Universiteit Leiden
The Netherlands

Expanding the chemical space of antibiotics produced by Paenibacillus and Streptomyces

Machushynets, N.V.

Chapter 3

Phylogeny and antibiotic producing potential of plant-associated *Paenibacillus*

Nataliia V. Machushynets, Somayah S. Elsayed, Chao Du, Vladyslav Lysenko, Mercedes de la Cruz, Olga Genilloud, Nathaniel I. Martin, Mark R. Liles, and Gilles P. van Wezel

Abstract

Antimicrobial resistance (AMR) is a pressing global issue, demanding the urgent identification of novel compounds to combat multidrug-resistant pathogens. Microbial natural products (NPs) are a rich source of potential therapeutic agents. Members of the genus *Paenibacillus* are renowned for their antibiotic-producing potential against diverse pathogens. Here, we employ Global Natural Products Social molecular networking (GNPS MN) combined with Mass Spec Query Language (MassQL) analysis to probe the chemical diversity of specialized metabolites from plant-associated *Paenibacillus*. By applying MassQL queries to tandem mass spectrometry data within GNPS MN, we enhance the molecular network with substructure information, facilitating the early prioritization and targeted isolation of bioactive natural products. Through the analysis of mass fragmentation patterns across 227 plant-associated *Paenibacillus* isolates, we discovered new bacitracin congeners, designated as paenitracins, with potent activity against Gram-positive pathogens, including vancomycin-resistant *Enterococcus faecium* E155. This study underscores the efficacy of our integrative approach in expediting the discovery of antimicrobial agents, offering promising solutions for addressing the challenge of antimicrobial resistance.

Introduction

In the 21st century, antimicrobial resistance (AMR) has emerged as a significant worldwide issue. Therefore, discovering new lead compounds is necessary to control outbreaks of multidrug-resistant human pathogens (WHO, 2017). Bacterial natural products (NPs) represent a rich reservoir of antibacterial compounds, many of which have been successfully translated into clinical applications (Newman & Cragg, 2020).

Firmicutes are well known for their potential to produce a wide range of antimicrobials, in particular, nonribosomally synthesized peptides (NRPs) (Aleti *et al.*, 2015, Cochrane & Vederas, 2016, Olishevska *et al.*, 2019). Numerous bioactive NRPs produced by *Bacillus* and *Paenibacillus* have been identified, including octapeptins, paenibacterins, polypeptins, tridecaptins, etc (Cochrane & Vederas, 2016, Olishevska *et al.*, 2019, Pilz *et al.*, 2023). Notably, certain NRPs, such as polymyxins and bacitracins, derived from these microorganisms, have found application in clinical settings (Falagas & Kasiakou, 2005, Johnson *et al.*, 1945). A large fraction of these natural products contain positively charged amino acids and possess potent antimicrobial activities against Gram-positive and Gram-negative pathogens (Li *et al.*, 2018, Cardoso *et al.*, 2021). The mode of action of cationic lipopeptides, like polymyxins, involves binding to the lipid A component of lipopolysaccharides (LPSs) in the outer membrane of Gram-negative bacteria. The positively charged Dab residues of polymyxins electrostatically interact with the negatively charged phosphate groups of the lipid A, facilitating binding and leading to membrane disruption and increased permeability (Mohapatra *et al.*, 2021). Other NRPs, like bacitracin, act as inhibitors of cell wall biosynthesis. Specifically, in the presence of a divalent metal ion, most commonly Zn^{2+}, bacitracin binds to and sequesters the membrane-associated phospholipid undecaprenyl pyrophosphate ($C_{55}PP$) (Stone & Strominger, 1971, Siewert & Strominger, 1967).

One of the challenges in NP discovery is the rapid identification of previously reported compounds, termed as structural dereplication (Hubert *et al.*, 2017). Liquid chromatography coupled with high-resolution mass spectrometry (LC-MS) is widely used at the early stages of dereplication studies. Molecular formulas of the detected molecules are determined through the exact mass using the Seven Golden Rules, Sirius 2, and MSFINDER software; and subsequently queried in various NP databases like DNP, UNPD, ChemSpider, and REAXYS; to annotate the molecular structures (Gaudêncio *et al.*, 2023)2023.

In order to further identify compounds in LC-MS-based metabolite profiling workflows, the introduction of the Global Natural Products Social Molecular Networking (GNPS MN) has appeared as a powerful tool to process tandem mass spectrometry (MS) data and perform molecular networking (MN) (Wang *et al.*, 2016). The MN concept is based on organizing and visualizing tandem MS data through a spectral similarity map, revealing the presence of homologous MS^2 fragmentations. As structurally related compounds share similar

fragmentation spectra, their nodes tend to gather together and create clusters of analogs (Nothias *et al.*, 2020). The known compounds are annotated by matching their MS^2 spectra against an available spectral library. Moreover, combination of MN with annotation methods, such as MS2LDA (van der Hooft *et al.*, 2017), NAP (da Silva *et al.*, 2018), or DEREPLICATOR (Mohimani *et al.*, 2017) accelerates the identification of unannotated ions (de Jonge *et al.*, 2022). Mass Spec Query Language (MassQL) has been developed based on advancements in molecular networking to explore underutilized MS/MS data further. MassQL captures the unique characteristics of MS data, including isotopic patterns, diagnostic fragmentation, and neutral loss; and establishes a comprehensive MS terminology for searching MS patterns across datasets. Additionally, NP prioritization strategies have been applied based on the combination of molecular networking with different sources of information (e.g., genomic, taxonomic, geographic, bioactivity, and/or spectral data) (Olivon *et al.*, 2017, Nothias *et al.*, 2018, Olivon *et al.*, 2020, Crüsemann *et al.*, 2017). Despite the ability of current tools to dereplicate NPs upstream of the isolation process, effective strategies for the early detection of potentially bioactive NPs within crude extracts are still lacking.

In the present study, we adopt an MN approach to provide insight into the chemical structures and diversity of *Paenibacillus* NRPs and facilitate the discovery of new natural products. We introduce an early-stage prioritization method that combines feature-based MN (FBMN) and MassQL search to target basic amino acid-containing peptides with potential antimicrobial properties. Basic amino acid signatures were detected and mapped onto a massive multi-informative molecular network for an extract library of 227 *Paenibacillus* plant-associated isolates. This led to the discovery of new bacitracin congeners, designated as paenitracins, with potent bioactivity against Gram-positive pathogens, including vancomycin-resistant *E. faecium* E155. Present study demonstrates that combining FBMN and MassQL is an effective strategy to accelerate the decoding of mass fragmentation patterns for the discovery of new bioactive natural products.

Results and Discussion

Taxonomic diversity of the strain collection

Considering the great pharmacological and biotechnological potential of the metabolites produced by *Paenibacillus*, in this work, we pursued the prospection of 227 *Paenibacillus* isolates from the Auburn University (USA) Plant-Associated Microbial strain collection. *Paenibacillus* spp. have been isolated from the root surface of a field-grown corn plant (*Z. mays*) and soybean (*G. max*) grown in Dunbar (Nebraska), Carrol (Iowa), Whitewater (Wisconsin), Sparta (Illinois), Troy (Ohio) and Tallassee (Alabama) (Table S1). Taxonomic affiliation of the isolated strains was conducted by comparing 16S rRNA gene sequences from the isolates with *Paenibacillus* type strains. To gain insights into the relatedness of the *Paenibacillus* isolates to known bacterial species, a maximum-likelihood tree was constructed based on 16S rRNA gene sequences (Figure 1).

Exploring the chemical diversity of specialized metabolites produced by *Paenibacillus*

In this study, we aimed to discover new bioactive NRPs and investigate the chemical diversity of the specialized metabolites produced by *Paenibacillus* isolates. For this purpose, 227 isolates were cultured on TSA media in 96 deep-well plates for 72 h and extracted with isopropyl alcohol (IPA) supplemented with 0.1% (v/v) formic acid (FA). Subsequently, crude extracts were tested against Gram-positive (*Staphylococcus aureus* ATCC 29213) and Gram-negative (*Escherichia coli* ATCC 25922) indicator strains to assess the antibacterial potential of the strain collection. The results of the antibacterial assays showed that nine crude extracts exhibited activity against *S. aureus*, while 22 crude extracts showed activity against *E. coli* (Figure 1, Table S1). These extracts were then subjected to LC-MS/MS analysis to identify the natural products accountable for the observed bioactivities. LC-MS/MS data were processed with MzMine 2 (Pluskal *et al.*, 2010) and exported for GNPS FBMN. Prior to uploading the data to GNPS, media-derived mass features were eliminated. A mass exclusion threshold of 300 *m/z* was implemented to minimize the occurrence of mass features originating from small metabolites that are not relevant to our study. FBMN analysis resulted in the generating molecular network consisting of 3096 parent ions (nodes) connected through 3692 edges (Figure 2). Characteristics of the natural products, such as annotation, *m/z* value and species-specific molecule production were visualized using Cytoscape 3.9.1 (Shannon *et al.*, 2003).

Figure 1. Phylogenetic tree (16S rRNA gene) obtained by neighbor-joining analysis of plant-associated *Paenibacillus* isolates and their closely related type strains (Tables S1, S2). The tree was rooted using *Bacillus subtilis* IAM 12118 as an outgroup. The results of the antibacterial screenings were mapped onto the tree. The tree was generated using Geneious v2024.0.3.

Mass spectrometry-based molecular networking allows clustering of molecules with similar MS/MS fragmentation patterns, which stem from the similarity in their structures (Wang *et al.*, 2016). The resulting clusters allow us to visually explore the metabolites produced by a given strain(s), allowing rapid dereplication of known compounds and visualization of their unknown structural analogs (Nothias *et al.*, 2020). Initially, we aimed to dereplicate the known metabolites by matching the experimentally derived MS/MS fragmentation patterns with the annotated and curated MS/MS spectra within the GNPS database. However, with

the spectral library search, we did not get any accurate hits for NRPs from *Paenibacillus*. Therefore, we created an in-house database of lipopeptides from *Paenibacillus* and used it to annotate the known metabolites based on the exact mass and MS/MS spectra of the detected features. We were able to annotate the lipopeptides belonging to the families of fusaricidins, polymyxins, and tridecaptins (Table S3, Figure 2, Figures S1-S3). These dereplicated mass features were used to propagate the annotation to other connected mass features. Despite the identification of a substantial number of putative congeners of known lipopeptides in the network, the majority of the produced metabolites remained unidentified.

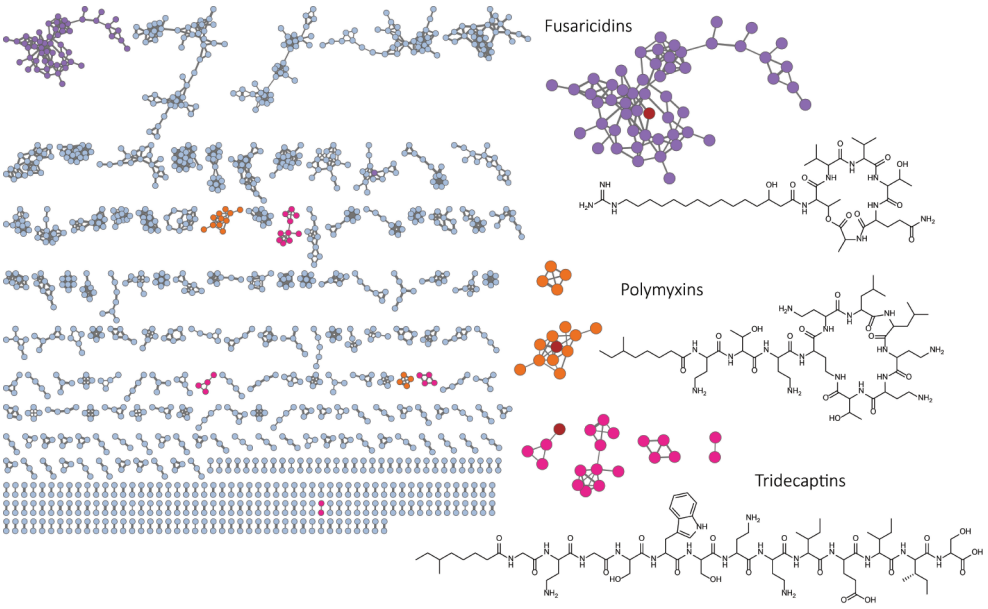

Figure 2. Molecular network of the mass features detected in the bacterial extracts from a collection of 227 plant-associated *Paenibacillus* isolates. Each node represents a parent ion detected in one or more strains. The annotated molecular families have been amplified. The colors of the nodes highlight molecular families that have at least one node annotated by an in-house database search (red node).

Targeting known cationic lipopeptides using MassQL-assisted fragment search

Several detected classes of lipopeptides exhibited potent bioactivities that are often associated with the overall positive charge of the molecules (Table S3). The positive charge enhances electrostatic interactions with the negatively charged components on the surface of bacterial cells, such as the cell membrane or lipopolysaccharides (LPSs). This leads to membrane disruption, permeabilization, and, ultimately, destruction of bacterial cells (Straus & Hancock, 2006). Therefore, we sought to detect the lipopeptides with potential bioactive properties that contained one or multiple positively charged amino acid residues. To accomplish this, a feature-based spectral similarity network was paired with a MassQL (Jarmusch *et al.*, 2022)

search query to target the MS/MS spectra that contained the most common positively charged amino acids. The fragment search of *b* ions of 2,4-diaminobuturic acid (Dab) (*m/z* 101.0709), ornithine (Orn) (*m/z* 115.0871), lysine (Lys) (*m/z* 129.1027), arginine (Arg) (*m/z* 157.1089) and the immonium ion of histidine (His) (*m/z* 110.0718) was conducted with the accuracy of *m/z* 0.005 and at minimum intensity of 5% (Figure 3).

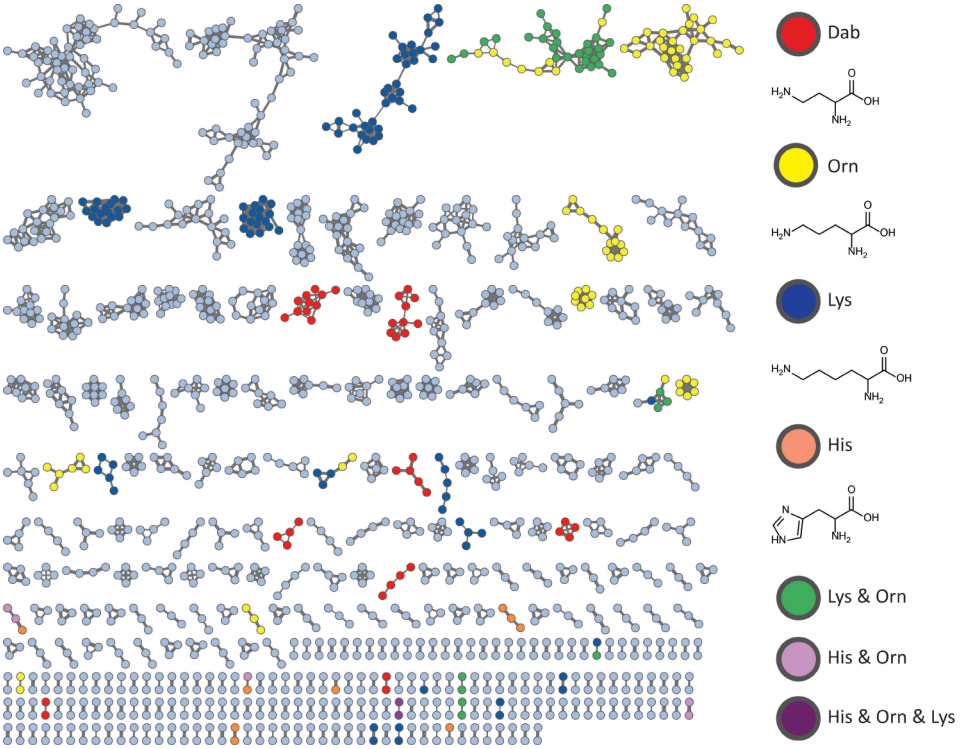

Figure 3. MassQL-annotated molecular network of the ions detected in the bacterial extracts from the collection of 227 plant-associated *Paenibacillus* isolates. The colors of the nodes highlight parent ions that contain basic amino acid residues in the structures.

MassQL search of Dab-containing parent ions highlighted previously dereplicated polymyxin and tridecaptin spectral families (Figure 4). The identification of different polymyxin (Vogler & Studer, 1966, Jian Li *et al.*, 2006) and tridecaptin (Bann *et al.*, 2021) congeners within numerous *Paenibacillus* spp. extracts underscored the chemical diversity of our strain collection. Despite the structural heterogeneity of polymyxins and tridecaptins, a shared characteristic is the occurrence of multiple Dab amino acid residues.

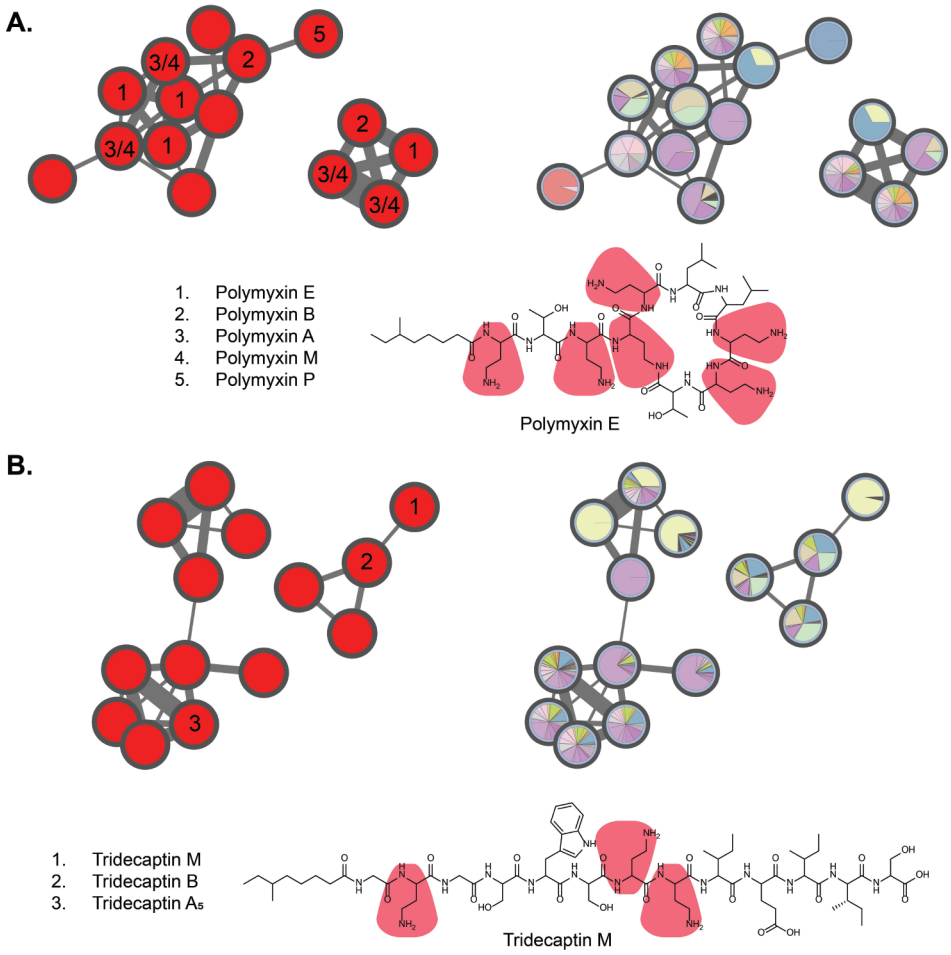

Figure 4. Molecular families of polymyxins (A) and tridecaptins (B) were highlighted with MassQL search as Dab-containing lipopeptides. On the same molecular family, pie charts were mapped to the nodes to represent the relative precursor ion intensities within each extract. Note that both polymyxins and tridecaptins contain multiple Dab amino acid residues in their structures.

Furthermore, MassQL search revealed a spectral family of compounds that contained two distinct amino acid residues, namely ornithine and lysine (Figure 5A). This cluster represented triply-charged ions with a mass range of 1500–1650 Da. Interestingly, the compounds from this cluster were produced exclusively by two isolates from the entire strain collection, namely *Paenibacillus* sp. JJ-602 and *Paenibacillus* sp. JJ-1115. These compounds could not be dereplicated by spectral library and in-house database search. Extensive literature search and manual MS/MS dereplication resulted in the annotation of these compounds as paenibacterin A congeners (Li *et al.*, 2018) (Figure 5B).

