Electronic Supplementary Material (ESI) for Organic & Biomolecular Chemistry.

Supplementary information

Genomics-Driven Discovery of a Linear Lipopeptide Promoting Host Colonization by Endofungal Bacteria

Sarah P. Niehs[1], Kirstin Scherlach[1], Christian Hertweck[1,2]

[1]Department of Biomolecular Chemistry, Leibniz-Institute for Natural Product Research and Infection Biology (HKI), Beutenbergstrasse 11a, 07745 Jena, Germany
[2]Friedrich Schiller University, 07743 Jena, Germany

Table of Contents

Supplemental methods

Media

All media were sterilized at 120 °C for 20 minutes (Tab. S1).

Identification of the holrhizin biosynthetic gene cluster using bioinformatics tools

Genome data was analyzed using AntiSMASH 4.0[1] and NRPSpredictor2[2].
Putative genes belonging to the holrhizin gene cluster are annotated in table S3. The amino acid sequence of each gene was compared to the Swiss-prot database or the protein data base (pdb) of NCBI.

NCBI[3] search revealed homologous genes to RBRH01792 and surrounding genes in *Burkholderia* sp. b13, *Burkholderia* sp. b14, and *B. endofungorum*. The disrupted NRPS genes are probably due to incomplete genomic sequences (Figure S1).

General analytical methods

HPLC/MS: Agilent Technologies 1100 Series HPLC with a UV detector (Agilent 1100 G1315 DAD Detector) using a reversed phase phenomenex column (Luna, C18(2), 10 µ, 250x4.6 mm) and a gradient flow of 10–100% MeCN for 20 min in destilled water (+0.1% formic acid), flow rate: 1 mL min^{-1}, mass spectrometer: Bruker HCT ultra with electrospray ion source and ion trap mass analyzer.

MS/MS (tandem mass spectrometry): QExactive Orbitrap High Performance Benchtop LC/MS (ThermoFisher, Germany) with an electron spray ion source and an Accela HPLC System, C18 column (Accucore C18 2.6 µm, 100 x 2.1 mm, Thermo Fisher Scientic, Germany), solvents: MeCN and destilled water (both supplemented with 0.1% formic acid), flow rate: 0.2 mL min^{-1}; program: 0–10 min 5–98% MeCN, hold 4 min 98% MeCN, 14–14.1 min 98% to 5% MeCN, hold 6 min at 5% MeCN.

Preparative HPLC: Shimadzu LC-8A HPLC system with a VP250/21 Nucleodur column (C18, HTec, 5 µm), a flow rate of 15 mL min^{-1}, UV detection at 190 nm, and a gradient of MeCN/destilled H_2O with 0.01% TFA ranging from 40% to 100% MeCN in 25 min.

NMR: Recorded in DMSO-d_6 on Bruker 600 MHz Avance III Ultra Shield (Bruker, Germany); signal reference: respective solvent signals.

IR: Using a JASCO spectrometer FT/IR-4600 (Jasco, Germany).

Physicochemical data of holrhizin A

White amorphous solid: (-)-ESI-MS *m/z* 815 $[M + H]^-$, (+)-ESI-MS *m/z* 817 $[M + H]^+$. HRESI/MS: *m/z* $[M + H]^+$ = 817.5057 (calculated for $C_{42}H_{69}N_6O_{10}$ $[M + H]^+$, 817.5070). UV (PDA): λ_{max}= 190 nm.

Supplemental tables

Tab. S1. Composition of used media in this study.

Medium or *medium additive*	**Composition (L^{-1})**
MGY+M9 medium	10 g glycerol, 1.25 g yeast extract (technical yeast extract, BD, Bacto®), 960 mL distilled water, after sterilization add: 20 mL M9 salt A, 20 mL M9 salt B
NAG agar	Nutrient agar (Bacto, BD), 1% glycerol
PDA	Potato dextrose agar (Bacto, BD)
Double selection agar	24 g agar, 10 g glycerol, 20 mL M9 salt A, 20 mL M9 salt B, 2 g L -amino acid mix, 1 g D,L-4-chlorophenylalanine, sterilization, add 2 mL vitamin solution, 1 mL trace element solution, 16.8 mL of 100 mM L-leucine solution, 5 mL of 60 mM L -histidine solution, 10 mL of 100 mM L-lysine solution, 10 mL of 40 mM L -tryptophan solution, 10 mL of 40 mM L-methionine solution
King's B medium	20 g proteose peptone Nr.3 (Bacto®, BD), 1.5 g K_2HPO_4 , 1.5 g $MgSO_4$ x 7 H_2O, 20 g glycerol, adjust pH to 7.3
VK medium	5 g glycerol, 10 g yeast extract, 10 g corn starch, 10 g corn step solids, and 10 g $CaCO_3$, pH 6.5
M9 salt solution A	350 g K_2HPO_4, 100 g KH_2PO_4
M9 salt solution B	29.4 g sodium citrate, 50 g $(NH_4)_2SO_4$, 5 g $MgSO_4$
Vitamin solution	10 mg folic acid, 6 mg biotin, 200 mg *p*-aminobenzoic acid, 1 g thiamine-HCl, 1.2 g pantothenic acid, 1 g riboflavin, 2.3 g nicotinic acid, 12 g pyridoxine HCl, 0.1 mg vitamin B_{12}
L-Amino acid mix	2 g of each L-amino acid: alanine, asparagine, cysteine, glutamate, isoleucine, serine, tyrosine, arginine, aspartate, glutamine, glycine, proline, threonine, valine
Trace element solution	40 mg $ZnCl_2$, 200 mg $FeCl_3$ x 6 H_2O, 10 mg $CuCl_2$ x 2 H_2O, 10 mg $MnCl_2$ x 4 H_2O, 10 mg $Na_2B_4O_7$ x 10 H_2O, 10 mg $(NH_4)_6Mo_7O_{24}$ x 4 H_2O

Tab. S2. Prediction of NRPS modules according to NRPSpredictor2[2]; C – condensation, A – adenylation, T – peptidyl carrier protein, TE – thioesterase.

Modules of RBRH01792	**Specificity code (NRPSpredictor2[2])**	**Predicted amino acid sequence**	**Amino acid sequence of holrhizin A**
M1 (C-A-P)	LNNAFDASTFEAWLIVGGDINGYGPTESTTFTTT DAFWIGGTFK	Val (100%)	Val
M2 (C-A-P)	LAQAFDASVFEMTLILAGEFNAYGPTETTVCVSA DAFTIAAVCK	Phe (90%)	Phe
M3 (C-A-P)	YFLTFDPCVRDGSILTSGEVNQYGPSECTMASTW DPRSLSQMAK	Leu (40%), for large clusters: Asp, Asn, Glu, Gln	Glu
M4 (C-A-P)	LNTSFDATTFETFLLFGGELHVYGPTETVTYASW DAFFLGVTYK	Ile (90%)	Ile
M5 (C-A-P)	FWATFDLSIYEVNTNMAGECNLYGPSESTTYSTW DLYNNALTYK	Ala (100%)	Ala
M6 (C-A-P)	LNTSFDATTFETFLLFGGELHVYGPTETVTYASW DAFFLGVTYK	Ile (90%)	Ile

Tab. S3. Proteins encoded up- and downstream of the holrhizin gene cluster with NRPS gene RBRH01792; Id - Identity; Si - Similarity.

Gene RBRH	Length [bp]	Putative protein	Closed characterized orthologous protein (organism)	Accession number	Id/ Si
01801	1,497	Transporter, MFS superfamily	Multidrug resistance protein MdtG (*Enterobacter lignolyticus*)	E3GC98.1	29%/ 27%
01800	1,452	6-Phospho-gluconate dehydrogenase	Apoform of dimeric 6-phosphogluconate dehydrogenase (*Klebsiella pneumoniae*)	2ZYG_A	70%/ 81%
01799	804	Sensory box/GGDEF family protein	Pamora dimeric phosphodiesterase apo-form MorA (*Pseudomonas aeruginosa*)	4RNI_A	29%/ 45%
			Oxygen sensor protein DosP (*Escherichia coli*)	DOSP_ECOLI	24%/ 44%
01798	2,187	Enoyl-CoA hydratase	Fatty acid beta-oxidation multienzyme complex (*Pseudomonas fragi*)	1WDK_A	29%/ 49%
01796	1,101	DNA topoisomerase I	Bacterial topoisomerase Ib (*Deinococcus radiodurans*)	3M4A_A	44%/ 59%
01793	447	Transposase	Putative transposase; no fully characterizied protein structure for comparison	-	-
01792	19,776	Non-ribosomal peptide synthetase	Tyrocidine synthase 3 (*Brevibacillus parabrevis*)	O30409	38%/ 55%
00214	1,536	Glycerol kinase	Glycerol kinase (*Escherichia coli*)	1BU6_O	59%/ 75%
00215	711	Glycerol uptake facilitator protein	Glycerol facilitator Glpf (*Escherichia coli*)	1FX8_A	33%/ 48%
00218	1,233	Phosphoesterase	Uncharacterized metallophosphoesterase Cj0846 (*Campylobacter jejuni* subsp. *jejuni*)	Y846_CAMJE	31%/ 49%
00219	339	Acetyltransferase, GNAT family	Putative acetyltransferase (*Thermus thermophilus*)	2Z10_A	39%/ 54%
00221	1,062	Lipopolysaccharide heptosyltransferase	Heptosyltransferase Waac (*Escherichia coli*)	2GT1_A	31%/ 51%

Tab. S4. ^{1}H- (600 MHz) and ^{13}C- (150 MHz) NMR shifts of holrhizin A.

