Discovery and Characterization of Myxobacterial and Actinobacterial Natural Products by Metabolome and Genome Mining Approaches

Dissertation
zur Erlangung des Grades
des Doktors der Naturwissenschaften
der Naturwissenschaftlich-Technischen Fakultät
der Universität des Saarlandes

von
Christine Walt
Saarbrücken
2022

Tag des Kolloquiums:	23. März 2023
Dekan:	Prof. Dr. Ludger Santen
Berichterstatter:	Prof. Dr. Rolf Müller
	Prof. Dr. Anna K.H. Hirsch
Vorsitz:	Prof. Dr. Gregor Jung
Akad. Mitarbeiter:	Dr. Stefan Boettcher

Diese Arbeit entstand unter der Anleitung von Prof. Dr. Rolf Müller am Institut für Pharmazeutische Biotechnologie der Naturwissenschaftlich-Technischen Fakultät der Universität des Saarlandes von November 2018 bis Dezember 2022.

“You are not meant to do what is easy.

You are meant to challenge yourself.”

- Justin Timberlake

Danksagung

Zuerst möchte ich meinem Doktorvater Prof. Dr. Rolf Müller meinen Dank aussprechen für das Ermöglichen meiner Promotion in seiner Arbeitsgruppe MINS mitsamt den unglaublich vielfältigen Gerätschaften, sodass ich mein analytisches Potential ergiebig ausschöpfen konnte und so als Wissenschaftlerin wachsen konnte. Weiterhin bedanke ich mich bei meiner wissenschaftlichen Begleiterin Prof. Dr. Anna Hirsch und meinem Betreuer Dr. Daniel Krug, die mir vor allem während der Thesis Komitee Meetings nützliche Denkanstöße gegeben und mir mit ihrer wissenschaftlichen Erfahrung beigestanden haben.

Ein riesiges Dankeschön geht an meine Betreuerin im Labor Dr. Chantal Bader für deine Geduld, das Teilen deiner wissenschaftlichen Expertise und deine moralische Unterstützung. Du hast immer ein offenes Ohr für mich und hast stets die richtigen Worte gefunden, um mich zu beruhigen, zu motivieren und mir zu helfen über mich hinauszuwachsen. Ich bin so glücklich diesen Weg mit dir gegangen zu sein und unsere Freundschaft bleibt über Raum und Zeit erhalten. Weiterhin bedanke ich mich von Herzen bei meinem Partner fürs Leben Dr. Nicolas Frank dafür, dass du mich ohne Worte verstehst und mich bei jeglichen Schwierigkeiten im Labor und im Privaten ohne Zögern unterstützt. Mit dir an meiner Seite habe ich das Gefühl alles schaffen zu können. Du bist mein Fels in der Brandung und machst mich mit jedem Tag glücklicher. Ich freue mich schon auf unsere gemeinsame Zukunft als Wissenschaftler und Paar. Außerdem möchte ich mich bei Dr. Fabian Panter, Sophia Panter und Patrick Haack für eure stetige Unterstützung während der Promotion bedanken, dass ihr mich immer aufgefangen habt, bevor ich falle (wortwörtlich im Falle von Fabian). Ich bin sehr froh euch als Freunde zu haben. In einer Welt voller Ablehnung seid ihr fünf mein Ort der Akzeptanz.

Natürlich gebührt ein besonderer Dank den Besuchern des Analytikbüros Dr. Jan Schlemmer, Dr. Sebastian Groß, Dr. Sebastian Adam alias Peter, Dr. Alexander Kiefer, Dr. Joachim Hug und Dr. Laura Franz. Mit euch konnte ich als Wissenschaftlerin wie als Person wachsen und erinnere mich gerne an die tolle Zeit mit euch zurück. Meinen best NMR buddies Sebastian Walesch und Joy Birkelbach will ich im Besonderen danken, dass wir die Höhen und Tiefen einer Promotion von Anfang an zusammen bewältigen konnten und uns gegenseitig unterstützt haben.

Ich möchte mich bei der gesamten Arbeitsgruppe bedanken für ein angenehmes Arbeitsklima, denn ich habe gelernt, dass man als Team deutlich einfacher gute Forschung machen kann. Vor allem geht ein Dankeschön an die neuen Analytiker Alexander Voltz und Christoph Porten. Ihr

habt mir mit eurer motivierten Art viel Unterstützung entgegengebracht, sodass ich mich während des letzten Jahrs besser auf meine Thesis konzentrieren konnte. Weiterhin danke ich Dr. Jake Haeckl, Dr. Markus Neuber, Selina Deckarm, Anna-Lena Huber, Carsten Seyfert, Sabine Backes, Kathrin Andres und Dr. Ronald Garcia für die tolle Zeit im Labor und während der Freizeit. Dankeschön an Jeenu Joy, Janetta Coetzee und Nestor Zaburannyi dafür, dass ihr so tolle Bürokollegen wart, wir so viel zusammen zu lachen hatten und ich mein Englisch deutlich verbessern konnte. Außerdem möchte ich meinen Dank Dr. Martin Empting, Dr. Valentin Jakob und Dr. Philine Kirsch aussprechen für eure Betreuung während meines Masters und eure Wertschätzung meines Beitrags in Form meiner allerersten Publikation. Vielen Dank auch an Dr. Chengzhang Fu und Yunsheng Gao für die wertvolle Zusammenarbeit an spannenden Projekten.

Schließlich möchte ich mich für die stetige moralische Unterstützung meiner Freunde und meiner Familie bedanken. Vielen Dank im Besonderen an meine engen Freunde Sonja Bleymehl (vor allem für die wöchentlichen Telefonate), Jacqueline Heier und Lara Jasinski. Ihr habt mir immer geholfen das Ziel im Blick zu behalten und das Beste aus mir herauszuholen. Herzlichen Dank an meine Eltern Johannes und Natalie Walt und meiner Schwester Oxana Ebert, dass ihr schon immer wusstet, dass ich für etwas Großes bestimmt bin und dass ihr mir ermöglicht habt unseren liebenswerten Merlin als Endorphin-Spender in mein Leben zu lassen. От всей души благодарю вас за сердечную поддержку и помощь в трудную минуту.

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 zur Veröffentlichung:

Publikationen

Walt, Christine; Bader, Chantal D.; Krug, Daniel; Müller, Rolf: Aggravated Cultivation of Myxobacteria Stimulates Secondary Metabolism. *Manuscript in preparation.*

Gao, Yunsheng; **Walt, Christine**; Bader, Chantal D.; Müller, Rolf: Sorangibactin: Genome Guided Discovery of a Novel Myxobacterial Siderophore. *Manuscript in preparation.*

Fu, Chengzhang; Liu, Yunkun; **Walt, Christine**; Bader, Chantal D.; Rasheed, Sari; Lukat, Peer; Neuber, Markus; Kalinina, Olga; Müller, Rolf: Unusual biosynthesis of alkylprolines and homo-amino acids in mycoplanecins: extremely potent anti-tuberculosis antibiotics targeting DnaN. *Manuscript in preparation.*

Publikationen, die nicht Teil dieser Arbeit sind

Kirsch, Philine; Jakob, Valentin; Elgaher, Walid A. M.; **Walt, Christine**; Oberhausen, Kevin; Schulz, Thomas F.; Empting, Martin (**2020**): Discovery of Novel Latency-Associated Nuclear Antigen Inhibitors as Antiviral Agents Against Kaposi's Sarcoma-Associated Herpesvirus. *ACS Chemical Biology* 15 (2), pp. 388–395.
DOI: 10.1021/acschembio.9b00845.

Hug, Joachim J.; Frank, Nicolas A.; **Walt, Christine**; Šenica, Petra; Panter, Fabian; Müller, Rolf (**2021**): Genome-Guided Discovery of the First Myxobacterial Biarylitide Myxarylin Reveals Distinct C–N Biaryl Crosslinking in RiPP Biosynthesis. *Molecules* 26 (24), p. 7483.
DOI: 10.3390/molecules26247483.

Bader, Chantal D.; Panter, Fabian; Garcia, Ronald; Tchesnokov, Egor P.; Haid, Sibylle; **Walt, Christine**; Spröer, Cathrin; Kiefer, Alexander F.; Götte, Matthias; Overmann, Jörg; Pietschmann, Thomas; Müller, Rolf (**2022**): Sandacrabins - Structurally Unique Antiviral RNA Polymerase Inhibitors from a Rare Myxobacterium. *Chemistry – A European Journal* 28 (10), e202104484.
DOI: 10.1002/chem.202104484.

Vij, Raghav; Tröger, Sophie; **Walt, Christine**; Haeckl, F. P. Jake; Müller, Rolf; Kniemeyer, Olaf; Brakhage, Axel; Hube, Bernhard; Brunke, Sascha (2022): S9.3d Antivirulence drug discovery to disarm Candida albicans with metabolites from myxobacteria. *In Med. Mycol.* 60 (Supplement_1), Article myac072S93d.
DOI: 10.1093/mmy/myac072.S9.3d.

Kohr, Michael; **Walt, Christine**; Dastbaz, Jan; Müller, Rolf; Kazmaier, Uli (2022): Total synthesis of Myxoprincomide, a secondary metabolite from *Myxococcus xanthus*. *Org. Biomol. Chem.*, published ahead of print.
DOI: 10.1039/D2OB02021A.

Tagungsbeiträge

Christine Walt, Chantal D. Bader, Daniel Krug and Rolf Müller (**2022**) Outside the Comfort Zone: Stimulating Myxobacterial Secondary Metabolism (Poster) *Promovierendentag der Naturwissenschaftlich-Technischen Fakultät, Saarbrücken, Germany*

Christine Walt, Chantal D. Bader, Daniel Krug and Rolf Müller (**2022**) Outside the Comfort Zone: Stimulating Myxobacterial Secondary Metabolism (Poster) *HIPS Symposium 2022, Saarbrücken, Germany*

Zusammenfassung

Myxobakterien und Actinobakterien sind fruchtbare Ressourcen neuartiger Naturstoffe (NS) mit interessanten Strukturen und vielversprechenden Bioaktivitäten. Im Rahmen dieser Arbeit wurden orthogonale Metabolom- und Genom-Mining Ansätze eingesetzt, um das biosynthetische Potenzial des Stammes MSr11367 zu erschließen und neue myxobakterielle NS zu entdecken. Zudem wurde die Erforschung der stark gegen den Tuberkuloseerreger wirksamen Mycoplanecine aus dem Actinobakterium *Actinoplanes awajinensis* vorgetrieben. Drei neue myxobakterielle Verbindungen namens Myxolutamid, Pentacyclic acid und Glucodiolic acid wurden durch Metabolom-Mining bei verschiedenen Kultivierungsbedingungen identifiziert, isoliert und ihre Strukturen aufgeklärt. Parallel dazu wurde ein orthogonaler Genom-Mining-Ansatz auf MSr11367 angewandt, was zur Entdeckung eines Coelibactin-ähnlichen nichtribosomalen Peptidsynthetase Genclusters führte. Nach heterologer Expression dieses Genclusters wurde der neue myxobakterielle Eisenchelator Sorangibactin isoliert und dessen Struktur wurde durch umfassende NMR-Experimente *de novo* aufgeklärt, wobei eine ungewöhnliche C-terminale γ-Thiolacton-Einheit entdeckt wurde. Schließlich führte die genomische Suche nach 4-Methylprolin-Stoffwechselwegen zur Wiederentdeckung der stark antituberkulösen Mycoplanecine. Die vier am besten produzierten Mycoplanecine wurden isoliert, ihre Struktur verifiziert und ihre Bindungsaffinität zu ihrem Target DnaN mittels Thermophorese bestimmt.

Abstract

Myxobacteria and actinobacteria have proven to be fruitful resources for novel natural products (NPs) with interesting scaffolds and promising bioactivities. In the course of this work, orthogonal metabolome and genome mining approaches were employed to exploit the biosynthetic potential of the myxobacterial strain MSr11367 for NP discovery. Additionally, studies on the highly antitubercular mycoplanecins produced by the actinobacteria *Actinoplanes awajinensis* were revived. Three novel myxobacterial compounds termed myxolutamid, pentacyclic acid and glucodiolic acid were successfully identified via metabolome mining using different cultivation conditions and their structures were *de novo* elucidated via NMR. In parallel, an orthogonal genome mining approach was applied to MSr11367 to further exploit the biosynthetic potential, which led to the discovery of a coelibactin-like nonribosomal peptide synthetase gene cluster. Heterologous expression of this biosynthetic gene cluster was followed by purification of a novel myxobacterial siderophore named sorangibactin, whose uncommon structure was *de novo* elucidated by NMR uncovering an unusual C-terminal γ-thiolactone moiety. Lastly, genome mining for 4-methylproline pathways resulted in the reawakening of the highly antitubercular mycoplanecins. The four most abundant mycoplanecins were isolated, their structures verified and their binding affinity to their molecular target DnaN determined using microscale thermophoresis.

Table of contents

Chapter 1

1. Introduction

Our modern world is increasingly intimidated by the rising antimicrobial resistance, especially of the so called ESKAPE (***E****nterococcus faecium,* ***S****taphylococcus aureus,* ***K****lebsiella pneumoniae,* ***A****cinetobacter baumannii,* ***P****seudomonas aeruginosa,* ***E****nterobacter* species) pathogens [1]. This global threat is aggravated through wrong application of commercially available antibiotics as well as the decline in the antibiotic development pipeline [2]. Natural products (NP), which have historically proven to be a promising source for antibiotics [3], address this need for novel innovative therapeutics. By the broadest definition, a natural product is any substance produced by a living organism originating from nature [4]. This includes both primary metabolites, which are essential for survival in terms of being part of basic metabolic pathways, as well as secondary metabolites, which usually have a broader range of purpose. In a narrower sense, NPs are seen as secondary metabolites resulting from prokaryotic or eukaryotic systems [5]. NPs can derive from all kingdoms of life such as morphine isolated from the plant *Papaver somniferum* [6], penicillin G produced by the fungus *Penicillium rubens* [7], the sandarazols as secondary metabolites of the myxobacterium *Sandaracinus* sp. [8], isoprenyl glycerol ether 1 produced by the archaeon *Thermococcus* [9], and areno-bufagin isolated from the skin of the toad *Bufo arenarum* [10] (Figure 1A). Some well-known functions of secondary metabolites are to defend the producer from hazardous environment and predators, to maintain a valuable symbiosis with other organisms, to communicate with one another using signaling molecules or to transport essential metal ions into the cells [11] (Figure 1B).

NPs and their semisynthetic derivatives proved to be a fruitful source of new drugs as they represent 32% of the 1394 small-molecule drugs that have been approved for the treatment of human diseases during the last four decades [12]. Thus, the discovery of hitherto unknown NPs is essential for the development of drug leads with improved physicochemical, pharmacokinetic and pharmacodynamics properties, especially considering antimicrobial drugs two thirds of which are either NPs or NP derived [13]. Among the various sources of NPs, microbial organisms and in particular myxobacteria and actinobacteria harbor a great untapped potential for the production of novel chemical scaffolds [14,15].

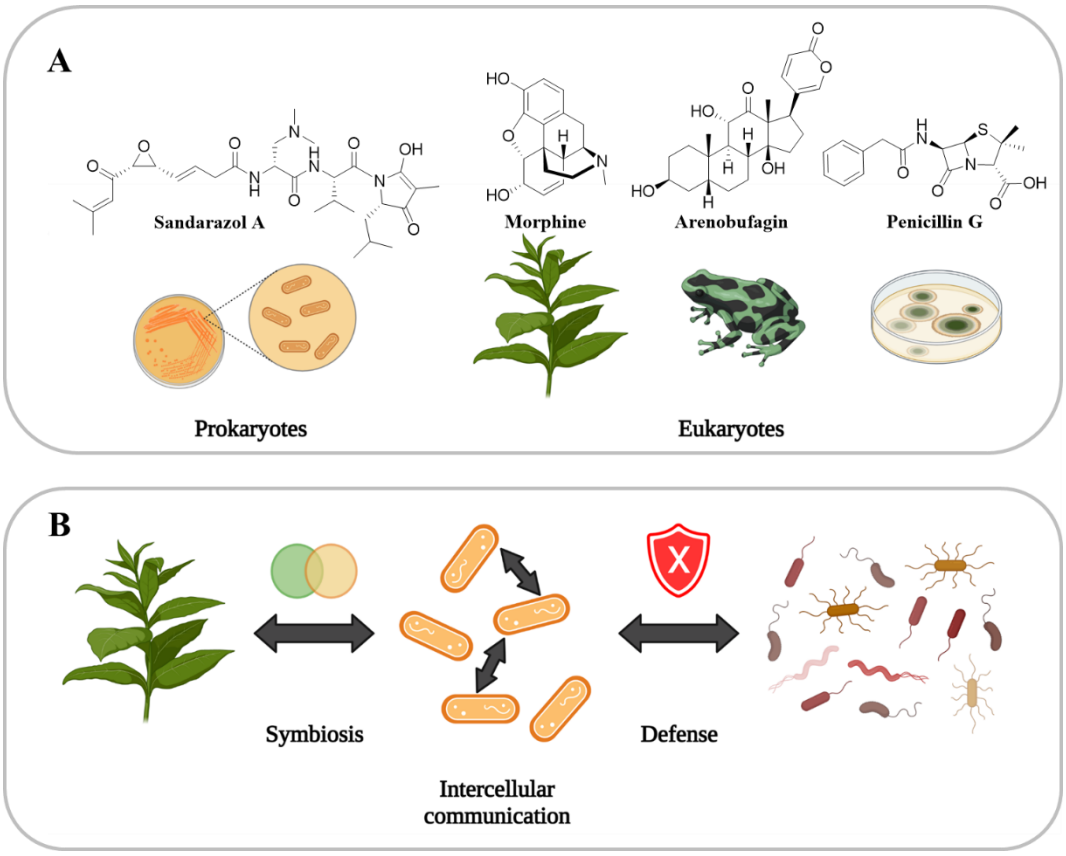


Figure 1: A) Prokaryotes (bacteria or archaea) and selected eukaryotes (plants, fungi or animals) as secondary metabolite producers, e.g. *sandarazols produced by* Sandaracinus sp., *morphine isolated from* Papaver somniferum, *arenobufagin isolated from skin of* Bufo arenarum, *penicillin G produced by* Penicillium rubens*; B) secondary metabolites utilized for attraction of symbiotic organisms, for cell-cell communication and for defense against competing organisms.*

1.1 Myxobacteria as a blooming resource for natural products

Myxobacteria, another term for the order *Myxococcales*, are Gram-negative, soil-dwelling δ-proteobacteria with a complex life cycle and the ability of coordinated swarming in gliding motions as prominent hallmarks [16]. Challenging starvation conditions induce the production of signaling molecules that activate the formation of spore-containing fruiting bodies [17]. Remarkably, the large genomes of myxobacteria that reach sizes beyond 10 Mbps, *e.g.* the most prominent representatives *Minicystis rosea* with 16 Mbps [18] and *Sorangium cellulosum* So0157-2 with 14.8 Mbps [19], contain many biosynthetic gene clusters (BGCs). These are forming the basis for the biosynthesis of many structurally diverse natural products via complex biosynthetic routes [20], which are encoded by two or more clustered genes. The order *Myxococcales* can be taxonomically subdivided in the 3 suborders *Sorangiineae*, *Cystobacterineae* and *Nannocystineae* [21]. According to the state of May 2018, the suborder *Sorangiineae* comprises 4 families and 11 genera, *Cystobacterineae* are subdivided in 4 families with a total of 12 genera and *Nannocystineae* feature 2 families and 6 genera [22].

1.1.1 Prominent myxobacterial natural products from different genera

Myxobacteria are recognized as promising producers for novel NPs that exhibit diverse structural features and biological activities [23,24]. A selected subset of published myxobacterial secondary metabolites of the past 40 years spanning 8 genera across the two suborders *Cystobacterineae* and *Sorangiineae* is listed in Figure 2.

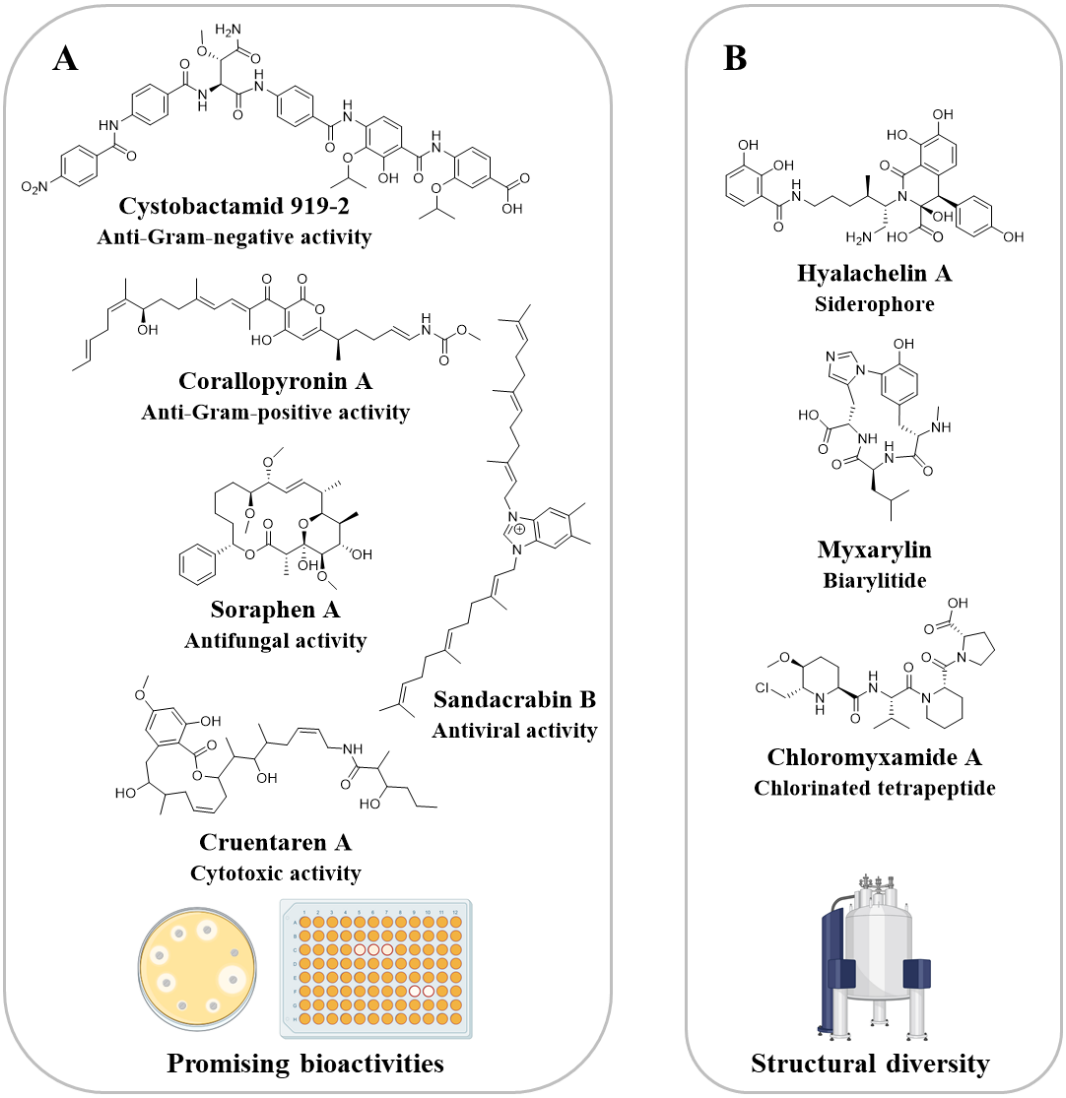


Figure 2: A) Selected natural products with promising bioactivities and B) structural diversity from different myxobacterial genera; A: cystobactamid 919-2 as a non-ribosomal peptide synthetase (NRPS) product with anti-Gram-negative activity [25]; soraphen A as a polyketide synthase (PKS) product with antifungal activity [26]; corallopyronin A as a product of a trans-AT-type PKS/NRPS hybrid cluster with anti-Gram-positive activity [27]; cruentaren A as a PKS product with cytotoxic activity [28]; sandacrabin B as a terpenoid alkaloid with antiviral activity [29]; B: hyalachelin A as a catecholate-type siderophore comprising an unusual cyclic moiety [30]; myxarylin as a ribosomally synthesized and post-translationally modified peptide (RiPP) with an uncommon C-N-biaryl crosslink [31]; chloromyxamide A as an NRPS product with a peculiar chloromethyl pipecolic acid moiety [32].

These compounds represent secondary metabolites from 8 different myxobacterial genera covering anti-Gram-negative and anti-Gram-positive, antifungal, cytotoxic and antiviral

bioactivities as well as several unusual chemical scaffolds. Cystobactamids were isolated from *Cystobacter* sp. and show both anti-Gram-positive as well as anti-Gram-negative activities. Since anti-Gram-negative activities are less frequently found and address the most pressing medical need, cystobactamids were prioritized for their activity against *Escherichia coli* and *Acinetobacter baumannii* in the lower µg/mL concentration range through inhibition of bacterial topoisomerases such as DNA gyrase and topoisomerase IV [25]. Soraphen A is known for its activity against many fungi (16 different test organisms) and was first detected as product of a polyketide synthase (PKS) of the *Sorangium cellulosum* strain So ce26 [26]. The trans-AT-type PKS/NRPS hybrid product corallopyronin A was isolated from *Corallococcus coralloides* and shows anti-Gram-positive bioactivity against *Staphylococcus aureus* including methicillin-resistant strains (MRSA) by acting as a bacterial RNA synthesis inhibitor [27]. The PKS product of *Byssovorax cruenta* cruentaren A is a benzolactone with cytotoxic activity against various human cancer cell lines as inhibitor of mitochondrial F_1-ATPases [28]. Sandacrabin B is a terpenoid-alkaloid RNA polymerase inhibitor that was isolated from *Sandaracinus defensii* and shows potent antiviral activity against the human pathogenic coronavirus HCoV229E [29]. Furthermore, not only already assigned bioactivities, but also structurally interesting secondary metabolites, where the actual ecological function or other beneficial functions for humans still need to be uncovered are of interest. Hyalachelin A is a catecholate-type siderophore with an unusual cyclic moiety isolated from *Hyalangium minutum* [30]. The ribosomally synthesized and post-translationally modified peptide (RiPP) myxarylin was purified from *Pyxidicoccus fallax* An d48 and is the first myxobacterial biarylitide that features an uncommon C-N biaryl crosslink [31]. Finally, chloromyxamide A was identified from *Myxococcus* sp. MCy10608, which features a rarely observed chloromethyl pipecolic acid moiety, and was isolated from the alternative producer MCy10615 [32].

These NP discoveries illustrate the myxobacterial potential as resource for novel secondary metabolites and motivate the application of different strategies to further exploit their metabolome and genome. On the one hand, the interdependence between taxonomic distance and chemical diversity of natural products, as inferred from a subset of approximately 2300 myxobacterial strains, encourages the exploration of myxobacterial strains from new genera rather than focusing on the same genus [33]. The intricacy of this tactic is clearly the identification and prioritization of target compounds by selection of metabolome, genome or bioactivity examination (Section 1.4). On the other hand, the so called one strain many compounds (OSMAC) approach emboldens the in-depth analysis of the secondary metabolome of single bacterial strains, especially for strains with a distinct biosynthetic potential represented

by the presence of many unknown BGCs [34], as exemplified for *Myxococcus xanthus* (*M. xanthus*) DK1622 (Section 1.1.2.). The idea of the OSMAC approach is to grow a single strain under different environmental conditions and subsequently isolate many different substances instead of using several different strains to purify only few molecules produced under a single condition. In both approaches, a crucial step is a reliable dereplication to avoid repurification of known natural products and identify novel congeners in a timely manner.

1.1.2 Natural products from the myxobacterial model strain *M. xanthus* DK1622

The strain *Myxococcus xanthus* DK1622 is one of the best studied myxobacterial organisms and is frequently utilized as an expression host of exogenous BGCs originating from other myxobacteria that are more difficult to handle [35]. Genomic mining exposed at least 18 biosynthetic gene clusters (BGCs) encoding assembly lines of PKS, NRPS and hybrid products thereof [36]. Thereby, for two-thirds of the predicted BGC pathways no NPs could be related yet, which shows the huge untapped biosynthetic potential. Figure 3 thus displays the known secondary metabolite inventory of a single strain that was well-studied.

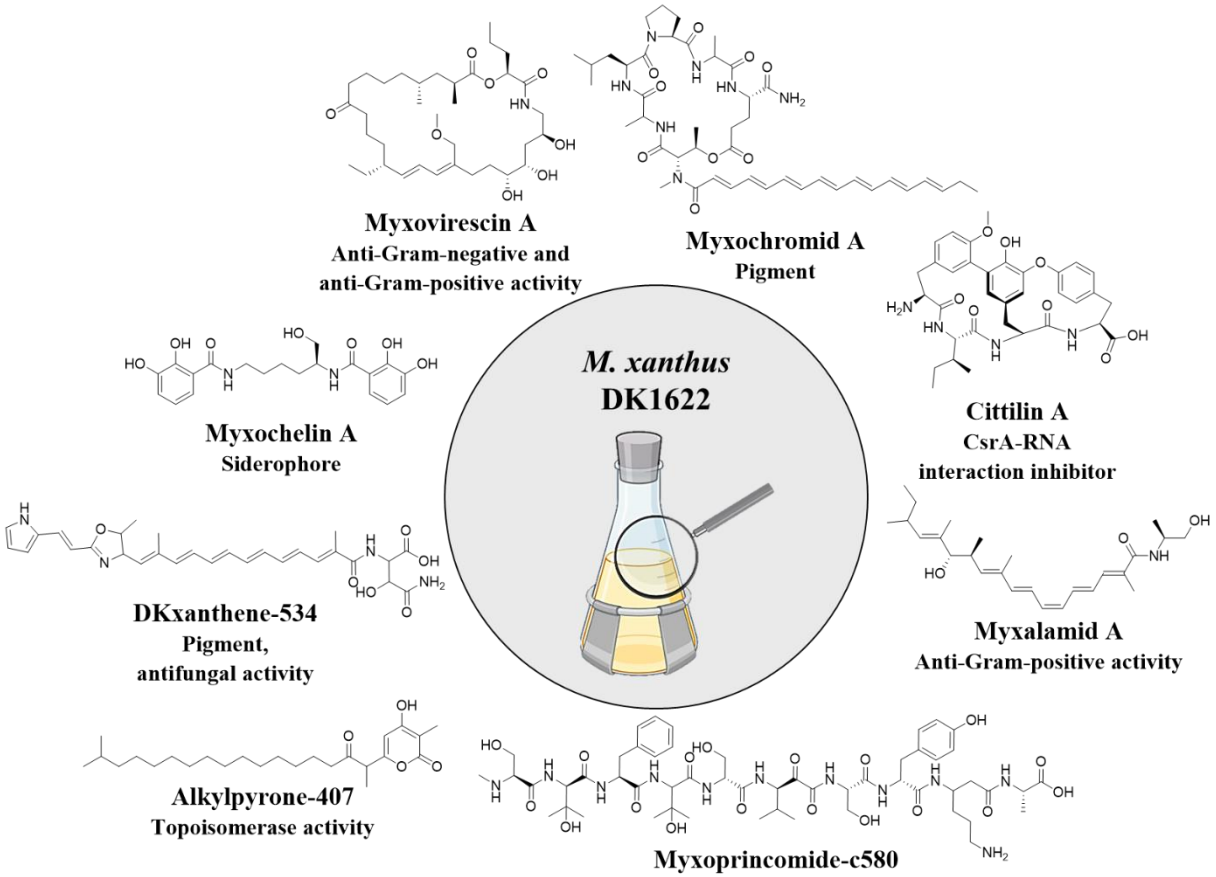


Figure 3: Natural products produced by Myxococcus xanthus DK1622 showing promising bioactivities and diverse structures [37]; the pigment myxochromid A as PKS/NRPS hybrid product [38]; the CsrA-RNA interaction inhibitor cittilin A as RiPP [39]; myxalamid A as PKS/NRPS hybrid product with anti-Gram-positive activity [40]; myxoprincomide-c580 as NRPS/PKS hybrid product [41]; alkylpyrone-407 as type III PKS product with topoisomerase activity [42]; DKxanthene-534 as PKS/NRPS hybrid product and pigment with moderate antifungal activity [43,44]; the catechole siderophore myxochelin [45]; myxovirescin A as PKS/NRPS hybrid product with anti-Gram-negative and anti-Gram-positive bioactivity [46].

Although all these compounds were found to be produced by *M. xanthus* DK1622, many were initially discovered in other *Myxococcus* species [37]. Myxochromid A, a PKS/NRPS hybrid product, was originally identified in *Myxococcus virescens* as a novel pigment, whereby the

derivative myxochromid B was found to be the main product in *Myxococcus flavescens* [38]. The cittilins, a class of bicyclic tetrapeptide RiPPs with unusual aryl-oxygen-aryl ether bonds and biaryl linkage, were firstly isolated from *M. xanthus* Mx x48. Cittilin A was hereby reported to be a promising CsrA-RNA interaction inhibitor [39,47,48]. The myxalamids are PKS/NRPS hybrid products with activity against several Gram-positive bacteria, molds and yeasts, that were firstly described for *M. xanthus* Mx x12 [40]. Myxoprincomides are NRPS/PKS products that were discovered via a metabolome mining approach and isolated from *M. xanthus* DK1622 [41]. A genomic survey of type III PKS biosynthetic pathways yielded in the discovery of the alkylpyrones in *M. xanthus* DK1622, which feature *in vitro* activity against *E.coli* topoisomerase IV [42]. The DKxanthenes are structurally interesting pigments produced by *M. xanthus* DK1050 as PKS/NRPS hybrid compounds, are important for developmental sporulation [43] and show moderate antifungal activity against *Aspergillus niger*, *Candida albicans*, and *Rhizopus stolonifer* [44]. The siderophore myxochelin A was firstly isolated from *Angiococcus disciformis* An d30 and shows weak anti-Gram-positive activity [45], whereby the myxochelin family was also found to be produced by *M. xanthus* DK1622 [37]. *Myxococcus virescens* Mx v48 was found to produce an antibiotic class named myxovirescin, PKS/NRPS hybrid products exhibiting bioactivities against several Gram-positive and particularly also Gram-negative pathogens such as enterobacteria [46]. These findings underpin the potential of in-depth investigation of the secondary metabolome of multiproducer strains to exploit the full biosynthetic potential instead of just sporadic purification of few novel compounds per strain. This idea is further supported by the four novel natural products that were isolated from the myxobacterial strain MSr11367, which belongs to the suborder *Sorangiineae* (see Chapters 2 and 3).

1.2 Actinobacteria as a long-standing resource for natural products

Apart from the myxobacteria described in the section above, the large phylum *actinobacteria* proved to be a promising resource for many natural products [49,50]. In contrast to myxobacteria, actinobacteria are Gram-positive filamentous bacteria with mycelia formation througout their lifestyle, which also have a similarly high guanine-plus-cytosine content in their genomes and are generally found in related natural habitats that span from aquatic to terrestrial environments [51]. Within the *bacteria* domain, they exhibit one of the largest taxonomic units including 5 subclasses, 6 orders and 12 suborders as described by Barka *et al.* in 2016 [52] and

have genome sizes of up to 11.9 Mbp as reported for *Streptomyces bingchenggensis* [53]. Their high biosynthetic potential has been extensively investigated yielding the isolation of many bioactive secondary metabolites and the establishment of about two-thirds of all bioactive natural products with pharmaceutical importance [54]. Figure 4 shows a selection of antibiotics in clinical use that were originally reported from actinobacterial sources.