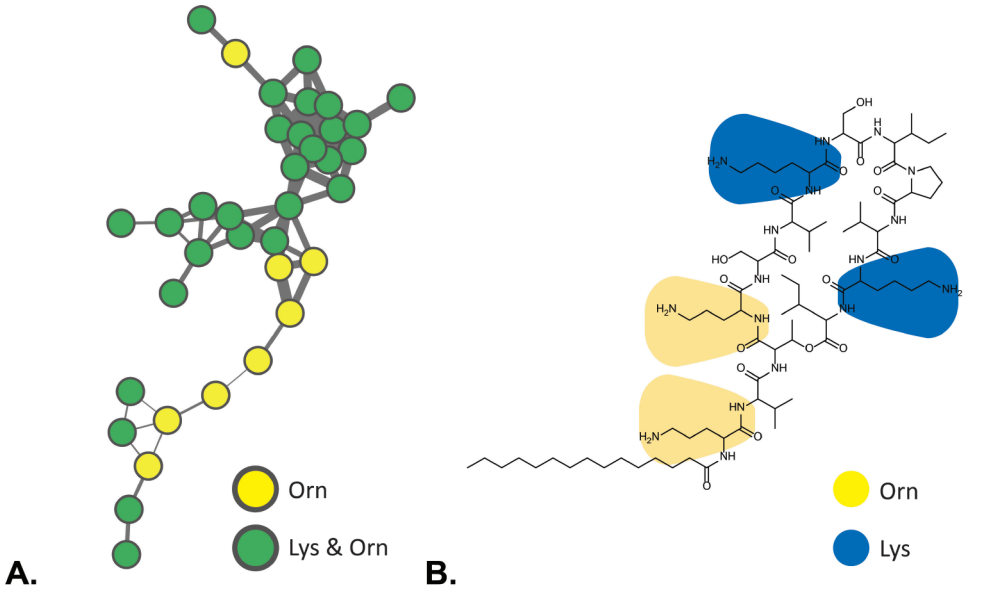

Figure 5. A. MassQL-annotated molecular family of paenibacterins. B. Chemical structure of paenibacterin A.

Multiple positively charged lysine and ornithine amino acid residues contribute to the overall charge of paenibacterins and enhance interactions with negatively charged LPS from the outer membrane of Gram-negative bacteria (Knolhoff *et al.*, 2015). Altogether, these results demonstrated the power of MassQL-assisted search of positively charged amino acids for the detection of bioactive cationic lipopeptides.

Molecular family of Lys-containing compounds contained new paenilipoheptin lipopeptides

In addition to the annotated lipopeptide classes, we identified multiple unknown molecular families, that contained compounds with basic amino acids such as lysine, ornithine, and histidine (Figure 3). We have focused on a molecular family of Lys-containing compounds, the production of which was specific to a limited number of isolates (Figure S4). The Lys-containing molecular family could be distinguished into three sub-families (Figure 6A).

Sub-family 1 represents doubly-charged ions with a mass range of 1080 to 1150 Da. A literature search revealed similarities between these compounds and the recently discovered lipopeptide paenilipoheptin A, produced by *P. polymyxa* E681 (Vater *et al.*, 2018). One of the nodes has an *m/z* 562.3133, which corresponds to the exact mass of paenilipoheptin. However, the MS/MS fragmentation pattern indicated differences in the amino acid sequences. The discrepancy lies in amino acid residues 2 and 4, which were annotated in the article as Dab and Val, while we identified them as Lys and Ala, respectively. In our high-resolution MS/MS

spectra, we did not observe fragment ions corresponding to Dab or Val, while we detected a clear ion corresponding to a Lys residue (*m/z* 129.1021), together with the fragment ions Ala-Trp (*m/z* 258.1224) and Ala-Phe (*m/z* 219.1122) (Figure S5). Further analysis of the *b* ions of the parent mass with *m/z* 562.3133 corroborated our finding and yielded the lipopeptide sequence FA($C_{13}H_{26}$)–Ser–Lys–Trp–Ala–Phe–Tyr–Glu (Figure S5). It is worth noting that low-resolution masses were used to assign the structure of paenilipoheptin A (Vater *et al.*, 2018), which might result in misassignments. Other nodes in sub-family 1 had mass differences relative to *m/z* 562.3133 that indicate different fatty acid chain lengths.

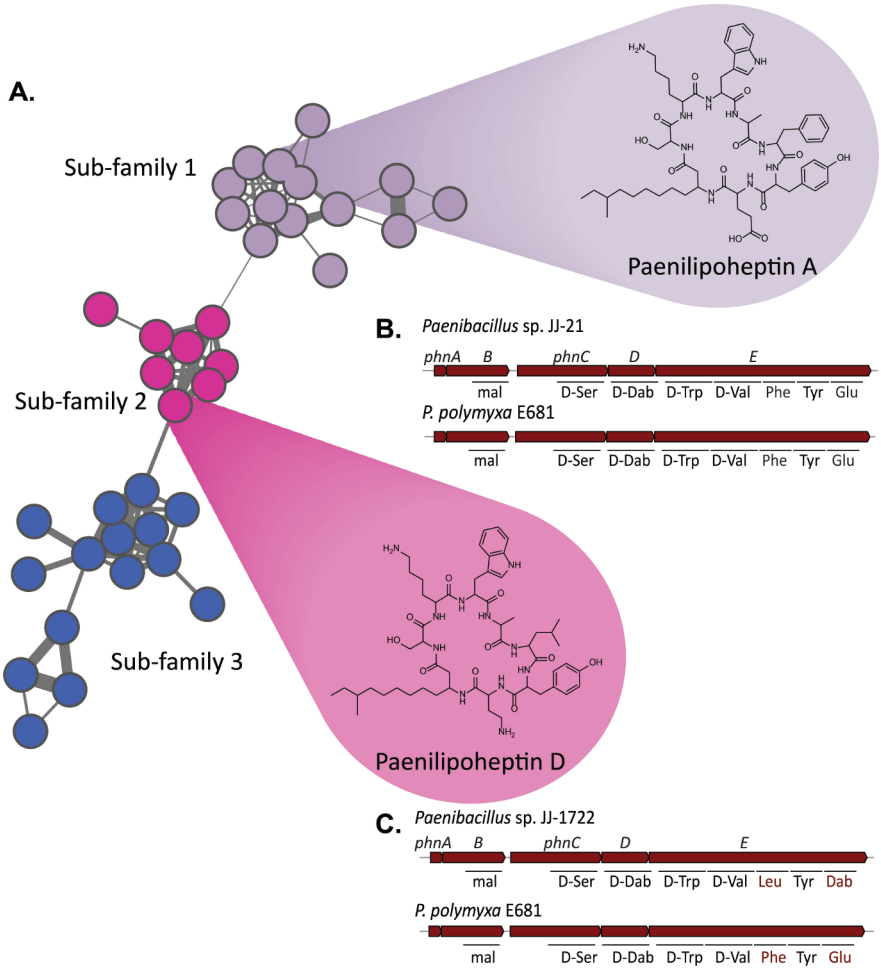

Figure 6. Molecular family of the Lys-containing specialized metabolites highlighted by MassQL search. Exploring the mass features from sub-family 1 and 2 led to the discovery of paenilipoheptin A and D, respectively (**A**). The BGCs of paenilipoheptin A from the genome of *Paenibacillus* sp. JJ-21 (**B**) and paenilipoheptin D from the genome of *Paenibacillus* sp. JJ-1722 (**C**) were compared to the reference BGC from *P. polymyxa* E681. Note that the NRPS assembly line of paenilipoheptin D differs from the one of the paenilipoheptin A in modules 5 and 7 (highlighted in red).

The node with *m/z* 555.305 in sub-family 1 of the Lys-containing molecular family was linked to sub-family 2 via the node with *m/z* 523.8231. A mirror plot of the MS/MS spectra revealed that the two mass features shared numerous peaks in the low mass region (Figure S6). The common fragment ions correspond to the amino acid residues Lys, Tyr and Trp; together with the dipeptide fragments Ser-Lys, Lys-Trp and Trp-Ala. Interestingly, in addition to Lys, a fragment ion of Dab was detected in the MS/MS spectrum of the mass feature with *m/z* 523.8231. At the same time, no fragment ions for Glu or Phe were detected. These observations suggested that compounds from the second sub-family contained two positively charged amino acids, namely Lys and Dab, thereby representing new congeners of paenilipoheptin. Upon further investigation of the *b* ions of the parent mass with *m/z* 530.8307 our analysis yielded the lipopeptide sequence FA($C_{13}H_{26}$)–Ser–Lys–Trp–Ala–Leu–Tyr–Dab (Figure S7).

Characterization of new paenilipoheptins

Paenibacillus sp. JJ-21 and *Paenibacillus* sp. JJ-1722 stood out by producing substantial amounts of paenilipoheptins, and we, therefore, used these strains for the up-scale fermentation and isolation of new paenilipoheptins. *Paenibacillus* sp. JJ-21 and *Paenibacillus* sp. JJ-1722 were grown in 10 L of MHB medium and the biomass was collected by centrifugation. Further, specialized metabolites were extracted from the cells with isopropyl alcohol (IPA) supplemented with 0.1% (v/v) formic acid. The most abundant compounds, with *m/z* 562.3133 and *m/z* 530.8307, were isolated from the extracts of *Paenibacillus* sp. JJ-21 and JJ-1722 through multiple rounds of high-performance liquid chromatography (HPLC) and designated as paenilipoheptin A (**1**) and D (**2**), respectively. Compound **1** showed a molecular ion peak for an $[M+H]^+$ ion having *m/z* 1123.6180 (calcd. for $C_{59}H_{83}N_{10}O_{12}$, 1123.6192) in the ESI-HRMS spectrum (Figure S8). Analysis of 1D and 2D NMR spectra of compound **1** indicated the presence of seven amino acids: Ser, Lys, Trp, Ala, Phe, Tyr, Glu, and β-amino fatty acyl chain (Table S4, Figure 7A, Figures S9–S15). The β-amino fatty acid was identified as a 3-amino-10-methyldodecanoic acid upon detailed NMR analysis. The NMR structure confirmed that paenilipoheptin had in the structure Ala instead of Val and Lys instead of Dab. ESI-HRMS of compound **2** showed a molecular ion peak for an $[M+H]^+$ ion having *m/z* 1060.6555 (calcd. for $C_{55}H_{86}N_{11}O_{10,}$ 1060.6559) (Figure S16). The ^{1}H NMR spectra of **1** and **2** were similar, indicating similar structures. Due to the low yield of **2**, the NMR signals were quite weak. Nevertheless, it was possible to confirm the presence of Leu and Dab instead of Phe and Glu in the amino acid positions 5 and 7, respectively, in comparison with **1** (Table S5, Figure 7B, Figures S17–S22).

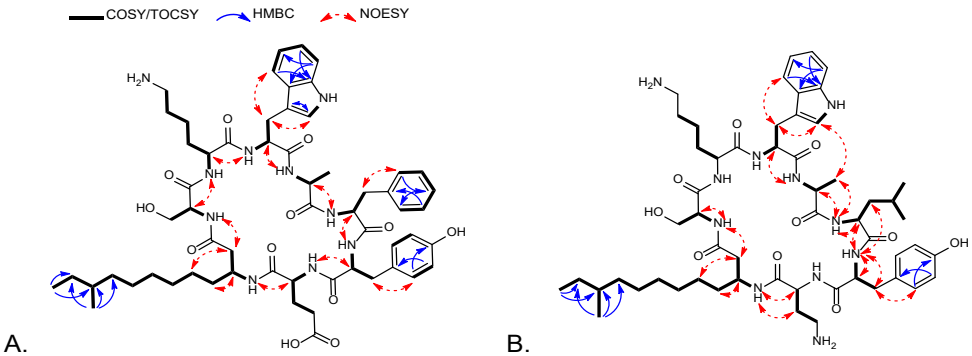

Figure 7. Correlations obtained by COSY, HMBC, and NOESY measurements in NMR of paenilipoheptin A (A) and paenilipoheptin D (B).

The inhibitory activity of the isolated specialized metabolites was evaluated against a panel of indicator strains by broth dilution assays (Table 1). Paenilipoheptin A and D showed moderate bioactivity against the Gram-positive *E. faecium* E980, *S. aureus* ATCC 29213 and *B. subtilis* 168 with an MIC of 32 µg/mL. The isolated compounds did not inhibit the growth of the Gram-negative strain at any tested concentration (MIC values of >32 µg/mL).

Table 1. In vitro minimum inhibitory concentrations (MICs) of paenilipoheptin A (1) and paenilipoheptin D (2).

	MIC (µg/mL)		
Indicator strain	(**1**)	(**2**)	Van/PmxB
E. faecium E980	32	32	0.5[a]
S. aureus ATCC 29213	32	32	0.5[a]
B. subtilis 168	32	32	0.25[a]
E. coli ATCC 25922	>32	>32	2[b]

[a] MIC value for vancomycin (Van)

[b] MIC value for polymyxin (PmxB)

Identification of the BGCs for paenilipoheptins

To characterize the paenilipoheptin BGCs the genomes of *Paenibacillus* sp. JJ-21 and *Paenibacillus* sp. JJ-1722 were sequenced using the PacBio platform. Assembly of the PacBio reads with Falcon (version 1.8.1) (Chin *et al.*, 2016) resulted in single contigs of 6.2 and 6.1 Mb for *Paenibacillus* sp. JJ-21 and *Paenibacillus* sp. JJ-1722, respectively. The genomes of *Paenibacillus* sp. JJ-21 and *Paenibacillus* sp. JJ-1722 were analyzed using antiSMASH 7.1.0 (Blin *et al.*, 2023) to identify the paenilipoheptin BGCs. The architecture of the paenilipoheptin BGCs in the genomes of *Paenibacillus* sp. JJ-21 and *Paenibacillus* sp. JJ-1722 was similar to

that of the *P. polymyxa* E681, from which paenilipoheptin A was previously identified (Vater *et al.*, 2018). An in silico analysis of the A domain substrate specificity of these BGCs was conducted with antiSMASH 7.1.0 (Blin *et al.*, 2023). The predicted amino acid composition of paenilipoheptin A of *Paenibacillus* sp. JJ-21 was identical to the one of *P. polymyxa* E681, which was D-Ser–D-Dab–D-Trp–D-Val–L-Phe–L-Tyr–L-Glu (Figure 6B). We also used antiSMASH 7.1.0 to predict which modules in the BGC of paenilipoheptin contained epimerization domains, i.e. domains that catalyze the conversion from L- to D-amino acids. Modules 1, 2, 3, and 4 were predicted to contain epimerization domains. The amino acid sequence predicted for paenilipoheptin D differed from the one predicted for paenilipoheptin A in positions 5 and 7, and was D-Ser–D-Dab–D-Trp–D-Val–L-Leu–L-Tyr–L-Dab (Figure 6C).

Discovery of potent bacitracin congeners within the molecular family of the lysine-containing peptides

The third sub-family comprises doubly-charged molecular ions with significantly larger masses compared to paenilipoheptins (Figure 8A). The edge connecting a node in sub-family 3 (*m/z* 712.8836) with the node due to paenilipoheptin D (**2**, *m/z* 530.8307) has a cosine score of 0.6, indicating low similarity. A mirror plot comparing the MS/MS spectra of the two mass features showed no matching dipeptide fragment ions, and the only shared ions were due to a Lys residue and fragments of Trp (Figure S23).

Notably, MS/MS spectra of all the mass features within this sub-family exhibited an ion with an *m/z* 199.0897. Manual dereplication efforts were undertaken to annotate these compounds, however, these mass features could not be matched to any of the previously reported microbial natural products. Analysis of the MS/MS spectrum of the parent mass with *m/z* 712.8836 revealed that the fragment ion with an *m/z* 199.0897 has a molecular formula of $C_9H_{15}N_2OS$, which corresponds to a thiazoline-containing substructure commonly found in the peptide antibiotic bacitracin, so far exclusively identified from the *Bacillus* genus (Qin *et al.*, 2020). The common b_2, b_3, b_4, and b_5 ions between the parent mass with *m/z* 712.8836 and bacitracin A (Suleiman *et al.*, 2017) revealed that their pentapeptide side chains are identical. The differences observed in b_6-b_{11} ions indicate that the structural distinctions between the two metabolites lie in the heptapeptide ring. We did not observe fragment ions corresponding to Orn, Phe or His, which are found in the structure of bacitracin A, while we detected a clear ion corresponding to a Trp residue (*m/z* 187.0864). Moreover, the dipeptide fragment ions Trp-Leu (*m/z* 300.171), Trp-Thr (*m/z* 288.1352), and Leu-Leu (*m/z* 227.1764) were detected. By tracing the series of both *b* and *y* ions, we elucidated the amino acid sequence of the compound with *m/z* 712.8836 as (Ile-Cys)–Leu/Ile–Glu–Leu/Ile–Lys–Leu/Ile–Leu/Ile–Trp–Thr–Asp–Asn (Figure S24). The MS/MS studies used did not allow us to discriminate between Leu and Ile in positions 3, 5, 7 and 8 of the peptide amino acid sequence. Notably, this sequence diverges from any known bacitracin analogues, indicating that we are dealing with a new variant of bacitracin.

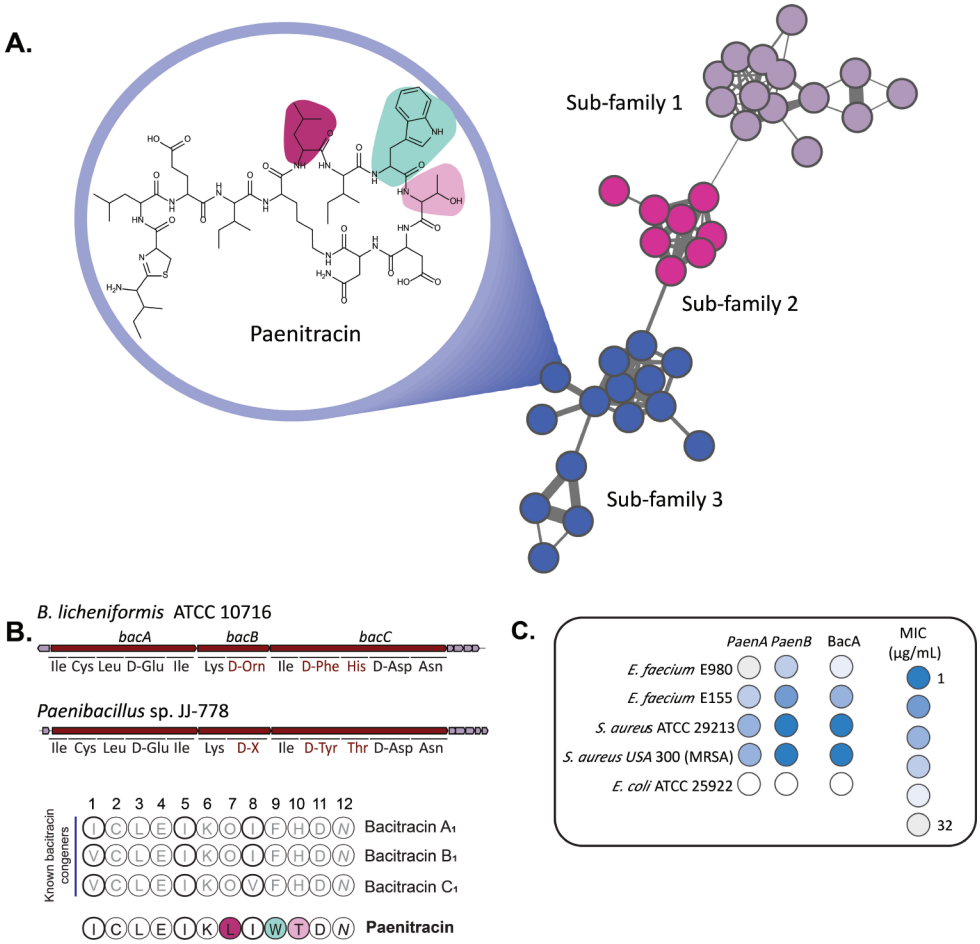

Figure 8. Molecular family of the Lys-containing specialized metabolites highlighted by MassQL search. A. Exploring the mass features from sub-family 3 led to the discovery of new bacitracin analog, designated as paenitracin. **B.** The BGC and its predicted amino acid substrates of new paenitracin from the genome of *Paenibacillus* sp. JJ-778 was compared to the reference BGC from *B. licheniformis* ATCC 10716. Comparison of the amino acid sequence of paenitracin with the sequences found in previously characterized bacitracin structures. The colored amino acid units show amino acids that are present solely in the structure of paenitracin. **C.** MIC data of paenitracin A and B, and commercial bacitracin against a panel of human pathogenic bacteria.

