Comprehensive Mining of Underexplored Microbes as a Potent Driver for Natural Product Discovery

Dissertation

zur Erlangung des Grades

des Doktors der Naturwissenschaften

der Naturwissenschaftlich-Technischen Fakultät

der Universität des Saarlandes.

von

Patrick Alexander Haack

Saarbrücken

2023

Tag des Kolloquiums:	26. April 2024
Dekan:	Prof. Dr.-Ing. Michael Vielhaber
Berichterstatter:	Prof. Dr. Rolf Müller
	Prof. Dr. Andriy Luzhetskyy
Vorsitz:	Prof. Dr. Uli Kazmaier
Akad. Mitarbeiter:	Dr. Demian Dietrich

Diese Arbeit entstand unter der Anleitung von Prof. Dr. Rolf Müller am Helmholtz-Institut für Pharmazeutische Forschung Saarland (HIPS) von November 2016 bis Oktober 2023.

Outside of a dog, a book is man's best friend.

Inside of a dog, it's too dark to read.

Groucho Marx

Danksagung

Als Erstes möchte ich mich bei meinem Doktorvater Prof. Dr. Rolf Müller für die Aufnahme in die Arbeitsgruppe, das entgegengebrachte Vertrauen und die beständige Unterstützung während des gesamten Promotionsstudiums bedanken. Zudem möchte ich Prof. Dr. Andriy Luzhetskyy für die wissenschaftliche Begleitung und die Erstellung des Zweitgutachtens der Dissertation danken. Mein besonderer Dank gilt Dr. Daniel Krug für seine tatkräftige Betreuung sowie seine positive und motivierende Einstellung.

Ein großes Dankeschön geht auch an Christina Decker und Ellen Merckel für ihre Organisation und die (meistens) gute Laune, sowie an alle anderen die das HIPS am Laufen halten: Christian Zeuner, Frank Jakob, Michael Roth, Mark Caspari sowie alle technischen Assistenten.

Als nächstes möchte ich mich bei meinen liebgewonnenen Freunden, Leidensgenossen und Affentennisspielern für jede Menge Dopamin und die unvergessliche gemeinsame Zeit bedanken. Vor allem die Strategiemeetings mit den Shinobis Jan und Seb, die Beratung in Wissenschaft, Geld und Politik von Dangerscientist Fabian, die Power meiner ständigen mit-Autorin und Endgegnerin im Dissertation-Korrigieren Chantal und die Kameradschaft und das Korrekturlesen von Jetpunk Weltmeister Niceolas haben mich durch diese Zeit getragen. An dieser Stelle ist der unermüdliche Einsatz von Prisma-Christine besonders hervorzuheben. Ohne sie wäre diese Arbeit immer noch nicht fertig und ich werde diesen selbstlosen Dienst niemals vergessen.

Meiner ganzen Familie gilt ein besonderer Dank für ihre bedingungslose Unterstützung und die gedankliche Freiheit, die sie leben und mir mitgegeben haben.

Mein letzter und größter Dank geht an meine Frau Yvonne: Danke für Deine unendlichen Geduld und dass du immer an meiner Seite stehst.

Zusammenfassung

Obwohl mikrobielle Naturstoffe (NS) sich seit Jahrzehnten als wichtige Quelle für pharmazeutisch relevante Substanzen erweisen, gestaltet sich die Entdeckung neuer bioaktiver Verbindungen zunehmend als aufwändig. Die vorliegende Arbeit beschreibt verschieden Strategien zur Identifikation neuer Wirkstoffe sowie deren anschließende Charakterisierung. Die Kombination aus bisher unterrepräsentierten bakteriellen Stämmen und Genom- sowie Metabolom-Mining Techniken erwies sich dabei als Erfolgsfaktor: Die Strukturfamilie der Thiamyxine wurde aus dem *Myxococcaceae* Stamm MCy9487 isoliert. Sie umschließt vier thiazolin- und thiazolreiche Peptide von denen zwei offenkettige und zwei zyklische Derivate sind. Beide zyklische Verbindungen erweisen sich durch potente Aktivität gegen RNA-Viren als vielversprechende Startpunkte für weitere Entwicklungen. Sesbanimid R ist der erste aus magnetotaktischen Bakterien isolierte NS und weist gute Aktivität gegen Karzinom Zelllinien auf. Diese Substanz wurde einem Biosynthesegencluster des Stammes *Magnetospirillum gryphiswaldense* durch gezielte Inaktivierung des Clusters und anschließendes statistisches Filtern des Metaboloms eindeutig zugeordnet. Neben der Suche nach bisher unentdeckten Sekundärmetaboliten (SM) befasst sich die vorliegende Arbeit mit einem Vergleich von Kultivierungs- und Analysetechniken. Es konnte gezeigt werden, dass Variation von Kultivierung und auch instrumenteller Analytik das Potential neue SM zu entdecken steigert.

Summary

Although microbial natural products (NPs) have proven to be an important source for pharmaceutically relevant substances, the discovery of novel bioactive compounds has grown increasingly difficult. This thesis describes different strategies for the identification of new active substances and their subsequent characterization. The application of genome and metabolome mining techniques to underrepresented bacterial strains was the key success factor therein: The compound family of the thiamyxins was isolated from the *Myxococcaceae* strain MCy9487. It is composed of four thiazoline- and thiazole-rich peptides, two of which are linear and two cyclized derivatives. Both cyclic derivatives were determined to be promising candidates for further development due to their potent activity against RNA-viruses. Sesbanimide R is a compound with good activity against human carcinoma cells and the first NP to be isolated from magnetotactic bacteria. It was unambiguously assigned to a biosynthetic gene cluster of *Magnetospirillum gryphiswaldense* by targeted inactivation of the cluster and subsequent statistical filtering of the metabolome. Besides the search for hitherto undiscovered secondary metabolites, this work also presents a comparative study of cultivation conditions and analytical setups. It was shown, that variation of culture conditions and analytical setups can significantly increase the potential of discovering new secondary metabolites.

Vorveröffentlichungen aus dieser Dissertation

Teile dieser Arbeit wurden vorab mit Genehmigung der Naturwissenschaftlich-Technischen Fakultäten, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen veröffentlicht oder sind derzeit in Vorbereitung zur Veröffentlichung:

Publikationen

Expanding the Scope of Detectable Microbial Natural Products by Complementary Analytical Methods and Cultivation Systems

Chantal D. Bader[1], Patrick A. Haack[1], Fabian Panter, Daniel Krug, and Rolf Müller*

Journal of Natural Products 2021 Feb 26;84(2):268-277

DOI: 10.1021/acs.jnatprod.0c00942

Sesbanimide R, a Novel Cytotoxic Polyketide Produced by Magnetotactic Bacteria

Ram Prasad Awal[1], Patrick A. Haack[1], Chantal D. Bader, Cornelius N. Riese, Dirk Schüler and Rolf Müller*

mBio 2021 May 18;12(3):e00591-21

DOI: 10.1128/mBio.00591-21

Thiamyxins: Structure and Biosynthesis of Myxobacterial RNA-Virus Inhibitors

Patrick A. Haack[1], Dr. Kirsten Harmrolfs[1], Dr. Chantal D. Bader[1], Dr. Ronald Garcia, Dr. Antonia P. Gunesch, Dr. Sibylle Haid, Dr. Alexander Popoff, Alexander Voltz, Dr. Heeyoung Kim, Prof. Dr. Ralf Bartenschlager, Prof. Dr. Thomas Pietschmann, Prof. Dr. Dr. Rolf Müller*

Angew. Chem. Int. Ed. 2022 Dec, 61, e202212946

DOI: 10.1002/anie.202212946

Publikationen die nicht Teil dieser Arbeit sind

In vivo and in vitro reconstitution of unique key steps in cystobactamid antibiotic biosynthesis

Sebastian Groß[1], Bastien Schnell[1], Patrick A. Haack, David Auerbach and Rolf Müller*

Nature Communications 2021 Mar 16;12(1):1696

DOI: 10.1038/s41467-021-21848-3

[1] These Authors contributed equally to the manuscript

Table of Contents

1 Introduction

1.1 Natural Products and their Pharmaceutical Applications

Natural Products (NPs) or secondary metabolites are defined as small molecules produced by biological sources such as plants, animals, fungi and bacteria.[1] Secondary metabolites in contrast to primary metabolites are not strictly necessary for the survival of the host, but still provide an evolutionary advantage.[2] NPs have been used throughout human history for the treatment of various illnesses. Herbal remedies, for example, are a staple of traditional Chinese medicine, which dates back to about 1,100 B.C.[3,4] The discoveries of penicillin in 1928 and streptomycin in 1945 ushered in the age of microbial natural products and triggered immense efforts from pharmaceutical companies to isolate novel NPs from various microbial sources. The great chemical diversity of natural products and the associated biological activities led to them being an invaluable source for pharmaceuticals.[2] Per the most recent report of Newman and Cragg only about a third of newly approved small molecule drugs from 1981 to September 2019 are purely synthetic drugs. Two thirds of new drugs are either unaltered NPs, botanical drugs or they are based on NP such as NP derivatives, NP mimics or synthetic drugs containing a NP pharmacophore.[5]

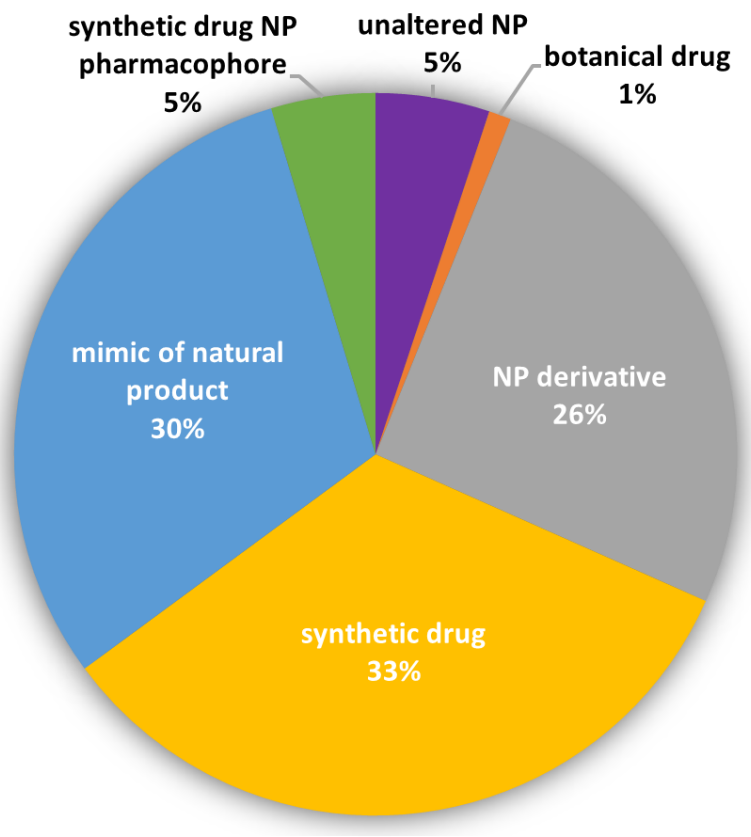


Figure **1**: This highlights the importance that natural products play to this day in various fields of medicine such as treatment of infections or cancer.

Although NPs boast a large variety of therapeutical applications, antibiotics like the penicillins are likely the most well-known pharmaceutical use of NP drugs.[2] The first commercially available antibiotics were the synthetic sulfonamides, which were discovered as an ingredient of cloth dye in the 1930s.[6] The first NP antibiotic was penicillin G, which was discovered in 1928, but was first applied as a drug to treat a *streptococcal* meningitis infection in 1942.[7] In the following 20 years many of the classes of antibiotics that we know today were developed. These include the β-lactams (1941), the aminoglycosides (1944), chloramphenicol (1949), the tetracyclines (1950), the lincosamides (1952), the macrolides (1952), the streptogramins (1952) the glycopeptides (1956), the ansamycins (1957) and the quinolones (1962).[8]

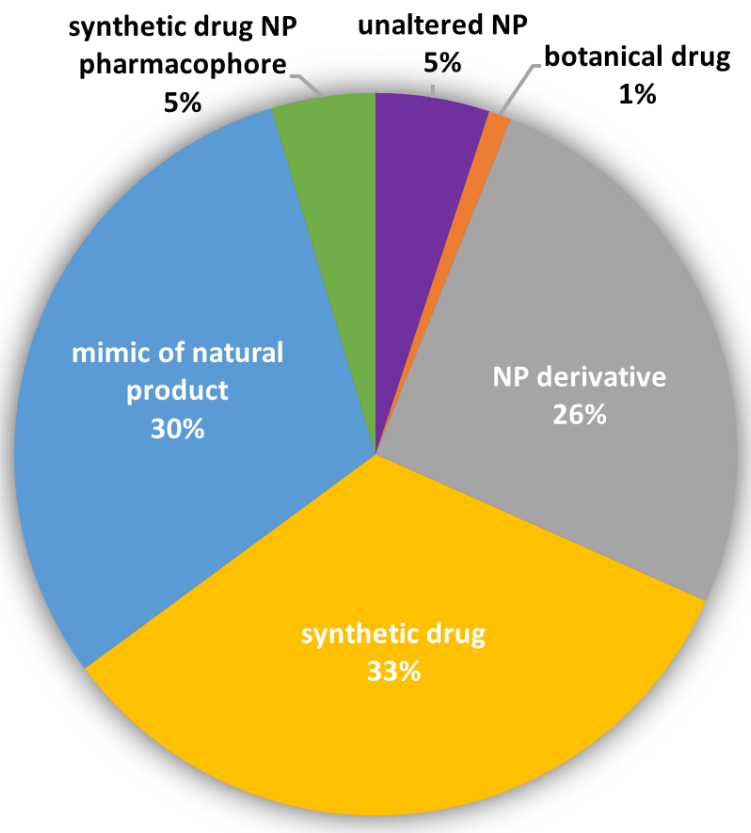


Figure 1: Classification of newly approved drugs from 1981 to 2019 by source. Adapted from [9]

This period, extending until the 1960s, is consequently known as the 'golden age' of antibiotic discovery. Although many second, third and fourth generation derivatives of already-known antibiotic compounds were developed during the period between the 1960s and 2000s, the discovery rate of new antibiotic classes dropped to almost zero.[10] As the stream of new compounds started to ebb, the realization arose that the classical methods of isolating compounds from microbial extracts by activity-guided approaches had been exhausted. Additionally, the plethora of recently discovered treatment options led to the belief that all future needs for antibiotic compounds were satisfied, which resulted in many pharmaceutical companies shutting down their NP research facilities. This presumption soon proved to be false due to the rising frequency of antimicrobial resistance (AMR) in human pathogens. Efforts to find antibiotic classes with novel targets, that could circumvent AMR, were increased again at the end of the second millennium, which led to the development of several novel antibiotic classes such as daptomycin and fidaxomicin. In spite of this, nosocomial infections with multi resistant pathogens are increasing on a global scale and the need for new lead structures is now more pressing than ever.[11,12] Prominent examples of NP antibiotics are displayed in **Figure 2**.

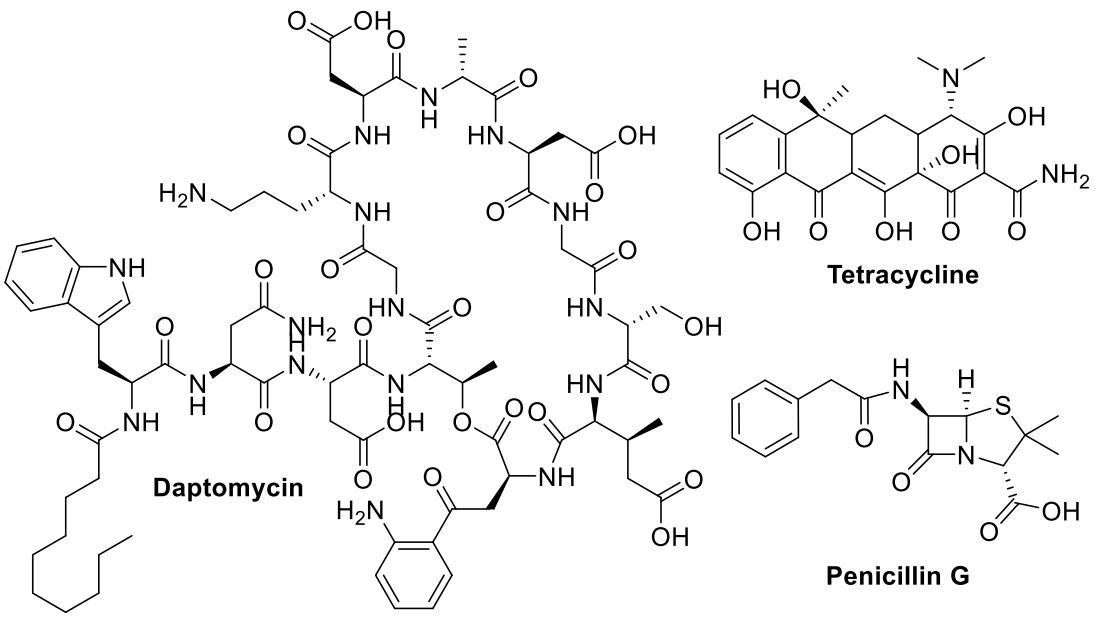


Figure 2: Prominent examples of natural product antibiotics.[13–15]

As is the case for antibiotics, NPs also make up a significant part of antitumor drugs on the market today. Most of these are of microbial origin such as mitomycin C from *Streptomyces caespitosus*, bleomycin from *Streptomyces verticillus*, actinomycin D from *Actinomyces antibioticus*, daunorubicin from *Streptomyces peucetius* and romidepsin from *Chromobacterium violaceum*.[16–20] Another major source of NP anticancer drugs is represented by plants, yielding vincristine from *Catharanthus roseus*, taxol from *Taxus brevifolia*, podophyllotoxin from *Podophyllum peltatum* and camptothecin from *Camptotheca acuminata* for example.[21–24] A representative overview of NP-derived antitumor drugs is depicted in **Figure 3**. One problem in developing antitumor drugs from NPs is that they are often too cytotoxic to be usable without further modifications. This is often accomplished my medicinal chemistry or creating antibody–drug-conjugates (ADCs) for target specific application.[25]

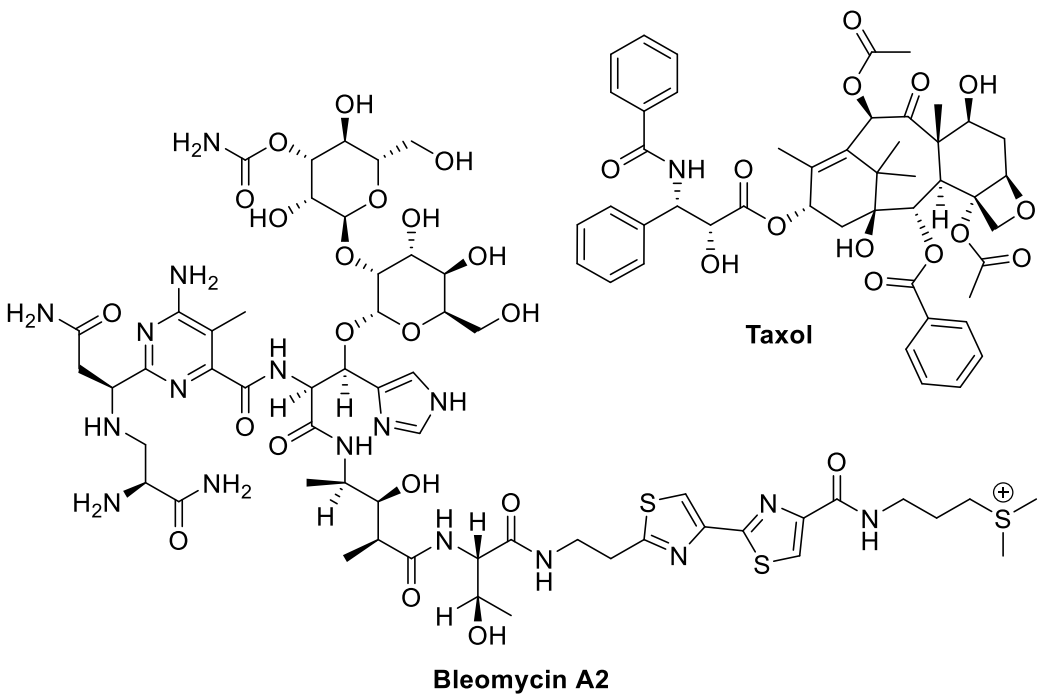


Figure 3: Natural product derived anticancer agents bleomycin and taxol.

Finally, NP based drugs also cover many other fields of application such as the antimalarial drugs artemisinin from *Artemisia annua* and quinine from *Cinchona pubescens*, analgesics like aspirin from the bark of the willow tree and morphine form *Papaver somniferum*, and the immunosuppressant cyclosporin from *Tolypocladium inflatum* as shown in **Figure 4**. [26–31]

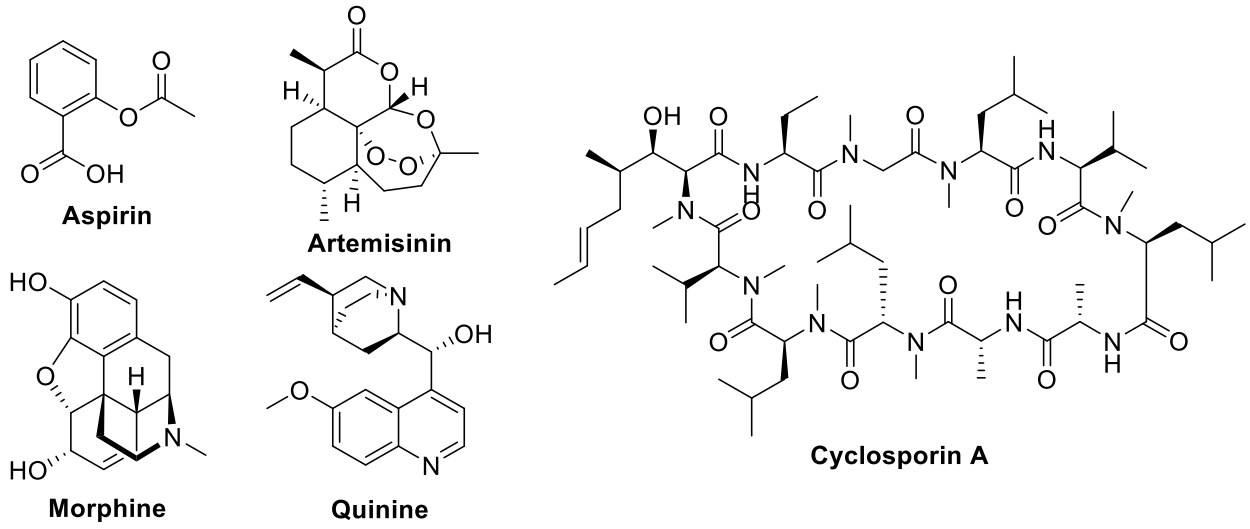


Figure 4: Well-known natural products displaying their structural diversity and wide-ranging applications: The analgesics aspirin and morphine, the antimalarial drugs artemisinin and quinine and the immunosuppressant cyclosporin A.

1.2 Myxobacteria as Sources of Secondary Metabolites

During the golden age of NP discovery bacteria and fungi were the main sources for secondary metabolites. Amongst bacteria, only three groups were considered "good" producers: The order *Actinomycetales*, the genera of *Bacillus* and the *Pseudomonads*. After most so-called "low hanging fruit" secondary metabolites from these groups had been discovered and rediscovery was becoming more frequent, other sources for bioactive compounds were needed. The resulting search was at least partially successful, revealing two novel fertile resources in cyanobacteria and the myxobacteria outlined here.[32] Myxobacteria are Gram-negative *δ-proteobacteria*. They occur ubiquitously in soil and have several characteristics unique among bacterial organisms. They can glide across surfaces and use this to form multicellular swarms. These swarms exhibit predatory behavior and use exoenzymes and secondary metabolites to lyse biological macromolecules and even whole microorganisms. Under harsh conditions they form so called fruiting bodies (**Figure 5**), that contain desiccation- and heat-resistant myxospores. The genomes of myxobacteria are amongst the largest of the bacterial kingdom and contain many biosynthetic gene clusters (BGCs).[33] As myxobacteria are in constant competition for habitats and nutrients with other microorganisms, they have evolved to be highly prolific in producing secondary metabolites with a variety of anti-microbial activities. Of particular interest are molecules with activity against Gram-negative bacteria, such as the Nonribosomal peptide – polyketide hybrid corramycin, isolated from *Corallococcus coralloides*, and displaying activity against *Escherichia coli* and *Salmonella typhimurium*.[34] Furthermore myxobacteria have also been shown to produce molecules with anti-Gram-positive bioactivity, for example the compound class of myxalamids. These NRPS-PKS hybrids were first isolated from *Myxococcus xanthus* and their main component myxalamid B exhibits activity against *Nocardia* species, *Corynebacterium mediolanum, Arthrobacter rubellus, Bacillus* species as well as activity against several molds and yeasts.[35,36] Another myxobacterial compound with antifungal activity is myxothiazol. This compound exhibits activity against *Pythium debaryanum, Piricularia oryzae* and *Trichophyton mentagrophytes,* is biosynthezised by an NRPS-PKS hybrid assembly line and was first isolated from *Myxococcus fulvus.*[37,38] Apart from compounds with antibacterial and antifungal activity, an important section of natural products research is concerned with discovering molecules with cytotoxicity against mammalian cells, which might be used in cancer therapy. Aurafuron A and B, polyketides isolated from *Stigmatella aurantiaca*, exhibit cytotoxic activity against a mouse fibroblast cell line.[39] Furthermore, the cytotoxic macrolides epothilones, discovered in 1995 and isolated in 1996 from *Sorangium cellulosum*, have been the starting point for several semisynthetic and synthetic anticancer agents.[40–42] An analog of epothilone B, ixabepilone, trade name ixempra®, is in use as a therapeutic against breast cancer and has

been proven to prolong progression-free survival in combination with capecitabine in comparison to capecitabine alone.[43] As demonstrated on a global scale during the COVID-19 pandemic, compounds with antiviral activity are also sorely needed and myxobacterial natural products are a promising source. The cyclic depsipeptides chondramides, initially isolated from *Chondromyces crocatus,* have shown high affinity to the SARS-CoV-2 spike protein receptor binding domain during *in silico* docking studies.[44,45] The recent discovery of the RNA polymerase inhibitors the sandacrabins, have proven once more the biosynthetic potential of myxobacteria. These antiviral terpenoid-alkaloids were isolated in 2022 from *Sandaracinus defensii* and feature potent activity against human pathogenic coronavirus 229E.[46,47] In chapter 2 we present the discovery of a family myxobacterial depsipeptides, the thiamyxins, with potent antiviral activity.

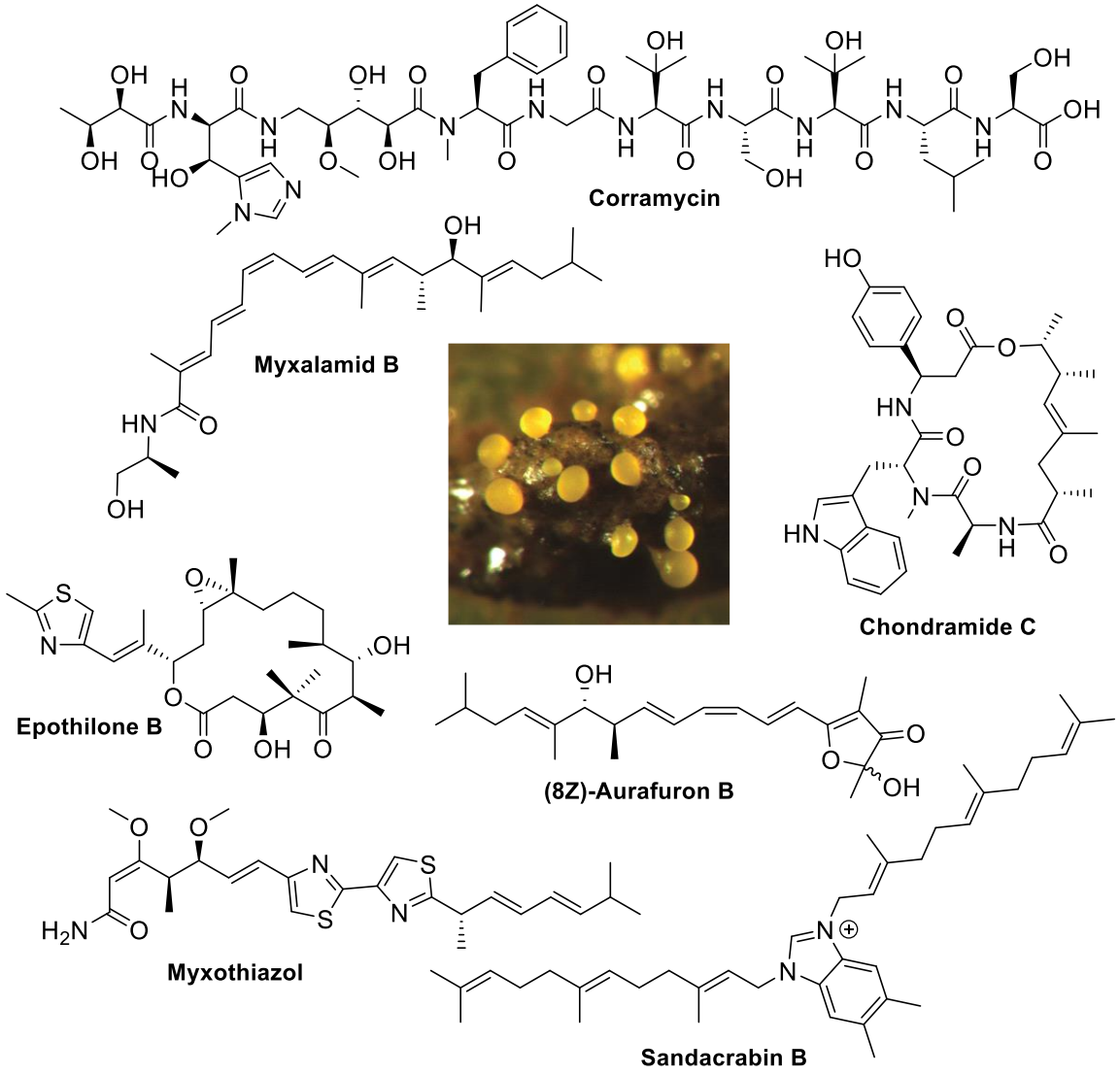


Figure 5: Fruiting bodies of Myxococcus xanthus and myxobacterial natural products representing various bioactivities: Corramycin with activity against Gram-negative bacteria; myxalamid B with anti-Gram-positive activity; myxothiazol as a representative of antifungal activity; (8Z)-aurafurone B and epothilone B as compounds with cytotoxic activity and chondramide C and sandacrabin B as antiviral compounds.[34,35,37,39,41,44,46]

1.3 Magnetotactic Bacteria

The umbrella term magnetotactic bacteria (MTB) describes a phylogenetically diverse group of prokaryotes. They were first discovered in 1963 by Salvatore Bellini and most of them belong to the *Proteobacteria* phylum, specifically to the classes of *α*-, *δ*- and *γ-proteobacteria*.[48,49] MTB garnered the interest of the scientific community as they are able to form highly unique organelles out of magnetite (Fe_3O_4) and greigite (Fe_3S_4). These organelles are called magnetosomes and allow the direction of movement of MTB to be influenced by the earth's magnetic field. This unique ability is called magnetotaxis and is thought to serve a similar function as chemotaxis by aiding MTB to maintain an optimal position in chemical concentration gradients. As most MTB are microaerophilic, aquatic microorganisms, the ability to passively align with and swim along magnetic field lines is thought to enable them to find a region of optimal oxygen concentration more readily.[50,51]

Similar to myxobacteria, MTB are a ubiquitous group of prokaryotes that have initially been explored due to an unusual morphological ability, but may now come under focus as a potential source for novel natural products. It was previously shown that phylogenetic distance correlates to the production of chemically distinct secondary metabolites and it was proven that the genomes of MTB may contain biosynthetic gene clusters.[48,52] This highlights MTB as a novel and promising source for the discovery of novel chemical scaffolds. In chapter 3 we present Sesbanimide R, the first bioactive secondary metabolite to be isolated from MTB.

1.4 Natural Product Biosynthesis

Two major biosynthesis routes for bacterial secondary metabolites are non-ribosomal peptide synthetases (NRPS) and polyketide synthases (PKS). Both are multimodular mega-enzymes that represent assembly lines and connect small monomers to form the final molecules. NRPS systems assemble peptides from single amino acids, including non-proteinogenic amino acids, and PKS assembly lines build polyketides from acyl units, similar to fatty acid synthases. NRPS as well as PKS products are often modified by tailoring enzymes, that can be part of the mega-enzyme complexes or independent proteins. Modifications include but are not limited to hydroxylation, methylation and heterocyclisation.[53]

1.4.1 Non-Ribosomal Peptide Synthetases

NRPS are organized in modules, with each module consisting of several domains with distinct functions. Each module consists at least of an adenylation (A) -domain, a peptide carrier protein (PCP) -domain, and

a condensation (C) -domain. One module is typically responsible for the introduction of one amino acid, with each step following the following general principle: A substrate is selected by the A-domain, activated through adenosine triphosphate (ATP) -dependent adenylation and loaded onto the phosphopantetheine (PPant) arm of the PCP-domain. The growing peptide chain connected to the upstream PCP-domain is then connected to the newly loaded substrate on the downstream PCP-domain. This step is catalyzed by the C-domain, which forms the peptide bond between the amino group of the nascent molecule and the thioester-carbonyl group of the next amino acid to be added. (**Figure 6**C) This is repeated for each module of the BGC until the last module releases the nascent molecule via a hydrolysis reaction. This step is usually catalyzed by a thioesterase (TE) -domain.[54] Examples of other, optional domains include heterocyclization, oxidation and epimerization-domains, the latter of which are able to epimerize L-amino acids into the respective D-amino acids. (**Figure 6**D)

Prominent examples of myxobacterial compounds containing heterocycles similar to the compound family described in chapter 2 and featuring NRPS modules in their biosynthetic machinery include the myxothiazols, isolated in 1981 from *Myxococcus fulvus,* and the melithiazols, isolated in 1999 from *Melittangium lichenicola*, *Archangium gephyra*, and *Myxococcus stipitatus*. Both compound families are electron transport inhibitors with antifungal activity.[55–58] Another example are the cystothiazoles, which were isolated in 1998 from *Cystobacter fuscus* and inhibit the growth of fungi as well as human tumor cells.[59,60] The BGCs of these compounds also contain PKS modules, a common phenomenon, as the underlying biochemical principles involved in NRPS and PKS biosynthesis are quite similar.[61] In the following section the biosynthetic logic of polyketide synthases will be explained in more detail.

1.4.2 Polyketide Synthases

Like the NRPS biosynthetic machinery type I PKS assembly lines follow a modular logic, where each module is responsible for the incorporation of an individual building block in the nascent molecule. Malonyl-CoA, which is the most common building block, or acetyl-CoA, which is the most common starter unit, is transferred to the PPant arm of an acyl carrier protein (ACP) -domain by an acyltransferase (AT) -domain. A nucleophilic attack of the ACP-bound malonyl C2 carbon to the nascent molecule leads to the formation of a C-C bond via a decarboxylative Claisen condensation. This reaction is catalyzed by a ketosynthase (KS) -domain and the growing chain is transferred to ACP-domain of the current module. The AT-domain of the following module then selects the next extender unit and transfers it to the corresponding ACP-domain, where the next Claisen condensation is then catalyzed by the KS domain of that module. (Figure 6A) The resulting β-ketones of each elongation step can be further reduced by additional non-obligatory domains, which affect the building block added in the previous module. The first reduction is catalyzed by a keto

reductase (KR) -domain forming a β-alcohol, the stereochemistry of which is determined by the type of the KR domain. The resulting hydroxy group can then be further reduced by a dehydratase (DH) -domain to form an α-β-enoyl group, which in a last step might be reduced to a

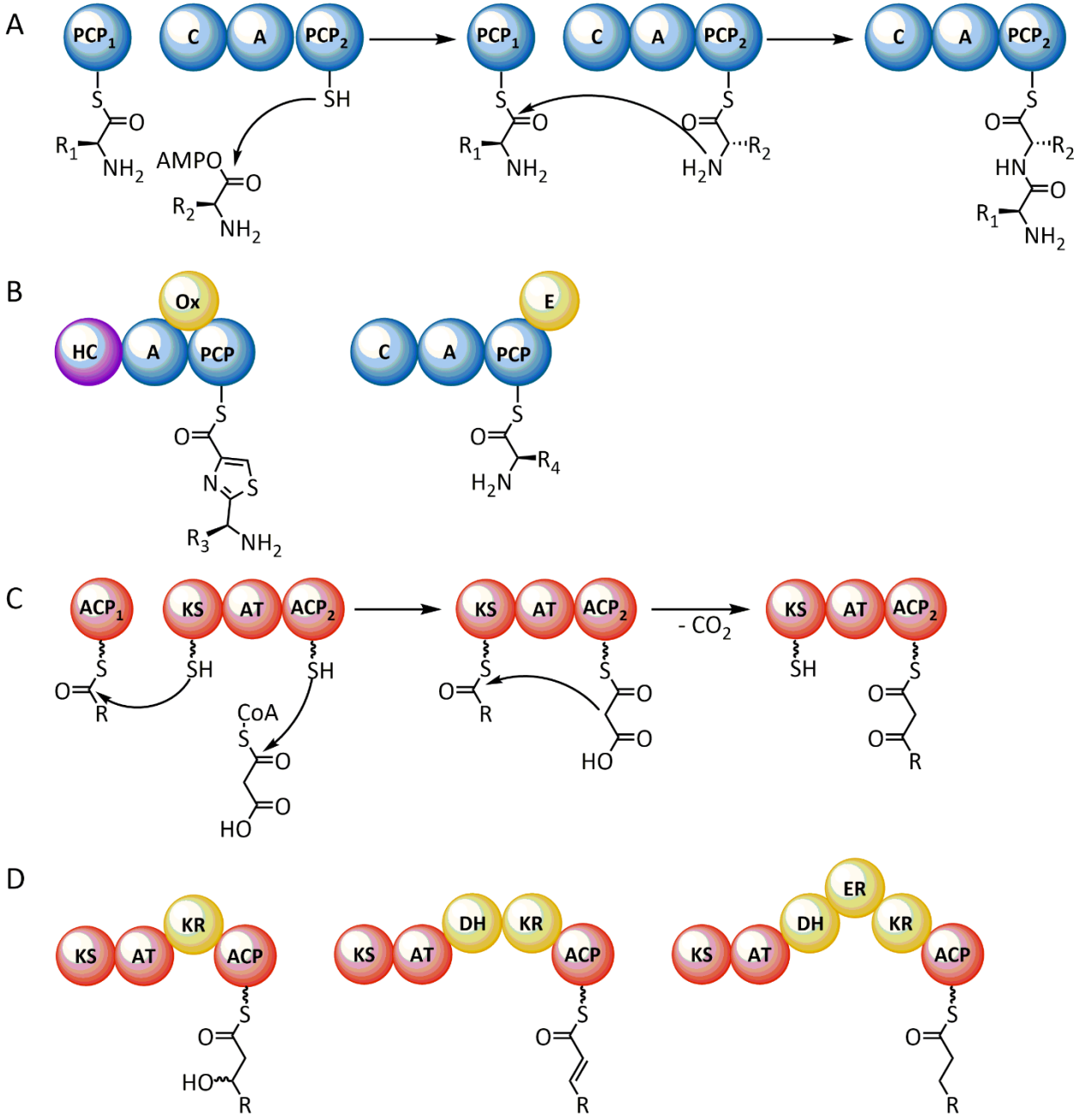


Figure 6: Schematic representations of different aspects of nonribosomal peptide and polyketide biosynthesis. A: The core reaction of nonribosomal peptide biosynthesis. B: Examples of NRPS modules with optional modifying domains that catalyze heterocycle formation and subsequent oxidation or incorporation of D-amino acids into the nascent molecule. C: The core reaction of polyketide biosynthesis. D: Examples of the different redox states of the polyketide extender unit depending on the optional redox domains present in the module.

saturated CH_2-CH_2 group by an enoyl reductase (ER) -domain. (**Figure 6**B) The variable inclusion of these three domains along with other optional modifying domains such at methyl-transferase and β-branching domains enable a high structural diversity for PKS products. Next to type I PKS, polyketides can also be produced by type II and type III PKS systems. In contrast to type I PKS, type II and III PKS work iteratively, where type II PKS consist of separate monofunctional enzymes and type III systems do not require an ACP-bound substrate, but can use free acyl-CoA esters as building blocks.[62–65]

Type I PKS can be further separated into so-called *cis*-AT PKS and *trans*-AT PKS.[66] While both types have the same non-iterative modular structure, they are thought to have evolved separately from each other. The major difference between *cis*- and *trans*-AT PKS assembly lines is the location of AT functionality. In *trans*-AT BGCs one to two AT domains are encoded on a separate open reading frame (orf) and are recruited by so-called docking domains located in each module to catalyze the Claisen condensation reaction. In the case of two AT domains encoded in the cluster, one has a proofreading functionality to remove stalled intermediates from the assembly line. Other than the AT functionality, several other peculiarities are common in *trans*-AT PKS biosynthetic pathways. Among these are non-elongating modules, modules divided onto two proteins and β-branching modules. These unusual features have the effect that the colinearity rule connecting module architecture and polyketide structure in *cis*-AT PKS cannot be applied to *trans*-AT PKS. The structure of *trans*-AT PKS products can however be predicted with a high degree of accuracy by clustering the KS domains of the BGC in a phylogenetic tree, as evolutionarily related KS domains of *trans*-AT PKS assembly lines are known to accept similar intermediates.[67,68]

Among the best-studied bacterial *trans*-AT PKS products is bacillaene, which was first isolated from *Bacillus subtilis*.[69] The elucidation of the bacillaene biosynthesis has shed light on many principles fundamental to *trans*-AT PKS logic.[68,70] Another prominent compound family of the *trans*-AT PKS class are the disorazols. First isolated in 1994 from *S. cellulosum*, the biosynthetic machinery of these potent cytotoxic compounds was the first myxobacterial *trans*-AT PKS to be expressed heterologously, thereby allowing for yield optimization and having the potential for the generation of new derivatives.[71,72] One of the most potent myxobacterial antibiotics ever identified and another example of *trans*-AT PKS products are the sorangicins. They were first isolated in 1987 from *Sorangium cellulosum* and exhibit activity against Gram-positive as well as Gram-negative bacteria.[73,74]

Apart from multimodular mega-enzymes, other prominent sources for secondary metabolites include terpenoid- as well as ribosomally synthesized and post-translationally modified peptide biosynthetic

pathways. As this work is focused on NRPS (thiamyxins - chapter 2) and *trans*-AT PKS (sesbanimide R - chapter 3) compounds, these additional biosynthetic routes will not be elaborated further here.

1.5 Approaches and Challenges of Natural Product Discovery

The different approaches to natural products discovery can broadly be categorized by three themes: activity-guided, genome mining- and metabolome mining-based approaches. Activity-guided approaches have historically employed the following general concept: Microbial or plant crude extracts were screened against test organisms in order to identify interesting bioactivities. The extract was then fractioned by its physiochemical properties through chromatography and the fractions were screened again to connect a UV or MS signal to the activity. Thus, the compounds could be isolated and further studies such as structure elucidation, mode of action determination and in some cases clinical development could be performed. During the period between 1940 and 1970 actinobacteria were intensively subjected to this kind of bioactivity screening. This led to the discovery of a vast number of NPs, including the antitumor agent actinomycin and the antibiotics vancomycin, streptothricin and streptomycin, the last of which is most famous for its use in tuberculosis treatment.[75–79]

Although activity-based approaches delivered many of the first clinically used antibiotics during the golden age of antibiotic discovery, this progress was slowed down, as rediscovery increasingly became an issue. Therefore, scientists turned to emerging technologies in order to access novel secondary metabolite scaffolds: The advancement of genome sequencing enabled the targeted search for biosynthetic gene clusters and increasingly sensitive analytical methods coupled with big data processing capabilities paved the way for statistical evaluation of whole microbial metabolomes.[80,81]

In contrast to the random nature of activity-guided screenings, genome mining approaches rely on the identification of biosynthetic gene clusters and targeted selection of the most interesting ones for further studies in order to connect them to their respective secondary metabolite products. Out of the vast pool of biosynthetic gene clusters with unknown functions, those that might be responsible for the production of antimicrobial secondary metabolites are of particular interest. A successful strategy in selecting these BGCs has been to identify self-resistance genes in the vicinity of the core cluster, as employed in the discovery of the topoisomerase inhibiting pyxidicyclines.[82] Once a BGC has been identified, *in silico* analysis of the cluster can be employed to make detailed predictions about its biosynthetic logic, the precursors involved in the biosynthesis and even on the final structure of the secondary metabolite.[83,84]

It has been shown that the genomes of many well-known producers of secondary metabolites contain considerably more biosynthetic gene clusters than the number of compounds that have been isolated

from them. The reason for this can be manifold, but most likely these genes are simply not expressed in sufficient levels under standard laboratory cultivation conditions.[85] To increase transcription levels of these "cryptic" cluster and thereby hopefully boost the production of the associated natural product, heterologous promotors can be inserted into the genome of the host strain. This has recently led to the identification of the sandarazols, plasmid-borne cytotoxins isolated from *Sandaracinus defensii*. MSr10575.[86] As an alternative to genetic manipulation of the natural host strain, cloning of the cluster into a heterologous production host has also proven to be a viable strategy. This has the added advantage that additional genetic manipulations to determine details of the biosynthesis, further improvement of production titers and even formation of entirely novel congeners are possible.[87]

Another way of dealing with strains while avoiding rediscovery issues, is to analyze their secondary metabolome with various sophisticated instrumental analytics. The most common analytical setups today include a combination of high-performance liquid chromatography (HPLC) and mass spectrometry (MS). Liquid chromatography (LC) is a separation technique employing a stationary and a mobile phase with different physicochemical properties. The components of a complex mixture, such as a culture broth extract, interact differently with both phases based on their own chemistry and can therefore be separated. Common LC setups, that allow for the separation of analytes based on their polarity, are normal phase LC, employing a polar stationary phase and a non-polar liquid phase and reverse phase LC, with a non-polar stationary phase and polar liquid phase. Other chromatography methods include size exclusion chromatography, affinity chromatography and ion-exchange chromatography.[88] As with the different LC setups, there also exist a vast amount of different mass spectrometry analyzers: Quadrupole mass filters measure the mass to charge (*m/z*) ratio of the ions by filtering them through oscillating electrical fields; Time-of-flight (TOF) mass spectrometers use flight tubes and measure the time ions require to reach a detector, which correlates to their *m/z*; and Fourier transform mass spectrometers measure the image current of ions passing by a detector in a periodic function and use Fourier transformation to calculate their mass. FT-MS is able to reach the highest resolution of these example followed by TOF-MS and then Quadrupole-MS. To analyze molecules via mass spectrometry, they first need to be converted into ions. In natural products research, a commonly employed ionization technique is electrospray ionization, as it can seamlessly couple liquid chromatography separation to several mass analyzers with very little fragmentation of the analyte.[89] The LC part of such a coupling has a meaningful influence on the separation capability of the setup, which allows for a large degree of focus on a specific group of compounds based on their chemical properties, but at same time also discriminates against others. The same holds true for the compound detection via mass spectrometry. Different types of ionization as well

as detection methods will allow the detection of incredibly low quantities of some metabolites while not detecting others at all. During large-scale screening approaches, the analytical setup is therefore usually highly standardized in order to generate robust datasets for statistical treatment.[81]

The application and combination of genome mining and metabolome mining techniques has led to the discovery of a number of novel secondary metabolites that would have remained elusive in standard activity-based approaches: Myxoprincomide, an NRPS/PKS product isolated from *Myxococcus xanthus* DK1622, was identified by targeted gene inactivation of a BGC and subsequent application of statistical filters to the secondary metabolomes of the wild type and the mutant strains.[90] Another example are the fulvuthiacenes, a family of NRPS/PKS compounds isolated from *Myxococcus fulvus*, which were identified by combining statistical filtering and spectral networking of high resolution LC-MS data.[91] Finally rowithocin, a polyketide isolated from *Sorangium cellulosum* MSr2476, was identified during a correlation study between taxonomic distance and chemical diversity combined with target masses associated with activity against *staphylococcusl*[52]

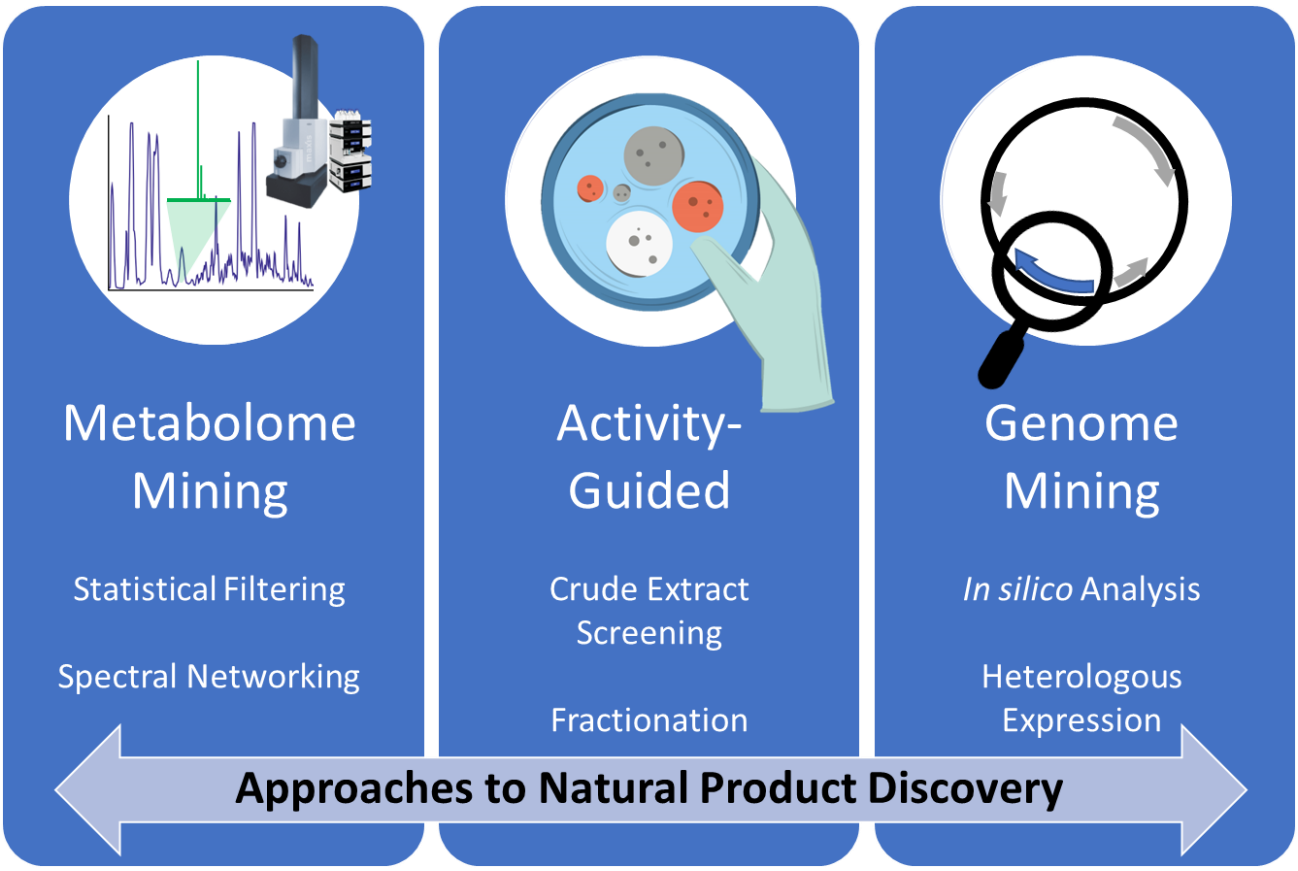


Figure 7: Schematic representation of the three main approaches to natural products discovery and examples of their implementation. Petri dish image obtained from[92].

1.6 Outline of this work

The primary aim of the work presented in this thesis was the discovery of novel natural products from microbial sources. In order to achieve this rather broadly characterized goal, underinvestigated microbial sources such as the *Myxococcaceae* strain MCy9487 and the even more elusive magnetotactic bacterium *Magnetospirillum gryphiswaldense* were explored. To aid compound discovery and biosynthesis elucidation efforts, metabolome and genome mining techniques such as *in silico* analysis of the biosynthetic gene cluster and principal component analysis were employed. As a result of these efforts, two bioactive secondary metabolites were isolated and characterized on a chemical and biochemical level. Furthermore, a fundamental question in metabolome screening approaches was challenged by comparing two different analytical setups – direct infusion Fourier transform ion cyclotron resonance mass spectrometry and ultra-high-performance liquid chromatography coupled with time-of-flight mass spectrometry - by the chemical space they cover.

The first part of this thesis is focused on the discovery and characterization of the thiamyxins. This family of cyclic depsipeptides was isolated from the *Myxococcaceae* strain MCy9487 and two of the four congeners exhibit potent activity against RNA viruses including the human pathogenic corona virus hCoV-229E. The aim of this study was to elucidate the structure of these compounds, determine their stereochemistry and generate a biosynthesis model based on in silico analysis of the cluster and isotopically labeled precursor feeding.

The second part of this thesis describes the discovery of the first secondary metabolite to be isolated from a magnetotactic bacterium. Sesbanimide R is a *trans*-AT PKS product isolated from *Magnetospirillum gryphiswaldense* and exhibiting cytotoxic activity against several carcinoma cell lines. It was connected to its biosynthetic machinery by targeted deletion, activation experiments and subsequent application of statistical filters to extracts of the wildtype and mutant strains. In *silico* analysis, mass spectrometry, nuclear magnetic resonance spectroscopy and Marfey's derivatization were used to provide a concise picture of its structure, stereochemistry, and biosynthesis.

The final part of this thesis is dedicated to a comparative study of different analytical setups and cultivation conditions. The aim of which was to quantify the discriminatory impact the analytical system has on metabolomics studies. Nine myxobacterial strains were cultivated in liquid and on solid medium and the extracts of the cultures were analyzed with DI-FTICR an uHPLC-TOFMS. Comparative analysis of the detected features revealed an expected but nevertheless striking gap in coverage of the respective chemical spaces. Furthermore, MS^2 analysis and molecular networking were applied in order to determine

the unique secondary metabolome of each strain, which can be used to prioritize features for further analysis.

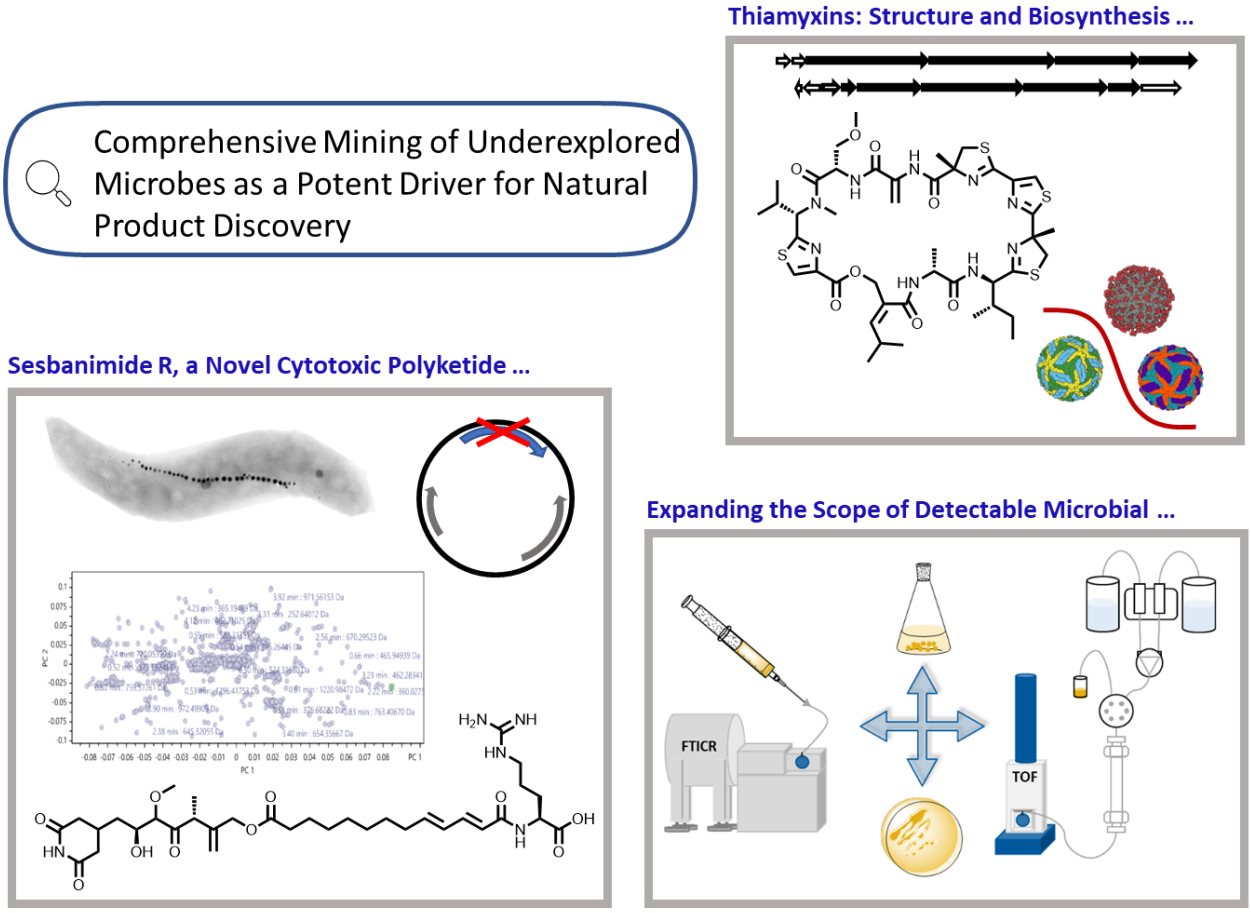


Figure 8: Representation of the different approaches to natural products discovery presented in this thesis.

1.7 References

[1] Nature Publishing Group, *Nat Chem Biol* **2007**, *3*, 351.

[2] L. Katz, R. H. Baltz, *J Ind Microbiol Biotechnol* **2016**, *43*, 155–176.

[3] D. Normile, *Science* **2003**, *299*, 188–190.

[4] A. Y. Leung, *Toxicol Pathol* **2006**, *34*, 319–326.

[5] D. J. Newman, G. M. Cragg, *J Nat Prod* **2020**, DOI 10.1021/acs.jnatprod.9b01285.

[6] H. OTTEN, *Journal of Antimicrobial Chemotherapy* **1986**, *17*, 689–690.

[7] A. Fleming, *The Lancet* **1943**, *242*, 434–438.

[8] J. Bérdy, *J Antibiot (Tokyo)* **2012**, *65*, 385–395.

[9] D. J. Newman, G. M. Cragg, *J Nat Prod* **2020**, *83*, 770–803.

[10] M. Hutchings, A. Truman, B. Wilkinson, *Curr Opin Microbiol* **2019**, *51*, 72–80.

[11] C. Walsh, T. Wencewicz, T. A. Wencewicz, *Antibiotics: Challenges - Mechanisms - Opportunities*, ASM Press, Washington, DC, **2016**.

[12] N. Khardori, C. Stevaux, K. Ripley, *The Indian Journal of Pediatrics* **2020**, *87*, 39–42.

[13] A. FLEMING, *Br Med Bull* **1944**, *2*, 4–5.

[14] G. M. Eliopoulos, S. Willey, E. Reiszner, P. G. Spitzer, G. Caputo, R. C. Moellering, *Antimicrob Agents Chemother* **1986**, *30*, 532–535.

[15] C. R. Stephens, L. H. Conover, F. A. Hochstein, P. P. Regna, F. J. Pilgrim, K. J. Brunings, R. B. Woodward, *J Am Chem Soc* **1952**, *74*, 4976–4977.

[16] S. WAKAKI, H. MARUMO, K. TOMIOKA, G. SHIMIZU, E. KATO, H. KAMADA, S. KUDO, Y. FUJIMOTO, *Antibiot Chemother (Northfield)* **1958**, *8*, 228–40.

[17] H. Umezawa, K. Maeda, T. Takeuchi, Y. Okami, *J Antibiot (Tokyo)* **1966**, *19*, 200–9.

[18] C. E. DALGLIESH, A. R. TODD, *Nature* **1949**, *164*, 830–830.

[19] B. Camerino, G. Palamidessi, *Gazz Chim Ital* **1960**, *90*, 1802–1815.

[20] H. UEDA, H. NAKAJIMA, Y. HORI, T. FUJITA, M. NISHIMURA, T. GOTO, M. OKUHARA, *J Antibiot (Tokyo)* **1994**, *47*, 301–310.

[21] M. C. Wani, H. L. Taylor, M. E. Wall, P. Coggon, A. T. McPhail, *J Am Chem Soc* **1971**, *93*, 2325–2327.

[22] I. S. JOHNSON, J. G. ARMSTRONG, M. GORMAN, J. P. BURNETT, *Cancer Res* **1963**, *23*, 1390–427.

[23] V. Podwyssotzki, *Archiv für Experimentelle Pathologie und Pharmakologie* **1880**, *13*, 29–52.

[24] M. E. Wall, M. C. Wani, C. E. Cook, K. H. Palmer, A. T. McPhail, G. A. Sim, *J Am Chem Soc* **1966**, *88*, 3888–3890.

[25] L.-A. Giddings, D. J. Newman, *J Ind Microbiol Biotechnol* **2013**, *40*, 1181–1210.

[26] D. J. COHEN, *Ann Intern Med* **1984**, *101*, 667.

[27] J. M. Liu, M. Ni, J. Fan, Y. Tu, Z. Wu, Y. Wu, W. Zhou, *Acta Chimi* **1979**, *37*, 129–143.

[28] H. Dreser, *Archiv für die gesamte Physiologie des Menschen und der Tiere* **1899**, *76*, 306–318.

[29] G. Lockermann, *J Chem Educ* **1951**, *28*, 277.

[30] A. M. Clark, *Pharm Res* **1996**, *13*, 1133–1144.

[31] E. Stone, *Philos Trans R Soc Lond* **1763**, *53*, 195–200.

[32] H. Reichenbach, *J Ind Microbiol Biotechnol* **2001**, *27*, 149–156.

[33] S. C. Wenzel, R. Müller, *Nat Prod Rep* **2009**, *26*, 1385.

[34] C. Couturier, S. Groß, A. von Tesmar, J. Hoffmann, S. Deckarm, A. Fievet, N. Dubarry, T. Taillier, C. Pöverlein, H. Stump, M. Kurz, L. Toti, S. Haag Richter, D. Schummer, P. Sizun, M. Hoffmann, R. Prasad Awal, N. Zaburannyi, K. Harmrolfs, J. Wink, E. Lessoud, T. Vermat, V. Cazals, S. Silve, A. Bauer, M. Mourez, L. Fraisse, C. Leroi-Geissler, A. Rey, S. Versluys, E. Bacqué, R. Müller, S. Renard, *Angewandte Chemie International Edition* **2022**, *61*, DOI 10.1002/anie.202210747.

[35] K. GERTH, R. JANSEN, G. REIFENSTAHL, G. HÖFLE, H. IRSCHIK, B. KUNZE, H. REICHENBACH, G. THIERBACH, *J Antibiot (Tokyo)* **1983**, *36*, 1150–1156.

[36] B. Silakowski, G. Nordsiek, B. Kunze, H. Blöcker, R. Müller, *Chem Biol* **2001**, *8*, 59–69.

[37] K. GERTH, H. IRSCHIK, H. REICHENBACH, W. TROWITZSCH, *J Antibiot (Tokyo)* **1980**, *33*, 1474–1479.

[38] W. Trowitzsch, G. Reifenstahl, V. Wray, G. Höfle, *Journal of Antibiotics Tokyo* **1980**, *33*, 1480–1490.

[39] B. Kunze, H. Reichenbach, R. Müller, G. Höfle, *J Antibiot (Tokyo)* **2005**, *58*, 244–251.

[40] D. M. Bollag, P. A. McQueney, J. Zhu, O. Hensens, L. Koupal, J. Liesch, M. Goetz, E. Lazarides, M. Woods, C. M. Woods, *Cancer Res* **1995**, *55*, 2325–2333.

[41] G. Höfle, N. Bedorf, H. Steinmetz, D. Schomburg, K. Gerth, H. Reichenbach, *Angewandte Chemie International Edition in English* **1996**, *35*, 1567–1569.

[42] F. Y. F. Lee, R. Borzilleri, C. R. Fairchild, A. Kamath, R. Smykla, R. Kramer, G. Vite, *Cancer Chemother Pharmacol* **2008**, *63*, 157–166.

[43] E. S. Thomas, H. L. Gomez, R. K. Li, H.-C. Chung, L. E. Fein, V. F. Chan, J. Jassem, X. B. Pivot, J. v. Klimovsky, F. H. de Mendoza, B. Xu, M. Campone, G. L. Lerzo, R. A. Peck, P. Mukhopadhyay, L. T. Vahdat, H. H. Roché, *Journal of Clinical Oncology* **2007**, *25*, 5210–5217.

[44] B. KUNZE, R. JANSEN, F. SASSE, G. HÖFLE, H. REICHENBACH, *J Antibiot (Tokyo)* **1995**, *48*, 1262–1266.

[45] R. A. Fernandez, M. T. Quimque, K. I. Notarte, J. A. Manzano, D. Y. Pilapil, V. N. de Leon, J. J. San Jose, O. Villalobos, N. H. Muralidharan, M. M. Gromiha, S. Brogi, A. P. G. Macabeo, *J Biomol Struct Dyn* **2022**, *40*, 12209–12220.

[46] C. D. Bader, F. Panter, R. Garcia, E. P. Tchesnokov, S. Haid, C. Walt, C. Spröer, A. F. Kiefer, M. Götte, J. Overmann, T. Pietschmann, R. Müller, *Chemistry – A European Journal* **2022**, *28*, DOI 10.1002/chem.202104484.

[47] N. A. Frank, Advancing Myxobacterial Natural Product Discovery by Combining Genome and Metabolome Mining with Organic Synthesis, Universität des Saarlandes, **2022**.

[48] A. Araujo, F. Abreu, K. Silva, D. Bazylinski, U. Lins, *Mar Drugs* **2015**, *13*, 389–430.

[49] Bellini S, *University of Pavia, Institute of Microbiology, Pavia* **1963**.

[50] D. A. Bazylinski, C. T. Lefèvre, D. Schüler, in *The Prokaryotes*, Springer Berlin Heidelberg, Berlin, Heidelberg, **2013**, pp. 453–494.

[51] S. Klumpp, D. Faivre, *Eur Phys J Spec Top* **2016**, *225*, 2173–2188.

[52] T. Hoffmann, D. Krug, N. Bozkurt, S. Duddela, R. Jansen, R. Garcia, K. Gerth, H. Steinmetz, R. Müller, *Nat Commun* **2018**, *9*, 803.

[53] M. A. Fischbach, C. T. Walsh, *Chem Rev* **2006**, *106*, 3468–3496.

[54] C. T. Walsh, *Nat Prod Rep* **2016**, *33*, 127–135.

[55] K. GERTH, H. IRSCHIK, H. REICHENBACH, W. TROWITZSCH, *J Antibiot (Tokyo)* **1980**, *33*, 1474–1479.

[56] B. Silakowski, H. U. Schairer, H. Ehret, B. Kunze, S. Weinig, G. Nordsiek, P. Brandt, H. Blöcker, G. Höfle, S. Beyer, R. Müller, *Journal of Biological Chemistry* **1999**, *274*, 37391–37399.

[57] F. SASSE, B. BÖHLENDORF, M. HERMANN, B. KUNZE, E. FORCHE, H. SIEINMEIZ, G. HÖFLE, H. REICHENBACH, *J Antibiot (Tokyo)* **1999**, *52*, 721–729.

[58] S. Weinig, H.-J. Hecht, T. Mahmud, R. Müller, *Chem Biol* **2003**, *10*, 939–952.

[59] M. Ojika, Y. Suzuki, A. Tsukamoto, Y. Sakagami, R. Fudou, T. Yoshimura, S. Yamanaka, *Journal of Antibiotics Tokyo* **1998**, *51*, 275–281.

[60] Z. FENG, J. QI, T. TSUGE, Y. OBA, T. KOBAYASHI, Y. SUZUKI, Y. SAKAGAMI, M. OJIKA, *Biosci Biotechnol Biochem* **2005**, *69*, 1372–1380.

[61] S. L. Wenski, S. Thiengmag, E. J. N. Helfrich, *Synth Syst Biotechnol* **2022**, *7*, 631–647.

[62] C. Hertweck, *Angewandte Chemie International Edition* **2009**, *48*, 4688–4716.

[63] K. J. Weissman, in *Comprehensive Natural Products III*, Elsevier, **2020**, pp. 4–46.

[64] M. A. Fischbach, C. T. Walsh, *Chem Rev* **2006**, *106*, 3468–3496.

[65] P. Caffrey, *Chem Biol* **2005**, *12*, 1060–1062.

[66] X. Sun, Y. Yuan, Q. Chen, S. Nie, J. Guo, Z. Ou, M. Huang, Z. Deng, T. Liu, T. Ma, *Nat Commun* **2022**, *13*, 5541.

[67] E. J. N. Helfrich, R. Ueoka, A. Dolev, M. Rust, R. A. Meoded, A. Bhushan, G. Califano, R. Costa, M. Gugger, C. Steinbeck, P. Moreno, J. Piel, *Nat Chem Biol* **2019**, *15*, 813–821.

[68] E. J. N. Helfrich, J. Piel, *Nat Prod Rep* **2016**, *33*, 231–316.

[69] P. S. PATEL, S. HUANG, S. FISHER, D. PIRNIK, C. AKLONIS, L. DEAN, E. MEYERS, P. FERNANDES, F. MAYERL, *J Antibiot (Tokyo)* **1995**, *48*, 997–1003.

[70] J. Moldenhauer, X.-H. Chen, R. Borriss, J. Piel, *Angewandte Chemie International Edition* **2007**, *46*, 8195–8197.

[71] Q. Tu, J. Herrmann, S. Hu, R. Raju, X. Bian, Y. Zhang, R. Müller, *Sci Rep* **2016**, *6*, 21066.

[72] R. Jansen, H. Irschik, H. Reichenbach, V. Wray, G. Höfle, *Liebigs Ann Chem* **1994**, *1994*, 759–773.

[73] H. Irschik, M. Kopp, K. J. Weissman, K. Buntin, J. Piel, R. Müller, *ChemBioChem* **2010**, *11*, 1840–1849.

[74] H. IRSCHIK, R. JANSEN, K. GERTH, G. HÖFLE, H. REICHENBACH, *J Antibiot (Tokyo)* **1987**, *40*, 7–13.

[75] L. Katz, R. H. Baltz, *J Ind Microbiol Biotechnol* **2016**, *43*, 155–176.

[76] S. A. Waksman, H. B. Woodruff, *Exp Biol Med* **1940**, *45*, 609–614.

[77] S. A. Waksman, H. B. Woodruff, *Exp Biol Med* **1942**, *49*, 207–210.

[78] A. Schatz, E. Bugle, S. A. Waksman, *Exp Biol Med* **1944**, *55*, 66–69.

[79] M. H. MCCORMICK, J. M. MCGUIRE, G. E. PITTENGER, R. C. PITTENGER, W. M. STARK, *Antibiot Annu* **1955**, *3*, 606–11.

[80] N. Ziemert, M. Alanjary, T. Weber, *Nat Prod Rep* **2016**, *33*, 988–1005.

[81] D. Krug, R. Müller, *Nat Prod Rep* **2014**, *31*, 768.

[82] F. Panter, D. Krug, S. Baumann, R. Müller, *Chem Sci* **2018**, *9*, 4898–4908.

[83] K. Blin, S. Shaw, A. M. Kloosterman, Z. Charlop-Powers, G. P. van Wezel, M. H. Medema, T. Weber, *Nucleic Acids Res* **2021**, *49*, W29–W35.

[84] E. J. N. Helfrich, R. Ueoka, A. Dolev, M. Rust, R. A. Meoded, A. Bhushan, G. Califano, R. Costa, M. Gugger, C. Steinbeck, P. Moreno, J. Piel, *Nat Chem Biol* **2019**, *15*, 813–821.

[85] S. C. Wenzel, R. Müller, *Mol Biosyst* **2009**, *5*, 567.

[86] F. Panter, C. D. Bader, R. Müller, *Angewandte Chemie International Edition* **2021**, *60*, 8081–8088.

[87] J. J. Hug, D. Krug, R. Müller, *Nat Rev Chem* **2020**, *4*, 172–193.

[88] L. R. Snyder, J. J. Kirkland, J. W. Dolan, *Introduction to Modern Liquid Chromatography, 3rd Edition*, Wiley, **2009**.

[89] G. L. Glish, R. W. Vachet, *Nat Rev Drug Discov* **2003**, *2*, 140–150.

[90] N. S. Cortina, D. Krug, A. Plaza, O. Revermann, R. Müller, *Angewandte Chemie International Edition in English* **2012**, *51*, 811–816.

[91] F. Panter, D. Krug, R. Müller, *ACS Chem Biol* **2019**, *14*, 88–98.

[92] “https://www.vecteezy.com/free-vector/petri-dish ,” **n.d.**

Chapter 2

-

Thiamyxins: Structure and Biosynthesis of Myxobacterial RNA-Virus Inhibitors

Patrick A. Haack[a]+, Kirsten Harmrolfs[a]+, Chantal D. Bader[a]+, Ronald Garcia[a], Antonia P. Gunesch[b], Sibylle Haid[b], Alexander Popoff[a], Alexander Voltz[a], Heeyoung Kim[c], Ralf Bartenschlager[c], Thomas Pietschmann[b], and Rolf Müller[a]*

Previously published in: Angewandte Chemie **2022**

DOI: 10.1002/anie.202212946

[a] P.A. Haack, Dr. K. Harmrolfs, Dr. C.D. Bader, Dr. R. Garcia, Dr. A. Popoff, Alexander Voltz, Prof. Dr. Dr. R. Müller
Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI) and Department of Pharmacy, Saarland University, Saarbrücken, Germany
German center for infection research (DZIF), Braunschweig, Germany
*E-Mail: Rolf.Mueller@helmholtz-hips.de

[b] Dr. A. P. Gunesch, Dr. S. Haid, Prof. Dr. T. Pietschmann
Institute of Experimental Virology, TWINCORE, Centre for Experimental and Clinical Infection Research, a joint venture between the Medical School Hannover (MHH) and the Helmholtz Centre for Infection Research (HZI), Hannover, Germany, German Center for Infection Research, Hannover-Braunschweig Partner Site, and Cluster of Excellence RESIST (EXC 2155)

[c] Dr. H. Kim, Prof. Dr. R. Bartenschlager
Department of Infectious Diseases, Molecular Virology, Heidelberg University, German Center for Infection Research, Heidelberg Partner Site and Division of Virus-Associated Carcinogenesis, German Cancer Research Center (DKFZ), German Center for Infection Research (DZIF), Heidelberg, Germany

[+] These authors contributed equally. * Corresponding author.

