **A**, pACYC-*XBJ1_2367* only expressed at 16°C. **B**, pACYC-*XBJ1_2367* only expressed at 30°C. **C**, pACYC-*XBJ1_2367* co-expressed with pCDF-*Leup* at 16°C. **D**, pACYC-*XBJ1_2367* co-expressed with pCDF-*Leup* at 30°C. **E,** Dose dependence production of bovienimides **5** and **6** against different concentrations of IPTG. Relative abundance was determined by comparing 5-mL cultures with the same extraction and injection methods. *n* =3 biological replicates. *p < 0.05, **p < 0.01, ***p < 0.001. Two-tailed t test.

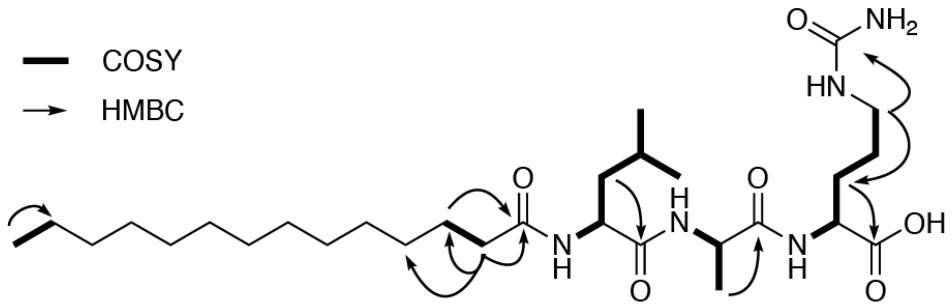


Figure S4 Key COSY and HMBC correlations of bovienimide B (**6**).

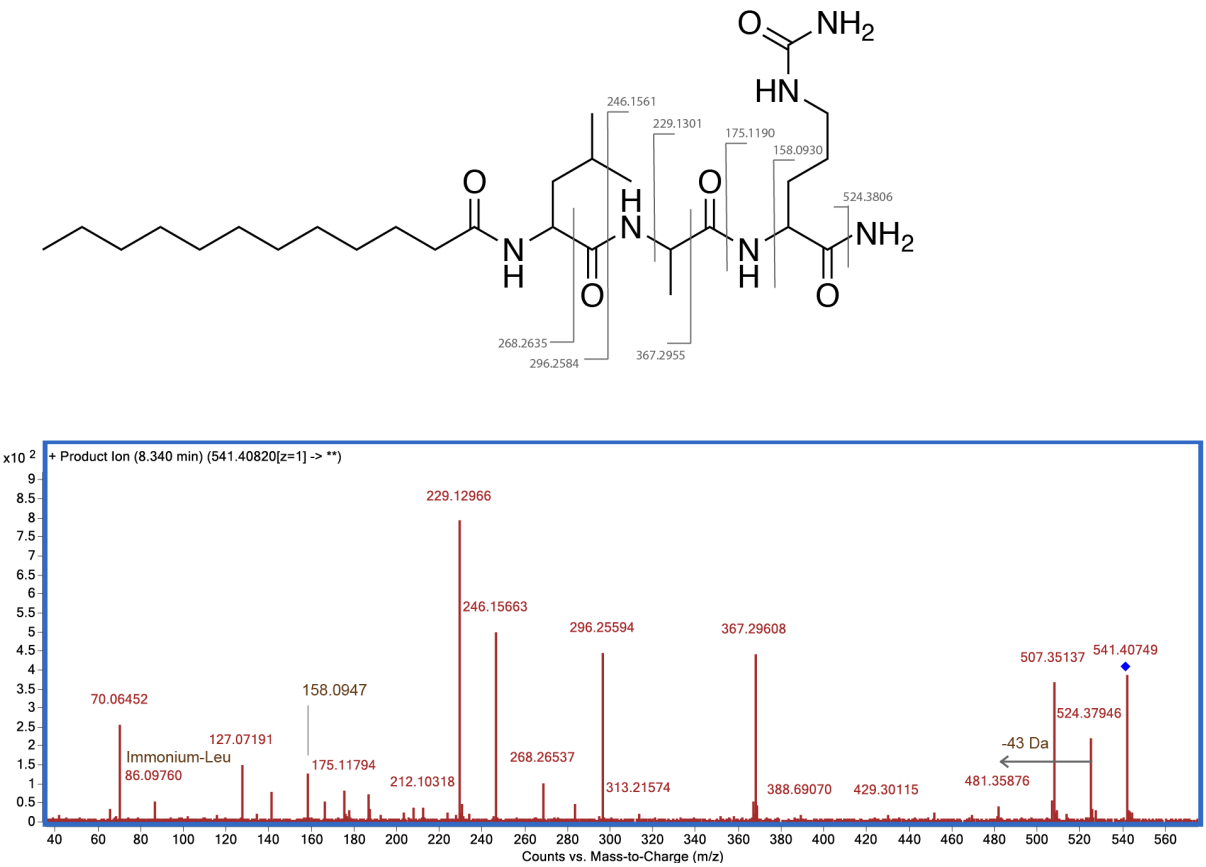
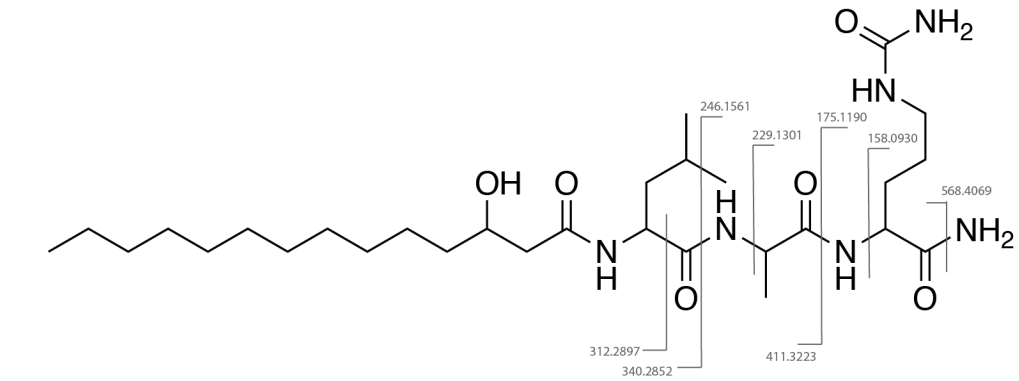


Figure S5 Tandem MS assignment of proposed compound **1**. Calculated exact mass of peptide fragment ions are shown on the structure.



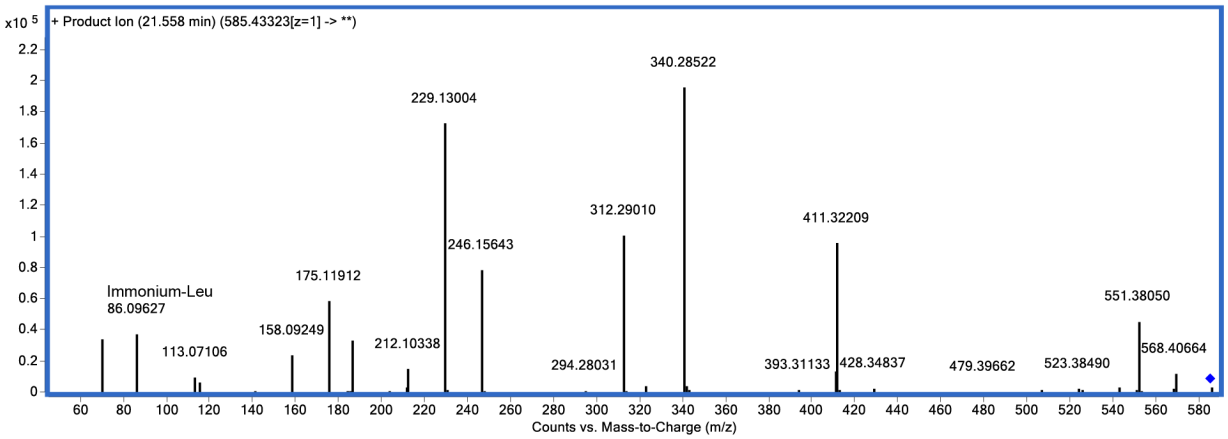


Figure S6 Tandem MS assignment of proposed compound **2**. Calculated exact mass of peptide fragment ions are shown on the structure. The position and stereochemistry of the hydroxyl group on the acyl chain was not determined.

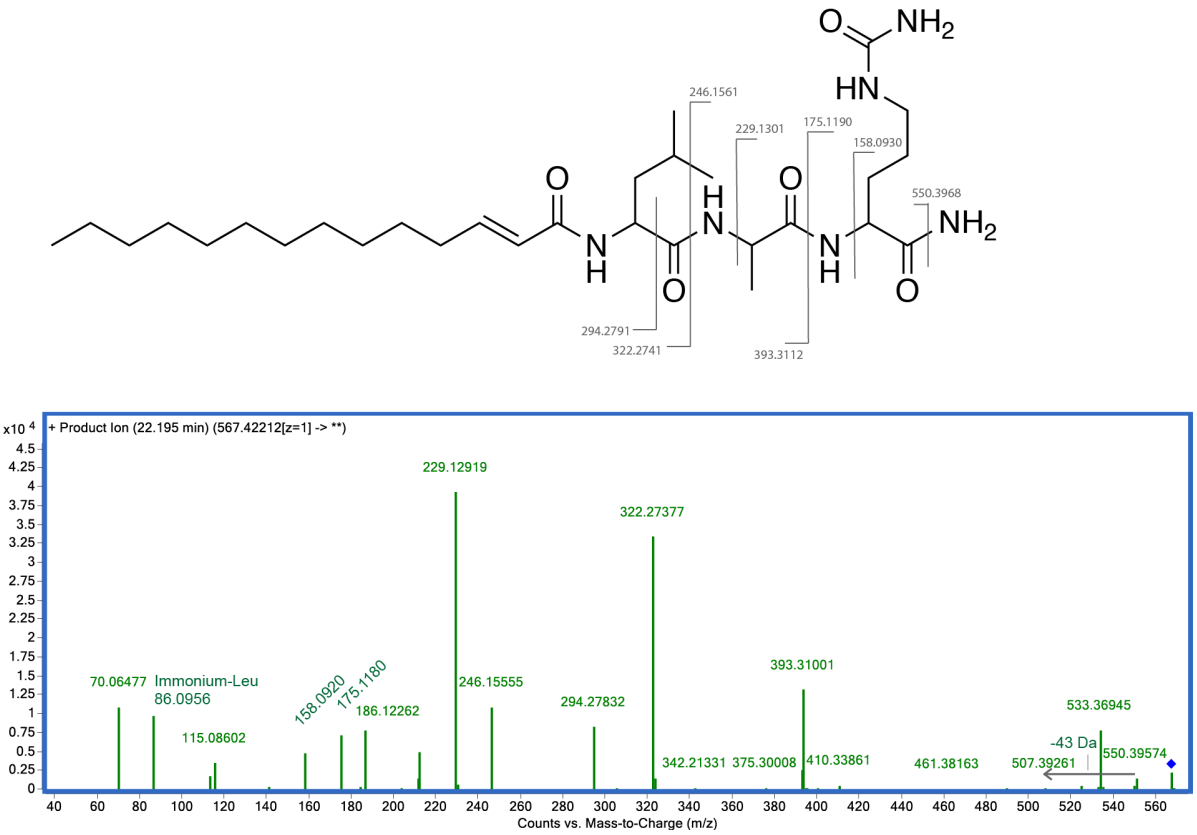


Figure S7 Tandem MS assignment of proposed compound **3**. Calculated exact mass of peptide fragment ions are shown on the structure. Double bond position and configuration was not determined.

