Analysis of the Myxobacterial Secondary Metabolome Facilitated by Supercritical Fluid Technologies and Magnetic Resonance Mass Spectrometry

Dissertation

zur Erlangung des Grades

des Doktors der Naturwissenschaften

der Naturwissenschaftlich-Technischen Fakultät

der Universität des Saarlandes

von

Chantal Bader

Saarbrücken

2021

Tag des Kolloquims: 30.04.2021

Dekan: Prof. Dr. Jörn Walter

Berichterstatter: Prof. Dr. Dr. Rolf Müller

Prof. Dr. Uli Kazmaier

Prof. Dr. Angelika Vollmar

Vorsitz: Prof. Dr. Claus-Michael Lehr

Akad. Mitarbeiter: Priv.-Doz. Dr. Martin Frotscher

Für meine Oma Elfriede,
damit ich niemals vergesse, dass ich fliegen kann.

Danksagung

Als erstes möchte ich mich bei meinem Doktorvater Prof. Dr. Dr. Rolf Müller für die Möglichkeit bedanken, meine Doktorarbeit in seinem Arbeitskreis anfertigen zu dürfen. Neben der Bereitstellung des unfassbar umfangreichen analytischen Fuhrparks, hast du mir die Möglichkeit gegeben mich wissenschaftlich weiterzuentwickeln und auch eigene Ideen verfolgen zu können. Ich möchte mich bei dir vor allem für dein mir entgegengebrachtes Vertrauen bedanken, sowie dafür, dass du trotz deines vollen Terminplans immer als Ansprechpartner verfügbar warst.

Des Weiteren möchte ich Prof. Dr. Uli Kazmaier für seine Unterstützung als wissenschaftlicher Begleiter und für die Übernahme des Zweigutachtens bedanken. Die hervorragende Zusammenarbeit mit verschiedenen Mitgliedern deines Arbeitskreises wird für mich immer ein gutes Beispiel dafür sein, dass Wissenschaft keine Konkurrenz, sondern vor allem ein miteinander sein soll.

Außerdem möchte ich mich bei meinen beiden Betreuern Judith Hoffmann und Daniel Krug bedanken. Danke Judith, dass du mich während meiner Diplomarbeit so gut auf die Promotion vorbereitet und mir das von mir während des Studiums so gehasste NMR letztendlich doch verständlich gemacht hast. Daniel, dir möchte ich danken, dass du meine menschliche und technische Weiterentwicklung mit der Übergabe so vieler Geräteverantwortlichkeiten und des Besuchs von internationalen Konferenzen stark vorangetrieben hast. Die wissenschaftlichen Diskussionen mit dir und deine Hilfe beim Anfertigen von Manuskripten waren für mich von unermesslichem Wert. Deine unerschütterlich positive Einstellung wird immer ein Vorbild für mich sein.

Bei meinem gesamten Arbeitskreis und externen Kooperationspartnern möchte ich mich für zahlreiche wissenschaftliche Diskussionen, aber auch das gute menschliche Zusammenspiel bedanken. Besonderer Dank gilt hierbei meinen Bürokollegen Fabian Panter, Nicolas Frank und Patrick Haack. Ohne euch hätte ich nur halb so viele spannende Projekte, aber vor allem auch viel weniger Spaß an meiner Promotion gehabt. Die ausgezeichneten wissenschaftlichen Diskussionen über Agarplattenexperimente waren es definitiv wert gefühlt jede Woche Ärger von den Putzfrauen zu bekommen. Des Weiteren möchte ich mich ganz herzlich bei Michael Hoffmann und David Auerbach bedanken. Ohne euch wäre ich immer noch nicht in der Lage eine Schraube intuitiv richtig herum zu drehen und hätte niemals herausgefunden wieviel Spaß man in und mit der Wissenschaft haben kann. Besonderer Dank gilt auch Laura Franz, Jan Schlemmer, Sebastian Groß, Jake Haeckl und Sebastian Peter. Ihr alle seid nicht nur Arbeitskollegen, sondern Freunde. Einen eigenen Abschnitt möchte ich Christine Walt widmen. Zum einen, weil ich weiß, dass du deswegen auf jeden Fall ausrasten wirst und zum anderen, weil ich es unglaublich schätze wie gut wir zusammenarbeiten. Deine überschwängliche Art ist für mich immer wieder eine Quelle der Motivation und ich bin froh dich zu meinen Freunden zählen zu können.

Zum Schluss möchte ich mich noch bei meinen Eltern und meiner ganzen Familie danken. Danke dafür, dass ihr mich immer auf meinem Weg immer unterstützt habt und mich so nehmt wie ich bin. Ich bin unendlich dankbar für eine Familie wie euch!

Zuallerletzt gilt ein besonderer Dank meinem Dreamteam-partner. Du lässt mich wissenschaftlich und menschlich über mich hinauswachsen und bist und bleibst etwas ganz Besonderes für mich.

Veröffentlichungen 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 der Veröffentlichung.

Bereits veröffentlichte Publikationen

Chantal D. Bader, Markus Neuber, Fabian Panter, Daniel Krug and Rolf Müller: Supercritical Fluid Extraction Enhances Discovery of Secondary Metabolites from Myxobacteria. *Anal Chem.* **2020**, Dec 1;92(23): 15403-15411.

Chantal D. Bader, Patrick A. Haack, Fabian Panter, Daniel Krug and Rolf Müller: Expanding the Scope of Detectable Microbial Natural Products by Complementary Analytical Methods and Cultivation Systems. *J Nat Prod.* **2021**, Jan 15; online ahead of print.

Publikationen in Vorbereitung der Veröffentlichung

Fabian Panter, **Chantal D. Bader** and Rolf Müller: The sandarazols are cryptic and structurally unique plasmid encoded toxins from a rare myxobacterium. *Manuscript accepted by Angewandte Chemie, online preview DOI: 10.1002/anie.202014671.*

Chantal D. Bader, Fabian Panter, Ronald Garcia, Christine Walt, Sibylle Haid, Cathrin Spröer, Alexander Kiefer, Thomas Pietschmann, Jörg Overmann and Rolf Müller: Sandacrabins - Antiviral and Insecticidal Farnesylated Benzimidazoles Produced by a Rare Myxobacterium. *Manuscript in preparation under consideration for patenting.*

Chantal D. Bader, Fabian Panter, Alexander Kiefer, Sibylle Haid, Judith Hoffmann, Thomas Pietschmann and Rolf Müller: Characterization of the Sandaramines, Myxobacterial Alkaloid Terpenoids accessible by Supercritical Fluid Chromatography. *Manuscript in preparation under consideration for patenting.*

Weitere Publikationen, die nicht Teil dieser Arbeit sind

Chantal D. Bader, Fabian Panter and Rolf Müller: In depth natural product discovery – Myxobacterial strains that provided multiple secondary metabolites. *J.Biotechadv.* **2020**, March-April; **39**; 107480.

Alexander Kiefer, **Chantal D. Bader**, Jana Held, Anna Esser, Ran Rybniker, Martin Empting, Rolf Müller and Uli Kazmaier: Synthesis of New Cyclomarin Derivatives and Their Biological Evaluation towards *Mycobacterium Tuberculosis* and *Plasmodium Falciparum*. *Chem. Eur. J.* **2019**, July 3; **25**(37); 8894-8902.

Tadeja Lukežič, Antoine Abou Fayad, **Chantal Bader**, Kirsten Harmrolfs, Johannes Bartuli, Sebastian Groß. Urška Lešnik, Fabienne Hennessen, Jennifer Herrmann, Špela Pikl, Hrvoje Petković and Rolf Müller: Engineering Atypical Tetracycline Formation in *Amycolatopsis sulphurea* for the Production of Modified Chelocardin Antibiotics. *ACS Chem. Biol.* **2019**, February 12; **14**(3); 468-477.

Rolf Müller, Tadeja Lukežič, Maja Remškar, Nestor Zaburannyi, **Chantal Bader**, Asfandyar Sikandar, Jesko Köhnke (**2018**). *Gene cluster for the biosynthetic production of tetracycline compounds in a heterologous host* (PCT/EP2018/086740). European Patent Office. https://patentscope.wipo.int/search/en/detail.jsf?docId=WO2019122400&tab=PCTBIBLIO

Joachim J. Hug, **Chantal D. Bader**, Maja Remškar, Katarina Cirnski and Rolf Müller: Concepts and Methods to Access Novel Antibiotics from Actinomycetes. *Antibiotics.* **2018**, May 22, **7**(2); 44.

Henni-Karoliina Ropponen, **Chantal D. Bader**, Eleonora Diamanti, Boris Illarionov, Matthias Rottmann, Markus Fischer, Matthias Witschel, Rolf Müller and Anna K. H. Hirsch: Unravelling the Active Ingredients from a 2-Aminothiazole DMSO Stock Solution with Antimalarial Activity. *Manuscript submitted to ChemMedChem.*

Ram Awal, Patrick A. Haack, **Chantal D. Bader**, Cornelius Riese, Dirk Schüler and Rolf Müller: Sesbanimide R, a novel cytotoxic polyketide produced by magnetotactic bacteria. *Manuscript submitted to mBio.*

Fabian Panter, **Chantal D. Bader** and Rolf Müller: Synergizing the potential of bacterial genomics and metabolomics to find novel antibiotics. *Manuscript submitted to ChemSci.*

Chantal Bader, Matthias Witt, Aiko Barsch, Patrick Haack, Daniel Krug and Rolf Müller. MRMS aXelerate for targeted metabolomics profiling of myxobacterial extracts. *Application Note.* https://lcms.labrulez.com/paper/7532

Reem K. Fathalla, Wolfgang Fröhner, **Chantal D. Bader**, Mohammad Abdel-Halim, Rolf Müller, Christian Ducho and Matthias Engel: Diarylpyrrolidinediones as novel, non-substrate competitive inhibitors of bacterial UDP-N-acetylglucosamine 1-carboxyvinyltransferase (MurA) - biological evaluation and mechanistic studies. *Manuscript in preparation.*

Lucija Slemc, Martina Avbelj, Martin Šala, **Chantal D. Bader**, Špela Pikl, Rolf Müller and Hrvoje Petković: Function of OxyS enzyme in C5 and C6 hydroxylation in tetracycline biosynthesis. *Manuscript in preparation.*

Tagungsbeiträge

Chantal Bader (**2020**) Purification of myxobacterial natural products supported by supercritical CO_2. (Oral communication) *Waters SFC/UPC^2 virtuelles Anwendertreffen*

Chantal Bader, Patrick Haack, Fabian Panter, Matthias Witt, Daniel Krug and Rolf Müller (**2019**) Magnetic resonance mass spectrometry profiling of myxobacterial extracts – higher resolution, deeper insights? (Poster presentation) *67th Conference of the America Society for Mass Spectrometry (ASMS) Atlanta, United States of America.*

Chantal Bader, Markus Neuber, Daniel Krug and Rolf Müller (**2019**) Supercritical fluid extraction for myxobacterial secondary metabolomics. (Oral communication) *Vereinigung für Allgemeine und Angewandte Mikrobiologie (VAAM) Workshop, Jena, Germany.*

Chantal Bader, Fabian Panter, Alexander Kiefer, Judith Hoffmann, Ronald Garcia and Rolf Müller (**2019**) Characterization and (bio-)synthesis of the Sandaramines – novel myxobacterial terpenes with activity against *Pseudomonas aeruginosa* (Oral communication) *VIII. Symposium of the Interdisciplinary Graduate School of Natural Product Research, Saarbrücken, Germany*

Chantal Bader, Patrick Haack, Fabian Panter, Daniel Krug und Rolf Müller (**2018**) DI-FTICR and LC-qTOF for the comprehensive profiling of myxobacterial secondary metabolomes (Oral communication and poster presentation) *Vereinigung für Allgemeine und Angewandte Mikrobiologie (VAAM) Workshop, Frankfurt am Main, Germany.*

Chantal Bader, Patrick Haack, Fabian Panter, Daniel Krug and Rolf Müller (**2018**) DI-FTICR and LC-qTOF for the comprehensive profiling of myxobacterial secondary metabolomes (Poster presentation) *3rd European Conference on Natural Products, Gesellschaft für Chemische Technik und Biotechnologie e.V. (DECHEMA), Frankfurt am Main, Germany.*

Chantal Bader, **Patrick Haack**, Fabian Panter, Matthias Witt, Daniel Krug and Rolf Müller (**2018**) DI-FTICR and LC-qTOF for the comprehensive profiling of myxobacterial secondary metabolomes (Poster presentation) *European Mass Spectrometry Conference (EMSC), Saarbrücken, Germany*

Chantal Bader, Patrick Haack, Fabian Panter, Matthias Witt, Daniel Krug and Rolf Müller (**2018**) DI-FTICR and LC-qTOF for the comprehensive profiling of myxobacterial secondary metabolomes (Poster presentation) *14th Annual Conference of the Metabolomics Society (Metabolomics), Seattle, United States of America.*

Chantal Bader, Patrick Haack, Fabian Panter, Matthias Witt, Daniel Krug and Rolf Müller (**2018**) Comparison of DI-FTICR and UPLC-qTOF for the comprehensive profiling of myxobacterial secondary metabolomes (Poster presentation) *8th Helmholtz Institut für Pharmazeutische Forschung Saarland (HIPS) Symposium, Saarbrücken, Germany*

Chantal Bader, Patrick Haack, Fabian Panter, Matthias Witt, Daniel Krug and Rolf Müller (**2018**) DI-FTICR and UPLC-qTOF for the comprehensive profiling of myxobacterial secondary metabolomes (Poster Presentation) *VII. Symposium of the Interdisciplinary Graduate School of Natural Product Research, Saarbrücken, Germany*

Chantal Bader, Jennifer Herrmann, Peer Lukat, Angela Kling and Rolf Müller (**2017**) (Re-) Discovery of compounds targeting DnaN (Poster presentation) *Vereinigung für Allgemeine und Angewandte Mikrobiologie (VAAM) Workshop, Tübingen, Germany.*

Chantal Bader, Jennifer Herrmann, Peer Lukat, Angela Kling and Rolf Müller (**2017**) (Re-) Discovery of compounds targeting DnaN. (Oral communication) *Summer School der European Society of Clinical Microbiology and Infectious Diseases (ESCMID), Borstel, Germany.*

Chantal Bader, Jennifer Herrmann, Peer Lukat, Angela Kling and Rolf Müller (**2017**) (Re-) Discovery of compounds targeting DnaN – A new target for antibiotic development (Poster presentation) *VI. Symposium of the Interdisciplinary Graduate School of Natural Product Research, Saarbrücken, Germany*

Zusammenfassung

Myxobakterien sind eine wertvolle Quelle für biologisch aktive und strukturell diverse Naturstoffe (NP). Neben Steigerung der NP-Produktion durch Aktivierung ihrer Biosynthesemaschinerie wurde in dieser Arbeit der Einfluss von unterrepräsentierten Techniken in diversen Schritten des myxobakteriellen NP-Entdeckungsprozesses evaluiert. Dabei wurden bakterielle Extrakte durch Direktinjektion an einem Fourier-Transformations-Ionenzyklotron-Massenspektrometer (DI-FTICR), sowie nach Extraktion und Chromatographie mit überkritischem CO_2 (SFE & SFC) analysiert. Die DI-FTICR-Analyse zeigte, dass eine immense Menge an NPs nicht detektierbar bleibt, wenn man die Nachweismethode auf gängige Methoden beschränkt. SFE erhöhte die biologischen Aktivitäten der Extrakte und ermöglichte die Isolierung eines neuen NP, der Chloroxanthinsäure. SFC zeigte überlegene Eigenschaften für die Isolierung von permanent geladenen NPs und wurde erfolgreich für die Isolierung natürlicher und synthetischer Sandaramine und Sandacrabine eingesetzt. Diese beiden NP-Familien werden von der neuen Spezies *Sandaracinus* MSr10575 produziert. Die Genomanalyse des Stammes zeigte ein Plasmid-kodiertes Biosynthese-Gencluster, dessen molekularbiologische Aktivierung die Sandarazol-Familie hervorbrachte. Alle drei aus *Sandaracinus* MSr10575 isolierten NP-Familien wiesen potente biologische Aktivitäten auf, wobei insbesondere die antivirale Aktivität der Sandaramine gegen das SARS-CoV2-Modellvirus CoV229E hervorzuheben ist.

Abstract

Myxobacteria are a valuable source of biologically active and highly diverse natural products (NP). Besides enhancing NP production by activating the underlying biosynthetic machinery, this thesis evaluates the impact of underrepresented techniques at several steps of the myxobacterial NP discovery workflow. Herein, direct injection on a Fourier transform ion cyclotron mass spectrometer (DI-FTICR), as well as supercritical fluid extraction and chromatography (SFE & SFC) were used to analyze and purify bacterial extracts. The DI-FTICR analysis revealed, that an immense number of secondary metabolites remains untapped when limiting the detection method to the typical analytical setups. SFE significantly enhanced detectable biological activities of the bacterial crude extracts and supported isolation of a new NP, chloroxanthic acid A. Finally, SFC showed superior properties for the isolation of permanently charged NPs and was successfully applied for the isolation of natural and synthetic sandaramines and sandacrabins. Those two NP-families are produced by the new myxobacterial species *Sandaracinus* MSr10575. The strains' genome analysis moreover revealed a plasmid-encoded biosynthetic gene cluster, whose genetic activation yielded the sandarazol family. All three NP families isolated from *Sandaracinus* MSr10575 exhibit potent biological activities, particularly the antiviral activity of the sandaramines against the SARS-CoV2 model virus CoV229E.

Table of Contents

<div align="center">

Chapter 1

</div>

Introduction

During the recent decades, antimicrobial resistance (AMR) of human pathogens against antibiotics and antifungal agents on the market is steadily increasing. Especially pathogens of the "ESKAPE"-panel, which includes *Escherichia coli*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumanii*, *Pseudomonas aeruginosa* and *Enterobacter* sp. are known to be more and more resistant to the existing antibiotics.[1] Concerning this trend, a switch in antibiotic usage as well as the development of new substances targeting these pathogens is unequivocally regarded as urgently needed.[2] Most antibiotics have been developed between 1950 and 1980. This era is also called the golden age of antibiotics, as 10 different classes of antibacterial substances were developed and brought to the market during that time.[3] Figure 1 summarizes main antibiotic classes on the market with their time of development, highlighting the innovation gap in this field.
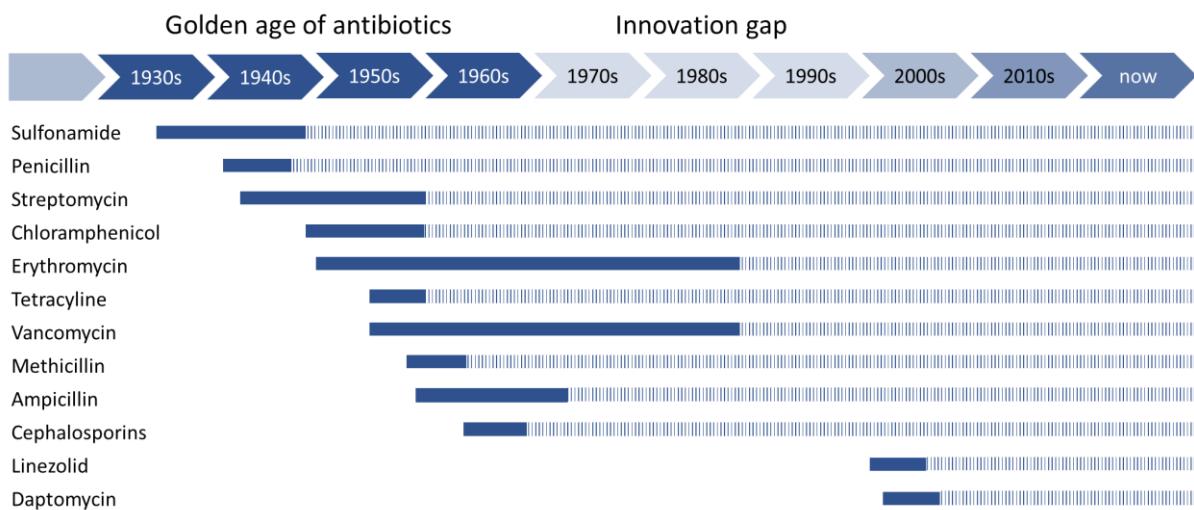
Figure 1 Development (blue bars) as well as resistance development (dashed bars) of current antibiotics.[4]

During the golden age of antibiotics, various natural products with intriguing antimicrobial activities were isolated and their structure elucidated, but neglected afterwards due to deficiencies in their pharmacokinetic or -dynamic properties.[5] In the 1960s, the market for antibiotics had reached a saturation point, having a good selection of antibiotics at hand to treat most diseases, which led to a decrease in public interest in antibiotics discovery. In the following years, no novel antibiotics were brought to the market as the problem of bacterial infections seemed to be solved. Recent pandemics such as the ebola or the covid-19 pandemic may also rise the impression, that bacterial infections only play an underrepresented role compared to viral infections. However, it should not be forgotten that such viral pandemics also affect the antibiotics resistance situation, as administration of antibiotics is often necessary to treat co-infections or secondary bacterial infections. For covid-19 for example, 71.9% of the hospitalized patients received antibiotics.[6] Even antibiotics such as Vancomycin that were kept as reserve for the treatment of infections with multiresistant pathogens have therefore decreased in efficiency, as they need to be administered more and more frequently.[7] The intensified development of AMR at the end of the 20th century led to a refocusing on antibiotic discovery, which however has not been fully redeveloped yet.[4] Most of the antibiotics shown in Figure 1 such as β-lactams,

tetracyclines and macrolides are derived from natural product scaffolds.[8] The number of isolated bioactive natural products isolated exceeds the number of antibiotics on the market derived from natural product scaffolds by far, which is exemplified by the isolation of approximately 2250 bioactive NPs from rare actinomycetes until 2005 alone.[9]

1.1 Natural products

Natural products (NP) are defined as organic molecules derived from secondary metabolism of a living organism. In contrast to the essential primary metabolism of a cell, which is dedicated to growth, development and reproduction, the secondary metabolism is involved in the long-term preservation of the organism's survivability.[10] In this context, NPs (or also called secondary or specialized metabolites) can act as communication molecules to attract other organisms for example as coloring agents and fragrances, competitive weapons in particular antibiotics, repellants and toxins or transport molecules like siderophores.[11] The diversity of their biological functions correlates perfectly with the chemical diversity of NP and the broad spectrum of NP-producing organisms.[12]

1.1.1 Natural product producers

NPs are either produced by prokaryotic organisms (bacteria and archaea) or eukaryotic organisms such as fungi, plants or even animals.[13] Animal-derived natural products most commonly have been described in venomous animals such as snakes, spiders or bees. However, not all animal-derived NPs are directly synthesized by the animal itself, they often take up and store NPs from other organisms like microbes or plants.[14] As most animals secret their NPs to defend themselves against higher developed organisms and not against microbes, they often show toxic effects on humans.[15] Those toxic NPs address highly specific macromolecular targets in the human body. The interaction between the NP and their molecular target have proven to be useful to study receptors, ion channels and enzymes, but also to develop drugs based on that interaction.[16] One example where an animal NP has been successfully used to develop a novel class of drugs is the nonapeptide teprotide, which was isolated from the snake *Bothrops jararaca* (see Figure 2).
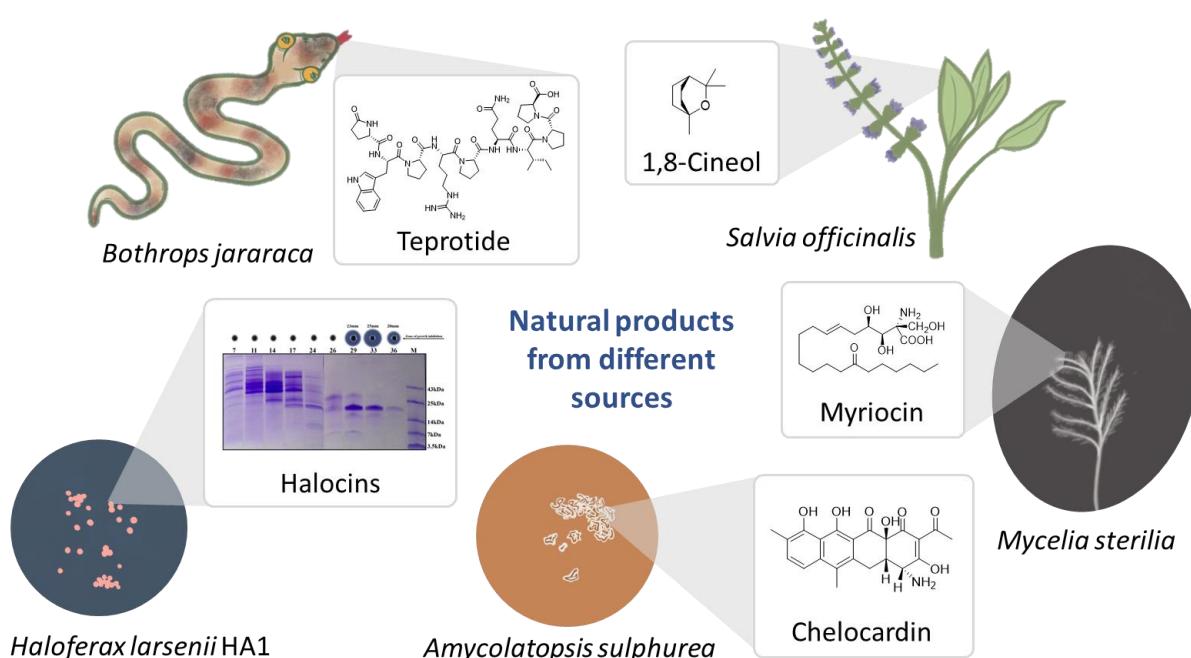

Figure 2 Examples for natural product producers and compound classes found from these.[17–21]

During the investigation of teprotides' pharmacological activities, Bianchi *et al*. observed an inhibition of the angiotensin converting enzyme (ACE).[17] Although teprotide showed good activity, its usage as drug was hindered by low oral bioavailability and cost expense. Further investigations of the teprotide binding site allowed the design of the approved drug captopril which became a founding member of numerous ACE inhibitors employed in human therapy of high blood pressure.[22]

Contrary to animals, plants have to defend themselves against herbivores, but they also benefit from higher organisms as seed distributors, making many plant natural products non-toxic for humans.[23] Amongst the NP producers, plants have been used for medical purposes for the longest time .[24] Their application ranges from headache and cough to viral and bacterial infections, as well as cancer. Commonly, plant NPs are not used as purified drugs, but rather as plant extract containing a mixture of pharmaceutically active substances.[25] Several applications benefit from the use of these mixtures, as different ingredients affect different symptoms of one disease. *Salvia officinalis* L. for example mainly contains essential oils like thujon and 1,8-cineol (Figure 2), as well as phenolic compounds such as apigenin or caffeic acid.[18] It is commonly used to treat the common cold, where the patient benefits from the monoterpene 1,8-cineol exhibiting expectorant, antibiotic, antiviral and anti-inflammatory effects. Furthermore, the phenols contained in *S. officinalis* extracts lead to improvement of inflammations in mouth and throat.[26] Even though medicinal plants are often well studied, there are still novel biological activities discovered in recent years exemplified by the use of *S. officinalis* L. in oncology and neurodegenerative diseases. In this context, researchers have shown that the composition of the extracts including different types of secondary metabolites and accordingly their medical use, depends on the extraction method chosen for processing the plant material.[18]

When compared to the long-studied plants, microorganisms more recently came into focus for natural product discovery. Among them, Archaea represent the least studied class of NP producers. Their ability to survive in extreme environments, such as hot springs, the dead sea or extreme pH conditions on the one hand make them an interesting target for studying their highly adapted secondary metabolism, but also limit their accessibility as well as cultivability on the other hand.[27] In these extreme environments, they profit from a situation with few competitors being present, resulting in a proportionally low tendency to produce secondary metabolites.[28] The archaeocins, such as halocin depicted in Figure 2, are a group of bacteriocin antibiotic peptides and the diketopiperazines, small cyclic NRPs, are rare examples of small molecule NPs found in archaea.[19,26] Contrary to Archaea, most fungi and bacteria live in a very competitive environment. Consequently, it is advantageous for them to produce a variety of antimicrobial NPs as defense mechanism against competitors.[29,30] The probably best-known fungus is *Penicillium chrysogenum,* from which penicillin was discovered in 1928 as one of the first antibiotics.[31] Fungal natural products show a high structural diversity, as well as interesting biological activities besides antimicrobial properties.[30] The unusual amino acid myriocin for example was isolated from the thermophilic fungus *Mycelia sterilia*.[21] It served as a lead structure to develop fingolimod, which was approved by the FDA in 2010 for the treatment of multiple sclerosis.[32] The vast fungal kingdom consisting of approximately 1.5 million species, besides the possibility to grow fungi in a laboratory in various scales and their accessibility to genetic manipulations emphasize fungi as an interesting target for discovering bioactive agents.[33]

Bacterial natural products research also makes use of those three properties (high number of representatives, growth under laboratory conditions, genetic manipulability), that are shared between fungi and bacteria.[34] As bacteria often live in symbiotic relationships e.g. in the human gut, they do not only produce NPs for their own defense against predators or to compete for nutritional sources, but

also to defend and interact with their hosts. One recent example, where such a symbiotic interaction was used to discover a new antibiotic is darobactin. This peptide natural product is synthesized by a *Photorhabdus* sp. living in the digestive systems of entomopathogenic nematodes, to prevent the host nematode from bacterial infections.[35] *Photorhabdus*, however, is one of the less investigated bacterial species.[36] Most NPs that were discovered to date stem from actinobacteria especially *Streptomycetes* or firmicutes such as *Bacilli*.[37] Examples for natural products discovered from *Streptomycetes* and Bacilli are the broad-spectrum antibiotic chelocardin, which is produced by *Amycolatopsis sulphurea*[20] or the Class II Lantibiotic Pseudomycoicidin from *Bacillus pseudomycoides*.[38] Nevertheless, rediscovery rate of already known NPs in intensively investigated species like *Bacilli* and *Streptomycetes* is high, wherefore underexploited bacteria such as myxobacteria gain more and more of interest. [39]

Myxobacteria

Myxobacteria are ubiquitous, soil-living bacteria. They belong to the Gram–negative bacteria and phylogenetically classify amongt the δ-proteobacteria.[40] Besides their fascinating morphological behavior, which includes swarming, predating and the formation of fruiting bodies, they belong to the bacteria with the largest genomes.[41] In those extraordinary large genomes with up to 16 MBp, myxobacteria feature a variety of biosynthetic gene clusters (BGC), which encode biochemical pathways for the production of diverse secondary metabolites.[42] One of the most widely known examples of a secondary metabolite produced by a myxobacterium is epothilone, an antineoplastic NP produced by different *Sorangium cellulosum* strains, which inspired the semisynthetic anti-cancer drug ixabepilone.[43] *S. cellulosum* is grouped among the *Sorangiineae*, one of the only three myxobacterial suborders besides *Cystobacterineae* and *Nannocystineae*.[44] It represents - to date - one of the most prolific myxobacterial secondary metabolite producing species. However, it is hardly surprising that *Sorangium* provided most secondary metabolites so far, as this genus is highly overrepresented in most myxobacterial strain collections.[37] This myxobacteria isolation bias was presumably introduced by the discovery of the epothilones, which raised hope to find more such promising NPs from this genus. Nevertheless, Hoffmann *et al.* highlighted that the chances for novel compound discovery are higher, when investigating taxonomically distant myxobacteria.[45] Many screening approaches have intuitively relied on this finding beforehand and focused on the isolation of novel bacterial strains to identify previously unknown secondary metabolites. These strains are usually characterized concerning their metabolome and biological activities of their crude extracts. Although the strains are well characterized in these first initial screening steps, the characterization of only one NP family per strain remains a standard.[37] The biosynthetic potential of myxobacteria on the other hand is far from exploited and neglecting a strain after the discovery of a single NP family appears to miss tremendous opportunity for the discovery of additional new natural products, considering their biosynthetic potential.[46] Table 1 summarizes myxobacterial strains, which have been investigated in depth so far and provided multiple secondary metabolites including their classification on suborder and family level, underlining this hypothesis.

Table 1 Examples for myxobacterial strains, their classification and natural products they produce adapted from Bader *et al.*. [37]

Strain	Classification	NPs produced by the strain
Myxococcus xanthus **DK1622**	*Cystobacterineae; Myxoccaceae*	DKxanthenes[47], myxovirescins*, myxalamids*, myxochelins*, cittilins*, myxochromids*[48], myxoprincomids[49], homospermidin lipids[50], alkylpyrones*[51]

Sorangium cellulosum So ce12	*Sorangiineae; Polyangiaceae*	Sorangicins[52], disorazoles[53], chivosazol[54], sorangiolids[55], sulfangolids[56]
Sorangium cellulosum So ce56	*Sorangiineae; Polyangiaceae*	Etnangien*, chivosazoles*[57], myxochelins*[58], eremophilene[59]
Sorangium cellulosum So ce481	*Sorangiineae; Polyangiaceae*	Lanyamycin, soraphens*[60]
Sorangium cellulosum So ce895	*Sorangiineae; Polyangiaceae*	Thuggacins[61], sorazinones[62]
Sorangium cellulosum So ce1525	*Sorangiineae; Polyangiaceae*	Chlorotonil[63], sulfangolid, chivosazoles*, disorazoles*, sorangicins*, soraphens*, sorangilides*[56]
Stigmatella aurantiaca Sg a15	*Cystobacterineae; Archangiaceae*	Stigmatellin, myxalamids*[64], aurachins[65], volatiles like methyltridecan-4-one[66], myxochelins*,myxochromides*, DKxanthenes*, rhizopodin*[67]
Stigmatella aurantiaca DW4/3-1	*Cystobacterineae; Archangiaceae*	Dawenol, myxothiazol*[68], melithiazol*[69], myxochromides**[70], aurafuron[71], volatiles like methyltridecan-4-one[66]
Stigmatella aurantiaca MYX-030	*Cystobacterineae; Archangiaceae*	Myxocoumarins, myxothiazols*, aurachins*[72]
Chondromyces crocatus Cm c5	*Sorangiineae; Polyangiaceae*	Chondramides[73], ajudazols[74], chondrochlorens[75], dialkylpyrazines*[76], crocapeptins[77], crocagins[78], crocadepsins, thuggacins*, crocacins*[79]
Corallococcus coralloides B035	*Cystobacterineae; Myxoccaceae*	Corallorazines, corallopyronins*[80]
Enhygromyxa salina SWB007	*Nannocystineae; Nannocystacece*	Salimabromides[46,81], enhygrolides[82]
Labilithrix luteola DSM 27648	*Sorangiineae; Labilitrichaceae*	Labindols, 9H-carbazole*, 3-chloro-9H-carbazole*, 4-hydroxymethylquinolone*, 3,6-dibenzylpyrazin-2(1H)-one*[60]
Haliangium ochraceum DSM 14365	*Nannocystineae; Kofleriaceae*	Haliangicins[83,84], haliamides**[46]
Pyxidicoccus fallax An d48	*Cystobacterineae; Myxoccaceae*	Tubulysins[85,86], pyxidicyclins[87]
Hyalangium minutum NOCB-2T (DSM 14724)	*Cystobacterineae; Archangiaceae*	Hyaladione[88], hyafurones, hyapyrrolines, hyapyrones[89]
Cystobacter fuscus MCy9118	*Cystobacterineae; Archangiaceae*	Cystomanamides[90], macyranones, althiomycin*, roimatacene*, myxochelins*[91]
Nannocystis pusilla Ar i7	*Nannocystineae; Nannocystaceae*	Pyrronazols, 1-hydroxyphenazin-6-yl-α-d-arabinofuranoside[92]

* Compounds originally isolated from another strain, but also produced by the strain mentioned.

** Compound family originated from another strain, but new derivatives isolated from the strain mentioned.

1.1.2 Chemical and biosynthetic diversity of myxobacterial natural products

The natural products described in table 1 cover a large chemical space, which can be subdivided into four main classes according to the biosynthetic origin of the NPs. Those four principal classes are polyketides (PKs), non-ribosomal peptides (NRPs), ribosomally produced and posttranslationally modified peptides (RiPPs) and terpenes.[40] Two examples for myxobacterial secondary metabolite structures deriving from each class are shown in Figure 3. Each type of biosynthetic pathway and its underlying enzymatic machinery will be described in detail in the following chapters.

1 Polyketide synthase (PKS) NPs

KS AT KR ACP

Sorangicins

Sulfangolids

2 Nonribosomal peptide synthetase (NRPS) NPs

A PCP C

Crocapeptins

Myxoprincomides

3 Ribosomally synthesized and posttranslationally modified (RiPPs) NPs

Y Y I Y

Cittilins

Crocagins

4 Terpenes

Cystodienoic acid

Germacradienol
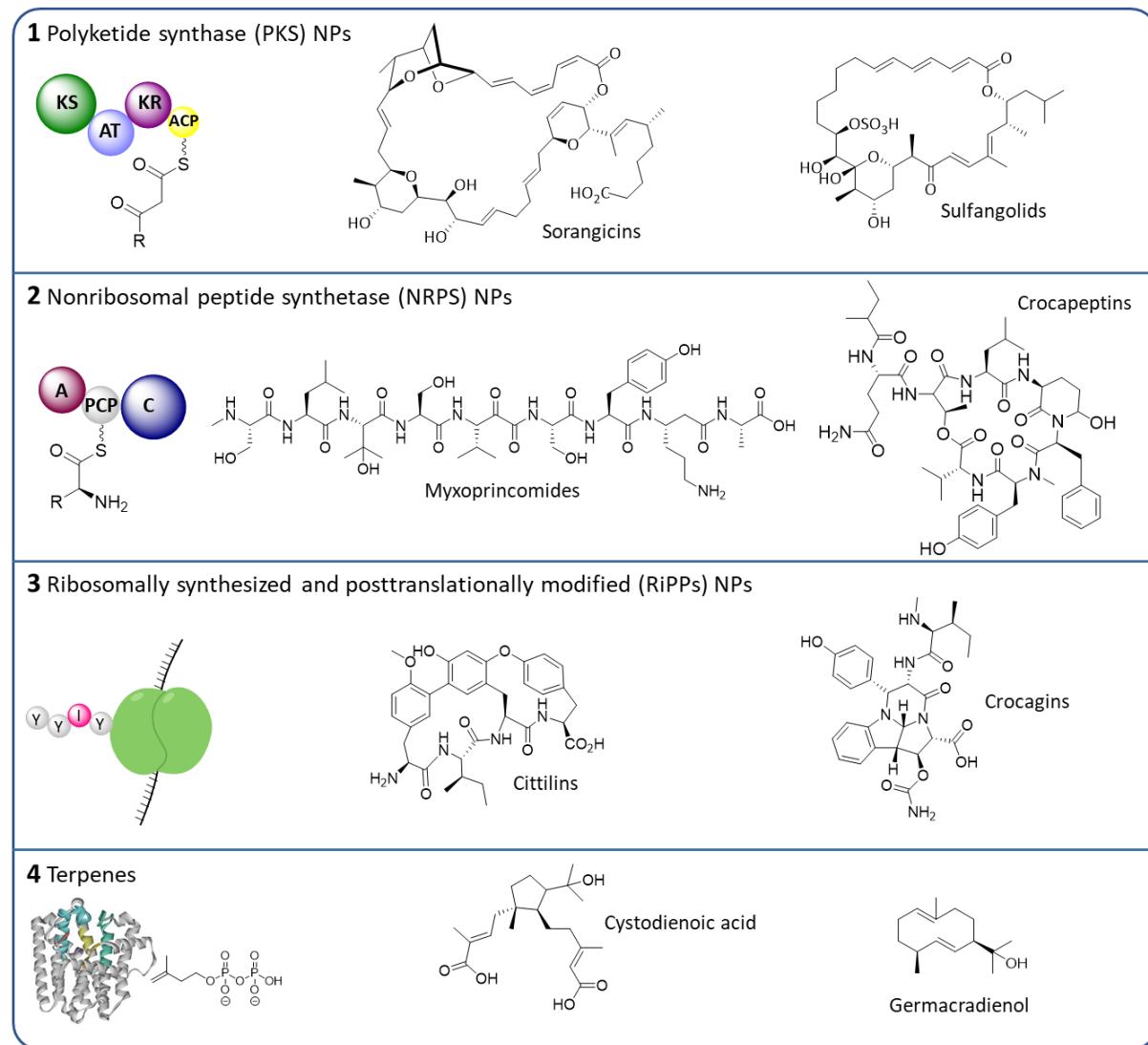
Figure 3 Classification of bacterial natural products by their biosynthetic origin with examples for each class that were found to be produced by myxobacteria.[49,56,77,78,93–97]

Polyketides

Most natural products isolated to date from myxobacteria are classified as PKs or NRPs.[98] Among the polyketide NPs, many promising biological activities can be found. The sorangicins for example, polyketides isolated from a *S. cellulosum* strain, show an inhibition of the bacterial RNA polymerase and are therefore active against Gram–positive bacteria with MIC values lower than 0.1 µg/mL.[52] In addition to the antibacterial activities, cytotoxic activities are often found for PKs. One example for highly cytotoxic secondary metabolites are the stigmatellins. Those PKs isolated from a *Stigmatella aurantiaca* strain bind to the cytochrome bc-1 complex, which leads to inhibition of the respiratory

chain and subsequent cell death.[99,100] Structure elucidation of polyketides, however is challenging, as besides many stereocenters, they often feature highly reactive functional groups such as epoxides, polyenes or lactones, leading to instability during the purification process. Sulfangolids for example were already mentioned in the publication describing the chivosazoles in 1997, but complete structure elucidation could only be achieved in 2012 due to their instability.[56,101] In general, there are three types of polyketide synthases (type I-III). In type I PKS systems, the biosynthetic enzymes are organized as multimodular megaenzymes. Those enzymes consist of modules, which are further organized in domains, the enzymatically active subunits of the megasynthase. Each domain in a type I PKS system has a specific function during the biosynthesis of a new molecule (see Figure 4).[102–105]
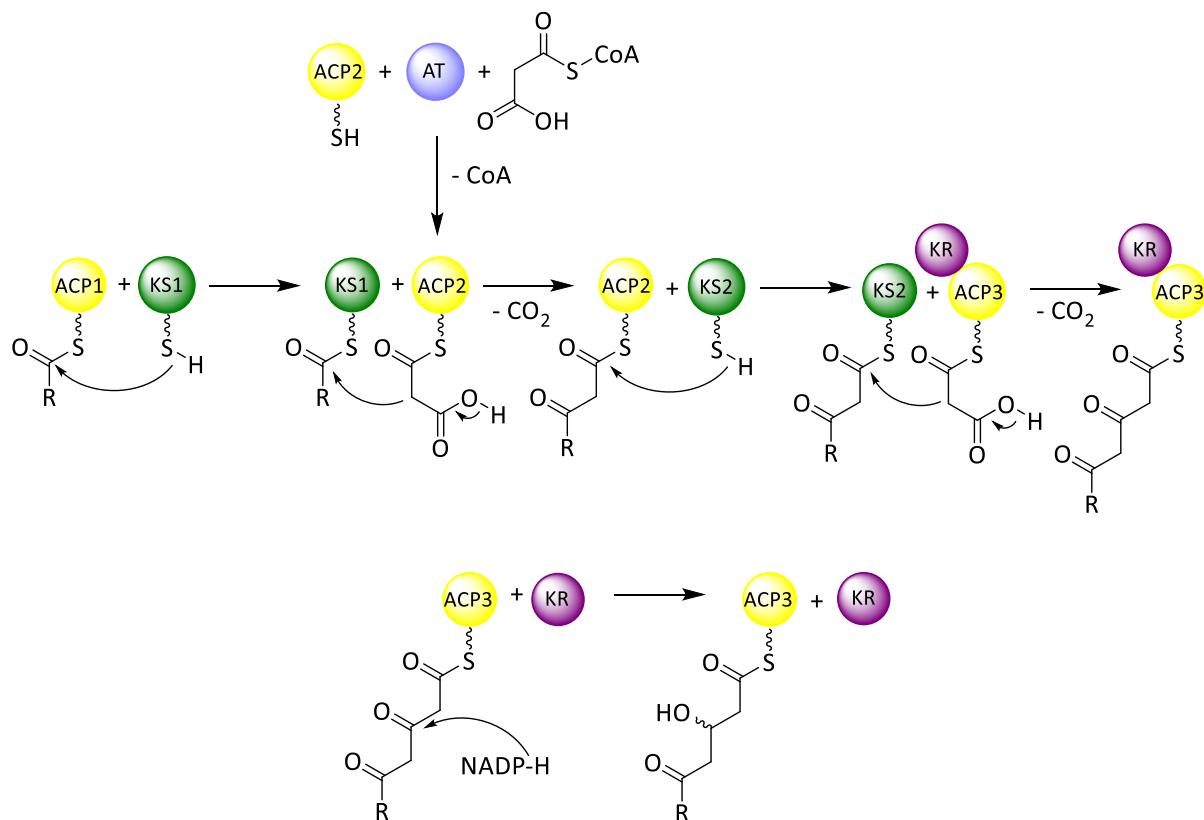
Figure 4 Schematic illustration of the loading process of malonyl-CoA to an ACP by the AT, as well as the elongation process of a type I PKS system by two subunits. Subunit one: Elongation via Claisen condensation without further modifications, Subunit two: Elongation via Claisen-condensation and subsequent reduction to an alcohol by a KR.[105,106]

The acyl carrier protein (ACP) is responsible for the binding of the molecule to the megaenzyme complex. This enzyme requires an acyltransferase (AT) domain for loading of the starter unit to its phosphopantetheinyl (Ppant) arm, as well as the passing on of the elongated molecule to the Ppant arm of the next ACP. The elongation process is performed by a ketosynthase (KS) domain, which typically uses malonyl-CoA for C_2-unit elongation of the molecule. However, alternative extender units such as methylmalonyl-CoA or methoxymalonyl-CoA may also be incorporated. The extender unit used is hereby determined by specificity of the AT-domain. After each elongation step, modifications are directly introduced by different domains. Consequently, an alcohol function can be introduced by a ketoreductase (KR) domain. The resulting alcohol can be dehydrated to a double-bond by a dehydratase (DH) domain, which can optionally be hydrogenated by an enoyl-reductase (ER) domain forming a saturated C_2-unit. The interaction of these different tailoring steps assembled in different module compositions alone enables a broad polyketide NP product spectrum.[107]

In addition to that, the iterative-working type II and III systems broaden the diversity of PK secondary metabolites. Those systems consist of a so-called minimal PKS system. Type II PKS systems are just assembled by three domains: one KSα and one ACP, which work iteratively and a chain length factor (CSL), which controls the total length of the synthesized polyketide chain. All tailoring steps in type II systems are performed on the initially synthesized 1,3 polyketone backbone.[108] One common tailoring reaction in type II PKS products is cyclization, resulting in polyaromatic NPs such as the chelocardins, which were isolated from a *Streptomyces* strain.[20] The also iteratively working type III PKS systems do not even require an ACP domain, as they can act directly on acyl CoA substrates. In contrary to the PKS type II systems, in those systems a chalcone/stilbene synthase (CHS) is responsible for the polyketide chain formation. PKS type III products often are polyaromates, which are formed by cyclisation of the primarily formed 1,3 polyketone backbone.[109,110]

Nonribosomal peptides

Non ribosomal peptide (NRP) natural products represent the second largest group of secondary metabolites besides the polyketide NPs. They are also synthesized by multimodular megaenzymes, the nonribosomal peptide synthetases. The key NRP biosynthesis domains are the adenylation (A) domain, the peptidyl carrier protein (PCP) and the condensation (C) domain, as represented in Figure 5. The first amino acid of the peptide backbone is converted into an AMP ester by the A domain, which is subsequently loaded onto the phosphopantetheinyl arm of the PCP domain. Two amino acids bound to PCP domains can then be condensed by the C domain of the NRPS resulting in a PCP loaded dipeptide. After further elongation steps, the final peptide is released by a thioesterase (TE) or a reduction (Red) domain.[111,112] For elongation non-canonical amino acids can also be used, as only the presence of a carboxylic acid and an amino group in the building block are necessary for the elongation reaction.[113]
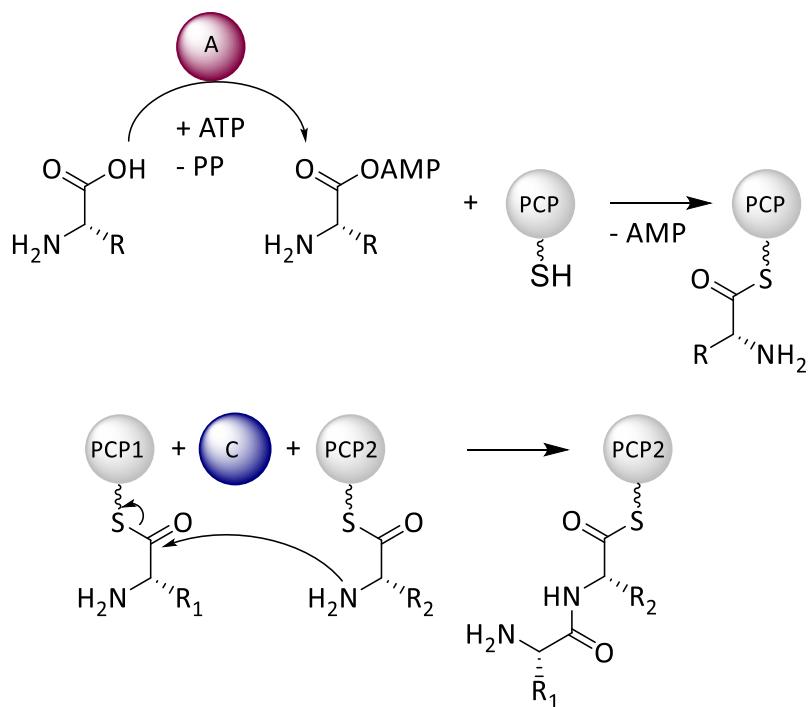
Figure 5 Schematic illustration of the amino-acid loading process by the A domain onto the Ppant of the PCP domain, as well as the peptide coupling of the PCP loaded amino acids by the C domain. Wavy line represents the Ppant arm.[111,112]

Some secondary metabolites remain as linear peptides exemplified by the myxoprincomides[49], but often NRPS products are cyclized during their release via a TE domain or in a post NRPS tailoring reaction.[114] One example for a cyclized NRPS product are the crocapeptins, which were isolated from *Chrondromyces crocatus* Cm c5.[77] The modular NRP biosynthesis incorporating various amino acids combined with the possibility for tailoring reactions result in a variety of diverse chemical structures, whereof many feature intriguing biological activities.[98] The crocapeptins for example were found to be serine protease inhibitors.[77] Other cyclic peptides such as the cyclomarins showed good activities against *Mycobacterium tuberculosis* and *Plasmodium falciparum,* making them promising candidates in the drug development pipeline for tuberculosis and malaria.[115]

Ribosomally produced and posttranslationally modified peptides

Another group of peptide natural products are the ribosomally synthesized and posttranslationally modified peptide (RiPP) NPs. As already indicated by their name, their peptide backbone (also called precursor peptide) is synthesized by the ribosome instead of an NRPS megaenzyme complex. This precursor peptide consists of a core peptide, which is later on incorporated in the NP itself and a leader peptide and/or a follower peptide, which are necessary for recognition of the peptide by modifying enzymes or for cellular transport (see Figure 6).[116] The chemical diversity of RiPPs is, on one hand, generated by their highly variable molecular weight, including rather small peptides like the cittilins or crocagins but also extremely large peptides for instance the lantibiotic nisin incorporating 34 amino acids.[117] On the other hand, the variations introduced by the posttranslational modification steps are diverse. Common posttranslational modifications carried out here are dehydrations, cyclodehydrations, prenylations, oxidative crosslinking, and cyclizations.[116,118] After introduction of these modifications, the cleavage of the leader peptide results in the mature RiPP.[119] Although relatively few RiPPS products are described from myxobacteria yet, representatives isolated from *Streptomyces* showed interesting biological activities, such as the antibiotic bottromycin which is an inhibitor of bacterial polysomes.[120]
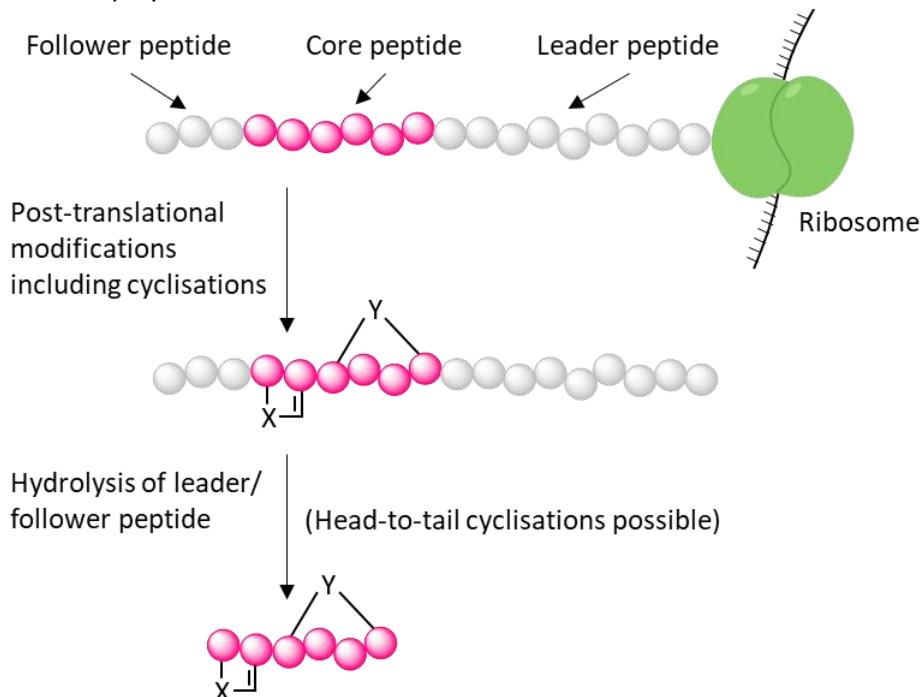
Figure 6 Schematic illustration of the biosynthesis of ribosomally synthesized and posttranslationally modified peptides.[116]

Terpenoid natural products

Finally, myxobacteria are also able to produce a variety of terpenoid secondary metabolites, also called isoprenoids.[66] Similar to RiPPs, terpenes are underrepresented amongst the natural products described from myxobacteria. One of the reasons for this may be, that terpenes are often poorly ionizable with electrospray ionization (ESI), which is the most common ionization technique used in bacterial secondary metabolomics workflows.[121–123] The general pathway underlying the terpene biosynthesis in myxobacteria is the mevalonate pathway, essential for the biosynthesis of the terpene key building blocks isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP).[124] As depicted in Figure 8, the mevalonate pathway starts with acetyl-CoA, which is elongated by a C_2-unit via an acetoactyl-CoA thiolase. Subsequently, another acetyl-CoA is condensed with the resulting acetoacetyl-CoA by an HMG-CoA synthase. 3-hydroxy-3-methylglutaryl-CoA is reduced giving mevalonate. Two kinases catalyze dephosphorylation of mevalonate. Final decarboxylation results in DMAPP, which can by isomerized to IPP via an IPP isomerase. IPP and DMAPP are utilized as universal building blocks to form mono-, sesqui-, di-, tri and tetra-terpenes (shown in Figure 8). [66,125]
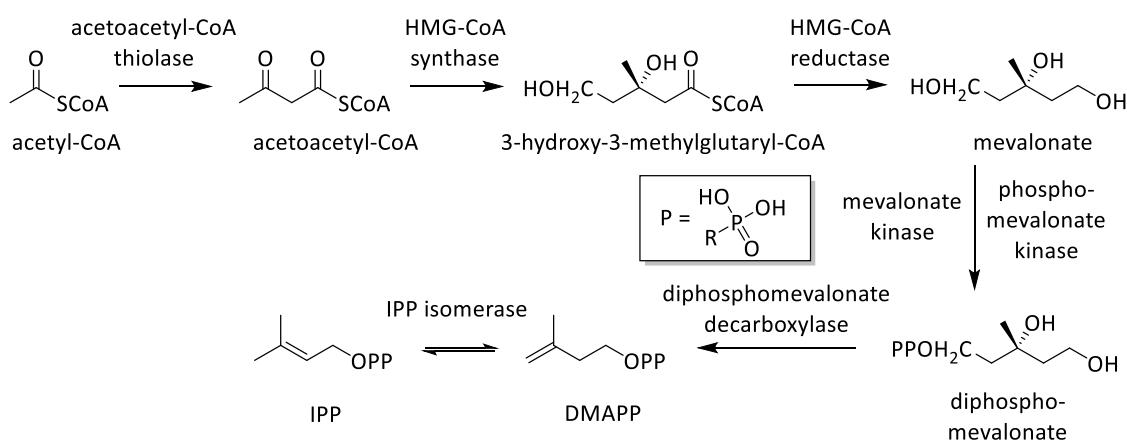
Figure 7 Mevalonate pathway for the biosynthesis of IPP and DMAPP in myxobacteria.[66]

Geranyl pyrophosphate (GPP), farnesyl pyrophosphate (FPP) and geranylgeranyl pyrophosphate (GGPP), linear isoprenoids that consist of a variable number of IPP units, are enzymatically formed by prenyl transferases.[126] Those enzymes are capable of catalyzing the synthesis of either *trans*- or *cis*-configured products, wherefore they are also classified as either *trans*- or *cis*-isoprenyl pyrophosphate synthases (IPPSs). All *trans*-IPPSs show two common motifs consisting of three aspartic acids (DDxxD motifs), which are involved in substrate binding and catalysis.[127] They bind Mg^{2+}, which serves as a cofactor during the catalysis process. On amino-acid sequence level, *cis*-IPPSs show few similarities with *trans*-IPPSs and lack the DDxxD motifs.[128] However, they share the requirement of Mg^{2+} for catalysis. In general, *trans*-IPPSs have been isolated from more organisms than *cis*- IPPSs. Products from *trans*-IPPSs have been found with various chain lengths, whereas products from *cis*- IPPSs commonly exceed a chain length of at least C_{55}.[129] Besides the linear isoprenoids, there are several terpenoid natural products featuring cyclic structures such as steroids.[130] Examples for terpenoid cyclases responsible for cyclisation of the linear isoprenoid core structures, are the squalene cyclase, the pentalenene synthase, the 5-*epi*-aristolochene synthases and the trichodiene synthase.[129] They are involved in the synthesis of cholesterol, the antibiotic pentalenolactone or the antifungal capsidiol.[131–134] In contrast to the other three enzymes that cyclize FPP, the squalene cyclase cyclizes the dimerization product of FPP. All cyclases contain conserved aspartate-rich motifs, which highlights their similarity to *trans*-IPPSs.[129]
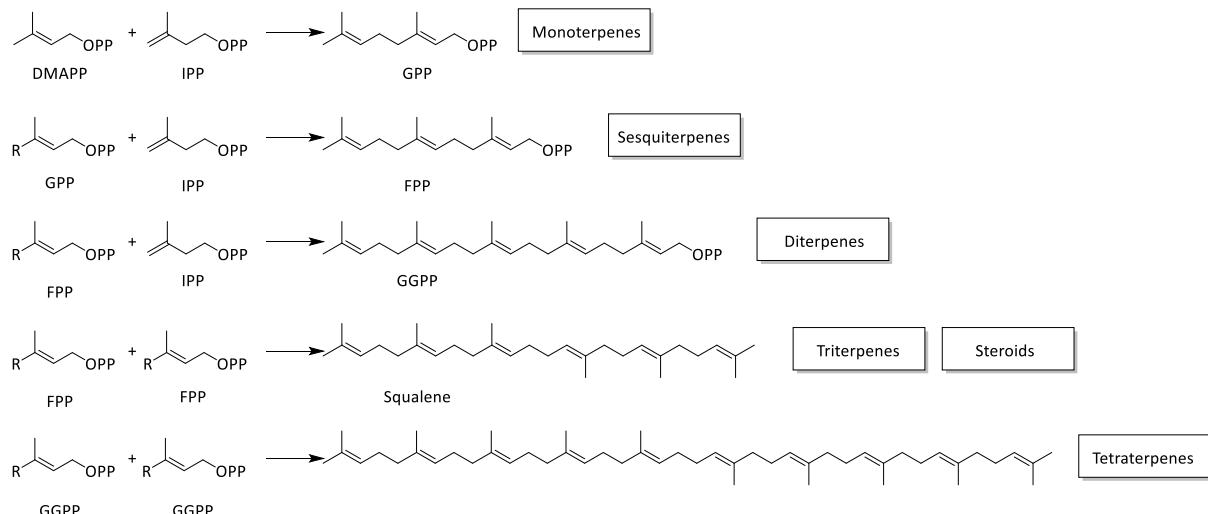
Figure 8 Linear isoprenoids formed by condensation of DMAPP and IPP catalyzed by *trans*-prenyl transferases.[126]

The number of myxobacterial terpenes identified to date is low, compared to NRPS or PKS products although myxobacterial genomes generally feature the corresponding biosynthetic pathways. However, geosmin, germacradienol and cysotdienoic acid, as depicted in Figure 3 are examples for terrestrial myxobacterial terpenes. All of them are cyclic terpenes, whereby geosmin and gemacradienol derive from an FPP body and therefore count as sesquiterpenes.[94] Cystodienoic acid was isolated from a *Cystobacter ferrugineus* strain and consists of a 20-carbon atom skeleton, which suggests a composition of two GPP subunits. Its biosynthesis, however, has not been elucidated yet.[95] Enhygromic acid, which is shown in Figure 9, is a diterpene produced by a marine myxobacterium, *Enhygromxya s*p.. It is an acrylic acid derivative, featuring a rare polycyclic carbon skeleton and showed antibacterial activity against *Bacillus subtilis* besides cytotoxic activity against melanoma cells.[135]
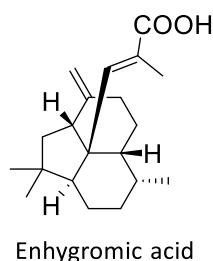
Enhygromic acid

Figure 9 Structure of enhygromic acid, a marine myxobacterial terpenoid NP.[135]

Hybrid natural products

The diversity of NPs is further increased by the possibility to combine two of these biosynthetic pathways, which is relatively common for natural products in NRPS-PKS hybrids.[136] The combination of these biosynthetic pathways is possible, because of the involvement of terminal carboxylic acids as functional groups involved in various biosynthetic reactions that can be funneled into the other assembly line respectively. Both reactions use a carboxylic acid moiety as subunit for molecule extensions by condensation reactions. NRPS-PKS hybrid biosynthetic pathways are widely distributed among myxobacteria and provide diverse variants of natural products as exemplified by the disorazoles, the DKxanthenes, the macyranones and the aetheramides, shown in Figure 10.[47,91,137,138]

Disorazol A

DKxanthene 534

Macyranone A

Aetheramide A
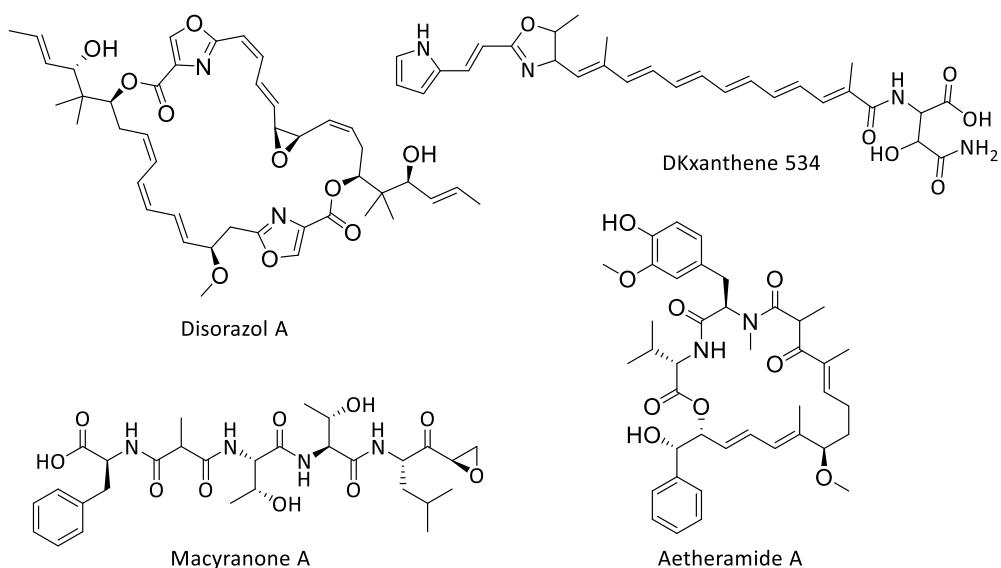
Figure 10 Selected examples for hybrid NRPS-PKS products.[47,53,91,138]

Terpene biosynthesis can be also be combined with NRPS or PKS biosynthesis. However, such hybrid pathways are more abundant in fungi than in bacteria. One example for a PKS-terpene hybrid are the meroterpenoic chevalones, which exhibit an antituberculotic activity in the low µM range. They were isolated from a *Neosartorya* species, which is the sexual state of an *Aspergillus,* in a bioactivity-guided approach.[139] Hybrid terpene NRPS products are exemplified by the alkaloidal terpenoid flavunoidine which was found by genome mining in an *Aspergillus* strain.[140] Combination of more than two pathways is also possible, which is exemplified by the myxobacterial hybrid terpene NRPS PKS product leupyrrin, where the isoprenoid substructure is incooporated into its macrolide core structure.[141,142] Structures of the three terpene hybrids are shown in Figure 11. The large diversity of already described NPs produced by myxobacteria, requires complex analyses when attempting the isolation of to date unknown secondary metabolites. Approaches and prioritization strategies carried out to access novel NPs are described in the following chapters.

Flavunoidine

Chevalone
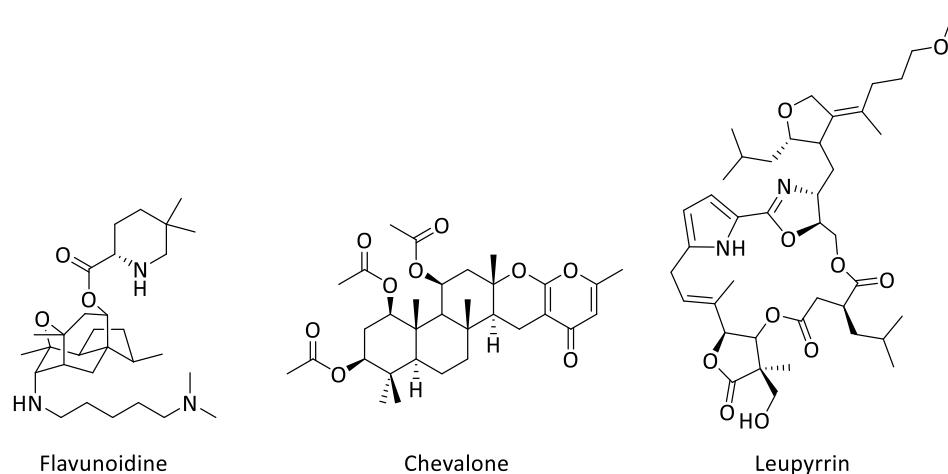
Leupyrrin

Figure 11 The hybrid terpene NRPS, terpene PKS and terpene NRPS PKS products flavunoidine, chevalone and leupyrrin.[139,140,142]

1.2 Approaches to access novel natural products

The genomic potential of myxobacteria in particular highlights that a strain is far from exploited when only isolating one class of NPs per strain.[46] Genomics investigation of 36 sequenced myxobacteria exemplarily revealed 843 BGCs with lower than 75% similarity scores to characterized clusters.[143] However, this genomic potential usually is not reflected in the strain's metabolic profile, as some of these BGCs are largely repressed or not expressed at all ("cryptic"). Other secondary metabolites are produced in such limited quantities, that they either do not reach the limit of detection of the employed analytics workflows or are not accessible for isolation and structure elucidation due to stability issues.[144] In general, there are three different approaches to access novel secondary metabolites, namely bioactivity-guided isolation and metabolome-and genome- guided approaches.[145] Especially genome- and metabolome driven approaches address current issues in NP discovery and attempt to fully exploit the genetic potential of a strain. One crucial step in all of the three approaches is careful dereplication of the acquired data. Dereplication herein depicts a comparative analysis of a potentially new secondary metabolite or its BGC with already characterized natural products or BGC using public libraries.[146] This step is necessary to prevent re-exploitation of hitherto described secondary metabolites. The following chapters will provide an overview about advantages and drawbacks of the three different strategies for natural product discovery.

1.2.1 Bioactivity-guided natural products discovery

Correlating biological activities to specific secondary metabolites present in a bacterium's extract is one of the oldest strategies applied for the discovery novel secondary metabolites.[145] For antibiotics discovery, crude extracts are either screened in a filter-disc assay, where the extracts are pipetted onto a filter-disc which is then applied to an agar plate subsequent or prior to inoculation with a pathogen strain or directly pipetted to a liquid pathogen culture.[147] After determination of the antimicrobial activities of the crude extracts, the extract displaying a desired activity is fractionated for correlation of the activity to specific natural products. This fractionation commonly is performed using rp-HPLC and can either be performed blind with subsequent offline detection or while directly acquiring mass spectrometric or nuclear magnetic resonance (NMR) data online to obtain further information about the active ingredients[148] elemental composition. Bioactivity-guided approaches on one hand enhance the discovery of secondary metabolites featuring a desired biological activity, but on the other hand strongly limit the detectable natural products to a predefined activity spectrum and those produced in relatively high amounts.

1.2.2 Metabolome-guided natural products discovery

Metabolomics investigations are a helpful tool to support bioactivity- and genomics-based natural product discovery but have also been used successfully as a standalone method for the discovery of secondary metabolites.[48,49] Metabolomics-based NP discovery relies on the hypothesis that secondary metabolism is associated with high energy-consumption for a bacterium, wherefore each individual secondary metabolite has to serve a specific purpose for the producing organism.[149] Many secondary metabolites are dedicated to the creation of a competitive advantage for the host bacterium within its microbial community. Therefore – to name an example scenario - chances should be high to find an antibiotic or antifungal metabolite from a soil bacterium.[150] Bacterial secondary metabolomics studies make use of this assumption, by aiming for a comprehensive overview about the organisms' capability of secondary metabolite production at a specific time-point.[151] One of the earliest discovered myxobacterial secondary metabolites following a metabolome-guided NP discovery approach are the

aurachins. In contrast to the more commonly used LC-coupled ultraviolet (UV) detection at that time, they were identified by investigation of the bacterium's extracts with ^{1}H-NMR experiments (see Figure 12).[65] Later, biological characterization of the compounds showed antibiotic effects against Gram-positive bacteria and pathogenic yeast strains.[152,153] UV-based detection methods are commonly used during chromatographic separation of the bacterial extract using liquid chromatography (LC). One example where a UV-LC hyphenation was successfully applied for secondary metabolite isolation in a metabolome-based detection and isolation approach is dawenol (Figure 12). This NP is produced by different *Stigmatella aurantiaca* strains and shows an UV absorption band with a maximum at 362 nm and three shoulders at 328, 344 and 383 nm, which is characteristic for six conjugated double bonds.[68] Since polyenes often show promising antimicrobial activities, searching for an absorption band with this shape appears as rationally justified.[154] UV-based screenings are biased towards molecules with strong chromophores, wherefore nowadays most metabolomics studies rely on the detection by mass spectrometry (MS).[145] Most commonly, MS is used as a hyphenated system coupled to LC and offers besides its high sensitivity the possibility to obtain information about the elemental composition of the target metabolites.[155] In case of chlorotonil, this information derived from the characteristic isotopic pattern was successfully used for prioritization of a NP class, later showing potent activities against *Plasmodium falciparum*, the malaria-causing parasite. Chlorotonil, as indicated by its name, shows a twofold chlorination. This chlorination was detected by investigation of the isotopic pattern of the substance by *hr*MS and translated into the sum formula proposed during the discovery process of the NP (shown in Figure 12).[63] The three main detection methods (UV, NMR, MS) are described in the following paragraphs in more detail.
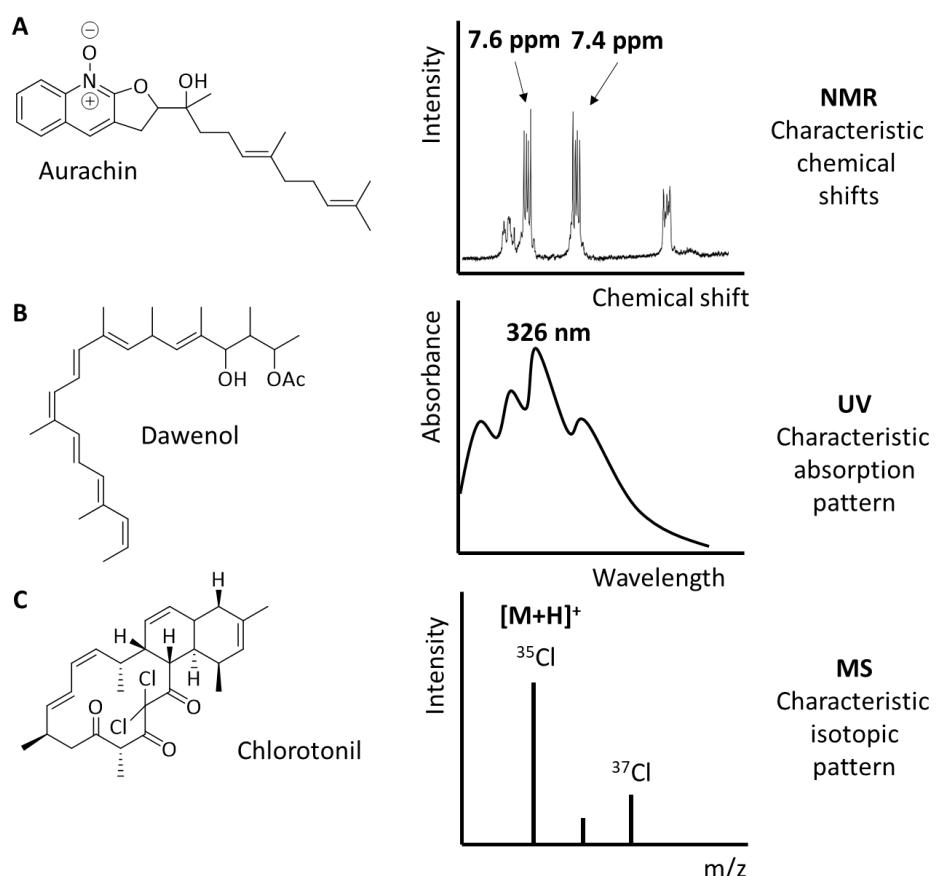
Figure 12 Examples of myxobacterial secondary metabolites isolated in a metabolomics-guided approach with corresponding detection methods used for compound prioritization.[63,65,68]

NMR-guided natural products discovery

The most common application of NMR in natural product research is the generation of 1D and 2D spectra for *de-novo* structure elucidation or for structural confirmation of synthetic or known NPs. Besides that, NMR spectroscopy is also used for determination of protein-ligand interactions or metabolomics-guided NP discovery.[156] Three of those four applications use purified samples, whereas NMR based metabolomics studies are carried out on crude extracts or pre-purified fractions, that represent complex mixtures.[157] The complexity of myxobacterial crude extracts is comparably high, wherefore examples such as the aurachins discovered by ^{1}H-based crude extract screening are a notable exception.[65] One option to reduce this complexity is the hyphenation of NMR with LC.[158] LC-NMR, often employed as LC-SPE-NMR among the detection methods described herein is the only method that offers detailed structural information in early steps of the NP discovery workflow. Specific functional groups can already be determined based on their chemical shifts and NPs can therefore be prioritized based on functionality. However, even the state of the art NMR spectrometers with field strength exceeding 1 GHz cannot reach the sensitivity of mass spectrometers, wherefore most samples need to be enriched in flow-through probes such as SPE-NMR or CapNMR.[37,145] Six linear myxobacterial peptides, the macyranones A-F (shown in Figure 10) have already been prioritized for isolation and structure elucidation in a combined LC-SPE-NMR-MS workflow.[91] This respective study highlights two things that have to be considered, when carrying out an LC-NMR experiment. On one hand, hyphenation with MS can provide more structural information in an early dereplication process, as both sum formula and functional groups can be determined in one analysis.[158] On the other hand, even with this structural information, full *de-novo* structure elucidation is not possible, wherefore LC-NMR cannot replace upscaling in cultivation, purification and subsequent structure elucidation by multi-dimensional NMR spectra or X-ray crystallography.[91]

UV-guided natural products discovery

UV-detection is the oldest, cheapest and easiest to use among the three metabolomics detection methods described here. However, prior to detection with UV most secondary metabolites were separated using thin layer chromatography (TLC) and detected with spray reagents. Detection by UV is very reliable, non-destructive and capable of detecting a broad range of metabolites compared to spray reagents, but suffers from a strong dependency of its sensitivity correlated to the extinction coefficient of a molecule. Furthermore, only little structural information is provided.[159] Only a limited number of structural features do actually bring about characteristic UV-absorption bands, enabling low detection limits and predictions of structural features such as the conjugated double bond system in dawenol.[68] A major advantage of UV-detectors however, is their interconnectivity to LC-MS systems. Detection by UV furthermore is non-destructive, wherefore most LC-MS hyphenations also include a UV detector, which is incorporated prior to the destructive MS analysis.[151]

Mass spectrometry-guided natural product discovery

Mass spectrometry is one of the disciplines in analytical chemistry that has revolutionized metabolomics-based natural product discovery is nowadays heavily used in natural product research.[160] It can be sub-divided into two major disciplines, low-resolution and high-resolution mass spectrometry (see Figure 13). Low-resolution mass spectrometers such as quadrupoles or ion traps are commonly used when primary information for example the structure or at least the exact mass of a target molecule is already known.[161] They can be operated by users with limited training, are less expensive and often more robust compared to high-resolution mass spectrometers. Quadrupoles and ion traps represent the two spectrometer types of low-resolution mass spectrometers commonly used

in natural product research that enable the detection of specific mass to charge (*m/z*) ratios. However, these technologies are often used in combinations such as the sequential alignment of a quadrupole with an ion-trap or the successive connection of multiple quadrupoles as in case of a triple quadrupole spectrometer, a device commonly used for quantifications.[162] In quadrupole MS, ions are guided between four rod electrodes. Controlled by the generated overlay of a constant and an alternating electric field, only ions with a specific *m/z* ratio can pass the quadrupole according to the Mathieu equation. For a full scan mass spectrum, all defined *m/z* ratios are scanned one after another, wherefore depending on the width of the *m/z* range full scans can be comparably slow in quadrupole MS. In ion trap MS, the complete ion cloud is trapped in an electrostatic field. Controlled excitement which leads to outpulsing of ions with a specific *m/z* ratio is then used to generate the mass spectrum. Like in quadrupole MS, each full scan is therefore divided into several consecutive scans, resulting in a comparably long cycle time needed for one full scan.[163–165] High-resolution mass spectrometers such as time-of-flight (TOF), Fourier transform ion cyclotron resonance (FT-ICR) and orbitrap MS offer high sensitivity and mass accuracy. In NP discovery, they are mainly used to determine sum-formulas, in statistics based non-targeted metabolomics workflows and for accurate MS/MS based structure characterizations. However, their complexity makes them less robust and they are much more expensive in acquisition and maintenance compared to the low-resolution MS. High-resolution MS are often available as hyphenations with several quadrupoles, which are used as an ion filter before entering the analyzer or as collision cells for MS/MS analysis.[165,166]
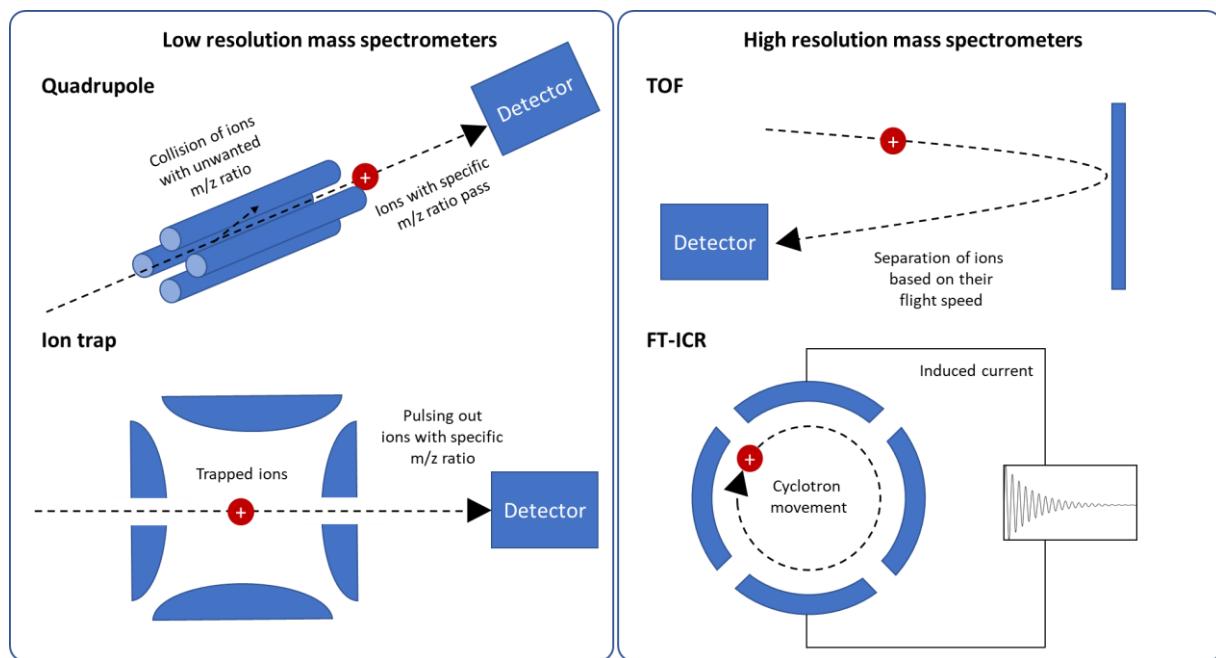
Figure 13 Four types of mass spectrometers mainly used in NP research and their principal function.[163,165]

All types of MS instruments can in principle be coupled to an LC system for prior analyte separation, but good chromatographic separation with very narrow peaks require a fast scan rate of the instrument. This may be one reason why TOF and orbitrap instruments are the preferred types of mass spectrometer among the high-resolution MS instruments in analytical UHPLC-MS hyphenations.[167,168] In TOF instruments, ions are separated based on their flight velocity, which is dependent on mass over charge ratio of the respective ions. The process of scanning different *m/z* values is therefore not applicable for those types of instruments, but rather the entire ion cloud can be measured, requiring only the time spent to pass the flight tube for one full scan.[165] Concerning resolution, orbitrap and FT-

ICR instruments provide the best results. They are more suitable for direct infusion setups, since fast scanning with those types of instruments is associated with a loss in resolution.[169] FT-ICR MS uses a strong, homogeneous magnetic field to trap the ion cloud. By applying an alternating voltage to the ions, a cyclotron resonance of the ions is induced, resulting in an induced current, which can be translated into a signal.[170]

The different types of mass spectrometers can be equipped with different categories of ion sources relying on electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI) or matrix assisted laser desorption ionization (MALDI). Those three represent the most commonly used ion sources in natural product research, but several more exotic ionization techniques for example desorption spray ionization (DESI) gain more and more interest in this field as it adds spatial imaging.[171] Each one of those ionization types relies on a different physio-chemical process, wherefore their application strongly depends on the analyte of interest. ESI is one of the softest ionization mechanisms and uses protonation of functional groups, wherefore it is particularly suitable for sensitive molecules like PKs and NRPs often rich in functional groups.[172] APCI as indicated by its name relies on a chemical reaction in the gas phase during the ionization process and is considered as hard ionization technique, leading to a more intense in source fragmentation of the analytes compared to ESI.[173] Nevertheless, ionization of molecules like terpenes, often poorly equipped with functional groups, however is more easily achieved with APCI compared to ESI. MALDI is also classified as hard ionization technique, since ions are generated through a laser beam which ionizes a matrix from which protons are transferred to the sample. MALDI is mainly used in imaging mass spectrometry (IMS) workflows as this type of ion source is suitable for 2D-MS. Thereby, the laser beam used for the ionization process is moved over a surface, which leads to generation of mass spectra containing pixels that can be translated into special mappings of NPs.[174]

Mass spectra of biological samples, especially when generated on high-resolution MS with ESI as ionization source are immensely rich in information. A non-targeted manual analysis often falls short in analyzing these large datasets, as it focuses on well-ionizing substances correlating with high peak intensities. In order to highlight less intense analytes, statistical analysis (e.g. PCA) is a helpful tool.[175,176] It is often used to show differences between single-crossover or overexpression mutants and their respective wild-type strain, but have also proven valuable for characterization of the secondary metabolome of myxobacterial strains when analyzing differences between related and taxonomically distant strains.[45,49] For statistical analysis, all detected signals are first grouped into so-called features (see Figure 14).[177] For feature detection, different signals are assigned to specific ion types based on the exact mass shift when comparing signals deriving from the same molecule.[176] This assignment is possible, as during the ESI process different ion types like the [M+H]$^+$ adduct, the [M+Na]$^+$ adduct or the [M+K]$^+$ adduct are generated and detected simultaneously, thus showing a timely correlation in elution during an LC-run.[178] In one feature, those ion types, as well as monoisotopic, primary and secondary isotopic peaks are combined and simplified as one data collection, a so-called bucket.[176]
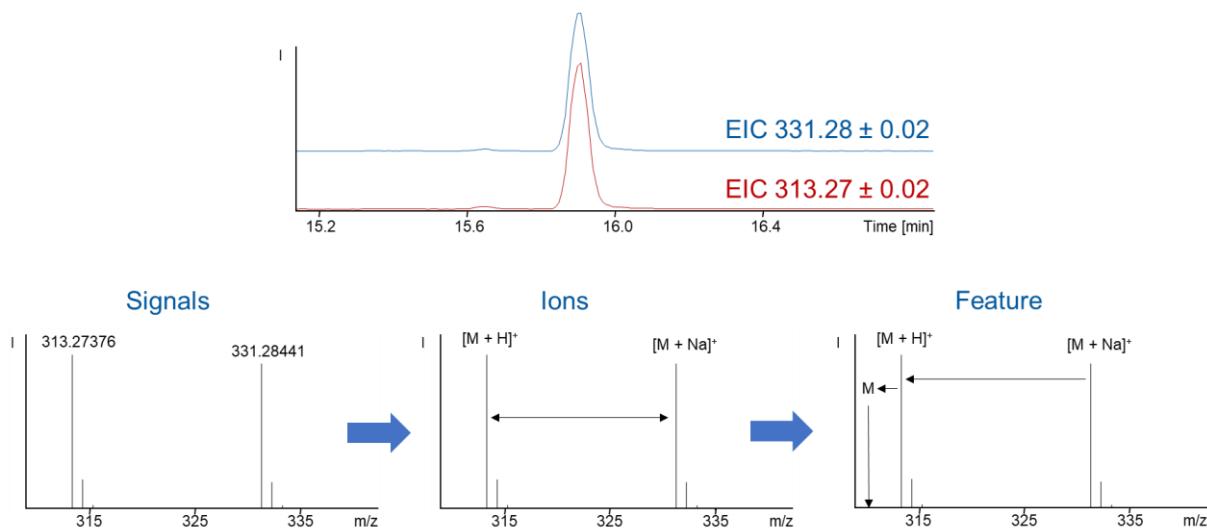
Figure 14 Dereplication steps involved in molecular feature generation from a signal detected with MS.[179]

The next step in data reduction during a statistical analysis workflow is subtraction of all features detected in a blank run. Those blank runs in myxobacterial metabolomics usually consist of a media blank which is treated in the same way as the myxobacterial cultures.[180] Myxobacteria require extremely complex media, often consisting of well-ionizing plant metabolites, wherefore this step is crucial to prevent follow-up of MS features which are not produced by the investigated myxobacterium. Even after bucketing of different signals into features and subtraction of blank features, myxobacterial extract datasets are still inherently complex as each extract consists of primary and secondary metabolites.[181] Each myxobacterium is capable of producing a variety of secondary metabolites, which commonly consist of several derivatives.[37] The fact that secondary metabolites usually do not consist of only one family member can be used for further data reduction. The Global Natural Product Social Molecular Networking (GNPS) is a free-of-charge online tool, which can be used for clustering metabolites based on their characteristic MS/MS fragmentation pattern. Members of one natural product family are easily assigned to each other.[182] Additionally, GNPS clustering provides various databases containing fragmentation pattern information about known metabolites, which can be matched with the imported datasets for dereplication.[183]

1.2.3 Genome-guided natural products discovery

Among the screening approaches available for NP discovery, genome-guided screening approaches provide a comprehensive overview about a strain's potential for secondary metabolite production. However, it is also one of the most recent approaches, since only the plummeting sequencing costs in recent decades allowed access to genomic data.[184] The first myxobacterial genome for example was only published in 2007.[58] These genomic data revealed that the biosynthetic potential of a strain is commonly not fully translated into the metabolome. A variety of bioinformatics and biotechnological tools have been developed since the availability of microbial genomes to access this hidden biosynthetic potential.[34] Most of them are based on the prediction of BGCs using tools such as antiSMASH,[185] which is possible due to the fact that most natural product biosynthesis genes are not located in disjointed places in the genome but rather cluster together. Manual annotation of BGC's is also possible and needed to verify hits generated with automated annotation tools, but laborious, time-consuming and not feasible in high-throughput workflows as a standalone method. There are other prediction tools for biosynthetic pathways besides antiSMASH, such as PRISM, ClustScan, the SBSPKS toolbox or CLUSEAN but they often are limited to a specific type of biosynthetic machinery or

are difficult to operate.[186,187] AntiSMASH on the other hand is able to annotate a large variety of biosynthetic gene clusters including NRPSs, PKSs, RiPPs and terpenes, as well as hybrids thereof.[185] After annotation of all BGCs in one bacterial strain, they need to be prioritized for further investigation as genetic manipulation followed by isolation and structure elucidation of the corresponding product is a laborious process not feasible for all BGCs in one strain. Prioritization includes a dereplication step, where BGCs already connected to known products are excluded from the pipeline. Subsequently, the BGCs could be ranked by the class of NP they are producing, BGC complexity and length, interesting tailoring enzymes or by resistance genes located in close neighborhood to the BGC.[145] The search for resistance genes relies on the hypothesis that the bacterium has to protect itself against the antibiotics it produces for survival in case it comprises the same target the antibiotic addresses. This strategy has already been used successfully for the isolation of the pyxidicyclins and alkylpyrones from two myxobacterial species.[51,87] Those two NP families were discovered by searching for so called pentapeptide repeat proteins (PRP) located in proximity to BGCs. The PRPs have been found to bind to DNA gyrase of the producing organisms and therefore protect the producer's gyrase (see Figure 15). This mechanism of resistance has been described for cystobactamid producing *Cystobacter* species before, allowing the targeted search for PRPs in other myxobacteria.[188,189]
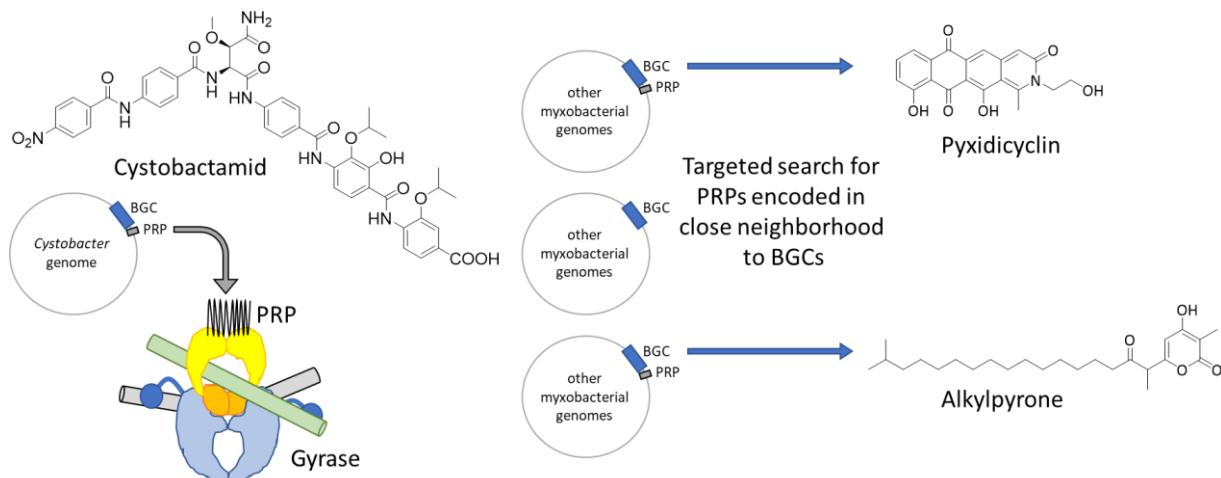
Figure 15 Self-resistance genome mining based on cystobactamid resistance mechanism reveals the pyxidicyclins and alkylpyrones as gyrase inhibitors.[51,87]

Once a BGC has been prioritized for in depth analysis, either inactivation experiments, overexpression mutations or heterologous expression can be carried out to connect the BGC to its corresponding NP (depicted in Figure 16).[145] The myxobacterial strain *Chondromyces crocatus* Cm c5 well exemplifies the impact of inactivation and overexpression mutations. Genomic investigation of this strain showed 18 BGCs either producing NRPs, PKs or hybrids thereof. Following further analysis of the genome in line with the metabolome, five BGCs thereof could be correlated to known products. Subsequently, three new secondary metabolites could be accessed by genetic manipulation.[37] Single-crossover mutagenesis led to the discovery of crocagins and crocadepsins and allowed their assignment to BGCs.[78,79] Insertion of a strong promotor in front of the crocapeptin BGC gave a mutant with 10-fold increased production of crocapeptin A1, making it accessible for purification and structure elucidation.[77] Newly isolated wild type bacteria, however, are often barely accessible by the genetic tools available, hindering the generation of inactivation or overexpression mutants. In this case, heterologous expression in the means of transfering the respective BGCs from the producer organism into a different host like *Myxococcus xanthus* which is accessible for genetic manipulations is another option to approach secondary metabolites associated to those BGCs.[190]
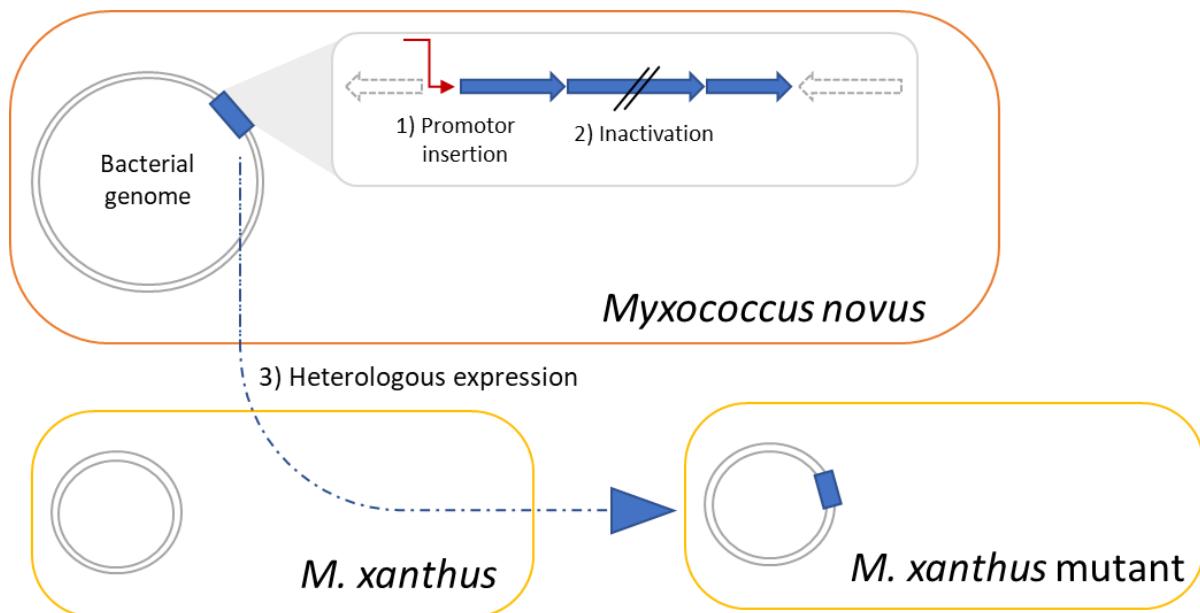
Figure 16 Schematic representation of three most common genetic tools used for the correlation of a BGC to its product.[190]

The output of inactivation and overexpression mutants as well as of heterologous expression experiments also relies on choice of the appropriate analytical setup to avoid detection failure. PCA based statistical comparison of LC-MS data of mutant and wild type extracts can further increase sensitivity and therefore help to detect also small changes in the metabolite profile.[49] In cases of very low production yield, subsequent optimization of the cultivation and extraction conditions or promotor optimization can be crucial to obtain sufficient amounts for structure elucidation of the novel secondary metabolite.

1.3 Extraction and secondary metabolite purification for structure elucidation

The methods described in chapter 1.3 are all suitable for metabolite detection and prioritization and already provide some useful information about the chemical properties of the molecule of interest. For a complete characterization, the metabolite of interest needs to be isolated in sufficient quantity for structure elucidation by either NMR or Xray crystallography as well as for evaluation of biological activity in subsequent biological assays.[191,192] Conventional workflows for isolation of myxobacterial secondary metabolites rely on production in either shake flasks or upscaling for production in fermenters. To enhance secondary metabolite production by blocking negative feed-back mechanisms and thus ease the downstream processing, an adsorber resin (e.g. XAD-16) can be added, which binds the secondary metabolites secreted into the supernatant of the culture and therefore circumvents negative feedback inhibition mechanisms.[193] Subsequently, cells and adsorber resin are extracted by solvent extraction (maceration). After evaporation of the extraction solvent, the crude extract is resuspended in a small amount of liquid and subjected to pre-purification steps such as liquid-liquid partitioning or chromatographic separation technologies such as reversed phase liquid chromatography (rp-LC), size-exclusion chromatography (SEC) or normal phase liquid chromatography (np-LC).[194] The high complexity of myxobacterial extracts usually requires the combination of two or more of these separation techniques to obtain the pure NP.[87,91] However, even with these chromatographic tools in hand, separation of highly complex NP families or purification of secondary metabolites bearing special features such as a permanent charge or instability-causing functional groups remains challenging.[195] The amounts produced during fermentation of a wild-type

myxobacterial strain in first isolation attempts are usually comparably low (less than one mg 12 L culture volume). To increase these yields, one can either optimize the production conditions to enhance the myxobacterium's capability of NP production or optimize the extraction process.[196,197] The implementation of supercritical fluid based technologies can help during two steps of the myxobacterial secondary metabolite isolation process. Supercritical fluid extraction (SFE) may help to enhance the extracted amounts of NPs leading to enriched crude extracts. Moreover, supercritical fluid chromatography (SFC) is a complementary method to rp-HPLC or SEC. It is therefore a useful tool, which can be applied for separation when the other chromatography reaches its limits as it relies on an orthogonal separation mechanism.[198] SFE and SFC furthermore come with decreased solvent consumption and are therefore often referred to as "green" technologies.[199] They are heavily used for the isolation and purification of plant metabolites, however underrepresented in bacterial NP research.[200] One exception here are cyanobacteria, where SFE based purifications have already have shown their impact.[201,202] The unique physicochemical properties of supercritical fluids are highlighted in chapter 1.4.1 and principal functionalities of SFE and SFC described in the subsequent chapters.

1.3.1 Supercritical fluid extraction

The supercritical state is defined as physical state of matter that is reached when a substance exceeds its so-called critical point. At this state, distinct gas and liquid phases do not exist anymore.[203] The supercritical fluid now features intrinsic properties from both states, it perfuses solids like a gas but shows a solvation like a liquid. Those intrinsic properties qualify supercritical fluids for usage in extraction processes as well as for chromatographic separations.[204] The temperature and pressure at which the critical point of a solvent is reached depends on the solvent itself. For most solvents, extreme temperatures and pressures are needed to reach the supercritical stage, wherefore they are not applicable in most technologies. The critical point of water for example is reached at 374 °C and 218 bars.[205] For supercritical CO_2 however, only comparably mild 31 °C and 74 bars are needed (see Figure 17).[206] Furthermore, CO_2 is abundant and arises as waste product during several processes, making it the solvent of choice for supercritical fluid technologies.[203]
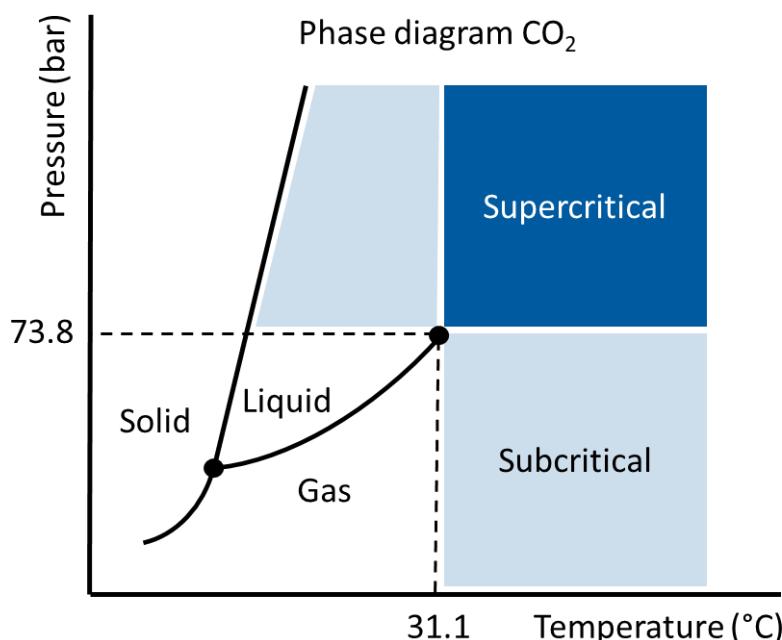
Figure 17 Phase diagram of CO_2. Lower point: triple point, upper point: critical point.[206]

The phase diagram of CO_2 however, only accounts for processes where CO_2 is the only solvent used. As CO_2 shows a polarity similar to hexane, its use alone as a solvent does not yield sufficient results in natural product research, as many secondary metabolites are quite polar.[196] Supercritical fluids used in NP research are therefore often mixtures of different solvents, for example CO_2 and ethanol.[207] In these mixtures, the supercritical point strongly depends on the concentration and type of co-solvent used.

Due to low viscosity and high diffusivity of supercritical CO_2, SFE is a very fast process. By selecting pressure and temperature for the extraction process, the extraction is selective to some extent.[208] A common SFE system contains four different modules, a pump, an automated back pressure regulator (ABPR), a column oven and a fraction collector (see Figure 18).[209] The pump is connected to a bottle of CO_2, which is pumped in liquid state and one or more co-solvent bottles. Before reaching the ABPR the liquid CO_2 and co-solvent are combined and mixed. The ABPR consists of a needle and a needle seat. The deeper the needle penetrates into the needle seat, the more back pressure is built up in the system. Fine adjustment of the needle allows accurate pressure regulation. After passing the ABPR, CO_2 and co-solvent reach the column oven, where they are heated to reach the critical point and transit to the supercritical state when entering the extraction vessel (cartridge).[209] The cartridge is prefilled with raw material, which needs to be completely dry and pulverized. Pulverization allows dense package and therefore better extraction efficiency. The material needs to be dry, as remaining water, which is not miscible with CO_2 may crystallize during the process and destroy the dense packaging, which in return lowers the extraction efficiency.[210] Most SFE systems offer the possibility to connect different co-solvents, allowing a sequential extraction of elutes with different polarities. Those can then by collected separately in the fraction collector.[211] One main advantage of SFE compared to conventional solvent extraction is the reduced amount of liquid solvent collected in the process. After leaving the column oven, the solvents are depressurized. At room temperature and atmospheric pressure, the CO_2 evaporates, retaining the elutes in the comparably small amount of co-solvent used for extraction.[209]
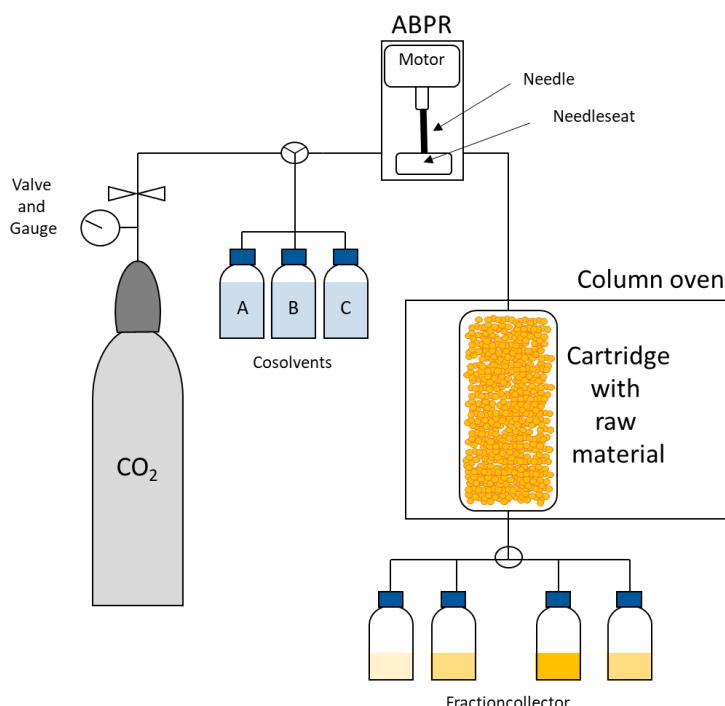
Figure 18 Principal composition of a supercritical fluid extraction system.

1.3.2 Supercritical fluid chromatography

SFC can be used in analytical scale as well as for compound purification in semipreparative and preparative scale. A principle preparative SFC system consists of the same modules like an SFE system. As only difference, the cartridge filled with raw material is replaced by a chromatographic column and an injection system is incorporated into the system between pump and ABPR.[212] Like in SFE, SFC methods are very fast compared to HPLC methods as they benefit from the high diffusivity of supercritical CO_2.[213] High chromatographic efficiency can furthermore be retained even when injecting large amounts of material. In principal, there are no limitations for the column material used in SFC except for pressure stability. There are specific SFC columns, but rp- and np-HPLC columns can also be mounted into the system and used in this type of chromatography.[214] The mobile phase preliminarily consists of supercritical CO_2. The polarity of the mobile phase can be modified by the usage of different co-solvents.   Prominent co-solvents are alcohols like methanol and ethanol or acetonitrile and ethyl acetate.[215] Similar to HPLC, the co-solvents can be further adjusted by modifying their pH value by addition of acid or base.[213] SFC can either be run isocratic or using a gradient. Here the high diffusivity of CO_2 allows fast changes in the percentage of co-solvent as well. Besides the choice of co-solvent, which has the biggest influence on selectivity, in SFC temperature and pressure can be regulated, having a not negligible effect on the chromatographic separation.[216] Modern SFC systems are as robust as HPLC systems. The high pressure under which the system needs to be operated, however makes the system more maintenance intensive, as all parts need to be "gas tight".

1.4 Outline of this work

The overarching aim of this work is the comprehensive characterization of the secondary metabolome of myxobacterial strains to facilitate the isolation and characterization of novel biologically active natural products. In this context, traditional methods applied in a typical natural product discovery workflow are evaluated and compared to hitherto underrepresented techniques in this field. The steps of cultivation, detection, extraction and purification are investigated in the course of this thesis to improve the natural product discovery workflow for myxobacterial strains. In total, nine myxobacterial strains are investigated, whereof the newly isolated *Sandaracinus* strain MSr10575 belonging to one of the underexplored myxobacterial genera and the already well-described myxobacterium *Myxococcus xanthus* DK1622 are used for natural product isolation.

The evaluation of different cultivation and detection methods is described in chapter 2. To enhance the production of novel secondary metabolites, two cultivations systems (agar plate and liquid cultures) are compared in a targeted and non-targeted statistical mass spectrometry workflow. Liquid cultivation hereby represents the traditional cultivation system, commonly used for the production of secondary metabolites in large scale. In contrast, Agar plate cultivation provides a much closer representation of the myxobacterial natural habitat as soil-living bacteria and fruiting body formation as well as swarming can be observed, which are unique myxobacterial morphological features. These morphological changes are likely to have consequences in the strain's secondary metabolome. In the course of the comparison of those two cultivation systems, the influence of the usage of different analytical setups for the subsequent metabolomics analysis is studied. The LC-ESI-coupled TOF analysis is herein used as representation of a standard metabolomics workflow available to most natural product laboratories. The results of this analysis are compared to a direct infusion MS workflow on an ultra-high-resolution mass spectrometer, the FT-ICR. This type of instrument offers the highest resolution among MS spectrometers on the market, but is only available in few natural product research groups. Following the investigation of the different cultivation and detection approaches, the extraction process, which is crucial for obtaining sufficient amounts for structure elucidation of each

novel compound, is evaluated in chapter 3. For decades, common extraction protocols for myxobacteria relied on classical solvent extraction. In this work, this traditional approach is compared to a supercritical fluid extraction (SFE) protocol. SFE is a technique commonly used for the extraction of plants and cyanobacteria, but the impact on extraction of secondary metabolites from different bacterial species has rarely been studied previously. Little solvent consumption and mild extraction conditions combined with good extraction efficiency make SFE an attractive extraction technique for myxobacterial secondary metabolites, which are often produced in small amounts and suffer from poor stability. As part of this work, the influence of different co-solvents on the extraction efficiency of myxobacterial secondary metabolites are studied.

The two evaluations of underrepresented techniques in a natural product discovery workflow described in chapter 2 and 3, as well as a genomic investigation of the new *Sandaracinus* strain MSr10575 delivers four new myxobacterial natural product families, which are described in this thesis. The enhanced extractability using SFE allows purification of chloroxanthic acid A from *Myxococcus xanthus* DK1622 (see chapter 3) and thereby highlights the impact of underrepresented methods on the accessibility of natural products from strains where most of the accessible metabolites seem to be characterized already. Supercritical fluid chromatography is successfully used to access two new natural product families from *Sandaracinus* MSr10575, the sandacrabins and the sandaramines. As each of these families contains members possessing a permanent positive charge, both of these alkaloid terpenoid families are not accessible by conventionally used purification methods.

The sandacrabins, which are described in chapter 4, are farnesylated dimethylbenzimidazoles displaying antiproliferative, insecticidal, antiviral and antibacterial activities. Three derivatives are accessible *via* isolation from the strains crude extract, whereas development of a total synthesis route for the sandacrabins extends the family by two additional members. Even though structurally related terpenylated benzimidazoles have been studied in chemical synthesis approaches before, this is the first description of bacterial derived members of this class. The symmetric sandaramines, consisting of two drimenol subunits coupled *via* a homospermidine linker, display antimicrobial, antiviral and antiproliferative activities and are described in chapter 5. Development of a total synthesis route for the sandaramines allows extension of the two membered family by two further derivatives providing insights into preliminary structure activity relationships of this natural product class. Genome sequencing of *Sandaracinus* MSr10575 finally helps to develop a concise biosynthesis hypothesis for the sandacrabins and sandaramines. Furthermore, it reveals that *Sandaracinus* MSr10575 contains a plasmid harboring a NRPS-PKS hybrid biosynthetic gene cluster (BGC). Genetic manipulation of this BGC leads to successful isolation and characterization of nine members of the third natural product family from this strain, the sandarazols, which are described in chapter 6. The antimicrobial and antiproliferative sandarazols contain intriguing structural features and very reactive functional groups such as an α-chlorinated ketone, an epoxyketone and a (*2R*)-2-amino-3-(*N,N*-dimethylamino)-propionic acid building block.

The efforts summarized in this thesis aim to testify to the ongoing innovation potential at each step of a natural product discovery workflow; an important contribution so that modern natural products research can continue to fulfill its role as a supplier of important new drug molecules.
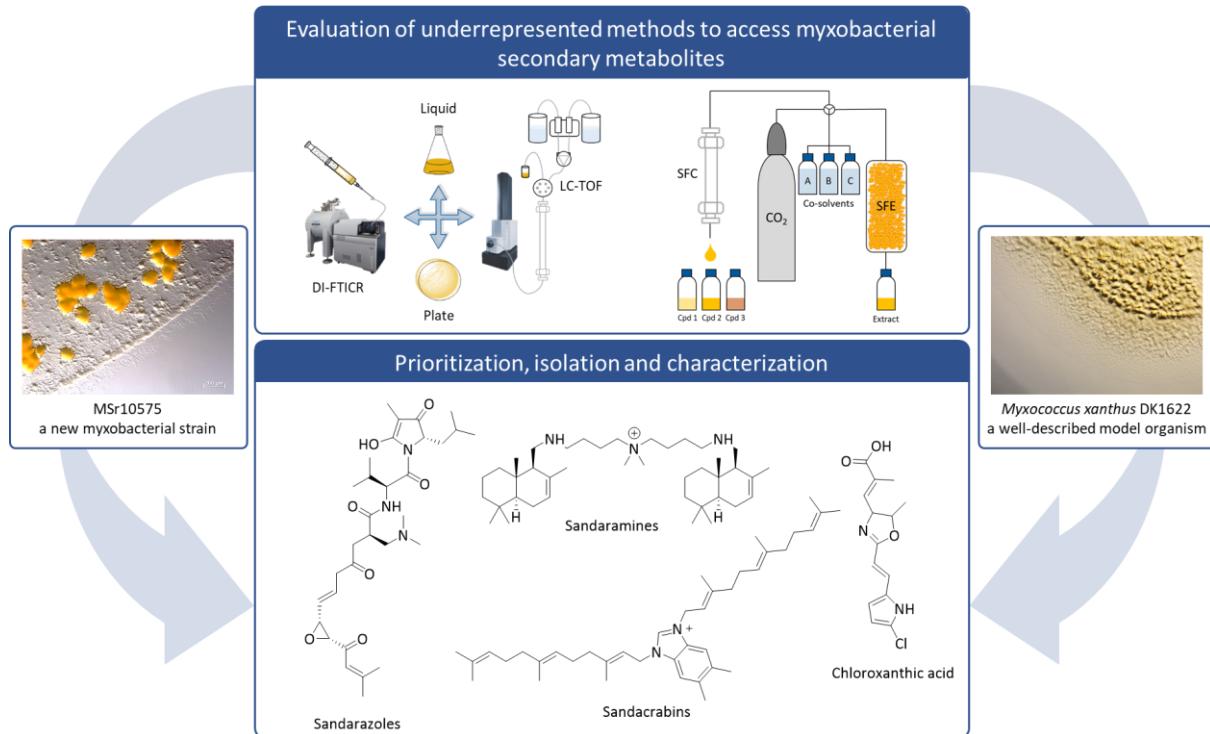
Figure 19 Underrepresented techniques evaluated and used to access four novel natural product families from the myxobacterial strains *Sandaracinus* MSr10575 and *Myxococcus xanthus* DK1622 in this thesis.