Author Contributions and Acknowledgments

Patrick A. Haack

This author contributed to the conception of this study, designed and performed experiments, and interpreted results. Cultivation of the producer strain, isolation of the thiamyxins, as well as identification of thiamyxin A was conducted by the author. Furthermore, the author performed *in-silico* analysis of the gene cluster, conceived the biosynthesis hypothesis and performed supporting feeding experiments. The author significantly contributed to the conception and writing of this manuscript.

Contributions by Others

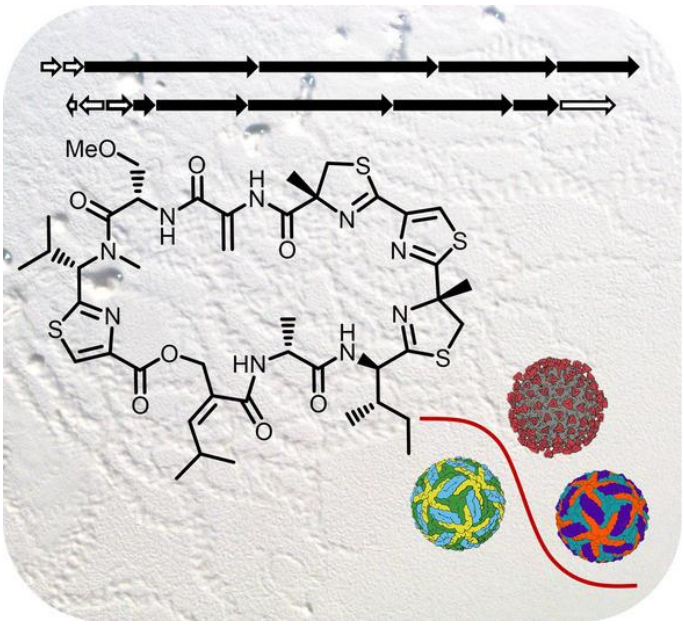
Kirsten Harmrolfs significantly contributed to the conception of this study, identified the thiamyxins as targets for isolation, performed NMR based structure elucidation of the compounds and contributed to the conception and writing of the manuscript. Chantal D. Bader contributed to the conception of this study, performed NMR based structure elucidation of the compounds and significantly contributed to the conception and writing of this manuscript. Ronald Garcia isolated, cultivated and classified the strain MCy9487. Antonia P. Gunesch, Sibylle Haid and Thomas Pietschmann performed and analyzed the antiviral assays against the corona model hCoV-229E-luc. Alexander Popoff and Alexander Voltz contributed to the isolation and NMR measurements of the thiamyxins. Heeyoung Kim and Ralf Bartenschlager performed and analyzed the antiviral assays against Dengue and Zika models DENV-R2A and ZIKV-H/PF/2019. Rolf Müller contributed to the conception and supervision of this study, as well as editing and proofreading of the manuscript.

Acknowledgments

The authors thank Silke Wenzel for primary analysis of the BGC, Nestor Zaburannyi for genome assembly, Joachim Hug for submitting the cluster to GenBank, Viktoria Schmitt, Alexandra Amann and Jennifer Herrmann for the antimicrobial and cytotoxicity assays, Susanne Kirsch-Dahmen for arranging the antiviral screenings, Joachim Wink, Azam Moradi, Heinrich Steinmetz and Stephan Hüttel for assistance with fermentation, Anne Kühnel and Christina Grethe for excellent technical assistance in the hCoV-229E-luc reporter virus assays, Volker Thiel for kind gift of the hCoV-229-luc virus, Charles M. Rice for provision of the Huh-7.5 cell line and Cathrin Spröer and Jörg Overmann for genome sequencing of MCy9487. T.P. received funds from the state of Lower Saxony within the frame of the rapid Coronavirus response programme (14-76103-184 CORONA-13/20) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2155 – project number 390874280.

2.1 Abstract

During our search for novel myxobacterial natural products, we discovered the thiamyxins: thiazole- and thiazoline-rich non-ribosomal peptide-polyketide hybrids with potent antiviral activity. We isolated four congeners of this unprecedented natural product family with the non-cyclized thiamyxin D fused to a glycerol unit at the C-terminus. Alongside their structure elucidation, we present a concise biosynthesis model based on biosynthetic gene cluster analysis and isotopically labelled precursor feeding. We report incorporation of a 2-(hydroxymethyl)-4-methylpent-3-enoic acid moiety by a GCN5-related N-acetyltransferase-like decarboxylase domain featuring polyketide synthase. The thiamyxins show potent inhibition of RNA-viruses in cell culture models of corona, zika and dengue virus infection. Their potency up to a half maximal inhibitory concentration of 560 nM combined with milder cytotoxic effects on human cell lines indicate a potential for further development of the thiamyxins.



2.2 Introduction

The severe-acute-respiratory-syndrome-coronavirus-2 (SARS-CoV-2) that was first identified in December 2019 in Wuhan, China, is the cause for the ongoing COVID-19 pandemic that is challenging global health administrations in an unprecedented way.[1] Due to the highly infectious nature of this pathogen, great efforts are being undertaken in developing medications to reduce and stop the spread of SARS-CoV-2.[3] However, it is by far not the only human pathogenic RNA virus with high economic and social burden.[4] Dengue fever, caused by infection with the dengue virus, on the one hand is a leading cause of severe illness and death in some Asian and Latin American countries.[5] Zika virus, on the other hand was found to be linked to congenital malformations in newborns and miscarriages due to intrauterine infection of the fetus with the virus.[6] Considering the recent pandemic, the possibility of future global health crises caused by viruses, but also ineffective treatment options against a variety of other viral diseases, it is of

the utmost importance to keep identifying new compounds with antiviral activities as potential basis for novel drug candidates.

Bacterial natural products (NPs) represent an ubiquitous source of novel chemistry, coming with a diverse range of biological activities.[7] They hereby form a great repository of finding promising lead structures in the search for new therapeutics.[8] Myxobacteria are predatory Gram–negative bacteria that have been intensely studied due to their biological uniqueness and extremely large genomes which are extraordinary rich in biosynthetic gene clusters (BGC) encoding diverse NPs.[9] An important aspect of compound discovery from myxobacteria is that the taxonomic distance correlates with distance in chemical diversity, making taxonomically distant strains the most promising source for the discovery of yet undescribed NP families.[10] As a result of this finding, we constantly aim to explore novel myxobacterial strains, such as *Myxococcaceae* strain MCy9487, towards their biosynthetic potential to produce biologically active NPs. In this manuscript, we report the structure of a family of unprecedented cyclic depsipeptides produced by this strain, which we called the thiamyxins, alongside the determination of their stereochemistry and biosynthesis, as well as their intriguing activities against RNA viruses.

2.3 Results and Discussion

MCy9487 is a myxobacterial strain of the family *Myxococcaceae* isolated from soil of the Saarland University campus. Phylogenetic analysis based on 16S RNA revealed the strain's distinct position within the *Myxococcus-Pyxidicoccus- Corallococcus* clade. Due to its yet unclassified genus and the interesting inhibitory activity profile (Table S1) of the crude extract, it was prioritized for chemical in depth-analysis. High performance liquid chromatography – high resolution mass spectrometry (HPLC-*hr*MS) analysis of the MCy9487 crude extract and nuclear magnetic resonance spectroscopy (NMR) analysis of pre-fractionated crude material, revealed a family of at least four putatively novel NPs. *hr*MS isotope pattern and NMR signal shifts indicated peptide type structures with some striking characteristics including thiazole and thiazoline substructures (Figure 1A). The thiazole scaffold is a key structure in drug discovery and medicinal chemistry, as it has been correlated with a broad variety of biological activities such as antiviral, anticancer and antibacterial activities.[11] This finding highly resembles observations from nature: Thiazoline and thiazole containing NPs were found to exhibit promising anticancer and antimicrobial properties as exemplified by the cyclic cyanobacterial NPs patellamide[2a], largazole[2d] and apratoxin[2e] or the myxobacterial NPs myxothiazol[2b] and thiangazole[2c] (Figure 1B and C). These examples made the thiamyxins promising targets for isolation.[2c,12] An optimized production and purification process using liquid/liquid partitioning and semi-preparative LC-MS led to the isolation of 1.5 – 2 mg each of four

different congeners belonging to the thiamyxin family (thiamyxin A-D, Figure 2). Isolated yields were < 0.2 mg/L for thiamyxin A, 0.6 mg/L for thiamyxin B, 1.4 mg/L for thiamyxin C and 1.6 mg/L for thiamyxin D. The structures of the isolated compounds were elucidated subsequently using a combination of NMR, high resolution (hr) MS and detailed configurational analysis, prior to evaluation of their biological activities against a broader panel of pathogens, including a panel of human pathogenic viruses.

Thiamyxin A and B were assigned a molecular formula of $C_{43}H_{59}N_9O_8S_4$ based on *hr*MS data. Interpretation of the 1D and 2D NMR spectra, alongside with their characteristic MS2 fragmentation pattern, revealed all thiamyxins to consist of the following peptide sequence: 2-(hydroxymethyl)-4-methylpent-3-enoic acid(HMMP)-(Ala)-(Ile)-Methyl-thiazoline(Me-thiazolin)-Thiazole-Me-Thiazoline-Dehydro-alanine(Dh-Ala)-O-Methylserine(O-Me-Ser)-N-Methylvaline(N-Me-Val)-Thiazole.

The occurrence of two thiazoline and two thiazole moieties was recognizable and eponymous for the thiamyxins (see Figure 2). Their consistent peptide sequence and exact mass but different retention time

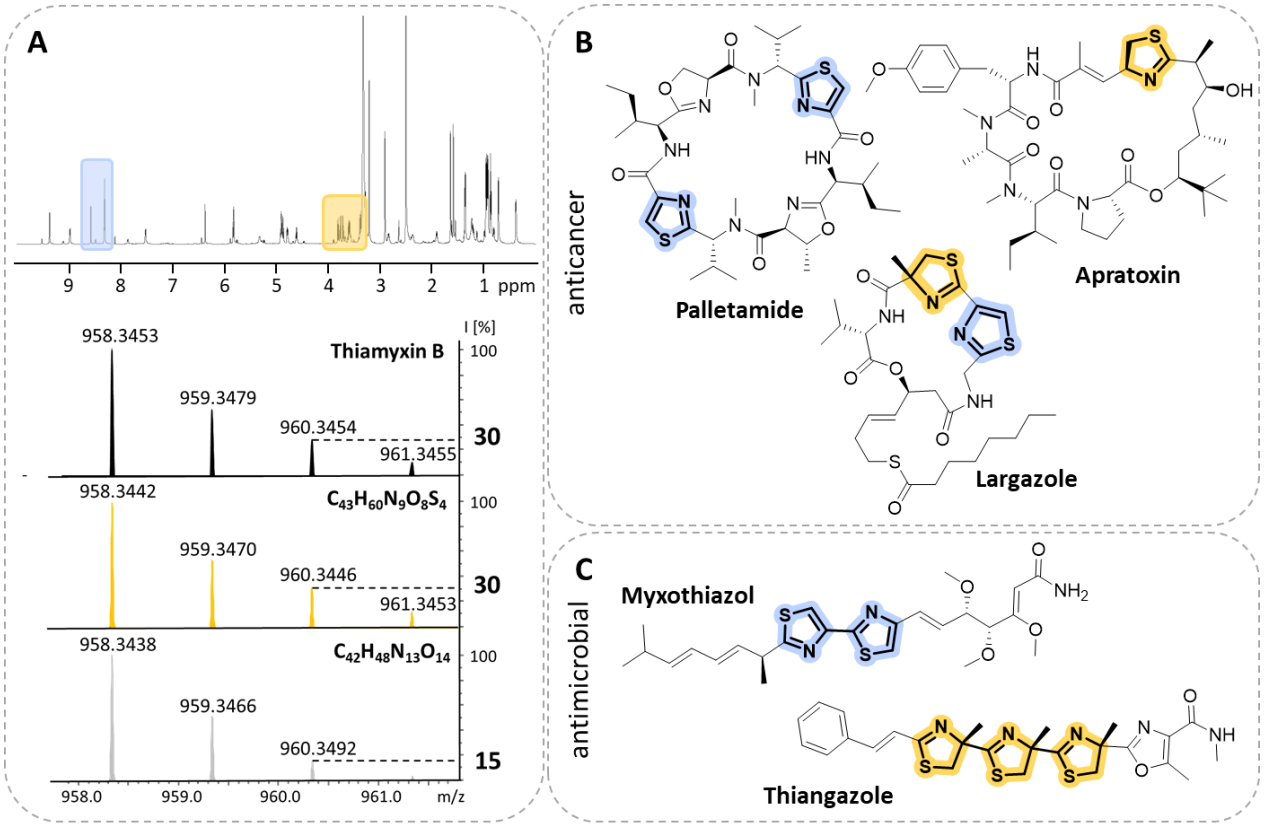


Figure 1: A: Characteristic chemical shifts (top, ^{1}H spectrum) and isotope pattern (bottom, mass spectrum) indicating the presence of thiazole and thiazoline in the thiamyxins. Mass spectra (top to bottom): measured spectrum of thiamyxin B (black), calculated spectrum for $C_{43}H_{60}N_9O_8S_4$ (yellow) and $C_{42}H_{48}N_{13}O_{14}$ (grey) as an example for a possible sum formulae highlighting the intensity shift of the second isotope peak caused by 34S. B: Natural products from cyanobacteria and C: myxobacteria comprising thiazoline (yellow) and thiazole (blue) units with their most prominent biological activity.[2]

on HPLC indicate thiamyxin A **1** and B **2** to be diastereomers. Specific fragment connectivity was obtained for all derivatives by detailed analysis of homonuclear and heteronuclear 2D NMR data (see SI). The characteristic chemical shift of the HMMP methylene in **1** at δ_H 4.90/4.86 and δ_C 67.0 ppm, alongside with its HMBC correlation to the thiazole carboxy function (δ_C 160.5 ppm) implies cyclisation between the C-terminal thiazole and the HMMP primary alcohol. The HMBC-correlation in **2** (δ_H 4.81/4.92 and δ_C 65.9 ppm to δ_C 160.4 ppm) is consistent with the finding for **1**. In contrary, thiamyxin C **3** shows a shielded chemical shift at this position of δ_H 4.00 and δ_C 63.1 ppm. This signal, in contrast to the corresponding signal in **1** and **2**, does not show any scalar coupling, indicating free rotatability of the HMMP methylene (Figure 1 SI). Together with the assigned molecular formula ($C_{43}H_{61}N_9O_9S_4$), thiamyxin C **3** was determined to be the ring open congener of **1** and **2**. The molecular mass of thiamyxin D **4** is further increased by $C_3H_6O_2$ when compared to **3** and the chemical shift of the HMMP methylene only shows a deviation of 0.1 p pm for δ_C and an exactly matched δ_H compared to **3**, affirming that thiamyxin D also belongs to the

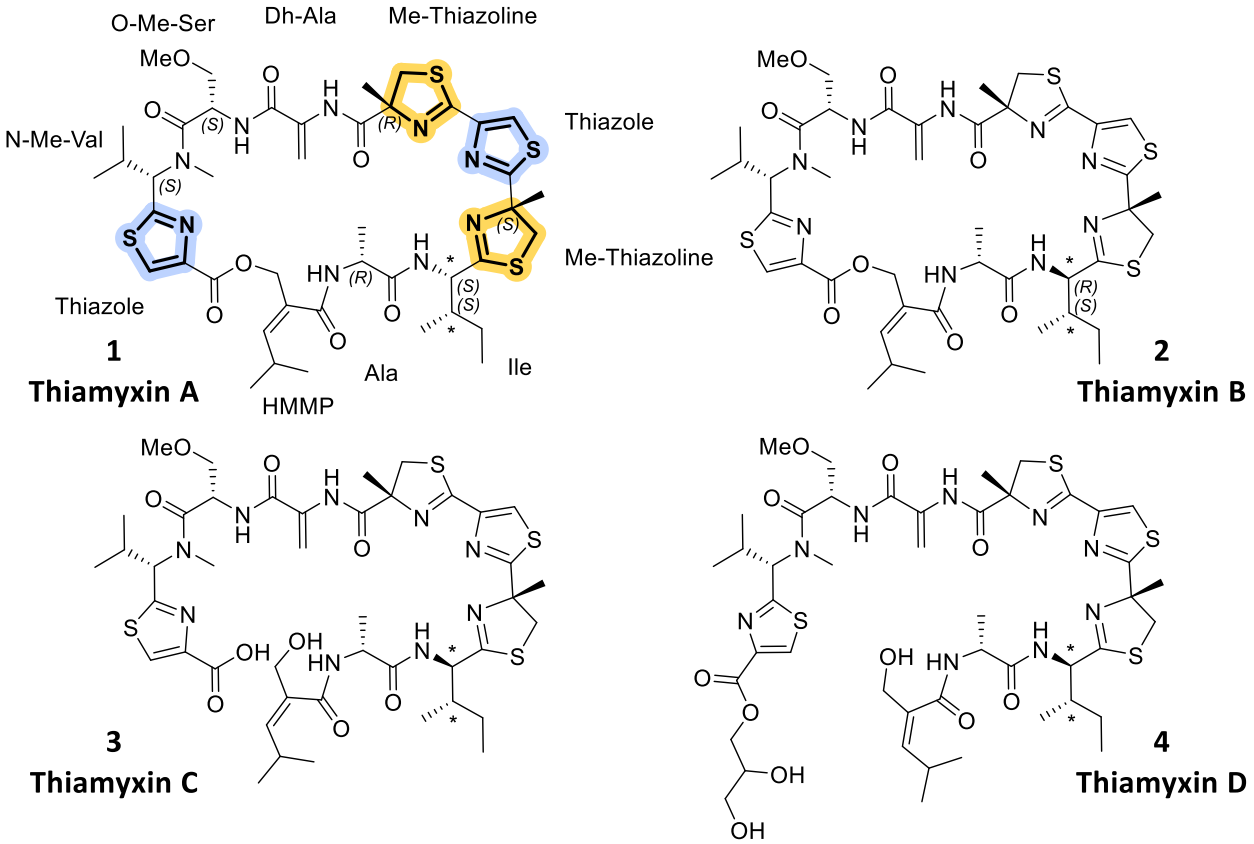


Figure 2: Chemical structures of the four congeners of the thiamyxin family (A-D). N-Me-Val: N-Methyl-Valine, O-Me-Ser: O-Methyl-Serine, Dh-Ala: Dehydroalanine, Me-Thiazoline: Methylthiazoline, Ala: Alanine, HMMP: 2-(hydroxymethyl)-4-methylpent-3-enoic acid. *Proposed stereochemistry based on observed chemical shifts and coupling constants, prevalence of L-Ile and D-*allo*-Ile, as well as the presence of an epimerization domain in module 4 and 5.

open chain derivatives. Thorough analysis of the 1D and 2D NMR spectra (Table 4-7 SI) finally revealed, that this derivative features an additional glycerol unit attached to the C-terminal carboxyl function (Figure 2 SI).

The stereochemical configuration of the individual aminoacids was elucidated using Marfey's method.[13] The Ala and O-Me-Ser stereocenters were assigned as *R* and *S,* respectively, by comparing the hydrolyzed thiamyxins to commercially available references (Table 1 SI). The configuration of the Me-thiazoline moieties were assigned by comparison to thiangazole, wherein all Me-thiazoline units previously were described to be derived from *R*-configured 2-Me-Cysteine.[2c] Chiral HPLC retention time comparison of the derivatized hydrolysis products of thiangazole and thiamyxin B showed the methyl-cysteine peaks with resembling retention times (Figure 10 SI), identifiying the Me-thiazoline stereocenters in the thiamyxins as *R*-configured as well. We observed demethylation of N-Me-Val under acidic conditions, wherefore the emerging Val was compared to a Val standard to assign the stereochemistry of this amino acid (see Figure 7 SI). The stereocenter of N-Me-Val was hereby assigned as *S*-configured. Additional NMR signal sets were detected for both cyclic derivatives **1** and **2**, in particular located surrounding the N-Me-Val signal sets. The ratio to the main signal set was constant (ca. 1:5) in both derivatives and in different isolated batches, indicating two conformational isomers even though NMR experiments at different temperatures and solvents did not yield a significant change of the conformer ratio (see Figure 5 and 6 SI). The Ile stereocenter was found to racemize during hydrolysis, which could not be prevented by optimizing the conditions. Similar effects were observed previously for stereocenters adjacent to thiazoline moieties in the bottromycin biosynthesis and the synthesis of peptide thiazolines. We believe that racemization is promoted by the neighboring Me-thiazoline in the thiamyxins.[14] We therefore take into account *in-silico* analysis of the biosynthetic domains for assignment of the Ile stereocenter. The presence of an epimerization domain in the Ile-incorporating module M4 and the C-domain of the following module M5, which is predicted to be a DLC domain, led to the conclusion that the Ile stereocenter is *R*-configured (see below). After careful analysis of the 2D NMR data, the corresponding signals revealed that thiamyxins **1** and **2** are diastereomers. Configurational assignment was achieved based on observed shift and coupling constant differences of isoleucine vs. allo-isoleucin (Figure 3 SI) [15]. Based on this analysis, thiamyxin A (**1**) and thiamyxin B (**2**) are epimers in the 2-Ile stereocenter. According to the high prevalence of L-Ile in natural products, we assume thiamyxin A (**1**) to incorporate L-isoleucine and thiamyxin B (**2**) D-*allo*-isoleucine, respectively. The open chain derivative **3** turned out to be a mixture of both isomers (ca. S:R 1:2 according to NMR) with exactly matching retention time on HPLC. The same holds true for thiamyxin D (**4**).

AntiSMASH analysis of the MCy9487 genome enabled identification of a PKS-NRPS hybrid gene cluster capable of thiamyxin biosynthesis (Figure 3). It consists of two PKS and nine NRPS modules, which are encoded on nine genes (thiA-thiH), as well as a cytochrome P450 dependent enzyme and a thioesterase domain encoded on thiI and thiJ, respectively. During bioinformatic analysis of the cluster we also found a fragmented but similar biosynthetic gene cluster in Corallococcus terminator, previously identified by Livingstone et al.[16] Modules 3 - 12 largely follow textbook NRPS biosynthesis logic.[17] The initiation of the biosynthesis, as proposed for modules 1 and 2 is unusual for PKS-NRPS systems but has been described in similar fashion previously for loading of isovaleryl- and isobutyryl-CoA by the GCN5-related N-acetyltransferase-like decarboxylase (GNAT) domain. A detailed description of the gene cluster organization, including all genes of the BGC and their closest homologues, can be found in the SI. As all attempts to genetically manipulate strain MCy9487 failed and we could not identify a genetically manipulable alternative producer, we propose the following biosynthesis model based on in-silico analysis of the BGC supported by feeding experiments with isotope labelled precursors:

The polyketide biosynthesis in thiamyxin is initiated by a GNAT domain. These domains have been shown to initiate polyketide biosynthesis by starter unit selection and decarboxylation, for example in the biosynthesis of the cytostatic polyketide gephyronic acid, as well as myxovirescin, rhizoxin, curacin and pederin.[18] The formal starter unit of the thiamyxin biosynthesis is isobutyryl-CoA, which is also the starter of the gephyronic acid biosynthesis. Isobutyryl-CoA can be derived either from valine or dimethylmalonyl-CoA. In the case of gephyronic acid dimethylmalonyl-ACP is decarboxylated to generate the isobutyryl starter unit.[18a] Analysis of the conserved residues of the thiamyxin GNAT domain confirmed the presence of arginine and threonine required for the decarboxylation. Analysis of the feeding experiment employing L-Methionine-(methyl-^{13}C) indicated up to seven methyl incorporations into thiamyxin C, whereas feeding with L-Valine-d_8 resulted in only one incorporation which is explained by the N-Methyl-Valine incorporated by module 11. Thus, no second valine seems to be incorporated as would be expected if isovaleryl-CoA formed from valine was the starter moiety. When feeding Methionine-(methyl-^{13}C), five of the observed incorporations can be accounted for by methyltransferases in modules 2, 6, 8, 10 and 11. To account for the remaining two methyl incorporations, we propose, that malonyl-CoA is bis-methylated to form dimethylmalonyl-CoA. A conserved domain search of the *thiA* gene, revealed a dimerization domain directly upstream of the GNAT domain. This dimerization domain is commonly found in methyl transferases.[19] We propose that the dimerization domain in combination with one of the methyltransferase domains present in the cluster is responsible for bis-methylating malonyl-CoA after it is loaded to the ACP in module 1 and thus this is the mechanism that forms the isobutyryl starter moiety.

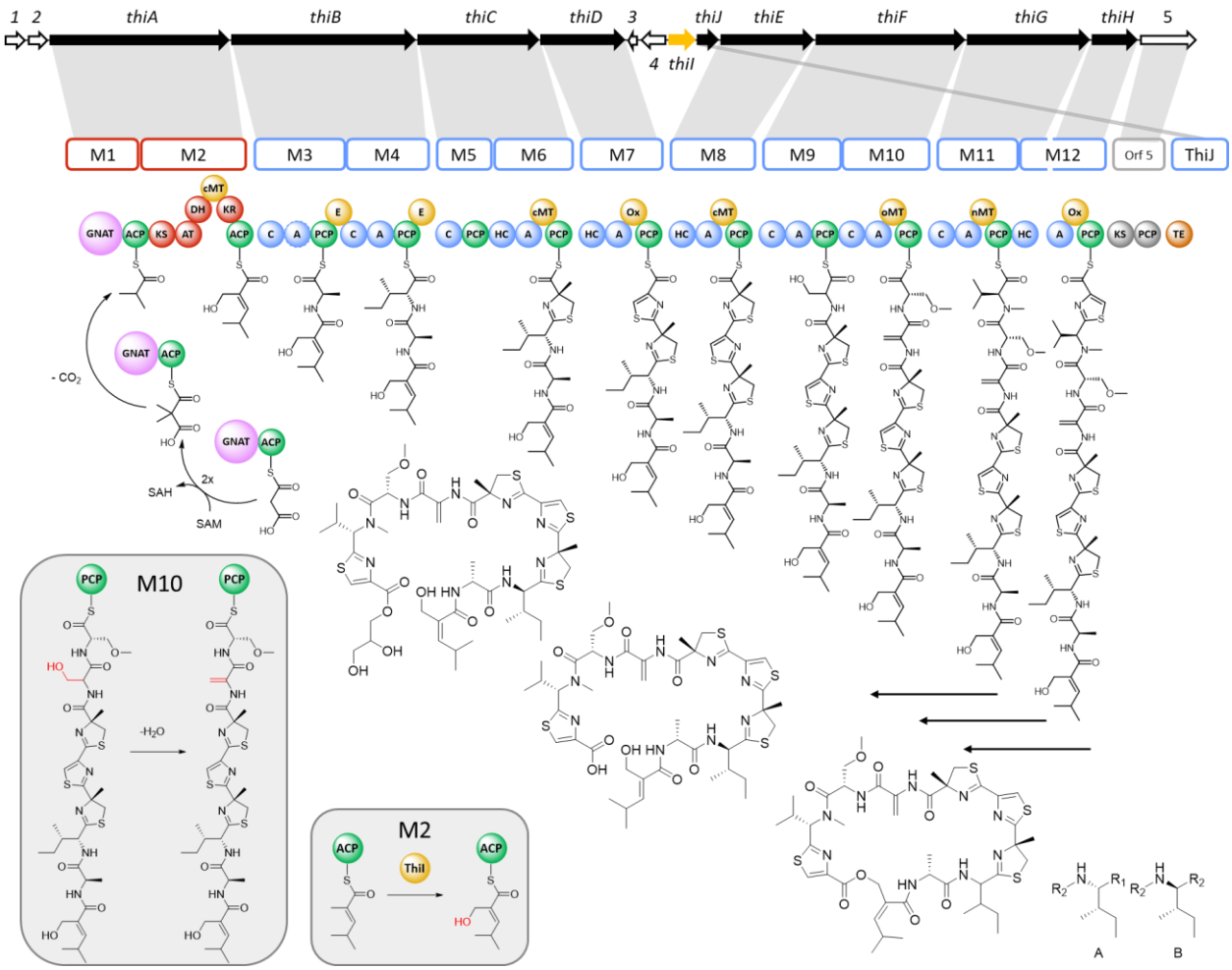


Figure 3: Proposed biosynthetic pathway for the thiamyxins (Map not drawn to scale). Core PKS modules are marked in red and core NRPS modules in blue, epimerization, methylation and oxidation domains are marked yellow and ACP and PCP domains in green. The thioesterase is shown in orange and the GNAT domain in pink. Modules proposed to be non-functional are marked in grey. The genes involved in the thiamyxin biosynthesis are marked in black and named *thi*A-J. The remaining genes with unknown or unassigned function are shown blank and named ORF1-5. The hydroxylation by ThiI is shown in in the right box and the water elimination of Serine to form Dehydroalanine (module 10) in the left box. Gene cluster color code: NRPS genes (blue), PKS genes (red). SAM = S-adenosyl methionine; SAH = S-adenosyl homocysteine.

Dimethylmalonyl-ACP is then decarboxylated by the GNAT and transferred to module 2. [18a] Module 2 extends the starter unit by one malonyl unit which we propose is subsequently methylated at the α-carbon by the cMT domain located in the same module, analogous to the gephyronic acid biosynthesis. [18c] This methyl group is then hydroxylated by the CyP450 encoded on *thiI*. This hydroxylation is required to take place on the assembly line as the resulting hydroxyl-group is necessary for final cyclisation by the type I thioesterase located on *thiJ*. Modules 3 and 4 contain epimerization domains, which indicates that these incorporate D-Ala and D-Ile. Module 5 consists only of a C and a PCP domain and does not incorporate any building block, although the C domain seems to be functional based on analysis of its active site residues

(see SI). The C domain is annotated as a DlC- domain, which are responsible for forming amide bonds with D-aminoacids. As the previously incorporated aminoacid is D-Ile, we propose that this C-domain has a transfer function that assists in loading of the intermediate onto the PCP-domain of the next module 6. In module 6 a heterocyclisation domain (HC) catalyzes the first cysteine incorporation followed by cyclisation and methylation to form Me-thiazoline as previously described in the bacillamide E biosynthesis.[20] The methyl function is introduced by another cMT domain found in the same module. The following module (7) introduces a thiazole and contains an oxidation domain that oxidizes the formed thiazoline introducing the double bond between the four and five position. Module 8 incorporates another Me-thiazoline, comparable to module 6. Modules 9 and 10 each introduce serine into the nascent molecule as confirmed by feeding experiments (SI). The serine incorporated in module 9, is dehydrated to Dh-Ala in the final product. A new clade of C domains, that are associated with dehydration reactions in NRPS assembly lines have recently been reported.[21] The condensation domain of module 10 was identified as belonging to this new clade of C domains by phylogenetic analysis. (See SI) We therefore propose that the C domain of module 10 facilitates the dehydration of the serine incorporated by module 9 to form dehydroalanine and also incorporates another serine. The second serine is subsequently O-methylated by an oMT domain. Module 11 introduces a valine which is N-methylated by an nMT domain in the same module. Module 12 is split on two genes: *thiG* and *thiH*. This module introduces a thiazole, analogous to module 7. The TE domain encoded on *thiJ*, finally releases the molecule from the assembly line by cyclic condensation with the hydroxyl function installed by the CYP-450 in module 2. The derivative thiamyxin D likely is a shunt product, created during the release process. We propose that the type I TE domain promiscuously accepts glycerol as substrate next to the HMMP-hydroxyl function required for the intramolecular cyclisation reaction. A similar glycerol-ester formation has been previously observed in tubulysin biosynthesis.[23] Thiamyxin C might also be created in a similar fashion by promiscuity for H_2O but could also be an artefact from the purification process. Downstream of the TE Domain a KS and PCP di-domain are encoded on orf5. Analysis of their active site residues revealed them to be non-functional, which is in line with our biosynthesis hypothesis, as there are no non-assigned biosynthetic reactions remaining to form the mature thiamyxins.

After determination of their chemical structure, the thiamyxins were evaluated against a broad panel of bacterial, fungal and viral pathogens, alongside with their antiproliferative effects on human cell lines (Table 1, Fig. 17-22 SI, Table 3 SI). A detailed description of the underlying assays and test organisms against which the thiamyxins were found inactive or only presented weak biological activities can be found in the SI. Against the antimicrobial test panel, the thiamyxins only presented weak activity at a minimal inhibitory concentration (MIC) of 32-64 µg/mL against two fungal test organisms (*Candida albicans* and *Mucor*

hiemalis) and two Gram-positive pathogens (*Bacillus subtilis* and *Micrococcus luteus*) (SI-Table 3). Thiamyxin D did not show effects in the tested concentrations.

Table 1: Antiviral and antiproliferative activities of the thiamyxins. Half maximal inhibitory concentrations against three RNA viruses (IC50) values determined simultaneously to half maximal cytotoxic concentrations (CC50) in infected cells. [1]positive control remdesivir IC50 = 5.6 nM [2][25] positive control ribavirin IC50 = 2.3 µM [3]positive control ribavirin IC50 = 2.5 µM nd = not determined

Test organism		**IC_{50} and corresponding CC_{50} [µM]**			
		Thiamyxin A	Thiamyxin B	Thiamyxin C	Thiamyxin D
hCov-229E-luc[1]	IC_{50}	2.47	2.39	>10	>10
Huh-7.5 Fluc infected	CC_{50}	>10	>10	>10	>10
DENV-R2A[2]	IC_{50}	nd	0.56	14.56	nd
Huh-7 infected with DENV-R2A	CC_{50}	nd	2.25	>50	nd
ZIKV-H/PF/2013[3]	IC_{50}	nd	1.07	>15	nd
Huh-7 infected with ZIKV-H/PF/2013	CC_{50}	nd	2.25	>50	nd

Thiamyxin B and C which are produced in higher amounts (0.6 and 1.4 mg/L isolated yield) and represent one cyclized and one open chain analogue were evaluated for activity in cell-based assays against the human pathogenic corona virus hCoV-229E and two representatives of the flavivirus genus, Dengue virus and Zika virus, for which we used the easy to measure reporter viruses hCoV-229E-luc, DENV-R2A and ZIKV-H/PF/2019, respectively.[24] These assays allow simultaneous determination of antiviral activity and cytotoxic effects on the chosen cell line (results see table 1). Due to their comparably low production rates of <0.2 mg/L isolated yield, thiamyxin A and D were only evaluated against a selected panel of test organisms.

The cyclized congeners thiamyxin A and B were found to effectively inhibit hCoV-229E-luc with a half maximal inhibitory concentration (IC_{50}) in the low micromolar range. This activity was highly decreased for the open chain analogues thiamyxin C and D, which showed IC_{50} values above our tested concentrations of around 20 µM.

This trend is also observed for DENV-R2A and ZIKV-H/PF/2019, where thiamyxin B was found to inhibit the two viruses with an IC_{50} of 560 nM and 1.07 µM, respectively. In constrast, thiamyxin C inhibited DENV-R2A with an IC_{50} of 14.56 µM and was completely inactive against ZIKV-H/PF/2019 (Table 1).

As depicted in Figure 4, we observed a potential application window for the two cyclized thiamyxins A and B in the HCov-229E assay, where we can see a clear separation of antiviral and cytotoxic activity with a more than 5-fold difference between IC_{50} and CC_{50}. This indicates a distinct mode of action for the antiviral activity compared to the cytotoxic effects. Among the three viral pathogens, DENV-R2A shows best inhibition with a 4-fold lower IC_{50} -value compared to HCov-229E-luc for thiamyxin B. In this assay,

however, the determined application window is smaller compared to HCoV-229E with only a 4-fold difference between IC_{50} and CC_{50}. The extend of cytotoxic effects of the thiamyxins therefore seem to be cell line dependent.

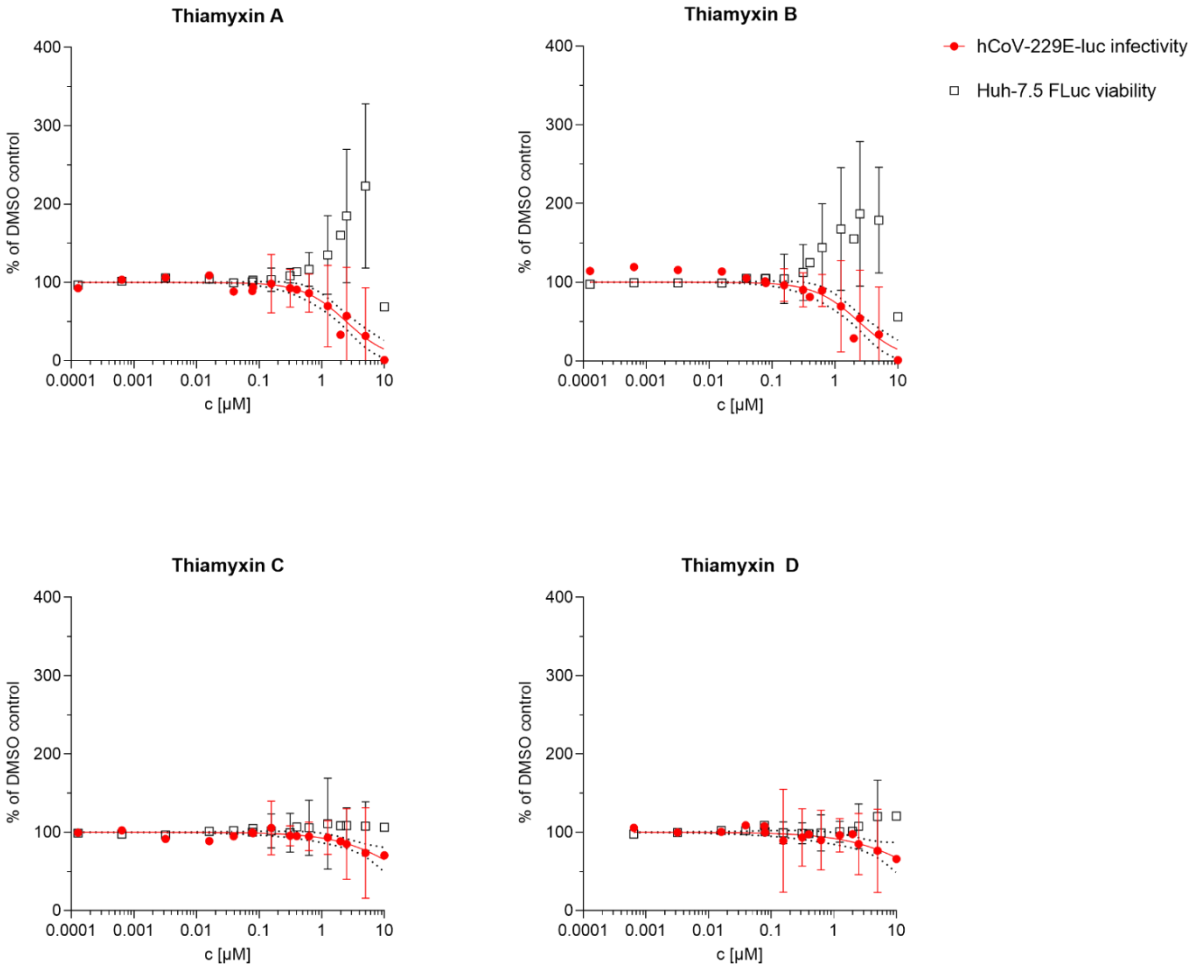


Figure 4: Antiviral activity of thiamyxin A-D against HCoV-229E-luc (red) when simultaneously determining their effect on Huh-7.5 Fluc cells (black). Renilla luciferase serves as reporter for viral load, firefly luciferase for cell viability. Measurements performed in technical duplicates of four independent biological experiments. Non-linear regression curves (red) are given with 95% confidence interval (black dots). Increase in cell viability over 100% caused by reduction in viral load.

2.4 Conclusion

In this study, we present the thiamyxins, a family of cyclic thiazole- and thiazoline rich non-ribosomal peptides, which we isolated from the myxobacterial strain MCy9487 discovered at the Saarland University campus. A combination of NMR analyses and detailed stereochemical configurational studies supported by bioinformatics analysis of the BGC responsible for their formation led to their complete stereochemical assignment. We developed a concise biosynthesis model for the thiamyxins, including the formation of their unusual isobutyryl starter unit, which presumably is formed in a similar fashion to gephyronic acid. Based on the underlying biosynthetic logic, thiamyxin B seems to be the main product of the thiamyxin BGC assembly line, which is also reflected in its higher production rate compared to thiamyxin A. The two open chain analogues thiamyxin C and D likely are shunt products, generated due to promiscuity of the release enzyme for gylcerol and water in competition with the preferred intramolecular HMMP-hydroxyl-function. Thiamyxin B also shows the most potent biological activity among the four characterized derivatives. It shows significant activity against RNA-viruses but also shows cytotoxic effects, which however seem to be cell-line dependent and can well be separated from the observed antiviral activity. Due to the comparably low production rates of the thiamyxins, biotechnological production optimization, biosynthetic engineering approaches or development of a total synthesis route would be of great interest allowing their in-depth evaluation against a broader panel of human pathogenic viruses and studying their cytotoxicity-profile on non-infected human cells. Those approaches would furthermore allow access to further derivatives, giving insights into the thiamyxins' structure-activity-relationship and evaluate pharmacokinetic properties of this interesting NP-class. In this study we could already observe significantly lower antiviral activities for the open chain analogue thiamyxin C compared to the two cyclic thiamyins A and B, but the data still suggest some residual antiviral activity at higher µM concentrations. This finding is particularly interesting, as most cyclic NRPs almost completely lose affinity to their target because their 3D structure is altered by opening the macrocycle. The resembling biological activities of thiamyxin A and B are unexpected, as their proposed incorporation of D- vs L-isoleucine results in a distinct conformation of the two macrocycles (see SI). Determining the thiamyxins' antiviral target – besides developing a total synthesis route for them – could provide deeper insights into the thiamyxins' pharmacophore and help to understand these observations.

In summary, the thiamyxins are a structurally unique class of NPs showing a potential application window as broad spectrum antivirals targeting human pathogenic RNA viruses and their desription herein paves the way for their further investigation.

2.5 References

[1] a) "WHO Coronavirus Disease (COVID-19) Dashboard | WHO Coronavirus Disease (COVID-19) Dashboard", to be found under https://covid19.who.int/, 2020; b) A. E. Gorbalenya, S. C. Baker, R. S. Baric, Raoul J. de Groot, C. Drosten, A. A. Gulyaeva, B. L. Haagmans, C. Lauber, A. M. Leontovich, B. W. Neuman, D. Penzar, St. Perlman, L. L. M. Poon, D. V. Samborskiy, I. A. Sidorov, I. Sola, J. Ziebuhr, *Nat. Microbiol.* **2020**, 5, 536–544.

[2] a) C. M. Ireland, A. R. Durso, R. A. Newman, M. P. Hacker, *J. Org. Chem*. **1982**, 47, 1807–1811; b) G. Thierbach, H. Reichenbach, *Antimicrob. Agents Chemother*. **1981**, 19, 504–507; c) B. Kunze, R. Jansen, L. Pridzun, E. Jurkiewicz, G. Hunsmann, G. Höfle, H. Reichenbach, *J Antibiot* **1993**, 46, 1752–1755; d) K. Taori, V. J. Paul, H. Luesch, *J. Am. Chem. Soc.* **2008**, 130, 1806–1807; e) H. Luesch, W. Y. Yoshida, R. E. Moore, V. J. Paul, T. H. Corbett, *J. Am. Chem. Soc.* **2001**, 123, 5418–5423.

[3] a) J. G. Rizk, D. N. Forthal, K. Kalantar-Zadeh, M. R. Mehra, C. J. Lavie, Y. Rizk, J. P. Pfeiffer, J. C. Lewin, *Drug Discov. Today* **2020**, 26, 593–603; b) J. H. Beigel, K. M. Tomashek, L. E. Dodd, A. K. Mehta, B. S. Zingman, A. C. Kalil, E. Hohmann, H. Y. Chu, A. Luetkemeyer, S. Kline, D. Lopez de Castilla, R. W. Finberg, K. Dierberg, V. Tapson, L. Hsieh, T. F. Patterson, R. Paredes, D. A. Sweeney, W. R. Short, G. Touloumi, D. C. Lye, N. Ohmagari, M.-D. Oh, G. M. Ruiz-Palacios, T. Benfield, G. Fätkenheuer, M. G. Kortepeter, R. L. Atmar, C. B. Creech, J. Lundgren, A. G. Babiker, S. Pett, J. D. Neaton, T. H. Burgess, T. Bonnett, M. Green, M. Makowski, A. Osinusi, S. Nayak, H. C. Lane, *N. Engl. J. Med*. **2020**, 383, 1813–1826; c) M. Jeyanathan, S. Afkhami, F. Smaill, M. S. Miller, B. D. Lichty, Z. Xing, *Nat. Rev. Immunol.* **2020**, 20, 615–632; d) S. P. Kaur, V. Gupta, *Virus Res*. **2020**, 288, 198114.

[4] R. G. Huber, X. N. Lim, W. C. Ng, A. Y. L. Sim, H. X. Poh, Y. Shen, S. Y. Lim, K. B. Sundstrom, X. Sun, J. G. Aw, H. K. Too, P. H. Boey, A. Wilm, T. Chawla, M. M. Choy, L. Jiang, P. F. de Sessions, X. J. Loh, S. Alonso, M. Hibberd, N. Nagarajan, E. E. Ooi, P. J. Bond, O. M. Sessions, Y. Wan, *Nat Commun* **2019**, 10, 1408.

[5] K. Wellekens, A. Betrains, P. de Munter, W. Peetermans, *Acta Clin. Belg.* **2022**, 77, 436–444.

[6] L. Schuler-Faccini, M. Del Campo, A. García-Alix, L. O. Ventura, J. A. Boquett, V. van der Linden, A. Pessoa, H. van der Linden Júnior, C. V. Ventura, M. C. Leal, T. W. Kowalski, L. Rodrigues Gerzson, C. Skilhan de Almeida, L. Santi, W. O. Beys-da-Silva, A. Quincozes-Santos, J. A. Guimarães, P. P. Garcez, J. d. A. Gomes, F. S. L. Vianna, A. Da Anjos Silva, L. R. Fraga, M. T. Vieira Sanseverino, A. R. Muotri, R. Da Lopes Rosa, A. M. Abeche, C. Marcolongo-Pereira, D. O. Souza, *Front. Genet.* **2022**, 13, 758715.

[7] S. Bernardini, A. Tiezzi, V. Laghezza Masci, E. Ovidi, *Nat. Prod. Res*. **2018**, 32, 1926–1950.

[8] M. S. Butler, *J. Nat. Prod*. **2004**, 67, 2141–2153.

[9] J. Herrmann, A. A. Fayad, R. Müller, *Nat. Prod. Rep.* **2017**, 34, 135–160.

[10]T. Hoffmann, D. Krug, N. Bozkurt, S. Duddela, R. Jansen, R. Garcia, K. Gerth, H. Steinmetz, R. Müller*, Nat. Commun*. **2018**, 9, 803.

[11] A. Ayati, S. Emami, A. Asadipour, A. Shafiee, A. Foroumadi, *Eur. J. Med.* **2015**, 97, 699–718.

[12] a) W. Cai, Q.-Y. Chen, L. H. Dang, H. Luesch, *ACS Med. Chem. Lett*. **2017**, 8, 1007–1012; b) Y. In, M. Doi, M. Inoue, T. Ishida, Y. Hamada, T. Shioiri, Acta crystallographica. Section C, *Crystal structure communications* **1994**, 50 (Pt 3), 432–434;

[13] K. Harada, K. Fujii, K. Hayashi, M. Suzuki, Y. Ikai, H. Oka, *Tetrahedron Lett*. **1996**, 37, 3001–3004.

[14] a) A. Sikandar, L. Franz, S. Adam, J. Santos-Aberturas, L. Horbal, A. Luzhetskyy, A. W. Truman, O. V. Kalinina, J. Koehnke, *Nat. Chem. Biol.* **2020**, 16, 1013–1018; b) P. Wipf, P. C. Fritch, *Tetrahedron Lett.* **1994**, 35, 5397–5400;

[15] Z. J. Anderson, C. Hobson, R. Needley, L. Song, M. S. Perryman, P. Kerby, D. J. Fox, *Org. Biomol. Chem*. **2017**, 15, 9372–9378.

[16] P. G. Livingstone, O. Ingleby, S. Girdwood, A. R. Cookson, R. M. Morphew, D. E. Whitworth, *Appl. Environ. Microbiol*. **2019**, 86, e01931-19.

[17] C. T. Walsh, *Nat. Prod. Rep*. **2016**, 33, 127–135.

[18] a) M. A. Skiba, C. L. Tran, Q. Dan, A. P. Sikkema, Z. Klaver, W. H. Gerwick, D. H. Sherman, J. L. Smith, *Structure* **2020**, 28, 63-74.e4; b) L. C. Gu, T. W. Geders, B. Wang, W. H. Gerwick, K. Hakansson, J. L. Smith, D. H. Sherman, *Science* **2007**, 318, 970–974; c) M. A. Skiba, A. P. Sikkema, W. D. Fiers, W. H. Gerwick, D. H. Sherman, C. C. Aldrich, J. L. Smith, *ACS Chem. Biol*. **2016**, 11, 3319–3327;

[19] S. Lu, J. Wang, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, M. Gwadz, Di Hurwitz, G. H. Marchler, J. S. Song, N. Thanki, R. A. Yamashita, M. Yang, D. Zhang, C. Zheng, C. J. Lanczycki, A. Marchler-Bauer, *Nucleic Acids Res*. **2020**, 48, D265-D268.

[20] K. Bloudoff, C. D. Fage, M. A. Marahiel, T. M. Schmeing, *Proc. Natl. Acad. Sci. U.S.A.* **2017**, 114, 95–100.

[21] a) S. Wang, Q. Fang, Z. Lu, Y. Gao, L. Trembleau, R. Ebel, J. H. Andersen, C. Philips, S. Law, H. Deng, *Angew. Chem. Int. Ed.* **2021**, 60, 3229–3237; b) S. Wang, W. D. G. Brittain, Q. Zhang, Z. Lu, M. H. Tong, K. Wu, K. Kyeremeh, M. Jenner, Y. Yu, S. L. Cobb, H. Deng, *Nat Commun* **2022**, 13, 62; c) J. B. Patteson, C. M. Fortinez, A. T. Putz, J. Rodriguez-Rivas, L. H. Bryant, K. Adhikari, M. Weigt, T. M. Schmeing, B. Li, J. *Am. Chem. Soc.* **2022**, 144, 14057–14070;

[22] L. Kjaerulff, R. Raju, F. Panter, U. Scheid, R. Garcia, J. Herrmann, R. Müller, *Angew. Chem. Int. Ed.* **2017**, 56, 9614–9618.

[23] Y. Chai, S. Shan, K. J. Weissman, S. Hu, Y. Zhang, R. Müller, *Chem. Biol.* **2012**, 19, 361–371.

[24] S. H. E. van den Worm, K. K. Eriksson, J. C. Zevenhoven, F. Weber, R. Züst, T. Kuri, R. Dijkman, G. Chang, S. G. Siddell, E. J. Snijder, V. Thiel, A. D. Davidson, *PLoS ONE* **2012**, 7, e32857.

[25] C. D. Bader, F. Panter, R. Garcia, E. P. Tchesnokov, S. Haid, C. Walt, C. Spröer, A. F. Kiefer, M. Götte, J. Overmann, T. Pietschmann, R. Müller, *Chemistry – A European Journal* **2022**, 28, e202104484.

Thiamyxins: Structure and Biosynthesis of Myxobacterial RNA-Virus-Inhibitors

Supporting Information

Patrick A. Haack[a]+, Kirsten Harmrolfs[a]+, Chantal D. Bader[a]+, Ronald Garcia[a], Antonia P. Gunesch[b], Sibylle Haid[b], Alexander Popoff[a], Alexander Voltz[a], Heeyoung Kim[c], Ralf Bartenschlager[c], Thomas Pietschmann[b], and Rolf Müller[a]*

Previously published in: Angewandte Chemie **2022**

DOI: 10.1002/anie.202212946

[a] P.A. Haack, Dr. K. Harmrolfs, Dr. C.D. Bader, Dr. R. Garcia, Dr. A. Popoff, Alexander Voltz, Prof. Dr. Dr. R. Müller
Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI) and Department of Pharmacy, Saarland University, Saarbrücken, Germany
German center for infection research (DZIF), Braunschweig, Germany
*E-Mail: Rolf.Mueller@helmholtz-hips.de

[b] Dr. A. P. Gunesch, Dr. S. Haid, Prof. Dr. T. Pietschmann
Institute of Experimental Virology, TWINCORE, Centre for Experimental and Clinical Infection Research, a joint venture between the Medical School Hannover (MHH) and the Helmholtz Centre for Infection Research (HZI), Hannover, Germany, German Center for Infection Research, Hannover-Braunschweig Partner Site, and Cluster of Excellence RESIST (EXC 2155)

[c] Dr. H. Kim, Prof. Dr. R. Bartenschlager
Department of Infectious Diseases, Molecular Virology, Heidelberg University, German Center for Infection Research, Heidelberg Partner Site and Division of Virus-Associated Carcinogenesis, German Cancer Research Center (DKFZ), German Center for Infection Research (DZIF), Heidelberg, Germany

† These authors contributed equally. * Corresponding author.

Table of Contents

S 2.1 General Experimental Procedures and Materials

S 2.1.1 Isolation and cultivation of MCy9487

The myxobacterial strain MCy9487 was isolated at the hillside of Saarland University Campus, Saarbrücken, Germany in January 2011 from a soil sample with leaf litters and other decaying plant materials. The strain was recognized on a standard mineral salts agar isolation medium for transparent swarming and whitish fruiting bodies appearing as mounds- and horn-like shapes. Phylogenetic analysis based on 16S rRNA gene sequence revealed the strain's distinct position within the *Myxococcus-Pyxidicoccus-Corallococcus* clade which shows 98.62 – 98.68% similarity with *Corallococcus coralloides* strain $M2^T$ (GenBank accession: NR_042329) and *C. exiguus* strain Cc $e167^T$ (GenBank accession: NR_042330), respectively. The detailed and valid strain description will be published in a separate taxonomic manuscript.

The fermentation was performed in 30 L VY/2S medium (g/L, w/v: 0.5% commercial fresh Baker's yeast, 0.01% $CaCl_2$. $2H_2O$, 0.2% HEPES, 0.5% soluble starch (Roth), pH 7.2, adjusted with KOH before autoclaving) supplemented with 2% (v/v) amberlite resin XAD-16 (Sigma). The cultivation was maintained under rotary shaking condition (160 r.p.m., 30°C, 10d). At the end of cultivation, the cells and resins were harvested together by centrifugation (8,000 r.p.m. 30mins, 4°C).

S 2.1.2 Isolation of the thiamyxins

The cell pellet and resin were extracted three times with methanol and dried using a rotary evaporator. The extract was then resolved in Methanol : H_20 (60:40) and extracted with hexane. The methanol was then removed by rotary evaporation and the remaining aqueous layer was extracted with chloroform. All four compounds were found in the chloroform partition, which was then dried, resolved in methanol and the compounds were purified using LC-MS system 2.

S 2.1.3 LC-MS systems

All analytical LC-MS measurements were performed on a Dionex Ultimate 3000 RSLC system using a BEH C18, 100 x 2.1 mm, 1.7 µm dp column (Waters, Germany), coupled to a maXis 4G hr-ToF mass spectrometer (Bruker Daltonics, Germany) using the Apollo ESI source. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75 µL/min before entering the mass spectrometer. The LC-MS data was analyzed using Bruker DataAnalysis version 4.2.

LC-MS system 1a – standard analytical measurements

Separation of 1 µl sample was achieved by a linear gradient from (A) H_2O + 0.1% FA to (B) ACN + 0.1% FA at a flow rate of 600 µL/min and 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5% B, followed by an increase to 95% B in 18 min to end up with a 2 min step at 95% B before reequilibration under the initial conditions. Mass spectra were acquired in centroid mode ranging from 150 – 2500 *m/z* at a 2 Hz scan rate.

LC-MS system 1b – Marfey's Method

Separation of 1 µl sample was achieved by a gradient from (A) H_2O + 0.1% FA to (B) ACN + 0.1% FA at a flow rate of 600 µL/min and 45 °C. The gradient was as follows: Ramp in 1 min from 5% B to 10% B, in 14 min to 35% B, in 7 min to 55% B and in 3 min to 80% B. This is followed by a 1 min step at 80% B before reequilibration with the initial conditions. Mass spectra were acquired in centroid mode ranging from 250 – 3000 *m/z* at a 2 Hz scan rate.

LC-MS system 1c – adapted Marfey's Method for improved separation of isoleucins

Separation of 1 µl sample was achieved by a gradient from (A) H_2O + 0.1% FA to (B) ACN + 0.1% FA at a flow rate of 600 µL/min and 45 °C. The gradient was as follows: Ramp in 1 min from 5% B to 27% B, in 37 min to 45% B and in 1 min to 80% B. This is followed by a 1 min step at 80% B before reequilibration with the initial conditions. Mass spectra were acquired in centroid mode ranging from 250 – 3000 *m/z* at a 2 Hz scan rate.

LC-MS system 2 – semi-preparative purification

The final purification was performed on a Dionex Ultimate 3000 SDLC low pressure gradient system using a Luna, 5u, C18(2), 100A, 250 x 100 mm column (Phenomenex). Separation of 80 µl sample was achieved by a gradient from (A) H_2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 5 mL/min and 45 °C. The gradient was as follows: A two min isocratic step at 40 %B, followed by a ramp to 80 %B in 15 min, a two min plateau, return to initial conditions in 1 min and re-equilibration for two min. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 0.525 mL/min before entering the Thermo Fisher Scientific ISQTM EM single quadrupole mass spectrometer. Mass spectra were acquired by selected ion monitoring (SIM) at *m/z* 479.68 $[M+2H]^{2+}$, 488.68 $[M+2H]^{2+}$, 525.70 $[M+2H]^{2+}$ and fraction collection times were set accordingly.

S 2.1.4 NMR measurements

NMR spectra were recorded on a 500 MHz *Avance III* (UltraShield) spectrometer or on a 700 MHz *Avance III* (Ascend) spectrometer by Bruker BioSpin MRI GmbH, each equipped with a Helium cooled CryoProbe (TCI), at 298 K, if not stated differently. Chemical shift values of ^{1}H- and ^{13}C-NMR spectra are reported in ppm relative to the residual solvent signal (DMSO-d_6, δ_H 2.50 ppm and δ_C 39.5 ppm) given as an internal standard. ^{13}C-signals were assigned via 2D-CH and CCH or CNH correlations (HSQC and HMBC), using Bruker standard pulse programs. HSQC experiments were optimized for $^{1}J_{C-H}$ = 145 Hz and HMBC experiments were optimized for $^{2,3}J_{C-H}$ = 6 Hz. Multiplicities are described using the following abbreviations: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, b = broad; coupling constants are reported in *Hz*.

S 2.1.5 Feeding experiments

4 x 50 mL cultures of Mx152 in M8-CyH medium (1.5 g/L casein peptone, 1.5 g/L yeast extract, 1.0 g/L glucose monohydrate, 1.0 g/L soy flour (degreased), 4.0 g/L starch, 1.0 g/L $CaCl_2$ dihydrate, 0.5 g/L $MgSO_4$ heptahydrate, 0.004 g/L Na-Fe-EDTA, 11.8 g/L HEPES pH 7.2, adjusted with NaOH before autoclaving) were inoculated with 5mL of preculture. Each culture was supplemented with 50 µL of a 0.1 M stock solution of one of the labeled aminoacids (L-Valine-d_8, L-Serine-2,3,3-d_3 and L-Methionine-(methyl-^{13}C)). One culture was supplemented with sterile H_2O as a control. After 24 h, 48 h and 72 h, an additional 50 µL of the stock solution was added. After incubation for 8 days at 30 °C and 180 rpm the cultures were harvested by centrifugation. The pellets were then extracted with 50 mL of methanol, dried using a rotary evaporator, redissolved in 1 mL of methanol, and analyzed using LC-MS system 1a.

S 2.1.6 Biological assays

Screening with hCoV-229E-luc reporter virus

Firefly luciferase-expressing Huh-7.5/Fluc cells were infected with a *Renilla* luciferase reporter virus of the alphacoronavirus HCoV-229E-luc one day after seeding ($2x10^4$ cells/well in a 96 well plate) in the presence of indicated thiamyxin concentrations or DMSO.[1] After 48h post inoculation and incubation of the cells at 33 °C and 5% CO_2, the virus inoculum was removed, cells were washed with PBS and lysed in 50 µl PBS/0,5% Triton X-100. Lysis of cells was further enhanced by freezing of the plates at -20 °C. To determine the cell viability, firefly luciferase activity was measured in 20 µl of the lysate, whereas residual virus replication/infection efficiency was measured in 20 µL of lysate via the *Renilla* luciferase activity (Berthold Centro plate luminometer version 2.02). Mean values were normalised to DMSO treated infected samples and standard deviations of technical duplicates of four independent biological experiments are depicted. Non-linear regression curve and confidence interval (confidence level 95%) were both calculated with GraphPad Prism 9 and half-maximal inhibitory (IC_{50}) and cytotoxic (CC_{50}) concentrations were interpolated from the curves. As previously described, remdesivir served as positive control for the assay giving an IC_{50} value of 5.6 nM.[2]

Screening with dengue and Zika virus

Huh-7 cells were seeded at 1.5×10^4 cells per well in 96 well plates. On the next day, cells were treated with 10 different concentrations of 3 fold serial diluted compounds ranging from 2.54 nM to 50 µM. Cells were inoculated with DENV2-R2A or ZIKV-R2A reporter virus using a multiplicity of infection (MOI) of 0.1. At 48 h post infection, cells were lysed and *Renilla* luciferase activity contained in cell lysates was quantified. To determine cytotoxicity of the compounds, cell viability was measured in parallel to antiviral activity test using the CellTiter Glo® Luminescent Cell Viability Assay (Promega) according to manufacturer's instruction. Values were normalized using non-treated solvent control (0.5% DMSO) and IC_{50} and CC_{50} values were calculated using non-linear regression dose response analysis of the Prism 7 software package (GraphPad Software). Ribavirin was used as positive control giving an IC_{50} value of 2.3 µM for DNV2-R2A and 2.5 µM for ZIKV-R2A, without showing effects on Huh-7 cells (>90% cell viability) up to 100 µM concentration.

S 2.2 Supporting Data

S 2.2.1 Bioactivity data of MCy9487 crude extract

Table 2: Activity of strain MCy9487 extracted after 10-d cultivation in CYH medium. Activity determined by serial dilution as previously described[3]. 0- No activity, A- 20 µl extract (lowest tested dilution) causing activity, H- 0.16 µl extract (highest tested dilution) causing activity.

Test organism	Activity score
Escherichia coli TolC	E
Escherichia coli DSM 1116	0
Pseudomonas aeruginosa PA14	0
Micrococcus luteus DSM 1790	E
Mycobacterium smegmatis mc2-155	A
Chromobacterium violaceum DSM 30191	C
Staphylococcus aureus Newman	D
Bacillus subtilis DSM 10	D
Mucor hiemalis DSM 2656	D
Candida albicans DSM 1665	C
Wickerhamomyces anomalus DSM 6766	H

S 2.2.2 NMR-based structure elucidation

Thiamyxin B is assigned a molecular formula of $C_{43}H_{59}N_9O_8S_4$ on the basis of HR-ESI-MS data ($[M+H]^+_{meas}$ = 958.3449, $[M+H]^+_{calc}$ = 958.3442, Δ = 0.7 ppm). Interpretation of the 1D and 2D NMR spectra (Table 5-8, Figure 26-43) —which is described in detail in the following—,alongside with their characteristic MS^2 fragmentation pattern, reveals thiamyxin B to consist of the following peptide sequence: 2-(hydroxymethyl)-4-methylpent-3-enoic acid(HMMP)—alanine(Ala)—isoleucine(Ile)—Methyl-thiazoline(Me-thiazoline)—Thiazole—Me-Thiazoline—Dehydro-alanine(Dh-Ala)—Methylserine(Me-Ser)—N-Methylvaline(N-Me-Val)—Thiazole. In the ^1H and HSQC spectra of thiamyxin B we detect four α-protons with characteristic chemical shifts of δ_H 4.61 (Ala), 4.78 (Ile), 4.89 (Me-Ser) and 5.31 ppm (N-Me-Val). We find two signals corresponding to the terminal protons of an exo double bond at δ_H = 6.37 and 5.81 ppm (Dh-Ala), as well as three methine groups at δ_H 8.30, 8.57 and 5.83 ppm, whereof the deshielded shift of the two protons at δ_H 8.30 and 8.57 indicate their incorporation in the thiazole units and the more shielded shift of the methine group at δ_H 5.83 ppm indicates it as the HMMP aliphatic double bond proton. There are three additional shielded methine groups detectable at δ_H 1.91 (Ile), 2.83 (HMMP) and 2.36 ppm (N-Me-Val), which are—based on their characteristic chemical shifts—branching positions in the aliphatic sidechains of the respective amino acids. Furthermore, we find five methylene groups at δ_H 1.36/1.21 (Ile),

3.74 and 3.38/3.81 (Me-Thiazoline), 3.59 (Me-Ser) and 4.89 ppm (HMMP). The shielded shifts of the carbon atoms corresponding to the two proton signals at δ_H 3.59 (Me-Ser) and 4.89 ppm (HMMP) in the HSQC spectra at δ_C 71.0 and 67.0 ppm indicate their allocation next to an oxygen atom, whereas the deshielded shift of the methylene at δ_H 1.36/1.21 ppm, besides its characteristic splitting pattern allows its assignment as part of the aliphatic chain in Ile next to a stereo center. We detect eleven methyl groups at δ_H 0.96 (HMMP), 0.96 (HMMP), 1.36 (Ala), 0.93 (Ile), 0.87 (Ile), 1.59 (Me-Thiazoline), 3.19 (Me-Ser), 0.38 (N-Me-Val), 0.71 (N-Me-Val), 2.90 (N-Me-Val) and 1.65 (Me-Thiazoline), whereof the shielded shift of the carbon at δ_C 58.4 ppm corresponding to δ_H 3.19 ppm (Me-Ser) points towards its allocation next to an oxygen atom and the slightly shielded shift of the carbon at δ_C 30.2 ppm corresponding to δ_H 2.90 (N-Me-Val) towards its allocation next to an nitrogen atom. Lastly, we find four broad signals in the proton spectrum at 7.51, 8.30, 9.36 and 8.97 ppm without corresponding signals in the HSQC, which we assign as peptide bond forming amino groups belonging to Ala, Ile, Dh-Ala and Me-Ser, respectively. Ala was identified as follows: The α-proton at δ_H 4.61 ppm shows COSY correlations to the methyl group at δ_H 4.61 ppm, as well as to the NH proton at δ_H 7.51 ppm. Both the α-proton and the methyl group reveal HMBC correlations to a quaternary carbon at δ_C 171.6 ppm, which was assigned as the amide function connecting Ala and Ile based on its HMBC correlations to the Ile α-proton. The Ala α-proton furthermore shows HMBC correlations to a quaternary carbon at δ_C 165.8 ppm assigned to HMMP, indicating their peptide bond connection between the respective functions. Ile was identified as follows: The Ile α-proton at δ_H 4.78 ppm shows COSY correlations to the methine group at δ_H 1.91 ppm, as well as to the NH proton at δ_H 8.30 ppm. The methine group in turn reveals COSY correlations to the methyl group at δ_H 0.93 and the methylene 1.36/1.21 ppm. The methylene group reveals COSY correlations to another methyl group at δ_H 0.87 ppm. Both the α-proton and the methine group reveal HMBC correlations to a quaternary carbon at δ_C 173.1 ppm, which was assigned as the thioamid function connecting Ile and Me-Thiazoline based on its HMBC correlations with the Me-Thiazoline methylene group. The first Me-Thiazoline was identified as follows: The Me-Thiazoline methylene group at δ_H 3.74 ppm reveals HMBC correlations to the methyl group at δ_H 1.65 ppm, as well as two quaternary carbons at δ_C 176.5 and 83.3 ppm. Based on the characteristic chemical shift of the quaternary carbon at δ_C 83.3 ppm and its correlations to both the methyl and methylene group, it was assigned as Me-Thiazoline α-carbon. The quaternary carbon at δ_C 176.5 ppm shows additional correlations with the Thiazole methine group at 8.30 ppm, indicating it as thiamid function connecting the first Me-Thiazoline with Thiazole. The first Thiazole was identified as follows: The first Thiazole methine group at δ_H 8.30 ppm shows HMBC correlations to two quaternary carbons at δ_C 147.5 and 163.8 ppm. Based on its shielded chemical shift and correlations with the second Me-Thiazoline methylene function at δ_H 3.38/3.81 ppm, the quaternary carbon at δ_C 163.8 ppm forms the thiamid bond

between Thiazole and the second Me-Thiazoline. The second Me-Thiazoline was identified as follows: The second Me-Thiazoline methylene group at δ_H 3.38/3.81 ppm shows HMBC correlations to the methyl group at δ_H 1.59 ppm, as well as two quaternary carbons at δ_C 172.9 and 84.9 ppm. Based on the characteristic chemical shift of the quaternary carbon at δ_C 84.9 ppm and its correlations to both the methyl and methylene group, it was assigned as Me-Thiazoline α-carbon. The quaternary carbon at δ_C 172.9 ppm shows additional correlations with the Dh-Ala NH proton at δ_H 9.36 ppm, indicating peptide bond formation here. Dh-Ala was identified as follows: The exo-double bond protons at δ_H 6.37 and 5.81 ppm show HMBC correlations to two quaternary carbons at δ_C 132.7 and 163.1 ppm. The characteristic chemical shift of the quaternary carbon at δ_C 132.7 identifies it as participating in the double bond, whereas the quaternary carbon at δ_C 163.1 ppm was identified as amid function connecting Dh-Ala with Me-Ser based on its characteristic chemical shift, as well as correlations to the Me-Ser α-proton. Me-Ser was identified as follows: The Me-Ser α-proton at δ_H 4.89 ppm shows COSY correlations to the methylene group at δ_H 3.59 ppm, as well as to the NH proton at δ_H 8.97 ppm. The methylene group reveals HMBC correlations to a methyl group at δ_H 3.19 ppm, which was assigned as methoxy function based on its characteristic chemical shifts and splitting pattern as a singlet. Both the α-proton and the methylene group reveal HMBC correlations to a quaternary carbon at δ_C 170.6 ppm, which was assigned as the amide function connecting Me-Ser and N-Me-Val based on its HMBC correlations to the N-Me-Val α-proton. N-Me-Val was identified as follows: The α-proton at δ_H 5.31 ppm shows COSY correlations to the methine group at δ_H 2.36 ppm. The methine group in turn reveals COSY correlations to two methyl groups at δ_H 0.38 and 0.71 ppm. For both the α-proton and the methine group the HMBC correlations to the quaternary carbon at δ_C 168.3 ppm do not exceed the limit of detection, but its characteristic chemical shift and correlations with the second Thiazole methine at 8.57 ppm allow its assignment as the amide function connecting N-Me-Val and the second Thiazole. Further proof for this assignment was given by the NMR spectra acquired for thiamyxin A (Table 6, Figure 31-34) where the respective correlations are detectable. The NMe group at δ_H 2.90 ppm, which was assigned based on its characteristic chemical shift, is located at the Val amino function based on its HMBC correlations to both the N-Me-Val α-proton and the Me-Ser amide function. The second Thiazole was identified as follows: The second Thiazole methine group at δ_H 8.57 ppm reveals HMBC correlations to two quaternary carbons at δ_C 145.2 and 160.5 ppm. Based on its shielded chemical shift and correlations with the HMMP methylene function at δ_H 4.89 ppm, the quaternary carbon at δ_C 160.5 ppm forms the ester bond between Thiazole and HMMP. HMMP was identified as follows: The HMMP methylene function participating in the ester bond at δ_H 4.89 ppm shows HMBC correlations to a methine group a δ_H 5.83 ppm, as well as two quaternary carbons at δ_C 129.2 and 165.8 ppm. Based on their characteristic chemical shifts, the quaternary carbon at δ_C 129.2 ppm and the methine group at δ_H

5.83 ppm participate in a double bond. The methine group reveals COSY correlations to another more deshielded methine group at δ_H 2.83 ppm, which in turn shows correlations with two methyl groups at δ_H 0.96 ppm.

Their consistent peptide sequence and exact mass but different retention time on HPLC-MS indicate thiamyxin A ($[M+H]^+_{meas}$ = 958.3448, $[M+H]^+_{calc}$ = 958.3442, Δ = 0.6 ppm) and B to be diastereomers. This finding is in-line with the identical correlations observed in the NMR data acquired for thiamyxin A and B, as well as their very similar chemical shifts, except in the Ala and Ile region (see Table 5 and 6). The detailed analysis of this effect is described in the main manuscript and depicted in Figure 4 in this supporting information. The characteristic chemical shift of the HMMP methylene at δ_H = 4.89 and δ_C = 67.0 ppm, alongside with its HMBC correlation to the thiazole carboxy function implies cyclisation between the C-terminal thiazole and HMMP primary alcohol for both thiamyxin A and B. In contrary, thiamyxin C (Table 7, Figure 35-39) shows a shielded chemical shift at this position of δ_H = 4.00 and δ_C = 63.1 ppm with no signal split, which—in accordance with its sum formula of $C_{43}H_{61}N_9O_9S_4$($[M+H]^+_{meas}$ = 976.3556, $[M+H]^+_{calc}$ = 976.3548, Δ = 0.8 ppm)—reveals it to be the hydrolyzed non-cyclized congener (Figure 1). The molecular mass of thiamyxin D is further increased by $C_3H_6O_2$ ($[M+H]^+_{meas}$ = 1050.3928, $[M+H]^+_{calc}$ = 1050.3916, Δ = 1.1 ppm) when compared to thiamyxin C. Chemical shifts of the HMMP methylene in thiamyxin D only shows a deviation of 0.1 ppm for δ_C and 0.0 ppm for δ_H when compared to the C derivative, affirming that thiamyxin D also belongs to the open chain derivatives. In comparison to thiamyxin A-C, HSQC spectra of thiamyxin D (Table 8, Figure 40-43) show two additional methylene groups at δ_H 4.16/4.28 and 3.41 ppm, as well as one additional methine group at δ_H 3.76 ppm. Their characteristic chemical shifts reveal them all as hydroxylated and COSY correlations the methylene at 4.16/4.28 with the methine function, as well as COSY correlations of the methine group with the second methylene group at 3.41 disseminates the additional part of the molecule to be glycerol. The methylene group at 4.16/4.28 ppm furthermore shows correlations to the N-Me-Val C-terminus, revealing the N-Me-Val quaternary carbon at 172.2 ppm to participate in the ester bond with the respective glycerol methylene.