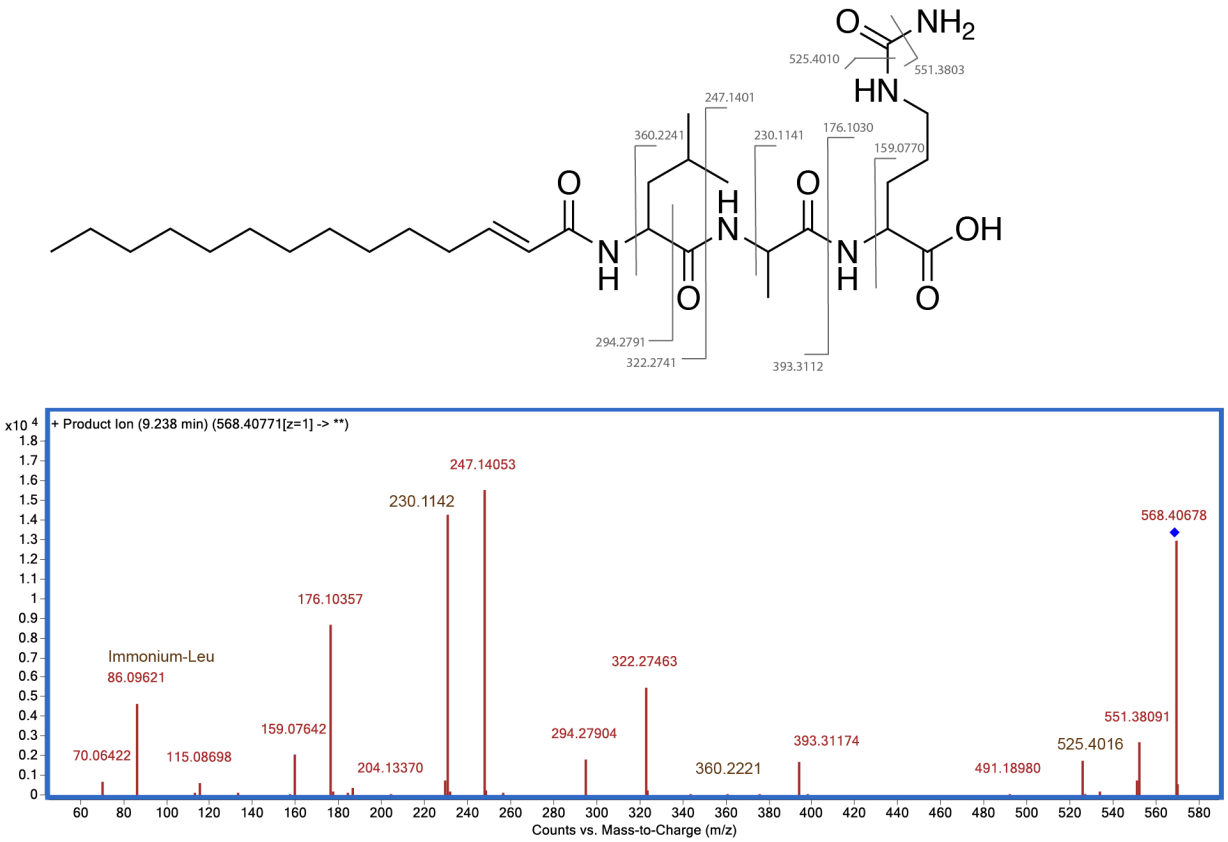


Figure S8 Tandem MS assignment of proposed compound **4**. Calculated exact mass of peptide fragment ions are shown on the structure. Double bond position and configuration was not determined.

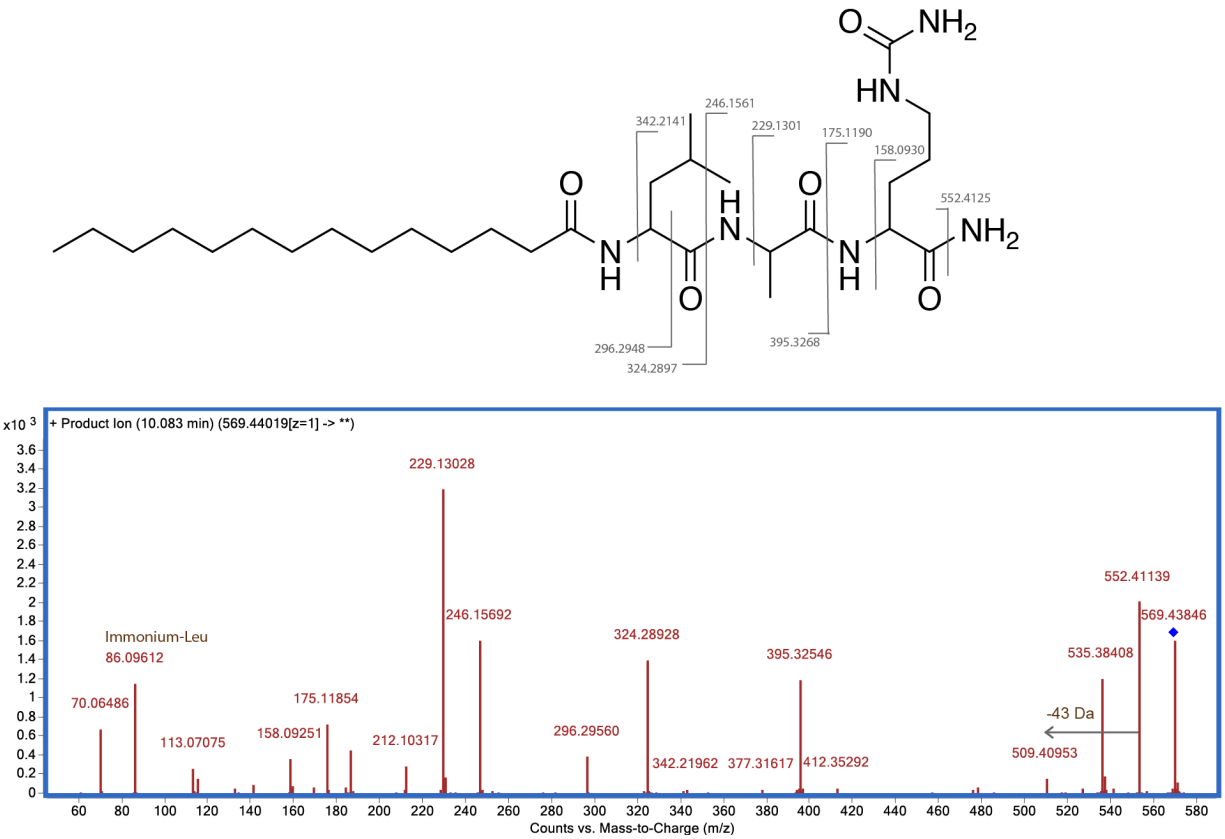


Figure S9 Tandem MS assignment of bovienimide A (**5**). Calculated exact mass of peptide fragment ions are shown on the structure.

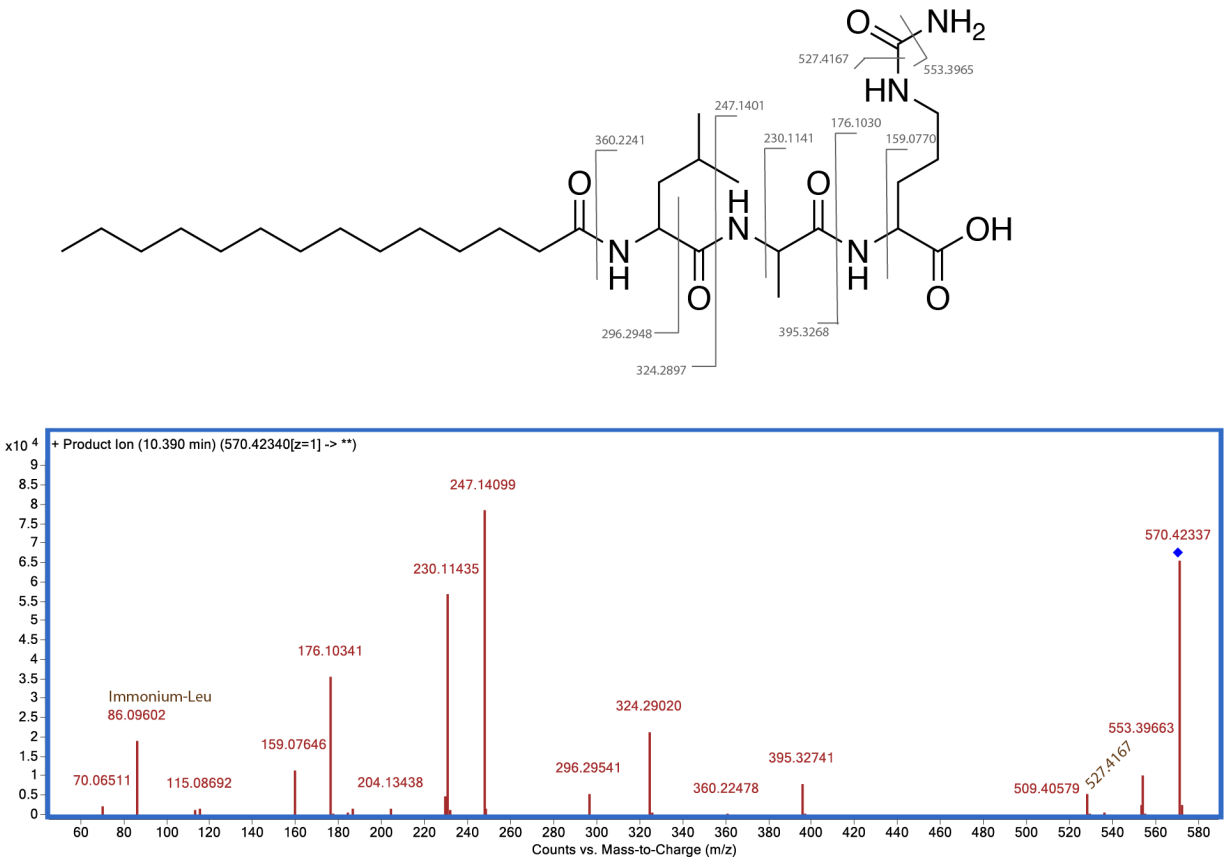


Figure S10 Tandem MS assignment of bovienimide B (**6**). Calculated exact mass of peptide fragment ions are shown on the structure.

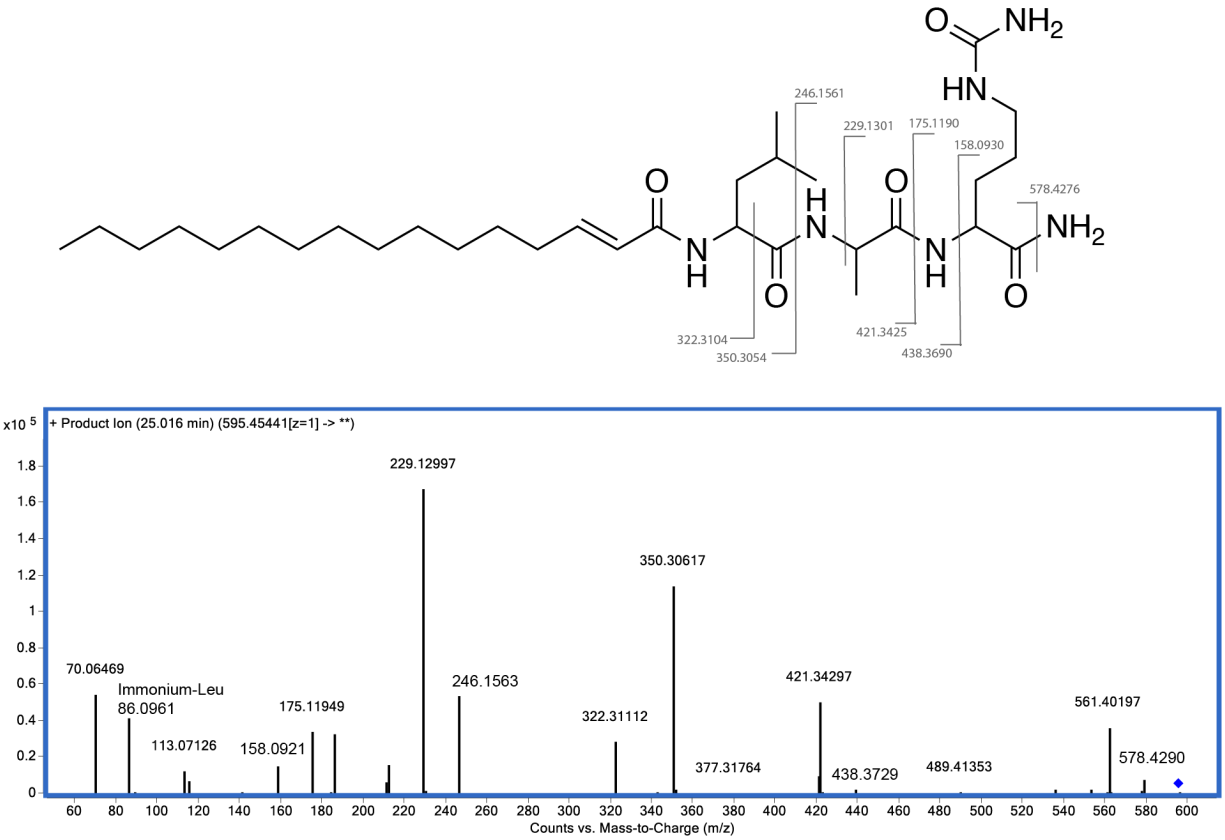


Figure S11 Tandem MS assignment of proposed compound **7**. Calculated exact mass of peptide fragment ions are shown on the structure. Double bond position and configuration was not determined.

