

Universiteit Leiden
The Netherlands

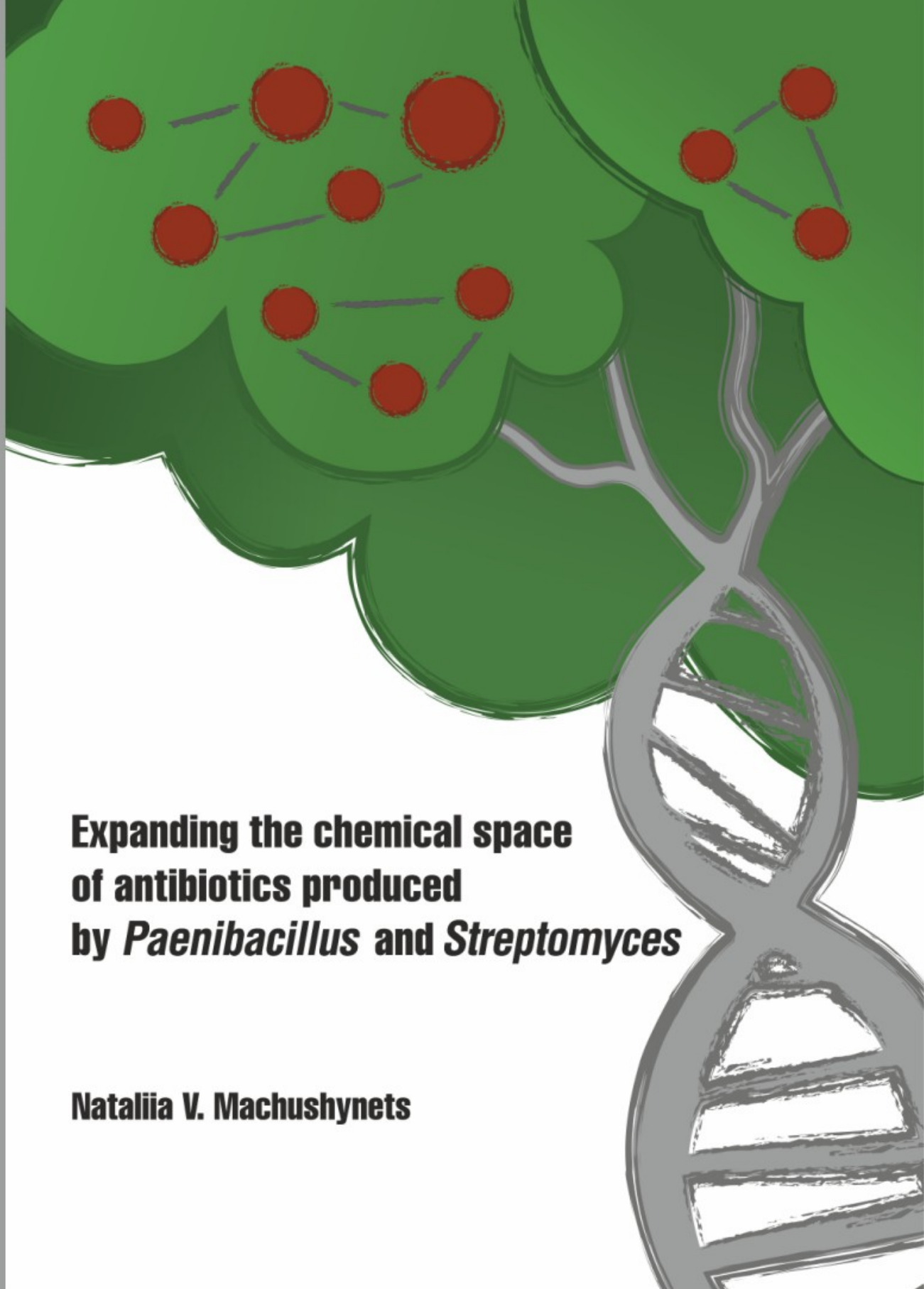
Expanding the chemical space of antibiotics produced by Paenibacillus and Streptomyces

Machushynets, N.V.

Citation
Machushynets, N. V. (2024, September 5). *Expanding the chemical space of antibiotics produced by Paenibacillus and Streptomyces*. Retrieved from https://hdl.handle.net/1887/4082475

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4082475

Note: To cite this publication please use the final published version (if applicable).



Expanding the chemical space of antibiotics produced by *Paenibacillus* and *Streptomyces*

Nataliia V. Machushynets

Expanding the chemical space of antibiotics produced by *Paenibacillus* and *Streptomyces*

Nataliia V. Machushynets

The work was supported by the NACTAR program of The Netherlands Organization for Scientific Research (NWO), Grant 16440.

Cover design: N.V. Machushynets, O.V. Podolets
Layout design: I.V. Zakharchuk
Printed by: Vezha Print, www.vezhaprint.com.ua

Copyright @ 2024, Nataliia V. Machushynets

Expanding the chemical space of antibiotics produced by *Paenibacillus* and *Streptomyces*

Proefschrif

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op donderdag 5 september 2024
klokke 13:00 uur

door

Nataliia V. Machushynets

geboren te Litohoscha, Oekraïne
in 1993

Promotores:

Prof. dr. G.P. van Wezel

Prof. dr. N.I. Martin

Promotiecommissie:

Prof. dr. A.H. Meijer

Prof. dr. H.P. Spaink

Prof. dr. A.T. Kovács

Prof. dr. J. Masschelein (Katholieke Universiteit Leuven, België)

Prof. dr. M.R. Liles (Auburn University, USA)

Contents

Chapter 1.	Thesis outline	6
Chapter 2.	General introduction	12
Chapter 3.	Phylogeny and antibiotic producing potential of plant-associated *Paenibacillus*	26
Chapter 4.	Discovery of novel glycerolated quinazolinones from *Streptomyces s*p. MBT27	86
Chapter 5.	Discovery of actinomycin L, a new member of the actinomycin family of antibiotics	112
Chapter 6.	nanoRAPIDS as an analytical pipeline for the discovery of novel bioactive metabolites in complex culture extracts at the nanoscale	148
Chapter 7.	Discovery and derivatization of tridecaptin antibiotics with altered host specificity and enhanced bioactivity	182
Chapter 8.	Chimeric biosynthesis unveils diverse tridecaptin variants in the *Paenibacillus* genus	218
Chapter 9.	General discussion	250
Nederlandse samenvatting		260
References		270
Curriculum vitae		296
Publications		298



Chapter 1

Thesis outline

Nature sustains seemingly unlimited resources of natural products (NPs), which are metabolites derived from living organisms, such as plants, animals, and microorganisms. Microorganisms are considered nature's medicine makers, producing biomolecules with varying functional activities such as antibiotics, antifungals, anticancer agents and immunosuppressants. About 70 % of clinically used antibiotics are produced by bacteria, with the majority originating from Actinobacteria (Katz & Baltz, 2016). Typically, these compounds are products of specialized metabolism and play a role in the survival of bacteria.

The evolution and widespread distribution of antibiotic resistance have made diseases that were once easily treatable challenging again. Unfortunately, the increase in antibiotic resistance is paralleled by a lack of success in discovering new antibacterial drugs (Brown & Wright, 2016). A significant challenge in antibiotic discovery is the repeated rediscovery of known compounds, often referred to as "low-hanging fruits" (Kolter & van Wezel, 2016). It has been estimated that only 3% of the natural products potentially encoded in bacterial genomes have been experimentally characterized (Gavriilidou *et al.*, 2022). This imbalance between the enormous potential and the limited exploration emphasizes the urgency for innovative strategies for antibiotic discovery.

There are two main approaches to antibiotic discovery: the traditional "top-down" method and the more recent "bottom-up" strategy known as genome mining (Figure 1) (Luo *et al.*, 2014). In the "top-down" approach, compounds are discovered from microorganisms without prior knowledge of their genetic makeup. Conversely, genome mining involves the systematic analysis of microbial genomes to identify biosynthetic gene clusters (BGCs) responsible for the biosynthesis of natural products. Genome mining facilitates prioritizing and subsequent activation or chemical synthesis of the predicted product. The goal of the work described in this thesis was to uncover previously undiscovered natural products with potent bioactivities. The thesis is divided into two sections based on the methodology employed for antibiotic discovery. The first section (**Chapters 3–6**) covers "top-down" approaches and the second section (**Chapters 7–8**) utilizes "bottom-up" strategies.

The research involves both chemistry and biology, e.g. (1) screening and prioritization of bacterial extract libraries to discover new bioactive compounds employing MS-based metabolomics and Global Natural Products Social molecular networking (GNPS MN); (2) activating silent BGCs in actinomycetes; (3) detection and isolation of bioactive compounds from a complex mixture; (4) identifying the structures of compounds by using spectroscopic techniques like nuclear magnetic resonance (NMR) or mass spectrometry (MS) and, where needed, complemented by X-ray crystallography; (5) organic synthesis of the discovered compounds and subsequent derivatization to enhance the bioactivity properties (6) testing antimicrobial activity of discovered compounds; (7) linking the compound to the biosynthetic gene clusters.

Chapter 2 provides a review of recent developments in the antibiotic discovery field. This review discusses (1) the contemporary challenges in antibiotic discovery; (2) advances in genome mining and in translating the genetic code into compounds; (3) strategies to expand the known chemical space and to activate the cryptic BGCs; (4) advances in LC-MS-based metabolomics for the prioritization and subsequent identification of bioactive compounds.

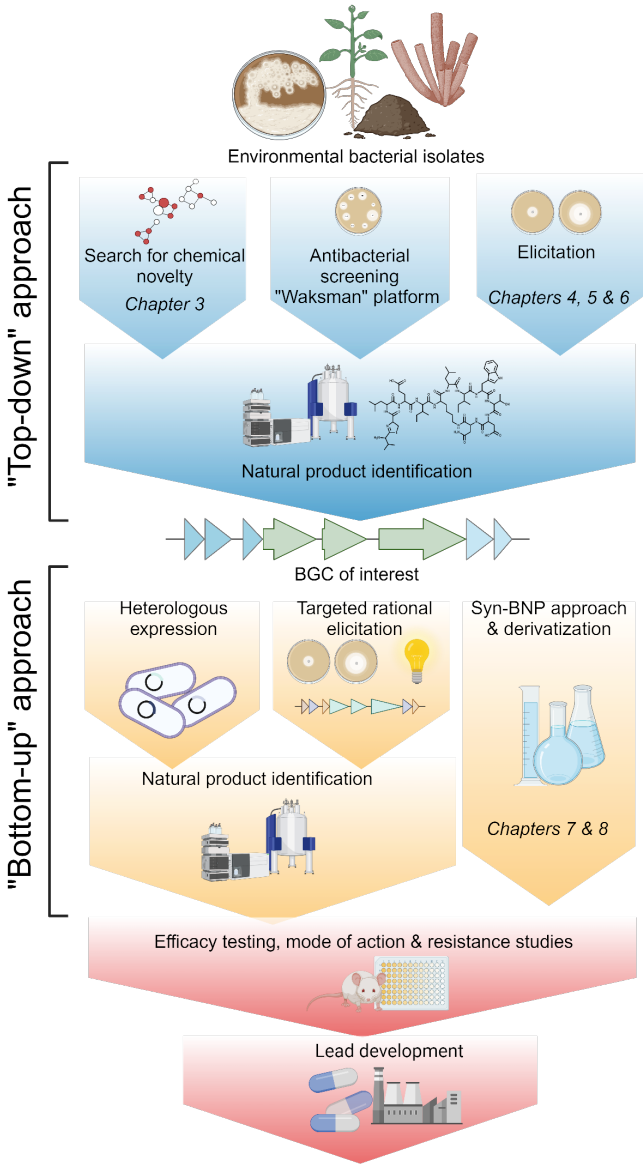


Figure 1. Summary of the major steps and processes in the antibacterial drug discovery pipeline.

Chapter 3 shows the application of MS-based metabolic profiling to facilitate the discovery of novel microbial NPs. We used MassQL-enhanced molecular networks (MN) to profile nonribosomal peptides (NRPs) produced by *Paenibacillus*. We propose an NP prioritization approach based on GNPS MN and MassQL that effectively targets basic amino acid-containing peptides within massive multi-informative MN. Compounds of interest are thereby highlighted based on the detection of the basic amino acid signatures and mapping the sub-structure information onto an MN acquired over 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 *Enterococcus faecium* E155.

In **Chapters 4** and **5**, different culture media were tested to increase secondary metabolite production and antimicrobial activity in *Streptomyces* sp. MBT27. LC-MS and molecular networking were subsequently used to profile the metabolomes elicited through these culture condition perturbations compared to the respective control experiments. LC-MS data were statistically analyzed using multivariate data analyses to differentiate chemical profiles and correlate specific *m/z* values to the bioactivity of the crude extracts. The up-scale fermentation in the resulting optimized culture media facilitated the isolation of new compounds from quinazolinone (**Chapter 4**) and actinomycin families (**Chapter 5**). The biosynthetic pathways of new compounds were proposed, and the corresponding BGC was identified through bioinformatics analysis.

Early-stage compound dereplication is critical to the discovery of bioactive natural products. In **Chapter 6,** we present an efficient analytical platform to allow the discovery of low-abundance compounds at medium throughput, based on at-line nanofractionation combined with molecular networking and bioactivity assays, which we call nanoRAPIDS. NanoRAPIDS encompasses analytical scale separation of natural extracts, post-column nanofractionation followed by high-throughput bioassay, automated mass spectrometry identification, and GNPS MN for dereplication and analog search. Proof of concept was obtained for families of iturins and surfactins in the extracts from *Bacillus*, whereby we readily identified compounds that act against Gram-positive and Gram-negative bacteria and fungi. Then, nanoRAPIDS was applied to address the eliciting role of catechol in the production of bioactive compounds, highlighting the individual mass features that explained the increased bioactivity in the crude extract.

In **Chapter 7,** we report the genome-guided discovery of new lipopeptide antibiotics tridecaptin A_5 and tridecaptin D. Through bioinformatic-guided organic synthesis and comprehensive antibacterial assays, we demonstrate that synthetic analogs Oct-TriA_5 and Oct-TriD exhibit unusual bioactivities not commonly associated with tridecaptins. The change in the antibacterial spectrum of Oct-TriA_5 could be explained solely by a Phe to Trp substitution compared to Oct-TriA_1, while Oct-TriD contained six substitutions. Screening

of tridecaptin analogs substituted at position 9 identified Oct-His9 as a potent congener with exceptional efficacy against colistin-resistant *Pseudomonas aeruginosa* and reduced hemolytic and cytotoxic properties. These findings underscore the significant potential of tridecaptin analogs as viable alternatives in combatting antibiotic-resistant bacterial pathogens.

Moreover, global genome mining of *Paenibacillus* nonribosomal peptide synthetase (NRPS) BGCs uncovered tridecaptin-like BGCs with unusual architecture. **Chapter 8** illustrates biosynthetic "crosstalk" among two gene clusters enabled de novo synthesis of chimeric tridecaptin B_1. We hypothesized that the truncated tridecaptin-like BGC in *Paenibacillus* sp. JJ-1683 might be involved in the chimeric biosynthesis of tridecaptin B_1 by biosynthesizing the first ten amino acids, upon which the last three are incorporated by the TriE synthase encoded by the full-length tridecaptin BGC. This result underscores chimeric biosynthesis as a strategy for the dynamic production of similar compounds with distinct chemistries and bioactive profiles.

Finally, **Chapter 9** presents a general discussion of the thesis, which also includes a summary of the most important data and observations and provides perspectives for future antibiotic discovery from microbial sources.



Chapter 2

General introduction

Pioneering and contemporary challenges in antibiotic discovery

The discovery and introduction of antibiotics into clinical use was the most remarkable medical breakthrough of the past century. Fleming's discovery of the first antibiotic, penicillin, in 1928 marked the beginning of the modern medical era of antibiotics (Katz & Baltz, 2016, Fleming, 2001). Inspired by Fleming's discovery, Selman Waksman started a systematic study of microbes as producers of antimicrobial compounds in the late 1930s (Lewis, 2012). The canonical "Waksman platform" for antibiotic discovery involved the identification of new bacterial isolates from the soil samples, antimicrobial screenings, and subsequent bioassay-guided purification of the bioactive compound (Waksman, 1945). Waksman discovered fifteen antibiotics made by soil-dwelling actinomycetes, including streptomycin, the first antibiotic against tuberculosis. Selman Waksman's pioneering work sparked a significant turning point in antibiotic discovery and opened the door to the golden era of antibiotic discovery. This period witnessed a surge in the identification and development of crucial antibiotics like bacitracin, tetracycline, polymyxin, vancomycin, and many others (Katz & Baltz, 2016) (Figure 1). More than half of all the classes of clinically used antibiotics came from actinomycetes, while the rest derived from fungi and Firmicutes (Katz & Baltz, 2016).

Soil bacteria produce a wide range of natural products that protect against insects, herbivores, and phytopathogens (including bacteria, fungi, nematodes, and viruses), as extensively highlighted in numerous reviews (Katz & Baltz, 2016, Bérdy, 2005, Newman & Cragg, 2020, Olishevska *et al.*, 2019). Several theories have been proposed to explain why soil microbes produce such a diverse array of bioactive natural products (NPs). The most likely explanation is that they have multiple functions and act as chemical weapons or signaling molecules; or mediate interactions with eukaryotic hosts such as insects and plants (Seipke *et al.*, 2012).

Following golden era of antibiotic discovery, the efficiency of classical screening programs dropped significantly (Figure 1), making them unprofitable for pharmaceutical companies (Baltz, 2008). Due to the high rediscovery rate, it was hypothesized that the biosynthetic potential of the traditional producers had been fully explored (Cooper & Shlaes, 2011, Hutchings *et al.*, 2019). Recent advancements in sequencing technologies have resulted in the availability of a repository of genome sequence information (Katz & Baltz, 2016, Baltz, 2008). Once thought to be largely devoid of new drugs, these technologies revealed evidence for a treasure trove of yet undiscovered molecules that lie hidden in the biosynthetically talented bacterial genera (Bentley *et al.*, 2002, Ikeda *et al.*, 2003, Wipat & Harwood, 1999). Next-generation sequencing of microbial genomes has rejuvenated the discovery and characterization of diverse natural products and revealed the still largely underexplored biosynthetic potential of bacteria (Bentley *et al.*, 2002).

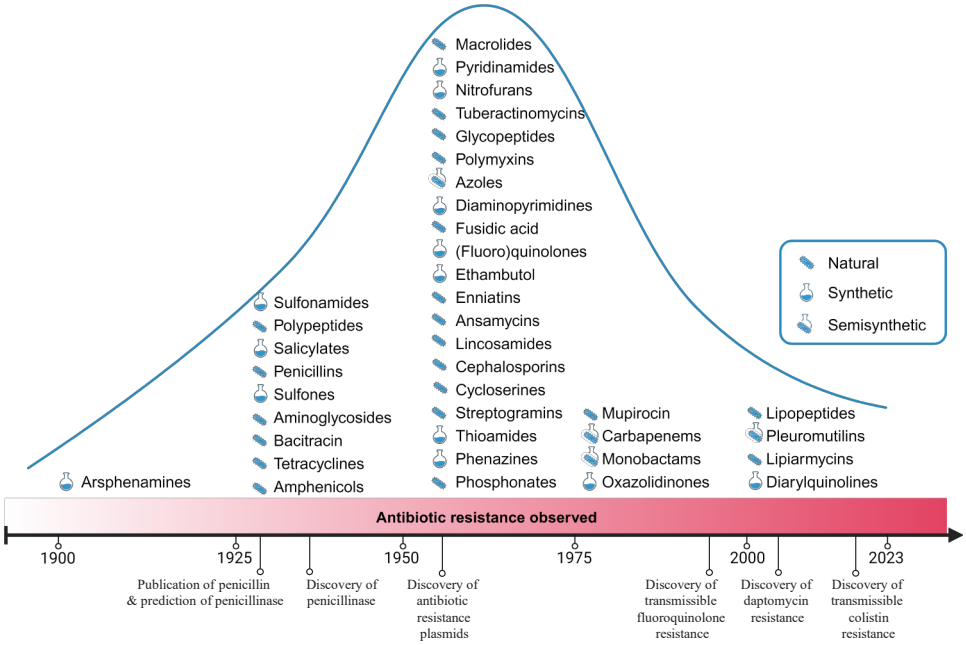


Figure 1. Timeline of antibiotic discovery and antibiotic resistance appearance. The icon marks the origin of the antibiotic class. The period between the 1950s and 1970s was the golden era of the discovery of novel antibiotic classes, but the development declined between the 1970s and the 1990s. Note that the increase in antibiotic resistance parallels a lack of success in discovering new antibacterial drugs.

Genome mining to evaluate the biosynthetic potential of bacteria

Thus, as the bacterial DNA ultimately encodes the production of all specialized metabolites, genome mining may offer a new route to natural product-based medicine discovery. In the bacterial genomes, the genes encoding biosynthetic pathways for natural products are usually organized in so-called biosynthetic gene clusters (BGCs) (Martin & Liras, 1989). It is efficient for microbes to have all the genes encoding a biosynthetic pathway in the same place; in this way, the entire production line can be activated by transcribing DNA from this genomic region. The toolkit for identifying BGCs involves methods to pinpoint key microbial sources, obtain high-quality genome data through sequencing technologies like Illumina, PacBio, and Oxford Nanopore, and analyze it using diverse bioinformatic tools (Mullowney *et al.*, 2023). Genome mining tools antiSMASH (Blin *et al.*, 2023) and PRISM (Skinnider *et al.*, 2020)2020 have been developed to identify and annotate microbial BGCs. With the help of antiSMASH (Blin *et al.*, 2023) we can identify BGCs in microbial genomes and classify them based on the type of molecule they are predicted to produce, such as polyketides (Ray & Moore, 2016, Donadio *et al.*, 2007, Shen, 2003), ribosomally synthesized and post-translationally modified peptides (RiPP) (Montalban-Lopez *et al.*, 2021), terpenes (Avalos *et al.*, 2022, Schulz &

Dickschat, 2007), and nonribosomal peptides (Sieber & Marahiel, 2005, Nivina *et al.*, 2019, Süssmuth & Mainz, 2017). PRISM uses pHMMs to detect BGCs and additionally predicts biological activity through machine learning (Skinnider *et al.*, 2020)2020. Alternative deep-learning genome mining software for BGC prediction, such as ClusterFinder (Cimermancic *et al.*, 2014), DeepBGC (Hannigan *et al.*, 2019), GECCO (Carroll *et al.*, 2021), and SanntiS (Sanchez *et al.*, 2023), uncovered undetectable putative BGCs that may code for natural products with novel bioactivities. BAGEL (van Heel *et al.*, 2013) and RODEO (Tietz *et al.*, 2017) are two genome mining tools designed explicitly for genome mining of RiPPs. Additionally, RiPPs may also be identified by their RiPP recognition Element (RRE), and the recently developed RRE-finder was incorporated into the shell of antiSMASH 6 (Kloosterman *et al.*, 2020b). Nevertheless, the diversity of RiPPs is vast, and the chemical space is expanding rapidly. To allow their detection, genome mining algorithms were developed to identify BGCs not captured using canonical rule-based annotation approaches (de Los Santos, 2019, Kloosterman *et al.*, 2020a, Merwin *et al.*, 2020, Tietz *et al.*, 2017). This includes the machine learning-based decRiPPter algorithm, which allowed the identification of a novel subclass of lanthipeptides that had escaped identification via existing algorithms (Kloosterman *et al.*, 2020a). The extensive list of the genome-mining software, tools, and databases utilized for NP discovery is available on the Secondary Metabolite Bioinformatics Portal (SMBP) (Weber & Kim, 2016).

The immense amount of genome data has prompted genome mining analyses conducted on a large scale. Rather than mining a single genome at a time, we can now incorporate big data approaches to mine a large pool of genomes or metagenomes for novel NPs (Medema & Fischbach, 2015). This approach is called "global genome mining" (Medema *et al.*, 2021). Bioinformatics tools that facilitate the analysis of big BGC datasets include the EvoMining (Sélem-Mojica *et al.*, 2019), ARTS (Alanjary *et al.*, 2017), BiG-SCAPE and the CORASON (Navarro-Muñoz *et al.*, 2020). Moreover, a highly scalable tool BiG-SLiCE was recently developed to deal with extensive data from global genome mining (Kautsar *et al.*, 2021b).

With the simultaneous sequencing of hundreds to thousands of microbial genomes becoming more common, the large quantities of BGC data resulting from this pose an opportunity and challenge. Databases such as antiSMASH-DB (Blin *et al.*, 2024)2024, IMG-ABC (Palaniappan *et al.*, 2020), and MIBiG (Terlouw *et al.*, 2023) play a critical role in analyzing BGCs by enabling the comparison of newly sequenced BGCs with those predicted and experimentally characterized in the past. In addition to the mentioned databased for exploring BGCs from the publicly available data, the recently established BiG-FAM database focuses on the gene cluster family (GCF) relationship of the detected BGCs and enables GCF-

based exploration and homology searching of >1.2 million BGCs harbored by >200,000 microbial genomes (Kautsar *et al.*, 2021a).

Translating the genetic code into compounds

Once BGCs of interest have been identified based on compound class, resistance genes, or phylogeny, predicting the specific final product remains challenging (Scherlach & Hertweck, 2021). For some natural product classes, we have a relatively good understanding of how the encoded biosynthetic machinery builds a natural product, and thus, we can attempt to predict the scaffold structure of the produced natural product. Among all natural product classes, the chemical structure can be predicted for NRPSs, modular polyketide synthase (PKSs), and RiPPs (Scherlach & Hertweck, 2021). The core structure and post-translational modifications (PTMs) of the RiPPs can be reliably predicted by RODEO (Tietz *et al.*, 2017), RiPPER (Moffat *et al.*, 2021), antiSMASH (Blin *et al.*, 2023) and PRISM (Skinnider *et al.*, 2020)2020.

Considering the scope of the thesis, the prediction principle will be illustrated based on the NRPS assembly line. NRPS machineries are complex, multi-domain enzymes that select and condensate amino acids iteratively to build up peptidic natural products (Sieber & Marahiel, 2005, Nivina *et al.*, 2019, Süssmuth & Mainz, 2017). To achieve nonribosomal peptide synthesis, minimal enzymatic complexes require four distinct domains called core or essential domains: condensation (C), adenylation (A), peptidyl carrier protein (PCP), and a terminal thioesterase (TE) domain (Figure 2) (Marahiel, 2016). Three domains make up the minimum elongation module, with C domains catalyzing condensation, A domains mediating building block selection through adenylation and subsequent thiolation, and non-catalytic PCP domains functioning as mechanical arms to which intermediate scaffolds are tethered between reactions (Winn *et al.*, 2016). An NRPS assembly line only requires a single TE domain, typically at the end of an NRPS enzyme, which is responsible for chain release (Mootz *et al.*, 2002).

As A domains govern building block selection, many NRP structure prediction algorithms attempt to predict the selected substrate from A domain sequence. Examples of such machine learning algorithms include NRPSpredictor2 (Röttig *et al.*, 2011), AdenylPred (Mongia *et al.*, 2023), SANDPUMA (Chevrette *et al.*, 2017), and AdenPredictor (Chevrette *et al.*, 2017), which were trained on A domain sequences of known specificity.

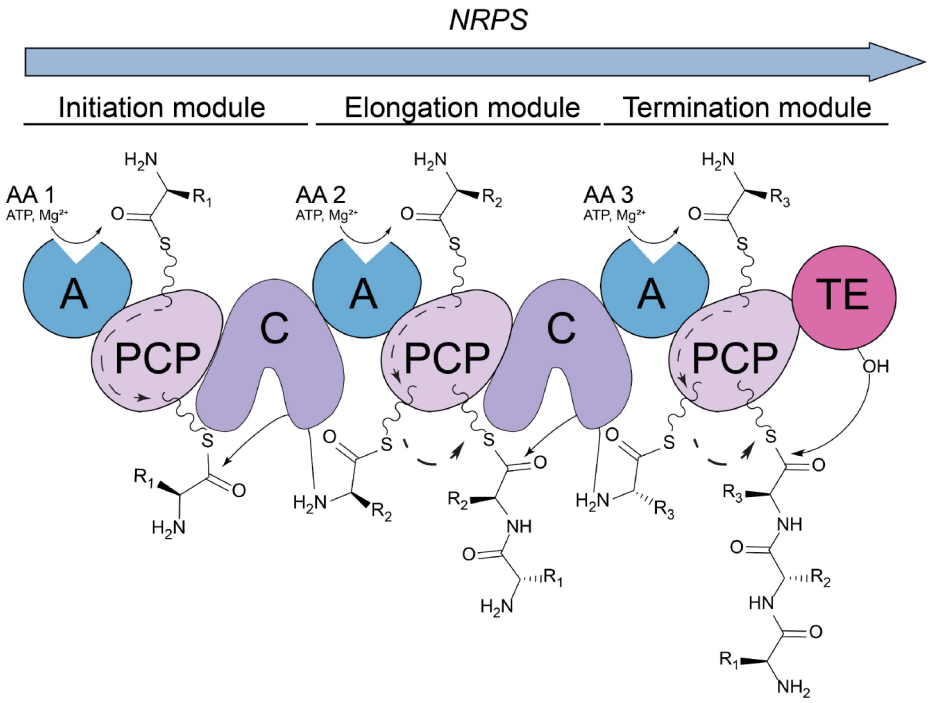


Figure 2. Schematic representation of NRP biosynthesis. The NRP biosynthesis involves the activation of amino acid substrates via ATP, facilitated by the adenylation domain (A). Afterward, the aminoacyl-AMP intermediate is captured by the thiol group of the flexible 4'-phosphopantetheine arm linked to a peptidyl carrier protein (PCP) domain. Condensation domains (C) subsequently facilitate peptide bond formation between thioester intermediates loaded onto neighboring PCP domains. Lastly, the final module hosts an additional thioesterase domain (TE), which is responsible for either hydrolysis or cyclization to release the product from the NRPS (adapted from Winn *et al.*, 2016).

However, the diversity and complexity of NRPs have made systematic investigation difficult, and consequently, most genetically encoded NRPs in bacteria have been overlooked. An alternative way to investigate the potential of novel scaffolds predicted by genome mining is via organic synthesis or chemoenzymatic total synthesis. The approach of synthetic bioinformatic natural product (syn-BNP) involves bioinformatic prediction of the natural product structure based on the genome sequence with subsequent organic synthesis of compounds, circumventing the need for bacterial culture and gene expression (Scherlach & Hertweck, 2021, Chu *et al.*, 2020). Concurrently, solid phase peptide synthesis (SPPS) of structurally diverse peptides has become rapid and economical, making NRPS gene clusters ideal candidates for a syn-BNP approach. This new methodology completely avoids the issues of isolation and scalability of target compounds. The syn-BNP approach was successfully applied to discover new antimicrobial entities from sequenced bacterial genomes or metagenomes, with an emphasis on the de novo discovery of novel clinically relevant NRP antibiotics. This approach led to the discovery of humimycin A and B (Chu *et al.*, 2016)2016,

kanglemycins (Peek *et al.*, 2018), macolacin (Wang *et al.*, 2022c), cilagicin (Wang *et al.*, 2022b) with potent bioactivity against multidrug-resistant (MDR) pathogens.

Examples of clinically used nonribosomal peptides

NRPs are among the most widespread groups of natural products with diverse chemical structures (Wang *et al.*, 2014). NRPs often contain both proteinogenic and nonproteinogenic amino acids, significantly expanding their chemical diversity and bioactivity spectrum (Caboche *et al.*, 2010). The rising issue of antibiotic resistance has led to increased interest in the therapeutic potential of NRP antibiotics (Hancock & Chapple, 1999). Peptide antibiotics currently used in clinics, such as bacitracin, daptomycin, polymyxin, and vancomycin constitute effective treatments for infections caused by multidrug-resistant pathogens (Figure 3) (Felnagle *et al.*, 2008, Liu *et al.*, 2019).

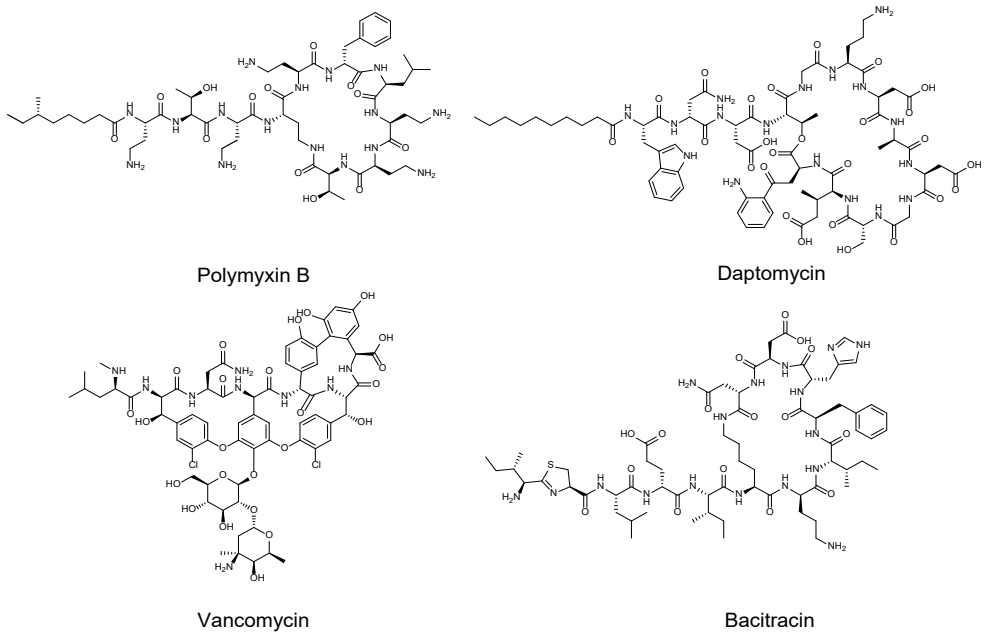


Figure 3. Structures of clinically used NRPs.

Polymyxins are active against Gram-negative bacteria and are typically only used to treat infections caused by multidrug-resistant organisms, including carbapenem-resistant *Enterobacterales*, *Acinetobacter baumannii,* and *Pseudomonas aeruginosa (Nang et al., 2021, Slingerland & Martin, 2024)*. The mode of action of polymyxins involves binding to the lipid A component of lipopolysaccharides (LPSs) in the outer membrane of Gram-negative bacteria. The positively charged 2,4-diaminobuturic acid residues of polymyxins electrostatically

interact with the negatively charged phosphate groups of lipid A, facilitating binding and leading to membrane disruption and increased permeability (Mohapatra *et al.*, 2021). In contrast, NRPs like bacitracin, vancomycin, and daptomycin are active preliminary against Gram-positive bacteria and act as inhibitors of cell wall biosynthesis at the different stages of this complex process (Buijs *et al.*, 2023). Daptomycin and vancomycin are considered "last resort" antibiotics for the treatment of severe Gram-positive infections (Gonzalez-Ruiz *et al.*, 2016, Griffith, 1981). Vancomycin targets and tightly binds lipid II, an essential building block of the cell wall, resulting in the cessation of cell wall biosynthesis (Williams & Butcher, 1981, Williams, 1984). The mode of action of daptomycin is multifaceted and involves the targeting and depolarization of bacterial cell membrane alongside the inhibition of cell wall biosynthesis by direct binding to lipid II (Kotsogianni *et al.*, 2021, Grein *et al.*, 2020). Other NRPs, like bacitracin, sequester the membrane-associated phospholipid undecaprenyl pyrophosphates (C_{55}PP), transport molecule that carries the building blocks of the peptidoglycan bacterial cell wall (Ming & Epperson, 2002). The binding prevents peptidoglycan synthesis, thereby inhibiting bacterial cell growth. The recent discovery of cilagicins, which exhibit a unique mode of action and bind both C_{55}PP and undecaprenyl phosphate (C_{55}P) involved in cell wall biosynthesis, highlights the significant untapped potential of nonribosomal peptides (NRPs) in combating antibiotic-resistant bacteria (Wang *et al.*, 2022b).

Exceptions to the rules of NRP biosynthesis

Still, translating genomic knowledge into molecules remains challenging. The diversity of NRPS-derived peptides is not only determined by the incorporation of different amino acid building blocks but also by an ever-increasing number of characterized tailoring reactions and enzymatic domains (McErlean *et al.*, 2019). Moreover, the promiscuity of NRPS modules allows the production of multiple compounds encoded by a single BGC (Duban *et al.*, 2022). In addition, a constantly expanding library of scientific literature unravels new examples of nonribosomal peptide synthetases exhibiting rare domains and noncanonical module organization, leading to unique microbial strategies of natural products biosynthesis (Chevrette *et al.*, 2020). Moreover, recent publications demonstrate a much greater mechanistic diversity in NRPS assembly lines than previously anticipated. Recent characterization of several NRPS megasynthases revealed examples of the loss of the co-linearity rule (Wenzel & Müller, 2005). It turned out that modules from bacterial NRPSs can be used repeatedly, skipped, or even split in the biosynthesis of a single product (Gewolb, 2002, Marahiel, 2009, Wenzel & Müller, 2005).

One notable example entails collaborative natural product biosynthesis, involving crosstalk between two separate BGCs. Such pairing of biosynthetic enzymes from distinct BGCs within the same biosynthetic pathway gives rise to increased chemical diversity

and an expanded scope of biological activity. Over the past few years, this phenomenon of collaborative biosynthesis has received increased attention, and specialized metabolites produced through such strategies are known as 'hybrid molecules' or 'chimeras' (Mevers *et al.*, 2019, Yin *et al.*, 2022). Chimeric biosynthesis has previously been observed in various NRPS and PKS (hybrid) systems, including the biosynthesis of azaphilone (Huang *et al.*, 2020), penilactones (Fan *et al.*, 2019), spirotryprostatin A (Yan *et al.*, 2019), and echinocandin B (Cacho *et al.*, 2012).

Despite the impact of genomics, it only delivers possible leads, which needs to be supported by experimental evidence provided by microbiology, analytical chemistry, and bioassays to purify and identify the NPs and determine their mode of action. Analytical chemistry techniques, such as high-resolution mass spectrometry and NMR (nuclear magnetic resonance spectroscopy), complement theoretical in silico predictions. They play a crucial role in bridging the gap between genomic potential and the actual production of metabolites (Caesar *et al.*, 2021).

Expanding chemical space and identification of bioactive natural products

Activation and characterization of cryptic BGCs

Many BGCs are not expressed during standard laboratory conditions, and thus, their cognate natural products escape attention (Nett *et al.*, 2009, Kolter & van Wezel, 2016). To unlock the full potential of bacteria, it's crucial to understand their ecological context, as this will provide clues to the activation mechanisms of their biosynthetic pathways (van Bergeijk *et al.*, 2020, Zhu *et al.*, 2014a). The signals that mediate the production of specialized metabolites are diverse and include physical cell-cell interactions, nutrient depletion, enzymatic conversion of precursors to active metabolites and microbial small molecules (Hoskisson & Fernández-Martínez, 2018, van der Heul *et al.*, 2018, Huang *et al.*, 2005). However, for many interactions, the signals and molecular mechanisms remain unknown. Several methods have been developed to boost the production of hidden natural products, including ribosome engineering, co-culture, varying culturing conditions, insertion of active promoters, and high-throughput elicitor screening (Bertrand *et al.*, 2014, Takano, 2006, Zhu *et al.*, 2014a, Bode *et al.*, 2002). The expression of bacterial BGCs can be manipulated by altering the culturing conditions. This is the basis of the one-strain–many-compounds (OSMAC) concept, meaning that a strain can produce a plethora of distinct compounds if exposed to different abiotic or biotic conditions (Bode *et al.*, 2002). In co-culturing, microbial partners are added to a single culture to trigger the expression of metabolic programs involved in defense or nutrient competition (Sugiyama *et al.*, 2015, Hoshino *et al.*, 2015, Bertrand *et al.*, 2014, Wu *et al.*, 2015b, Schroeckh *et al.*, 2009). As an alternative, high-throughput elicitor screening can be used to induce the expression of silent BGCs using the small compounds. This is an efficient platform that enables the discovery

of small molecule activators that lead to the induction of silent biosynthetic clusters as well as structural and functional elucidation of their products (Seyedsayamdost, 2014, Moon *et al.*, 2019a, Moon *et al.*, 2019b). Moreover, advances in synthetic biology enable the engineering of biological systems to express customized biosynthetic pathways. Synthetic biology tools can alter the genome of selected microorganisms to induce the expression of BGCs and the efficient production of natural products (Smanski *et al.*, 2016, Zhang *et al.*, 2019). The resulting "cell factory" provides improved genomic stability and allows the optimization of the targeted biosynthetic pathway in culturable and biochemically characterized systems.

LC-MS-based metabolomics methods for data acquisition

For metabolite profiling, culture extracts are analyzed by NMR spectroscopy, high-resolution mass spectrometry (HRMS), or combined methods involving upstream liquid chromatography (LC), such as LC-HRMS and more recent LC-NMR (Atanasov *et al.*, 2021, Wu *et al.*, 2015a). Compatibility of reversed-phase (RP) LC with biological samples allowed its utilization in combination with accurate mass MS (QToF or Orbitrap technologies) (Roux *et al.*, 2011). LC-MS established itself as the technology of choice for microbial metabolic profiling due to the widespread availability of suitable instrumentation and ease of both access and use, combined with good sensitivity, specificity, and resolving power. Considering the scope of the thesis, I will here focus mainly on LC-MS-based metabolomics. For NMR-based metabolomics the reader is referred to reviews elsewhere (Markley *et al.*, 2017, Wu *et al.*, 2016a, Grienke *et al.*, 2019), and likewise for reviews of MS-based approaches (Krug & Müller, 2014, Bouslimani *et al.*, 2014, Spraker *et al.*, 2020).

The key challenges scientists face when targeting the metabolome are analytical reproducibility, feature coverage, dynamic range, component annotation, data analysis, and processing large sample sets. LC-MS has enabled the detection of thousands of mass features within a single biological sample. Therefore, one of the main challenges in LC-MS-based metabolomics is the comprehensive and accurate analysis of large datasets. Freely available data preprocessing tools, such as XCMS (Smith *et al.*, 2006), Mzmine (Pluskal *et al.*, 2010), MetaboAnalyst (Pang *et al.*, 2022), MetAlign (Lommen, 2009), Metabolomic Analysis and Visualization Engine (MAVEN) (Clasquin *et al.*, 2012), MET-COFEA (Zhang *et al.*, 2014) as well as commercial software, have been developed to facilitate data processing. Moreover, the recently developed publicly available R-based web application Metabolomics Explorer (MetEx) enables users to quickly and intuitively analyze comprehensive metabolomics datasets. MetEx facilitates analysis of complex datasets, consisting of retention time, *m/z*, and MS intensity features, as a function of hundreds of conditions (Covington & Seyedsayamdost, 2021).

LC-MS-based metabolomics and the discovery of bioactive compounds

Metabolomics measures multiple metabolic responses in living systems, allowing simultaneous tracking of changes in hundreds of small molecules. However, the complexity and diversity of chemical structures make metabolomic analysis challenging. The methodologies employed to interpret high-throughput metabolomics data mainly derived from earlier emerging omics technologies. Classic approaches aim to assess group-wise differences, either in a univariate, parameter-by-parameter fashion, e.g., *t*-test, analysis of variance (ANOVA), or using multivariate techniques (e.g., MANOVA, ASCA, PCA, PLS) (Bartel *et al.*, 2013). Statistical models such as Pearson correlation, partial least squares (PLS), discriminant analysis (PCA-DA, PLS-DA, OPLS-DA), and hierarchical cluster analysis (HCA) are often used to link metabolite fingerprints to bioactivity data (Bartel *et al.*, 2013). The general idea behind supervised methods is to unravel distinct metabolite profiles that are associated with the observed bioactivity of the sample. A recent extension to the PLS repository is the orthogonal-PLS (OPLS) method. In the OPLS analysis, the data variation is split into the variance of interest and a noise part, which is unrelated to the response (Dettmer *et al.*, 2007). This leads to a simplified interpretability of the resulting components, allowing additional assessment of within- and between-group variance.

While the approaches mentioned above provide a list of the highlighted features, the information obtained is at the prediction level. Multiple analytical modules involving different bioassays and detection technologies should be linked to allow simultaneous bioactivity evaluation and identification of compounds in complex compound mixtures. Such approaches may help provide a more comprehensive insight into compounds responsible for the bioactivity of interest.

GNPS molecular networking and dereplication of natural products

In the pursuit of antibiotic discovery, a pivotal role is played by Global Natural Products Social molecular networking (GNPS) (Wang *et al.*, 2016). This open-access platform is a hub for sharing, processing, and visualizing tandem mass spectrometry data (Quinn *et al.*, 2017). The concept of molecular networking is based on the organization and visualization of 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 connect and create spectral families (Wang *et al.*, 2016). Classical molecular networking has recently been integrated with feature detection methods to introduce Feature-Based Molecular Networking (FBMN). This approach enables the discrimination of isomers within the molecular network and incorporates quantitative information generated by the feature detection tools (Nothias *et al.*, 2020).

Molecular networking lays the groundwork for bioactivity-based molecular networking. Combining the MN with metabolite "bioactivity scores" enables the detection and relative quantification of LC-MS/MS spectral features across chromatographic fractions. The "bioactivity scores" are calculated based on the Pearson correlation between feature intensity across samples and the bioactivity level associated with each sample (Nothias *et al.*, 2018). The correlation of the metabolite quantity and observed bioactivity facilitates the discovery of biologically active compounds within complex metabolomic datasets.

The major challenge in metabolomics is the identification of previously reported compounds, known as structural dereplication (Gaudencio & Pereira, 2015, Hubert *et al.*, 2017, Silver, 2011). Dereplication approaches include the determination of molecular formula using the Seven Golden Rules, Sirius 2, and MS-FINDER software and searching for compounds described by this formula in chemical structure databases (Gaudencio & Pereira, 2015, Mohamed *et al.*, 2016). The primary challenge of these approachs is the exponential increase in the number of potential formulas as the molecular mass of metabolites grows. Furthermore, existing chemical databases frequently contain numerous compounds with identical formulas, which complicates the annotation process.

Currently, the combination of molecular networking with annotation methods such as MS2LDA (van der Hooft *et al.*, 2016), Network Annotation Propagation (NAP) (da Silva *et al.*, 2018) or DEREPLICATOR (Mohimani *et al.*, 2018) accelerates the identification of unannotated ions. MolNetEnhancer (Ernst *et al.*, 2019) combines outputs from MS2LDA, NAP, DEREPLICATOR, and molecular networking, along with the automated chemical classification from ClassyFire (Djoumbou Feunang *et al.*, 2016), to assign structural features to chemical classes. Algorithms such as DEREPLICATOR or recently published VarQuest (Gurevich *et al.*, 2018) were designed to expand the library search to identify variants of known peptidic natural products. The newly developed tool Qemistree is employed to classify MS data into a tree structure (Tripathi *et al.*, 2021).

Future perspectives

A number of innovative methods have evolved to enhance the synergy between genomics and metabolomics for natural product discovery. Future developments for the processing of metabolome data will be focused more on continuous platform improvement, elucidation of structural identification, and functional interpretation of metabolomic data. As multi-omics technologies mature, there is an increasing need for data integration between platforms. One of the key challenges in the field of omics-based natural product discovery is the set-up of high-quality datasets to train deep learning algorithms and appropriate strategies for algorithm validation.

Despite the advances in the genomics and metabolomics fields, the current state of the art regarding BGC and natural product diversity limits our ability to guide drug discovery.

It is, therefore, necessary to further characterize the extant microbial diversity, pathways, and regulatory mechanisms for natural product biosynthesis. Moreover, characterizing the functions and modes of action of natural products in native microbial communities will be crucial for prioritizing BGCs and eliciting their expression. This should bring many more natural products to light, potentially contributing to future drug candidates.



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.

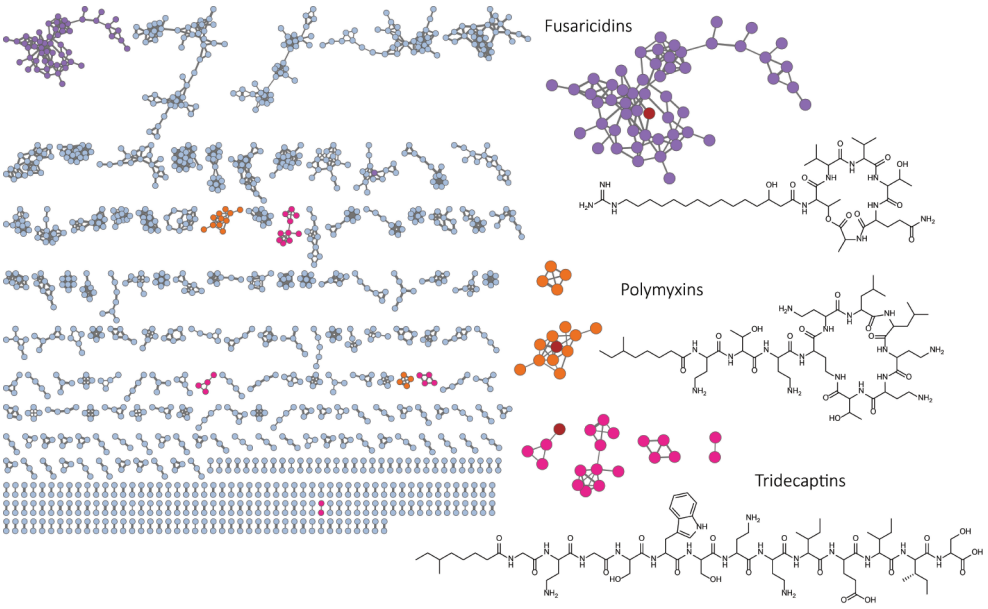


Figure 2. Molecular network of the mass features detected in the bacterial extracts from a collection of 227 plant-associated *Paenibacillus* isolates. Each node represents a parent ion detected in one or more strains. The annotated molecular families have been amplified. The colors of the nodes highlight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

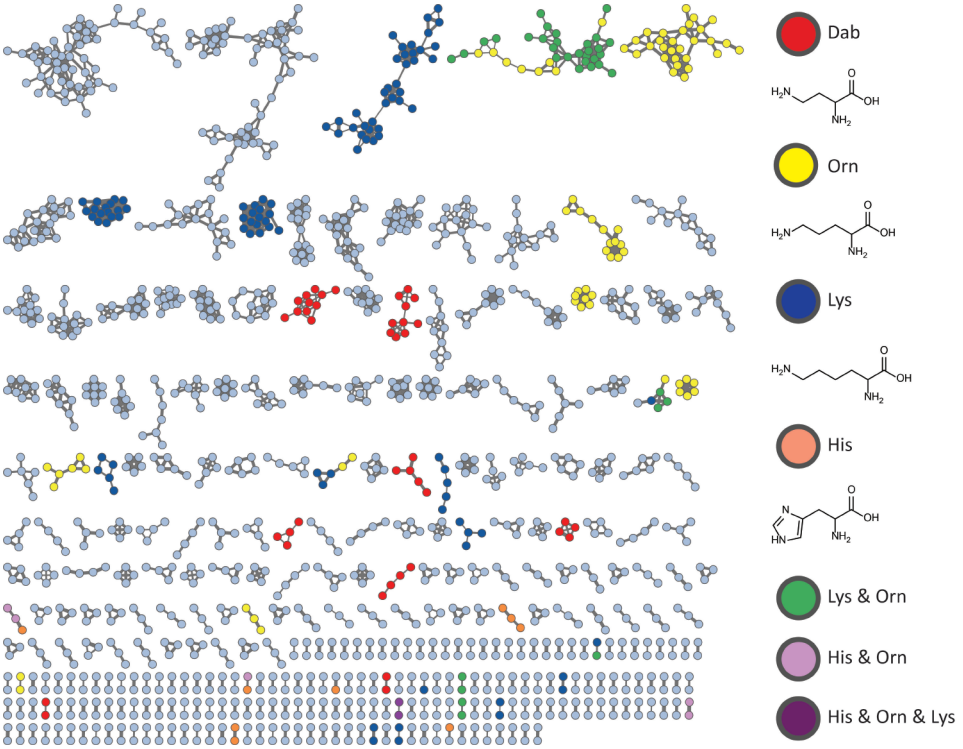


Figure 3. MassQL-annotated molecular network of the ions detected in the bacterial extracts from the collection of 227 plant-associated *Paenibacillus* isolates. The colors of the nodes highlight parent ions that contain basic amino acid residues in the structures.

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.

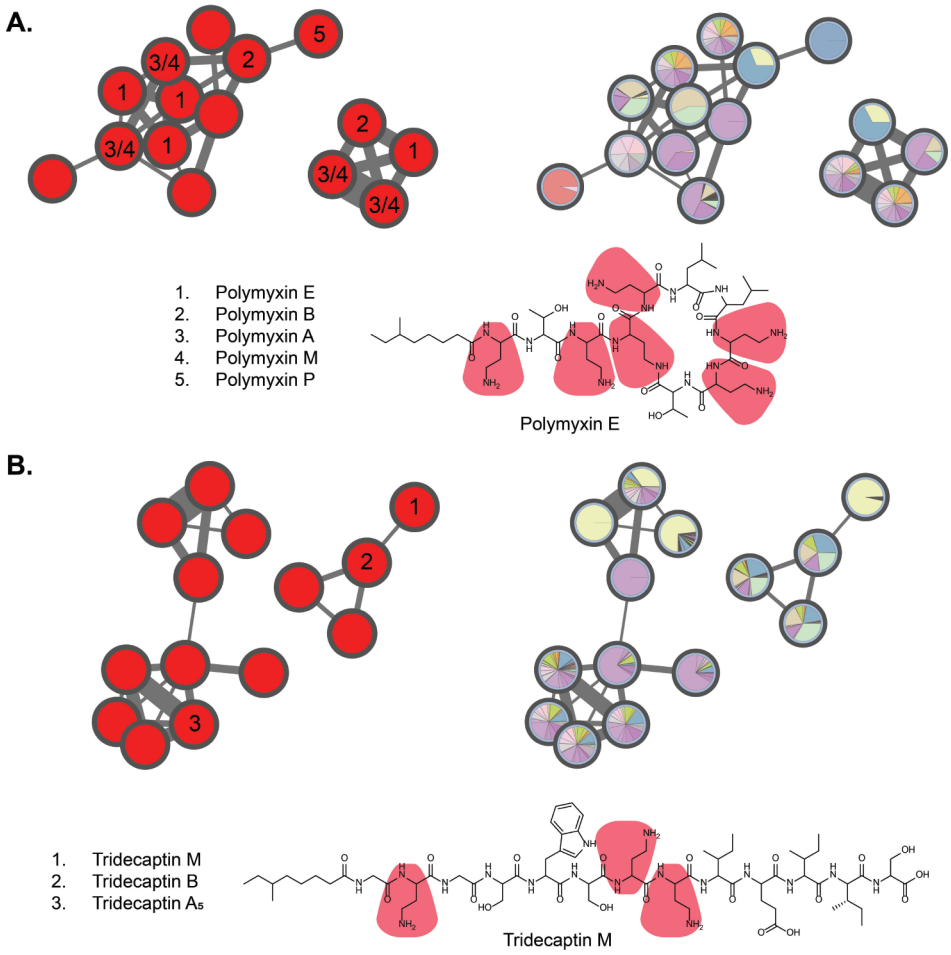


Figure 4. Molecular families of polymyxins (A) and tridecaptins (B) were highlighted with MassQL search as Dab-containing lipopeptides. On the same molecular family, pie charts were mapped to the nodes to represent the relative precursor ion intensities within each extract. Note that both polymyxins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

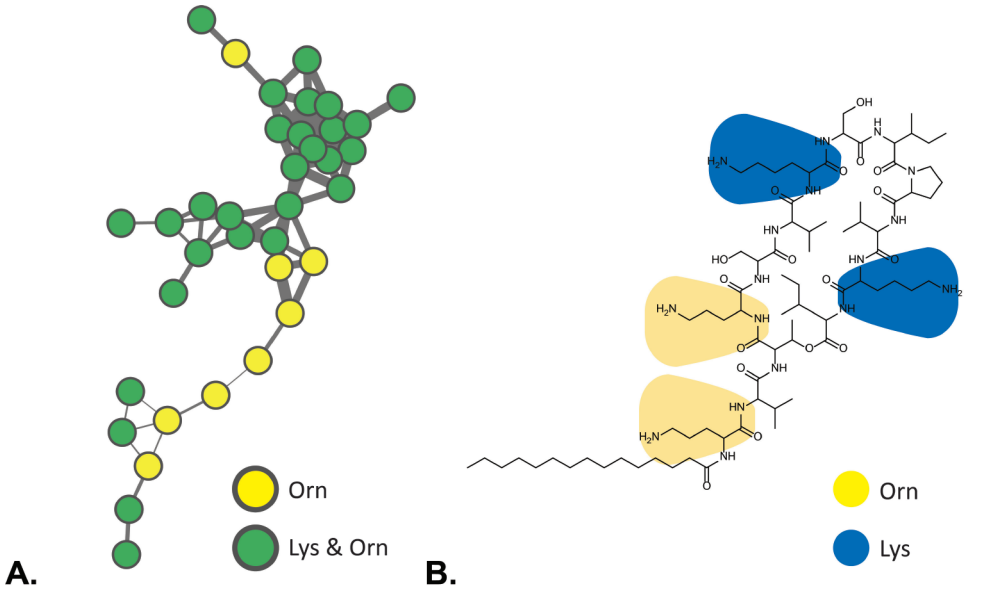


Figure 5. A. MassQL-annotated molecular family of paenibacterins. B. Chemical structure of paenibacterin A.

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.

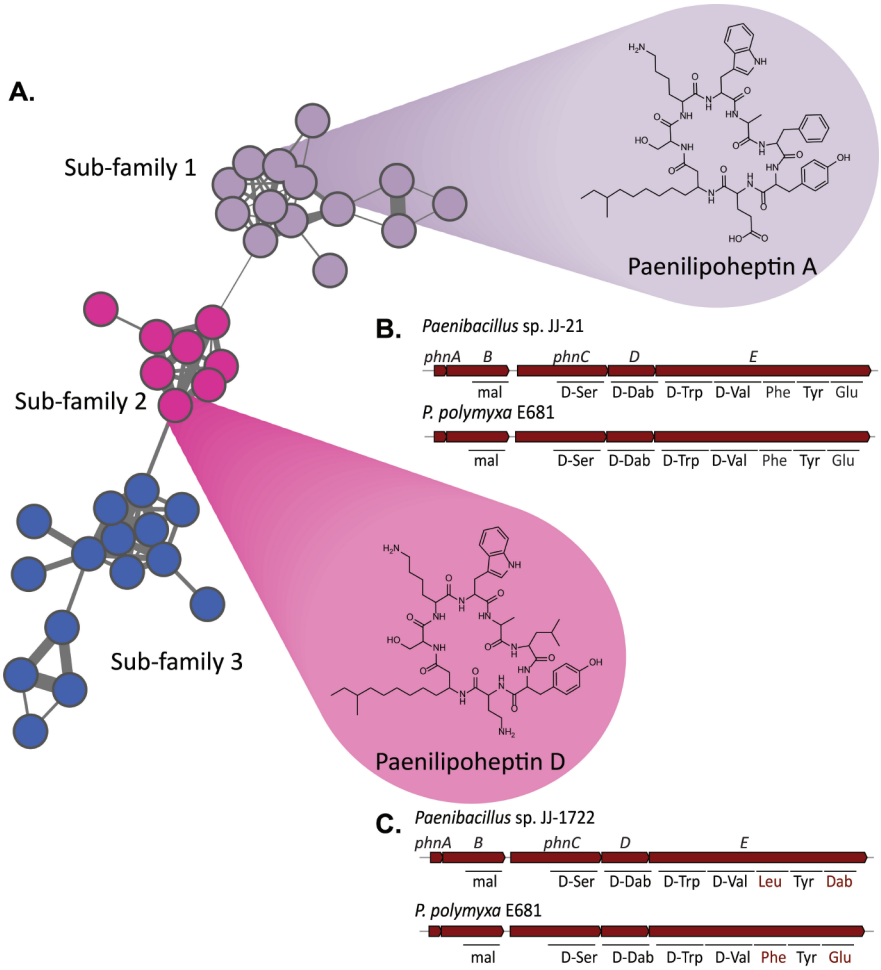


Figure 6. Molecular family of the Lys-containing specialized metabolites highlighted by MassQL search. Exploring the mass features from sub-family 1 and 2 led to the discovery of paenilipoheptin A and D, respectively (**A**). The BGCs of paenilipoheptin A from the genome of *Paenibacillus* sp. JJ-21 (**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).

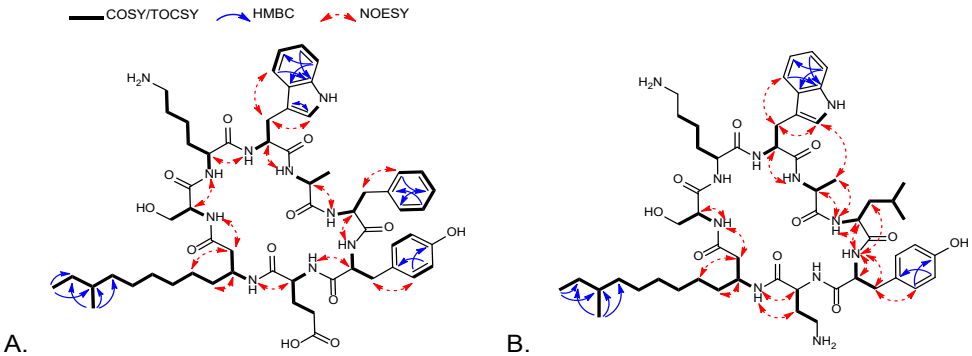


Figure 7. Correlations obtained by COSY, HMBC, and NOESY measurements in NMR of paenilipoheptin A (A) and paenilipoheptin D (B).

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.

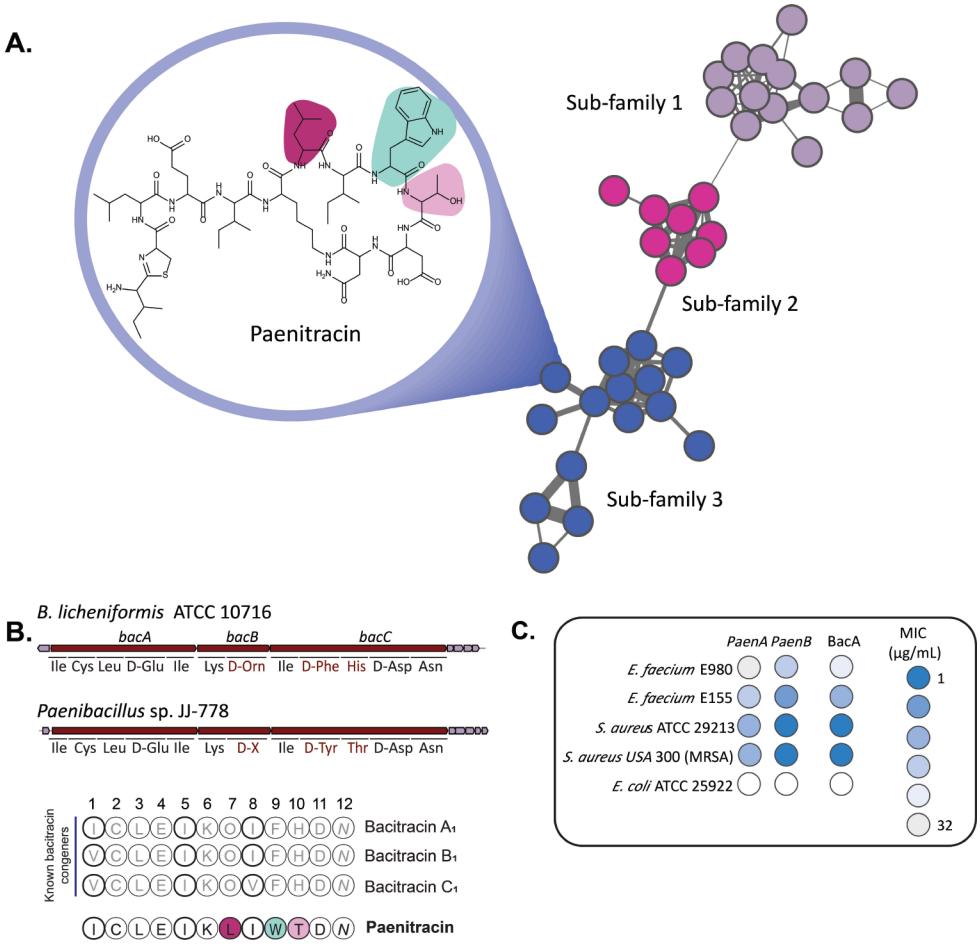


Figure 8. Molecular family of the Lys-containing specialized metabolites highlighted by MassQL search. A. Exploring the mass features from sub-family 3 led to the discovery of new bacitracin analog, designated as paenitracin. **B.** The BGC and its predicted amino acid substrates of new paenitracin from 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**).

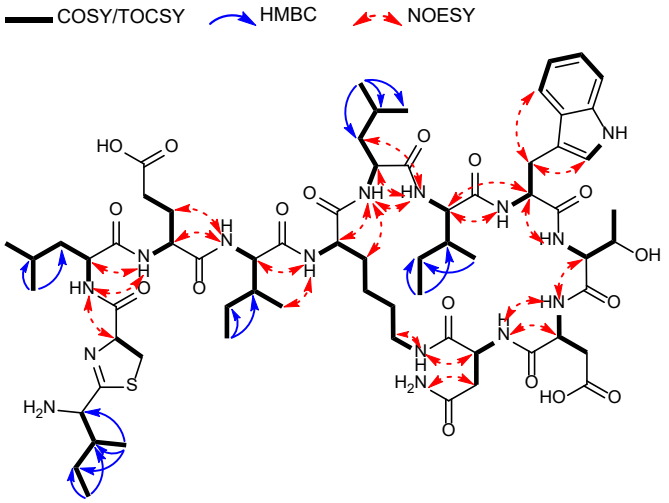


Figure 9. Correlations obtained by COSY, HMBC, and NOESY measurements in NMR of paenitracin A (3).

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).

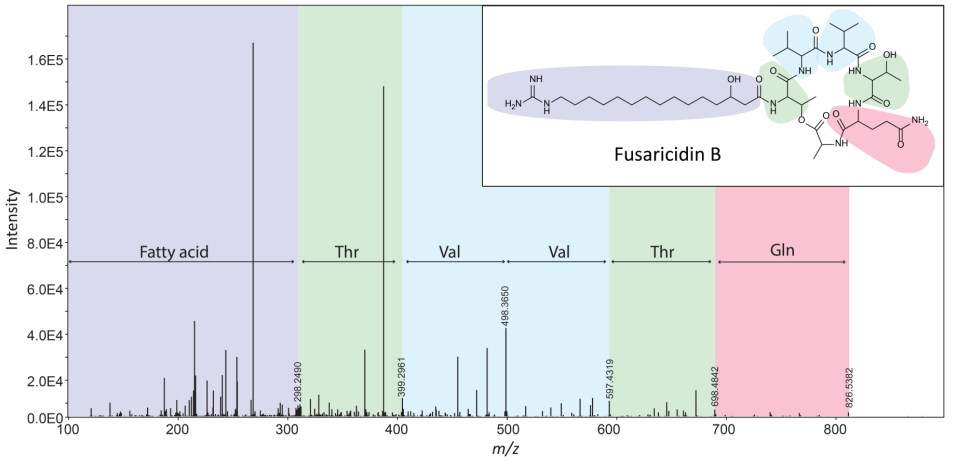


Figure S1. MS/MS spectrum of fusaricidin B (precursor ion $[M+2H]^{2+}$ *m/z* 449.2909). The assignment of the sequence of amino acid residues is based on the mass differences between the consecutive *b* ions.

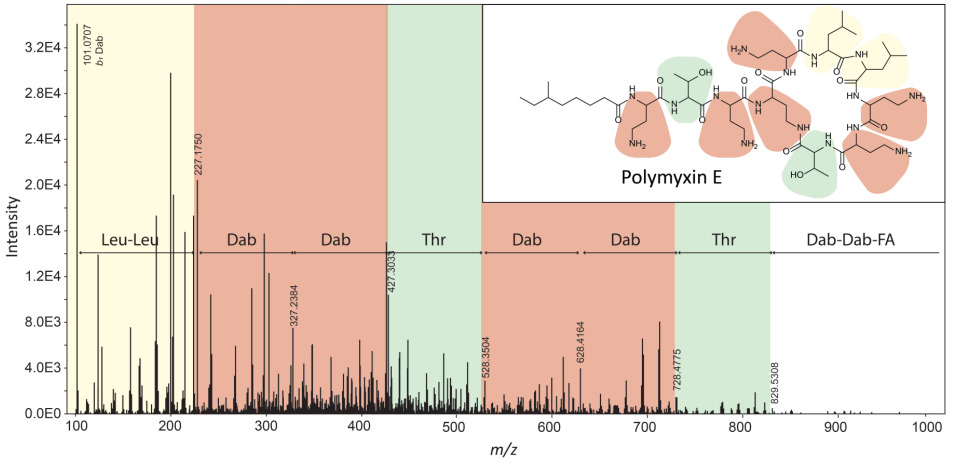


Figure S2. MS/MS spectrum of polymyxin E (precursor ion $[M+3H]^{3+}$ *m/z* 390.596). The assignment of the sequence of amino acid residues is based on the mass differences between the consecutive *b* ions.

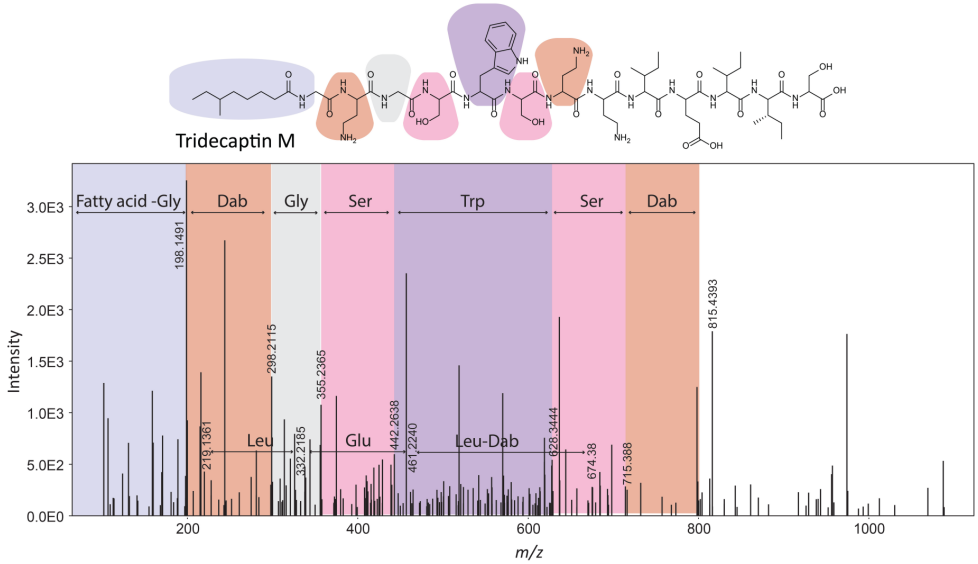


Figure S3. MS/MS spectrum of tridecaptin M (precursor ion $[M+3H]^{3+}$ *m/z* 496.9524). The assignment of the sequence of amino acid residues is based on the mass differences between the consecutive *b* ions.

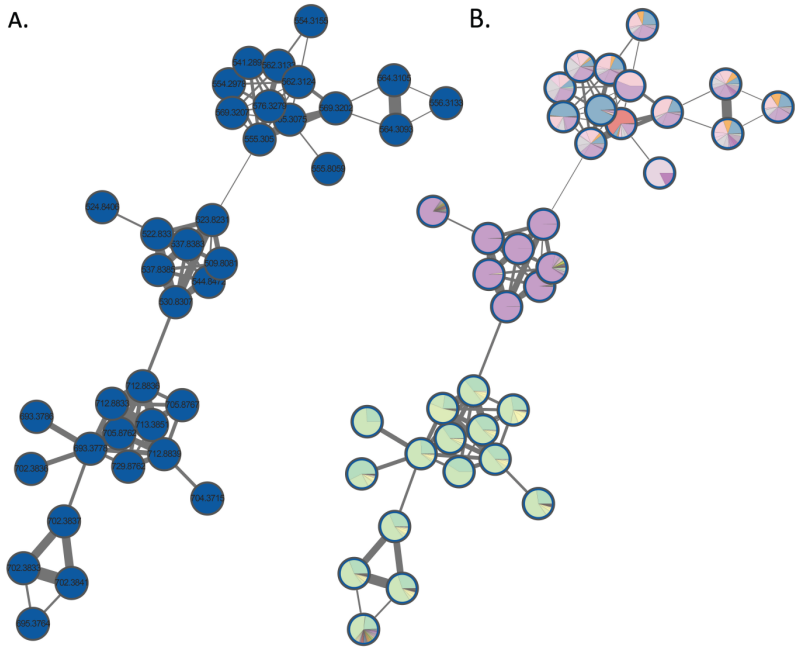


Figure S4. A. The MassQL-annotated molecular family of lysine-containing natural products. **B.** On the same molecular family, pie charts were mapped to the nodes to represent the relative precursor ion intensities within each extract.

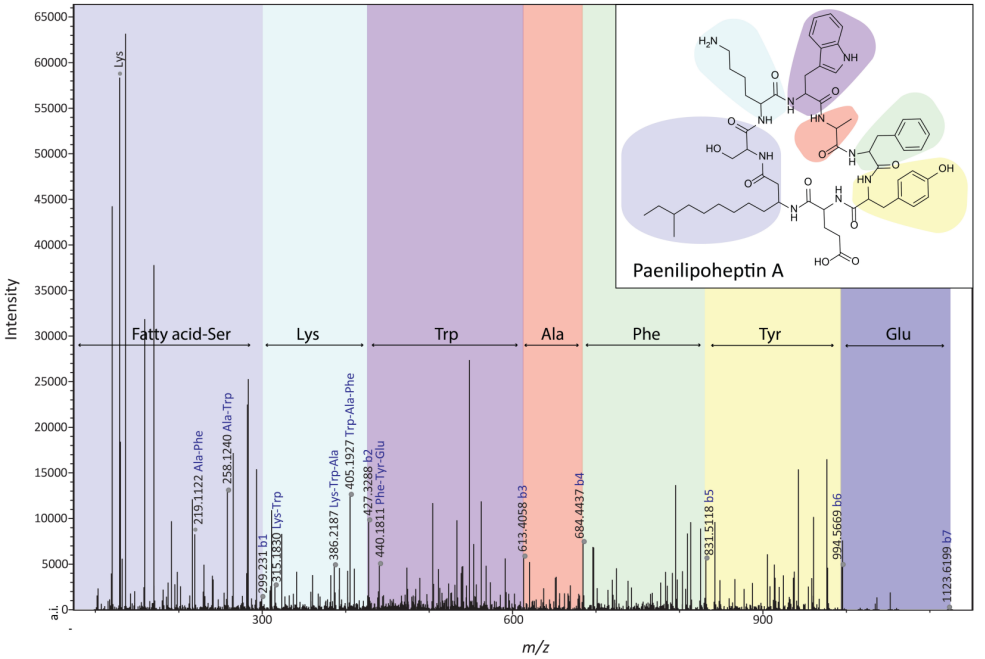


Figure S5. MS/MS spectrum of paenilipoheptin A (precursor ion $[M+2H]^{2+}$ *m/z* 562.3130). The assignment of the sequence of amino acid residues is based on the mass differences between the consecutive *b* ions.

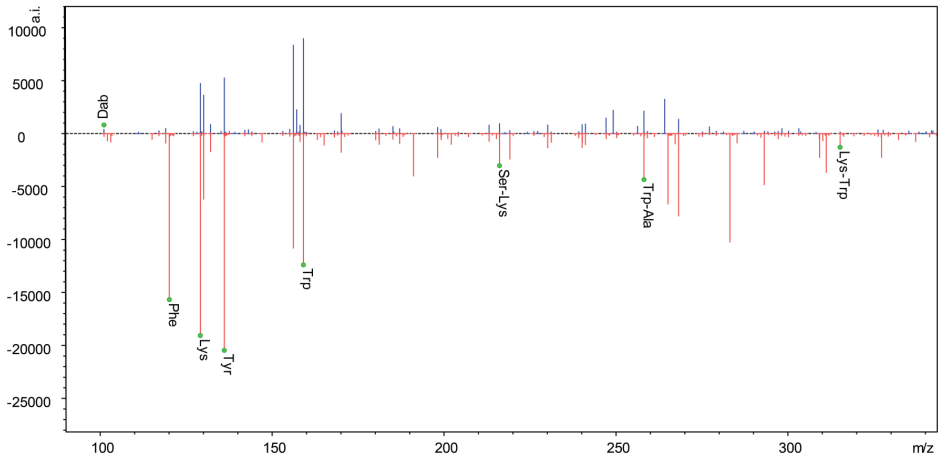


Figure S6. Direct MS/MS spectra comparison of the mass feature having *m/z* of 555.305 (red) with the mass feature having *m/z* of 523.8231 (blue).

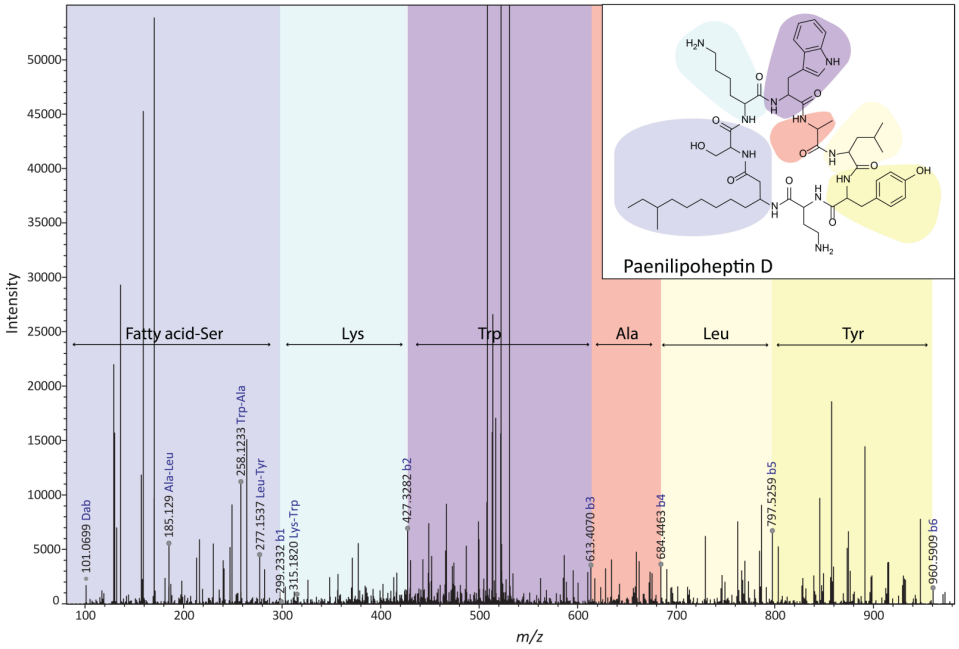


Figure S7. MS/MS spectrum of paenilipoheptin D (precursor ion $[M+2H]^{2+}$ *m/z* 530.8309). The assignment of the sequence of amino acid residues is based on the mass differences between the consecutive *b* ions.

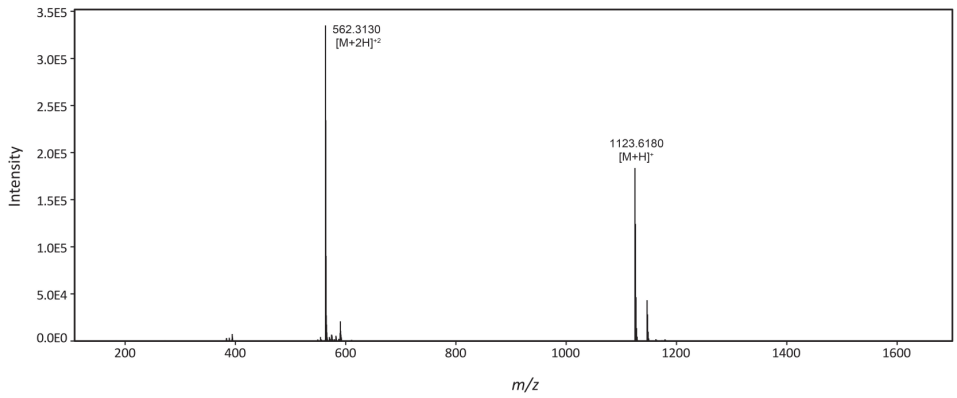


Figure S8. HRMS spectrum of 1.

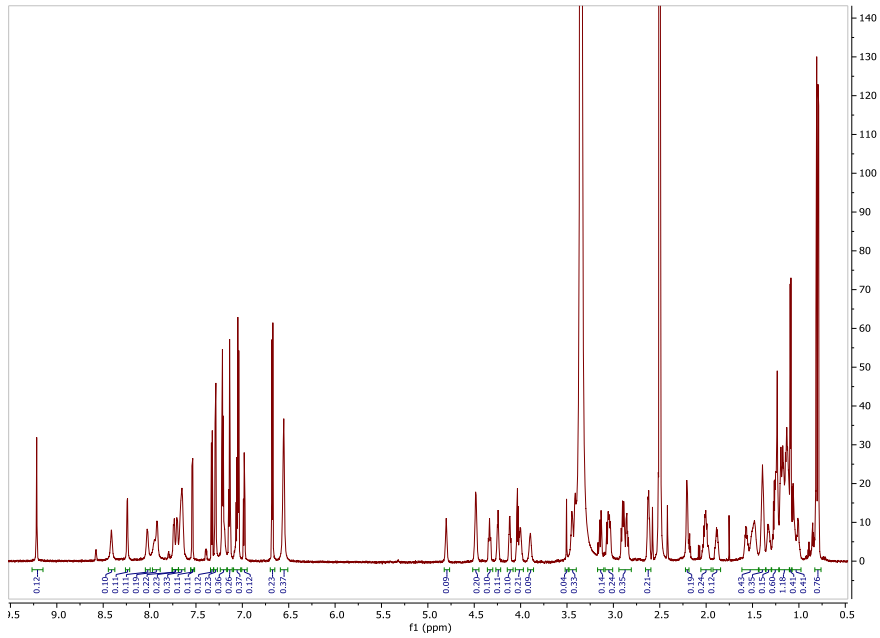


Figure S9. ^{1}H NMR spectrum of 1 (850 MHz, in DMSO-d$_6$).

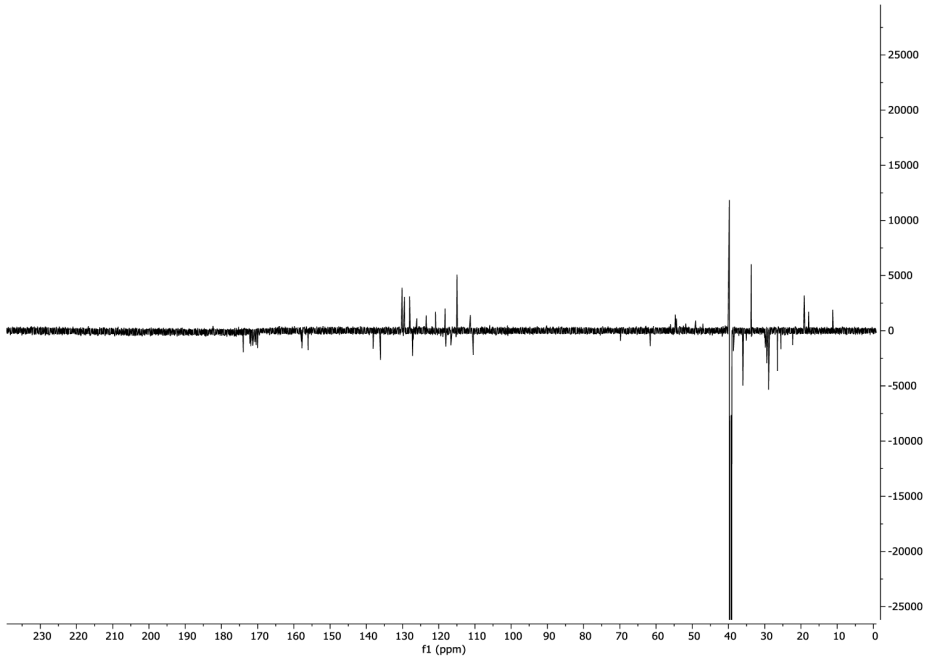


Figure S10. APT spectrum of 1 (850 MHz, in DMSO-d$_6$).