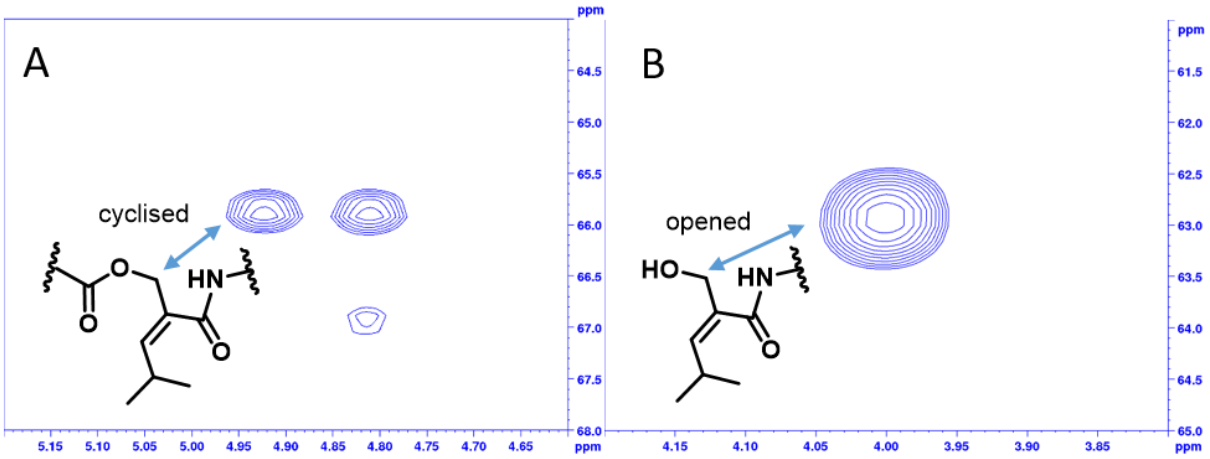


Figure 2: HSQC NMR signal of HMMP-CH_2 group of cyclized thiamyxin A (A) compared to open chain thiamyxin C (B) derivative.

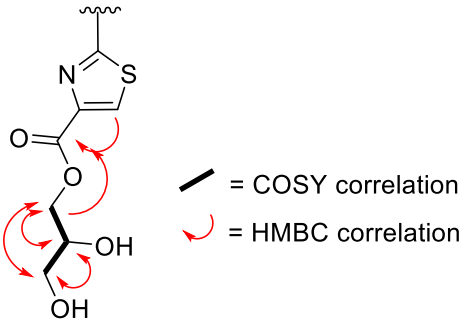


Figure 3: Glycerol attachment to thiazole unit in thiamyxin D derivative.

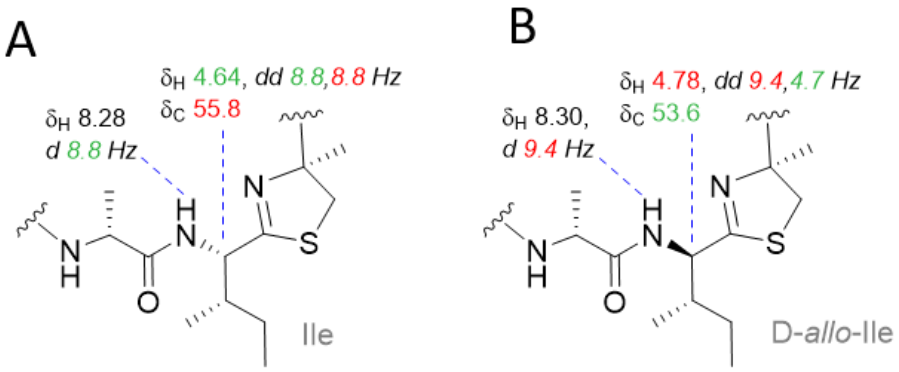


Figure 4: Analysis of observed shift and coupling constant differences of isoleucine vs. *allo*-isoleucin in thiamyxin A (A) and B (B). Larger values in red, smaller values in green are exactly matching the pattern shown by Anderson *et al.*, in 2017.[4]

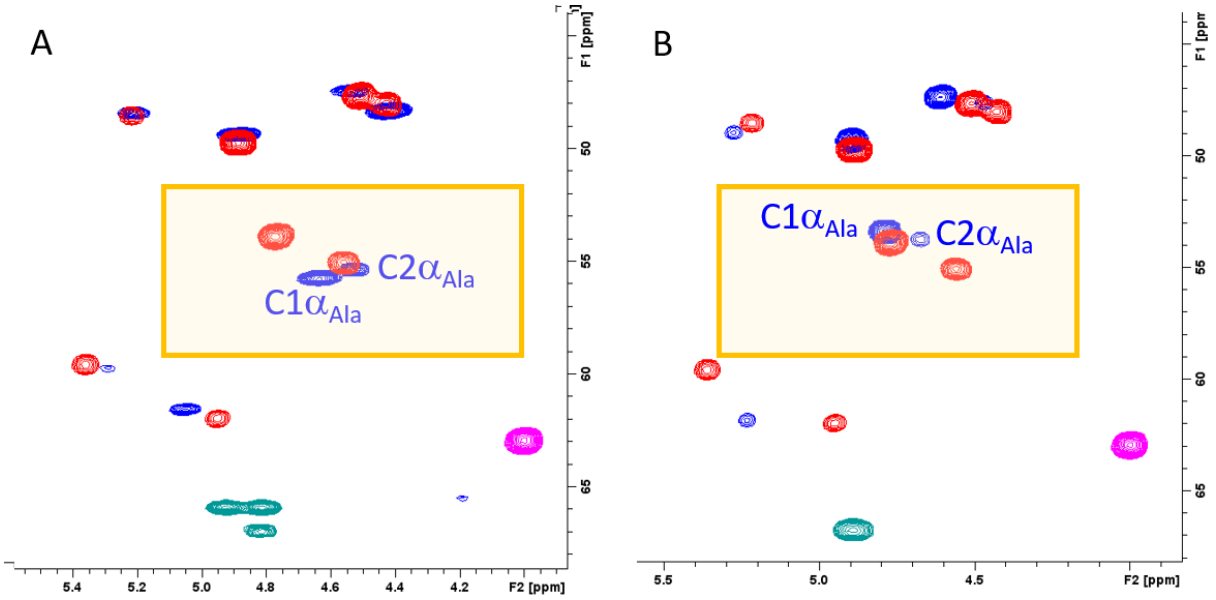


Figure 5: HSQC NMR section of thiamyxin A (A, blue/green signal set) and thiamyxin B (B, blue/green signal set), each overlapped with open chain derivative thiamyxin C (red/pink signal set). The highlighted signal sets were assigned to alpha-CH of the Alanin residue (Ala). Detailed analysis revealed both diastereomers (Ile and *allo*-Ile carrying derivative) in the open chain derivatives (see also Figure 3). The second signal set in thiamyxin A and B was assigned a second conformer (C1 and C2), as only occurring in the cyclized derivatives. In addition, the same ratio of signal sets in different isolated batches strongly indicates a stable equilibrium of two conformers.

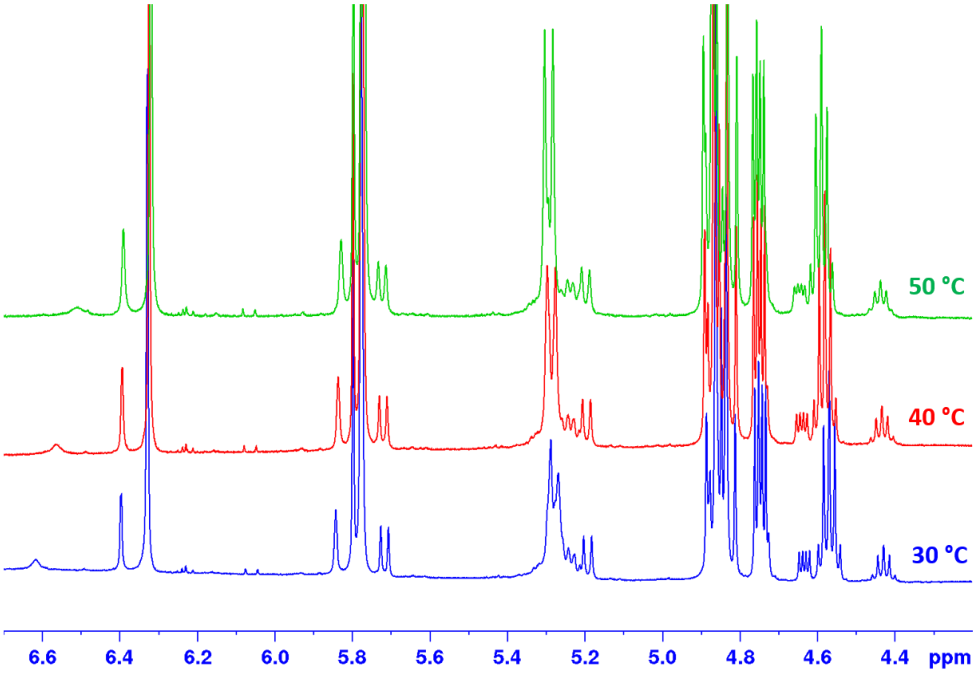


Figure 6: ^{1}H NMR section of thiamyxin B. Increasing measurement temperature does not cause a change in ration of different conformers.

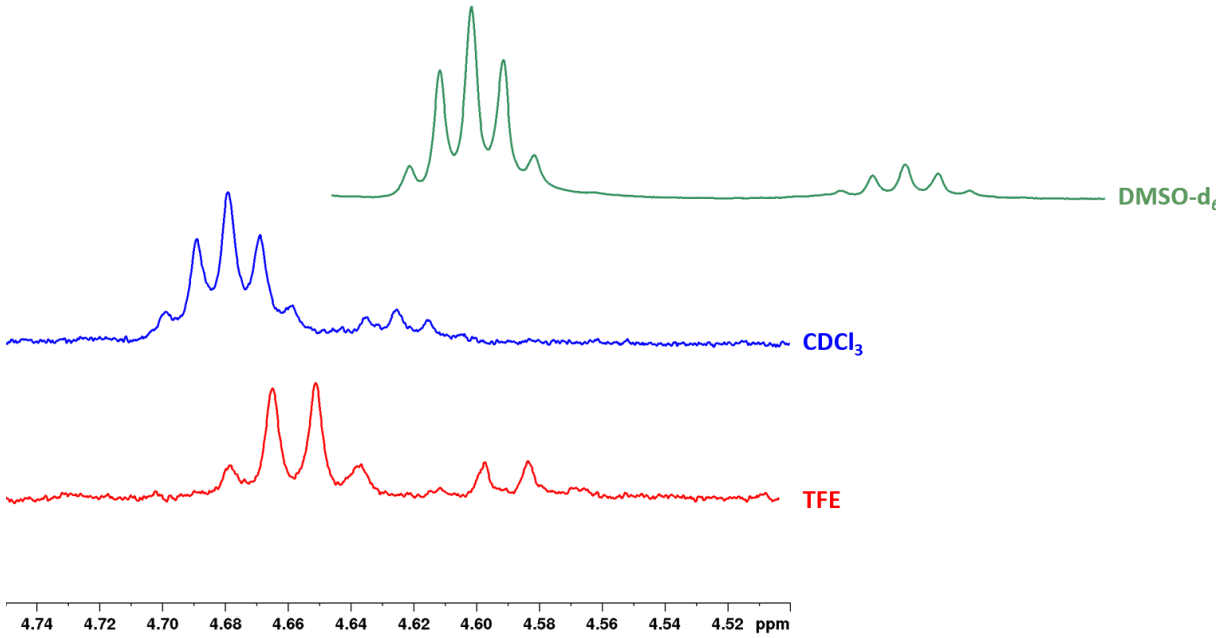


Figure 7: ^{1}H NMR section of thiamyxin B. Comparison of different solvents used for NMR measurements. No significant change in the ration of conformers was observed.

S 2.2.3 Stereochemistry assignment

The stereochemistry was assigned based using the Marfey's method.[5] Thiamyxin D was used for the stereochemistry assignment of Alanine, 4-methyl-thiazoline, O-methyl-serine and N-Methyl-valine. To attempt assignment of the isoleucine stereochemistry, thiamyxin A and B were hydrolyzed in addition to thiamyxin D and derivatized separately. The samples were analyzed using a longer gradient with improved separation for isoleucine and alloisoleucine. The retention times of references and thiamyxin hydrolysate and stereochemical assignment are summarized in table 1.

Table 3: Retention times and stereochemical assignment of the thiamyxin amino acids.

Reference	*L-FDLA*				*D-FDLA*				
Amino Acid	**Standard**	**TM D**	**TM A**	**TM B**	**Standard**	**TM D**	**TM A**	**TM B**	**configuration**
L-alanine	14.1	16.6	n/a	n/a	16.6	14.1	n/a	n/a	D/R
L-isoleucine	13.5	13.5	13.5	13.5	26.75	26.75	26.75	26.75	D/R – L/S
D-isoleucine	26.75	26.75	26.75	26.75	13.3	13.3	13.3	13.3	D/R – L/S
2-Me-Cysteine 1&2	10.4	10.4	n/a	n/a	9.6	9.6	n/a	n/a	L/R
O-Me-L-Serine	13.7	13.7	n/a	n/a	16.6	16.6	n/a	n/a	L/S
L-Valine	16.1	16.1	n/a	n/a	19.6	19.6	n/a	n/a	L/S

N-methyl-valine assignment

We were not able to detect N-methyl-valine after hydrolysis of the intact compounds, but instead observed a peak for L-valine. When subjecting the N-methyl-valine reference to the same conditions we were also able to detect Valine although in lower amounts (Figure 7). We assume that under acidic conditions the N-Methyl-Valine is demethylated and assigned the stereocenter as *S*-configured, in accordance with the detected L-Valine.

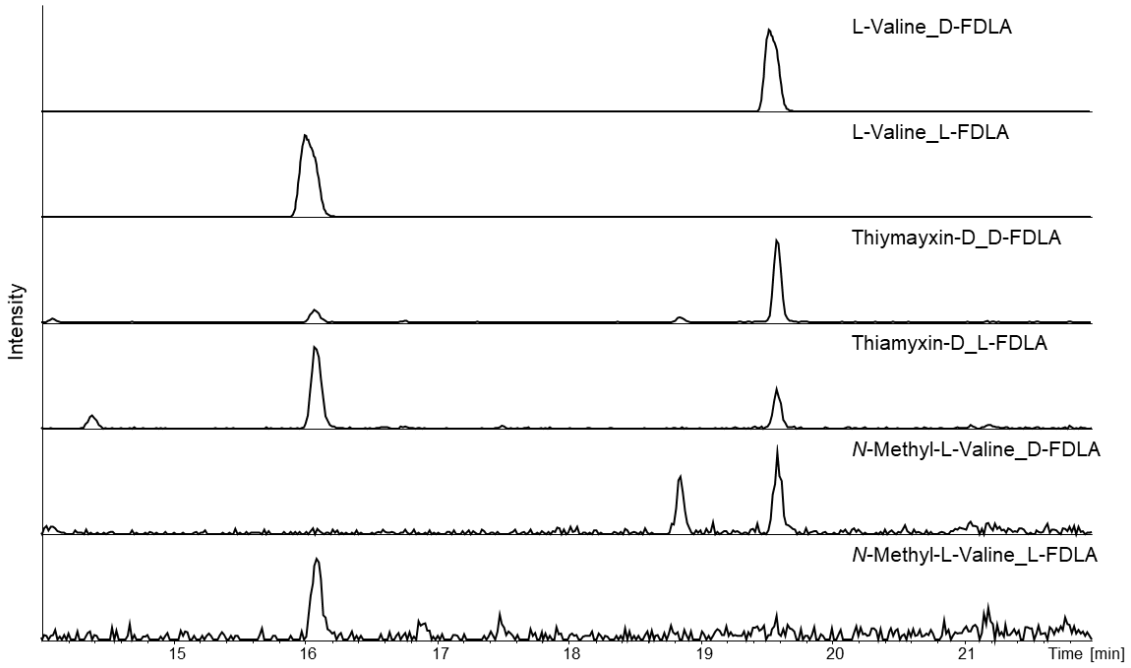


Figure 8: Extracted ion chromatograms for 412.183 *m/z* for derivatized L-Valine, thiamyxin D and N-methyl-L-valine.

Isoleucine assignment

To assign the stereochemistry of the isoleucine moiety, D- and L- isoleucine (I) as well as D- and L-alloisoleucine (AI) references were derivatized with D- and L- 1-flouro-2,4-dinitrophenyl-5-leucine-amide (FDLA). D-AI-D-FDLA, D-I-D-FDLA, L-AI-L-FDLA and L-I-L-FDLA have identical retention times,, as do D-AI-L-FDLA, D-I-L-FDLA, L-AI-D-FDLA and L-I-D-FDLA. When derivatized with D-FDLA, the thiamyxin A and B samples show peaks for D-AI-D-FDLA / D-I-D-FDLA and L-AI-D-FDLA / L-I-D-FDLA, however the peaks corresponding to D-AI-D-FDLA / D-I-D-FDLA are slightly larger. When derivatized with L-FDLA the thiamyxin A and B samples show peaks for L-I-L-FDLA / L-AI-L-FDLA and D-AI-L-FDLA / D-I-L-FDLA and again the peaks corresponding to D-AI-L-FDLA / D-I-L-FDLA are slightly larger. (Figure 8) We observe the corresponding peaks for D- and L- isoleucine for thiamyxin A as well as B due to a racemization reaction commonly occurring during acid hydrolysis for stereocenters adjacent to thiazoline moieties.[6] Even when extensively optimizing the reaction conditions, we were unable to fully retain Ile in its original configuration. As the differences in peak heights are not significant, we therefore were unable to assign D- or L-configuration of the thiamyxins based on this Marfeys analysis. Our NMR results however, clearly indicate thiamyxin B to incorporate Ile in *allo*-configuration (see Figure 3). We find an epimerization domain located in the isoleucine incorporating module so we assume the main product of the assembly line and all deriving shunt or degradation products (thiamyxin B, C and D) to incorporate D-*allo*-isoleucine. The high prevalence of L-Ile in natural products points towards its incorporation in thiamyxin A, but we cannot exclude a rather rare

incorporation of D-Ile here. The resembling biological activities of thiamyxin A and B would reason them both to be D-configured, as this would result in a more similar conformation of the macrocycle. As the target of the thiamyxins has not been identified yet, correlating their three dimensional structure with activity is highly speculative and only developing a total synthesis route for the thiamyxins will finally confirm the configuration of the Ile stereocenters.

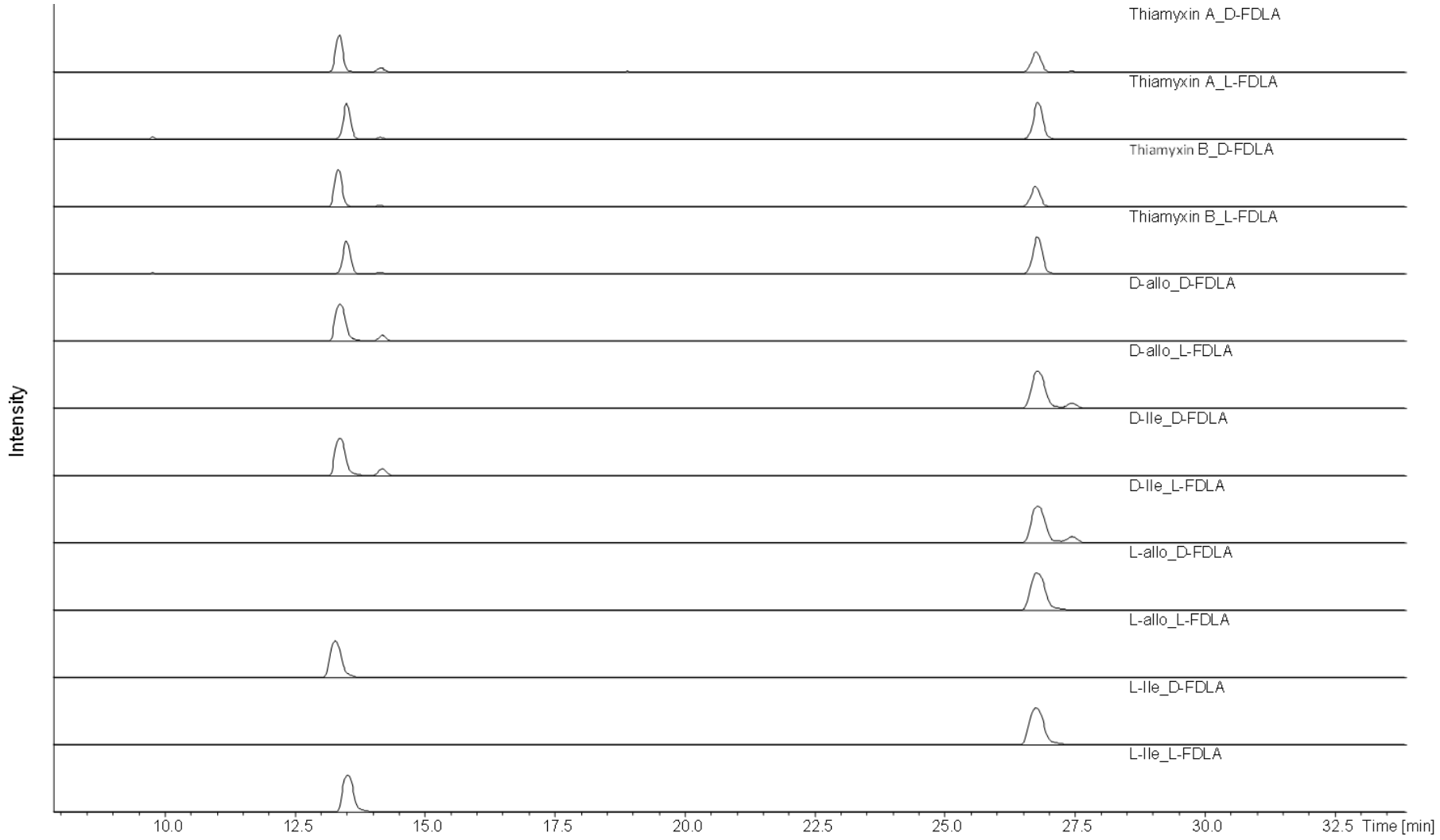


Figure 9: Extracted ion chromatograms for 426.198 *m/z* for derivatized D-*allo*-isoleucine, D-Isoleucine, L-alloisoleucine, L-isoleucine and thiamyxins A and B

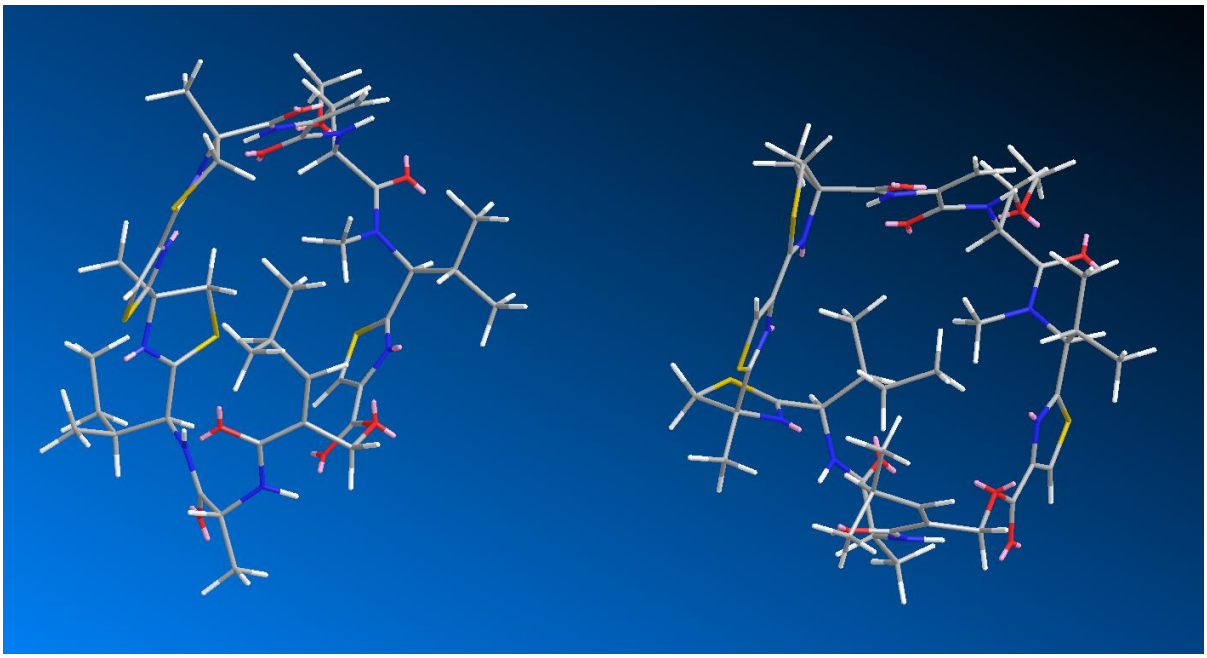


Figure 10: Lowest energy conformation of thiamyxin A (left) and thiamyxin B (right) as simulated using Chem3D (20.1.1).

Thiazoline assignment

To assign the stereochemistry of the 4-methyl-thiazoline stereocenters, we decided to use hydrolyzed thiangazole A as a reference due to inaccessibility issues of 2-methyl-cysteine. The peaks for 430.139 *m/z* of the hydrolyzed and derivatized samples of thiangazole A and thiamyxin D show identical retention times. As the stereocenters of the 4-methyl-thiazoline moieties in thiangazole A are all derived from *R*-configured 2-methylcysteine and there are no additional peaks with the relevant *m/z* in either sample we assign both 4-methyl-thiazoline stereocenters in the thiamyxins to be *R*-configured as well.

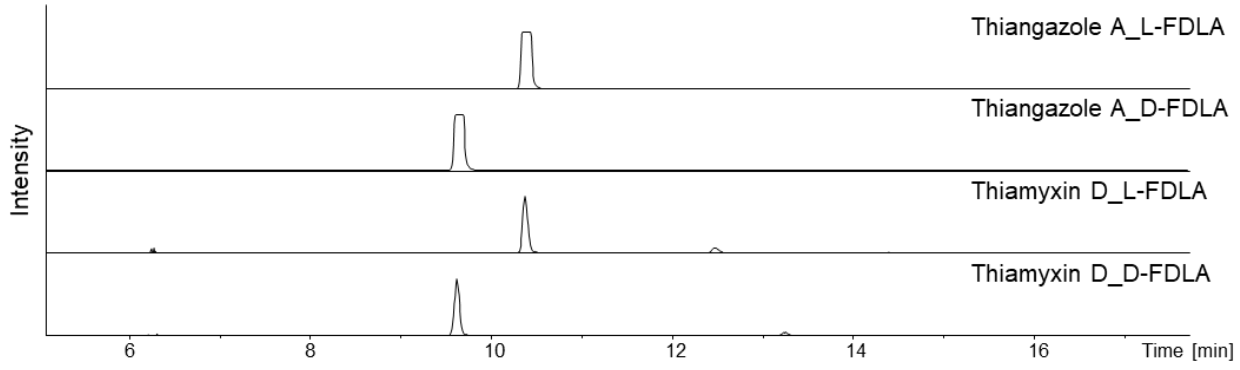


Figure 11: Extracted ion chromatograms for 430.139 *m/z* for derivatized thiangazole A and thiamyxin D.

Stability of thiamyxin A

To investigate if thiamyxin A can be converted to thiamyxin B, we treated thiamyxin A with pyridine, but could not detect any conversion. This finding supports our hypothesis that thiamyxin A indeed seems to be a side product of the assembly line rather than an artefact from the purification process.

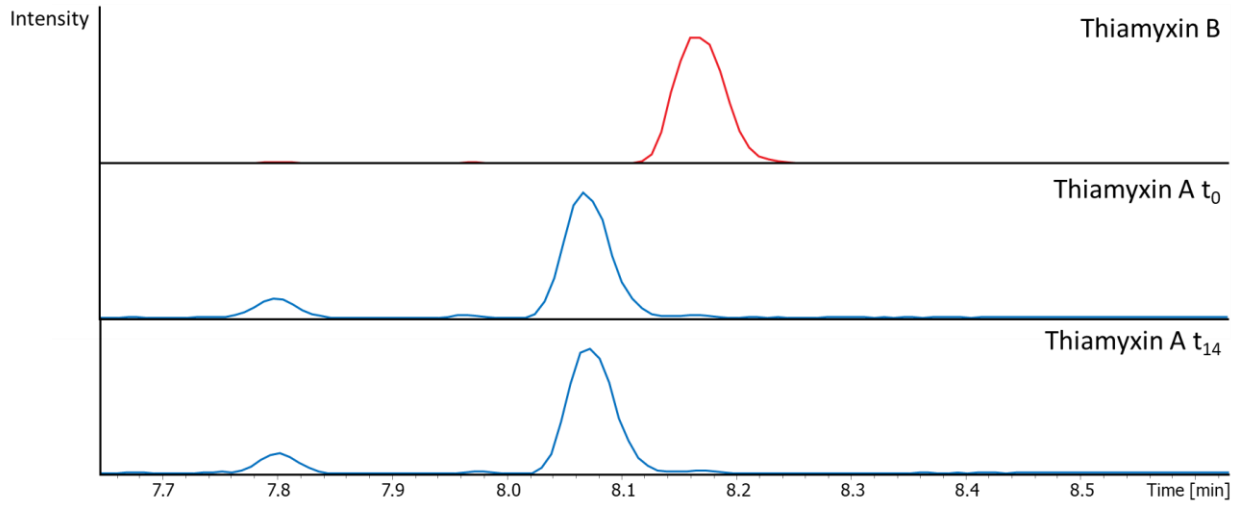


Figure 12: BPC of thiamyxin A (blue) in acqueous solution before base treatment (t_0) and after 14 h incubation at rt upon treatment with 5% pyridine. Thiamyxin B shown in red as a reference.

S 2.2.4 Gene cluster organization and proposed biosynthesis

Sequencing of the MCy9487 genome was performed as previously described for other myxobacteria.[2] The thiamyxin biosynthetic gene cluster was identified by submitting the MCy9487 genome to antiSMASH[7], analyzing the biosynthetic logic of the clusters and matching them to the elucidated structure. The pinpointed genes and closest orthologues can be found in Table 4. The orthologues were identified by BLAST[8] search based on the amino acid sequence of the query and the results table sorted by E-value. The top hit was chosen as representative orthologue. Most domains required for the biosynthesis hypothesis were identified by antiSMASH, however the oxidation domains in modules M7 and M12 as well as the O-methyltransferase domain in module M10 and the dimerization domain in module M1 were identified by submitting the corresponding sequences to CDD/SPARCLE[9]. The biosynthetic gene cluster with the corresponding annotations can be found under GenBank tag Thiamyxin_BGC_MCy9487 under GenBank accession number OP494098.

Table 4: Open reading frames in the thiamyxin biosynthetic gene cluster with proposed function and closest homologue.

Name	Size [AA]	Proposed function / homologue	organism	Accession number of closest protein homologue	Sequence Identity [%]
Orf 1	117	calmodulin	*Myxococcus fulvus*	WP_074958239.1	89.66
Orf 2	104	hypothetical protein	*Corallococcus terminator*	WP_147449015.1	62.14
ThiA	3022	polyketide synthase dehydratase domain-containing protein	*Corallococcus terminator*	WP_120539938.1	82.22
ThiB	3114	amino acid adenylation domain-containing protein	*Corallococcus terminator*	WP_120539939.1	84.04
ThiC	2025	AMP-binding protein	*Corallococcus terminator*	WP_147448653.1	84.85
ThiD	1393	amino acid adenylation domain-containing protein	*Hapalosiphon sp. MRB220*	WP_053457693.1	51.31
Orf 3	107	hypothetical protein	*Corallococcus terminator*	WP_120544219.1	61.95
Orf 4	396	TRC40/GET3/ArsA family transport-energizing ATPase	*Corallococcus terminator*	WP_120544218.1	96.97
ThiI	413	cytochrome P450	*Corallococcus terminator*	RKG77511.1	87.89
ThiJ	319	alpha/beta fold hydrolase	*Corallococcus terminator*	WP_158625188.1	84.64
ThiE	1555	amino acid adenylation domain-containing protein	*Corallococcus terminator*	WP_120544215.1	88.24
ThiF	2499	amino acid adenylation domain-containing protein	*Archangium primigenium*	WP_204490555.1	49.15
ThiG	2055	amino acid adenylation domain-containing protein	*Corallococcus terminator*	WP_120544021.1	85.29
ThiH	742	amino acid adenylation domain-containing protein	*Corallococcus terminator*	WP_120544021.1	83.11
Orf 5	940	type I polyketide synthase	*Corallococcus terminator*	WP_120544022.1	81.06

Orf 6	336	patatin-like phospholipase family protein	*Corallococcus terminator*	WP_120544023.1	87.99
Orf 7	333	radical SAM protein	*Corallococcus terminator*	WP_120543822.1	94.89
Orf 8	182	NUDIX domain-containing protein	*Corallococcus sp. ZKHCc1 1396*	WP_193348944.1	87.43
Orf 9	296	hypothetical protein	*Corallococcus terminator*	WP_120543827.1	87.50

Active site residues of module 5 (*thiC*)

Module 5 consists of a condensation and peptidyl carrier protein (PCP) domain. As there is no biosynthetic step assigned to this module, yet it the C domain contains an intact active site, we propose that its function is assisting the loading the nascent molecule onto the PCP-domain of the next module.

```
Feature 1
query          9 IYPLSPSQQGMLYESLA.[5].IHVEQLVWRM.[1].GPLDEA.[1].LERAWQQVVDRH.[1].VLRTCF.[4].QSEP 75
NP_624809   1194 ILPLTPLQEGLYFHSVY.[6].SYVEQQLLTL.[1].GEVDPG.[1].LAAAATRLLTLH.[1].NLAARF.[4].DGRV 1261
NP_251114   3250 VYPLTPMQEGLLLHTLL.[5].IYYMQDRYRI.[1].SPLDPE.[1].FAAAWQAVVARH.[1].ALRASF.[4].GETM 3316
Q939Z0      3022 IWPLSPLQEGLLFHAAD.[5].VYASMRTLAI.[1].GPLDVA.[1].FRASWTVLLDRH.[1].ALRASF.[4].SGEA 3088
WP_012266627  15 IYPLSPMQEGMLFHSLY.[5].IYCSQTLITL.[1].GEINLA.[1].FRQAWEKVVERH.[1].VLRTLF.[4].REKP 81
WP_012267918  17 IYPLSPMQQGMLFHSLY.[5].TYLSQIQITL.[1].GNLDIN.[1].FQQAWQKVVDRH.[1].ILRTCF.[4].TKQP 83
O68008      5091 IYPLANMQRGMLFHALE.[5].AYFEQMAINM.[1].GLIDER.[1].FAETFNDIMERH.[1].ILRASI.[4].TDEP 5157
Q70LM6      2584 VYTLTPLQEGMLFHSLY.[5].DYVVQLALKL.[1].HVNVEA     FSAAWQKVVERH.[1].ILRTSF.[4].LEKP 2649
Q04747         9 VYALTPMQEGMLYHAML.[5].SYFTQLELGI.[1].GAFDLE.[1].FEKSVNELIRSY.[1].ILRTVF.[4].LQKP 75

Feature 1                                                                                ##   #
query         76 VQVVL.[2].VSVTLRSHDL.[11].LEAELE.[1].DRLRGFKP.[2].APLMKLSLFR.[4].EARLVWTFHHLLLDG 149
NP_624809   1262 VSVLE.[2].REAPFTVLDR.[ 7].IRAHAE.[1].DRRAGFDL.[2].GPPMRYTLIR.[4].RHVLVQTVHHIVADG 1331
NP_251114   3317 LQVIH.[2].GRTRIEFLDW.[11].LQALHK.[1].EREAGFDL.[2].QPPFHLRLIR.[4].RYWFMMSNHHILIDA 3390
Q939Z0      3089 VQVIA.[2].VPPDWRETDL.[11].FDRLAA.[1].MHAERFDL.[2].APQLRLHLVR.[4].RYRLIFTSHHIVADG 3162
WP_012266627   82 LQIVR.[2].VDLPWDYQDW.[11].LDLLLE.[1].ERQQGFEL.[2].APLMRCLMIQ.[4].TYKFLCNHHHIILDG 155
WP_012267918   84 LQVVR.[2].VTLPWFNQDW.[11].FQELLT.[1].DKEQYFEL.[2].VPLIRCHLIR.[4].KYEFIHTGHHILLDG 157
O68008      5158 RNVII.[2].RKINLDYHDL.[11].IQAYRK.[1].DREKGFRL.[2].EPLIRAALMR.[4].SYTFIWTNHHILLDG 5231
Q70LM6      2650 HQVVH.[2].VKTFVERLDW.[11].LQTYLE.[1].DRKRGFDL.[2].PPLMRWTLIR.[4].TFQFVWSFHHMLLDG 2723
Q04747         76 RQVVL.[2].RKTKVHYEDI.[11].IERYKQ.[1].VQRQGFNL.[2].DILFKVAVFR.[4].QLYLVWSNHHIMMDG 149

Feature 1
query        150 WCLPLLLQEVLSCY.[13].SRPFRDYVAWLKR.[2].PEAARTFWAERLRA.[3].PTPLG.[ 4].QGEPG.[1].G 224
NP_624809   1332 WSVPPMLRTLLAEY.[ 8].LGGFPEHVRRLAA.[2].GAASDRVWDEQLAD     LPGPS      LIAEG     H 1393
NP_251114   3391 WCRGLLMNDFFEIY.[13].PPRYRDYIAWLQR.[2].LEQSRRWWSESLRG.[2].RPTLV.[10].AGESG.[1].M 3470
Q939Z0      3163 WSLPLILVDVLTAY.[11].ATSYRDFLAWVDR.[2].KGAAGQAWRTELAG.[2].EATHV.[ 3].GSIIT.[1].L 3233
WP_012266627  156 WSMPIIYQEVLGFY.[13].PRPYQDYIVWLQQ.[2].PSIAESFWQRTLEG.[3].PTPLR.[ 6].MKSEG.[1].P 232
WP_012267918  158 WATAILLKEVFDFY.[13].PRPYQDYINWLQQ.[2].QTEAERFWRKNLQG.[3].PTPLV.[ 6].PLVTQ.[1].K 234
O68008      5232 WSRGIIMGELFHMY.[13].ARPYSDYIGWLQQ.[2].KEAAKAYWRNYLSG.[3].KSPIS.[ 4].SSGHA.[1].Y 5306
Q70LM6      2724 WSTPIVFQDWQAFY.[13].IPPFSAYIAWLKR.[2].LEEAQQYWRDYLQG.[3].PTPLG.[ 5].GSAGQ.[1].K 2799
Q04747       150 WSMGVLMKSLFQNY.[13].GKPYSDYIKWLGK.[2].NEEAESYWSERLAG.[3].PSVLP.[ 3].PVKKD.[1].Y 223
```

Figure 13: Alignment of the C domain of module 5 (thiC) amino acid sequence (query) with reference sequences from CDD/SPARCLE database. The active site histidines and aspartic acid of the characteristic HHxxxD motif are highlighted in yellow.

Active site residues of orf 5

Orf 5 is located downstream of thiH of the thiamyxin gene cluster. It contains a ketosynthase (KS) and PCP domain. The KS domain is missing the first active site histidine (Figure 13). We were not able to assign a biosynthetic function to these domains and therefore presume them to be inactive.

```
Feature 1                #
query       158 GPSLGVQTACSTSLVAVHLACQSLLSG.[2].DMALAGAVSM.[5].RGYLHEEGSSLS.[4].CRAFDARANGTPFGSG 233
AAC38075    174 GPAMTVTTACSSSLVAMHLACRALQAG.[2].DMALAGGVNL.[5].LTIYMSQIRAIS.[4].CRVFDAAADGIVRGEG 249
AAB08104    177 GPSVLVDTACSGGLTALHLACQSLLVG.[2].RQALAAGSSL.[5].MMVTMSMMKFLS.[4].CYAFDERANGYARGEG 252
AAB53258    174 GPSYTVDSACSSSLYALEHAFRAIRDG.[2].DAAIVGGSNL.[5].VSLQFSRLGVLS.[4].CKSFDNSANGYARSEA 249
CAB06094    175 GPSIAVQTACSSSLVAVHLACLSLLSG.[2].DMALAGGSSL.[5].VGYFTSPGSMVS.[4].CRPFDVRADGTVFGSG 250
AAP42872    206 GPTATLDTACSGSLVALHLACQSLRGG.[2].SMALAGGVTV.[5].TFIGTGRGIGLP.[4].CRSFADGAEGIAFAEG 281
CAD19086    194 GPSIVVDTACSSSLVAVHLACQSLRSK.[2].DLAIAGGANL.[5].WSVAISKLQALA.[4].CKTFDARADGFVRAEG 269
CAE14178    201 GPSLQIDTACSSGLTALTQAVNSLRSG.[2].QQAIVGSVNL.[5].NMAAYYRAGMLS.[4].CRVFDADANGFVRGEG 276
AAF00958    197 GPCLSIDTACASSLVAVHQGIRSLRNR.[2].ELALVGGVNL.[5].ITISLSQSGMMS.[4].CKTFDASANGYVRGEG 272
BAB12210   2281 GPSMTIDTACSSSLVAIHLAYNALLNG.[2].DLALAGGVNI.[5].ISLIESRAHMLA.[4].CKTFDESANGMVRGEG 2356

Feature 1                                                                #
query       234 LGVVILKRLSEA.[6].IHAIIKGSALNNDGALKVGY.[4].GDSQARVIREALA.[7].SISYVEANGTGTPLGDTIEL 315
AAC38075    250 CGVTVLKRLADA.[6].IQAVIRGSAINQDGASAGQT.[3].ANAQAAVISQALK.[7].DIDYVEAHGTGTPLGDPIEL 330
AAB08104    253 VAVLLLKRLEDA.[6].IRAVIRGTGCNQDGKTPGIT.[3].SVSQEALIRSVYK.[7].DTTYVECHGTGTQAGDTTEA 333
AAB53258    250 IVVCFLQKAKDS.[2].VYAQLLHAKTNCDGYKEQGI.[4].GHIQKLLLREFYE.[7].ELEFVEAHGTGTRVGDPEEL 327
CAB06094    251 VGLVVLKPLAAA.[6].IHAVIRGSAINNDGSAKMGY.[4].PAAQADVIAEAHA.[7].TVSYVECHGTGTPLGDPIEI 332
AAP42872    282 AGVVLLERLSTA.[6].VLAVVRGSAIGQEGTNNGVS.[3].GPAQQRLIRQALA.[7].EIDAVEGQGTGGLLSDAVEA 362
CAD19086    270 CGTIVLKRLSDA.[6].IRALIRGSATNQDGHSQGLT.[3].GLTQQALLRQALQ.[7].QVSYIETHGTGTILGDPIEV 350
CAE14178    277 AICLFLKTQKQA.[6].IYGYVRASAVNHGGRANSLT.[3].PEQQIALVKDCLL.[7].QISYLEAHGTGTSLGDPIEF 357
AAF00958    273 CGVLILKTLSEA.[6].ILALLRGSAVNHNGAAAGLT.[3].GPAQQELLRQALA.[7].DVSYIEAHGTGTSLGDPIEL 353
BAB12210   2357 CGIVVLKRLSQA.[6].ILAKIYGTAVNHDGPSSGLT.[3].GQAQEKLLHQALK.[7].QIDYIEAHGTGTALGDPIEL 2437

Feature 1                          #
query       316 SALNRAF.[10].ALGSVKTNVGHLSVASGMAGLIKTVLALE.[1].QQLPASLNF.[17].TGLKGWP.[5].RRAGVS 406
AAC38075    331 SSLDSAF.[ 7].WVGSVKANMGHLDAAAGMASVIKTMMVLK.[1].AEVPAQLHL.[17].TAIESLP.[3].RLAGIS 416
AAB08104    334 SALSKVF.[ 8].LIGSVKTNIGHLEGASGLAGVVKSILMLE.[1].GVILPNRNF.[16].PTTLECW.[5].RRVSIN 421
AAB53258    328 LAIDEIF.[ 8].LLGSIKSNLGHSEPASGLCSIAKMCIAYT.[1].GYIPPNLNY.[14].MNVITDK.[5].GMSGIN 413
CAB06094    333 QGLRAAF.[11].VLGSVKSNIGHLEVAAGIAGLIKTILCLK.[1].KALPATLHY.[17].SKYGPWE.[4].RRAGVS 423
AAP42872    363 QALASVY.[10].LLGAVKSNLGHTQGASGVAGVIKTVQAMR.[1].GVLPRTLHT.[17].TAATPWP.[5].LRSGVS 453
CAD19086    351 SALSEVY.[10].ILGSVKTNVGHLEAAAGIAGLLKVVLALE.[1].GAVPKQLHF.[17].TEMSPWP.[5].RMGAVS 441
CAE14178    358 NALNEVF.[12].YIGSVKANIGHLEGAAGLAGIVKVLLMLQ.[1].KSIVPSAAF.[17].TEENSWR.[6].RFAGLS 451
AAF00958    354 NAIASVY.[ 7].YVASVKTNIGHLEAAAGMAGIIKTILILQ.[1].GEIPPHLHF.[17].TQNIPWP.[5].PIAGVS 441
BAB12210   2438 ESMSAVF.[ 9].IIGSVKTNLGHLEGAAGIAGLIKTVLALQ.[1].HKIPPHLHF.[17].VQGKPWD.[5].RIAGVS 2527
```

Figure 14: Alignment of the orf 5 KS Domain amino acid sequence (query) with reference sequences from CDD/SPARCLE database. The active site cysteine and histidines are highlighted in yellow.

Dimerization domain in thiA

A dimerization domain (red) commonly found within methyltransferase domains was identified upstream of the GNAT domain in module M1 of the thiamyxin gene cluster by submitting the aminoacid sequence of thiA to CD-search (Figure 14).

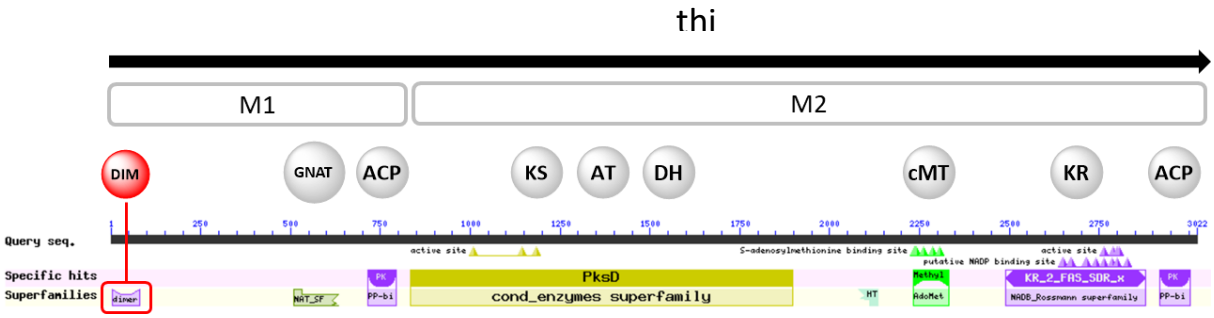


Figure 15: Output of the CDD/SPARCLE conserved domain search of thiA. The dimerization domain is marked in red, other domains in grey.

Alignment of GNAT Domains

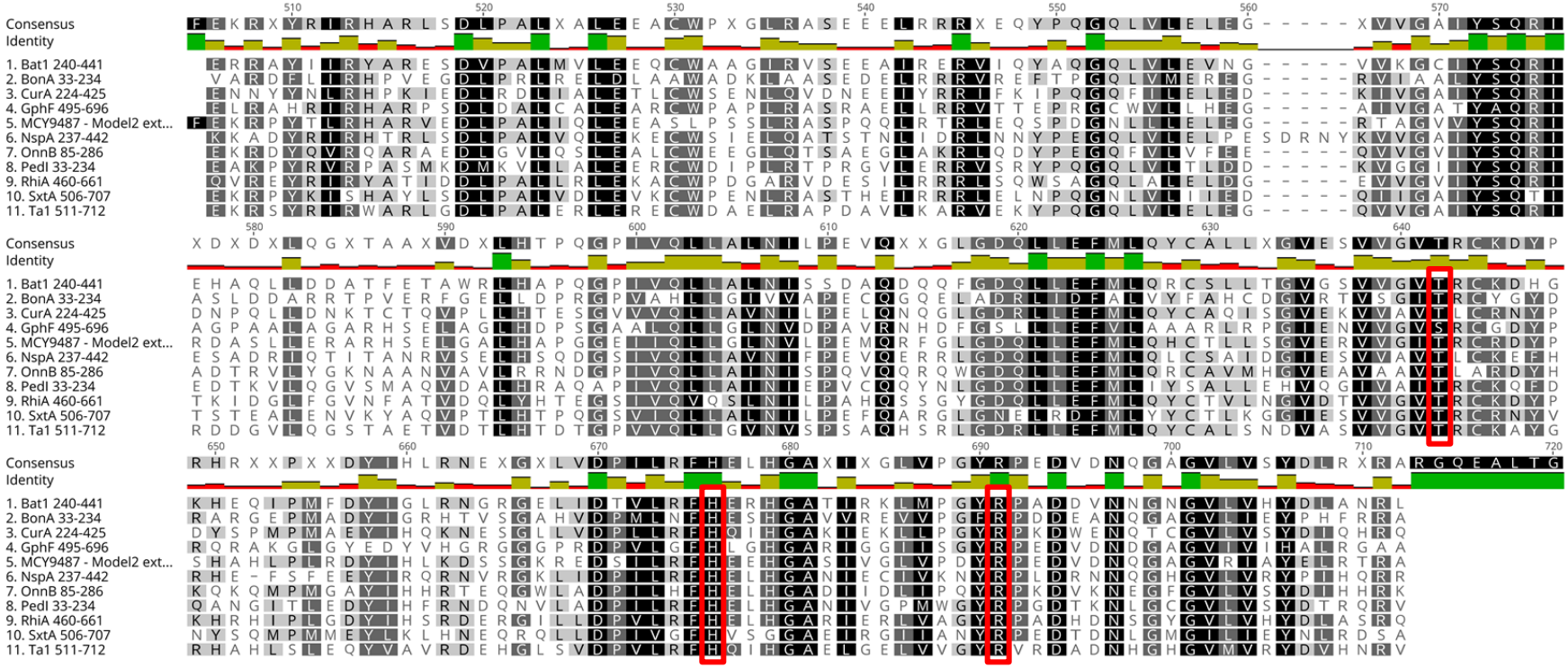


Figure 16: Alignment of GNAT domains from the thiamyxin (MCY9487 – Model 2), Gephyronic acid (GphF), Curacin A (CurA), Batumin (Bat1), Bongkrekic acid (BonA), Nosperin (NspA), Onnamide (OnnB), Pederin (PedI), Rhizoxin (RhiA), Saxitoxin (SxtA) and Myxovirescin A (Ta1) gene cluster, as performed by Skiba *et al*.[10] The conserved histidine, arginine and threonine required for decarboxylation are marked in red.

Phylogenetic analysis of C domains

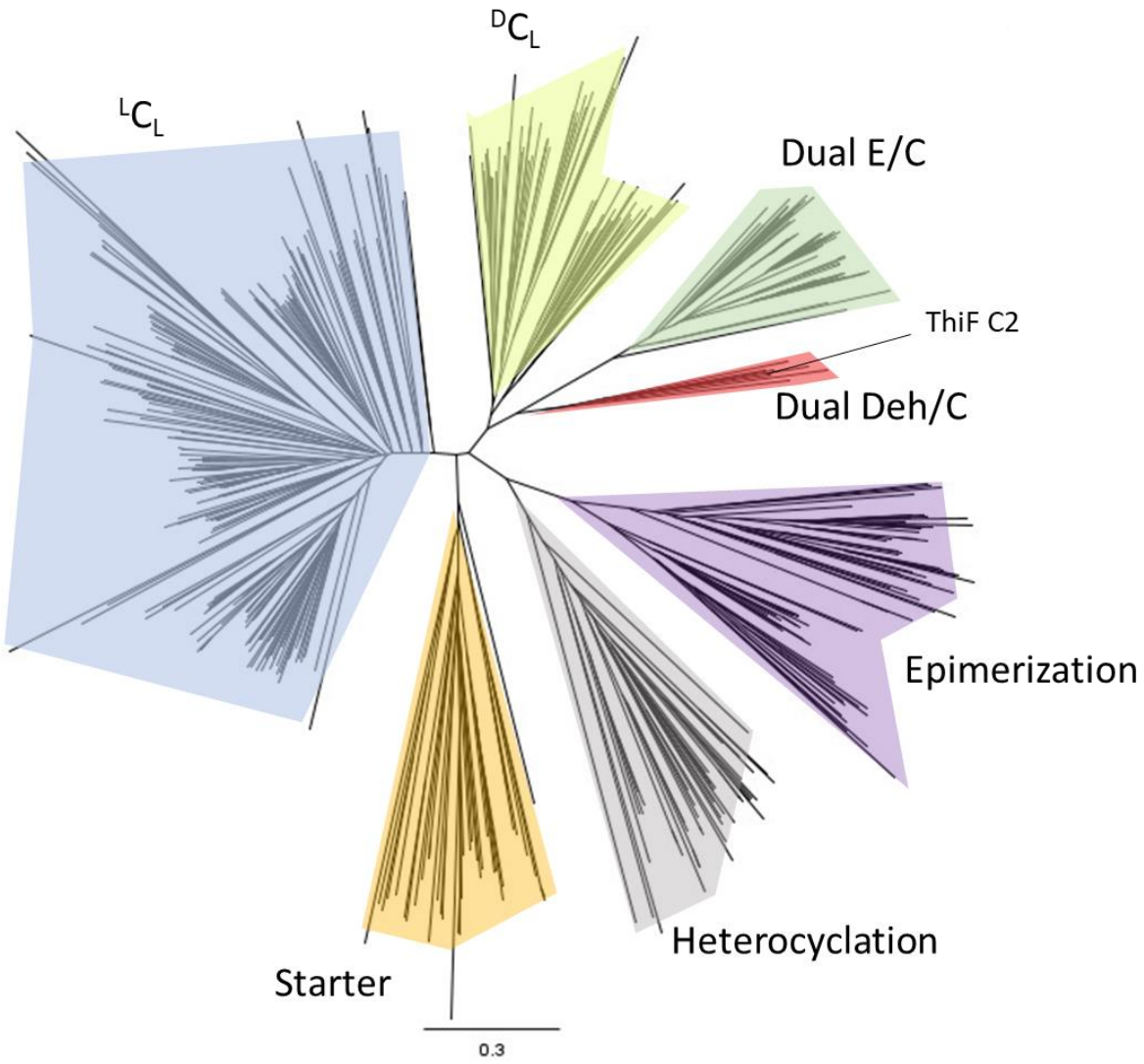


Figure 17: Phylogenetic tree of different C domain subtypes as previously described by Pogorevc *et al*.[11] Phylogenetic tree of all known C domain subtypes (LCL, DCL, Starter, Dual E/C, Epimerization and Heterocyclization domains), with the new Dual Deh/C subtype highlighted in red. The phylogeny was reconstructed using phyml, employing the JTT model of amino acid substitution and a gamma-distributed rate variation with four categories. The support values are based on 100-fold bootstrapping. The C domain list includes 525 domains from phylogenetic study by Rausch *et. al*.,[12], 10 C domains from the thiamyxin biosynthesis, as well as selected C domain examples from α,β-dehydro amino acid forming pathways: bleomycin (blmVI; Q9FB23), burriogladin (BgdA; MH170348), haerogladin (HgdA, MH170356), nocardicin (NocB; Q5J1Q6), hassallidin (HasO; K7VZQ9), syringomycin (SyrE; O85168), stenothricin (StenS; EFE73312.1). The C domain of module 10 of the thiamyxin biosynthesis encoded on thiF, belongs to the clade of dehydrating C domains.

S 2.2.5 Feeding Experiments

The isotopic patterns of the thiamyxins for the serine-d_3 supplemented cultures show intensified peaks at +2 Da, +3 Da, +4 Da and +5 Da compared to the monoisotopic peak, indicating 2 deuterium-labelled serine incorporations, one of which is reduced to dehydroalanine. The isotopic pattern of the thiamyxins for the Valine-d_8 supplemented culture show intensified peaks at +7 Da, +8 Da and +9 Da compared to the monoisotopic peak, indicating a deuterium-labelled valine incorporation. The alpha deuterium is acidic and exchanges with hydrogen in aqueous conditions which causes the +7 Da signal. The isotopic patterns for the methionine-(methyl-^{13}C) supplemented cultures show intensified peaks at +1 - +9 Da, indicating up to 7 methylations (see Figure 17).

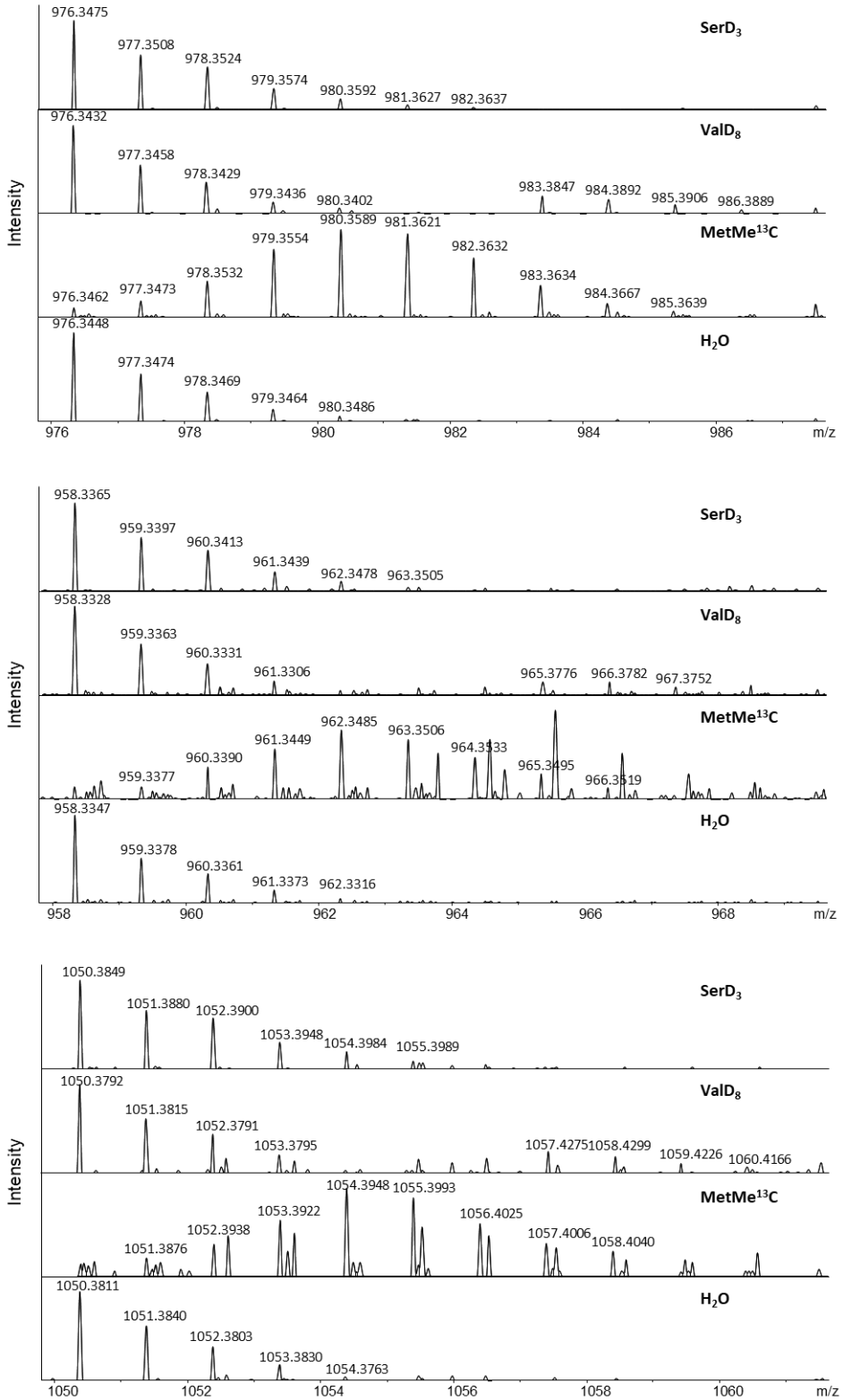


Figure 18: Zoomed in mass spectra for thiamyxin C (top), thiamyxin A/B (middle) and thiamyxin D (bottom) with L-serine-2,3,3-d_3, L-valine-d_8, and L-methionine-(methyl-^{13}C) feeding and control.

S 2.2.6 Antimicrobial activities

All thiamyxins except congener A, which was only produced in low amounts, were tested against two Gram-positive and two fungal pathogens. They were only found to have weak effects on our test panel with MIC values higher than 32 µg/mL (see Table 4).

Table 5: Antimicrobial activities of the thiamyxins.

Test organism	**MIC [µg/mL]**			
	thiamyxin A	thiamyxin B	thiamyxin C	thiamyxin D
Bacillus subtilis DSM-10	nd	> 64	64	> 64
Mucor hiemalis DSM-2656	nd	> 64	64	> 64
Candida. albicans DSM-1665	nd	64	> 64	> 64
Micrococcus luteus DSM-1790	nd	> 64	32	> 64

S 2.2.7 Antiviral activities

Half maximal inhibitory concentrations (IC_{50}) values presented in the main text were calculated based on the following curves determined for the inhibition of the respective virus (Figure 18-25). Half maximal cytotoxic concentration (CC_{50}) values presented here were determined simultaneously in the infected cell lines to see if the observed antiviral effects are overlapping with effects on the cell lines. IC_{50} values shown in the manuscript were determined in an independent experiment to circumvent effects of the viral infections on the cell lines.

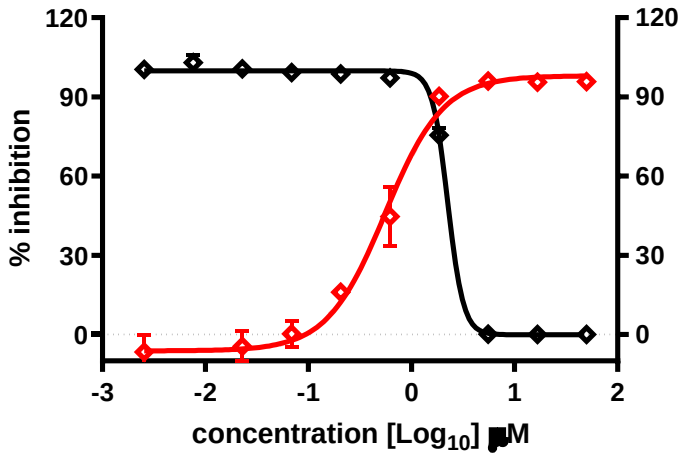


Figure 19: Inhibition of DENV-R2A (red, IC_{50} = 0.56 µM) and effects on cell viability (black, CC_{50} = 2.25 µM) of thiamyxin B.

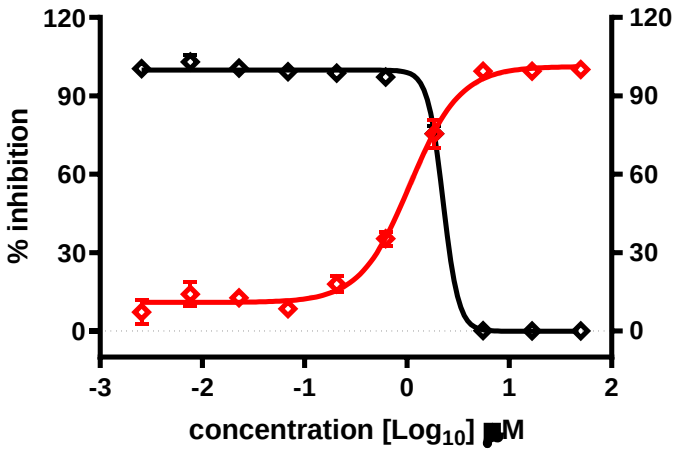


Figure 20: Inhibition of ZIKV-R2A (red, IC_{50} = 1.07 µM) and effects on cell viability (black, CC_{50} = 2.25 µM) of thiamyxin B.

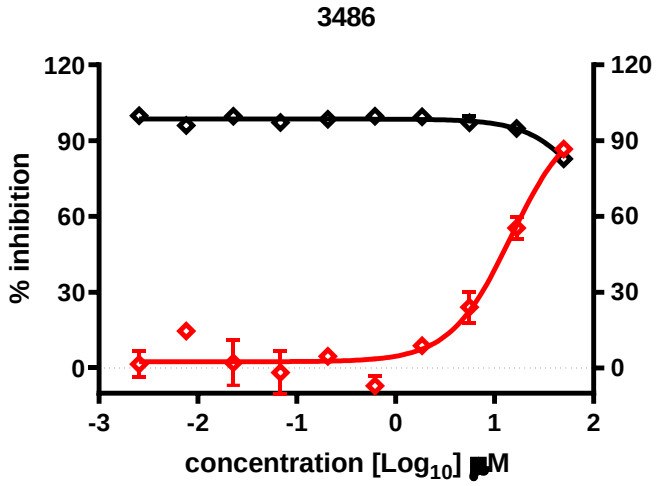


Figure 21: Inhibition of DENV-R2A (red, IC_{50} = 14.56 µM) and effects on cell viability (black, CC_{50} = >50 µM) of thiamyxin C.

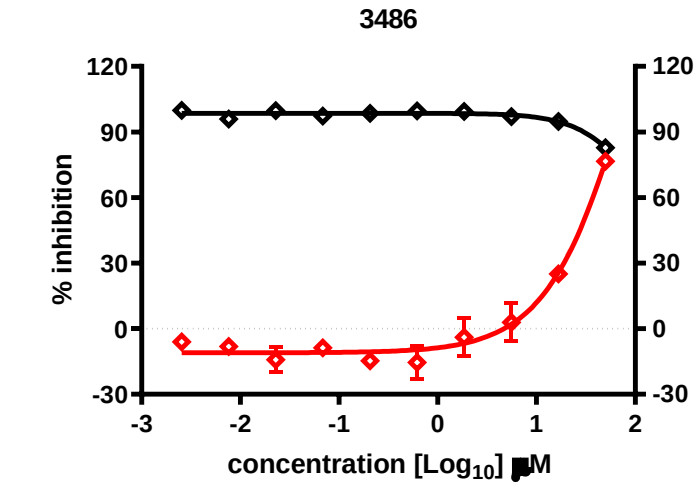


Figure 22: Inhibition of ZIKV-R2A (red, IC_{50} >15 µM) and effects on cell viability (black, CC_{50} >50 µM) of thiamyxin C.

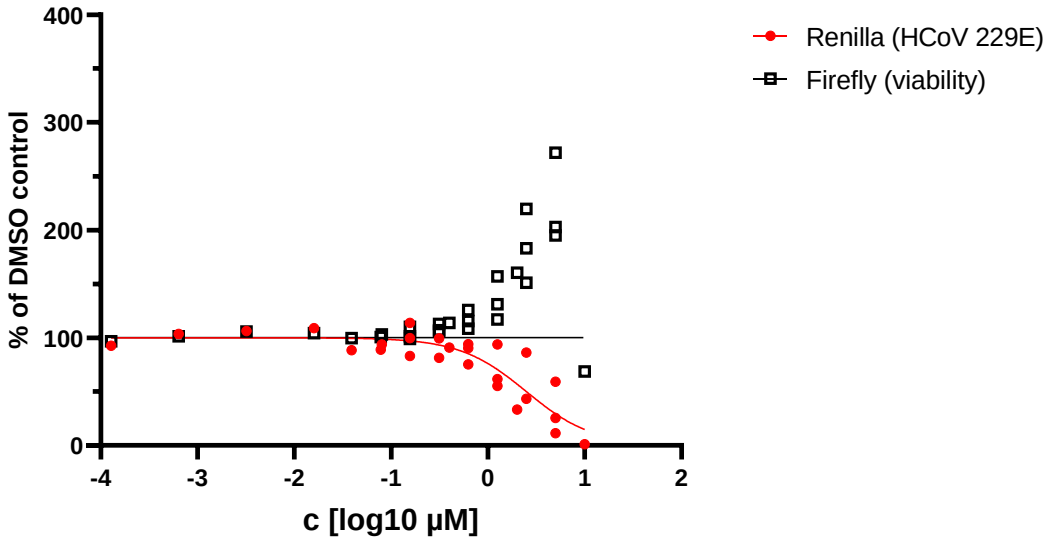


Figure 23: Inhibition of hCoV-229E-luc (red, IC_{50} = 2.47 µM) and effects on cell viability (black, CC_{50} >10 µM) of thiamyxin A.

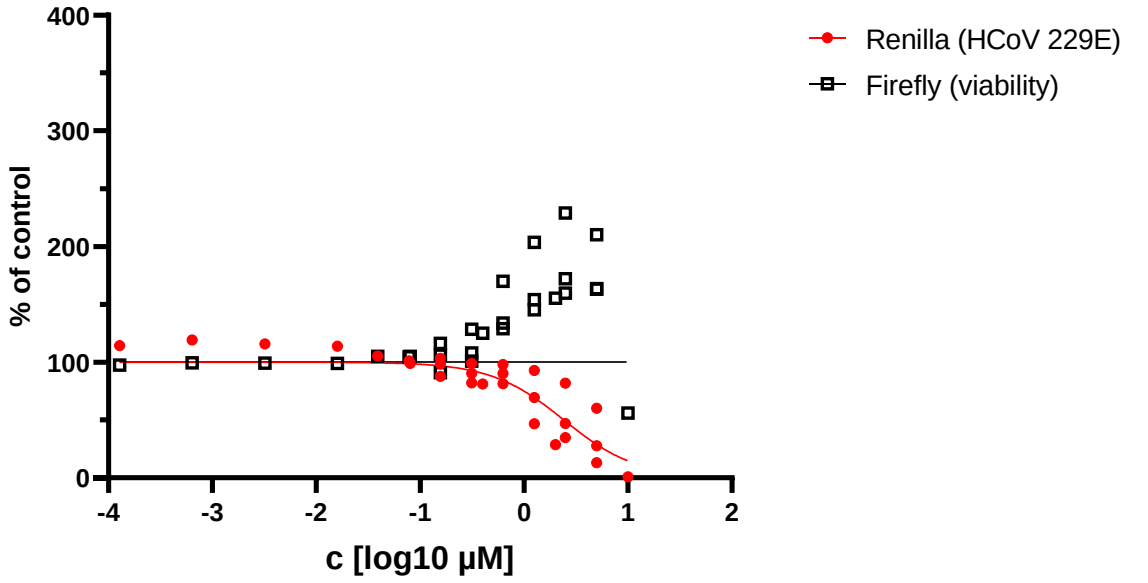


Figure 24: Inhibition of hCoV-229E-luc (red, IC_{50} = 2.39 µM) and effects on cell viability (black, CC_{50} >10 µM) of thiamyxin B.

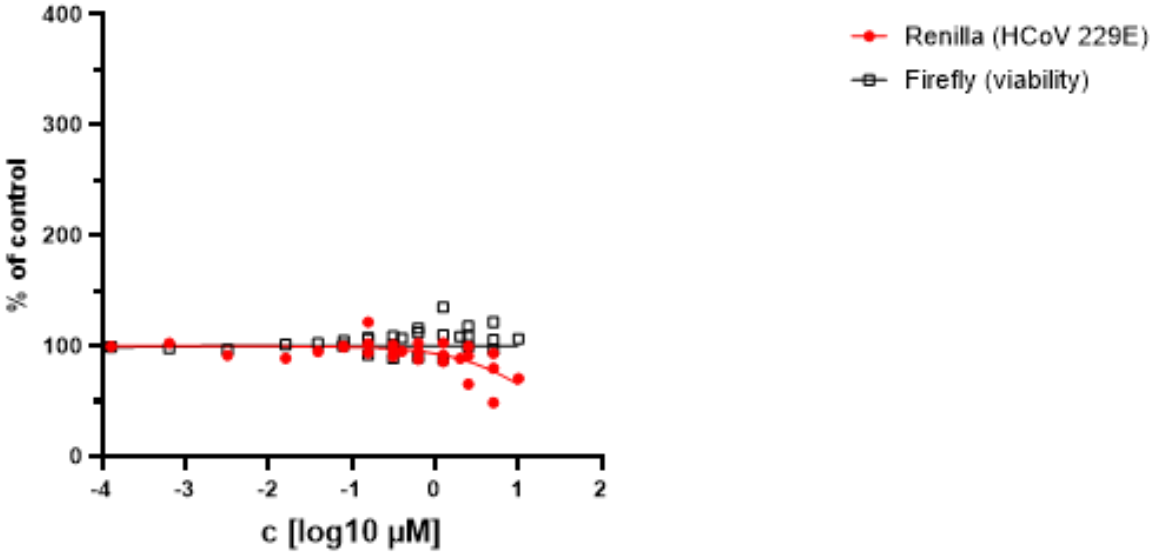


Figure 25: Inhibition of hCoV-229E-luc (red, IC_{50} ~ 20.52 µM) and effects on cell viability (black, CC_{50} >10 µM) of thiamyxin C.