Next, to isolate the new bacitracin analog, *Paenibacillus* sp. JJ-778 was cultured in a 10-liter volume of MHB and subjected to extraction using Diaion® HP-20 beads, followed by elution with 100% IPA supplemented with 0.1% (v/v) formic acid. The crude extract was subjected to multiple rounds of chromatographic separation, resulting in the isolation of compounds (**3**) and (**4**) with the same exact mass (Figures S25, S26). A molecular formula of $C_{67}H_{106}N_{15}O_{17}S$ was deduced for **3** and **4**, based on their exact mass ($[M+H]^+$ found *m/z* 1424.7584, calcd. 1424.7611). 1D and 2D NMR analysis of **3** confirmed the amino acid sequence obtained

through MS/MS analysis (Figure 9), and together with the molecular formula dictated by the accurate mass, the structure of **3** was determined as a new congener of bacitracin (Table S6, Figure 8A, Figures S27-S33). The ^{1}H NMR spectra of **3** and **4** were almost identical, suggesting that they are diastereomers (Figure S34). Accordingly, the new compounds were designated as paenitracin A (**3**) and B (**4**).

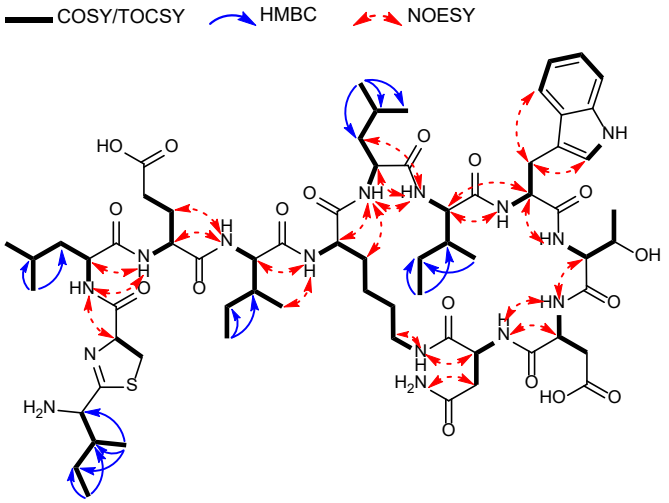

Figure 9. Correlations obtained by COSY, HMBC, and NOESY measurements in NMR of paenitracin A (3).

Bacitracin, originally isolated from *Bacillus licheniformis* and *B. subtilis*, targets primarily Gram-positive bacteria such as *Staphylococcus*, *Streptococcus*, *Enterococcus* and *Clostridium difficile* (André, 2005). Bacitracin and its congeners are synthesized by a nonribosomal peptide synthetase (NRPS) complex encoded by the bacitracin synthetase operon, which consists of *bacA*, *bacB*, and *bacC* genes. To explore the biosynthesis of new paenitracin, the genome of *Paenibacillus* sp. JJ-778 was sequenced and assembled using Falcon (version 1.8.1) (Chin *et al.*, 2016). Analysis using antiSMASH 7.1.0 (Blin *et al.*, 2023) readily identified a 12-modular gene cluster with 77% similarity to the bacitracin BGC of *B. licheniformis* ATCC 10716. In silico analysis of the A domain substrate specificity of BacA, BacB, and BacC predicted variations in amino acid composition at positions 7, 9, and 10 compared to the well-known bacitracin A. All known bacitracin congeners isolated and identified so far exhibit similar amino acid composition, with the Ile residues occasionally replaced by Val (Figure 8B). In contrast, paenitracin contains Leu, Trp and Thr instead of Orn, Phe and His at the amino acid positions 7, 9 and 10, respectively.

Bioactivity of new paenitracins

Antibacterial properties of new paenitracin A and B were assessed in a minimum inhibitory concentration (MIC) assay run against a panel of representative pathogenic strains. The

media used in the MIC assay was supplemented with 0.3 mM $ZnSO_4$ to provide the divalent metal ion required for bacitracin's activity (Buijs *et al.*, 2022). The activity of the paenitracin diastereomers was compared to that of commercially sourced bacitracin A, which served as a positive control. The results of the MIC assay revealed that paenitracin B exhibited activity against Gram-positives, with MICs ranging from 1 to 8 μg/mL that was comparable with the commercial bacitracin (Figure 8C). Importantly, paenitracin B exhibited activity against vancomycin-resistant *E. faecium* E155 with MICs of 2 μg/mL, which is lower compared to the MIC of the commercially available bacitracin A of natural origin. Considering the divergent structure of paenitracin as compared to the previously reported congeners the mode of action should be investigated further.

Conclusions

Extensive metabolomic analysis of 227 extracts derived from the plant-associated *Paenibacillus* isolates revealed underscored chemical diversity of the specialized metabolites and led to the discovery of novel paenitracin with potent bioactivity. Paenitracin possesses activity against Gram-positive ESKAPE pathogens, including vancomycin-resistant *E. faecium* E155. Molecular networking analysis, enhanced by the MassQL search of basic amino acids, revealed previously undescribed paenilipoheptins and paenitracins in the cultures of *Paenibacillus* environmental isolates. This substructure information facilitated the early prioritization and targeted isolation of bioactive natural products. This study illustrates that the integration of FBMN and MassQL represents an powerful approach for the elucidation of mass fragmentation patterns and the identification of novel bioactive natural products.

Materials and Methods

General experimental procedures

NMR spectra were recorded on a Bruker Ascend 850 MHz NMR spectrometer (Bruker BioSpin GmbH). Data were analyzed using MestReNova 14 software (Mestrelab Research, Santiago de Compostela, Spain). High-performance liquid chromatography (HPLC) purification was performed on a Waters preparative HPLC system composed of a 1525 pump, a 2707 autosampler, a 2998 PDA detector, and a Water fraction collector III or BESTA-Technik preparative HPLC system equipped with an ECOM Flash UV detector. All solvents and chemicals were of HPLC or LC-MS grade, depending on the experiment.

Isolation and identification of bacterial strains

Paenibacillus strains were obtained from the Auburn University Plant-Associated Microbial strain collection and had previously been isolated from the root surface of field-grown corn (*Zea mays*) and soybean (*Glycine max*) plants grown in Dunbar (Nebraska), Carrol (Iowa), Whitewater (Wisconsin), Sparta (Illinois), Troy (Ohio) and Tallassee (Alabama) (Table S1).

The phylogenetic tree was constructed using 16S rRNA gene sequences obtained through PCR amplification with universal bacterial primers 27F and 1492R (Lane, 1991). Each PCR amplicon was purified and used for Sanger sequencing. Subsequently, 16S rRNA gene sequences were assembled into consensus sequences and aligned with reference sequences. These include 220 sequences from our study, 41 sequences from type strains of *Paenibacillus* spp. downloaded from GenBank (Table S2) and sequence of *Bacillus subtilis* IAM 12118 (GenBank accession NR_112116.2). The alignment is done using ClustalOmega (v 1.2.2) (Sievers & Higgins, 2018). Aligned sequences are trimmed to the length of 1,452 bp to remove extending regions from a few sequences and also to preserve as many variable regions as possible. The trimmed alignment is re-aligned and a neighbor-joining consensus tree is generated using Geneious (v 2024.0.3) tree function, with Tamura-Nei genetic distance model and *B. subtilis* as outgroup. The generated tree is then uploaded to ITOL for visualization (Letunic & Bork, 2021).

Growth of *Paenibacillus* spp. and natural product extraction

Our method for cultivation and extraction of specialized metabolites was adapted from previously described procedures (Nguyen *et al.*, 2016)2016. Briefly, frozen stocks of 227 *Paenibacillus* spp. from the Auburn University strain collection were inoculated into 1600 µL of Tryptic Soy Agar (TSA, Bacto Soybean-Casein Digest Medium, 30 g/L) in 2.0 mL 96 deep well plates (Thermo Scientific, Nunc 2.0 mL DeepWell Plate). After 72 h, cultures were used to inoculate 200 µL Tryptic Soy Broth (TSB, Bacto Soybean-Casein Digest Medium, 30 g/L) with

a pin replicator in 96 well plates. These cultures were incubated overnight at 30 °C and used as an inoculum for the second 2.0 mL 96-deep well plate containing 600 µL TSA. Plates were sealed with 96 Well-Cap Mats (Thermo Scientific, Nunc 96 Well-Cap Mats), and incubated at 30 °C for 72 h. The cultures were extracted with 300 µL 100% isopropanol acidified with 0.1% formic acid. The plates were resealed with the same 96 Well-Cap Mats, sonicated for 10 min, and extracted for an additional 50 min. 200 µL of these crude extracts were transferred into a pre-washed 96 well plate (Agilent Technologies, 96 well plates, 0.5 mL, polypropylene) and lyophilized to dryness. The extraction protocol was repeated once more for a total extract volume of 400 µL. Dried samples were redissolved in 160 µL of methanol:water (1:1 v/v).

Genome sequencing, assembly and annotation

Paenibacillus sp. JJ-21, *Paenibacillus* sp. JJ-778, and *Paenibacillus* sp. JJ-1722 were grown in TSB at 30 °C and 220 rpm for 24 h. DNA was extracted as described (Kieser, 2000). DNA quality was verified by agarose gel electrophoresis. PacBio sequencing and assembly was performed by Novogene (UK). Generally, libraries were prepared using SMRTbell template prep kit (PacBio, USA) according to manufacturer instructions. Sequencing was then performed using PacBio Sequel platform in continuous long reads mode. Assembly was done using Falcon (version 1.8.1) (Chin *et al.*, 2016). BGCs in these genomes were annotated using AntiSMASH (version 7.1.0) (Blin *et al.*, 2023).

Data-dependent LC-ESI-HRMS/MS

LC-MS/MS acquisition was performed using Shimadzu Nexera X2 ultra-high-performance liquid chromatography (UPLC) system, with an attached photodiode array detector (PDA), coupled to Shimadzu 9030 QTOF mass spectrometer, equipped with a standard electrospray ionization (ESI) source unit. A total of 2 µL was injected into a Waters Acquity HSS C_{18} column (1.8 µm, 100 Å, 2.1 × 100 mm) and data acquisition was performed as previously described (van Bergeijk *et al.*, 2022). Briefly, the gradient used was 5% B for 1 min, 5–85% B for 9 min, 85–100% B for 1 min, and 100% B for 4 min. The column was re-equilibrated to 5% B for 3 min before the next run was started. All the samples were analyzed in positive polarity, using data-dependent acquisition mode. In this regard, full scan MS spectra (*m/z* 100–1700, scan rate 10 Hz, ID enabled) were followed by two data-dependent MS/MS spectra (*m/z* 100–1700, scan rate 10 Hz, ID disabled) for the two most intense ions per scan. The ions were fragmented using collision-induced dissociation (CID) with fixed collision energy (CE 20 eV) and excluded for 1 s before being re-selected for fragmentation. The parameters used for the ESI source were: interface voltage 4 kV, interface temperature 300 °C, nebulizing gas flow 3 L/min, and drying gas flow 10 L/min.

MZmine 2 parameters

Prior to statistical analysis, mzXML files were imported into Mzmine 2.53 (Pluskal *et al.*, 2010) and processed as previously described (van Bergeijk *et al.*, 2022). Briefly, mass ion peaks were detected for MS^1 and MS^2 at a noise level of 2.0E2 and 0.0E0, respectively (positive polarity, mass detector: centroid), and their chromatograms were built using ADAP chromatogram builder. The detected peaks were smoothed, and the chromatograms were deconvoluted (algorithm: local minimum search; chromatographic threshold: 90; search minimum in RT range: 0.05; minimum relative height: 1%; minimum ratio of peak top/edge: 2; peak duration: 0.03–3.00 min). The detected peaks were deisotoped (monotonic shape; maximum charge: 3; representative isotope: most intense). Peak lists from different samples were aligned (weight for RT = weight for m/z = 50; compare the isotopic pattern with a minimum score of 50%). Missing peaks detected in at least one of the samples were filled with the gap-filling algorithm (RT tolerance: 0.1 min, m/z tolerance was set to 0.002 m/z or 15 ppm). Duplicate peaks were filtered. Only the features with MS/MS data were exported to a GNPS-FBMN.

In addition, all features that originate from the culture medium were removed by retaining only features with an average peak intensity of at least 100 times greater in the bacterial extracts than in the culture medium extracts. Features that are present in the blanks with an intensity higher than 5000 were also removed. In the end, features with m/z less than 300 were filtered out.

Molecular network analysis and MassQL search

The resulting feature quantification table (CSV file) and MS/MS spectrum files (in mgf format) were uploaded to the GNPS webserver (http://gnps.ucsd.edu) (Wang *et al.*, 2016). The LC-MS/MS data for the 227 *Paenibacillus* isolates were analyzed using the Feature-Based Molecular Networking tool (FBMN) (Nothias *et al.*, 2020). Briefly, the precursor ion mass tolerance was set to 0.005 Da and the MS/MS fragment ion tolerance to 0.05 Da. A molecular network was then created where edges were filtered to have a cosine score above 0.5 and more than 3 matched peaks. The molecular networks were visualized using Cytoscape software version 3.9.1 (Shannon *et al.*, 2003) and displayed using an unweighted force-directed layout. The data are publicly accessible in the MassIVE repository (MSV000094386) and the networking results and parameters can be found at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=2f1e0ae1728142249ac4e841d5a72ef4 . The MassQL tool version 31.4 (Jarmusch *et al.*, 2022) was employed to search for specific MS/MS fragments that correspond to the *b* or immonium ions of positively charged amino acids. The fragment search of *b* ions of 2,4-diaminobuturic acid (m/z 101.0709), ornithine (m/z 115.0871), lysine (m/z 129.1027), and arginine (m/z 157.1089); together with the immonium ion of histidine (m/z 110.0718), was conducted with an accuracy of m/z 0.005 and minimum intensity of 5%.

Up-scale fermentation, extraction, and isolation

Paenibacillus sp. JJ-21, JJ-1722 and JJ-778 were grown at 30 °C on tryptic soy agar (TSA) for 72 h and three colonies were inoculated into TSB and incubated at 30 °C overnight. This inoculum (1%) was used to inoculate thirteen 2 L Erlenmeyer flasks containing 0.75 L of sterile MHB and fermented at 30 °C while shaking at 200 rpm for 72 h.

To extract the specialized metabolites produced by *Paenibacillus* sp. JJ-21 and JJ-1722, cells were collected by centrifugation (8000 rpm, 30 min, 4 °C) and washed with H_2O. Then, the cells were sonicated for 30 min and extracted with 100% isopropyl alcohol (IPA) supplemented with 0.1% (v/v) formic acid (FA) for 6 h. The crude extracts were collected, concentrated under reduced pressure and reconstituted in 50% ACN. The crude extract of *Paenibacillus* sp. JJ-21 was subjected to the preparative HPLC BESTA-Technik system with a Dr. Maisch Reprosil Gold 120 C18 column (25 × 250 mm, 10 µm), run at 12 mL/min, and eluted using an H_2O–ACN gradient of 30–70% in 49 min to yield paenilipoheptin A (**1**, 1 mg). The crude extract of *Paenibacillus* sp. JJ-1722 was subjected to the Waters preparative HPLC system with SunFire C18 column (10 µm, 100 Å, 19×150 mm) eluted with a H_2O–ACN gradient of 10–60% in 30 min, at a flow rate of 15 mL/min. The third fraction was further purified on a semi-preparative SunFire C_{18} column (5 µm, 100 Å, 10×250 mm), run at 3 mL/min, and eluted using a H_2O–ACN gradient of 25–35% in 30 min, to yield paenilipoheptin D (**2**, 0.3 mg).

To extract the specialized metabolites produced by *Paenibacillus* sp. JJ-778 supernatant was collected by centrifugation and subjected to extraction using Diaion® HP-20 (Resindion, Mitsubishi), followed by elution with 100% IPA supplemented with 0.1% (v/v) formic acid. The crude extract was adsorbed onto silica gel (pore size 60 Å, 70–230 mesh, Sigma Aldrich), and loaded on a silica column, which was eluted using varying gradients of EtOAc/MeOH. The fractions eluted with EtOAc-MeOH (1:3) supplemented with 0.1% (v/v) formic acid were combined, reconstituted in 50% ACN and injected into the preparative SunFire column (19 × 150 mm), which was eluted with H_2O:ACN gradient of 15–60% in 30 min, at a flow rate of 15 mL/min, to yield three fractions. Fraction 3 was further separated on the semi-preparative SunFire column (10 × 250 mm) at 3 mL/min using a H_2O–ACN gradient of 30–50% in 30 min to yield paenitracin A (**3**, 0.3 mg) and paenitracin B (**4**, 0.2 mg).

Antibacterial activity assays and MIC determination

For *E. coli* ATCC 25922 and *S. aureus* ATCC 29213, single colonies were inoculated into 20 mL of MHII broth medium in 250 mL Erlenmeyer flask and incubated overnight at 37 °C with shaking at 220 rpm and then diluted to obtain an assay inoculum of 5 x 10^5 CFU/mL. Aztreonam and vancomycin were used as reference positive controls for each pathogen. Absorbance at OD_{612} nm was measured at T_0 (zero time) using an EnVision™ Microplate Reader (PerkinElmer), and assay plates were incubated at 37 °C. After that, absorbance at

OD_{612} nm was measured at T_f (final time) using the same microplate reader equipment. The percentage of growth inhibition was calculated using the following normalization:

$$\%\text{Inhibition} = 100 \times \left\{1 - \frac{([T_{f\text{Sample}} - T_{0\text{Sample}}] - [T_{f\text{Blank}} - T_{0\text{Blank}}])}{([T_{f\text{Growth}} - T_{0\text{Growth}}] - [T_{f\text{Blank}} - T_{0\text{Blank}}])}\right\}$$

$T_{0\text{Sample}}$, the absorbance of the strain growth in the presence of sample measured at zero time;
$T_{f\text{Sample}}$, the absorbance of the strain growth in the presence of sample measured at final time;
$T_{0\text{Growth}}$, the absorbance of the strain growth in the absence of sample measured at zero time;
$T_{f\text{Growth}}$, the absorbance of the strain growth in the absence of sample measured at the final time;
$T_{0\text{Blank}}$, the absorbance of the broth medium (blank) measured at zero time;
$T_{f\text{Blank}}$, the absorbance of the broth medium (blank) measured at the final time.

The Genedata Screener software (Genedata, Inc., Basel, Switzerland) was used to process and analyze the data. The activity of the samples was expressed as a percentage of growth inhibition, where (-100%) represented the total growth inhibition of the target microorganism and 0% represented the total growth of the target microorganism in the assay. The Z' factor predicts the robustness of an assay by considering the mean and standard deviation of both positive and negative controls (Zhang *et al.*, 1999). The robust Z' factor (RZ' factor) is based on the Z' factor, but standard deviations and means are replaced by the robust standard deviations and medians, respectively. In all experiments performed in this work, the RZ' factor obtained was between 0.9173 and 0.9684.

For MIC tests, blood agar plates were inoculated with glycerol stocks of *E. coli* ATCC 25922, *S. aureus* ATCC 29213, *S. aureus* USA300 (clinical isolate from Texas Children's Hospital, USA), *B. subtilis* 168, *E. faecium* E155 and *E. faecium* E980 and incubated overnight at 37 °C. Following incubation, 3 mL of tryptic soy broth (TSB) was inoculated with an individual colony and incubated at 37 °C with shaking at 220 rpm until the cultures reached the exponential phase. The bacterial suspensions were then diluted 100-fold in cation-adjusted Mueller-Hinton broth (MHB) or Luria-Bertani (LB) broth supplemented with 0.3 M $ZnSO_4$ to reach a bacterial cell density of 10^6 CFU/mL. In parallel, test compounds were 2-fold serially diluted with cation-adjusted MHB or LB supplemented with 0.3 M $ZnSO_4$ in polypropylene 96-well plates to achieve a final volume of 50 µL per well. An equal volume of bacterial suspension (10^6 CFU/mL) was added to the wells. The well plates were sealed with an adhesive membrane and incubated for 20 hours at 37 °C. The reported MIC values are based on three technical triplicates and are defined as the lowest concentration of the compound that prevented the visible growth of bacteria.

Acknowledgments

We thank D. van der Horst for technical assistance with LC-MS analysis. We are grateful to Jesús Martín for assistance in compound annotation and for stimulating discussions. The work was supported by the NACTAR program of The Netherlands Organization for Scientific Research (NWO), Grant 16440 to G.P.vW., M.H.M. and N.I.M.

Supplementary information

Table S1. Overview of 227 *Paenibacillus* spp. from the Auburn University Plant-Associated Microbial strain collection.