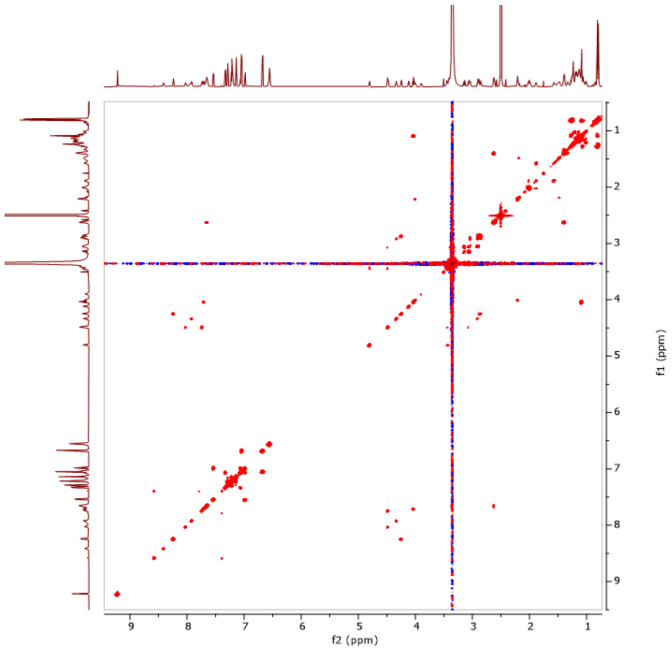


Figure S11. $^{1}H–^{1}H$ COSY spectrum of 1 (850 MHz, in DMSO-d_6).

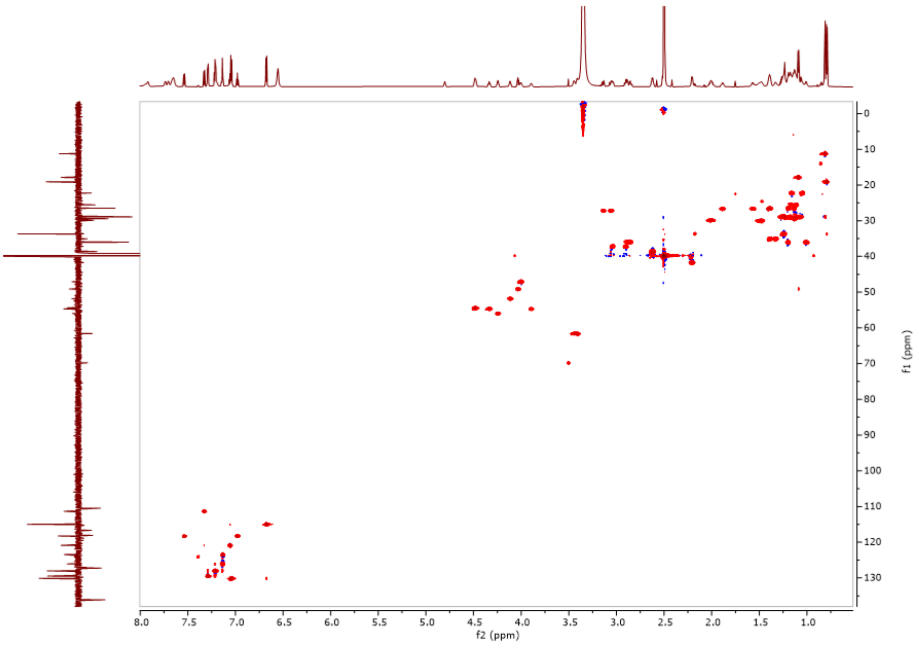


Figure S12. HSQC spectrum of 1 (850 MHz, in DMSO-d_6).

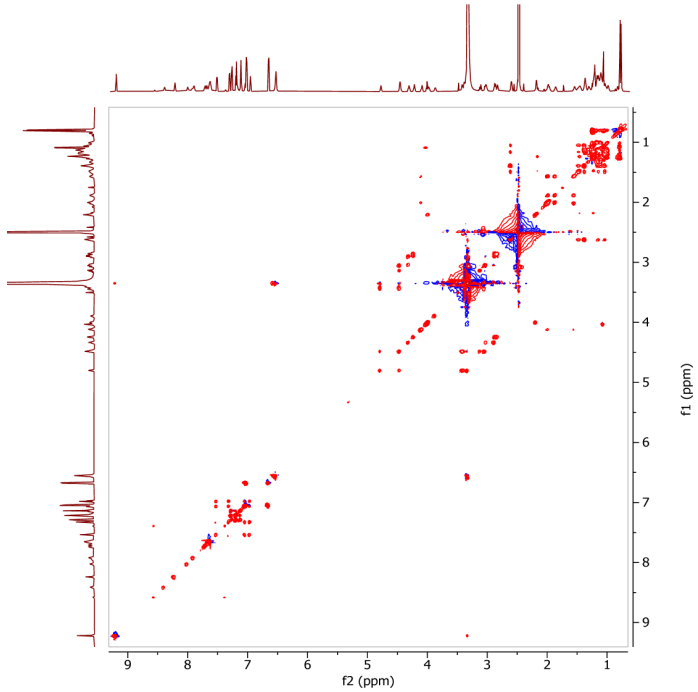


Figure S13. $^{1}H–^{1}H$ TOCSY spectrum of 1 (850 MHz, in DMSO-d_6).

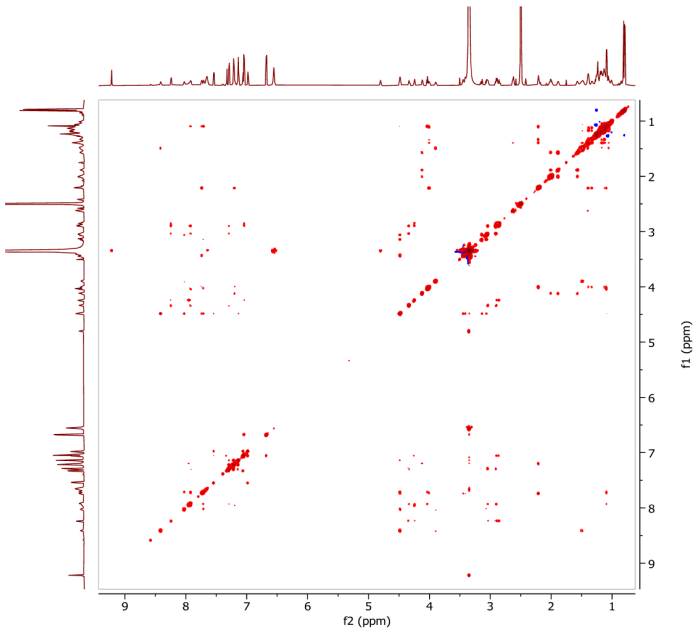


Figure S14. NOESY spectrum of 1 (850 MHz, in DMSO-d_6).

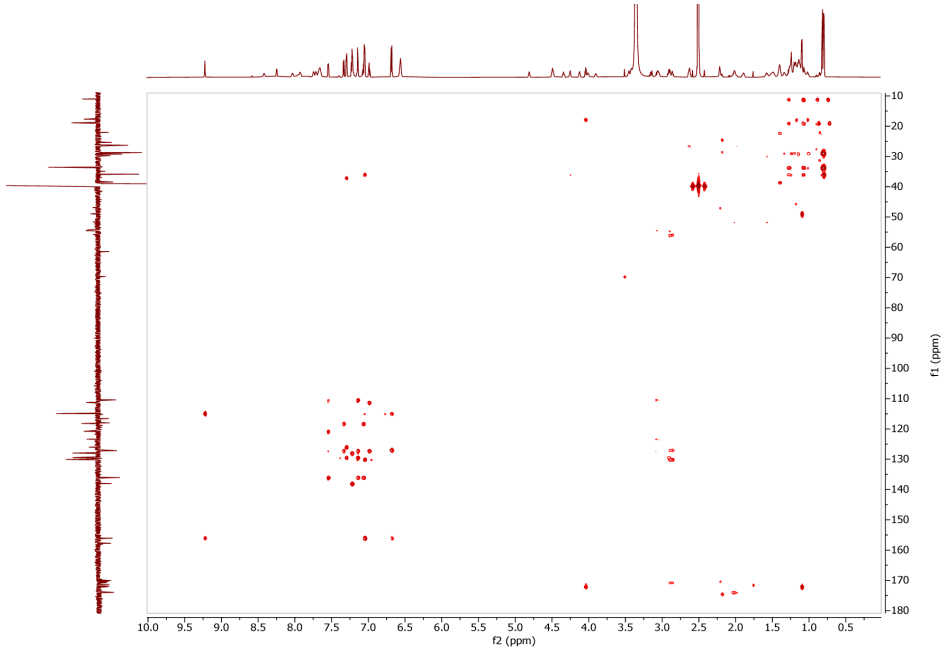


Figure S15. HMBC spectrum of 1 (850 MHz, in DMSO-d_6).

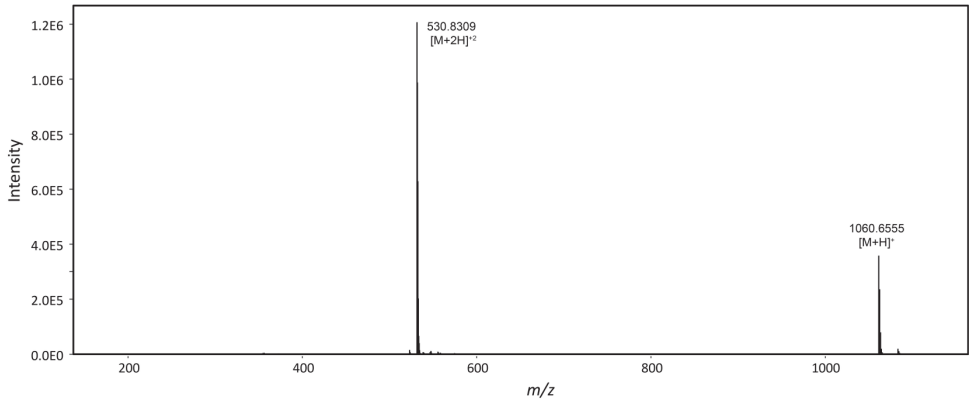


Figure S16. HRMS spectrum of 2.

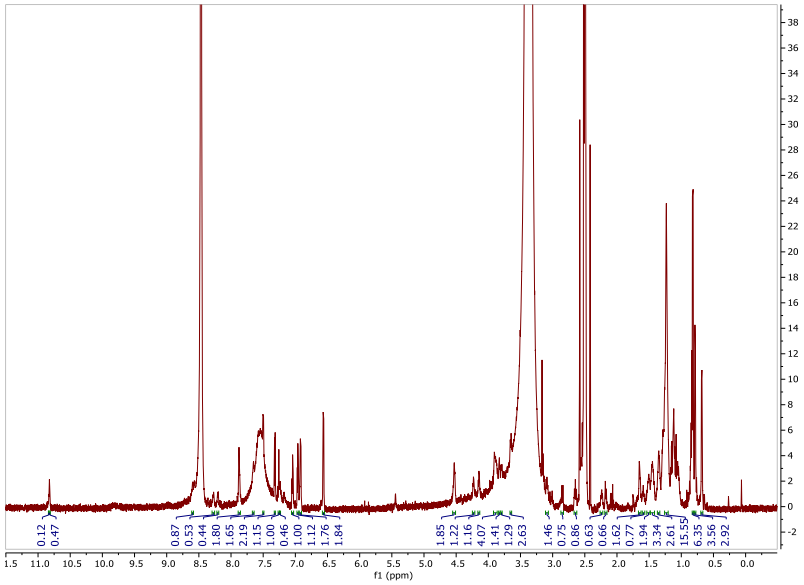


Figure S17. ^{1}H NMR spectrum of 2 (850 MHz, in DMSO-d_6).

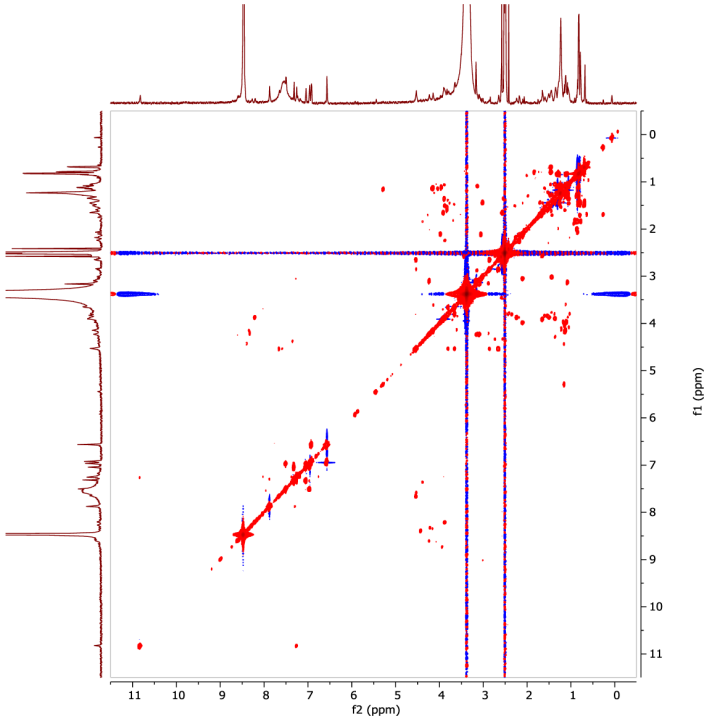


Figure S18. ^{1}H–^{1}H COSY spectrum of 2 (850 MHz, in DMSO-d_6).

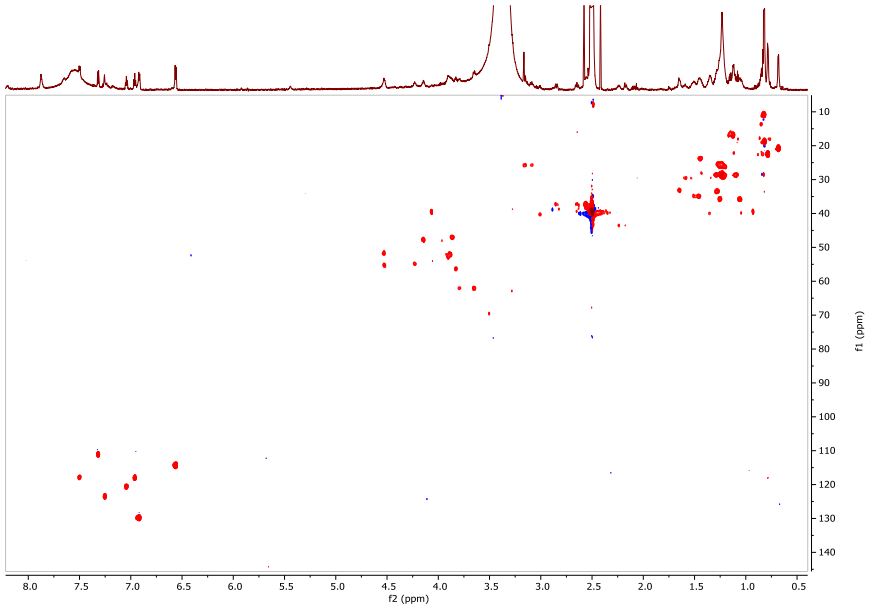


Figure S19. HSQC spectrum of 2 (850 MHz, in DMSO-d_6).

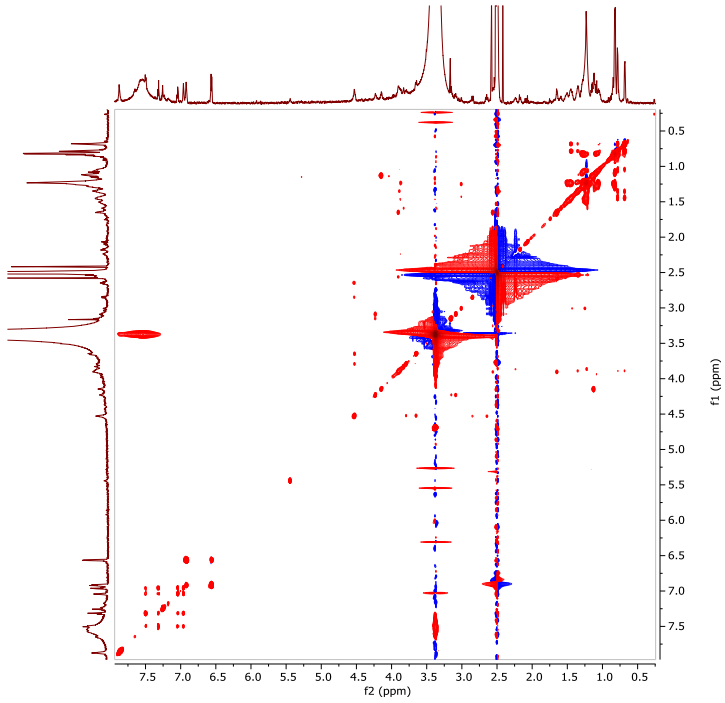


Figure S20. ^{1}H–^{1}H TOCSY spectrum of 2 (850 MHz, in DMSO-d_6).

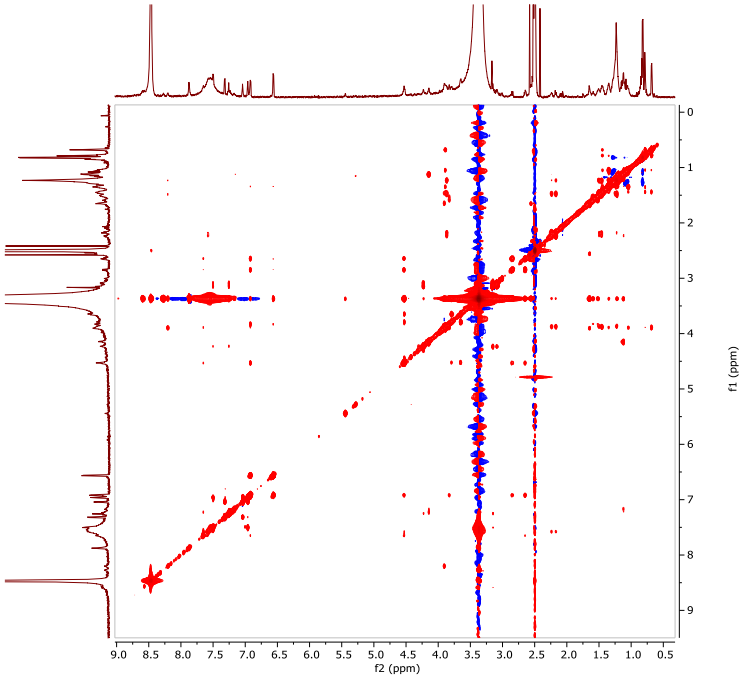


Figure S21. NOESY spectrum of 2 (850 MHz, in DMSO-d_6).

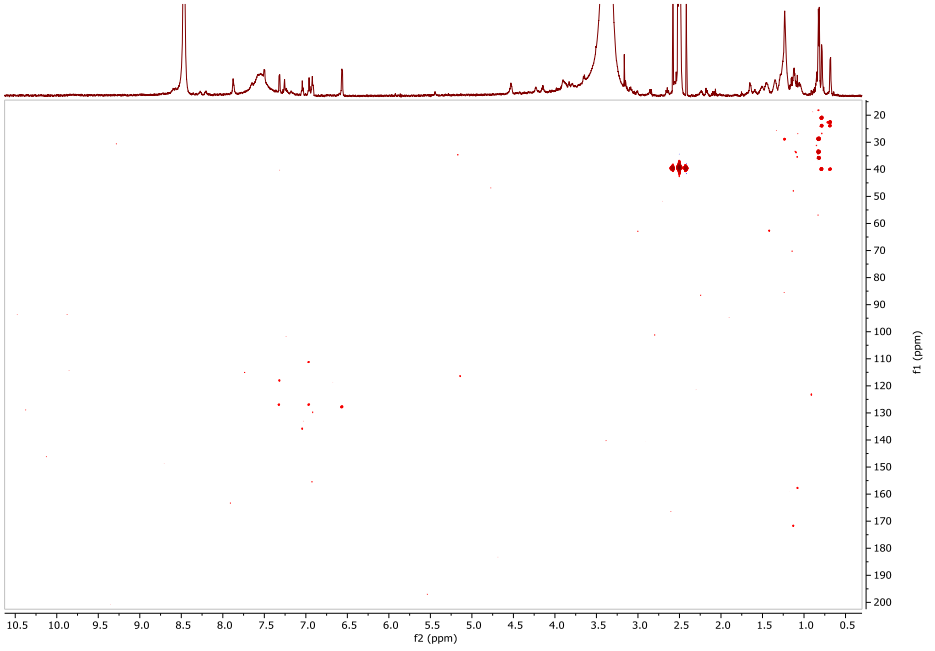


Figure S22. HMBC spectrum of 2 (850 MHz, in DMSO-d_6).

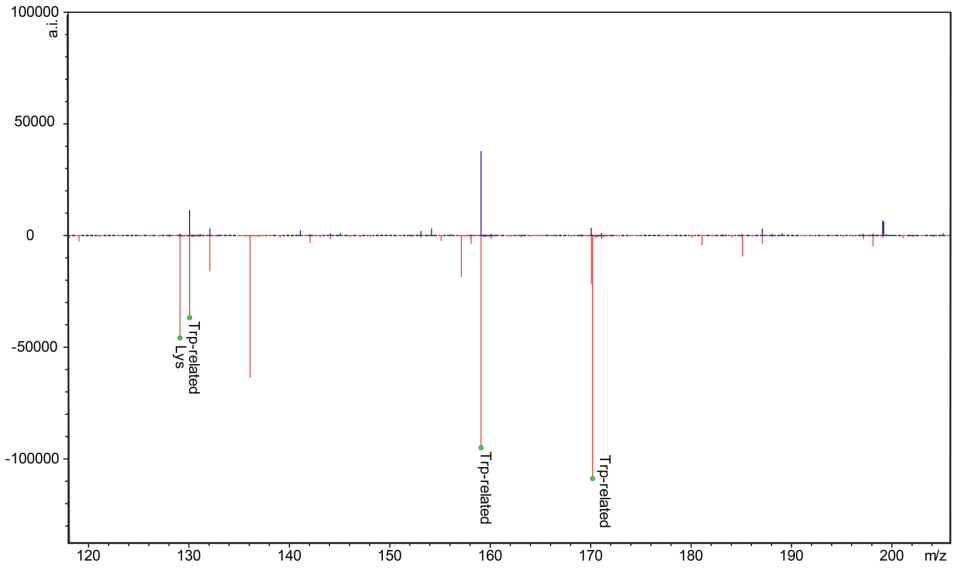


Figure S23. Direct MS/MS spectra comparison of the mass feature with *m/z* 712.8836 (blue) with the mass feature with m/z 530.8307 (red).

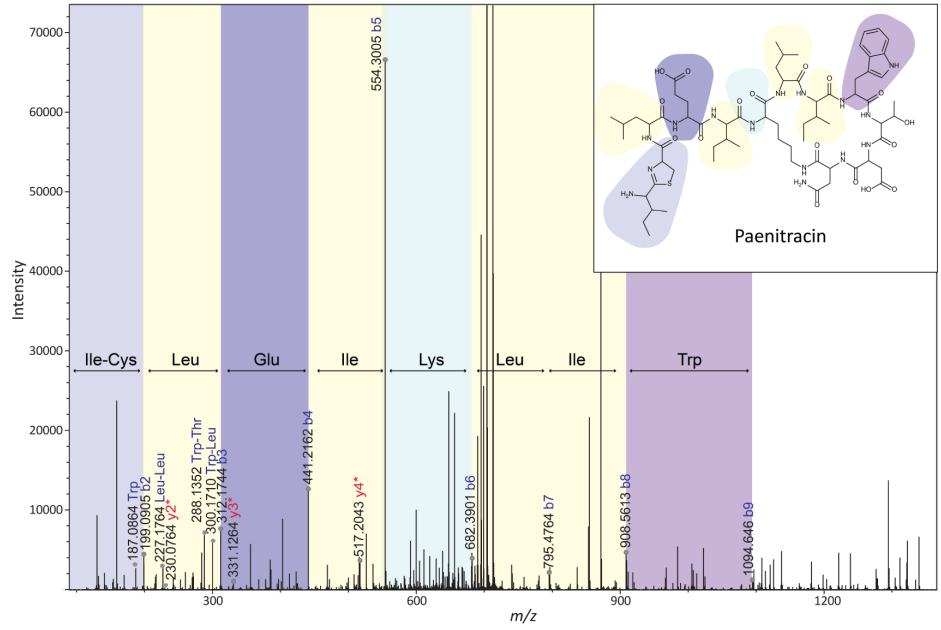


Figure S24. MS/MS spectrum of paenitracin A (precursor ion $[M+2H]^{2+}$ *m/z* 712.8843). The assignment of the sequence of amino acid residues is based on the mass differences between the consecutive *b* ions.

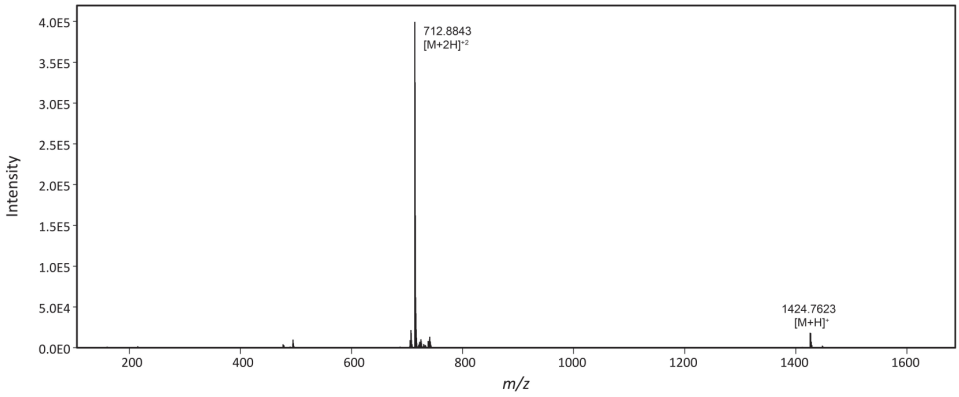


Figure S25. HRMS spectrum of 3.

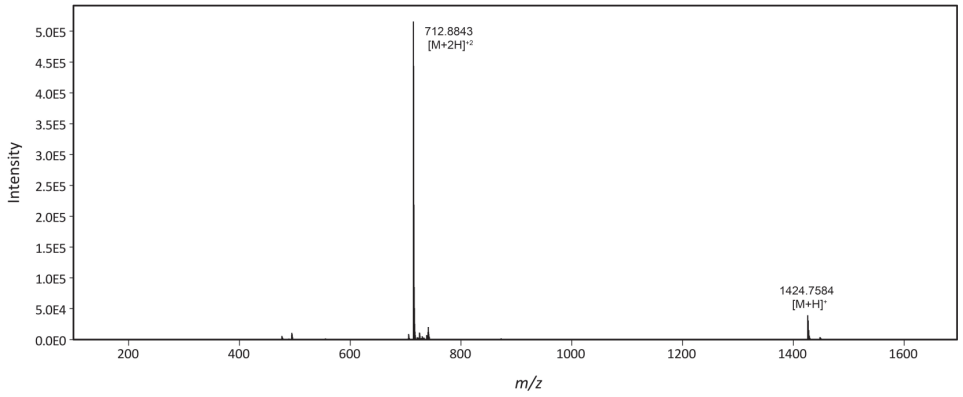


Figure S26. HRMS spectrum of 4.

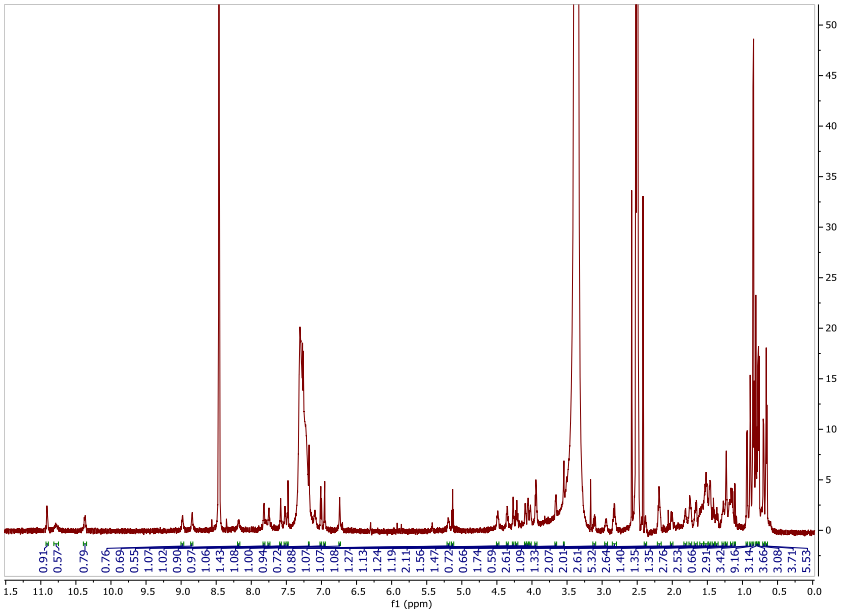


Figure S27. ^{1}H NMR spectrum of 3 (850 MHz, in DMSO-d$_6$).

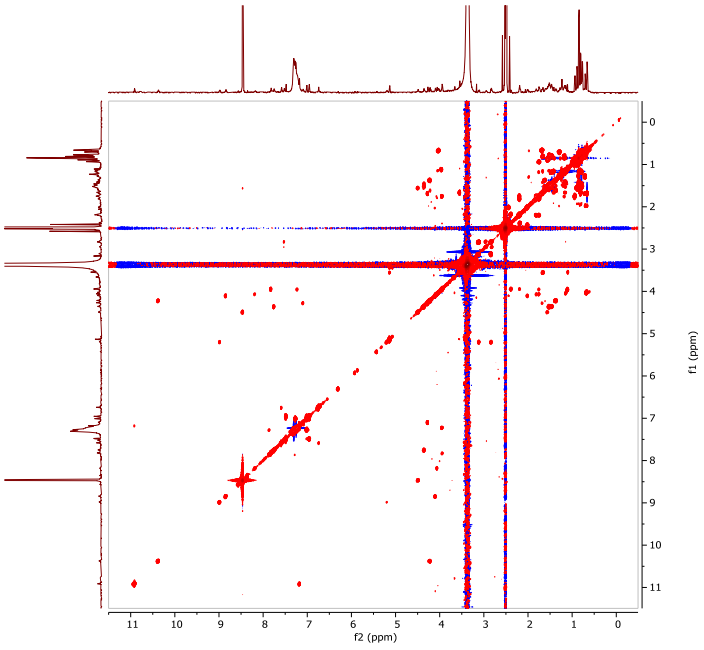


Figure S28. ^{1}H–^{1}H COSY spectrum of 3 (850 MHz, in DMSO-d$_6$).

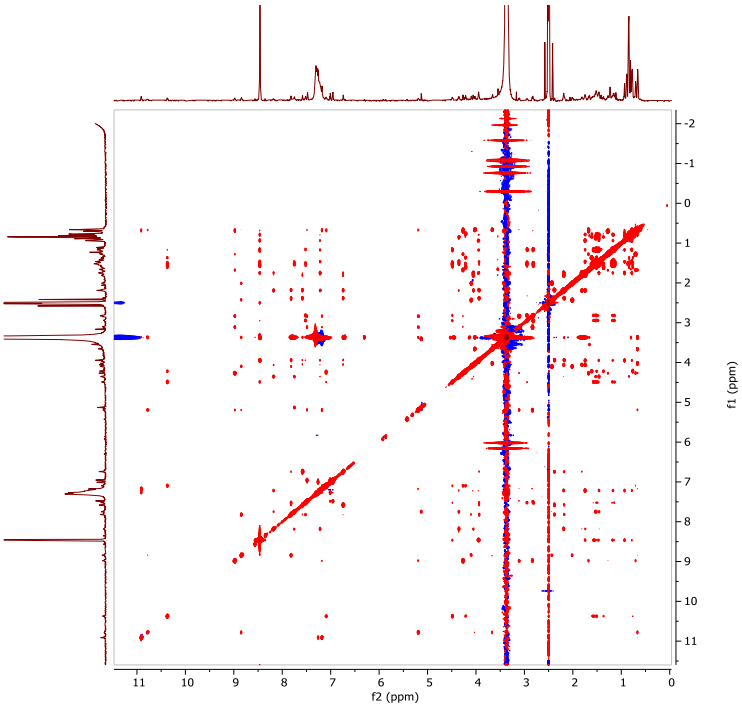


Figure S29. NOESY spectrum of 3 (850 MHz, in DMSO-d_6).

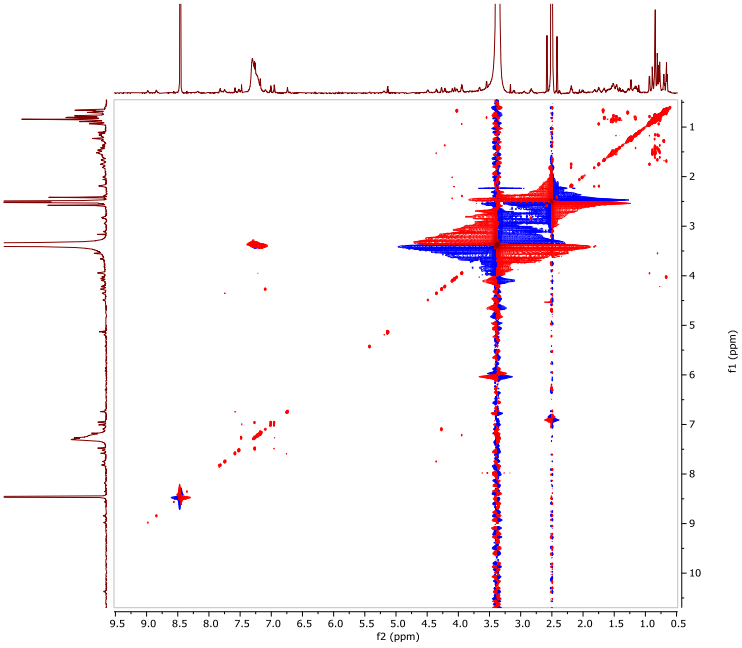


Figure S30. ^{1}H–^{1}H TOCSY spectrum of 3 (850 MHz, in DMSO-d_6).

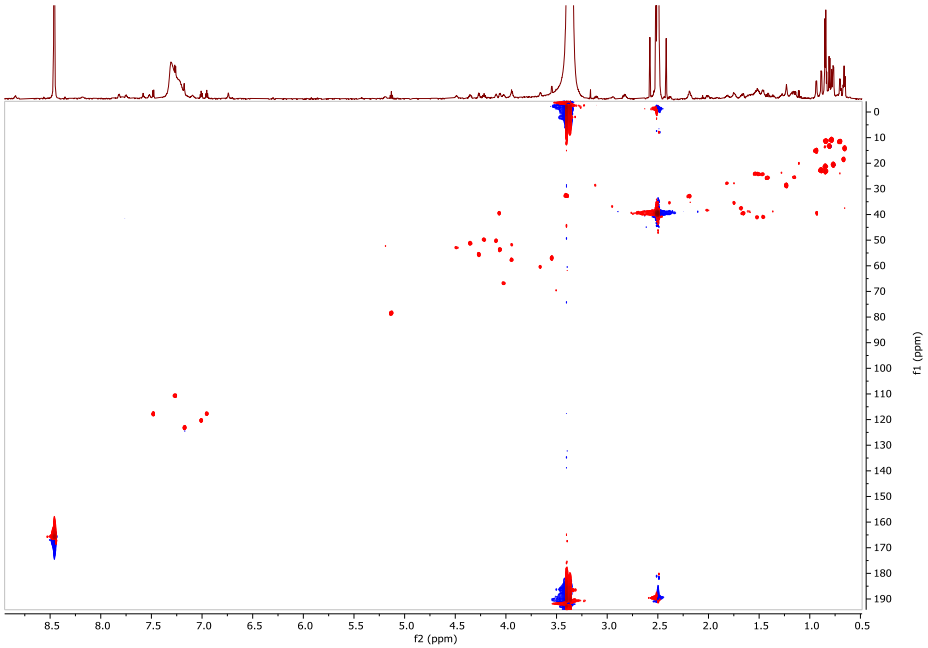


Figure S31. HSQC spectrum of 3 (850 MHz, in DMSO-d_6).

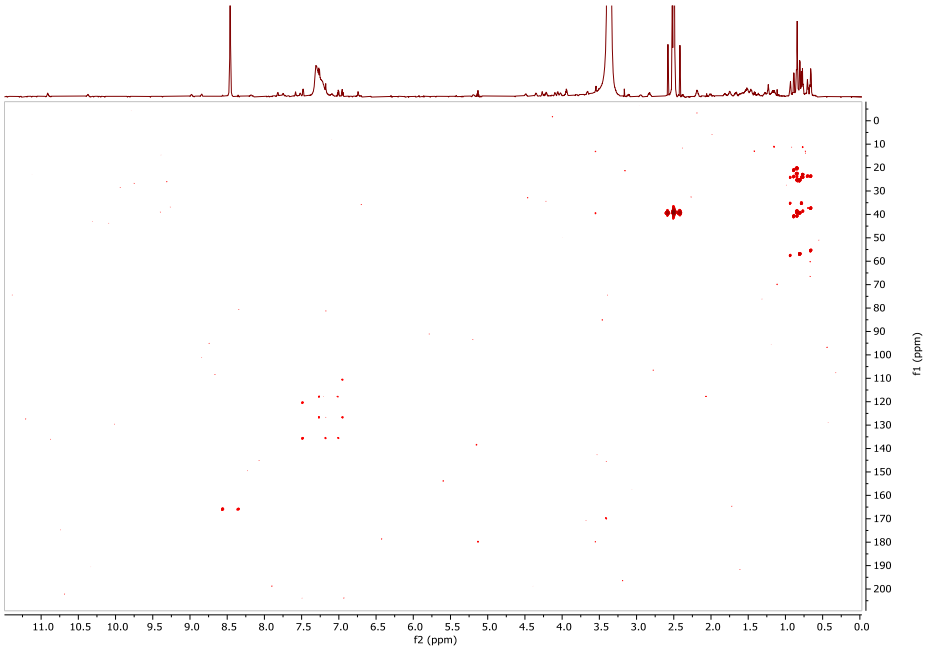


Figure S32. HBMC spectrum of 3 (850 MHz, in DMSO-d_6).

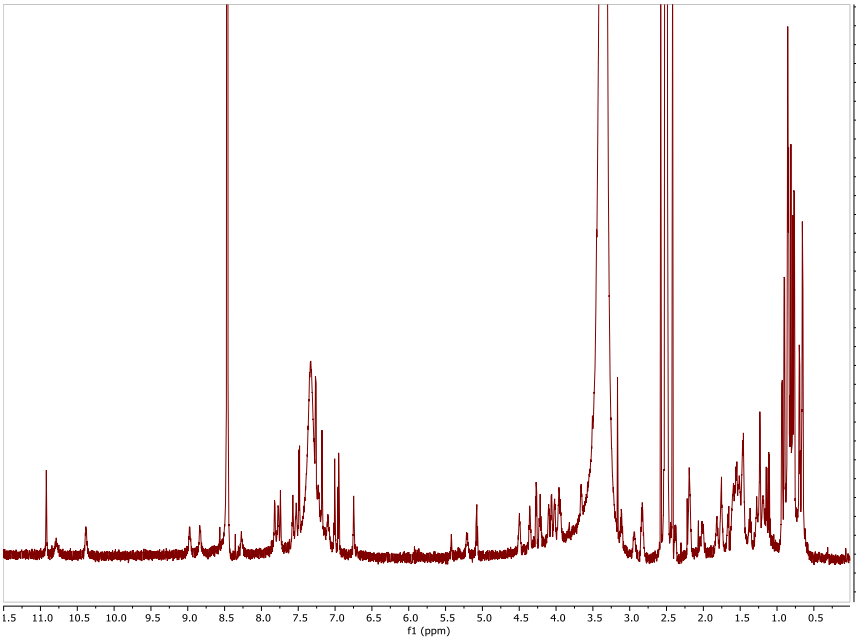


Figure S33. ^{1}H NMR spectrum of 4 (850 MHz, in DMSO-d$_6$).

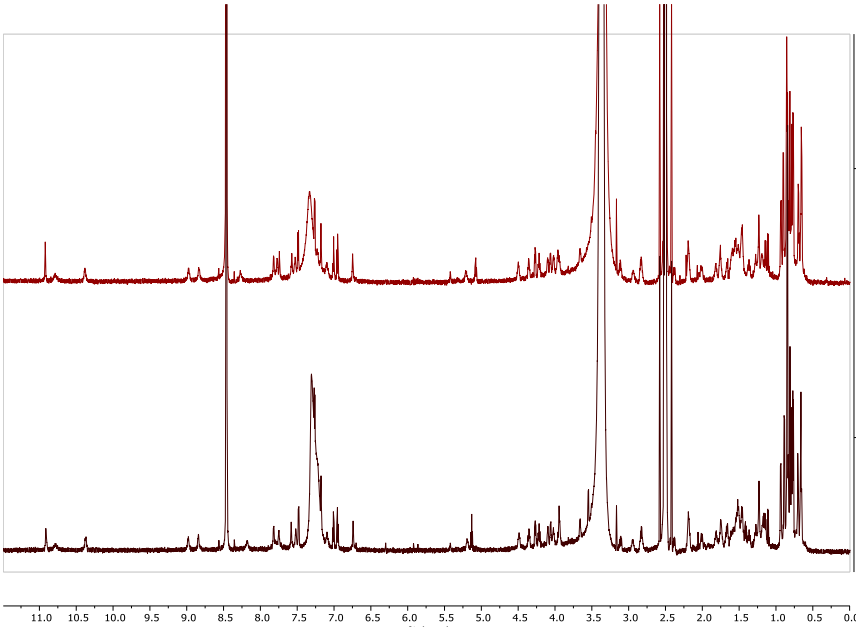


Figure S34. Stacked ^{1}H NMR spectra of 3 and 4 (850 MHz, in DMSO-d$_6$).



Chapter 4

Discovery of novel glycerolated quinazolinones from *Streptomyces* sp. MBT27

Nataliia V. Machushynets, Changsheng Wu, Somayah S. Elsayed, Thomas Hankemeier and Gilles P. van Wezel

This chapter was published as:
Machushynets N.V., Wu C., Elsayed S.S., Hankemeier T., & van Wezel G.P. (2019). Discovery of novel glycerolated quinazolinones from *Streptomyces* sp. MBT27. *J Ind Microbiol Biotechnol.,* **46**, 483–492.

Abstract

Actinobacteria are a major source of novel bioactive natural products. A challenge in the screening of these microorganisms lies in finding favorable growth conditions for secondary metabolite production and dereplication of known molecules. Here, we report that *Streptomyces* sp. MBT27 produces 4-quinazolinone alkaloids in response to elevated levels of glycerol, whereby quinazolinones A (**1**) and B (**2**) form a new sub-class of this interesting family of natural products. Global Natural Product Social molecular networking (GNPS) resulted in a quinazolinone-related network that included anthranilic acid (**3**), anthranilamide (**4**), 4(3H)-quinazolinone (**5**) and 2,2-dimethyl-1,2-dihydroquinazolin-4(3*H*)-one (**6**). Actinomycins D (**7**) and X_2 (**8**) were also identified in the extracts of *Streptomyces* sp. MBT27. The induction of quinazolinone production by glycerol combined with biosynthetic insights provides evidence that glycerol is integrated into the chemical scaffold. The unprecedented 1,4-dioxepane ring that is spiro-fused into the quinazolinone backbone, is most likely formed by intermolecular etherification of two units of glycerol. Our work underlines the importance of varying the growth conditions for the discovery of novel natural products and for understanding their biosynthesis.

Introduction

Actinobacteria are a major source of bioactive compounds, producing some two-thirds of all antibiotics as well as molecules with a wide variety of activities such as anticancer, antifungal and immunosuppressant (Barka *et al.*, 2016, Bérdy, 2005). Traditionally, microbial natural product (NP) discovery has been done via high-throughput screening followed by iterative bioassay-guided fractionation and structure elucidation. While such pipelines were extremely successful and delivered a plethora of therapeutic agents, in the modern era the large pharmaceutical companies moved out of NP-discovery programs due to high cost and chemical redundancy (Baltz, 2008, Cooper & Shlaes, 2011, Payne *et al.*, 2007). At the same time, the power of genome sequencing brought the recognition that microorganisms harbor a vast and yet untapped biosynthetic potential, and it rapidly became clear that the potential for metabolic diversity of even the best-studied model organisms as producers of natural products had been grossly underestimated (Bentley *et al.*, 2002, Cruz-Morales *et al.*, 2013, Ikeda *et al.*, 2003). How could these compounds have been missed by the very extensive high-throughput (HT) screening campaigns of the twentieth century? The answer is that many of the biosynthetic gene clusters (BGCs) discovered by genome mining are poorly expressed or cryptic under laboratory conditions (Kolter & van Wezel, 2016, Nett *et al.*, 2009). A drug-discovery pipeline that is rapidly gaining momentum involves combining genome mining with fluctuating the culturing conditions to achieve differential synthesis of NPs, followed by the metabolic profiling-based identification of the bioactivity of interest (Du & van Wezel, 2018, Wu *et al.*, 2015a). A major challenge thereby lies in finding the appropriate chemical triggers or ecological cues to elicit the production of cryptic antibiotics (Rutledge & Challis, 2015, Yoon & Nodwell, 2014, Zhu *et al.*, 2014a). The use of chemical elicitors is thereby a promising approach (Craney *et al.*, 2012, Zhu *et al.*, 2014a).

Manipulation of fermentation conditions for promising producer strains, known as the "one strain many compounds" (OSMAC) approach, is an effective way of enhancing the production of secondary metabolites (Romano *et al.*, 2018, Bode *et al.*, 2002). The regulatory networks that control the production of bacterial natural products respond strongly to changes in carbon, nitrogen or phosphate concentration (van der Heul *et al.*, 2018, Urem *et al.*, 2016). The re-isolation of known metabolites is a major bottleneck in the discovery of new bioactive natural products. A crucial step in this regard is the early identification of already known substances, to concentrate the efforts on the discovery of new ones, a process known as dereplication (Gaudencio & Pereira, 2015). Current dereplication strategies include hyphenated techniques, such as LC-MS, LC-NMR, LC-NMR-MS, and LC-SPE-NMR. Bioactivity fingerprinting has also been used to dereplicate natural products based on their biological modes of action (Schulze *et al.*, 2013), while molecular networking is a powerful tool for the visualization and dereplication of natural products (Yang *et al.*, 2013, Winnikoff

et al., 2014). Mass spectrometry-based molecular networking relies on the clustering of molecules based on similarities in their MS/MS fragmentation patterns, which depends on the structural features of the ionized molecules. The resulting clusters allow scientists to visually explore the metabolites produced by a given strain under a specific growth condition, allowing rapid dereplication of known compounds by automated spectral library searches, and to visualize their unknown structural analogs (Wang *et al.*, 2016).

In this work, we analyzed the potential of *Streptomyces* sp. MBT27 as a producer of natural products in response to changes in the carbon source. The strain had previously been identified as a promising producer of NPs (Zhu *et al.*, 2014b). Extensive fluctuations in the secondary metabolite profiles were observed depending on the carbon source used, and statistical methods combined with GNPS molecular networking identified a family of known, as well as novel quinazolinone compounds, in response to high concentrations of glycerol. Quinazolinones are heterocyclic compounds with a wide range of medical applications, such as antimicrobial, antiviral, antituberculosis and, as enzyme inhibitors (Kshirsagar, 2015, Hameed *et al.*, 2018, Jafari *et al.*, 2016). A combination of MS and NMR methods identified the novel quinazolinones A (**1**) and B (**2**), which further expands the chemical space of this rich family of natural products.

Results and Discussion

The influence of carbon sources on secondary metabolite production

Previous screening of our in-house actinomycete collection, obtained from remote mountain soils, showed that it is a promising source of new bioactive compounds (Zhu *et al.*, 2014b). Under specific growth conditions, these isolates exert potent inhibitory activity against the so-called ESKAPE pathogens *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacter* spp. (Zhu *et al.*, 2014b) In the current study, the potential of one of the *Streptomyces* strains in our collection, namely *Streptomyces* sp. MBT27, was analyzed to study the effect of carbon sources on its metabolite profile. Traditional approaches, such as changes in fermentation conditions, are known to induce significant changes in the microbial metabolome. Culture medium components, and particularly the carbon source, have major effects on the production of secondary metabolites (van Wezel *et al.*, 2009, Sanchez *et al.*, 2010, Martin *et al.*, 2011, Bode *et al.*, 2002). The differential production of small molecules is ideal for metabolomics studies, whereby metabolic variations are correlated statistically with bioactivity, thus facilitating the identification of the bioactive molecule of interest (Wu *et al.*, 2015a, Zhu *et al.*, 2014a). To establish the potential of *Streptomyces* sp. MBT27, the strain was grown in liquid minimal media containing different carbon sources, namely (percentages in w/v): 1% of both mannitol and glycerol, 1% mannitol, 2% mannitol, 1% glycerol, 2% glycerol, 1% glucose, 2% glucose, 1% fructose, 1% arabinose, or 1% *N*-acetylglucosamine (GlcNAc). The latter is an elicitor of antibiotic production, via metabolic interference with the global nutrient sensory network controlled by DasR (Swiatek-Polatynska *et al.*, 2015, Rigali *et al.*, 2008). Phosphate buffer was omitted from the MM as it repressed the production of secondary metabolites by *Streptomyces* sp. MBT27 (data not shown). To identify the secondary metabolites in the cultures, supernatants were extracted with EtOAc, and the resulting crude extracts were subjected to LC-MS analysis. Using the LC-MS data, a heat map with added hierarchical clustering was generated, to visualize the production of different metabolites under different culture conditions (Figure 1).

Hierarchical clustering analysis of the LC-MS data allows effective comparative analysis of metabolomics data, and the heat map revealed major differences in the metabolic profiles of *Streptomyces* sp. MBT27, whereby different groups of metabolites were enhanced depending on the carbon source used. Interestingly, not only the type of carbon source, but especially also the concentration resulted in large changes in the metabolic profiles. Doubling the concentration of either glycerol or glucose from 1 to 2% had a profound effect on the metabolic profile. Thin layer chromatography (TLC) was conducted to compare metabolic profiles of the 1% and 2% glycerol-grown cultures. Interestingly, this revealed that several fluorescent compounds were differentially produced in the extracts of 2% glycerol-grown

cultures relative to those produced in 1% glycerol (Figure 2A). To provide statistical relevance to the data, the metabolic profiles of 1% and 2% glycerol-grown cultures were compared using a volcano plot (Figure 2B). The volcano plot was then searched for the mass features which increased in production in 2% glycerol as compared to 1%. A compound with *m/z* of 281.1151 (**1**) stood out as its intensity had increased by around 7000 fold (*p*-value = 0.002) in cultures fermented in 2% glycerol as compared to 1%.

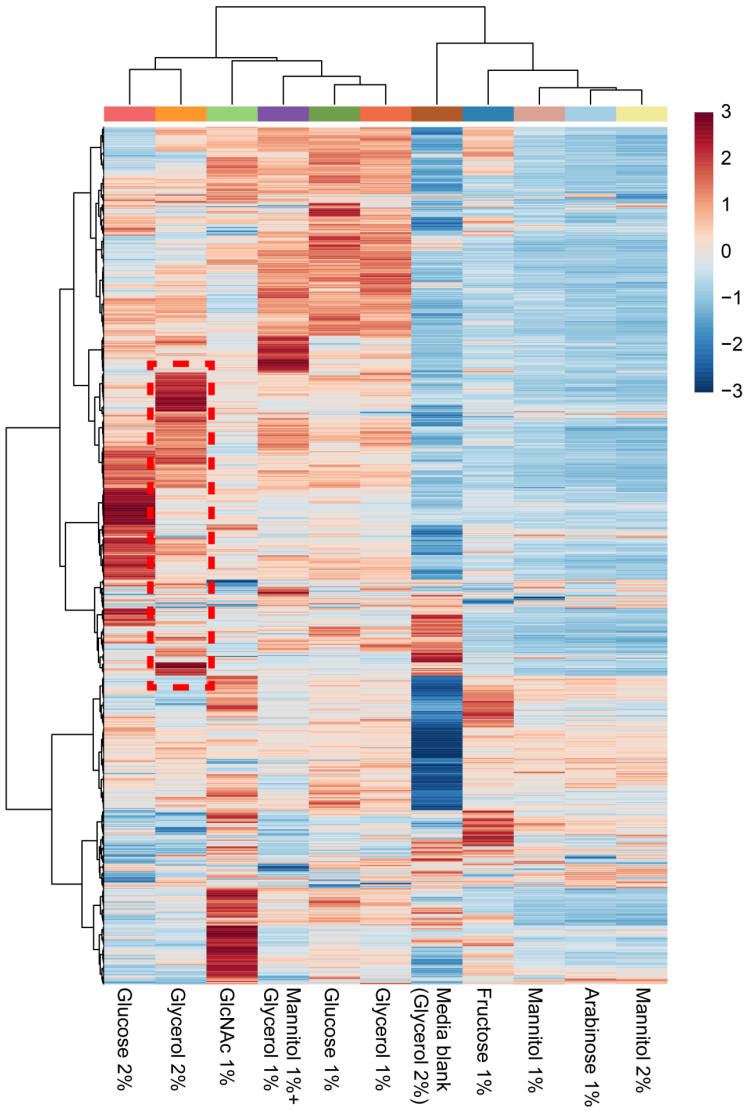


Figure 1. Analysis of secondary metabolites produced by *Streptomyces* sp. MBT27. The heat map depicts the relative abundance of the metabolites (rows) produced under different growth conditions (columns).

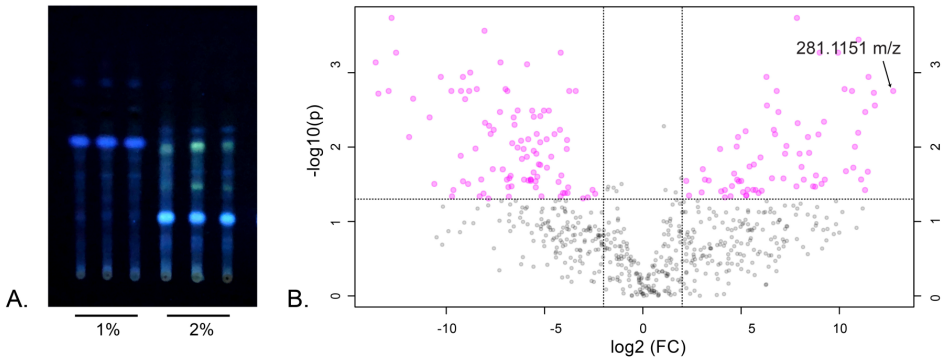


Figure 2. Secondary metabolite profiles found in 1% and 2% glycerol-grown cultures of *Streptomyces* sp. MBT27. **A.** Thin layer chromatography of extracts from the cultures grown in MM with either 1% or 2% glycerol as the carbon source. **B.** Volcano plot highlighting the differences in metabolite profiles. The *x* and *y* axes of the volcano plot represent the log2 fold changes and the corresponding -log10 FDR-adjusted *p*-value of all metabolites, respectively. Pink circles represent metabolites with an intensity difference of more than fourfold (*p*-value ≤ 0.05). Ions present in the left and right quadrants are associated with the 1% and 2% glycerol-grown cultures, respectively. Metabolites situated towards the left and right top quadrants represent values of large magnitude fold changes as well as high statistical significance.

Global Natural Product Social (GNPS) molecular networking (Wang *et al.*, 2016) was employed to detect the MS/MS structural relatedness among molecules in an automated manner; the software generates a molecular network wherein molecules with related scaffolds cluster together (Wang *et al.*, 2016). A network representing the ions detected in the crude extract of *Streptomyces sp.* MBT27 grown with 2% glycerol was constructed, revealing 183 nodes clustered in 19 spectral families (Figure 3). GNPS dereplication based on matching with its MS/MS spectral database highlighted some known metabolites. These included anthranilic acid (**3**), anthranilamide (**4**) (Shaaban *et al.*, 2012b), actinomycin D (**7**), and actinomycin X_2 (**8**) (Vater *et al.*, 2014). The annotation of the compounds was supported by comparison of the exact mass, fragmentation pattern, and UV spectra, with reference data. Moreover, crude extracts of MBT27 possessed antimicrobial activity against *B. subtilis*, which was most likely due to the expression of actinomycins (data not shown). Additional metabolites could be dereplicated through manual comparison of their spectral data against the microbial natural products database Antibase. These previously described metabolites were 4(3*H*)-quinazolinone (**5**) (Kornsakulkarn *et al.*, 2015) and 2,2-dimethyl-1,2-dihydroquinazolin-4(3*H*)-one (**6**) (Conti *et al.*, 2016).

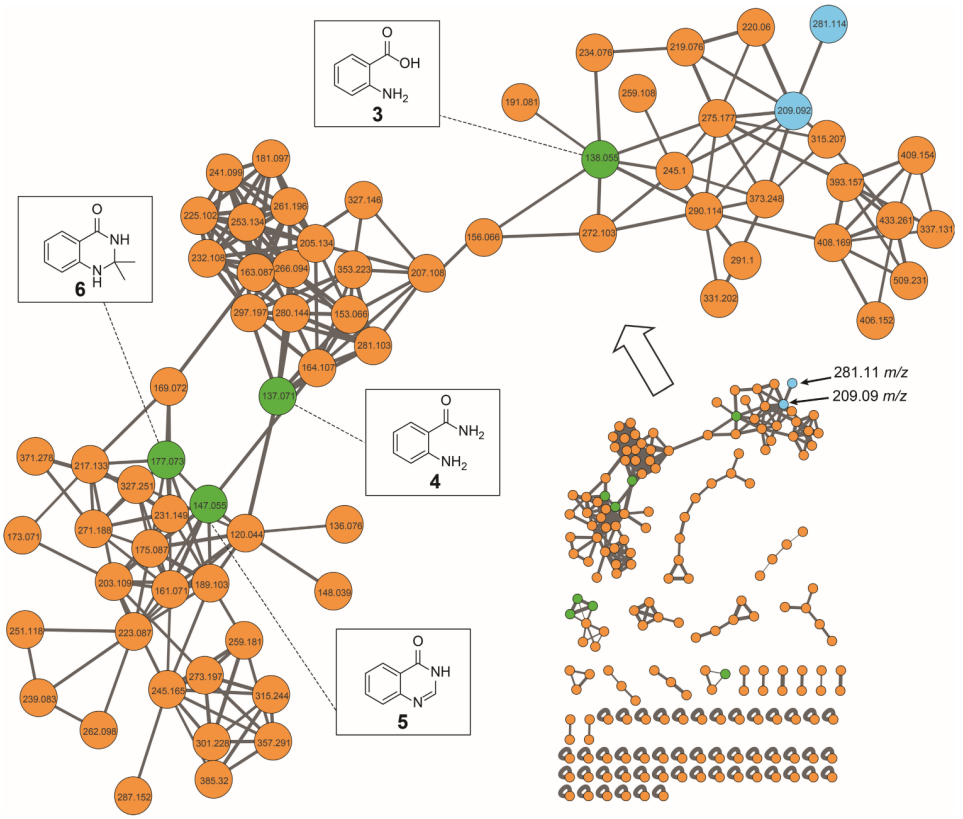


Figure 3. GNPS molecular network of the ions detected in the crude extract of *Streptomyces* sp. MBT27 grown in MM with 2% glycerol. Enlargement is made to the cluster including new metabolites. Orange nodes represent all the ions detected in the extract. Green nodes represent the dereplicated metabolites, while blue nodes represent the ions upregulated when glycerol concentration was increased from 1 to 2%.

Besides known molecules, the network also contained many mass features that could not be assigned to any of the previously identified metabolites. One of these was an ion with an *m/z* value of 281.1151 (**1**), which was very highly increased in intensity (7000-fold, *p*-value = 0.002) when the glycerol concentration was increased from 1% to 2%. It was closely connected (cosine score 0.91) to another upregulated ion with an *m/z* value of 209.0932 (**2**), which was also not previously reported. Both ions were part of the large spectral family comprising the annotated anthranilic acid (**3**), anthranilamide (**4**), and the 4-quinazolinones **5** and **6**, suggesting that they are structurally related metabolites. It is important to note that the concentrations of these compounds did not change significantly when the concentration of glucose was doubled from 1 to 2%, which suggests that the effect of glycerol is not due to growth stimulation (Figure S1).

The relationship between the increase in glycerol concentration and the strong increase in the production of the compounds **1** and **2** was investigated further by expanding the range of glycerol concentrations and analyzing the metabolic profiles. For this, *Streptomyces* sp. MBT27 was cultured in glycerol concentrations ranging from 1 to 4% (w/v). ANOVA, followed by a *post hoc* Tukey's HSD test, was performed to trace the variation in the production of **1** and **2** among the different culturing conditions. A box plot was used to visualize such variation (Figure S2). As observed earlier, the production of **1** and **2** was significantly increased when the glycerol concentration was increased from 1 to 2%. However, further increase in glycerol concentration (3% and 4%) did not lead to any significant increase in the production of **1** and **2**. Accordingly, MM with 2% glycerol was used to culture the bacteria, for the purpose of purification and identification of the new metabolites **1** and **2**.

Isolation and structure elucidation of novel quinazolinones

To elucidate the structure of **1** and **2**, large-scale fermentation was done to obtain larger quantities of the compounds. For this, *Streptomyces* sp. MBT27 was fermented in a total of 4 L of liquid MM supplemented with 2% glycerol, and the supernatant was extracted with EtOAc. Following repeated chromatographic isolation, compounds **1** and **2** were obtained as pure, colorless, amorphous powders. Both compounds were fluorescent, showing UV absorption maxima at 224 and 322 nm and at 226 and 347 nm, respectively. The final structures of **1** and **2** were determined by the combination of NMR and high-resolution MS (Figure 4, Figures S3–S8, S11–S16).

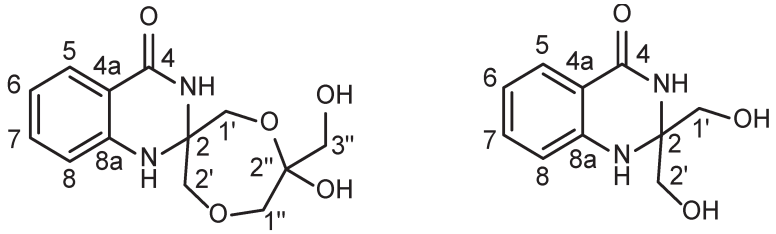


Figure 4. Chemical structures of quinazolinones A (1) and B (2).

The high resolution mass of *m/z* 281.1151 for an $[M+H]^+$ established a molecular formula of $C_{13}H_{16}N_2O_5$, with seven degrees of unsaturation, for **1** (yield 0.25 mg/L). The deduced molecular formula was corroborated by the ^{13}C NMR attached proton test (APT) spectrum that exhibited 13 carbons in total. On the other hand, the ^1H NMR spectrum of **1** (Table 1) presented four coupling aromatic signals at δ_H 7.88 (dd, J = 7.8, 1.2, H-5), 7.07 (t, J = 7.8, H-6), 7.46 (td, J = 7.8, 1.2, H-7), and 7.23 (brd, J = 7.8, H-8), indicating an *o*-disubstituted aromatic benzene system. The HMBC correlation (Figure 5) from H-5 to C-4 (δ_C 166.1)

confirmed one of the substituents on the benzene ring to be an ester/amide carbonyl group, while the downfield chemical shift of C-8a at δ_C 143.7 indicated that it was nitrogenated. Four O-bearing methylene groups at δ_H 4.43 (d, J = 9.0, H-1'a) and 4.24 (d, J = 9.0, H-1'b); 3.59 (d, J = 11.4, H-2'a) and 3.47 (d, J = 11.4, H-2'b); 3.86 (d, J = 12.6, H-1"a) and 3.80 (d, J = 12.6, H-1"b); and 3.66 (d, J = 3.0, H_2-3") were resolved by HSQC experiments. Three of these methylene groups were part of a 1,4-dioxepane ring system, which was established based on the HMBC correlations observed from H-1' to C-2 (δ_C 79.0) and C-2' (δ_C 64.6), from H-2' to C-1" (δ_C 65.7), and from H-1' to the hemiacetal C-2" (δ_C 102.9). The remaining oxymethylene group CH_2-3" was connected to C-2" based on the HMBC correlations observed from H_2-3" to C-1" and C-2". The two substructures obtained accounted for all of the oxygens and six out of the seven degrees of unsaturation required by the molecular formula of **1**. Accordingly, an additional ring including two nitrogen atoms was deduced to connect the two substructures, forming a 4-quinazolinone ring. Consequently, compound **1** was identified to be a *di*-glycerolated 4-quinozolinone, and was named quinazolinone A.

The $[M+H]^+$ molecular ion at *m/z* 209.0932 in the ESI-HRMS spectrum resulted in a molecular formula of $C_{10}H_{12}N_2O_3$ for **2** (yield 0.3 mg/L). The ^{13}C NMR spectrum presented 10 signals, which was consistent with the molecular formula (Table 1). The ^1H NMR spectrum of **2** resembled that of **1** in the downfield aromatic region, and the major difference is the absence of a set of signals for the additional glycerol unit. Further, HMBC experiment confirmed that **2** is a *mono*-glycerolated 4-quinazolinone (Figure 5). Compound **2** was named quinazolinone B.

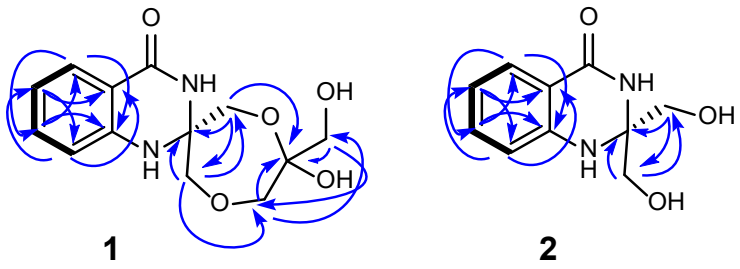


Figure 5. Key COSY (▬) and HMBC (⌒) correlations for 1 and 2.

The 4-quinazoline alkaloids represent an important class of nitrogen-containing heterocyclic compounds (Michael, 2007), which have highly diverse biological activities, such as antitumor, anti-inflammatory, antihypertensive, antimicrobial, anticonvulsant and antifungal activities (He *et al.*, 2017). Though the basic bicyclic core (the fused benzene ring and pyrimidine ring) is normally conserved in most of the naturally occurring 4-quinazoline derivatives (Michael, 2007), a variety of structural substitutions have oftentimes been found on the pyrimidine

ring of the 4-quinazolinone. The chemical structures of quinazolinones A and B and the glycerol-dependent production suggest that glycerol participates in the construction of the ring system. We propose that quinazolinones A and B form a new sub-branch in the family of quinazolines. Particularly, quinazolinone A contains an exciting seven-membered ring that, we believe, may be formed by intermolecular etherification of two units of glycerol, and this ring is further linked with the quinazolinone backbone by a spiro atom at C-2, which is unprecedented. Antimicrobial activities of compounds **1** and **2** were tested against *B. subtilis* and *E. coli*. None of the compounds showed antimicrobial activity against these indictor strains (data not shown).

Biosynthesis of quinazolinones A and B

In our study, the production of **1** and **2** was dramatically enhanced in the cultures with 2% glycerol. Elicitation by glycerol is an indication that glycerol may play a key role during their biosynthesis. The aromatic ring system of quinazoline alkaloids is known to be derived from anthranilic acid, which in turn is biosynthesized through the shikimate pathway (Dewick, 2009). Detection of anthranilamide in the extract of *Streptomyces* sp. MBT27 leads to the assumption that the biosynthesis of the isolated quinazolinones starts from anthranilic acid, which is then converted into anthranilamide (Figure 6). Further successive attachments of two molecules of glycerol then result in the formation of **2**, followed by **1**.

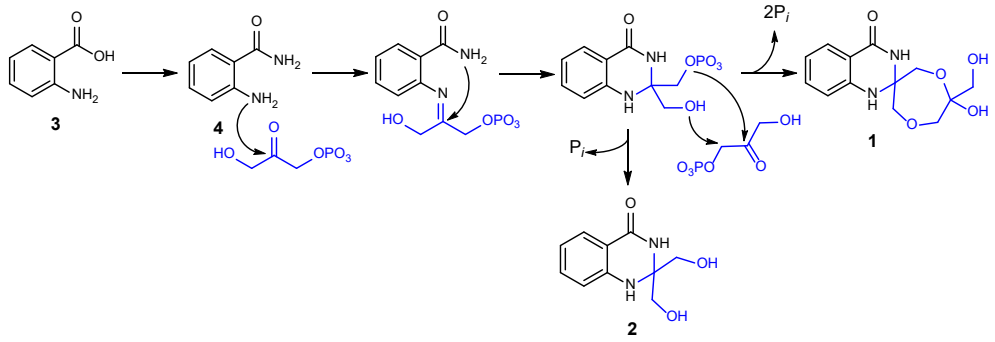


Figure 6. Proposed biosynthetic pathway for quinazolinone A (1) and B (2).

However, multiple biosynthetic routes have been proposed for quinazolinones, because the C-2 residue of the quinazoline ring may originate from various precursors. For example, the C-2 and the remaining non-aromatic part of the quinazolinone alkaloid chrysogine, produced by *Penicillium chrysogenum*, are biosynthesized from pyruvic acid via an nonribosomal peptide synthetase (NRPS) assembly line (Wollenberg *et al.*, 2017), while the non-aromatic part of the quinazoline alkaloid peganine, produced by the plant *Peganum harmala*, is derived from ornithine (Dewick, 2009). Furthermore, fumiquinazoline F originates from a

fungal nonribosomal peptide synthetase (TqaA) and is biosynthesized from anthranilic acid, L-tryptophan and L-alanine (Gao *et al.*, 2011).

In conclusion, the metabolic potential of *Streptomyces* sp. MBT27 was analyzed under different nutritional conditions, showing major changes in the metabolome, depending on the carbon source used. In particular, the relatively unspectacular change from 1 to 2% glycerol resulted in a surprisingly global change in the secondary metabolome, with some compounds changing by almost four orders of magnitude. The use of GNPS molecular networking, together with LC-MS-based dereplication, allowed us to annotate a cluster of quinazolinone-family molecules that connect to anthranilic acid and anthranilamide. Isolation and structure elucidation revealed the novel alkaloid natural products quinazolinone A and B. Based on the structures, the biosynthesis of the compounds most likely involves the conversion of anthranilic acid into anthranilamide, which is subsequently attached to two glycerol units to produce compounds **1** and **2**. Identification of the biosynthetic gene cluster and subsequent analysis of the individual enzymatic reactions should reveal the precise biosynthetic pathway for these exciting novel molecules.

Materials and Methods

Bacterial strain, growth conditions

Streptomyces sp. MBT27 was obtained from the Leiden University strain collection and had previously been isolated from the Qingling Mountains, Shanxi province, China (Zhu *et al.*, 2014b). Cultures were grown in triplicate in 100 mL Erlenmeyer flasks with 30 mL of liquid minimal medium (MM) (Kieser, 2000), supplemented with various carbon sources, and inoculated with 10 µL of 10^9/mL spore suspension. The carbon sources (percentages in w/v) were: 1% mannitol + 1% glycerol, 1% mannitol, 2% mannitol, 1% glycerol, 2% glycerol, 1% glucose, 2% glucose, 1% fructose, 1% arabinose or 1% *N*-acetylglucosamine (GlcNAc). The cultures were incubated in a rotary shaker at 30 °C at 220 rpm for seven days.

General experimental procedures

NMR spectra were recorded in deuterated methanol (CD_3OD) on a Bruker 600 MHz, and were referenced using the residual ^1H signal of deuterated solvent at 3.30 ppm. FT-IR was measured on Perkin-Elmer FT-IR Spectrometer Paragon 1000. UV measurements were performed using a Shimadzu UV mini-1240. Optical rotations were measured on a JASCO P-1010 polarimeter. HPLC purification was performed on Waters preparative HPLC system comprised of 1525 pump, 2707 autosampler, 2998 PDA detector, and Water fraction collector III. The columns used were SunFire C_{18} column (5 µm, 100 Å, 10 × 250 mm) and SunFire C_{18} column (10 µm, 100 Å, 19 × 150 mm). TLC was performed using aluminum plates coated with silica gel 60 F_{254} (Merck). All organic solvents and chemicals were of analytical or LCMS grade, depending on the experiment.

Metabolite profiling

Following fermentation, culture supernatants were extracted with ethyl acetate (EtOAc) and evaporated under reduced pressure. For LC-ESI-QTOFMS analyses, extracts were dissolved in MeOH to a final concentration of 1 mg/mL, and 1 µL was injected into Waters Acquity UPLC system equipped with Waters Acquity HSS C_{18} column (1.8 µm, 100 Å, 2.1 × 100 mm), which is coupled to Agilent 6530 QTOF MS equipped with Agilent Jet Stream ESI source (Agilent Technologies, Inc., Palo Alto, CA, USA). For the LC, solvent A was 95% H_2O, 5% acetonitrile (ACN) and 0.1% formic acid; solvent B was 100% ACN and 0.1% formic acid. The gradient used was 2% B for 1 min, 2–85% for 9 min, 85–100% for 1 min, and 100% for 3 min. The flow rate used was 0.5 mL/min. As for the MS, the following ESI source parameters were used: capillary voltage 3 kV, source temperature 325 °C, drying gas flow rate 10 L/min, and fragmentor 175 V. Full MS spectra were acquired in positive mode in the range of 100–1700 *m/z*, in the extended dynamic range mode. Internal reference masses of purine and Agilent HP-921 were continuously delivered to the ESI source through an Agilent 1260

isocratic pump. Thermo Instruments MS system (LTQ Orbitrap XL, Bremen, Germany) equipped with an electrospray ionization source (ESI) was used for LC-MS/MS analysis. The Waters Acquity UPLC system equipped with Waters Acquity PDA was run using a SunFire Waters C_{18} column (3.5 µm, 100 Å, 4.6 × 150 mm), at a flow rate of 0.9 mL/min. Solvent A was 95% H_2O, 5% acetonitrile (ACN) and 0.1% formic acid; solvent B was 100% ACN and 0.1% formic acid. The gradient used was 2% B for 1 min, 2–85% for 15 min, 85–100% for 3 min, and 100% for 3 min. As for the MS, the following ESI parameters were used: capillary voltage 5 V, spray voltage 3.5 kV, capillary temperature 300 °C, auxiliary gas flow rate 10 arbitrary units, and sheath gas flow rate 50 arbitrary units. Full MS spectra were acquired in the Orbitrap in positive mode at a mass range of 100–2000 *m/z*, and FT resolution of 30,000. Data-dependent MS^2 spectra were acquired in the ion trap for the three most intense ions using collision-induced dissociation (CID). The resulting chemical data were compared with those in SciFinder and Antibase (Laatsch, 2012a).

Large-scale fermentation and isolation of metabolites (1) and (2)

For large-scale fermentation, *Streptomyces* sp. MBT27 was grown in eight 2 L Erlenmeyer flasks, each containing 500 mL liquid MM supplemented with 2% w/v glycerol at 30 °C for seven days. The metabolites were extracted from the spent media of culture filtrates using EtOAc, and the solvent was subsequently evaporated under reduced pressure at 40 °C. The crude extract (1.4 g) was adsorbed onto 1.4 g silica gel (pore size 60 Å, 70–230 mesh, Sigma Aldrich), and loaded on silica column, which was eluted using gradient mixtures of *n*-hexane, acetone, and MeOH. The fractions eluted with acetone 100% were combined, reconstituted in MeOH, and injected into the preparative SunFire column (19 x 150 mm), which was eluted with H_2O:MeOH gradient of 7–100% in 30 min, at a flow rate of 15 mL/min, to yield five fractions. Quinazolinones A and B were further purified on the semi-preparative SunFire column (10 x 250 mm), run at 3 mL/min. Fraction 2 was eluted using H_2O:MeOH gradient of 20–40% in 20 min, to yield quinazolinone A (**1**, 1 mg). On the other hand, Fraction 3 was eluted using H_2O:MeOH gradient of 30–40% in 20 min, to yield quinazolinone B (**2**, 1.5 mg).

Quinazolinone A (1): colorless, amorphous powder; $[\alpha]_D^{20}$ 2.3 (c 0.1, MeOH); UV (MeOH) λ_{max} (log ε) 224 (4.09), 322 (2.98) nm; IR ν_{max} 3334, 2922, 1652, 1475, 1052 cm^{-1}; ^1H and ^{13}C NMR data, see Table 1; HRESIMS (positive mode) *m/z* 281.1130 $[M+H]^+$ (calcd. for $C_{13}H_{17}N_2O_5$, 281.1132).

Quinazolinone B (2): colorless, amorphous powder, UV (MeOH) λ_{max} (log ε) 226 (4.18), 347 (3.03) nm; IR ν_{max} 3350, 2950, 1649, 1526, 1049, 751 cm^{-1}; ^1H and ^{13}C NMR data, see Table 1; HRESIMS (positive mode) *m/z* 209.0915 $[M+H]^+$ (calcd. for $C_{10}H_{13}N_2O_3$, 209.0921).

Table 1. ^{1}H and ^{13}C NMR data for compounds 1 and 2[a].

NO.	1		2	
	δ_C	δ_H (mult., *J* in Hz)	δ_C	δ_H (mult., *J* in Hz)
2	79.0		73.3	
4	166.1		166.7	
4a	121.8		114.5	
5	128.8	7.88 (dd, *J* = 7.8, 1.2)	128.5	7.66 (dd, *J* = 7.8, 1.2)
6	123.0	7.07 (t, *J* = 7.8)	118.5	6.69 (td, *J* = 7.8, 1.2)
7	134.6	7.46 (td, *J* = 7.8, 1.2)	135.3	7.26 (td, *J* = 7.8, 1.2)
8	122.3	7.23 (brd, *J* = 7.8)	115.8	6.73 (brd, *J* = 7.8)
8a	143.7		148.6	
1'	73.9	4.43 (d, *J* = 9.0); 4.24 (d, *J* = 9.0)	64.5	3.67 (d, *J* = 10.8); 3.62 (d, *J* = 10.8)
2'	64.6	3.59 (d, *J* = 11.4); 3.47 (d, *J* = 11.4)	64.5	3.67 (d, *J* = 10.8); 3.62 (d, *J* = 10.8)
1"	65.7	3.86 (d, *J* = 12.6); 3.80 (d, *J* = 12.6)		
2"	102.9			
3"	61.9	3.66 (d, *J* = 3.0)		

[a] **1** and **2** were recorded in CD_3OD, at 298 K. All chemical shift assignments were done on the basis of 1D and 2D NMR techniques.

Computation of mass spectral networks

MS/MS raw data were converted into a 32-bit mzXML file using MSConvert (ProteoWizard) (Chambers *et al.*, 2012) and spectral networks were assembled using Global Natural Product Social molecular networking (GNPS) (https://gnps.ucsd.edu) as described (Wang *et al.*, 2016). For both parent and MS/MS fragment ions, the mass tolerance was set to 0.5 Da, while the minimum cosine score was set to 0.7. The data were clustered using MSCluster with a minimum cluster size of three spectra. The spectra in the network were also searched against GNPS spectral libraries. A minimum score of 0.5 was set for spectral library search, with at least two fragment peaks matching. Cytoscape 3.5.1. was used for visualization of the generated molecular network (Shannon *et al.*, 2003). The edge thickness was set to represent the cosine score, with thicker lines indicating higher similarity between nodes. LC-MS/MS data were deposited in the MassIVE Public GNPS data set (MSV000082988). The molecular networking job in GNPS can be found at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=8fd1fcfa0ff744a9808e80bc7be12115. The annotated MS/MS spectra were deposited in the GNPS spectral library for quinazolinone A (CCMSLIB00004684355), and B (CCMSLIB00004684354).

Statistical analysis

Prior to statistical analysis, mzXML files were imported into Mzmine 2.31 (Pluskal *et al.*, 2010) for data processing. Mass ion peaks were detected using the exact mass algorithm with a noise level set to 1.0×10^4. Afterward, chromatograms were built for the detected masses with a minimum time span of 0.05 min, *m*/*z* tolerance of 0.001 *m*/*z* and minimum height of 1.0×10^4. Chromatogram deconvolution was then performed using local minimum search algorithm (search minimum in RT range 0.1 min, chromatographic threshold 90%, minimum relative height 1%, minimum absolute height 1.0×10^4, minimum ratio of peak top/edge 2 and peak duration range 0.05–3 min). In the generated peak lists, isotopes were identified using isotopic peaks grouper (*m*/*z* tolerance 0.001 *m*/*z* and retention time tolerance 0.1 min), and variations in retention time were reduced using retention time normalizer (*m*/*z* tolerance 0.001 *m*/*z* and retention time tolerance 1 min). All the peak lists were subsequently aligned using join aligner (*m*/*z* tolerance 0.001 *m*/*z*, *m*/*z* weight 20, retention time tolerance 0.1 min, and retention time weight 20), and missing peaks were detected through gap filling using peak finder (intensity tolerance 1.0%, *m*/*z* tolerance 0.001 *m*/*z*, and retention time tolerance of 0.2 min). Finally, the aligned peak list was exported as a comma-separated file for statistical analysis.

Statistical analysis was performed using MetaboAnalyst (Chong *et al.*, 2018), where log transformation and pareto scaling was initially applied to the data. A heat map of all detected masses, among the different growth conditions, was generated in MetaboAnalyst, to which additional hierarchical clustering analysis (HCA) was performed using Euclidean distance measure and Ward clustering algorithm. Student's *t*-tests with multiple testing corrections (Benjamini-Hochberg-false discovery rate or FDR) were used to determine significant differences in the intensities of the metabolites, under two different growth conditions. The thresholds set for statistically significant differences were a fold change ≥ 4, together with FDR corrected *p*-value ≤ 0.05. Based on these criteria, a volcano plot was generated. To identify the difference in intensity of a single mass feature among multiple growth conditions, one-way ANOVA was performed, followed by a *post hoc* Tukey's honest significant difference (HSD) test.

Supplementary information

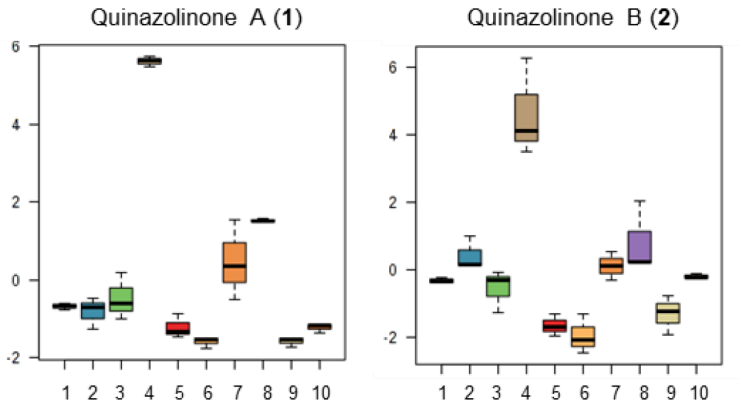


Figure S1. Effect of carbon source utilization on secondary metabolite profiles of *Streptomyces* sp. MBT27. Box plots show the relative intensities of quinazolinones A (**1**) and B (**2**) in cultures grown in MM with different carbon sources, namely: 1. mannitol 1% + glycerol 1%; 2. GlcNAc 1%; 3. glycerol 1%; 4. glycerol 2%; 5. mannitol 1%; 6. mannitol 2%; 7. glucose 1%; 8. glucose 2%; 9. arabinose 1%; 10. fructose 1%. Note the spectacular increase in production when the glycerol concentration is doubled.