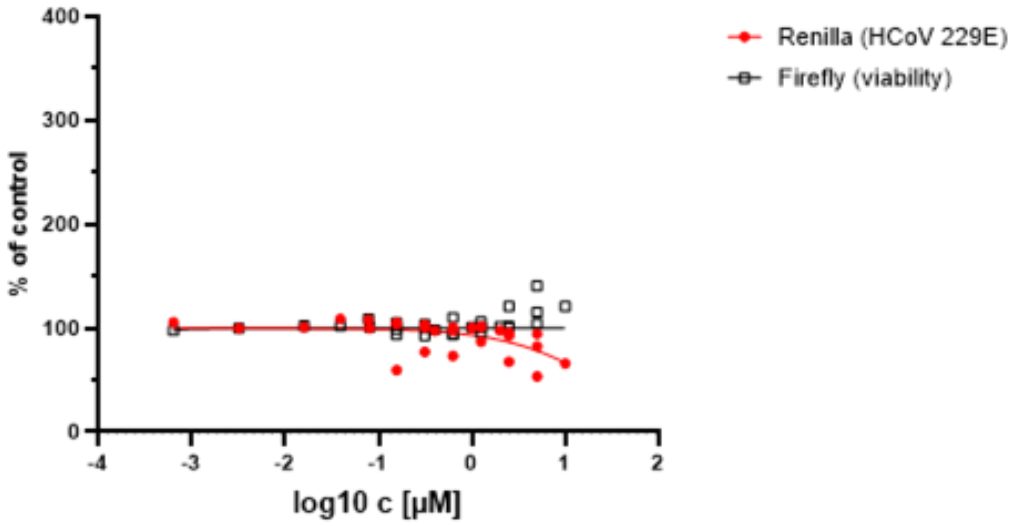


Figure 26: Inhibition of hCoV-229E-luc (red, IC_{50} ~ 23.06 µM) and effects on cell viability (black, CC_{50} >10 µM) of thiamyxin D.

S 2.2.8 NMR data

Thiamyxin B

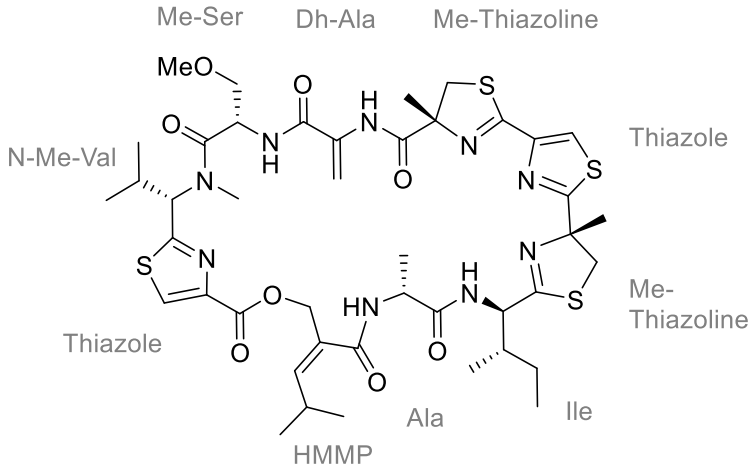


Table 6: NMR spectroscopic data of thiamyxin B in DMSO-d_6 at 700/175 MHz.

	NMR data in DMSO-*d_6*			
position	δ_C	δ_H (*J* in Hz)	COSY correlations	HMBC correlations
HMMP				
1	165.8	-	-	-
2	129.2	-	-	-
2’	67.0	4.90, 4.86, d (12.1)	3	1, 2, 3, *Thiazole*-1
3	146.6	5.83, s	2’, 4	1, 2, 2’,4, 4’, 5
4	27.8	2.83, dq (16.4, 6.7)	3, 4’, 5	2, 3, 4’, 5
4’	22.2	0.96, d (6.6)	4, 5	3, 4, 5
5	22.2	0.96, d (6.6)	4, 4’	3, 4, 4’
Ala				
1	171.6	-	-	-
2	47.6	4.61, quin (7.0)	3, NH	1, 3, *HMMP* -1
3	19.4	1.36, m*	2	1, 2
NH	-	7.51, bd (7.7)	2	-
Ile				
1	173.1	-	-	-

2	53.6	4.78, dd (9.4, 4.7)	3, NH	1, 3, 3', 4, *Ala*-1
3	37.5	1.91, m	2, 3', 4	1, 2, 3', 4, 5
3'	14.2	0.93, dd (10.5, 6.8)	3	2, 3, 4, 5
4	25.9	1.36, 1.21, m*	3, 5	3' ,3 ,5
5	11.4	0.87, t (7.4)	4	3, 4
NH	-	8.30, m*	2	2, *Ala*-1
Me-thiazoline				
1	176.5	-	-	-
2	83.3	-	-	-
3	43.3	3.74, dd (11.8, 39.7)	-	1, 2, Me, *Ile*-1
Me	28.3	1.65, s	-	1, 2, 3
Thiazole				
1	163.8	-	-	-
2	147.5	-	-	-
3	122.5	8.30, s*	-	1, 2, *Me-thiazoline*-1
Me-thiazoline				
1	*172.9*	-	-	-
2	*84.9*	-	-	-
3	*40.7*	*3.38, 3.81, d (11.7)*	-	*1, 2, Me,* Thiazole-1
Me	*25.5*	*1.59, s*	-	*1, 2, 3*
Dh-Ala				
1	163.1	-	-	-
2	132.7	-	-	-
3	102.6	6.37 *s*, 5.81 *s*	NH	1, 2
NH	-	9.36, s	3	1, 3, *Me-Thiazoline-1*
Me-Ser				
1	170.6	-	-	-
2	49.5	4.89, m*	3, NH	1, 3, *Dh-Ala*-1
3	71.0	3.59, m	2	1, 3, OMe
OMe	58.4	3.19, s	-	3
NH	-	8.97	2	2, 3, *Dh-Ala*-1

N-Me-Val				
1	168.3	-	-	-
2	59.7	5.31, bd (8.6)	3	nd
3	27.8	2.36, m	2, 3‘, 4	4, nd
3‘	18.3	0.38, d (6.4)	4	2, 3, 4
4	20.1	0.71, d (6.4)	3‘	2, 3, 3‘
NMe	30.2	2.90, s	-	2, *Me-Ser*-1
Thiazole				
1	160.5	-	-	-
2	145.2	-	-	-
3	130.6	8.57, s	-	1, 2, *N-Me-Val*-1

* Overlapping signals circumvent exact assignment; nd = not detectable

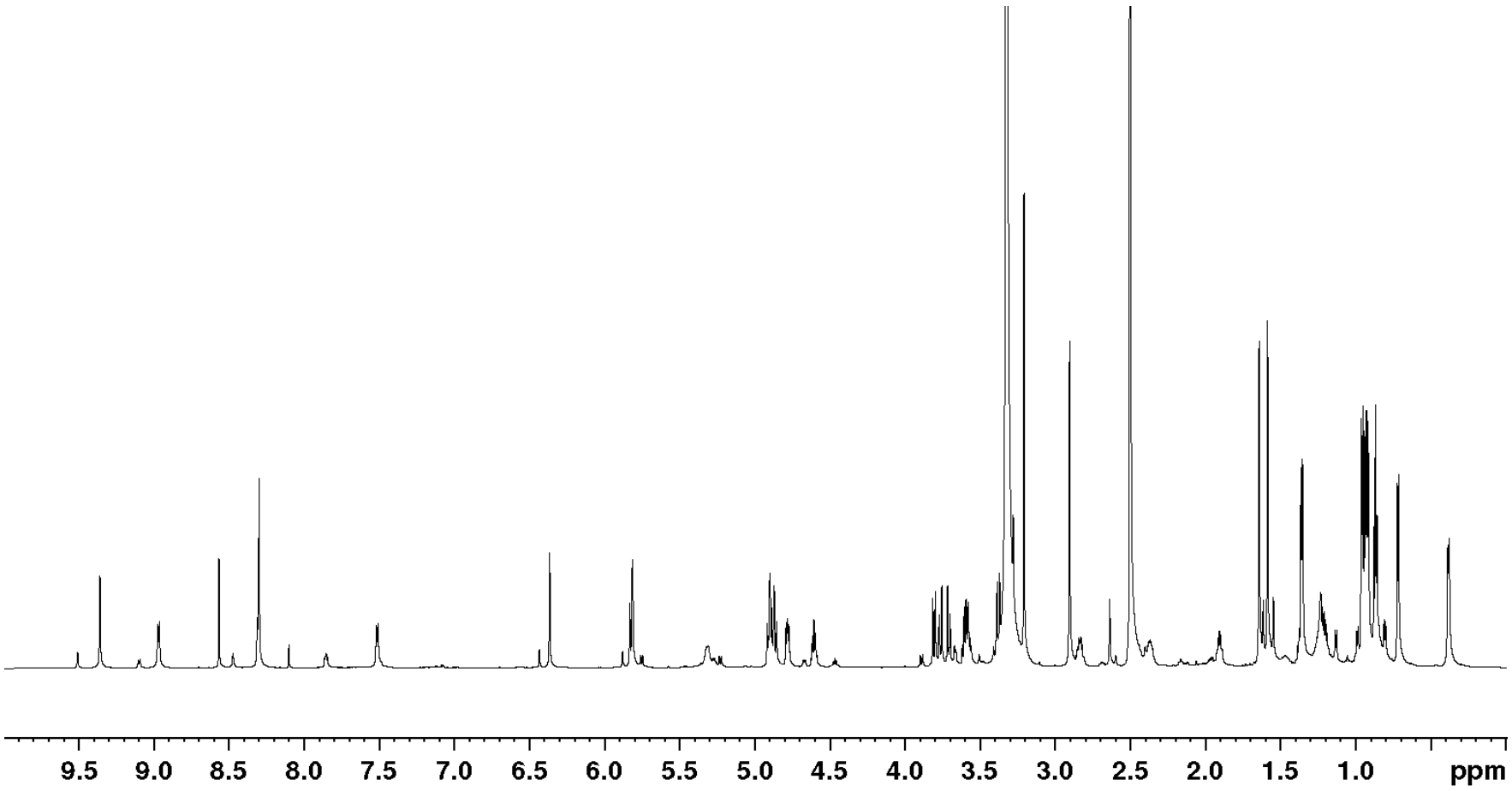


Figure 27: ^{1}H spectrum of thiamyxin B in DMSO-d_6 at 700 MHz.

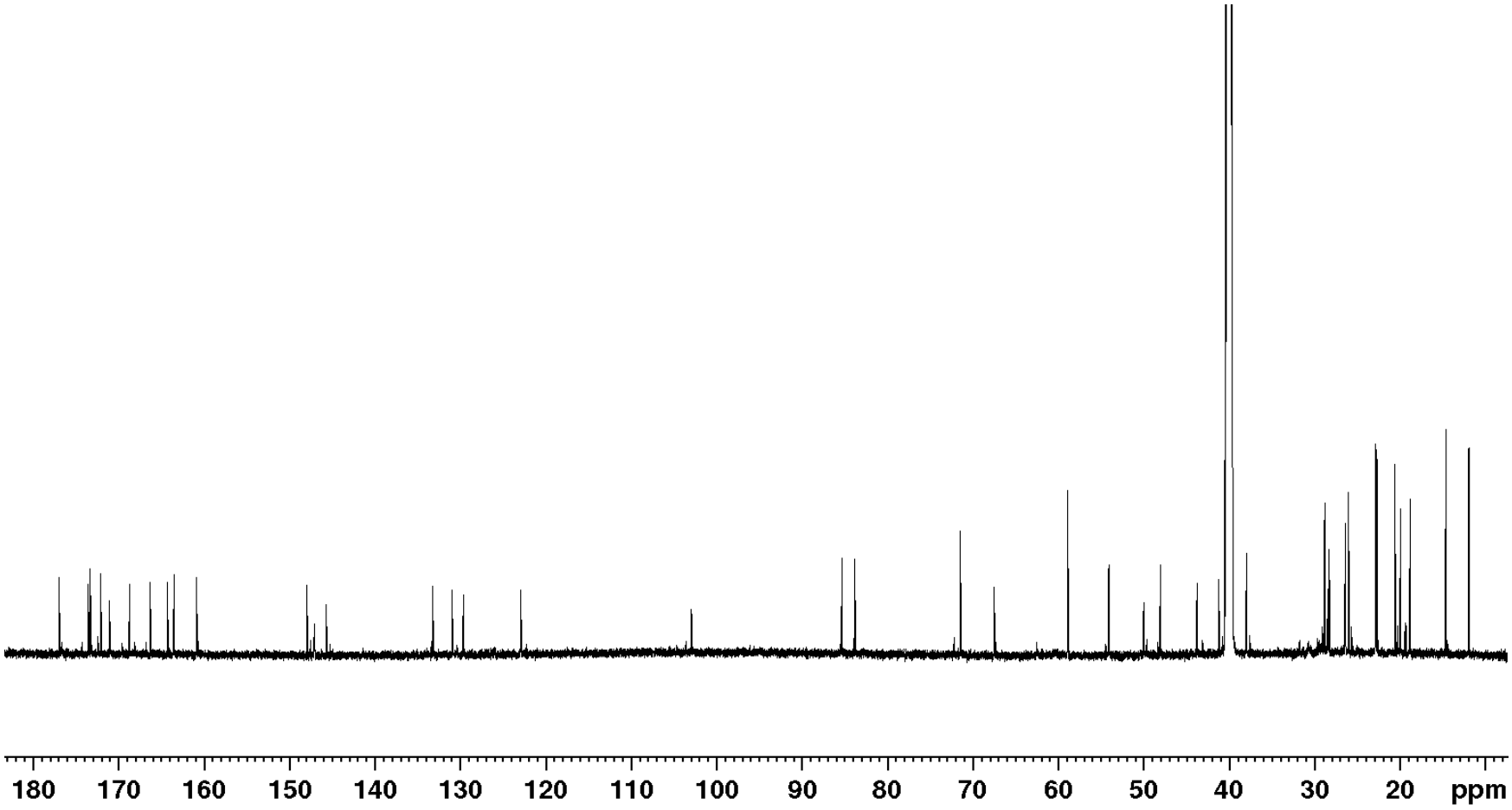


Figure 28: ^{13}C spectrum of thiamyxin B in DMSO-d_6 at 175 MHz.

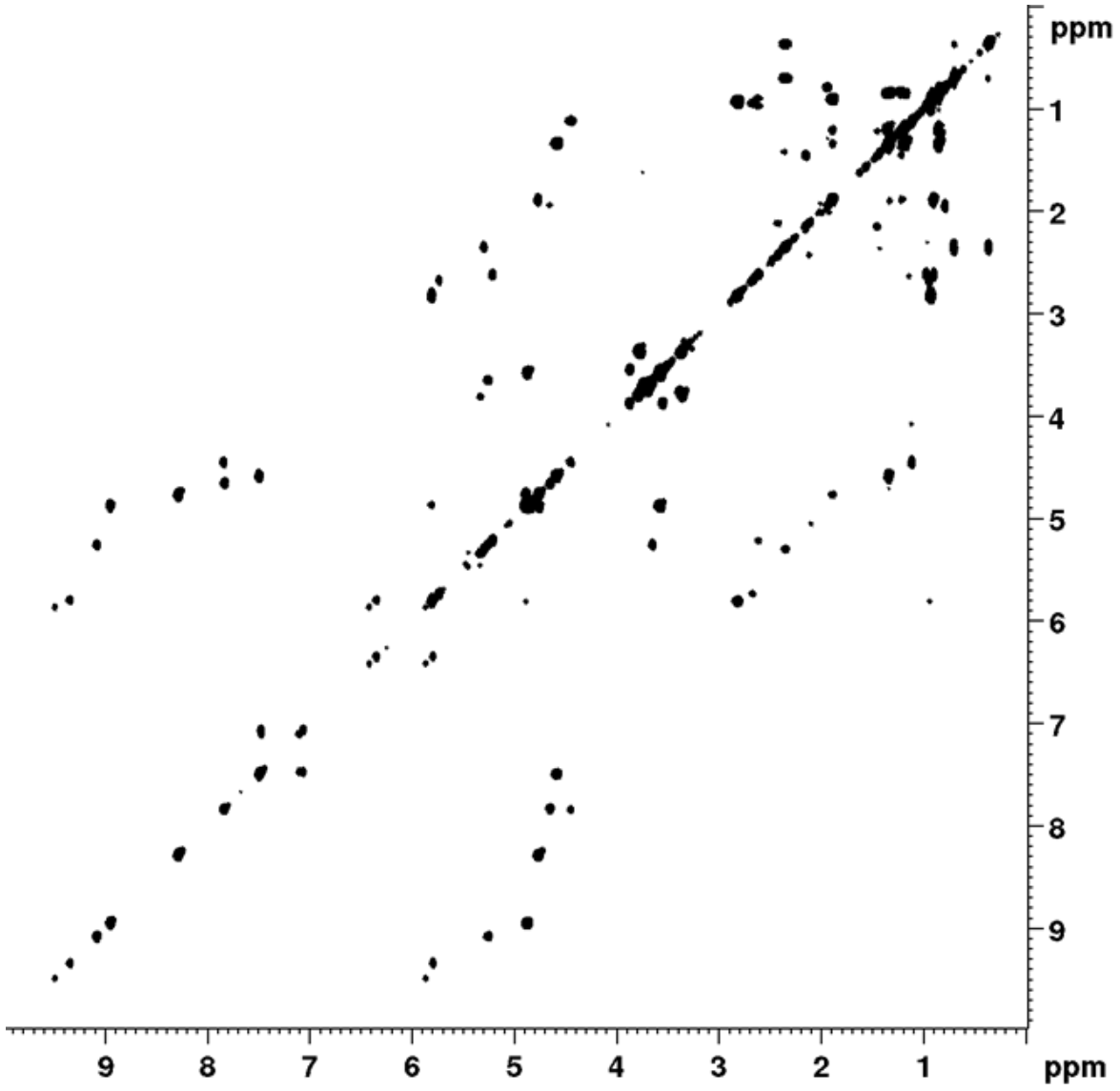


Figure 29: COSY spectrum of thiamyxin B in DMSO-d_6 at 700 MHz.

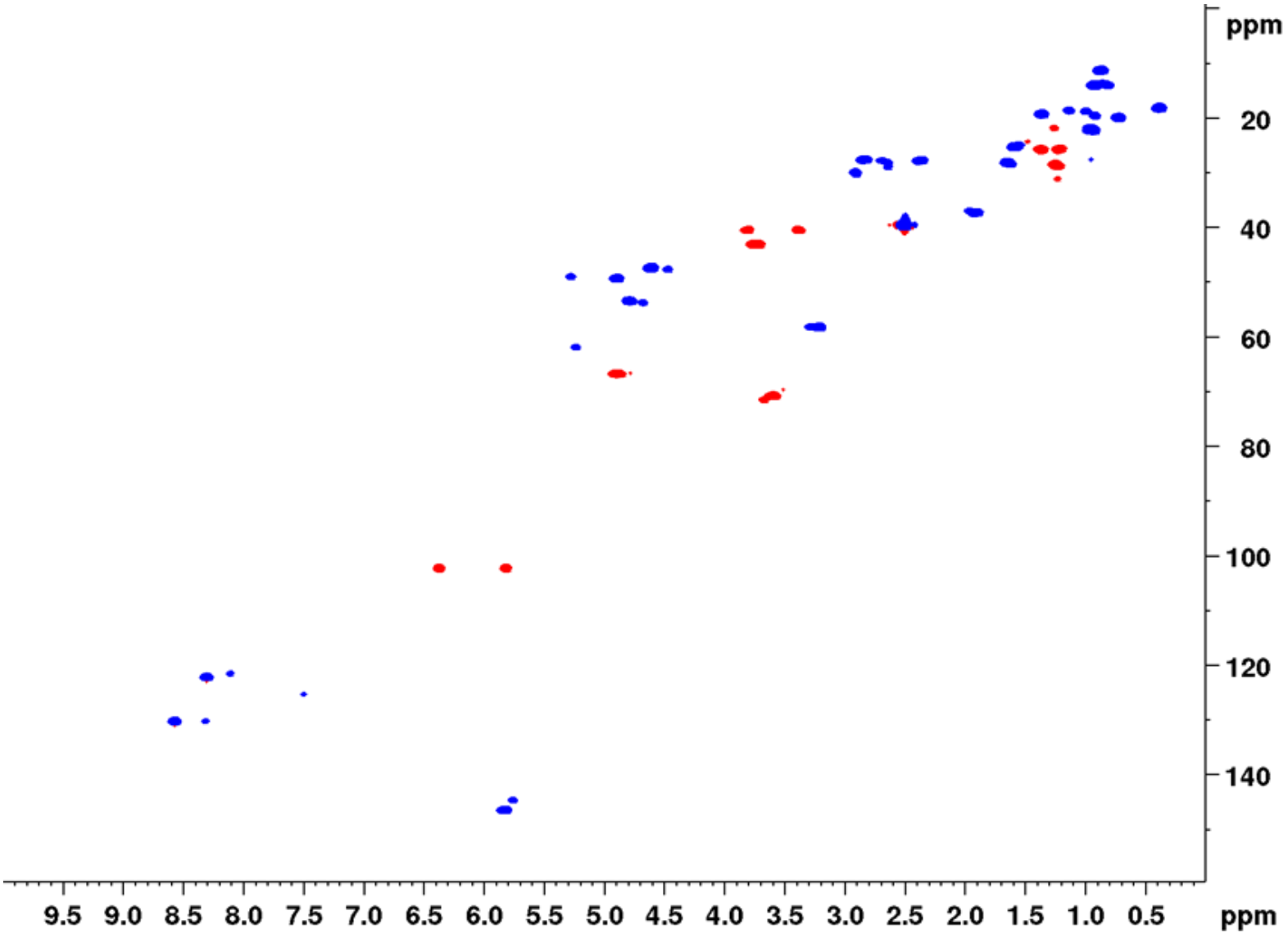


Figure 30: HSQC spectrum of thiamyxin B in DMSO-d_6 at 700/175 MHz.

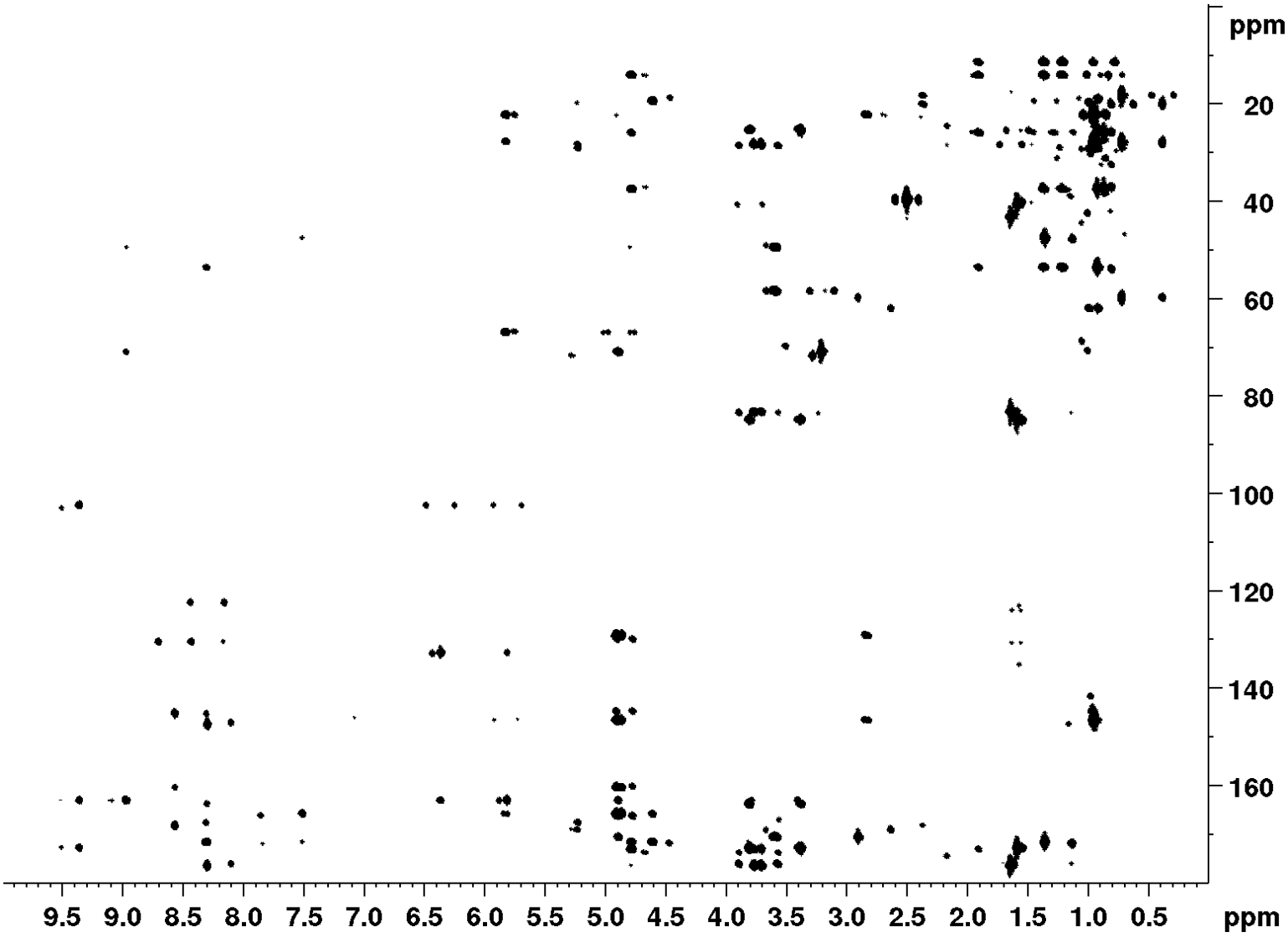


Figure 31: HMBC spectrum of thiamyxin B in DMSO-d_6 at 700/175 MHz.

Thiamyxin A

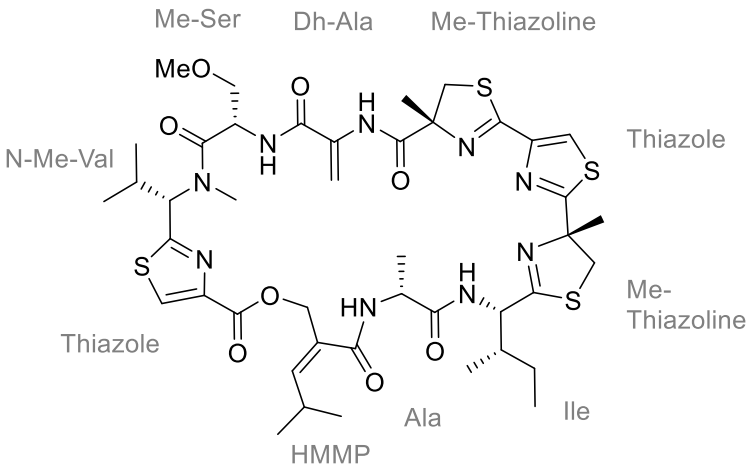


Table 7: NMR spectroscopic data of thiamyxin A in DMSO-d_6 at 500/125 MHz.

	NMR data in DMSO-*d*$_6$			
position	δ_C	δ_H (*J* in Hz)	COSY correlations	HMBC correlations
HMMP				
1	165.6	-	-	-
2	129.8	-	-	-
2’	65.9	4.81, 4.92, d (12.8)	3	1, 2, 3, *Thiazole*-1
3	142.6	5.76, d (10.0)	2’, 4	1, 2, 2’,4, 4’, 5
4	27.6	2.70, m	3, 4’, 5	2, 3, 4’, 5
4’	22.4	0.92, m*	4, 5	3, 4, 5
5	22.4	0.92, m*	4, 4’	3, 4, 4’
Ala				
1	171.6	-	-	-
2	48.4	4.42, quin (6.8)	3, NH	1, 3, *HMMP* -1
3	19.0	1.20, bd (7.0)	2	1, 2
NH	-	7.83, d (8.0)	2	-
Ile				
1	174.6	-	-	-
2	55.8	4.64, dd (8.8, 8.8)	3, NH	1, 3, 3’, 4, *Ala*-1
3	37.0	1.82, m	2, 3’, 4	1, 2, 3’, 4, 5
3’	15.1	0.91, dd (10.5, 6.8)	3	2, 3, 4, 5
4	24.3	1.49, 1.19, m*	3, 5	3‘ ,3 ,5

5	10.4	0.85, t (7.4)	4	3, 4
NH	-	8.26, d (8.8)	2	2, *Ala*-1
Me-thiazoline				
1	176.1	-	-	-
2	82.8	-	-	-
3	43.4	3.63, 3.50, m*	-	1, 2, Me, *Ile*-1
Me	27.2	1.67, s	-	1, 2, 3
Thiazole				
1	163.7	-	-	-
2	147.9	-	-	-
3	121.9	8.26, s*	-	1, 2, *Me-thiazoline*-1
Me-thiazoline				
1	*172.8*	-	-	-
2	*84.8*	-	-	-
3	*40.6*	*3.37, 3.78, d (11.6)*	-	*1, 2, Me,* Thiazole-1
Me	*25.2*	*1.56, s*	-	*1, 2, 3*
Dh-Ala				
1	162.9	-	-	-
2	132.9	-	-	-
3	102.5	6.38, 5.79, s	NH	1, 2
NH	-	9.28, s	3	1, 3, *Me-Thiazoline-1*
Me-Ser				
1	170.6	-	-	-
2	49.4	4.88, m*	3, NH	1, 3, *Dh-Ala*-1
3	70.8	3.58, m	2	1, 3, OMe
OMe	58.3	3.20, s	-	3
NH	-	8.93	2	2, 3, *Dh-Ala*-1
N-Me-Val				
1	168.1	-	-	-
2	61.7	5.05, bd (10.6)	3	1, 3, 3‘, 4
3	27.8	2.31, m	2, 3‘, 4	1, 3‘, 4
3‘	18.3	0.54, d (6.6)	4	2, 3, 4
4	19.7	0.74, d (6.4)	3‘	2, 3, 3‘
NMe	30.0	2.82, s	-	2, *Me-Ser*-1

Thiazole				
1	160.4	-	-	-
2	144.9	-	-	-
3	130.3	8.54, s	-	1, 2, *N-Me-Val*-1

* overlapping signals circumvent exact assignment; nd = not detectable

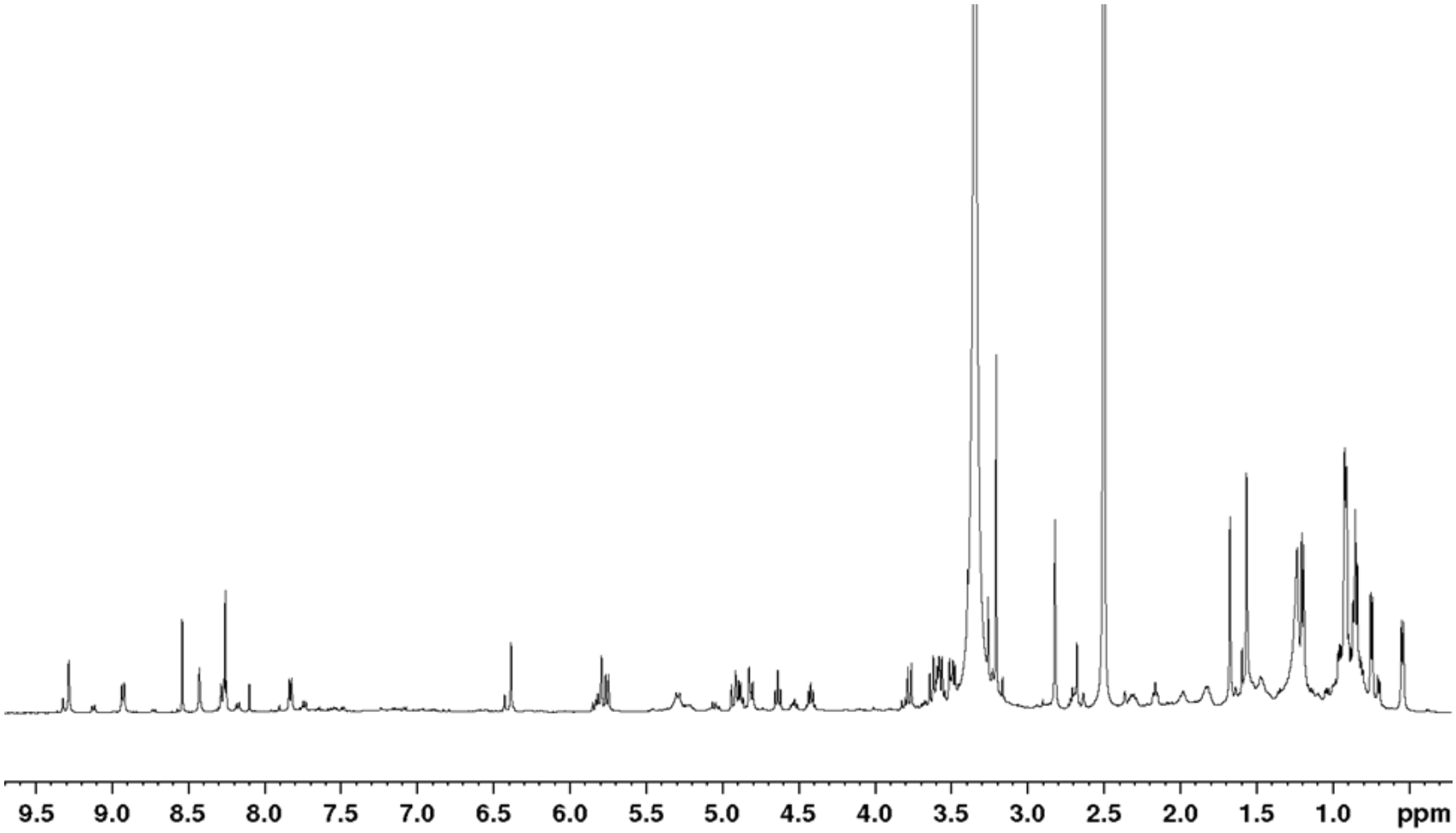


Figure 32: ^{1}H spectrum of thiamyxin A in DMSO-d_6 at 500 MHz.

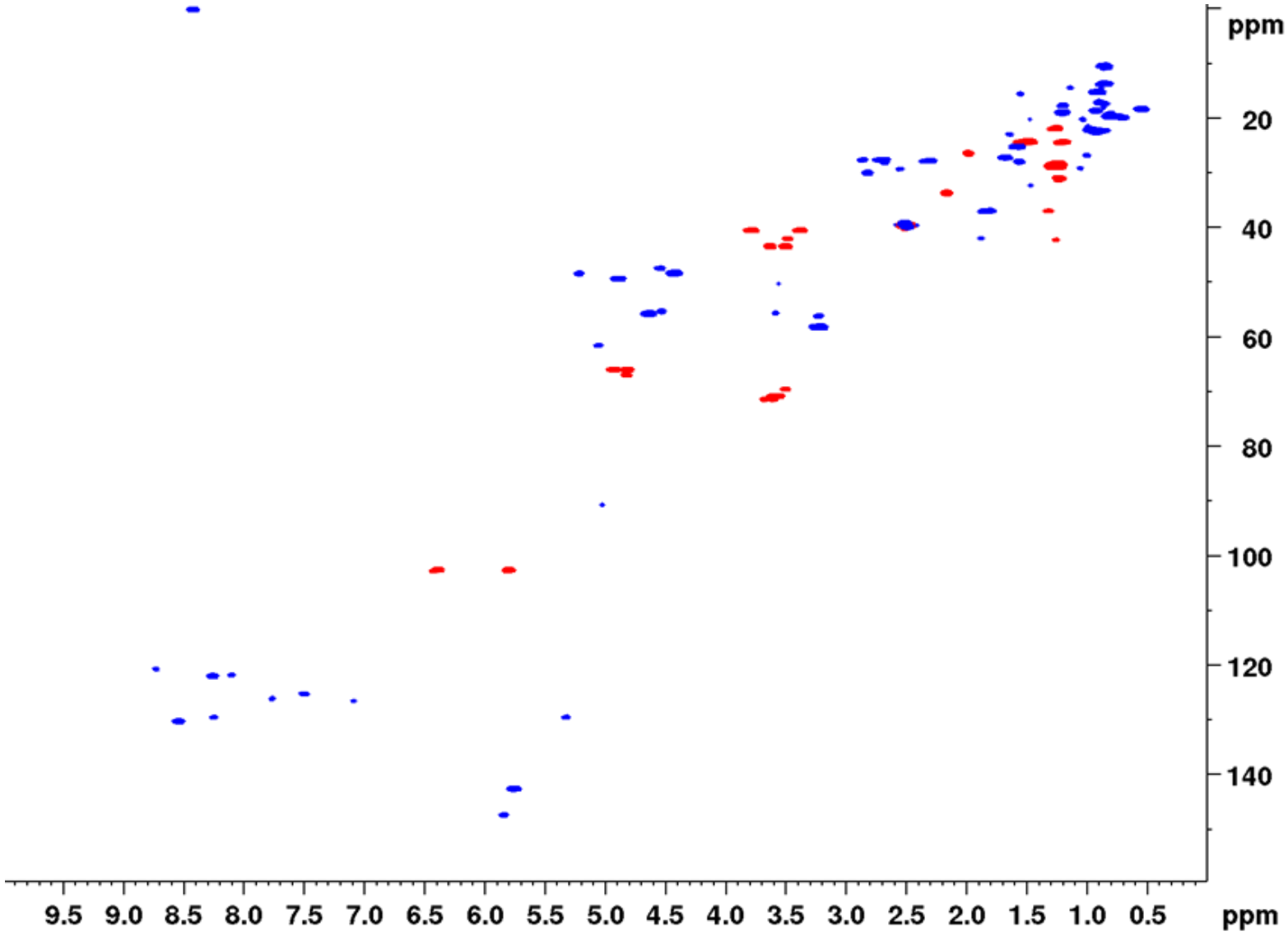


Figure 33: HSQC spectrum of thiamyxin A in DMSO-d_6 at 500/125 MHz.

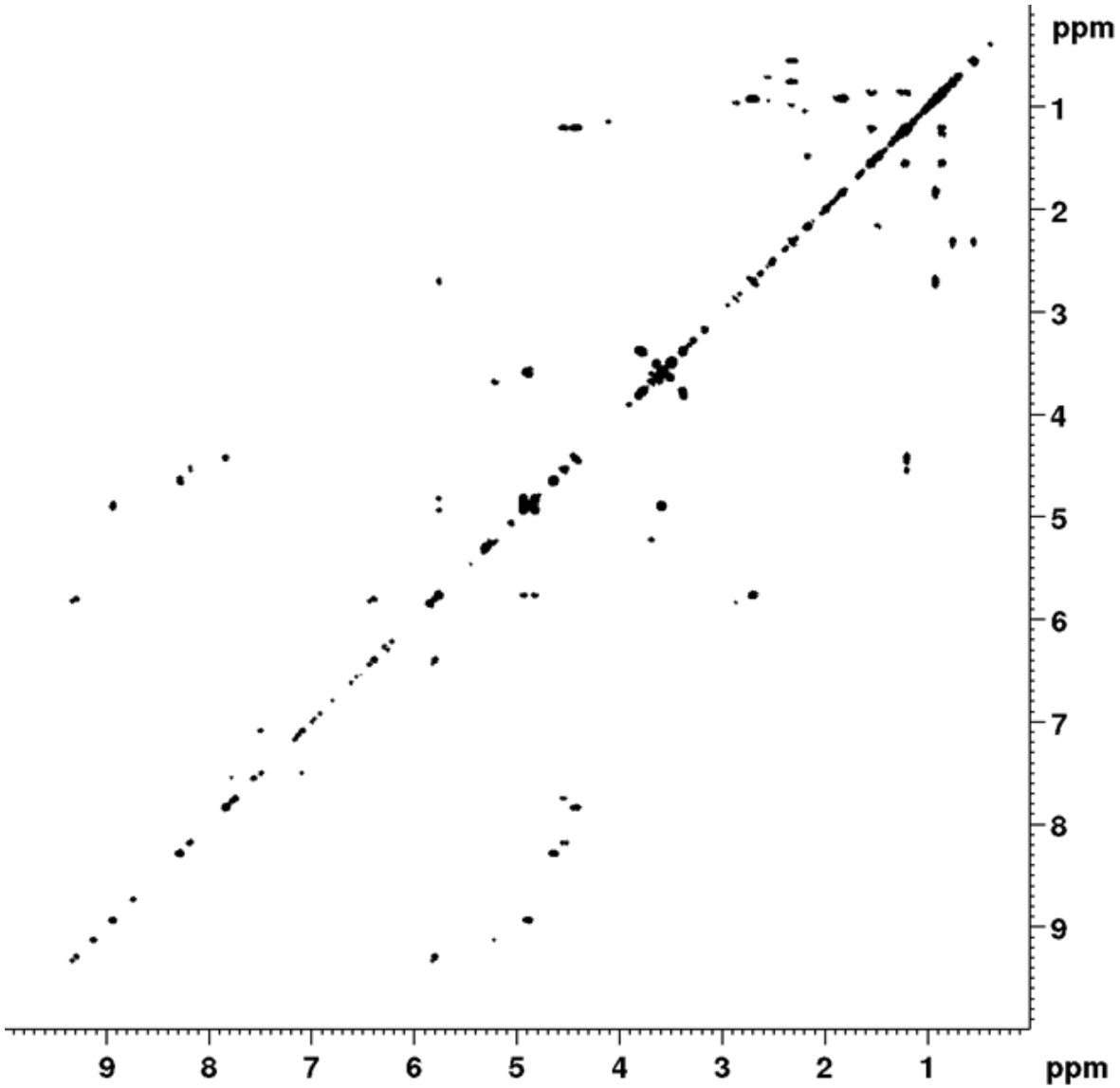


Figure 34: COSY spectrum of thiamyxin A in DMSO-d_6 at 500 MHz.

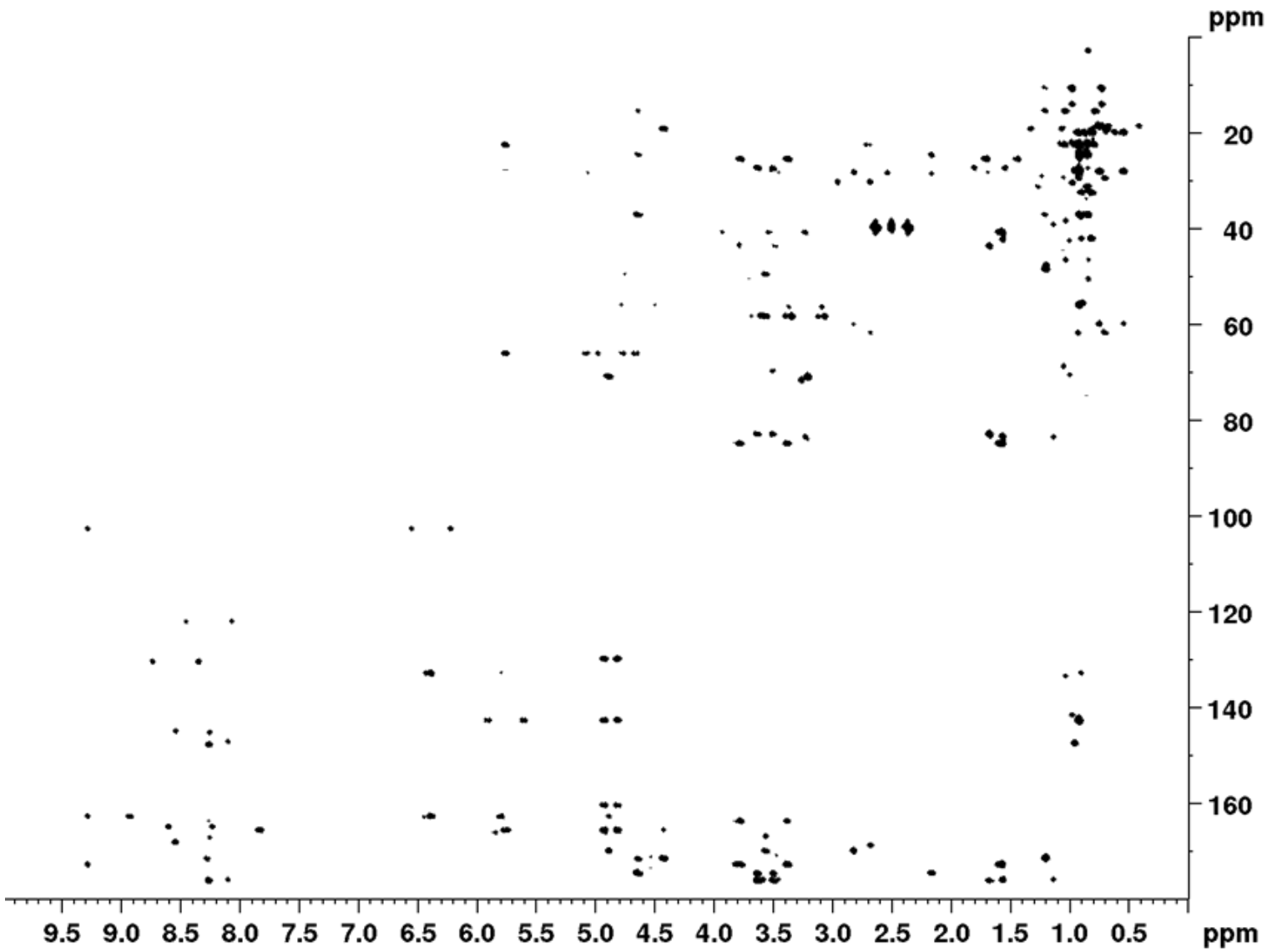


Figure 35: HMBC spectrum of thiamyxin A in DMSO-d_6 at 500/125 MHz.

Thiamyxin C

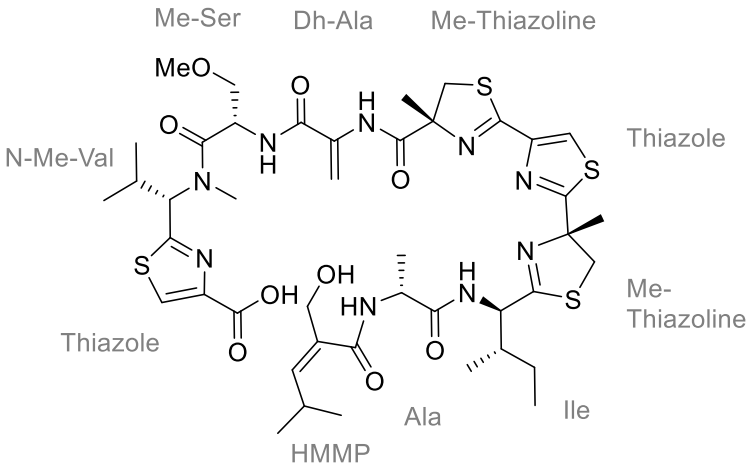


Table 8: NMR spectroscopic data of thiamyxin C in DMSO-d_6 at 700/175 MHz.

	NMR data in DMSO-d_6[xx]			
position	δ_C	δ_H (*J* in Hz)	COSY correlations	HMBC correlations
HMMP				
1	167.6	-	-	-
2	134.6	-	-	-
2’	63.1	4.00, m	3	1, 2, 3
3	139.8	5.43, bd (9.8)	2’, 4	1, 2, 2’,4, 4’, 5
4	27.3	2.81, m	3, 4’, 5	2, 3, 4’, 5
4’	22.7	0.90, m*	4, 5	3, 4, 5
5	22.7	0.90, m*	4, 4’	3, 4, 4’
Ala				
1	172.8	-	-	-
2	47.9	4.52, quin (7.3)	3, NH	1, 3, *HMMP* -1
3	18.4	1.30, q (7.4)	2	1, 2
NH	-	7.98, bs	2	-
Ile				
1	175.5	-	-	-
2	54.2	4.77, dd (8.9, 4.9)	3, NH	1, 3, 3’, 4, *Ala*-1
3	37.7	2.01, m	2, 3’, 4	2, 3’, 4, 5
3’	14.4	0.92, m*	3	2, 3, 4, 5
4	25.7	1.38, 1.20, m*	3, 5	3‘ ,3 ,5

5	11.4	0.85, t (7.8)	4	3, 4
NH	-	8.27, bd (8.8)	2	*Ala*-1
Me-thiazoline				
1	176.7	-	-	-
2	83.7	-	-	-
3	43.1	3.62, 3.51, m*	-	1, 2, Me, *Ile*-1
Me	27.4	1.64, s	-	1, 2, 3
Thiazole				
1	163.1	-	-	-
2	147.7	-	-	-
3	122.2	8.21, s*	-	2, *Me-thiazoline*-1
Me-thiazoline				
1	*172.8*	-	-	-
2	*85.0*	-	-	-
3	*40.4*	*3.39, 3.76, m**	-	*1, 2, Me,* Thiazole-1
Me	*25.9*	*1.54, s*	-	*1, 2, 3*
Dh-Ala				
1	163.1	-	-	-
2	133.2	-	-	-
3	103.0	6.37, 5.79, s	NH	1, 2
NH	-	9.30, s	3	1, 3, *Me-Thiazoline-1*
Me-Ser				
1	170.0	-	-	-
2	50.0	4.89, m*	3, NH	1, 3, *Dh-Ala*-1
3	70.8	3.56, m	2	1, 3, OMe
OMe	58.3	3.19, s	-	3
NH	-	8.85, t (7.8)	2	3, *Dh-Ala*-1
N-Me-Val				
1	172.2	-	-	-
2	59.8	5.36, bd (10.6)	3	n.d.
3	27.9	2.46, m	2, 3‘, 4	n.d.
3‘	18.9	0.72, t (7.8)	4	2, 3, 4
4	20.0	0.82, m*	3‘	2, 3, 3‘
NMe	29.8	2.92, s	-	2, *Me-Ser*-1

Thiazole				
1	176.4	-	-	-
2	147.7	-	-	-
3	122.3	8.19, s	-	1, 2

[xx] main diastereoisomer (D-*allo*-Ile carrying derivative) was analyzed; * overlapping signals circumvent exact assignment; n.d. = not detectable

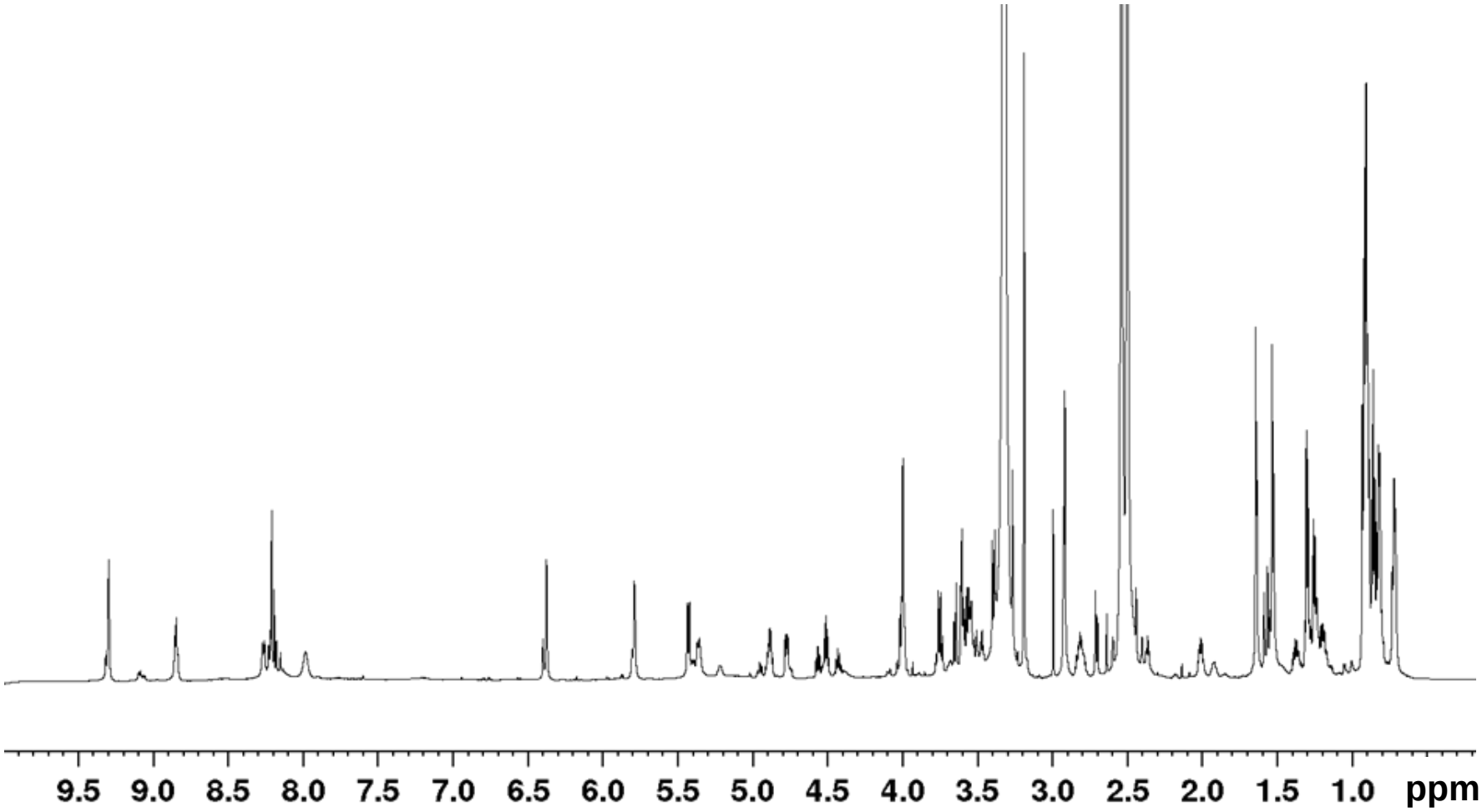


Figure 36: ^{1}H spectrum of thiamyxin C in DMSO-d_6 at 700 MHz.

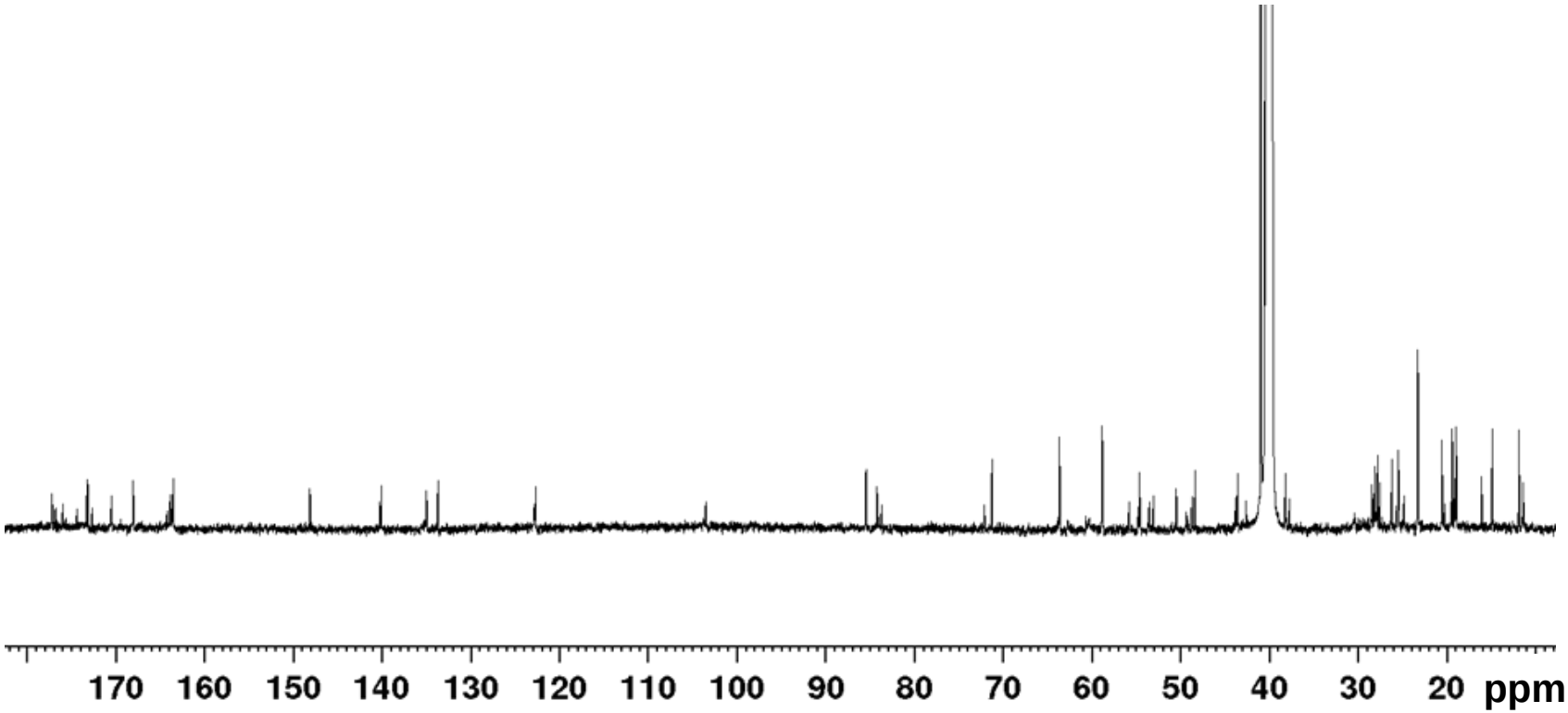


Figure 37: ^{13}C spectrum of thiamyxin C in DMSO-d_6 at 175 MHz.

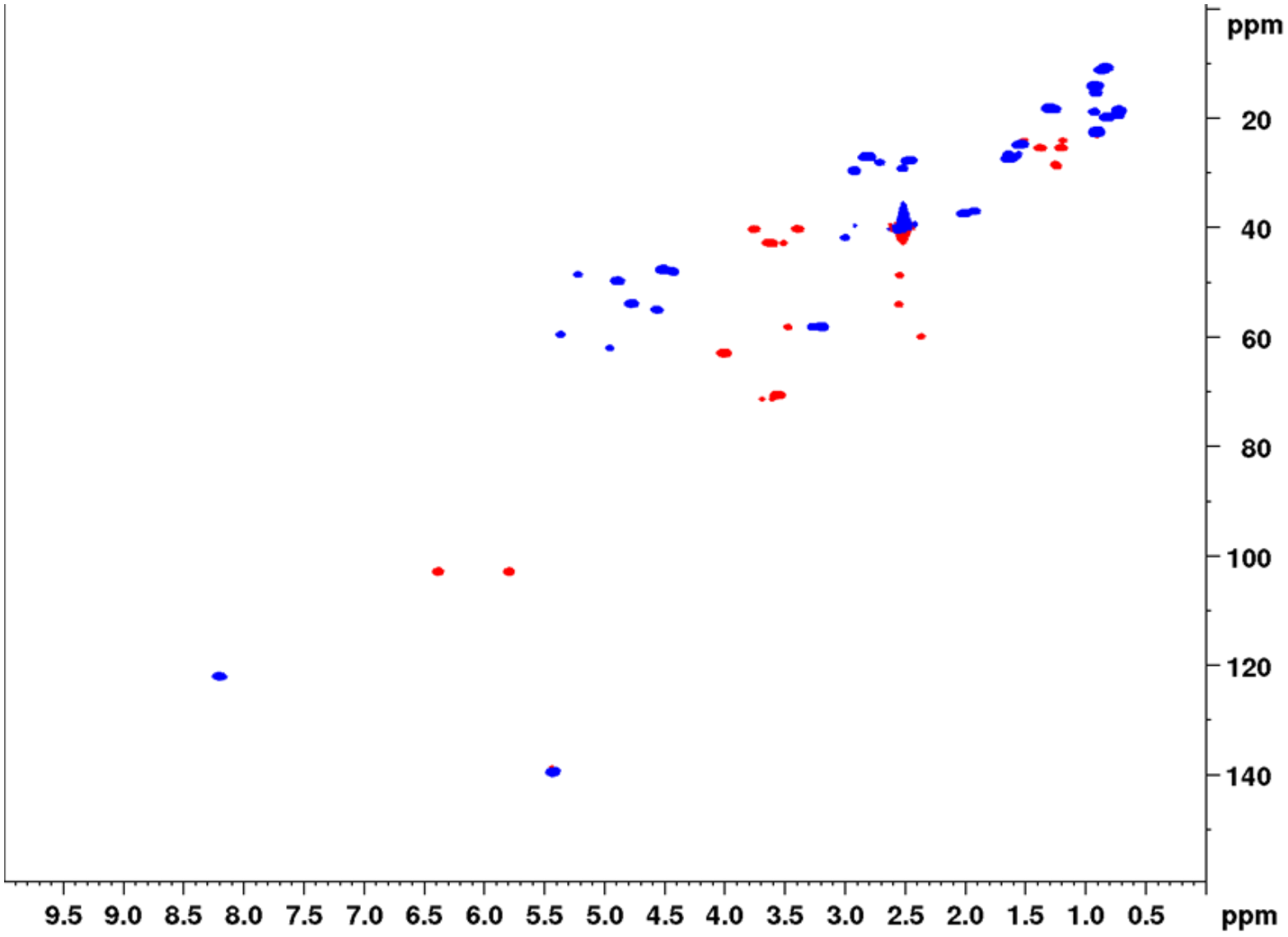


Figure 38: HSQC spectrum of thiamyxin C in DMSO-d_6 at 700/175 MHz.

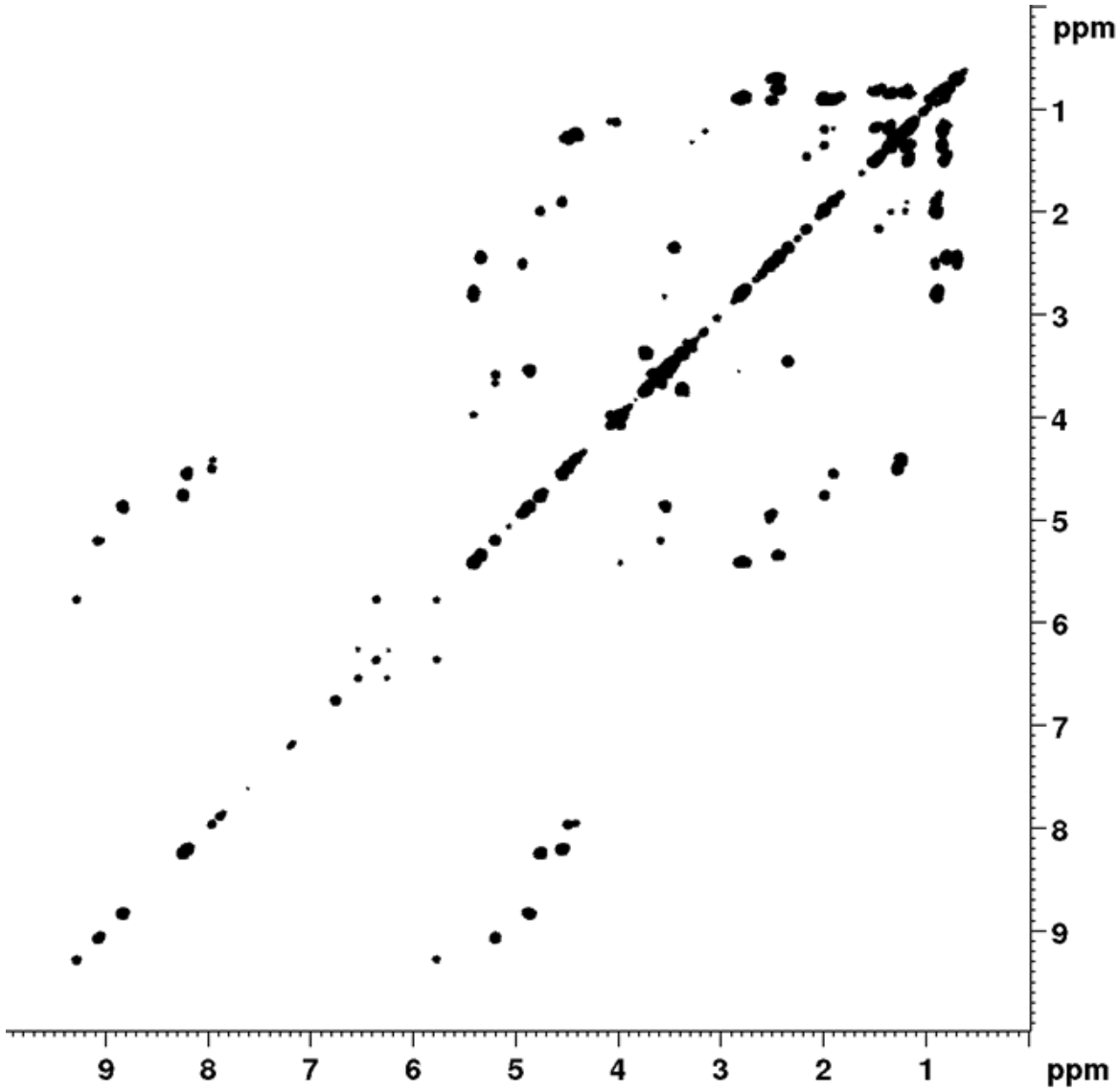


Figure 39: COSY spectrum of thiamyxin C in DMSO-d_6 at 700 MHz.

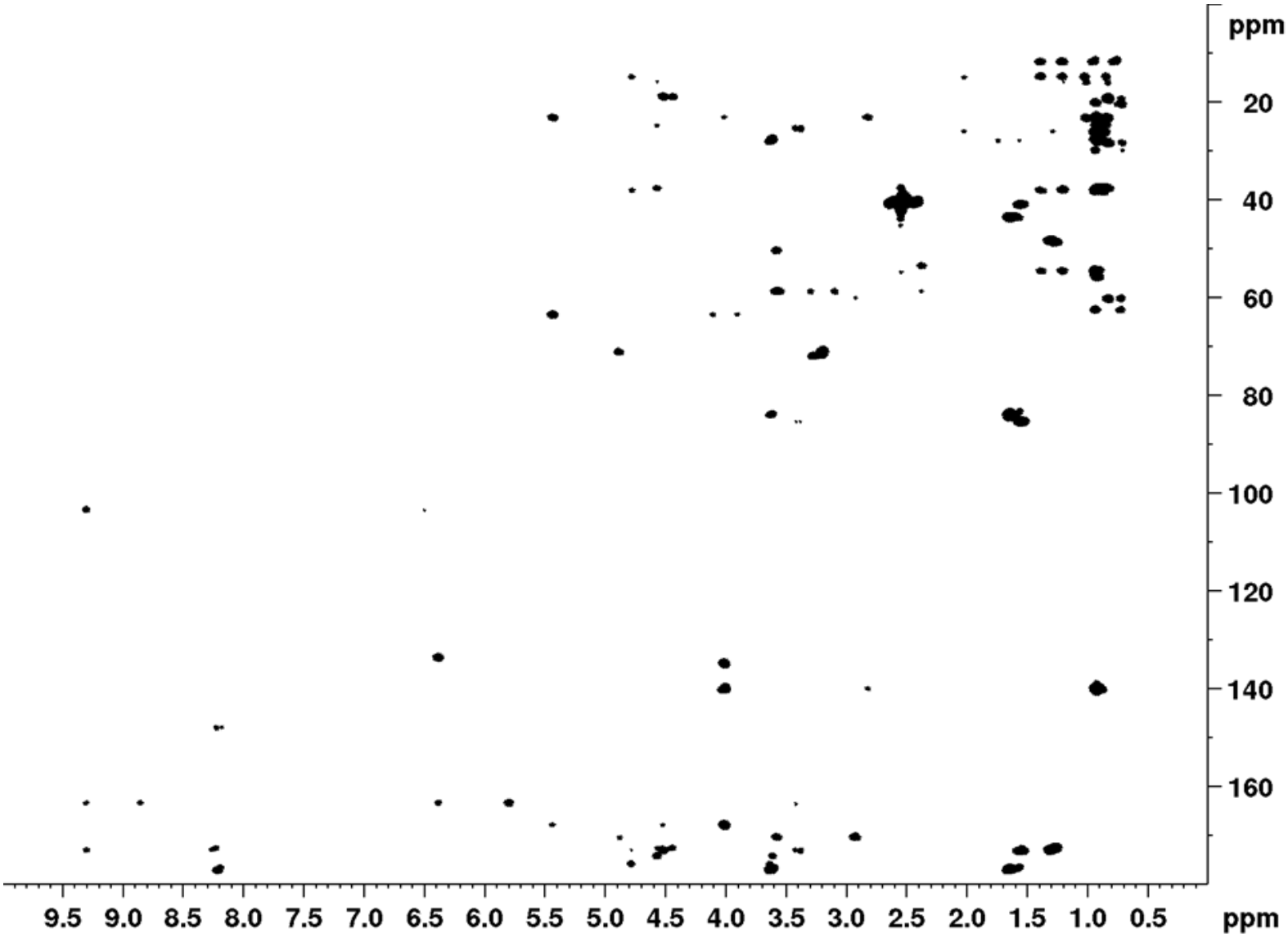


Figure 40: HMBC spectrum of thiamyxin C in DMSO-d_6 at 700/175 MHz.

Thiamyxin D

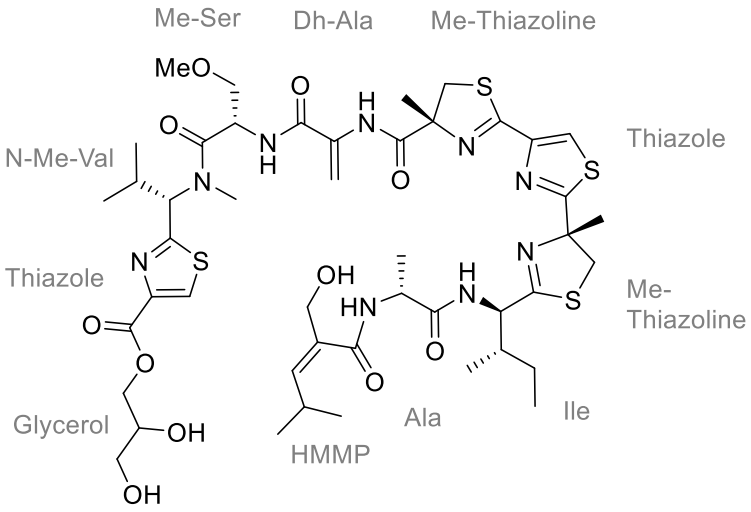


Table 9: NMR spectroscopic data of thiamyxin D in DMSO-d_6 at 700/175 MHz.

	NMR data in DMSO-d_6			
position	δ_C	δ_H (*J* in Hz)	COSY correlations	HMBC correlations
HMMP				
1	167.2	-	-	-
2	134.2	-	-	-
2’	63.0	4.00, s	3	1, 2, 3
3	139.5	5.45, bd (9.7)	2’, 4	1, 2, 2’,4, 4’, 5
4	27.0	2.83, m	3, 4’, 5	2, 3, 4’, 5
4’	22.4	0.91, m*	4, 5	3, 4, 5
5	22.4	0.91, m*	4, 4’	3, 4, 4’
Ala				
1	172.6	-	-	-
2	47.6	4.51, quin (7.3)	3, NH	1, 3, *HMMP*-1
3	18.0	1.29, m*	2	1, 2
NH	-	7.85, bd (7.4)	2	-
Ile				
1	175.0	-	-	-
2	53.8	4.76, m*	3, NH	1, 3, 3’, 4, *Ala*-1
3	37.4	2.00, m	2, 3’, 4	2, 3’, 4, 5
3’	16.4	0.90, m*	3	2, 3, 4, 5

4	25.4	1.37, 1.21, m*	3, 5	3‘ ,3 ,5
5	11.0	0.89, m*	4	3, 4
NH	-	8.21, m*	2	*Ala*-1
Me-thiazoline				
1	176.6	-	-	-
2	89.8	-	-	-
3	42.8	3.62, 3.51, m*	-	1, 2, Me, *Ile*-1
Me	22.1	1.14, s	-	1, 2, 3
Thiazole				
1	163.1	-	-	-
2	147.3	-	-	-
3	121.9	8.19, s*	-	2, *Me-thiazoline*-1
Me-thiazoline				
1	*172.6*	-	-	-
2	*84.8*	-	-	-
3	*40.4*	*3.40, 3.75, m**	-	*1, 2, Me*, Thiazole-1
Me	*24.7*	*1.07, s*	-	*1, 2, 3*
Dh-Ala				
1	162.7	-	-	-
2	132.9	-	-	-
3	102.8	6.38, 5.78, s	NH	1, 2
NH	-	9.30, s	3	1, 3, *Me-Thiazoline-1*
Me-Ser				
1	169.8	-	-	-
2	49.7	4.88, m*	3, NH	1, 3, *Dh-Ala*-1
3	70.7	3.49, 3.65, m	2	1, 3, OMe
OMe	58.0	3.19, s	-	3
NH	-	8.84, t (7.8)	2	3, *Dh-Ala*-1
N-Me-Val				
1	168.0	-	-	-
2	59.5	5.41, bd (10.8)	3	1, 3, 3‘, 4, NMe, *Me-Ser*-1
3	27.7	2.47, m*	2, 3‘, 4	1, 2, 3‘, 4
3‘	18.7	0.76, s	4	2, 3, 4
4	19.8	0.86, m*	3‘	2, 3, 3‘

NMe	29.8	2.94, s	-	1, 2, *Me-Ser*-1
Thiazole				
1	160.4	-	-	-
2	145.1	-	-	-
3	129.8	8.52, s	-	1, 2, *N-Me-Val*-1
Glycerol				
1	66.1	4.16, 4.28	2	2, 3, *Thiazole*-1
2	69.2	3.76	1, 3	1, 3
3	62.3	3.41	2	1, 2

* overlapping signals circumvent exact assignment; n.d. = not detectable

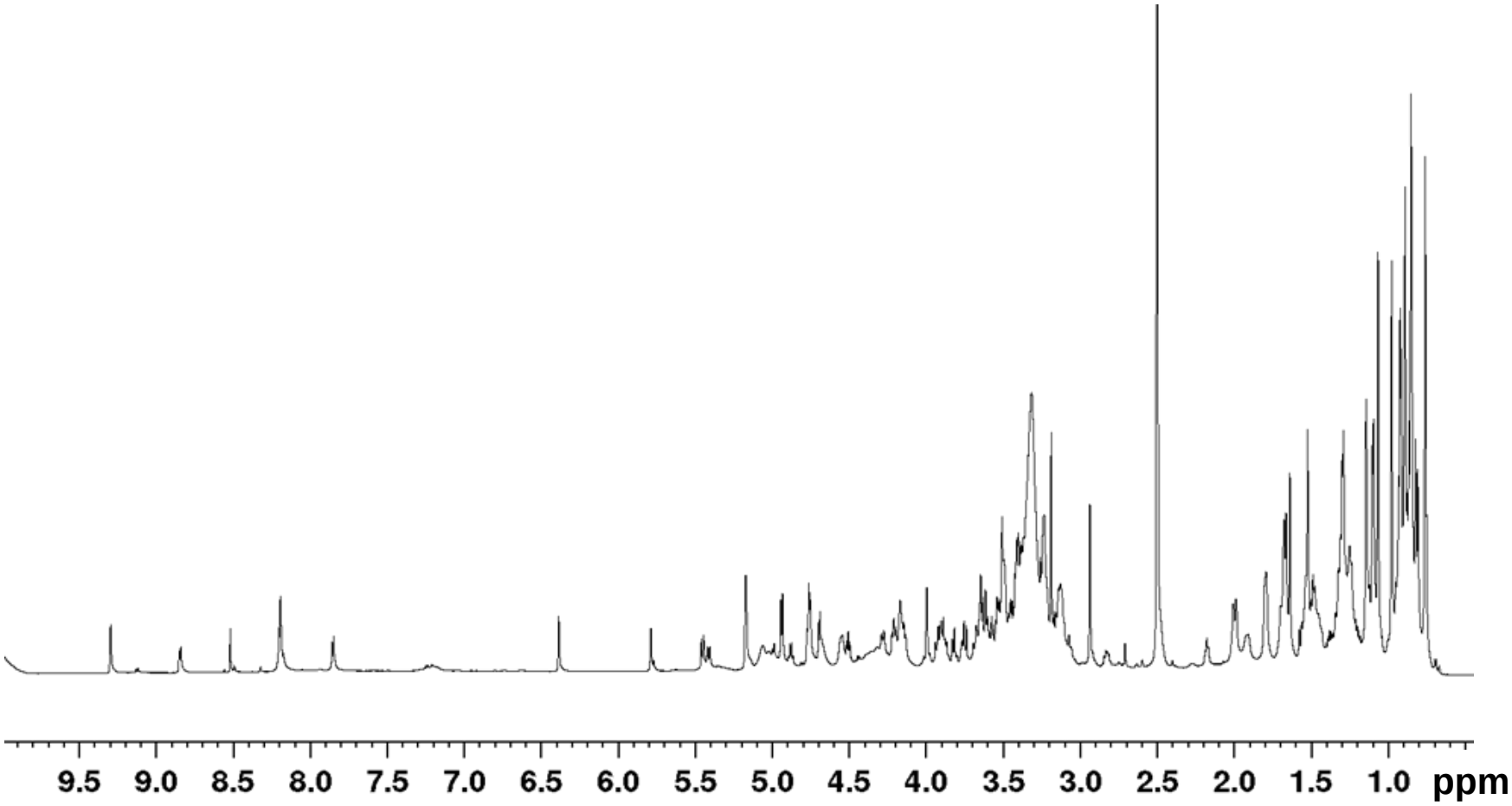


Figure 41: ^{1}H spectrum of thiamyxin D in DMSO-d_6 at 700 MHz.