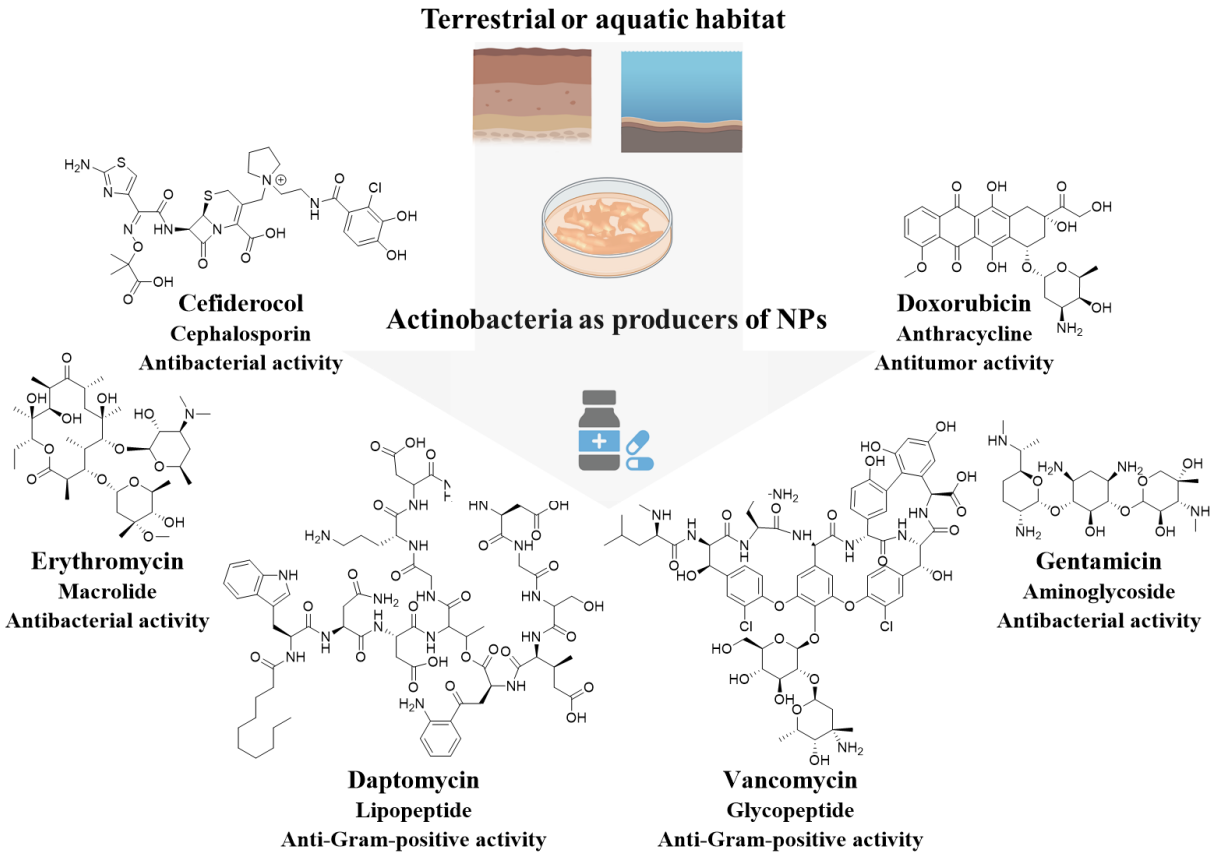


Figure 4: Selected antibiotics originally produced by actinobacteria that are in clinical use; cefiderocol as a cephalosporin with anti-Gram-positive and anti-Gram-negative activity [55]; doxorubicin as an anthracycline with antitumor activity [56]; erythromycin as a macrolidewith anti-Gram-positive and anti-Gram-negative activity [57]; daptomycin as a lipopeptide with anti-Gram-positive activity [58,59]; vancomycin as a glycopeptide with anti-Gram-positive activity [60]; gentamicin as an aminoglycoside with anti-Gram-positive and anti-Gram-negative activity [61].

1.3 Biosynthesis of myxobacterial and actinobacterial natural products

Exemplified by the compounds presented in Sections 1.1.1 and 1.1.2, most known myxobacterial natural products are biosynthesized by multimodular enzymes of NRPS, PKS and hybrids thereof, which are often encoded in large biosynthetic gene clusters (BGCs). These BGCs are consequently of immense interest for modern NP research [62,63]. Besides that, increasingly more ribosomally synthetized and post-translationally modified peptides and terpenes are discovered to be produced by myxobacteria, thus further expanding the structural variety. In the following, these four types of assembly lines are explained to give an insight in natural product biosynthesis.

NRP biosynthesis

Although PKS/NRPS hybrid BGCs are more commonly found in myxobacteria, several solely NRPS-related secondary metabolites were reported, such as cystobactamids [25], chloromyxamides [32], argyrins [64], myxovalargins [65], vioprolides [66] and thiangazoles [67]. In general, NRPs are products of an assembly line of modules producing a short peptide through catalysis of peptide bond formation between certain substrates such as both proteinogenic and non-proteinogenic amino acids. This biosynthesis consists of three stages: 1. substrate selection and loading, 2. peptide formation and elongation as well as 3. peptide release. Each assembly line consists of different modules that are specific for an amino acid and each module involves a certain number of domains with dedicated functions [68].

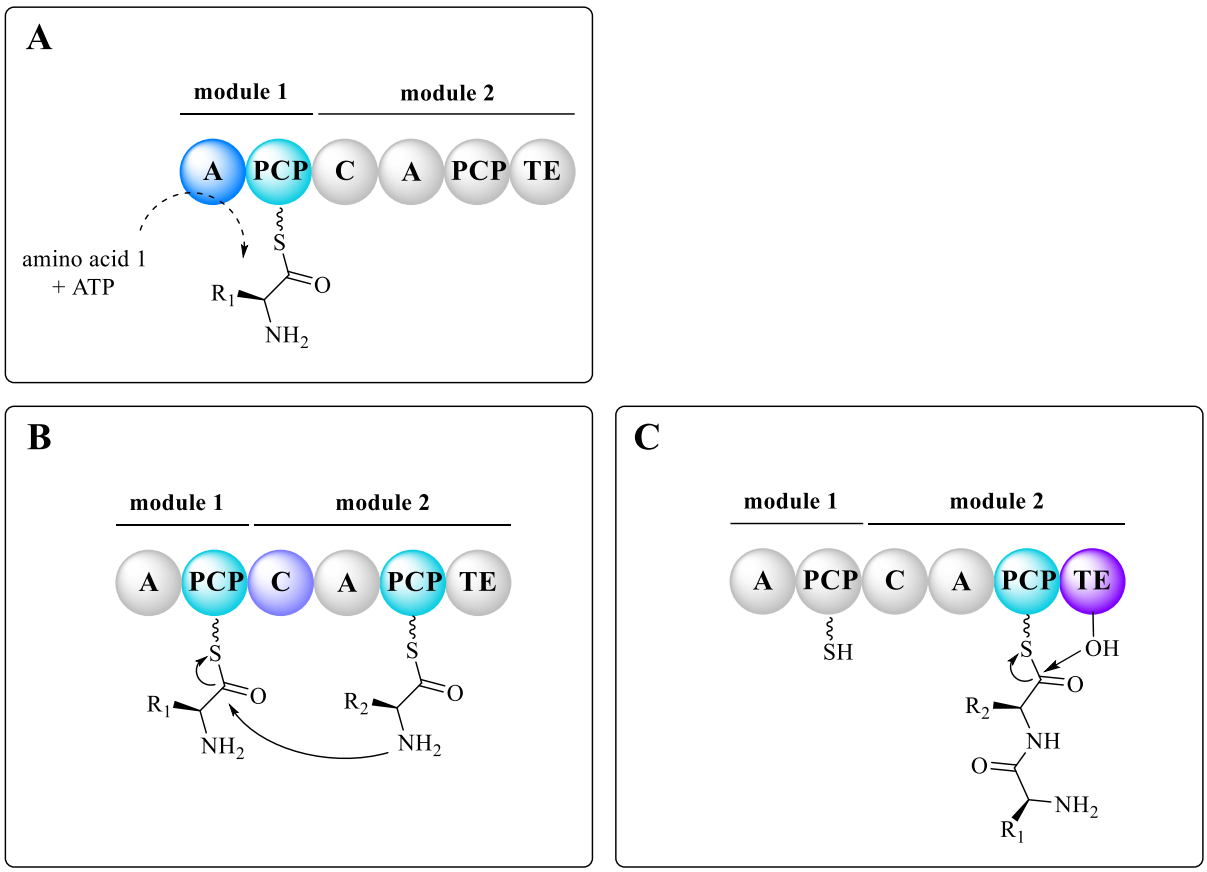


Figure 5: Schematic illustration of an NRPS assembly line with A) substrate selection and loading stage catalyzed by A domain, B) peptide formation and elongation stage catalyzed by C domain and C) peptide release stage catalyzed by TE domain; abbreviations are explained in the section below.

During the first stage the primary substrate is activated with adenosine triphosphate (ATP) by the adenylation domain (A), which forms an adenosylmonophosphatidyl (AMP) ester, which is then loaded on the 4'-phospho-pantethine sidechain of the peptidyl carrier protein (PCP). Subsequently, this building block is delivered by the PCP domain to the substrate-specific and highly stereo selective condensation (C) domain, where the actual peptide formation between two building blocks is catalyzed. The elongation of the peptide chain is carried out through repetition of the steps of activation, delivery and condensation of building blocks between adjacent modules. Finally, the peptide is released via the thioesterase (TE) domain, which cleaves off the thioester bond between the PCP domain by formation of an ester bond and the final substrate is released through hydrolysis or cyclization. Additional modification of the structure is enabled by optional domains like formylation, cyclization, oxidation, reduction, epimerization and methylation domains [69].

PK biosynthesis

Polyketide synthases are multi-domain enzymes that produce chemical scaffolds featuring polyketide functions. Three types of polyketide synthases have been reported, whereby type I PKS are large, highly modular proteins, type II PKS are aggregates of monofunctional proteins that act iteratively and type III PKS lack acyl carrier protein (ACP) domains. Known myxobacterial PKS products are soraphen [70], jerangolid [71], sorangicin [72], stigmatellin [73] and spirangien [74]. Hereafter, the type I PKS assembly line as producer of long, usually linear carbon chains with alcohol, olephine or methyl functions is explained. The corresponding biosynthesis consists again of three stages including loading, elongation/extension and termination through release of the substrate. Every module correlates with the addition of one dicarboxylic acid building block and contains domains that each perform a specific step in the condensation and modification of each added monomer [69].

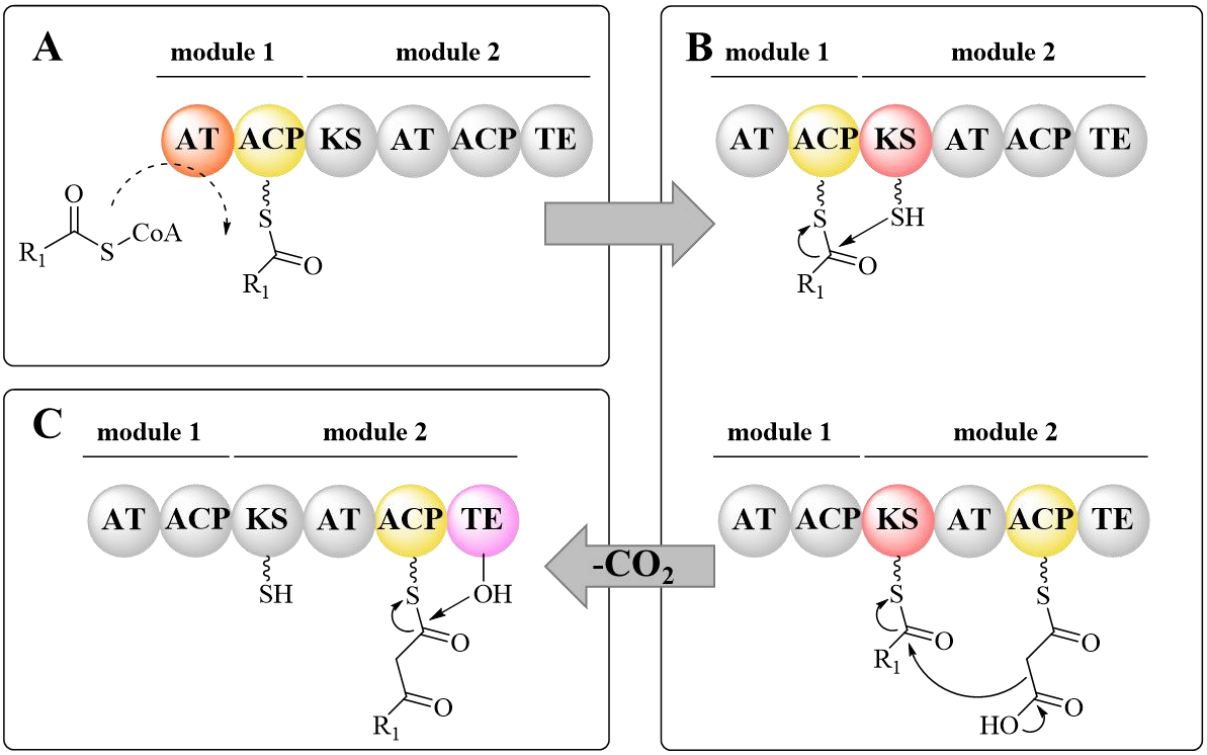


Figure 6: Schematic illustration of a PKS assembly line with A) loading stage catalyzed by AT domain, B) elongation and extension stage catalyzed by KS domain and C) release stage catalyzed by TE domain; abbreviations are explained in the section above.

During the first stage the substrate, usually acetyl-CoA or a derivative, is loaded on an ACP domain of the starter module via an acyltransferase (AT) domain Subsequently, the next building block, commonly malonyl-CoA or a derivative, is selected by the AT domain of the adjacent module that loads the building block onto the ACP domain of its module and the ketosynthase (KS) domain catalyzes the decarboxylative Claisen condensation between the two

building blocks. Now, the ACP bound elongated polyketide chain can repeatedly be extended by more elongation modules until the final substrate is released via a TE domain by thioester bond cleavage either through hydrolysis to obtain a free acid or an intramolecular ester formation resulting in a macrolactone. In addition, further optional domains can be part of an elongation module to derivatize the β-ketone moiety, *e.g.* ketoreductase (KR) to receive an alcohol forming a stereo center, dehydratase (DH) for double bond formation, methyl transferase (MT), epimerization (E) and enoyl reductase (ER) to obtain an alkane. The substrate might also be halogenated, oxidized, glycosylated or cyclized during the overall synthesis [75].

RiPP biosynthesis

RiPPs are ribosomally synthetized and post-translationally modified peptides that are also reported from myxobacteria like fulvocins [76], xanthacins [77], crocagins [78], cittilins [48] and myxarylin [31]. Throughout the RiPP biosynthesis, ribosomes firstly translate messenger RNA into a precursor peptide that usually consists of a core peptide and a preceding leader peptide that enables enzymatic processing as well as an optional recognition sequence at the C-terminus, which is used for excision and cyclization. Afterwards, the precursor peptide undergoes extensive post-translational modification and the resulting modified precursor peptide is finally released via proteolysis of the leader peptide and then exported as a mature peptide. Post-translational modifications include, among others, (cyclo-) dehydration, prenylation, epimerization, oxidation, decarboxylation, acetylation, methylation and cyclization [79].

Terpene biosynthesis

Plants in particular are the most common origin for terpenes. Nevertheless, several terpenoid-type secondary metabolites of myxobacterial origin like indiacens [80], salimyxins [81], cystodienoic acid [82], enhygromic acid [83], aurachins [84] and sandacrabins [29] have been published throughout the past decades. Unlike most other bacteria including actinobacteria, which produce terpenoids via the methylerythritol phosphate (MEP) pathway, in myxobacteria these compounds are mainly produced via the mevalonic acid (MVA) pathway. Both the MEP and MVA pathways make use of the activated building blocks isopentenyl diphosphate (IPP) and its isomer 3,3-dimethylallyldiphosphate (DMAPP). The terpene biosynthesis starts by joining together several building blocks of IPP and/or DMAPP to form a linear precursor such as the C_{10} monoterpene precursor Geranyl diphosphate (GPP), the C_{15} sesquiterpene precursor farnesyldiphosphate (FPP), the C_{20} diterpene precursor geranylgeranyl diphosphate (GGPP) or

longer C_{30} triterpene and C_{40} tetraterpene percursors that can be used during the biosynthetic pathway along with further polyterpenoids [85]. Subsequently, the linear terpenoid precursors are cyclized by terpene cyclases using carbocation chemistry to generate the mature terpene scaffolds, which are finally modified to yield mature products.

1.4 Mining approaches for natural product discovery

Nowadays, three main strategies are applied to yield new promising NPs, commonly named as the bioactivity-guided approach, genome mining and metabolome mining (Figure 7).The oldest and most common strategy is the bioactivity guided approach, which was used in a broad range of applications since the early 1900s. Its main advantage is the possibility to get a direct hint on the presence of bioactive components from any crude extract without the prior purification of single compounds. Nevertheless, frequent obstacles are the loss of activity after fractionation or even a synergistic effect between multiple compounds causing the observed activity. Furthermore, this approach is restricted to the available panel of test organisms and choice of the bioassays, which empowers the metabolome and genome mining approaches with the advancement of analytical instruments and methods to uncover previously unseen molecules.

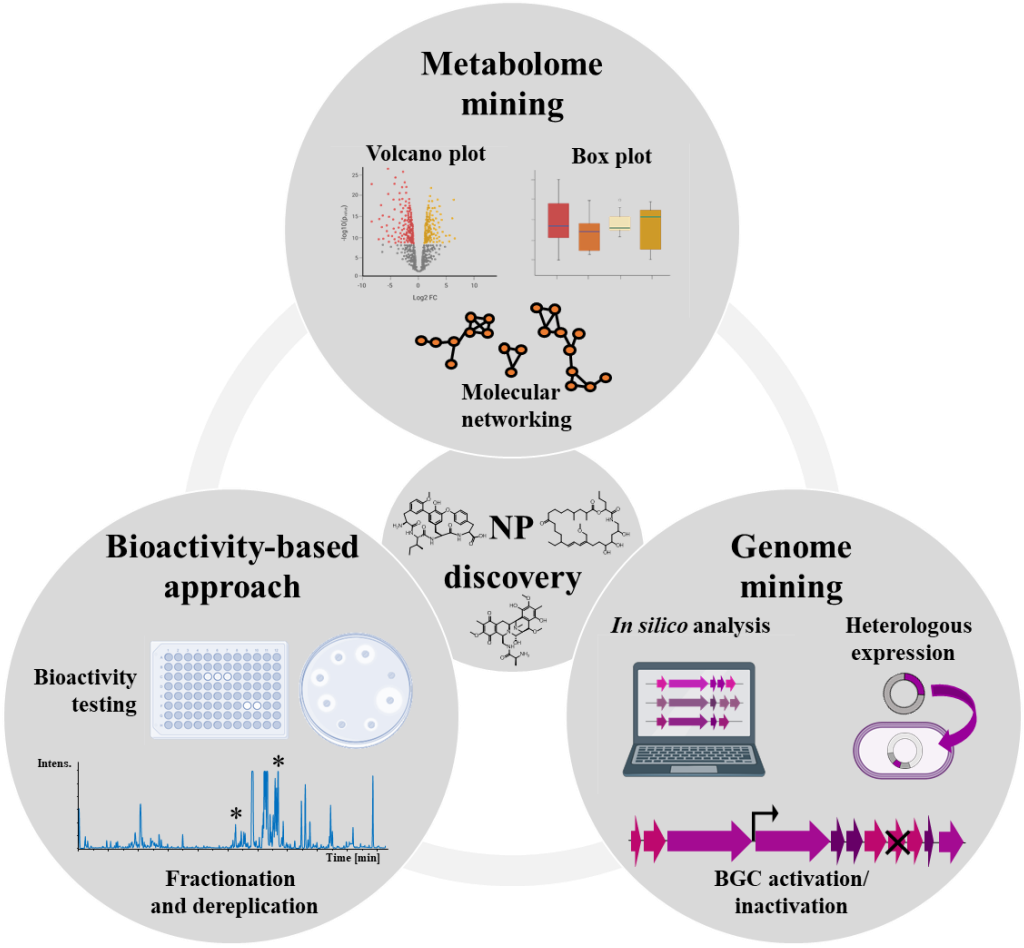


Figure 7: Approaches for natural product (NP) discovery: metabolome mining using statistical techniques visualized by volcano plots, box-whiskers plots and molecular networking; bioactivity-based approach using bioactivity testing, fractionation and dereplication; genome mining using in silico analysis, heterologous expression and biosynthetic gene cluster (BGC) activation or inactivation.

1.4.1 Metabolome mining approach for NP discovery

It has been found that a tremendous discrepancy between the genome-inscribed production capacity and actually known secondary metabolites exists for many producing organisms [86]. This means that many putatively produced secondary metabolites have not yet been detected nor identified and isolated. To approach this problem, metabolome mining as an analytics-centered approach enabling direct analysis of the detectable secondary metabolomes of producer organisms using *e.g.* mass spectrometry. Thus, a multitude of new secondary metabolites can potentially be uncovered, some of which could also exhibit interesting biological properties. In order to increase chances for the discovery of truly novel scaffolds, it is crucial to apply suitable filtering principles and perform comprehensive dereplication to direct efforts for NP isolation, structural elucidation and bioactivity characterization towards unknown candidates. For example, in a metabolomic-taxonomic study ~2300 myxobacterial strains from the HIPS/HZI in-house library revealed the distribution of both known and previously unidentified metabolites detected under laboratory cultivation conditions, thereby enabling large-scale comparison of production profiles in relation to myxobacterial taxonomy [33]. A clear correlation between taxonomic distance and the production of distinct secondary metabolite families was shown, thereby emphasizing that chances of discovering novel metabolites are much greater by examining strains from new genera rather than from additional representatives within the same genus or even species. This method paved the way for successful identification, purification and structure elucidation of a new myxobacterial secondary metabolite specific to the myxobacterial genus *Sorangium*, which was named rowithocin, featuring an uncommon phosphorylated polyketide scaffold [33].

Another encouraging example for the success of a metabolome mining strategy is the isolation and structure elucidation of two new natural products named fulvuthiacenes A and B from *Myxococcus fulvus* MCy9280 [87]. The fulvuthiacene biosynthetic pathway was characterized in detail and the evolutionary link between the fulvuthiacene and previously described myxothiazol hybrid PKS/NRPS gene clusters was established. This was enabled via modern analytical methods permitting in-depth analysis of the secondary metabolome in combination with a statistics-based filtering of relevant MS signals related to myxobacterial metabolism. Here, structural novelty can be anticipated typically by using MS/MS fragmentation data and database-assisted spectral networking such as Global Natural Products Social Molecular Networking (GNPS) [88]. Since its appearance spectral networking has been increasingly used to place single analytes into context by establishing their relationship to other molecules in the same sample or in public databases through spectral similarity.

These state-of-the-art analytical techniques were successfully applied in Chapter 2 to yield three structurally interesting novel NPs, the identification of a known derivative belonging to the nocardamine family and to highlight the extended biosynthetic potential of one single strain that was activated through altered cultivation conditions.

1.4.2 Genome mining approach for NP discovery

Genome mining is based on the in-depth analysis of the genome data of the organism of interest in order to highlight BGCs for the production of natural products, targeted gene inactivations or BGC activations, *e.g*. through promotor insertion. Differential profiling of the metabolome data of the wildtype and generated mutant strains can then be used in order to pinpoint a new natural product [41,89]. In particular, the latter is a crucial point, since even well studied myxobacterial strains like *M xanthus* on the one hand already yielded several different secondary metabolites, but on the other hand for 13 out of 18 predicted BGC pathways no NPs could be connected [90]. This discrepancy might be due to non-functionality of some BGCs or because of so called cryptic BGCs that produce insufficient compound yields, but might be potentially accessible for BGC activation using a promotor. Paradigmatically, the cryptic and structurally unique sandarazols were produced by *Sandaracinus* sp. after genetic activation of the respective BGC, which artificially induced the biosynthesis of these plasmid-encoded natural products [8].

Further examples for the successful identification and purification of bioactive myxobacterial NPs using the genome mining technique are the pyxidicyclines and the alkylpyrones. Panter *et al.* identified the pyxidicyclines via a self-resistance-guided screening approach, which made use of the co-location of candidate BGCs and genetically encoded determinants for host self-resistance to discover the potent topoisomerase inhibitors [91]. Hug *et al.* discovered the uncommon alkylpyrones as type III PKS products resulting from a pathway employing small homodimeric enzymes producing molecules with immense structural and functional diversity [42].

In Chapters 3 and 4, the genome mining approach paved the way for the discovery of a novel myxobacterial siderophore as well as the revival of the already known mycoplanecins and their development towards potential drugs against the causative agent of tuberculosis.

1.5 Comprehensive NMR analysis for *de novo* structure elucidation of NPs

Nuclear magnetic resonance (NMR) spectroscopy is a method commonly used to either verify a tentative structural proposal or to *de novo* determine the structure of an unknown organic compound. The requirement for this technique is the presence of an isotope with a nuclear spin different from zero namely a value of ±1/2, thus possible for nuclei with an odd number of protons or neutrons, *e.g.* ^{1}H, ^{13}C, ^{15}N, ^{19}F and ^{31}P. For simplification purposes, the ^{1}H-experiment is used to explain the basic principle of NMR. The ^{1}H proton of a molecule exhibits its own magnetic field via the spin of the positively charged nucleus. These nuclei spins have a random orientation without the influence of an external magnetic field and align parallel or antiparallel when placed into an external magnetic field. Thereby, the parallel orientation is the energetically favorable α-state and the antiparallel oriented spins are in the energetically unfavorable β-state respectively. When a radio frequency (rf) pulse reaches the energy difference between these two states, it can be used for excitation of the nuclei in the sample from α- to β-state (Figure 8A). During the spin transition back from β- to α-state, a specific/characteristic amount of energy is released, which can be detected as nuclear magnetic resonance in a free induction decay (FID) as a function of time. Ultimately, the final NMR spectrum is obtained as a function of frequency after a mathematical transformation known as Fourier transformation (FT) (Figure 8B). In the resulting ^{1}H-spectrum, the peaks appear at different frequencies in the spectrum due to their different local environment and are referenced to a fixed reference frequency, *e.g.* of tetramethylsilane (TMS) or of a deuterated solvent. The reference frequency of a specific ^{1}H-atom in the sample with respect to the reference signal is termed chemical shift and expressed in parts per million (ppm), whereby electronegative surrounding atoms expose a deshielding effect and thus cause a more downfield shift (higher ppm values) of this specific proton. Furthermore, the integration of peaks gives evidence about how many chemically equivalent protons of the compound generate this peak. Moreover, the number of neighboring protons has an influence on the peak splitting due to the spin-spin coupling phenomenon resulting in different multiplicity following the N+1 rule, so an adjacent methylene group would split the affected proton into a triplet. The distance between the split peaks can be used for coupling constant determination, which among others usually enables the assignment of conformational preferences of certain moieties of a compound [92].

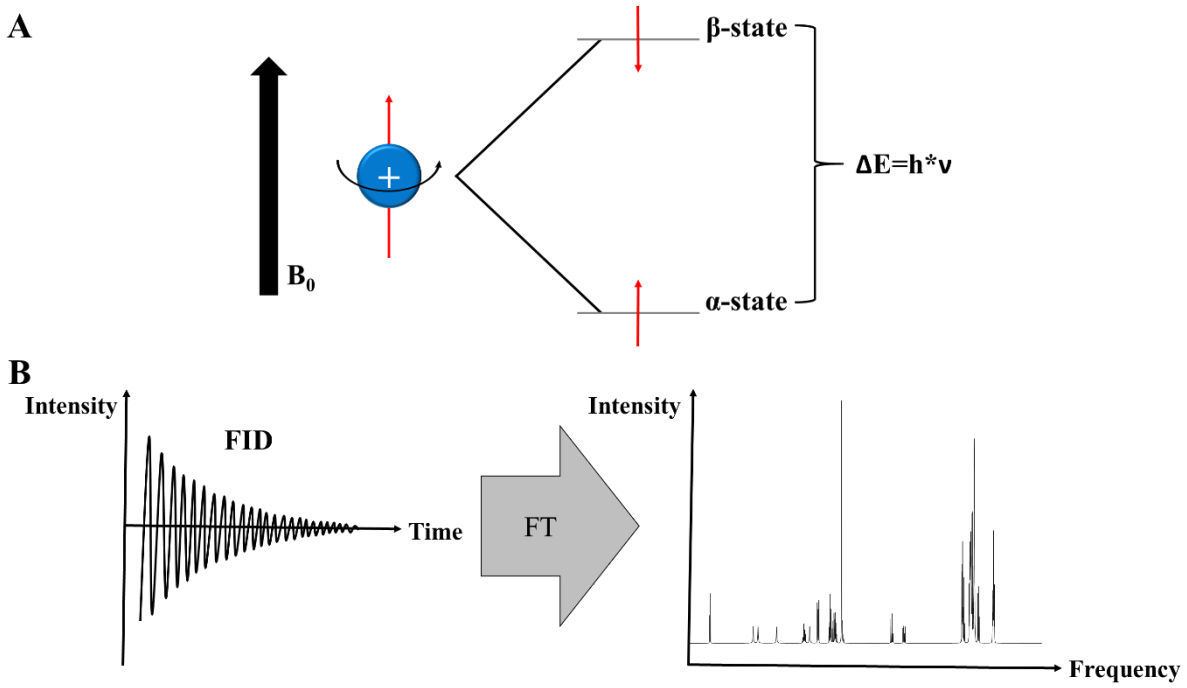


Figure 8: A) schematic illustration of the principle of NMR spectroscopy showing a nuclear spin that is oriented parallel to the external magnetic field B_0 and can be excited from α- to β-state by a specific radio frequency pulse to subsequently emit energy during spin transition from β- to α-state that is recorded as a free induction decay (FID); B) fourier transformation (FT) of the FID results in the NMR spectrum; ΔE=energy difference, h= Planck's constant, ν=frequency.

For a successful *de novo* structure elucidation a combination of different 1D and 2D NMR experiments is used to pinpoint a structural suggestion. Most commonly, ^{1}H- and ^{13}C-spectra are used as 1D experiments due to their presence in NP scaffolds with a sufficient amount of the respective NMR active isotope, which already enable an overview about the presence of various functional groups as part of the molecule. Nevertheless, the assignment is even more facilitated and accurate by utilization of 2D NMR experiments. ^{1}H-^{1}H-correlated spectroscopy (COSY) experiments mainly show ^{3}J-correlations between two protons of adjacent carbon moieties [93], whereas ^{1}H-^{13}C-heteronuclear single quantum coherence (HSQC) or heteronuclear multiple quantum coherence (HMQC) spectra exhibit cross peaks for ^{1}J-correlations of directly bound nuclei, in this case which proton is connected to which carbon atom. ^{1}H-^{13}C-heteronuclear multiple bond correlation (HMBC) experiments uncover ^{3}J-correlations between a proton and a carbon atom, but frequently also ^{2}J- and ^{4}J-couplings can be observed. Comprehensive NMR data sets comprising ^{1}H-, ^{13}C- (in case of sufficient amount of compound), ^{1}H-^{1}H-COSY, ^{1}H-^{13}C-HSQC and ^{1}H-^{13}C-HMBC experiments enabled the *de novo* structure elucidation presented in Chapters 2-4, whereby an additional ^{1}H-^{15}N-HMBC experiment was necessary in Chapter 3 to propose the structure. Furthermore, in many cases a total correlation spectroscopy (TOCSY) experiment gives additional information about which

signals belong to the same spin system and nuclear Overhauser enhancement spectroscopy (NOESY) and rotating frame Overhauser enhancement spectroscopy (ROESY) experiments use the nuclear Overhauser effect to obtain a 3D model of the molecule via spatial proximity [94].

The advantage of NMR spectroscopy is not only the gain of knowledge about the structure of a hitherto unknown compound, but also the ability to be used for structure verification, *e.g.* as a starting point of a structure-activity-relationship (SAR) study for substances with a promising bioactivity. Furthermore, in contrast to bioactivity testing or derivatization reactions for stereo center determination [95], NMR is a nondestructive method. Thus, the experiments can be repeated under different conditions to enhance the quality of the spectra as well as with different solvents in case of overlapping signals and the compound is still available for follow up reactions. Thereby, protic solvents like deuterated water and methanol are able to exchange acidic protons of the analyte, which causes the respective signal to disappear, *e.g.* in the case of the protons of hydroxyl and amino groups. This problem can be circumvented by the usage of aprotic solvents like deuterated chloroform, dimethylformamide, acetonitrile and dimethylsulfoxide.

Due to its many advantages, NMR analysis is ubiquitously present in the majority of scientific publications from the natural products field. In addition to the very standardized experiments described above using isolated pure compounds, NMR also provides opportunities for a variety of customized/unconventional methods, such as ^{1}H spectra of crude cell mass extracts as reported for the aurachins, which led to the discovery of the aurachins as myxobacterial quinolone antibiotics produced by *Stigmatella aurantiaca* [96] or even LC-coupled approaches such as the LC-solid phase extraction (SPE)-NMR-MS workflow ultimately leading to the structure-guided isolation of the macyranones produced by *Cystobacter fuscus* [97].

1.6 Outline of this thesis

This thesis is mainly focused on the discovery of structurally novel myxobacterial NPs by exploitation of the biosynthetic potential of the *Sorangiineae* strain MSr11367, whereas an additional part deals with the revival of mycoplanecins as potent antitubercular agents. Both metabolome and genome mining approaches are utilized to prioritize interesting target compounds or peculiar BGCs ultimately leading to the purification of new secondary metabolites and examination of their physico-chemical properties and biosynthesis (Figure 9). Furthermore, the benefit of focusing on one promising strain to purify several NPs from different structural classes instead of using several different strains for isolation of only few secondary metabolites each is demonstrated.

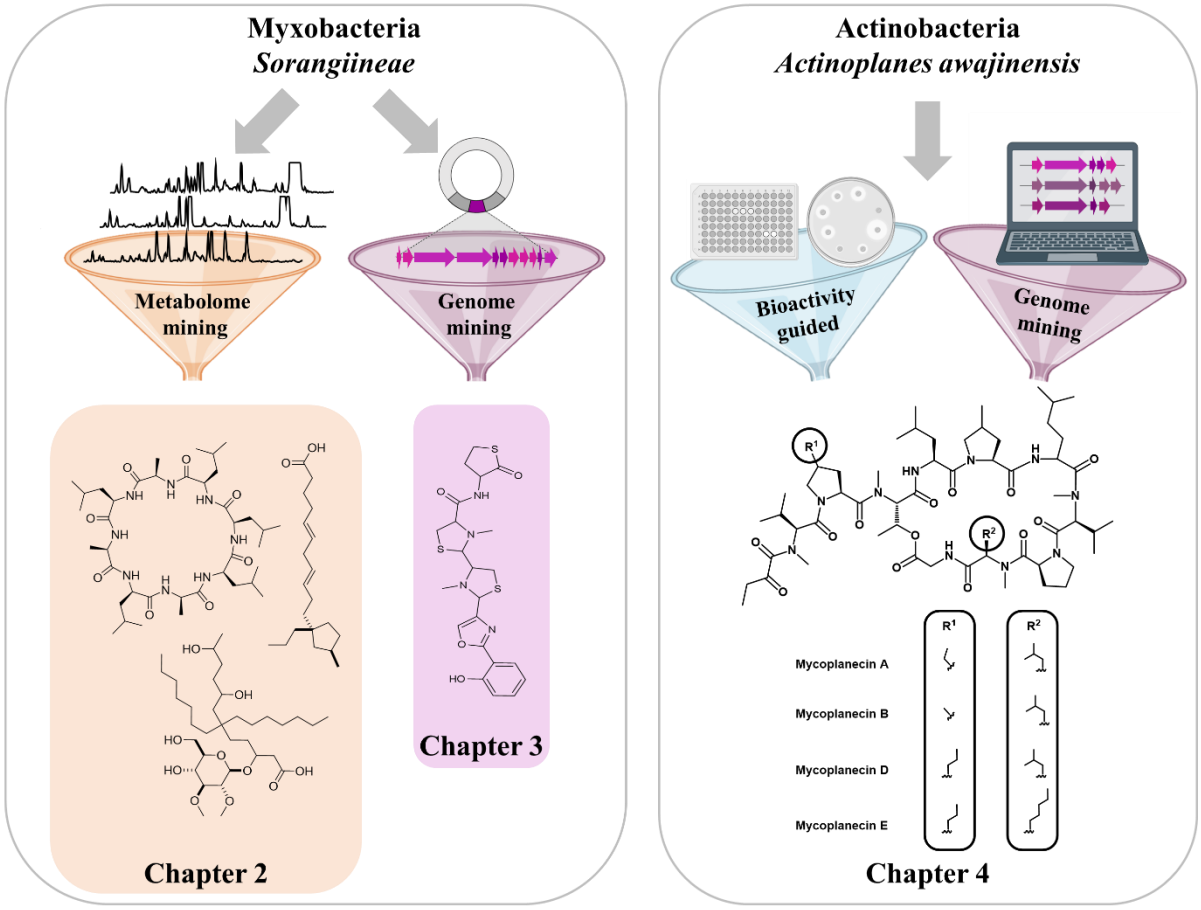


Figure 9: Overview of the three chapters of this thesis showing the respective metabolome, genome and bioactivity guided strategies, which resulted in the discovery of four novel structural classes and new derivatives belonging to the mycoplanecin family.