	Position	**δ_C [ppm]**	**δ_H [ppm]; Mult. (*J* [Hz])**
Fatty acid chain	C=O	172.2	-
	2	35.2	2.07; 2 H m
	3	25.3	1.44; 2 H m
	4/5	28.6/28.4	1.22; 4 H m
	6	31.3	1.22; 2 H m
	7	22.0	1.25; 2 H m
	8	13.9	0.84; 3 H m
L-Val	C=O	170.9	-
	2	57.6	4.09; 1 H dd (*8.7, 7.1*)
	3	30.3	1.87; 1 H sixt (*6.8*)
	4	19.3	0.74; 3 H dd (*9.7, 6.8*)
	5	17.9	0.74; 3 H dd (*9.7, 6.8*)
	NH	-	7.69; 1 H d (*8.9*)
L-Phe	C=O	171.0	-
	2	53.4	4.55; 1 H m
	3	37.4	3.01; 1 H dd (*14.1, 4.2*)
			2.77; 1 H dd (*14.0, 9.9*)
	4	137.7	-
	5	127.9	7.21; 1 H m
	6	129.2	7.23; 1 H m
	7	126.1	7.15; 1 H m
	NH	-	7.92; 1 H d (*8.1*)
L-Glu	C=O	171.2	-
	2	51.6	4.17; 1 H m
	3	27.3	1.77; 1 H m
			1.95; 1 H m
	4	31.6	2.18; 2 H m
	5	173.0	-
	NH	-	8.15; 1 H d (*7.7*)
L-Ile (A)	C=O	170.8	-
	2	56.6	4.19; 1 H m
	3	36.6	1.67; 1 H m
	4	15.2	0.80; 3 H m
	5	24.3	1.39; 2 H m
	6	11.0	0.78; 3 H m
	NH	-	7.84; 1 H d (*8.8*)

Tab. S4 continued. ^{1}H- (600 MHz) and ^{13}C- (150 MHz) NMR shifts of holrhizin A

	Position	δ_C [ppm]	δ_H [ppm]; Mult. (*J* [Hz])
L-Ala	C=O	172.1	-
	2	48.0	4.35; 1 H quin (*7.1*)
	3	18.1	1.18; 3 H m
	NH	-	8.07; 1 H d (*7.4*)
L-Ile (B)	C=O	172.8	-
	2	56.1	4.16; 1 H m
	3	36.4	1.76; 1 H m
	4	15.4	0.83; 3 H m
	5	24.5	1.39; 2 H m
	6	11.3	0.82; 3 H m
	NH	-	7.81; 1 H d (*8.4*)

Tab. S5. Retention times of amino acids derivatized with Marfey's reagent.

Derivatization with Marfey's reagent	Configuration	Retention time of amino acid standard [min.]	Retention time of hydrolyzed holrhizin A [min.]
Valine	D	21.4	
	L	14.7	15.0
Phenylalanine	D	25.8	
	L	21.4	21.8
Glutamate	D	7.1	
	L	6.4	6.5
Isoleucine	D	25.0	
	L	18.8	19.1
allo-Isoleucine	D	25.0	
	L	18.8	19.1
Alanine	D	12.6	
	L	9.8	10.1

Tab. S6. Retention times of amino acids derivatized with Sanger's reagent.

Derivatization with Sanger's reagent	Configuration	Retention time of standard [min.]	Retention time of hydrolyzed holrhizin A [min.]
Isoleucine	L	29.3	29.2
allo-Isoleucine	D, L	28.2	-

Tab. S7. Results of bioactivity testing against several bacterial and fungal strains; diameter of hole: 9 mm.

Bacterial/fungal strains	Zone of inhibition of holrhizin A [mm]	Zone of inhibition of solvent control [mm]
Bacillus subtilis 6633 B1	11	-
Staphylococcus aureus 511 B3	-	-
Escherichia coli 458 B4	-	-
Pseudomonas aeruginosa SG 137 B7	-	-
Pseudomonas aeruginosa K799/61 B9	-	10
MRSA *Staphylococcus aureus* 134/93 R9	10	-
VRE *Enterococcus faecalis* 1528 R10	10	-
Mycobacterium vaccae 10670 M4	11	-
Sporobolomyces salmonicolor 549 H4	-	10
Candida albicans H8	-	-
Penicillium notatum JP36 P1	-	10

Tab. S8. Oligonucleotides used for construction of holrhizin-deficient mutant.

Oligonucleotides	Base sequence 5' → 3'	PCR-amplified regions
1792_UP_fw	CTATAGGGCGAATTGGGTACAGCTGCGCCATTAAGTAAC	1,454 bp (flanking region I)
1792_UP_rv	ACATTCATCCAGCTGTTAGAAAGCGATG	
KAN_1792_fw	TCTAACAGCTGGATGAATGTCAGCTACTG	1,058 bp (kanamycin resistance cassette)
KAN_1792_rv	CCTATGTTGGTCAGAAGAACTCGTCAAG	
1792_DWN_fw	GTTCTTCTGACCAACATAGGCGAACTTC	425 bp (flanking region II)
1792_DWN_rv	CGGTGGCGGCCGCTCTAGAAGCAACAATGAGCCGATTTG	
1792_Aint_fw	ATCTTGGGCTCTGGTCATCC	Verification primers
1792_Aint_rv	ACGCAATGCAATATCGGCTT	

Supplemental figures

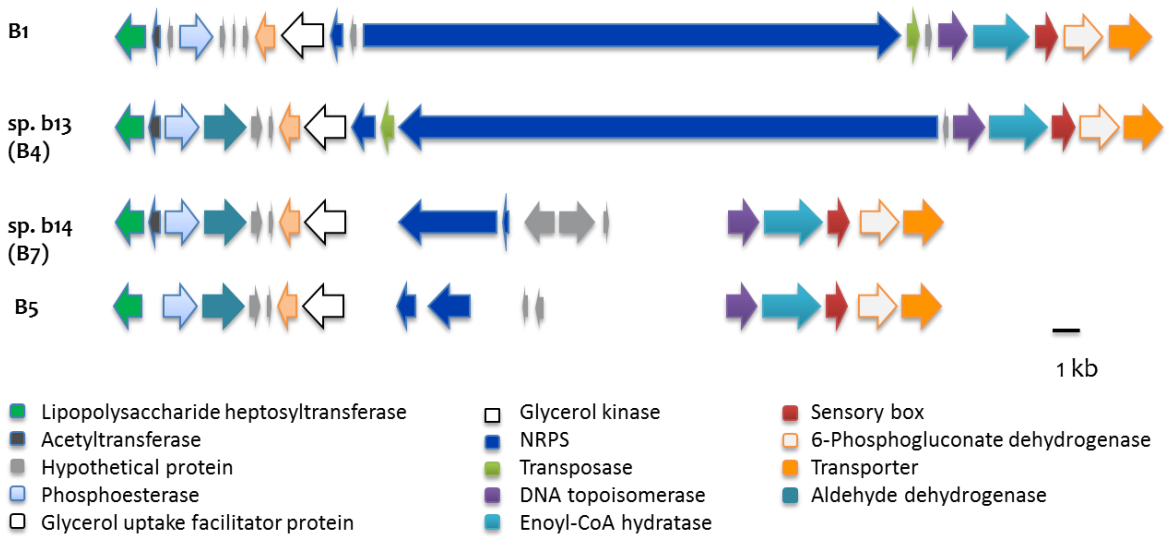
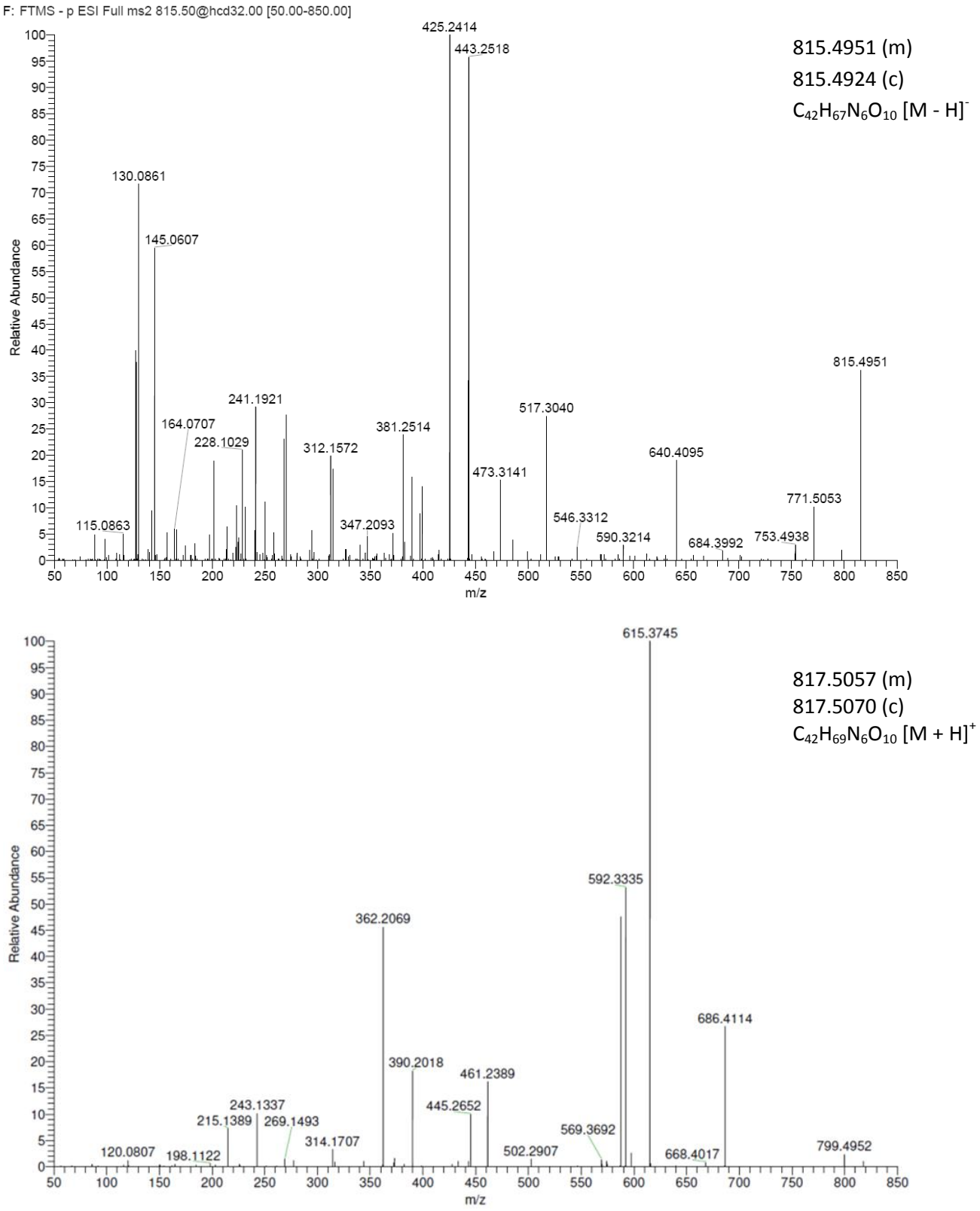