1.5 References

(1) Boucher, H. W.; Talbot, G. H.; Bradley, J. S.; Edwards, J. E.; Gilbert, D.; Rice, L. B.; Scheld, M.; Spellberg, B.; Bartlett, J. Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin. Infect. Dis.* **2009**, *48*, 1–12.

(2) Tommasi, R.; Brown, D. G.; Walkup, G. K.; Manchester, J. I.; Miller, A. A. ESKAPEing the labyrinth of antibacterial discovery. *Nat. Rev. Drug Discov.* **2015**, *14*, 529–542.

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

(4) Mohr, K. I. History of Antibiotics Research. *Curr. Top. Microbiol. Immunol.* **2016**, *398*, 237–272.

(5) Pidot, S. J.; Coyne, S.; Kloss, F.; Hertweck, C. Antibiotics from neglected bacterial sources. *Int. J. Med. Microbiol.* **2014**, *304*, 14–22.

(6) Langford, B. J.; So, M.; Raybardhan, S.; Leung, V.; Westwood, D.; MacFadden, D. R.; Soucy, J.-P. R.; Daneman, N. Bacterial co-infection and secondary infection in patients with COVID-19: a living rapid review and meta-analysis. *Clin. Microbiol. Infect.* [Online early access]. DOI: 10.1016/j.cmi.2020.07.016.

(7) Faron, M. L.; Ledeboer, N. A.; Buchan, B. W. Resistance Mechanisms, Epidemiology, and Approaches to Screening for Vancomycin-Resistant Enterococcus in the Health Care Setting. *J. Clin. Microbiol.* **2016**, *54*, 2436–2447.

(8) Pelaez, F. The historical delivery of antibiotics from microbial natural products - Can history repeat? *Biochem Pharmacol* **2006**, *71*, 981–990.

(9) Subramani, R.; Sipkema, D. Marine Rare Actinomycetes: A Promising Source of Structurally Diverse and Unique Novel Natural Products. *Marine drugs* **2019**, *17*.

(10) Sandler, J. S.; Fenical, W.; Gulledge, B. M.; Chamberlin, A. R.; La Clair, J. J. Fluorescent profiling of natural product producers. *Adv. Ceram. Mater.* **2005**, *127*, 9320–9321.

(11) Pichersky, E.; Gang, D. R. Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. *Trends Plant Sci.* **2000**, *5*, 439–445.

(12) Hong, J. Natural product diversity and its role in chemical biology and drug discovery. *Curr. Opin. Chem. Biol.* **2011**, *15*, 350–354.

(13) Michael, A. J. Polyamines in Eukaryotes, Bacteria, and Archaea. *J. Biol. Chem.* **2016**, *291*, 14896–14903.

(14) Mebs, D. Toxicity in animals. Trends in evolution? *Toxicon* **2001**, *39*, 87–96.

(15) Fan, Q.; Ma, J.; Xu, Q.; Zhang, J.; Simion, D.; Carmen, G.; Guo, C. Animal-derived natural products review: focus on novel modifications and applications. *Colloids and surfaces. B, Biointerfaces* **2015**, *128*, 181–190.

(16) Chamberlin, S. R.; Blucher, A.; Wu, G.; Shinto, L.; Choonoo, G.; Kulesz-Martin, M.; McWeeney, S. Natural Product Target Network Reveals Potential for Cancer Combination Therapies. *Front. Pharmacol.* **2019**, *10*, 557.

(17) Bianchi, A.; Evans, D. B.; Cobb, M.; Peschka, M. T.; Schaeffer, T. R.; Laffan, R. J. Inhibition by SQ 20881 of vasopressor response to angiotensin I in conscious animals. *Eur. J. Pharmacol.* **1973**, *23*, 90–96.

(18) Jakovljević, M.; Jokić, S.; Molnar, M.; Jašić, M.; Babić, J.; Jukić, H.; Banjari, I. Bioactive Profile of Various Salvia officinalis L. Preparations. *Plants (Basel, Switzerland)* **2019**, *8*.

(19) Besse, A.; Peduzzi, J.; Rebuffat, S.; Carré-Mlouka, A. Antimicrobial peptides and proteins in the face of extremes: Lessons from archaeocins. *Biochimie* **2015**, *118*, 344–355.

(20) Lukežič, T.; Lešnik, U.; Podgoršek, A.; Horvat, J.; Polak, T.; Šala, M.; Jenko, B.; Raspor, P.; Herron, P. R.; Hunter, I. S.*; et al.* Identification of the chelocardin biosynthetic gene cluster from *Amycolatopsis sulphurea*: a platform for producing novel tetracycline antibiotics. *Microbiology* **2013**, *159*, 2524–2532.

(21) He, Q.; Johnson, V. J.; Osuchowski, M. F.; Sharma, R. P. Inhibition of serine palmitoyltransferase by myriocin, a natural mycotoxin, causes induction of c-myc in mouse liver. *Mycopathologia* **2004**, *157*, 339–347.

(22) Nemec, K.; Schubert-Zsilavecz, M. Vom Teprotid zum Captopril. Rationales Design von ACE-Hemmern. *Pharmazie in unserer Zeit* **2003**, *32*, 11–16.

(23) Mack, R. N.; Lonsdale, W. M. Humans as Global Plant Dispersers: Getting More Than We Bargained For. *Bioscience* **2001**, *51*, 95–102.

(24) Petrovska, B. B. Historical review of medicinal plants' usage. *Pharmacognosy reviews* **2012**, *6*, 1–5.

(25) Varijakzhan, D.; Chong, C.-M.; Abushelaibi, A.; Lai, K.-S.; Lim, S.-H. E. Middle Eastern Plant Extracts: An Alternative to Modern Medicine Problems. *Molecules (Basel, Switzerland)* **2020**, *25*.

(26) Poulios, E.; Giaginis, C.; Vasios, G. K. Current State of the Art on the Antioxidant Activity of Sage (Salvia spp.) and Its Bioactive Components. *Planta medica* **2020**, *86*, 224–238.

(27) Charlesworth, J. C.; Burns, B. P. Untapped Resources: Biotechnological Potential of Peptides and Secondary Metabolites in Archaea. *Archaea (Vancouver, B.C.)* **2015**, *2015*, 282035.

(28) Rudolph, C.; Wanner, G.; Huber, R. Natural communities of novel archaea and bacteria growing in cold sulfurous springs with a string-of-pearls-like morphology. *Appl. Environ. Microbiol.* **2001**, *67*, 2336–2344.

(29) Schäberle, T. F.; Lohr, F.; Schmitz, A.; König, G. M. Antibiotics from myxobacteria. *Nat. Prod. Rep.* **2014**, *31*, 953–972.

(30) Jakubczyk, D.; Dussart, F. Selected Fungal Natural Products with Antimicrobial Properties. *Molecules (Basel, Switzerland)* **2020**, *25*.

(31) Houbraken, J.; Frisvad, J. C.; Samson, R. A. Fleming's penicillin producing strain is not *Penicillium chrysogenum* but *P. rubens*. *IMA Fungus* **2011**, *2*, 87–95.

(32) Meca-Lallana, J.; Ayuso, T.; Martínez-Yelamos, S.; Durán, C.; Contreras Martín, Y.; Herrera Navarro, N.; Pérez Sempere, A.; Álvarez-Cermeño, J. C.; Millán Pascual, J.; Meca-Lallana, V.*; et al.* Effectiveness of Fingolimod versus Natalizumab as Second-Line Therapy for Relapsing-Remitting Multiple Sclerosis in Spain: Second-Line GATE Study. *Eur. Neurol.* **2020**, 1–9.

(33) Heitman, J. Microbial Pathogens in the Fungal Kingdom. *Fungal Biol. Rev.* **2011**, *25*, 48–60.

(34) Hug, J. J.; Krug, D.; Müller, R. Bacteria as genetically programmable producers of bioactive natural products. *Nat. Rev. Chem.* **2020**, *4*, 172–193.

(35) Imai, Y.; Meyer, K. J.; Iinishi, A.; Favre-Godal, Q.; Green, R.; Manuse, S.; Caboni, M.; Mori, M.; Niles, S.; Ghiglieri, M.*; et al.* A new antibiotic selectively kills Gram-negative pathogens. *Nature* **2019**, *576*, 459–464.

(36) Bode, H. B. Entomopathogenic bacteria as a source of secondary metabolites. *Curr. Opin. Chem. Biol.* **2009**, *13*, 224–230.

(37) Bader, C. D.; Panter, F.; Müller, R. In depth natural product discovery - Myxobacterial strains that provided multiple secondary metabolites. *Biotechnol. Adv.* **2020**, *39*, 107480.

(38) Basi-Chipalu, S.; Dischinger, J.; Josten, M.; Szekat, C.; Zweynert, A.; Sahl, H.-G.; Bierbaum, G. Pseudomycoicidin, a Class II Lantibiotic from Bacillus pseudomycoides. *Appl. Environ. Microbiol.* **2015**, *81*, 3419–3429.

(39) Lewis, K. The Science of Antibiotic Discovery. *Cell* [Online early access]. DOI: 10.1016/j.cell.2020.02.056.

(40) 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.

(41) Berleman, J. E.; Kirby, J. R. Deciphering the hunting strategy of a bacterial wolfpack. *FEMS Microbiol. Rev.* **2009**, *33*, 942–957.

(42) Wenzel, S. C.; Müller, R. Myxobacteria—''microbial factories' for the production of bioactive secondary metabolites. *Mol. Biosyst.* **2009**, *5*, 567–574.

(43) Müller, R. Biosynthesis and heterologous production of epothilones. In *The Epothilones, an Outstanding Family of Anti-Tumor Agents*; Mulzer, J., Ed.; Springer: New York, 2009; pp 29–53.

(44) Ludwig, W.; Schleifer, K. H.; Reichenbach, H.; Stackebrandt, E. A Phylogenetic Analysis of the Myxobacteria *Myxococcus fulvus*, *Stigmatella aurantiaca*, *Cystobacter fuscus*, *Sorangium cellulosum* and *Nannocystis exedens*. *Arch. Microbiol.* **1983**, *135*, 58–62.

(45) Hoffmann, T.; Krug, D.; Bozkurt, N.; Duddela, S.; Jansen, R.; Garcia, R.; Gerth, K.; Steinmetz, H.; Müller, R. Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat. Commun.* **2018**, *9*, 803.

(46) Amiri Moghaddam, J.; Crüsemann, M.; Alanjary, M.; Harms, H.; Dávila-Céspedes, A.; Blom, J.; Poehlein, A.; Ziemert, N.; König, G. M.; Schäberle, T. F. Analysis of the Genome and Metabolome of Marine Myxobacteria Reveals High Potential for Biosynthesis of Novel Specialized Metabolites. *Sci. Rep.* **2018**, *8*, 16600.

(47) Meiser, P.; Bode, H. B.; Müller, R. The unique DKxanthene secondary metabolite family from the myxobacterium *Myxococcus xanthus* is required for developmental sporulation. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19128–19133.

(48) Krug, D.; Zurek, G.; Revermann, O.; Vos, M.; Velicer, G. J.; Müller, R. Discovering the Hidden Secondary Metabolome of *Myxococcus xanthus*: a Study of Intraspecific Diversity. *Appl. Environ. Microbiol.* **2008**, *74*, 3058–3068.

(49) Cortina, N. S.; Krug, D.; Plaza, A.; Revermann, O.; Müller, R. Myxoprincomide: a natural product from *Myxococcus xanthus* discovered by comprehensive analysis of the secondary metabolome. *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 811–816.

(50) Hoffmann, M.; Auerbach, D.; Panter, F.; Hoffmann, T.; Dorrestein, P. C.; Müller, R. Homospermidine Lipids: A compound class specifically formed during fruiting body formation of *Myxococcus xanthus* DK1622. *ACS Chem. Biol.* **2018**, *13*, 273–280.

(51) Hug, J. J.; Panter, F.; Krug, D.; Müller, R. Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria. *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 319–334.

(52) Irschik, H.; Jansen, R.; Gerth, K.; Hofle, G.; Reichenbach, H. The sorangicins, novel and powerful inhibitors of eubacterial RNA polymerase isolated from myxobacteria. *J. Antibiot.* **1987**, *40*, 7–13.

(53) Jansen, R.; Irschik, H.; Reichenbach, H.; Wray, V.; Höfle, G. Disorazoles, highly cytotoxic metabolites from the Sorangicin-producing bacterium *Sorangium cellulosum,* strain So ce12. *Liebigs Ann. Chem.* **1994**, *1994*, 759–773.

(54) Irschik, H.; Jansen, R.; Gerth, K.; Höfle, G.; Reichenbach, H. Chivosazol A, a new inhibitor of eukaryotic organisms isolated from myxobacteria. *J. Antibiot.* **1995**, *48*, 962–966.

(55) Irschik, H.; Jansen, R.; Gerth, K.; Höfle, G.; Reichenbach, H. Sorangiolid A, a new antibiotic isolated from the myxobacterium *Sorangium cellulosum* So ce 12. *J. Antibiot.* **1995**, *48*, 886–887.

(56) Zander, W.; Irschik, H.; Augustiniak, H.; Herrmann, M.; Jansen, R.; Steinmetz, H.; Gerth, K.; Kessler, W.; Kalesse, M.; Höfle, G.*; et al.* Sulfangolids, macrolide sulfate esters from *Sorangium cellulosum*. *Chem. Eur. J.* **2012**, *18*, 6264–6271.

(57) Pradella, S.; Hans, A.; Sproer, C.; Reichenbach, H.; Gerth, K.; Beyer, S. Characterisation, genome size and genetic manipulation of the myxobacterium Sorangium cellulosum So ce56. *Archives of Microbiology* **2002**, *178*, 484–492.

(58) Schneiker, S.; Perlova, O.; Kaiser, O.; Gerth, K.; Alici, A.; Altmeyer, M. O.; Bartels, D.; Bekel, T.; Beyer, S.; Bode, E.*; et al.* Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol.* **2007**, *25*, 1281–1289.

(59) Schifrin, A.; Ly, Thuy T B; Günnewich, N.; Zapp, J.; Thiel, V.; Schulz, S.; Hannemann, F.; Khatri, Y.; Bernhardt, R. Characterization of the Gene Cluster CYP264B1-*geoA* from *Sorangium cellulosum* So ce56: Biosynthesis of (+)-Eremophilene and Its Hydroxylation. *ChemBioChem* **2015**, *16*, 337–344.

(60) Mulwa, L.; Jansen, R.; Praditya, D.; Mohr, K.; Wink, J.; Steinmann, E.; Stadler, M. Six Heterocyclic Metabolites from the Myxobacterium Labilithrix luteola. *Molecules* **2018**, *23*, 542.

(61) Steinmetz, H.; Irschik, H.; Kunze, B.; Reichenbach, H.; Höfle, G.; Jansen, R. Thuggacins, macrolide antibiotics active against Mycobacterium tuberculosis: Isolation from myxobacteria, structure elucidation, conformation analysis and biosynthesis. *Chem. Eur. J.* **2007**, *13*, 5822–5832.

(62) Jansen, R.; Sood, S.; Mohr, K. I.; Kunze, B.; Irschik, H.; Stadler, M.; Müller, R. Nannozinones and Sorazinones, Unprecedented Pyrazinones from Myxobacteria. *J. Nat. Prod.* **2014**, *77*, 2545–2552.

(63) Gerth, K.; Steinmetz, H.; Höfle, G.; Jansen, R. Chlorotonil A, ein Macrolid mit einzigartiger gem-Dichlor-1,3-dionfunktion aus Sorangium cellulosum, So ce1525. *Angew. Chem.* **2008**, *120*, 610–613.

(64) Kunze, B.; Kemmer, T.; Höfle, G.; Reichenbach, H. Stigmatellin, a new antibiotic from *Stigmatella aurantiaca* (Myxobacterales). I. Production, physico-chemical and biological properties. *J. Antibiot.* **1984**, *37*, 454–461.

(65) Kunze, B.; Höfle, G.; Reichenbach, H. The aurachins, new quinoline antibiotics from myxobacteria: Production, physico-chemical and biological properties. *J Antibiot (Tokyo)* **1987**, *40*, 258–265.

(66) Dickschat, J. S.; Bode, H. B.; Wenzel, S. C.; Müller, R.; Schulz, S. Biosynthesis and identification of volatiles released by the myxobacterium *Stigmatella aurantiaca*. *ChemBioChem* **2005**, *6*, 2023–2033.

(67) Pistorius, D.; Müller, R. Discovery of the rhizopodin biosynthetic gene cluster in Stigmatella aurantiaca Sg a15 by genome mining. *ChemBioChem* **2012**, *13*, 416–426.

(68) Söker, U.; Kunze, B.; Reichenbach, H.; Höfle, G. Dawenol, a new polyene metabolite from the myxobacterium *Stigmatella aurantiaca*. *Z. Naturforsch. B* **2003**, *58*, 1024–1026.

(69) Weinig, S.; Hecht, H.-J.; Mahmud, T.; Müller, R. Melithiazol biosynthesis: further insights into myxobacterial PKS/NRPS systems and evidence for a new subclass of methyl transferases. *Chem. Biol.* **2003**, *10*, 939–952.

(70) Wenzel, S. C.; Kunze, B.; Höfle, G.; Silakowski, B.; Scharfe, M.; Blöcker, H.; Müller, R. Structure and Biosynthesis of Myxochromides S$_{1-3}$ in *Stigmatella aurantiaca*: Evidence for An Iterative Bacterial Type I Polyketide Synthase and for Module Skipping in Nonribosomal Peptide Biosynthesis. *ChemBioChem* **2005**, *6*, 375–385.

(71) Kunze, B.; Reichenbach, H.; Müller, R.; Höfle, G. Aurafuron A and B, new bioactive polyketides from *Stigmatella aurantiaca* and *Archangium gephyra* (myxobacteria). *J. Antibiot.* **2005**, *58*, 244–251.

(72) Gulder, T. A. M.; Neft, S.; Schüz, T.; Winkler, T.; Gees, R.; Böhlendorf, B. The myxocoumarins A and B from *Stigmatella aurantiaca* strain MYX-030. *Beilstein J. Org. Chem.* **2013**, *9*, 2579–2585.

(73) Sasse, F.; Kunze, B.; Gronewold, T. M.; Reichenbach, H. The chondramides: cytostatic agents from myxobacteria acting on the actin cytoskeleton. *J Natl Cancer Inst* **1998**, *90*, 1559–1563.

(74) Jansen, R.; Kunze, B.; Reichenbach, H.; Höfle, G. The ajudazols A and B, novel isochromanones from *Chondromyces crocatus* (Myxobacteria): Isolation and structure elucidation. *Eur. J. Org. Chem.* **2002**, 917–921.

(75) Jansen, R.; Kunze, B.; Reichenbach, H.; Höfle, G. Chondrochloren A and B, New β-Amino Styrenes from *Chondromyces crocatus* (Myxobacteria). *Eur. J. Org. Chem.* **2003**, *2003*, 2684–2689.

(76) Nawrath, T.; Gerth, K.; Müller, R.; Schulz, S. Volatile Methyl Esters of Medium Chain Length from the Bacterium *Chitinophaga* Fx7914. *Chemistry & Biodiversity* **2010**, *7*, 2228–2253.

(77) Viehrig, K.; Surup, F.; Harmrolfs, K.; Jansen, R.; Kunze, B.; Müller, R. Concerted action of P450 plus helper protein to form the amino-hydroxy-piperidone moiety of the potent protease inhibitor crocapeptin. *J. Am. Chem. Soc.* **2013**, *135*, 16885–16894.

(78) Viehrig, K.; Surup, F.; Volz, C.; Herrmann, J.; Abou Fayad, A.; Adam, S.; Kohnke, J.; Trauner, D.; Müller, R. Structure and biosynthesis of crocagins: polycyclic postranslationally modified ribosomal peptides from *Chondromyces crocatus*. *Angew. Chem.* **2017**, 1–5.

(79) Surup, F.; Viehrig, K.; Rachid, S.; Plaza, A.; Maurer, C. K.; Hartmann, R. W.; Müller, R. Crocadepsins-Depsipeptides from the Myxobacterium *Chondromyces crocatus* Found by a Genome Mining Approach. *ACS Chem. Biol.* **2018**, *13*, 267–272.

(80) Schmitz, A.; Kehraus, S.; Schäberle, T. F.; Neu, E.; Almeida, C.; Roth, M.; König, G. M. Corallorazines from the myxobacterium *Corallococcus coralloides*. *J. Nat. Prod.* **2014**, *77*, 159–163.

(81) Felder, S.; Dreisigacker, S.; Kehraus, S.; Neu, E.; Bierbaum, G.; Wright, P. R.; Menche, D.; Schäberle, T. F.; König, G. M. Salimabromide: unexpected chemistry from the obligate marine myxobacterium *Enhygromxya salina*. *Chem. Eur. J.* **2013**, *19*, 9319–9324.

(82) Felder, S.; Kehraus, S.; Neu, E.; Bierbaum, G.; Schäberle, T. F.; König, G. M. Salimyxins and enhygrolides: antibiotic, sponge-related metabolites from the obligate marine myxobacterium *Enhygromyxa salina*. *ChemBioChem* **2013**, *14*, 1363–1371.

(83) Fudou, R.; Iizuka, T.; Yamanaka, S. Haliangicin, a novel antifungal metabolite produced by a marine myxobacterium. 1. Fermentation and biological characteristics. *J. Antibiot.* **2001**, *54*, 149–152.

(84) Sun, Y.; Feng, Z.; Tomura, T.; Suzuki, A.; Miyano, S.; Tsuge, T.; Mori, H.; Suh, J. W.; Iizuka, T.; Fudou, R*.; et al.* Heterologous Production of the Marine Myxobacterial Antibiotic Haliangicin and Its Unnatural Analogues Generated by Engineering of the Biochemical Pathway. *Sci. Rep.* **2016**, *6*, 22091.

(85) Chai, Y.; Pistorius, D.; Ullrich, A.; Weissman, K. J.; Kazmaier, U.; Müller, R. Discovery of 23 natural tubulysins from *Angiococcus disciformis* An d48 and *Cystobacter* SBCb004. *Chem. Biol.* **2010**, *17*, 296–309.

(86) Sasse, F.; Steinmetz, H.; Heil, J.; Höfle, G.; Reichenbach, H. Tubulysins, new cytostatic peptides from myxobacteria acting on microtubuli. Production, isolation, physico-chemical and biological properties. *J. Antibiot.* **2000**, *53*, 879–885.

(87) Panter, F.; Krug, D.; Baumann, S.; Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **2018**, *9*, 4898–4908.

(88) Okanya, P. W.; Mohr, K. I.; Gerth, K.; Steinmetz, H.; Huch, V.; Jansen, R.; Müller, R. Hyaladione, an S-methyl cyclohexadiene-dione from *Hyalangium minutum*. *J. Nat. Prod.* **2012**, *75*, 768–770.

(89) Okanya, P. W.; Mohr, K. I.; Gerth, K.; Kessler, W.; Jansen, R.; Stadler, M.; Müller, R. Hyafurones, hyapyrrolines, and hyapyrones: polyketides from *Hyalangium minutum*. *J. Nat. Prod.* **2014**, *77*, 1420–1429.

(90) Etzbach, L.; Plaza, A.; Garcia, R.; Baumann, S.; Müller, R. Cystomanamides: structure and biosynthetic pathway of a family of glycosylated lipopeptides from myxobacteria. *Org. Lett.* **2014**, *16*, 2414–2417.

(91) Keller, L.; Plaza, A.; Dubiella, C.; Groll, M.; Kaiser, M.; Müller, R. Macyranones: Structure, Biosynthesis, and Binding Mode of an Unprecedented Epoxyketone that Targets the 20S Proteasome. *J. Am. Chem. Soc.* **2015**, *137*, 8121–8130.

(92) Jansen, R.; Sood, S.; Huch, V.; Kunze, B.; Stadler, M.; Müller, R. Pyrronazols, metabolites from the myxobacteria *Nannocystis pusilla* and *N. exedens*, are unusual chlorinated pyrone-oxazole-pyrroles. *J. Nat. Prod.* **2014**, *77*, 320–326.

(93) Irschik, H.; Kopp, M.; Weissman, K. J.; Buntin, K.; Piel, J.; Müller, R. Analysis of the sorangicin gene cluster reinforces the utility of a combined phylogenetic/retrobiosynthetic analysis for deciphering natural product assembly by trans-AT PKS. *ChemBioChem* **2010**, *11*, 1840–1849.

(94) Dickschat, J. S.; Bode, H. B.; Mahmud, T.; Müller, R.; Schulz, S. A novel type of geosmin biosynthesis in myxobacteria. *J. Org. Chem.* **2005**, *70*, 5174–5182.

(95) Raju, R.; Mohr, K. I.; Bernecker, S.; Herrmann, J.; Müller, R. Cystodienoic acid: A new diterpene isolated from the myxobacterium *Cystobacter sp. J. Antibiot.* **2015**, *68*, 473–475.

(96) Cane, D. E.; Watt, R. M. Expression and mechanistic analysis of a germacradienol synthase from Streptomyces coelicolor implicated in geosmin biosynthesis. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 1547–1551.

(97) Hug, J. J.; Dastbaz, J.; Adam, S.; Revermann, O.; Koehnke, J.; Krug, D.; Müller, R. Biosynthesis of Cittilins, Unusual Ribosomally Synthesized and Post-translationally Modified Peptides from Myxococcus xanthus. *ACS Chem. Biol.* **2020**, *15*, 2221–2231.

(98) Herrmann, J.; Fayad, A. A.; Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **2017**, *34*, 135–160.

(99) Gaitatzis, N.; Silakowski, B.; Kunze, B.; Nordsiek, G.; Blöcker, H.; Höfle, G.; Müller, R. The biosynthesis of the aromatic myxobacterial electron transport inhibitor stigmatellin is directed by a novel type of modular polyketide synthase. *J. Biol. Chem.* **2002**, *277*, 13082–13090.

(100) Beyer, S.; Kunze, B.; Silakowski, B.; Müller, R. Metabolic diversity in myxobacteria: identification of the myxalamid and the stigmatellin biosynthetic gene cluster of *Stigmatella aurantiaca* Sg a15 and a combined polyketide-(poly)peptide gene cluster from the epothilone producing strain *Sorangium cellulosum* So ce90. *Biochim. Biophys. Acta* **1999**, *1445*, 185–195.

(101) Jansen, R.; Irschik, H.; Reichenbach, H.; Höfle, G. Antibiotics from gliding bacteria, LXXX. Chivosazoles A-F: Novel antifungal and cytotoxic macrolides from *Sorangium cellulosum* (Myxobacteria). *Liebigs Ann. Chem.* **1997**, *1997*, 1725–1732.

(102) Zhang, Q.; Pang, B.; Ding, W.; Liu, W. Aromatic Polyketides Produced by Bacterial Iterative Type I Polyketide Synthases. *ACS Catal.* **2013**, *3*, 1439–1447.

(103) Chen, X.; Ji, R.; Jiang, X.; Yang, R.; Liu, F.; Xin, Y. Iterative type I polyketide synthases involved in enediyne natural product biosynthesis. *IUBMB Life* **2014**, *66*, 587–595.

(104) Hari, Taylor P A; Labana, P.; Boileau, M.; Boddy, C. N. An evolutionary model encompassing substrate specificity and reactivity of type I polyketide synthase thioesterases. *ChemBioChem* **2014**, *15*, 2656–2661.

(105) Keating-Clay, A. T. The structures of type I polyketide synthases. *Nat. Prod. Rep.* **2012**, *29*, 1050–1073.

(106) Zheng, J.; Keatinge-Clay, A. T. The status of type I polyketide synthase ketoreductases (REVIEW). *Med. Chem. Commun.* **2013**, *2013*, 34–40.

(107) Chen, H.; Du, L. Iterative Polyketide Biosynthesis by Modular Polyketide Synthases in Bacteria. *Appl. Microbiol. Biotechnol.* **2016**, *100*, 541–557.

(108) Dreier, J.; Khosla, C. Mechanistic analysis of a type II polyketide synthase. Role of conserved residues in the beta-ketoacyl synthase-chain length factor heterodimer. *Biochemistry* **2000**, *39*, 2088–2095.

(109) Chemler, J. A.; Buchholz, T. J.; Geders, T. W.; Akey, D. L.; Rath, C. M.; Chlipala, G. E.; Smith, J. L.; Sherman, D. H. Biochemical and structural characterization of germicidin synthase: Analysis of a type III polyketide synthase that employs acyl-ACP as a starter unit donor. *J. Am. Chem. Soc.* **2012**, *134*, 7359–7366.

(110) Funa, N.; Ohnishi, Y.; Ebizuka, Y.; Horinouchi, S. Alteration of reaction and substrate specificity of a bacterial type III polyketide synthase by site-directed mutagenesis. *Biochemical Journal* **2002**, *367*, 781–789.

(111) Chen, W.-H.; Li, K.; Guntaka, N. S.; Bruner, S. D. Interdomain and intermodule organization in epimerization domain containing nonribosomal peptide synthetases. *ACS Chem. Biol.* [Online early access]. DOI: 10.1021/acschembio.6b00332.

(112) Finking, R.; Marahiel, M. A. Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.* **2004**, *58*, 453–488.

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

(114) Trauger, J. W.; Kohli, R. M.; Mootz, H. D.; Marahiel, M. A.; Walsh, C. T. Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature* **2000**, *407*, 215–218.

(115) Kiefer, A.; Bader, C. D.; Held, J.; Esser, A.; Rybniker, J.; Empting, M.; Müller, R.; Kazmaier, U. Synthesis of New Cyclomarin Derivatives and Their Biological Evaluation towards Mycobacterium Tuberculosis and Plasmodium Falciparum. *Chem. Eur. J.* **2019**, *25*, 8894–8902.

(116) Dunbar, K. L.; Mitchell, D. A. Revealing nature's synthetic potential through the study of ribosomal natural product biosynthesis. *ACS Chem. Biol.* **2013**, *8*, 473–487.

(117) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.*; et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **2013**, *30*, 108–160.

(118) Truman, A. W. Cyclisation mechanisms in the biosynthesis of ribosomally synthesised and post-translationally modified peptides. *Beilstein J. Org. Chem.* **2016**, *12*, 1250–1268.

(119) Ortega, M. A.; van der Donk, Wilfred A. New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-translationally Modified Peptide Natural Products. *Cell Chem. Biol.* **2016**, *23*, 31–44.

(120) Otaka, T.; Kaji, A. Mode of action of bottromycin-A2 - effect of bottromycin-A2 on polysomes. *FEBS Lett.* **1983**, *153*, 53–59.

(121) Krug, D.; Zurek, G.; Schneider, B.; Garcia, R.; Müller, R. Efficient mining of myxobacterial metabolite profiles enabled by liquid chromatography-electrospray ionization-time-of-flight mass spectrometry and compound-based principal component analysis. *Anal. Chim. Acta* **2008**, *624*, 97–106.

(122) Ney, I.; Jäger, E.; Herderich, M.; Schreier, P.; Schwab, W. Analysis of Phosphorylated Terpenes by Electrospray Ionization Liquid Chromatography–Tandem Mass Spectrometry. *Phytochemical Analysis* **1996**, *7*, 233–236.

(123) Kumar, S.; Singh, A.; Kumar, B. Identification and characterization of phenolics and terpenoids from ethanolic extracts of Phyllanthus species by HPLC-ESI-QTOF-MS/MS. *J Pharm Anal* **2017**, *7*, 214–222.

(124) Bode, H. B.; Ring, M. W.; Schwär, G.; Altmeyer, M. O.; Kegler, C.; Jose, I. R.; Singer, M.; Müller, R. Identification of additional players in the alternative biosynthesis pathway to isovaleryl-CoA in the myxobacterium *Myxococcus xanthus*. *ChemBioChem* **2009**, *10*, 128–140.

(125) Dairi, T. Studies on biosynthetic genes and enzymes of isoprenoids produced by actinomycetes. *J. Antibiot.* **2005**, *58*, 227–243.

(126) Dewick, P. M. The biosynthesis of C5-C25 terpenoid compounds. *Nat. Prod. Rep.* **2002**, *19*, 181–222.

(127) Chang, S.-Y.; Ko, T.-P.; Chen, A. P.-C.; Wang, A. H.-J.; Liang, P.-H. Substrate binding mode and reaction mechanism of undecaprenyl pyrophosphate synthase deduced from crystallographic studies. *Protein Sci.* **2004**, *13*, 971–978.

(128) Lu, S.; Wang, J.; Chitsaz, F.; Derbyshire, M. K.; Geer, R. C.; Gonzales, N. R.; Gwadz, M.; Di Hurwitz; Marchler, G. H.; Song, J. S.*; et al.* CDD/SPARCLE: The Conserved Domain Database in 2020. *Nucleic Acids Res.* **2020**, *48*.

(129) Liang, P.-H.; Ko, T.-P.; Wang, A. H.-J. Structure, mechanism and function of prenyltransferases. *Eur. J. Clin. Invest.* **2002**, *269*, 3339–3354.

(130) Helfrich, E. J. N.; Lin, G.-M.; Voigt, C. A.; Clardy, J. Bacterial terpene biosynthesis: challenges and opportunities for pathway engineering. *Beilstein J. Org. Chem.* **2019**, *15*, 2889–2906.

(131) Wei, J. H.; Yin, X.; Welander, P. V. Sterol Synthesis in Diverse Bacteria. *Front. Microbiol.* **2016**, *7*.

(132) Satoshi Takamatsu; Lian-Hua Xu; Shinya Fushinobu; Hirofumi Shoun; Mamoru Komatsu; David E Cane; Haruo Ikeda. Pentalenic acid is a shunt metabolite in the biosynthesis of the pentalenolactone family of metabolites: hydroxylation of 1-deoxypentalenic acid mediated by CYP105D7 (SAV_7469) of Streptomyces avermitilis. *J Antibiot* **2011**, *64*, 65–71.

(133) Brodelius, M.; Lundgren, A.; Mercke, P.; Brodelius, P. E. Fusion of Farnesyldiphosphate Synthase and Epi-Aristolochene Synthase, a Sesquiterpene Cyclase Involved in Capsidiol Biosynthesis in Nicotiana Tabacum. *FEBS J* **2002**, *269*.

(134) Back, K.; Nah, J.; Lee, S. B.; Song, J. H.; Shin, D. H.; Kim, H. Y. Cloning of a Sesquiterpene Cyclase and Its Functional Expression by Domain Swapping Strategy. *Mol. Cells* **2000**, *10*.

(135) Tomura, T.; Nagashima, S.; Yamazaki, S.; Iizuka, T.; Fudou, R.; Ojika, M. An Unusual Diterpene-Enhygromic Acid and Deoxyenhygrolides From a Marine Myxobacterium, Enhygromyxa Sp. *Marine drugs* **2017**, *15*.

(136) Du, L.; Shen, B. Biosynthesis of hybrid peptide-polyketide natural products. *Curr. Opin. Drug Discov. Devel.* **2001**, *4*, 215–228.

(137) Rox, K.; Rohde, M.; Chhatwal, G. S.; Müller, R. Disorazoles block group A streptococcal invasion into epithelial cells via interference with the host factor ezrin. *Cell Chem. Biol.* **2017**, *24*, 159–170.

(138) Plaza, A.; Garcia, R.; Bifulco, G.; Martinez, J. P.; Hüttel, S.; Sasse, F.; Meyerhans, A.; Stadler, M.; Müller, R. Aetheramides A and B, potent HIV-inhibitory depsipeptides from a myxobacterium of the new genus "*Aetherobacter*". *Org. Lett.* **2012**, *14*, 2854–2857.

(139) Rajachan, O.-A.; Kanokmedhakul, K.; Sanmanoch, W.; Boonlue, S.; Hannongbua, S.; Saparpakorn, P.; Kanokmedhakul, S. Chevalone C analogues and globoscinic acid derivatives from the fungus Neosartorya spinosa KKU-1NK1. *Phytochemistry* **2016**, *132*, 68–75.

(140) Yee, D. A.; Kakule, T. B.; Cheng, W.; Chen, M.; Chong, C. T. Y.; Hai, Y.; Hang, L. F.; Hung, Y.-S.; Liu, N.; Ohashi, M.*; et al.* Genome Mining of Alkaloidal Terpenoids from a Hybrid Terpene and Nonribosomal Peptide Biosynthetic Pathway. *J. Am. Chem. Soc.* **2020**, *142*, 710–714.

(141) Kopp, M.; Irschik, H.; Gemperlein, K.; Buntin, K.; Meiser, P.; Weissman, K. J.; Bode, H. B.; Müller, R. Insights into the complex biosynthesis of the leupyrrins in *Sorangium cellulosum* So ce690. *Mol. Biosyst.* **2011**, *7*, 1549–1563.

(142) Bode, H. B.; Irschik, H.; Wenzel, S. C.; Reichenbach, H.; Müller, R.; Höfle, G. The leupyrrins: A structurally unique family of secondary metabolites from the myxobacterium *Sorangium cellulosum*. *J. Nat. Prod.* **2003**, *66*, 1203–1206.

(143) Gregory, K.; Salvador, L. A.; Akbar, S.; Adaikpoh, B. I.; Stevens, D. C. Survey of Biosynthetic Gene Clusters from Sequenced Myxobacteria Reveals Unexplored Biosynthetic Potential. *Microorganisms* **2019**, *7*.

(144) Baral, B.; Akhgari, A.; Metsä-Ketelä, M. Activation of microbial secondary metabolic pathways: Avenues and challenges. *Synth. Syst. Biotechnol.* **2018**, *3*, 163–178.

(145) Hug, J. J.; Bader, C. D.; Remškar, M.; Cirnski, K.; Müller, R. Concepts and Methods to Access Novel Antibiotics from Actinomycetes. *Antibiotics* **2018**, *7*, 44.

(146) Gaudêncio, S. P.; Pereira, F. Dereplication: racing to speed up the natural products discovery process. *Nat. Prod. Rep.* **2015**, *32*, 779–810.

(147) Heller, A. A.; Spence, D. M. A rapid method for post-antibiotic bacterial susceptibility testing. *PLoS ONE* **2019**, *14*.

(148) Hostettmann, K.; Wolfender, J. L.; Terreaux, C. Modern screening techniques for plant extracts. *Pharm Biol* **2001**, *39 Suppl 1*, 18–32.

(149) Abdelmohsen, U. R.; Grkovic, T.; Balasubramanian, S.; Kamel, M. S.; Quinn, R. J.; Hentschel, U. Elicitation of secondary metabolism in actinomycetes. *Biotechnol. Adv.* **2015**, *33*, 798–811.

(150) Herrmann, J.; Lukezic, T.; Kling, A.; Baumann, S.; Hüttel, S.; Petkovic, H.; Müller, R. Strategies for the discovery and development of new antibiotics from natural products: Three case studies. *Curr. Top. Microbiol. Immunol.* **2016**, *398*, 339–363.

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

(152) Li, X. W.; Herrmann, J.; Zang, Y.; Grellier, P.; Prado, S.; Müller, R.; Nay, B. Synthesis and biological activities of the respiratory chain inhibitor aurachin D and new ring versus chain analogues. *Beilstein J. Org. Chem.* **2013**, *9*, 1551–1558.

(153) Höfle, G.; Böhlendorf, B.; Fecker, T.; Sasse, F.; Kunze, B. Semisynthesis and antiplasmodial activity of the quinoline alkaloid aurachin E. *J. Nat. Prod.* [Online early access]. DOI: 10.1021/np8004612.

(154) Keller, L.; Siqueira-Neto, J. L.; Souza, J. M.; Eribez, K.; LaMonte, G. M.; Smith, J. E.; Gerwick, W. H. Palstimolide A: A Complex Polyhydroxy Macrolide with Antiparasitic Activity. *Molecules* **2020**, *25*.

(155) Bumpus, S. B.; Kelleher, N. L. Accessing natural product biosynthetic processes by mass spectrometry. *Curr. Opin. Chem. Biol.* **2008**, *12*, 475–482.

(156) Carlomagno, T. NMR in natural products: understanding conformation, configuration and receptor interactions. *Nat. Prod. Rep.* **2012**, *29*, 536–554.

(157) Robinette, S. L.; Brüschweiler, R.; Schroeder, F. C.; Edison, A. S. NMR in metabolomics and natural products research: two sides of the same coin. *Acc. Chem. Res.* **2012**, *45*, 288–297.

(158) Lin, Y.; Schiavo, S.; Orjala, J.; Vouros, P.; Kautz, R. Microscale LC-MS-NMR platform applied to the identification of active cyanobacterial metabolites. *Anal. Chem.* **2008**, *80*, 8045–8054.

(159) Avula, B.; Wang, Y.-H.; Smillie, T. J.; Mabusela, W.; Vincent, L.; Weitz, F.; Khan, I. A. Comparison of LC–UV, LC–ELSD and LC–MS Methods for the Determination of Sesquiterpenoids in Various Species of Artemisia. *Chromatographia* **2009**, *70*, 797–803.

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

(161) Son, S. Y.; Lee, S.; Singh, D.; Lee, N.-R.; Lee, D.-Y.; Lee, C. H. Comprehensive Secondary Metabolite Profiling Toward Delineating the Solid and Submerged-State Fermentation of *Aspergillus oryzae* KCCM 12698. *Front. Microbiol.* **2018**, *9*, 1076.

(162) Byrdwell, W. C.; Neff, W. E. Dual parallel electrospray ionization and atmospheric pressure chemical ionization mass spectrometry (MS), MS/MS and MS/MS/MS for the analysis of triacylglycerols and triacylglycerol oxidation products. *Rapid Commun. Mass Spectrom.* **2002**, *16*, 300–319.

(163) Griffiths, J. A brief history of mass spectrometry. *Anal. Chem.* **2008**, *80*, 5678–5683.

(164) Scheltema, R. A.; Hauschild, J.-P.; Lange, O.; Hornburg, D.; Denisov, E.; Damoc, E.; Kuehn, A.; Makarov, A.; Mann, M. The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer. *Mol. Cell Proteomics* **2014**, *13*, 3698–3708.

(165) Haag, A. M. Mass Analyzers and Mass Spectrometers. *Adv. Exp. Med. Biol.* **2016**, *919*, 157–169.

(166) Züllig, T.; Köfeler, H. C. High resolution mass spectrometry in lipidomics. *Mass Spectrom. Rev.* [Online early access]. DOI: 10.1002/mas.21627.

(167) Holcapek, M.; Kolarova, L.; Nobilis, M. High-performance liquid chromatography-tandem mass spectrometry in the identification and determination of phase I and phase II drug metabolites. *Anal. Bioanal. Chem* **2008**, *391*, 59–78.

(168) Gallardo, E.; Barroso, M.; Queiroz, J. A. LC-MS: a powerful tool in workplace drug testing. *Drug Test Anal* **2009**, *1*, 109–115.

(169) Ghaste, M.; Mistrik, R.; Shulaev, V. Applications of Fourier Transform Ion Cyclotron Resonance (FT-ICR) and Orbitrap Based High Resolution Mass Spectrometry in Metabolomics and Lipidomics. *Int. J. Mol. Sci.* **2016**, *17*.

(170) Nikolaev, E. N.; Kostyukevich, Y. I.; Vladimirov, G. N. Fourier transform ion cyclotron resonance (FT ICR) mass spectrometry: Theory and simulations. *Mass Spectrom. Rev.* **2016**, *35*, 219–258.

(171) Zenobi, R. Ionization methods in mass spectrometry. *Anal. Chem.* **2015**, *87*, 3543.

(172) Konermann, L.; Metwally, H.; Duez, Q.; Peters, I. Charging and supercharging of proteins for mass spectrometry: recent insights into the mechanisms of electrospray ionization. *The Analyst* **2019**, *144*, 6157–6171.

(173) Kiontke, A.; Billig, S.; Birkemeyer, C. Response in Ambient Low Temperature Plasma Ionization Compared to Electrospray and Atmospheric Pressure Chemical Ionization for Mass Spectrometry. *Int J Anal Chem* **2018**, *2018*, 5647536.

(174) Israr, M. Z.; Bernieh, D.; Salzano, A.; Cassambai, S.; Yazaki, Y.; Suzuki, T. Matrix-assisted laser desorption ionisation (MALDI) mass spectrometry (MS): basics and clinical applications. *Clin. Chem. Lab. Med.* [Online early access]. DOI: 10.1515/cclm-2019-0868.

(175) Hoffmann, T.; Krug, D.; Hüttel, S.; Müller, R. Improving natural products identification through targeted LC-MS/MS in an untargeted secondary metabolomics workflow. *Anal. Chem.* **2014**, *86*, 10780–10788.

(176) Zhang, F.; Ge, W.; Ruan, G.; Cai, X.; Guo, T. Data-Independent Acquisition Mass Spectrometry-based Proteomics and Software Tools: A Glimpse in 2020. *Proteomics* **2020**, e1900276.

(177) Phelan, V. V. Feature-Based Molecular Networking for Metabolite Annotation. *Methods Mol. Biol.* **2020**, *2104*, 227–243.

(178) Kruve, A.; Kaupmees, K. Adduct Formation in ESI/MS by Mobile Phase Additives. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 887–894.

(179) Bader, C. D.; Haack, P. A.; Panter, F.; Krug, D.; Müller, R. Expanding the Scope of Detectable Microbial Natural Products by Complementary Analytical Methods and Cultivation Systems. *Journal of natural products* [Online early access]. DOI: 10.1021/acs.jnatprod.0c00942. Published Online: Jan. 15, 2021.

(180) Krug, D.; Zurek, G.; Schneider, B.; Bässmann, C.; Müller, R. *Analysis of secondary metabolites from myxobacteria using ESI-TOF-MS and PCA* March 2007, 2007.

(181) Barsch, A.; Zurek, G.; Krug, D.; Cortina, N. S.; Müller, R. *Challenges in Metabolomics addressed by targeted and untargeted UHR-Q-TOF analysis*, 2010.

(182) 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.

(183) Vargas, F.; Weldon, K. C.; Sikora, N.; Wang, M.; Zhang, Z.; Gentry, E. C.; Panitchpakdi, M. W.; Caraballo-Rodríguez, A. M.; Dorrestein, P. C.; Jarmusch, A. K. Protocol for Community-created Public MS/MS Reference Spectra Within the Global Natural Products Social Molecular Networking Infrastructure. *Rapid Commun. Mass Spectrom.* **2020**, e8725.

(184) Amoutzias, G. D.; Chaliotis, A.; Mossialos, D. Discovery Strategies of Bioactive Compounds Synthesized by Nonribosomal Peptide Synthetases and Type-I Polyketide Synthases Derived from Marine Microbiomes. *Marine drugs* **2016**, *14*.

(185) Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **2019**, W81-W87.

(186) Fedorova, N. D.; Moktali, V.; Medema, M. H. Bioinformatics approaches and software for detection of secondary metabolic gene clusters. *Methods Mol. Biol.* **2012**, *2012*, 23–45.

(187) Skinnider, M. A.; Merwin, N. J.; Johnston, C. W.; Magarvey, N. A. PRISM 3: Expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res* **2017**, *45*, W49-W54.

(188) Moreno, M.; Elgaher, W.; Herrmann, J.; Schläger, N.; Hamed, M.; Baumann, S.; Müller, R.; Hartmann, R.; Kirschning, A. Synthesis and biological evaluation of cystobactamid 507: a bacterial topoisomerase inhibitor from *Cystobacter sp*. *Synlett* **2015**, *26*, 1175–1178.

(189) Baumann, S.; Herrmann, J.; Raju, R.; Steinmetz, H.; Mohr, K. I.; Hüttel, S.; Harmrolfs, K.; Stadler, M.; Müller, R. Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *Angew. Chem. Int. Ed.* **2014**, *53*, 14605–14609.

(190) Sucipto, H.; Pogorevc, D.; Luxenburger, E.; Wenzel, S. C.; Müller, R. Heterologous production of myxobacterial α-pyrone antibiotics in *Myxococcus xanthus*. *Metab. Eng.* **2017**, *44*, 160–170.

(191) Breton, R. C.; Reynolds, W. F. Using NMR to identify and characterize natural products. *Nat. Prod. Rep.* **2013**, *30*, 501–524.

(192) Nicolaou, K. C.; Snyder, S. A. Chasing molecules that were never there: misassigned natural products and the role of chemical synthesis in modern structure elucidation. *Angew. Chem. Int. Ed.* **2005**, *44*, 1012–1044.

(193) Phillips, T.; Chase, M.; Wagner, S.; Renzi, C.; Powell, M.; DeAngelo, J.; Michels, P. Use of in situ solid-phase adsorption in microbial natural product fermentation development. *J. Ind. Microbiol. Biotechnol.* **2013**, *40*, 411–425.

(194) Majik, M. S.; Gawas, U. B.; Mandrekar, V. K. Analytical methods for natural products isolation. In *Advances in biological science research: A practical approach*; Meena, S. N., Naik, M. M., Eds.; Academic Press, an imprint of Elsevier: London, 2019; pp 395–409.

(195) Gerth, K.; Müller, R. Moderately thermophilic myxobacteria: novel potential for the production of natural products isolation and characterization. *Environ. Microbiol.* **2005**, *7*, 874–880.

(196) Bart, H.-J. Extraction of Natural Products from Plants – An Introduction. *Industrial Scale Natural Products Extraction*; John Wiley & Sons, Ltd; pp 1–25.

(197) Wakefield, J.; Hassan, H. M.; Jaspars, M.; Ebel, R.; Rateb, M. E. Dual Induction of New Microbial Secondary Metabolites by Fungal Bacterial Co-cultivation. *Front. Microbiol.* **2017**, *8*, 1284.

(198) Gibitz Eisath, N.; Sturm, S.; Stuppner, H. Supercritical Fluid Chromatography in Natural Product Analysis - An Update. *Planta Med.* [Online early access]. DOI: 10.1055/s-0037-1599461 (accessed January 26, 2018).

(199) Lemasson, E.; Bertin, S.; West, C. Use and practice of achiral and chiral supercritical fluid chromatography in pharmaceutical analysis and purification. *J Sep Sci* **2016**, *39*, 212–233.

(200) Wrona, O.; Rafińska, K.; Możeński, C.; Buszewski, B. Supercritical Fluid Extraction of Bioactive Compounds from Plant Materials. *J. AOAC Int.* **2017**, *100*, 1624–1635.

(201) Pyo, D.; Shin, H. Supercritical fluid extraction of microcystins from cyanobacteria. *Anal. Chem.* **1999**, *71*, 4772–4775.

(202) Yang, X.; Li, Y.; Li, Y.; Ye, D.; Yuan, L.; Sun, Y.; Han, D.; Hu, Q. Solid Matrix-Supported Supercritical CO_2 Enhances Extraction of γ-Linolenic Acid from the Cyanobacterium Arthrospira (Spirulina) platensis and Bioactivity Evaluation of the Molecule in Zebrafish. *Marine drugs* **2019**, *17*.

(203) Machida, H.; Takesue, M.; Smith, R. L. Green chemical processes with supercritical fluids: Properties, materials, separations and energy. *J Supercrit Fluids* **2011**, *60*, 2–15.

(204) Carlès, P. A brief review of the thermophysical properties of supercritical fluids. *J Supercrit Fluids* **2010**, *53*, 2–11.

(205) Zhang, Y.; Zhang, G.; Zhang, A.; Jin, Y.; Ru, R.; Tian, M. Frosting Phenomenon and Frost-Free Technology of Outdoor Air Heat Exchanger for an Air-Source Heat Pump System in China: An Analysis and Review. *Energies* **2018**, *11*, 2642.

(206) Wells, S. L.; DeSimone, J. CO2 Technology Platform: An Important Tool for Environmental Problem Solving. *Angew. Chem. Int. Ed.* **2001**, *40*, 518–527.

(207) Esquivel-Hernández, D. A.; Rodríguez-Rodríguez, J.; Cuéllar-Bermúdez, S. P.; García-Pérez, J. S.; Mancera-Andrade, E. I.; Núñez-Echevarría, J. E.; Ontiveros-Valencia, A.; Rostro-Alanis, M.; García-García, R. M.; Torres, J. A.*; et al.* Effect of Supercritical Carbon Dioxide Extraction Parameters on the Biological Activities and Metabolites Present in Extracts from Arthrospira platensis. *Marine drugs* **2017**, *15*.

(208) Rafińska, K.; Pomastowski, P.; Rudnicka, J.; Krakowska, A.; Maruśka, A.; Narkute, M.; Buszewski, B. Effect of solvent and extraction technique on composition and biological activity of Lepidium sativum extracts. *Food Chem.* **2019**, *289*, 16–25.

(209) Khaw, K.-Y.; Parat, M.-O.; Shaw, P. N.; Falconer, J. R. Solvent Supercritical Fluid Technologies to Extract Bioactive Compounds from Natural Sources: A Review. *Molecules (Basel, Switzerland)* **2017**, *22*.

(210) RAMSEY, E. D. *Analytical Supercritical Fluid Extraction Techniques*; Springer Netherlands: Dordrecht, 2012.

(211) Portet-Koltalo, F.; Oukebdane, K.; Dionnet, F.; Desbène, P. L. Optimisation of supercritical fluid extraction of polycyclic aromatic hydrocarbons and their nitrated derivatives adsorbed on highly sorptive diesel particulate matter. *Anal. Chim. Acta* **2009**, *651*, 48–56.

(212) Novakova, L.; Perrenoud, A. G.-G.; Francois, I.; West, C.; Lesellier, E.; Guillarme, D. Modern analytical supercritical fluid chromatography using columns packed with sub-2 μm particles: a tutorial. *Anal. Chim. Acta* **2014**, *824*, 18–35.

(213) Lemasson, E.; Bertin, S.; Hennig, P.; Lesellier, E.; West, C. Comparison of ultra-high performance methods in liquid and supercritical fluid chromatography coupled to electrospray ionization - mass

spectrometry for impurity profiling of drug candidates. *Journal of chromatography. A* **2016**, *1472*, 117–128.

(214) West, C.; Lemasson, E.; Bertin, S.; Hennig, P.; Lesellier, E. An improved classification of stationary phases for ultra-high performance supercritical fluid chromatography. *J. Chromatogr. A* **2016**, *1440*, 212–228.

(215) Khater, S.; Anand, S.; Mohanty, D. *In silico* methods for linking genes and secondary metabolites: The way forward. *Synth. Syst. Biotechnol.* **2016**, *1*, 80–88.

(216) Lesellier, E.; West, C. The many faces of packed column supercritical fluid chromatography--a critical review. *J. Chromatogr. A* **2015**, *1382*, 2–46.

Chapter 2

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

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

Affiliation

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 to the manuscript

Contributions and Acknowledgements

Author's effort:

The author significantly contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. The author cultivated the strains in liquid culture and on agar plates and acquired LC-MS, DI-MS, LC-MS/MS and DI-MS/MS data for all myxobacterial strains examined in this study. The author performed targeted analysis, as well as non-targeted statistical analysis of the acquired data and generated GNPS molecular networks for all strains. Furthermore, the author contributed significantly to conceiving and writing this manuscript.

Contributions by others:

Patrick A. Haack contributed to the conception of this study, to the design and performance of experiments, 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 conception and supervision of this study, as well as editing and proofreading of the manuscript.

2.1 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.
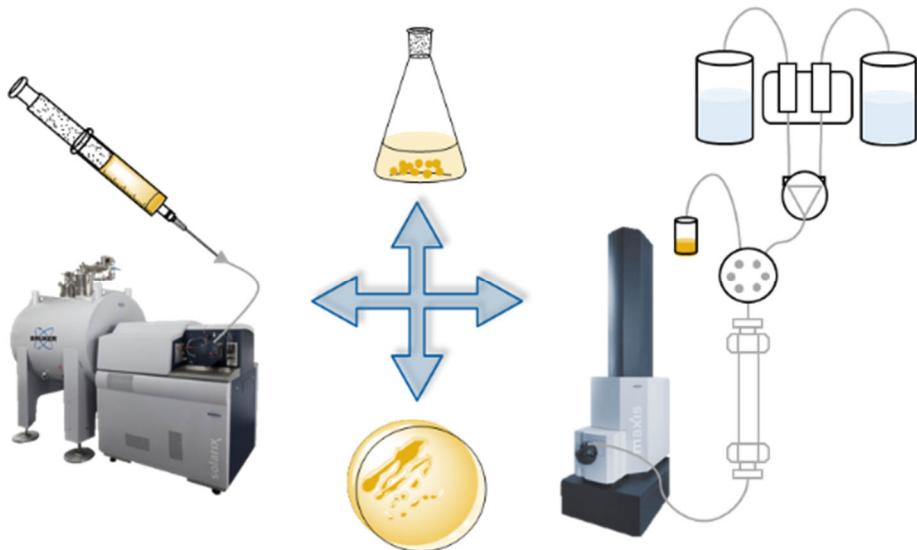
Figure 1 Table of contents graphic showing the principle design of the study

2.2 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. MS[2] 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.

2.3 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 2). 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.
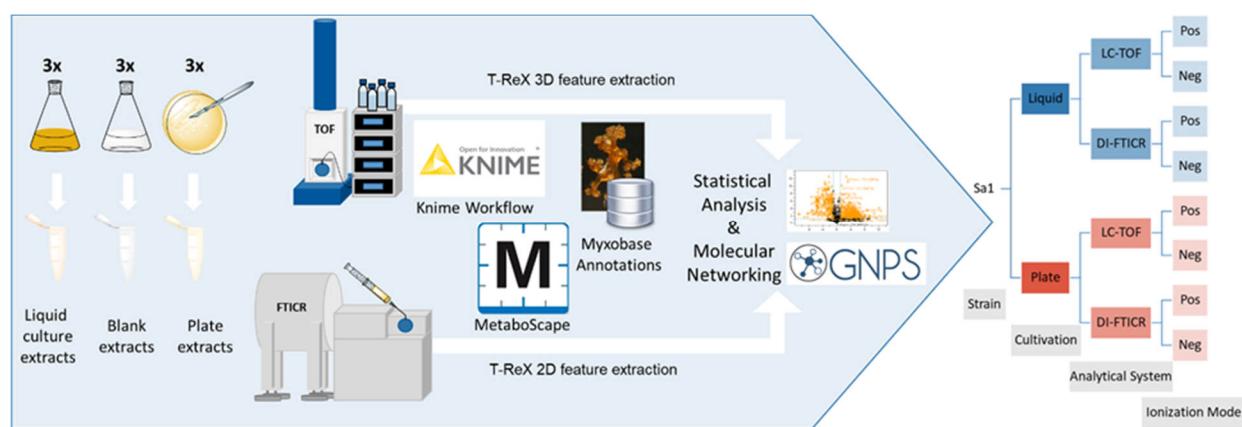
Figure 2 General workflow (left side) and measurements conducted per strain (right side).

2.3.1 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.

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
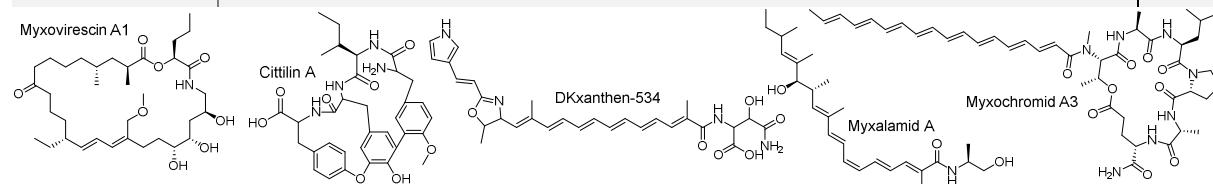
Generally, most annotations can be found for the combination of liquid cultivation and LC-TOF analysis, showing that the standard cultivation and analytical setup is suitable for a broad range of metabolites. As our in-house database strongly relies on input of natural products isolated after cultivation 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.

2.3.2 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 3). The absolute amount of molecular features detectable in negative mode is generally lower than in positive mode (Figure 3 and 4). 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).
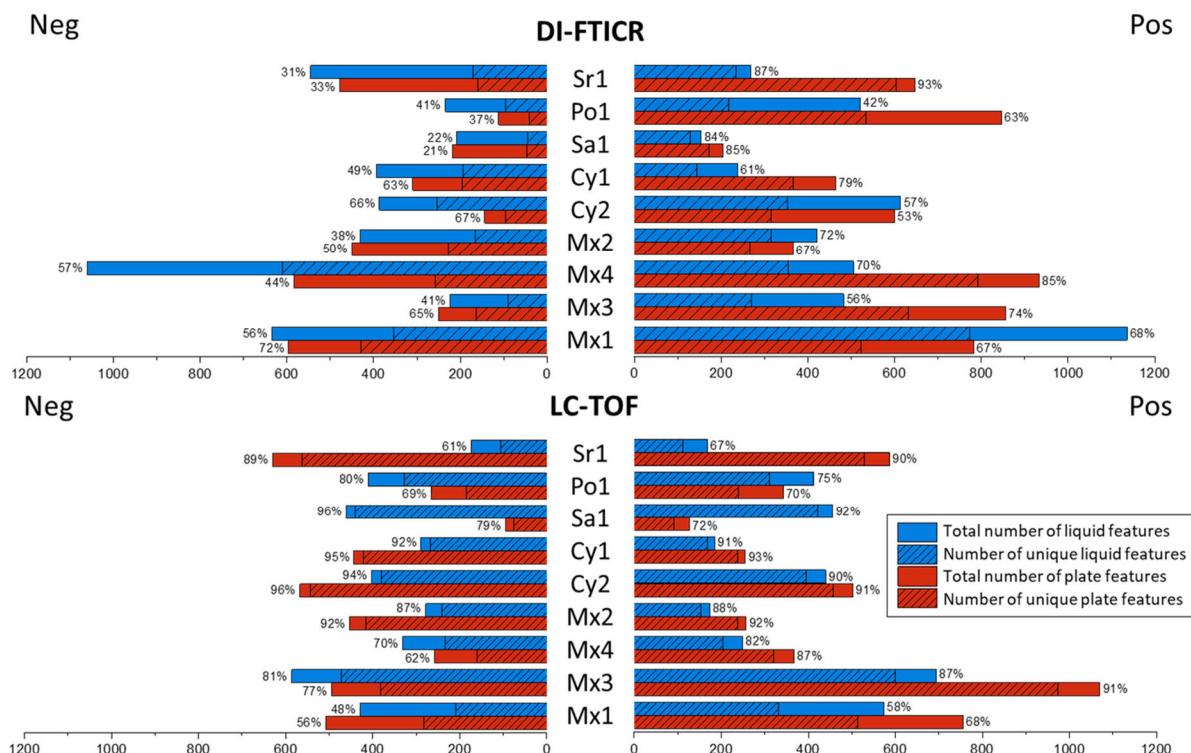
Figure 3 Features detected for the nine different myxobacterial strains comparing plate and liquid cultivation. Blue: Liquid features, Red: Plate features. The percentage of unique features to one of the cultivation systems is marked as dashes and shown at the respective end of the bar. Bars to the left represent molecular features detected in negative mode and bars to the right features detected in positive mode.

The absolute number of features detectable is highly strain dependent, also accounting for the percentage of features uniquely found in one cultivation system (Figure 3). 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 4). 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.
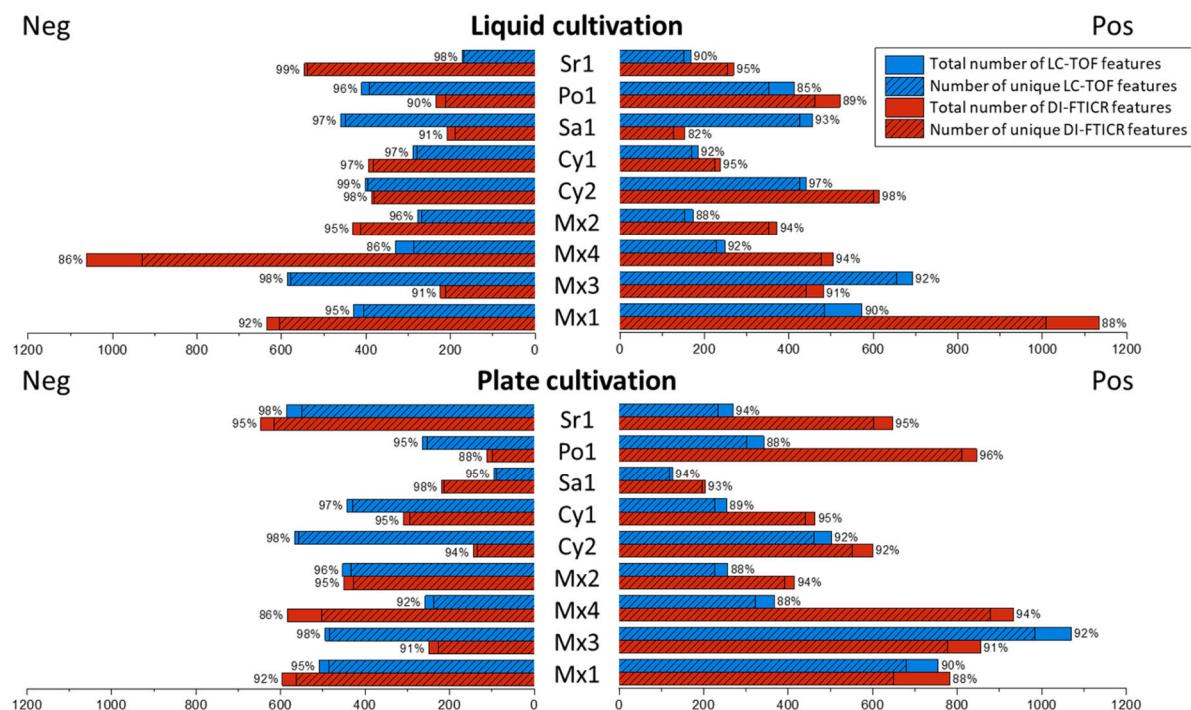
Figure 4 Features detected for the nine different myxobacterial strains comparing the two analytical systems. Blue: LC-TOF features, Red: DI-FTICR features. Percentage of unique features to one of the analytical systems shown at the respective end of a bar and marked as dashes. Bars to the left represent molecular features detected in negative mode and bars to the right features in positive mode.

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.

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.

2.3.3 MS²-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 5).
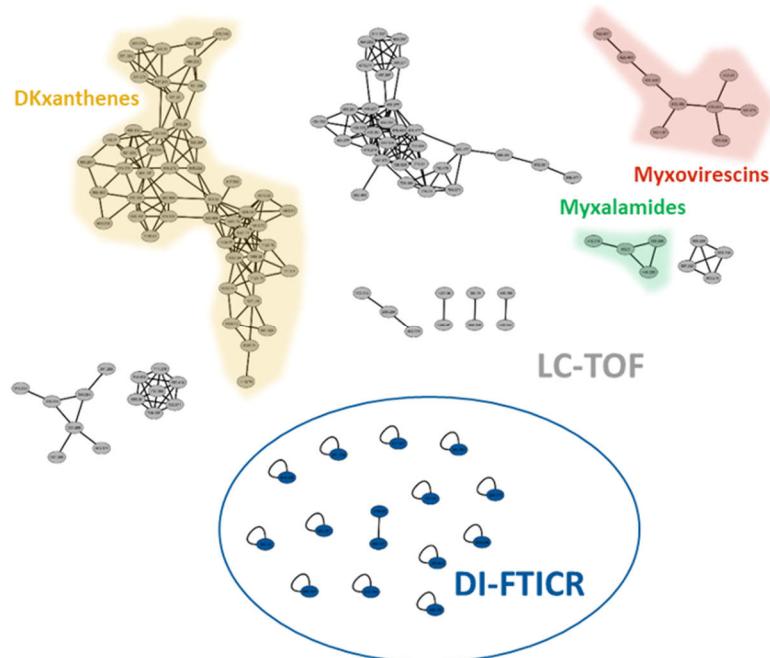
Figure 5 Molecular network of Mx1 features detected with LC TOF (grey nodes) and eight of the most intense features uniquely found with DI-FTICR (blue nodes). Known compound families produced by this strain marked in red (Myxovirescins), yellow (DKxanthenes) and green (Myxalamides).

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, MS² 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.

2.3.4 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 6). 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.
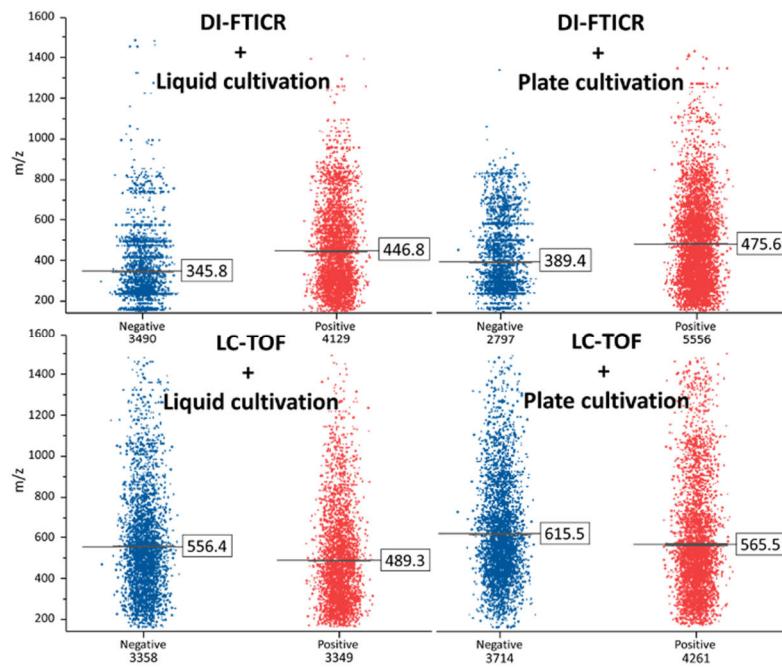
Figure 6 *m/z* distribution for negative (blue) and positive (red) ionization mode. Molecular features from all strains were grouped according to the analytical setup and cultivation condition. Mean *m/z* value +/- standard deviation indicated by the horizontal bars. Total number of features detected for each cultivation condition and analytical setup combination, across all strains for positive and negative ionization measurements indicated below the scatter plot

2.3.5 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 defence 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 7).[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.
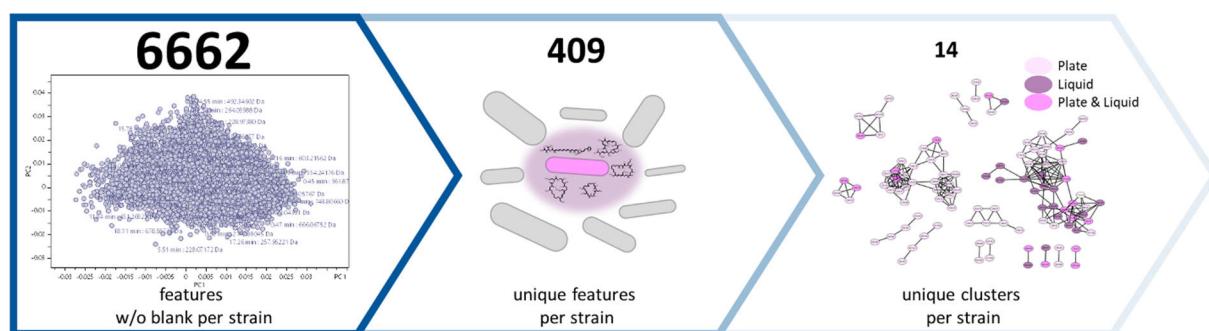
Figure 7 Prioritization workflow to cope with the large number of molecules deriving from primary metabolism. Molecular features depicted are the average number of molecular features detected per strain. Molecular network shown on right side derives from strain Sr1. Nodes depicted in light purple derive 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.

2.4 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.

2.5 Experimental Section

2.5.1 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.

2.5.1 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.

2.5.2 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) ddH$_2$O 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.

2.5.3 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).

2.5.4 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.

2.6 References

(1) U.S. Centers for Disease Control and Prevention. *Antibiotic Resistance Threats in the United States, 2019*.

(2) Bernardini, S.; Tiezzi, A.; Laghezza Masci, V.; Ovidi, E. Natural products for human health: an historical overview of the drug discovery approaches. *Nat. Prod. Res.* **2018**, *32*, 1926–1950.

(3) Baker, D. D.; Chu, M.; Oza, U.; Rajgarhia, V. The value of natural products to future pharmaceutical discovery. *Nat. Prod. Rep.* **2007**, *24*, 1225–1244.

(4) Bader, C. D.; Panter, F.; Müller, R. In depth natural product discovery - Myxobacterial strains that provided multiple secondary metabolites. *Biotechnol. Adv.* **2020**, *39*, 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. Antibiotics from myxobacteria. *Nat. Prod. Rep.* **2014**, *31*, 953–972.

(7) Herrmann, J.; Fayad, A. A.; Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **2017**, *34*, 135–160.

(8) Dehhaghi, M.; Mohammadipanah, F.; Guillemin, G. J. Myxobacterial natural products: An under-valued source of products for drug discovery for neurological disorders. *Neurotoxicology* **2018**, *66*, 195–203.

(9) Steinmetz, H.; Irschik, H.; Kunze, B.; Reichenbach, H.; Höfle, G.; Jansen, R. Thuggacins, macrolide antibiotics active against Mycobacterium tuberculosis: Isolation from myxobacteria, structure elucidation, conformation analysis and biosynthesis. *Chem. Eur. J.* **2007**, *13*, 5822–5832.

(10) Steinmetz, H.; Mohr, K. I.; Zander, W.; Jansen, R.; Gerth, K.; Müller, R. Indiacens A and B: prenyl indoles from the *myxobacterium Sandaracinus amylolyticus*. *J. Nat. Prod.* **2012**, *75*, 1803–1805.

(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. Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 175–184.

(14) Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **2019**, W81-W87.

(15) Fraser, C. M.; Eisen, J. A.; Salzberg, S. L. Microbial genome sequencing. *Nature* **2000**, *406*, 799–803.

(16) Fraser, C. M.; Eisen, J. A.; Nelson, K. E.; Paulsen, I. T.; Salzberg, S. L. The Value of Complete Microbial Genome Sequencing (You Get What You Pay For). *J. Bacteriol.* **2002**, *184*, 6403–6405.

(17) Wenzel, S. C.; Müller, R. The biosynthetic potential of myxobacteria and their impact on drug discovery. *Curr. Opin. Drug Discov. Devel.* **2009**, *12*, 220–230.

(18) Jensen, P. R.; Chavarria, K. L.; Fenical, W.; Moore, B. S.; Ziemert, N. Challenges and triumphs to genomics-based natural product discovery (Review). *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 203–209.

(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.* Co-culture of Marine Invertebrate-Associated Bacteria and Interdisciplinary Technologies Enable Biosynthesis and Discovery of a New Antibiotic, Keyicin. *ACS Chem. Biol.* **2017**, *12*, 3093–3102.

(20) Piddock, L. J. V. Teixobactin, the first of a new class of antibiotics discovered by iChip technology? *J Antimicrob Chemother* **2015**, *70*, 2679–2680.

(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. Co-cultivation of *Sorangium cellulosum* strains affects cellular growth and biosynthesis of secondary metabolite epothilones. *FEMS Microbiol. Ecol.* **2013**, *85*, 358–368.

(22) Panter, F.; Krug, D.; Baumann, S.; Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **2018**, *9*, 4898–4908.

(23) Saito-Shida, S.; Hamasaka, T.; Nemoto, S.; Akiyama, H. Multiresidue determination of pesticides in tea by liquid chromatography-high-resolution mass spectrometry: Comparison between Orbitrap and time-of-flight mass analyzers. *Food chemistry* **2018**, *256*, 140–148.

(24) Fels, H.; Dame, T.; Sachs, H.; Musshoff, F. Liquid chromatography-quadrupole-time-of-flight mass spectrometry screening procedure for urine samples in forensic casework compared to gas chromatography-mass spectrometry. *Drug Test. Anal.* **2017**, *9*, 824–830.

(25) Boesl, U. Time-of-flight mass spectrometry: Introduction to the basics. *Mass Spectrom. Rev.* **2017**, *36*, 86–109.

(26) Kim, J.; Choi, J. N.; Kim, P.; Sok, D. E.; Nam, S. W.; Lee, C. H. LC-MS/MS Profiling-Based Secondary Metabolite Screening of *Myxococcus xanthus*. *J. Microbiol. Biotechnol.* **2009**, *19*, 51–54.

(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.* Molecular networking as a dereplication strategy. *J. Nat. Prod.* **2013**, *76*, 1686–1699.

(28) Hoffmann, T.; Krug, D.; Hüttel, S.; Müller, R. Improving natural products identification through targeted LC-MS/MS in an untargeted secondary metabolomics workflow. *Anal. Chem.* **2014**, *86*, 10780–10788.

(29) Sleighter, R. L.; Hatcher, P. G. The application of electrospray ionization coupled to ultrahigh resolution mass spectrometry for the molecular characterization of natural organic matter. *J. Mass Spectrom.* **2007**, *42*, 559–574.

(30) Brown, S. C.; Kruppa, G.; Dasseux, J.-L. Metabolomics applications of FT-ICR mass spectrometry. *Mass Spectrom. Rev.* **2005**, *24*, 223–231.

(31) Sommella, E.; Conte, G. M.; Salviati, E.; Pepe, G.; Bertamino, A.; Ostacolo, C.; Sansone, F.; Prete, F. D.; Aquino, R. P.; Campiglia, P. Fast Profiling of Natural Pigments in Different Spirulina (Arthrospira platensis) Dietary Supplements by DI-FT-ICR and Evaluation of their Antioxidant Potential by Pre-Column DPPH-UHPLC Assay. *Molecules (Basel, Switzerland)* **2018**, *23*.

(32) Wu, Z.; Rodgers, R. P.; Marshall, A. G. Characterization of vegetable oils: detailed compositional fingerprints derived from electrospray ionization fourier transform ion cyclotron resonance mass spectrometry. *J. Agric. Food Chem.* **2004**, *52*, 5322–5328.

(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. 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.

(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. Petroleum crude oil characterization by IMS-MS and FTICR MS. *Anal. Chem.* **2009**, *81*, 9941–9947.

(36) Annesley, T. M. Ion Suppression in Mass Spectrometry. *Clin. Chem.* **2003**, *49*, 1041–1044.

(37) Southam, A. D.; Payne, T. G.; Cooper, H. J.; Arvanitis, T. N.; Viant, M. R. Dynamic range and mass accuracy of wide-scan direct infusion nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry-based metabolomics increased by the spectral stitching method. *Anal. Chem.* **2007**, *79*, 4595–4602.

(38) Furey, A.; Moriarty, M.; Bane, V.; Kinsella, B.; Lehane, M. Ion suppression; a critical review on causes, evaluation, prevention and applications. *Talanta* **2013**, *115*, 104–122.

(39) Muñoz-Dorado, J.; Marcos-Torres, F. J.; García-Bravo, E.; Moraleda-Muñoz, A.; Pérez, J. Myxobacteria: moving, killing, feeding, and surviving together. *Front. Microbiol.* **2016**, *7*, 781.

(40) Hoffmann, M.; Auerbach, D.; Panter, F.; Hoffmann, T.; Dorrestein, P. C.; Müller, R. Homospermidine Lipids: A compound class specifically formed during fruiting body formation of *Myxococcus xanthus* DK1622. *ACS Chem. Biol.* **2018**, *13*, 273–280.

(41) Gerth, K.; Irschik, H.; Reichenbach, H.; Trowitzsch, W. The myxovirescins, a family of antibiotics from Myxococcus virescens (Myxobacterales). *J. Antibiot.* **1982**, *35*, 1454–1459.

(42) Hug, J. J.; Dastbaz, J.; Adam, S.; Revermann, O.; Koehnke, J.; Krug, D.; Müller, R. Biosynthesis of Cittilins, Unusual Ribosomally Synthesized and Post-translationally Modified Peptides from Myxococcus xanthus. *ACS Chem. Biol.* **2020**, *15*, 2221–2231.

(43) Meiser, P.; Bode, H. B.; Müller, R. The unique DKxanthene secondary metabolite family from the myxobacterium *Myxococcus xanthus* is required for developmental sporulation. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19128–19133.

(44) Gerth, K.; Jansen, R.; Reifenstahl, G.; Höfle, G.; Irschik, H.; Kunze, B.; Reichenbach, H.; Thierbach, G. The myxalamids, new antibiotics from *Myxococcus xanthus* (Myxobacterales) I. production, physico-chemical and biological properties, and mechanism of action. *J Antibiot* **1983**, *36*, 1150–1156.

(45) Trowitzsch-Kienast, W.; Gerth, K.; Wray, V.; Reichenbach, H.; Höfle, G. Antibiotika aus Gleitenden Bakterien, LV – Myxochromid A: Ein hochungesättigtes Lipopeptidlacton aus *Myxococcus virescens*. *Liebigs Ann. Chem.* **1993**, *1993*, 1233–1237.

(46) Krug, D.; Garcia, R.; Müller, R. Myxobakterielle Naturstofffabriken. *Biospektrum* **2020**, *26*, 32–36.

(47) Krug, D.; Zurek, G.; Revermann, O.; Vos, M.; Velicer, G. J.; Müller, R. Discovering the Hidden Secondary Metabolome of *Myxococcus xanthus*: a Study of Intraspecific Diversity. *Appl. Environ. Microbiol.* **2008**, *74*, 3058–3068.

(48) Nordström, A.; Want, E.; Northen, T.; Lehtiö, J.; Siuzdak, G. Multiple ionization mass spectrometry strategy used to reveal the complexity of metabolomics. *Anal. Chem.* **2008**, *80*, 421–429.

(49) Oss, M.; Kruve, A.; Herodes, K.; Leito, I. Electrospray ionization efficiency scale of organic compounds. *Anal. Chem.* **2010**, *82*, 2865–2872.

(50) Huffman, B. A.; Poltash, M. L.; Hughey, C. A. Effect of polar protic and polar aprotic solvents on negative-ion electrospray ionization and chromatographic separation of small acidic molecules. *Anal. Chem.* **2012**, *84*, 9942–9950.

(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.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837.

(52) Hoffmann, T.; Krug, D.; Bozkurt, N.; Duddela, S.; Jansen, R.; Garcia, R.; Gerth, K.; Steinmetz, H.; Müller, R. Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat. Commun.* **2018**, *9*, 803.

(53) Irschik, H.; Reichenbach, H.; Höfle, G.; Jansen, R. The thuggacins, novel antibacterial macrolides from Sorangium cellulosum acting against selected Gram-positive bacteria. *J Antibiot.* **2007**, *60*, 733–738.

(54) 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.

2.7 Supporting information

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

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

Affiliation

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 to the manuscript

S2.1 Extract generation, LC-MS data acquisition and GNPS-based secondary metabolite clustering

S2.1.1 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 S1 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*

S2.1.2 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 S2 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)	

S2.1.3 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.

S2.1.4 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.

S2.2 Conditions of the analytical measurements

S2.2.1 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
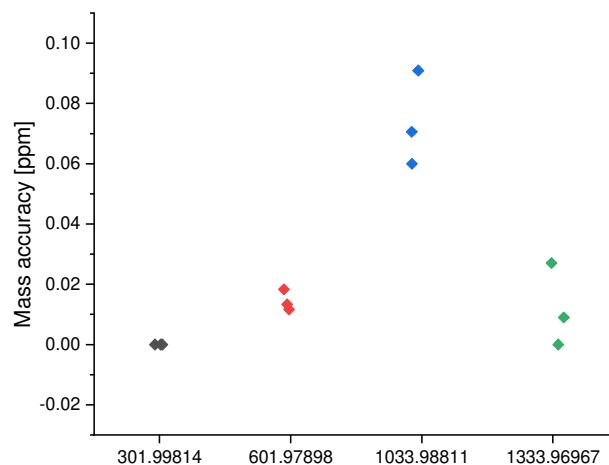
Figure S1. Mass accuracy for the four components of the tune mix in the m/z range from 200-1500 on the three days of measurements in negative ionization mode.

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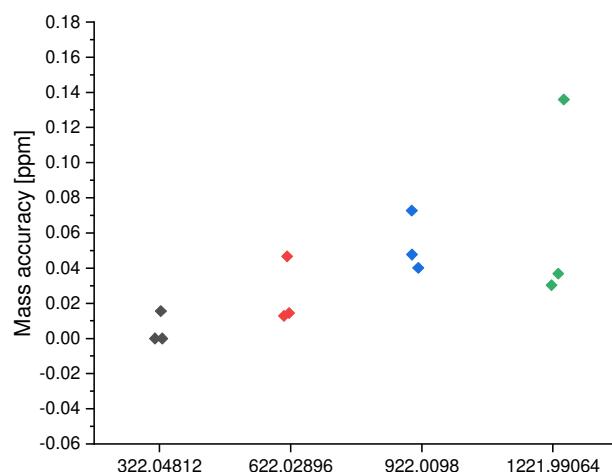
Figure S2 Mass accuracy for the four components of the tune mix in the m/z range from 200-1500 on the three days of measurements in positive ionization mode.

S2.2.2 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
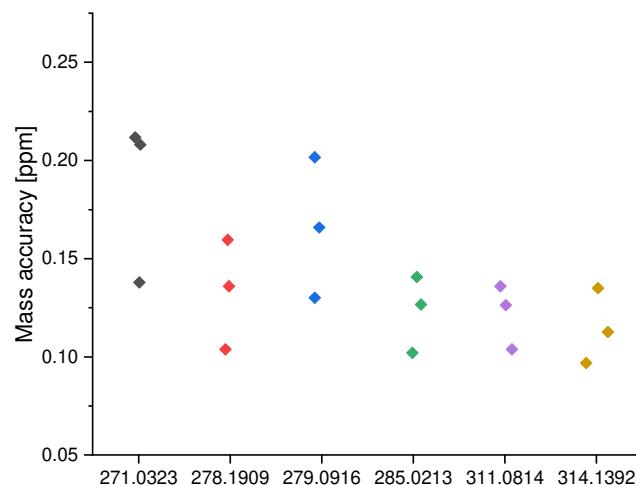
Figure S3 Mass accuracy for the six components of the in-house in positive ionization mode.

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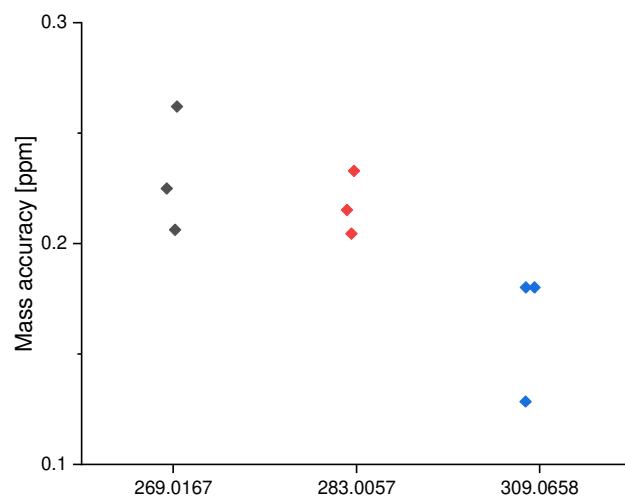
Figure S4 Mass accuracy for the six components of the in-house in positive ionization mode.

S2.3 Methodology for statistical metabolome filtering

S2.3.1 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.
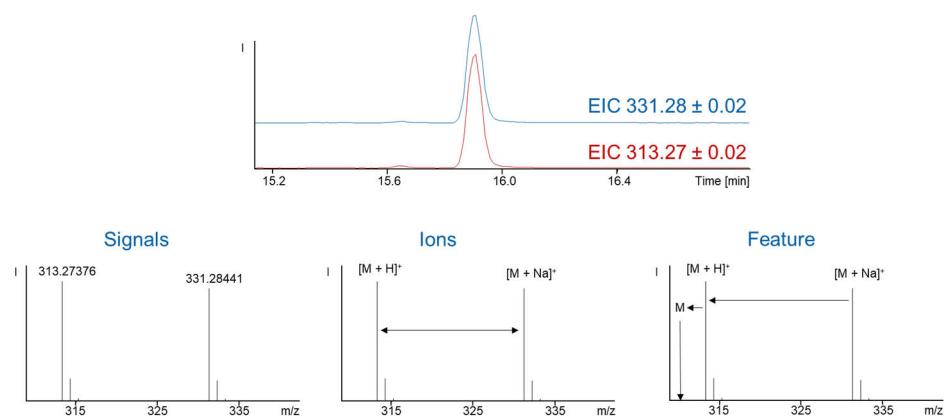
Figure S5 Difference between signals, ions and features exemplified for m/z 331.28 and 313.27.

S2.3.2 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.

S2.4 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.*
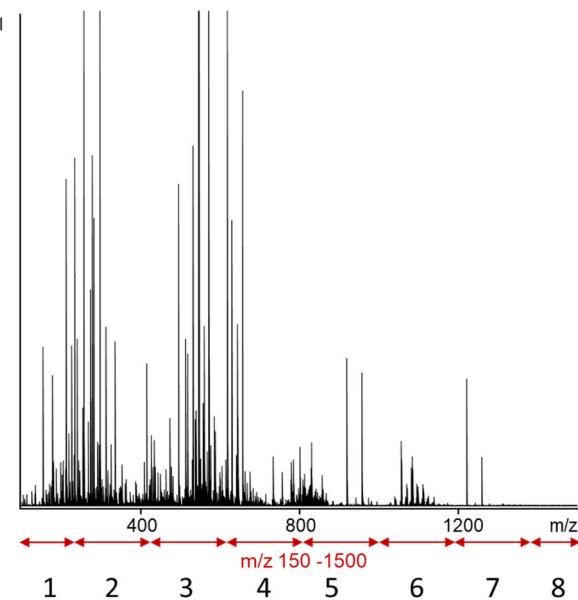
Figure S6 Segments used for the DI-FTICR measurements.

S2.5 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.

S2.6 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 7 Known myxobacterial secondary metabolites, detected in the extracts used for this study.

Mx122	LC-TOF+Liquid		LC-TOF+Plate		DI-FTICR+Liquid		DI-FTICR+Plate		ΔRT [min]
	Δppm	ion type	Δppm	ion type	Δppm	ion type	Δppm	ion type	
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

SoCe26	LC-TOF+Liquid		LC-TOF+Plate		DI-FTICR+Liquid		DI-FTICR+Plate		ΔRT [min]
	Δppm	ion type	Δppm	ion type	Δppm	ion type	Δppm	ion type	
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

SbSr044	LC-TOF+Liquid		LC-TOF+Plate		DI-FTICR+Liquid		DI-FTICR+Plate		ΔRT [min]
	Δppm	ion type	Δppm	ion type	Δppm	ion type	Δppm	ion type	
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

	LC-TOF+Liquid		LC-TOF+Plate		DI-FTICR+Liquid		DI-FTICR+Plate		
	Δppm	ion type	Δppm	ion type	Δppm	ion type	Δppm	ion type	ΔRT [min]
Corallopyronin A2			0,7	[M-H$_2$O+H]$^+$					0,04
Salinimyxantin B			0,8	[M+H]$^+$			1,8	[M+K]$^+$	0,07

Mx152	LC-TOF+Liquid		LC-TOF+Plate		DI-FTICR+Liquid		DI-FTICR+Plate		
	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$_2$O+H]$^+$	N/A
Soraphen C2α							2,1	[M-H$_2$O+H]$^+$	N/A

S2.7.2 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.
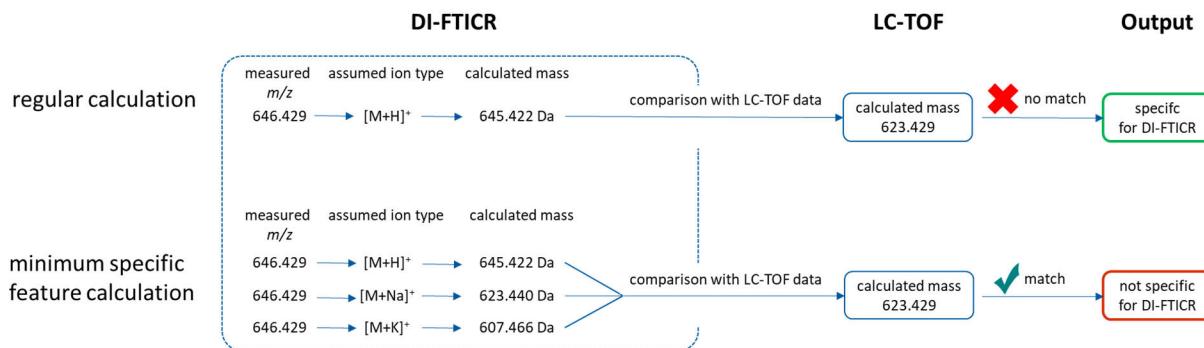
Scheme S1. Minimum number of unique feature calculation for the comparison of LC-TOF and DI-FTICR in positive ionization mode.

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	$[M+H]^+$ com	$[M+H]^+$ spec	$[M+Na]^+$ com	$[M+Na]^+$ spec	$[M+K]^+$ com	$[M+K]^+$ spec	total sum commons	total specifics absolute	total 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%

S2.8 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.
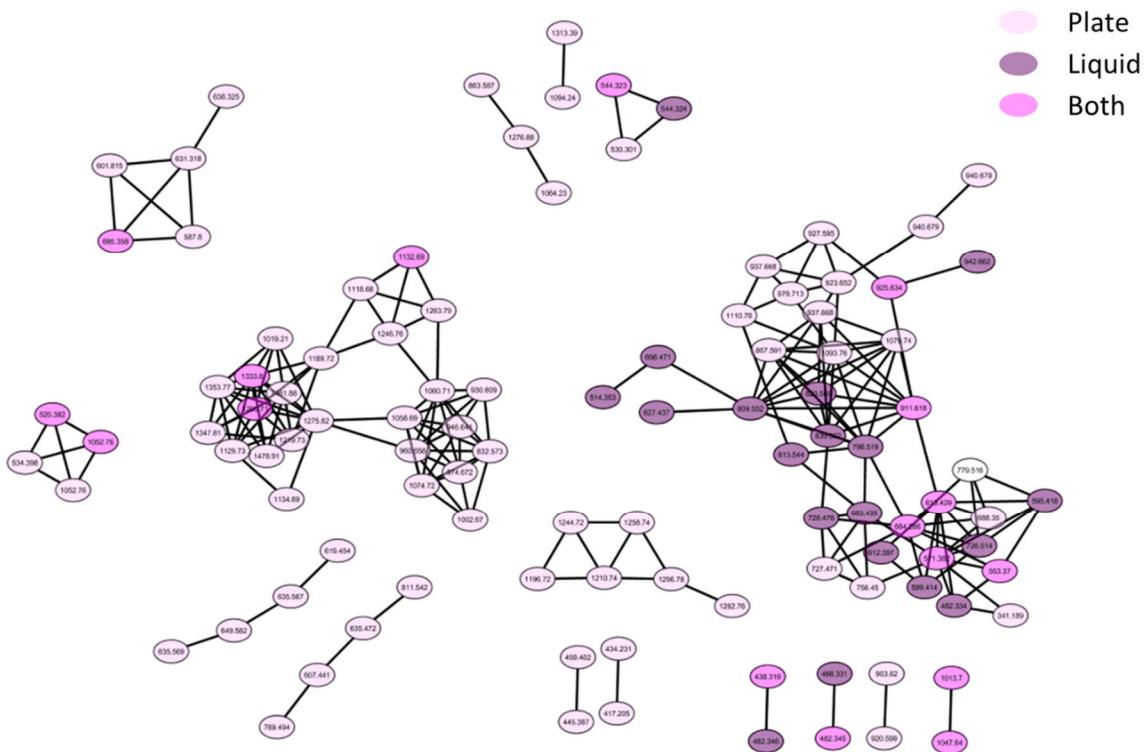
Figure S7 Molecular network of unique metabolites produced by strain MSr11367.
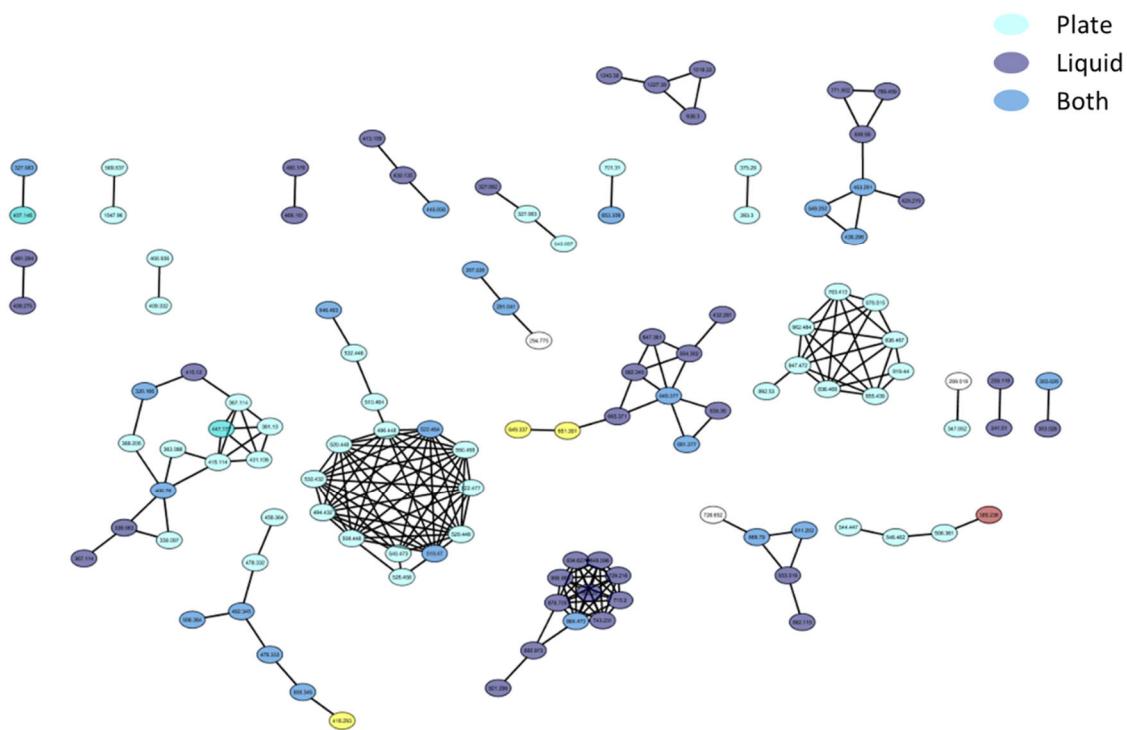
Figure S8 Molecular network of unique metabolites produced by *Myxococcus fulvus* MCy9270.
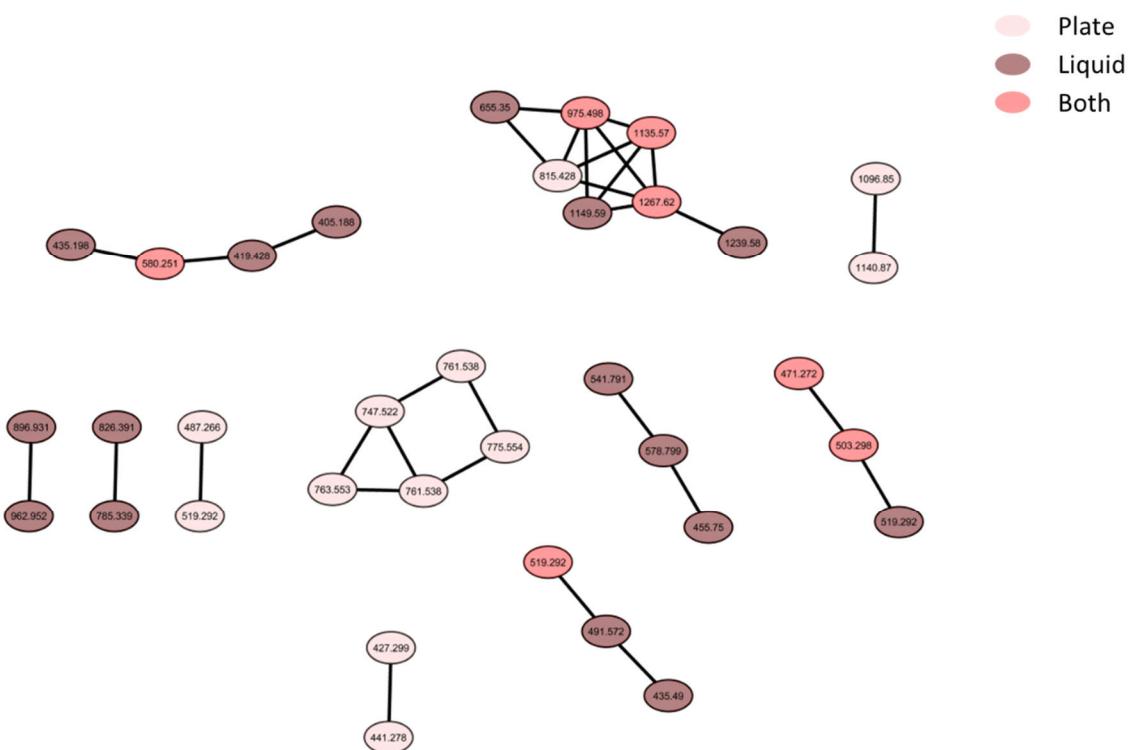
Figure S9 Molecular network of unique metabolites produced by *Sorangium cellulosum* SoCe 26.
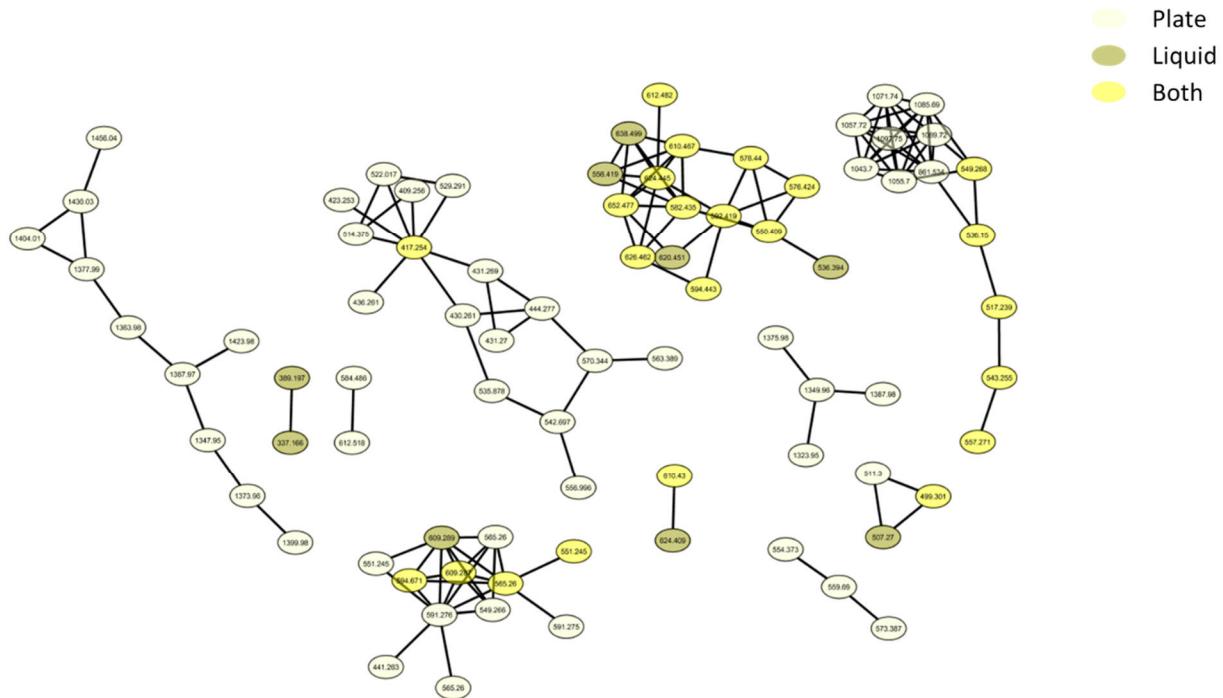
Figure S10 Molecular network of unique metabolites produced by *Myxococcus xanthus* DK1622.
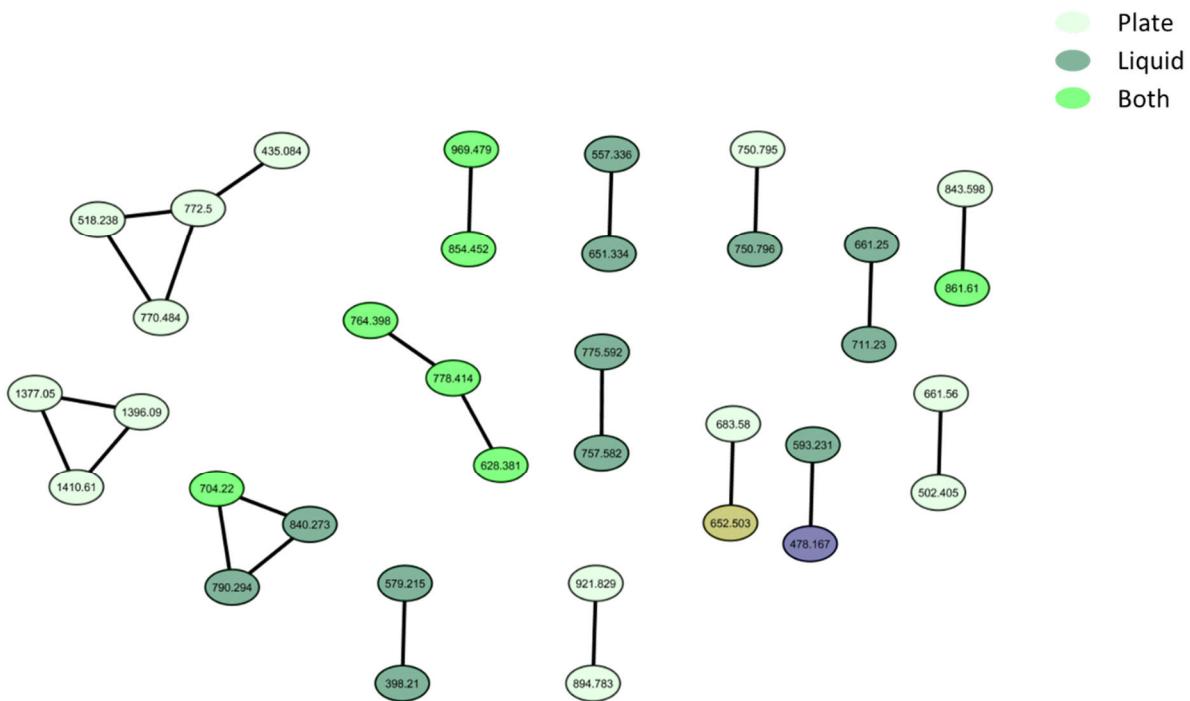
Figure S11 Molecular network of unique metabolites produced by *Pyxidicoccus* sp. MCy10649.
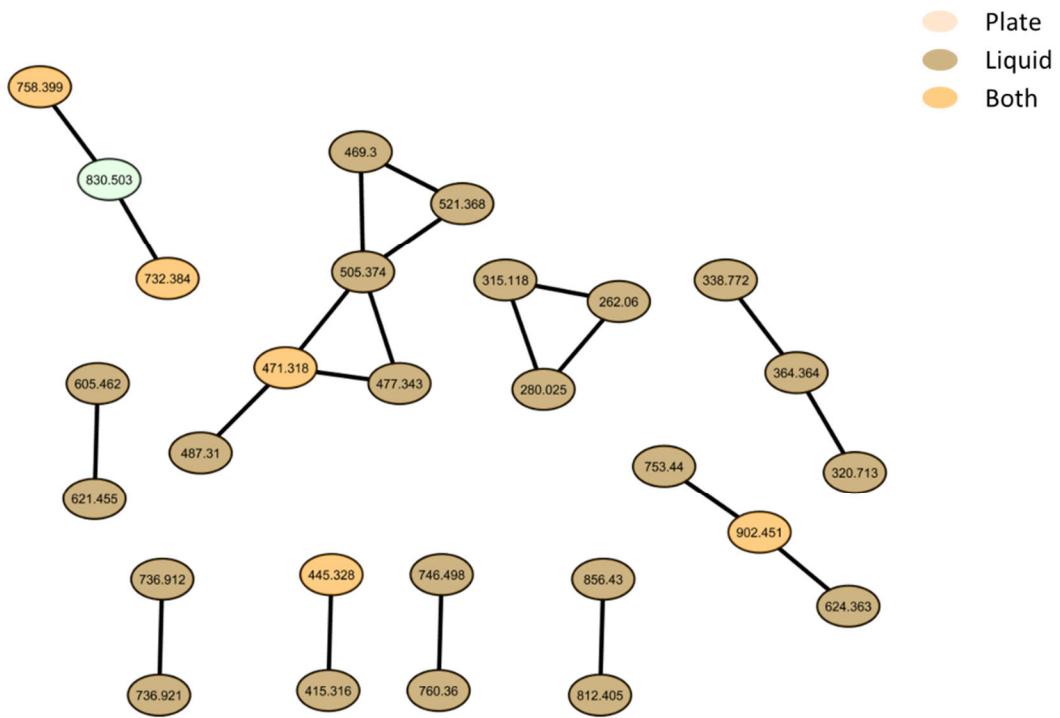
Figure S12 Molecular network of unique metabolites produced by MSr10575.
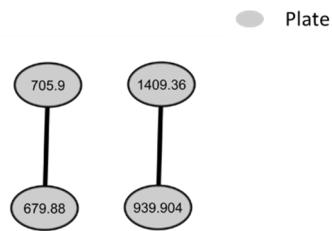
Figure S13 Molecular network of unique metabolites produced by MSr11770.
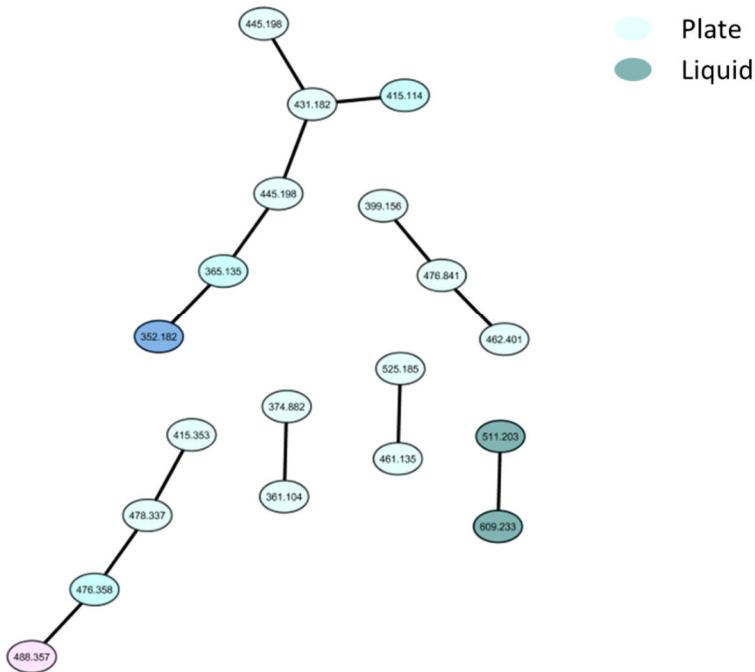
Figure S14 Molecular network of unique metabolites produced by MCy9487.