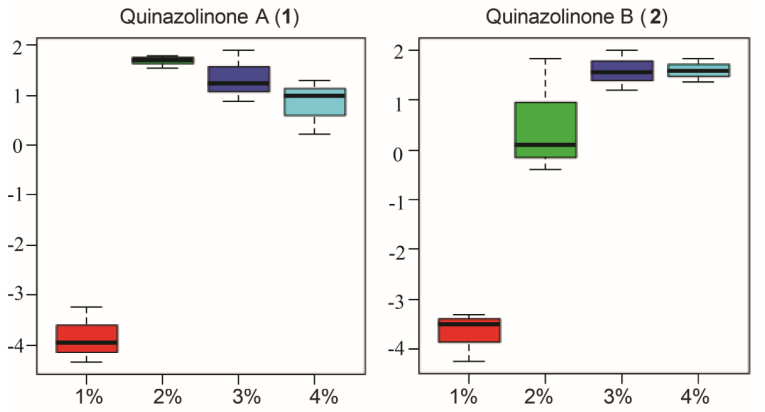


Figure S2. Changes in the secondary metabolite profiles of *Streptomyces* sp. MBT27 depending on the glycerol concentration. Changes in the secondary metabolite profiles of *Streptomyces* sp. MBT27 depending on the glycerol concentration. Box plots show the relative intensities of quinazolinones A (**1**) and B (**2**) in the cultures with added 1%, 2%, 3% or 4% (w/v) glycerol.

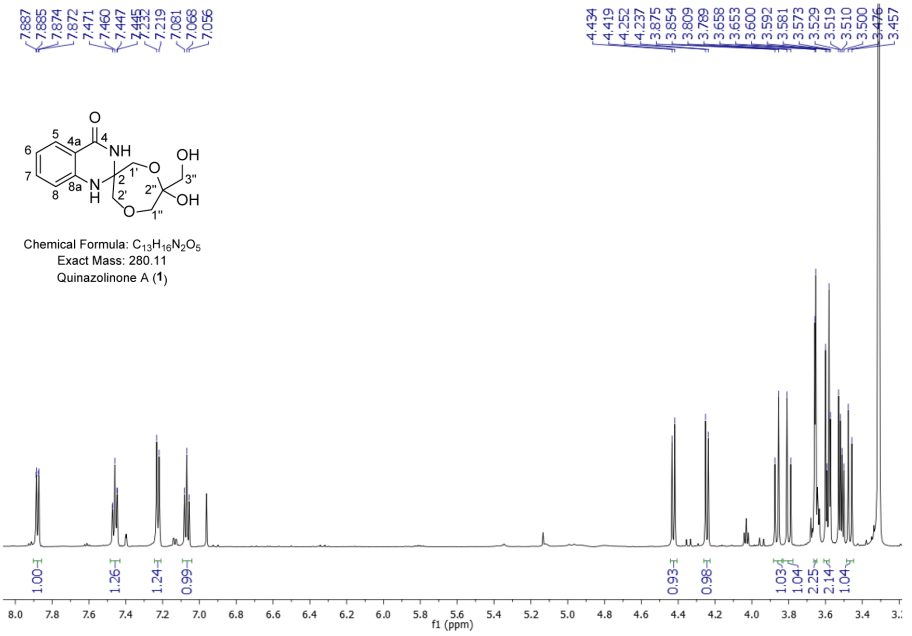


Figure S3. ^{1}H NMR spectrum of quinazolinone A (1) in CD_3OD.

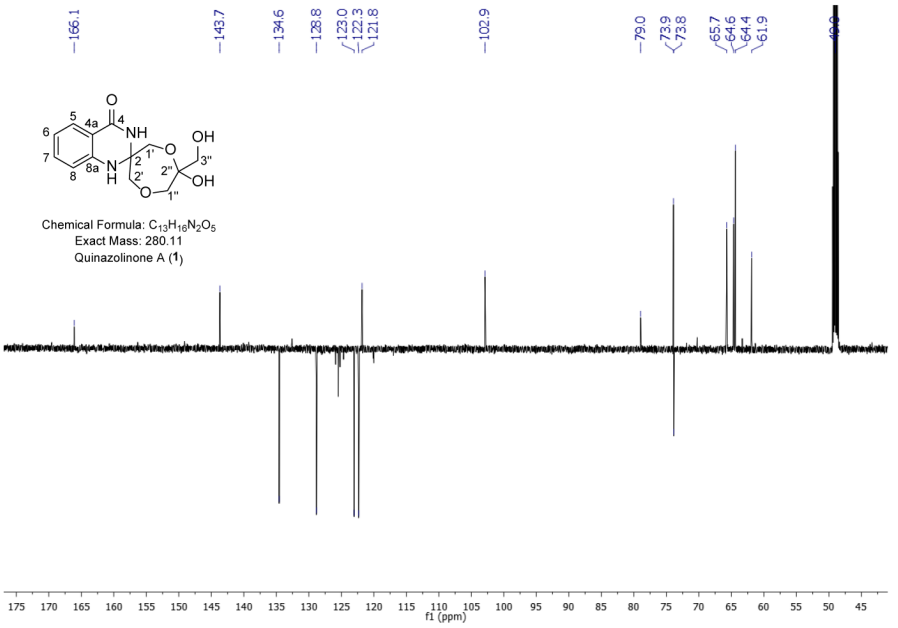


Figure S4. APT spectrum of quinazolinone A (1) in CD_3OD.

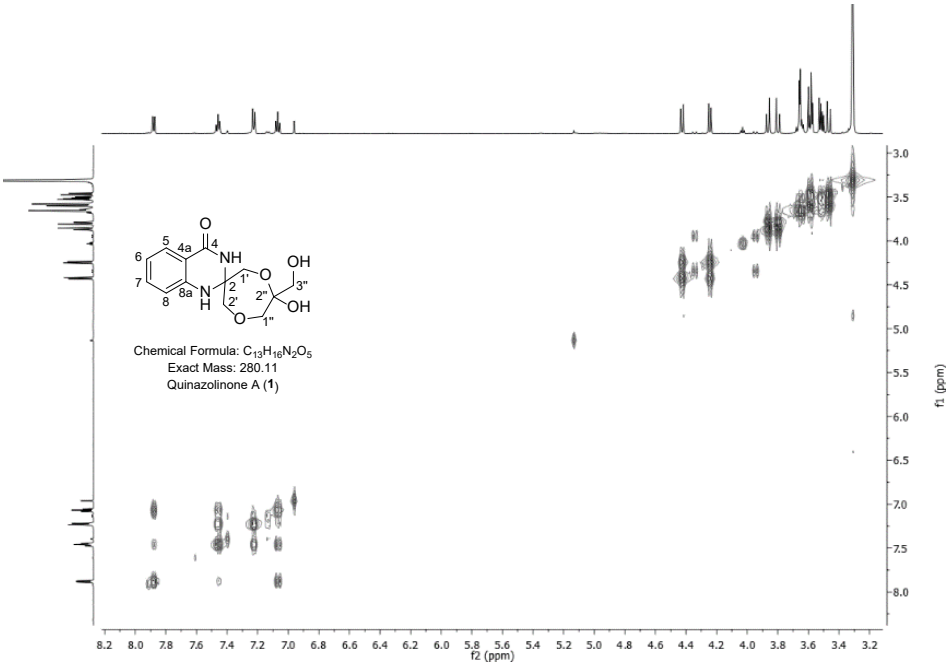


Figure S5. ^{1}H-^{1}H COSY spectrum of quinazolinone A (1) in CD_3OD.

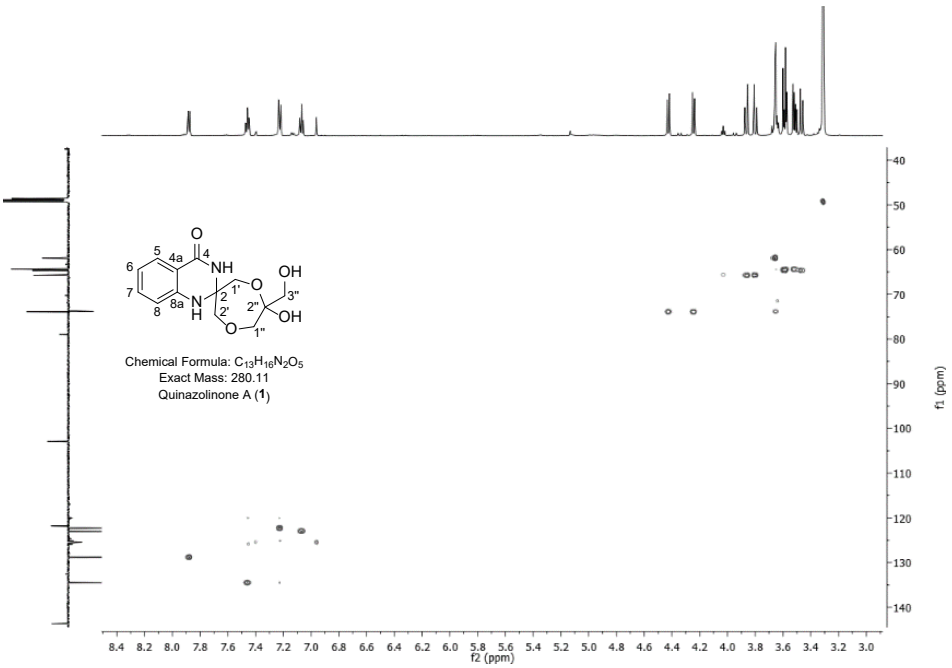


Figure S6. HSQC spectrum of quinazolinone A (1) in CD_3OD.

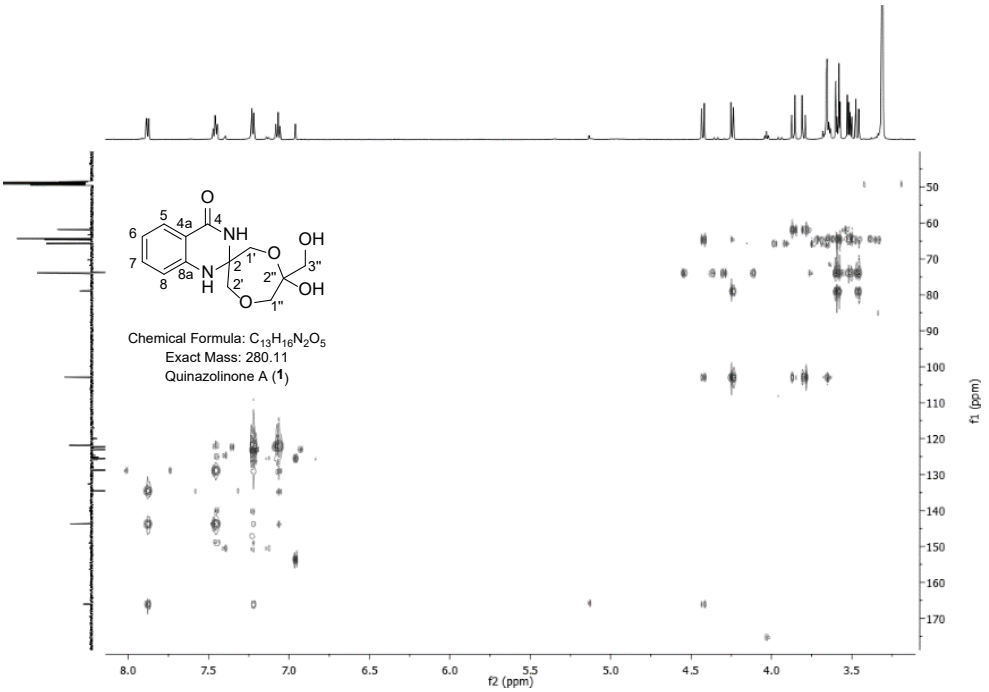


Figure S7. HMBC spectrum of quinazolinone A (1) in CD_3OD.

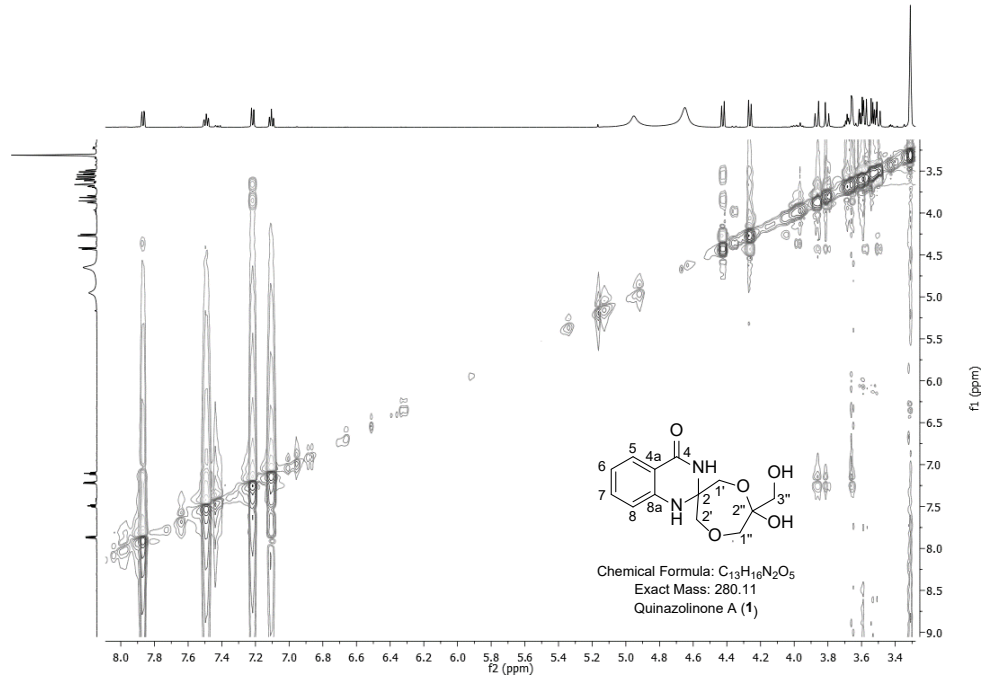


Figure S8. ^1H-^1H NOSEY spectrum of quinazolinone A (1) in CD_3OD.

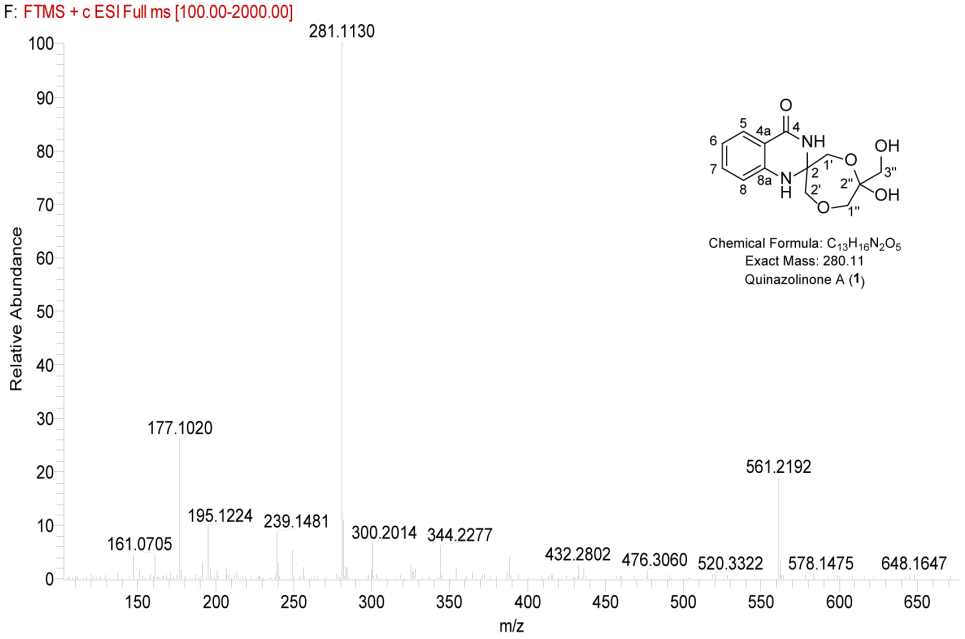


Figure S9. HRMS spectrum of quinazolinone A (1).

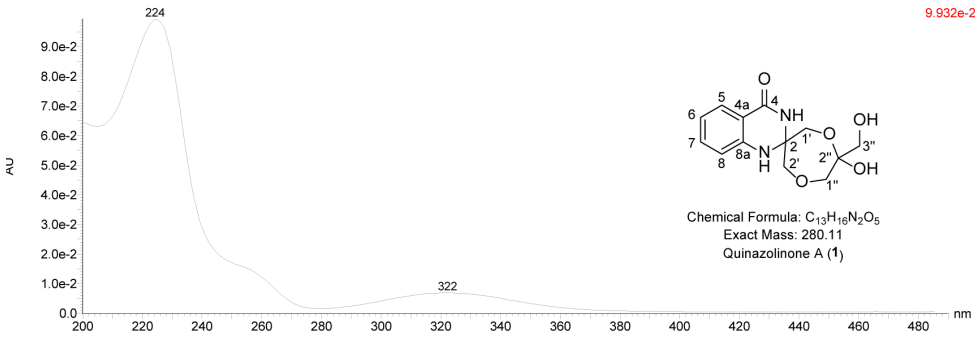


Figure S10. UV spectrum of quinazolinone A (1).

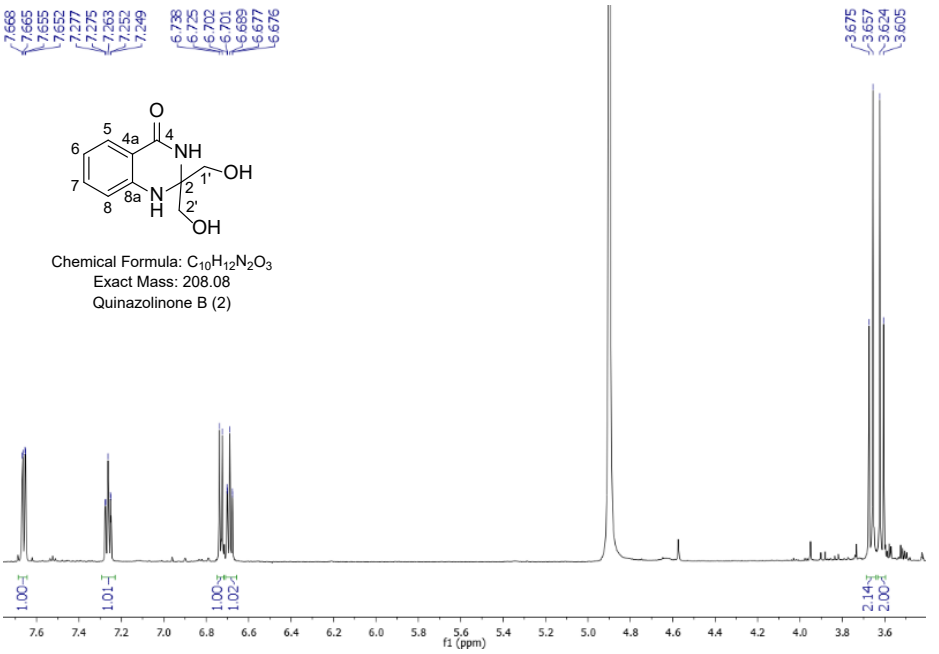


Figure S11. ^{1}H NMR spectrum of quinazolinone B (2) in CD_3OD.

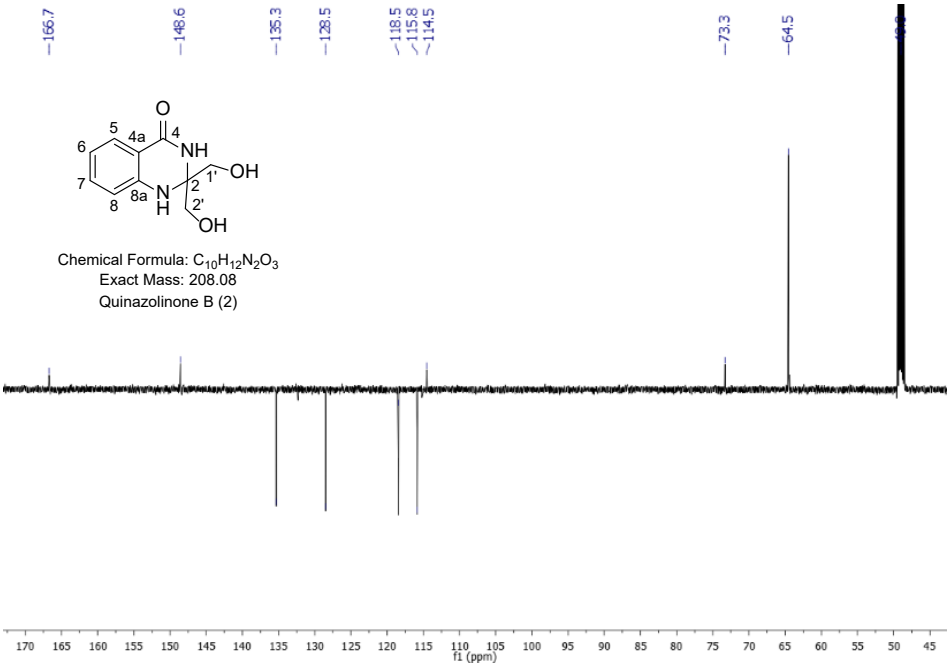


Figure S12. APT spectrum of quinazolinone B (2) in CD_3OD.

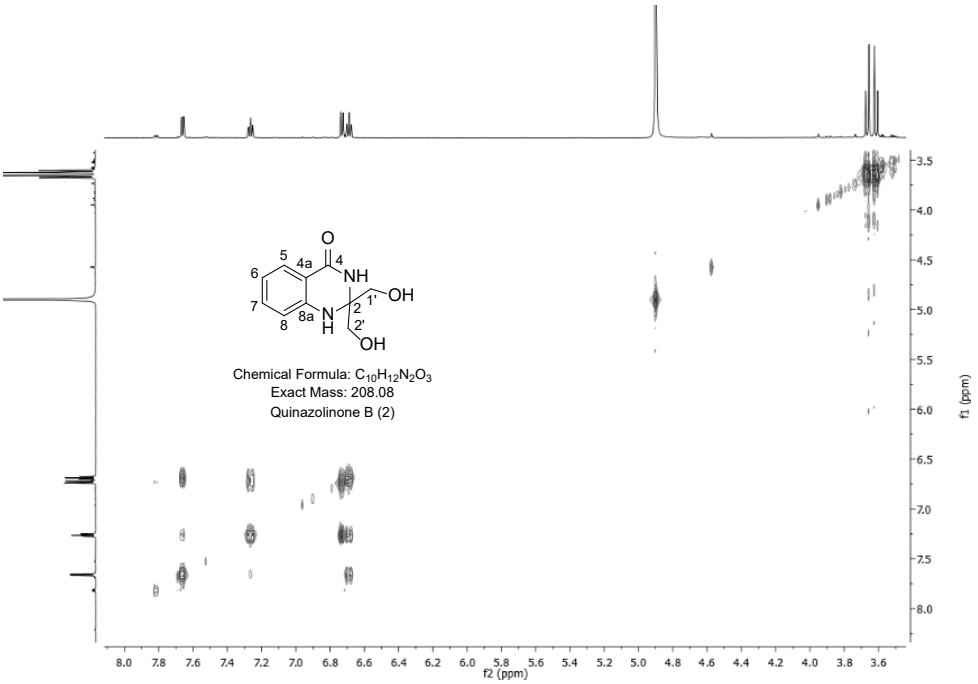


Figure S13. ^{1}H-^{1}H COSY spectrum of quinazolinone B (2) in CD_3OD.

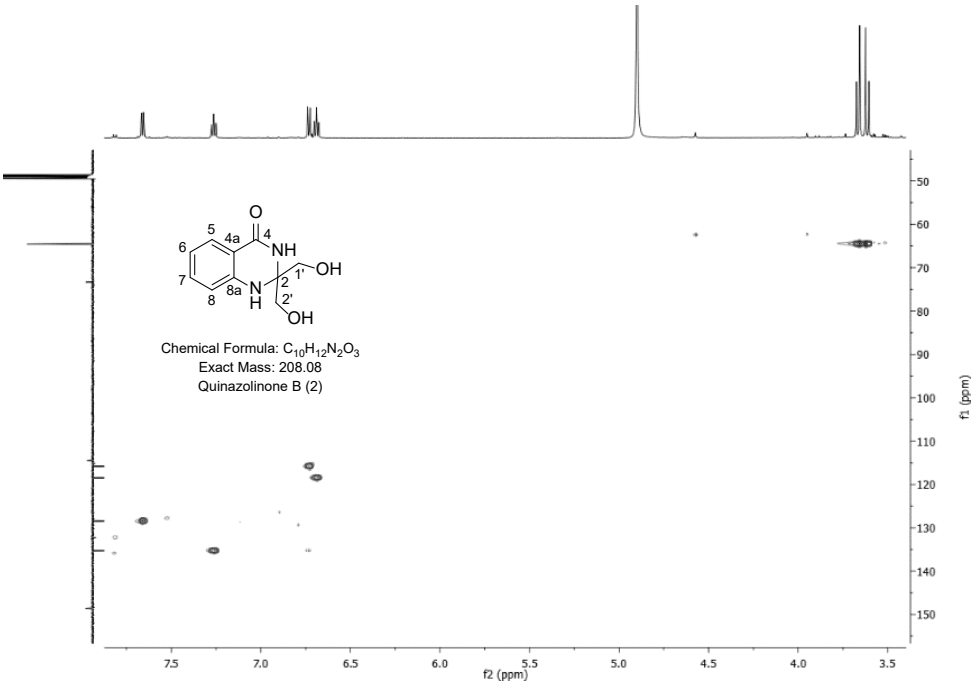


Figure S14. HSQC spectrum of quinazolinone B (2) in CD_3OD.

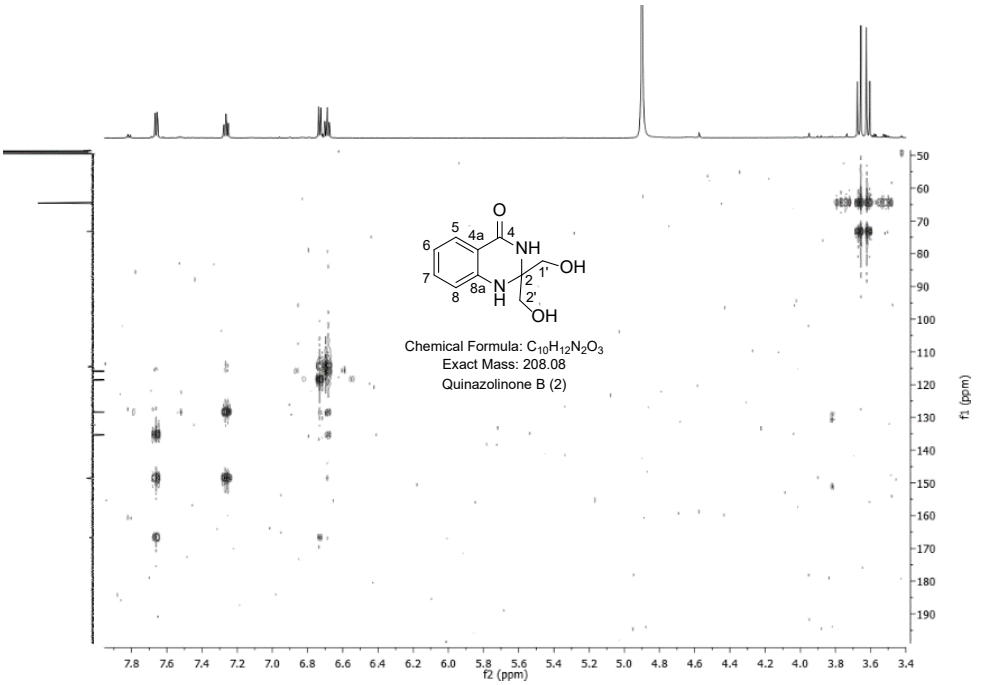


Figure S15. HMBC spectrum of quinazolinone B (2) in CD_3OD.

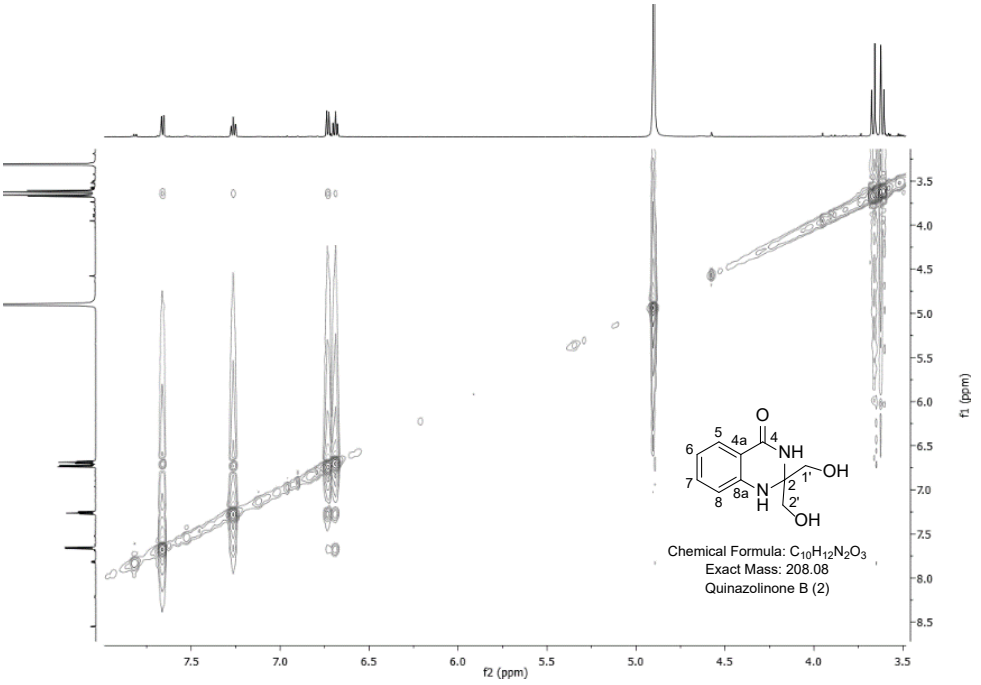


Figure S16. ^1H-^1H NOSEY spectrum of quinazolinone B (2) in CD_3OD.

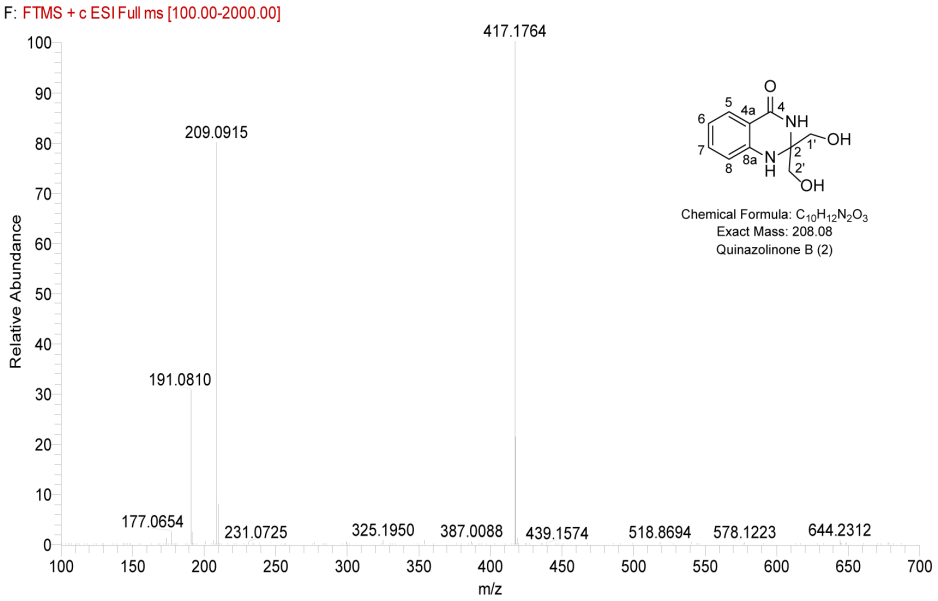


Figure S17. HRMS spectrum of quinazolinone B (2).

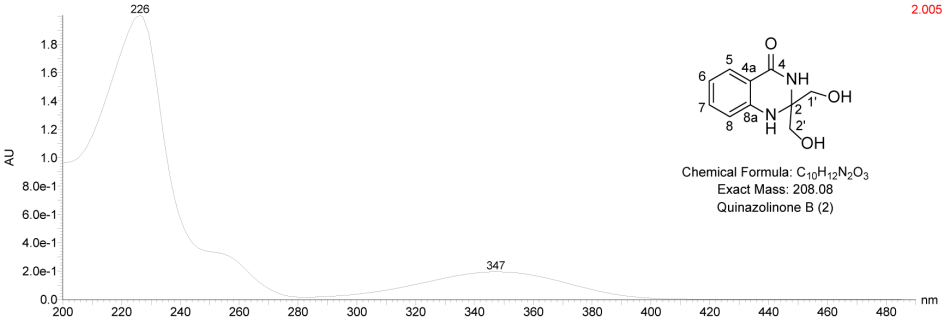


Figure S18. UV spectrum of quinazolinone B (2).



Chapter 5

Discovery of actinomycin L, a new member of the actinomycin family of antibiotics

Nataliia V. Machushynets, Somayah S. Elsayed, Chao Du,
Maxime A. Siegler, Mercedes de la Cruz, Olga Genilloud,
Thomas Hankemeier and Gilles P. van Wezel

This chapter was published as:
Machushynets N.V., Elsayed S.S., Du C., Siegler M.A., de la Cruz M., Genilloud O., Hankemeier T., & van Wezel G.P. (2022). Discovery of actinomycin L, a new member of the actinomycin family of antibiotics. *Sci Rep*., **12**, 2813.

Abstract

Streptomycetes are major producers of bioactive natural products, including the majority of the naturally produced antibiotics. While much of the low-hanging fruit has been discovered, it is predicted that less than 5% of the chemical space of natural products has been mined. Here, we describe the discovery of the novel actinomycins L_1 and L_2 produced by *Streptomyces* sp. MBT27, via application of metabolic analysis and molecular networking. Actinomycins L_1 and L_2 are diastereomers, and the structure of actinomycin L_2 was resolved using NMR and single-crystal X-ray crystallography. Actinomycin L is formed via spirolinkage of anthranilamide to the 4-oxoproline moiety of actinomycin X_2, prior to the condensation of the actinomycin halves. Such a structural feature has not previously been identified in naturally occurring actinomycins. Adding anthranilamide to cultures of the actinomycin X_2 producer *Streptomyces antibioticus*, which has the same biosynthetic gene cluster as *Streptomyces* sp. MBT27, resulted in the production of actinomycin L. This supports a biosynthetic pathway whereby actinomycin L is produced from two distinct metabolic routes, namely those for actinomycin X_2 and for anthranilamide. Actinomycins L_1 and L_2 showed significant antimicrobial activity against Gram-positive bacteria. Our work shows how new molecules can still be identified even in the oldest of natural product families.

Introduction

Considering the emerging crisis of antibiotic resistance that spreads among bacterial pathogens and increasing incidence of cancer, the search for new, efficient and less toxic drugs remains a priority (Wright, 2015, Silver, 2011). Actinobacteria have been the source for the majority of the antibiotics in use today (Barka *et al.*, 2016, Bérdy, 2005). Of the Actinobacteria, members of the genus *Streptomyces* produce over half of all currently characterized antibiotics (Barka *et al.*, 2016). Genome sequencing revealed that Actinobacteria have much more biosynthetic potential to produce bioactive molecules than originally anticipated, with even the model organisms harboring many so-called cryptic or silent biosynthetic gene clusters (BGCs) that specify yet unknown compounds (Bentley *et al.*, 2002, Cruz-Morales *et al.*, 2013, Ikeda *et al.*, 2003). Triggering the expression of silent BGCs by genetic and cultivation-based techniques should facilitate unlocking this yet unexplored chemical diversity, allowing the discovery of novel molecules (Rutledge & Challis, 2015, van Bergeijk *et al.*, 2020). This strategy relies on altering the regulatory networks of the producing organism in response to fluctuating culturing conditions, such as carbon, nitrogen or phosphate concentration (Urem *et al.*, 2016, van der Heul *et al.*, 2018, Sanchez *et al.*, 2010). Manipulation of fermentation conditions of promising producer strains, known as "one strain many compounds" (OSMAC) approach, is effective in enhancing secondary metabolites production (Bode *et al.*, 2002, Romano *et al.*, 2018). Novel secondary metabolites have been discovered via modification of cultivation parameters, including nutrients (Yoon & Nodwell, 2014, Zhu *et al.*, 2014a), and the addition of chemical elicitors (Craney *et al.*, 2012, Rigali *et al.*, 2008).

Metabolic profiling of crude extracts obtained under different growth conditions represents a challenging analytical task since these mixtures are composed of hundreds of natural products. Therefore, metabolomics, particularly those based on mass spectrometry (MS), became more and more valuable and greatly increased the efficiency of such screenings (Demarque *et al.*, 2020). Supervised statistical methods are able to classify a response like a biological activity, and to determine the most discriminant metabolite(s) related to such response (O'Leary *et al.*, 2016). Moreover, simultaneous dereplication of differentially expressed compounds is implemented into the drug-discovery pipelines in order to avoid rediscovery of already known compounds (Gaudencio & Pereira, 2015). MS-based metabolomics provides important information on the distribution of the metabolites that are present in complex mixtures, but the identification of their structures is complicated. For this purpose, the Global Natural Products Social molecular networking (GNPS MN) was developed, applying both molecular networking and automated searches of tandem mass spectrometry (MS/MS) fragmentation spectra against spectral libraries, to identify structural relationships between metabolites (Nguyen *et al.*, 2013, Wang *et al.*, 2016). This greatly

facilitates the annotation and dereplication of known molecules.

Actinomycin is a DNA-targeting antibiotic and anticancer compound discovered in 1940 by Waksman and Woodruff and in fact, the first antibiotic that was isolated from an Actinobacteria (Waksman & Woodruff, 1940). Actinomycins are produced by various *Streptomyces* strains and are composed of a chromophore group and two pentapeptide chains with a variable composition of amino acids (Figure S1) (Katz, 1967). Actinomycins D, $X_{0\beta}$ and X_2 are usually simultaneously produced and differ from each other by substitutions on the proline residue in their pentapeptide lactone rings, while members of the actinomycin C complex vary in their D-valine residues (Crnovcic *et al.*, 2017). The pentapeptide precursors are biosynthesized by a nonribosomal peptide synthetase (NRPS) assembly line, and actinomycins are formed through oxidative condensation of two 3-hydroxy-4-methylanthranilic acid (4-MHA) pentapeptide lactones (PPLs) (Crnovčić *et al.*, 2014).

In this work, we report the discovery of new actinomycin analogs, actinomycin L_1 and L_2, from the extracts of *Streptomyces* sp. MBT27. Multivariate data analysis combined with molecular networking indicated that the antimicrobial activity of the extracts correlated with novel actinomycins L_1 and L_2 and known actinomycins D, $X_{0\beta}$ and X_2. NMR and single-crystal X-ray crystallography revealed that an anthranilamide moiety was linked through a spiro-center to a proline residue in the structure of actinomycins L_1 and L_2. Such a structural feature has not previously been identified in naturally occurring actinomycins.

Results

The influence of carbon sources on bioactivity and actinomycin production

Streptomyces sp. MBT27 is a gifted natural product producer that was isolated from Qinling mountains in China, with potent antibacterial activity against various MDR (multi-drug resistant) bacteria (Zhu *et al.*, 2014a). We previously showed that the strain, among others produces the novel quinazolinones A and B (Machushynets *et al.*, 2019). To investigate the antibiotic activity of *Streptomyces* sp. MBT27 the strain was fermented in minimal medium (MM) with either of the following carbon sources (percentages in w/v): 1% of both mannitol and glycerol, 1% mannitol, 2% mannitol, 1% glycerol, 2% glycerol, 1% glucose, 2% glucose, 1% fructose, 1% arabinose, or 1% *N*-acetylglucosamine (GlcNAc). Supernatants of *Streptomyces* sp. MBT27 cultures were extracted with ethyl acetate and bioactivity assays were performed against *Bacillus subtilis* 168. Interestingly, the carbon sources had a huge effect on the antimicrobial activity (Figure S2). Particularly strong antimicrobial activity was observed when the culture medium was supplemented with glycerol + mannitol, glucose 1%, glycerol, fructose or GlcNAc; as compared to when mannitol or arabinose were used as the carbon sources.

In order to investigate the metabolic differences due to nutritional supplementation and correlate that to the antimicrobial activity, LC-MS-based metabolomics was performed. Initially, the LC-MS data were explored by unsupervised Principal Component Analysis (PCA). The first two PCs accounted for 37% and 16%, respectively, of the total data variation. PCA analysis failed to show significant metabolic separation in relation to the observed bioactivity (Figure 1A). The supervised Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) was then applied to discriminate the samples based on their ability to inhibit *B. subtilis* (Figure 1B). The cross-validation metrics of the model (R^2Y = 0.748 and Q^2Y = 0.676) indicated that the model has good reliability and ability of prediction. A permutation test was performed (n = 100) and the resulting R^2Y and Q^2Y values were significantly lower (p values < 0.01 for both), which indicated that there was no overfitting in the model (Figure S3) (Chong *et al.*, 2019). The OPLS-DA loadings S-plot revealed the most discriminative features between active and inactive groups (Figure 1C).

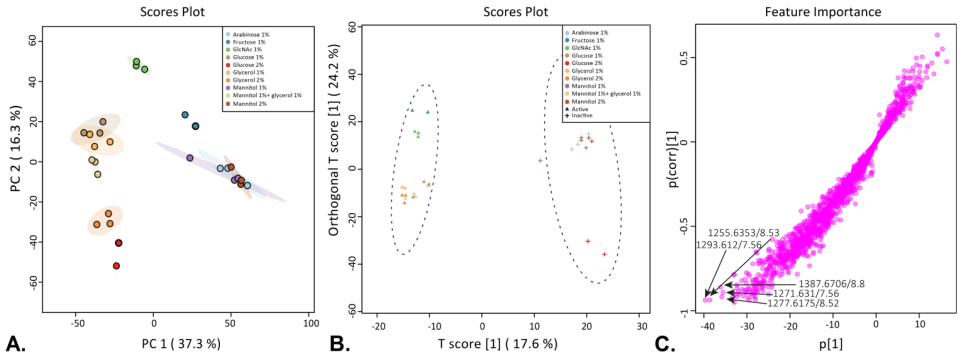


Figure 1. Differential production of metabolites depending on the carbon source.

A. PCA score plot of *Streptomyces* sp. MBT27 metabolites produced in cultures with different carbon sources, namely, 1% arabinose, 1% fructose, 1% GlcNAc, 1% glucose, 2% glucose, 1% glycerol, 2% glycerol, 1% mannitol, 1% mannitol + 1% glycerol and 2% mannitol (%ages in w/v). **B.** OPLS-DA score plot. Triangles and crosses represent samples of active and inactive groups respectively, circular areas represent the 95% confidence region of each group. **C.** OPLS-DA loadings S-plot. Arrows indicate the most discriminative features that positively correlate with the bioactive group.

The mass features that correlated best to the bioactivity were *m/z* 1387.6706 (8.8 min), *m/z* 1255.6353 (8.53 min), *m/z* 1277.6175 (8.52 min), *m/z* 1271.631 (7.56 min) and *m/z* 1293.612 (7.56 min) (Figure 1C). Dereplication of those features was performed through comparison of the UV spectra, accurate masses, isotope distribution and fragmentation patterns obtained in MS/MS analysis against the chemistry databases Reaxys, ChemSpider, and the microbial natural products database Antibase (Laatsch, 2012b). This allowed us to annotate the mass features with *m/z* 1255.6353 and 1277.6175 as the $[M+H]^+$ and $[M+Na]^+$ adduct ions of actinomycin D, respectively, while the mass features with *m/z* 1271.631 and 1293.612 were annotated as $[M+H]^+$ and $[M+Na]^+$ adduct ions of actinomycin $X_{0\beta}$, respectively (Crnovčić *et al.*, 2014). However, the mass feature with an *m/z* value of 1387.6706 $[M+H]^+$ could not be matched to any of the previously reported microbial natural products.

Global Natural Product Social (GNPS) molecular networking (Wang *et al.*, 2016) was subsequently employed to detect MS/MS-based structural relatedness among features in an automated manner. The web-based platform generates a molecular network wherein features with related scaffolds cluster together. Cytoscape 3.7.2 was used for visualization of the generated molecular networks (Shannon *et al.*, 2003). A network representing the ions detected in the crude extract of *Streptomyces sp.* MBT27 grown with 1% glycerol was constructed, revealing 172 nodes clustered in 10 spectral families (Figure 2). The molecular network revealed an actinomycin spectral family containing actinomycin D, X_2 and $X_{0\beta}$. Moreover, the same spectral family included a yet unidentified compound with *m/z* 1387.67. It was closely connected (cosine score > 0.7) to the known actinomycins, suggesting that the molecule was a novel actinomycin. Statistical analysis showed that the extracts with stronger antimicrobial activity contained higher concentrations of actinomycins X_2, $X_{0\beta}$, D and the

new compound, in comparison with the less active ones (ANOVA, $p < 0.05$; Figure 2). It is important to note that actinomycins were only detected in the bioactive extracts.

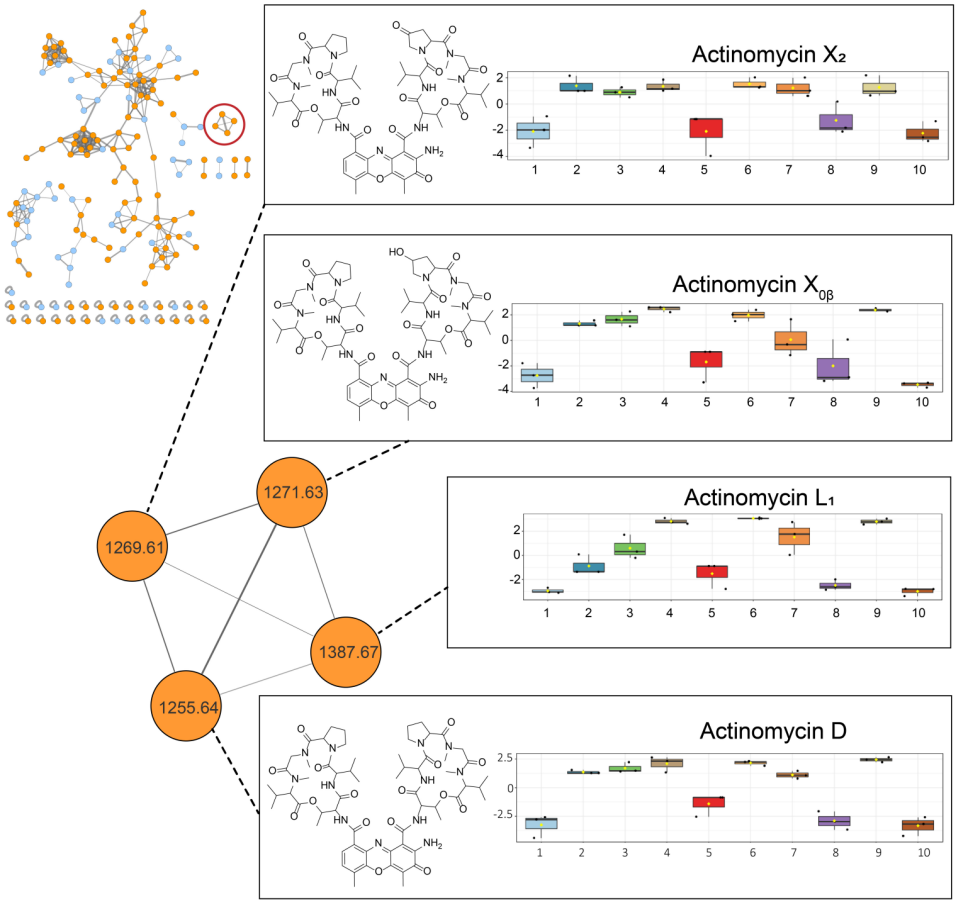


Figure 2. GNPS molecular network of the ions detected in the crude extract of *Streptomyces* sp. MBT27. Cultures were grown for seven days in MM with 1% glycerol. Orange nodes represent ions of the metabolites produced by *Streptomyces* sp. MBT27, while blue nodes represent those of the media components. The actinomycin spectral family is enlarged. Results of ANOVA statistical analysis were mapped onto the molecular network to illustrate the differential production of actinomycin cluster members under various growth conditions. Box plots represent relative intensities of actinomycins X_2, $X_{0\beta}$, and D after log transformation and pareto scaling; together with a compound with an *m/z* value of 1387.67, in cultures grown in MM with the following carbon sources: 1. 1% arabinose; 2. 1% fructose; 3. 1% GlcNAc; 4. 1% glucose; 5. 2% glucose; 6. 1% glycerol; 7. 2% glycerol; 8. 1% mannitol; 9. 1% mannitol + 1% glycerol; 10. 2% mannitol (%ages in w/v).

Large scale fermentation and NMR

To allow identification and structural analysis of the likely novel actinomycin analog, we performed large-scale fermentation of *Streptomyces* sp. MBT27 followed by bioactivity-guided fractionation. The purification process resulted in the isolation of two compounds

(**1**) and (**2**), with the same mass (Figures S4, S5). The NMR spectra of the two compounds were very similar, suggesting that they were diastereomers (Figures S6-S12). Based on 1D and 2D NMR analysis of **1**, together with the molecular formula and degrees of unsaturation dictated by the accurate mass, the structure of the isolated diastereomers was determined as a variant of actinomycin D, whereby an aminal was formed between the amino group of anthranilamide moiety and keto group at the γ position of the proline residues (Figure 3). The prolyl substitution position is the same as that of the hydroxyl and keto groups in actinomycins $X_{0\beta}$ and X_2, respectively. The new actinomycin analog was designated actinomycin L (with L standing for Leiden, the city of its discovery).

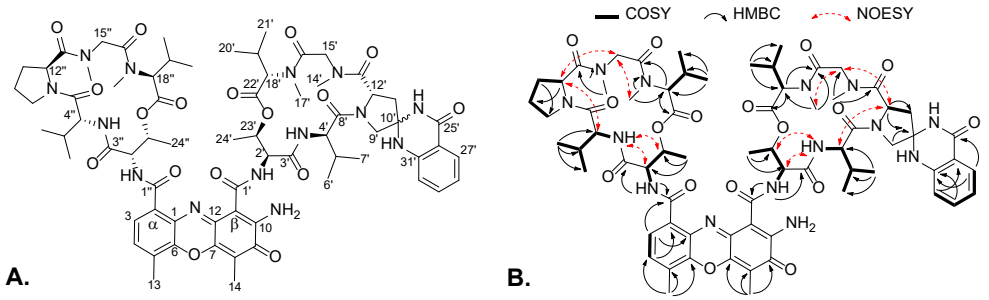


Figure 3. Chemical structures of the new actinomycins. Shown are actinomycin L_1 (10′*S*) (**1**) and L_2 (10′*R*) (**2**) (**A**) and the key COSY, HMBC and NOESY correlations for **1** (**B**).

The second stereoisomer (**2**) was crystallized successfully. Single-crystal X-ray diffraction confirmed the structure obtained for (**1**) based on NMR, and established the absolute configuration to be 2′*S*, 2″*S*, 4′*R*, 4″*R*, 10′*R*, 12′*S*, 12″*S*, 18′*S*, 18″*S*, 23′*R*, 23″*R* by anomalous-dispersion effects in diffraction measurements on the crystal (Figure 4).

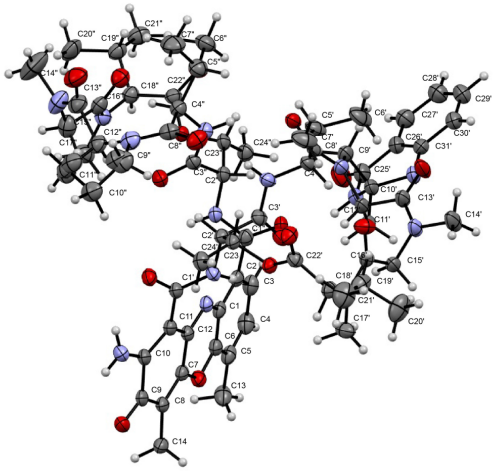


Figure 4. X-ray ORTEP drawing of the crystal structure of compound 2.

As the absolute configuration of the amino acid residues in **2** was consistent with that of previously reported actinomycins (Wang *et al.*, 2017), and considering that the two isomers stemmed from the aminal formation at C-10′, compound **1** is inevitably the 10′*S* isomer of actinomycin L.

Biosynthesis of actinomycin L

Actinomycins D (or X_1), X_2, and $X_{0\beta}$ detected in the extracts of *Streptomyces* sp. MBT27 are members of the actinomycin X complex. Recently it was shown that actinomycins $X_{0\beta}$ and X_2 are formed through the sequential oxidation of the γ-prolyl carbon by the cytochrome P450 enzyme saAcmM (Semsary *et al.*, 2018, Liu *et al.*, 2017). Based on its structure, actinomycin L is most likely formed through an aminalization reaction between the two amino groups of anthranilamide and the γ-keto group on the proline residue of actinomycin X_2. Accordingly, its production should be arrested when one of the precursors is not available. Interestingly, *Streptomyces* sp. MBT27 produced actinomycin L in very low amounts when grown with fructose (1% w/v) as the sole carbon source. Moreover, ANOVA statistical analysis showed that anthranilamide was produced in equally low amounts under the same growth conditions (ANOVA, $p < 0.05$; Figure 5). Under conditions where *Streptomyces* sp. MBT27 produced actinomycin L, namely when grown in MM with 1% GlcNAc, 1% glucose, 1% glycerol, 2% glycerol or in 1% mannitol + 1% glycerol, the strain invariably produced both actinomycin X_2 and anthranilamide. However, under conditions where actinomycin X_2 was produced but not anthranilamide, the strain failed to produce actinomycin L (Figure 5).

We, therefore, wondered if anthranilamide may be a precursor for the biosynthesis of actinomycin L. To test this hypothesis, we performed a feeding experiment, whereby anthranilamide was added to cultures of *Streptomyces* sp. MBT27 grown in MM with 1% fructose, where virtually no actinomycin L was produced. Analysis of the supernatant of the cultures via LC-MS revealed that actinomycin L was readily produced when anthranilamide was added, but not without it (Figure 6A). This strongly suggested that anthranilamide is required for the production of actinomycin L. However, extracts of *Streptomyces* sp. MBT27 fermented with 1% fructose and additional anthranilic acid contained both anthranilamide and actinomycin L (Figure 6A). This suggests that indeed anthranilic acid is converted into anthranilamide, which in turn is incorporated into actinomycin L.

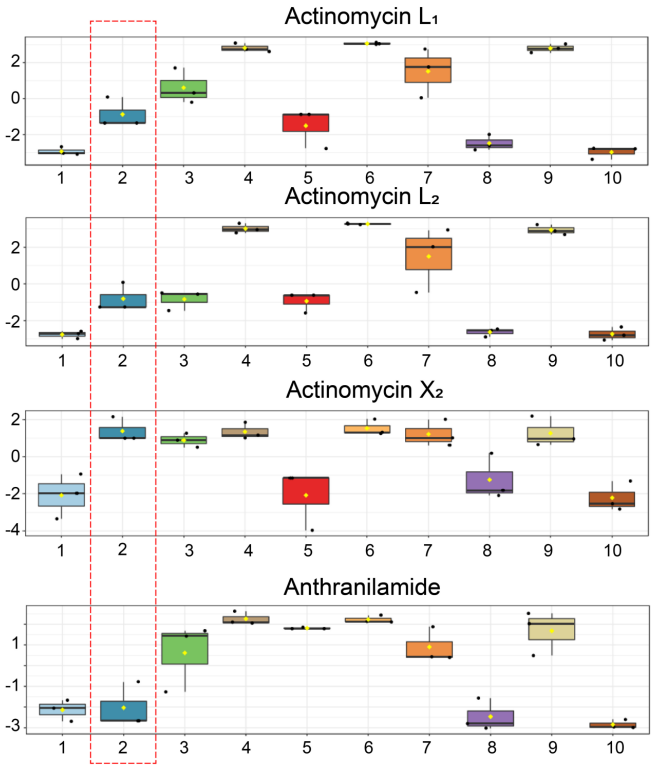


Figure 5. Box plots showing the relative intensities of actinomycin L_1, L_2 and X_2 and anthranilamide after log transformation and pareto scaling. Cultures of *Streptomyces* sp. MBT27 were grown for seven days in MM with the following carbon sources: 1. 1% arabinose; 2. 1% fructose; 3. 1% GlcNAc; 4. 1% glucose; 5. 2% glucose; 6. 1% glycerol; 7. 2% glycerol; 8. 1% mannitol; 9. 1% mannitol +1 % glycerol; 10. 2% mannitol (% ages in w/v). Red box indicates the abundance of actinomycin L_1, L_2 and X_2 and anthranilamide in the cultures grown with fructose (1% w/v). Note that *Streptomyces* sp. MBT27 produced actinomycin L and anthranilamide in very low amounts when fermented with fructose (1% w/v) as the sole carbon source.

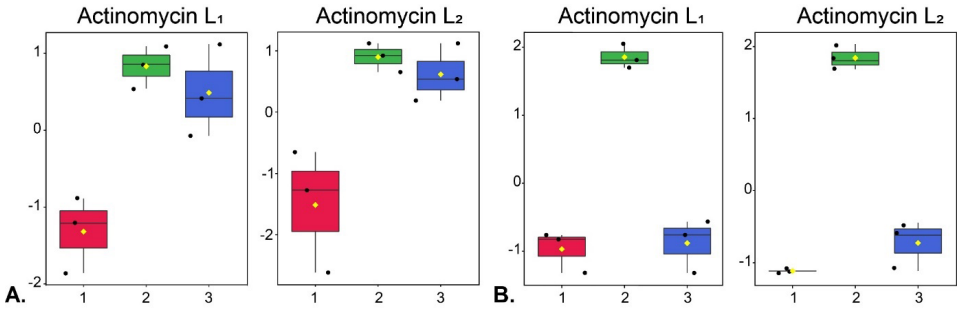


Figure 6. Anthranilamide is required for the biosynthesis of actinomycins L_1 and L_2. Box plots show the relative intensities of actinomycin L_1 and L_2 after log transformation and pareto scaling in the cultures of *Streptomyces* sp. MBT27 (**A**) and *S. antibioticus* (**B**) fermented for seven days in MM with fructose (1% w/v) (1), fed with 0.7 mM anthranilamide (2) and 0.7 mM anthranilic acid (3). Note that *S. antibioticus* produces actinomycin L exclusively in the presence of anthranilamide and not with anthranilic acid; conversely, *Streptomyces* sp. MBT27 is able to convert anthranilic acid into anthranilamide, enabling the production of actinomycin L.

In order to unambiguously verify that actinomycin L was the product of anthranilamide and actinomycin X_2, we conducted another biotransformation experiment, now feeding anthranilamide to *S. antibioticus* IMRU 3720, which is a known producer of actinomycins X_2 and $X_{0\beta}$, but fails to produce actinomycin L under any condition tested. In line with our hypothesis, *S. antibioticus* IMRU 3720 also failed to produce anthranilamide under any of the growth conditions (Figure S13). Excitingly, LC-MS analysis revealed the production of actinomycin L by *S. antibioticus* IMRU 3720 when anthranilamide was fed to the cultures, but never without anthranilamide (Figure 6B). This validates the concept that anthranilamide is a key precursor of actinomycin L. Conversely, when anthranilic acid instead of anthranilamide was added to cultures of *S. antibioticus* IMRU 3720, we failed to detect actinomycin L and anthranilamide (Figure 6B).

The oxidation of the proline residue in actinomycins $X_{0\beta}$ and X_2 occurs following the formation of the two halves of actinomycin, known as 4-MHA PPLs, and prior to the condensation of these halves to form actinomycin (Semsary *et al.*, 2018). Taking this into account, we reasoned that anthranilamide should be incorporated into the actinomycin halves prior to condensation. To check this, 3-hydroxy-4-methylbenzoic acid (4-MHB) was added to cultures of *Streptomyces* sp. MBT27 and of *S. antibioticus* IMRU 3720. 4-MHB is a structural analog of 4-MHA that replaces 4-MHA as a starter unit in the nonribosomal assembly of the actinomycin halves (Semsary *et al.*, 2018). When 4-MHB replaces 4-MHA, 4-MHB containing PPLs accumulate because they cannot react with each other to give a phenoxazinone ring, as is the case with 4-MHA PPLs (Keller, 1984). LC-MS analysis of the 4-MHB-supplemented extracts showed the appearance of the previously reported 4-MHB-containing pentapeptide lactones PPL 1, PPL 0, and PPL 2, and new PPL, designated as PPL 3 (Figures S14–S17, Table S1). The exact mass and fragmentation pattern of PPL 3 was consistent with a 4-MHB containing PPL wherein an anthranilamide moiety had been attached to the proline residue (Figure S17).

Taken together, the feeding experiments convincingly show that actinomycin L is formed through the reaction of anthranilamide with the 4-keto group on the proline residue in the pentapeptide lactone. Moreover, results of the feeding experiments with 3-hydroxy-4-methylbenzoic acid show that this reaction occurs prior to the condensation of the pentapeptide lactones into actinomycin L (Figure 7).

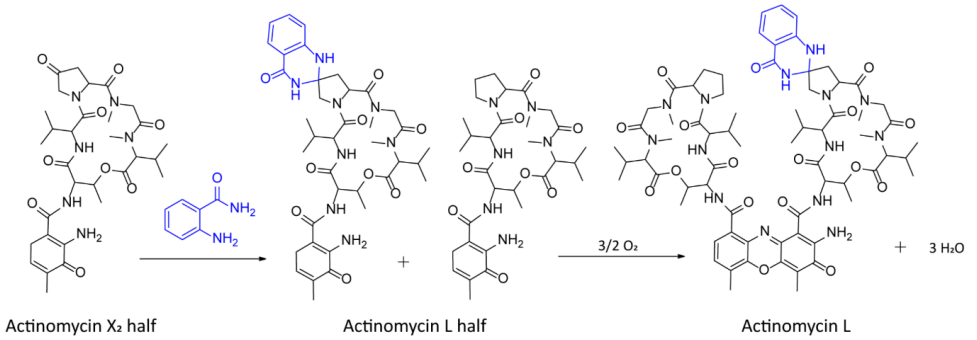


Figure 7. Proposed biosynthetic pathway for actinomycin L. We propose that actinomycin L is formed through the reaction of anthranilamide (blue) with the 4-oxoproline moiety of actinomycin X_2 prior to the condensation of two 4-MHA PPLs into the actinomycin L.

Identification of the actinomycin BGC in *Streptomyces* sp. MBT27

To characterize the BGC responsible for actinomycin biosynthesis and compare the genes with those found in known actinomycin BGCs, *Streptomyces* sp. MBT27 was sequenced using the PacBio platform (GenBank accession number: PRJNA602665). Assembly of the PacBio reads resulted in two contigs of 8.4 Mb and 0.13 Mb in length. Analysis using AntiSMASH 6 (Blin *et al.*, 2021) readily identified the actinomycin BGC in the 8.4 Mb contig. A comparison with the actinomycin X_2 BGC from *S. antibioticus* showed that all genes were highly conserved between the two clusters (Table S2, Figure S18). This strongly suggests that the actinomycin BGC does not specify the observed modifications in the actinomycin structure, and is not responsible for the production of anthranilamide. We have not yet identified the enzyme for the predicted conversion of anthranilic acid to anthranilamide.

Bioactivity of isolated compounds (MIC)

Bioactivity assays were carried out for the actinomycins, to test their ability to act as antibiotics. As expected, the compounds showed selective antibacterial activity against Gram-positive pathogens, and none of the actinomycins presented any activity against *E. coli* ATCC 25922 or *K. pneumoniae* ATCC 700603 (Table 1). All tested compounds showed antibacterial activity against Gram-positive bacteria with MIC values ranging from 1 to 16 µg/mL. Actinomycin L_1 showed somewhat higher bioactivity than actinomycin L_2, while both compounds showed slightly higher MICs than actinomycin D and Actinomycin X_2.

Table 1. Antibacterial activities of compounds expressed as minimal inhibitory concentrations (MIC)[#].

Strain	Actinomycin X_2	Actinomycin D	Actinomycin L_1	Actinomycin L_2
S. aureus MB5393 (methicillin-resistant)	2-4	1-2	4-8	8-16
S. aureus ATCC 29213 (methicillin-sensitive)	1-2	1-2	2-4	n.a.
E. faecium (vancomycin-sentitive)*	1-2	1-2	4-8	8-16
E. faecium VanB (vancomycin-resistant)*	1-2	1-2	4-8	8-16
S. epidermidis *	2-4	2-4	4-8	n.a.
E. coli ATCC 25922	>128	>128	>128	n.a
K. pneumoniae ATCC 700603	>128	>128	>128	n.a

(*), clinical isolates; n.a.: data not available. [#] all concentrations given in μg/mL.

Discussion

Actinomycin was the first antibiotic identified in Actinobacteria (Waksman & Woodruff, 1940). The well-established actinomycin structure is composed of a heterocyclic chromophore and two cyclic PPLs. Biosynthetically, PPL is biosynthesized by an NRPS assembly line with the 4-MHA as the initiating unit (Jones, 1987, Pfennig *et al.*, 1999). 4-MHA is derived from 3-hydroxy-4-methylkynurenine (4-MHK), which is formed by methylation of 3-hydroxykynurenine (3-HK) (Crnovcic *et al.*, 2017). Our work surprisingly revealed a novel structure within the extensively studied actinomycin family, namely that of actinomycin L, which arises via attachment of an anthranilamide moiety to the γ-carbon of one of the proline residues through aminal formation. ANOVA statistical analysis proved that production of anthranilamide is the limiting factor in the biosynthesis of actinomycin L. Feeding experiments with anthranilamide suggested that actinomycin L is formed through the spontaneous reaction of anthranilamide with the 4-oxoproline site of actinomycin X_2 prior to the condensation of the two 4-MHA PPLs into actinomycin L. To the best of our knowledge, the attachment of anthranilamide to a 4-oxoproline moiety is a novel observation.

The actinomycin BGC of *Streptomyces* sp. MBT27 harbors the same genes as that of *S. antibioticus,* with high homology between the genes, which strongly suggests that the modification of actinomycin X_2 to actinomycin L is not encoded by the BGC itself. Indeed, we anticipate that anthranilamide is derived from anthranilic acid in *Streptomyces* sp. MBT27, whereby anthranilic acid in turn is biosynthesized through the shikimate pathway (Dewick, 2009). Anthranilic acid is a commonly produced primary metabolite in *Streptomyces*, while anthranilamide is less common (Shaaban *et al.*, 2012a, Biabani *et al.*, 1998, Ham & Kim, 2018). The actinomycin X_2 producer *S. antibioticus* IMRU 3720 fails to convert anthranilic acid into anthranilamide, which explains why actinomycin L was also not detected in the extracts. However, actinomycin L was produced when we fed cultures of *S. antibioticus* IMRU 3720 with additional anthranilamide, which is fully in line with our proposed biosynthetic pathway. Thus, actinomycin L is an example of a natural product that requires the joining of two separate metabolic pathways, and this is a concept that needs more attention. After all, scientists rely increasingly on heterologous expression and synthetic biology approaches (Kolter & van Wezel, 2016), and these will likely fail if genes are required that do not fall within the main BGC.

The production of actinomycins by *Streptomyces* spp. is strongly influenced by the carbon source, whereby the preferred carbon source varies from strain to strain (Katz *et al.*, 1957, Haque *et al.*, 1995, Theobald *et al.*, 2000, Dalili & Chau, 1988, Williams & Katz, 1977). D-galactose favors actinomycin production in *Streptomyces antibioticus* over arabinose, xylose, glucose, fructose and rhamnose (Katz *et al.*, 1957), while glycerol was the optimal carbon source for actinomycin production by *S. antibioticus* Tü 6040 and *S. antibioticus* SR15.4

(Theobald *et al.*, 2000, Haque *et al.*, 1995). In the case of *Streptomyces* sp. MBT27, growth on MM with glycerol, GlcNAc, fructose and glucose (all 1% w/v) as sole carbon sources were the best carbon sources to promote the production of actinomycins. However, increasing the glucose concentration to 2% blocked the production of actinomycins. Glucose was previously reported to repress the transcription of the gene for hydroxykynureninase, which is involved in the formation of the main actinomycin precursor 4-MHA (Katz *et al.*, 1957). Importantly, in our experiments the carbon source not only promoted the overall production levels, but also contributed to the chemical diversity of the actinomycins, including the production of actinomycin L. This coincided with the production of anthranilamide, an essential substrate to form this novel actinomycin variant.

In the 21st century, genome mining and renewed drug discovery efforts have revealed that Actinobacteria may produce many more molecules than was expected (Nett *et al.*, 2009). What is important to note is that this also applies to well-known families of molecules and in extensively studied model organisms. Examples are the highly rearranged cryptic polyketide lugdunomycin that belongs to the family of angucyclines (Wu *et al.*, 2019), the new glycopeptide corbomycin with a novel mode of action (Culp *et al.*, 2020), as well as the discovery of coelimycin (Gomez-Escribano *et al.*, 2012) and a novel branch of the actinorhodin biosynthetic pathway (Wu *et al.*, 2017) in the model organism *Streptomyces coelicolor*. The discovery of actinomycin L provides another interesting example that we have not yet exhausted the known part of the chemical space. Indeed, the isolation of these novel actinomycins underlines that the biosynthetic potential of Actinobacteria still has major surprises in store, and that we can expect that new molecules can be discovered even within extensively studied microbes and compound classes.

Materials and Methods

General experimental procedures

Optical rotation, FT-IR and UV were measured as previously described (Elsayed *et al.*, 2020). NMR spectra were recorded on a Bruker Ascend 850 MHz NMR spectrometer (Bruker BioSpin GmbH). Data was analyzed using MestReNova 14 software (Mestrelab Research, Santiago de Compostela, Spain). The structures of molecules were drawn using ChemDraw Professional version 16.0 (Perkin-Elmer Informatics). HPLC purification was performed on Waters preparative HPLC system as described (Machushynets *et al.*, 2019). All solvents and chemicals were of HPLC or LC-MS grade, depending on the experiment.

Bacterial strains, growth conditions and metabolite extraction

Streptomyces sp. MBT27 was obtained from the Leiden University strain collection and had previously been isolated from the Qingling Mountains, Shanxi province, China (Zhu *et al.*, 2014a). Cultures were grown in triplicate in 100 mL Erlenmeyer flasks containing 30 mL of liquid minimal medium (MM) (Kieser, 2000), supplemented with various carbon sources, and inoculated with 10 µL of 10^9/mL spore suspension. The carbon sources (percentages in w/v) were: 1% mannitol + 1% glycerol, 1% mannitol, 2% mannitol, 1% glycerol, 2% glycerol, 1% glucose, 2% glucose, 1% fructose, 1% arabinose or 1% *N*-acetylglucosamine (GlcNAc). The cultures were incubated in a rotary shaker at 30 °C and 220 rpm for seven days. Following fermentation, culture supernatants were extracted with ethyl acetate (EtOAc) and evaporated under reduced pressure. In the series of feeding experiments *Streptomyces* sp. MBT27, *S. antibioticus* IMRU 3720 and *S. chrysomallus* ATCC11523 were fermented in MM supplemented with 1% fructose and 0.7 mM anthranilamide. For the directed biosynthesis of non-natural actinomycin X halves *Streptomyces* sp. MBT27 and *S. antibioticus* IMRU 3720 were grown in MM with 1% w/v fructose, 0.7 mM anthranilamide and 0.7 mM 3-hydroxy-4-methylbenzoic acid (4-MHB) and extracted with EtOAc.

Genome sequencing, assembly and annotation

Streptomyces sp. MBT27 was grown in YEME at 30 °C and 220 rpm for 48 hours. DNA was extracted from *Streptomyces* sp. MBT27 as described (Kieser, 2000). DNA quality was verified by agarose gel electrophoresis. PacBio sequencing and assembly was performed by Novogene (UK). Generally, library was 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 Flye (version 2.8.1) (Kolmogorov *et al.*, 2019). Biosynthetic gene clusters (BGCs) in this genome were annotated using AntiSMASH 6.0 (Blin *et al.*, 2021). The actinomycin BGC was then extracted

and compared with the same cluster from *S. antibioticus* IMRU 3720 using clinker version 0.0.20) with default settings (Gilchrist & Chooi, 2021).

Up-scale fermentation, extraction and fractionation

Large-scale fermentation, extraction end fractionation were performed as previously described (Machushynets *et al.*, 2019). The fractions eluted with *n*-hexane–acetone (1:1) was subjected to a SunFire C_{18} column (10 μm, 100 Å, 19×150 mm) eluted with a H_2O–MeOH gradient of 50–100% in 20 min, at a flow rate of 15 mL/min. The fraction containing the actinomycins was collected and further purified on semi-preparative SunFire C_{18} column (5 μm, 100 Å, 10×250 mm), run at 3 mL/min, and eluted using a H_2O–MeOH gradient of 70–100% in 20 min, to yield actinomycins L_1 (1, 2.9 mg), L_2 (2, 1.3 mg), X_2 (3, 1 mg), $X_{0\beta}$ (4, 2.9 mg), and D (5, 3.2 mg).

Actinomycin L_1 (**1**): red amorphous powder; $[\alpha]_D^{25}$ -38 (c 0.2, MeOH); UV (MeOH) λ_{max} (log ε) 211 (1.957), 312 (0.124), 427.5 (0.152), 438 (0.151) nm; IR ν_{max} 3301, 2963, 2921, 2859, 1740, 1662, 1644, 1585, 1521, 1465, 1405, 1300, 1262, 1191, 1097 cm^{-1}; ^1H and ^{13}C NMR data, see Table 2; HRESIMS (positive mode) *m/z* 1387.6681 $[M+H]^+$ (calcd. for $C_{69}H_{91}N_{14}O_{17}$, 1387.6681).

Actinomycin L_2 (2): red amorphous powder; $[\alpha]_D^{25}$ -54 (c 0.2, MeOH); UV (MeOH) λ_{max} (log ε) 226 (2.189), 364 (0.186) nm, 426.5 (0.367) nm; IR ν_{max} 3308, 2943, 2929, 2831, 1748, 1662, 1644, 1585, 1448, 1406, 1302, 1262, 1191, 1113 cm^{-1}; HRESIMS (positive mode) *m/z* 1387.6674 $[M+H]^+$ (calcd. for $C_{69}H_{91}N_{14}O_{17}$, 1387.6681). The X-ray diffraction experiment on a crystal grown from MeOH further confirmed the structure and determined the absolute configuration of compound **2** (Figure 4) (CCDC 2110000). Structure visualization and image preparation was done using Mercury version 4.3.0 (Macrae *et al.*, 2020).

Table 2. ^{1}H and ^{13}C NMR data of 1 in $CDCl_3$ at 298 K[a].

Position	δ_C, type	δ_H, mult. (*J* in Hz)	Position	δ_C, type	δ_H, mult. (*J* in Hz)
1	129.5, C		**22′**	167.4, C	
2	130.9, C		**23′**	74.5, CH	5.29, qd (6.3, 3.1)
3	125.9	7.75, m	**24′**	17.5, CH_3	1.23, d (6.3)
4	130.0, CH	7.44, dd (7.6, 0.9)	**25′**	163.1, C	
5	128.5, C		**26′**	114.3, C	
6	140.6, C		**27′**	128.6, CH	7.90, dd (7.9, 1.6)
7	144.7, C		**28′**	118.7, CH	6.81, dt (0.9, 7.9)
8	113.6, C		**29′**	133.6, CH	7.30, dt (1.6, 7.9)
9	178.9, C		**30′**	113.9, CH	6.67, dd (7.9, 0.9)
10	ND		**31′**	146.7, C	
11	ND		**1″**	166.4, C	
12	ND		**2″**	55.5, CH	4.69, dd (6.6, 2.7)
13	15.1, CH_3	2.58, d (0.9)	**NH-2″**		7.64, d (6.6)
14	7.7, CH_3	2.26, s	**3″**	168.6, C	
1′	166.1, C		**4″**	58.6, CH	3.57, m
2′	54.7, CH	4.45, dd (6.4, 3.1)	**NH-4″**		7.75, m
NH-2′		7.38, d (6.4)	**5″**	31.8, CH	2.10, m
3′	168.7, C		**6″**	18.9, CH_3	1.10, d (6.7)
4′	57.9, CH	3.56, m	**7″**	19.2, CH_3	0.91, d (6.7)
NH-4′		8.28, d (5.4)	**8″**	ND	
5′	31.0, CH	2.17, m	**9″**	47.5, CH_2	3.82, m 3.75, m
6′	18.8, CH_3	1.13, d (6.7)	**10″**	22.8, CH_2	2.29, m 2.07, m
7′	19.1, CH_3	0.87, d (6.7	**11″**	31.1, CH_2	2.84, m 1.86, dd (11.9, 6.9)
8′	ND		**12″**	56.3, CH	5.95, d (9.2)
9′	60.5, CH_2	4.49, d (13.4) 4.07, d (13.4)	**13″**	172.9, C	
10′	76.8, C		**14″**	34.9, CH_3	2.89, s
11′	43.1, CH_2	2.94, m 2.25, m	**15″**	51.3, CH_2	4.72, d (17.1) 3.65, d (17.1)
12′	56.7, CH	6.27, d (10.5)	**16″**	165.4, C	
13′	173.1, C		**17″**	39.4, CH_3	2.93, s
14′	35.1, CH_3	2.93, s	**18″**	71.4, CH	2.70, m
15′	51.1, CH_2	4.36, d (17.1) 3.58, d (17.1)	**19″**	27.0, CH	2.69, m
16′	166.3, C		**20″**	19.0, CH_3	0.75, d (6.3)
17′	39.1, CH_3	2.95, s	**21″**	21.7, CH_3	0.98, d (6.0)
18′	71.3, CH	2.67, m	**22″**	167.2, C	
19′	26.9, CH	2.65, m	**23″**	75.2, CH	5.22, qd (6.3, 2.7)
20′	19.0, CH_3	0.75, d (6.8)	**24″**	17.9, CH_3	1.30, d (6.3)
21′	21.6, CH_3	0.96, d (6.4)			

[a] ^{1}H 850 MHz and ^{13}C NMR resonances inferred from HSQC and HMBC spectra.

Single crystal X-ray crystallography

All reflection intensities were measured at 100(2) K using a SuperNova diffractometer (equipped with Atlas detector) with Cu Kα radiation (λ = 1.54178 Å) under the program CrysAlisPro (Version CrysAlisPro 1.171.39.29c, Rigaku OD, 2017). The same program was used to refine the cell dimensions and for data reduction. The structure was solved with the program SHELXS-2018/3 and was refined on F^2 with SHELXL-2018/3 (Sheldrick, 2015). Analytical numeric absorption correction using a multifaceted crystal model was applied using CrysAlisPro. The temperature of the data collection was controlled using the system Cryojet (manufactured by Oxford Instruments). Prior to mounting the chosen single crystal on the diffractometer, crystals were placed in some Parabar 10312 on a microscope slide and cooled via a cold $N_2(g)$ stream in order to protect the crystal from potential solvent loss. The H atoms were placed at calculated positions (unless otherwise specified) using the instructions AFIX 13, AFIX 23, AFIX 43, AFIX 137 or AFIX 147 with isotropic displacement parameters having values 1.2 or 1.5 U_{eq} of the attached C or O atoms. The H atoms attached to N2, N1', N2', N6', N7', N1" and N2" were found from difference Fourier maps, and their coordinates were refined pseudofreely using the DFIX instruction in order to keep the N–H bonds within an acceptable range. The structure is partly disordered. The asymmetric unit contains one molecule of the target compound as well as some significant amount of lattice MeOH solvent molecules. Six solvent molecules were modelled as ordered with four being fully occupied and two with partial occupancy factors of 0.918(11) and 0.487(10). The contribution of the remaining amount of disordered lattice solvent molecules has been removed using the SQUEEZE procedure in Platon (Spek, 2009). Furthermore, the atoms C9", C10" and C11" are disordered over two orientations, and the occupancy factor of the major component of the disorder refines to 0.533(10). The absolute configuration has been established by anomalous-dispersion effects in diffraction measurements on the crystal, and the Flack and Hooft parameters refine to 0.08(5) and 0.06(5), respectively. The model has chirality S, S, R, R, R, S, S, S, R, R at C2', C2", C4', C4", C10'. C12', C18', C18", C23', C23", respectively. Structure visualization and image preparation was done using Mercury (Macrae *et al.*, 2020).

Antimicrobial activity assay and MIC determination

The antimicrobial activity of the compounds was tested in liquid inhibition assays against seven pathogens including Gram-negative and Gram-positive bacteria (*Escherichia coli* ATCC25922, *Klebsiella pneumoniae* ATCC700603, methicillin-resistant *Staphylococcus aureus* MB5393, methicillin-sensitive *Staphylococcus aureus* ATCC29213, linezolid-resistant *Staphylococcus epidermidis* (clinical isolate), vancomycin-sensitive *Enterococcus faecium* (clinical isolate), and vancomycin-resistant *Enterococcus faecium* VanB (clinical isolate), as described (Audoin *et al.*, 2013). Each compound was serially diluted in DMSO with a dilution

factor of 2 to test 10 concentrations starting at 128 µg/mL in all the antimicrobial assays. The MIC was defined as the lowest concentration of compound that inhibited ≥95% of the growth of a microorganism after overnight incubation. The Genedata Screener software (Genedata, Inc., Basel, Switzerland) was used to process and analyze the data and to calculate the RZ' factor in the assay that was between 0.90 and 0.98 supporting its robustness.

LC-MS/MS analysis

For LC-MS analyses, extracts were dissolved in MeOH to a final concentration of 1 mg/mL, and 1 µL was injected into Waters Acquity UPLC system coupled to Agilent 6530 QTOF MS. Samples were analyzed according to the protocol that was previously published (Machushynets *et al.*, 2019). LC-MS/MS acquisition of the pure compounds was performed using Shimadzu Nexera X2 UHPLC system coupled to Shimadzu 9030 QTOF mass spectrometer as previously described (Xiao *et al.*, 2020). LC–MS/MS acquisition for molecular networking was performed using Thermo Instruments MS system (LTQ Orbitrap XL, Bremen, Germany) equipped with an electrospray ionization source (ESI) as described (Machushynets *et al.*, 2019).

Computation of mass spectral networks

MS/MS raw data were converted to a 32-bit mzXML file using MSConvert (ProteoWizard) (Chambers *et al.*, 2012) and spectral networks were assembled using Global Natural Product Social molecular networking (GNPS) (https://gnps.ucsd.edu) as described (Wang *et al.*, 2016). Briefly, the precursor ion mass tolerance was set to 2.0 Da and a MS/MS fragment ion tolerance of 0.5 Da, while the minimum cosine score was set to 0.7. The data were clustered using MSCluster with a minimum cluster size of three spectra. The spectra in the network were also searched against GNPS spectral libraries. A minimum score of 0.7 was set for spectral library search, with at least six fragment peaks matching. Cytoscape 3.7.2 was used for visualization of the generated molecular networks (Shannon *et al.*, 2003). In Cytoscape the detected chemical space is displayed as nodes and edges (nodes correspond to a specific consensus spectrum; edges represent significant pairwise alignment between nodes). The edge thickness was set to represent the cosine score, with thicker lines indicating higher similarity between nodes. LC-MS/MS data were deposited in the MassIVE Public GNPS data set (MSV000085106). The molecular networking job in GNPS can be found at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=0c9153470404488d8927289139f875d3. The annotated MS/MS spectra were deposited in the GNPS spectral library for actinomycin L_1 (CCMSLIB00005718892) and L_2 (CCMSLIB00005718891).