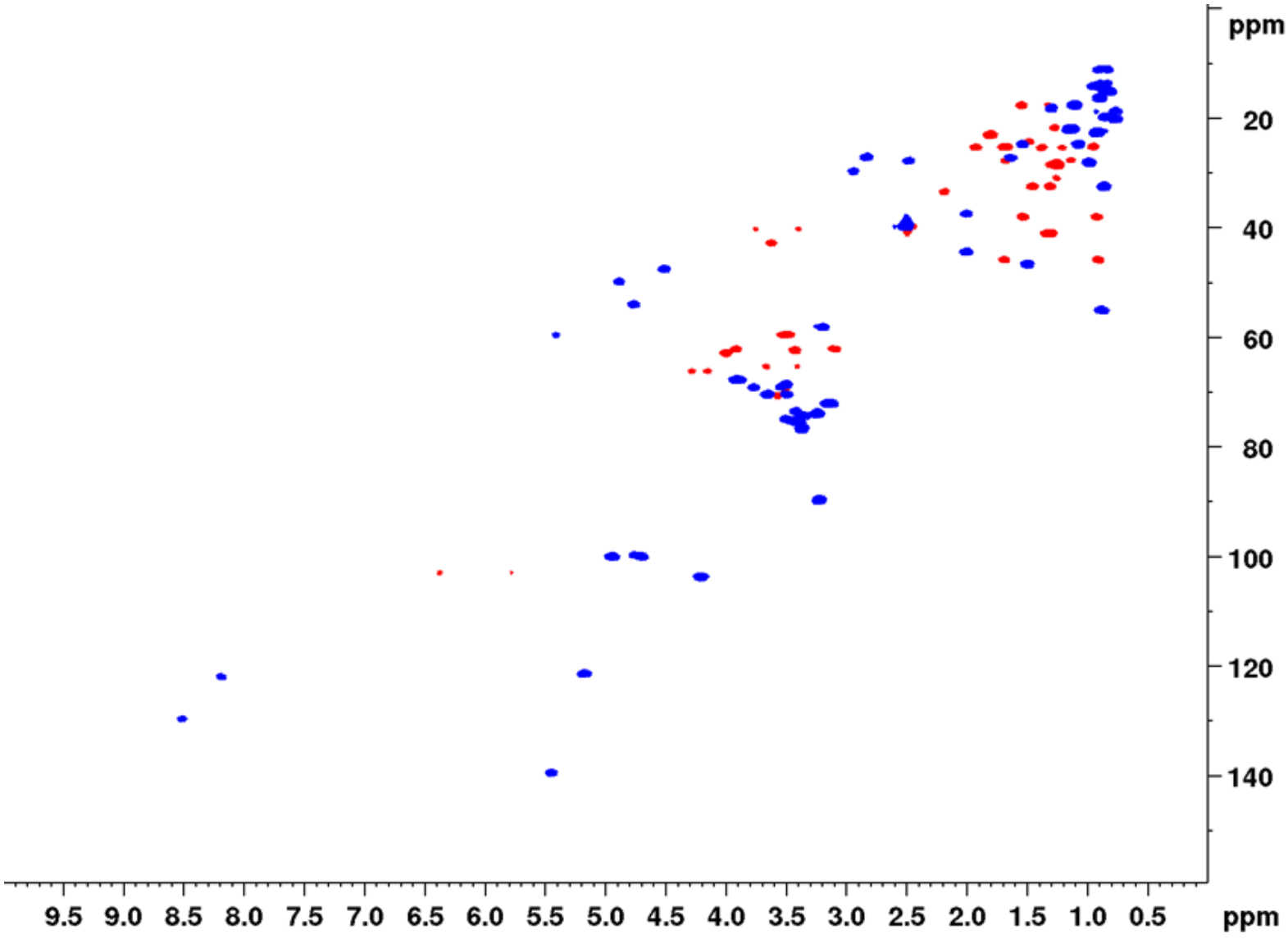


Figure 42: HSQC spectrum of thiamyxin D n DMSO-d_6 at 700/175 MHz.

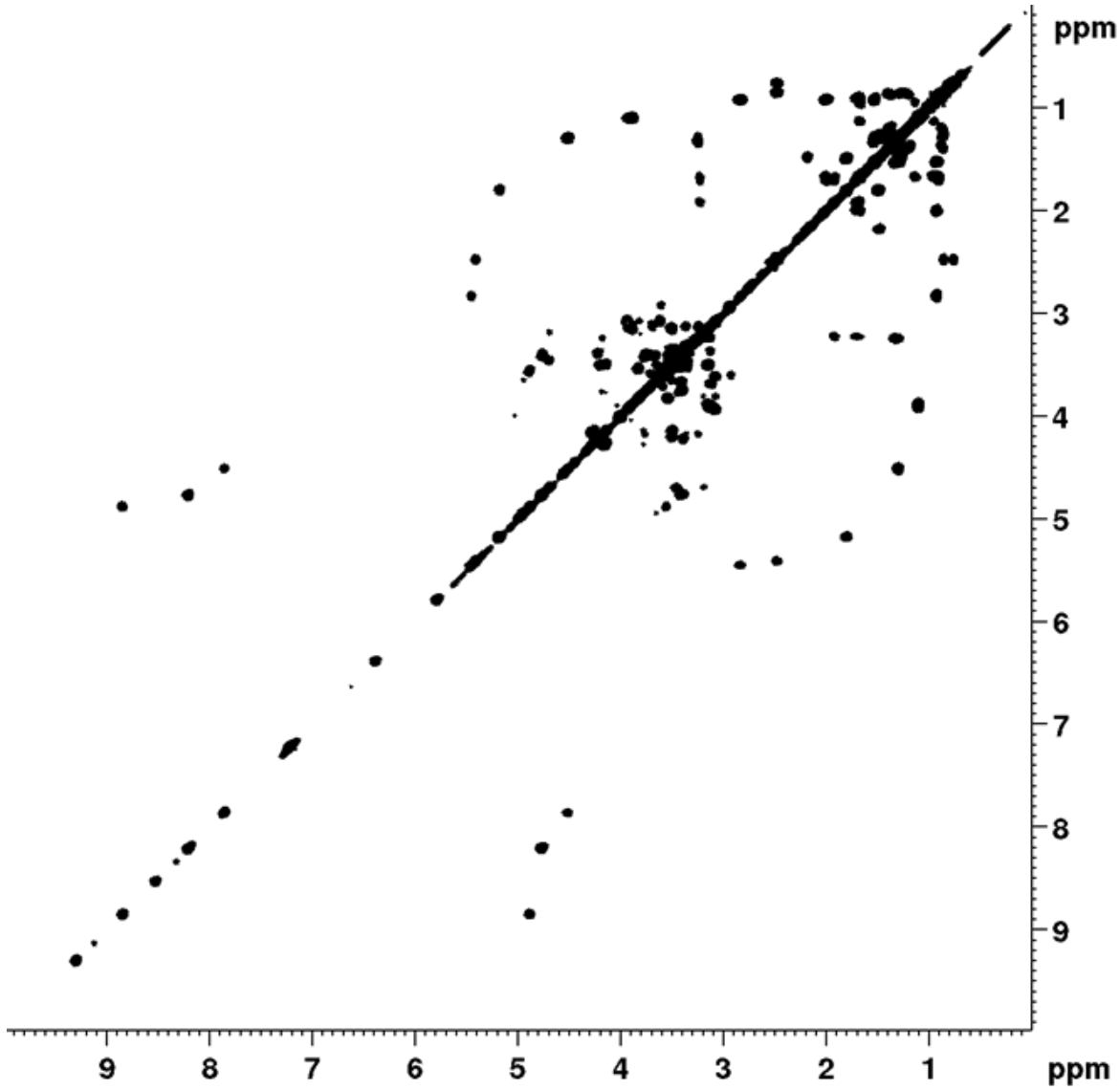


Figure 43: COSY spectrum of thiamyxin D in DMSO-d_6 at 700 MHz.

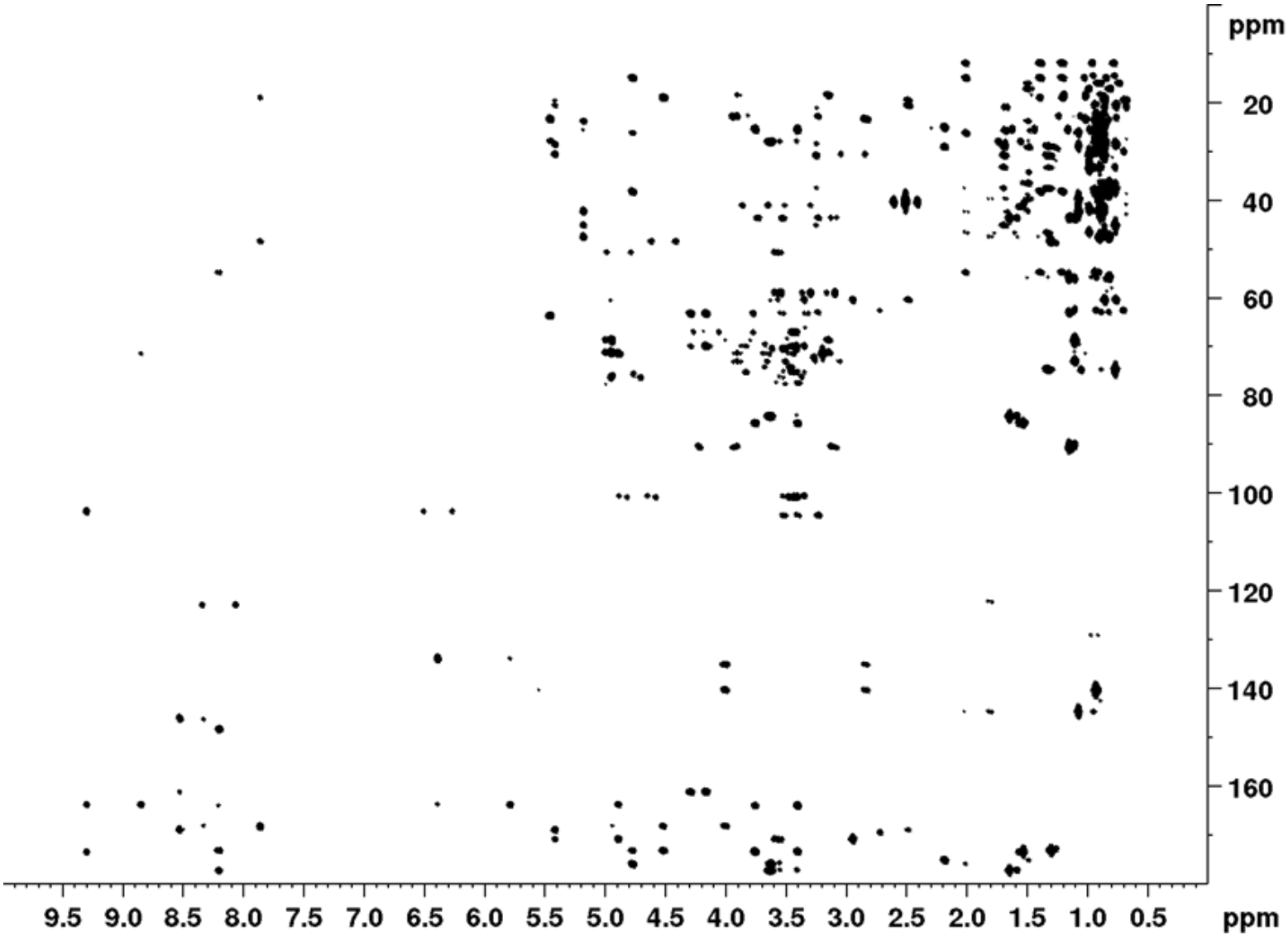


Figure 44: HMBC spectrum of thiamyxin D in DMSO-d_6 at 700/175 MHz.

S 2.3 References

[1] S. H. E. van den Worm, K. K. Eriksson, J. C. Zevenhoven, F. Weber, R. Züst, T. Kuri, R. Dijkman, G. Chang, S. G. Siddell, E. J. Snijder, V. Thiel, A. D. Davidson, *PLoS ONE* **2012**, *7*, e32857.

[2] C. D. Bader, F. Panter, R. Garcia, E. P. Tchesnokov, S. Haid, C. Walt, C. Spröer, A. F. Kiefer, M. Götte, J. Overmann, T. Pietschmann, R. Müller, *Chemistry – A European Journal* **2022**, *28*, e202104484.

[3] Okoth Dorothy A, J. J. Hug, R. Garcia, C. Spröer, J. Overmann, R. Müller, *Molecules (Basel, Switzerland)* **2020**, *25*, 2676.

[4] Z. J. Anderson, C. Hobson, R. Needley, L. Song, M. S. Perryman, P. Kerby, D. J. Fox, *Org. Biomol. Chem.* **2017**, *15*, 9372–9378.

[5] a) K. Fujii, Y. Ikai, T. Mayumi, H. Oka, M. Suzuki, K. Harada, *Anal. Chem.* **1997**, *69*, 3346–3352; b) R. Bhushan, H. Bruckner, *Amino Acids* **2004**, *27*, 231–247;

[6] P. Wipf, P. C. Fritch, *Tetrahedron Lett.* **1994**, *35*, 5397–5400.

[7] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2019**, W81-W87.

[8] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J. Mol. Biol.* **1990**, *215*, 403–410.

[9] S. Lu, J. Wang, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, M. Gwadz, Di Hurwitz, G. H. Marchler, J. S. Song, N. Thanki, R. A. Yamashita, M. Yang, D. Zhang, C. Zheng, C. J. Lanczycki, A. Marchler-Bauer, *Nucleic Acids Res.* **2020**, *48*, D265-D268.

[10] M. A. Skiba, C. L. Tran, Q. Dan, A. P. Sikkema, Z. Klaver, W. H. Gerwick, D. H. Sherman, J. L. Smith, *Structure* **2020**, *28*, 63-74.e4.

[11] D. Pogorevc, Y. Tang, M. Hoffmann, G. Zipf, H. S. Bernauer, A. Popoff, H. Steinmetz, S. C. Wenzel, *ACS Synth. Biol.* **2019**, *8*, 1121–1133.

[12] C. Rausch, I. Hoof, T. Weber, W. Wohlleben, D. H. Huson, *BMC Evol. Biol.* **2007**, *7*, 78–92.

Chapter 3

-

Sesbanimide R, a Novel Cytotoxic Polyketide Produced by Magnetotactic Bacteria

Ram Prasad Awal[a]+ Patrick A. Haack[b,c]+, Chantal D. Bader[b,c], Cornelius N. Riese[a], Dirk Schüler [a]* and Rolf Müller[b,c]*

Previously published in: mBio **2021**

DOI: 10.1128/mbio.00591-21

[a] Department of Microbiology, University of Bayreuth, Bayreuth, Germany
*E-Mail: Dirk.Schueler@uni-bayreuth.de
[b] Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research, Saarland University Campus, 66123 Saarbrücken, Germany
*E-Mail: Rolf.Mueller@helmholtz-hips.de.
[c] Department of Pharmacy, Saarland University, 66123 Saarbrücken, Germany

[+] These authors contributed equally. * Corresponding authors.

Author Contributions and Acknowledgments

Patrick A. Haack

This author designed and performed experiments and interpreted results. Extraction and statistical analysis of the crude extract as well as identification and subsequent purification of sesbanimide R was conducted by the author. Furthermore, the author performed *in-silico* analysis of the gene cluster, generated the biosynthesis hypothesis, and contributed to structure elucidation and stereochemical assignment of the compound. The author significantly contributed to the conception and writing of this manuscript.

Contributions by Others

Ram Prasad Awal contributed to the conception of this study, designed and performed micro- and molecular biological experiments, interpreted results, and significantly contributed to the conception and writing of this manuscript. Chantal D. Bader performed NMR structure elucidation, contributed to stereochemical assignment of the sesbanimide R and writing of the manuscript. Cornelius N. Riese contributed to the large-scale fermentation of the producer strain. Dirk Schüler and Rolf Müller contributed to the conception and supervision of this study, as well as editing and proofreading of the manuscript.

Acknowledgments

This study was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 692637 to D.S.), the Deutsche Forschungsgemeinschaft (INST 91/374-1 LAGG to D.S.), and the Federal Ministry of Education and Research (BMBF) (grant MagBioFab to D.S.).

We thank Alexandra Amann and Stefanie Schmidt for performing the cytotoxicity assays and Daniel Krug and Fabian Panter for helpful discussion and critical review of the manuscript.

D.S., R.M., and R.P.A. conceived and designed research, R.P.A., P.A.H., C.D.B., and C.N.R. performed research, R.P.A., P.A.H., C.D.B., R.M., and D.S. analyzed data, and R.P.A., P.A.H., C.D.B., D.S., and R.M. wrote the paper. All authors read and approved the final manuscript.

We declare no competing interest.

3.1 Abstract

Genomic information from various magnetotactic bacteria suggested that besides their common ability to form magnetosomes, they potentially also represent a source of bioactive natural products. By using targeted deletion and transcriptional activation, we connected a large biosynthetic gene cluster (BGC) of the *trans*-acyltransferase polyketide synthase (*trans*-AT PKS) type to the biosynthesis of a novel polyketide in the alphaproteobacterium *Magnetospirillum gryphiswaldense*. Structure elucidation by mass spectrometry and nuclear magnetic resonance spectroscopy (NMR) revealed that this secondary metabolite resembles sesbanimides, which were very recently reported from other taxa. However, sesbanimide R exhibits an additional arginine moiety the presence of which reconciles inconsistencies in the previously proposed sesbanimide biosynthesis pathway observed when comparing the chemical structure and the potential biochemistry encoded in the BGC. In contrast to the case with sesbanimides D, E, and F, we were able to assign the stereocenter of the arginine moiety experimentally and two of the remaining three stereocenters by predictive biosynthetic tools. Sesbanimide R displayed strong cytotoxic activity against several carcinoma cell lines.

Importance

The findings of this study contribute a new secondary metabolite member to the glutarimide-containing polyketides. The determined structure of sesbanimide R correlates with its cytotoxic bioactivity, characteristic for members of this family. Sesbanimide R represents the first natural product isolated from magnetotactic bacteria and identifies this highly diverse group as a so-far-untapped source for the future discovery of novel secondary metabolites.

Keywords

glutarimide-containing polyketides, cytotoxic activity, trans-AT polyketide synthase, magnetotactic bacteria

3.2 Introduction

Magnetotactic bacteria (MTB) share the ability to biomineralize membrane enclosed organelles consisting of either magnetite (Fe_3O_4) or greigite (Fe_3S_4) crystals, called magnetosomes, which enable the cells to navigate within Earth's magnetic field [1]. Studies on MTB so far have focused mainly on understanding magnetosome structure, biosynthesis, and biological function as well as exploring the potential utility of magnetosomes as magnetic nanoparticles for various applications, such as magnetic imaging or magnetic hyperthermia, magnetosome-based immunoassays, and as nano carriers in magnetic drug targeting and multifunctional nanomaterials with versatile functional moieties [2–7].

Apart from their common ability to form magnetosomes, MTB represent a highly heterogeneous group of prokaryotes. They are abundant and widespread in the sediments of many diverse aquatic ecosystems, ranging from freshwater to hypersaline habitats [8, 9], and besides a multitude of free-living, single-celled MTB, multicellular and even ectosymbiotic members of this group have been discovered [10-12]. MTB are known to have diverse and versatile lifestyles, and members of this group are found in many different classes of eubacteria [13–15]. Within the last years, a wealth of genomic information has been obtained by conventional genomics, metagenomics, and single-cell genomics [14–20]. We have recently shown that chances for the discovery of novel secondary metabolites clearly correlate with the increasing phylogenetic distance of the microorganisms under study [21]. Because of their huge ecological, metabolic, phylogenetic, and genomic diversity, producers of such interesting natural products might also be expected among MTB. Indeed, Araujo *et al*. [22] first noted the presence of typical secondary metabolite biosynthetic gene clusters (BGCs), such as putative polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs), in the genomes of several MTB. However, this so far has remained an untapped source for discoveries, largely owing to the fact that most of these bacteria are not tractable; many cannot be cultured in the laboratory.

One of the few MTB that can be cultivated reasonably well and is genetically tractable is the alphaproteobacterium *Magnetospirillum gryphiswaldense* [23–26], which previously served as a model in many studies on magnetotaxis, organelle biosynthesis, and magnetite biomineralization [2, 27]. Interestingly, several putative BGCs for secondary metabolites were tentatively predicted in its genome [22, 26]. This prompted us to investigate in more detail the strains' biosynthetic capability using a combination of molecular and analytical methods.

In this study, we focused on the role of a *trans*-acyltransferase PKS (*trans*-AT PKS) BGC in *M. gryphiswaldense*, which we identified as a homologue of the sesbanimide gene cluster described by Kačar *et al*. in parallel to our studies [28, 60]. We set out to unambiguously assign the corresponding

secondary metabolite from *M. gryphiswaldense* by markerless deletion of the gene cluster, to isolate the polyketide product and to elucidate its structure. Furthermore, we devised a model for sesbanimide biosynthesis that complements the one suggested by Kačar *et al*. [28, 60] and revealed the new sesbanimide R as a missing link between the sesbanimide biosynthesis pathways when compared across several taxa [29]. In addition, we demonstrate cytotoxic activity of the novel sesbanimide congener.

3.3 Results and Discussion

Identification, deletion, and transcriptional activation of a *trans*-AT PKS gene cluster.

Using the antiSMASH tool[30], we identified several secondary metabolite gene clusters in the *M. gryphiswaldense* genome. Three gene clusters were predicted to encode the biosynthesis of a putative lasso peptide, an aryl polyene, and a homoserine lactone (see Table S1 in the supplemental material). In addition, a large (69,942 bp) gene cluster was predicted to encode a putative *trans*-AT PKS. It has a conspicuously high G+C content (66.7% versus 63.2% of the entire genome) and comprises 30 open reading frames (ORFs), which were tentatively assigned to various constituents of a *trans*-AT PKS.

To study the function of the cluster, we deleted the three putative core-biosynthetic genes (MSR-1_15630 to MSR-1_15650) encoding two large PKSs and a monooxygenase and a gene (MSR-1_15620) encoding an acyltransferase. The deletion comprised 41,295 bp and yielded a Δ*trans*-at-pks strain (Fig. S1). Growth of the Δ*trans*-at-pks strain was essentially wild type-like, with slightly increased doubling times during growth under aerobic conditions (Fig. 1A). Mutant cells were indistinguishable from the wild type with respect to length and shape (Fig. 1B to E). Cultures of the Δ*trans*-at-pks strain exhibited a lower magnetic response (C_{mag} = 1.17; wild type, C_{mag} = 1.3; C_{mag} is a light-scattering parameter for the semiquantitative estimation of average magnetic alignment of cells[31]). Transmission electron microscopy (TEM) of wild-type (Fig. 1D) and Δ*trans*-at-pks (Fig. 1E) cells showed that the strains formed magnetosomes in about same numbers and with similar average sizes (Fig. 1F and G); however, both smaller. (<25 nm) and larger (>60 nm) particles were more frequent in the Δ*trans*-at-pks strain than in the wild type (Fig. 1F), which might explain the slightly lower magnetic response. To identify the biosynthetic product(s) of the *trans*-AT PKS cluster, the wild-type and Δ*trans*-at-pks strains were cultivated under aerobic, microaerobic, and anaerobic conditions in flask standard medium (FSM), and extracts of these strains were compared using principal-component analysis (Fig. S2) as previously described [32]. Under microoxic and anoxic conditions, which are known to favor magnetosome biosynthesis [33, 34], there were no significant differences detectable between the mutant and the wild type. However, in the extract of the wild-type strain grown under aerobic conditions that are known to inhibit magnetosome formation [33, 34], we identified a compound with a mass

of 691.38 Da which was absent from the Δ*trans*-at-pks mutant strain. Yields of the target compound obtained from wild-type cultures proved insufficient for the isolation and subsequent elucidation of its structure by nuclear magnetic resonance spectroscopy (NMR). Since we hypothesized that the low production might be due to poor expression of biosynthetic genes, we attempted to enhance their expression by transcriptional activation. To this end, a DNA fragment of 145 bp harboring a putative native promoter in front of MSR-1_15600 (ORF7) was replaced with a 64-bp

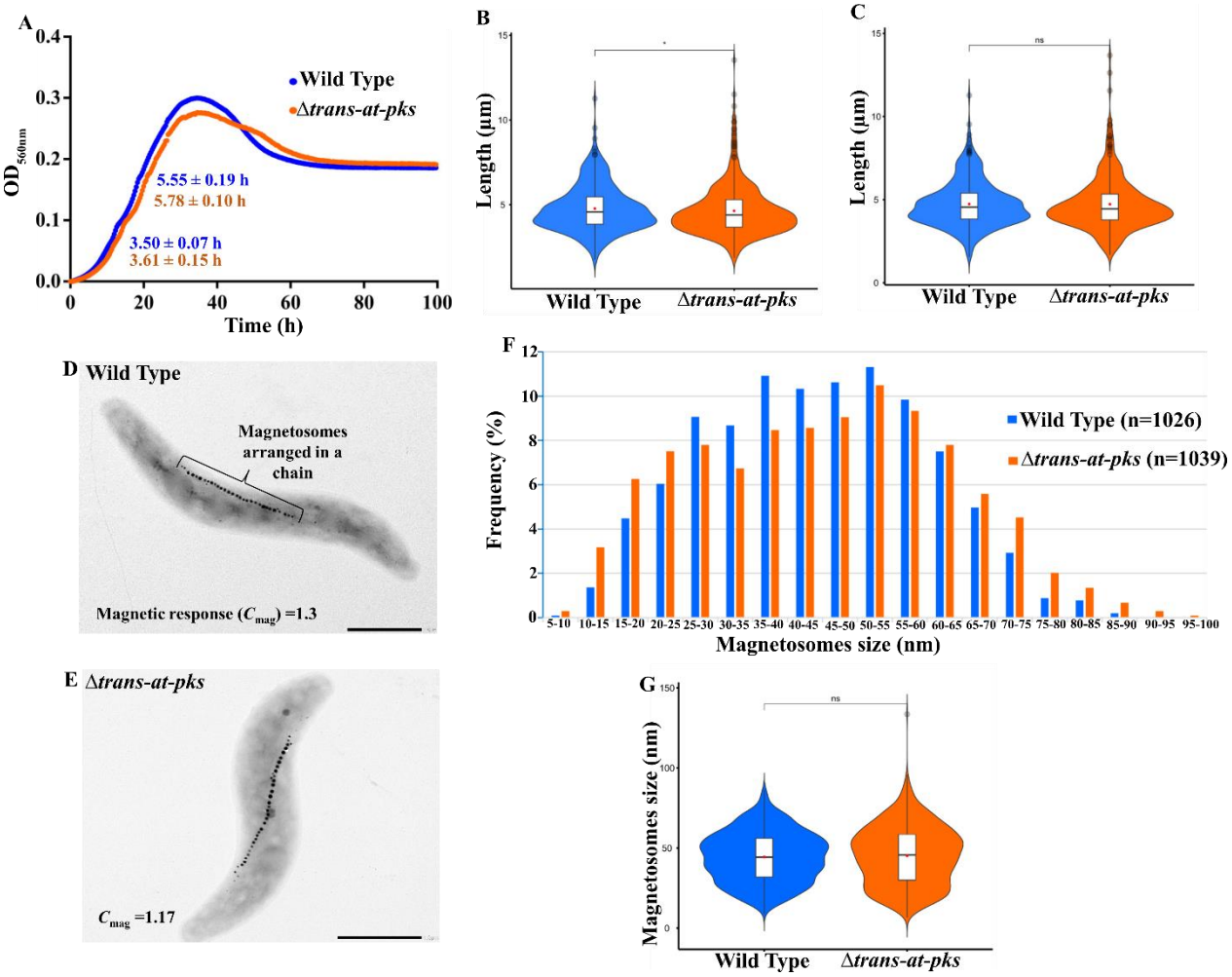


Figure 1: (A) Growth of the wild type and Δ*trans-at-pks* strains under aerobic conditions where the target compound was produced. Each growth curve represents the average of two individual growth curves. The doubling time (T_d) (mean±SD) for each strain is given in the graph for the first and second part of the diauxic growth curve. (B) Cell length of wild type (mean=4.77±1.37 µm, n=312) and Δ*trans-at-pks* (mean=4.64±1.46 µm, n=504) grown under aerobic conditions. (C) Cell length of wild type (mean=4.73±1.37 µm, n=347) and Δ*trans-at-pks* (mean=4.72±1.6 µm, n=354) grown under microaerobic conditions. TEM images of wild type (D) and Δ*trans-at-pks* (E). Analysis of magnetosomes size distribution in wild type (mean=44.45±15.59 nm, n=1026) and Δ*trans-at-pks* (mean=45.18 ±18.29 nm, n=1039) (F, G).

fragment containing the stronger constitutive promoter $P_{mamDC45}$ [35] and the optimized ribosomal binding site (oRBS), yielding the $P_{mamDC45}$-trans-at-pks strain (Fig. S3). $P_{mamDC45}$ is an optimized version of the native

promoter P_{mamDC}, which drives transcription of the *mamGFDC* operon involved in magnetosome biosynthesis of *M. gryphiswaldense* [36], and was shown to enhance the expression of a foreign gene 8-fold compared to that obtained with P_{mamDC} [35]. Indeed, mass spectra obtained by liquid chromatography-mass spectrometry (LC-MS) from extracts of the $P_{mamDC45}$-*trans*-at-pks strain showed a 7-fold-increased intensity of the target mass, suggesting a successful transcriptional activation of the gene cluster (Fig. 2A). As the yield of the compound obtained from shake flasks cultures of the $P_{mamDC45}$-*trans*-at-pks strain was still too low for the isolation of the corresponding natural product, we scaled its production up to a 10-liter fermentor, which provided enhanced aeration and growth of the culture. This approach enabled the isolation of 2 mg of the compound by semipreparative high-performance liquid chromatography (HPLC) and the elucidation of its structure using MS and NMR spectrometry.

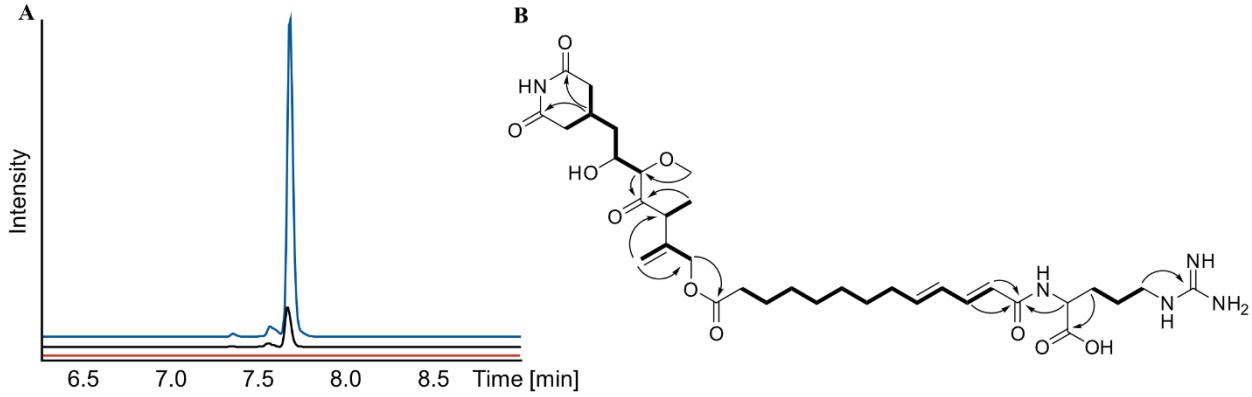


Figure 2: (A) Extracted ion chromatograms for *m/z* 692.38 $[M+H]^+$ showing the difference in compound production. In the strain Δ*trans-at-pks* (red), the production of the compound was abolished. In the promoter-activated strain $P_{mamDC45}$-*trans-at-pks* (blue) the production was increased ca. 7-fold (AUC: 6513288) in comparison to the wild type (black) (AUC: 880064). (B) NMR elucidated structure of sesbanimide R with the most relevant COSY and HMBC correlations.

De novo structure elucidation.

High-resolution electrospray ionization mass spectrometry (HRESI-MS) analysis of the compound (Fig. 2B) shows an $[M+H]^+$ signal at *m/z* 692.3869 (calc. 692.3865 Δ = 20.6 ppm) consistent with the neutral sum formula of $C_{34}H_{53}N_5O_{10}$ and containing 11 double bond equivalents (DBEs). The ^{1}H NMR and heteronuclear single quantum coherence (HSQC) spectra of the compound (Table 1) revealed a signal characteristic for an aliphatic exo double bond at $\delta(^1H)$ of 5.18 and 4.96 ppm (H-13), which shows heteronuclear multiple bond correlation (HMBC) correlations to a methylene group at $\delta(^1H)$ of 4.66 displayed as a singlet (H-10), as well as a methine group at $\delta(^1H)$ of 3.71 ppm (H-8). The quartet of the methine group shows correlation spectroscopy (COSY) correlations to one methyl group at $\delta(^1H)$ of 1.19 ppm (H-12). This methyl group exhibits HMBC correlations to the quaternary carbon participating in the exo double bond at $\delta(^{13}C)$ of 144.5 ppm (C-9) besides a ketone at 213.2 ppm (C-7), indicating location of the methyl group and the

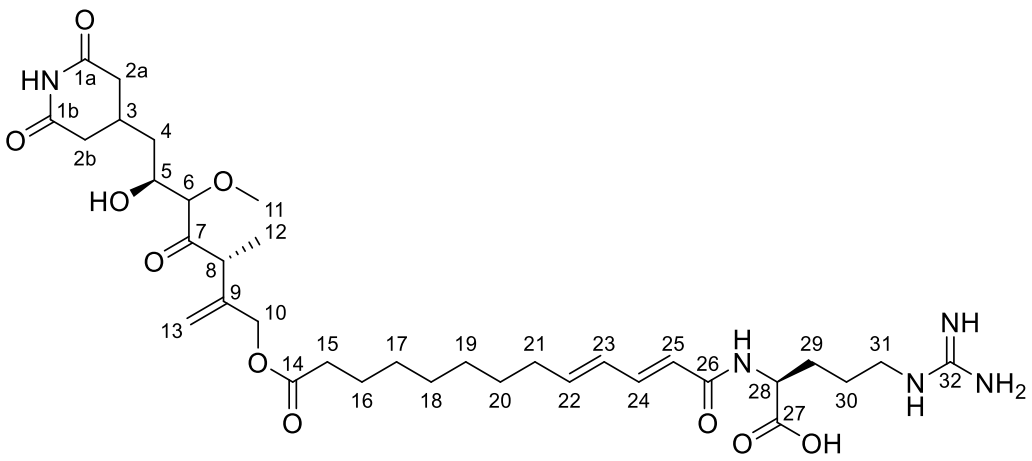
methine group between the ketone (C-7) and the exo double bond (C-9 and C-13). On the other side of the ketone, another methine group with a chemical shift of $\delta(^1H)$ of 3.66 ppm (H-6) is found, which is indicated by the HMBC correlations of those two groups. The moderately deshielded shift of this methine group in line with HMBC correlations to a methyl group with a moderately deshielded shift of $\delta(^1H)$ of 3.40 ppm (H-11) reveals this part as a methoxy function. The methine group at $\delta(^1H)$ of 3.66 ppm (H-6) furthermore shows COSY correlations to a second methine group at $\delta(^1H)$ of 3.98 ppm (H-5). Its deshielded chemical shift suggests hydroxylation of this methine. It shows COSY correlations to a diastereotopic methylene group at $\delta(^1H)$ of 1.49 and 1.73 ppm (H-4), which is located next to a methine group at 2.34 ppm (H-3) based on their COSY correlations. This methine group exhibits further COSY correlations to two diastereotopic methylene groups at d (1H) of 2.36 and 2.68 (H-2a) and of 2.33 and 2.70 ppm (H-2b) with almost identical chemical shifts, wherefore they have to be located in almost identical chemical surroundings. They do not reveal any further COSY correlations but do reveal HMBC correlations to two quaternary carbons and atd (13 C) of 174.6 ppm (C-1a/b), which also show correlations to the methine group. Based on the sum formula of the molecule and the two-dimensional (2D) NMR data, the methine, the two methylenes, and the two quaternary carbons therefore likely are arranged as glutarimide, with substitution in position 4. There are no further correlations of any glutarimide participating functional groups; as a result, this part depicts one end of the molecule.

Besides correlations of the H-10 methylene to the partial structure described above, it shows HMBC correlations to a quaternary carbon at $\delta(^{13}C)$ of 174.6 ppm (C-14). The deshielded chemical shift of this quaternary carbon suggests an ester bond in this position, which was confirmed by a saponification reaction (Fig. 3). The following seven methylene groups are arranged in a straight aliphatic chain, based on their chemical shifts and COSY as well as HMBC correlations. The deshielded chemical shift of the last of these seven methylene groups at $\delta(^1H)$ of 2.18 ppm (H-21) and its signals displayed as quartet suggest that it is followed by the first of four dienone double bond protons at $\delta(^1H)$ of 6.10, 6.22, 7.12, and 6.02 ppm (H-22-25). The two double bonds are conjugated based on COSY correlations of the four dienone double bond protons and their deshielded carbon chemical shifts at $\delta(^{13}C)$ of 144.1, 129.7, 142.3, and 122.9 ppm (C-22-25). The last two double bond protons at $\delta(^1H)$ of 7.12 and 6.02 ppm (H-24 and H-25) show HMBC correlations to a quaternary carbon at $\delta(^{13}C)$ of 168.3 ppm (H-26). Its characteristic chemical shift and additional correlations of an alpha proton (H-28) to this quaternary carbon (C-26) reveal it as acid function of an amide bond. The alpha proton at $\delta(^1H)$ of 4.40 ppm (H-28) belongs to arginine, which was confirmed by Marfey's analysis (59) in addition to the following NMR correlations. It shows HMBC correlations to a free carboxylic acid function at δ(13C) of 177.6 ppm (C-27) and two methylene groups

and at $\delta(^1H)$ of 1.92 and 1.63 ppm (H-29 and H-30), which themselves show HMBC correlations to a third more moderately deshielded chemically shifted methylene group at $\delta(^1H)$ of 3.22 ppm (H-31). Its characteristic chemical shift and correlations to a quaternary carbon at $\delta(^{13}C)$ of 158.4 ppm (C-32) reveal the coupling to the guanidine moiety of the molecule here, which marks the other end of compound. In addition to the NMR data, we observed a fragment of *m/z* 397.245 $[M+H]^+$ with LC-MS after saponification of the ester (Fig. 3). The size and sum formula of this fragment correspond to the arginine plus aliphatic chain containing part of the molecule and confirms the elucidated structure. This structure is highly similar

Table 1: NMR spectroscopic data for Sesbanimide R in methanol-d_4 at 500/125 MHz

#	$\Delta^{13}C$ [PPM]	$\Delta^{1}H$ [PPM], MULT (*J* [HZ])	COSY	HMBC
1A	174.6	-	-	-
1B	174.6	-	-	-
2A	39.1	2.36, 2.68, m	2b, 3	1a, 3, 4
2B	37.7	2.33, 2.70, m	2a, 3	1b, 3, 4
3	28.4	2.34, m	2a, 2b, 4	1a, 1b, 2a, 2b, 4, 5
4	38.9	1.49, 1.73, m	3, 5	2a, 2b, 3, 5, 6
5	71.7	3.98, dt (10.4, 3.1)	4, 6	3, 4, 6, 7
6	90.6	3.66, d (3.1)	5	4, 5, 7, 11
7	213.2	-	-	-
8	46.9	3.71, q (7.0)	12	7, 9, 10, 12, 13
9	144.5	-	-	-
10	66.8	4.66, s	13	8, 9, 13, 14
11	60.6	3.40, s	-	6
12	16.6	1.19, d (7.0)	8	7, 8, 9, 13
13	113.9	4.96, 5.18, s	10	8, 9, 10
14	174.6	-	-	-
15	34.8	2.38, t (7.5)	16	14, 16, 17
16	26.0	1.63, m	15, 17	14, 15, 17, 18
17	30.4	1.34, m	16, 18	15, 16, 18, 19
18	30.1	1.34, m	17, 19	16, 17, 19, 20
19	30.6	1.30, m	18, 20	17, 18, 20, 21
20	29.7	1.45, m	19, 21	18, 19, 21, 22
21	33.8	2.18, q (7.1)	20, 22	19, 20, 22, 23
22	144.1	6.10, dt (7.2, 15.1)	21, 23	20, 21, 23, 24, 25
23	129.7	6.22, dd (10.8, 15.1)	22, 24	21, 22, 24, 25
24	142.3	7.12, dd (10.7, 15.1)	23, 25	22, 23, 25, 26
25	122.9	6.02, d (15.1)	24	22, 23, 24, 26
26	168.3	-	-	-
27	177.6	-	-	-
28	54.9	4.40, dd (5.3, 7.7)	nd	26, 27, 29, 30
29	30.9	1.74, 1.92, m	30	27, 28, 30, 31
30	26.0	1.63, m	29, 31	28, 29, 31
31	41.9	3.22, m	30	29, 30, 32
32	158.4	-	-	-



to the structure of sesbanimide F, which became available at a late stage of our work in a study by Kačar *et al*. (28, 60). However, our compound contains an additional terminal arginine (R) moiety. Hence, we used the name sesbanimide R.

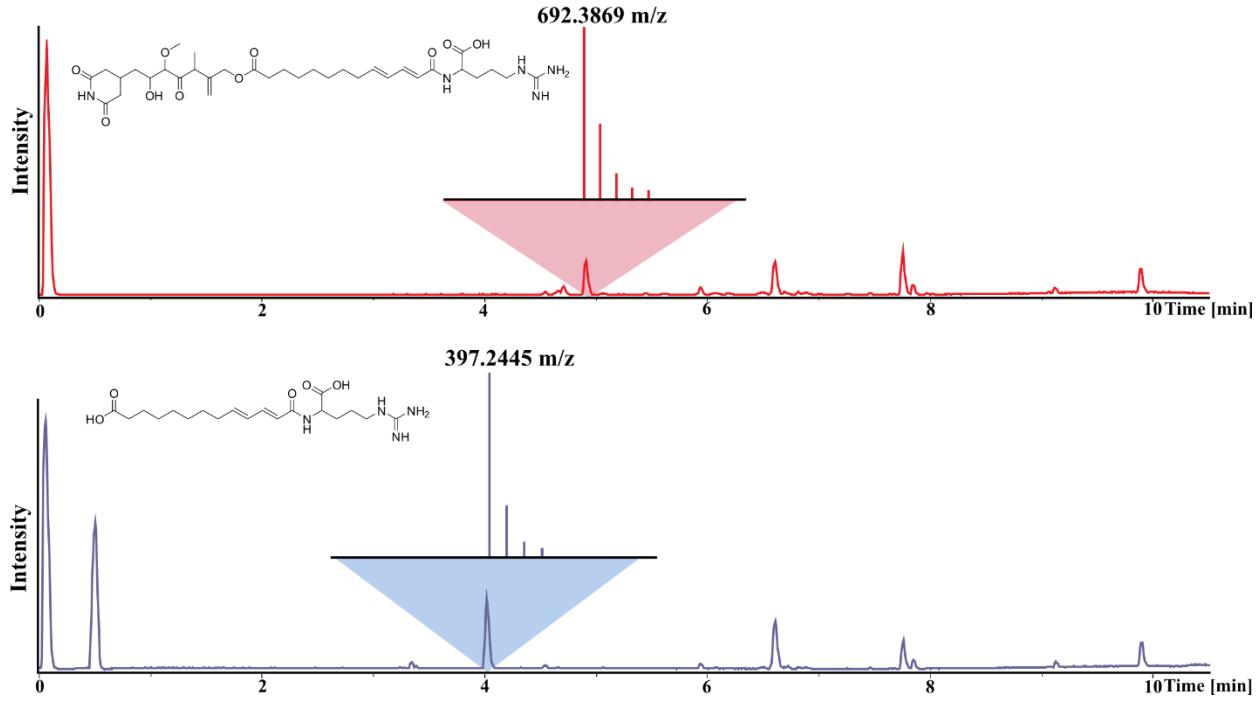


Figure 3: Saponification of sesbanimide R to confirm the NMR elucidated structure. Base peak chromatogram of a sesbanimide R sample before (top) and after (bottom) treatment with NaOH. A fragment with an *m/z* of 397.2445 $[M+H]^+$ was observed, which corresponds to the arginine containing part of the molecule after ester hydrolysis.

Determination of the sesbanimide R stereochemistry.

The vicinal coupling constant of 15.1 Hz for both aliphatic double bonds suggests an *E*-configuration of both double bonds. Marfey's analysis and comparison to commercially available L- and D-arginine standards revealed the arginine from sesbanimide R to be *S* configured, as the hydrolysis product of sesbanimide R has the same retention times as L-arginine when derivatized with fluorodinitro-phenyl-5-L-leucine amide (L-FDLA) and D-FDLA, respectively (Table 2). Due to instabilities of the molecule under acidic and basic conditions and selectivity issues between the free hydroxyl groups and the glutarimide, Mosher esterification experiments, which were carried out to elucidate the configuration of the remaining stereocenters, were not successful. When adding 10 or fewer equivalents of pyridine to the reaction mixture in chloroform, we observed complete degradation of the molecule. When performing the experiment with pure pyridine, the hydroxyl group underwent fast elimination after formation of the

Table 2: Retention times for L- and D-arginine standards as well as sesbanimide R after derivatization with L- and D-FDLA to determine the stereochemistry of the arginine moiety of sesbanimide R.

	Retention Time		
Sample	**D-FDLA [min]**	**L-FDLA [min]**	**assignment**
D-arginine	10.85	10.35	
L-arginine	10.35	10.85	
Hydrolysis of sesbanimide R	10.88	10.36	*S*-configured

respective Mosher ester. We were therefore not able to determine the absolute stereochemical configuration of the molecule experimentally and speculate on the stereochemistry based on *in silico* analysis of the BGC. The *trans*ATor tool predicts the structure of *trans*-AT polyketides according to the substrate specificities of the involved keto-synthase (KS) domains [37]. The top five hits of the tool predict the KS domain of module 4 to accept D-OH, while the sequence-based stereochemistry prediction for the ketoreductase (KR) domain of module 3 was inconclusive. We therefore predict the stereocenter at C-5 to be *S* configured. Xie *et al*. (38) recently suggested that all C-methyltransferases in *trans*-AT PKS assembly lines generate (2*R*)-2-methyl-3-ketoacyl-acyl carrier protein (3-ketoacyl-ACP) intermediates and that (2*S*)-2-methyl-3-hydroxyacyl-ACP intermediates are produced by epimerizing A2-type KR domains [38]. As there is no KR domain present in module 4, we propose that the stereocenter at C-8 is *R* configured. The stereocenter at C-6 is likely generated by a cytochrome P450 (cyP450) enzyme (SbnE), but we were not able to make a prediction for its stereochemistry. These predictions are speculative, and further experiments are required to fully elucidate the stereochemistry of sesbanimide R.

***In silico* analysis of the gene cluster and biosynthesis hypothesis.**

A detailed annotation of the BGC was carried out (Table S2). Besides using antiSMASH [30] for cluster and domain identification, additional information was gained by submitting the translated protein sequences to the *trans*ATor tool [37] (Table S3). Finally, the conserved domain search tool CDD [58] was used to identify domains that were not identified by antiSMASH. The core biosynthetic gene cluster (BGC) spans over 39 kbp and consists of the two large PKS genes, *sbnO* (MSR-1_15630) and *sbnQ* (MSR-1_15650), as well as one monooxygenase-encoding gene, *sbnP* (MSR-1_15640). The core cluster is flanked by two acyltransferase (AT) domains encoded by *sbnA* and *sbnN*. SbnN was identified as an *in-trans* acyltransferase and SbnA as an *in-trans* acylhydrolase. Several additional biosynthetic genes are encoded up- and downstream of the core BGC: an asparagine synthase accompanied by an ACP domain (*sbnJ* and *sbnK*), a beta-branching cassette [39] (*sbnF-I*), a cytochrome P450 enzyme (*sbnE*), a methyltransferase (*sbnD*), and a stand-alone acyl-coenzyme A (acyl-CoA) dehydrogenase (*sbnX*). ORFs 6, 8, 9, 11, 14, and 15 encode transport-associated proteins putatively responsible for exporting sesbanimide R out of the cell. ORFs 2, 3, 5, and 7 encode regulatory proteins putatively responsible for controlling BGC expression and thereby sesbanimide production. ORFs 4, 10, 12, and 13 were annotated as encoding hypothetical proteins with unknown function.

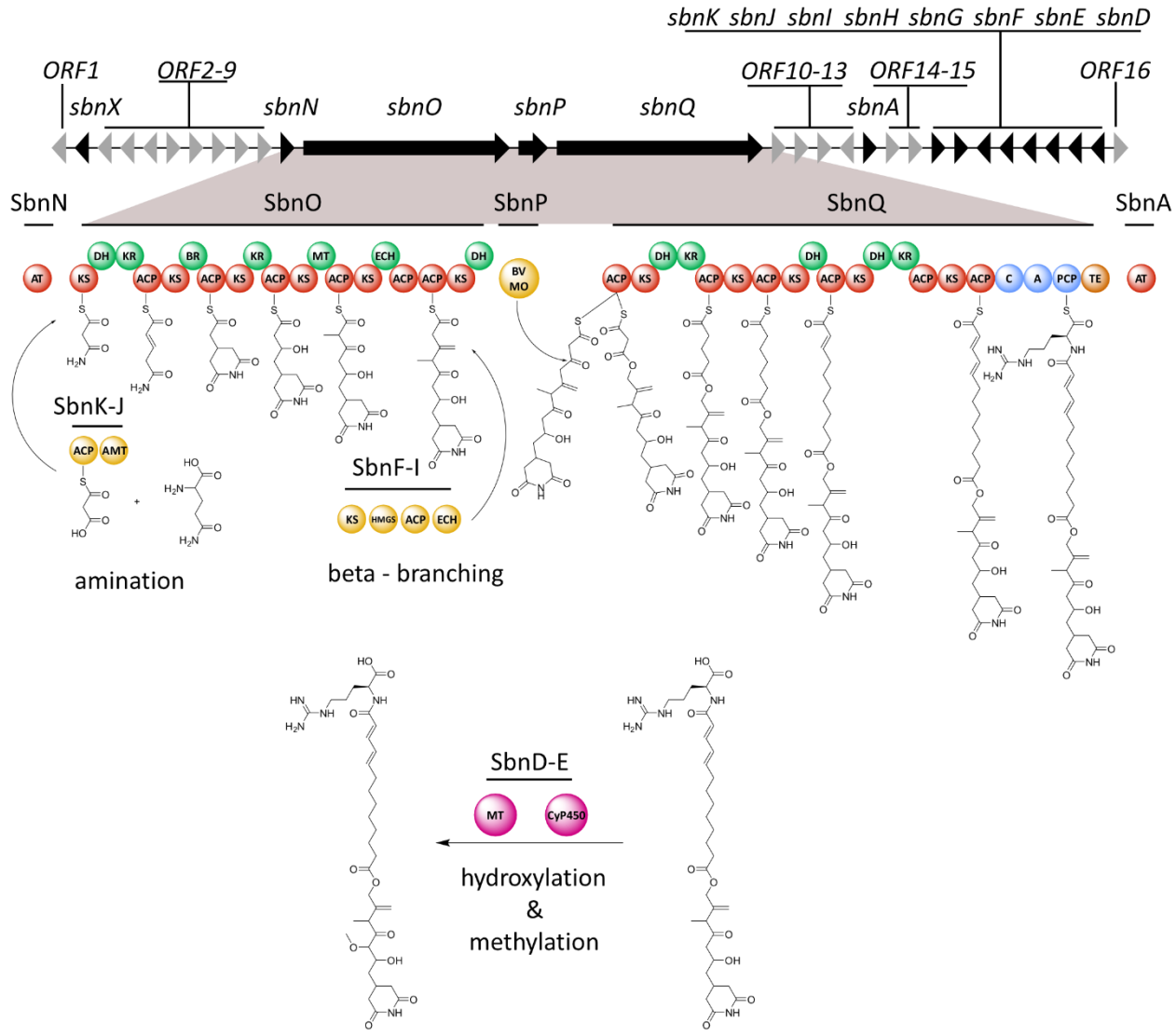


Figure 4: Proposed biosynthetic pathway for sesbanimide R. Core PKS modules are marked in red, core NRPS modules in blue and the thioesterase in orange. DH and KR modules of the core assembly line are marked in green. The amidotransferase, beta-branching cassette and Bayer-Villiger Monooxygenase are marked in yellow and the tailoring methyltransferase and Cyp450 enzyme in pink. The genes involved in the sesbanimide R biosynthesis are marked in black and named *sbnA-X.* The remaining genes with unknown or unassigned functioned are marked in grey and named *ORF1-16*.

All KS domains of SbnO and SbnQ possess the active-site cysteine and histidine, except for the nonelongating KS5 of SbnQ, which is missing the first histidine [40]. All ACP domains of SbnO and SbnQ possess the canonical active-site serine. The catalytic triad of serine, tyrosine, and asparagine is present in all KR domains of SbnO and SbnQ. The dehydratase (DH1) of SbnO and DH1 and DH2 of SbnQ contain the conserved HXXXGXXXXP motive, which is missing in DH2 of SbnO and DH3 of SbnQ (Fig. S4A to D). We

therefore propose the following biosynthesis scheme, based on *in silico* analysis of the BGC and considering the elucidated chemical structure of sesbanimide R (Fig. 4).

Initially, an amino group is transferred to ACP bound malonate by SbnJ [41]. The starter moiety is then transferred to the first module of SbnO. Modules one and two of the assembly line then form the glutarimide moiety as previously described for the biosynthesis of gladiofungin [42]. Modules three to five elongate the nascent molecule according to the substrate specificity prediction for their KS domains. Exomethylene moiety incorporation by module five has previously been described for several trans-AT PKS biosyntheses [43]. The domains required for exomethylene formation (ECH domain, tandem ACP domains, and a beta-branching cassette [*sbnF-I*]) are all present in the cluster. Module six is found split onto the genes *sbnO* and *sbnQ*, which are separated by *sbnP*, encoding a flavin-binding monooxygenase. Such a split module, containing a monooxygenase and a DH domain missing the conserved HXXXGXXXXP motif, has been shown to incorporate oxygen into polyketide backbones. The monooxygenase accepts thioesters bearing β-keto groups and acts as a Baeyer-Villiger mono-oxygenase (BVMO) to generate malonyl esters [29]. The module is thus likely responsible for the ester formation in sesbanimide R. The second part of the molecule is synthesized by the PKS megasynthase SbnQ. Judging by the structure formula of sesbanimide R, we propose that module seven or eight performs one iterative PKS elongation step and thus incorporates a second malonyl-CoA building block into the final molecule. The DH and KR domains are proposed to act in *trans*, to biosynthesize the saturated part and double bonds present in sesbanimide R. An enoylreductase (ER) domain would also be needed to fully reduce the incorporated C-2 unit, but this domain is not encoded on *sbnQ*. We therefore propose that this function is carried out by SbnX, which was identified as an acyl-CoA dehydrogenase. The terminal NRPS module on *sbnQ* was predicted to incorporate L-arginine by NRPS predictor 2, which fits well with the elucidated structure [44]. We propose that the methoxy group at C6 is incorporated by a cytochrome P450 enzyme and an Fkbm family methyltransferase, encoded by *sbnE* and *sbnD*, respectively.

Taken together, our devised biosynthesis scheme for sesbanimide R is very similar to the pathway suggested in parallel by Kačar *et al*. for the biosynthesis of sesbanimide F from *Stappia indica* PHM037 *[28, 60]*. The main differences between the two BGC lie in the distribution of DH and KR domains in SbnQ, the presence of three additional transport-associated genes in the *M. gryphiswaldense* cluster, and a phosphopantetheinyl transferase in the *Stappia indica* cluster which is absent in the *M. gryphiswaldense* BGC. Notably, the final products from the strains under investigation by Kačar *et al.* do not contain the terminal arginine moiety observed in sesbanimide R, even though the corresponding biosynthetic gene cluster contains the L-arginine-incorporating NRPS module [28, 60]. We speculate that the BGC from *M. gryphiswaldense* responsible for sesbanimide R formation is an evolutionary intermediate in a

developmental line leading to the sesbanimide gene cluster from PHM037 and PHM038 or that these clusters may carry a nonfunctional NRPS module. A conserved domain search of the adenylation (A) domain of the NRPS modules from *M. gryphiswaldense* and S. indica PHM037 revealed that the active sites are likely intact in both cases. In the case of the *S. indica* domain, however, the residues just before the active site seem to be unusual for A-domains. They were identified because they do not match the alignment against the reference A-domains from the CDD database [58] (Fig. S5A and B). Kačar *et al*. speculated that the arginine moiety is cleaved rapidly after the biosynthesis, so that the corresponding analogues are not detectable with the applied analytical conditions [28, 60]. As we were able to detect sesbanimide R, which was also relatively stable, we suggest as an alternative explanation that the uncommon residues close to the active-site residues might result in an inactive A-domain in the *S. indica* cluster and that, therefore, no arginine is incorporated. Additionally, we did not detect any of the sesbanimides (A, B, C, D, E, and F) which were observed by Kačar *et al*. [28, 60] in *M. gryphiswaldense*. A possible explanation might be that the tailoring steps resulting in the formation of sesbanimides A, D, C, and E occur only if no arginine moiety is present. However, until further insight is gained into the biosynthesis of these compounds, the reason(s) for the discrepancy in product composition remains elusive.

Cytotoxicity.

Sesbanimides have been associated with strong antitumor/cytotoxic activities, which is common for polyketides containing a glutarimide moiety [45, 46]. We therefore tested sesbanimide R *in vitro* against cell lines of liver carcinoma (HepG2), endocervical adenocarcinoma (KB3.1), colon carcinoma (HCT-116), and lung carcinoma (A549). The 50% inhibitory concentration (IC_{50}) values against HePG2 (23 nM; 95% confidence interval [CI], 9 to 65), HCT-116 (39 nM; 95% CI, 28 to 54), and KB3.1 (20 nM; 95% CI, 15 to 28) are comparable to those of sesbanimide F, which exhibits a compound concentration that produces 50% cell growth inhibition (GI_{50}) of 20 nM against A549 cells [28, 60]. To better compare sesbanimides R and F, we also tested sesbanimide R against A549 cells, which resulted in an IC_{50} of 30 nM (95% CI, 21 to 40). These results indicate that the arginine moiety has no effect on cytotoxicity, which falls well within the range commonly observed for glutarimide-containing polyketides and other members of the sesbanimide compound family [28, 46, 47, 60].

3.4 Conclusion

We unambiguously assigned a new member of the sesbanimide compound family to a *trans*-AT polyketide synthase biosynthetic gene cluster from *Magnetospirillum gryphiswaldense* by inactivation and overexpression of the cluster and statistical analysis of the strains' metabolome.

Sesbanimide R belongs to the sesbanimide family of natural products. We suggest a biosynthesis pathway which is largely in line with the one proposed in a parallel study for sesbanimides A, C, D, E, and F [28, 60]. In contrast to these compounds, sesbanimide R contains a terminal arginine moiety, which perfectly matches the *in silico* predictions of the BGC.

Sesbanimides were isolated originally from the seeds of *Sesbania drummondii* [48] and later from marine agrobacteria, indicating that symbiotic microorganisms are the actual sources for these metabolites rather than the plant [28, 49, 60], a finding which is further supported by our study. Sesbanimide R is of interest owing to its cytotoxic bioactivity against several carcinoma cell lines, which is a characteristic of glutarimide-containing polyketides [45, 46]. The potent cytotoxic activity makes it a candidate for further investigations regarding its mode of action and development as an antitumor agent. As in other bacteria, the role of sesbanimide R for the physiology and fitness of *M. gryphiswaldense* in its freshwater habitat remains elusive and requires further investigations.

Sesbanimide R is the first natural product identified and isolated from a magnetotactic bacterium. In addition to its well-established property to produce biogenic magnetic nanoparticles, it makes the tractable strain *M. gryphiswaldense* highly interesting also as a producer of secondary metabolites. Since numerous biosynthetic gene clusters encoding putative polyketide synthases and nonribosomal peptide synthetases are present in the genomes of many different MTB (Table S4), our study sets the stage for exploring this highly diverse group of prokaryotes as a potential source for the future discovery of novel secondary metabolites.

3.5 Materials and Methods

In silico analysis of the genome of magnetotactic bacteria and bioinformatics methods

The *Magnetospirillum gryphiswaldense* genome (accession no. CP027527) and genomes of other magnetotactic bacteria were screened for secondary metabolites biosynthetic gene clusters using the bioinformatic tool *antibiotics & Secondary Metabolite Analysis Shell* (antiSMASH version 5.1.2)[30,57]. The amino acid sequence was aligned with the Basic Local Alignment Search Tool (BLASTp) against the publicly available database to find homologous proteins and to predict the functions of the ORFs. The presence of homologous ORFs in PHM037/PHM038 strains[28,61] was searched using the software Geneious Prime® (Biomatters Ltd., Auckland, New Zealand, 2020.0.3) (http://www.geneious.com). Furthermore, PKS and NRPS domain architecture and specificities present in the cluster were considered using TransATor (http://transator.ethz.ch) and Pfam database[54].

Bacterial strains and culture conditions

Escherichia coli was grown in lysogeny broth (LB) at 37°C and shaking at 180 rpm. Donor strain *E. coli* WM3064 (W. Metcalf, unpublished) was cultivated with 0.1 mM DL-α, Ɛ-diaminopimelic acid (DAP). *M. gryphiswaldense* was grown micro-aerobically at 28°C in modified flask standard medium (FSM)[33] with moderate agitation at 120 rpm, if not mentioned otherwise. Optical density (OD) and magnetic response (C_{mag}) of *M. gryphiswaldense* strains were determined photometrically at 565 nm as reported earlier[31]. Antibiotic selection was achieved by the addition of kanamycin at a concentration of 5 µg/ml (*M. gryphiswaldense*), and 25 µg/ml (*E. coli*). For Agar media, 1.5% (wt/vol) agar added to the liquid culture medium. Strains and vectors used in this study are given in Table S5.

Molecular and genetic techniques

Oligonucleotides (Table S5) were purchased from Sigma-Aldrich (Steinheim, Germany). Chromosomal DNA of *M. gryphiswaldense* was isolated using a kit from Zymo Research, USA. Plasmids were constructed by standard recombinant techniques as described below. All constructs and selected amplicon from the mutants were sequenced by Macrogen Europe (Amsterdam, Netherlands).

Construction of markerless site-specific deletion and activation of *trans*-AT PKS cluster

Markerless in-frame deletion of core-biosynthetic biosynthetic genes of *trans*-AT PKS cluster and insertion of a promoter in front of the cluster were conducted using homologous recombination-based on counter

selection systems described previously[55]. For the construction of the deletion plasmid, homologous regions of ca. 1.6 kb located upstream of MSR-1_15620 (Locus tag) including the first three codons of MSR-1_15620 and downstream of MSR-1_15650 with its last three codons were amplified from gDNA of *M. gryphiswaldense* using a proof-reading DNA polymerase and primer pairs RPA595/RPA596 and RPA597/RPA598. The PCR-products were purified from agarose gel using a gel extraction kit (Zymo Research, USA) and cloned into pORFM [55] digested with *Sal*I and *Not*I by Gibson assembly [56].

For activation of the *trans*-AT PKS cluster, the strong promoter $P_{mamDC45}$ with the spacing-optimized ribosome binding site (oRBS) was amplified from pAP150 [35] with primer pair RPA939/940. Homologous arms consisting of ca. 1.5 kb of C-terminus of MSR1_15590 and N-terminus of MSR1_15600 were amplified from gDNA of *M. gryphiswaldense* using primer pairs RPA937/938 and RPA940/941. The purified PCR products were assembled into pORFM[55], digested with *Sal*I and *Not*I by Gibson assembly[56] with the $P_{mamDC45}$-oRBS in between the two homologous arms. 5 µl of the Gibson assembly reaction was transformed into chemically competent *E. coli* DH5α[57], and the presence of the cloned fragment was confirmed by colony PCR using a pair of RPA484/485. The plasmid was isolated from the correct clone using Zymo Research kit, USA, and sequenced by Macrogen Europe (Armsterdam, Netherlands).

Conjugation

Plasmid transfer by biparental conjugation was performed with donor strain *E. coli* WM3064 consisting of the verified construct and *M. gryphiswaldense* as the acceptor strain as reported previously.[25] In-frame markerless chromosomal deletion and insertion were generated following the conjugative transfer of the plasmid to *M. gryphiswaldense* and homologous recombination utilizing GalK-based counterselection as previously described.[55] Successful deletion and insertion yielded Δ*trans-at-pks* and $P_{mamDC45}$-*trans-at-pks* strains, respectively. The mutants were confirmed by PCR using primers (Table S5) specific to sequences adjacent to the homologous regions, and verified by Sanger-sequencing of the amplicons.

Growth curve and cell length analysis

For growth analyses, the strains were grown in 24 well plates (Sarstedt, Nümbrecht) in 1 ml FSM[33] in a microplate reader (infinite 200Pro, Tecan, Switzerland) with an automated reading of absorbance (560 nm) every 20 min for 150 cycles under aerobic conditions at 28°C shaking at 140 rpm. Absorbance values were corrected using FSM medium as blank. Cell length of the strains was estimated with the ImageJ plugin MicrobeJ 5.13i (58) using the 'SHAPElength' cell shape descriptor. Analysis of cell length was done as reported previously (59).

Transmission electron microscopy

For TEM analysis, strains (wild type, Δ*trans-at-pks*, and $P_{mamDC45}$-*trans-at-pks*) cultivated in 6 well plates (Sarstedt, Nümbrecht) under micro-oxic conditions at 24°C for 48 hrs, fixed in formaldehyde (1.8%), adsorbed onto carbon-coated copper grids (F200-CU Carbon support film, 200 Mesh, Electron Microscopy Sciences, Hatfield, UK), and washed three times with ddH_2O. TEM was performed on JEM-2100 (JEOL Ltd., Tokyo, Japan) with an accelerating voltage of 80 kV. Images were captured with a Gatan Model 782 ES500W Erlangshen CCD camera (Gatan Inc., Pleasanton, USA) with the software Digital MicrographTM 1.80.70 (Gatan Inc., Pleasanton, USA). For data analysis and measurements, the software ImageJ Fiji V1.50c[60] was used.

Cultivation of strains for statistical analysis of the metabolome

For the screening of secondary metabolites, *M. gryphiswaldense* and Δ*trans-at-pks* strains were cultivated at 28°C in FSM medium[33] with an initial OD_{565} of 0.01 under aerobic, microoxic and anaerobic conditions in 500 ml baffled Erlenmeyer flasks, and Duran Laboratory flasks with rubber-stopper containing 50 ml medium, and in 250ml Duran Laboratory flasks containing 240 ml degassed medium with rubber-stoppers, respectively. 1 mL (v/v) sterile amberlite resin XAD-16 (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany) was added to the culture grown under aerobic and micro-oxic conditions and 5 ml (v/v) XAD-16 into the culture grown under anaerobic conditions. The culture under aerobic condition was agitated at 150 rpm. The cells and the resin were harvested together by centrifugation after 60 hours of incubation before extraction.

To access the activation of the cluster, wild type, $P_{mamDC45}$-*trans-at-pks*, and Δ*trans-at-pks* strains were cultivated under aerobic condition at 28°C in 100 ml FSM medium in 1L baffled Erlenmeyer flask with starting OD_{565} of 0.01 at 150 rpm. The culture was supplemented with 2 ml (v/v) sterile amberlite resin XAD-16 (Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany). After 60 hours of incubation, the cells and resin were harvested together by centrifugation.

Fermenter cultivation

Up-scale cultivation of the $P_{mamDC45}$-*trans-at-pks* strain was done in a 10 L BioFlow® 320 Fermenter (Eppendorf)[34] under aerobic conditions with an initial OD_{565} of 0.04. A 900 ml culture grown under aerobic conditions was used as an inoculum for 9 L culture which was supplemented with 200 ml (v/v) XAD-16. The cells and resin were harvested together by centrifugation after 60 hours of incubation and dried in Lyophiliser before extraction.

Extraction of the cell pellet and resin

For screening of the target compound(s), the cell pellet and resin of each culture was extracted with 50 mL methanol for 1h. The extract was then dried and resolved in 2mL methanol. This extract was then centrifuged for 5 min at 215 g and diluted 1:10 prior to analysis with LC-MS system 1a and processing with metaboscape 5.0 (bruker).

Extraction and isolation of sesbanimide R

The dry cells and resin from the upscaled fermentation were extracted three times with 500mL methanol. The extract was subsequently partitioned between hexane, ethylacetate and water. Sesbanimide R was detected solely in the aqueous layer. This layer was then dried and resuspended in methanol. Sesbanimide R was isolated from this pre-purified extract using LC-MS system 2. During purification, it became apparent, that sesbanimide R is unstable during prolonged exposure to light and oxygen simultaneously. Therefore, all purification steps were carried out with minimal exposure to light.

LC-MS systems:

All analytical LC-MS measurements were performed on a Dionex Ultimate 3000 RSLC system using a BEH C18, 100 x 2.1 mm, 1.7 µm particle diameter (dp) column (Waters, Germany), coupled to a maXis 4G hr-ToF mass spectrometer (Bruker Daltonics, Germany) using the Apollo ESI source. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75 µL/min before entering the mass spectrometer.

LC-MS system 1a – standard measurements:

Separation of 1 µl sample was achieved by a linear gradient from H2O + 0.1 % formic acid (FA) (solvent A) to acetonitrile (ACN) + 0.1 % FA (solvent B) at a flow rate of 600 µL/min and 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5 % B, followed by an increase to 95 % B in 18 min to end up with a 2 min step at 95 % B before reequilibration under the initial conditions. Mass spectra were acquired in centroid mode ranging from 150 – 2500 *m/z* at a 2 Hz scan rate.

LC-MS system 1b – Marfey's Method:

Separation of 1 µl sample was achieved by a gradient from (A) H2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 µL/min and 45 °C. The gradient was as follows: Ramp in 1 min from 5% B to 10%B, in 14 min to 35% B, in 7 min to 55% B and in 3 min to 80 % B. This is followed by a 1 min step at 80% B before

reequilibration with the initial conditions. Mass spectra were acquired in centroid mode ranging from 250 – 3000 *m/z* at a 2 Hz scan rate.

LC-MS system 1c – MS/MS measurements:

Separation of 1 µl sample was achieved by a linear gradient from (A) H_2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 600 µL/min and 45 °C. The gradient was initiated by a 0.5 min isocratic step at 5 % B, followed by an increase to 95 % B in 18 min to end up with a 2 min step at 95 % B before reequilibration under the initial conditions. Mass spectra were acquired in centroid mode ranging from 150 – 2500 *m/z* at a 2 Hz scan rate. Ions were selected for fragmentation by scheduled precursor list and the collision energy was determined by mass-and charge state dependent stepping from 25 to 60 eV.

LC-MS system 2

The final purification was performed on a Dionex Ultimate 3000 SDLC low pressure gradient system using a Luna, 5u, C18(2), 100A, 250 x 100 mm column (Phenomenex). Separation of 50 µl sample was achieved by a gradient from (A) H_2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 5 mL/min and 45 °C. The gradient was as follows: A two min isocratic step at 5 %B, followed by a ramp to 35 %B in three min, ramp to 50 %B in 20 and to 95%B in 1 min. This 3 min wash step was followed by a return to initial conditions in 1 min and reequilibration for 3 min. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 0.525 mL/min before entering the Thermo Fisher Scientific ISQ^{TM} EM single quadrupole mass spectrometer. Mass spectra were acquired by selected ion monitoring (SIM) at *m/z* 692.38 $[M+H]^+$.

Statistical analysis

Duplicates of wild type, and Δ*trans-at-pks* cultures were measured twice with LC-MS system 1a. Feature finding and bucketing was performed with the following parameters: Minimum Intensity: 5000; Minimum spectra for extraction: 5; minimum spectra for recursive feature extraction: 3. Recursive Feature extraction was performed when a feature was present in 2 out of 8 analyses and features were included in the bucket table when present in 3 out of 8 analyses after recursive feature extraction. Principal component analysis was performed to find differences between the two groups (wild type, and Δ*trans-at-pks*) with four analyses in each group. The PCA results were normalized with a logarithmic algorithm to account for low intensity features. Features that accounted for the largest difference between the data sets were reevaluated in the raw data.