Fig. S1. Distribution of holrhizin NRPS gene cluster in different endofungal symbionts of *R. microsporus*; NRPS, non-ribosomal peptide synthetase; B1 - *B. rhizoxinica* HKI 454; B4 - *Burkholderia* spp. HKI402; B7 - *Burkholderia* spp. HKI403; B5 - *B. endofungorum* HKI 456.

Consecutive numbering of *Burkholderia* strains according to Lackner *et. al.* [4]

B4 and B7 are publically available as *Burkholderia* sp. b13 and *Burkholderia* sp. b14, respectively.



815.4951 (m)
815.4924 (c)
$C_{42}H_{67}N_6O_{10}$ $[M - H]^-$

817.5057 (m)
817.5070 (c)
$C_{42}H_{69}N_6O_{10}$ $[M + H]^+$

Fig. S2. MS/MS fragmentation pattern of holrhizin A as $[M - H]^-$ 815.5 (collision energy 32%; above) and as $[M + H]^+$ 817.5 (collision energy 12%; below) with masses according to LC/MS measurement (m) and calculation (c).

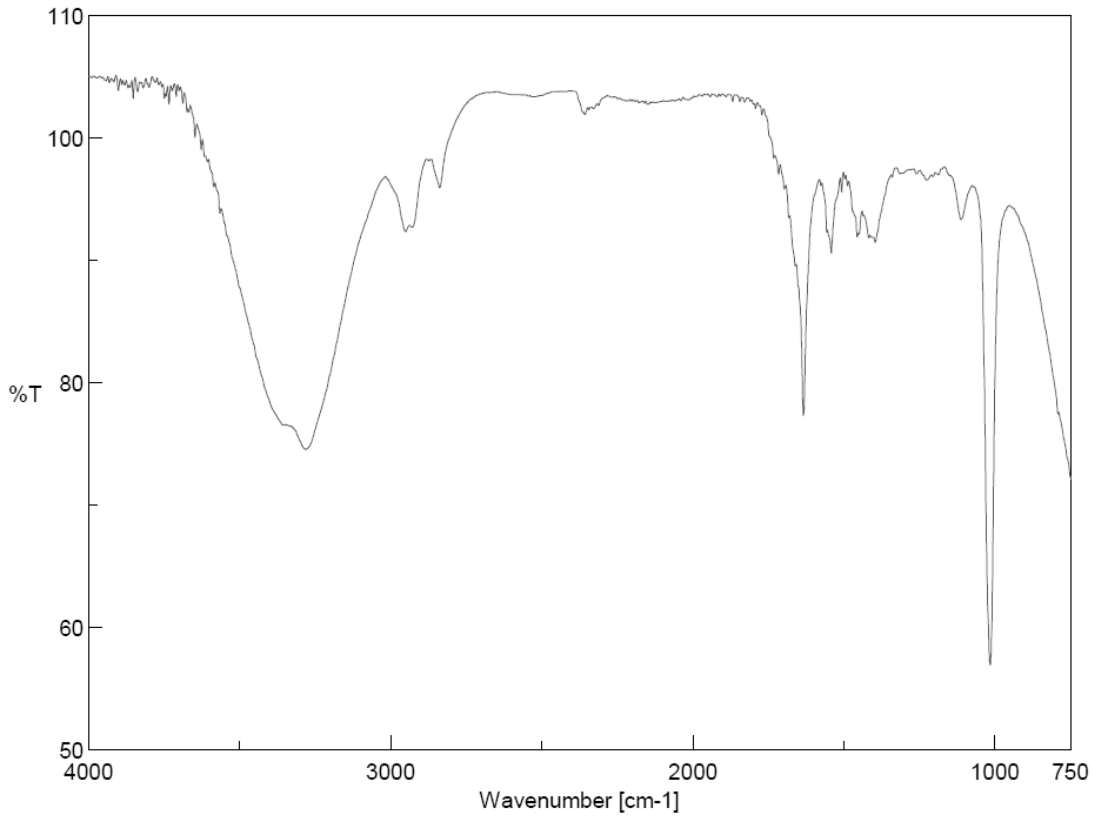


Fig. S3. IR spectrum of holrhizin A.

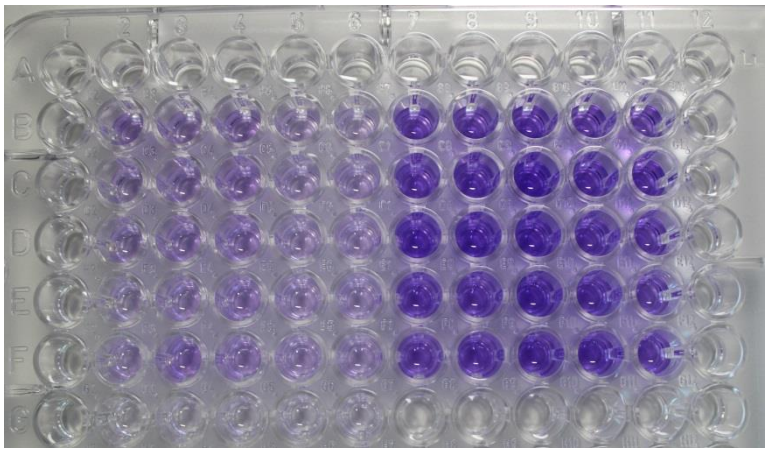


Fig. S4. Biofilm determination assay after addition of ethanol; left: wild type (columns 2-6) and right: holrhizin-deficient strain (columns 7-11).