In Chapter 2, a comprehensive analysis of the myxobacterial metabolome is carried out to correlate the effects of altered cultivation conditions to the secondary metabolite profile of the novel *Sorangiineae* strain MSr11367 with the overall goal to trigger natural product biosynthesis. Metabolome mining is performed via statistical analysis for cultivation using

different temperatures (18 °C, 30 °C, 37 °C), illuminations (white light, darkness), additives in sub-lethal doses (rifampicin, chloramphenicol, trimethoprim, lanthanum chloride, ethanol and soil extract) as well as oxygen limitation. In-depth analysis of these data is performed to uncover effects on up- or downregulation of metabolites that are already produced in the standard setup. Conditions that mimic the natural habitat of myxobacteria like the addition of soil extract are evaluated in terms of the number of highly reproducible signals corresponding to NPs that are new in comparison to standard cultivation. These undertakings lead to the discovery of three structurally novel secondary metabolites as revealed in Chapter 2—myxolutamid, pentacyclic acid and glucodiolic acid—including their purification and *de novo* structure elucidation.

The biosynthetic potential of strain MSr11367 is further exploited in Chapter 3 using an orthogonal genome mining approach to search for peculiar BGCs potentially featuring intriguing biosynthesis pathways that might be responsible for the production of novel NPs. Here, the investigation centers around a coelibactin-like nonribosomal peptide synthetase (NRPS) gene cluster due to the presence of several modules that are predicted to be responsible for the incorporation and subsequent cyclization of cysteines, which foreshadows the biosynthesis of a promising compound. Heterologous expression of the respective BGC in the host *Myxococcus xanthus* DK1622 is employed to facilitate the purification of a novel myxobacterial siderophore named sorangibactin followed by *de novo* structure determination using comprehensive 1D and 2D NMR experiments including a ^{1}H-^{15}N-HMBC experiment. Hereby an unusual C-terminal γ-thiolactone moiety is revealed as a striking structural feature, which is not represented by an apparent module in the BGC. These findings allow creation of a first hypothesis for the biosynthetic pathway and pave the way for further biosynthesis studies of such unusual features, which might potentially help to also solve the to date still tentative structure of coelibactin, which is solely predicted by bioinformatics analyses with unassigned stereochemistry.

Chapter 4 deals with the mycoplanecins, a family of antitubercular compounds discovered in the early 1980 from actinobacteria. The “reawakening” of these compounds by discovery of the respective BGC by mining (actinobacterial) genomes for 4-methylproline pathways is described, utilizing the insight that alkylproline moieties proved to play an important role in DnaN binding and metabolic stability for the structurally related griselimycins. The project includes the purification of the four most abundant mycoplanecins and performing a full set of 1D and 2D NMR experiments for each derivative to prove/confirm the respective structures and to amend previous studies where only 1D spectra of the MPs are shown. Moreover, re-evaluation of the bioactivity is reported, revealing a 167 fold better bioactivity against the

attenuated *M. tuberculosis* strain H37Ra and a 20 folds stronger bioactivity against the virulent *M. tuberculosis* strain H37Rv of mycoplanecin A in comparison to griselimycin. In addition, microscale thermophoresis is performed in order to verify DnaN as the cellular target of mycoplanecins and to obtain the binding affinities of several derivatives. A matching biosynthesis proposal is also devised as part of Chapter 4 and experiments towards drug target validation are described.

The three chapters outlined above should serve to manifest the successful application of metabolome and genome mining strategies to discover structurally novel NPs on the one hand and to find and resolve biosynthetic pathways of bioactive compounds on the other hand. Structure elucidation using NMR plays an essential role in each of these studies, either in terms of *de novo* structure determination or for structure verification.

1.7 References

1. Oliveira, D.M.P. de; Forde, B.M.; Kidd, T.J.; Harris, P.N.A.; Schembri, M.A.; Beatson, S.A.; Paterson, D.L.; Walker, M.J. Antimicrobial Resistance in ESKAPE Pathogens. *Clin. Microbiol. Rev.* **2020**, *33*, e00181-19, doi:10.1128/CMR.00181-19.
2. Stephens, L.J.; Werrett, M.V.; Sedgwick, A.C.; Bull, S.D.; Andrews, P.C. Antimicrobial innovation: a current update and perspective on the antibiotic drug development pipeline. *Future Med. Chem.* **2020**, *12*, 2035–2065, doi:10.4155/fmc-2020-0225.
3. Moloney, M.G. Natural Products as a Source for Novel Antibiotics. *Trends in Pharmacological Sciences* **2016**, *37*, 689–701, doi:10.1016/j.tips.2016.05.001.
4. Nature Publishing Group. All natural. *Nat. Chem. Biol.* **2007**, *3*, 351, doi:10.1038/nchembio0707-351.
5. Bérdy, J. Bioactive Microbial Metabolites. *J Antibiot* **2005**, *58*, 1–26, doi:10.1038/ja.2005.1.
6. Schmitz, R. Friedrich Wilhelm Sertürner and the discovery of morphine. *Pharm. Hist.* **1985**, *27*, 61–74.
7. 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, doi:10.5598/imafungus.2011.02.01.12.
8. Panter, F.; Bader, C.D.; Müller, R. The Sandarazols are Cryptic and Structurally Unique Plasmid-Encoded Toxins from a Rare Myxobacterium*. *Angew. Chem. Int. Ed Engl.* **2021**, *60*, 8081–8088, doi:10.1002/anie.202014671.

9. Gonthier, I.; Rager, M.-N.; Metzger, P.; Guezennec, J.; Largeau, C. A di-O-dihydrogeranylgeranyl glycerol from Thermococcus S 557, a novel ether lipid, and likely intermediate in the biosynthesis of diethers in Archæa. *Tetrahedron Letters* **2001**, *42*, 2795–2797, doi:10.1016/S0040-4039(01)00305-7.
10. K. K. CHEN; H. JENSEN; A. L. CHEN. THE PHYSIOLOGICAL ACTION OF THE PRINCIPLES ISOLATED FROM THE SECRETION OF BUFO ARENARUM. *J Pharmacol Exp Ther* **1933**, *49*, 1–13.
11. Demain, A.L.; Fang, A. The Natural Functions of Secondary Metabolites. In *History of modern biotechnology*; Fiechter, A., Ed.; Springer: Berlin, London, 2000; pp 1–39, ISBN 978-3-540-67793-2.
12. Newman, D.J.; Cragg, G.M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83*, 770–803, doi:10.1021/acs.jnatprod.9b01285.
13. Bhawna Chopra; Ashwani Kumar Dhingra. Natural products: A lead for drug discovery and development. *Phytotherapy Research* **2021**, *35*, 4660–4702, doi:10.1002/ptr.7099.
14. 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, doi:10.1002/ange.200703993.
15. Kim, B.S.; Moon, S.S.; Hwang, B.K. Isolation, identification, and antifungal activity of a macrolide antibiotic, oligomycin A, produced by Streptomyces libani. *Can. J. Bot.* **1999**, *77*, 850–858, doi:10.1139/b99-044.
16. Wrótniak-Drzewiecka, W.; Brzezińska, A.J.; Dahm, H.; Ingle, A.P.; Rai, M. Current trends in myxobacteria research. *Ann. Microbiol.* **2015**, doi:10.1007/s13213-015-1104-3.
17. Bhat, S.; Ahrendt, T.; Dauth, C.; Bode, H.B.; Shimkets, L.J. Two lipid signals guide fruiting body development of *Myxococcus xanthus*. *mBio* **2014**, *5*, 13, doi:10.1128/mBio.00939-13.
18. Pal, S.; Sharma, G.; Subramanian, S. Complete genome sequence and identification of polyunsaturated fatty acid biosynthesis genes of the myxobacterium *Minicystis rosea* DSM 24000T. *BMC Gen.* **2021**, *22*, 655, doi:10.1186/s12864-021-07955-x.
19. Han, K.; Li, Z.-F.; Peng, R.; Zhu, L.-P.; Zhou, T.; Wang, L.-g.; Li, S.-g.; Zhang, X.-b.; Hu, W.; Wu, Z.-H.; et al. Extraordinary expansion of a Sorangium cellulosum genome from an alkaline milieu. *Scientific Reports* **2013**, *3*, 2101, doi:10.1038/srep02101.

20. 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, doi:10.1038/nbt1354.
21. Garcia, R.; Gerth, K.; Stadler, M.; Dogma, JR, Irineo J.; Muller, R. Expanded phylogeny of myxobacteria and evidence for cultivation of the 'unculturables'. *Mol. Phylogenet. Evol.* **2010**, *57*, 878–887, doi:10.1016/j.ympev.2010.08.028.
22. Mohr, K.I. Diversity of Myxobacteria-We Only See the Tip of the Iceberg. *Microorganisms* **2018**, *6*, doi:10.3390/microorganisms6030084.
23. Herrmann, J.; Fayad, A.A.; Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **2017**, *34*, 135–160, doi:10.1039/C6NP00106H.
24. Garcia, R.; Krug, D.; Müller, R. Discovering natural products from myxobacteria with emphasis on rare producer strains in combination with improved analytical methods. In *Methods in Enzymology*: *Complex Enzymes in Microbial Natural Product Biosynthesis* Part A: Overview Articles and Peptides; David A. Hopwood, Ed., 2009; pp 59–91.
25. 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, doi:10.1002/anie.201409964.
26. Gerth, K.; Bedorf, N.; Irschik, H.; Höfle, G.; Reichenbach, H. The soraphens: a family of novel antifungal compounds from *Sorangium cellulosum* (Myxobacteria). I. Soraphen $A_{1\alpha}$: fermentation, isolation, biological properties. *J. Antibiot.* **1994**, *47*, 23–31, doi:10.7164/antibiotics.47.23.
27. Irschik, H.; Jansen, R.; Höfle, G.; Gerth, K.; Reichenbach, H. The corallopyronins, new inhibitors of bacterial RNA synthesis from Myxobacteria. *J. Antibiot.* **1985**, *38*, 145–152.
28. Kunze, B.; Sasse, F.; Wieczorek, H.; Huss, M. Cruentaren A, a highly cytotoxic benzolactone from Myxobacteria is a novel selective inhibitor of mitochondrial F-1-ATPases. *FEBS Lett.* **2007**, *581*, 3523–3527, doi:10.1016/j.febslet.2007.06.069.
29. Bader, C.D.; Panter, F.; Garcia, R.; Tchesnokov, E.P.; Haid, S.; Walt, C.; Spröer, C.; Kiefer, A.F.; Götte, M.; Overmann, J.; et al. Sandacrabins - Structurally Unique Antiviral RNA Polymerase Inhibitors from a Rare Myxobacterium. *Chemistry – A European Journal* **2022**, *28*, e202104484, doi:10.1002/chem.202104484.

30. Nadmid, S.; Plaza, A.; Lauro, G.; Garcia, R.; Bifulco, G.; Müller, R. Hyalachelins A-C, unusual siderophores isolated from the terrestrial myxobacterium *Hyalangium minutum*. *Org. Lett.* **2014**, *16*, 4130–4133, doi:10.1021/ol501826a.
31. Hug, J.J.; Frank, N.A.; Walt, C.; Šenica, P.; Panter, F.; Müller, R. Genome-guided Discovery of the First Myxobacterial Biarylitide Myxarylin Reveals Distinct C–N Biaryl Crosslinking in RiPP Biosynthesis **2021**, doi:10.20944/preprints202111.0467.v1.
32. 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, doi:10.1002/anie.201808028.
33. 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, doi:10.1038/s41467-018-03184-1.
34. Bader, C.D.; Panter, F.; Müller, R. In depth natural product discovery - Myxobacterial strains that provided multiple secondary metabolites. *Biotechnol. Adv.* **2020**, *39*, 107480, doi:10.1016/j.biotechadv.2019.107480.
35. Yang, Y.-J.; Singh, R.P.; Lan, X.; Zhang, C.-S.; Li, Y.-Z.; Li, Y.-Q.; Sheng, D.-H. Genome editing in model strain *Myxococcus xanthus* DK1622 by a Site-Specific Cre/loxP Recombination System. *Biomolecules* **2018**, *8*, doi:10.3390/biom8040137.
36. Goldman, B.S.; Nierman, W.C.; Kaiser, D.; Slater, S.C.; Durkin, A.S.; Eisen, J.A.; Ronning, C.M.; Barbazuk, W.B.; Blanchard, M.; Field, C.; et al. Evolution of sensory complexity recorded in a myxobacterial genome. *Proc. Nat. Acad. Sci. USA* **2006**, *103*, 15200–15205, doi:10.1073/pnas.0607335103.
37. 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, doi:10.1128/AEM.02863-07.
38. 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, doi:10.1002/jlac.1993199301200.
39. *Cittilins: Bicyclic Isotrityrosines from Myxococcus xanthus*: *24th General meeting German Chemists' Society;* Trowitzsch-Kienast, W., Ed., Hamburg, Germany, Sep. 5-11, 1993.

40. 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, doi:10.7164/antibiotics.36.1150.
41. 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, doi:10.1002/anie.201106305.
42. 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, doi:10.1007/s10295-018-2105-6.
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, doi:10.1073/pnas.0606039103.
44. Hyun, H.; Lee, S.; Lee, J.S.; Cho, K. Genetic and Functional Analyses of the DKxanthene Biosynthetic Gene Cluster from Myxococcus stipitatus DSM 14675. *Journal of Microbiology and Biotechnology* **2018**, *28*, 1068–1077, doi:10.4014/jmb.1802.02045.
45. Kunze, B.; Bedorf, N.; Kohl, W.; Höfle, G.; Reichenbach, H. Myxochelin A, a new iron-chelating compound from *Angiococcus disciformis* (Myxobacterales). Production, isolation, physico-chemical and biological properties. *J. Antibiot.* **1989**, *42*, 14–17, doi:10.7164/antibiotics.42.14.
46. Gerth, K.; Irschik, H.; Reichenbach, H.; Trowitzsch, W. The myxovirescins, a family of antibiotics from *Myxococcus virescens* (Myxobacterales). *J Antibiot (Tokyo)* **1982**, *35*, 1454–1459, doi:10.7164/antibiotics.35.1454.
47. Maurer, C.K.; Fruth, M.; Empting, M.; Avrutina, O.; Hossmann, J.; Nadmid, S.; Gorges, J.; Herrmann, J.; Kazmaier, U.; Dersch, P.; et al. Discovery of the first small-molecule CsrA-RNA interaction inhibitors using biophysical screening technologies. *Future Med. Chem.* **2016**, *8*, 931–947, doi:10.4155/fmc-2016-0033.
48. 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, doi:10.1021/acschembio.0c00430.
49. Wink, J. Special Issue: "Actinobacteria and Myxobacteria-Important Resources for Novel Antibiotics". *Microorganisms* **2020**, *8*, doi:10.3390/microorganisms8101464.

50. Manivasagan, P.; Venkatesan, J.; Sivakumar, K.; Kim, S.-K. Pharmaceutically active secondary metabolites of marine actinobacteria. *Microbiol. Res.* **2014**, *169*, 262–278, doi:10.1016/j.micres.2013.07.014.
51. Servin, J.A.; Herbold, C.W.; Skophammer, R.G.; Lake, J.A. Evidence Excluding the Root of the Tree of Life from the Actinobacteria. *Mol Biol Evol* **2008**, *25*, 1–4, doi:10.1093/molbev/msm249.
52. Essaid Ait Barka; Parul Vatsa; Lisa Sanchez; Nathalie Gaveau-Vaillant; Cedric Jacquard; Hans-Peter Klenk; Christophe Clément; Yder Ouhdouch; Gilles P. van Wezel. Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol. Mol. Biol. Rev.* **2016**, *80*, 1–43, doi:10.1128/MMBR.00019-15.
53. Zhou, Z.; Gu, J.; Li, Y.-Q.; Wang, Y. Genome plasticity and systems evolution in Streptomyces. *BMC Bioinformatics* **2012**, *13*, 1–17, doi:10.1186/1471-2105-13-S10-S8.
54. Tiwari, K.; Gupta, R.K. Rare actinomycetes: A potential storehouse for novel antibiotics. *Crit. Rev. Biotechnol.* **2012**, *32*, 108–132, doi:10.3109/07388551.2011.562482.
55. Sato, T.; Yamawaki, K. Cefiderocol: Discovery, Chemistry, and In Vivo Profiles of a Novel Siderophore Cephalosporin. *Clin Infect Dis.* **2019**, *69*, S538-S543, doi:10.1093/cid/ciz826.
56. Tacar, O.; Sriamornsak, P.; Dass, C.R. Doxorubicin: an update on anticancer molecular action, toxicity and novel drug delivery systems. *J. Pharm. Pharmacol.* **2013**, *65*, 157–170, doi:10.1111/j.2042-7158.2012.01567.x.
57. Wolfe, A.D.; Hahn, F.E. Erythromycin: Mode of Action. *Science* **1964**, *143*, 1445–1446, doi:10.1126/science.143.3613.1445.
58. Debono, M.; Abbott, B.J.; Molloy, R.M.; Fukuda, D.S.; Hunt, A.H.; Daupert, V.M.; Counter, F.T.; Ott, J.L.; Carrell, C.B.; Howard, L.C. Enzymatic and chemical modifications of lipopeptide antibiotic A21978C: the synthesis and evaluation of daptomycin (LY146032). *J Antibiot* **1988**, *41*, 1093–1105, doi:10.7164/antibiotics.41.1093.
59. Baltz, R.H. Daptomycin: mechanisms of action and resistance, and biosynthetic engineering. *Curr. Opin. Chem. Biol.* **2009**, *13*, 144–151, doi:10.1016/j.cbpa.2009.02.031.
60. Griffith, R.S. Introduction to Vancomycin. *Clinical Infectious Diseases* **1981**, *3*, S200-S204, doi:10.1093/clinids/3.supplement_2.s200.
61. Marvin J. Weinstein, George M. Luedemann, Edwin M. Oden, Gerald H. Wagman, Jean Pierre Rosselet, Joseph A. Marquez, Carmine T. Coniglio, William Charney, Hershel L.

Herzog, Jack Black. Gentamicin,1 a New Antibiotic Complex from Micromonospora. *J. Med. Chem.* **1963**, *6*, 463–464, doi:10.1021/jm00340a034.

62. 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.
63. Baltz, R.H. Natural product drug discovery in the genomic era: realities, conjectures, misconceptions, and opportunities. *J. Ind. Microbiol.* **2018**, *46*, 281–299, doi:10.1007/s10295-018-2115-4.
64. Pogorevc, D.; Tang, Y.; Hoffmann, M.; Zipf, G.; Bernauer, H.S.; Popoff, A.; Steinmetz, H.; Wenzel, S.C. Biosynthesis and Heterologous Production of Argyrins. *ACS Synth. Biol.* **2019**, *8*, 1121–1133, doi:10.1021/acssynbio.9b00023.
65. Irschik, H.; Gerth, K.; Kemmer, T.; Steinmetz, H.; Reichenbach, H. The Myxovalargins, new peptide antibiotics from *Myxococcus fulvus* (Myxobacterales) I. Cultivation, isolation, and some chemical and biological properties. *J. Antibiot.* **1983**, *36*, 6–12.
66. Yan, F.; Auerbach, D.; Chai, Y.; Keller, L.; Tu, Q.; Hüttel, S.; Glemser, A.; Grab, H.A.; Bach, T.; Zhang, Y.; et al. Biosynthesis and Heterologous Production of Vioprolides: Rational Biosynthetic Engineering and Unprecedented 4-Methylazetidinecarboxylic Acid Formation. *Angew. Chem. Int. Ed. Engl.* **2018**, *57*, 8754–8759, doi:10.1002/anie.201802479.
67. Jansen, R.; Kunze, B.; Reichenbach, H.; Jurkiewicz, E.; Hunsmann, G.; Höfle, G. Antibiotics from Gliding Bacteria, XLVII. Thiangazole: A novel inhibitor of HIV-1 from *Polyangium spec. Liebigs Ann. Chem.* **1992**, *1992*, 357–359, doi:10.1002/jlac.199219920163.
68. Marahiel, M.A. A structural model for multimodular NRPS assembly lines. *Nat. Prod. Rep.* **2016**, *33*, 136–140, doi:10.1039/c5np00082c.
69. Weissman, K.J.; Müller, R. A brief tour of myxobacterial secondary metabolism. *Bioorg. Med. Chem.* **2009**, *17*, 2121–2136, doi:10.1016/j.bmc.2008.11.025.
70. Wenzel, S.C.; Williamson, R.M.; Grünanger, C.; Xu, J.; Gerth, K.; Martinez, R.A.; Moss, S.J.; Carroll, B.J.; Grond, S.; Unkefer, C.J.; et al. On the biosynthetic origin of methoxymalonyl-acyl carrier protein, the substrate for incorporation of “glycolate” units into ansamitocin and soraphen A. *J. Am. Chem. Soc.* **2006**, *128*, 14325–14336, doi:10.1021/ja064408t.
71. Julien, B.; Tian, Z.Q.; Reid, R.; Reeves, C.D. Analysis of the ambruticin and jerangolid gene clusters of *Sorangium cellulosum* reveals unusual mechanisms of polyketide biosynthesis. *Chem. Biol.* **2006**, *13*, 1277–1286, doi:10.1016/j.chembiol.2006.10.004.

72. 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, doi:10.1002/cbic.201000313.
73. 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, doi:10.1016/s0167-4781(99)00041-x.
74. Frank, B.; Knauber, J.; Steinmetz, H.; Scharfe, M.; Blöcker, H.; Beyer, S.; Müller, R. Spiroketal polyketide formation in *Sorangium*: Identification and analysis of the biosynthetic gene cluster for the highly cytotoxic spirangienes. *Chem. Biol.* **2007**, *14*, 221–233, doi:10.1016/j.chembiol.2006.11.013.
75. Wenzel, S.C.; Müller, R. Myxobacterial natural product assembly lines: fascinating examples of curious biochemistry. *Nat. Prod. Rep.* **2007**, *24*, 1211–1224, doi:10.1039/b706416k.
76. Hirsch, H.J.; Tsai, H.; Geffers, I. Purification and effects of fulvocin C, a bacteriocin from *Myxococcus fulvus Mx f16*. *Arch. Microbiol.* **1978**, *119*, 279–286.
77. Mccurdy, H.D.; MacRae, T.H. Xanthacin. A bacteriocin of Myxococcus xanthus fb. *Can. J. Microbiol.* **1974**, *20*, 131–135, doi:10.1139/m74-021.
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, doi:10.1002/anie.201612640.
79. 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, doi:10.1039/C2NP20085F.
80. 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, doi:10.1021/np300288b.
81. 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, doi:10.1002/cbic.201300268.

82. 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, doi:10.1038/ja.2015.8.
83. Tomura, T.; Nagashima, S.; Yamazaki, S.; Iizuka, T.; Fudou, R.; Ojika, M. An Unusual Diterpene-Enhygromic Acid and Deoxyenhygrolides From a Marine Myxobacterium, *Enhygromyxa s*p. *Mar. Drugs* **2017**, *15*, doi:10.3390/md15040109.
84. Pistorius, D.; Li, Y.; Sandmann, A.; Müller, R. Completing the puzzle of aurachin biosynthesis in *Stigmatella aurantiaca* Sg a15. *Mol. Biosyst.* **2011**, *7*, 3308–3315, doi:10.1039/c1mb05328k.
85. 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, doi:10.3762/bjoc.15.283.
86. Wenzel, S.C.; Müller, R. Myxobacteria—''microbial factories' for the production of bioactive secondary metabolites. *Mol. Biosyst.* **2009**, *5*, 567–574, doi:10.1039/b901287g.
87. 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, doi:10.1021/acschembio.8b00948.
88. 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, doi:10.1038/nbt.3597.
89. 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, doi:10.1039/c3np70127a.
90. Wenzel, S.C.; Muller, R. Myxobacteria—''microbial factories' for the production of bioactive secondary metabolites. *Mol. Biosyst.* **2009**, *5*, 567–574, doi:10.1039/b901287g.
91. 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, doi:10.1039/C8SC01325J.
92. *Annual Reports on NMR Spectroscopy;* Webb, G.A., Ed.; Academic Press, 2011, ISBN 0066-4103.

93. Aue, W.P.; Bartholdi, E.; Ernst, R.R. Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J. Chem. Phys.* **1976**, *64*, 2229–2246, doi:10.1063/1.432450.
94. Schirmer, R.E.; Noggle, J.H.; Davis, J.P.; Hart, P.A. Determination of molecular geometry by quantitative application of the nuclear Overhauser effect. *Adv. Ceram. Mater.* **1970**, *92*, 3266–3273, doi:10.1021/ja00714a005.
95. 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, doi:10.1016/0040-4039(96)00484-4.
96. Kunze, B.; Höfle, G.; Reichenbach, H. The aurachins, new quinoline antibiotics from myxobacteria: production, physico-chemical and biological properties. *J. Antibiot.* **1987**, *40*, 258–265.
97. Keller, L.; Plaza, A.; Dubiella, C.; Groll, M.; Kaiser, M.; Muller, R. Macyranones: Structure, Biosynthesis, and Binding Mode of an Unprecedented Epoxyketone that Targets the 20S Proteasome. *Adv. Ceram. Mater.* **2015**, *137*, 8121–8130, doi:10.1021/jacs.5b03833.

Chapter 2

2. Aggravated Cultivation of Myxobacteria Stimulates Secondary Metabolism

Christine Walt[1,2], Chantal D. Bader[1,2], Daniel Krug[1,2] and Rolf Müller[1,2,3,*]

[1]Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Saarland University, Campus E8 1, 66123 Saarbrücken, Germany

[2]German Center for Infection Research (DZIF), Partner Site Hannover-Braunschweig, 38124 Braunschweig, Germany

[3]Helmholtz International Lab for anti-infectives Campus E8 1, 66123 Saarbrücken, Germany

[*]Author to whom correspondence should be addressed.

Contributions and acknowledgements

Author's effort:

The author significantly contributed to the concept of this study as well as the design and performance of experiments and subsequent interpretation of the results. Cultivation, statistical analysis, compound purification and *de novo* structure elucidation was carried out by the author. Moreover, the author was significantly involved in conceiving and writing of this manuscript.

Contributions by others:

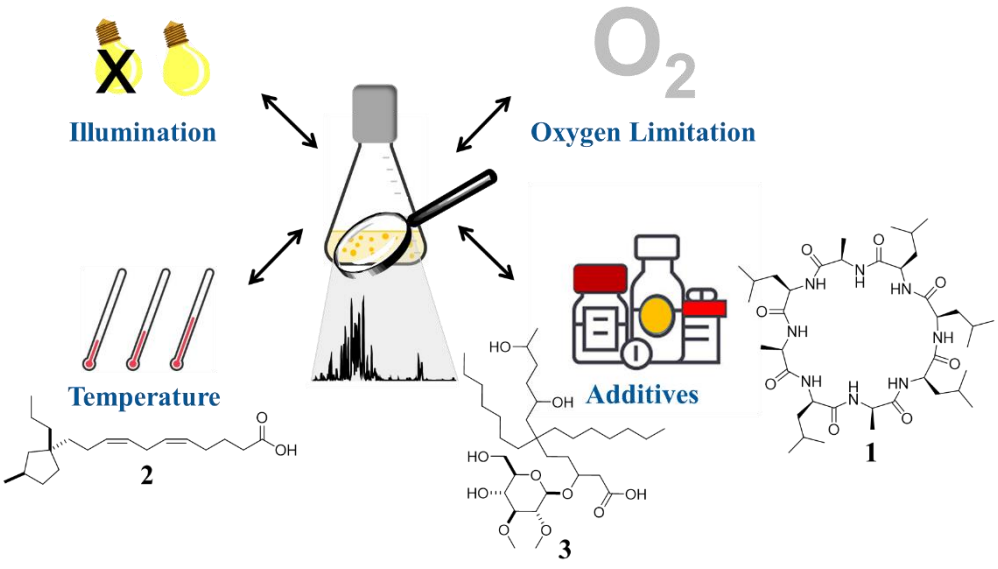
Chantal Bader contributed to the concept creation of the study, verified the *de novo* structure elucidation and significantly contributed to conceiving and writing the manuscript. Jake Haeckl created the reproducibility heatmap. Daniel Krug and Rolf Müller contributed to conception and supervision of this study with subsequent editing and proofreading of the manuscript.

Abstract

Cultivation of actinomycetes under aggravated conditions has proven a fruitful method to trigger the expression of cryptic biosynthetic gene clusters (BGCs). This method, however, was rarely applied to explore the hidden biosynthetic potential of myxobacteria. In this manuscript we present a comprehensive analysis of the myxobacterial strains MSr11367 metabolome using aggravated cultivation conditions. Statistical analysis of cultivation under different temperatures (18 °C, 30 °C, 37 °C), illuminations (white light, exclusion of light), influence of additives in sub-lethal concentrations (rifampicin, chloramphenicol, trimethoprim, $LaCl_3$, ethanol and soil extract) and oxygen limitation were investigated. The most significant influence on feature abundance was found for cultivation with oxygen limitation (73 upregulated and 39 downregulated features), whereby the highest amount of new features was detected for supplementation of soil extract (542 features) or $LaCl_3$ (352 features) and for oxygen limitation (325 features). Thereby the most unique features were observed for soil extract (243 features) and $LaCl_3$ (98 features) supplementation. The reproducibility of features was determined best for addition of soil extract, lanthanum chloride and with oxygen limitation. Finally, we were able to structurally elucidate three unprecedented secondary metabolites–myxolutamid, pentacyclic acid and glucodiolic acid–enabled by cultivation with addition of soil extract, cultivation at 18 °C and lanthanum chloride supplementation, respectively.

Keywords: myxobacteria; natural products; metabolome mining; macrocycles, peptides

TOC



2.1 Introduction

Myxobacteria are recognized as a promising source for natural products (NPs), known to encode one of the highest numbers of biosynthetic gene clusters (BGCs) responsible for production of NPs with chemically diverse structures and intriguing biological activities [1–3]. Under standard cultivation conditions applied in most NP labs, many of those BGCs were found to have no assigned NPs and thus are cryptic, which might be due to poor expression titers, very low or even completely abolished production rates or hindered detection of the corresponding NPs [4,5]. Commonly, access to these cryptic BGCs is attempted by their genetic manipulation introducing a promotor for activation of the BGC in the native host [6]. This technique, however, is laborious, low throughput and limited to genetically amenable organisms. Many myxobacteria, especially those belonging to underexplored genera—as the strain presented in this study—are lacking genetic tools to perform BGC activation [7]. An alternative approach to upregulate secondary metabolism, is cultivation under altered conditions, which—in addition to its feasibility for organisms that cannot be genetically manipulated—more globally affects secondary metabolism instead of targeting only one specific BGC. This method was already successfully applied to actinomycetes by Tanaka *et al.*, where Rifampin supplementation as well as further condition variations led to an overproduction of antibiotics like actinorhodin, actinomycin, streptomycin, formycin and erythromycin by activation of their BGCs [8]. Furthermore, the one strain many compounds (OSMAC) approach often utilizes additives such as organic solvents [9] and heavy metals [10] or regulation of oxygen supply [11] and pH , whereby the latter has an impact on permeability of cell [12]. For addition of rare earth elements in sub-lethal concentration exemplarily activation of cryptic gene clusters in *Streptomyces coelicolor* [13] was reported. The addition of rare earth elements deems especially impactful on soil dwelling bacteria—to which also myxobacteria belong—due to low levels of rare earth elements present in their natural environment but often lacking in common cultivation media. Variations in myxobacterial cultivation setups are rather uncommon. Seventeen selected studies describing the discovery of new NPs in the last 5 years all apply moderate shaking at 30 °C and only slight variations in the media composition (SI) [3,6,14–28]. The antifungal soraphens, isolated from the *Sorangium cellulosum* strain So ce26, are one of the few examples, where modifications in temperature had been correlated with production rates in myxobacteria. An increase of the cultivation temperature to 40 °C resulted in complete depletion of their production [29]. Other "stressful" conditions such as constant illumination were found to induce starvation-like scenarios *e.g.* fruiting body formation and sporulation as reported for *Stigmatella aurantiaca* [30]. Inspired by those examples, we aimed to diversify the detectable metabolome

of the myxobacterial strain MSr11367 belonging to the suborder of *Sorangiineae*, which showed an impressive biosynthetic potential of 62 predicted BGCs, but was only found to produce a very limited number of metabolites under standard conditions. Variations in temperature, illumination, additives and oxygen supply were studied in an untargeted metabolomics approach and utilized to gain access to three unprecedented NPs, whose structures are presented herein.

2.2 Results and discussion

According to the prediction by antiSMASH [31], the myxobacterial strain MSr11367 comprises 62 BGCs including 15 polyketide synthase (PKS), 10 non-ribosomal peptide synthetase (NRPS), 8 terpene, 10 ribosomally synthesized and post-translationally modified peptide (RiPP), 15 hybrids thereof and 4 other BGCs (SI). This genomic potential should be exploited intensively by cultivation under aggravating cultivation conditions. In this study, we decided upon 12 different conditions including variations of temperature (18 °C, 30 °C, 37 °C), illumination (white light, exclusion of light), additives in sub-lethal concentrations (rifampicin, chloramphenicol, trimethoprim, lanthanum chloride, ethanol and soil extract) and oxygen supply. The three different temperatures were selected to cover temperature fluctuations in the natural habitat to reevaluate the optimum for production of NPs by this strain. Different illumination experiments were utilized as stimulation adapted to the extremes: cultivation under complete light exclusion to mimic soil conditions and permanent light exposure as long-term stimulation of the strain. Addition of rifampicin, trimethoprim and chloramphenicol were selected based on their different mode of action. Rifampicin is known to be a transcription inhibitor [32], whereas the mode of action of trimethoprim is inhibition of the folic acid metabolism [33] and chloramphenicol is a translation inhibitor [34]. Rare soil elements or soil extract was supplemented to mimic the conditions in nature. 2 % ethanol was added to create an altered environment for the bacteria without actually killing them. Reduction in shaking was performed to limited oxygen supply and lower shear forces on the production of NPs.

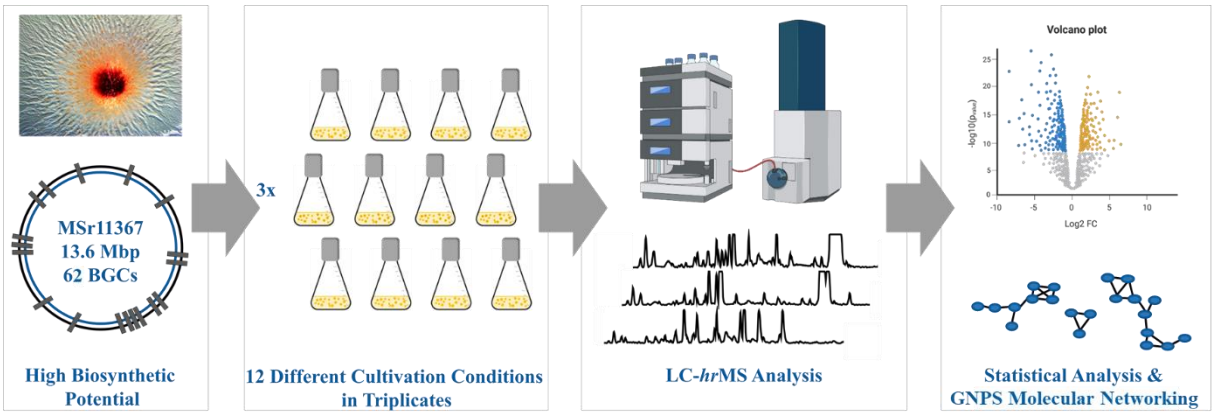


Figure 1: General workflow of metabolome mining approach using statistical analysis for selection of promising cultivation conditions under which cryptic BGCs are activated.