Statistical analysis

Prior to statistical analysis, mzXML files, which were converted using Shimadzu LabSolutions Postrun Analysis, were imported into Mzmine 2.31 (Pluskal *et al.*, 2010) and processed as

previously described (Machushynets *et al.*, 2019). The aligned peak list was exported as a comma-separated file for statistical analysis. Statistical analysis was performed using MetaboAnalyst (Chong *et al.*, 2018), where log transformation and pareto scaling was initially applied to the data. The normalized data were subjected to principal components analysis (PCA) and orthogonal partial least squares discriminant analysis (OPLS-DA). The quality of the models was evaluated with the relevant R^2 and Q^2. To identify the difference in intensity of a single mass feature among multiple growth conditions, one-way ANOVA was performed, followed by a *post hoc* Tukey's honest significant difference (HSD) test.

Acknowledgments

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. Accurate mass of $[M+H]^+$ and product ions of the PPLs analyzed be ESI-QTOF MS/MS.

Compound	Scan #, t_R (min)	Composition	Proposed ion	Measurement (m/z)	Calculated (m/z)	Difference (ppm)
PPL0	2973 @5.29	$C_{31}H_{46}N_5O_9$	$[M+H]^+$	632.3170	632.3295	-18.99
		$C_{31}H_{44}N_5O_8$	$[M+H-H_2O]^+$	614.3177	614.3189	-1.2
		$C_{25}H_{33}N_4O_7$	$[M+H-(H\text{-}MeVal\text{-}OH)]^+$	501.2347	501.2349	0.65
		$C_{22}H_{28}N_3O_6$	$[M+H-(H\text{-}Sar\text{-}MeVal\text{-}OH)]^+$	430.1969	430.1978	-0.84
		$C_{19}H_{35}N_4O_6$	$[(H\text{-}Val\text{-}OxoPro\text{-}Sar\text{-}MeVal\text{-}OH)+H]^+$	415.2552	415.2556	0.214
		$C_{19}H_{33}N_4O_5$	$[(H\text{-}Val\text{-}OxoPro\text{-}Sar\text{-}MeVal\text{-}OH)+H\text{-}H_2O]^+$	397.2441	397.2450	-1.124
		$C_{14}H_{26}N_3O_5$	$[(H\text{-}Hyp\text{-}Sar\text{-}MeVal\text{-}OH)+H]^+$	316.1870	316.1872	0.957
		$C_{14}H_{24}N_3O_4$	$[(H\text{-}Hyp\text{-}Sar\text{-}MeVal\text{-}OH)+H\text{-}H_2O]^+$	298.1761	298.1766	-0.110
		$C_{13}H_{22}N_3O_4$	$[(H\text{-}Val\text{-}HyPro\text{-}Sar)+H]^+$	284.1606	284.1610	0.413
		$C_{12}H_{12}NO_3$	$[M+H-(H\text{-}Val\text{-}Hyp\text{-}Sar\text{-}MeVal\text{-}OH)]^+$	218.0816	218.0817	1.973
				213.1233		
		$C_9H_{19}N_2O_3$	$[(H\text{-}Sar\text{-}MeVal\text{-}OH)+H]^+$	203.1391	203.1395	0.399
		$C_6H_{14}NO_2$	$[(H\text{-}MeVal\text{-}OH)+H]^+$	132.1015	132.1024	-3.067
		$C_4H_6NO_2$	$[Thr+H]^+$	100.0389	100.0398	-4.048

PPL1	#3129 @5.60	$C_{31}H_{46}N_5O_8$	$[M+H]^+$	616.3326	616.3346	-2.417
		$C_{31}H_{44}N_5O_7$	$[M+H-H_2O]^+$	598.3248	598.3241	2.131
		$C_{25}H_{33}N_4O_6$	[M+H-(H-MeVal-OH)]$^+$	485.2392	485.2400	-0.538
		$C_{22}H_{28}N_3O_5$	[M+H-(H-Sar-MeVal-OH)]$^+$	414.2025	414.2028	0.368
		$C_{19}H_{35}N_4O_5$	[(H-Val-Pro-Sar-MeVal-OH)+H]$^+$	399.2607	399.2607	1.261
		$C_{19}H_{33}N_4O_4$	[(H-Val-Pro-Sar-MeVal-OH)+H-H_2O]$^+$	381.2504	381.2501	2.014
		$C_{17}H_{21}N_2O_4$	[M+H-(H-Pro-Sar-MeVal-OH)]$^+$	317.1496	317.1501	0.052
		$C_{14}H_{26}N_3O_4$	[(H-Pro-Sar-MeVal-OH)+H]$^+$	300.1916	300.1923	-0.609
		$C_{14}H_{24}N_3O_3$	[(H-Pro-Sar-MeVal-OH)+H-H_2O]$^+$	282.1809	282.1817	-1.127
		$C_{13}H_{22}N_3O_3$	[(H-Val-Pro-Sar)+H]$^+$	268.1657	268.1661	0.492
		$C_9H_{19}N_2O_3$	[(H-Sar-MeVal-OH)+H]$^+$	203.1392	203.1395	0.891
				187.1078		
		$C_8H_{13}N_2O_2$	[(H-Pro-Sar)+H]$^+$	169.0968	169.0977	-2.095
		$C_6H_{14}NO_2$	[(H-MeVal-OH)+H]$^+$	132.1016	132.1024	-2.310
		$C_4H_6NO_2$	[Thr+H]$^+$	100.0391	100.0398	-2.049

PPL2	#3708 @6.60	$C_{31}H_{44}N_5O_9$	$[M+H]^+$	630.3132	630.3139	-0.245
		$C_{31}H_{42}N_5O_8$	$[M+H-H_2O]^+$	612.3027	612.3033	-0.146
			??$[M+H-CO]^+$	602.5191		
		$C_{25}H_{33}N_4O_8$	$[M+H-(H\text{-}MeVal)]^+$	517.2286	517.2298	-1.335
		$C_{25}H_{31}N_4O_7$	$[M+H-(H\text{-}MeVal\text{-}OH)]^+$	499.2190	499.2192	0.549
		$C_{22}H_{26}N_3O_6$	$[M+H-(H\text{-}Sar\text{-}MeVal\text{-}OH)]^+$	428.1814	428.1821	-0.495
		$C_{19}H_{33}N_4O_6$	$[(H\text{-}Val\text{-}OxoPro\text{-}Sar\text{-}MeVal\text{-}OH)+H]^+$	413.2395	413.2400	0.094
		$C_{19}H_{31}N_4O_5$	$[(H\text{-}Val\text{-}OxoPro\text{-}Sar\text{-}MeVal\text{-}OH)+H\text{-}H_2O]^+$	395.2293	395.2294	1.021
		$C_{14}H_{24}N_3O_5$	$[(H\text{-}OxoPro\text{-}Sar\text{-}MeVal\text{-}OH)+H]^+$	314.1711	314.1715	0.168
				300.1555		
		$C_{14}H_{22}N_3O_4$	$[(H\text{-}Val\text{-}OxoPro\text{-}Sar\text{-}MeVal\text{-}OH)+H\text{-}H_2O]^+$	296.1605	296.1610	0.059
		$C_{13}H_{20}N_3O_4$	$[(H\text{-}Val\text{-}OxoPro\text{-}Sar)+H]^+$	282.1448	282.1453	-0.116
		$C_{12}H_{12}NO_3$	$[M+H-(H\text{-}Val\text{-}Hyp\text{-}Sar\text{-}MeVal\text{-}OH)]^+$	218.0815	218.0817	1.514
		$C_9H_{19}N_2O_3$	$[(H\text{-}Sar\text{-}MeVal\text{-}OH)+H]^+$	203.1391	203.1395	0.399
				183.0761		
		$C_6H_{14}NO_2$	$[(H\text{-}MeVal\text{-}OH)+H]^+$	132.1017	132.1024	-1.554
		$C_4H_6NO_2$	$[Thr+H]^+$	100.0391	100.0398	-2.049

PPL3	#3312 @5.96	$C_{38}H_{50}N_7O_9$	$[M+H]^+$	748.3663	748.3670	-0.204
		$C_{38}H_{48}N_7O_8$	$[M+H-H_2O]^+$	730.3561	730.3564	0.290
				702.3638		
		$C_{32}H_{37}N_6O_7$	$[M+H-(H\text{-}MeVal\text{-}OH)]^+$	617.2716	617.2723	-0.363
				596.3200		
		$C_{26}H_{39}N_6O_6$	$[(H\text{-}Val\text{-}AntPro\text{-}Sar\text{-}MeVal\text{-}OH)+H]^+$	531.2928	531.2931	0.453
		$C_{26}H_{37}N_6O_5$	$[(H\text{-}Val\text{-}AntPro\text{-}Sar\text{-}MeVal\text{-}OH)+H\text{-}H_2O]^+$	513.2820	513.2825	0.010
				485.2873		
		$C_{21}H_{30}N_5O_5$	$[(H\text{-}AntPro\text{-}Sar\text{-}MeVal\text{-}OH)+H]^+$	432.2247	432.2246	1.283
		$C_{21}H_{28}N_5O_4$	$[(H\text{-}AntPro\text{-}Sar\text{-}MeVal\text{-}OH)+H\text{-}H_2O]^+$	414.2137	414.2141	0.288
		$C_{20}H_{26}N_5O_4$	$[(H\text{-}Val\text{-}AntPro\text{-}Sar)+H]+$	400.1982	400.1984	0.673
		$C_{17}H_{21}N_4O_3$	$[(H\text{-}Val\text{-}AntPro)+H]+$	329.1607	329.1613	
		$C_{17}H_{21}N_2O_4$	$[M+H-(H\text{-}Pro\text{-}Sar\text{-}MeVal\text{-}OH)]^+$	317.1493	317.1501	-0.894
		$C_{15}H_{17}N_4O_3$	$[(H\text{-}Pro\text{-}Sar)+H]^+$	301.1295	301.1300	-0.056
				284.1032		
				230.0925		
		$C_9H_{19}N_2O_3$	$[(H\text{-}Sar\text{-}MeVal\text{-}OH)+H]^+$	203.1390	203.1395	-0.093
				175.0868		
		$C_6H_{14}NO_2$	$[(H\text{-}MeVal\text{-}OH)+H]^+$	132.1014	132.1024	-3.824
		$C_4H_6NO_2$	$[Thr+H]^+$	100.0401	100.0398	7.947

Table S2. Gene organization of the actinomycins biosynthetic gene cluster in *Streptomyces* sp. MBT27 and similarities to corresponding protein sequences encoded by orthologues in the *S. antibioticus* IMRU 3720 biosynthetic gene cluster.

S. antibioticus IMRU 3720 gene	Function	*Streptomyces* sp. MBT27 actinomycin cluster ORF	Identity	Similarity
saacmT	Hypothetical protein	ORF01	0.80	0.89
saacmS	Hypothetical protein	ORF02	0.84	0.93
saacmR	MbtH-like protein	ORF03	0.86	0.94
saacmD	4-MHA carrier protein AcmACP	ORF04	0.60	0.69
saacmA	Peptide synthetase ACMS I	ORF05	0.71	0.78
saacmB	Peptide synthetase ACMS II	ORF06	0.70	0.78
saacmC	Peptide synthetase ACMS III	ORF07	0.77	0.85
saacmE	Hypothetical protein	ORF08	0.91	0.93
saacmF	Aryl formamidase	ORF09	0.79	0.82
saacmG	Tryptophan 2,3-dioxygenase	ORF10	0.83	0.89
saacmK	Kynureninase	ORF11	0.86	0.91
saacmL	Methyltransferase	ORF12	0.88	0.93
saacmM	Cytochrome P450	ORF13	0.85	0.88
saacmN	Ferredoxin	ORF14	0.69	0.78
saacmO	LmbU-like protein	ORF15	0.72	0.80
saacmP	TetR family transcriptional regulator	ORF16	0.77	0.86
saacmQ	Siderophore-interacting protein	ORF17	0.76	0.85
saacmrA	ABC transporter ATPase subunit	ORF18	0.89	0.93
saacmrB	ABC 2-type transporter	ORF19	0.93	0.96
saacmrC	UvrA-like protein	ORF20	0.86	0.90

Table S3. X-ray crystallography data for 2.

	xs2313a
Crystal data	
Chemical formula	$C_{69}H_{90}N_{14}O_{17}$·5.405(CH_4O)
M_r	1560.87
Crystal system, space group	Trigonal, $P3_221$
Temperature (K)	100
a, c (Å)	18.4731 (2), 42.4216 (5)
V (Å^3)	12537.1 (3)
Z	6
Radiation type	Cu $K\alpha$
μ (mm^{-1})	0.77
Crystal size (mm)	0.23 × 0.15 × 0.07
Data collection	
Diffractometer	SuperNova, Dual, Cu at zero, Atlas
Absorption correction	Analytical *CrysAlis PRO* 1.171.40.53 (Rigaku Oxford Diffraction, 2019) Analytical numeric absorption correction using a multifaceted crystal model based on expressions derived by (Clark & Reid, 1995). Empirical absorption correction using spherical harmonics, implemented in SCALE3 ABSPACK scaling algorithm.
T_{min}, T_{max}	0.897, 0.957
No. of measured, independent and observed [$I > 2\sigma(I)$] reflections	90598, 14994, 13477
R_{int}	0.043
(sin θ/λ)max (Å^{-1})	0.598
Refinement	
$R[F^2 > 2\sigma(F^2)]$, $wR(F^2)$, S	0.046, 0.128, 1.03
No. of reflections	14994
No. of parameters	1091
No. of restraints	79

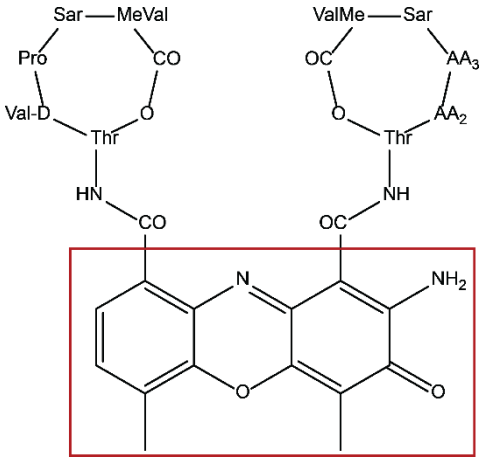


Figure S1. Chemical structure of actinomycin. Actinomycins are composed of a chromophore group highlighted with the red square and two pentapeptide chains, where amino acid AA2 and AA3 differs depending on the actinomycin group.

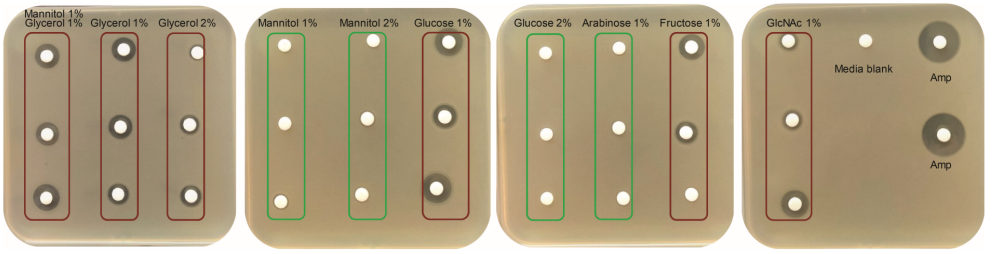


Figure S2. Representative results of antimicrobial activity of *Streptomyces* sp. MBT27 extracts against *B. subtilis 168*. *Streptomyces* sp. MBT27 was fermented in minimal medium (MM) with different carbon sources, namely (percentages in w/v): 1% of both mannitol and glycerol, 1% mannitol, 2% mannitol, 1% glycerol, 2% glycerol, 1% glucose, 2% glucose, 1% fructose, 1% arabinose, or 1% *N*-acetylglucosamine (GlcNAc), and extracted with ethyl acetate. The red boxes indicate active extracts of *Streptomyces* sp. MBT27, while the green boxes represent extracts without antibacterial activity against *B. subtilis 168*. Ampicillin was used as a positive control. Note, that extracts of cultures fermented with different carbon sources showed different bioactivity profiles.

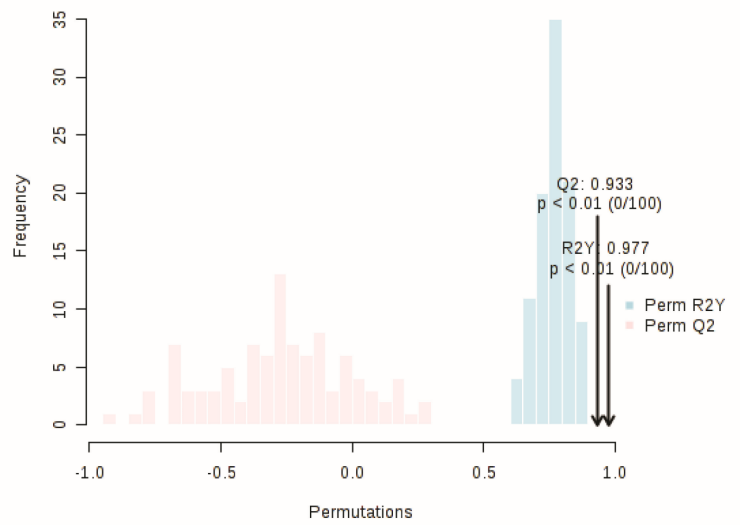


Figure S3. Permutation validation of OPLS-DA model.

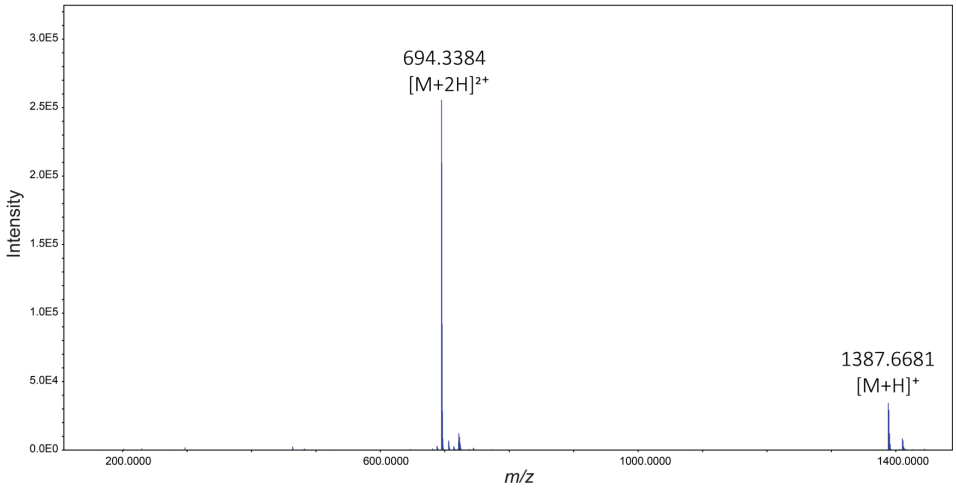


Figure S4 HRMS spectrum of 1.

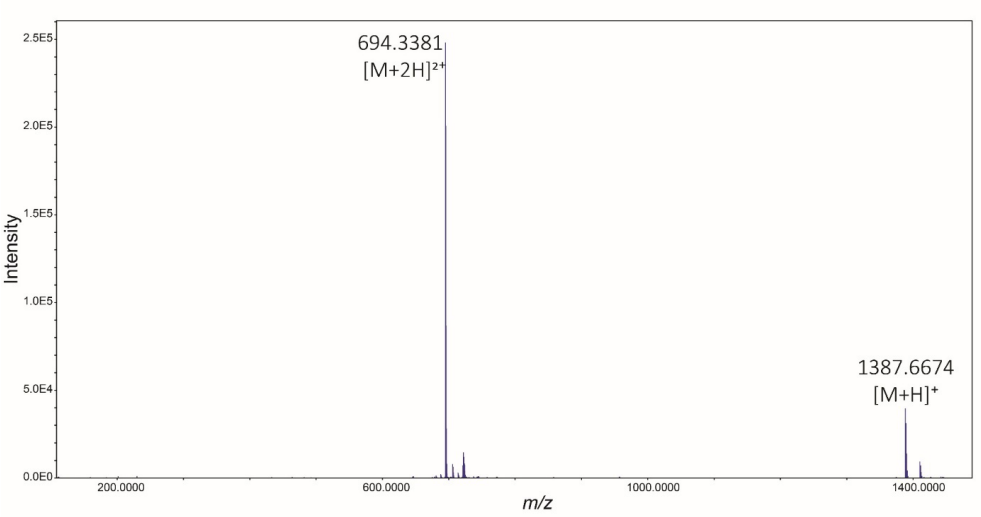


Figure S5. HRMS spectrum of 2.

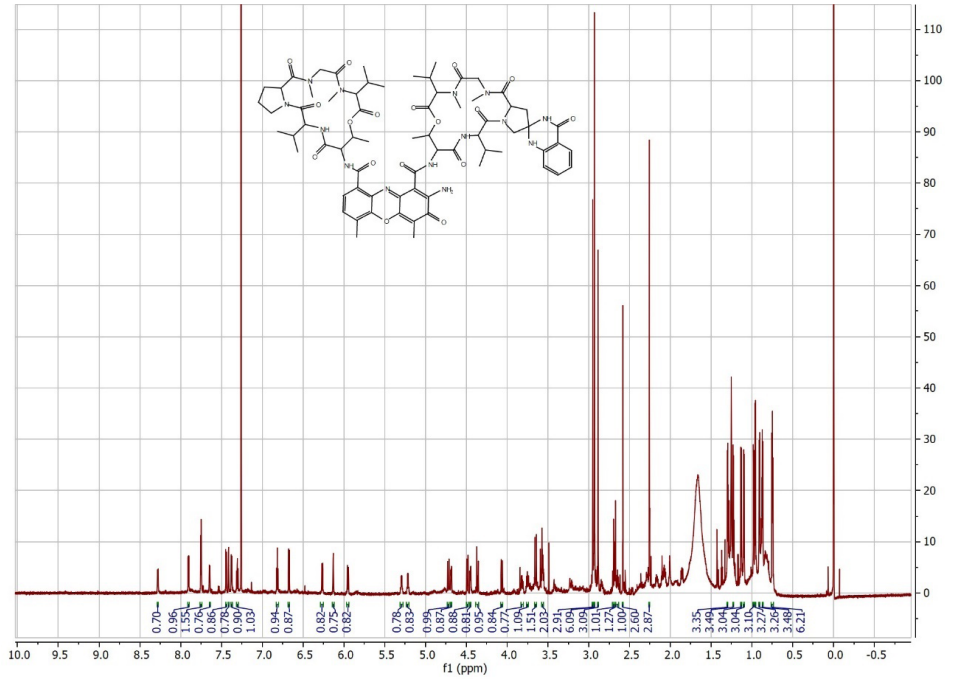


Figure S6. ^{1}H NMR spectrum of 1 (850 MHz, in $CDCl_3$ with TMS).

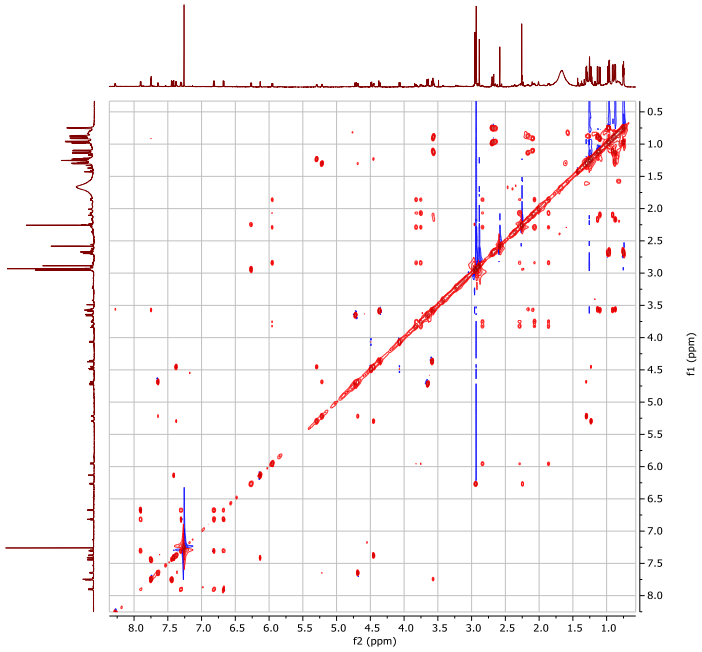


Figure S7. $^{1}H–^{1}H$ TOCSY spectrum of 1 (850 MHz, in CDCl3 with TMS).

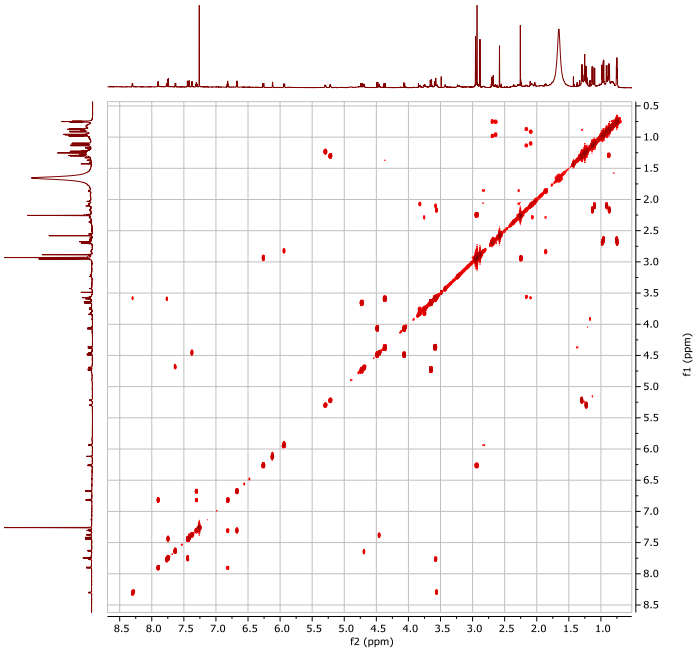


Figure S8. $^{1}H–^{1}H$ COSY spectrum of 1 (850 MHz, in CDCl3 with TMS).

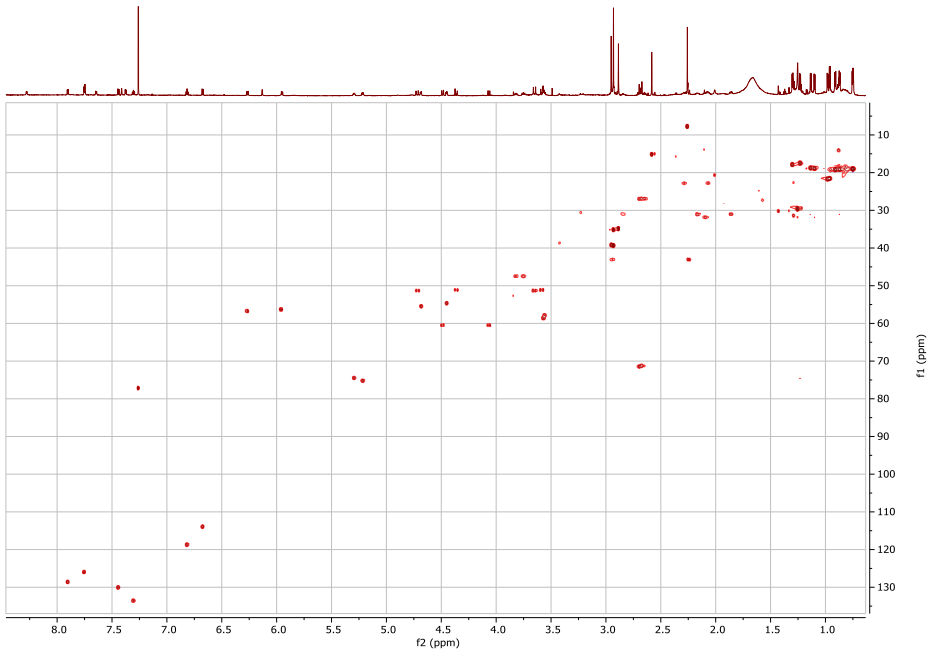


Figure S9. HSQC spectrum of 1 (850 MHz, in $CDCl_3$ with TMS).

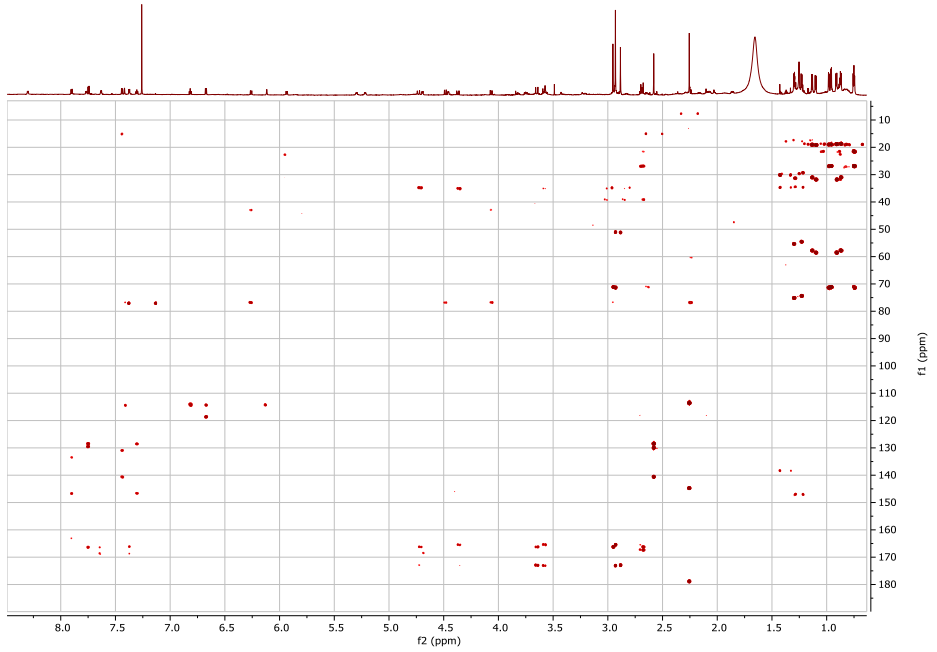


Figure S10. HMBC spectrum of 1 (850 MHz, in $CDCl_3$ with TMS).

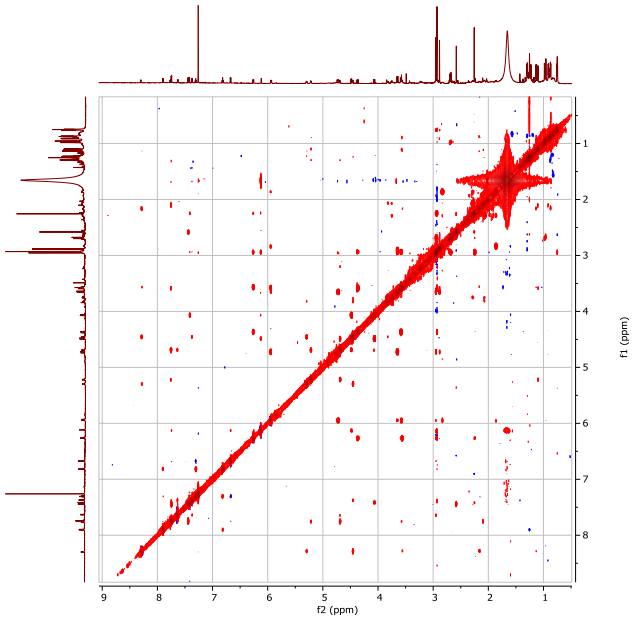


Figure S11. NOESY spectrum of 1 (850 MHz, in $CDCl_3$ with TMS).

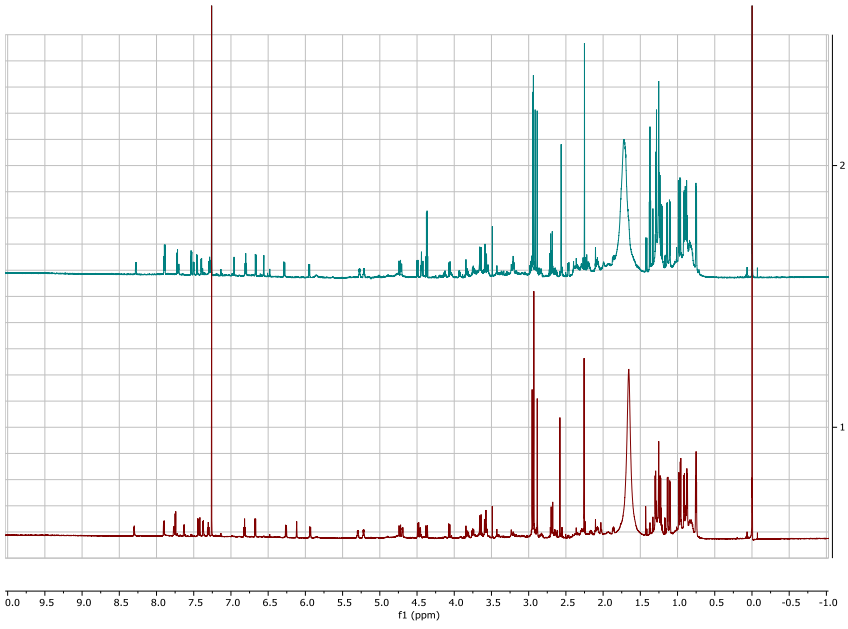


Figure S12. Stacked ^{1}H NMR spectra of 1 and 2 (850 MHz, in CDCl3 with TMS).

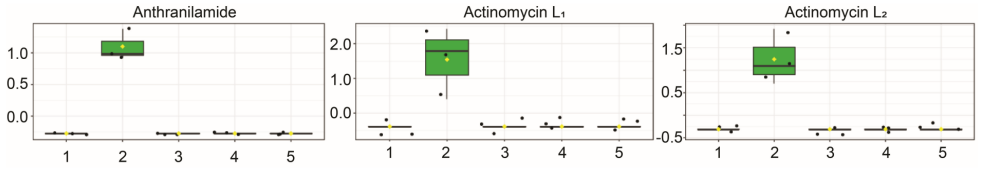


Figure S13. *S. antibioticus* is incapable to produces actinomycin L unless anthranilamide is supplied. Box plots showing the relative intensities of anthranilamide, actinomycin L_1 and L_2 after log transformation and pareto scaling in the cultures of *S. antibioticus* fermented for seven days in MM with different carbon sources 1. 1% fructose; 2. 1% fructose + 0.7 mM anthranilamide; 3. 1% glycerol; 4. 1% mannitol + 1 % glycerol; 5. 2% glycerol (all %ages in w/v).

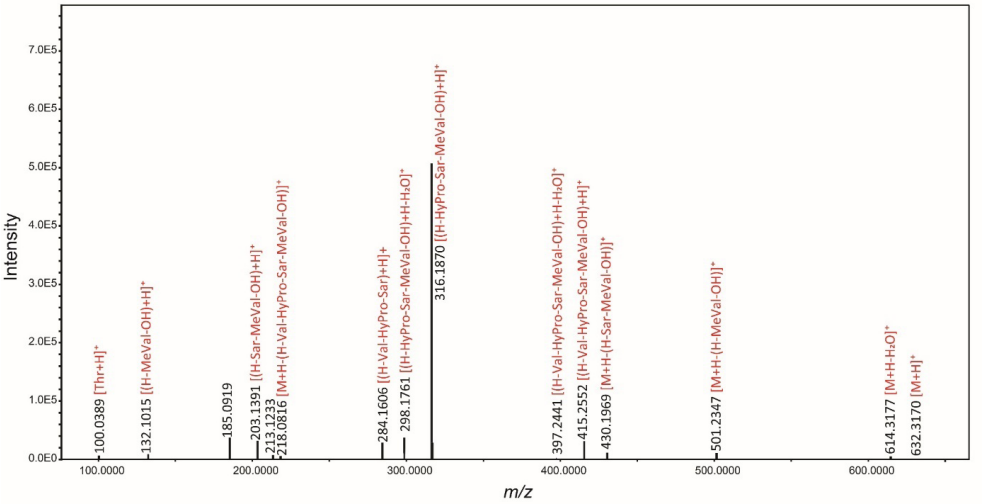


Figure S14. QTOF MS/MS spectrum of 4-MHB-containing pentapeptide lactone PPL 0.

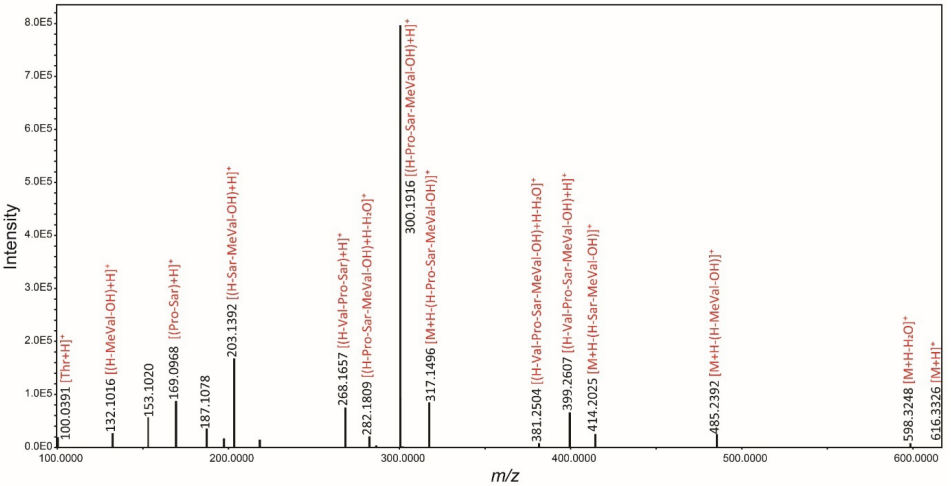


Figure S15. QTOF MS/MS spectrum of 4-MHB-containing pentapeptide lactone PPL 1.

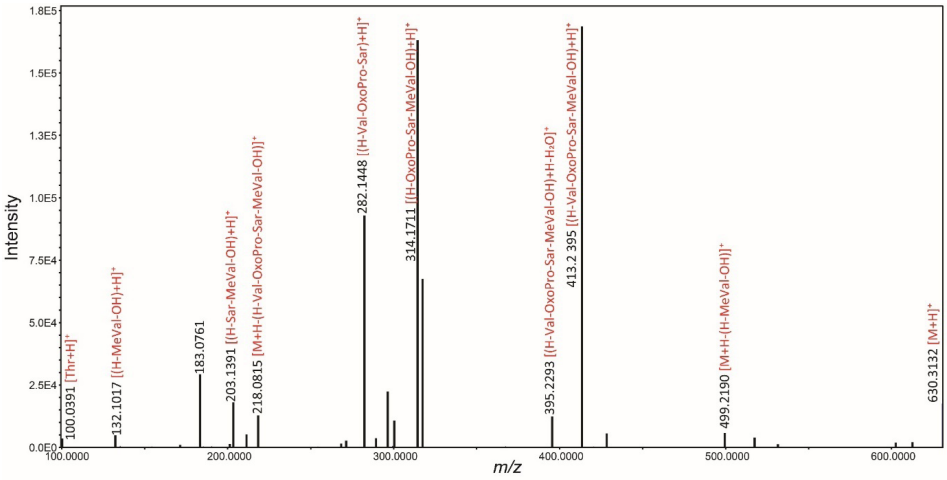


Figure S16. QTOF MS/MS spectrum of 4-MHB-containing pentapeptide lactone PPL 2.

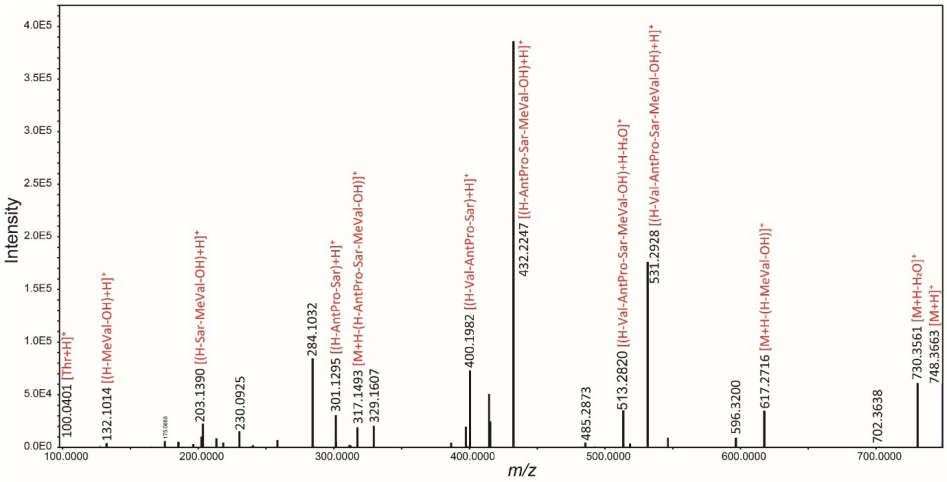


Figure S17. QTOF MS/MS spectrum of 4-MHB-containing pentapeptide lactone PPL 3.

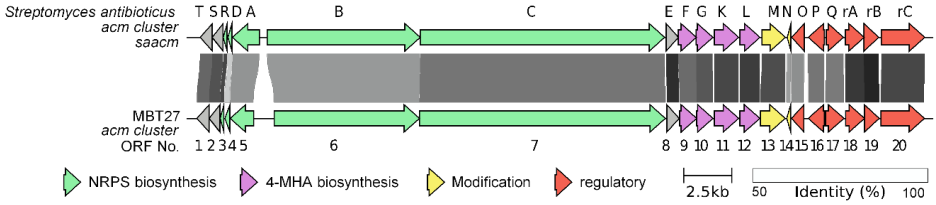


Figure S18. Alignment of actinomycin biosynthetic gene cluster from *S. antibioticus* and *Streptomyces* sp. MBT27.Grey arrows indicate genes with unknown function. Grey bars connecting homologues pairs from the two cluster, the identity of two genes is indicated by different levels of greyness.



Chapter 6

nanoRAPIDS as an analytical pipeline for the discovery of novel bioactive metabolites in complex culture extracts at the nanoscale

Isabel Nuñez Santiago*, Nataliia V. Machushynets*, Marija Mladic, Doris A. van Bergeijk, Somayah S. Elsayed, Thomas Hankemeier, Gilles P. van Wezel

This chapter was published as:
Nuñez I.S.*, Machushynets N.V.*, Mladic M., van Bergeijk D.A., Elsayed S.S., Hankemeier T., van Wezel G.P. (2024). nanoRAPIDS as an analytical pipeline for the discovery of novel bioactive metabolites in complex culture extracts at the nanoscale. *Commun. Chem.,* **7** (71).

*These authors contributed equally.

Abstract

Microbial natural products form the basis of most of the antibiotics used in the clinic. The vast majority has not yet been discovered, among others because the hidden chemical space is obscured by previously identified (and typically abundant) antibiotics in culture extracts. Efficient dereplication is therefore key to the discovery of our future medicines. Here we present an analytical platform for the efficient identification and prioritization of low abundance bioactive compounds at nanoliter scale, called nanoRAPIDS. NanoRAPIDS encompasses analytical scale separation and nanofractionation of natural extracts, followed by the bioassay of interest, automated mass spectrometry identification, and Global Natural Products Social molecular networking (GNPS) for dereplication. As little as 10 µL crude extract is fractionated into 384 fractions. First, bioactive congeners of iturins and surfactins were identified in *Bacillus,* based on their bioactivity. Subsequently, bioactive molecules were identified in an extensive network of angucyclines elicited by catechol in cultures of *Streptomyces* sp. MBT84. This allowed the discovery of a highly unusual N-acetylcysteine conjugate of saquayamycin, despite low production levels in an otherwise abundant molecular family. These data underline the utility and broad application of the technology for the prioritization of minor bioactive compounds in complex extracts.

Introduction

The increasing number of antibiotic-resistant bacteria combined with the decreasing number of new antibiotics discovered are becoming serious threats to public health. The traditional source of antibiotics are natural products produced by microorganisms, whereby some 70% are produced by Actinobacteria (Barka *et al.*, 2016, Bérdy, 2005). The canonical antibiotic discovery approach, known as the "Waksman platform" (Wright, 2014), involves high-throughput screening of Actinobacteria for antimicrobial activity, but while this approach was highly successful in the golden era of antibiotic discovery, nowadays, that approach fails to identify truly novel classes of compounds (Baltz, 2008). Still, we have likely only exploited a fraction of the chemical space of microbial natural products (Genilloud, 2014). Indeed, the molecules that are currently described (van Santen *et al.*, 2019) likely only represent 3% of the extant chemical space of natural products (Gavriilidou *et al.*, 2022). The challenge lies in the development of technologies to access the undiscovered chemical space of natural products. Genome sequencing revealed that *Streptomyces* genomes each contain up to 60 or even more biosynthetic gene clusters (BGCs) that specify natural products (Nett *et al.*, 2009). Bioinformatic tools such as antiSMASH (Blin *et al.*, 2023) allow visualizing these BGCs and predicting the classes of molecules they produce. However, many of these BGCs are expressed at a very low level under routine laboratory conditions, and hence their cognate products cannot be elucidated. Actinobacteria are usually grown in isolation and in laboratory media that do not accurately mimic the complexity of their natural habitat. The control of BGCs is most likely tightly tied to the competitive ecological conditions in which antibiotic production evolved (Chevrette *et al.*, 2020, Cornforth & Foster, 2013).

Several methods have been developed to boost the production of hitherto hidden natural products, including ribosome engineering, co-cultivation, OSMAC (one-strain many-compounds), promoter replacement and high-throughput elicitor screens (van Bergeijk *et al.*, 2020, Yoon & Nodwell, 2014). Metabolites are identified by liquid chromatography coupled to mass spectrometry (LC-MS) and NMR spectroscopy. Statistical models such as Pearson correlation, partial least squares (PLS), discriminant analysis (PCA-DA, PLS-DA, OPLS-DA) and hierarchical cluster analysis (HCA) are then used to link metabolite fingerprints to bioactivity data (Ory *et al.*, 2019). Identifying the bioactive compounds of interest is often time-consuming and expensive, while efficient dereplication is essential to avoid wasting time and resources on the rediscovery of known compounds. Global Natural Products Social (GNPS) molecular networking developed in the Dorrestein laboratory (Wang *et al.*, 2016) is an important addition to the toolbox. Although these approaches provide lists of highlighted features, the information obtained is not immediately linked to bioactivity. Several analytical modules involving different bioassays and detection technologies can be linked to allow simultaneous bioactivity evaluation and identification of compounds present in small

amounts (analytical scale) in complex compound mixtures. Such approaches include the at-line high-resolution nanofractionation, which could complement previously mentioned strategies and provide a more comprehensive insight of the compounds responsible for the activity (Mladic *et al.*, 2016, Zietek *et al.*, 2017).

One of the major challenges is the lack of truly novel chemical scaffolds, which may serve as the basis for the antibiotics of tomorrow. We discovered lugdunomycin that is produced by the soil-derived *Streptomyces* sp. QL37 (Wu *et al.*, 2019). Lugdunomycin is a highly rearranged member of the angucyclines, the largest family of type II polyketides (Fotso *et al.*, 2008, Kharel *et al.*, 2012), and has a radically different chemical scaffold as compared to other polyketides. Its discovery underlines that novel chemical space may still lie hidden even in highly explored chemical families.

To allow more efficient mining of microbial extracts for novel chemical space, we developed an analytical platform for the rapid and efficient identification and dereplication of bioactive metabolites based on nanofractionation (nanoRAPIDS). NanoRAPIDS combines at-line high-resolution nanofractionation (Mladic *et al.*, 2018) coupled to LC-MS/MS and featured-based molecular networking (FBMN) (Nothias *et al.*, 2020) in a single analysis. The implementation of Mzmine (Pluskal *et al.*, 2010) in our platform allows for automatic peak selection and correlation of the bioactivity with the appropriate mass, while the same workflow performs FBMN. The nanoRAPIDS workflow was validated by identifying a suite of iturins and surfactins produced by *Bacillus* sp. 90A-23 based on their bioactivity against *Escherichia coli*, *Bacillus subtilis* and/or *Aspergillus niger* in nanoliter fractions. We then applied the technology to prioritise bioactive angucyclines produced by *Streptomyces* sp. MBT84 following challenge by catechol. This afforded the isolation of a hitherto unknown N-acetylcysteine conjugate of saquayamycin N.

Results and Discussion

Development of the nanoRAPIDS platform

The challenge we sought to overcome is that bioactivity-guided fractionation is typically frustrated by highly abundant known antibiotics, which obscures bioactive compounds that are produced at low levels. This is particularly true for screening of small culture volumes, such as those obtained through HT cultivation. Therefore, we developed an antibiotic discovery platform based on nanofractionation (Mladic *et al.*, 2016), mass spectrometry and bioactivity assays. This set-up was combined with automated data pre-processing in MZmine 2.53 (Pluskal *et al.*, 2010) and feature-based molecular networking (Nothias *et al.*, 2020). The pipeline is designated nanoRAPIDS, for Reliable Analytical Platform for Identification and Dereplication of Specialized metabolites based on nanofractionation. The schematic overview of the proposed workflow is shown in Figure 1.

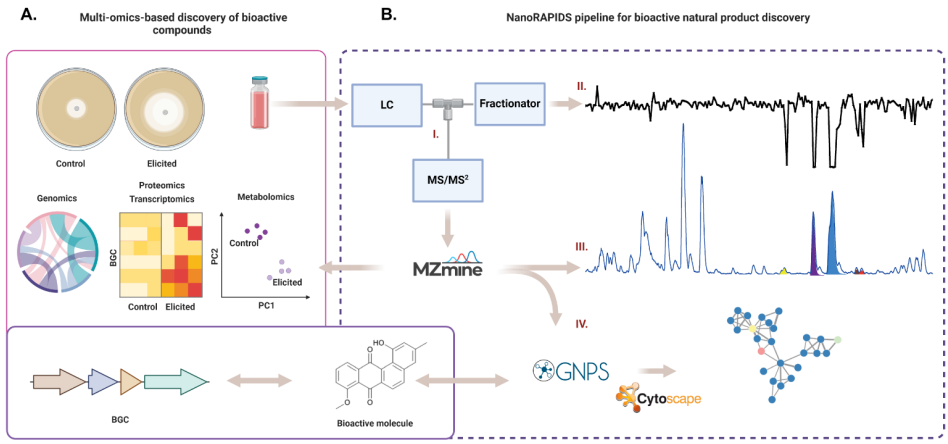


Figure 1. Schematic overview of the nanoRAPIDS platform. A. Classical omics-based pipeline for discovery of bioactive compounds. Multi-omics approaches advance the field of microbial natural products research. However, despite significant improvements in metabolite profiling, the unambiguous identification of each individual NP in an extract using generic methods remains challenging. **B.** Integration of the nanoRAPIDS pipeline into the multi-omics workflow allows for detection and dereplication of the compounds contributing to the bioactivity of the complex extract. The nanoRAPIDS pipeline starts with at-line nanofractionation (I) and parallel MS/MS², followed by bioactivity assays of the individual nanofractions through the resazurin reduction assay (II) to determine their retention time (RT). The MS/MS² data are automatically pre-processed in MZmine to determine the *m/z* values of the bioactive features (III). The peak list is then subjected to GNPS molecular networking and dereplication. Molecular network visualization and automatic bioactive features mapping is conducted in Cytoscape (IV).

The nanoRAPIDS pipeline makes use of an at-line nanofractionation setup (Mladic *et al.*, 2018). First, extracts are separated by liquid chromatography (LC) and after the post-column split, the largest part of the flow is fractionated using high-resolution nanofractionation, while a small part is subjected to mass spectrometry (MS and MS/MS), allowing the

parallel data collection of all fractions. Next, the individual fractions are tested against target microorganisms such as Gram-positive, Gram-negative bacteria and/or fungi with a resazurin reduction assay. Since the at-line nanofractionation is performed at a 6 sec resolution, the retention time of eluting compounds from the LC separation is retained in the bioactivity chromatograms. After at-line nanofractionation, the LC-MS/MS spectral data are automatically processed with Mzmine (Pluskal *et al.*, 2010), which gives the information on the *m/z* values of the bioactive fraction(s) constituents and their MS/MS spectra. Pre-processed MS^2 data are subsequently uploaded onto the GNPS web-platform (Wang *et al.*, 2016) for feature-based molecular networking (Nothias *et al.*, 2020). Molecular networking is then performed to visualize all ions of molecules detected and fragmented during an MS experiment and identify the chemical relationships between them in a semi-quantitative fashion. Additionally, GNPS automatically performs a spectral library search for known molecules in the molecular networks, if their MS/MS spectra are available in public MS/MS spectral libraries. The final step of the analysis consists of mapping of the bioactive compounds within the obtained network in Cytoscape, based on their *m/z* and retention time (Figure 1B).

Validation of nanoRAPIDS by screening *Bacillus* sp. 90A-23 extracts

Bacillus sp. 90A-23 produces several bioactive compounds with broad spectrum activity, many of which are well characterised, including iturins and surfactins that both consist of families of molecules. To validate the platform, crude extracts were subjected to analysis using nanoRAPIDS (Figure 2). Automated feature detection and correlation guided by high-resolution bioactivity assays identified eight mass features with *m/z* 1029.5404 at 16.32 min, *m/z* 1043.5591 at 18.05 min, *m/z* 1057.5642 at 19.22 min, *m/z* 1071.5880 at 20.64 min, *m/z* 1071.5880 at 20.95 min, *m/z* 1008.6698 at 27.91 min, *m/z* 1022.6782 at 28.84 min and *m/z* 1036.6921 at 29.47 min (Table 1). It is important to note that some compounds may remain undetected due to the limitation of the ionization method from the MS system.

The filtered peak list generated in MZmine was exported for GNPS molecular networking (Nothias *et al.*, 2020). The network consisted of 852 nodes, which were clustered in 92 spectral families and 134 individual nodes. The bioactive compounds with *m/z* 1029.5404 at 16.32, *m/z* 1043.5591 at 18.05 min, *m/z* 1057.5642 at 19.22 min, *m/z* 1071.5880 at 20.64 min, *m/z* 1071.5880 at 20.95 min were found in a 19-node cluster (Figure 2C). Search each MS/MS spectrum against the GNPS spectral libraries resulted in the annotation of *m/z* 1043.5591, *m/z* 1057.5642, *m/z* 1071.5880 as iturin A2, A3/4/5 and A6/7, respectively. Annotation of these compounds was supported by direct comparison of the MS/MS spectra with the standards of iturins from GNPS by mirror plot using mMass software (Strohalm *et al.*, 2010) (Figures S2–S4).

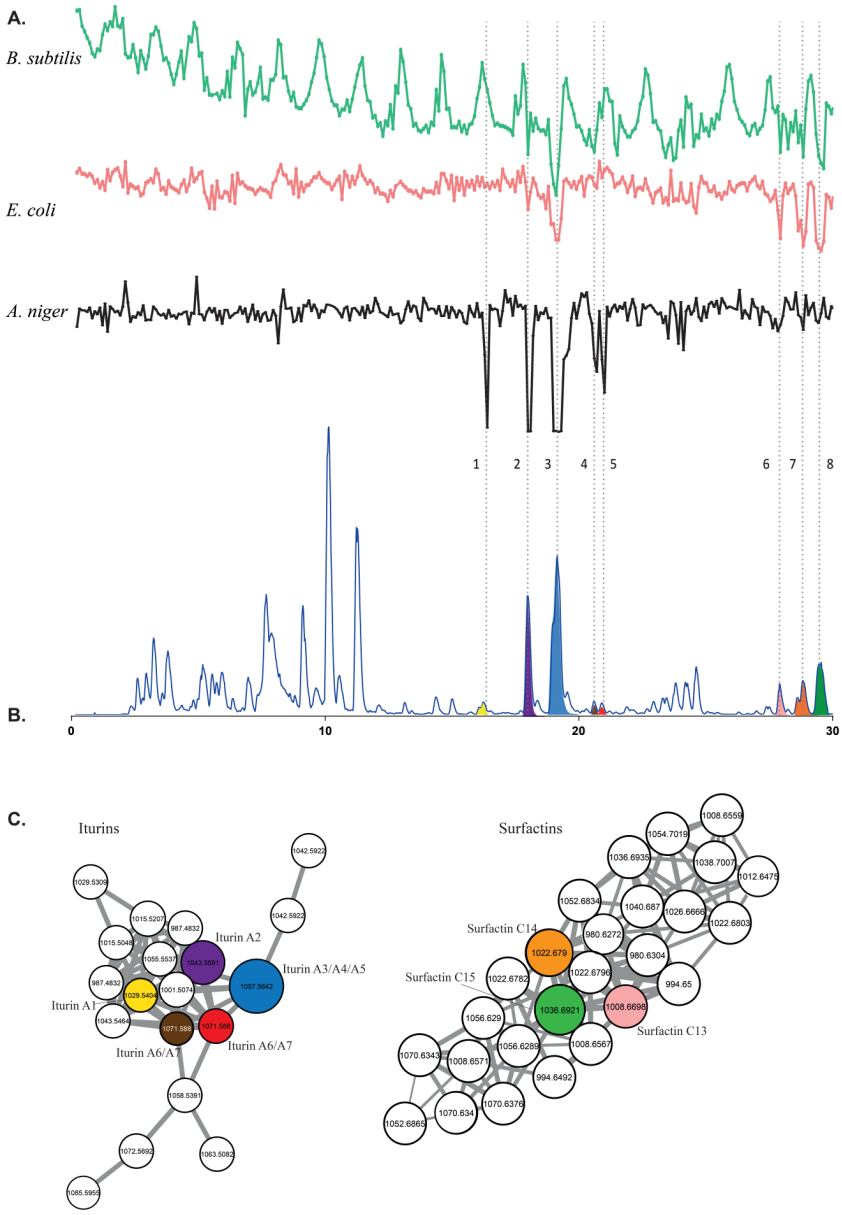


Figure 2. NanoRAPIDS screen of extract of *Bacillus* sp. 90A-23. A. Chromatograms representing the results of the resazurin reduction assay against *E. coli*, *B. subtilis* and *A. niger* for each nanofraction plotted against time the nanofraction was collected. Fractions were collected with 6 seconds resolution onto 384-well plates after 10 µL injection of crude extract obtained from *Bacillus* sp. 90A-23, at 15 mg/mL concentrations. **B.** The LC-MS data presented as Base Peak Chromatogram (BPC) with extracted ion currents of the *m/z* values found in the MS spectra corresponding to the bioactive peaks. **C.** Subfamilies containing the bioactive iturins (left) and surfactins (right) as identified by GNPS-molecular networking of culture extracts. Colors of the nodes correspond to the colors of the chromatographic peaks and highlight the mass features detected in the bioactive nanofractions. The node size represents the relative abundance of ion. The edge thickness attribute is an interpretation of cosine similarity score, with thicker lines indicating higher similarity.

The compound with *m/z* 1029.5404 at 16.32 min was annotated as iturin A1 using the Antibase database (Laatsch, 2011) and manual comparison of MS/MS spectra with those in the literature (Pathak & Keharia, 2014). These amphiphilic compounds are characterized by a peptide ring of seven amino acid residues, closed by a β-amino fatty acid (Figure 3).

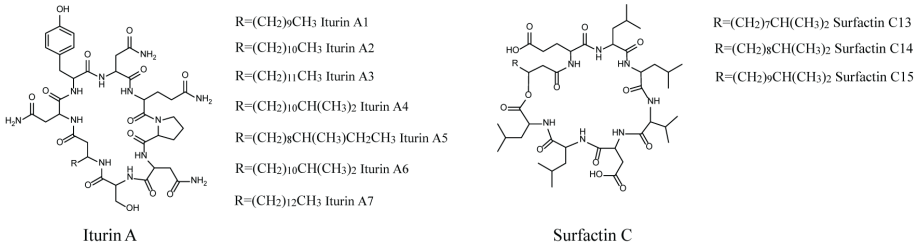


Figure 3. Bioactive compounds found in the extracts of *Bacillus* sp. 90A-23. Structures of iturins and surfactins detected in the extracts of *Bacillus* sp. 90A-23. R, site where the fatty acid chain is linked to the peptidic backbone.

Iturins are a family of lipopeptides produced by various Bacilli species (Sood *et al.*, 2014, Chung *et al.*, 2000, Hertlein *et al.*, 2016). The 14 Da difference between molecules with *m/z* 1029.5404, *m/z* 1043.5591, *m/z* 1057.5642 and *m/z* 1071.5880 is explained by variation in length of the fatty acid chain, containing 13, 14, 15 or 16 carbons, respectively (Figure S5). The major peaks in the bioactivity chromatogram at 18.05 and 19.22 min corresponded to the most intense compounds, which were dereplicated as iturin A2 (*m/z* 1043.5591) and iturin A3/A4 (*m/z* 1057.5642). Moreover, 12 more mass spectral features were also identified as iturins based on analog searches. The bioactive features with *m/z* 1008.6698 at 27.91 min, *m/z* 1022.6782 at 28.84 min and *m/z* 1036.6921 at 29.47 min were detected within the surfactin spectral family (Figure 2C).

Table 1. Bioactive compounds identified in extracts of *Bacillus* sp. 90A-23.

Peak	Mass (*m/z*)	RT, min	Molecular formula	Identified compound
1	1029.5404	16.32	$C_{47}H_{72}O_{14}N_{12}$	(C_{13}) Iturin A1 (Pathak & Keharia, 2014)
2	1043.5591	18.05	$C_{48}H_{74}O_{14}N_{12}$	(C_{14}) Iturin A2 (Ye *et al.*, 2012)
3	1057.5642	19.22	$C_{49}H_{76}O_{14}N_{12}$	(C_{15}) Iturin A3/A4/A5 (Pathak & Keharia, 2014)
4	1071.5880	20.64	$C_{50}H_{78}O_{14}N_{12}$	(C_{16}) Iturin A6/A7 (Caldeira *et al.*, 2011)
5	1071.5880	20.95	$C_{50}H_{78}O_{14}N_{12}$	(C_{16}) Iturin A6/A7 (Caldeira *et al.*, 2011)
6	1008.6698	27.91	$C_{51}H_{89}N_7O_{13}$	Surfactin Leu/Ile7 C13 (Pathak & Keharia, 2014, Ma *et al.*, 2016, Nanjundan *et al.*, 2019)

7	1022.6782	28.84	$C_{52}H_{91}N_7O_{13}$	Surfactin Leu/Ile7 C14 (Pathak & Keharia, 2014, Ma *et al.*, 2016, Nanjundan *et al.*, 2019)
8	1036.6921	29.47	$C_{53}H_{93}N_7O_{13}$	Surfactin Leu/Ile7 C15 (Pathak & Keharia, 2014, Ma *et al.*, 2016, Nanjundan *et al.*, 2019)

The molecule with *m/z* 1008.6698 was annotated as surfactin Leu/Ile7 C13, that with *m/z* 1022.6782 as surfactin Leu/Ile7 C14 and the compound with *m/z* 1036.6921 as surfactin Leu/Ile7 C15 (Ma *et al.*, 2016) (Figures S6–S9). An additional 23 other spectral features also belonged to the same spectral family.

In the initial screen, *Bacillus* sp. 90A-23 showed bioactivity against both bacteria and fungi. NanoRAPIDS allowed us to readily establish which detected individual mass features contributed to the bioactive nanofraction. This showed that the initially detected antifungal and antibacterial activity of the crude extract was mostly caused by iturins and surfactins (Table 1, Figure 2).

Identification of a *N*-acetylcysteine conjugate in a complex network of angucyclines in *Streptomyces* sp. MBT84

Streptomycetes typically produce a large variety of angucyclines, which are the largest family of type II polyketides, many of which with significant bioactivity as antibiotics and/or anticancer agents (Kharel *et al.*, 2012). We, therefore, identified this molecular family as an ideal challenge for nanoRAPIDS, seeking to identify minor molecules with bioactivity without prior pre-purification. We recently showed that there is still a large chemical space that remains to be elucidated (van Bergeijk *et al.*, 2022, Wu *et al.*, 2019). *Streptomyces* sp. MBT84 produces a particularly large number of angucyclines and derivatives, many of which are produced specifically when catechol is added as elicitor to the culture media (van Bergeijk *et al.*, 2022). To identify the nature of the metabolites produced by *Streptomyces* sp. MBT84 in response to catechol, the strain was streaked on MM agar plates and the metabolites were then extracted from the spent agar after five days of growth using ethyl acetate (EtOAc). The crude extracts of the catechol-grown cultures showed increased bioactivity against *B. subtilis* 168, as compared to those grown under control conditions (van Bergeijk *et al.*, 2022). To identify which mass features correlated to the increased bioactivity of *Streptomyces* sp. MBT84, crude extracts were screened with the nanoRAPIDS platform (Figure 4).

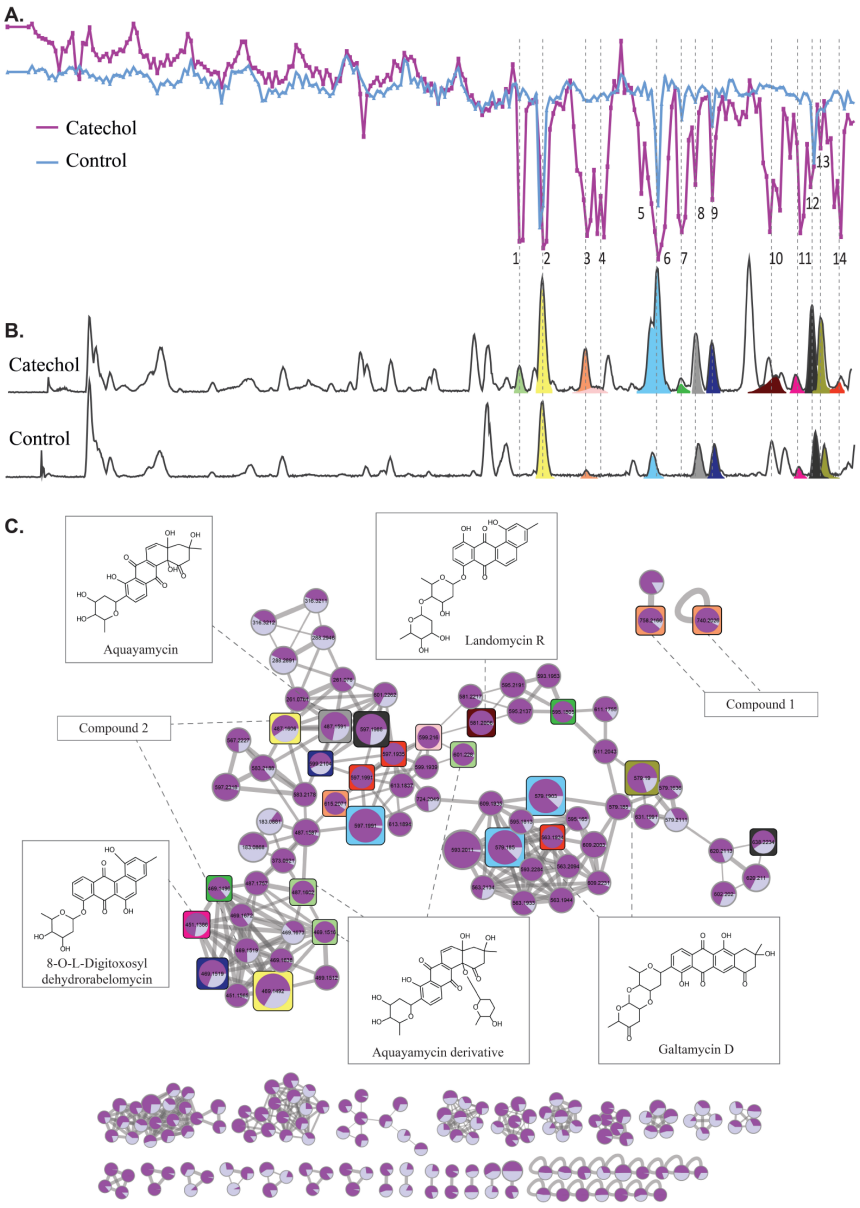


Figure 4. NanoRAPIDS screen of extracts of *Streptomyces* sp. MBT84 in the presence and absence of catechol. **A.** Resazurin reduction assay against *B. subtilis* 168 for each nanofraction plotted against retention time. Fractions were collected with 6 seconds resolution onto 384-well plates after 10 µL injection of crude extract obtained from *Streptomyces* sp. MBT84 with catechol (purple) or without (blue), at 4 mg/mL concentrations. **B.** LC-MS data presented as base peak chromatogram (BPC) with extracted ion currents of the *m/z* values found in the MS spectra corresponding to the bioactive peaks produced in the presence (top) and absence (bottom) of catechol. **C.** Molecular network representing the bioactive molecules identified by GNPS-molecular networking and MolNetEnhancer in the extracts. The colors of the squares correspond to the colors of the chromatographic peaks and highlight the mass features detected in the bioactive nanofractions. The node size represents the relative abundance of the ion, in dark purple withing the catechol-grown sample and in light purple the control. The edge thickness attribute is an interpretation of cosine similarity score, with thicker lines indicating higher similarity.

Following nanofractionation, the chromatogram showing the bioactivity against *B. subtilis* 168 showed 14 peaks with negative maxima (Figure 4A). Automated peak detection combined with antimicrobial activity assays allowed the selection of the individual bioactive features (Figure 4B), which were mapped onto the enhanced feature-based molecular network. Combining the output from molecular networking, MS2LDA and in silico annotation, the MolNetEnhancer (Ernst *et al.*, 2019) workflow predicted an angucycline molecular family (Figure 4C). The bioactive angucyclines were annotated by manual comparison of the molecular formula and monoisotopic mass using the Dictionary of Natural Products (Table 2). Notably, all selected features were more abundant in the extracts from cultures supplemented with catechol.

Multiple mass features were detected in the bioactive nanofractions, mostly due to in-source fragmentation (Table 2). For example, for peak 1 three mass features were selected, namely, *m/z* 601.228, *m/z* 487.1601 and *m/z* 469.1515 that correspond to the mass of protonated aquayamycin derivative, in-source sugar loss and subsequent water loss. Two of the 14 bioactivity peaks, namely 3 and 6, were caused by co-elution of two different compounds. To our knowledge, the compounds with *m/z* 758.2166 and *m/z* 615.2071, responsible for the bioactivity of peak 3, have not been previously described. Interestingly, the mass feature with *m/z* 758.2166 did not cluster with the angucycline spectral family and was classified as an aminocyclitol glycoside based on the enhanced molecular network annotation (Figure 4C).

Table 2. List of mass features detected in the bioactive nanofractions of *Streptomyces* sp. MBT84 extract.

Peak	Mass (*m/z*)	Quasi molecular ion	Molecular formula	Compound annotation
1	601.228	$[M+H]^+$	$C_{31}H_{37}O_{12}$	Aquayamycin derivative (Rohr *et al.*, 1993)
	487.1601	$[M+H\text{-}C_6H_{10}O_2]^+$	$C_{25}H_{27}O_{10}$	
	469.1515	$[M+H\text{-}C_6H_{10}O_2\text{-}H_2O]^+$	$C_{25}H_{25}O_9$	
2	487.1606	$[M+H]^+$	$C_{25}H_{27}O_{10}$	Aquayamycin/fridamycin A (Sezaki *et al.*, 1968)
	469.1492	$[M+H\text{-}H_2O]^+$	$C_{25}H_{25}O_9$	
3	758.2166	$[M+H]^+$	$C_{36}H_{40}NO_{15}S$	Unknown
	740.2026	$[M+H\text{-}H_2O]^+$	$C_{36}H_{38}NO_{14}S$	
	615.2071	$[M+H]^+$	$C_{31}H_{35}O_{13}$	Unknown
4	599.216	$[M+H]^+$	$C_{31}H_{35}O_{12}$	Saquayamycin C1/landomycin D/grincamycin K (Uchida *et al.*, 1985, Akhter *et al.*, 2018, Lai *et al.*, 2018)
5	NA			

6	597.199	$[M+H]^+$	$C_{31}H_{33}O_{12}$	Fridamycin D/saquayamycin A1/saquayamycin B1 (Maskey *et al.*, 2003, Uchida *et al.*, 1985)
	579.1902	$[M+H-H_2O]^+$	$C_{31}H_{31}O_{11}$	
	579.1849	$[M+H]^+$	$C_{31}H_{31}O_{11}$	Galtamycin D (van Bergeijk *et al.*, 2022)
7	595.1804	$[M+H]^+$	NA	Unknown angucycline-like compound
	469.1495	$[M+H]^+$	$C_{25}H_{25}O_9$	Antibiotic PD 116740 derivative/atramycin A/ landomycin I/urdamycin B derivative (Zheng *et al.*, 2021, Erb *et al.*, 2009, Rohr & Zeeck, 1987)
8	487.159	$[M+H]^+$	$C_{25}H_{27}O_{10}$	Aquayamycin/fridamycin A (Sezaki *et al.*, 1968, Krohn & Baltus, 1988)
9	599.2103	$[M+H]^+$	$C_{31}H_{35}O_{12}$	Saquayamycin C1/landomycin D/grincamycin K (Uchida *et al.*, 1985, Henkel *et al.*, 1990, Lai *et al.*, 2018)
	469.1518	$[M+H-C_6H_{10}O_3]^+$	$C_{25}H_{25}O_9$	
10	581.2005	$[M+H]^+$	$C_{31}H_{33}O_{11}$	Landomycin R (Shaaban *et al.*, 2011)
11	451.1385	$[M+H]^+$	$C_{25}H_{23}O_8$	8-O-L-Digitoxosyldehydrorabelomycin (Rix *et al.*, 2005)
12	597.1988	$[M+H]^+$	$C_{31}H_{33}O_{12}$	Fridamycin D/saquayamycin A1/saquayamycin B_1 (Maskey *et al.*, 2003, Uchida *et al.*, 1985)
	638.2234	$[M+H+MeCN]^+$	$C_{33}H_{36}NO_{12}$	
13	579.19	$[M+H]^+$	$C_{31}H_{31}O_{11}$	Galtamycin D (van Bergeijk *et al.*, 2022)
14	597.1935	$[M+H]^+$	$C_{31}H_{33}O_{12}$	Fridamycin D/saquayamycin A1/saquayamycin B_1 (Maskey *et al.*, 2003, Uchida *et al.*, 1985)
	597.199	$[M+H]^+$	$C_{31}H_{33}O_{12}$	
	563.1933	$[M+H-H_2O-O]^+$	$C_{31}H_{31}O_{10}$	

Large-scale fermentation was performed to elucidate the structure of this compound. *Streptomyces* sp. MBT84 was grown confluently on 12 x 12 mm MM agar plates with 50 µM catechol, and the metabolites were extracted with EtOAc. Following chromatographic separation, **1** (red amorphous powder) and **2** (yellow amorphous powder) were isolated. Compound **1** had a molecular formula of $C_{36}H_{39}NO_{15}S$, with 18 degrees of unsaturation based on HRESIMS (high resolution-electrospray ionization-mass spectrometry) analysis. Detailed

NMR analysis showed that **1** is a derivative of saquayamycin B_1 (Figures S10–S13) (Uchida *et al.*, 1985). This derivative was named saquayamycin N (Figure 5). The derivatization is at C-5, which showed in the NMR a signal for a tetrasubstituted carbon (δ_c 162.6) instead of a methine group in saquayamycin B_1. This was established through the HMBC correlations from H_2-4 and H-6, which is now a singlet proton (δ_H 6.65, s), to C-4a and C-5 (Figure S14). An *N*-acetylcysteine moiety was identified as the remaining part of **1** based on the observed remaining NMR signals and correlations, together with the established molecular formula. The *N*-acetylcysteine was connected to C-5 of the saquayamycin B core through the sulphur atom based on the HMBC and NOESY correlations observed from the protons of the methylene group in *N*-acetylcysteine to C-5 and H-6, respectively (Figures S14–S15,S22). The absolute configuration of compound **1** could not be determined due to the limited amount of material purified. The configuration is most likely 3*R*, 4a*S*, 12b*S*, as was consistently observed for all saquayamycins identified so far (Uchida *et al.*, 1985, Shaaban *et al.*, 2012a). For the *N*-acetylcysteine moiety, an L-cysteine would be expected. The molecular formula of compound **2** was established to be $C_{25}H_{26}O_{10}$, with 13 degrees of unsaturation, based on HRESIMS analysis. 1D and 2D NMR analysis (Figures S16–S22) revealed that **2** is the previously known fridamycin A (Krohn & Baltus, 1988).

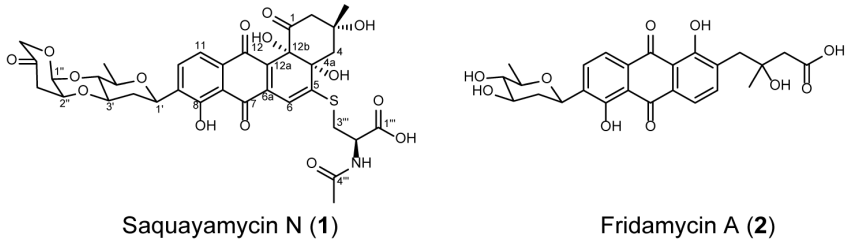


Figure 5. Structures of saquayamycin N (1) and fridamycin (2) isolated from *Streptomyces* sp. MBT84. Compound **1** was, to the best of our knowledge, previously unreported and was named as saquayamycin N. The structure of the isolated compound **2** corresponds to fridamycin A.

The isolated saquayamycin N and fridamycin A were tested for antibacterial activity. Compounds **1** and **2** showed moderate activity against *B. subtilis* 168, with an MIC of 125 µg/mL, indicating low bioactivity. The observed bioactivity for **1** and **2** aligned with the data obtained in the resazurin reduction assay and corroborated the contribution of saquayamycin N and fridamycin A to the bioactivity observed in the peaks 2 and 3, respectively (Figure 4). The resazurin reduction assay in the nanofractionation setup only has a short incubation time and this is radically different from—and more sensitive then—the MIC assays which are based on overnight cultures. To validate the contribution of saquayamycin N and fridamycin A to the activity detected in the crude extracts, the purified compounds were pipetted in a 384 well plate and the resazurin reduction assay was performed as previously described. The selected

concentrations were based on the MIC values, namely 125 µg/mL (MIC value), 12.5 µg/mL and 1.25 µg/mL concentrations (sub-MIC values). The MIC and sub-MIC concentration of 12.5 µg/mL resulted in peaks with significant negative maxima for both saquayamycin N and fridamycin A (Figure S23). The potential of these two isolated molecules should be addressed by further investigation of their anticancer activity, as the compounds from the angucycline family are well-known for their anticancer properties (Korynevska *et al.*, 2007).

NanoRAPIDS is a pipeline based on at-line nanofractionation technology, which allows to establish which enables the detection of individual mass features contribute to the bioactivity of nanofractions. If multiple fractions with bioactivity are identified, GNPS will allow their linkage if indeed they would belong to the same (known or unknown) family. Thus, the nanoRAPIDS platform has the potential to go beyond the discovery of known specialized metabolites and facilitates the identification of new natural product families. Nevertheless, the discovery of novel bioactive natural products remains a substantial challenge in the field of natural product discovery.

The nanoRAPIDS platform is an efficient and cost-effective technology based on nanofractionation, which allows the identification of low-abundance molecules, and in particular antibiotics, in small culture extracts against a background of abundantly present known antibiotics. The powerful combination of analytical LC-MS/MS, bioassays, open bioinformatic tools, including MZmine, GNPS Web-platform and Cytoscape, creates an attractive pipeline for natural product discovery. Only minimal amounts of crude extract are required for LC-MS/MS injection, allowing a significant reduction of fermentation volumes. The analysis of a single sample using the nanoRAPIDS platform requires approximately 24 hours. Therefore, multiple extracts obtained from a range of culturing conditions can be rapidly tested for bioactive compounds and dereplicated, which is particularly powerful when combined with elicitation approaches in antibiotic discovery (van Bergeijk *et al.*, 2020, Yoon & Nodwell, 2014). Proof of concept for the technology was obtained by identification of the bioactive members of the iturin and surfactin molecular families in extracts of *Bacillus* sp. 90A-23, and then successfully applied to discover the angucycline saquayamycin N with a unique cysteine moiety in extracts of *Streptomyces* sp. MBT84. The method is versatile and can be readily applied to screen nanoscale-level extracts or biological mixtures against a range of indicator strains.

Materials and Methods

Bacterial strains and culturing conditions

Bacillus sp. 90A-23 was obtained from the Auburn strain collection and had previously been isolated from the peanut rhizosphere in Tallassee, Alabama, USA. The strain was fermented in 100 mL Erlenmeyer flasks with 30 mL of tryptic soy broth (TSB) at 30 °C at 220 rpm for 3 days. *Streptomyces* sp. MBT84 (van Bergeijk *et al.*, 2022) was obtained from the Leiden University strain collection and had been isolated from soil samples collected from the Qinling mountains (Shanxi province, China) (Zhu *et al.*, 2014a). *Streptomyces* sp. MBT84 was grown confluently on minimal media (MM) agar plates with and without 100 µM catechol for five days at 30 °C. *Escherichia coli* strain AS19-RlmA (Liu & Douthwaite, 2002), *Bacillus subtilis* 168 (Barbe *et al.*, 2009)2009 and *Aspergillus niger* N402 (Wu *et al.*, 2015b) were used as indicator strains for bioactivity assays. The bacterial strains were grown on Luria-Bertani (LB) agar at 37 °C for 24 h, *A. niger* was propagated on complete medium (CM) agar at 25°C for 48 h.

Metabolite extraction

Following fermentation of *Bacillus* sp. 90A-23, 1.5 g of Diaion® HP-20 (Resindion, Mitsubishi) was added to the cultures and stirred overnight at 4°C. The resin was collected by filtration, washed with distilled water, resuspended in 30 mL methanol and maintained at 4°C for 4 h. The resulting methanol fraction was collected and concentrated under reduced pressure to a final concentration of 15 mg/mL. *Streptomyces* sp. MBT84 was grown on MM agar plates for five days, after which the agar was cut into small pieces and soaked overnight in ethyl acetate (EtOAc) (van Bergeijk *et al.*, 2022). The solvent was evaporated at room temperature and the dry extracts were dissolved in MeOH to a final concentration of 4 mg/mL.

Liquid chromatography, at-line nanofractionation and mass spectrometry

Liquid chromatography separation, subsequent at-line nanofractionation and parallel mass spectrometry analysis were performed in an automated fashion. For LC-MS analyses, 10 µL were injected into Waters Acquity UPLC system equipped with XBridge Peptide BEH C_{18} column (5 µm, 300 Å, 4.6×100 mm). The flow rate used was 0.6 mL/min. Solvent A was 0.1% formic acid in water; solvent B was 0.1% formic acid in ACN. The gradient started with 100% A, and was from the start linearly increased to 50% B over 20 min, and then linearly increased to 90% B over 5 min, and then held at 90% for 5 min, and then increased to 100% B over 1 min and held at 100% B for 10 min. After the column, the flow was split at a 1:9 ratio. The smaller fraction was sent to the Waters Acquity photodiode array detector followed by a Thermo Instruments MS system (LTQ Orbitrap XL, Bremen, Germany) equipped with an electrospray ionization source (ESI). As for the MS, the following ESI parameters were used (Machushynets *et al.*, 2022): capillary voltage 5 V, spray voltage 3.5 kV, capillary temperature

300 °C, auxiliary gas flow rate 10 arbitrary units, and sheath gas flow rate 50 arbitrary units. Full MS spectra were acquired with the Orbitrap in positive mode at a mass range of 100–2000 *m/z*, and FT resolution of 30,000. Data-dependent MS^2 spectra were acquired in the ion trap for the three most intense ions using collision-induced dissociation (CID). The larger fraction was sent to the chip-based nano-electrospray ionization source/fractionation robot (NanoMate Triversa, Advion BioSciences). The nanofractions were collected every six seconds onto black 384-well plates (Greiner Bio One, Alphen aan den Rijn, The Netherlands) in a serpentine pattern. Each sample was collected into 350 wells of the 384-well plate. After the nanofractionation, the plates were dried under vacuum. In order to determine the delay between the MS trace and the bioassay trace, nalidixic acid was dissolved in distilled water to a final concentration of 0.5 mg/mL, and 1 µL was fractionated (Figure S1).