NMR Spectra

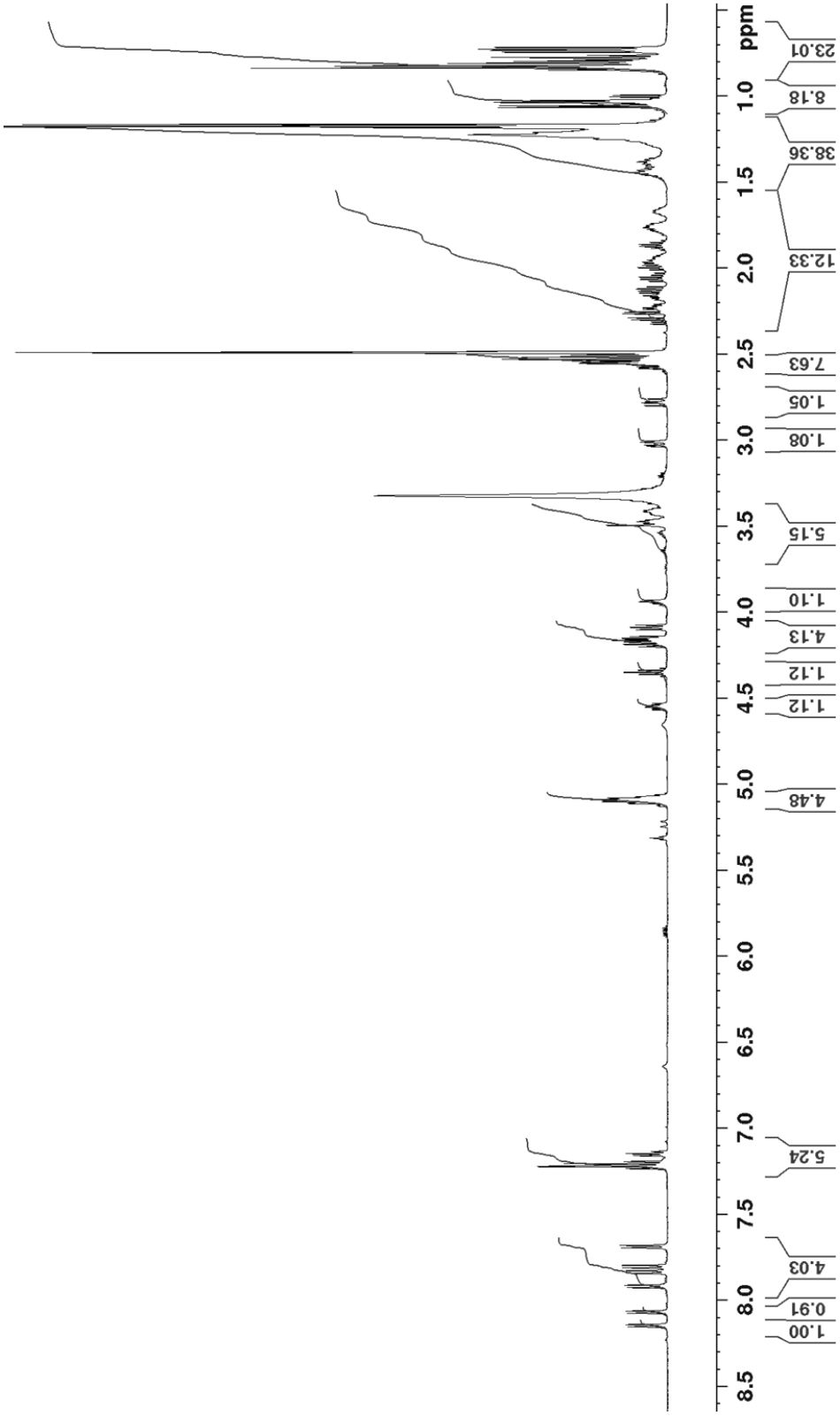


Fig. S5. ^{1}H-NMR spectrum of holrhizin A, recorded at 600 MHz in DMSO-d_6.

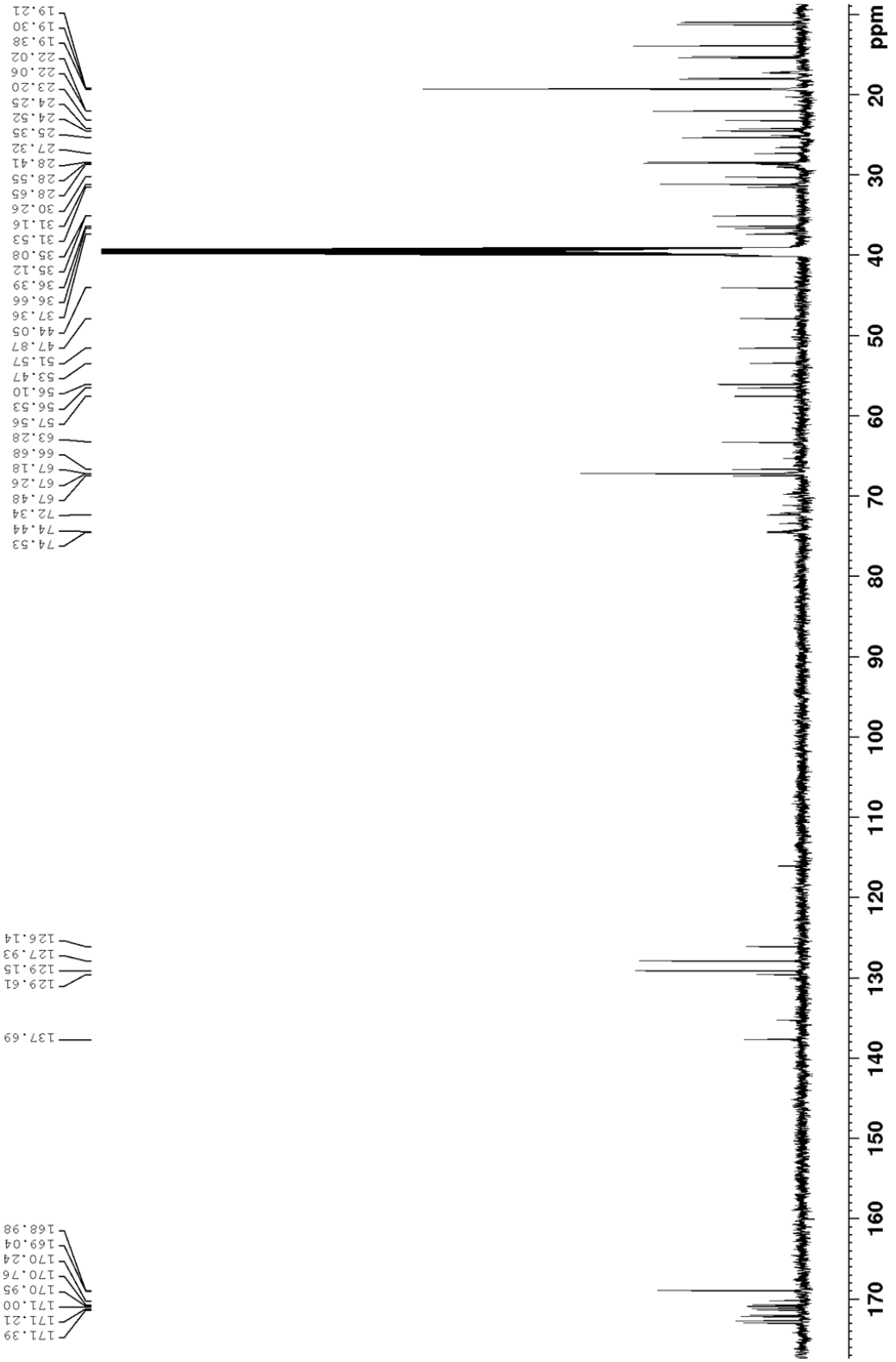


Fig. S6. ^{13}C-NMR spectrum of holrhizin A, recorded at 150 MHz in DMSO-d_6.

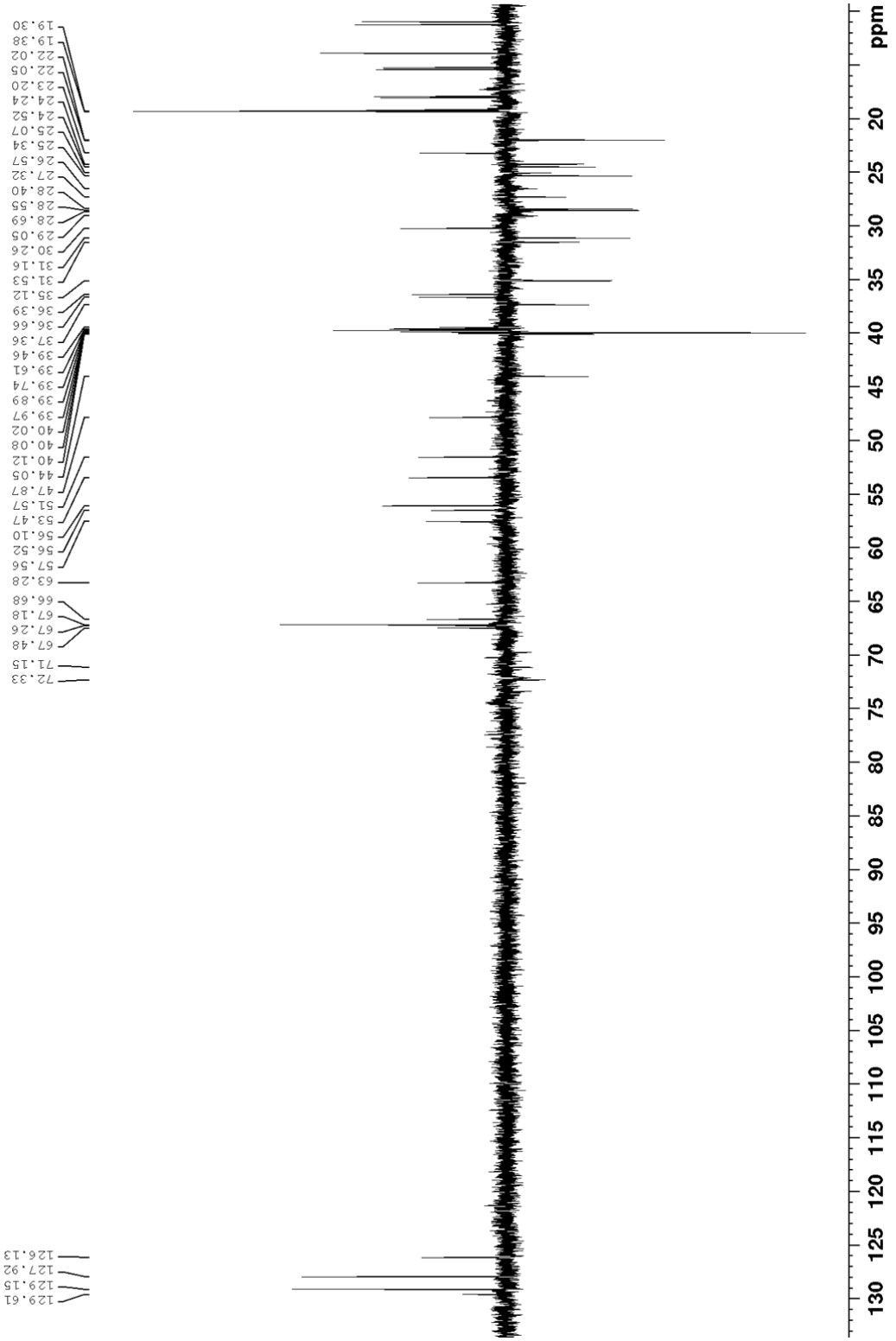


Fig. S7. DEPT-135-NMR spectrum of holrhizin A, recorded at 150 MHz in DMSO-d_6.

