

## **Scytodecamide from the Cultured *Scytonema* sp. UIC 10036 Expands the Chemical and Genetic Diversity of Cyanobactins**

Camila M. Crnkovic,<sup>[a,b]</sup> Jana Braesel,<sup>[a]</sup> Aleksej Kronic,<sup>[a]</sup> Alessandra S. Eustáquio,<sup>[a]</sup> Jimmy Orjala\*<sup>[a]</sup>

<sup>[a]</sup> Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, Illinois 60612 (USA)

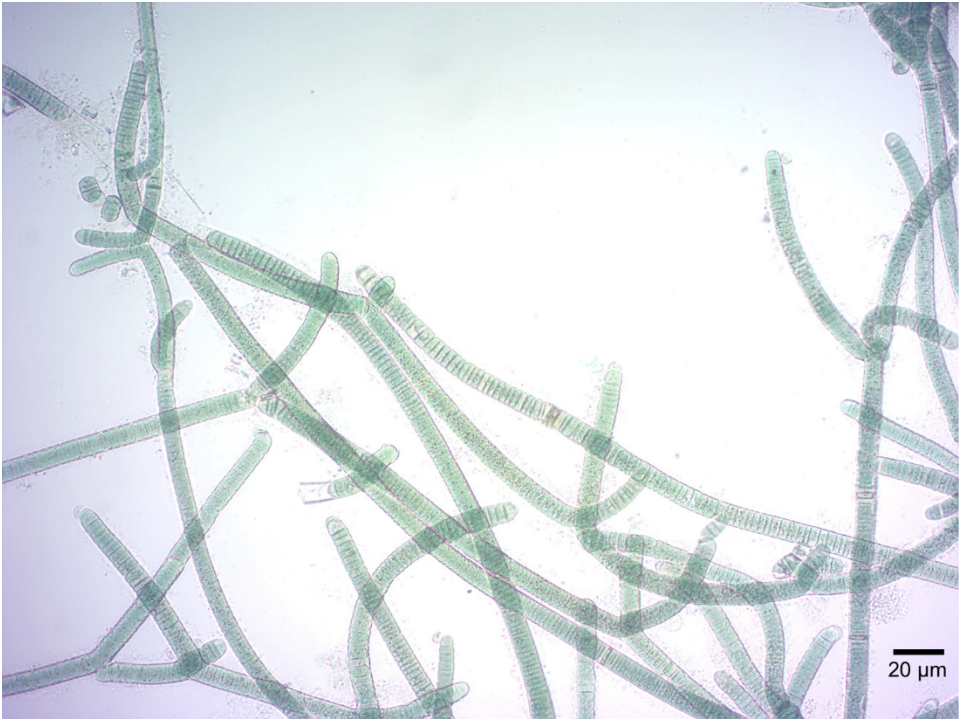
<sup>[b]</sup> CAPES Foundation, Ministry of Education of Brazil, Brasília, Federal District 70040-020 (Brazil)

# **Supporting Information**

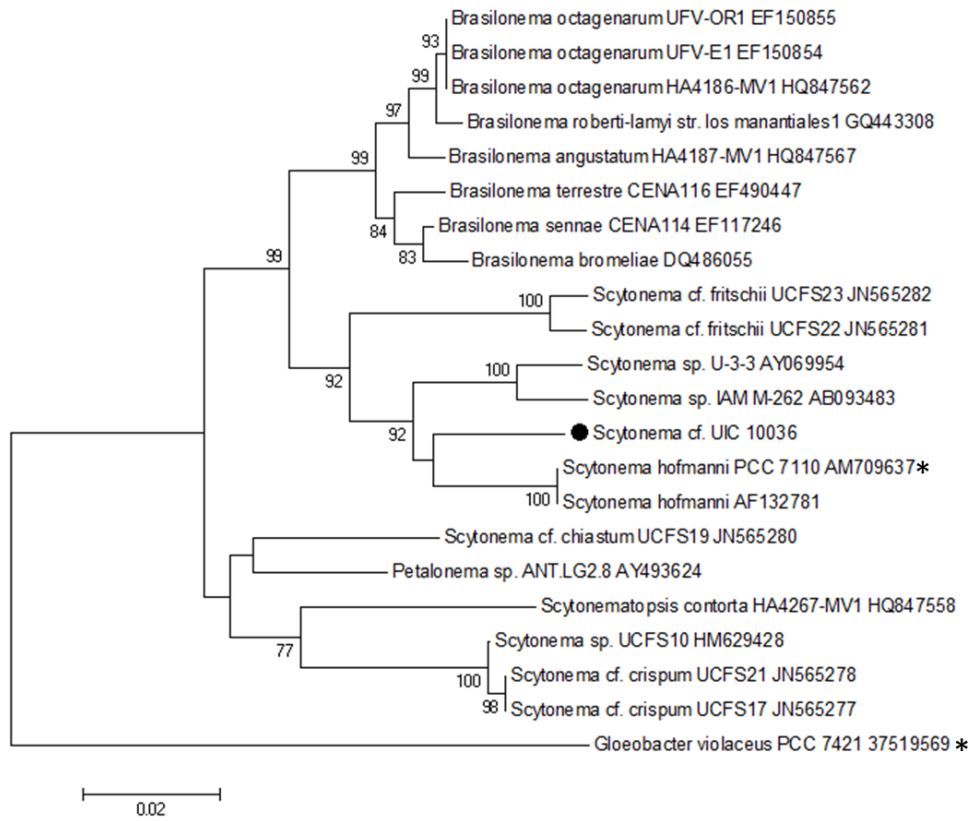
Table of contents:

- S1.** Photomicrograph (40x, bright field) of *Scytonema* sp. UIC 10036
- S2.** Phylogenetic analysis of *Scytonema* sp. UIC 10036
- S3.** MS/MS spectrum of **1**
- S4.** <sup>1</sup>H NMR spectrum (900 MHz, methanol-d<sub>4</sub>) of **1**
- S5.** <sup>1</sup>H NMR spectrum (900 MHz, methanol-d<sub>3</sub>) of **1**
- S6.** DEPTQ spectrum (226 MHz, methanol-d<sub>4</sub>) of **1**
- S7.** COSY spectrum (900 MHz, methanol-d<sub>4</sub>) of **1**
- S8.** COSY spectrum (900 MHz, methanol-d<sub>3</sub>) of **1**
- S9.** TOCSY spectrum (900 MHz, methanol-d<sub>4</sub>) of **1**
- S10.** Edited HSQC spectrum (900 MHz, methanol-d<sub>4</sub>) of **1**
- S11.** HMBC spectrum (900 MHz, methanol-d<sub>4</sub>) of **1**
- S12.** HMBC spectrum (900 MHz, methanol-d<sub>3</sub>) of **1**
- S13.** <sup>1</sup>H NMR spectrum (900 MHz, pyridine-d<sub>5</sub> with TFA vapour) of **1**
- S14.** DEPTQ spectrum (226 MHz, pyridine-d<sub>5</sub> with TFA vapour) of **1**
- S15.** COSY spectrum (900 MHz, pyridine-d<sub>5</sub> with TFA vapour) of **1**
- S16.** TOCSY spectrum (900 MHz, pyridine-d<sub>5</sub> with TFA vapour) of **1**
- S17.** Edited HSQC spectrum (900 MHz, pyridine-d<sub>5</sub> with TFA vapour) of **1**
- S18.** HMBC spectrum (900 MHz, pyridine-d<sub>5</sub> with TFA vapour) of **1**
- S19.** Band-selective HMBC spectrum (900 MHz, pyridine-d<sub>5</sub> with TFA vapour) of **1**
- S20.** Marfey's analysis of **1**
- S21.** Protein BLAST statistics for putative *scd* biosynthetic gene cluster
- S22.** Sequence comparison of known cyanobactin precursor peptides and ScdE
- S23.** Comparison of ScdMT from *Scytonema* sp. UC 10036 and PatA from *Prochloron didemni*
- S24.** Percent identity (similarity) shared between the DUF domains of ScdA and ScdMT with PatA and three PatA homologues.
- S25.** Percent identity (similarity) shared between the DUF domains of ScdA and ScdMT with PatG and three PatG homologues.