Marfey's method to elucidate the stereochemistry of the arginine moiety

100 µg of sesbanimide R was dissolved in 100µL 6N HCL and incubated for 45 min at 110°C. It was subsequently dried under N_2 stream and redissolved in 110µL dH_2O. This was then split into 2 x 50 µL and 20 µL L-FDLA or D-FDLA and 20µL of $NaHCO_3$ were added. The reaction was shaken at 700 rpm and 40°C for 2h and then stopped with the addition of 10µL 2N HCl. The reaction was then diluted with 300µL acetonitrile, centrifuged and analyzed using LC-MS system 1b. The same reaction and measurement, without the hydrolysis, were performed with L- and D-arginine as a reference. The retention times of the derivatized standards were compared to the derivatized samples to assign the stereochemistry.

Saponification of sesbanimide R

50 µg of sesbanimide R were dried and redisolved in 100 µL 2 M NaOH. The reaction was stopped instantly by adding 200 µL1 M HCL and an aliquot of the solution was diluted 1:5 with acetonitrile and analyzed on LC-MS system 1a.

Cytotoxicity assays with HCT-116, HepG2, KB3.1 cells and A549 cells

The cell lines were obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ) and cultured under conditions recommended by the depositor. Cells were grown and diluted to 5 x 10^4 cells per well of 96-well plates in 180 µL complete medium. After 2 h of equilibration, the cells were treated with a serial dilution of sesbanimide R in methanol. 20 µL of 5 mg/mL MTT (thiazolyl blue tetrazolium bromide) in PBS was added to each well after growing the cells for five days. The cells were further incubated for 2 h at 37°C, before the supernatant was discarded. Subsequently, the cells were washed with 100 µL PBS and treated with 100 µL 2-propanol/10 N HCl (250:1) to dissolve formazan granules. Cell viability was measured as percentage relative to the respective methanol control by measuring the absorbance at 570 nm with a microplate reader (Tecan Infinite M200Pro). GraphPad Prism was used for sigmoidal curve fitting to determine the IC50 values as well as the calculation confidence intervals.

3.6 References

[1] Bazylinski DA, Lefèvre CT, Schüler D. 2013. Magnetotactic Bacteria, p. 453–494. In Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (ed), The Prokaryotes. Springer Berlin Heidelberg, Berlin, Heidelberg.

[2] Uebe R, Schüler D. 2016. Magnetosome biogenesis in magnetotactic bacteria. Nat Rev Microbiol 14:621–637. doi:10.1038/nrmicro.2016.99.

[3] McCausland HC, Komeili A. 2020. Magnetic genes: Studying the genetics of biomineralization in magnetotactic bacteria. PLoS Genet 16:e1008499. doi:10.1371/journal.pgen.1008499.

[4] Lee J-H, Huh Y-M, Jun Y-w, Seo J-w, Jang J-t, Song H-T, Kim S, Cho E-J, Yoon H-G, Suh J-S, Cheon J. 2007. Artificially engineered magnetic nanoparticles for ultra-sensitive molecular imaging. Nat Med 13:95–99. doi:10.1038/nm1467.

[5] Hergt R, Dutz S, Röder M. 2008. Effects of size distribution on hysteresis losses of magnetic nanoparticles for hyperthermia. J Phys Condens Matter 20:385214. doi:10.1088/0953-8984/20/38/385214.

[6] Sun J, Li Y, Liang X-J, Wang PC. 2011. Bacterial Magnetosome: A Novel Biogenetic Magnetic Targeted Drug Carrier with Potential Multifunctions. J Nanomater 2011:469031–469043. doi:10.1155/2011/469031.

[7] Mickoleit F, Lanzloth C, Schüler D. 2020. A Versatile Toolkit for Controllable and Highly Selective Multifunctionalization of Bacterial Magnetic Nanoparticles. Small 16:e1906922. doi:10.1002/smll.201906922.

[8] Lefèvre CT, Bazylinski DA. 2013. Ecology, diversity, and evolution of magnetotactic bacteria. Microbiol Mol Biol Rev 77:497–526. doi:10.1128/MMBR.00021-13.

[9] Lin W, Pan Y, Bazylinski DA. 2017. Diversity and ecology of and biomineralization by magnetotactic bacteria. Environ Microbiol Rep 9:345–356. doi:10.1111/1758-2229.12550.

[10] Wenter R, Wanner G, Schüler D, Overmann J. 2009. Ultrastructure, tactic behaviour and potential for sulfate reduction of a novel multicellular magnetotactic prokaryote from North Sea sediments. Environ Microbiol 11:1493–1505. doi:10.1111/j.1462-2920.2009.01877.x.

[11] Abreu F, Morillo V, Nascimento FF, Werneck C, Cantão ME, Ciapina LP, Almeida LGP de, Lefèvre CT, Bazylinski DA, Vasconcelos ATR de, Lins U. 2014. Deciphering unusual uncultured magnetotactic multicellular prokaryotes through genomics. ISME J 8:1055–1068. doi:10.1038/ismej.2013.203.

[12] Monteil CL, Vallenet D, Menguy N, Benzerara K, Barbe V, Fouteau S, Cruaud C, Floriani M, Viollier E, Adryanczyk G, Leonhardt N, Faivre D, Pignol D, López-García P, Weld RJ, Lefevre CT. 2019.

Ectosymbiotic bacteria at the origin of magnetoreception in a marine protist. Nat Microbiol 4:1088–1095. doi:10.1038/s41564-019-0432-7.

[13] Lefèvre CT, Viloria N, Schmidt ML, Pósfai M, Frankel RB, Bazylinski DA. 2012. Novel magnetite-producing magnetotactic bacteria belonging to the Gammaproteobacteria. ISME J 6:440–450. doi:10.1038/ismej.2011.97.

[14] Lin W, Zhang W, Zhao X, Roberts AP, Paterson GA, Bazylinski DA, Pan Y. 2018. Genomic expansion of magnetotactic bacteria reveals an early common origin of magnetotaxis with lineage-specific evolution. ISME J 12:1508–1519. doi:10.1038/s41396-018-0098-9.

[15] Lin W, Zhang W, Paterson GA, Zhu Q, Zhao X, Knight R, Bazylinski DA, Roberts AP, Pan Y. 2020. Expanding magnetic organelle biogenesis in the domain Bacteria. Microbiome 8:152. doi:10.1186/s40168-020-00931-9.

[16] Jogler C, Wanner G, Kolinko S, Niebler M, Amann R, Petersen N, Kube M, Reinhardt R, Schüler D. 2011. Conservation of proteobacterial magnetosome genes and structures in an uncultivated member of the deep-branching Nitrospira phylum. Proc Natl Acad Sci U S A 108:1134–1139. doi:10.1073/pnas.1012694108.

[17] Kolinko S, Jogler C, Katzmann E, Wanner G, Peplies J, Schüler D. 2012. Single-cell analysis reveals a novel uncultivated magnetotactic bacterium within the candidate division OP3. Environ Microbiol 14:1709–1721. doi:10.1111/j.1462-2920.2011.02609.x.

[18] Kolinko S, Wanner G, Katzmann E, Kiemer F, Fuchs BM, Schüler D. 2013. Clone libraries and single cell genome amplification reveal extended diversity of uncultivated magnetotactic bacteria from marine and freshwater environments. Environ Microbiol 15:1290–1301. doi:10.1111/1462-2920.12004.

[19] Kolinko S, Richter M, Glöckner F-O, Brachmann A, Schüler D. 2014. Single-cell genomics reveals potential for magnetite and greigite biomineralization in an uncultivated multicellular magnetotactic prokaryote. Environ Microbiol Rep 6:524–531. doi:10.1111/1758-2229.12198.

[20] Lefèvre CT, Trubitsyn D, Abreu F, Kolinko S, Jogler C, Almeida LGP de, Vasconcelos ATR de, Kube M, Reinhardt R, Lins U, Pignol D, Schüler D, Bazylinski DA, Ginet N. 2013. Comparative genomic analysis of magnetotactic bacteria from the Deltaproteobacteria provides new insights into magnetite and greigite magnetosome genes required for magnetotaxis. Environ Microbiol 15:2712–2735. doi:10.1111/1462-2920.12128.

[21] Hoffmann T, Krug D, Bozkurt N, Duddela S, Jansen R, Garcia R, Gerth K, Steinmetz H, Müller R. 2018. Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. Nat Commun 9:803. doi:10.1038/s41467-018-03184-1.

[22] Araujo ACV, Abreu F, Silva KT, Bazylinski DA, Lins U. 2015. Magnetotactic bacteria as potential sources of bioproducts. Mar Drugs 13:389–430. doi:10.3390/md13010389.

[23] Schleifer KH, Schüler D, Spring S, Weizenegger M, Amann R, Ludwig W, Köhler M. 1991. The Genus Magnetospirillum gen. nov. Description of Magnetospirillum gryphiswaldense sp. nov. and Transfer of Aquaspirillum magnetotacticum to Magnetospirillum magnetotacticum comb. nov. Systematic and Applied Microbiology 14:379–385. doi:10.1016/S0723-2020(11)80313-9.

[24] Schüler D, Köhler M. 1992. The isolation of a new magnetic spirillum. Zentralblatt für Mikrobiologie 147:150–151. doi:10.1016/S0232-4393(11)80377-X.

[25] Schultheiss D, Schüler D. 2003. Development of a genetic system for Magnetospirillum gryphiswaldense. Arch Microbiol 179:89–94. doi:10.1007/s00203-002-0498-z.

[26] Zwiener Theresa, Dziuba Marina, Mickoleit Frank, Rückert Christian, Busche Tobias, Kalinowski Jörn, Uebe René, Schüler Dirk. Towards a 'chassis' for bacterial magnetosome biosynthesis: Genome streamlining of Magnetospirillum gryphiswaldense by multiple deletions. Microb Cell Fact.

[27] Schüler D, Monteil CL, Lefevre CT. 2020. Magnetospirillum gryphiswaldense. Trends Microbiol 28:947–948. doi:10.1016/j.tim.2020.06.001.

[28] Kačar D, Cañedo LM, Rodríguez P, Gonzalez E, Galán B, Schleissner C, Leopold-Messer S, Piel J, Cuevas C, La Calle F de, García JL. 2020. Identification of trans-AT polyketide clusters in two marine bacteria reveals cryptic similarities between distinct symbiosis factors. doi:10.1101/2020.09.18.303172.

[29] Meoded RA, Ueoka R, Helfrich EJN, Jensen K, Magnus N, Piechulla B, Piel J. 2018. A Polyketide Synthase Component for Oxygen Insertion into Polyketide Backbones. Angew Chem Int Ed Engl 57:11644–11648. doi:10.1002/anie.201805363.

[30] Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81-W87. doi:10.1093/nar/gkz310.

[31] Schüler D, Uhl R, Bäuerlein E. 1995. A simple light scattering method to assay magnetism in Magnetospirillum gryphiswaldense. FEMS Microbiology Letters 132:139–145. doi:10.1111/j.1574-6968.1995.tb07823.x.

[32] Cortina NS, Krug D, Plaza A, Revermann O, Müller R. 2012. Myxoprincomide: a natural product from Myxococcus xanthus discovered by comprehensive analysis of the secondary metabolome. Angew Chem Int Ed Engl 51:811–816. doi:10.1002/anie.201106305.

[33] Heyen U, Schüler D. 2003. Growth and magnetosome formation by microaerophilic Magnetospirillum strains in an oxygen-controlled fermentor. Appl Microbiol Biotechnol 61:536–544. doi:10.1007/s00253-002-1219-x.

[34] Riese CN, Uebe R, Rosenfeldt S, Schenk AS, Jérôme V, Freitag R, Schüler D. 2020. An automated oxystat fermentation regime for microoxic cultivation of Magnetospirillum gryphiswaldense. Microb Cell Fact 19:206. doi:10.1186/s12934-020-01469-z.

[35] Borg S, Hofmann J, Pollithy A, Lang C, Schüler D. 2014. New vectors for chromosomal integration enable high-level constitutive or inducible magnetosome expression of fusion proteins in Magnetospirillum gryphiswaldense. Appl Environ Microbiol 80:2609–2616. doi:10.1128/AEM.00192-14.

[36] Schübbe S, Würdemann C, Peplies J, Heyen U, Wawer C, Glöckner FO, Schüler D. 2006. Transcriptional organization and regulation of magnetosome operons in Magnetospirillum gryphiswaldense. Appl Environ Microbiol 72:5757–5765. doi:10.1128/AEM.00201-06.

[37] Helfrich EJN, Ueoka R, Dolev A, Rust M, Meoded RA, Bhushan A, Califano G, Costa R, Gugger M, Steinbeck C, Moreno P, Piel J. 2019. Automated structure prediction of trans-acyltransferase polyketide synthase products. Nat Chem Biol 15:813–821. doi:10.1038/s41589-019-0313-7.

[38] Xie X, Khosla C, Cane DE. 2017. Elucidation of the Stereospecificity of C-Methyltransferases from trans-AT Polyketide Synthases. J. Am. Chem. Soc. 139:6102–6105. doi:10.1021/jacs.7b02911.

[39] Blin K, Shaw S, Kautsar SA, Medema MH, Weber T. 2020. The antiSMASH database version 3: increased taxonomic coverage and new query features for modular enzymes. Nucleic Acids Res. doi:10.1093/nar/gkaa978.

[40] Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A. 2020. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res 48:D265-D268. doi:10.1093/nar/gkz991.

[41] Kjaerulff L, Raju R, Panter F, Scheid U, Garcia R, Herrmann J, Müller R. 2017. Pyxipyrrolones: Structure Elucidation and Biosynthesis of Cytotoxic Myxobacterial Metabolites. Angew Chem Int Ed Engl 56:9614–9618. doi:10.1002/anie.201704790.

[42] Nguyen T, Ishida K, Jenke-Kodama H, Dittmann E, Gurgui C, Hochmuth T, Taudien S, Platzer M, Hertweck C, Piel J. 2008. Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. Nat Biotechnol 26:225–233. doi:10.1038/nbt1379.

[43] Lešnik U, Lukežič T, Podgoršek A, Horvat J, Polak T, Šala M, Jenko B, Harmrolfs K, Ocampo-Sosa A, Martínez-Martínez L, Herron PR, Fujs Š, Kosec G, Hunter IS, Müller R, Petković H. 2015. Construction of a new class of tetracycline lead structures with potent antibacterial activity through biosynthetic engineering. Angew Chem Int Ed Engl 54:3937–3940. doi:10.1002/anie.201411028.

[44] Niehs SP, Kumpfmüller J, Dose B, Little RF, Ishida K, Flórez LV, Kaltenpoth M, Hertweck C. 2020. Insect-Associated Bacteria Assemble the Antifungal Butenolide Gladiofungin by Non-Canonical Polyketide Chain Termination. Angew Chem Int Ed Engl 59:23122–23126. doi:10.1002/anie.202005711.

[45] Helfrich EJN, Piel J. 2016. Biosynthesis of polyketides by trans-AT polyketide synthases. Nat Prod Rep 33:231–316. doi:10.1039/c5np00125k.

[46] Röttig M, Medema MH, Blin K, Weber T, Rausch C, Kohlbacher O. 2011. NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res. 39:W362-W367. doi:10.1093/nar/gkr323.

[47] Bhushan R, Bruckner H. 2004. Marfey's reagent for chiral amino acid analysis: A review. Amino Acids 27:231–247. doi:10.1007/s00726-004-0118-0.

[48] Gorst-Allman CP, Steyn PS, Vleggaar R, Grobbelaar N. 1984. Structure elucidation of sesbanimide using high-field n.m.r. spectroscopy. J Chem Soc Perk T 1:1311. doi:10.1039/p19840001311.

[49] Rajski SR, Shen B. 2010. Multifaceted modes of action for the glutarimide-containing polyketides revealed. Chembiochem 11:1951–1954. doi:10.1002/cbic.201000370.

[50] Sugawara K, Nishiyama Y, Toda S, Komiyama N, Hatori M, Moriyama T, Sawada Y, Kamei H, Konishi M, Oki T. 1992. Lactimidomycin, a new glutarimide group antibiotic. Production, isolation, structure and biological activity. J Antibiot (Tokyo) 45:1433–1441. doi:10.7164/antibiotics.45.1433.

[51] Powell RG, Smith CR, Weisleder D, Matsumoto G, Clardy J, Kozlowski J. 1983. Sesbanimide, a potent antitumor substance from Sesbania drummondii seed. J. Am. Chem. Soc. 105:3739–3741. doi:10.1021/ja00349a081.

[52] Acebal C, Alcazar R, Cañedo LM, La Calle F de, Rodriguez P, Romero F, Fernandez Puentes JL. 1998. Two marine Agrobacterium producers of sesbanimide antibiotics. J Antibiot (Tokyo) 51:64–67. doi:10.7164/antibiotics.51.64.

[53] Gorst-Allman CP, Steyn PS, Vleggaar R, Grobbelaar N. 1984. Structure elucidation of sesbanimide using high-field n.m.r. spectroscopy. J. Chem. Soc., Perkin Trans. 1:1311. doi:10.1039/P19840001311.

[54] Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279-85. doi:10.1093/nar/gkv1344.

[55] Raschdorf O, Plitzko JM, Schüler D, Müller FD. 2014. A tailored galK counterselection system for efficient markerless gene deletion and chromosomal tagging in Magnetospirillum gryphiswaldense. Appl Environ Microbiol 80:4323–4330. doi:10.1128/AEM.00588-14.

[56] Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 6:343–345. doi:10.1038/nmeth.1318.

[57] Hanahan D. 1983. Studies on transformation of Escherichia coli with plasmids. Journal of Molecular Biology 166:557–580. doi:10.1016/s0022-2836(83)80284-8.

[58] Ducret A, Quardokus EM, Brun YV. 2016. MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis. Nat Microbiol 1:16077. doi:10.1038/nmicrobiol.2016.77.

[59] Pfeiffer D, Toro-Nahuelpan M, Awal RP, Müller F-D, Bramkamp M, Plitzko JM, Schüler D. 2020. A bacterial cytolinker couples positioning of magnetic organelles to cell shape control. Proc Natl Acad Sci U S A. doi:10.1073/pnas.2014659117.

[60] Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez J-Y, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: an open-source platform for biological-image analysis. Nat Methods 9:676–682. doi:10.1038/nmeth.2019.

[61] Kačar D, Cañedo LM, Rodríguez P, González EG, Galán B, Schleissner C, Leopold-Messer S, Piel J, Cuevas C, de la Calle F, García JL. 18 March 2021. Identification of *trans*-AT polyketide clusters in two marine bacteria reveals cryptic similarities between distinct symbiosis factors. Soc Appl Microbiol. doi:10.1111/1462-2920.15470

3.7 Supplemental Material

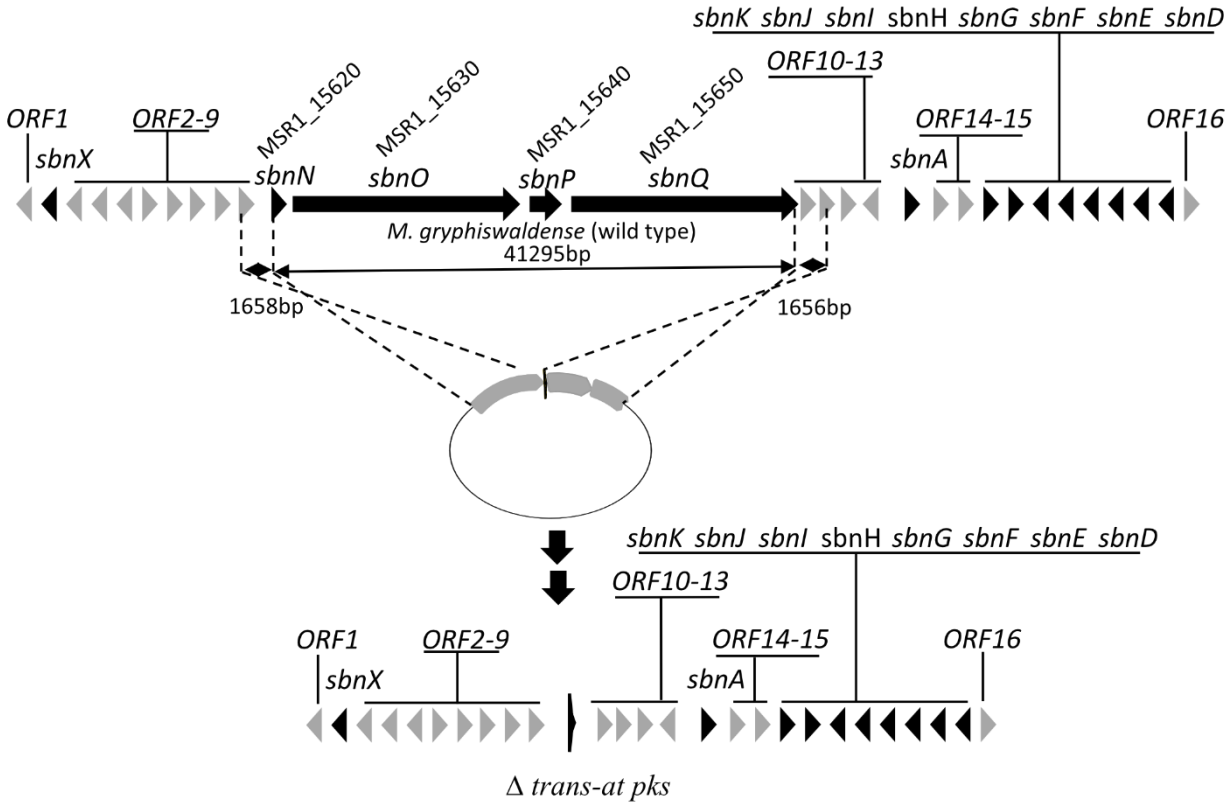


Figure S 1*:* Simplified schematic illustration of an in-frame deletion of core-biosynthetic genes of *trans*-AT PKS cluster in *Magnetospirillum gryphiswaldense*, yielding the Δ*trans-at-pks* strain.

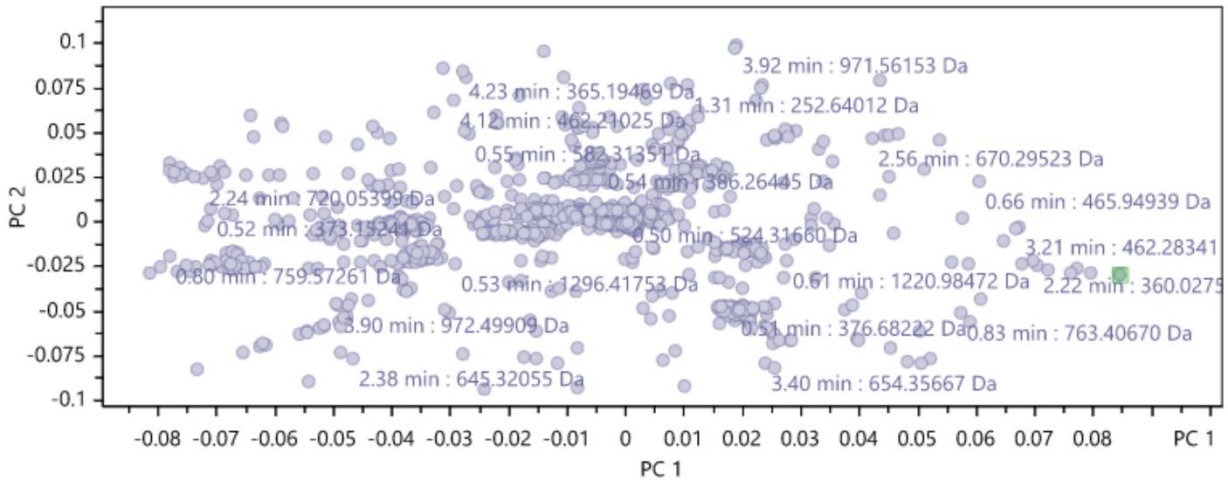


Figure S 2: Principal-component analysis of the extracts of the *M. gryphiswaldense* wild type and Δ*trans-at-pks* strain. The more a component is responsible for a difference between the data sets, the farther it is to the right (wild type) or to the left (Δ*trans-at-pks*) from the center; the highlighted feature represents the target mass 692.38 *m/z*, which was assigned to the *trans*-AT PKS cluster and was found only in the wild-type strain.

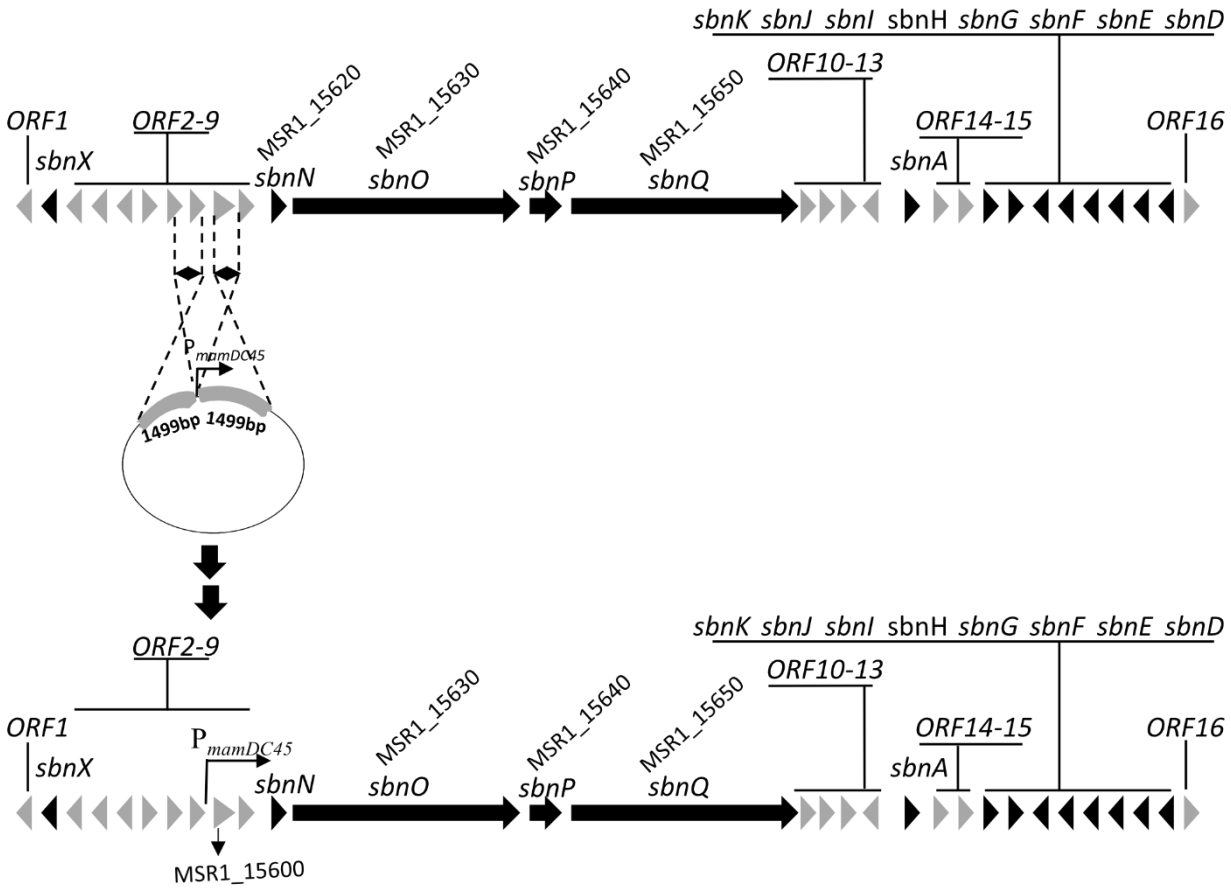


Figure S 3: Simplified schematic illustration of the insertion of promoter $P_{mamDC45}$ in front of MSR-1_15600 in the *trans*-AT PKS cluster in *Magnetospirillum gryphiswaldense*, generating the $P_{mamDC45}$-*trans-at-pks* strain.

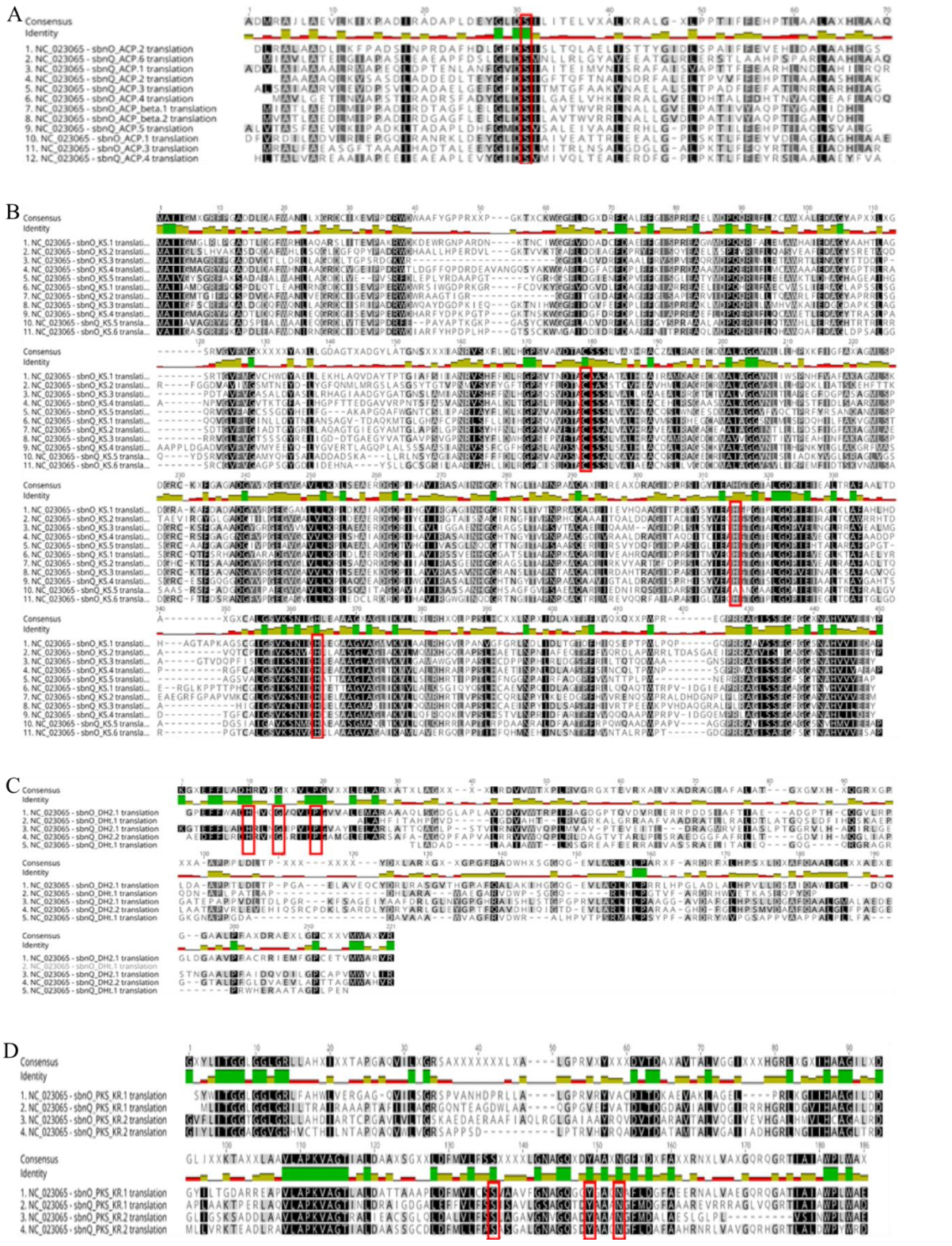


Figure S 4: (A) Alignment of ACP domains of *sbnO* and *sbnQ*. All ACPs contain the active-site serine. (B) Alignment of KS domains of *sbnO* and *sbnQ*. All KS domains contain the conserved cysteine and two histidines except for KS5 of *sbnQ*. (C) Alignment of DH domains of *sbnO* and *sbnQ*. DH2 (DHt.1) of *sbnO* and DH3 (DHt.1) of *sbnQ* are missing the conserved HXXXGXXXXP motif. (D) Alignment of KR domains of *sbnO* and *sbnQ*. All domains contain the conserved active-site residues serine, tyrosine, and asparagine.

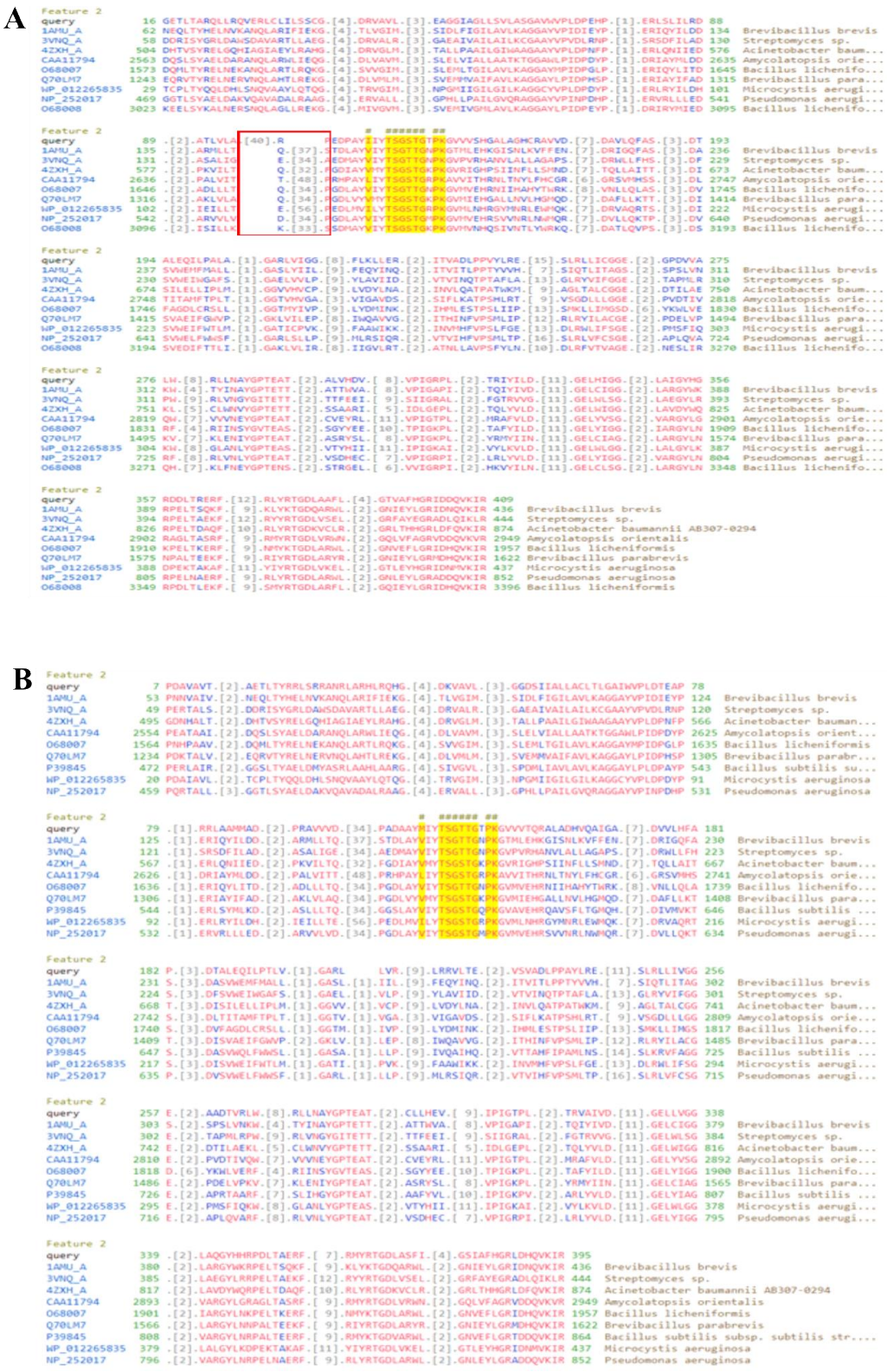


Figure S 5: (A) Alignment of the amino acid sequence of the A-domain from the last module of SbnQ from *S. indica* PHM037 with the reference domains of the CDD search tool. The active-site residues are highlighted in yellow. The residues just before the active site (red box) do not match the reference sequences in length. (B) Alignment of the

amino acid sequence of the A-domain from the last module of SbnQ from *M. gryphiswaldense* with the reference domains of the CDD search tool. The active-site residues are highlighted in yellow.

Table S 1: Putative secondary metabolites gene clusters present in the genome of *Magnetospirillum gryphiswaldense* with locus tag and size.

Gene Clusters	Locus tags
Lasso peptide	MSR-1_06390-06890
Aryl polyene	MSR-1_09890-10640
Homoserine lactone	MSR-1_16040-16260
Trans-AT PKS	MSR-1_15520-15810

Table S 2: Annotation of the sesbanimide gene cluster in Magnetospirillum gryphiswaldense and comparison the sesbanimide cluster in PHM037 and PHM038

	Putative function/ homologue	Accession number of closest protein homologue	Cover/Pairwise identity [%]	Corresponding gene in PHM037/ PHM038
Orf1	CoA transferase	WP_024079852	100/100	-
Orf2	response regulator	WP_024079854	100/100	-
Orf3	response regulator receiver modulated diguanylate cyclase	CDK98834	100/100	-
Orf4	Hpt domain-containing protein	WP_024079856	73.5/100	-
Orf5	diguanylate cyclase	WP_024079857	90.4/100	-
Orf6	amino acid ABC transporter substrate-binding protein	WP_024079858	100/100	-
Orf7	response regulator	WP_024079859	100/100	-
Orf8	cyclic peptide export ABC transporter	WP_024079860	100/100	SbnL
Orf9	cyclic peptide export ABC transporter	WP_024079861	100/100	SbnM
Orf10	Band 7 protein	CDK98845	100/100	SbnR
Orf11	ABC transporter substrate-binding protein	WP_024079867	100/100	SbnS
Orf12	hypothetical protein	WP_024079868	100/100	SbnT
Orf13	DUF-697 domain-containing protein	WP_024079869	94.7/100	SbnU
Orf14	putative lysine-arginine-ornithine-binding periplasmic protein	CDK98850	100/100	-
Orf15	cation:dicarboxylase symporter family transporter	WP_024079872	88.7/99.8	-
Orf16	metallophosphoesterase	WP_041633527	100/100	SbnC
SbnA	acyltransferase domain-containing protein	WP_158497748	73.8/99.7	SbnA
SbnD	Fkbm family methyltransferase	WP_158497749	87.2/100	SbnD
SbnE	cytochrome P450	WP_024079879	100/100	SbnE
SbnF	hydroxymethylglutaryl-CoA synthase family protein	WP_024079878	100/100	SbnF
SbnG	acyl carrier protein	WP_024079877	100/100	SbnG
SbnH	decarboxylase	WP_024079876	100/100	SbnH

SbnI	enoyl-CoA hydratase/ isomerase	WP_024079875	100/100	SbnI
SbnJ	asparagine synthase (glutamine-hydrolyzing)	WP_024079874	100/100	SbnJ
SbnK	acyl carrier protein	WP_106002060	100/96.3	SbnK
SbnN	ACP S-malonyltransferase	WP_024079862	100/100	SbnN
SbnO	SDR family NAD(P)-dependent oxidoreductase	WP_024079863	100/100	SbnO
SbnP	monooxygenase	OJX77727	99.5/87.0	SbnP
SbnQ	non-ribosomal peptide synthetase	WP_024079865	100/100	SbnQ
SbnX	acyl-CoA dehydrogenase	WP_024079853	100/100	SbnX

Table S 3: Substrate specificities as predicted by the *trans*ATor tool and sorted according to E-value. The predictions that fit the structure and biosynthesis proposal are highlighted in bold.

	prediction		**e value**	**score**
KS1 (sbnO)	Clade 95	various specificities	1.1E-179	590.7
	Clade 109	completely reduced	1.0E-176	580.7
	Clade 8	**unusual starter (AMT/Succinate)**	2.8E-217	714.5
	Clade 136	β D-OH	2.8E-178	585.8
	Clade 96	various specificies (mainly α-Me)	2.0E-177	583.2
KS2 (sbnO)	Clade_64	non-elongating (double bonds (mostly z-configured))	9.1E-218	715.9
	Clade 5	amino acids (oxa/thia)	2.9E-151	497.0
	Clade 82	**double bonds (mostly e-configured)**	5.8E-151	495.9
	Clade 115	**β-keto or double bonds**	7.8E-149	488.8
	Clade 99	**double bonds (e-configured)**	4.4E-147	483.4
KS3 (sbnO)	Clade 95	various specificities	2.1E-149	490.9
	Clade 109	completely reduced	1.4E-141	464.9
	Clade 96	various specificies (mainly α-Me)	4.3E-166	545.8
	Clade 12	**vinylogous chain branching**	2.5E-155	510.2
	Clade 136	β D-OH	5.0E-153	502.6
KS4 (sbnO)	**Clade_7**	**ß D-OH**	1.2E-207	683.1
	Clade_110	**ß D-OH or double bonds (e-configured)**	8.7E-212	696.5
	Clade_140	**β D-OH**	9.2E-209	686.7
	Clade_137	**β D-OH**	8.9E-207	680.0
	Clade_62	**β D-OH (some with α L-Me)**	1.6E-201	662.6
KS5 (sbnO)	Clade_68	α L-OH/Me β D-OH	8.6E-174	571.1
	Clade_104	ß OMe or ß Me double bond	8.4E-170	557.7
	Clade_21	**α Me reduced/keto/D-OH**	2.5E-194	638.7
	Clade_74	**α Me reduced/keto/D-OH**	7.3E-193	633.9
	Clade_23	**α-Me**	5.9E-180	591.3
KS6 (sbnO)	Clade_104	ß OMe or ß Me double bond	1.9E-177	582.9
	Clade_86	α-L-Me red or OH	6.7E-169	554.9
	Clade_14	exomethyl/exoester	9.3E-196	643.5

	Clade_2	α-Me shifted double bond or OH	3.6E-174	572.2
	Clade_73	**exomethylene**	1.1E-169	557.3
KS7 (sbnQ)	**Clade_35**	**oxidative rearrangement**	4.7E-214	704.0
	Clade_95	various specificities	9.4E-167	548.0
	Clade_25	completely reduced	8.3E-173	567.8
	Clade_96	various specificies (mainly α-Me)	3.7E-172	565.8
	Clade_136	β D-OH	8.9E-169	554.5
KS8 (sbnQ)	Clade_95	various specificities	2.6E-181	596.0
	Clade_25	**completely reduced**	1.2E-201	662.9
	Clade_108	shifted double bonds	5.2E-191	627.9
	Clade_96	various specificies (mainly α-Me)	9.3E-188	617.2
	Clade_136	β D-OH	2.2E-173	569.7
KS9 (sbnQ)	**Clade_25**	**completely reduced**	9.4E-217	712.6
	Clade_108	shifted double bonds	2.9E-204	671.6
	Clade_96	various specificies (mainly α-Me)	1.2E-190	626.7
	Clade_11	shifted double bonds	6.5E-189	621.0
	Clade_136	β D-OH	5.4E-185	607.9
KS10 (sbnQ)	**Clade_82**	**double bonds (mostly e-configured)**	1.5E-227	748.3
	Clade_125	**double bonds (e-configured)**	6.4E-223	733.0
	Clade_129	**double bonds (e-configured)**	6.5E-213	700.1
	Clade_101	**double bonds**	9.2E-213	699.5
	Clade_99	**double bonds (e-configured)**	9.9E-213	699.7
KS11 (sbnQ)	**Clade_76**	**non-elongating (double bonds)**	1.7E-179	589.9
	Clade_142	**non-elongating (various)**	2.7E-167	549.6
	Clade_101	double bonds	1.6E-162	534.0
	Clade_90	β-keto or double bonds	4.1E-161	529.5
	Clade_115	β-keto or double bonds	8.7E-160	524.9

Table S 4: Putative open reading frames (ORFs) encoding polyketide synthases (PKSs), non-ribosomal peptide synthetases (NRPSs), or hybrid in PKS and NRPS gene clusters from different magnetotactic bacteria. *indicates that genome assembly of these species is not complete, and the number of Open Reading Frames (ORFs) might be variable in the final analysis.

Species	strain	Class	PKS	NRPS	Hybrid	Trans-AT PKS
Magnetospirillum gryphiswaldense	MSR-1	α	0	0	0	1 Trans-AT PKS
Magnetospirillum magneticum	AMB-1	α	1 T1PKS	1 NRPS-like	0	0
***Magnetospira* sp.**	QH-2	α	0	1 NRPS	0	0
***Magnetospira* sp.**	ME-1	α	1 T1PKS	1 NRPS-like	0	0
***Magnetovibrio blakemorei* ***	MV-1	α	0	0	0	1 Trans-AT PKS
***Magnetospirillum* sp.**	XM-1	α	1 T1PKS	1 NRPS-like	0	0
***Magnetofaba australis* ***	IT-1	α	1 T1PKS	0	0	0
***Magnetospirillum magnetotacticum* ***	MS-1	α	1 T1PKS	1 NRPS-like	0	0
***Magnetospirillum* sp. ***	SO-1	α	1 T1PKS	1 NRPS-like	0	0
***Magnetospirillum marisnigri* ***	SP-1	α	1 T3PKS	1 NRPS-like	0	0
***Magnetospirillum* sp. ***	mag 15-1	α	1 T1PKS	1 NRPS-like	0	1 Trans-AT PKS
***Magnetospirillum* sp. ***	64-120	α	0	0		1 Trans-AT PKS
Magneto-ovoid bacterium	MO-1	α	0	1 NRPS / 1 NRPS-like	0	0
Desulfovibrio magneticus	RS-1	δ	0	1 NRPS-like	0	0
***Desulfamplus magnetovallimortis* ***	BW-1	δ	2 T1PKS	0	0	0
Candidatus* Magnetoglobus multicellularis str.	Araruama	δ	1 T1PKS	5 NRPS / 2 NRPS-like		2 Trans-AT PKS like
Ectothiorhodospiraceae bacterium	BW-2	γ	T2PKS	1 NRPS-like	2 NRPS/T1PKS	
Gamma Proteobacterium	SS-5	γ	0	1 NRPS	2 NRPS/T1PKS	1 Trans-AT PKS
***Candidatus* Magnetobacterium casensis ***	MYR-1	Nitrospirae	0	1 NRPS-like	0	0
***Candidatus* Magnetobacterium bavaricum ***	TM-1	Nitrospirae	0	1 NRPS-like	0	0
***Candidatus* Magnetoovum chiemensis ***	CS-04	Nitrospirae	1 T1PKS	1 NRPS	0	0

Table S 5: Strains, vectors and primers used in this study.

Strain or vector	Relevant characteristic (s)	Reference and/or source
Strains		
E. coli		
DH5α	Host for cloning; F^+ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* ϕ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^-m_K^+$), λ^-	(1)
WM3064	Conjugation strain; *thrB1004 pro thi rpsL hsdS lacZ*ΔM15 *RP4-1360* Δ(*araBAD*)*567* Δ*dapA1341*::[*erm pir*]	William Metcalf at UIUC
M. gryphiswaldense		
Wild type	MSR-1 R3/S1; Rif^R, Sm^R	(2)
Δ*trans-at-pks*	Core-biosynthetic genes of trans-AT PKS deletion strain (MSR-1_15620-15650)	This study
$P_{mamDC45}$-*trans-at-pks*	Strain with chromosomally inserted $P_{mamDC45}$ promoter in front of MSR-1_15600	This study
Vectors		
pORFM	universal in-frame deletion/in-frame fusion vector for GalK-based counterselection; *npt galK tetR mob*RK2	(3)
pORFM-Δ*trans-at-pks*	Vector for chromosomal deletion of core-biosynthetic genes of trans-AT PKS gene cluster (MSR-1_15620-15650)	This study
pORFM-$P_{mamDC45}$-*trans-at-pks*	Vector for insertion of a promoter 1xPmamDC45-oRBS in front of MSR-1_15600	This study

Primers

No. (RPA)	Primer name	Sequence 5'-3'
Site-specific chromosomal deletion/insertion by homologous recombination		
595	F1_del_Trnpks-nrps	gtcattactggatctatcaacaggagtcctgcagtaggatgagcatcgccgctttcctggg
596	R1_del_Trnpks-nrps	gctggatcggttagcccgaggctttcatgcatggcctccttcgc
597	F2_del_Trnpks-nrps	ggaggccatgcatgaaagcctcgggctaaccgatccagcataatatg
598	R2_del_Trnpks-nrps	gcggcagcgtgaagctagcatcactagtctagcgcagcaggtcatcgatggagcgg
599	sq_Rev1_Trnpks_nrps	cccatacaggcggtcaaacag
600	sq_For2_Trnpks_nrps	cgaggtggtgttcgtggtc
601	sq_Rev2_Trnpks_nrps	ggcttttggcgatgatctgc
602	sq_For3_Trnpks_nrps	accgctacatcatcgtcgac
937	F1_Pro_in_PKS_NRPS	caggaaagacttaagctgcagtagggcccgggtgatggtcgcc
938	R1_Pro_in_PKS_NRPS	gagaactaagagctagtaaagcgaaaaagtcttacttgtcttgtccggcg
939	F2_Pro_in_PKS_NRPS	cgccatcgccggacaagacaagtaagacttttcgctttactagctc
940	R2_Pro_in_PKS_NRPS	ggtgctgacgagacgaagaacatgcatatgctgatctcctaagcttcgc
941	F3_Pro_in_PKS_NRPS	ccctgcgaagcttaggagatcagcatatgcatgttcttcgtctcgtcagc
942	R3_Pro_in_PKS_NRPS	ctctagactaaagcttatcgaattcctagccagaaccgtatagaacaattcg
943	Ck_barA_CT_R	catgtcgttgccgtcaagc
944	Ck_P_PKS_NRPS_F	gaccttgtacgaatgctgcc
945	Ck_P_PKS_NRPS_R	cgcatcaattcgtggtccag
946	Ck_yojil1_NT_F	gcaccttggaaaatctcggc
484	sq_bk_pORFM_F	gccactcatcgcagtctagc
485	sq_pORFM_bk_Rev	tctgcggactggctttctac

(1) Hanahan D. 1983. Studies on transformation of Escherichia coli with plasmids. Journal of Molecular Biology 166:557–580. doi:10.1016/s0022-2836(83)80284-8.

(2) Schultheiss D, Kube M, Schüler D. 2004. Inactivation of the flagellin gene flaA in Magnetospirillum gryphiswaldense results in nonmagnetotactic mutants lacking flagellar filaments. Appl Environ Microbiol 70:3624–3631. doi:10.1128/AEM.70.6.3624-3631.2004.

(3) Raschdorf O, Plitzko JM, Schüler D, Müller FD. 2014. A tailored galK counterselection system for efficient markerless gene deletion and chromosomal tagging in Magnetospirillum gryphiswaldense. Appl Environ Microbiol 80:4323–4330. doi:10.1128/AEM.00588-14.

Chapter 4

-

Expanding the Scope of Detectable Microbial Natural Products by Complementary Analytical Methods and Cultivation Systems

Chantal D. Bader[a]+, Patrick A. Haack[a]+, Fabian Panther[a], Daniel Krug and Rolf Müller[a]*

Previously published in: Journal of Natural Products **2021**

DOI: 10.1021/acs.jnatprod.0c00942

[a] Department Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), German Center for Infection Research (DZIF, Partnersite Hannover-Braunschweig) and Department of Pharmacy, Saarland University Campus E8.1 , 66123 Saarbrücken (Germany)

[+] These authors contributed equally. * Corresponding author. rolf.mueller@helmholtz-hips.de, Tel. +49 681 98896 3000, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), University Campus E8.1, 66123 Saarbrücken, Germany

Author Contributions and Acknowledgments

Patrick A. Haack

This author contributed to the conception of this study, designed and performed experiments, and interpreted results. TOF-MS and FT-ICR-MS measurements were performed by the author for all strains examined in this study. Furthermore, the author performed statistical analysis of the data and defined parameters for data comparison. Moreover, the author significantly contributed to the conception and writing of this manuscript.

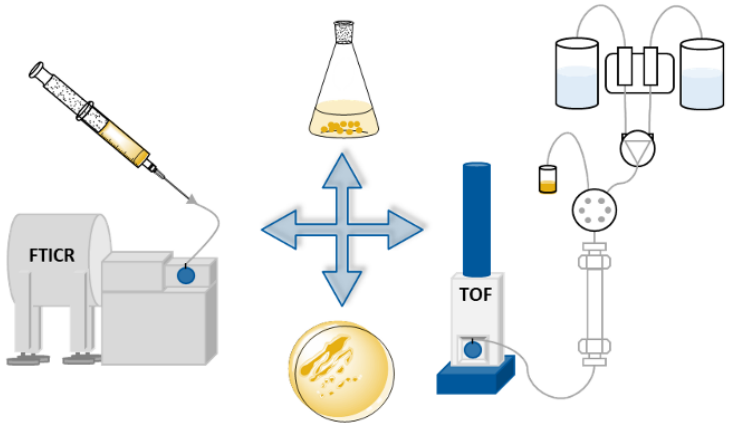
Contributions by Others

Chantal D. Bader contributed to the conception of this study, to the design and performance of experiments, as well as data evaluation and interpretation, laboratory work and writing of this manuscript. Fabian Panter contributed by developing a KNIME workflow for data comparison and the subsequent calculations for the non-targeted analysis, as well as editing of the manuscript. Daniel Krug and Rolf Müller contributed to the conception and supervision of this study, as well as editing and proofreading of the manuscript.

Acknowledgments

We thank Jake F.P. Haeckl and Alexander Kiefer for proofreading of the manuscript.

4.1 Abstract



Recent advances in genome sequencing have unveiled a large discrepancy between the genome-encoded capacity of microorganisms to produce secondary metabolites and the number detected. In this work, a two-platform mass spectrometry analysis for the comprehensive secondary metabolomics characterization of nine myxobacterial strains, focusing on extending the range of detectable secondary metabolites by diversifying analytical methods and cultivation conditions, is presented. Direct infusion measurements of crude extracts on a Fourier-Transform Ion Cyclotron Resonance mass spectrometer are compared to Time-Of-Flight device coupled to liquid chromatography measurements. Both methods are successful in detecting known metabolites, whereas statistical analysis of unknowns highlights their complementarity: Strikingly, 82-99% of molecular features detected with one setup were not detectable with the other. Metabolite profile differences from our set of strains grown in liquid culture versus their swarming colonies on agar plates were evaluated. The detection of up to 96% more molecular features when both liquid and plate cultures were analyzed translates into increased chances to identify new secondary metabolites. Discrimination between primary and secondary metabolism in combination with GNPS molecular networking revealed strain Mx3 as particularly promising for the isolation of novel secondary metabolites among the nine strains investigated in this study.

4.2 Introduction

The search for natural products from bacterial sources is experiencing a resurgence, as the need for novel antimicrobial lead structures is becoming more urgent.[1–3] Compared to well-studied microorganisms such as streptomycetes and bacilli, myxobacteria represent an ubiquitous yet underexplored source of natural products.[4,5] Previously identified and isolated myxobacterial secondary metabolites exhibit a wide range of biological activities as well as structural diversity.[6–10] Many of the compounds isolated early on were identified by activity-guided isolation, while genomics and metabolomics driven approaches have recently

come into focus.[11–13] This is mainly caused by advances in sequencing technology, as well as in silico annotation and prediction tools, that supplied an unprecedented amount of ready to use genomic data.[14–16] These data revealed that the number of biosynthetic gene clusters (BGCs) encoding secondary metabolites far exceeds the number of compounds characterized to date.[17,18] Common approaches to access these hidden secondary metabolites include the activation of underlying biosynthetic pathways. This is approached by changing the cultivation conditions, introducing environmental challenges or by genetically manipulating strains.[19–22] While this relatively slow methodology has proven to be successful for compound discovery, one often over-looked limiting factor for compound detection and identification is the analytical setup applied during microbial extract analysis.

Common analytical approaches include liquid chromatography - mass spectrometry (LC-MS), gas chromatography - MS (GC-MS) and matrix assisted laser desorption ionization - MS (MALDI-MS). LC-coupled time of flight (TOF)-MS is among the most frequently used high-resolution analytical approaches in bacterial secondary metabolomics. The main advantages of TOF-MS are its high dynamic range, high sensitivity, high mass accuracy and easily established automation.[23–25] TOF mass spectrometers are also well suited to be coupled to liquid chromatography due to their fast scan rates. This offers the advantage to link mass signals with LC retention times that serve as an orthogonal identifier, which increases the confidence of annotations and allows association of adducts based on peak congruence.[26–28] Although this is capable of detecting a wide range of microbial secondary metabolites, it has some inherent limitations that need to be considered. The choice of eluents as well as column material types adds a discriminatory effect that limits the detectable chemical space. An inadequate choice of the chromatographic conditions can lead to the dilution of compounds in the LC gradient, rather than the desired concentration due to chromatographic retention. To bypass these factors, mass spectra can also be generated using direct infusion (DI) into the mass spectrometer. In order to use DI-MS as a profiling tool, mass spectrometers need ultra-high resolution to differentiate between near-isobaric substances.[29] A Fourier transform ion cyclotron (FTICR) MS offers ultra-high resolution as well as a low limit of detection and high precision.[30] Nevertheless, the ultra-high mass accuracy of the system requires a slower scan rate than TOF instruments, which makes them less suitable for LC-MS hyphenation. Therefore, direct infusion setups are a plausible choice for FTICR instruments. While DI-FTICR applications remain underrepresented in bacterial secondary metabolomics, there are several applications in fields ranging from biomarker identification to petroleomics already utilizing the strengths of DI-FTICR to rapidly and accurately analyze complex mixtures.[31–35] A drawback of DI-MS applications is the higher prevalence of ion suppression effects.[36] Moreover, the injection of complex mixtures like bacterial extracts generates a dense ion cloud, which may cause reduction in data quality due to so-called space-charge-effects. To minimize these effects

in the FTICR cell, a spectral stitching method as developed by Southam *et al*. can be applied. Herein, the quadrupole incorporated in the device is used as an ion filter for a specific *m/z* value range. The whole *m/z* range can thus be divided into smaller segments, entering the FT-ICR cell one after another.[37]

Myxobacteria require complex cultivation media, which contain molecules with high ionization efficiency including a variety of polar compounds and lipids. This may be one of the reasons why LC-TOF is commonly employed as the analytical method of choice. The LC method separates all components across a polarity gradient, resulting in reduced ion suppression effects.[38] Rich and complex liquid media, however differ greatly from the ecological niche of myxobacteria which are soil living organisms. A practical approach to cultivate myxobacteria under conditions that resemble their natural habitat is to grow myxobacteria on agar plates. This has a big influence on their unique morphology, as they are only able to form their characteristic fruiting bodies or show their characteristic swarming behavior when grown on solid medium.[39] Those changes in phenotype are accompanied by immense changes in the secondary metabolite profiles.[40]

This work aims at extending the detectable space of microbial secondary metabolites, using the characterization of the metabolomes of nine different myxobacterial strains as an example. Furthermore, the influence of different cultivation conditions on the metabolites produced is disclosed and the possible impact of DI-FTICR as a hitherto non-standard analytical setup in bacterial secondary metabolomics research evaluated. The production of known myxobacterial metabolites in two different cultivation systems as well as their detection using DI-FTICR and LC-TOF platforms was investigated. All metabolites detected were analyzed in a non-targeted metabolomics workflow. MS2 spectra in combination with molecular networking were used to further investigate metabolites only detectable with one of the analytical setups. The generated data is ultimately used to distinguish between the primary and the secondary metabolome.

4.3 Results and Discussion

In order to extend the scope of detection for myxobacterial secondary metabolites, we generated a total number of eight different analyses per strain (Figure 1). Each myxobacterial strain was cultivated in liquid medium and on agar plates. Subsequently, the two extracts per strain were analyzed by LC-TOF and DI-FTICR in positive and negative ionization mode. This cascade therefore allows comparison from four perspectives: the ionization polarity, the analytical systems, the cultivation system and the metabolome

of the different strains.

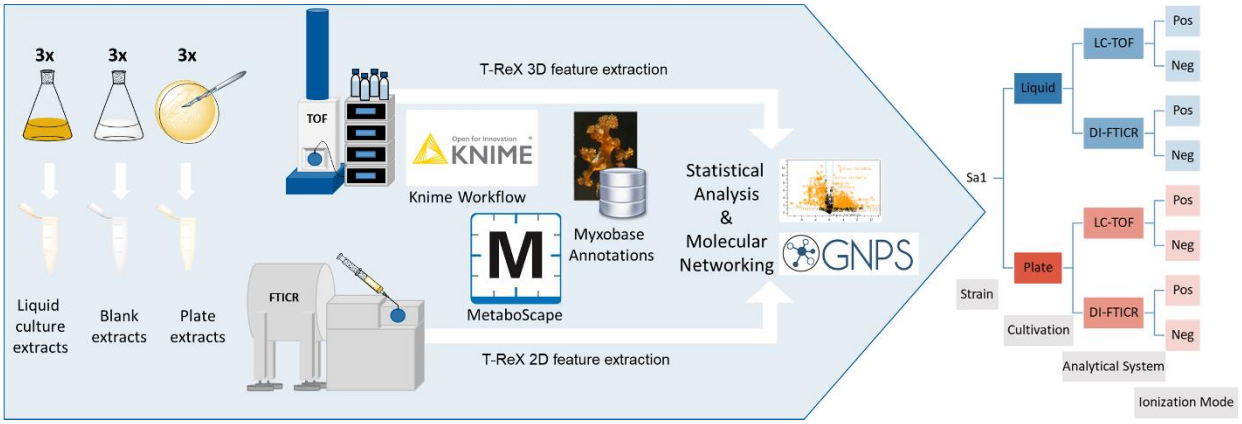


Figure 1: General workflow (left side) and measurements conducted per strain (right side).

Targeted Metabolomics Comparison of the Two Cultivation and Analytical Approaches.

To assess the production of known compounds under the two cultivation conditions as well as their detectability with the two analytical approaches, the recorded data were annotated using analytical details from our in-house database, Myxobase. As most of these metabolites were formerly detected in positive ionization mode, negative ionization data was excluded here. For annotations to be accepted as valid, a mass accuracy window of 5 ppm was used and retention time deviations up to 0.2 min permitted. There is no second dimension, such as retention time, for the DI-FTICR measurements, which increases the risk of false positive annotations. To mitigate this risk, the detection of another member from its secondary metabolite family was taken as proof that the strain really is capable of its production. Furthermore, hits that were annotated in the LC-TOF measurements were considered as valid annotations. Most of the strains studied here are new isolates and have therefore not been investigated in terms of secondary metabolite richness. This explains why there are few annotations for most of the strains, making them interesting subjects for screening approaches aimed at novel natural product discovery. Table 1 shows secondary metabolites detectable in Mx1 extracts.[41-47] This strain is one of the best-characterized myxobacterial strains and produces several natural product families comprising multiple family members, demonstrated by the high number of 14 annotated features.[47] Additional evidence for correct annotations was gained by antiSMASH annotations of the corresponding known biosynthetic gene clusters (BGCs) to confirm the strains ability to produce the detected secondary metabolites. Detailed annotation results for the targeted investigation for all other strains can be found in the SI.

When comparing the metabolic effects of the cultivation conditions on agar plate and in liquid medium within the LC-TOF datasets, more than 75% of the known compounds produced with liquid cultivation were also found in the extracts from plate cultures. Analyzing DI-FTICR data for differences between the

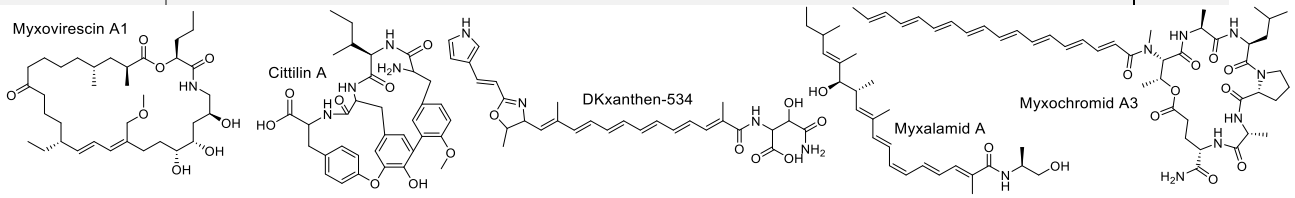
two cultivation systems also shows that 8 out of 9 metabolites can be detected in plate extracts as well as in liquid extracts. Myxovirescin A was only found in the liquid extract and DKxanthene-534 only in the plate extract. Otherwise, all annotated compounds were found in both extracts. Therefore, the general picture emerging from comparison of the two analytical approaches is that they can be regarded as similarly capable to detect known myxobacterial compound classes.

When analyzing the individual members of secondary metabolite families, however, not all derivatives found with the LC-TOF setting were detectable with DI-FTICR. Furthermore, major differences in the ion types that are detected were observed. In the LC-TOF analyses $[M+H]^+$ represents the most abundant ion type. $[M+Na]^+$ adducts were often detected and can be assigned to the other adducts by the chromatographic peak profile. In the DI-FTICR measurements $[M+Na]^+$ and $[M+K]^+$ ions are the most abundant ions, likely because the salts are not separated from the metabolites as it would happen during an LC run. Frequently, $[M+H]^+$ cannot be found in the DI-FTICR analyses at all. Although this does not affect conclusions from the measurements in principle, operators need to be aware of this in order to avoid false annotations.

Generally, most annotations can be found for the combination of liquid cultivation and LC-TOF analysis, showing that the standard cultivation and analytical setup is suitable for a broad range of metabolites. As our in-house database strongly relies on input of natural products isolated after cultivation in liquid medium and detected by LC-TOF, our analyte data collection is biased towards compounds detectable with exactly this setup. In order to make a less biased statement on which setup leads to detection of the largest share of secondary metabolites, including unknowns, anon-targeted metabolomics analysis was conducted.

Table 1: Known myxobacterial secondary metabolites detectable in Mx1 extracts by analysis with LC-TOF and DI-FTICR. The ion types assigned by automated annotations were manually confirmed. Only hits with mass deviation below 5 ppm and retention time deviation (if given) below 0.2 min were considered as valid

	LC-TOF+Liquid		*LC-TOF+Plate*		*DI-FTICR+Liquid*		*DI-FTICR+Plate*		
	Δ ppm	**Ion type**	**Δ ppm**	**Ion type**	**Δ ppm**	**Ion type**	**Δ ppm**	**Ion type**	**ΔRT [min]**
Myxovirescin A	3	$[M+H]^+$	3	$[M+H]^+$	0.7	$[M+K]^+$			0.08
Myxovirescin B	2.7	$[M+H]^+$							0.1
Myxovirescin C	2.6	$[M-MeOH+H]^+$	0.6	$[M-MeOH+H]^+$	2.3	$[M+Na]^+$	3.3	$[M+K]^+$	0.11
Myxovirescin G	2.3	$[M+H]^+$			2.6	$[M+K]^+$	3.1	$[M+Na]^+$	0.04
Cittilin A	2.7	$[M+H]^+$							0.02
DKxanthen-534	4	$[M+H]^+$	0.7	$[M+H]^+$			0	$[M+Na]^+$	0.04
Dkxanthen-544	3.5	$[M+H]^+$							0
Dkxanthen-548	3.8	$[M+H]^+$	0.4	$[M+H]^+$	3.1	$[M+K]^+$	0.5	$[M+H]^+$	0.16
Dkxanthen-574	3.3	$[M+H]^+$	0.1	$[M+H]^+$	4.5	$[M+H]^+$	4.8	$[M+H]^+$	0.18
Dkxanthen-560	2.5	$[M+2H]^{2+}$	1.5	$[M+2H]^{2+}$					0.03
Myxalamid A	3.4	$[M-H_2O+H]^+$	0.6	$[M-H_2O+H]^+$	0.6	[M+Na]+	0.7	[M+Na]+	0.06
Myxalamid B	3.4	$[M-H_2O+H]^+$	0.6	$[M+H]^+$	2.9	$[M+K]^+$	2	[M+Na]+	0.08
Myxalamid C	3.3	$[M-H_2O+H]^+$	0.5	$[M+H]^+$					0.08
Myxochromid A3	1.4	$[M+H]^+$	0.6	$[M+H]^+$	0	[M+Na]+	0.4	[M+Na]+	0.18



Non-targeted Comparison of the Two Cultivation and Analytical Approaches.

For non-targeted comparison of our measurements, the detected signals were grouped into molecular features to collect different ion types belonging to one metabolite as a first data reduction step. Subsequently, all molecular features from one analysis were compared to those measured in the same ionization mode, either to compare the two cultivations or the two analytical approaches. In that way, the percentage of features only present in one of the analyses (later on referred to as unique features) was calculated.

In positive ionization mode between 53 and 93% plate unique molecular features (DI FTICR) were found. In the liquid cultivation experiment, the percentage of unique molecular features is 42-87%. Cultivation on plate therefore delivers more unique features that can be detected in positive mode than cultivation in liquid (see Figure 2). The absolute amount of molecular features detectable in negative mode is generally lower than in positive mode (Figure 2 and 3). In negative ionization mode, 22-62% plate-unique and 21-72% liquid-unique molecular features (DI-FTICR) were found. Hence, there are more liquid unique molecular features than plate unique molecular features detectable in negative mode. Comparing our results of the DI-FTICR measurements to the results of the LC-TOF measurements, the percentage of unique molecular features is higher in the latter. In positive ionization mode, 70-93% plate-unique molecular features and 58-92% liquid-unique molecular features (LC-TOF) were observed. For measurements conducted in negative mode, the percentage of unique features is in a similar range, giving 56-96% plate-unique molecular features and 48-96% liquid-unique molecular features (LC-TOF).

The absolute number of features detectable is highly strain dependent, also accounting for the percentage of features uniquely found in one cultivation system (Figure 2). However, a trend that more positive than negative ionizing substances are detectable in the plate extracts was noted (6 out of 9 strains with DI-FTICR and 7 out of 9 with LC-TOF). Most importantly, our study shows that a minimum of 21% and a maximum of 93% of information is lost when using a single cultivation system, as those features are uniquely produced under one cultivation condition. Our results confirm that cultivation on a plate significantly changes the metabolite profile. Although known secondary metabolites are detectable in both cultivation systems, they only represent a small fraction of the metabolome. They are usually produced in dominant amounts, which also triggered their isolation. The production of novel secondary metabolites may be below the limit of detection in liquid cultures, but can be enhanced by a change in the cultivation system. Plate cultivation should therefore be seen as an important complementary method for isolation of new secondary metabolites rather than an occasional amendment to bacterial cultivation in liquid medium.

When comparing the two analytical setups to each other within the context of one cultivation system, in positive mode we find an average of 88-94% LC-TOF unique molecular features in the plate extracts (see Figure 3). With DI-FTICR we find 88-96% features that we cannot detect with the other method. Those results highlight the comparably small overlap of the two systems, and at the same time reveal the chemical space not captured, when solely using one of the analytical methods. Analysis of the liquid extracts also shows low overlap of the two analytical systems, which is consistent with the results obtained for the plate extracts. In the liquid extracts we detect 85-97% unique LC-TOF molecular features and 82-98% unique DI-FTICR molecular features in positive mode. In negative mode, we find 86-99% unique LC-TOF molecular features and 86-99% unique DI-FTICR molecular features for liquid cultivation. Plate cultivation shows similar results, giving 92-98% unique LC-TOF molecular features and 86-98% unique DI-FTICR molecular features. Our analysis therefore suggests that the analytical system chosen for the detection of metabolites has a considerably larger influence on the observable metabolome than the choice of cultivation system.

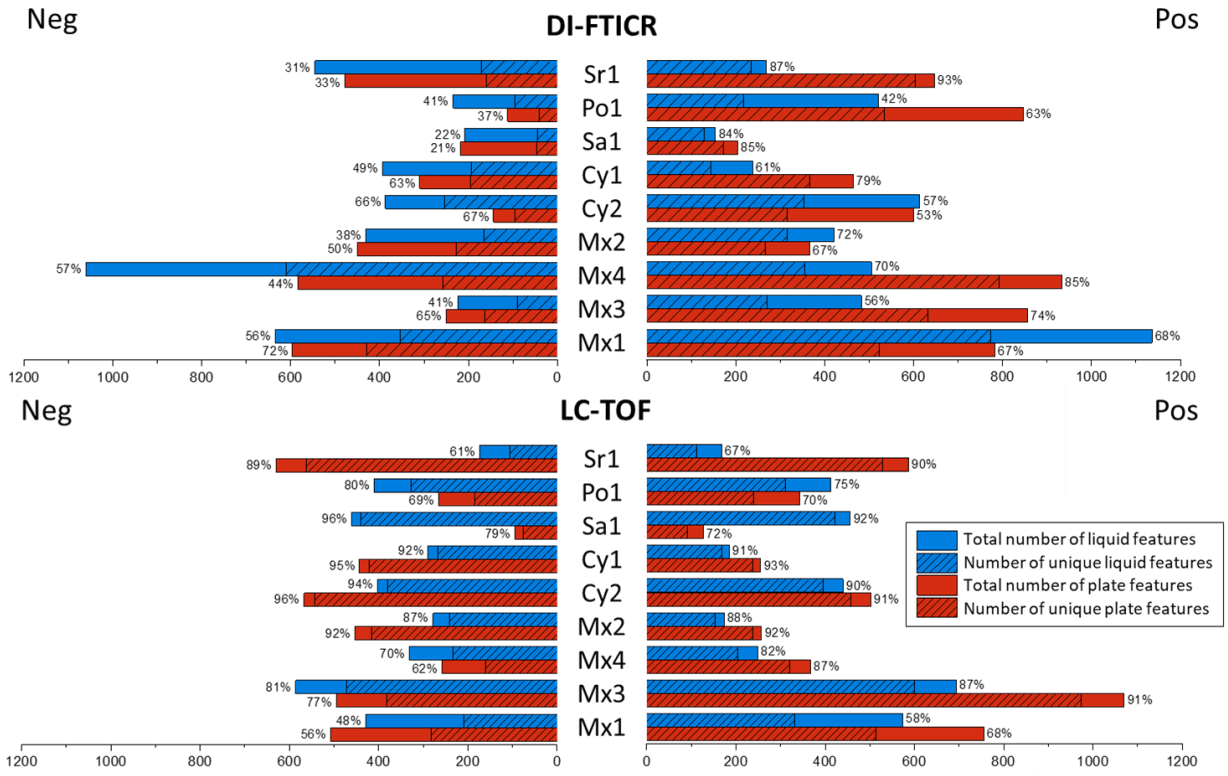


Figure 2: Features detected for the nine different myxobacterial strains comparing plate and liquid cultivation. Blue: Liquid features, Red: Plate features. The percentage of unique features to one of the cultivation systems is marked as dashes and shown at the respective end of the bar. Bars to the left represent molecular features detected in negative mode and bars to the right features detected in positive mode.