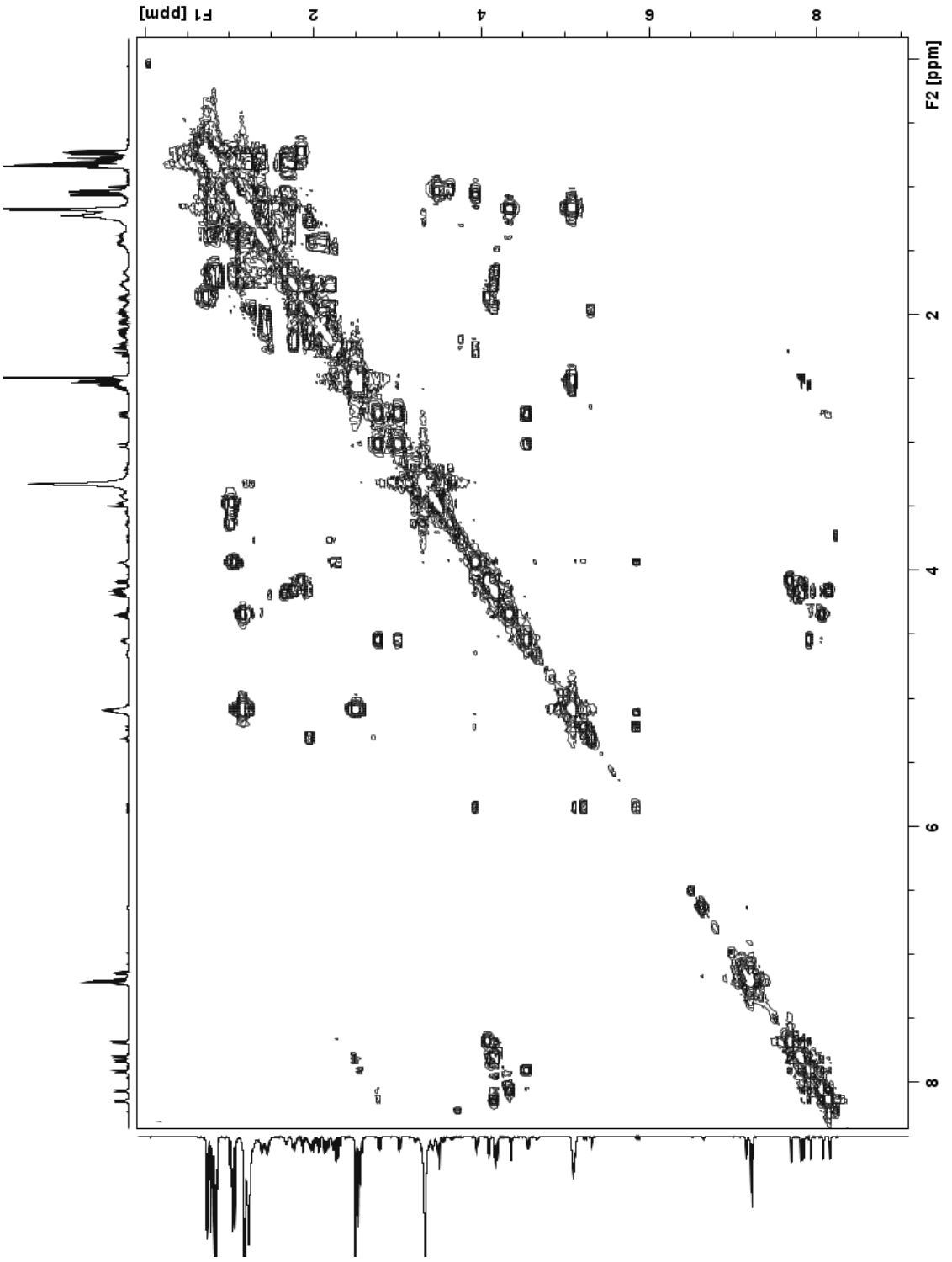


Fig. S8. ^{1}H-^{1}H COSY spectrum of holrhizin A, recorded at 600 MHz in DMSO-d_6.

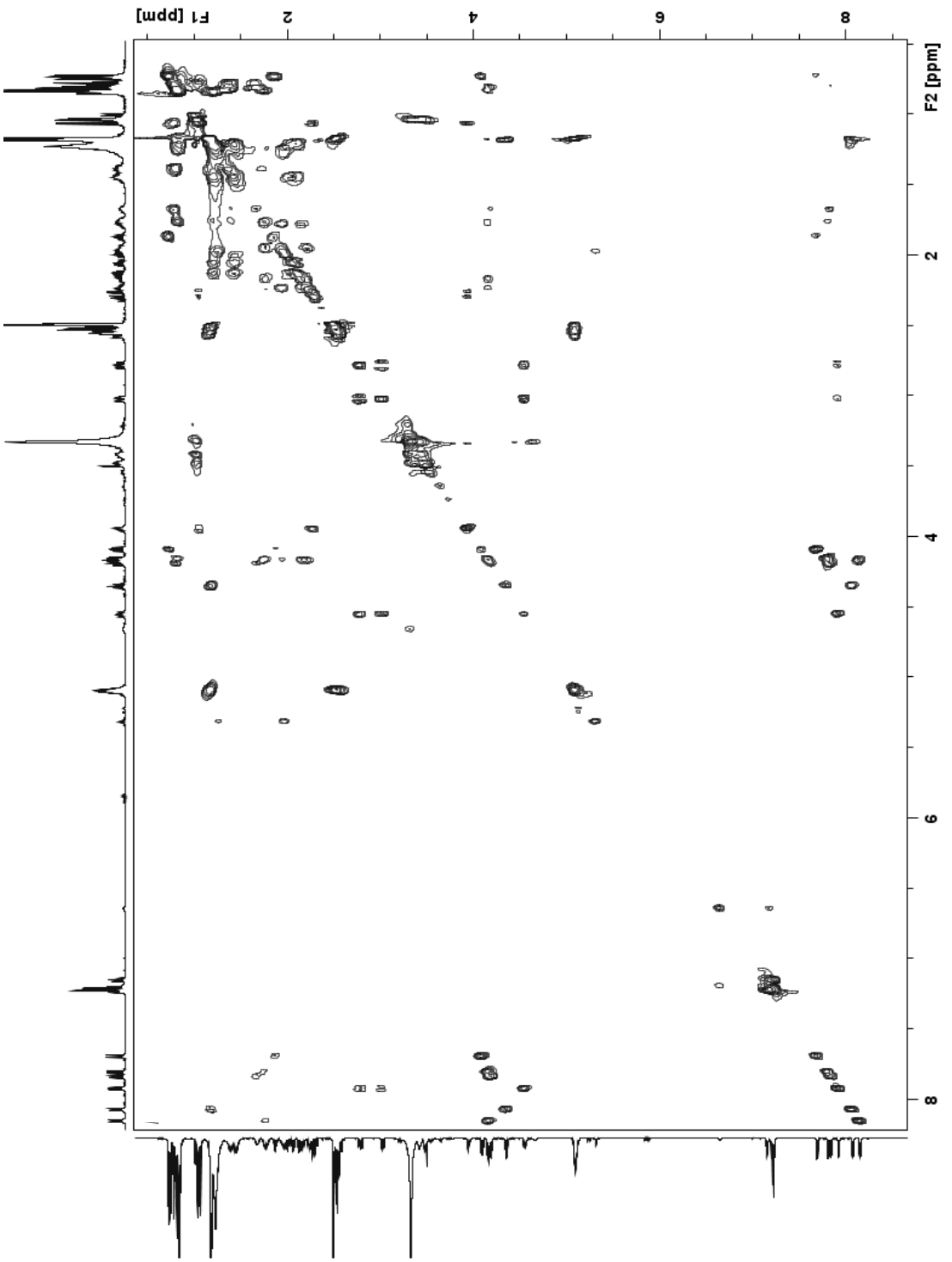


Fig. S9. ^{1}H-^{1}H TOCSY spectrum of holrhizin A, recorded at 600 MHz in DMSO-d_6.

