Discovery of a lagriamide polyketide by integrated genome mining, isotopic labeling, and untargeted metabolomics

Claire H. Fergusson[a], Julia Saulog[a], Bruno S. Paulo[b], Darryl Wilson[a], Dennis Y. Liu[a], Nicholas J. Morehouse[c], Sam Waterworth[d], John Barkei[d], Christopher A. Gray[c], Jason C. Kwan[d], Alessandra S. Eustaquio[b], and Roger G. Linington*[a]*

[a] Department of Chemistry, Simon Fraser University, Burnaby, BC Canada

[b] Department of Pharmaceutical Sciences and Center for Biomolecular Sciences, University of Illinois at Chicago, Chicago, IL USA

[c] Department of Biological Sciences, University of New Brunswick, Saint John, NB Canada

[d] Division of Pharmaceutical Sciences, University of Wisconsin, Madison, WI USA

*Corresponding author

Supplementary Information

2

substates are shown as blue gradient boxes. The purple diamonds highlight Fisher projection predictions, including KR-based hydroxyl group stereochemistry prediction and amino acid configuration, which we added at the bottom of the chart also in purple. The analysis predicted the second DH domain in LgbC to be a pyran synthase (PS, green). However, this does not match our DH alignments which showed the presence of a DxxxQ motif in all DHs whereas PSs lack this motif (Figure S13). (C) Summary of sequence alignment-based prediction of KR type and TransATor KR predictions.

Discovery of Lagriamide B

Isolation of strain RL17-338-BIF-B

The producing organism, RL17-338-BIF-B, was isolated from rhizome associated with plant roots collected near Soames Hill, Gibsons, British Columbia, at an altitude of 172 m. The strain was originally isolated on BIF medium (3 g KH_2PO_4, 29.2 g $CuSO_4$, 8.5 g $NiSO_4*7H_2O$, 5 g L-threonine, 100 mg fusaric acid, 100 mg bacitracin, 100 mg cycloheximide, 100 mg nystatin, 100 mg econazole) and further isolation was performed on LB medium (Casein Peptone 10 g/L, Yeast extract 5 g/L, sodium chloride 5 g/L). Working stocks were stored in 1:1 glycerol/water in cryomicrocentrifuge tubes at -80°C.

Identification of RL17-338-BIF-B

DNA extraction and sequencing was performed according to an established protocol as previously reported[1]. This strain was previously assigned a NCBI Genbank accession number of MK373475.

RL17-338-BIF-B was grown in LB medium and DNA was extracted from turbid overnight culture using Promega Wizard Genomic DNA Purification Kit according to Gram-negative bacteria protocol. DNA concentrations were measured using Spectramax i3X plate reader (Molecular Devices). Rehydration buffer from DNA purification kit was used as blank in three technical replicates and extracted DNA was used for PCR experiments.

PCR reactions occurred in 50 µL volumes containing 25 µL 2 × MasterMix with dye (ABM), 2.5 µL of 8F and 1492R primers (0.15 µM final concentration), 20 µL of nuclease-free water, and template DNA. Cycling conditions were 35 cycles at 95°C for 1 minute, 52.5°C for 1.5 minutes, 72°C for 1.5 minutes, and final extension at 72°C for 10 minutes. PCR products were confirmed by standard gel electrophoresis (1% agarose in 1 × Tris-acetate-EDTA [TAE] buffer) at 100 V for 30 minutes (Bio-Rad). PCR products were purified using QIAQuick PCR purification kit (Qiagen). DNA concentrations were measured using SpectraMax i3x plate reader. PCR products were sequenced by UBC-NAPS sequencing service (University of British Columbia, Vancouver, BC) with 8F and 1492R primers. Contigs were generated by CAP3. BLASTn searches of 16S rRNA sequence database were completed to determine top BLAST homolog and percent identity.

Genome sequencing of RL17-338-BIF-B and assignment as *Paraburkholderia acidicola*

For genome sequencing, RL17-338-BIF-B was grown in 10 mL BYP medium in a 14-mL round-bottom tube at 30°C and 220 rpm for two days. DNA was isolated using the GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich) according to the manufacturer's protocol. Specifications of the DNA submitted for sequencing were $Abs_{260}/Abs_{280} = 1.90$ and $Abs_{260}/Abs_{230} = 2.00$ and concentration of approximately 100 ng/µL as measured using a Nanodrop (Thermo Scientific NanoDrop TM 1000 Spectrophotometer). Sequencing and assembly was first performed using Illumina technology. All samples were indexed using IDT 10bp UDI indices and sequenced in NextSeq 2000 platform by the SeqCenter (former MiGS, Pittsburgh, PA). The Illumina shotgun library produced around 1.5 million reads in a total of 0.6 Gb data. Reads were quality controlled and adapter-trimmed with bcl2fastq (v. 2.20.0.422). The paired reads from the adapter-trimmed files were then *de novo* assembled using Unicycler (v 0.4.8)[2] and statically recorded with QUAST (v. 5.0.2)[3]. The Illumina-only assembly yielded 84 contigs and 7.71 Mbp. To improve the assembly, we next used the long read sequencing platform Oxford Nanopore and reconstructed the genome by applying Illumina and Nanopore hybrid assembly methodology. The DNA submitted for Nanopore sequencing had following specifications, a DNA concentration of 80 ng/µL measured by Qubit, $Abs_{260}/Abs_{280} = 1.8$ and $Abs_{260}/Abs_{230} = 2.0$ measured by Nanodrop. The DNA sample was processed by the SeqCenter (former MiGS, Pittsburgh, PA) giving a library of 1.06 million reads

spanning 0.9 Gb data. The long reads were quality controlled, and adapter trimmed using porechop (v 0.2.3_seqan2.1.1). Hybrid assembly with the Illumina data was performed with Unicycler (v 0.4.8) using default parameters. Assembly statistics and annotation were recorded with QUAST (v 5.0.2) and Prokka (v 1.14.5)[4] respectively. Hybrid assembly of Illumina and Nanopore reads yielded three large contigs of 4.52 Mbp (#1, likely chromosome 1), 2.99 Mbp (#2, circular, chromosome 2), and 216 kbp (#3), along with eight shorter contigs of 5.23 kbp, 5.18 kbp, 2.90 kbp, 192 bp, 191 bp, 173 bp, 142 bp, and 128 bp. The total genome length was 7.72 Mbp with a G+C content of 61.81%. The whole genome shotgun project of strain RL17-338-BIF-B has been deposited at DDBJ/ENA/GenBank under the accession code JAOALG000000000, and BioProject ID number PRJNA875462. The version described in this manuscript is JAOLG010000000 and can be readily found on NCBI website.

During the genome sequence submission, the taxonomy identification level was reassigned by GenBank from *P. megapolitana* (previously predicted based on 16S rRNA identity) to *Paraburkholderia acidicola* because it was found that RL17-338-BIF-B has an average nucleotide identity (ANI) of 97.164% to the genome of *Paraburkholderia acidicola* with 86.7% genome coverage. We confirmed this assignment by comparing the genome of RL17-338-BIF-B with a deposited *Paraburkholderia megapolitana* genome (BioProject ID: PRJNA554389) and a *Paraburkholderia acidicola* genome (PRJNA361317). Using the ANI algorithm available at https://www.ezbiocloud.net/tools/ani[5] we detected a genome identity of 92.17% and for *P. megapolitana* and 97.18% identity with *P. acidicola*. Since, the *P. acidicola* genome exhibited higher identity and coverage to the RL17-338-BIF-B genome, we accepted the reassignment suggested by the NCBI curation system.

For reference, *Paraburkholderia acidicola* was recently reclassified from what was in the past referred to as *Pseudomonas acidophila.* This reclassification was based on a phylogenomic tree construction of concatenated sequences of 118 conserved proteins, and through conserved signature indels[6].

Genome sequencing of ATCC 31363

The BGC was discovered during a routine inspection of the related lagriamide BGC. It was noted that all genes within the cluster were most closely related to genes within the *Pseudomonas acidophila* ATCC 31363 strain. Closer inspection revealed the presence of homologs for all *lga* genes across several contigs. Attempts were made to improve the genome assembly using the publicly available reads (SRR5380803): Briefly, reads were trimmed with Trimmomatic (v0.36) against the TruSeq3-PE library with 3 leading and trailing nucleotides (nt), a sliding window of 4:15 and minimum length of 36 nt. Trimmed reads were assembled with SPAdes (v3.12). Reassembly did not improve the fragmentation of the BGC region, so attempts were made to manually stitch the BGC with paired-end mapping. Briefly, trimmed reads were mapped back to the assembled contigs using bbmap. Connectivity between contigs was inferred using Cytoscapeviz with default parameters. Although there was not enough overlap to manually stitch the contigs together, the results suggested that all contigs with *lga*-like genes were contiguous.

Following this, the *P. acidophila* ATCC 31363 strain was ordered. The strain arrived as a desiccated pellet and rehydrated with 1 mL nutrient broth. The strain was streaked on Nutrient agar and incubated overnight at 28°C. A single colony was picked from the plate and inoculated into 5 mL sterile nutrient broth and incubated overnight at 28°C. A gram stain and follow up discontinuous streak were performed to confirm culture purity. A volume of 1.5 mL of the overnight culture was centrifuged at 5000 rpm for 3 minutes. The supernatant was removed and 500 µL of Buffer A (0.315 g Tris, 0.117 g NaCl, 0.08 g EDTA, 10 mL MilliQ water, filter sterilized), 210 µL of 20% SDS (2 g in 10 mL MilliQ water, filter sterilized), 1 µL RNAase A, 20 µL protease and 10 µL lysozyme (100 mg/mL) was added. The pellet was gently resuspended via pipetting with wide-bore tips. The mix

was incubated at 37°C for 1 hour on a shaker set to the lowest setting. A volume of 500 µL of phenol/chloroform/isoamyalcohol (25:24:1) was added, mixed via inversion then centrifuged at 7200g for 3 mins at 4°C. The pellet of DNA was then moved from the aqueous layer to a new Eppendorf tube. A 10% (v/v) volume of 3M sodium acetate and an equal volume of isopropanol was added and mixed by gentle inversion. The sample tube was then placed in a -80°C freezer for 10 minutes. The DNA was then transferred to a new Eppendorf tube and 500 µL 100% ethanol was added to wash the DNA. The ethanol was removed via pipetting and replaced with fresh ethanol, which was subsequently removed via pipetting. Excess ethanol was evaporated on a 50°C heating block for 10 minutes. The DNA was resuspended in 100 µL sterile water. The DNA was quantified with a Nanodrop, with a resultant concentration of 1539.4 ng/µL with an 260/280 of 1.79 and a 260/230 of 1.79. The DNA was then sent to the University of Wisconsin Biotechnology Center DNA Sequencing Facility for Nanopore MinION sequencing.

Nanopore MinION long reads were assembled along with the existing trimmed Illumina short reads using the hybrid assembly option in SPAdes (v3.12). This resulted in two closed chromosomes of approximately 4.3 Mbp and 2.8 Mbp in size. The lagriamide-like BGC was identified on the larger of the two chromosomes. Direct comparison with the lagriamide BGC revealed that the synteny of the BGC relative to the lga BGC was largely identical except for two extra DH domains on LgaB and G respectively (identified previously to be non-functional), an absent KR domain on LgaD, and the P450 gene (lgaJ) appeared to have been substituted with a gene encoding a 4'PP transferase.

Identification and in silico analyses of the lagriamide B gene cluster

Analysis of the obtained RL17-338-BIF-B genome using antiSMASH 6.0[7] followed by manual curation revealed 18 biosynthetic gene clusters spread across the two largest contigs (#1, #2) (**Table S1**), Region 1.2 included a hybrid PKS-NRPS BGC with sequence and synteny similarity to the previously reported lagriamide (*lga*) gene cluster (MiBIG accession code: BGC0001646) with 70% of nucleotide pairwise identity (**Figure 1**, main text)[8]. PKS-NRPS domains were compared via multi-sequence alignments using standard options on Geneious Prime (v 2022.2.2) accessory genes were further analyzed using BLAST.

Table S1: Biosynthetic gene clusters identified by antiSMASH and manual curation in the genome of *P. acidicola* RL17-338-BIF-B.

BGC	Region	Type	From	To	Similarity
1	1.1	RiPP-like	41,163	52,002	
2	1.2a	Terpene	586,591	604,487	
3	**1.2b**	***Trans*-AT-PKS, NRPS**	**615,124**	**705,074**	**Lagriamide**
4	1.3	NRPS	1,182,301	1,252,547	
5	1.4	Terpene	2,646,363	2,667,430	
6	2.1	Terpene	4,143	35,687	
7	2.2	Redox-cofactor	116,404	138,591	
8	2.3	β-lactone	304,818	336,297	
9	2.4	RiPP-like	571,011	581,362	
10	2.5a	NRPS	756,284	789,295	Sulfazecin
11	2.5b	Arylpolyene, Ectoine	798,962	809,063	Bulgecin
12	2.5c	β-lactam	809,130	828,759	
13	2.6	NRPS	1,271,776	1,325,050	Haereomegapolitanin
14	2.7	T1PKS	1,834,314	1,879,027	
15	2.8	RiPP-like	2,043,724	2,054,647	
16	2.9	Phosphonate	2,164,890	2,206,593	
17	2.10	Homoserinelactone	2,331,154	2,351,747	
18	2.11	T1PKS-NRPS-HglE-KS	2,389,077	2,468,130	

Table S2: BLASTp results (as of Mar. 2024) of encoded proteins in the *lgb* BGC.

Gene name	Protein size (aa)	Top BLAST hit	Species, Identity, Accession Code
lgbA	3591	hybrid *trans*-AT PKS-NRPS LgaA	*Burkholderia gladioli*, 74.57%, AXA20090.1
lgbB	6520	hybrid *trans*-AT PKS-NRPS	*Paraburkholderia acidicola*, 96.09%, PCE28606.1
lgbC	6959	*trans*-AT PKS	*Paraburkholderia acidicola*, 95.32%, WP_096716360.1
lgbD	888	*trans*-AT PKS	*Paraburkholderia acidicola*, 98.65%, WP_170043813.1
lgbE	379	acyl-carrier-protein S-malonyltransferase	*Paraburkholderia acidicola*, 99.21%, PCE28609.1
lgbF	461	enoylreductase (ER)	*Paraburkholderia acidicola*, 99.33%, WP_207556501.1
lgbG	9226	*trans*-AT PKS	*Paraburkholderia acidicola*, 97.05%, WP_096716363.1
lgbL	131	nuclear transport factor 2 family protein	*Paraburkholderia acidicola*, 100%, WP_096716364.1
lgbH	452	mate efflux transporter	*Paraburkholderia acidicola*, 98.67%, WP_096716365.1
lgbI	338	acyltransferase domain-containing protein	*Paraburkholderia acidicola*, 99.41%, WP_096716366.1
lgbM	247	hypothetical protein	*Paraburkholderia acidicola*, 99.60%, WP_096716367.1
lgbN	258	4'-phosphopantetheinyl transferase	*Paraburkholderia acidicola*, 99.61%, WP_096716368.1
lgbK	271	ketoreductase domain-containing protein	*Paraburkholderia acidicola*, 100%, WP_096716369.1

Table S3: Biosynthetic gene cluster labeling prediction for *P. acidicola* RL17-338-BIF-B.

Cluster	Type	A	P	M	G
1.1	RiPP-like				
1.2 a	Terpene				
1.2 b	***trans*AT-PKS, NRPS**	**15-16**	**0**	**5**	**2**
1.3	NRPS	>1	0-1	0	8
1.4	Terpene				
2.1	Terpene				
2.2	redox-cofactor				
2.3	betalactone				
2.4	RiPP-like				
2.5a	NRPS-like, NRPS			1	3
2.5b	bulgecin (not predicted by antiSMASH)				
2.5c	arylpolyene				
2.6	NRPS	4			5
2.7	T1-PKS	>1			1-2
2.8	RiPP-like				
2.9	Phosphonate				
2.10	Hserlactone				
2.11	NRPS, T1-PKS, hglE-KS	>4			5

Parallel well-plate fermentation and SIL incorporation

Burkholderia Yeast Peptone (BYP, 10 g/L starch, 4 g/L yeast extract, 2 g/L peptone, 3 g/L NaCl per litre of water) medium was used for all SIL fermentation experiments and follow protocol by McCaughey et al[9] with some adaptations. The SIL feedstock compounds (Sigma-Aldrich Canada), [99% 1-13C]acetate, [99% 1-^{13}C]propionate, [99% methyl-^{13}C]methionine and [98% 1-^{15}N]glutamate, and the unlabeled version of each compound (Sigma-Aldrich Canada) were prepared as stock solutions in Milli-Q water and sterilized by filtration using 0.2 µm polyethersulfone filter (Canadian Life Sciences). Bacterial inoculum of *P. acidicola* RL17-338-BIF-B was prepared by streaking frozen stock onto LB agar plate (Casein Peptone 10 g/L, Yeast extract 5 g/L, sodium chloride 5 g/L, 15 g/L agar). Single colonies were selected to inoculate a 10 mL liquid culture of BYP media. After 24 hours of growth, 20 µL of inoculum was added to the inner 16 wells of the 24-well microtiter plates. A filtered SIL compound or corresponding unlabeled compound was then added to each of the inner 16 wells. The final concentrations of each of these wells were 30 mM [1-^{13}C]acetate, 30 mM [1-^{13}C]propionate, 5 mM [methyl-^{13}C]methionine, and 10 mM [1-^{15}N]glutamate. 20-200 µL of stock solution was added to each well to reach desired concentration of each SIL or unlabeled precursor in each well. The same culture was used for all each of the four replicate wells of each feedstock condition. The outer eight wells that contained 2 mL of BYP media were included as controls to account for metabolic changes that may occur due to adding the precursor compound.

Microtiter plates contain SIL supplemented bacterial cultures were shaken at 200 rpm and maintained at 27.0 °C for five days. After 5 days, the cultures were extracted by adding 2 mL of Optima methanol (Thermo Fisher Scientific). 1 mL from each well was transferred to 1.5 mL microcentrifuge tubes (Eppendorf Canada Ltd), sonicated for 5 minutes, and centrifuged for 1 minute at 16,000g. The MeOH:H$_2$O supernatant was injected directly onto the UPLC-ESI-qTOF MS system (Waters Corporation Synapt G2Si qTOF) or diluted with an equal volume of methanol to maintain the most intense signals in the chromatogram in an optimal range for sensitivity and mass accuracy.

MS data acquisition for IsoAnalyst

The supernatants were subjected to chromatographic separation and MS analysis. Chromatography was performed on a Waters I-Class Acquity UPLC system (Acquity HSS T3 1.8 µm, 2.1 × 100 mm) using a linear gradient (solvent A: H$_2$O + 0.01% FA; solvent B: ACN + 0.01% FA) of 5-98% B over 5.8 minutes, a hold at 98% B for 0.3 minutes, followed by a 1.8-minute re-equilibration at 5% B. All mass spectra were acquired using a Waters Synapt G2Si qTOF running MassLynx v.4.1 SCN941. MS data were acquired in data-independent acquisition mode (MS^E). The MS detector range was set to 50-1500 *m/z* in positive mode, with a capillary voltage of 3.5 kV, and a desolvation temperature of 200 °C. IsoAnalyst calculates isotopologue ratios using a scan-by-scan comparison of centroided isotopologue peaks from the MS^1 data. To obtain the required scan rate for IsoAnalyst we used a 0.2 -s scan rate with alternating MS^1 and MS^2 scans. As detailed by McCaughey et al[9], we used a similar MS^1 scan rate for data processed in IsoAnalyst to ensure adequate MS^1 coverage across chromatographic peaks for each analyte.

IsoAnalyst pre-processing

Raw data files from the UPLC-MS were processed using a customized workflow developed in collaboration with Waters Corporation. The initial peak detection and MS^E deconvolution software package MSeXpress 2.0 was employed to generate peak lists of precursor ions with their associated product ions. Func001 files of MSeXpress 2.0 in .csv tabular inputs with "FunctionScanIndex", "RT", "MZ", and "Intensity" are used in IsoAnalyst. Feature lists are also generated by MSeXpress 2.0 as CPPIS files. Lastly, an input specification file is required. These file types are explained in detail in the README.md file of IsoAnalyst in the GitHub repository (https://github.com/liningtonlab/isoanalyst).

Using IsoAnalyst protocol and MZmine to visualize isotope labelling data

The IsoAnalyst protocol as described above was repeated on *P. acidicola* ATCC 31363 with [1-^{15}N]glutamate labeled and unlabeled conditions to help identify lagriamide B in our in-house strain *P. acidicola* RL17-338-BIF-B. From this, summary report files were generated using the IsoAnalyst software package, providing a list of ^{15}N labeled mass features for each strain.