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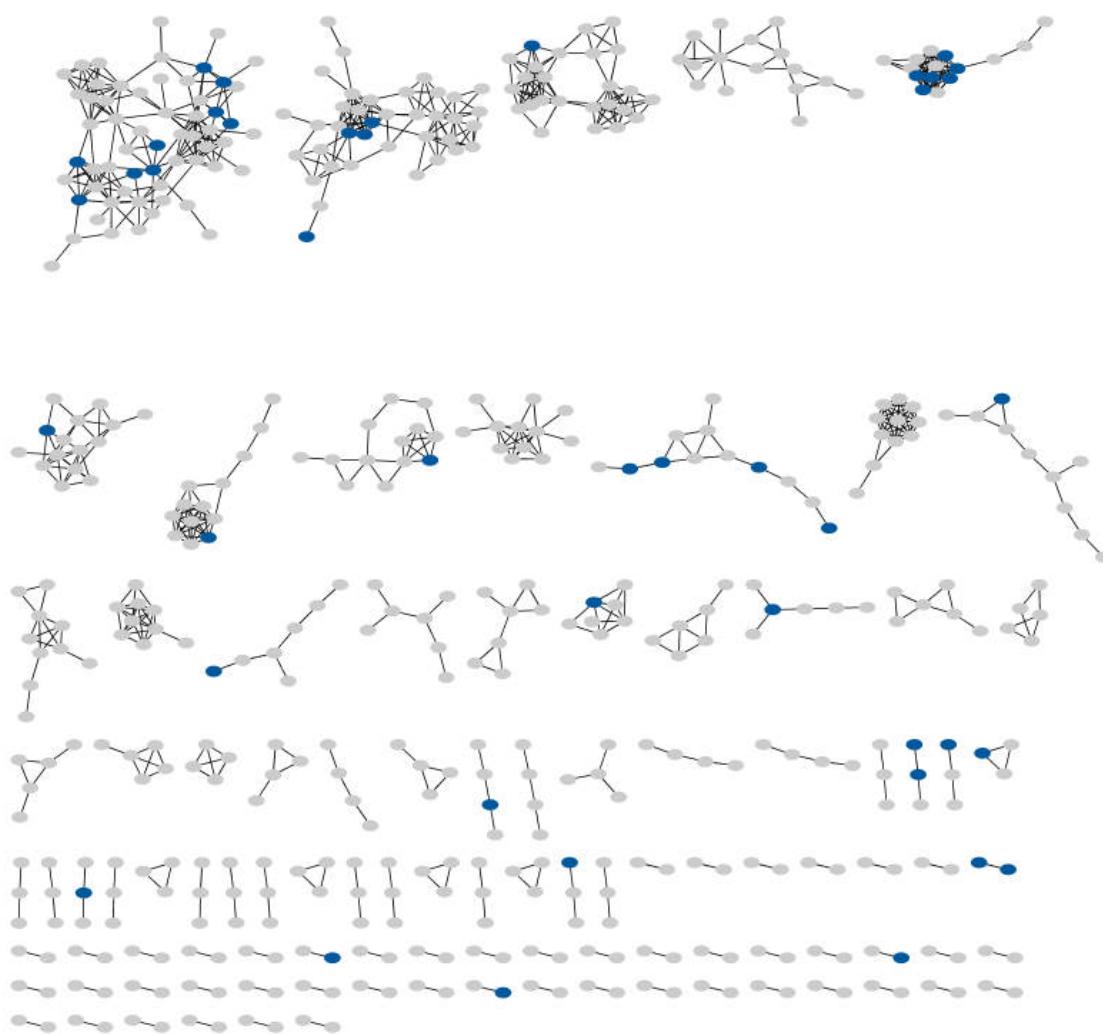
Figure S15 Molecular network of all strain unique metabolite with feautures only detectable with LC-TOF highlighted in blue and features detectable with LC-TOF and DI-FTICR in grey.

S2.9 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.

Chapter 3

Supercritical Fluid Extraction Enhances Discovery of Secondary Metabolites from Myxobacteria

Chantal D. Bader, Markus Neuber, Fabian Panter, Daniel Krug and Rolf Müller

Affiliation

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)

Contributions and Acknowledgements

Author's effort:

The author significantly contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. The author performed supercrititcal fluid and conventional extraction, acquired LC-MS for all myxobacterial strains examined in this study and performed targeted, as well as non-targeted statistical analysis of the acquired data. The author isolated and structurally elucidated chloroxanthic acid A. Furthermore, the author contributed significantly to conceiving and writing this manuscript.

Contributions by others:

Markus Neuber contributed by cultivation of the strains and proofreading of the manuscript. Fabian Panter contributed by isolation of indiacene A and B and their structure elucidation, as well as editing of the manuscript. Stefanie Schmidt and Viktoria Schmitt performed the antimicrobial assays. Daniel Krug and Rolf Müller contributed by conception and supervision of this study, as well as editing and proofreading of the manuscript.

3.1 Abstract

Supercritical fluid extraction (SFE) is widely used for the isolation of natural products from plants, but its application in efforts to identify structurally and physicochemically often dissimilar microbial natural products is limited to date. In this study we evaluated the impact of SFE on the extractability of myxobacterial secondary metabolites aiming to improve the prospects of discovering novel natural products. We investigated the influence of different co-solvents on the extraction efficiency of secondary metabolites from three myxobacterial strains as well as the antimicrobial activity profiles of the corresponding extracts. For each known secondary metabolite we found extraction conditions using SFE leading to superior yields in the extracts compared to conventional solvent extraction. Compounds with a logP higher than 3 showed best extraction efficiency using 20% EtOAc as a co-solvent, whereas compounds with logP values lower than 3 were better extractable using more polar co-solvents like MeOH. Extracts generated with SFE showed increased antimicrobial activities including the presence of activities not explained by known myxobacterial secondary metabolites, highlighting the advantage of SFE for bioactivity-guided isolation. Moreover, non-targeted metabolomics analysis revealed a group of chlorinated metabolites produced by the well-studied model myxobacterium *Myxococcus xanthus* DK1622 which were not accessible previously due to their low concentration in conventional extracts. The enriched SF extracts were used for isolation and subsequent structure elucidation of chloroxanthic acid A as founding member of a novel secondary metabolite family. Our findings encourage the increased utilization of SFE as part of future microbial natural products screening workflows.
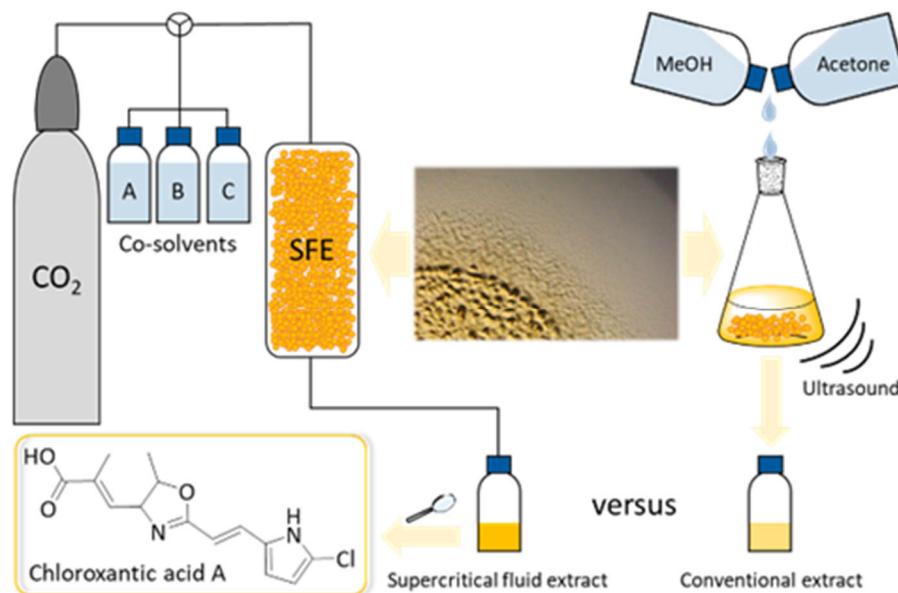
Figure 1 Table of contents graphic showing the principle design of the study

3.2 Introduction

Over the years, supercritical fluid extraction (SFE) has contributed to efficient extraction processes of plant based natural products for pharmaceutical and food applications.[1,2] SFE is often characterized as a "green technology" as its benefits include comparably low energy and liquid solvent consumption, relying mainly on supercritical carbon dioxide (CO_2) as extraction solvent.[3] Furthermore it profits from the unique intrinsic properties of a supercritical fluid, showing liquid–like solvation because of its liquid-like density, combined with increased mass transfer between the extracted material and the supercritical fluid because of its gas–like viscosity and diffusivity.[4] These unique properties allowed extraction and isolation of a diverse set of natural products from plants, such as flavonoids and phenolic acids from *Medicago sativa L.* (Alfalfa)[5] or piperine from *Piper nigrum L.*.[6] In microbial natural product research however, the application of SFE appears limited to the isolation of natural products from cyanobacteria. The structural diversity among cyanobacterial metabolites isolated with SFE is high, comprising the epoxide containing polycyclic natural product chaetoglobosin, the xanthone sydowinin B, the nitrophenyl containing mycolutein and luteoreticulin, the glycosylated macrolactam elaiophylin and the polyunsaturated fatty acid γ-linoleic acid.[7,8] All of these compounds extractable with dichloromethane in a conventional solvent extraction approach can be enriched with alcohol-modified supercritical CO_2 extraction.[8]

Some of the metabolites successfully isolated with SFE from plants or cyanobacteria are not only characterized by their unique chemical properties but exhibit potent biological activities such as the cytotoxic chaetoglobosins.[9,10] The comparison of chemical profiles and biological activities of SF extracts to conventional solvent extracts has also been carried out successfully for various plants like *Satureja montana* or *Sideritis scardica* Griseb..[11,12] Compared to soxhlet and hydrodistillation extracts of *Satureja montana*, SF extracts of this traditional medicine plant show stronger activities against *Bacillus cereus* and *B. subtilis* as well as unique activities against *Salmonella enteritidis* and *Colletotrichum coffeanum*.[13] These results highlight the potential impact of SFE on natural products research, as SFE was not only able to generate extracts with improved biological activities, but also enabled generation of extracts showing previously undetected biological activities.

In studies by Silva *et al.*, only supercritical CO_2 without the addition of co-solvent was used, which limits the supercritical fluid extraction exclusively to highly non-polar metabolites, as CO_2 itself is highly hydrophobic.[13] Esquivel-Hernández *et al.* have shown for the cyanobacterium *Arthrospira platensis* that from all variable extraction parameters like pressure, temperature, co-solvent and the introduction of static, dispersant and dynamic states during the extraction process, the choice of co-solvent has the highest impact on the extraction efficiency.[14] Pressure and temperature have to be chosen appropriately to exceed the supercritical point of CO_2 at 304.12 K and 73.7 bar.[14,15] Subsequently, they have to be kept high enough, to preserve the CO_2 in supercritical state. For choosing temperature and pressure, it has to be taken into account that increasing percentage of co-solvent increases pressure and temperature required to reach the supercritical point of the mixture.[16] The exact supercritical point of the respective mixture herein is strongly dependent on the chosen co-solvent. The most widely used co-solvents in SFE are alcohols such as ethanol (EtOH) and methanol (MeOH).[3]

For conventional solvent extraction, the choice of extraction solvent is also the most varied parameter. The extraction process can be enhanced by exposing the extraction mixture to microwaves, ultrasound or heat. Microwave-assisted extraction or heat supported extractions like soxhlet extraction, are not applicable for metabolites suffering from poor thermal stability, given for a broad range of natural products.[17] SFE in comparison offers the possibility to keep the temperature relatively low due to the

additional pressure applied, without loss in extraction efficiency.[16] On the other hand, conventional solvent extraction methods have one main advantage: they come without expensive instrumentation.

Most microbial natural product isolation protocols rely on maceration as conventional solvent extraction method.[18–20] To access a preferably diverse set of microbial natural products, Ito *et al.* exemplarily used a mixture of dichloromethane:methanol:water (64:36:8).[18] Novel secondary metabolites from myxobacteria isolated in the past years have almost exclusively been extracted with classical solvent extraction methods, using acetone as most commonly used solvent. Besides acetone, only MeOH and ethyl acetate (EtOAc) were used.[21–26]

Despite its many advantages over solvent extraction, SFE is underrepresented in microbial natural product research and has never been described for the extraction of myxobacterial cells. Most publications dealing with SFE furthermore only report its utilization for previously described natural products, like caffeine or capsaicin.[3] As a representative example for natural product extraction from microbial secondary metabolism we therefore set out to investigate the metabolome of three different myxobacterial strains using SFE in comparison to conventional ultrasound-assisted extraction (UAE) with acetone and MeOH. To study the influence of different co-solvents on the extraction of known myxobacterial metabolites, we performed targeted metabolomics-based examination of the SF extracts. Furthermore, all extracts were compared with regards to detection of antimicrobial activities. As a proof of concept, a novel secondary metabolite only detectable in trace amounts in UA extracts of the well-studied model organism *Myxococcus xanthus* DK1622 was isolated and structurally elucidated.

2.3 Experimental section

All cultivation conditions, analytical methods and statistical data treatment procedures are described in detail in the Supporting Information.

3.3 1 Cultivation and sample preparation

Three myxobacterial strains (*M. xanthus* DK1622, *Sorangium* sp. MSr11367 and *Sandaracinus* sp. MSr10575) were grown in 2 liter shaking flasks containing 400 mL of cultivation medium for 10 days at 30 °C and 180 rpm. An autoclaved aqueous suspension of Amberlite XAD-16 resin was added to all production media to a volumetric concentration of 2% [v/v]. The cell pellets and resin were lyophilized prior to further processing. *M. xanthus* DK1622 yielded 8.7 g, *Sorangium* sp. MSr11367 8.7 g and *Sandaracinus* sp. MSr10575 2.7 g of dry cell pellet and resin. The combined dry pellet containing cells and resin was then pulverized using an electric coffee grinder and split into two equal batches. One batch was extracted by UAE whilst the other batch was extracted with SFE. Blank extracts were generated by extracting the medium with resin.

3.3.2 Extraction procedure

. SFE was carried out on a Waters® MV-10 ASFE® system equipped with 10 mL analytical extraction cartridges. The pulverized pellets were densely packed into the cartridges and topped up with glass beads to avoid pressure drops due to inconsistencies in the packing. The extraction process was carried out at 200 bar and 40 °C using a total flow rate of 10 mL/min. The extraction was done in five steps, each one performed with an initial dynamic extraction step for 5 minutes followed by a static extraction step for 5 minutes and a second dynamic extraction step for another 5 minutes. The pellets were first extracted with 100% CO_2. Afterwards, the pellet was extracted with 20% EtOAc as co-solvent, followed by 20% isopropanol (^{i}PrOH), 20% MeOH and as a final step the pellet was flushed with 50% MeOH. The MeOH make-up flow was set to a total organic solvent flow of 3 mL/min. The resulting solvent consumption was therefore 30 mL per step, except the extraction step with 50% MeOH where 50 mL

MeOH was used. All steps were collected and treated separately. The solvents were evaporated under reduced pressure and the residues were dissolved in 4 mL MeOH each. UAE consisted of two subsequent extraction steps, which were adapted from the most common extraction procedures conducted for myxobacterial natural products from 2018-2020 (see chapter 3.1 SI). [21–24,27–34] Firstly, 100 mL MeOH was added and the pellet sonicated for 1 h. After removal of the MeOH, 100 mL acetone was added, and the pellet was sonicated for another 1 h. The two extracts were combined, and the solvent was removed under reduced pressure. The residues were dissolved in 4 mL MeOH each. All extracts were diluted 1:100 prior to UHPLC-MS measurements.

3.3.3 Isolation procedure

Chloroxanthic acid A was isolated by semipreparative HPLC following a fractionation of the crude extract by liquid-liquid partitioning. Purification was performed on a Dionex Ultimate 3000 SDLC low pressure gradient system equipped with a Phenomonex Luna® 5 µm C18(2) 100 Å LC Column (250 x 10 mm) thermostated at 45 °C. Separation was achieved using a linear gradient from 95% (A) ddH$_2$O with 0.1% formic acid to 70% (B) acetonitrile with 0.1% formic acid over 29 minutes. The compound was detected by UV absorption at 210 and 280 nm.

3.3.4 MS and NMR analysis of crude extracts and pure compound

UHPLC-qTOF measurements were performed in duplicates as technical replicates on a Dionex Ultimate 3000 SL system coupled to a Bruker maXis 4G UHRqTOF. The mobile phase consisted of (A) ddH$_2$O 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 C18 column (100 x 2.1 mm, 1.7 µm d$_p$) was used. Flow rate was set to 0.6 mL/min and the column heated to 45 °C. The LC flow was split to 75 µL/min before entering the mass spectrometer, which was externally calibrated to a mass accuracy of < 1 ppm. External calibration was achieved with sodium formate clusters. Mass spectra were acquired in centroid mode ranging from 150-2500 m/z at a 2 Hz scan rate. Capillary voltage was set to + 4000 V for measurements in positive ionization mode. Dry gas was set to a flow rate of 5 L/min at 200 °C.

NMR spectra were recorded on a Bruker Ascend 700 spectrometer equipped with a 5 mm TXI cryoprobe (^{1}H at 700 MHz, ^{13}C at 175 MHz). All observed chemical shift values (δ) are given in in ppm and coupling constant values (J) in Hz. Standard pulse programs were used for HMBC, HSQC and gCOSY experiments. HMBC experiments were optimized for $^{2,3}J_{C-H}$ = 6 Hz. The spectra were recorded in methanol-d_4 and chemical shifts of the solvent signals at δ_H 3.31 ppm and δ_C 49.2 ppm were used as reference signals for spectra calibration. To increase sensitivity, the measurements were conducted in a 5 mm Shigemi tube (Shigemi Inc., Allison Park, PA 15101, USA).

3.3.5 Statistical analysis and annotations

Statistical interpretation of the LC-MS data for targeted and non-targeted metabolomics analysis was carried out with MetaboScape 4.0 and 5.0 (Bruker). The minimal intensity threshold for feature detection by the built-in T-Rex 3D algorithm was set to 1.5 x 10^4 and the maximum charge was set to three. The minimal group size for creating batch features to five. Known myxobacterial secondary metabolites were annotated using our in-house database containing myxobacterial secondary metabolites (Myxobase). Overlap of the extraction methods as well as features uniquely found in one analysis were calculated via t-Test analysis in MetaboScape 4.0.

3.3.6 Determination of antimicrobial activities

For determination of the extracts' antimicrobial activities, all three biological replicates were combined and 20 µL of the saturated methanolic extract was pipetted to 150 µL microbial test culture and serially diluted in 96 well plates. The microbial test panel consisted of *Staphylococcus aureus* Newman, *Escherichia coli* DSM111, *E. coli* TolC, *Bacillus subtilits*, *Micrococcus luteus*, *Pseudomonas aeruginosa* PA14, *Candida albicans*, *Pichia anomala*, *Mucor hiemalis* and *Mycobacterium smegmatis* Mc2155.

3.4 Results and discussion

We generated a total number of six different extracts per strain. In case of the SF extraction, the pellet was extracted sequentially with pure CO_2, followed by the addition of 20% EtOAc, *i*PrOH and MeOH as co-solvents. The co-solvents were used in ascending polarity. As the last step 50% MeOH was used in addition to the 20% MeOH step. When SF extracts are referred to in the following, this is indicated by the percentage and type of cosolvent used. Conventional extraction is referred to as ultrasound assisted extraction (UAE).

3.4.1 Influence of the co-solvent on the extraction of myxobacterial secondary metabolites

All extracts were investigated in a targeted metabolomics workflow and known myxobacterial secondary metabolites were annotated from our in-house data base (Myxobase). As *Sandaracinus* MSr10575 and *Sorangium* strain MSr11367 are new myxobacterial strains that were recently isolated in our working group, a lower number of metabolites is known compared to the well-studied model myxobacterium *M. xanthus* DK1622.[35,36] In *M. xanthus* DK1622 extracts, five different natural product families, namely the cittilins, myxovirescins, DKxanthenes, myxalamides and myxochromids, are detected. Representatives of these natural product families are shown in Figure 2 (A-E). The remaining derivatives can be found in in the SI. Myxochelins and myxoprincomides, two additional natural product families known from this strain, are not produced under the chosen cultivation conditions. Besides those secondary metabolites, the strain is found to produce riboflavin. As there is a controversial debate if riboflavin should be taken as primary or secondary metabolite (sometimes classified as pseudo-secondary metabolite), the production and extraction is just mentioned for completeness but not discussed in more detail here.[37] Both described cittilin derivatives (A and B) could be detected in the extracts, with the highest intensity found in the 50% MeOH extracts.[38,39] Cittilins are also extractable with 20% MeOH as well as UAE and cittilin A can also be detected in lower intensities in the *i*PrOH extract. The family of myxovirescins contains more than 12 members, whereof six different derivatives (A-C and G-I) are detectable in our analysis.[40] The highest intensities of all myxovirescin derivatives are found in the EtOAc extracts. Contrary to the cittilins that are not extractable with pure CO_2, the myxovirescins are also extractable without the addition of any co-solvent in SFE. They are also present in low amounts in the 20% and 50% MeOH extracts. Due to the serial design of the experiment, this indicates that the myxovirescins could be extracted using MeOH as a co-solvent, but the compounds were already exhaustively extracted during the EtOAc step. Another abundant and derivative-rich family of secondary metabolites are the DKxanthenes (DKx).[41] In our analysis, we detect seven DKx derivatives, DKx 520, 526, 534, 544, 548, 560 and 574. Except for DKx 526 all derivatives show the highest intensities in the 50% MeOH extracts. DKx 526 on the other hand shows highest intensities in the EtOAc extract. All described myxalamide derivatives (A-D and K) are detectable in our extraction analysis, with highest intensities in the EtOAc extracts. The *i*PrOH extracts show higher intensities than the UAE as well.[42,43] From the myxochromide compound family, only one derivative

(A3) is observed.[44] Here, we find highest intensities in the EtOAc extracts. Myxochromide A3 is also found in the ^{i}PrOH and both of the MeOH extracts. In those extracts the detected intensities of the secondary metabolite are on a similar level as in the UAE, indicating that only SF extraction using EtOAc is superior in extraction efficiency compared to UAE in this case. I In the *Sandaracinus* MSr10575 extracts, three known myxobacterial secondary metabolite families are detectable (Figure 2 F-H).
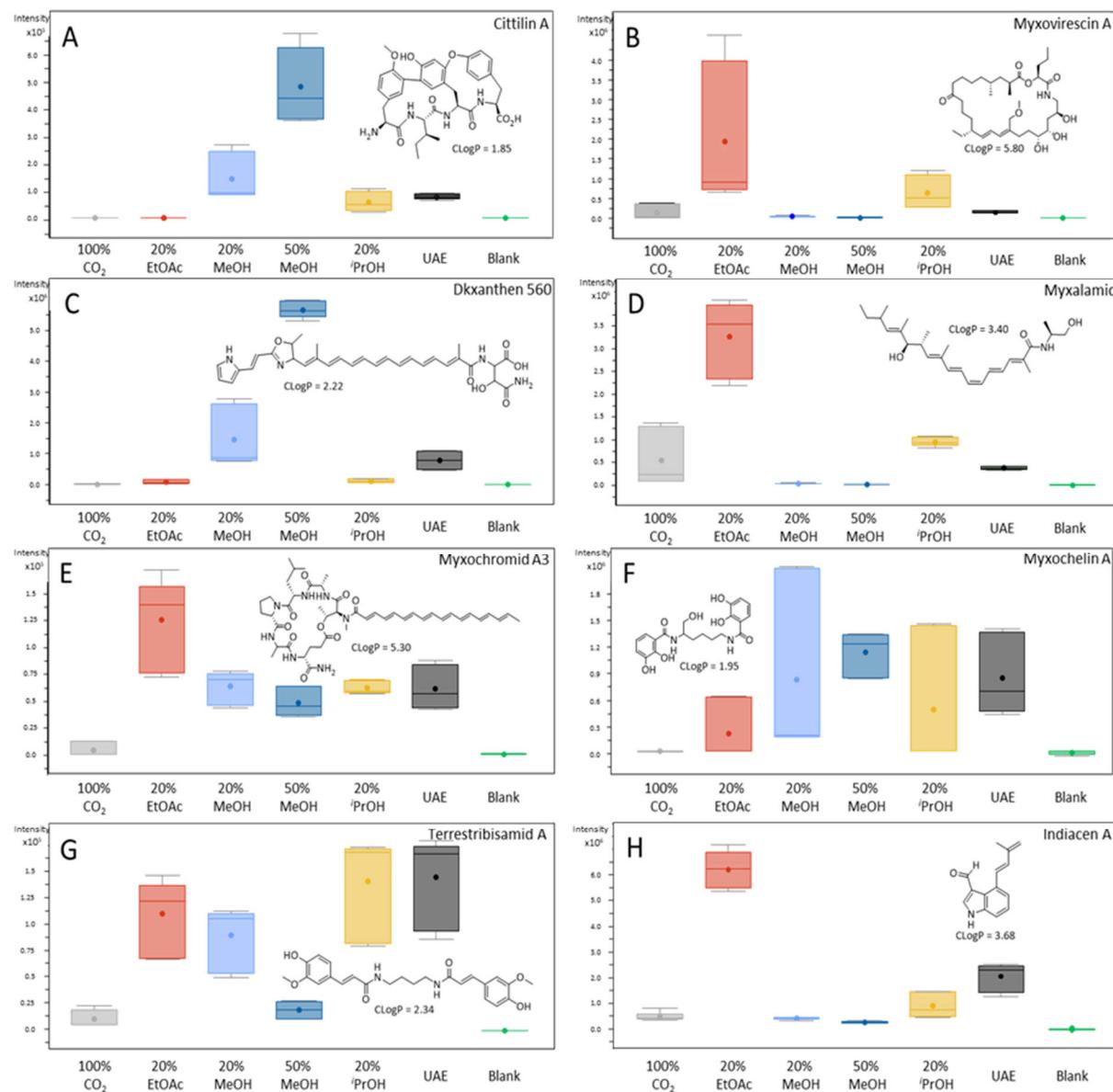
Figure 2 Influence of the co-solvent on the supercritical fluid extraction of secondary metabolites from M. xanthus DK1622 (**A-E**) and MSr10575 (**F-H**). Detected intensities are shown as box plots. From left to right: gray 100% CO_2, red 20% EtOAc, light blue 20% MeOH, dark blue 50% MeOH, yellow 20% ^{i}PrOH, black UAE and green combined blank (SFE+UAE).

From the two natural myxochelins A and B only myxochelin A is observed in the extracts. Highest intensities are found in the 20% MeOH extracts, but the compound is present in all other extracts except the 100% CO_2 extract. Terrestribisamid A is the only known secondary metabolite in our extracts showing the highest intensity in one of the UA extracts.[45] Nevertheless, this compound is also extractable with SFE, whereby ^{i}PrOH shows the best extraction efficiency. Indiacen A and B, two secondary metabolites originally isolated from the myxobacterial strain *Sandaracinus amylolyticus* NOSO-4T also belonging to the family of *Sandaracineae* are also found in the extracts of MSr10575.[46] The Indiacens show highest intensities in the EtOAc extracts. As those prenylated indole aldehydes with

activity against Gram–positive and Gram–negative bacteria as well as the fungus *Mucor hiemalis* are the only reported secondary metabolites from another *Sandaracinus* strain to date, we wanted to validate our initial hit based on exact mass and retention time and isolated both derivatives. Crystallization followed by X-ray crystallography of the isolated compounds confirms that MSr10575 is indeed able to produce Indiacens (see SI).

All known secondary metabolites described above are extractable with UAE as a conventional extraction method and SFE. Concerning the different co-solvents investigated, EtOAc and MeOH are found to be the most efficient co-solvents for extraction. MeOH-based SF extracts show specificity for more polar myxobacterial secondary metabolites, while EtOAc as a co-solvent leads to increased extraction of more nonpolar secondary metabolites. Furthermore, extraction with 50% MeOH as co-solvent seems to be superior to the extraction with 20% MeOH. Due to the relatively high pressure of 200 bar in our experiments, the CO_2/MeOH mixture remains in the supercritical state even when increasing to an equipotential portion of the co-solvent. This is reflected in the high extraction efficiency for more polar natural products such as the cittilins and the DKxanthenes.[47,48] Our data suggests that secondary metabolites such as myxovirescin A (CLogP =5.80), myxalamid A (CLogP = 3.40), myxochromid A3 (CLogP = 5.30) and indiacen A (CLogP = 3.68) exceeding a logP of 3 show best extraction using 20% EtOAc as a co-solvent. Secondary metabolites such as cittilin A (CLogP = 1.85), DKxanthen 560 (CLogP = 2.22), myxochelin A (CLogP = 2.34) and terrestribisamid A (CLogP = 2.34) with logP values below 3 are better extractable using more polar and protic co-solvents such as *i*PrOH and MeOH (see **Error! Reference source not found.**2). This logP dependency can also be observed during consecutive UA extraction with different solvents (see chapter 3.4 SI).

3.4.2 Antimicrobial activities of the SF and UA extracts

All of the three examined myxobacterial strains show potent activities against at least three test organisms of our antimicrobial test panel (see **Error! Reference source not found.**3). Extracts of *Sandaracinus* MSr10575 display the highest and broadest activities, with antimicrobial activities of at least one of the extracts against each of the test organisms. The conventional UA extract shows activities against *S. aureus*, *E. coli*, *E. coli* TolC (representing an efflux deficient *E. coli* strain), *B. subtilis*, *M. luteus*, *P. aeruginosa* and *M. smegmatis*. All of these activities are also found in the 20% MeOH and 50% MeOH SF extracts, indicating that the responsible metabolites are relatively polar. Interestingly, the 100% CO_2 extracts show a selective activity against Gram-positive pathogens as well as *C. albicans*. As those activities do not correlate to the ones observed in other extracts, it is very likely that they are caused by a different family of natural products. Furthermore, we observe activities of the *i*PrOH based SF extract against *M. luteus*, *M. smegmatis* and *C. albicans*. The EtOAc extract on the contrary shows a selective activity against *M. smegmatis*. Besides the activities of the *i*PrOH extract against *M. luteus* and *C. albicans* all activities are comparably strong, showing an inhibition down to a dilution of the extract to 0.1% of the starting concentration. This indicates that the selectivity of antimicrobial activities does not result from the inaccuracy of the assay or the presence of compound yields close to their minimum inhibitory concentration (MIC), but rather the presence of several different antimicrobial natural product classes in the extract. From the known myxobacterial secondary metabolites produced by this strain, only indiacen is known for antimicrobial activity.[46] Although it was not tested against *M. smegmatis* itself, it exhibits activities against two related mycobacterial species. By far the highest intensities of indiacens are measured in our EtOAc extracts. This finding correlates with the selective inhibition of *M. smegmatis* by the EtOAc extract. Nevertheless, the strong other antimicrobial activities

cannot be explained by the presence of already described secondary metabolites alone, making this strain a target for further secondary metabolite extraction campaigns.
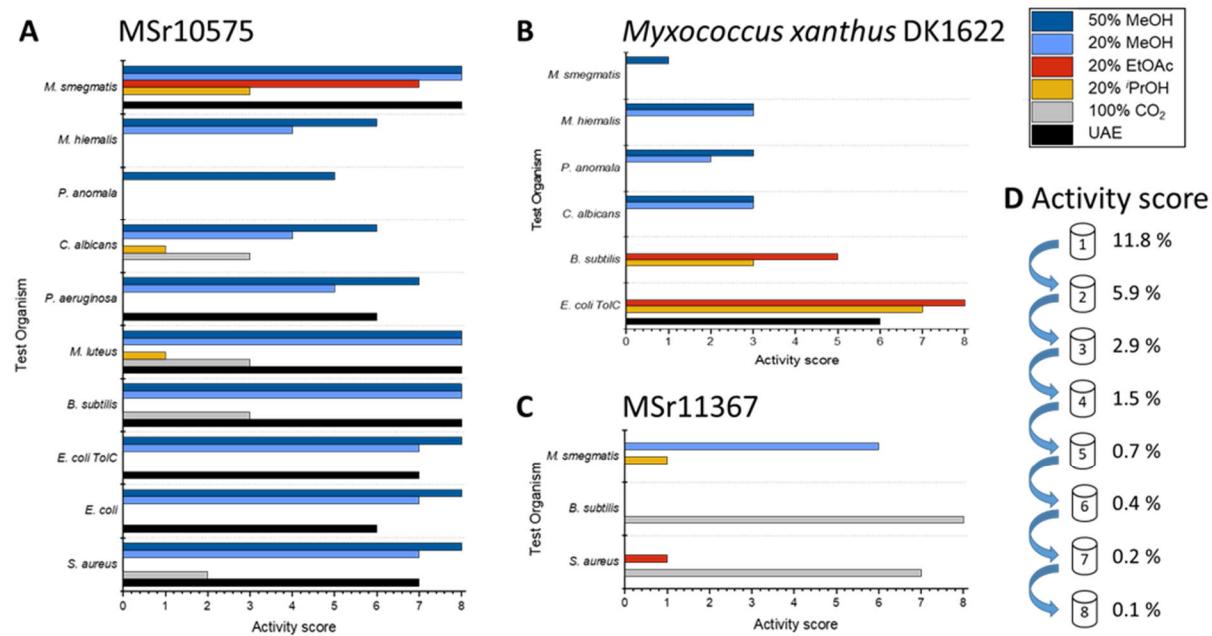
Figure 3 Antimicrobial activity score of myxobacterial SF extracts in comparison to conventional UA extraction (black). **A** MSr10575, **B** M. xanthus DK1622, **C** MSr11367. Extract with 50% MeOH as co-solvent is shown in dark blue, 20% MeOH in light blue, 20% EtOAc in red, 20% ^{i}PrOH in yellow and 100% CO_2 in grey. **D** Activity score with correlating concentration of the crude extract showing an inhibition of the test organism.

The UA extract of *M. xanthus* DK1622 shows activities against *E. coli* TolC, explained by the presence of myxovirescins and myxalamids.[49,50] Anti-*E. coli* TolC activity was also found for the EtOAc and ^{i}PrOH extracts where myxovirescins and myxalamids are present in higher intensities than in the UAE. Furthermore, those two extracts show activity against *B. subtilis*, explained by the presence of the same natural product classes. In the UA extract the concentration of those secondary metabolites very likely does not reach the minimal inhibitory concentration, wherefore no activity against *B. subtilis* is detectable. We additionally observe antifungal activities of the SF MeOH extracts as well as antimycobacterial activity only observed for the 50% MeOH extract. None of these activities can be correlated to known secondary metabolites detectable in the crude extract, highlighting that the potential of secondary metabolite discovery even in well-described strains like *M. xanthus* DK1622 is far from exhausted.

MSr11367 only shows few activities compared to the other two strains. Contrary to the other myxobacterial strains we observe no antimicrobial activities in the UA extract at all. The 100% CO_2 extract of this strain, however, shows the highest antimicrobial activities when just comparing the 100% CO_2 extracts of the different strains and represents the only case where we can detect an activity selectively in the extract without co-solvent. Furthermore, we observe a strong activity of the 20% MeOH extract against *M. smegmatis* besides weak activities of the ^{i}PrOH extract against *M. smegmatis* and of the EtOAc extract against *S. aureus*. In all our extracts from MSr11367 we were not able to detect any known myxobacterial natural product, wherefore these activities offer the possibility for bioactivity-guided isolation of novel secondary metabolites.

In summary, for all three myxobacterial strains we were able to enhance the detectability of biological activities using SFE instead of UAE. Furthermore, each of the strains showed additional activities for the SF extracts, not observable for the UA extracts. This increase in detectable antimicrobial activities may stem from extraction of a larger dynamic range as well as enrichment of certain active secondary metabolites by introducing some selectivity to the extraction process. This selectivity can as well be introduced by using solvents with different polarity for UAE (see 3.3 SI).

3.4.3 Non-targeted metabolomics investigation of the extracts

Besides the targeted metabolomics investigation of the three myxobacterial strains' extracts as well as their biological characterization, we sought to access the total chemical complexity of the different extracts to fully characterize overlaps and differences of the two extraction systems as well as the influence of the different co-solvents. We therefore performed a non-targeted metabolomics analysis and compared features detected in the UAE with those detected in the SF extracts. Furthermore, we compared features detected in the individual SF extracts resulting from the use of different co-solvents. The results of this investigation for strain MSr10575 are depicted in Figure 4 as Euler diagrams. Results for the two other strains can be found in the SI.
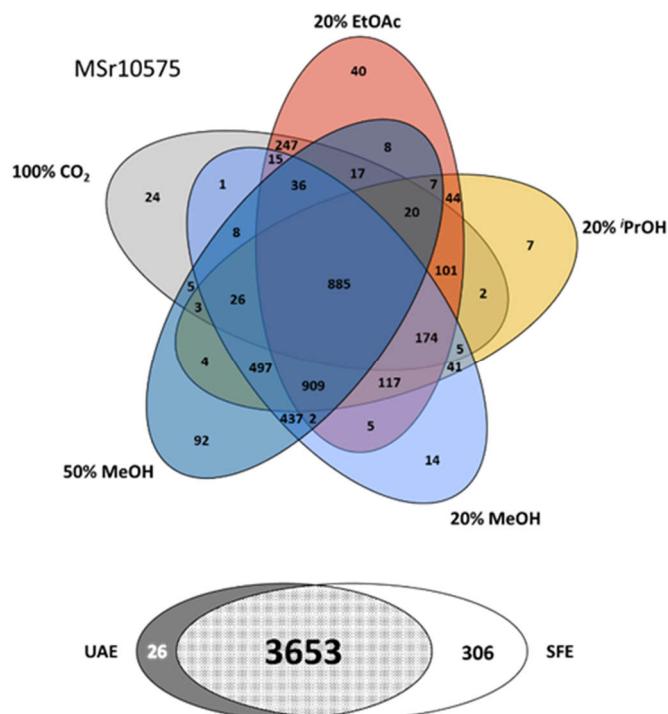
Figure 4 Upper part: Euler diagram showing features found in the SF extracts from MSr10575 sorted by co-solvent. Intersections of the extracts generated using different co-solvents are shown in the same colors as used in Fig. 1 and 2. Lower part: Euler diagram showing features unique for UAE or SFE and features extractable with both methods.

The overlap of features found with UAE and SFE is very high. 92% of the total features can be extracted with both methods. Looking at the features uniquely extractable with one of the methods, about 12% more features are only present in the SF extract than in the UA extract. If it is only a matter of accessing as many features as possible with one method, SFE should therefore be the method of choice. The comparison of the different co-solvents used for the extraction shows that the number of metabolites extractable with only one co-solvent is relatively low. Among the different co-solvents, extraction with 50% MeOH yielded the most features. In general, most of the metabolites are detectable in all of the SF extracts (885 features) or in all of the SF extracts except the CO_2 extract (909 features). Omitting

the first extraction step with 100% CO_2 would only result in a data loss of less than 1% as all other metabolites are also accessible with the addition of co-solvents and should therefore be preferred if the extraction protocol needs to be shortened.

During the non-targeted investigation of the SF extracts of *M. xanthus* DK1622 a group of metabolites raised our interest. They all show significantly higher intensities in the SF extract using EtOAc as a co-solvent and an isotopic pattern characteristic for chlorination as depicted in **Error! Reference source not found.**5. Halogenated natural products are well-known to exhibit diverse biological activities, making those features particularly interesting for further investigation. The presence of 30 features with characteristic chlorination-indicating isotope pattern furthermore indicated, that they belong to a relatively large family of secondary metabolites (see SI). According to MS analysis all family members incorporate a minimum of two nitrogen and two oxygen atoms. To investigate the chemical structure of those compounds, we isolated the main derivative, later named chloroxanthic acid A. As the other compounds are only produced in comparably low amounts, they were not accessible to full structure elucidation by NMR.

3.4.4 Structure elucidation of chloroxanthic acid A

HRESIMS of chloroxanthic acid A shows an [M+H]+ peak at *m/z* 295.0848 consistent with the molecular formula $C_{14}H_{15}ClN_2O_3$ (*m/z* calcd for [M+H]+ 295.0844) containing 8 double-bond equivalents (DBE). The ^{1}H-NMR spectrum of chloroxanthic acid A exhibits five signals characteristic for double bond protons at δ = 7.14 (1H, d, *J* = 16.18 Hz), δ = 6.38 (2H, m), δ = 6.18 (1H, d, *J* = 16.09 Hz), δ = 6.03 (1H, d, *J* = 3.66 Hz). Furthermore, two methyl groups at δ = 1.94 (3H, d, *J* = 1.44 Hz) and δ = 1.43 (3H, dd, *J* = 7.00 Hz) and two methine groups at δ = 4.55 (1H, dd, *J* = 9.30 , *J* = 7.18 Hz) and δ = 4.46 (1H, m) are observed. Their coupling constants as well as COSY correlations reveal the two protons at δ = 7.14 and δ = 6.18 to be part of an aliphatic double bond in *E*-configuration. Their down-field chemical shift suggests participation in a bigger conjugated π-system. HMBC correlations show their location between two different substructures (substructure A and B) of the molecule (see Figure 5). COSY and HMBC correlations reveal, that substructure A consists of the two double bond protons at δ = 6.03 and δ = 6.38 as well as two quaternary carbons at δ = 129.8 and δ = 120.2. Their characteristic chemical shifts suggest that substructure A is a pyrrole ring with chlorination at the quaternary carbon at δ = 120.2 and the quaternary carbon at δ = 129.8 is the linker to the aliphatic double bond. The splitting pattern of the two protons of the pyrrole moiety as well as their COSY correlations confirm their allocation next to each other, leaving only two possibilities for substitution of the pyrrole ring. The downfield shifts of both the chlorine bearing carbon at δ = 120.2 and the quaternary carbon at δ = 129.8 finally reveal chlorination in 5 position of the pyrrole ring. In substructure B the two aliphatic double bond protons only show HMBC correlations to one quaternary carbon at δ = 166.8. Its characteristic chemical shift as well as further HMBC correlations to the two methine groups at δ = 4.55 and δ = 4.46 suggest an oxazoline core structure. COSY correlations reveal a substitution with the methyl group at δ = 4.46 in the 5 position. The 4 position of the oxazoline moiety is linked to 2-methylacrylic acid, which are confirmed by COSY and HMBC correlations.
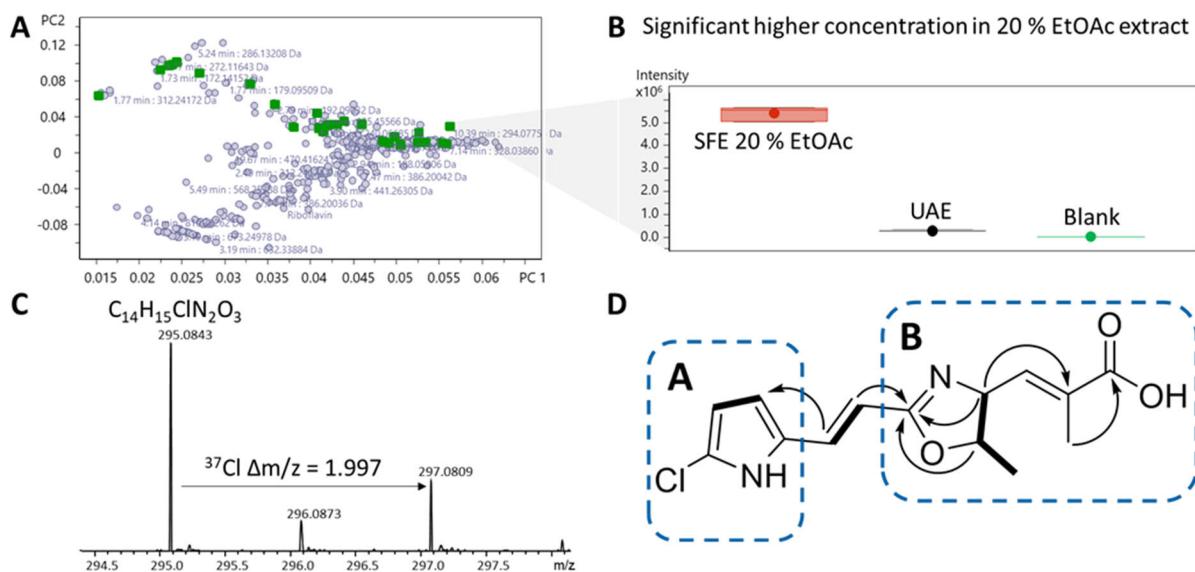
Figure 5 Chlorinated metabolites from M. xanthus DK1622. **A** PCA plot of the comparison between UAE and SFE 20% EtOAc extracts. Chlorinated features are highlighted in green. **B** Box plot for chloroxanthic acid A highlighting the differences in intensities between UAE and SFE EtOAC extract. **C** Zoom into the isotopic pattern of the $[M+H]^+$ peak of chloroxanthic acid A showing the characteristic mass shift of the ^{37}Cl isotope. **D** Structure of chloroxanthic acid A. Most relevant COSY correlations are marked bold and most relevant HMBC correlations are shown as arrows. Substructures A and B are marked in dashed blue boxes.

Chloroxanthic acid A shows partial structural similarity to the DKxanthenes that were already described from *M. xanthus* DK1622 before.[41] The coupling constant of $^3J = 7.2$ Hz indicate a *Z* configuration of the two protons participating in the chloroxanthic acid A oxazoline ring, which is in line with the configuration determined for the DKxanthenes. The DKxanthenes are produced by a hybrid polyketide synthase – non-ribosomal peptide synthetase (PKS-NRPS) gene cluster.[51] The two NRP parts of the molecules are linked by a polyketide chain consisting of an aliphatic conjugated double bond system in variable chain lengths. Chloroxanthic acid A only comprises the first NRP part of the DKxanthenes, elongated with one methylmalonyl-CoA unit of the PK part. To the best of our knowledge, halogenated DKxanthene derivatives have not been described previously, nor is a halogenase part of the reported biosynthetic gene clusters of those natural products.[41,51] The isolation and structure elucidation of chloroxanthic acid A highlights the impact of applying deviant analytical chemistry approaches for the discovery of novel natural products from already well-studied bacterial strains. Besides the setup chosen for the detection of secondary metabolites as shown by the discovery of the myxoprincomides by Cortina *et al.,* the extraction method has a tremendous influence on the accessibility of natural products.

3.5 Conclusion

SFE in microbial secondary metabolomics is a promising yet underexplored technique for the discovery and extraction of new natural products. In our study we were able to show that for each known metabolite in our three myxobacterial strains, supercritical CO_2 extraction efficiency was comparable or superior to conventional ultrasound assisted extraction. We developed an SFE protocol suitable for all metabolites that resulted in higher extraction yields for most target compounds. The choice of co-solvent herein plays an important role. For more polar compounds such as the cittilins or DKxanthenes, MeOH is the best co-solvent for extraction amongst the co-solvents tested in our study. Here we also observe that the use of 50% MeOH is superior to 20% MeOH. For myxobacterial natural products with a logP higher than 3, namely the myxovirescins, myxalamides, indiacens and myxochromids, the most

efficient co-solvent for extraction is EtOAc. Estimating the polarity of the secondary metabolite to be extracted therefore is crucial for the choice of a matching co-solvent. SFE offers the possibility to extract secondary metabolites from bacteria while simultaneously introducing some degree of selectivity, whereas solvent extractions have a much broader extraction scope. Non-targeted analysis of the extracts underpin the notion that the choice of co-solvent (or rather the use of co-solvent at all) plays the most important role in SFE. Only less than one percent of myxobacterial compounds are extractable with pure CO_2 alone. Besides the many compounds extractable under all conditions, most myxobacterial secondary metabolites in this study are extractable when adding a polar co-solvent like ^{i}PrOH or MeOH. In comparison to UAE we find 12% more features uniquely extractable with SFE, highlighting that SFE should be preferred over standard UAE using only MeOH and acetone as solvents in order to access as many metabolites as possible. We also observe an increase in antimicrobial activities of the SF extracts compared to the standard UA extracts. Importantly, some antimicrobial activities could only be seen in the SF extracts indicating that certain bioactive metabolites can only be detected and isolated using the SF-based approach. This is exemplified by strain MSr11367, where we could not observe any bioactivities from the UAE at all.

Our non-targeted analysis of the extracts shows that the overall number of metabolites that are exclusively extractable with SFE is comparably low. Nevertheless, this result highlights that the most powerful feature of SFE is its capability to increase the metabolite concentration in the resulting extract fractions thereby making new metabolites accessible. This finding is also reflected in the increased bioactivities seen in SF extracts. Many secondary metabolites are generally detectable with highly sensitive mass spectrometry but may not give rise to bioactivity in assays as these often suffer from comparably low sensitivity, especially in case of cell based assays aiming to identify antifungals and antibacterials. The use of SFE can therefore help to detect and prioritize new natural products based on their antimicrobial activities. Furthermore, our study shows that SFE can be used to prioritize secondary metabolites for isolation and structure elucidation based on their chemical properties, as demonstrated here for chloroxanthic acid A. In this case, SFE achieved concentrations in the extract at which isolation and structure elucidation became feasible, without the necessity to upscale to very high culture volumes. Especially for myxobacteria – where the change in cultivation conditions during upscaling often results in a decrease of secondary metabolite production - SFE thus paves the way to access secondary metabolites that have not been uncovered to date.[52] Taken together, the advantageous features of SFE encourage its use for a wider range of microbial natural products applications in the future.

3.6 References

(1) Rafińska, K.; Pomastowski, P.; Rudnicka, J.; Krakowska, A.; Maruśka, A.; Narkute, M.; Buszewski, B. Effect of solvent and extraction technique on composition and biological activity of Lepidium sativum extracts. *Food Chem.* **2019**, *289*, 16–25.

(2) Manjare, S. D.; Dhingra, K. Supercritical fluids in separation and purification: A review. *Mater. Sci. Technol.* **2019**, *2*, 463–484.

(3) Khaw, K.-Y.; Parat, M.-O.; Shaw, P. N.; Falconer, J. R. Solvent Supercritical Fluid Technologies to Extract Bioactive Compounds from Natural Sources: A Review. *Molecules (Basel, Switzerland)* **2017**, *22*.

(4) Chollet, C.; Boutet-Mercey, S.; Laboureur, L.; Rincon, C.; Méjean, M.; Jouhet, J.; Fenaille, F.; Colsch, B.; Touboul, D. Supercritical fluid chromatography coupled to mass spectrometry for lipidomics. *J. Mass Spectrom.* **2019**, *54*, 791–801.

(5) Wrona, O.; Rafińska, K.; Walczak-Skierska, J.; Możeński, C.; Buszewski, B. Extraction and Determination of Polar Bioactive Compounds from Alfalfa (Medicago sativa L.) Using Supercritical Techniques. *Molecules (Basel, Switzerland)* **2019**, *24*.

(6) Dutta, S.; Bhattacharjee, P. Enzyme-assisted supercritical carbon dioxide extraction of black pepper oleoresin for enhanced yield of piperine-rich extract. *J Biosci Bioeng* **2015**, *120*, 17–23.

(7) Yang, X.; Li, Y.; Li, Y.; Ye, D.; Yuan, L.; Sun, Y.; Han, D.; Hu, Q. Solid Matrix-Supported Supercritical CO_2 Enhances Extraction of γ-Linolenic Acid from the Cyanobacterium Arthrospira (Spirulina) platensis and Bioactivity Evaluation of the Molecule in Zebrafish. *Marine drugs* **2019**, *17*.

(8) Cocks, S.; Wrigley, S. K.; Chicarelli-Robinson, M. I.; Smith, R. M. High-performance liquid chromatograpy comparison of supercritical-fluid extraction and solvent extraction of microbial fermentation products. *J. Chromatogr. A* **1995**, *697*, 115–122.

(9) Knudsen, P. B.; Hanna, B.; Ohl, S.; Sellner, L.; Zenz, T.; Döhner, H.; Stilgenbauer, S.; Larsen, T. O.; Lichter, P.; Seiffert, M. Chaetoglobosin A preferentially induces apoptosis in chronic lymphocytic leukemia cells by targeting the cytoskeleton. *Leukemia* **2014**, *28*, 1289–1298.

(10) Li, B.; Gao, Y.; Rankin, G. O.; Rojanasakul, Y.; Cutler, S. J.; Tu, Y.; Chen, Y. C. Chaetoglobosin K induces apoptosis and G2 cell cycle arrest through p53-dependent pathway in cisplatin-resistant ovarian cancer cells. *Cancer Lett.* **2015**, *356*, 418–433.

(11) Tadić, V.; Bojović, D.; Arsić, I.; Dorđević, S.; Aksentijevic, K.; Stamenić, M.; Janković, S. Chemical and antimicrobial evaluation of supercritical and conventional Sideritis scardica Griseb., Lamiaceae extracts. *Molecules (Basel, Switzerland)* **2012**, *17*, 2683–2703.

(12) Wrona, O.; Rafińska, K.; Możeński, C.; Buszewski, B. Supercritical Fluid Extraction of Bioactive Compounds from Plant Materials. *J. AOAC Int.* **2017**, *100*, 1624–1635.

(13) Silva, F. V. M.; Martins, A.; Salta, J.; Neng, N. R.; Nogueira, J. M. F.; Mira, D.; Gaspar, N.; Justino, J.; Grosso, C.; Urieta, J. S.*; et al.* Phytochemical profile and anticholinesterase and antimicrobial activities of supercritical versus conventional extracts of Satureja montana. *J. Agric. Food Chem.* **2009**, *57*, 11557–11563.

(14) Esquivel-Hernández, D. A.; Rodríguez-Rodríguez, J.; Cuéllar-Bermúdez, S. P.; García-Pérez, J. S.; Mancera-Andrade, E. I.; Núñez-Echevarría, J. E.; Ontiveros-Valencia, A.; Rostro-Alanis, M.; García-García, R. M.; Torres, J. A.*; et al.* Effect of Supercritical Carbon Dioxide Extraction Parameters on the Biological Activities and Metabolites Present in Extracts from Arthrospira platensis. *Marine drugs* **2017**, *15*.

(15) Aresta, M. *Carbon dioxide as chemical feedstock*; Green chemistry; Wiley-VCH: Weinheim, 2010.

(16) Pourmortazavi, S. M.; Hajimirsadeghi, S. S. Supercritical fluid extraction in plant essential and volatile oil analysis. *J. Chromatogr. A* **2007**, *1163*, 2–24.

(17) Alvarez-Sanchez, B.; Priego-Capote, F.; de Castro,M. D. L. Metabolomics analysis II. Preparation of biological samples prior to detection. *Trac-Trends Anal Chem* **2010**, *29*, 120–127.

(18) Tatsuya Ito; Miyako Masubuchi. Dereplication of microbial extracts and related analytical technologies. *J Antibiot* **2014**, *67*, 353–360.

(19) Seidel, V. Initial and Bulk Extraction of Natural Products Isolation. In *Natural products isolation,* 3rd ed. / edited by Satyajit D. Sarker, Lutfun Nahar; Sarker, S. D., Nahar, L., Eds.; Methods in molecular biology, 1064-3745 864; Humana Press: New York, 2012; pp 27–41.

(20) Zhang, L. Integrated Approaches for Discovering Novel Drugs From Microbial Natural Products. In *Natural Products*; Zhang, L., Demain, A. L., Eds.; Humana Press Inc: [New York], 2005; pp 33–55.

(21) 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.

(22) Zhang, F.; Braun, D. R.; Rajski, S. R.; DeMaria, D.; Bugni, T. S. Enhypyrazinones A and B, Pyrazinone Natural Products from a Marine-Derived Myxobacterium Enhygromyxa sp. *Marine drugs* **2019**, *17*.

(23) Surup, F.; Chauhan, D.; Niggemann, J.; Bartok, E.; Herrmann, J.; Koeck, M.; Zander, W.; Stadler, M.; Hornung, V.; Müller, R. Activation of the NLRP3 inflammasome by hyaboron, a new asymmetric boron-containing macrodiolide from the Myxobacterium *Hyalangium minutum*. *ACS Chem. Biol.* **2018**, *13*, 2981–2988.

(24) Gorges, J.; Panter, F.; Kjaerulff, L.; Hoffmann, T.; Kazmaier, U.; Müller, R. Structure, Total Synthesis, and Biosynthesis of Chloromyxamides: Myxobacterial Tetrapeptides Featuring an Uncommon 6-Chloromethyl-5-methoxypipecolic Acid Building Block. *Angew. Chem. Int. Ed. Engl.* **2018**, *57*, 14270–14275.

(25) Kjaerulff, L.; Raju, R.; Panter, F.; Scheid, U.; Garcia, R.; Herrmann, J.; Müller, R. Pyxipyrrolones: Structure elucidation and biosynthesis of cytotoxic myxobacterial metabolites. *Angew. Chem. Int. Ed.* **2017**, *56*, 9614–9618.

(26) Nadmid, S.; Plaza, A.; Garcia, R.; Müller, R. Cystochromones, unusual chromone-containing polyketides from the myxobacterium *Cystobacter* sp. MCy9104. *J. Nat. Prod.* **2015**, *78*, 2023–2028.

(27) Okoth Dorothy A.; Hug, J. J.; Garcia, R.; Spröer, C.; Overmann, J.; Müller, R. 2-Hydroxysorangiadenosine: Structure and Biosynthesis of a Myxobacterial Sesquiterpene–Nucleoside. *Molecules (Basel, Switzerland)* **2020**, *25*.

(28) Oueis, E.; Klefisch, T.; Zaburannyi, N.; Garcia, R.; Plaza, A.; Müller, R. Two Biosynthetic Pathways in Jahnella thaxteri for Thaxteramides, Distinct Types of Lipopeptides. *Org. Lett.* **2019**, *21*, 5407–5412.

(29) Hug, J. J.; Panter, F.; Krug, D.; Müller, R. Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria. *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 319–334.

(30) Sester, A.; Winand, L.; Pace, S.; Hiller, W.; Werz, O.; Nett, M. Myxochelin- and Pseudochelin-Derived Lipoxygenase Inhibitors from a Genetically Engineered Myxococcus xanthus Strain. *J. Nat. Prod.* **2019**, *82*, 2544–2549.

(31) Panter, F.; Krug, D.; Baumann, S.; Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **2018**, *9*, 4898–4908.

(32) Hoffmann, T.; Krug, D.; Bozkurt, N.; Duddela, S.; Jansen, R.; Garcia, R.; Gerth, K.; Steinmetz, H.; Müller, R. Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat. Commun.* **2018**, *9*, 803.

(33) Dehhaghi, M.; Tan, V.; Heng, B.; Mohammadipanah, F.; Guillemin, G. J. Protective Effects of Myxobacterial Extracts on Hydrogen Peroxide-induced Toxicity on Human Primary Astrocytes. *Neuroscience* **2019**, *399*, 1–11.

(34) Lee, C.; Park, S.; Ayush, I.; Cho, K.; Kim, S. S.; Kang, I.; Choe, W.; Kim, Y.-S.; Yoon, K.-S. Effects of Myxococcus fulvus KYC4048 Metabolites on Breast Cancer Cell Death. *J. Microbiol. Biotechnol.* **2018**, *28*, 765–775.

(35) Bader, C. D.; Panter, F.; Müller, R. In depth natural product discovery - Myxobacterial strains that provided multiple secondary metabolites. *Biotechnol. Adv.* **2020**, *39*, 107480.

(36) Wenzel, S. C.; Müller, R. Myxobacteria—''microbial factories' for the production of bioactive secondary metabolites. *Mol. Biosyst.* **2009**, *5*, 567–574.

(37) Abbas, C. A.; Sibirny, A. A. Genetic control of biosynthesis and transport of riboflavin and flavin nucleotides and construction of robust biotechnological producers. *Microbiol Mol Biol Rev* **2011**, *75*, 321–360.

(38) Hug, J. J.; Dastbaz, J.; Adam, S.; Revermann, O.; Koehnke, J.; Krug, D.; Müller, R. Biosynthesis of Cittilins, Unusual Ribosomally Synthesized and Post-translationally Modified Peptides from Myxococcus xanthus. *ACS Chem. Biol.* **2020**, *15*, 2221–2231.

(39) Krug, D.; Zurek, G.; Revermann, O.; Vos, M.; Velicer, G. J.; Müller, R. Discovering the Hidden Secondary Metabolome of *Myxococcus xanthus*: a Study of Intraspecific Diversity. *Appl. Environ. Microbiol.* **2008**, *74*, 3058–3068.

(40) Trowitzsch, W.; Wray, V.; Gerth, K.; Höfle, G. Structure of myxovirescin A, a new macrocylic antibiotic from gliding bacteria. *J. Chem. Soc., Chem. Commun.* **1982**, 1340.

(41) Meiser, P.; Bode, H. B.; Müller, R. The unique DKxanthene secondary metabolite family from the myxobacterium *Myxococcus xanthus* is required for developmental sporulation. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19128–19133.

(42) Jansen, R.; Reifenstahl, G.; Gerth, K.; Reichenberg, H.; Höfle, G. Myxalamide A,B,C und D, eine Gruppe homologer Antibiotika aus *Myxococcus xanthus* Mx x12 (Myxobacterales). *Liebigs Ann. Chem.* **1983**, 1081–1095.

(43) Kim, J.; Choi, J. N.; Kim, P.; Sok, D. E.; Nam, S. W.; Lee, C. H. LC-MS/MS Profiling-Based Secondary Metabolite Screening of *Myxococcus xanthus*. *J. Microbiol. Biotechnol.* **2009**, *19*, 51–54.

(44) Trowitzsch Kienast, W.; Gerth, K.; Reichenbach, H.; Höfle, G. Myxochromid A: Ein hochungesättigtes Lipopeptidlacton aus *Myxococcus virescens*. *Liebigs Ann. Chem.* **1993**, 1233–1237.

(45) Iwasa, K.; Takahashi, T.; Nishiyama, Y.; Moriyasu, M.; Sugiura, M.; Takeuchi, A.; Tode, C.; Tokuda, H.; Takeda, K. Online structural elucidation of alkaloids and other constituents in crude extracts and cultured cells of Nandina domestica by combination of LC-MS/MS, LC-NMR, and LC-CD analyses. *J. Nat. Prod.* **2008**, *71*, 1376–1385.

(46) Steinmetz, H.; Mohr, K. I.; Zander, W.; Jansen, R.; Gerth, K.; Müller, R. Indiacens A and B: prenyl indoles from the *myxobacterium Sandaracinus amylolyticus*. *J. Nat. Prod.* **2012**, *75*, 1803–1805.

(47) Helka Turunen. CO2-Balance in the Atmosphere and CO2-Utilisation: an engineering approach. Academic dissertation, University of Oulu, Linnanmaa, 2011.

(48) RAMSEY, E. D. *Analytical Supercritical Fluid Extraction Techniques*; Springer Netherlands: Dordrecht, 2012.

(49) Gerth, K.; Jansen, R.; Reifenstahl, G.; Höfle, G.; Irschik, H.; Kunze, B.; Reichenbach, H.; Thierbach, G. The myxalamids, new antibiotics from *Myxococcus xanthus* (Myxobacterales). I. Production, physico-chemical and biological properties, and mechanism of action. *J. Antibiot.* **1983**, *36*, 1150–1156.

(50) Xiao, Y.; Gerth, K.; Müller, R.; Wall, D. Myxobacterium-produced antibiotic TA (myxovirescin) inhibits type II signal peptidase. *Antimicrob. Agents Chemother.* **2012**, *56*, 2014–2021.

(51) Meiser, P.; Weissman, K. J.; Bode, H. B.; Krug, D.; Dickschat, J. S.; Sandmann, A.; Müller, R. DKxanthene biosynthesis–understanding the basis for diversity-oriented synthesis in myxobacterial secondary metabolism. *Chem. Biol.* **2008**, *15*, 771–781.

(52) Hug, J. J.; Bader, C. D.; Remškar, M.; Cirnski, K.; Müller, R. Concepts and Methods to Access Novel Antibiotics from Actinomycetes. *Antibiotics* **2018**, *7*, 44.

3.7 Supporting information

Supercritical Fluid Extraction Enhances Discovery of Secondary Metabolites from Myxobacteria

Chantal D. Bader, Markus Neuber, Fabian Panter, Daniel Krug and Rolf Müller

Affiliation

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)

Data with limited visibility on print media such as raw NMR data can be found on the enclosed storage medium.

S3.1   Extract generation for statistical analysis and bioactivity profiling

S3.1 Myxobacterial culture media

All culture 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. Cultivations were performed in triplicates as biological replicates.

S3.1.1 Seed cultures

As seed culture, *M. xanthus* DK1622 was cultivated in CTT medium, MSr10575 in 2-SWT medium and MSr11367 in SHG-P medium, each in 100 mL culture volume in a 300 mL shaking flasks. The strains were grown at 30°C and 180 rpm on an Orbitron shaker for three days prior to inoculation of the production cultures.

S3.1.2 Production cultures

As production culture, *M. xanthus* DK1622 was cultivated in SHG-P medium, MSr10575 in 2-SWT medium and MSr11367 in PYGS medium, each in 400 mL culture volume in 2000 mL shaking flasks. The strains were grown at 30°C and 180 rpm on an orbital shaker for ten days prior to inoculation of the production cultures. Inoculation was performed using 5 % [v/v] of the seed cultures. To adsorb the metabolites Amberlite XAD16 resin was added to all production media to a volumetric concentration of 2 % [v/v]. At the end of fermentation, both resin and cells were harvested together by centrifugation (8000 rpm, 4 °C, 20 min).

S3.1.3 Media recipes

Table S1 Medium recipe of CTT medium

Ingredient	Supplier	Amount [mM]	pH adjusted to
Bacto Casitone	Difco	1.0 %	7.2 (KOH)
Tris-HCl		10.0	
KH$_2$PO$_4$		1.0	
MgSO$_4$ x 7 H$_2$O	Grüssing	8.0	

Table S2 Medium recipe of PYGS medium.

Ingredient	Supplier	Amount [%]	pH adjusted to
Potato starch	Sigma-Aldrich	1.0	7.6 (KOH)
Fresh baker's yeast	DHW	2.0	
Glucose	Roth	0.5	
CaCl$_2$ x 2 H$_2$O	VWR Chemicals	0.1	
MgSO$_4$ x 7 H$_2$O	Grüssing	0.1	
TRIS-HCl	Roth	11.9 (50 mM)	

Table S3 Medium recipe of 2-SWT medium.

Ingredient	Supplier	Amount [%]	pH adjusted to
Bacto tryptone	Difco	0.3	7.0 (KOH)
Soytone	BD	0.1	
Glucose	Roth	0.2	
Soluble starch	Roth	0.2	
Maltose monohydrate	Roth	0.1	
Cellobiose	MP-Biomedicals	0.2	
CaCl$_2$ x 2 H$_2$O	VWR Chemicals	0.05	
MgSO$_4$ x 7 H$_2$O	Grüssing	0.1	
HEPES	Roth	10 mM	

Table S4 Medium recipe of SHG-P medium.

Ingredient	Supplier	Amount [%]	pH adjusted to
Corn flour	Sigma-Aldrich	2	7.5 (KOH)
Soluble starch	Roth	1	
Glycerol	Sigma-Aldrich	2	
CaCO$_3$		0.2	

S3.2 Calculation of the Analytical GREEnness

SFE and UAE were compared regarding their environmental friendliness and human safety using the AGREE Analytical GREEnness calculator.[1] Here 12 principles of green analytical chemistry are transformed into a 0-1 scale, as shown below for both extractions. 1 would represent an optimal analytical chemistry process, whereas 0 represents a process which does not meet the criteria. Weight for all criteria was set to 2 to rank all criteria equally.

S3.2.1 SF extraction

Analytical Greenness report sheet

29/09/2020 17:18:23

0.56

Criteria	Score	Weight
1. Direct analytical techniques should be applied to avoid sample treatment.	0.3	2
2. Minimal sample size and minimal number of samples are goals.	0.65	2
3. If possible, measurements should be performed in situ.	0.0	2
4. Integration of analytical processes and operations saves energy and reduces the use of reagents.	0.8	2
5. Automated and miniaturized methods should be selected.	1.0	2
6. Derivatization should be avoided.	1.0	2
7. Generation of a large volume of analytical waste should be avoided, and proper management of analytical waste should be provided.	0.17	2
8. Multi-analyte or multi-parameter methods are preferred versus methods using one analyte at a time.	1.0	2
9. The use of energy should be minimized.	0.5	2
10. Reagents obtained from renewable sources should be preferred.	0.5	2
11. Toxic reagents should be eliminated or replaced.	0.0	2
12. Operator's safety should be increased.	0.8	2

Figure S1 Analytical GREEnness evaluation of SFE.

S3.2.2 UA extraction

Analytical Greenness report sheet

29/09/2020 17:17:30

0.48

| Criteria	Score	Weight
1. Direct analytical techniques should be applied to avoid sample treatment.	0.3	2
2. Minimal sample size and minimal number of samples are goals.	0.65	2
3. If possible, measurements should be performed in situ.	0.0	2
4. Integration of analytical processes and operations saves energy and reduces the use of reagents.	0.6	2
5. Automated and miniaturized methods should be selected.	0.5	2
6. Derivatization should be avoided.	1.0	2
7. Generation of a large volume of analytical waste should be avoided, and proper management of analytical waste should be provided.	0	2
8. Multi-analyte or multi-parameter methods are preferred versus methods using one analyte at a time.	1.0	2
9. The use of energy should be minimized.	0.91	2
10. Reagents obtained from renewable sources should be preferred.	0.0	2
11. Toxic reagents should be eliminated or replaced.	0.0	2
12. Operator's safety should be increased.	0.8	2

Figure S2 Analytical GREEnness evaluation of UAE.

All criteria influenced by sample preparation and subsequent analytical measurements are set to the same values, as those steps are the same for SF and UA extraction. For the two extraction procedures therefore main differences are found in criteria number 5 and 7, as UAE is a fully manual process, whereas SFE is automated. Also the amount of waste generated is reduced for SFE, as at least 50% of the solvent used is CO_2 which can be recycled from other industrial processes. The difference of the total ranking of both extractions however is lowered as UAE requires less than 10% of the energy that is required for SFE (9).

S3.3 Development of the UA extraction protocol

S3.3.1 Extraction conditions

Classical solvent extraction (using one solvent stirring at room temperature) is still the most widespread extraction protocol used for natural product extraction from microorganisms. However, to the best of our knowledge there is no publication available, comparing the parameters chosen for solvent extraction of microorganisms used in recent years. This was only done for plant secondary metabolites, where the extraction procedures greatly differ due to the absence of media components. The UAE protocol used in this study we therefore developed adapted from the extraction protocols used for isolation of novel myxobacterial secondary metabolites and extract generation for comparison of biological functions in the past two years.[2–12] Except for one study, all used MeOH or acetone (or both) for the extraction, wherefore we decided to use a combination of both solvents. Concerning the solid/liquid ratio, those publications used between 4 and 20 % of extraction solvent related to the culture volume for compound isolation. However, as many publications do not mention the amount of solvent used per culture volume or solid mass at all, we decided to take the same amount of solvent as the culture volume, to guarantee exhaustive extraction. This may not represent the optimal method concerning "greenness", but we wanted to circumvent limiting the extraction efficiency by using not enough solvent. In most of the cases the extraction time chosen for compound extraction for subsequent isolation is also not mentioned, but for analytical extraction 60 minutes was the common timespan chosen per solvent, wherefore we took this as widely accepted time for extraction.

S3.3.2 Extraction solvent

To investigate if MeOH and acetone really are the best solvents for extraction of myxobacterial secondary metabolites we carried out a targeted metabolomics comparison of extracts generated with the same solvents as previously used as co-solvents for SFE. For this purpose we re-grew the three myxobacterial strains used in our study in their respective media (later on referred as batch 2) and extracted aliquots of 100 mg dry, pulverized pellet separately with 1.5 mL hexane, EtOAc, ^{i}PrOH, acetone or MeOH following otherwise the same conditions for extraction and analytical evaluation as used for UAE before. As those results stem from a second cultivation and the extracts were generated with a different volume of solvent used for extraction, intensity levels of the detected secondary metabolites as well as biological activities of those extracts cannot directly be compared to the results of the first batch and should be treated as a separate individual evaluation. The box plots showing extraction efficiency for secondary metabolites detectable in the extracts are shown in figure S3-S22.
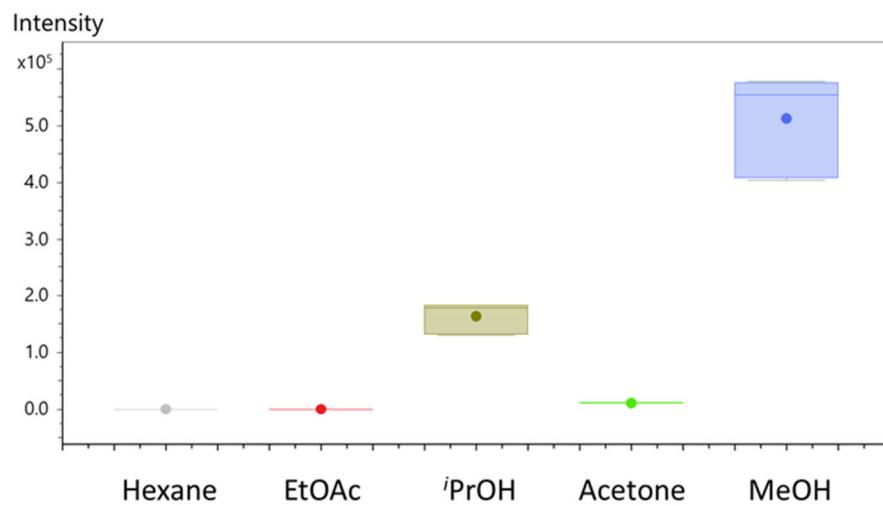
Figure S3 Box plot for cittilin A from *M. xanthus* DK1622 comparing different solvents for UAE.
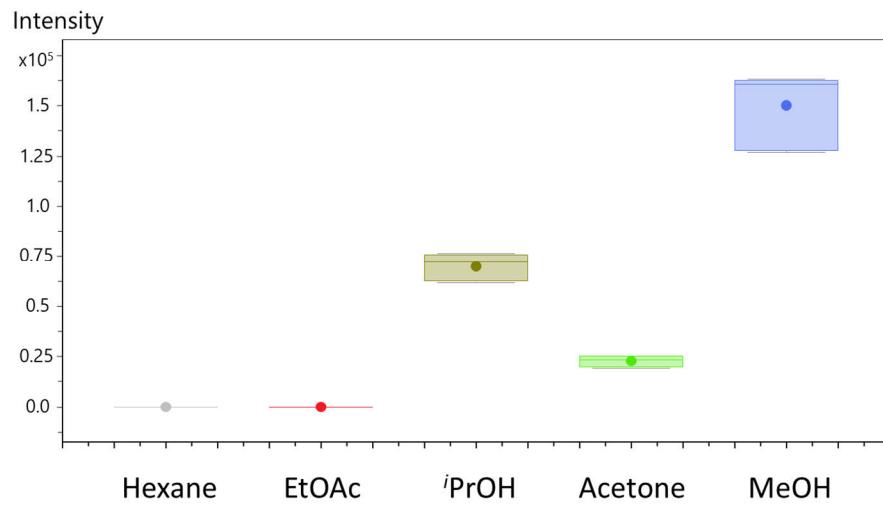
Figure S4 Box plot for riboflavin from *M. xanthus* DK1622 comparing different solvents for UAE.
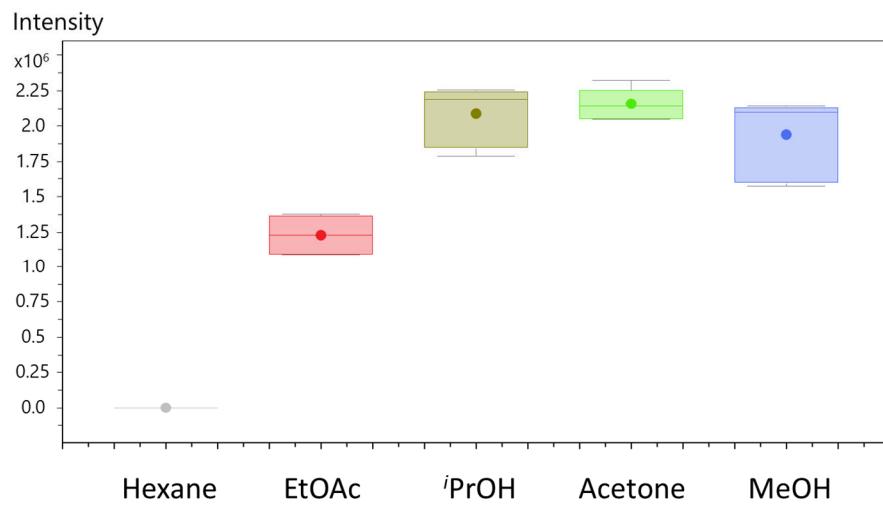
Figure S5 Box plot for myxovirescin A from *M. xanthus* DK1622 comparing different solvents for UAE.
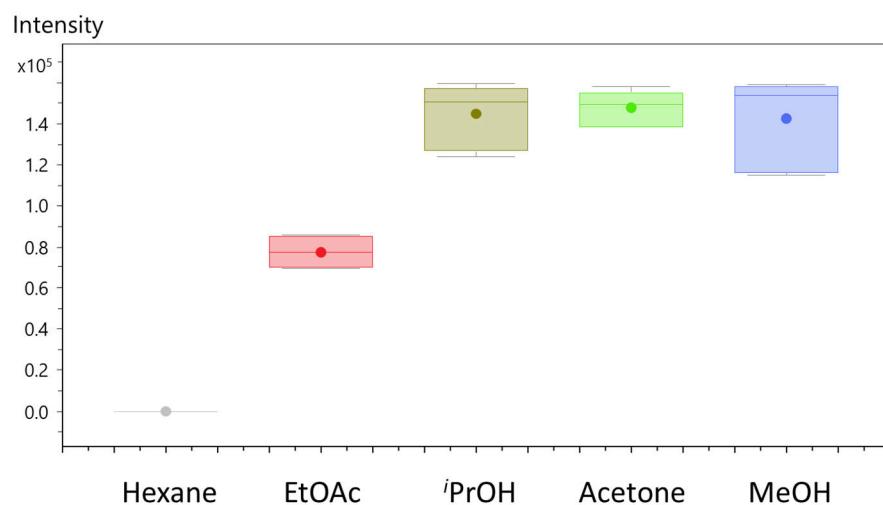
Figure S6 Box plot for myxovirescin B from *M. xanthus* DK1622 comparing different solvents for UAE.
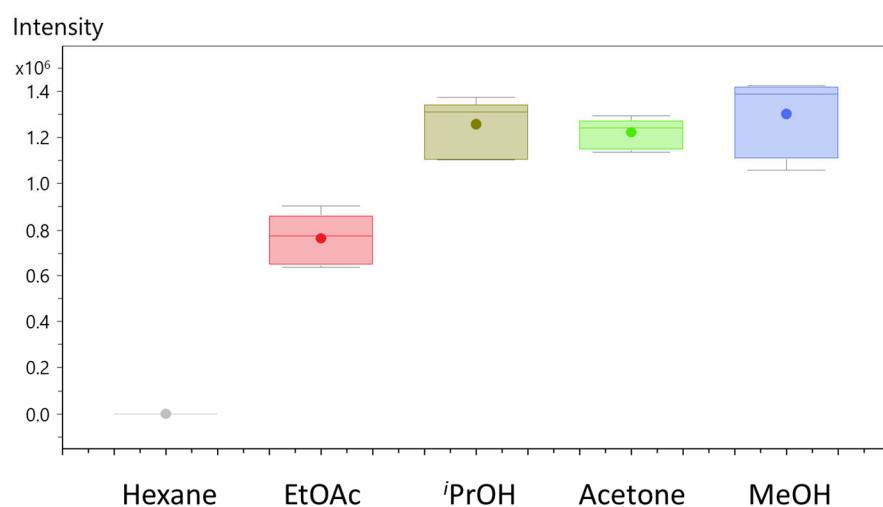
Figure S7 Box plot for myxovirescin C from *M. xanthus* DK1622 comparing different solvents for UAE.
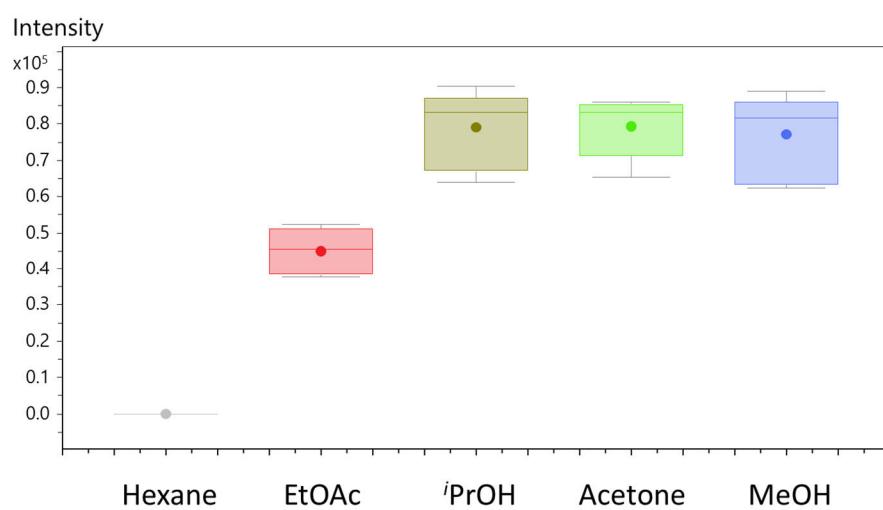
Figure S8 Box plot for myxovirescin G from *M. xanthus* DK1622 comparing different solvents for UAE.
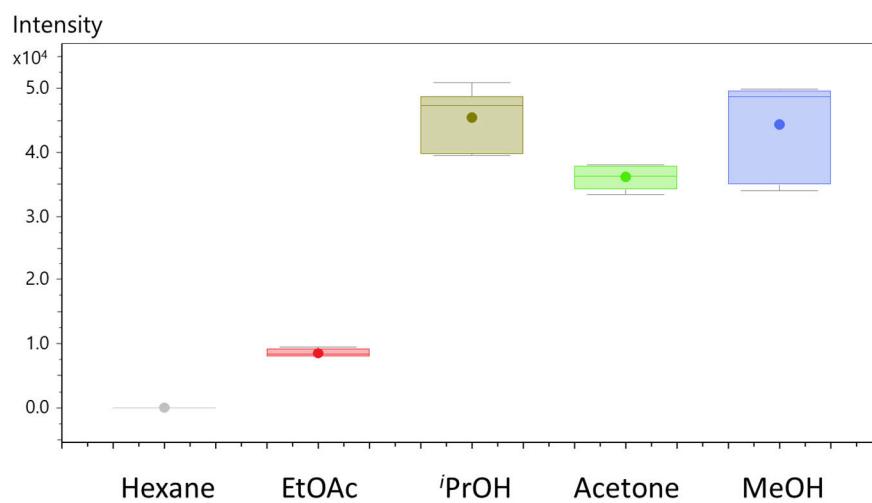
Figure S9 Box plot for myxovirescin variant KP641 from *M. xanthus* DK1622 comparing different solvents for UAE.
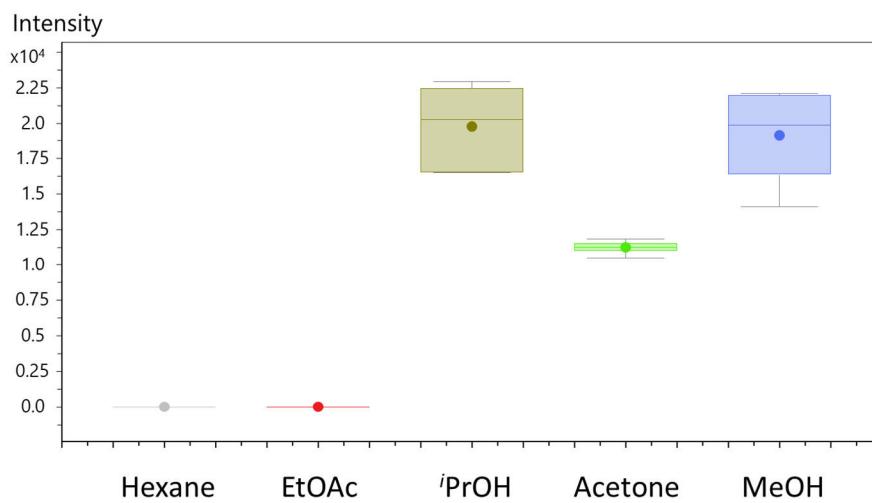
Figure S10 Box plot for chloroxanthic acid A from *M. xanthus* DK1622 comparing different solvents for UAE.
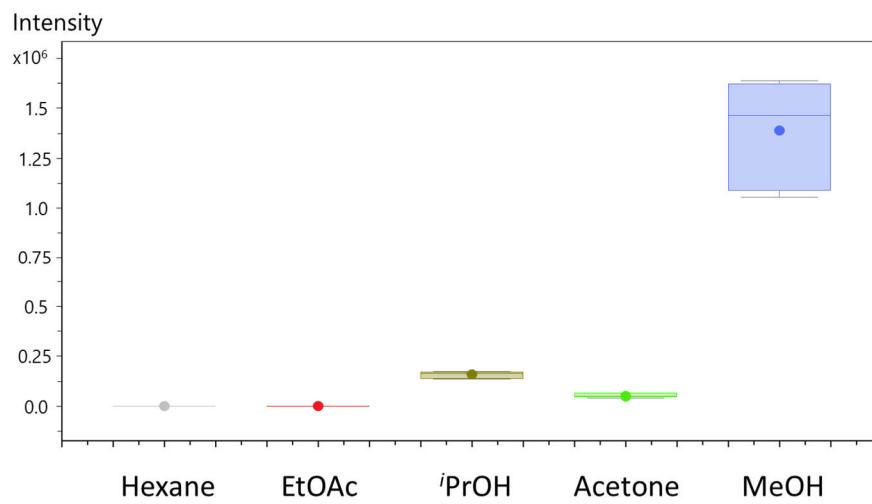
Figure S11 Box plot for DKxanthen 560 from *M. xanthus* DK1622 comparing different solvents for UAE.
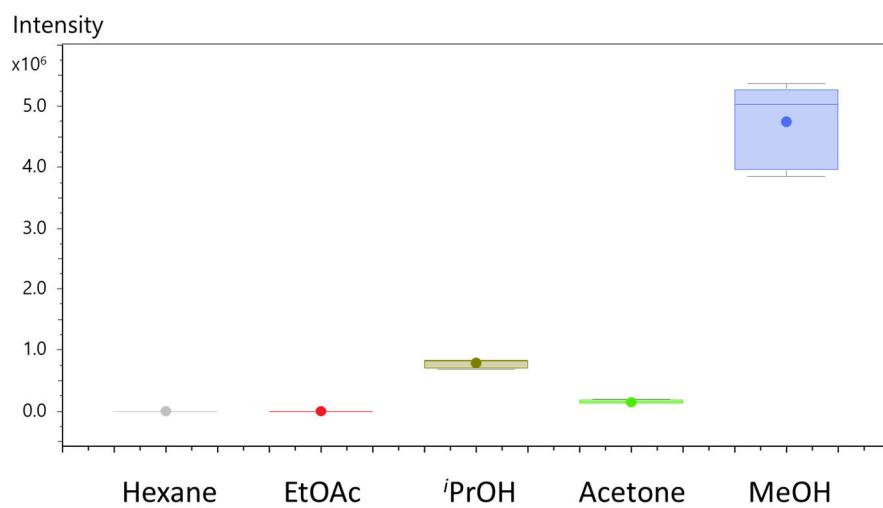
Figure S12 Box plot for DKxanthen 534 from *M. xanthus* DK1622 comparing different solvents for UAE.
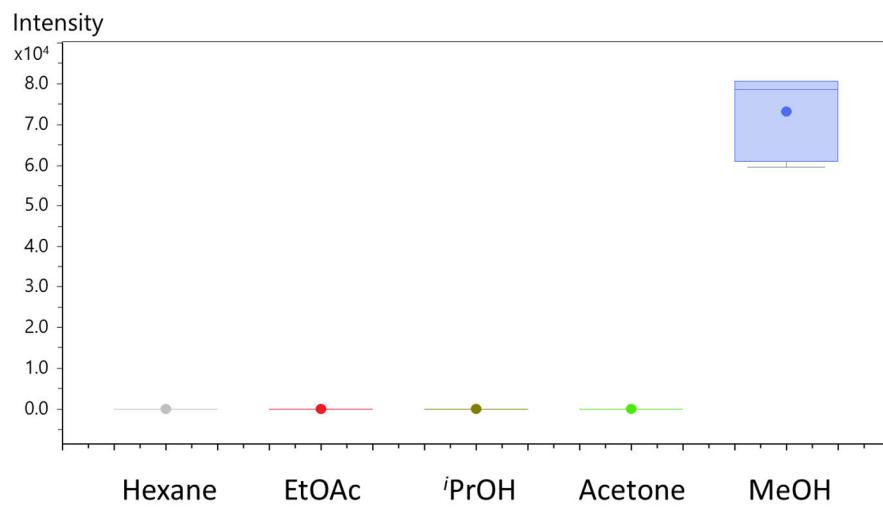
Figure S13 Box plot for DKxanthen 548 from *M. xanthus* DK1622 comparing different solvents for UAE.
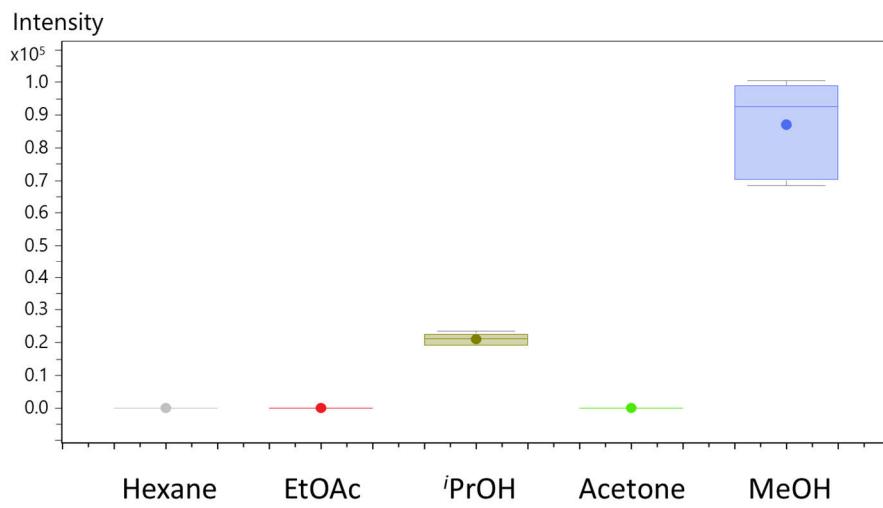
Figure S14 Box plot for DKxanthen 574 from *M. xanthus* DK1622 comparing different solvents for UAE.
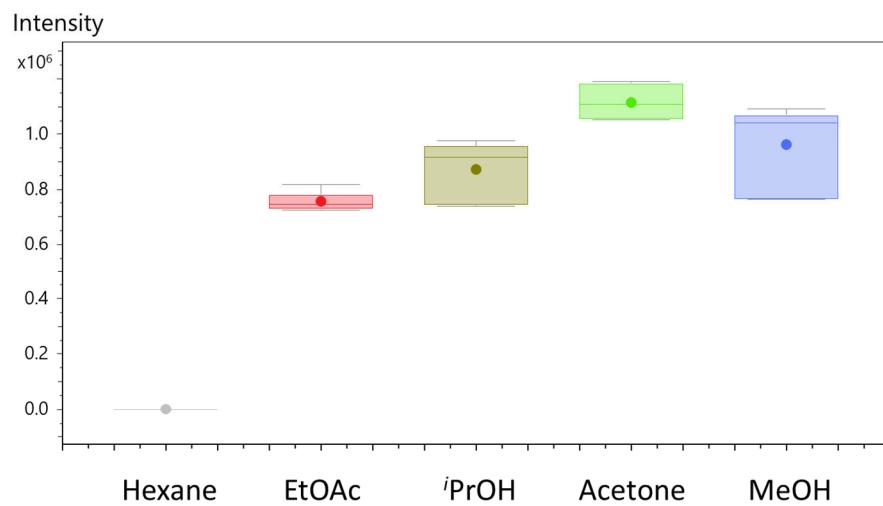
Figure S15 Box plot for myxalamid A from *M. xanthus* DK1622 comparing different solvents for UAE.
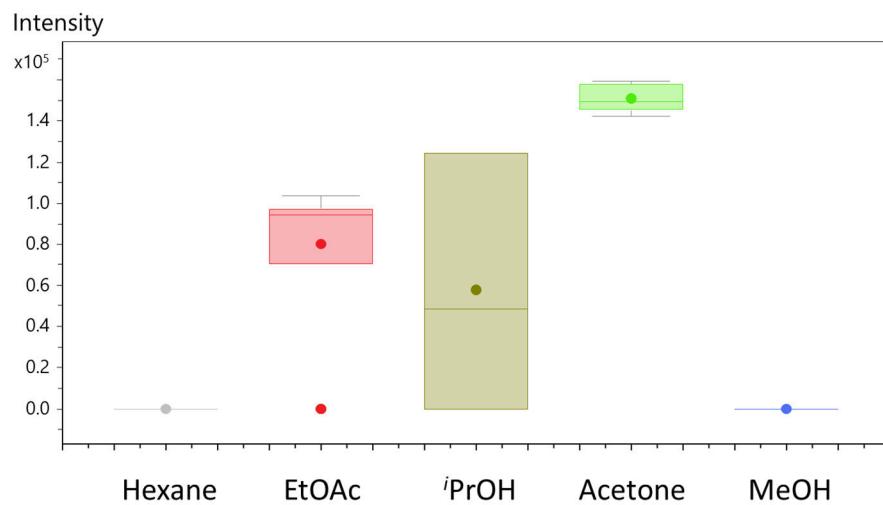
Figure S16 Box plot for myxalamid B from *M. xanthus* DK1622 comparing different solvents for UAE.
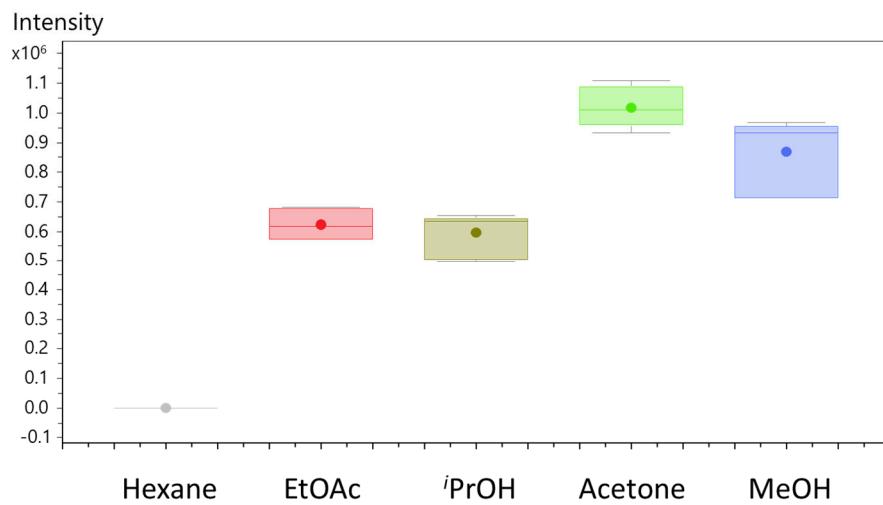
Figure S17 Box plot for myxalamid C from *M. xanthus* DK1622 comparing different solvents for UAE.
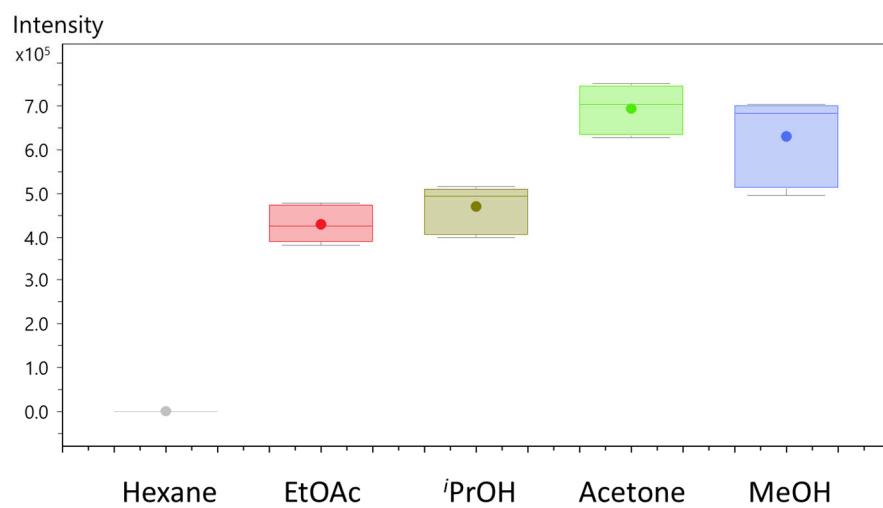
Figure S18 Box plot for myxalamid D from *M. xanthus* DK1622 comparing different solvents for UAE.
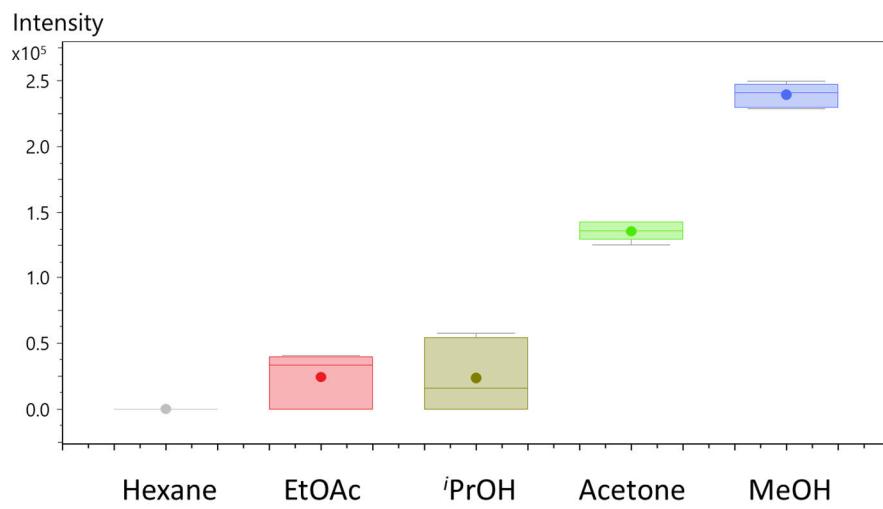
**Figure S19** Box plot for myxochromid A3 from *M. xanthus* DK1622 comparing different solvents for UAE.
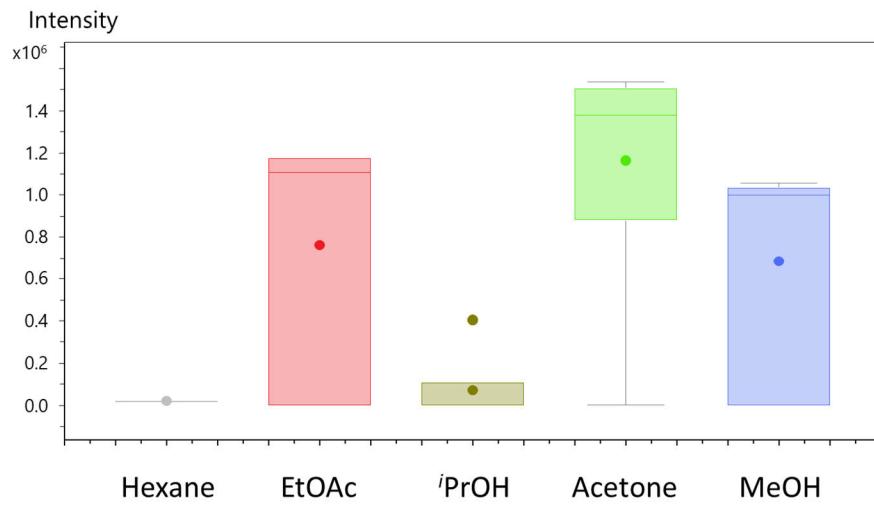

Figure S20 Box plot for indiacene A from MSr10575 comparing different solvents for UAE.
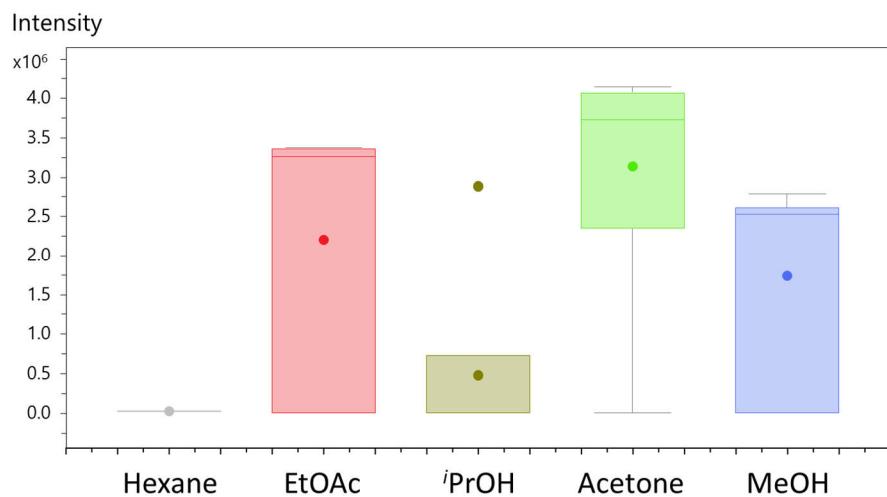
Figure S21 Box plot for indiacene B from MSr10575 comparing different solvents for UAE.
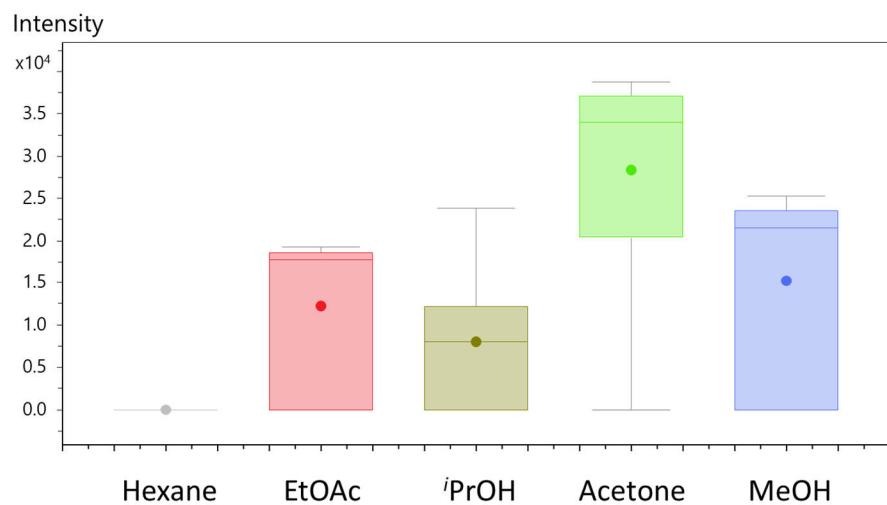
Figure S22 Box plot for terrestribisamid from MSr10575 comparing different solvents for UAE.

Our results support, that acetone and MeOH are indeed the best solvents for UA extraction in a targeted metabolomics workflow. In all cases they show equal if not superior extractability compared to the other three solvents. Myxochelin A, myxovirescin variant KP643, DKxanthen-520 and DKxanthen-526 could not be detected in the extracts were different solvents for UAE were evaluated, which may be caused by differences in production levels in between the two cultivation batches.

S3.3.3 Antimicrobial activities of the UA extracts generated with different solvents

We evaluated the antimicrobial activities of the UA extracts generated with different solvents against a test panel of eight microbial strains. As reference we used a standard extract generated by consecutive extraction with MeOH and acetone (Std 2. Batch) comparable to the UA extract used for comparison with SFE. The antimicrobial activities of the different extracts are separately shown in figure S23-25 for the three myxobacterial strains.
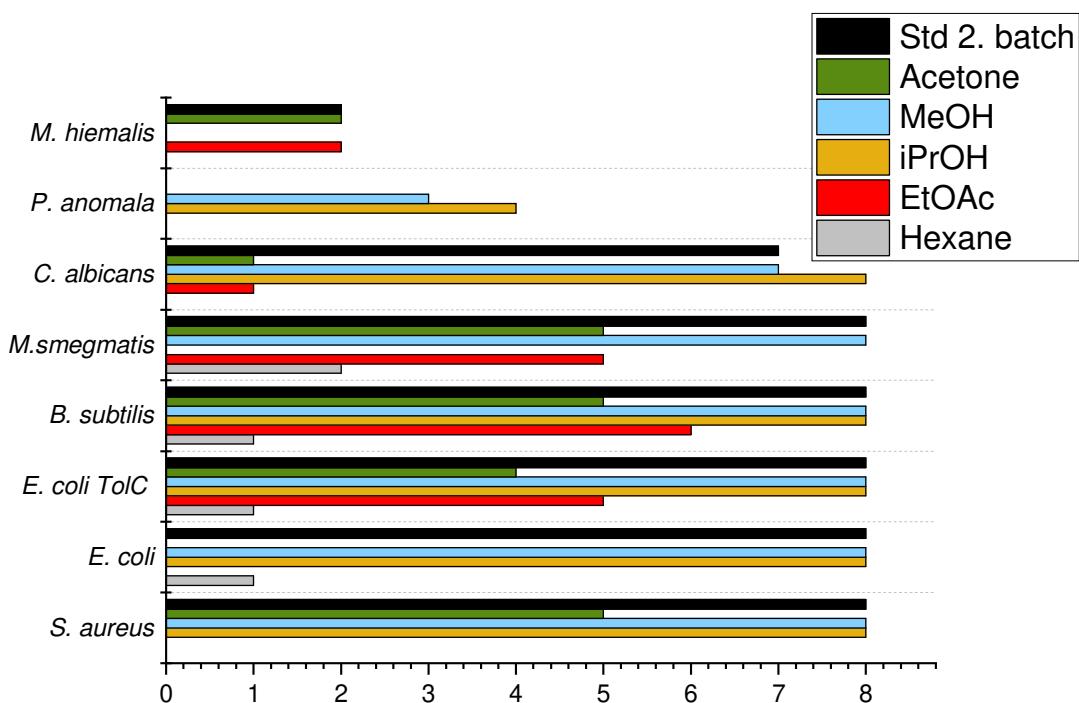
Figure S23 Antimicrobial activities of UA extracts from MSr10575 generated with different solvents.
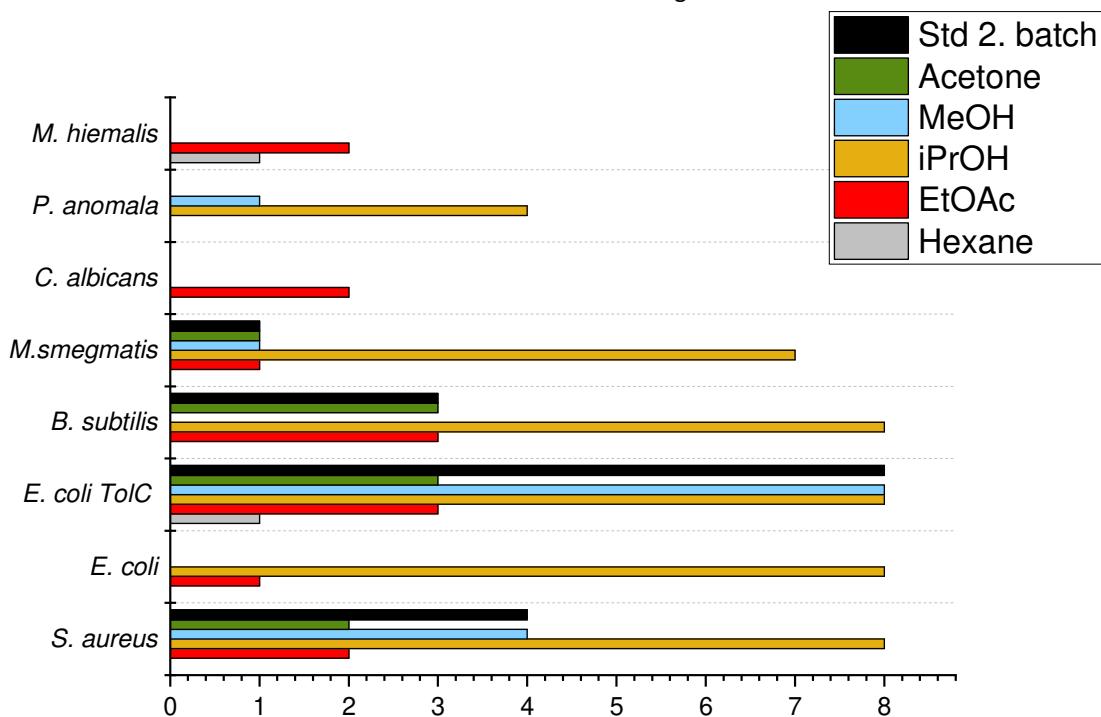
Figure S24 Antimicrobial activities of UA extracts from MSr11367 generated with different solvents.
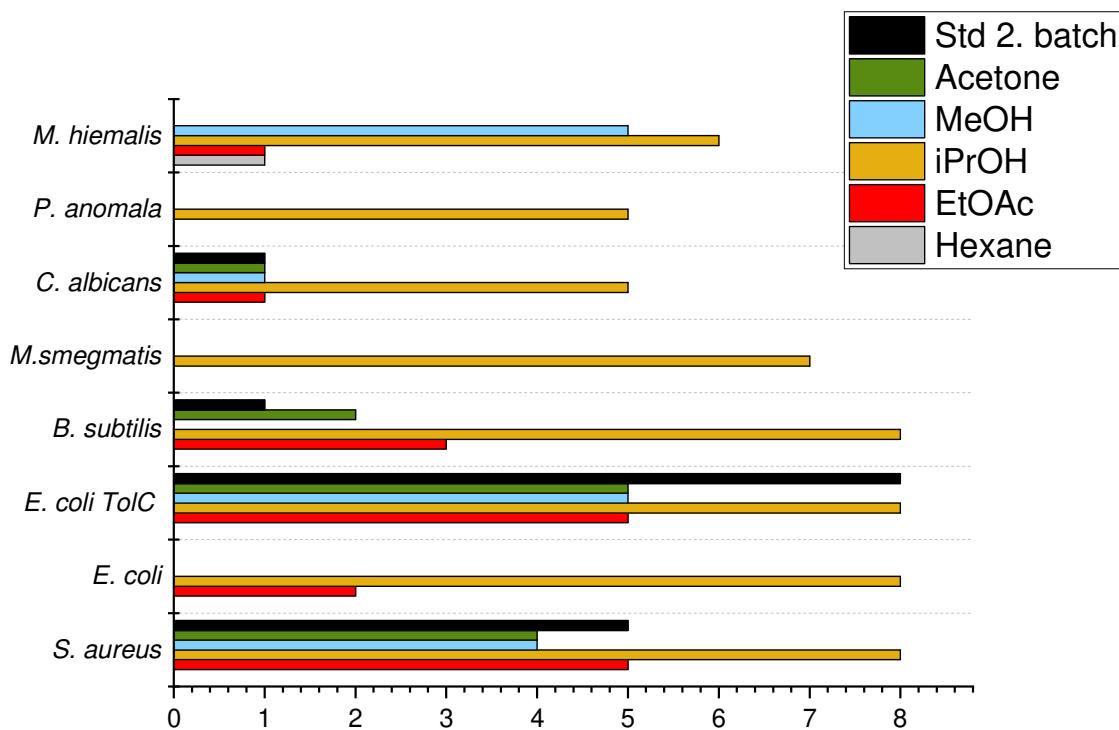
Figure S25 Antimicrobial activities of UA extracts from *M. xanthus* DK1622 generated with different solvents.

We observe similar effects on the antimicrobial activities when generating UA extracts with different solvents compared to SF extraction. We find higher activities of some of the extracts than the standard UA extracts, as well as some antimicrobial activities only detectable using specific solvents. Using a different solvent may therefore introduce some selectivity to the extraction process and generate extracts enriched in bioactive secondary metabolites, which highlights that limiting an extraction process to one solvent may hinder the detection of some antimicrobial activities or decrease the observed activity. We observe good antimicrobial activities of the ^{i}PrOH extracts for all three myxobacterial strains and even an activity of *M. xanthus* DK1622 extracts against *M. smegmatis* and *P. anomala*, not observable for any other solvent. This finding highlights ^{i}PrOH as interesting UA extraction solvent for bioactivity-guided isolation and structure elucidation of novel natural products, which is in contrast to the SF extraction. Here the ^{i}PrOH showed the lowest activities besides the pure CO_2 extracts. However, this finding for SFE may stem from the consecutive extraction protocol chosen, where biologically active secondary metabolites may already be extracted during the 20% EtOAc step.