Isolate	Top hit	Pairwise similarity (%)	Crop	Rhizoplane/ Endophyte	Root #	City	State	Bioactivity against *E.coli* ATCC 25922	Bioactivity against *S. aureus* ATCC 29213
JJ-6	*P. xylanexedens*	99.39	Corn	Endophyte	1	Dunbar	Nebraska	–	–
JJ-7	*P. typhae*	98.70	Corn	Endophyte	1	Dunbar	Nebraska	–	–
JJ-15	*P. amylolyticus*	99.59	Corn	Endophyte	3	Dunbar	Nebraska	–	–
JJ-16	*P. peoriae*	99.59	Corn	Endophyte	3	Dunbar	Nebraska	+	–
JJ-18*	*P. alginolyticus*	99.22	Corn	Endophyte	3	Dunbar	Nebraska	–	–
JJ-21*	*P. peoriae*	99.66	Corn	Endophyte	3	Dunbar	Nebraska	+	–
JJ-29	*P. illinoisensis*	99.65	Corn	Endophyte	4	Dunbar	Nebraska	–	–
JJ-33	*P. lautus*	99.37	Corn	Endophyte	4	Dunbar	Nebraska	–	–
JJ-39	*P. terrigena*	99.25	Corn	Endophyte	6	Dunbar	Nebraska	–	–
JJ-42	*P. pectinilyticus*	98.78	Corn	Endophyte	6	Dunbar	Nebraska	–	–
JJ-47	*P. cineris*	100.00	Corn	Endophyte	7	Dunbar	Nebraska	–	–
JJ-49	*P. lautus*	99.16	Corn	Endophyte	7	Dunbar	Nebraska	–	–
JJ-56	*P. amylolyticus*	99.45	Corn	Endophyte	9	Dunbar	Nebraska	–	–
JJ-59	*P. aceris*	98.61	Corn	Endophyte	9	Dunbar	Nebraska	–	–
JJ-60	*P. etheri*	99.65	Corn	Endophyte	9	Dunbar	Nebraska	–	–
JJ-72	*P. amylolyticus*	99.80	Corn	Endophyte	12	Dunbar	Nebraska	–	–

JJ-73	*P. terrigena*	99.04	Corn	Endophyte	12	Dunbar	Nebraska	–	–
JJ-77	*P. aceris*	99.65	Corn	Endophyte	12	Dunbar	Nebraska	–	–
JJ-90	*P. typhae*	98.70	Corn	Endophyte	14	Dunbar	Nebraska	–	–
JJ-92	*P. lautus*	99.30	Corn	Endophyte	14	Dunbar	Nebraska	–	–
JJ-93	*P. aceris*	98.75	Corn	Endophyte	14	Dunbar	Nebraska	–	–
JJ-95	*P. aceris*	99.29	Corn	Endophyte	15	Dunbar	Nebraska	–	–
JJ-99	*P. aceris*	98.75	Corn	Endophyte	15	Dunbar	Nebraska	–	–
JJ-102	*P. xylanexedens*	99.39	Corn	Endophyte	16	Dunbar	Nebraska	–	–
JJ-106	*P. marinesediminis*	99.59	Soybean	Endophyte	1	Dunbar	Nebraska	–	–
JJ-121	*P. lautus*	99.23	Soybean	Endophyte	5	Dunbar	Nebraska	–	–
JJ-148	*P. aceris*	98.68	Soybean	Endophyte	11	Dunbar	Nebraska	–	–
JJ-159	*P. lautus*	99.23	Soybean	Endophyte	13	Dunbar	Nebraska	–	–
JJ-174	*P. amylolyticus*	99.25	Soybean	Endophyte	15	Dunbar	Nebraska	–	–
JJ-180	*P. aceris*	98.75	Soybean	Endophyte	16	Dunbar	Nebraska	–	–
JJ-193	*P. terrigena*	99.52	Corn	Endophyte	1	Dunbar	Nebraska	–	–
JJ-194	*P. amylolyticus*	99.32	Corn	Endophyte	3	Dunbar	Nebraska	–	–
JJ-195	*P. peoriae*	99.72	Corn	Endophyte	3	Dunbar	Nebraska	+	+
JJ-216	*P. pocheonensis*	99.04	Corn	Rhizoplane	1	Dunbar	Nebraska	–	–
JJ-223	*P. xylanilyticus*	98.37	Corn	Rhizoplane	1	Dunbar	Nebraska	–	–
JJ-226	*P. peoriae*	99.79	Corn	Rhizoplane	2	Dunbar	Nebraska	+	–
JJ-227	*P. peoriae*	99.79	Corn	Rhizoplane	2	Dunbar	Nebraska	+	+
JJ-228	*P. peoriae*	99.79	Corn	Rhizoplane	2	Dunbar	Nebraska	+	–
JJ-231	*P. illinoisensis*	99.86	Corn	Rhizoplane	3	Dunbar	Nebraska	–	–

JJ-232	*P. dongdonensis*	99.19	Corn	Rhizoplane	3	Dunbar	Nebraska	–	–
JJ-235	*P. amylolyticus*	99.66	Corn	Rhizoplane	3	Dunbar	Nebraska	–	–
JJ-237	*P. amylolyticus*	99.66	Corn	Rhizoplane	3	Dunbar	Nebraska	–	–
JJ-246	*P. oenotherae*	98.63	Corn	Rhizoplane	5	Dunbar	Nebraska	–	–
JJ-253	*P. pectinilyticus*	98.78	Corn	Rhizoplane	6	Dunbar	Nebraska	–	–
JJ-268	*P. tritici*	99.73	Corn	Rhizoplane	9	Dunbar	Nebraska	–	–
JJ-270	*P. tritici*	100.00	Corn	Rhizoplane	9	Dunbar	Nebraska	–	–
JJ-272	*P. tritici*	98.90	Corn	Rhizoplane	9	Dunbar	Nebraska	–	–
JJ-287	*P. agaridevorans*	97.41	Corn	Rhizoplane	13	Dunbar	Nebraska	–	–
JJ-305	*P. illinoisensis*	99.72	Corn	Rhizoplane	16	Dunbar	Nebraska	–	–
JJ-310	*P. endophyticus*	99.52	Soybean	Rhizoplane	1	Dunbar	Nebraska	–	–
JJ-311	*P. amylolyticus*	98.64	Soybean	Rhizoplane	1	Dunbar	Nebraska	–	–
JJ-317	*P. 'taohuashanense'*	99.86	Soybean	Rhizoplane	1	Dunbar	Nebraska	–	–
JJ-324	*P. lautus*	99.37	Soybean	Rhizoplane	2	Dunbar	Nebraska	–	–
JJ-329	*P. 'yonginensis'*	97.47	Soybean	Rhizoplane	3	Dunbar	Nebraska	–	–
JJ-331	*P. 'taohuashanense'*	98.66	Soybean	Rhizoplane	3	Dunbar	Nebraska	–	–
JJ-332	*P. catalpae*	99.59	Soybean	Rhizoplane	3	Dunbar	Nebraska	–	–
JJ-333	*P. jamilae*	99.86	Soybean	Rhizoplane	3	Dunbar	Nebraska	–	+
JJ-337	*P. illinoisensis*	99.79	Soybean	Rhizoplane	4	Dunbar	Nebraska	–	–
JJ-338	*P. lautus*	99.37	Soybean	Rhizoplane	4	Dunbar	Nebraska	–	–
JJ-340	*P. thiaminolyticus*	99.79	Soybean	Rhizoplane	4	Dunbar	Nebraska	–	–
JJ-342	*P. terrigena*	99.52	Soybean	Rhizoplane	4	Dunbar	Nebraska	–	–

JJ-352	*P. aceris*	98.61	Soybean	Rhizoplane	6	Dunbar	Nebraska	–	–
JJ-355	*P. illinoisensis*	99.72	Soybean	Rhizoplane	6	Dunbar	Nebraska	–	–
JJ-364	*P. urinalis*	99.35	Soybean	Rhizoplane	7	Dunbar	Nebraska	–	–
JJ-365	*P. odorifer*	99.51	Soybean	Rhizoplane	7	Dunbar	Nebraska	–	–
JJ-372	*P. aceris*	99.30	Soybean	Rhizoplane	8	Dunbar	Nebraska	–	–
JJ-380	*P. rigui*	98.18	Soybean	Rhizoplane	10	Dunbar	Nebraska	–	–
JJ-385	*P. silagei*	97.09	Soybean	Rhizoplane	11	Dunbar	Nebraska	–	–
JJ-391	*P. illinoisensis*	99.65	Soybean	Rhizoplane	12	Dunbar	Nebraska	–	–
JJ-393	*P. terrigena*	99.17	Soybean	Rhizoplane	13	Dunbar	Nebraska	–	–
JJ-402	*P. terrigena*	99.79	Soybean	Rhizoplane	14	Dunbar	Nebraska	–	–
JJ-408	*P. terrigena*	99.04	Soybean	Rhizoplane	15	Dunbar	Nebraska	–	–
JJ-410	*P. tritici*	99.93	Soybean	Rhizoplane	15	Dunbar	Nebraska	–	–
JJ-411	*P. aceris*	98.68	Soybean	Rhizoplane	15	Dunbar	Nebraska	–	–
JJ-412	*P. lautus*	99.23	Soybean	Rhizoplane	15	Dunbar	Nebraska	–	–
JJ-413	*P. 'taohuashanense'*	98.80	Soybean	Rhizoplane	16	Dunbar	Nebraska	–	–
JJ-415	*P. catalpae*	99.59	Soybean	Rhizoplane	16	Dunbar	Nebraska	–	–
JJ-447	*P. ehimensis*	96.56	Corn	Rhizoplane	12	Dunbar	Nebraska	–	–
JJ-450	*P. pectinilyticus*	99.20	Soybean	Rhizoplane	5	Dunbar	Nebraska	–	–
JJ-460	*P. pectinilyticus*	99.78	Corn	Endophyte	9	Dunbar	Nebraska	–	–
JJ-467	*P. pectinilyticus*	98.85	Soybean	Rhizoplane	5	Dunbar	Nebraska	–	–
JJ-471	*P. sabinae*	99.46	Soybean	Rhizoplane	15	Dunbar	Nebraska	–	–
JJ-483	*P. panacisoli*	99.65	Corn	Rhizoplane	2	Carrol	Iowa	–	–
JJ-497	*P. amylolyticus*	99.59	Corn	Rhizoplane	3	Carrol	Iowa	–	–
JJ-514	*P. dongdonensis*	99.12	Corn	Rhizoplane	4	Carrol	Iowa	–	–

JJ-526	*P. lautus*	99.21	Corn	Rhizoplane	6	Carrol	Iowa	–	–
JJ-534	*P. barcinonensis*	99.11	Corn	Rhizoplane	7	Carrol	Iowa	–	–
JJ-539	*P. cucumis*	100.00	Corn	Rhizoplane	8	Carrol	Iowa	–	–
JJ-541	*P. panacisoli*	99.72	Corn	Rhizoplane	8	Carrol	Iowa	–	–
JJ-602	*P. alvei*	99.19	Corn	Endophyte	5	Carrol	Iowa	–	–
JJ-624	*P. alkaliterrae*	99.33	Corn	Endophyte	7	Carrol	Iowa	–	–
JJ-631	*P. lautus*	99.09	Corn	Endophyte	8	Carrol	Iowa	–	–
JJ-633	*P. lautus*	99.09	Corn	Endophyte	8	Carrol	Iowa	–	–
JJ-667	*P. aceris*	98.68	Corn	Rhizoplane	3	Whitewater	Wisconsin	–	–
JJ-676	*P. 'taohuashanense'*	98.59	Corn	Rhizoplane	5	Whitewater	Wisconsin	–	–
JJ-690	*P. lautus*	99.58	Corn	Rhizoplane	7	Whitewater	Wisconsin	–	–
JJ-702	*P. tundrae*	99.46	Corn	Rhizoplane	8	Whitewater	Wisconsin	–	–
JJ-718	*P. 'taohuashanense'*	98.73	Corn	Rhizoplane	10	Whitewater	Wisconsin	–	–
JJ-754	*P. panacisoli*	99.59	Corn	Endophyte	5	Whitewater	Wisconsin	–	–
JJ-764	*P. castaneae*	97.70	Corn	Endophyte	6	Whitewater	Wisconsin	–	–
JJ-773	*P. amylolyticus*	99.59	Corn	Endophyte	7	Whitewater	Wisconsin	–	–
JJ-778*	*P. amylolyticus*	99.80	Corn	Endophyte	8	Whitewater	Wisconsin	–	–
JJ-779	*P. lactis*	99.51	Corn	Endophyte	8	Whitewater	Wisconsin	–	–
JJ-782	*P. amylolyticus*	99.73	Corn	Endophyte	8	Whitewater	Wisconsin	–	–
JJ-788	*P. amylolyticus*	99.52	Corn	Endophyte	9	Whitewater	Wisconsin	–	–
JJ-794	*P. lautus*	99.44	Corn	Endophyte	10	Whitewater	Wisconsin	–	–
JJ-798	*P. 'taohuashanense'*	98.87	Corn	Endophyte	10	Whitewater	Wisconsin	–	–
JJ-816	*P. barcinonensis*	99.11	Corn	Rhizoplane	3	Carrol	Iowa	–	–
JJ-820	*P. tundrae*	99.59	Corn	Rhizoplane	6	Carrol	Iowa	–	–

JJ-824	*P. illinoisensis*	99.86	Corn	Rhizoplane	9	Carrol	Iowa	–	–
JJ-845	*P. jamilae*	99.52	Corn	Endophyte	10	Whitewater	Wisconsin	+	–
JJ-870	*P. castaneae*	99.32	Soybean	Rhizoplane	4	Whitewater	Wisconsin	–	–
JJ-890	*P. contaminans*	99.93	Soybean	Rhizoplane	7	Whitewater	Wisconsin	–	–
JJ-955	*P. silagei*	98.84	Soybean	Endophyte	7	Whitewater	Wisconsin	–	–
JJ-960	*P. aestuarii*	95.04	Soybean	Endophyte	8	Whitewater	Wisconsin	–	–
JJ-961	*P. marinisediminis*	99.52	Soybean	Endophyte	8	Whitewater	Wisconsin	–	–
JJ-963	*P. tritici*	99.57	Soybean	Endophyte	8	Whitewater	Wisconsin	–	–
JJ-964	*P. lautus*	99.44	Soybean	Endophyte	8	Whitewater	Wisconsin	–	–
JJ-985	*P. castaneae*	99.39	Soybean	Rhizoplane	2	Whitewater	Wisconsin	–	–
JJ-1000	*P. xylanexedens*	99.86	Corn	Rhizoplane	1	Sparta	Illinois	–	–
JJ-1004	*P. barcinonensis*	98.69	Corn	Rhizoplane	1	Sparta	Illinois	–	–
JJ-1057	*P. amylolyticus*	99.80	Corn	Rhizoplane	8	Sparta	Illinois	–	–
JJ-1059	*P. lautus*	99.23	Corn	Rhizoplane	8	Sparta	Illinois	–	–
JJ-1069	*P. barcinonensis*	98.69	Corn	Endophyte	1	Sparta	Illinois	–	–
JJ-1074	*P. amylolyticus*	99.86	Corn	Endophyte	1	Sparta	Illinois	–	–
JJ-1099	*P. odorifer*	98.58	Corn	Endophyte	6	Sparta	Illinois	–	–
JJ-1103	*P. xylanilyticus*	100.00	Corn	Endophyte	6	Sparta	Illinois	–	–
JJ-1115	*P. amylolyticus*	99.59	Corn	Endophyte	9	Sparta	Illinois	–	–
JJ-1181	*P. terrigena*	99.17	Soybean	Rhizoplane	8	Sparta	Illinois	–	–
JJ-1187	*P. illinoisensis*	99.79	Soybean	Rhizoplane	8	Sparta	Illinois	–	–
JJ-1277	*P. dongdonensis*	99.12	Corn	Rhizoplane	5	Sparta	Illinois	–	–
JJ-1283	*P. pocheonensis*	98.97	Corn	Endophyte	9	Sparta	Illinois	–	–

JJ-1311	*P. amylolyticus*	99.58	Soybean	Endophyte	6	Sparta	Illinois	–	–
JJ-1319	*P. barcinonensis*	99.17	Soybean	Rhizoplane	2	Carrol	Iowa	–	–
JJ-1343	*P. barcinonensis*	99.21	Soybean	Rhizoplane	5	Carrol	Iowa	–	–
JJ-1371	*P. catalpae*	99.66	Soybean	Rhizoplane	9	Carrol	Iowa	–	–
JJ-1402	*P. susongensis*	99.17	Soybean	Endophyte	2	Carrol	Iowa	–	–
JJ-1405	*P. susongensis*	99.17	Soybean	Endophyte	2	Carrol	Iowa	–	–
JJ-1410	*P. odorifer*	98.78	Soybean	Endophyte	3	Carrol	Iowa	–	–
JJ-1415	*P. selenitireducens*	99.23	Soybean	Endophyte	4	Carrol	Iowa	–	–
JJ-1425	*P. granivorans*	96.71	Soybean	Endophyte	5	Carrol	Iowa	–	–
JJ-1461	*P. barcinonensis*	99.24	Corn	Rhizoplane	1	Troy	Ohio	–	–
JJ-1465	*P. cucumis*	99.93	Corn	Rhizoplane	1	Troy	Ohio	–	–
JJ-1470	*P. amylolyticus*	99.45	Corn	Rhizoplane	2	Troy	Ohio	–	–
JJ-1476	*P. massiliensis*	99.86	Corn	Rhizoplane	2	Troy	Ohio	–	–
JJ-1487	*P. pocheonensis*	98.97	Corn	Rhizoplane	3	Troy	Ohio	–	–
JJ-1501	*P. lautus*	99.37	Corn	Rhizoplane	6	Troy	Ohio	–	–
JJ-1516	*P. susongensis*	99.24	Corn	Rhizoplane	8	Troy	Ohio	–	–
JJ-1522	*P. barcinonensis*	99.31	Corn	Endophyte	2	Troy	Ohio	–	–
JJ-1524	*P. tritici*	99.71	Corn	Endophyte	2	Troy	Ohio	–	–
JJ-1525	*P. odorifer*	99.93	Corn	Endophyte	2	Troy	Ohio	–	–
JJ-1531	*P. rhizoplanae*	98.21	Corn	Endophyte	3	Troy	Ohio	–	–
JJ-1544	*P. barcinonensis*	99.17	Corn	Endophyte	5	Troy	Ohio	–	–
JJ-1549	*P. massiliensis*	99.73	Corn	Endophyte	5	Troy	Ohio	–	–
JJ-1553	*P. cucumis*	99.79	Corn	Endophyte	6	Troy	Ohio	–	–
JJ-1555	*P. barcinonensis*	99.31	Corn	Endophyte	6	Troy	Ohio	–	–

JJ-1564	*P. cucumis*	99.59	Corn	Endophyte	7	Troy	Ohio	–	–	
JJ-1570	*P. susongensis*	99.17	Corn	Endophyte	8	Troy	Ohio	–	–	
JJ-1580	*P. peoriae*	99.72	Soybean	Rhizoplane	1	Troy	Ohio	+	–	
JJ-1582	*P. peoriae*	99.65	Soybean	Rhizoplane	1	Troy	Ohio	–	–	
JJ-1586	*P. cucumis*	100.00	Soybean	Rhizoplane	2	Troy	Ohio	–	–	
JJ-1587	*P. illinoisensis*	99.72	Soybean	Rhizoplane	2	Troy	Ohio	–	–	
JJ-1588	*P. amylolyticus*	99.52	Soybean	Rhizoplane	2	Troy	Ohio	–	–	
JJ-1591	*P. taichungensis*	99.73	Soybean	Rhizoplane	2	Troy	Ohio	–	–	
JJ-1595	*P. odorifer*	98.48	Soybean	Rhizoplane	3	Troy	Ohio	–	–	
JJ-1596	*P. barcinonensis*	99.17	Soybean	Rhizoplane	3	Troy	Ohio	–	–	
JJ-1598	*P. odorifer*	99.80	Soybean	Rhizoplane	3	Troy	Ohio	–	–	
JJ-1601	*P. illinoisensis*	99.79	Soybean	Rhizoplane	3	Troy	Ohio	–	–	
JJ-1602	*P. typhae*	98.18	Soybean	Rhizoplane	3	Troy	Ohio	–	–	
JJ-1603	*P. peoriae*	99.86	Soybean	Rhizoplane	3	Troy	Ohio	+	–	
JJ-1604	*P. polymyxa*	100.00	Soybean	Rhizoplane	3	Troy	Ohio	+	–	
JJ-1611	*P. typhae*	97.08	Soybean	Rhizoplane	4	Troy	Ohio	–	–	
JJ-1612	*P. barcinonensis*	99.24	Soybean	Rhizoplane	4	Troy	Ohio	–	–	
JJ-1614	*P. peoriae*	99.65	Soybean	Rhizoplane	4	Troy	Ohio	+	–	
JJ-1615	*P. typhae*	99.45	Soybean	Rhizoplane	4	Troy	Ohio	–	–	
JJ-1620	*P. massiliensis*	99.86	Soybean	Rhizoplane	5	Troy	Ohio	–	–	
JJ-1623	*P. dongdonensis*	99.12	Soybean	Rhizoplane	5	Troy	Ohio	–	–	
JJ-1631	*P. dongdonensis*	99.27	Soybean	Rhizoplane	6	Troy	Ohio	–	–	
JJ-1638	*P. peoriae*	99.72	Soybean	Rhizoplane	7	Troy	Ohio	+	+	
JJ-1639	*P. cucumis*	100.00	Soybean	Rhizoplane	7	Troy	Ohio	–	–	