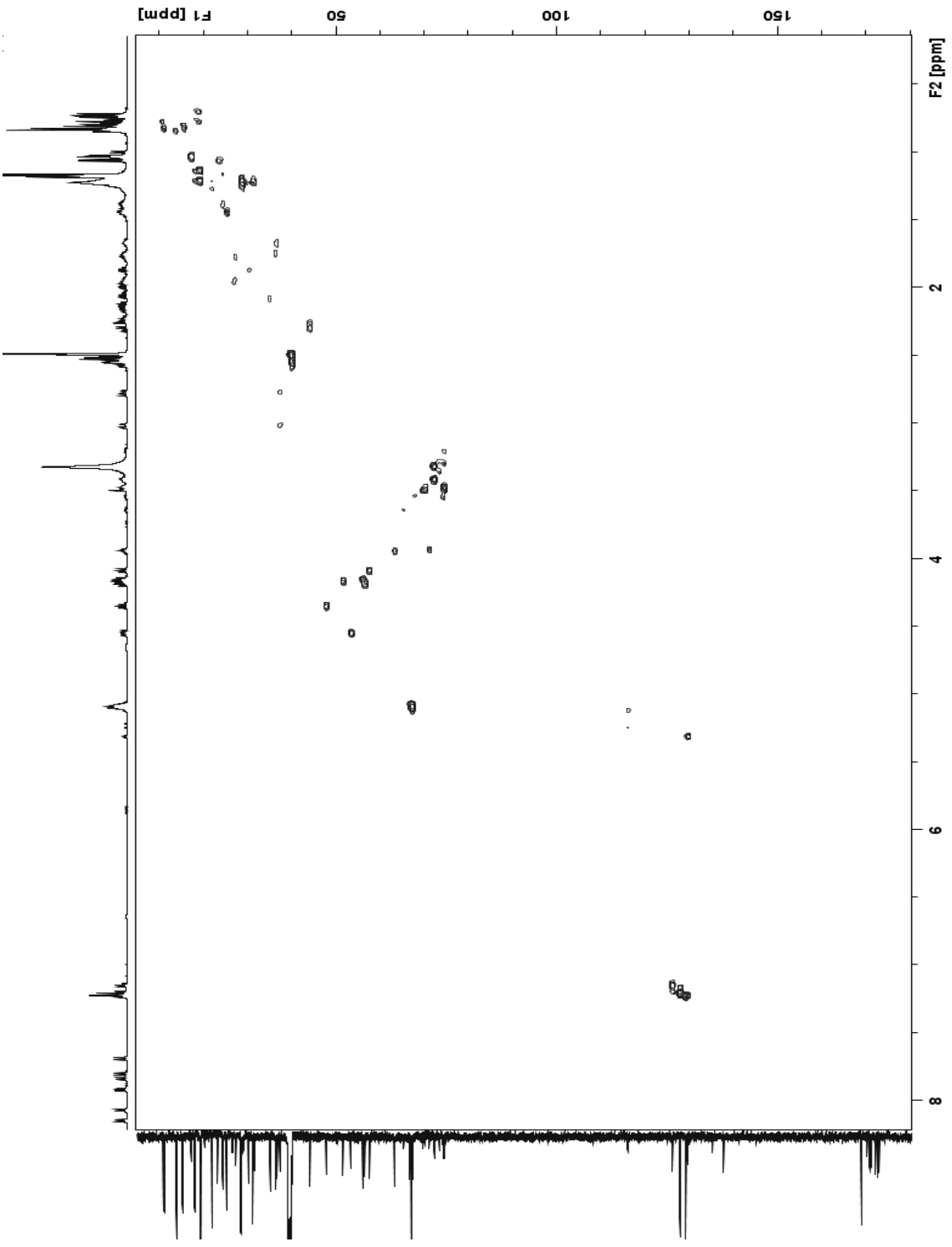


Fig. S10. ^{1}H-^{13}C HSQC spectrum of holrhizin A, recorded at 600 MHz (^{1}H) and 150 MHz (^{13}C) in DMSO-d_6.

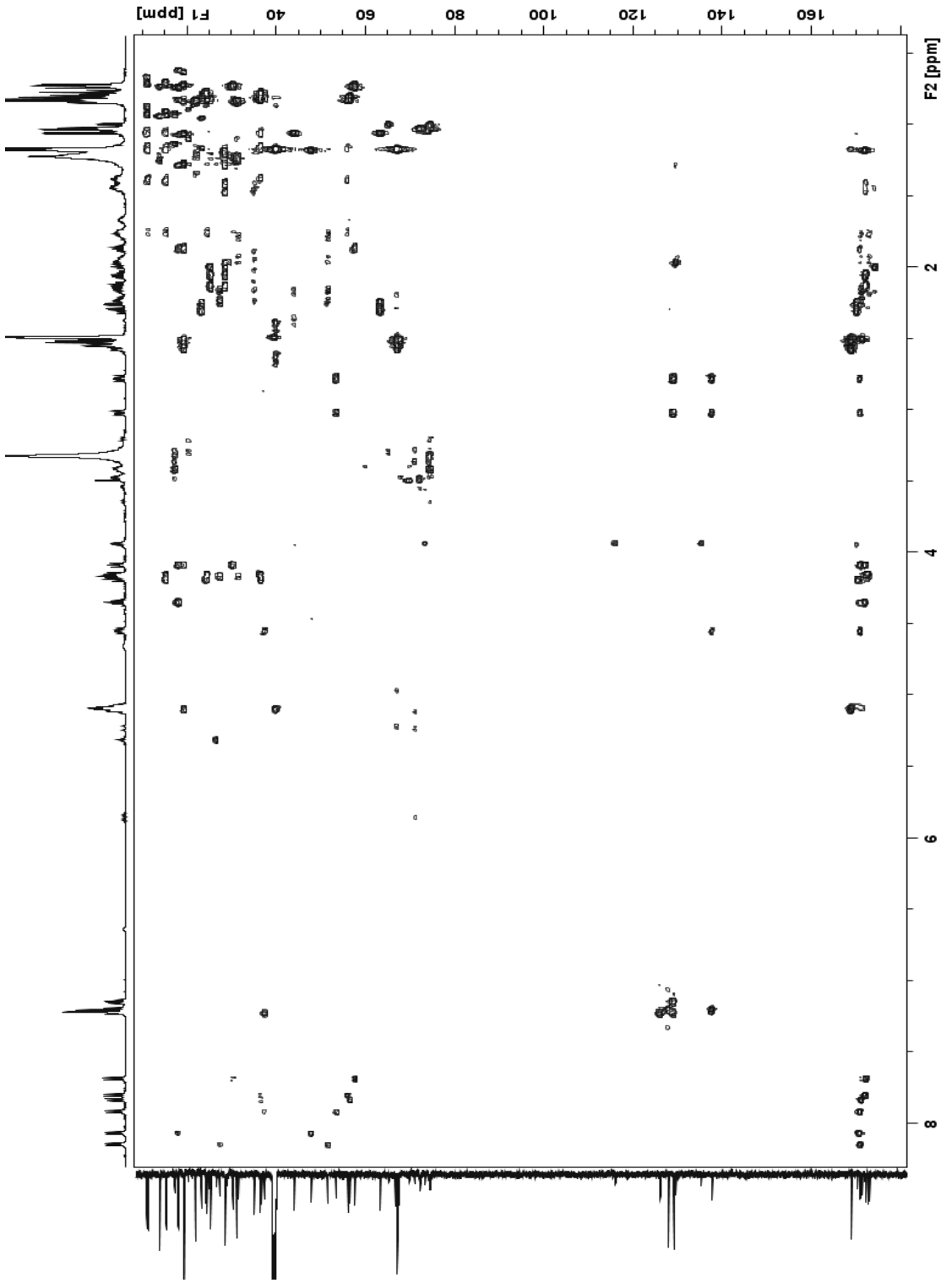


Fig. S11. ^{1}H-^{13}C HMBC spectrum of holrhizin A, recorded at 600 MHz (^{1}H) and 150 MHz (^{13}C) in DMSO-d_6.

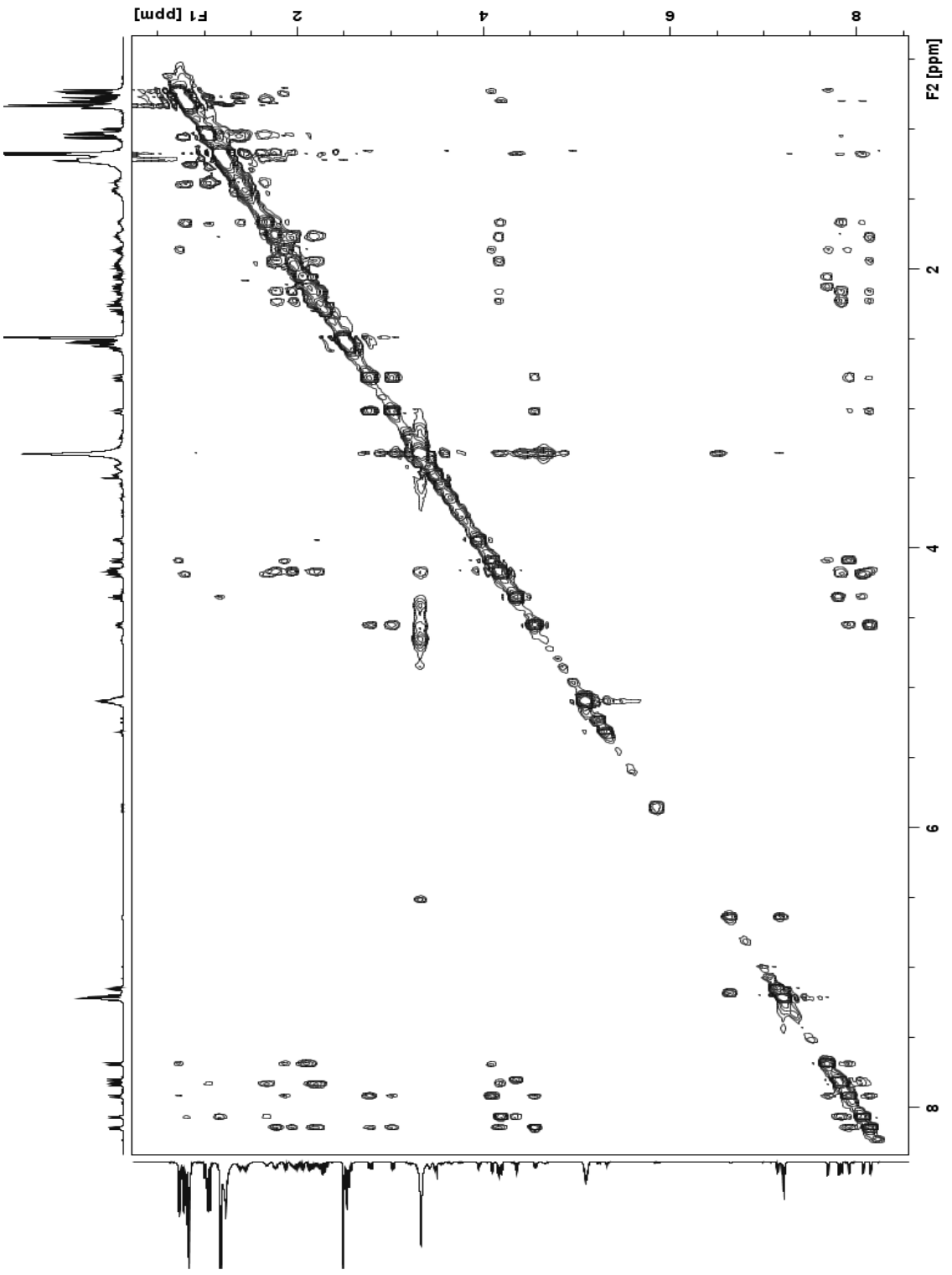


Fig. S12. ^{1}H-^{1}H NOESY spectrum of holrhizin A, recorded at 600 MHz in DMSO-d_6.