Resazurin reduction assay

The antibacterial and antifungal activities of the nanofractionated *Bacillus* sp. 90A-23 extract were determined in a resazurin reduction assay. Indicator strains *B. subtilis* 168 and *E. coli* ASD19 were cultured in liquid LB medium at 37°C to exponential phase and diluted in LB to the final OD_{600} of 0.005. Resazurin solution was then added to the cells to a final concentration of 15 µg/mL and 25 µL of this mixture were directly pipetted into each well of 384-well plates containing dried nanofractions. Controls were pipetted in the last column of the plate: positive control (cells with resazurin and ampicillin (0.2 mg/mL)), growth control (cells and resazurin), blank (LB with resazurin), and sterility control (only LB). Directly after this, the plates were spun down for 10 s at 1500 rpm and subsequently incubated without the lids in the plastic bags in a humidified incubator at 37°C for 2-3 h (Mladic *et al.*, 2018).

Spores of *A. niger* N402 were harvested from CM agar plates, pregerminated in complete medium for 5 h at 30 °C and diluted in CM to the final titer of 1 x 10^5 spores/mL. Each well of 384-well plate with dried nanofraction was filled with 25 µL of *A. niger* spore solution. Amphotericin B (0.02 mg/mL) was added to the *A. niger* spores in the positive control wells. Subsequently, the plates were centrifuged and incubated at 30 °C overnight as explained above. Finally, resazurin was added to each well to a final concentration of 60 µg/mL and incubated for 45 min at 37 °C. After the incubation, the fluorescence was measured as a single point measurement using a Victor3 Plate Reader (PerkinElmer, Inc, Waltham, MA, USA) at 550 nm excitation and 590 nm emission wavelengths. The fluorescent readout was normalized by dividing each value measured with the median of all the values obtained in a single measurement. Subsequently, the bioactivity chromatograms were plotted in GraphPad Prism 8 software (La Jolla, CA, USA) in a graph showing the normalized response of each nanofraction versus time at which each nanofraction was collected.

Alignment of chromatograms

Alignment of the different chromatographic plots was conducted through a semi-automated process. Initially, the delay introduced by the tubing setup was determined by the injection

of 0.5 µg of nalidixic acid in the nanoRAPIDS setup and *E. coli* ASD19 was used as indicator strain (Figure S1). Once the delay was quantified in 0.3 minutes, and the bioactivity chromatogram was adjusted based on this determined delay. Subsequently, both the mass spectrometry (MS) chromatogram, utilizing automatic peak detection, and the bioactivity chromatogram were standardized to the same size. Given the identical time function, the chromatograms were then aligned to the center, ensuring accurate synchronization of the two datasets. This semi-automated alignment procedure aimed to enhance precision in data interpretation and visualization, facilitating meaningful comparisons between the MS and bioactivity chromatograms.

Feature-based GNPS networking

A molecular network was created with the feature-based molecular networking workflow (Nothias *et al.*, 2020) on Global Natural Product Social molecular networking (GNPS) (Wang *et al.*, 2016). Prior to feature-based GNPS networking, the MS files in mzXML format were imported into MZmine 2.53 (Pluskal *et al.*, 2010) for data processing. Mass ion peaks were detected using the centroid algorithm with a noise level set to 1.0×10^4 for MS scans and 0 for MS^2 scans. Afterward, chromatograms were built for the detected masses with a minimum group size of 10, *m/z* tolerance of 0.001 *m/z*, group intensity threshold of 1.0×10^4 and minimum height of 5.0×10^4. Chromatogram deconvolution was then performed using local minimum search algorithm (search minimum in RT range 0.1 min, chromatographic threshold 90%, minimum relative height 1%, minimum absolute height 1.0×10^4, minimum ratio of peak top/edge two and peak duration range 0.05–3 min). A feature with its MS^2 scans was paired using 0.05 *m/z* range and a retention time range of 1 min. In the generated peak lists, isotopes were identified using isotopic peaks grouper (*m/z* tolerance 0.001 *m/z* and retention time tolerance 0.1 min). Duplicate peaks were filtered using single feature and old average mode by deleting the duplicate rows with *m/z* tolerance of 0.001 *m/z* in RT range 0.05 min and with *m/z* tolerance of 1 *m/z* in RT range 0.05 min, respectively. The resulting peak list was filtered with a rows filter to keep only the peaks with MS^2 scans. The filtered peak list was exported to the feature quantification table and the MS^2 spectral summary file for feature-based molecular networking. On GNPS the precursor ion mass tolerance was set to 0.02 Da and the MS/MS fragment ion tolerance to 0.9 Da. A molecular network was then created where edges were filtered to have a cosine score above 0.7 and more than 3 matched peaks. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 3 matched peaks. To enhance chemical structural information within the molecular network, information from *in silico* structure annotations from GNPS Library Search, Network Annotation Propagation (da Silva *et al.*, 2018), Dereplicator (Mohimani *et al.*, 2017) were incorporated into the network using the GNPS MolNetEnhancer workflow (Ernst *et al.*, 2019). Chemical class annotations were performed using the ClassyFire chemical ontology (Djoumbou Feunang *et al.*, 2016). The molecular networks were visualized using

Cytoscape software (Shannon *et al.*, 2003). The molecular networking job from the extract of *Bacillus* sp. 90A-23 can be publicly accessed at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=1ba65a809b17413694522c27b0d952df. The molecular networking job from the extract of *Streptomyces* sp. MBT84 can be publicly accessed at https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=15e986d4356f41b4962baab398c571fa. LC-MS/MS data were deposited in the MassIVE Public GNPS data set (MSV000094019).

Large-scale fermentation, extraction and fractionation

Large-scale fermentation of *Streptomyces* sp. MBT84, extraction and fractionation were performed as previously described (van Bergeijk *et al.*, 2022). Briefly, *Streptomyces* sp. MBT84 was grown on solid MM supplemented with 50 µM catechol, 1% glycerol and 0.5% mannitol at 30 °C for five days. Agar plates were extracted with EtOAc and 1.6 g of crude extract were obtained. This extract was adsorbed onto 1.6 g silica gel (pore size 60 Å, 70–230 mesh, Sigma Aldrich), and loaded on a silica column, followed by gradient elution using mixtures of *n*-hexane, EtOAc, and MeOH. The fractions that eluted with 50% MeOH: 50% EtOAc were combined with the fraction that eluted with 100% MeOH, and reconstituted in 50% ACN: 50% H_2O. The resulting fraction was subjected to a SunFire C_{18} column (10 µm, 100 Å, 19 × 150 mm) and eluted with H_2O (solvent A) and ACN (solvent B) with 0.1% formic acid in a gradient of solvent B 30–70% in 30 min, at a flow rate of 15 mL/min. Fractions A, B and C were collected every 10 min. The fraction A was further purified on semi-preparative SunFire C_{18} column (5 µm, 100 Å, 10 × 250 mm), run at 3 mL/min, and eluted using a H_2O and ACN with 0.1% formic acid gradient of 47–50% in 20 min, to yield compound **1** (0.8 mg) and compound **2** (2.6 mg). NMR data were acquired on Bruker Ascend 850 NMR spectrometer (850 MHz) and interpreted using the MestreNova V.14 software.

Saquayamycin N (**1**): red amorphous powder; ^{1}H and ^{13}C NMR data, see Table S1; HRESIMS (positive mode) *m/z* 758.2103 $[M+H]^+$ (calcd. for $C_{36}H_{40}NO_{15}S$, 758.21187).

Fridamycin A (2): yellow amorphous powder; ^{1}H and ^{13}C NMR data, see Table S2; HRESIMS (positive mode) m/z 487.1606 $[M+H]^+$ (calcd. for $C_{25}H_{27}O_{10}$, 487.1604).

MIC tests

The minimum inhibitory concentration (MIC) was determined by the broth microdilution method (Uiterweerd *et al.*, 2020). A stock solution of compounds **1** and **2** was made by dissolving them in MeOH to a concentration of 1 mg/mL. *B. subtilis* 168 was grown from an overnight culture until an OD_{600} of 0.3 in LB and diluted until a concentration of bacteria of 1×10^6 CFU/mL in fresh broth. Two-fold serial dilutions of compounds **1** and **2** were tested in the concentration range of 250 µg/mL to 0.12 µg/mL in triplicates. Growth control wells containing 100 µL of 5×10^5 CFU/mL were included without test compounds. After overnight incubation at 37 °C, inhibition was defined as no visible growth compared to the growth observed in the control wells.

Acknowledgments

We are grateful to Mark Liles for providing *Bacillus* sp. 90A-23. The work was supported by NACTAR grant 14439 from the Netherlands Organization for Scientific Research (NWO) to G.P.v.W.

Supplementary information

Table S1. NMR data of 1 (in CD_3OD at 298 K).

Position	δ_H, mult. (J in Hz)[a]	δ_C, type[b]
1		206.7, C
2	a: 2.83, d (13.2) b: 2.64, dd (13.2, 2.5)	53.3, CH_2
3		77.3, C
4	2.11, m	46.3, CH_2
4a		84.4, C
5		162.6, C
6	6.65, s	107.4, CH
6a		139.6, C
7		190.0, C
7a		115.5, C
8		158.7, C
9		138.2, C
10	7.86, d (7.8)	134.5, CH
11	7.55, d (7.8)	119.7, CH
11a		132.6, C
12		183.3, C
12a		135.3, C
12b		79.5, C
3-Me	1.24, d (1.2)	30.2, CH_3
1′	5.01, d (11.1)	72.6, CH
2′	a: 2.37, m b: 1.51, m	37.8, CH_2
3′	3.84, m	77.9, CH
4′	3.49, t (9.0)	75.7, CH
5′	3.57, m	75.6, CH
6′	1.35, d (6.0)	17.7, CH_3
1″	5.20, d (2.8)	92.8, CH
2″	4.37, m	72.6, CH
3″	a: 2.79, dd (17.4, 2.7) b: 2.55, dd (17.4, 3.7)	40.9, CH_2
4″		210.1, C
5″	4.72, q (6.8)	78.9, CH
6″	1.31, d (6.8)	16.5, CH_3

1‴		173.4, C
2‴	4.81, m	51.9, CH
3‴	a: 3.59, m b: 3.21, m	33.5, CH_2
4‴		173.4, C
5‴	2.03, s	22.5, CH_3

[a] ^1H 850 MHz, [b] ^{13}C 213 MHz

Table S2. NMR data of 2 (in CD_3OD at 298 K).

Position	δ_H, mult. (J in Hz)[a]	δ_C, type[b]
1		175.4, C
2	2.54, d (5.7)	46.4, CH_2
3		73.0, C
4	a: 3.10, d (13.6) b: 3.01, d (13.6)	41.2, CH_2
4a		136.1, C
5	7.70, m	140.8, CH
6	7.68, m	119.6, CH
6a		132.9, C
7		189.2, C
7a		116.5, C
8		159.7, C
9		139.6, C
10	7.83, d (7.8)	134.2, CH
11	7.72, d (7.8)	120.1, CH
11a		133.1, C
12		189.1, C
12a		116.6, C
12b		162.4, C
3-Me	1.30, s	27.0, CH_3
1′	4.85[c]	72.5, CH
2′	a: 2.45, ddd (12.9, 4.9, 2.1) b: 1.41, m	40.9, CH_2
3′	3.71, ddd (11.3, 8.9, 4.9)	73.6, CH
4′	3.06, t (8.9)	78.8, CH
5′	3.46, dq (8.9, 6.0)	77.7, CH
6′	1.40, d (6.0)	18.7, CH_3

[a] ^1H 850 MHz, [b] ^{13}C 213 MHz, [c] the signal was suppressed with the water peak suppression

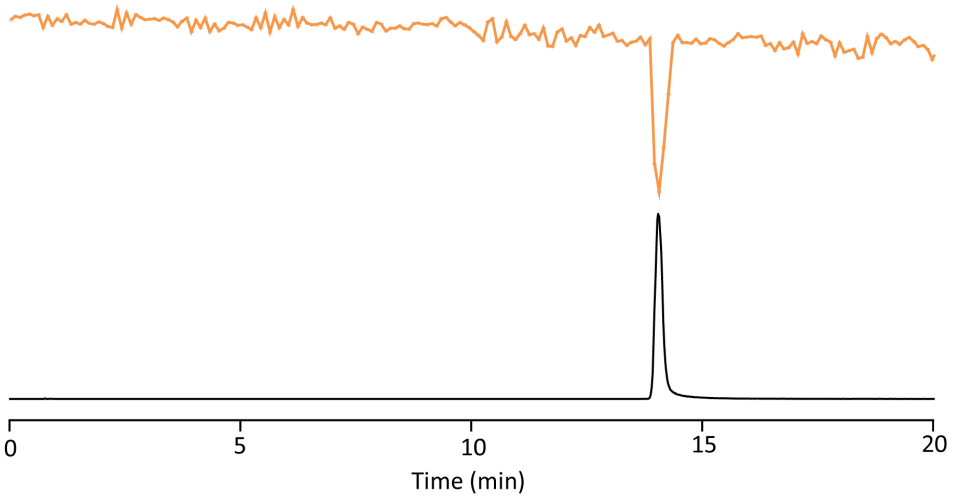


Figure S1. Delay calculation by correlation of mass spectrometry data with bioassay data. A. The bioactivity chromatogram of fractionated nalidixic acid against *E. coli* ASD19. The peak with the negative maxima indicates the growth inhibition of *E. coli*. **B.** The extracted ion chromatogram (EIC) of nalidixic acid. The time difference between the peak of bioactivity and MS chromatogram is determined as 0.3 min.

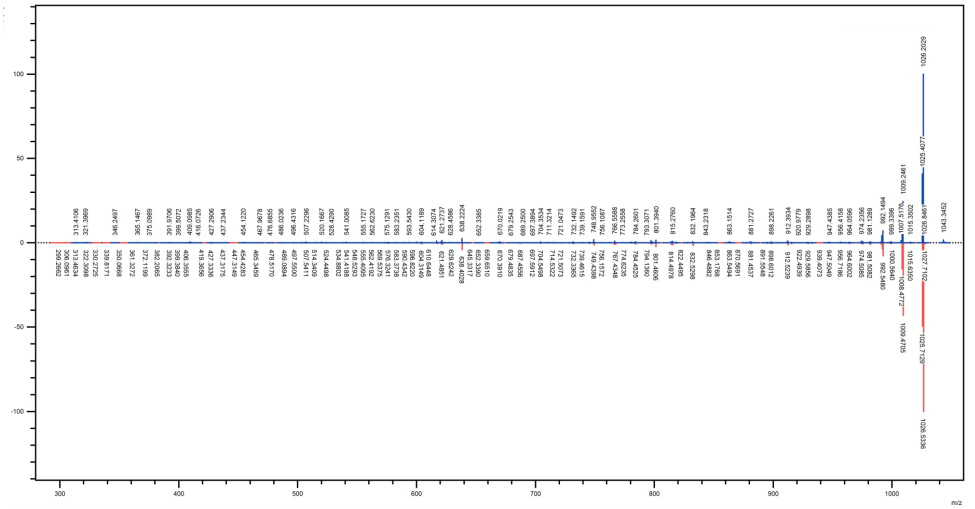


Figure S2. Direct MS/MS comparison of iturin A2 from the extract of *Bacillus* sp. 90A-23 with iturin A2 spectrum from GNPS spectral library (CCMSLIB00000086198).

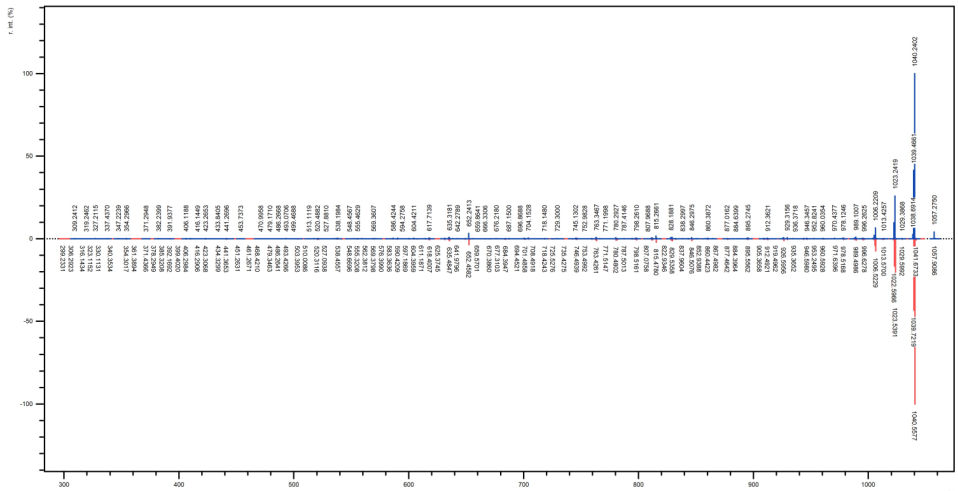


Figure S3. Direct MS/MS comparison of iturin A3/A4/A5 from the extract of *Bacillus* sp. 90A-23 with iturin A3/A4/A5 spectrum from GNPS spectral library (CCMSLIB00000086200).

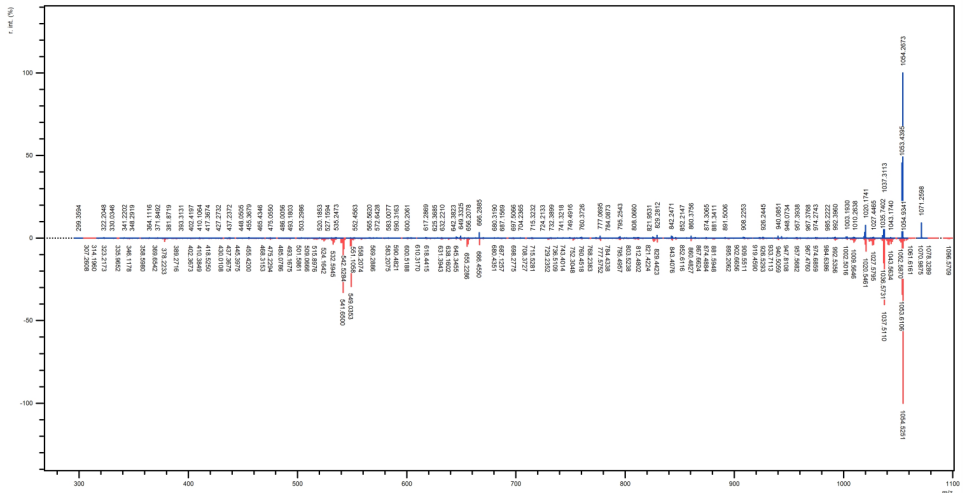


Figure S4. Direct MS/MS comparison of iturin A6/7 from the extract of *Bacillus* sp. 90A-23 with iturin A6/7 spectrum from GNPS spectral library (CCMSLIB00000086202).

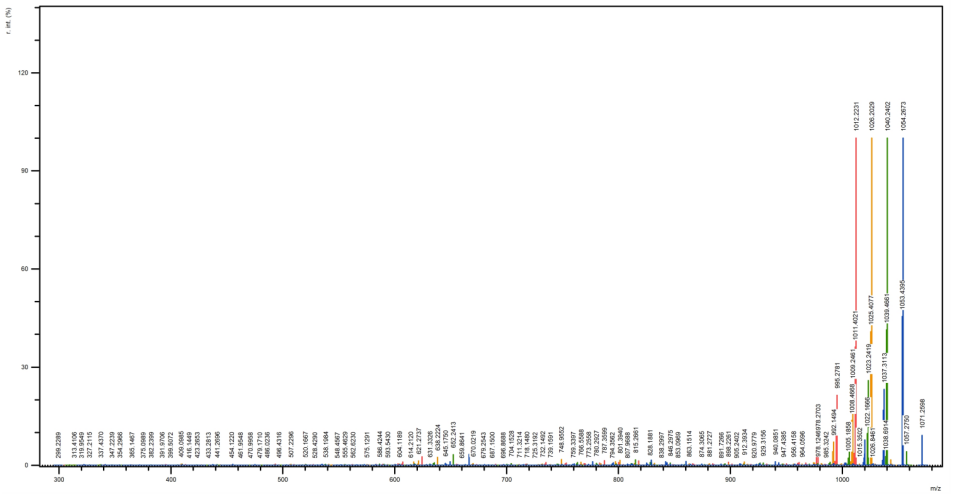


Figure S5. Direct MS/MS comparison of the iturin A1, iturin A2, iturin A3/A4/A5 and iturin A6/7 spectra from *Bacillus* sp. 90A-23 extract.

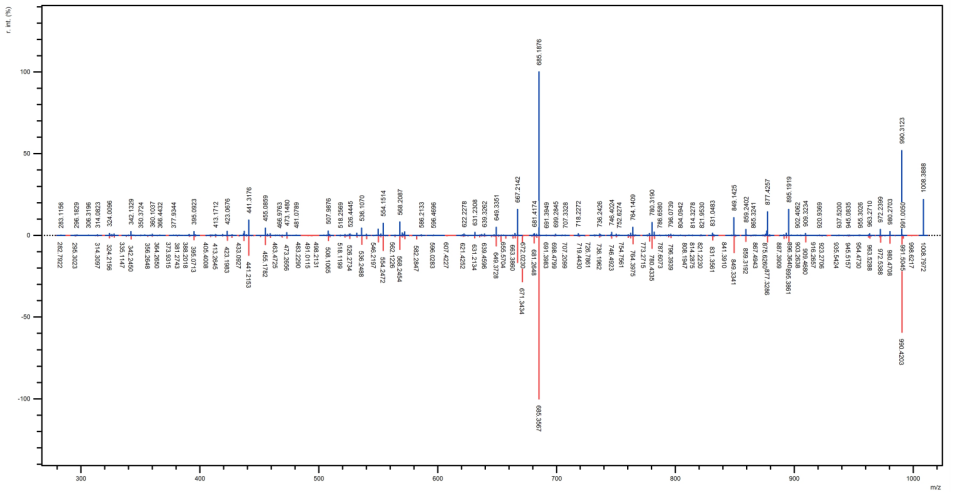


Figure S6. Direct MS/MS comparison of surfactin Leu/Ile7 C13 from the extract of *Bacillus* sp. 90A-23 with surfactin Leu/Ile7 C13 spectrum from GNPS spectral library (CCMSLIB00000006885).

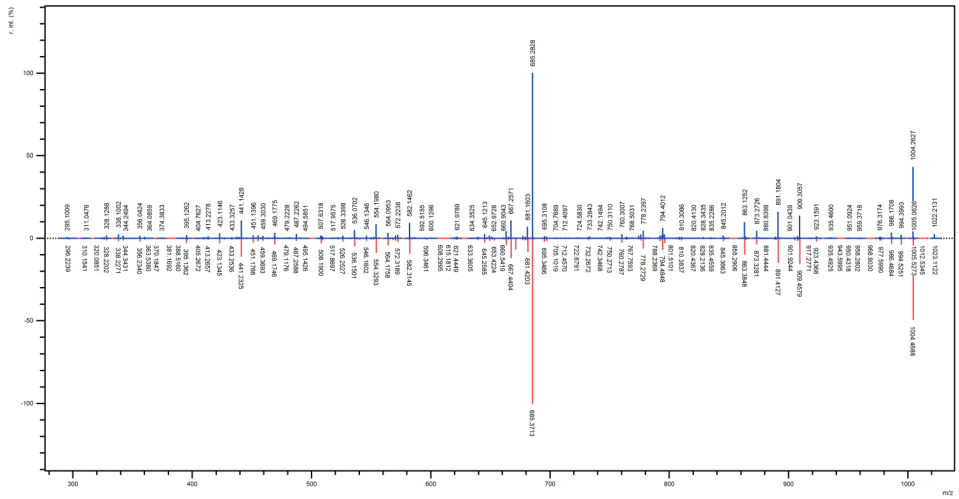


Figure S7. Direct MS/MS comparison of surfactin Leu/Ile7 C14 from the extract of *Bacillus* sp. 90A-23 with surfactin Leu/Ile7 C14 spectrum from GNPS spectral library (CCMSLIB00000006890).

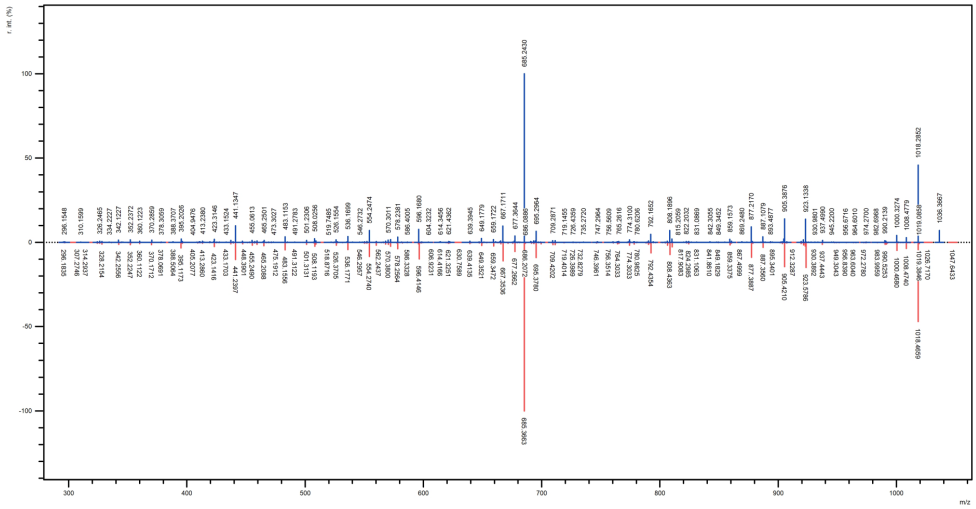


Figure S8. Direct MS/MS comparison of surfactin Leu/Ile7 C15 from the extract of *Bacillus* sp. 90A-23 with surfactin Leu/Ile7 C15 spectrum from GNPS spectral library (CCMSLIB00000006895).

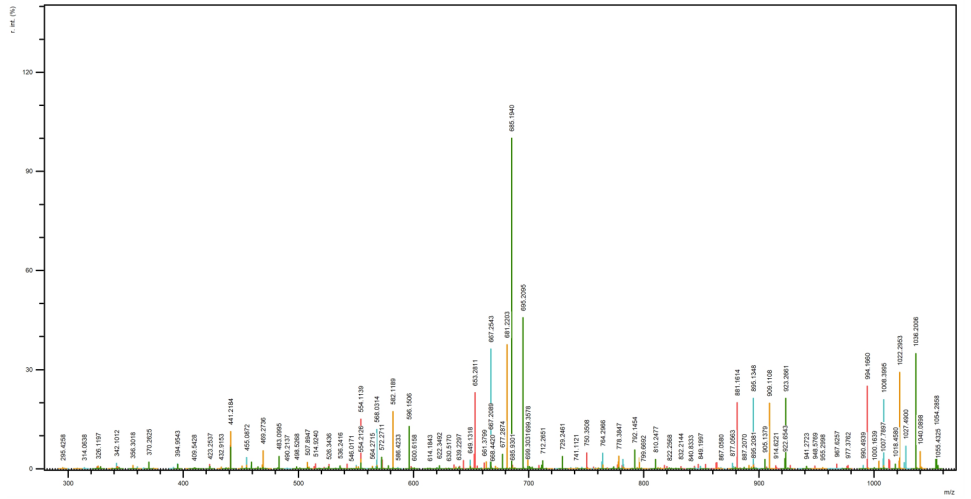


Figure S9. Direct MS/MS comparison of the C12, C13, C14 and C15 linear surfactins Leu/Ile7 from *Bacillus* sp. 90A-23 extract.

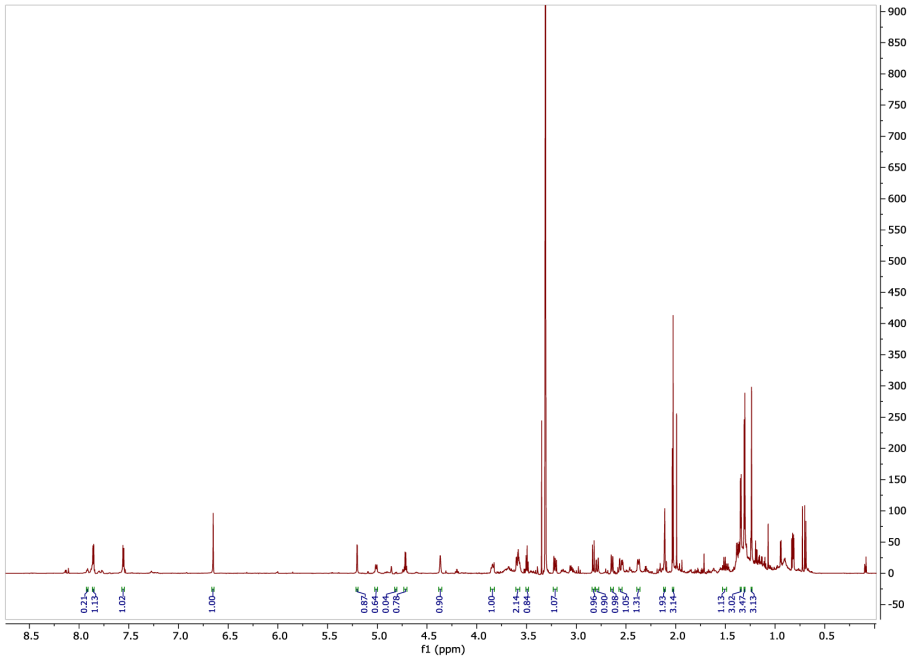


Figure S10. ^{1}H NMR spectrum of saquayamycin N (1) (850 MHz, in CD_3OD).

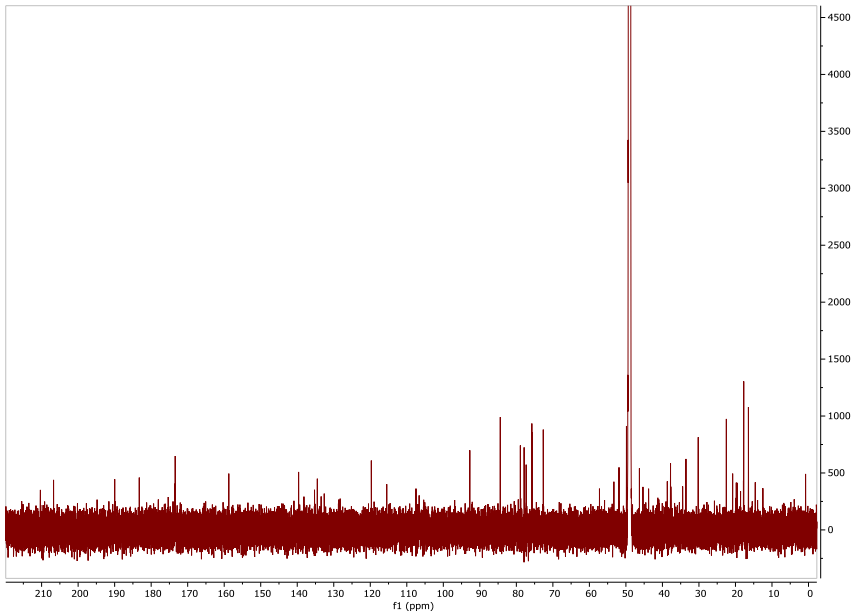


Figure S11. ^{13}C NMR spectrum of saquayamycin N (1) (213 MHz, in CD_3OD).

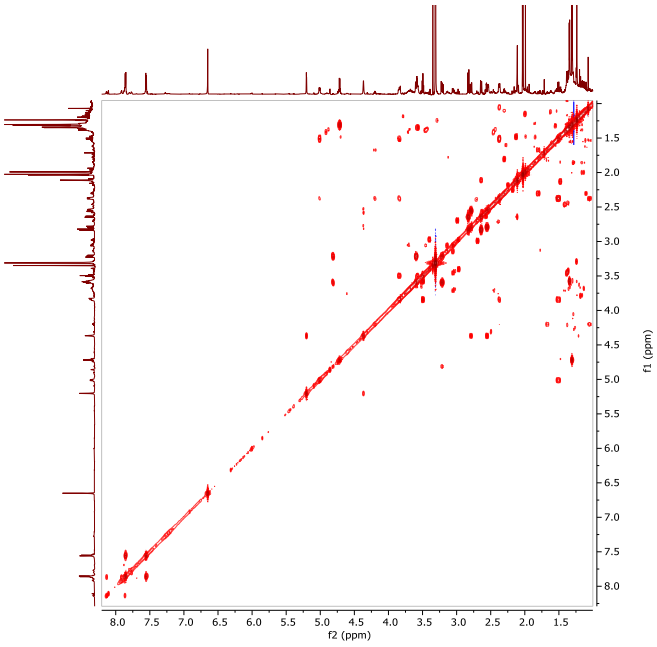


Figure S12. COSY spectrum of saquayamycin N (1) (850 MHz, in CD_3OD).

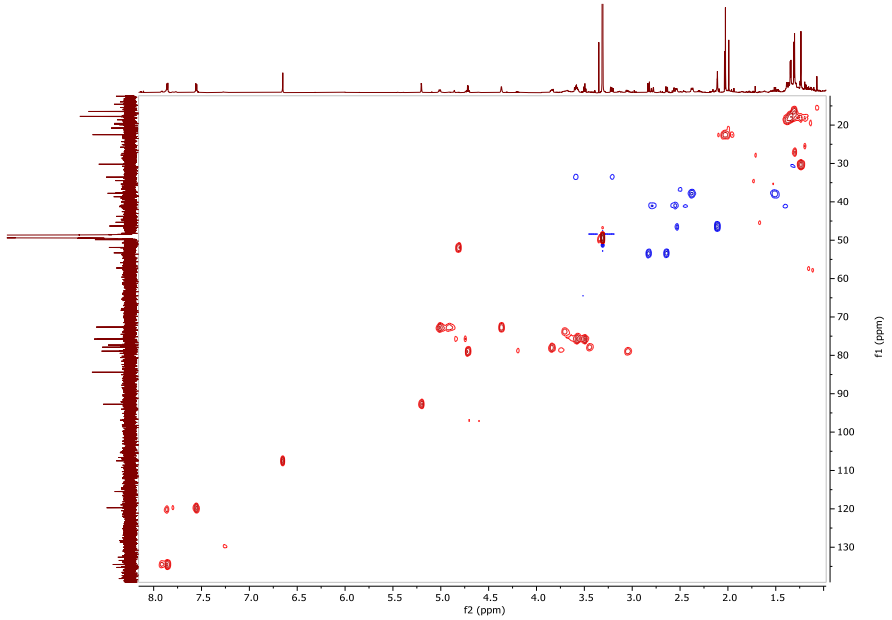


Figure S13. Multiplicity-edited HSQC spectrum of saquayamycin N (1) (850 MHz, in CD_3OD).

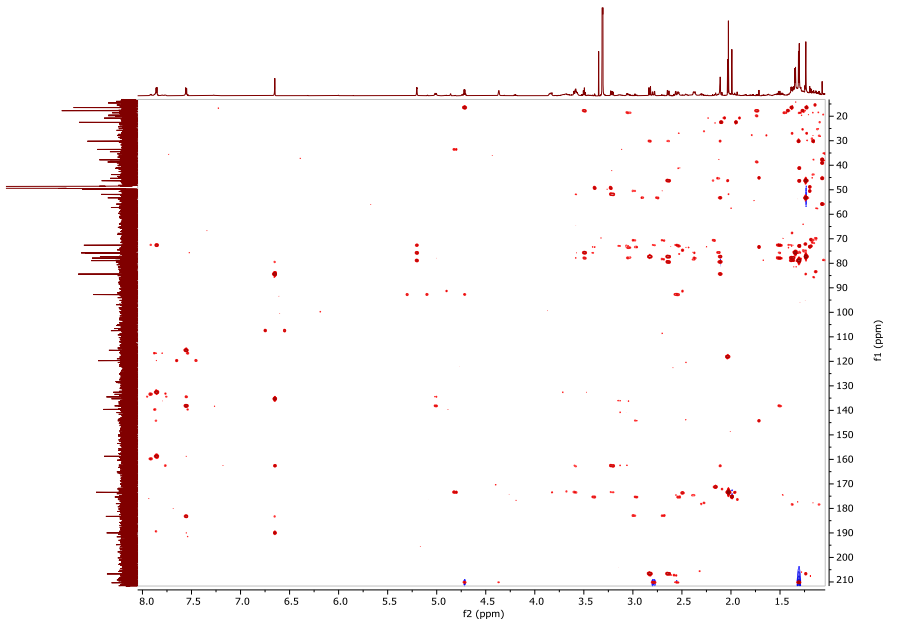


Figure S14. HMBC spectrum of saquayamycin N (1) (850 MHz, in CD_3OD).

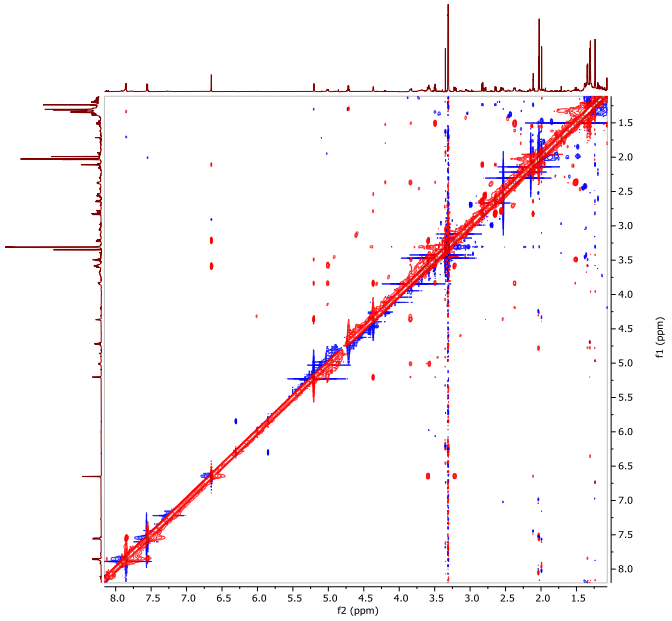


Figure S15. NOESY spectrum of saquayamycin N (1) (850 MHz, in CD_3OD).

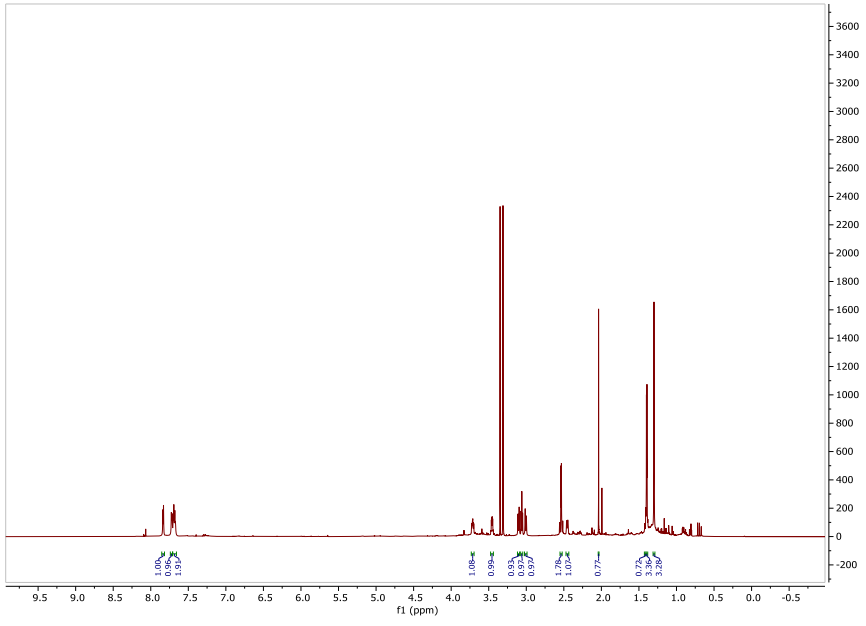


Figure S16. ^1H NMR spectrum of fridamycin A (2) (850 MHz, in CD_3OD).

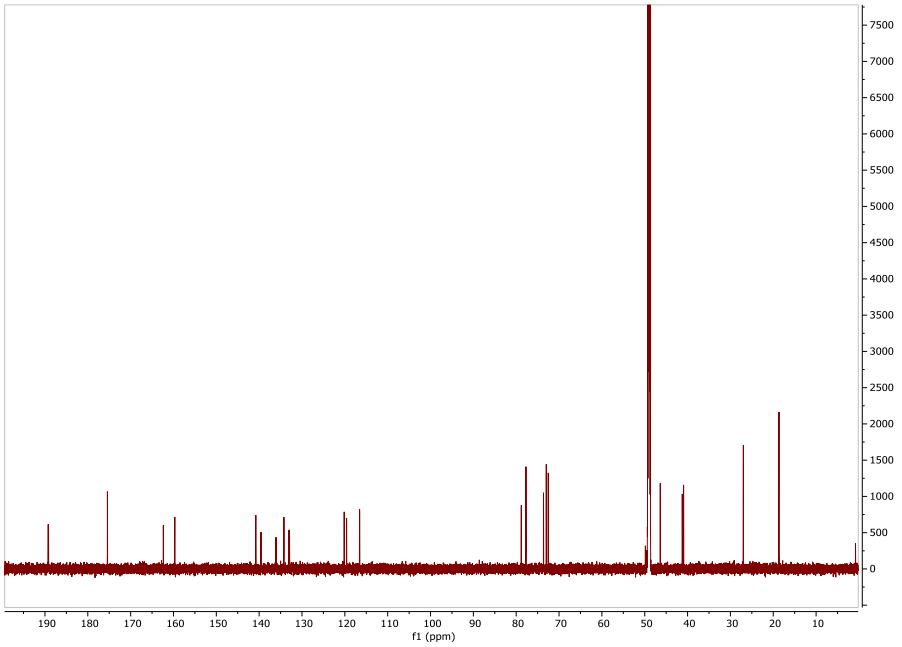


Figure S17. ^{13}C NMR spectrum of fridamycin A (2) (213 MHz, in CD_3OD).

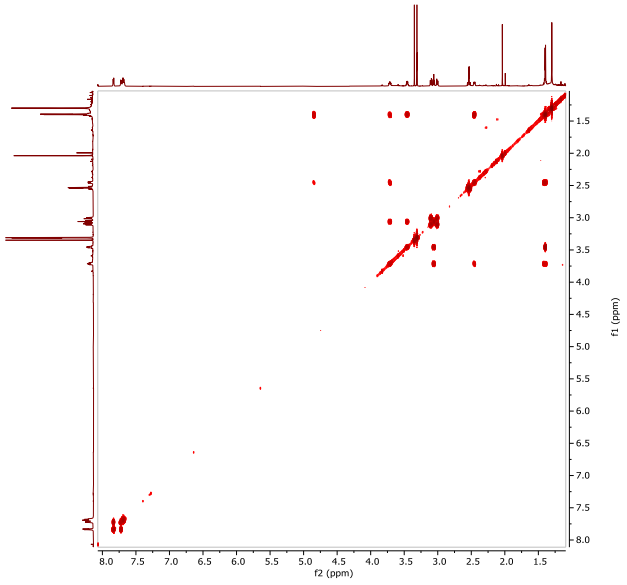


Figure S18. COSY spectrum of fridamycin A (2) (850 MHz, in CD_3OD).

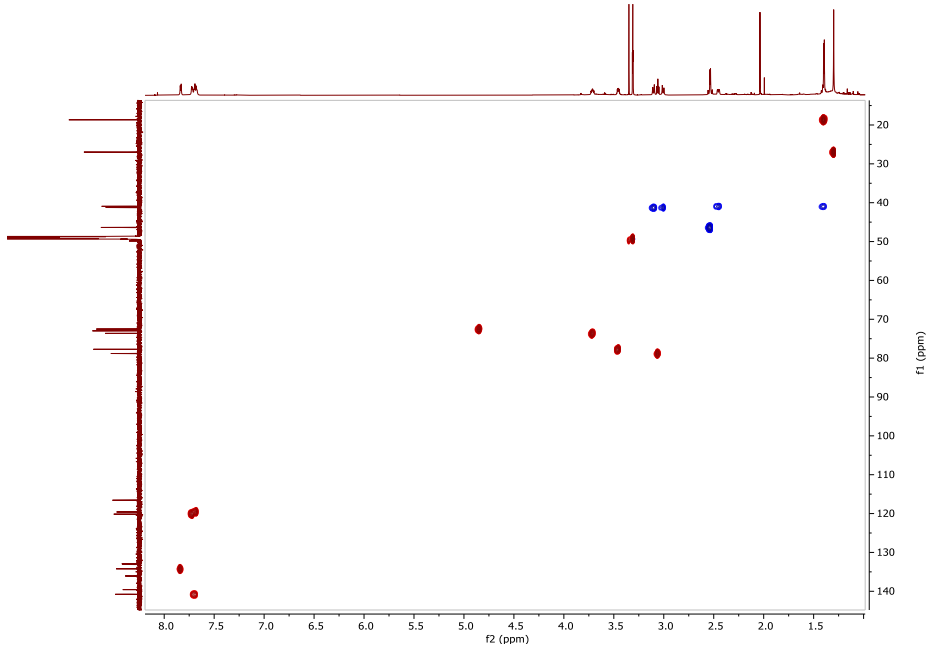


Figure S19. Multiplicity-edited HSQC spectrum of fridamycin A (2) (850 MHz, in CD_3OD).

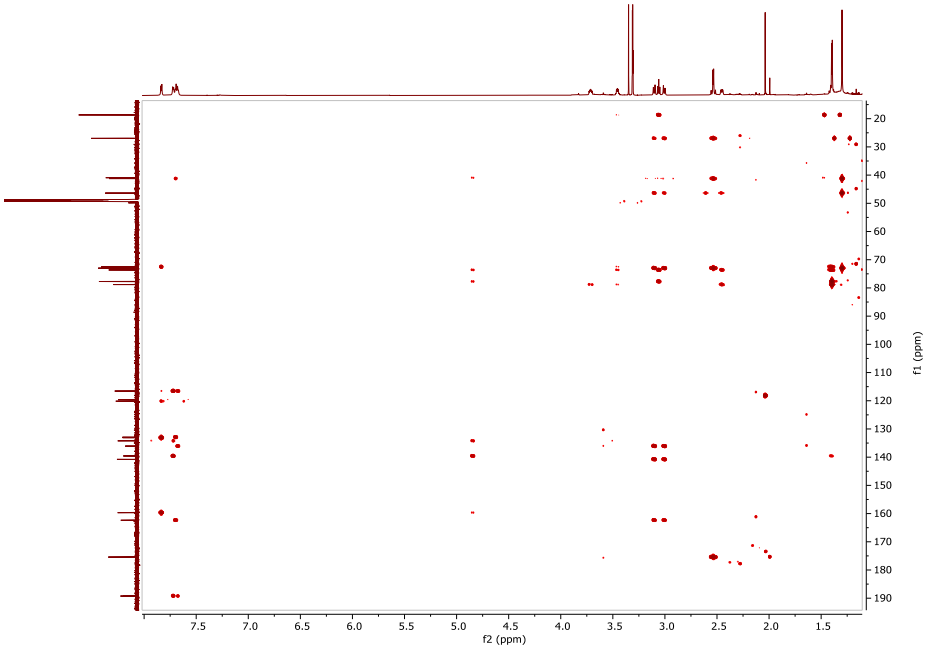


Figure S20. HMBC spectrum of fridamycin A (2) (850 MHz, in CD_3OD).

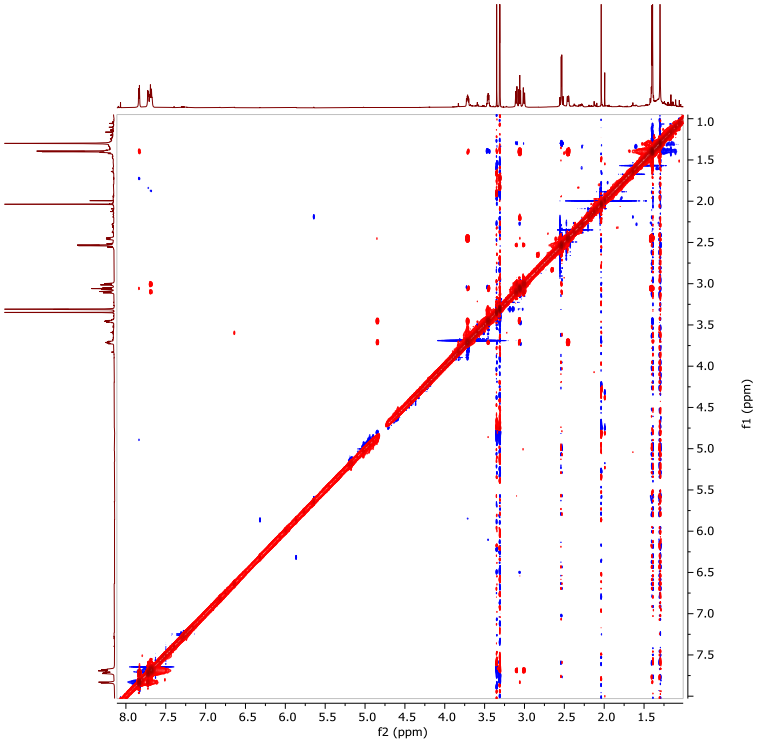


Figure S21. NOESY spectrum of 1 fridamycin A (2) (850 MHz, in CD_3OD).

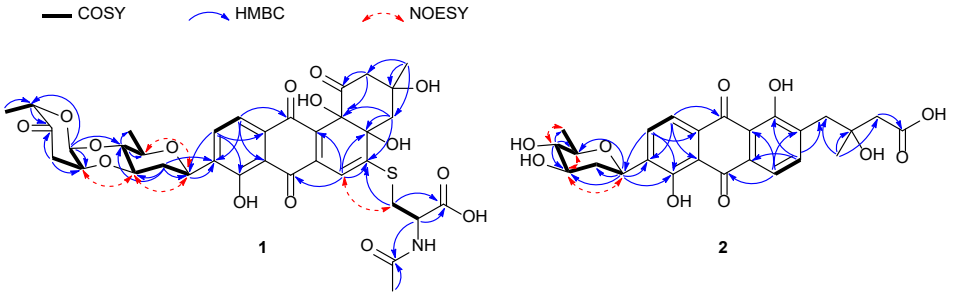


Figure S22. Correlations obtained by COSY, HMBC and NOESY measurements in NMR of saquayamycin N (1) and fridamycin A (2).

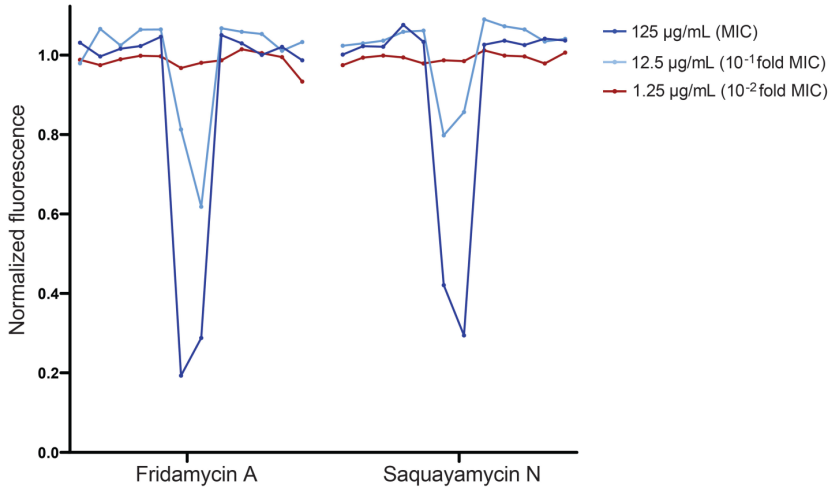
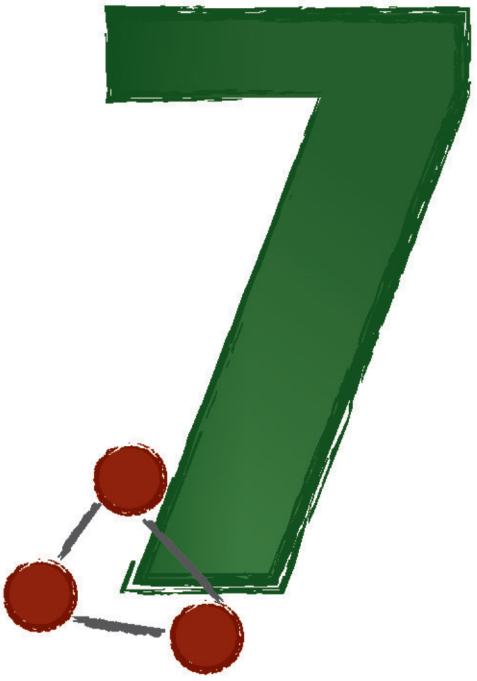


Figure S23. Results of the resazurin reduction assay for the purified saquayamycin N (1) and fridamycin A (2) against *B. subtilis* 168. Compounds were tested in 125 µg/mL, 12.5 µg/mL and 1.25 µg/mL concentration. Note, that the MIC and sub-MIC concentration (12.5 µg/mL) resulted in the drop of fluorescence.



Chapter 7

Discovery and derivatization of tridecaptin antibiotics with altered host specificity and enhanced bioactivity

Nataliia V. Machushynets, Karol Al Ayed, Barbara R. Terlouw, Chao Du, Ned P. Buijs, Joost Willemse, Somayah S. Elsayed, Julian Schill, Vincent Trebosc, Michel Pieren, Francesca M. Alexander, Stephen A. Cochrane, Mark R. Liles, Marnix H. Medema, Nathaniel I. Martin and Gilles P. van Wezel

This chapter was published as:
Machushynets N.V., Ayed K.A., Terlouw B.R., Du C., Buijs N.P., Willemse J., Elsayed S.S., Schill J., Trebosc V., Pieren V., Alexander F.M., Cochrane S.A., Liles M.R., Medema M.H, Martin N.I. & van Wezel G.P.(2024). Discovery and derivatization of tridecaptin antibiotics with altered host specificity and enhanced bioactivity. *ACS Chem Bio.*

Abstract

The prevalence of multidrug-resistant (MDR) pathogens combined with a decline in antibiotic discovery presents a major challenge for health care. To refill the discovery pipelines, we need to find new ways to uncover new chemical entities. Here, we report the global genome mining-guided discovery of new lipopeptide antibiotics tridecaptin A_5 and tridecaptin D, which exhibit unusual bioactivities within their class. The change in antibacterial spectrum of Oct-TriA_5 was explained solely by a Phe to Trp substitution as compared to Oct-TriA_1, while Oct-TriD contained 6 substitutions. Metabolomics analysis of producer *Paenibacillus* sp. JJ-21 validated the predicted amino acid sequence of tridecaptin A_5. Screening of tridecaptin analogs substituted at position 9 identified Oct-His9 as a potent congener with exceptional efficacy against *Pseudomonas aeruginosa* and reduced hemolytic and cytotoxic properties. Our work highlights the promise of tridecaptin analogs to combat MDR pathogens.

Introduction

The overuse of antibiotics over the past many decades has contributed to the rapid emergence and spread of antimicrobial resistance. Combined with the decline in the number of new clinically approved antibacterial drugs, infections caused by resistant bacteria are frequently difficult to treat (Wright, 2015). The ever-increasing prevalence of multidrug-resistant (MDR) bacteria is recognized by the World Health Organization (WHO) as an imminent threat to human health, particularly the Gram-negative critical priority pathogens *Acinetobacter baumannii, Pseudomonas aeruginosa* and *Enterobacteriaceae* (WHO, 2017). Thus, there is an urgent need for novel antimicrobial agents, and the challenge lies in finding the undiscovered gems that may form the basis for our future medicines (Newman & Cragg, 2007, Davies & Davies, 2010). Traditional high-throughput screening is becoming less attractive due to the issue of dereplication, rediscovering compounds that had already been identified before (Kolter & van Wezel, 2016, Lewis, 2013). However, next-generation sequencing has revealed that microbial genomes still harbor a huge underexplored biosynthetic potential of bacteria (Medema *et al.*, 2015, Gavriilidou *et al.*, 2022).

Polymyxins are a promising class of antibiotics that are used as a last line of therapy against MDR pathogens (Chiu *et al.*, 2022), and are primarily produced by *Paenibacillus* species. With resistance against polymyxins inevitably increasing, we urgently need to search for potent alternatives. *Bacillus* and *Paenibacillus* produce a range of other classes of lipopeptides that act on MDR Gram-negative pathogens, including brevicidines, laterocidines, relacidines, paenibacterins and tridecaptins (Al Ayed *et al.*, 2023, Cochrane & Vederas, 2016, Olishevska *et al.*, 2019, Raaijmakers *et al.*, 2010, Bann *et al.*, 2021). Of these, tridecaptins stand out as they possess a mechanism of action that is distinct from the aforementioned lipopeptides. In addition to interacting with lipopolysaccharides (LPS), tridecaptins also bind to Gram-negative lipid II and in doing so cause the disruption of the proton motive force (Cochrane *et al.*, 2015, Buijs *et al.*, 2023, Medeiros-Silva *et al.*, 2019, Cochrane *et al.*, 2016). This unique dual mechanism of action reduces the likelihood of cross-resistance (Jangra *et al.*, 2020). Since the discovery of tridecaptin A in 1978, multiple tridecaptin variants with modifications at either the N-terminal fatty acid moiety or in the amino acid sequence have been reported. Tridecaptin A (Shoji *et al.*, 1978), tridecaptin B (Cochrane *et al.*, 2015), tridecaptin C (Shoji J. *et al.*, 1979) and tridecaptin M (Jangra *et al.*, 2019b) are predominantly active against Gram-negative bacteria, while the recently discovered tridecaptin G (da Costa *et al.*, 2022) has a broad-spectrum activity. Studies on natural tridecaptins include research on their biosynthesis, structure-activity relationships, their mechanism of action and their synergy with other antibiotics (Bann *et al.*, 2021). Structure-activity studies formed the basis for drug development studies to enhance their potency (Ballantine *et al.*, 2019, Cochrane *et al.*, 2014), expand their bioactivity spectrum (Wang *et al.*, 2022a), or increase their stability (Ballantine *et al.*, 2018).

In this regard, approaches that combine genome mining with variation in culturing conditions has proven to be a valuable way to achieve differential synthesis of NPs, followed by the metabolic profiling-based identification of the bioactivity of interest (O'Leary *et al.*, 2016). However, a major challenge is to find the appropriate chemical triggers or ecological cues to elicit the production of cryptic antibiotics (Zhu *et al.*, 2014a, Xu *et al.*, 2019, van der Meij *et al.*, 2017, van Bergeijk *et al.*, 2020). An alternative way to investigate the potential of novel scaffolds predicted by genome mining is via organic synthesis or chemoenzymatic total synthesis (Wang *et al.*, 2022a, Wang *et al.*, 2022b, Wang *et al.*, 2022c). Genome mining tools such as antiSMASH (Blin *et al.*, 2023) allow interrogation of microbial genomes for the presence of biosynthetic gene clusters (BGCs) that specify the biosynthesis of natural products and, subsequently, predict the types of compounds that are derived from them, such as polyketides (Ray & Moore, 2016), ribosomally synthesized and post-translationally modified peptides (RiPPs) (Montalban-Lopez *et al.*, 2021), terpenes (Avalos *et al.*, 2022) or nonribosomal peptides (NRPs) (Süssmuth & Mainz, 2017). In the case of nonribosomal peptide synthetases (NRPS), NRPS modules and the amino acids predicted to be incorporated by the A domain in each module can be predicted using NRPS-specific prediction algorithms, such as NRPSpredictor2 (Röttig *et al.*, 2011) or SANDPUMA (Chevrette *et al.*, 2017) or machine learning techniques trained on a set of A domains with known specificities (Mongia *et al.*, 2023). With the power of modern synthetic organic chemistry and the increasing accuracy of natural product structure prediction algorithms, it is increasingly possible to chemically synthesize new bioactive molecules based on BGC sequences (Wang *et al.*, 2022b).

Here, we report a global genome-mining approach that led to the discovery of novel tridecaptins. Bioinformatics analysis of 785 bacterial genomes identified novel tridecaptins tridecaptin A_5 and tridecaptin D. Unexpectedly, two synthetic analogs Oct-TriD and Oct-Tri$A_{5,}$ efficiently killed both Gram-negative and Gram-positive bacteria. The increased bioactivity spectrum of tridecaptin Oct-TriA_5 compared to Oct-TriA_1 also correlated with hemolytic and cytotoxic activity and could be attributed to a single amino acid substitution at position 9. In contrast to Oct-TriA_1, the broadened antimicrobial spectrum of Oct-TriA_5 and Oct-TriD is caused by its increased membrane disruptive capacity on Gram-positive pathogens. Subsequent screening of other amino acids at position 9 led to the discovery of the potent tridecaptin analog Oct-TriHis9 with reduced hemolytic and cytotoxic properties and potent activity against *P. aeruginosa*.

Results and Discussion

Large-scale network analysis and bioinformatic prediction of novel tridecaptin BGCs

Paenibacillus sp. JJ-21 is a gifted natural product producer isolated from corn rhizosphere and identified as an antibacterial metabolite producer based on its antagonistic potential against polymyxin-resistant *Escherichia coli* ATCC 25922 harboring plasmid-mediated *mcr-1* resistance gene. The *Paenibacillus* sp. JJ-21 genome was sequenced using the PacBio platform. Assembly of the PacBio reads with Falcon (version 1.8.1) (Wang *et al.*, 2016) resulted in a single contig of 6.2 Mb (GenBank accession number: CP132974). The *Paenibacillus* sp. JJ-21 genome was analyzed using antiSMASH 6.0.1 (Blin *et al.*, 2021) to obtain an overview of the predicted BGCs. AntiSMASH predicted a total of 18 BGCs, including three for NRPSs (Figure 1A). These NRPS BGCs were predicted to encode the biosynthesis of fusaricidin, tridecaptin and polymyxin BGCs, respectively, based on their high similarity (identical gene order and >75% nucleotide sequence identity) to characterized clusters (listed in MIBiG) (Terlouw *et al.*, 2023). An in silico analysis of the A domain substrate specificity of these BGCs was conducted with the software tool PARAS (v0.0.4, available at https://github.com/BTheDragonMaster/paras) to predict the amino acid composition of peptide scaffolds (Table S1). Surprisingly, the amino acid sequence predicted from the tridecaptin BGC differed from the well-known tridecaptin A_1 and contained Trp instead of Phe in position 9 (Table S2).

To gain insight into the chemical space of NRPSs and specifically tridecaptin BGCs, we bioinformatically analyzed 785 complete genomes from *Paenibacillus* spp. for BGCs encoding NRPSs. The genome of *Paenibacillus* sp. JJ-21 was included as a reference genome in the data set. To visualize the diversity, distribution, and NRPS novelty, a BiG-SCAPE (Navarro-Muñoz *et al.*, 2020) sequence similarity network (SSN) was then constructed, which consisted of 4,367 NRPS BGCs forming 1609 gene cluster families (GCFs) (Figure 1B).

Genetic variation of BGCs within GCFs is often directly associated with structural differences between their molecular products, and even small chemical variations can lead to different biological activities (Medema *et al.*, 2014). We also visualized the modular architecture of NRPS assembly lines, revealing the order and the number of modules for each NRPS synthase that, in turn, helped to determine the class of the molecules they encode. The BiG-SCAPE sequence similarity network highlighted various known classes of NRPS BGCs, such as polymyxins (three NRPSs, 10 modules), tridecaptins (two NRPSs, 13 modules), paenilipoheptins (three NRPSs, seven modules), fusaricidins (one NRPS, six modules), paenibacterins (four NRPSs, 13 modules), octapeptins (three NRPS, eight modules), bacitracins (three NRPS, 12 modules) and cilagicins (three NRPS, 12 modules), but about twenty percent of the detected BGCs could not be linked to known compounds (Terlouw *et al.*, 2023).

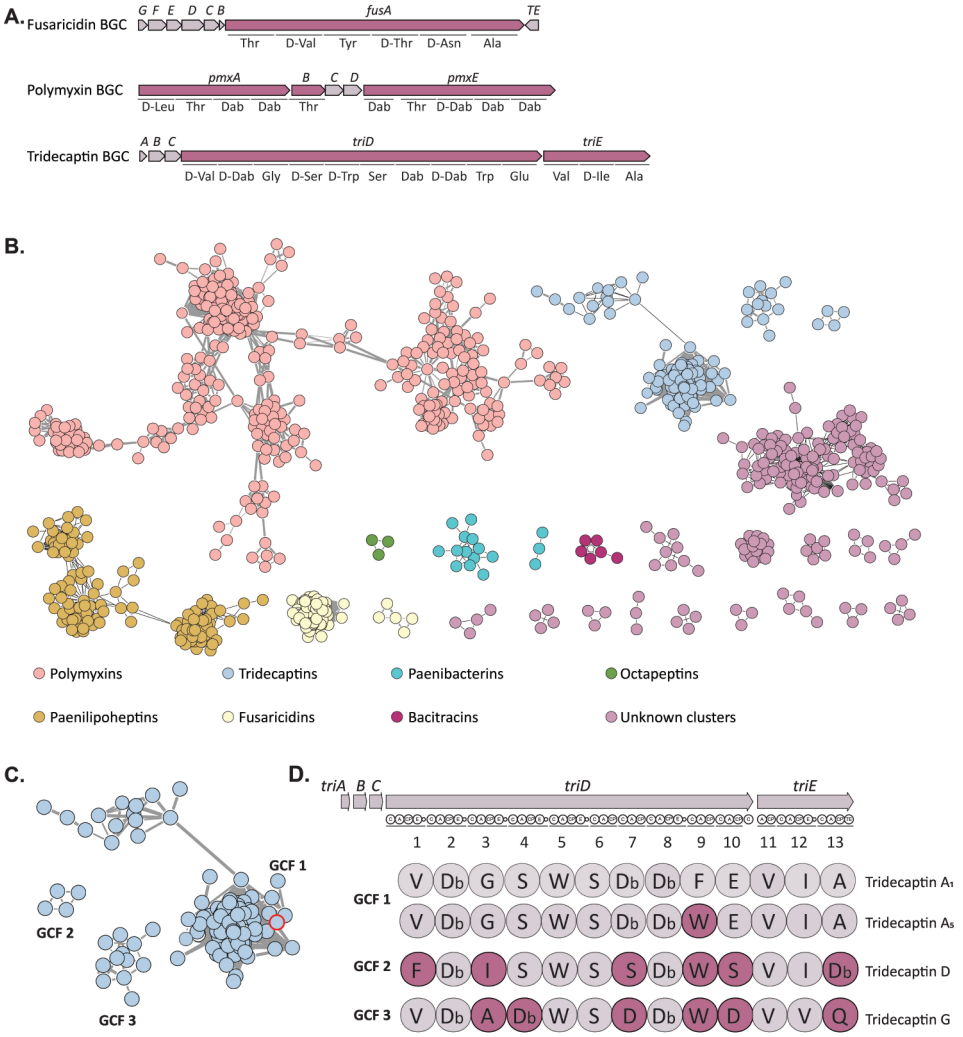


Figure 1. ***Paenibacillus*** **sp. JJ-21 BGC sequence similarity network and tridecaptin BGC analysis. A.** Organization of the NRPS BGCs predicted in the *Paenibacillus* sp. JJ-21 genome that encode the biosynthesis of lipopeptides. **B**. BiG-SCAPE sequence similarity network (SSN) (c 0.25) containing validated NRPS BGCs of *Paenibacillus* spp. visualized in Cytoscape. Each node represents one NRPS BGC identified by antiSMASH. Singletons and single links are not shown. **C.** Enlarged gene cluster families (GCFs) of tridecaptins BGCs. Red circle indicates the tridecaptin BGC detected in the genome of *Paenibacillus* sp. JJ-21. **D.** Scheme of the tridecaptin BGC and amino acid sequences of tridecaptin A_1, new tridecaptin A_5, tridecaptin D and tridecaptin G. The sequences of their peptide moieties were predicted from the active site sequences of their A domains. Note, the peptide sequences of tridecaptin A_5 and the well-known tridecaptin A_1 differ exclusively in one amino acid at position 9. Db: 2,4-diaminobutyric acid.

Surprisingly, three GCFs contained BGCs with 13 NRPS modules and were predicted as tridecaptin BGCs based on the antiSMASH (Blin *et al.*, 2021) and BiG-SCAPE (Navarro-Muñoz *et al.*, 2020) results (Figure 1C). The assembly line of tridecaptins is subdivided into two peptide synthetases, namely TriD and TriE, consisting of ten and three modules, respectively

(Lohans *et al.*, 2014). Notably, substrate specificity predictions by PARAS indicated that the tridecaptin BGCs within GCF 1 not only encoded the previously characterized tridecaptin A_1, but also a new analog, designated tridecaptin A_5, which should incorporate Trp in position 9, and was detected in the genome of *Paenibacillus* sp. JJ-21 (Figure 1D, Table S3). Additionally, we discovered another new structurally distinct analog of tridecaptins, which we designated tridecaptin D, and is predicted to be encoded by the tridecaptin BGCs of GCF 2. GCF 3 represented BGCs corresponding to the recently published tridecaptin G (da Costa *et al.*, 2022). Tridecaptin D differs from tridecaptin A_1 at six amino acid positions, with five positions of the molecule featuring amino acids that are unique among all known tridecaptins identified so far, namely Phe1, Ile3, Ser7, Ser10 and Dab13. We used antiSMASH (Blin *et al.*, 2021) to predict which modules in the BGC for tridecaptin D contained epimerization domains, *i.e.* domains that catalyze the conversion from L- to D-amino acids. Modules 1–5 and 8 of TriD and module 12 of TriE, which incorporate Phe, Dab, Ile, Ser, Trp, Dab, and Ile, respectively, were all predicted to contain epimerization domains (Table S3).

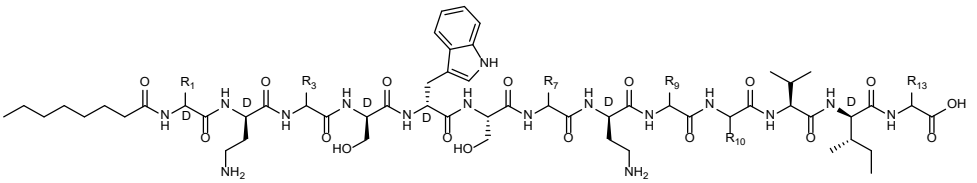
Identification of tridecaptin A_5 in vivo using metabolomics

To validate the bioinformatic prediction of the amino acid sequence of tridecaptin A_5, the lipopeptide was detected in the cultures of *Paenibacillus* sp. JJ-21 and analyzed by LC-MS/MS. For this, *Paenibacillus* sp. JJ-21 was fermented in liquid TSB media, biomass was collected by centrifugation and, subsequently, extracted with acidified 70% IPA. LC-MS/MS analysis identified fusaricidins and polymyxins as the major specialized metabolites. Considering that all known tridecaptins contain several positively charged diaminobutyric acids (Dab), we searched for the MS/MS spectra that contained a *b* ion with the *m/z* 101.0709. Besides polymyxins that also contained multiple Dab residues, we detected a compound with an *m/z* 539.9712, corresponding to an $[M+3H]^{3+}$ ion. On the basis of the biosynthetic features derived from the genome sequence of *Paenibacillus* sp. JJ-21, we initiated *de novo* sequencing of the compound with the *m/z* 539.9712 from the *y* and *b* ion series (Figure S1). The MS/MS spectrum showed a low-mass ion region containing a Trp immonium ion with an *m/z* 159.0912, as well as internal fragment ions. Fragment analysis of the *y* and *b* ions yielded an amino acid sequence that matched that predicted for tridecaptin A_5, which is FA-D-Val–D-Dab–Gly–D-Ser–D-Trp–L-Ser–L-Dab–D-Dab–L-Trp–L-Glu–L-Val–D-Ile–L-Ala. The MS/MS spectrum revealed the presence of a hydroxy-containing C11 fatty acid in tridecaptin A_5, similar to the previously reported 3-hydroxy-methyldecanoic fatty acids found in tridecaptin A_3 and tridecaptin A_4 (Lohans *et al.*, 2014). The MS/MS studies did not allow us to discriminate between Leu, Ile and *allo*-Ile in position 12 of the tridecaptin sequence. Here, the assignment was performed according to the structure prediction from our genome mining studies and comparison with the literature data (Lohans *et al.*, 2014). Additionally, the MS results would not allow optical isomers to be distinguished. Therefore, the configurations

of the residues of tridecaptin A_5 were predicted from the domain organization of the modules along the assembly lines.

Chemical synthesis and antimicrobial testing of tridecaptin analogs

Because of the low levels of production of tridecaptin A_5 by *Paenibacillus* sp. JJ-21, we decided to produce synthetic analogs of the peptides by solid phase peptide synthesis (SPPS). The added value of this approach is that it also allows the synthesis and direct comparison of analogs, chosen based on the substrate specificity and stereochemical predictions generated from the analysis of the tridecaptin BGCs. Previously, it was demonstrated that a synthetic analog of tridecaptin A_1 ($TriA_1$) wherein the N-terminal acyl moiety was replaced with octanoic acid retained the full activity of the natural product against ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *A. baumannii*, *P. aeruginosa*, and *Enterobacter* species) (Cochrane *et al.*, 2014). For this reason, in the present studies the chiral lipid tail of the tridecaptins was also replaced with the octanoyl chain, which is more accessible than natural lipid and enables larger quantities of peptide to be obtained for biological studies. Linear synthetic analogs of tridecaptin A_1, tridecaptin A_5 and tridecaptin D were synthesized and designated as Oct-$TriA_1$ (**1**), Oct-$TriA_5$ (**2**) and Oct-TriD (**3**), respectively (Table S4, Figure 2, Figures S2–S4).



		R_1	R_3	R_7	R_9	R_{10}	R_{13}
Oct-$TriA_1$	(1)	D-Val	Gly	Dab	Phe	Glu	Ala
Oct-$TriA_5$	(2)	D-Val	Gly	Dab	Trp	Glu	Ala
Oct-TriD	(3)	D-Phe	D-Ile	Ser	Trp	Ser	Dab

Figure 2. Structures of synthetic tridecaptin A variants Oct-$TriA_1$ (1), Oct-$TriA_5$ (2) and Oct-TriD (3).

Subsequently, Oct-$TriA_1$, Oct-$TriA_5$ and Oct-TriD were assayed for activity against the ESKAPE pathogens. MIC values were determined using broth-dilution assays (Table 1). Oct-$TriA_1$ displayed potent activity against most of the Gram-negative test-strains, but was less active against Gram-positive bacteria, similar to known tridecaptins (i.e., tridecaptins A, B, C, and M) (Cochrane *et al.*, 2014). Surprisingly, Oct-TriD showed significant bioactivity against the Gram-positive species *S. aureus, E. faecalis and E. faecium*, with an MIC between 4–8 μg/mL, in addition to moderate activity against Gram-negative pathogens. To the best of our knowledge, Oct-TriD is the first tridecaptin analog that is predominantly active against Gram-positive bacteria. Moreover, Oct-$TriA_5$ showed broad-spectrum bioactivity, similar to

the recently published synthetic analog syn-CNRL5 and natural tridecaptin G (Wang *et al.*, 2022a, da Costa *et al.*, 2022). Specifically, Oct-TriA$_5$ exhibited activity against Gram-negative ESKAPE pathogens, with MICs ranging from 2 to 16 μg/mL. Notably, a single amino acid substitution at position 9 from Phe in Oct-TriA$_1$ to Trp in Oct-TriA$_5$ changed the bioactivity profile of Oct-TriA$_5$ from being limited to Gram-negative bacteria to broad-spectrum activity.

In addition, tridecaptins were tested for their hemolytic and cytotoxic activity (Table 1). Oct-TriA$_5$ and Oct-TriD showed increased hemolytic and cytotoxic activity compared to Oct-TriA$_1$, suggesting that the altered spectrum of activity of Oct-TriA$_5$ and Oct-TriD is due to unspecific membrane lysis properties and emphasizing the importance of the amino acid in position 9 for selectivity towards antibacterial activity.

Mode of action on Gram-positive and Gram-negative bacteria

Tridecaptin A$_1$ exerts its bactericidal effect on Gram-negative bacteria by binding to lipid II on the surface of the inner membrane and by disrupting the proton motive force (Cochrane *et al.*, 2016). Tridecaptin A$_1$ binds to both Gram-positive and Gram-negative lipid II, although it has a much higher affinity for the Gram-negative analog (Cochrane *et al.*, 2014, Bann *et al.*, 2019). To assess if Oct-TriA$_5$ analogs bind to lipid II, we performed in vitro lipid II antagonization assays. Gram-positive lipid II, containing lysine at position 3 of the pentapeptide, was prepared by total chemical synthesis (Dong *et al.*, 2018, Karak *et al.*, 2024). The bioactivity of Oct-TriA$_5$ against *S. aureus* USA300 (MRSA) was evident with a MIC of 8 μg/mL. However, the addition of lipid II significantly reduced the efficacy of Oct-TriA$_5$, as the growth of *S. aureus* USA300 (MRSA) was not inhibited at the concentration of 8x MIC (Figure S5). This indicates that Oct-TriA$_5$ binds to lipid II of Gram-positive bacteria.

To monitor the effect of tridecaptins on the membrane permeability of Gram-positive and Gram-negative bacteria we performed a LIVE/DEAD staining assay with SYTO 9 and propidium iodide with *S. aureus* ATCC 29213 and *E.coli* ATCC 25922, as published previously (Tenconi *et al.*, 2018). Microscopic analysis demonstrated that the vast majority of *S. aureus* cells, when exposed to Oct-TriA$_5$ and Oct-TriD, exhibited red staining comparable to the positive control (Figure 3).

Table 1. In vitro minimum inhibitory concentrations (MICs) of Oct-TriA$_1$, Oct-TriA$_5$ and Oct-TriD[a]

Strain		Species	MIC			
			Oct-TriA$_1$ (1)	Oct-TriA$_5$ (2)	Oct-TriD(3)	Colistin
BV18	UAMS-1625	*S. aureus*	32	8	16	> 64
BV249	ATCC 29213	*S. aureus*	32	8	8	> 64
BV1402	1126387	*S. aureus*	32	8	16	> 64
BV29	ATCC 29212	*E. faecalis*	32	8	8	> 64
BV99	ATCC 25922	*E. faecalis*	32	8	8	> 64
BV144	IHMA 890472	*E. faecium*	16	8	4	> 64
BV159	ATCC 51559	*E. faecium*	32	8	4	> 64
BV94	IHMA 777621	*A. baumannii*	8	4	8	64
BV160	ATCC 17978	*A. baumannii*	8	4	16	1
BV374	HUMC1	*A. baumannii*	16	8	16	1
BV23	ATCC 25922	*E. coli*	2	2	32	1
BV1475	1220120	*E. coli*	2	4	16	0.5
BV1469	1214245	*E. coli*	2	4	16	32
BV306	ATCC 13883	*K. pneumoniae*	4	4	8	1
BV1445	1227947	*K. pneumoniae*	4	4	> 64	0.5
BV1447	1228586	*K. pneumoniae*	4	4	64	64
BV34	PAO1	*P. aeruginosa*	64	16	> 64	2
BV1551	1226072	*P. aeruginosa*	8	8	32	0.5
BV1544	1218019	*P. aeruginosa*	4	4	32	> 64
Hemolysis EC50 [μg/mL]			102.2	6.9	2.2	n/a
IC50 on HepG2 w/o FCS [μg/mL]			109.6	11.4	6.2	n/a

[a] MIC values reported in units of μg/mL

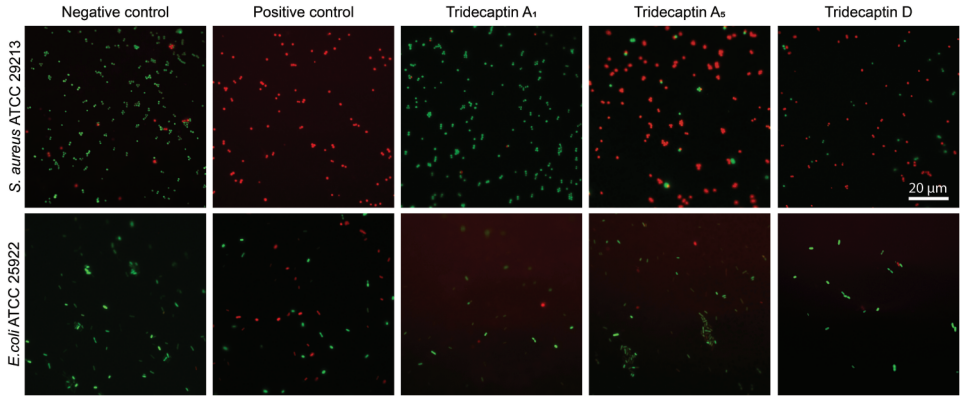


Figure 3. LIVE/DEAD staining of S. *aureus* ATCC 29213 and *E. coli* ATCC 25922 cells in the control conditions and after being exposed to nisin/polymyxin B or tridecaptin analogs Oct-TriA$_1$, Oct-TriA$_5$ and Oct-TriD. Green color indicates the cells with intact membrane, while cell with compromised membrane are stained in red. Note, after the treatment with Oct-TriA$_5$ and Oct-TriD *S. aureus* cells were stained red indicating the membrane disruption.

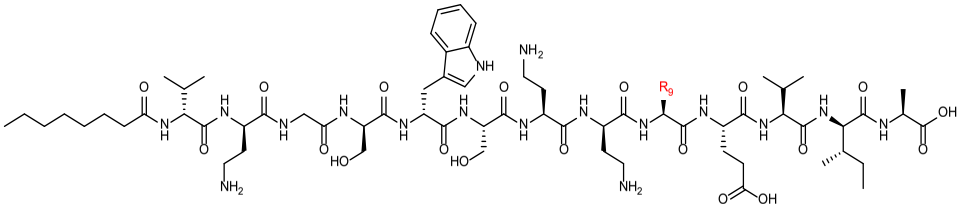
S. aureus cells incubated with Oct-TriA$_5$ and Oct-TriD had lysed, indicative of nonspecific membrane disruption, while those incubated with Oct-TriA$_1$ were still alive, suggesting that Oct-TriA$_1$ did not have any influence on the membrane permeability. Treatment of *E. coli* cells with Oct-TriA$_1$, Oct-TriA$_5$ or Oct-TriD did affect the membrane integrity, while after the same treatment with polymyxin (4 µg/mL; 2 × MIC) resulted in death of half of the *E.coli* cells. These data strongly suggest that Oct-TriA$_5$ and Oct-TriD act on Gram-positive bacteria by disrupting the cellular membrane. This is most likely explained by the presence of Trp9 in Oct-TriA$_5$ instead of Phe9 in Oct-TriA$_1$, that made the Oct-TriA$_5$ more hydrophobic and, therefore, increased its membrane disruption capacity.

Amino acid substitutions in tridecaptin A at position 9 impacting bioactivity and cytotoxicity

Considering the striking difference in the spectrum of activity between tridecaptin A$_1$ (active against Gram-negatives) and A$_5$ (active against both Gram-positives and Gram-negatives), we decided to look into the impact of changes in position 9 on bioactivity. To do so, we synthesized a series of analogs wherein residue 9 was changed into either Gly, Ala, Val, Ile, Ser, Tyr, Glu, His or Dab, and compared these new variants to Oct-TriA$_1$ and Oct-TriA$_5$ for bioactivity, as well as its hemolytic and cytotoxic activity (Table S4, Figures S6–S14).

The synthetic analogs exhibited significant variation in their MIC values against the pathogens tested (Table 2). Hemolytic and cytotoxic activities also differed significantly, which provided further proof of concept that indeed the residue in position 9 plays a critical role in the hemolytic and cytotoxic activity of tridecaptin A (Table 2).

Table 2. In vitro minimum inhibitory concentrations (MICs) of synthetic tridecaptin analogs [a].