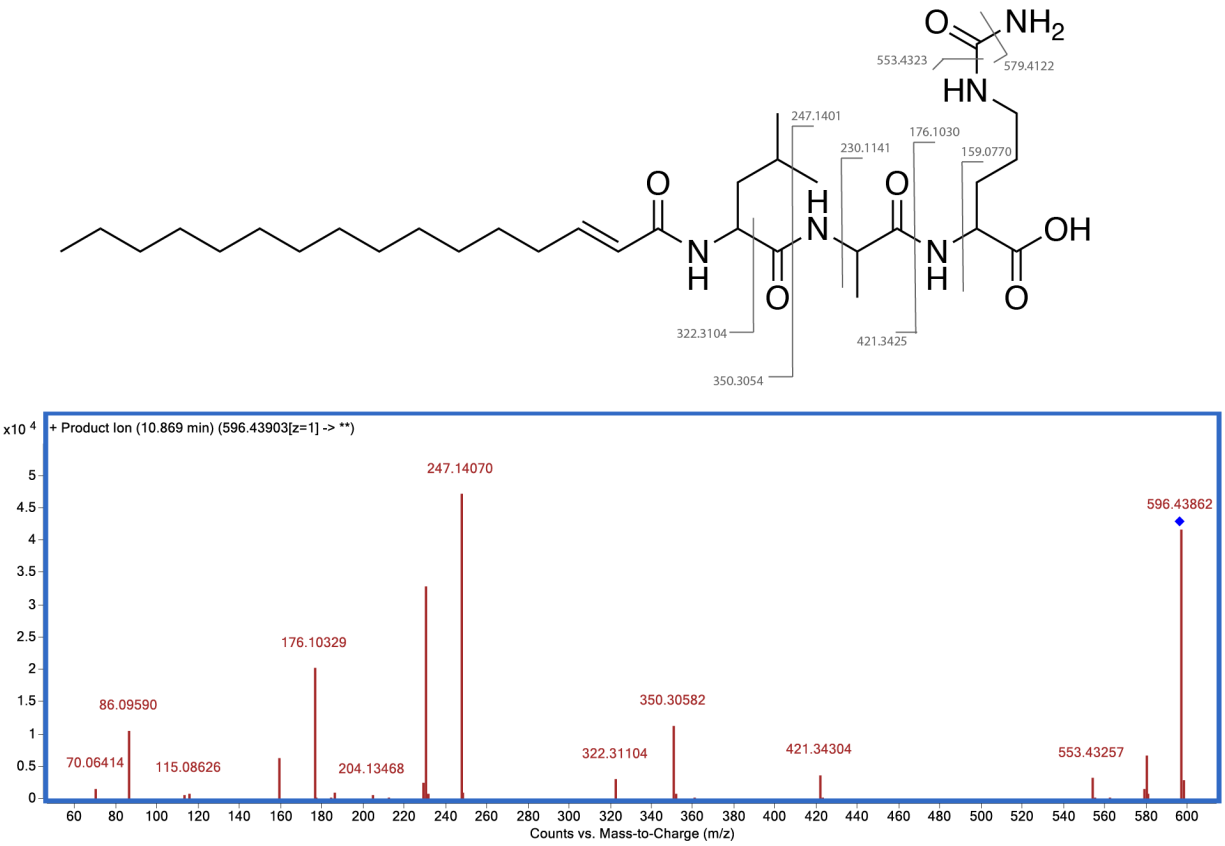


Figure S12 Tandem MS assignment of proposed compound **8**. Calculated exact mass of peptide fragment ions are shown on the structure. Double bond position and configuration was not determined.

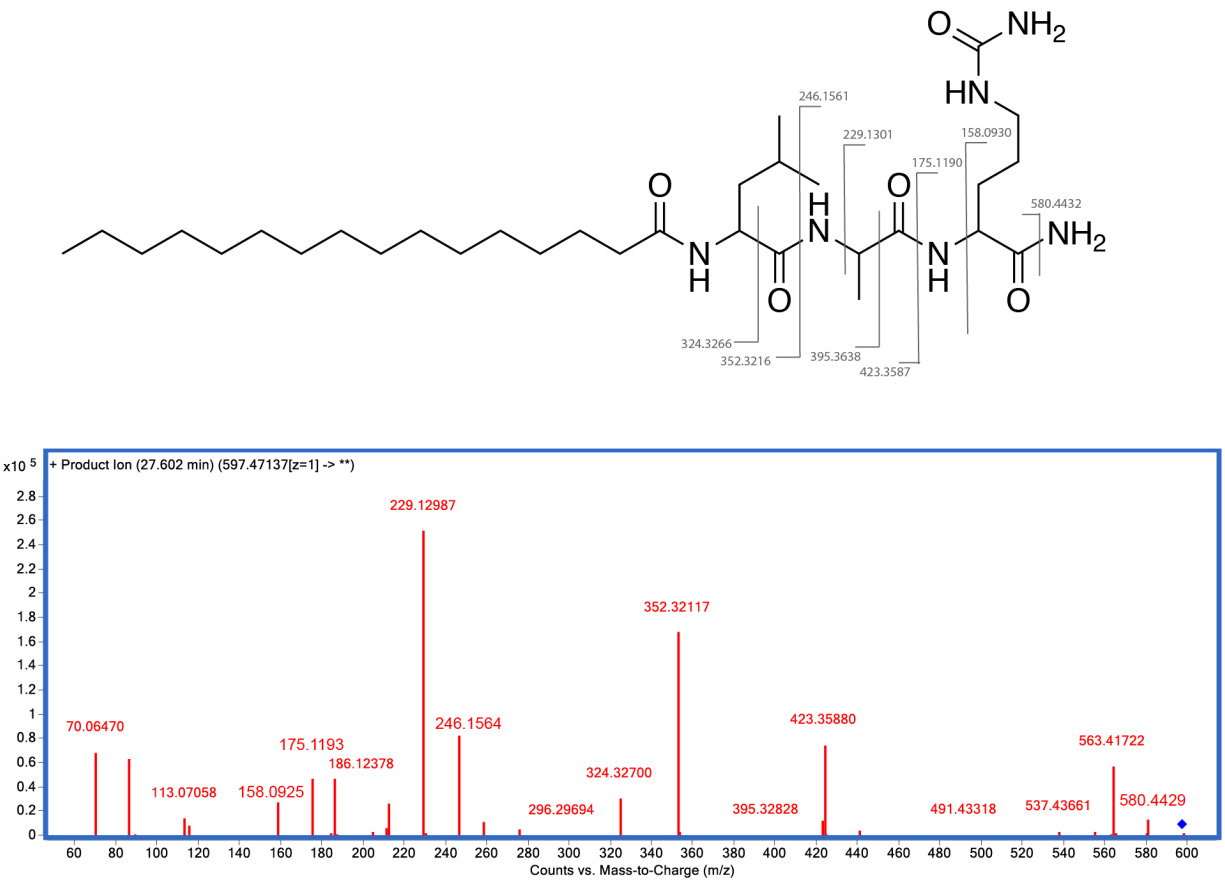


Figure S13 Tandem MS assignment of proposed compound **9**. Calculated exact mass of peptide fragment ions are shown on the structure.

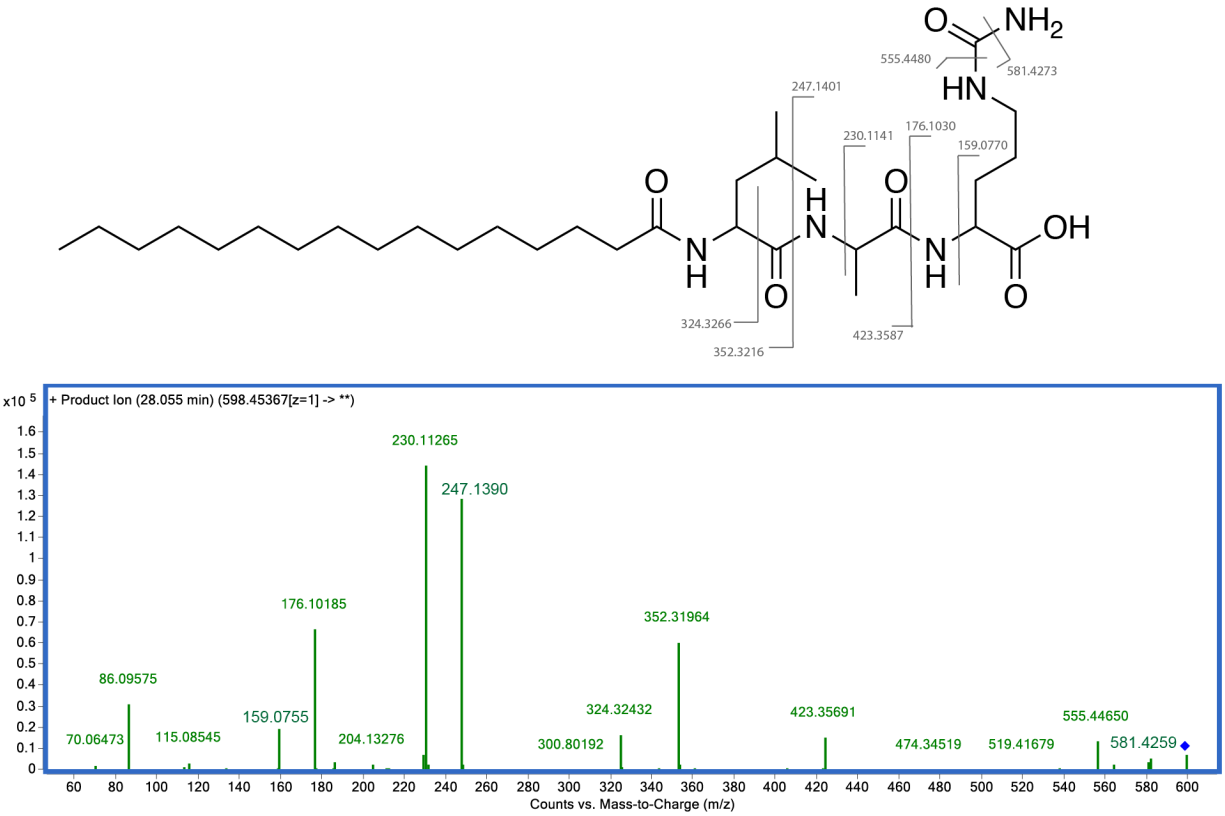


Figure S14 Tandem MS assignment of proposed compound **10**. Calculated exact mass of peptide fragment ions are shown on the structure.