To visualize the unlabeled versus labeled mass features of both strains, MS data of the unlabeled glutamate supplemented bacterial cultures of *P. acidicola* RL17-338-BIF-B and *P. acidicola* ATCC 31363 were processed using MZmine version 2.53 (http://mzmine.github.io/), yielding the traces in **Figure 2A**, main text. Peak detection was done using mass detection and the ADAP chromatogram builder module using MS level 1, an *m/z* tolerance of 0.02 *m/z* or 10.0 ppm, and a minimum group size of four scans. For both strains, we used a mass detection (noise level) of 1×10^3, a group intensity threshold of 5×10^3, and a minimum highest intensity of 1×10^3. Deconvolution was then done using a baseline cut-off algorithm with a minimum peak height of 5×10^2, peak duration range of 0–10, and baseline level of 2×10^2. Total ion current (TIC) plots were generated from the deconvoluted feature lists: for each strain, we generated a TIC plot in which all extracted mass features were selected and a second TIC plot in which only ^{15}N labeled features (as dictated from the IsoAnalyst results file) in the mass range 600–800 Da were manually selected.

Cultivation of *P. acidicola* RL17-338-BIF-B

P. acidicola RL17-338-BIF-B was cultivated on a rotary shaker (200 rpm, 27.0 °C) in 2.8 L Fernbach flasks containing 1 L of BYP medium (10 g/L starch, 4 g/L yeast extract, 2 g/L peptone, 3 g/L NaCl) with 20.0 g of Amberlite XAD-16 adsorbent resin per liter.

Extraction of *P. acidicola* RL17-338-BIF-B

Cells and resin were removed by vacuum filtration using Whatman glass microfiber filters and washed with deionized water. Each liter of resin and cells were extracted with 250 mL of 1:1 MeOH:DCM and the organic extract was removed from the cells by vacuum filtration.

The crude organic extract was subjected to solid-phase extraction using a Teledyne C$_{18}$ cartridge (10 g) and eluted using a step gradient of MeOH/H$_2$O solvent mixtures (10% MeOH, 20% MeOH, 30% MeOH, 40% MeOH, 60% MeOH, 80% MeOH, 100% MeOH) and finally with ethyl acetate to afford seven fractions via CombiFlash. The 10% MeOH fraction was discarded, and remaining six were dried in vacuo using SpeedVac evaporator.

Purification of lagriamide B from *P. acidicola* RL17-338-BIF-B

Analysis by reverse phase (RP) liquid-chromatography mass spectrometry (LCMS) revealed that the lagriamides were present in the 80% and 100% MeOH fractions, with stronger UV signals present for lagriamide B in the 100% MeOH fraction when visualized by extract ion chromatograms. HPLC separations were performed on an Agilent 1200 series HPLC equipped with a binary pump and a diode array detector. Reverse phase (RP) high-performance liquid chromatography (HPLC) was used to isolate lagriamide B using an isocratic separation (34% MeCN + 0.02% formic acid, 66% H$_2$O + 0.02% formic acid, 2.0 mL/min). Lagriamide B was isolated by threshold intensity (Phenomenex Synergi Fusion-RP, 250 mm × 4.6 mm, 10 μm).
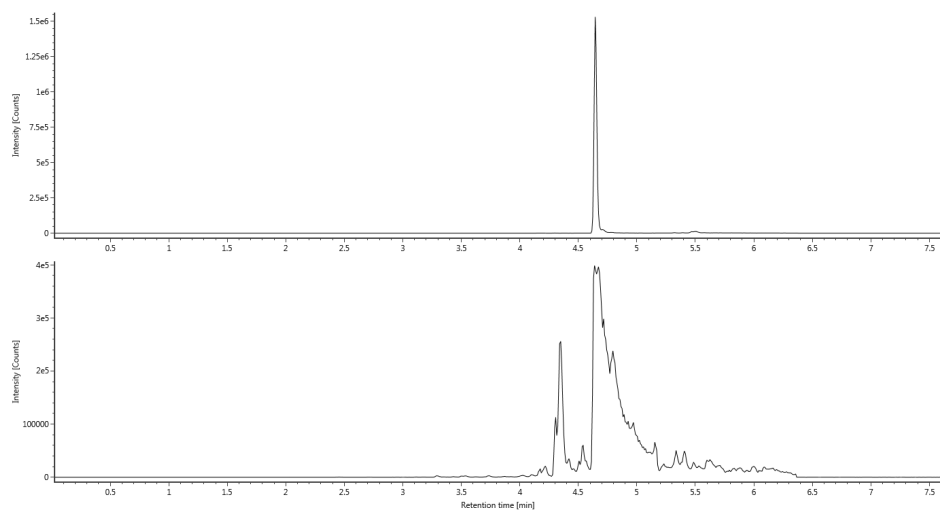
Figure S1: Extract ion chromatogram of lagriamide B of isolated sample (top) and in *P. acidicola* extract (bottom).
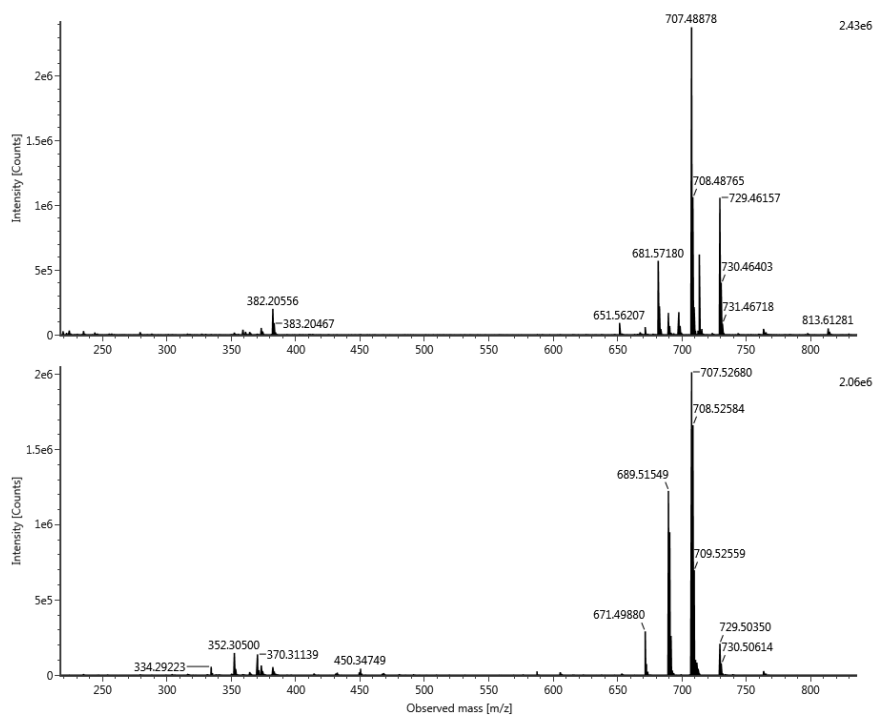
Figure S2: Mass spectra of lagriamide B of isolated sample (top) and in *P. acidicola* extract (bottom).

Chemical Characterization of Lagriamide B

NMR spectra were acquired on a Bruker AVANCE III 600 MHz spectrometer, with a 5 mm TCI cryoprobe, and referenced to residual solvent proton and carbon signals. UV-vis spectrum was obtained on an Agilent Cary 300 spectrophotometer (Agilent Technologies). Optical rotations were measured on a Model 341 Polarimeter (PerkinElmer). HR-ESI-MS and MS^2 fragmentation spectra were recorded on a SYNAPT UPLC-ESI-qTOF (Waters Corporation).

Lagriamide B (**2**) $[\alpha]^D_{25}$= +42.4 (c 0.05, MeOH); UV (MeOH) λmax (log ε) = 217 (5.07) nm. See Table S4 for NMR shifts. HR-ESI-MS *m/z* 707.4835 [M+H]$^+$ (calculated for $C_{39}H_{66}N_2O_9$ *m/z* 707.4841 [M+H]$^+$).
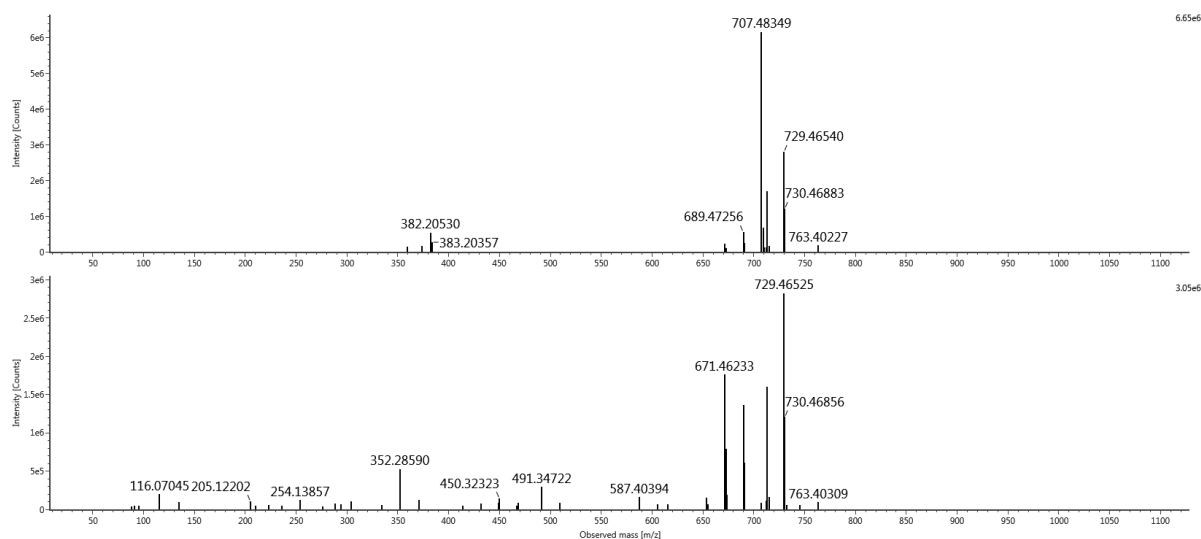
Figure S3: Lagriamide B mass spectra (top) and MS^2 (bottom).
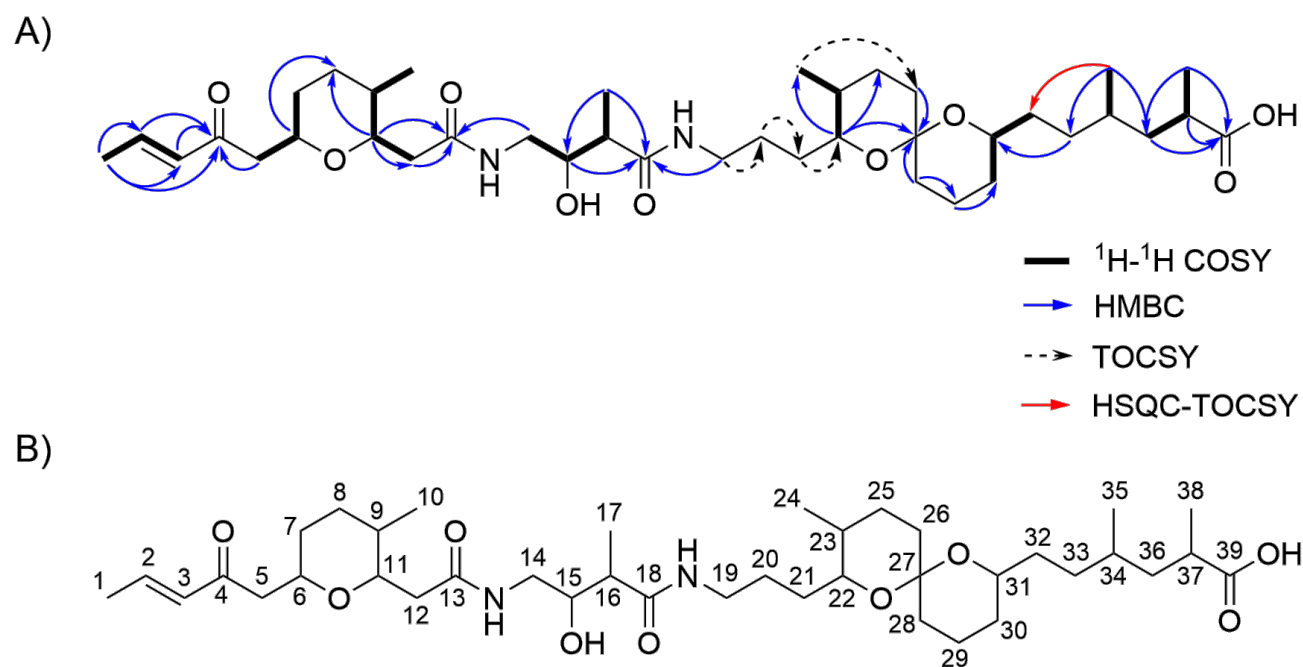
Figure S4: A) Selected 2D NMR correlations and B) Planar structure for lagriamide B.

Planar Structure Elucidation of Lagriamide B

The molecular formula of lagriamide B was established as $C_{39}H_{66}N_2O_9$ based on the HR-ESI-MS ([M+H]$^+$ *m/z* 707.4841, **Figure S3**), indicating 8 degrees of unsaturation. In contrast, the molecular formula of lagriamide A is $C_{41}H_{68}N_2O_{10}$ ([M+H]$^+$ *m/z* 749.4949), with 9 degrees of unsaturation.[8]

Analyses of 1D and 2D NMR spectra (**Figure S16–Figure S23**) showed strong alignment with the count of atoms in the molecular formula for lagriamide B. The ^{1}H and phase-sensitive HSQC spectra revealed 2 × amide NH, 1 × primary CH_3, 5 × secondary CH_3, and 12 × CH, including two olefinic signals. There are also 16 × CH_2 signals, with pairs of diastereotopic methylene protons deciphered from the HSQC. The ^{13}C and HMBC spectra further revealed 5 × qC signals, including one carboxylic acid, one ketone, and two amide carbonyls. These signals accounted for all carbon atoms, 65 of 66 protons, both nitrogen atoms, 5 of 9 oxygen atoms, and 5 of 8 degrees of unsaturation. The final proton and one oxygen atom were assigned together as a hydroxy group based on five HSQC signals consistent with oxygenated methines (δ_H 3.06–3.49 ppm, δ_C 66.1–73.8 ppm). Further, two of these methine protons had reciprocal HMBC correlations to their carbon atoms, which suggested they were adjacent to the same oxygen atom, thus accounting for the final atoms in the molecular formula.

This analysis identified differences in moieties between lagriamide B and A. Evidently, lagriamide B lacks signals consistent with an epoxy-ketone as found in lagriamide A (δ_H 3.10–3.34 ppm, δ_C 53.7–59.7 ppm).[8] Moreover, while lagriamide A has only one olefinic proton (δ_H 4.87 ppm), there are two signals in lagriamide B which are more deshielded, at 6.83 (H2) and 6.10 (H3) ppm.

To solve the full planar structure of lagriamide B and further elucidate the structural variations between the two lagriamides, we identified key signals in the ^{1}H-^{1}H COSY, HMBC, TOCSY, and HSQC-TOCSY spectra, summarized in **Figure S4A**. COSY correlations from both olefinic protons to a primary methyl at 1.85 (H1) ppm established a disubstituted trans-alkene (J_{H2-H3} = 15.8 Hz). Further, HMBC signals from H1, H2, and H3 to 198.0 (C4) linked these as an α, β-unsaturated ketone forming one terminus of the molecule. An HMBC correlation between C4 and one of the oxygenated methine protons 4.01 (H6) linked it to the next spin system, which bore similar signals and correlations to the tetrahydropyran ring system in lagriamide A.

^{1}H-^{1}H COSY correlations from H6 to diastereotopic methylenes at 1.16 (H7), 1.64 (H7'), 2.55 (H5), and 2.69 ppm (H5') linked their carbon atoms as one chain with CH_2-5 assigned between the ketone and methine due to its chemical shifts and HMBC correlations to the ketone carbon. Separately, there were sequential COSY correlations from secondary methyl 0.74 (H10) to methine 1.76 (H9), oxygenated methine 4.11 (H11), and finally to diastereotopic methylene 2.18 (H12) and 2.47 (H12'). Reciprocal HMBC correlations between H6 and 73.2 (C11), and shared HMBC correlations from both oxygenated methine protons to 26.1 (C8), completed this section as a trisubstituted tetrahydropyran ring system. Sequential HMBC correlations from H11 to C12 and H12/H12' to 170.6 (C13) closed the spin system at the C-terminus of an amide. From this, it was clear that this region of the molecule is highly similar to one end of lagriamide A, except with an alkene instead of an epoxide in the C2 and C3 positions. To emphasise these large similarities with few key differences between the lagriamides, a comparison of ^{13}C chemical shifts is presented in **Figure S5**.

Investigating correlations to the carboxylic acid signal in lagriamide B—the moiety which forms the other terminus in lagriamide A—leads to a similar finding. HMBC correlations to the carboxylic acid 178.1 (C39) from the methyl at 1.01 (H38), methine 2.32 (H37), and methylene protons 1.22 and 1.41 (H36 and H36') revealed a key difference of lacking two methylenes adjacent to the acid as found in lagriamide A at its C39 and C40 positions. Further, a shared HMBC correlation to methylene 41.0 (C36) from both H38 and 0.81 (H35) established a 1,3-dimethyl relationship, explaining the lack of the same olefin found at the C36 and C37 positions of lagriamide A.

From positions C13 to C35, the chemical shifts and correlations are highly similar between the two lagriamides, with near-zero or zero-ppm differences in ^{13}C chemical shifts. In lagriamide B, an HMBC cross-peak to C13 connected the amide at its N-terminus to diastereotopic methylene 3.01 (H14) and 3.16 (H14'). Sequential COSY correlations from H14 to oxygenated methine 3.49 (H15), methine 2.32 (H16), and up to methyl 0.95 (H17) established a short chain with a hydroxy group at C15 and a methyl group on C16. HMBC correlations from H15, H16, and H17 to amide carbon 174.2 (C18) closed the spin system at the C-terminus of the second amide. Concurrently, an HMBC cross-peak between methylene 3.05 (H19) and C18 connected it to the next section.

Assembling the dioxaspiro ring system in lagriamide B required TOCSY data. Sequential correlations from H19 to methylene 1.36 (H20) and 1.68 (H20'), to methylene 1.20 (H21) and 1.60 (H21'), then to 3.06 (H22) placed a short chain between the amide N-terminus and the oxygenated methine. HMBC correlations to qC 94.6 (C27) from oxygenated methine H22 and multiple aliphatic methylene signals (δ_H 1.28–1.53) were indicative of heterospiro rings. With the tetrahydropyran, this accounted for the three remaining degrees of unsaturation. A COSY cross-peak between 0.77 (H24) and 1.20 (H23) placed a methyl group at 34.5 (C23). An HMBC correlation from H22 to the methyl 17.9 (C24) reinforced its placement next to C22. The five methylenes within the spiro (CH$_2$-25 to 26, CH$_2$-28 to 30) were assigned through analyses of HMBC and TOCSY data. The oxygenated methine at 3.41 (H31) and its COSY cross-peaks closed the second ring of the spiro and connected it to the exocyclic methylene at 33.0 (C32). Lacking an HMBC correlation to C32, an HSQC-TOCSY from H35 to C32 was needed to confirm the placement of the methylenes CH$_2$-32 and CH$_2$-33 as exocyclic and in between the C31 and C34 methines. Finally, an HMBC correlation from H35 to methylene 32.3 (C33), then another from its protons 1.06 and 1.50 (H33 and H33') to the methine at 68.4 (C31) formed the connection between carboxylic acid terminus to the dioxaspiro system like in lagriamide A, thus completing the planar structure of lagriamide B. A summary of chemical shifts assignments, COSY correlations, and ^{1}H-^{13}C HMBC correlations is presented in **Table S4**.
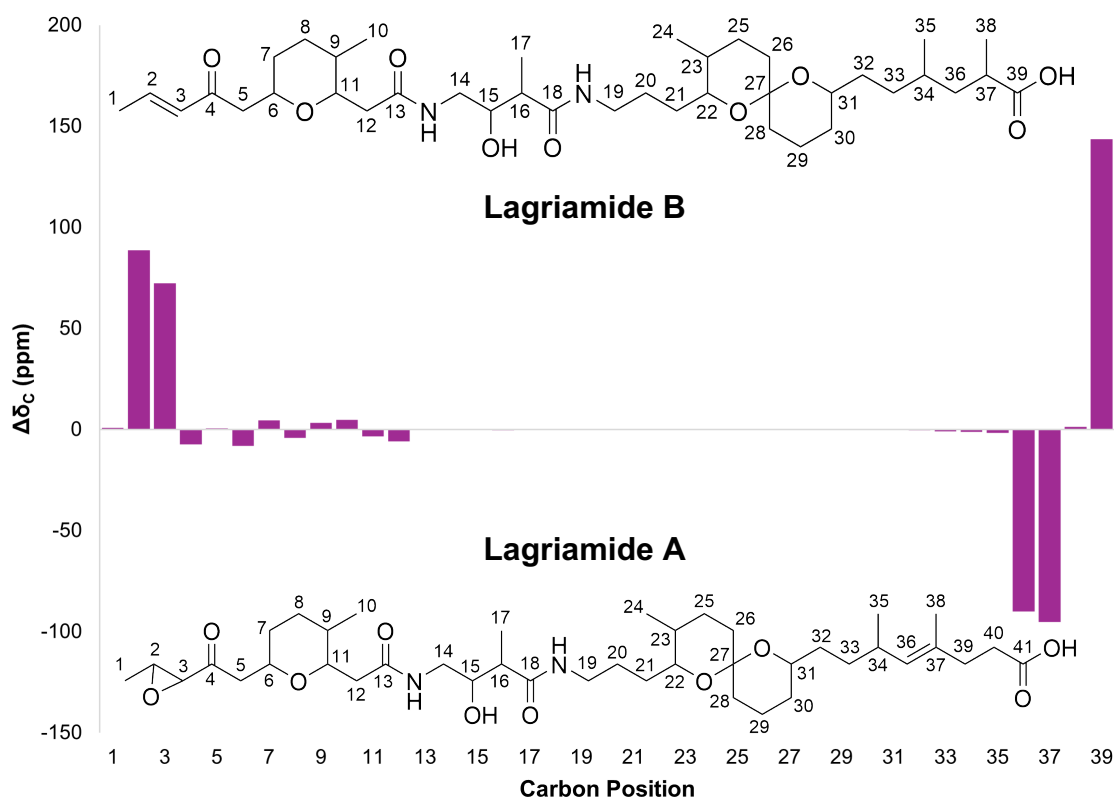

Figure S5: Chemical shift differences between carbons (ppm) as sequentially numbered in lagriamide B (top) and A[8] (bottom). Both assignments were made in DMSO-d_6 at 150 MHz.