Strain		MIC									
		Oct-Gly9 (4)	Oct-Ala9 (5)	Oct-Val9 (6)	Oct-Ile9(7)	Oct-Ser9 (8)	Oct-Tyr9 (9)	Oct-Glu9 (10)	Oct-His9 (11)	Oct-Dab9(12)	Colistin
UAMS-1625	*S. aureus*	> 64	> 64	> 64	> 64	> 64	> 64	> 64	> 64	> 64	> 64
ATCC 29213	*S. aureus*	> 64	> 64	> 64	64	> 64	> 64	> 64	> 64	> 64	> 64
1126387	*S. aureus*	> 64	> 64	> 64	> 64	> 64	> 64	> 64	> 64	> 64	> 64
ATCC 29212	*E. faecalis*	> 64	> 64	64	64	> 64	64	> 64	> 64	> 64	> 64
ATCC 25922	*E. faecalis*	> 64	> 64	> 64	64	> 64	64	> 64	> 64	> 64	> 64
IHMA 890472	*E. faecium*	> 64	> 64	> 64	32	> 64	32	> 64	> 64	> 64	> 64
ATCC 51559	*E. faecium*	> 64	> 64	64	64	> 64	64	> 64	> 64	> 64	> 64
IHMA 777621	*A. baumannii*	> 64	> 64	32	16	> 64	32	> 64	> 64	> 64	64
ATCC 17978	*A. baumannii*	> 64	> 64	32	16	> 64	32	> 64	> 64	> 64	1
HUMC1	*A. baumannii*	> 64	> 64	32	16	> 64	32	> 64	> 64	> 64	1
ATCC 25922	*E. coli*	16	16	4	2	32	2	> 64	16	> 64	1
1220120	*E. coli*	16	8	4	2	32	4	64	8	> 64	0.5
1214245	*E. coli*	16	16	4	2	64	2	> 64	16	> 64	32
ATCC 13883	*K. pneumoniae*	16	8	2	2	64	4	> 64	16	> 64	1
1227947	*K. pneumoniae*	32	16	8	4	> 64	4	> 64	32	> 64	0.5

1228586	*K. pneumoniae*	8	8	8	4	8	4	> 64	8	> 64	64
PAO1	*P. aeruginosa*	64	64	64	64	> 64	64	> 64	32	> 64	2
1226072	*P. aeruginosa*	32	32	16	16	32	4	> 64	4	8	0.5
1218019	*P. aeruginosa*	> 64	> 64	16	16	32	4	> 64	2	> 64	> 64
Hemolysis EC50 [μg/mL]		>128	>128	>128	>128	>128	65.6	>128	>128	>128	n/a
IC50 on HepG2 w/o FCS [μg/mL]		>128	>128	>128	110.2	>128	64	>128	>128	>128	>128

[a] MIC values reported in units of μg/mL

Of particular interest was Oct-His9, which efficiently inhibited colistin-resistant *P. aeruginosa* 1218019, with an MIC of 2 μg/mL. Furthermore, Oct-His9 showed lower hemolytic and cytotoxic activity compared to Oct-TriA_1. Conversely, synthetic analogs Oct-Gly9, Oct-Ala9, Oct-Ser9, Oct-Glu9 and Oct-Dab9 showed no antibacterial activity against Gram-positive bacteria at a concentration of 64 μg/mL and were moderately active against Gram-negative pathogens, with MICs ranging from 8 to 64 μg/mL. Similar to Oct-TriA_5 and Oct-TriD, variants Oct-Tyr9 and Oct-Ile9 showed moderate bioactivity against Gram-positives, and high cytotoxic and hemolytic activity. Synthetic analogs with Val and Ile at position 9 shared bioactivity, cytotoxic and hemolytic profiles with Oct-TriA_1. Taken together, our data highlight position 9 as a key residue for the spectrum and degree of bioactivity, as well as the cytotoxicity of tridecaptin A.

We here report the discovery of new lipopeptide antibiotics tridecaptin A_5 and tridecaptin D. Bioinformatics analysis of 785 complete *Paenibacillus* genomes revealed the BGC for tridecaptin A_5 in the genome of *Paenibacillus* sp. JJ-21. Thorough examination of the adenylation domains encoded by the tridecaptin A_5 BGC allowed us to predict the amino acid sequence of the peptide scaffold; this was then matched to the compound produced by *Paenibacillus* sp. JJ-21. Subsequent comparison of the tridecaptin A_5 BGC with that of tridecaptin A_1 revealed that they differed exclusively in module 9 of TriD. Based on the PARAS substrate specificity predictions we synthesized tridecaptin analogs Oct-TriA_5, Oct-TriD and reference Oct-TriA_1 to evaluate their antibacterial activity against a panel of ESKAPE pathogens. Surprisingly, while tridecaptins are primarily active against Gram-negative bacteria, Oct-TriA_5 possessed significant activity against both Gram-positive and Gram-negative bacteria, while Oct-TriD was primarily active against Gram-positive pathogens. Particularly striking is that a single amino acid substitution at position 9, from Phe in Oct-TriA_1 to Trp in Oct-TriA_5, changes the bioactivity profile from Gram-negative to broad-spectrum activity. Through the screening of analogs substituted at position 9, we subsequently identified the highly potent tridecaptin, Oct-His9, that exhibits effective activity against *P. aeruginosa* and, at the same time, possesses reduced hemolytic and cytotoxic properties. The observed decrease in cytotoxicity and the promising antibacterial properties, provide strong grounds for further investigation and development of these analogs as alternative antibiotics for combating infectious diseases associated with antibiotic-resistant bacterial pathogens.

Materials and Methods

Global genome mining of NRPSs

Genomes of all *Paenibacillus* spp. available from RefSeq (release 213) (O'Leary *et al.*, 2016) were downloaded from the NCBI FTP site. All genomes with more than 400 contigs were considered as low-quality assemblies and were removed from the collection. All genomes were analyzed using AntiSMASH (version 6.0.1) (Blin *et al.*, 2021) to obtain BGC predictions. These predictions were then used as input into BiG-SCAPE (version 1.1.4) (Navarro-Muñoz *et al.*, 2020), for the creation of a sequence similarity network, with distance matrix cutoff set to 0.25. The resulting full network was visualized by Cytoscape (3.9.1) (Shannon *et al.*, 2003). To predict the specificity of the adenylation domains of NRPS BGCs in this study, we used the software package PARAS (v0.0.4, available at https://github.com/BTheDragonMaster/paras). PARAS is an adenylation domain predictor that uses structure-guided sequence alignments to extract the active site prior to making a prediction.

Genome sequencing, assembly and annotation of *Paenibacillus* sp. JJ-21

Paenibacillus sp. JJ-21 was grown in tryptic soy broth (TSB) at 30 °C and 220 rpm for 24 h. DNA was extracted from *Paenibacillus* sp. JJ-21 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 this genome were annotated using AntiSMASH (version 6.0.1) (Blin *et al.*, 2021).

Growth of *Paenibacillus* sp. JJ-21 and specialized metabolites extraction

Paenibacillus sp. JJ-21 was obtained from the Auburn University Plant-Associated Microbial strain collection and had previously been isolated from the root surface of a field-grown corn plant (*Zea mays*) grown in Dunbar, Nebraska, and maintained as a viable cryostock in a -80 °C freezer. *Paenibacillus* sp. JJ-21 was grown at 30 °C on tryptic soy agar (TSA) for 72 h and two to three colonies were inoculated into TSB and incubated overnight at 30 °C. This inoculum (1%) was transferred to 1 L Erlenmeyer flasks containing 400 mL of sterile TSB and fermented at 30 °C while shaking at 220 rpm for 96 h. Cells were collected by centrifugation (5000 × *g*, 30 min, 4 °C) and extracted for 6 h with 70% isopropyl alcohol (IPA) supplemented with 0.1% (v/v) formic acid (FA). The crude extracts were clarified by centrifugation, then the solvent was evaporated under reduced pressure and reconstituted in 50% MeOH.

LC-MS/MS analysis

For LC-MS analyses, extracts were dissolved in 50% MeOH to a final concentration of 2 mg/mL, and 1 µL was injected into Shimadzu Nexera X2 UHPLC system coupled to a Shimadzu 9030 QTOF mass spectrometer, and data acquisition was performed as previously described (van Bergeijk *et al.*, 2022). Briefly, 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.

HRMS analyses of the synthetic peptides were performed on a Thermo Scientific Dionex UltiMate 3000 HPLC system with a Phenomenex Kinetex C_{18} (2.1 x 150 mm, 2.6 µm) column at 35 °C and equipped with a diode array detector. The following solvent system, at a flow rate of 0.3 mL/min, was used: solvent A, 0.1 % formic acid in water; solvent B, 0.1% formic acid in acetonitrile. Gradient elution was as follows: 95:5 (A/B) for 1 min, 95:5 to 5:95 (A/B) over 9 min, 5:95 to 2:98 (A/B) over 1 min, 2:98 (A/B) for 1 min, then reversion back to 95:5 (A/B) over 2 min, 95:5 (A/B) for 1 min. This system was connected to a Bruker micrOTOF-Q II mass spectrometer (ESI ionization) calibrated internally with sodium formate.

Peptide synthesis

Fmoc-L-Dab(Boc)-OH, Fmoc-D-Dab(Boc)-OH, Fmoc-D-Ile-OH, Fmoc-allo-Ile-OH and 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate (HATU) were purchased from Combi-Blocks. All other Fmoc-amino acids, the Fmoc-Ala-Wang resin and Fmoc-Glu(OtBu)-Wang resin were purchased from P3 BioSystems. 2-Chlorotrityl chloride resin (CTC) was purchased from Iris Biotech. Octanoic acid was purchased from Alfa Aesar. ((1H-Benzo[d][1,2,3]triazol-1-yl)oxy)tris(dimethylamino) phosphonium hexafluorophosphate (BOP), N,N-Diisopropylcarbodiimide (DIC) and triisopropylsilane (TIPS) were purchased from Manchester Organics. Diisopropylethylamine (DIPEA), piperidine, trifluoroacetic acid (TFA) and dimethyl sulfoxide (DMSO) were purchased from Carl Roth. Dichloromethane (CH_2Cl_2) and petroleum ether were purchased from VWR Chemicals. Acetonitrile (MeCN), dimethylformamide (DMF) and methyl tertiary-butyl ether (MTBE) were purchased from Biosolve.

Resin swelling

The resin was swollen in 10 mL of DMF for 300 s prior to the first coupling.

Automated coupling protocol

Step	Function	Duration/Temperature
1	Initial Deprotection N-terminus	15 s at 60°C then 30 s at 70°C
2	Deprotection N-terminus	15 s at 60°C then 180 s at 70°C
3	Wash (DMF)	RT
4	Wash (DMF)	RT
5	Wash (DMF)	RT
6	Coupling amino acid	15 s at 60°C then 300 s at 70°C

After coupling of the final residue on the synthesizer, the resin was washed with DCM, filtered and treated with 3 mL of TFA : TIPS : H_2O (95 : 2.5 : 2.5, v/v) for 90 min. The reaction mixture was filtered through cotton, the filtrate was precipitated in MTBE : petroleum ether (1 : 1, v/v) and centrifuged (4500 rpm, 5 min). The pellet was then resuspended in MTBE : petroleum ether (1 : 1, v/v) and centrifuged again (4500 rpm, 5 min). Finally the pellet containing the crude lipopeptide was dissolved in tBuOH : H_2O (1 : 1, v/v) and lyophilized overnight. The crude mixtures were subsequently purified by RP-HPLC. Fractions were assessed by HPLC and LC-MS and product containing fractions were pooled, frozen and lyophilized to yield the pure lipopeptides in >95% purity (determined by HPLC).

General procedure for manual solid phase peptide synthesis

The peptides were made on a 0.25 mmol scale on either preloaded Wang resin or 2-Chlorotrityl chloride (CTC) resin. The syntheses of Oct-TriA$_1$ and Oct-TriA$_5$ were performed on preloaded Fmoc-Ala-Wang resin (0.29 mmol/g loading). The synthesis of Oct-TriD was performed by loading Fmoc-Dab(Boc)-OH on CTC resin. Resin loading was determined to be 0.50 mmol/g. All couplings with the exception of Fmoc-d-*allo*-Ile-OH and Fmoc-d-Ile-OH were performed using 4 eq. of amino acid or fatty acid, benzotriazol-1-yloxytris(dimethylamino) phosphonium hexafluorophosphate (BOP) (4 eq.) and N,N-diisopropylethylamine (DiPEA) (8 eq.) in 10 mL of DMF for 1 h at RT, under nitrogen flow. Fmoc-d-*allo*-Ile and Fmoc-d-Ile-OH were coupled by treating the resin with 2 eq. of the amino acid, 2 eq. of BOP and 4 eq. of DiPEA in 10 mL of DMF overnight at RT. Fmoc group removal was performed by treating the resin with 10 mL of piperidine : DMF (1 : 4, v/v) for 5 min and then again for 15 min. Final sidechain deprotection and cleavage from the resin was carried out by treating the resin with 10 mL of TFA : TIPS : H_2O (95 : 2.5 : 2.5, v/v) for 90 min. The reaction mixture was filtered through cotton, the filtrate was precipitated in MTBE : petroleum ether (1 : 1, v/v) and centrifuged (4500 rpm, 5 min). The pellet was then resuspended in MTBE : petroleum ether (1 : 1, v/v) and centrifuged again (4500 rpm, 5 min). Finally the pellet containing the crude lipopeptide was dissolved in tBuOH:H_2O (1 : 1, v/v) and lyophilized overnight. The crude mixtures were subsequently purified by RP-HPLC. Fractions were assessed by HPLC and LC-

MS and product containing fractions were pooled, frozen and lyophilized to yield the pure lipopeptides in >95% purity (determined by HPLC).

General procedure for automated solid phase peptide synthesis

Position 9 analogs were made on a 0.05 mmol scale on preloaded Fmoc-Ala-CTC resin (0.68 mmol/g) using a CEM Liberty Blue automated peptide synthesizer with microwave irradiation. Couplings were performed at 0.125 M concentration using 5 eq. of amino acid, 5 eq. of 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate (HATU) and 10 eq. of diisopropylethylamine (DIPEA). Fmoc group removal was performed using piperidine : DMF (1 : 4, v/v). Cleavage from resin along with global deprotection and subsequent purification by RP-HPLC was performed as described above for the manually synthesized peptides.

Prep RP-HPLC purification

Peptides were purified using a BESTA-Technik system with a Dr. Maisch Reprosil Gold 120 C_{18} column (25 × 250 mm, 10 µm) and equipped with a ECOM Flash UV detector monitoring at 214 nm and 254 nm. The following solvent system, at a flow rate of 12 mL/min, was used: solvent A, 0.1 % TFA in water/acetonitrile 95/5; solvent B, 0.1 % TFA in water/acetonitrile 5/95. Gradient elution was as follows: 100:0 (A/B) for 5 min, 100:0 to 50:50 (A/B) over 50 min, 50:50 to 0:100 (A/B) for 3min, then reversion back to 100:0 (A/B) over 1 min, 100:0 (A/B) for 5 min.

Analytical RP-HPLC

Shimadzu Prominence-i LC-2030 system with a Dr. Maisch ReproSil Gold 120 C_{18} column (4.6 × 250 mm, 5 µm) at 30 °C and equipped with a UV detector monitoring 214 nm and 254 nm. The following solvent system, at a flow rate of 1 mL/min, was used: solvent A, 0.1 % TFA in water/acetonitrile 95/5; solvent B, 0.1 % TFA in water/acetonitrile 5/95. Method A: Gradient elution was as follows: 100:0 (A/B) for 2 min, 100:0 to 0:100 (A/B) over 23 min, 0:100 (A/B) for 1 min, 0:100 (A/B) then reversion back to 100:0 (A/B) over 1 min, 100:0 (A/B) for 3 min. Method B: Gradient elution was as follows: 100:0 (A/B) for 2 min, 100:0 to 50:50 (A/B) over 45 min, 50:50 (A/B) to 0:100 (A/B) over 1 min, 0:100 (A/B) for 6 min then reversion back to 100:0 (A/B) over 1min, 100:0 (A/B) for 5min. Method C: Gradient elution was as follows: 100:0 (A/B) for 2 min, 100:0 to 0:100 (A/B) over 53 min, 0:100 (A/B) for 1 min, 0:100 (A/B) then reversion back to 100:0 (A/B) over 1 min, 100:0 (A/B) for 4 min.

Antimicrobial testing

All minimum inhibitory concentrations (MICs) were determined according to Clinical and Standards Laboratory Institute (CLSI) guidelines. In brief, agar plates were inoculated from

glycerol stocks and incubated overnight at 35 °C. Peptides DMSO stocks were dispensed into 96-well round bottom plates using a TECAN D300e dispenser. Colistin sulfate (Sigma-Aldrich) was dissolved in water and serially diluted in cation-adjusted Mueller Hinton broth (10 µL per well). The final concentrations of the test compounds were 0.06 to 64 µg/mL. Inocula were prepared by direct colony suspension in NaCl at 0.5 MacFarland which was further diluted 200-fold in cation-adjusted Mueller Hinton broth for a target inoculum of 2 x 10^5 colony-forming units (CFUs)/mL. 100 µL of the prepared inoculum was added to each well of the prepared plates containing the test compounds. Plates were sealed with parafilm and incubated for 20 hours at 35 °C. Images of all plates were recorded and MIC was assessed visually.

Hemolysis assays

Peptides, dissolved in DMSO, were dispensed into 96-well round bottom microplates using a TECAN D300e dispenser and subsequently diluted with 100 µL PBS. PBS or 2 % Triton X-100 were used as negative and positive controls, respectively. Rabbit red blood cells (BioConcept) were diluted in PBS to a final concentration of 2 % and 100 µL were added to each well of the prepared 96-well plates. Final concentration of the test compounds were 0.25 to 128 µg/mL. Plates were incubated at 37 °C for 1 h and afterwards subjected to centrifugation at 1000 x *g* for 5 min. Then 30 µL of supernatant was transferred to a new round bottom plate and absorbance was measured at 405 nm (TECAN Infinite F200). EC50s were calculated after blank subtraction and normalization to the Triton control.

Cytotoxicity determination

HepG2 cells were grown in EMEM (Sigma) supplemented with 2 mM glutamine and 10 % fetal bovine serum (Fisher) at 37 °C, 5 % CO_2. 20'000 cells per well were seeded into clear tissue culture treated 96 well plates and incubated for 24 hours. Peptides, dissolved in DMSO, were dispensed into 96 deep well plates (TECAN D300e dispenser) and diluted to a final concentration of 1 to 128 µg/mL using fresh medium without serum. The next day, the old media was removed from the cells and 200 µL of the prepared compound dilutions were added per well. Plates were incubated for another 24 hours before assessing the cell viability using the CellTiter-Glo kit (Promega) according to the manufacturers protocol. IC50s were calculated after normalization to untreated control.

Lipid II antagonization assay

The minimum inhibitory concentrations (MICs) of Oct-TriA_1, Oct-TriA_5 and vancomycin against *S. aureus* USA300 (MRSA) were determined according to Clinical and Standards Laboratory Institute (CLSI) guidelines. Gram-positive lipid II, containing lysine at position 3

of the pentapeptide, was prepared by total chemical synthesis (Dong *et al.*, 2018, Karak *et al.*, 2024). Lipid II in 1 : 1 chloroform : methanol (v/v) was added to a polypropylene 96-well plate (5-fold molar excess compared to test antibiotics) and the organic solvent was allowed to evaporate. Oct-TriA$_5$ and vancomycin in MHB (50 µL, 16xMIC) were added to the wells with the 5-fold molar excess of pure Gram-positive lipid II in triplicates and to the control wells without lipid II. Single colony of *S. aureus* USA300 (MRSA) from a fresh blood agar plate was suspended in TSB and grown to an OD_{600} of 0.5. The bacterial culture then was diluted in MHB with 0.002% polysorbate-80 to reach 10^6 CFU/mL and 50 µL were mixed with test compounds to achieve the final concentration of 8 x MIC for all tested compounds. The plate was incubated at 37 °C for 18 h with constant shaking (600 rpm) and subsequently inspected for visible bacterial growth. To visualize the viability of the indicator strain, resazurin sodium salt solution (Merck, Dorset, UK) was added to the cultures to achieve the final concentration of 0.0015% (w/v) and incubated at room temperature for 1 h.

LIVE/DEAD staining and confocal microscopy

This assay was performed according to a previously described procedure (Li *et al.*, 2020, Zacchetti *et al.*, 2018). Briefly, overnight cultures of *E. coli* ATCC 25922 and *S. aureus* ATCC 29213 were diluted to an OD_{600} of 0.2 in MHB and mixed with 1 × MIC of Oct-TriA$_1$, Oct-TriA$_5$ and Oct-TriD. Nisin and polymyxin B were used as a positive controls at a concentration of two-fold MIC. At the same time, green fluorescent nucleic acid stain SYTO 9 (excitation, 450–490 nm; emission, 500–550 nm) and red fluorescent nucleic acid stain propidium iodide (excitation, 574–599 nm; emission, 612–682 nm) (LIVE/DEAD Baclight Bacterial Viability Kit, Invitrogen) were added to the above cell suspensions. After mixing and incubating for 15 min in the dark at room temperature, cells were briefly sedimented via centrifugation and resuspended in fresh MHB. Then the cell suspensions were loaded on 1.5% agarose pads and analyzed by Zeiss Imager M2 microscope. Confocal images were obtained at 4 random locations for each sample and visualized using FIJI version 1.51H (Schindelin *et al.*, 2012).

Acknowledgments

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. A-domain specificity analysis of NRPS BGCs from the genome of *Paenibacillus* sp. JJ-21.

Domain	Residues in the binding pocket	Amino acid prediction	Probability
Fusaricidin A1	DFWNIGMVH	Threonine	0.716357143
Fusaricidin A2	DAFWLGCTF	Valine	0.795809524
Fusaricidin A3	DASTLAGVC	Tyrosine	0.78
Fusaricidin A4	DFWNIGMVH	Threonine	0.716357143
Fusaricidin A5	DLTKIGEVG	Asparagine	0.96
Fusaricidin A6	DFPNFCIVY	Alanine	0.77
Polymyxin A1	DVGEISSID	2,4-diaminobutyric acid	1
Polymyxin A2	DFWNIGMVH	Threonine	0.716357143
Polymyxin A3	DVGEISSID	2,4-diaminobutyric acid	1
Polymyxin A4	DVGEISSID	2,4-diaminobutyric acid	0.99
Polymyxin A5	DVGEISSID	2,4-diaminobutyric acid	0.99
Polymyxin A6	DAWIVGAIV	Leucine	0.716357143
Polymyxin A7	DFWNIGMVH	Threonine	0.18
Polymyxin A8	DVGEISAID	2,4-diaminobutyric acid	0.716357143
Polymyxin A9	DVGEISAID	2,4-diaminobutyric acid	0.99
Polymyxin A10	DFWNIGMVH	Threonine	0.99
Tridecaptin A1	DAFWLGGTF	Valine	0.661904762
Tridecaptin A2	DVGEISSID	2,4-diaminobutyric acid	0.99
Tridecaptin A3	DILQMGMVW	Glycine	0.98

Tridecaptin A4	DVWHFSLVD	Serine	0.99
Tridecaptin A5	DAWAFAGVA	Tryptophan	0.63
Tridecaptin A6	DVWHFSLVD	Serine	0.99
Tridecaptin A7	DVGEISSID	2,4-diaminobutyric acid	0.99
Tridecaptin A8	DVGEISSID	2,4-diaminobutyric acid	0.99
Tridecaptin A9	DAWAFAGVA	Tryptophan	0.63
Tridecaptin A10	DAKDLGVVD	Glutamic acid	0.66
Tridecaptin A11	DAFWLGGTF	Valine	0.621571429
Tridecaptin A12	DAFFLGITF	Isoleucine	0.815
Tridecaptin A13	DVFWLGGTF	Alanine	0.72

Table S2. A-domain specificity comparison of reference tridecaptin A_1 BGC and tridecaptin cluster from the genome of *Paenibacillus* sp. JJ-21.

Domain	Residues in the binding pocket		Amino acid prediction	
	Tridecaptin A_1 reference BGC BGC0000449	Tridecaptin A_5 BGC of *Paenibacillus* sp. JJ-21	Tridecaptin A_1 reference BGC BGC0000449	Tridecaptin A_5 BGC of *Paenibacillus* sp. JJ-21
TriA1	DAFWLGGTF	DAFWLGGTF	Valine/Glycine	Valine
TriA2	DVGEISSID	DVGEISSID	2,4-diaminobutyric acid	2,4-diaminobutyric acid
TriA3	DILQMGMVW	DILQMGMVW	Glycine	Glycine
TriA4	DVWHFSLVD	DVWHFSLVD	Serine	Serine
Tri A5	DAWAFAGVA	DAWAFAGVA	Tryptophan /Phenylalanine	Tryptophan
TriA6	DVWHFSLVD	DVWHFSLVD	Serine	Serine
TriA7	DVGEISSID	DVGEISSID	2,4-diaminobutyric acid	2,4-diaminobutyric acid
TriA8	DVGEISSID	DVGEISSID	2,4-diaminobutyric acid	2,4-diaminobutyric acid
TriA9	DAWTFAGVA	DAWAFAGVA	Phenylalanine /Valine/ Isoleucine	Tryptophan
TriA10	DAKDLGVVD	DAKDLGVVD	Glutamic acid	Glutamic acid
TriA11	DAFWLGGTF	DAFWLGGTF	Valine	Valine
TriA12	DAFFLGITF	DAFFLGITF	Isoleucine	Isoleucine
TriA13	DVFWLGGTF	DVFWLGGTF	Alanine	Alanine

Table S3. Predicted amino-acid specificities of the adenylation domains present in representative BGCs of tridecaptin gene cluster families (GCFs) in a BiG-SCAPE sequence similarity network.

Accession number	GCF	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	Compound
NZ_CP092831.1. region009	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Phe	Glu	Val	D-Ile	Ala	Tridecaptin A_1
NZ_CP048793.1. region008	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Phe	Glu	Val	D-Ile	Ala	Tridecaptin A_1
NZ_CP061172.1. region008	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Phe	Glu	Val	D-Ile	Ala	Tridecaptin A_1
NZ_CP073683.1. region012	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Phe	Glu	Val	D-Ile	Ala	Tridecaptin A_1
NZ_CP097778.1. region009	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Phe	Glu	Val	D-Ile	Ala	Tridecaptin A_1
NZ_JTHP01000234.1. region001	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Phe	Glu	Val	D-Ile	Ala	Tridecaptin A_1
NZ_CP023711.1. region001	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_CP025957.1. region010	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	NA	Val	D-Ile	Ala	Tridecaptin A_5
NZ_CP084033.1. region008	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_CP086373.1. region009	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NC_023037.2. region009	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_CP011420.1. region010	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	NA	Tridecaptin A_5
NZ_CP011512.1. region010	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_CP015423.1. region015	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_CP017968.3. region011	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5

NZ_CP073682.1. region011	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_CP097767.1. region001	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_FOYG01000003.1. region001	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_JAFIBS010000001.1. region012	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_JMLR01000007.1. region003	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_NOLA01000018.1. region001	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_OXKC02000014.1. region001	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_POVS01000007.1 .region001	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
*Paenibacillus*_sp._MBT _JJ-21. region011	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_AMQU01000029.1. region001	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_CP040829.1. region014	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_CP049598.1. region019	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_CP097766.1. region009	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_CP097771.1. region013	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_JAACZY010000021.1. region001	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_JAAMNS010000007.1. region001	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_JNCB01000011.1. region006	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_JTHO01000008.1. region001	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5

NZ_JWJJ01000007.1. region001	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NZ_QVPU01000010.1. region001	1	D-Val	D-Dab	Gly	D-Ser	D-Trp	Ser	Dab	D-Dab	Trp	Glu	Val	D-Ile	Ala	Tridecaptin A_5
NC_015690.1. region011	2	D-Phe	D-Dab	D-Ile	D-Ser	D-Trp	Ser	Ser	D-Dab	Trp	Ser	Val	D-Ile	Dab	Tridecaptin D
NC_016935.1. region011	2	D-Phe	D-Dab	D-Ile	D-Ser	D-Trp	Ser	Ser	D-Dab	Trp	Ser	Val	D-Ile	Dab	Tridecaptin D
NC_017672.3. region011	2	D-Phe	D-Dab	D-Ile	D-Ser	D-Trp	Ser	Ser	D-Dab	Trp	Ser	Val	D-Ile	Dab	Tridecaptin D
NZ_JAM-AVM010000004.1. region001	3	D-Val	D-Dab	D-Ala	D-Dab	D-Trp	Ser	Asp	D-Dab	Trp	Asp	Val	D-Val	Gln	Tridecaptin G

Table S4. HPLC and HRMS analysis of peptides.

Peptide	Name	Chemical Formula	Calcd. Exact mass	Mass found	Calcd.	Overall yield [%]
1	Oct-TriA$_1$	$C_{72}H_{113}N_{17}O_{19}$	1519,8399	760,9278 $[M+2H]^{2+}$	760,9272	19
2	Oct-TriA$_5$	$C_{74}H_{114}N_{18}O_{19}$	1558,8508	780,4332 $[M+2H]^{2+}$	780,4327	23
3	Oct-TriD	$C_{80}H_{120}N_{18}O_{19}$	1636,8977	819,4565 $[M+2H]^{2+}$	819,4562	51
4	Oct-Gly9	$C_{65}H_{107}N_{17}O_{19}$	1429,7929	715,9046 $[M+2H]^{2+}$	715,9038	10
5	Oct-Ala9	$C_{66}H_{109}N_{17}O_{19}$	1443,8086	722,9120 $[M+2H]^{2+}$	722,9116	11
6	Oct-Val9	$C_{68}H_{113}N_{17}O_{19}$	1471,8399	736,9278 $[M+2H]^{2+}$	736,9272	11
7	Oct-Ile9	$C_{69}H_{115}N_{17}O_{19}$	1485,8555	743,9357 $[M+2H]^{2+}$	743,9351	16
8	Oct-Ser9	$C_{66}H_{109}N_{17}O_{20}$	1459,8035	730,9098 $[M+2H]^{2+}$	730,9090	16
9	Oct-Tyr9	$C_{72}H_{113}N_{17}O_{20}$	1535,8348	768,9256 $[M+2H]^{2+}$	768,9247	16
10	Oct-Glu9	$C_{68}H_{111}N_{17}O_{21}$	1501,8140	751,9150 $[M+2H]^{2+}$	751,9143	12
11	Oct-His9	$C_{69}H_{111}N_{19}O_{19}$	1509,8304	755,9232 $[M+2H]^{2+}$	755,9225	16
12	Oct-Dab9	$C_{67}H_{112}N_{18}O_{19}$	1472,8351	737,4254 $[M+2H]^{2+}$	737,4249	12

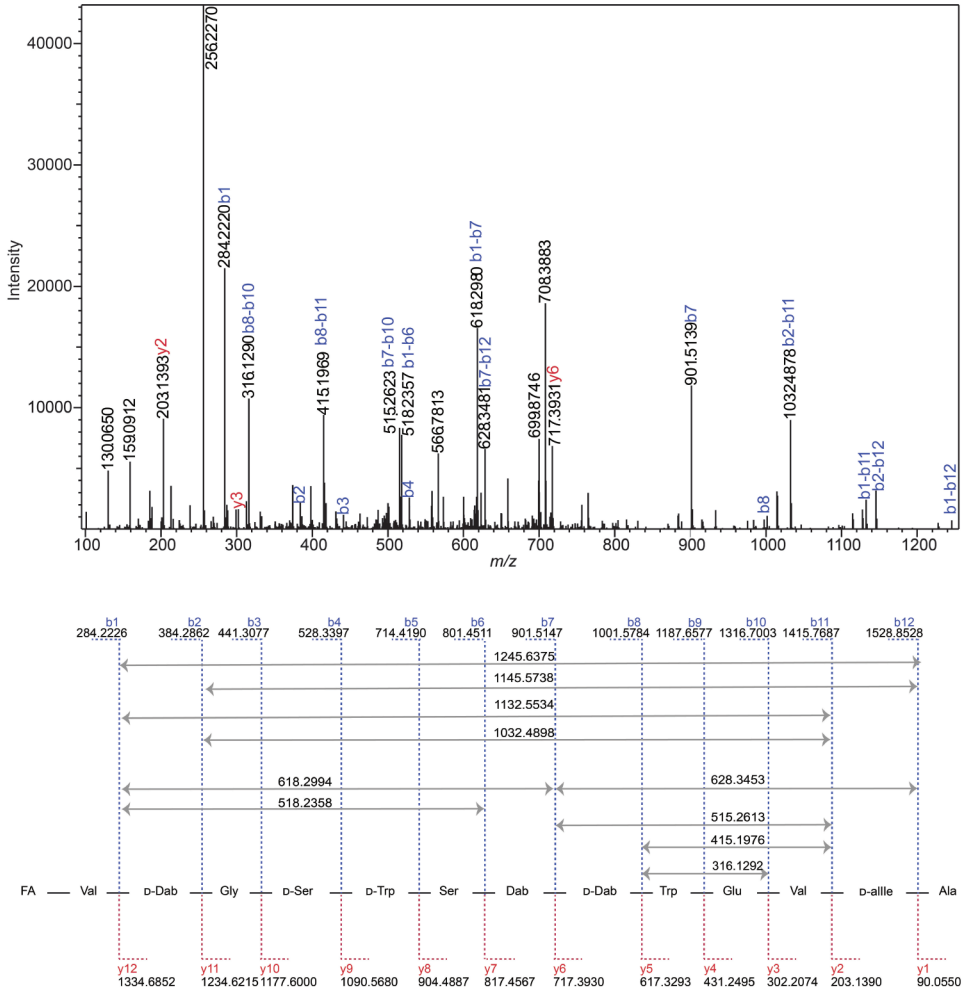


Figure S1. MS/MS spectrum of tridecaptin A_5 produced by of *Paenibacillus* sp. JJ-21 (precursor ion $[M+3H]^{3+}$ *m/z* 539.9712). The assignment of the sequence of amino acid residues at the top of the spectrum is based on the mass differences between the consecutive *y* and *b* ions.

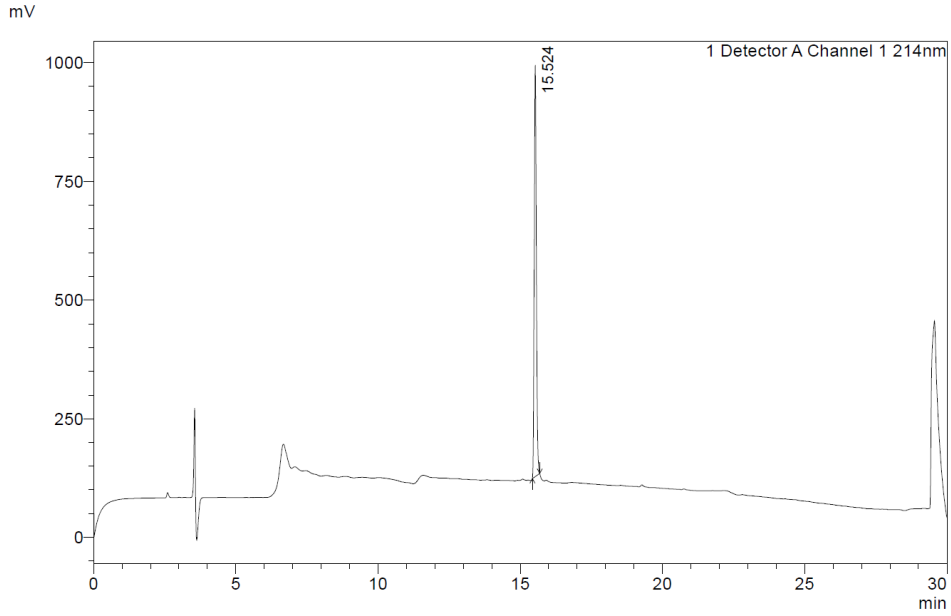


Figure S2. HPLC trace showing the reinjection of purified Oct-TriA$_1$ (1). The peptide eluted as a single peak at 15.524 min using the HPLC method A.

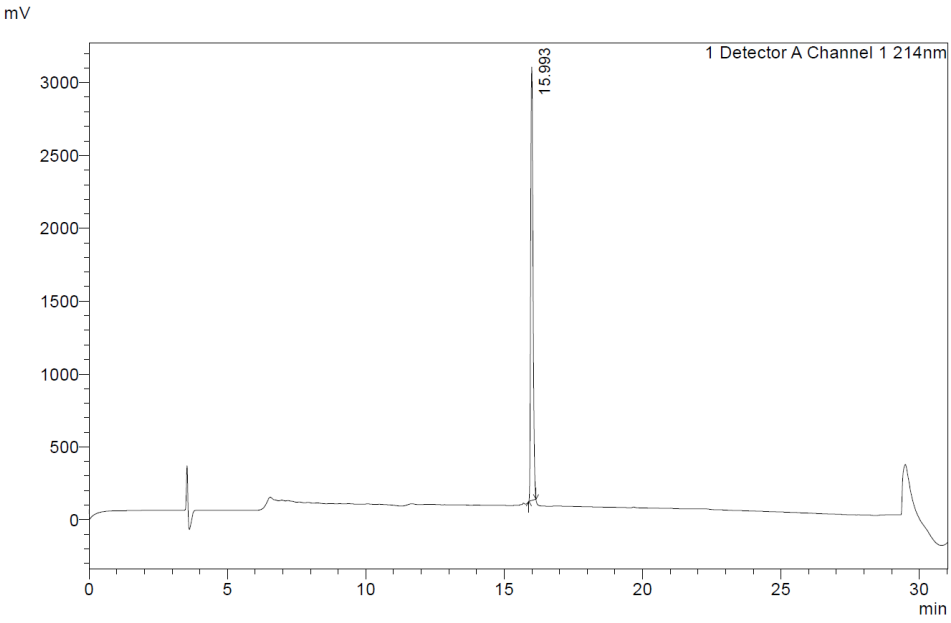


Figure S3. HPLC trace showing the reinjection of purified Oct-TriA5 (2). The peptide eluted as a single peak at 15.993 min using the HPLC method A.

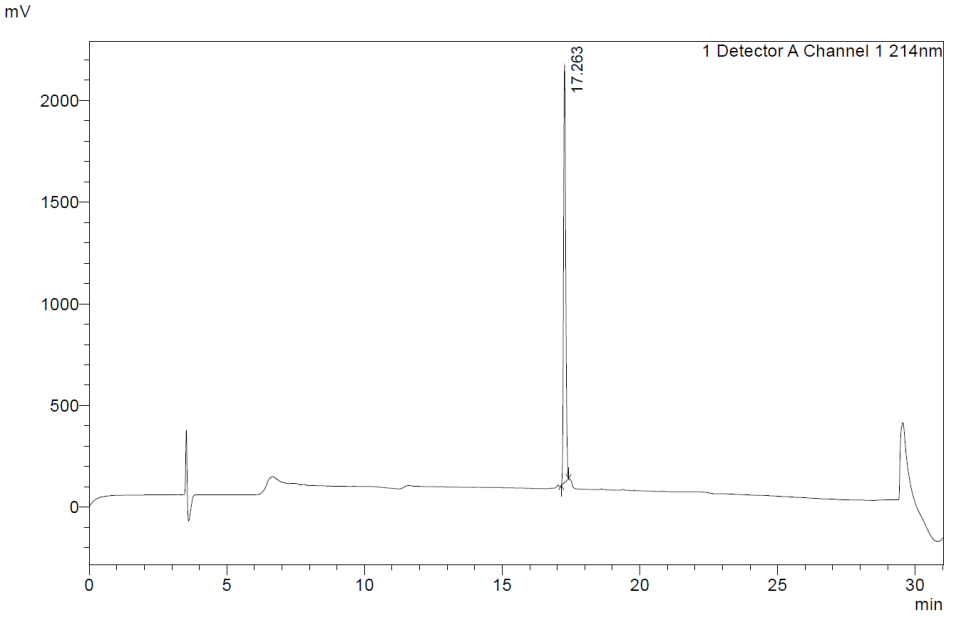


Figure S4. HPLC trace showing the reinjection of purified Oct-TriD (3). The peptide eluted as a single peak at 17.263 min using the HPLC method A.

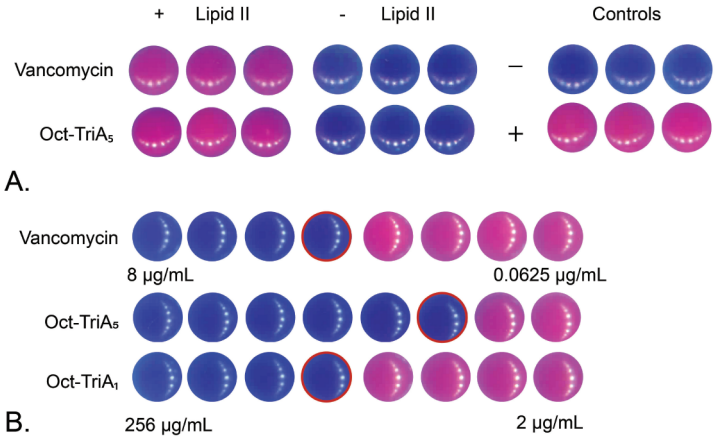


Figure S5. A. Gram-positive lipid II binding assays with *S. aureus* USA300 (MRSA) to show lipid II binding of Oct-TriA$_5$. Vancomycin was used as a positive control for lipid II binding. Resazurin was used to visualize the viability of the indicator strain (pink colour indicates growth and blue means inhibition of growth). Note, that addition of lipid II significantly reduced the efficacy of Oct-TriA$_5$ and vancomycin, as the growth of *S. aureus* USA300 (MRSA) was not inhibited at the concentration of 8x MIC. This indicates that Oct-TriA$_5$ binds to lipid II of Gram-positive bacteria. **B.** MIC assays for Oct-TriA$_1$, Oct-TriA$_5$ and vancomycin against *S. aureus* USA300. The wells with the MIC of the test compounds are highlighted with red edge.

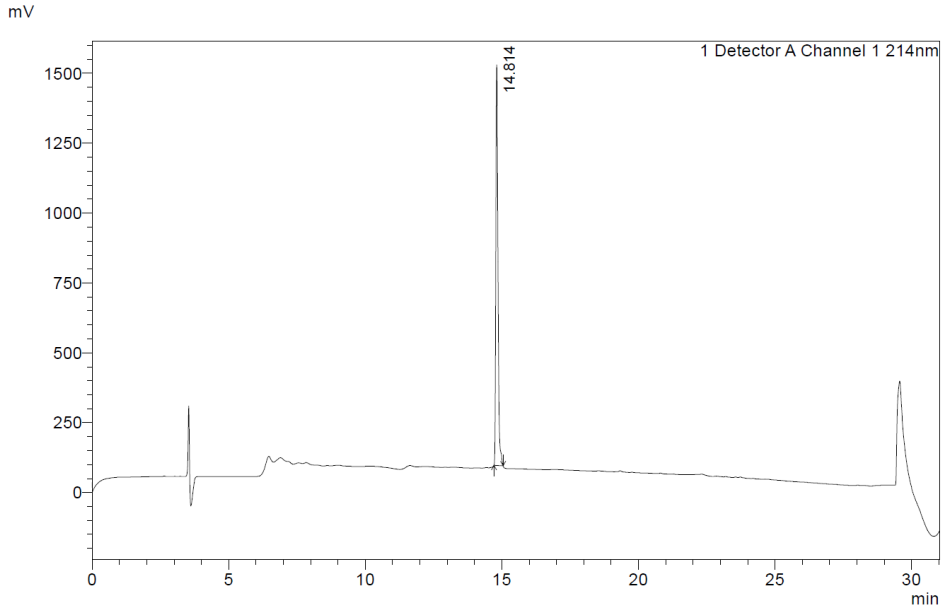


Figure S6. HPLC trace showing the reinjection of purified Oct-Gly9 (4). The peptide eluted as a single peak at 14.814 min using the HPLC method A.

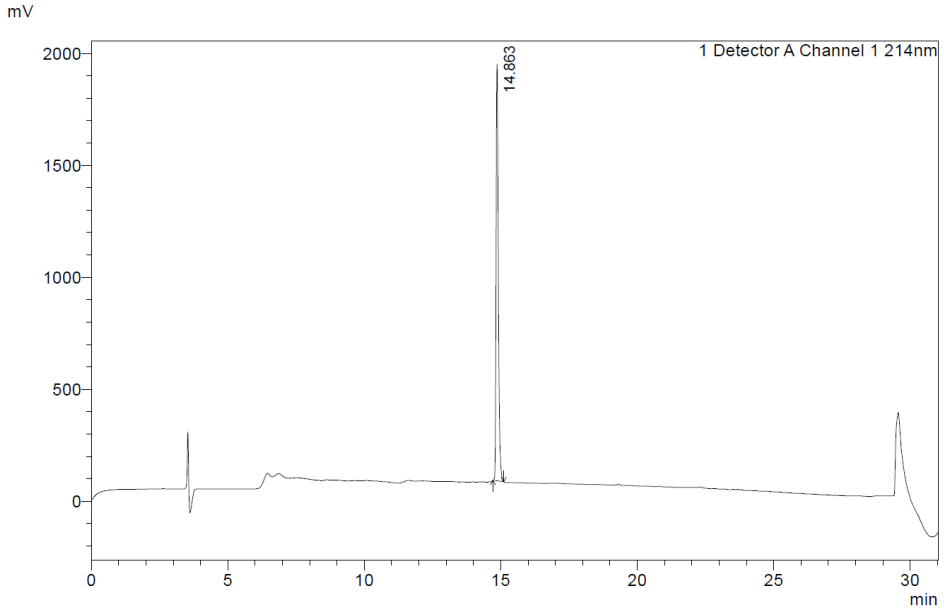


Figure S7. HPLC trace showing the reinjection of purified Oct-Ala9 (5). The peptide eluted as a single peak at 14.863 min using the HPLC method A.

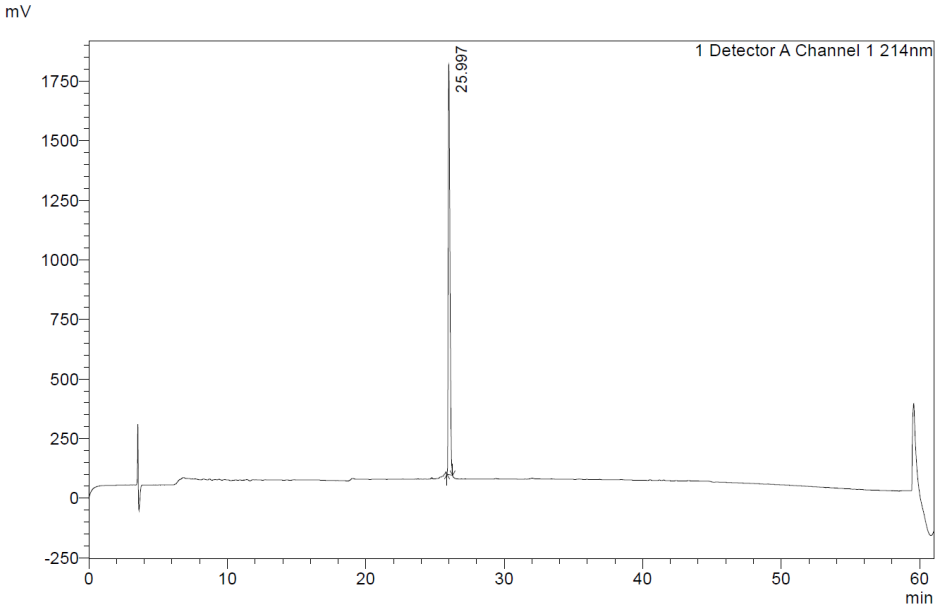


Figure S8. HPLC trace showing the reinjection of purified Oct-Val9 (6). The peptide eluted as a single peak at 25.997 min using the HPLC method C.

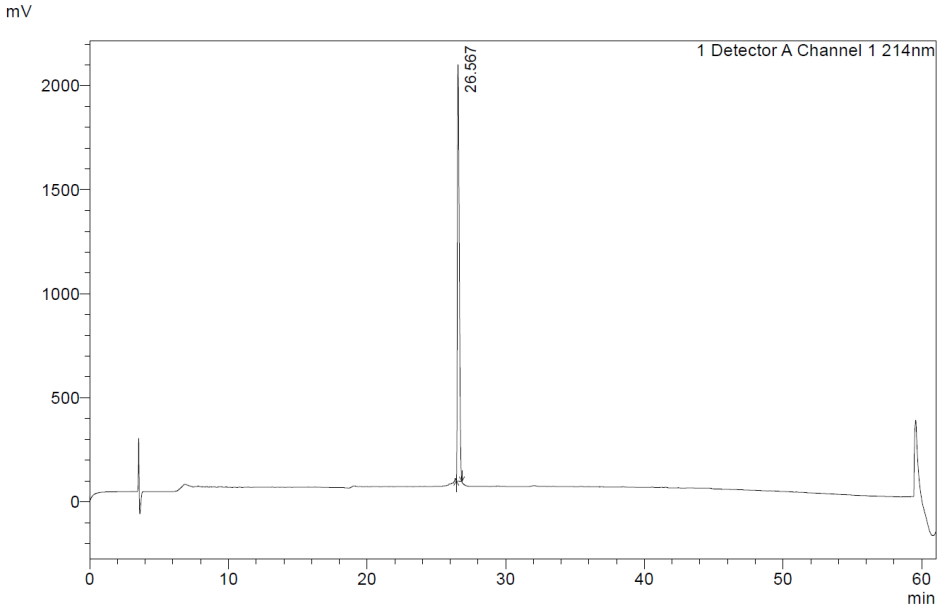


Figure S9. HPLC trace showing the reinjection of purified Oct-Ile9 (7). The peptide eluted as a single peak at 26.567 min using the HPLC method C.

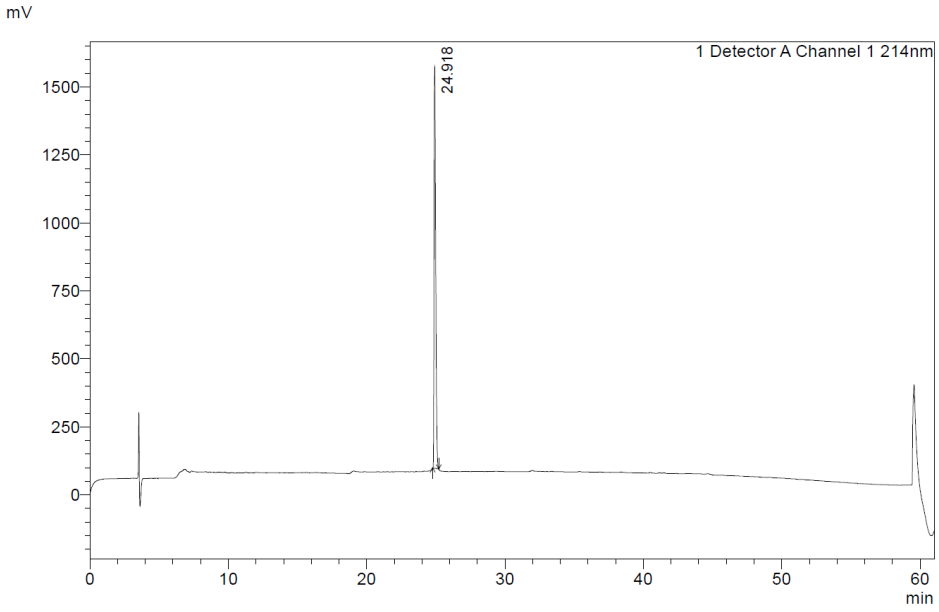


Figure S10. HPLC trace showing the reinjection of purified Oct-Ser9 (8). The peptide eluted as a single peak at 24.918 min using the HPLC method C.

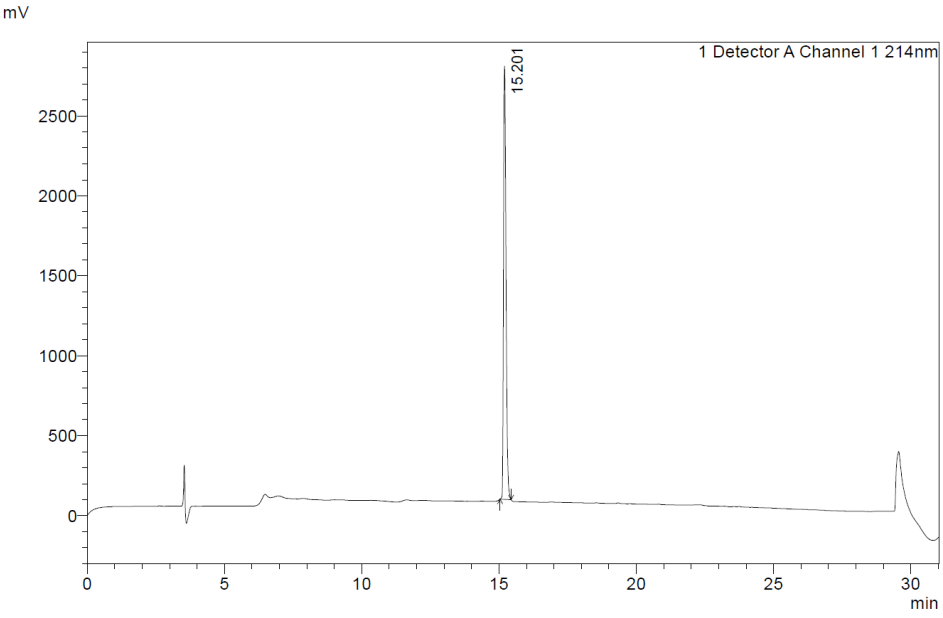


Figure S11. HPLC trace showing the reinjection of purified Oct-Tyr9 (9). The peptide eluted as a single peak at 15.201 min using the HPLC method A.

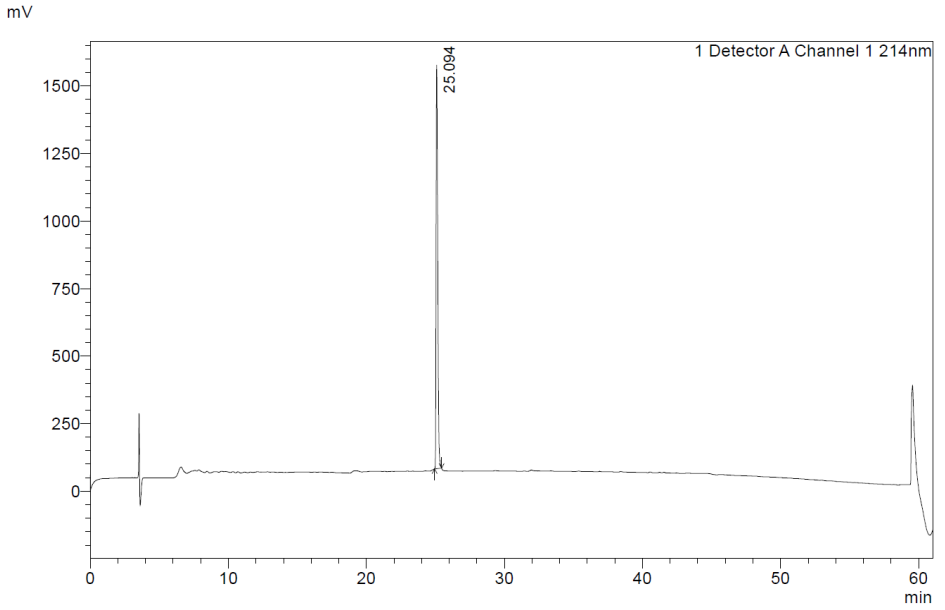


Figure S12. HPLC trace showing the reinjection of purified Oct-Glu9 (10). The peptide eluted as a single peak at 25.094 min using the HPLC method C.

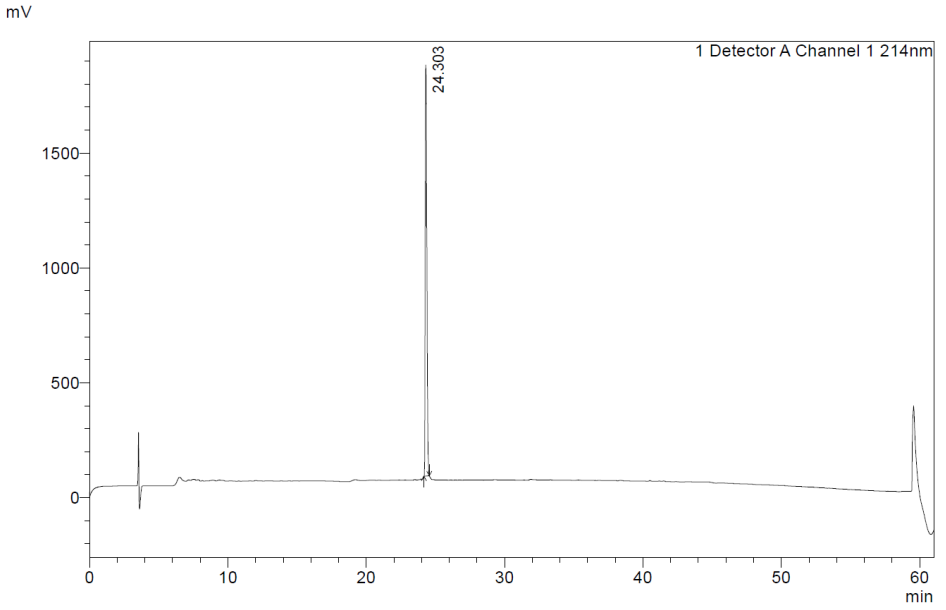


Figure S13. HPLC trace showing the reinjection of purified Oct-His9 (11). The peptide eluted as a single peak at 24.303 min using the HPLC method C.

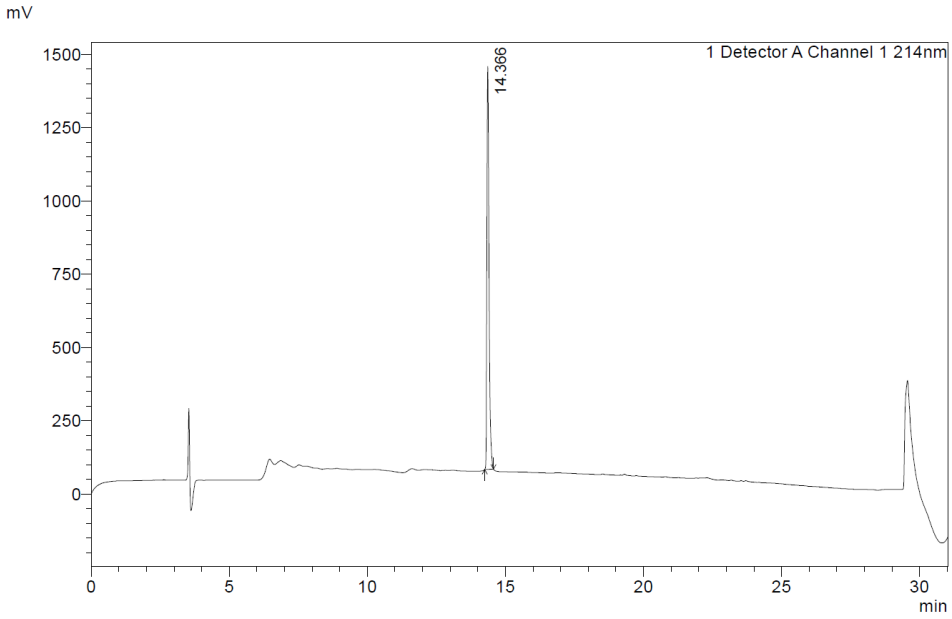
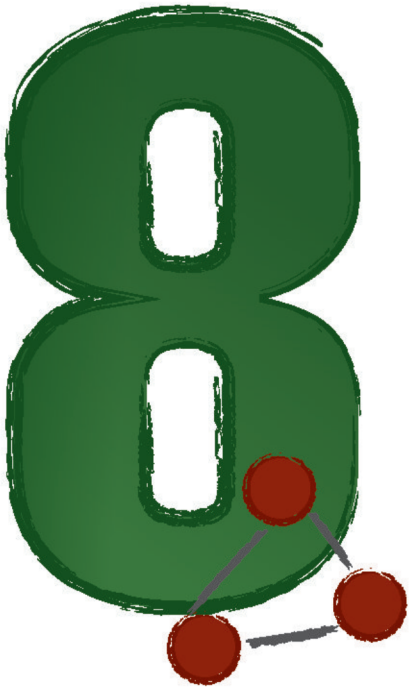


Figure S14. HPLC trace showing the reinjection of purified Oct-Dab9 (12). The peptide eluted as a single peak at 14.366 min using the HPLC method A.



Chapter 8

Chimeric biosynthesis unveils diverse tridecaptin variants in the *Paenibacillus* genus

Nataliia V Machushynets, Barbara R. Terlouw, Chao Du, Karol Al Ayed, Somayah S Elsayed, Julian Schill, Vincent Trebosc, Michel Pieren, Mark R. Liles, Marnix H Medema, Nathaniel I Martin, Gilles P van Wezel

Abstract

Modern sequencing technologies offer deep insight into the vast potential of *Paenibacillus* species to produce an extensive range of nonribosomal peptides (NRPs). Despite the expanding knowledge of the diverse microbial strategies used to produce these natural products, translating genomic information into bioactive lead compounds remains a complex challenge. Here, we show that chimeric biosynthesis occurs within the family of tridecaptin antibiotics. Through genome mining, numerous new tridecaptin biosynthetic gene clusters (BGCs) were identified in *Paenibacillus* genomes. Surprisingly, 15 strains harbor both a full tridecaptin BGC and a second BGC that specifies a truncated decamer. The encoded NRPS domain architectures suggested the capability of these strains to produce multiple tridecaptin variants through collaborative action between this tridecaptin-like NRPS and a shorter NRPS encoded within the 'main' tridecaptin BGC. Indeed, *Paenibacillus* sp. JJ-1683 was found to produce both tridecaptin A_5 and tridecaptin B, providing an unusual example of chimeric biosynthesis of NRPS lipopeptide antibiotics. These findings reveal chimeric biosynthesis as a strategy enabling strains to dynamically produce similar compounds with distinct chemistry and bioactivity profiles.

Introduction

Demand for new antibiotics is at an all-time high due to the emergence of antimicrobial resistance among human pathogens that poses a serious but largely unchallenged global health threat (WHO, 2019). With the advent of increasingly affordable DNA sequencing technologies and the development of advanced bioinformatic tools, bacterial genome mining for novel natural products is poised to play an important role in addressing this threat (Baltz, 2021, Palazzotto & Weber, 2018). In recent years, genome mining has revealed evidence for an expansive trove of yet-undiscovered drug-like molecules hidden in the genomes of biosynthetically talented bacterial genera (Katz & Baltz, 2016, Medema *et al.*, 2021, Scherlach & Hertweck, 2021). One such genus is *Paenibacillus*, which includes many bacterial species commonly found in soil that are often associated with plant roots. Members of this genus can offer protection against insects, herbivores, and phytopathogens, including other bacteria, fungi, nematodes, and viruses, by producing a wealth of natural products (Olishevska *et al.*, 2019, Grady *et al.*, 2016).

Genes encoding biosynthetic pathways for natural products are frequently clustered in bacterial genomes (Medema & Fischbach, 2015). With the help of genome mining tools such as antiSMASH, such biosynthetic gene clusters (BGCs) can be identified in microbial (meta) genomes and classified based on the type of molecule they are predicted to produce, such as polyketides, ribosomally synthesized and post-translationally modified peptides (RiPPs), terpenes, and nonribosomal peptides (Blin *et al.*, 2023). For some natural product classes, it is well-understood how the encoded biosynthetic machinery builds a natural product, and thus, the core scaffold structure of the produced natural product can be (partially) predicted. This is especially true for nonribosomal peptides (NRPs): short peptides that are oligomerized in an mRNA- and ribosome-independent manner. Members of the *Paenibacillus* genus produce a plethora of NRPs, including polymyxins, octapeptins, paenibacterins, fusaricidins, and tridecaptins (Cochrane & Vederas, 2016).

NRP synthetase (NRPS) assembly lines select and condense amino acids iteratively to assemble peptides (Sieber & Marahiel, 2005). To achieve nonribosomal peptide synthesis, minimal enzymatic complexes require four distinct domains called core or essential domains: condensation (C), adenylation (A), peptidyl carrier protein (PCP) and a terminal thioesterase (TE) domain (Marahiel, 2016). The C, A, and PCP domains make up functional modules, with C domains catalyzing condensation, A domains mediating building block selection through adenylation and subsequent thiolation, and non-catalytic PCP domains functioning as mechanical arms to which intermediate scaffolds are tethered between reactions. An NRPS assembly line only requires a single TE domain, typically at the end of an NRPS enzyme, which is responsible for chain release. As A domains govern building block selection, many NRP structure prediction algorithms attempt to predict the selected substrate from the A

domain sequence. Examples of such machine learning algorithms include NRPSpredictor2 (Röttig *et al.*, 2011), AdenylPred (Mongia *et al.*, 2023), SANDPUMA (Chevrette *et al.*, 2017), and AdenPredictor (Mongia *et al.*, 2023), which were trained on A domain sequences of known specificity (Jenke-Kodama & Dittmann, 2009).

However, translating genomic information into predicted molecular structures remains challenging (Udwary *et al.*, 2021). The diversity of NRPS-derived peptides is not only determined by the incorporation of different amino acid building blocks but also by an ever-increasing number of characterized tailoring reactions and enzymatic domains (Wenski *et al.*, 2022). In addition, a constantly expanding library of scientific literature describes the unraveling of new examples of nonribosomal peptide synthetases exhibiting rare domains and noncanonical domain and module organizations, leading to unique microbial strategies to synthesize novel natural products (Duban *et al.*, 2022).

One notable example entails collaborative natural product biosynthesis, involving crosstalk between two separate BGCs. Such pairing of biosynthetic enzymes from distinct BGCs within the same biosynthetic pathway can give rise to increased chemical diversity and an expanded scope of biological activity. Over the past few years, this phenomenon of collaborative biosynthesis has received increased attention, and specialized metabolites produced through such strategies have been referred to in the literature as 'hybrid molecules' or 'chimeras' (Mevers *et al.*, 2019, Yin *et al.*, 2022). Chimeric biosynthesis has previously been observed in various NRPS and PKS (hybrid) systems, including the biosynthesis of leucanicidin and endophenaside C (Wu *et al.*, 2016b), actinomycin L (Machushynets *et al.*, 2022), azaphilone (Huang *et al.*, 2020), penilactones (Fan *et al.*, 2019), spirotryprostatin A (Yan *et al.*, 2019), and echinocandin B (Cacho *et al.*, 2012).

In a prior study, we applied global genome mining for an in-depth examination of the diversity of NRPS BGCs across 785 *Paenibacillus* genomes. BiG-SCAPE analysis revealed numerous uncharacterized gene cluster families (GCFs), which led to the discovery of the new tridecaptin variants A_5 and D. Tridecaptins are linear nonribosomal antibacterial peptides produced by various *Paenibacillus* spp., with potent activity against multidrug-resistant Gram-negative bacteria (Cochrane *et al.*, 2014, Cochrane *et al.*, 2015, Jangra *et al.*, 2019b). Tridecaptins are open-chain tridecapeptides with a β-hydroxy fatty acid tail at the N terminus. Numerous natural tridecaptin variants (tridecaptins A, B, C, M, and G) differ from each other in the fatty acid tail and/or amino acid composition (Bann *et al.*, 2021). Typically, tridecaptin BGCs comprise five genes encoding different proteins that are responsible for the synthesis and secretion of the active peptide: TriA (a putative thioesterase), TriB and TriC (ABC transporters), TriD (a 10-module NRPS) and TriE (a 3-module NRPS) (da Costa *et al.*, 2022). The peptide scaffold is assembled by the cooperative action of TriD and TriE, with TriD incorporating the first ten amino acids and TriE incorporating the last three (Bann *et al.*, 2021).

Here, we present evidence for the widespread chimeric biosynthesis of tridecaptin variants in the genus *Paenibacillus* through a combination of genome mining and experimental validation. We found that 15 different *Paenibacillus* genomes harbored both a full-length tridecaptin BGC and a separate, tridecaptin-like NRPS at a different locus. We hypothesized that these strains are capable of producing multiple tridecaptin variants through chimeric biosynthesis. Specifically, we propose that one variant is produced by the full-length BGC and the second by collaborative action between the tridecaptin-like NRPS at the secondary locus and the shorter NRPS in the full-length BGC. We verified this hypothesis by sequencing the genome of *Paenibacillus* JJ-1683, a strain which we experimentally showed to produce both tridecaptin A_5 and tridecaptin B_1. We demonstrated that *Paenibacillus* JJ-1683 contains both loci that match the bioinformatics predictions for the NRPS enzymes involved in the speculated biosynthetic routes for both tridecaptins. As such, this work provides an example of chimeric biosynthesis of NRPS lipopeptide antibiotics, and hints at chimeric biosynthesis as a strategy to swap between the production of two similar compounds with slightly different chemistry and bioactivity profiles.

Results and Discussion

Bioinformatics prediction of novel tridecaptin BGCs

We previously demonstrated that members of the genus *Paenibacillus* have great biosynthetic potential to produce NRP-like natural products. BiG-SCAPE (Navarro-Muñoz *et al.*, 2020) networking analysis revealed gene cluster families (GCFs) of numerous unknown NRPS-like BGCs alongside previously discovered polymyxins, paenilipoheptins, fusaricidins, paenibacterins, octapeptins, bacitracins, cilagicins, and tridecaptins (**Chapter 7**). To characterize the putative structures of the tridecaptins produced by *Paenibacillus* sp. strains, we first identified all tridecaptin-associated GCFs from our BiG-SCAPE analysis (Figure 1A). This yielded 7 GCFs that comprise tridecaptin-associated BGCs across 117 strains. Notably, some of these BGCs were incomplete, as they were found on the contigs of heavily fragmented genomes, and full tridecaptin clusters could not always be derived.

To maximize the discovery of tridecaptin BGCs, including those that did not cluster into GCFs, we identified all adenylation-domain-containing proteins encoded by the BGCs of the 785 publicly available *Paenibacillus* genomes. We then extracted the A domains and their active site signatures and predicted their substrate selectivity with the adenylation domain substrate prediction algorithm PARAS (Table S1). BGCs with ten or more modules were retained, and from these, putative tridecaptin assembly lines with associated core structure predictions were manually assigned based on building block predictions and the position of modules containing epimerization domains. This yielded 79 putative tridecaptin assembly lines (Table S2). Of these, 53 were full-length clusters containing all 13 modules, while the other 26 had one or more missing modules. Among the 53 full-length clusters, 10 were found to be unique.

Next, we computationally cross-referenced the predicted structures with all known tridecaptin peptide scaffolds (Table 1) and found full-length predicted scaffolds corresponding to tridecaptins G_4, A_1/A_3, A_5 and D (Table S2). Additionally, we predicted the structures of nine potentially novel tridecaptins, seven of which were full-length (Figure 1B; Table S2). This included two putative variants of tridecaptin A encoded in three *Paenibacillus kribbensis* strains, which are predicted to contain an alanine at position 3 as only previously observed in tridecaptin G (da Costa *et al.*, 2022). Notably, the second variant encoded by two of the strains was predicted to incorporate a glycine moiety in module 6, which recognizes a serine in all known tridecaptin variants. Other predicted novel tridecaptins include a tridecaptin B_γ/C_{β_1} variant in *Paenibacillus peoriae* KCTC 3763, and a tridecaptin A_5 variant encoded in seven strains of *Paenibacillus polymyxa*. Finally, we found a putative tridecaptin M variant encoded in *Paenibacillus* sp. M-152 and *Paenibacillus* sp. lzh-N1, for which our predictions matched the six characterized amino acids of tridecaptin M_{11}. Tridecaptin M_{11} was originally isolated from *Paenibacillus* sp. M-152 in a very low yield, and its amino acid sequence was only partially characterized through MS/MS analysis (Jangra *et al.*, 2019a). We here propose the sequence of its full amino acid scaffold as D-Val–D-Dab–D-Ser–D-Gly–D-Trp–L-Ser–L-Dab–D-Dab–L-Phe–L-Glu-L-Ile–D-Ile–L-Ser, based on our genome mining for the producing strain.

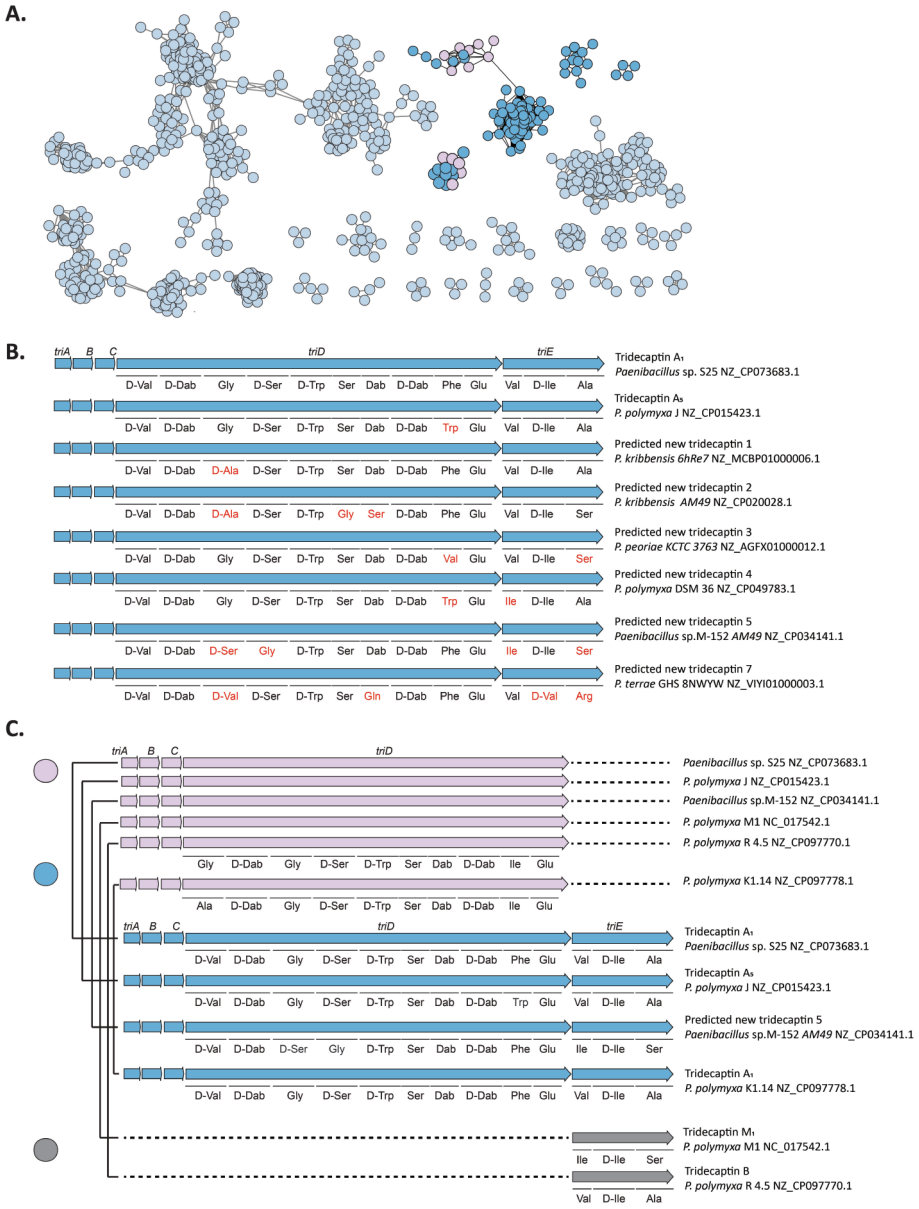


Figure 1. Sequence similarity network of NRPS BGCs and tridecaptin BGC analysis. A. BiG-SCAPE sequence similarity network (SSN) and tridecaptin-like GCFs visualized in Cytoscape (c 0.25). Each node represents one tridecaptin BGC identified by antiSMASH 6.0. Blue: full-length tridecaptin BGCs. Pink: truncated tridecaptin BGCs. **B.** New tridecaptin variants predicted in this study. The A domains and their active site signatures were extracted, and their substrate selectivity was predicted using the adenylation domain substrate prediction algorithm PARAS. BGCs with ten or more modules were retained, and from these, putative tridecaptin clusters assembly lines with associated core structure predictions were manually assigned. The amino acids highlighted in red differ from those in the structures of known tridecaptins. **C.** Gene architectures and amino acid scaffolds of full-length (blue) and truncated ten-(pink) and three-modular (grey) tridecaptin BGCs. Black connectors indicate BGCs that co-occur in the same genome. Collaboration between the two clusters, where the truncated cluster provides the first 10 amino acids and the full-length cluster provides the last three, leads to tridecaptin B production.

Table 1. Amino acid scaffolds of known tridecaptins.

Compound	1[D]	2[D]	3[D]	4[D]	5[D]	6	7	8[D]	9	10	11	12[D]	13
$TriA_1/A_3$	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Phe	Glu	Val	Ile	Ala
$TriA_2$	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Phe	Glu	Val	Val	Ala
$TriA_4$	Val	Dab	Gly	Ser	Phe	Ser	Dab	Dab	Phe	Glu	Val	Ile	Ala
$TriA_5$	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala
$TriB_1$	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	Val	Ile	Ala
TriBα	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	Val	Ile	Ser
TriBβ / M_2	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	Val	Val	Ser
TriBγ	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Val	Glu	Val	Val	Ser
TriBδ	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Val	Glu	Val	Ile	Ser
TriCα1/Cα2	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Phe	Glu	Val	Val	Ser
TriCβ1	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Phe	Glu	Val	Ile	Ser
TriD	Phe	Dab	Ile	Ser	Trp	Ser	Ser	Dab	Trp	Ser	Val	Ile	Dab
TriE	Ala	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala
$TriG_1$	Ile	Dab	Ala	Dab	Trp	Ser	Asp	Dab	Phe	Asp	Val	Val	Arg
$TriG_1$	Leu	Dab	Ala	Dab	Trp	Ser	Asp	Dab	Phe	Asp	Val	Val	Arg
$TriG_2$	Val	Dab	Ala	Dab	Trp	Ser	Asp	Dab	Phe	Asp	Val	Val	Arg
$TriG_3$	Ile	Dab	Ala	Dab	Trp	Ser	Asp	Dab	Trp	Asp	Val	Val	Arg
$TriG_3$	Leu	Dab	Ala	Dab	Trp	Ser	Asp	Dab	Trp	Asp	Val	Val	Arg
$TriG_4$	Val	Dab	Ala	Dab	Trp	Ser	Asp	Dab	Trp	Asp	Val	Val	Arg
$TriM_1$	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	Ile	Ile	Ser
$TriM_5$	Gly	Dab	Gly	Ser	Phe	Ser	Dab	Dab	Ile	Glu	Ile	Val	Ser
$TriM_6$	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	Ile	Val	Ser
$TriM_7$	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Val	Glu	Ile	Ile	Ser
$TriM_8$	Gly	Dab	Gly	Ser	Phe	Ser	Dab	Dab	Ile	Glu	Ile	Ile	Ser
$TriM_{11}$	ND	ND	ND	ND	ND	ND	ND	Dab	Phe	Glu	Ile	Ile	Ser

[D] Modules incorporate a D-amino acid. ND, not determined.

Discovery of tridecaptin BGCs with unusual architecture

Some of the detected BGCs differed from the reference tridecaptin A_1 not only by the predicted amino acid composition but also by their architecture. These truncated BGCs contained the ten-modular NRPS gene corresponding to the *triD* gene but lacked the terminal 3-module *triE* gene (Figure 1C). Importantly, these truncated tridecaptin clusters were not predicted to encode the TE domain that is usually located in the TriE synthase and ensures the release of the final product from its carrier protein. Besides the genome sequences being of high quality, the truncated ten-module tridecaptin clusters always co-occurred with a full-length BGC in the same genome that was predicted to encode a different tridecaptin (Figure 1C). The seven remaining ten-modular tridecaptin BGCs, which were not flagged by BiG-SCAPE as part of this family, were also co-localized either with full-length tridecaptin BGCs or with another truncated tridecaptin cluster that only contained a predicted *triE* ortholog.

Interestingly, the predicted amino acid scaffolds for some of the truncated tridecaptins perfectly matched the N-terminal parts of the scaffolds of previously isolated tridecaptins B_1/M_1. These truncated BGCs always co-localized with the full-length BGCs predicted to produce tridecaptin A_1 (*Paenibacillus* sp. S25, *P. peoriae* ZBSF16, *P. polymyxa* R 4.24), tridecaptin A_5 (*P. polymyxa* J, *P. polymyxa* EB4 G3, P. *polymyxa* MEZ6, *P. polymyxa* R 6.14, *P. polymyxa* HY96-2, *P. polymyxa* SQR-21) or tridecaptin M_{11} (*Paenibacillus* sp. M-152, *Paenibacillus* sp. lzh-N1, *P. polymyxa* CJX518). The authors who previously isolated these compounds ascribed the production of tridecaptins B_1/M_1 to full-length tridecaptin BGCs within the genomes of their producing strains, *P. polymyxa* NRRL B-30507 and *Paenibacillus* sp. M-152, respectively (Cochrane *et al.*, 2015, Jangra *et al.*, 2019b). However, both groups noticed discrepancies between the adenylation domain predictions and the observed scaffolds for these full-length clusters, which instead corresponded to tridecaptin A_1 (Table S3) and the putative tridecaptin M_{11} (Table S4), respectively. Specifically, in *P. polymyxa* NRRL B-30507, the greatest discrepancies were observed for adenylation domains 1 and 9, which the authors predicted to select valine and phenylalanine, respectively, rather than the glycine and isoleucine, which are found in the peptide scaffolds of tridecaptin B_1 and M_1 (Tables S3, S4). Similarly, glycine (position 1) and isoleucine (position 9)-recognizing adenylation domains are found in the TriD peptide synthase of the truncated tridecaptin cluster in the genome of *Paenibacillus* sp. M-152 (Table S2).

We hypothesized that the truncated BGC in *Paenibacillus* sp. M-152 might be involved in the chimeric biosynthesis of tridecaptin M_1 by biosynthesizing the first ten amino acids, upon which the last three are incorporated by the TriE synthase encoded by the full-length tridecaptin M_{11} BGC (Figure 1C; Table S4). This seems particularly likely considering that both tridecaptins M_1 and M_{11} were isolated from this strain. At this point, we strongly suspected that the tridecaptin B_1 biosynthesis in *P. polymyxa* NRRL B-30507 was also chimeric, especially considering the wide-spread co-occurrence of full-length and

truncated BGCs among *Paenibacillus* genomes that would encode for tridecaptin A_1 and B_1, respectively. Unfortunately, the genome of this strain is too heavily fragmented to confirm that it also harbors a truncated tridecaptin BGC that might be responsible for the chimeric biosynthesis of tridecaptin B_1.

Screening of tridecaptin B producers using large-scale metabolomics

To strengthen our hypothesis that chimeric biosynthesis also occurs for tridecaptin A/B_1-producing strains, we set out to identify and sequence a *Paenibacillus* strain that produced both tridecaptin A and B. For this purpose, we screened a collection of 227 plant-associated *Paenibacillus* isolates (see Materials and Methods) for producers of tridecaptin B. To accomplish this, the isolates were grown in 96 well plates, and extracts of the cultures were analyzed using liquid chromatography coupled with mass spectrometry (LC-MS). A mass features having an *m/z* of 1458.8315, 729.9194 or 486.9487 corresponding to an $[M+H]^+$, $[M+2H]^{2+}$ or $[M+3H]^{3+}$ ions of tridecaptin B_1, respectively; were searched in the obtained LC-MS dataset. Eventually, a triply charged ion peak having an *m/z* of 486.9487, eluting at a retention time of 5.09 min, was detected and annotated as tridecaptin B_1 based on its MS/MS spectrum assignment (Figure 2). We found that tridecaptin B_1 was produced by seven *Paenibacillus* isolates, namely *Paenibacillus* sp. JJ-195, *Paenibacillus* sp. JJ-1650, *Paenibacillus* sp. JJ-333, *Paenibacillus* sp. JJ-1683, *Paenibacillus* sp. JJ-1604, *Paenibacillus* sp. JJ-1667, and *Paenibacillus* sp. JJ-1669. Importantly, all tridecaptin B_1 producers also produced the previously discovered tridecaptin A_5, which was characterized by the MS/MS analysis of its triply charged ion peak having an *m/z* of 539.9715 (Figure 2). Interestingly, a total of 28 tridecaptin A_5 producers were detected in the collection; therefore, 21 strains exclusively produced tridecaptin A_5.