S3.3.4 Consecutive extraction

We were interested in comparison of UA extractions using single solvents with consecutive extractions comparable to the protocol performed with SFE. We therefore extracted an aliquot of *M. xanthus* DK1622 pellet first with hexane, then EtOAc, *i*PrOH, MeOH and finally with acetone. We also performed UA extraction with a mixture of an equal volume of each solvents as reference. The results for the six different secondary metabolite families detected in *M. xanthus* DK1622 extracts are shown in figure S26-S31.
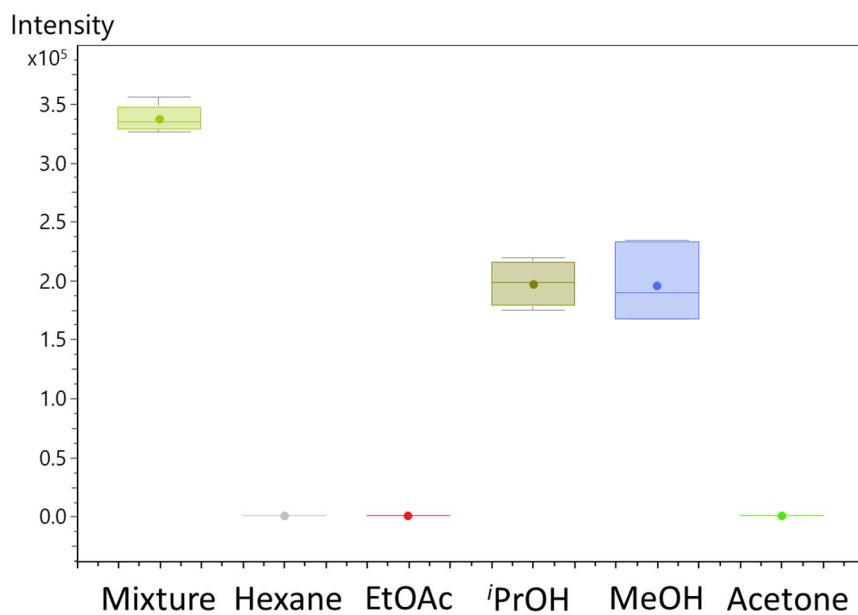
Figure S26 Box plot for cittilin A from *M. xanthus* DK1622 comparing consecutive extraction for UAE.
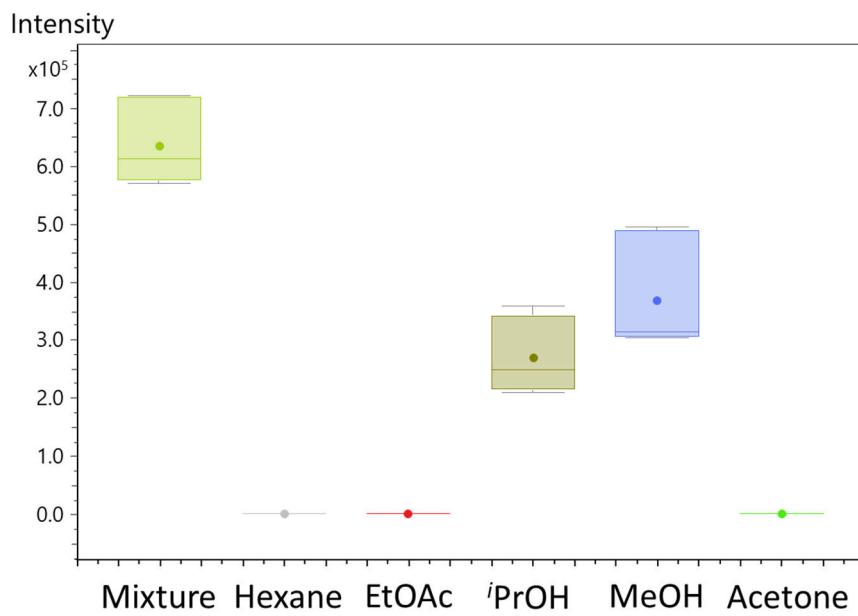
Figure S27 Box plot for DKxanthen 560 from *M. xanthus* DK1622 comparing consecutive extraction for UAE.
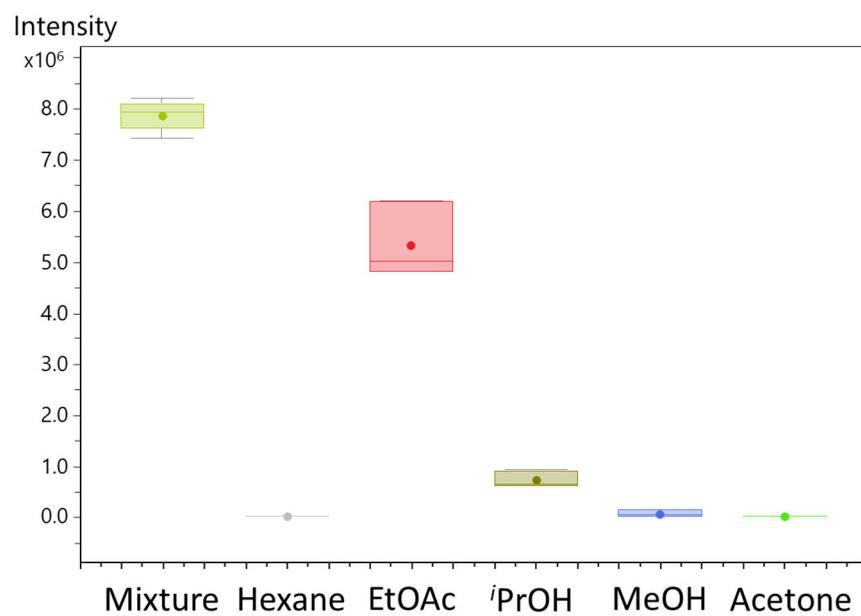
Figure S28 Box plot for myxalamid A from *M. xanthus* DK1622 comparing consecutive extraction for UAE.
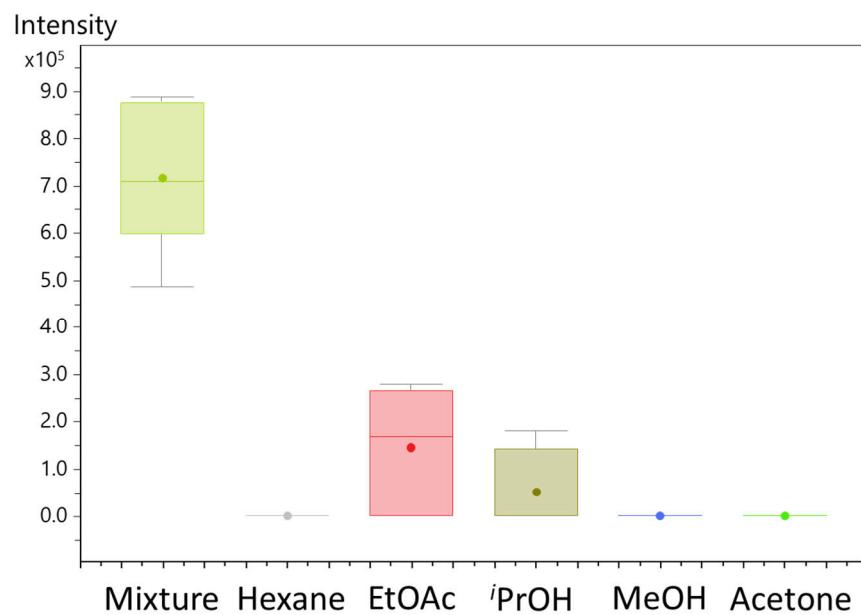
Figure S29 Box plot for myxochromid A3 from *M. xanthus* DK1622 comparing consecutive extraction for UAE.

Intensity
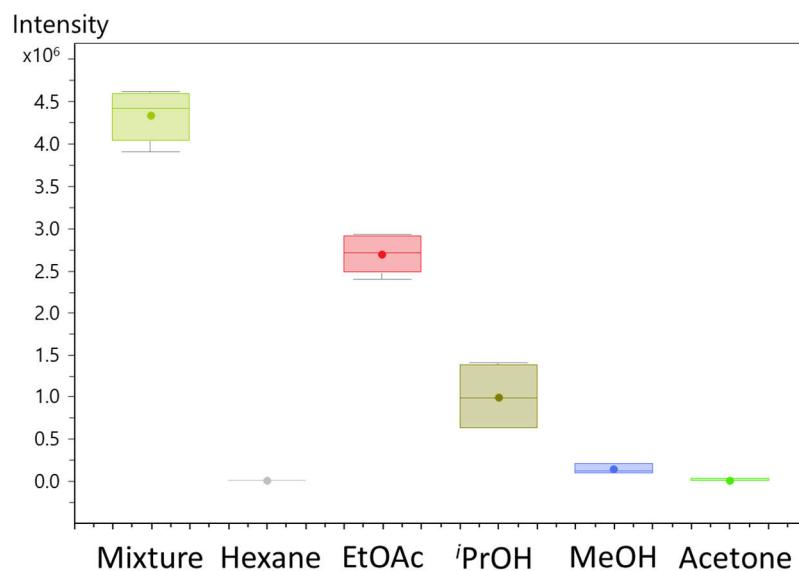
Figure S30 Box plot for myxovirescin A from *M. xanthus* DK1622 comparing consecutive extraction for UAE.

Intensity
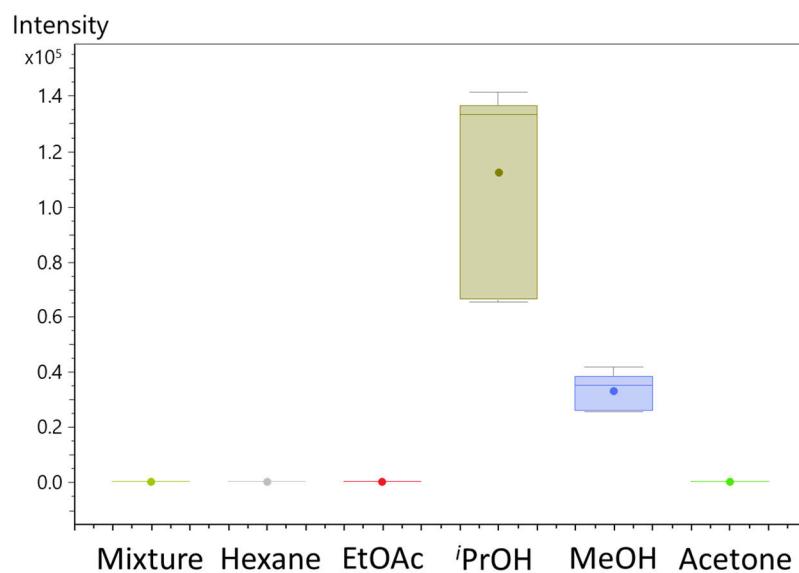
Figure S31 Box plot for riboflavin from *M. xanthus* DK1622 comparing consecutive extraction for UAE.

The secondary metabolite distribution when performing consecutive UA extraction with the five different solvents shows highly similar profiles compared to consecutive SF extraction. Secondary metabolites with a logP higher than 3 already show extraction with EtOAc, whereas secondary metabolites with a logP lower than 3 are mainly found in the *i*PrOH and MeOH extraction step. In contrary to SFE however, none of the metabolites was extractable under the most non-polar condition using hexane. All metabolites were already exhaustively extracted after the MeOH extraction step, as we were not able to detect any of them in the acetone extracts but have shown beforehand that different myxovirescin and myxalamid derivatives, as well as myxochromid A3 are extractable with acetone. Consecutive UA extraction is therefore also feasible for conducting a primary fractionation step during the extraction process, leading to enrichment of different secondary metabolites in the fractions based on their logP values.

S3.4 Targeted metabolomics investigation of known myxobacterial secondary metabolites.

S3.4.1 Box plots of secondary metabolites extractable with SFE from *M. xanthus* DK1622

This chapter including figure S32- S47 completes the box plots shown in figure 1 in the main text for the remaining derivatives of each family of secondary metabolites produced by *M. xanthus* DK1622.
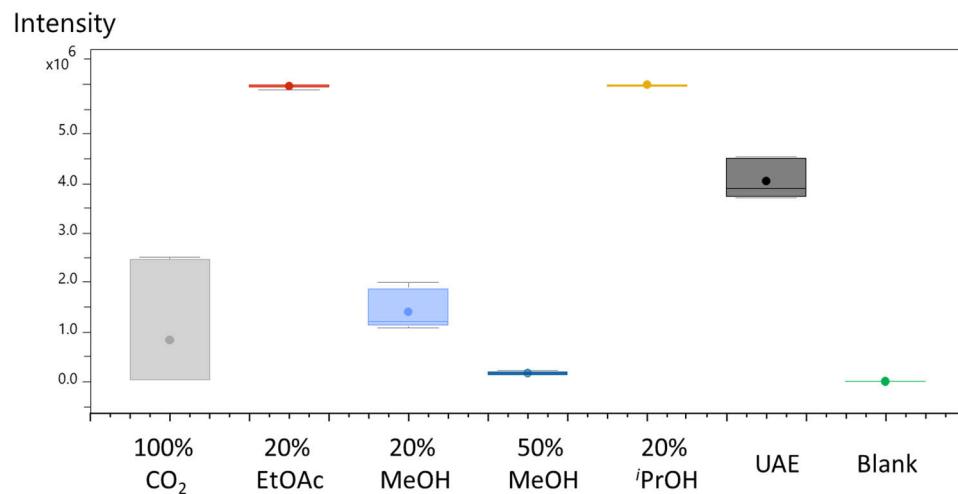
Figure S32 Box plot for riboflavin from *M. xanthus* DK1622.
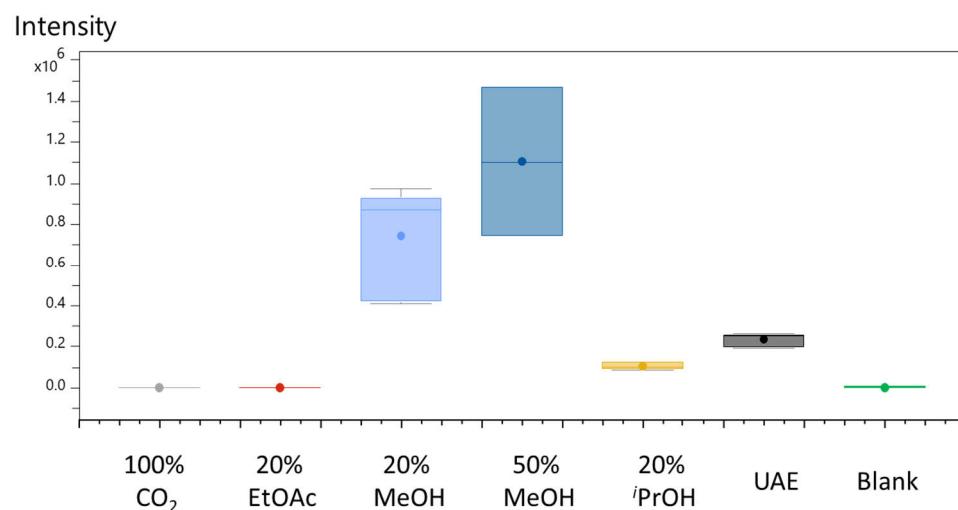
Figure S33 Box plot for chloroxanthic acid A from *M. xanthus* DK1622 (SFE vs UAE).

Intensity
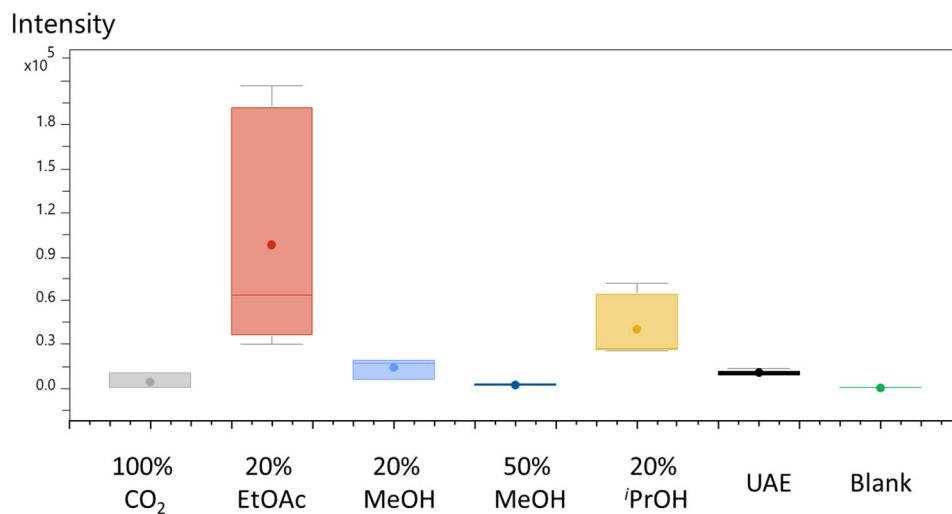
Figure S34 Box plot for myxovirecin variant KP643 from *M. xanthus* DK1622.

Intensity
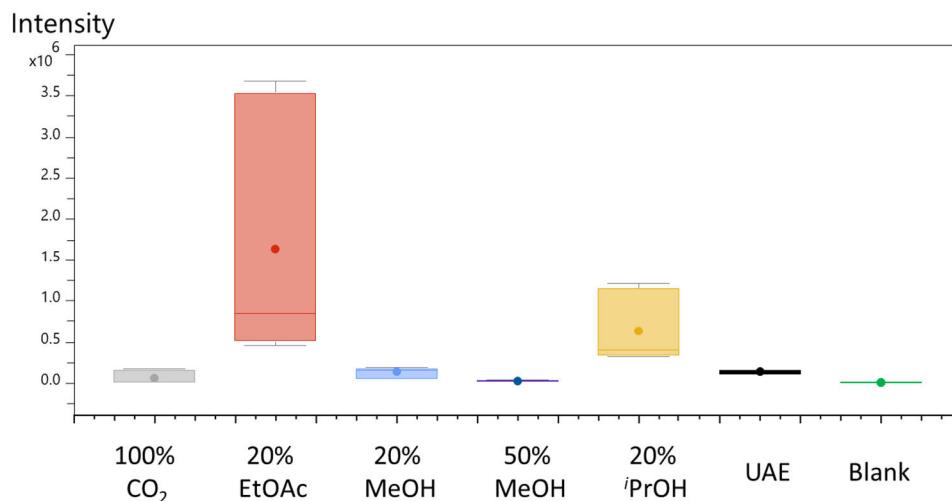
Figure S35 Box plot for myxovirescin variant KP641 from *M. xanthus* DK1622.

Intensity
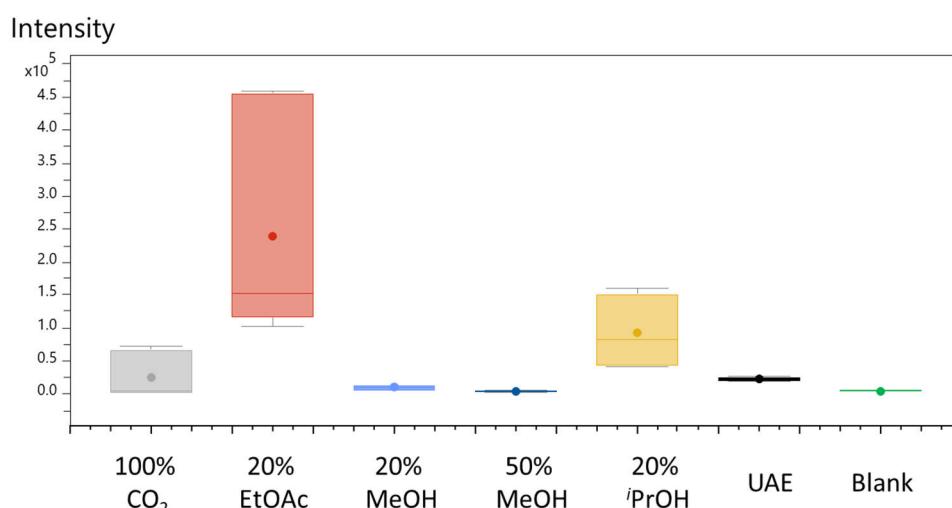
Figure S36 Box plot for myxovirescin B from *M. xanthus* DK1622.

Intensity
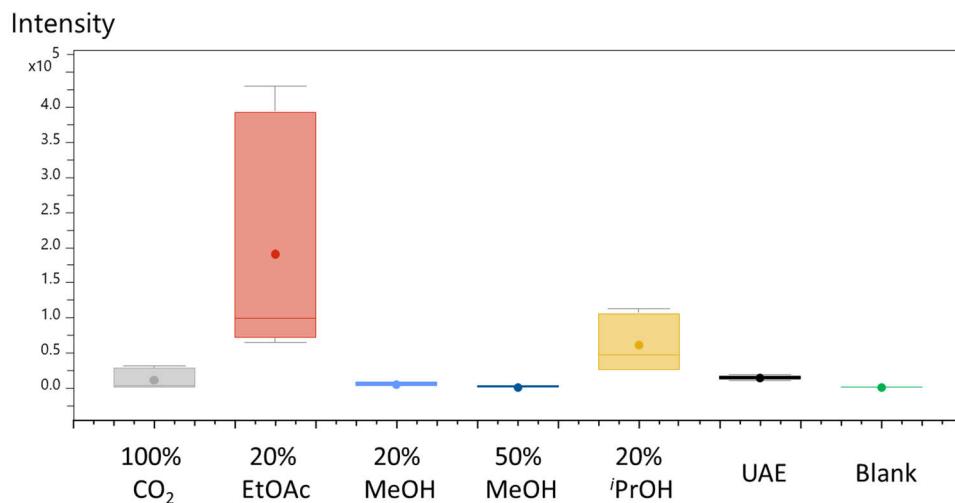
Figure S37 Box plot for myxovirescin C from *M. xanthus* DK1622.

Intensity
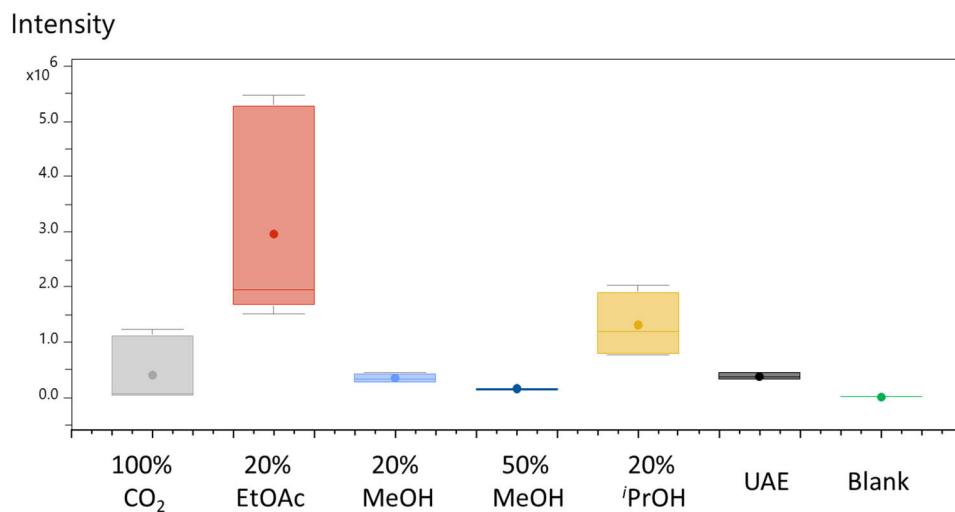
Figure S38 Box plot for myxovirescin G from *M. xanthus* DK1622 (SFE vs UAE).

Intensity
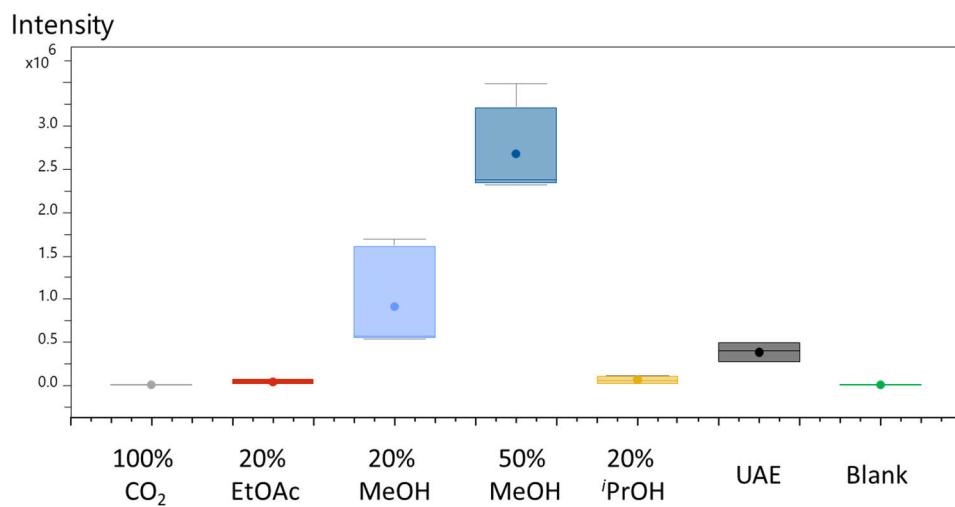
Figure S39. Box plot for Dkxanthen-543 from *M. xanthus* DK1622.
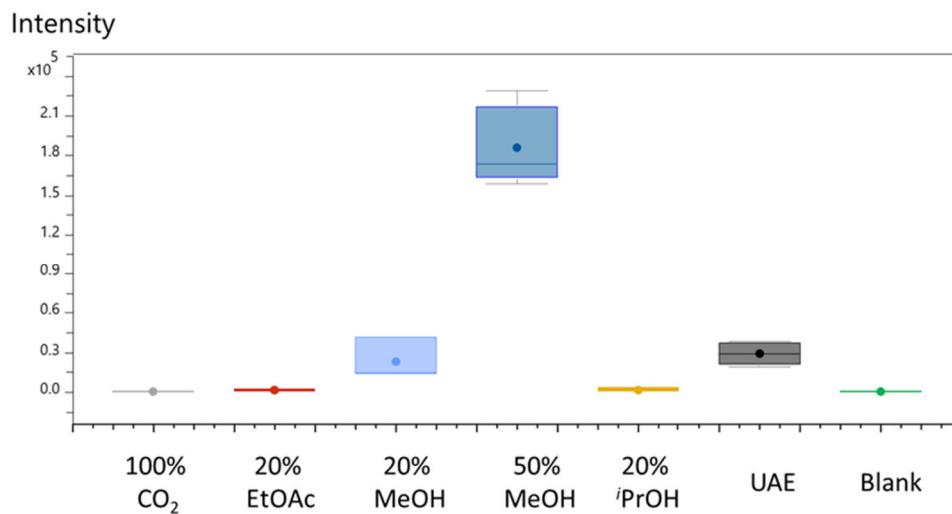
Figure S40 Box plot for DKxanthen 520 from *M. xanthus* DK1622.
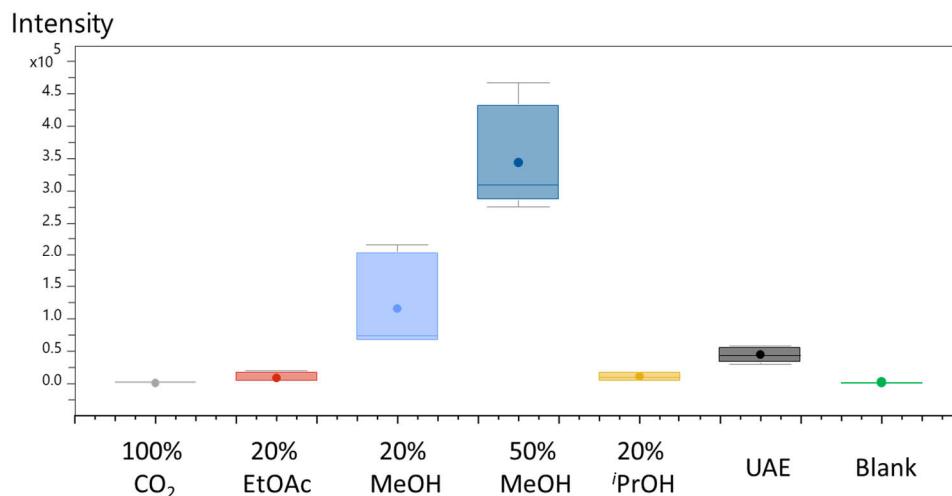
Figure S41 Box plot for DKxanthen 548 from *M. xanthus* DK1622.
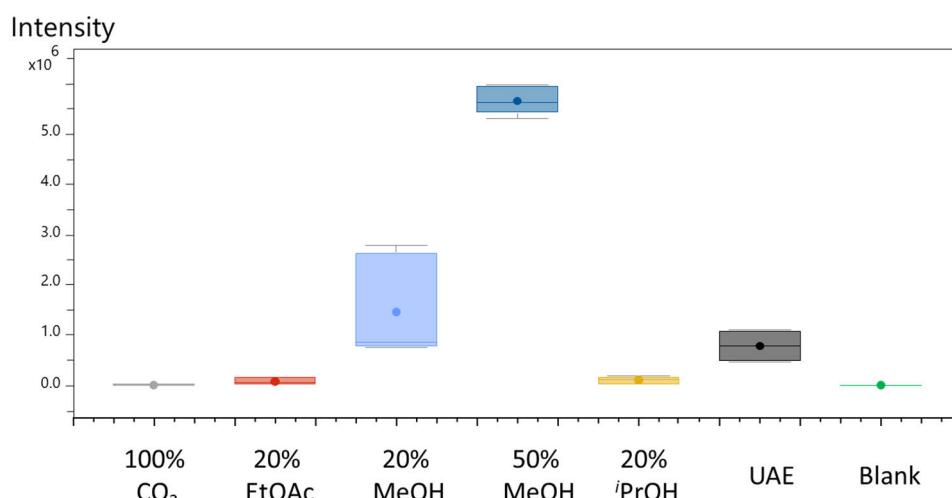
Figure S42 Box plot for DKxanthen 560 from *M. xanthus* DK1622.
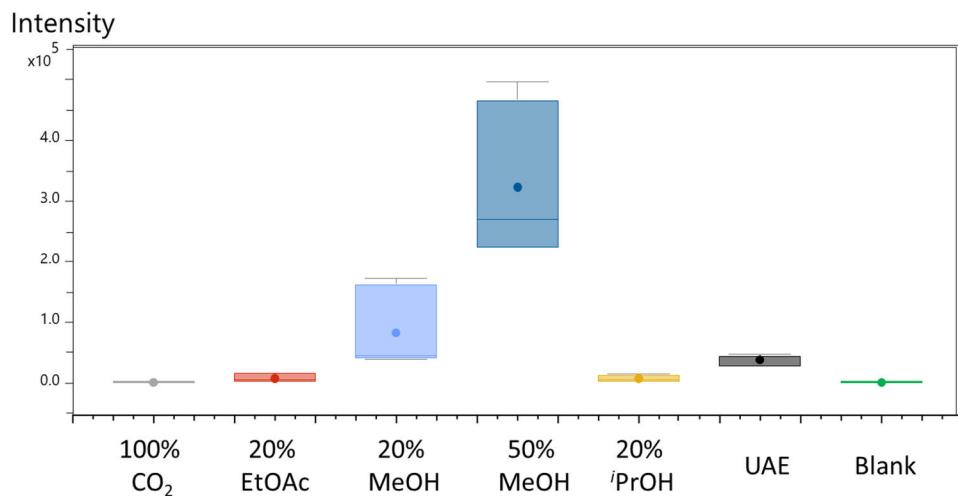
Figure S43 Box plot for DKxanthen 574 from *M. xanthus* DK1622.
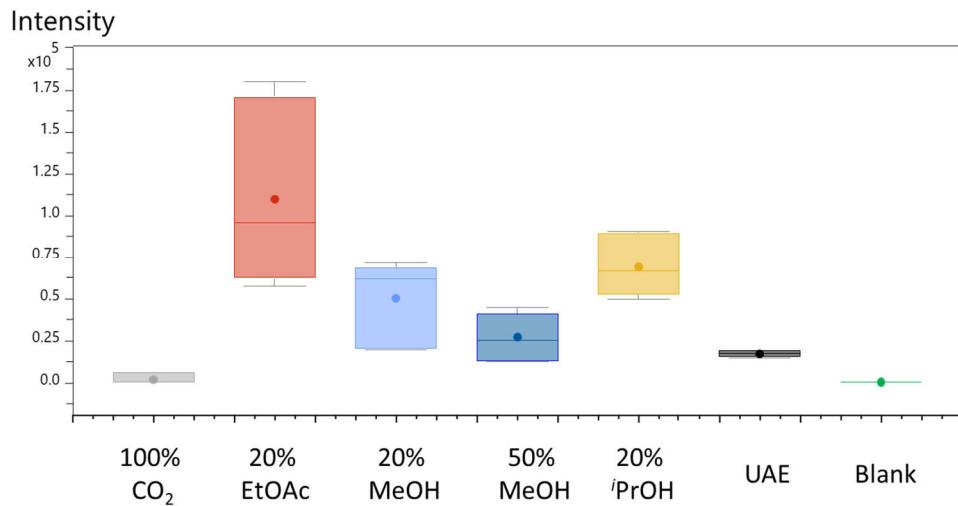
Figure S44 Box plot for DKxanthen 526 from *M. xanthus* DK1622.
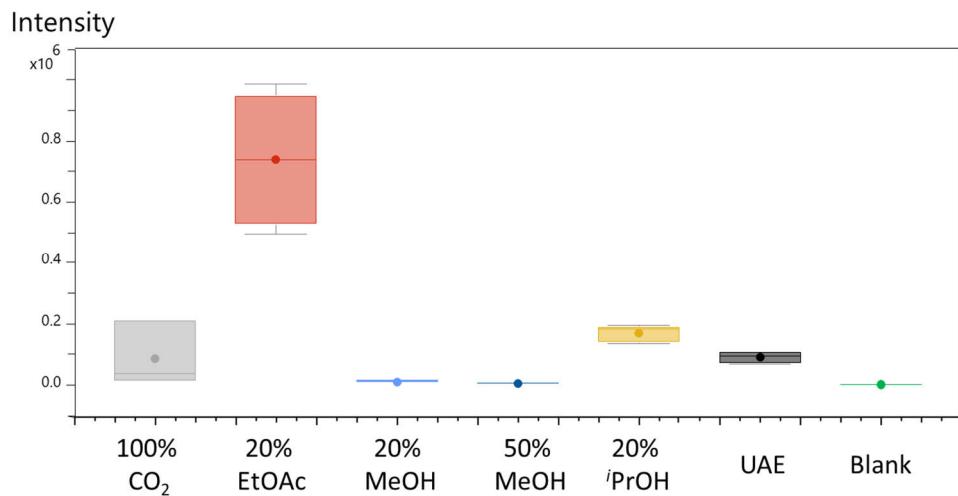
Figure S45 Box plot for myxalamid B from *M. xanthus* DK1622.

Intensity
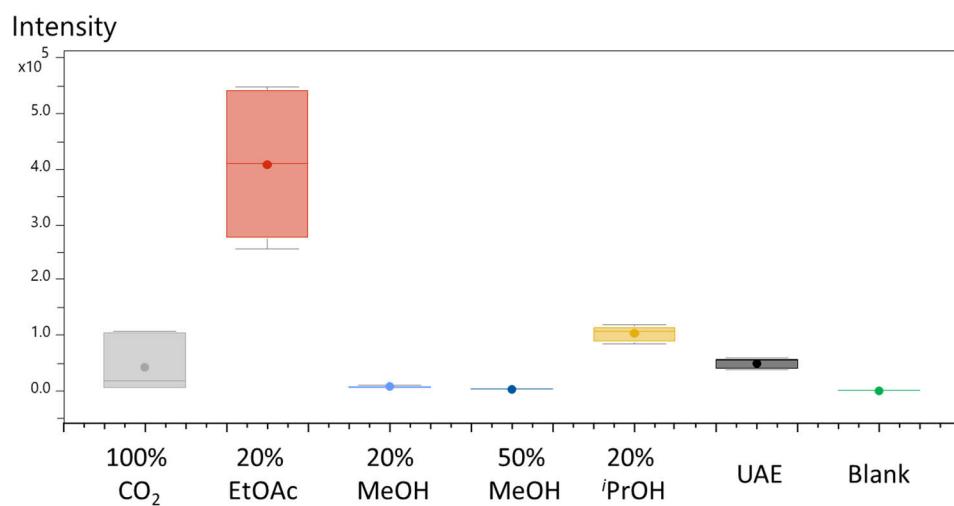
Figure S46 Box plot for myxalamid C from *M. xanthus* DK1622.

Intensity
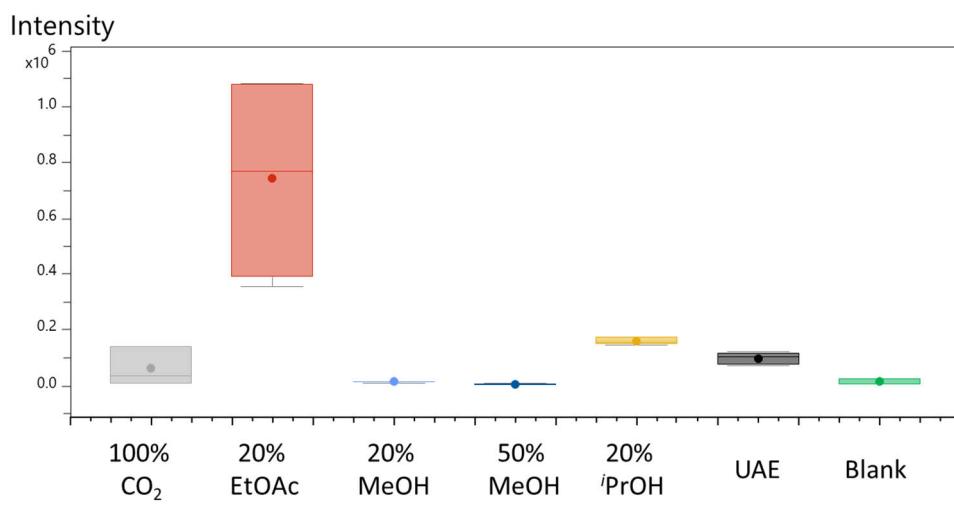
Figure S47 Box plot for myxalamid 430 from *M. xanthus* DK1622.

S3.4.2 Blox plots of secondary metabolites extractable with SFE from MSr10575

This chapter including figure S48 completes the box plots shown in figure 1 in the main text for the remaining derivatives of each family of secondary metabolites produced by MSr10575.
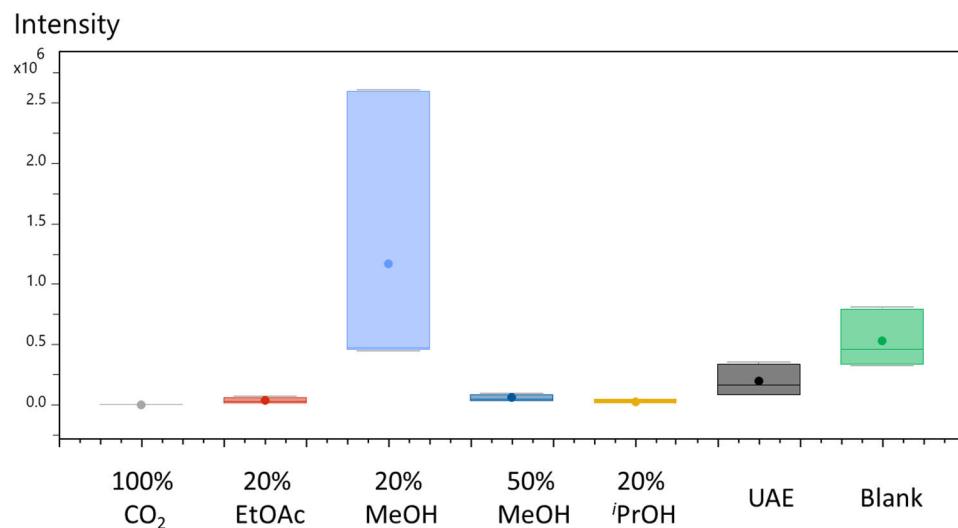
Figure S48 Box plot for indiacene B from MSr10575.

S3.5 Non-targeted metabolomics investigation

S3.5.1 Comparison of features detected in the UAE and SFE
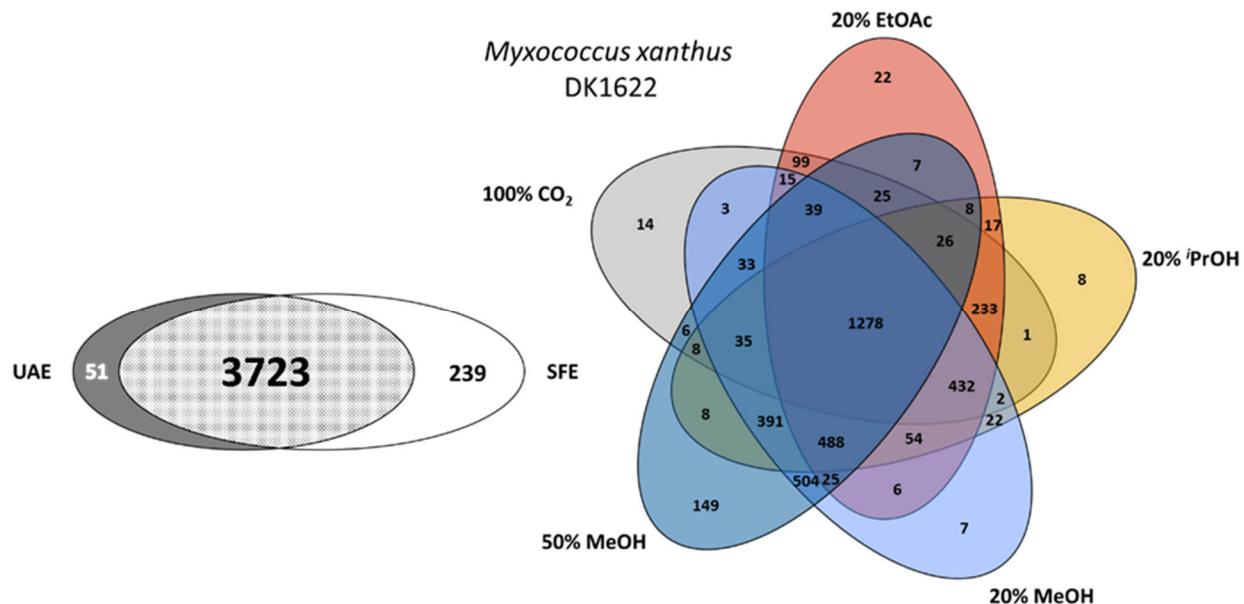
Figure S49 Euler diagrams for *M. xanthus* DK1622 Left-hand side: Features unique for UAE or SFE and features extractable with both methods. Right-hand side: Features found in the SF extracts from MSr10575 sorted by co-solvent used during the extraction process.
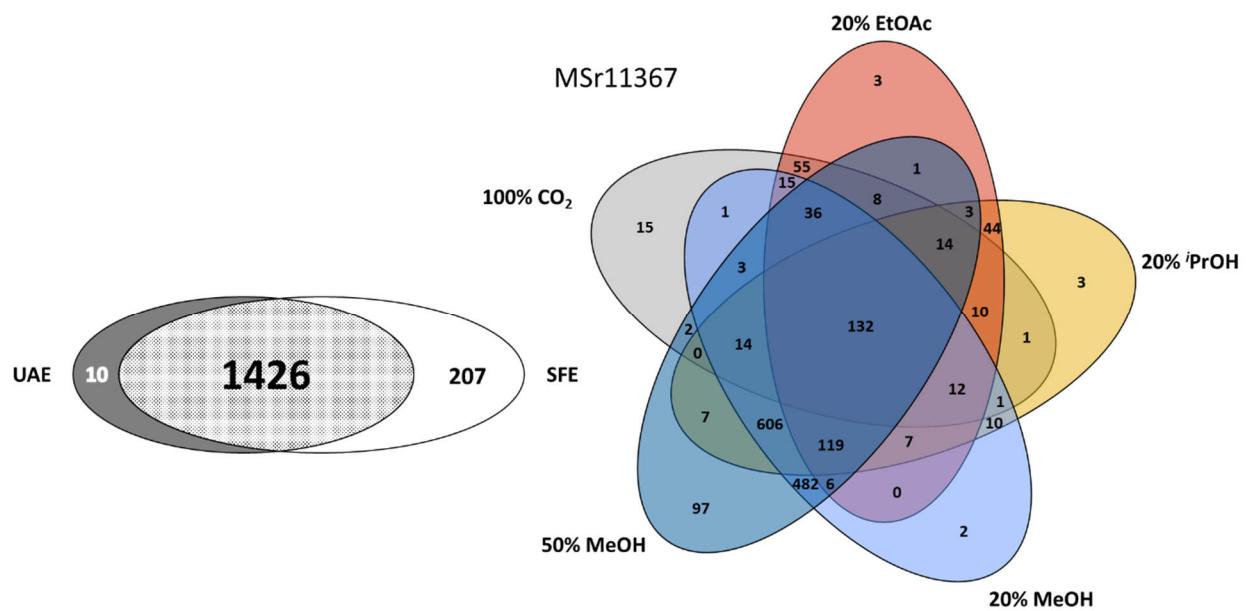
Figure S50 Euler diagrams for MSr11367. Left-hand side: Features unique for UAE or SFE and features extractable with both methods. Right-hand side: Features found in the SF extracts from MSr10575 sorted by co-solvent used during the extraction process.

S3.5.2 Chlorinated metabolites from *M. xanthus* DK1622

Table S5. List of chlorinated molecular features detected with elevated intensities in the SF extracts.

	m/z meas.	Molecular Formula (M)	Intensity UAE/SFE [%]	RT
1	295.0848	C14H15ClN2O3	5	10.39
2	313.0954	C14H17ClN2O4	11	7.44
3	329.0459	C14H14Cl2N2O3	62	7.14
4	337.1682	C18H25ClN2O2	52	7.32
5	265.0741	C13H13ClN2O2	47	10.32
6	363.1474	C19H23ClN2O3	48	12.28
7	329.0459	C14H14Cl2N2O3	3	11.78
8	335.1526	C18H23ClN2O2	3	7.48
9	313.0954	C14H17ClN2O4	19	8.54
10	313.0954	C14H17ClN2O4	0	4.31
11	363.1475	C19H23ClN2O3	2	7.49
12	355.0616	C16H16Cl2N2O3	61	12.41
13	227.0584	C10H11ClN2O2	0	4.23
14	335.1162	C17H19ClN2O3	2	7.26
15	281.1055	C14H17ClN2O2	0	5.25
16	335.1162	C17H19ClN2O3	39	12.95
17	269.1055	C13H17ClN2O2	0	4.99
18	349.1318	C18H21ClN2O3	50	7.81
19	361.1319	C19H21ClN2O3	3	12.77
20	349.1319	C18H21ClN2O3	2	19.40
21	255.0534	C11H11ClN2O3	0	4.43
22	309.1368	C16H21ClN2O2	3	10.64
23	355.0616	C16H16Cl2N2O3	3	12.41
24	347.1164	C18H19ClN2O3	2	12.01
25	241.0741	C11H13ClN2O2	0	4.23
26	373.1319	C20H21ClN2O3	3	12.96
27	269.1054	C13H17ClN2O2	0	5.48
28	307.1211	C16H19ClN2O2	5	13.03
29	340.6873	C35H54ClN3O8	73	13.10
30	311.1525	C16H23ClN2O2	8	11.41

S3.6 Isolation of myxobacterial natural products described in this study

S3.6.1 Isolation of the Indiacenes A and B from *Sandaracinus* sp. MSr10575

Large scale cultivation and extraction

The *Sandaracinus* strain MSr10575 was fermented in 50 ml 2-SWT medium as a seed culture flasks on an Orbitron shaker at 160 rpm and 30°C. The translucid culture medium becomes orange and opaque after 7 to 11 days of fermentation. This pre-culture was used to inoculate 6 x 2L 2-SWT medium supplemented with 2% XAD-16 resin suspension in sterilized water in 6 x 5L baffled shake flasks on an

Orbiton shaker at 160 rpm and 30°C. Fermentation was complete after 16 days. Cells and XAD-16 resin were harvested by centrifugation on a Beckmann Avanti J-26 XP with the JLA 8.1 rotor at 6000 rcf. Combined resin and cells were freeze dried and subsequently extracted using 2x 500 ml of a 2:1 mixture of methanol and chloroform. The combined extracts were concentrated on a rotary evaporator and partitioned between methanol and hexane. The indiacenes remained in the methanol phase, which was dried with a rotary evaporator and the residue partitioned between water and chloroform. The indiacenes retained in the chloroform phase, which was concentrated and stored in an air-tight glass vial at -20°C until further processing.

Centrifugal partition chromatography (CPC) pre-purification of the compounds

Pre-purification and fractionation of indiacene A and B was achieved by CPC. We used a Gilson CPC 100 device (Gilson purification S.A.S.) connected to a Varian ProStar Solvent delivery module and a Varian ProStar 2 Channel UV detector. Fraction collection is done with a Foxy Jr. auto sampler (Isco). Used biphasic solvent system is the buffered ARIZONA solvent system consisting of a 1:1:1:1 mixture of 10 mM Tris x HCl buffer pH 8.0 (Sigma), methanol (analytical grade, Fluka), EtOAc (analytical grade, Fluka) and hexane (analytical grade, Fluka). After equilibration of the system and loading of the sample, the aqueous phase is used as a stationary phase for 40 min at 2 ml/min and a fraction size of 4 ml. The indiacenes elute simultaneously in fractions 35-55, that are pooled for fine purification.

Purification of Indiacene A and B by HPLC

Purification of indiacene A and B was carried out using a Dionex Ultimate 3000 SDLC low pressure gradient system on a Waters Acquity CSH C18 250x10mm 5µm column with the eluents H_2O + 1% FA as A and ACN + 1% FA as B, a flow rate of 5 ml/min and a column thermostated at 30°C. Indiacene A and B are detected by UV absorption at 316 nm and purification is done by time dependent fraction collection. Separation is started with a plateau at 60% A for 2 minutes followed by a ramp to 24% A during 24 minutes and a ramp to 0% A during 1 minute. The A content is kept at 0% A for 2 minutes. The A content is ramped back to starting conditions during 30 seconds and the column is re-equilibrated for 2 minutes. After evaporation, the indiacenes A and B are obtained as orange solids.

S3.6.2 Isolation of Chloroxanthic acid A

Large scale cultivation and extraction

M. xanthus DK1622 was fermented in 50 ml CTT medium medium as a seed culture flasks on an Orbiton shaker at 160 rpm and 30°C. The translucid culture medium becomes dark yellow and opaque after 5 to 7 days of fermentation. This pre-culture was used to inoculate 3 x 300 mL SHG-P medium supplemented with sterile-filtered 7.5 µg/ml Fe-EDTA solution and 2% XAD-16 resin suspension in sterilized water in 3 x 1L shake flasks on an Orbiton shaker at 160 rpm and 30°C. Fermentation was complete after 10 days. Cells and XAD-16 resin were harvested by centrifugation on a Beckmann Avanti J-26 XP with the JLA 8.1 rotor at 6000 rcf. Combined resin and cells were freeze dried and subsequently extracted using SFE with 20% EtOAc as co-solvent. SFE parameters were consistent with the ones used for in the analytical setup (see "material and methods, extraction procedure" in the main text). The combined extracts were concentrated on a rotary evaporator and partitioned between methanol and hexane. Chloroxanthic acid A remained in the methanol phase, which was dried with a rotary evaporator and the residue partitioned between water and chloroform. Chloroxanthic acid retained in the chloroform phase, which was concentrated and stored in an air-tight glass vial at -20°C until further processing. Further isolation steps are described in the main text of the manuscript.

S3.7 Structure elucidation of the natural products described in this study

S3.7.1 Indiacene A and B

Crystallization conditions for indiacene A and B

The two indiacene secondary metabolites are dissolved in a minimal amount of methanol, transferred in a glass vial und kept in the fridge at 4°C without a lid. After two weeks, we see yellowish-orange needle shaped crystals appearing in the residual solvent. These crystals diffracted in a small molecule X-ray crystallograph, which allowed the acquisition of X-ray crystal structures. The data were collected at low temperature on a BrukerAXS X8Apex CCD diffractometer operating with graphite-monochromatized Mo Kα radiation. Frames of 0.5° oscillation were exposed; deriving reflections in the θ range of 2 to 27° with a completeness of ~99%. Structure solving and full least-squares refinement with anisotropic thermal parameters of all non-hydrogen atoms were performed using SHELX.[13] All relevant data concerning the crystal structures can be found in this section. Crystallographic data for the structure have been deposited with the Cambridge Crystallographic Data Centre, CCDC, 12 Union Road, Cambridge CB21EZ, UK. Copies of the data can be obtained free of charge on quoting the repository number CCDC 2035083 for indiacene A and 2035084 for indiacene B under www.ccdc.cam.ac.uk/data_request/cif.

Crystal structure parameters for indiacene A
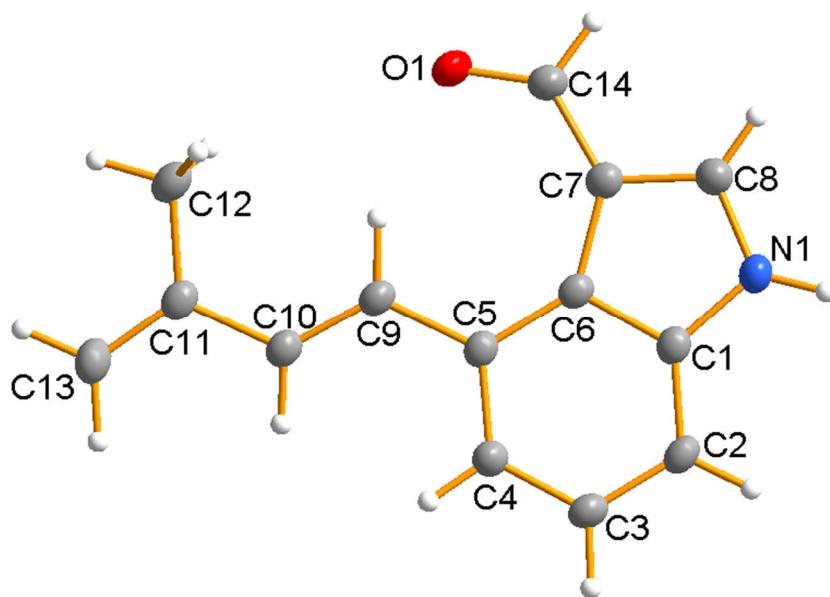
Figure S51 Electron density map calculated from the indiacene A diffraction pattern.

Empirical formula	$C_{14}H_{13}NO$
Formula weight	211.25 Da
Temperature	133(2) K
Wavelength	1.54178 Å
Crystal system	Orthorhombic
Space group	$P\ 2_12_12_1$

Unit cell dimensions	a = 5.8884(3) Å α = 90°.
	b = 13.0563(6) Å β = 90°.
	c = 14.2977(6) Å γ = 90°.
Volume	1099.22(9) Å^3
Z	4
Density (calculated)	1.277 Mg/m^3
Absorption coefficient	0.635 mm^{-1}
F(000)	448
Crystal size	0.131 x 0.093 x 0.015 mm^3
Theta range for data collection	4.586 to 80.478°.
Index ranges	-7 ≤ h ≤ 7, -16 ≤ k ≤ 16, -18 ≤ l ≤ 18
Reflections collected	21720
Independent reflections	2388 [R (int) = 0.0408]
Completeness to theta = 67.679°	99.4 %
Absorption correction	Semi-empirical from equivalents
Max. and min. transmission	0.7543 and 0.6472
Refinement method	Full-matrix least squares on F2
Data / restraints / parameters	2388 / 0 / 197
Goodness-of-fit on F2	1.072
Final R indices [I>2σ (I)]	R1 = 0.0278, wR2 = 0.0674
R indices (all data)	R1 = 0.0284, wR2 = 0.0679
Absolute structure parameter	0.08(8)
Largest diff. peak and hole	0.157 and -0.146 e.Å^{-3}

Table S6 Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters (Å^2 x10^3) for indiacene A. U (eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)
O(1)	6198(2)	5389(1)	2194(1)	29(1)
N(1)	6737(2)	4902(1)	5414(1)	25(1)
C(1)	4766(2)	5470(1)	5304(1)	22(1)
C(2)	3252(2)	5777(1)	5998(1)	25(1)
C(3)	1425(2)	6368(1)	5722(1)	26(1)
C(4)	1137(2)	6642(1)	4780(1)	24(1)
C(5)	2604(2)	6312(1)	4070(1)	21(1)
C(6)	4477(2)	5698(1)	4340(1)	20(1)
C(7)	6383(2)	5209(1)	3869(1)	23(1)
C(8)	7688(2)	4762(1)	4569(1)	26(1)
C(9)	2237(2)	6611(1)	3091(1)	23(1)
C(10)	369(2)	7090(1)	2772(1)	24(1)
C(11)	-22(2)	7416(1)	1807(1)	24(1)
C(12)	1745(3)	7184(1)	1081(1)	31(1)
C(13)	-1939(3)	7919(1)	1595(1)	32(1)
C(14)	7121(2)	5086(1)	2915(1)	25(1)

Table S7 Bond lengths [Å] and angles [°] for indiacene A.

bond lengths [Å]

O(1)-C(14)	1.2303(18)
N(1)-C(8)	1.3434(19)
N(1)-C(1)	1.3864(17)
N(1)-H(5)	0.93(3)
C(1)-C(2)	1.3922(19)
C(1)-C(6)	1.4202(18)
C(2)-C(3)	1.382(2)
C(2)-H(1)	0.98(2)
C(3)-C(4)	1.4040(19)
C(3)-H(2)	0.98(2)
C(4)-C(5)	1.4010(19)
C(4)-H(3)	0.96(2)
C(5)-C(6)	1.4168(18)
C(5)-C(9)	1.4694(18)
C(6)-C(7)	1.4566(19)
C(7)-C(8)	1.390(2)
C(7)-C(14)	1.4410(18)
C(8)-H(4)	0.93(2)
C(9)-C(10)	1.3446(19)
C(9)-H(6)	1.01(2)
C(10)-C(11)	1.4621(18)

C(10)-H(7)	0.95(2)
C(11)-C(13)	1.341(2)
C(11)-C(12)	1.501(2)
C(12)-H(8)	1.00(2)
C(12)-H(9)	0.97(2)
C(12)-H(10)	0.98(2)
C(13)-H(11)	0.98(2)
C(13)-H(12)	1.02(2)
C(14)-H(13)	0.962(19)

bond angles in [°]

C(8)-N(1)-C(1)	108.69(11)
C(8)-N(1)-H(5)	127.6(15)
C(1)-N(1)-H(5)	123.2(15)
N(1)-C(1)-C(2)	127.55(12)
N(1)-C(1)-C(6)	108.79(11)
C(2)-C(1)-C(6)	123.66(12)
C(3)-C(2)-C(1)	117.11(12)
C(3)-C(2)-H(1)	122.8(12)
C(1)-C(2)-H(1)	120.1(12)
C(2)-C(3)-C(4)	120.71(13)
C(2)-C(3)-H(2)	119.3(11)
C(4)-C(3)-H(2)	120.0(11)
C(5)-C(4)-C(3)	122.81(13)
C(5)-C(4)-H(3)	120.2(12)
C(3)-C(4)-H(3)	117.0(12)
C(4)-C(5)-C(6)	117.16(12)
C(4)-C(5)-C(9)	121.18(12)
C(6)-C(5)-C(9)	121.64(12)
C(5)-C(6)-C(1)	118.46(12)
C(5)-C(6)-C(7)	136.21(12)
C(1)-C(6)-C(7)	105.31(11)
C(8)-C(7)-C(14)	117.92(13)
C(8)-C(7)-C(6)	106.12(12)
C(14)-C(7)-C(6)	135.96(12)
N(1)-C(8)-C(7)	111.05(13)
N(1)-C(8)-H(4)	121.9(13)
C(7)-C(8)-H(4)	127.1(13)
C(10)-C(9)-C(5)	124.51(12)
C(10)-C(9)-H(6)	119.3(12)
C(5)-C(9)-H(6)	116.2(12)
C(9)-C(10)-C(11)	125.74(13)
C(9)-C(10)-H(7)	121.3(12)
C(11)-C(10)-H(7)	113.0(12)

C(13)-C(11)-C(10)	119.29(13)
C(13)-C(11)-C(12)	121.72(13)
C(10)-C(11)-C(12)	119.00(13)
C(11)-C(12)-H(8)	109.8(12)
C(11)-C(12)-H(9)	113.3(13)
H(8)-C(12)-H(9)	105.8(18)
C(11)-C(12)-H(10)	112.7(13)
H(8)-C(12)-H(10)	106.6(17)
H(9)-C(12)-H(10)	108.1(18)
C(11)-C(13)-H(11)	120.5(13)
C(11)-C(13)-H(12)	122.1(13)
H(11)-C(13)-H(12)	117.5(18)
O(1)-C(14)-C(7)	128.65(13)
O(1)-C(14)-H(13)	117.8(12)
C(7)-C(14)-H(13)	113.6(12)

Table S8 Anisotropic displacement parameters (Å^2x 10^3) for indiacene A.

	U^{11}	U^{22}	U^{33}	U^{23}	U^{13}	U^{12}
O(1)	32(1)	37(1)	19(1)	-4(1)	1(1)	2(1)
N(1)	28(1)	27(1)	21(1)	2(1)	-2(1)	2(1)
C(1)	25(1)	21(1)	20(1)	1(1)	-2(1)	-1(1)
C(2)	30(1)	28(1)	16(1)	0(1)	0(1)	-2(1)
C(3)	27(1)	30(1)	20(1)	-3(1)	2(1)	-1(1)
C(4)	24(1)	28(1)	21(1)	-2(1)	-1(1)	1(1)
C(5)	23(1)	22(1)	18(1)	-1(1)	-2(1)	-2(1)
C(6)	24(1)	20(1)	18(1)	-2(1)	0(1)	-3(1)
C(7)	26(1)	21(1)	22(1)	-1(1)	1(1)	0(1)
C(8)	27(1)	26(1)	25(1)	0(1)	1(1)	3(1)
C(9)	26(1)	24(1)	18(1)	-2(1)	0(1)	-2(1)
C(10)	27(1)	26(1)	19(1)	-1(1)	-1(1)	-1(1)
C(11)	29(1)	23(1)	19(1)	-2(1)	-3(1)	-5(1)
C(12)	38(1)	33(1)	21(1)	1(1)	2(1)	1(1)
C(13)	31(1)	40(1)	24(1)	3(1)	-7(1)	-2(1)
C(14)	26(1)	24(1)	23(1)	-4(1)	2(1)	1(1)

Table S9 Hydrogen coordinates (x 10^4) and isotropic displacement parameters (Å^2x 10^3) for indiacene A.

	x	y	z	U(eq)
H(1)	3520(30)	5582(16)	6649(15)	37(5)
H(2)	340(40)	6609(15)	6195(14)	31(5)
H(3)	-100(30)	7093(14)	4636(13)	29(4)
H(4)	9060(40)	4422(15)	4497(14)	31(5)
H(5)	7350(40)	4740(17)	5994(18)	44(6)
H(6)	3500(30)	6433(15)	2638(15)	34(5)
H(7)	-870(40)	7247(14)	3178(14)	33(5)
H(8)	1180(40)	7387(16)	451(15)	38(5)
H(9)	2110(40)	6461(17)	1035(15)	39(5)
H(10)	3160(40)	7562(17)	1183(15)	38(5)
H(11)	-2250(40)	8127(16)	946(16)	38(5)
H(12)	-3140(40)	8090(18)	2089(16)	44(6)
H(13)	8510(30)	4705(14)	2854(13)	29(4)

Crystal structure parameters for indiacene B.
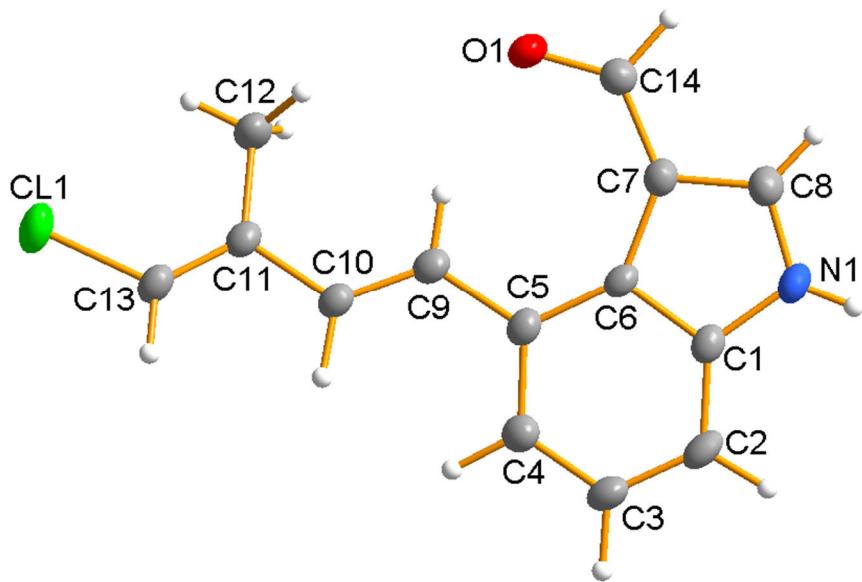
Figure S52 Electron density map calculated from the indiacene B diffraction pattern.

Empirical formula	$C_{14}H_{12}NOCl$
Formula weight	245.70 Da
Temperature	152(2) K
Wavelength	0.71073 Å
Crystal system	Orthorhombic
Space group	P $2_12_12_1$
Unit cell dimensions	a = 6.3460(4) Å α = 90°.

	b = 12.9149(9) Å β = 90°.
	c = 14.3042(10) Å γ = 90°.
Volume	1172.34(14) $Å^3$
Z	4
Density (calculated)	1.392 Mg/m^3
Absorption coefficient	0.307 mm^{-1}
F(000)	512
Crystal size	0.402 x 0.125 x 0.072 mm^3
Theta range for data collection	2.125 to 27.972°.
Index ranges	-8 ≤ h ≤ 6, -17 ≤ k ≤ 17, -18 ≤ l ≤ 18
Reflections collected	18425
Independent reflections	2822 [R (int) = 0.0561]
Completeness to theta = 25.242°	100.0 %
Absorption correction	Semi-empirical from equivalents
Max. and min. transmission	0.7456 and 0.7062
Refinement method	Full-matrix least squares on F2
Data / restraints / parameters	2822 / 0 / 202
Goodness-of-fit on F2	1.050
Final R indices [I>2σ (I)]	R1 = 0.0370, wR2 = 0.0723
R indices (all data)	R1 = 0.0575, wR2 = 0.0802
Absolute structure parameter	-0.04(4)
Largest diff. peak and hole	0.202 and -0.204 e.$Å^{-3}$

Table S10 Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters ($Å^2$x 10^3) for indiacene B. U (eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)
Cl(1)	11853(1)	1972(1)	9149(1)	38(1)
N(1)	3467(4)	5208(2)	4042(2)	29(1)
O(1)	3466(3)	4533(2)	7232(1)	31(1)
C(1)	5331(4)	4676(2)	4181(2)	24(1)
C(2)	6890(5)	4457(2)	3531(2)	29(1)
C(3)	8591(5)	3891(2)	3837(2)	30(1)
C(4)	8723(5)	3551(2)	4760(2)	29(1)
C(5)	7209(5)	3795(2)	5431(2)	24(1)
C(6)	5458(5)	4387(2)	5135(2)	21(1)
C(7)	3569(4)	4814(2)	5570(2)	24(1)
C(8)	2447(5)	5286(2)	4855(2)	29(1)
C(9)	7399(5)	3444(2)	6405(2)	25(1)
C(10)	9116(5)	3022(2)	6782(2)	24(1)
C(11)	9310(5)	2687(2)	7758(2)	24(1)
C(12)	7397(5)	2682(3)	8369(2)	34(1)
C(13)	11212(5)	2404(2)	8038(2)	28(1)

Table S11 Bond lengths [Å] and angles [°] for indiacene B.

bond lengths [Å]

Cl(1)-C(13)	1.733(3)
N(1)-C(8)	1.335(4)
N(1)-C(1)	1.382(4)
N(1)-H(5)	0.84(3)
O(1)-C(14)	1.237(3)
C(1)-C(2)	1.387(4)
C(1)-C(6)	1.418(4)
C(2)-C(3)	1.375(4)
C(2)-H(1)	0.99(3)
C(3)-C(4)	1.394(4)
C(3)-H(2)	1.01(3)
C(4)-C(5)	1.394(4)
C(4)-H(3)	1.03(3)
C(5)-C(6)	1.414(4)
C(5)-C(9)	1.470(4)
C(6)-C(7)	1.458(4)
C(7)-C(8)	1.387(4)
C(7)-C(14)	1.433(4)
C(8)-H(4)	0.92(3)
C(9)-C(10)	1.333(4)
C(9)-H(6)	0.96(3)
C(10)-C(11)	1.466(4)

C(10)-H(7)	1.05(3)
C(11)-C(13)	1.323(4)
C(11)-C(12)	1.496(4)
C(12)-H(8)	0.92(4)
C(12)-H(9)	1.00(4)
C(12)-H(10)	0.91(4)
C(13)-H(11)	0.99(3)
C(14)-H(12)	1.01(3)

bond angles in [°]

C(8)-N(1)-C(1)	109.1(2)
C(8)-N(1)-H(5)	128(2)
C(1)-N(1)-H(5)	123(2)
N(1)-C(1)-C(2)	128.0(3)
N(1)-C(1)-C(6)	108.5(2)
C(2)-C(1)-C(6)	123.5(3)
C(3)-C(2)-C(1)	117.0(3)
C(3)-C(2)-H(1)	124.7(16)
C(1)-C(2)-H(1)	118.2(16)
C(2)-C(3)-C(4)	121.1(3)
C(2)-C(3)-H(2)	118.5(17)
C(4)-C(3)-H(2)	120.2(17)
C(5)-C(4)-C(3)	122.7(3)
C(5)-C(4)-H(3)	118.7(17)
C(3)-C(4)-H(3)	118.5(17)
C(4)-C(5)-C(6)	117.2(2)
C(4)-C(5)-C(9)	121.8(3)
C(6)-C(5)-C(9)	121.0(3)
C(5)-C(6)-C(1)	118.4(2)
C(5)-C(6)-C(7)	136.3(2)
C(1)-C(6)-C(7)	105.3(2)
C(8)-C(7)-C(14)	117.6(3)
C(8)-C(7)-C(6)	105.9(2)
C(14)-C(7)-C(6)	136.4(2)
N(1)-C(8)-C(7)	111.1(3)
N(1)-C(8)-H(4)	121(2)
C(7)-C(8)-H(4)	128(2)
C(10)-C(9)-C(5)	125.2(3)
C(10)-C(9)-H(6)	118.9(18)
C(5)-C(9)-H(6)	115.8(18)
C(9)-C(10)-C(11)	125.1(3)
C(9)-C(10)-H(7)	120.4(14)
C(11)-C(10)-H(7)	114.4(14)
C(13)-C(11)-C(10)	116.5(3)
C(13)-C(11)-C(12)	124.2(3)

C(10)-C(11)-C(12)	119.3(3)
C(11)-C(12)-H(8)	110(2)
C(11)-C(12)-H(9)	111(2)
H(8)-C(12)-H(9)	103(3)
C(11)-C(12)-H(10)	110(2)
H(8)-C(12)-H(10)	111(3)
H(9)-C(12)-H(10)	112(3)
C(11)-C(13)-Cl(1)	125.5(2)
C(11)-C(13)-H(11)	124.4(17)
Cl(1)-C(13)-H(11)	110.0(17)
O(1)-C(14)-C(7)	128.6(3)
O(1)-C(14)-H(12)	115.5(17)
C(7)-C(14)-H(12)	115.9(17)

Table S12 Anisotropic displacement parameters (Å^2x 10^3) for indiacene B.

	U^{11}	U^{22}	U^{33}	U^{23}	U^{13}	U^{12}
Cl(1)	39(1)	48(1)	27(1)	12(1)	-9(1)	-4(1)
N(1)	39(2)	29(1)	19(1)	3(1)	-4(1)	5(1)
O(1)	32(1)	40(1)	21(1)	1(1)	4(1)	4(1)
C(1)	30(2)	20(1)	20(1)	0(1)	-3(1)	-1(1)
C(2)	42(2)	31(2)	16(1)	-1(1)	0(1)	-2(1)
C(3)	32(2)	38(2)	20(1)	-3(1)	3(1)	-2(1)
C(4)	28(2)	34(2)	24(2)	-3(1)	-2(1)	1(1)
C(5)	27(2)	24(1)	20(1)	-1(1)	-3(1)	-2(1)
C(6)	27(2)	21(1)	16(1)	-1(1)	0(1)	-3(1)
C(7)	26(2)	23(1)	22(1)	0(1)	-3(1)	0(1)
C(8)	33(2)	27(2)	27(2)	1(1)	-3(1)	6(1)
C(9)	28(2)	27(2)	20(1)	-2(1)	0(1)	-1(1)
C(10)	28(2)	27(2)	18(1)	0(1)	0(1)	-1(1)
C(11)	31(2)	24(2)	18(1)	-2(1)	-1(1)	0(1)
C(12)	31(2)	46(2)	25(2)	6(1)	4(1)	5(2)
C(13)	33(2)	31(2)	20(1)	4(1)	-1(1)	-2(1)
C(14)	26(2)	25(2)	26(1)	-1(1)	0(1)	1(1)

Table S13 Hydrogen coordinates (x 10^4) and isotropic displacement parameters ($Å^2$x 10^3) for indiacene B.

	x	y	z	U(eq)
H(1)	6690(40)	4700(20)	2882(19)	23(7)
H(2)	9700(50)	3680(20)	3360(20)	32(8)
H(3)	9940(50)	3070(30)	4940(20)	39(9)
H(4)	1170(60)	5620(20)	4890(20)	35(9)
H(5)	3080(50)	5420(20)	3520(20)	36(9)
H(6)	6200(50)	3570(20)	6800(20)	36(9)
H(7)	10500(40)	2930(20)	6383(18)	24(7)
H(8)	6380(70)	2250(30)	8120(20)	63(13)
H(9)	6690(60)	3370(30)	8370(30)	60(11)
H(10)	7740(50)	2480(30)	8960(30)	50(10)
H(11)	12490(50)	2440(20)	7650(20)	34(9)
H(12)	1310(50)	5240(20)	6560(20)	31(8)

S3.7.2 Chromatographic, mass spectrometric and NMR data of chloroxanthic acid A
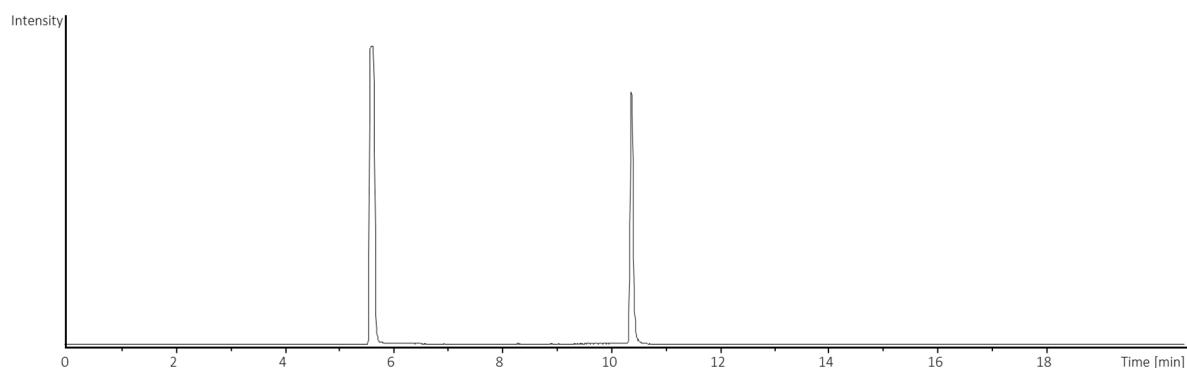
Figure S53 EIC at 295.08 for chloroxanthic acid A.

The two peaks visible in the EIC at 295.08 for chloroxanthic acid A likely derive from the *E/Z* isomers of chloroxanthic acid A, based on the high difference in retention time on rp-HPLC. As the double bond in position 9 is thermodynamically more stable and was determined to be in *E*-configuration based on the coupling constant of 16.1 Hz of proton 9, most likely those isomers derive from isomerization of the double bond in position 2 in aqueous solution. We only were able to generate NMR data for the mayor derivative, corresponding to the earlier eluting peak due to general low production of the compounds. In the NMR spectra we however are able to observe a second minor broad doublet at δ^1H = 6.46 showing ca. 30% of the intensity of the mayor peak for proton 3. Therefore, we likely already observe a fast isomerization during NMR analysis in methanolic solution, although the isomers are nicely separable by semipreparative HPLC. The coupling constant of 8.4 Hz for the minor derivative points towards a *Z*-configuration of the minor and an *E*-configuration of the mayor isomer. Due to overlapping signals of the mayor proton signal at position 3 with the proton at position 13, the coupling constant of the proton belonging to the mayor isomer could not be determined.
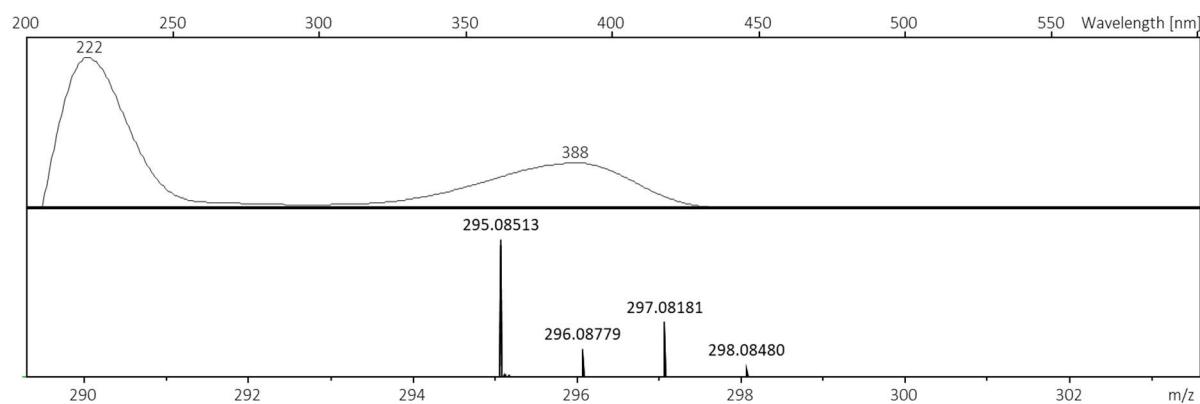
Figure S54 UV-absorption and isotopic pattern for [M+H]+ of chloroxanthic acid A.
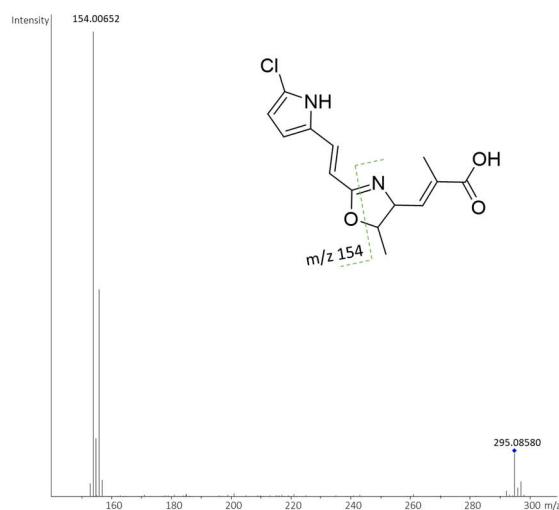
Figure S55 MS/MS Fragmentation pattern of chloroxanthic acid A.

Table S14 NMR signals of chloroxanthic acid A. Carbon numbers are shown in structure below.

Carbon No.	δ ¹H [ppm]	Multiplicity, *J* [Hz] and number of protons	δ ¹³C [ppm]	COSY correlations	HMBC correlations
1	-	-	174.5	-	-
2	-	-	135.6	-	-
3	6.38	m*, 1H	136.9	4	-
4	4.55	dd, 1H, 9.3, 7.2	71.4	3,5	5,6,8
5	4.45	m, 1H	83.2	4,6	3,4,8
6	1.43	dd, 1H, nd	20.3	5	4,5
7	1.94	d, 1H, 1.44	14.0	-	1,2,3,4,5
8	-	-	166.9	-	-
9	6.18	d, 1H, 16.1	108.2	10	18,10,11
10	7.13	d, 1H, 16.2	131.8	9	8,9,11,12,13
11	-	-	129.8	-	-
12	6.03	d, 1H, 3.7	109.0	13	11,13,14
13	6.38	d*, 1H	115.6	12	10,11,12
14	-	-	120.3	-	-
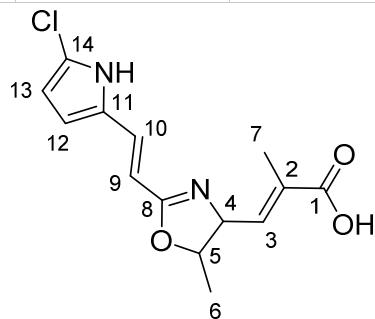
References

(1) Pena-Pereira, F.; Wojnowski, W.; Tobiszewski, M. AGREE-Analytical GREEnness Metric Approach and Software. *Analytical chemistry* **2020**, *92*, 10076–10082.

(2) Okoth Dorothy A.; Hug, J. J.; Garcia, R.; Spröer, C.; Overmann, J.; Müller, R. 2-Hydroxysorangiadenosine: Structure and Biosynthesis of a Myxobacterial Sesquiterpene–Nucleoside. *Molecules (Basel, Switzerland)* **2020**, *25*.

(3) Zhang, F.; Braun, D. R.; Rajski, S. R.; DeMaria, D.; Bugni, T. S. Enhypyrazinones A and B, Pyrazinone Natural Products from a Marine-Derived Myxobacterium Enhygromyxa sp. *Marine drugs* **2019**, *17*.

(4) Oueis, E.; Klefisch, T.; Zaburannyi, N.; Garcia, R.; Plaza, A.; Müller, R. Two Biosynthetic Pathways in Jahnella thaxteri for Thaxteramides, Distinct Types of Lipopeptides. *Org. Lett.* **2019**, *21*, 5407–5412.

(5) 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.

(6) Hug, J. J.; Panter, F.; Krug, D.; Müller, R. Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria. *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 319–334.

(7) Korp, J.; Winand, L.; Sester, A.; Nett, M. Engineering Pseudochelin Production in *Myxococcus xanthus*. *Appl. Environ. Microbiol.* **2018**, *84*, e01789-18.

(8) Gorges, J.; Panter, F.; Kjaerulff, L.; Hoffmann, T.; Kazmaier, U.; Müller, R. Structure, Total Synthesis, and Biosynthesis of Chloromyxamides: Myxobacterial Tetrapeptides Featuring an Uncommon 6-Chloromethyl-5-methoxypipecolic Acid Building Block. *Angew. Chem. Int. Ed. Engl.* **2018**, *57*, 14270–14275.

(9) Panter, F.; Krug, D.; Baumann, S.; Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **2018**, *9*, 4898–4908.

(10) Hoffmann, T.; Krug, D.; Bozkurt, N.; Duddela, S.; Jansen, R.; Garcia, R.; Gerth, K.; Steinmetz, H.; Müller, R. Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat. Commun.* **2018**, *9*, 803.

(11) Dehhaghi, M.; Tan, V.; Heng, B.; Mohammadipanah, F.; Guillemin, G. J. Protective Effects of Myxobacterial Extracts on Hydrogen Peroxide-induced Toxicity on Human Primary Astrocytes. *Neuroscience* **2019**, *399*, 1–11.

(12) Lee, C.; Park, S.; Ayush, I.; Cho, K.; Kim, S. S.; Kang, I.; Choe, W.; Kim, Y.-S.; Yoon, K.-S. Effects of Myxococcus fulvus KYC4048 Metabolites on Breast Cancer Cell Death. *Journal of Microbiology and Biotechnology* **2018**, *28*, 765–775.

(13) Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. Sect. A* **2008**, *64*, 112–122.

Chapter 4

Sandacrabins - Antiviral and Insecticidal Farnesylated Benzimidazoles Produced by a Rare Myxobacterium

Manuscript in preparation

Chantal D. Bader,[1] Fabian Panter,[1] Ronald Garcia,[1] Christine Walt,[1] Sibylle Haid,[2] Cathrin Spröer,[3] Alexander Kiefer,[1] Thomas Pietschmann,[2] Jörg Overmann[3] and Rolf Müller *[1]

Affiliation

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

[2] 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), Feodor-Lynen-Str. 7, 30625 Hannover, Germany

[3] Leibniz-Institut DSMZ – Deutsche Sammlung von Mikroorganismen und Zellkulturen, Inhoffenstraße 7, 38124 Braunschweig.

Contributions and Acknowledgements

Author's effort:

The author contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. The author performed isolation and structure elucidation of the sandacrabins, acquired MS^n spectra, performed the sandacrabins synthesis and the insecticidal assays. Furthermore, the author contributed to conceiving and writing this manuscript.

Contributions by others:

Fabian Panter contributed by genomic analysis including development of a biosynthesis hypothesis and proofreading of the manuscript. Ronald Garcia contributed by isolation of *Sandaracinus* MSr10575 and its morphological description besides its phylogenetic classification. Christine Walt helped with the purification of the sandacrabins. Sibylle Haid and Thomas Pietschmann performed and analyzed the antiviral assays. Cathrin Spröer and Jörg Overmann sequenced the genome of *Sandaracinus* MSr10575. Stefanie Schmidt and Alexandra Amann performed the antimicrobial and cytotoxicity assays. Nestor Zaburannyi assembled *Sandaracinus* MSr10575s' genome. Alexander Kiefer helped with the chemical synthesis. Rolf Müller contributed to conception and supervised this study. He also edited and proofread the manuscript.

4.1 Abstract

In this study we report the isolation, structure elucidation and total synthesis of three terpenoid alkaloids, the sandacrabins, alongside with the first description of their producing organism *Sandaracinus* MSr10575. We expand the myxobacterial *Sandaracineae* family by its second member and present the strains' genome sequence. Furthermore, we identify dimethylbenzimidazol, deriving from cobalamine biosynthesis, as key intermediate in the sandacrabin biosynthesis. Antimicrobial, antiviral, insecticidal and cytotoxic activity profiling of the sandacrabins shows that all sandacrabins except for sandacrabin A exhibit good insecticidal activity against *Acyrthosiphon pisum* and antiproliferative activity against HCT-116 cells. Sandacrabin B-E furthermore reveal antiviral activity against the SARS-CoV2 model virus CoV229E. Chemical modification of the sandacrabin core scaffold shows that the moderate antiproliferative activity of sandacrabin B can be increased to an IC_{50} of 0.3 μg/mL against HCT116 cells in the mono-farnesylated derivative sandacrabin D. The non-cytotoxic sandacrabin A in contrary, exhibits the best antimicrobial activities with an MIC of 32 μg/mL against *Bacillus subtilis*, which highlights the diversity of biological activities within this natural product family.
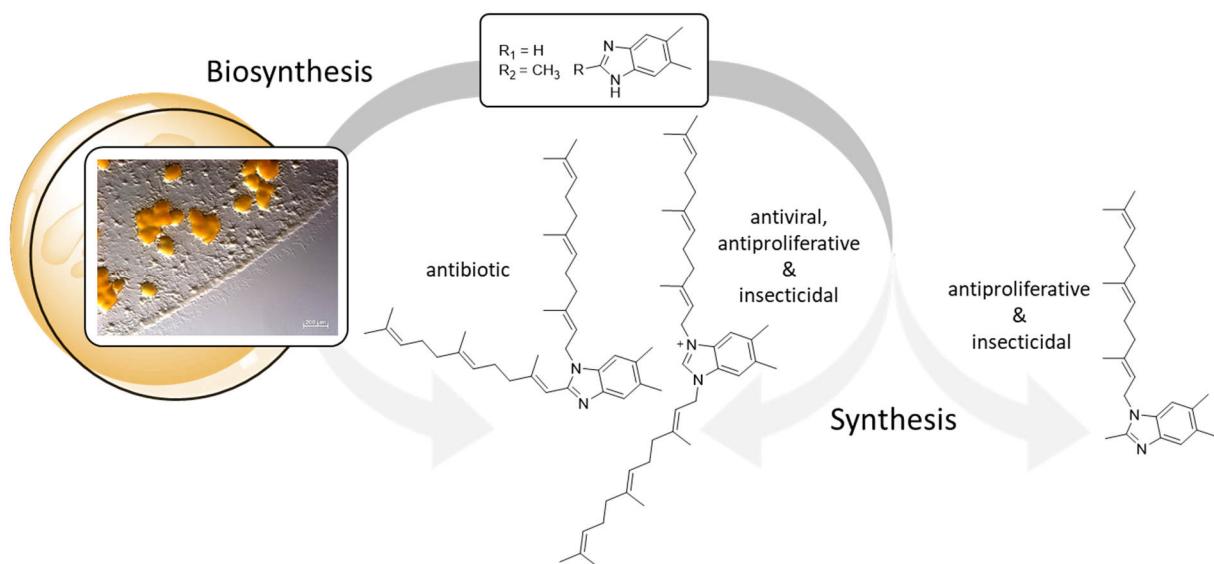
Figure 1 Table of contents graphic showing the key aspects of the study.

4.2 Introduction

Natural products (NPs) often serve as a rational starting point for drug development as they commonly show intriguing biological activities based on their complex chemical scaffolds optimized during evolutionary processes.[1] The evolving viral pandemics such as the coronavirus disease or swine flu pandemic and the progressive spread of antibacterial resistances, for example the propagation of multidrug resistant tuberculosis, led to an increased recurrence in natural product research to supply the drug discovery pipelines. Besides the evaluation of already described NPs, characterization of novel NPs is very much needed.[2,3] Among the NPs, bacterial secondary metabolites have always greatly contributed to the stream of natural product based drug leads, especially in anti-infective research and oncology.[4] Myxobacteria, an order of soil-dwelling δ-proteobacteria, are a prolific source of such structurally diverse NPs, which are known to exhibit a broad range of potent antimicrobial, cytotoxic and anti-parasitic activities.[5] Especially newly isolated myxobacterial strains that show sufficient phylogenetic distance from well-described genera have proven to be a fruitful source of novel NPs.[6]

As part of our continuous screening efforts of novel myxobacterial strains for yet undescribed natural products, the myxobacterial strain *Sandaracinus* species MSr10575 gained our attention. The only member of the *Sandaracinus* family so far is *Sandaracinus amylolyticus* type strain NOSO-4(T), a strain investigated on genomic level for its starch degrading properties with emphasis on alpha-amylases.[7] Secondary metabolome screening of *Sandaracinus amylolyticus* NOSO-4(T) furthermore led to isolation of two prenyl indols (indiacen A and B) in an bioactivity-guided isolation approach. The purified secondary metabolites were found to exhibit both antibacterial and antifungal properties.[8] Production of indiacen A and B is also observed and confirmed in *Sandaracinus* sp. MSr10575, where we additionally detected myxochelin A and terrestribisamid A in the strain's extracts.[9] In depth analysis of the *Sandaracinus* MSr10575 metabolome by high performance liquid chromatography coupled to mass spectrometry (HPLC-MS) additionally revealed a group of three peaks at *m/z* 541.45, 555.47 and 569.48 in positive ESI ionization mode. The three corresponding secondary metabolites were named sandacrabins A-C. According to sum formula predictions based on HRMS data, their sum formula lacks oxygen atoms commonly present in polypeptide and polyketide NPs, which suggested their classification among the alkaloid family of terpenoid NPs. This class of myxobacterial natural products is of special interest, as terpenoids are significantly underrepresented within the known myxobacterial secondary metabolites, despite their biosynthetic gene clusters (BGCs) being not particularly rare in myxobacterial genomes.[10] Besides some small terpenoid NPs, such as geosmin and germacradienol,[11] which were first discovered and isolated from actinobacteria, only few myxobacterial terpenes such as salimyxin, cystodienoic acid, enhygromic acid and the aurachins were isolated from myxobacteria.[12–15] Among those, only the strongly bioactive aurachins belong to the class of alkaloid terpenoids, making the sandacrabins a promising target for isolation and structure elucidation.[15,16]

4.3 Results

4.3.1Strain description

Sandaracinus MSr10575 was isolated in 2013 from the HZI soil collection (formerly GBF soil collection). The strain was found to form swarm colonies on a standard myxobacterial mineral salts agar baited with filter paper.[17] In axenic culture, the strain forms yellowish-orange colonies with flare edges towards the colony margin. Hump-shaped aggregates of orange cells are sometimes formed on the colony surface. Vegetative cells are slender rod-shaped with blunted ends capable of starch degradation (see Figure 2). Repeated sub-cultivations of the farthest swarm edge allowed the purification of the MSr10575 strain.

Figure 2 Growth morphology of MSr10575 on agar plates.

Based on 16S rRNA gene phylogenetic analysis MSr10575 was positioned among the *Sorangiineae* suborder in the *Sandaracinus* clade (see Figure 3). The 16S rRNA sequencing and subsequent BLAST (Basic Local Alignment Search Tool) search, revealed closest similarity of the strain (99.4%) to *Sandaracinus amylolyticus* NOSO4(T) (GenBank accession number: KP306728).[18] *Sandaracinus* sp. MSr10575 exhibits starch degrading properties comparable to *Sandaracinus amylolyticus* NOSO4(T), with the respective alpha-amylases present in its genome (see SI). We find five alpha amylase type genes as well as an alpha 1-6 glucosidase gene, responsible for 1-6 glyosidic bonds in amylopectin for example. Its phylogenetic positioning in the *Sandaracinus* clade based on 16S rRNA was confirmed by genome sequencing using PacBio genome sequencing technology. The respective genome sequence of the strain will be deposited in GenBank alongside with the publication of this manuscript. The *Sandaracinus amylolyticus* type strain *Sandaracinus amylolyticus* NOSO-4(T) is the only member of the *Sandaracinus* clade characterized so far, which indicates that *Sandaracinus* sp. MSr10575 belongs to yet underexplored (myxo)bacterial genera.[18]
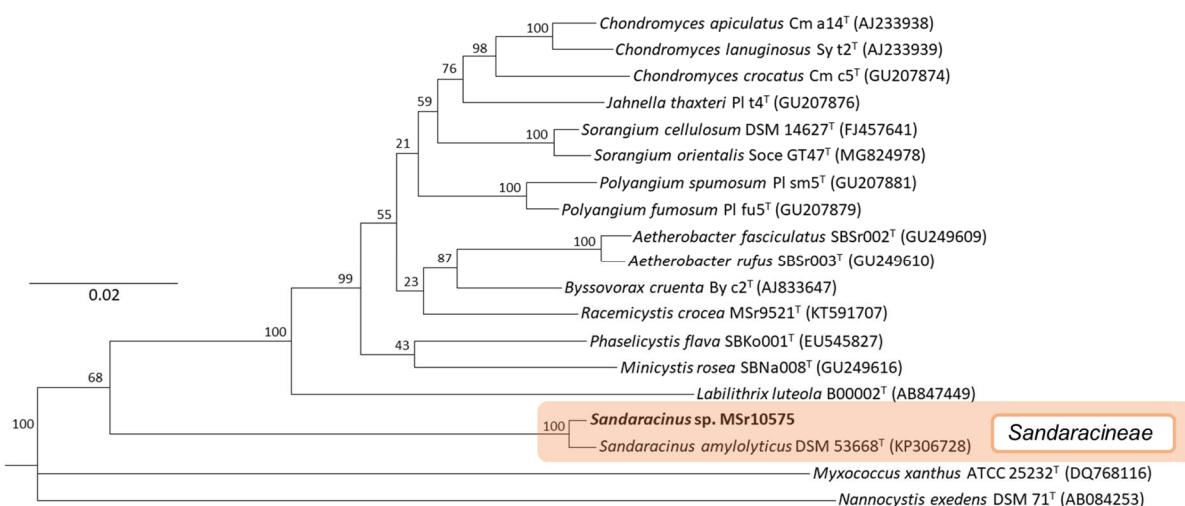
Figure 3 Phylogenetic classification of *Sandaracinus species* MSr10575.[19]

4.3.2 Isolation and structure elucidation of the sandacrabins

Sandacrabin A was purified by semi-preparative reverse phase (rp) HPLC from the dried hexane layer obtained by partitioning of a *Sandaracinus* sp. MSr10575 crude extract between methanol and hexane. The high lipophilicity of sandacrabin B and C along with their structural similarity, however, required a four-step purification process (see Figure 4). After separation of sandacrabin A from the crude extract, the remaining methanol layer was dried and partitioned between water and chloroform. The dried chloroform layer was first subjected to centrifugal partitioning chromatography (CPC). Separation of sandacrabin B and C from the respective CPC fractions was subsequently achieved by supercritical fluid chromatography (SFC) prior to isolation of the secondary metabolites by semi-preparative rp HPLC. HRESI-MS analysis of sandacrabin A showed an [M+H]$^+$ signal at *m/z* 541.4514 (calc. 541.4516 Δ = 0.4 ppm) consistent with the sum formula of $C_{38}H_{57}N_2$ containing 12 double bond equivalents (DBEs).
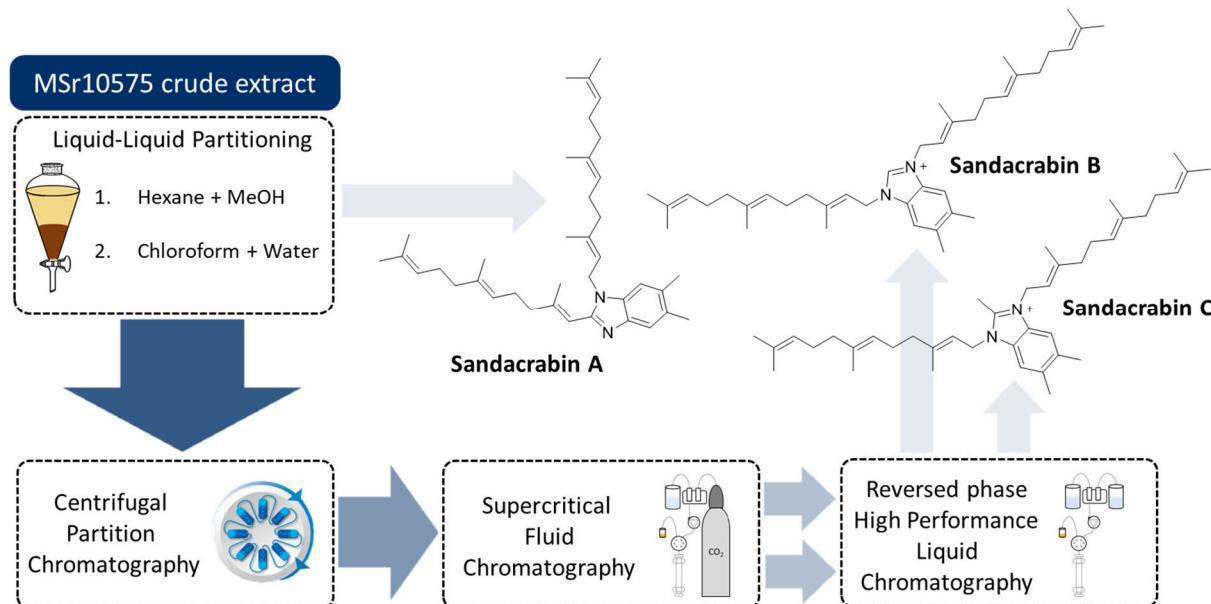
Figure 4 Structural formulae of sandacrabin A, B and C and the respective methods used for their purification.

1D and 2D NMR spectra of sandacrabin A (tables see SI) suggested a benzimidazole core structure of the molecule. The splitting pattern of the two aromatic protons in line with their COSY correlations furthermore revealed a 5,6-dimethyl substitution of this unit. COSY and HMBC correlations showed arrangement of the remaining methylene and methyl groups in two farnesyl moieties. Based on the chemical shift of the methylene group at $\delta(^1H)$ = 4.76 ppm and its correlation to the quaternary carbon at $\delta(^{13}C)$ = 134.1 ppm, the first farnesyl side chain is connected to the 5,6 dimethyl benzimidazole (DMB) core structure at position 1. HMBC correlations of the aliphatic double bond proton at $\delta(^1H)$ = 6.22 ppm, as well as its downfield shift showed, that the second farnesyl side chain substitutes the 2-position of DMB (see Figure 5A). The high-field shift of all methyl carbons indicates an all-E configuration of sandacrabin A. HRESI-MS analysis of sandacrabin B and C showed an $[M]^+$ signal at m/z 555.4514 (calc. 555.4673 Δ = 0.2 ppm) corresponding to the sum formula of $C_{39}H_{59}N_2$ and 569.4825 (calc. 569.4829 Δ = 0.7 ppm) corresponding to the sum formula of $C_{40}H_{61}N_2$, both containing 12 DBEs, comparable to sandacrabin A. 1D and 2D NMR-spectra of sandacrabin B and C (tables see SI), revealed a C_{2v}-symmetry (see Figure 5A) and a bisfarnesylated DMB core structure for both derivatives. The C_{2v}-symmetry of the molecules, as well as their permanent positive charge, indicated a 1,3-substitution of the DMB unit. In contrary to sandacrabin B, 1D and 2D NMR-spectra of sandacrabin C revealed a methylation in 2-position of the DMB core.

Tautomerization of the imidazole double bond as well the symmetry axis passing through this part of the molecule however hindered a detection of the corresponding proton signal at this position in sandacrabin B, wherefore HRESI-MS^3 spectra of the three sandacrabins were used to provide additional proof for their structures (see Figure 5B).
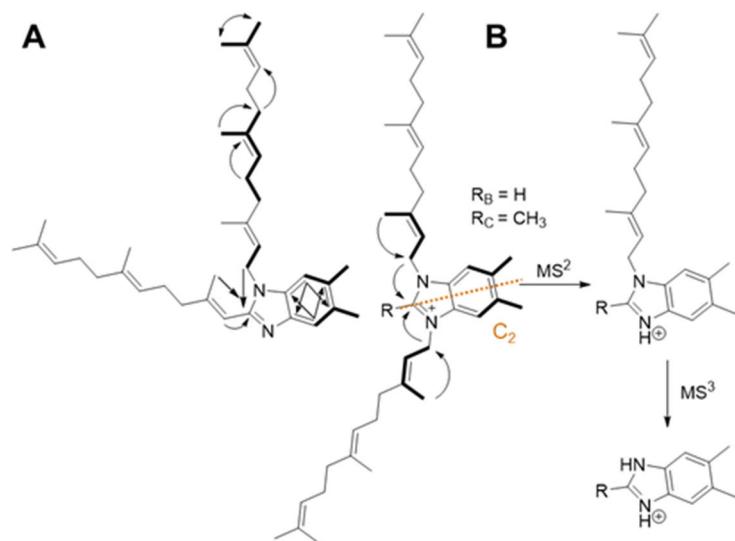
Figure 5 A Key NMR correlations used for structure elucidation. COSY correlations: bold line. HMBC correlations: indicated with arrows. Dashed orange line: symmetry axis of sandacrabin B. **B** Prominent MS^n fragments of sandacrabin B and C.

4.3.3 Biosynthesis of the sandacrabins

To the best of our knowledge, there are no farneslyated 5,6-dimethyl benzimidazoles described from bacteria yet, wherefore we were interested in elucidating their biosynthetic origin. The sandacrabin core unit (DMB) likely derives from the respective pathway that supplies this biosynthetic precursor to the cobalamin (vitamin B12) biosynthesis pathway.[9] The BluB enzyme, involved in this pathway oxidizes riboflavin to form DMB.[10] On the other hand, DMB biosynthesis can also derive from 5-aminoimidazole ribotide, which is converted to 1H-benzo[*d*]imidazole-5-ol by BzaA, D and F (see Figure 6). Subsequent methylation steps are carried out by BzaC and BzaD, before the resultant 5-methoxy-1H-benzo[d]imidazole is converted to DMB by BzaE.[9] Analogues for these enzymes can be found in the genome of *Sandaracinus* sp. MSr10575 (see SI), which underlines the ability of the strain to synthesize DMB.
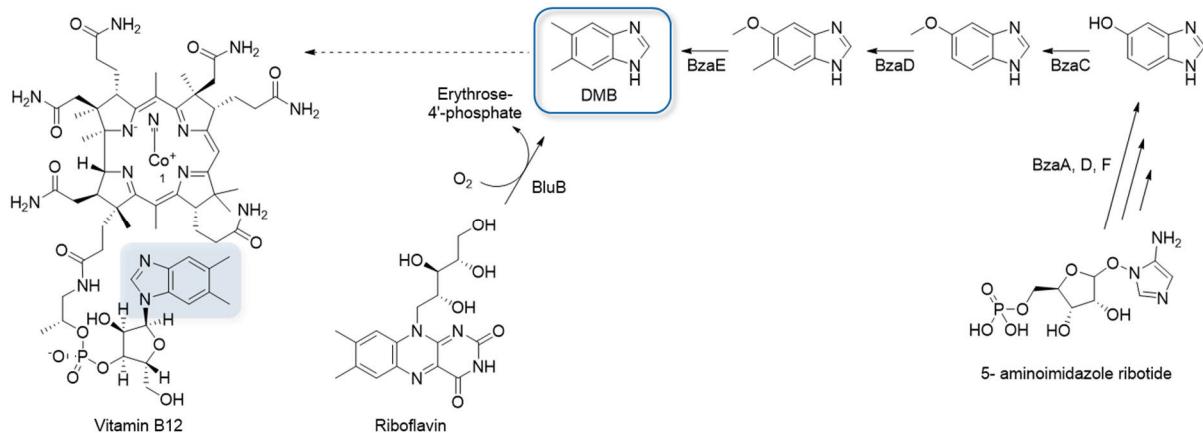
Figure 6 Biosynthesis of 5,6–dimethyl benzimidazole (DMB), the sandacrabins core structure, adapted from Mattes *et al.*.[20]

Bacteria can contain a reversible condensation enzyme that fuses formate or other carboxylic acids to 1,2-diamino-4,5-dimethylbenzol to form 5,6 dimethyl benzimidazole. This enzyme is likely to work similarly to GTP cyclohydrolase type enzymes such as RIB1, an enzyme that removes a C_1 unit bound to two vicinal nitrogen atoms as formate from GTP in a hydrolysis reaction.[10] This reaction likely plays a key part in biosynthesis of the DMB-derived core structures of sandacrabin A and C that may be produced by condensation of the 1,2-diamino-4,5-dimethylbenzol moiety with farnesoic acid or acetic acid, respectively. To obtain their mature structure, the different DMB core structures would need to be substituted by an *N*-farnesyl transferase that transfers one farnesyl sidechain to the precursor of sandacrabin A, while transferring two farnesyl residues to each of the two DMB ring nitrogen atoms of sandacrabin B and C (see Figure 7A).
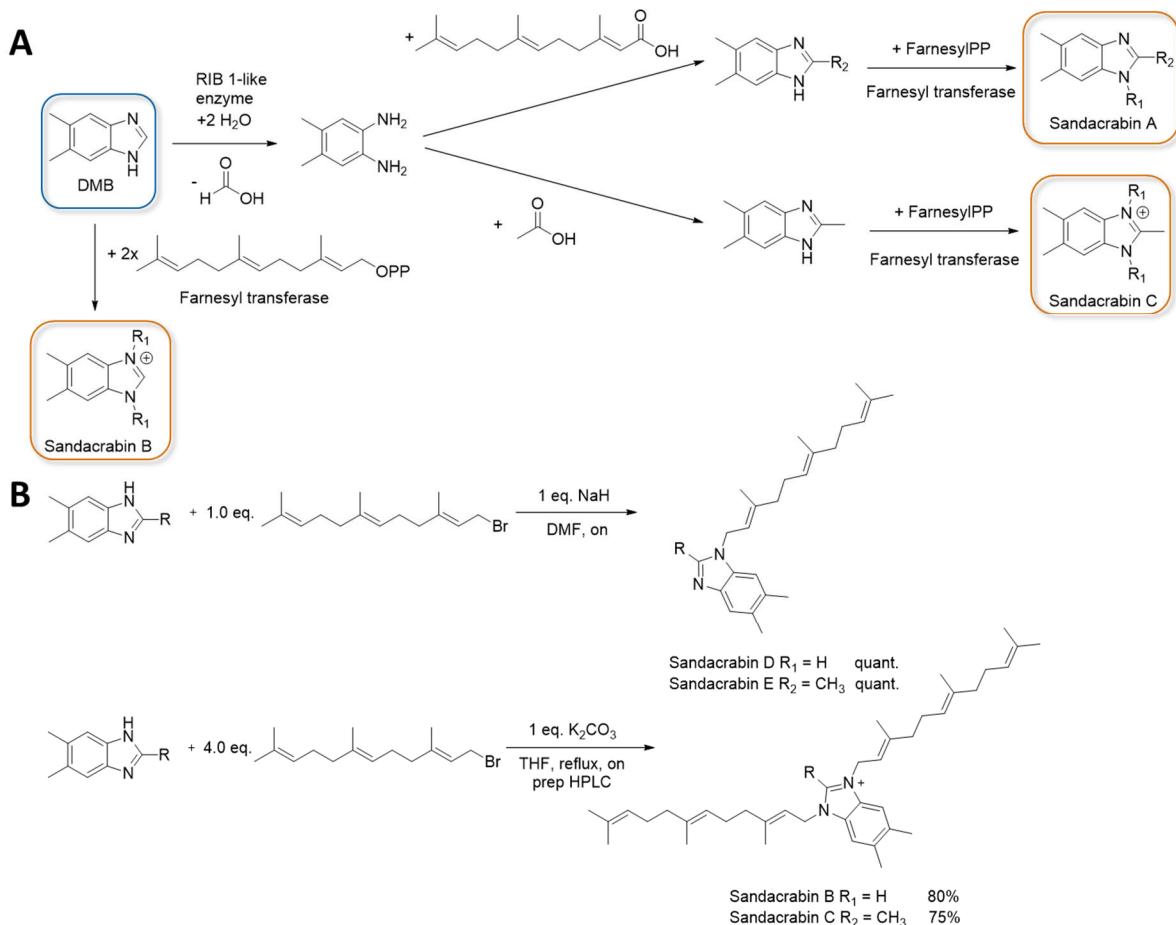
Figure 7 A Putative sandacrabin biosynthesis starting from DMB. **B** Synthesis of Sandacrabin B and C (lower part), as well as their mono-farneslyated derivatives sandacrabin D and E (upper part).

As the biosynthetic machinery supplying the DMB precursor was found encoded in close proximity to the cobalamin biosynthesis pathway, the complete sandacrabin biosynthetic machinery is not clustered in a single BGC as it is commonly observed in PKS or NRPS BGCs.[21] Furthermore, prenyl transferase enzymes such as the missing *N*-farnesyl transferase are not exceptionally rare enzymes in myxobacteria, wherefore we were unable to pinpoint the genes responsible for sandacrabin biosynthesis.[11] An overview about putative prenyl transferases detected in the *Sandaracinus* MSr10575 genome can be found in the SI. Although we were not able to pinpoint the exact prenyl transferase involved in the sandacrabins biosynthesis, the presence of several of these biosynthesis enzymes underlines our biosynthesis hypothesis. Further studies to explore sandacrabin biosynthesis by heterologous expression, gene inactivation experiments and in-vitro reconstitution of the biosynthetic machinery would need to be done to fully characterize the biosynthetic pathway and validate our biosynthesis hypothesis.

4.3.4 Synthesis and biological evaluation of the sandacrabins

Despite the novelty of bacterial farnesylated 5,6-dimethyl benzimidazole biosynthesis, several synthetic approaches have already investigated benzimidazoles featuring various substituent patterns.[9,10,11] They were found to exhibit a broad range of pharmacological activities from antimicrobial and anticancer to anthelmintic, insecticidal and anti-inflammatory activities,[9,11] which raised our interest in assessing the pharmacological activities of the sandacrabins as well. The isolated yields of sandacrabin B and C were comparably low, so we first aimed for a total synthesis of sandacrabin B and C (see Figure 7B), wherefore we adapted the synthesis route from the general

experimental procedures for the synthesis of benzimidazoliums described by Lim *et al.*.[22] Decrease in temperature, while changing the deprotonating agent from K_2CO_3 to NaH during the S_N-type reaction also allowed us to obtain mono-farnesylated versions called sandacrabin D and E in quantitative yields. Sandacrabin D and E were used to study the influence of the second farnesylation on the biological activities of the sandacrabins (see Figure 7B). It is worth mentioning that we detect only trace amounts of those derivatives in the extracts of MSr10575 (see SI), showing that the sandacrabin biosynthesis is highly optimized for generating bis-farnesylated sandacrabins.

Minimal inhibitory concentrations (MICs) for sandacrabin A-E were determined against a panel of Gram-positive, Gram-negative, as well as fungal pathogens (see Table 1). Furthermore, half maximal inhibitory concentrations (IC_{50}) against HCT-116 cancer cells and antiviral activity against CoV229E in Huh-7.5 were assessed. Hu-7.5 cells constitutively expressing a firefly luciferase reporter gene were therefore infected by a renilla luciferase CoV229E reporter virus. This renilla-firefly luciferase dual assay allows assessment of reduction in viral replication, while monitoring the cell viability in parallel. Insecticidal activity was determined by applying 0.5 µl of an acetonic sandacrabin dilution to the posterior segment of adult *Acyrthosiphon pisum* adapted from the procedure described by Ahumada *et al.*.[23] After assessing the minimal insecticidal concentration in a small test group of three insects in a dilution series from 5-0.05 µg sandacrabin applied per insect, the death rate of 20 insects at this concentration was evaluated (see Table 1).

Table 1 Antimicrobial, insecticidal, antiviral and antiproliferative activities of the sandacrabins.

Test organism	MIC [µg/mL]				
	Sandacrabin A	Sandacrabin B	Sandacrabin C	Sandacrabin D	Sandacrabin E
Bacillus subtilis DSM-10	32	64	128	64	64
Staphylococcus aureus Newman	128	64-128	64-128	64	64
Candida. albicans DSM-1665	> 64	> 64	> 64	> 64	> 64
Pichia anomala	> 64	> 64	> 64	> 64	> 64
Escherichia coli BW25113	> 64	> 64	> 64	> 64	> 64
Escherichia coli ΔacrB JW0451-2	> 64	> 64	> 64	> 64	> 64
Pseudomonas aeruginosa PA14	> 64	> 64	> 64	> 64	> 64
Mycobacterium smegmatis mc^2 155	> 64	> 64	> 64	64	64
	IC_{50} [µg/mL]				
HCT-116	> 37	2.1	20.0	0.3	1.4
	Death rate at 0.5 µg/insect [%]				
Acyrthosiphon pisum	10	100	90	100	90

All sandacrabins showed moderate activity against *B. subtilis* and *S. aureus* in the antimicrobial assays, with sandacrabin A as the best antibiotic derivative exhibiting an MIC of 32 µg/mL against *B. subtilis.* In contrary to the three natural sandacrabins, the sandacrabins D and E show an additional weak activity against *M. smegmatis* at 64 µg/mL. However, the more promising biological activity of the sandacrabins are their antiproliferative and insecticidal properties against HCT-116 cells and *A. pisum*. Especially sandacrabin D and E, which are only mono-farnesylated, show strong antiproliferative

activity of 0.3 and 1.4 µg/mL, respectively. Among the natural sandacrabins, sandacrabin B exhibits the best antiproliferative activity (2.1 µg/mL), similar to the biological activities of aurachin D against HCT-116 cells (2.2 µg/mL).[14] All sandacrabins except for sandacrabin A, which did not show insecticidal activity up to 5 µg/insect, were found to be toxic to *A. pisum* at 0.5 µg/insect. Sandacrabin B and D displayed a 100% death rate, whereas sandacrabin C and E both showed a 90% death rate. This finding points towards a slight reduction in insecticidal activity by methylation of the benzimidazol core structure, but the results would need confirmation using a larger test group.