Table S4: Lagriamide B NMR shifts in DMSO at 150 MHz (^{13}C) and 600 MHz (^{1}H).

Carbon	^{13}C (ppm)	^{1}H (ppm)	Multiplicity [J (Hz)]	^{1}H-^{1}H COSY	HMBC
1	18.0	1.85	dd [1.6, 6.8]	2, 3	2, 3, 4
2	142.3	6.83	dq [15.9, 6.8]	1, 3	1, 4
3	132.0	6.10	dq [15.8, 1.7]	1, 2	1, 4, 5
4	198.0	–	–	–	–
5	45.2	2.55	dd [6.03, 15.33]	5', 6	4, 6, 7
		2.69	dd [7.01, 15.37]	5, 6	4, 6, 7
6	66.1	4.01	dddd [2.9, 6.5, 6.5, 9.3]	5, 5', 7	4, 5, 8, 11
7	29.9	1.16	dddd [4.2, 10.0, 13.5, 13.5]	6	6, 8
		1.64	m		8
8	26.1	1.33	dddd [3.5, 3.5, 3.9, 12.6]		6
		1.56	dddd [3.9, 11.3, 11.3, 14.9]		6
9	32.5	1.76	dddq [3.7, 12.2, <1, 6.4]	10, 11	10, 11
10	16.3	0.74	d [6.94]	9	8, 9, 11
11	73.2	4.11	ddd [5.0, 5.0, 9.3]	9, 12, 12'	6, 8, 12, 13
12	33.8	2.18	dd [5.4, 14.4]	11	9, 11, 13
		2.47	dd [9.8, 14.2]	11	9, 11, 13
13	170.6	–	–	–	–
14	43.0	3.01	m	15	13, 14, 15, 16
		3.16	m		13
15	71.6	3.49	ddd [6.6, 10.6, 10.6]	14, 16	16, 17, 18
16	43.1	2.32	m	15, 17	15, 17, 18
17	14.4	0.95	d [7.0]	16	15, 16, 18
18	174.2	–	–	–	–
19	38.7	3.05	m	20, 20'	18, 20, 21
20	25.7	1.36	m	19	19, 22
		1.68	m	19	19, 22
21	29.9	1.20	m	22	20, 22
		1.60	m	22	20, 22
22	73.8	3.06	ddd [2.1, 9.7, 9.7]	21, 21', 23	24, 25, 27
23	34.5	1.20	dddq [<1, 8.8, 15.2, 6.1]	22, 24	22
24	17.9	0.77	d [6.5]	23	22, 23, 25
25	27.6	1.42	m		21, 22, 26, 27
		1.45	m		22, 27
26	35.5	1.36	m		22, 27
		1.49	m		22, 23, 27
27	94.6	–	–	–	–
28	34.9	1.28	m		26, 27, 30, 31
		1.48	m		26, 27, 31
29	18.8	1.44	m		27, 30, 31
		1.73	m		30
30	30.8	1.05	m [10.7, …]	31	30, 31
		1.53	m [<1, …]		28, 31, 32
31	68.4	3.41	dddd [10.7, <1, 3.9, 6.3]	30, 32, 32'	
32	33.0	1.30	m [3.9, …]	31	31
		1.37	m [6.3, …]	31	31
33	32.3	1.06	m		30, 31, 34, 35
		1.50	m		32
34	30.1	1.45	m	35	31, 36
35	19.5	0.81	d [6.2]	34	33, 34, 36
36	41.0	1.22	m	37	33, 34, 37, 39
		1.41	m	37	33, 34, 35, 37, 38, 39
37	37.1	2.32	m	36, 36', 38	34, 36, 38, 39
38	17.3	1.01	d [6.9]	37	36, 37, 39
39	178.1	–	–	–	–

Configurational Analysis of Lagriamide B

To solve the complete absolute configuration of lagriamide B, we combined predictions made from the genome sequence, information from NMR data, and predictions made using DP4 analysis. Unfortunately, attempts to form Mosher's ester derivatives were unsuccessful due to issues with compound stability. General computation methods for DP4 and full NMR spectra are provided at the end of this section.

DP4 calculations of lagriamide B fragments A-D

Figure S6–Figure S9 showcase the simulated diastereomers of lagriamide B fragments A-D. In all cases the structures indicated in black were simulated, while the structures indicated in red are the corresponding enantiomeric set and were not simulated but are shown for convenience. Nuclei indicated by an orange sphere (^{1}H/^{13}C) were excluded from the DP4 probability calculation because they were either 1) exchangeable protons with variable chemical shifts; or 2) no longer comparable to the corresponding nuclei in lagriamide B. The latter reason for exclusion from the DP4 calculation is particularly important, since as a direct result of truncation the termini of the fragments end up being in significantly different magnetic environments than the corresponding nuclei in lagriamide B (e.g. what used to be a methylene in lagriamide B is now a methyl group in some fragments of lagriamide B). **Table S5–Table S8** indicate the relative DP4 probabilities expressed as a percentage for the ^{1}H, ^{13}C, or combined ^{1}H and ^{13}C chemical shifts. Several notable points are worth discussing. First, fragments A-C showed an unambiguous match to a single diastereomer (or its corresponding enantiomer) as indicated by a bold entry in each fragments respective table (**Table S5–Table S7**).

DP4 analysis of fragment A indicated a very high probability for the 6S_9S_11R diastereomer, or its enantiomer (6R_9R_11S). DP4 analysis of fragment B unambiguously showed strong agreement for the 15S_16R diastereomer, or its enantiomer (15R_16S).

Figure S6: Simulated structures and their enantiomeric set for fragment A of lagriamide B.

Table S5: DP4 probabilities for fragment A of lagriamide B.

Compound	DP4 ^{1}H	DP4 ^{13}C	DP4 ^{1}H + ^{13}C
A_6S_9R_11S	0.0%	0.0%	0.0%
A_6R_9R_11S	0.7%	0.4%	0.0%
A_6S_9S_11S	0.8%	0.0%	0.0%
A_6S_9R_11R	**98.4%**	**99.6%**	**100.0%**

Figure S7: Simulated structures and their enantiomeric set for fragment B.

Table S6: DP4 probabilities for fragment B of lagriamide B.

Compound	DP4 ^{1}H	DP4 ^{13}C	DP4 ^{1}H + ^{13}C
B_15S_16R	**83.6%**	**90.6%**	**98.0%**
B_15R_16R	16.4%	9.4%	2.0%

While the ^{13}C was ambiguous for fragment C, the combined ^{1}H and ^{13}C DP4 probability was unambiguous for the 22S_23R_27R_31S diastereomer. Unfortunately, DP4 analysis of fragment D yielded ambiguous results with the ^{1}H and ^{13}C DP4 probabilities contradicting each other. This may be due to the fragment no longer adequately reflecting the conformational ensemble of the alkyl chain/acid when it is attached to the spirocyclic ring system. This result prompted us to perform a DP4 calculation on a larger fragment using the relative configuration ascertained from the DP4 analysis of fragment C and varying the configuration of the dimethyl functionalities within the alkyl/acid side chain.
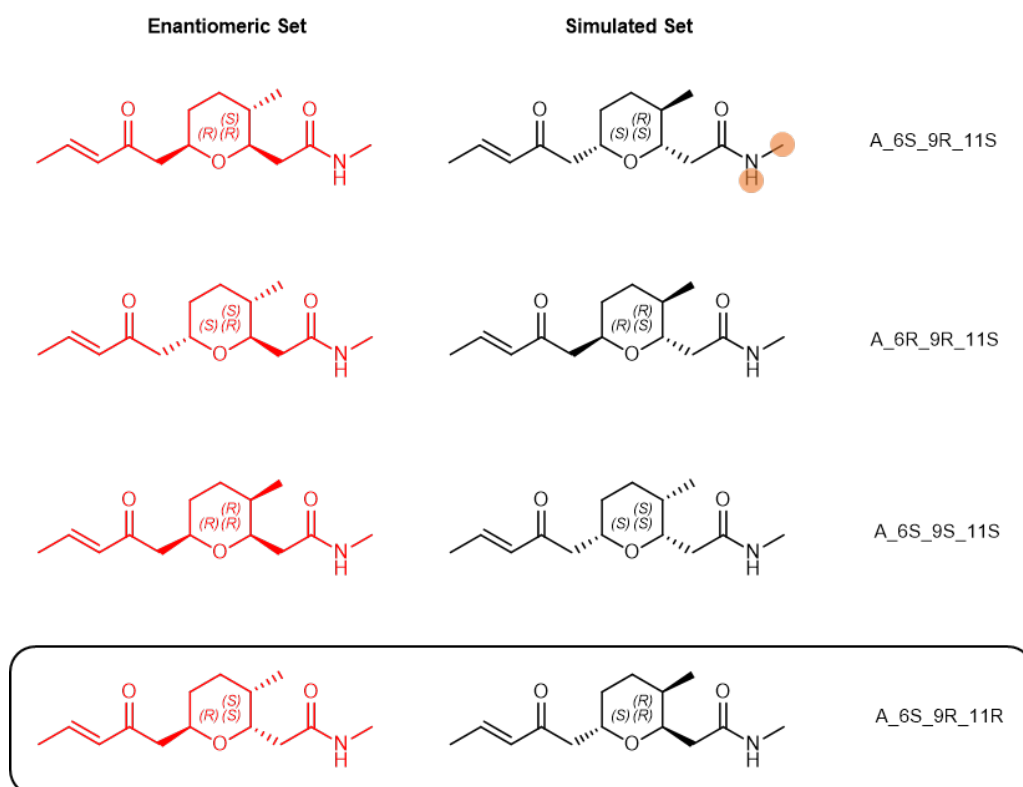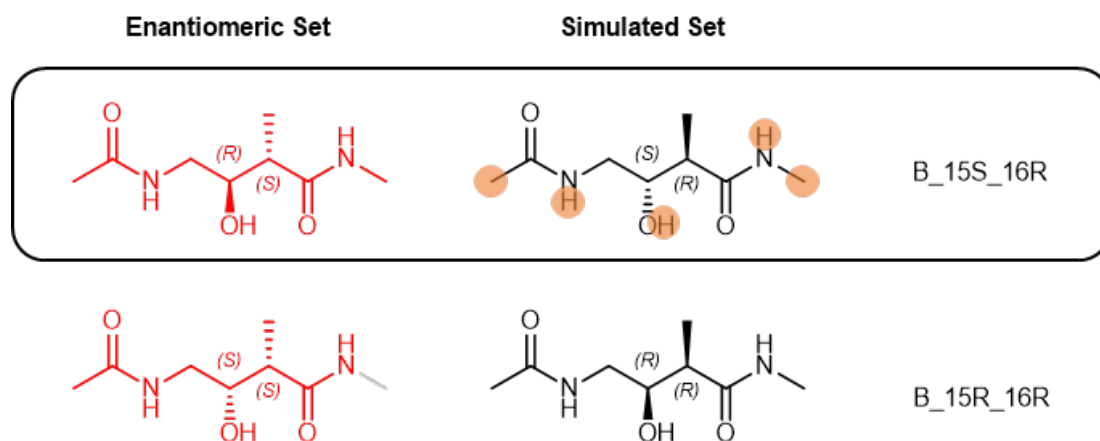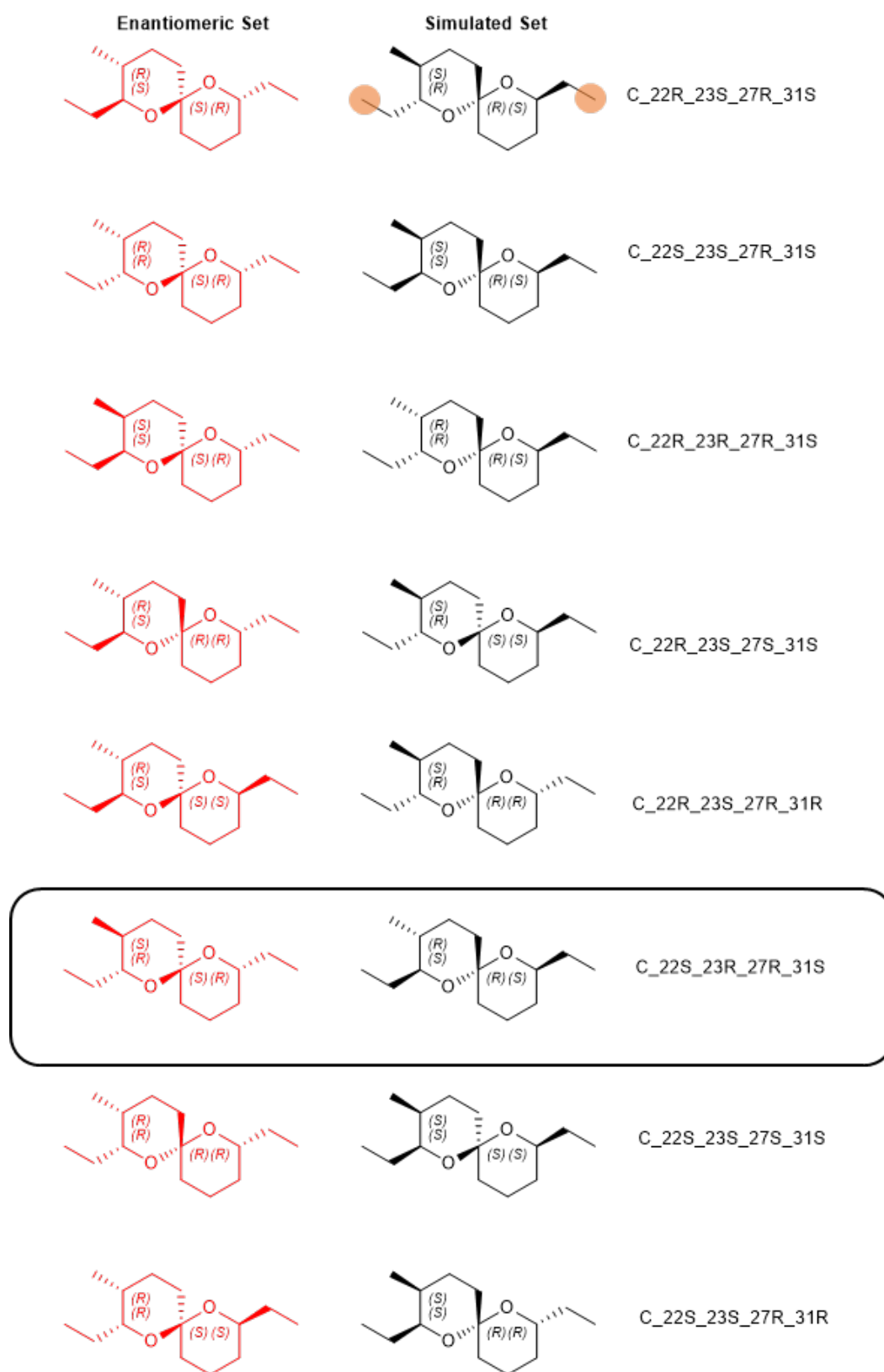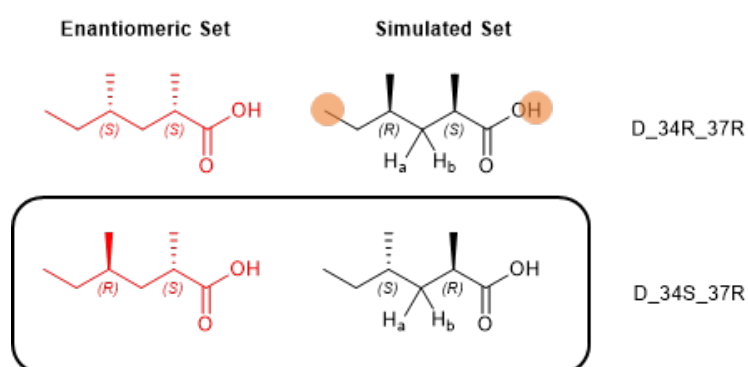
Figure S8: Simulated structures and their enantiomeric set for fragment C of lagriamide B.

Table S7: DP4 probabilities for fragment C of lagriamide B.

Compound	DP4 ^{1}H	DP4 ^{13}C	DP4 ^{1}H + ^{13}C
C_22R_23S_27R_31S	0.1%	2.3%	0.0%
C_22S_23S_27R_31S	0.0%	0.0%	0.0%
C_22R_23R_27R_31S	0.1%	2.6%	0.0%
C_22R_23S_27S_31S	0.0%	1.1%	0.0%
C_22R_23S_27R_31R	0.0%	60.8%	0.0%
C_22S_23R_27R_31S	**99.9%**	**31.3%**	**100.0%**
C_22S_23S_27S_31S	0.0%	1.8%	0.0%
C_22S_23S_27R_31R	0.0%	0.0%	0.0%

Figure S9: Simulated structures and their enantiomeric set for fragment D of lagriamide B.

Table S8: DP4 Probabilities for fragment D of lagriamide B.

Compound	DP4 ^{1}H	DP4 ^{13}C	DP4 ^{1}H + ^{13}C
D_34R_37R	32.6%	53.7%	36.0%
D_34S_37R	67.4%	46.3%	64.0%

Figure S10 showcases the simulated diastereomers of lagriamide B fragment C+D. The structures indicated in black were simulated, while the structures indicated in red are the corresponding enantiomeric set and were not simulated but are shown for convenience. Nuclei indicated by an orange sphere (^{1}H/^{13}C) were excluded from the DP4 probability calculation as they were no longer comparable to the corresponding nuclei in lagriamide B (see discussion for DP4 analysis of fragments A-D). Due to the flexible nature of these diastereomers, the conformational search was expanded to 50,000 MD runs/energy minimizations.

Table S9 shows the calculated DP4 probabilities for each of the calculated diastereomers from **Figure S10**. ^{1}H+^{13}C DP4 probability indicates a moderate probability that the configuration is 22S_23R_27R_31S_34R_37R (or its enantiomer). However, the ^{13}C DP4 probability is ambiguous on its own. On the basis of the data we can conclude that the relationship of the 1,3 dimethyl group is almost certainly 1,3-syn, and that the relative configuration of the C+D fragment is most likely 22S_23R_27R_31R_34R_37R (or its enantiomer).
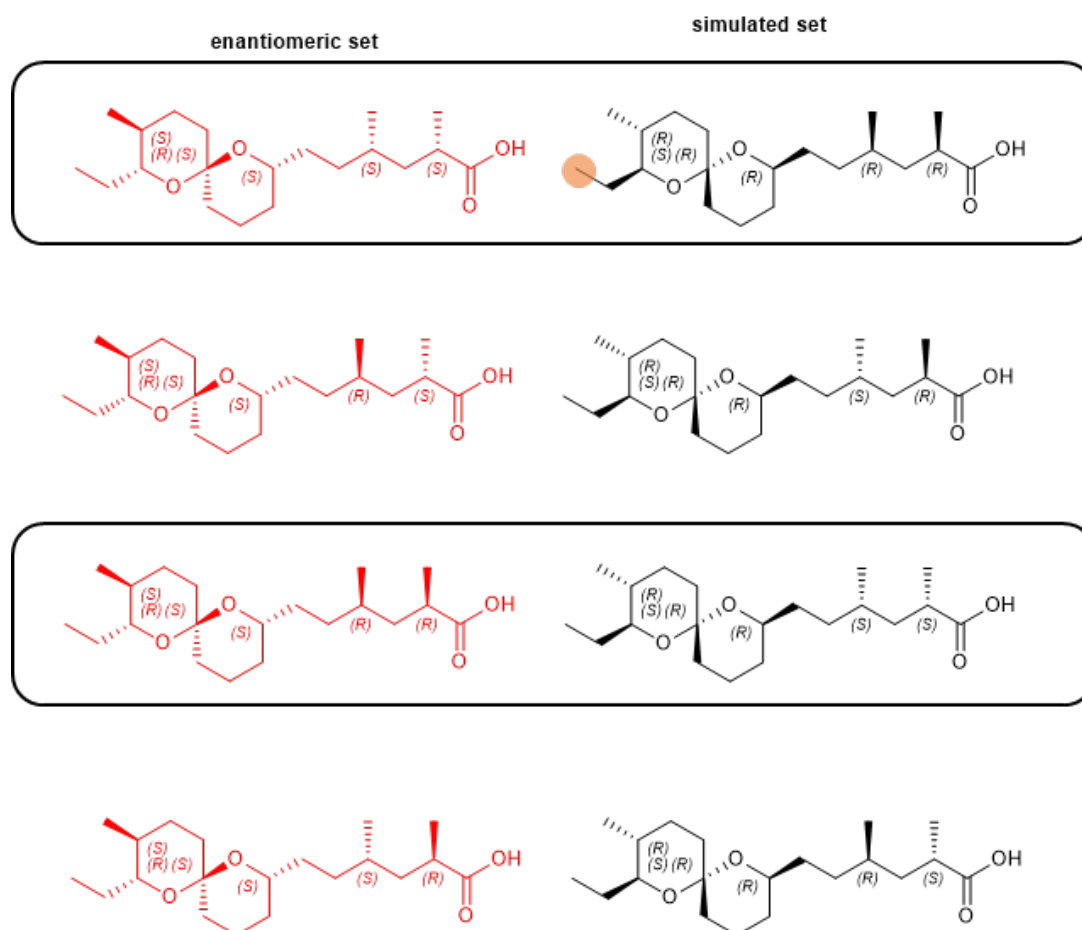
Figure S10: Simulated structures and their enantiomeric set for fragments C+D.