For the comprehensive metabolite analysis, biological triplicates and technical duplicates were investigated by LC-*hr*MS measurements with subsequent statistical analysis using MetaboScape and GNPS molecular networking (Figure 1). The associated impact on the secondary metabolite profile in terms of up- or downregulation of NP production was inspected, especially regarding reproducibility, number and yield of novel features.

Influence on secondary metabolism and reproducibility

After subtraction of blank features, we detected a total number of 2178 features in all samples, whereof about half (1157 features) were not present in the 30 °C cultivation. Among those, we first evaluated the influence of the different conditions by calculating features that are up- or down-regulated under said condition (Figure 2). For the two deviating temperature setups, a reduction in temperature was found to have a greater influence on feature abundance than increasing the temperature: 30 features were found to be upregulated and 61 downregulated. At 37 °C only 6 were upregulated and 42 downregulated. For different illuminations, cultivation with white light led to a greater change of feature abundance than with light exclusion: upregulation of 16 and downregulation of 71 features. Cultivation in the dark resulted in upregulation of 32 and downregulation of no features. Oxygen limitation caused 73 upregulated and 39 downregulated features. Among the addition of antibiotics in sub-lethal concentrations, supplementation of trimethoprim was found to have the most influence on feature abundance with upregulation of 36 and downregulation of 58 features. Addition of chloramphenicol resulted in upregulation of 26 and downregulation of 62 features, whereas rifampicin showed upregulation of 32 and downregulation of 33 features. For the remaining setups with supplementation, addition of soil extract showed the greatest change in feature abundance resulting in 55 upregulated and 62 downregulated features. Lanthanum chloride

supplementation led to 61 upregulated and 49 downregulated features and ethanol addition to 23 upregulated and 56 downregulated features (SI).

The up- and downregulation of features sheds light on the impact of the altered conditions on the features that are also present during standard cultivation. The highest number of upregulated features could be observed during cultivation with oxygen limitation. This indicates that the shaking speed of 180 rpm during standard cultivation might be too high for this strain to produce increased yields. On the contrary, the condition with the highest amount of downregulated features is cultivation under permanent illumination, which seems to be a rather stressful condition that decreases the amount of produced features. Finally, cultivation in the dark is the closest to standard cultivation, since no features have been downregulated and solely 32 features were upregulated. In conclusion, even with the maximum number of approximately 70 up- and downregulated features the impact of the altered cultivation conditions is rather moderate concerning features that are also present in the standard setup.

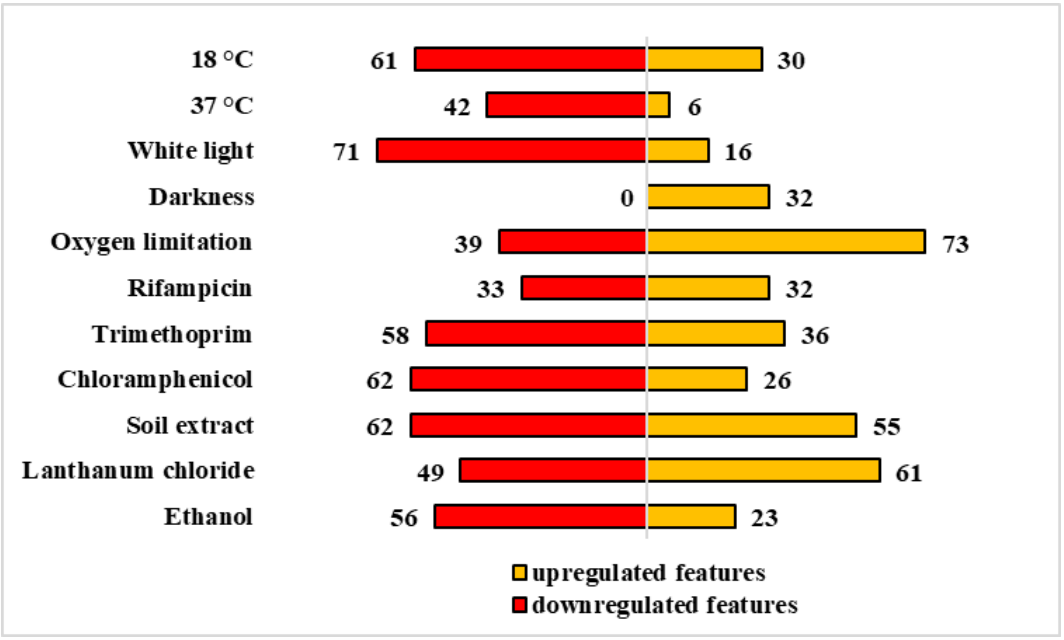


Figure 2: Detailed analysis among all cultivation conditions of up and down regulated features in comparison to standard cultivation.

To exploit the unprecedented biosynthetic potential of MSr11367, we focused on the features that are completely absent when cultivating under standard conditions. When examining those features, addition of soil extract (542 features), $LaCl_3$ supplementation (352 features) and cultivation with oxygen limitation (325 features) were found to yield the most metabolites absent under standard conditions. Furthermore, 256 new features were detected for chloramphenicol addition, 151 features for cultivation at 18 °C, 113 features for addition of

ethanol, 110 features for trimethoprim supplementation and 81 features for permanent illumination with white light as well as rifampicin addition in sub-lethal concentrations. The least effect on the presence of new features absent during standard cultivation was found for cultivation under light exclusion (28 features) and 37 °C cultivation (2 features) (Figure 3). Furthermore, the number of unique features that were only detected under one specific condition were determined to be the highest for soil extract supplementation (243 features) and $LaCl_3$ addition (98 features). Moreover, 69 unique features were observed for ethanol addition, 66 features for oxygen limitation, 59 features for rifampicin addition and 44 features for trimethoprim supplementation. Another 38 unique features were found for illumination with white light, 36 features for cultivation at 18 °C and 28 features for addition of chloramphenicol. Cultivation under light exclusion (6 features) and cultivation at 37 C (no unique features) turned out to be the least productive conditions for unique features (Figure 3).

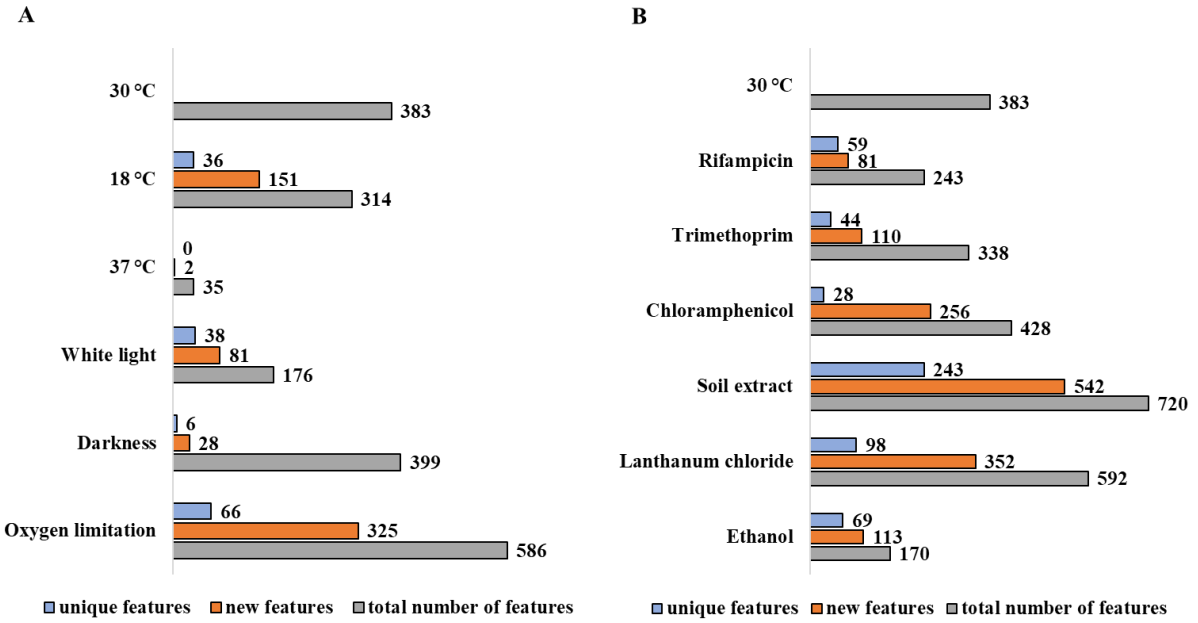


Figure 3: Detailed analysis among all cultivation conditions of total number of detected features, novel features in comparison to standard cultivation and features that are unique to the specific condition; A) physiological variations; B) different supplements; unique features detected under one specific cultivation condition highlighted in blue, new features in comparison to standard cultivation highlighted in orange and total number of detected features highlighted in gray.

Regarding evaluation of the temperature optimum, cultivation at lower temperatures seem to trigger the production of more novel features than during standard cultivation, while 37 °C led to solely 2 new features. It is worth noting that cultivation at 18 °C however is highly sensitive to contaminations, which might impede large scale cultivation for NP purification. Thus, 30 °C seems to remain the cultivation optimum regarding temperature. Cultivation in the dark shows high similarity to standard cultivation, whereas permanent illumination triggers production of

novel features. Nevertheless, white light favored contamination during cultivation as already observed for reduction of the temperature, making it a less advantageous approach for NP isolation. One of the most promising conditions investigated in this study is oxygen limitation, which seemed to have the highest impact on the metabolome among the three different physiological conditions. Among the addition of sub-lethal concentration of antibiotics, chloramphenicol shows the highest number of novel features followed by trimethoprim and rifampicin. The effect on the metabolome during cultivation with addition of ethanol is in the same range as antibiotic supplementation (113 features vs 81 for rifampicin and 110 for trimethoprim), whereby addition of chloramphenicol resulted in the highest number of new features (256 features) herein. Further supplements such as soil extract and $LaCl_3$ yielded the highest number of novel features. Summing up, the most promising results could be observed for conditions that try to better mimic the natural environment in soil such as adding soil extract, $LaCl_3$ and limiting oxygen intake.

Finally, we were also interested in calculating which features are unique for only one condition. The highest amount of new unique features were found for soil extract supplement and addition of lanthanum chloride with 243 and 98 unique features. This intriguingly high amount of 243 unique features only detectable under one condition supports a diversifying approach when aiming to explore the full secondary metabolome of a strain instead of focusing only on one condition.

Reproducibility

Since the retrieval of any new detected feature (replicability or reproducibility) is of utmost importance for NP purification, especially when working with a living organism, the reliable production of novel features was investigated in detail. For all cultivation conditions a statistical analysis of biological triplicates and technical duplicates was carried out to evaluate the reproducibility of each feature. The results of this calculation are visualized in the gradient bar chart below (Figure 4), whereby best reproducibility (6/6) of novel features is observed in the cultures supplemented with soil extract, $LaCl_3$ and using oxygen limitation.

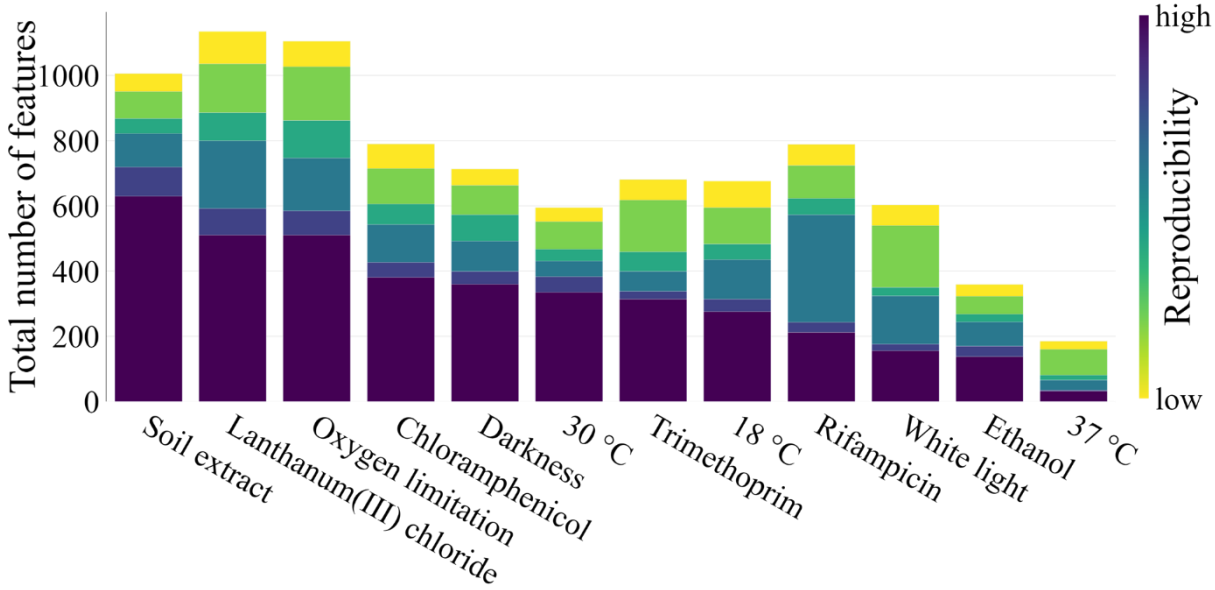


Figure 4: Statistical analysis of biological triplicates and technical duplicates for feature reproducibility across the different cultivation conditions with total amount of features arranged along the y-axis (ranging from dark blue (highest reproducibility 6/6) over green to yellow (lowest reproducibility 1/6)); supplementation with soil extract, lanthanum chloride and cultivation with oxygen limitation show the best reproducibility.

Starting with the variation of physiological conditions—especially for cultivation at higher temperatures—the number of features in comparison to standard cultivation decreases drastically, especially the number of highly reproducible features (6/6). Whereas cultivation at lower temperatures exhibits several reproducible features, but still fewer highly reproducible features compared to the standard setup. Light exclusion shows very similar numbers of highly reproducible features as the standard setup, while exhibiting few additional features with moderate reproducibility. Permanent illumination (white light) leads to a lower amount of highly reproducible features, while maintaining a similar amount of total features. Oxygen limitation exhibits a higher number of features with good or moderate reproducibility in comparison to standard cultivation. Coming to the cultivation with additives, supplementation of chloramphenicol shows a slightly higher amount of highly reproducible features than during standard cultivation and trimethoprim as well as rifampicin result in slightly fewer highly reproducible features. Cultivation with supplementation of ethanol results in a drastic decrease of features, especially highly reproducible features, whereas soil extract and lanthanum chloride addition show the highest total number of features with the best reproducibility for the majority of features across all conditions.

In conclusion, the cultivation under different conditions resulted in a rather moderate effect on the abundance of features already present during standard cultivation. Several conditions still yielded the production of new features with a promising reproducibility. Nevertheless, some

conditions revealed to be less appealing like cultivation in the dark, which is very similar to the standard setup with an overall comparable reproducibility of those features and little influence on production of new features. Thus cultivation under light exclusion is an unprofitable condition regarding isolation of novel features. Moreover, cultivation at 37 °C shows low reproducibility, less features in comparison to standard cultivation and rarely contributes novel features, which makes it an unfavorable condition for production of novel secondary metabolites. 18 °C seems more promising regarding novel features with good reproducibility, however as already mentioned contamination issues make it less appealing. Cultivation under white light yields in only few novel features with less reproducible features in comparison to standard cultivation, which leads to a deprioritization of this condition. The most promising physiologic condition seems to be cultivation under oxygen limitation showing high numbers of reproducible features and promising upregulation of unprecedented features. Among the antibiotics supplementation, the highest amount of novel features can be observed for addition of chloramphenicol with quite several more reproducible features in comparison to the standard setup, however rifampicin and trimethoprim supplementation resulted in less upregulated new molecules with a slightly lower amount of reproducible features. However, antibiotic supplementation revealed to be less appealing in comparison to other conditions. Especially rifampicin is known to have a high spontaneous mutation frequency for its target protein RNA polymerase, which has an impact on transcription and product profiles [35]. Furthermore, ethanol addition yielded in minor novel features with an overall unbeneficial reproducibility, which is not promising for purification of new NPs. Fortunately, $LaCl_3$ supplementation uncovers a beneficial potential for activation of new feature production with valuable reproducibility and above all, soil extract addition is the most promising condition with the highest number of novel features, with the highest amount of reproducible features and was prioritized for novel NP isolation besides optimization of the $LaCl_3$ condition for better yields.

Determination of target features

Using our untargeted approach for investigation of the metabolome of MSr11367, we first evaluated already known NPs and their production among the different cultivation conditions. The only previously described NPs we could detect in all conditions are the nocardamines, which is in line with the manually inspected antiSMASH predictions of known BGCs, highlighting the strains potential to discover new chemistry [36]. Besides some fully described congeners of this NP family, we detect dehydroxynocardamine B the structure of which was only proposed by Stubbendieck *et al.* through comparison of the fragmentation patterns and

predicted sum formulas to dehydroxynocardamine previously [37]. We herein present a full NMR data set of this derivative confirming its structure proposed by Stubbendieck *et al.*. Our cultivation setup showed rather low production rates for dehydroxynocardamine B (approximately 0.2 mg/L; SI). However, dehydroxynocardamine B was produced under standard conditions and rifampicin supplement all with medium reproducibility (Figure 5) and was purified from large scale cultivation at 30 °C. Furthermore, dehydroxynocardamine A was detected in several conditions (30 °C, darkness, chloramphenicol/rifampicin/soil extract addition, oxygen limitation, permanent illumination) as well as nocardamine (chloramphenicol/rifampicin/$LaCl_3$/soil extract addition, oxygen limitation, permanent illumination) (SI).

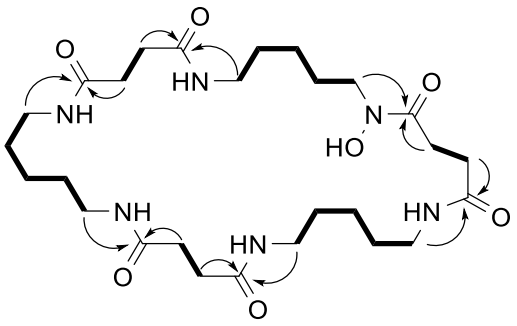


Figure 5: Chemical structure of dehydroxynocardamine B with most relevant COSY (bold) and HMBC (arrows) correlations.

After evaluation of known NPs, we prioritized three unprecedented NPs with additional striking characteristics like prominent increase in production through addition of soil extract for myxolutamid (**1**), temperature dependence for pentacyclic acid (**2**) and unique occurrence in one of the twelve cultivation conditions for glucodiolic acid (**3**) for isolation and structure elucidation.

Myxolutamid

The first target mass with *m/*z 779.5384 is produced under all conditions except for modification of the temperature (Figure 6). Especially, during cultivation in the dark, addition of chloramphenicol and soil extract the production is upregulated in comparison to the standard cultivation setup. Since, the reproducibility of new features is the most promising during cultivation with soil extract supplementation, this was the method of choice for purification of this target feature with a reproducibility of 6/6. After large scale cultivation of MSr11367 with addition of soil extract, subsequent separation steps of the crude extract and semi-preparative workup of respective fractions, the structure of the pure NP was determined using 1D and 2D NMR analysis and Marfey's-analysis.

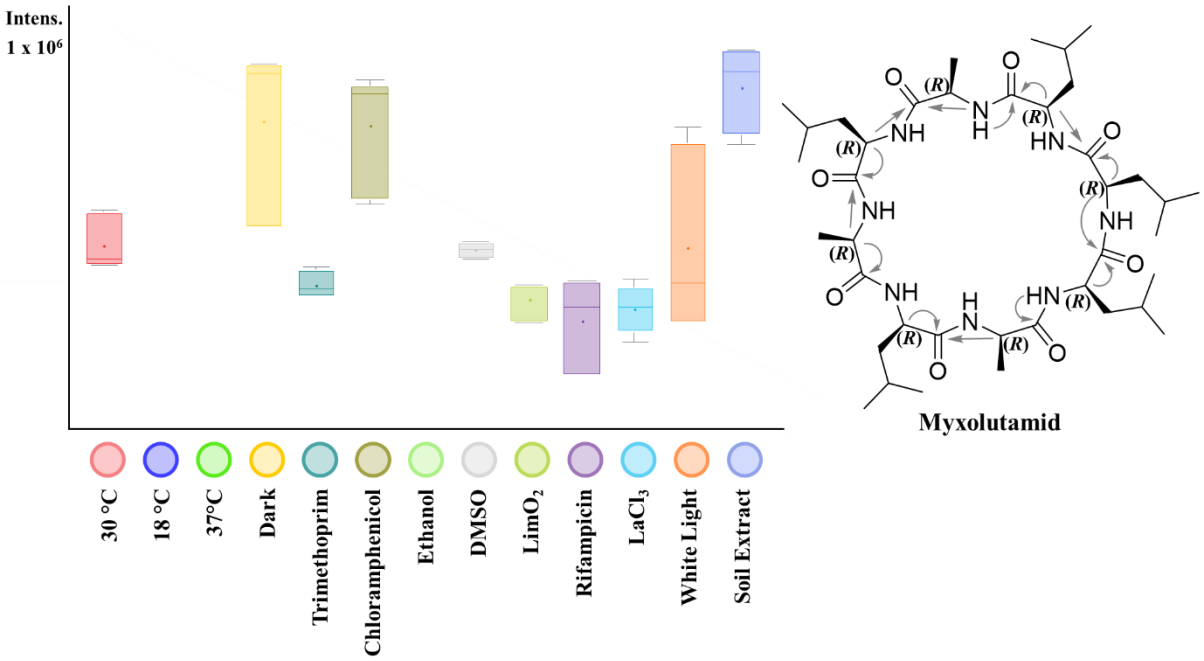


Figure 6: Box-and-whisker diagram showing intensity levels of m/z 779.5384 among the different cultivation conditions: myxolutamid (1) is produced under several conditions with best production through addition of soil extract (light purple); median and mean values and 25% and 75% quartiles illustrated within box with grey whiskers representing lower and upper extremes; relevant HMBC (gray arrows) correlations.

Structure elucidation of myxolutamid. 1D and 2D NMR experiments suggest a macrocycle comprising eight amino acids. Further, COSY and HMBC correlations uncover five leucines as well as three alanines with a predicted amino acid sequence via HMBC correlations from the respective α-protons to the adjacent carbonyl groups of Leu-Leu-Leu-Ala-Leu-Ala-Leu-Ala. Noteworthy, only one correlation from the α-proton at $\delta(^{1}H)$ = 4.06 ppm to the carbonyl group at $\delta(^{13}C)$ = 174.1 ppm was not detected. Yet, the determined sum formula, DBEs and number of amide protons as well as their integration of one lead to a cyclic scaffold. Additionally, the amino acid sequence is further supported by MS^2 fragmentation of linear derivatives of myxolutamid (SI). The stereochemistry of the amino acids was determined following Marfey's derivatization protocol and comparison of the retention times of the diastereomeres to reference substances. All three incorporated alanines and all five leucines were determined to be D-configured.

When inspecting congeners of **1** assigned via molecular networking using GNPS, several derivatives of smaller ring sizes or linear peptides were produced under different cultivation conditions, *e.g.* addition of soil extract, supplement of ethanol, addition of rifampicin, with light exclusion, with oxygen limitation as well as standard cultivation condition (Figure 7). Here, the MS^2 fragmentation behavior of linear congeners are specifically helpful to underpin the predicted amino acid sequence of **1**.

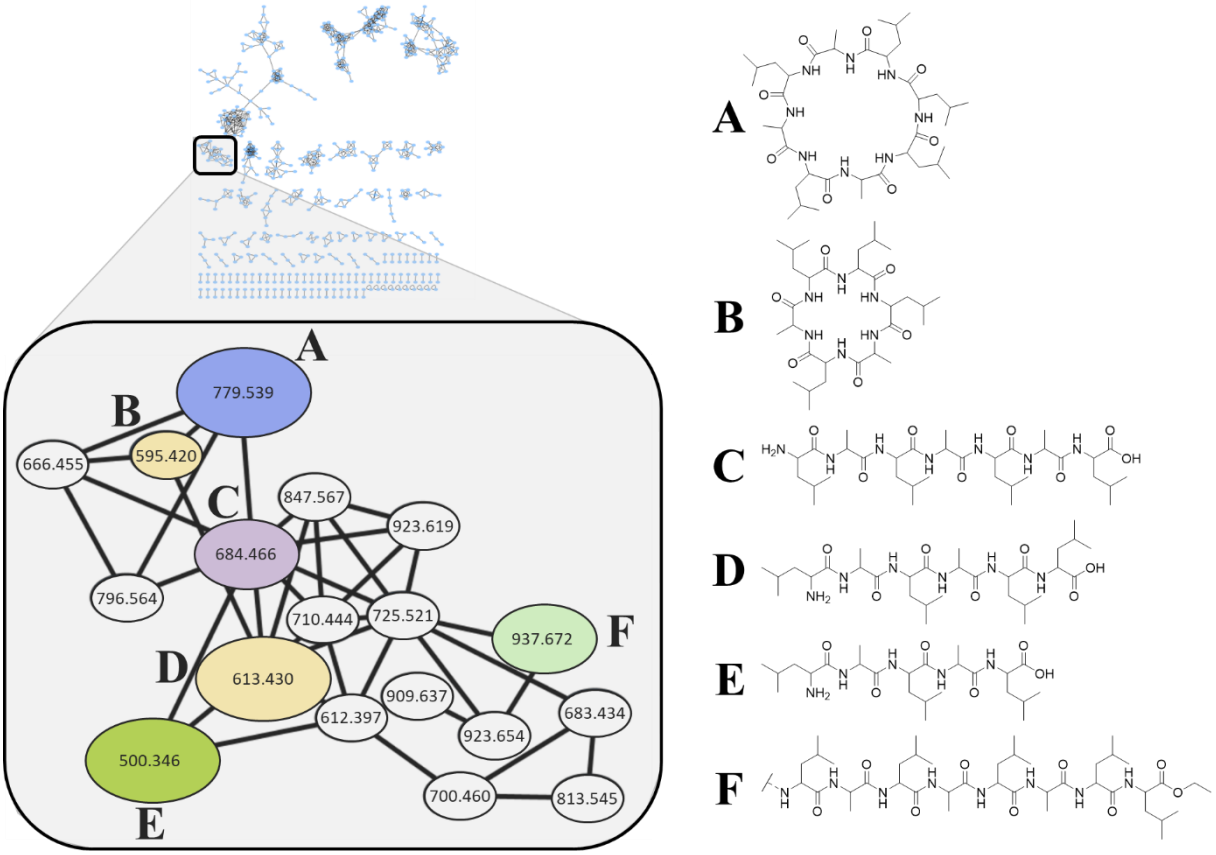


*Figure 7: GNPS network of cultivation of MSr11367 under 12 different cultivation conditions with zoomed view in the MS cluster showing derivatives of **1** (**A**) aggregated due to similar MS^2 fragmentation pattern; chemical structures predicted for **B**-**F** via MS^2 fragmentation; unclear N-terminus for **F**; node size reflects production rate and color indicates different conditions: **A** produced best during cultivation with soil extract, **B** and **D** during light exclusion, **C** during rifampicin supplementation, **E** during oxygen limitation, **F** during ethanol addition; grey nodes equal low feature production.*

The smallest derivative among this molecular network myxolutamid E belongs to an $[M+H]^+$ peak at *m/z* 500.3455 with a predicted sum formula $C_{24}H_{46}N_5O_6^+$ (*m/z* calcd for $[M+H]^+$ 500.3443 Δ2.48 ppm). The respective MS^2 fragmentation pattern is pointing towards a linear NP with an amino acid sequence of Leu-Ala-Leu-Ala-Leu. Interestingly, myxolutamid E was produced the best during cultivation with oxygen limitation and downregulated when adding soil extract. Another linear derivative, myxolutamid D, corresponds to an $[M+H]^+$ peak at *m/z* 613.4284 with a predicted sum formula $C_{30}H_{57}N_6O_7^+$ (*m/z* calcd for $[M+H]^+$ 613.4283 Δ0.06 ppm) with a predicted amino acid sequence of Leu-Ala-Leu-Ala-Leu-Leu. The respective cyclic analog myxolutamid B exhibits an $[M+H]^+$ peak at *m/z* 595.4169 with a predicted sum formula $C_{30}H_{55}N_6O_6^+$ (*m/z* calcd for $[M+H]^+$ 595.4178 Δ1.38 ppm). Myxolutamids B and D were produced best during light exclusion and soil extract supplementation led to downregulation. Furthermore, another linear congener, myxolutamid C, shows an $[M+H]^+$ peak at *m/z* 684.4653 with a predicted sum formula $C_{33}H_{62}N_7O_8^+$ (*m/z* calcd for $[M+H]^+$ 685.4654

Δ0.24 ppm), whereby the MS^2 fragmentation points towards the amino acid sequence Leu-Ala-Leu-Ala-Leu-Ala-Leu. The superior cultivation condition for myxolutamid C was supplementation with rifampicin, whereby soil extract addition reduced the production. Finally, the linear derivative myxolutamid F displays an $[M+H]^+$ peak at *m/z* 937.6705 in line with the sum formula $C_{48}H_{89}N_8O_{10}^+$ (*m/z* calcd for $[M+H]^+$ 937.6696 Δ0.95 ppm). The corresponding MS^2 fragmentation pattern suggests the partial amino acid sequence to be ?-Leu-Ala-Leu-Ala-Leu-Ala-Leu-Leu-Et with an unpredictable N-terminus. Myxolutamid F was only produced during ethanol supplementation.

To the best of our knowledge, compound **1** is the first eight-membered cyclopeptide with only D-amino acids incorporated. The closest structural similarity might be found to valinomycin, a cyclododecadepsipeptide consisting of D-valine, D-α-hydroxyvaleric acid, L-valine and L-lactic acid [38,39]. Valinomycin was isolated from *Streptomyces fulvissimus* and is known for its ion transport behavior of mitochondrial systems due to its high affinity to potassium ions [40]. Furthermore, an *in-vitro* bioactivity against *Mycobacterium tuberculosis* was reported [38], whereas for **1** no activity against our panel of Gram-positive and Gram-negative bacteria as well as fungi could be determined. In contrast to **1**, valinomycin features both ester bonds as well as amid bonds instead of amid bonds solely, which explains the affiliation to depsipeptides. Moreover, **1** consists of D-configured amino acids, whereas valinomycin includes D- and L-configured moieties.

Pentacyclic acid

Another highly temperature-dependent feature with at *m/z* 307.2639, where the production increases by 2-fold when cultivating at 18 °C in comparison to the standard cultivation and vanishes completely at 37 °C, caught our attention (Figure 8). As cultivation at 18 °C however was highly prone to contaminations and production levels were still satisfactory, we decided to perform large scale cultivation at 30 °C. Separation of the crude extract and semi-preparative purification allowed to *de novo* NMR structure elucidation of pentacyclic acid.

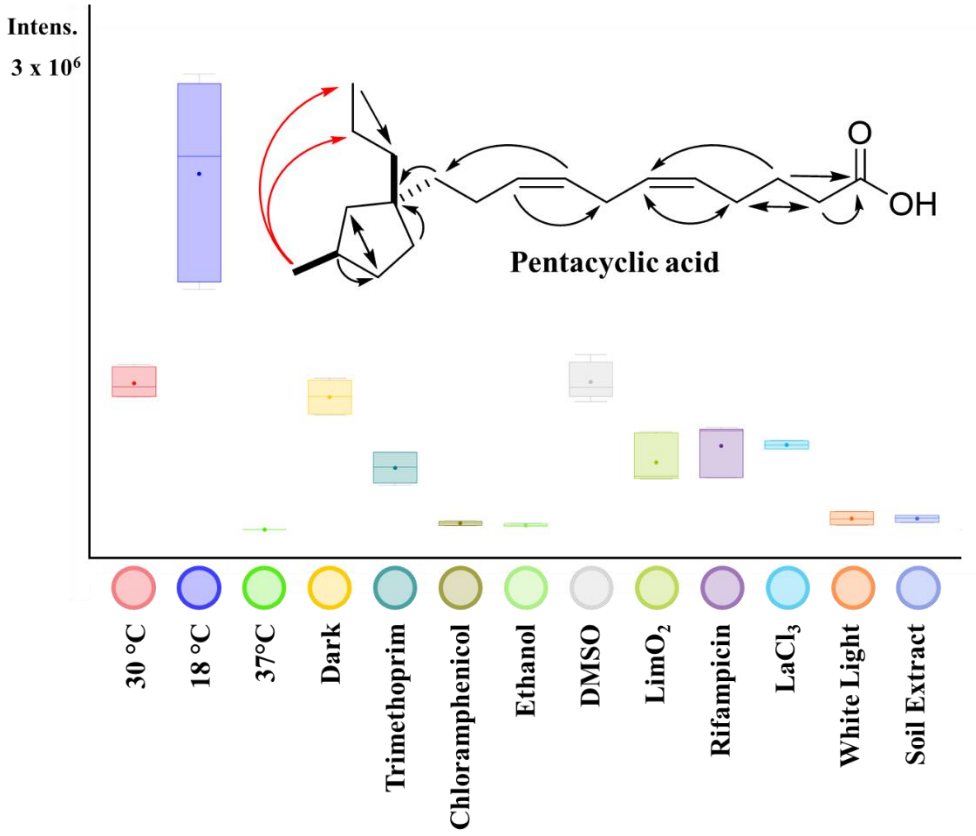


*Figure 8: Box-and-whisker diagram showing intensity levels of m/z 307.2639 among the different cultivation conditions: pentacyclic acid (**2**) shows a temperature dependent intensity variation, cultivation at 18 °C (lilac) improves production by 2-fold, whereas production is completely depleted at 37 °C (green); median and mean values and 25% and 75% quartiles illustrated within box with grey whiskers representing lower and upper extremes; relevant HMBC (black arrows) and NOESY (red arrows) correlations.*

Structure elucidation of pentacyclic acid. The collective structural information received from 1D and 2D NMR experiments including a NOESY experiment indicate an uncommon pentacyclic moiety methylated on position 15 with two alkyl chains linked via a quaternary carbon at position 12 (Figure 8). One of the substituents at the quaternary carbon was assigned as propyl function and the other substituent as built up of six methylene groups separated by two double-bonds at position 5 and 8 and a terminal carboxyl function with a characteristic shift of $\delta(^{13}C) = 178.4$ ppm. This is underpinned by COSY correlations between the methylene signal of the propyl group at $\delta(^{1}H) = 1.29$ $\delta(^{13}C) = 33.2$ ppm and the methylene signal of the

cyclopentyl ring at $\delta(^1H)$ = 1.18 ppm. Furthermore, the HMBC correlation between the methylene group belonging to the aliphatic chain at $\delta(^1H)$ = 1.36 ppm and the quaternary carbon support the predicted structural moiety. The NOESY experiment was used to examine the relative stereochemistry at positions 12 and 15. Respective results point towards a cis-configuration of the methyl group to the propyl chain (SI). Moreover, the two double bonds of **2** are cis-configured, which is supported by the finding of Pfeffer *et al.*, where the ^{13}C-shift of the allylic methylene group between two double-bonds in $CDCl_3$ was calculated to be found at $\delta(^{13}C)$ = 25.7 ppm for all cis- and at $\delta(^{13}C)$ = 35.7 ppm for all trans-configuration [41]. The respective signal of **2** is located at $\delta(^{13}C)$ = 26.7 ppm in methanol-d_4 with a deviation of 1 Hz to the all-cis-configuration prediction.