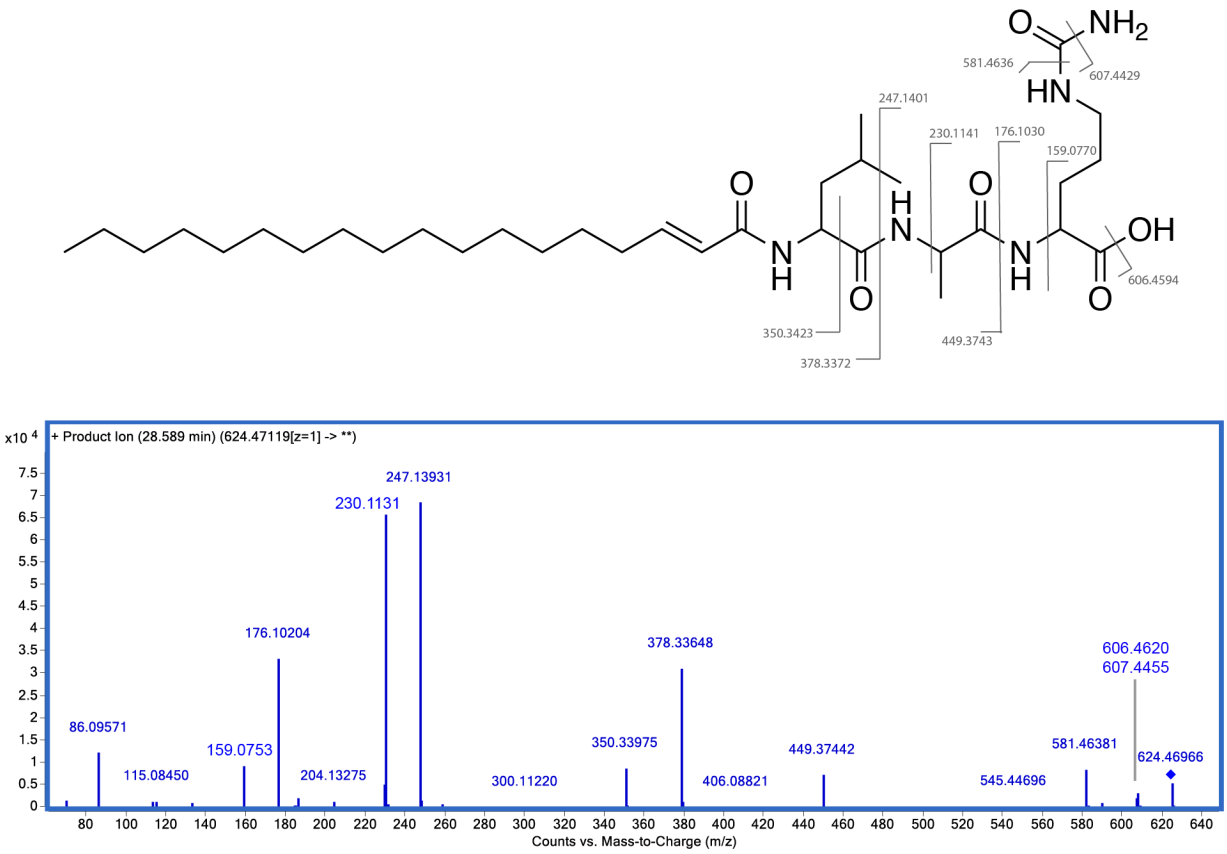


Figure S15 Tandem MS assignment of proposed compound **11**. Calculated exact mass of peptide fragment ions are shown on the structure. Double bond position and configuration was not determined.

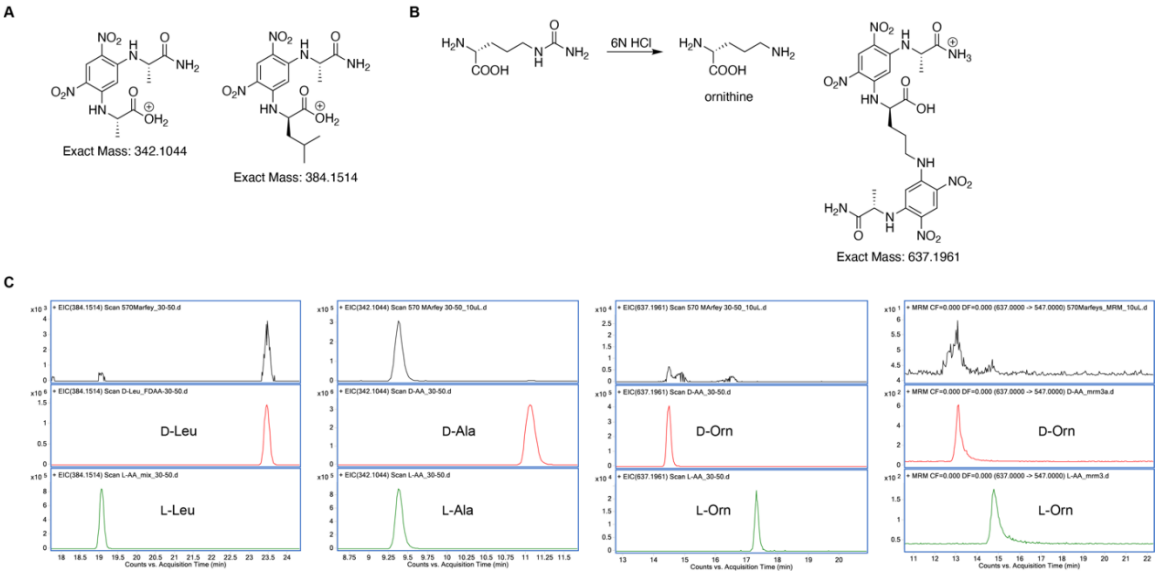


Figure S16 Configuration analysis of bovienimide B (**6**) using Marfey's reagent. (A) Structures and exact masses of alanine and leucine FDAA adduct. (B) Hydrolysis of citrulline results in ornithine, which reacted with two equivalents of FDAA to form double addition adduct. (C) LC-MS analysis of derivatized standard amino acids (middle and bottom) and hydrolyzed bovienimide (top). The first three panels were EICs analyzed by HPLC-QTOF with a 10 ppm extraction window. The fourth panel was detected by dynamic multiple reaction monitoring (dMRM) mode. (637.0 -> 547.0, CE = 10).

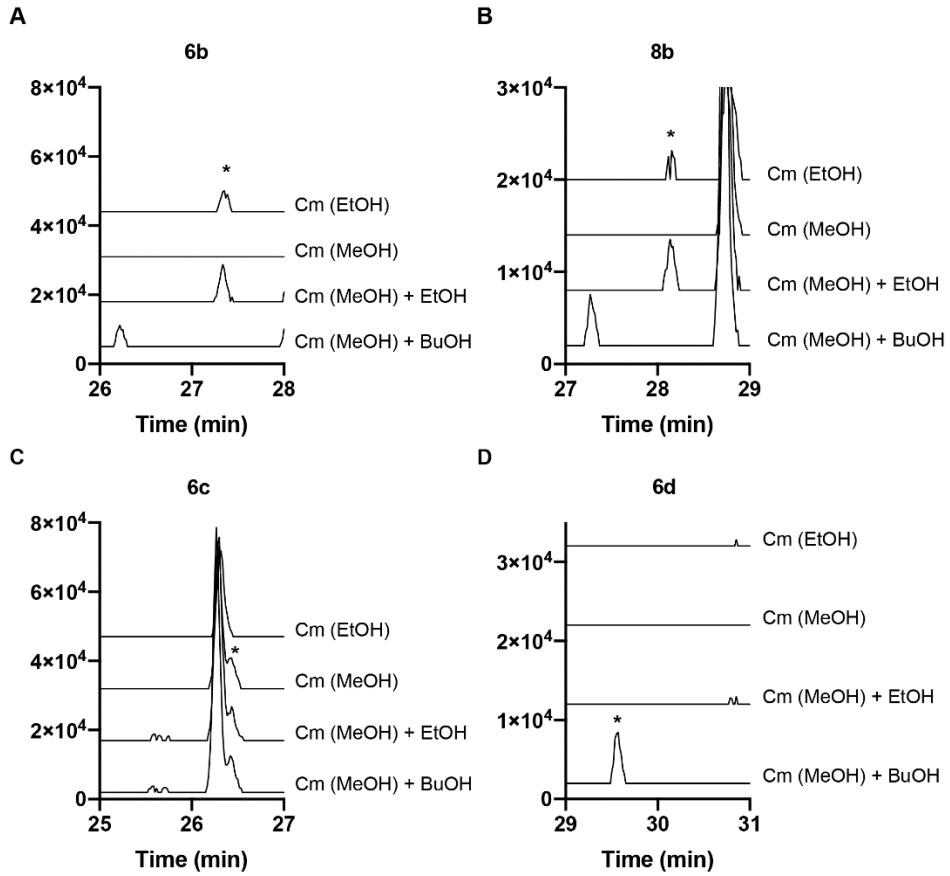


Figure S17 EICs of methyl (**6c**), ethyl (**6b**, **8b**) and butyl esters (**6d**) in *E. coli* expressing *XBJ1_2367* in the presence of exogenous alcohols. Chloramphenicol stock solutions are prepared in methanol (Cm (MeOH)) and ethanol (Cm (EtOH)), with the addition of ethanol (+EtOH) or *n*-butanol (BuOH).

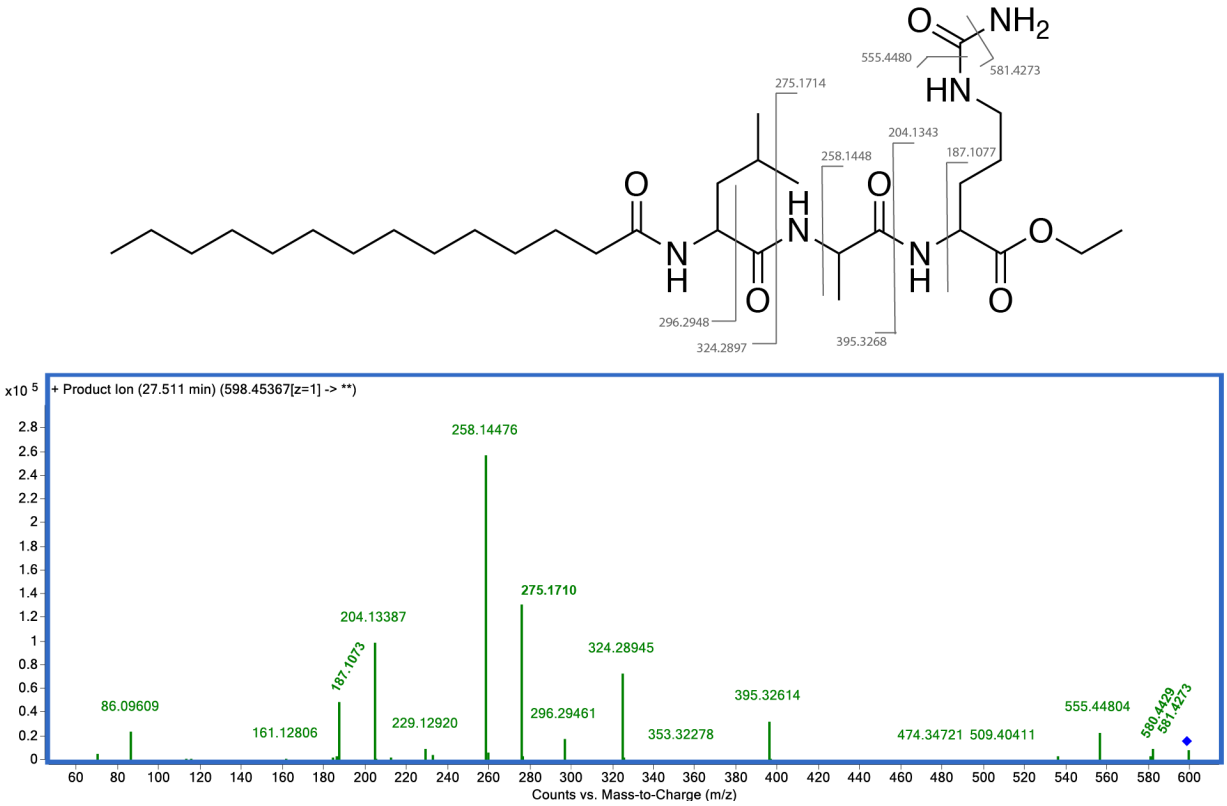


Figure S18 Tandem MS assignment of proposed compound **6b**. Calculated exact mass of peptide fragment ions are shown on the structure.