Table S9: DP4 probabilities for fragments C+D of lagriamide B.

Compound	DP4 ^{1}H	DP4 ^{13}C	DP4 ^{1}H + ^{13}C
CD_22S_23R_27R_31R_34R_37R	**83.7**	**32.5%**	**75.2%**
CD_22S_23R_27R_31R_34S_37R	0.1%	1.3%	0.0%
CD_22S_23R_27R_31R_34S_37S	14.9%	60.1%	24.6%
CD_22S_23R_27R_31R_34R_37S	1.3%	6.1%	0.2%

The relative configurations of regions A-C in lagriamide B are very likely to be as indicated in **Figure S6–Figure S8**, and the relative configuration of the dimethyl groups in region D is almost certainly 1,3-syn as per **Figure S10** and **Table S9**. Finally, the likely configuration of the C+D fragment is 22S_23R_27R_31R_34R_37R (or its enantiomer).

Relative configurations of lagriamide B regions A and C using NMR experiments
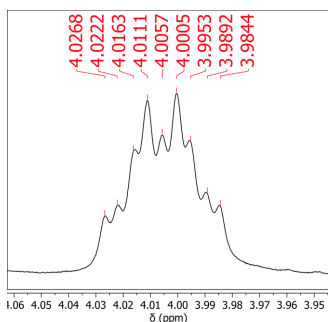

**Zoom into ^{1}H NMR of lagriamide B (Figure S16):** H6 (δ4.01)

The relative configuration of the tetrahydropuran ring system (region A) was determined using ROESY correlations and key coupling constants between protons in the ring system. The set of coupling constants between H6 and its neighbouring protons includes one large coupling constant ($^3J_{\text{H5,5'-H6}}$ = 6.5 Hz, $^3J_{\text{H6-H7}}$ = 9.3 Hz, $^3J_{\text{H6-H7'}}$ = 2.9 Hz), which is only possible if H6 is in an axial position.

A ROESY correlation between H9 and H11 (**Figure S21**) precludes a 1,2-diaxial relationship and tells us that at least one of H9/H11 is in an equatorial position. A 1D TOCSY selectively irradiating H10 (**Figure S24**) separated H9 from overlapping signals found in the ^{1}H NMR, allowing us to evaluate the coupling constants to H9. Excluding its coupling constant to the methyl protons of C10, H9 has one large and two small coupling constants ($^3J_{\text{H9-H10}}$ = 6.4 Hz, $^3J_{\text{H8-H9}}$ = 12.2 Hz, $^3J_{\text{H9-H11}}$ ≤ 3.7 Hz, $^3J_{\text{H8'-H9}}$ ≤ 3.7 Hz). The sole large coupling constant must arise from a 1,2-diaxial relationship with one of the protons on the neighbouring methylene C8. Therefore, H9 is in an axial position, while H11 is in an equatorial position.
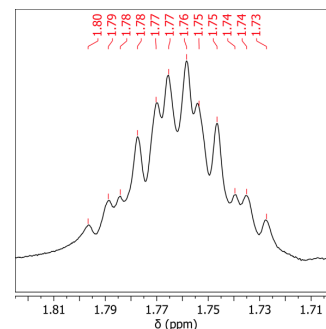
Zoom into 1D TOCSY selectively irradiating H10 (Figure S24): H9 (δ0.76)

Key ROESY correlations as depicted in **Figure 3B** of the main text—in particular the correlations between H6/H7', H6/H8 on one face of the six-membered ring and between H9/H7, H9/H8', H9/H11 on the other face—support this relative configuration (**Figure S21**). As such, we determined through NMR data that the absolute configuration of region A can be either 6*S*_9*R*_11*R* or 6*R*_9*S*_11*S*.
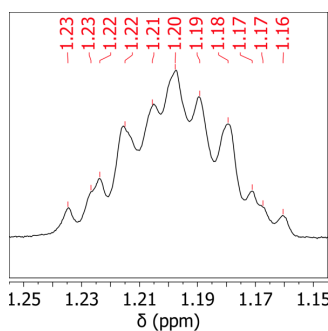
Zoom into 1D TOCSY selectively irradiating H24 (Figure S25): H23 (δ1.20)

The relative configuration of the spiroacetal system (region C) was determined in a similar manner. A 1D TOCSY selectively irradiating H24 (**Figure S25**) separated H23 from overlapping signals. This revealed the presence of one small and two large coupling constants ($^3J_{\text{H23-H25 or H25'}}$ = 11.0 Hz, $^3J_{\text{H23-H25 or 25'}}$ = 4.8 Hz, $^3J_{\text{H22-H23}}$ = 11.0 Hz) in addition to its coupling constant to the methyl protons of C24 ($^3J_{\text{H23-H24}}$ = 6.7 Hz). This indicates that H22 and H23 have a 1,2-diaxial relationship.

Although the H31 signal was sufficiently discrete from other signals in the ^{1}H NMR spectrum (**Figure S16**), the complexity of its multiplet prevented us from discerning its coupling constants. Therefore, to determine the relative configuration at C31, we performed a set of experiments selectively decoupling H31 from each of its neighbours—H30, H30', H32, and H32' (**Figure 3C, Figure S29**). These show that H31 has a large coupling constant to H30 ($^3J_{\text{H30-H31}}$ = 10.7 Hz), indicating that H31 must be in an axial position.

There is a ROESY correlation between the two protons adjacent to each oxygen atom of the spiroacetal, H22 and H31 (**Figure S21**), as confirmed by a selective 1D ROESY experiment irradiating H31 (**Figure S28**). This key correlation limits the spiro centre at C27 to the relative configuration depicted in **Figure 3B**, which is the only possibility that puts H22 and H31 in close enough proximity as to observe an nOe signal. Therefore, we determined through NMR data that the absolute configuration of region C can be either 22*R*_23*S*_27*S*_31*S* or 22*S*_23*R*_27*R*_31*R*.

Predictions of lagriamide B absolute configurations using BGC analysis

We predicted the absolute configurations of hydroxyl groups based on alignments of KR domains (**Figure S11**) and attributed the relative Fischer configuration to them according to the predicted KR type A or B (**Figure 1**, main text). In summary, the presence or absence of a diagnostic aspartate residue, results either in D-configured alcohols (B-type KR) or L-configured alcohols (A-type KR)[10,11,12]. Based on the alignment we predicted the alcohols stereocenters as indicated in **Figure 1**. It is not currently possible to predict the configuration of methylations catalyzed by cMT domains in *trans*-AT PKS systems, and the configurations of stereocenters modified after PKS processing are also undefined.

Supplementary PKS Domain Analyses

Two KS domains (the last KS on LgaG and the last KS on LgaC) are predicted to be inactive (KS^0) based on mutations on the serine-histidine-histidine tryad responsible for decarboxylative condensation present on motifs C̲SSSLV, H̲GTGT and NIGH̲ (**Figure S13**). Although the KS domains at the C-termini of LgaA and LgaB display all the conserved active site motifs (C-H-H tryad), they appear to be nonfunctional based on the observed lagriamide structure. We found that these two KS domains have mutations in the CSSSLV motif and in the NIGH motif. Thus, the KS domains at the C-termini of LgaA and LgaB were assigned as inactive despite their conserved catalytic tryad. Non-elongating KSs having conserved active site residues have been observed in several other *trans*-AT PKSs such as in etnangien[13], and in patelazolles biosynthesis[14].

The dehydratase (DH) domain sequence alignment (**Figure S14**) showed the presence of the catalytic aspartic acid in PXLLD̲XXXQ/H postulated to donate a proton to the β-hydroxy group and the catalytic histidine present in a H̲xxxGxxxxP motif for all individual cases, which is inconsistent with the presence of a hydroxy group at position 4 (further converted into α,β-unsaturated carbonyl), and hydroxyl group at C-10. After careful inspection we found mutations on the conserved GLXYGP for the second DH in LgaG (GWSHGP), and a mutation on the HXXXGXXXXP motif (HXXXGXXXXS) for DH5 of LgaG$_P$. However, it is unclear if these mutations will lead to inactive domains.
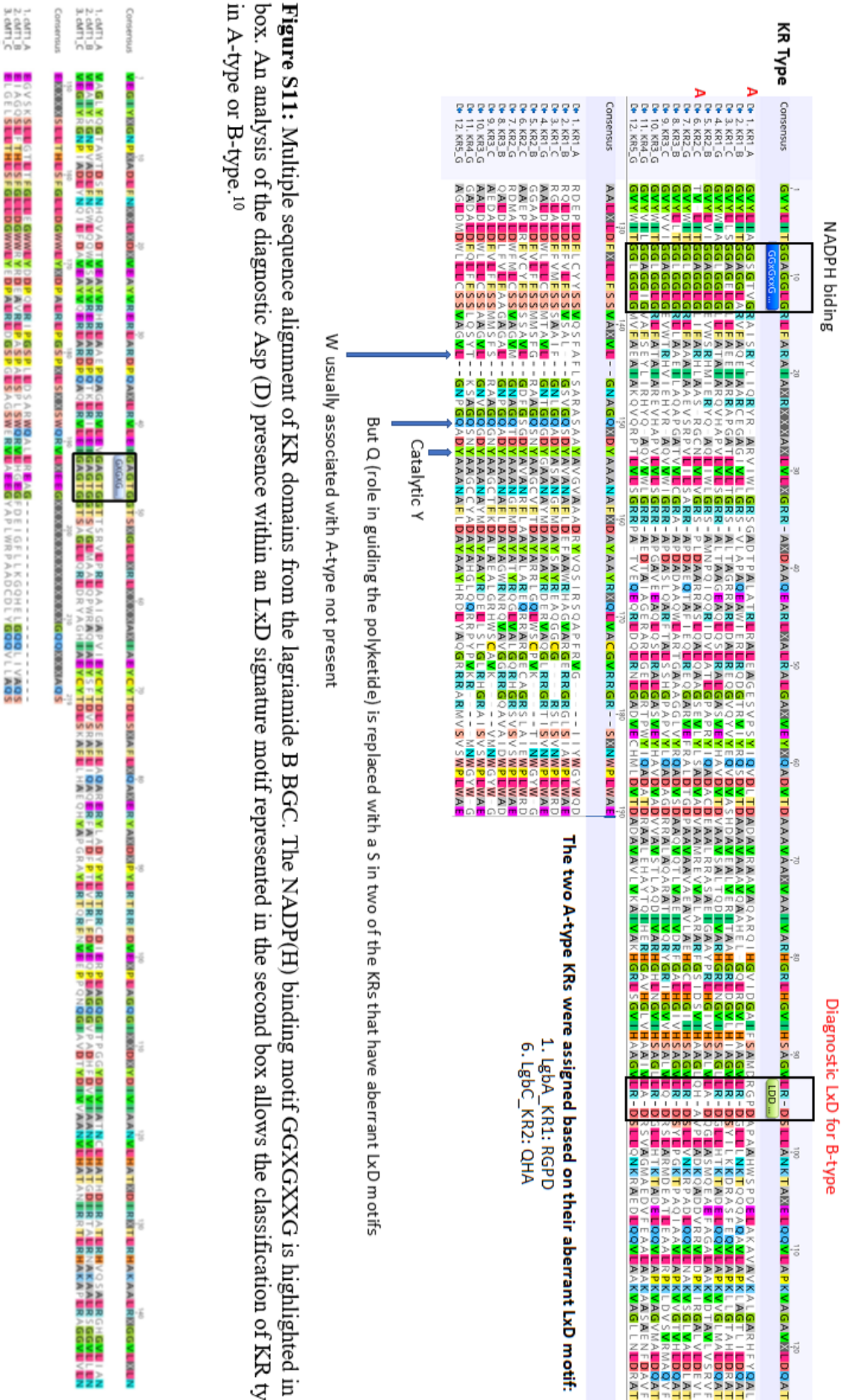
Figure S11: Multiple sequence alignment of KR domains from the lagriamide B BGC. The NADP(H) binding motif GGxGXXG is highlighted in a box. An analysis of the diagnostic Asp (D) presence within an LxD signature motif represented in the second box allows the classification of KR types in A-type or B-type.[10]

Figure S12: Multiple sequence alignment of cMT domains from the lagriamide B BGC. The SAM-binding site motif GxGXGXG is highlighted in a box.
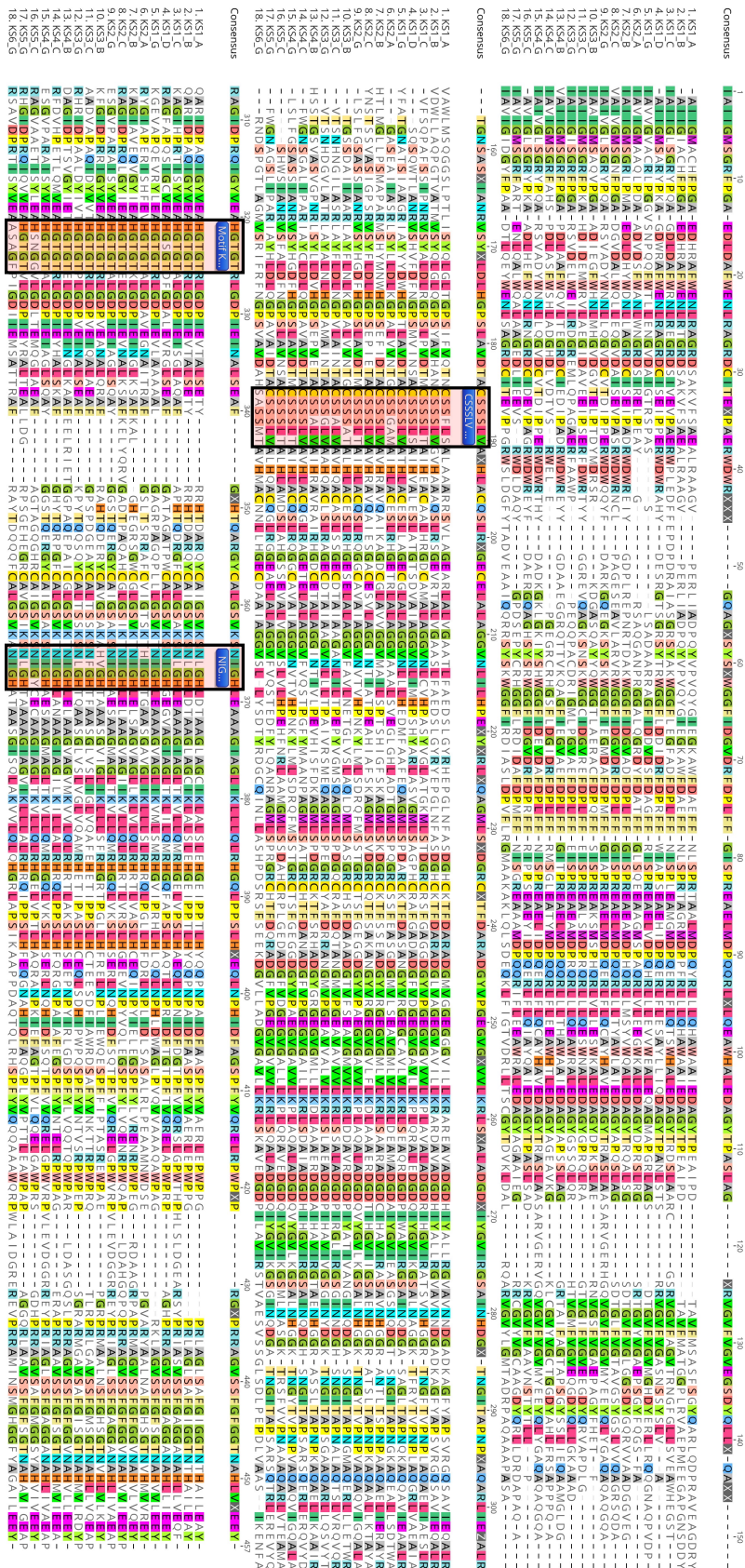
Figure S13: Multiple sequence alignment of the amino acids of KS domains from the lagriamide B BGC. The conserved motifs that include the Cys-His-His catalytic triad (CSSSLV, HGTGT, NIGH) essential for decarboxylative condensation are highlighted in boxes[15].
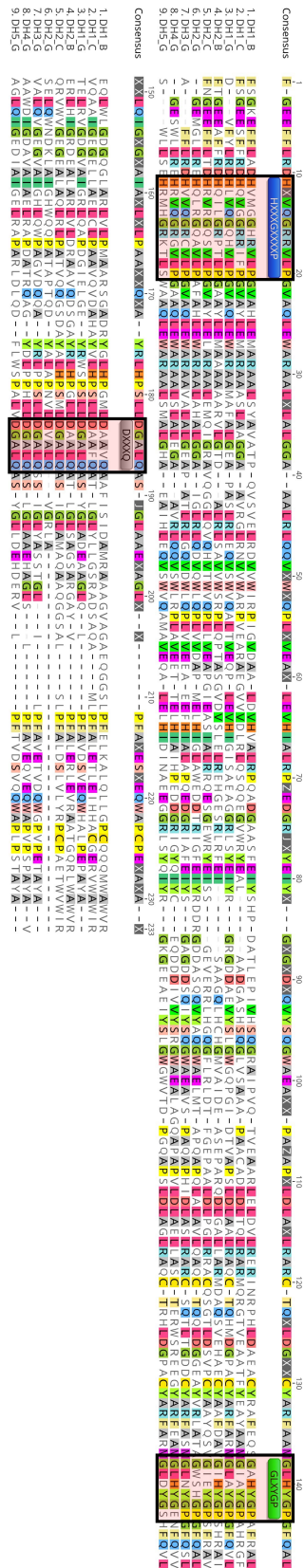
Figure S14: Multiple sequence alignment of DH domains from the lagriamide B BGC. The catalytic Asp (PXLLDXXXQ/H) and His residues (HXXXGXXXXP) are highlighted in their respective motifs, along with the conserved motif GLXYGP where the Tyr may aid in β-hydroxyl group binding.[13,16]
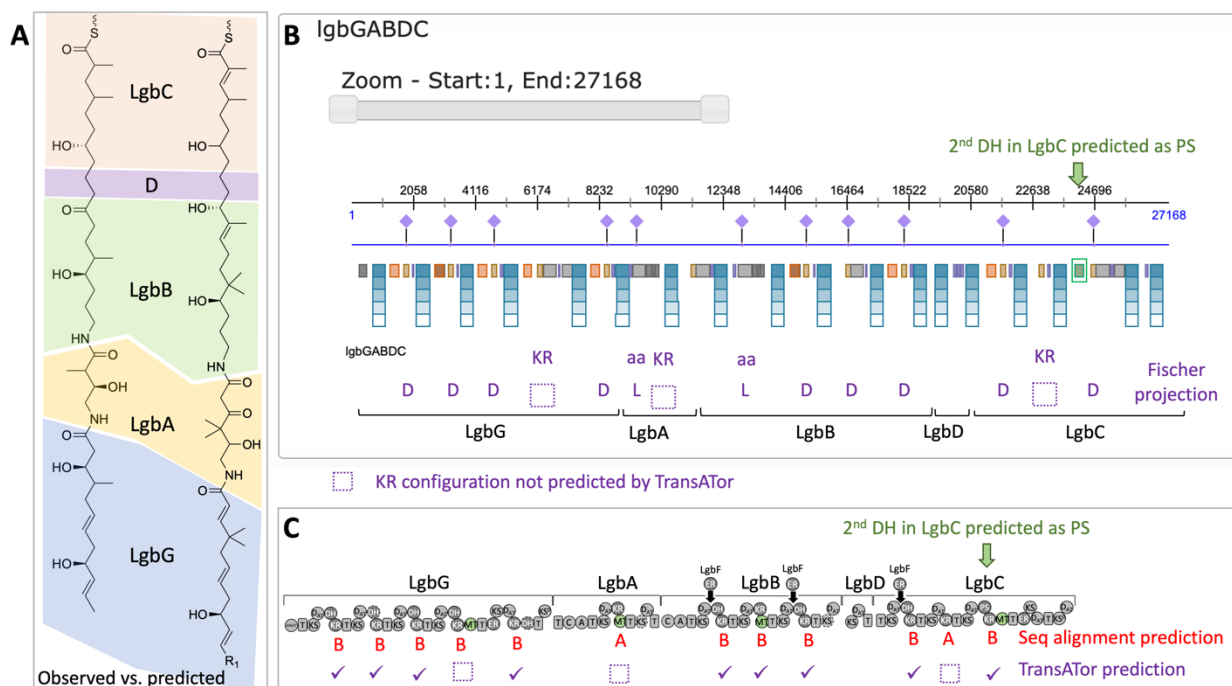
Figure S15: TransATor analysis for the *lgb* PKS and PKS-NRPS proteins. (A) Observed (left) vs. predicted (right) growing chain on last module of LgbC. Color represents observed and predicted structures for each Lgb PKS. (B) TranATor trans PKS-NRPS annotation result.[17] The primary protein sequence of the Lgb PKS-NRPS was used as input in the predicted order of biosynthesis that matches the observed structure (LgbGABDC). The chart shows an interactive protein scale with colored boxes representing HMM-based annotations. The top 5 predicted KS substates are shown as blue gradient boxes. The purple diamonds highlight Fisher projection predictions, including KR-based hydroxyl group stereochemistry prediction and amino acid configuration, which we added at the bottom of the chart also in purple. The analysis predicted the second DH domain in LgbC to be a pyran synthase (PS, green). However, this does not match our DH alignments which showed the presence of a DxxxQ motif in all DHs whereas PSs lack this motif (**Figure S14**). (C) Summary of sequence alignment-based prediction of KR type and TransATor KR predictions.