We also observed a significant reduction in viral replication when cells infected with the SARS-CoV2 model virus CoV229E were treated with sandacrabin B-E (see Figure 8). As observed for the insecticidal and cytotoxic activities sandacrabin A did not show any effect on viral replication. Although sandacrabin B and C exhibited cytotoxic effects at the highest test concentration, they both show a therapeutic window for use as antiviral medicines. Interestingly, sandacrabin D and E, which showed the highest antiproliferative activities against HCT-116 cells, did not exhibit toxic effects on Huh-7 cells.
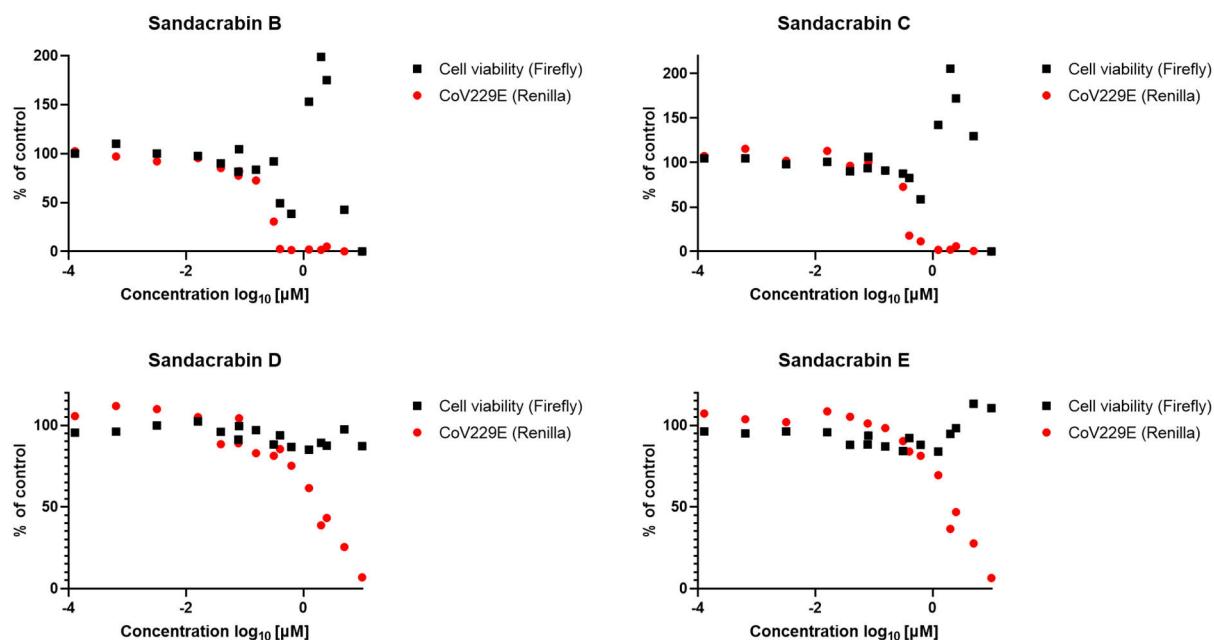
Figure 8 Antiviral activities of Sandacrabin B-E against CoV229E displayed as reduction in viral replication with simultaneous determination of the cell viability of the Huh-7 host cells.

4.4 Discussion

With this study we extend the myxobacterial *Sandaracinus* family by its second member, *Sandaracinus* species MSr10575 and report its genome sequence. Analysis of the strain's metabolome revealed three alkaloid terpenoids that were named the sandacrabins. Their isolation and structure elucidation showed that those farnesylated DMBs belong to the family of alkaloid terpenoids. Substituted benzimidazoles have already been studied extensively in synthetic chemistry approaches towards their antimicrobial, antiviral, antiproliferative and insecticidal activities, which we could also observe for the sandacrabins. Compared to the sandacrabins, synthetic mono-geranylated 5,6-dimethylbenzimidazole and mono-farnesylated 1-methylbenzimidazole derivatives, which were studied in a *Tribolium* chitin synthetase inhibition assay for generation of novel insecticidal compounds, showed highest structural similiarities.[24] Bis-terpenylated benzimidazoles however, are rarely found in literature. One bis-prenylated benzimidazole was generated by Holtgrewe *et al.* as intermediate for studying the

rearrangement of electron-rich *N*-allyldibenzotetraazafulvalenes. Instead of DMB it incorporates benzimidazole and its biological activities were never evaluated.[25]

Even though several synthetic approaches already generated various compounds with benzimidazole core structures, to the best of our knowledge the sandacrabins represent the first terpenylated benzimidazoles of bacterial origin, wherefore we developed a concise biosynthesis hypothesis for them. Analysis of the potential genes involved, revealed that sandacrabin biosynthesis likely consists of two steps: generation of the different DMB core structures and subsequent farnesylation by a prenyltransferase. Cobalamine biosynthesis also requires the synthesis of DMB, wherefore the biosynthesis pathway that supplies DMB to Vitamin B12 biosynthesis likely also supplies the sandacrabin biosynthesis with the respective DMB precursors. Removal of the DMB C_1, bound to the two vicinal nitrogen atoms, would allow subsequent generation of the two sandacrabin derivatives A and C, whereas sandacrabin B incorporates the native DMB. Besides the genes involved in cobalamine biosynthesis, we could identify several genes encoding for prenyltransferases in the *Sandaracinus* MSr10575, supporting our biosynthesis hypothesis.

Sandacrabin biosynthesis seems to be optimized towards the biosynthesis of sandacrabin A, which we find to be produced in highest amounts in comparison to the other natural derivatives. Driven by the intriguing biological activities described for synthetic benzimidazole derivatives, we developed a chemical synthesis route for sandacrabin B and C produced in relatively low yields, allowing full characterization of their biological activities. We furthermore synthesized their mono-farnesylated derivatives sandacrabin D and E, which could only be detected in trace amounts in *Sandaracinus* sp. MSr10575 crude extracts.

The observed broad-spectrum activities of the different sandacrabins derivatives, which exhibit antibiotic, insecticidal, antiviral and antiproliferative effects generally point towards a defensive function of sandacrabins for *Sandaracinus* MSr10575. The insecticidal activity against *A. pisum* alongside with their straight forward synthesis indicates sandacrabins B-E as potential candidates for agricultural use, as their production could easily be upscaled. The detected reduction in viral replication of the SARS-CoV2 model virus CoV229E for sandacrabin B is one more example of a benzimidazole exhibiting antiviral activities, much needed in times of evolving viral pandemics such as the covid-19 pandemic. As sandacrabin B also showed cytotoxic effects on HCT-116 cells, further medicinal chemistry optimization would be needed to advance the sandacrabins as an antiviral drug candidates though. As sandacrabin A in contrary to the other derivatives exhibits focused activity against *B. subtilis* without showing any cytotoxicity up to 37 µg/ml and no insecticidal activity, modifications in the DMB substitution pattern (especially at position 2) could help to overcome these obstacles. The broad biological activities, which interestingly are different for the various sandacrabin derivatives, once more highlight the intriguing biosynthetic potential of rare myxobacteria and their suitability to isolate novel natural products for supplying the drug discovery pipeline.

4.5 References

(1) Dias, D. A.; Urban, S.; Roessner, U. A historical overview of natural products in drug discovery. *Metabolites* **2012**, *2*, 303–336.

(2) Romanelli, R. J.; Azar, K. M.J.; Sudat, S.; Hung, D.; Frosch, D. L.; Pressman, A. R. The Learning Health System in Crisis: Lessons from the Novel Coronavirus Disease Pandemic. *J. Mayocpico* [Online early access]. DOI: 10.1016/j.mayocpiqo.2020.10.004.

(3) Sergeev, R. S.; Kavaliou, I.; Sataneuski, U.; Gabrielian, A.; Rosenthal, A.; Tartakovsky, M.; Tuzikov, A. Genome-wide Analysis of MDR and XDR Tuberculosis from Belarus: Machine-learning Approach. *IEEE/ACM Trans. Comput. Biol. Bioinform.* [Online early access]. DOI: 10.1109/TCBB.2017.2720669. http://dx.doi.org/10.1109/TCBB.2017.2720669 (accessed August 30, 2017).

(4) Thomford, N. E.; Senthebane, D. A.; Rowe, A.; Munro, D.; Seele, P.; Maroyi, A.; Dzobo, K. Natural Products for Drug Discovery in the 21st Century: Innovations for Novel Drug Discovery. *Int. J. Mol. Sci.* **2018**, *19*.

(5) Herrmann, J.; Fayad, A. A.; Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **2017**, *34*, 135–160.

(6) Hoffmann, T.; Krug, D.; Bozkurt, N.; Duddela, S.; Jansen, R.; Garcia, R.; Gerth, K.; Steinmetz, H.; Müller, R. Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat. Commun.* **2018**, *9*, 803.

(7) Sharma, G.; Khatri, I.; Subramanian, S. Complete genome of the starch-degrading myxobacteria Sandaracinus amylolyticus DSM 53668T. *Genome Biol. Evol.* **2016**, 2520–2529.

(8) Steinmetz, H.; Mohr, K. I.; Zander, W.; Jansen, R.; Gerth, K.; Müller, R. Indiacens A and B: prenyl indoles from the *myxobacterium Sandaracinus amylolyticus*. *J. Nat. Prod.* **2012**, *75*, 1803–1805.

(9) Bader, C. D.; Neuber, M.; Panter, F.; Krug, D.; Müller, R. Supercritical Fluid Extraction Enhances Discovery of Secondary Metabolites from Myxobacteria. *Analytical chemistry* [Online early access]. DOI: 10.1021/acs.analchem.0c02995.

(10) Gregory, K.; Salvador, L. A.; Akbar, S.; Adaikpoh, B. I.; Stevens, D. C. Survey of Biosynthetic Gene Clusters from Sequenced Myxobacteria Reveals Unexplored Biosynthetic Potential. *Microorganisms* **2019**, *7*, 181.

(11) Schifrin, A.; Khatri, Y.; Kirsch, P.; Thiel, V.; Schulz, S.; Bernhardt, R. A single terpene synthase is responsible for a wide variety of sesquiterpenes in *Sorangium cellulosum* So ce56. *Org. Biomol. Chem.* **2016**, *14*, 3385–3393.

(12) Felder, S.; Kehraus, S.; Neu, E.; Bierbaum, G.; Schäberle, T. F.; König, G. M. Salimyxins and enhygrolides: antibiotic, sponge-related metabolites from the obligate marine myxobacterium *Enhygromyxa salina*. *ChemBioChem* **2013**, *14*, 1363–1371.

(13) Raju, R.; Mohr, K. I.; Bernecker, S.; Herrmann, J.; Müller, R. Cystodienoic acid: A new diterpene isolated from the myxobacterium *Cystobacter sp. J. Antibiot.* **2015**, *68*, 473–475.

(14) Tomura, T.; Nagashima, S.; Yamazaki, S.; Iizuka, T.; Fudou, R.; Ojika, M. An Unusual Diterpene—Enhygromic Acid and Deoxyenhygrolides from a Marine Myxobacterium, *Enhygromyxa* sp. *Mar. Drugs* **2017**, *15*.

(15) Kunze, B.; Höfle, G.; Reichenbach, H. The aurachins, new quinoline antibiotics from myxobacteria: Production, physico-chemical and biological properties. *J Antibiot (Tokyo)* **1987**, *40*, 258–265.

(16) Li, X. W.; Herrmann, J.; Zang, Y.; Grellier, P.; Prado, S.; Müller, R.; Nay, B. Synthesis and biological activities of the respiratory chain inhibitor aurachin D and new ring versus chain analogues. *Beilstein J. Org. Chem.* **2013**, *9*, 1551–1558.

(17) Garcia, R.; Müller, R. Family *Polyangiaceae*. In *The Prokaryotes*; Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F., Eds.; Springer-Verlag: Berlin Heidelberg, 2014; pp 247–279.

(18) Mohr, K. I.; Garcia, R. O.; Gerth, K.; Irschik, H.; Müller, R. *Sandaracinus amylolyticus* gen. nov., sp. nov., a starch-degrading soil myxobacterium, and description of *Sandaracinaceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* **2012**, *62*, 1191–1198.

(19) Gaffield, W. Circular dichroism, optical rotatory dispersion and absolute configuration of flavanones, 3-hydroxyflavanones and their glycosides: Determination of aglycone chirality in flavanone glycosides. *Tetrahedron* **1970**, *26*, 4093–4108.

(20) Mattes, T. A.; Deery, E.; Warren, M. J.; Escalante-Semerena, J. C. Cobalamin Biosynthesis and Insertion. In *Encyclopedia of inorganic and bioinorganic chemistry*; Scott, R. A., Ed.; John Wiley and Sons, Inc: [Hoboken, NJ], 2011; pp 1–24.

(21) Walsh, C. T. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Nat. Prod. Rep.* **2016**, *33*, 127–135.

(22) Lim, C.-H.; Ilic, S.; Alherz, A.; Worrell, B. T.; Bacon, S. S.; Hynes, J. T.; Glusac, K. D.; Musgrave, C. B. Benzimidazoles as Metal-Free and Recyclable Hydrides for CO_2 Reduction to Formate. *J. Am. Chem. Soc.* **2019**, *141*, 272–280.

(23) Ahumada, M. I.; Chorbadjian, R. A. Laboratory assays of the insecticidal activity of cyantraniliprole and imidacloprid on Brevicoryne brassicae , Myzus persicae (Hemiptera: Aphididae) and Trialeurodes vaporariorum (Hemiptera: Aleyrodidae) pests species and a biological control agent Chrysoperla defreitasi (Neuroptera: Chrysopidae). *Chil. j. agric. res.* **2019**, *79*, 658–663.

(24) Cohen, E.; Kuwano, E.; Eto, M. The use of a Tribolium chitin synthetase assay in studying the effects of benzimidazoles with a terpene moiety and related compounds. *Agric. Biol. Chem.* **1984**, *48*, 1617–1620.

(25) Holtgrewe, C.; Diedrich, C.; Pape, T.; Grimme, S.; Hahn, F. E. Rearrangement of Electron-RichN-Allyldibenzotetraazafulvalenes –An Experimental and Theoretical Study. *Eur. J. Org. Chem.* **2006**, *2006*, 3116–3124.

4.6 Supporting information

Sandacrabins - Antiviral and Insecticidal Farnesylated Benzimidazoles Produced by a Rare Myxobacterium

Manuscript in preparation

Chantal D. Bader,[1] Fabian Panter,[1] Ronald Garcia,[1] Christine Walt,[1] Sibylle Haid,[2] Cathrin Spröer,[3] Alexander Kiefer,[1] Thomas Pietschmann,[2] Jörg Overmann[3] and Rolf Müller *[1]

Affiliation

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

[2]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), Feodor-Lynen-Str. 7, 30625 Hannover, Germany

[2]Leibniz-Institut DSMZ – Deutsche Sammlung von Mikroorganismen und Zellkulturen, Inhoffenstraße 7, 38124 Braunschweig

Data with limited visibility on print media such as raw NMR data can be found on the enclosed storage medium.

S4.1   Cultivation conditions and extract generation

S4.1.1 Production and seed cultures

For fermentative production of the sandacrabins, *Sandaracinus* sp. MSr10575 was cultivated in a 50 ml pre-culture using a 300 ml shake flask containing 50 ml of 2SWT medium for 8 days at 30°C and 160 rpm. Sandacrabin production was achieved by fermentation of 6 x 5L shake flasks containing each 2L of 2SWYT medium and 40 ml of a 50% v/v sterilized XAD-16 resin. Each of these flasks was inoculated with one of the 50ml pre-cultures and fermented at 30°C and 160 rpm on an orbitron shaker for 2 weeks until the fermentation broth turns brightly orange.

S4.1.2 Extraction

Cells and XAD-16 resin were harvested by centrifugation and the residual medium was discarded. The XAD-16 cell mixture was repeatedly extracted using 50/50 mixtures of methanol/chloroform that were combined. The chloroform was evaporated under reduced pressure and the extract partitioned between methanol and hexane. Both fractions were dried under vacuum. The dry methanol layer was subsequently partitioned between chloroform and water and the chloroform phase was again dried again. The dried hexane layer was re-dissolved in methanol and used for the purification of sandacrabin A by semi-preparative reverse phase HPLC. The dried chloroform layer containing sandacrabin B and C was subjected to CPC separation on a Gilson CPC 100 SCPC system using a variation of the ARIZONA solvent system containing a 1:1:1:1 amount of TRIS buffer pH8, methanol, ethyl acetate and hexane. The lighter, more hydrophobic organic layer elutes a pre-purified fraction containing the sandacrabins.

S4.1.3 Media recipes

Table S1 Medium recipe of 2SWT medium

2SWT – Medium			
Amount	Ingredient	Concentration	Supplier
3 g/L	Tryptone	-	BD
1 g/L	Soytone	-	BD
3.5 g/L	Soluble Starch	-	Roth
4 g/L	Maltose Monohydrate	-	
2 g/L	Glucose	-	Roth
10 g/L	Starch (soluble)	-	Roth
0.5 g/L	$CaCl_2$	-	Sigma Aldrich
1 g/L	$MgSO_4$ x $7H_2O$	-	Grüssing
10 ml/L	TRIS x HCl ph8	1M	Sigma Aldrich
100 µL/L	Sterile Vit. B12 solution (added after autoclaving)	1 mg/ml	Roth
200 µL/L	Sterile FeEDTA solution (added after autoclaving)	8 mg/ml	Sigma Aldrich
Dissolved in milli-Q. Water, pH adjusted to 7.2 with 1N KOH			

Table S2 Medium recipe of 2SWYT medium.

<div align="center">

2SWYT – Medium

</div>

Amount	Ingredient	Concentration	Supplier
3 g/L	Tryptone	-	BD
1 g/L	Soytone	-	BD
3.5 g/L	Soluble Starch	-	Roth
4 g/L	Maltose Monohydrate	-	
10 g/L	Baker's yeast (alive)	-	
2 g/L	Glucose	-	Roth
10 g/L	Starch (soluble)	-	Roth
0.5 g/L	$CaCl_2$	-	Sigma Aldrich
1 g/L	$MgSO_4$ x $7H_2O$	-	Grüssing
10 ml/L	TRIS x HCl ph8	1M	Sigma Aldrich
100 µL/L	Sterile Vit. B12 solution (added after autoclaving)	1 mg/ml	Roth
200 µL/L	Sterile FeEDTA solution (added after autoclaving)	8 mg/ml	Sigma Aldrich
	Dissolved in milli-Q Water, pH adjusted to 7.2 with 1N KOH		

S4.2   Analytical methods

S4.2.1 UHPLC-*hr*MS measurements

UHPLC-qTOF measurements were carried out on a Dionex Ultimate 3000 SL system using a Bruker maXis 4G UHRqTOF for mass spectrometric analysis. The mobile phase consisted of (A) ddH2O with 0.1% formic acid and (B) acetonitrile with 0.1% formic acid. A linear gradient from 95-5% A in B at a flow rate of 0.6 mL/min was used. Dionex Ultimate 3000 SL system was equipped with a Waters Acquity BEH C18 column (100 x 2.1 mm, 1.7 µm dp). The column was heated to 45 °C and the sample injection volume 1 µL. The LC flow was split to 75 µL/min before entering the mass spectrometer, which was externally calibrated to a mass accuracy of < 1 ppm on sodium formate clusters. Mass spectra in a mass range from 150-2500 m/z were acquired in centroid mode at a 2 Hz scan rate. Measurements were conducted in positive ionization mode, using a capillary voltage of + 4000 V. Dry gas was set to a flow rate of 5 L/min at 200 °C.

S4.2.2 MS^n measurements

MS^2 and MS^3 experiments are carried out in positive ionization mode on a SolariX XR 7T FT-ICR (Bruker Daltonics, Bremen, Germany) equipped with an Apollo II ESI source. The mass spectrometer was externally calibrated to a mass accuracy below 1 ppm by injecting the LC/MS Calibration standard for ESI-TOF (Agilent). The samples were injected with a preinstalled syringe pump at 2.0 µl min^{-1} flowrate. 16 scans were performed accumulating for 50 ms. The data size was set to 2 M. The capillary voltage was set to -4500 V and the dry gas to a flow rate of 4 L/min at 220 °C. CID energy is set to 22 eV for sandacrabin A and 26 eV for sandacrabin B and C. SORI energy for the second fragmentation step after isolation of the generated fragment precursor ions is set to 1.29 for sandacrabin A, 1.40 for sandacrabin B and 1.26 for sandacrabin C.

S4.2.3 NMR measurements

NMR spectra were recorded on a Bruker UltraShield 500 or a Bruker Ascend 700 spectrometer equipped with a 5 mm TCI cryoprobe (^{1}H at 500 and 700 MHz, ^{13}C at 125 and 175 MHz). All observed chemical shift values (δ) are given in in ppm and coupling constant values (J) in Hz. Standard pulse programs were used for HMBC, HSQC, and gCOSY experiments. HSQC experiments were optimized for ^1JC-H = 145 Hz and HMBC experiments were optimized for $^{2,3}J$C-H = 6 Hz. The spectra were recorded in methanol-d_4 and chemical shifts of the solvent signals at δ_H 3.31 ppm and δ_C 49.2 ppm were used as reference signals for spectra calibration. To increase sensitivity, the measurements were conducted in a 5 mm Shigemi tube (Shigemi Inc., Allison Park, PA 15101, USA).

S4.3 Isolation procedure

Separation of sandacrabin B and C from the respective CPC fractions was achieved by supercritical fluid chromatography (SFC) prior to isolation of the secondary metabolites by semi-preparative reverse phase (rp) HPLC. Sandacrabin A was directly purified from the hexane extracts using semi-preparative HPLC.

S4.3.1 Centrifugal partition chromatography (CPC)

Pre-purification and fractionation of the different sandacrabins into four fractions is achieved on a Gilson CPC 100 device (Gilson purification S.A.S.) connected to a Varian ProStar Solvent delivery module and a Varian ProStar 2 Channel UV detector. Fraction collection is done with a Foxy Jr. autosampler (Isco). Used biphasic solvent system is the buffered ARIZONA solvent system consisting of a 1:1:1:1 mixture of 10 mM Tris x HCl buffer pH 8.0 (Sigma), methanol (analytical grade, Fluka), EtOAc (analytical grade, Fluka) and hexane (analytical grade, Fluka). After equilibration of the system and loading of the sample, the aqueous phase is used as a stationary phase for 40 min at 2 ml/min and a fraction size of 4 ml. Then the organic phase is used as a stationary phase in dual mode for 40 min at 2 ml/min and a fraction size of 4 ml. The CPC process is followed up on the UV detector at the wavelengths 220nm and 254nm. Subsequently the sandacrabins were detected in their respective fractions using our standard analytical UHPLC-hrMS.

S4.3.2 Supercritical fluid chromatography (SFC)

Sandacrabin B was separated from sandacrabin C on a Waters Prep 15 SFC System equipped with a 5 µm Torus Diol 130 Å OBD Prep Column 250 x 19 mm thermostated at 40°C. Separation was achieved using a linear gradient of 5-50 MeOH with 0.5 % triethylamine as a co-solvent over 10 minutes. The percentage of co-solvent was kept at 50 % for 1 minute before returning to 5 % in 1 minute and re-equilibration for 3 minutes. Flow rate was set to 15 mL/min and backpressure to 110 bar.

S4.3.3 Semi-preparative HPLC

The final purification step for all sandacrabins was performed on a Dionex Ultimate 3000 SDLC low pressure gradient system equipped with a Waters BEH 5 µm C18(2) 100 Å, LC Column 250 x 10 mm thermostated at 45 °C coupled to a Thermo Fisher Scientific ISQ™ EM single quadrupole mass spectrometer. Separation was achieved using a linear gradient from 15% (B) acetonitrile with 0.1% formic acid to 100% (B) ddH$_2$O with 0.1% formic acid over 20 minutes. Detection of the compounds was achieved by UV absorption at 210 nm and selected ion monitoring (SIM) at m/z 541.5, 555.5 and 569.5.

S4.4 *In-silico* analysis of the biosynthesis genes

To narrow down the biosynthetic origin of the sandacrabins we performed a global analysis of enzymes responsible for producing DMB and of farnesyl transferases in *Sandaracinus sp.* MSr10575. To support our biosynthesis hypothesis, we looked for the BluB BzaA and BzaE protein homologs that would potentially be involved in biosynthesis of the DMB core unit on the MSr10575 chromosome. Furthermore, we screened the bacterial chromosome for prenyl transferases that would transform DMB into the mature sandacrabins. Locus tags described here are shown with respect to the MSr10575 chromosome deposited on GenBank under Accession number to be announced (TBA).

Putative sandacrabin biosynthesis protein	locus-tag on the MSr10575 chromosome
BluB homolog 1	TBA
BzaA homolog 1	TBA
BzaE homolog 1	TBA
BzaE homolog 2	TBA
BzaE homolog 3	TBA
BzaE homolog 4	TBA
BzaE homolog 5	TBA
Farnesyl transferase homolog 1	TBA
Octaprenyl transferase homolog 1	TBA
Octaprenyl transferase homolog 2	TBA
Octaprenyl transferase homolog 3	TBA

It is very likely that the sandacrabin biosynthesis enzymes are among the proteins we discovered in silico. Still, as mutations in MSr10575 are difficult to achieve and we still lack the possibility to perform double crossover mutagenesis, we are unable to confirm the enzyme's involvement in the sandacrabin biosynthesis experimentally.

S4.5 Structure elucidation of the sandacrabins

S4.5.1 Full description of the *de-novo* structure elucidaton

HRESI-MS analysis of sandacrabin A showed an [M+H]$^+$ signal at *m/z* 541.4514 (calc. 541.4516 Δ = 0.4 ppm) consistent with the sum formula of $C_{38}H_{57}N_2$ containing 12 double bond equivalents (DBEs). The ^{1}H-NMR and HSQC spectra of sandacrabin A revealed two signals characteristic for aromatic double bonds at $\delta(^1H)$ = 7.37 (^{1}H, s) $\delta(^{13}C)$ = 119.6 and $\delta(^1H)$ = 7.15 (^{1}H, s) $\delta(^{13}C)$ = 111.4 ppm. Furthermore, six signals correlating to aliphatic double bond protons were detected at $\delta(^1H)$ = 6.22 (1H, s), 5.21 (1H, m), 5.16 (1H, m), 5.09 (1H, m) 5.00 (1H, m) and 4.99 ppm (1H, m). The proton spectrum exhibited 10 methyl groups at $\delta(^1H)$ = 2.38 (3H, s), 2.36 (3H, s), 2.05 (3H,m), 1.88 (3H, bs), 1.66 (3H, bs), 1.63 (3H, d, J = 1.06), 1.62 (3H, d, J = 1.06), 1.58 (3H, bs), 1.53 (3H, bs) and 1.51 ppm (3H, bs), as well as nine methylene groups at 4.76 (2H, d, J = 6.73), 2.33 (2H, m), 2.30 (2H, m), 2.11 (2H, m), 2.09 (2H, m), 2.06 (2H,m), 2.01 (2H, m), 1.91 (2H, m) and 1.82 ppm (2H, m). Based on their characteristic chemical shifts at $\delta(^{13}C)$ = 151.3, 142.2, 134.1, 133.0 and 132.3 ppm as well as HMBC correlations to the two aromatic double bond protons at $\delta(^1H)$ = 7.37 and 7.15 ppm, five quaternary carbons could be allocated in the aromatic ring system. The downfield shift of the three quaternary carbons at $\delta(^{13}C)$ 151.3, 142.2 and 134.1 ppm suggested a benzimidazole core structure of sandacrabin A. The splitting pattern of the two aromatic protons in line with COSY correlations of the methyl group at $\delta(^1H)$ = 2.36 ppm to the aromatic proton at $\delta(^1H)$ = 7.37 ppm and the methyl group at $\delta(^1H)$ = 2.38 ppm to the aromatic proton at $\delta(^1H)$ = 7.15

ppm revealed a 5,6 dimethyl substitution of the benzimidazole unit. COSY and HMBC correlations showed arrangement of the remaining methylene and methyl groups in two farnesyl moieties. Based on the chemical shift of the methylene group at $\delta(^1H)$ = 4.76 ppm and its correlation to the quaternary carbon at $\delta(^{13}C)$ = 134.1 ppm, the first farnesyl side chain is connected to the 5,6 dimethyl benzimidazole (DMB) core structure at position 1. HMBC correlations of the aliphatic double bond proton at $\delta(^1H)$ = 6.22 ppm, as well as its downfield shift showed, that the second farnesyl side chain substitutes the 2-position of the benzimidazole core structure. The high-field chemical shift of all methyl carbons pointed towards an all-*trans* configuration of the sandacrabins.

HRESI-MS analysis of sandacrabin B showed a $[M]^+$ signal at *m/z* 555.4674 (calc. 555.4673 Δ = 0.2 ppm) consistent with the sum formula of $C_{39}H_{59}N_2$ containing 12 DBEs like sandacrabin A. 1D and 2D NMR-spectra of sandacrabin B, revealed a C2v-symmetry of the molecule (see Figure 3). Characteristic chemical shifts of two aromatic protons at $\delta(^1H)$ = 7.64 (2H, s), two methyl groups at $\delta(^1H)$ = 2.48 (6H, s) and four quaternary carbons at $\delta(^{13}C)$ = 138.5 and 131.4 ppm as well as their COSY and HMBC correlations showed that sandacrabin B also contains a DMB core structure, comparable to sandacrabin A. COSY and HMBC correlations of further eight methyl groups, six aliphatic double bond protons, as well as ten methylene groups suggested two farnesyl substituents of the DMB core structure as well. The C2v-symmetry of the molecule, as well as its permanent positive charge indicated a 1,3- substitution of the DMB unit. A carbon signal at $\delta(^{13}C)$ = 140.2 ppm showed correlations to the two starter methylene groups of the farnesyl branches at $\delta(^1H)$ = 5.07 ppm, revealing its allocation at the two position in the benzimidazole core structure. Tautomerization of the imidazole double bond as well as passage of the symmetry axis at this part of the molecule however, hindered a detection of the corresponding proton signal at this position. HRESI-MS analysis of sandacrabin C showed a $[M]^+$ signal at *m/z* 569.4825 (calc. 569.4829 Δ = 0.7 ppm) consistent with the sum formula of $C_{40}H_{61}N_2$ containing 12 DBEs like sandacrabin A and B. 1D and 2D NMR-spectra revealed a quaternary carbon at $\delta(^{13}C)$ = 150.5 ppm, as well as its substituent, a methyl group at $\delta(^1H)$ = 2.81 and $\delta(^{13}C)$ = 10.4 ppm. Besides the quaternary carbon at $\delta(^{13}C)$ = 140.2 ppm in sandacrabin B, the remaining carbon and proton chemical shifts were identical for the two derivatives. In line with the C2v symmetry of sandacrabin C, this observation revealed a methylation in 2-position of the 5,6 dimethyl benzimidazole core. HRESI-MS[3] of the three sandacrabins gave additional proof for their structures proposed based on the 1D and 2D NMR spectra. They all showed loss of one farnesyl side chain. For sandacrabin B an additional peak at *m/z* 147.1 corresponding to its 5,6–dimethyl benzimidazole core structure was detected in the positive ionization ESI MS[3] spectra, respectively for sandacrabin C a peak at *m/z* 161.1074 corresponding to 1,5,6–trimethyl benzimidazole.

S4.5.2 NMR spectroscopic, *hr*ESI-MS and *hr*ESI-MS³ data of the natural sandacrabins

Table S3 NMR spectroscopic data for sandacrabin A in methanol-d_4 at 500/125 MHz.

#	Δ ¹³C [PPM]	Δ ¹H [PPM], MULT (*J* [HZ])	COSY	HMBC
1	132.4	-	-	-
2	125.5	5.09, m	3, 29, 30	3, 4, 29, 30
3	27.9	2.09, m	2, 4	1, 2, 4
4	41.0	2.01, m	3	2, 3, 5, 31
5	137.0	-	-	-
6	124.8	5.21, m	7, 31	4, 5, 7, 8, 31
7	27.4	2.30, m	6, 8, 31	5, 6, 8, 9
8	41.5	2.33, m	7	6, 7, 9, 10, 11, 32
9	136.8	-	-	-
10	113.2	6.22, s	-	8, 9, 11, 32
11	151.3	-	-	-
12	142.2	-	-	-
13	119.6	7.37, s	16, 33	15, 17, 33
14	132.3	-	-	-
15	134.0	-	-	-
16	111.4	7.15, s	13, 34	12, 14, 34
17	134.1	-	-	-
18	42.9	4.76, d (6.73)	19, 35	11, 17, 19, 20, 21
19	121.1	5.16, m	18, 35	18, 21, 35
20	140.4	-	-	-
21	40.5	2.06, m	22	19, 20, 22, 23, 35
22	27.1	2.11, m	21, 23, 36	20, 21, 23, 24
23	124.8	5.00, m	23, 36	21, 22, 25, 36
24	136.5	-	-	-
25	40.9	1.82, m	26	23, 24, 26, 27, 36
26	27.9	1.91, m	25, 27, 37	24, 25, 27, 28
27	125.6	4.99, m	26, 37, 38	25, 26, 37, 38
28	132.1	-	-	-
29	17.9	1.58, bs	2, 30	1, 2, 30
30	26.1	1.63, d (1.06)	2, 3, 29	1, 2, 29
31	16.4	1.66, bs	6, 7	4, 5, 6
32	19.4	2.05, m	-	8, 10
33	20.6	2.36, s	13	13, 14, 15
34	20.8	2.38, s	16	14, 15, 16
35	16.9	1.88, bs	18, 19	19, 20, 21
36	16.3	1.51, bs	22, 23	23, 24, 25
37	26.0	1.62, d (1.06)	26, 27, 38	27, 28, 38
38	17.9	1.53, bs	27, 37	27, 28, 37
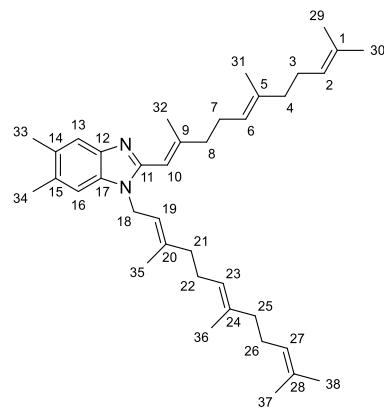
Table S4 NMR spectroscopic data for sandacrabin B in methanol-d_4 at 500/125 MHz.

#	Δ ^{13}C [PPM]	Δ ^{1}H [PPM], MULT (*J* [HZ])	COSY	HMBC
1, 28	132.0	-	-	-
2, 27	125.2	5.01	3/26, 29/37, 30/38	3/26, 4/25, 29/37, 30/38
3, 26	27.6	1.95	2/27, 4/26, 29/37	1/28, 2/27, 4/25, 5/24
4, 25	40.6	1.85	3/26	2/27, 3/26, 5/24, 6/23, 31/38
5, 24	136.5	-	-	-
6, 23	124.5	5.05	7/22, 8/21, 31/36	4/25, 7/22, 8/21, 31/38
7, 22	26.8	2.18	6/23, 8/21, 31/36	6/23, 8/21, 9/20
8, 21	40.3	2.17	7/22	6/23, 7/22, 9/20, 10/19, 32/35
9, 20	145.8	-	-	-
10, 19	117.2	5.49	11/18, 32/35	8/21, 11/18, 32/35
11, 18	45.9	5.07	10/19, 32/35	9/20, 10/19, 12/17, 39
12, 17	131.4	-	-	-
13, 16	114.1	7.64	33/34	12/17, 14/15, 16/13, 33/34
14, 15	138.5	-	-	-
29, 37	25.7	1.64	2/27, 3/26, 30/38	1/28, 2/27, 30/38
30, 38	17.6	1.56	2/27, 29/37	1/28, 2/27, 29/37
31, 36	15.9	1.56	6/23, 7/22	4/25, 5/24, 6/23
32, 35	16.6	1.92	10/19, 11/18	8/21, 9/20, 10/19
33, 34	20.6	2.48	13/16	13/16, 14/15
39	140.2	n.d.	n.d.	n.d.
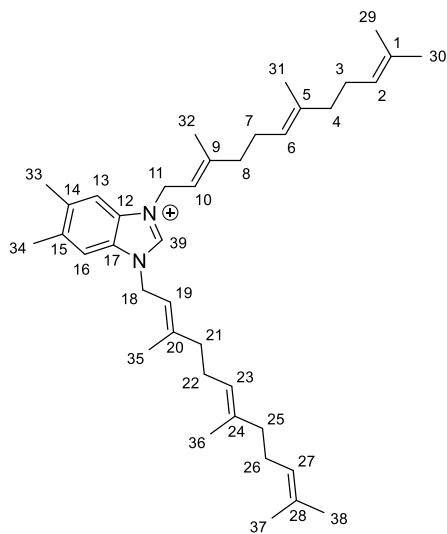
Table S5 NMR spectroscopic data for sandacrabin C in methanol-d_4 at 500/125 MHz.

#	Δ ^{13}C [PPM]	Δ ^{1}H [PPM], MULT (J [HZ])	COSY	HMBC
1, 28	132.2	-	-	-
2, 27	125.4	5.01	3/26, 29/37, 30/38	3/26, 4/25, 29/37, 30/38
3, 26	27.9	1.92	2/27, 4/25, 29/37	1/28, 2/27, 4/25, 5/24
4, 25	41.0	1.85	3/26	2/27, 3/26, 5/25, 6/23, 31/36
5, 24	137.0	-	-	-
6, 23	124.7	5.01	7/22, 31/36	3/26, 4/25, 7/22, 8/12, 31/36
7, 22	27.0	2.16	6/22, 8/21, 31/36	5/24, 6/23, 8/21
8, 21	40.4	2.13	7/22	6/23, 7/22, 9/20, 10/19, 32/35
9, 20	143.9	-	-	-
10, 19	118.1	5.26	11/18, 32/35	8/21, 11/18, 32/35
11, 18	44.8	5.09	10/19, 32/35	9/20, 10/19, 12/17, 32/35, 39
12, 17	131.3	-	-	-
13, 16	113.9	7.58	33/34	12/17, 16/13, 14/15, 33/34
14, 15	138.0	-	-	-
29, 37	17.9	1.54	2/27, 3/26, 30/38	1/28, 2/27, 30/38
30, 38	26.0	1.64	2/27, 29/37	1/28, 2/27, 29/37
31, 36	16.2	1.54	6/23, 7/22	4/25, 5/24, 6/23
32, 35	16.9	1.95	10/19, 11/18	8/21, 9/20, 10/19, 11/18
33, 34	20.7	2.46	13/16	13/16, 14/15
39	150.5	-	-	-
40	10.4	2.81	-	12/17
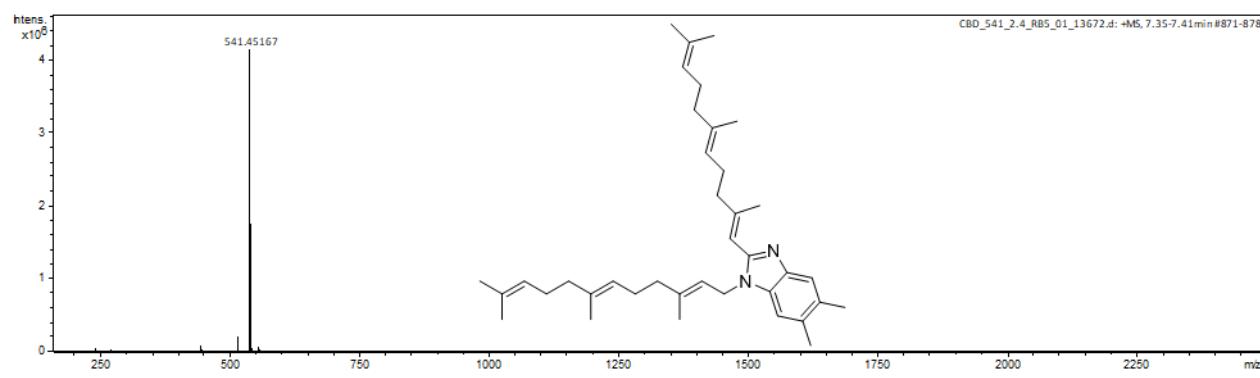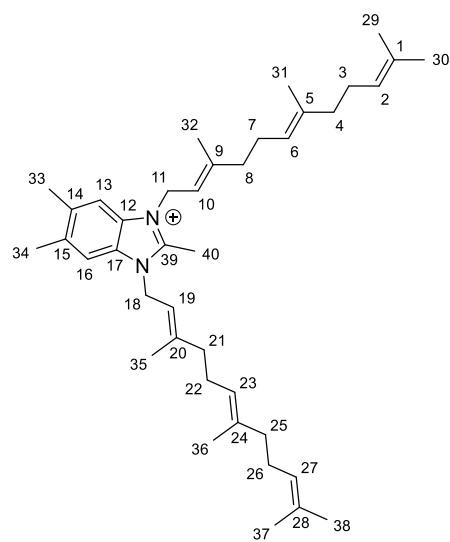
Figure S1 *hr*ESI-MS spectrum of sandacrabin A.
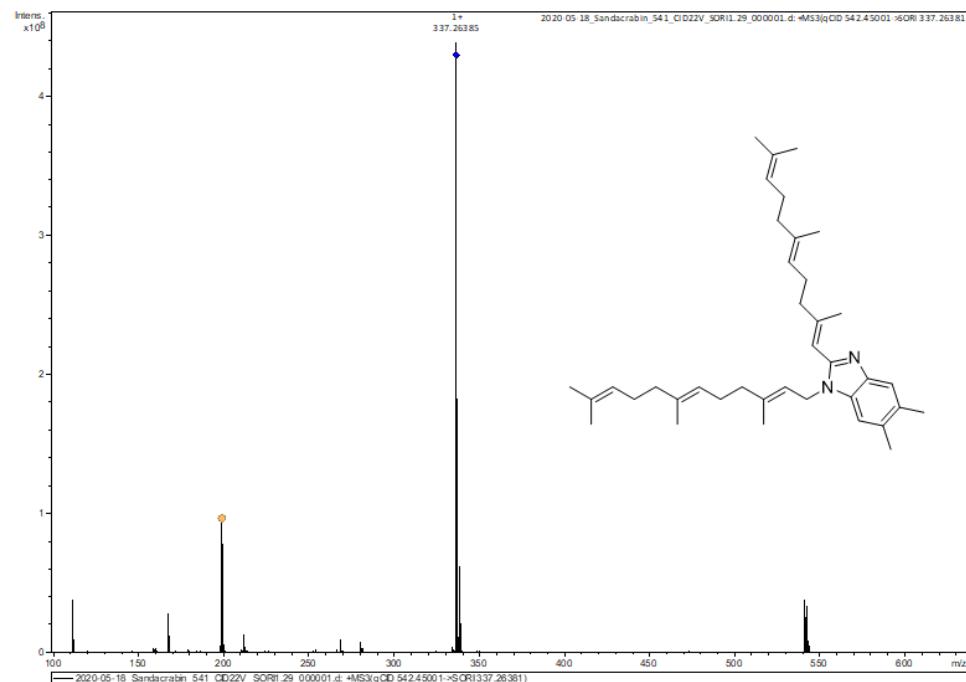
Figure S2 *hr*ESI-MS^3 spectrum of sandacrabin A.
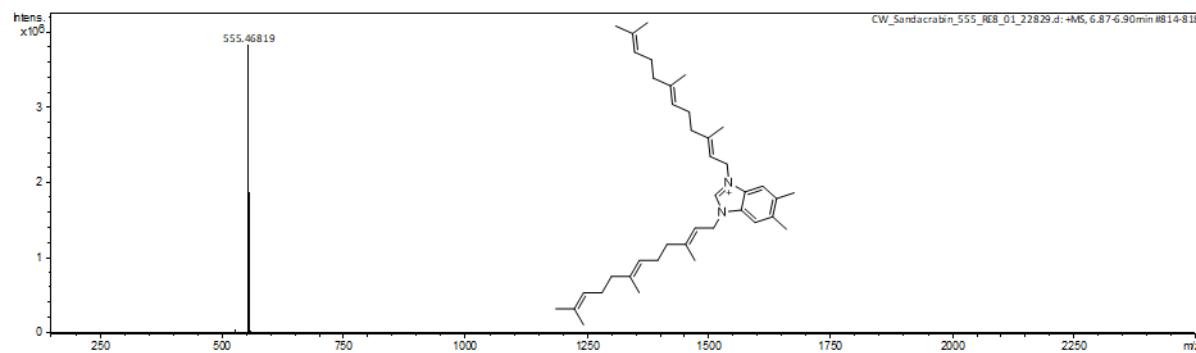
Figure S3 *hr*ESI-MS spectrum of sandacrabin B.
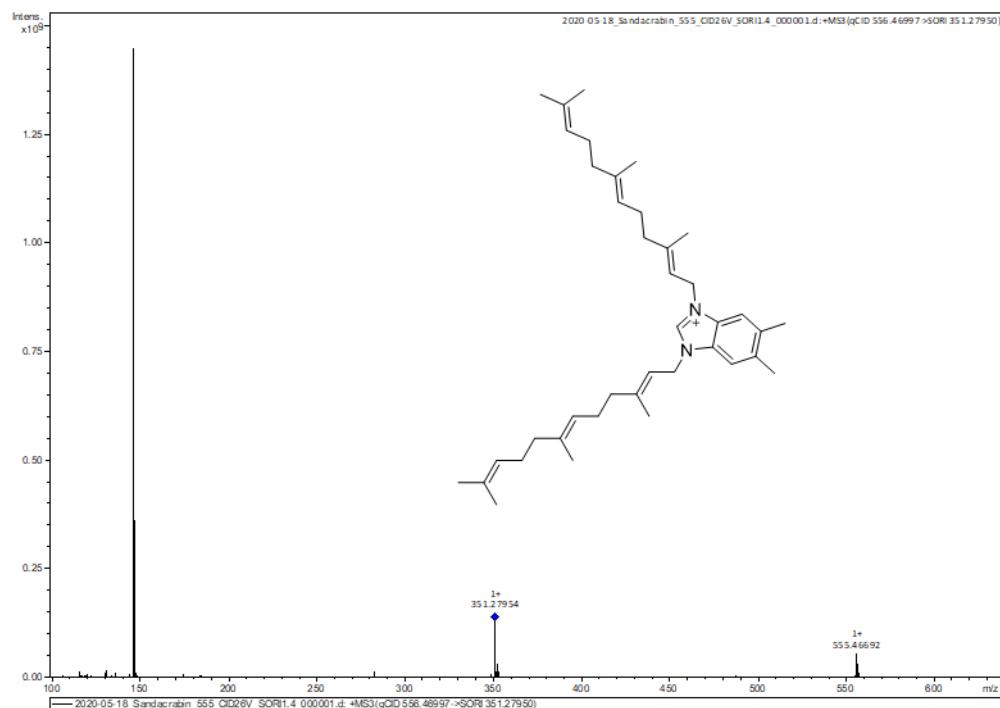
Figure S4 *hr*ESI-MS^3 spectrum of sandacrabin B.
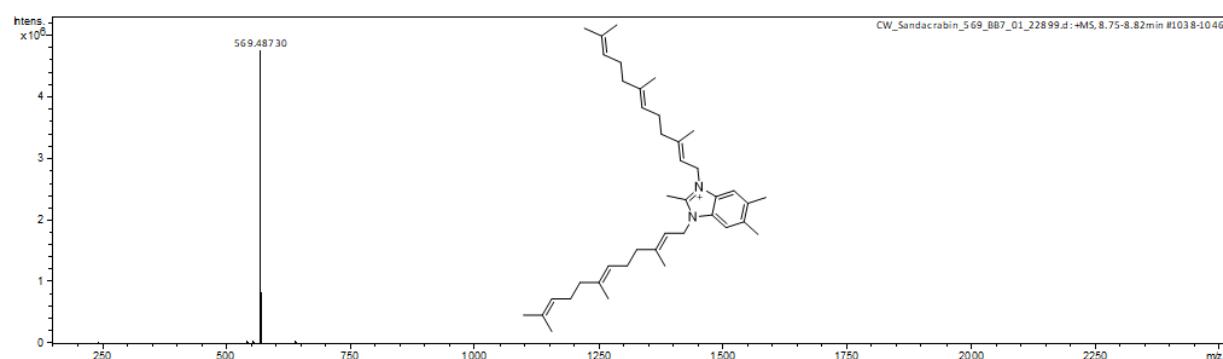
Figure S5 *hr*ESI-MS spectrum of sandacrabin C.
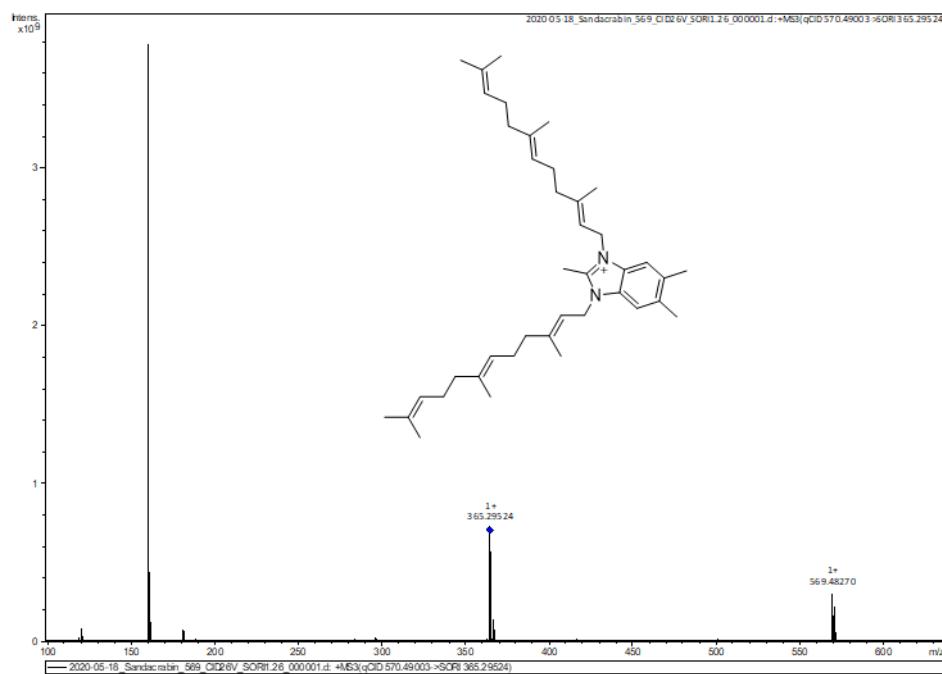
Figure S6 *hr*ESI-MS³ spectrum of sandacrabin C.

- Mono sandacrabin B
- Mono sandacrabin C
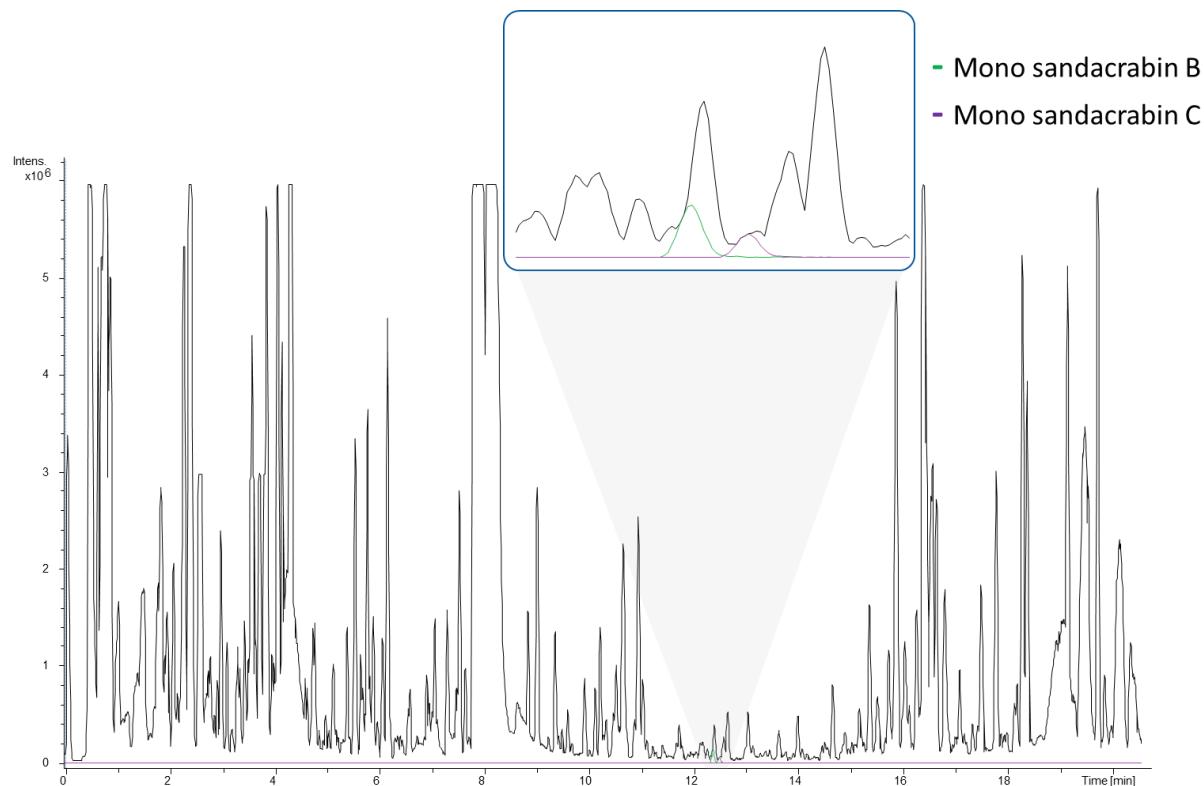
Figure S7 LC-MS analysis of MSr10575 crude extract. Base peak chromatogram (BPC, black trace), extracted ion chromatograms (EIC) for monofarnesylated sandacrabin B (green trace) at 351.2790 ±0.005 Da and monofarnesylated sandacrabin C (violet trace) at 365.2951 ±0.005 Da.

S4.6 Chemical synthesis of the sandacrabins

S4.6.1 General experimental procedures

General procedure A: monofarnesylation

Mono-farnesylation of the respective benzimidazole core structures was achieved adapted from the procedure described by Kishida et al..[5] The benzimidazole (1.0 equiv) in DMF was added dropwise to a mixture of DMF and a 90% oil suspension of NaH (1.0 equiv) at room temperature. The resulting mixture was stirred for 1 h, before farnesylbromide (1.0 equiv) in DMF was added dropwise at 5-10 °C. The 0.1 M solution was stirred overnight at room temperature. After complete conversion (LC-MS), the reaction was quenched with water. The mixture was extracted with EtOAc twice, before the combined organic layers were washed with brine and dried over Na_2SO_4. Concentration of the solvent *in vacuo* afforded the desired product as light yellow oils.

General procedure B: bis-farnesylation

Bis-farneylation of the respective benzimidazole core structures was achieved adapted from the procedure described by Lim et al..[1] 2 mL of THF was added to a crimp vial equipped with a magnetic stir bar followed by the addition of the benzimidazole (1 equiv), farnesylbromide (4 equiv) and finally potassium carbonate (1 equiv). This suspension was subsequently heated at 80 °C over night. After cooling down the reaction to room temperature, the solvent was evaporated under reduced pressure. The residue was dissolved in EtOAc and the mixture extracted twice with brine. The combined organic layers were dried over Na_2SO_4, the EtOAc was evaporated under reduced pressure and the residue re-dissolved in methanol before purification by semipreparative HPLC.

2,5,6-trimethyl-1H-benzo[*d*]imidazole (1)

2,5,6-trimethyl-1H-benzo[d]imidazole was synthesized like previously described by Llona-Minguez et al..[2] 4,5-dimethylbenzene-1,2-diamine (1.02 g, 7.36 mmol) was added to 20 mL (73.6 mmol) acetic acid in a crimp vial equipped with a magnetic stir bar. The crimp vial was sealed and heated at 145 °C for 30 min. The reaction mixture was allowed to cool to room temperature and then poured into 40 mL of saturated sodium carbonate. The precipitate was filtered off and washed with water repeatedly. The precipitate was dried *in vacuo* giving the desired benzimidazole as light brown solid in quantitative yields.
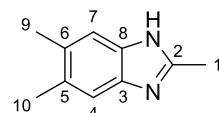
¹H NMR (500 MHz, DMSO-d_6): δ = 2.26 (s, 6H, 9-H, 10-H), 2.43 (s, 3H, 1-H), 7.20 (bs, 2H, 4-H,7-H), 11.96 (bs, 1H, NH)

¹³C NMR (125 MHz, DMSO-d_6): δ = 14.6 (s, 1-C), 19.9 (s, 9-C, 10-C), 110.9 (bs, 7-C), 118.1 (bs, 4-C), 129.2 (bs, 5-C, 6-C), 133.0 (bs, 8-C), 142.2 (bs, 3-C), 150.3 (s, 2-C)

HRMS (ESI):   *m/z* calculated 161.1073 for $C_{10}H_{13}N_2$; found 161.1077 Δ-0.37 ppm.

IR (neat, cm^{-1}): ν = 2979, 2920, 1452, 1432, 1380, 1307, 1103, 1020, 1001, 867, 854, 806, 678, 654
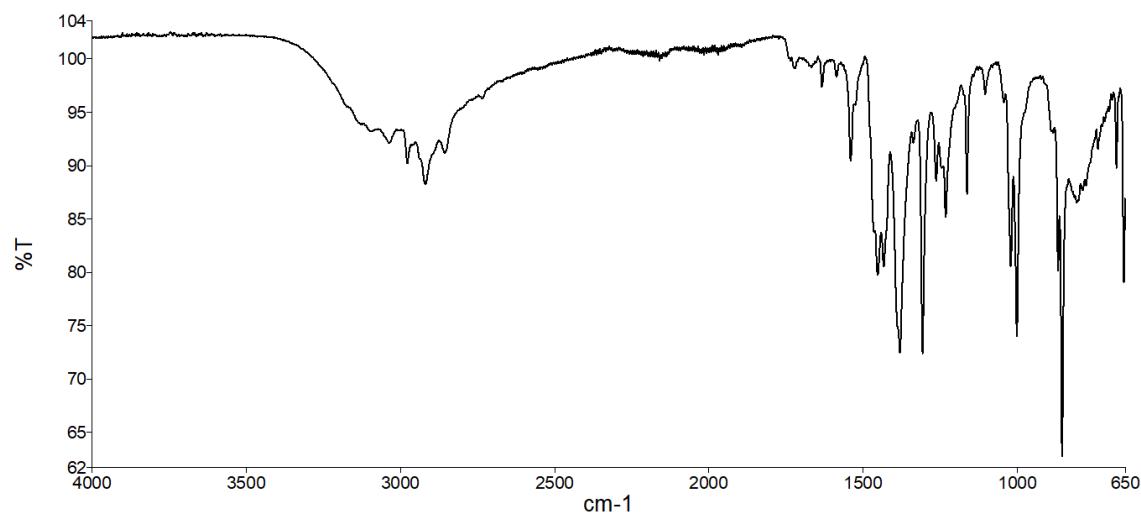
Figure S8 IR spectrum of **1**.

5,6-dimethyl-1-((2E,6E)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1H-benzo[d]imidazole (2)

According to **GP A**, 5,6-dimethyl-1H-benzo[d]imidazole (26.3 mg, 0.18 mmol) was reacted with farnesylbromide (50.0 mg, 0.18 mmol) utilizing NaH (4.1 mg, 0.18 mmol) in 1.8 mL DMF. The reaction resulted in 5,6-dimethyl-1-((2E,6E)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1H-benzo[d]imidazole (**2**) in quantitative yields as a pale yellow oil.
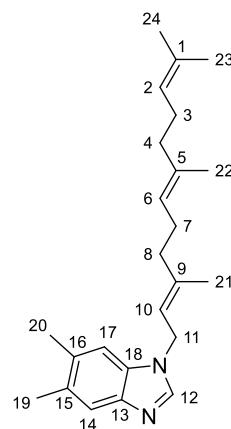
HRMS (ESI):   m/z calculated 351.2795 for $C_{24}H_{35}N_2$; found 351.2795 Δ -0.01 ppm.

UHPLC (18min):   Rt = 11.98 min

Table S6 NMR spectroscopic data for 2 in methanol-d_4 at 700/175 MHz.

#	Δ ^{13}C [PPM]	Δ ^{1}H [PPM], MULT (*J* [HZ])	COSY	HMBC
1	132.1	-	-	-
2	125.6	4.99	3, 23	3, 4, 23, 24
3	27.8	1.93	2, 4	1, 2, 4, 5
4	40.9	1.84	3	2, 3, 5, 6, 22
5	136.6	-	-	-
6	124.9	5.00	7, 22	4, 7, 8, 22
7	27.1	2.11	6, 8	4, 5, 6, 8, 9, 22
8	40.5	2.07	7	6, 7, 9, 10, 21
9	142.4	-	-	-
10	120.2	5.34	11, 21	8, 11, 21
11	43.9	5.37	10, 21	8, 9, 10, 12, 13, 18, 21
12	143.4	7.93	-	13, 18
13	142.9	-	-	-
14	120.2	7.41	19	16, 18, 19
15	132.6	-	-	-
16	133.5	-	-	-
17	111.8	7.19	20	13, 15, 20
18	133.5	-	-	-
19	20.5	2.32	14	14, 15, 16
20	20.9	2.35	17	15, 16, 17
21	16.7	1.83	10, 11	8, 9, 10, 11
22	16.3	1.53	6	4, 5, 6
23	26.1	1.61	2	1, 2, 24
24	17.9	1.52	-	1, 2, 23

2,5,6-trimethyl-1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (3)

According to **GP A**, 2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (28.8 mg, 0.18 mmol) was reacted with farnesylbromide (50.0 mg, 0.18 mmol) utilizing NaH (4.1 mg, 0.18 mmol) in 1.8 mL DMF. The reaction resulted in 2,5,6-trimethyl-1-((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**3**) in quantitative yields as a pale orange oil.
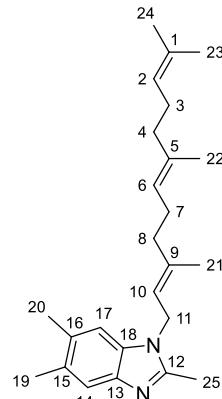
HRMS (ESI):   m/z calculated 365.2951 for $C_{25}H_{37}N_2$; found 365.2957 Δ -1.45 ppm.

UHPLC (18min):   Rt = 12.06 min

Table S7 NMR spectroscopic data for 3 in methanol-d_4 at 700/175 MHz.

#	Δ ^{13}C [PPM]	Δ ^{1}H [PPM], MULT (*J* [HZ])	COSY	HMBC
1	132.2	-	-	-
2	125.5	4.99	3, 23	3, 4, 23, 24
3	27.9	1.91	2, 4, 23, 24	1, 2, 4, 5
4	40.9	1.83	3	2, 3, 5, 6, 22
5	136.6	-	-	-
6	124.9	5.00	7, 22	4, 7, 8, 22
7	27.1	2.10	6, 8, 22	5, 6, 8, 9
8	40.5	2.06	7	6, 7, 9, 10, 21
9	140.9	-	-	-
10	120.5	5.14	11, 21	8, 11, 21
11	42.9	4.72	10, 21	9, 10, 12, 13, 18
12	152.2	-	-	-
13	141.3	-	-	-
14	119.1	7.30	19	16, 18, 19
15	132.1	-	-	-
16	132.6	-	-	-
17	111.3	7.12	20	13, 14, 15, 20
18	134.7	-	-	-
19	20.5	2.32	14	14, 15, 16
20	20.8	2.35	17	15, 16, 17
21	16.8	1.87	10, 11	8, 9, 10
22	16.3	1.52	6, 7	4, 5, 6
23	26.0	1.62	2, 3, 24	1, 2, 3, 24
24	17.9	1.53	2, 3, 23	1, 2, 23
25	13.6	2.51	-	

5,6-dimethyl-1,3-bis((2E,6E)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1H-benzo[d]imidazolium (4)

According to **GP B**, 5,6-trimethyl-1H-benzo[d]imidazole (6.6 mg, 0.05 mmol) was reacted with farnesylbromide (50.0 mg, 0.18 mmol) utilizing K_2CO_3 (6.2 mg, 0.05 mmol) in 2.0 mL THF. The reaction resulted in 5,6-dimethyl-1,3-bis((2E,6E)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1H-benzo[d]imidazoleium (**4**) (14.0 mg, 0.04 mmol, 80%) as a pale yellow oil.
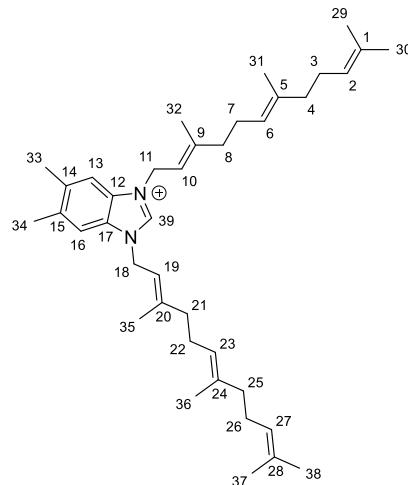
HRMS (ESI):   m/z calculated 555.4673 for $C_{39}H_{59}N_2$; found 555.4669 Δ0.76ppm

UHPLC (9min):   Rt = 8.76 min

Table S8 NMR spectroscopic data for 4 in methanol-d_4 at 500/125 MHz.

#	Δ ^{13}C [PPM]	Δ ^{1}H [PPM], MULT (J [HZ])	COSY	HMBC
1, 28	132.3	-	-	-
2, 27	125.5	5.01	3/26, 29/37, 30/38	3/26, 4/25, 29/37, 30/38
3, 26	27.9	1.95	2/27, 4/26, 29/37	1/28, 2/27, 4/25, 5/24
4, 25	40.9	1.85	3/26	2/27, 3/26, 5/24, 6/23, 31/38
5, 24	136.9	-	-	-
6, 23	124.7	5.06	7/22, 8/21, 31/36	4/25, 7/22, 8/21, 31/38
7, 22	27.1	2.19	6/23, 8/21, 31/36	5/24, 6/23, 8/21, 9/20
8, 21	40.6	2.17	7/22	6/23, 7/22, 9/20, 10/19, 32/35
9, 20	146.1	-	-	-
10, 19	117.5	5.49	11/18, 32/35	8/21, 11/18, 32/35
11, 18	46.3	5.08	10/19, 32/35	9/20, 10/19, 12/17, 39
12, 17	131.7	-	-	-
13, 16	114.5	7.64	33/34	12/17, 14/15, 16/13, 33/34
14, 15	138.8	-	-	-
29, 37	26.1	1.64	2/27, 3/26, 30/38	1/28, 2/27, 30/38
30, 38	17.9	1.56	2/27, 29/37	1/28, 2/27, 29/37
31, 36	16.3	1.56	6/23, 7/22	4/25, 5/24, 6/23
32, 35	16.9	1.93	10/19, 11/18	8/21, 9/20, 10/19
33, 34	20.8	2.48	13/16	13/16, 14/15
39	140.9	9.27	-	-

2,5,6-trimethyl-1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazol-3-ium (5)

According to **GP B**, 2,5,6-trimethyl-1*H*-benzo[*d*]imidazole (**1**) (7.3 mg, 0.05 mmol) was reacted with farnesylbromide (50.0 mg, 0.18 mmol) utilizing K_2CO_3 (6.2 mg, 0.05 mmol) in 2.0 mL THF. The reaction resulted in 2,5,6-dimethyl-1,3-bis((2*E*,6*E*)-3,7,11-trimethyldodeca-2,6,10-trien-1-yl)-1*H*-benzo[*d*]imidazole (**5**) (13.7 mg, 0.04 mmol, 75%) as a pale orange oil.
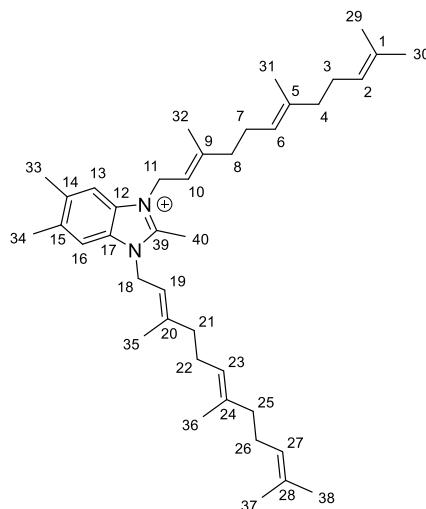
HRMS (ESI):   m/z calculated 569.4829 for $C_{40}H_{61}N_2$; found 569.4822.

UHPLC (9min):   Rt = 8.79 min

Table S9 NMR spectroscopic data for 5 in methanol-*d₄* at 500/125 MHz.

#	Δ ^{13}C [PPM]	Δ ^{1}H [PPM], MULT (*J* [HZ])	COSY	HMBC
1, 28	132.4	-	-	-
2, 27	125.4	5.00	3/26, 30/38	3/26, 4/25, 29/37, 30/38
3, 26	27.9	1.93	2/27, 4/25, 29/37, 30/38	1/28, 2/27, 4/25, 5/24
4, 25	40.9	1.84	3/26	2/27, 3/26, 5/25, 6/23
5, 24	136.8	-	-	-
6, 23	124.7	5.02	7/22, 31/36	4/25, 7/22, 8/12, 31/36
7, 22	27.0	2.14	6/22, 8/21	5/24, 6/23, 8/21, 9/20
8, 21	40.5	2.13	7/22, 10/19, 32/35	6/23, 7/22, 9/20, 10/19, 32/35
9, 20	143.9	-	-	-
10, 19	118.0	5.26	8/21, 11/18	8/21, 11/18, 32/35
11, 18	44.9	5.09	10/19, 32/35	9/20, 10/19, 12/17, 32/35, 39
12, 17	131.1	-	-	-
13, 16	113.9	7.58	33/34	12/17, 16/13, 14/15, 33/34
14, 15	138.0	-	-	-
29, 37	17.9	1.54	2/27, 3/26, 30/38	1/28, 2/27, 30/38
30, 38	26.1	1.64	3/26, 29/37	1/28, 2/27, 29/37
31, 36	16.3	1.54	6/23	4/25, 5/24, 6/23

32, 35	16.9	1.95	8/21, 11/18	8/21, 9/20, 10/19, 11/18
33, 34	20.8	2.46	13/16	13/16, 14/15
39	150.5	-	-	-
40	10.8	2.85	-	12/17, 39

S4.7 Biological assays

S4.7.1 MTT assay

Human HCT-116 colon carcinoma cells (ACC-581) were bought from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ) and were cultured under the conditions recommended by the depositor. To determine the cytotoxic activities of the sandacrabins, cells from actively growing cultures were harvested and seeded at 5 x 10^4 cells per well in a 96 CELLBind® surface well plate in 120 µL 90% modified McCoy's 5A medium with 10% h.i. fetal bovine serum (FBS). After 2 h of equilibration, the cells were treated with the compounds in a serial dilution. After 5 days of incubation at 37 °C, 20 µL of 5 mg/mL thiazolyl blue tetrazolium bromide (MTT) in PBS was added. After discarding the medium, 100 µL of a 2-propanol 10 N HCl mixture (250:1) was added to dissolve formazan granules. A microplate reader (EL808, Bio-Tek Instruments Inc.) was used to determine the absorbance at 570 nm.

S4.7.2 Antimicrobial activities

All microorganisms used for the biological assays were obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ) or were part of our in-house strain collection and were cultured under the conditions recommended by the depositor. Bacterial cultures were prepared in MHB medium inoculated from the strain grown on agar plate (2.9 g/L beef infusion solids, 17.5 g/L casein hydrolysate, 1.5 g/L starch at pH 7.4).The compounds were diluted serially in sterile 96 well-plates before adding the bacterial cell suspension. The bacteria were grown for 24 h at 37 °C. Growth inhibition was inspected visually. MIC_{50} values were determined relative to the respective control samples by sigmoidal curve fitting.

S4.7.3 Antiviral activity

Huh-7.5 cells constitutively expressing a firefly luciferase reporter gene were infected with a renilla luciferase CoV229E reporter virus in the presence of 10 µM compound. 48h post inoculation, cells were lysed in PBS/0.5% Triton-X100 and infection efficiency was determined by renilla luciferase assay, whereas cell survival was analyzed by firefly luciferase assay.

S4.7.4 Induction of the lysosomal apoptosis pathway

Lysosomal apoptosis induction was determined by DNA ladder assay adapted from Saadat *et al..*[3] Human HCT-116 colon carcinoma cells (ACC-581) were cultivated in a 6 well plate containing 2 mL 90% modified McCoy's 5A medium with 10% h.i. fetal bovine serum (FBS) each. After 48 h of equilibration, the cells were treated with the compounds in MeOH at 2 x MIC. As negative control pure MeOH and as positive control 500 µM H_2O_2 was used. After 24h incubation at 37 °C, floating apoptotic cells were harvested by pipetting the supernatant into tubes and centrifuging the supernatant at 5000 rpm for 5 minutes. The supernatant was discarded and the empty wells containing adherent cells were incubated with 500 µL lysis buffer. After 10 minutes of incubation at room temperature, the lysate cells from the wells were pipetted into the respective tubes. The combined cells in lysis buffer were incubated at 65

°C for 5 minutes. After cooling to room temperature, 700 μL chloroform-isoamyl alcohol was added and centrifuged at 12000 rpm for 5 minutes. The aquatic upper phase was transferred into fresh tubes and 100 μL of 3.9M $(NH_4)_2SO_4$ solution was added to precipitate residual protein and centrifuged at 12000 rpm for 5 minutes. The pellet is discarded and 800μL of Isopropanol at -20°C were added to the supernatant in a fresh tube to precipitate the DNA. The tubes were inverted gently before another centrifugation step at 12000 rpm for 5 minutes. The supernatant was discarded and the pellet air-dried for 30 minutes. The dry DNA was dissolved in 50 μL distilled water. The samples were separated by electrophoresis on a 1.5% agarose gel containing 1 μL/100 mL Roti-Safe DNA gel stain (Roth). The gel was examined by ultraviolet gel documentation.

Figure S9. Gel documentation picture of the apoptosis assay, pos indicating H_2O_2 positive control, neg indicating negative control, x belonging to assays not related to sandacrabins that were not cropped to show integrity of the agarose gel and lanes 1 and 2 belonging to mono sandacrabin B. Lad refers to the 1kb DNA ladder purchased from Thermo scientific.

S.4.7.5 Insecticidal activity

Acyrthosiphon pisum were obtained from DendroShop Terraristik and maintained as non-treated culture on pea sprouts prior to retrieval for the insecticidal assays. Insecticidal activity against *A. pisum* was determined as direct spray insecticidal activity adapted from the procedure described by Ahumanda *et al.*.[4] 0.5 microliter of insecticide solution in acetone were applied with a transfer pipette (BRAND, Germany) directly onto the posterior segment of the adult *A. pisum* in a petri dish. As negative control pure acetone and as positive control 0.01 μg/mL imidacloprid (PESTANAL®, Sigma-Aldrich) in acetone was used. Twenty replicates were tested for each insecticide. The exposed insects were maintained at 20 °C until evaluation of their survival via tactile stimulation after one hour. The acetone control group showed a death rate of 10%, whereas the group treated with imidacloprid showed a 100% death rate.

References

(1) Lim, C.-H.; Ilic, S.; Alherz, A.; Worrell, B. T.; Bacon, S. S.; Hynes, J. T.; Glusac, K. D.; Musgrave, C. B. Benzimidazoles as Metal-Free and Recyclable Hydrides for CO2 Reduction to Formate. *Journal of the American Chemical Society* **2019**, *141*, 272–280.

(2) Llona-Minguez, S.; Höglund, A.; Jacques, S. A.; Johansson, L.; Calderón-Montaño, J. M.; Claesson, M.; Loseva, O.; Valerie, N. C. K.; Lundbäck, T.; Piedrafita, J.*; et al.* Discovery of the First Potent and Selective Inhibitors of Human dCTP Pyrophosphatase 1. *J. Med. Chem.* **2016**, *59*, 1140–1148.

(3) Rahbar Saadat, Y.; Saeidi, N.; Zununi Vahed, S.; Barzegari, A.; Barar, J. An update to DNA ladder assay for apoptosis detection. *BioImpacts : BI* **2015**, *5*, 25–28.

(4) Ahumada, M. I.; Chorbadjian, R. A. Laboratory assays of the insecticidal activity of cyantraniliprole and imidacloprid on Brevicoryne brassicae , Myzus persicae (Hemiptera: Aphididae) and Trialeurodes vaporariorum (Hemiptera: Aleyrodidae) pests species and a biological control agent Chrysoperla defreitasi (Neuroptera: Chrysopidae). *Chil. j. agric. res.* **2019**, *79*, 658–663.

Chapter 5

Characterization of the Sandaramines, Myxobacterial Alkaloid Terpenoids accessible by Supercritical Fluid Chromatography

Manuscript in preparation

Chantal D. Bader[1], Fabian Panter[1], Alexander Kiefer[2], Sibylle Haid [3], Judith Hoffmann[1], Thomas Pietschmann[3] and Rolf Müller[1]

Affiliations

[1]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)

[2]Department Drug Design and Optimization, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI) and Department of Pharmacy, Saarland University Campus E8.1, 66123 Saarbrücken (Germany)

[3] 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), Feodor-Lynen-Str. 7, 30625 Hannover, Germany

Contributions and Acknowledgements

Author's effort:

The author contributed to the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. The author performed extraction and purification of the sandaramines, as well as their structure elucidation. Furthermore, the author conducted the chemical synthesis of the molecules and contributed to conceiving and writing this manuscript.

Contributions by others:

5.1 Abstract

During our screening endeavors searching for bioactive myxobacterial natural products we isolated the sandaramines from *Sandaracinus* strain MSr10575 with supercritical fluid chromatography. Their chemical structure consists of two drimenol moieties, which are linked by homospermidine. Sandaramine B features a quaternary ammonium ion, rarely found in bacterial secondary metabolites. Gene disruption experiments revealed the biosynthetic gene cluster (BGC) of the sandaramines comprising a class II terpene cyclase, responsible for the drimenol formation. Further examination of the BGC led to the development of a bio-inspired total synthesis route for the sandaramines. After formation of the drimenol building block in four steps, a sequence of oxidations allows reductive amination for coupling of the homospermidine linker with different drimenol skeletons in one step. Biological activity profiling revealed moderate antimicrobial activities, good anti-proliferative activity against HCT-116 cells and inhibition of the SARS-CoV2 model virus CoV229E for sandaramine A and C.
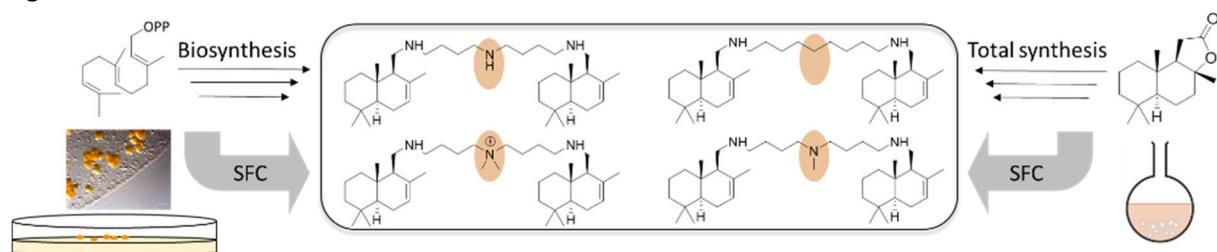
Figure 1 Table of contents graphic showing the principle design of the study.

5.2 Introduction

Recent progression of antimicrobial resistances such as the spread of multidrug resistant tuberculosis and the occurrence of pandemic viral infections caused by pathogens such as SARS-CoV2 have dramatically increased the request for novel antimicrobial and antiviral small molecule drugs.[1,2] Natural products (NPs) have streamlined the discovery workflow of novel drugs to address this demand and defend infectious diseases since ancient times, as they often feature intriguing antibacterial, antifungal and antiviral activities.[3] Traditionally, mainly plant extracts were used to treat a diverse set of diseases ranging from caries to candidosis or influenza.[4–6] Even though they have proven their use in practical use and controlled clinical trials, the present pharmaceutical demand for defined compound-activity correlations increasingly shifted towards the isolation of individual NPs and further development of drug lead candidates based on natural products core structures.[7] This development is reasonable, as a single active ingredient can be characterized more easily and undesirable side-effects can be circumvented.[8] Furthermore, the isolated molecules can then be optimized with regards to pharmacodynamic and -kinetic parameters, which is barely possible for crude extracts.

The immense complexity of bacterial crude extracts, however, requires sophisticated separation techniques for isolation and purification of a single natural product and often more than one purification step.[9] Common NP purification workflows rely on solvent extraction with subsequent crude extract fractionation *via* liquid-liquid partitioning, normal phase chromatography or size exclusion chromatography with a final reversed phase high pressure liquid chromatography (rp-HPLC) purification step to get to the pure NP.[10] Due to the chemical complexity of bacterial NPs, those techniques however occasionally reach their limits. Purification of molecules featuring a permanent charge in aqueous solution which frequently are found to exhibit intriguing antimicrobial activities, for example often show bad separation properties in rp-HPLC.[11] Such compounds are therefore commonly separated by hydrophilic interaction liquid chromatography (HILIC) or ion exchange chromatography. HILIC is however most suitable for the separation of polar or ionic molecules and ion exchange limited to mixtures only containing one ion type. Both techniques fail when it comes to purification of lipophilic NPs featuring a charge in aqueous solution from complex bacterial crude extracts.[12,13] An alternative for those two separation principles is supercritical fluid chromatography (SFC). It is commonly seen as hybrid technique between gas and liquid chromatography, as it uses supercritical CO_2 modified by the addition of an organic co-solvent as eluent.[14] SFC, besides supercritical fluid extraction, which has already been successfully used for myxobacterial extractions (see chapter 2), belongs to the "green" technologies and therefore gained more and more attention in recent years, as it replaces majority of the organic solvents used by CO_2, which can be recycled from industrial processes.[15,16] SFC has furthermore proven to be beneficial for the separation of a variety of non-polar as well as polar molecules.[17] Due to miscibility constraints with supercritical CO_2, SFC separations are performed water-free. It is therefore suitable for the separation of molecules that are charged in aqueous solution. Even though SFC has already been successfully used for the isolation of plant NPs, it still remains highly underrepresented in bacterial NP research.[18]

In this study, we isolated a novel class of myxobacterial terpenoid polyamine NPs called sandaramines by application of supercritical fluid chromatography to myxobacterial extracts. Differences between the biological activities of the two natural sandaramine derivatives as well as investigation of their biosynthetic origin prompted development of a bio-inspired total synthesis, additionally shedding light on the absolute configuration of the terpenoid skeleton. Furthermore, we assessed the antimicrobial, antiproliferative and antiviral activities of the purified natural products and their synthetic analogs and found them to be highly active against HCT-116 cells and the SARS-CoV2 model virus CoV229E.

5.3 Results and discussion

In the context of our search for novel bioactive natural products we started an antimicrobial screening campaign of myxobacterial strains towards their capability to produce to date unknown secondary metabolites. The myxobacterial strain *Sandaracinus* strain MSr10575 gained substantial interest because of its crude extracts' intriguing biological activities against a panel of Gram-positive and negative bacteria, as well as the tuberculosis model strain *Mycobacterium smegmatis,* which could only partly be explained by the two secondary metabolites already described from this strain namely the sandacrabins and indiacens (see chapter 2 and 4).[16]

Two not yet described secondary metabolites, termed sandaramines A (**1**) and B (**2**) could be correlated to the remaining antimicrobial activities. *hr*ESI-MS of sandaramine A (**1**) showed a peak at *m/z* 568.5567 [M+H]$^+$ consistent with the molecular formula $C_{38}H_{69}N_3$ (calc. 568.5564, Δ = 0.48 ppm), whereas *hr*ESI-MS of sandaramine B (**2**) showed an ion peak at *m/z* 596.5879 [M]$^+$ consistent with the molecular formula $C_{40}H_{74}N_3$ (calc. 569.5877, Δ = 0.32 ppm). Both molecules contain six double bond equivalents (DBE). Optimization of the liquid cultivation medium for fermentation in 2 L shake flasks allowed upscaled production and subsequent bioactivity-guided isolation of the compounds by supercritical fluid chromatography. We obtained 2.7 mg of **1** and 1.8 mg of **2** from 6 L cultivation volume. Both sandaramine A and B showed intense tailing and co-elution on rp-HPLC under analytical UHPLC conditions, which impedes separation of the two derivatives by rp-HPLC (see Figure 2D). These two issues could not be solved, even when modifying pH of the eluents by the addition of up to 0.5% triethylamine (TEA) or 0.5% formic acid (FA). We furthermore observe ion suppression effects of sandaramine B to sandaramine A, due to their co-elution. SFC separation in contrary allowed a 300-fold injection volume compared to UHPLC, without losing baseline separation of the two derivatives (see Figure 2). Best separation was achieved by using MeOH with 0.1% TEA as co-solvent on a Torus Diol column. Exact separation conditions for the two measurements can be found in the SI. Figure 2 furthermore highlights the ion suppression in rp-HPLC separations that disappears in the SFC measurements, which is also reflected in the isolated amounts of 1.5 times sandaramine A compared to B. Based on the peak areas on rp-HPLC shown in figure 2D we previously estimated that sandaramine B would be the main derivative produced in highest amounts.

5.3.1 Structure elucidation of the sandaramines

The 1D NMR spectra of **1** in CDCl$_3$ suggested a C$_2$-symmetry of the molecule. The ^{1}H NMR spectrum shows four methyl signals at δ 0.76 (6H, s), 0.86 (6H, s), 0.88 (6H, s), and 1.73 (6H, d, *J* = 0.73 Hz) ppm as well as three methine groups at 1.19 (2H, m), 1.75 (2H, bs) and 5.44 ppm (2H, q, *J* = 0.73 Hz) characteristic for one double bond proton. Three out of nine methylene groups of the molecules were found to be in close proximity to a nitrogen atom due to their characteristic down-field shifts: one diastereotopic methylene group at δ 2.47 (2H, dd, *J* = 12.2, 7.31) and 2.71 (2H, dd, *J* = 12.2, 1.62), one at 2.63 (4H, m) and one at 2.56 ppm (4H, m). A detailed analysis of the 2D NMR data obtained from HSQC, HSQC-TOCSY, HMBC and COSY (see SI) showed that the molecule contains two bicyclic terpene moieties, which could be identified as drimenol core structure, connected by a homospermidine linker (see Figure 2). The homospermidine linker was identified by COSY correlations of the two methylene groups at δ 2.64 and 2.71 ppm to two additional methylene groups at δ 1.44 and 1.94 ppm. The position of the linker was verified by HMBC correlations of the diastereotopic methylene group at 2.47 and 2.71 ppm to the methylene group at 2.56 ppm and HMBC correlations between the proton at 2.63 ppm to its corresponding carbon at 49.7 ppm.
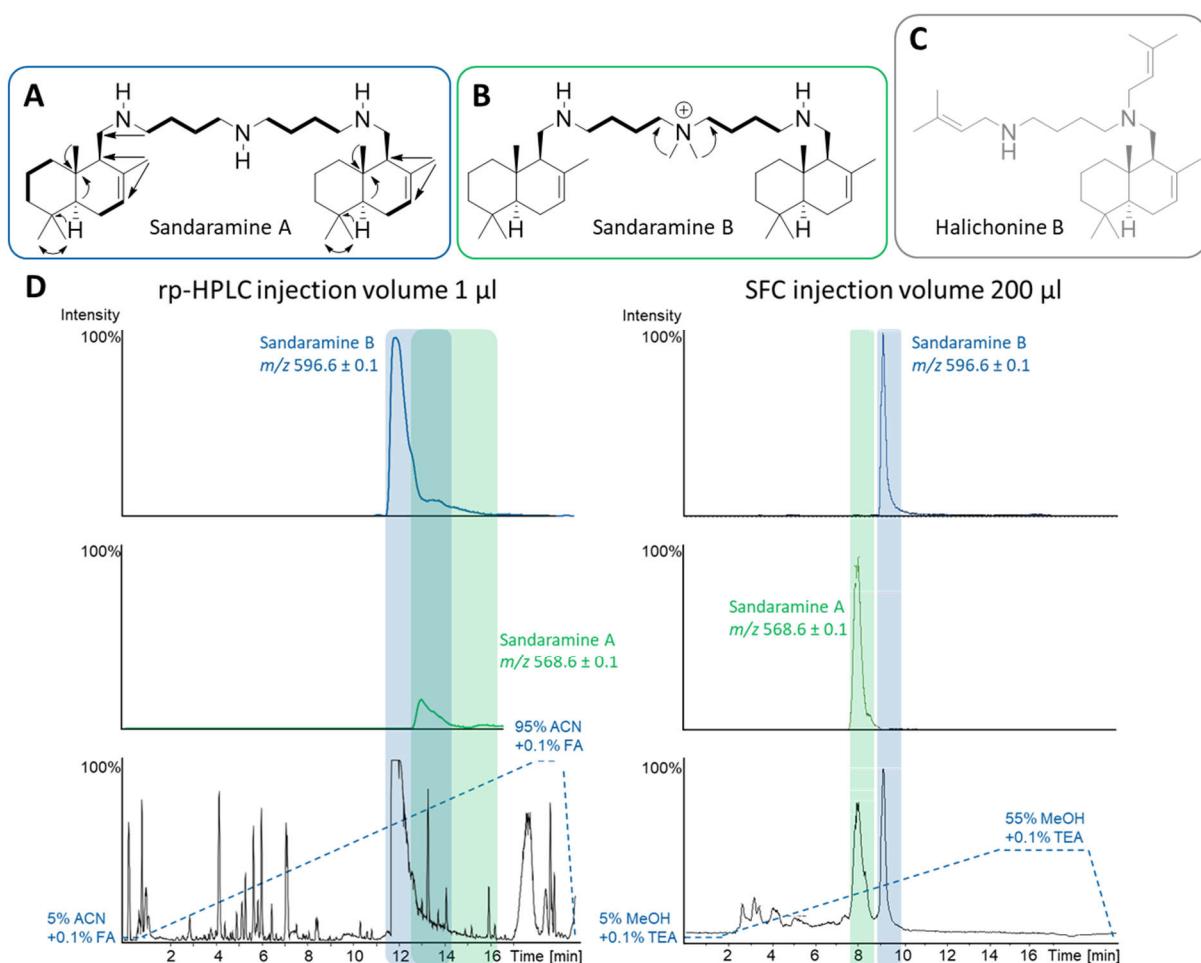
Figure 2 A and **B** Structure formulae of sandaramine A and B elucidated by NMR spectroscopy displaying key NMR correlations. COSY correlations are marked bold and HMBC correlations as arrows. **C** Halichonine B as their most similar natural product described. **D** Analysis of an MSr10575 crude extract with SFC and rp-HPLC. Blue trace: extracted ion chromatogram (EIC) sandaramine A, green trace: sandaramine B, black trace: base peak chromatogram (BPC).

*hr*ESI-MS of sandaramine B (**2**) showed an ion peak at *m/z* 596.5879 [M]$^+$ consistent with the molecular formula $C_{40}H_{74}N_3$ (calc. 569.5877, Δ = 0.32 ppm) containing six DBE. 1D and 2D NMR spectra revealed a C_2-symmetry and two terpene moieties linked by homospermidine in sandaramine B (**2**) comparable to sandaramine A. The central nitrogen in sandaramine B (**2**) however, is substituted by two additional methyl groups at δ 3.50 ppm, which influences the chemical shift of the two neighboring methylene groups. All other NMR signals only show a proton chemical shift deviation below 0.03 ppm when comparing the chemical shifts in sandaramine A to the respective chemical shifts in sandaramine B.

Although sandaramines represent new NP scaffolds, there are some similarities to known compounds. The natural product family that is structurally most similar to the sandaramines are the halichonines, marine sponge-derived terpenoid alkaloids (see Figure 2C).[19] However, among the three natural halichonine derivatives only halichonine B incorporates drimenol and none of the halichonines shows a symmetry axis nor a permanent positive charge. Instead of homospermidine, the drimenol unit in halichonine B furthermore is coupled to prenylated putrescine.[20] The biosynthetic machinery responsible for halichonine formation has not been described yet, which additionally increased our interest in elucidating the sandaramine biosynthesis.

5.3.2 Biosynthesis of the sandaramines

Based in the structure formula of the sandaramines, they seem to employ two distinct types of biosynthetic machineries for their biosynthesis (Figure 3). The bicyclic drimenol scaffold clearly derives from terpene biosynthesis, while the central sym-homospermidine part is biosynthesized from putrescine using polyamine biosynthesis enzymes. All genes involved in sandaramine biosynthesis except for the methyl transferase converting sandramine A to sandaramine B are encoded in two biosynthetic loci. The Terpene cluster locus spanning the genes *sanA* to *sanI* is responsible for drimenol biosynthesis as well as coupling drimenol to the sym-homospermidine, while the homospermidine synthase locus spanning the genes *hssI* to *hssIII* is responsible for formation of the sym-homospermidine precursor (Figure 3).

The drimenol biosynthesis is well-characterized in plants such as *Verbena officinalis*, where a class II terpene synthase was found to be responsible for its formation.[21] Furthermore, bacterial SHC proteins have already been shown to be able to perform drimenol biosynthesis. *In-vitro* experiments, incubating an SHC from *Alicyclobacillus acidocaldarius* with farnesol, verify drimenol to be a component of the bacterial secondary metabolism.[22] The drimenol biosynthesis starts from a farnesyl pyrophosphate precursor that is cyclized to the bicyclic ring system by the class II terpene cyclase in a two-step reaction cascade.[23,24] Mechanistically, these class II terpene cyclases of the squalene/hopene cyclase (SHC) type do not hydrolyze pyrophosphate bonds, but rather protonate a terminal double bond to form a tertiary cationic carbon species. This reactive cation subsequently triggers two cyclisation reactions by intramolecular electrophilic addition to form the bicyclic drimenol pyrophosphate core (see Figure 3).[25] After hydrolysis of the pyrophosphate, the bicyclic alcohol namely drimenol is formed. This substrate that can be further oxidized to the corresponding aldehyde drimenal using CYP P450 type enzymes. This reaction sequence was already hypothesized in the polygodial biosynthesis pathway of the plant *Persicaria hydropiper*.[26]

To assess if the same biosynthesis machinery responsible for sandaramine biosynthesis, we analyzed the genome of *Sandaracinus* strain MSr10575 for the presence of a class II terpene cyclase of SHC type. The only gene encoding such an biosynthesis enzyme was found clustered with a cytochrome P450 type enzyme. This biosynthetic gene cluster is thus likely to be responsible for both drimenol biosynthesis and its subsequent oxidation to drimenal by the cytochrome P450 enzyme. [26]

The linker between the two drimenol moieties in the molecule derives from sym-homospermidine, a biogenic amine that is known to be biosynthesized from spermidine and putrescine. Spermidine is a common secondary metabolite in bacteria that is biosynthesized from methionine and putrescine *via* spermidine synthases. [27] Contrary to spermidine synthases, the sym-homospermidine biosynthetic machinery is more rarely found in bacteria. Biosynthetic enzymes involved in sym-homospermidine production evolved from deoxyhypusine biosynthesis genes as evidenced by plant alkaloid biosynthesis genes.[28] A deoxyhypusine biosynthesis gene responsible for sym-homospermidine biosynthesis, for which we find an homolog in *Sandaracinus* sp. Sr10575, has already been described for the homospermidine lipids from *Myxococcus xanthus*.[29]

To couple the sandaramines' drimenal moieties with the sym-homospermidine linker, the bacterium most likely performs reductive amination reactions. According to our biosynthesis hypothesis, this reductive amination is carried out by a NAD dependent reductase enzymes that reduces the imine bonds that form after condensation of the drimenal to homospermidine to yield the final product. This condensation and reductive amination reaction sequence is reminiscent both of the biosynthesis step to form pipecolic acid as well as cyclisation reactions of homospermidine.[30,31] The biosynthetic transformation creating the quaternary ammonium salt in sandaramine B (**2**) would then be performed

by a SAM–dependent methyl transferase. A sandaramine derivative containing a tertiary amine, which constitutes the intermediate in this transformation, is observed in trace amounts by LC-MS (see SI) and could not be isolated in sufficient yields for structure confirmation by NMR.

Unfortunately, the sandaramine methyl transferase is neither encoded in the locus containing the homospermidine synthase nor the terpene locus containing the drimenol biosynthesis genes. It is therefore hard to pinpoint the location of the gene involved in sandaramine biosynthesis as methyl transferases are rather common in bacterial genomes. We find over 90 methyl transferases in the MSr10575 genome which could be responsible for methylation of sandaramine A (**1**) (see SI). As the methyl transferase located closest to the sandaramine terpene biosynthesis locus is located 20kb upstream of the sandaramine terpene biosynthesis locus, we however were not able to pinpoint the exact enzyme responsible for this transformation.
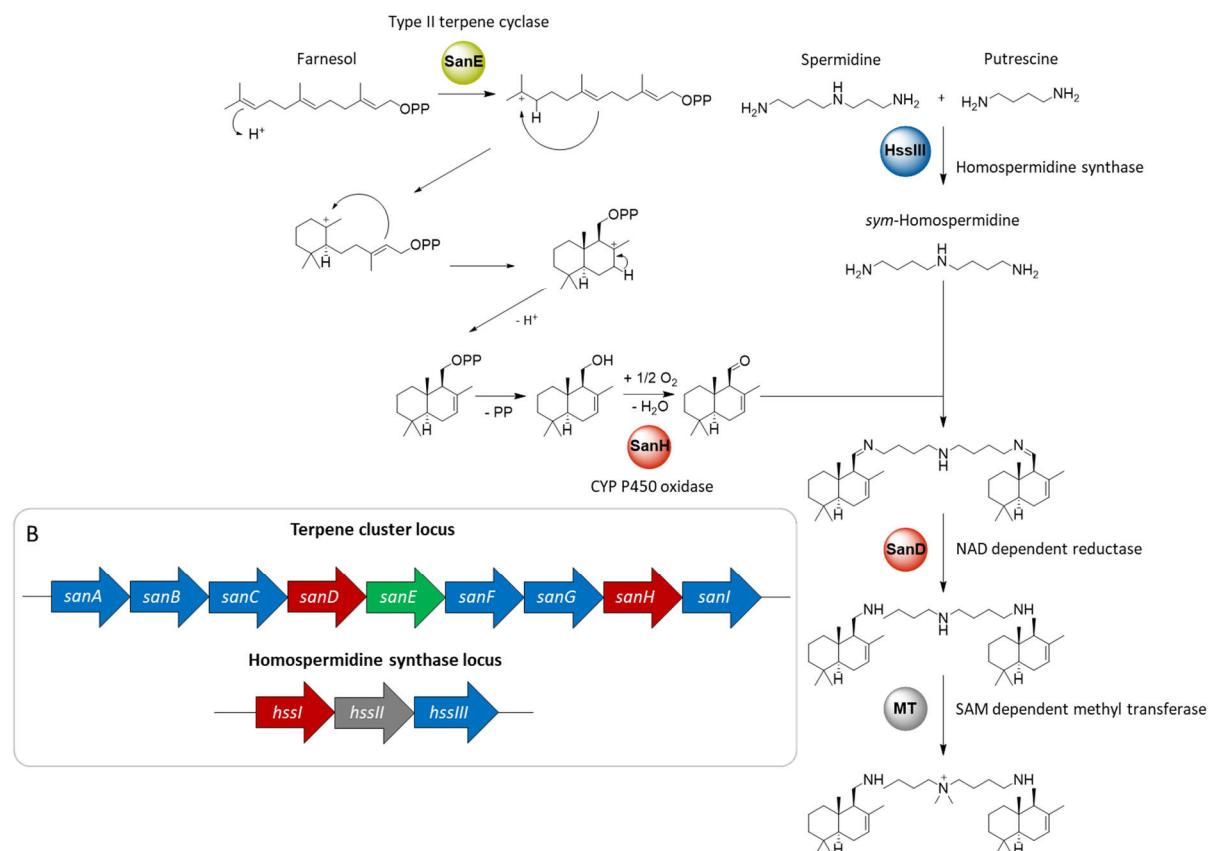
Figure 3 Biosynthesis proposal for the sandaramines. **A** Scheme of the biosynthesis steps and enzymes responsible for sandaramine formation. **B** Biosynthetic genes involved in the sandaramine biosynthesis. Detailed description of the *san and hss* genes involved can be found in the SI.

While the biosynthetic origin of the homospermidine unit was already extensively investigated in the biosynthesis of the homospermidine lipids, biosynthesis of the decalin systems of class II terpene synthase origin as the one we see in drimenol production during sandaramine biosynthesis is evaluated to a lesser extent in bacteria.[29] In order to obtain further evidence for the direct correlation between the terpene gene cluster locus and biosynthesis of the drimenol precursor we sought to inactivate the terpene cluster locus by single crossover inactivation of the *sanE* gene.

We therefore created a pCR2.1 TOPO plasmid derivative harboring a 1 kb DNA fragment from the *sanE* gene for homologous recombination based gene inactivation.[32] The terpene cluster unit is encoded as a single operon and putatively forms a single mRNA transcript. It is thus likely that by targeting and disrupting the class II terpene cyclase gene *sanE* with this single crossover integration we might

inactivate the downstream biosynthesis genes of the operon by a polar effect. Still, as our primary readout is the production of sandaramines and our target strain *Sandaracinus* sp. MSr10575 is inaccessible to genetic manipulation attempts other than single crossover mutagenesis, integrative inactivation of *sanE* to evaluate its effect on sandaramine production is the best way to obtain further evidence of the involvement of this terpene synthase operon in sandaramine biosynthesis. Evaluation of the MSr10575 *sanE*::pCR2.1 *sanE* mutant strains' secondary metabolome by LC-MS revealed the sandaramines to be missing, wherefore we can conclude that this terpene cluster locus is likely involved in sandaramine biosynthesis (see SI).

5.3.3 Total synthesis of the sandaramines

Since crystallization experiments did not lead to sandaramine single crystals, we sought to elucidate the absolute stereochemistry of the sandaramines, *via* ex chiral pool synthesis. Structurally related natural products such as the halichonines were described to incorporate (-)-drimenol.[19] We therefore assumed, that the drimenol moieties in sandaramine A (**1**) and B (**2**) likely feature the same stereochemistry. Development of a total synthesis route for the sandaramines would be useful to confirm this assumption, while providing the possibility to study the influence of various linkers on the sandaramines' biological activities.

We established a bio-inspired chemical synthesis route for the sandaramines, depicted in Figure 4. To allow late stage linker exchange, the synthesis starts with the preparation of (-)-drimenal **9** from (3a*R*)-(+)-sclareolide **5**. First, the lactone **5** is treated with methyllithium at −78 °C, giving quantitatively ketone **6**.[33] The following Baeyer-Villiger-Oxidation led to acetate **7**.[34] Formation of the *endo* double bond was achieved by acid catalyzed dehydration using *p*TsOH at room temperature.[35] Subsequent saponification gave (-)-drimenol **8** in 55 % yield over two steps. Final DMP-oxidation of **8** delivered aldehyde **9**, which can be coupled with/to various diamine linkers.
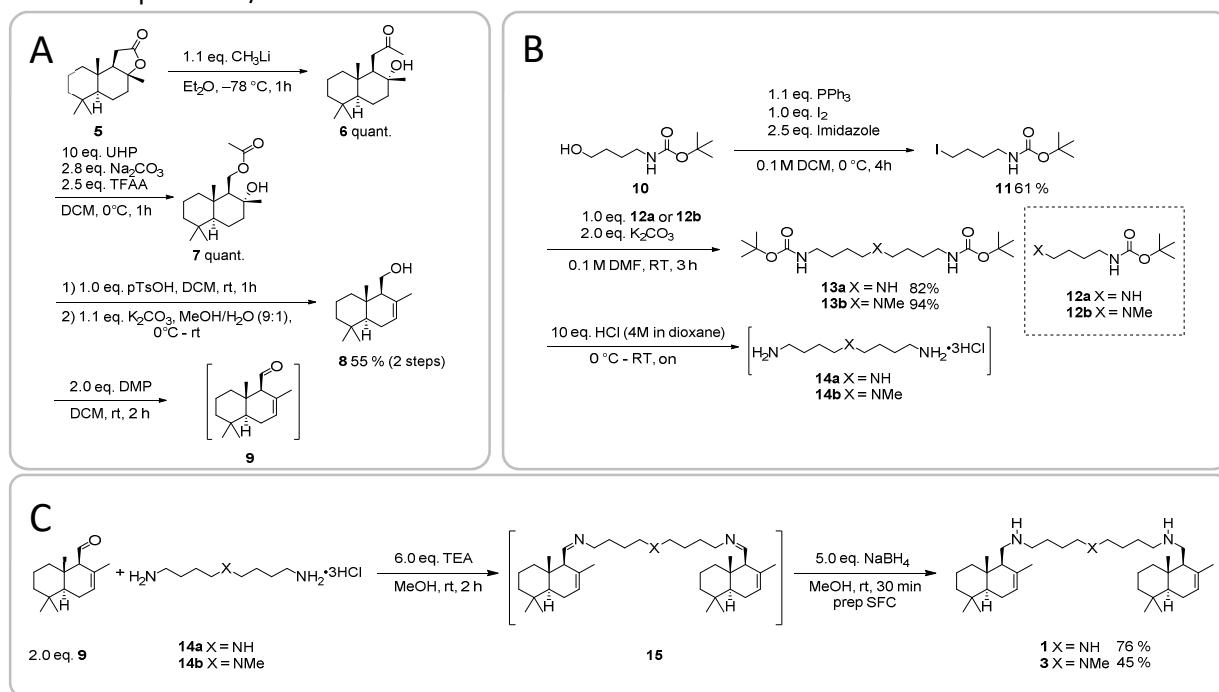
Figure 4 Synthesis route for sandaramine A (**1**), C (**3**) and D (**4**). **A** Synthesis of the drimenal skeleton. **B** Synthesis of the different linkers for synthesis of non-natural sandaramine A (**1**) and C (**3**). **C** Reductive amination reactions.

For synthesis of the secondary amine linker in sandaramine A (**1**), and sandaramine C (**3**), which features a tertiary amine, we firstly converted Boc-protected amino alcohol (**10**) into *tert*-butyl (4-iodobutyl)carbamate (**11**).[36,37] Nucleophilic substitution between **12a**-**b** and iodine **11** under basic conditions, followed by subsequent Boc-deprotection, gave linker **13a** with 82% yield and linker **13b** with 94% yield over two steps.

Final reductive amination was optimized in a two-step process. First, imine formation was carried out under basic conditions in MeOH at ambient temperature. After 2h, sodium borohydride was added to obtain sandaramine A (**1**) and C (**3**). Under these conditions, superior isolated yields of 76% for sandaramine A (**1**) and 45% for sandaramine C (**3**) after preparative SFC purification could be achieved. Comparison between the HSQC spectra of the natural and synthetic sandacrabin A (**1**) shows maximal deviations of $\Delta^1H = 0.06$ and $\Delta^1H = 0.5$ ppm, respectively. This confirms the incorporation of two drimenol moieties in identical configuration. Evaluation of synthetic and natural sandaramine CD spectra gave final proof that they incorporate (-)-drimenol (see SI).

5.3.4 Biological activities of the sandaramines

Minimal inhibitory concentrations (MICs) for sandaramine A (**1**), B (**2**), C (**3**) and D (**4**) were determined against a panel of Gram-positive and Gram-negative bacteria, as well as some fungi including various pathogens (see Table 1). Furthermore, half maximal inhibitory concentrations (IC$_{50}$) against HCT-116 cancer cells and antiviral activity represented by reduction in viral replication in the SARS-CoV2 model virus CoV229E were determined.

Table 1 Antimicrobial and cytotoxic activities of sandaramine A-D (**1-4**) and their respective structures.

	MIC [µg/mL]			
	Sand A	Sand B	Sand C	Sand D
Candida albicans	64±0	64±0	48±16	64±0
Pseudomonas aeruginosa	24±8	48±16	24±8	≥64
Staphylococcus aureus	48±16	48±16	48±16	64±0
Mycobacterium smegmatis	32±0	16±0	32±0	64±0
	IC$_{50}$ [µg/mL]			
HCT-116	0.1±0.1	10.9±3.5	0.4±0.2	1.1±0.3
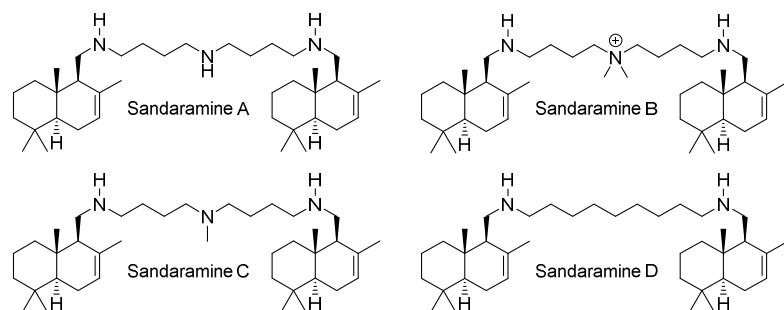
All sandaramines showed weak activity against *Staphylococcus aureus* and *Candida albicans* in the antimicrobial test panel (48-64 µg/mL). Sandaramine D (**4**) showed no biological activity against *Pseudomonas aeruginosa* up to a concentration of 64 µg/mL, whereas sandaramine A (**1**) and C (**3**) both exhibited relatively good biological activities of 24 µg/mL. However, the most promising activity of the sandaramines are their antiproliferative activity as well as their antiviral activity against CoV229E. Sandaramine A (**1**) and C (**3**) showed antiproliferative activities against HCT-116 with IC50 values of

130 ng/mL and 390 ng/mL, respectively, whereas sandaramine D (**1**) and sandaramine B (**2**) display only moderate cytotoxicities (1 µg/mL and 10 µg/mL, respectively). Therefore, the antiproliferative properties of the molecules do not seem to correlate with the antimicrobial activities.

The antiproliferative activity of the structurally related halichonines could already be correlated to an induction of lysosomal apoptosis.[19,20] To investigate if the antiproliferative effects of sandaramine A and C result from lysosomal apoptosis induction as well, we performed a DNA ladder assay (see SI). Cells treated with sandaramines do not show extensive DNA degradation, which suggests that the cytotoxic effects of the sandaramines are not caused by induction of the lysosomal apoptosis pathway but rather by another mechanism.

Additionally to the antiproliferative effects observed for HCT116 cells, sandaramine A (**1**) and C (**3**) exhibited promising reduction in viral replication (see Figure 5). For sandaramine B (**2**) and D (**4**) only a slight reduction in viral replication was detected. The antiviral activity was assessed in Huh-7.5 cells constitutively expressing a firefly luciferase reporter gene. Those cells were infected by a renilla luciferase CoV229E reporter virus. This renilla-firefly luciferase dual assay therefore allows simultaneous assessment of reduction in viral replication and the cell viability. Along with the significant reduction of the viral load at 2 µM for sandaramine A and C, we observe an increase in cell viability of Hu-7.5 cells. This deviation between the effect on HU-7.5 cells and on CoV229E indicates, that there is a therapeutic window for the sandaramines as antiviral drugs.
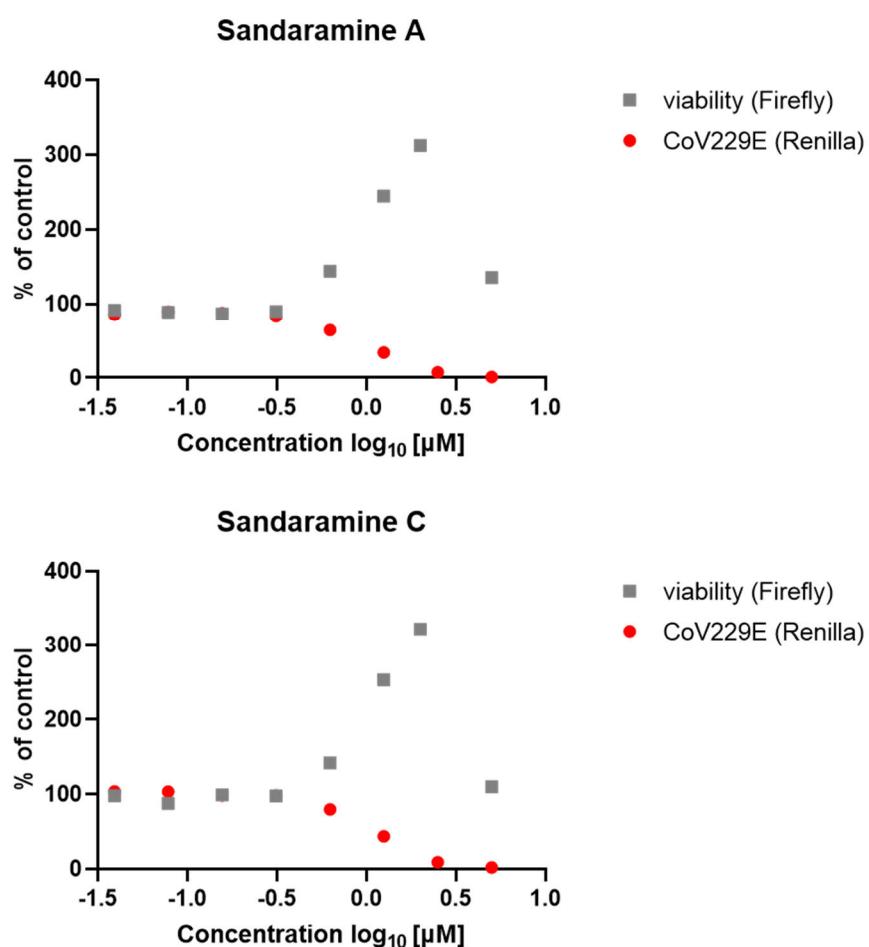
Figure 5 Antiviral activities of sandaramine A (**1**) and C (**3**) against the SARS-CoV2 model CoV229E represented as reduction of viral reproduction (red) in Hu-7.5 cells (grey).