JJ-1640	*P. peoriae*	99.72	Soybean	Rhizoplane	7	Troy	Ohio	+	–
JJ-1648	*P. turicensis*	97.36	Soybean	Rhizoplane	8	Troy	Ohio	–	–
JJ-1650	*P. jamilae*	99.88	Soybean	Rhizoplane	8	Troy	Ohio	+	–
JJ-1652	*P. peoriae*	99.72	Soybean	Rhizoplane	9	Troy	Ohio	+	–
JJ-1653	*P. cucumis*	100.00	Soybean	Rhizoplane	9	Troy	Ohio	–	–
JJ-1667	*P. odorifer*	99.86	Soybean	Endophyte	1	Troy	Ohio	–	–
JJ-1669	*P. odorifer*	98.58	Soybean	Endophyte	1	Troy	Ohio	–	–
JJ-1678	*P. cucumis*	100.00	Soybean	Endophyte	2	Troy	Ohio	–	–
JJ-1679	*P. cucumis*	100.00	Soybean	Endophyte	2	Troy	Ohio	–	–
JJ-1680	*P. massiliensis*	99.86	Soybean	Endophyte	2	Troy	Ohio	–	–
JJ-1683*	*P. jamilae*	99.73	Soybean	Endophyte	3	Troy	Ohio	+	–
JJ-1684	*P. cucumis*	99.31	Soybean	Endophyte	3	Troy	Ohio	–	–
JJ-1688	*P. massiliensis*	99.86	Soybean	Endophyte	3	Troy	Ohio	–	–
JJ-1692	*P. typhae*	99.59	Soybean	Endophyte	4	Troy	Ohio	–	–
JJ-1693	*P. amylolyticus*	99.52	Soybean	Endophyte	4	Troy	Ohio	–	–
JJ-1701	*P. typhae*	99.66	Soybean	Endophyte	5	Troy	Ohio	–	–
JJ-1703	*P. typhae*	99.59	Soybean	Endophyte	5	Troy	Ohio	–	–
JJ-1709	*P. massiliensis*	99.93	Soybean	Endophyte	5	Troy	Ohio	–	–
JJ-1715	*P. peoriae*	99.72	Soybean	Endophyte	6	Troy	Ohio	+	+
JJ-1720	*P. peoriae*	99.72	Soybean	Endophyte	6	Troy	Ohio	+	+
JJ-1722*	*P. peoriae*	99.72	Soybean	Endophyte	7	Troy	Ohio	+	+
JJ-1723	*P. dongdonensis*	99.27	Soybean	Endophyte	7	Troy	Ohio	–	–
JJ-1724	*P. peoriae*	99.72	Soybean	Endophyte	7	Troy	Ohio	+	+
JJ-1725	*P. amylolyticus*	99.66	Soybean	Endophyte	7	Troy	Ohio	–	–

JJ-1729	*P. polymyxa*	100.00	Soybean	Endophyte	8	Troy	Ohio	–	–
JJ-1736	*P. typhae*	99.59	Soybean	Endophyte	8	Troy	Ohio	–	–
JJ-1742	*P. cucumis*	100.00	Soybean	Endophyte	9	Troy	Ohio	–	–
JJ-1743	*P. polymyxa*	99.93	Soybean	Endophyte	9	Troy	Ohio	+	–
JJ-1744	*P. typhae*	99.65	Soybean	Endophyte	9	Troy	Ohio	–	–
JJ-1747	*P. polymyxa*	99.65	Soybean	Endophyte	10	Troy	Ohio	+	+
JJ-1755	*P. graminis*	100.00	Soybean	Rhizoplane	5	Carrol	Iowa	–	–
JJ-1759	*P. catalpae*	99.66	Soybean	Rhizoplane	8	Carrol	Iowa	–	–
JJ-1762	*P. alginolyticus*	99.29	Soybean	Endophyte	4	Carrol	Iowa	–	–
JJ-1774	*P. susongensis*	99.21	Corn	Rhizoplane	2	Troy	Ohio	–	–
JJ-1781	*P. barcinonensis*	99.03	Corn	Rhizoplane	5	Troy	Ohio	–	–
JJ-1783	*P. sonchi*	98.64	Corn	Rhizoplane	6	Troy	Ohio	–	–
JJ-1789	*P. dongdonensis*	99.27	Soybean	Rhizoplane	2	Troy	Ohio	–	–
JJ-1817	*P. rhizoplanae*	98.35	Corn	Endophyte	6	Troy	Ohio	–	–
JJ-1819	*P. contaminans*	99.93	Soybean	Endophyte	1	Troy	Ohio	–	–
JJ-1831	*P. typhae*	97.33	Soybean	Endophyte	9	Troy	Ohio	–	–
JM-929	*P. odorifer*	99.51	Cotton	Endophyte	n.a	Tallassee	Alabama	–	–
JM-1337	*P. cucumis*	99.86	n.a	Endophyte	n.a	Tallassee	Alabama	–	–
JM-1355	*P. barcinonensis*	99.16	n.a	Endophyte	n.a	Tallassee	Alabama	–	–
JM-1368	*P. barcinonensis*	99.07	n.a	Endophyte	n.a	Tallassee	Alabama	–	–
JM-1419	*P. cucumis*	99.86	n.a	Endophyte	n.a	Tallassee	Alabama	–	–
JM-1420	*P. cucumis*	99.86	n.a	Endophyte	n.a	Tallassee	Alabama	–	–
AP-66	*P. silvae*	99.50	n.a	n.a	n.a	n.a	n.a	–	–

* Whole genome sequence is available

Table S2. List of *Paenibacillus* type strains used for the phylogenetic tree construction.

Strain	GenBank accession number
P. massiliensis subsp. panacisoli DSM 21345	AB245384.1
P. ehimensis DSM 11029	AY116665.1
P. illinoisensis DSM 11733	NR_115624.1
P. selenitireducens KCTC 33157	NR_133807.1
P. cucumis DSM 101601	NR_149778.1
P. thiaminolyticus DSM 7262	AB073197.1
P. terrigena DSM 21567	AB248087.1
P. granivorans	AF237682.1
P. turicensis DSM 14349	AF378694.1
P. agaridevorans DSM 1355	AJ345023.1
P. cineris DSM 16945	AJ575658.1
P. barcinonensis DSM 15478	AJ716019.1
P. lactis DSM 15596	AY257868.1
P. xylanilyticus DSM 17255	AY427832.1
P. alkaliterrae DSM 17040	AY960748.1
P. urinalis DSM 22281	EF212892.1
P. contaminans BCRC 17728	EF626690.1
P. taichungensis DSM 19942	EU179327.1
P. peoriae DSM 8320	EU391157.1
P. xylanexedens DSM 21292	EU558281.1
P. tundrae DSM 21291	EU558284.1
P. typhae DSM 25190	NR_109462.1
P. pocheonensis DSM 23906	NR_112565.1
P. alginolyticus DSM 5050	NR_115595.1
P. sonchi CCBAU 83901	NR_115751.1
P. aestuarii DSM 23861	NR_116365.1
P. catalpae DSM 24714	NR_118012.1
P. taohuashanense DSM 25809	NR_118393.1
P. sabinae T27	NR_121732.2
P. dongdonensis DSM 27607	NR_134112.1
P. susongensis JCM 19951	NR_134118.1
P. endophyticus LMG 27297	NR_135705.1
P. oenotherae JCM 19573	NR_136822.1

P. etheri DSM 29760	NR_148622.1
P. aceris DSM 24950	NR_156841.1
P. rhizoplanae DSM 103963	NR_156842.1
P. tritici CECT 9125	NR_157638.1
P. solisilvae JCM 32513	NR_175458.1
P. lautus DSM 3035	NZ_BIMF01000051.1
P. amylolyticus NBRC 15957	NZ_BIMJ01000009.1
P. rigui WPCB173	NZ_NMQW01000006.1

Table S3. Annotation of the known specialized metabolites detected in the extracts of 227 plant-associated *Paenibacillus* isolates. For the metabolites highlighted in bold, annotation was performed on the MS/MS level.

Parent ion *m/z*	Adduct ion	Compound	Family	Bioactivity
442.2839	$[M+2H]^{2+}$	Fusaricidin A	Fusaricidins	Gram +, Fungi (Kajimura & Kaneda, 1996)
449.2909	$[M+2H]^{2+}$	Fusaricidin B/Antibiotic LI-F05A/LI-F06A	Fusaricidins	Gram +, Fungi (Ryu *et al.*, 2017)
449.2921	$[M+2H]^{2+}$	Fusaricidin B/Antibiotic LI-F05A/LI-F06A	Fusaricidins	Gram +, Fungi (Ryu *et al.*, 2017)
456.300	$[M+2H]^{2+}$	Antibiotic LI-F05B/LI-F06B/LI-F08A	Fusaricidins	Gram +, Fungi (Ryu *et al.*, 2017)
463.3076	$[M+2H]^{2+}$	Antibiotic LI-F08B	Fusaricidins	Gram +, Fungi (Ryu *et al.*, 2017)
463.3077	$[M+2H]^{2+}$	Antibiotic LI-F08B	Fusaricidins	Gram +, Fungi (Ryu *et al.*, 2017)
466.2835	$[M+2H]^{2+}$	Antibiotic LI-F07A	Fusaricidins	Gram +, Fungi (Kajimura & Kaneda, 1996)
466.2848	$[M+2H]^{2+}$	Antibiotic LI-F07A	Fusaricidins	Gram +, Fungi (Kajimura & Kaneda, 1996)
473.2915	$[M+2H]^{2+}$	Antibiotic LI-F07B	Fusaricidins	Gram +, Fungi (Ryu *et al.*, 2017)
481.2895	$[M+2H]^{2+}$	Fusaricidine D	Fusaricidins	Gram +, Fungi (Ryu *et al.*, 2017)
883.5606	$[M+H]^{+}$	Fusaricidin A	Fusaricidins	Gram +, Fungi (Kajimura & Kaneda, 1996)

911.5916	$[M+H]^+$	Antibiotic LI-F08A	Fusaricidins	Gram +, Fungi (Kajimura & Kaneda, 1996)
911.5921	$[M+H]^+$	Antibiotic LI-F08A	Fusaricidins	Gram +, Fungi (Kajimura & Kaneda, 1996)
925.6076	$[M+H]^+$	Antibiotic LI-F08B	Fusaricidins	Gram +, Fungi (Ryu *et al.*, 2017)
945.5767	$[M+H]^+$	Antibiotic LI-F07B	Fusaricidins	Gram +, Fungi (Ryu *et al.*, 2017)
947.5557	$[M+H]^+$	Fusaricidin C	Fusaricidins	Gram +, Fungi (Kajimura & Kaneda, 1996)
961.5712	$[M+H]^+$	Fusaricidine D	Fusaricidins	Gram +, Fungi (Ryu *et al.*, 2017)
961.5718	$[M+H]^+$	Fusaricidine D	Fusaricidins	Gram +, Fungi (Ryu *et al.*, 2017)
381.9119	$[M+3H]^{3+}$	Polymyxin M2/A1	Polymyxins	Gram – (Martin *et al.*, 2003)
385.9241	$[M+3H]^{3+}$	Polymyxin E	Polymyxins	Gram – (Ikai *et al.*, 1998)
386.5838	$[M+3H]^{3+}$	Polymyxin M1/A2	Polymyxins	Gram – (Martin *et al.*, 2003)
386.5839	$[M+3H]^{3+}$	Polymyxin M1/A2	Polymyxins	Gram – (Martin *et al.*, 2003)
390.5958	$[M+3H]^{3+}$	Polymyxin E	Polymyxins	Gram – (Ikai *et al.*, 1998)
390.596	$[M+3H]^{3+}$	Polymyxin E	Polymyxins	Gram – (Ikai *et al.*, 1998)

397.9119	$[M+3H]^{3+}$	Polymyxin P1	Polymyxins	Gram – (Niu *et al.*, 2013)
401.9238	$[M+3H]^{3+}$	Polymyxin B	Polymyxins	Gram – (Pittenauer *et al.*, 2006)
579.371	$[M+2H]^{2+}$	Polymyxin M1	Polymyxins	Gram – (Martin *et al.*, 2003)
579.3713	$[M+2H]^{2+}$	Polymyxin M1	Polymyxins	Gram – (Martin *et al.*, 2003)
585.3898	$[M+2H]^{2+}$	Polymyxin E7	Polymyxins	Gram – (Ikai *et al.*, 1998)
596.3656	$[M+2H]^{2+}$	Polymyxin P1	Polymyxins	Gram – (Niu *et al.*, 2013)
602.382	$[M+2H]^{2+}$	Polymyxin B5	Polymyxins	Gram – (Pittenauer *et al.*, 2006)
610.3812	$[M+2H]^{2+}$	Polymyxin B6	Polymyxins	Gram – (Pittenauer *et al.*, 2006)
486.9491	$[M+3H]^{3+}$	Tridecaptin B_1	Tridecaptins	Gram – (Cochrane *et al.*, 2015)
496.9524	$[M+3H]^{3+}$	Tridecaptin M	Tridecaptins	Gram – (Jangra *et al.*, 2019b)
744.9239	$[M+2H]^{2+}$	Tridecaptin M	Tridecaptins	Gram – (Jangra *et al.*, 2019b)
789.9477	$[M+2H]^{2+}$	Tridecaptin A_3	Tridecaptins	Gram – (Lohans *et al.*, 2012)
539.9715	$[M+3H]^{3+}$	Tridecaptin A_5	Tridecaptins	Gram -, Gram +

Table S4. NMR data of paenilipoheptin A (1), measured in DMSO-d_6 at 298 K*.

	Residue	NH	H_α (C_α, type)	H_β (C_β, type)	Other
1	**Fatty acid**	7.20	2.21 (41.7, CH_2)	4.00 (47.2, CH)	1 C: (170.4) 4 CH_2: 1.39, 1.33 (35.11) 5-8 CH_2: 1.05-1.25 (25.5) 9 CH_2: 1.20, 1.00 (36.30) 10 CH: 1.25 (33.8) 11 CH_2: 1.26, 1.07 (29.0) 12 CH_3: 0.81 (11.18) 13 CH_3: 0.79 (19.07)
2	**1-Ser**	7.74	4.48 (54.4, CH)	3.44, 3.41 (61.63, CH_2)	1C: ND OH: 4.80
3	**2-Lys**	8.42	3.90 (54.7, CH)	1.49 (30.0, CH_2)	1C: ND 4 CH_2: 1.16, 1.05 (22.3) 5 CH_2: 1.39 (26.66) 6 CH_2: 2.62 (38.8) NH_2: ND
4	**3-Trp**	8.02	4.48 (54.4, CH)	3.13, 3.06 (27.17, CH_2)	1 C: ND 4 C: (110.5) 5 CH: 7.14 (123.5) 6 C: (136.1) 7 CH: 7.33 (111.3) 8 CH: 7.06 (120.9) 9 CH: 6.98 (118.27) 10 CH: 7.54 (118.3) 11 C: (127.2) NH: 10.81
5	**4-Ala**	7.70	4.03 (49.14, CH)	1.09 (17.85, CH_3)	1 C: (172.1)
6	**5-Phe**	7.92	4.33 (54.7, CH)	3.05, 2.90 (37.2, CH_2)	1 C: ND 4 C: 138.1 5 & 9 CH: 7.29 (129.5) 6 & 8 CH: 7.22 (128.0) 7 CH: 7.14 (126.1)
7	**6-Tyr**	8.24	4.25 (55.9, CH)	2.89, 2.85 (36.0, CH_2)	1 C: (170.75) 4 C: (127.05) 5 & 9 CH: 7.05 (130.1) 6 & 8 CH: 6.67 (115.0) 7 C: (156.3) OH: 9.22
8	**7-Glu**	7.95	4.11 (51.8, CH)	1.88, 1.57 (26.6, CH_2)	1 C: ND 4 CH_2: 2.00 (29.9) 5 C: (174.0)

* ^{1}H 850 MHz and ^{13}C 212.5 MHz

ND: not determined under these experimental conditions.

Table S5. NMR data of paenilipoheptin D (2), measured in DMSO-d_6 at 298 K*.

	Residue	NH	H_α(C_α, type)	H_β(C_β, type)	Other
1	**Fatty acid**	8.20	2.23, 2.17 (43.5, CH_2)	3.87 (47.1, CH)	4 CH_2: 1.49 (34.9) 5-8 CH_2: 1.23 (28.7) 9 CH_2: 1.24 (35.7) 10 CH: 1.28 (33.5) 11 CH_2: 1.26, 1.07 (29.0) 12 CH_3: 0.83 (10.9) 13 CH_3: 0.82 (18.8)
2	**1-Ser**	7.58	4.53 (55.3, CH)	3.79, 3.65 (62.0, CH_2)	
3	**2-Lys**	ND	3.83 (56.3, CH)	1.59, 1.53 (29.5, CH_2)	4 CH_2: 1.21, 1.12 (22.2) 5 CH_2: 1.34 (29.6) 6 CH_2: ND NH_2: ND
4	**3-Trp**	8.60	4.23 (54.8, CH)	3.16, 3.09 (25.7, CH_2)	4 C: ND 5 CH: 7.26, (123.5) 6 NH: 10.8 7 C: 135.8 8 CH: 7.32 (111.1) 9 CH: 7.04 (120.6) 10 CH: 6.96 (118.1) 11 CH: 7.50 (117.9) 12 C: 126.9
5	**4-Ala**	7.16	4.15 (47.7, CH)	1.12 (16.9, CH_3)	
6	**5-Leu**	7.23	3.89 (52.1, CH)	1.35, 1.05 (40.0, CH_2)	4 CH: 1.45 (23.8) 5 CH_3: 0.68 (20.8) 6 CH_3: 0.78 (22.4)
7	**6-Tyr**	7.65	4.53 (51.7, CH)	2.85, 2.65 (37.3, CH_2)	4 C: 127.7 5 & 9 CH: 6.92 (129.8) 6 & 8 CH: 6.56 (114.3) 7 C: 155.5
8	**7-Dab**	ND	3.91 (52.8, CH)	1.65 (33.2, CH_2)	4 CH_2: 2.56 (37.4) NH_2: ND

* ^{1}H 850 MHz and ^{13}C chemical shifts inferred from HSQC and HMBC spectra

ND: not determined under these experimental conditions (and also all carbonyl carbons).

Table S6. NMR data of paenitracin A (3), measured in DMSO-d_6 at 298 K*.

	Residue	NH	H_α(C_α, type)	H_β(C_β, type)	Other
1	**1-Ile**	ND	3.55,d (57.0, CH)	1.66 (39.5, CH)	1 C: ND 4 CH_2: 1.41, 1.15 (25.5) 5 CH_3: 0.84 (11.3) 6 CH_3: 0.81 (13.3)
2	**2-Cys**		5.13 (78.5, CH)	3.41 (32.7, CH_2)	
3	**3-Leu**	7.75	4.35 (51.9, CH)	1.52, 1.46 (41.0, CH_2)	4 CH: 1.54 (24.0) 5 CH_3: 0.85 (21.2) 6 CH_3: 0.89 (22.8)
4	**4-Glu**	8.18	4.06 (53.6, CH)	1.81, 1.75 (27.7, CH_2)	4 CH_2: 2.19 (32.8, CH_2)
5	**5-Ile**	7.22	3.95 (57.6, CH)	1.75 (35.4, CH)	4 CH_2: 1.46, 1.17 (24.3) 5 CH_3: 0.78 (10.8) 6 CH_3: 0.94 (15.0)
6	**6-Lys**	8.46	4.49 (52.9, CH)	1.42, 1.15 (25.5, CH_2)	4 CH_2: 1.54 (24.0) 5 CH_2: 1.47 (24.3) 6 CH_2: 2.95, 2.82 (36.8) NH: 7.52
7	**7-Leu**	1037	4.22 (49.8, CH)	1.61, 1.36 (38.7, CH_2)	4 CH: 1.51 (24.1) 5 CH_3: 0.77 (20.4) 6 CH_3: 0.85 (22.9)
8	**8-Ile**	7.09	4.26 (55.6, CH)	1.68 (37.5, CH)	4 CH_2: 1.28, 0.70 (23.9) 5 CH_3: 0.70 (11.5) 6 CH_3: 0.66 (14.1)
9	**9-Trp**	8.98	5.18 (52.2, CH)	3.11, 2.83 (28.5, CH_2)	4 C: ND 5 CH: 7.18 (123.2) 6 NH: 10.9 7 C: 135.5 8 CH: 7.26 (110.7) 9 CH: 7.01 (120.3) 10 CH: 6.95 (117.7) 11 CH: 7.48 (117.8) 12 C: 126.7
10	**10-Thr**	10.78	3.66 (60.4, CH)	4.02 (66.8, CH)	4 CH_3: 0.67 (18.5)
11	**11-Asp**	8.84	4.09 (50.2, CH)	2.43, 2.01 (38.2, CH_2)	
12	**12-Asn**	7.82	3.94 (51.8, CH)	2.38, 2.18 (35.2, CH_2)	NH_2: 7.58, 6.74

* ^{1}H 850 MHz and ^{13}C chemical shifts inferred from HSQC and HMBC spectra

ND: not determined under these experimental conditions (and also all carbonyl carbons).