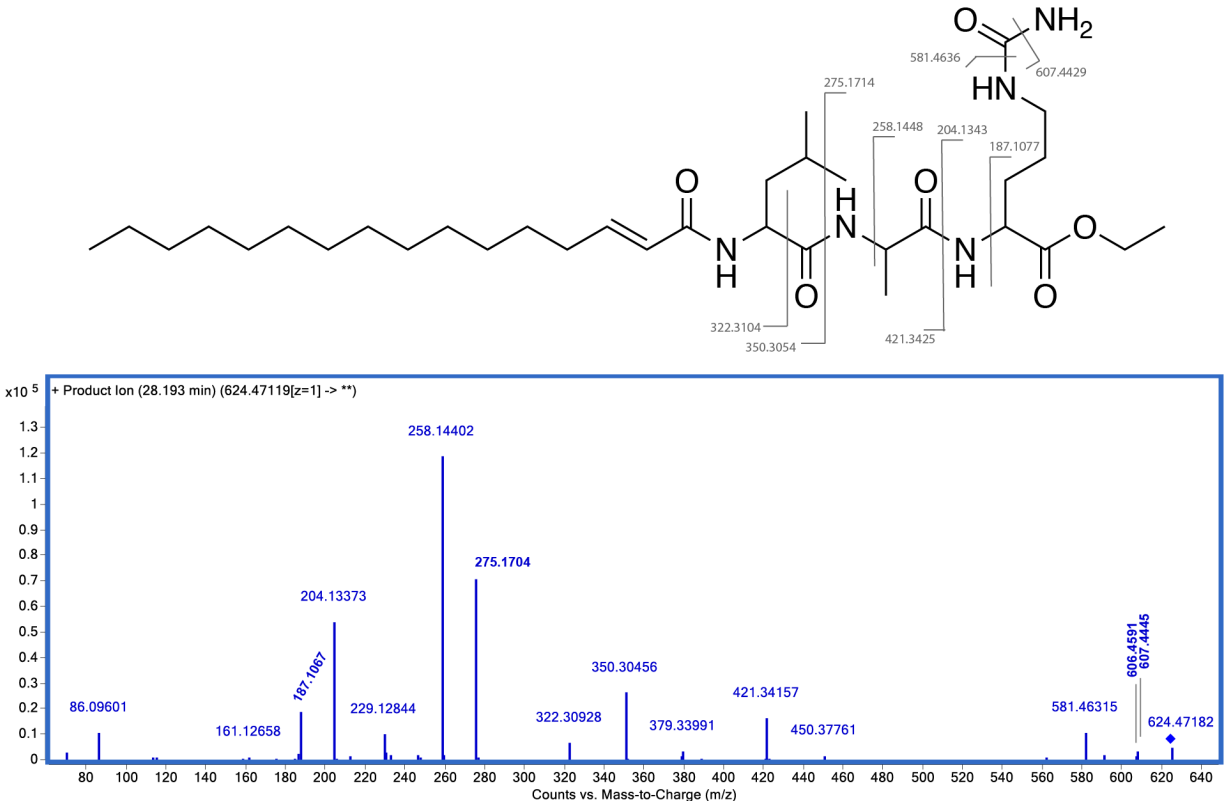


Figure S19 Tandem MS assignment of proposed compound **8b**. Calculated exact mass of peptide fragment ions are shown on the structure. Double bond position and configuration was not determined.

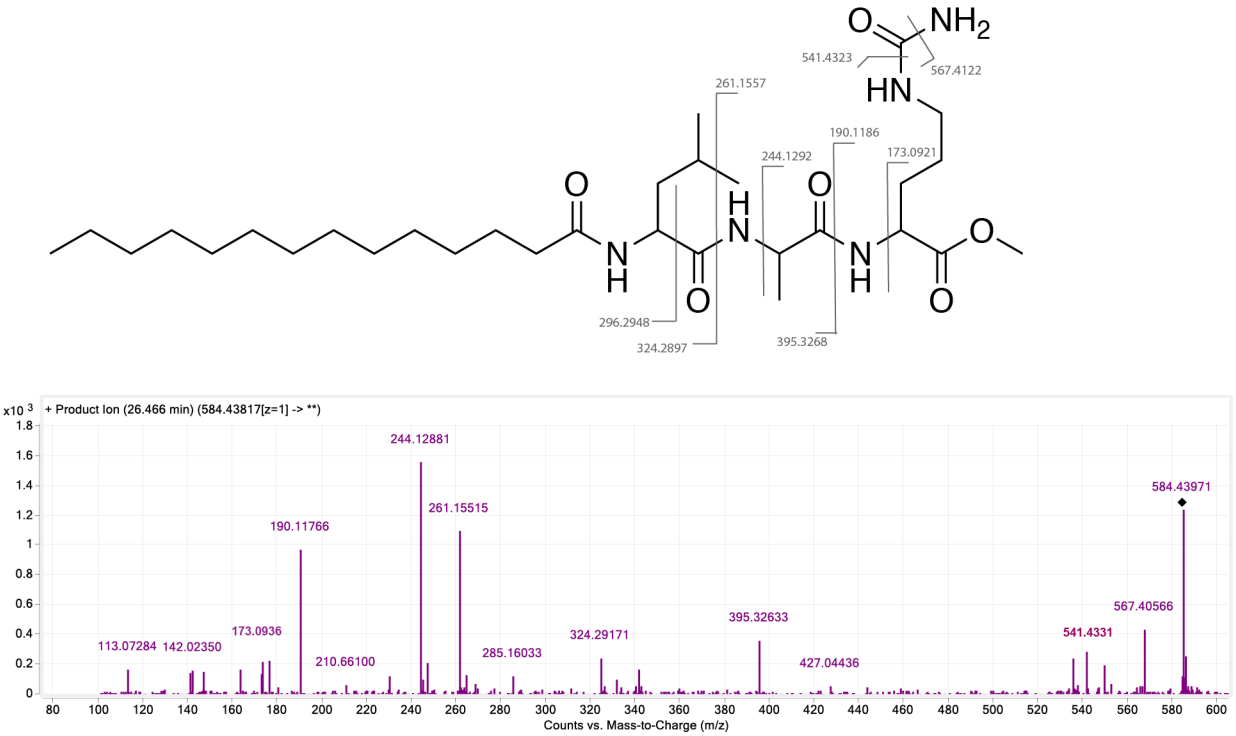


Figure S20 Tandem MS assignment of proposed compound **6c**. Calculated exact mass of peptide fragment ions are shown on the structure.

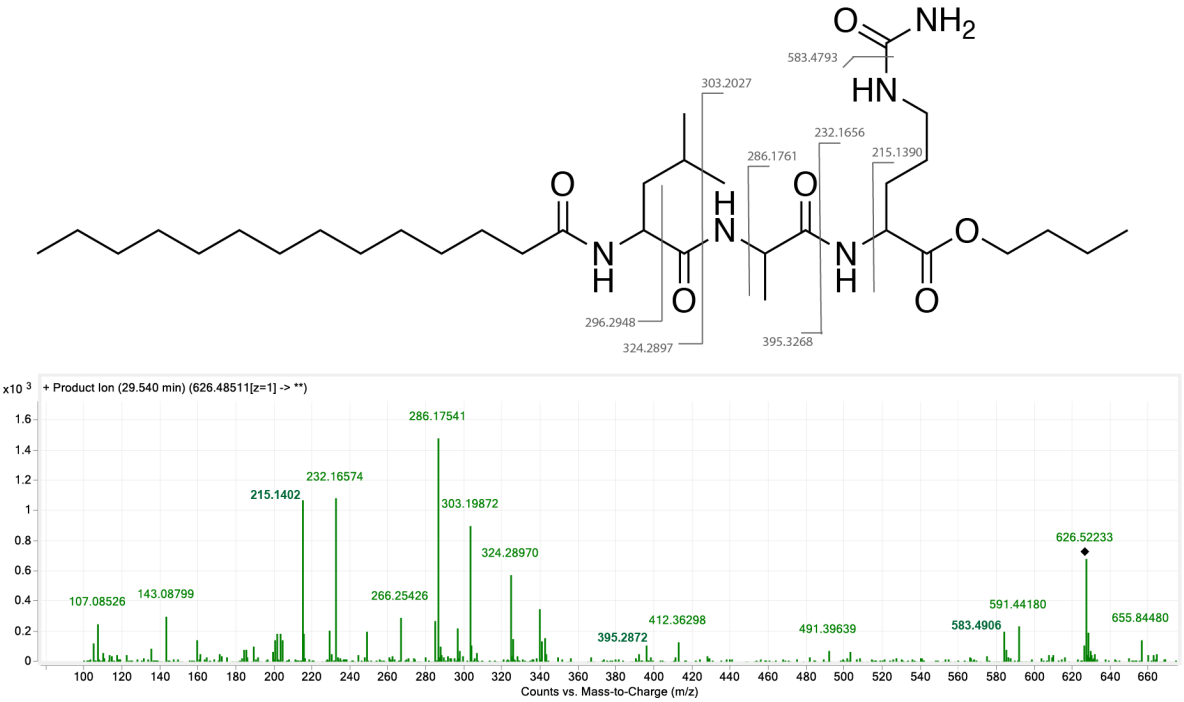


Figure S21 Tandem MS assignment of proposed compound **6d**. Calculated exact mass of peptide fragment ions are shown on the structure.

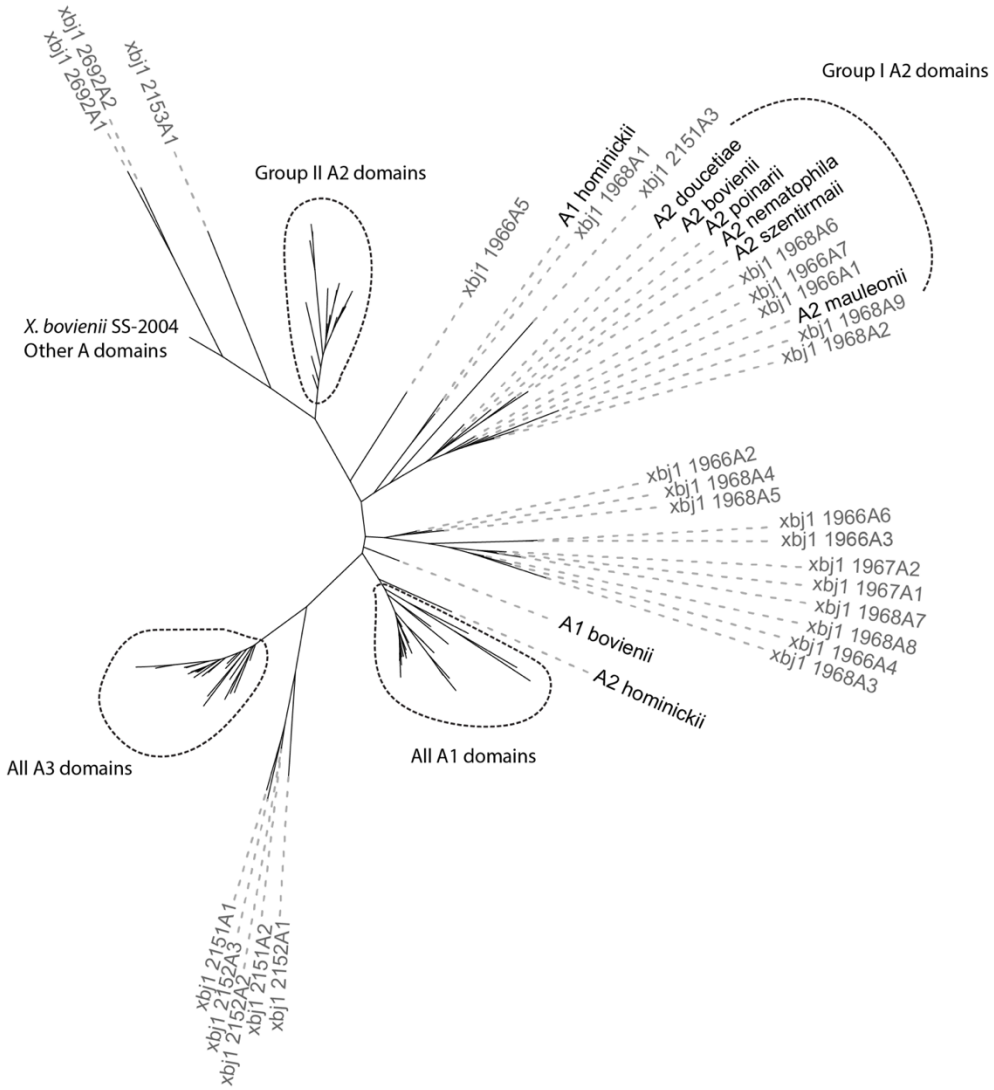


Figure S22 Maximum likelihood tree of all annotated adenylation domains in *X. bovienii* SS-2004 and *XBJ1_2367* homologues from different *Xenorhabdus* species. Genes from *X. bovienii* SS-2004 are labeled as xbj1 based on MaGe genome annotations. *xbj1_2151* (*paxC*) and *xbj1_2152* (*paxB*) belong to PAX-peptide synthetase.