5.4 Summary and outlook

The myxobacterial strain *Sandaracinus* sp. MSr10575 first gained our attention due to strong activities of its bacterial crude extract against a panel of Gram-positive and Gram-negative bacteria, as well as the tuberculosis model strain *Mycobacterium smegmatis*. After subsequent upscaling cultivations, we were able to isolate two natural sandaramine derivatives using supercritical fluid chromatography and correlate them to the activities observed in the crude extract. Structure elucidation by NMR revealed them to be C_2 symmetric alkaloid terpenoid natural products whose structure have not been described yet. Their chemical structure consists of two drimenol moieties that are coupled *via* a homospermidine linker. In sandaramine B the secondary amine in the middle of the homospermidine linker is methylated twice, giving the molecule a permanent positive charge.

The sandaramines are structurally related to the sponge derived halichonines, of which halichonine B features the same terpene core structure. In contrary to the sandaramines, the halichonines were isolated by preparative thin layer chromatography.[19] Terpenoid polyamines are generally hard to purify with commonly used rp-HPLC protocols. The combination of their permanent positive charge in aqueous solution and the high lipophilicity of the terpene core structure impedes efficient separation and leads to strong LC peak tailing. SFC, which we used for the isolation of the sandaramines, is still underrepresented in most microbial natural products isolation workflows.[10] When compared to preparative TLC, SFC purifications are highly automated and the number of theoretical plates as well as the separation efficiency can be improved by the use of a variety of chromatographic columns.[14] We highlighted in this study that this technique offers good separation even for permanently charged natural products.

After isolation of the two natural sandaramines, we were furthermore interested in their biosynthetic origin. The sandaramine chemical structure suggests, that two different types of biosynthetic machinery are involved in their formation. The key biosynthetic steps are performed by *sanE* and *hssIII* coding for a type II terpene cyclase (*sanE*) and a homospermidine synthase (*hssIII*), respectively. Inactivation of the respective terpene cyclase *sanE* provided evidence for this hypothesis, as it led to complete loss in sandaramine production by the sandaramine producer strain MSr10575. This terpene cyclase is responsible for cyclisation of farnesol to drimenol, which (after its oxidation to drimenal) is coupled to the homospermidine linker *via* reductive amination. Our biosynthesis hypothesis for the sandaramine suggests, that the halichonines, which have formerly been isolated from the marine sponge *Halichondria okada* either derive from a similar biosynthesis machinery or are potentially also bacterial natural products produced by a sponge symbiont.[19] This assumption is supported by a metagenomics analysis of this sponge, which has already shown, that other secondary metabolites isolated from it rather originate from symbiotic bacteria than from the sponge itself.[38] Among the sponge symbionts, Abe *et al.* found deltaproteobacteria and detected a homolog of the saframycin Mx1 synthetase B while examining the metagenomics data set towards genes involved in secondary metabolite biosynthesis. It is therefore possible, that myxobacteria also belong to the symbionts of *Halichondria okada*. Hoffmann *et al.* made similar observations for another class of myxobacterial secondary metabolites, the microsclerodermins. Those secondary metabolites where first isolated from a marine sponge. Later on, they were also found to be produced by a set of terrestrial myxobacteria, which indicated that they could also be produced by symbiontic bacteria and not the sponge itself.[39]

Due to the structural similarities of the halichonines with the sandaramines, we assumed that they likely also feature the same stereochemical configuration, which was determined to be (-)-drimenol.[20] To prove this assumption, we developed an ex-chiral-pool total synthesis starting from sclareolide **6**.

After formation of drimenal **9** in four steps, we synthesized the secondary amine sandaramine A (**1**), as well as the tertiary amine sandaramine C (**3**), which is only present in trace amounts in the bacteriums' crude extract. Furthermore, we generated sandaramine D (**4**) which features a carbon linker. Comparison of HSQC and CD spectra of the synthetic and natural sandaramine A (**1**) confirmed, that they incorporate (-)drimenol **8** as observed in the halichonines.

Extended biological characterization of the natural and synthetic sandaramines showed, that they are not only active against several antimicrobial pathogens, but also exhibit intriguing antiviral activities against the SARS-CoV2 model CoV229E as well as antiproliferative effects. The sandaramines' producer strain *Sandaracinus sp.* MSr10575 therefore likely synthesizes these secondary metabolites as defense mechanism against microbial and eukaryotic competitors living in the same ecological niche. Even though the sandaramines bear some structural similarity to the halichonines, they seem to have a different mode of action. We could not detect apoptosis induction *via* the lysosomal pathway for cells treated with sandaramines, which is the proposed mode of action for the antiproliferative halichonines.[20] From the four different synthetic and natural sandaramine derivatives synthesized or isolated in this study, sandaramine A (**1**) and C (**3**) revealed the "best" biological activities. The synthetic derivative sandaramine D (**4**) only showed moderate cytotoxicity and slight antimicrobial activity. The central nitrogen in the homospermidine linker of the sandaramines and its substitution pattern therefore seems to play a crucial role for the molecules' biological activities and should optimally be a secondary or tertiary amine.

Our initial findings, especially the good antiviral activities of sandaramine A (**1**) and C (**3**), may serve as starting point for further medicinal chemistry based optimization approaches for the development of novel antiviral drugs. For the producer strain, the cytotoxic effects of the sandaramines may be desired effects as defense against eukaryotes such as amoebae or yeast or there was no need for the bacterium to optimize these effects, as most myxobacteria do not live as eukaryotic symbionts. From a pharmaceutical point of view, however, the therapeutic window should still be optimized towards less cytotoxic but still antiviral derivatives in order to proceed with sandaramines as an antiviral drug lead.

5.5 References

(1) Dousa, K. M.; Kurz, S. G.; Bark, C. M.; Bonomo, R. A.; Furin, J. J. Drug-Resistant Tuberculosis: A Glance at Progress and Global Challenges. *Infectious disease clinics of North America* **2020**, *34*, 863–886.

(2) Romanelli, R. J.; Azar, K. M.J.; Sudat, S.; Hung, D.; Frosch, D. L.; Pressman, A. R. The Learning Health System in Crisis: Lessons from the Novel Coronavirus Disease Pandemic. *Mayo Clinic Proceedings: Innovations, Quality & Outcomes* [Online early access]. DOI: 10.1016/j.mayocpiqo.2020.10.004.

(3) Dias, D. A.; Urban, S.; Roessner, U. A historical overview of natural products in drug discovery. *Metabolites* **2012**, *2*, 303–336.

(4) Martin, K. W.; Ernst, E. Herbal medicines for treatment of fungal infections: a systematic review of controlled clinical trials. *Mycoses* **2004**, *47*, 87–92.

(5) Palombo, E. A. Traditional Medicinal Plant Extracts and Natural Products with Activity against Oral Bacteria: Potential Application in the Prevention and Treatment of Oral Diseases. *Evidence-based complementary and alternative medicine : eCAM* **2011**, *2011*, 680354.

(6) Lin, L.-T.; Hsu, W.-C.; Lin, C.-C. Antiviral Natural Products and Herbal Medicines. *Journal of Traditional and Complementary Medicine* **2014**, *4*, 24–35.

(7) Bruno David; Jean-Luc Wolfender; Daniel A. Dias. The pharmaceutical industry and natural products: historical status and new trends. *Phytochem Rev* **2015**, *14*, 299–315.

(8) Thomford, N. E.; Senthebane, D. A.; Rowe, A.; Munro, D.; Seele, P.; Maroyi, A.; Dzobo, K. Natural Products for Drug Discovery in the 21st Century: Innovations for Novel Drug Discovery. *International journal of molecular sciences* **2018**, *19*.

(9) Pauli, G. F.; Chen, S.-N.; Friesen, J. B.; McAlpine, J. B.; Jaki, B. U. Analysis and Purification of Bioactive Natural Products: The AnaPurNa Study. *Journal of natural products* **2012**, *75*, 1243–1255.

(10) Hug, J. J.; Bader, C. D.; Remškar, M.; Cirnski, K.; Müller, R. Concepts and Methods to Access Novel Antibiotics from Actinomycetes. *Antibiotics* **2018**, *7*, 44.

(11) Joyce, M. D.; Jennings, M. C.; Santiago, C. N.; Fletcher, M. H.; Wuest, W. M.; Minbiole, K. P. Natural product-derived quaternary ammonium compounds with potent antimicrobial activity. *The Journal of antibiotics* **2016**, *69*, 344–347.

(12) Buszewski, B.; Noga, S. Hydrophilic interaction liquid chromatography (HILIC)--a powerful separation technique. *Analytical and bioanalytical chemistry* **2012**, *402*, 231–247.

(13) Liu, S.; Yu, H.; Zhang, X.; Cai, Y.-Q. Reversed-phase ion-pair solid-phase extraction and ion chromatography analysis of pyrrolidinium ionic liquid cations in environmental water samples. *Journal of separation science* [Online early access]. DOI: 10.1002/jssc.202000234.

(14) Majewski, W.; Valery, E.; Ludemann-Hombourger, O. Principle and Applications of Supercritical Fluid Chromatography. *Journal of Liquid Chromatography & Related Technologies* **2005**, *28*, 1233–1252.

(15) Machida, H.; Takesue, M.; Smith, R. L. Green chemical processes with supercritical fluids: Properties, materials, separations and energy. *J Supercrit Fluids* **2011**, *60*, 2–15.

(16) Bader, C. D.; Neuber, M.; Panter, F.; Krug, D.; Müller, R. Supercritical Fluid Extraction Enhances Discovery of Secondary Metabolites from Myxobacteria. *Analytical chemistry* [Online early access]. DOI: 10.1021/acs.analchem.0c02995.

(17) Lesellier, E. Usual, unusual and unbelievable retention behavior in achiral supercritical fluid chromatography: Review and discussion. *Journal of Chromatography A* **2020**, *1614*, 460582.

(18) Gibitz Eisath, N.; Sturm, S.; Stuppner, H. Supercritical Fluid Chromatography in Natural Product Analysis - An Update. *Planta Med.* [Online early access]. DOI: 10.1055/s-0037-1599461 (accessed January 26, 2018).

(19) Ohno, O.; Chiba, T.; Todoroki, S.; Yoshimura, H.; Maru, N.; Maekawa, K.; Imagawa, H.; Yamada, K.; Wakamiya, A.; Suenaga, K.*; et al.* Halichonines A, B, and C, novel sesquiterpene alkaloids from the marine sponge Halichondria okadai Kadota. *Chemical communications (Cambridge, England)* **2011**, *47*, 12453–12455.

(20) Hayakawa, I.; Nakamura, T.; Ohno, O.; Suenaga, K.; Kigoshi, H. Synthesis and structure-activity relationships for cytotoxicity and apoptosis-inducing activity of (+)-halichonine B. *Organic & biomolecular chemistry* **2015**, *13*, 9969–9976.

(21) Kwon, M.; Cochrane, S. A.; Vederas, J. C.; Ro, D.-K. Molecular cloning and characterization of drimenol synthase from valerian plant (Valeriana officinalis). *FEBS letters* **2014**, *588*, 4597–4603.

(22) Hoshino, T.; Kumai, Y.; Kudo, I.; Nakano, S.-i.; Ohashi, S. Enzymatic cyclization reactions of geraniol, farnesol and geranylgeraniol, and those of truncated squalene analogs having C20 and C25 by recombinant squalene cyclase. *Org. Biomol. Chem.* **2004**, *2*, 2650–2657.

(23) Porter, N. J.; Christianson, D. W. Binding of the Microbial Cyclic Tetrapeptide Trapoxin A to the Class I Histone Deacetylase HDAC8. *ACS chemical biology* **2017**, *12*, 2281–2286.

(24) Christianson, D. W. Structural and Chemical Biology of Terpenoid Cyclases. *Chemical reviews* **2017**, *117*, 11570–11648.

(25) Dickschat, J. S. Bacterial terpene cyclases. *Nat. Prod. Rep.* **2015**, *33*, 87–110.

(26) Henquet, M. G. L.; Prota, N.; van der Hooft, J. J. J.; Varbanova-Herde, M.; Hulzink, R. J. M.; Vos, M. de; Prins, M.; Both, M. T. J. de; Franssen, M. C. R.; Bouwmeester, H.*; et al.* Identification of a drimenol synthase and drimenol oxidase from Persicaria hydropiper, involved in the biosynthesis of insect deterrent drimanes. *The Plant journal : for cell and molecular biology* **2017**, *90*, 1052–1063.

(27) Tofalo, R.; Cocchi, S.; Suzzi, G. Polyamines and Gut Microbiota. *Frontiers in nutrition* **2019**, *6*, 16.

(28) Moll, S.; Anke, S.; Kahmann, U.; Hansch, R.; Hartmann, T.; Ober, D. Cell-specific expression of homospermidine synthase, the entry enzyme of the pyrrolizidine alkaloid pathway in Senecio vernalis, in comparison with its ancestor, deoxyhypusine synthase. *Plant Physiol.* **2002**, *130*, 47–57.

(29) Hoffmann, M.; Auerbach, D.; Panter, F.; Hoffmann, T.; Dorrestein, P. C.; Müller, R. Homospermidine Lipids: A compound class specifically formed during fruiting body formation of *Myxococcus xanthus* DK1622. *ACS Chem. Biol.* **2018**, *13*, 273–280.

(30) Gatto, G. J.; Boyne, M. T.; Kelleher, N. L.; Walsh, C. T. Biosynthesis of pipecolic acid by RapL, a lysine cyclodeaminase encoded in the rapamycin gene cluster. *J. Am. Chem. Soc.* **2006**, *128*, 3838–3847.

(31) Ober, D.; Hartmann, T. Homospermidine synthase, the first pathway-specific enzyme of pyrrolizidine alkaloid biosynthesis, evolved from deoxyhypusine synthase. *PNAS* **1999**, *96*, 14777–14782.

(32) Panter, F.; Bader, C. D.; Müller, R. *The sandarazols are cryptic and structurally unique plasmid encoded toxins from a rare myxobacterium*; BioRxIV, 2020.

(33) Zhang, S.; Li, D.; Song, Z.; Zang, C.; Zhang, L.; Song, X.; Li, S. "Carbon Assimilation" Inspired Design and Divergent Synthesis of Drimane Meroterpenoid Mimics as Novel Fungicidal Leads. *Journal of agricultural and food chemistry* **2017**, *65*, 9013–9021.

(34) Kuchkova, K. I. A Short Efficient Synthesis of 11-Monoacetate of Drimane-8α,11-diol from Norambreinolide. *Synthesis* **1997**, *1997*, 1045–1048.

(35) Li, D.; Zhang, S.; Song, Z.; Li, W.; Zhu, F.; Zhang, J.; Li, S. Synthesis and bio-inspired optimization of drimenal: Discovery of chiral drimane fused oxazinones as promising antifungal and antibacterial candidates. *European journal of medicinal chemistry* **2018**, *143*, 558–567.

(36) Ina, H.; Ito, M.; Kibayashi, C. Enantioselective Total Synthesis of the Macrocyclic Spermidine Alkaloid (−)-Oncinotine. *J. Org. Chem.* **1996**, *61*, 1023–1029.

(37) Takeda, Y.; Narita, K.; Katoh, T. Total Synthesis of Marine Sesquiterpene Quinones (+)-Cyclospongiaquinone-1 and (-)-Dehydrocyclospongiaquinone-1 with a Tetracyclic Benzo[a]xanthene Skeleton. *Eur. J. Org. Chem.* **2017**, *2017*, 901–907.

(38) Kikuchi, S.; Okada, K.; Cho, Y.; Yoshida, S.; Kwon, E.; Yotsu-Yamashita, M.; Konoki, K. Isolation and structure determination of lysiformine from bacteria associated with marine sponge Halichondria okadai. *Tetrahedron* **2018**, *74*, 3742–3747.

(39) Hoffmann, T.; Müller, S.; Nadmid, S.; Garcia, R.; Müller, R. Microsclerodermins from terrestrial myxobacteria: An intriguing biosynthesis likely connected to a sponge symbiont. *J. Am. Chem. Soc.* **2013**, *45*, 16904–16911.

5.6 Supporting information

Characterization of the Sandaramines, Myxobacterial Alkaloid Terpenoids accessible by Supercritical Fluid Chromatography

Manuscript in preparation

Chantal D. Bader[1], Fabian Panter[1], Alexander Kiefer[2], Sibylle Haid [3], Judith Hoffmann[1], Thomas Pietschmann[3] and Rolf Müller[1]

Affiliations

[1]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)

[2]Department Drug Design and Optimization, Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI) and Department of Pharmacy, Saarland University Campus E8.1, 66123 Saarbrücken (Germany)

[3] 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), Feodor-Lynen-Str. 7, 30625 Hannover, Germany

Data with limited visibility on print media such as raw NMR data can be found on the enclosed storage medium.

S5.1 Myxobacterial growth conditions and analytical setup

S5.1.1 Myxobacterial culture media

Table S1 Recipe for 2-SWT medium

2SWT – Medium			
Amount	Ingredient	Concentration	Supplier
3 g/L	Tryptone	-	BD
1 g/L	Soytone	-	BD
3.5 g/L	Soluble Starch	-	Roth
4 g/L	Maltose Monohydrate	-	
2 g/L	Glucose	-	Roth
10 g/L	Starch (soluble)	-	Roth
0.5 g/L	$CaCl_2$	-	Sigma Aldrich
1 g/L	$MgSO_4$ x $7H_2O$	-	Grüssing
10 mL/L	TRIS x HCl ph8	1M	Sigma Aldrich
100 µL/L	Sterile Vit. B12 solution (added after autoclaving)	1 mg/mL	Roth
200 µL/L	Sterile FeEDTA solution (added after autoclaving)	8 mg/mL	Sigma Aldrich
Dissolved in milli-Q. Water, pH adjusted to 7.2 with 1N KOH			

Table S2 Recipe for S15 medium

S15 – Medium			
Amount	Ingredient	Concentration	Supplier
3 g/L	Tryptone	-	BD
1 g/L	Soytone	-	BD
3.5 g/L	Soluble Starch	-	Roth
4 g/L	Maltose Monohydrate	-	
2 g/L	Glucose	-	Roth
10 g/L	Starch (soluble)	-	Roth
0.5 g/L	$CaCl_2$	-	Sigma Aldrich
1 g/L	$MgSO_4$ x $7H_2O$	-	Grüssing
10 mL/L	TRIS x HCl ph8	1M	Sigma Aldrich
100 µL/L	Sterile Vit. B12 solution (added after autoclaving)	1 mg/ml	Roth
200 µL/L	Sterile FeEDTA solution (added after autoclaving)	8 mg/ml	Sigma Aldrich
Dissolved in milli-Q. Water, pH adjusted to 7.2 with 1N KOH			

The myxobacterial strain MSr10575 was kept in agar culture broth for storage over short amounts of time and for cloning. The agar media used are S15 agar and S15 soft agar, which is prepared by adding 14 g/L agarose and 8 g/L agarose (BD) to S15 medium preparations before autoclaving.

S5.1.2 Small scale myxobacterial fermentations for acquisition of UPLC-HRMS data

Cultures for UHPLC/hrMS analysis were grown in 100 mL Erlenmeyer flasks containing 50 ml of 2-SWT medium and were inoculated with 1 mL of pre culture. After inoculation, the medium was supplemented with 2% of sterile XAD-16 adsorber resin (Sigma Aldrich) suspension in water. The strains were grown at 30 °C for 12-15 days on an orbitron shaker at 160 rpm. Fermentation was complete once the pale yellow medium turned orange and spherical myxobacterial bio-film clumps were visible in the shake flask.

S5.1.3 Extraction procedures of myxobacterial fermentation cultures in analytical scale

The frozen cell pellet and XAD was lyophilized and transferred into a 100 mL Erlenmeyer flask and a magnetic stirrer was added. 50 mL of analytical reagent grade methanol were added onto the pellet and the mixture was stirred for 60 min. The supernatant was filtered with a 125 micron folded filter and subsequently transferred into a 100 mL round bottom flask. The solvent was evaporated under reduced pressure. The residue was taken up in 1 mL of methanol (Chromasolv HPLC grade, Sigma Aldrich) and transferred into a 1.5 ml Eppendorf tube. This tube was centrifuged with Heareus biofuge pico at 15000 rpm for 15 minutes at 15 °C to remove residual insolubilities. The extracts were diluted 10fold prior to UHPLC-hrMS measurements.

S5.1.4 Standardized UHPLC MS method for screening of myxobacterial crude extracts

UHPLC-hrMS measurements were conducted on a Dionex Ultimate 3000 RSLC system coupled to a maXis 4G qTOF mass spectrometer (Bruker Daltonics, Bremen, Germany). Chromatographic separation of 1 μL myxobacterial extracts was achieved on a Waters BEH C18, 100 x 2.1 mm, 1.7 μm dp column at a flow rate of 600 μL/min and 45 °C using a linear gradient starting with an initial isocratic step at 95 % H_2O + 0.1 % FA (A) for 0.5 min, followed by an increase to 95 % ACN + 0.1 % FA (B) in 18 min, an isocratic step at 95 % B for 1 min and subsequent reequilibration to starting conditions. Pure compounds were measured under the same conditions on a Waters BEH C18, 50 x 2.1 mm, 1.7 μm dp column shortening the gradient from 18 min to 6 min. After UV detection with a DAD detector ranging from 200 to 600 nm, the LC flow was split to 75 μL/min before ionization with an Apollo II ESI source. Mass spectra were recorded in centroid mode with a detection range from 150 to 2000 m/z and a scan rate of 2 Hz. Calibration was done automatically before every LC-MS run by injecting sodium formate. All MS analyses were acquired in the presence of the lock masses $C_{12}H_{19}F_{12}N_3O_6P_3$; $C_{18}H_{19}O_6N_3P_3F_2$ and $C_{24}H_{19}F_{36}N_3O_6P_3$ which generate the $[M+H]^+$ ions of 622.028960; 922.009798 and 1221.990638 used for subsequent lock mass calibration.

S5.2 Molecular biology protocols

S5.2.1 Isolation of genomic DNA for single molecule real-time sequencing

To isolate total DNA for sequencing purposes, phenol-chloroform gDNA extraction as described in the following protocol was used.

1) Spin down 50 mL of fresh myxobacterial culture 6000 rcf 10 min.

2) Discard the supernatant.

3) Wash the cells once with SET Buffer, centrifuge at 6000 rcf 10 min.

4) Resuspend cell pellet in 5ml SET Buffer.

5) Add 100 µL of Lysozym (50 mg/mL in ddH$_2$O) stock solution as well as 50µL RNAse A (10 mg/mL in ddH$_2$O) stock solution.

6) Incubate at 37 °C for 15 min

7) Add 300 µL Proteinase K solution (10 mg/mL 50 mM Tris 1mM CaCl$_2$) invert several times and add 600 µL 10% SDS solution.

8) Incubate at 55°C for 2 h, invert every 15 min.

9) Add even volume (6 mL) of phenol/chloroform/isoamylalcohol (25:24:1) and swing the tube for 60 min at 5 rpm.

10) Centrifuge the mixture at 6000 rcf for 5 min at room temperature.

11) Transfer the upper phase into a new tube using a cut end 1mL tip.

12) Repeat step 9)-11) two more times.

13) Add 1/10 of the total volume of 3M NaOAc solution pH 5.5 and mix well by inverting several times.

14) Add 2.5 volumes of ice cold ethanol (100% technical purity, -20°C) and invert the tube several times, DNA precipitation should be visible as a cotton like fog in the tube.

15) Spool the DNA on a sealed Pasteur pipette.

16) Rinse the DNA with 70% ethanol (-20°C).

17) Air dry the DNA for at least 15 minutes (dry DNA will become completely translucid).

18) Resuspend dried DNA in 0.5 mL of ddH$_2$O and keep the Eppendorf tube at room temperature for 24 hours.

19) Check the genomic DNA on an agarose gel.

S5.2.2 *In-silico* analysis of the genes involves in the sandaramine biosynthesis

The sandaramine biosynthetic pathway consists of the biosyntheses of the two building blocks of sandaramine, namely the drimane backbone and the homospermidine linker. The biosynthetic machinery of the homospermidine linker is related to deoxyhypusine synthase like proteins that perform the homospermidine biosynthesis from the primary metabolites putrescine and spermidine.[1]
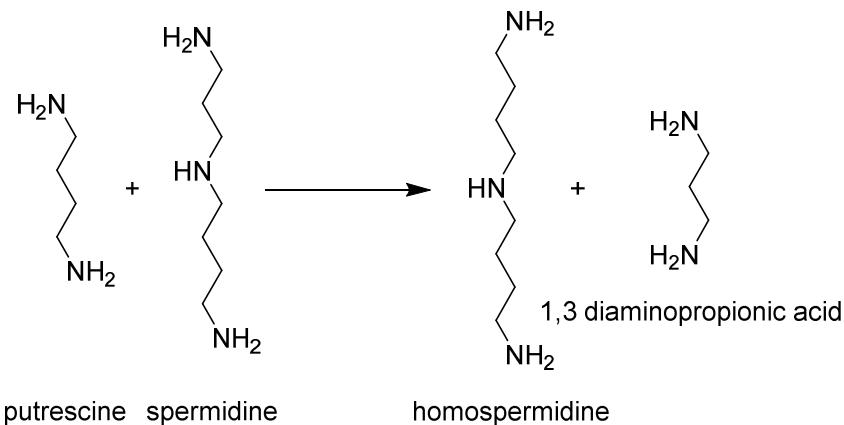
Figure S1 Reaction scheme of the homospermidine synthase.

Blast searches in the MSr10575 genome for the homospermidine synthase of *Myxococcus xanthus* DK1622 (CDS 1643 on RefSeq NC_008095.1) revealed the MSr10575 homospermidine synthase coupled to its AraC like regulator and a putative damage inducible protein DinB. As there are no terpene biosynthesis genes in the surrounding genetic regions of the MSr10575 homospermidine synthase, the biosynthesis of the drimane backbone had to be encoded elsewhere in the genome. Every coding sequence in the homospermidine biosynthesis operon was extracted translated and searched with the blastp algorithm against the RefSeq non-redundant protein sequence database at NCBI.[2]

Table S3 Blast analysis of the homospermidine synthase operon against the non-redundant sequences database responsible for homospermidine linker biosynthesis

CDS Name	Length [bp]	Closest homologue [Organism of origin]	Identity [%] and alignment length [AA]	Proposed function	Accession Nr.
HssI	747	AraC family transcriptional regulator [Reyranella sp.]	51.6; 217	regulation	TAJ36350
HssII	513	damage-inducible protein DinB [Sandaracinus amylolyticus]	91.8; 170	-	WP_053238380
HssIII	1008	deoxyhypusine synthase [Sandaracinus amylolyticus]	97.3; 335	homospermidine biosynthesis	WP_053238381

Drimenol, or drimane alcohol, is biosynthesized from farnesol or farnesyl pyrophosphate (farnesyl-PP) by a class II terpene cyclase. Thus, instead of removing the pyrophosphate group from the terpenoid precursor to create a cation that will initiate the cyclization reactions as in limonene biosynthesis (a class I terpene synthase) this class II terpene cyclase iniciates farnesyl-PP cyclization by creating a cation by protonating the terminal C-C double bond (Figure S2).

Class I terpene cyclisation in limonene biosynthesis

geranyl pyrophosphate   Limonene
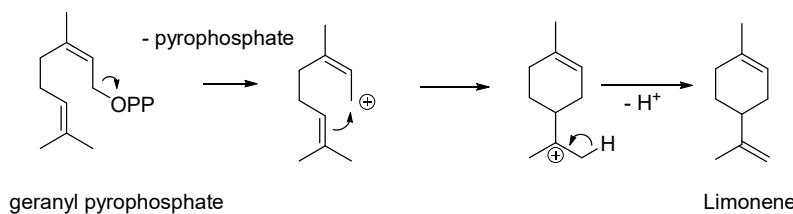
Class II terpene cyclisation in drimenol biosynthesis

farnesyl pyrophosphate   drimenol pyrophosphate
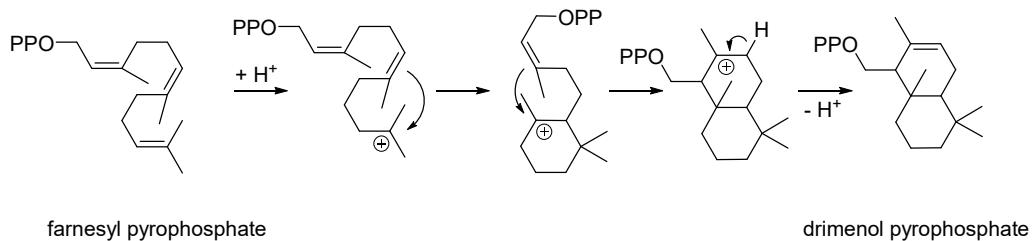
Figure S2 Reaction scheme of limonene cyclization and drimenol cyclization as examples for class I and class II terpene cyclization reactions.

So even though Terpene cyclases catalyze seemingly similar cyclizations, the starting reactions for these cyclizations, e.g the removal of a pyrophosphate group vs protonation of a C-C double bond are very different. It is therefore possible to discern these terpene cyclase classes in blast alignments.[3]

Table S4 Blast analysis of the sandaramine biosynthesis operon against the non-redundant sequences database responsible for homospermidine linker biosynthesis

CDS Name	Length [bp]	Closest homologue [Organism of origin]	Identity [%] and alignment length [AA]	Proposed function	Accession Nr.
sanA	1134	Isopentenyl-diphosphate delta-isomerase, FMN-dependent [Sandaracinus amylolyticus]	94.7; 376	Terpenoid precursor biosynthesis	AKF09146
sanB	2625	hydroxymethylglutaryl-CoA reductase, degradative [Sandaracinus amylolyticus]	97.5; 875	Terpenoid precursor biosynthesis	WP_075097668
sanC	1176	aromatic ring-hydroxylating dioxygenase subunit alpha [Sandaracinus amylolyticus]	97.3; 392	-	WP_075097667
sanD	1125	NAD(P)/FAD-dependent oxidoreductase [Sandaracinus amylolyticus]	97.3; 374	Reductive amination	WP_053236222
sanE	3438	Squalene--hopene cyclase [Chromatiales bacterium USCg_Taylor]	50.5; 1103	Farnesyl-PP cyclization	OOO02642
sanF	1176	Delta(14)-sterol reductase [Nicotiana attenuata]	75.0, 369	Reductive amination	XP_019250816
sanG	3432	FAD-dependent oxidoreductase [Sandaracinus amylolyticus]	96.6; 1146	Reductive amination	WP_053236219
sanH	1779	cytochrome P450 [Sandaracinus amylolyticus]	97.6; 591	drimenol oxiation	WP_083457980
sanI	564	hypothetical protein [Sandaracinus amylolyticus]	96.3; 187	-	WP_075097666

The only class II terpene cyclase found in the MSr10575 genome is the protein encoded by *sanE,* which is most likely responsible for cyclization of the drimenol core structure. This protein is encoded on *sanE*, a class II squalene-hopene cyclase type protein. It is embedded in an operon spanning 9 proteins (*sanA* to *sanI*) that not only contain the biosynthetic machinery for drimenol cyclization but also additional copies of proteins involved in farnesyl-PP biosynthesis (SanA and SanB) as well as the redox proteins putatively responsible for catalyzing the reductive amination coupling the drimenol to the homospermidine to form the sandaramines. As this operon seems to be not only responsible for drimenol biosynthesis and sandaramine assembly but also for the biosynthesis of steroids that are cyclized via the same mechanism, we attempted an inactivation of *sanE* by homologous recombination to evaluate whether this would abolish sandaramine production. The two operons are available in GenBank under accession number XXX for the homospermidine biosynthesis operon and XXX for the drimenol biosynthesis and sandaramine assembly operon (data will be deposited upon publication of the respective article).

S5.2.3 Gene Disruption of *sanE*

Integrative inactivation of *sanE* to evaluate its effect on sandaramine production was achieved by single crossover mutagenesis. Therefore, a 921 bp fragment of the *sanE* gene was PCR amplified (Phusion Polymerase, Thermo Scientific), A-tailed and cloned into a pCR2.1 vector backbone originating from the TOPO-TA cloning kit.Primers utilized for PCR amplification of the *sanE* fragment are SHC fragment fwd. and SHC fragment rev.. Correct clones are selected after plasmid preparation by alkaline lysis followed by control digestion using the restriction endonuclease SphI.

Table S5 Primers used for single crossover inactivation of the *sanE* gene

primer name	3' to 5' primer sequence
SHC fragment fwd	ACTGCGCTCGATGGGGTTCC
SHC fragment rev	CCGGGGAAGATCTCGTCGCG
SHC integration verif fwd	TGGGCGAGTGTCGATCTGCG
SHC integration verif rev	CCAGACCCAGCCCTCGAAGC
pCR2.1 in	CCTCTAGATGCATGCTCGAGC
pCR2.1 out	GGATCCACTAGTAACGGCCG

Integration of this vector into the genome of MSr10575 was achieved by electroporation followed by single crossover integration into said chromosome to form MSr10575:: pCR2.1 *sanE* fragment. Correct integration of the pCR2.1 vector into *sanE* is verified by PCR using the verif primers and the outward facing primers on the pCR2.1 vector pCR2.1 in and pCR2.1 out. Single crossover integration was confirmed by PCR. The mutant did not produce sandaramines any more as verified by LC-MS.
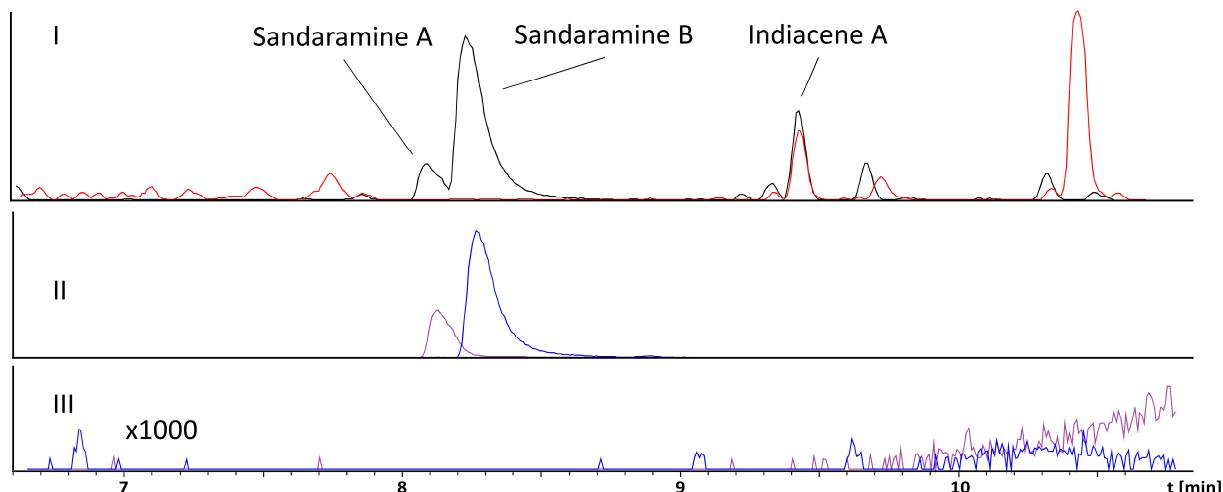
Figure S3 I) Chromatogram of MSr10575 wild type (black trace) and MSr10575 :: pCR2.1 sanE fragment (red trace) extract II) EIC traces for sandaramine A [M+H]$^+$ (568.56 ± 0,05 Da, purple) and sandaramine B [M]$^+$ (569.59 ± 0.05 Da, blue) in MSr10575 wild type extract and III) EIC traces for sandaramine A [M+H]$^+$ (568.56 ± 0.05 Da, purple) and sandaramine B [M]$^+$ (569.59 ± 0.05 Da, blue) in MSr10575 :: pCR2.1 *sanE* fragment extract (1000 fold magnification)

S5.3 Isolation and structure elucidation of sandaramine A and B

S5.3.1 General NMR setup

NMR spectra were recorded on a Bruker UltraShield 500 spectrometer or a Bruker Ascend 700 equipped with a 5 mm TCI cryprobe each. Chloroform-*d* and methanol-*d4* were used as solvents and chemical shifts of the solvent signals at 3.31/49.2 (MeOD-*d4*) or 7.24/77.2 (CDCl₃) were considered as reference signals. [4] Chemical shift values (*δ*) are given in ppm and the coupling constants (*J*) in Hz. *g*COSY, HSQC-dept, HSQC-TOCSY and HMBC experiments were recorded using standard pulse programs. To increase sensitivity 5 mm Shigemi tubes (Shigemi Inc., Allison Park, PA 15101, USA) optimized for CHCl₃ or MeOH were used.

S5.3.2 Isolation of sandaramine A and B from MSr10575

Cultivation conditions for production of the sandaramines

For the production of the sandaramines the myxobacterial strain MSr10575 was grown in 25 mL of CYH-medium (4 g/L Soy flour, 3 g/L Casitone, 2 g/L Glucose, 8 g/L Soluble starch, 1.5 g/L Yeast extract, 1 g/L CaCl₂ x 2 H₂O, 1 g/L MgSO₄ x 7 H₂O, 50 mM HEPES adjusted to pH 7.2 with KOH) for 5 days. Those precultures were subsequently used to inoculate 2 L of RG5-medium (0.5 g/L Soy peptone, 0.5 g/L Soytone, 2 g/L Soy flour, 1 g/L Corn steep solids, 0.5 g/L Yeast extract, 8 g/L soluble starch, 5 g/L Baker's Yeast, 2 g/L Gluten from wheat, 1 g/L MgSO₄ x 7 H₂O, 1 g/L CaCl₂ x 2 H₂O, 50 mM HEPES adjusted to pH 7.2 with KOH). Growing for 14 days at 30 °C and 160 rpm, the production cultures were spiked with two percent of XAD adsorber resin (Amberlite XAD-16, Sigma) at day 3. The cultures were harvested by centrifugation, using XAD and cell pellet for the extraction.

Isolation procedure

XAD and cell pellet were extracted with 3 x 300 mL of methanol followed by 1 x 300 mL of acetone. The thus obtained extracts were combined and the solvents evaporated, yielding the crude extract as orange oil. The extract was further processed by liquid-liquid partitioning between methanol and chloroform. Sandaramine A and B were purified by preparative supercritical fluid chromatography using a Waters SFC 20 system equipped with a Waters Torus Diol 150 x 19 mm, 5 um d_P column from the concentrated chloroform extract. The system was operated with a flow rate of 15 mL/min at 45 °C, 120 bar using a linear gradient from 5-55% with methanol with 0.1 % TEA as co-solvent. Mass triggered fraction collection yielded 2.7 mg of sandaramine A and 1.8 mg of sandaramine B as light yellow oil.

S5.3.3 NMR spectra of the isolated natural sandaramines

Table S6 NMR spectroscopic data for sandaramine A (**1**) in chloroform-*d* at 500/125 MHz isolated from MSr10575 extracts.

#	δ ^{13}C [ppm]	δ ^{1}H [ppm], mult (*J* [Hz])	COSY	HMBC
1, 1'	49.7	2.63, 4H, m	2/2', 3/3', 4/4'	1/1', 2/2', 3/3',
2, 2'	27.9	1.53, 4H, m*	1/1', 3/3', 4/4'	1/1', 3/3', 4/4'
3, 3'	27.9	1.53, 4H, m*	1/1', 2/2', 4/4'	1/1', 2/2', 4/4'
4, 4'	50.1	2.56, 4H, m	1/1', 2/2', 3/3'	2/2', 3/3', 5/5'
5, 5'	48.2	2.47/2.71, 4H, dd (12.2, 7.3)	5/5', 6/6'	6/6', 7/7', 15/15'
6, 6'	55.2	1.75, 2H, bs	5/5'	7/7', 8/8', 9/9'
7, 7'	134.2	-	-	-
8, 8'	123.0	5.44, 2H, bdt	9/9', 10,10' 16/16'	7/7', 9/9', 10/10', 16/16'
9, 9'	23.7	1.98/1.84*	8/8', 9/9', 10/10' 16/16'	6/6', 7/7', 8/8', 10/10', 15/15'
10,10'	49.9	1.19, 4H, m	8/8', 9/9', 17/17'	6/6', 7/7', 8/8', 9/9', 11/11', 12/12', 13/13', 14/14', 15/15', 17/17', 18/18'
11, 11'	32.9	-	-	-
12, 12'	42.2	1.16/1.41, 4H*, m*	12/12', 13/13', 18/18'	10/10', 11/11', 13/13', 14/14', 17/17', 18/18'
13, 13'	18.8	1.49, 4H*, m*	12/12', 14/14'	11/11', 12/12', 14/14', 15/15'
14, 14'	39.3	1.07/1.94, 4H, td (13.1, 3.6)	13/13', 14/14', 19/19'	6/6', 10/10', 12/12', 13/13', 15/15', 19/19'
15, 15'	36.2	-	-	-
16, 16'	22.0	1.73, 6H*, bd (0.7)	8/8'	6/6', 7/7', 8/8', 9/9'
17, 17'	33.2	0.85, 6H, s	10/10', 18/18'	10/10', 11/11', 12/12', 13/13', 15/15', 18/18'
18, 18'	21.9	0.88, 6H, s	12/12', 17/17'	10/10', 11/11', 12/12', 17/17'
19, 19'	14.0	0.76, 6H, s	6/6', 14/14'	6/6', 7/7', 14/14', 15/15'

* overlapping signals
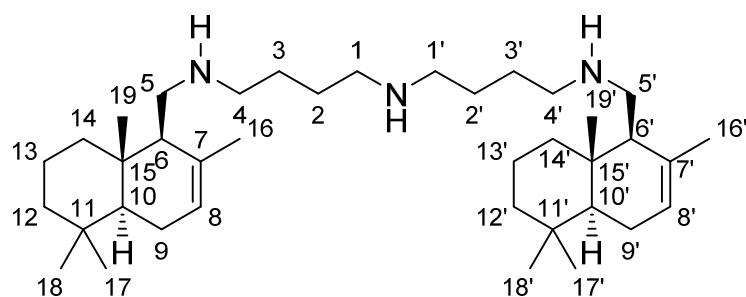
Table S7 NMR spectroscopic data for sandaramine B (**2**) in methanol-d_4 at 500/125 MHz isolated from MSr10575 extracts.

#	δ ^{13}C [ppm]	δ ^{1}H [ppm], mult (J [Hz])	COSY	HMBC
1, 1'	65.1	3.31*	2/2', 3/3'	1/1', 2/2', 3/3', 20/20'
2, 2'	21.7	1.80, 4H, m*	1/1', 3/3'	1/1', 3/3', 4/4'
3, 3'	27.3	1.56, 4H, m*	1/1', 2/2', 4/4'	1/1', 2/2', 4/4'
4, 4'	50.3	2.65, 4H, m*	3/3'	2/2', 3/3', 5/5'
5, 5'	49.2	2.48/2.72, 4H, dd (12.3, 7.5)	5/5', 6/6'	6/6', 7/7', 15/15'
6, 6'	56.4	1.78, 2H, m	5/5', 8/8'	7/7', 8/8', 9/9'
7, 7'	135.1	-	-	-
8, 8'	124.3	5.46, 2H, m	9/9', 10,10' 16/16'	7/7', 9/9', 10/10', 16/16'
9, 9'	25.0	2.00/1.89, 4H, m*	8/8', 9/9', 10/10' 16/16'	6/6', 7/7', 8/8', 10/10', 15/15'
10,10'	51.5	1.21, 4H, m*	8/8', 9/9', 17/17'	6/6', 7/7', 8/8', 9/9', 11/11', 12/12', 13/13', 14/14', 15/15', 17/17', 18/18'
11, 11'	34.1	-	-	-
12, 12'	43.6	1.21/1.44, 4H*, m*	12/12', 13/13', 17/17'	10/10', 11/11', 13/13', 14/14', 17/17', 18/18'
13, 13'	20.0	1.48/1.61, 4H*, m*	12/12', 14/14'	11/11', 12/12', 14/14', 15/15'
14, 14'	40.5	1.13/1.96, 4H, m*	13/13', 14/14', 19/19'	6/6', 10/10', 12/12', 13/13', 15/15', 19/19'
15, 15'	37.6	-	-	-
16, 16'	22.3	1.75, 6H, bs	8/8'	6/6', 7/7', 8/8', 9/9'
17, 17'	33.9	0.88, 6H, s	10/10', 18/18'	10/10', 11/11', 12/12', 13/13', 15/15', 18/18'
18, 18'	22.4	0.91, 6H, s	12/12', 17/17'	10/10', 11/11', 12/12', 17/17'
19, 19'	14.5	0.79, 6H, s	5/5', 6/6', 14/14'	6/6', 7/7', 14/14', 15/15'
20, 20'	51.5	3.07	1/1'	1/1', 20/20'
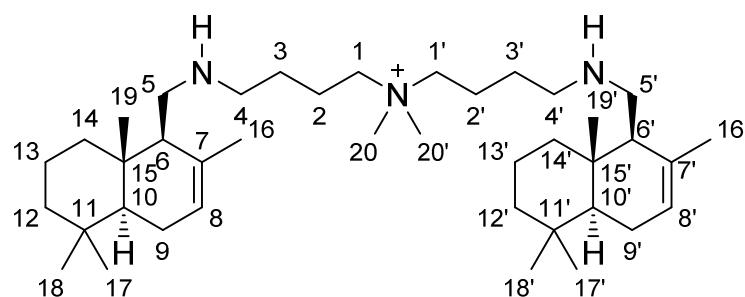
* overlapping signals

Table S8 NMR spectroscopic data for sandaramine B (**2**) in chlorofom-*d* at 500/125 MHz isolated from MSr10575 extracts.

#	δ ^{13}C [ppm]	δ ^{1}H [ppm], mult (*J* [Hz])	COSY
1, 1'	63.7	3.43, 4H, m	2/2'
2, 2'	20.5	1.80, 4H, m*	1/1', 3/3'
3, 3'	26.2	1.56, 4H, m*	2/2', 4/4'
4, 4'	48.8	2.68, 4H, m*	3/3'
5, 5'	48.0	2.47/2.69, 4H, dd (11.9, 7.3)	5/5', 6/6'
6, 6'	55.1	1.72, 2H, m	5/5', 8/8'
7, 7'	n.d.	-	-
8, 8'	123.1	5.47, 2H, m	9/9', 10,10' 16/16'
9, 9'	23.5	1.86/1.99, 4H, m*	8/8', 9/9', 10/10' 16/16'
10,10'	49.8	1.20, 4H, m*	8/8', 9/9', 12/12', 17/17'
11, 11'	n.d.	-	-
12, 12'	42.0	1.17/1.47, 4H*, m*	12/12', 13/13', 18/18'
13, 13'	18.6	1.56, 4H, m*	12/12', 14/14'
14, 14'	39.2	1.05/1.92, 4H, dt (13.3, 3.3)	12/12', 13/13', 14/14', 19/19'
15, 15'	n.d.	-	-
16, 16'	21.9	1.73, 6H, bs	8/8'
17, 17'	33.1	0.88, 6H, s	10/10', 18/18'
18, 18'	21.8	0.90, 6H, s	12/12', 17/17'
19, 19'	14.0	0.78, 6H, s	5/5', 6/6', 10/10', 14/14'
20, 20'	50.9	3.24, 6H, bs	-
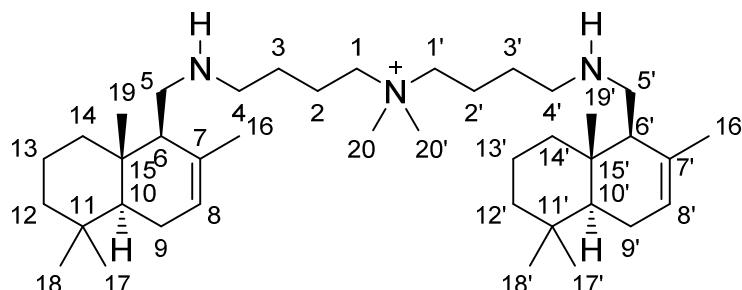
* overlapping signals

S5.4 Total synthesis and characterization of the reaction intermediates

S5.4.1 General experimental procedures

General procedure A: DMP oxidation and reductive amination for sandaramine D

DMP (2.0 equiv) was added in small portions to a stirred solution of ((1S,8aS)-2,5,5,8a-tetramethyl-1,4,4a,5,6,7,8,8a-octahydronaphthalen-1-yl)methanol (**8**) in DCM (0.1 M) at 0 °C. After 120 minutes, the reaction was quenched with sat. Na_2SO_3 solution and the resulting mixture was extracted with CH_2Cl_2 twice. The combined organic extracts were washed with sat. $NaHCO_3$ solution and brine before dried over Na_2SO_4. Concentration of the solvent in vacuum afforded a residue which was used without further purification. To a solution of aldehyde and amine (0.5 equiv) in DCE (0.1 M), $NaBHOAc_3$ (5.0 equiv) was added portionwise at rt. After complete conversion (LC-MS), the mixture was diluted with DCM, washed with sat. $NaHCO_3$ solution and brine, dried over Na_2SO_4 and the solvent was removed under reduced pressure. The crude products were then purified *via* supercritical fluid chromatography following the protocol described for the isolation of sandaramine A (**1**) and B (**2**).

General procedure B: DMP oxidation and reductive amination for sandaramine A (1) and C (3)

DMP (2.0 equiv) was added in small portions to a stirred solution of ((1S,8aS)-2,5,5,8a-tetramethyl-1,4,4a,5,6,7,8,8a-octahydronaphthalen-1-yl)methanol (**8**) in DCM (0.1 M) at 0 °C. After 120 minutes, the reaction was quenched with sat. Na_2SO_3 solution and the resulting mixture was extracted with CH_2Cl_2 twice. The combined organic extracts were washed with sat. $NaHCO_3$ solution and brine before dried over Na_2SO_4. Concentration of the solvent in vacuo afforded a residue which was used without further purification. The aldehyde was mixed with the amine (0.5 equiv) in MeOH (0.1 M) and triethylamine (6 equiv) and stirred for 120 minutes for imine **15** formation. Subsequently, $NaBH_4$ (5.0 equiv) was added portionwise at rt. After complete conversion (LC-MS), the mixture was diluted with DCM, washed with sat. $NaHCO_3$ solution and brine, dried over Na_2SO_4 and the solvent was removed under reduced pressure. The crude products were then purified via supercritical fluid chromatography following the protocol described for the isolation of sandaramine A (**1**) and B (**2**).

1-((1*R*,2*R*,4a*S*,8a*S*)-2-hydroxy-2,5,5,8a-tetramethyldecahydronaphthalen-1-yl)propan-2-one (6)

(3aR,5aS,9aS,9bR)-3a,6,6,9a-tetramethyldecahydronaphtho[2,1-b]furan-2(1H)-one **(5)** (2.00 g, 7.99 mmol) was diluted in 30 mL anhydrous Et_2O. After cooling to −78 °C methyllithium (5.49 mL, 8.79 mmol, 1.6 M in Et_2O) was added dropwise. After stirring for 60 minutes at this temperature the reaction mixture was quenched with sat. NH_4Cl solution. The organic phase was washed with brine and dried over Na_2SO_4 before the solvent was evaporated under reduced pressure. Flash chromatography (SiO₂, PE:EA 7:3) afforded the desired ketone **6** (2.13 g, 7.99 mmol, quant.) as colorless oil in quantitative yields.
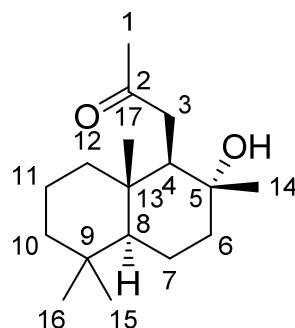
TLC: PE:EA = 7:3; R$_f$(**6**) = 0.25.

¹H NMR (500 MHz, CDCl₃): δ = 0.79 (s, 3 H, 16-H), 0.79 (s, 3 H, 17-H), 0.88 (s, 3 H, 15-H), 0.91 (dd, 1 H, *J* = 12.5, 3.7 Hz, 12-H'), 1.02 (dd, *J* =12.3, 2.2 Hz, 1 H, 8-H), 1.11 (s, 3 H, 14-H), 1.13 - 1.19 (m, 2 H, 10-H'), 1.25 - 1.30 (m, 1 H, 7-H'), 1.33 − 1.39 (m, 1 H, 12-H''), 1.34 − 1.39 (m, 1 H, 10-H''), 1.39 − 1.45 (m, 1 H, 11-H'), 1.38 − 1.45 (m, 2 H, 6-H), 1.55 (dt, *J*= 13.8, 3.4 Hz, 1 H, 11-H''), 1.69 (dq, 1 H, *J* = 13.7, 3.7 Hz, 7-H''), 1.90 − 1.95 (m, 1 H, 6-H''), 1.92 − 1.96 (m, 1 H, 4-H), 2.20 (s, 3 H, 1-H), 2.45 (dd, *J*= 17.9, 4.1 ppm, 1 H, 3-H'), 2.54 (dd, *J*= 17.9, 4.1 ppm, 1 H, 3H'') ppm

¹³C NMR (125 MHz, CDCl₃): δ = 15.7 (s, 1 C, 16-C), 18.4 (s, 1 C, 11-C), 20.6 (s, 1 C, 7-C), 21.4 (s, 1 C, 17-C), 32.1 (s, 1 C, 14-C), 30.3 (s, 1 C, 1-C), 33.2 (s, 1 C, 9-C), 33.3 (s, 1 C, 15-C), 38.3 (s, 1 C, 13-C), 39.3 (s, 1 C, 12-C), 39.5 (s, 1 C, 3-C), 41.7 (s, 1 C, 10-C), 44.6 (s, 1 C, 6-C), 55.8 (s, 1 C, 8-C), 55.9 (s, 1 C, 4-C), 73.1 (s, 1 C, 5-C), 210.3 (s, 1 C, 2-C)

HRMS (ESI):   289.2144 *m/z* calculated for [M+Na]⁺; found 289.2153 (Δ 3.1 ppm).

((1S,2R,4aS,8aS)-2-hydroxy-2,5,5,8a-tetramethyldecahydronaphthalen-1-yl)methyl acetate (7)

To a solution of urea peroxide (1.84 g, 19.52 mmol) and Na_2CO_3 (0.58 g, 5.46 mmol) in 15 mL DCM, trifluoroacetic acid anhydride (0.69 mL, 4.88 mmol) was added dropwise at 0 °C. The resulting mixture was stirred for 5 minutes at this temperature and 5 minutes at rt. Afterwards a solution of ketone **6** (0.50 g, 1.95 mmol) in 4.0 mL DCM was added at 0 °C and the mixture was stirred for one hour. Quenching by addition of sat. Na_2SO_3, washing by sat. $NaHCO_3$ and brine, drying over Na_2SO_4 and evaporation of the solvent afforded the acetate **7** (0.55 g, 1.95 mmol, quant.) as colorless oil in quantitative yields.
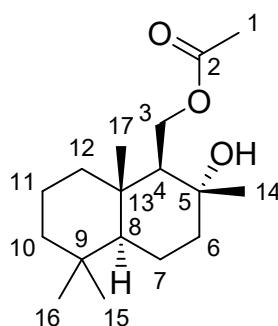
TLC: PE:EA = 7:3, R_f(**7**) = 0.30.

^{1}H NMR (400 MHz, CDCl$_3$): δ = 0.81 (s, 3 H, 17-H), 0.87 (s, 3 H, 16-H), 0.89 (s, 3 H, 15-H), 0.96 (dd, *J* =12.2, 2.1 Hz, 1 H, 8-H), 1.05 (dt, 1 H, *J* = 13.1, 3.2 Hz, 12-H'), 1.11 - 1.19 (m, 2 H, 10-H'), 1.19 (s, 3 H, 14-H), 1.24 - 1.27 (m, 1 H, 7-H'), 1.37 – 1.42 (m, 1 H, 10-H''), 1.42 – 1.49 (m, 1 H, 11-H'), 1.44 – 1.50 (m, 2 H, 6-H), 1.53 (t, 1 H, *J* = 4.8 Hz, 4-H), 1.60 (dt, *J*= 13.5, 3.4 Hz, 1 H, 11-H''), 1.64 – 1.70 (m, 1 H, 12-H''), 1.64 – 1.70 (m, 1 H, 7-H''), 1.90 (dt, 1 H, *J* = 12.7, 3.2 Hz, 6-H''), 2.05 (s, 3 H, 1-H), 4.25 (dd, *J*= 11.7, 5.4 Hz, 1 H, 3-H'), 4.36 (dd, *J*= 11.8, 4.3 Hz, 1 H, 3H'') ppm

^{13}C NMR (100 MHz, CDCl$_3$): δ = 15.8 (s, 1 C, 16-C), 18.4 (s, 1 C, 11-C), 20.3 (s, 1 C, 7-C), 21.3 (s, 1 C, 1-C), 21.5 (s, 1 C, 17-C), 24.6 (s, 1 C, 14-C), 33.2 (s, 1 C, 9-C), 33.4 (s, 1 C, 15-C), 38.1 (s, 1 C, 13-C), 39.7 (s, 1 C, 12-C), 41.7 (s, 1 C, 10-C), 43.9 (s, 1 C, 6-C), 55.7 (s, 1 C, 8-C), 59.9 (s, 1 C, 4-C), 62.6 (s, 1 C, 3-C), 72.6 (s, 1 C, 5-C), 171.3 (s, 1 C, 2-C)

HRMS (ESI):   305.2093 *m/z* calculated for [M+Na]$^+$; found 305.2097 (Δ 1.4 ppm).

((1S,2R,4aS,8aS)-2-hydroxy-2,5,5,8a-tetramethyloctacahydronaphthalen-1-yl)methanol (8)

p-Toluenesulfonic acid monohydrate (0.34 g, 1.77 mmol) was added in one portion to a solution of acetate **7** (0.50 g, 1.77 mmol) in 17 mL DCM at rt. After complete conversion (TLC), the solution was diluted with DCM and washed with water, sat. $NaHCO_3$ solution and brine. The organic phase was dried over Na_2SO_4 and the solvent was removed under reduced pressure. The crude product was purified by column chromatography (SiO$_2$, PE:Et$_2$O 4:1) to give olefin (0.23 g, 0.88 mmol, 50%) and the corresponding alcohol **8** (40 mg, 0.18 mmol, 10%) as colorless oils. To a solution of the olefinic acetate (0.23 g, 0.88 mmol) in 8.8 mL MeOH/H$_2$O (9:1) K_2CO_3 (0.13 g, 0.94 mmol) directly was added at 0 °C.

After complete conversion (TLC), the solvent was evaporated and the residue was diluted with water. After acidification with 1M $KHSO_4$ solution, the aqueous phase was extracted with Et_2O twice. Washing the organic phase with brine, drying over Na_2SO_4 and removing the solvent under reduced pressure gave the pure alcohol **8** (0.18 g, 0.81 mmol, 94%) without further purification.
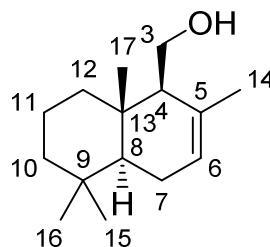
TLC: PE:EA = 7:3, R_f(**8**) = 0.45.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 0.84 (s, 3 H, 15-H), 0.86 (s, 3 H, 16-H), 0.88 (s, 3 H, 17-H), 1.17 (dd, *J* =11.8, 4.4 Hz, 1 H, 8-H), 1.06 (td, 1 H, *J* = 13.2, 3.5 Hz, 12-H'), 1.11 - 1.19 (m, 2 H, 10-H'), 1.78 (s, 3 H, 14-H), 1.82 - 1.90 (m, 1 H, 7-H'), 1.40 – 1.44 (m, 1 H, 10-H''), 1.44 – 1.49 (m, 1 H, 11-H'), 5.52 – 1.55 (m, 2 H, 6-H), 1.82 – 1-87 (m, 1 H, 4-H), 1.56 (dt, *J*= 13.6, 3.0 Hz, 1 H, 11-H''), 1.93 – 1.98 (m, 1 H, 12-H''), 1.96 – 2.03 (m, 1 H, 7-H''), 3.72 (dd, *J*= 11.2, 5.0 Hz, 1 H, 3-H'), 3.85 (dd, *J*= 11.3, 3.2 Hz, 1 H, 3H'') ppm

^{13}C NMR (125 MHz, $CDCl_3$): δ = 14.9 (s, 1 C, 16-C), 18.8 (s, 1 C, 11-C), 22.0 (s, 1 C, 14-C), 22.1 (s, 1 C, 17-C), 23.6 (s, 1 C, 7-C), 32.9 (s, 1 C, 9-C), 33.4 (s, 1 C, 15-C), 36.1 (s, 1 C, 13-C), 39.9 (s, 1 C, 12-C), 42.1 (s, 1 C, 10-C), 49.9 (s, 1 C, 8-C), 57.2 (s, 1 C, 4-C), 60.9 (s, 1 C, 3-C), 124.1 (s, 1 C, 6-C), 133.0 (s, 1 C, 5-C)

HRMS (ESI):   245.1882 *m/z* calculated for [M+Na]$^+$; found 245.1886 (Δ 1.6 ppm).

tert-butyl (4-iodobutyl)carbamate (11)

I_2 (279 mg, 1.1 mmol) was added to a solution of PPh_3 (317 mg, 1.2 mmol) and imidazole (187 mg, 2.8 mmol) in 8.0 mL DCM at 0 °C. *tert*-butyl-(4-hydroxybutyl)carbamate (**10**) (208 mg, 1.1 mmol) was added to the mixture at 0 °C and the mixture was stirred for 4 hours at rt. 15 mL sat. aq. Na_2SO_3 was added and the resulting mixture stirred for 30 min. The aqueous solution was extracted with 30 mL DCM . The organic layer was washed with 30 mL sat. aq. Na_2SO_3 (30mL) and brine, then dried over Na_2SO_4. The solvent was removed under reduced pressure, and flash chromatography (SiO_2, PE:EA 7:3) afforded the desired *tert*-butyl (4-iodobutyl)carbamate (**11**) as yellow oil (200 mg, 0.67 mmol, 61%), which coincided with literature.[5]
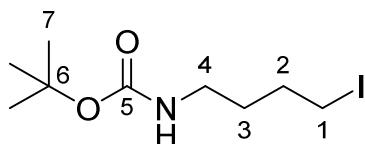
^{1}H NMR (500 MHz, $CDCl_3$): δ = 1.42 (s, 9 H, 7-H), 1.54 - 1.62 (m, 2 H, 2-H), 1.78 - 1.84 (m, 2 H, 3-H), 3.12 (bq, 2H, 1-H), 3.16 - 3.21 (m, 2H, 4-H), 4.63 (bs, NH)

^{13}C NMR (125 MHz, $CDCl_3$): δ = 6.2 (s, 1 C, 1-C), 28.3 (s, 3 C, 7-C), 30.6 (s, 1 C, 2-C), 30.9 (s, 1 C, 3-C), 39.3 (s,1 C, 4-C), 79.1 (s, 1 C, 6-C), 155.8 (s, 1 C, 5-C)

di-*tert*-butyl (azanediylbis(butane-4,1-diyl))dicarbamate (13a)

A mixture of 1.0 eq. *tert*-butyl (4-iodobutyl)carbamate (**11**) (75 mg, 0.25 mmol), 1.0 eq. *tert*-butyl (4-aminobutyl)carbamate (**12a**) (47 mg, 0.25 mmol) and 2.0 eq. K_2CO_3 (69 mg, 0.50 mmol) in 2.5 mL DMF was stirred under N_2 at rt for 3 hours. The mixture was poured into 50 mL ice-water and extracted with
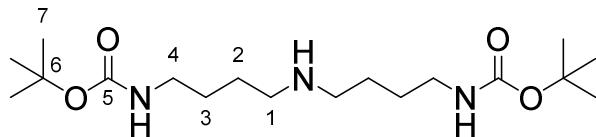
30 mL EtOAc (four times). The combined organic layers were washed with water, dried over Na_2SO_4 and concentrated in vacuum. Rapid chromatography on a short column of silica gel (SiO_2, $CHCl_3$:MeOH:TEA 80:18:2) gave the desired product **13a** (74 mg, 0.2 mmol, 82%) as a pale yellow oil.

^{1}H NMR (500 MHz, $CDCl_3$): δ = 1.44 (s, 18 H, 7-H), 1.49 - 1.55 (m, 4 H, 2-H), 1.49 - 1.55 (m, 4 H, 3-H), 2.60 (bt, *J* = 6.5, 4H, 1-H), 3.07 - 3.17 (m, 4H, 4-H), 4.88 (bs, NH)

^{13}C NMR (125 MHz, $CDCl_3$): δ = 27.4 (s, 2 C, 2-C), 27.9 (s, 2 C, 3-C), 28.4 (s, 6 C, 7-C), 40.4 (s, 2 C, 4-C), 49.5 (s, 2 C, 1-C), 79.0 (s, 2 C, 6-C), 156.0 (s, 2 C, 5-C)

HRMS (ESI):   360.2830 *m/z* calculated for [M+H]$^+$; found 360.2850 (Δ 1.8 ppm).

di-*tert*-butyl (methylazanediylbis(butane-4,1-diyl))dicarbamate (13b)

A mixture of 1.0 eq. *tert*-butyl (4-iodobutyl)carbamate (**11**) (57 mg, 0.19 mmol), 1.0 eq. *tert*-butyl (4-methyl,4-aminobutyl)carbamate (**12a**) (36 mg, 0.19 mmol) and 2.0 eq. K_2CO_3 (52 mg, 0.38 mmol) in 1.9 mL DMF was stirred under N_2 at RT for 3 hours. The mixture was poured into 50 mL ice-water and extracted with 30 mL EtOAc (four times). The combined organic layers were washed with water, dried over Na_2SO_4 and concentrated under reduced pressure. Rapid chromatography on a short column of silica gel ($CHCl_3$:MeOH: TEA 80:18:2) gave the product **13b** (67 mg, 0.18 mmol, 94%) as a pale yellow oil.
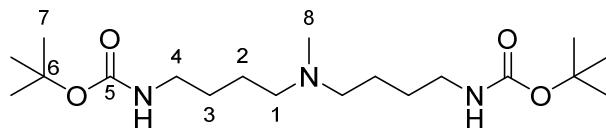
^{1}H NMR (500 MHz, $CDCl_3$): δ = 1.44 (s, 18 H, 7-H), 1.47 - 1.52 (m, 4 H, 2-H), 1.47 - 1.52 (m, 4 H, 3-H), 2.33 (bt, *J* = 6.2, 4H, 1-H), 3.08 - 3.17 (m, 4H, 4-H), 5.02 (bs, NH)

^{13}C NMR (125 MHz, $CDCl_3$): δ = 27.9 (s, 2 C, 3-C), 24.6 (s, 2 C, 2-C), 28.4 (s, 6 C, 7-C), 40.5 (s, 2 C, 4-C), 57.4 (s, 2 C, 1-C), 78.9 (s, 2 C, 6-C), 156.0 (s, 2 C, 5-C)

HRMS (ESI):   374.3013 *m/z* calculated for [M+H]$^+$; found 374.3012 (Δ 0.3 ppm).

Sandaramine D (4)

According to **GP A**, alcohol **8** (39 mg, 0.18 mmol) was reacted with DMP (150 mg, 0.35 mmol) in 1.7 mL DCM. Reductive amination, utilizing nonane-1,9-diamine (14 mg, 0.09 mmol) and NaBHOAc₃ (188 mg, 0.89 mmol) resulted in N_1,N_9-bis(((1S,8aS)-2,5,5,8a-tetramethyl-1,4,4a,5,6,7,8,8a-octahydronaphthalen-1-yl)methyl)nonane-1,9-diamine (**4**) (7 mg, 0.01 mmol, 14%) as a pale yellow oil.
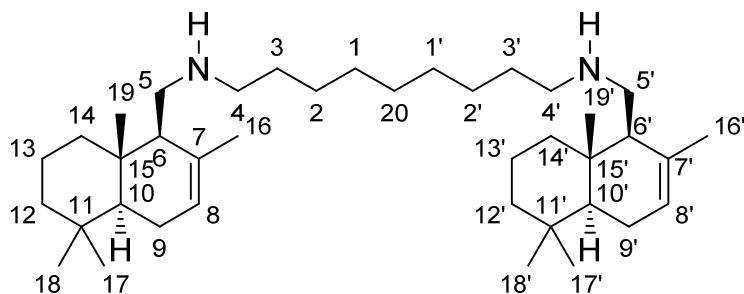
HRMS (ESI):   567.5612 *m/z* calculated for [M+H]⁺; found 567.5608 (Δ 0.6 ppm), 284.2842 *m/z* calculated for [M+2H]²⁺; found 284.2838 (Δ 1.7 ppm).

UHPLC (6 min gradient): 5.21 min

Table S9 NMR spectroscopic data for synthetic sandaramine D (**4**) in chloroform-*d* at 500/125 MHz.

#	δ ¹³C [ppm]	δ ¹H [ppm], mult (*J* [Hz])	COSY	HMBC
1, 1'	29.7	1.30, 4H, m*	2/2', 3/3', 20/20'	2/2', 3/3', 20/20'
2, 2'	27.5	1.30, 4H, m*	1/1', 3/3'	1/1', 3/3', 4/4', 20/20'
3, 3'	30.0	1.49, 4H, m*	2/2', 4/4'	1/1', 2/2', 4/4'
4, 4'	50.5	2.54/2.64, 4H, m*	3/3', 4/4', 5/5'	2/2', 3/3', 5/5'
5, 5'	48.4	2.72/2.49, 4H, dd (12.2, 7.5)	4/4', 5/5', 6/6'	4/4', 6/6', 7/7'
6, 6'	55.3	1.80, 2H, m	5/5', 8/8', 16/16',	7/7', 8/8', 10/10', 14/14'
7, 7'	134.4	-	-	-
8, 8'	123.2	5.45, 2H, bdt	6/6', 9/9', 10,10' 16/16'	6/6', 9/9', 10/10', 16/16'
9, 9'	23.9	1.86/1.98, 4H, m*	8/8', 9/9', 10/10' 16/16'	7/7', 8/8', 10/10', 15/15'
10,10'	50.1	1.20, 4H, m*	8/8', 9/9', 12/12'	9/9', 11/11', 12/12', 15/15', 19/19'
11, 11'	33.2	-	-	-
12, 12'	42.4	1.18/1.42, 4H*, m*	12/12', 13/13', 14/14', 18/18'	10/10', 11/11', 12/12', 13/13', 14/14', 17/17'
13, 13'	19.0	1.54/1.46, 4H, dt (13.6, 3.1)	12/12', 14/14'	11/11', 12/12', 14/14', 15/15', 17/17'
14, 14'	39.5	1.09/1.94, 4H td (13.1, 3.8)	13/13', 14/14', 19/19'	10/10', 12/12', 13/13', 15/15', 19/19'
15, 15'	36.4	-	-	-

#	δ 13C [ppm]	δ 1H [ppm], mult (J [Hz])	COSY	HMBC
16, 16'	22.2	1.74, 6H, bs	6/6', 8/8'	6/6', 7/7', 8/8', 9/9'
17, 17'	33.4	0.86, 6H, s	12/12', 18/18'	10/10', 11/11', 12/12', 18/18'
18, 18'	22.2	0.88, 6H, s	10/10', 12/12', 17/17'	10/10', 11/11', 12/12', 17/17'
19, 19'	14.2	0.76, 6H, s	6/6', 10/10', 14/14'	6/6', 10/10', 11/11', 14/14', 15/15'
20	29.7	1.30, 6H, bs*	1/1', 2/2'	1/1', 2/2', 3/3'

* overlapping signals

Sandaramine A (1)

According to **GP B**, alcohol **8** (4 mg, 18 µmol) was reacted with DMP (15 mg, 35 µmol) in 1.7 mL DCM. Simultaneously di-*tert*-butyl (azanediylbis(butane-4,1-diyl))dicarbamate (**13a**) (3 mg, 9 µmol) was treated with HCl (23 µL, 90 µmol, 4 M in dioxane) for deprotection. Reductive amination, utilizing the HCl salt of **14a**, triethylamine (7.5 µL, 5 µmol) and NaBH$_4$ (1.7 mg, 45 µmol) resulted in sandaramine A (**1**) (8.60 mg, 15 µmol, 76%) as a pale yellow oil.
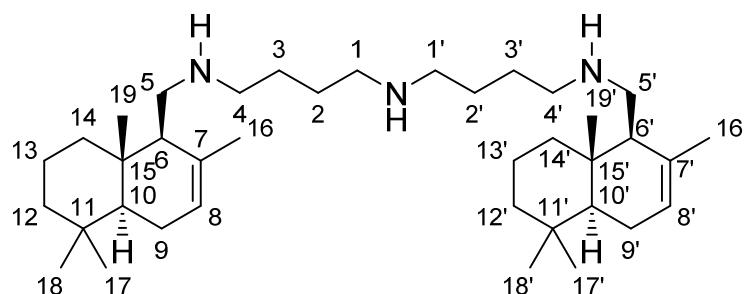
HRMS (ESI):   568.5564 *m/z* calculated for [M+H]$^+$; found 568.5558 (Δ 1.2 ppm), 284.7819 *m/z* calculated for [M+2H]$^{2+}$; found 284.7816 (Δ 1.1 ppm).

UHPLC (6 min gradient): 4.54 min

Table S10 NMR spectroscopic data for synthetic sandaramine A (**1**) in chloroform-*d* at 700/175 MHz.

#	δ 13C [ppm]	δ 1H [ppm], mult (J [Hz])	COSY	HMBC
1, 1'	49.2	2.69, 4H, m*	2/2', 3/3', 4/4'	1/1', 2/2', 3/3', 4/4'
2, 2'	27.7	1.55, 4H, m*	1/1', 3/3', 4/4'	1/1', 3/3', 4/4'
3, 3'	27.7	1.55, 4H, m*	1/1', 2/2', 4/4'	1/1', 2/2', 4/4'
4, 4'	49.9	2.59, 4H, m	1/1', 2/2', 3/3'	1/1', 2/2', 3/3', 5/5'
5, 5'	48.1	2.47/2.71, 4H, dd (12.3, 7.4)*	5/5', 6/6'	4/4', 6/6', 7/7', 15/15'
6, 6'	55.1	1.75, 2H, bs	5/5'	7/7', 8/8', 9/9'
7, 7'	133.9	-	-	-
8, 8'	123.2	5.44, 2H, bdt	9/9', 16/16'	6/6', 9/9', 10/10', 16/16'
9, 9'	23.7	1.98/1.85, 4H, m	8/8', 9/9', 10/10' 16/16'	6/6', 7/7', 8/8', 10/10', 15/15'
10,10'	49.8	1.18, 4H, m	8/8', 9/9', 12/12'	6/6', 7/7', 8/8', 9/9', 11/11', 12/12', 13/13', 14/14', 15/15', 17/17', 18/18'

#		δ ¹³C	δ ¹H	COSY	HMBC

11, 11'	32.9	-		-	-
12, 12'	42.2	1.16/1.41, 4H*, m*	12/12', 13/13', 18/18'	10/10', 11/11', 13/13', 14/14', 17/17', 18/18'
13, 13'	18.7	1.46, 4H*, m*	12/12', 14/14'	11/11', 12/12', 14/14', 15/15'
14, 14'	39.3	1.06/1.92, 4H, btd (12.8)	13/13', 14/14', 19/19'	6/6', 10/10', 12/12', 13/13', 15/15', 19/19'
15, 15'	36.2	-	-	-
16, 16'	22.0	1.72, 6H*, bs	8/8'	6/6', 7/7', 8/8', 9/9'
17, 17'	33.2	0.86, 6H, s	10/10', 18/18'	10/10', 11/11', 12/12', 13/13', 15/15', 18/18'
18, 18'	21.9	0.87, 6H, s	12/12', 17/17'	10/10', 11/11', 12/12', 17/17'
19, 19'	14.0	0.76, 6H, s	6/6', 14/14'	6/6', 10/10', 14/14', 15/15'

* overlapping signals

Sandaramine C (3)

According to **GP B**, alcohol **8** (4 mg, 18 µmol) was reacted with DMP (15 mg, 35 µmol) in 1.7 mL DCM. Simultaneously di-tert-butyl (methylazanediylbis(butane-4,1-diyl))dicarbamate (**13b**) (3 mg, 9 µmol) was treated with HCl (23 µL, 90 µmol, 4 M in dioxane) for deprotection. Reductive amination, utilizing the HCl salt of **13b**, triethylamine (7.5 µL, 5 µmol) and NaBH$_4$ (1.7 mg, 45 µmol) resulted in sandaramine C (**3**) (4.7 mg, 8 µmol, 45%) as a pale yellow oil.
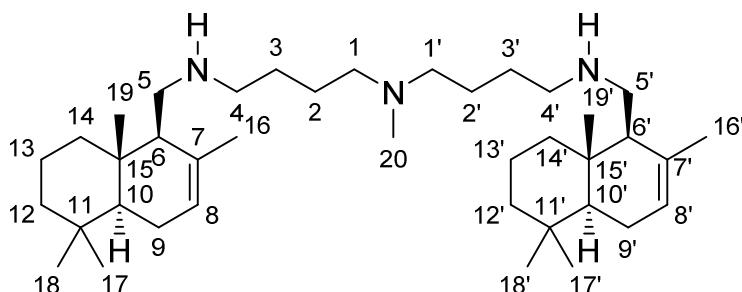
HRMS (ESI): 582.57208 *m/z* calculated for [M+H]⁺; found 582.57121 (Δ 1.5 ppm).

UHPLC (9 min gradient): 4.59 min

Table S11 NMR spectroscopic data for synthetic sandaramine C (**3**) in chloroform-*d* at 500/125 MHz.

#	δ ¹³C [ppm]	δ ¹H [ppm], mult (*J* [Hz])	COSY	HMBC
1, 1'	56.4	2.50, 4H, m	2/2'	2/2', 3/3', 20/20'
2, 2'	23.6	1.63, 4H, m*	1/1', 3/3'	3/3'
3, 3'	23.6	2.08, 4H, m*	2/2', 4/4'	1/1', 2/2', 4/4'
4, 4'	49.0	3.10/3.02, 4H, m*	3/3', 5/5'	2/2', 3/3', 4/4', 5/5', 6/6'
5, 5'	46.9	2.87/3.02, 4H, dd (13.4, 7.6)	4/4', 5/5', 6/6'	4/4', 5/5', 6/6', 7/7'
6, 6'	51.8	2.30, 2H, m	5/5', 8/8', 14/14', 16/16',	5/5', 7/7', 8/8'
7, 7'	131.1	-	-	-

8, 8'	124.8	5.46, 2H, m	6/6', 9/9', 10,10' 16/16'	6/6', 9/9', 10/10', 16/16'
9, 9'	23.8	1.80/1.97, 4H, m*	8/8', 9/9', 10/10' 16/16'	6/6',7/7', 8/8', 10/10', 15/15'
10,10'	49.3	1.24, 4H, m*	8/8', 9/9'	9/9', 11/11', 12/12', 15/15', 19/19'
11, 11'	33.2	-	-	-
12, 12'	42.0	1.15/1.36, 4H*, m*	12/12', 13/13', 18/18'	10/10', 11/11', 12/12', 13/13', 14/14', 17/17'
13, 13'	18.8	1.49, 4H, m*	12/12', 14/14'	11/11', 12/12', 14/14', 15/15', 17/17'
14, 14'	39.2	1.28/1.80, 4H, m*	13/13', 14/14', 19/19'	10/10', 11/11', 12/12', 13/13', 14/14', 15/15', 19/19'
15, 15'	36.6	-	-	-
16, 16'	22.6	1.92, 6H, bs	6/6', 8/8'	6/6', 7/7', 8/8'
17, 17'	33.1	0.82, 6H, s	12/12', 18/18'	10/10', 11/11', 12/12', 18/18'
18, 18'	22.0	0.83, 6H, s	12/12', 17/17'	10/10', 11/11', 12/12', 17/17'
19, 19'	14.0	0.70, 6H, s	5/5', 6/6', 10/10', 14/14'	6/6', 10/10', 11/11', 14/14', 15/15'
20	40.4	2.19, 6H, bs	-	1/1'

* overlapping signals
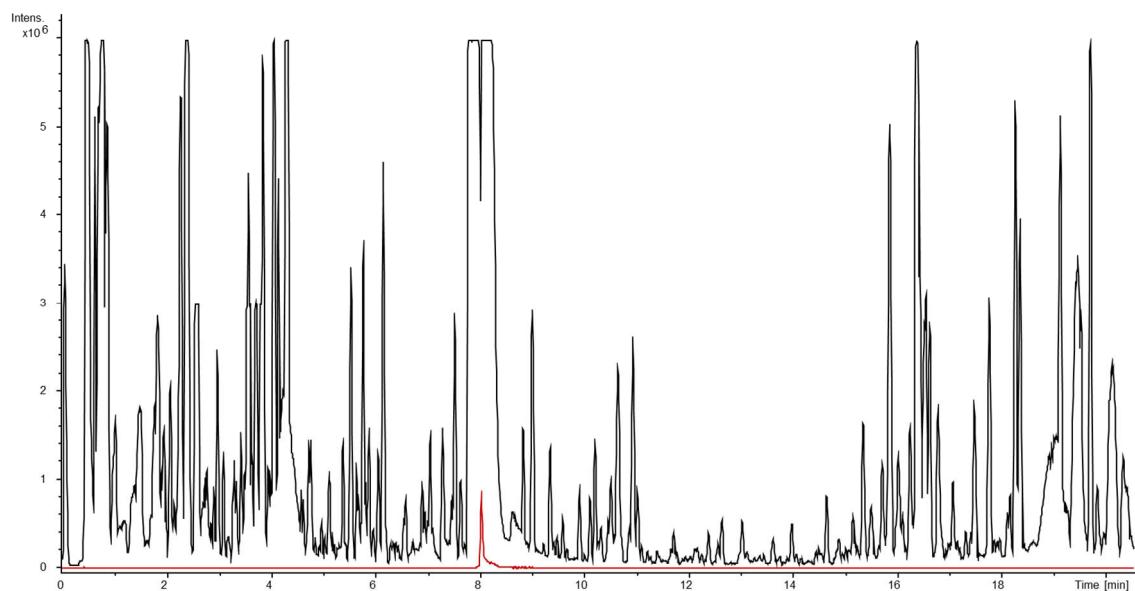
Figure S4 Base peak chromatogram (BPC) of MSr10575 crude extract and extracted ion chromatogramm (EIC) at 582.5719 ± 0.02 Da for [M+H]$^+$ of sandaramine C. The two overlaying peaks in the BPC display sandaramine A and B.

S5.5 Biological activity evaluation of the sandaramines

S5.5.1 Determination of antimicrobial activities

P. aeruginosa PA14 DSM50071, *Staphylococcus aureus* Newman, *Candida albicans* DSM-1665 and *Mycobacterium smegmatis* were obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ) or were part of our internal strain collection. 20 µL of a 960 µg/mL compound stock solution in methanol was mixed with 280 µL of an overnight culture or plate culture of the test organism which is scraped of and diluted to approximately 10^6 CFU. The mixture was serially diluted with MHB Medium for *P. aeruginosa* and *S. aureus*, M7H9 medium for *M. smegmatis* or Myc 2.0 medium for *C. albicans* in a 96 well plate, using 150 µL dilution 1 and 150 µL medium respectively, discarding the remaining 150 µL after the last dilution step. Plates were sealed with parafilm and incubated at 37 °C or 30 °C for *C. albicans* respectively. After 18 h the minimal inhibitory concentration (MIC) was determined by manual inspection. MIC values were determined in duplicates with 64 µg/mL as highest concentration determined and solvent as internal control. As positive control for *P. aeruginosa* Ciprofloxacin-HCl (MIC 0.01-0.05 µg/mL), for *S. aureus* Vancomycin (MIC 1 µg/mL), for *M. smegmatis* Rifampicin (MIC 16 µg/mL) and for *C. albicans* Amphotericin B (MIC 0.2 µg/mL) was used.

S5.5.2 Cytotoxicity Assay with HCT-116 cells

The HCT-116 cell line was obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ) and was cultured under conditions recommended by the depositor. Cells were grown and diluted to 6×10^3 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 the sandaramine dervatives in methanol. The assay was performed in duplicates for each tested compounds with solvent as internal control. 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 measuring the absorbance at 570 nm with a microplate reader (Tecan Infinite M200Pro). Sigmoidal curve fitting was used to determine the IC_{50} values based on the absorbance measurements.

S5.5.3 DNA Ladder Assay

Lysosomal apoptosis induction was determined by DNA ladder assay adapted from Saadat *et al..*[6] Human HCT-116 colon carcinoma cells (ACC-581) were cultivated in a 6 well plate containing 2 mL 90% modified McCoy's 5A medium with 10% h.i. fetal bovine serum (FBS) each. After 48 h of equilibration, the cells were treated with the compounds in MeOH at 2 x MIC. As negative control pure MeOH and as positive control 500 µM H_2O_2 was used. After 24h incubation at 37 °C, floating apoptotic cells were harvested by pipetting the supernatant into tubes and centrifuging the supernatant at 5000 rpm for 5 minutes. The supernatant was discarded and the empty wells containing adherent cells were incubated with 500 µL lysis buffer. After 10 minutes of incubation at room temperature, the lysate cells from the wells were pipetted into the respective tubes. The combined cells in lysis buffer were incubated at 65 °C for 5 minutes. After cooling to room temperature, 700 µL chloroform-isoamyl alcohol was added and centrifuged at 12000 rpm for 5 minutes. The aquatic upper phase was transferred into fresh tubes and 100 µL of 3.9M $(NH_4)_2SO_4$ solution was added to precipitate residual protein and centrifuged at 12000 rpm for 5 minutes. The pellet is discarded and 800µL of Isopropanol at -20°C were added to the supernatant in a fresh tube to

precipitate the DNA. The tubes were inverted gently before another centrifugation step at 12000 rpm for 5 minutes. The supernatant was discarded and the pellet air-dried for 30 minutes. The dry DNA was dissolved in 50 µL distilled water. The samples were separated by electrophoresis on a 1.5% agarose gel containing 1 µL/100 mL Roti-Safe DNA gel stain (Roth). The gel was examined by ultraviolet gel documentation.

Figure S5 Gel documentation picture of the apoptosis assay, pos indicating H_2O_2 positive control, neg indicating negative control, x belonging to assays not related to sandaramines that were not cropped to show integrity of the agarose gel and lanes 1, 2 and 3 belonging to sandaramine A and C respectively. Lad refers to the 1kb DNA ladder purchased from Thermo scientific.

All lanes displaying treatment of cells with sandaramines do not show extensive DNA degradation. Therefore, sandaramines do not seem to induce apoptosis via the lysosomal pathway as their DNA is mostly intact as in the negative control.

S5.5.4 Antiviral activity against CoV229E

Firefly luciferase-expressing Huh-7.5/FLuc cells were infected with a Renilla luciferase reporter virus of the alphacoronavirus HCoV 229E one day after seeding (2x10e4 cells/well in a 96 well plate) in the presence of indicated concentrations of the inhibitor/drug/chemical. 48h after inoculation and incubation of the cells at 33°C and 5% CO2, the virus inoculum was removed, cells were washed twice in 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. 20 µl of the lysate was used to measure cell viability via the firefly luciferase signal and 20 µl of the lysate was used to analyse virus replication/infection efficiency via the renilla luciferase signal respectively. Mean values and standard deviations of triplicate measurements are depicted. As positive control K22 and as negative control DMSO was used. Antiviral activity was distinguished from cytotoxic activity targeting <6.5% CoV229E replikation while increasing the cell viability with >150% as optimum.

References

(1) Hoffmann, M.; Auerbach, D.; Panter, F.; Hoffmann, T.; Dorrestein, P. C.; Müller, R. Homospermidine Lipids: A compound class specifically formed during fruiting body formation of *Myxococcus xanthus* DK1622. *ACS Chem. Biol.* **2018**, *13*, 273–280.

(2) O'Leary, N. A.; Wright, M. W.; Brister, J. R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.*; et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**, *44*, D733-45.

(3) Christianson, D. W. Structural and Chemical Biology of Terpenoid Cyclases. *Chemical reviews* **2017**, *117*, 11570–11648.

(4) Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. NMR Chemical Shifts of Common Laboratory Solvents as Trace Impurities. *J. Org. Chem.* **1997**, *62*, 7512–7515.

(5) Crich, D.; Rahaman, M. Y. Thiomaleic anhydride: a convenient building block for the synthesis of alpha-substituted gamma- and delta-lactones through free-radical addition, nucleophilic ring opening, and subsequent thiocarboxylate manipulation. *The Journal of organic chemistry* **2009**, *74*, 6792–6796.

(6) Rahbar Saadat, Y.; Saeidi, N.; Zununi Vahed, S.; Barzegari, A.; Barar, J. An update to DNA ladder assay for apoptosis detection. *BioImpacts : BI* **2015**, *5*, 25–28.

Chapter 6

The Sandarazols Are Cryptic and Structurally Unique Plasmid Encoded Toxins from a Rare Myxobacterium

Fabian Panter[1], Chantal D. Bader[1] and Rolf Müller[1]

Affiliations

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

Contributions and Acknowledgements

Author's effort:

The author significantly contributed to this manuscript by performing the structure elucidation of the sandarazols by generating NMR spectra and their evaluation. The author furthermore contributed to elucidation of the molecules stereochemistry by performing the Moshers ester reaction and interpreting their results. The author contributed by writing and editing this manuscript.

Contributions by others:

Fabian Panter did the conception of this study, designed and performed experiments, evaluated and interpreted resulting data. Fabian Panter generated the overexpression mutants and isolated the sandarazols. He elucidated their stereochemistry and developed a biosynthesis hypothesis. Furthermore, Fabian Panter wrote this manuscript. Stefanie Schmidt and Alexandra Amann conducted the antimicrobial profiling and the cytotoxicity assays. Nestor Zaburannyi assembled the Genome of MSr10575 and Daniel Krug contributed by revision of the manuscript. Rolf Müller contributed to conception and supervision of this study, as well as editing and proofreading of the manuscript.

6.1 Abstract

Soil dwelling bacteria such as myxobacteria outcompete other microorganisms by using growth inhibitory secondary metabolites. In this work, we describe a new plasmid found in *Sandaracinus* sp. MSr10575 named pSa001 spanning 209.7 kbp that harbours a cryptic secondary metabolite biosynthesis gene cluster (BGC). Activation of this BGC by homologous recombination mediated exchange of the native promoter sequence against a vanillate inducible system led to production and subsequent isolation and structure elucidation of novel secondary metabolites, the sandarazols A-G. The sandarazols contain intriguing structural features and very reactive functional groups such as an α-chlorinated ketone, an epoxyketone and a (*2R*)-2-amino-3-(*N,N*-dimethylamino)-propionic acid building block. In depth investigation of the underlying biosynthetic machinery led to a concise biosynthetic model for the new compound family, including several uncommon biosynthesis steps. The chlorinated congener sandarazol C shows an IC_{50} value of 0.5 µM against HCT 116 cells and a MIC of 14 µM against *Mycobacterium smegmatis*, which points at the sandarazols' potential function as defensive secondary metabolites or toxins.
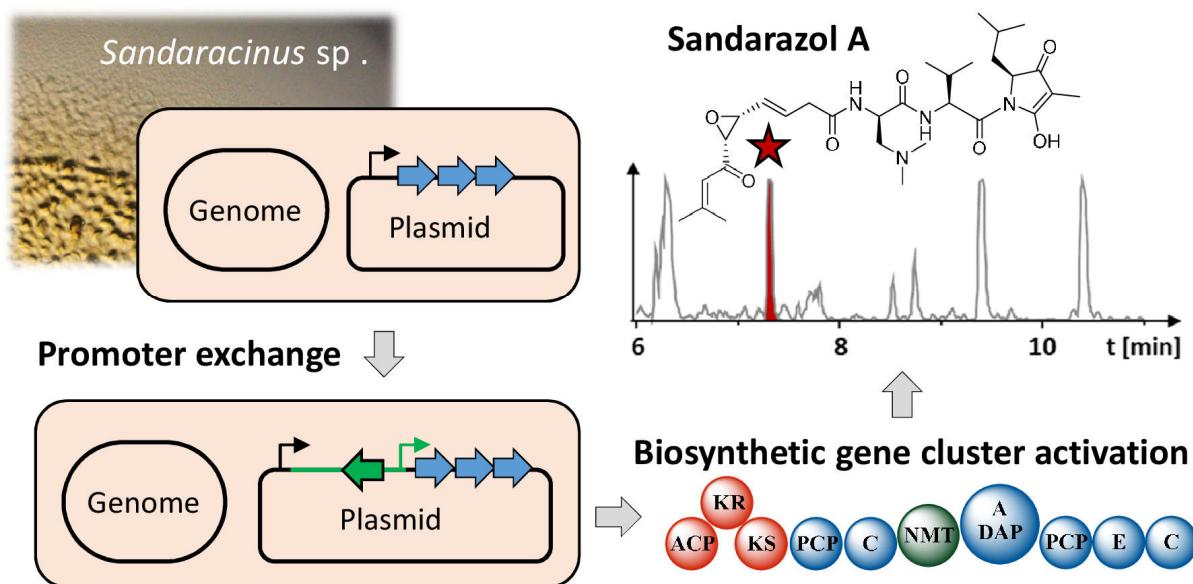

Figure 1 Table of contents graphic showing the principle design of the study.

6.2 Introduction

Bacterial antibiotic resistance genes cause antimicrobial resistance (AMR) in pathogens thereby leading to enormous challenges in the treatment of infectious diseases in the clinics. [1] AMR genes are often encoded on plasmids, horizontally transferrable elements employed by bacteria to survive antibiotics treatment and effectively spread AMR. [2] In addition to that, AMR genes co-localized with antibiotic biosynthesis gene clusters (BGCs) are important for self-resistance during antibiotic production in bacteria capable of respective biosyntheses. Similarly to plasmid mediated exchange of antibiotic resistance, bacteria are able to gain an advantage over competing microorganisms by exchanging secondary metabolite BGCs in horizontal gene transfer events. [3] The mechanism for plasmid mediated exchange of secondary metabolite BGCs is known for plasmid-encoded toxins, with the anthrax causing *Bacillus anthracis* as the most prominent example. [4] The *Bacillus* plasmid is shared between different bacilli and thereby transmits the anthrax toxin virulence factor. Exchange mechanisms of such plasmids that encode toxins enable the respective bacteria to acquire the ability for chemical warfare. As bacteria are competing with other microbes for nutrients in their ecological niches, the encoded chemical warfare molecules optimized by evolution are attractive targets in the search for novel bioactive natural products. For actinobacteria it has already been shown that defensive secondary metabolites can be encoded as multimodular BGCs on autonomously replicating plasmids. A BGC responsible for production of the highly cytotoxic mycolactones for example is also encoded on a large plasmid. [5] *Mycobacterium ulcerans* produces this cytotoxin and releases it during the infection process of human skin, while the bacterium feeds on skin cells. Streptomycetes also host plasmid borne antibiotics BGCs as exemplified by the esmeraldins, a series of phenazine antibiotics. [6] With respect to the total output of natural products and natural product diversity, bacteria have emerged as major players as they are prolific producers of biologically active secondary metabolites. [7,8] Besides the well-described bacterial phyla such as actinobacteria, firmicutes and cyanobacteria that are responsible for the majority of identified biologically active natural products to date, Gram-negative proteobacteria such as myxobacteria have also shown great promise for the discovery of novel bioactive secondary metabolites. [9–11] Especially novel and phylogenetically distant bacterial genera among the myxobacteria, such as the recently identified genus *Sandaracinus* studied herein, hold promise for finding interesting novel natural products chemistry. [12] In recent years, availability of cheap and reliable DNA sequencing technologies along with the development of BGC prediction tools such as antiSMASH depict the theoretical genetically encoded bacterial secondary metabolome as to date largely underexploited. [13,14] However, many BGCs that can be identified *in-silico* are 'cryptic' under laboratory conditions. In this case, the production of the corresponding secondary metabolite is completely repressed or remains below the detection limit of modern analytical tools such as liquid chromatography (LC) coupled mass spectrometry (MS) instrumentation. Activating such cryptic clusters by genetic tools holds promise for discovering new chemistry, but currently remains a non-automatable, labor-intensive process. [15] Therefore, a strong focus on rational prioritization of BGCs to be activated by heterologous expression or gene cluster overexpression using heterologous promoters remains crucial to focus discovery efforts on BGCs encoding for bioactive natural products. [16] Methods employed for BGC prioritization so far include expected target-guided mining for potential self-resistance genes and exploring biosynthetic gene cluster complexity for modular BGC architectures as well as the presence of certain tailoring enzymes such as halogenases or epoxidases. [17] Still, presence a BGC on a plasmid as an example for a BGC that is likely to be a mobile genetic element has not yet been used as a means for BGC prioritization. As part of our ongoing efforts to isolate taxonomically diverse myxobacteria we isolated a novel myxobacterial strain belonging to the *Sandaracinus* clade.

Surprisingly, we found that the strain's genome contains a circular 209.7 kbp plasmid called pSa001 that features a large polyketide synthase (PKS) non-ribosomal peptide synthetase (NRPS) hybrid BGC in its sequence. In this work, we show that activation of this BGC by genetic engineering of this strain leads to production of several derivatives of the sandarazols, a previously unknown family of natural products. These compounds turned out to be potent and chemically novel toxins biosynthesized by a series of intriguing biosynthetic steps.

6.3 Results and discussion

6.3.1 Biosynthetic gene cluster identification and activation

Myxobacteria are Gram-negative δ-proteobacteria with a rich and diverse secondary metabolism. [10,11] A survey of known and unknown natural products in 2300 myxobacteria showed structural novelty to be clearly correlated to phylogenetic distance and thus heavily reliant on in depth investigation of novel myxobacterial genera. [12] We thus chose to dive into the secondary metabolome of a myxobacterial strain called MSr10575 that was recently isolated in-house. The strain belongs to the *Sandaracinus* genus, a rare myxobacterial genus little studied for secondary metabolism with the type strain *Sandaracinus amylolyticus* NOSO-4T. [18] This strain was the only member in the Sorangineae clade until MSr10575 was isolated and provided the indiacens A and B, two prenylated indole secondary metabolites. [19] We chose to sequence the genome of strain MSr10575 using single molecule real time sequencing technology.[20] We subsequently annotated the BGCs in its sequence by antiSMASH to obtain an overview about the strains' secondary metabolite production potential. [13] Sequencing coverage analysis and genome assembly revealed that the genome of MSr10575 consists not only of a bacterial chromosome of 10.75 Mbp but also of an autonomously replicating plasmid of 209.7 kbp, which we named pSa001 (see SI). As sequencing coverage per base on the pSa001 plasmid is twice as large as observed for the bacterial chromosome, we assume a median value of two pSa001 copies per MSr10575 cell (see SI). The overall GC content of the plasmid is 70% and the codon usage bias favors GC rich codons over AT rich codons for the same amino acid at a rate of 3.6 to 1. As these values do not differ significantly from the corresponding parameters for the MSr10575 chromosome, the plasmid seems well adapted to its host strain (see SI). The presence of this plasmid is distinctive for this myxobacterial strain as autonomously replicating plasmids are rarely observed in myxobacteria. The only other example of a characterized myxobacterial plasmid is pMF1 from *Myxococcus fulvus*. [21] In comparison to pMF1, pSa001 is not only significantly larger, but also encodes a large type 1 *in-trans* acyl transferase (*trans*-AT) PKS-NRPS hybrid BGC, which was named sandarazol (*szo*) BGC (see SI). Besides the *szo* BGC, the plasmid contains five ORFs encoding putative transposases and one ORF encoding a putative integrase indicating its ability to integrate into (or transfer parts of its sequence into) foreign bacterial genomes after conjugation. It might thus serve as a BGC shuttle vector for horizontal gene transfer of the BGC. Cultivation of wild type MSr10575 followed by extraction and liquid chromatography – tandem mass spectrometry (LC-MS²) analysis and GNPS based spectral networking of the bacterial metabolome did not reveal a family of secondary metabolites matching the BGC architecture on the plasmid in expected secondary metabolite size and fragmentation pattern (see SI). [22] We therefore assumed the plasmid borne *szo* BGC to be 'cryptic' as seen in many other cases such as the pyxidicyclines or taromycin. [23,24] All the coding regions in the *szo* cluster spanning from *szoA* to *szoO* are encoded on the same DNA strand and the intergenic regions seemed too small to contain important elements other than ribosome binding sites. We therefore reasoned that the *szo* cluster is encoded as a single transcriptional unit even though it spans 44.5 kbp. Activating the *szo* BGC with a single promoter exchange in line with the experiments described in promoter exchange BGC

activation in myxobacteria seemed therefore a reasonable strategy to unlock the corresponding natural products. [23,25] As we set out to create a plasmid for *szo* cluster overexpression by single crossover promoter exchange, we chose a plasmid with a pBelobac replication origin, as this would allow the entire 209.7 kbp pSa001 to be extracted and transfered into *E. coli* as replicative plasmid for further investigations into the plasmids replication mechanism. [26] While extracting the pBeloBacSa001 into *E. coli* is possible, retransformation of this plasmid into a plasmid cured version of MSr10575 needs further optimization. To investigate the secondary metabolite products of the BGC, we chose BGC activation in the native host MSr10575 first, as it was assumed MSr10575 would be best suited to produce the corresponding small molecule product. We activated the *szo* cluster on the plasmid by promoter exchange against the vanillate promoter and repressor cassette, since this tool achieved overexpression of BGCs in several other myxobacteria including *Myxococcus xanthus* and *Pyxidicoccus fallax* (see SI). [23,27] Compared to other promoters, the vanillate promoter and repressor system shows tight control of BGC expression as well as strong secondary metabolite production upon BGC induction in myxobacteria. [28]
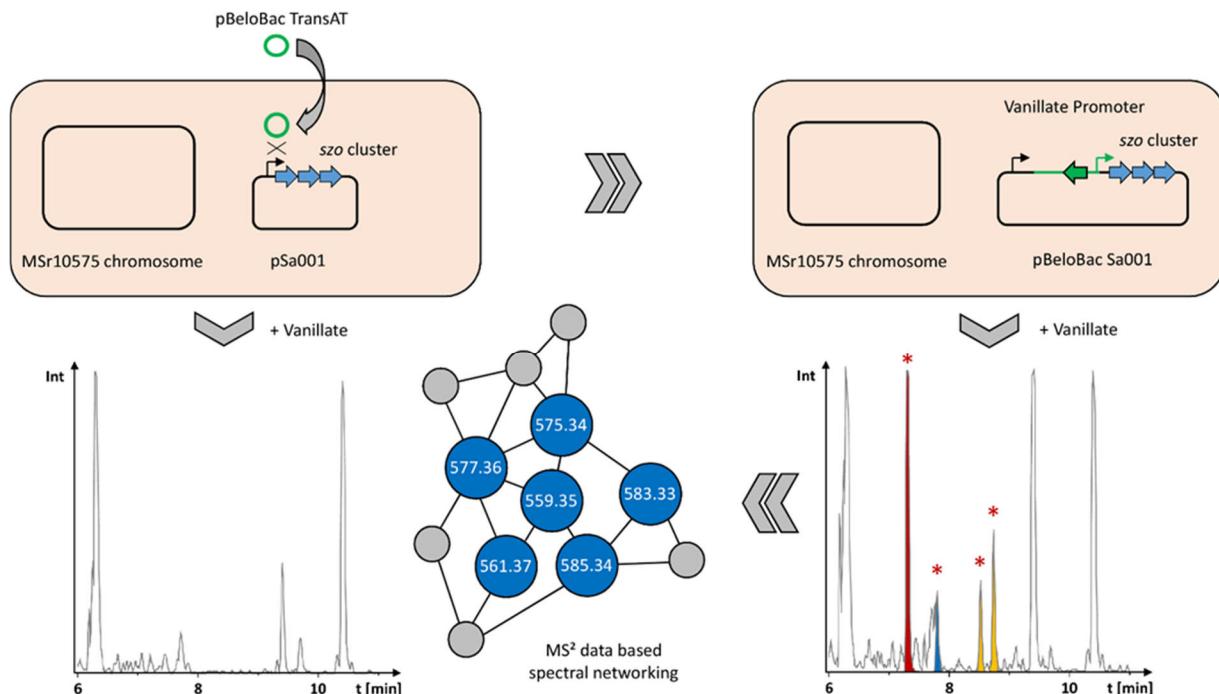
Figure 2 Schematic overview of the activation of the plasmid borne *szo* BGC in wild type *Sandaracinus sp.* MSr10575 by homologous recombination, the corresponding changes in the extracts' LC-MS chromatograms and LC-MS² based spectral networking for identification of the produced sandarazols.

To exchange the promoter of the *szo* cluster, the first 2 kbp of the *szo* BGC were PCR-amplified, fused to a vanillate promoter and repressor cassette and ligated into the commercial pBeloBac11. This plasmid featuring a BeloBac origin and a kanamycin resistance in its backbone was named pBeloBac-TransAT (see SI). Promoter exchange of the *szo* clusters' native promoter against the vanillate cassette is achieved by homologous recombination of pBeloBac-TransAT with the pSa001 plasmid after electroporation (Figure 2). MSr10575 clones harboring the recombined fused plasmid called pBeloBacSa001 were selected on 25 µg/mL kanamycin and genotypically verified by PCR (see SI). To obtain a comprehensive overview about the metabolic differences between MSr10575 wild type and MSr10575::pBeloBacSa001, analytical scale fermentation cultures were prepared in triplicates, extracted and subjected to UHPLC-qTOF analysis. All detected LC-MS features in the triplicate analyses were grouped into 2D retention time - exact mass windows, so-called buckets, and all MSr10575 wild type derived liquid chromatography – mass spectrometry (LC-MS) features were subtracted from the

detected LC-MS features using principal component analysis (PCA).[29,30] These features were then used for selective acquisition of LC-MS² data covering exclusively the mutant derived LC-MS features that are subsequently used for spectral networking in GNPS.[22] We thus obtained spectral networks comprising all LC-MS features linked to activation of the sandarazol BGC (see Figure 2, SI).

6.3.2 Isolation and structure elucidation

The novel secondary metabolites observed in this PCA-based analysis of LC-MS features were named the sandarazols. Besides sandarazol A (**1**; 575.344 Da [M+H]⁺; $C_{30}H_{47}N_4O_7$, Δ = 0.64 ppm), we were able to identify and characterize seven structural variants from appearance of their characteristic MS signals, approximately half of which showed the characteristic isotope pattern for chlorine containing compounds such as sandarazol C (**3**; 583.3256 Da [M+H]⁺; $C_{29}H_{48}ClN_4O_6$, Δ = 0.07 ppm). [31] The sandarazols A, B, C and F (**1,2,3** and **6**) were isolated from vanillate-induced large-scale cultures of MSr10575::pBeloBac Sa001 using a sequence of different techniques such as liquid/liquid extraction, countercurrent partition chromatography (CPC) and HPLC under N_2 atmosphere (see SI, Figure 3). It is worth noting that not even trace amounts of the sandarazols A-G (**1-7**) can be detected in the MSr10575 wild type extracts indicating the *szo* cluster to be most likely fully cryptic in the wild type strain under laboratory conditions applied in our study. The structures of sandarazol D,E (**4,5**) and G (**7**) were assigned by comparison of their MS² spectra to the other derivatives (see SI). As the compounds are sensitive to oxygen and/or strongly acidic or basic pH values, compound isolation and purification was performed in ammonium formate buffered eluents and under constant N_2 stream wherever possible (see Figure 3).
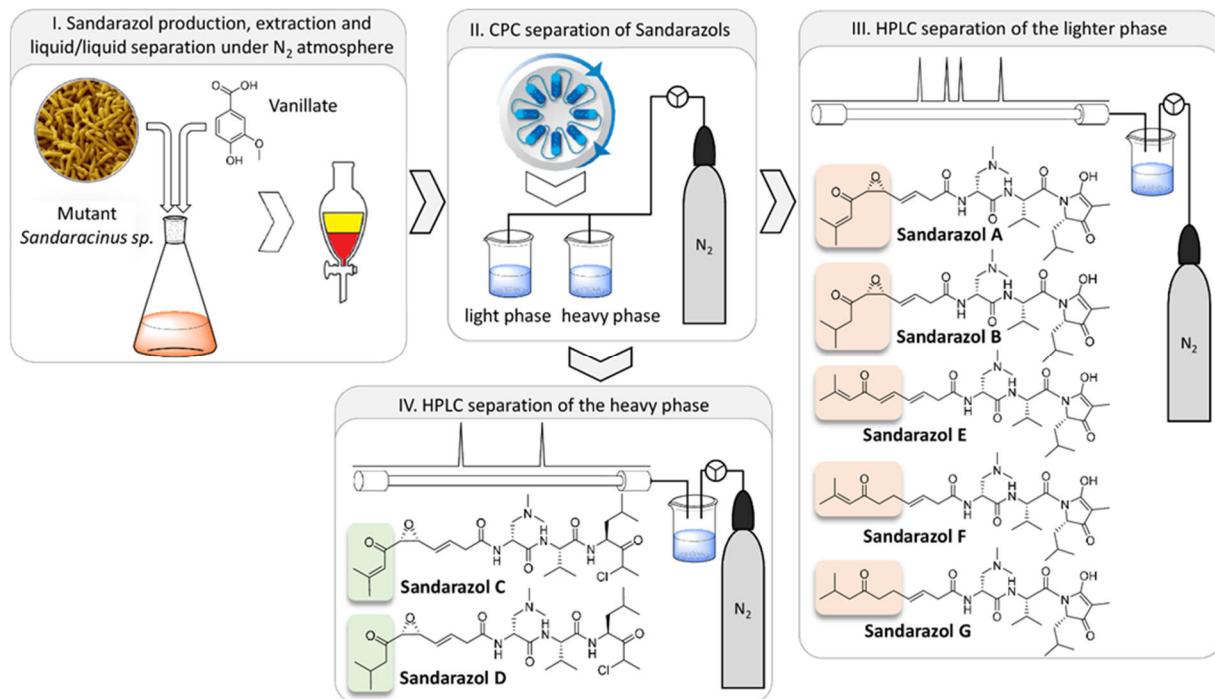
Figure 3 Workflow and isolation scheme for preparation of the sandarazols from *Sandaracinus* sp. MSr10575::pBeloBacSa001 including structure formulae for all structurally elucidated sandarazol derivatives.

¹H, ¹³C and HSQC-dept. spectra of **1** reveal two α-protons based on their characteristic chemical shift at δ^1H = 4.78 and 4.45 ppm (see Figure 4). COSY and HMBC correlations of the first methine group at δ^1H = 4.78 ppm to one methylene group, two methyl groups with a downfield $\delta^{13}C$ chemical shift of 45.2 ppm and a quaternary carbon with a characteristic amide shift, show that this α-proton is part of 2-amino-3-(*N,N*-dimethyl amino)-propionic acid (Me₂Dap). The second α-proton at δ^1H = 4.45 ppm

exhibits COSY and HMBC correlations to one methine group, two methyl groups and another amide function. The second amino acid of the molecule was determined to be a valine. HMBC correlations of the valine α-proton to the Me$_2$Dap carboxylic acid function suggests their connection via the valine N-terminus. Further HMBC correlations of the Me$_2$Dap α-proton to another amide function reveal N-terminal elongation by the polyketide part of the molecule. Characteristic chemical shifts of two protons at δ^1H = 3.47 ppm and 3.46 ppm with correlations to two vinylic double bond protons at one side and one epoxide on the other side, suggest **1** to contain an epoxyketone, as well as unsaturation in β,γ position relative to the amide bond. Typical J_{H-H} coupling values of 15.4 Hz for the vinylic protons at the double bond indicate *E* configuration. [32] 1D and 2D NMR spectra reveal a trisubstituted vinyl group is conjugated to the ketone side of the epoxyketone. The C-terminal end of valine show correlations to another methine group with characteristic chemical shifts close to an α-proton shift, suggesting further elongation of the molecule by another amino acid. This amino acid is determined to be leucine coupled to a propionate unit from 1D and 2D NMR data. The combination of a low field proton shift and high field carbon shift of an additional methyl group at δ^1H = 1.60 and δ^{13}C = 5.7 ppm in this part of the molecule, reveal heterocyclization of a propionate-extended valine cyclized to a tetramic acid, which also serves as connection between the valine and leucine building blocks (see Figure 3,4).
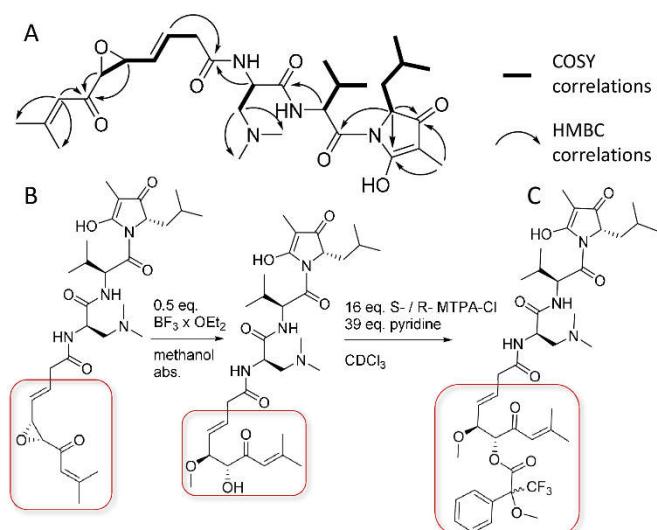
Figure 4 A) NMR correlations important for structure elucidation B) Reaction sequence to open the epoxide and form a methanol adduct C) Mosher's esterification of **1** used to elucidate the epoxide's configuration.