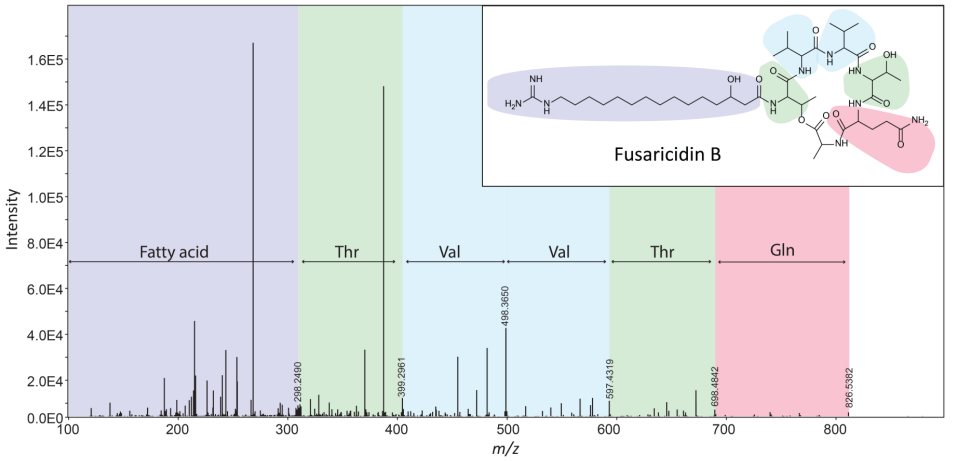

Figure S1. MS/MS spectrum of fusaricidin B (precursor ion $[M+2H]^{2+}$ *m/z* 449.2909). The assignment of the sequence of amino acid residues is based on the mass differences between the consecutive *b* ions.

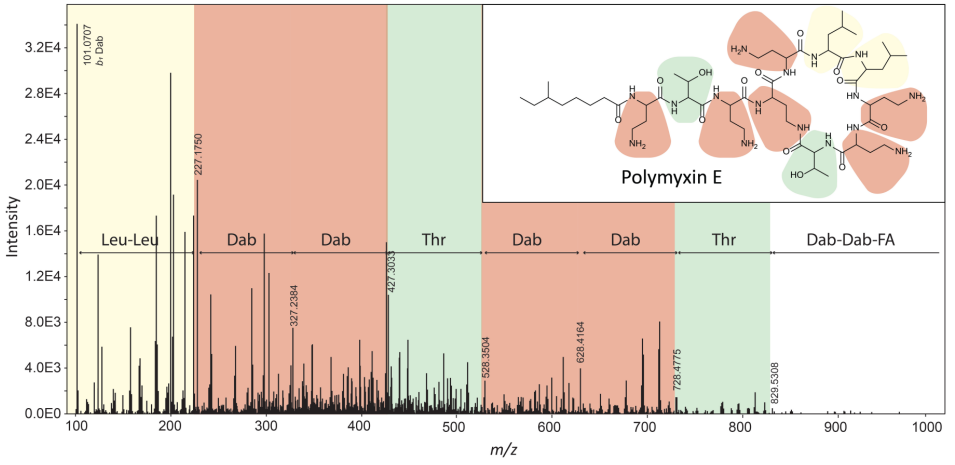

Figure S2. MS/MS spectrum of polymyxin E (precursor ion $[M+3H]^{3+}$ *m/z* 390.596). The assignment of the sequence of amino acid residues is based on the mass differences between the consecutive *b* ions.

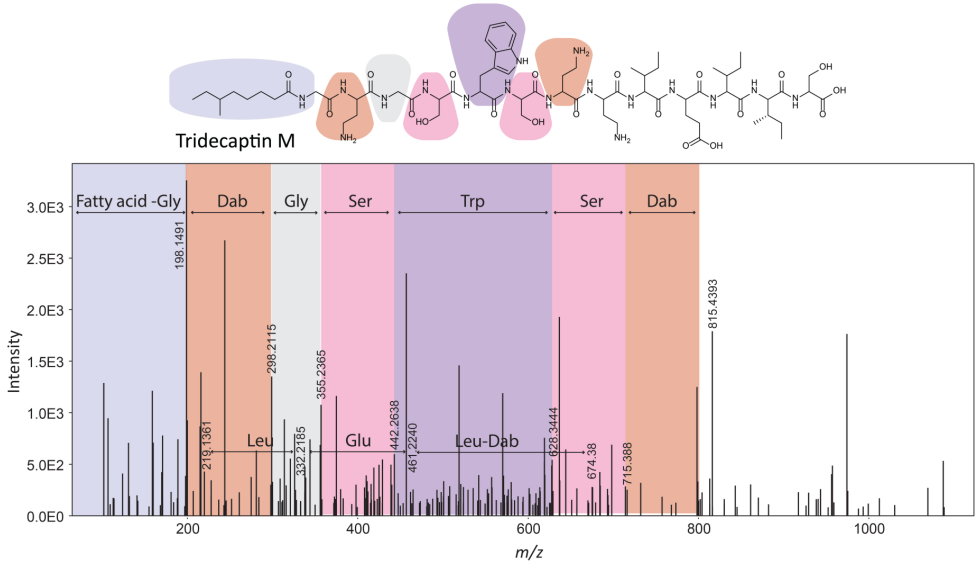

Figure S3. MS/MS spectrum of tridecaptin M (precursor ion $[M+3H]^{3+}$ *m/z* 496.9524). The assignment of the sequence of amino acid residues is based on the mass differences between the consecutive *b* ions.

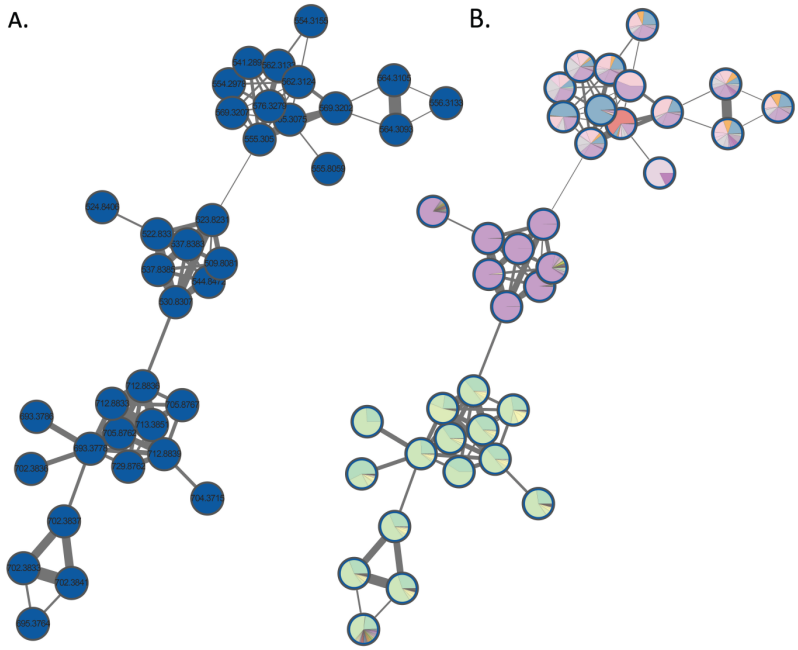

Figure S4. A. The MassQL-annotated molecular family of lysine-containing natural products. **B.** On the same molecular family, pie charts were mapped to the nodes to represent the relative precursor ion intensities within each extract.

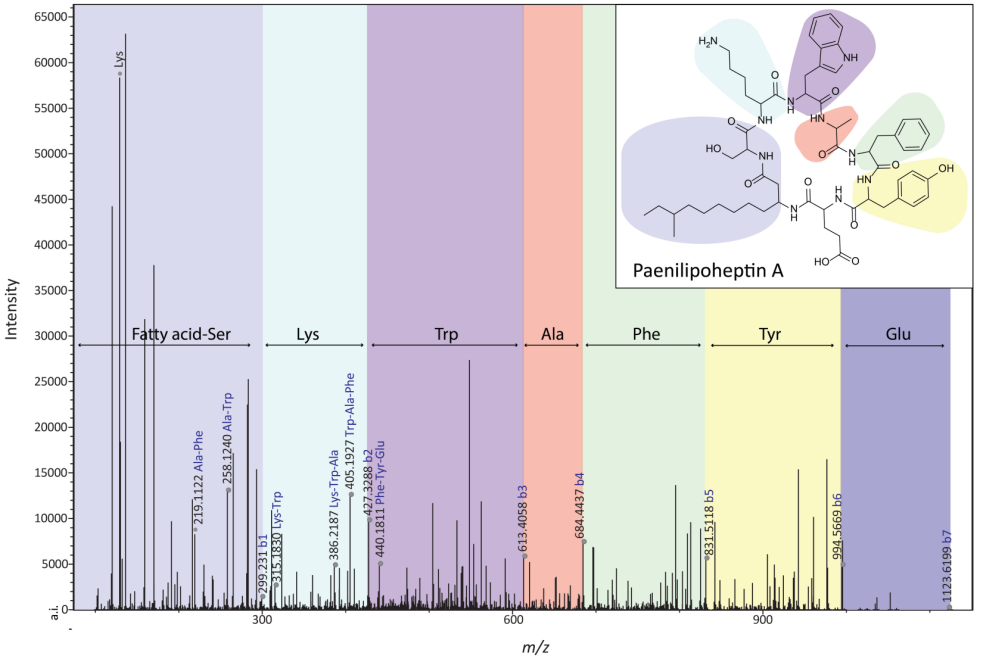

Figure S5. MS/MS spectrum of paenilipoheptin A (precursor ion $[M+2H]^{2+}$ *m/z* 562.3130). The assignment of the sequence of amino acid residues is based on the mass differences between the consecutive *b* ions.

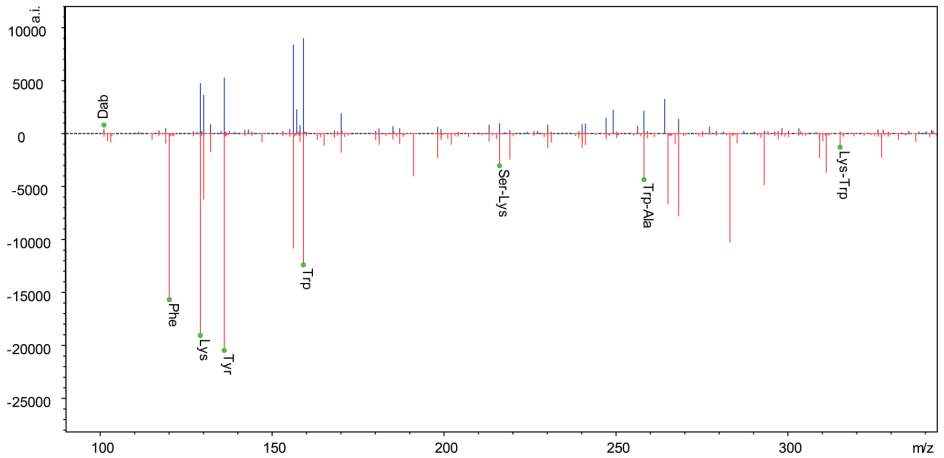

Figure S6. Direct MS/MS spectra comparison of the mass feature having *m/z* of 555.305 (red) with the mass feature having *m/z* of 523.8231 (blue).

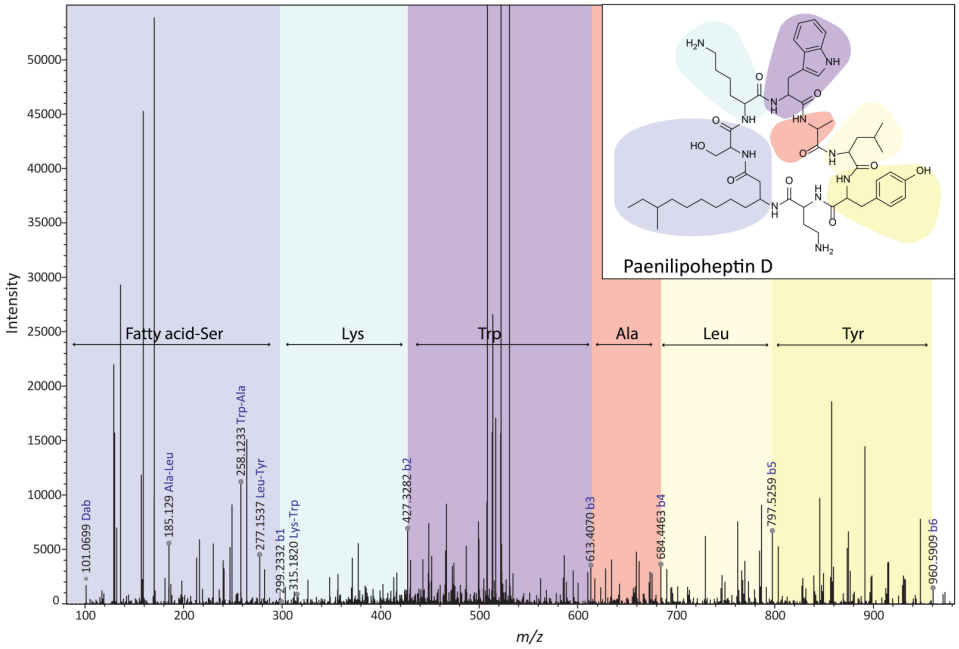

Figure S7. MS/MS spectrum of paenilipoheptin D (precursor ion $[M+2H]^{2+}$ *m/z* 530.8309). The assignment of the sequence of amino acid residues is based on the mass differences between the consecutive *b* ions.

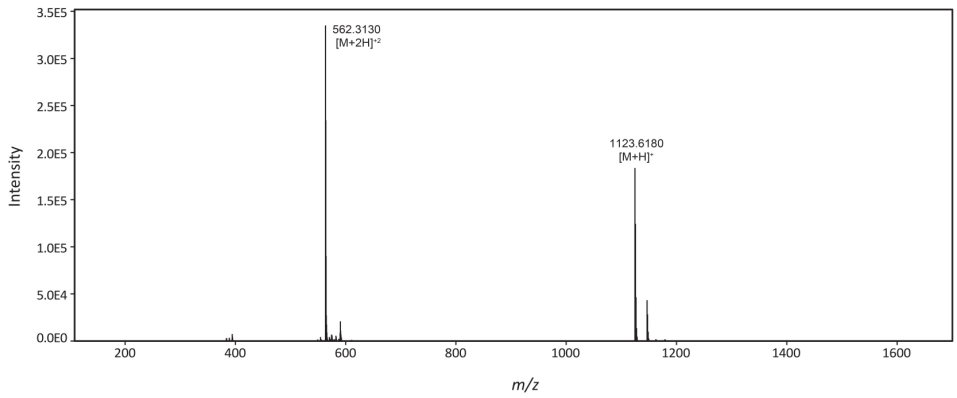

Figure S8. HRMS spectrum of 1.

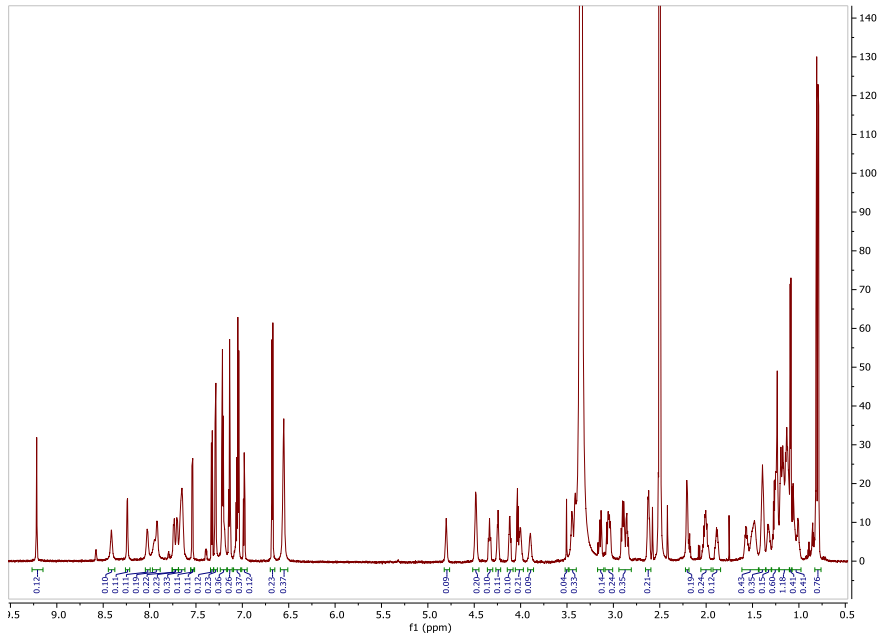

Figure S9. ^{1}H NMR spectrum of 1 (850 MHz, in DMSO-d$_6$).

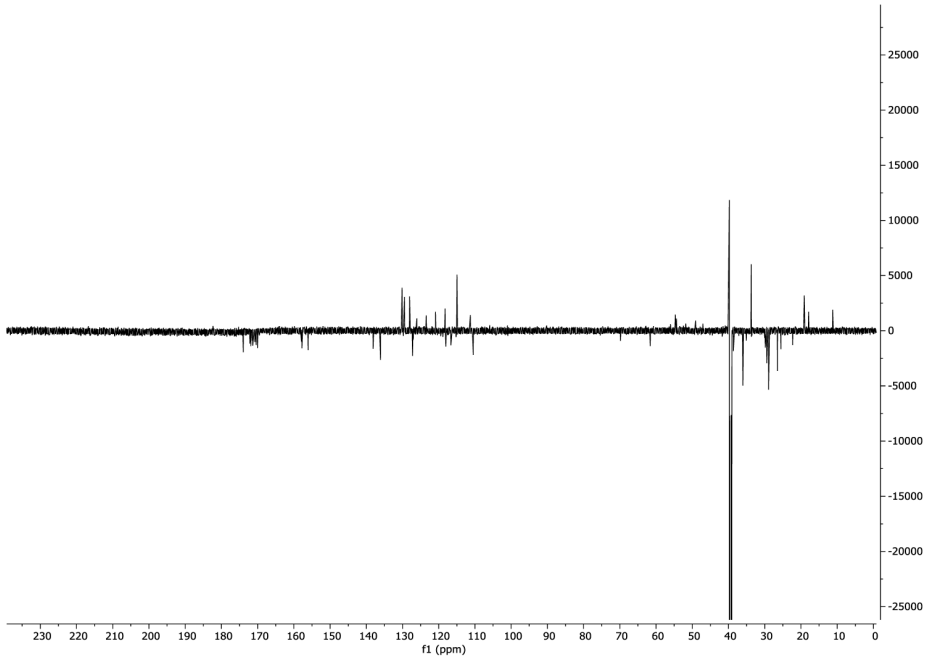

Figure S10. APT spectrum of 1 (850 MHz, in DMSO-d$_6$).

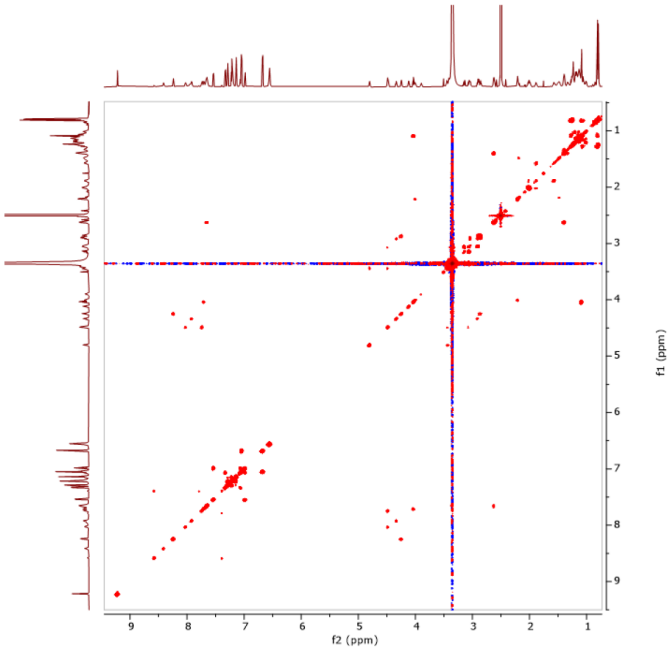

Figure S11. $^{1}H–^{1}H$ COSY spectrum of 1 (850 MHz, in DMSO-d_6).