S1. Photomicrograph (40x, bright field) of *Scytonema* sp. UIC 10036

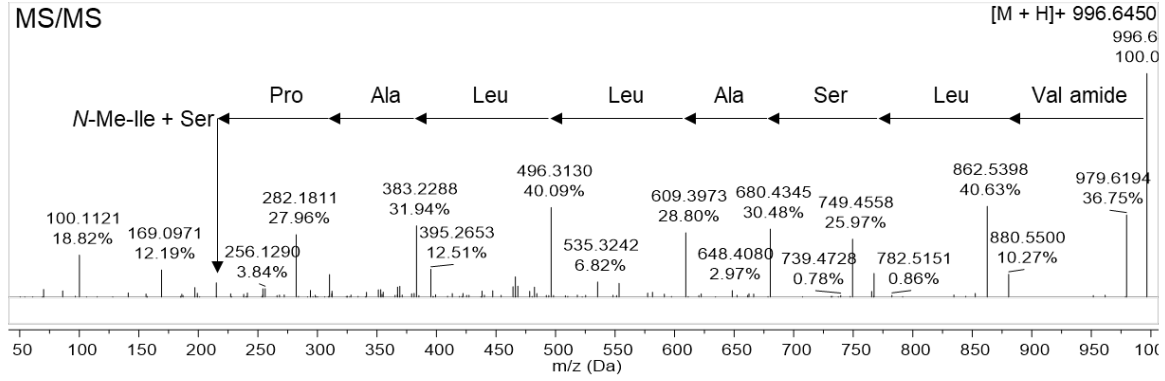


S2. Phylogenetic analysis of *Scytonema* sp. 10036

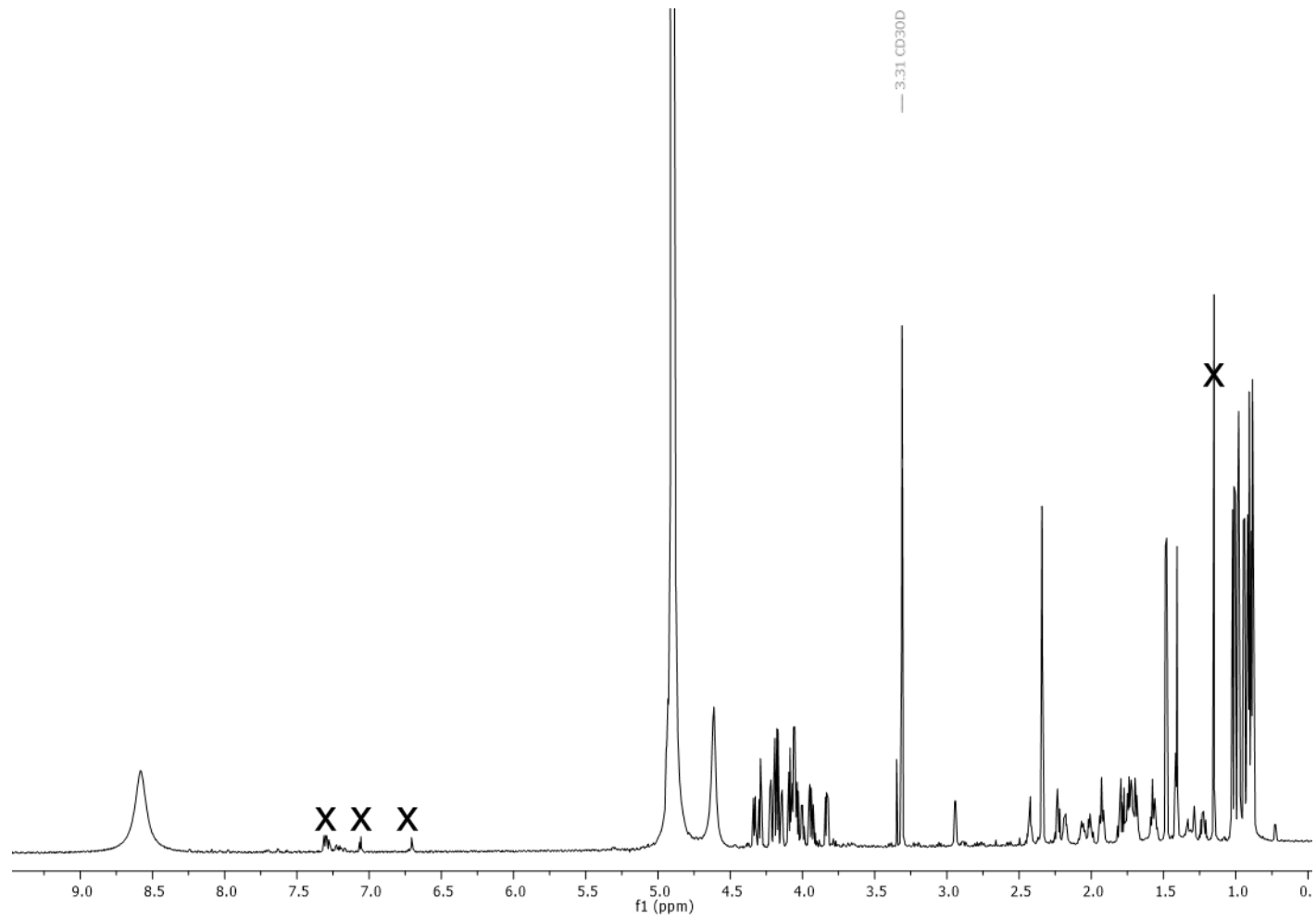


\* Reference strains are indicated by an asterisk

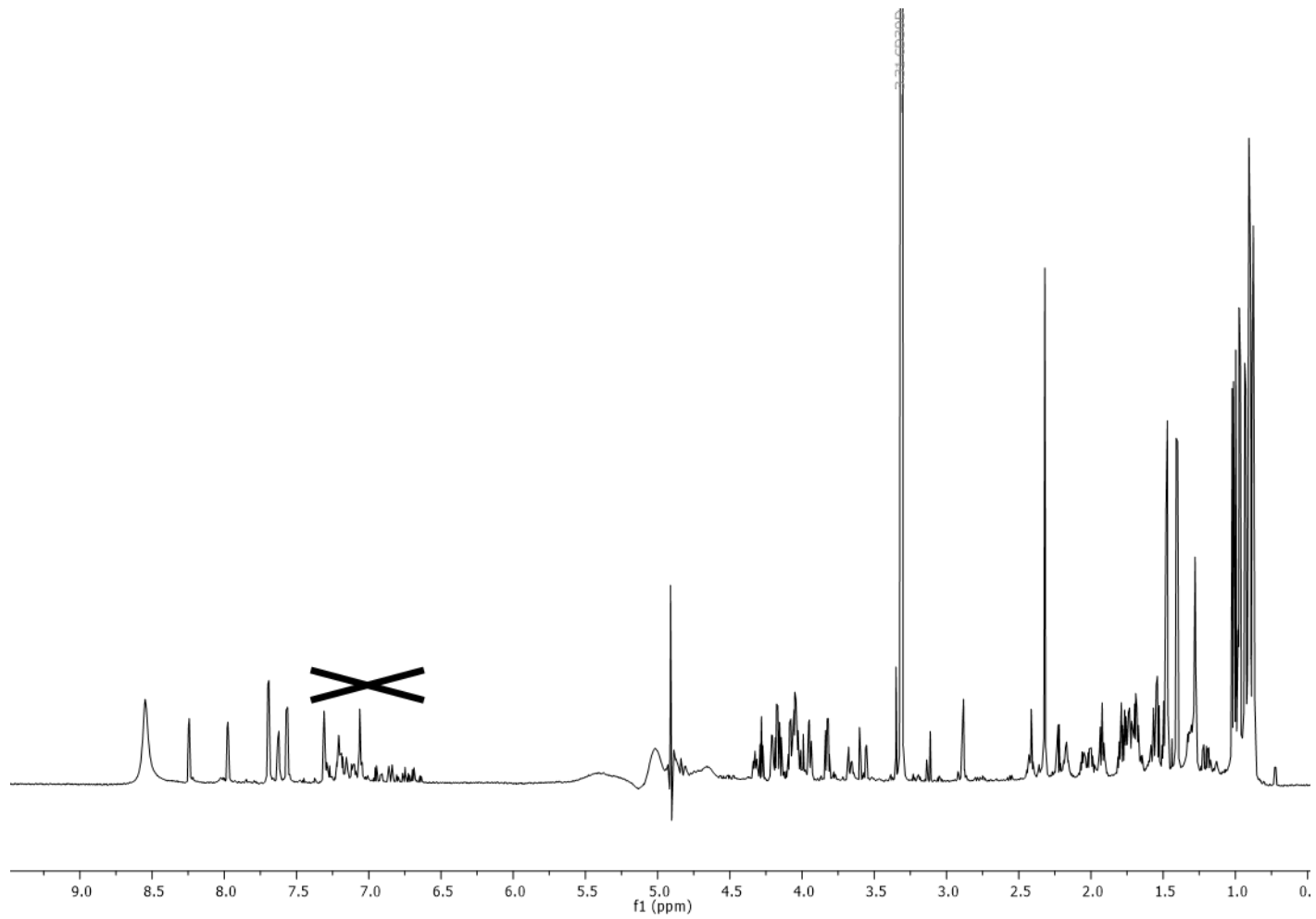
### S3. MS/MS spectrum of 1



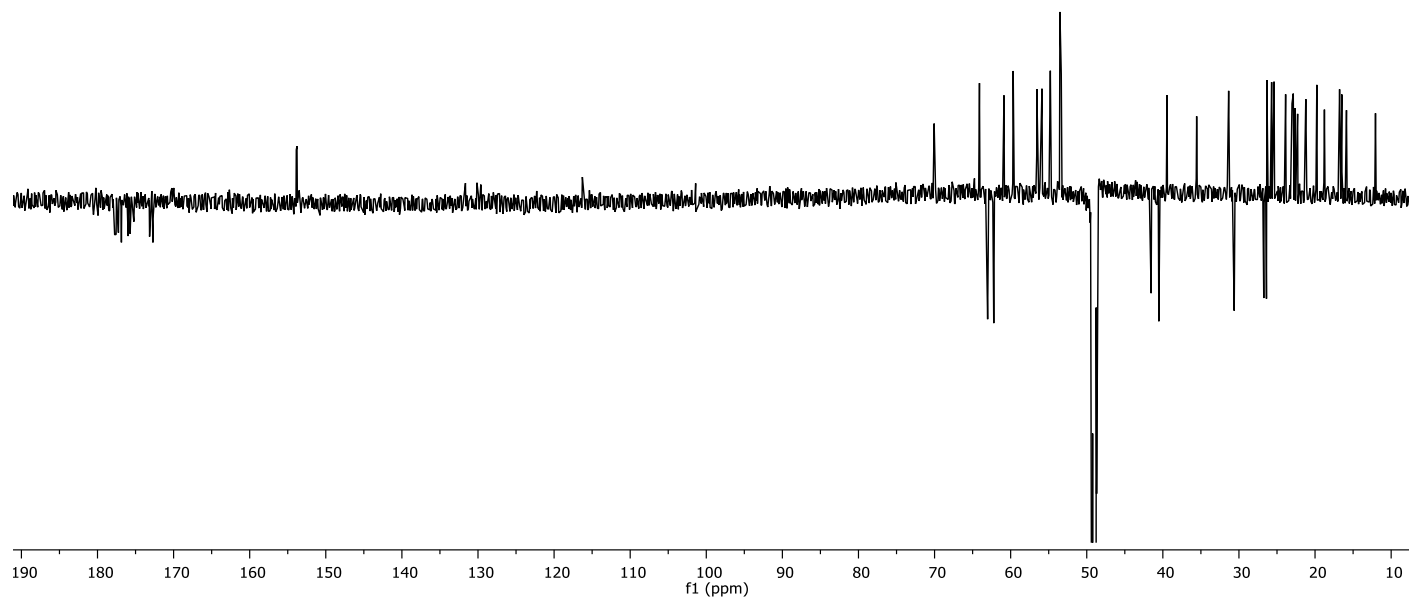
S4.  $^1\text{H}$  NMR spectrum (900 MHz, methanol- $d_4$ ) of 1. X: impurities



S5.  $^1\text{H}$  NMR spectrum (900 MHz, methanol- $\text{d}_3$ ) of 1. X: impurities