Mass Spectra of Holrhizins B, C, and D

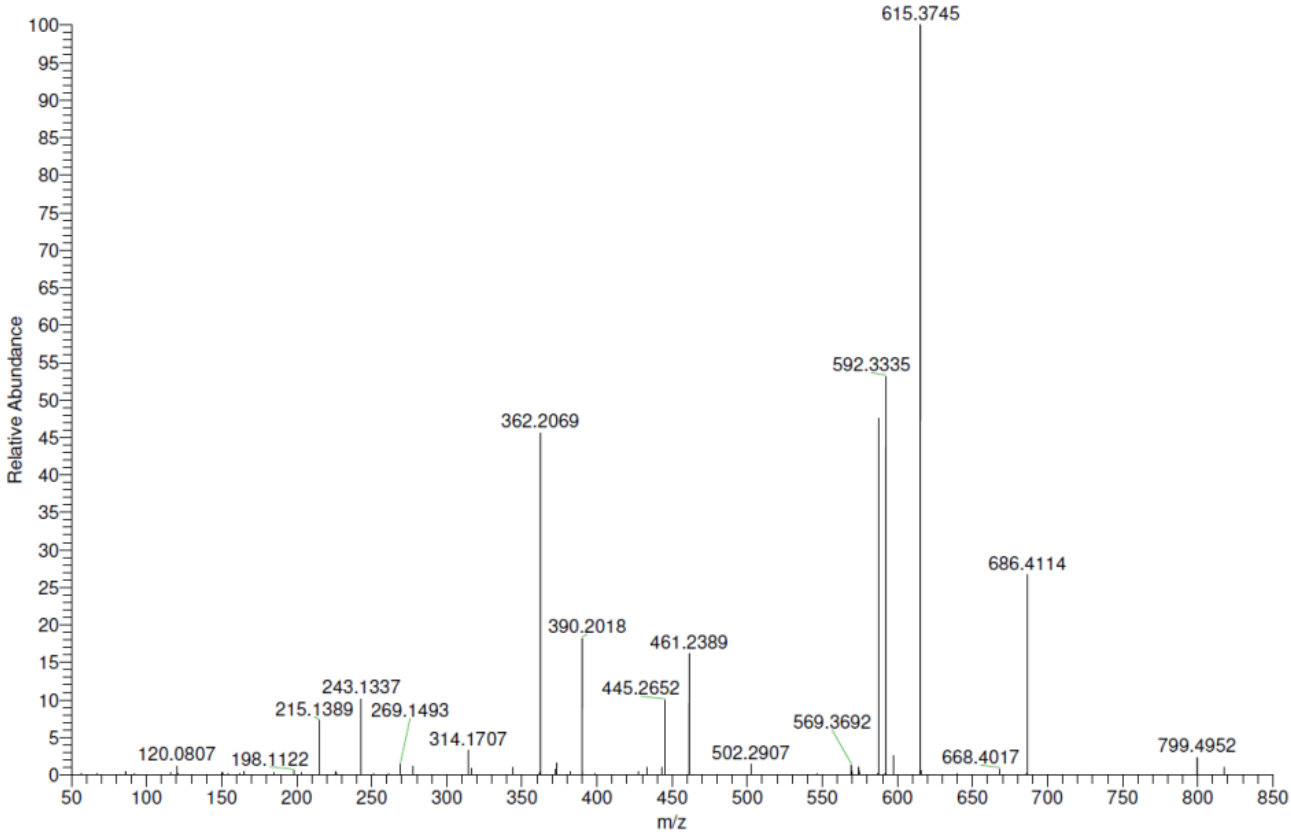


Fig. S13. MS/MS spectra of holrhizin **A** and its congeners **B**-**D** with masses according to measurement (m) and calculation (c); MS fragmentation as [M + H]$^+$; collision energy 12–15%.

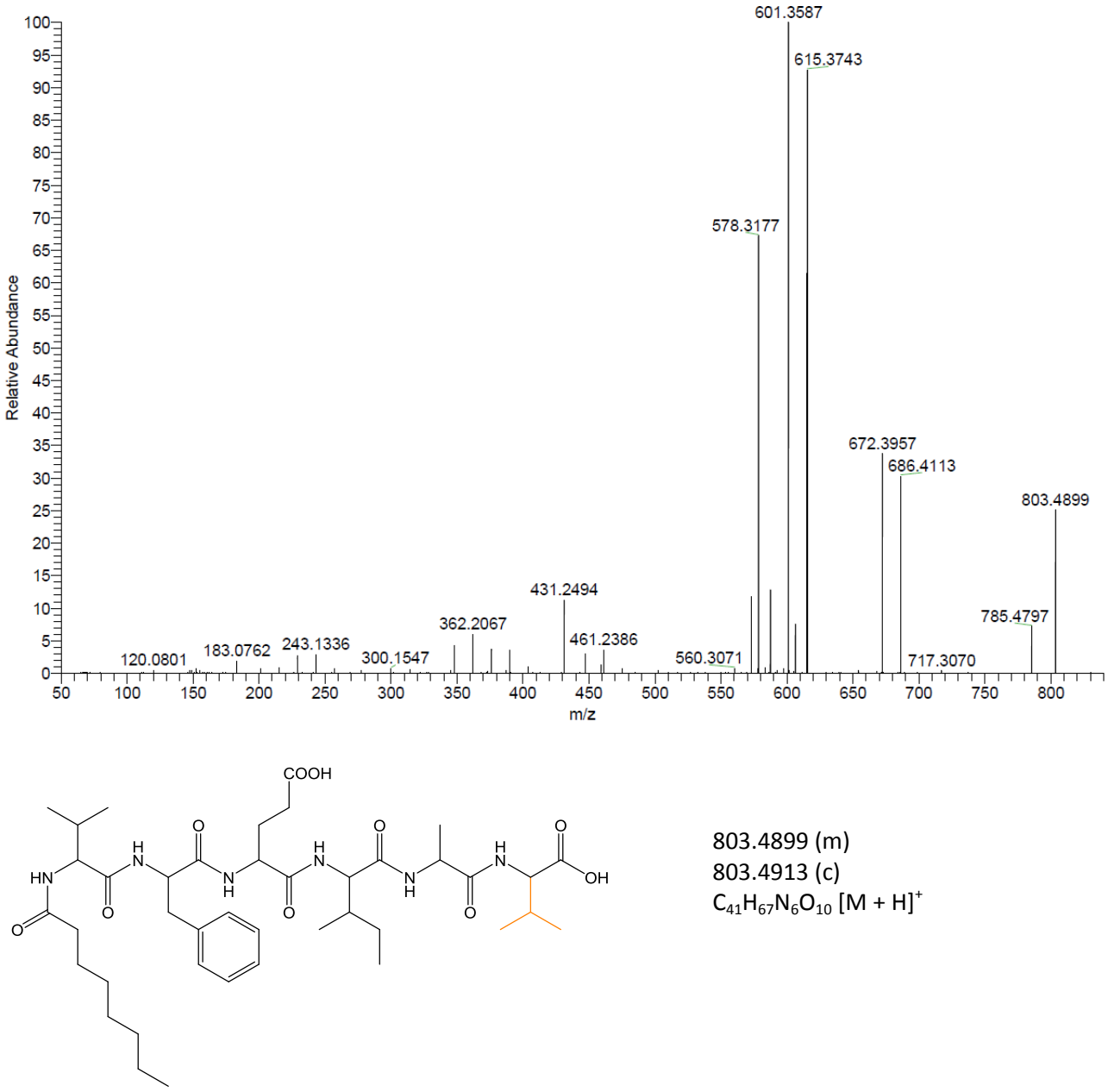


Fig. S13 continued. MS/MS spectra of holrhizin **A** and its congeners **B**-**D** with masses according to measurement (m) and calculation (c); MS fragmentation as $[M + H]^+$; collision energy 12–15%.