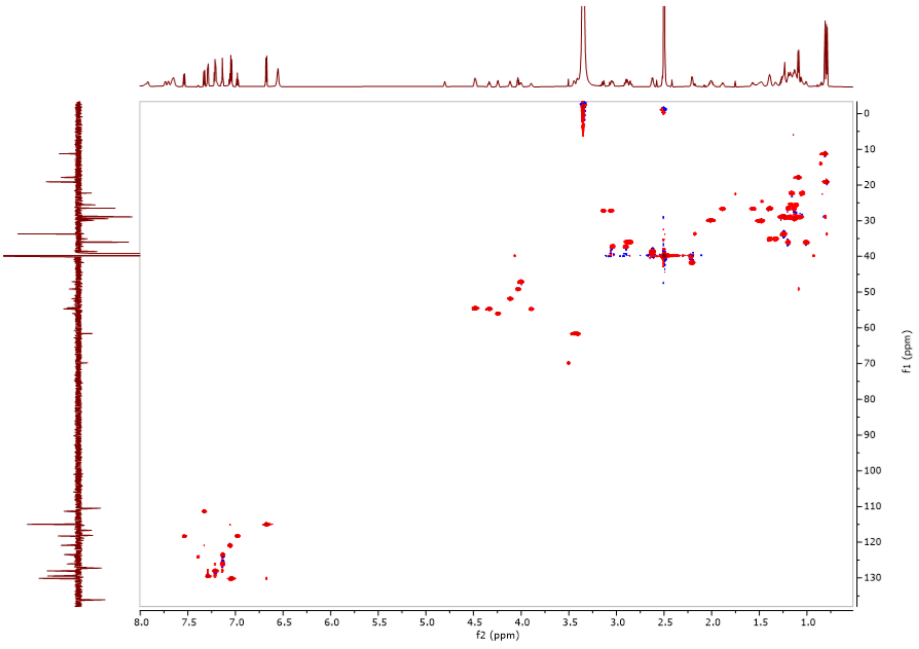

Figure S12. HSQC spectrum of 1 (850 MHz, in DMSO-d_6).

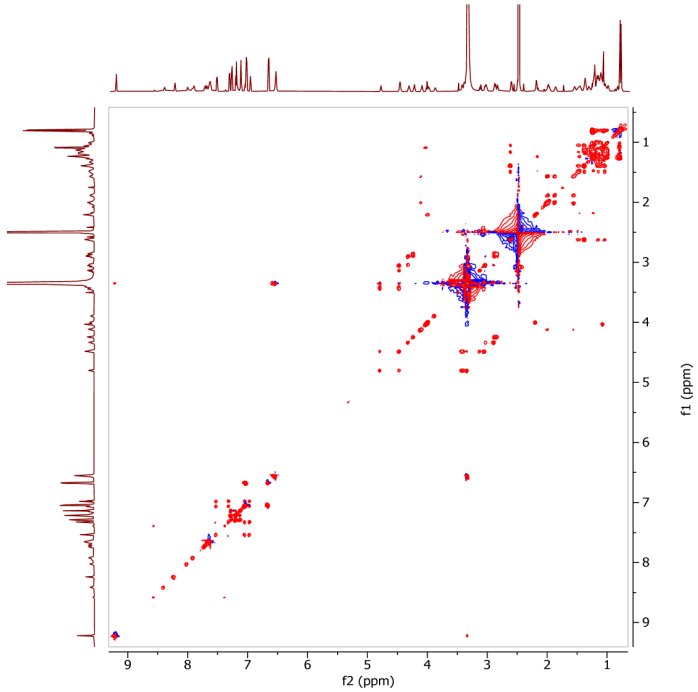

Figure S13. $^{1}H–^{1}H$ TOCSY spectrum of 1 (850 MHz, in DMSO-d_6).

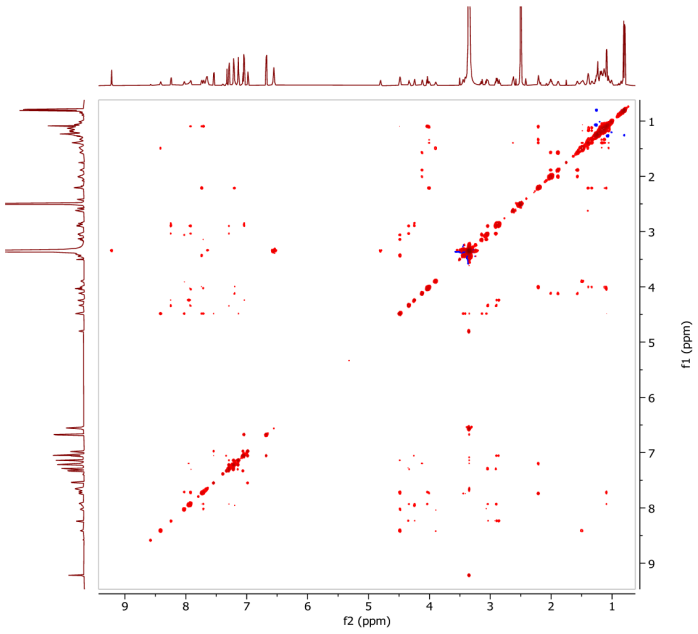

Figure S14. NOESY spectrum of 1 (850 MHz, in DMSO-d_6).

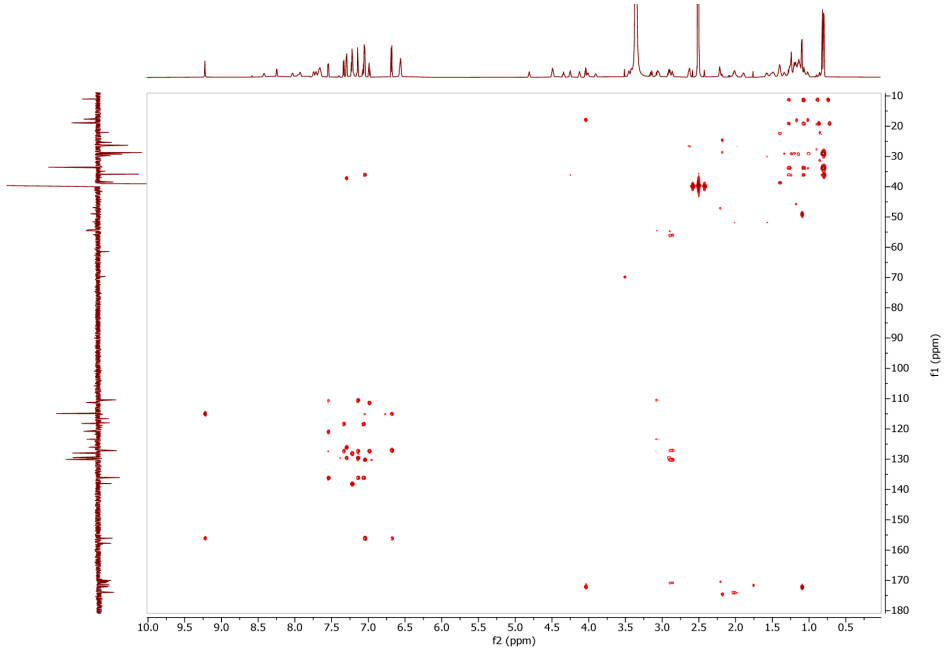

Figure S15. HMBC spectrum of 1 (850 MHz, in DMSO-d_6).

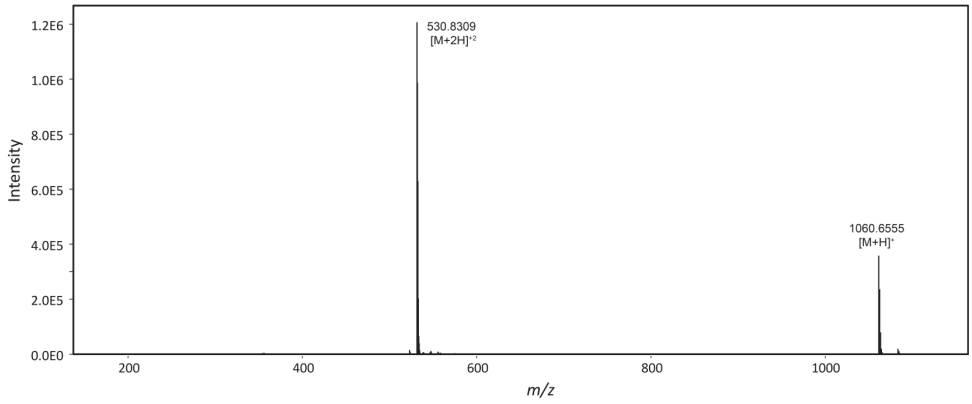

Figure S16. HRMS spectrum of 2.

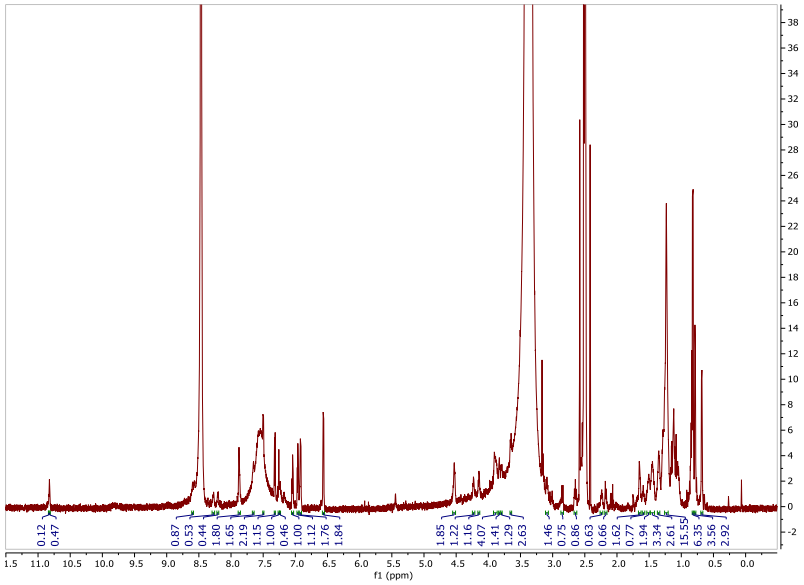

Figure S17. ^{1}H NMR spectrum of 2 (850 MHz, in DMSO-d_6).

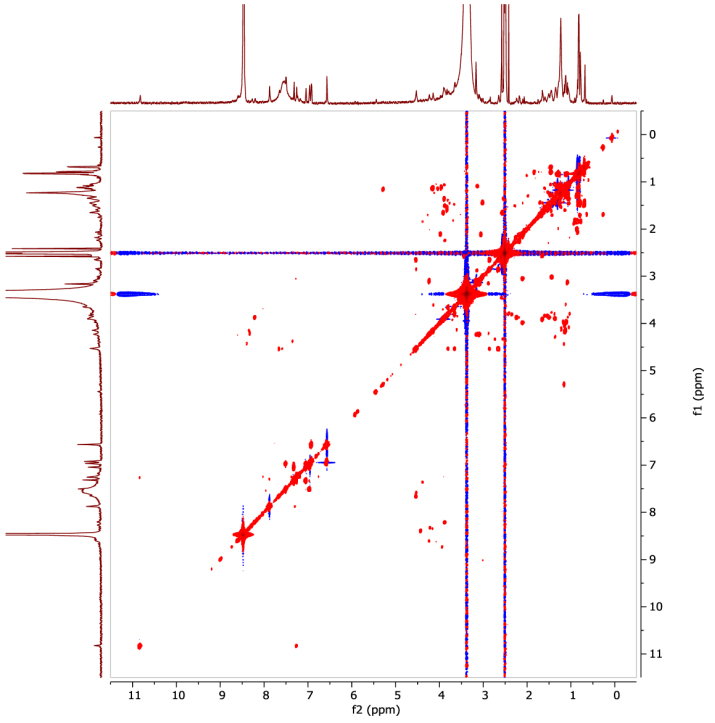

Figure S18. ^{1}H–^{1}H COSY spectrum of 2 (850 MHz, in DMSO-d_6).

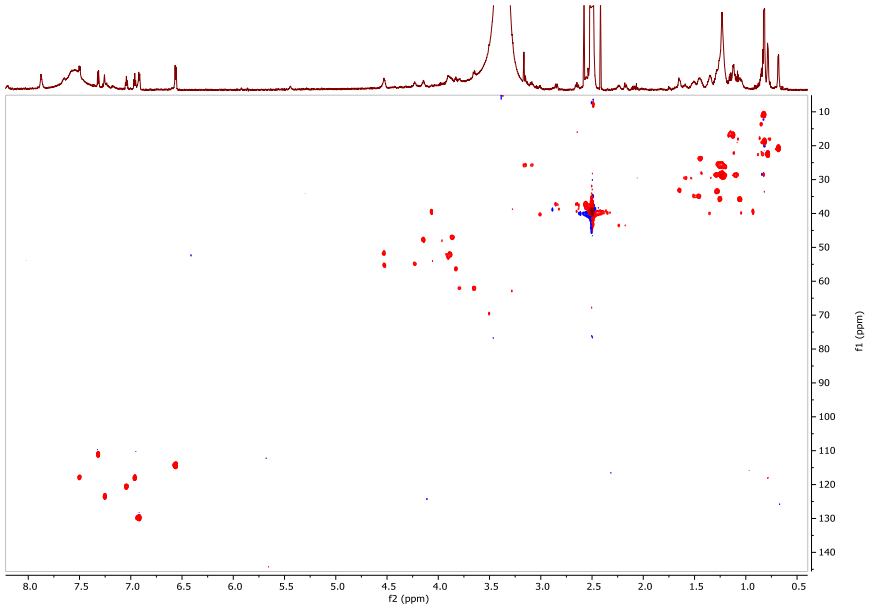

Figure S19. HSQC spectrum of 2 (850 MHz, in DMSO-d_6).

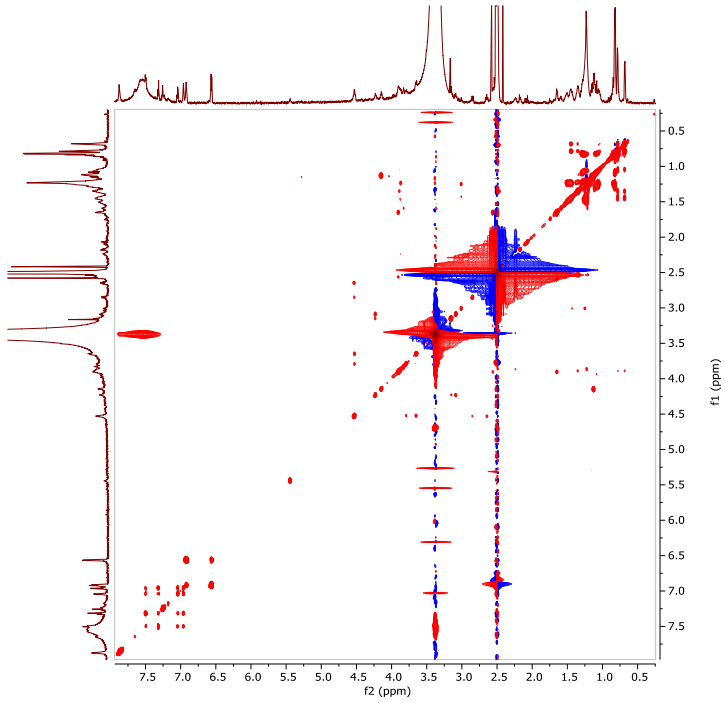

Figure S20. ^{1}H–^{1}H TOCSY spectrum of 2 (850 MHz, in DMSO-d_6).

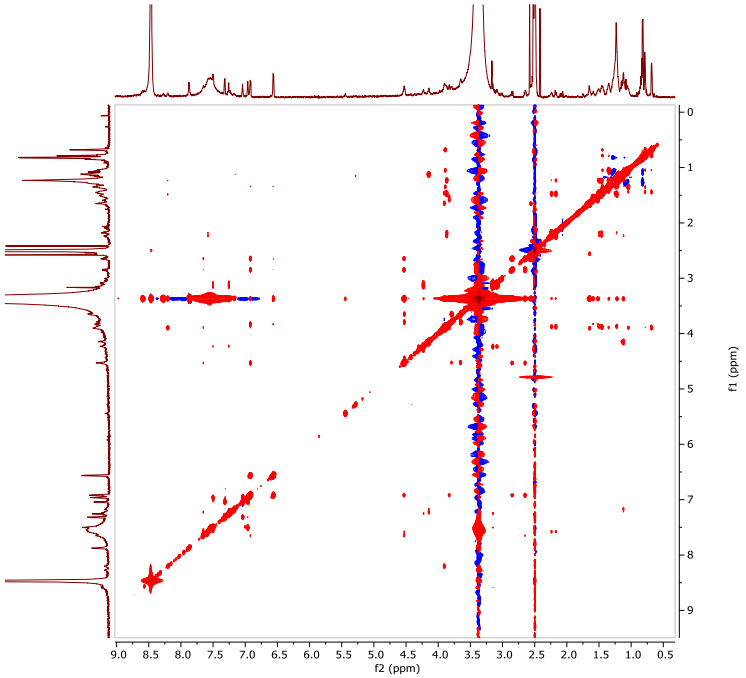

Figure S21. NOESY spectrum of 2 (850 MHz, in DMSO-d_6).

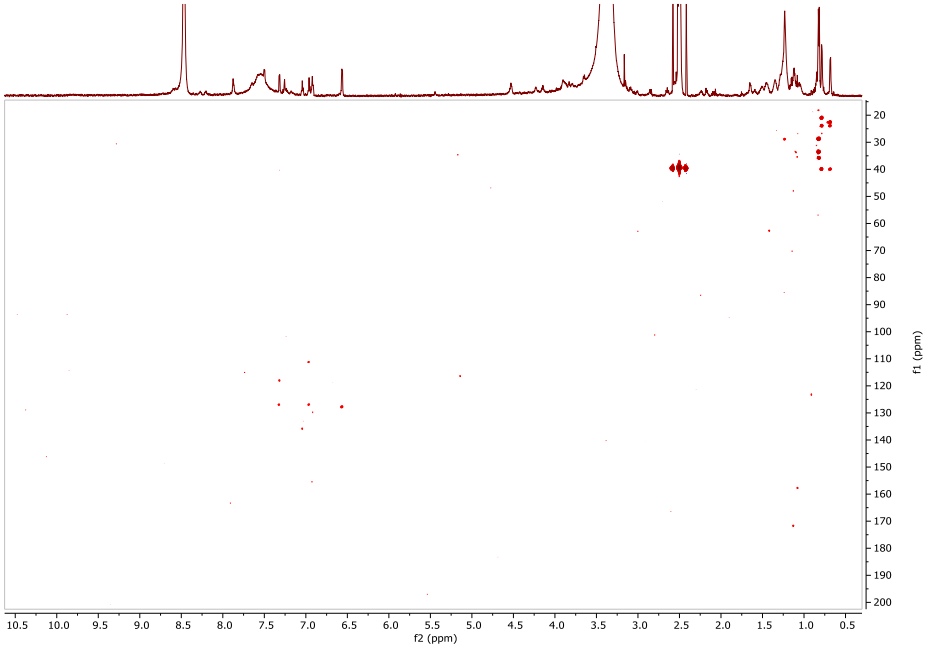

Figure S22. HMBC spectrum of 2 (850 MHz, in DMSO-d_6).