General computational methods for DP4 analysis

In-silico generation of initial geometries for conformational search

SMILES strings for each of the diastereomers for the lagriamide B fragments (see **Figure S6–Figure S10**) were converted to molecule objects in RDKit 2017.09.[18] RDKit's distance geometry algorithm was used to generate an initial geometry as an SDF file for each compound. Each of the initial geometries were then parametrized for the MMFF94[19] forcefield with sdf2tinkerxyz[20].

Conformational search protocol

Using the Tinker Molecular Modelling Package[21,22] each diastereomer (see **Figure S6–Figure S10**) was subjected to 10,000-50,000 (see individual cases below) gas-phase molecular dynamics (MD) simulations of 1 picosecond at 1000K using the Berendsen temperature coupling bath, with a time step of 1 femtosecond. The first iteration started from the randomly generated geometry as described above, and the resultant geometry was then energy minimized using a Newton-Raphson minimization algorithm to an atom root-mean-square gradient of 0.01 kcal/mol/Å. Subsequent MD steps repeated the above process starting from the energy minimized geometry of each preceding step until 10,000-50,000 (see individual cases below) iterations were reached. Throughout the course of the simulations, stereochemistry was enforced, and a soft harmonic constraint (k = 0.005 kcal/deg^2) was applied to the torsion angles of all double bonds when they rotated beyond 45° from their ideal geometry defined in the SMILES string (*i.e. E* or *Z*). Together, this ensured that the correct stereochemistry was maintained, and prevented double bonds from isomerizing during conformational sampling while ensuring maximum conformational flexibility.

Conformation sorting and duplicate elimination

After completion of the conformational sampling protocol the resultant 10,000-50,000 conformations were sorted from lowest to highest energy. Conformations with an energy greater than 10.0 kcal/mol relative to the lowest energy conformation identified for each individual molecule were discarded. Finally, all pairs of conformers were overlaid, and their pair-wise atom RMS displacements were calculated. If the heavy atom root-mean-square displacement was < 0.10 Å between any pair of conformers, the higher energy conformer was discarded.

Quantum mechanical calculations

All quantum mechanical calculations were carried out with the Gaussian software package revision e.01[23]. All MMFF optimized conformers of each diastereomer within 5.0-10.0 kcal/mol (see below for energy limit in each case) of their respectively lowest energy conformer were subjected to NMR GIAO shielding constant calculations at the PCM(solvent=dmso)-mPW1PW91/6-311G(d) level of theory. Single point quantum mechanical energies were calculated at the PCM(solvent=dmso)-M06-2X/6-31G(d,p) level of theory. This level of theory was chosen on the basis of Goodman and co-workers' findings that this approach results in a cost-effective and reliable approach to DP4 calculations[24]. Shielding constants for each nuclei (σ^x) of each conformer were converted to chemical shifts (δ^x, ppm) by calculation of the ^{1}H and ^{13}C shielding constants of tetramethylsilane (σ^0) at the PCM(solvent=dmso)-mPW1PW91/6-311G(d) level of theory followed by substitution of the appropriate quantities into **equation 1**.

$$\text{Equation 1}: \delta^x = \frac{\sigma^0 - \sigma^x}{1 - \sigma^0/10^6}$$

Subsequently, Boltzmann averaged chemical shifts were calculated for each nuclei in a particular diastereomer using **equation 2**. In this equation δ^x_{calcd} is the Boltzmann averaged chemical shift in ppm relative to tetramethylsilane, δ^x_i is the chemical shift of nucleus x, E_i is the energy of conformer i relative to the lowest energy conformer identified (i.e. the putative global minimum), T is 298.15 K, and R is the ideal gas constant (8.3145 J/(mol K)).

$$\text{Equation 2}: \delta^x_{calcd} = \frac{\sum_i \delta^x_i e^{(-\frac{E_i}{RT})}}{\sum_i e^{(-\frac{E_i}{RT})}}$$

We limited the averaging process for the NMR chemical shifts for each diastereomer to conformers within 2.39 kcal/mol (10 kJ/mol) (M06-2X/6-31G(d,p) energy) relative to the lowest energy conformer identified for a particular diastereomer.

DP4 calculations and statistical parameters

All DP4 probability calculations were performed using a script published by Goodman and co-workers[25], and the following statistical models were used for the ^{1}H shifts and ^{13}C shifts respectively[24].

^{13}C shifts:
Mean1: -0.138097
StDev1: 1.970539
^{1}H shifts:
Mean1: 0.007649
StDev1: 0.165801

Calculation parameters for lagriamide B fragments A-D

Number MD iterations: 10,000
Conformational search energy cutoff: 10 kcal/mol (all of these conformers subjected to QM calculations)
RMS atom displacement cutoff: 0.10 Å
QM theory level for conformer energies: PCM(solvent=dmso)-M06-2X/6-31G(d,p)
QM theory level for NMR shielding constants: PCM(solvent=dmso)-mPW1PW91/6-311G(d)
NMR conformer energy cutoff: 2.39 kcal/mol

Calculation parameters for lagriamide B fragments C+D

Number MD iterations: 50,000
Conformational search energy cutoff: 5 kcal/mol (all of these conformers subjected to QM calculations)
RMS atom displacement cutoff: 0.10 Å
QM theory level for conformer energies: PCM(solvent=dmso)-M06-2X/6-31G(d,p)
QM theory level for NMR shielding constants: PCM(solvent=dmso)-mPW1PW91/6-311G(d)
NMR conformer energy cutoff: 2.39 kcal/mol