Compound **2** features a conspicuous structure that shows low similarity to already reported NPs. To the best of our knowledge no polyunsaturated fatty acids with pentacyclic moieties produced by bacteria are reported to date. The pentacyclic-substructure featuring a quaternary carbon incorporated in **2** was described in pentacyclic triterpenes of the hopanoid type, which are part of the membrane [42], as well as the diterpene cystodienoic acid [43]. Instances for the polyunsaturated fatty acids incorporating to a cyclic substructures are (2*E*,4*E*)-5-cyclohexylpenta-2,4-dienoic acid, which is part of alisamycin produced by *Streptomyces* sp. featuring a terminal cyclohexyl moiety [44], and imidacin with a cyclopropyl part in the structure isolated from *Stigmatella aurantiaca* [45]. So far, no antimicrobial or cytotoxic activity could be observed for **2**, thus we suppose a function as a signaling molecule between the myxobacteria or as a membrane building block. This means that a clear assignment as primary or secondary metabolite is not trivial hindering a simple identification of the biosynthetic genes responsible for pentacyclic acid formation.

Glucodiolic acid

Lastly, we investigated a feature for cultivation with appendage of $LaCl_3$ which is completely absent under standard cultivation conditions and generally only observed in three out of our twelve conditions (Figure 9). This compound was produced during cultivation in the dark as well as addition of soil extract. It is noteworthy that the production rate is low for all of these conditions (approximately 0.4 mg/L for cultivation with $LaCl_3$), so we further optimized production by optimizing the media composition. EPM medium was found to provide sufficient production for purification and structure elucidation (production rate approximately 3 mg/L). Interestingly, a trace element solution is added to EPM medium including cobalt, copper and manganese salts, which might be responsible for the boost in production of this feature, likely showing similar effects compared to the addition of $LaCl_3$. Separation and following semi-preparative workup of the crude extract yielded glucodiolic acid, which was *de novo* structure elucidated using 1D and 2D NMR experiments.

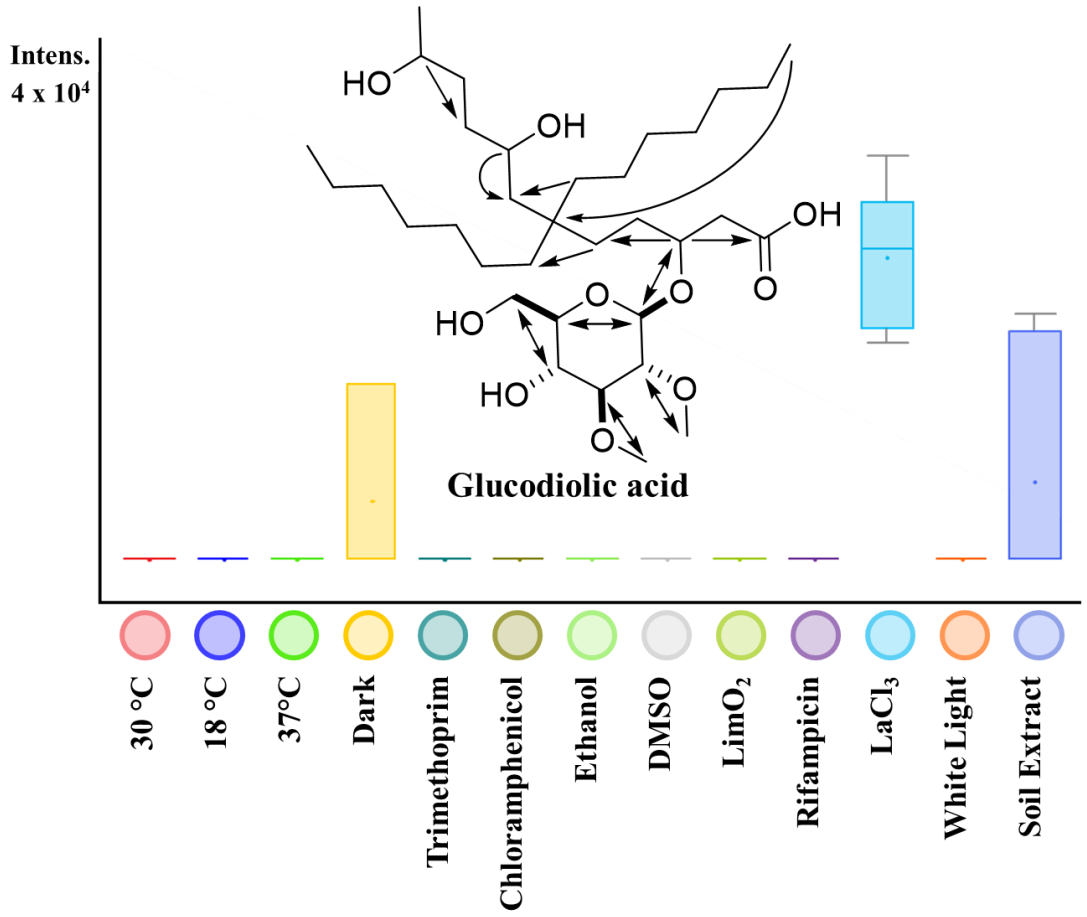
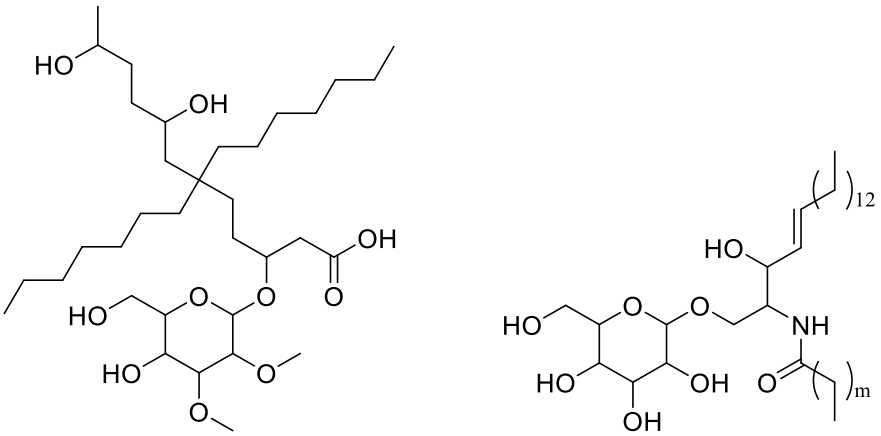


*Figure 9: Box-and-whisker diagram showing intensity levels of m/z 635.4722 among the different cultivation conditions: glucodiolic acid (**3**) is absent under standard conditions (red) but induced by supplementation of LaCl3 (light blue); median and mean values and 25% and 75% quartiles illustrated within box with grey whiskers representing lower and upper extremes; relevant HMBC correlations (arrows).*

Structure elucidation of glucodiolic acid. The structure proposed using 1D and 2D NMR experiment data comprises a PKS part with a carboxyl group and two hydroxyl functions as well as a dimethylated hexose in β-position and an uncommon moiety consisting of two C_7H_{15}-chains connected via a quaternary carbon. The β-methine group in relation to the free acid function features a shift of $\delta(^1H)$ = 4.08 ppm due to the deshielding hexose and exhibits HMBC correlations to the carboxyl group with a characteristic shift of $\delta(^{13}C)$ = 177.2 ppm, to the anomeric carbon of the sugar at $\delta(^1H)$ = 4.38 ppm as well as to the δ-methylene group at $\delta(^1H)$ = 1.45 ppm. The latter is adjacent to the unprecedented quaternary carbon at $\delta(^{13}C)$ = 33.3 ppm, which connects the two C_7H_{15}-alkyl chains with the remaining PKS part containing two hydroxyl functions at position 8 and 11, which feature characteristic downfield shifts of $\delta(^1H)$ = 3.50 and 3.71 ppm. All signal belonging to the two C_7H_{15}-alkyl chains show HMBC correlations to the quaternary carbon in the center of the molecule. Coming to the hexose, corresponding COSY and HMBC correlations prove the sugar to be a twice methylated hexose. Furthermore, the coupling constants assigned to the hexose moiety point towards a β-glucose configuration with a deviation of less than 0.32 Hz to β-D-glucopyranose measured in D_2O by Roslund *et al.* [46]. The stereo centers of the hydroxylated methine groups are inaccessible so far due to the close neighborhood to each other only separated by two methylene groups, thus hindering a prediction via shift changes after a derivatization, *e.g.* Mosher esterification [47]. However, a synthetic approach might help in the future to solve this issue, but this would be out of the scope of this study.

The proposed structure is underpinned by the fragmentation behavior of **3** in scheduled precursor MS^2 measurements and selected MS^n (SI). This data especially supports the hypothesis of two C_7H_{15}-moieties at the quaternary carbon instead of two different chain lengths. Only few glycosylated NPs were isolated from myxobacteria like the PKS/NRPS hybrid products cystomanamides C and D that are N-glycosylated and isolated from *Cystobacter fuscus* [48] as well as the PKS product chivosazol A with a glycosidically bound 6-deoxyglucose produced by *Sorangium cellulosum* [49]. Furthermore, glucodiolic acid's striking structural feature is the quaternary carbon, that is rather rare in NP discovery besides the structurally distant steroids like cholesterol [50].



Glucodiolic acid **Cerebroside**

Figure 10: Structural comparison of glucodiolic acid and cerebrosides.

The highest scaffold similarity might be found for so called cerebrosides (Figure 10), that are glycosphingolipides usually found in muscles and brain tissue, but they were also isolated from yeast as cell membrane building blocks [51]. In contrast to **3**, cerebrosides do not contain quaternary carbons and exhibit amide bonds as well as double bonds that are not present in glucodiolic acid. Thus, compound **3** seems to be structurally special among NPs, which raises the question of its function. Our in-house antimicrobial panel showed no bioactivity or cytotoxic activity, so this compound's purpose is no antibiotic or toxin function. Concluding, no clear assignment as primary or secondary metabolite is possible at this point, as the responsible biosynthetic gene cluster could not be identified so far.

2.3 Materials and methods

Cultivation of strain MSr11367

For the statistical analysis MSr11367 was cultivated in 100 mL scale in triplicates with respective blanks for 2 weeks in 2-SWT medium. 1% of a preculture grown for 4 days in S15 medium was used for inoculation and after 3 days 2% XAD-16 adsorber resin (Amberlite XAD-16, Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany) was added. Standard cultivation condition were 30 °C temperature and 180 rpm shaker speed. Small scale cultivation experiments were carried out under 12 conditions: different temperatures (18 °C, 30 °C, 37 °C), illuminations (white light, darkness), influence of additives in sub-lethal concentrations (50 µg/mL rifampicin, 50 µg/mL chloramphenicol, 25 µg/mL trimethoprim, 0.5 mM $LaCl_3$, 2% ethanol and 1% soil extract) and oxygen limitation (150 mL culture volume, 90 rpm). Antibiotics stock solutions were prepared in DMSO and $LaCl_3$ x 6 H_2O was dissolved in water and sterile filtrated. Soil extract was obtained through lyophilization of a soil sample (pond in Schiffweiler, Germany) and super critical fluid extraction using a Waters MV-10 ASFE system. Approximately 36 mg of dried soil extract were dissolved in 4 mL of water/methanol 1:1, sterile filtrated and 1 mL thereof was added to 100 mL liquid culture. After 2 weeks all cultures were harvested by centrifugation for 4 min at 4000 rpm and decantated to retain the cell pellet with XAD-16. Cells and adsorber resin were extracted with 40 mL 1:1 methanol/acetone stirring overnight. The filtrated extract was dried under reduced pressure and redissolved in 1 mL methanol.

Isolation procedure

Compound **1** was isolated from 10 L large scale cultivation of MSr11367 in 2-SWT medium supplemented with soil extract upscaled from analytic conditions to 25 x 25 mL extraction vessels at 30 °C. Cells and adsorber resin were extracted with 2 L 1:1 methanol/acetone, stirring overnight. The filtrated extract was dried and redissolved in methanol to be preseparated by liquid-liquid partitioning using hexane, chloroform and ethyl acetate (SI). The combined chloroform phase was fractionated by size exclusion chromatography using a gravitation column filled with Sephadex LH 20 as stationary phase and methanol as mobile phase. Respective fractions containing **1** were further purified on a Dionex ultimate 3000 series system coupled to a Thermo Scientific ISQ EM single quadrupole MS system with a HESI ion source (Thermo Fisher Scientific, Waltham, MA, USA). More details can be found in the SI.

Compound **2** was purified from 10 L large scale cultivation of MSr11367 in 2-SWT medium without additives at 30 °C. After methanol/acetone extraction and liquid-liquid partitioning the

desired compound was found in the hexane phase, which was further separated using size exclusion chromatography as described for compound 1. Subsequent purification of **2** was achieved using a Dionex ultimate 3000 series system (Thermo Fisher Scientific, Waltham, MA, USA) linked to a Bruker Daltonics HCT plus ion trap system with standard ESI ion source (Bruker Daltonics, Billerica, MA, USA). More details can be found in the SI.

For the purification of **3**, MSr11367 was cultivated on a large scale in EPM medium. Afterwards, the methanol/acetone extract was separated via liquid-liquid partitioning, with the chloroform phase dissolving the target compound. The first purification step was centrifugal partition chromatography (CPC) using a Gilson CPC 100 device linked to a Varian ProStar Solvent delivery module and a Varian ProStar 2 Channel UV detector operating at 220 and 254 nm. Compound **3** containing fractions were eluted with reverse mode and further purified with a Dionex ultimate 3000 series system coupled to a Thermo Scientific ISQ EM single quadrupole MS system with a HESI ion source (Thermo Fisher Scientific, Waltham, MA, USA).

UHPLC-*hr*MS measurements

UHPLC-*hr*MS experiments were carried out on a Dionex UltiMate 3000 rapid separation liquid chromatography (RSLC) system (Thermo Fisher Scientific, Waltham, MA, USA) coupled to a Bruker Daltonics maXis 4G UHR-TOF device (Bruker Daltonics, Billerica, MA, USA). Separation is reached using an Acquity BEH C18 column (100 mm × 2.1 mm, 1.7 µm dp) (Waters, Eschborn, Germany) equipped with a Waters VanGuard BEH C18 1.7 µm guard in combination with a linear 5–95% gradient of acetonitrile with 0.1% formic acid in ddH2O with 0.1% formic acid for 18 min with detection by a diode array detector at 200–600 nm.. The flow rate was adjusted to 0.6 mL/min, the thermostat to 45 °C and the LC flow is split to 75 µL/min before entering the mass spectrometer. The ultra high resolution time of flight mass spectrometer utilizes an ESI source and following MS settings: capillary voltage 4000 V, end plate off-set -500 V, nebulizer gas pressure 1 bar, dry gas flow rate 5 L/min, dry gas temperature 200 °C, mass scan range *m/z* 150–2500. Calibration was performed on the masses of the first isotope signal of sodium formate clusters in Quadratic + HPC mode, whereby the clusters are created in the ion source after co-injection of a 1 mM sodium hydroxide in 1:1 water:isopropanol + 0.1% FA mix. For MS/MS measurements a collision energy of 5 eV was used in CID fragmentation mode. LC-MS data were analysed with DataAnalysis 4.4 (Bruker Daltonics, Billerica, MA, USA).

MS^n measurements

DI-FTICR experiments were conducted on a Bruker SolariX XR 7T device (Bruker Daltonics, Billerica, MA, USA) after external calibration to ensure a mass accuracy below 1 ppm. Samples were diluted to 1:100 with methanol and injected via direct infusion using a preinstalled syringe pump at flowrate of 2.0 µL/min. All measurements were acquired in positive ESI ionization mode with same source parameters as used for maXis 4G. For full scan measurements 16 scans were acquired accumulating for 50 ms and a data size of 1 M. MS^n fragmentation experiments were carried out with 64 scans accumulating for 200 ms, 1 M data size and 1.3 eV SORI fragmentation.

Metabolomics analysis and molecular networking

Statistical analysis of the myxobacterial extracts was carried out using Metaboscape 2021b (Bruker Daltonics, Billerica, MA, USA). Threshold was set to 15,000 counts with a minimum peak length of 5 spectra for LC-MS data. Spectra were recalibrated on sodium formate cluster mix. Features were accepted when detected in 5/6 samples and 0/6 blanks. T-test results were used to generate scheduled precursor lists for subsequent tandem MS measurements with the Bruker maXis 4G platform. Molecular networking of the SPL-MS/MS spectra was performed with GNPS using preset standard parameters and visualized with Cytoscape 3.7.1. The assignment of the secondary metabolite regions was done by antiSMASH version 6.1.1.

NMR analysis

The NMR data set consisting of 1 and 2D spectra for structure elucidation of **1**, **2** and **3** were recorded on a Bruker Avance III (Ultrashield) 500 MHz spectrometer equipped with a 5 mm TCI cryoprobe and respective measurements for dehydroxynocardamine B were conducted on a Bruker Avance III (Ascend) 700 MHz spectrometer. Further details can be found in SI.

Stereochemical analysis of 1

Configuration of the amino acids incorporated in compound **1** was evaluated via Marfey's assay after hydrolysis of the amid bonds, derivatization of the free amino acids and chromatographic comparison to derivatized reference leucine and alanine. Samples were measured on the maXis 4G UHR-TOF (Bruker Daltonics, Billerica, MA, USA) coupled to a Dionex Ultimate 3000 SL system (Thermo Fisher Scientific, Waltham, MA, USA). Analytes were separated by a Waters Acquity BEH C18 column (50x2.1 mm, 1.7 µm) using the following multi-step gradient with

A=0.1% formic acid in ddH2O and B=0.1% formic acid in acetonitrile: 5-10% B in 1 min, 10-35% B in 14 min, 35-55% in 7 min, 55-80% in 3 min, hold at 80% for 1 min and reequilibration at 5% B for 5 min. The MS settings are equivalent to the UHPLC-*hr*MS measurements besides an adapted mass range detection of 150 to 1000 *m/z*.

2.4 Conclusion

We have applied altered cultivation conditions to the myxobacterial strain MSr11367, which were found to lead to a significant up- or downregulation of feature production rates or even the detection of novel features that are not observable under standard cultivation conditions. The highest amount of novel features could be detected for addition of soil extract, lanthanum chloride and limited oxygen supply. Taking also reproducibility aspects into consideration, which play an important role for subsequent isolation and structure elucidation of those NPs, addition of soil extract as well as lanthanum chloride were found especially promising. This indicates that mimic the natural environment in soil seems to have the most promising positive effect on the reproducible production of features absent under standard cultivation conditions. Based on this analysis, we successfully isolated three structurally novel metabolites, namely myxolutamid, pentacyclic acid and glucodiolic acid, highlighting the use of aggregated cultivation conditions for the discovery of unprecedented metabolites from myxobacteria. However, the responsible gene clusters remain elusive, so it is worth noting that with this method it is not possible to distinguish between primary and secondary metabolites. One possible way to address this problem in the future would be conducting transcriptomics or proteomics analysis using the different cultivation conditions to correlate RNA or protein expression levels to production titers and subsequently use this data to uncover their biosynthetic machinery. Strikingly, all three structures show very low similarity to already known metabolites.

2.5 References

1. Baumann, S.; Herrmann, J.; Raju, R.; Steinmetz, H.; Mohr, K.I.; Huttel, S.; Harmrolfs, K.; Stadler, M.; Muller, R. Cystobactamids: myxobacterial topoisomerase inhibitors exhibiting potent antibacterial activity. *ANGEWANDTE CHEMIE-INTERNATIONAL EDITION* **2014**, *53*, 14605–14609, doi:10.1002/anie.201409964.
2. 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, doi:10.1002/anie.201808028.
3. Bader, C.D.; Panter, F.; Garcia, R.; Tchesnokov, E.P.; Haid, S.; Walt, C.; Spröer, C.; Kiefer, A.F.; Götte, M.; Overmann, J.; et al. Sandacrabins - Structurally Unique Antiviral RNA Polymerase Inhibitors from a Rare Myxobacterium. *Chemistry – A European Journal* **2022**, *28*, e202104484, doi:10.1002/chem.202104484.
4. 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. *J. Nat. Prod.* **2021**, doi:10.1021/acs.jnatprod.0c00942.
5. Scherlach, K.; Hertweck, C. Triggering cryptic natural product biosynthesis in microorganisms. *Org. Biomol. Chem.* **2009**, *7*, 1753–1760, doi:10.1039/b821578b.
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, doi:10.1039/C8SC01325J.
7. Gemperlein, K.; Rachid, S.; Garcia, R.O.; Wenzel, S.C.; Müller, R. Polyunsaturated fatty acid biosynthesis in myxobacteria: Different PUFA synthases and their product diversity. *Chem. Sci.* **2014**, *5*, 1733–1741, doi:10.1039/C3SC53163E.
8. Tanaka, Y.; Kasahara, K.; Hirose, Y.; Murakami, K.; Kugimiya, R.; Ochi, K. Activation and products of the cryptic secondary metabolite biosynthetic gene clusters by rifampin resistance (*rpoB*) mutations in actinomycetes. *J. Bacteriol.* **2013**, *195*, 2959–2970, doi:10.1128/JB.00147-13.
9. Chen, G.; Wang, G.Y.; Li, X.; Waters, B.; Davies, J. Enhanced production of microbial metabolites in the presence of dimethyl sulfoxide. *J. Antibiot.* **2000**, *53*, doi:10.7164/antibiotics.53.1145.
10. Weinberg, E.D. Roles of trace metals in transcriptional control of microbial secondary metabolism. *Biol Metals* **1990**, *2*, 191–196, doi:10.1007/BF01141358.

11. Bode, H.B.; Bethe, B.; Höfs, R.; Zeeck, A. Big Effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity. *ChemBioChem* **2002**, *3*, 619, doi:10.1002/1439-7633(20020703)3:7<619:AID-CBIC619>3.0.CO;2-9.
12. Pan, R.; Bai, X.; Chen, J.; Zhang, H.; Wang, H. Exploring Structural Diversity of Microbe Secondary Metabolites Using OSMAC Strategy: A Literature Review. *Front. Microbiol.* **2019**, *10*, 294, doi:10.3389/fmicb.2019.00294.
13. Tanaka, Y.; Hosaka, T.; Ochi, K. Rare earth elements activate the secondary metabolite-biosynthetic gene clusters in Streptomyces coelicolor A3(2). *J. Antibiot.* **2010**, *63*, 477–481, doi:10.1038/ja.2010.53.
14. 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, doi:10.1021/acschembio.7b00900.
15. 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, doi:10.1021/acschembio.8b00948.
16. 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, doi:10.1021/acschembio.7b00816.
17. Hug, J.J.; Panter, F.; Krug, D.; Müller, R. Genome mining reveals uncommon alkylpyrones as type III PKS products from myxobacteria. *Journal of Industrial Microbiology & Biotechnology* **2018**, doi:10.1007/s10295-018-2105-6.
18. Bader, C.D.; Neuber, M.; Panter, F.; Krug, D.; Müller, R. Supercritical Fluid Extraction Enhances Discovery of Secondary Metabolites from Myxobacteria. *Anal. Chem.* **2020**, *92*, 15403–15411, doi:10.1021/acs.analchem.0c02995.
19. 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* **2020**, *25*, 2676, doi:10.3390/molecules25112676.
20. Li, Y.; Zhuo, L.; Li, X.; Zhu, Y.; Wu, S.; Hu, W.; Shen, T.; Li, Y.-Z.; Wu, C. Myxadazoles, Myxobacterium-Derived Isoxazole-Benzimidazole Hybrids with Cardiovascular Activities. *Angew. Chem.* **2021**, doi:10.1002/ange.202106275.
21. Hug, J.J.; Frank, N.A.; Walt, C.; Šenica, P.; Panter, F.; Müller, R. Genome-Guided Discovery of the First Myxobacterial Biarylitide Myxarylin Reveals Distinct C–N Biaryl

Crosslinking in RiPP Biosynthesis. *Molecules* **2021**, *26*, 7483, doi:10.3390/molecules26247483.

22. Frank, N.A.; Széles, M.; Akone, S.H.; Rasheed, S.; Hüttel, S.; Frewert, S.; Hamed, M.M.; Herrmann, J.; Schuler, S.M.M.; Hirsch, A.K.H.; et al. Expanding the Myxochelin Natural Product Family by Nicotinic Acid Containing Congeners. *Molecules* **2021**, *26*, 4929, doi:10.3390/molecules26164929.
23. Popoff, A.; Hug, J.J.; Walesch, S.; Garcia, R.; Keller, L.; Müller, R. Structure and Biosynthesis of Myxofacyclines: Unique Myxobacterial Polyketides Featuring Varing and Rare Heterocycles *. *Chemistry* **2021**, *27*, 16654–16661, doi:10.1002/chem.202103095.
24. Panter, F.; Bader, C.D.; Müller, R. The Sandarazols are Cryptic and Structurally Unique Plasmid-Encoded Toxins from a Rare Myxobacterium*. *Angew. Chem. Int. Ed Engl.* **2021**, *60*, 8081–8088, doi:10.1002/anie.202014671.
25. Okoth, D.A.; Hug, J.J.; Mándi, A.; Kurtán, T.; Garcia, R.; Müller, R. Structure and biosynthesis of sorangipyranone - a new γ-dihydropyrone from the myxobacterial strain MSr12020. *J. Ind. Microbiol.* **2021**, *48*, doi:10.1093/jimb/kuab029.
26. Hug, J.J.; Kjaerulff, L.; Garcia, R.; Müller, R. New Deoxyenhygrolides from *Plesiocystis pacifica* Provide Insights into Butenolide Core Biosynthesis. *Mar. Drugs* **2022**, *20*, 72, doi:10.3390/md20010072.
27. Panter, F.; Popoff, A.; Garcia, R.; Krug, D.; Müller, R. Myxobacteria of the Cystobacterineae Suborder Are Producers of New Vitamin K2 Derived Myxoquinones. *Microorganisms* **2022**, *10*, 534, doi:10.3390/microorganisms10030534.
28. Okoth, D.A.; Hug, J.J.; Garcia, R.; Müller, R. Three New Stigmatellin Derivatives Reveal Biosynthetic Insights of Its Side Chain Decoration. *Molecules* **2022**, *27*, doi:10.3390/molecules27144656.
29. Gerth, K.; Bedorf, N.; Irschik, H.; Höfle, G.; Reichenbach, H. The soraphens: a family of novel antifungal compounds from *Sorangium cellulosum* (Myxobacteria). I. Soraphen $A_{1\alpha}$: fermentation, isolation, biological properties. *J. Antibiot.* **1994**, *47*, 23–31, doi:10.7164/antibiotics.47.23.
30. Qualls, G.T.; Stephens, K.; White, D. Light-stimulated morphogenesis in the fruiting myxobacterium *Stigmatella aurantiaca*. *Science* **1978**, *201*, 444–445.
31. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; van Wezel, G.P.; Medema, M.H.; Weber, T. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **2021**, *49*, W29-W35, doi:10.1093/nar/gkab335.

32. Mosteller, R.D.; Yanofsky, C. Transcription of the tryptophan operon in Escherichia coli: Rifampicin as an inhibitor of initiation. *J. Mol. Biol.* **1970**, *48*, 525–531, doi:10.1016/0022-2836(70)90064-1.
33. S. Benham Kahn; Sandra A. Fein; Isadore Brodsky. Effects of trimethoprim on folate metabolism in man. *Clinical Pharmacology & Therapeutics* **1968**, *9*, 550–560, doi:10.1002/cpt196895550.
34. ALLISON, J.L.; HARTMAN, R.E.; HARTMAN, R.S.; Wolfe, A.D.; CIAK, J.; Hahn, F.E. Mode of action of chloramphenicol. VII. Growth and multiplication of Escherichia coli in the presence of chloramphenicol. *J. Bacteriol.* **1962**, *83*, 609–615, doi:10.1128/jb.83.3.609-615.1962.
35. Baltz, R.H. Spontaneous and induced mutations to rifampicin, streptomycin and spectinomycin resistances in actinomycetes: mutagenic mechanisms and applications for strain improvement. *J Antibiot* **2014**, *67*, 619–624, doi:10.1038/ja.2014.105.
36. 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, doi:10.1038/nbt.3597.
37. Stubbendieck, R.M.; May, D.S.; Chevrette, M.G.; Temkin, M.I.; Wendt-Pienkowski, E.; Cagnazzo, J.; Carlson, C.M.; Gern, J.E.; Currie, C.R. Competition among Nasal Bacteria Suggests a Role for Siderophore-Mediated Interactions in Shaping the Human Nasal Microbiota. *Appl. Environ. Microbiol.* **2019**, *85*, doi:10.1128/AEM.02406-18.
38. Hans Brockmann; Günter Schmidt-Kastner. Valinomycin I, XXVII. Mitteil. über Antibiotica aus Actinomyceten. *Chemische Berichte* **1955**, *88*, 57–61, doi:10.1002/cber.19550880111.
39. Shemyakin, M.M.; Vinogradova, E.I.; Feigina, M.Y.; Aldanova, N.A.; Loginova, N.F.; Ryabova, I.D.; Pavlenko, I.A. The structure-antimicrobial relation for valinomycin depsipeptides. *Experientia* **1965**, *21*, 548–552, doi:10.1007/BF02138991.
40. Moore, C.; Pressman, B.C. Mechanism of action of valinomycin on mitochondria. *Biochemical and Biophysical Research Communications* **1964**, *15*, 562–567, doi:10.1016/0006-291X(64)90505-4.
41. Philip E. Pfeffer; Francis E. Luddy; Joseph Unruh; James N. Shoolery. Analytical13C NMR: A rapid, nondestructive method for determining the cis,trans composition of catalytically treated unsaturated lipid mixtures. *Journal of the American Oil Chemists' Society* **1977**, *54*, 380–386, doi:10.1007/BF02802040.

42. Kannenberg, E.L.; Poralla, K. Hopanoid biosynthesis and function in bacteria. *Naturwissenschaften* **1999**, *86*, 168–176.
43. 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, doi:10.1038/ja.2015.8.
44. Franco, C.M.; Maurya, R.; Vijayakumar, E.K.; Chatterjee, S.; Blumbach, J.; Ganguli, B.N. Alisamycin, a new antibiotic of the manumycin group. I. Taxonomy, production, isolation and biological activity. *J. Antibiot.* **1991**, *44*, doi:10.7164/antibiotics.44.1289.
45. Krug, D. Natural product biosynthesis in myxobacteria: Studies on enzymatic versatility and secondary metabolite diversity. Dissertation; Saarland University, Saarbrücken, 2009.
46. Roslund, M.U.; Tähtinen, P.; Niemitz, M.; Sjöholm, R. Complete assignments of the 1H and 13C chemical shifts and JH,H coupling constants in NMR spectra of d-glucopyranose and all d-glucopyranosyl-d-glucopyranosides. *Carbohydr. Res.* **2008**, *343*, 101–112, doi:10.1016/j.carres.2007.10.008.
47. 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, doi:10.1038/nprot.2007.354.
48. 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, doi:10.1021/ol500779s.
49. 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.
50. Bode, H.B.; Zeggel, B.; Silakowski, B.; Wenzel, S.C.; Reichenbach, H.; Müller, R. Steroid biosynthesis in prokaryotes: identification of myxobacterial steroids and cloning of the first bacterial 2,3(*S*)-oxidosqualene cyclase from the myxobacterium *Stigmatella aurantiaca*. *Mol. Microbiol.* **2003**, *47*, 471–481, doi:10.1046/j.1365-2958.2003.03309.x.
51. Takakuwa, N.; Kinoshita, M.; Oda, Y.; Ohnishi, M. Existence of cerebroside in Saccharomyces kluyveri and its related species. *FEMS Yeast Res* **2002**, *2*, 533–538, doi:10.1111/j.1567-1364.2002.tb00120.x.

S2. Supporting information

Aggravated Cultivation of Myxobacteria Stimulates Secondary Metabolism

Christine Walt[1,2], Chantal D. Bader[1,2], Daniel Krug[1,2] and Rolf Müller[1,2,3,*]

[1]Helmholtz-Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), Saarland University, Campus E8 1, 66123 Saarbrücken, Germany

[2]German Center for Infection Research (DZIF), Partner Site Hannover-Braunschweig, 38124 Braunschweig, Germany

[3]Helmholtz International Lab for anti-infectives Campus E8 1, 66123 Saarbrücken, Germany

[*]Author to whom correspondence should be addressed.

S2.1 Myxobacterial natural products isolated between 2018 and 2022

Table S1: Selected list of myxobacterial natural products isolated between 2018 and 2022 under standard cultivation conditions.

Compound	Source	Tempera-ture	Medium	Year of isolation	Reference
Crocadepsins	*Chondromyces crocatus* Cm c5	30 °C	Pol 0.3	2018	[1]
Fulvuthiacenes	*Myxococcus fulvus* MCy9280	30 °C	VYGS	2018	[2]
Homospermidin lipids	*Myxococcus xanthus* DK1622	30 °C	CTT; VY/2	2018	[3]
Pyxidicyclines	*Pyxidicoccus fallax* An d48	30 °C	YM	2018	[4]
Alkylpyrones	*Myxococcus xanthus* DK1622	30 °C	CTT	2019	[5]
Chloroxanthic acid A	*Myxococcus xanthus* DK1622	30 °C	SHG-P	2020	[6]
2-Hydroxysorangiadenosine	*Vitiosangium cumulatum* MCy10943^{T}	30 °C	CyHv3	2020	[7]
Myxadazoles	*Myxococcus* sp. SDU36	30 °C	VY/2	2021	[8]
Myxarylin	*Pyxidicoccus fallax* An d48	30 °C	VY	2021	[9]
Myxochelins N1-N3	*Corallococcus sp.* MCy9049	30 °C	VY/2S	2021	[10]
Myxofacyclines	*Corallococcus* sp. MCy9072	30 °C	CyH	2021	[11]

Sandarazols	*Sandaracinus* sp. MSr10575	30 °C	2SWT	2021	[12]
Sorangipyranone	*Polyangiaceae* MSr12020	30 °C	VY/2	2021	[13]
Deoxyenhygrolides	*Plesiocystis pacifica* DSM 14875^{T}	30 °C	RG244	2022	[14]
Myxoquinones	*Corallococcus sp. strain* MCy9049	30 °C	PYGS	2022	[15]
Sandacrabins	*Sandaracinus defensii* MSr10575	30 °C	2SWT	2022	[16]
Stigmatellins	*Vitiosangium cumulatum* MCy10943^{T}	30 °C	CYHv3	2022	[17]

S2.2 Media composition

Table S2: Media composition of S15, 2-SWT and EPM medium.

Medium	Ingredients	Supplier	Amount [g/L]	pH (adjusted with KOH)
S15	Soymeal	Heusel	4	7.4
	Glucose	Sigma-Aldrich	2	
	Soluble starch	Roth	8	
	$CaCl_2$ x 2 H_2O	Sigma-Aldrich	1	
	$MgSO_4$ x 7 H_2O	Sigma-Aldrich	1	
	HEPES	Roth	11.9 (50 mM)	
	Fe-EDTA	Fluka	8 mg	
2-SWT	Tryptone	BD	3	7.0
	Soytone	BD	1	
	Glucose	Sigma-Aldrich	2	
	Soluble starch	Roth	2	
	Maltose monohydrate	Sigma-Aldrich	1	
	Cellobiose	MP Biomedicals	2	
	$CaCl_2$ x 2 H_2O	VWR Chemicals	0.5	
	$MgSO_4$ x 7 H_2O	Grüssing	1	
	HEPES	Roth	2.38	
EPM	Potato starch	Sigma-Aldrich	2	7.2
	Glucose	Sigma-Aldrich	2	
	Soy powder	Roth	2	
	Skim milk powder	Sigma-Aldrich	1	
	$MgSO_4$ x 7 H_2O	Grüssing	1	
	$CaCl_2$ x 2 H_2O	VWR Chemicals	1	
	Fe-EDTA	Fluka	8 mg	
	Trace elements	-	1 mL	

	Vitamin B12	Roth	0.5 mg	
Trace elements solution	$MnCl_2$ x 4 H_2O	Sigma-Aldrich	0.1	-
	$CoCl_2$ x 6 H_2O	Sigma-Aldrich	0.02	
	$CuSO_4$	Sigma-Aldrich	0.01	
	Na_2MoO_4 x 2H_2O	Sigma-Aldrich	0.01	
	$ZnCl_2$	Roth	0.02	
	LiCl	Sigma-Aldrich	0.005	
	$SnCl_2$ x 2 H_2O	Sigma-Aldrich	0.005	
	H_3BO_3	Roth	0.01	
	KBr	Sigma-Aldrich	0.02	
	KI	Sigma-Aldrich	0.02	

S2.3 BGC assignment by antiSMASH

Table S3: AntiSMASH version 6.1.1 assignment of BGCs of MSr11367.