In contrast to **1**, HSQC spectra of **2** reveal a missing double bond in the terminal moiety of the polyketide part of the molecule (Figure 3). Sandarazol F differs in the polyketide part of the molecule, as HSQC spectra reveal the two methines in the epoxide motif to be replaced by two double bond protons. In **3**, however, the polyketide part equates the one of sandarazol A, but 1D and HSQC spectra show a missing tetramic acid moiety. The methyl group at δ^{13}C = 6.0 in **1**, is shifted to δ^{13}C = 20.7 ppm in **3**. In line with the *hr*MS spectra of **3**, chlorination of the molecule is observed at the α-position of the elongated leucine, confirmed by the characteristic chemical shifts and COSY as well as HMBC correlations of the surrounding methyl and methine groups.

All sandarazol derivatives contain five stereocenters, two of which are linked by the epoxide. The chlorinated derivatives **3** and **4** feature an additional stereocenter at the chlorinated carbon atom, which racemizes fast at room temperature (see SI). Its original stereo configuration can therefore not be determined as racemization occurs already under fermentation conditions. The stereo centers contained in the L-valine (L-Val), L-leucine (L-Leu) and (*2R*)-2-amino-3-(*N,N*-dimethyl amino)-propionic

acid (D- Me$_2$Dap) building blocks were confirmed by Marfey's analysis using commercially available standards (see SI). [33,34] The stereocenters in the polyketide part of the molecule are determined by the orientation of the epoxide. Direct configurative assignment was not possible, as established derivatization methods like the Mosher's method rely on reaction of functional groups like alcohols. [35] To tackle this issue, the epoxide was transformed into an alcohol by Lewis acid catalyzed addition of methanol to the epoxide (Figure 4). The resulting secondary alcohol inherits its stereochemistry from the epoxide due to retention of the epoxide's stereochemistry in the S$_N$2 based ring opening reaction (Figure 4). Subsequent Mosher's esterification of the secondary alcohol revealed the alcohol to be in *R*- configuration and thus the epoxide in sandarazols to be *R,R*- configured. With full structure elucidation, the sandarazols' absolute configuration as well as the *szo* biosynthetic gene cluster at hand we were able to devise a biosynthetic model for the sandarazols.

6.3.3 Biosynthesis of the sandarazols

Sandarazol biosynthesis starts with a *trans*-AT PKS starter module on SzoE that loads an acetate unit onto the first acyl carrier protein (ACP). The peculiarity of this first module is the presence of a β-branching ACP, which acts at the site where the β-branching cassette SzoK to SzoO adds a β-methyl branch to the first PKS extension (Figure 5 B). [36] In contrast to standard β-branching cassettes like PyxK to PyxO in the pyxipyrrolone BGC consisting of an ACP, a ketosynthase (KS), a hydroxymethylglutaryl–CoA synthase (HMG) and two enoyl-CoA dehydratases (ECH), the latter ECH domain is replaced with SzoO, a thioesterase (TE). [37] This enzyme is able to replace the decarboxylation function of the second ECH as described in the curacin biosynthesis. [38] While most branched chain end moieties in polyketide synthase are biosynthesized by incorporation of an isovaleryl- or an isoamyl- starter unit as in fulvuthiacene biosynthesis for example, bongkrekic acid biosynthesis contains a similar branched chain starter unit biosynthesis based on a β-branching cassette. [30,39] As the *szo* cluster does not encode enoyl reductase (ER) domains, the standalone short chain reductase protein *szoA* is likely responsible for the reduction steps leading to the production of the saturated tail group seen in the sandarazol variants **2**, **4** and **7**. The enzyme is also likely to be responsible for the other α,β enoyl reduction reaction next to the ketone in **6** and **7** (see Figures 3, 5). Module 2 is most likely either skipped or its ACP solely takes part in a transacylation reaction to forward the polyketide chain to module 3, as we see no polyketide extension in this module even though both the KS and the ACP in this module seem functional based on sequence alignment analysis. The following modules three to five incorporate three units of malonyl-CoA by successive decarboxylative Claisen condensation to attach 6 carbon atoms to the growing polyketide chain. In this process, we observe consecutive formation of two double bonds through β-ketoreduction by a ketoreductase (KR) and subsequent dehydration most likely performed by the in-*trans*-acting dehydratase SzoD in modules four and five. In-*trans*-acting dehydratases have already been described in *trans*-AT PKS pathways such as the CurE and CorN proteins from curacin and corallopyronin respectively. [40,41] The KR domain located in module 3 is most likely responsible for β-ketoreduction in module 4, as module 4 does not contain a KR domain and no β-ketoreduction is observed in module 3. Both double bonds formed in modules 4 and five are isomerized from an α,β-double bond to a β,γ-double bond by the double bond shifting DH$_2$ domain encoded on *szoF*. [42] Similarly to the double bond shifting domains in corallopyronins, the double bond shifting DH$_2$ on SzoF contains the first DH domain consensus motif but lacks the second DxxxQ consensus motif, a characteristic that is conserved in double bond shifting DH domains (see SI). [41,43] The subsequent NRPS module on *szoF* contains an adenylation (A) domain with a specificity code close to serine specificity according to NRPS predictor 2. This domain is assumed to accept (*2S*)-2,3-diamino-propionic acid (L-Dap) that is then incorporated into the nascent chain (Figure 5). [44] Biosynthesis of L-Dap is presumed

in analogy to the described pathway in *Staphylococcus aureus*. [45] The proteins carrying out the corresponding biosynthetic steps in MSr10575 are the SbnA homolog SzoB and the SbnB homolog SzoJ. SbnA is a pyridoxal phosphate (PLP)-dependent aminotransferase type enzyme that connects the amino group of glutamate to replace the phosphate moiety of an *O*-phosphoserine molecule from the bacterium's primary metabolism (see Figure 5 D). [46] SbnB is an NAD dependent oxidase that releases α-ketoglutarate from this intermediate and forms L-Dap. A SAM dependent *N*-methyl transferase domain in module 6 subsequently transfers two methyl groups to the free amino group of the attached L-Dap building block to form L- Me$_2$Dap, which is subsequently epimerized to D- Me$_2$Dap by the following epimerase domain (Figure 5). In contrast to most *N*-methyl transferase domains built into NRPS systems, which transfer a methyl group onto the α-nitrogen atom that will later form the peptide bond, this methylation reaction consecutively transfers two methyl groups to the nitrogen atom in 3 position of L-Dap. [47] The following two modules incorporate L-Val and L-Leu according to NRPS textbook logic. [48] Incorporation of all three amino acids was proven by feeding of stable isotope labelled precursors (see SI). Module 9 is a PKS module, that incorporates another malonyl-CoA unit by decarboxylative Claisen condensation and adds an α-methyl branch via a SAM-dependent C-methyl transferase domain to the molecule (Figure 5).

We did not find any non-chlorinated, non-cyclized sandarazols in the culture broth and cyclisation of an amide nitrogen with a carboxylic acid function to form the tetramic acid heterocycle is very unlikely to occur spontaneously. We therefore suspect that chlorination of sandarazols by the halogenase SzoJ occurs on the assembly line. This reaction is likely to govern, whether sandarazol is released from the megasynthase as a chlorinated open chain molecule prone to decarboxylation like seen in **3**, or as a non-chlorinated cyclized product like **1** (Figure 5). Halogenation by SzoJ occurs via the accepted FAD-depended halogenation mechanism similar to synthetic α-chlorination of β-keto acids. [49,50] This chlorinated ACP-bound intermediate is released from the assembly line by the TE and quickly loses its terminal carboxylic acid moiety by decarboxylation, a reaction occurring spontaneously in α-chlorinated β-keto acids. If chlorination does not occur, the molecule is subsequently cleaved off the assembly line by a TE domain with a much slower rate, thus cyclizing the product to form a 5-membered ring system as seen for example in **1** and **2** (see Figure 5). To finalize sandarazol biosynthesis, epoxidation needs to occur next to the ketone moiety in the polyketide part of the molecule. Catalysis of such a reaction is likely performed by FAD-dependent monooxygenase SzoP that putatively performs a reaction similar to the one performed by the FAD-dependent epoxidizing styrene monooxygenase. [51,52] We assume that sandarazols are finally exported into the surrounding medium by an ABC exporter of broad range specificity encoded somewhere else in the MSr10575 genome, as no specific exporter system is encoded within or in close proximity to the sandarazol BGC or elsewhere the pSa001 plasmid.
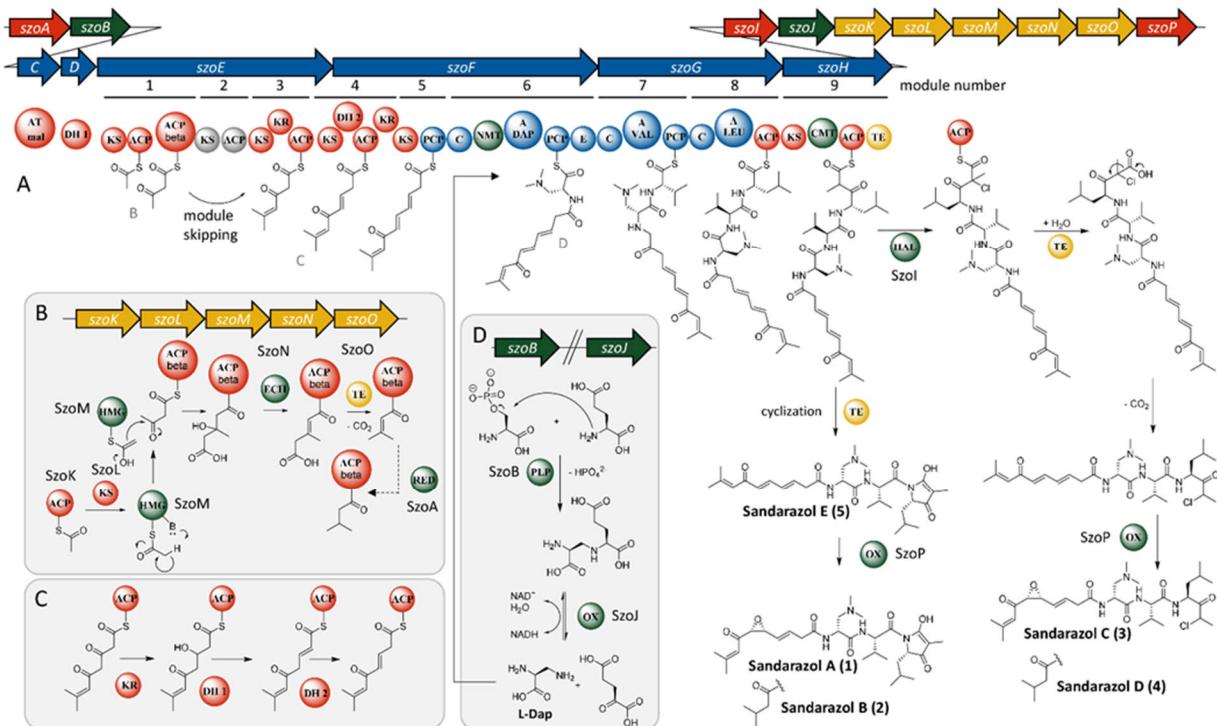
Figure 5 A) Scheme of the proposed sandarazol BGC and sandarazol biosynthesis by its megasynthase including the corresponding tailoring reactions B) β-branching reaction cascade to form the branched chain tail group C) Reaction cascade that leads to the formation of two isomerized double bonds D) Reaction sequence that supplies the amino acid L-Dap to module 6 of the assembly line (arrows: genes; circles: domains; A: adenylation; ACP/PCP: acyl-/peptidyl carrier protein; AT: acyl transferase; C: condensation; CMT/NMT: C/N methyl transferase; DH: dehydratase; E: epimerization; ECH: enoyl-CoA hydratase; HAL: halogenase; HMG: hydroxymethylglutaryl-CoA synthase; KR: ketoreductase; KS: ketosynthase; PLP: pyridoxal phosphate dependent enzyme; RED: reductase; TE: thioesterase; Ox: oxidation)

6.3.4 Biological activity of the sandarazols

Due to the low stability of the sandarazols in aerobic environments, antimicrobial and anti-proliferative activities were only determined for **1** and **3** as representatives for one chlorinated and one cyclized member of the sandarazol compound family respectively. Under the specified assay conditions, the non-chlorinated **1** displays only very limited biological activity, while the chlorinated **3** shows prominent cytotoxicity as well as antibiotic activity against *Mycobacterium smegmatis* and a variety of other indicator bacteria for Gram-positive pathogens (see Table 1).

Table 1 Antimicrobial and cytotoxic activities of sandarazol A (**1**) and C (**3**) as minimum inhibitory concentrations (MIC) and inhibitory concentrations at 50% inhibition (IC50).

Microbial strain	MIC 1	MIC 3
C. albicans	>110 µM	110 µM
P. anomala	>110 µM	110 µM
C. freundii	>110 µM	>110 µM
A. baumanii	>110 µM	>110 µM
S. aureus	110 µM	55 µM
B. subtilis	>110 µM	55 µM
E. coli	>110 µM	>110 µM
P. aeruginosa	>110 µM	55 µM
M. smegmatis	>110 µM	14 µM
Cell line	**IC$_{50}$ sandarazol A**	**IC$_{50}$ sandarazol C**
Human colon cancer HCT116	>64 µM	0.5 µM

We observe that the biological activity of the sandarazols mainly stems from the chlorinated sandarazols such as **3**, which shows both the best antimicrobial and anti-proliferative activity. Its cytotoxicity of 0.5 µM against HCT116 cells might make **3** a promising candidate for compound optimization and mode of action studies. Such approaches directed towards improvements of the sandarazols' stability against oxygen would probably increase the observable IC$_{50}$ values, as the values shown here cannot be determined under anaerobic conditions that stabilize the sandarazols. The antimicrobial and anti-proliferative activities of **3** indicate that those secondary metabolites enhance the bacterium's capability to defend itself against competitors like other bacteria or eukaryotic microorganisms. The position of the *szo* BGC on the replicative plasmid pSa001 indicates that this bacterial self-defense strategy can be transferred between bacteria via conjugation.

6.4 Conclusion and outlook

In this work, we identified pSa001 as the second myxobacterial autonomously replicating plasmid known to date. It was found in the newly isolated *Sandaracinus* sp. MSr10575, the second myxobacterial isolate belonging to the *Sandaracinus* clade. The 209.7 kbp pSa001 plasmid itself encodes the *szo* biosynthetic gene cluster as well as five transposase- and one integrase-type ORFs. Therefore, the plasmid's content including the *szo* BGC may be transferred between bacterial species, not only by conjugative transfer of the plasmid as an autonomously replicating unit, but also by transposition or integration into the acceptor strain's genome. We chose to investigate and characterize the pSa001 borne *szo* BGC that turned out to be cryptic under laboratory conditions. Previous examples of such plasmid borne secondary metabolite pathways have shown to be responsible for esmeraldin and mycolactone production, both of which show significant biological activity. [5,6] The biosynthetic origin of these two compounds led to the theory that the BGC encoded on pSa001 also encodes the biosynthetic machinery for toxin production. After activation of the BGC located on the plasmid pSa001 by promoter exchange, we observed the production of the sandarazols, a series of type1-*trans*-AT PKS-NRPS hybrid secondary metabolites. The chemical structure of the sandarazols features a permanent positive charge at physiological conditions at the tertiary amine, a Michael acceptor system, an epoxyketone as well as an alpha chlorinated ketone. These compounds are therefore prone to display a variety of different reactivities that might well translate into diverse biological activities. More precisely, the sandarazols show promising anti-bacterial and anti-proliferative activities. Following in-depth investigation of the *szo* BGC, we developed a concise biosynthetic model, explaining the sandarazol biosynthesis on its megasynthase protein. Furthermore, we shed light on a variety of uncommon biosynthetic steps such as the first description of incorporation of Me$_2$Dap into natural products, polyketide β-branching, isomerization of two consecutive double bonds by a single shifting DH-domain and α-chlorination governed release of the sandarazols from the assembly line, all of which warrant further investigations of biosynthetic details. The discovery of the sandarazols does not only highlight the capability of myxobacteria to biosynthesize diverse and biologically active secondary metabolites, it also emphasizes the promise of activating biosynthetic gene clusters by inducible heterologous promoters. Vanillate-inducible promoter systems have proven once more to be able to unlock a larger fraction of the cryptic secondary metabolome of myxobacteria in promoter exchange experiments. [9] The sandarazols combine biological activity with both their intriguing chemical structure and biosynthesis. Thus, natural products encoded on other bacterial plasmids might feature similar intriguing properties. The discovery of sandarazols therefore highlights plasmid encoded BGCs as valuable targets in the search for novel biologically active natural products in the future. As for myxobacteria, encoding toxin BGCs on plasmids that are transferrable between themselves may provide myxobacteria a competitive edge in their ecological niche. This transfer of toxin production machinery would make particular sense, as myxobacteria do not live individually but rather as part of a swarm, which shows wolf pack like predatory behavior. [53] Thus, distribution of the genetic means for toxin production among its kind should be advantageous for these microorganisms in their collective predation and survival strategies.

6.5 References

(1) Naylor, N. R.; Atun, R.; Zhu, N.; Kulasabanathan, K.; Silva, S.; Chatterjee, A.; Knight, G. M.; Robotham, J. V. Estimating the burden of antimicrobial resistance: a systematic literature review. *Antimicrob. Resist. Infect. Control* **2018**, *7*, 58.

(2) Brenciani, A.; Morroni, G.; Pollini, S.; Tiberi, E.; Mingoia, M.; Varaldo, P. E.; Rossolini, G. M.; Giovanetti, E. Characterization of novel conjugative multiresistance plasmids carrying cfr from linezolid-resistant Staphylococcus epidermidis clinical isolates from Italy. *J. Antimicrob. Chemother.* **2016**, *71*, 307–313.

(3) Soucy, S. M.; Huang, J.; Gogarten, J. P. Horizontal gene transfer: Building the web of life. *Nat. Rev. Genet.* **2015**, *16*, 472.

(4) Luna, V. A.; King, D. S.; Peak, K. K.; Reeves, F.; Heberlein-Larson, L.; Veguilla, W.; Heller, L.; Duncan, K. E.; Cannons, A. C.; Amuso, P.*; et al.* Bacillus anthracis Virulent Plasmid pX02 Genes Found in Large Plasmids of Two Other Bacillus Species. *J. Clin. Microbiol.* **2006**, *44*, 2367–2377.

(5) Stinear, T. P.; Mve-Obiang, A.; Small, P. L.; Frigui, W.; Pryor, M. J.; Brosch, R.; Jenkin, G. A.; Johnson, P. D.; Davies, J. K.; Lee, R. E.*; et al.* Giant plasmid-encoded polyketide synthases produce the macrolide toxin of Mycobacterium ulcerans. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 1345–1349.

(6) Rui, Z.; Ye, M.; Wang, S.; Fujikawa, K.; Akerele, B.; Aung, M.; Floss, H. G.; Zhang, W.; Yu, T.-W. Insights into a divergent phenazine biosynthetic pathway governed by a plasmid-born esmeraldin gene cluster. *Chemistry & biology* **2012**, *19*, 1116–1125.

(7) Cragg, G. M.; Newman, D. J. Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta* **2013**, *1830*, 3670–3695.

(8) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, *79*, 629–661.

(9) Wenzel, S. C.; Müller, R. Myxobacteria—''microbial factories' for the production of bioactive secondary metabolites. *Mol. Biosyst.* **2009**, *5*, 567–574.

(10) Bader, C. D.; Panter, F.; Müller, R. In depth natural product discovery - Myxobacterial strains that provided multiple secondary metabolites. *Biotechnol. Adv.* **2020**, *39*, 107480.

(11) Herrmann, J.; Fayad, A. A.; Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **2017**, *34*, 135–160.

(12) Hoffmann, T.; Krug, D.; Bozkurt, N.; Duddela, S.; Jansen, R.; Garcia, R.; Gerth, K.; Steinmetz, H.; Müller, R. Correlating chemical diversity with taxonomic distance for discovery of natural products in myxobacteria. *Nat. Commun.* **2018**, *9*, 803.

(13) Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X.; Schwalen, C. J.; Kautsar, S. A.; Suarez Duran, H. G.; Los Santos, E. L. C. de; Kim, H. U.; Nave, M.*; et al.* antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* **2017**, *45*, W36-W41.

(14) Cimermancic, P.; Medema, M. H.; Claesen, J.; Kurita, K.; Wieland Brown, L. C.; Mavrommatis, K.; Pati, A.; Godfrey, P. A.; Koehrsen, M.; Clardy, J.*; et al.* Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. *Cell* **2014**, *158*, 412–421.

(15) Ross, A. C.; Gulland, L. E. S.; Dorrestein, P. C.; Moore, B. S. Targeted capture and heterologous expression of the Pseudoalteromonas alterochromide gene cluster in Escherichia coli represents a promising natural product exploratory platform. *ACS Synth. Biol.* **2015**, *4*, 414–420.

(16) Mao, D.; Okada, B. K.; Wu, Y.; Xu, F.; Seyedsayamdost, M. R. Recent advances in activating silent biosynthetic gene clusters in bacteria. *Curr. Opin. Microbiol.* **2018**, *45*, 156–163.

(17) Hug, J. J.; Bader, C. D.; Remškar, M.; Cirnski, K.; Müller, R. Concepts and Methods to Access Novel Antibiotics from Actinomycetes. *Antibiotics* **2018**, *7*, 44.

(18) Mohr, K. I.; Garcia, R. O.; Gerth, K.; Irschik, H.; Müller, R. *Sandaracinus amylolyticus* gen. nov., sp. nov., a starch-degrading soil myxobacterium, and description of *Sandaracinaceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* **2012**, *62*, 1191–1198.

(19) Steinmetz, H.; Mohr, K. I.; Zander, W.; Jansen, R.; Gerth, K.; Müller, R. Indiacens A and B: prenyl indoles from the *myxobacterium Sandaracinus amylolyticus*. *J. Nat. Prod.* **2012**, *75*, 1803–1805.

(20) Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.*; et al.* Real-time DNA sequencing from single polymerase molecules. *Science* **2009**, *323*, 133–138.

(21) Feng, J.; Chen, X. J.; Sun, X.; Wang, N.; Li, Y. Z. Characterization of the replication origin of the myxobacterial self-replicative plasmid pMF1. *Plasmid* **2012**, *68*, 105–112.

(22) 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.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837.

(23) Panter, F.; Krug, D.; Baumann, S.; Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **2018**, *9*, 4898–4908.

(24) Yamanaka, K.; Reynolds, K. A.; Kersten, R. D.; Ryan, K. S.; Gonzalez, D. J.; Nizet, V.; Dorrestein, P. C.; Moore, B. S. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 1957–1962.

(25) Cortina, N. S.; Krug, D.; Plaza, A.; Revermann, O.; Müller, R. Myxoprincomide: a natural product from *Myxococcus xanthus* discovered by comprehensive analysis of the secondary metabolome. *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 811–816.

(26) Shizuya, H.; Birren, B.; Kim, U. J.; Mancino, V.; Slepak, T.; Tachiiri, Y.; Simon, M. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 8794–8797.

(27) Hug, J. J.; Panter, F.; Krug, D.; Müller, R. Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria. *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 319–334.

(28) Iniesta, A. A.; García-Heras, F.; Abellón-Ruiz, J.; Gallego-García, A.; Elías-Arnanz, M. Two systems for conditional gene expression in *Myxococcus xanthus* inducible by isopropyl-ß-D-thiogalactopyranoside or vanillate. *J. Bacteriol.* **2012**, *194*, 5875–5885.

(29) Hoffmann, T.; Krug, D.; Hüttel, S.; Müller, R. Improving natural products identification through targeted LC-MS/MS in an untargeted secondary metabolomics workflow. *Anal. Chem.* **2014**, *86*, 10780–10788.

(30) 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.

(31) Meusel, M.; Hufsky, F.; Panter, F.; Krug, D.; Müller, R.; Böcker, S. Predicting the presence of uncommon elements in unknown biomolecules from isotope patterns. *Anal. Chem.* **2016**, *88*, 7556–7566.

(32) H. Duddeck. 4. Determination of Absolute and Relative Configuration: 4.1. Determination of Relative Configuration by Nuclear Magnetic Resonance Methods (I). *Methods of organic chemistry (Houben-Weyl)*.

(33) Harada, K.-i.; Fujii, K.; Hayashi, K.; Suzuki, M.; Ikai, Y.; Oka, H. Application of d,l-FDLA derivatization to determination of absolute configuration of constituent amino acids in peptide by advanced Marfey's method. *Tetrahedron Lett.* **1996**, *37*, 3001–3004.

(34) Marfey, P. Determination of D-amino acids. II. Use of a bifunctional reagent, 1,5-difluoro-2,4-dinitrobenzene. *Carlsberg Res. Commun.* **1984**, *49*, 591–596.

(35) Hoye, T. R.; Jeffrey, C. S.; Shao, F. Mosher ester analysis for the determination of absolute configuration of stereogenic (chiral) carbinol carbons. *Nat. Protoc.* **2007**, *2*, 2451–2458.

(36) Maloney, F. P.; Gerwick, L.; Gerwick, W. H.; Sherman, D. H.; Smith, J. L. Anatomy of the β-branching enzyme of polyketide biosynthesis and its interaction with an acyl-ACP substrate. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 10316–10321.

(37) Kjaerulff, L.; Raju, R.; Panter, F.; Scheid, U.; Garcia, R.; Herrmann, J.; Müller, R. Pyxipyrrolones: Structure elucidation and biosynthesis of cytotoxic myxobacterial metabolites. *Angew. Chem. Int. Ed.* **2017**, *56*, 9614–9618.

(38) Gehret, J. J.; Gu, L. C.; Gerwick, W. H.; Wipf, P.; Sherman, D. H.; Smith, J. L. Terminal alkene formation by the thioesterase of curacin a biosynthesis structure of a decarboxylating thioesterase. *J. Biol. Chem.* **2011**, *286*, 14445–14454.

(39) Moebius, N.; Ross, C.; Scherlach, K.; Rohm, B.; Roth, M.; Hertweck, C. Biosynthesis of the respiratory toxin bongkrekic acid in the pathogenic bacterium Burkholderia gladioli. *Chem. Biol.* **2012**, *19*, 1164–1174.

(40) Akey, D. L.; Razelun, J. R.; Tehranisa, J.; Sherman, D. H.; Gerwick, W. H.; Smith, J. L. Crystal structures of dehydratase domains from the curacin polyketide biosynthetic pathway. *Structure* **2010**, *18*, 94–105.

(41) Pogorevc, D.; Panter, F.; Schillinger, C.; Jansen, R.; Wenzel, S. C.; Müller, R. Production optimization and biosynthesis revision of corallopyronin A, a potent anti-filarial antibiotic. *Metab. Eng.* **2019**, *55*, 201–211.

(42) Gay, D. C.; Spear, P. J.; Keatinge-Clay, A. T. A double-hotdog with a new trick: structure and mechanism of the trans-acyltransferase polyketide synthase enoyl-isomerase. *ACS chemical biology* **2014**, *9*, 2374–2381.

(43) Kusebauch, B.; Busch, B.; Scherlach, K.; Roth, M.; Hertweck, C. Functionally Distinct Modules Operate Two Consecutive α,ß ->ß,γ Double-Bond Shifts in the Rhizoxin Polyketide Assembly Line. *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 1460–1464.

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

(45) Beasley, F. C.; Cheung, J.; Heinrichs, D. E. Mutation of L-2,3-diaminopropionic acid synthase genes blocks staphyloferrin B synthesis in *Staphylococcus aureus*. *BMC Microbiol.* **2011**, *11*, 199.

(46) Kobylarz, M. J.; Grigg, J. C.; Takayama, S.-i. J.; Rai, D. K.; Heinrichs, D. E.; Murphy, M. E.P. Synthesis of L-2,3-Diaminopropionic Acid, a Siderophore and Antibiotic Precursor. *Chemistry & biology* **2014**, *21*, 379–388.

(47) Hornbogen, T.; Riechers, S.-P.; Prinz, B.; Schultchen, J.; Lang, C.; Schmidt, S.; Mügge, C.; Turkanovic, S.; Süssmuth, R. D.; Tauberger, E*.; et al.* Functional characterization of the recombinant N-methyltransferase domain from the multienzyme enniatin synthetase. *ChemBioChem* **2007**, *8*, 1048–1054.

(48) Walsh, C. T. Insights into the chemical logic and enzymatic machinery of NRPS assembly lines. *Nat. Prod. Rep.* **2016**, *33*, 127–135.

(49) Yeh, E.; Blasiak, L. C.; Koglin, A.; Drennan, C. L.; Walsh, C. T. Chlorination by a long-lived intermediate in the mechanism of flavin-dependent halogenases. *Biochemistry* **2007**, *46*, 1284–1292.

(50) Guan, X.; An, D.; Liu, G.; Zhang, H.; Gao, J.; Zhou, T.; Zhang, G.; Zhang, S. Enantioselective α-chlorination of β-keto esters and amides catalyzed by chiral imidodiphosphoric acids. *Tetrahedron Lett.* **2018**, *59*, 2418–2421.

(51) Kantz, A.; Gassner, G. T. Nature of the reaction intermediates in the flavin adenine dinucleotide-dependent epoxidation mechanism of styrene monooxygenase. *Biochemistry* **2011**, *50*, 523–532.

6. S a n d a r a z o l s

P a g e | 241

(52) Morrison, E.; Kantz, A.; Gassner, G. T.; Sazinsky, M. H. Structure and mechanism of styrene monooxygenase reductase: new insight into the FAD-transfer reaction. *Biochemistry* **2013**, *52*, 6063–6075.

(53) Muñoz-Dorado, J.; Marcos-Torres, F. J.; García-Bravo, E.; Moraleda-Muñoz, A.; Pérez, J. Myxobacteria: moving, killing, feeding, and surviving together. *Front. Microbiol.* **2016**, *7*, 781.

6. S a n d a r a z o l s

P a g e | 241

6.6 Supporting information

The Sandarazols Are Cryptic and Structurally Unique Plasmid Encoded Toxins from a Rare Myxobacterium

Fabian Panter[1], Chantal D. Bader[1] and Rolf Müller[1]

Affiliations

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

Data with limited visibility on print media such as raw NMR data can be found on the enclosed storage medium.

S6.1 Myxobacterial growth conditions

S6.1.1 Myxobacterial culture media

Table S1 Recipe for 2-SWT medium

2SWT – Medium			
Amount	Ingredient	Concentration	Supplier
3 g/L	Tryptone	-	BD
1 g/L	Soytone	-	BD
3.5 g/L	Soluble Starch	-	Roth
4 g/L	Maltose Monohydrate	-	
2 g/L	Glucose	-	Roth
10 g/L	Starch (soluble)	-	Roth
0.5 g/L	$CaCl_2$	-	Sigma Aldrich
1 g/L	$MgSO_4$ x $7H_2O$	-	Grüssing
10 mL/L	TRIS x HCl ph8	1M	Sigma Aldrich
100 µL/L	Sterile Vit. B12 solution (added after autoclaving)	1 mg/mL	Roth
200 µL/L	Sterile FeEDTA solution (added after autoclaving)	8 mg/mL	Sigma Aldrich
Dissolved in milli-Q. Water, pH adjusted to 7.2 with 1N KOH			

Table S2 Recipe for S15 medium

S15 – Medium			
Amount	Ingredient	Concentration	Supplier
3 g/L	Tryptone	-	BD
1 g/L	Soytone	-	BD
3.5 g/L	Soluble Starch	-	Roth
4 g/L	Maltose Monohydrate	-	
2 g/L	Glucose	-	Roth
10 g/L	Starch (soluble)	-	Roth
0.5 g/L	$CaCl_2$	-	Sigma Aldrich
1 g/L	$MgSO_4$ x $7H_2O$	-	Grüssing
10 mL/L	TRIS x HCl ph8	1M	Sigma Aldrich
100 µL/L	Sterile Vit. B12 solution (added after autoclaving)	1 mg/mL	Roth
200 µL/L	Sterile FeEDTA solution (added after autoclaving)	8 mg/mL	Sigma Aldrich
Dissolved in milli-Q. Water, pH adjusted to 7.2 with 1N KOH			

Table S3 Recipe for 2SWYT medium

2SWYT – Medium			
Amount	Ingredient	Concentration	Supplier
3 g/L	Tryptone	-	BD
1 g/L	Soytone	-	BD
3.5 g/L	Soluble Starch	-	Roth
4 g/L	Maltose Monohydrate	-	
10 g/L	Baker's yeast (alive)	-	
2 g/L	Glucose	-	Roth
10 g/L	Starch (soluble)	-	Roth
0.5 g/L	$CaCl_2$	-	Sigma Aldrich
1 g/L	$MgSO_4$ x $7H_2O$	-	Grüssing
10 mL/L	TRIS x HCl ph8	1M	Sigma Aldrich
100 µL/L	Sterile Vit. B12 solution (added after autoclaving)	1 mg/mL	Roth
200 µL/L	Sterile FeEDTA solution (added after autoclaving)	8 mg/mL	Sigma Aldrich
Dissolved in milli-Q Water, pH adjusted to 7.2 with 1N KOH			

The myxobacterial strain MSr10575 was kept in agar culture both for storage over short amounts of time and for cloning. The agar media used are S15 agar and S15 soft agar, which is prepared by adding 14 g/L agarose and 8 g/L agarose (BD) to S15 medium preparations before autoclaving.

S6.1.2 Myxobacterial fermentation conditions for LC-MS analysis

Cultures for UHPLC-*hr*MS analysis are grown in 300 mL shake flasks containing 50 mL of 2SWT medium for *Sandaracinus sp.* MSr10575 inoculated with 1 mL of pre culture. Media for mutant MSr10575 strains were supplemented with 50 mg/L kanamycin (Roth) and 1 mM of aqueous sterile filtrated K-vanillate solution if the corresponding strain's vanillate promotor is to be induced. After inoculation the medium is supplemented with 2% of sterile XAD-16 adsorber resin (Sigma Aldrich) suspension in water to bind secondary metabolites in the culture medium and limit sandarazol auto toxicity. Small scale cultures were grown for 10-12 days. After fermentation the culture is pelleted in a 50 mL falcon at 6000 rcf for 10 minutes using an Eppendorf falcon table centrifuge and stored at -20 °C until further use.

S6.2 Analytical methods used in this work

S6.2.1 Metabolite extraction procedure for analytical scale extractions

The frozen cell pellet is transferred into a 100 mL Erlenmeyer flask and a magnetic stirrer is added. 50 mL of acetone (fluka analytical grade, redistilled in house) are added onto the pellet and the mixture is stirred for 60 min on a magnetic stirrer. The acetone extract is left to settle in order to sediment cell debris and XAD resin for a second extraction step. The supernatant is filtered with a 125 micron folded filter keeping cell pellet and XAD-16 resin in the Erlenmeyer flask for a second extraction step. The residual pellet and XAD-16 resin is extracted again with 30 mL of distilled acetone for 60 min on a magnetic stirrer and filtered through the same folded filter. The combined extracts are transferred into a 100 mL round bottom flask. The acetone is evaporated using a rotary evaporator at 260 mbar and 40 °C water bath temperature. The residual water is evaporated at 20 mbar until the residue in the flask is completely dry. The residue is taken up in 550 μL of methanol (Chromasolv HPLC grade, Sigma Aldrich) and transferred into an 1.5 mL Eppendorf tube. This tube is centrifuged with a Hitachi table centrifuge at 15000 rpm for 2 minutes to remove residual insolubilities such as salts, cell debris and XAD fragments. The residual extract is diluted 1:10 for UHPLC-*hr*MS analysis.

S6.2.2 Standardized UHPLC MS conditions

UPLC-*hr*MS analysis is performed on a Dionex (Germering, Germany) Ultimate 3000 RSLC system using a Waters (Eschborn, Germany) BEH C18 column (50 x 2.1 mm, 1.7 μm) equipped with a Waters VanGuard BEH C18 1.7 μm guard column. Separation of 1 μL sample is 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 a column temperature of 45 °C. Gradient conditions are as follows: 0 – 0.5 min, 5% B; 0.5 – 18.5 min, 5 – 95% B; 18.5 – 20.5 min, 95% B; 20.5 – 21 min, 95 – 5% B; 21-22.5 min, 5% B. UV spectra are recorded by a DAD in the range from 200 to 600 nm. The LC flow is split to 75 μL/min before entering the Bruker Daltonics maXis 4G *hr*ToF mass spectrometer (Bremen, Germany) equipped with an Apollo II ESI source. Mass spectra are acquired in centroid mode ranging from 150 – 2500 m/z at a 2 Hz full scan rate. Mass spectrometry source parameters are set to 500 V as end plate offset; 4000 V as capillary voltage; nebulizer gas pressure 1 bar; dry gas flow of 5 l/min and a dry temperature of 200

°C. Ion transfer and quadrupole settings are set to funnel RF 350 Vpp.; multipole RF 400 Vpp as transfer settings and ion energy of 5 eV as well as a low mass cut of 300 m/z. Collision cell is set to 5.0 eV and pre-pulse storage time is set to 5 μs. Spectra acquisition rate is set to 2 Hz. Calibration is done automatically before every LC-MS run by injection of sodium formate and calibration on the respective clusters formed in the ESI source. All MS analyses are acquired in the presence of the lock masses ($C_{12}H_{19}F_{12}N_3O_6P_3$, $C_{18}H_{19}O_6N_3P_3F_2$ and $C_{24}H_{19}F_{36}N_3O_6P_3$) which generate the [M+H]$^+$ ions of 622.0289; 922.0098 and 1221.9906.

S6.2.3 Methodology for statistics based metabolome filtering

In order to detect all metabolites appearing after genetic manipulation of MSr10575 we compare wild type MSr10575 to the sandarazol cluster activation mutant in an unbiased principal component analysis (PCA) based statistical analysis adapted from Panter et al..[1] For this purpose, LC-MS chromatograms of 3 independent cultivations are measured as 2 technical replicates each, giving a total number of 6 LC-*hr*MS chromatograms per strain. To obtain all molecular features in the 6 LC-*hr*MS chromatograms of the bacterial extracts of the induced mutant strains and the 6 LC-*hr*MS chromatograms of the corresponding wild type strain extracts, the T-ReX-3D molecular feature finder implemented in Bruker Metaboscape 4.01 is used. Compound detection parameters intensity threshold is set to 10000, mz threshold to 0.005 Da and minimum compound length to 4 spectra. PCA t-test tables are created with the built in PCA t-test routine and filtered according to 6 appearances in the sandarazol cluster activation mutant extract chromatograms and 0 appearances in the MSr10575 wild type extract chromatograms. The t-test table from the bucketing as well as the processed scheduled precursor list used to acquire SPL-MS/MS data will be supplied upon request.

S6.2.4 Acquisition parameters for acquiring high-resolution tandem MS data

LC and MS conditions for SPL guided MS/MS data acquisitions are kept constant according to section standardized UHPLC-MS conditions. MS/MS data acquisition parameters are set to exclusively fragment scheduled precursor list entries. SPL tolerance parameters for precursor ion selection are set to 0.2 minutes and 0.05 m/z in the SPL MS/MS method. The method picks up to 2 precursors per cycle, applies smart exclusion after 5 spectra and performs CID and MS/MS spectra acquisition time ramping. 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. MS full scan acquisition rate is set to 2 Hz and MS/MS spectra acquisition rates are ramped from 1 to 4 Hz for precursor ion intensities of 10 kcts. to 1000 kcts..

S6.2.5 Spectral networking parameters for GNPS clustering

All supporting GNPS clustering data presented here is created based on exported .mzML files from the UHPLC-*hr*MS² chromatograms using the parameters specified in the experimental section of the main text. The MS/MS chromatograms are exported containing all MS/MS data as an .mzML data file and uploaded to the GNPS server at University of California San Diego via FileZilla FTP upload to ftp:// ccms-ftp01.ucsd.edu and all acquired SPL MS/MS spectra are used for spectral network creation. [2] A molecular network is created using the online workflow at GNPS. The data is filtered by removing all MS/MS peaks within +/- 17 Da of the precursor m/z. MS/MS spectra are window filtered by choosing only the top 6 peaks in the +/- 50 Da window throughout the spectrum. The data is then clustered with a parent mass tolerance of 0.05 Da and a MS/MS fragment ion tolerance of 0.1 Da to create consensus spectra. Further, consensus spectra that contained less than 1 spectra are discarded. A network is then

created where edges are filtered to have a cosine score above 0.65 and more than 4 matched peaks. Further edges between two nodes are kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. [2] The dataset is downloaded from the server and subsequently visualized using Cytoscape 3.7.2.

S6.2.6 Identification of putative sandarazols from bulk MS/MS data

As we had the *szo* biosynthetic gene cluster at hand before starting tandem MS experiments on both the MSr10575 wild type as well as the MSr10575 : pBeloBacSa001 plasmid we could assume building blocks that should be incorporated into the sandarazols final structure from the predicted incorporation specificities of the BGC modules. Browsing the created metabolome dependent tandem MS data we could not find a group of secondary metabolites prominently featuring a loss of a valine or leucine fragment which led us to the assumption that the corresponding BGC is putatively 'cryptic'. After activation of the *szo* BGC we observed the molecular masses discussed below, all of which show a similar fragmentation pattern, which leads to grouping in the GNPS software indicating them to be a family of secondary metabolites. Moreover, all of these secondary metabolites feature a prominent and well visible loss of a valine fragment indicating them, to be of peptidic origin featuring a valine. As the *szo* BGC contains a valine specific module we assumed this group of secondary metabolites to be dependent on the *szo* BGC.

S6.2.7 Identification of minor sandarazols and sandarazol fragments by GNPS analysis

In this GNPS based spectral networking approach the largest set of clustered MS/MS spectra represents the sandarazol compound family. As one would expect, this GNPS cluster contains the sandarazols described in this work as well as a number of additional minor derivatives of the sandarazols or sandarazol fragments that we could not isolate or structurally elucidate due to their low production titers. Neutral formulas given here are calculated using the built-in tool in Bruker Compass Data analysis.

Table S4 List of all MS signals tied to the sandarazol MS/MS cluster created using the GNPS software

measured parent mass [M+H]$^+$	putative neutral formula	compound name
575.341 Da	$C_{30}H_{46}N_4O_7$	Sandarazol A
577.365 Da	$C_{30}H_{48}N_4O_7$	Sandarazol B
583.325 Da	$C_{29}H_{47}N_4O_6Cl$	Sandarazol C
585.343 Da	$C_{29}H_{49}N_4O_6Cl$	Sandarazol D
559.349 Da	$C_{30}H_{46}N_4O_6$	Sandarazol E
561.365 Da	$C_{30}H_{48}N_4O_6$	Sandarazol F
563.380 Da	$C_{30}H_{50}N_4O_6$	Sandarazol G
589.359 Da	$C_{29}H_{53}N_4O_6Cl$	N.A.
551.357 Da	$C_{28}H_{46}N_4O_7$	N.A.
549.365 Da	$C_{28}H_{44}N_4O_7$	N.A.
579.376 Da	$C_{30}H_{50}N_4O_7$	N.A.
593.354 Da	$C_{30}H_{48}N_4O_8$	Sandarazol A hydrolyzed (diol)
571.362 Da	$C_{29}H_{51}N_4O_5Cl$	N.A.
569.346 Da	$C_{29}H_{49}N_4O_5Cl$	N.A.
545.370 Da	$C_{30}H_{48}N_4O_5$	N.A.
543.354 Da	$C_{30}H_{46}N_4O_5$	N.A.
603.375 Da	$C_{32}H_{50}N_4O_7$	N.A.
770.448 Da	$C_{44}H_{59}N_5O_7$	N.A.
768.433 Da	$C_{44}H_{57}N_5O_7$	N.A.
772.466 Da	$C_{44}H_{61}N_5O_7$	N.A.
802.395 Da	$C_{44}H_{56}N_5O_7Cl$	N.A.
804.412 Da	$C_{44}H_{58}N_5O_7Cl$	N.A.
806.426 Da	$C_{44}H_{60}N_5O_7Cl$	N.A.

S6.2.8 MS² spectra analysis to assign the sandarazol structures not elucidated by NMR

A couple of sandarazol derivatives that were not structurally elucidated by NMR were assigned by analyzing their MS/MS spectra. The MS/MS spectra were acquired according to the parameters described in the MS/MS parameter description.
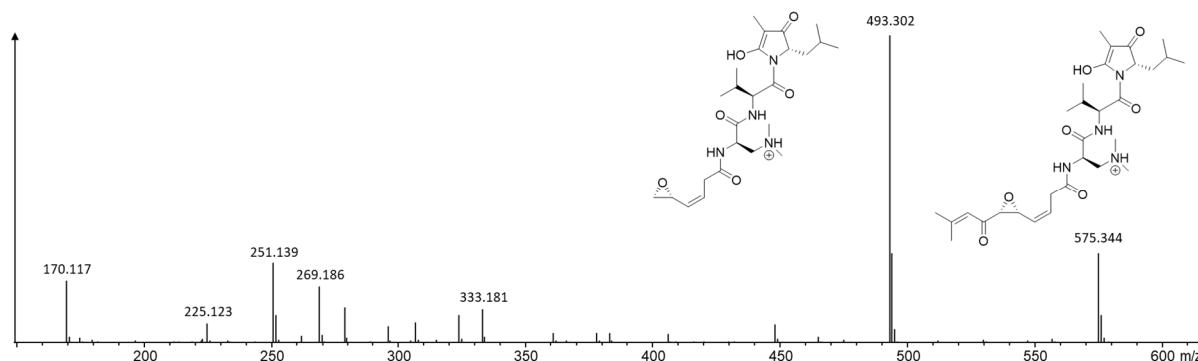
Figure S1 Annotated MS² spectrum of sandarazol A
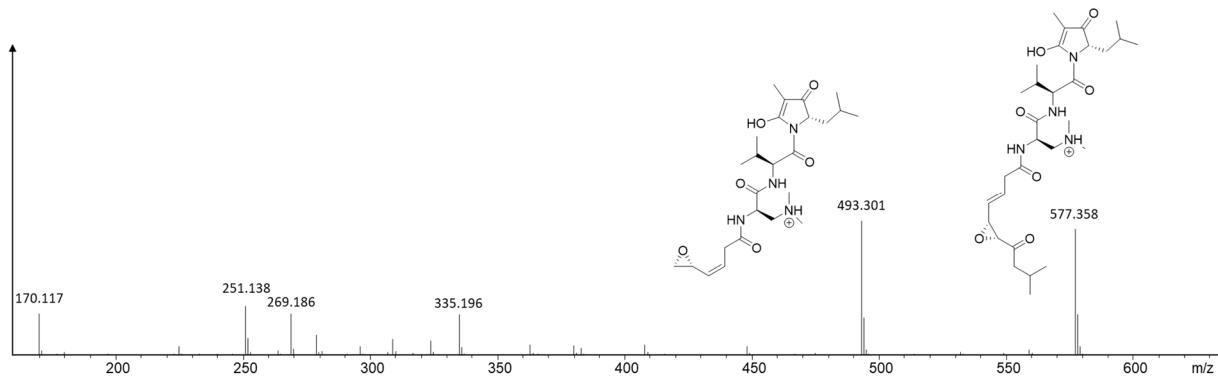
Figure S2 Annotated MS² spectrum of sandarazol B
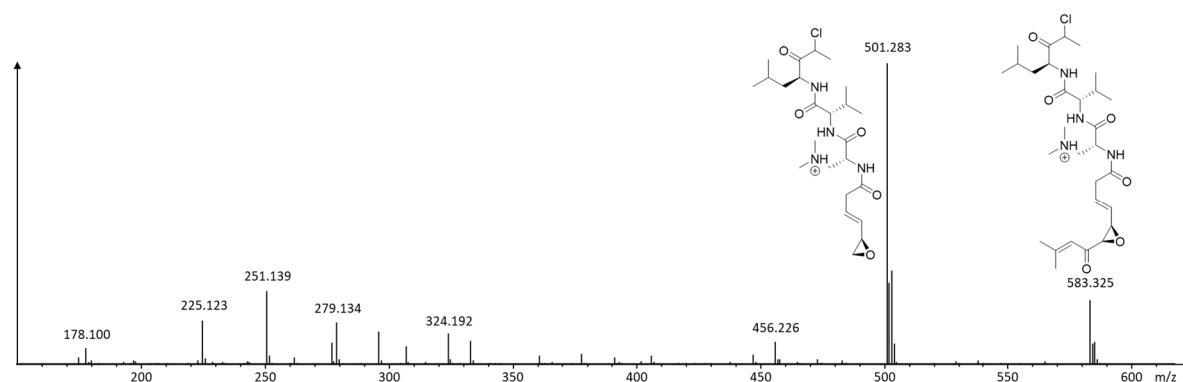
Figure S3 Annotated MS² spectrum of sandarazol C
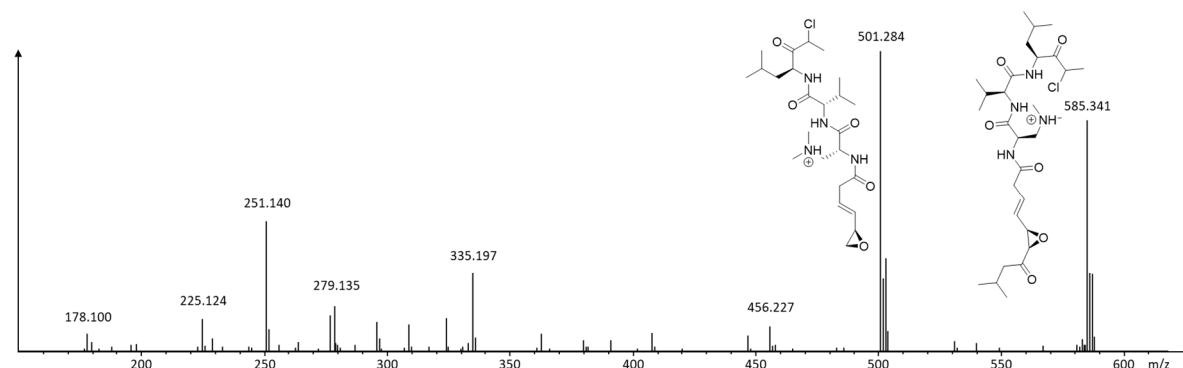
Figure S4 Annotated MS² spectrum of sandarazol D
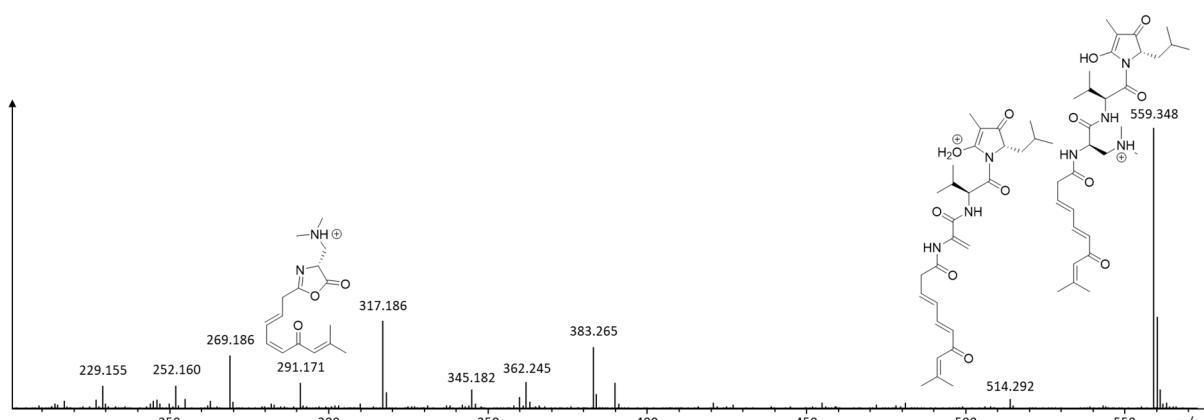
Figure S5 Annotated MS² spectrum of sandarazol E
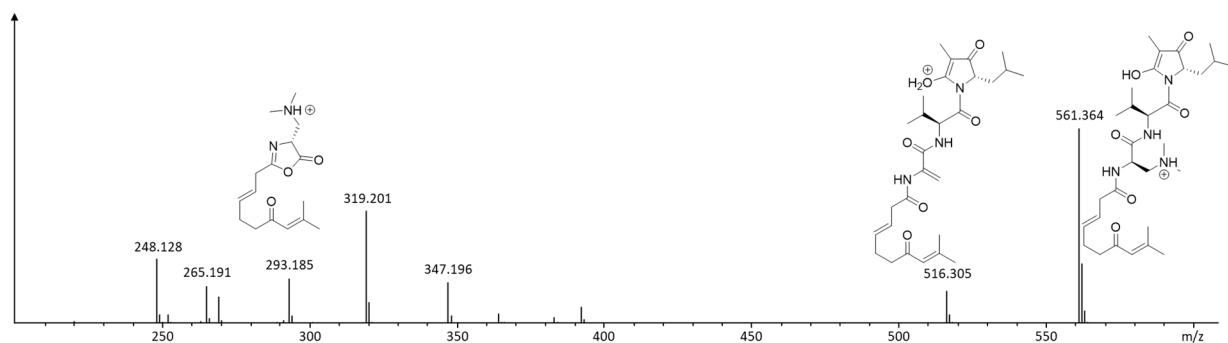
Figure S6 Annotated MS² spectrum of sandarazol F
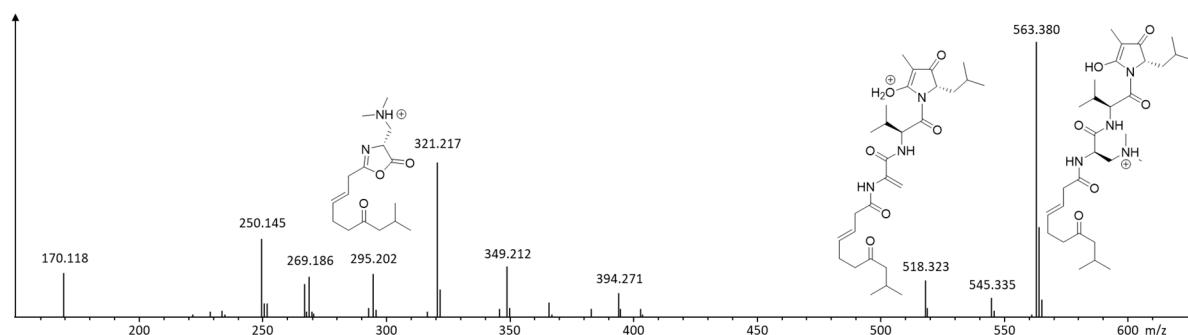
Figure S7 Annotated MS² spectrum of sandarazol G

Sandarazol A, B, C and F were structurally elucidated using multidimensional NMR analyses. The structural formulae of sandarazol D, E and G can be inferred on the basis of the above described NMR spectra. First of all, we can clearly see in the MS² spectra of sandarazol A, B and C, if the structure contains an epoxyketone, as the bond between the epoxide and the ketone is preferentially fragmented. This means that the difference between sandarazol A and B in the structural formulae is the same difference as the difference between sandarazol C and D as both sandarazols show an identical MS² fragment with a terminal epoxide. In sandarazols E, F and G that are all lacking one oxygen atom, we can clearly see that this terminal epoxide fragment does no longer exist. It is therefore safe to assume that the missing oxygen atom is indeed the epoxide oxygen atom in the sandarazols' epoxyketone moiety. In the three MS2 spectra we can see a fragment that is either 291.171, 293.185 or 295.202 Da representing the PKS part of the molecule that still contains the one double bond present in all sandarazols. As we known the structural formula of sandarazol F that has only one additional double bond (for which there would be 2 regioisomers), we can assume that sandarazol E has two additional double bonds and is therefore the direct precursor of sandarazol A, while sandarazol G has no additional double bonds and has therefore only one possible structural formula. As all sandarazols are produced by the same assembly line that is stereospecific, we assume the stereo centers in all sandarazol derivatives to have the same configuration. It is worth mentioning that the stereocenter at the position of the chlorine in sandarazol C and D is rapidly racemizing and we are therefore unable to determine its configuration.

S6.2.9 Stable isotope labelling experiments

In order to take a deeper look at the sandarazol biosynthesis we decided to feed L-Serine to an induced analytical culture of MSr10575 :: pBeloBacSa001 to see whether we observe incorporation into the

sandarazols. This would equally mean that the uncommon amino acid (2R)-2-Amino-3-(N,N-dimethyl amino)-propanoic acid (D-DimeDap) provides from serine.

575.343

Sandarazol A control

575.343

Sandarazol A L-Ser-d_3 feeding
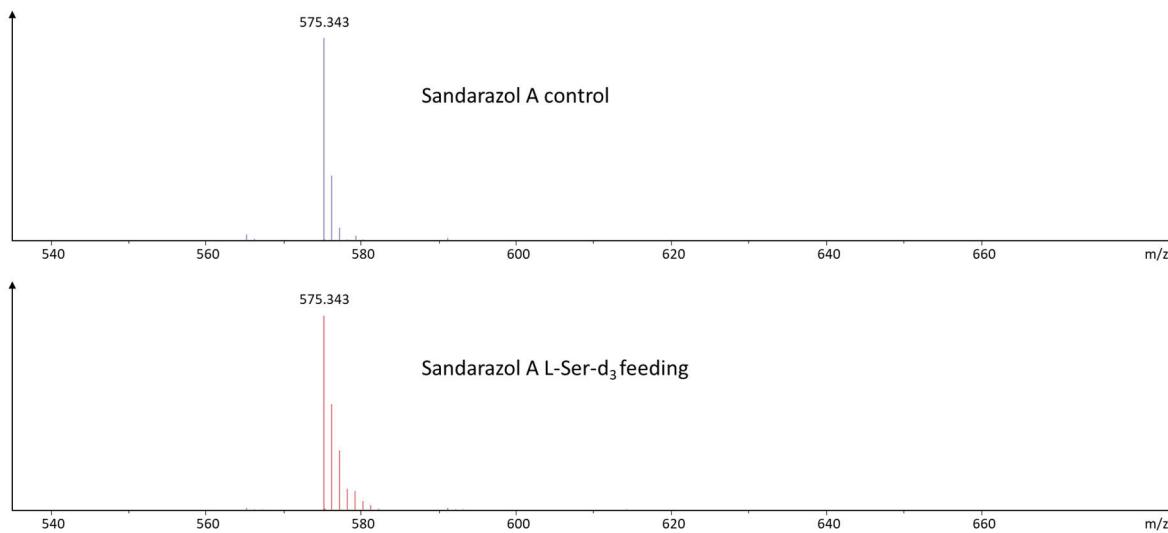
Figure S8 Influence of feeding of L-serine-d_3 on the isotope pattern of sandarazol A

583.324

Sandarazol C control

583.324

Sandarazol C L-Ser-d_3 feeding
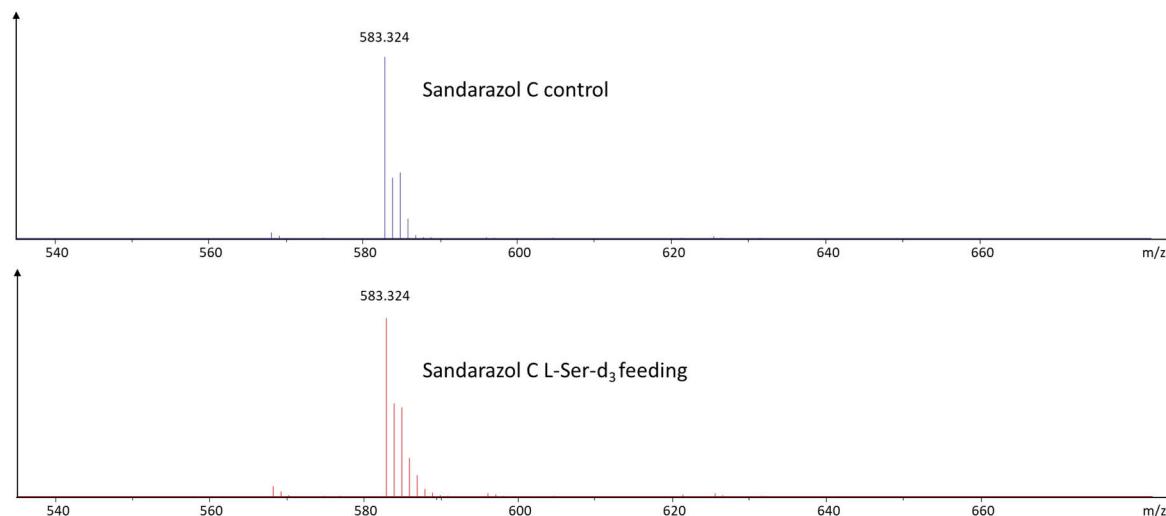
Figure S9 Influence of feeding of L-serine-d_3 on the isotope pattern of sandarazol C

We see a clear shift of sandarazols towards higher masses both in sandarazol A and sandarazol C. This means that there is additional evidence that L-Me$_2$Dap is indeed synthesized from L-serine as proposed in the sandarazol biosynthesis. The sandarazol labelling study with L-serine-d_3 is not as crisp as the labelling study with L-leucine-d_3, as serine is metabolized rather quickly. So, in addition to the washing out effect that removed deuterium substituents from α-protons leading to production of L-serine-d_2 from L-serine-d_3 (and which can also be observed for Val-d_8 below) partial degradation of the molecule can produce the corresponding d_1 species. We therefore end up with a mixture of d_0, d_1, d_2 and d_3 species all of which may be incorporated into the sandarazol scaffold which explains the isotope patterns seen above.

Furthermore, we also fed stable isotope labelled L-valine and L-leucine in order to prove incorporation of these amino acids by the sandarazol assembly line. To check the integration of L-valine, L-Valine-d_8 was fed to an induced culture of MSr10575 :: pBeloBacSa001.
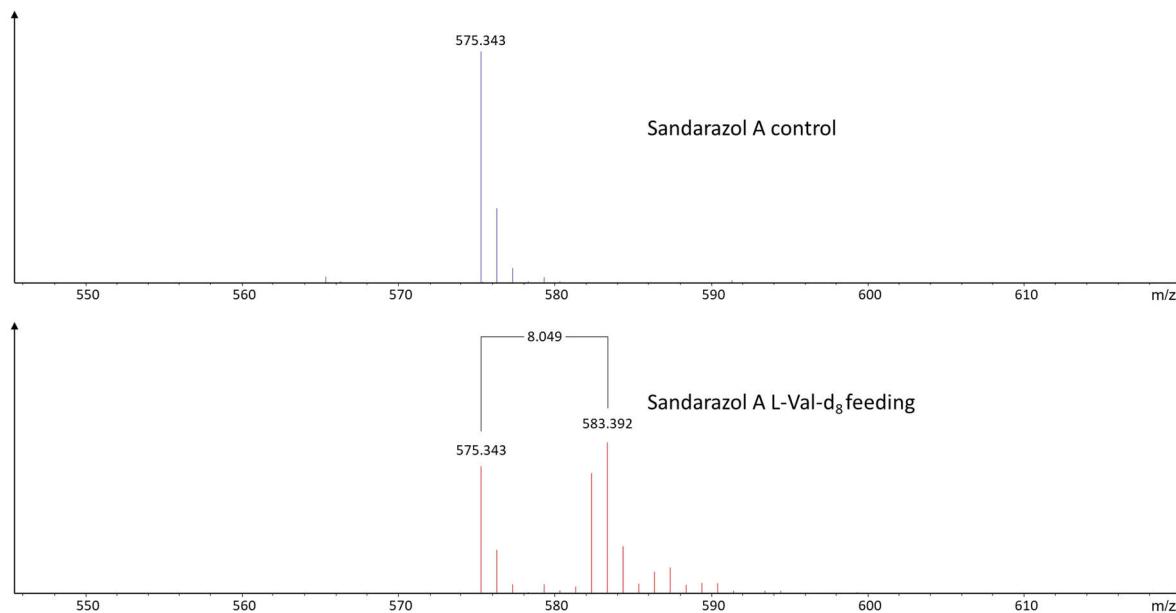
Figure S10 Influence of feeding of L-valine-d_8 on the isotope pattern of sandarazol A

We can see a clear +8 da shift upon feeding of L-Val-d_8, which is in line with incorporation of Valine into the sandarazols. The small amount of +7 shift is caused by D-H exchange at the alpha position of the Valine, which is often observed in feeding experiments. To check the integration of L-Leucine, L-Leucine-d_3 was fed to an induced culture of MSr10575 :: pBeloBacSa001.
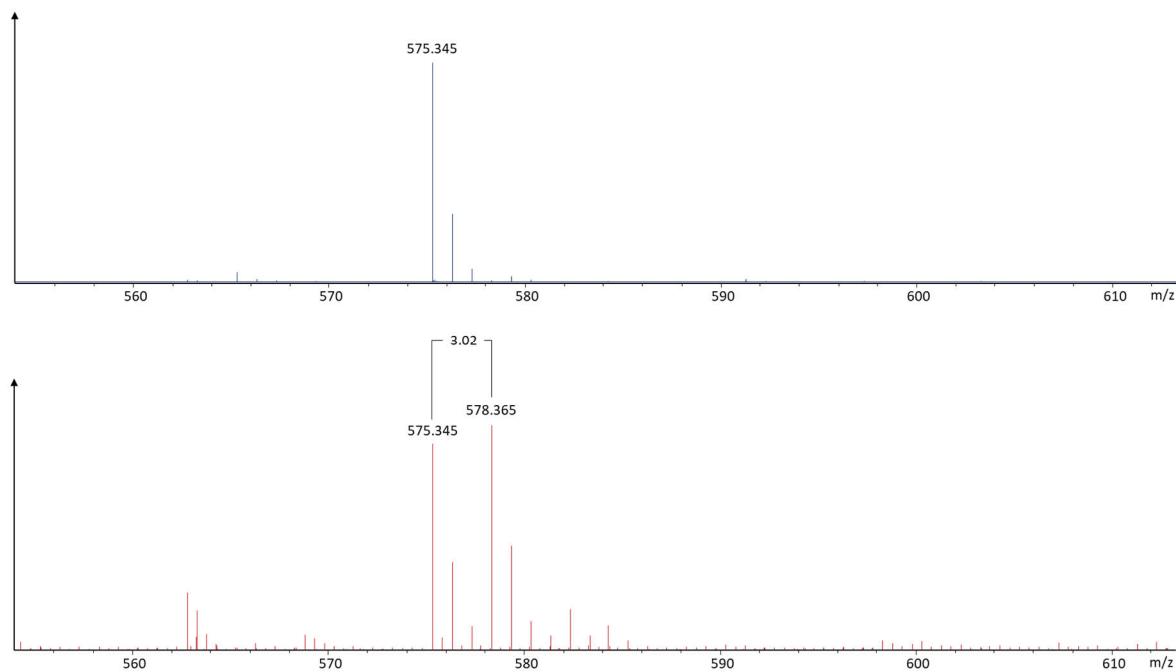
Figure S11 Influence of feeding of L-leucine-d_3 on the isotope pattern of sandarazol A

We observe a mass shift of +3 Daltons upon feeding of L-Leucine-d_3 which indicates incorporation of the labelled leucine into the sandarazols during biosynthesis.

S6.3 Genome sequencing of *Sandaracinus sp.* MSr10575

S6.3.1 Preparation of genomic DNA for PacBio sequencing of MSr10575

To isolate total DNA for sequencing purposes such as PacBio sequencing, phenol-chloroform gDNA extraction is used.

1) Spin down 50 mL of fresh myxobacterial culture 6000 rcf 10 min
2) Discard the supernatant
3) Wash the cells once with SET Buffer, centrifuge at 6000 rcf 10 min
4) Resuspend cell pellet in 5 mL SET Buffer
5) Add 100 µL of lysozyme (50 mg/mL in ddH$_2$O) stock solution as well as 50 µL RNAse A (10 mg/mL in ddH$_2$O) stock solution
6) Add 300 µL Proteinase K solution (10 mg/mL 50 mM Tris 1 mM CaCl$_2$) invert several times and add 600µL 10% SDS solution
7) Incubate at 55 °C for 2 h, invert every 15 min
8) Add even Volume (6 mL) of Phenol/Chloroform/Isoamylalcohol (25:24:1) and swing the tube for 60 min at 5 rpm
9) Centrifuge the mixture at 6000 rcf for 5 min at room temperature
10) Transfer the upper phase into a new tube using a cut end 1 mL tip
11) Add even Volume (6 mL) of Phenol/Chloroform/Isoamylalcohol (25:24:1) and swing the tube for 60 min at 5 rpm
12) Centrifuge the mixture at 6000 rcf for 5 min at room temperature
13) Transfer the upper phase into a new tube using a cut end 1 mL tip
14) Add even Volume (6 mL) of Chloroform/Isoamylalcohol (24:1) and swing the tube for 60 min at 5 rpm
15) Centrifuge the mixture at 6000 rcf for 5 min at room temperature
16) Transfer the upper phase into a new tube using a cut end 1 mL tip
17) Add 1/10 of the total volume of 3 M NaOAc solution pH 5.5 and mix well by inverting several times
18) Add 2.5 Volumes of ice cold ethanol (100% technical purity, -20 °C) and invert the tube several times, DNA precipitation should be visible as a cotton like fog in the tube
19) Spool the DNA on a sealed Pasteur pipette
20) Rinse the DNA with 70% Ethanol (cold , -20 °C)
21) Air dry the DNA for at least 15 minutes (Dry DNA will become completely translucid)
22) Resuspend dried DNA in 0.5 mL of ddH$_2$O and keep the Eppendorf tube at room temperature for 24 Hours

S6.3.2 Results of PacBio sequencing of MSr10575 wild type

The wild type strain *Sandaracinus sp.* MSr10575 was sequenced using a PacBio RS II device at the DSMZ using a single SMRT cell. The raw sequence reads were assembled in the SMRT portal software as recommended by the manufacturer. The MSr10575 genome sequence consists of a closed circular bacterial chromosome spanning 10.754 Mbp and a closed circular extrachromosomal mega plasmid spanning 209.732 kbp. Sequence coverage across the circular MSr10575 bacterial chromosome is approximately 65 fold. Coverage of the bacterial mega plasmid containing the *szo* biosynthetic gene cluster is approximately 130 fold indicating a plasmid copy number per cell of about 2 as bases on the mega plasmid are about twice as likely to be sequenced than bases on the genome. The presence of this extrachromosomal mega plasmid can be easily seen even before assembly by the distribution of sequence coverage depth. Assembly of the genomic DNA confirms this finding as it produces two closed unitigs with the above mentioned characteristics.

Depth Of Coverage

Depth of coverage distribution
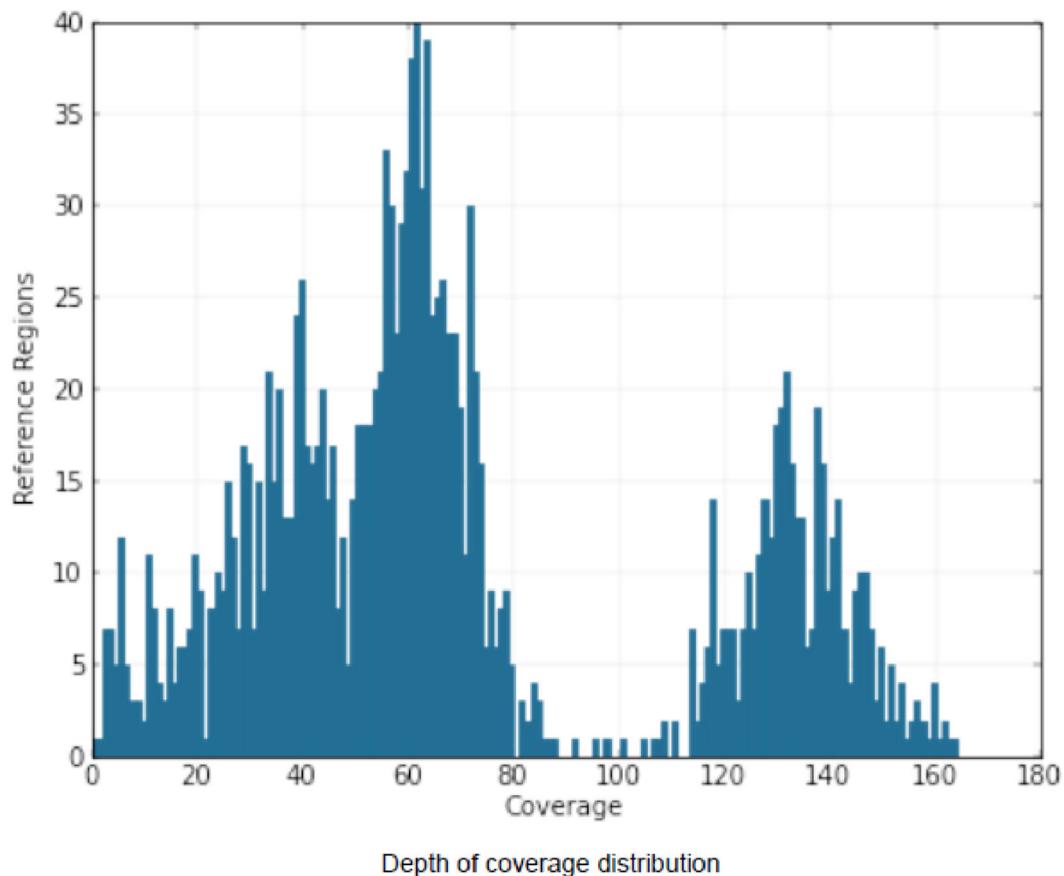
Figure S12 Sequence coverage depth distribution per base revealing the presence of an extrachromosomal element with a copy number of approximately 2

S6.3.3 Codon usage analysis for the MSr10575 chromosome, the pSa001 plasmid and the szo gene cluster

To compare the DNA of MSr10575s Chromosome to its native plasmid pSa001 as well as the corresponding *szo* BGC we performed in depth codon usage analysis.

Table S5 Codon usage analysis for the MSr10575 chromosomal DNA

MSr10575 chromosomal DNA			
Total GC		71%	
Amino Acid (AA)	Freq.	%	
A:	446,706	13.60%	
C:	43,918	1.30%	
D:	210,780	6.40%	
E:	216,177	6.60%	
F:	94,839	2.90%	
G:	289,370	8.80%	
H:	71,343	2.20%	
I:	130,988	4.00%	
K:	50,475	1.50%	
L:	319,459	9.70%	
M:	57,556	1.70%	
N:	47,978	1.50%	
P:	190,035	5.80%	
Q:	77,333	2.40%	
R:	302,988	9.20%	
S:	175,636	5.30%	
T:	175,830	5.30%	
V:	276,434	8.40%	
W:	45,569	1.40%	
Y:	57,756	1.80%	
*:	8,749	0.30%	
AA_Group	Freq.	%	
Acidic:	426957	13.00%	
Basic:	424806	12.90%	
Charged:	851763	25.90%	
Polar_Uncharged:	624020	19.00%	
Hydrophobic:	1850956	56.30%	
GC-rich:	1229099	37.40%	
AT-rich:	382036	11.60%	
Codon	AA	Codon usage per AA	Freq
GCA	A	3.90%	17,528

GCC	A	19.90%	88,883
GCG	A	74.10%	331,147
GCT	A	2.00%	9,148
TGC	C	89.30%	39,218
TGT	C	10.70%	4,700
GAC	D	73.40%	154,779
GAT	D	26.60%	56,001
GAA	E	8.40%	18,206
GAG	E	91.60%	197,971
TTC	F	98.90%	93,795
TTT	F	1.10%	1,044
GGA	G	8.00%	23,240
GGC	G	63.00%	182,206
GGG	G	19.60%	56,606
GGT	G	9.40%	27,318
CAC	H	88.00%	62,797
CAT	H	12.00%	8,546
ATA	I	0.10%	151
ATC	I	98.50%	129,043
ATT	I	1.40%	1,794
AAA	K	2.60%	1,294
AAG	K	97.40%	49,181
CTA	L	0.30%	803
CTC	L	64.60%	206,223
CTG	L	31.70%	101,292
CTT	L	1.10%	3,430
TTA	L	0.00%	53
TTG	L	2.40%	7,658
ATG	M	100%	57,556
AAC	N	92.40%	44,336
AAT	N	7.60%	3,642
CCA	P	1.90%	3,587
CCC	P	36.00%	68,493
CCG	P	58.30%	110,801
CCT	P	3.80%	7,154
CAA	Q	4.20%	3,212
CAG	Q	95.80%	74,121
AGA	R	0.50%	1,384
AGG	R	1.40%	4,327
CGA	R	6.20%	18,776
CGC	R	69.30%	209,886
CGG	R	14.40%	43,521
CGT	R	8.30%	25,094
AGC	S	37.50%	65,925
AGT	S	1.20%	2,074
TCA	S	0.90%	1,546
TCC	S	8.60%	15,157

TCG	S	50.60%	88,949
TCT	S	1.10%	1,985
ACA	T	1.20%	2,110
ACC	T	41.50%	73,025
ACG	T	55.60%	97,746
ACT	T	1.70%	2,949
GTA	V	0.40%	1,155
GTC	V	46.60%	128,748
GTG	V	51.80%	143,167
GTT	V	1.20%	3,364
TGG	W	100%	45,569
TAC	Y	85.10%	49,169
TAT	Y	14.90%	8,587
TAA	*	1.40%	126
TAG	*	10.60%	928
TGA	*	88.00%	7,695

Table S6 Codon usage analysis for the MSr10575 borne pSa001 plasmid

pSa001 plasmid DNA		
Total GC		70%
Amino Acid (AA)	Freq.	%
A:	5,874	13.90%
C:	896	2.10%
D:	2,814	6.70%
E:	2,623	6.20%
F:	1,205	2.90%
G:	3,729	8.80%
H:	982	2.30%
I:	1,524	3.60%
K:	430	1.00%
L:	3,884	9.20%
M:	599	1.40%
N:	658	1.60%
P:	2,564	6.10%
Q:	859	2.00%
R:	3,955	9.40%
S:	2,547	6.00%
T:	2,251	5.30%
V:	3,367	8.00%
W:	588	1.40%
Y:	715	1.70%
*:	112	0.30%

AA_Group	Freq.	%
Acidic:	5437	12.90%
Basic:	5367	12.70%
Charged:	10804	25.60%
Polar_Uncharged:	8514	20.20%
Hydrophobic:	23334	55.30%
GC-rich:	16122	38.20%
AT-rich:	4532	10.70%

Codon	AA	Codon usage per AA	Freq
GCA	A	9.00%	528
GCC	A	22.10%	1,297
GCG	A	62.80%	3,691
GCT	A	6.10%	358
TGC	C	86.20%	772
TGT	C	13.80%	124
GAC	D	70.00%	1,969
GAT	D	30.00%	845
GAA	E	15.90%	416
GAG	E	84.10%	2,207
TTC	F	92.40%	1,114
TTT	F	7.60%	91
GGA	G	13.20%	492
GGC	G	55.50%	2,068
GGG	G	18.30%	681
GGT	G	13.10%	488
CAC	H	81.80%	803
CAT	H	18.20%	179
ATA	I	1.60%	24
ATC	I	89.80%	1,369
ATT	I	8.60%	131
AAA	K	10.50%	45
AAG	K	89.50%	385
CTA	L	1.30%	49
CTC	L	61.60%	2,393
CTG	L	27.00%	1,049
CTT	L	4.60%	179
TTA	L	0.30%	10
TTG	L	5.30%	204
ATG	M	100%	599
AAC	N	82.70%	544
AAT	N	17.30%	114
CCA	P	4.30%	110
CCC	P	30.50%	782
CCG	P	56.70%	1,454
CCT	P	8.50%	218
CAA	Q	13.20%	113

CAG	Q	86.80%	746
AGA	R	1.10%	45
AGG	R	2.60%	101
CGA	R	10.00%	395
CGC	R	56.70%	2,242
CGG	R	17.40%	689
CGT	R	12.20%	483
AGC	S	30.40%	775
AGT	S	4.60%	116
TCA	S	2.80%	71
TCC	S	15.30%	390
TCG	S	43.70%	1,114
TCT	S	3.20%	81
ACA	T	4.00%	91
ACC	T	35.20%	793
ACG	T	55.20%	1,243
ACT	T	5.50%	124
GTA	V	3.10%	105
GTC	V	46.90%	1,579
GTG	V	45.10%	1,517
GTT	V	4.90%	166
TGG	W	100%	588
TAC	Y	80.40%	575
TAT	Y	19.60%	140
TAA	*	3.60%	4
TAG	*	8.00%	9
TGA	*	88.40%	99

Table S7 Codon usage analysis for the *szo* BGC

szo Cluster on pSa001

Total GC		71%	

Amino Acid (AA)	Freq.	%	
A:	2,511	14.80%	
C:	212	1.30%	
D:	1,127	6.70%	
E:	1,061	6.30%	
F:	473	2.80%	
G:	1,541	9.10%	
H:	449	2.70%	
I:	647	3.80%	
K:	151	0.90%	
L:	1,742	10.30%	
M:	202	1.20%	
N:	219	1.30%	
P:	964	5.70%	

Q:	316	1.90%
R:	1,663	9.80%
S:	960	5.70%
T:	820	4.80%
V:	1,370	8.10%
W:	212	1.30%
Y:	276	1.60%
*:	23	0.10%

AA_Group	Freq.	%
Acidic:	2188	12.90%
Basic:	2263	13.40%
Charged:	4451	26.30%
Polar_Uncharged:	3015	17.80%
Hydrophobic:	9662	57.00%
GC-rich:	6679	39.40%
AT-rich:	1766	10.40%

Codon	AA	Codon usage per AA	Freq
GCA	A	8.20%	205
GCC	A	21.60%	543
GCG	A	65.20%	1,636
GCT	A	5.10%	127
TGC	C	85.80%	182
TGT	C	14.20%	30
GAC	D	71.40%	805
GAT	D	28.60%	322
GAA	E	13.50%	143
GAG	E	86.50%	918
TTC	F	94.90%	449
TTT	F	5.10%	24
GGA	G	11.90%	183
GGC	G	55.90%	862
GGG	G	18.40%	283
GGT	G	13.80%	213
CAC	H	84.20%	378
CAT	H	15.80%	71
ATA	I	1.40%	9
ATC	I	92.10%	596
ATT	I	6.50%	42
AAA	K	9.30%	14
AAG	K	90.70%	137
CTA	L	1.10%	19
CTC	L	67.60%	1,178
CTG	L	24.50%	426
CTT	L	3.30%	58

TTA	L	0.20%	3
TTG	L	3.30%	58
ATG	M	100%	202
AAC	N	82.60%	181
AAT	N	17.40%	38
CCA	P	2.70%	26
CCC	P	30.40%	293
CCG	P	60.10%	579
CCT	P	6.80%	66
CAA	Q	9.80%	31
CAG	Q	90.20%	285
AGA	R	0.70%	11
AGG	R	1.40%	23
CGA	R	9.40%	157
CGC	R	60.60%	1,008
CGG	R	15.80%	262
CGT	R	12.10%	202
AGC	S	28.10%	270
AGT	S	2.70%	26
TCA	S	0.90%	9
TCC	S	16.00%	154
TCG	S	49.40%	474
TCT	S	2.80%	27
ACA	T	2.30%	19
ACC	T	36.30%	298
ACG	T	56.70%	465
ACT	T	4.60%	38
GTA	V	2.30%	31
GTC	V	50.60%	693
GTG	V	43.50%	596
GTT	V	3.60%	50
TGG	W	100%	212
TAC	Y	80.80%	223
TAT	Y	19.20%	53
TAA	*	0.00%	0
TAG	*	0.00%	0
TGA	*	100%	23

S6.3.4 Results of Illumina sequencing of MSr10575 mutants

In order to confirm successful homologous recombination in MSr10575 we prepared MSr10575 :: pBeloBac Sa001 genomic DNA and sent it to GMAK (Braunschweig, Germany) for Illumina sequencing. Results confirm that on the one hand, the extrachromosomal element in MSr10575 remains a plasmid and is not recombined into the MSr10575 bacterial chromosome and on the other hand that the homologous recombination to create pBeloBac Sa001 did not create any errors such as SNP's or short insertion/deletions.

S6.4 *In-silico* analysis of the sandarazol biosynthetic gene cluster

The sandarazol biosynthetic gene cluster is a type 1 Trans-AT Polyketide synthase (PKS) non-ribosomal peptide synthethase (NRPS) hybrid biosynthetic gene cluster located on a natural plasmid spanning 209,732 bp in *Sandaracinus sp.* MSr10575 wild type. The biosynthetic gene cluster with the corresponding annotations can be found under GenBank tag Sandaracinus strain MSr10575 pSa001 plasmid under GenBank accession number MW053453. Based on this GenBank file, the BGC will be available at the MiBiG database.

S6.4.1 Overview over the pSa001 plasmid
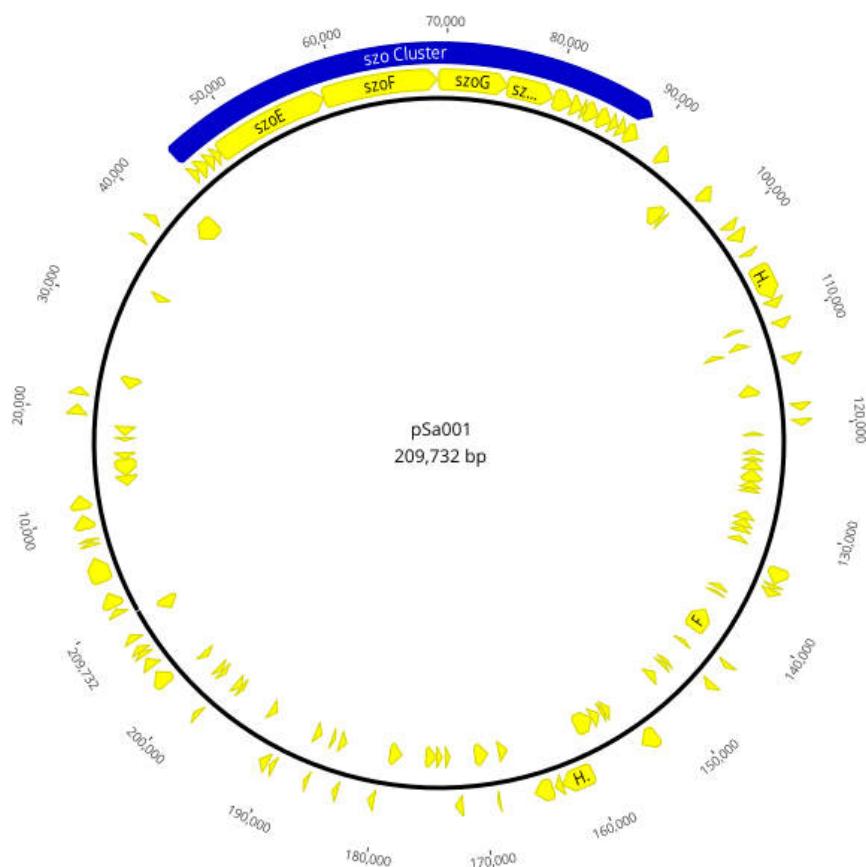
Figure S13 Schematic view of the pSa001 plasmid in MSr10575

Table S8 Tabulated view of all detected CDS on the pSa001 plasmid

Predicted protein product	CDS start	CDS end	Length	reading direction
Hypothetical protein CDS	424	1,092	669	forward
Hypothetical protein CDS	1,239	2,678	1,440	forward
Hypothetical protein CDS	3,780	6,122	2,343	forward
Hypothetical protein CDS	7,308	7,649	342	forward
Hypothetical protein CDS	7,748	8,023	276	forward
Hypothetical protein CDS	8,914	10,161	1,248	forward
Hypothetical protein CDS	10,784	12,055	1,272	forward
Hypothetical protein CDS	12,353	13,375	1,023	reverse

Serine/threonine protein kinase CDS	13,372	15,174	1,803	reverse
Hypothetical protein CDS	15,171	15,899	729	reverse
Hypothetical protein CDS	17,118	17,507	390	reverse
Hypothetical protein CDS	17,721	18,662	942	reverse
Serine/threonine protein kinase CDS	19,526	20,536	1,011	forward
Hypothetical protein CDS	21,401	22,156	756	forward
Transposase family protein CDS	22,701	23,870	1,170	reverse
Hypothetical protein CDS	32,461	33,102	642	reverse
Hypothetical protein CDS	36,823	37,218	396	forward
FG-GAP repeat-HVR domain containing protein CDS	38,898	39,500	603	forward
Hypothetical protein CDS	40,535	42,781	2,247	reverse
szoA CDS	44,462	45,190	729	forward
szoB CDS	45,218	46,258	1,041	forward
szoC CDS	46,255	47,103	849	forward
szoD CDS	47,100	47,843	744	forward
szoE CDS	47,836	58,599	10,764	forward
szoF CDS	58,596	69,254	10,659	forward
szoG CDS	69,290	75,655	6,366	forward
szoH CDS	75,652	80,040	4,389	forward
szoI CDS	80,103	81,854	1,752	forward
szoJ CDS	81,863	82,915	1,053	forward
szoK CDS	82,912	83,196	285	forward
szoL CDS	83,193	84,515	1,323	forward
szoM CDS	84,512	85,747	1,236	forward
szoN CDS	85,744	86,487	744	forward
szoO CDS	86,591	87,379	789	forward
szoP CDS	87,428	88,675	1,248	forward
chitinase CDS	90,944	92,113	1,170	forward
Hemolysin-type calcium binding region CDS	93,669	95,156	1,488	reverse
Hypothetical protein CDS	95,305	95,604	300	reverse
Hypothetical protein CDS	96,300	97,517	1,218	forward
Hypothetical protein CDS	100,082	100,789	708	forward
Oxidoreductase, NmrA-like CDS	101,215	102,189	975	forward
Hypothetical protein CDS	103,311	103,757	447	forward
Hypothetical protein CDS	104,927	108,178	3,252	forward
Hypothetical protein CDS	108,191	109,087	897	forward
Hypothetical protein CDS	109,621	109,992	372	reverse
transposase CDS	110,176	111,045	870	forward
Activator of HSP90 ATPase 1 family protein CDS	111,170	111,658	489	reverse
arsR CDS	111,637	112,071	435	reverse

Hypothetical protein CDS	113,516	114,496	981	forward
Hypothetical protein CDS	115,666	116,799	1,134	reverse
Cobyrinic acid ac-diamidesynthase CDS	118,113	118,772	660	forward
Hypothetical protein CDS	119,528	120,130	603	forward
Hypothetical protein CDS	120,567	121,022	456	reverse
Hypothetical protein CDS	122,528	122,926	399	reverse
Hypothetical protein CDS	123,247	123,663	417	reverse
Hypothetical protein CDS	123,660	124,586	927	reverse
Peptidase, M23/M37 family CDS	124,583	125,872	1,290	reverse
Hypothetical protein CDS	125,869	126,546	678	reverse
Hypothetical protein CDS	126,566	127,024	459	reverse
Hypothetical protein CDS	129,053	129,844	792	reverse
Hypothetical protein CDS	129,857	130,693	837	reverse
Hypothetical protein CDS	130,745	131,302	558	reverse
Hypothetical protein CDS	131,693	132,355	663	reverse
Hypothetical protein CDS	133,584	135,074	1,491	forward
Hypothetical protein CDS	135,169	135,609	441	forward
Hypothetical protein CDS	135,650	136,342	693	forward
Hypothetical protein CDS	137,252	137,734	483	reverse
Hypothetical protein CDS	138,098	138,316	219	reverse
FG-GAP repeat/HVR domain protein CDS	140,368	142,821	2,454	reverse
Hypothetical protein CDS	143,491	144,018	528	forward
Hypothetical protein CDS	144,309	144,659	351	reverse
Hypothetical protein CDS	145,832	146,539	708	forward
Hypothetical protein CDS	147,166	147,534	369	reverse
Hypothetical protein CDS	147,544	147,918	375	reverse
Sigma54 specific transcriptional regulator, Fis family CDS	148,954	149,781	828	reverse
Integrase CDS	152,839	154,389	1,551	forward
transcriptional regulator CDS	155,087	155,770	684	reverse
Glyoxalase /bleomycin resistance protein/ dioxygenase CDS	155,775	156,173	399	reverse
Hypothetical protein CDS	156,401	157,324	924	reverse
Serine/threonine protein kinase CDS	157,324	159,282	1,959	reverse
Hypothetical protein CDS	159,642	162,587	2,946	forward
Serine/threonine specific protein phosphatase (Putative) CDS	162,596	163,369	774	forward
Hypothetical protein CDS	163,605	165,287	1,683	forward
Hypothetical protein CDS	166,983	167,897	915	reverse
Hypothetical protein CDS	168,518	168,739	222	forward

Hypothetical protein CDS	169,022	170,371	1,350	reverse
ThiJ/PfpI domain protein CDS	172,051	172,707	657	forward
Hypothetical protein CDS	173,007	173,519	513	reverse
lipoprotein CDS	173,852	174,481	630	reverse
Hypothetical protein CDS	174,478	175,581	1,104	reverse
Hypothetical protein CDS	178,139	179,374	1,236	reverse
Hypothetical protein CDS	180,285	180,782	498	forward
Transposase CDS	183,745	184,122	378	forward
Transposase IS66 CDS	184,134	184,844	711	reverse
Hypothetical protein CDS	185,405	185,839	435	reverse
ferredoxin/ferredoxin--NADP reductase CDS	186,640	186,900	261	forward
Resolvase domain protein CDS	187,006	187,677	672	reverse
Hypothetical protein CDS	189,926	190,441	516	forward
Hypothetical protein CDS	190,677	191,483	807	forward
Hypothetical protein CDS	192,307	192,924	618	reverse
Hypothetical protein CDS	196,545	196,949	405	reverse
Hypothetical protein CDS	196,943	197,599	657	reverse
Hypothetical protein CDS	198,400	198,837	438	forward
Hypothetical protein CDS	199,053	199,622	570	reverse
Hypothetical protein CDS	199,716	200,141	426	reverse
Hypothetical protein CDS	201,803	202,342	540	reverse
Hypothetical protein CDS	202,505	204,013	1,509	forward
Hypothetical protein CDS	204,523	205,509	987	forward
Transcriptional regulator, MarR family CDS	205,861	206,316	456	forward
Hypothetical protein CDS	206,313	206,825	513	forward
pdxH CDS	207,468	208,112	645	forward
Transposase CDS	208,247	209,497	1,251	reverse

S6.4.2 *In-silico* blast analysis of the sandarazol biosynthetic gene cluster

Every coding sequence in the sandarazol biosynthesis gene cluster (*szo* cluster) was extracted translated and searched with the blastp algorithm against the RefSeq non-redundant protein sequence database at NCBI. [3]

Table S9 Tabulated blastP results for the CDS regions present in the szo biosynthetic gene cluster

CDS Name	Length [AA]	Closest homologue [Organism of origin]	Identity [%] and alignment length [AA]	Proposed function	Accession Nr.
szoA	242	Cytochrome C reductase subunit [Cystobacter fuscus]	32.2; 192	enoyl-reductase	WP_002625886
szoB	346	2,3 diaminopropionate biosynthesis protein SbnA [Acidobacterium sp.]	52.5; 318	diaminopropionate biosynthesis	PYS24491
szoC	282	ACP-acyltransferase [Ralstonia solanacerum]	50.0; 280	malonyl-CoA transfer	WP_064046451
szoD	247	Enoyl-CoA hydratase [Bacillus megatherium]	45.4; 247	*in-trans* dehydration	WP_097812162
szoE	3587	Acyl transferase domain containing Protein [Dendosporobacter quercicolus]	37.8; 1875	Polyketide biosynthesis	SDL84501
szoF	3552	Polyketide synthase PksM [Streptomyces leeuwenhoekii]	33.5; 1882	Polyketide biosynthesis	CQR60407
szoG	2121	Amino acid adenylation domain containing Protein [Brevibacillus sp.]	36.3; 2108	Amino acid incorporation	WP_116333041
szoH	1462	Type 1 Polyketide synthase [Stigmatella aurantiaca]	37.7; 1462	Polyketide biosynthesis	WP_075005930
szoI	583	Halogenase [Burkholderia Pseudomallei]	45.1; 531	Halogenase	WP_119566518
szoJ	350	2,3 diaminopropionate biosynthesis protein SbnB [Acidobacterium sp.]	45.8; 345	diaminopropionate biosynthesis	PYS24498
szoK	94	Acyl carrier protein [Duganella sp.]	43.8; 88	β-branching	WP_056154626
szoL	440	Beta-ketoacyl synthase [Salinispora arenicola]	40.2; 425	β-branching	WP_018792602
szoM	411	Hydroxymethylglutaryl-CoA synthase protein PyxM [Pyxidicoccus fallax]	62.1; 411	β-branching	ASA76639
szoN	243	Enoyl-CoA hydratase/isomerase family protein [Streptomyces sp.]	38.7; 243	β-branching	WP_100840951
szoO	262	Thioesterase [Sorangium cellulosum]	37.0; 262	β-branching	WP_012232991
szoP	415	Monooxygenase [Nitrospora sp.]	29.0; 400	epoxidation	PLY22143

S6.4.3 Analysis of the shifting DH domain on *szoF*

During sandarazol Biosynthesis, we observe the consecutive formation of two β, γ double bonds in module 4 and 5 of the sandarazol biosynthesis. The *in-trans* dehydratase protein SzoD is solely able to form α, β unsaturated precursor molecules as it is the case for CurE and CorN in curacin and corallopyronin biosynthesis respectively for example. [4,5] As these other dehydratase proteins, SzoD has a standard enoyl-CoA dehydratase fold responsible for the formation of α,β unsaturated carbonyls from 3-hydroxycarbonyls. The corresponding β, γ double bonds are subsequently created by the double bond shifting DH domain on the SzoF protein. Double bond shifting DH domains are characterized by the presence of an HxxxGxxxxP motif at the start of the domain and a DxxxQ/H motif at the end of said domain. In shifting DH domains on the other hand, the HxxxGxxxxP is mostly completely present, while the DxxxQ/H is absent.

Table S10 Alignment of the double bond shifting sandarazol DH domain to other shifting (marked with an asterisk *) and non-shifting DH domains

```
Domain name             Motif 1 sequence        Motif 2 sequence

SzoF_DH*                ...HRVHGARVVPG...        ...DGATL...

CorL_DH*                ...HEVFGRPLFPT...        ...NGLLM...

CorJ_DH*                ...HTVLGQRVLLG...        ...DGVIV...

NspC_DH*                ...HTLLGDRVLLG...        ...NSAFL...

RhiE_DH*                ...HQFNHRRILLG...        ...NSAFL...

BaeR_DH*                ...HQFSGEPVLVG...        ...NSAYL...

DifK_DH*                ...HLVFGKPALMG...        ...NSCYM...

EryAII_DH               ...HVVGGRTLVPG...        ...DAVAQ...

DH_Consensus            ...HXXXGXXXXPG...        ...DXXXQ/H...
```

As we can see from this analysis, the SzoF shifting DH domain seems to be an intermediate of the shifting DH domains in corallopyronin on the proteins CorL and CorJ. [4] Therefore, sandarazol biosynthesis seems to contain a dehydratation followed by double bond shifting reaction sequence as proposed in the sandarazol biosynthesis.

S6.5 Genetic manipulation protocols applied to MSr10575

S6.5.1 List of primers used in this study

Table S11 List of all primers used in this study

Primers used for the creation of pBeloBac TransAT	
MSr10575 VanR NotI fwd	GAGCGGCCGCCAGTGTGATG
MSr10575 VanR rev	CGAATTACTGCGCGCCACAAACATATGCGTTTCCTCGCATCGTGG
MSr10575 TransAT fwd	CCACGATGCGAGGAAACGCATATGTTTGTGGCGCGCAGTAATTCG
MSr10575 TransAT NotI rev	CGGCGGCCGCGATGTCGGCTCGCCAA
Van primer int fwd	TCGCCGCTCTTGATCTCGCC
pBeloBac Sa001 int rev	CCACCCGAGCACCTCTTCCG

S6.5.2 PCR reactions and cycler protocols

Amplification of the sandarazol cluster start as well as the vanillate promotor and repressor cassette as well as the overlap extension PCR step was done with the proofreading Phusion polymerase (Thermo scientific). PCR mix and thermocycler program can be found below.

Thermo scientific Phusion Polymerase

Phusion Polymerase PCR mix:
5 µL GC Buffer
2.5 µL dNTP's (2 mM)
1.25 µL DMSO
0.5 µL Primer fwd and rev (100mM)
0.2 µL Phusion DNA Polymerase
0.5 µL gDNA Template (~50 ng/µL)
14.55 µL ddH$_2$O

Phusion Polymerase cycler program

Table S12 Phusion DNA Polymerase PCR program

Step	Time [min:s]	Temperature [°C]
Initial denaturation	4:00	95
	0:30	98
Cycle, repeat 30x	0:15	63
	0:30 per Kbp to amplify	72
Final elongation	10	72
Store	forever	8

S6.5.3 Creation of the pBeloBac TransAT plasmid for single crossover integration

The plasmid is based on the commercially available pBeloBac11 (NEB Biolabs). In a first step, the chloramphenicol resistance on the pBeloBac Backbone is replaced by a kanamycin resistance cassette using standard restriction cloning to form pBeloBacKan.
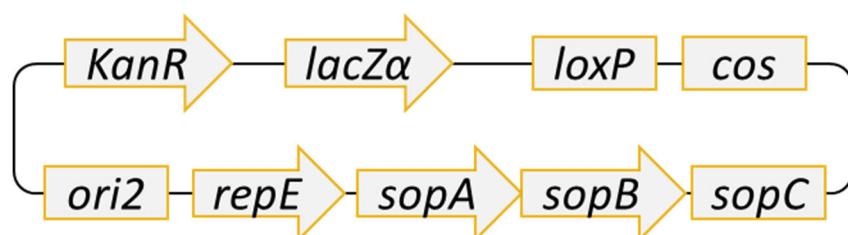

Figure S14 Schematic image of pBeloBac11 kan

The Vanillic acid promotor and repressor cassette are amplified using the above described Phusion polymerase based PCR protocol and the VanR primer pair as described in the primer list from the pFPVan template plasmid described in the Paper describing the discovery of the pyxidicyclines. [6] The sandarazol plasmid cluster fragment necessary for homologous recombination based promotor exchange in front of the sandarazol cluster is amplified using the same PCR protocol from MSr10575

genomic DNA prepared according to the described phenol/chloroform DNA preparation protocol. The two described fragments are fused by overlap extension PCR using the overlapping sequences on the MSr10575 VanR rev and MSr10575 TransAT fwd primers. The fused PCR product is integrated into pBeloBac11 kan using the NotI restriction endonuclease by conventional restriction cloning, thereby replacing the lacZα gene that is not needed. The resulting plasmid is called pBeloBac TransAT. Ligations are performed with T4 DNA ligase (Thermo scientific) according to the manufacturers' instructions.
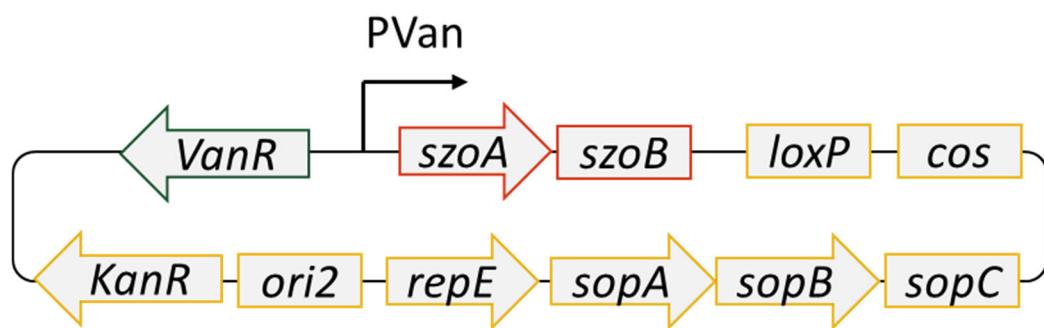
Figure S15 Schematic image of pBeloBac TransAT

The plasmid is confirmed by restriction analysis and prepared in larger amounts for electro transformation into *Sandaracinus sp.* MSr10575.

S6.5.4 Transformation of *Sandaracinus sp.* MSr10575

As transformation of wild type *Sandaracinus* sp*.* MSr10575 was not yet established we devised an electro transformation protocol using pMycoMar Kan, a mariner element based transposon that contains a kanamycin resistance between its IR elements. Good transformation efficiency was obtained using the following electroporation protocol.

Transformation protocol

1) Centrifuge 2 mL of MSr10575 culture in S15 Medium at OD_{600} of approx. 0.8 at 8000 rpm for 2 minutes with an Eppendorf tube table centrifuge.
2) Wash the residual cell pellet 2 times with 1 mL autoclaved ddH_2O and discard the supernatant.
3) Resuspend cells in 50 µL of ddH_2O, add 5 to 10 µL of plasmid solution (prepared from E. coli with Thermo Scientific Miniprep Kit) at a conc. of 0.3-0.4 ng/µL and transfer the solution into an electroporation cuvette.
4) Electroporation at 1000 V, 400 Ω, 25 µF and 1 mm cuvette length settings for optimum electroporation efficiency.
5) Flush out the cells with 1 mL fresh S15 medium and transfer the cell suspension into a 2 mL Eppendorf tube.
6) Incubate the cells for 5 h on a shaker thermostatic to 37 °C at 300 rpm.
7) Plate the cells on Kan50 (or Kan75) S15 agar after mixing the cell suspension with 3 mL of Kan50 S15 soft agar.
8) Orange, spherical clones appear in the soft agar layer after 9-14 days in the 30 °C incubator.

S6.5.5 Integration of pBeloBac TransAT into the native plasmid of MSr10575

The pBeloBac TransAT plasmid is repeatedly transformed into MSr10575 wild type by electroporation according to the devised protocol. The resulting transformants are grown in kanamycin 50 µg/mL supplemented S15 medium and genomic DNA of the transformants is prepared.
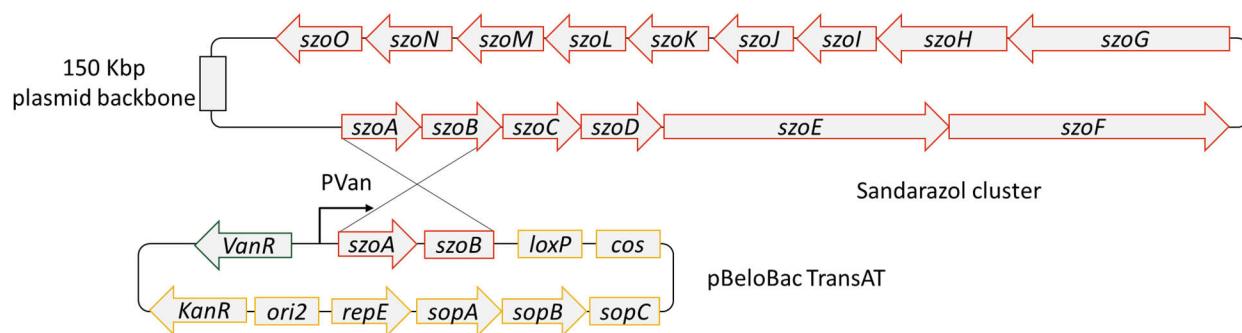

Figure S16 Single crossover integration event that integrates pBeloBac TransAT into the MSr10575 wild type plasmid to form pBeloBac Sa001

Successful integration of pBeloBac TransAT into the MSr10575 wild type plasmid is checked by PCR reaction using the Van primer int fwd and pBeloBac Sa001 int rev primers. After successful homologous recombination of the pBeloBac plasmid with the 209 Kbp pBeloBac Sa001 megaplasmid, MSr10575 clones show a 2359 bp PCR product in this PCR reaction. The strain showing this PCR product is genome

sequenced by Illumina to confirm successful homologous recombination on the megaplasmid to form pBeloBac Sa001.
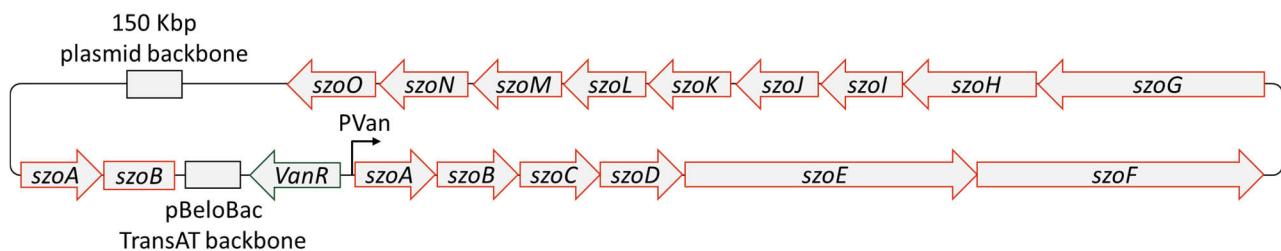

Figure S17 Organisation of the szo cluster on pBeloBac Sa001

S6.6 Isolation and structure elucidation of the sandarazols

S6.6.1 Isolation of the sandarazols from MSr10575 :: pBeloBac Sa001

Cultivation and extraction

The *Sandaracinus* strain MSr10575 :: pBeloBac Sa001 is fermented in 50 mL 2SWT medium as a seed culture flasks on an Orbiton shaker at 160 rpm and 30 °C. The translucid culture medium becomes orange and opaque after 7 to 11 days of fermentation. This pre-culture is used to inoculate 6 x 2 L 2SWYT medium supplemented with 2 % XAD-16 resin suspension in sterilized water in 6 x 5 L baffled shake flasks on an Orbiton shaker at 160 rpm and 30 °C. Directly after inoculation the culture is induced with 1 mM sterile filtrated K-vanillate solution. Fermentation is complete after 16 days. Cells and XAD-16 resin are harvested by centrifugation on a Beckmann Avanti J-26 XP with the JLA 8.1 rotor at 6000 rcf. Combined resin and cells are freeze dried and subsequently extracted using 2x 500 mL of a 2:1 mixture of methanol and chloroform. The combined extracts are concentrated on a rotary evaporator and partitioned between methanol and hexane. The sandarazols remain in the methanol phase. The methanol phase is dried with a rotary evaporator and the residue is partitioned between water and chloroform. The sandarazols are retained in the chloroform phase. The chloroform phase is concentrated and stored in an air-tight glass vial at -20 °C until further processing.

Centrifugal partition chromatography (CPC) pre-purification of the compounds

Pre-purification and fractionation of the different sandarazols into four fractions is achieved by CPC. We use a Gilson CPC 100 device (Gilson purification S.A.S.) connected to a Varian ProStar Solvent delivery module and a Varian ProStar 2 Channel UV detector. Fraction collection is done with a Foxy Jr. autosampler (Isco). Used biphasic solvent system is the buffered ARIZONA solvent system consisting of a 1:1:1:1 mixture of 10 mM Tris x HCl buffer pH 8.0 (Sigma), methanol (analytical grade, Fluka), EtOAc (analytical grade, fluka) and hexane (analytical grade, fluka). After equilibration of the system and loading of the sample, the aqueous phase is used as a stationary phase for 40 min at 2 mL/min and a fraction size of 4 ml. Then the organic phase is used as a stationary phase in Dual Mode for 40 min at 2 mL/min and a fraction size of 4 mL. The CPC process is followed up on the UV detector at the wavelengths 220nm and 254nm. After checking the fractions using our standard analytical LC-*hr*MS setup we create a total of 3 fractions of sandarazols, one containing the larger sandarazols of above 700 Da, one containing the chlorinated smaller sandarazols and one containing the smaller non-chlorinated sandarazols below 700 Da.

Purification of sandarazol A, B and F by HPLC

Purification of sandarazol A, B and F is carried out using a Dionex Ultimate 3000 SDLC low pressure gradient system on a Waters Acquity CSH C18 250x10mm 5µm column with the eluents H_2O 10 mM ammonium formate pH 7 as A and ACN + 10 mM ammonium formate pH 7 as B, a flow rate of 5 mL/min and a column thermostatic at 30 °C. Sandarazol A and B are detected by UV absorption at 254 nm and purification is done by time dependent fraction collection. Separation is started with a plateau at 60% A for 2 minutes followed by a ramp to 24% A during 24 minutes and a ramp to 0% A during 1 minute. The A content is kept at 0% A for 2 minutes. The A content is ramped back to starting conditions during 30 seconds and the column is re equilibrated for 2 minutes. After evaporation, the sandarazols A, B and F are obtained as pale yellowish amorphous solids. After purification, LC-hrMS analysis shows a single peak with an exact mass of 575.344, 577.360 and 561.365 [M+H]$^+$ for sandarazol A, B and F respectively.

Purification of Sandarazol C by HPLC

Purification of sandarazol C is carried out using a Dionex Ultimate 3000 SDLC low pressure gradient system on a Waters Acquity CSH C18 250x10mm 5µm column with the eluents H_2O 10 mM ammonium formate pH 7 as A and ACN + 10 mM ammonium formate pH 7 as B, a flow rate of 5 mL/min and a column thermostatic at 30 °C. Sandarazol C is detected by UV absorption at 254 nm and purification is done by time dependent fraction collection. Separation is started with a plateau at 50% A for 2 minutes followed by a ramp to 20% A during 24 minutes and a ramp to 0% A during 1 minute. The A content is kept at 0% A for 2 minutes, before ramping back to starting conditions during 30 seconds and column re equilibration for 2 minutes. After solvent evaporation, sandarazol C is obtained as dark yellow amorphous solids. After purification, LC-hrMS analysis shows a single peak with an exact mass of 583.325 [M+H]$^+$ for sandarazol C.