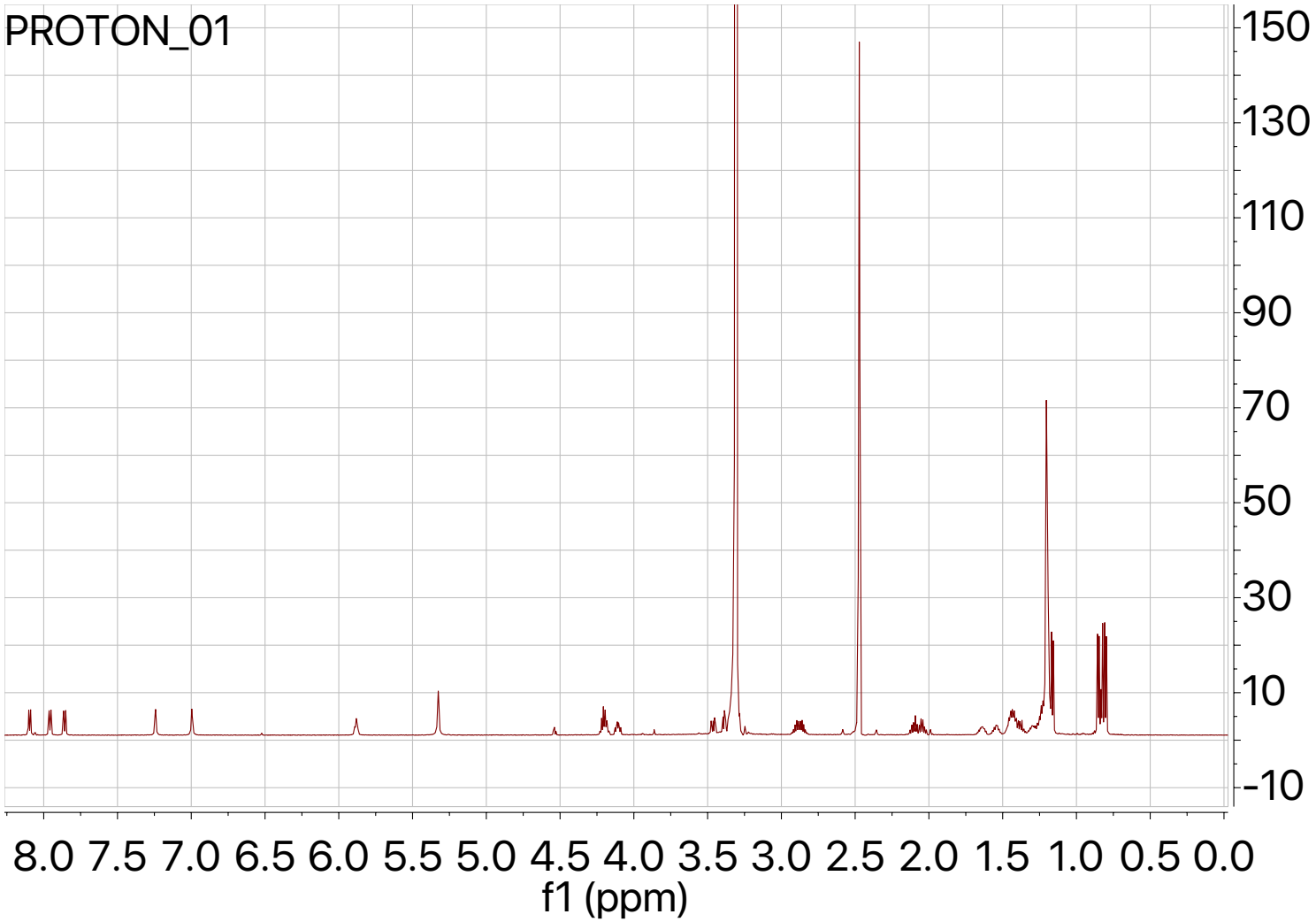


Figure S23 ^{1}H NMR spectrum of bovienimide A (**5**), 600 MHz, DMSO-d_6, 25 °C

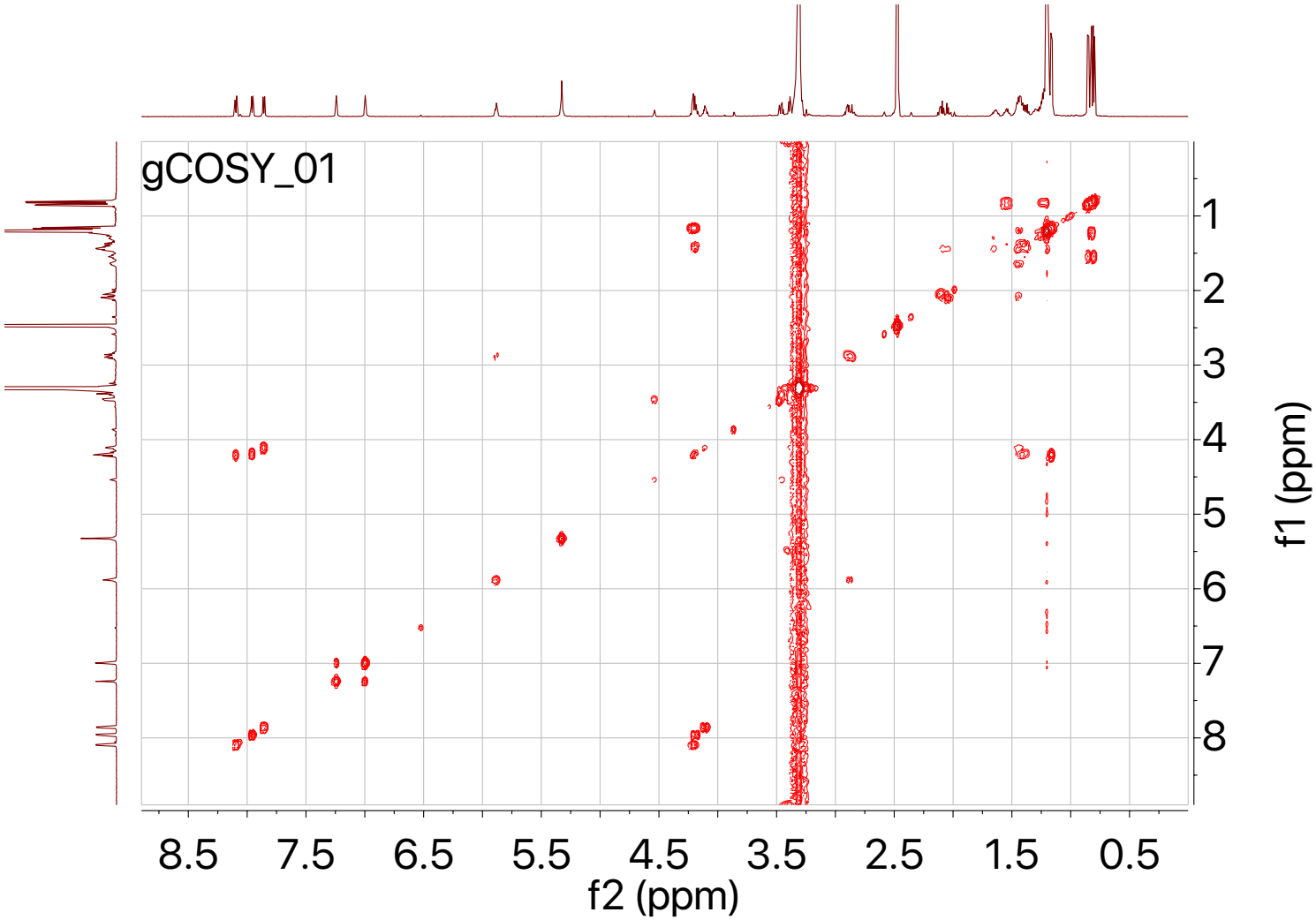


Figure S24 gCOSY NMR spectrum of bovienimide A (**5**), 600 MHz, DMSO-d_6, 25 °C

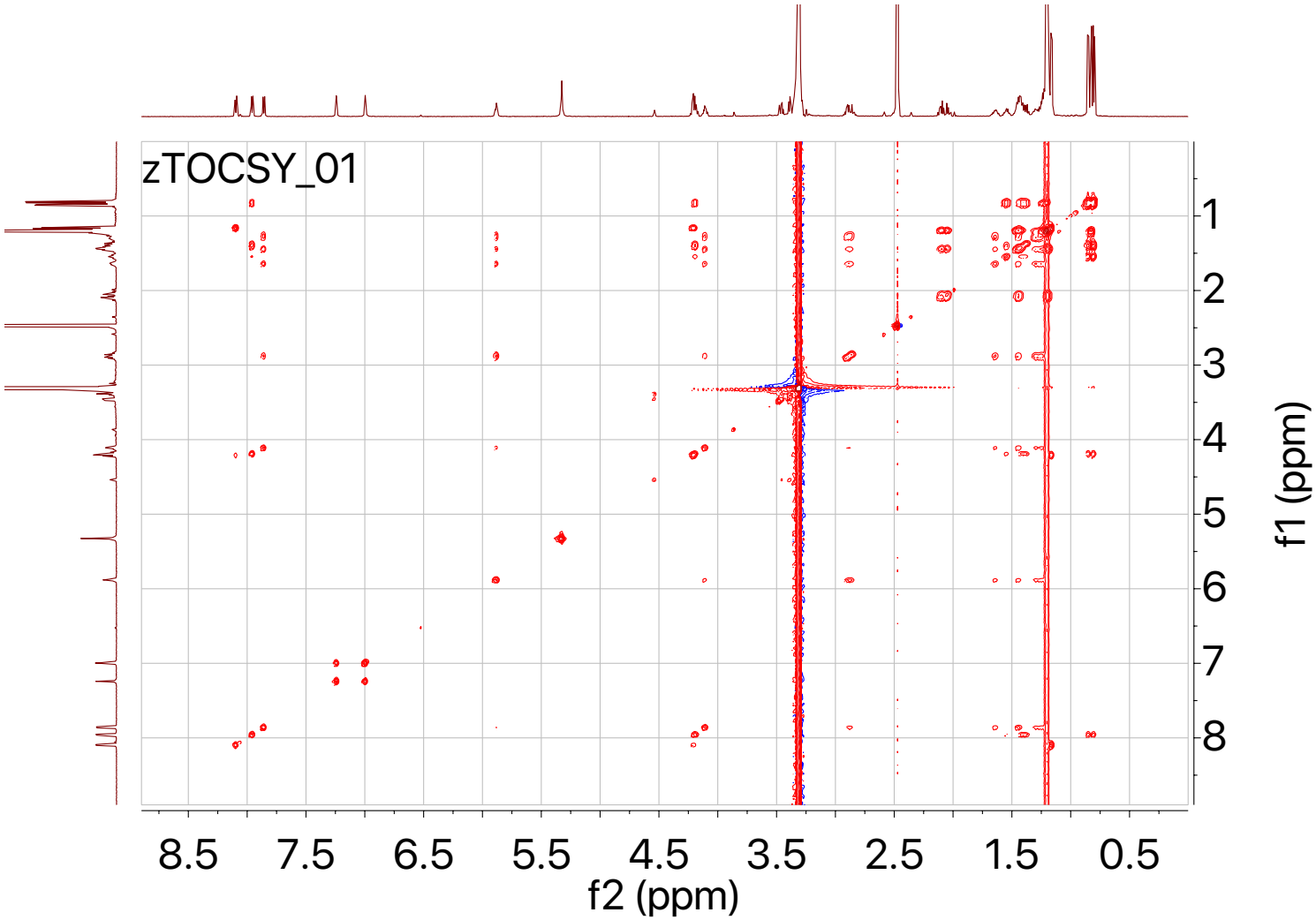


Figure S25 zTOCSY NMR spectrum of bovienimide A (**5**), 600 MHz, DMSO-d_6, 25 °C