NMR Spectra of Lagriamide B
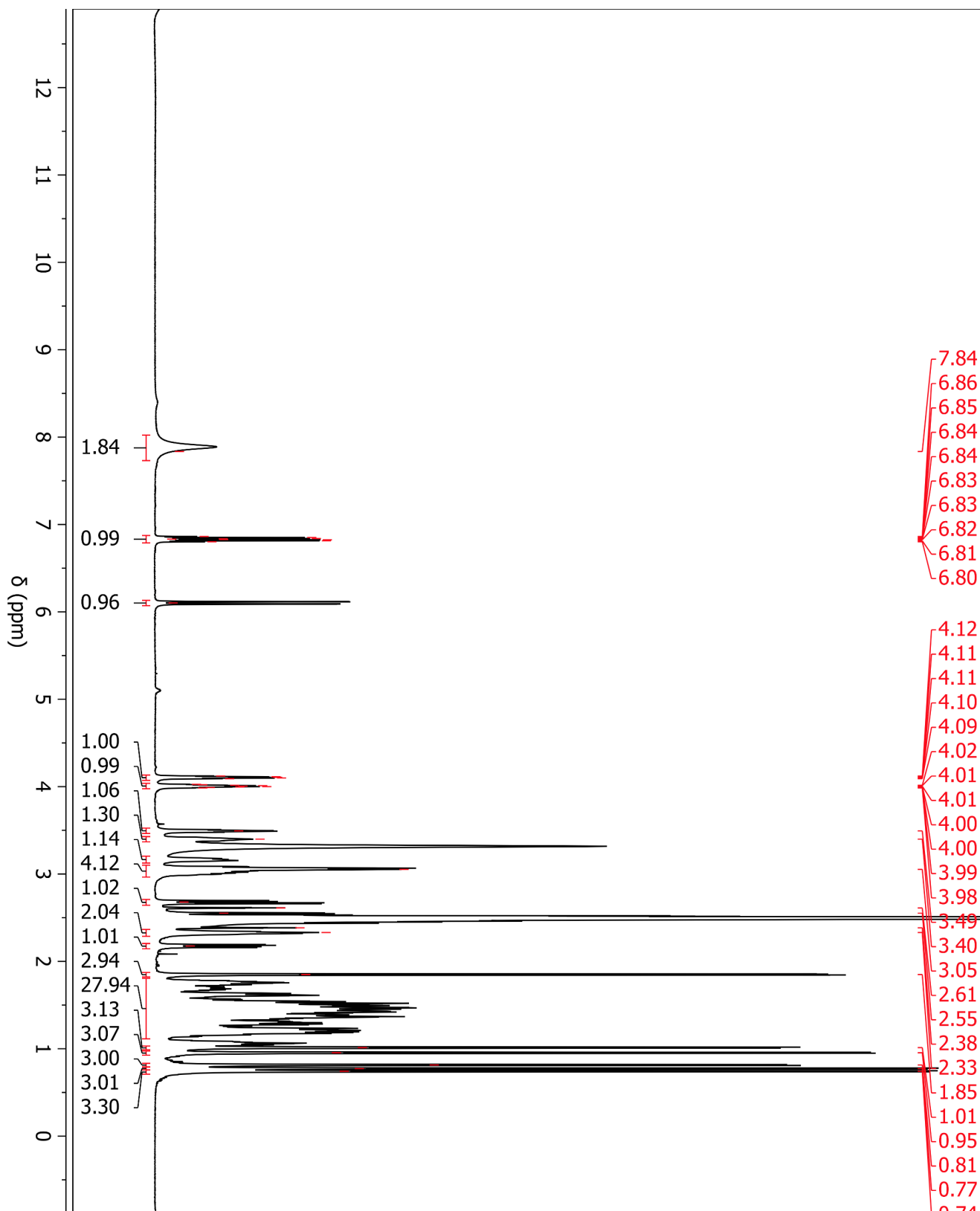
Figure S16: ^{1}H NMR spectrum for lagriamide B acquired in DMSO-d$_6$ at 600 MHz.
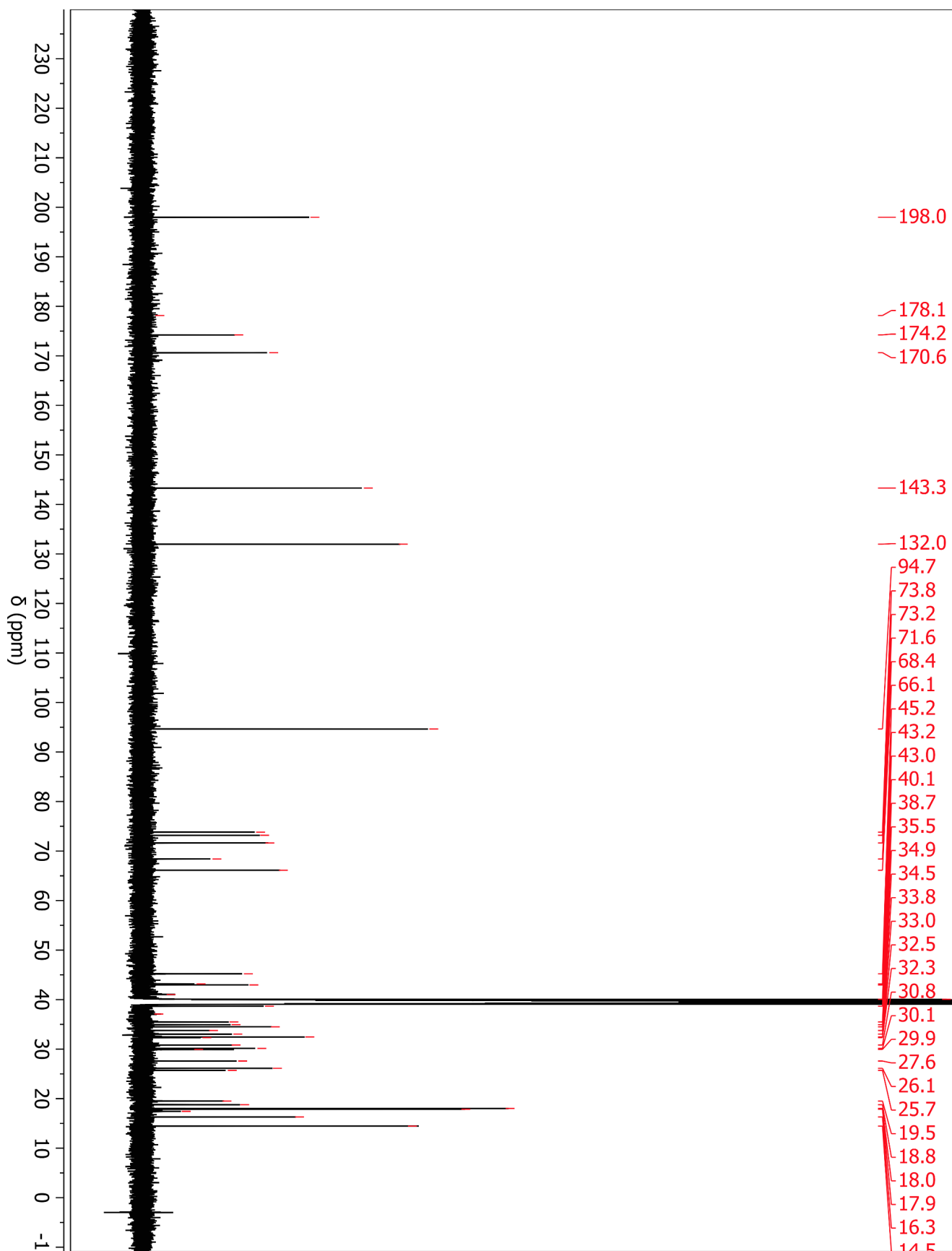
Figure S17: ^{13}C NMR spectrum for lagriamide B acquired in DMSO-d_6 at 150 MHz.
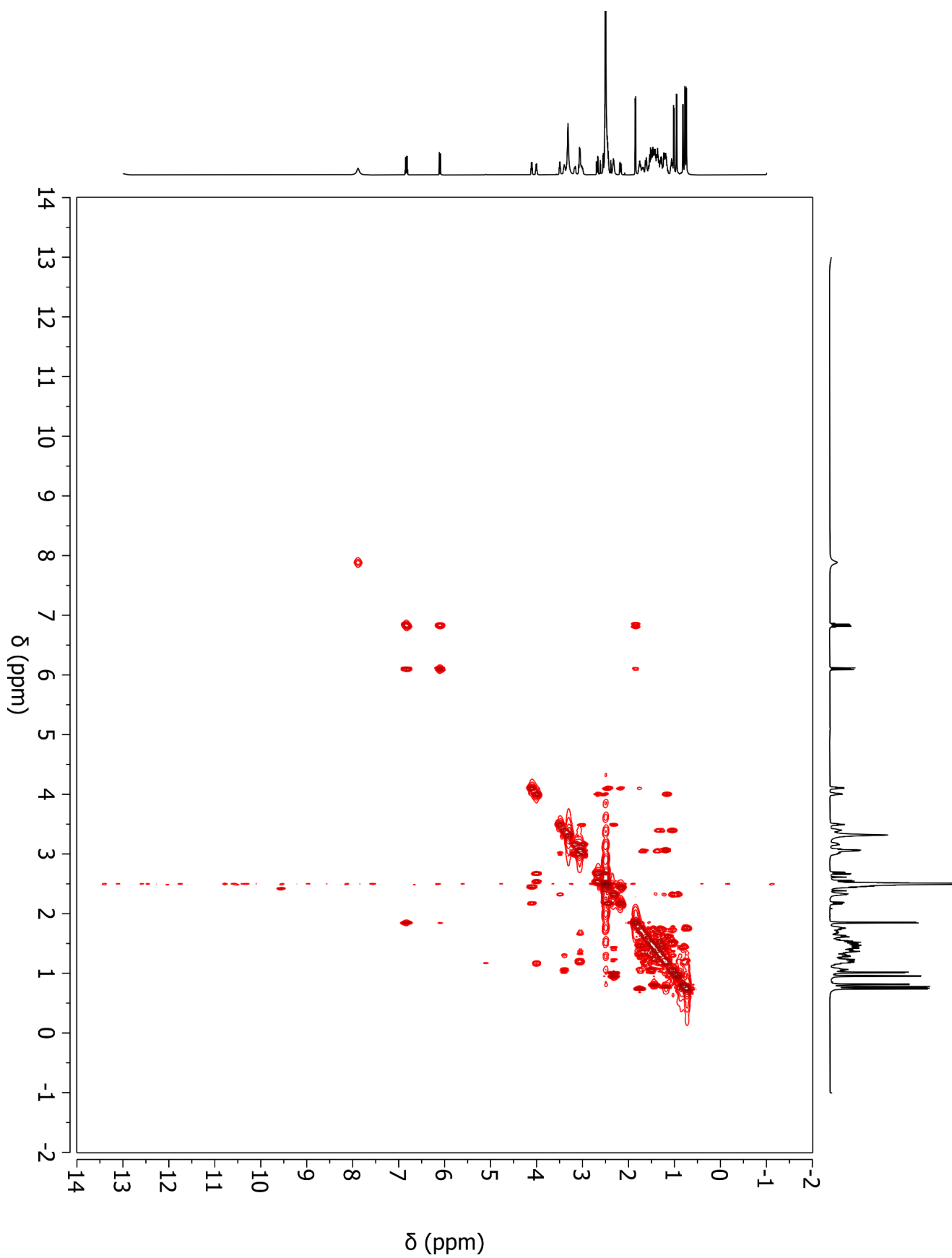
Figure S18: ^{1}H-^{1}H COSY of lagriamide B acquired in DMSO-d_6 at 600 MHz.
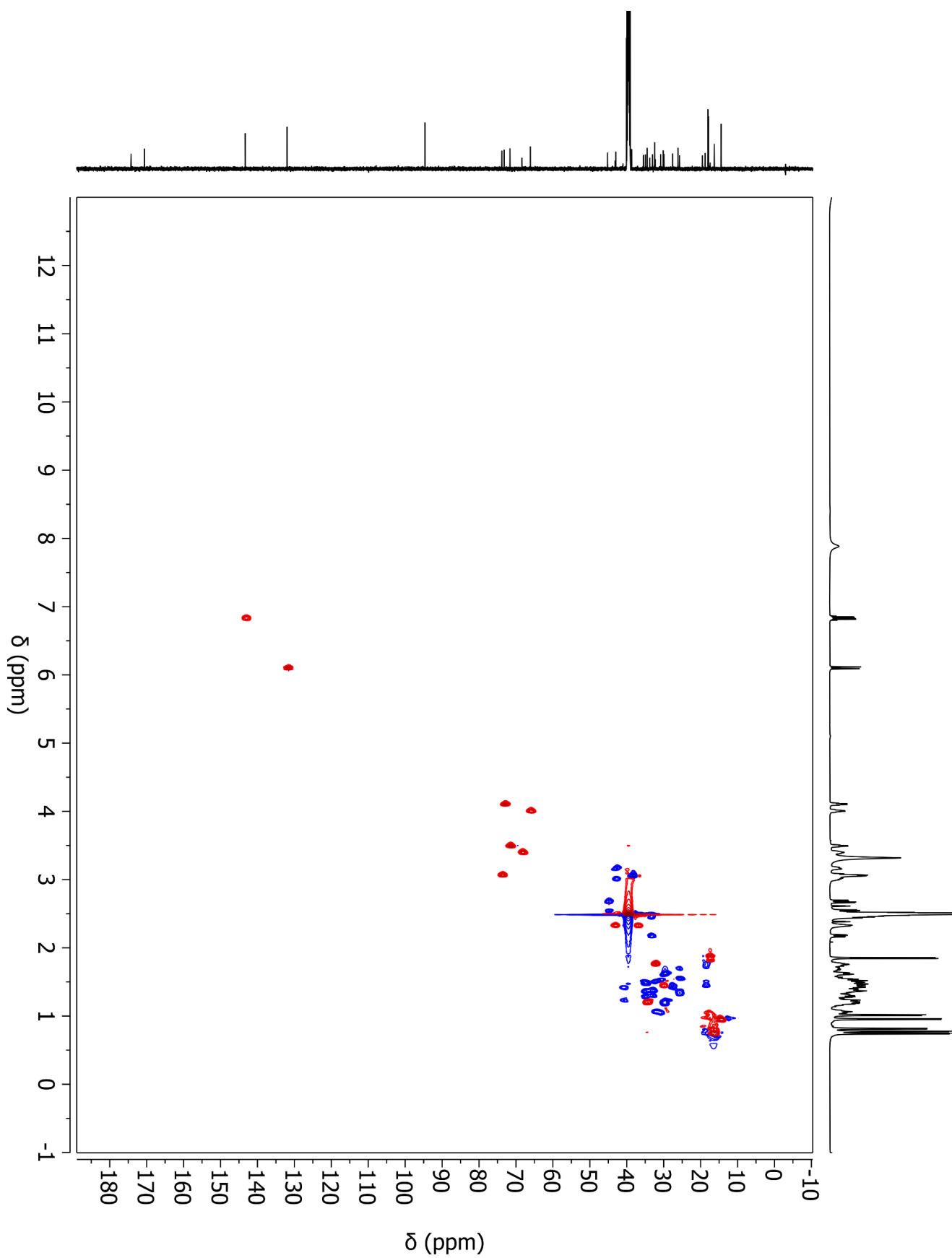
Figure S19: HSQC of lagriamide B in DMSO-*d₆* at 600 MHz
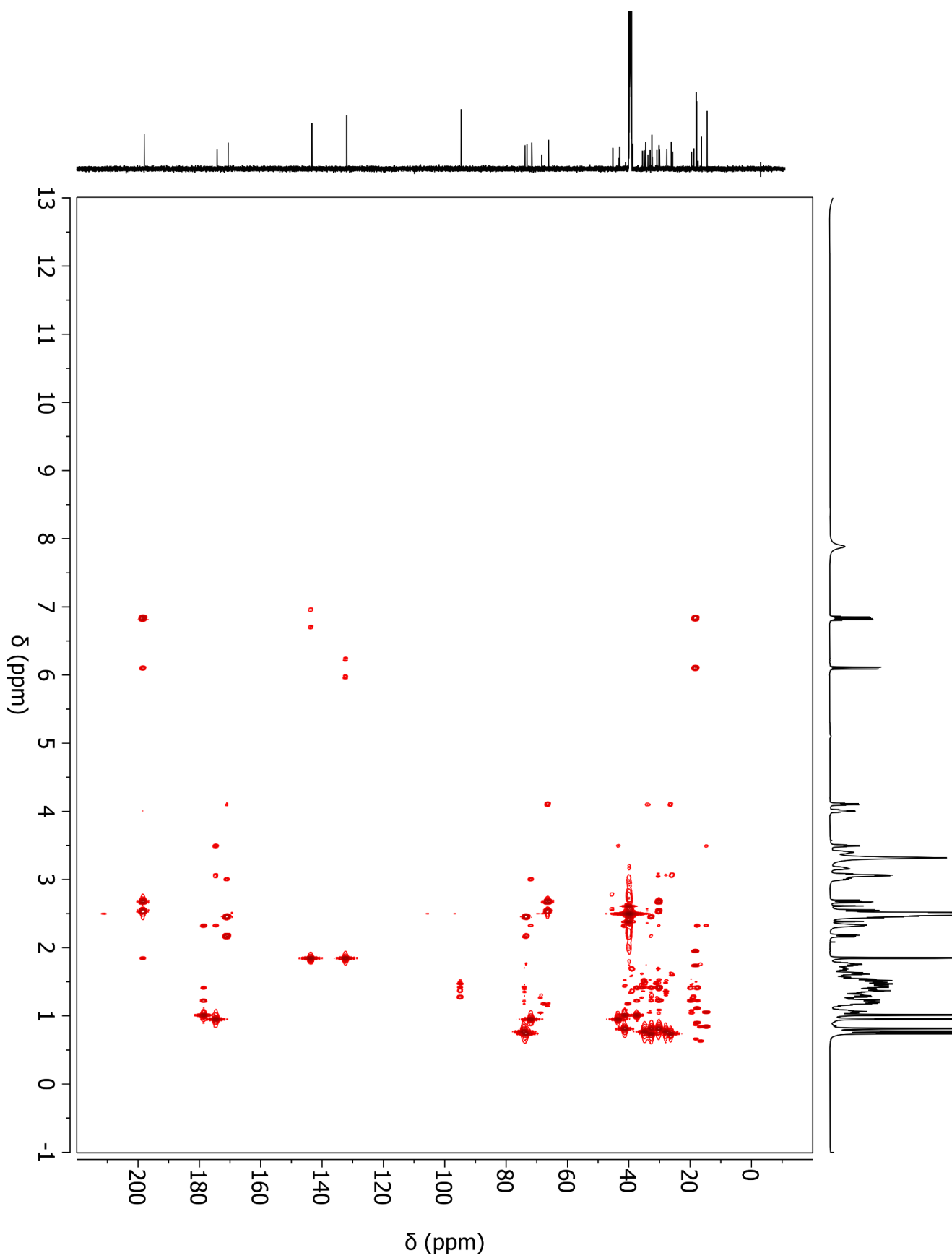
Figure S20: HMBC of lagriamide B in DMSO-d_6 at 600 MHz
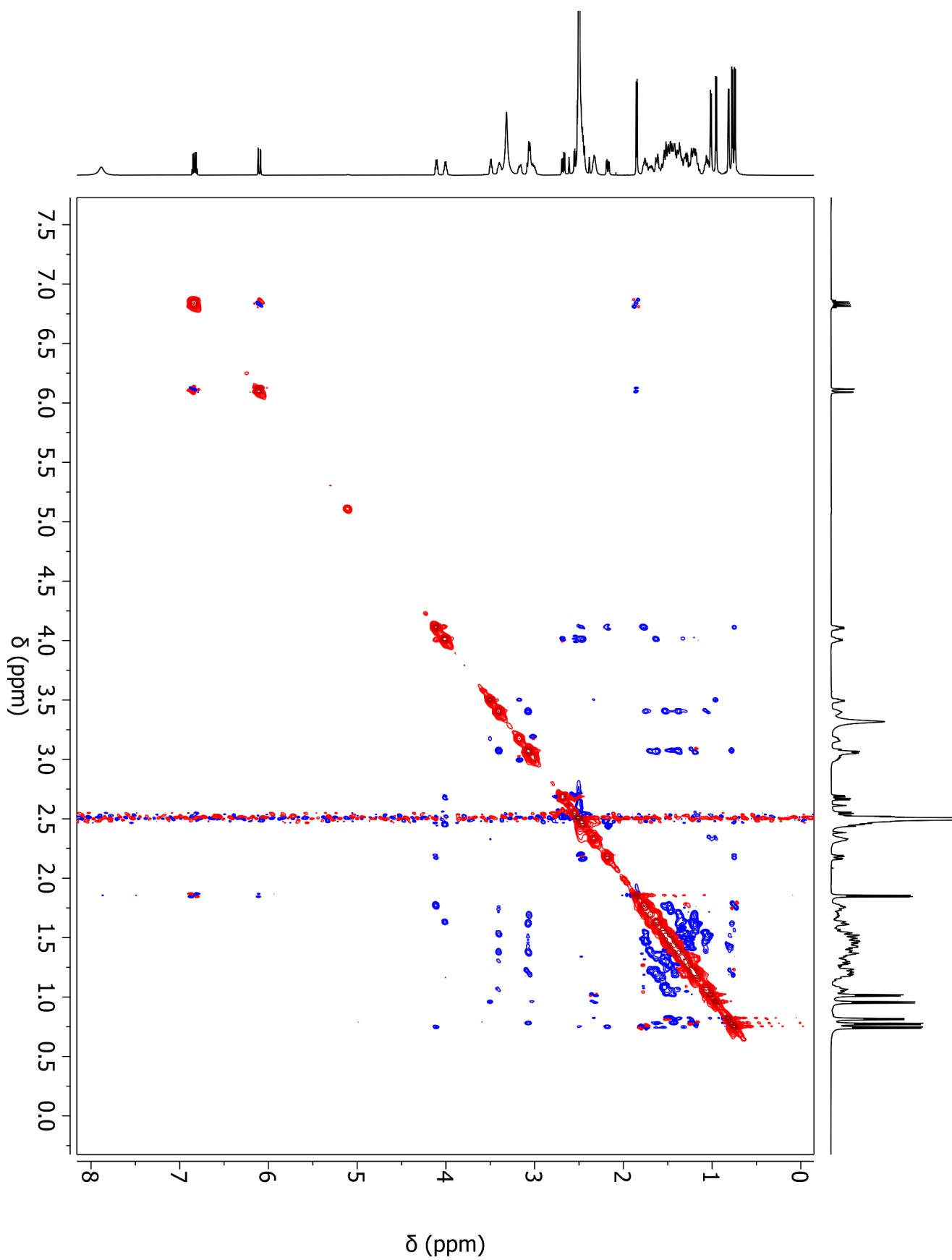
Figure S21: ROESY of lagriamide B in DMSO-d_6 at 600 MHz
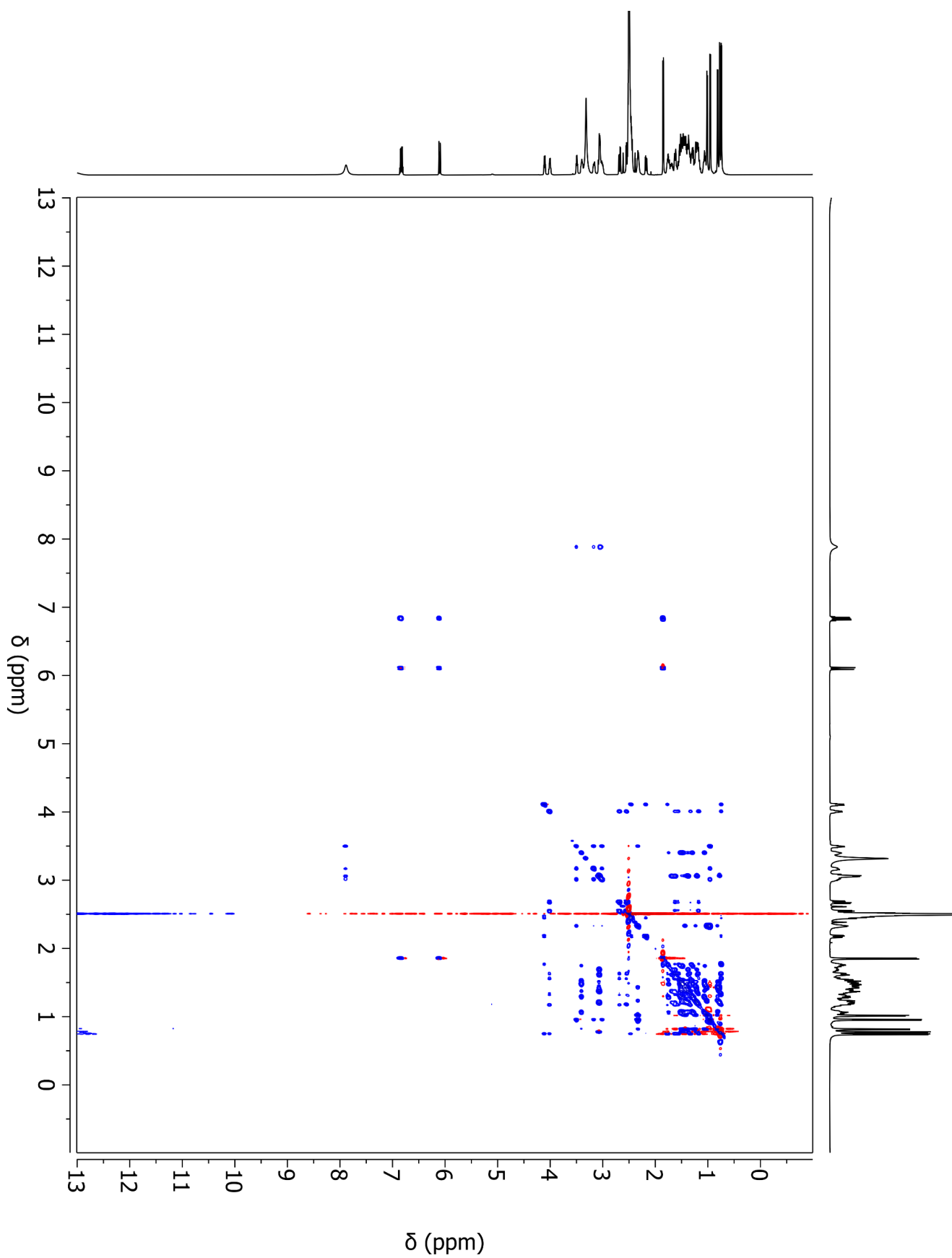
Figure S22: TOCSY of lagriamide B in DMSO-*d₆* at 600 MHz
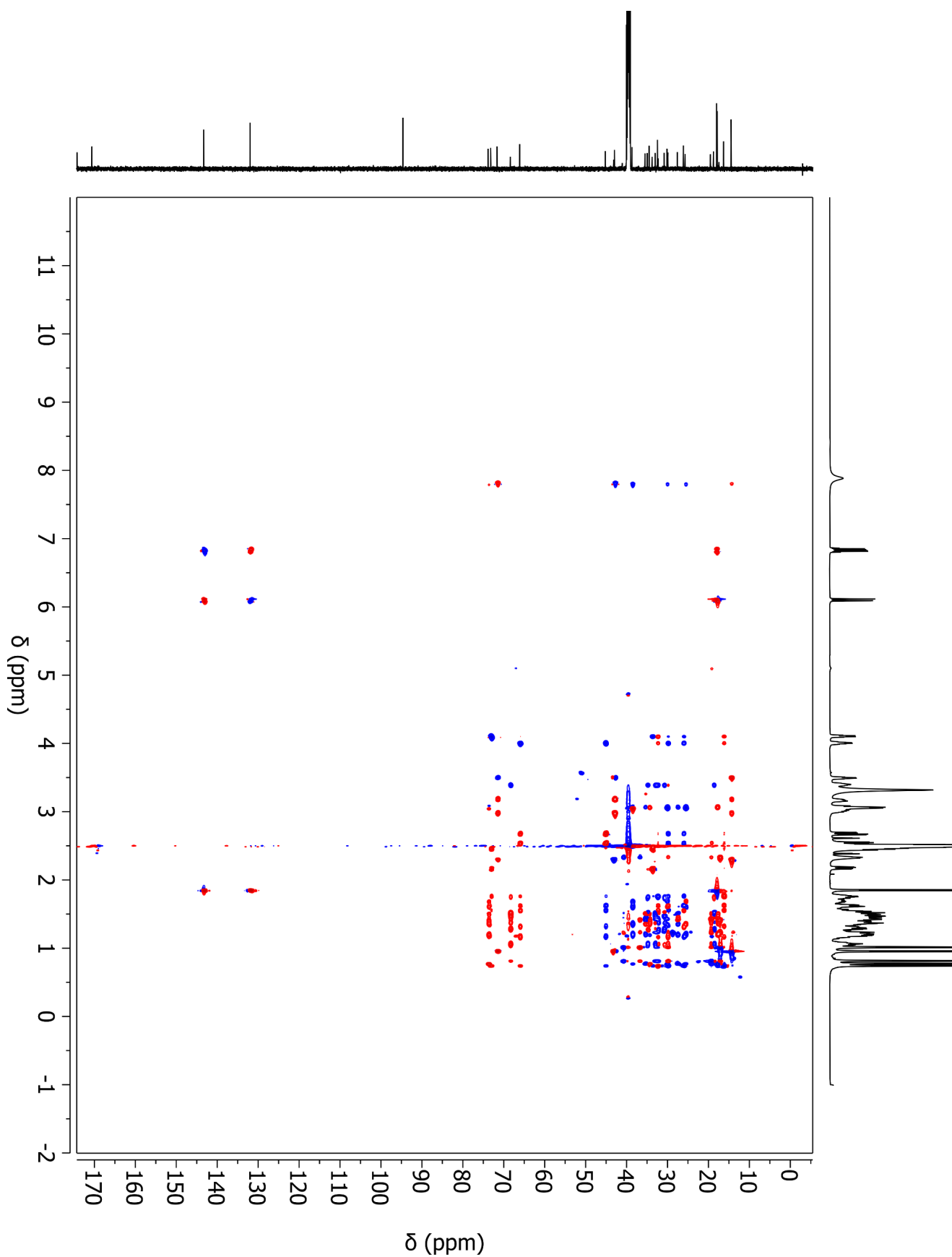
Figure S23: HSQC-TOCSY of lagriamide B in DMSO-d_6 at 600 MHz
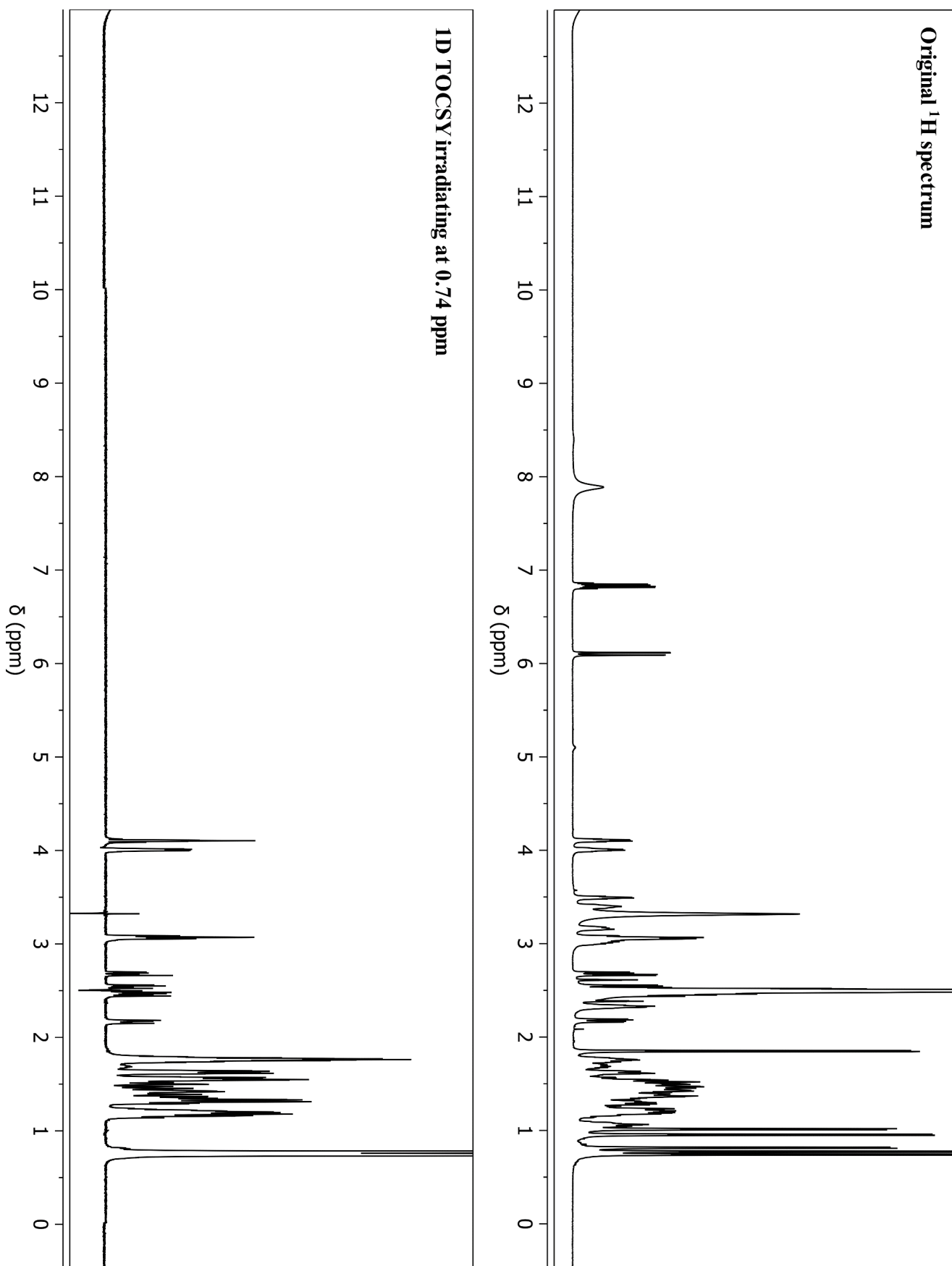
Figure S24: 1D Selective Gradient TOCSY of lagriamide B in DMSO-d_6 at 600 MHz, obtained by irradiating at 0.74 ppm
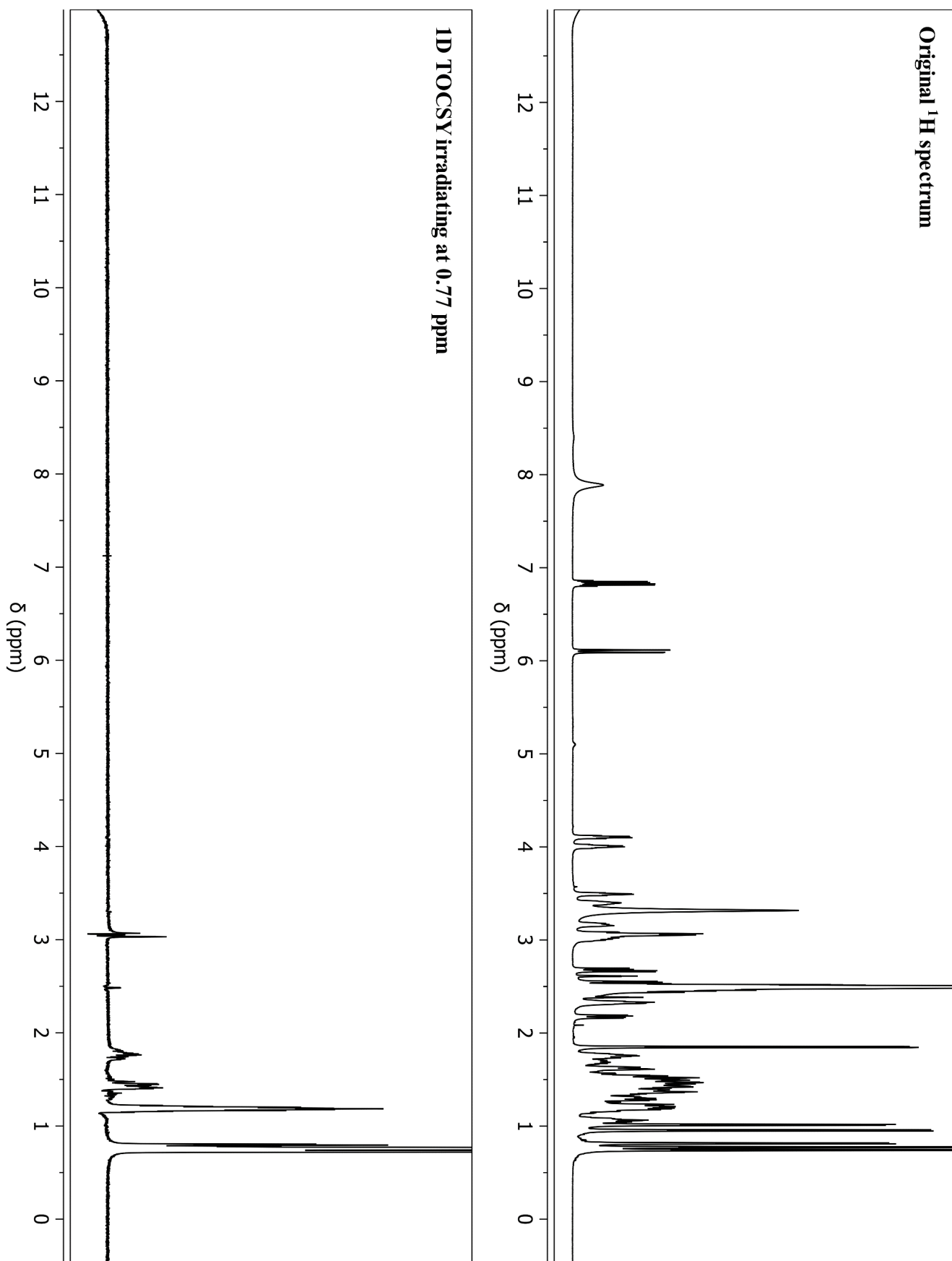
Figure S25: 1D Selective Gradient TOCSY of lagriamide B in DMSO-d6 at 600 MHz, obtained by irradiating at 0.77 ppm
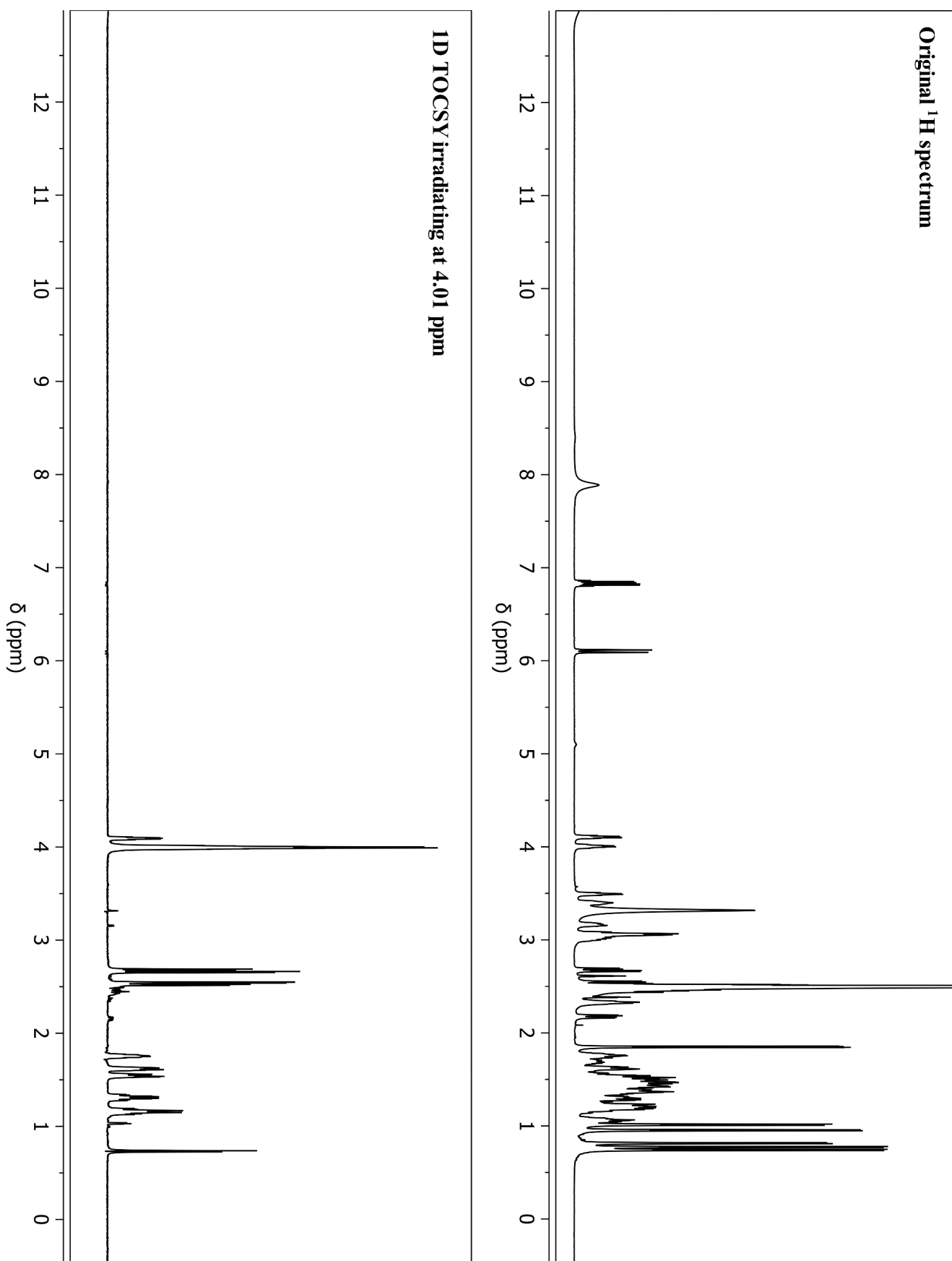
Figure S26: 1D Selective Gradient TOCSY of lagriamide B in DMSO-d_6 at 600 MHz, obtained by irradiating at 4.01 ppm
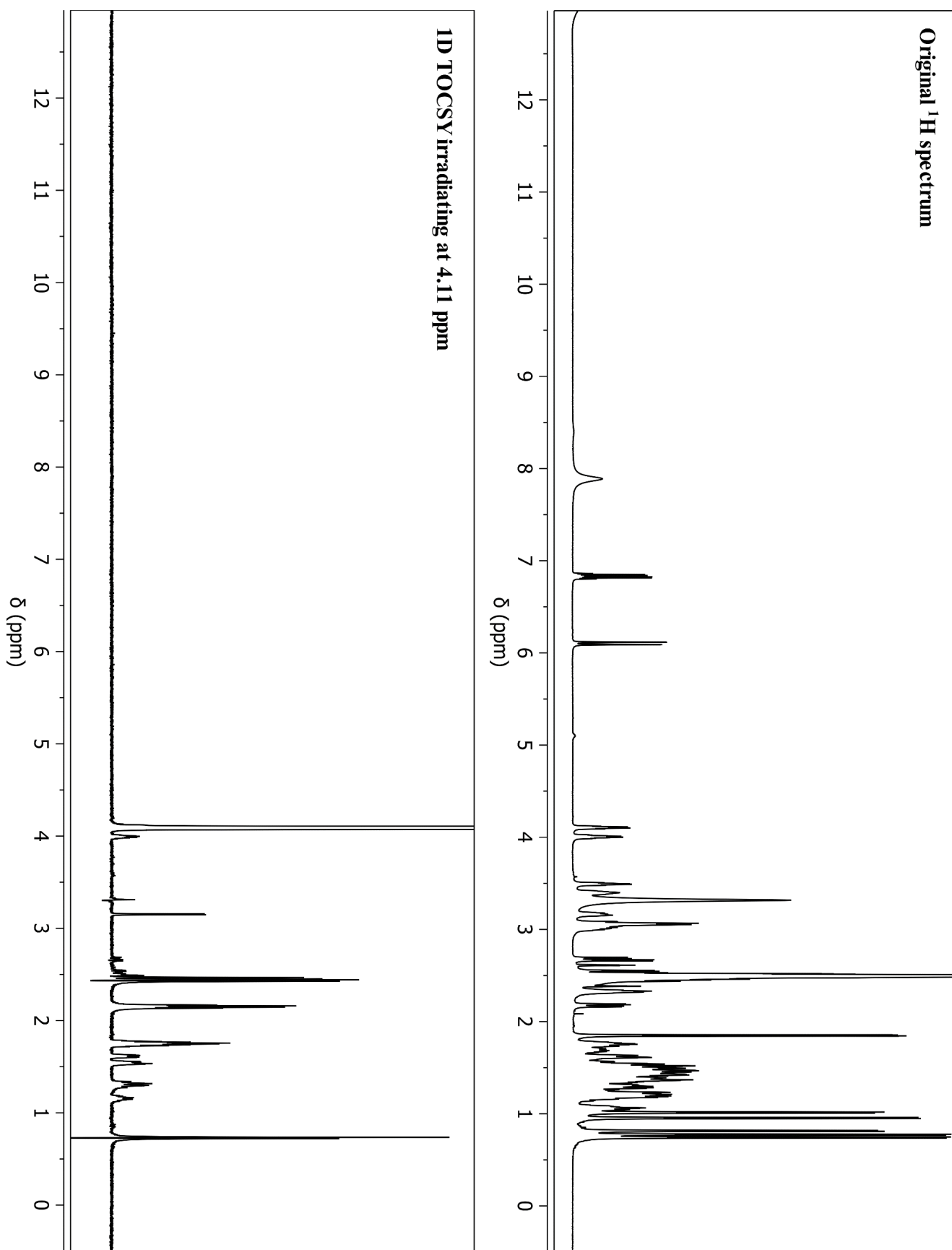
Figure S27: 1D Selective Gradient TOCSY of lagriamide B in DMSO-d_6 at 600 MHz, obtained by irradiating at 4.11 ppm
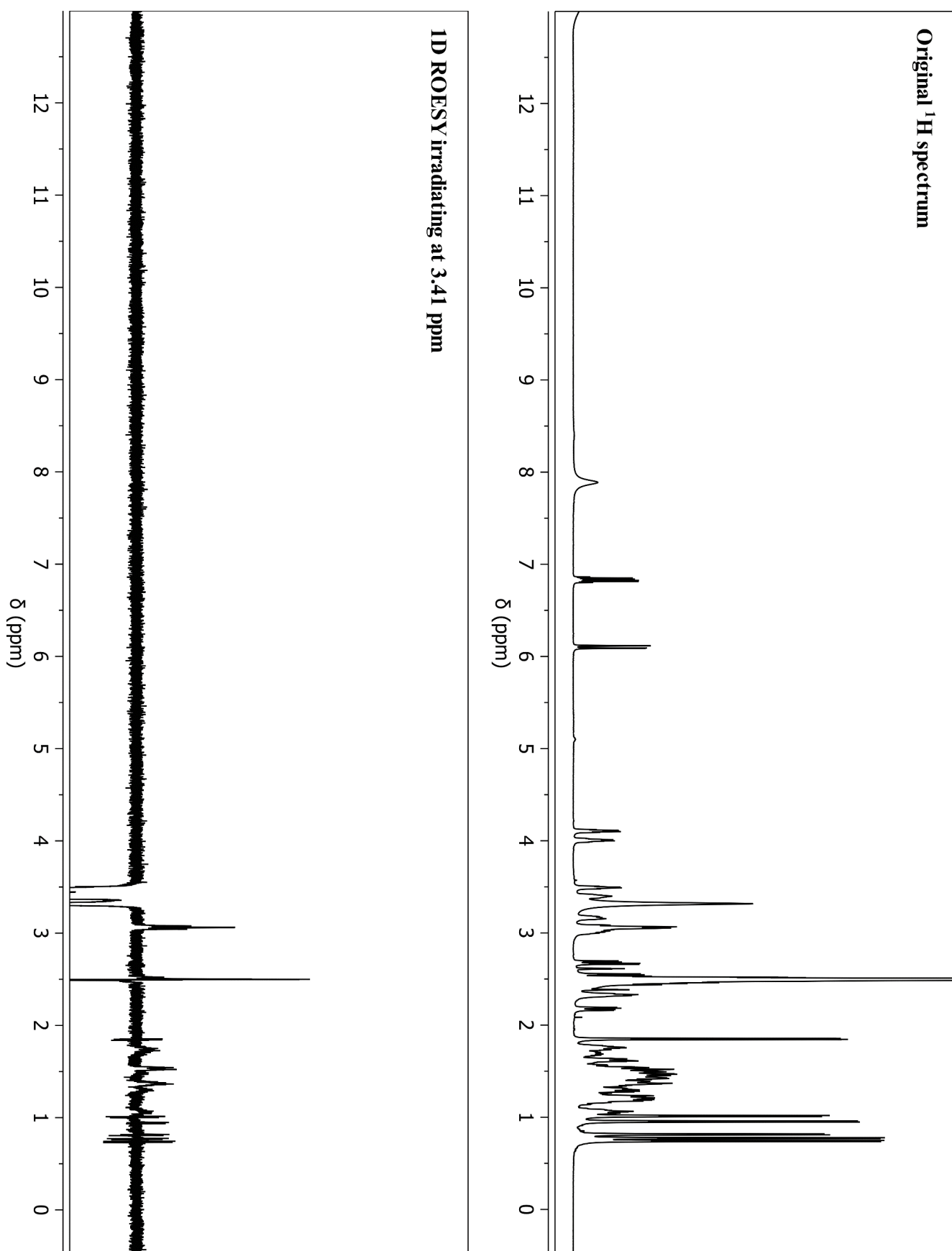
Figure S28: 1D Selective Gradient ROESY of lagriamide B in DMSO-d_6 at 600 MHz, obtained by irradiating at 3.41 ppm
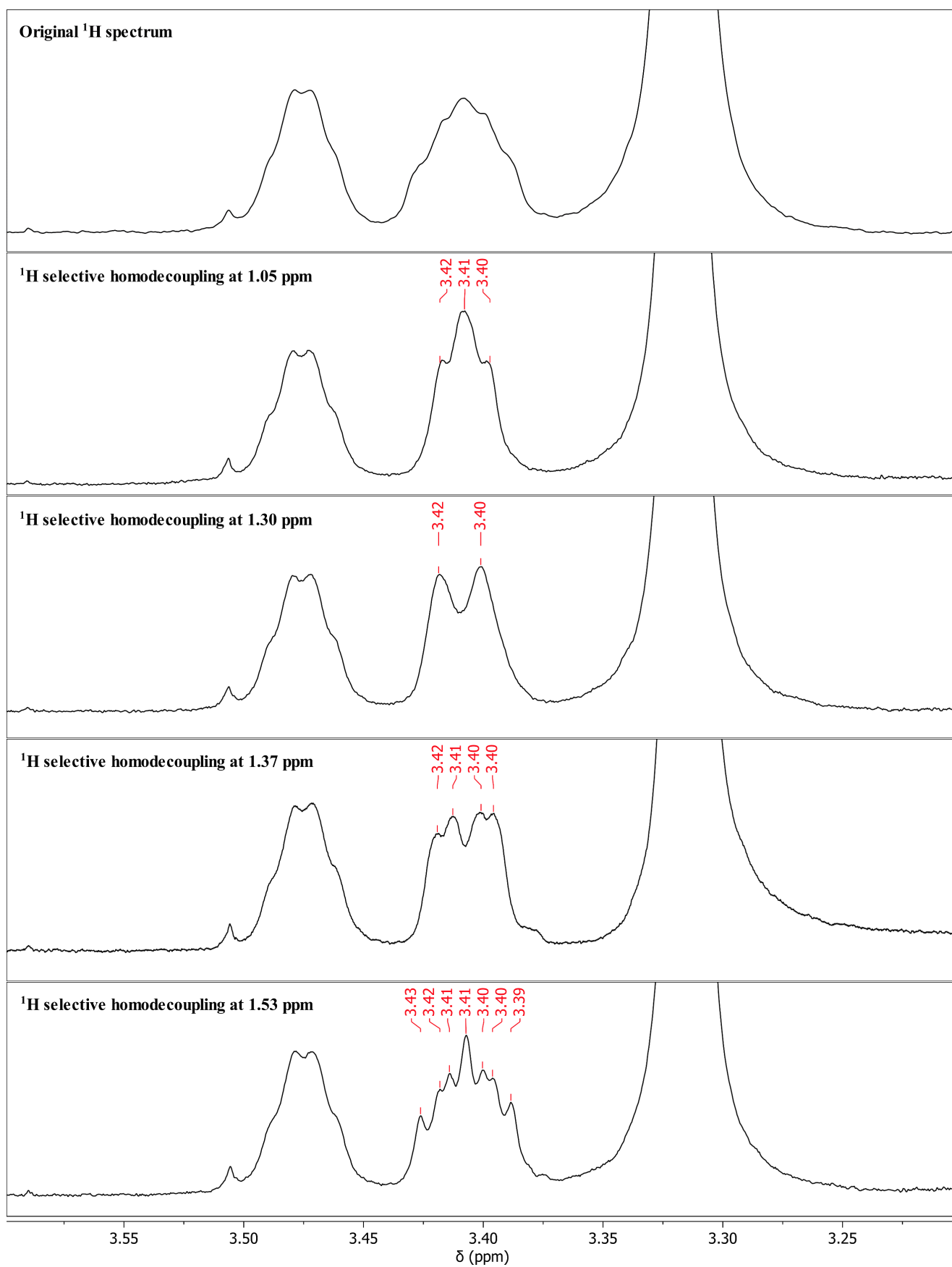
Figure S29: 1D-^{1}H selective homodecoupling spectra of lagriamide B in DMSO-d_6 at 600 MHz.