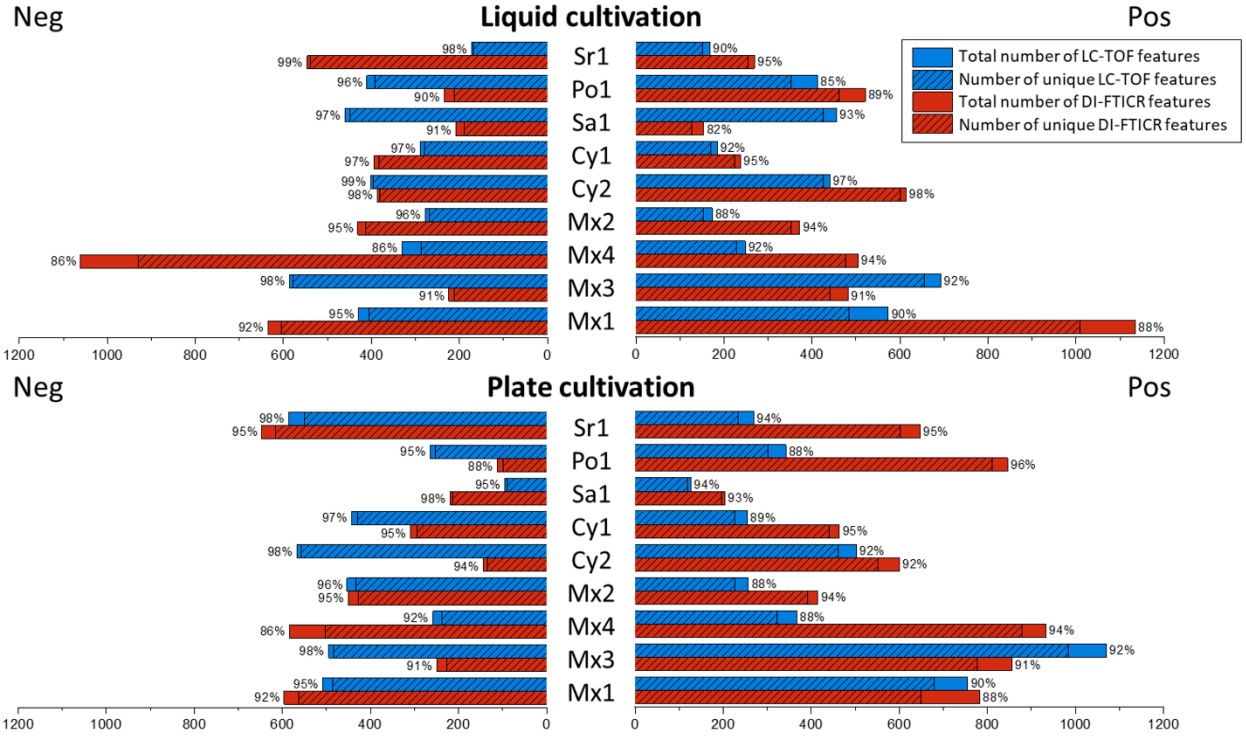


Figure 3: Features detected for the nine different myxobacterial strains comparing the two analytical systems. Blue: LC-TOF features, Red: DI-FTICR features. Percentage of unique features to one of the analytical systems shown at the respective end of a bar and marked as dashes. Bars to the left represent molecular features detected in negative mode and bars to the right features in positive mode.

Nevertheless, as already shown for the targeted metabolomics analysis, the ion type detectable with DI-FTICR may differ from the ion-type detectable with LC-TOF. This also has a large influence on the output of a non-targeted analysis, as the detection of several ion-types is crucial for molecular feature generation. Calculation of exact masses for LC-TOF analysis in MetaboScape is relatively reliable, as most of the times several different ion types including the $[M+H]^+$, are observed. During DI-FTICR measurements, however, often only one ion type for each analyte was detected. The observed ion types in the DI-FTICR measurements corresponded most often to $[M+Na]^+$ or $[M+K]^+$, but without observation of the associated $[M+H]^+$ ion, the software automatically assigned these features as the $[M+H]^+$. This resulted in inaccurate exact mass calculations for the compounds studied (see Scheme S1 in SI).

In order to estimate the influence of this phenomenon on our non-targeted analysis, the minimum number of specific features (number of features that are guaranteed unique to one instrument setup) was calculated for strain Mx1 in addition to our original calculation. For each MS feature in the Mx1 DI-FTICR experiment, we calculated the exact mass for all three most abundant ion types ($[M+Na]^+$, $[M+K]^+$ and $[M+H]^+$) instead of just using the assumed ion type $[M+H]^+$. This resulted in three exact mass results for each individual detected metabolite. To calculate the overlap between the DI-FTICR measurements and

the LC-TOF measurements, the list of exact masses from the LC-TOF experiments was then subtracted from the list of exact masses in the DI-FTICR experiments. After subtraction, 53-74% of the observed features in the DI-FTICR measurements did not have a corresponding LC-TOF exact mass match and therefore are found to be uniquely detectable with one approach. Even though this number is lower than the initial 88-90% features only detectable with one analytical approach for Mx1, this analysis confirms that when considering mismatches during statistical analysis the number of unique features is still high. This calculation likely underestimates the amount of unique features as the same feature could be matched more than one time. We therefore consider our initial calculations as more suitable for a practical assessment of analytical system complementarity. Moving beyond this minimum number of unique features calculation, MS^2 fragmentation and subsequent GNPS-clustering was performed, which provides extended molecular feature comparison.

MS^2-clustering of DI-FTICR Unique Features.

To determine whether molecular features produced by the strain Mx1 uniquely found with DI FTICR really constitute an extension of the detectable chemical space, 18 DI-FTICR unique molecular features found in our prior analysis were fragmented. Their fragmentation pattern was compared to the fragmentation pattern of all molecular features found for Mx1 in the LC-TOF analysis. None of these 18 molecular features found uniquely with DI-FTICR clusters to molecular features found with LC-TOF (see Figure 4).

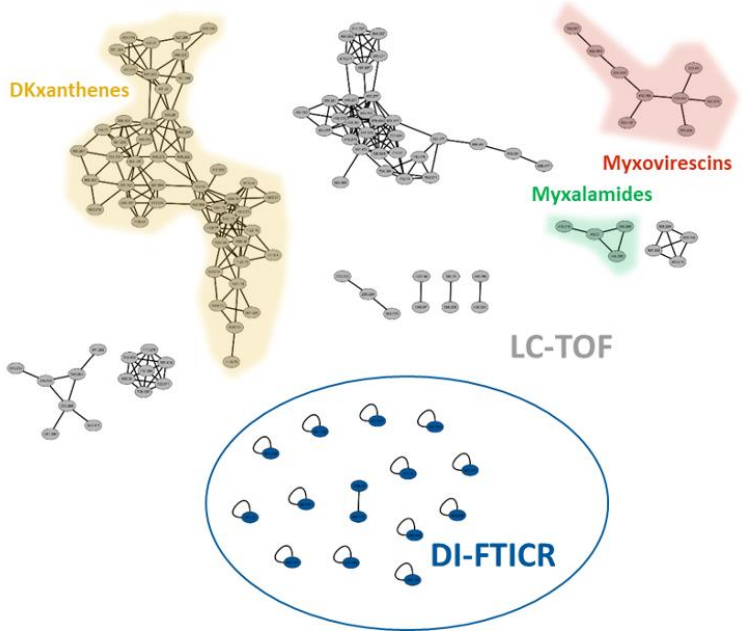


Figure 4: Molecular network of Mx1 features detected with LC TOF (grey nodes) and eight of the most intense features uniquely found with DI-FTICR (blue nodes). Known compound families produced by this strain marked in red (Myxovirescins), yellow (DKxanthenes) and green (Myxalamides).

This result suggests that substances we can detect exclusively with DI-FTICR are not simply derivatives of compound families that can also be observed with LC-TOF. Therefore, DI-FTICR analysis of bacterial extracts unlocks previously undetected metabolites. Final characterization of the chemical structures of all molecular features found with DI-FTICR nevertheless could not be performed with our current setup, as fragmentation is non-automated and only possible for a limited number of molecular features. Although only a limited set of molecular features uniquely detectable by DI-FTICR were fragmented, MS^2 based fragmentation pattern clustering of two of the DI-FTICR unique molecular features was observed. Such substances are of special interest, since secondary metabolite families usually have several family members.

Comparison of the Influence of the Ionization Mode.

In previous studies, Nordström et al. have reported that they achieved a 90% increase in detected ion species when per-forming ESI-MS analysis in both ionization modes instead of just using one mode in a human serum extract analysis.[48] We were therefore interested to assess whether the influence of the ionization mode is comparably high for myxobacterial metabolite detection. Across all nine strains and after subtraction of blank features, 17,295 molecular features were detected in positive and 13,359 molecular features in negative ionization mode. In positive ionization mode, more molecular features were consistently detected (Figure 5). Only for the combination of LC-TOF and liquid cultivation are the number of detected molecular features in the same range. This discrepancy is likely due to the variance in efficiencies between positive and negative ionization for different compounds.[49,50] It seems likely that due to their chemical structure, featuring many hydroxy, carbonyl and amine groups, a majority of myxobacterial metabolites has a higher ionization efficiency in positive ionization mode. When comparing positive and negative ionization between the two analytical setups, a difference in the mean *m/z* values of the measured molecular features was observed. In the DI-FTICR experiments, the mean *m/z* of all molecular features detected with positive ionization is about 20-30 % higher than the mean *m/z* of all molecular features detected with negative ionization. For LC-TOF data, the opposite effect, with the mean *m/z* of all features measured in negative mode being about 10 % higher, was observed.

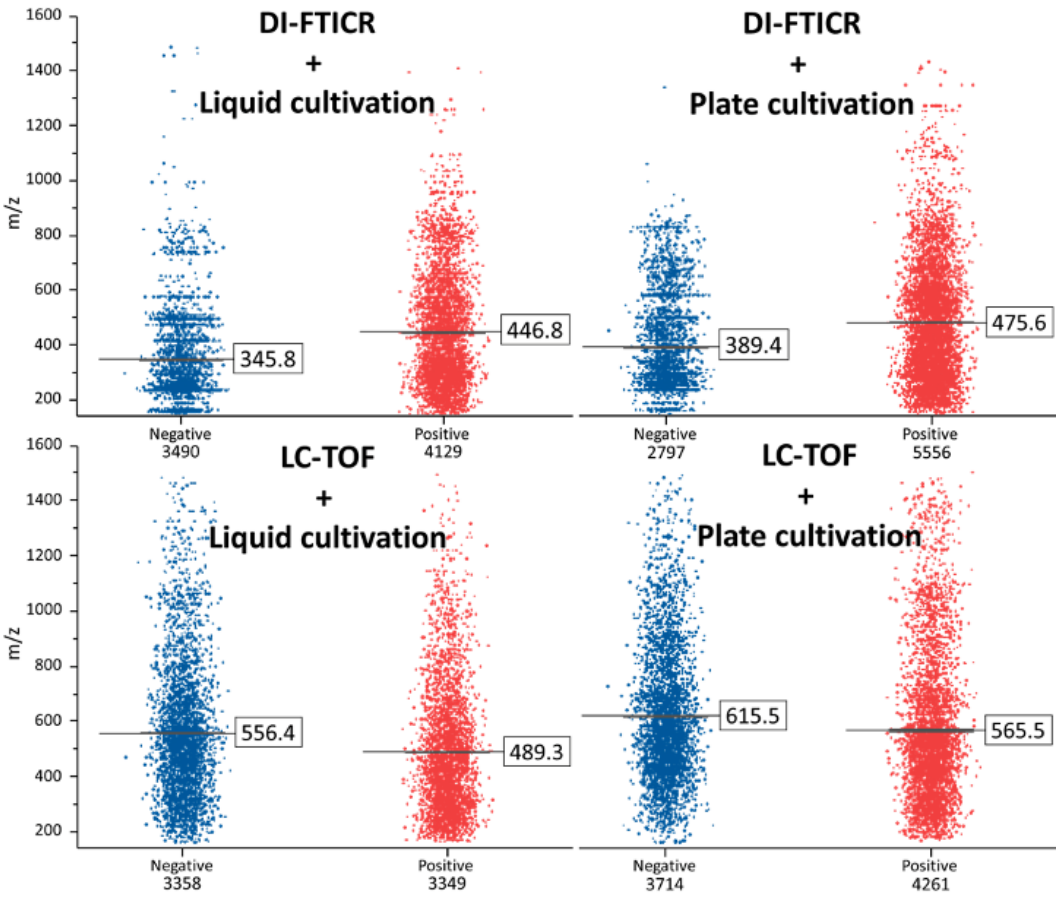


Figure 5: m/z distribution for negative (blue) and positive (red) ionization mode. Molecular features from all strains were grouped according to the analytical setup and cultivation condition. Mean m/z value +/- standard deviation indicated by the horizontal bars. Total number of features detected for each cultivation condition and analytical setup combination, across all strains for positive and negative ionization measurements indicated below the scatter plot.

Characterization of the Unique Secondary Metabolome for each Myxobacterial Strain.

In order to cope with the large number of features in our analyses and to prioritize features for isolation and structure elucidation, we devised a workflow to estimate which molecules stem from primary and which from secondary metabolism. This distinction is of great importance when searching for antimicrobial natural products, as secondary metabolites are often produced as a defense mechanism against competitors. A large number of metabolites detectable in a metabolomics investigation, however stem from primary metabolism. The presented workflow enables us to remove all features of primary metabolism and focus on the unique strain metabolome facilitating the discovery of novel secondary metabolites. As fully automated MS^2 fragmentation is only possible on the LC-TOF system and our previous analysis showed that more myxobacterial metabolites ionize in positive mode, we only performed this analysis for the data generated with positive LC-TOF data. All molecular features detected in one strains' analysis were compared to all the features of the other strains. Each features found in at least one other

strain is removed and the remaining features are submitted to MS^2 analysis and GNPS-based molecular networking (Figure 6).[51] By subtracting all molecular features that appear in several analyses, most of the compounds belonging to the primary metabolism are removed, as these would appear in most of the analyzed strains. We likely also remove frequently occurring secondary metabolites from the analysis. Secondary metabolites appearing in several strains are, in principle, of lesser interest than a unique molecular family, as the likelihood of discovering novel chemistry is decreased in these cases and rediscovery of already described secondary metabolites is more frequent. Our workflow allows for the calculation of a shared metabolome between two or more strains to cope with strains showing high similarity on phylogenetic level and therefore are likely to also share biosynthetic gene clusters.[52] This may then be used to correlate features to BGCs based on their presence in two or more strains.

After removal of common molecular features from each strains analysis the average of 6662 features detected per strain (combined plate and liquid) could be reduced by 94% to an average of 409 features per strain. Remarkably, the amount of unique features per strain is very diverse. The strain with the lowest amount of unique features is Mx2 showing 16 unique molecular features, whereas Mx3 also belonging to the family of Myxococcaceae shows 1242 unique molecular features. After selectively targeting only these molecular features for an MS^2 experiment and GNPS clustering, a molecular network of solely unique molecular features across all strains was obtained.

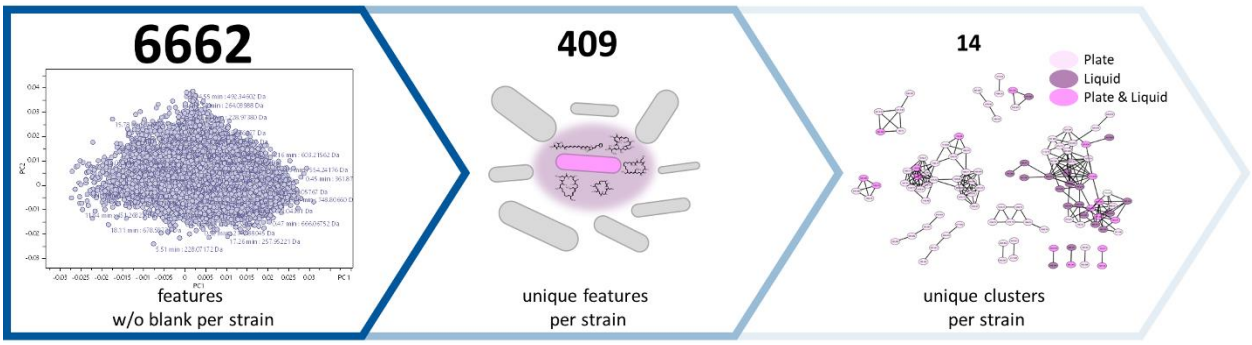


Figure 6: Prioritization workflow to cope with the large number of molecules deriving from primary metabolism. Molecular features depicted are the average number of molecular features detected per strain. Molecular network shown on right side derives from strain Sr1. Nodes depicted in light purple derive from plate cultivation, dark purple nodes from liquid cultivation and pink nodes can be found in both analyses.

Analyzing the molecular network of those features, 2/3 of the MS-clusters comprise features solely produced by one strain. 1/3 of the MS-clusters contain mixed features from different strains, meaning that the different strains produce variants of related natural product families. This finding is reflected at the genetic level, as many strains contain BGCs of the same BGC family with less than 100% similarity.[53] In total, we find 81 MS clusters where all features are produced by a single strain. Interestingly, 35 MS-

clusters thereof solely contain molecular features that are plate-unique molecular features. Thirteen MS-clusters only contain molecular features that are just produced in liquid culture. For more than half of the secondary metabolite families in our analysis the change of cultivation conditions therefore is an all-or-nothing criterion. Under one condition a set of derivatives is produced, whereas under the other condition the BGC remains inactive. Most of the unique secondary metabolites were detectable with both LC-TOF and DI-FTICR (see Figure S15). The LC-TOF unique features furthermore are distributed over the MS-clusters. Only one MS-cluster, consisting of two isomers with the same exact mass was uniquely detectable with LC-TOF. This analysis therefore underlines in addition to the targeted analysis, that most of the secondary metabolite families detectable with LC-TOF are also detectable with DI-FTICR.

4.4 Conclusion

In this work, we showed that both LC-TOF as well as DI-FTICR can be used to detect myxobacterial secondary metabolites in a targeted metabolomics workflow. Certain derivatives of known metabolites were only found when cultivating in liquid medium and subsequently analyzing the corresponding extracts with LC-TOF. Nevertheless, this analysis is severely biased, as secondary metabolites known so far and used for the annotation are mainly found in LC-MS experiments. In the DI-FTICR setup $[M+Na]^+$ and $[M+K]^+$ adducts are often the only observed ion-type, which should be kept in mind for further analysis and when searching for and annotating specific metabolites. The non-targeted analysis of our data showed that the two cultivation methods (agar plates and liquid culture) are complementary and 21-93% of the myxobacterial metabolome cannot be detected when applying a single cultivation method. The amount of information lost depends on the strain investigated as well as on the analytical setup employed. The choice of the analytical system used for myxobacterial metabolomics has a bigger impact on the output of the study than the cultivation system. 82-99% of molecular features are unique for one of the two analytical systems that were compared in this study. MS^2 spectra of a subset of unique DI-FTICR features furthermore confirmed that many features uniquely detectable in DI-FTICR measurements belong to a different chemical space than the features detected in LC-TOF analyses. In general, positive ionization mode allows the detection of more myxobacterial molecular features than negative ionization mode. A difference in the mean *m/z* value detectable depending on the analytical setup used was observed, giving a lower mean *m/z* value detectable with DI-FTICR than with LC-TOF. The immense amount of data created from this metabolomics analysis requires further prioritization strategies to optimize secondary metabolite identification. One suitable approach presented in this work is to exclude molecular features found in more than one strain. This results in over 90% data reduction. Remaining molecular features can

be further grouped with GNPS molecular networking. This whole workflow results – for the set of example strains presented here - in a manageable amount of 14 unique MS-cluster families per strain that can be readily followed up. Comparing the nine myxobacterial strains investigated, Mx3 showed the highest potential for metabolomics-guided isolation and structure elucidation because of its high amount of unique MS-clusters in addition to the few known natural products it produces.

In conclusion, this study underlines that besides the choice of cultivation, the analytical system influences the results of any microbial secondary metabolome analysis. In targeted workflows, one well-chosen setup may be considered sufficient to detect all metabolites of interest. In a non-targeted workflow however, using at least two analytical setups as well as different cultivation conditions leads to an increase of the metabolome coverage. The additional molecular features uncovered by complementary growth conditions and analytical setups directly translate into increased chances for the discovery of new natural products. Thus, we argue that the simplification of screening workflows should not come at the cost of decreased metabolomics characterization, as this not only paints an incomplete picture of the bacterial secondary metabolome complexity but also results in missed compelling target molecules.

4.5 Experimental Section

General experimental procedures

All materials, including the bacterial strains used for this analysis (as diverse set of representatives of different myxobacterial families), bacterial cultivation media, fermentation protocols, related sample preparation and experimental procedures are described in detail in the supporting information.

Cultivation and Sample Preparation

Myxobacterial strains were cultivated in liquid culture as well as on agar plates using CYH medium. All cultivations were done in triplicate. Cells from the agar plate cultures were scraped off, and liquid cultures were centrifuged to separate cells and supernatant, respectively. Prior to extraction with methanol, the cell pellets were lyophilized. Blank samples were generated by lyophilizing and extracting the cultivation medium treated in the same way as the bacterial fermentation cultures.

DI-FTICR and LC-TOF Analysis

All MS measurements were performed in duplicates. DI-FTICR measurements were performed using positive as well as negative ionization on the Bruker SolariX XR 7T. Samples were diluted 1:200 with methanol prior to measurements and the mass spectrometer was externally calibrated to a mass accuracy below 1 ppm. The samples were injected with a pre-installed syringe pump at 2.0 µl min^{-1} flowrate. 110 scans were performed within 4 minutes accumulating for 150 ms. The total *m/z* range from 100-1500 was

divided in eight segments that were later combined to a full spectrum. Collision RF Amplitude was optimized for each segment. The data size was set to 4 M with a 2 s transient. LC-TOF measurements were performed on a Dionex Ultimate 3000 SL system coupled to a Bruker maXis 4G UHRTOF. The mobile phase consisted of (A) *dd*H_2O with 0.1% formic acid and (B) acetonitrile with 0.1% formic acid. For separation a linear gradient from 5-95% B in A on a Waters Acquity BEH C_{18} column (100 x 2.1 mm, 1.7 µm dp) was used. The flow rate was set to 0.6 mL min^{-1} and the column thermostated at 45 °C. Extracts were diluted 1:20 prior to measurements. The LC flow was split to 75 µL min^{-1} before entering the mass spectrometer, which was externally calibrated to a mass accuracy below 1 ppm. Mass spectra were acquired in centroid mode ranging from 150-2500 *m/z* at a 2 Hz scan rate. Further details of the analytical setups are described in the SI.

Statistical Analysis and Annotations

Statistical interpretation for targeted and non-targeted metabolomics analysis was carried out with MetaboScape 4.0 (Bruker). The minimal intensity threshold for molecular feature detection was set to 1.5 x 10^4 for DI-LC-TOF data and 6 x 10^6 for DI-FTICR data. The maximum charge was set to three (positive and negative) and the minimal group size for creating batch features to five. Known metabolites were annotated using an in-house database containing myxobacterial secondary metabolites (Myxobase).

MS^2 Analysis and Molecular Networking

Scheduled precursor list -guided MS^2 spectra were generated on the LC-TOF system using the same chromatographic conditions as for MS measure-ments. For enhanced spectra coverage, a specified precursor list was used, containing only features originating from the bacteria.[54] MS^2 spectra generated on the FT-ICR system were recorded manually and parameters were optimized for each precursor ion. A complete list of CID energies can be found in the SI. MS^2 data was uploaded to the global natural products social molecular networking (GNPS) server at the University of California, San Diego.[51] The clustered dataset was visualized using Cytoscape 3.7.2.

4.6 References

[1] Baker, D. D.; Chu, M.; Oza, U.; Rajgarhia, V. *Nat. Prod. Rep.* **2007**, *24*, 1225–1244, DOI: 10.1039/b602241n.

[2] Bernardini, S.; Tiezzi, A.; Laghezza Masci, V.; Ovidi, E. *Nat. Prod. Res.* **2018**, *32*, 1926–1950, DOI: 10.1080/14786419.2017.1356838.

[3] U.S. Centers for Disease Control and Prevention. *Antibiotic Resistance Threats in the United States, 2019*.

[4] Bader, C. D.; Panter, F.; Müller, R. *Biotechnol. Adv.* **2020**, *39*, 107480, DOI: 10.1016/j.biotechadv.2019.107480.

[5] Wenzel, S. C.; Müller, R. Myxobacteria - unique microbial secondary metabolite factories. In *Comprehensive Natural Products Chemistry II, Vol 2: Structural Diversity II - Secondary Metabolite Sources, Evolution and Selected Molecular Structures*; Moore, B.S., Ed.; Elsevier: Oxford, 2010; pp 189–222.

[6] Schäberle, T. F.; Lohr, F.; Schmitz, A.; König, G. M. *Nat. Prod. Rep.* **2014**, *31*, 953–972, DOI: 10.1039/c4np00011k.

[7] Herrmann, J.; Fayad, A. A.; Müller, R. *Nat. Prod. Rep.* **2017**, *34*, 135–160, DOI: 10.1039/C6NP00106H.

[8] Dehhaghi, M.; Mohammadipanah, F.; Guillemin, G. J. *Neurotoxicology* **2018**, *66*, 195–203, DOI: 10.1016/j.neuro.2018.02.017.

[9] Steinmetz, H.; Irschik, H.; Kunze, B.; Reichenbach, H.; Höfle, G.; Jansen, R. *Chem. Eur. J.* **2007**, *13*, 5822–5832, DOI: 10.1002/chem.200700269.

[10] Steinmetz, H.; Mohr, K. I.; Zander, W.; Jansen, R.; Gerth, K.; Müller, R. *J. Nat. Prod.* **2012**, *75*, 1803–1805, DOI: 10.1021/np300288b.

[11] Corre, C.; Challis, G. L. Exploiting Genomics for New Natural Product Discovery in Prokaryotes. In *Comprehensive Natural Products Chemistry II, Vol 2: Structural Diversity II - Secondary Metabolite Sources, Evolution and Selected Molecular Structures*; Moore, B.S., Ed.; Elsevier: Oxford, 2010; pp 429–453.

[12] Adamek, M.; Spohn, M.; Stegmann, E.; Ziemert, N. Mining Bacterial Genomes for Secondary Metabolite Gene Clusters. In *Antibiotics: Methods and protocols / edited by Peter Sass*; Sass, P., Ed.; Springer protocols 1520; Humana Press: New York, 2017; pp 23–47.

[13] Bachmann, B. O.; van Lanen, S. G.; Baltz, R. H. *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 175–184, DOI: 10.1007/s10295-013-1389-9.

[14] Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. *Nucleic Acids Res.* **2019**, W81-W87, DOI: 10.1093/nar/gkz310.

[15] Fraser, C. M.; Eisen, J. A.; Salzberg, S. L. *Nature* **2000**, *406*, 799–803.

[16] Fraser, C. M.; Eisen, J. A.; Nelson, K. E.; Paulsen, I. T.; Salzberg, S. L. *J. Bacteriol.* **2002**, *184*, 6403–6405, DOI: 10.1128/JB.184.23.6403–6405.2002.

[17] Wenzel, S. C.; Müller, R. *Curr. Opin. Drug Discov. Devel.* **2009**, *12*, 220–230.

[18] Jensen, P. R.; Chavarria, K. L.; Fenical, W.; Moore, B. S.; Ziemert, N. *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 203–209, DOI: 10.1007/s10295-013-1353-8.

[19] Adnani, N.; Chevrette, M.; Adibhatla, S. N.; Zhang, F.; Yu, Q.; Braun, D. R.; Nelson, J.; Simpkins, S. W.; McDonald, B. R.; Myers, C. L. *et al. ACS Chem. Biol.* **2017**, *12*, 3093–3102, DOI: 10.1021/acschembio.7b00688.

[20] Piddock, L. J. V. *J Antimicrob Chemother* **2015**, *70*, 2679–2680, DOI: 10.1093/jac/dkv175.

[21] Li, P. F.; Li, S. G.; Li, Z. F.; Zhao, L.; Wang, T.; Pan, H. W.; Liu, H.; Wu, Z. H.; Li, Y. Z. *FEMS Microbiol. Ecol.* **2013**, *85*, 358–368, DOI: 10.1111/1574-6941.12125.

[22] Panter, F.; Krug, D.; Baumann, S.; Müller, R. *Chem. Sci.* **2018**, *9*, 4898–4908, DOI: 10.1039/C8SC01325J.

[23] Saito-Shida, S.; Hamasaka, T.; Nemoto, S.; Akiyama, H. *Food chemistry* **2018**, *256*, 140–148, DOI: 10.1016/j.foodchem.2018.02.123.

[24] Fels, H.; Dame, T.; Sachs, H.; Musshoff, F. *Drug testing and analysis* **2017**, *9*, 824–830, DOI: 10.1002/dta.2039.

[25] Boesl, U. *Mass spectrometry reviews* **2017**, *36*, 86–109, DOI: 10.1002/mas.21520.

[26] Kim, J.; Choi, J. N.; Kim, P.; Sok, D. E.; Nam, S. W.; Lee, C. H. *J. Microbiol. Biotechnol.* **2009**, *19*, 51–54, DOI: 10.4014/jmb.0711.

[27] Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Xueting, L.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; Felicio, R. de; Fenner, A. *et al. J. Nat. Prod.* **2013**, *76*, 1686–1699, DOI: 10.1021/np400413s.

[28] Hoffmann, T.; Krug, D.; Hüttel, S.; Müller, R. *Anal. Chem.* **2014**, *86*, 10780–10788, DOI: 10.1021/ac502805w.

[29] Sleighter, R. L.; Hatcher, P. G. *J. Mass Spectrom.* **2007**, *42*, 559–574, DOI: 10.1002/jms.1221.

[30] Brown, S. C.; Kruppa, G.; Dasseux, J.-L. *Mass Spectrom. Rev.* **2005**, *24*, 223–231, DOI: 10.1002/mas.20011.

[31] Sommella, E.; Conte, G. M.; Salviati, E.; Pepe, G.; Bertamino, A.; Ostacolo, C.; Sansone, F.; Prete, F. D.; Aquino, R. P.; Campiglia, P. *Molecules (Basel, Switzerland)* **2018**, *23,* DOI: 10.3390/molecules23051152.

[32] Wu, Z.; Rodgers, R. P.; Marshall, A. G. *J. Agric. Food Chem.* **2004**, *52*, 5322–5328, DOI: 10.1021/jf049596q.

[33] Park, K. H.; Kim, M. S.; Baek, S. J.; Bae, I. H.; Seo, S.-W.; Kim, J.; Shin, Y. K.; Lee, Y.-M.; Kim, H. S. *Plant methods* **2013**, *9*, 15, DOI: 10.1186/1746-4811-9-15.

[34] Witt, M.; Barsch, A.; Wolff, J.; Krug, D.; Hoffmann, T.; Müller, R. *Ultrafast Statistical Profiling of Bacterial Metabolite Extracts*, 2013.

[35] Fernandez-Lima, F. A.; Becker, C.; McKenna, A. M.; Rodgers, R. P.; Marshall, A. G.; Russell, D. H. *Anal. Chem.* **2009**, *81*, 9941–9947, DOI: 10.1021/ac901594f.

[36] Annesley, T. M. *Clin. Chem.* **2003**, *49*, 1041–1044, DOI: 10.1373/49.7.1041.

[37] Southam, A. D.; Payne, T. G.; Cooper, H. J.; Arvanitis, T. N.; Viant, M. R. *Anal. Chem.* **2007**, *79*, 4595–4602, DOI: 10.1021/ac062446p.

[38] Furey, A.; Moriarty, M.; Bane, V.; Kinsella, B.; Lehane, M. *Talanta* **2013**, *115*, 104–122, DOI: 10.1016/j.talanta.2013.03.048.

[39] Muñoz-Dorado, J.; Marcos-Torres, F. J.; García-Bravo, E.; Moraleda-Muñoz, A.; Pérez, J. *Front. Microbiol.* **2016**, *7*, 781, DOI: 10.3389/fmicb.2016.00781.

[40] Hoffmann, M.; Auerbach, D.; Panter, F.; Hoffmann, T.; Dorrestein, P. C.; Müller, R. *ACS Chem. Biol.* **2018**, *13*, 273–280, DOI: 10.1021/acschembio.7b00816.

[41] Gerth, K.; Irschik, H.; Reichenbach, H.; Trowitzsch, W. *J. Antibiot.* **1982**, *35*, 1454–1459.

[42] Hug, J. J.; Dastbaz, J.; Adam, S.; Revermann, O.; Koehnke, J.; Krug, D.; Müller, R. *ACS Chem. Biol.* **2020**, *15*, 2221–2231, DOI: 10.1021/acschembio.0c00430.

[43] Meiser, P.; Bode, H. B.; Müller, R. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19128–19133, DOI: 10.1073/pnas.0606039103.

[44] Gerth, K.; Jansen, R.; Reifenstahl, G.; Höfle, G.; Irschik, H.; Kunze, B.; Reichenbach, H.; Thierbach, G. *J Antibiot* **1983**, *36*, 1150–1156, DOI: 10.7164/antibiotics.36.1150.

[45] Trowitzsch-Kienast, W.; Gerth, K.; Wray, V.; Reichenbach, H.; Höfle, G. *Liebigs Ann. Chem.* **1993**, *1993*, 1233–1237, DOI: 10.1002/jlac.1993199301200.

[46] Krug, D.; Garcia, R.; Müller, R. *Biospektrum* **2020**, *26*, 32–36, DOI: 10.1007/s12268-020-1334-1.

[47] Krug, D.; Zurek, G.; Revermann, O.; Vos, M.; Velicer, G. J.; Müller, R. *Appl. Environ. Microbiol.* **2008**, *74*, 3058–3068, DOI: 10.1128/AEM.02863-07.

[48] Nordström, A.; Want, E.; Northen, T.; Lehtiö, J.; Siuzdak, G. *Anal. Chem.* **2008**, *80*, 421–429, DOI: 10.1021/ac701982e.

[49] Oss, M.; Kruve, A.; Herodes, K.; Leito, I. *Anal. Chem.* **2010**, *82*, 2865–2872, DOI: 10.1021/ac902856t.

[50] Huffman, B. A.; Poltash, M. L.; Hughey, C. A. *Anal. Chem.* **2012**, *84*, 9942–9950, DOI: 10.1021/ac302397b.

[51] Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T. *et al. Nat. Biotechnol.* **2016**, *34*, 828–837, DOI: 10.1038/nbt.3597.

[52] Hoffmann, T.; Krug, D.; Bozkurt, N.; Duddela, S.; Jansen, R.; Garcia, R.; Gerth, K.; Steinmetz, H.; Müller, R. *Nat. Commun.* **2018**, *9*, 803, DOI: 10.1038/s41467-018-03184-1.

[53] Irschik, H.; Reichenbach, H.; Höfle, G.; Jansen, R. *J Antibiot.* **2007**, *60*, 733–738, DOI: 10.1038/ja.2007.95.

[54] Panter, F.; Krug, D.; Müller, R. *ACS Chem. Biol.* **2019**, *14*, 88–98, DOI: 10.1021/acschembio.8b00948.

Expanding the scope of detectable microbial natural products by complementary analytical methods and cultivation systems

Supporting Information

Chantal D. Bader[a]+, Patrick A. Haack[a]+, Fabian Panther[a], Daniel Krug and Rolf Müller[a]*

Previously published in: Journal of Natural Products **2021**

DOI: 10.1021/acs.jnatprod.0c00942

[a] Department Microbial Natural Products, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), German Center for Infection Research (DZIF, Partnersite Hannover-Braunschweig) and Department of Pharmacy, Saarland University Campus E8.1 , 66123 Saarbrücken (Germany)

+ These authors contributed equally. * Corresponding author.

.

Table of Contents

S 4.1 Extract generation, LC-MS data acquisition and GNPS-based secondary metabolite clustering

S 4.1.1 Myxobacterial strains used for the analysis

The following strains were chosen as a diverse set of representatives of different myxobacterial families. In-house isolated non-published strains that are not fully classified yet are marked with *.

Table S 1: Myxobacterial strains used in this study

Strain	ID in this study	Family
***Myxococcus xanthus* DK1622**	Mx1	*Myxococcaceae*
MCy11770*	Cy1	*Myxococcaceae*
***Simulacricoccus ruber* MCy10636**	Mx2	*Myxococcaceae*
***Pyxidicoccus* sp. MCy10649**	Cy2	*Myxococcaceae*
MSr10575*	Sa1	*Sandaracineae*
***Sorangium cellulosum* SoCe 26**	Po1	*Polyangiaceae*
***Myxococcus fulvus* MCy9270**	Mx3	*Myxococcaceae*
MSr11367*	Sr1	Presumably member of novel family
MCy9487*	Mx4	*Myxococcaceae*

S 4.1.2 Myxobacterial culture medium

All cultivations (liquid and agar plates) were performed using CYH medium (see Table S2). For preparation of agar plates 15 g/L agar (BD) was added to the liquid medium before autoclaving. All media were prepared using deionized water and autoclaved at 121 °C, 2 bar for 20 min. Sterile-filtered Fe-EDTA and sterile-filtered vitamin B_{12} were added after autoclaving of the medium.

Table S 2: Medium recipe of CYH medium.

Ingredient	Supplier	Amount [g/L]	pH adjusted to
Bacto Casitone	BD	1.5	7.3 (NaOH)
Yeast extract	BD	1.5	
Glucose	Roth	1.0	
Soybean meal	Heusel	1.0	
Soluble Starch	Roth	4.0	
CaCl2 x 2 H2O	VWR Chemicals	1.0	
MgSO4 x 7 H2O	Grüssing	0.5	
HEPES	Roth	11.9 (50 mM)	

S 4.1.3 Myxobacterial fermentation conditions

Liquid cultures were grown in 300 mL shake flasks containing 100 mL of CYH medium and were inoculated with 2 mL of pre culture. All strains were grown at 30 °C for 14 days on an Orbitron shaker at 160 rpm. Plate cultures were grown on 25 mL agar plates at 30 °C for 14 days. Agar plates were inoculated with 100 µL of concentrated pre-culture and spread using a Drigalski spatula.

S 4.1.4 Metabolite extraction procedure for analytical scale extractions

After centrifugation, the cells were extracted by adding 50 mL MeOH. They were then sonicated for 10 min prior to shaking for 50 min at 180 rpm. The cells were separated from the extract by filtration. Afterwards, the solvent was evaporated under reduced pressure and the extracts re-dissolved in 1 mL MeOH.

S 4.1.5 Conditions of the analytical measurements

DI-FTICR

Main parameters used for the measurements are described in the material and methods part of the manuscript. Additional to that, the capillary voltage was set to -4500 V for positive ionization and -+4500 V for negative ionization. Dry gas was set to a flow rate of 4 L/min at 220 °C.

Calibration and performance evaluation of DI-FTICR measurements

Calibration was performed by injecting the LC/MS Calibration standard for ESI-TOF (Agilent). The instrument was calibrated after switching from standby to operate mode each day prior to measurements. Calibration files for the tune mix measurements were saved and used to evaluate the performance of the

instrument during the period of measurements (see Table S3, Table S4, Figure S21 and Figure S2). For measurement of the calibration files, 16 scans were performed accumulating for 200 ms with a total m/z range from 150-1500. The data size was set to 1 M with a 0.7 s transient.

The average mass accuracy after calibration was calculated to be 0.25 ppm for the negative ionization measurements and 0.38 ppm for the positive ionization measurements.

The resolving power across all measurements for the strain Mx1 was 2,758,378 – 35,217 across the range of 150 – 1500 m/z. In comparison to the ginseng measurements of Park *et al.*[1] we reach a higher maximum resolving power, which probably is caused by the lower m/z minimum chosen. The lower resolving power in the higher m/z regions is caused by the lower magnetic field strength of 7T sustained by our FT-ICR system.

Table S3: Results of the QC for DI-FTICR measurements in negative ionization mode.

Day	Mass (theoretical)	Mass (measured)	Δ Mass [Da]	Mass accuracy [ppm]	Resolving power	Intensity
1	301.9981	301.9981	0.0000	0.0000	156267	3715893504
2	301.9981	301.9981	0.0000	0.0000	139390	3521025536
3	301.9981	301.9981	0.0000	0.0000	157730	3221839104
1	601.9790	601.9791	-0.0001	0.1329	78260	9681099776
2	601.9790	601.9791	-0.0001	0.1827	74396	17996009472
3	601.9790	601.9791	-0.0001	0.1163	78887	11770544128
1	1033.9881	1033.9874	0.0007	0.7060	44989	2448982272
2	1033.9881	1033.9872	0.0009	0.9091	45656	3756234240
3	1033.9881	1033.9875	0.0006	0.5996	44772	2798772992
1	1333.9697	1333.9698	-0.0001	0.0900	35532	2314893312
2	1333.9697	1333.9700	-0.0004	0.2699	36169	3455701760
3	1333.9697	1333.9697	0.0000	0.0000	35528	2485162496

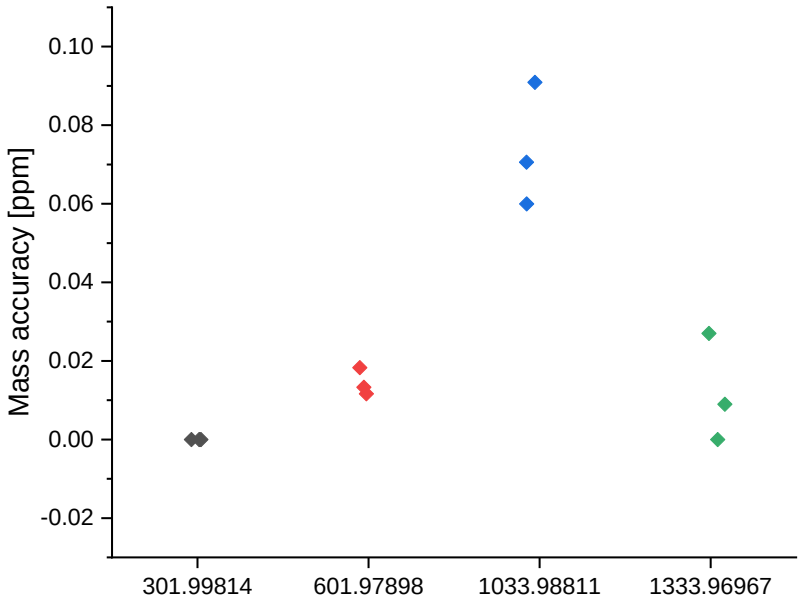


Figure S1: Mass accuracy for the four components of the tune mix in the m/z range from 200-1500 on the three days of measurements in negative ionization mode.

Table S4: Results of the QC for DI-FTICR measurements in positive ionization mode.

Day	Mass (theoretical)	Mass (measured)	Δ Mass [Da]	Mass accuracy [ppm]	Resolving power	Intensity
1	322.0481	322.0482	-0.0001	0.1553	397445	930811072
2	322.0481	322.0481	0.0000	0.0000	399417	1289422848
3	322.0481	322.0481	0.0000	0.0000	399615	1139737344
1	622.0290	622.0287	0.0003	0.4662	206119	9647129600
2	622.0290	622.0290	-0.0001	0.1286	209357	7608689644
3	622.0290	622.0291	-0.0001	0.1447	209939	8804069352
1	922.0098	922.0091	0.0007	0.7267	141960	6287506432
2	922.0098	922.0094	0.0004	0.4013	142304	5230539776
3	922.0098	922.0094	0.0004	0.4772	140587	6044235776
1	1221.9906	1221.9923	-0.0017	1.3584	107728	2381585920
2	1221.9906	1221.9910	-0.0004	0.3028	107084	2147749632
3	1221.9906	1221.9911	-0.0005	0.3683	108466	2680427264

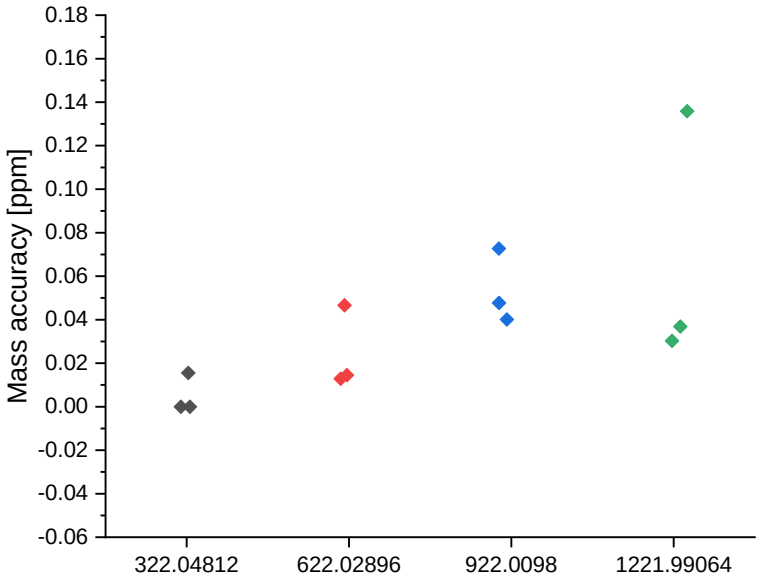


Figure S2: Mass accuracy for the four components of the tune mix in the m/z range from 200-1500 on the three days of measurements in positive ionization mode.

LC-TOF

Main parameters used for the measurements are described in the material and methods part of the manuscript. Additionally, the sample injection volume was 1 µL. Capillary voltage was set to +4000 V in positive and +2500 V in negative ionization mode. Dry gas was set to a flow rate of 5 L/min at 200 °C. For MS/MS measurements SPL tolerance parameters for precursor ion selection are set to 0.2 min and 0.05 m/z. CID energy is ramped from 35 eV for 500 m/z to 45 eV for 1000 m/z and 60 eV for 2000 m/z.

Due to the high dynamic range of TOF spectrometers the *m/z* scan range was set from 150-2500 which is also the value used in our standard screening method. As we wanted to compare our standard method and not a method optimized for this comparison, we kept the high *m/z* cut-off at 2500. To create accurate mass spectra in the low *m/z* range, our TOF instruments would have to be retuned, which would results in a lower mass accuracy in the higher *m/z* range. We therefore kept the standard cut-off at *m/z* 150 for the TOF measurements. The FT-ICR, however, has a more limited dynamic range, wherefore we manually evaluated the highest and lowest *m/z* value observable in our extracts for the DI-FTICR measurements in one full scan. No ions with an *m/z* higher than 1500 or lower than *m/z* 100 was observed, wherefore we chose these values as cut-off. The overlapping range of 150-1500 *m/z* was set as limits for the feature finding algorithm, which was applied uniformly to all measurements. All calculations shown in the manuscript were only performed for features in this scan range, which is shared between the two setups. Any features lower than 150 and higher than 1500 would therefore not appear in any comparison and would not influence the results.

Calibration and performance evaluation of LC-TOF measurements

The instrument was calibrated after switching from standby to operate mode each day prior to measurements using sodium formate. All spectra were calibrated on the sodium formate cluster, which was also injected in the beginning of each chromatographic run. Additionally, lock mass calibration was applied. An in-house standard mixture was measured at regular intervals between each series of measurements. This was used to evaluate the performance of the instrument during the period of measurements. For the standard mixture, the same conditions as for the sample measurements were used, except for the length of gradient, which was shortened from 18 minutes to 9 minutes. All LC measurements were performed within a time period of 72 h. The average mass accuracy after calibration was calculated to be 2.04 ppm for the negative ionization measurements and 1.48 ppm for the positive ionization measurements.

Table S5: Results of the QC for LC-TOF measurements in positive ionization mode.

Testmix #	Mass (theoretical)	Mass (measured)	Δ Mass [Da]	Mass accuracy [ppm]	AUC	RT
1	279.0916	279.0912	-0.0004	1.3006	2433542	2.81
2	279.0916	279.0911	-0.0005	1.3799	2611725	2.82
3	279.0916	279.0910	-0.0006	1.2666	2513730	2.82
1	271.0323	271.0320	-0.0004	1.2633	1206557	2.82
2	271.0323	271.0318	-0.0006	1.5960	1211159	2.83
3	271.0323	271.0318	-0.0006	0.9677	1217103	2.83
1	285.0213	285.0209	-0.0004	1.6590	1058387	3.22
2	285.0213	285.0210	-0.0003	2.0809	1057026	3.23
3	285.0213	285.0209	-0.0004	1.0210	1067704	3.22
1	311.0814	311.0810	-0.0004	1.0383	2293771	4.05
2	311.0814	311.0811	-0.0003	1.6320	2493051	4.05
3	311.0814	311.0810	-0.0004	1.1269	2373666	4.05
1	278.1909	278.1904	-0.0004	2.0173	4870030	4.98
2	278.1909	278.1904	-0.0005	2.1178	4947451	4.99
3	278.1909	278.1903	-0.0006	1.4069	5016843	5.00
1	314.1392	314.1389	-0.0003	1.3598	1868371	6.22
2	314.1392	314.1389	-0.0004	2.0993	1774190	6.22
3	314.1392	314.1388	-0.0004	1.3497	1769561	6.22

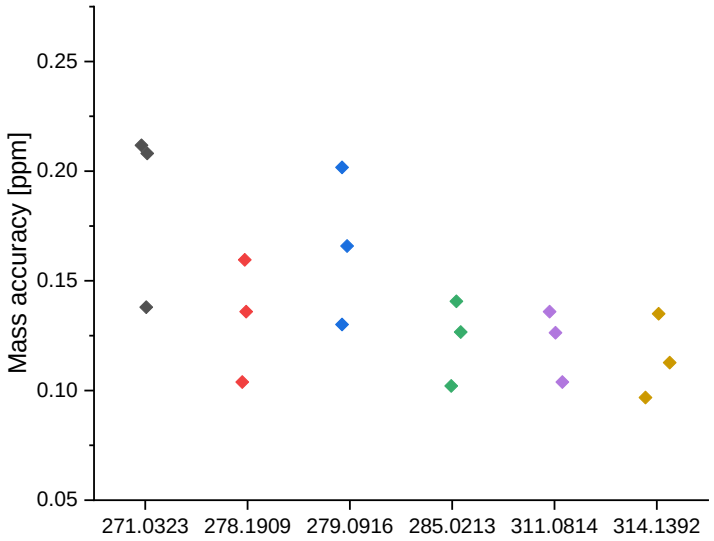


Figure S3: Mass accuracy for the six components of the in-house in positive ionization mode.

Table S6: Results of the QC for LC-TOF measurements in negative ionization mode.

Testmix #	Mass (theoretical)	Mass (measured)	Δ Mass [Da]	Mass accuracy [ppm]	AUC	RT
1	269.0167	269.0173	0.0006	2.0631	1402327	2.85
2	269.0167	269.0174	0.0007	2.0459	1667004	2.84
3	269.0167	269.0173	0.0006	1.2845	1735250	2.84
1	283.0057	283.0062	0.0006	2.6207	1387172	3.24
2	283.0057	283.0063	0.0006	2.1519	1521210	3.24
3	283.0057	283.0063	0.0007	1.8022	1687087	3.23
1	309.0658	309.0662	0.0004	2.2489	962758	4.07
2	309.0658	309.0663	0.0006	2.3286	1127675	4.06
3	309.0658	309.0663	0.0006	1.8022	1161140	4.06

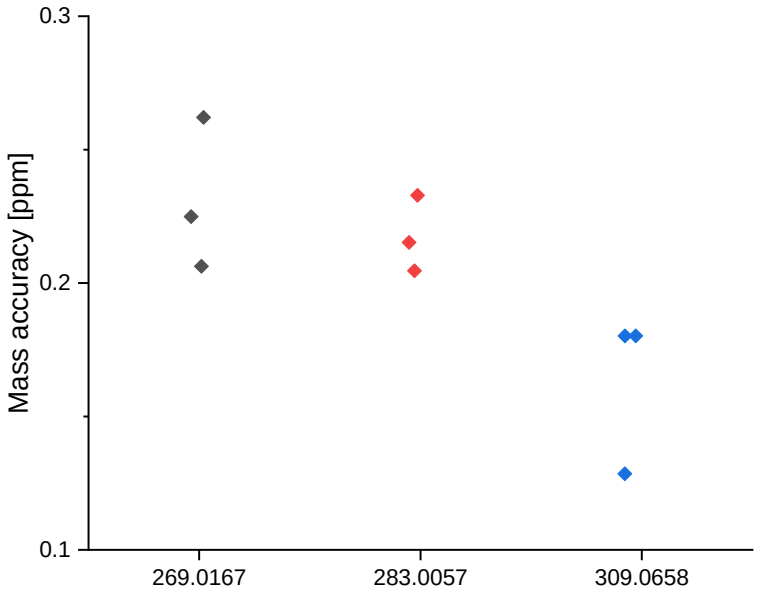


Figure S4: Mass accuracy for the six components of the in-house in positive ionization mode.

S 4.1.6 Methodology for statistical metabolome filtering

Feature definition

For statistical analysis all detected signals of the MS spectra were grouped into features. We used the TRex feature finding algorithms build in into Bruker Metaboscape 4.0.1 build 594. For two-dimensional data such as the DI measurements on the FTICR, TRex 2D was used while three-dimensional data such as the LC-TOF datasets were processed using TRex 3D. Features were defined as, the difference between signal, ion and feature as exemplified for a compound (Cpd X) showing an m/z at 331.28 and 313.28 Da (see Figure S5). In general, without any data processing an MS spectrum only consists of different signals detected at different m/z. During the ESI process different ions can be generated, for example the $[M+H]^+$ for Cpd X at 313.27 and $[M+Na]^+$ 331.28. Those ions can further be grouped to features by calculating the exact difference between the different ions. Further proof for the cohesiveness of the different ions can be obtained by comparing the peak progression of the ions, which correlate if the signals belong to the same compound.

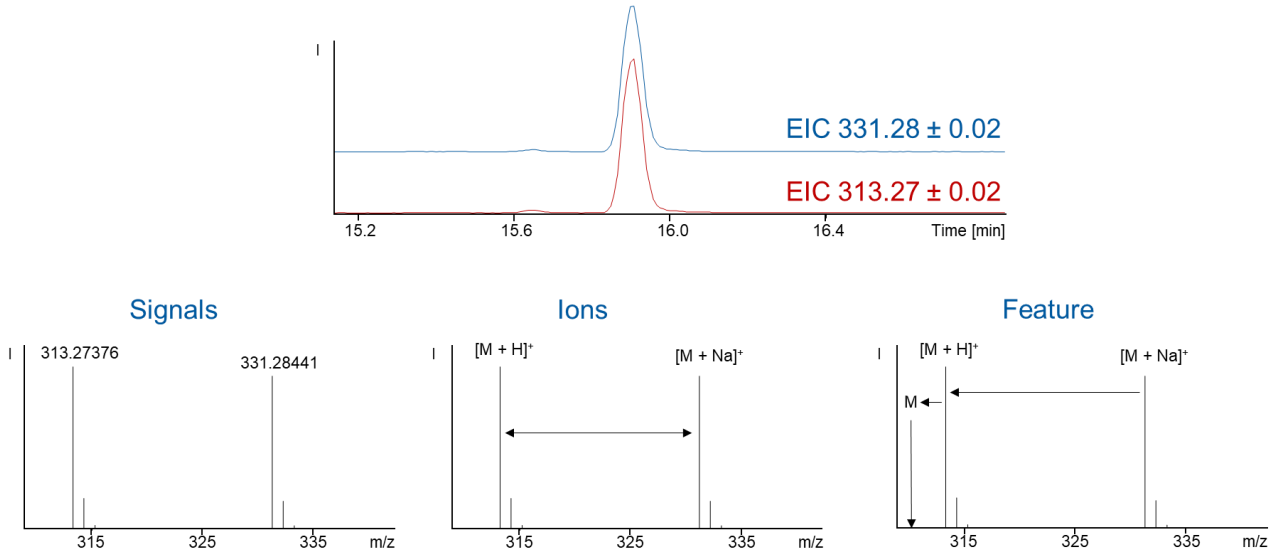


Figure S5: Difference between signals, ions and features exemplified for m/z 331.28 and 313.27.

S 4.1.7 Spectral networking details for the acquired tandem MS data

Molecular networks were created using a parent mass tolerance of 0.05 Da and a fragment ion tolerance of 0.1 Da. Cosine score of edges considered to network was set to 0.65 and the minimum matched fragment peaks was set to 4.

S 4.2 Segmented DI-FTICR method

As described in the material and methods part of the manuscript, for DI-FTICR measurements the mass spectrum was divided into eight segments (see Figure S6), using the quadrupole of the system as a filter. The collision RF Amplitude was optimized for each segment, resulting in 1000 Vpp for the segment from 150-200 m/z (1) and 800-1000 m/z (5), 1300 Vpp from 200-600 m/z (2,3) and 1000-1200 (6), 800 Vpp for 600-800 m/z (4) and 1800 Vpp for 1200-1500 (7,8). The spectra acquired in each segment were later stitched together by the FTMS control software (Bruker) to obtain a full spectrum from 150-1500 m/z.

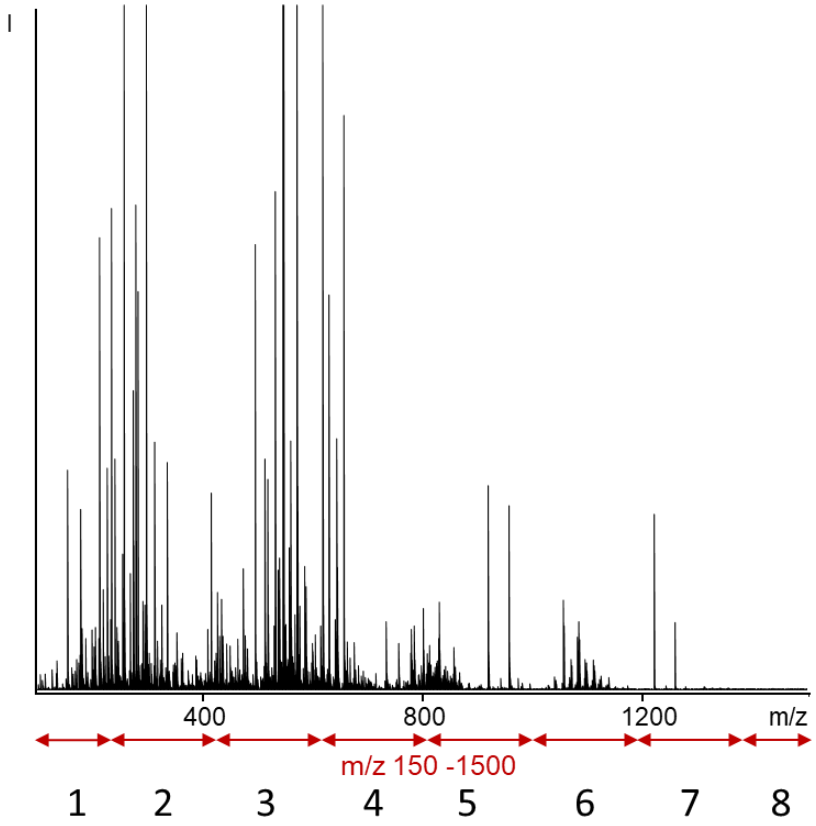


Figure S6: Segments used for the DI-FTICR measurements.

S 4.3 KNIME-workflow for feature comparison

Comparison of molecular feature content of the different analyses was done using SQL queries on an SQL server after upload of molecular feature data including retention time (RT, for LC based analyses) exact Mass (m), Intensity (I) and bucketing counts for feature presence in blank and sample analyses. Automated iterations of these SQL queries were performed using the KNIME software framework. A detailed description of the bucketing process is described in the discovery process of the fulvuthiacenes.[2] Manual access to the SQL server for functionality control was realized with NaviCat v12.1.27. Tabulated feature data containing the bucketing information is extracted from Metaboscape as a csv file. Automated SQL server upload and SQL queries are realized in the KNIME software framework v4.0.2. All KNIME workflows used in the comparison will be provided upon request.

KNIME comparison of the different MS based analyses is done as follows:

Upon upload, all features occurring in both blank and bacterial extract, which is indicated by a bucketing count of non-zero among the blanks, are omitted from upload to remove all medium associated LC-MS signals. Next, overlap of two different LC-MS based analyses is determined using grouping queries in SQL with a retention time tolerance of 0.2 min and a mass tolerance of 0.02 Da. Overlap of two different DI-MS based analyses is determined using a similar grouping query with a mass tolerance of 0.005 Da. The overlap of LC-MS and DI-MS analyses is determined using a grouping query with a mass tolerance of 0.02 Da. Tables containing all grouped molecular features containing information about retention time (RT, for LC based analyses) exact mass (m) and intensity (I) can be provided as csv files for each MS analysis under each cultivation condition.

S 4.4 Annotations of known secondary metabolites

All measurements were annotated with our in-house database Myxobase containing known myxobacterial secondary metabolites. As the analytical data was mainly deposited for positive ionization, negative ionization measurements were disregarded for this analysis. Compounds that were only detected with one setup, were manually checked for other adduct ions of the same compound in the other setups. Only hits with a mass deviation of less than 5 ppm and a retention time deviation of 0.2 min were considered accurate.

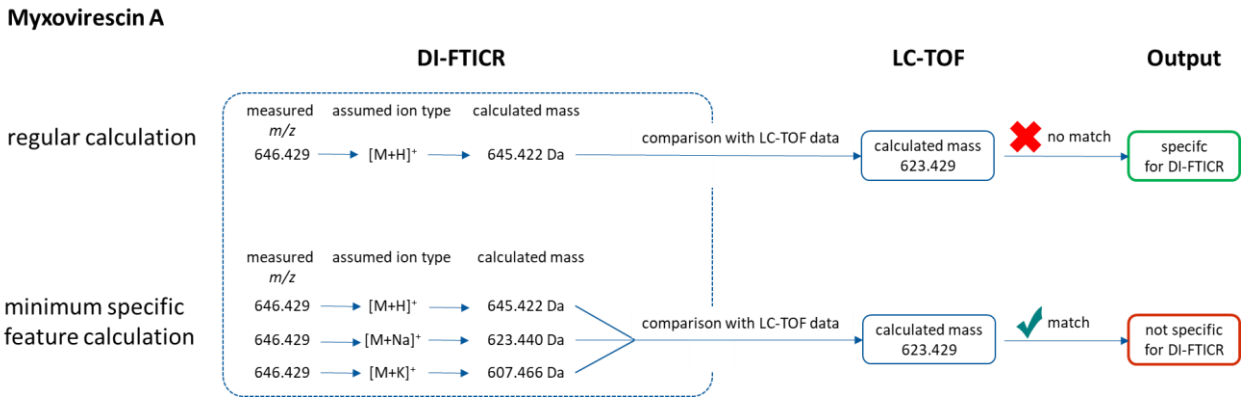
Table S7: Known myxobacterial secondary metabolites, detected in the extracts used for this study.

	LC-TOF+Liquid		**LC-TOF+Plate**		**DI-FTICR+Liquid**		**DI-FTICR+Plate**		
Mx122	**Δppm**	**ion type**	**Δppm**	**ion type**	**Δppm**	**ion type**	**Δppm**	**ion type**	**ΔRT [min]**
Dkxanthen-520	0,3	$[M+H]^+$							0,06
Dkxanthen-548	1	$[M+H]^+$	2	$[M+H]^+$					0,15
Homospermidin Lipid 521	0,4	$[M+H]^+$	2,3	$[M+H]^+$			1,1	[M+Na]+	0,05
Phenalamid A1	1,9	$[M+H]^+$	0,2	$[M-H_2O+H]^+$	2,9	[M+Na]+	1	[M+Na]+	0,01
Phenalamid B	0,3	$[M+H]^+$	2,6	$[M+H]^+$					0,22
Melithiazol A	0	[M+Na]+	0,4	[M+Na]+	0	$[M-MeOH+H]^+$	0,6	[M+Na]+	0,78
Melithiazol B			0,8	$[M-MeOH+H]^+$					0,04
Cystobactamid A1	0,8	$[M+H]^+$							0
Cystobactamid C	0,5	$[M+H]^+$							0,05
Myxalamid B			1,2	$[M+H]^+$	1,8	[M+Na]+			0,22
Soraphen B2α							3	$[M-H_2O+H]^+$	N/A
Soraphen C2α							2,1	$[M-H_2O+H]^+$	N/A
	LC-TOF+Liquid		**LC-TOF+Plate**		**DI-FTICR+Liquid**		**DI-FTICR+Plate**		
SoCe26	**Δppm**	**ion type**	**Δppm**	**ion type**	**Δppm**	**ion type**	**Δppm**	**ion type**	**ΔRT [min]**
Soraphen A	0,1	$[M-H_2O+H]^+$	0,6	$[M-H_2O+H]^+$	0,4	[M+Na]+	0,1	[M+Na]+	0,01
Soraphen V/C	1,5	$[M-2H_2O+H]^+$	0,3	$[M-H_2O+H]^+$	1,4	$[M+Na]^+$	1,2	$[M+K]^+$	0.01/0.15
Soraphen F2β	1,9	$[M-H_2O+H]^+$	0	$[M+NH_4]^+$	2,9	$[M+K]^+$	0,8	[M+Na]+	0,07
Soraphen A1γ/δ	0,8	$[M-H_2O+H]^+$	0,9	$[M-H_2O+H]^+$	2,8	[M+Na]+	2,4	[M+Na]+	0,05
Jerangolid A					2,1	[M+Na]+			N/A
Jerangolid E					4,8	$[M+NH_4]^+$			N/A
	LC-TOF+Liquid		**LC-TOF+Plate**		**DI-FTICR+Liquid**		**DI-FTICR+Plate**		
SbSr044	**Δppm**	**ion type**	**Δppm**	**ion type**	**Δppm**	**ion type**	**Δppm**	**ion type**	**ΔRT [min]**
Ambruticin Pre-S-OMe					0,9	$[M-H_2O+H]^+$			N/A

Ambruticin S			4,5	$[M-H_2O+H]^+$				0,98
Ambruticin VS-3 N-Oxid					2,6	$[M+H]^+$		N/A
Corallopyronin A2			0,7	$[M-H_2O+H]^+$				0,04
Salinimyxantin B			0,8	$[M+H]^+$			1,8	$[M+K]^+$	0,07
	LC-TOF+Liquid		**LC-TOF+Plate**		**DI-FTICR+Liquid**		**DI-FTICR+Plate**		
Mx152	**Δppm**	**ion type**	**Δppm**	**ion type**	**Δppm**	**ion type**	**Δppm**	**ion type**	**ΔRT [min]**
Pyrrolnitrin	0,5	$[M+H]^+$	0,6	$[M+H]^+$					0,02
Myxothiazol A	0,2	$[M-MeOH+H]^+$	0,1	$[M-MeOH+H]^+$			1,7	$[M+Na]^+$	0,04
Myxothiazol A Methylester	0,7	$[M+H]^+$	1,8	$[M+H]^+$	3,2	$[M+Na]^+$	0,3	$[M+H]^+$	0,09
Soraphen A							1,8	$[M-2H_2O+H]^+$	N/A
Soraphen C2α							2,1	$[M-H_2O+H]^+$	N/A

S 4.5 Minimum specific feature calculation

In order to evaluate the influence of ion types on our analyses, we performed a minimum specific feature calculation. With this we attempted to determine the number of features that were guaranteed unique to one instrument setup.



Scheme S1: Minimum number of unique feature calculation for the comparison of LC-TOF and DI-FTICR in positive ionization mode.

The features from the data sets Mx1_pos_liquid_LC, Mx1_pos_liquid_DI, Mx1_pos_plate_LC and Mx1_pos_plate_DI were matched against each other (LC vs. DI for liquid and plate respectively) to find common and setup specific features. The calculation was performed three times, under the assumption that all ions are either $[M+H]^+$, $[M+K]^+$ or $[M+Na]^+$. The sum of all commons was then subtracted from the total number of features to find the minimum number of specific features. (Table S8 and Scheme S1) This

calculation assumes no overlap between ion types and therefore underestimates the true number of instrument specific features.

Table S8: Results of the minimum number of unique feature calculation for the comparison of LC-TOF and DI-FTICR in positive ionization mode.

setup	total	$[M+H]^+$		$[M+Na]^+$		$[M+K]^+$		total		
		common	specific	common	specific	common	specific	sum commons	specifics absolute	specifics relative
Liquid + LC-TOF	573	88	485	93	480	86	487	267	306	53%
Liquid + DI-FTICR	1135	127	1008	220	915	133	1002	480	655	58%
Plate + LC-TOF	755	75	680	96	659	27	728	198	557	74%
Plate + DI-FTICR	783	133	650	167	616	51	732	351	432	55%

S 4.6 Molecular networks of unique secondary metabolites

Molecular networks of secondary metabolites found to be uniquely produced by one of the myxobacterial strains investigated in our analysis are listed in Figure S7-Figure S14. Self-loops (features without similar fragmentation pattern to any of the other features) were excluded from our analysis. For *Simulacricoccus ruber* MCy10636 no molecular network is shown as no features clustered to another.

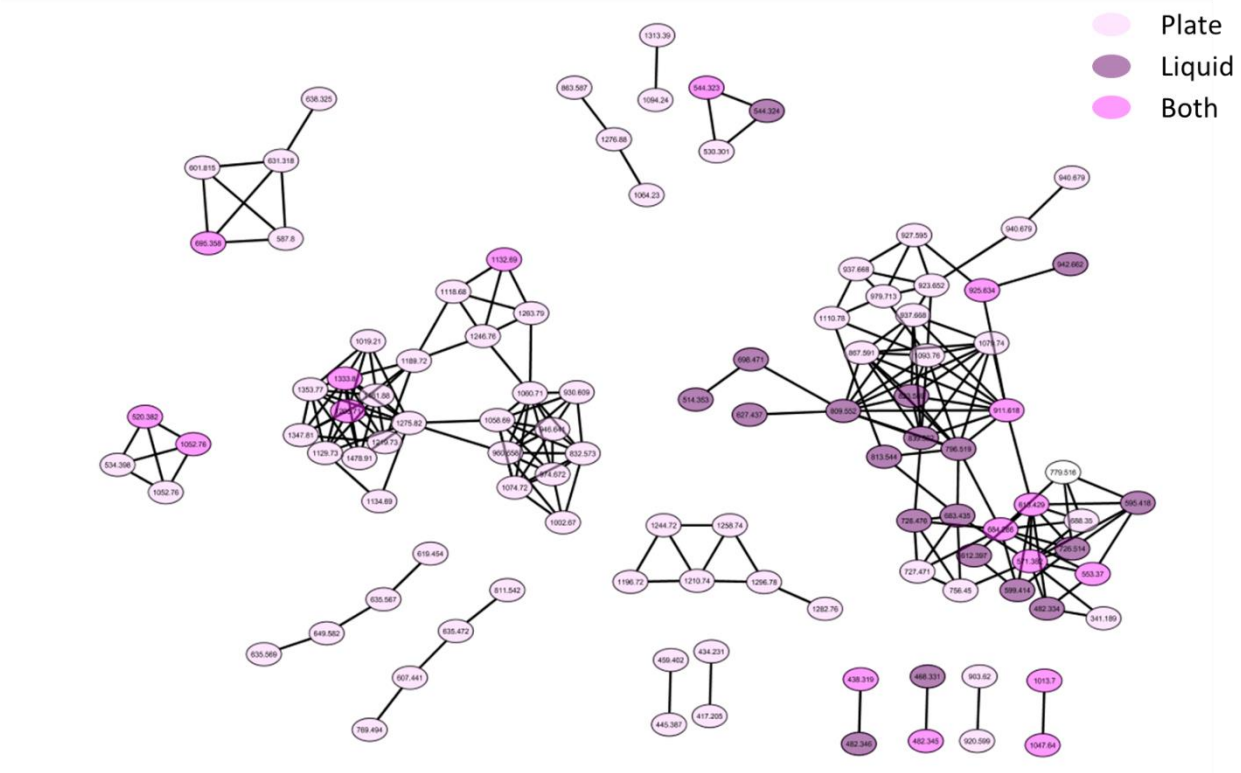


Figure S7: Molecular network of unique metabolites produced by strain MSr11367.

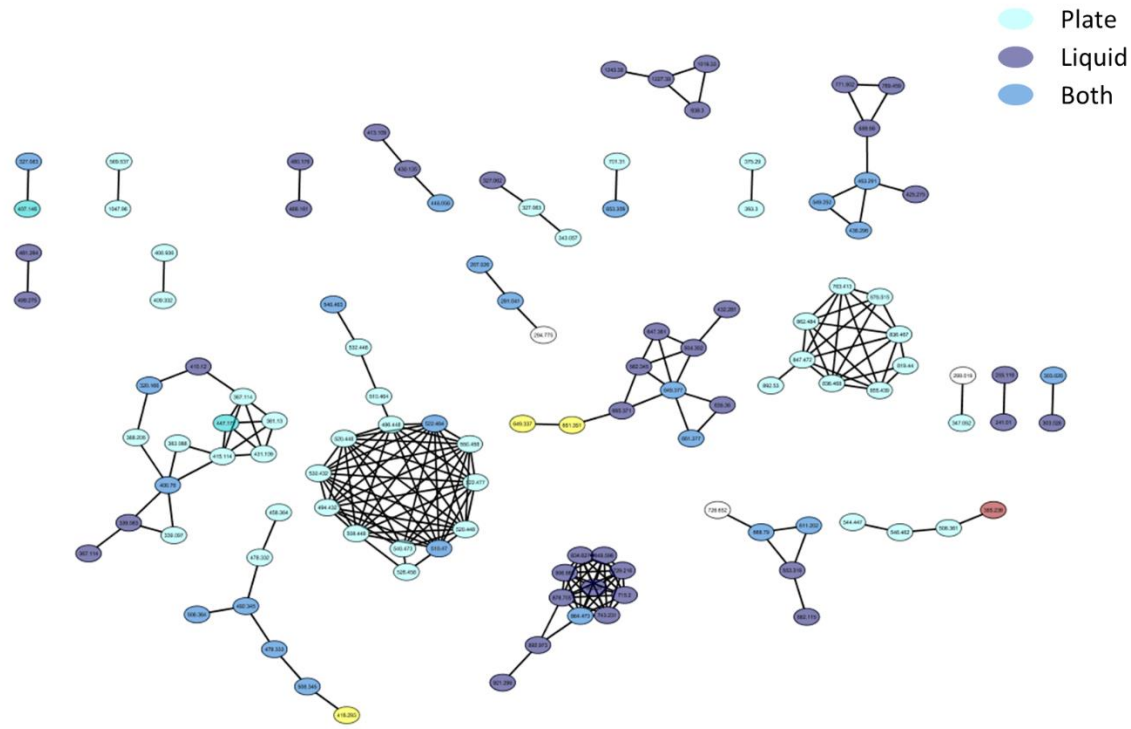


Figure S8: Molecular network of unique metabolites produced by Myxococcus fulvus MCy9270.