Region	Type
1	T1PKS
2	T1PKS, NRPS
3	T1PKS, NRPS
4	T1PKS
5	RRE-containing
6	T1PKS, NRPS, lanthipeptide-class-i
7	ladderane
8	NRPS
9	RRE-containing
10	terpene
11	T1PKS
12	NRPS, T1PKS, betalactone
13	NRPS, T1PKS
14	terpene
15	T1PKS
16	NRPS, T1PKS
17	NRPS, T1PKS
18	T1PKS, NRPS
19	arylpolyene, ladderane
20	T1PKS, hgIE-KS
21	RiPP-like
22	terpene
23	T1PKS
24	terpene
25	lassopeptide
26	thioamitides
27	NRPS, terpene
28	T1PKS
29	T1PKS

30	NRPS
31	NRPS, T1PKS, betalactone
32	T1PKS
33	terpene
34	T1PKS, NRPS, transAT-PKS, NRPS-like, lanthipeptide-class-i
35	RiPP-like
36	T2PKS, transAT-PKS, PKS-like, lanthipeptide-class-ii
37	RiPP-like
38	T1PKS
39	NRPS
40	NRPS
41	T2PKS
42	T1PKS, NRPS
43	NRPS-like
44	NRPS-like
45	hgIE-KS
46	NRPS
47	RiPP-like
48	Siderophore
49	PKS-like, T1PKS
50	PKS-like
51	Terpene
52	NRPS, T1PKS
53	Terpene
54	NRPS, NRPS-like
55	redox-cofactor
56	NRPS
57	hgIE-KS
58	Indole
59	Lanthipeptide-class-i
60	NRPS

61	Terpene
62	NRPS, T1PKS

S2.4 Number and reproducibility of new features

Table S4: Total number of features per condition, number of new features in comparison to standard cultivation at 30 °C and unique features for a specific condition at an intensity threshold of 1,500 and a presence of 5/6 and 6/6 per condition.

Condition	Total number of features per condition	Number of new features in comparison to standard cultivation	Number of unique features per condition
30 °C	383	-	-
18 °C	314	151	36
37 °C	35	2	0
White light	176	81	38
Darkness	399	28	6
Oxygen limitation	586	325	66
Rifampicin	243	81	59
Trimethoprim	338	110	44
Chloramphenicol	428	256	28
Soil extract	720	542	243
Lanthanum chloride	592	352	98
Ethanol	170	113	69

S2.5 Up- and downregulation of features in comparison to 30 °C

Table S5: Number of up- and down-regulated features in comparison to standard cultivation at 30 °C for each condition.

Condition	Number of upregulated features in comparison to standard cultivation	Number of downregulated features in comparison to standard cultivation
18 °C	30	61
37 °C	6	42
White light	16	71
Darkness	32	0
Oxygen limitation	73	39
Rifampicin	32	33
Trimethoprim	36	58
Chloramphenicol	26	62
Soil extract	55	62
Lanthanum chloride	61	49
Ethanol	23	56

18 °C

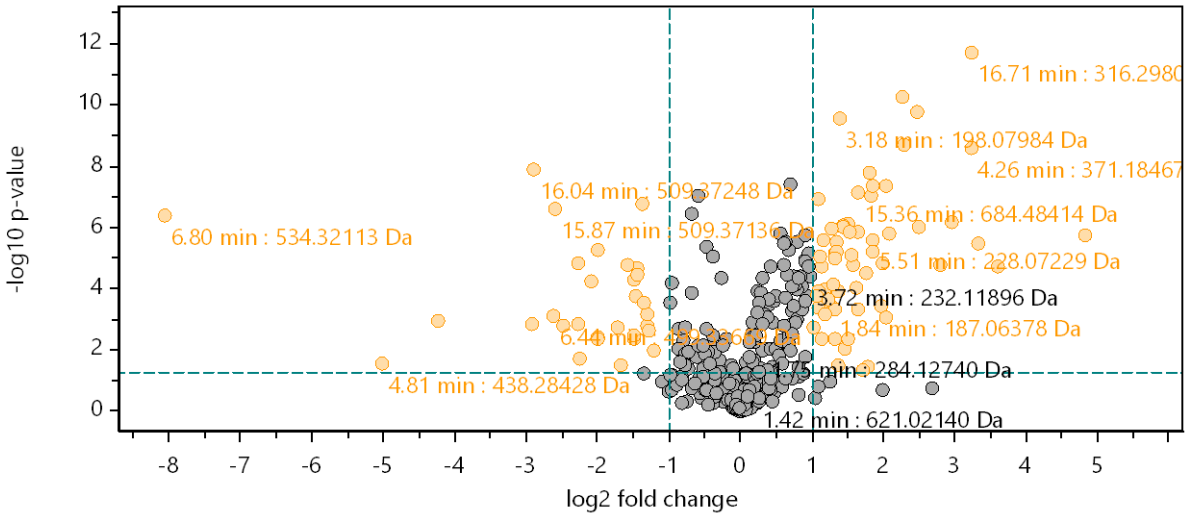


Figure S1: Volcano plot of significance (-log10 p-value) versus fold change (log2 fold change); data points appearing towards the top of the plot indicate high significance; data points towards the left side of the plot indicate upregulation and towards the right side downregulation of features during cultivation at 18 °C.

37 °C

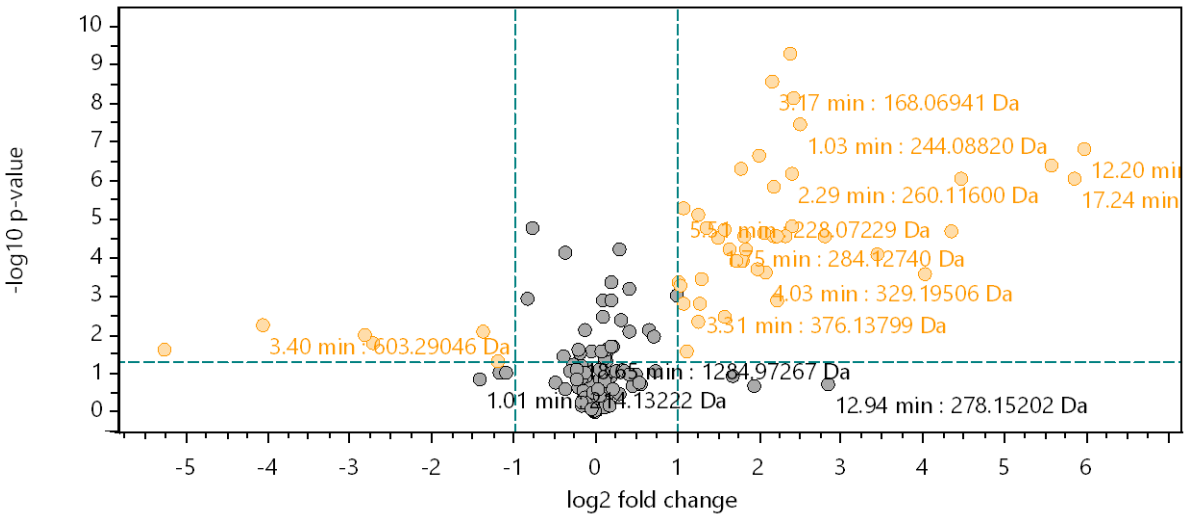


Figure S2: Volcano plot of significance (-log10 p-value) versus fold change (log2 fold change); data points appearing towards the top of the plot indicate high significance; data points towards the left side of the plot indicate upregulation and towards the right side downregulation of features during cultivation at 37 °C.

White light

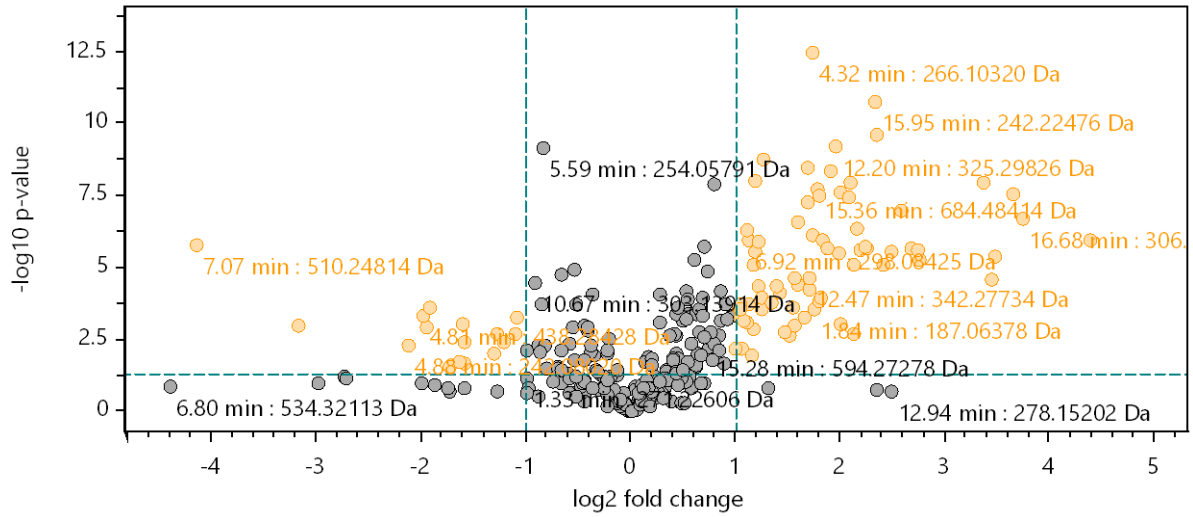


Figure S3: Volcano plot of significance (-log10 p-value) versus fold change (log2 fold change); data points appearing towards the top of the plot indicate high significance; data points towards the left side of the plot indicate upregulation and towards the right side downregulation of features during cultivation with white light.

Darkness

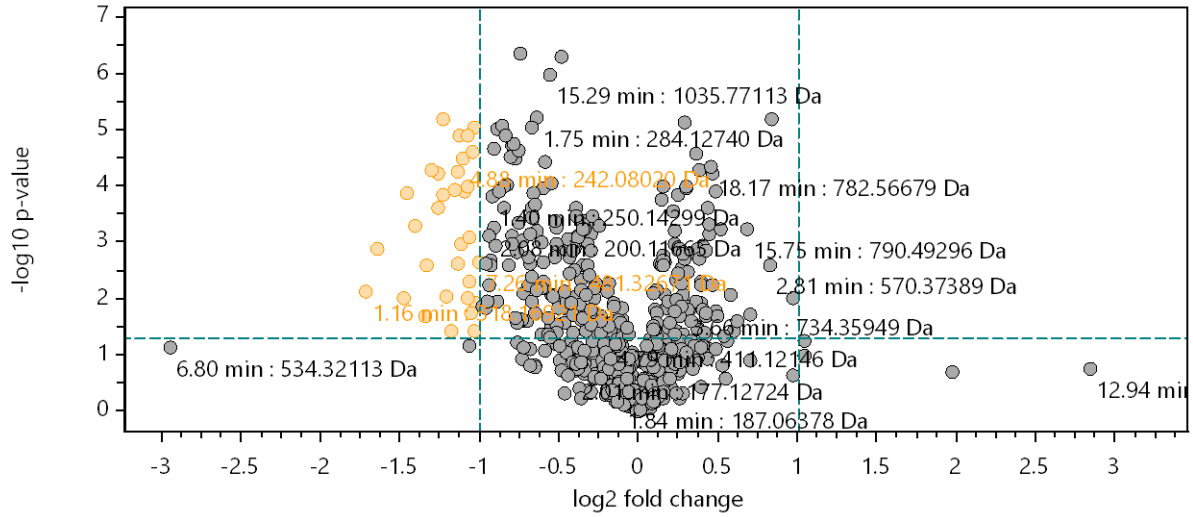


Figure S4: Volcano plot of significance (-log10 p-value) versus fold change (log2 fold change); data points appearing towards the top of the plot indicate high significance; data points towards the left side of the plot indicate upregulation and towards the right side downregulation of features during cultivation in darkness.

Oxygen limitation

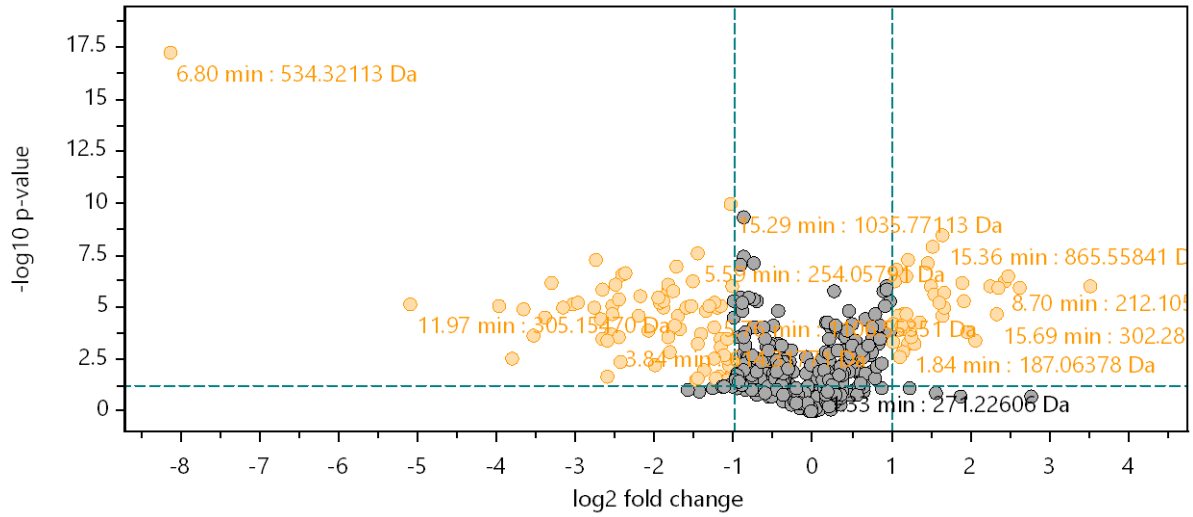


Figure S5: Volcano plot of significance (-log10 p-value) versus fold change (log2 fold change); data points appearing towards the top of the plot indicate high significance; data points towards the left side of the plot indicate upregulation and towards the right side downregulation of features during cultivation with oxygen limitation.

Rifampicin

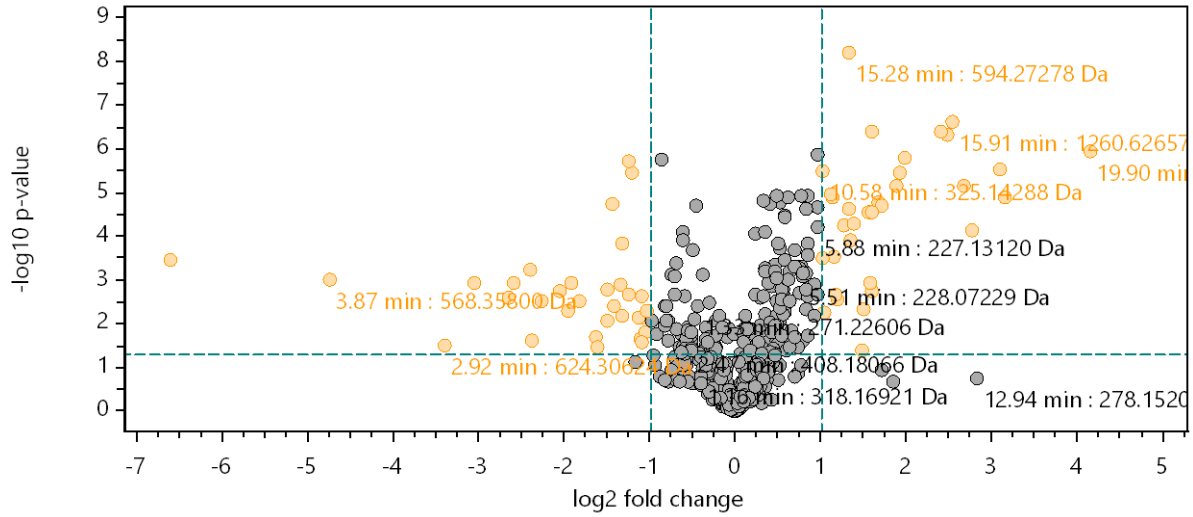


Figure S6: Volcano plot of significance (-log10 p-value) versus fold change (log2 fold change); data points appearing towards the top of the plot indicate high significance; data points towards the left side of the plot indicate upregulation and towards the right side downregulation of features during cultivation with addition of 50 µg/mL Rifampicin.

Trimethoprim

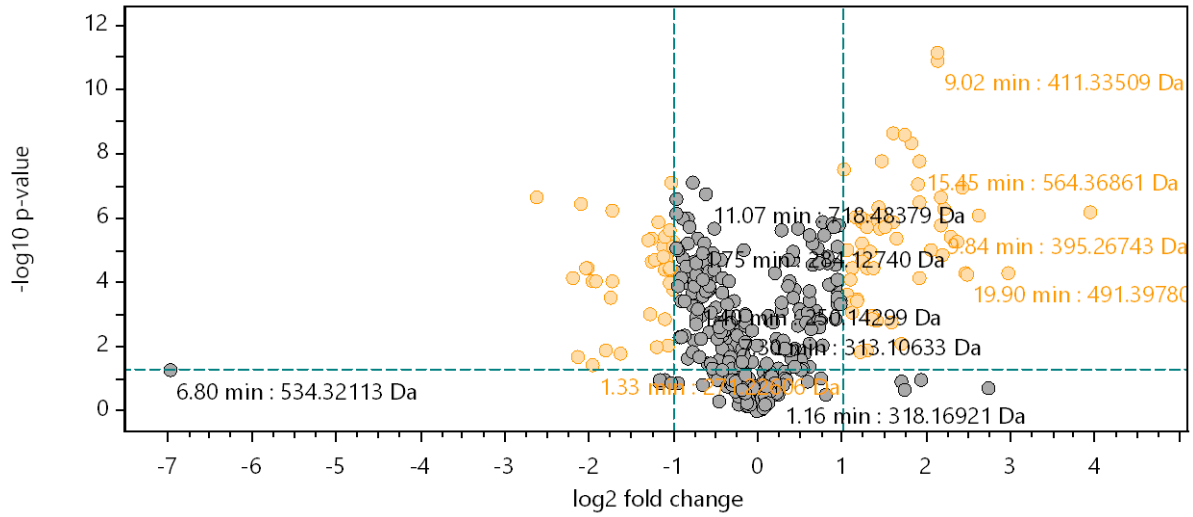


Figure S7: Volcano plot of significance (-log10 p-value) versus fold change (log2 fold change); data points appearing towards the top of the plot indicate high significance; data points towards the left side of the plot indicate upregulation and towards the right side downregulation of features during cultivation with addition of 25 µg/mL Trimethoprim.

Chloramphenicol

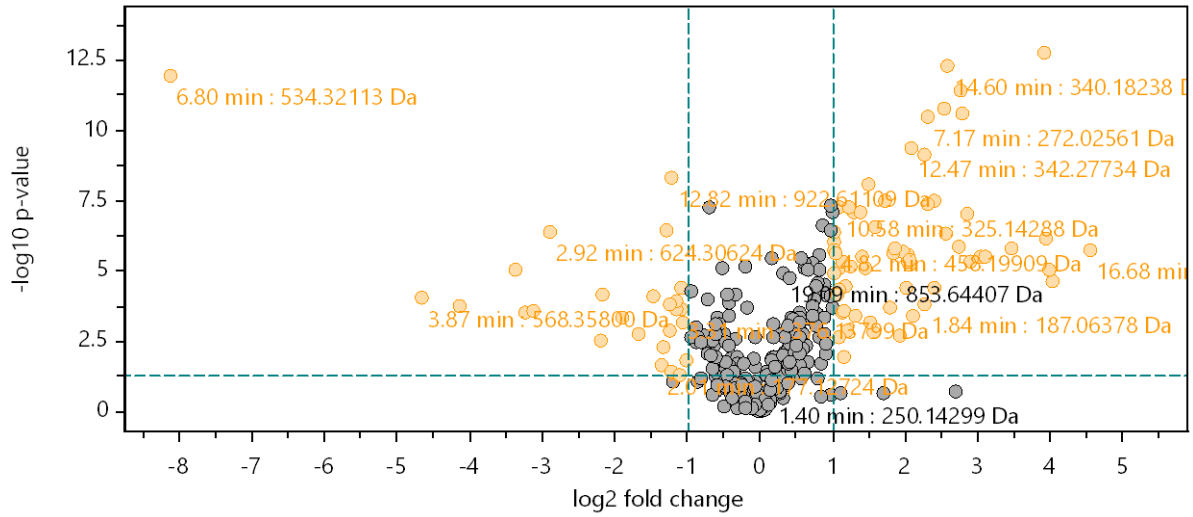


Figure S8: Volcano plot of significance (-log10 p-value) versus fold change (log2 fold change); data points appearing towards the top of the plot indicate high significance; data points towards the left side of the plot indicate upregulation and towards the right side downregulation of features during cultivation with addition of 50 µg/mL Chloramphenicol.

Soil extract

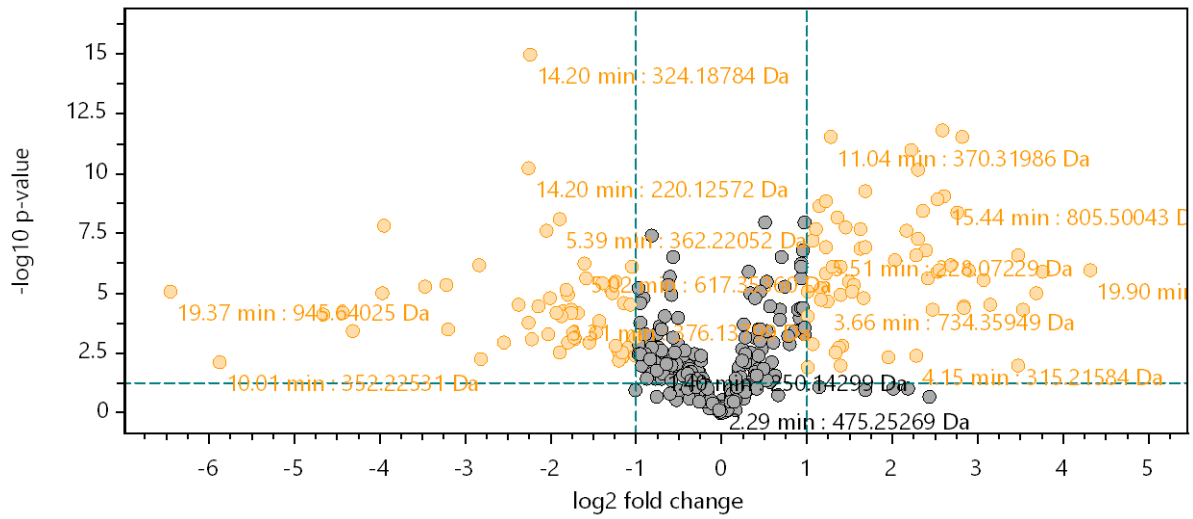


Figure S9: Volcano plot of significance (-log10 p-value) versus fold change (log2 fold change); data points appearing towards the top of the plot indicate high significance; data points towards the left side of the plot indicate upregulation and towards the right side downregulation of features during cultivation with addition of soil extract.

Lanthanum chloride

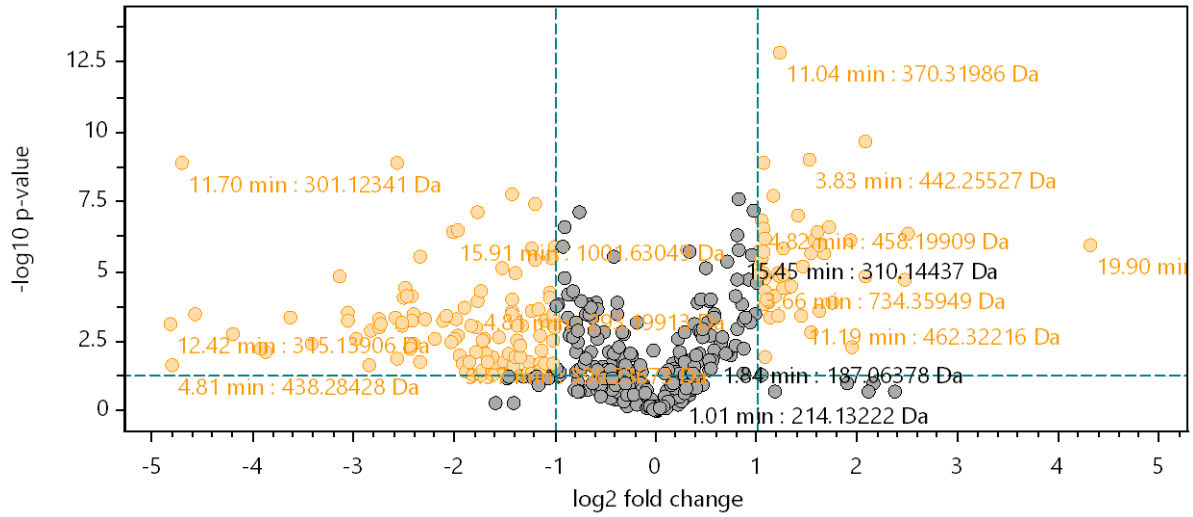


Figure S10: Volcano plot of significance (-log10 p-value) versus fold change (log2 fold change); data points appearing towards the top of the plot indicate high significance; data points towards the left side of the plot indicate upregulation and towards the right side downregulation of features during cultivation with addition of 0.5 mM lanthanum chloride.

Ethanol

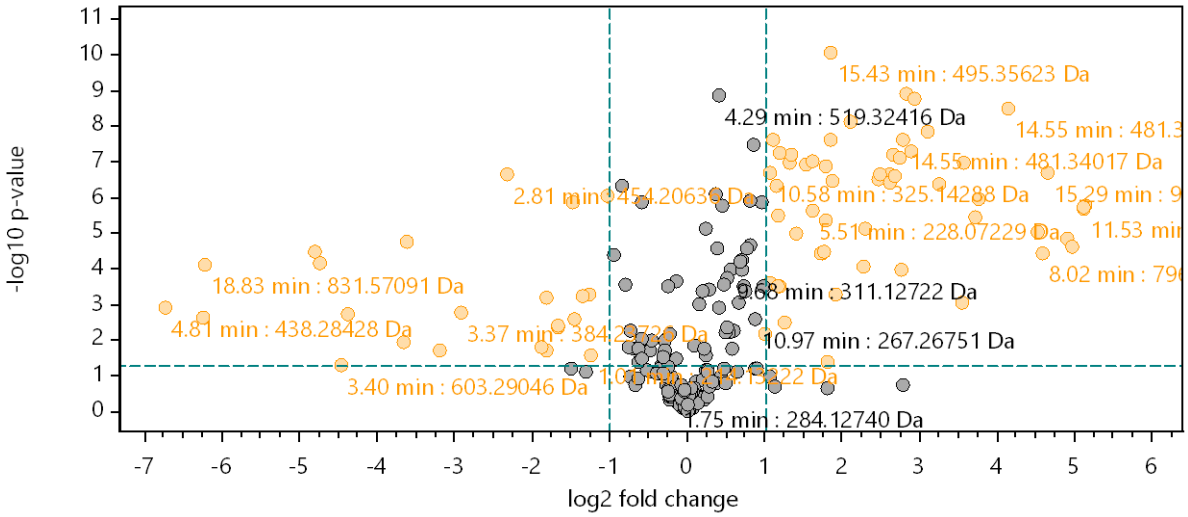


Figure S11: Volcano plot of significance (-log10 p-value) versus fold change (log2 fold change); data points appearing towards the top of the plot indicate high significance; data points towards the left side of the plot indicate upregulation and towards the right side downregulation of features during cultivation with addition of 2% ethanol.

S2.6 Cultivation and isolation

Soil extract generation

The soil sample from a pond in Schiffweiler, Germany was extracted at 40 °C, 200 bar and a total flowrate of 10 mL/min. The first extraction step was performed using 20% ethyl acetate as co-solvent along with CO_2, whereby in the second step 50% methanol was used for sufficient flushing. Both stages included an initial dynamic phase, a subsequent static extraction step and a final dynamic part for 5 min each. Additionally, methanol make-up flow was set to 3 mL/min. For small scale cultivation one 10 mL extraction vessel was used and for the large scale cultivation 25 of 25 mL extraction vessels.

Liquid liquid partitioning

The filtrated extract was dried and redissolved in methanol to be preseparated by liquid-liquid partitioning using three times hexane. Respective methanol phase was dried and redissolved in water to be extracted three times with chloroform and ethyl acetate each.

Size exclusion chromatography

Size exclusion chromatography was performed using a gravitation column with 3.5 cm diameter and 104 cm length (packing height around 75 cm). The stationary phase consisted of Sephadex LH 20 and mobile phase is methanol, whereby the flow rate was adapted to approximately 246 µL/min assuming a volume per drop of 61.5 µL, fractionating overnight.

Centrifugal partitioning chromatography

For proper separation, the solvent system hexane/acetone/water 5:4:1 with hexane/acetone subjected in forward mode and water/acetone in reverse mode was used. Compound **3** containing fractions eluted with reverse mode.

HPLC workup

Compound **1**: Mass spectrometric measurements were carried out on a Thermo Scientific ISQ EM system was used. Separation was achieved via a Waters XBridge Peptide BEH C18 OBD prep column (250 x 10 mm, 5 µm) applying a gradient from 50 to 95% acetonitrile in water with 0.1 % formic acid for 29.5 min at a flow rate of 5 mL/min. For each injection the column was equilibrated for 5 min at 5% and flushed with 95% acetonitrile in water with 0.1 % formic

acid for 2 min. Column thermostat was maintained at 45 °C. Single ion monitoring was utilized for detection of the target mass at *m/z* 779.5 with 10 V CID voltage.

Compound **2**: Mass spectrometric measurements were done on a Bruker Daltonics HCT plus ion trap system. Separation was reached with a Waters XBridge Peptide BEH C18 column (250 x 4.6 mm, 5 µm) at a flow rate of 6.0 mL/min with a mobile phase including (A) 0.1 % formic acid in ddH_2O and (B) 0.1 % formic acid in acetonitrile. Following linear gradient was applied: 0-2 min hold at 5 % B, 2-2.5 min increasing to 50 % B, 2.5-32 min increasing to 95 %, reequilibration at 5 % B for 5 min.

Compound **3**: Mass spectrometric measurements were performed on a Thermo Scientific ISQ EM system was used. Pure compound isolation was achieved with a Waters Peptide BEH C18 OBD prep column (250 x 10 mm, 5 µm) column with a mobile phase including (A) 0.1 % formic acid in ddH_2O and (B) 0.1 % formic acid in acetonitrile at 5 mL/min and following linear gradient: 0-2 min hold at 5 % B, 2-32 min increasing to 95 %, reequilibration at 5 % B for 5 min. Single ion monitoring was adapted to *m/z* 635.5 with 10 V CID voltage.

Stereochemical analysis of 1 using Marfey's analysis

For the hydrolysis step, 0.1 mg of **1** was dissolved in 100 µL of 3 M HCl for 45 °C at 110 °C. The dry residue was split in two 50 µL aliquots in ddH_2O. To both samples 60 µL of 1 M $NaHCO_3$ as well as 20 µL of the derivatization agent (1 % 1-fluroro-2,4-dinitrophenyl-5-leucine-amide solution in acetone [D-FDLA and L-FDLA]) were added and incubation was carried out at 40 °C and 700 rpm for 2 h. The reaction was quenched by addition of 60 µL 3 M HCl and diluted with 150 µL of both methanol and acetonitrile to be measured on the maXis 4G UHR-TOF (Bruker Daltonics, Billerica, MA, USA) coupled to a Dionex Ultimate 3000 SL system (Thermo Fisher Scientific, Waltham, MA, USA).

NMR analysis

The NMR data set consisting of 1 and 2D spectra for structure elucidation of **1**, **2** and **3** were recorded on a Bruker Avance III (Ultrashield) 500 MHz spectrometer equipped with a 5 mm TCI cryoprobe and respective measurements for dehydroxynocardamine B were conducted on a Bruker Avance III (Ascend) 700 MHz spectrometer. All measurements were carried out using standard pulse programs. Observed chemical shifts (δ) are specified in ppm and coupling constants (J) in Hz. For dehydroxynocardamine B, **2** and **3** respective spectra were calibrated on the intrinsic chemical shifts of the solvent at δ(1H) = 3.31 ppm and δ(13C) = 49.2 ppm as well as δ(1H) = 2.50 ppm and δ(13C) = 39.5 ppm for **1**.

S2.7 Production of dehydroxynocardamine B

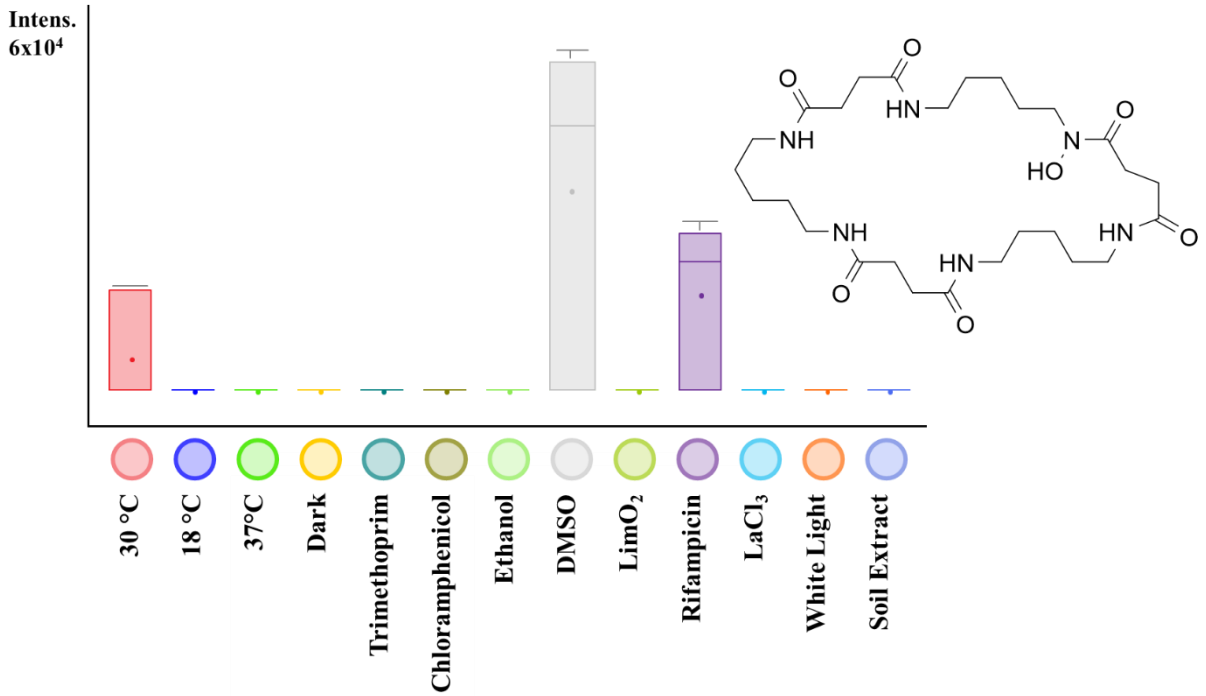


Figure S12: Box-and-whisker diagram showing production rates of a specific feature among the different cultivation conditions: dehydroxynocardamine B was only produced in small amounts in two biological replicates for cultivation at 30 °C, DMSO addition and Rifampicin, whereby DMSO supplementation was a control for the addition of the three antibiotics from DMSO stock solutions; median and mean values and 25% and 75% quartiles illustrated within box with grey whiskers representing lower and upper extremes.