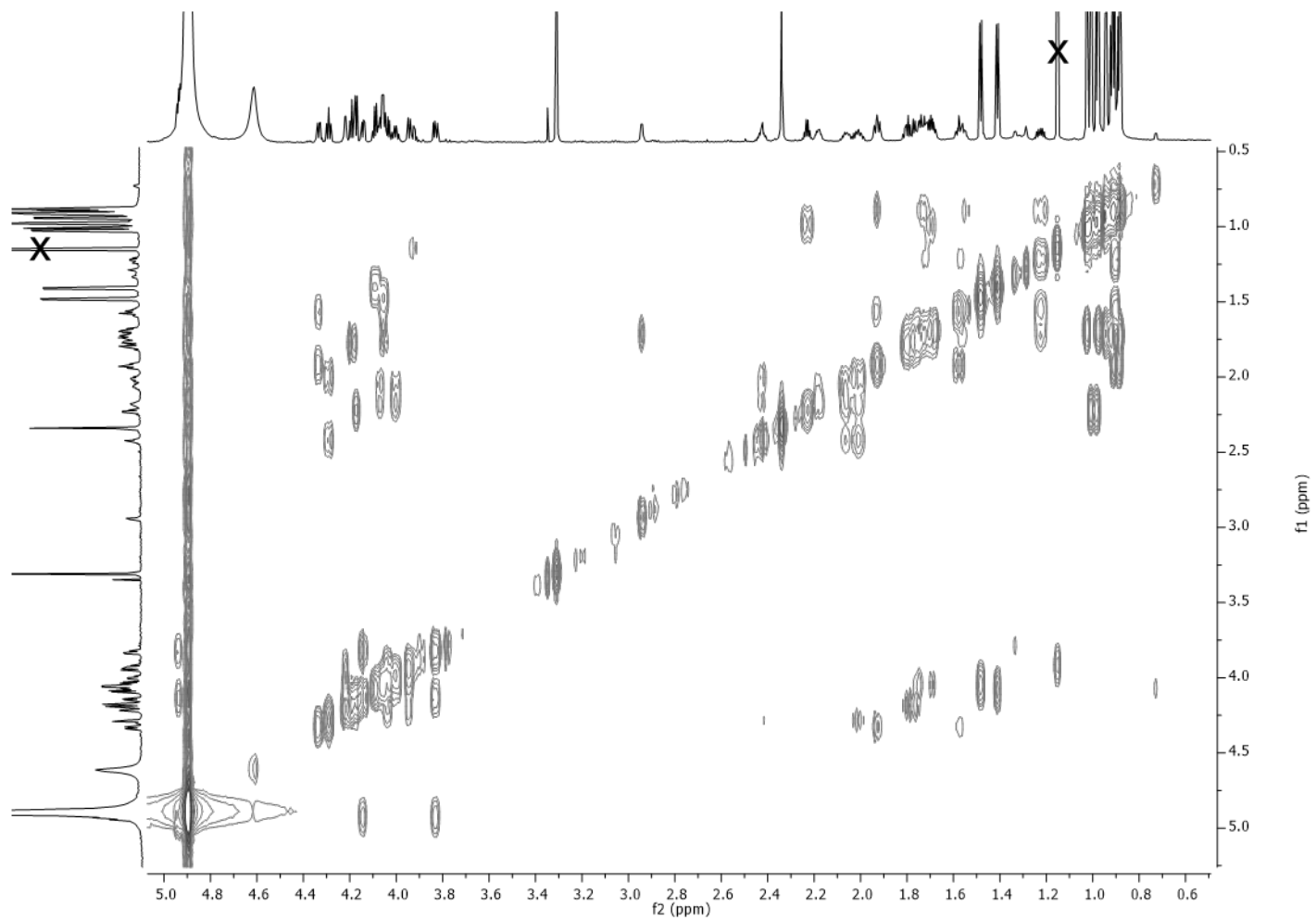


S6. DEPTQ spectrum (226 MHz, methanol-d<sub>4</sub>) of **1**

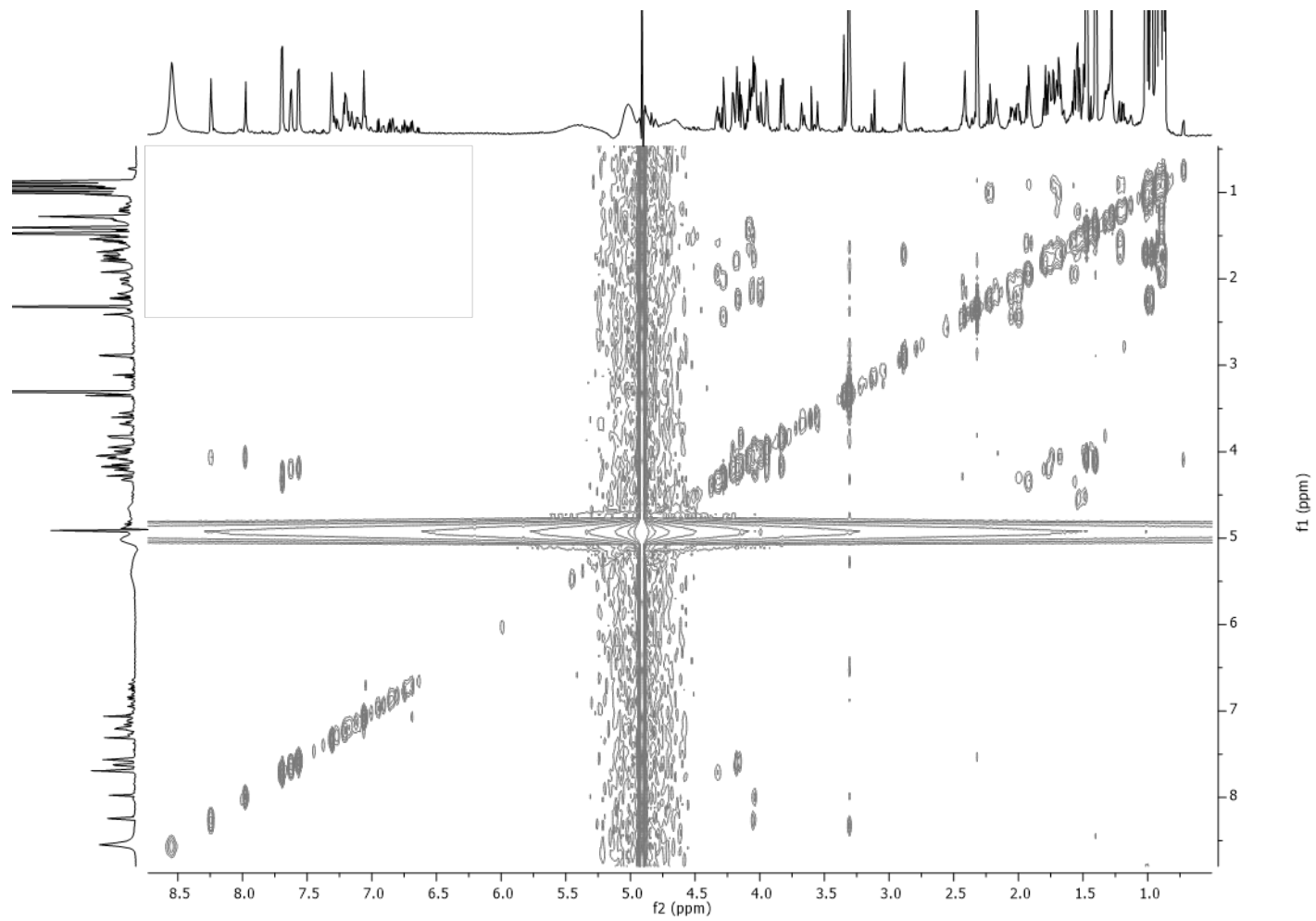




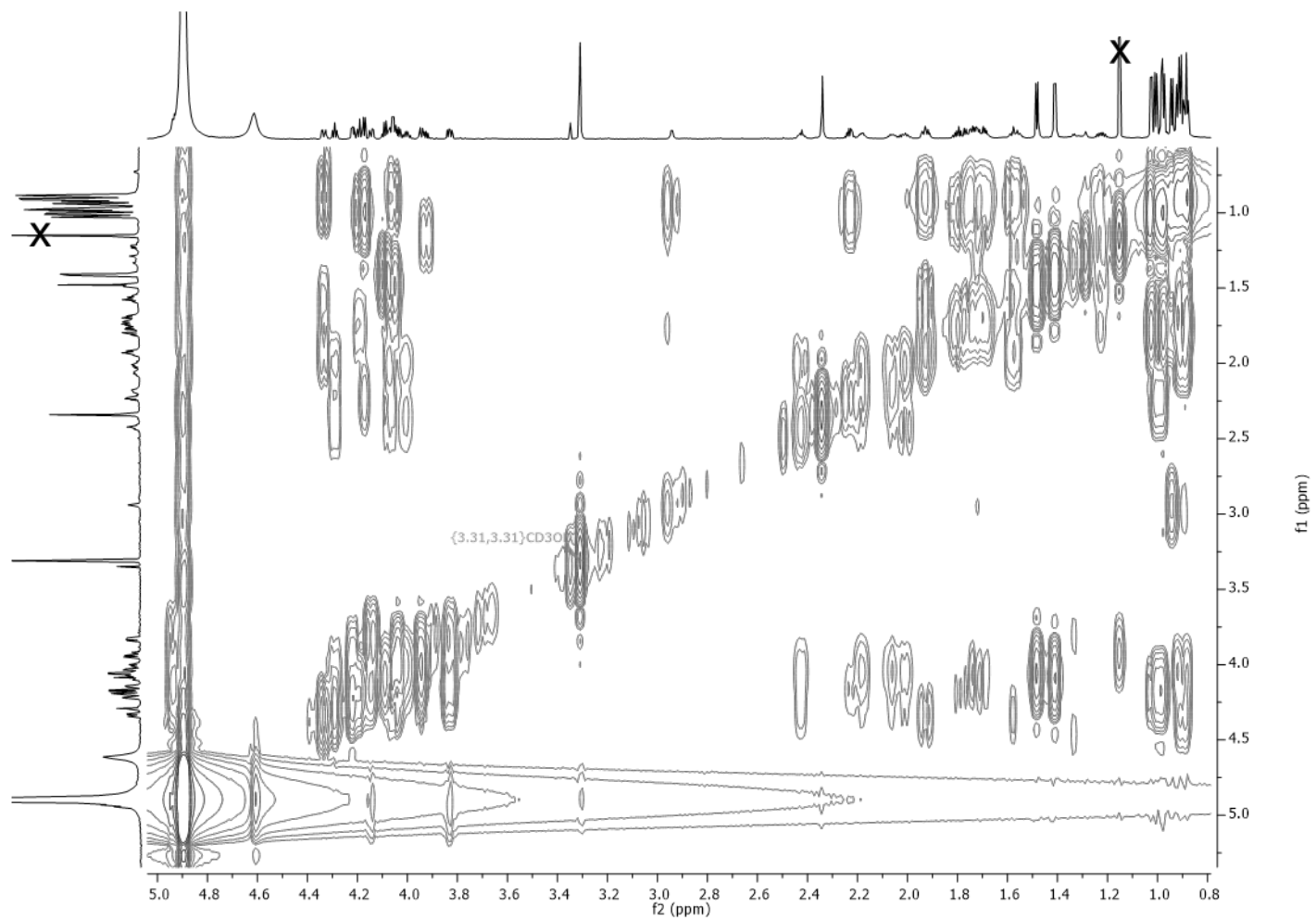
S7. COSY spectrum (900 MHz, methanol-d<sub>4</sub>) of 1. X: impurities