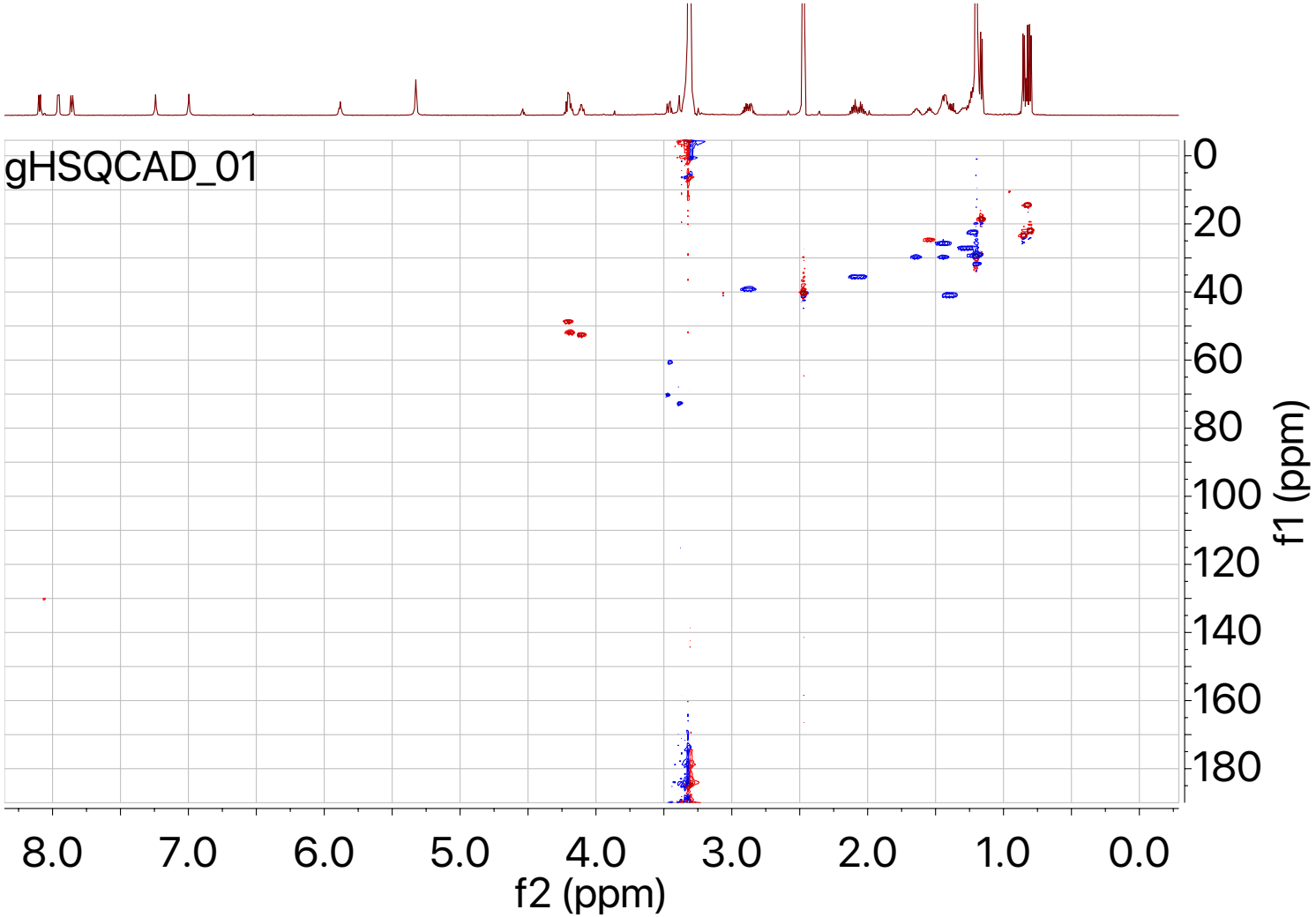


Figure S26 ^{1}H-^{13}C gHSQC NMR spectrum of bovienimide A (**5**), 600 MHz, DMSO-d_6, 25 °C

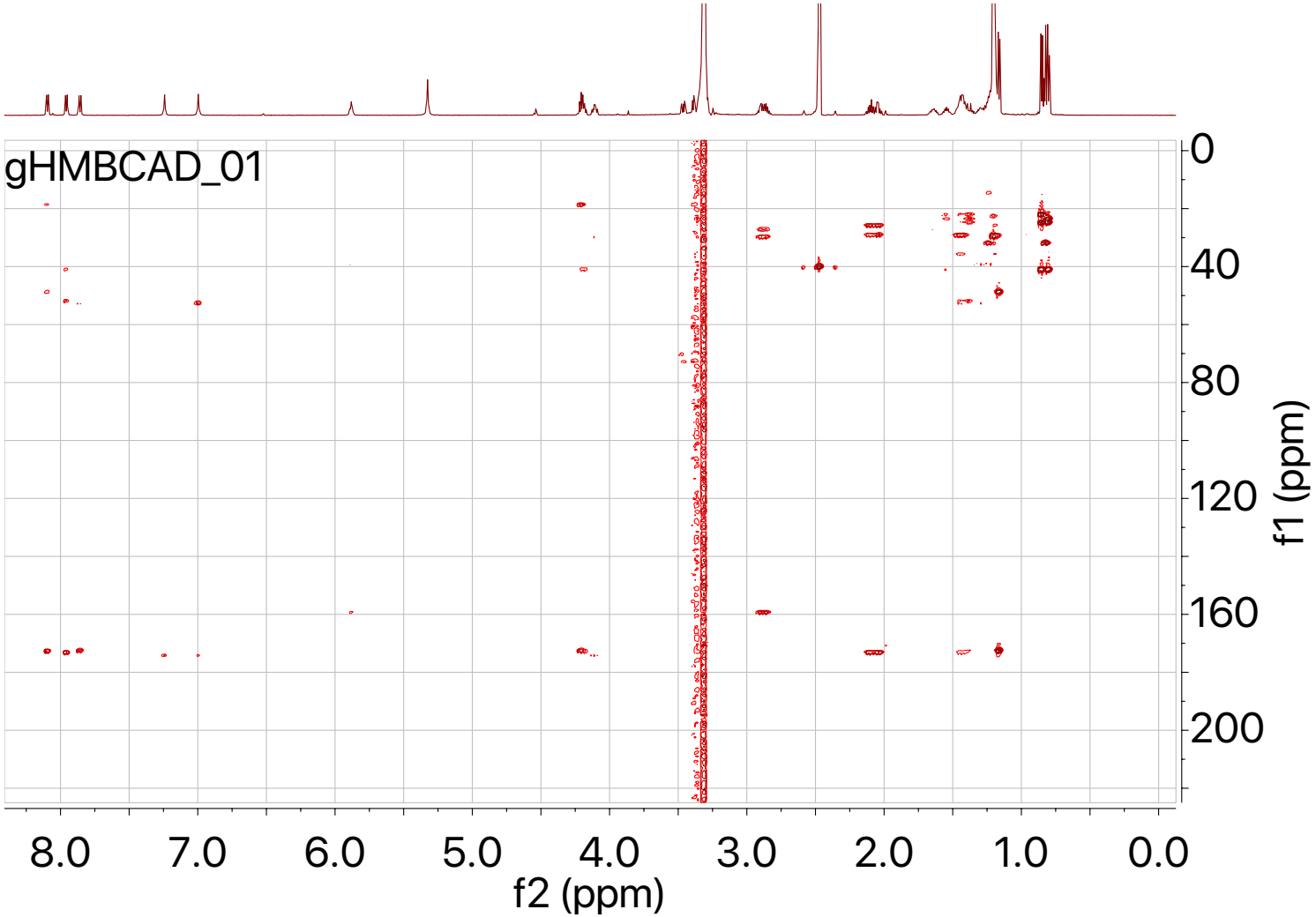


Figure S27 ^{1}H-^{13}C gHMBC NMR spectrum of bovienimide A (**5**), 600 MHz, DMSO-d_6, 25 °C