Of the isolates that produced both tridecaptin B_1 and tridecaptin A_5, *Paenibacillus* sp. JJ-1683 was selected for further analysis to assess whether the production of tridecaptin B_1 could potentially be encoded by two collaborating BGCs involved in chimeric biosynthesis. The genome of *Paenibacillus* sp. JJ-1683 was sequenced using the PacBio platform. Assembly of the PacBio reads with Falcon (version 1.8.1) (Lohans *et al.*, 2014) resulted in a single contig of 5.8 Mb. The *Paenibacillus* sp. JJ-1683 genome was analyzed using antiSMASH 6.0.1 (Blin *et al.*, 2021) to obtain an overview of the predicted BGCs. Indeed, two tridecaptin BGCs were found in the genome of *Paenibacillus* sp. JJ-1683 (Table S2) that produced tridecaptin B_1, together with tridecaptin A_5. The first, full-length, BGC encoded both ten-modular TriD and three-modular TriE synthases, while the second BGC was truncated and encoded exclusively a ten-modular TriD ortholog. Conversely, *Paenibacillus* sp. JJ-21 harbors solely the full-length tridecaptin BGC predicted to encode the production of tridecaptin A_5 (Table S2). Notably, tridecaptin A_5 was detected in the extract of this strain, whereas tridecaptin B_1 was not found (Figure 2A).

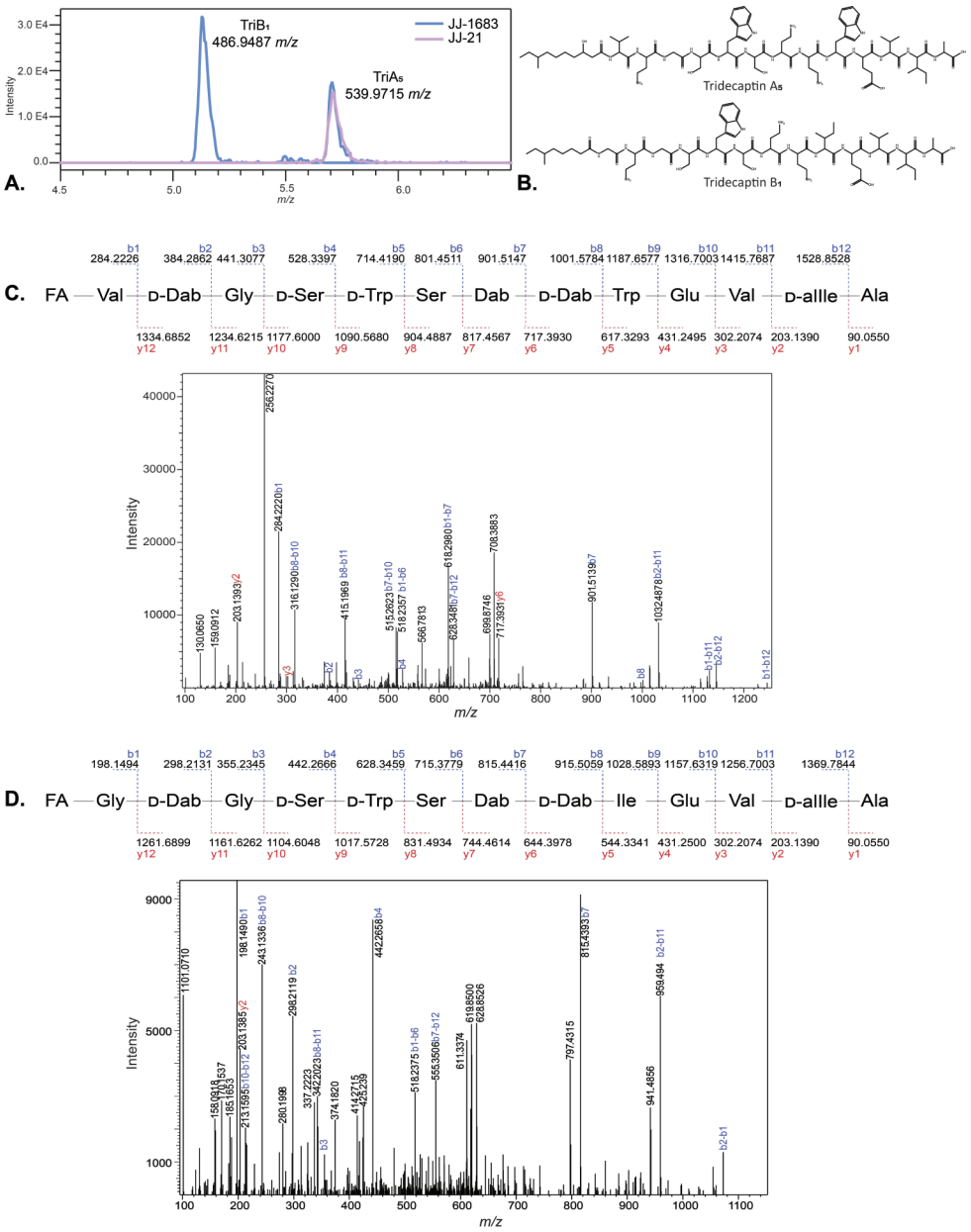


Figure 2. Extracted ion chromatograms (EICs) (A) and chemical structures (B) of tridecaptin A_5 (539.9715 *m/z*) and B_1 (486.9487 *m/z*). Note that, *Paenibacillus* sp. JJ-1683 produced both tridecaptin B_1 and tridecaptin A_5, while *Paenibacillus* sp. JJ-21 produced solely tridecaptin A_5. The assignment of the sequence of amino acid residues of tridecaptin A_5 **(C)** and tridecaptin B_1 **(D)** was conducted based on the mass differences between the consecutive *y* and *b* ions.

Production of tridecaptin B_1 exclusively by the strain that harbors two tridecaptin BGCs strongly suggests that both the truncated and full-length clusters are functional and required for the biosynthesis of this hybrid tridecaptin (Figure 3). This is further supported by the widespread co-occurrence of full-length and truncated tridecaptin BGCs: we found evidence of chimeric tridecaptin biosynthesis in 18 *Paenibacillus* strains in total (Table S2). Out of all 61 genomes in which tridecaptin BGCs were detected, this represents around 25% of all putative tridecaptin producers. A partially related mechanism has been shown previously for the biosynthesis of fungal NRPs, in which a single NRPS is involved in two separate biosynthesis pathways (Zhao *et al.*, 2016). Some families of *Pseudomonas* lipopeptides display a "split" BGC configuration, with the initiator NRPS gene at a genomic location distant from the operon with the other NRPS genes (Cesa-Luna *et al.*, 2023).

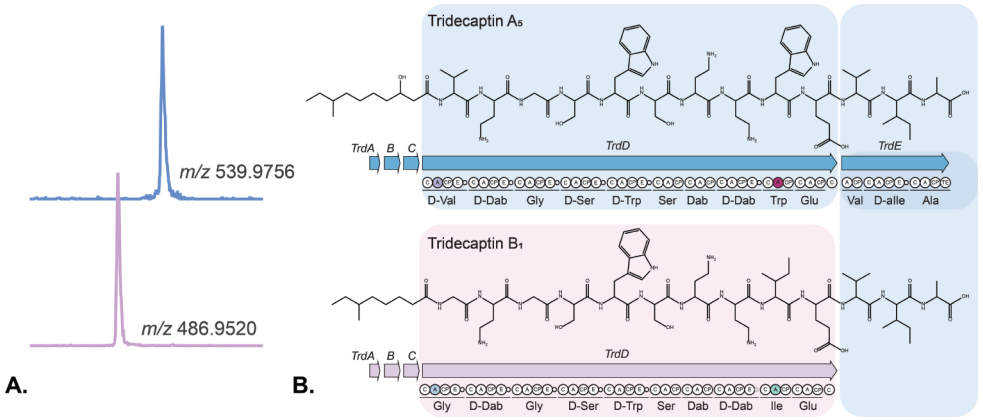


Figure 3. EICs of tridecaptin A_5 (blue) and tridecaptin B_1 (pink) co-produced by *Paenibacillus* sp. JJ-1683 (A) and proposed biosynthetic pathway of tridecaptin B_1 (B).

Bioactivity profile of synthetic tridecaptins A_5 and B_1 analogs

Chimeric biosynthesis is a meaningful mechanism of bacterial and fungal secondary metabolism, which has the potential to further increase the structural diversity of encoded molecules and subsequent biological activity. To assess whether tridecaptin diversity caused by chimeric biosynthesis also translates to a strain's potential for functional diversity, we were also motivated to test tridecaptins A_5 and B_1 for bioactivity against a panel of common human pathogens.

Due to the limited production of tridecaptins A_5 and B_1 by *Paenibacillus* sp. JJ-1683, we developed synthetic analogs for both lipopeptides. Previous studies have shown that the natural N-terminal lipid tail of tridecaptin A_1 can be replaced with octanoic acid without impacting the antimicrobial properties of the peptide (Lohans *et al.*, 2014). Thus, to simplify the synthetic procedure, the chiral lipid tails found in the natural lipopeptides were replaced with commercially available octanoic acid, yielding the synthetic tridecaptin analogs Oct-TriA_5 and Oct-TriB_1 (Table S5, Figures S1, S2). As we previously reported, the synthetic

analogs of tridecaptin A_5, Oct-TriA_5, have unusual broad-spectrum activity against Gram-positive and Gram-negative ESKAPE pathogens (**Chapter 7**). On the other hand, Oct-TriB_1 displays moderate activity against Gram-negative strains and is not active against Gram-positive bacteria (Table 2). The selective activity observed for Oct-TriB_1 is consistent with the antimicrobial spectrum known for several natural tridecaptins (i.e., tridecaptins A, C, and M) (Bann *et al.*, 2021, Jangra *et al.*, 2019b).

Table 2. *In vitro* minimum inhibitory concentrations (MICs) of Oct-TriB_1 and Oct-TriA_5.

Strain		Species	MIC[a]		
			Oct-TriB_1	Oct-TriA_5	Colistin
BV18	UAMS-1625	*S. aureus*	> 64	8	> 64
BV249	ATCC 29213	*S. aureus*	> 64	8	> 64
BV1402	1126387	*S. aureus*	> 64	8	> 64
BV29	ATCC 29212	*E. faecalis*	> 64	8	> 64
BV99	ATCC 25922	*E. faecalis*	> 64	8	> 64
BV144	IHMA 890472	*E. faecium*	> 64	8	> 64
BV159	ATCC 51559	*E. faecium*	> 64	8	> 64
BV94	IHMA 777621	*A. baumannii*	> 64	4	64
BV160	ATCC 17978	*A. baumannii*	> 64	4	1
BV374	HUMC1	*A. baumannii*	> 64	8	1
BV23	ATCC 25922	*E. coli*	16	2	1
BV1475	1220120	*E. coli*	16	4	0.5
BV1469	1214245	*E. coli*	16	4	32
BV306	ATCC 13883	*K. pneumoniae*	8	4	1
BV1445	1227947	*K. pneumoniae*	8	4	0.5
BV1447	1228586	*K. pneumoniae*	4	4	64
BV34	PAO1	*P. aeruginosa*	>64	16	2
BV1551	1226072	*P. aeruginosa*	16	8	0.5
BV1544	1218019	*P. aeruginosa*	>64	4	> 64
Hemolysis EC50 [μg/mL]			>128	6.9	n/a
IC50 on HepG2 w/o FCS [μg/mL]			>128	11.4	n/a

[a] MIC values reported in units of μg/mL

In conclusion, this study is the first to reveal the biosynthetic potential of the *Paenibacillus* genus to produce hybrid tridecaptin antibiotics. Through BiG-SCAPE analysis we gained insight into the diversity of tridecaptin BGCs from *Paenibacillus* spp. and discovered new

tridecaptin BGCs including molecules with unusual biosynthetic architectures. Using LC-MS analysis and genome mining studies, we revised the biosynthetic pathway of tridecaptin B_1 and demonstrated that the biosynthesis of tridecaptin B_1 is hybrid and involves two separate BGCs. Tridecaptin B_1 assembly line provides an unusual example of chimeric biosynthesis of NRPS lipopeptide antibiotics.

Materials and Methods

Global genome mining of NRPSs

Genomes of all *Paenibacillus* spp. available from RefSeq (release 213) (O'Leary *et al.*, 2016) were downloaded from the NCBI FTP site. All genomes with more than 400 contigs were considered low-quality assemblies and were removed from the collection, leaving a set of 785 *Paenibacillus* genomes. All genomes were analyzed using AntiSMASH (version 6.0.1) (Blin *et al.*, 2021) to obtain biosynthetic gene cluster predictions. These predictions were then used as input for BiG-SCAPE (version 1.1.4) (Navarro-Muñoz *et al.*, 2020) analysis with a distance matrix cutoff set to 0.25. The resulting network of gene cluster families was visualized using Cytoscape (3.9.1) (Shannon *et al.*, 2003).

To detect tridecaptin BGCs in our dataset, we used the adenylation domain predictor PARAS (v0.0.1, available at https://github.com/BTheDragonMaster/parasect) to extract adenylation domains and predict their substrate specificity. PARAS is an adenylation domain predictor that uses an HMM to extract the A-domain active site and was trained on over 2800 adenylation domains from fungi and bacteria. Importantly, PARAS also reports the confidence score of a prediction. The *triD* and *triE* NRPS genes are always positioned on the same DNA strand, with typically only 165 bp between them. Keeping this in mind, we identified putative tridecaptin NRPS enzymes by sorting A-domain containing proteins into 'protein blocks': proteins that are encoded side by side on the genome on the same strand, with a maximum of 1000 bp between them. Then, we counted the adenylation domains for each protein block, and only considered those protein blocks that had ten or more adenylation domains. We chose a cutoff of ten instead of thirteen to also capture potentially truncated clusters, some of which we had detected with our BiG-SCAPE analysis. Then, we predicted the adenylation domains of all remaining protein blocks with PARAS (v0.0.1), and manually selected putative tridecaptin-producing BGCs based on predicted substrate selectivity and epimerization domain presence in modules 1-5, 8, and 12 as predicted by antiSMASH (Blin *et al.*, 2023).

Finally, for the three genomes, that contained a truncated ten-modular tridecaptin BGC but not a full-length tridecaptin BGC, we manually inspected the antiSMASH results to search for NRPS BGCs encoding fewer than ten NRPS modules to see if these might contribute to tridecaptin biosynthesis as well. We found three three-modular BGCs with an epimerization domain in their second module (corresponding to module 12 in full-length tridecaptin BGCs) and predicted their adenylation domain specificity as previously described.

Growth of *Paenibacillus* spp. and specialized metabolite extraction

Our method for cultivation and extraction of specialized metabolites was adapted from previously described procedures (Nguyen *et al.*, 2016)2016. Briefly, 227 *Paenibacillus* spp. from the Auburn University Plant-Associated Microbial strain collection were inoculated into 600 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) 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).

LC-MS/MS analysis

For LC-MS analyses 1 µL of crude extract was injected into a Shimadzu Nexera X2 UHPLC system coupled to a Shimadzu 9030 QTOF mass spectrometer, and data acquisition was performed as previously described (van Bergeijk *et al.*, 2022). Briefly, 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.

Genome sequencing, assembly and annotation of *Paenibacillus* sp. JJ-1683

Paenibacillus sp. JJ-1683 and *Paenibacillus* sp. JJ-21 were grown in TSB (Bacto Soybean-Casein Digest Medium, 30 g/L) at 30 °C and 220 rpm for 24 h. DNA was extracted from *Paenibacillus* sp. JJ-1683 and *Paenibacillus* sp. JJ-21 as described (Kieser, 2000). DNA quality was verified by agarose gel electrophoresis. PacBio sequencing and assembly was performed by Novogene (UK). Generally, the library was prepared using a 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.*, 2014). BGCs in this genome were annotated using AntiSMASH (version 6.0.1) (Blin *et al.*, 2021).

Solid phase peptide synthesis

Peptides were prepared on 0.25 mmol scale on either preloaded Wang resin or CTC resin. The synthesis of Oct-TriB was performed on preloaded Fmoc-Ala-Wang resin (0.29 mmol/g loading). All couplings with the exception of Fmoc-d-*allo*-Ile-OH and Fmoc-d-Ile-OH were performed using 4 eq. of amino acid or fatty acid, benzotriazol-1-yloxytris(dimethylamino) phosphonium hexafluorophosphate (BOP) (4 eq.) and N,N-diisopropylethylamine (DiPEA) (8 eq.) in 10 mL of DMF for 1 h at RT, under nitrogen flow. Fmoc-d-allo-Ile and Fmoc-d-Ile-OH were coupled by treating the resin with 2 eq. of the amino acid, 2 eq. of BOP and 4 eq. of DiPEA in 10 mL of DMF overnight at RT. Fmoc group removal was performed by treating the resin with 10 mL of piperidine : DMF (1 : 4, v/v) for 5 min and then again for 15 min.

Final sidechain deprotection and cleavage from the resin was carried out by treating the resin with 10 mL of TFA : TIPS : H_2O (95 : 2.5 : 2.5, v/v) for 90 min. The reaction mixture was filtered through cotton, the filtrate was precipitated in MTBE : petroleum ether (1 : 1, v/v) and centrifuged (4500 rpm, 5 min). The pellet was then resuspended in MTBE : petroleum ether (1 : 1, v/v) and centrifuged again (4500 rpm, 5 min). Finally the pellet containing the crude lipopeptide was dissolved in tBuOH:H_2O (1 : 1, v/v) and lyophilized overnight. The crude mixtures were subsequently purified by RP-HPLC. Fractions were assessed by HPLC and LC-MS and product containing fractions were pooled, frozen and lyophilized to yield the pure lipopeptides in >95% purity (determined by HPLC).

Peptide synthesis

Fmoc-L-Dab(Boc)-OH, Fmoc-D-Dab(Boc)-OH, Fmoc-D-Ile-OH, Fmoc-allo-Ile-OH and 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate (HATU) were purchased from Combi-Blocks. All other Fmoc-amino acids, the Fmoc-Ala-Wang resin and Fmoc-Glu(OtBu)-Wang resin were purchased from P3 BioSystems. 2-Chlorotrityl chloride resin (CTC) was purchased from Iris Biotech. Octanoic acid was purchased from Alfa Aesar. ((1H-Benzo[d][1,2,3]triazol-1-yl)oxy)tris(dimethylamino) phosphonium hexafluorophosphate (BOP), N,N-Diisopropylcarbodiimide (DIC) and triisopropylsilane (TIPS) were purchased from Manchester Organics. Diisopropylethylamine (DIPEA), piperidine, trifluoroacetic acid (TFA) and dimethyl sulfoxide (DMSO) were purchased from Carl Roth. Dichloromethane (CH_2Cl_2) and petroleum ether were purchased from VWR Chemicals. Acetonitrile (MeCN), dimethylformamide (DMF) and methyl tertiary-butyl ether (MTBE) were purchased from Biosolve.

Resin swelling

The resin was swollen in 10 mL of DMF for 300 s prior to the first coupling.

Automated coupling protocol

Step	Function	Duration/Temperature
1	Initial Deprotection N-terminus	15 s at 60°C then 30 s at 70°C
2	Deprotection N-terminus	15 s at 60°C then 180 s at 70°C
3	Wash (DMF)	RT
4	Wash (DMF)	RT
5	Wash (DMF)	RT
6	Coupling amino acid	15 s at 60°C then 300 s at 70°C

After coupling of the final residue on the synthesizer, the resin was washed with DCM, filtered and treated with 3 mL of TFA : TIPS : H_2O (95 : 2.5 : 2.5, v/v) for 90 min. The reaction mixture was filtered through cotton, the filtrate was precipitated in MTBE : petroleum ether (1 : 1, v/v) and centrifuged (4500 rpm, 5 min). The pellet was then resuspended in MTBE : petroleum

ether (1 : 1, v/v) and centrifuged again (4500 rpm, 5 min). Finally the pellet containing the crude lipopeptide was dissolved in tBuOH : H_2O (1 : 1, v/v) and lyophilized overnight. The crude mixtures were subsequently purified by RP-HPLC. Fractions were assessed by HPLC and LC-MS and product containing fractions were pooled, frozen and lyophilized to yield the pure lipopeptides in >95% purity (determined by HPLC).

Prep RP-HPLC purification

Peptides were purified using a BESTA-Technik system with a Dr. Maisch Reprosil Gold 120 C_{18} column (25 × 250 mm, 10 µm) and equipped with a ECOM Flash UV detector monitoring at 214 nm and 254 nm. The following solvent system, at a flow rate of 12 mL/min, was used: solvent A, 0.1 % TFA in water/acetonitrile 95/5; solvent B, 0.1 % TFA in water/acetonitrile 5/95. Gradient elution was as follows: 100:0 (A/B) for 5 min, 100:0 to 50:50 (A/B) over 50 min, 50:50 to 0:100 (A/B) for 3min, then reversion back to 100:0 (A/B) over 1 min, 100:0 (A/B) for 5 min.

Analytical RP-HPLC

Shimadzu Prominence-i LC-2030 system with a Dr. Maisch ReproSil Gold 120 C_{18} column (4.6 × 250 mm, 5 µm) at 30 °C and equipped with a UV detector monitoring 214 nm and 254 nm. The following solvent system, at a flow rate of 1 mL/min, was used: solvent A, 0.1 % TFA in water/acetonitrile 95/5; solvent B, 0.1 % TFA in water/acetonitrile 5/95. Method A: Gradient elution was as follows: 100:0 (A/B) for 2 min, 100:0 to 0:100 (A/B) over 23 min, 0:100 (A/B) for 1 min, 0:100 (A/B) then reversion back to 100:0 (A/B) over 1 min, 100:0 (A/B) for 3 min. Method B: Gradient elution was as follows: 100:0 (A/B) for 2 min, 100:0 to 50:50 (A/B) over 45 min, 50:50 (A/B) to 0:100 (A/B) over 1 min, 0:100 (A/B) for 6 min then reversion back to 100:0 (A/B) over 1min, 100:0 (A/B) for 5 min.

HRMS analysis

HRMS analyses were performed on a Thermo Scientific Dionex UltiMate 3000 HPLC system with a Phenomenex Kinetex C_{18} (2.1 x 150 mm, 2.6 µm) column at 35 °C and equipped with a diode array detector. The following solvent system, at a flow rate of 0.3 mL/min, was used: solvent A, 0.1 % formic acid in water; solvent B, 0.1% formic acid in acetonitrile. Gradient elution was as follows: 95:5 (A/B) for 1 min, 95:5 to 5:95 (A/B) over 9 min, 5:95 to 2:98 (A/B) over 1 min, 2:98 (A/B) for 1 min, then reversion back to 95:5 (A/B) over 2 min, 95:5 (A/B) for 1 min. This system was connected to a Bruker micrOTOF-Q II mass spectrometer (ESI ionization) calibrated internally with sodium formate.

Antimicrobial testing

All minimum inhibitory concentrations were determined according to Clinical and Standards Laboratory Institute (CLSI) guidelines. In brief, agar plates were inoculated from glycerol stocks and incubated overnight at 35 °C. Peptides DMSO stocks were dispensed into 96-well round bottom plates using a TECAN D300e dispenser. Colistin sulfate (Sigma-Aldrich) was dissolved in water and serially diluted in cation-adjusted Mueller Hinton broth (10 µL per well). The final concentrations of the test compounds were 64–0.06 µg/mL. Inocula were prepared by direct colony suspension in NaCl at 0.5 MacFarland, which was further diluted 200-fold in cation-adjusted Mueller Hinton broth for a target inoculum of 2 x 10^5 colony forming units/mL. 100 µL of the prepared inoculum was added to each well of the prepared plates containing the test compounds. Plates were sealed with parafilm and incubated for 20 h at 35 °C. Images of all plates were recorded and MIC was assessed visually.

Acknowledgments

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. Active site conservation among *Paenibacillus* tridecaptin adenylation domains. Active sites were extracted using PARAS. Residues that deviate from the most commonly found active site for that substrate are indicated in bold. Confidence shows the average confidence of PARAS predictions for that specific active site. The number of domains with that active site among tridecaptin BGCs is also given. Rows shown in color correspond to active sites found in the BGCs of *Paenibacillus sp.* M-152 and *Paenibacillus polymyxa* NRRL B-30507.

(Predicted) substrate	Active site signature	Confidence	Number of domains
Valine	DAFWLGGTFK	1.00	95
	DAFWIGGTFK	0.97	4
	DAFWLGGAFK	0.77	1
	DAFFIGATFK	0.36	1
2,4-diaminobutyric acid	DVGEISSIDK	1.00	200
	DVGEISAIDK	0.96	10
	DVGETGTIDK	0.43	1
Alanine	DVFWLGGTFK	0.77	48
	DVTNFAIIYK	0.81	2
Tryptophan	DAWAFAGVAK	0.95	83
	DAWAVAGVAK	0.80	22
	DAWAFAGIAK	0.80	7
Serine	DVWHFSLVDK	1.00	147
	DVWHYSLVDK	0.99	2
	DVWHFSMVEK	0.55	1
Aspartic acid	DMKDVGVVDK	0.34	4
Arginine	DPEDIGHIDK	0.33	3
Glycine	DILQMGMVWK	1.00	81
Glutamic acid	DAKDLGVVDK	0.95	67
Isoleucine	DAFFLGITFK	0.95	47
	DAFFLGVTFK	0.91	25
	DAFFLGATFK	0.42	4
	DGFFLGVVYK	0.81	2
	DGFFLGVVFK	0.85	1

Phenylalanine	DAWTFAGVAK	0.73	8
	DAWTVAGVAK	0.78	2
	DAWTIAAICK	0.92	2
	DAWTIAGVAK	0.75	2
	DAWTVAAVAK	0.87	1
	DAWTFAAVAK	0.88	1
Glutamine	DAQDLGVVDK	0.59	1
Leucine	DAWLLGAIVK	0.91	1

Table S2. Predicted *Paenibacillus* tridecaptin biosynthetic gene clusters.

In fifteen genomes, both a full-length (13-modular) and a truncated (10-modular) tridecaptin BGC can be found. In three genomes, two truncated (10-modular and 3-modular) tridecaptin BGCs can be found. Novel tridecaptins are highlighted in green. The tridecaptin BGC from the genomes of *Paenibacillus* sp. JJ-1683 and *Paenibacillus* sp. JJ-21 are highlighted in bold.

Strain	Accession number	1[D]	2[D]	3[D]	4[D]	5[D]	6	7	8[D]	9	10	11	12[D]	13	Compound
P. elgii MER 157	NZ_JAMAVM010000004.1. region001	Val	Dab	Ala	Dab	Trp	Ser	Asp	Dab	Trp	Asp	Val	Val	Arg	Tridecaptin G_4
P. elgii SWL-W8	NZ_JAAIVH010000001.1. region006	Val	Dab	Ala	Dab	Trp	Ser	Asp	Dab	Trp	Asp	Val	Val	Arg	Tridecaptin G_4
P. jamilae KACC 10925	NZ_QVPU01000010.1. region001	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. kribbensis 6hRe76	NZ_MCBP01000006.1. region002	Val	Dab	Ala	Ser	Trp	Ser	Dab	Dab	Phe	Glu	Val	Ile?	Ala	New Tridecaptin 1
P. kribbensis AM49	NZ_CP020028.1.region007	Val	Dab	Ala	Ser	Trp	Gly	Ser	Dab	Phe	Glu	Val	Ile?	Ser	New Tridecaptin 2
P. kribbensis PS04	NZ_CP041731.1.region012	Val	Dab	Ala	Ser	Trp	Gly	Ser	Dab	Phe	Glu	Val	Ile?	Ser	New Tridecaptin 2
P. mucilaginosus K02	NC_017672.3.region011	Phe	Dab	Ile	Ser	Trp	Ser	Ser	Dab	Trp	Ser	Val	Ile	Dab	Tridecaptin D
P. mucilaginosus KNP414	NC_015690.1.region011	Phe	Dab	Ile	Ser	Trp	Ser	Ser	Dab	Trp	Ser	Val	Ile	Dab	Tridecaptin D
P. peoriae HS311	NZ_CP011512.1.region010	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. peoriae KCTC 3763	NZ_AGFX01000012.1. region001	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Val	Glu	Val	Ile	Ser	New Tridecaptin 3
P. peoriae ZBSF16	NZ_CP092831.1.region009	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Phe	Glu	Val	Ile	Ala	Tridecaptin A_1/A_3
P. peoriae ZBSF16	NZ_CP092831.1.region010	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	---	---	---	Tridecaptin B_1
P. peoriae ZF390	NZ_CP061172.1.region008	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Phe	Glu	Val	Ile	Ala	Tridecaptin A_1/A_3
P. polymyxa 2020	NZ_CP049598.1.region019	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Ile	Ile	Ala	New Tridecaptin 4
P. polymyxa A18	NZ_JWJJ01000007.1.region001	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. polymyxa ATCC 15970	NZ_CP011420.1.region010	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	---	Tridecaptin A_5

P. polymyxa C12	NZ_CP023711.1.region001	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. polymyxa CICC 10580	NZ_JNCB01000011.1. region006	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. polymyxa CJX518	NZ_CP029370.1.region007	Val	Dab	Ser	Gly	Trp	Ser	Dab	Dab	Phe	Glu	Ile	Ile	Ser	New Tridecaptin 5
P. polymyxa CJX518	NZ_CP029370.1.region008	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	---	---	---	Tridecaptin M_1
P. polymyxa CR1	NC_023037.2.region009	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. polymyxa DCB2-2	NZ_RXYF01000012.1. region001	---	---	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. polymyxa DSM 292	NZ_OXKC02000014.1. region001	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. polymyxa DSM 36	NZ_CP049783.1.region018	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Ile	Ile	Ala	New Tridecaptin 4
P. polymyxa E681	NZ_CP048793.1.region008	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Phe	Glu	Val	Ile	Ala	Tridecaptin A_1/A_3
P. polymyxa EB4 G3	NZ_CP084033.1.region008	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. polymyxa EB4 G3	NZ_CP084033.1.region009	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	---	---	---	Tridecaptin B_1
P. polymyxa G25-124	NZ_LDIE01000050.1. region001	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Ile	Ile	Ala	New Tridecaptin 4
P. polymyxa HK4	NZ_JAACZY010000021.1. region001	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Ile	Ile	Ala	New Tridecaptin 4
P. polymyxa HY96-2	NZ_CP025957.1.region010	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	---	Val	Ile	Ala	Tridecaptin A_5
P. polymyxa HY96-2	NZ_CP025957.1.region011	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	---	---	---	Tridecaptin B_1
P. polymyxa J	NZ_CP015423.1.region015	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. polymyxa J	NZ_CP015423.1.region017	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	---	---	---	Tridecaptin B_1

P. polymyxa KCCM 40454	NZ_POVT01000002.1. region001	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Ile	Ile	Ala	New Tridecaptin 4
P. polymyxa K 1.14	NZ_CP097778.1.region009	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Phe	Glu	Val	Ile	Ala	Tridecaptin A_1/A_3
P. polymyxa K 1.14	NZ_CP097778.1.region011	Ala	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	---	---	---	New Tridecaptin 6
P. polymyxa LY214	NZ_NOLA01000018.1. region001	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. polymyxa M1	NC_017542.1.region008	---	---	---	---	---	---	---	---	---	---	Ile	Ile	Ser	Tridecaptin M_1
P. polymyxa M1	NC_017542.1.region009	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	---	---	---	Tridecaptin M_1
P. polymyxa MEZ6	NZ_CP086373.1.region009	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. polymyxa MEZ6	NZ_CP086373.1.region010	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	---	---	---	Tridecaptin B_1
P. polymyxa NCTC10343	NZ_UGSC01000001.1. region012	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Ile	Ile	Ala	New Tridecaptin 4
P. polymyxa NRRL B-30509	NZ_JTHO01000008.1. region001	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. polymyxa R 3.13	NZ_CP097771.1.region013	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. polymyxa R 4.24	NZ_CP097769.1.region005	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Phe	Glu	Val	Ile	Ala	Tridecaptin A_1/A_3
P. polymyxa R 4.24	NZ_CP097769.1.region006	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	---	---	---	Tridecaptin B_1
P. polymyxa R 4.5	NZ_CP097770.1.region005	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	---	---	---	Tridecaptin B_1
P. polymyxa R 4.5	NZ_CP097770.1.region007	---	---	---	---	---	---	---	---	---	---	Val	Ile	Ala	Tridecaptin B_1
P. polymyxa R 5.31	NZ_CP097767.1.region001	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. polymyxa R 6.14	NZ_CP097766.1.region009	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. polymyxa R 6.14	NZ_CP097766.1.region010	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	---	---	---	Tridecaptin B_1
P. polymyxa SC2	NC_014622.2.region008	---	---	---	---	---	---	---	---	---	---	Ile	Ile	Ser	Tridecaptin M_1
P. polymyxa SC2	NC_014622.2.region009	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	---	---	---	Tridecaptin M_1
P. polymyxa SQR-21	NZ_CP006872.1.region009	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. polymyxa SQR-21	NZ_CP006872.1.region010	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	---	---	---	Tridecaptin B_1
P. polymyxa YC0573	NZ_CP017968.3.region011	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5

P. polymyxa ZF129	NZ_CP040829.1.region014	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
P. polymyxa ZJ-9	NZ_CP092779.1.region010	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Ile	Ile	Ala	New Tridecaptin 4
Paenibacillus sp. EKM211P	NZ_JAAMNS010000007.1. region001	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Ile	Ile	Ala	New Tridecaptin 4
Paenibacillus sp. F4	NZ_POVS01000007.1. region001	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
Paenibacillus sp. ICGEB2008	NZ_AMQU01000029.1. region001	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
***Paenibacillus* sp. JJ-1683**	**Paenibacillus sp._MBT_ JJ-1683.region008**	**Val**	**Dab**	**Gly**	**Ser**	**Trp**	**Ser**	**Dab**	**Dab**	**Trp**	**Glu**	**Val**	**Ile**	**Ala**	**Tridecaptin A_5**
***Paenibacillus* sp. JJ-1683**	**Paenibacillus sp._MBT_ JJ-1683.region009**	**Gly**	**Dab**	**Gly**	**Ser**	**Trp**	**Ser**	**Dab**	**Dab**	**Ile**	**Glu**	---	---	---	**Tridecaptin B_1**
***Paenibacillus* sp. JJ-21**	**Paenibacillus sp._MBT_ JJ-21.region011**	**Val**	**Dab**	**Gly**	**Ser**	**Trp**	**Ser**	**Dab**	**Dab**	**Trp**	**Glu**	**Val**	**Ile**	**Ala**	**Tridecaptin A_5**
Paenibacillus sp. M-152	NZ_CP034141.1.region008	Val	Dab	Ser	Gly	Trp	Ser	Dab	Dab	Phe	Glu	Ile	Ile	Ser	New Tridecaptin 5
Paenibacillus sp. M-152	NZ_CP034141.1.region009	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	---	---	---	Tridecaptin M_1
Paenibacillus sp. PvR133	NZ_JAFIBS010000001.1. region012	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
Paenibacillus sp. S02	NZ_CP073682.1.region011	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
Paenibacillus sp. S25	NZ_CP073683.1.region012	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Phe	Glu	Val	Ile	Ala	Tridecaptin A_1/A_3
Paenibacillus sp. S25	NZ_CP073683.1.region014	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	---	---	---	Tridecaptin B_1
Paenibacillus sp. UKAQ 18	NZ_JAKFSF010000002.1. region001	---	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Phe	Glu	Val	---	---	Tridecaptin A/B/C
Paenibacillus sp. UNCCL52	NZ_JMLR01000007.1. region003	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
Paenibacillus sp. cl130 CL130	NZ_FOYG01000003.1. region001	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Trp	Glu	Val	Ile	Ala	Tridecaptin A_5
Paenibacillus sp. lzh-N1	NZ_CP025696.1.region011	Gly	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Ile	Glu	---	---	---	Tridecaptin M_1

Paenibacillus sp. lzh-N1	NZ_CP025696.1.region012	Val	Dab	Ser	Gly	Trp	Ser	Dab	Dab	Phe	Glu	Ile	Ile	Ser	New Tridecaptin 5
P. terrae sp. GHS.8NWYW.5	NZ_VIYI01000003.1.region001	Val	Dab	Val	Ser	Trp	Ser	Gln	Dab	Phe	Glu	Val	Val	Arg	New Tridecaptin 7
P. terrae sp. HPL-003	NC_016641.1.region014	---	Dab	Gly	Trp	Leu	Ser	Dab	Dab	Phe	Glu	Val	Ile	Ala	New Tridecaptin 8
P. terrae sp. HPL-003	NC_016641.1.region010	Val	Dab	Gly	Ser	Trp	Ser	Val	Gln	Ala	---	---	---	---	New Tridecaptin 9
P. terrae sp. NRRL B-30644	NZ_JTHP01000234.1. region001	Val	Dab	Gly	Ser	Trp	Ser	Dab	Dab	Phe	Glu	Val	Ile	Ala	Tridecaptin A_1/A_3

[D] Modules incorporate a D-amino acid. ND, not determined.

Table S3. Bioinformatic analysis of tridecaptin BGC of *Paenibacillus polymyxa* NRRL B-30507.

Module number	Tridecaptin B_1	Active site signature	Tridecaptin A_5	Predicted aa (PARAS)	Predicted aa (Cochrane *et al.*, 2015)
1	Gly	DAFWLGGTFK	Val	Val (1.00)	Val
2	Dab	DVGEISSIDK	Dab	Dab (1.00)	Orn
3	Gly	DILQMGMVWK	Gly	Gly (1.00)	Gly
4	Ser	DVWHFSLVDK	Ser	Ser (1.00)	Ser
5	Trp	DAWAFAGVAK	Trp	Trp (0.94)	Phe
6	Ser	DVWHFSLVDK	Ser	Ser (1.00)	Ser
7	Dab	DVGEISSIDK	Dab	Dab (1.00)	Orn
8	Dab	DVGEISSIDK	Dab	Dab (1.00)	Orn
9	Ile	DAWAFAGVAK	Trp	Trp (0.94)	Phe
10	Glu	DAKDLGVVDK	Glu	Glu (0.96)	Glu
11	Val	DAFWLGGTFK	Val	Val (1.00)	Val
12	Ile	DAFFLGITFK	Ile	Ile (0.93)	Ile
13	Ala	DVFWLGGTFK	Ala	Ala (0.73)	Val

The tridecaptin BGC in *P. polymyxa* NRRL B-30507 matches the structure of tridecaptin A_5, not that of tridecaptin B_1. For PARAS predictions, prediction confidence (scale from 0.0-1.0) is shown between brackets.

Table S4. Bioinformatic analysis of tridecaptin BGC of *Paenibacillus* sp. M-152.

The predicted scaffold for truncated tridecaptin M_1 matches the isolated structure from *Paenibacillus sp*. M-152. The postulated structure of tridecaptin M_{11} is also shown, which matches with the full-length BGC. For PARAS predictions, prediction confidence (scale from 0.0-1.0) is shown between brackets.

Module number	M_1	Predicted aa truncated (PARAS)	Active site signature truncated	M_{11}	Predicted aa full (PARAS)	Predicted aa full (Jangra *et al.*, 2019b)	Active site signature full
1	Gly	Gly (1.00)	DILQMGMVWK	ND	Val (1.00)	Ile/Val	DAFWLGGTFK
2	Dab	Dab (1.00)	DVGEISSIDK	ND	Dab (1.00)	Dab	DVGEISSIDK
3	Gly	Gly (1.00)	DILQMGMVWK	ND	Ser (1.00)	Ser	DVWHFSLVDK
4	Ser	Ser (1.00)	DVWHFSLVDK	ND	Gly (1.00)	Gly	DILQMGMVWK
5	Trp	Trp (0.95)	DAWAFAGVAK	ND	Trp (0.95)	Trp	DAWAFAGVAK
6	Ser	Ser (1.00)	DVWHFSLVDK	ND	Ser (1.00)	Ser	DVWHFSLVDK
7	Dab	Dab (1.00)	DVGEISSIDK	ND	Dab (1.00)	Dab	DVGEISSIDK
8	Dab	Dab (1.00)	DVGEISSIDK	Dab	Dab (1.00)	Dab	DVGEISSIDK
9	Ile	Ile (0.98)	DAFFLGITFK	Phe	Phe (0.75)	Phe	DAWTIAGVAK
10	Glu	Glu (0.95)	DAKDLGVVDK	Glu	Glu (0.95)	Glu	DAKDLGVVDK
11	Ile			Ile	Ile (0.95)	Ile/aIle	DAFFLGITFK
12	Ile			Ile	Ile (0.94)	Ile/aIle	DAFFLGVTFK
13	Ser			Ser	Ser (1.00)	Ser	DVWHFSLVDK

Table S5. HPLC and HRMS analysis of the synthetic peptides prepared in this study.

Peptide	Name	Chemical formula	Calcd. exact mass	Mass found	Calcd.	Overall yield [%]
1	Oct-TriB_1	$C_{66}H_{109}N_{17}O_{19}$	1443.8086	722.9122 $[M+2H]^{2+}$	722.9116	27
2	Oct-TriA_5	$C_{74}H_{114}N_{18}O_{19}$	1558.8508	780.4332 $[M+2H]^{2+}$	780.4327	23

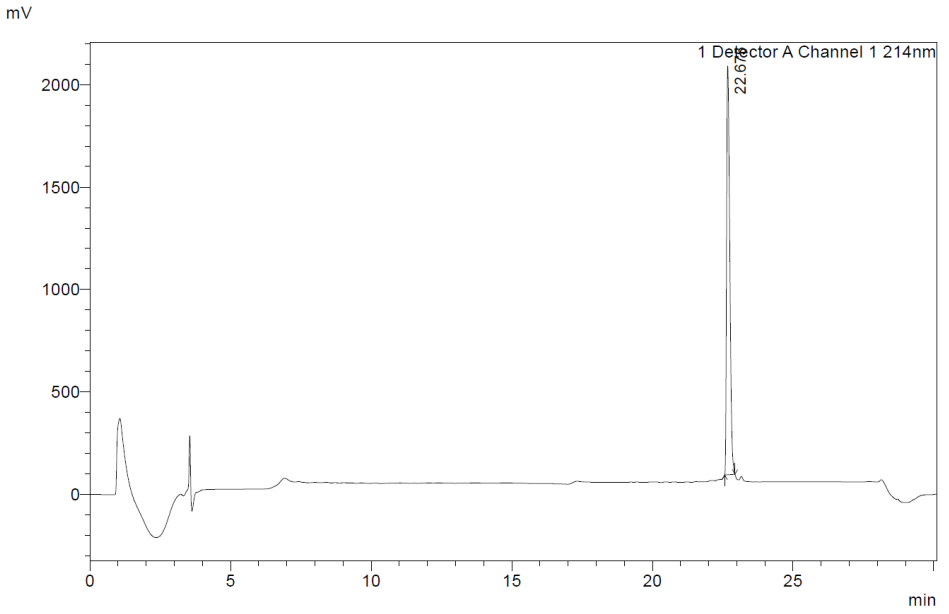


Figure S1. HPLC trace showing the reinjection of purified Oct-TriB$_1$. The peptide eluted as a single peak at 22.675 min using the HPLC method B.

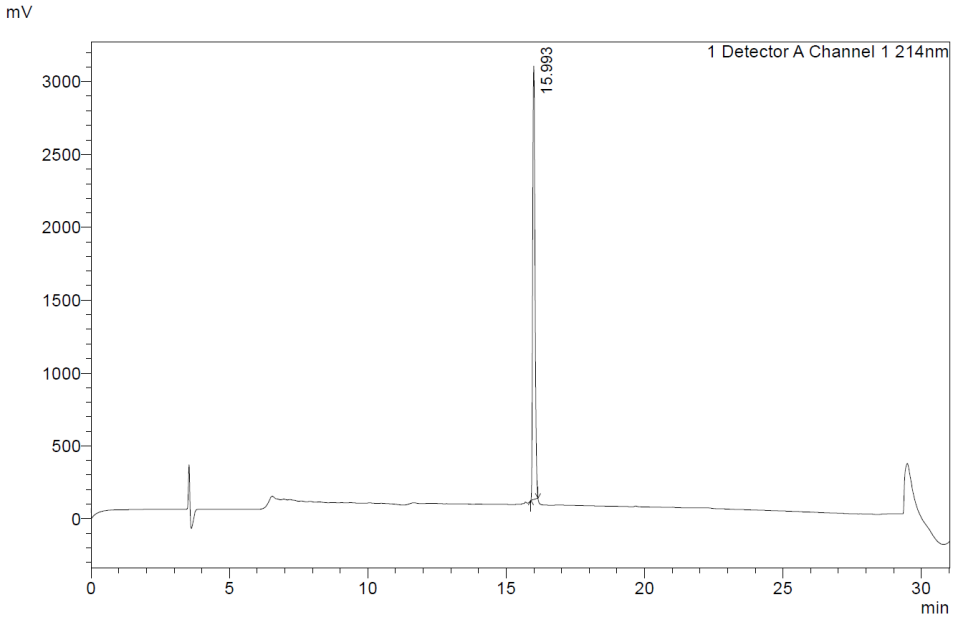


Figure S2. HPLC trace showing the reinjection of purified Oct-TriA$_5$. The peptide eluted as a single peak at 15.993 min using the HPLC method A.



Chapter 9

General discussion

Bacteria are invaluable sources of natural products (NPs), owing to their remarkable capacity to produce a plethora of both primary and specialized metabolites. As a result, bacterial fermentation has long been used to manufacture NPs that have pharmaceutical and agrochemical value. The extensive array of bioactivities exhibited by bacterial natural products is a result of the evolutionarily diversified metabolism, with a main portion still waiting to be discovered (Gavriilidou *et al.*, 2022). The limited rate of antibiotic discovery, coupled with the escalating threat of antibiotic resistance, underscores the pressing need for innovative approaches to fully harness the diversity of bacterial natural products (Silver, 2011, Lewis, 2013).

In this thesis, approaches to aid the discovery of novel antibiotics from bacteria are presented, focusing on new strategies from the perspectives of microbiology, chemistry and bioinformatics. This study encompasses exploring the chemical diversity of specialized metabolites from *Paenibacillus* and *Streptomyces*, eliciting the production of bioactive natural products, identifying novel biosynthetic gene clusters (BGCs), and devising strategies to prioritize bioactive molecules for subsequent isolation, structure elucidation, or organic synthesis.

GNPS molecular networking for the discovery of new bioactive compounds

Global Natural Products Social molecular networking (GNPS MN) is a powerful tool to process and analyze tandem mass spectrometry (MS^2) data (Wang *et al.*, 2016). GNPS MN has been successfully applied to assist in the discovery of new microbial compounds (Nguyen *et al.*, 2016, Wu *et al.*, 2019, Männle *et al.*, 2020) to support strain prioritization, culture conditions, and extraction methods for drug discovery purposes (Crüsemann *et al.*, 2017, Floros *et al.*, 2016), microbial interaction investigations (Watrous *et al.*, 2012, Vallet *et al.*, 2017) and chemical diversity profiling (Purves *et al.*, 2016, Nguyen *et al.*, 2013). 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 spectral data) (Olivon *et al.*, 2017, Nothias *et al.*, 2018, Olivon *et al.*, 2020, Crüsemann *et al.*, 2017). In **Chapter 3**, we introduce an early-stage prioritization method that combines feature-based GNPS MN and MassQL to prioritize peptides that contain basic amino acids (2,4-diaminobutyric acid, lysine, ornithine, histidine, and arginine) with potential antimicrobial properties. Basic amino acid signatures were detected and mapped into massive multi-informative MN acquired over an extract library of 227 *Paenibacillus* isolates from the Auburn University Plant-Associated Microbial strain collection. This led to the discovery of new paenilipoheptins and new bacitracin congeners, designated as paenitracins (Figure 1).

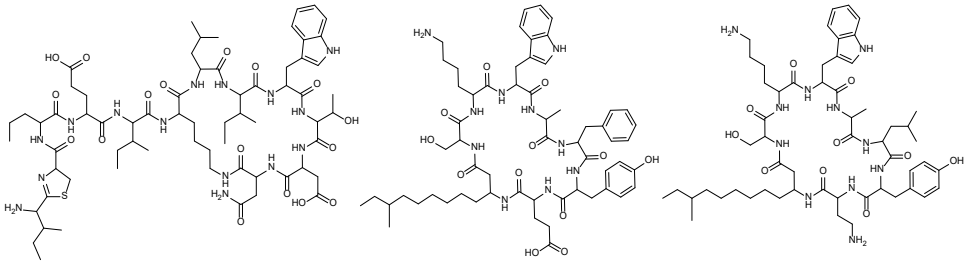


Figure 1. Overview of the compounds discovered by the MassQL MN integrative approach (Chapter 3).

Paenitracin differs from the known bacitracin A at three amino acid positions featuring amino acids that are unique among all bacitracins identified so far, namely Leu7, Trp9, and Thr10 (Suleiman *et al.*, 2017). Despite its significant structural divergence, paenitracin exhibits potent bioactivity against Gram-positive pathogens, including vancomycin-resistant *Enterococcus faecium* E155. Paenitracin is produced by *Paenibacillus* species, as opposed to previously reported bacitracins that have been isolated from *Bacillus* (Johnson *et al.*, 1945, Stone & Strominger, 1971). We used an adopted GNPS MN approach to provide insight into chemical structures and diversity of *Paenibacillus* specialized metabolites and to facilitate the discovery of new natural products.

Chemical elicitors in combination with metabolomics

In our exploration of natural product diversity, metaphorically we are likely still "looking through a keyhole", which only offers a restricted view of the entire chemical landscape of bacterial specialized metabolites. Bacteria do not produce their whole spectrum of specialized metabolites under routine laboratory conditions. Many BGCs are activated only under specific environmental conditions, and the cognate signals are likely missing under laboratory culturing conditions (Demain & Fang, 2000). Understanding the signals and the corresponding regulatory mechanisms will allow the development of eliciting approaches, which are essential for natural product discovery.

Streptomycetes have a complex lifestyle that begins with the germination of a spore, which grows out to form a multicellular network known as the vegetative mycelium. In response to stressors like nutrient depletion, a portion of the mycelium is autolytically degraded, releasing nutrients into the environment for the development of aerial hyphae and spores (Jakimowicz & van Wezel, 2012, Flärdh & Buttner, 2009). Antibiotic production by streptomycetes is linked to the growth phase (Rigali *et al.*, 2008, Barka *et al.*, 2016). Therefore, the modulation of growth conditions plays a critical role in shaping the metabolic output and, consequently, the diversity of bioactive compounds synthesized by these bacteria (van der Heul *et al.*, 2018, Urem *et al.*, 2016). Novel quinazolinone A and B were discovered following

alterations of the growth conditions during cultivation of *Streptomyces* sp. MBT27 (**Chapter 4**). We discovered that the production of quinazolinone alkaloids was induced in response to glycerol. The chemical structures of quinazolinones A and B and the glycerol-dependent production suggest that glycerol participates in the construction of the ring system (Figure 2). Our data indicate that quinazolinones A and B form a new sub-branch in the family of quinazolines. Identification and an in-depth analysis of the biosynthetic gene cluster should reveal the precise biosynthetic pathway for these novel molecules.

Moreover, carbon sources had a significant effect on the antimicrobial activity of *Streptomyces* sp. MBT27 (**Chapter 5**). Antimicrobial activity was most pronounced when the culture medium was supplemented with glycerol + mannitol, glucose, glycerol or fructose. Statistical analysis showed that the bioactivity correlated with elevated production of actinomycins, the first antibiotics ever discovered in an actinomycete (Katz, 1967). Importantly, carbon utilization not only affected the overall production levels of specialized metabolites, but also contributed to the chemical diversity of the actinomycins, whereby production of the novel compound actinomycin L was observed in the cultures grown with 1% glucose, 1% and 2% glycerol, 1% mannitol +1 % glycerol (Figure 2). The isolation of actinomycin L in such a well-studied family of NPs underlines that new molecules can be discovered even within extensively studied microbes and compound classes.

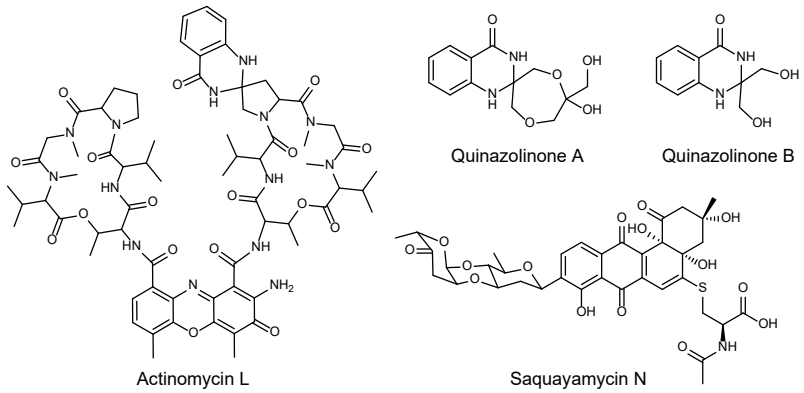


Figure 2. Overview of the compounds discovered by the elicitation techniques (Chapters 4, 5, and 6).

To enhance early-stage prioritization of bioactive compounds, we developed an analytical platform for the rapid and efficient identification and dereplication of bioactive metabolites based on nanofractionation (nanoRAPIDS) (**Chapter 6**). NanoRAPIDS combines at-line high-resolution nanofractionation coupled to LC-MS/MS and featured-based molecular networking (FBMN) in a single analysis. The nanoRAPIDS platform allows the identification of low-abundance molecules against a background of abundantly present known antibiotics.

We used the platform to correlate antimicrobial bioactivities to compounds in extracts of *Streptomyces* sp. MBT84 grown in the presence or absence of catechol as the elicitor (van Bergeijk *et al.*, 2022). The majority of the bioactivity peaks were associated with mass features annotated as angucyclines, which predominantly contributed to the increased bioactivity of the crude extract of *Streptomyces* sp. MBT84 when grown with catechol (van Bergeijk *et al.*, 2022). This afforded the isolation of a novel angucycline, saquayamycin N. The combination of various elicitation strategies with multi-omics approaches and nanoRAPIDS platform offers a powerful workflow to prioritize bioactive metabolites for isolation and assist in 'decrypting' cryptic BGCs.

Global genome mining as a way to access hidden chemical diversity

Genome mining methodologies complement the capabilities of analytical chemistry techniques by accessing specialized metabolites that may not be produced under any tested growth conditions (Sayed *et al.*, 2020, Gavriilidou *et al.*, 2022, Lloyd Karen *et al.*, 2018). Genome mining-guided natural product discovery involves detecting novel BGCs, predicting the chemical structures of the molecules of interest and estimating their bioactivity based on these structures (Medema & Fischbach, 2015). The exploration of bacterial nonribosomal peptide synthetases (NRPSs) through global-genome mining has shown a substantial genomic potential within Firmicutes, with the majority of NRPSs remaining unexplored (Li *et al.*, 2018). *Paenibacillus* spp. are important producers of nonribosomal peptides (NRPs), including polymyxins, octapeptins, pelgipeptins, fusaricidins, tridecaptins, and cilagicins (Cochrane & Vederas, 2016). These NRPs exhibit a broad range of activities, ranging from antibacterial and antifungal to anticancer and antiviral. To gain insight into the chemical diversity of NRPs and discover new potentially bioactive natural products, we analyzed 785 complete publicly available *Paenibacillus* genomes using BiG-SCAPE (Navarro-Muñoz *et al.*, 2020). (**Chapters 7** and **8**). Further in silico analysis using PARAS facilitated the prediction of amino acid compositions for peptide scaffolds, leading to the discovery of ten novel tridecaptin BGCs with variations in amino acid composition compared to known scaffolds.

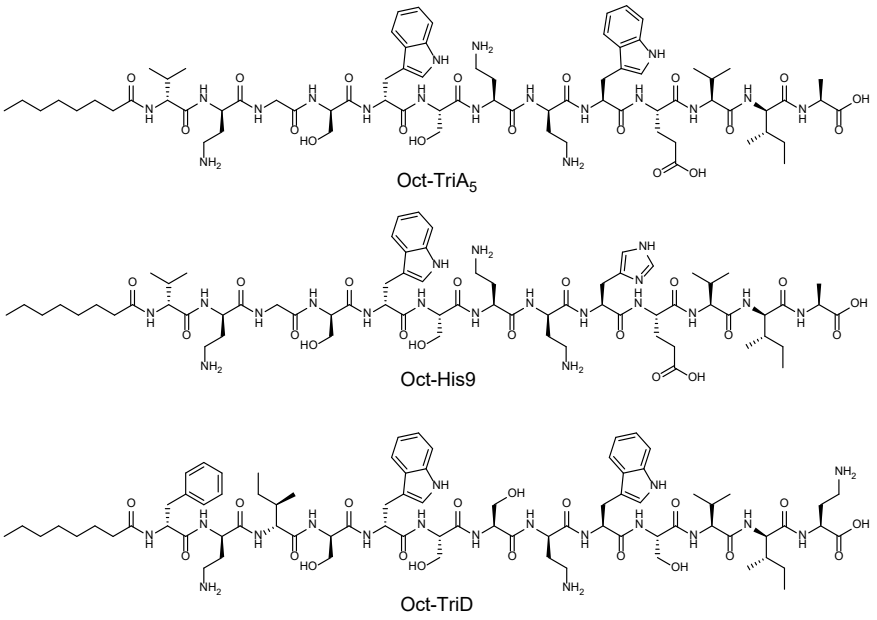


Figure 3. Syn-BNPs discovered by global genome mining of *Paenibacillus* NRPs.

We identified significant diversification within the tridecaptin family, whereby chemically synthesized congeners demonstrated distinct bioactivity profiles compared to known tridecaptins (**Figure 3**). Tridecaptin A_5 was discovered through global genome mining, identified by metabolomic studies, and produced synthetically. Considering the striking difference in the spectrum of activity between known tridecaptin A_1 (active against Gram-negatives) and new tridecaptin A_5 (active against both Gram-positives and Gram-negatives), we were able to pinpoint a specific residue (position 9) responsible for altering the antibacterial spectrum and cytotoxicity profile of these compounds. Single amino acid substitution in position 9 led to the discovery of Oct-His9, which efficiently inhibited colistin-resistant *Pseudomonas aeruginosa* and showed lower hemolytic and cytotoxic activity compared to Oct-$TriA_1$. Global genome mining also revealed Oct-TriD, that contained 6 substitutions as compared to Oct-$TriA_1$ and showed significant bioactivity against *E. faecium*. Thus, it is important to continue to explore tridecaptins towards the development of effective antibacterial agents. The approach of synthetic-bioinformatic natural product (syn-BNP) represents a cutting-edge methodology at the intersection of computational biology, synthetic chemistry, and microbiology (Chu *et al.*, 2020). Integrating single amino acid substitution strategies into syn-BNP development enables researchers to systematically explore the structure-activity relationships of natural products, facilitating the rational design of analogs with enhanced therapeutic potential.

Chimeric biosynthesis of tridecaptins and actinomycin

Analysis of *Paenibacillus* genomes using BiG-SCAPE revealed that some of the tridecaptin BGCs differ not only by the predicted amino acid composition but also by their architecture. These "truncated BGCs" contain the 10-modular NRPS *triD* gene but lack the gene encoding the terminal 3-modular TriE synthase. Suggestively, the truncated tridecaptin-like BGC always co-localizes either with a full-length BGC containing all 13 modules or with another truncated tridecaptin-like BGC that encodes a 3-modular TriE synthase. Considering the widespread co-occurrence of full-length and truncated BGCs, we hypothesized that both tridecaptin BGCs are functional and participate in the biosynthesis of so-called "chimeric tridecaptins". Through MS-based metabolomic analysis of 227 plant-associated *Paenibacillus* spp. we detected the chimeric tridecaptin B in the crude extracts of seven *Paenibacillus* isolates. Genome sequencing of one such isolate, *Paenibacillus* spp. JJ-1683, identified both full-length and truncated tridecaptin BGCs. Therefore, we conclude that tridecaptin B is a product of collaborative biosynthesis, where the NRPS of the truncated cluster provides the first ten amino acids, while the full-length cluster contributes the last three.

Another instance of collaborative biosynthesis is seen in the production of actinomycin L (**Chapter 5**). The biosynthesis of actinomycin L requires genes outside of the actinomycin cluster for the production of anthranilamide, one of the precursors of actinomycin L. The feeding experiments convincingly showed that actinomycin L is formed through the reaction of anthranilamide with the 4-keto group on the proline residue in the pentapeptide lactone prior to the condensation of the pentapeptide lactones into actinomycin L. The enzyme responsible for converting anthranilic acid to anthranilamide remains to be identified. Chimeric biosynthesis is a mechanism of bacterial and fungal secondary metabolism, which has the potential to increase further the structural diversity of encoded molecules and subsequent biological activity.

Conclusions and future perspectives

Integration of diverse data types, such as metabolomics, genomic, transcriptomic, and phenotypic data, provides a holistic insight into the antibiotic-producing potential of bacteria and to prioritize the antibiotics they produce. To make new scientific breakthroughs in this important field, we need to find ways to explore undiscovered natural product diversity. Artificial intelligence (AI) techniques are a very promising addition to the available tools and enable combining information from multiple sources to uncover complex relationships between molecular structures, biological functions, and bioactivity (Mullowney *et al.*, 2023). AI developments are expected to cause a revolution in the discovery of bioactive NPs and the design of NP-inspired drugs and their future development.

Besides finding novel chemical entities, microbes are biofactories for the development of exogenous metabolites, peptides, and proteins through recombinant DNA

technology (Pham *et al.*, 2019). The future steps in metabolic engineering should involve genetic manipulations of NP biosynthetic powerhouses like S*treptomyces* (Whitford *et al.*, 2021), so as to develop novel and complex NPs in a more sustainable manner.

This research has demonstrated that a combined use of metabolomic and genomic approaches is an effective way to discover new NPs. By integrating genomic insights with metabolomics analyses, we have discovered a rich diversity of chemical structures and bioactivities even within known families of molecules, including quinazolinones, actinomycins, saquayamycins, tridecaptins, paenilipoheptins, and bacitracins. This strategy not only facilitates the identification of previously unexplored natural products but also offers insights into their biosynthetic pathways and potential applications.

By harnessing the collective power of AI technologies, natural product chemistry, synthetic biology, and genome mining strategies, we can unlock the vast reservoir of microbial biodiversity and pave the way for the development of next-generation antibiotics to combat emerging infectious diseases and address the growing threat of antimicrobial resistance.



Nederlandse samenvatting

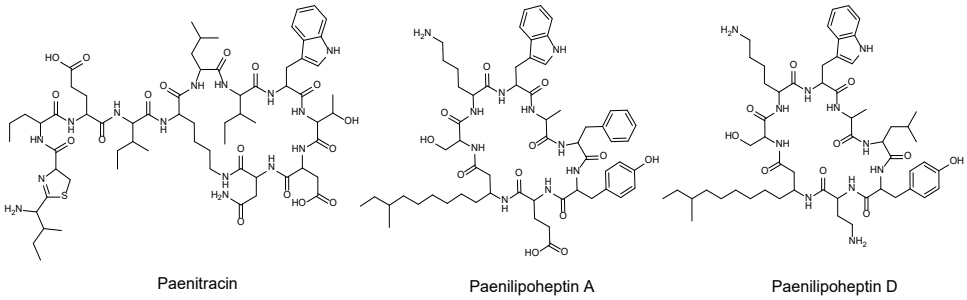
Bacteriën zijn onmisbare en schier onuitputtelijke bronnen van natuurstoffen, wat te danken is aan hun opmerkelijke vermogen om een overvloed aan zowel primaire als gespecialiseerde metabolieten te produceren. Als gevolg hiervan wordt bacteriële fermentaties al lange tijd gebruikt voor de productie van natuurstoffen en derivaten daarvan die toepassing vinden in de landbouw, biotechnologie of de kliniek. Het uitgebreide scala aan bioactiviteiten van natuurstoffen is het resultaat van diversificatie door evolutie van het metabolisme, waarbij duidelijk is dat het grootste deel van de chemische ruimte van natuurstoffen nog ontdekt moet worden (Gavriilidou et al., 2022). De beperkte snelheid van het ontdekken van antibiotica, in combinatie met de toenemende dreiging van antibioticaresistentie, zorgt ervoor dat de noodzaak hoog is om innovatieve benaderingen te ontwikkelen om de diversiteit van bacteriële natuurstoffen volledig te benutten (Silver, 2011; Lewis, 2013).

In dit proefschrift worden benaderingen gepresenteerd om de ontdekking van nieuwe antibiotica uit bacteriën te ondersteunen, met nieuwe strategieën gebaseerd op microbiologie, chemie en bioinformatica. Dit onderzoek richtte zich op het verkennen en uitnutten van de chemische diversiteit van natuurstoffen van *Paenibacillus* en *Streptomyces*, het activeren van de productie van bioactieve natuurstoffen, het identificeren van nieuwe biosynthetische genclusters (BGC's) en het bedenken van strategieën om bioactieve moleculen te prioriteren voor daaropvolgende isolatie, structuuropheldering of organische synthese. Tot slot wed van veel van de moleculen gekeken naar de bioactiviteit en mogelijke toekomstige ontwikkeling tot nieuw medicijn.

GNPS-moleculaire netwerken voor de ontdekking van nieuwe bioactieve verbindingen

Global Natural Products Social molecular networking (GNPS MN) is een krachtig instrument om tandemmassaspectrometrie (MS^2) gegevens te verwerken en analyseren (Wang et al., 2016). GNPS MN is met succes toegepast om te assisteren bij de ontdekking van nieuwe microbiële verbindingen (Nguyen et al., 2016; Wu et al., 2019; Männle et al., 2020) om de prioritering van productiestammen, groeiomstandigheden en extractiemethoden voor het ontdekken van medicijnen te ondersteunen (Crüsemann et al., 2017; Floros et al., 2016). Tevens vindt het toepassing in het onderzoeken van microbiële interacties (Watrous et al., 2012; Vallet et al., 2017) en profilering van chemische diversiteit (Purves et al., 2016; Nguyen et al., 2013). Bovendien zijn prioriteringsstrategieën voor het ontdekken van nieuwe natuurstoffen toegepast op basis van de combinatie van moleculaire netwerken met verschillende informatiebronnen (bijv. genomische, taxonomische, geografische, bioactiviteits- en spectrale gegevens) (Olivon et al., 2017; Nothias et al., 2018; Olivon et al., 2020; Crüsemann et al., 2017). In **Hoofdstuk 3** wordt een methode geïntroduceerd voor het in een vroeg stadium prioriteren van moleculen, waarbij feature-based GNPS MN en MassQL wordt gecombineerd om peptiden die basische aminozuren bevatten (2,4-diaminoboterzuur, lysine, ornithine, histidine en arginine) en mogelijk interessante antimicrobiële eigenschappen hebben, te prioriteren. Annotatie van basische aminozuren werd gedaan op basis van multifunctionele MNs verkregen uit een bibliotheek van extracten van 227 isolaten uit de Auburn University stammenverzameling

bestaande uit plant-geassocieerde *Paenibacillus* stammen. Dit leidde tot de ontdekking van nieuwe paenilipoheptins en nieuwe bacitracinvarianten, aangeduid als paenitracins (Figuur 1).



Figuur 1. Overzicht van de verbindingen ontdekt door de MassQL Molecular Networking aanpak (Hoofdstuk 3).

Paenitracin verschilt van het bekende bacitracin A op drie posities, met aminozuren die uniek zijn ten opzichte van alle tot nu toe geïdentificeerde bacitracins, namelijk Leu7, Trp9 en Thr10 (Suleiman et al., 2017). Ondanks de significante structurele veranderingen vertoont paenitracine krachtige bioactiviteit tegen Gram-positieve pathogenen, waaronder vancomycine-resistente *Enterococcus faecium* E155. Paenitracine wordt geproduceerd door *Paenibacillus*-soorten, in tegenstelling tot eerder beschreven bacitracins die geïsoleerd werden uit *Bacillus* (Johnson et al., 1945, Stone & Strominger, 1971). We gebruikten een aangenomen GNPS MN-benadering om inzicht te krijgen in de chemische structuren en diversiteit van gespecialiseerde metabolieten van *Paenibacillus* en om de ontdekking van nieuwe natuurstoffen te vergemakkelijken.

Chemische elicitors in combinatie met metabolomics

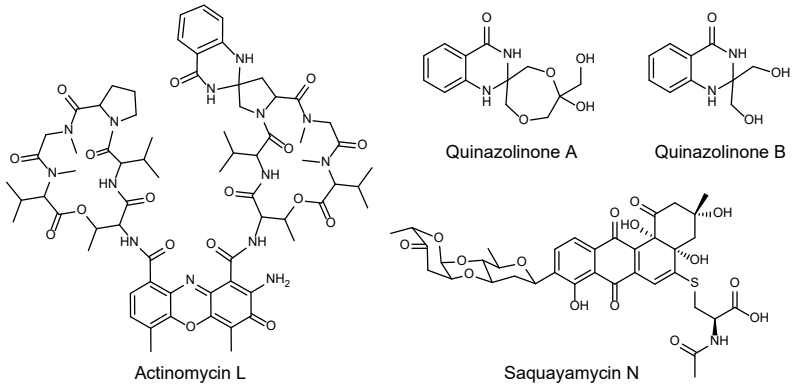
Ons streven om de diversiteit van natuurstoffen te begrijpen, laat zich vergelijken met het kijken door een sleutelgat, wat slechts een beperkte blik biedt op het uitgestrekte chemische landschap van bacteriële gespecialiseerde metabolieten. Bacteriën laten hun volledige spectrum aan gespecialiseerde metabolieten niet zien onder standaard laboratoriumomstandigheden. Veel biosynthetische genclusters (BGC's) blijven inactief of 'slapend' totdat ze worden getriggerd door specifieke omgevingsignalen, die vaak afwezig zijn onder laboratoriumomstandigheden (Demain & Fang, 2000). Het ontcijferen van deze signalen en regelmechanismen is cruciaal voor het ontwikkelen van strategieën die noodzakelijk zijn voor het aanzetten van de biosynthese van natuurstoffen.

Streptomyceten hebben een complexe levensstijl, beginnend met een spore die ontkiemt en vervolgens uitgroeit tot een multicellulair netwerk dat bekend staat als het vegetatieve mycelium. IN reactie op stressfactoren zoals uitputting van voedingsstoffen ondergaat een deel van het mycelium autolytische afbraak, waarbij voedingsstoffen in de omgeving vrijkomen om de groei van luchthyfen en sporen te ondersteunen (Jakimowicz & van Wezel, 2012, Flärdh & Buttner, 2009). De productie van antibiotica door streptomyceten

is nauw verbonden met de levenscyclus (Rigali et al., 2008, Barka et al., 2016). Daarbij beïnvloedt het moduleren van de groeiomstandigheden het metabolisme aanzienlijk en daarmee ook de diversiteit van bioactieve verbindingen die gesynthetiseerd worden door deze bacteriën (van der Heul et al., 2018, Urem et al., 2016).

Nieuwe quinazolinon A en B zijn ontdekt na het veranderen van de groeiomstandigheden bij het kweken van *Streptomyces* sp. MBT27 (**Hoofdstuk 4**). De productie van quinazolinon alkaloïden werd sterk verbeterd door de toevoeging van glycerol als C-bron. De chemische structuren van quinazolinon A en B en hun glycerol-afhankelijke productie suggereren dat glycerol bijdraagt aan de constructie van het ringsysteem (Figuur 2). Quinazolinone A en B vertegenwoordigen een nieuwe tak in de familie van quinazolinones. Identificatie en een uitgebreide analyse van het biosynthetische gencluster zullen naar verwachting het precieze biosynthetische pad voor deze nieuwe moleculen onthullen.

Bovendien heeft de specifieke koolstofbron een significante invloed op de antimicrobiële activiteit van *Streptomyces* sp. MBT27 (**Hoofdstuk 5**). De antimicrobiële activiteit piekte wanneer glycerol + mannitol, glucose, glycerol of fructose werden gebruikt als C-bron. Statistische analyse liet een directe correlatie zien tussen bioactiviteit en verhoogde productie van actinomycins; dit zijn de eerste antibiotica die zijn ontdekt in een actinomyceet (Katz, 1967). Opmerkelijk is dat koolstofgebruik niet alleen de algehele productieniveau's van gespecialiseerde metabolieten beïnvloedde, maar ook bijdroeg aan de chemische diversiteit van de actinomycines. Productie van de nieuwe verbinding actinomycine L werd waargenomen in culturen met 1% glucose, 1% of 2% glycerol, of 1% mannitol + 1% glycerol (Figuur 2). De isolatie van het nieuwe actinomycine L binnen zo'n goed bestudeerde familie van natuurstoffen benadrukt het potentieel voor het ontdekken van nieuwe moleculen, zelfs binnen uitgebreid bestudeerde microben en verbindingenklassen.



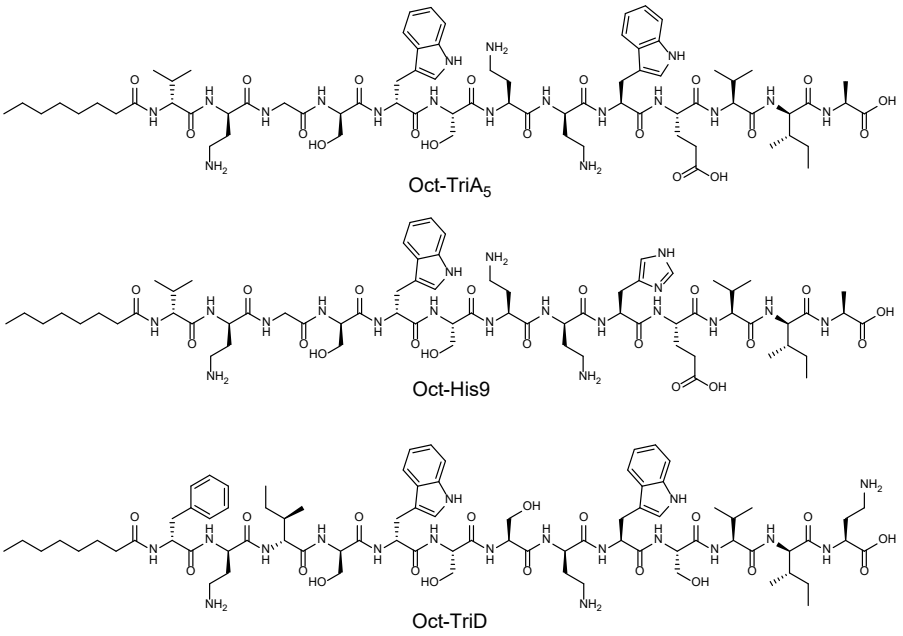
Figuur 2. Overzicht van nieuwe verbindingen die zijn ontdekt na het aanpassen van de groeicondities (Hoofdstukken 4, 5 en 6).

Om de vroegtijdige prioritering van bioactieve verbindingen te verbeteren, hebben we een analytisch platform ontwikkeld voor de snelle en efficiënte identificatie en dereplicatie van bioactieve metabolieten op basis van nanofractionering (nanoRAPIDS) (**Hoofdstuk 6**).

NanoRAPIDS combineert hoge resolutie nanofractionering met LC-MS/MS en functionele moleculaire netwerken (FBMN) in één enkele analyse. Het nanoRAPIDS-platform maakt de identificatie van moleculen met lage concentraties mogelijk tegen een achtergrond van bekende antibiotica, die vaak juist in hoge concentraties voorkomen. We gebruikten het platform om antimicrobiële bioactiviteiten te correleren met verbindingen in extracten van *Streptomyces* sp. MBT84 gekweekt in cultures met of zonder catechol als elicitor (van Bergeijk et al., 2022). De meerderheid van de pieken van bioactiviteit was geassocieerd met mass features die behoorden bij angucyclins, die voornamelijk bijdroegen aan de verhoogde bioactiviteit van het ruwe extract van *Streptomyces* sp. MBT84 wanneer het gekweekt werd met catechol (van Bergeijk et al., 2022). Dit leidde tot de isolatie van een nieuw angucycline, saquayamycine N. De combinatie van verschillende elicitatiestrategieën met multi-omics-benaderingen en het nanoRAPIDS-platform biedt een krachtige workflow om bioactieve metabolieten te prioriteren voor isolatie en bij te dragen aan het 'ontcijferen' van cryptische BGC's.

Grootschalig genoomonderzoek als een manier om verborgen chemische diversiteit te benaderen

Methoden om genomen te sequencen en te annoteren complementeren analytisch chemische techniekendoor toegang te bieden tot gespecialiseerde metabolieten die mogelijk niet worden geproduceerd in het laboratorium (Sayed et al., 2020, Gavriilidou et al., 2022, Lloyd Karen et al., 2018). Het analyseren van genomen als basis voor het ontdekken van natuurstoffen omvat het detecteren van nieuwe BGC's, het voorspellen van de chemische structuren van de relevante moleculen en het inschatten van hun bioactiviteit op basis van deze structuren (Medema & Fischbach, 2015). De verkenning van bacteriële niet-ribosomale peptidesynthetasen (NRPS's) via grootschalige genoomanalyses heeft een aanzienlijk genoompotentieel binnen Firmicutes aangetoond, waarbij de functie van de meerderheid van de NRPS's onderbelicht blijft (Li et al., 2018*). Paenibacillus* spp. zijn belangrijke producenten van niet-ribosomale peptiden (NRPs), waaronder polymyxins, octapeptins, pelgipeptins, fusaricidins, tridecaptins en cilagicins (Cochrane & Vederas, 2016). Deze NRPs vertonen een breed scala aan activiteiten, variërend van antibacterieel en antischimmel tot antikanker en antiviraal. Om inzicht te krijgen in de chemische diversiteit van NRPs en nieuwe potentieel bioactieve natuurstoffen te ontdekken, hebben we 785 volledige, openbaar beschikbare genomen van *Paenibacillus* geanalyseerd met behulp van BiG-SCAPE (Navarro-Muñoz et al., 2020). (Hoofdstukken 7 en 8). Verdere in silico analyse met behulp van het algoritme PARAS maakte mogelijk om de aminozuursamenstelling voor peptide-scaffolds beter te voorspellen, wat heeft geleid tot de ontdekking van tien nieuwe tridecaptin-BGC's met variaties in aminozuursamenstelling in vergelijking met bekende scaffolds.



Figuur 3. Syn-BNPs ontdekt door wereldwijde genoomontginning van *Paenibacillus* NRPs.

We zagen daarbij significante diversificatie binnen de tridecaptin familie, waarbij chemisch gesynthetiseerde varianten verschillende bioactiviteitsprofielen vertoonden ten opzichte van bekende tridecaptins (Figuur 3). Tridecaptin A_5 werd ontdekt door analyse van genoomdatabasesanalyses, vervolgens geïdentificeerd door metaboloomstudies en tot slot synthetisch geproduceerd. Gezien het opvallende verschil in het activiteitsspectrum tussen de bekende tridecaptin A_1 (actief tegen Gram-negatieven) en de nieuwe tridecaptin A_5 (actief tegen zowel Gram-positieve als Gram-negatieve bacteriën), konden we een specifiek residu (positie 9) identificeren dat een sleutelrol speelt in het veranderen van het antibacteriële spectrum en het profiel van de cytotoxiciteit van deze verbindingen. Een enkele aminozuursubstitutie op positie 9 leidde tot de ontdekking van Oct-His9, die colistine-resistente *Pseudomonas aeruginosa* efficiënt remde en lagere hemolytische en cytotoxische activiteit vertoonde in vergelijking met Oct-$TriA_1$. Wereldwijde genoomontginning onthulde ook Oct-TriD, dat 6 substituties bevatte in vergelijking met Oct-$TriA_1$ en aanzienlijke bioactiviteit vertoonde tegen *E. faecium*. Het is daarom belangrijk om tridecaptins verder te blijven verkennen voor de ontwikkeling van effectieve antibacteriële middelen. De benadering om synthetische biologie te combineren met bioinformatica om natuurstoffen te analyseren en produceren (syn-BNP) vertegenwoordigt een geavanceerde methodologie op het snijvlak van computationele biologie, synthetische chemie en microbiologie (Chu et al., 2020). Het integreren van strategieën voor aminozuursubstituties bij syn-BNP stelt onderzoekers in

staat systematisch de structuur-activiteitsrelaties van natuurstoffen te analyseren, om zo rational design van analogen met verbeterd therapeutisch potentieel te vergemakkelijken.

Chimere biosynthese van tridecaptins en actinomycine

Analyse van *Paenibacillus* genomen met behulp van BiG-SCAPE heeft aangetoond dat sommige van de tridecaptin BGC's niet alleen verschillen in de voorspelde aminozuursamenstelling, maar ook in hun architectuur. Deze "verkorte BGC's" bevatten het 10-modulaire NRPS triD-gen maar missen het gen dat codeert voor de terminale 3-modulaire TriE-synthase. Het is suggestief dat de verkorte tridecaptin-achtige BGC's altijd samenvallen met ofwel een volledige BGC met alle 13 modules of met een andere verkorte tridecaptinachtige BGC die een 3-modulaire TriE-synthase codeert. Gezien het wijdverbreide voorkomen van zowel een volledige als een verkorte BGC, kan worden aangenomen dat beide tridecaptin BGC's functioneel zijn en deelnemen aan de biosynthese van zogenaamde "chimerische tridecaptins". Via MS-gebaseerde metaboloomanalyse van 227 plant-geassocieerde *Paenibacillus* spp. detecteerden we de chimerische tridecaptine B in de ruwe extracten van zeven *Paenibacillus*-isolaten. Genoomsequentie van een dergelijk isolaat, *Paenibacillus* spp. JJ-1683, identificeerde zowel volledige als verkorte tridecaptine BGC's. Daarom concluderen we dat tridecaptine B een product is van collaboratieve biosynthese, waarbij het NRPS van het verkorte cluster de eerste tien aminozuren levert, terwijl enzymen gecodeerd door het volledige cluster zorgen voor de toevoeging van de laatste drie aminozuren (**Hoofdstuk 8**). Een ander geval van collaboratieve biosynthese is te zien in de productie van actinomycine L (**Hoofdstuk 5**). De biosynthese van actinomycine L vereist genen van buiten het actinomycine-cluster voor de productie van anthranilamide, een van de precursors van actinomycine L. De voedingsexperimenten toonden overtuigend aan dat actinomycine L wordt gevormd door de reactie van anthranilamide met de 4-keto-groep op het prolinerestant in de pentapeptidelacton vóór de condensatie van de pentapeptidelactonen tot actinomycine L. Het enzym dat verantwoordelijk is voor de omzetting van anthranilzuur naar anthranilamide moet nog worden geïdentificeerd. Chimere biosynthese is een mechanisme van secundair metabolisme van bacteriën en schimmels, dat het potentieel heeft om de structurele diversiteit van gecodeerde moleculen verder te vergroten en de daaropvolgende biologische activiteit.