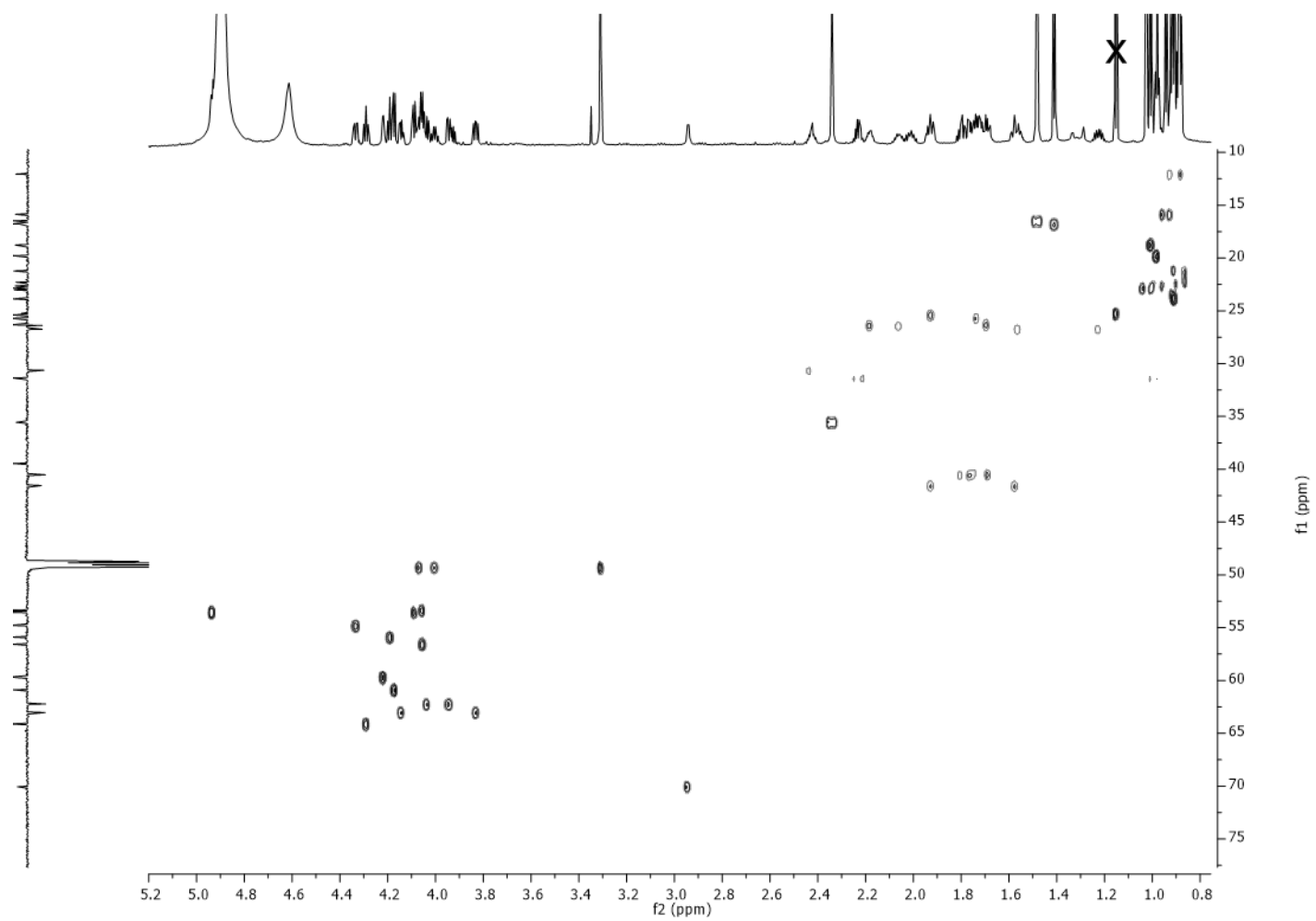
**S8.** COSY spectrum (900 MHz, methanol-d<sub>3</sub>) of **1**



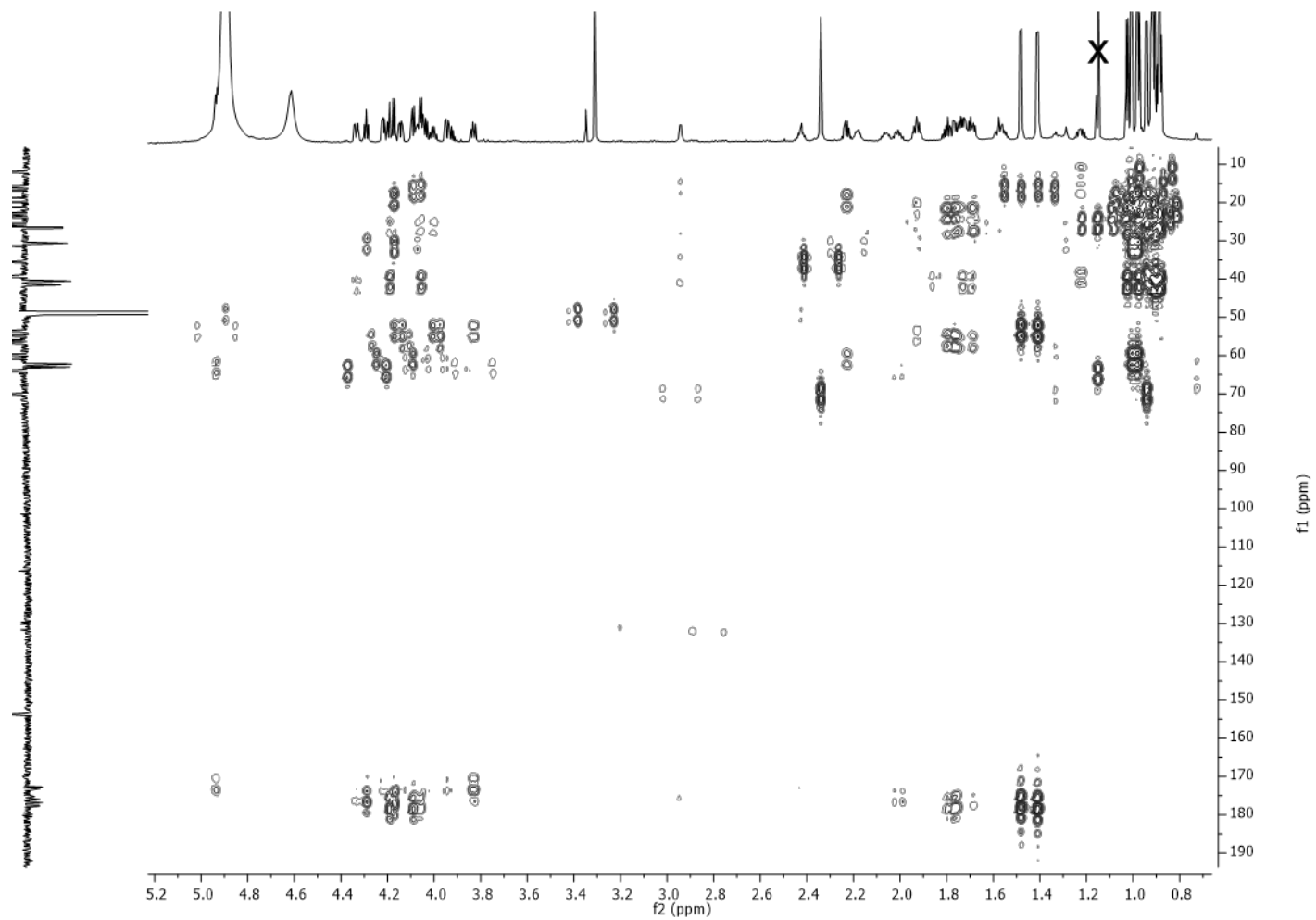
S9. TOCSY spectrum (900 MHz, methanol-d<sub>4</sub>) of 1. X: impurities