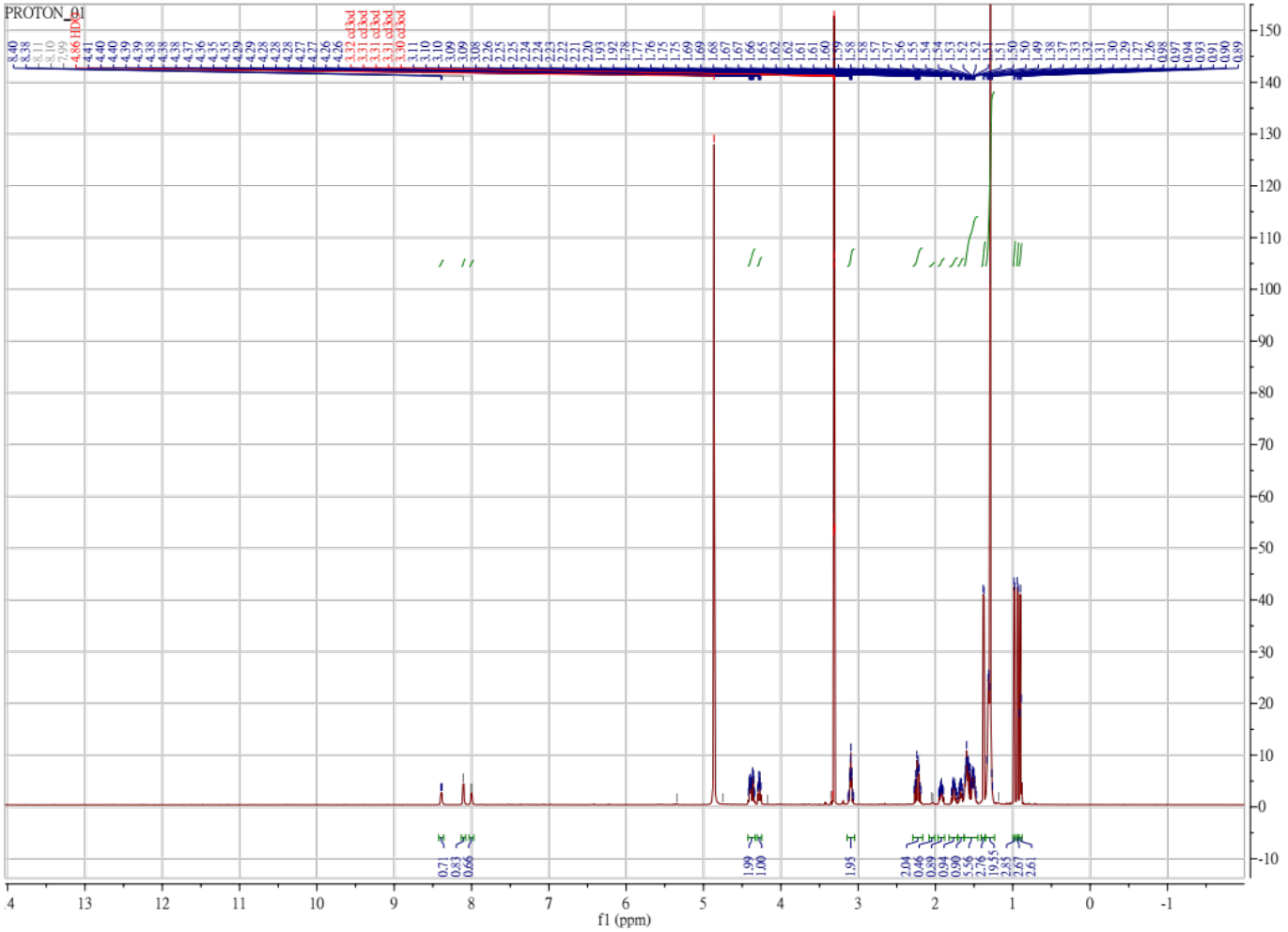


Figure S28 ^{1}H NMR spectrum of bovienimide B (**6**), 600 MHz, MeOD-d_4, 25 °C

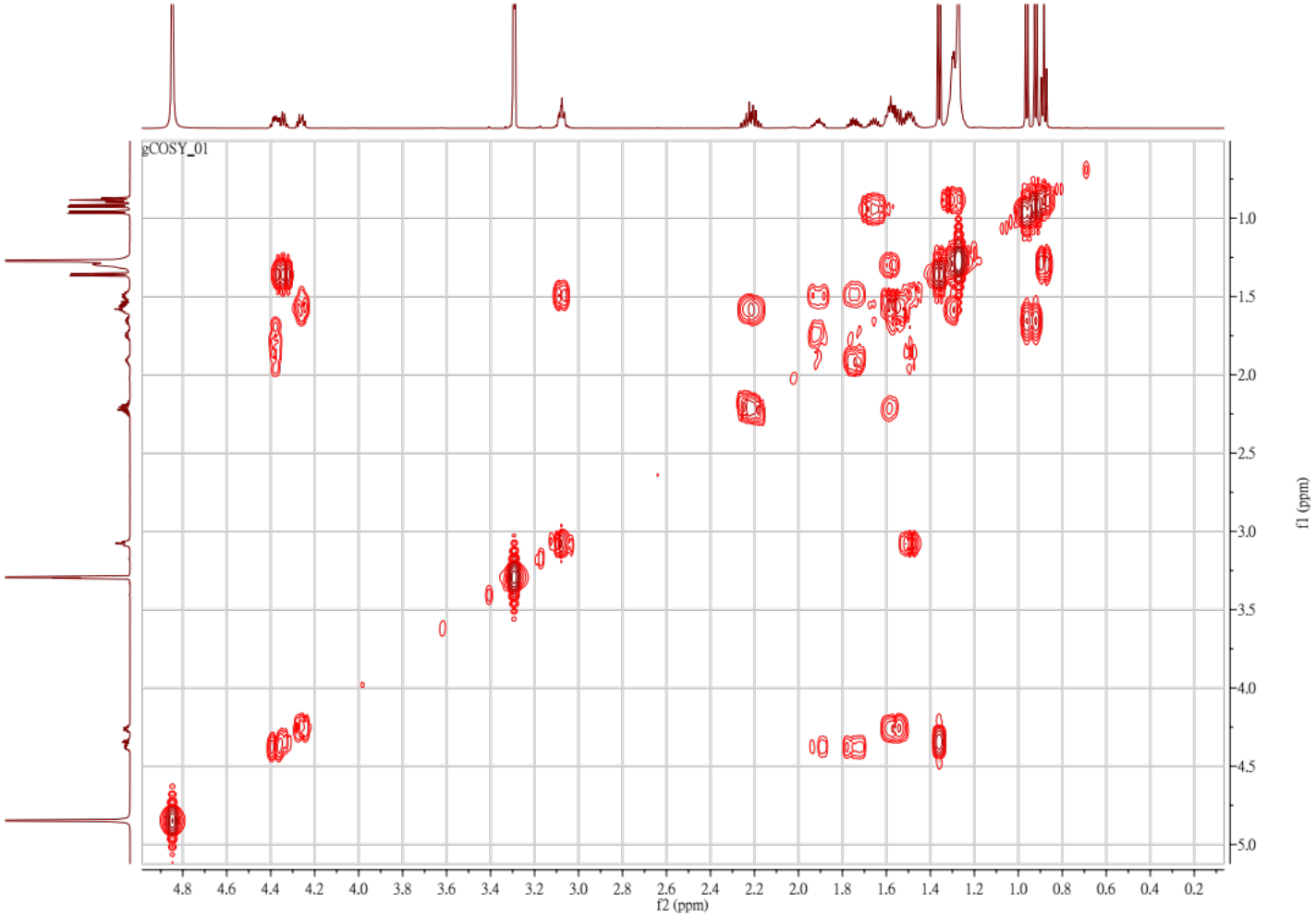


Figure S29 gCOSY NMR spectrum of bovienimide B (**6**), 600 MHz, MeOD-d_4, 25 °C

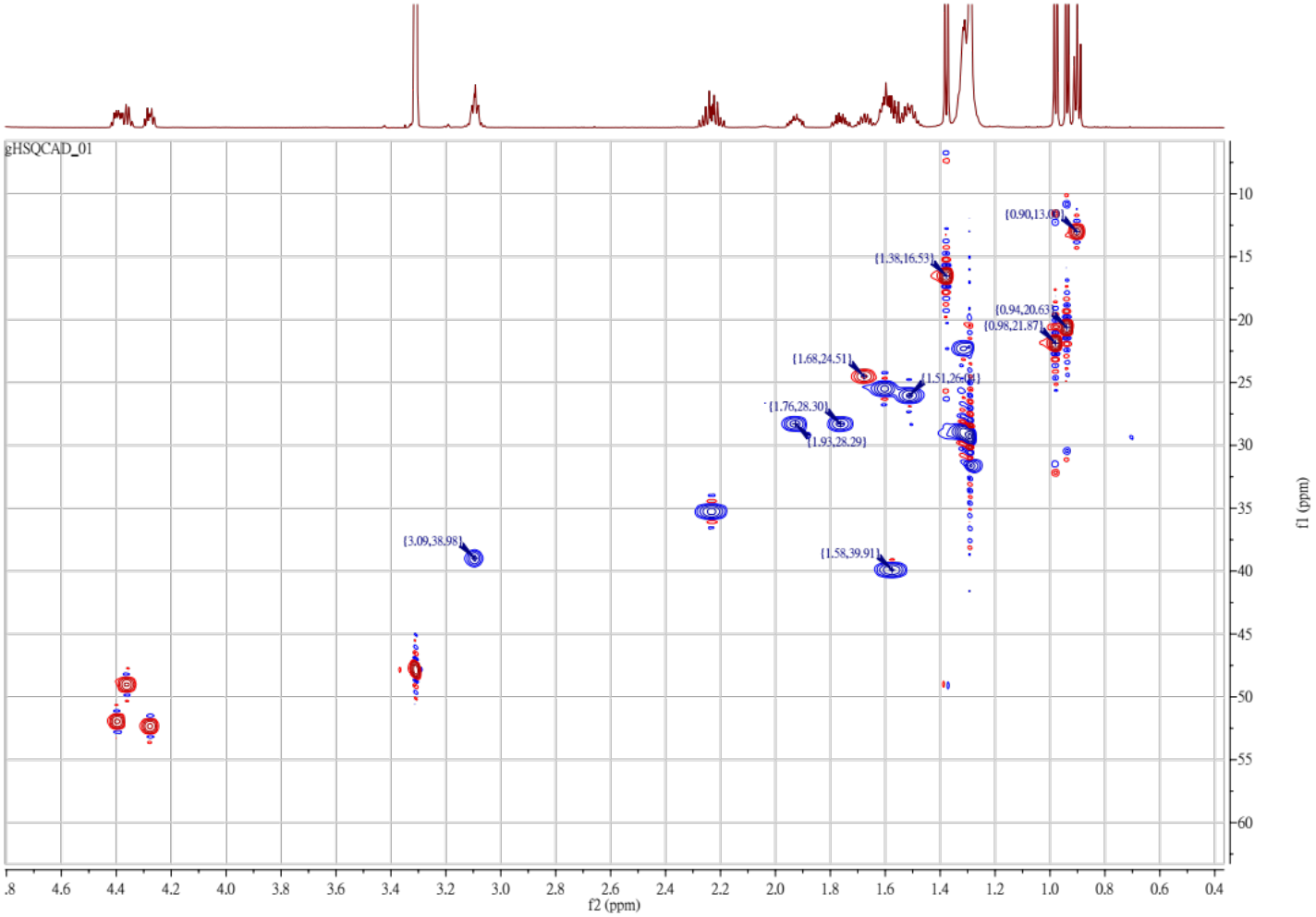


Figure S30 ^{1}H-^{13}C gHSQC NMR spectrum of bovienimide B (**6**), 600 MHz, MeOD-d_4, 25 °C

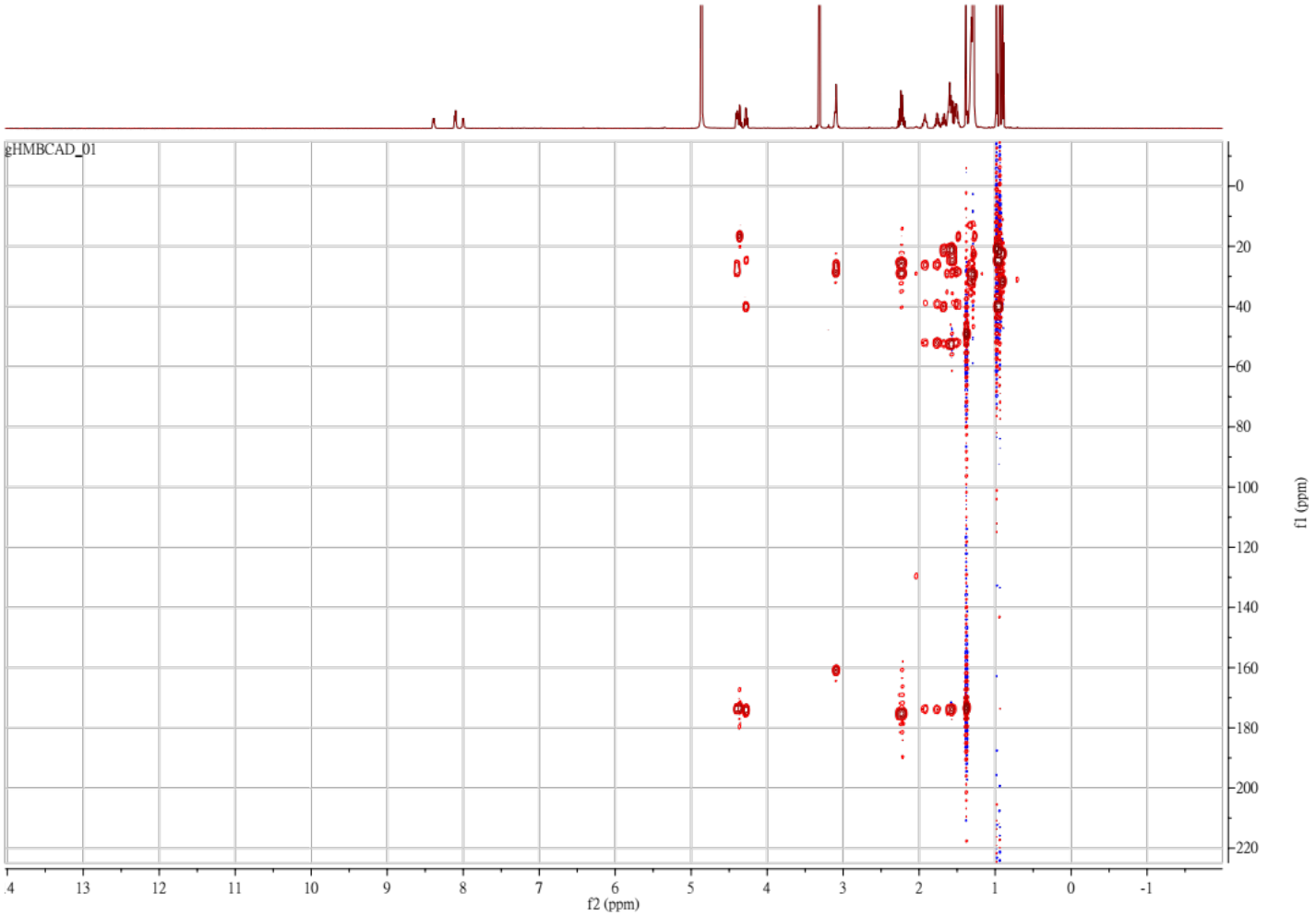


Figure S31 ^{1}H-^{13}C gHMBC NMR spectrum of bovienimide B (**6**), 600 MHz, MeOD-d_4, 25 °C

Supplementary References

Kroeze, W.K., Sassano, M.F., Huang, X.P., Lansu, K., McCorvy, J.D., Giguere, P.M., Sciaky, N., and Roth, B.L. (2015). PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. Nat. Struct. Mol. Biol. *22*, 362-369.