S6.6.2 NMR based structure elucidation

NMR conditions and spectroscopic data

1D and 2D NMR data used for structure elucidation of the sandarazol derivatives is acquired in Methanol-d_4 on a Bruker Ascend 700 spectrometer equipped with a 5 mm TXI cryoprobe (^{1}H at 700 MHz, ^{13}C at 175 MHz). All observed chemical shift values (δ) are given in in ppm and coupling constant values (J) in Hz. Standard pulse programs were used for HMBC, HSQC, and gCOSY experiments. HMBC experiments were optimized for $^{2,3}J_{C-H}$ = 6 Hz. The spectra were recorded in methanol-d_4 and chemical shifts of the solvent signals at δ^H 3.31 ppm and δ^C 49.2 ppm were used as reference signals for spectra calibration. To increase sensitivity, the measurements were conducted in a 5 mm Shigemi tube (Shigemi Inc., Allison Park, PA 15101, USA). The NMR signals are grouped in the tables below and correspond to the numbering in the schemes corresponding to every table. All structure formulae devised by NMR will be made publicly available under their corresponding name in NPatlas. [7,8]
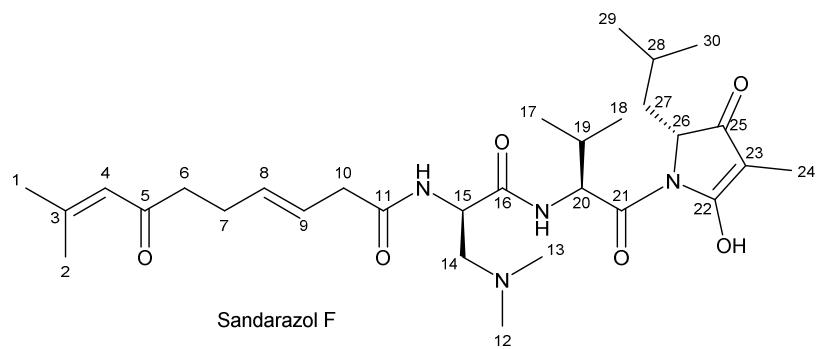
Sandarazol F

Table S13 NMR spectroscopic data for sandarazol F in methanol-d_4.

#	δ ^{13}C [ppm]	δ ^{1}H [ppm], mult (J [Hz])	COSY	HMBC
1	27.7	1.92 (bd, 1.1)[+]	2, 4	2, 3, 4
2	20.8	2.13 (3H, bd, 1.1)	1, 4	1, 3, 4
3	157.0	-	-	-
4	124.5	6.19 (1H, quin, 1.3)	1, 2	1, 2, 5
5	202.3	-	-	-
6	44.1	2.56 (2H, bt, 7.3)	7	5, 7, 8
7	28.0	2.32 (2H, bdd, 14.0, 7.0)	6, 8	5, 6, 8, 9
8	135.1	5.69* (2H, m)	7, 9	10
9	124.1	5.62* (2H, m)	8, 10	7
10	40.5	3.02 (2H, q, 6.4)	9	9, 12,
11	174.7	-	-	-
12	44.9	2.63 (3H, bs)	-	13, 14
13	44.9	2.63 (3H, bs)	-	12, 14
14	60.4	2.93*[+], 3.11*[+]	15	12, 13
15	50.6	4.78* (1H, bt)	14	16
16	171.1	-	-	-
17	19.8	0.98*[+]	19	18, 19
18	19.8	0.98*[+]	19	17, 19
19	31.8	2.19, m[+]	17, 18	17, 18
20	61.5	4.47*[+]	-	21
21	171.9	-	-	-
22	176.1	-	-	-
23	96.9	-	-	-
24	5.7	1.60 (3H, s)	-	22, 23, 25
25	188.8	-	-	-
26	61.4	4.15 (1H, q, 3.0)	27	-
27	39.9	1.92*[+], 1.77*[+]	26	29, 30
28	24.8	1.72*[+]	-	-
29	23.6	0.97*[+]	-	27
30	23.6	0.97*[+]	-	27

* overlapping signals hindering determination of multiplicity and coupling constants

[+] overlapping signals hindering integration
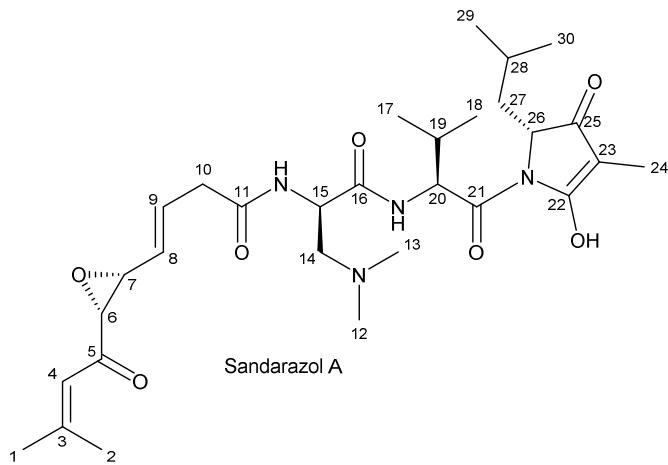
Table S14 NMR spectroscopic data for sandarazol A in methanol-d_4.

#	δ ^{13}C [ppm]	δ ^{1}H [ppm], mult (J [Hz])	COSY	HMBC
1	28.3	1.97 (3H, d, 1.2)	2, 4	2, 3, 4
2	21.6	2.18[+] (d, 1.1)	1, 4	1, 3, 4
3	162.3	-	-	-
4	119.6	6.24 (1H, quin, 1.0)	1, 2	1, 2, 5
5	197.0	-	-	-
6	62.4	3.46* (1H)	7	3, 4, 5, 7, 8
7	59.1	3.47* (1H)	6, 8	6, 8, 9
8	131.5	5.45 (1H, dd, 15.4, 7.7)	7, 9	6, 7, 9, 10, 11
9	131.6	6.19 (1H, bdt, 7.2)	8, 10	7, 8, 10, 11
10	40.3	3.14* (2H)	9	8, 9, 11
11	173.6	-	-	-
12	45.2	2.58 (3H, s)	-	13, 14
13	45.2	2.58 (3H, s)	-	12, 14
14	60.8	2.87*, 3.04* (2H)	15	12, 13, 15
15	51.2	4.78 (1H, bt)	14	11, 14, 16
16	171.6	-	-	-
17	20.5	0.98 (3H, d, 6.8)	19	18, 19, 21
18	17.4	0.86 (3H, d, 6.6)	19	17, 19, 21
19	32.2	2.18*[+]	17, 18	17, 18, 21
20	61.8	4.45*[+]	-	16, 19
21	171.5	-	-	-
22	176.5	-	-	-
23	97.1	-	-	-
24	6.0	1.58 (3H, s)	26	22, 23, 25, 26
25	189.2	-	-	-
26	61.8	4.13 (1H, dd, 5.7, 3.3)	24, 27	21, 22, 24, 25, 27, 28
27	40.0	1.92, 1.75 (2H, m)	26	25, 26, 28, 29, 30
28	25.0	1.71 (1H, quin, 6.5)	27, 29, 30	26, 27, 29, 30
29	24.6	0.81 (3H, d, 6.5)	28, 30	27, 28, 30
30	24.2	0.86[+] (3H, d, 6.6)	28, 29	27, 28, 29

* overlapping signals hindering determination of multiplicity and coupling constants

[+] overlapping signals hindering integration
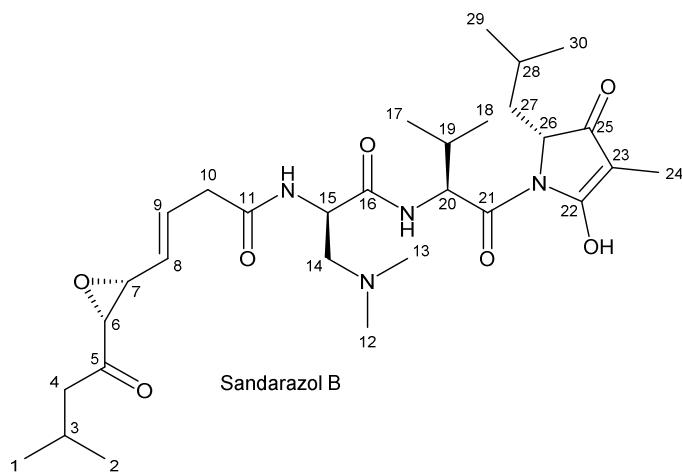
Sandarazol B

Table S15 NMR spectroscopic data for sandarazol B in methanol-d_4.

#	δ ^{13}C [ppm]	δ ^{1}H [ppm], mult (J [Hz])	COSY	HMBC
1	22.8	0.92*+	3	2, 3, 4
2	22.8	0.92*+	3	1, 3, 4
3	25.1	2.12*+	1, 2, 4	1, 2, 4
4	47.3	2.29*+, 2.39*+	3	1, 2, 3, 5
5	208.1	-	-	-
6	61.5	3.47 (1H, d, 1.7)	7	5, 7, 8
7	58.4	3.53* (1H, bdd, 7.8, 1.6)	6, 8	5, 6, 8, 9
8	130.9	5.44 (1H, dd, 15.4, 7.7)	7, 9	6, 7, 10
9	131.3	6.19* (1H, m)	8, 10	7, 10
10	40.0	3.15*+	9	8, 9
11	173.2	-	-	-
12	44.9	2.59 (3H, s)	-	13, 14
13	44.9	2.59 (3H, s)	-	12, 14
14	60.5	2.87*, 3.03* (2H)	15	12, 13, 15
15	50.8	4.75* (1H)	-	11, 14, 16
16	171.4	-	-	-
17	20.1	0.98*+	18, 19	18, 19
18	17.2	0.86*+	17, 19	17, 19
19	32.2	2.18*+	17, 18	17, 18, 21
20	61.8	4.45*+	-	16, 19
21	n.d.	-	-	-
22	176.3	-	-	-
23	97.0	-	-	-
24	5.7	1.59 (3H, bs)	-	22, 23, 25
25	188.8	-	-	-
26	61.4	4.15 (1H, dd, 5.6, 3.2)	27	25, 27, 28
27	39.7	1.91+, 1.76+ (m)	26	25, 26, 28, 29, 30
28	24.8	1.71*+	27, 29	27, 29, 30
29	24.4	0.81 (3H, d, 6.5)	28, 30	27, 28, 30
30	24.0	0.86*+	28, 29	27, 28, 29

* overlapping signals hindering determination of multiplicity and coupling constants

+ overlapping signals hindering integration
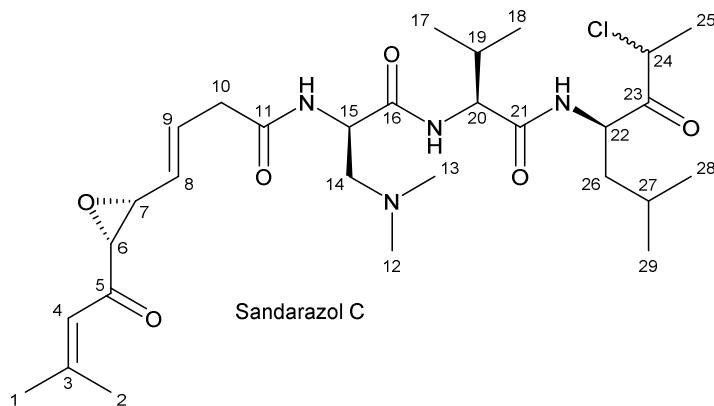
Table S16 NMR spectroscopic data for sandarazol C in methanol-d_4.

#	δ ^{13}C [ppm]	δ ^{1}H [ppm], mult (J [Hz])	COSY	HMBC
1	21.5	2.19 (3H, bs)	2, 4	2, 3, 4, 5
2	28.2	1.97 (3H, d, 1.1)	1, 4	1, 3, 4
3	162.3	-	-	-
4	119.6	6.24 (1H, dquin, 5.1, 1.3)	1, 2	1, 2, 5
5	197.2	-	-	-
6	62.4	3.44* (1H)	7	4, 5, 7, 8, 9
7	59.2	3.44* (1H)	6, 8	5, 6, 8, 9
8	131.1	5.40 (1H, ddd, 15.4, 7.9)	7, 9	6, 7, 9, 10
9	131.9	6.12 (1H, dt, 15.4, 7.1)	8, 10	7, 8, 11, 10
10	40.0	3.09 (2H, m)	9	8, 9, 11
11	173.5	-	-	-
12	45.8	2.32 (3H, s)	-	13, 14
13	45.8	2.32 (3H, s)	-	12, 14
14	61.3	2.59 (2H, m)	15	12, 13, 15, 16
15	53.1	4.50 (1H, q, 7.3)	-	11, 14, 16
16	173.9	-	-	-
17	19.9	0.99*+	18, 19	18, 19
18	18.2	0.94*+	17, 19	17, 19
19	31.5	2.21*+	17, 18	17, 18, 21
20	60.1	4.23*+	-	16, 19
21	180.7	-	-	-
22	55.6	4.74 (1H, m)	25, 26	23, 26, 27
23	204.6	-	-	-
24	56.3	4.79 (1H, m)	25	25
25	20.7	1.53 (3H; dd, 8.8, 6.8)	24	24, 26, 22
26	39.8	1.64 (2H, m)	22, 27	22, 23, 27, 28, 29
27	26.2	1.72 (1H, m)	26, 28, 29	22, 26, 28, 29
28	23.8	0.96*+	27, 29	26, 27, 29
29	21.7	0.93*+	27, 28	26, 27, 28

* overlapping signals hindering determination of multiplicity and coupling constants

+ overlapping signals hindering integration

Structure elucidation based on relevant chemical shifts and correlations

The ^{1}H spectrum of sandarazol A showed nine methyl groups at δ^1H = 0.81 (3H, d, 6.5), 0.86 (6H, d, 6.6), 0.98 (3H, d, 6.8), 1.58 ppm (3H, s), 1.97 (3H, d, 1.2), 2.18 (d, 1.1) and 2.58 (6H, bs). Furthermore, the spectrum revealed three methylene groups at δ^1H = 1.75/1.92 (2H, m), 2.87/3.04 (2H) and 3.14 (2H) ppm. Two methine groups were found at δ^1H = 4.78 (1H, bt) and 4.45 ppm, whose characteristic chemical shift revealed them as α-protons. COSY correlations of the methine group at δ^1H = 4.45 ppm to another methine group at 2.19 ppm and HMBC correlations to two of the nine methyl groups at 0.98 and 0.86 ppm, as well as HMBC correlations of those groups to a quaternary carbon at δ^{13}C = 171.5 ppm revealed the first amino acid to be valine. The second α-proton at δ^1H = 4.78 ppm showed COSY correlations to a diastereotopic methylene group at δ^1H = 2.87/3.04, which shows HMBC correlations to two methyl groups at δ^1H = 2.58 ppm and one quaternary carbon at δ^{13}C = 171.6 ppm. The ^{13}C downfield chemical shift of the methylene group at δ^{13}C = 60.8 ppm, as well as the downfield shift of the two methyl groups at δ^{13}C = 45.2 ppm, besides the correlations described beforehand, revealed the second amino acid to be 2-Amino-3-(N,N-dimethyl amino)-propanoic acid. HMBC correlations of the valine methine group at δ^1H = 4.45 ppm to the acid function of the 2-Amino-3-(N,N-dimethyl amino)-propanoic acid suggested their amide connection at valine the α-proton.

HMBC correlations of the 3N,3N-dimethyl-2,3-diaminopropionic acid α-proton to a quaternary carbon at δ^{13}C = 173.6 ppm suggested its participation in another amide function. HMBC and COSY correlations of two downfield methine groups at δ^1H = 6.19 and 5.45 ppm, as well as one methylene group at δ^1H = 3.14 ppm revealed unsaturation in β position of the PKS part of the molecule. The aliphatic double bond proton at δ^1H = 5.45 ppm showed COSY correlations to a methine group at δ^1H = 3.47 ppm which could, along with its neighboring methine group at δ^1H = 3.46 ppm, be determined as epoxide function due to their characteristic chemical shifts. The downfield chemical shift of the quaternary carbon at δ^{13}C = 197.0 ppm which shows HMBC correlations to the two epoxide methines revealed the next functional group as a ketone. This ketone shows further HMBC correlations to one aliphatic double bond proton at δ^1H = 6.24 ppm, as well as two methyl groups at δ^1H = 2.18 and 1.97 ppm. Only one additional quaternary carbon at δ^{13}C = 162.3 ppm showed further HMBC correlations to the two methyl groups and the aliphatic double bond proton, revealing the N-terminal end of sandarazol A here.

The carboxylic acid of valine shows correlations to a methine group at δ^1H = 4.13 ppm, suggesting elongation through an amide bond in this part of the molecule as well. The methine group exhibits COSY correlations to a diastereotopic methylene group at δ^1H = 1.92/1.75 ppm and HMBC correlations to one methine and two methyl groups at δ^1H = 1.71, 0.86 and 0.81 ppm. Their characteristic chemical shifts as well as further correlations to a quaternary carbon at δ^{13}C = 189.2 ppm revealed this part of the molecule to derive from leucine. The downfield chemical shift of the quaternary carbon at δ^{13}C = 189.2 ppm and further HMBC correlations of the methine group at δ^1H = 4.13 ppm to two quaternary carbons at δ^{13}C = 97.1 and 176.5 as well as a methyl group at δ^1H = 1.58 ppm suggested further elongation of the molecule in C-terminal direction of the leucine. The characteristic chemical shift of the methyl group at δ^1H = 1.58 ppm and δ^{13}C = 6.0 ppm, alongside with the sum formula of sandarazol A revealed, that the elongated leucine forms a five-membered heterocycle at the nitrogen atom of the amide bond to valine.

Sandarazol B shows high similarity to sandarazol A, as it only differs in the terminal part of the PKS part of the molecule. Instead of the quaternary carbon at δ^{13}C = 162.3 ppm and the aliphatic double bond

proton at δ^1H = 6.24 ppm, the HSQC spectrum of sandarazol B reveals one diastereotopic methylene and one methine group at δ^1H = 2.29/, 2.39 and 2.12 ppm, indicating saturation of the terminal double bond.

In sandarazol C, the PKS part is the same as in sandarazol A. It is fused to 3*N*,3*N*-dimethyl-2,3-diaminopropionic acid and valine as well. The leucine derived heterocycle however is missing in sandarazol C. The methine group showing HMBC correlations to the carboxylic acid of valine is downfield shifted to δ^1H = 4.74 ppm, but the subsequent leucine part shows similar chemical shifts and correlations like in sandarazol A.. The methyl group at δ^{13}C = 6.0 in sandarazol A however, is shifted to δ^{13}C = 20.7 ppm in sandarazol C, revealing its terminal position of an aliphatic chain now. It exhibits correlations to a methine group at δ^1H = 4.79 ppm and a quaternary carbon at δ^{13}C = 204.6 ppm, a ketone deriving from the leucine carboxylic acid. The sum formula of sandarazol C in line with the chemical shifts in this part of the molecule revealed chlorination at the methine group at δ^1H = 4.79 ppm.

Sandarazol F also shows high similarity to sandarazol A, as it only differs in the middle part of the PKS part of the molecule. The two epoxide bearing methines are replaced by two methylene groups at δ^1H = 2.56 and 2.32 ppm in sandarazol F, revealing missing epoxidation of the molecule.

### S6.6.3 Elucidation of the absolute stereochemistry

Marfey's analysis protocol

To determine absolute configurations of amino acids the derivatization method Marfey's analysis is employed. Approximately 50 µg of the peptide to analyze is dried in a glass vial at 110 °C. 100 µL 6N HCl are added, the vial is filled with N2 gas and incubated at 110 °C for 45 minutes to 20 hours, depending on peptide stability, for peptide hydrolysis. The vial is subsequently opened and containing fluid is dried at 110 °C. The residue is taken up in 100 µL milliQ H2O and split into two 2 mL Eppendorf tubes. To each tube 20 µL of 1 N NaHCO3 is added as well as 20 µL of D- respective L-FDLA as a 1% solution in acetone. The mixture is incubated at 40 °C and centrifuged at 700 rpm. 10 µL 2 N HCL solution is added to quench the reaction and 300 µL ACN is added to obtain a total volume of 400 µL. The Eppendorf tube is centrifuged at 15000 rpm in a table centrifuge and transferred into a conical HPLC vial.

Separation is done on Waters Acquity BEH C-18 1.7 µ, 100 mm x 2.1 mm at the maxis 4G coupled UHPLC-*hr*MS system. Eluents are water dd. + 0.1% formic acid as A and acetonitrile dest. + 0.1% formic acid as B. Column is thermostatic at 45 °C and volume flow rate is set to 600 µL/min. The gradient is a linear gradient from 5 to 10% B over 1 min, followed by a linear gradient to 35% B over 14 min, followed by a linear gradient over 6 min to 50% B and linear gradient to 80% B. 80% B are held for one minute and the column is reconditioned with a linear gradient of 1 min to 5% B. Detection of the Marfey's derivatives is done by mass spectrometry and UV detection at 340 nm. Identification of the correct stereochemistry of the amino acid is done via comparison of retention times to FDLA derivatized standards. [9]

Marfey's analysis of sandarazol A

Sandarazol A poses some significant challenges to Marfey's analysis as on the one hand, Leucine has to be released from the compound following a retro aldol decomposition of the terminal heterocyclic system as seen in Scheme S1.
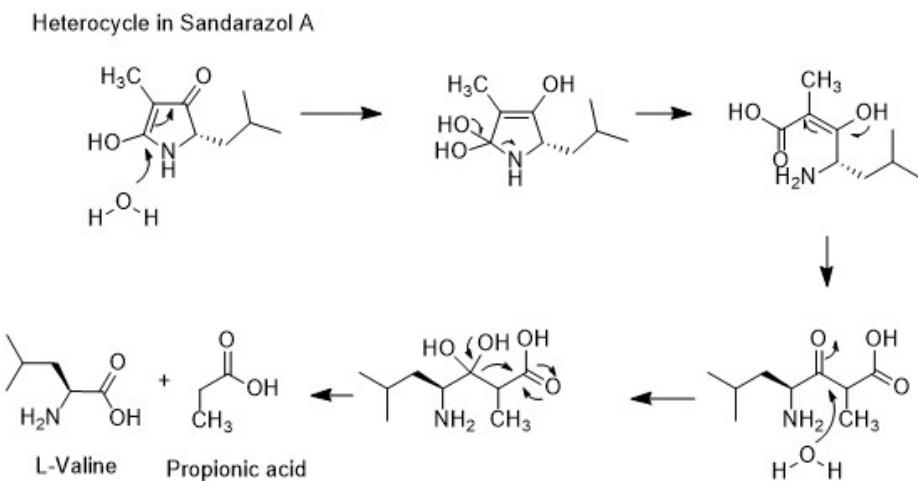
Scheme S1 Acid catalyzed retro aldol reaction that releases L-Valine from the terminal heterocycle in sandarazol A

This reaction requires rather harsh hydrolysis conditions. On the other hand configuration assignment of the 2-Amino-3-(N,N-dimethyl amino)-propanoic acid (L-DimeDap) has to be done following very mild hydrolysis conditions as this amino acid is very prone to racemization even under acidic conditions. Thus, one hydrolysis reaction is set up according to the protocol above using 6N HCl for 15 min at 110 °C to avoid DimeDap racemization and one Marfey's reaction is set up using conc. (37%, 12N) HCl for 12 h at 110 °C to get sufficient turnover in the retro aldol reaction. Used standard substances are L-Valin and L-Leucin (Sigma Aldrich) as well as Boc-L-DimeDap (Boc-(2S)-2-Amino-3-(N,N-dimethyl amino)-propanoic acid, Thermo Fischer) that are all commercially available. It is worth noticing that the Boc protection group will be removed in the acid treatment of the reference compound and therefore the reference amino acid in the Marfey's reaction is L-DimeDap. The retention times for all Marfey's derivatized amino acids are tabulated below.

Table S17 Retention times and exact masses of the Marfeys derivatives

Sample/Standard	FDLA derivative	Exact mass [M+H]$^+$	Retention time	Detected species
L-Leu	D-FDLA	420.204 Da	21.04 min	L-Leu-D-FDLA
L-Leu	L-FDLA	420.204 Da	17.67 min	L-Leu-L-FDLA
L-Val	D-FDLA	412.189 Da	19.57 min	L-Val-D-FDLA
L-Val	L-FDLA	412.189 Da	16.00 min	L-Val-L-FDLA
Boc-L-DimeDap	D-FDLA	427.201 Da	8.96 min	L-DimeDap-D-FDLA
Boc-L-DimeDap	L-FDLA	427.201 Da	9.23 min	L-DimeDap-L-FDLA
Sandarazol A	D-FDLA	420.204 Da	21.03 min	L-Leu-D-FDLA
Sandarazol A	D-FDLA	412.189 Da	19.56 min	L-Val-D-FDLA
Sandarazol A	D-FDLA	427.201 Da	9.23 min	D-DimeDap-D-FDLA
Sandarazol A	L-FDLA	420.204 Da	17.67 min	L-Leu-L-FDLA
Sandarazol A	L-FDLA	412.189 Da	16.02 min	L-Val-L-FDLA
Sandarazol A	L-FDLA	427.201 Da	8.94 min	D-DimeDap-L-FDLA

From the measured retention times as well as from the fact that enantiomers cannot differ in retention time on an achiral HPLC column as the one we used, we can conclude sandarazol A to contain D-DimeDap, L-Val and L-Leu. This is also in accordance with the modules arrangement in the sandarazol biosynthetic gene cluster as the only module containing an epimerization domain is the one introducing DimeDap into the sandarazols.
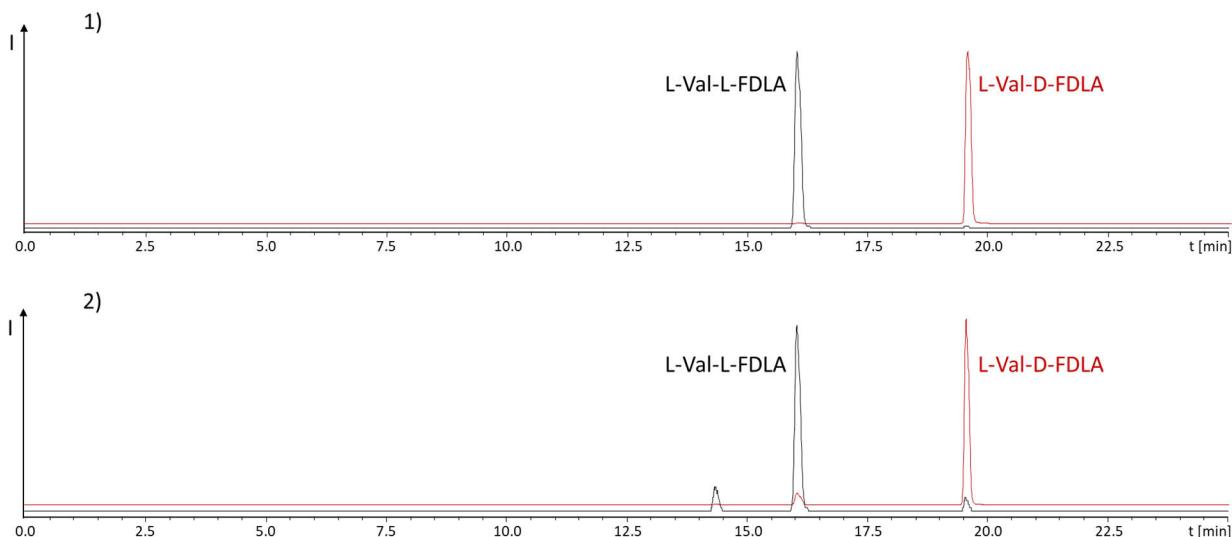

Figure S18 1) Derivatization of L-Val standard with L-FDLA (black trace) and D-FDLA (red trace). 2) Derivatization of hydrolyzed sandarazol with L-FDLA (black trace) and D-FDLA (red trace). The traces represent EIC chromatograms at 412.189 Da. Peak retention time comparison reveals the Valine present in sandarazol to be L- configured.
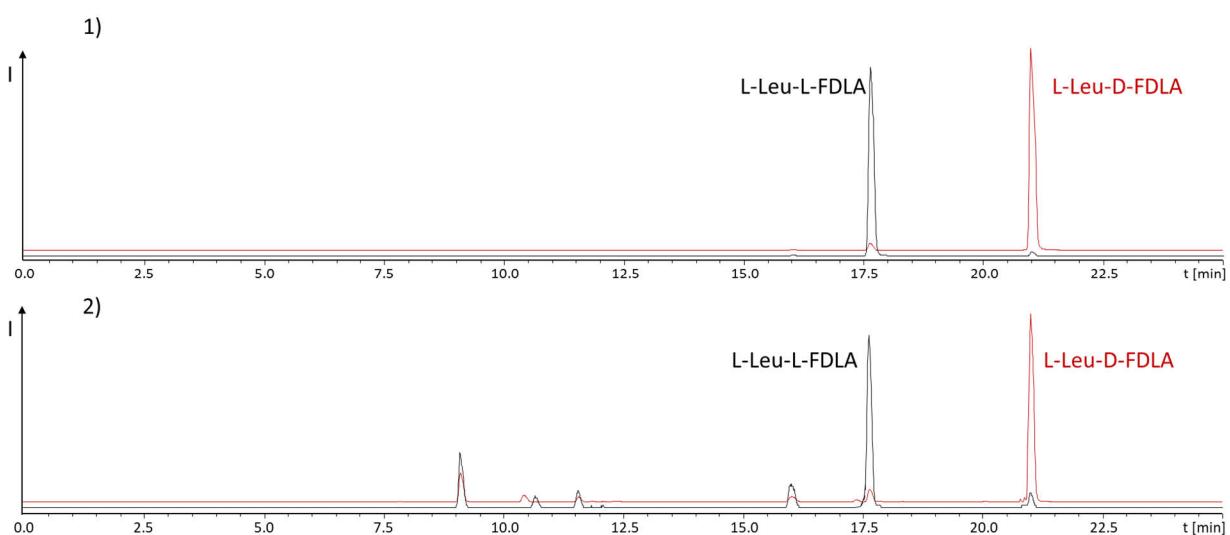
Figure S19 1) Derivatization of L-Leu standard with L-FDLA (black trace) and D-FDLA (red trace). 2) Derivatization of hydrolyzed sandarazol with L-FDLA (black trace) and D-FDLA (red trace). The traces represent EIC chromatograms at 420.204 Da. Peak retention time comparison reveals the Leucine present in sandarazol to be L- configured.
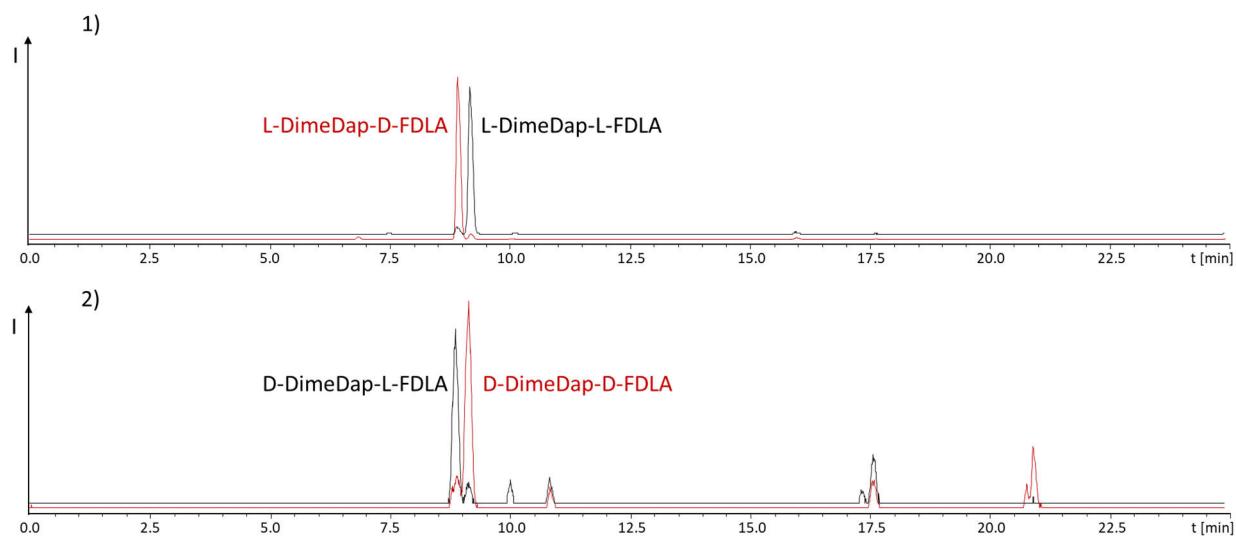
Figure S20 1) Derivatization of L-DimeDap standard with L-FDLA (black trace) and D-FDLA (red trace). 2) Derivatization of hydrolyzed sandarazol with L-FDLA (black trace) and D-FDLA (red trace). The traces represent EIC chromatograms at 412.189 Da. Peak retention time comparison reveals the DimeDap present in sandarazol to be D- configured as it shows inverse identity to the peaks in the L-DimeDap standards and enantiomers elute at the same retention time.

The stereo center on the chlorinated carbon

The chlorinated sandarazols have one additional stereo center at the chlorinated carbon in the molecule. Unfortunately, we were not able to determine its original stereochemistry as this carbon racemizes fast, even in the culture broth. Sandarazol C and D always show two distinct LC-MS peaks of a constant peak area ratio that represent the R- and the S-configuration at this carbon atom. After isolation of both of these peaks by LC we again get both peaks with the same peak area ratio meaning the isomerization happens fast at room temperature. Since the non-chlorinated sandarazols do not show this behavior, this splitting in the LC must occur at the stereo center at the chlorinated carbon that racemizes and creates diastereomers in that process.
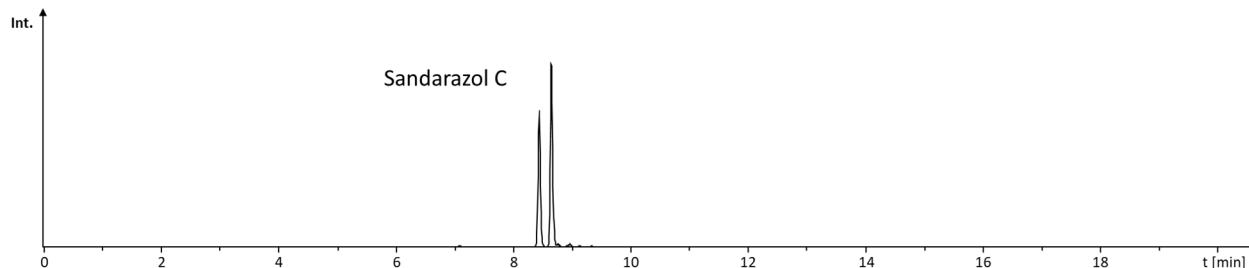

Figure S21 Chromatogram displaying the equilibrium distribution of sandarazol C under HPLC conditions, the chromatogram displays an extracted ion chromatogram at 583.325 Da to visualize sandarazol C with R- and S-config at the Cl bound carbon atom.

For that reason we concluded that natural sandarazol C and D are mixtures of both occurring epimers. The rapid racemization process did not allow us to assign the earlier or the latter peak to either the *S*- or *R*- form as racemization occurs faster than compound isolation. Epimerization was also observed during NMR analysis as represented in Figure S 22, reflected in a double set of signals for the neighboring ketone and methyl group. Similar signal intensities indicate an equilibrium between the two epimers. Carbon signals and proton signals could not be correlated due to insufficient resolution in the HMBC spectrum.
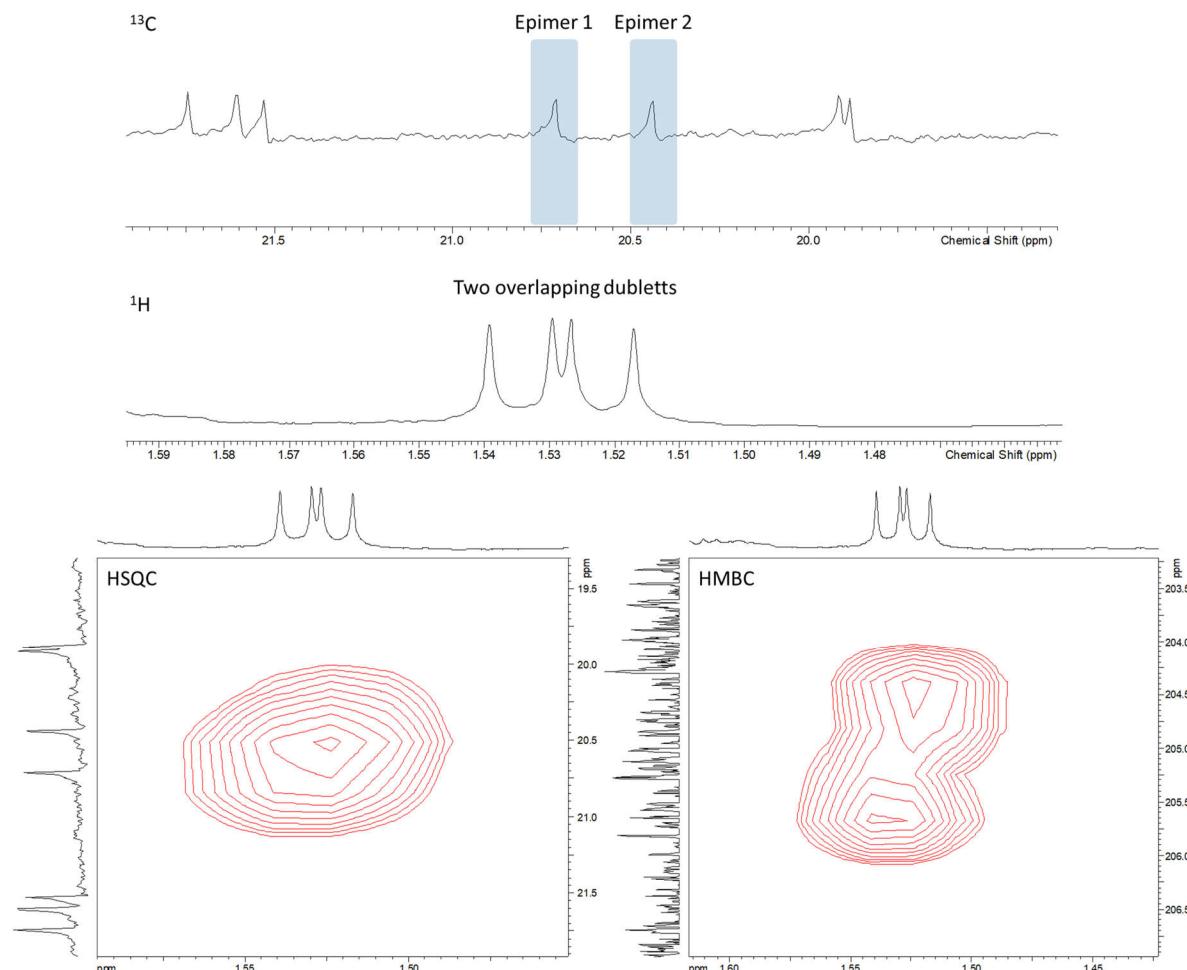

Figure S 22 1D and 2D NMR spectra of sandarazol C showing the epimerization of the chlorinated carbon. Upper spectra show zoom into the 1D spectra highlighting the signals of the neighboring methyl group. HSQC spectrum shows signal of the methyl group and HMBC spectra of the ketone.

Stereo chemical elucidation of the epoxide stereo center

Most sandarazols except for sandarazols E and F feature an epoxy ketone in their chemical structure. The ketone is introduced by an asymmetric FAD monooxygenase type epoxidation putatively catalyzed by the flavin containing protein SzoP. [10] As there are no stereochemistry prediction tools available to predict the stereo chemical outcome of such a reaction based on the protein sequence, the stereochemistry of the epoxide has to be elucidated by chemical methods. While there are methods available to assign the stereochemistry of a secondary alcohol to its respective stereochemistry, the tool for direct stereo chemical elucidation of epoxides is limited to X-ray crystal structure analysis. Still, as nucleophilic ring opening of epoxide structures is always *anti* based on its mechanism, an epoxide transfers its stereochemistry to the corresponding alcohol in this type of an epoxide ring opening

reaction. The resulting alcohol can then be structurally elucidated using Mosher's esterification method. [11] To use this, sandarazol A is transformed into methoxy-sandarazol A in a Lewis catalyzed epoxide ring opening reaction. The resulting methoxy-sandarazol A is purified by HPLC on a Phenomenex C18 biphenyl column under nitrogen gas using water + 10 mM Ammonium formate (AmFo) and methanol + 10 mM AmFo as eluents A and B. Separation is started with a plateau at 50% A for 2 minutes followed by a ramp to 32% A during 24 minutes and a ramp to 0% A during 1 minute. The A content is kept at 0% A for 2 minutes. The A content is ramped back to starting conditions during 30 seconds and the column is re equilibrated for 2 minutes. After evaporation, the methoxy-sandarazol A is obtained as pale yellowish amorphous solid.
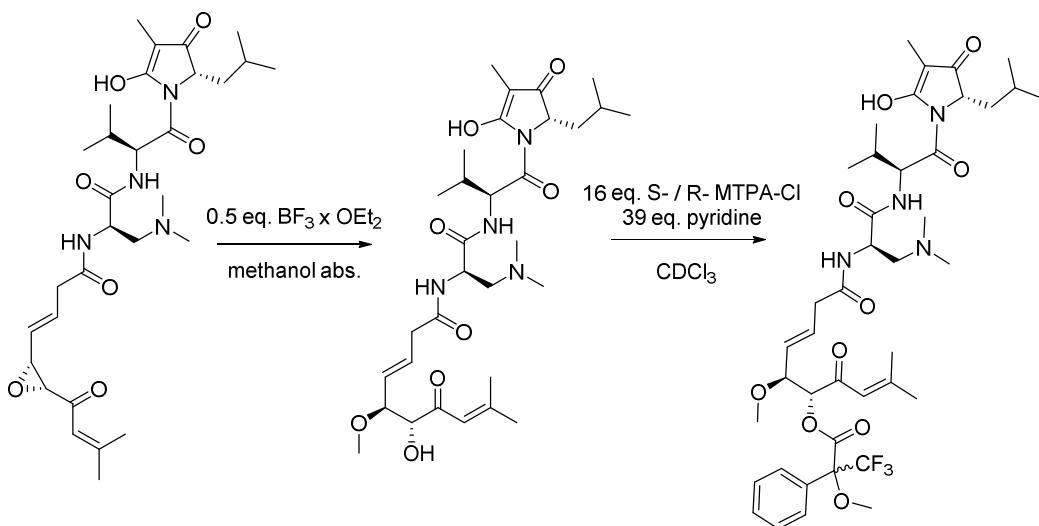

Figure S23 Epoxide ring opening reaction and Mosher's esterification for determination of the sandarazoles' epoxide stereochemistry

The resulting methoxy-sandarazol A features only one alcohol function that can be used in Mosher's method that inherited its stereochemistry from the epoxide. [11,12] Methoxy-sandarazol A is transferred into the corresponding S- and R- Mosher's ester by incubation with 16 eq. of the corresponding MTPA chloride and 39 eq. pyridine in chloroform. Configuration of the hydroxyl-group and therefore configuration of the epoxide is determined by the proton shifts of the S- and R- Moshers esters to be R.

Table S18 Chemical shifts of S- and R-ester of methoxy-sandarazol A and resulting assignment of R1 and R2.

Functional group	No.	δ S-ester [ppm]	δ R-ester [ppm]	Δ ppm	Hz (700 MHz)	
CH₃	1	2.1	2.14	-0.04	-28	R1
CH₃	2	1.85	1.96	-0.11	-77	-259
CH	3	6.09	6.31	-0.22	-154	
CH	4	5.43	5.39	0.04	28	R2
CH	5	4.13	4.05	0.08	56	+553
CH₃	6	3.32	3.26	0.06	42	
CH	7	5.55	5.46	0.09	63	
CH	8	5.89	5.82	0.07	49	
CH₂	9	3.11	3.04	0.07	49	
CH	10	5.01	5.01	0	0	
CH₂	11a	3.39	3.38	0.01	7	
CH₂	11b	3.55	3.55	0	0	
CH₃	12	2.95	2.96	-0.01	-7	
CH₃	13	2.95	2.96	-0.01	-7	
CH	14	5.64	5.64	0	0	
CH	15	2.19	2.19	0	0	
CH₃	16	1.01	1.01	0	0	
CH₃	17	0.91	0.9	0.01	7	
CH₃	18	1.69	1.77	-0.08	-56	
CH	19	4.93	4.94	-0.01	-7	
CH₂	20a	1.6	1.45	0.15	105	
CH₂	20b	1.87	1.86	0.01	7	
CH	21	1.45	1.32	0.13	91	
CH₃	22	0.74	0.65	0.09	63	
CH₃	23	0.71	0.58	0.13	91	

S6.8 Biological assay conditions

Human HCT-116 colon carcinoma cells (ACC-581) were received from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ) and were cultured under the conditions recommended by the depositor. To determine the cytotoxic activities of the sandarazols, cells from actively growing cultures were harvested and seeded at 5×10^4 cells per well in a 96 CELLBind® surface well plate in 120 µL 90% modified McCoy's 5A medium with 10% h.i. fetal bovine serum (FBS). After 2 h of equilibration, the cells were treated with the compounds in a serial dilution. After 5 days of incubation at 37 °C, 20 µL of 5 mg/mL thiazolyl blue tetrazolium bromide (MTT) in PBS was added. After discarding the medium, 100 µL of a 2-propanol 10 N HCl mixture (250:1) was added to dissolve formazan granules. A microplate reader (EL808, Bio-Tek Instruments Inc.) was used to determine the absorbance at 570 nm.

All microorganisms used for the biological assays were obtained from the German Collection of Microorganisms and Cell Cultures (Deutsche Sammlung für Mikroorganismen und Zellkulturen, DSMZ) or were part of our in-house strain collection and were cultured under the conditions recommended by the depositor. Bacterial cultures were prepared in MHB (2.9 g/L beef infusion solids, 17.5 g/L casein hydrolysate, 1.5 g/L starch at pH 7.4), M7H9 (0.5 g/L ammonium sulfate, 2.5 g/L disodium phosphate, 1.0 g/L monopotassium phosphate, 0.1 g/L sodium citrate, 0.05 g/L magnesium sulfate, 0.0005 g/L calcium chloride, 0.001 g/L zinc sulfate, 0.001 g/L copper sulfate, 0.04 g/L ferric ammonium citrate, 0.50 g/L *L*-glutamic acid, 0.001 g/L pyridoxine, 0.0005 g/L biotin at pH 6.6) or Myc 2.0 medium

inoculated from the strain grown on agar plate. The compounds were diluted serially in sterile 96 well-plates before adding the bacterial cell suspension. The bacteria were grown for 24 h at RT, 30 °C or 37 °C. Growth inhibition was inspected visually. MIC50 values were determined relative to the respective control samples by sigmoidal curve fitting. Positive controls used for the respective microbial test strains are listed in table S 19.

Table S 19 Microbial test strains and positive controls used for MIC determination.

Microbial strain	Control
C. albicans	Amphotericin B
P. anomala	Amphotericin B
C. freundii	Ciprofloxacin-HCl
A. baumanii	Ciprofloxacin-HCl
S. aureus	Vancomycin
B. subtilis	Vancomycin
E. coli	Ciprofloxacin-HCl
P. aeruginosa	Ciprofloxacin-HCl
M. smegmatis	Rifampicin

S6.9 Availability of data on a preprint server

An earlier version of this manuscript has been uploaded to the preprint server BioRxIV and this preliminary version of the manuscript, the supporting information and all raw NMR data can be accessed there. [13]

References

(1) 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.

(2) 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.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837.

(3) O'Leary, N. A.; Wright, M. W.; Brister, J. R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.*; et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**, *44*, D733-45.

(4) Pogorevc, D.; Panter, F.; Schillinger, C.; Jansen, R.; Wenzel, S. C.; Müller, R. Production optimization and biosynthesis revision of corallopyronin A, a potent anti-filarial antibiotic. *Metab. Eng.* **2019**, *55*, 201–211.

(5) Chang, Z.; Sitachitta, N.; Rossi, J. V.; Roberts, M. A.; Flatt, P.; Jia, J.; Sherman, D. H.; Gerwick, W. H. Biosynthetic pathway and gene cluster analysis of Curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod.* **2004**, *67*, 1356–1367.

(6) Panter, F.; Krug, D.; Baumann, S.; Müller, R. Self-resistance guided genome mining uncovers new topoisomerase inhibitors from myxobacteria. *Chem. Sci.* **2018**, *9*, 4898–4908.

(7) Linington, R. G. npatlas - The Natural Products Atlas. https://www.npatlas.org/.

(8) van Santen, J. A.; Jacob, G.; Singh, A. L.; Aniebok, V.; Balunas, M. J.; Bunsko, D.; Neto, F. C.; Castaño-Espriu, L.; Chang, C.; Clark, T. N.*; et al.* The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery. *ACS Cent. Sci.* [Online early access]. DOI: 10.1021/acscentsci.9b00806.

(9) Marfey's reagent: Past, present, and future uses of 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide.

(10) Morrison, E.; Kantz, A.; Gassner, G. T.; Sazinsky, M. H. Structure and mechanism of styrene monooxygenase reductase: new insight into the FAD-transfer reaction. *Biochemistry* **2013**, *52*, 6063–6075.

(11) Sullivan, G. R.; Dale, J. A.; Mosher, H. S. Correlation of configuration and fluorine-19 chemical shifts of .alpha.-methoxy-.alpha.-trifluoromethylphenyl acetate derivatives. *J. Org. Chem.* **1973**, *38*, 2143–2147.

(12) Hoye, T. R.; Jeffrey, C. S.; Shao, F. Mosher ester analysis for the determination of absolute configuration of stereogenic (chiral) carbinol carbons. *Nat. Protoc.* **2007**, *2*, 2451–2458.

(13) Panter, F.; Bader, C. D.; Müller, R. *The sandarazols are cryptic and structurally unique plasmid encoded toxins from a rare myxobacterium*; BioRxIV, 2020.

<div align="center">

Chapter 7

</div>

Discussion

A classical myxobacterial natural product discovery workflow as depicted in Figure 1 consists of six main steps: collection of soil samples (1.), isolation of specific myxobacterial strains from those samples, their cultivation and/or genome sequencing (2.), prioritization of one or a group of NPs for isolation (3.), extraction of the respective NP (4.), its purification (5.) and its characterization on structural level as well as determination of its pharmaceutical properties (6.). In this thesis, nine different myxobacterial strains were investigated on analytical scale and two strains were prioritized for subsequent isolation and characterization of four NP families: the chloroxanthic acids, the sandacrabins, the sandaramines and the sandarazols. All of those four families were accessed by applying variant methods on different stages of the NP discovery workflow with emphasis on underrepresented techniques. The chloroxanthic acids were found while evaluating different extraction procedures for microbial extract generation. The sandacrabins were discovered in a metabolome-guided approach, whereas the sandaramines were obtained in a bioactivity-guided approach. The sandarazols finally are representatives of the genome-guided discovery process. The following chapters summarize mayor obstacles tackled during the discovery of these four new NP families and discuss the suitability of the methods applied for other NP families, as well as further possible approaches applicable at the respective step.
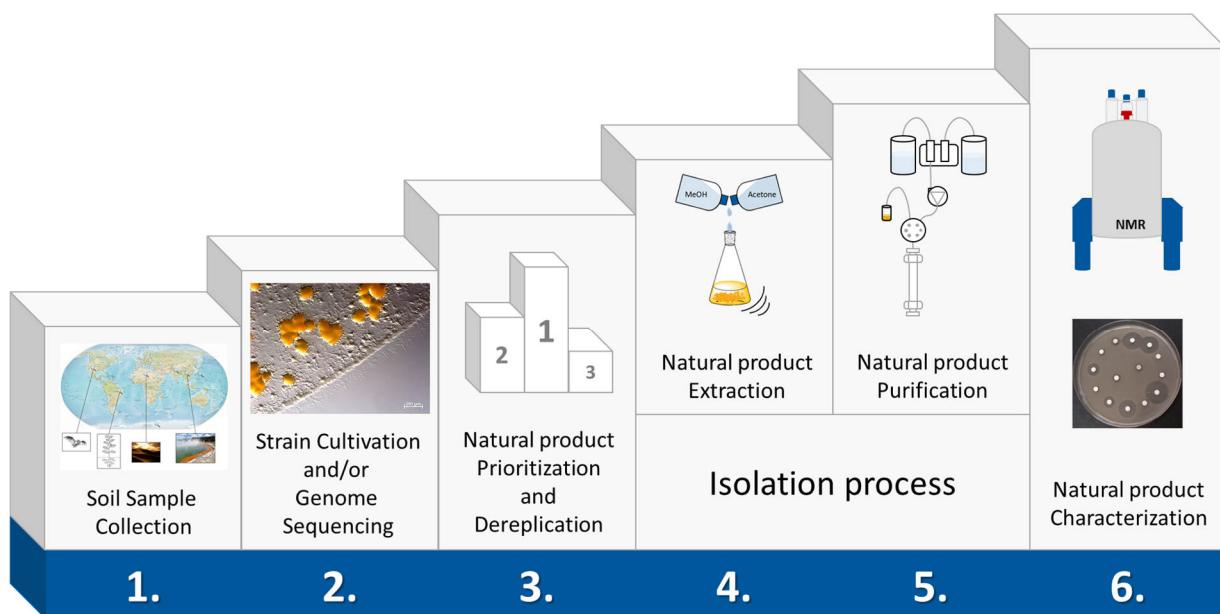

Figure 1 Six steps of a classical myxobacterial natural product discovery workflow.

7.1 Accessing, sequencing and cultivating myxobacterial strains

Three out of four NP families described in this thesis were discovered by in-depth analysis of *Sandaracinus* strain MSr10575 (see Chapter 2-6), isolated from a soil sample in 2013. It has not been described beforehand and was only found to produce few known NPs.[1] This finding already highlights one of the main advantages of the isolation of new myxobacterial strains for the discovery of novel secondary metabolites, especially if they belong to underexplored genera: most of their metabolome is not known yet. Even under standard cultivation conditions, they often produce a variety of novel secondary metabolites. Members of underrepresented genera are furthermore well-suited for bioactivity-guided isolation approaches, as their chemical distance to well-described genera reduces rediscovery of already known biologically active metabolites with this approach. But even well-described myxobacterial strains still harbor potential for the discovery of novel NPs, as demonstrated by the discovery of the chloroxanthic acids from *Myxococcus xanthus* DK1622 (see Chapter 2). Isolation from such well-studied strains however may require more sophisticated methods, as secondary metabolites accessible with traditional methods are often already described.

For well-described strains as *M. xanthus* DK1622, genome sequencing data often is already available, whereas for new strains as *Sandaracinus* MSr10575 this information first has to be obtained. Previously investigated strains therefore ease the start when attempting to analyze the strain in genome-guided approaches. For new isolates one main challenge is obtaining a good coverage during genome sequencing to receive a closed genome sequence necessary for reliable BGC prediction. As genome sequencing is hard to achieve in one piece, it is divided into several smaller fragments with overlapping regions, which are assembled afterwards into bigger contigs or a full genome.[2] If the DNA was not appropriately sequenced in these overlapping regions, it cannot be assembled correctly. If this area is a non-coding region, an interruption of the sequence may remain without consequences for a NP discovery workflow. In case a BGC is encoded in this region, it may however no longer be detectable by BGC prediction tools such as AntiSMASH.[3] In this study a full genome sequence for the new isolate *Sandaracinus* MSr10575 was obtained, which was later on used to assign the BGC of the sandarazols (chapter 6) and sandaramines (chapter 5) to the respective NPs. Furthermore, the *Sandaracinus* MSr10575 full genome sequence was used to develop a concise biosynthesis hypothesis for those two NPs, as well as for a third new NP, the sandacrabins (chapter 4). Obtaining a full genome sequence is therefore crucial to estimate the biosynthetic potential of a new myxobacterial strain, investigate the biosynthetic origin of novel NPs and explore gene clusters not expressed under standard laboratory conditions.

While genome-driven NP discovery approaches can also be performed for non-isolated bacterial strains in metagenomics analysis,[4] metabolome-based approaches require separation of single strains from their cohabitants. The respective strains are commonly first cultivated on plates for their isolation and subsequently cultivated in shake flasks using a liquid cultivation medium. While this cultivation system supports the following isolation-for-structure-elucidation workflow and can be upscaled to relatively high culture volumes, it fails to mimic the natural habitat of soil living bacteria and therefore often leads to suppression of secondary metabolite production.[5] Chapter 2 highlights how much a strains metabolome already changes when only supplementing liquid by plate cultivation. In contrary to the commonly used liquid cultivation, myxobacteria may develop their natural morphology including swarming and the formation of fruiting bodies on agar plates, which is suppressed in liquid medium. This has a major impact on their metabolome, as shown by Hoffmann *et al.*.[6] This influence is also observed for the sandaramines, which were first only detected when growing the strain on agar

plates. Initial production in liquid cultures was relatively low for this NP family. Subsequent variations in the composition of the liquid medium however also enabled production in large-scale liquid cultures.

Different growth morphology in liquid medium and on agar plates is not the only factor differing under laboratory conditions from natural soil conditions (see Figure 2). Dependent on the soil layer the bacteria are living in, they may be exposed to more or less light and oxygen. For secondary metabolite production, light sources may therefore be changed to limited wavelengths or cultivation may also be performed completely in the dark. Oxygen supply can be adapted by varying the culture in relation to the flask volume or adjusting the shaking speed. Dependent on the location the strain is living in, it also may require different temperatures optimal for secondary metabolite production. Finally yet importantly, in their natural habitat myxobacterial strains do not live as monocultures, but are exposed to a variety of competing residents.[7] Rational strategies to enhance secondary metabolite production therefore are co-cultivation systems.[8] Our study in chapter 2 highlights the tremendous effects of changing the cultivation system and emphasizes how important it is to investigate the other methods described above to mimic the natural myxobacterial habitat. The strategies and parameters summarized in figure 2 therefore represent further approaches which might be used in the future to access even more NPs from *Sandaracinus* MSr10575 and *M. xanthus* DK1622 than the four NP families described in this thesis.
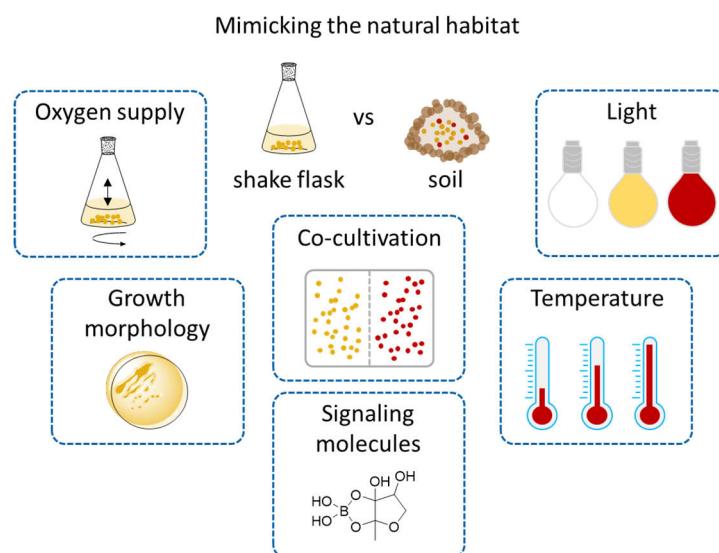

Figure 2 Strategies and parameters that may be adapted to mimic the natural myxobacterial habitat in soil to enhance production of novel secondary metabolites.

7.2 Prioritization and detection of novel secondary metabolites

After successful cultivation of the myxobacterial strain of interest, NPs have to be selected and prioritized for isolation. This on one hand can be done by using genome information, as carried out for the sandarazols, or on the other hand by exploring the strains metabolome, as performed for the sandaramines, sandacrabins and chloroxanthic acid A. The prioritization step starts with exclusion of known secondary metabolites and their family members (dereplication). The dereplication process should be performed as early as possible in the NP discovery workflow and requires access to preferably comprehensive libraries of known NPs.[9] Even though there is some progress in making such information publicly available as exemplified by the npatlas[10], many databases are still hidden behind paywalls, which highly obstruct this step in NP discovery workflows. In case of *Sandaracinus* MSr10575 few known NPs were detectable in its extracts (see chapter 3 and 4). In comparison to such new strains

with few annotations, isolation of novel NPs from extensively investigated strains such as *M. xanthus* DK1622 bears significant challenges.[11] As described in chapter 7.1, they often produce a variety already described secondary metabolites. Those metabolites often are relatively easy to detect, but interfere with the detection of the remaining not yet explored NPs, which often are only produced in small amounts, are highly unstable or do require non-standard detection or extraction methods. With the isolation of chloroxanthic acid A from this strain however, deviating from standard settings (in this case the extraction method) has shown that these strains still bear great potential for isolation of new secondary metabolites. Figure 3 depicts known secondary metabolites produced by *M. xanthus* DK1622, which needed to be excluded during dereplication highlighting those isolated in recent years.
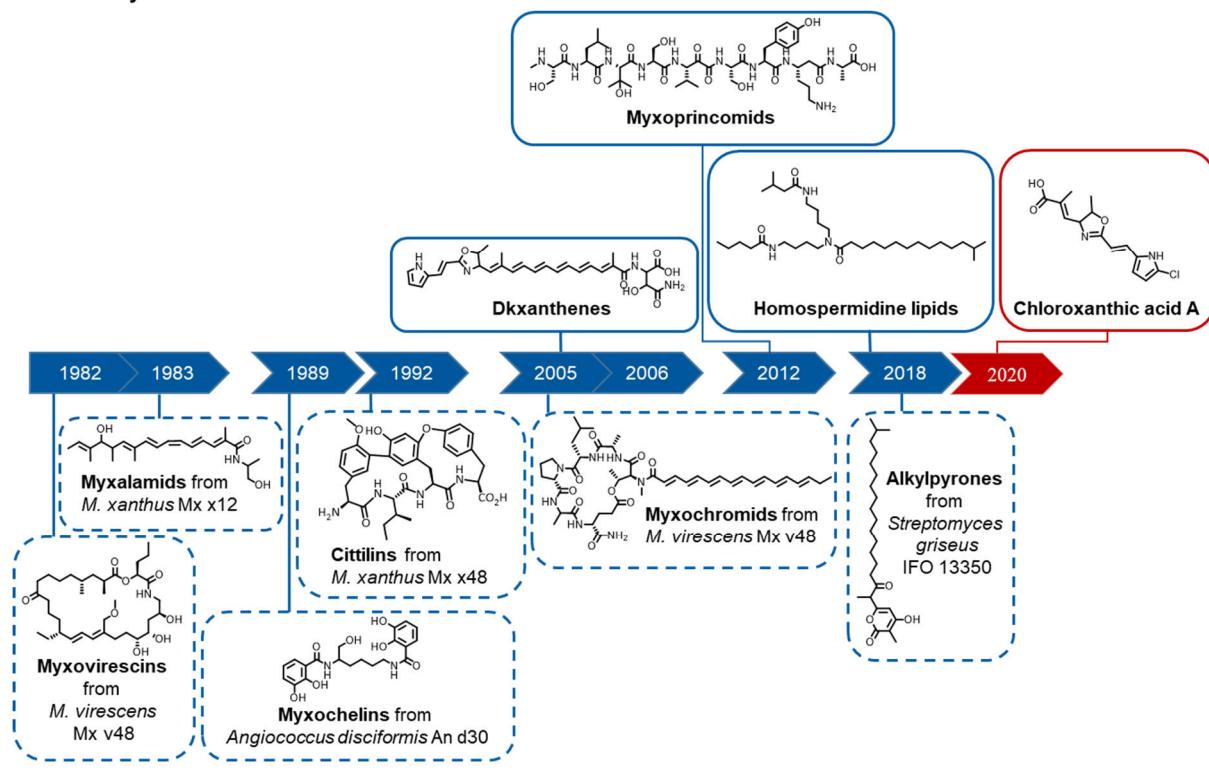
Figure 3 Secondary metabolites produced by (blue dashed) and isolated from (blue and red) *M. xanthus* DK1622 adapted from Bader *et al.*.[11] Chloroxanthic acid (red) added to the elucidated structures in this thesis.

The success of the dereplication process, as well as the detection of novel secondary metabolites is highly dependent on the analytical method applied to the bacterium's crude extracts. One detection method deviating from standard setups is DI-FTICR (see chapter 2). Its low overlap to the commonly used LC-TOF setup for metabolome analysis highlights the immense chemical space still waiting for exploration. Even though DI-FTICR has proven its impact on analytical scale for the discovery of new NPs, isolation based on this detection method is still challenging, as most purification methods also rely on rp-HPLC. If metabolites are detectable by DI-FTICR but not LC-TOF, one reason may be too strong retention on the chromatographic column. When the subsequent purification process also relies on similar retention principles, these metabolites will stay non-accessible. Therefore, it is not only crucial to diversify the analytical detection methods, but also the purification methods.

7.3 Supercritical fluid techniques as keys to novel metabolites

The isolation process can be subdivided into two main parts: the extraction from cell material, as well as adsorber resin if applicable and the purification process mainly using chromatographic separation techniques. Compared to the high diversity of chromatographic principles available, diversity is low in extraction protocols from bacterial raw material. Most variations just are performed by adapting the solvent used to the estimated polarity of the NPs of interest. For myxobacteria in particular, stirring the raw material with MeOH and acetone remains the standard method since accessing this group of bacteria for secondary metabolite purification. The extraction process however, is one of the crucial steps in sample preparation, especially when it comes to a large-scale isolation process and should therefore carefully be evaluated.[12] An efficient extraction protocol is key for getting enough pure compound, even when multiple rounds of chromatographic purification are required subsequently.

Contrary to extraction from bacterial material, in plant secondary metabolomics the diversity of established methods for their extraction is much more diverse.[13] Supercritical fluid extraction herein has already shown to be superior to conventional solvent extraction.[14] In chapter 2, a possible transfer of this finding to myxobacterial extractions was evaluated, which was found to be true. Most known myxobacterial secondary metabolites showed superior extraction with SFE in comparison to the conventionally used UAE. The new secondary metabolite chloroxanthic acid A was even only accessible due to its better extraction efficiency with SFE. The sandacrabins and sandaramines were also found to be extractable with SFE (see Figure 4). In their cases however, SFE was not better suitable for extraction. Interestingly, sandacrabin A did not show any enrichment using a specific solvent for extraction and is one of the rare cases, showing good extractability with pure CO_2. Sandacrabin B and C, which are structurally very similar and which both bear a permanent positive charge, showed highly similar extraction profiles. They were both best extractable with SFE using 50% MeOH as co-solvent. For the sandaramines, MeOH was the optimal co-solvent too. For the sandaramine derivative with the permanent positive charge (sandaramine B) 50% MeOH was the optimal co-solvent among the different co-solvents evaluated in this study. As the biosynthetic machinery responsible for sandarazol production was found to be inactive under laboratory conditions, sandarazoles were not detected in the SF extracts.
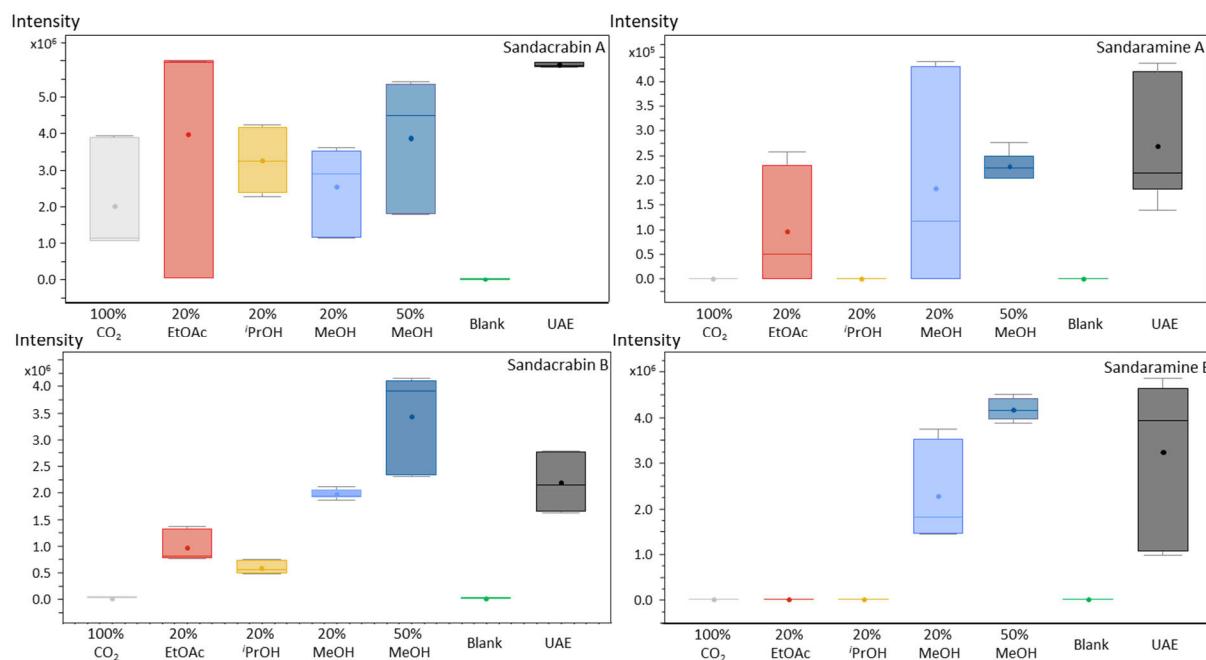
Figure 4 Box plots of SF extraction in comparison to conventional UA extraction for sandaramine and sandacrabin A and B.

The sandacrabins, the sandaramines and the sandarazoles (after activation of their respective biosynthetic machinery) were produced in good quantities by *Sandaracinus* MSr10575 and showed good extraction efficiency with the conventional myxobacterial extraction protocol, wherefore their extraction was performed with conventional solvent extraction using MeOH and acetone. Contrary to the chloroxanthic acids, where extraction efficiency has been the limiting factor for structure elucidation beforehand, these three new NPs presented themselves as challenging during the subsequent purification steps. Due to the oxygen sensitivity of the sandarazols, they had to be purified under a constant N_2 flow to avoid degradation. Furthermore, with over 20 family members, they belong to a very rich secondary metabolite family. Separation of these highly similar NPs from each other and from the sandacrabins and sandaramines, which introduced a second challenge to the separation process by their permanent positive charge, could therefore only be achieved by combining different chromatographic techniques (see Figure 5).

The first chromatographic purification step of *Sandaracinus* MSr10575 crude extracts was performed with centrifugal partition chromatography (CPC). CPC does not belong to the extensively used chromatographic techniques in NP research,[15] but its underlying principles often are applied when partitioning an extract between different solvents, which is a step commonly implemented before chromatographic separation. During CPC fractionation of *Sandaracinus* MSr10575 extracts, the sandacrabins, sandaramines and sandarazoles could be separated from each other. Furthermore, CPC allowed direct separation of sandacrabin A from its two derivatives bearing a permanent positive charge (sandacrabin B and C). CPC is one of the separation technologies showing good performance for permanently charged molecules as exemplified for berberine, as they exhibit a high affinity to the aqueous phase.[16] In case of the sandaramines and sandacrabins alkalizing the solvent system used in CPC deprotonated the derivatives without permanent charge, wherefore they passed into the organic phase and were nicely separable from those bearing a quaternary ammonium function.
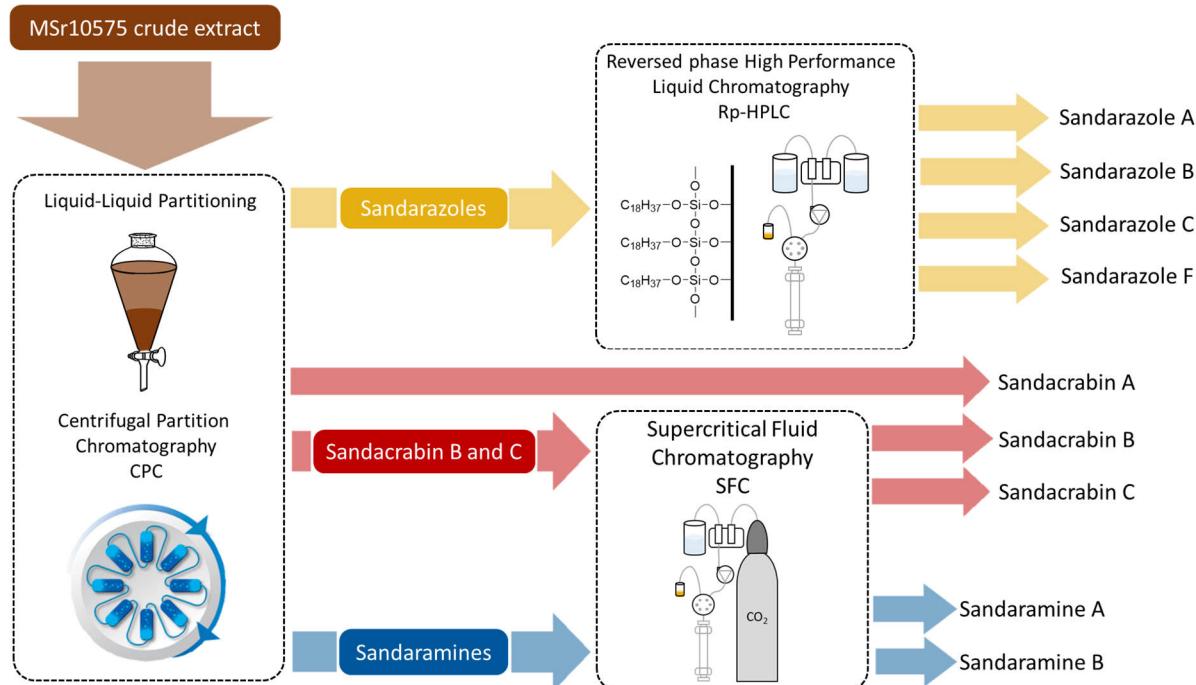
Figure 5 Chromatographic techniques used for separation of the three natural product families under investigation in this thesis produced by *Sandaracinus* MSr10575.

For the sandarazoles the second and final purification step, which allowed separation of the different family members as well as partition from remaining impurities from the crude extract, was performed using rp-HPLC. In comparison to the sandarazoles, the sandaramines and sandacrabins are relatively non-polar. This may be one of the reasons, why purification with rp-HPLC turned out challenging for them. Even after extensive method optimization, separation of the structurally highly similar sandacrabin B and C could not be achieved with rp-HPLC. The sandaramines exhibited extensive tailing on rp columns, which hindered upscaling during the isolation process. Both of these challenges could be overcome by switching the chromatographic separation method to SFC (see chapter 4 and 5). Although SFC has already been used for the purification of plant NPs, similar to SFE it remains an underrepresented technique when it comes to the separation of bacterial NPs. The underlying separation mechanisms in SFC are orthogonal to rp-HPLC and do not resemble the ones in normal phase HPLC, wherefore it can nicely be combined with those two chromatographic principles as well as CPC to achieve separation of complex mixtures such as myxobacterial crude extracts. Development of an SFC purification protocol for the sandacrabins allowed separation of the two derivatives only differing in one methyl group. Furthermore, good peak shapes for the sandaramines were achieved with SFC, which supported up-scaling of the isolation process to obtain sufficient amount for the subsequent structure elucidation and biological characterization. To the best of our knowledge, SFC has not been used for the isolation of NPs featuring a quaternary ammonium function to date. The successful isolation of the sandaramines and the sandacrabins with SFC however emphasizes to extend its use for NP purification, especially when it comes to charged molecules with high lipophilicity.

7.4 NMR structure elucidation supported by MS[2]

While after chromatographic separation several different NPs or family members may be collected, *de novo* structure elucidation by NMR remains a one-compound-at-a-time process. It is therefore one of the bottle necks when trying to speed up NP discovery workflows.[17] For *de novo* structure elucidation of complete NP families it is helpful to acquire MS[2] data and use these data sets for structure elucidation, once one family member has been elucidated by NMR. Additional confirmation of the structure hypothesis based on MS[2] data can then be acquired by investigation of the NPs biosynthesis. After finding the respective BGC responsible for the biosynthesis of one NP family, modifications possible by additional genes or promiscuity of some modules can be narrowed down and compared to the respective MS[2] spectra.[18] Furthermore, NP family members can be assigned to one family by comparing similarities in their MS[2] spectra with global natural products social molecular networking (GNPS). This step is additionally useful for dereplication, as GNPS does not only allow to compare generated MS[2] spectra, but supports library matching.[19] Family members of already described NPs can therefore be deprioritized, while derivatives of new families can be prioritized for isolation based on structural differences and production yield. MS[2] spectra however, will not replace NMR for structure elucidation, as only for proteinogenic amino acids fragmentation rules are completely understood, while most of the gas phase reactions are not completely predictable.[20] At least one member of each new secondary metabolite family described in this thesis was structurally elucidated using NMR. Both sandaramines and sandacrabins only consist of relatively few family members, and most of them are produced in decent amount. The different derivatives where therefore all isolated and submitted to NMR for structural confirmation. In case of the sandacrabins, NMR data alone were not elusive and their MS[2] fragmentation pattern was used to essentially underpin one structural proposal. The sandarazoles are one example where isolation and structure elucidation by NMR is not be feasible for all derivatives due to their oxygen sensitivity. Furthermore, they consist of a great number of derivatives and production levels between the different family members are highly different. By using characteristic MS[2] fragmentation patterns and their biosynthesis hypothesis, the structure of seven family members could be determined instead of just the structures of the four main derivatives sandarazol A, B, C and F which were subjected to NMR. This study once more highlights, that combining the two technologies MS and NMR with knowledge about the biosynthesis of the secondary metabolites is a very powerful approach to investigate a complete NP family.

Structure elucidation is the main step of the whole characterization step, as NPs are often published without further investigation of their biosynthesis or biological activities but very rarely NPs are described by correlating BGCs to biological activities without describing the chemical structure of the respective NPs. One of those rare examples where scientist could correlate the biosynthetic origin of a NP with its effect without solving the NPs structure is colibactin.[21] Colonialization of *E. coli* bacteria bearing the colibactin BGC could be associated with the development of colorectal cancer, even though the molecule itself was not isolated and structurally elucidated at that time.[22,23] In case of colibactin its aero sensitivity hindered structure elucidation for a long period of time, an instability also observed for the sandarazoles.[24] However, for the development of a drug on the market from a NP, its structure has to be elucidated to perform medicinal chemistry optimization which is required for most of the bacterial secondary metabolites. Obstacles such as oxygen sensitivity may be overcome with this optimization by bioisoteric replacement of the respective functional groups for example.

7.5 An appeal for early-stage natural product synthesis

De novo structure elucidation of the non-derivatized NPs by NMR only provides information about the relative stereochemical configuration of a molecule. For investigation of the absolute configuration of a molecule however, at least one stereo center has to be completely characterized. For peptidic NPs, such as the sandarazoles this is possible using a Marfeys derivatization reaction, where implementation of a second stereo center in the product allows separation and comparison of retention times of the two possible diastereomers with rp-HPLC. For secondary metabolites not incorporating amino acids, or other structural parts with the respective standards in both configurations available, Moshers esterification provides valuable information. This reaction was also performed to elucidate the non-peptide stereocenters in the sandarazoles. However, this derivatization reaction is only possible when free functional groups, such as hydroxyl functions are available for the reaction. For NPs without such groups (exemplarily the sandaramines), the only method remaining to elucidate the absolute stereochemistry of the NP is total synthesis of the possible stereoisomers and subsequent comparison with the NP itself. Information about the biosynthetic origin of a molecule can simplify the synthesis, as it may already point towards a more likely configuration. This configuration can then be prioritized for synthesis instead of accessing all possible stereoisomers. Bioinformatic tools, estimating the configuration of a stereo center based on the biosynthetic enzymes involved in its formation, mainly use prediction of functionality of specific domains. For NRPS prediction, they consider additionally encoded epimerase domains, which convert e.g. the native L-amino acids into their D isomer.[25] For PKS, hidden Markov models can be used to predict the configuration of secondary alcohols or α-methyl branches.[26] Another hint towards one configuration of the NP under investigation is comparison with already described NPs. Configuration of the halichonines, showing structural similarity to the sandaramines, already pointed towards a similar configuration, which was confirmed by total synthesis of sandaramine A later on (Figure 6).

Total synthesis, especially when performed at an early stage of the NP discovery workflow cannot only be used for assigning the configuration of a secondary metabolite. It additionally reveals possible mistakes during the structure elucidation process, which are unfortunate but rather common as summarized in a review by Chhetri *et al..* [27] Even though they highlight that taking precautions such as questioning unusual structural features and comparing experimental to predicted spectra, *de novo* structure elucidation still remains an error-prone process.[27] One example for a structural revision of a quaternary ammonium salt is ageloxime D, a strong cytotoxic NP isolated from a marine sponge (see Figure 6).[28] Besides the wrong assignment of its terpenoid skeleton configuration, it was first thought to incorporate a purine unit. Total synthesis of ageloxime D and comparison of the synthetic NMR spectra and the natural ones revealed, that ageloxime incorporates a aminopyrimidine core structure instead.[28] Those structures are hard to distinguish by NMR spectroscopy, as they consist of neighboring quaternary carbons, which are only limited accessible by 2D NMR correlation spectroscopy. In case of ageloxime D, the many nitrogen atoms participating in the ring system furthermore interrupt long range correlations visible in the respective HMBC spectra. It is therefore understandable that the structure was not elucidated correctly at first. In case of sandacrabin B and C similar obstacles were observed. The NMR data alone allowed two possible structures as represented in Figure 6.

In case of the sandaracrabins, the symmetry axis dividing the benzimidazole core structure besides delocalization of the positive charge complicated the detectability of the proton at C_1, which first let to the conclusion that this position has to be substituted. Integrals of the first CH_2 group substituting the benzimidazole nitrogen atoms and exact same chemical shifts of the farnesyl side chain however pointed towards a symmetry of the molecule. Total synthesis of the symmetric structure proposal and comparison with the NP was therefore used to confirm the structural formula of sandacrabin B.
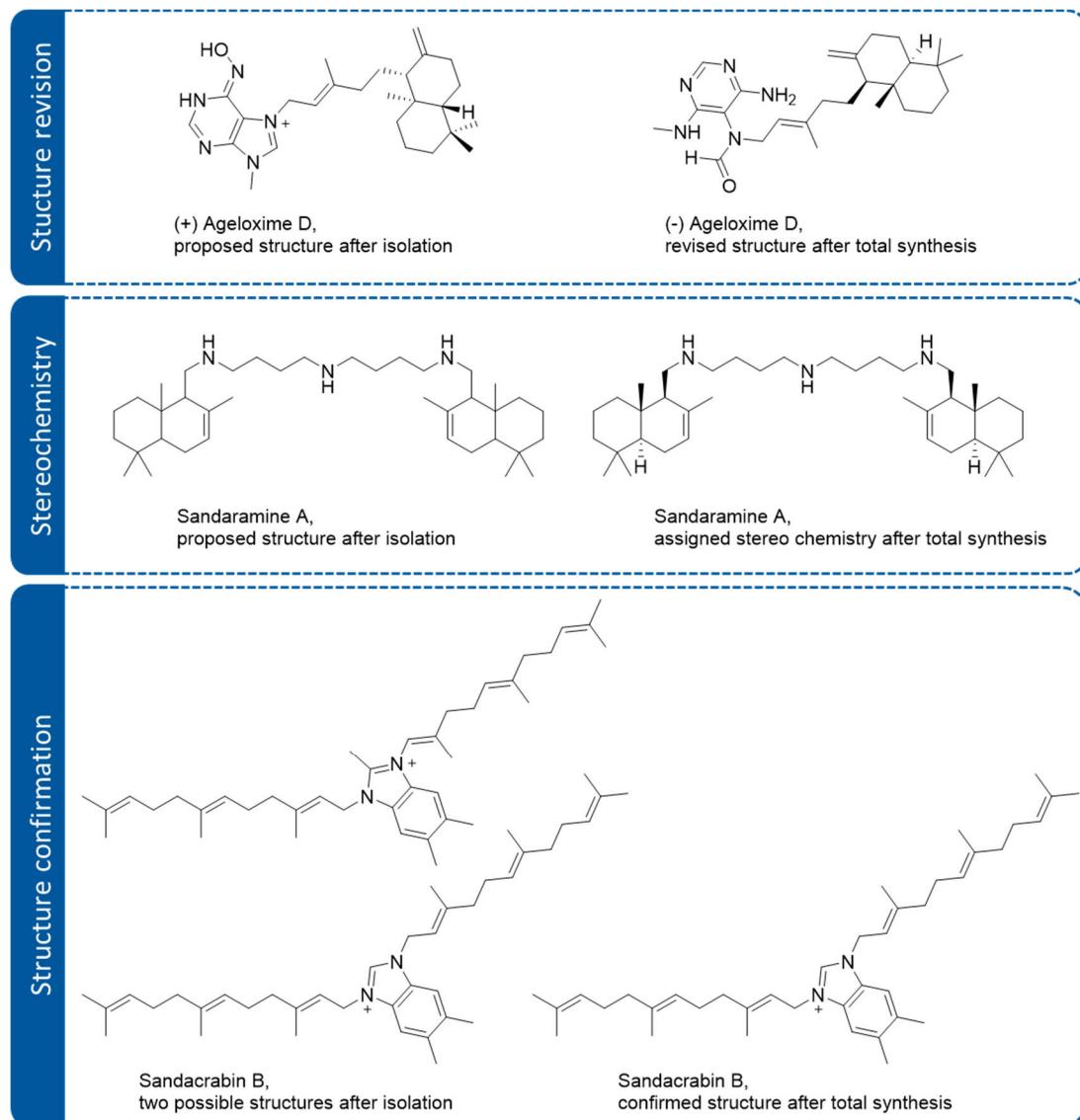

Figure 6 Structure of ageloxime D reported after isolation and revised structure after total synthesis (top)[28] In comparison to ageloxime D, direct total synthesis after isolation allowed elucidation of the sandaramine stereochemistry (middle) and discernibility between two structure proposals for sandacrabin B (bottom) before publication.

Early-stage NP synthesis has proven a useful strategy for both structure confirmation, as well as for elucidation of a molecules stereochemical configuration in this thesis. Additionally, observed biological activities of the purified NPs could be confirmed. This is of special importance when it comes to bacterial NPs, as impurities from the complex bacterium's crude extract can lead to false positive hits. Even though synthesis products may also contain impurities causing biological activities, it is highly unlikely that they are the same as contained in the NP purified from the bacterial extract. This evaluation therefore also helps to increase the reliability of biological characterizations, not only of the structure elucidation.

7.6 Translating the natural product function into human use

The biological characterization of a pure NP commonly is the last step of a NP discovery workflow, as the assays performed within are destructive. While marking some end-point in the discovery process, this step also determines the future of a NP, from an academic as well as industrial view. If very potent activities are observed, there is interest in medicinal chemistry optimization and development into a pharmaceutical drug or optimization towards an insecticide that can be used for crop protection. In case no activities could be determined in the first screens, the question remains why the bacterium produces this specific NP, which needs to be clarified in follow-up studies. They for example may serve as signaling molecules and help to understand morphological development or also induce the production of other secondary metabolites when added to a culture of a different bacterium.[29,30] As biological functions for the bacterium are diverse, the biological characterization should be as comprehensive as possible. Those screens however are laborious and good reproducibility needs to be guaranteed, wherefore most groups focus on a subset of test organisms. If no activities against this subset were observed, the respective NPs are sent to cooperation partners to be involved in their screening campaigns or stored in NP libraries to be tested later on.[31]

Most defensive purposes for the bacterium displayed as antimicrobial, antiproliferative or insecticidal activities are directly translatable into a possible human use (see Figure 7).[32,33] Insecticidal activities of NOs such as the sandacrabins for example are a good starting point for the development of insecticides used in agricultures.[34] Other eukaryotes living in soil the bacterium may target with NPs toxic to eucaryotic cells, are protozoa or nematodes. They both are however not only soil inhabitants but also pathogenic organisms.[35] Examples for pathogenic protozoa are the Malaria casing *Plasmodium falciparum* or the Leishmaniosis causing *Leishmania major*.[36,37] Protozoan diseases and helminthiasis belong to the neglected diseases, wherefore they are underrepresented in most screening campaigns.[38] Commonly, cancer cells are used instead to detect anti-eucaryotic effects, as such effects may hint towards a suitability of the NP as anticancer drug.[39] However, it is obvious that non-symbiotic bacteria produce the NP rather as a defense against worms or protozoa instead of defending itself against cancer cells or higher developed organisms. From the four NP families accessed in this thesis, anti-eucaryotic effects where observed for the sandacrabins, sandarazoles and sandaramines.

Besides their cytotoxic activities, antibacterial activities were observed for the sandacrabins, sandarazoles and sandaramines. Chances to observe such activities of bacterial NPs, which may serve as starting point towards development as an antibiotic, are relatively high, as bacteria have to defend themselves against a variety of bacterial competitors in their natural environment. *M. smegmatis* exemplarily, for which we find the highest antibacterial activities of the substances investigated in this thesis, belongs to the fast growing mycobacteria found in soil.[40] From a myxobacterial point of view, it is therefore reasonable to produce antibiotics targeting this organism. It belongs to the family of mycobacteria, which involves the tuberculosis causing *Mycobacterium tuberculosis*, wherefore chances are increased for substances active against *M. smegmatis* to address a target also available in *M. tuberculosis.*[40]
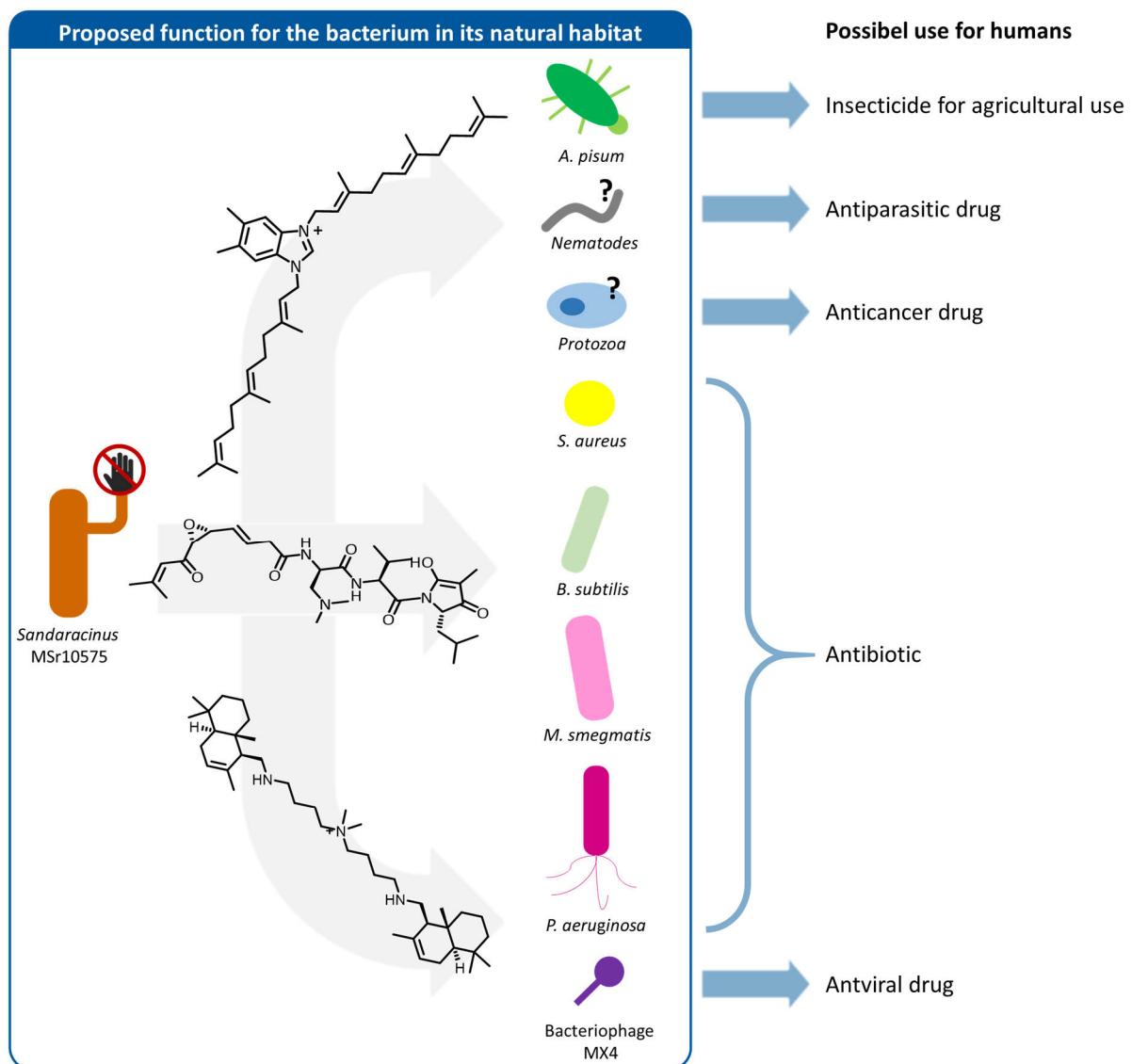
Figure 7 Proposed function of the secondary metabolites isolated and characterized in this thesis and transfer possibilities to human use.

Another activity detected for the sandaramines and sandacrabins is their potential to inhibit viral replication. Although human viruses such as SARS-CoV2 are not capable of infecting bacterial cells, bacteria have to defend themselves against bacterial viruses, the so-called bacteriophages.[41] One example of a myxobacterial phage against which myxobacteria have to defend themselves, is MX4.[42] It is likely, that when observing antiviral activities of NPs, they evolutionary were developed as defense mechanism against bacteriophages. With this underlying rational, antiviral screens of NPs for their development towards an antiviral drug, which are becoming more and more urgent with the increased incidence of viral pandemics, are very reasonable.[43]

In contrary to the demand of the pharmaceutical industry for highly selective drugs, soil living bacteria often produce defensive NPs with broad-spectrum activities as exemplified for the diverse biological activities observed for the sandaramines and sandacrabins comprising antiviral, insecticidal, cytotoxic and antibiotic effects. NPs featuring a quaternary ammonium function are already known to exhibit such broad-spectrum activities (see Figure 8).[44] The plant-derived NP berberine for example, shows antiviral activities besides broad band antibiotic activity and cytotoxic effects against HCT-8 cells.[16,45] However, there are only few NPs with a quaternary ammonium function described to date.[44] Examples

for additional plant-derived NPs are the antiplasmodial magnocurarine and the antibiotic latifolian A, which was also found to induce apoptosis in eukaryotic cells.[46] Furthermore, there is one quaternary ammonium NP, quinocitrinine, with antibiotic and antitumor activities.[47] With this thesis the class of quaternary ammonium bearing NPs was extended by two more bacterial secondary metabolite families, the sandaramines and sandacrabins. Their structure elucidation and biological characterization is exceptionally interesting, as among the quaternary ammonium featuring NPs, bacterial ones are rare.[44]
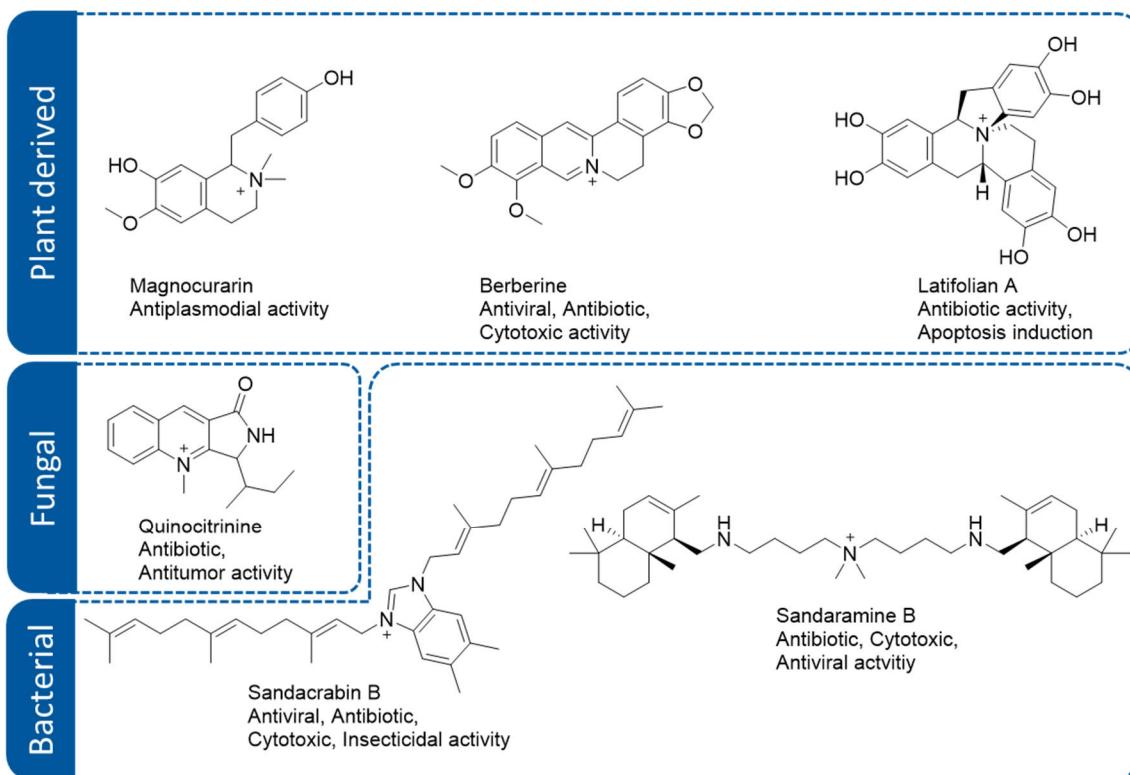

Figure 8 Natural products from different sources featuring a quaternary ammonium function and their respective biological activities.[44]

The diverse biological activities of the sandacrabins and sandaramines open up several starting points for their development towards antibacterial, antiviral or antiproliferative drugs. However, those broad activities are also a limiting factor for their development, as pharmaceutical drugs should display very selective activities. Especially for a use as antibiotic, antiviral substance or insecticide, the cytotoxic effects have to be reduced or preferably completely removed. Further medicinal chemistry development will show if the initial activities can be pinpointed to one selective activity by chemical modifications of the NPs.

7.7 Final conclusion

For bridging the gap between genomically-encoded bacterial potential and secondary metabolites actually isolated, most approaches focus on genetic manipulation of the respective strains. As shown in the previous chapters however, each step of a NP discovery workflow leaves room for variants and may be optimized. Looking at the diversity of chemical structures produced by myxobacteria alone it is reasonable that methods used for the isolation and structure elucidation of one NP be less suited for another one. Approaches to access novel bacterial secondary metabolites should therefore try to broaden the methods applied to the microbes under investigation. Optimizing only one parameter, such as the detection method or the extraction protocol, can already drastically increase the number of detectable or isolatable NPs. If multiple parameters are optimized in this workflow, it is likely that their effects even add up. Furthermore, all techniques applied will also influence the outcome of genome manipulation experiments, not only wild type strains, as the methodology applied after the manipulation itself is the same as when searching for novel NPs in wild type organisms. If extractability of NPs is increased, heterologous expression experiments will also benefit from the increased yields during extraction for example. Instead of only focusing on the genome, the techniques applied subsequently in the NP discovery workflow should be carefully evaluated to generate the best outcome and to detect and isolate the products of BGCs not accessible before. Furthermore, strains that already have provided one or even several NPs before shouldn't be neglected, as their biosynthetic potential mostly is still not fully elucidated. This accounts even for well-described strains such as *M. xanthus* DK1622, but especially holds true for novel strains which are commonly neglected after collecting "the low hanging fruits" in the means of easily detectable and isolatable metabolites. Secondary metabolites produced in low quantities may be even more interesting as those produced in large amounts, as they also might be optimized towards strong biological activities by the producing organism. Looking at the NPs described in this thesis, three out of four families exhibit intriguing biological activities. Only one of those however, was isolated in a bioactivity-guided isolation approach. This finding once again highlights that myxobacterial NPs are prone for finding bioactive molecules, even without prioritizing biologically active crude extracts or secondary metabolites with resistance genes in close neighborhoods to their BGCs. Limiting the detectable secondary metabolites to only one desired biological activity may also lead to excluding a variety of interesting NPs. Nowadays, we are not only facing just an antibiotic crisis but also viral pandemics. Furthermore, those infections often go hand in hand with other diseases, for example secondary bacterial infections during or after a viral infection, which seems to play an important role in corona mortality or viruses triggering cancer development, as the case in cervical cancer triggered by the human papilloma virus.[48,49] NP infection research should therefore try to keep an open focus, especially when it comes to the isolation of novel NPs. Nature provides us a nearly inexhaustible source of NPs, as the producing organisms will always adapt faster than we are capable of characterizing its products. With broadening the techniques applied to investigate those NPs, we are capable of making pharmaceutical use of an increasing amount of its intriguing chemical space, but we will likely never completely keep up with nature's pace. Instead of seeing how little we already have characterized, we should however stay curious about what still awaits us.

7.8 References

(1) Bader, C. D.; Neuber, M.; Panter, F.; Krug, D.; Müller, R. Supercritical Fluid Extraction Enhances Discovery of Secondary Metabolites from Myxobacteria. *Analytical chemistry* [Online early access]. DOI: 10.1021/acs.analchem.0c02995.

(2) Chandra Shekhar Pareek; Rafal Smoczynski; Andrzej Tretyn. Sequencing technologies and genome sequencing. *J Appl Genetics* **2011**, *52*, 413–435.

(3) Blin, K.; Wolf, T.; Chevrette, M. G.; Lu, X.; Schwalen, C. J.; Kautsar, S. A.; Suarez Duran, H. G.; Los Santos, E. L. C. de; Kim, H. U.; Nave, M.*; et al.* antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* **2017**, *45*, W36-W41.

(4) Garcia, R.; La Clair, J. J.; Müller, R. Future Directions of Marine Myxobacterial Natural Product Discovery Inferred from Metagenomics. *Mar. Drugs* **2018**, *16*.

(5) Hug, J. J.; Bader, C. D.; Remškar, M.; Cirnski, K.; Müller, R. Concepts and Methods to Access Novel Antibiotics from Actinomycetes. *Antibiotics* **2018**, *7*, 44.

(6) Hoffmann, M.; Auerbach, D.; Panter, F.; Hoffmann, T.; Dorrestein, P. C.; Müller, R. Homospermidine Lipids: A compound class specifically formed during fruiting body formation of *Myxococcus xanthus* DK1622. *ACS Chem. Biol.* **2018**, *13*, 273–280.

(7) Vartoukian, S. R.; Palmer, R. M.; Wade, W. G. Strategies for culture of 'unculturable' bacteria. *FEMS Microbiol Lett* **2010**, *309*, 1–7.

(8) Marmann, A.; Aly, A. H.; Lin, W.; Wang, B.; Proksch, P. Co-cultivation-a powerful emerging tool for enhancing the chemical diversity of microorganisms. *Mar. Drugs* **2014**, *12*, 1043–1065.

(9) Tatsuya Ito; Miyako Masubuchi. Dereplication of microbial extracts and related analytical technologies. *J Antibiot* **2014**, *67*, 353–360.

(10) Linington, R. G. npatlas - The Natural Products Atlas. https://www.npatlas.org/.

(11) Bader, C. D.; Panter, F.; Müller, R. In depth natural product discovery - Myxobacterial strains that provided multiple secondary metabolites. *Biotechnol. Adv.* **2020**, *39*, 107480.

(12) Alvarez-Sanchez, B.; Priego-Capote, F.; de Castro,M. D. L. Metabolomics analysis II. Preparation of biological samples prior to detection. *Trac-Trends Anal Chem* **2010**, *29*, 120–127.

(13) Azmir, J.; Zaidul, I. S.M.; Rahman, M. M.; Sharif, K. M.; Mohamed, A.; Sahena, F.; Jahurul, M. H.A.; Ghafoor, K.; Norulaini, N. A.N.; Omar, A. K.M. Techniques for extraction of bioactive compounds from plant materials: A review. *J. Food Eng.* **2013**, *117*, 426–436.

(14) Pourmortazavi, S. M.; Hajimirsadeghi, S. S. Supercritical fluid extraction in plant essential and volatile oil analysis. *J. Chromatogr. A* **2007**, *1163*, 2–24.

(15) Duval, J.; Destandau, E.; Pecher, V.; Poujol, M.; Tranchant, J.-F.; Lesellier, E. Selective enrichment in bioactive compound from Kniphofia uvaria by super/subcritical fluid extraction and centrifugal partition chromatography. *Journal of chromatography. A* **2016**, *1447*, 26–38.

(16) Kim, J.-B.; Bang, B.-H. Isolation and Purification of Berberine in Cortex Phellodendri by Centrifugal Partition Chromatography. *The Korean Journal of Food And Nutrition* **2014**, *27*, 532–537.

(17) McAlpine, J. B.; Chen, S.-N.; Kutateladze, A.; MacMillan, J. B.; Appendino, G.; Barison, A.; Beniddir, M. A.; Biavatti, M. W.; Bluml, S.; Boufridi, A.*; et al.* The value of universally available raw NMR data for transparency, reproducibility, and integrity in natural product research. *Natural product reports* **2019**, *36*, 35–107.

(18) 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.

(19) Vargas, F.; Weldon, K. C.; Sikora, N.; Wang, M.; Zhang, Z.; Gentry, E. C.; Panitchpakdi, M. W.; Caraballo-Rodríguez, A. M.; Dorrestein, P. C.; Jarmusch, A. K. Protocol for Community-created Public MS/MS Reference Spectra Within the Global Natural Products Social Molecular Networking Infrastructure. *Rapid Commun. Mass Spectrom.* **2020**, e8725.

(20) Tobias Kind; Oliver Fiehn. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanal Rev* **2010**, *2*, 23–60.

(21) Bian, X.; Plaza, A.; Zhang, Y.; Müller, R. Two more pieces of the colibactin genotoxin puzzle from *Escherichia coli* show incorporation of an unusual 1-aminocyclopropanecarboxylic acid moiety. *Chem. Sci.* **2015**, *6*, 3154–3160.

(22) Homburg, S.; Oswald, E.; Hacker, J.; Dobrindt, U. Expression analysis of the colibactin gene cluster coding for a novel polyketide in *Escherichia coli*. *FEMS Microbiol. Lett.* **2007**, *275*, 255–262.

(23) Putze, J.; Hennequin, C.; Nougayrede, J. P.; Zhang, W.; Homburg, S.; Karch, H.; Bringer, M. A.; Fayolle, C.; Carniel, E.; Rabsch, W.*; et al.* Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. *Infect. Immun.* **2009**, *77*, 4696–4703.

(24) Wernke, K. M.; Xue, M.; Tirla, A.; Kim, C. S.; Crawford, J. M.; Herzon, S. B. Structure and bioactivity of colibactin. *Bioorganic & Medicinal Chemistry Letters* **2020**, *30*, 127280.

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

(26) Kitsche, A.; Kalesse, M. Configurational assignment of secondary hydroxyl groups and methyl branches in polyketide natural products through bioinformatic analysis of the ketoreductase domain. *ChemBioChem* **2013**, *14*, 851–861.

(27) Chhetri, B. K.; Lavoie, S.; Sweeney-Jones, A. M.; Kubanek, J. Recent trends in the structural revision of natural products. *Nat. Prod. Rep.* [Online early access]. DOI: 10.1039/c8np00011e.

(28) Paulsen, B.; Fredriksen, K. A.; Petersen, D.; Maes, L.; Matheeussen, A.; Naemi, A.-O.; Scheie, A. A.; Simm, R.; Ma, R.; Wan, B.*; et al.* Synthesis and antimicrobial activities of N6-hydroxyagelasine analogs and revision of the structure of ageloximes. *Bioorganic & Medicinal Chemistry* **2019**, *27*, 620–629.

(29) Meiser, P.; Bode, H. B.; Müller, R. The unique DKxanthene secondary metabolite family from the myxobacterium *Myxococcus xanthus* is required for developmental sporulation. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 19128–19133.

(30) Polkade, A. V.; Mantri, S. S.; Patwekar, U. J.; Jangid, K. Quorum Sensing: An Under-Explored Phenomenon in the Phylum Actinobacteria. *Front. Microbiol.* **2016**, *7*, 131.

(31) Hughes, J. P.; Rees, S.; Kalindjian, S. B.; Philpott, K. L. Principles of early drug discovery. *Br. J. Pharmacol.* **2011**, *162*, 1239–1249.

(32) Lilic, M.; Chen, J.; Boyaci, H.; Braffman, N.; Hubin, E. A.; Herrmann, J.; Müller, R.; Mooney, R.; Landick, R.; Darst, S. A.*; et al.* The antibiotic sorangicin A inhibits promoter DNA unwinding in a *Mycobacterium tuberculosis* rifampicin-resistant RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* [Online early access]. DOI: 10.1073/pnas.2013706117.

(33) Jenke-Kodama, H.; Müller, R.; Dittmann, E. Evolutionary mechanisms underlying secondary metabolite diversity. In *Natural compounds as drugs*; Petersen, F., Amstutz, R., Eds.; Progress in drug research v. 65; Birkhäuser; [London : Springer: Basel, 2008; pp 119–140.

(34) Mishra, P.; Tripathi, A.; Dikshit, A.; Pandey, A. Insecticides Derived from Natural Products: Diversity and Potential Applications. In *Natural bioactive products in sustainable agriculture [electronic resource]*; Singh, J., Yadav, A. N., Eds.; Springer: Singapore, 2020; pp 83–99.

(35) B. S. Griffiths. Microbial-feeding nematodes and protozoa in soil: Their effectson microbial activity and nitrogen mineralization in decomposition hotspots and the rhizosphere. *Plant Soil* **1994**, *164*, 25–33.

(36) Zhang, W. W.; Karmakar, S.; Gannavaram, S.; Dey, R.; Lypaczewski, P.; Ismail, N.; Siddiqui, A.; Simonyan, V.; Oliveira, F.; Coutinho-Abreu, IV*; et al.* A second generation leishmanization vaccine with a markerless attenuated Leishmania major strain using CRISPR gene editing. *Nature communications* **2020**, *11*.

(37) Prajapati, S. K.; Ayanful-Torgby, R.; Pava, Z.; Barbeau, M. C.; Acquah, F. K.; Cudjoe, E.; Kakaney, C.; Amponsah, J. A.; Obboh, E.; Ahmed, A. E.*; et al.* The transcriptome of circulating sexually committed Plasmodium falciparum ring stage parasites forecasts malaria transmission potential. *Nat Commun* **2020**, *11*.

(38) George, S.; Joy, T. M.; Kumar, A.; Panicker, K. N.; George, L. S.; Raj, M.; Leelamoni, K.; Nair, P. Prevalence of Neglected Tropical Diseases (Leishmaniasis and Lymphatic Filariasis) and Malaria Among a Migrant Labour Settlement in Kerala, India. *Journal of immigrant and minority health* **2019**, *21*, 563–569.

(39) Sarker, S. D.; Nahar, L.; Miron, A.; Guo, M. Anticancer natural products. In *Medicinal natural products: A disease-focused approach / editors, Satyajit Dey Sarker, Lutfun Nahar*; Sarker, S. D., Nahar, L., Eds.; Annual reports in medicinal chemistry 55; Academic Press: Amsterdam, 2020; pp 45–75.

(40) Shiloh, M. U.; DiGiuseppe Champion, P. A. To catch a killer. What can mycobacterial models teach us about Mycobacterium tuberculosis pathogenesis? *Antimicrobials* **2009**, *13*, 86–92.

(41) Secor, P. R.; Sweere, J. M.; Michaels, L. A.; Malkovskiy, A. V.; Lazzareschi, D.; Katznelson, E.; Rajadas, J.; Birnbaum, M. E.; Arrigoni, A.; Braun, K. R.*; et al.* Filamentous bacteriophage promote biofilm assembly and function. *Cell Host & Microbe* **2015**, *18*, 549–559.

(42) Vasse, M.; Wielgoss, S. Bacteriophages of *Myxococcus xanthus*, a Social Bacterium. *Viruses* **2018**, *10*.

(43) Verma, S.; Twilley, D.; Esmear, T.; Oosthuizen, C. B.; Reid, A.-M.; Nel, M.; Lall, N. Anti-SARS-CoV Natural Products With the Potential to Inhibit SARS-CoV-2 (COVID-19). *Frontiers in pharmacology* **2020**, *11*.

(44) Joyce, M. D.; Jennings, M. C.; Santiago, C. N.; Fletcher, M. H.; Wuest, W. M.; Minbiole, K. P. Natural product-derived quaternary ammonium compounds with potent antimicrobial activity. *The Journal of antibiotics* **2016**, *69*, 344–347.

(45) Stermitz, F. R.; Lorenz, P.; Tawara, J. N.; Zenewicz, L. A.; Lewis, K. Synergy in a medicinal plant: antimicrobial action of berberine potentiated by 5'-methoxyhydnocarpin, a multidrug pump inhibitor. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 1433–1437.

(46) Rochfort, S. J.; Towerzey, L.; Carroll, A.; King, G.; Michael, A.; Pierens, G.; Rali, T.; Redburn, J.; Whitmore, J.; Quinn, R. J. Latifolians A and B, novel JNK3 kinase inhibitors from the Papua New Guinean plant Gnetum latifolium. *J. Nat. Prod.* **2005**, *68*, 1080–1082.

(47) Xiao-Fei Shang; Susan L. Morris-Natschke; Ying-Qian Liu; Xiao Guo; Xiao-Shan Xu; Masuo Goto; Jun-Cai Li; Guan-Zhou Yang; Kuo-Hsiung Lee. Biologically active quinoline and quinazoline alkaloids part I. *Medicinal Research Reviews* **2018**, *38*, 775–828.

(48) Langford, B. J.; So, M.; Raybardhan, S.; Leung, V.; Westwood, D.; MacFadden, D. R.; Soucy, J.-P. R.; Daneman, N. Bacterial co-infection and secondary infection in patients with COVID-19: a living rapid review and meta-analysis. *Clin. Microbiol. Infect.* [Online early access]. DOI: 10.1016/j.cmi.2020.07.016.

(49) Burd, E. M. Human papillomavirus and cervical cancer. *Clinical microbiology reviews* **2003**, *16*, 1–17.