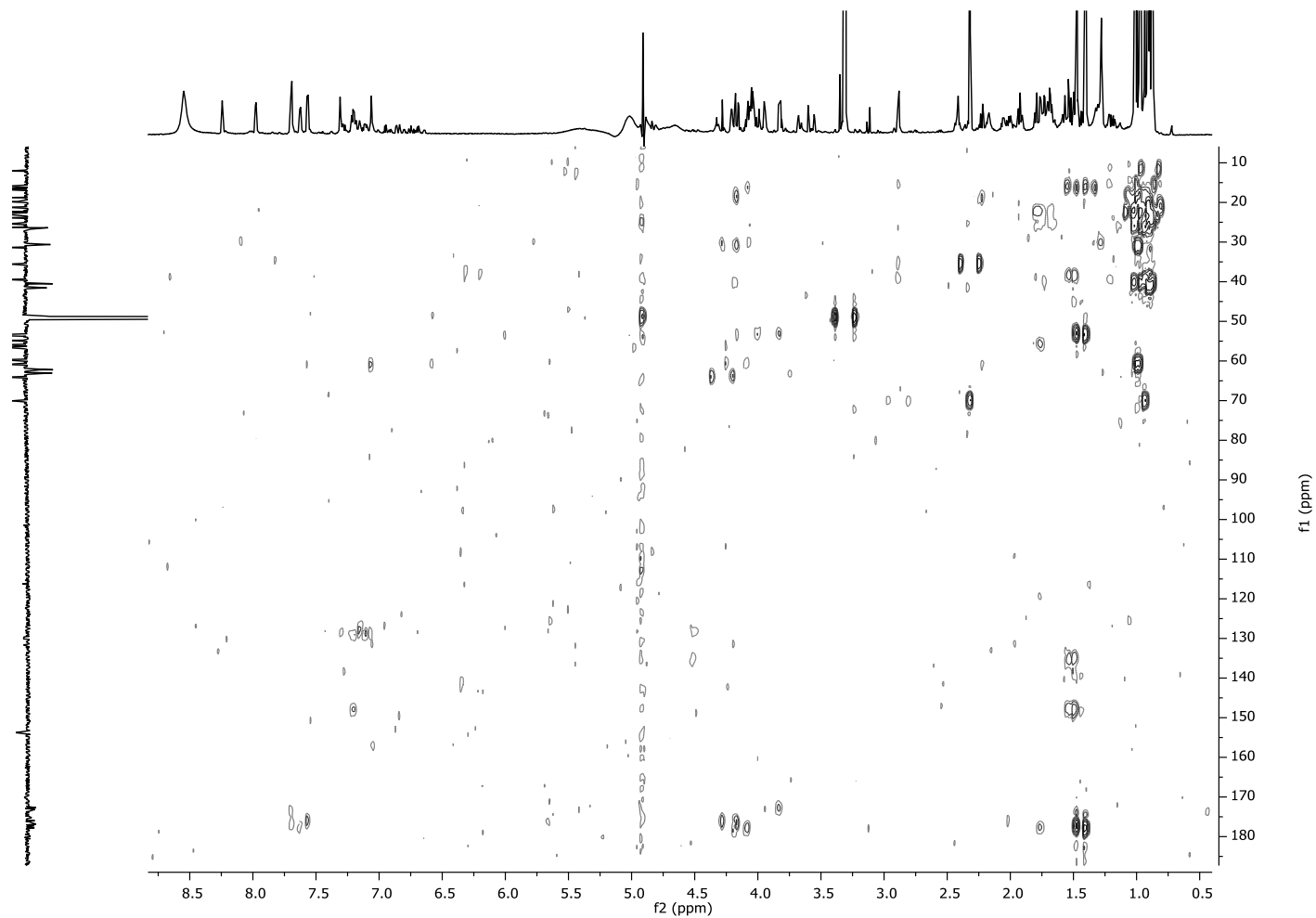
**S10.** Edited HSQC spectrum (900 MHz, methanol-d<sub>4</sub>) of **1**. X: impurities



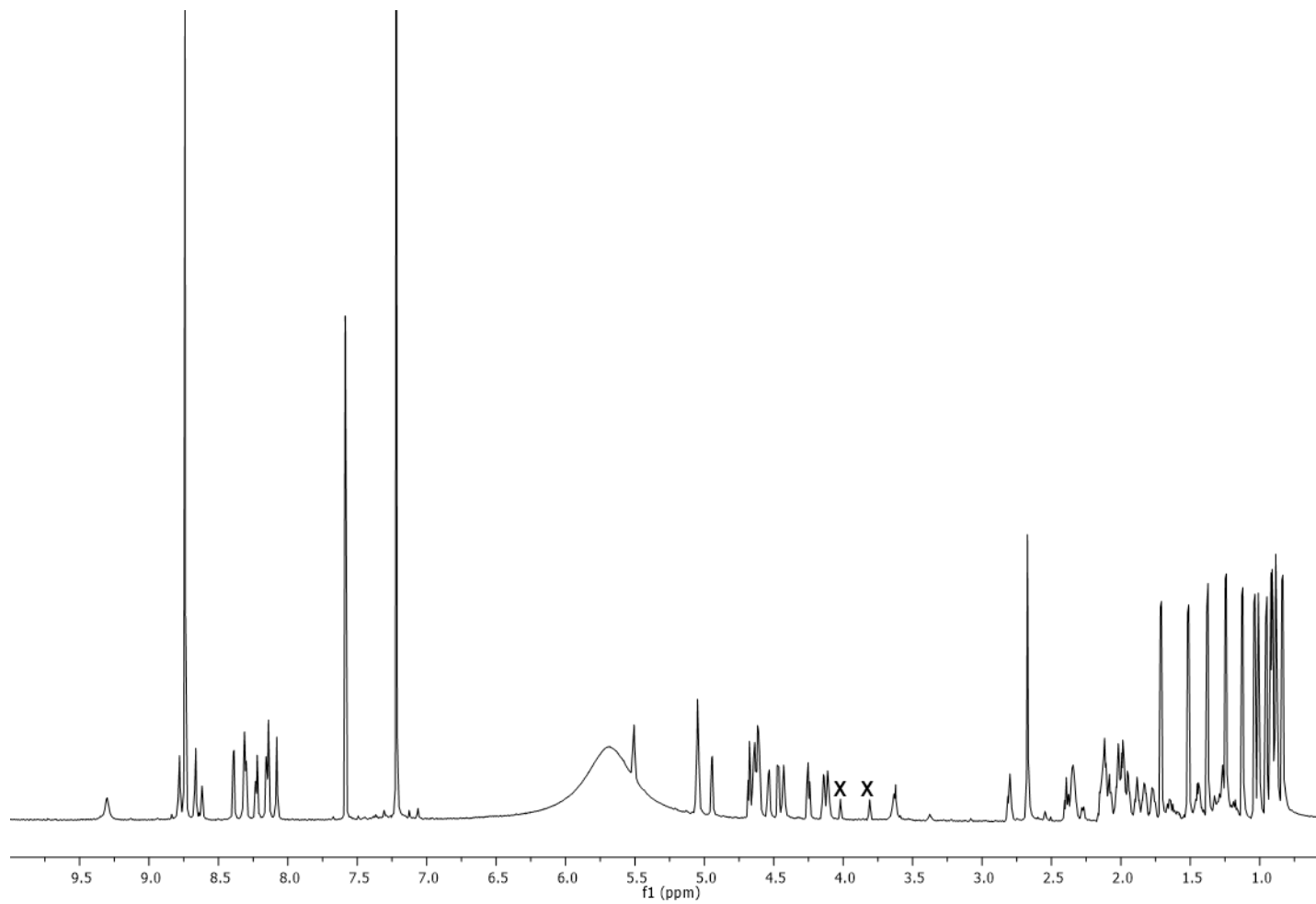
S11. HMBC spectrum (900 MHz, methanol-d<sub>4</sub>) of 1. X: impurities



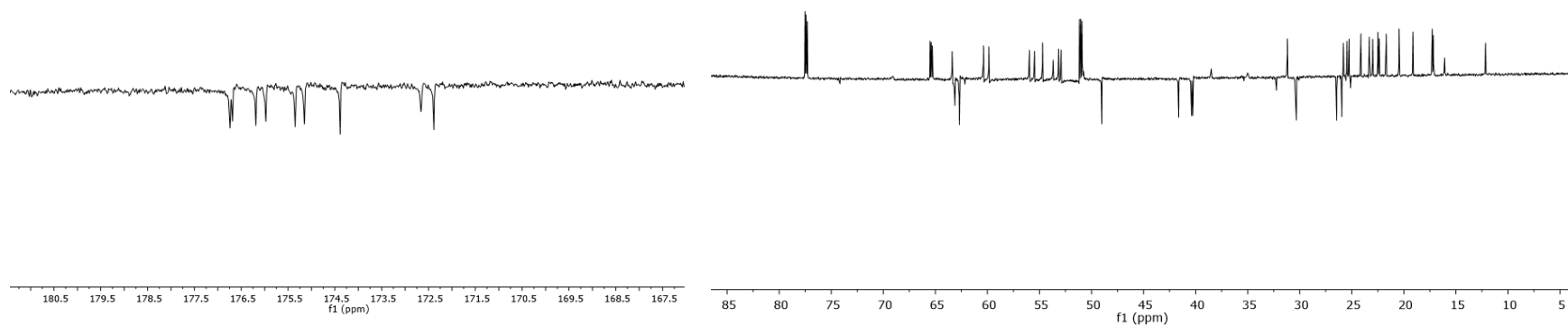
S12. HMBC spectrum (900 MHz, methanol-d<sub>3</sub>) of **1**