S2.8 Purification procedure

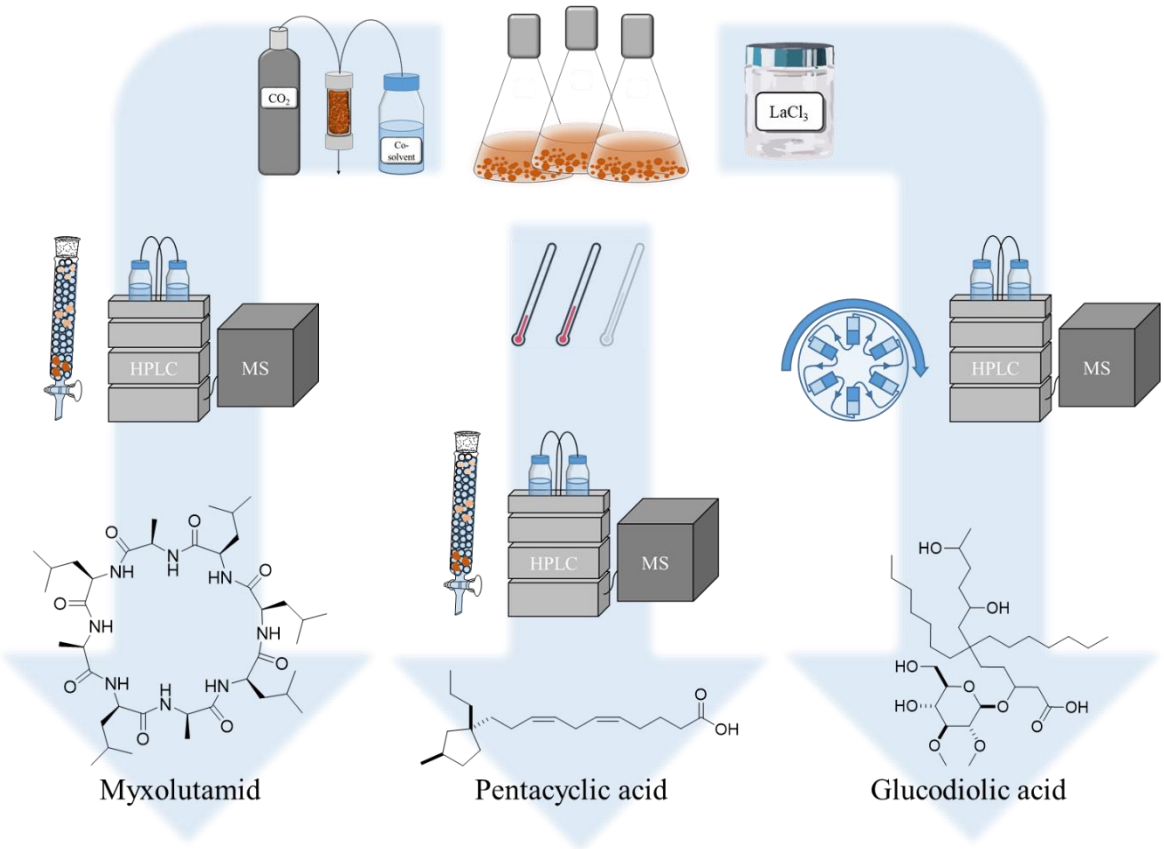


Figure S13: Workflow for purification of the three target compounds myxolutamid, pentacyclic acid and glucodiolic acid using three different cultivation conditions and two different preseparation steps before semi-preparative workup and de novo structure elucidation using NMR.

S2.9 Structure determination

S2.9.1 Dehydroxynocardamine B.

Structure elucidation of dehydroxynocardamine B. HRESIMS uncovers an $[M+H]^+$ peak for dehydroxynocardamine B at *m/z* 569.3666 matching the molecular formula $C_{27}H_{49}N_6O_7$ (*m/z* calcd for $[M+H]^+$ 569.3657) featuring 7 double-bond equivalents (DBE). Inspection of ^{1}H-NMR in methanol-d_4 shows seven methylene signals at $\delta(^1H)$ = 3.60 (2H, t, *J* = 6.49 Hz), $\delta(^1H)$ = 3.16 (10H, q, *J* = 6.49 Hz), $\delta(^1H)$ = 2.78 (2H, t, *J* = 6.96 Hz), $\delta(^1H)$ = 2.47 (10H, m), $\delta(^1H)$ = 1.63 (2H, quint, *J* = 6.92 Hz), $\delta(^1H)$ = 1.50 (10H, m) and $\delta(^1H)$ = 1.35 (6H, m) ppm. Signal integration in the ^{1}H-spectrum as well as COSY and HMBC correlations point towards a macrocyclic structure with a high symmetry. The low field shift methylene signals at $\delta(^1H)$ = 3.60 and 2.78 ppm indicate the close neighborhood to the sole hydroxamate group in contrast to the remaining five amide functions. The carbonyl functions are separated through either a dimethylene-linker or a pentamethylene-linker. All ^{1}H-shifts are comparable to the already published dehydroxynocardamine with a deviation of 0.03 ppm [18].

Table S6: NMR spectroscopic data of dehydroxynocardamine B in methanol-d_4 at 700/175 MHz.

#	δ ^{13}C [ppm]	δ ^{1}H [ppm], mult (*J* [Hz])	COSY	HMBC
1,4,1',4',1'',4''	174.3	-	-	-
2,3,3',2'',3''	31.8	2.47, m	5,5',9',5'',9''	1,2,3,4,1',2',3', 4',1'',2'',3'',4''
5,5',9',5'',9''	39.8	3.16, q (6.49)	2,3,6,3',6',8', 2'',3'',6'',8''	1,4,6,7,1',4',6', 7',8',1'',4'',6'', 7'',8''
6,6',8',6'',8''	29.5	1.50, m	5,7,5',7',9',5'', 7'',9''	5,6,7,8,5',6',7', 8',9',5'',6'',7'', 8'',9''
7,7',7''	24.6	1.35, m	6,6',8',6'',8''	5,6,5',6',8',9', 5'',6'',8'',9''
8	26.9	1.63, quint (6.92)	9	6,9,6',8',6'',8''
9	48.2	3.60, t (6.49)	8	1,4,8,1',4',1'', 4''
2'	28.6	2.78, t (6.96)	3'	1,2,3,4,1',3',4', 1'',2'',3'',4''

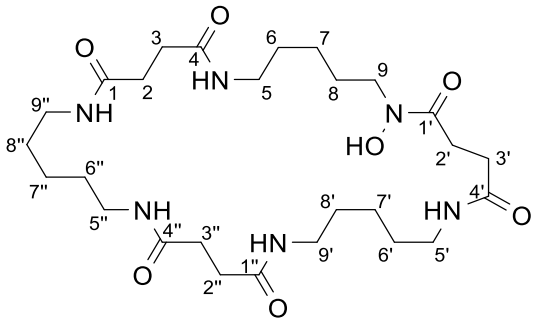


Figure S14: Chemical structure of dehydroxynocardamine B with numbering of relevant atoms.

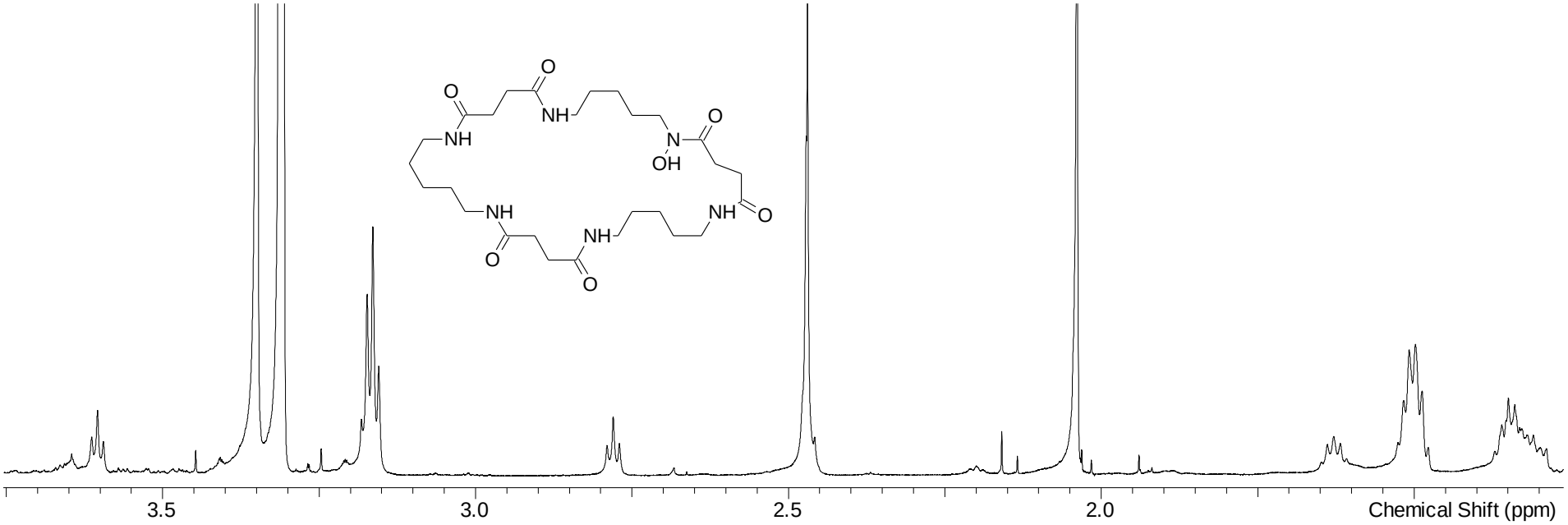
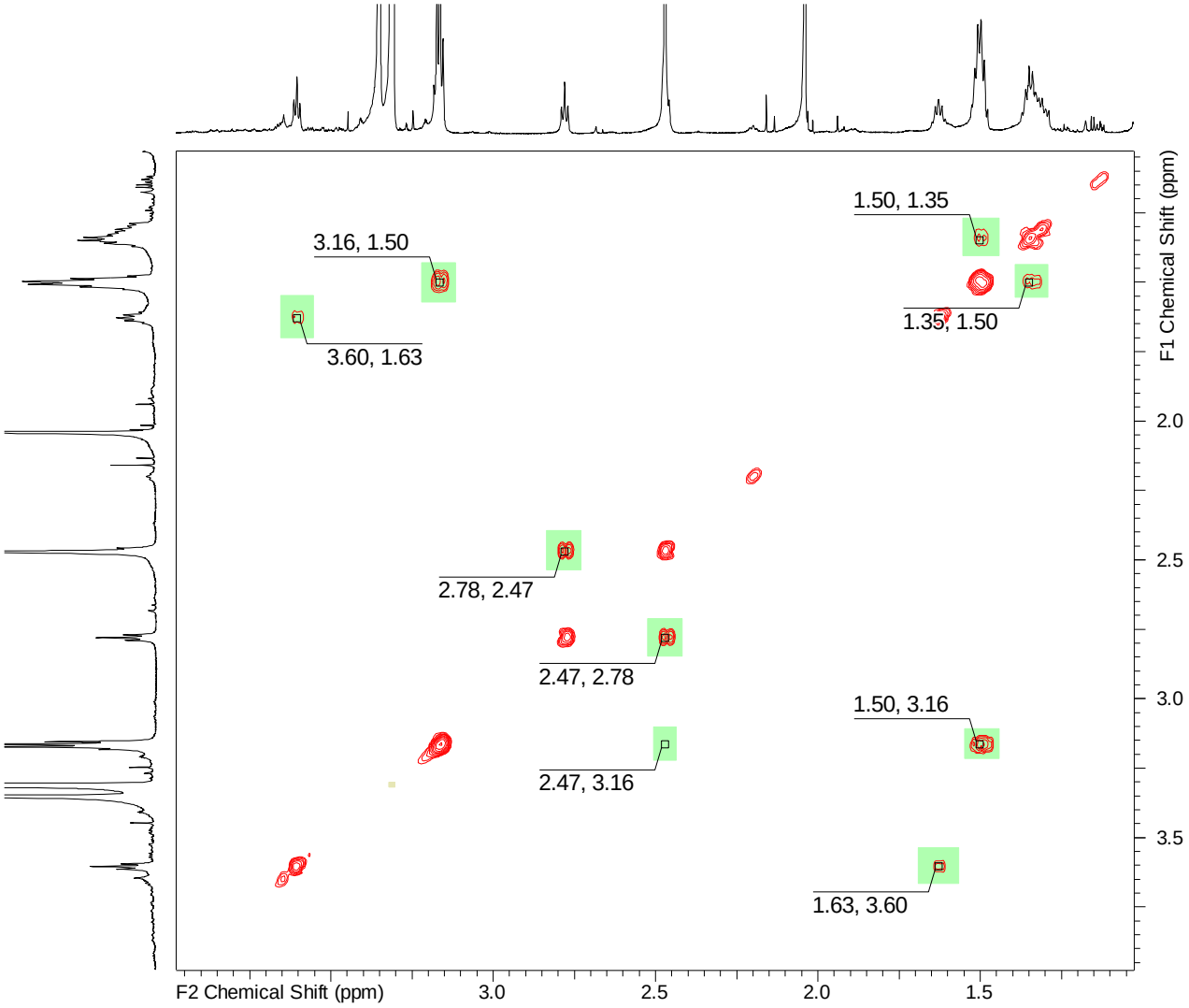


Figure S15: ^{1}H-spectrum of dehydroxynocardamine B in methanol-d_4 at 700 MHz.



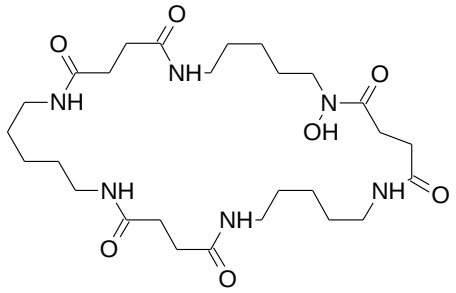
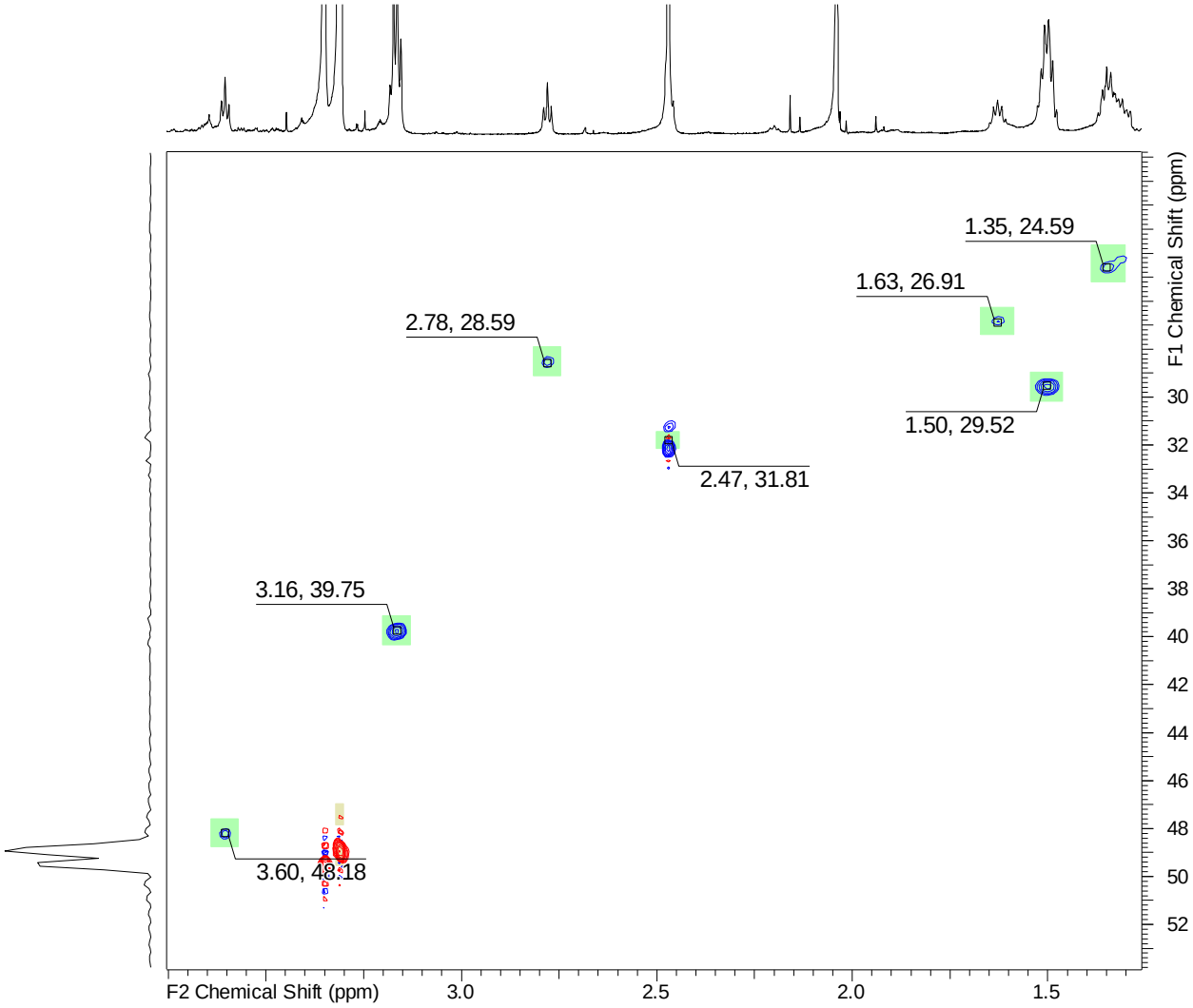


Figure S16: COSY-spectrum of dehydroxynocardamine B in methanol-d_4 at 700 MHz.



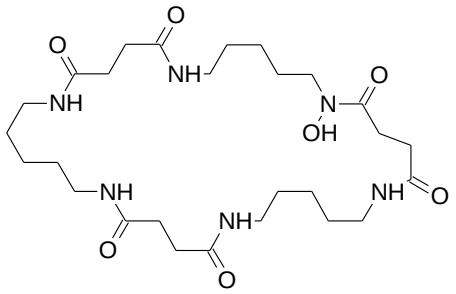
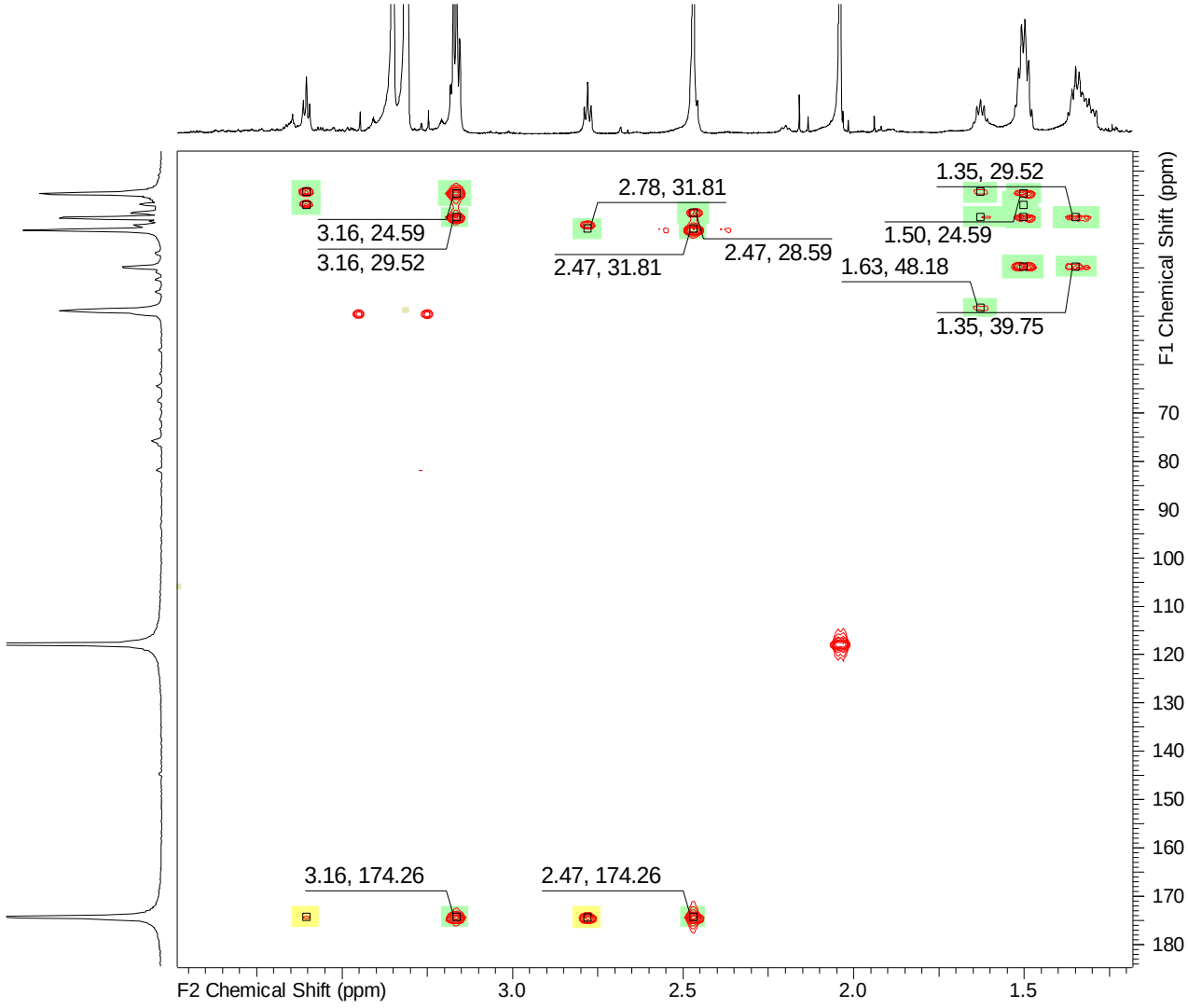


Figure S17: HSQC-spectrum of dehydroxynocardamine B in methanol-d_4 at 700 MHz (^{1}H)/175 MHz (^{13}C).



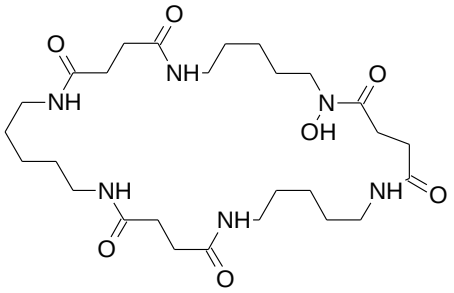


Figure S18: HMBC-spectrum of dehydroxynocardamine B in methanol-d_4 at 700 MHz (^{1}H)/175 MHz (^{13}C).

Further derivatives of the nocardamine family

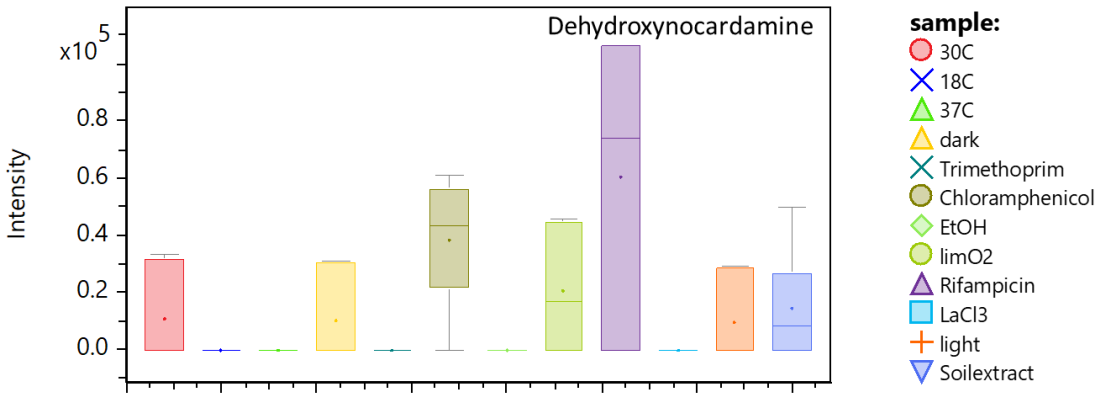


Figure S19: Box-and-whisker diagram showing intensity levels of m/z 585.359 among the different cultivation conditions: Dehydroxynocardamine A is produced under several conditions with best production through addition of rifampicin (purple); median and mean values and 25% and 75% quartiles illustrated within box with grey whiskers representing lower and upper extremes.

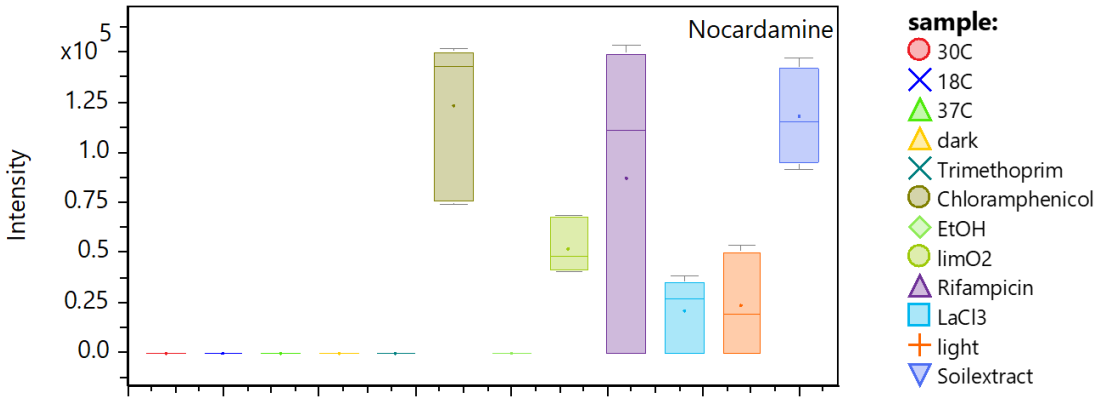


Figure S20: Box-and-whisker diagram showing intensity levels of m/z 601.355 among the different cultivation conditions: nocardamine is produced under several conditions with best production through addition of chloramphenicol (olive), rifampicin (purple) and soil extract (light purple); median and mean values and 25% and 75% quartiles illustrated within box with gray whiskers representing lower and upper extremes.

S2.9.2 Myxolutamid

Structure elucidation of myxolutamid. HRESIMS exhibits an $[M+H]^+$ peak for compound **1** at *m/z* 779.5384 associated with the molecular formula $C_{39}H_{71}N_8O_8$ (*m/z* calcd for $[M+H]^+$ 779.5389) featuring 9 double-bond equivalents (DBE). ^{1}H-NMR and HSQC experiments in DMSO-d_6 display characteristic signals for an octapeptid with eight α-proton signals at $\delta(^1H)$ = 4.64 (1H, m) $\delta(^{13}C)$ = 49.2, $\delta(^1H)$ = 4.37 (1H, m) $\delta(^{13}C)$ = 49.2, $\delta(^1H)$ = 4.24 (1H, brdd, *J* = 11.63, 3.85 Hz) $\delta(^{13}C)$ = 50.2, $\delta(^1H)$ = 4.19 (1H, dd, *J* = 11.44, 2.98 Hz) $\delta(^{13}C)$ = 50.9, $\delta(^1H)$ = 4.06 (1H, m) $\delta(^{13}C)$ = 55.2, $\delta(^1H)$ = 3.98 (1H, m) $\delta(^{13}C)$ = 53.1, $\delta(^1H)$ = 3.95 (1H, m) $\delta(^{13}C)$ = 51.3 and $\delta(^1H)$ = 3.76 (1H, m) $\delta(^{13}C)$ = 52.3 ppm. Moreover, four diastereotopic methylene signals can be observed at $\delta(^1H)$ = 1.74, 1.35 (2H, m) $\delta(^{13}C)$ = 39.8, $\delta(^1H)$ = 1.68, 1.41 (2H, m) $\delta(^{13}C)$ = 41.4, $\delta(^1H)$ = 1.65, 1.55 (2H, m) $\delta(^{13}C)$ = 38.9 and $\delta(^1H)$ = 1.55, 1.35 (2H, m) $\delta(^{13}C)$ = 40.0 ppm as well as one methylene signals at $\delta(^1H)$ = 1.62 (2H, m) $\delta(^{13}C)$ = 40.4 ppm. Five methine signals are located at $\delta(^1H)$ = 1.75 (1H, m) $\delta(^{13}C)$ = 24.6, $\delta(^1H)$ = 1.73 (1H, m) $\delta(^{13}C)$ = 24.5, $\delta(^1H)$ = 1.57 (1H, m) $\delta(^{13}C)$ = 24.8, $\delta(^1H)$ = 1.52 (1H, m) $\delta(^{13}C)$ = 23.7 and $\delta(^1H)$ = 1.44 (1H, m) $\delta(^{13}C)$ = 24.6 ppm. The HSQC spectrum shows three methyl signals at $\delta(^1H)$ = 1.27 (3H, d , *J* = 7.40 Hz) $\delta(^{13}C)$ = 16.6, $\delta(^1H)$ = 1.24 (3H, m) $\delta(^{13}C)$ = 16.4 and $\delta(^1H)$ = 1.10 (3H, d, *J* = 6.41 Hz) $\delta(^{13}C)$ = 16.1 ppm and additionally at $\delta(^1H)$ = 0.91-0.75 ppm (22H, m) signals can be found in the ^{1}H spectrum corresponding to ten methyl groups. Finally, eight amid proton signals are depicted at $\delta(^1H)$ = 8.89 (1H, brs), $\delta(^1H)$ = 8.61 (1H, brs), $\delta(^1H)$ = 8.58 (1H, brs), $\delta(^1H)$ = 7.50 (1H, brd, *J* = 7.93 Hz), $\delta(^1H)$ = 7.46 (1H, brd, *J* = 9.92 Hz), $\delta(^1H)$ = 7.42 (1H, brs), $\delta(^1H)$ = 7.27 (1H, brs) and $\delta(^1H)$ = 7.01 (1H, m) ppm with an integration of one each, which points towards a macrocycle comprising eight amino acids. Combination of COSY and HMBC correlations uncover the presence of five leucines as well as three alanines, whereby the amino acid sequence was predicted via HMBC correlations from the respective α-protons to the adjacent carbonyl groups. It is worth noting that only one correlation from the α-proton at $\delta(^1H)$ = 4.06 ppm to the carbonyl group at $\delta(^{13}C)$ = 174.1 ppm is missing. However, the predicted sum formula, DBEs and number of amide protons with their integration of one strongly point towards a cyclic scaffold. Thus, starting from the α-proton signal with the most downfield shift at $\delta(^1H)$ = 4.64 ppm the sequence of the macrocyclic myxolutamid is found to be Leu-Leu-Leu-Ala-Leu-Ala-Leu-Ala.