Bioactivity Screening Methods and Results

Antifungal screening of lagriamide B

Antifungal activity against *C. albicans* (ATCC 14053) and *S. cerevisiae* (ATCC 9763) was evaluated using a microbroth dilution antibiotic susceptibility assay modified from McCulloch et al[26]. Immediately prior to use, stock solutions of test compounds were prepared at the desired concentration in sterile-filtered dimethyl sulfoxide (DMSO, 40 µL) and diluted with either Difco Sabouraud Dextrose Broth (SDB; *C. albicans*, 960 µL) or Yeast Mold broth (YM; *S. cerevisiae*, 960 µL) (Becton Dickinson, Mississauga, Ontario). The resulting test solutions (100 µL; 4% DMSO) were transferred to the non-peripheral wells of a clear, non-tissue culture 96-well microtiter plate in triplicate (BD Falcon, Becton Dickinson, Mississauga, Ontario). Wells were then inoculated with suspensions of either *C. albicans* or *S. cerevisiae* (100 µL; 1 x 10^6 colony forming unit [CFU]/mL), to obtain a cell density of 5 x 10^5 CFU/mL. Sterile water (200 µL) was added to all perimeter wells to reduce evaporation from experimental wells. Each plate contained 3 negative control wells (4% DMSO in appropriate broth [100 µL] inoculated with appropriate fungi [100 µL; 1 x 10^6 CFU/mL]) and 3 untreated blank wells (2% DMSO in appropriate broth [200 µL]). Initial and final optical densities (OD) were determined for each well by recording absorbance at 600 nm immediately before and after incubation for 24 hours at 37°C using a Molecular Devices Emax microplate reader (Molecular Devices; Sunnyvale, CA, USA). Initial OD readings were subtracted from the final readings for each well to obtain the change in OD (ΔOD). ΔOD values were corrected for background absorbance of the media by subtracting the mean ΔOD readings of the blanks from the mean ΔOD readings of the control and test wells. The percentage inhibition of fungal growth is defined as:

$$[1 - (\text{mean test } \Delta OD/\text{mean negative control } \Delta OD)] \times 100$$

Antifungal activity against *A. niger* (DSM 737) and *P. lilacinum* (DSM 846) were assessed in a similar fashion. Fungal strains were inoculated onto fresh Potato Dextrose Broth (PDB) agar and incubated at 37°C for 3-5 days. Colonies were then covered with 5 mL of double distilled water with 0.1% tween 80 and carefully rubbed with a sterile loop. The suspension was transferred by pipet to a sterile tube and spores were counted using a haemocytometer chamber. The spore solution was adjusted to 2.5 x 10^5 CFU/mL. Samples were prepared in 96-well plates as described using Difco Potato Dextrose Broth and wells were inoculated with suspensions of either *A. niger* or *P. lilacinum*. Following a 24-hour incubation at 37°C, wells were examined under microscope to determine the MIC.