**S13.**  $^1\text{H}$  NMR spectrum (900 MHz, pyridine- $d_5$  with TFA vapour) of **1**. X: impurities

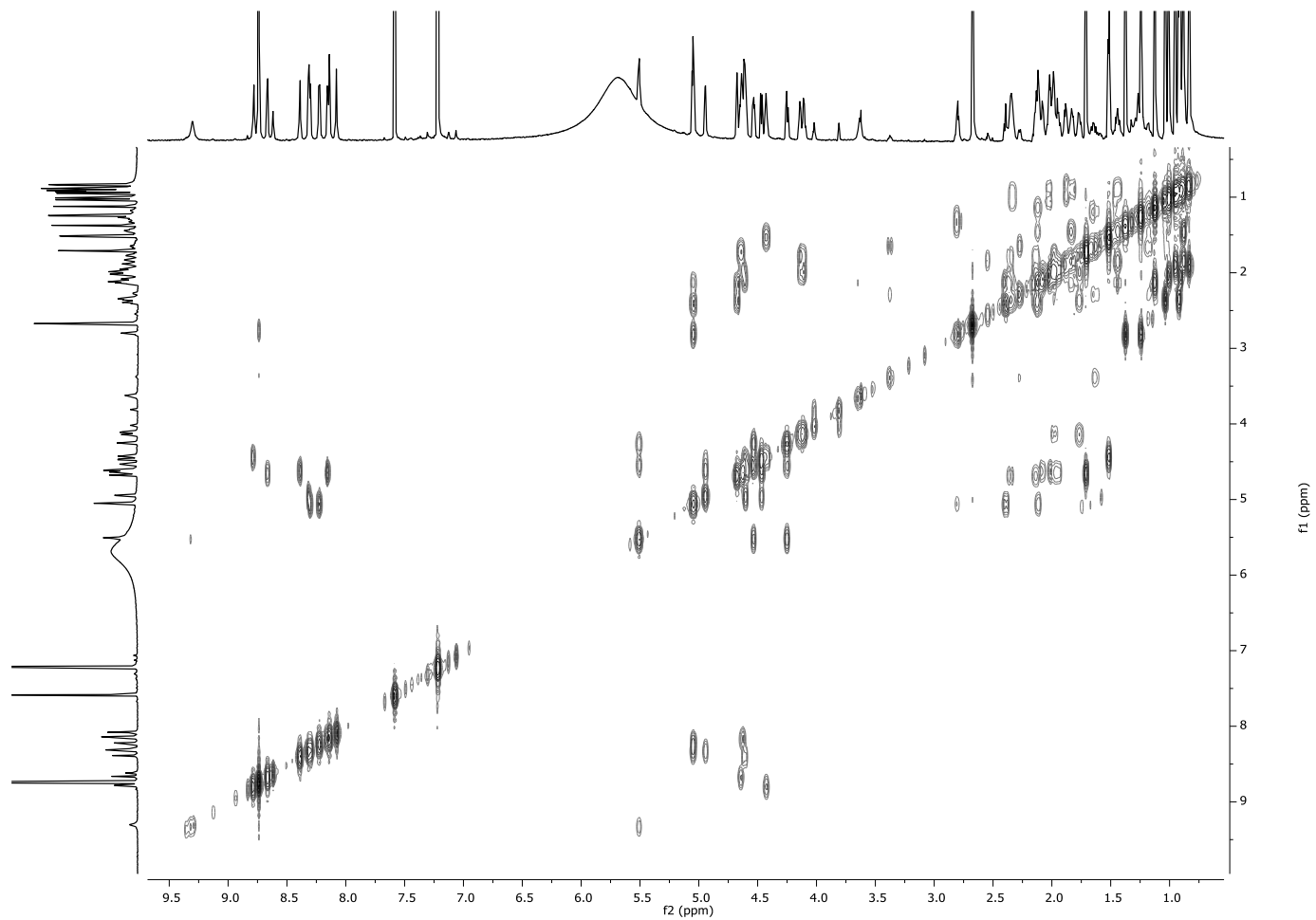


**S14.** DEPTQ spectrum (226 MHz, pyridine-d<sub>5</sub> with TFA vapour) of **1**

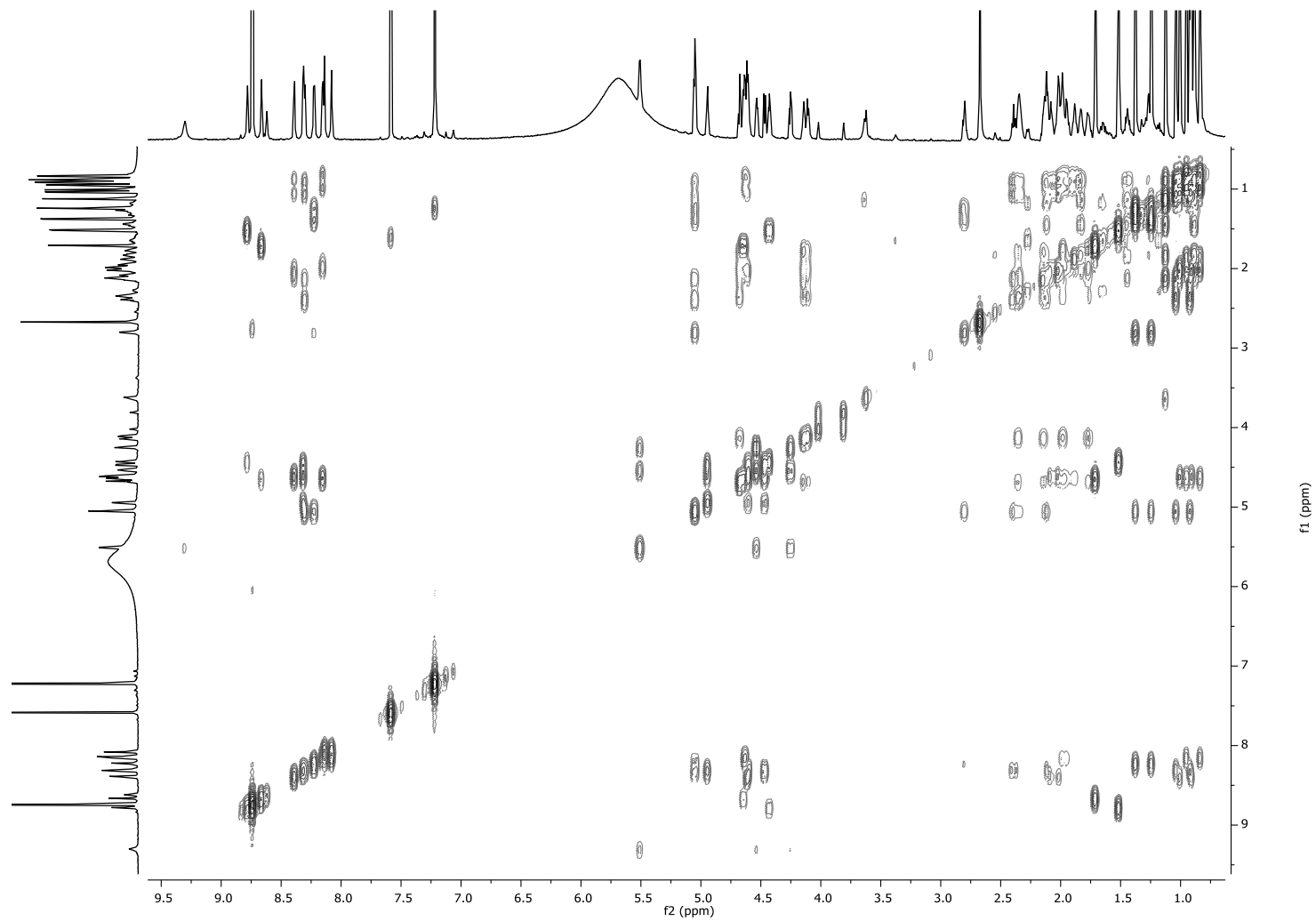




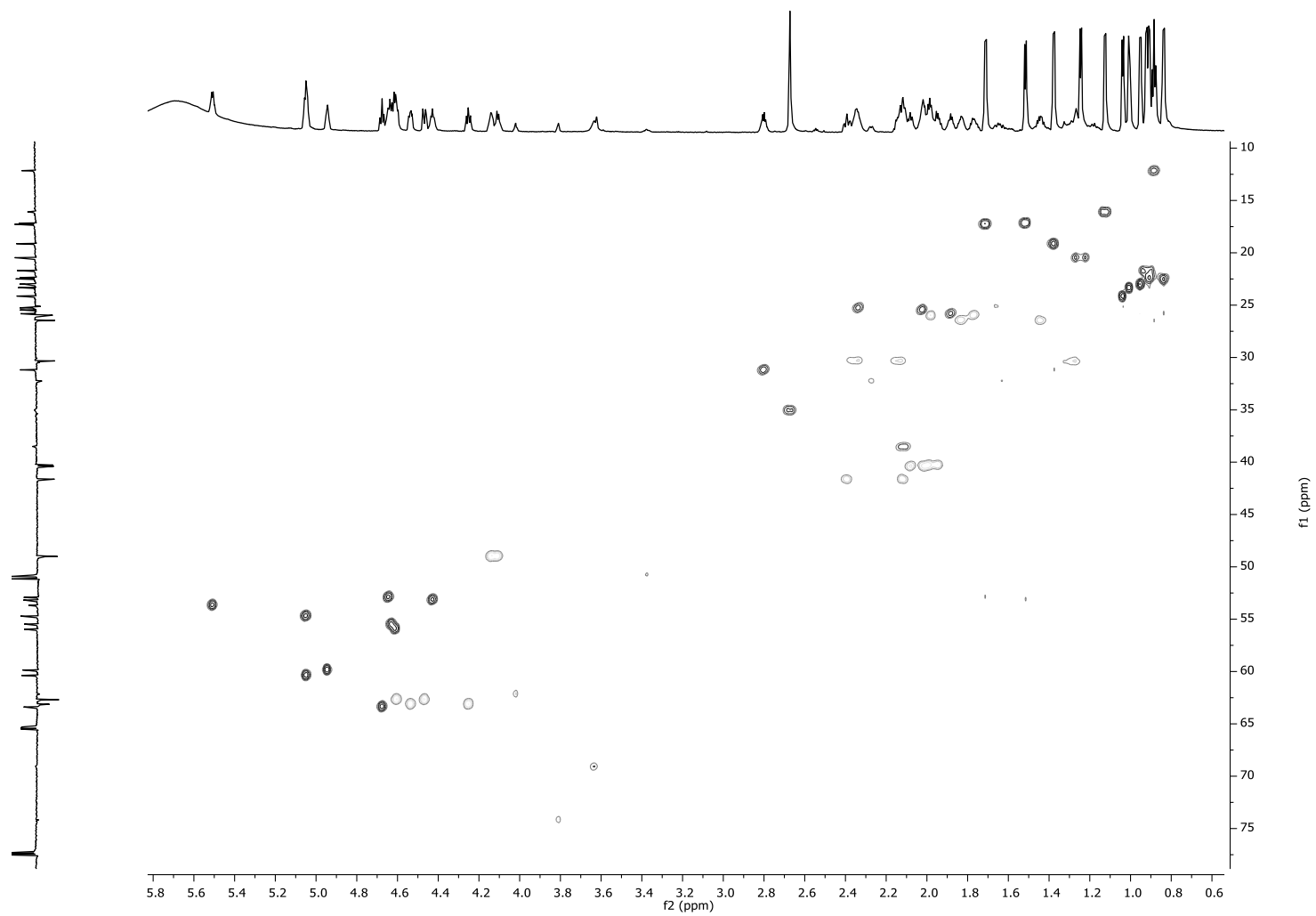
S15. COSY spectrum (900 MHz, pyridine-d<sub>5</sub> with TFA vapour) of **1**