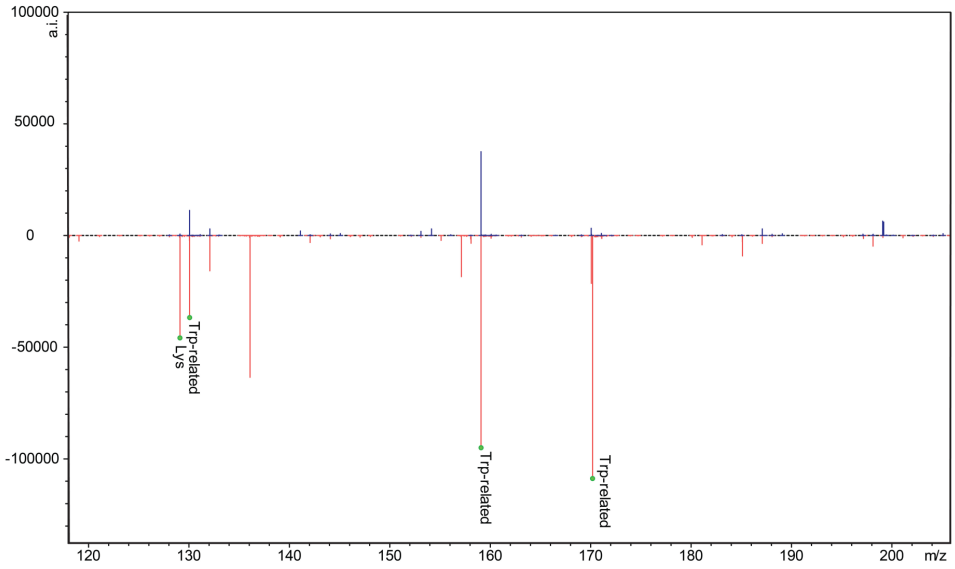

Figure S23. Direct MS/MS spectra comparison of the mass feature with *m/z* 712.8836 (blue) with the mass feature with m/z 530.8307 (red).

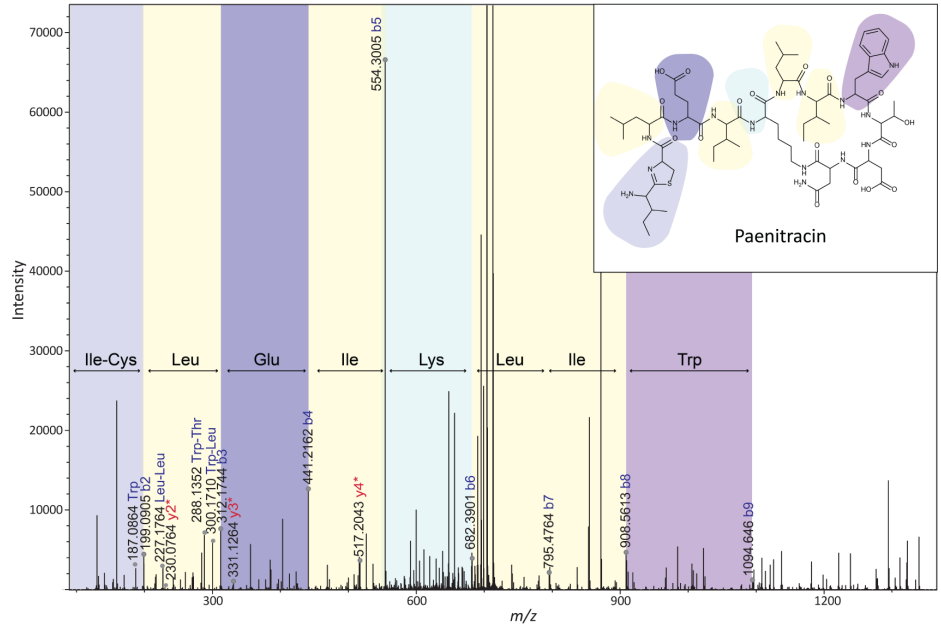

Figure S24. MS/MS spectrum of paenitracin A (precursor ion $[M+2H]^{2+}$ *m/z* 712.8843). The assignment of the sequence of amino acid residues is based on the mass differences between the consecutive *b* ions.

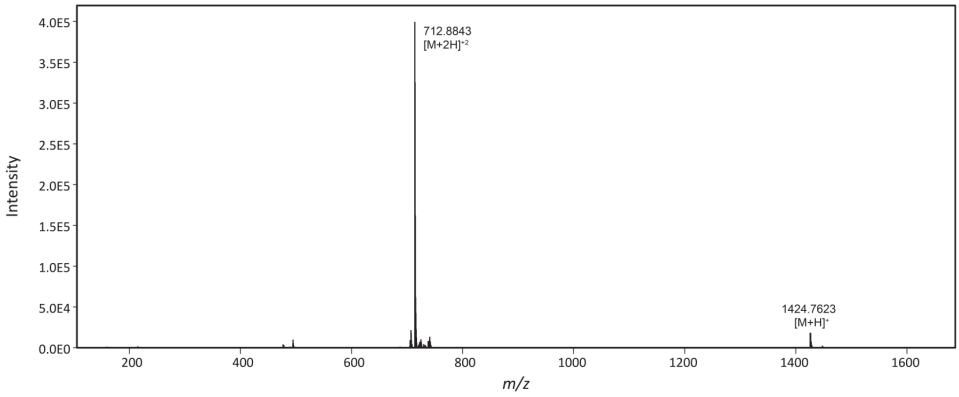

Figure S25. HRMS spectrum of 3.

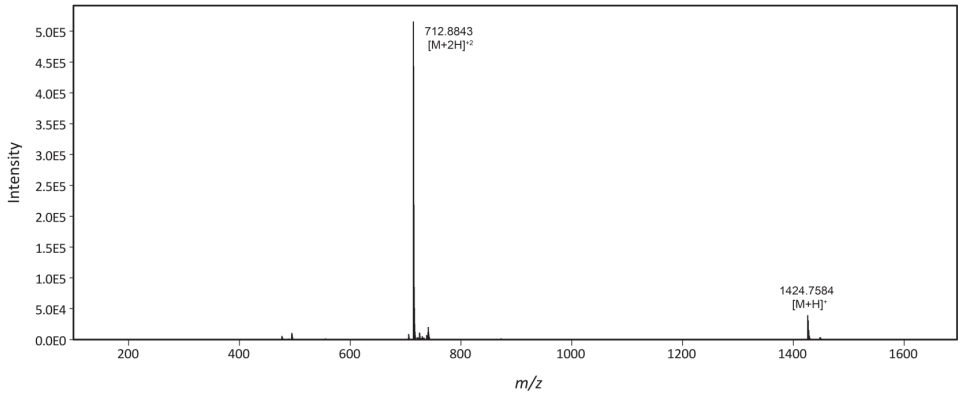

Figure S26. HRMS spectrum of 4.

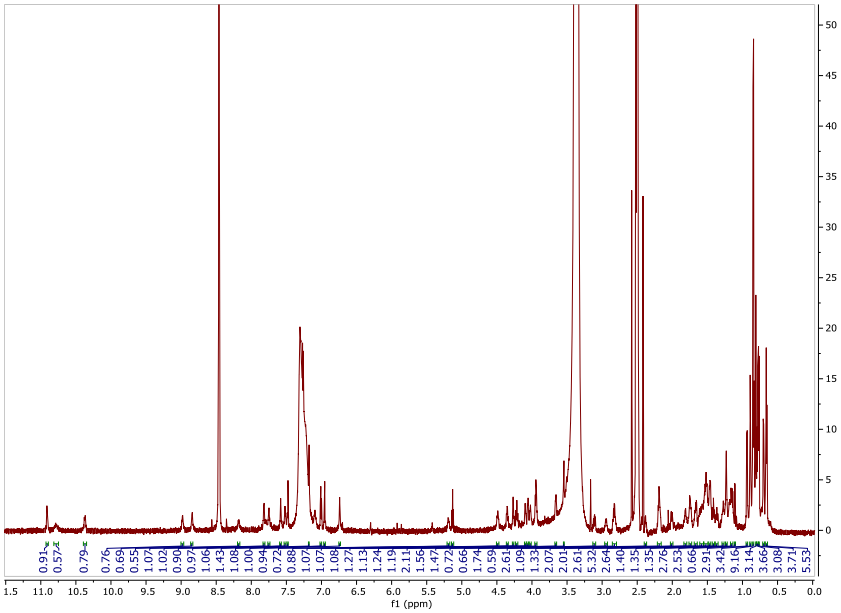

Figure S27. ^{1}H NMR spectrum of 3 (850 MHz, in DMSO-d$_6$).

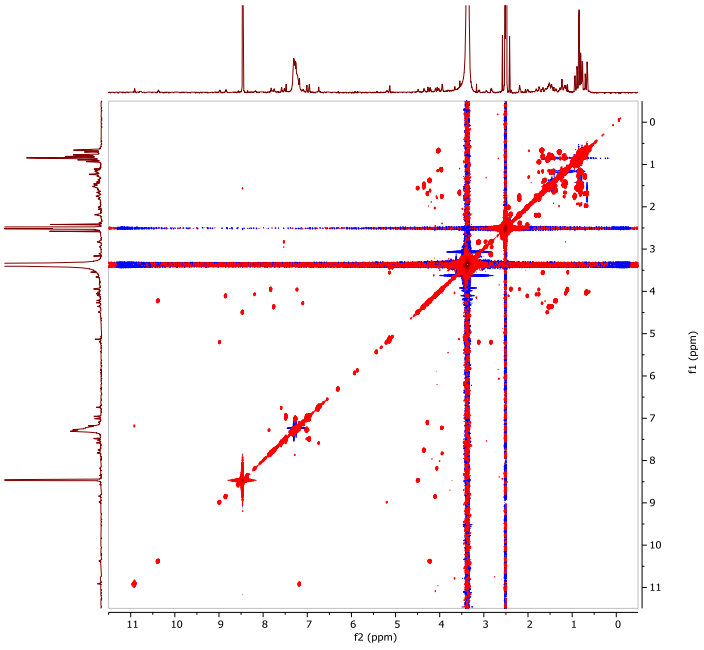

Figure S28. ^{1}H–^{1}H COSY spectrum of 3 (850 MHz, in DMSO-d$_6$).

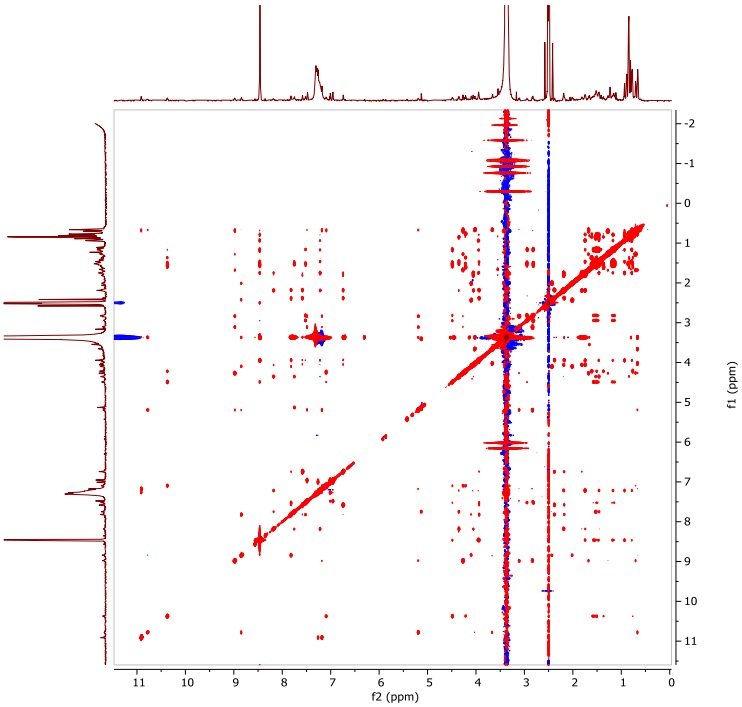

Figure S29. NOESY spectrum of 3 (850 MHz, in DMSO-d_6).

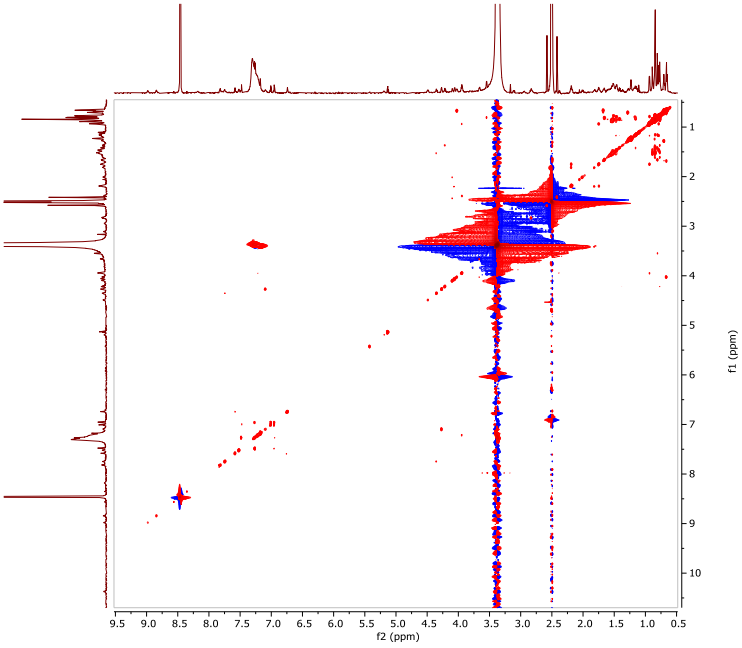

Figure S30. ^{1}H–^{1}H TOCSY spectrum of 3 (850 MHz, in DMSO-d_6).

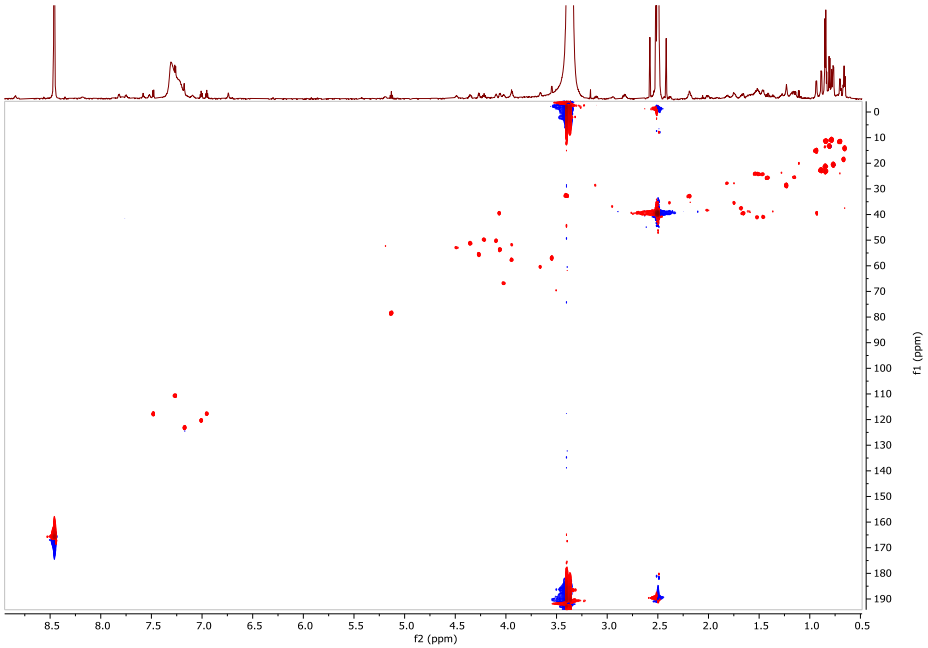

Figure S31. HSQC spectrum of 3 (850 MHz, in DMSO-d_6).

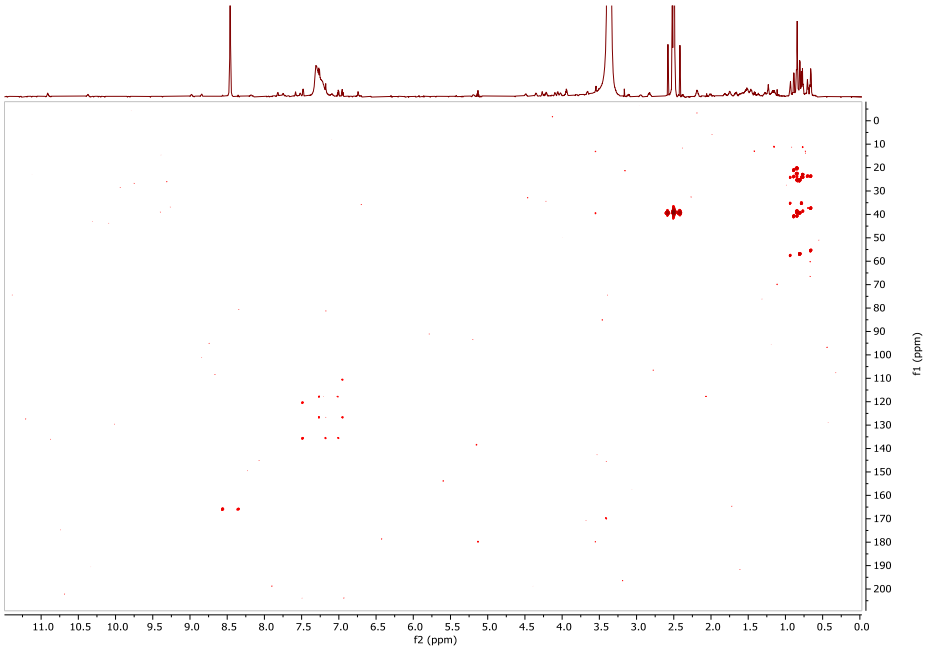

Figure S32. HBMC spectrum of 3 (850 MHz, in DMSO-d_6).

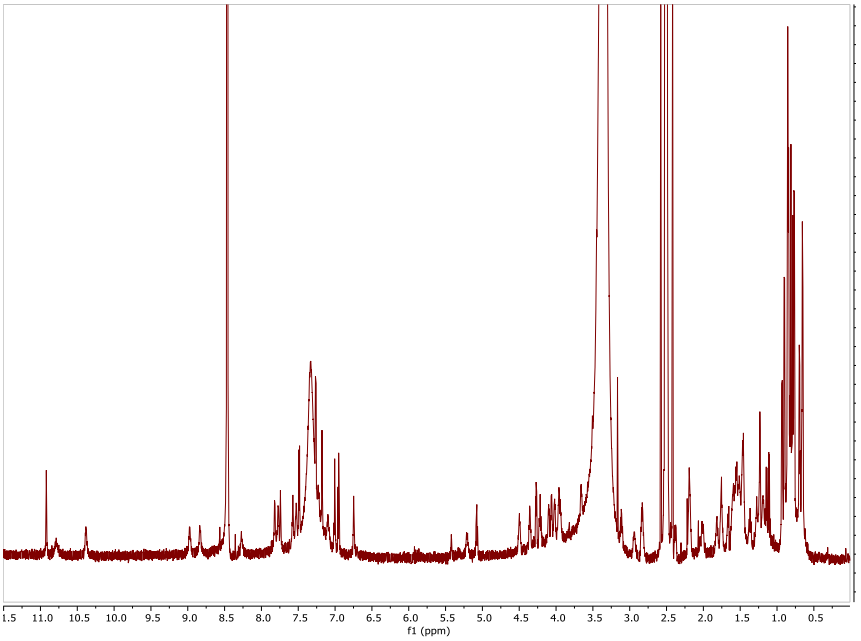

Figure S33. ^{1}H NMR spectrum of 4 (850 MHz, in DMSO-d$_6$).

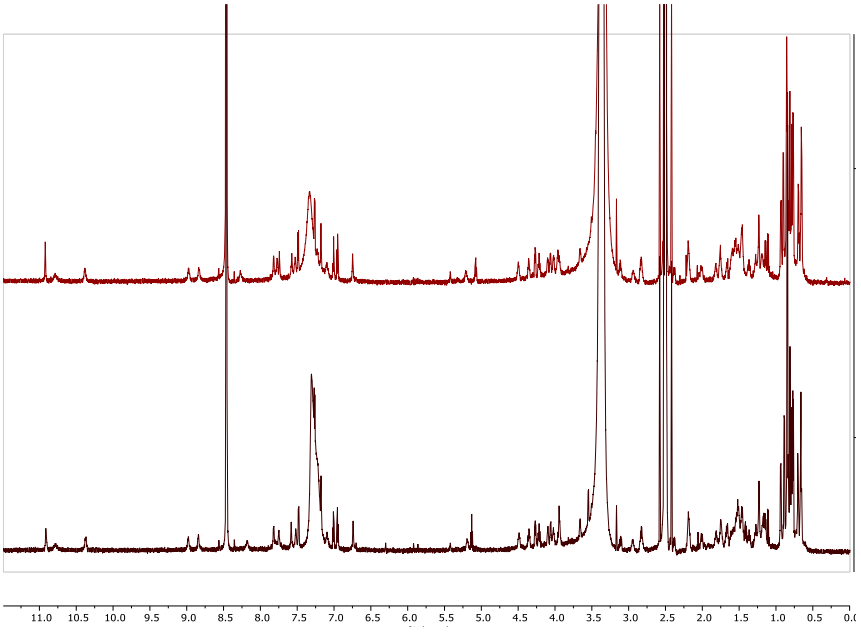

Figure S34. Stacked ^{1}H NMR spectra of 3 and 4 (850 MHz, in DMSO-d$_6$).