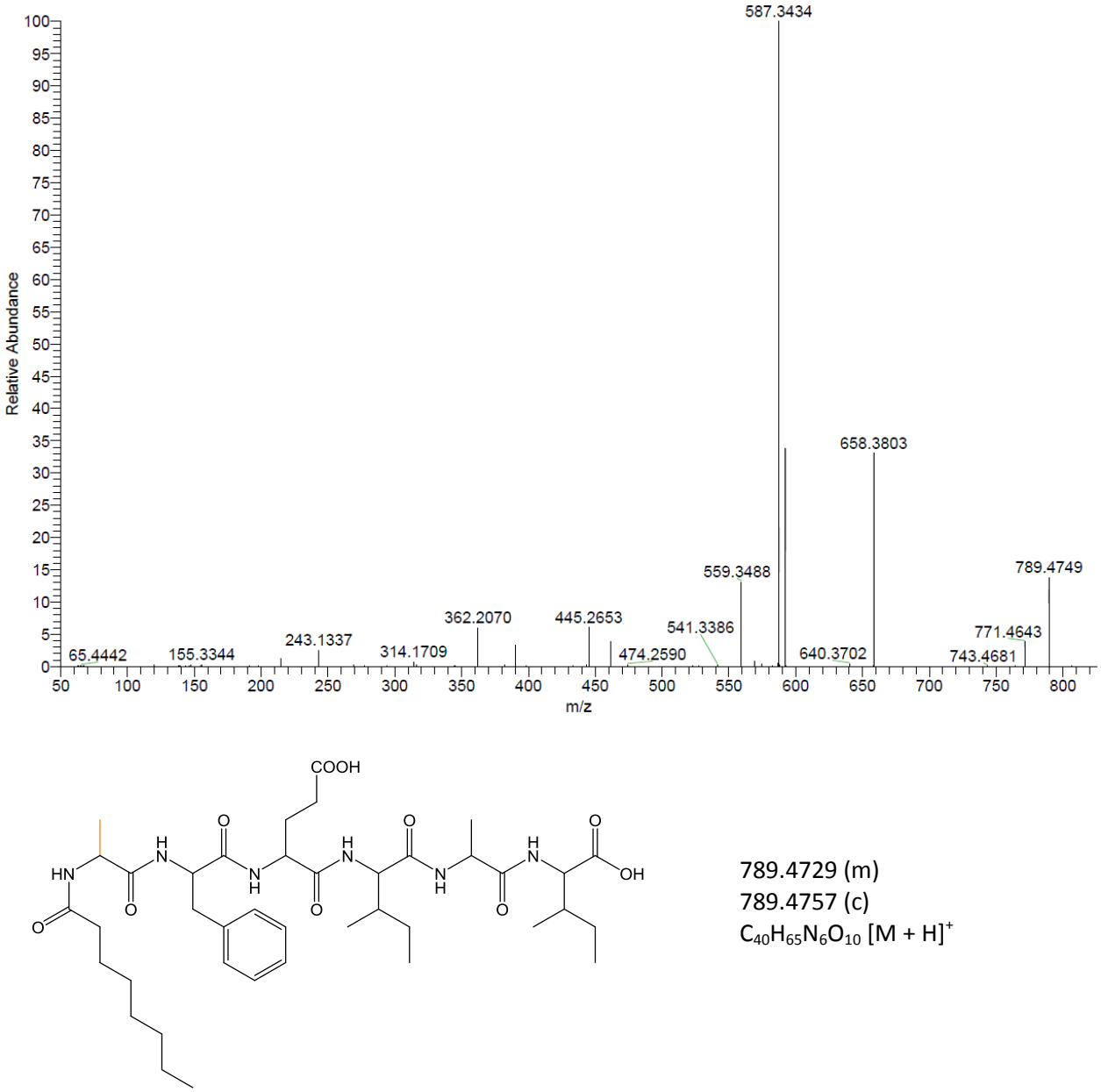
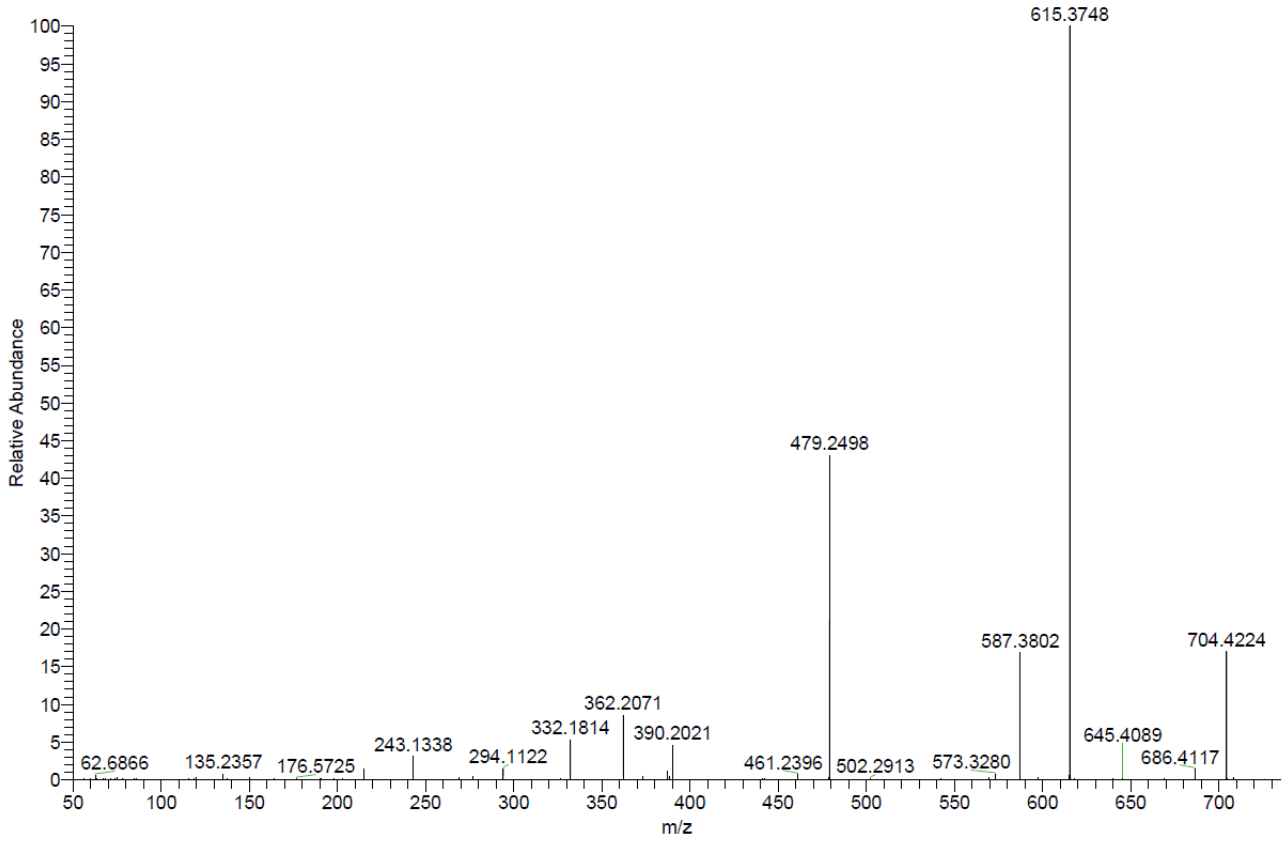
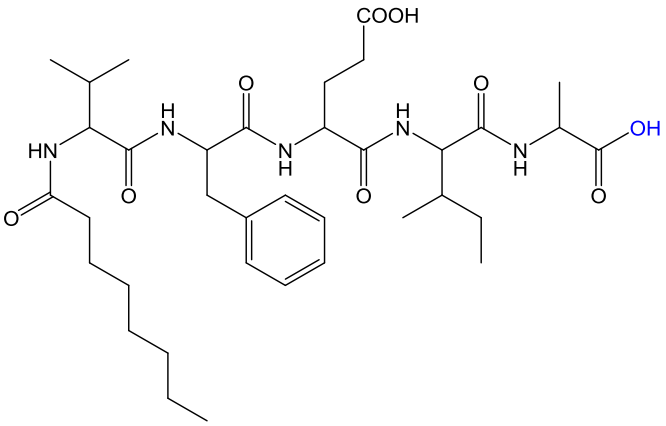


Fig. S13 continued. MS/MS spectra of holrhizin **A** and its congeners **B**-**D** with masses according to measurement (m) and calculation (c); MS fragmentation as $[M + H]^+$; collision energy 12–15%.





704.4224 (m)
704.4229 (c)
$C_{36}H_{58}N_5O_9$ $[M + H]^+$

Fig. S13 continued. MS/MS spectra of holrhizin **A** and its congeners **B**-**D** with masses according to measurement (m) and calculation (c); MS fragmentation as $[M + H]^+$; collision energy 12–15%.

Supplemental references

1 K. Blin, T. Wolf, M. G. Chevrette, X. Lu, C. J. Schwalen, S. A. Kautsar, H. G. Suarez Duran, E. L. C. de los Santos, H. U. Kim, M. Nave, J. S. Dickschat, D. A. Mitchell, E. Shelest, R. Breitling, E. Takano, S. Y. Lee, T. Weber and M. H. Medema, *Nucleic Acids Res.*, 2017, **45**, W36–W41.

2 M. Röttig, M. H. Medema, K. Blin, T. Weber, C. Rausch and O. Kohlbacher, *Nucleic Acids Res.*, 2011, **39**, W362–W367.

3 N. Resource Coordinators, *Nucleic Acids Res.*, 2017, **45**, D12.

4 G. Lackner, N. Möbius, K. Scherlach, L. P. Partida-Martinez, R. Winkler, I. Schmitt and C. Hertweck, *Appl. Environ. Microbiol.*, 2009, **75**, 2982–2986.