Conclusies en toekomstperspectieven

De integratie van diverse soorten data, zoals metabolomics, genomics, transcriptomics en fenotypische gegevens, biedt een holistisch inzicht in het antibiotica producerende potentieel van bacteriën en om de antibiotica die ze produceren te prioriteren. Om nieuwe wetenschappelijke doorbraken te maken in dit belangrijke veld, moeten we manieren vinden om onontdekte diversiteit van natuurstoffen te verkennen. Kunstmatige intelligentie cq

artificial inteligence (AI) is een veelbelovende toevoeging aan de beschikbare tools en stelt ons in staat informatie uit meerdere bronnen te combineren om complexe relaties tussen moleculaire structuren, biologische functies en bioactiviteit te identificeren (Mullowney et al., 2023). Ontwikkelingen in de AI zullen naar verwachting een revolutie teweeg brengen in het ontdekken van bioactieve natuurstoffen en het ontwerp van op deze moleculen geïnspireerde medicijnen en hun toekomstige ontwikkeling. Naast het vinden van nieuwe chemische entiteiten zijn microben biofabrieken voor de ontwikkeling van metabolieten, peptiden en eiwitten door middel van recombinant DNA technologie (Pham et al., 2019). De toekomstige stappen in metabolic engineering zouden genetische manipulaties van biosynthetische krachtpatsers zoals *Streptomyces* moeten omvatten (Whitford et al., 2021), om zo nieuwe en complexe natuurstoffen op een duurzamere manier te ontwikkelen. Dit onderzoek heeft aangetoond dat een gecombineerd gebruik van metabolomics en genomics benaderingen een effectieve manier is om nieuwe natuurstoffen te ontdekken. Door genomische inzichten te integreren met metaboloomanalyses, hebben we een rijke diversiteit aan chemische structuren en bioactiviteiten ontdekt, zelfs binnen bekende families van moleculen, waaronder quinazolinones, actinomycins, saquayamycins, tridecaptins, paenilipoheptins en bacitracins. Deze strategie vergemakkelijkt niet alleen de identificatie van eerder onontdekte natuurstoffen, maar biedt ook inzichten in de biosynthetische routes en hun potentiële toepassingen. Door gebruik te maken van de collectieve kracht van AI-technologieën, natuurstofchemie, synthetische biologie en genoomanalyse, kunnen we het uitgebreide reservoir van microbiële biodiversiteit ontgrendelen en paden effenen naar de ontwikkeling van antibiotica van de volgende generatie om zo infectieziekten te bestrijden en het groeiende gevaar van antimicrobiële resistentie aan te pakken.



References

Akhter, N., Liu, Y., Auckloo, B.N., Shi, Y., Wang, K., Chen, J., *et al.* (2018) Stress-driven discovery of new angucycline-type antibiotics from a marine *Streptomyces pratensis* NA-ZhouS1. *Mar Drugs* **16**: 192.

Al Ayed, K., Zamarbide Losada, D., Machushynets, N.V., Terlouw, B., Elsayed, S.S., Schill, J., *et al.* (2023) Total synthesis and structure assignment of the relacidine lipopeptide antibiotics and preparation of analogues with enhanced stability. *ACS Infect Dis* **9**: 739-748.

Alanjary, M., Kronmiller, B., Adamek, M., Blin, K., Weber, T., Huson, D., *et al.* (2017) The Antibiotic Resistant Target Seeker (ARTS), an exploration engine for antibiotic cluster prioritization and novel drug target discovery. *Nucleic Acids Res* **45**: W42-W48.

Aleti, G., Sessitsch, A., and Brader, G. (2015) Genome mining: prediction of lipopeptides and polyketides from *Bacillus* and related Firmicutes. *Comput Struct Biotechnol J* **13**: 192-203.

André, B., (2005) *Antimicrobial agents: antibacterials and antifungals*. ASM Press, Washington, D.C.

Atanasov, A.G., Zotchev, S.B., Dirsch, V.M., Orhan, I.E., Banach, M., Rollinger, J.M., *et al.* (2021) Natural products in drug discovery: advances and opportunities. *Nat Rev Drug Discov* **20**: 200-216.

Audoin, C., Bonhomme, D., Ivanisevic, J., Cruz, M.D.l., Cautain, B., Monteiro, M.C., *et al.* (2013) Balibalosides, an original family of glucosylated sesterterpenes produced by the mediterranean sponge *Oscarella balibaloi*. *Mar Drugs* **11**: 1477-1489.

Avalos, M., Garbeva, P., Vader, L., van Wezel, G.P., Dickschat, J.S., and Ulanova, D. (2022) Biosynthesis, evolution and ecology of microbial terpenoids. *Nat Prod Rep* **39**: 249-272.

Ballantine, R.D., Li, Y.X., Qian, P.Y., and Cochrane, S.A. (2018) Rational design of new cyclic analogues of the antimicrobial lipopeptide tridecaptin A_1. *Chem Comm* **54**: 10634-10637.

Ballantine, R.D., McCallion, C.E., Nassour, E., Tokajian, S., and Cochrane, S.A. (2019) Tridecaptin-inspired antimicrobial peptides with activity against multidrug-resistant Gram-negative bacteria. *Medchemcomm* **10**: 484-487.

Baltz, R.H. (2008) Renaissance in antibacterial discovery from actinomycetes. *Curr Opin Pharmacol* **8**: 557-563.

Baltz, R.H. (2021) Genome mining for drug discovery: progress at the front end. *J Ind Microbiol Biotechnol* **48**: kuab044.

Bann, S.J., Ballantine, R.D., and Cochrane, S.A. (2021) The tridecaptins: non-ribosomal peptides that selectively target Gram-negative bacteria. *RSC Med Chem* **12**: 538-551.

Bann, S.J., Ballantine, R.D., McCallion, C.E., Qian, P.-Y., Li, Y.-X., and Cochrane, S.A. (2019) A chemical-intervention strategy to circumvent peptide hydrolysis by D-stereoselective peptidases. *J Med Chem* **62**: 10466-10472.

Barbe, V., Cruveiller, S., Kunst, F., Lenoble, P., Meurice, G., Sekowska, A., *et al.* (2009) From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. *Microbiology* **155**: 1758-1775.

Barka, E.A., Vatsa, P., Sanchez, L., Gavaut-Vaillant, N., Jacquard, C., Meier-Kolthoff, J., *et al.* (2016) Taxonomy, physiology, and natural products of the *Actinobacteria*. *Microbiol Mol Biol Rev* **80**: 1-43.

Bartel, J., Krumsiek, J., and Theis, F.J. (2013) Statistical methods for the analysis of high-throughput metabolomics data. *Comput Struct Biotechnol J* **4**: e201301009.

Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., *et al.* (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417**: 141-147.

Bérdy, J. (2005) Bioactive microbial metabolites. *J Antibiot* **58**: 1-26.

Bertrand, S., Bohni, N., Schnee, S., Schumpp, O., Gindro, K., and Wolfender, J.-L. (2014) Metabolite induction via microorganism co-culture: a potential way to enhance chemical diversity for drug discovery. *Biotechnol Adv* **32**: 1180-1204.

Biabani, M.A., Baake, M., Lovisetto, B., Laatsch, H., Helmke, E., and Weyland, H. (1998) Anthranilamides: new antimicroalgal active substances from a marine *Streptomyces* sp. *J Antibiot* **51**: 333-340.

Blin, K., Shaw, S., Augustijn, H.E., Reitz, Z.L., Biermann, F., Alanjary, M., *et al.* (2023) antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res* **51**: W46-W50.

Blin, K., Shaw, S., Kloosterman, A.M., Charlop-Powers, Z., van Wezel, G.P., Medema, Marnix H., and Weber, T. (2021) antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* **49**: W29-W35.

Blin, K., Shaw, S., Medema, M.H., and Weber, T. (2024) The antiSMASH database version 4: additional genomes and BGCs, new sequence-based searches and more. *Nucleic Acids Res* **52**: D586-D589.

Bode, H.B., Bethe, B., Hofs, R., and Zeeck, A. (2002) Big effects from small changes: possible ways to explore nature's chemical diversity. *ChemBioChem* **3**: 619-627.

Bouslimani, A., Sanchez, L.M., Garg, N., and Dorrestein, P.C. (2014) Mass spectrometry of natural products: current, emerging and future technologies. *Nat Prod Rep* **31**: 718-729.

Brown, E.D., and Wright, G.D. (2016) Antibacterial drug discovery in the resistance era. *Nature* **529**: 336-343.

Buijs, N., Vlaming, H.C., van Haren, M.J., and Martin, N.I. (2022) Synthetic studies with bacitracin A and preparation of analogues containing alternative zinc binding groups. *ChemBioChem* **23**: e202200547.

Buijs, N.P., Matheson, E.J., Cochrane, S.A., and Martin, N.I. (2023) Targeting membrane-bound bacterial cell wall precursors: a tried and true antibiotic strategy in nature and the clinic. *Chem Commun* **59**: 7685-7703.

Caboche, S., Leclère, V., Pupin, M., Kucherov, G., and Jacques, P. (2010) Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity. *J Bacteriol* **192**: 5143-5150.

Cacho, R.A., Jiang, W., Chooi, Y.H., Walsh, C.T., and Tang, Y. (2012) Identification and Characterization of the Echinocandin B Biosynthetic Gene Cluster from Emericella rugulosa NRRL 11440. *J Am Chem Soc* **134**: 16781-16790.

Caesar, L.K., Montaser, R., Keller, N.P., and Kelleher, N.L. (2021) Metabolomics and genomics in natural products research: complementary tools for targeting new chemical entities. *Nat Prod Rep* **38**: 2041-2065.

Caldeira, A.T., Santos Arteiro, J.M., Coelho, A.V., and Roseiro, J.C. (2011) Combined use of LC-ESI-MS and antifungal tests for rapid identification of bioactive lipopeptides produced by *Bacillus amyloliquefaciens* CCMI 1051. *Process Biochem* **46**: 1738-1746.

Cardoso, P., Glossop, H., Meikle, T.G., Aburto-Medina, A., Conn, C.E., Sarojini, V., and Valery, C. (2021) Molecular engineering of antimicrobial peptides: microbial targets, peptide motifs and translation opportunities. *Biophys Rev* **13**: 35-69.

Carroll, L.M., Larralde, M., Fleck, J.S., Ponnudurai, R., Milanese, A., Cappio, E., and Zeller, G. (2021) Accurate *de novo* identification of biosynthetic gene clusters with GECCO. *bioRxiv*: 2021.2005.2003.442509.

Cesa-Luna, C., Geudens, N., Girard, L., Roo, V.D., Maklad, H.R., Martins, J.C., *et al.* (2023) Charting the Lipopeptidome of Nonpathogenic Pseudomonas. *mSystems* **8**: e00988-00922.

Chambers, M.C., Maclean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., *et al.* (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* **30**: 918–920.

Chevrette, M.G., Aicheler, F., Kohlbacher, O., Currie, C.R., and Medema, M.H. (2017) SANDPUMA: ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across *Actinobacteria*. *Bioinformatics* **33**: 3202-3210.

Chevrette, M.G., Gutiérrez-García, K., Selem-Mojica, N., Aguilar-Martínez, C., Yañez-Olvera, A., Ramos-Aboites, H.E., *et al.* (2020) Evolutionary dynamics of natural product biosynthesis in bacteria. *Nat Prod Rep* **37**: 566-599.

Chin, C.A.O., Peluso, P., Sedlazeck, F.A.-O., Nattestad, M., Concepcion, G.T., Clum, A., *et al.* (2016) Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* **13**: 1050–1054.

Chiu, S., Hancock, A.M., Schofner, B.W., Sniezek, K.J., Soto-Echevarria, N., Leon, G., *et al.* (2022) Causes of polymyxin treatment failure and new derivatives to fill the gap. *J Antibiot* **75**: 593-609.

Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., *et al.* (2018) MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res* **46**: W486-W494.

Chong, J., Yamamoto, M., and Xia, J. (2019) MetaboAnalystR 2.0: from raw spectra to biological insights. *Metabolites* **9**: 57.

Chu, J., Vila-Farres, X., Inoyama, D., Ternei, M., Cohen, L.J., Gordon, E.A., *et al.* (2016) Discovery of MRSA active antibiotics using primary sequence from the human microbiome. *Nat Chem Biol* **12**: 1004-1006.

Chu, J.A.O., Koirala, B.A.O., Forelli, N., Vila-Farres, X., Ternei, M.A., Ali, T., *et al.* (2020) Synthetic-bioinformatic natural product antibiotics with diverse modes of action. *J. Am. Chem. Soc.* **142**: 14158–14168.

Chung, Y.R., Kim, C.H., Hwang, I., and Chun, J. (2000) *Paenibacillus koreensis* sp. nov., a new species that produces an iturin-like antifungal compound. *Int J Syst Evol Microbiol* **50**: 1495-1500.

Cimermancic, P., Medema, M.H., Claesen, J., Kurita, K., Brown, L.C.W., Mavrommatis, K., *et al.* (2014) Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **158**: 412-421.

Clark, R.C., and Reid, J.S. (1995) The analytical calculation of absorption in multifaceted crystals. *Acta Crystallogr* **51**: 887-897.

Clasquin, M.F., Melamud, E., and Rabinowitz, J.D. (2012) LC-MS data processing with MAVEN: a Metabolomic Analysis and Visualization Engine. *Curr Protoc Bioinformatics* **37**: 14.11.

Cochrane, S.A., Findlay, B., Bakhtiary, A., Acedo, J.Z., Rodriguez-Lopez, E.M., Mercier, P., and Vederas, J.C. (2016) Antimicrobial lipopeptide tridecaptin A_1 selectively binds to Gram-negative lipid II. *Proc Natl Acad Sci USA* **113**: 11561-11566.

Cochrane, S.A., Lohans, C.T., Brandelli, J.R., Mulvey, G., Armstrong, G.D., and Vederas, J.C. (2014) Synthesis and structure–activity relationship studies of N-terminal analogues of the antimicrobial peptide tridecaptin A_1. *J Med Chem* **57**: 1127-1131.

Cochrane, S.A., Lohans, C.T., van Belkum, M.J., Bels, M.A., and Vederas, J.C. (2015) Studies on tridecaptin B_1, a lipopeptide with activity against multidrug resistant Gram-negative bacteria. *Org Biomol Chem* **13**: 6073-6081.

Cochrane, S.A., and Vederas, J.C. (2016) Lipopeptides from *Bacillus* and *Paenibacillus* spp.: A gold mine of antibiotic candidates. *Med Res Rev* **36**: 4-31.

Conti, R., Chagas, F.O., Caraballo-Rodriguez, A.M., Melo, W.G., do Nascimento, A.M., Cavalcanti, B.C., *et al.* (2016) Endophytic actinobacteria from the brazilian medicinal plant *Lychnophora ericoides* Mart. and the biological potential of their secondary metabolites. *Chem Biodivers* **13**: 727–736.

Cooper, M.A., and Shlaes, D. (2011) Fix the antibiotics pipeline. *Nature* **472**: 32.

Cornforth, D.M., and Foster, K.R. (2013) Competition sensing: the social side of bacterial stress responses. *Nat Rev Microbiol* **11**: 285-293.

Covington, B.C., and Seyedsayamdost, M.R. (2021) MetEx, a metabolomics explorer application for natural product discovery. *ACS Chem Biol* **16**: 2825-2833.

Craney, A., Ozimok, C., Pimentel-Elardo, S.M., Capretta, A., and Nodwell, J.R. (2012) Chemical perturbation of secondary metabolism demonstrates important links to primary metabolism. *Chem Biol* **19**: 1020-1027.

Crnovcic, I., Ruckert, C., Semsary, S., Lang, M., Kalinowski, J., and Keller, U. (2017) Genetic interrelations in the actinomycin biosynthetic gene clusters of *Streptomyces antibioticus* IMRU 3720 and *Streptomyces chrysomallus* ATCC11523, producers of actinomycin X and actinomycin C. *Adv Appl Bioinform Chem* **2017**: 29-46.

Crnovčić, I., Semsary, S., Vater, J., and Keller, U. (2014) Biosynthetic rivalry of o-aminophenol-carboxylic acids initiates production of hemi-actinomycins in *Streptomyces antibioticus*. *RSC Adv* **4**: 5065-5074.

Crüsemann, M., O'Neill, E.C., Larson, C.B., Melnik, A.V., Floros, D.J., da Silva, R.R., *et al.* (2017) Prioritizing natural product diversity in a collection of 146 bacterial strains based on growth and extraction protocols. *J Nat Prod* **80**: 588-597.

Cruz-Morales, P., Vijgenboom, E., Iruegas-Bocardo, F., Girard, G., Yanez-Guerra, L.A., Ramos-Aboites, H.E., *et al.* (2013) The genome sequence of *Streptomyces lividans* 66 reveals a novel tRNA-dependent peptide biosynthetic system within a metal-related genomic island. *Genome Biol Evol* **5**: 1165-1175.

Culp, E.J., Waglechner, N., Wang, W., Fiebig-Comyn, A.A., Hsu, Y.P., Koteva, K., *et al.* (2020) Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling. *Nature* **578**: 582-587.

da Costa, R.A., Andrade, I.E.P.C., Pinto, O.H.B., de Souza, B.B.P., Fulgêncio, D.L.A., Mendonça, M.L., *et al.* (2022) A novel family of non-secreted tridecaptin lipopeptide produced by *Paenibacillus elgii*. *Amino Acids* **54**: 1477-1489.

da Silva, R.R., Wang, M., Nothias, L.F., van der Hooft, J.J.J., Caraballo-Rodríguez, A.M., Fox, E., *et al.* (2018) Propagating annotations of molecular networks using in silico fragmentation. *PLOS Comput Biol* **14**: e1006089.

Dalili, M., and Chau, P.C. (1988) Production of actinomycin D with immobilized *Streptomyces parvullus* under nitrogen and carbon starvation conditions. *Biotechnol Lett* **10**: 331-336.

Davies, J., and Davies, D. (2010) Origins and evolution of antibiotic resistance. *Microbiol Mol Biol Rev* **74**: 417-433.

de Jonge, N.F., Mildau, K., Meijer, D., Louwen, J.J.R., Bueschl, C., Huber, F., and van der Hooft, J.J.J. (2022) Good practices and recommendations for using and benchmarking computational metabolomics metabolite annotation tools. *Metabolomics* **18**: 103.

de Los Santos, E.L. (2019) NeuRiPP: Neural network identification of RiPP precursor peptides. *Sci Rep* **9**: 13406.

Demain, A.L., and Fang, A., (2000) The natural functions of secondary metabolites. In: History of modern biotechnology I. A. Fiechter (ed). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1-39.

Demarque, D.P., Dusi, R.G., de Sousa, F.D.M., Grossi, S.M., Silvério, M.R.S., Lopes, N.P., and Espindola, L.S. (2020) Mass spectrometry-based metabolomics approach in the isolation of bioactive natural products. *Sci Rep* **10**: 1051.

Dettmer, K., Aronov, P.A., and Hammock, B.D. (2007) Mass spectrometry-based metabolomics. *Mass Spectrom Rev* **26**: 51-78.

Dewick, P.M., (2009) *Medicinal natural products: a biosynthetic approach,* p. 395-396. John Wiley & Sons Ltd, Chichester.

Djoumbou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., *et al.* (2016) ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J Cheminform* **8**: 61.

Donadio, S., Monciardini, P., and Sosio, M. (2007) Polyketide synthases and nonribosomal peptide synthetases: the emerging view from bacterial genomics. *Nat Prod Rep* **24**: 1073-1109.

Dong, Y.Y., Wang, H., Pike, A.C.W., Cochrane, S.A., Hamedzadeh, S., Wyszyński, F.J., *et al.* (2018) Structures of DPAGT1 explain glycosylation disease mechanisms and advance TB antibiotic design. *Cell* **175**: 1045-1058.e1016.

Du, C., and van Wezel, G.P. (2018) Mining for microbial gems: integrating proteomics in the postgenomic natural product discovery pipeline. *Proteomics* **18**: e1700332.

Duban, M., Cociancich, S., and Leclère, V. (2022) Nonribosomal peptide synthesis definitely working out of the rules. *Microorganisms* **10**: 577.

Elsayed, S.S., Genta-Jouve, G., Carrión, V.J., Nibbering, P.H., Siegler, M.A., de Boer, W., *et al.* (2020) Atypical spirotetronate polyketides identified in the underexplored genus *Streptacidiphilus*. *J Org Chem* **85**: 10648-10657.

Erb, A., Krauth, C., Luzhetskyy, A., and Bechthold, A. (2009) Differences in the substrate specificity of glycosyltransferases involved in landomycins A and E biosynthesis. *Appl Microbiol Biotechnol* **83**: 1067-1076.

Ernst, M., Kang, K.B., Caraballo-Rodríguez, A.M., Nothias, L.-F., Wandy, J., Chen, C., *et al.* (2019) MolNetEnhancer: enhanced molecular networks by integrating metabolome mining and annotation tools. *Metabolites* **9**: 144.

Falagas, M.E., and Kasiakou, S.K. (2005) Colistin: the revival of polymyxins for the management of multidrug-resistant Gram-negative bacterial infections. *Clin. Infect Dis* **40**: 1333-1341.

Fan, J., Liao, G., Kindinger, F., Ludwig-Radtke, L., Yin, W.-B., and Li, S.-M. (2019) Peniphenone and penilactone formation in *Penicillium crustosum* via 1,4-Michael additions of ortho-quinone methide from hydroxyclavatol to γ-butyrolactones from crustosic acid. *J Am Chem Soc* **141**: 4225-4229.

Felnagle, E.A., Jackson, E.E., Chan, Y.A., Podevels, A.M., Berti, A.D., McMahon, M.D., and Thomas, M.G. (2008) Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol Pharm* **5**: 191-211.

Flärdh, K., and Buttner, M.J. (2009) *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat Rev Microbiol* **7**: 36-49.

Fleming, A. (2001) On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of *B. influenzae*. 1929. *Br J Exp Pathol* **79**: 780-790.

Floros, D.J., Jensen, P.R., Dorrestein, P.C., and Koyama, N. (2016) A metabolomics guided exploration of marine natural product chemical space. *Metabolomics* **12**: 145.

Fotso, S., Mahmud, T., Zabriskie, T.M., Santosa, D.A., and Proteau, P.J. (2008) Rearranged and unrearranged angucyclinones from Indonesian *Streptomyces* spp. *J Antibiot* **61**: 449-456.

Gao, X., Chooi, Y.H., Ames, B.D., Wang, P., Walsh, C.T., and Tang, Y. (2011) Fungal indole alkaloid biosynthesis: genetic and biochemical investigation of the tryptoquialanine pathway in *Penicillium aethiopicum*. *J Am Chem Soc* **133**: 2729–2741.

Gaudêncio, S.P., Bayram, E., Lukić Bilela, L., Cueto, M., Díaz-Marrero, A.R., Haznedaroglu, B.Z., *et al.* (2023) Advanced methods for natural products discovery: bioactivity

screening, dereplication, metabolomics profiling, genomic sequencing, databases and informatic tools, and structure elucidation. *Mar Drugs* **21**: 308.

Gaudencio, S.P., and Pereira, F. (2015) Dereplication: racing to speed up the natural products discovery process. *Nat Prod Rep* **32**: 779-810.

Gavriilidou, A., Kautsar, S.A., Zaburannyi, N., Krug, D., Müller, R., Medema, M.H., and Ziemert, N. (2022) Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat Microbiol* **7**: 726-735.

Genilloud, O. (2014) The re-emerging role of microbial natural products in antibiotic discovery. *Antonie Van Leeuwenhoek* **106**: 173-188.

Gewolb, J. (2002) Working Outside the Protein-Synthesis Rules. *Science* **295**: 2205-2207.

Gilchrist, C.L.M., and Chooi, Y.-H. (2021) clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* **37**: 2473-2475.

Gomez-Escribano, J.P., Song, L., Fox, D.J., Yeo, V., Bibb, M.J., and Challis, G.L. (2012) Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P_1, a metabolic product of the cpk gene cluster of *Streptomyces coelicolor* M145. *Chem Sci* **3**: 2716-2720.

Gonzalez-Ruiz, A., Seaton, R.A., and Hamed, K. (2016) Daptomycin: an evidence-based review of its role in the treatment of Gram-positive infections. *Infect Drug Resist* **15**: 47-58.

Grady, E.N., MacDonald, J., Liu, L., Richman, A., and Yuan, Z.-C. (2016) Current knowledge and perspectives of Paenibacillus: a review. *Microb cell fact* **15**: 203.

Grein, F., Müller, A., Scherer, K.M., Liu, X., Ludwig, K.C., Klöckner, A., *et al.* (2020) Ca2+-Daptomycin targets cell wall biosynthesis by forming a tripartite complex with undecaprenyl-coupled intermediates and membrane lipids. *Nat Commun* **11**: 1455.

Grienke, U., Foster, P.A., Zwirchmayr, J., Tahir, A., Rollinger, J.M., and Mikros, E. (2019) ^{1}H NMR-MS-based heterocovariance as a drug discovery tool for fishing bioactive compounds out of a complex mixture of structural analogues. *Sci Rep* **9**: 11113.

Griffith, R.S. (1981) Introduction to vancomycin. *Rev Infect Dis* **3**: S200-S204.

Gurevich, A., Mikheenko, A., Shlemov, A., Korobeynikov, A., Mohimani, H., and Pevzner, P.A. (2018) Increased diversity of peptidic natural products revealed by modification-tolerant database search of mass spectra. *Nat Microbiol* **3**: 319-327.

Ham, Y., and Kim, T.J. (2018) Anthranilamide from *Streptomyces* spp. inhibited *Xanthomonas oryzae* biofilm formation without affecting cell growth. *Appl Biol Chem* **61**: 673-680.

Hameed, A., Al-Rashida, M., Uroos, M., Ali, S.A., Arshia, Ishtiaq, M., and Khan, K.M. (2018) Quinazoline and quinazolinone as important medicinal scaffolds: a comparative patent review (2011-2016). *Expert Opin Ther Pat* **28**: 281-297.

Hancock, R.E.W., and Chapple, D.E. (1999) Peptide antibiotics. *Antimicrob Agents Chemother* **43**: 1317-1323.

Hannigan, G.D., Prihoda, D., Palicka, A., Soukup, J., Klempir, O., Rampula, L., *et al.* (2019) A deep learning genome-mining strategy for biosynthetic gene cluster prediction. *Nucleic Acids Res* **47**: e110-e110.

Haque, S.F., Sen, S.K., and Pal, S.C. (1995) Nutrient optimization for production of broad spectrum antibiotic by *Streptomyces antibioticus* SR15.4. *Acta Microbiol Immunol Hung* **42**: 155-162.

He, D., Wang, M., Zhao, S., Shu, Y., Zeng, H., Xiao, C., *et al.* (2017) Pharmaceutical prospects of naturally occurring quinazolinone and its derivatives. *Fitoterapia* **119**: 136–149.

Henkel, T., Rohr, J., Beale, J.M., and Schwenen, L. (1990) Landomycins, new angucycline antibiotics from *Streptomyces* sp. I. Structural studies on landomycins A-D. *J Antibiot* **43**: 492-503.

Hertlein, G., Seiffert, M., Gensel, S., Garcia-Gonzalez, E., Ebeling, J., Skobalj, R., *et al.* (2016) Biological role of paenilarvins, iturin-like lipopeptide secondary metabolites produced by the honey bee pathogen *Paenibacillus larvae*. *PLoS One* **11**: e0164656.

Hoshino, S., Wakimoto, T., Onaka, H., and Abe, I. (2015) Chojalactones A–C, cytotoxic butanolides isolated from *Streptomyces* sp. cultivated with mycolic acid containing bacterium. *Org Lett* **17**: 1501-1504.

Hoskisson, P.A., and Fernández-Martínez, L.T. (2018) Regulation of specialised metabolites in Actinobacteria - expanding the paradigms. *Environ Microbiol Rep* **10**: 231-238.

Huang, J., Shi, J., Molle, V., Sohlberg, B., Weaver, D., Bibb, M.J., *et al.* (2005) Cross-regulation among disparate antibiotic biosynthetic pathways of *Streptomyces coelicolor*. *Molecular microbiology* **58**: 1276-1287.

Huang, X., Zhang, W., Tang, S., Wei, S., and Lu, X. (2020) Collaborative biosynthesis of a class of bioactive azaphilones by two separate gene clusters containing four PKS/NRPSs with transcriptional crosstalk in fungi. *Angew Chem Int Ed Engl* **59**: 4349-4353.

Hubert, J., Nuzillard, J.-M., and Renault, J.-H. (2017) Dereplication strategies in natural product research: How many tools and methodologies behind the same concept? *Phytochem Rev* **16**: 55-95.

Hutchings, M.I., Truman, A.W., and Wilkinson, B. (2019) Antibiotics: past, present and future. *Curr Opin Microbiol* **51**: 72-80.

Ikai, Y., Fau, O.H., Fau, H.J., Fau, K.N., Fau, M.T., Fau, S.M., and Harada, K. (1998) Total structures of colistin minor components. *J Antibiot* **51**: 492-498.

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., *et al.* (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nat Biotechnol* **21**: 526-531.

Jafari, E., Khajouei, M.R., Hassanzadeh, F., Hakimelahi, G.H., and Khodarahmi, G.A. (2016) Quinazolinone and quinazoline derivatives: recent structures with potent antimicrobial and cytotoxic activities. *Res Pharm Sci* **11**: 1-14.

Jakimowicz, D., and van Wezel, G.P. (2012) Cell division and DNA segregation in *Streptomyces*: how to build a septum in the middle of nowhere? *Molecular microbiology* **85**: 393-404.

Jangra, M., Kaur, M., Podia, M., Tambat, R., Singh, V., Chandal, N., *et al.* (2019a) Purification and biological activity of natural variants synthesized by tridecaptin M gene cluster and in vitro drug-kinetics of this antibiotic class. *Sci Rep* **9**: 18870.

Jangra, M., Kaur, M., Tambat, R., Rana, R., Maurya Sushil, K., Khatri, N., *et al.* (2019b) Tridecaptin M, a new variant discovered in mud bacterium, shows activity against

colistin- and extremely drug-resistant *Enterobacteriaceae*. *Antimicrob Agents Chemother* **63**: e00338-00319.

Jangra, M., Raka, V., and Nandanwar, H. (2020) *In vitro* evaluation of antimicrobial peptide tridecaptin M in combination with other antibiotics against multidrug resistant *Acinetobacter baumannii*. *Molecules* **25**: 3255.

Jarmusch, A., Aron, A., Petras, D., Phelan, V., Bittremieux, W., Acharya, D., *et al.*, (2022) A universal language for finding mass spectrometry data patterns. In.: bioRxiv, pp.

Jenke-Kodama, H., and Dittmann, E. (2009) Bioinformatic perspectives on NRPS/PKS megasynthases: Advances and challenges. *Nat Prod Rep* **26**: 874-883.

Jian Li, Roger L Nation, John D Turnidge, Robert W Milne, Kingsley Coulthard, Craig R Rayner, and Paterson, D.L. (2006) Colistin: the re-emerging antibiotic for multidrug-resistant Gram-negative bacterial infections. *Lancet Infect Dis* **6**: 589-601.

Johnson, B.A., Anker, H., and Meleney, F.L. (1945) Bacitracin: a new antibiotic produced by a member of the *B. subtilis g*roup. *Science* **102**: 376-377.

Jones, G.H. (1987) Actinomycin synthesis in *Streptomyces antibioticus*: enzymatic conversion of 3-hydroxyanthranilic acid to 4-methyl-3-hydroxyanthranilic acid. *J Bacteriol* **169**: 5575-5578.

Kajimura, Y., and Kaneda, M. (1996) Fusaricidin A, a new depsipeptide antibiotic produced by *Bacillus polymyxa* KT-8. Taxonomy, fermentation, isolation, structure elucidation and biological activity. *J Antibiot* **49**: 129-135.

Karak, M., Cloonan, C.R., Baker, B.R., Cochrane, R.V.K., and Cochrane, S.A. (2024) Optimizations of lipid II synthesis: an essential glycolipid precursor in bacterial cell wall synthesis and a validated antibiotic target. *Beilstein J Org Chem* **20**: 220-227.

Katz, E., (1967) Actinomycin. In: Biosynthesis. D. Gottlieb & P.D. Shaw (eds). Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 276-341.

Katz, E., Pienta, P., and Sivak, A. (1957) The role of nutrition in the synthesis of actinomycin. *Appl Microbiol* **6**: 236–241.

Katz, L., and Baltz, R.H. (2016) Natural product discovery: past, present, and future. *J Ind Microbiol Biotechnol* **43**: 155-176.

Kautsar, S.A., Blin, K., Shaw, S., Weber, T., and Medema, M.H. (2021a) BiG-FAM: the biosynthetic gene cluster families database. *Nucleic Acids Res* **49**: D490-D497.

Kautsar, S.A., van der Hooft, J.J.J., de Ridder, D., and Medema, M.H. (2021b) BiG-SLiCE: A highly scalable tool maps the diversity of 1.2 million biosynthetic gene clusters. *Gigascience* **10**: giaa154.

Keller, U. (1984) Acyl pentapeptide lactone synthesis in actinomycin-producing streptomycetes by feeding with structural analogs of 4-methyl-3-hydroxyanthranilic acid. *J Biol Chem* **259**: 8226-8231.

Kharel, M.K., Pahari, P., Shepherd, M.D., Tibrewal, N., Nybo, S.E., Shaaban, K.A., and Rohr, J. (2012) Angucyclines: biosynthesis, mode-of-action, new natural products, and synthesis. *Nat Prod Rep* **29**: 264-325.

Kieser, T., Bibb, M. J., Buttner, M. J., Chater, K. F. & Hopwood, D. A, (2000) *Practical Streptomyces genetics*. John Innes Foundation.

Kloosterman, A.M., Cimermancic, P., Elsayed, S.S., Du, C., Hadjithomas, M., Donia, M.S., *et al.* (2020a) Expansion of RiPP biosynthetic space through integration of pan-genomics and machine learning uncovers a novel class of lanthipeptides. *PLoS Biol* **18**: e3001026.

Kloosterman, A.M., Shelton, K.E., van Wezel, G.P., Medema, M.H., and Mitchell, D.A. (2020b) RRE-Finder: a genome-mining tool for class-independent RiPP discovery. *mSystems* **5**: e00267-00220.

Knolhoff, A.M., Zheng, J., McFarland, M.A., Luo, Y., Callahan, J.H., Brown, E.W., and Croley, T.R. (2015) Identification and structural characterization of naturally-occurring broad-spectrum cyclic antibiotics isolated from *Paenibacillus*. *J Am Soc Mass Spectrom* **26**: 1768-1779.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019) Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37**: 540-546.

Kolter, R., and van Wezel, G.P. (2016) Goodbye to brute force in antibiotic discovery? *Nat Microbiol* **1**: 15020.

Kornsakulkarn, J., Saepua, S., Srijomthong, K., Rachtawee, P., and Thongpanchang, C. (2015) Quinazolinone alkaloids from actinomycete *Streptomyces* sp. BCC 21795. *Phytochem Lett* **12**: 6–8.

Korynevska, A., Heffeter, P., Matselyukh, B., Elbling, L., Micksche, M., Stoika, R., and Berger, W. (2007) Mechanisms underlying the anticancer activities of the angucycline landomycin E. *Biochem Pharmacol* **74**: 1713-1726.

Kotsogianni, I., Wood, T.M., Alexander, F.M., Cochrane, S.A., and Martin, N.I. (2021) Binding studies reveal phospholipid specificity and its role in the Calcium-dependent mechanism of action of daptomycin. *ACS Infect Dis* **7**: 2612-2619.

Krohn, K., and Baltus, W. (1988) Synthesis of rac- and ent-fridamycin E. *Tetrahedron* **44**: 49-54.

Krug, D., and Müller, R. (2014) Secondary metabolomics: the impact of mass spectrometry-based approaches on the discovery and characterization of microbial natural products. *Nat Prod Rep* **31**: 768-783.

Kshirsagar, U.A. (2015) Recent developments in the chemistry of quinazolinone alkaloids. *Org Biomol Chem* **13**: 9336-9352.

Laatsch, H., (2011) AntiBase 2011: the natural compound identifier. In.: Wiley-Vch, pp.

Laatsch, H., (2012a) Antibase 2012, the Natural Compound Identifier. In. Weinheim, Germany: Wiley-VCH pp.

Laatsch, H., (2012b) Antibase Version 4.0 - The Natural Compound Identifier. In. New York: Wiley-VCH Verlag GmbH & Co., pp.

Lai, Z., Yu, J., Ling, H., Song, Y., Yuan, J., Ju, J., *et al.* (2018) Grincamycins I - K, cytotoxic angucycline glycosides derived from marine-derived Actinomycete *Streptomyces lusitanus* SCSIO LR32. *Planta Med* **84**: 201-207.

Lane, D., (1991) *16S/23S rRNA sequencing,* p. 115–175. Wiley.

Letunic, I., and Bork, P. (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**: W293-W296.

Lewis, K. (2012) Recover the lost art of drug discovery. *Nature* **485**: 439-440.

Lewis, K. (2013) Platforms for antibiotic discovery. *Nat Rev Drug Discov* **12**: 371-387.

Li, Y.-X., Zhong, Z., Zhang, W.-P., and Qian, P.-Y. (2018) Discovery of cationic nonribosomal peptides as Gram-negative antibiotics through global genome mining. *Nat Commun* **9**: 3273.

Li, Z., Chakraborty, P., de Vries, R.H., Song, C., Zhao, X., Roelfes, G., *et al.* (2020) Characterization of two relacidines belonging to a novel class of circular lipopeptides that act against Gram-negative bacterial pathogens. *Environ Microbiol* **22**: 5125-5136.

Liu, M., and Douthwaite, S. (2002) Resistance to the macrolide antibiotic tylosin is conferred by single methylations at 23S rRNA nucleotides G748 and A2058 acting in synergy. *Proc Natl Acad Sci U S A* **99**: 14658-14663.

Liu, M., Grkovic, T., Liu, X., Han, J., Zhang, L., and Quinn, R.J. (2017) A systems approach using OSMAC, Log P and NMR fingerprinting: An approach to novelty. *Synth Syst Biotechnol* **2**: 276-286.

Liu, Y., Ding, S., Shen, J., and Zhu, K. (2019) Nonribosomal antibacterial peptides that target multidrug-resistant bacteria. *Nat Prod Rep* **36**: 573-592.

Lloyd Karen, G., Steen Andrew, D., Ladau, J., Yin, J., and Crosby, L. (2018) Phylogenetically novel uncultured microbial cells dominate Earth microbiomes. *mSystems* **3**: e00055-00018.

Lohans, C.T., Huang, Z., van Belkum, M.J., Giroud, M., Sit, C.S., Steels, E.M., *et al.* (2012) Structural characterization of the highly cyclized lantibiotic paenicidin A via a partial desulfurization/reduction strategy. *J Am Chem Soc* **134**: 19540-19543.

Lohans, C.T., van Belkum, M.J., Cochrane, S.A., Huang, Z., Sit, C.S., McMullen, L.M., and Vederas, J.C. (2014) Biochemical, structural, and genetic characterization of tridecaptin A_1, an antagonist of *Campylobacter jejuni*. *ChemBioChem* **15**: 243-249.

Lommen, A. (2009) MetAlign: interface-driven, versatile metabolomics tool for hyphenated full-scan mass spectrometry data preprocessing. *Anal Chem* **81**: 3079-3086.

Luo, Y., Cobb, R.E., and Zhao, H. (2014) Recent advances in natural product discovery. *Curr Opin Biotechnol* **30**: 230-237.

Ma, Y., Kong, Q., Qin, C., Chen, Y., Chen, Y., Lv, R., and Zhou, G. (2016) Identification of lipopeptides in *Bacillus megaterium* by two-step ultrafiltration and LC-ESI-MS/MS. *AMB Express* **6**: 79.

Machushynets, N.V., Elsayed, S.S., Du, C., Siegler, M.A., de la Cruz, M., Genilloud, O., *et al.* (2022) Discovery of actinomycin L, a new member of the actinomycin family of antibiotics. *Sci Rep* **12**: 2813.

Machushynets, N.V., Wu, C., Elsayed, S.S., Hankemeier, T., and van Wezel, G.P. (2019) Discovery of novel glycerolated quinazolinones from *Streptomyces* sp. MBT27. *J Ind Microbiol Biotechnol* **46**: 483-492.

Macrae, C.F., Sovago, I., Cottrell, S.J., Galek, P.T.A., McCabe, P., Pidcock, E., *et al.* (2020) Mercury 4.0: from visualization to analysis, design and prediction. *J Appl Crystallogr* **53**: 226-235.

Männle, D., McKinnie Shaun, M.K., Mantri Shrikant, S., Steinke, K., Lu, Z., Moore Bradley, S., *et al.* (2020) Comparative genomics and metabolomics in the genus nocardia. *mSystems* **5**.

Marahiel, M.A. (2009) Working outside the protein-synthesis rules: insights into non-ribosomal peptide synthesis. *J Pept Sci* **15**: 799-807.

Marahiel, M.A. (2016) A structural model for multimodular NRPS assembly lines. *Nat Prod Rep* **33**: 136-140.

Markley, J.L., Brüschweiler, R., Edison, A.S., Eghbalnia, H.R., Powers, R., Raftery, D., and Wishart, D.S. (2017) The future of NMR-based metabolomics. *Curr Opin Biotechnol* **43**: 34-40.

Martin, J.F., Sola-Landa, A., Santos-Beneit, F., Fernandez-Martinez, L.T., Prieto, C., and Rodriguez-Garcia, A. (2011) Cross-talk of global nutritional regulators in the control of primary and secondary metabolism in *Streptomyces*. *Microb Biotechnol* **4**: 165-174.

Martin, M.F., and Liras, P. (1989) Organization and expression of genes involved in the biosynthesis of antibiotics and other secondary metabolites. *Annu Rev Microbiol* **43**: 173-206.

Martin, N.I., Hu, H., Moake, M.M., Churey, J.J., Whittal, R., Worobo, R.W., and Vederas, J.C. (2003) Isolation, structural characterization, and properties of mattacin (polymyxin M), a cyclic peptide antibiotic produced by *Paenibacillus kobensis* M. *J Biol Chem* **278**: 13124-13132.

Maskey, R.P., Helmke, E., and Laatsch, H. (2003) Himalomycin A and B: isolation and structure elucidation of new fridamycin type antibiotics from a marine *Streptomyces* isolate. *J Antibiot* **56**: 942-949.

McErlean, M., Overbay, J., and Van Lanen, S. (2019) Refining and expanding nonribosomal peptide synthetase function and mechanism. *J Ind Microbiol Biotechnol* **46**: 493-513.

Medeiros-Silva, J., Jekhmane, S., Breukink, E., and Weingarth, M. (2019) Towards the native binding modes of antibiotics that target Lipid II. *ChemBioChem* **20**: 1731-1738.

Medema, M.H., Cimermancic, P., Sali, A., Takano, E., and Fischbach, M.A. (2014) A systematic computational analysis of biosynthetic gene cluster evolution: lessons for engineering biosynthesis. *PLOS Comput Biol* **10**: e1004016.

Medema, M.H., de Rond, T., and Moore, B.S. (2021) Mining genomes to illuminate the specialized chemistry of life. *Nat Rev Genet* **22**: 553-571.

Medema, M.H., and Fischbach, M.A. (2015) Computational approaches to natural product discovery. *Nat Chem Biol* **11**: 639-648.

Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K., *et al.* (2015) Minimum Information about a Biosynthetic Gene cluster. *Nat Chem Biol* **11**: 625-631.

Merwin, N.J., Mousa, W.K., Dejong, C.A., Skinnider, M.A., Cannon, M.J., Li, H., *et al.* (2020) DeepRiPP integrates multiomics data to automate discovery of novel ribosomally synthesized natural products. *Proc Natl Acad Sci U S A* **117**: 371-380.

Mevers, E., Saurí, J., Helfrich, E.J.N., Henke, M., Barns, K.J., Bugni, T.S., *et al.* (2019) Pyonitrins A–D: chimeric natural products produced by *Pseudomonas protegens*. *J Am Chem Soc* **141**: 17098-17101.

Michael, J.P. (2007) Quinoline, quinazoline and acridone alkaloids. *Nat Prod Rep* **24**: 223–246.

Ming, L.-J., and Epperson, J.D. (2002) Metal binding and structure–activity relationship of the metalloantibiotic peptide bacitracin. *J Inorg Biochem* **91**: 46-58.

Mladic, M., Slagboom, J., Kool, J., Vonk, F.J., van Wezel, G.P., and Richardson, M.K. (2018) Detection and identification of antibacterial proteins in snake venoms using at-line nanofractionation coupled to LC-MS. *Toxicon* **155**: 66-74.

Mladic, M., Zietek, B.M., Iyer, J.K., Hermarij, P., Niessen, W.M.A., Somsen, G.W., *et al.* (2016) At-line nanofractionation with parallel mass spectrometry and bioactivity assessment for the rapid screening of thrombin and factor Xa inhibitors in snake venoms. *Toxicon* **110**: 79-89.

Moffat, A.D., Santos-Aberturas, J., Chandra, G., and Truman, A.W. (2021) A user guide for the identification of new RiPP biosynthetic gene clusters using a RiPPER-based workflow. *Methods Mol Biol* **2296**: 227-247.

Mohamed, A., Nguyen, C.H., and Mamitsuka, H. (2016) Current status and prospects of computational resources for natural product dereplication: a review. *Briefings in Bioinform* **17**: 309-321.

Mohapatra, S.S., Dwibedy, S., and Padhy, I. (2021) Polymyxins, the last-resort antibiotics: mode of action, resistance emergence, and potential solutions. *J Biosci* **46**: 85.

Mohimani, H., Gurevich, A., Mikheenko, A., Garg, N., Nothias, L.-F., Ninomiya, A., *et al.* (2017) Dereplication of peptidic natural products through database search of mass spectra. *Nat Chem Biol* **13**: 30-37.

Mohimani, H., Gurevich, A., Shlemov, A., Mikheenko, A., Korobeynikov, A., Cao, L., *et al.* (2018) Dereplication of microbial metabolites through database search of mass spectra. *Nat Commun* **9**: 4035.

Mongia, M., Baral, R., Adduri, A., Yan, D., Liu, Y., Bian, Y., *et al.* (2023) AdenPredictor: accurate prediction of the adenylation domain specificity of nonribosomal peptide biosynthetic gene clusters in microbial genomes. *Bioinformatics* **39**: i40-i46.

Montalban-Lopez, M., Scott, T.A., Ramesh, S., Rahman, I.R., van Heel, A.J., Viel, J.H., *et al.* (2021) New developments in RiPP discovery, enzymology and engineering. *Nat Prod Rep* **38**: 130-239.

Moon, K., Xu, F., and Seyedsayamdost, M.R. (2019a) Cebulantin, a cryptic lanthipeptide antibiotic uncovered using bioactivity-coupled HiTES. *Angew Chem Int Ed Engl* **58**: 5973-5977.

Moon, K., Xu, F., Zhang, C., and Seyedsayamdost, M.R. (2019b) Bioactivity-HiTES unveils cryptic antibiotics encoded in actinomycete bacteria. *ACS Chem Biol* **14**: 767-774.

Mootz, H.D., Schwarzer, D., and Marahiel, M.A. (2002) Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBioChem* **3**: 490-504.

Mullowney, M.W., Duncan, K.R., Elsayed, S.S., Garg, N., van der Hooft, J.J.J., Martin, N.I., *et al.* (2023) Artificial intelligence for natural product drug discovery. *Nat Rev Drug Discov* **22**: 895-916.

Nang, S.C., Azad, M.A.K., Velkov, T., Zhou, Q., and Li, J. (2021) Rescuing the last-line polymyxins: achievements and challenges. *Pharmacol Rev* **73**: 679-728.

Nanjundan, J., Ramasamy, R., Uthandi, S., and Ponnusamy, M. (2019) Antimicrobial activity and spectroscopic characterization of surfactin class of lipopeptides from *Bacillus amyloliquefaciens* SR1. *Microb Pathog* **128**: 374-380.

Navarro-Muñoz, J.C., Selem-Mojica, N., Mullowney, M.W., Kautsar, S.A., Tryon, J.H., Parkinson, E.I., *et al.* (2020) A computational framework to explore large-scale biosynthetic diversity. *Nat Chem Biol* **16**: 60-68.

Nett, M., Ikeda, H., and Moore, B.S. (2009) Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat Prod Rep* **26**: 1362-1384.

Newman, D.J., and Cragg, G.M. (2007) Natural products as sources of new drugs over the last 25 years. *J Nat Prod* **70**: 461-477.

Newman, D.J., and Cragg, G.M. (2020) Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J Nat Prod* **83**: 770-803.

Nguyen, D.D., Melnik, A.V., Koyama, N., Lu, X., Schorn, M., Fang, J., *et al.* (2016) Indexing the *Pseudomonas* specialized metabolome enabled the discovery of poaeamide B and the bananamides. *Nat Microbiol* **2**: 16197.

Nguyen, D.D., Wu, C.H., Moree, W.J., Lamsa, A., Medema, M.H., Zhao, X., *et al.* (2013) MS/MS networking guided analysis of molecule and gene cluster families. *Proc Natl Acad Sci U S A* **110**: E2611-2620.

Niu, B., Vater, J., Rueckert, C., Blom, J., Lehmann, M., Ru, J.-J., *et al.* (2013) Polymyxin P is the active principle in suppressing phytopathogenic *Erwinia* spp. by the biocontrol rhizobacterium *Paenibacillus polymyxa* M-1. *BMC Microbiol* **13**: 137.

Nivina, A., Yuet, K.P., Hsu, J., and Khosla, C. (2019) Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chem Rev* **119**: 12524-12547.

Nothias, L.-F., Nothias-Esposito, M., da Silva, R., Wang, M., Protsyuk, I., Zhang, Z., *et al.* (2018) Bioactivity-based molecular networking for the discovery of drug leads in natural product bioassay-guided fractionation. *J Nat Prod* **81**: 758-767.

Nothias, L.-F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A., *et al.* (2020) Feature-based molecular networking in the GNPS analysis environment. *Nat Microbiol* **17**: 905-908.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., *et al.* (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**: D733-745.

Olishevska, S., Nickzad, A., and Déziel, E. (2019) *Bacillus* and *Paenibacillus* secreted polyketides and peptides involved in controlling human and plant pathogens. *Appl Microbiol Biotechnol* **103**: 1189-1215.

Olivon, F., Allard, P.-M., Koval, A., Righi, D., Genta-Jouve, G., Neyts, J., *et al.* (2017) Bioactive natural products prioritization using massive multi-informational molecular networks. *ACS Chem Biol* **12**: 2644-2651.

Olivon, F., Retailleau, P., Desrat, S., Touboul, D., Roussi, F., Apel, C., and Litaudon, M. (2020) Isolation of picrotoxanes from austrobuxus carunculatus using taxonomy-based molecular networking. *J Nat Prod* **83**: 3069-3079.

Ory, L., Nazih, E.H., Daoud, S., Mocquard, J., Bourjot, M., Margueritte, L., *et al.* (2019) Targeting bioactive compounds in natural extracts - development of a comprehensive workflow combining chemical and biological data. *Anal Chem* **1070**: 29-42.

Palaniappan, K., Chen, I.M.A., Chu, K., Ratner, A., Seshadri, R., Kyrpides, N.C., *et al.* (2020) IMG-ABC v.5.0: an update to the IMG/Atlas of biosynthetic gene clusters knowledgebase. *Nucleic Acids Res* **48**: D422-D430.

Palazzotto, E., and Weber, T. (2018) Omics and multi-omics approaches to study the biosynthesis of secondary metabolites in microorganisms. *Curr Opin Microbiol* **45**: 109-116.

Pang, Z., Zhou, G., Ewald, J., Chang, L., Hacariz, O., Basu, N., and Xia, J. (2022) Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. *Nat Protoc* **17**: 1735-1761.

Pathak, K.V., and Keharia, H. (2014) Identification of surfactins and iturins produced by potent fungal antagonist, *Bacillus subtilis* K1 isolated from aerial roots of banyan (*Ficus benghalensis*) tree using mass spectrometry. *3 Biotech* **4**: 283-295.

Payne, D.J., Gwynn, M.N., Holmes, D.J., and Pompliano, D.L. (2007) Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat Rev Drug Discov* **6**: 29-40.

Peek, J., Lilic, M., Montiel, D., Milshteyn, A., Woodworth, I., Biggins, J.B., *et al.* (2018) Rifamycin congeners kanglemycins are active against rifampicin-resistant bacteria via a distinct mechanism. *Nat Commun* **9**: 4147.

Pfennig, F., Schauwecker, F., and Keller, U. (1999) Molecular characterization of the genes of actinomycin synthetase I and of a 4-methyl-3-hydroxyanthranilic acid carrier protein involved in the assembly of the acylpeptide chain of actinomycin in *Streptomyces*. *J Biol Chem* **274**: 12508-12516.

Pham, J.V., Yilma, M.A., Feliz, A., Majid, M.T., Maffetone, N., Walker, J.R., *et al.* (2019) A review of the microbial production of bioactive natural products and biologics. *Front Microbiol* **10**: 1404.

Pilz, M., Cavelius, P., Qoura, F., Awad, D., and Brück, T. (2023) Lipopeptides development in cosmetics and pharmaceutical applications: A comprehensive review. *Biotechnol Adv* **67**: 108210.

Pittenauer, E., Zehl, M., Belgacem, O., Raptakis, E., Mistrik, R., and Allmaier, G. (2006) Comparison of CID spectra of singly charged polypeptide antibiotic precursor ions obtained by positive-ion vacuum MALDI IT/RTOF and TOF/RTOF, AP-MALDI-IT and ESI-IT mass spectrometry. *J Mass Spectrom* **41**: 421-447.

Pluskal, T., Castillo, S., Villar-Briones, A., and Oresic, M. (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**: 395.

Purves, K., Macintyre, L., Brennan, D., Hreggviðsson, G.Ó., Kuttner, E., Ásgeirsdóttir, M.E., *et al.* (2016) Using molecular networking for microbial secondary metabolite bioprospecting. *Metabolites* **6**: 2.

Qin, F., Zhang, H., Liu, H., and Zhang, X. (2020) High performance liquid chromatography-quadrupole/Time of Flight–tandem mass spectrometry for the characterization of components in bacitracin. *Chromatographia* **83**.

Quinn, R.A., Nothias, L.-F., Vining, O., Meehan, M., Esquenazi, E., and Dorrestein, P.C. (2017) Molecular networking as a drug discovery, drug metabolism, and precision medicine strategy. *Trends Pharmacol Sci* **38**: 143-154.

Raaijmakers, J.M., De Bruijn, I., Nybroe, O., and Ongena, M. (2010) Natural functions of lipopeptides from *Bacillus* and *Pseudomonas*: more than surfactants and antibiotics. *FEMS Microbiol Rev* **34**: 1037-1062.

Ray, L., and Moore, B.S. (2016) Recent advances in the biosynthesis of unusual polyketide synthase substrates. *Nat Prod Rep* **33**: 150-161.

Rigali, S., Titgemeyer, F., Barends, S., Mulder, S., Thomae, A.W., Hopwood, D.A., and van Wezel, G.P. (2008) Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. *EMBO Rep* **9**: 670-675.

Rix, U., Wang, C., Chen, Y., Lipata, F.M., Remsing Rix, L.L., Greenwell, L.M., *et al.* (2005) The oxidative ring cleavage in jadomycin biosynthesis: a multistep oxygenation cascade in a biosynthetic black box. *ChemBiochem* **6**: 838-845.

Rohr, J., Schoenewolf, M., Udvarnoki, G., Eckardt, K., Schumann, G., Wagner, C., *et al.* (1993) Investigations on the biosynthesis of the angucycline group antibiotics aquayamycin and the urdamycins A and B. Results from the structural analysis of novel blocked mutant products. *J Org Chem* **58**: 2547-2551.

Rohr, J., and Zeeck, A. (1987) Metabolic products of microorganisms. 240. Urdamycins, new angucycline antibiotics from *Streptomyces fradiae*. II. Structural studies of urdamycins B to F. *J Antibiot* **40**: 459-467.

Romano, S., Jackson, S.A., Patry, S., and Dobson, A.D.W. (2018) Extending the "One Strain Many Compounds" (OSMAC) principle to marine microorganisms. *Mar Drugs* **16**: 244.

Röttig, M., Medema, M.H., Blin, K., Weber, T., Rausch, C., and Kohlbacher, O. (2011) NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Res* **39**: W362-W367.

Roux, A., Lison, D., Junot, C., and Heilier, J.-F. (2011) Applications of liquid chromatography coupled to mass spectrometry-based metabolomics in clinical chemistry and toxicology: A review. *Clin. Biochem.* **44**: 119-135.

Rutledge, P.J., and Challis, G.L. (2015) Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol* **13**: 509-523.

Ryu, J., Kim, J.M., Lee, C.W., and Kim, S.W. (2017) Structural analysis and enhanced production of fusaricidin from *Paenibacillus kribbensis* CU01 isolated from yellow loess. *J Basic Microbiol* **57**: 525-535.

Sanchez, S., Chavez, A., Forero, A., Garcia-Huante, Y., Romero, A., Sanchez, M., *et al.* (2010) Carbon source regulation of antibiotic production. *J Antibiot* **63**: 442-459.

Sanchez, S., Rogers, J.D., Rogers, A.B., Nassar, M., McEntyre, J., Welch, M., *et al.* (2023) Expansion of novel biosynthetic gene clusters from diverse environments using SanntiS. *bioRxiv*: 2023.2005.2023.540769.

Sayed, A.M., Hassan, M.H.A., Alhadrami, H.A., Hassan, H.M., Goodfellow, M., and Rateb, M.E. (2020) Extreme environments: microbiology leading to specialized metabolites. *J Appl Microbiol* **128**: 630-657.

Scherlach, K., and Hertweck, C. (2021) Mining and unearthing hidden biosynthetic potential. *Nat Commun* **12**: 3864.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nat Microbiol* **9**: 676-682.

Schroeckh, V., Scherlach, K., Nutzmann, H.W., Shelest, E., Schmidt-Heck, W., Schuemann, J., *et al.* (2009) Intimate bacterial-fungal interaction triggers biosynthesis of archetypal polyketides in *Aspergillus nidulans*. *Proc Natl Acad Sci U S A* **106**: 14558-14563.

Schulz, S., and Dickschat, J.S. (2007) Bacterial volatiles: the smell of small organisms. *Nat Prod Rep* **24**: 814-842.

Schulze, C.J., Bray, W.M., Woerhmann, M.H., Stuart, J., Lokey, R.S., and Linington, R.G. (2013) "Function-first" lead discovery: mode of action profiling of natural product libraries using image-based screening. *Chem Biol* **20**: 285–295.

Seipke, R.F., Kaltenpoth, M., and Hutchings, M.I. (2012) Streptomyces as symbionts: an emerging and widespread theme? *FEMS Microbiol Rev* **36**: 862-876.

Sélem-Mojica, N., Aguilar, C., Gutiérrez-García, K., Martínez-Guerrero, C.E., and Barona-Gómez, F. (2019) EvoMining reveals the origin and fate of natural product biosynthetic enzymes. *Microb Genom* **5**.

Semsary, S., Crnovčić, I., Driller, R., Vater, J., Loll, B., and Keller, U. (2018) Ketonization of proline residues in the peptide chains of actinomycins by a 4-oxoproline synthase. *ChemBioChem* **19**: 706-715.

Seyedsayamdost, M.R. (2014) High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proc Natl Acad Sci U S A* **111**: 7266-7271.

Sezaki, M., Hara, T., Ayukawa, S., Takeuchi, T., and Okami, Y. (1968) Studies on a new antibiotic pigment, aquayamycin. *J Antibiot* **21**: 91-97.

Shaaban, K.A., Ahmed, T.A., Leggas, M., and Rohr, J. (2012a) Saquayamycins G–K, cytotoxic angucyclines from *Streptomyces* sp. including two analogues bearing the aminosugar rednose. *J Nat Prod* **75**: 1383-1392.

Shaaban, K.A., Shepherd, M.D., Ahmed, T.A., Nybo, S.E., Leggas, M., and Rohr, J. (2012b) Pyramidamycins A-D and 3-hydroxyquinoline-2-carboxamide; cytotoxic benzamides from *Streptomyces* sp. DGC1. *J Antibiot* **65**: 615–622.

Shaaban, K.A., Srinivasan, S., Kumar, R., Damodaran, C., and Rohr, J. (2011) Landomycins P-W, cytotoxic angucyclines from *Streptomyces cyanogenus* S-136. *J Nat Prod* **74**: 2-11.

Shannon, P., Markiel, A., Ozier, O., Baliga, N., Wang, J., Ramage, D., *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**: 2498-2504.

Sheldrick, G.M. (2015) SHELXT - integrated space-group and crystal-structure determination. *Acta Crystallogr* **71**: 3–8.

Shen, B. (2003) Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Curr Opin Chem Biol* **7**: 285-295.

Shoji, J., Hinoo, H., Sakazaki, R., Kato, T., Akisaka, Y., Mayama, M., *et al.* (1978) Isolation of tridecaptins A, B and C (studies on antibiotics from the genus *Bacillus*. XXIII). *J Antibiot* **31**: 646-651.

Shoji J., Kato T., Terabe S., and R., K. (1979) Resolution of peptide antibiotics, cerexins and tridecaptins, by high performance lliquid chromatography. *J Antibiot* **32** 313-319.

Sieber, S.A., and Marahiel, M.A. (2005) Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem Rev* **105**: 715-738.

Sievers, F., and Higgins, D.G. (2018) Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci* **27**: 135-145.

Siewert, G., and Strominger, J.L. (1967) Bacitracin: an inhibitor of the dephosphorylation of lipid pyrophosphate, an intermediate in the biosynthesis of the peptidoglycan of bacterial cell walls. *Proc Natl Acad Sci U S A* **57**.

Silver, L.L. (2011) Challenges of antibacterial discovery. *Clin Microbiol Rev* **24**: 71-109.

Skinnider, M.A., Johnston, C.W., Gunabalasingam, M., Merwin, N.J., Kieliszek, A.M., MacLellan, R.J., *et al.* (2020) Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nat Commun* **11**: 6058.

Slingerland, C.J., and Martin, N.I. (2024) Recent advances in the development of polymyxin antibiotics: 2010–2023. *ACS Infect Dis* **10**: 1056–1079.

Smanski, M.J., Zhou, H., Claesen, J., Shen, B., Fischbach, M.A., and Voigt, C.A. (2016) Synthetic biology to access and expand nature's chemical diversity. *Nat Rev Microbiol* **14**: 135-149.

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **78**: 779-787.

Sood, S., Steinmetz, H., Beims, H., Mohr, K.I., Stadler, M., Djukic, M., *et al.* (2014) Paenilarvins: Iturin family lipopeptides from the honey bee pathogen *Paenibacillus larvae*. *ChemBiochem* **15**: 1947-1955.

Spek, A.L. (2009) Structure validation in chemical crystallography. *Acta Crystallogr* **65**: 148-155.

Spraker, J.E., Luu, G.T., and Sanchez, L.M. (2020) Imaging mass spectrometry for natural products discovery: a review of ionization methods. *Nat Prod Rep* **37**: 150-162.

Stone, K.F., and Strominger, J.L. (1971) Mechanism of action of bacitracin: complexation with metal ion and C_{55}-isoprenyl pyrophosphate. *Proc Natl Acad Sci U S A* **68**.

Straus, S.K., and Hancock, R.E.W. (2006) Mode of action of the new antibiotic for Gram-positive pathogens daptomycin: Comparison with cationic antimicrobial peptides and lipopeptides. *Biochim Biophys Acta* **1758**: 1215-1223.

Strohalm, M., Kavan, D., Novak, P., Volny, M., and Havlicek, V. (2010) mMass 3: a cross-platform software environment for precise analysis of mass spectrometric data. *Anal Chem* **82**: 4648-4651.

Sugiyama, R., Nishimura, S., Ozaki, T., Asamizu, S., Onaka, H., and Kakeya, H. (2015) 5-Alkyl-1,2,3,4-tetrahydroquinolines, new membrane-interacting lipophilic metabolites produced by combined culture of *Streptomyces nigrescens* and *Tsukamurella pulmonis*. *Org Lett* **17**: 1918-1921.

Suleiman, S.A., Song, F., Su, M., Hang, T., and Song, M. (2017) Analysis of bacitracin and its related substances by liquid chromatography tandem mass spectrometry. *J Pharm Anal* **7**: 48-55.

Süssmuth, R.D., and Mainz, A. (2017) Nonribosomal peptide synthesis - principles and prospects. *Angew Chem Int Ed Engl* **56**: 3770-3821.

Swiatek-Polatynska, M.A., Bucca, G., Laing, E., Gubbens, J., Titgemeyer, F., Smith, C.P., *et al.* (2015) Genome-wide analysis of *In vivo* binding of the master regulator DasR in *Streptomyces coelicolor* identifies novel non-canonical targets. *PLoS One* **10**: e0122479.

Takano, E. (2006) Gamma-butyrolactones: *Streptomyces* signalling molecules regulating antibiotic production and differentiation. *Curr Opin Microbiol* **9**: 287-294.

Tenconi, E., Traxler, M.F., Hoebreck, C., van Wezel, G.P., and Rigali, S. (2018) Production of prodiginines is part of a programmed cell death process in *Streptomyces coelicolor*. *Front Microbiol* **9**: 1742.

Terlouw, B.R., Blin, K., Navarro-Muñoz, J.C., Avalon, N.E., Chevrette, M.G., Egbert, S., *et al.* (2023) MIBiG 3.0: a community-driven effort to annotate experimentally validated biosynthetic gene clusters. *Nucleic Acids Res* **51**: D603-D610.

Theobald, U., Schimana, J., and Fiedler, H. (2000) Microbial growth and production kinetics of *Streptomyces antibioticus* Tü 6040. *Antonie van Leeuwenhoek* **78**: 307–313.

Tietz, J.I., Schwalen, C.J., Patel, P.S., Maxson, T., Blair, P.M., Tai, H.-C., *et al.* (2017) A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nat Chem Biol* **13**: 470-478.

Tripathi, A., Vázquez-Baeza, Y., Gauglitz, J.M., Wang, M., Dührkop, K., Nothias-Esposito, M., *et al.* (2021) Chemically informed analyses of metabolomics mass spectrometry data with Qemistree. *Nat Chem Biol* **17**: 146-151.

Uchida, T., Imoto, M., Watanabe, Y., Miura, K., Dobashi, T., Matsuda, N., *et al.* (1985) Saquayamycins, new aquayamycin-group antibiotics. *J Antibiot* **38**: 1171-1181.

Udwary, D.W., Otani, H., and Mouncey, N.J. (2021) New keys to unlock the treasure trove of microbial natural products. *Nat Rev Microbiol* **19**: 683-683.

Uiterweerd, M.T., Santiago, I.N., van der Heul, H., van Wezel, G.P., and Minnaard, A.J. (2020) Iso-maleimycin, a constitutional isomer of maleimycin, from *Streptomyces* sp. QL37. *Eur J Org Chem* **2020**: 5145-5152.

Urem, M., Swiatek-Polatynska, M.A., Rigali, S., and van Wezel, G.P. (2016) Intertwining nutrient-sensory networks and the control of antibiotic production in *Streptomyces*. *Molecular microbiology* **102**: 183-195.

Vallet, M., Vanbellingen, Q.P., Fu, T., Le Caer, J.-P., Della-Negra, S., Touboul, D., *et al.* (2017) An integrative approach to decipher the chemical antagonism between the competing endophytes *Paraconiothyrium* variabile and *Bacillus subtilis*. *J Nat Prod* **80**: 2863-2873.

van Bergeijk, D.A., Elsayed, S.S., Du, C., Santiago, I.N., Roseboom, A.M., Zhang, L., *et al.* (2022) The ubiquitous catechol moiety elicits siderophore and angucycline production in *Streptomyces*. *Commun Chem* **5**.

van Bergeijk, D.A., Terlouw, B.R., Medema, M.H., and van Wezel, G.P. (2020) Ecology and genomics of Actinobacteria: new concepts for natural product discovery. *Nat Rev Microbiol* **18**: 546-558.

van der Heul, H.U., Bilyk, B.L., McDowall, K.J., Seipke, R.F., and van Wezel, G.P. (2018) Regulation of antibiotic production in Actinobacteria: new perspectives from the post-genomic era. *Nat Prod Rep* **35**: 575-604.

van der Hooft, J.J.J., Wandy, J., Barrett, M.P., Burgess, K.E.V., and Rogers, S. (2016) Topic modeling for untargeted substructure exploration in metabolomics. *Proc Natl Acad Sci U S A* **113**: 13738-13743.

van der Hooft, J.J.J., Wandy, J., Young, F., Padmanabhan, S., Gerasimidis, K., Burgess, K.E.V., *et al.* (2017) Unsupervised discovery and comparison of structural families across multiple samples in untargeted metabolomics. *Anal Chem* **89**: 7569-7577.

van der Meij, A., Worsley, S.F., Hutchings, M.I., and van Wezel, G.P. (2017) Chemical ecology of antibiotic production by actinomycetes. *FEMS Microbiol Rev* **41**: 392-416.

van Heel, A.J., de Jong, A., Montalban-Lopez, M., Kok, J., and Kuipers, O.P. (2013) BAGEL3: Automated identification of genes encoding bacteriocins and (non-)bactericidal posttranslationally modified peptides. *Nucleic Acids Res* **41**: W448-453.

van Santen, J.A., Jacob, G., Singh, A.L., Aniebok, V., Balunas, M.J., Bunsko, D., *et al.* (2019) The natural products atlas: an open access knowledge base for microbial natural products discovery. *ACS Cent Sci* **5**: 1824-1833.

van Wezel, G.P., McKenzie, N.L., and Nodwell, J.R., (2009) Applying the genetics of secondary metabolism in model actinomycetes to the discovery of new antibiotics. In.: Academic Press, pp. 117–141.

Vater, J., Crnovcic, I., Semsary, S., and Keller, U. (2014) MALDI-TOF mass spectrometry, an efficient technique for in situ detection and characterization of actinomycins. *J Mass Spectrom* **49**: 210–222.

Vater, J., Herfort, S., Doellinger, J., Weydmann, M., Borriss, R., and Lasch, P. (2018) Genome mining of the lipopeptide biosynthesis of *Paenibacillus polymyxa* E681 in combination with mass spectrometry: discovery of the lipoheptapeptide paenilipoheptin. *ChemBioChem* **19**: 744-753.

Vogler, K., and Studer, R.O. (1966) The chemistry of the polymyxin antibiotics. *Experientia* **22**: 345-354.

Waksman, S.A. (1945) Microbial antagonisms and antibiotic substances. *Arch Intern Med* **76**: 255-256.

Waksman, S.A., and Woodruff, H.B. (1940) Bacteriostatic and bactericidal substances produced by a soil *Actinomyces*. *Proc Soc Exp Biol Med* **45**: 609-614.

Wang, H., Fewer, D.P., Holm, L., Rouhiainen, L., and Sivonen, K. (2014) Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc Natl Acad Sci U S A* **111**: 9259-9264.

Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., *et al.* (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **34**: 828-837.

Wang, Q., Zhang, Y., Wang, M., Tan, Y., Hu, X., He, H., *et al.* (2017) Neo-actinomycins A and B, natural actinomycins bearing the 5H-oxazolo[4,5-b]phenoxazine chromophore, from the marine-derived *Streptomyces* sp. IMB094. *Sci Rep* **7**: 3591.

Wang, Z., Koirala, B., Hernandez, Y., and Brady, S.F. (2022a) Discovery of *Paenibacillaceae* family Gram-negative-active cationic lipopeptide antibiotics using evolution-guided chemical synthesis. *Org Lett* **24**: 4943-4948.

Wang, Z., Koirala, B., Hernandez, Y., Zimmerman, M., and Brady, S.F. (2022b) Bioinformatic prospecting and synthesis of a bifunctional lipopeptide antibiotic that evades resistance. *Science* **376**: 991-996.

Wang, Z., Koirala, B., Hernandez, Y., Zimmerman, M., Park, S., Perlin, D.S., and Brady, S.F. (2022c) A naturally inspired antibiotic to target multidrug-resistant pathogens. *Nature* **601**: 606-611.

Watrous, J., Roach, P., Alexandrov, T., Heath, B.S., Yang, J.Y., Kersten, R.D., *et al.* (2012) Mass spectral molecular networking of living microbial colonies. *Proc Natl Acad Sci U S A* **109**: E1743-E1752.

Weber, T., and Kim, H.U. (2016) The secondary metabolite bioinformatics portal: computational tools to facilitate synthetic biology of secondary metabolite production. *Synth Syst Biotechnol* **1**: 69-79.

Wenski, S.L., Thiengmag, S., and Helfrich, E.J.N. (2022) Complex peptide natural products: Biosynthetic principles, challenges and opportunities for pathway engineering. *Synth Syst Biotechnol* **7**: 631-647.

Wenzel, S.C., and Müller, R. (2005) Formation of novel secondary metabolites by bacterial multimodular assembly lines: deviations from textbook biosynthetic logic. *Curr Opin Microbiol* **9**: 447-458.

Whitford, C.M., Cruz-Morales, P., Keasling, J.D., and Weber, T.A.-O. (2021) The design-build-test-learn cycle for metabolic engineering of *Streptomycetes*. *Essays Biochem* **65**: 1744-1358.

WHO, (2017) Prioritization of pathogens to guide discovery, research and development of new antibiotics for drug-resistant bacterial infections, including tuberculosis.

WHO, (2019) No time to wait: securing the future from drug-resistant infections.

Williams, D.H. (1984) Structural studies on some antibiotics of the vancomycin group, and on the antibiotic-receptor complexes, by proton NMR. *Acc Chem Res* **17**: 364-369.

Williams, D.H., and Butcher, D.W. (1981) Binding site of the antibiotic vancomycin for a cell-wall peptide analog. *J Am Chem Soc* **103**: 5697-5700.

Williams, W.K., and Katz, E. (1977) Development of a chemically defined medium for the synthesis of actinomycin D by *Streptomyces parvulus*. *Antimicrob Agents Chemother* **11**: 281.

Winn, M., Fyans, J.K., Zhuo, Y., and Micklefield, J. (2016) Recent advances in engineering nonribosomal peptide assembly lines. *Natural product reports* **33**: 317-347.

Winnikoff, J.R., Glukhov, E., Watrous, J., Dorrestein, P.C., and Gerwick, W.H. (2014) Quantitative molecular networking to profile marine cyanobacterial metabolomes. *J Antibiot* **67**: 105–112.

Wipat, A., and Harwood, C.R. (1999) The Bacillus subtilis genome sequence: the molecular blueprint of a soil bacterium. *FEMS Microbiol Ecol* **28**: 1-9.

Wollenberg, R.D., Saei, W., Westphal, K.R., Klitgaard, C.S., Nielsen, K.L., Lysoe, E., *et al.* (2017) Chrysogine biosynthesis is mediated by a two-module nonribosomal peptide synthetase. *J Nat Prod* **80**: 2131–2135.

Wright, G.D. (2014) Something old, something new: revisiting natural products in antibiotic drug discovery. *Can J Microbiol* **60**: 147-154.

Wright, G.D. (2015) Solving the antibiotic crisis. *ACS Infect Dis* **1**: 80-84.

Wu, C., Choi, Y.H., and van Wezel, G.P. (2016a) Metabolic profiling as a tool for prioritizing antimicrobial compounds. *J Ind Microbiol Biotechnol* **43**: 299-312.

Wu, C., Ichinose, K., Choi, Y.H., and van Wezel, G.P. (2017) Aromatic polyketide GTRI-02 is a previously unidentified product of the *act* gene cluster in *Streptomyces coelicolor* A3(2). *ChemBiochem* **18**: 1428-1434.

Wu, C., Kim, H.K., van Wezel, G.P., and Choi, Y.H. (2015a) Metabolomics in the natural products field - a gateway to novel antibiotics. *Drug Discov Today Technol* **13**: 11-17.

Wu, C., Medema, M.H., Läkamp, R.M., Zhang, L., Dorrestein, P.C., Choi, Y.H., and van Wezel, G.P. (2016b) Leucanicidin and endophenasides result from methyl-rhamnosylation by the same tailoring enzymes in *Kitasatospora* sp. MBT66. *ACS Chem Biol* **11**: 478-490.

Wu, C., van der Heul, H.U., Melnik, A.V., Lubben, J., Dorrestein, P.C., Minnaard, A.J., *et al.* (2019) Lugdunomycin, an angucycline-derived molecule with unprecedented chemical architecture. *Angew Chem Int Ed Engl* **58**: 2809-2814.

Wu, C., Zacchetti, B., Ram, A.F., van Wezel, G.P., Claessen, D., and Hae Choi, Y. (2015b) Expanding the chemical space for natural products by *Aspergillus-Streptomyces* co-cultivation and biotransformation. *Sci Rep* **5**: 10868.

Xiao, X., Elsayed, S.S., Wu, C., van der Heul, H.U., Metsä-Ketelä, M., Du, C., *et al.* (2020) Functional and structural insights into a novel promiscuous ketoreductase of the lugdunomycin biosynthetic pathway. *ACS Chem Biol* **15**: 2529-2538.

Xu, F., Wu, Y., Zhang, C., Davis, K.M., Moon, K., Bushin, L.B., and Seyedsayamdost, M.R. (2019) A genetics-free method for high-throughput discovery of cryptic microbial metabolites. *Nat Chem Biol* **15**: 161-168.

Yan, D., Chen, Q., Gao, J., Bai, J., Liu, B., Zhang, Y., *et al.* (2019) Complexity and diversity generation in the biosynthesis of fumiquinazoline-related peptidyl alkaloids. *Org Lett* **21**: 1475-1479.

Yang, J.Y., Sanchez, L.M., Rath, C.M., Liu, X., Boudreau, P.D., Bruns, N., *et al.* (2013) Molecular networking as a dereplication strategy. *J Nat Prod* **76**: 1686–1699.

Ye, Y., Li, Q., Fu, G., Yuan, G., Miao, J., and Lin, W. (2012) Identification of antifungal substance (Iturin A2) produced by *Bacillus subtilis* B47 and its effect on southern corn leaf blight. *J Integr Agric* **11**: 90-99.

Yin, S., Liu, Z., Shen, J., Xia, Y., Wang, W., Gui, P., *et al.* (2022) Chimeric natural products derived from medermycin and the nature-inspired construction of their polycyclic skeletons. *Nat Commun* **13**: 5169.

Yoon, V., and Nodwell, J.R. (2014) Activating secondary metabolism with stress and chemicals. *J Ind Microbiol Biotechnol* **41**: 415-424.

Zacchetti, B., Andrianos, A., van Dissel, D., de Ruiter, E., van Wezel, G.P., and Claessen, D. (2018) Microencapsulation extends mycelial viability of *Streptomyces lividans* 66 and increases enzyme production. *BMC Biotechnol* **18**: 13.

Zhang, J.-H., Chung, T.D.Y., and Oldenburg, K.R. (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen* **4**: 67-73.

Zhang, J.J., Tang, X., and Moore, B.S. (2019) Genetic platforms for heterologous expression of microbial natural products. *Nat Prod Rep* **36**: 1313–1332.

Zhang, W., Chang, J., Lei, Z., Huhman, D., Sumner, L.W., and Zhao, P.X. (2014) MET-COFEA: A liquid chromatography/mass spectrometry data processing platform for metabolite compound feature extraction and annotation. *Anal Chem* **86**: 6245-6253.

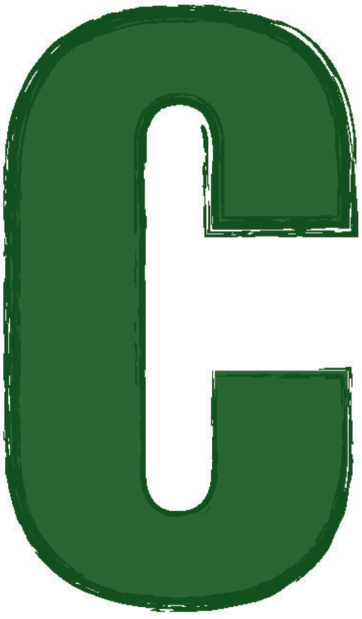
Zhao, M., Lin, H.C., and Tang, Y. (2016) Biosynthesis of the α-nitro-containing cyclic tripeptide psychrophilin. *J Antibiot* **69**: 571-573.

Zheng, C., Xie, T., He, H., and Gao, S. (2021) Asymmetric total synthesis of PD-116740. *Org Lett* **23**: 469-473.

Zhu, H., Sandiford, S.K., and van Wezel, G.P. (2014a) Triggers and cues that activate antibiotic production by actinomycetes. *J Ind Microbiol Biotechnol* **41**: 371-386.

Zhu, H., Swierstra, J., Wu, C., Girard, G., Choi, Y.H., van Wamel, W., *et al.* (2014b) Eliciting antibiotics active against the ESKAPE pathogens in a collection of actinomycetes isolated from mountain soils. *Microbiology* **160**: 1714–1725.

Zietek, B.M., Mladic, M., Bruyneel, B., Niessen, W.M.A., Honing, M., Somsen, G.W., and Kool, J. (2017) Nanofractionation platform with parallel mass spectrometry for identification of CYP1A2 inhibitors in metabolic mixtures. *SLAS Discov* **23**: 283-293.



Curriculum vitae



Nataliia Volodymyrivna Machushynets was born in Litohoscha, Ukraine, on December 15, 1993. She completed her high school education at the Kivertsivska Gymnasium in Kivertsi, Ukraine, graduating with a gold medal in 2010. Afterward, she enrolled in a Bachelor of Biology program at Taras Shevchenko Kyiv National University. During her undergraduate studies, Nataliia interned at the Institute of Molecular Biology and Genetics of the National Academy of Sciences of Ukraine in Kyiv and earned her Bachelor's degree in Biology in 2014.

Driven by her passion for research, Nataliia earned a "Scholarship for talented Ukrainian students" from the TOKU-E company to pursue a Master's in Microbial Biotechnology and Health at Leiden University. During her Master's program, she conducted research internships in the labs of dr. Young Choi and prof. Gilles van Wezel, focusing on the LC-MS based metabolomics and isolation of antibacterial compounds from *Streptomyces,* and gained valuable experience in co-cultivation of Streptomyces and fungi. She obtained her Master's degree in 2017. During her Master's, Nataliia worked as a database specialist at the ResearchAnt Foundation.

In 2017, Nataliia took on the role of a scientific analyst at the Institute of Biology (IBL) at Leiden University for the Syngenopep project. She specialized in de novo peptide sequencing and nanofractionation of scorpion and snake venoms to identify new bioactive peptides.

In 2018, Nataliia started her PhD at the Institute of Biology (IBL) at Leiden University as part of the NACTAR program, which aimed to discover new antibiotics. Under the supervision of prof. van Wezel, she focused on identifying novel antimicrobial leads from *Streptomyces* and *Paenibacillus,* employing LC-MS-based metabolomic and bioinformatic approaches. Her doctoral work is detailed in her thesis.

Currently, she is continuing as a Postdoctoral Fellow under the supervision of prof. van Wezel at Leiden University and prof. Marnix Medema at Wageningen University, focusing on bioactive ribosomally synthesized peptides in the plant and human microbiome.



Publications

Machushynets, N.V., Elsayed, S.S., Du, C., Lysenko, V., de la Cruz, M., Genilloud, O., Martin, N.I., Liles, M.R., and van Wezel, G.P. Phylogeny and antibiotic producing potential of plant-associated *Paenibacillus (In preparation)*.

Nuñez Santiago, I.*, **Machushynets, N.V.***, Mladic, M., van Bergeijk, D.A., Elsayed, S.S., Hankemeier, T., and van Wezel, G.P. (2024) nanoRAPIDS as an analytical pipeline for the discovery of novel bioactive metabolites in complex culture extracts at the nanoscale. *Commun Chem* **7**: 71.

Machushynets, N.V., Al Ayed, K., Terlouw, B.R., Du, C., Buijs, N.P., Willemse, J., Elsayed, S.S., Schill, J., Trebosc, V., Pieren, M., Alexander, F.M., Cochrane, S.A., Liles, M.R., Medema, M.H., Martin, N.I., and van Wezel, G.P. (2024) Discovery and derivatization of tridecaptin antibiotics with altered host specificity and enhanced bioactivity. *ACS Chemical Biology* **19**: 1106-1115.

Machushynets, N.V., Elsayed, S.S., Du, C., Siegler, M.A., de la Cruz, M., Genilloud, O., Hankemeier, T., and van Wezel, G.P. (2022) Discovery of actinomycin L, a new member of the actinomycin family of antibiotics. *Sci Rep* **12**: 2813.

Machushynets, N.V., Wu, C., Elsayed, S.S., Hankemeier, T., and van Wezel, G.P. (2019) Discovery of novel glycerolated quinazolinones from *Streptomyces* sp. MBT27. *J Ind Microbiol Biotechnol* **46**: 483-492.

Al Ayed, K., Zamarbide Losada, D., **Machushynets, N.V.**, Terlouw, B., Elsayed, S.S., Schill, J., Trebosc, V., Pieren, M., Medema, M.H., van Wezel, G.P., and Martin, N.I. (2023) Total synthesis and structure assignment of the relacidine lipopeptide antibiotics and preparation of analogues with enhanced stability. *ACS Infect Dis* **9**: 739-748.