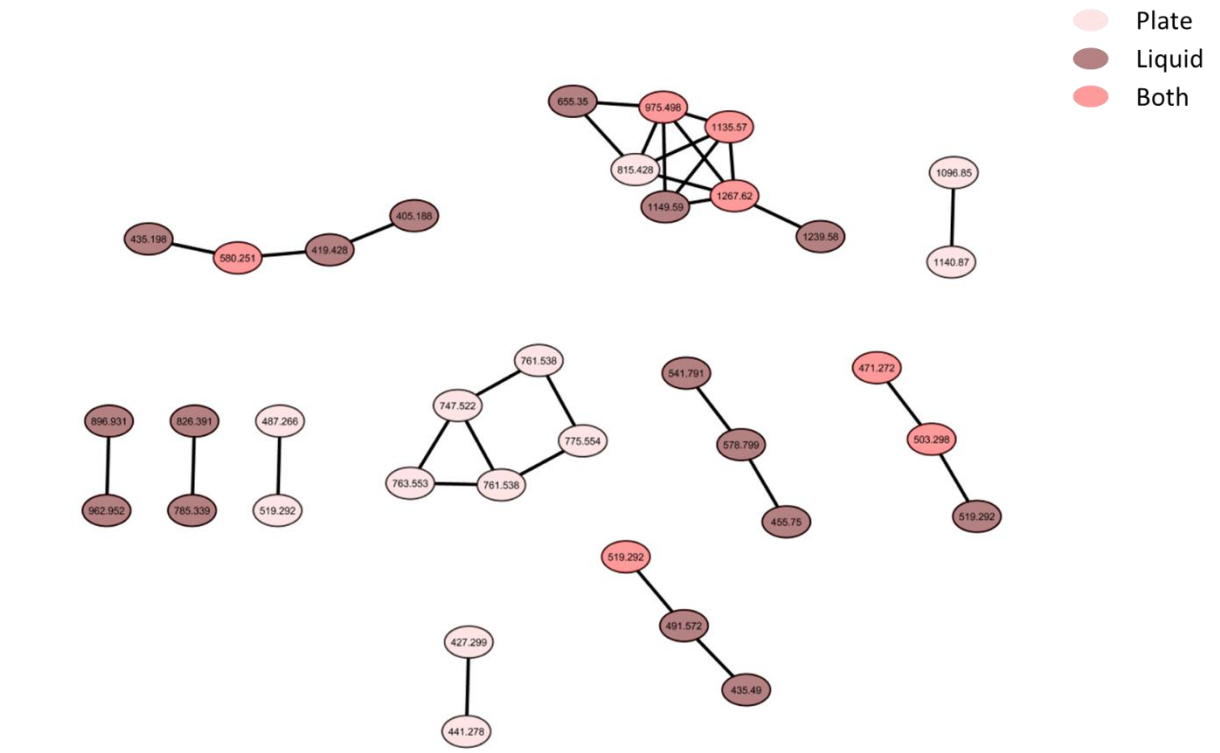


Figure S9: Molecular network of unique metabolites produced by Sorangium cellulosum SoCe 26.

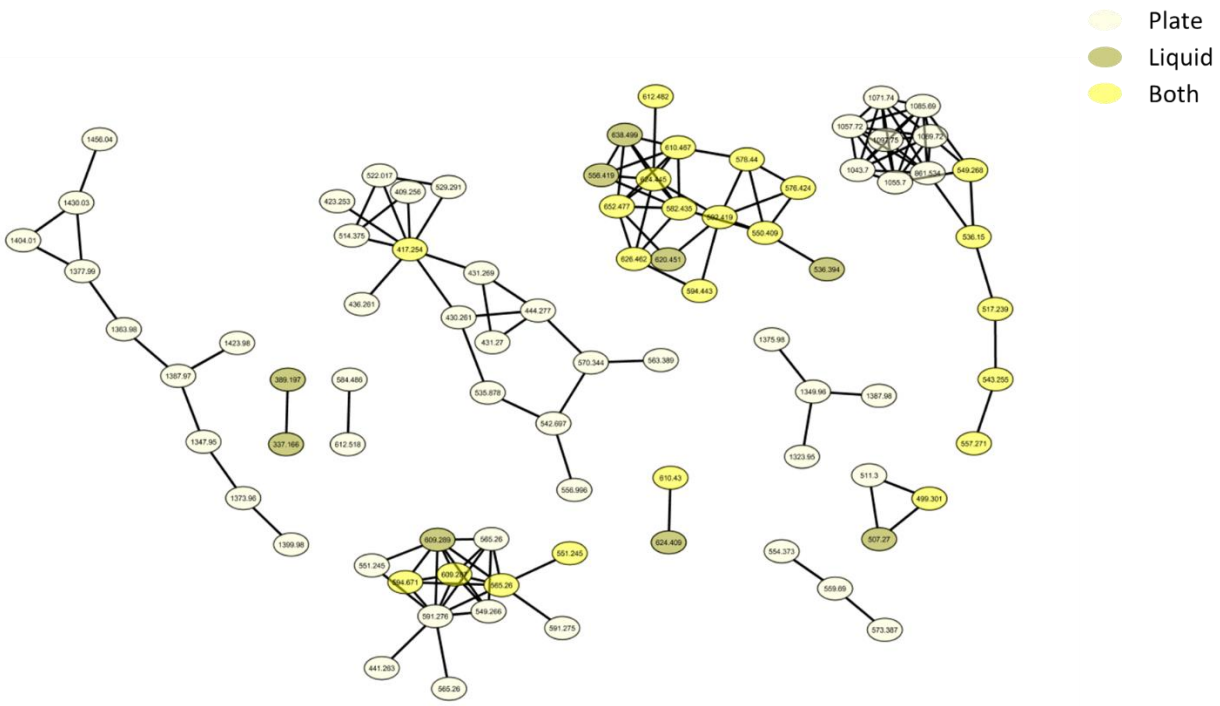


Figure S10: Molecular network of unique metabolites produced by Myxococcus xanthus DK1622.

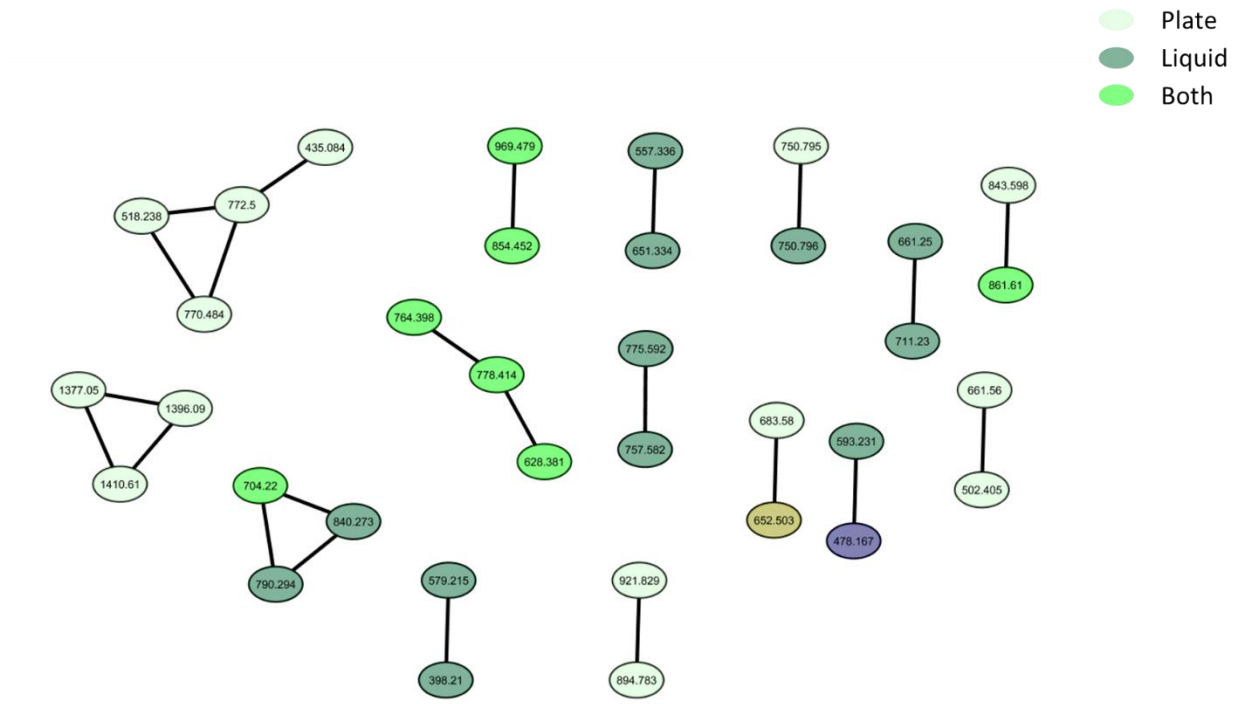


Figure S11: Molecular network of unique metabolites produced by Pyxidicoccus sp. MCy10649.

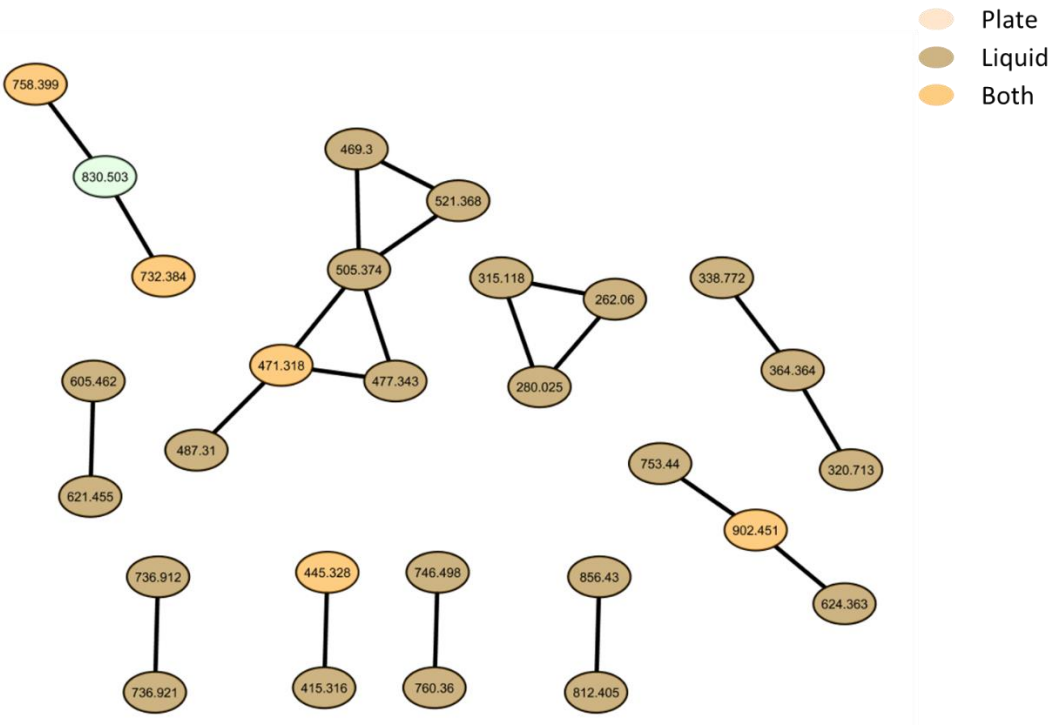


Figure S12: Molecular network of unique metabolites produced by MSr10575.

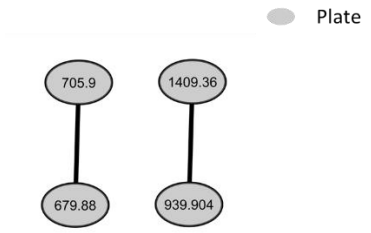


Figure S13: Molecular network of unique metabolites produced by MSr11770.

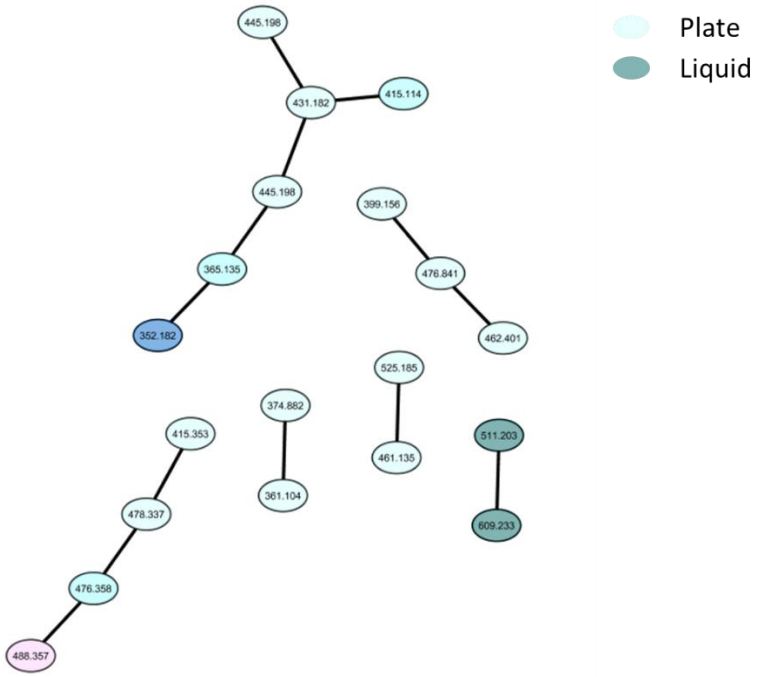


Figure S14: Molecular network of unique metabolites produced by MCy9487.

We furthermore were interested if there are any secondary metabolite clusters only detectable with LC-TOF, which would indicate that specific chemistry is only detactable with this setup or if the LC-TOF unique features are distributed among different secondary metabolite families. We therefore created a molecular network of all strain unique metabolites and highlighted the LC-TOF unique features (see figure S15). Interestingly, those features are distributed over the whole network. Most of the times we only find one member of a secondary metabolite family that is only detectable with LC-TOF. Some of the bigger clusters include more unique LC-TOF features, but for those big clusters there are always more features detectable with both systems than unique features. There is only one cluster consisting of two nodes, which is only detactable with LC-TOF. This cluster however consists of two features with the same exact mass, but

different retention times, why they likely are just isomers.This analysis therefore underlines in addition to the targeted analysis, that most of the secondary metabolite families detectable with LC-TOF are also detectable with DI-FTICR.

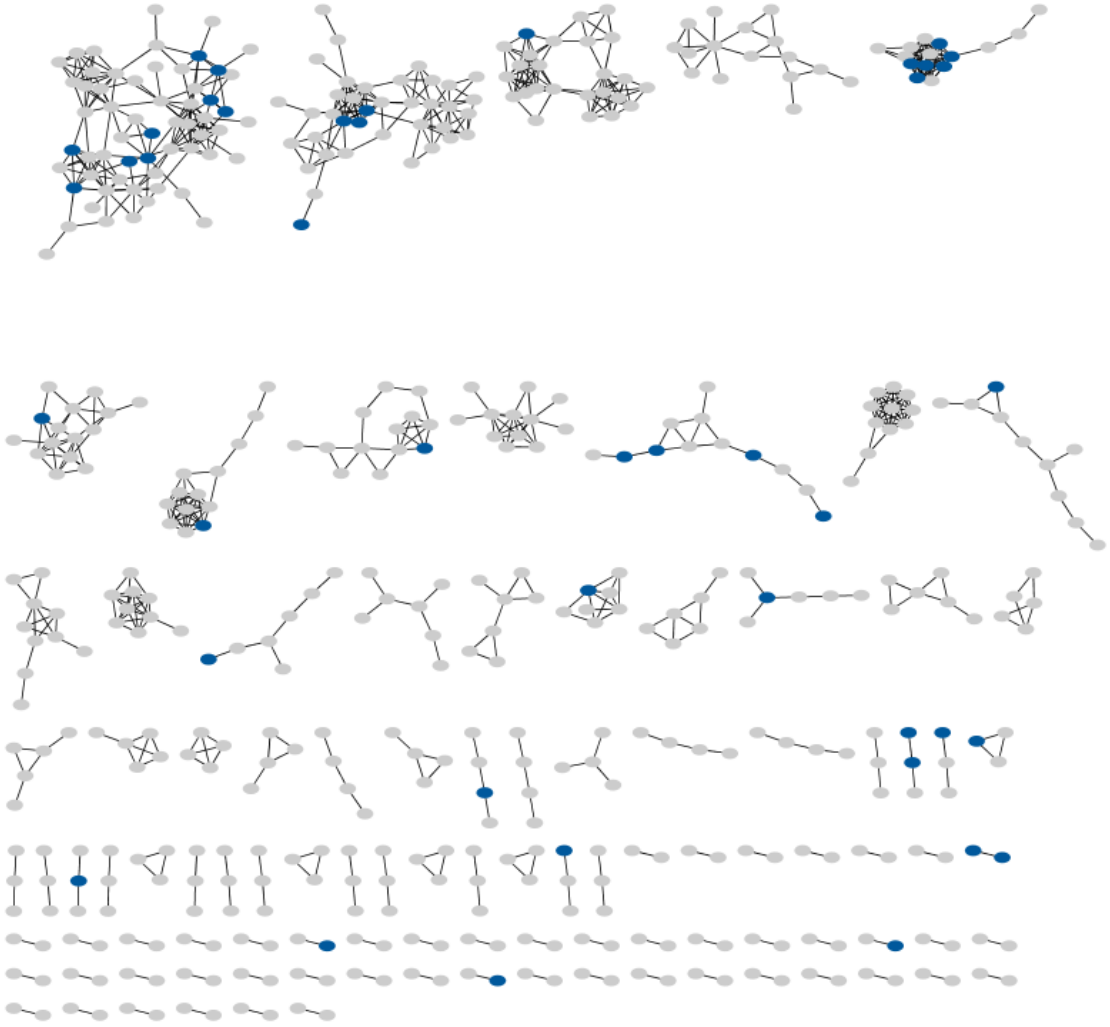


Figure S15: Molecular network of all strain unique metabolite with feautures only detectable with LC-TOF highlighted in blue and features detectable with LC-TOF and DI-FTICR in grey.

S 4.7 References

[1] Park, K. H.; Kim, M. S.; Baek, S. J.; Bae, I. H.; Seo, S.-W.; Kim, J.; Shin, Y. K.; Lee, Y.-M.; Kim, H. S. Simultaneous molecular formula determinations of natural compounds in a plant extract using 15 T Fourier transform ion cyclotron resonance mass spectrometry. Plant methods 2013, 9, 15.

[2] Panter, F.; Krug, D.; Müller, R. Novel Methoxymethacrylate Natural Products Uncovered by Statistics-Based Mining of the Myxococcus fulvus Secondary Metabolome. ACS Chem. Biol. 2019, 14, 88–98.

5. Discussion

The central theme of this thesis revolves around bacterial natural products and new strategies for accessing the bacterial secondary metabolome. The potential of underexplored bacterial strains to produce novel natural products (NP) was successfully exploited in chapters two and three of this work, describing the isolation and characterization of the thiamyxins from a myxobacterium and sesbanimide R from a magnetotactic bacterium. The third chapter is based on a "thinking outside of the standard analytical toolbox" approach to demonstrate the inherent limitations of conventional metabolome mining. Liquid chromatography - mass spectrometry with subsequent statistical data analysis, as well as genome mining with *in silico* analysis of biosynthetic gene clusters played vital roles during this research. The following chapter summarizes the broader impact of lessons learned in the individual projects by highlighting the importance of investigating previously underexplored microbial strains, critically assessing different cultivation conditions and analytical setups, and advocates for the usage of diverse methodological approaches when searching for novel natural products.

5.1 The potential of underexplored bacteria as NP sources

The most well-known bacterial NP producers have been exploited almost to the point of depletion. *Actinomycetes*, particularly *Streptomyces,* suffer from an extremely high rediscovery rate of already known NPs, making it challenging to identify unprecedented NP families and not only new congeners of already described NP classes.[1] Rediscovery can be mitigated by robust dereplication efforts early in the screening process and the likelihood of identifying new compounds can be increased by focusing on rare bacterial taxa as phylogenetic distance was shown to correlate with chemical diversity.[2,3] It is thought that of all bacterial species present on earth only a fraction has so far been cultivated under laboratory conditions and, while not all of these uncultured taxa will have a high capacity to produce natural products, this nevertheless highlights an immense potential still awaiting discovery.[4,5]

A drawback of focusing on underexplored bacteria, is that there is often little experience in cultivating and genetically manipulating these species and activating their full secondary metabolome under laboratory conditions is therefore far from trivial.[2]One promising strategy to overcome these issues is to repurpose strains that were thus far only investigated for presenting particularly interesting phenotypes in

microbiology studies and not for their NP production capacity. Since the handling of many of these strains is already established in laboratories, most difficulties in their management are already known and possibly eliminated, thus the work can focus on activating their biosynthetic machinery and characterizing the correlating NPs. Myxobacteria were a prime example of such a group of microorganisms. Originally studied for their interesting morphology and predatory behavior, researchers were able to leverage their knowledge of cultivation and genetic manipulation. Today myxobacteria are established as prolific producers of natural products. This is exemplified in this work by the *Myxococcaceae* strain MCy9487 of the *Myxococcus-Pyxidicoccus-Corallococcus* clade, but of a yet unclassified genus, from which the antiviral thiamyxins were isolated (see chapter 2). Another example of a taxonomically distinct strain that was a source for novel chemistry is *Sandaracinus sp*. MSr10575 from which sandarazol, a rare myxobacterial plasmid encoded cytotoxic natural product was recently isolated.[6] Both of these examples highlight that taxonomic diversity is key in identifying novel natural products, yet despite these efforts, the vast amount of myxobacterial taxa yet remains uncultured and thereby unexplored.[7,8] These findings have increased the drive of the scientific community to realize the potential of myxobacteria and other microbial families and strive towards utilizing them as sources of natural products. As highlighted by Challinor *et al.*, these sources include anaerobic bacteria, pathogens, and microbes with symbiotic life cycles.[2] Isolated in 2010 from *Clostridium cellolyticum*, closthioamide was the first natural product to be isolated from obligate anaerobic bacteria. It was only discovered after its biosynthesis was induced by adding aqueous soil extract to the culture broth and it exhibits strong activity against methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant *Enterococcus faecalis*.[9] The polyene enacyloxin IIa was first isolated in 1982 from *Frateuria* sp. W-315 and later also connected to a hybrid NRPS–PKS gene cluster from *Burkholderia ambifaria.* It exhibits activity against Gram-positive and Gram-negative pathogens such as *Enterococcus faecalis* as well as *Acinetobacter baumannii* and exemplifies that pathogens themselves can be a source for antibiotic natural products.[10,11] An example of a natural product that was isolated from a symbiont is tilivalline, which was first isolated in 1982 from *Klebsiella oxytoca,* a human gut commensal.[12] It exhibits cytotoxic activity and therefore can cause antibiotic associated hemorrhagic colitis (AAHC) when antibiotic resistant *K. oxytoca* overgrows the healthy human gut microbiome as a result of antibiotic therapy.[13] The identification of tilivalline as a contributing factor to AAHC enabled the identification of salicylic and acetylsalicylic acid as inhibitors of tilivalline production in liquid culture.[14] This might pave the way for therapeutic applications of salicylic and acetylsalicylic acid as treatments for AAHC. Finally, the fascinating tripartite symbiosis between attine ants, various fungi such as *Basidionmycota* of the order *Agaricales* and filamentous actinobacteria of the genus *Pseudonocardia* is a remarkable example of an ecological system

that could be exploited in the search for bioactive natural products. In this mutualistic relationship, the ants tend “fungal gardens” and use the fungi to break down plant matter, which they then can use as a food source.[15] The *Pseudonocardia* spp. live in specialized anatomical compartments of the ants and produce natural products that can protect the fungus from other predatory fungi.[16] In 2009, dentigerumycin was isolated from from *Pseudonocardia spp.,* and was shown to have potent activity against the parasitic fungus *Escovopsis* sp.[17] Prior to the study presented in chapter 3, magnetotactic bacteria have been largely neglected with regards to their potential of NP production. Their model organism *Magnetospirillum gryphiswaldense*, has been extensively studied for the intriguing ability to form magnetosomes, its handling under laboratory conditions is well established, the toolkit for genetic manipulation of the strain is available and in 2015 Araujo *et al.* demonstrated the potential of several magnetotactic species to produce NPs.[18,19] This enabled us to study its biosynthetic potential and ultimately lead to the isolation of the cytotoxic sesbanimide R, the first NP isolated from a magnetotactic

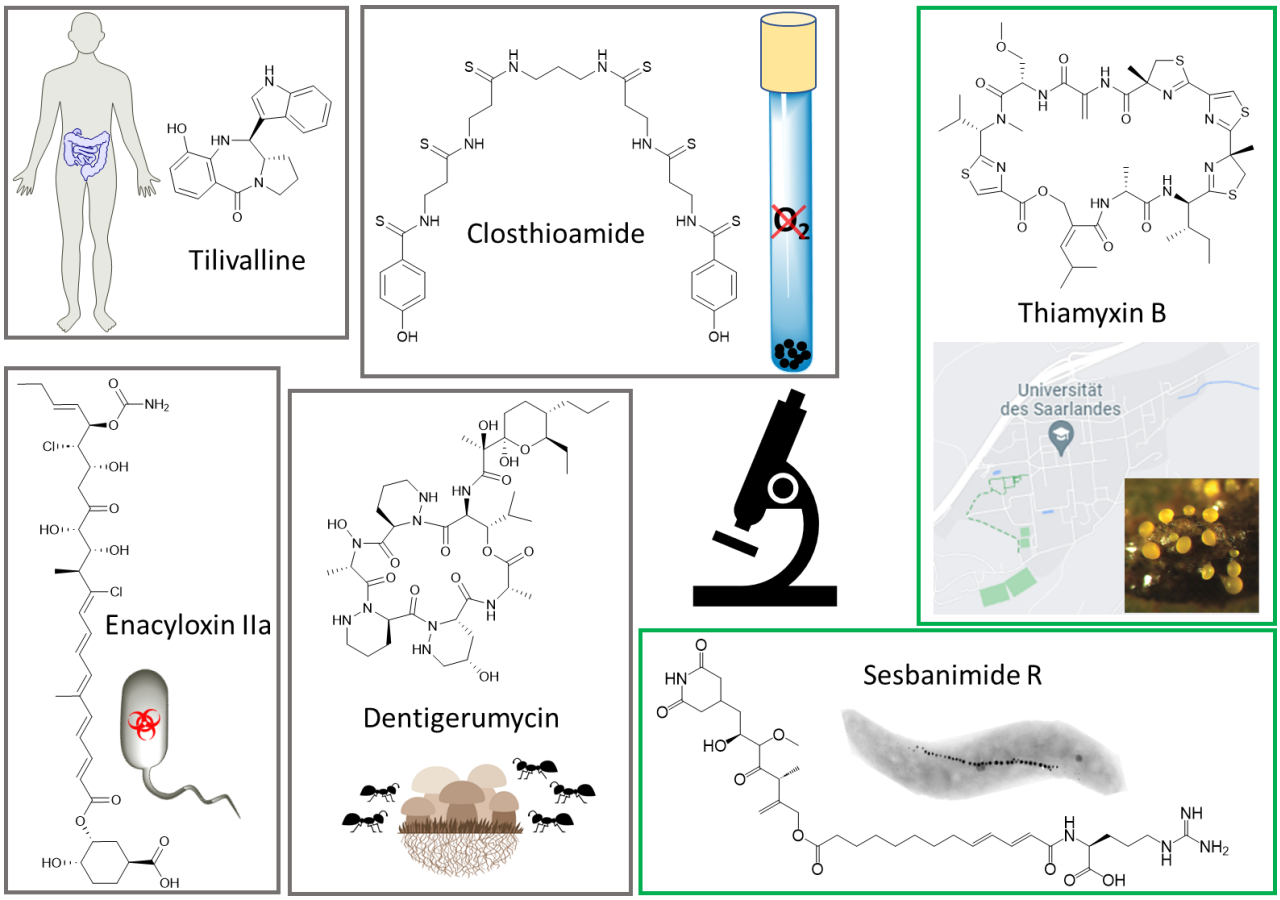


Figure 1: Chemical structures of selected natural products originating from neglected bacteria or underexplored habitats. Tilivalline and dentigerumycin from the symbiotic *Klebsiella oxytoca and Pseudonocardia spp* respectively; closthioamide from the anaerobic *Clostridium cellolyticum*; enacyloxin IIa from the pathogenic *Burkholderia ambifaria*; thiamyxin B from the myxobacterial strain MCy9487 isolated from our own university campus and sesbanimide R from *Magnetospirillum gryphiswaldense*.

bacterium. Furthermore, we were able to associate sesbanimide R to a biosynthetic gene cluster in the *M. gryphiswaldense* genome, adding magnetotactic bacteria to the pool of potential NP sources. All of these examples highlight, that not only underexplored families and genera, but also unique and underexplored ecological niches should be considered as starting points for natural product research.

5.2 Environmental triggers induce the production of natural products

Besides accessing bacteria that are challenging to culture under laboratory conditions, the investigation of biosynthetic gene clusters that are considered "cryptic" holds great potential for discovering novel and potentially bioactive NPs.[20] Genetic manipulation, as exemplified in chapter 3, and heterologous expression can be considered as brute force strategies that compel the strain to produce specific products. Although this has led to the discovery of several intriguing NPs, these interventions require availability of the respective genome sequences and appropriate tools for manipulation. Furthermore they are often associated with a loss in cell viability and in the case of heterologous expression the remaining uncertainty if the obtained product is the true product of the full assembly line.[21] Therefore, alternative, less invasive strategies are sought to persuade microbes to produce the desperately desired NPs, by utilizing triggers that are suspected to be involved in enhancing production in the native environment. These strategies may involve co-cultivation of the producing organism with other microbes, the addition of environmental factors into the growth medium, or the modification of the media composition to mimic the natural environment more closely.

Notable examples of this strategy include the isolation of myxarylin in 2021 from *Pyxidicoccus fallax*, achieved by adding autoclaved *Escherichia coli* or yeast cultures to the growth medium.[22] Another instance is the discovery of the alkaloid aspergicin, which was obtained from the co-culture of two mangrove epiphytic *Aspergillus* species, and shows activity against Gram-positive and Gram-negative bacteria.[23] Finally, the isolation of the previously highlighted closthioamide from *Clostridium cellulolyticum*, was only made possible by inducing its production through the addition of aqueous soil extract to the growth medium.[9] These examples highlight the efficacy of environmental triggers in uncovering previously unknown natural products. Especially for myxobacteria there is an opportunity for comprehensive studies with robust statistical analysis to determine different conditions optimized for the production of natural products. In the comparative study outlined in chapter 4 we demonstrated that even a simple change in the growth medium, from liquid to solid, leads to a striking alteration in the metabolic

profiles of the investigated strains. Liquid cultivation of bacteria offers many advantages in terms of scalability, growth speed and ease of the extraction process, while plate cultivation better emulates the natural habitat of myxobacteria, as they are predominately dwelling in soil, which enables the formation of morphological phenotypes such as fruiting bodies that are absent in liquid cultivation. This approach underscores the possibilities of harnessing the inherent capability of bacteria to adapt to their surroundings to induce the production of natural products. Another application of environmental triggers could be to analyze the production of sesbanimide R while cultivating *M. gryphiswaldense* under the influence of different magnetic fields, as we already observed a difference in magnetic response between the wildtype strain and deletion mutants.

In summary, to maximize the access to the genomic potential of microbes, a combination of targeted activation through genetic manipulation, as described in chapter 3 for sesbanimide R, and induction through environmental triggers as an untargeted approach should be considered.

5.3 Leveraging methodical diversity to access the bacterial secondary metabolome

The quest for novel bioactive natural products begins with the search for a suitable source, the microbial strain, and culminates in the isolation of the desired natural product, usually from a complex matrix. Every step of this journey significantly impacts the detectability and subsequent isolation of the target compounds. The selection of the microbial strain and the optimization of its cultivation conditions form the foundation of this endeavor, while each subsequent step adds a discriminatory layer. (Figure 2 and 3) In the following paragraphs, the impact of different extraction methods, chromatography systems and mass spectrometers along with their ion sources, are discussed in the context of natural products research.

5.3.1 The first step: Extraction

During the cultivation of potential NP producers, adsorber resin is often added to the culture broth to capture natural products that are excreted by the cells. Subsequent solvent extraction is commonly employed to elute any produced NPs, primary metabolites, and media components from the resin. The choice of solvent and adsorber resin significantly influences the selectivity of the extraction, as the polarity of the solvent largely determines the compounds that can be bound and extracted. The purification of sesbanimide R (chapter 3) for example, was greatly enhanced by a liquid-liquid partition step following the initial methanolic extraction. Liquid-liquid partitioning is one of the only separation methods which does

not reduce the total number of analytes, which makes it a useful addition to almost any purification workflow. Partitioning the extract between hexane, ethyl-acetate and water resulted in the collection of sesbanimide R in the aqueous layer. This could subsequently be dried and resuspended in methanol to obtain an extract that was already sufficiently pure enough for semi-preparative HPLC separation without the need for additional pre-purification steps. In the identification and isolation of the thiamyxins, a standard methanolic extraction was employed due to the, at this time, unknown chemistry of the target compounds and the inclusion of an additional solvent such as chloroform might have increased the extraction yield or enabled the identification of more members of the compound family. The following examples further highlight diverse extraction methods that can improve the coverage of bacterial metabolomes:

For some compounds a classical solvent extraction is highly inefficient, as exemplified by the discovery of chloroxanthic acid A, which was isolated from *Myxococcus xanthus* DK1622. Due to the low concentration of this compound in conventional extracts, it could only be obtained in sufficient amounts for structure elucidation by enrichment via super-critical-fluid-extraction (SFE).[24] SFE commonly employs supercritical CO_2 as a solvent, benefitting from its excellent matrix penetration and compound solvation, therefore combining the advantages of gases and liquids. Furthermore, SFE is considered a green extraction method, as it eliminates the need for organic solvents, which are burdensome to dispose of.[25,26] Other extraction methods that minimize solvent usage include pressurized liquid extraction and microwave assisted extraction (MAE). By using microwaves to rapidly heat the sample, both solvent consumption and extraction time can be reduced.[27] Chen *et al*. recently used an optimized MAE protocol to extract several bioactive terpenoids from lotus plumule.[28] Further factors that impact extraction efficiency include the surface area of the extraction matrix (for solid extraction), the miscibility of the solvent and growth medium (for liquid-liquid extraction), the solvent-to-solute ratio, extraction temperature and extraction duration.[29] Considering these factors, it is of vital importance to tailor the extraction process according to specific goals. For compound screening approaches, it might be advisable to use several extraction methods to cover a chemical space as wide as possible, while for isolation purposes of a specific compound class with known physico-chemical properties a highly selective extraction method is desired to simplify subsequent purification steps.

5.3.2 The effect of different separation techniques

The chromatographic separation of bacterial extracts plays a crucial role in natural products discovery for several reasons:[30] In screening studies, analytical-scale separation is employed to obtain a comprehensive overview of the secondary metabolome of the source organisms, as extensively demonstrated in chapter

4. Information about the retention time of compounds can be used as an identifier for library matching and early dereplication of target compounds, which is essential to avoid rediscovery. Conversely, to obtain the amount of pure compound necessary for structure elucidation and biological evaluation, various methods of preparative and semi-preparative scale separation can be applied. For natural products, a commonly used chromatographic separation principle is high-performance liquid chromatography (HPLC), which serves as an essential part of each chapter of this thesis, as it is highly flexible and does not require complex sample preparation.[31] HPLC utilizes a small particle size of 2–50 µm with pore sizes typically between 8 and 12 nm. This results in superior resolving power when compared to flash chromatography.[32] There are a multitude of different stationary and mobile phases that can be selected based on the properties of the analytes and the requirements of the study. One of the oldest forms of liquid chromatography is normal phase chromatography (NPC), which employs a polar stationary phase (e.g. silica), an apolar mobile phase (e.g. hexane), and mostly relies on hydrogen-bonding and dipole interaction of the analyte with the stationary phase.[33,34] More commonly used in current HPLC applications, is reverse phase chromatography (RPC). RPC employs an apolar stationary phase (e.g. C_{18} chains fused to silica) and a polar or moderately polar mobile phase (methanol or acetonitrile mixed with water). The analytes are separated based on the strength of their hydrophobic interaction with the stationary phase.[35] RPC often uses gradient elution to decrease the polarity of the mobile phase over the course of a run. Most biological small molecules have a good solubility in aqueous solvents (since cytoplasm is also aqueous) and RPC setups can be easily coupled to mass spectrometers which are commonly used in analytical and preparative workflows.[36] These advantages make RPC the most commonly used HPLC method in natural products research and RPC coupled with mass spectrometry (MS) also served as the key method for the screening and isolation workflows described in chapters 2, 3 and 4 of this work. However, relying solely on one type of separation principle also bears the risk of overlooking potential target compounds. Molecules might co-elute with other compounds that overload the detector, they could be insoluble in the mobile phase, or they might even permanently adhere to the stationary phase and be missed completely. When screening new strains for potential targets for isolation, it is therefore advisable to employ at least two orthogonal separation principles when analyzing the crude extract.

As described in the recent review by Brandão *et al*. two LC systems can also be coupled in one setup to achieve a comprehensive separation.[37] Although this technique has some limitations, such as maintaining the resolution that was obtained in the first separation during the second separation, it has been successfully applied to analyze several complex extracts in untargeted analysis: For instance, the bioactive

components of hops were analyzed by coupling RPC with RPC using a shifted gradient in the second dimension.[38] Metabolic profiling of *Glycyrrhiza glabra* using a RPC x RPC setup with a microbore cyano column and a superficially porous octadecylsilica column enabled the detection of 120 compounds and the tentative identification of 37 of them.[39] In another study, *Schisandra chinensis,* which is used in traditional Chinese medicine, was investigated for bioactive components by coupling an immobilized liposome column with a silica monolithic column.[40]

Additional types of LC separation mechanisms, that can be used to (pre-)separate the complex crude extract fractions of microbial culture broths, include ion exchange, size exclusion and affinity chromatography. Ion exchange chromatography is based on the interaction of ions in the sample with stationary charges on the solid phase. Ions of opposite charge are retained on the column, while ions of the same charge are eluted with the solvent. While this type of separation is typically not practicable for the screening of crude extracts due to its binary nature, in a recent study, a cation exchange adsorber resin has shown beneficial for enhancing the isolation of several novel darobactin derivatives.[41] Size exclusion chromatography (SEC) separates analytes based on their physical size and shape. The column contains porous particles that allow small molecules to penetrate deeper into the pores, while larger molecules show less interactions and elute faster. As this is a low-resolution separation, it is primarily used as a pre-purification step or for isolating compound mixtures such as flavonoids from bamboo leaves.[42] During the purification of sesbanimide R, a pre-purification step using size exclusion chromatography could have potentially increased the efficiency of the isolation procedure. However, as sesbanimide R was observed to degrade under prolonged exposure to oxygen and light, the long duration of an SEC pre-purification made it unsuitable. Finally, affinity chromatography employs ligands attached to the stationary phase that can form covalent bonds with certain target molecules, while other molecules do not interact with them. The target molecules can then be eluted by using a wash buffer that disrupts the covalent bonds. In this case only knowledge of the binding property of the target analytes is required, while other factors such as size or polarity play only a minimal role. For these reasons, this technique is frequently used in nucleic acid and protein purification.[43] Recently, affinity chromatography has also been applied in natural products screening efforts, for example by loading crude extract onto a bio-affinity column containing immobilized target molecules. These molecules can capture affinity binders from the extract, which are subsequently eluted for further analysis.[44] One of the major findings of the study described in chapter 4, was that we were able to detect a significant number of analytes only with the direct infusion setup, compared to the RPC setup. It stands to reason that the chromatographic separation significantly influenced the detectable analytical space of the extracts. However, in order to isolate sufficient amounts of a target mass only

detected with direct infusion some sort of scalable separation is required. One possible approach could be to compare the measurable spectrum of several different separation principles and find a method that does not discriminate against the desired target.

Although the methods described above are commonly used for specific purification purposes, they can also be useful for the analysis of bacterial crude extracts. Since crude extracts usually are highly complex, separating them into fractions prior to the initial screening can facilitate the detection of smaller amounts of each analyte. In activity-based assays, basic fractionation at an early stage can also reveal activities that might otherwise be overshadowed by more dominant effects. Furthermore, once a target compound has been identified for purification, it is advisable to use an orthogonal separation method for pre-purification (e.g. preparative size exclusion chromatography) in order to increase the efficiency of the final purification (e.g. semi-preparative reverse phase chromatography). In the isolation of sesbanimide R (chapter 3) and the thiamyxins (chapter 2), a high degree of purity was already achieved after respective liquid-liquid partitioning protocols, eliminating the need for additional pre-purification steps.

In summary, as with extraction methods, it is of the utmost importance to tailor the separation process to the aim of the study: general screening studies for example, require highly flexible methods that can separate a broad spectrum of compounds. Here a combination of RP-HPLC and SEC HPLC can be used to vastly increase the physico-chemical space covered, as their separation principles are based on polarity and compound size respectively, which significantly reduces the likelihood that a whole class of compounds remains completely undetected. On the other hand, when a specific compound or NP-class is targeted, highly discriminatory methods such as affinity chromatography become much more valuable, as they can drastically reduce the number of purification steps required to obtain a pure compound. For these reasons a clear idea of what each study is trying to achieve, and knowledge of the limitations and biases of each method are essential to any separation workflow.

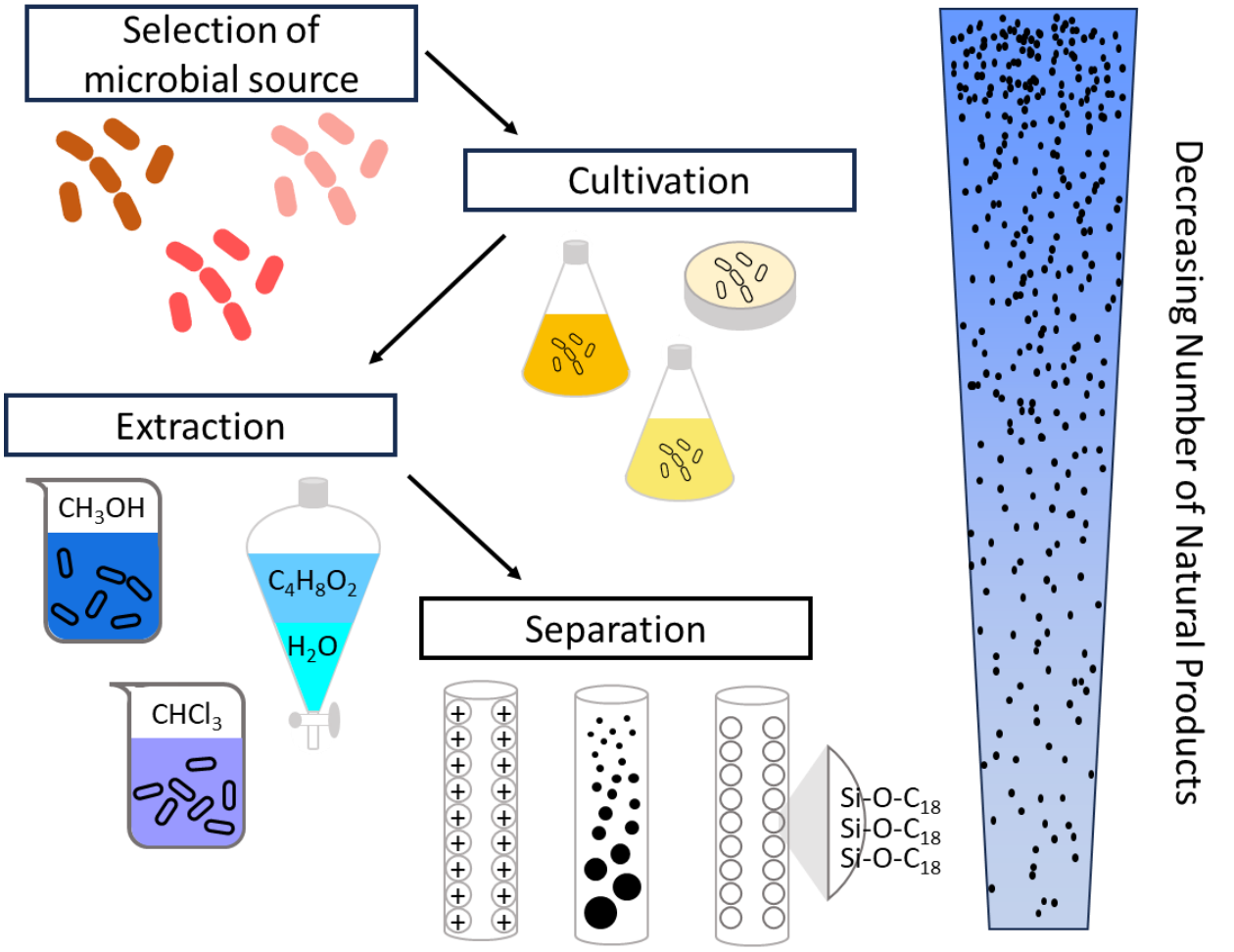


Figure 2*:* Schematic representation of the first part of a compound identification workflow (microbe → cultivation → extraction → separation) indicating the loss of metabolic information introduced by each step.

5.3.3 Ionization – the gateway to mass spectrometry

To gain knowledge of the components of a crude extract, they need to be detected using appropriate analytical techniques. Most HPLC setups accomplish this using diode-array-detectors (DAD), which rely on the absorption of analytes in the visible and ultraviolet spectrum.[45] While these detectors are inexpensive and robust, their sensitivity is highly compound dependent, and they are limited in the separation of spectra from coeluting NPs. Therefore, mass spectrometry (MS), which offers enhanced specificity compared to DADs and provides information on the molecular weight of the analytes, has emerged as an invaluable tool for the analysis of complex extracts.[46] Mass spectrometers are mainly used in tandem with DADs, as the information provided by each detector is complementary. Depending on the ionization mode, mass spectrometers are easily coupled to an HPLC flow, but can also be used via direct infusion (DI) (chapter 4). The principles and challenges of these systems are discussed in the following section.

The most limiting part of a mass spectrometric analysis is converting analytes into gas-phase ions, as this step determines which analytes can reach the mass analyzer ultimately leading to their detection. For the analysis of complex mixtures such as myxobacterial extracts, a high degree of source-fragmentation as resulting from so called hard ionization techniques such as electron ionization (EI) is usually undesirable as it complicates the assignment of the fragments to their parent molecules. Natural products, in particular, are not well suited for EI as most NPs are thermally unstable or not volatile, making them incompatible with the source conditions required for EI.[47] Therefore, soft ionization techniques are preferred for the analysis of natural products. These techniques, such as electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), and matrix assisted laser desorption ionization (MALDI) operate under milder conditions, avoiding extensive in-source fragmentation and thermal degradation of the analytes.[48]

ESI is a particularly popular ionization technique as it is well suited for coupling to liquid chromatography which is prevalent in NP research. ESI also enables the generation of multiply charged molecules thereby dramatically extending the mass range capability of the mass spectrometer. This was already observed for the thiamyxins in chapter 2, which were measured with two charges and becomes even more important for larger molecules such as peptides or proteins, for which ESI has become the most commonly applied ionization technique for this reason.[49] Within the scope of this work, ESI has proven to be an incredibly versatile ionization technique and was employed as the sole ionization method in both analytical setups (direct infusion FT-ICR and LC-TOF) used in the comparison study in chapter 4. One drawback of ESI is the occurrence of so-called ion suppression.[50] Due to the limited number of protons available for ionization within each droplet, analytes that ionize well tend to receive more protons compared to others. As a result, their true quantities are not accurately represented by their intensities in the resulting mass spectrum. This limitation makes ESI unsuitable for quantitative analysis, and during screening workflows, it can lead to “missing” detection of a molecule, because it co-elutes with a particularly well ionizing analyte. This likely also was a major limiting factor of the study in chapter 4, as ion suppression affects direct infusion setups to a larger degree than liquid chromatography coupled setups. In this study, no specific steps were implemented to reduce ion suppression, although the inclusion of a desalting step during sample preparation could be considered. Introducing an additional step would, however, prolong the total time required per sample, thereby compromising one of the benefits associated with using a direct infusion setup. Additionally, given the comparisons we already made regarding growth conditions, separation, ionization polarity, and mass detectors, introducing different ionization sources as a fourth parameter was not considered. Future studies should explore the expansion of the detectable metabolome by employing

other ionization techniques such as APCI, which suffers less from ion suppression effects and can access a different chemical space.[51,52]

Another technique called matrix-assisted laser desorption ionization (MALDI)[53] utilizes a matrix irradiated with a laser for ionization. The matrix absorbs most of the laser's energy, resulting in minimal fragmentation of the analytes, therefore making it highly suitable for the ionization of NPs.[54] MALDI offers many advantages such as low cost and fast, easy sample preparation. However, it cannot be easily coupled to a separation system, making it unsuitable for compound purification, similar to the direct infusion ESI setup employed in chapter 4. Furthermore, the matrix background can limit the sensitivity for compounds smaller than 1 kDa and the laser can degrade photosensitive compounds.[48] MALDI is particularly useful in imaging studies, enabling the analysis of specific regions within the sample and providing the visual representation of the spatial distribution of NPs and other molecules. For example, MALDI imaging can potentially be used to further study the magnetosome formation of *Magnetosprillium gryphiswaldense* by analyzing the spatial and temporal distribution of the metabolites involved in magnetosome formation. Furthermore, MALDI imaging is also used to analyze the distribution of pharmaceuticals in animal models such as zebrafish, which might be an interesting next step in the development of the thiamyxins as antiviral therapeutics.

Similar to extraction and separation, the selection of the ionization method and its parameters, have a major impact on the detectable metabolites in a natural product screening workflow. The choice between positive ionization and negative ionization mode alone can greatly influence the number of molecular features that are detectable. As demonstrated in chapter 4, most myxobacterial NPs tend to ionize slightly better in the positive ionization mode, which is consistent with the findings of Nielsen *et al.*, who tested 719 natural product reference standards and observed a higher detection rate of 93% in positive ESI mode compared to 77% in negative ESI mode.[55] However, this does not imply that negative ionization can be disregarded. The initial stages of untargeted NP screenings should always include both ionization modes to cover a chemical space as broad as possible. While the biases introduced by extraction and separation methods can relatively easily be mitigated by expanding the process through additional orthogonal methods, the ionization method used presents a unique consideration. If a target compound can only be detected by an ionization technique that cannot easily be coupled to HPLC separation, such as MALDI or the direct infusion ESI setup outlined in chapter 4, subsequent purification of this compound becomes be exceedingly burdensome. In such cases, an approach involving blind fractionation and subsequent analysis of the fractions would likely be employed. Nevertheless, this approach is valuable, as it unveils exactly

these kinds of compounds that would have remained concealed in traditional standardized screening workflows.

5.3.4 The last filter: mass analysis

After the generation of analyte ions, determining their mass-to-charge ratio (m/z) becomes crucial. Different types of mass analyzers employ different principles to achieve m/z determination. Quadrupole mass filters are likely the most commonly used mass analyzers. As the name suggests, they selectively allow ions within a specific *m/z* range to pass through by manipulating their flight path using oscillating electric fields. By adjusting the direct current (DC) and radio frequency (RF) voltages applied to the quadrupole mass filter, the entire mass spectrum can be scanned.[56] A single quadrupole mass spectrometer coupled to a semipreparative HPLC system was also used in the purification process of sesbanimide R and the thiamyxins described in chapters 3 and 2 respectively. Using an inline mass analyzer offers tight process control and eliminates the need for subsequent control measurements. However, it should be noted that a portion of the LC flowthrough (typically around 10 %) is used for mass analysis, which reduces the final compound yield. This loss of material to the mass spectrometer was negligible in these cases, as the production titers of these compounds were sufficient. Particularly for sesbanimide R, a fast and accurate purification was preferred due to the previously mentioned sensitivity to light and oxygen. Once the purification is stable, the split into the mass spectrometer can also be removed to mitigate the loss of final product, although this also nullifies the tight process control. Quadrupole mass filters can also be used upstream of other mass analyzers to enhance their sensitivity, mass resolution, and mass accuracy and/or to serve as collision cells for MS^2 experiments.[57] Quadrupole ion traps are closely related to regular quadrupole mass filter, with the difference being that ions are trapped by an RF field and then sequentially ejected based on their m/z.[56] Of particular note is the capability ion trap instruments to measure ions of positive and negative polarity simultaneously, which greatly increases the potential metabolomic coverage when used during screening workflows.[58] Because of the robustness of these instruments, ESI-ion trap mass spectrometers are particularly useful for quick and reliable analysis of extract fractions when purifying compounds based solely on UV-VIS detection, as it allows the exclusion of impurities that do not absorb in the UV-VIS spectrum.

Another widely used mass spectrometry technique is time-of-flight mass spectrometry (TOF). This technique involves accelerating ions and measuring the time they require to reach a detector, which corresponds to their m/z values. Lighter ions of the same charge reach higher velocities thus exhibiting a shorter flight time compared to heavier ions of the same charge.[59] TOF instruments coupled with two quadrupole mass filters are known as Q*q*TOF mass spectrometers. These instruments were used in the

initial screening of crude extracts and subsequent dereplication in all three chapters of this thesis, due to their high resolution and mass accuracy. However, one limitation of TOF instruments is the inability to rapidly switch polarities due to the design of the flight tube. Consequently, this prohibits simultaneous measurement of positive and negative ions and therefore requires two separate measurements, if both polarities are required.

Accurate mass determination (≤ 5 ppm) enables the determination of sum formulas for compounds, which is essential for structure elucidation purposes. This was especially important during the structure elucidation of sesbanimide R, as several possible sum formulas were identified based on Q*q*TOF data. Only by conducting measurements with even higher mass accuracy, we successfully determined the correct sum formula, unveiling an uncommonly high nitrogen content, indicating the incorporation of an arginine moiety. This breakthrough was accomplished by using a Fourier-transform ion cyclotron resonance mass spectrometer (FT-ICR) which, together with Orbitrap mass spectrometers, falls within the realm of Fourier transform mass spectrometry (FT-MS). FT-ICR instruments employ the principle of a penning trap, in which the rotation frequency of the trapped ions is measured as free induction decay (FID). By performing the mathematical operation of Fourier transformation on the FID, a mass spectrum can be obtained.[60] The exceptional resolution of FT-ICR MS allows its application in NP screening workflows even without prior HPLC separation, as demonstrated by direct infusion (DI) MS in chapter 4 of this thesis. This approach offers faster screening times per sample and eliminates the discriminatory effect of the separation method. However, the removal of the separation step can lead to pronounced ion suppression as well as space charge effects. In this work, space charge was mitigated by employing a spectral stitching method that only allowed a certain mass range to enter the penning trap at a time, thereby reducing the number of different ion species that could negatively impact the measurements. Both FT-ICR and orbitrap are capable of MS^n experiments by capturing ions after an initial fragmentation step and subjecting them to another round of fragmentation. This is used to generate fragmentation trees, where fragments are followed and assigned to the larger fragment they are originating from. This could be particularly useful in the case of cyclic peptides such as the thiamyxins. In contrast to linear peptides, which often have an easy-to-follow fragmentation patterns, the macrocycle of cyclic peptides often leads to rather unpredictable fragmentation. The combination of high mass accuracy, resolving power and MS^n capability positions FT-ICR mass analyzers and Orbitraps as the most powerful mass spectrometers available today. A coupling with HPLC systems requires an increase of their scan rate, which comes with a decrease in resolution to a similar range as TOF-instruments. Considering the high cost of FT-ICR instruments, it is therefore not surprising that their application remains limited to niches. such as native protein mass spectrometry. In

these highly delicate measurements, proteins are measured through nano ESI MS, while preserving their native quaternary structure.[61] This method was previously used to identify 96 binding partners of 32 putative malarial targets by incubating NP fragments with the respective protein targets and subsequently measuring the complexes through native FT-ICR MS.[62] Furthermore, an automated and highly efficient combination of ligand fishing approaches and an HPLC-MS-based NP screening setup, was recently developed. In this workflow, a cyanobacterial extract was first separated by HPLC, and then a target protein was infused orthogonally to generate protein-ligand complexes, which were directly measured by native MS. The same procedure was repeated without the protein infusion to obtain MS/MS data of the natural products. By comparing retention times and mass shifts of the two measurements, Reher *et al.* successfully identified a serine protease inhibitor family termed rivulariapeptolides.[63] This direct coupling of native protein-ligand MS to natural product MS represents a promising future addition to both NP screening approaches discussed in chapter 4.

Like the previously discussed filters, the choice of mass analyzer has a profound impact on the compounds that can be detected. While the earlier methods primarily excluded compounds based on their chemical properties, mass analyzers primarily discriminate based on the *m/z* of the analyte ions and the sensitivity of the analyzer. Once a sufficient quantity of an analyte is ionized, the upper and lower mass limits become the only factors limiting the detection of this analyte. Considering this, Q*q*TOF instruments emerge as the most suitable for metabolite screening workflows, offering a combination of high sensitivity, sub 5 ppm mass accuracy, and the widest dynamic range among the discussed mass analyzers. Nevertheless, it is advisable to complement Q*q*TOF measurements with ultra-high-resolution instruments such as Orbitrap and FT-ICR mass spectrometers, whenever available, to bridge the gap in sensitivity and mass accuracy. Due to their relatively low cost, quadrupole instruments, including ion traps, are better suited for rapid control measurements and for coupling with semi-preparative HPLC systems employed in purification processes.

Most of the different extraction, separation and analytical methods discussed here, can be combined with each other based on the specific requirements of the compounds or study. For the work presented in this thesis, RP-HPLC coupled to an ESI-Q*q*TOF mass spectrometer has proven invaluable in the early stages of target identification. The highly standardized and robust methods enabled dereplication against our in-house compound library as well as statistical filtering of the results. For the isolation of the thiamyxins as well as sesbanimide R liquid-liquid extraction protocols followed by reverse-phase HPLC coupled to a single quadrupole mass detector have proven to be highly successful. Especially for sesbanimide R a short and efficient purification was essential due to the previously mentioned photo- and oxygen sensitivity. And,

while not yet applicable as a general screening workflow, the rather specialized direct infusion - FT-ICR measurements performed in chapter 4 revealed a stunning and previously unseen chemical space that awaits to be further explored.

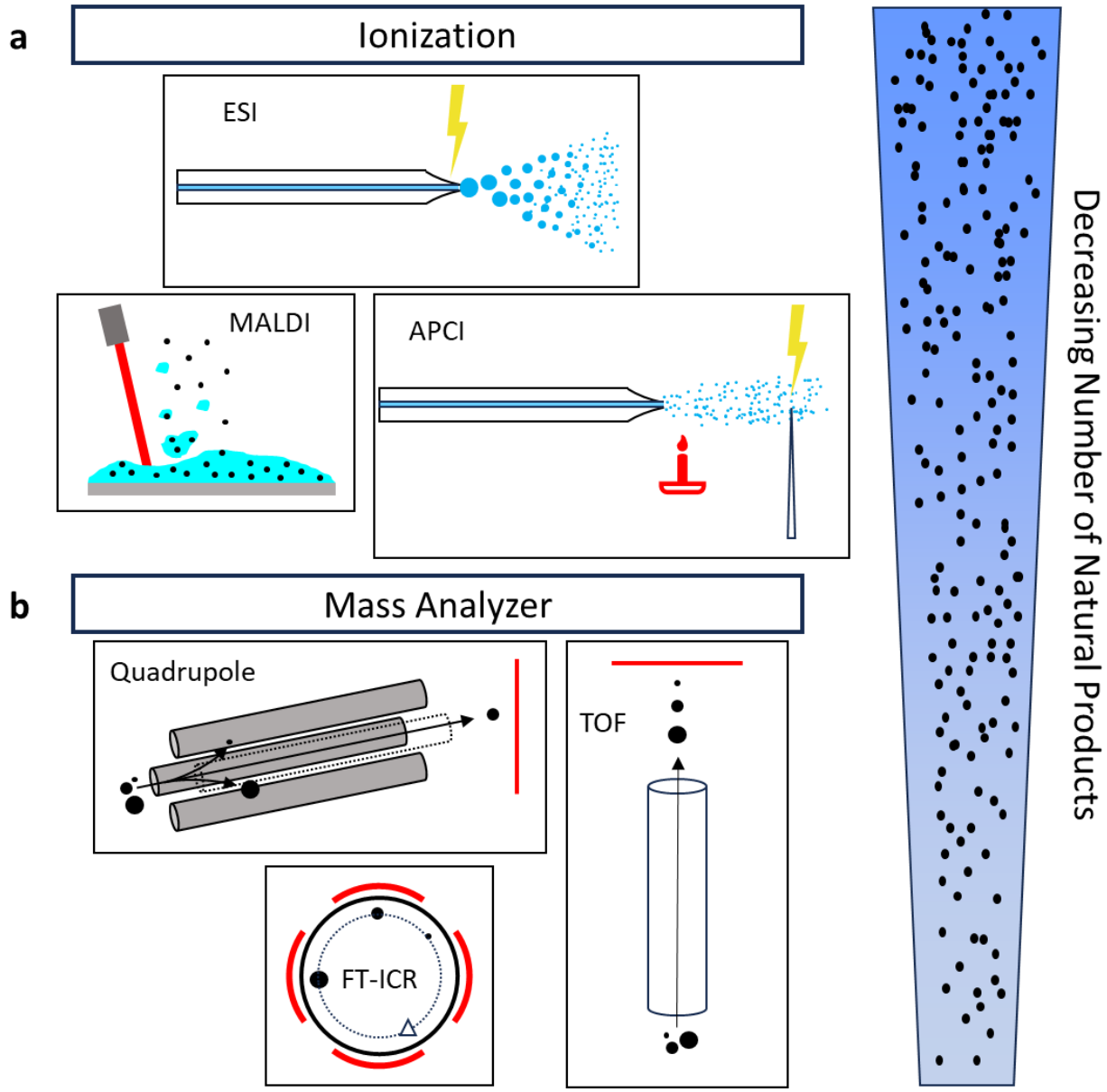


Figure 3: Schematic representation of the mass spectrometry aspect of a compound identification workflow (ionization & mass analysis) indicating the loss of metabolic information introduced by each step. **a**: top: ESI; bottom left: MALDI; bottom right: APCI - **b**: top left: single quadrupole; bottom left: FT-ICR; right: TOF

5.4 Conclusion

This thesis focused on the exploration of natural product discovery from unconventional bacterial sources using cutting-edge analytical techniques. It resulted in the discovery, isolation, and characterization of two biologically relevant compounds: the thiamyxin compound family and sesbanimide R, a derivative of the sesbanimides. The thiamyxins showed promising activity against RNA viruses, including SARS-CoV-2 and could serve as a starting point for further development towards therapeutic applications. Sesbanimide R exhibits cytotoxic effects in cell-based assays and is a prime example of finding a novel natural product in a previously disregarded ecological niche.

Another key finding of this study was the substantial enhancement of metabolomic data obtained during screening workflows by employing complementary orthogonal analysis methods, effectively increasing the metabolomic coverage nearly 2-fold. These results strongly support the notion that a wealth of untapped natural products awaits discovery. However, the challenge lies in cultivating new and phylogenetically diverse bacterial strains and allocating sufficient resources to utilize advanced analytical tools, thereby expanding the coverage of chemical space.

Overall, this research underscores the importance of leveraging state-of-the-art analytical approaches and investing in the exploration of novel bacterial sources to further uncover the vast potential of natural products.

5.5 References

[1] A. G. Atanasov, S. B. Zotchev, V. M. Dirsch, C. T. Supuran, *Nat Rev Drug Discov* **2021**, *20*, 200–216.

[2] V. L. Challinor, H. B. Bode, *Ann N Y Acad Sci* **2015**, *1354*, 82–97.

[3] T. Hoffmann, D. Krug, N. Bozkurt, S. Duddela, R. Jansen, R. Garcia, K. Gerth, H. Steinmetz, R. Müller, *Nat Commun* **2018**, *9*, 803.

[4] E. J. Stewart, *J Bacteriol* **2012**, *194*, 4151–4160.

[5] M. W. Hahn, U. Koll, J. Schmidt, **2019**, pp. 313–351.

[6] F. Panter, C. D. Bader, R. Müller, *Angewandte Chemie International Edition* **2021**, *60*, 8081–8088.

[7] S. C. Wenzel, R. Müller, *Mol Biosyst* **2009**, *5*, 567.

[8] K. Mohr, *Microorganisms* **2018**, *6*, 84.

[9] T. Lincke, S. Behnken, K. Ishida, M. Roth, C. Hertweck, *Angewandte Chemie International Edition* **2010**, *49*, 2011–2013.

[10] T. WATANABE, K. IZAKI, H. TAKAHASHI, *J Antibiot (Tokyo)* **1982**, *35*, 1141–1147.

[11] E. Mahenthiralingam, L. Song, A. Sass, J. White, C. Wilmot, A. Marchbank, O. Boaisha, J. Paine, D. Knight, G. L. Challis, *Chem Biol* **2011**, *18*, 665–677.

[12] N. Mohr, H. Budzikiewicz, *Tetrahedron* **1982**, *38*, 147–152.

[13] G. Schneditz, J. Rentner, S. Roier, J. Pletz, K. A. T. Herzog, R. Bücker, H. Troeger, S. Schild, H. Weber, R. Breinbauer, G. Gorkiewicz, C. Högenauer, E. L. Zechner, *Proceedings of the National Academy of Sciences* **2014**, *111*, 13181–13186.

[14] A. von Tesmar, M. Hoffmann, A. Abou Fayad, S. Hüttel, V. Schmitt, J. Herrmann, R. Müller, *ACS Chem Biol* **2018**, *13*, 812–819.

[15] T. R. Schultz, S. G. Brady, *Proceedings of the National Academy of Sciences* **2008**, *105*, 5435–5440.

[16] T. R. Ramadhar, C. Beemelmanns, C. R. Currie, J. Clardy, *J Antibiot (Tokyo)* **2014**, *67*, 53–58.

[17] D.-C. Oh, M. Poulsen, C. R. Currie, J. Clardy, *Nat Chem Biol* **2009**, *5*, 391–393.

[18] A. Araujo, F. Abreu, K. Silva, D. Bazylinski, U. Lins, *Mar Drugs* **2015**, *13*, 389–430.

[19] D. Schultheiss, D. Schüler, *Arch Microbiol* **2003**, *179*, 89–94.

[20] C. D. Bader, F. Panter, R. Müller, *Biotechnol Adv* **2020**, *39*, 107480.

[21] L. Huo, J. J. Hug, C. Fu, X. Bian, Y. Zhang, R. Müller, *Nat Prod Rep* **2019**, *36*, 1412–1436.

[22] J. J. Hug, N. A. Frank, C. Walt, P. Šenica, F. Panter, R. Müller, *Molecules* **2021**, *26*, 7483.

[23] F. Zhu, G. Chen, X. Chen, M. Huang, X. Wan, *Chem Nat Compd* **2011**, *47*, 767–769.

[24] C. D. Bader, M. Neuber, F. Panter, D. Krug, R. Müller, *Anal Chem* **2020**, *92*, 15403–15411.

[25] G. Mazzoccanti, S. Manetto, A. Ciogli, C. Villani, F. Gasparrini, *TrAC Trends in Analytical Chemistry* **2022**, *147*, 116511.

[26] T. A. Berger, *J Chromatogr A* **2016**, *1459*, 136–144.

[27] C. Sparr Eskilsson, E. Björklund, *J Chromatogr A* **2000**, *902*, 227–250.

[28] W. Xiong, X. Chen, G. Lv, D. Hu, J. Zhao, S. Li, *J Pharm Anal* **2016**, *6*, 382–388.

[29] Q.-W. Zhang, L.-G. Lin, W.-C. Ye, *Chin Med* **2018**, *13*, 20.

[30] D. GI. Kingston, *J Nat Prod* **1979**, *42*, 237–260.

[31] J.-L. Wolfender, *Planta Med* **2009**, *75*, 719–734.

[32] Z. Latif, S. D. Sarker, **2012**, pp. 255–274.

[33] S. R. Abbott, *J Chromatogr Sci* **1980**, *18*, 540–550.

[34] J. J. Pesek, M. T. Matyska, R. I. Boysen, Y. Yang, M. T. W. Hearn, *TrAC Trends in Analytical Chemistry* **2013**, *42*, 64–73.

[35] E. Cruz, M. R. Euerby, C. M. Johnson, C. A. Hackett, *Chromatographia* **1997**, *44*, 151–161.

[36] R. F. C. Mantoura, C. A. Llewellyn, *Anal Chim Acta* **1983**, *151*, 297–314.

[37] P. F. Brandão, A. C. Duarte, R. M. B. O. Duarte, *TrAC Trends in Analytical Chemistry* **2019**, *116*, 186–197.

[38] E. Sommella, F. Pagano, E. Salviati, M. Chieppa, A. Bertamino, M. Manfra, M. Sala, E. Novellino, P. Campiglia, *J Sep Sci* **2018**, *41*, 1548–1557.

[39] Y. F. Wong, F. Cacciola, S. Fermas, S. Riga, D. James, V. Manzin, B. Bonnet, P. J. Marriott, P. Dugo, L. Mondello, *Electrophoresis* **2018**, *39*, 1993–2000.

[40] S. Wang, C. Wang, X. Zhao, S. Mao, Y. Wu, G. Fan, *Anal Chim Acta* **2012**, *713*, 121–129.

[41] C. E. Seyfert, C. Porten, B. Yuan, S. Deckarm, F. Panter, C. D. Bader, J. Coetzee, F. Deschner, K. H. M. E. Tehrani, P. G. Higgins, H. Seifert, T. C. Marlovits, J. Herrmann, R. Müller, *Angewandte Chemie International Edition* **2023**, *62*, DOI 10.1002/anie.202214094.

[42] S. Ye, F. Pan, L. Yao, H. Fang, Y. Cheng, Z. Zhang, Y. Chen, A. Zhang, *Chem Biodivers* **2022**, *19*, DOI 10.1002/cbdv.202200506.

[43] N. E. Labrou, *Journal of Chromatography B* **2003**, *790*, 67–78.

[44] J. Guo, H. Lin, J. Wang, Y. Lin, T. Zhang, Z. Jiang, *J Pharm Biomed Anal* **2019**, *165*, 182–197.

[45] W. J. Lough, I. W. Wainer, *High Performance Liquid Chromatography: Fundamental Principles and Practice*, CRC Press, **1995**.

[46] J. J. Pitt, *Clin Biochem Rev* **2009**, *30*, 19–34.

[47] M. A. Strege, *J Chromatogr B Biomed Sci Appl* **1999**, *725*, 67–78.

[48] K. V Sashidhara, Rosaiah Jammikuntla N, *Nat Prod Commun* **2007**, *2*, 193–202.

[49] D. P. Donnelly, C. M. Rawlins, C. J. DeHart, L. Fornelli, L. F. Schachner, Z. Lin, J. L. Lippens, K. C. Aluri, R. Sarin, B. Chen, C. Lantz, W. Jung, K. R. Johnson, A. Koller, J. J. Wolff, I. D. G. Campuzano, J. R. Auclair, A. R. Ivanov, J. P. Whitelegge, L. Paša-Tolić, J. Chamot-Rooke, P. O. Danis, L. M. Smith, Y. O. Tsybin, J. A. Loo, Y. Ge, N. L. Kelleher, J. N. Agar, *Nat Methods* **2019**, *16*, 587–594.

[50] H. Mei, Y. Hsieh, C. Nardo, X. Xu, S. Wang, K. Ng, W. A. Korfmacher, *Rapid Communications in Mass Spectrometry* **2003**, *17*, 97–103.

[51] L. Elviri, F. Speroni, M. Careri, A. Mangia, L. S. di Toppi, M. Zottini, *J Chromatogr A* **2010**, *1217*, 4120–4126.

[52] M. Holčapek, R. Jirásko, M. Lísa, *J Chromatogr A* **2010**, *1217*, 3908–3921.

[53] R. Silva, N. Lopes, D. Silva, *Planta Med* **2016**, *82*, 671–689.

[54] M. Karas, R. Krüger, *Chem Rev* **2003**, *103*, 427–440.

[55] K. F. Nielsen, M. Månsson, C. Rank, J. C. Frisvad, T. O. Larsen, *J Nat Prod* **2011**, *74*, 2338–2348.

[56] P. H. Dawson, *Quadrupole Mass Spectrometry and Its Applications*, Elsevier, **2013**.

[57] I. V. Chernushevich, A. V. Loboda, B. A. Thomson, *Journal of Mass Spectrometry* **2001**, *36*, 849–865.

[58] H. Carlsson, A. Vaivade, P. Emami Khoonsari, J. Burman, K. Kultima, *Journal of Chromatography B* **2022**, *1195*, 123200.

[59] U. Boesl, *Mass Spectrom Rev* **2017**, *36*, 86–109.

[60] E. N. Nikolaev, Y. I. Kostyukevich, G. N. Vladimirov, *Mass Spectrom Rev* **2016**, *35*, 219–258.

[61] A. C. Leney, A. J. R. Heck, *J Am Soc Mass Spectrom* **2017**, *28*, 5–13.

[62] H. Vu, L. Pedro, T. Mak, B. McCormick, J. Rowley, M. Liu, A. Di Capua, B. Williams-Noonan, N. B. Pham, R. Pouwer, B. Nguyen, K. T. Andrews, T. Skinner-Adams, J. Kim, W. G. J. Hol, R. Hui, G. J. Crowther, W. C. Van Voorhis, R. J. Quinn, *ACS Infect Dis* **2018**, *4*, 431–444.

[63] R. Reher, A. T. Aron, P. Fajtová, P. Stincone, B. Wagner, A. I. Pérez-Lorente, C. Liu, I. Y. Ben Shalom, W. Bittremieux, M. Wang, K. Jeong, M. L. Matos-Hernandez, K. L. Alexander, E. J. Caro-Diaz, C. B. Naman, J. H. W. Scanlan, P. M. M. Hochban, W. E. Diederich, C. Molina-Santiago, D. Romero, K. A. Selim, P. Sass, H. Brötz-Oesterhelt, C. C. Hughes, P. C. Dorrestein, A. J. O'Donoghue, W. H. Gerwick, D. Petras, *Nat Commun* **2022**, *13*, 4619.