Table S10: Fungal target panel strains and growth conditions

Strain Name	Strain Designation	BSL	Growth Medium	Growth Condition
Aspergillus niger	DSM 737	1	PDB	37°C
Candida albicans	ATCC 14053	2	SDB	37°C
Purpureocillium lilacinum	DSM 846	2	PDB	37°C
Saccharomyces cerevisiae	ATCC 9763	1	YM	37°C

Antifungal activities for lagriamide B against fungal screening panel (MIC$_{50}$, µM)

Lagriamide B was screened for antifungal activity against two yeasts, *Candida albicans* and *Saccharomyces cerevisiae*, and two filamentous fungi, *Aspergillus niger* and *Purpureocillium lilacinum*. The natural product showed no inhibition against the yeasts or *P. lilacinum* at the top concentration (100 µM); however, it showed selective inhibition against *A. niger* as observed by the inhibition of spore germination at a test concentration of 12.5 µM.

Antimicrobial screening methods and data

Antimicrobial susceptibility testing for lagriamide B (**2**) against a 17-member bacterial target panel (**Table S11**) were performed using a miniaturized high-throughput assay adapted from the broth microdilution method outlined by the Clinical and Laboratory Standards Institute (CLSI). Bacterial test strains were individually grown on fresh Nutrient Broth (NB, ATCC Medium 3) agar, Tryptic Soy Broth (TSB, ATCC Medium 18) agar or Brain Heart Infusion (BHI, ATCC Medium 44) agar, respectively, as recommended by the American Type Culture Collection (ATCC) cultivation protocol. Individual colonies were used to inoculate 3 mL of sterile NB, TSB or BHI media and grown overnight with shaking (200 RPM; 37 °C). *Listeria ivanovii* (ATCC BAA-139) was incubated overnight but not shaken (37°C; 5% CO_2). Saturated overnight cultures were diluted in their respective media according to turbidity to achieve approximately 5×10^5 CFU/mL of final inoculum density and dispensed into sterile clear polystyrene 384-well microplates (Thermo Scientific™ 265202) with a final screening volume of 30 µL. *L. ivanovii* was diluted with and grown in Haemophilus Test Medium (HTM; ATCC Medium 2167). DMSO solutions of test compounds and antibiotic controls were prepared as 1:1 dilution series and pinned into each assay plate (200 nL) using a high-throughput pinning robot (Tecan Freedom EVO 100; V&P Scientific Pin-tool) to achieve final screening concentrations ranging from 128 µM to 3.91 nM per compound. In each 384-well plate; lane 1 was reserved for DMSO vehicle and culture medium; lane 2 reserved for DMSO vehicle, culture medium, and target bacteria; lanes 23 and 24 reserved for antibiotic controls, DMSO vehicle, culture medium, and target bacteria. After compound pinning, assay plates were read as t_0 at OD_{600} using an automated plate reader (BioTek Synergy Neo2), sealed with lid and placed in a humidity-controlled incubator at 37°C for 18-20 hours before optical density is obtained for t_{20}. *L. ivanovii* was incubated in a separate incubator (37°C; 5% CO_2). Resulting growth curves for each dilution series were used to determine the MIC values for all test compounds following standard procedures.

Table S11: Bacterial target panel strains and culture conditions.

Strain Name	Strain Designation	BSL	Growth Medium	Growth Condition
Gram-Positive				
Bacillus subtilis	ATCC 6051	1	NB	37°C
Enterococcus faecalis	ATCC 29212	2	BHI	37°C
Enterococcus faecium	ATCC 6569	2	BHI	37°C
Listeria ivanovii	BAA-139	1	BHI-A; HTM	37°C; 5% CO_2
Staphylococcus aureus (Methicillin-Resistant)	BAA-44	2	TSB	37°C
Staphylococcus aureus (Methicillin-Sensitive)	ATCC 29213	2	TSB	37°C
Staphylococcus epidermidis	ATCC 14990	1	TSB	37°C
Gram-Negative				
Escherichia coli	K-12 MG1655	1	NB	37°C
Klebsiella aerogenes	ATCC 35029	1	NB	37°C
Klebsiella pneumoniae	ATCC 700603	2	NB	37°C
Ochrobactrum anthropi	ATCC 49687	1	TSB	37°C
Providencia alcalifaciens	ATCC 9886	1	TSB	37°C
Pseudomonas aeruginosa	ATCC 27853	2	TSB	37°C
Salmonella enterica	ATCC 13311	2	NB	37°C
Shigella sonnei	ATCC 25931	2	NB	37°C
Vibrio cholerae	A1552 El Tor	2	TSB	37°C
Yersinia pseudotuberculosis	ATCC 6904	2	BHI	37°C

Table S12: Antibacterial activities of lagriamide B.

Minimum Inhibitory Concentrations (MIC; µM)					
Strain	Lagriamide B (**2**)	Ampicillin	Azithromycin	Norfloxacin	Rifamycin
Gram-Positive					
B. subtilis	>128[*]	0.063	1	1	0.13
E. faecalis	>128	4	4	8	2
E. faecium	>128	4	8	2	8
L. ivanovii	>128	0.5	2	32	N/A[**]
S. aureus (Methicillin-Resistant)	>128	>4	>64	>32	2
S. aureus (Methicillin-Sensitive)	>128	2	1	4	0.063
S. epidermidis	>128				
Gram-Negative					
E. coli	>128	4	8	0.25	8
K. aerogenes	>128	>4	8	0.5	8
K. pneumoniae	>128	>4	32	16	>8
O. anthropi	>128	>4	0.5	4	4
P. alcalifaciens	>128	>4	64	0.25	>8
P. aeruginosa	>128	>4	32	2	>8
S. enterica	>128	0.5	8	0.25	4
S. sonnei	>128	4	4	0.25	2
V. cholerae	>128	>4	0.5	0.25	>8
Y. pseudotuberculosis	>128	32	8	0.5	8

[*] MIC value not observed below the highest tested concentration.
[**] Not tested.

Morphological analysis of lagriamide B mechanism

Compound plate preparation

Lagriamide B was solubilized using DMSO at a stock concentration of 19.2 mM. 40 µL of the resulting stock solution were transferred to a 384-well polypropylene compound storage plate (Thermo Scientific NUNC 264573) alongside other samples and control compounds. A dilution series of lagriamide B was then prepared by pipetting 20 µL of the stock solution to 20 µL of DMSO vehicle on the subsequent row. This process was repeated 14 times to dilute the original 19.2 mM solution by half for every row to ~ 8 µM in row P; the excess volume was discarded. Columns 1 and 24 were reserved for DMSO vehicle. Columns 2 and 23 were reserved for positive control two-fold dilution series of latrunculin B (A2: 25.6 mM to P2: ~ 8 µM) and monasterol (A23: 25.6 mM to P23: ~ 8 µM).

U-2 OS cell culturing

Human osteosarcoma epithelial cells (U-2 OS; ATCC HTB-96) were passaged in McCoy's 5A (ATCC 30-2007) medium supplemented with 10% FBS (ATCC 30-2020) and 1% Pen/Strep (ATCC 30-2300) at 37ºC, 5% CO_2. Upon 80% confluence, cells were trypsinized (ATCC 30-2101) following standard procedure and cell concentration determined using a BioRad TC20 automated cell counter. Trypsinized cells were then seeded into 384-well microscopy assay plates (Corning

4681) at a volume of 40 µL, ~ 800 cells/well using a BioTek EL406 automated plate washer/dispenser, before incubation at 37ºC, 5% CO_2 overnight.

Cell Painting high-content imaging

Test compounds were transferred from the compound plate to their corresponding overnight assay plates at a volume of 200 nL/well using a Tecan Freedom EVO 100 high-throughput liquid handling robot with attached V&P Scientific pin-tool (384 pins). Pinned assays plates then were placed back into the incubator at 37ºC, 5% CO_2 overnight.

Following overnight incubation, all assay plates were treated, stained, and imaged according to the protocol outlined by the PerkinElmer PhenoVue Cell Painting kit (PING11). In brief, 30 µL of medium were aspirated from each well in the assay plate and replaced with a staining solution containing PhenoVue 641 Mitochondrial Stain (500 nM final concentration) and PhenoVue Dye Diluent A (HBSS + 1% BSA). Assay plates were then incubated at 37ºC, 5% CO_2 for 30 min in the dark, before fixation using 16% PFA (Fisher Scientific 50980488; 3.2% final concentration). Fixed assay plates were incubated at RT for 20 min, before wash with HBSS and permeabilization using a 0.1% Triton X-100 (Sigma-Aldrich T8787-250ML) solution, followed by another 20 min incubation at RT. After a second HBSS wash, 30 µL of a second staining solution was added to each well containing the following dyes at their respective final concentrations: PhenoVue Fluor 555-WGA (43.7 nM), PhenoVue Fluor 488-Concanavalin A (960 nM), PhenoVue Fluor 568-Phalloidin (33 nM), PhenoVue Hoechst 33342 Nuclear stain (8.12 µM), PhenoVue 512 Nucleic acid stain (3 µM), and PhenoVue Dye Diluent A (HBSS + 1% BSA). After a final incubation period of 30 min at RT, assay plates were lidded with adhesive aluminum foil (Fisher Scientific 12-56-398) prior to imaging via a Molecular Devices ImageXpress micro XLS automated microscope (20X objective; channel: DAPI, FITC, TRITC, Cy5, and Texas Red).

Data normalization, analysis, and feature selection

Cell painting images were processed using a modified protocol adapted from Bray et al.[27] and proceeded as follows. Cell segmentation and feature extraction via CellProfiler[28] produced 2,090 raw features on the basis of cell count and fluorescence intensity, granularity, texture, area, shape, and overlap for each of the five channels. Raw feature values for each sample were normalized to the DMSO vehicle control using the robust histogram-difference method (HistDiff)[29]. In order to identify potential modes of action from sample compounds, a reference library containing 4,400 compounds with known biological function (TargetMol, L4000) were screened in cell painting at 50 µM with the corresponding image data processed and normalized as described above, then combined with the sample compound data. The CP score[30] was calculated as the square root of the sum of the squares of all 2,090 feature values for each sample and reference library compound. All sample compounds and reference library compounds with CP scores at or below the DMSO vehicle controls were removed, resulting in 1,265 samples with non-zero feature profiles. Feature selection was performed via the Fast Correlation-Based Filtering (FCBF) algorithm on Orange[31] to identify features with maximal correlation distribution while removing redundant features, resulting in a reduced subset of 435 features. The trimmed dataset, containing a combined 1,265 sample and reference library compounds multiplied by 435 unique features (CP fingerprints) were then hierarchically clustered using Spearman correlations and Ward linkage parameters to group like fingerprints together. Hierarchical clustering trees and heatmaps were visualized using Orange.
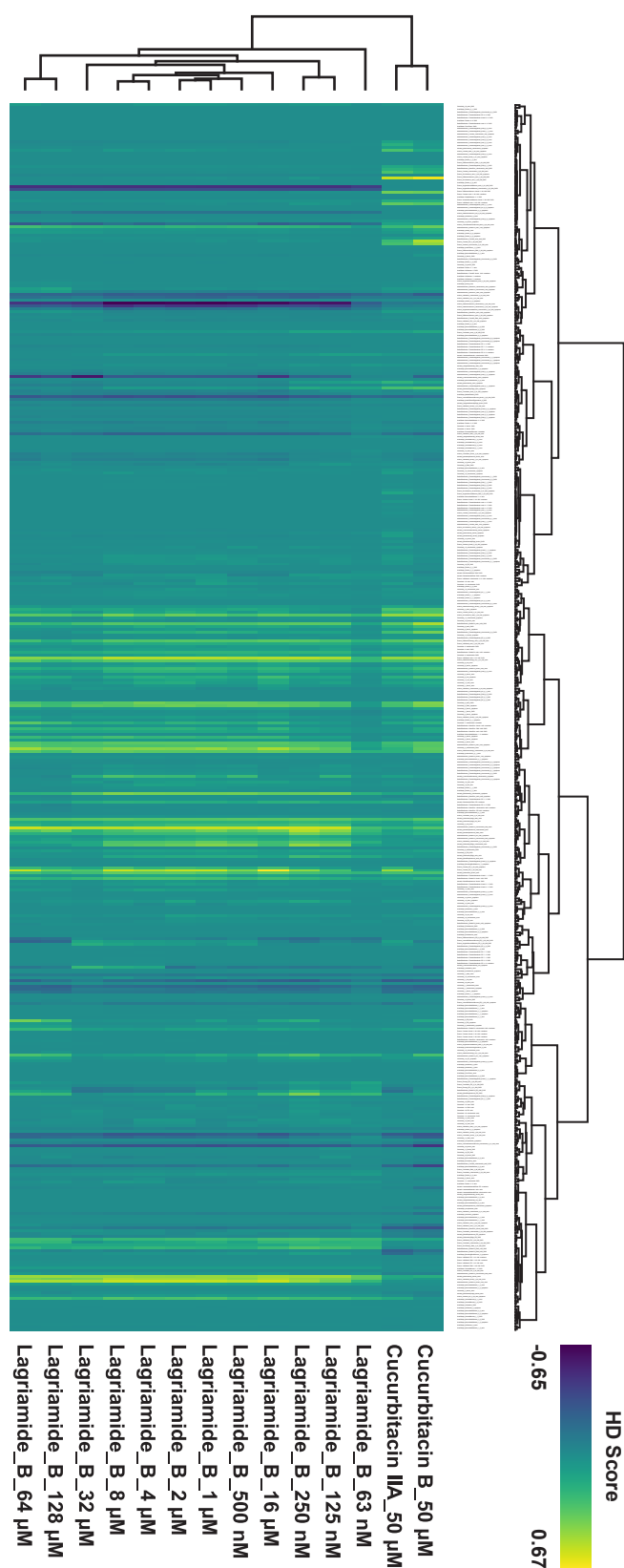
Figure S30: Hierarchical clustering heat map of Cell Painting features for lagriamide B assayed as a two-fold dilution series. Horizontal axis shows hierarchical clustering dendrogram between lagriamide B and TargetMol standards. Vertical axis shows hierarchical clustering dendrogram of a reduced subset of 435 CellProfiler features. Yellow shading indicates increasing HistDiff values compared to vehicle control while deep blue shading indicates decreasing HistDiff values compared to vehicle control.

Supplementary references

1. C. H. Fergusson, J. M. F. Coloma, M. C. Valentine, F. P. J. Haeckl and R. G. Linington, *Appl. Environ. Microbiol.*, 2020, **86**, e00354-20.
2. R. R. Wick, L. M. Judd, C. L. Gorrie and K. E. Holt, *PLoS Comput. Biol.*, 2017, **13**, e1005595.
3. A. Gurevich, V. Saveliev, N. Vyahhi and G. Tesler, *Bioinformatics*, 2013, **29**, 1072–1075.
4. T. Seemann, *Bioinformatics*, 2014, **30**, 2068–2069.
5. S.-H. Yoon, S.-M. Ha, S. Kwon, J. Lim, Y. Kim, H. Seo and J. Chun, *Int. J. Syst. Evol. Microbiol.*, 2017, **67**, 1613–1617.
6. R. Bashudev and G. Radhey S, *Int. J. Syst. Evol. Microbiol.*, 2021, **71**, 005011.
7. K. Blin, S. Shaw, Z. Charlop-Powers, G. P. van Wezel, M. H. Medema and T. Weber, *Nucleic Acids Res.*, 2021, **49**, W29–W35.
8. L. V. Flórez, K. Scherlach, I. J. Miller, A. Rodrigues, J. C. Kwan, C. Hertweck and M. Kaltenpoth, *Nat. Commun.*, 2018, **9**, 2478.
9. C. S. McCaughey, J. A. van Santen, J. J. J. van der Hooft, M. H. Medema and R. G. Linington, *Nat. Chem. Biol.*, 2022, **18**, 295–304.
10. P. Caffrey, *ChemBioChem*, 2003, **4**, 654–657.
11. R. Reid, M. Piagentini, E. Rodriguez, G. Ashley, N. Viswanathan, J. Carney, D. V. Santi, C. R. Hutchinson and R. McDaniel, *Biochemistry*, 2003, **42**, 72–79.
12. D. Menche, F. Arikan, O. Perlova, N. Horstmann, W. Ahlbrecht, S. C. Wenzel, R. Jansen, H. Irschik and R. Müller, *J. Am. Chem. Soc.*, 2008, **130**, 14234–14243.
13. E. J. N. Helfrich and J. Piel, *Nat. Prod. Rep.*, 2016, **33**, 231–316.
14. J. C. Kwan, M. S. Donia, A. W. Han, E. Hirose, M. G. Haygood and E. W. Schmidt, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 20655–20660.
15. A. T. Keatinge-Clay, *Nat. Prod. Rep.*, 2012, **29**, 1050–1073.
16. C. R. Valenzano, Y.-O. You, A. Garg, A. Keatinge-Clay, C. Khosla and D. E. Cane, *J. Am. Chem. Soc.*, 2010, **132**, 14697–14699.
17. E. J. N. Helfrich, R. Ueoka, A. Dolev, M. Rust, R. A. Meoded, A. Bhushan, G. Califano, R. Costa, M. Gugger, C. Steinbeck, P. Moreno and J. Piel, *Nat. Chem. Biol.*, 2019, **15**, 813–821.
18. RDKit: Open-Source Cheminformatics Software, www.rdkit.org.
19. T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 490–519.
20. P. Tosco, T. Balle and F. Shiri, *J. Mol. Model.*, 2011, **17**, 3021–3023.
21. J. W. Ponder and F. M. Richards, *J. Comput. Chem.*, 1987, **8**, 1016–1024.
22. J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardère, M. J. Schnieders, J.-P. Piquemal, P. Ren and J. W. Ponder, *J. Chem. Theory Comput.*, 2018, **14**, 5273–5289.
23. Gaussian 09, Revision E.01. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian Inc., Wallingford CT, 2016.
24. K. Ermanis, K. E. B. Parkes, T. Agback and J. M. Goodman, *Org. Biomol. Chem.*, 2019, **17**, 5886–5890.
25. K. Ermanis, K. E. B. Parkes, T. Agback and J. M. Goodman, *Org. Biomol. Chem.*, 2016, **14**, 3943–3949.

26   M. W. B. McCulloch, F. Berrue, B. Haltli and R. G. Kerr, *J. Nat. Prod.*, 2011, **74**, 2250–2256.

27   M.-A. Bray, S. Singh, H. Han, C. T. Davis, B. Borgeson, C. Hartland, M. Kost-Alimova, S. M. Gustafsdottir, C. C. Gibson and A. E. Carpenter, *Nat. Protoc.*, 2016, **11**, 1757–1774.

28   C. McQuin, A. Goodman, V. Chernyshev, L. Kamentsky, B. A. Cimini, K. W. Karhohs, M. Doan, L. Ding, S. M. Rafelski, D. Thirstrup, W. Wiegraebe, S. Singh, T. Becker, J. C. Caicedo and A. E. Carpenter, *PLoS Biol.*, 2018, **16**, e2005970.

29   M. H. Woehrmann, W. M. Bray, J. K. Durbin, S. C. Nisam, A. K. Michael, E. Glassey, J. M. Stuart and R. S. Lokey, *Mol. Biosyst.*, 2013, **9**, 2604–2617.

30   S. K. Hight, T. N. Clark, K. L. Kurita, E. A. McMillan, W. Bray, A. F. Shaikh, A. Khadilkar, F. P. J. Haeckl, F. Carnevale-Neto, S. La, A. Lohith, R. M. Vaden, J. Lee, S. Wei, R. S. Lokey, M. A. White, R. G. Linington and J. B. MacMillan, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, **119**, e2208458119.

31   J. Demšar, T. Curk, A. Erjavec, Č. Gorup, T. Hočevar, M. Milutinovič, M. Možina, M. Polajnar, M. Toplak, A. Starič, M. Štajdohar, L. Umek, L. Žagar, J. Žbontar, M. Žitnik, B. Zupan, *J. Mach. Learn. Res.*, 2013, **14**, 2349–2353.