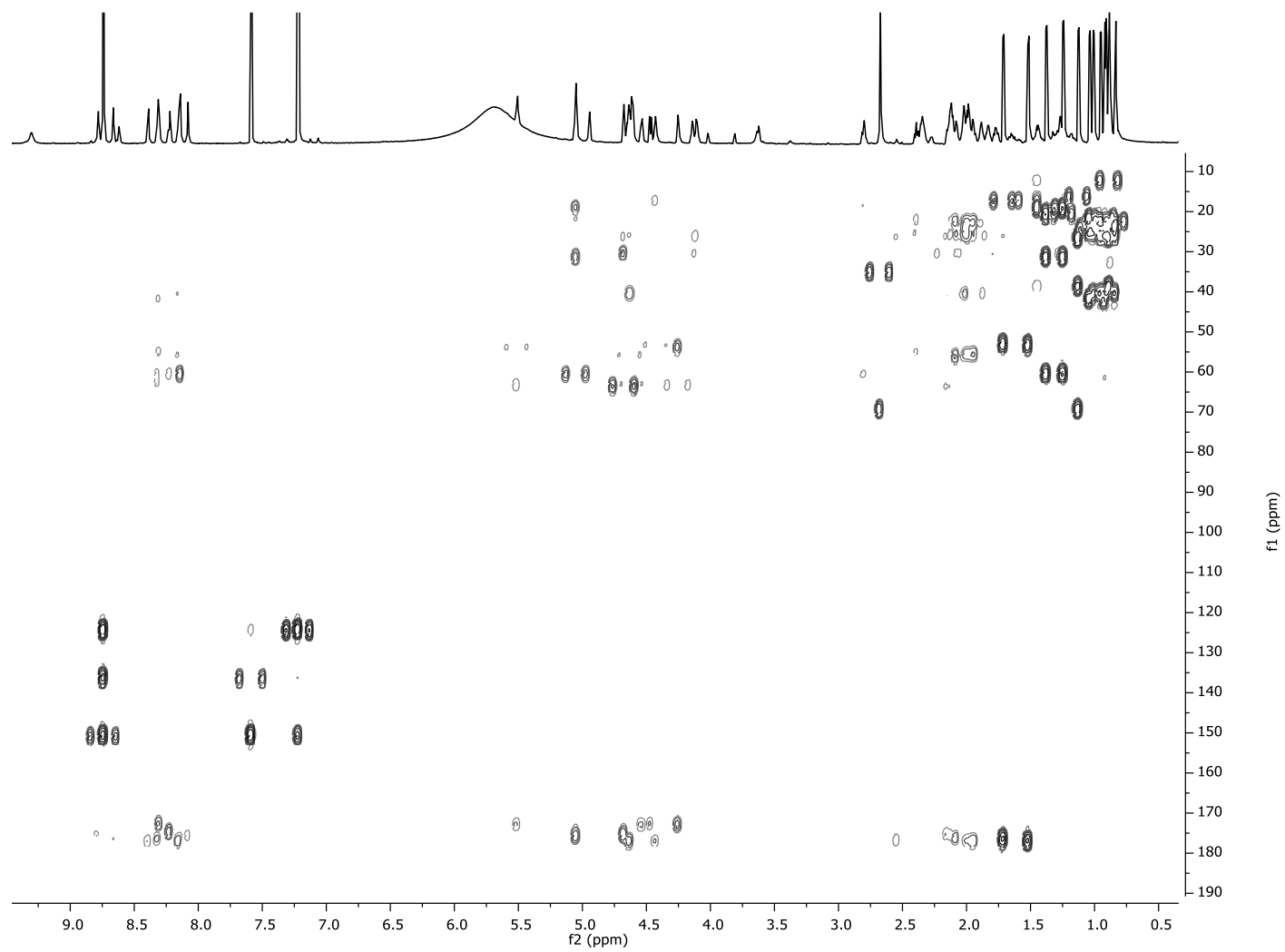
**S16.** TOCSY spectrum (900 MHz, pyridine-d<sub>5</sub> with TFA vapour) of **1**



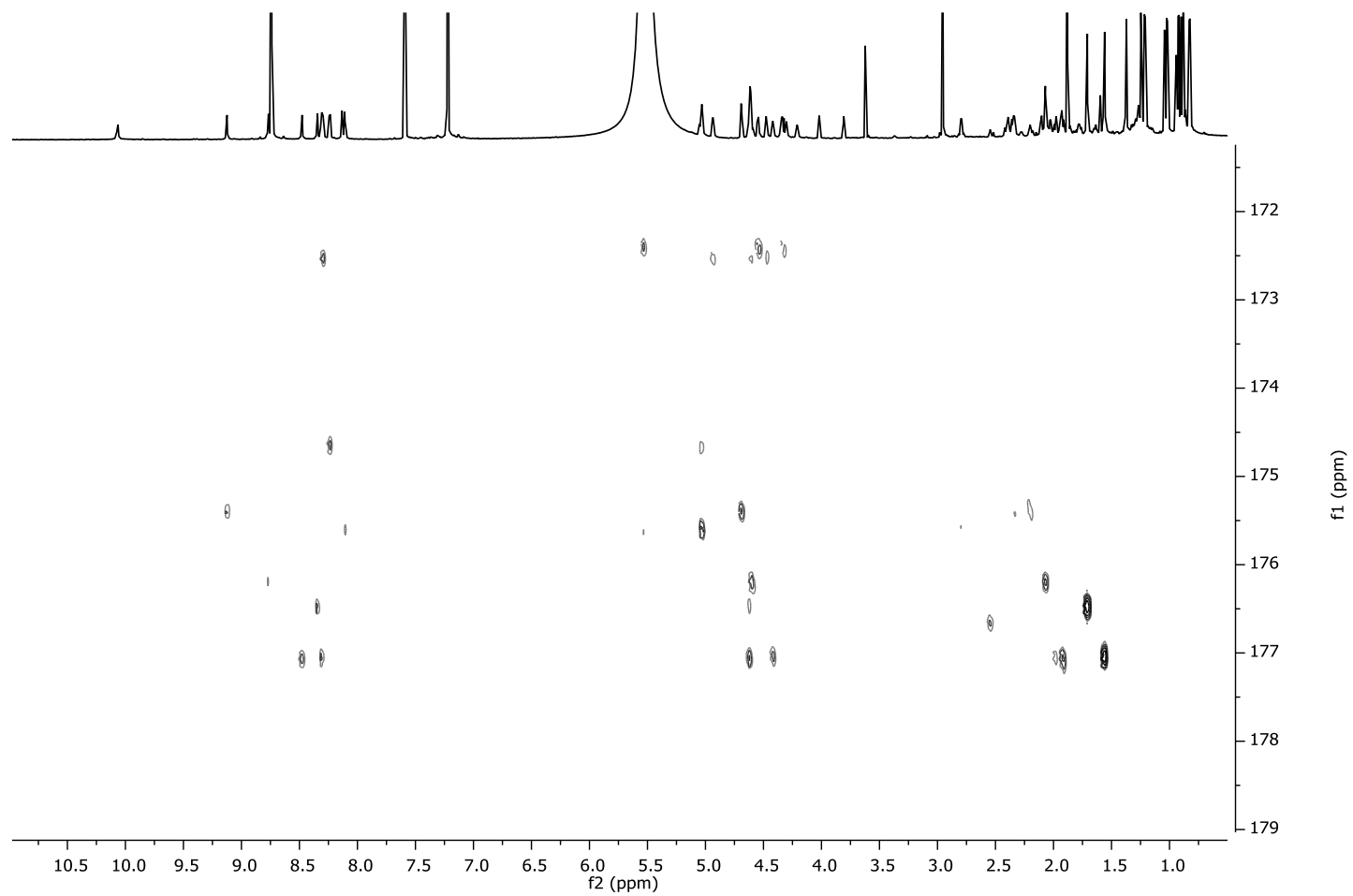
S17. Edited HSQC spectrum (900 MHz, pyridine-d<sub>5</sub> with TFA vapour) of **1**



S18. HMBC spectrum (900 MHz, pyridine-d<sub>5</sub> with TFA vapour) of **1**



**S19.** Band-selective HMBC spectrum (900 MHz, pyridine-d<sub>5</sub> with TFA vapour) of **1**

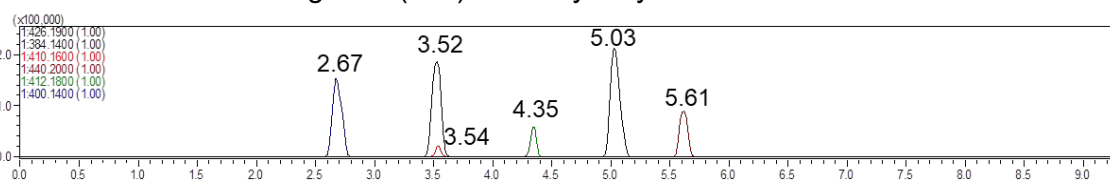


## S20. Marfey's analysis of **1**

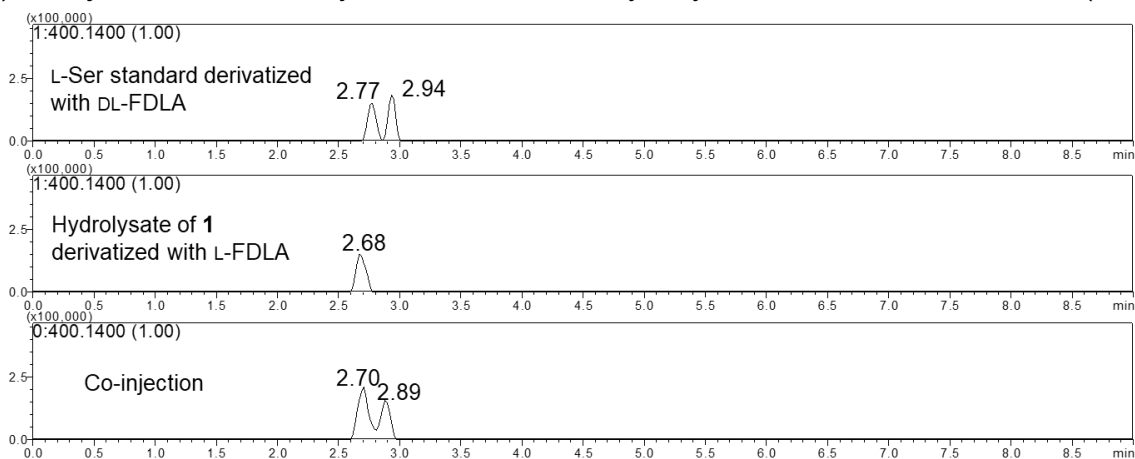
Amino acid ( <i>m/z</i> )	Retention time (min)			Assignment
	L	D	Measured	
Val (412.18)	4.35	5.76	4.35	L
Ser (400.14)	2.70	2.89	2.67	L*
Leu (426.19)	5.03	6.57	5.03	L
Ala (384.14)	3.55	4.36	3.52	L
Pro (410.16)	3.56	4.16	3.54	L
<i>N</i> -Me-Ile (440.2)	5.58	6.47	5.61	L*
<i>allo-N</i> -Me-Ile (440.2)	5.64	6.52	--	

\*Verified by co-injection (see below)

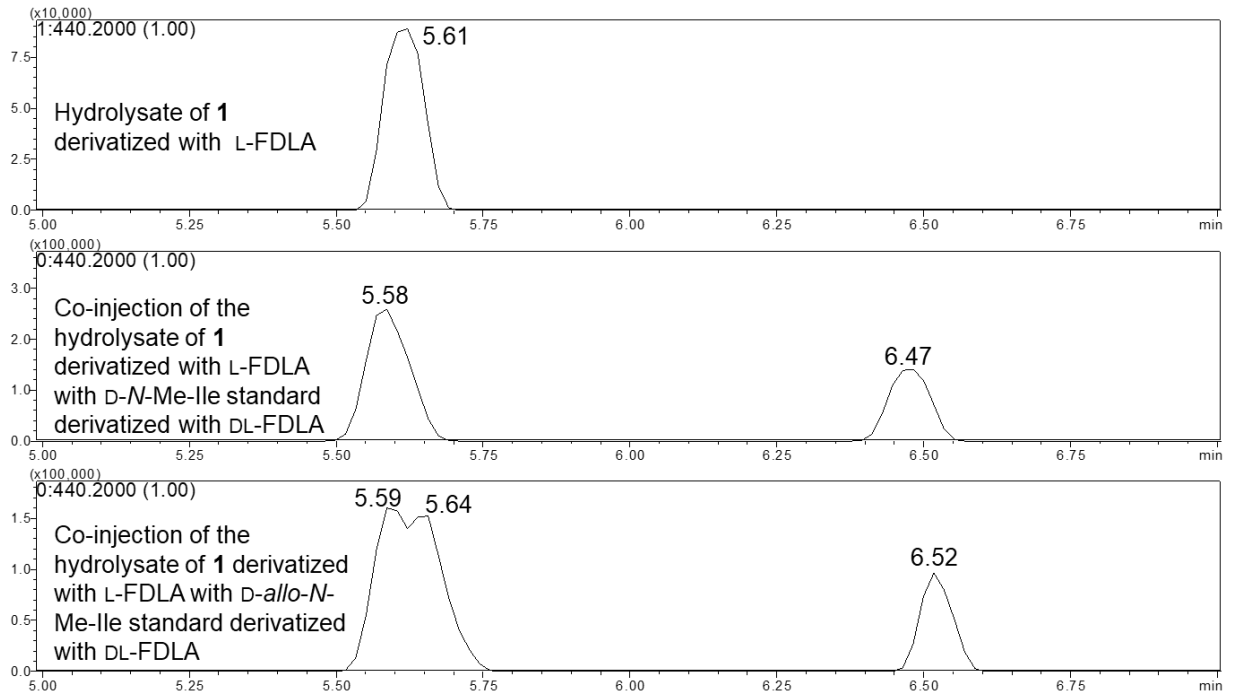
### a) Extracted ion chromatograms (EIC) of the hydrolysate of **1** derivatized with L-FDLA



### b) Co-injection of the Marfey's derivatives of the hydrolysate of **1** and L-Ser standard (EIC)



b) Co-injections of the Marfey's derivatives of the hydrolysate of **1** and D-N-Me-Ile or D-*allo*-N-Me-Ile standards (EIC)



**S21. BLASTP\*** results for putative *scd* biosynthetic gene cluster

<b>Protein</b>	<b>Length (aa)</b>	<b>Description</b>	<b>Organism</b>	<b>Accession number</b>	<b>Identity/coverage (%)</b>
ORF-2	188	hypothetical protein	<i>Scytonema hofmannii</i>	WP_017742275.1	84/100
ORF-1	119	hypothetical protein	<i>Pseudomonas</i> sp. 286	WP_122523577.1	27/71
ScdC	56	cyanobactin biosynthesis PatC/TenC/TruC family protein	<i>Nostoc linckia</i>	WP_099072933.1	82/87
ScdB	65	cyanobactin biosynthesis system PatB/AcyB/McaB family protein	<i>Nostoc linckia</i>	WP_099068421.1	91/100
ScdMT	587	PatA/PatG family cyanobactin maturation protease (Blast identifies two methyltransferase domains at the N-terminus of this <i>Cyanothece</i> protein)	<i>Cyanothece</i> sp. PCC 7822	WP_013335048.1	79/100
ScdA	697	PatA/PatG family cyanobactin maturation protease	<i>Nostoc linckia</i>	WP_099068419.1	77/100
ScdE	70	hypothetical protein	<i>Cyanothece</i> sp. PCC 7822	WP_013335046.1	63/95
ScdTA	627	asparagine synthase (glutamine-hydrolyzing)	<i>Nostocales cyanobacterium</i> HT-58-2	WP_087538154.1	89/99
ORF+1	82	hypothetical protein	<i>Scytonema hofmannii</i>	WP_017742276.1	99/100
ORF+2	415	CCA tRNA nucleotidyltransferase	<i>Tolypothrix bouteillei</i>	WP_038080351.1	90/99

\*protein-protein blast against refseq protein database, accessed April 2019



**Table S22. Sequence comparison of known cyanobactin precursor peptides and ScdE.**

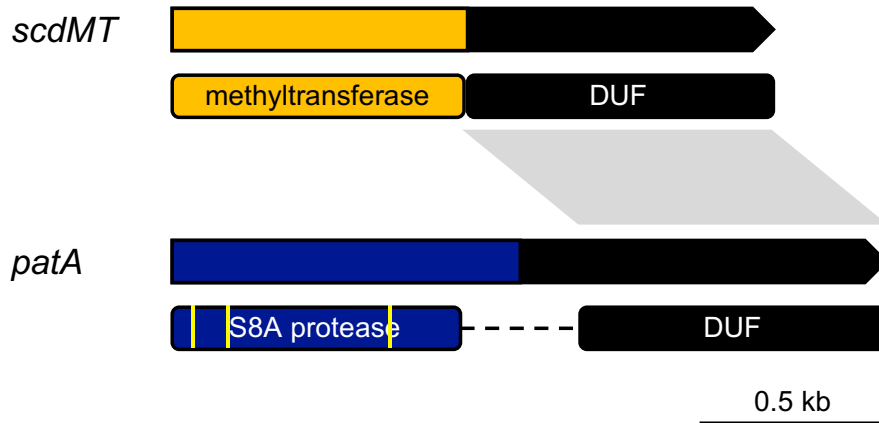
Protein	Organism	Accession number	Leader sequence	RSI	Spacer	RSII	Core	RSIII	Compound
ScdE	<i>Scytonema</i> sp. UIC10036	This study	MRPENQKENRCLQPKLSEPMARSHAGM	PVDLSEEEL	TAEVVNG	GVFAS GVTAS	MYGGTDG <sup>1</sup> ISPALLASLV <sup>2</sup>	----- -----	scytodecamide
AgeE	<i>Microcystis aeruginosa</i> PCC 9432	WP_043998581.1	MDKKNILPHQGKPVLRRTNGKLPSH	LAELSEEAL	GGA	GMDAS GADAS GADAS	FFPC FFPVC FFPC	SYD---- SYD---- SYDDGDA	aeruginosamide
LynE	<i>Lynbya</i> sp. PCC 8106	WP_071941443.1	MDKKNILPHQGKPVLRRTNGKLPSH	LAELSEEAL	GGN	GVDAS GVDAS GVDAS	ACMPCYP VCMPCYP VCMPCYP	SYD--- SYD--- SYDDAE	aestuaramide
McaE	<i>Microcystis aeruginosa</i> PCC 9809	CCI23769.1	MDKKNLLPNQGAPVIRGISGKLPSH	LAELSEEAL	GGN	GLEAS GAEAS GAEAS	YTSSIC VLATFC VTVTIC	AFD---- AFD---- AFDGDEA	microcyclamide
PagE6	<i>Planktothrix agardhii</i> NIES-596	AED99432.1	MTKKNLKPQQAAPVQREINTTSSES	-----	GTST	GLTPH	INPYLYP	FAGDDAE	prenylagaramide
PatE1	<i>Prochloron didemni</i>	AAY21154.1	MNKKNILPQQQPVIRLITAGQLSSQ	LAELSEEAL	GDA	GLEAS GVEPS	VTACITFC ITVCISVC	--AYD-- --AYDGE	patellamide
TenE	<i>Nostoc spongiaeforme</i> var. <i>tenu</i> str. Carmeli	ACA04484.1	MDKKNILPQQGKPVIRITGQLPSF	LAELSEEAL	GDA	GVGAS GAGAS GAGAS GAGAS	ATGCMC ATGCMC ATACAC ATACAC	--AYD-- --AYD-- --AYD-- --AYE--	tenucyclamide
ThcE4	<i>Cyanothece</i> sp. PCC 7425	WP_012626005.1	MDLQNLPPQSQPIQRATAGQLPTE	LAELTEEAL	NNES	AVLAS	SCDCSLYGGCESC	SYEGDEAE	cyanothecamide
TriG	<i>Trichodesmium erythraeum</i>	WP_044136730.1	MGKKNIQPNSSQPVFRSLVARPA	LEELREENL	TEGNQGHGPLAN	GPGPS	GDGLHPRLCSCS	-YDGDE	trichamide

TruE1	uncultured <i>Prochloron</i> sp. 06037A	ACA04491.1	MNKKNILPQLGQPVIRLTAGQLSSQ	LAELSEEAL	G	GVDAS GVDAS	TLPVPTLC TVPTLC	SYD---- SYDD---	trunkamide
VirE	<i>Oscillatoria</i> <i>nigroviridis</i> PCC 7112	AFZ08006.1	MNKKNILPNPGKPVIRGISGKLPSY	LAELSEEAL	GDA	GADAS SVDAS SVDAS SVDAS	FIC FIC FIC FIC	----- ----- ----- SVDGDA	viridisamide

<sup>1</sup>This putative core peptide was not identified.

<sup>2</sup>Scytodecamide.

**S23.** Comparison of ScdMT from *Scytonema* sp. UC 10036 and PatA from *Prochloron didemni* (accession number AAY21150.1). Predicted domains are shown below the genes. The dotted line indicates the approx. 120 aa linker between the protease and DUF domains. Catalytic triad of the S8A protease is highlighted in yellow. The homolog region is connected by a grey area.



**S24.** Percent identity (similarity) shared between the DUF domains of ScdA and ScdMT with PatA and three PatA homologues.

	<b>AgeA-DUF</b>	<b>PatA-DUF</b>	<b>TruA-DUF</b>	<b>VirA-DUF</b>
<b>ScdA-DUF</b>	72% (92%)	72% (92%)	72% (92%)	72% (91%)
<b>ScdMT-DUF</b>	65% (89%)	65% (89%)	65% (89%)	66% (88%)

**S25.** Percent identity (similarity) shared between the DUF domains of ScdA and ScdMT with PatG and three PatG homologues.

	<b>AgeG-DUF</b>	<b>PatG-DUF</b>	<b>TruG-DUF</b>	<b>VirG-DUF</b>
<b>ScdA-DUF</b>	52% (80%)	53% (80%)	52% (80%)	54% (81%)
<b>ScdMT-DUF</b>	55% (82%)	55% (81%)	53% (81%)	54% (81%)