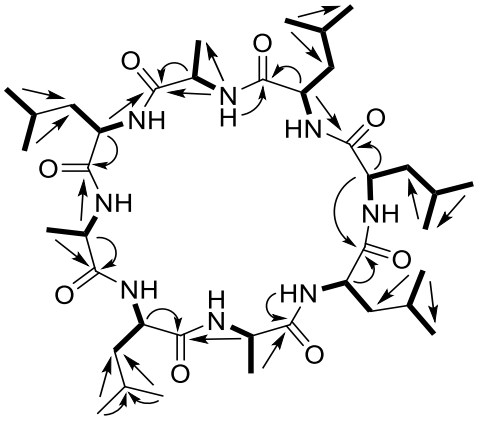


*Figure S21: Chemical structure of **1** with selected COSY (bold) and HMBC (arrows) correlations.*

*Table S7: NMR spectroscopic data of **1** in DMSO-d_6 at 500/125 MHz.*

#	δ ^{13}C [ppm]	δ ^{1}H [ppm], mult (*J* [Hz])	COSY	HMBC
1	174.5	-	-	-
α1	49.2	4.64, m	β_1, NH_1	1,2
β1	41.4	1.68, 1.41, m	α_1, γ_1, $\delta_{1'}$	1
γ1	24.6	1.44, m	β_1, δ_1, $\delta_{1'}$	-
δ1	23.7	0.81, m	γ_1	β_1, $\delta_{1'}$
δ1'	21.1	0.77, m	β_1, γ_1	-
NH1	-	7.01, m	α_1	-
2	171.8	-	-	-
α2	53.1	3.98, m	β_2, NH_2	1,3
β2	39.8	1.74, 1.35, m	α_2, γ_2, δ_2	α_2
γ2	24.5	1.73, m	β_2, δ_2, $\delta_{2'}$	-
δ2	21.1	0.83, m	β_2, γ_2	γ_2, $\delta_{2'}$
δ2'	23.3	0.89, m	γ_2	β_2
NH2	-	7.42, brs	α_2	-
3	172.3	-	-	-
α3	50.2	4.24, brdd (11.63, 3.85)	β_3, NH_3	1
β3	40.4	1.62, m	α_3	α_3
γ3	24.6	1.75, m	δ_3, $\delta_{3'}$	-
δ3	21.1	0.83, m	γ_3	β_3, $\delta_{3'}$
δ3'	23.3	0.89, m	γ_3	-
NH3	-	7.46, brd (9.92)	α_3	4
4	172.7	-	-	-
α4	51.3	3.95, m	β_4, NH_4	4, β_4, 5
β4	16.6	1.27, d (7.40)	α_4	4, α_4
NH4	-	8.62, brs	α_4	-
5	175.6	-	-	-
α5	55.2	4.06, m	β_5, γ_5, NH_5	5
β5	40.0	1.55, 1.35, m	α_5, δ_5, $\delta_{5'}$	α_5

γ_5	24.8	1.57, m	-	5
δ_5	23.7	0.81, m	β_5	β_5, γ_5
$\delta_{5'}$	23.5	0.87, m	β_5	β_5, γ_5
NH_5	-	8.89, brs	α_5	-
6	174.1	-	-	-
α_6	49.2	4.37, m	β_6, NH_6	6, 7
β_6	16.1	1.10, d (6.41)	α_6	6
NH_6	-	7.27, brs	α_6	-
7	171.8	-	-	-
α_7	50.9	4.19, dd (11.44, 2.98)	β_7, NH_7	7, 8
β_7	38.9	1.65, 1.55, m	α_7, $\delta_{7'}$	α_7
γ_7	23.7	1.52, m	δ_7, $\delta_{7'}$	-
δ_7	23.7	0.81, m	γ_7	β_7
$\delta_{7'}$	23.5	0.85, m	β_7, γ_7	β_7, γ_7
NH_7	-	7.50, brd (7.93)	α_7	8
8	173.3	-	-	-
α_8	52.3	3.76, m	β_8, NH_8	8, β_8
β_8	16.4	1.24, m	α_8	8, α_8
NH_8	-	8.58, brs	α_8	8, α_8, β_8, 1

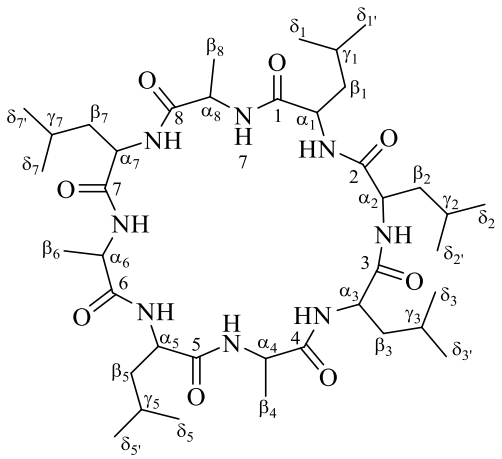


*Figure S22: Chemical structure of **1** with numbering of relevant atoms.*

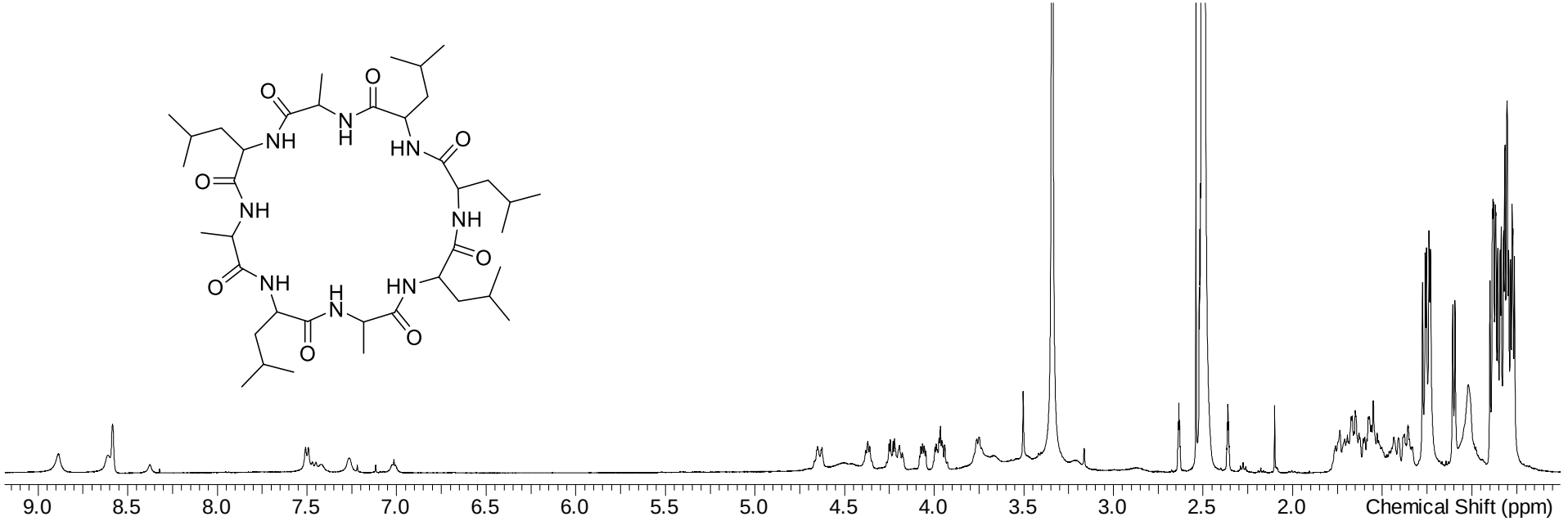
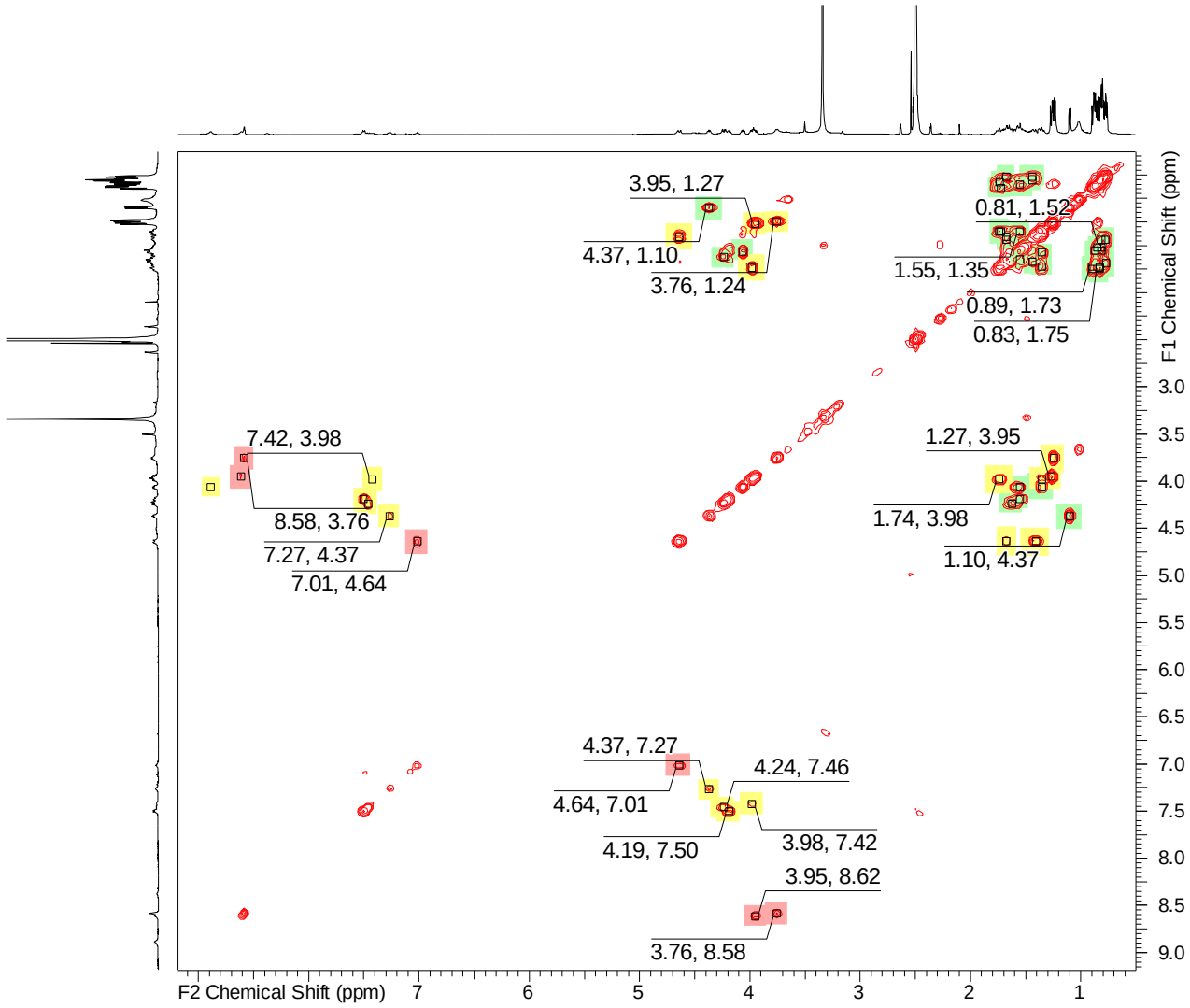


*Figure S23: ^{1}H-spectrum of **1** in DMSO-d_6 at 500 MHz.*



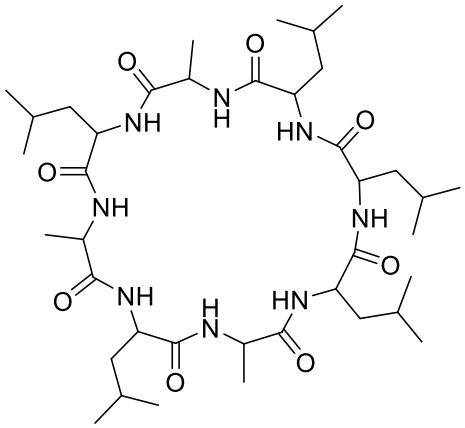
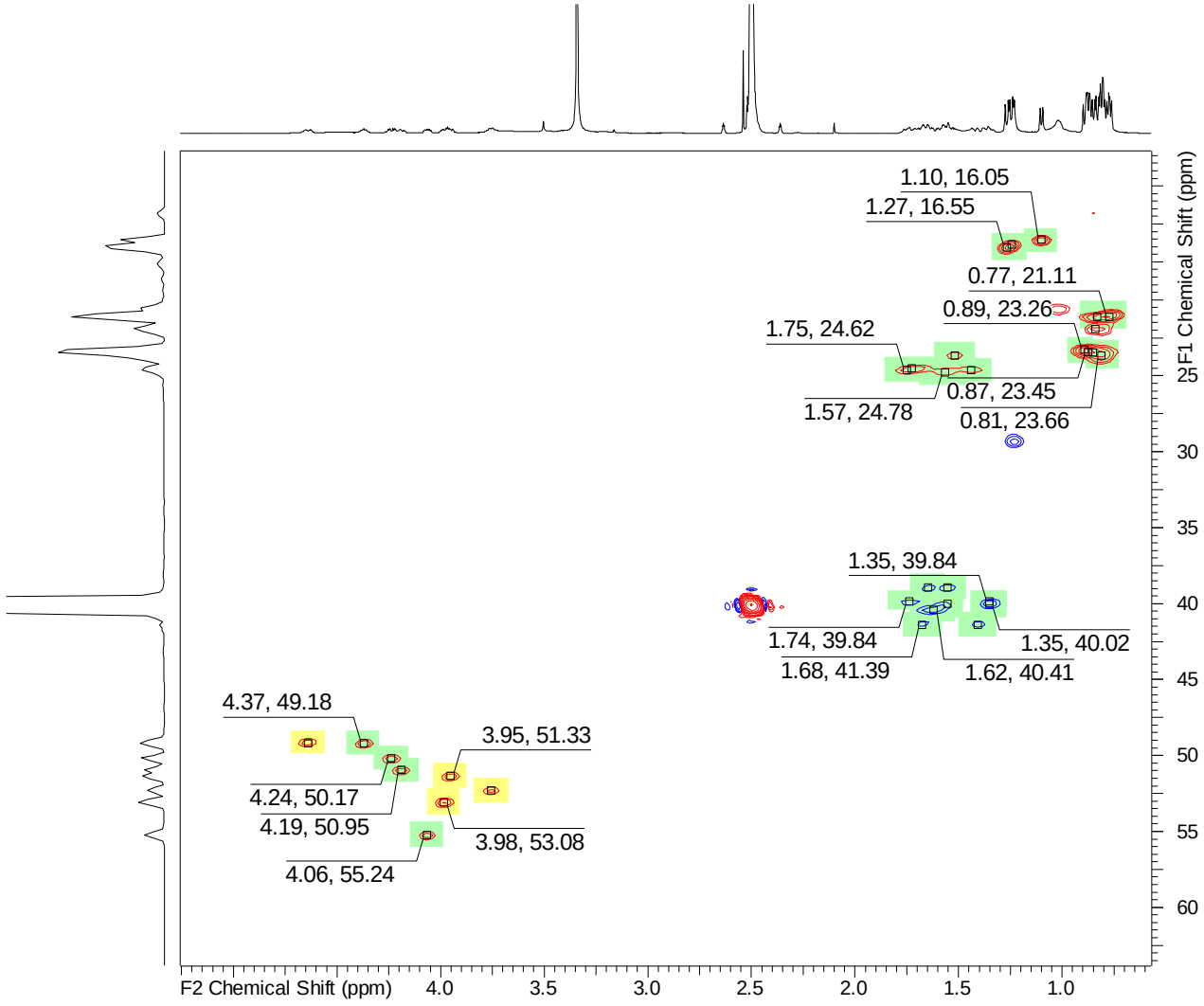


*Figure S24: COSY-spectrum of **1** in DMSO-d_6 at 500 MHz.*



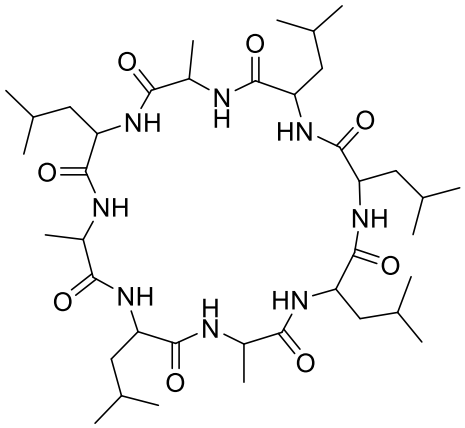
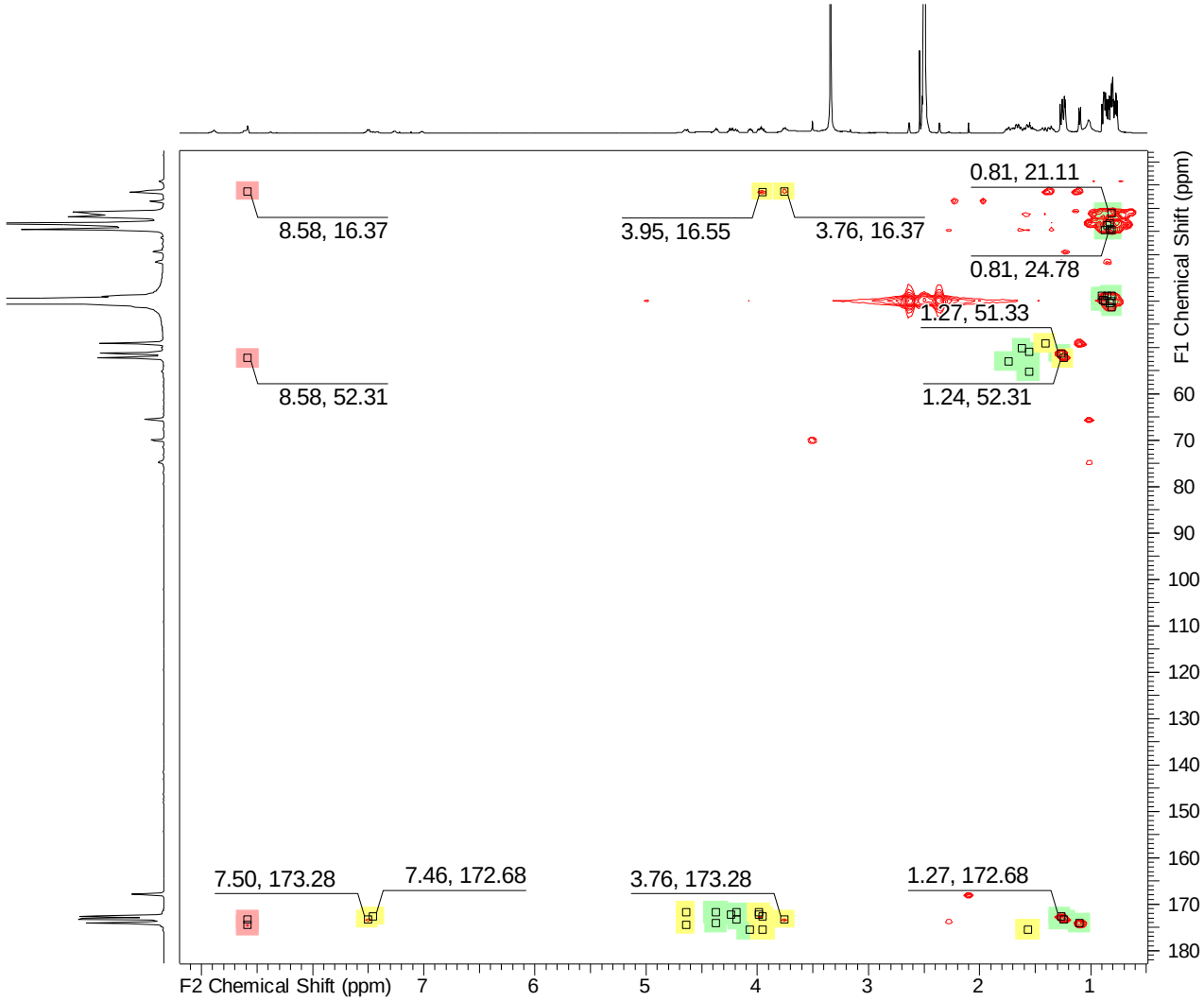


*Figure S25: HSQC-spectrum of **1** in DMSO-d_6 at 500 MHz (^{1}H)/125 MHz (^{13}C).*



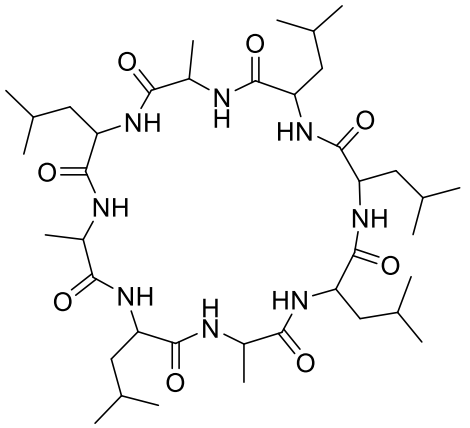


*Figure S26: HMBC-spectrum of **1** in DMSO-d_6 at 500 MHz (^{1}H)/125 MHz (^{13}C).*

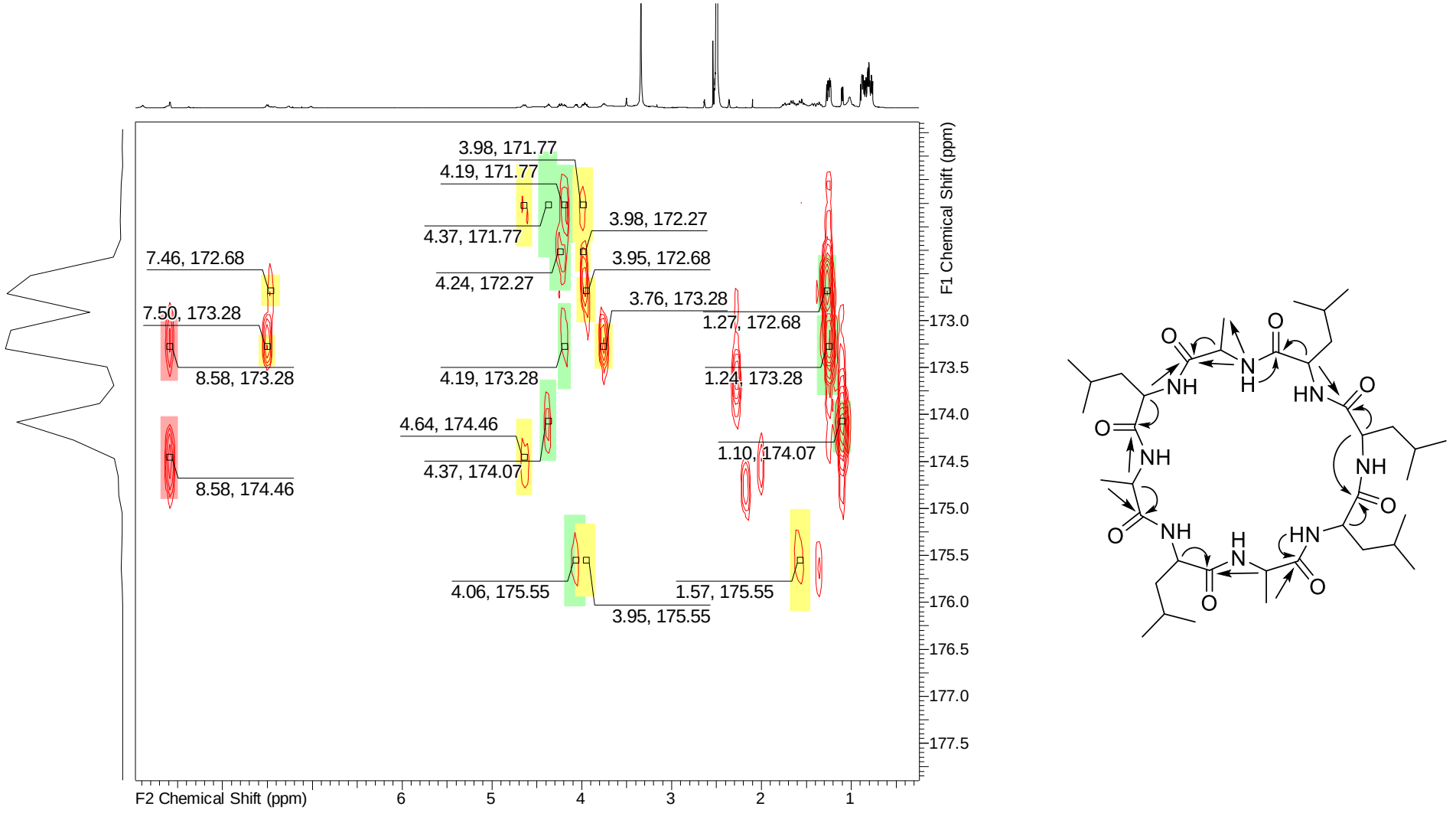


*Figure S27: Zoomed in HMBC-spectrum of **1** in DMSO-d_6 at 500 MHz (^{1}H)/125 MHz (^{13}C) for amino acid sequence determination.*

MS^2 measurement for myxolutamid

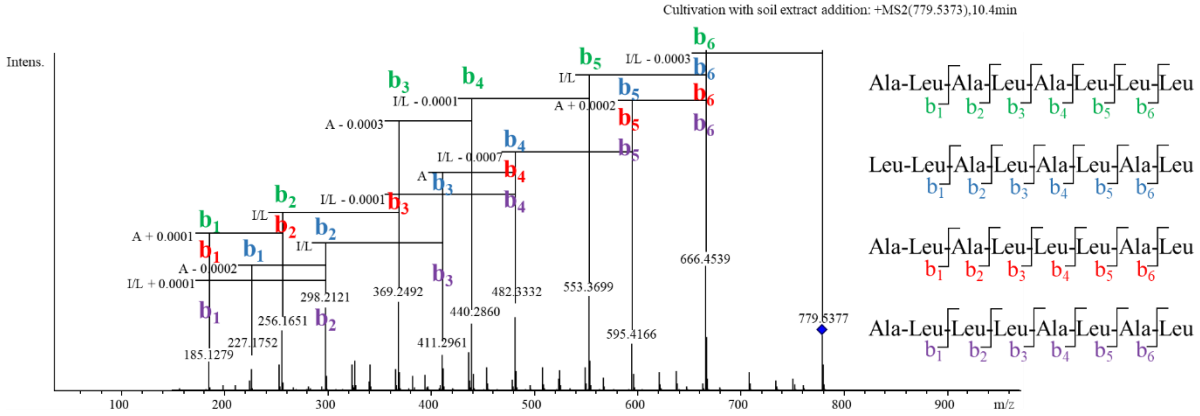


*Figure S28: MS^2 fragment spectrum for **1** with a precursor ion at m/z 779.537 with annotated amino acid eliminations after ring opening comprising isoleucine/leucine and alanine (b-fragments series in different colors), predicted amino acid sequence is Ala-Leu-Ala-Leu-Ala-Leu-Leu-Leu.*

Marfey's analysis myxolutamid

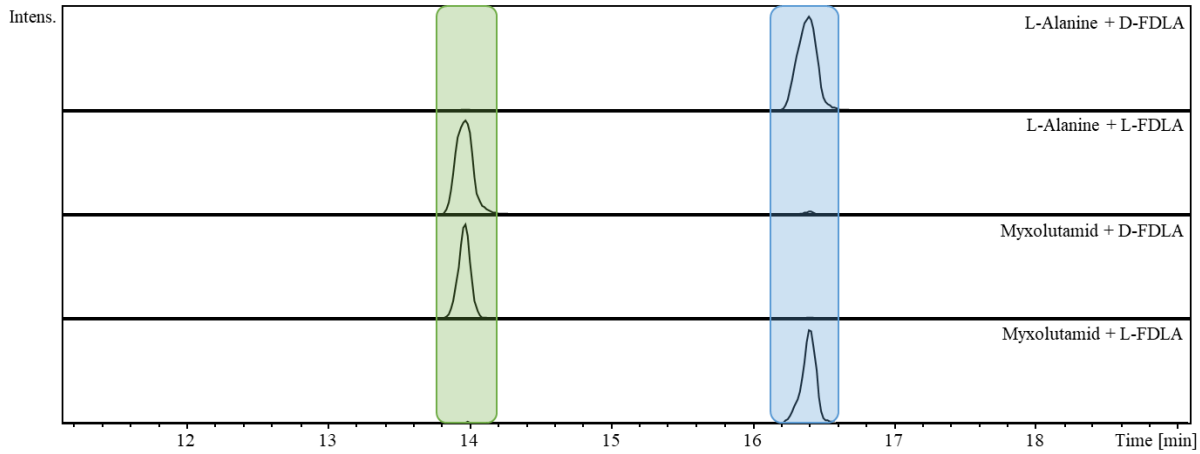


*Figure S29: Marfey's derivatization of reference L-Alanine derivatized with both D-FDLA and L-FDLA (upper 2 chromatograms) and **1** with both D-FDLA and L-FDLA (lower 2 chromatograms) for retention time comparison represented as EICs of m/z 384.151; D+L-FDLA/L+D-FDLA pointing towards contrary configuratios of the amino acid in comparison to FDLA highlighted in blue and L+L-FDLA/D+D-FDLA pointing towards same configuration of the amino acid in comparison to FDLA in green.*

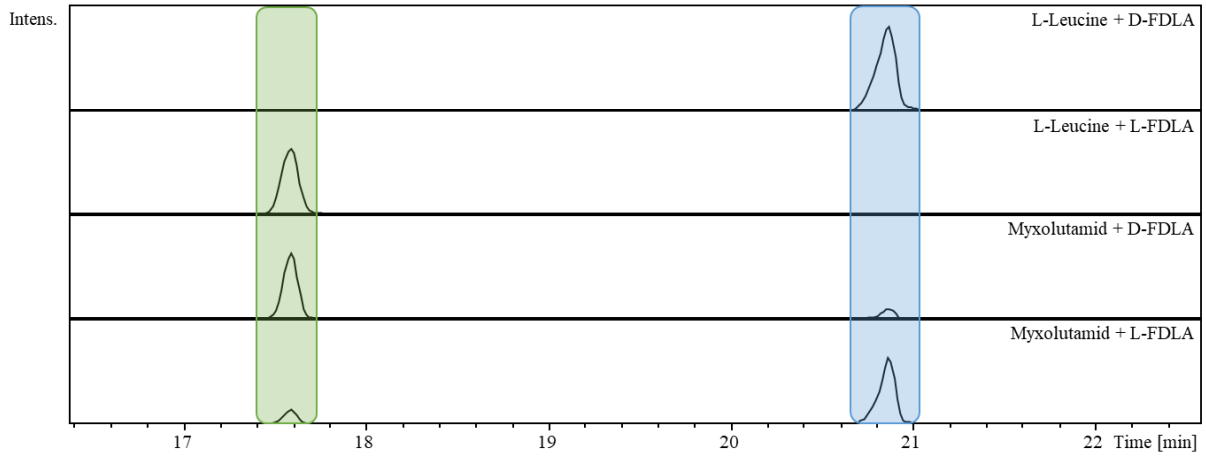


*Figure S30: Marfey's derivatization of reference L-Leucine derivatized with both D-FDLA and L-FDLA (upper 2 chromatograms) and **1** with both D-FDLA and L-FDLA (lower 2 chromatograms) for retention time comparison represented as EICs of m/z 426.198; D+L-FDLA/L+D-FDLA pointing towards contrary configuratios of the amino acid in comparison to FDLA highlighted in blue and L-L-FDLA/D-D-FDLA pointing towards same configuration of the amino acid in comparison to FDLA in green.*

It is worth noting that during hydrolysis of the amide bonds slight isomerization of the acidic α-proton in the case of leucine often occur. For NP **1** the isomerization peak is 5 to 7 times lower in comparison to the main peak, thus it is more likely that the last leucine is D-configured as well.

Further derivatives

Detailed inspection of the MS cluster assigned to **1** in the respective molecular network generated by GNPS uncovered several congeners produced under different cultivation conditions, *e.g.* addition of soil extract, supplement of ethanol, with light exclusion as well as standard cultivation condition. In-depth investigation of the respective HRESIMS experiments leads to the assumption of presence of smaller ring sizes as well as linear derivatives. An $[M+H]^+$ peak at *m/z* 666.4556 matching the molecular formula $C_{33}H_{60}N_7O_7^+$ (*m/z* calcd for $[M+H]^+$ 666.4549 Δ1.15 ppm) points towards a seven membered ring after exclusion of one leucine in comparison to compound **1**.

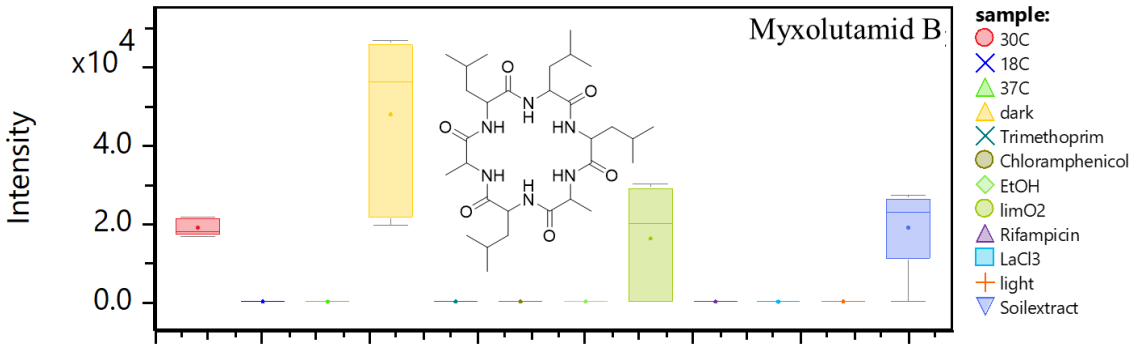


Figure S31: Box-and-whisker diagram showing intensity levels of m/z 595.417 among the different cultivation conditions: myxolutamid B is produced under several conditions with best production through light exclusion (yellow); median and mean values and 25% and 75% quartiles illustrated within box with grey whiskers representing lower and upper extremes.

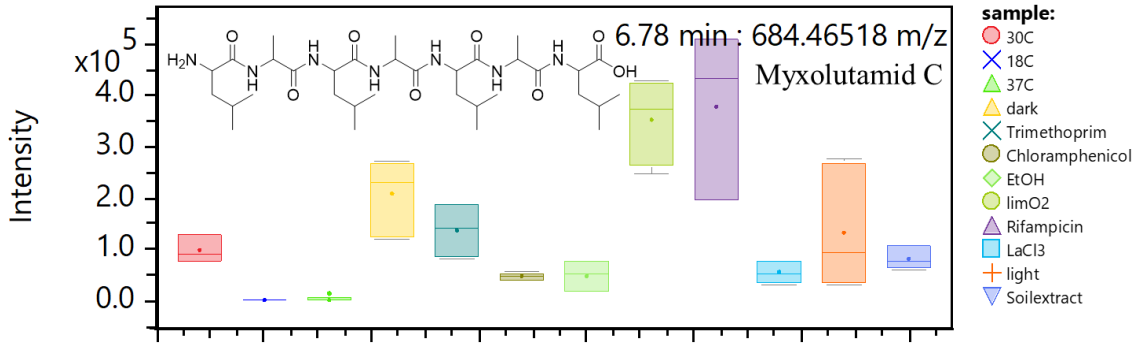


Figure S32: Box-and-whisker diagram showing intensity levels of m/z 684.465 among the different cultivation conditions: myxolutamid C is produced under several conditions with best production through oxygen limitation (olive) and rifampicin addition (purple); median and mean values and 25% and 75% quartiles illustrated within box with grey whiskers representing lower and upper extremes.

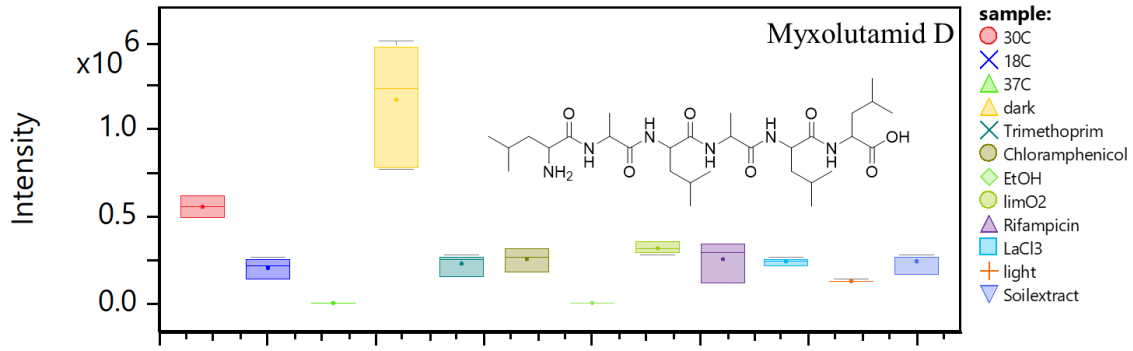


Figure S33: Box-and-whisker diagram showing intensity levels of m/z 613.428 among the different cultivation conditions: myxolutamid D is produced under several conditions with best production through light exclusion (yellow); median and mean values and 25% and 75% quartiles illustrated within box with grey whiskers representing lower and upper extremes.

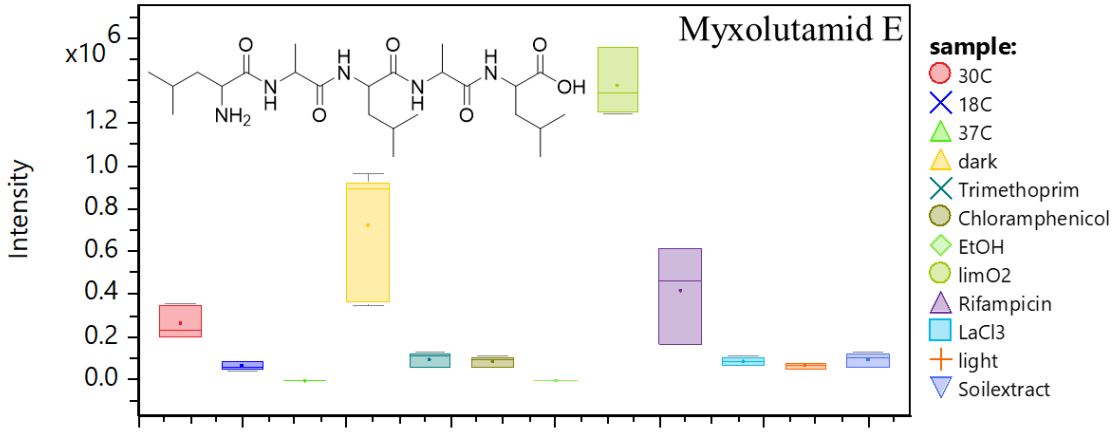


Figure S34: Box-and-whisker diagram showing intensity levels of m/z 500.344 among the different cultivation conditions: myxolutamid E is produced under several conditions with best production through oxygen limitation (olive); median and mean values and 25% and 75% quartiles illustrated within box with grey whiskers representing lower and upper extremes.

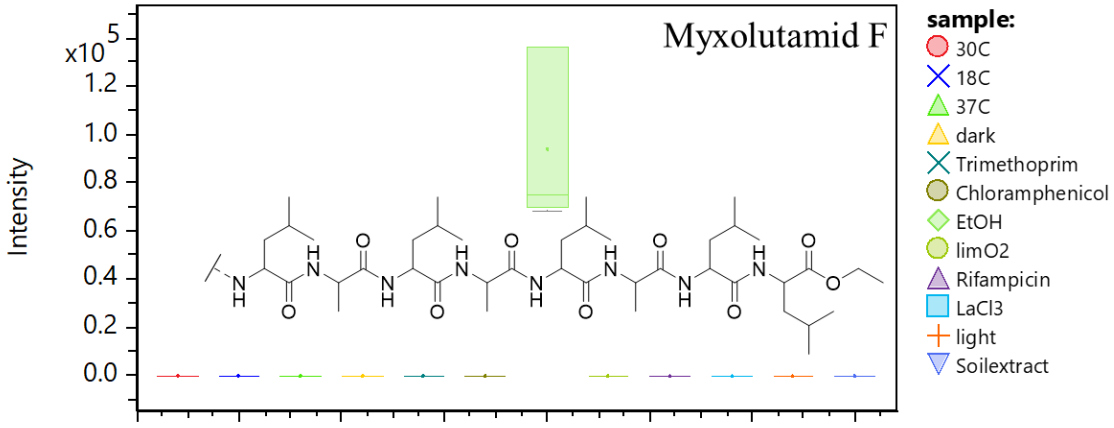


Figure S35: Box-and-whisker diagram showing intensity levels of m/z 937.670 among the different cultivation conditions: myxolutamid F is produced under several conditions with best production through ethanol addition (green); median and mean values and 25% and 75% quartiles illustrated within box with grey whiskers representing lower and upper extremes.

MS^2 fragmentation of myxolutamid derivatives

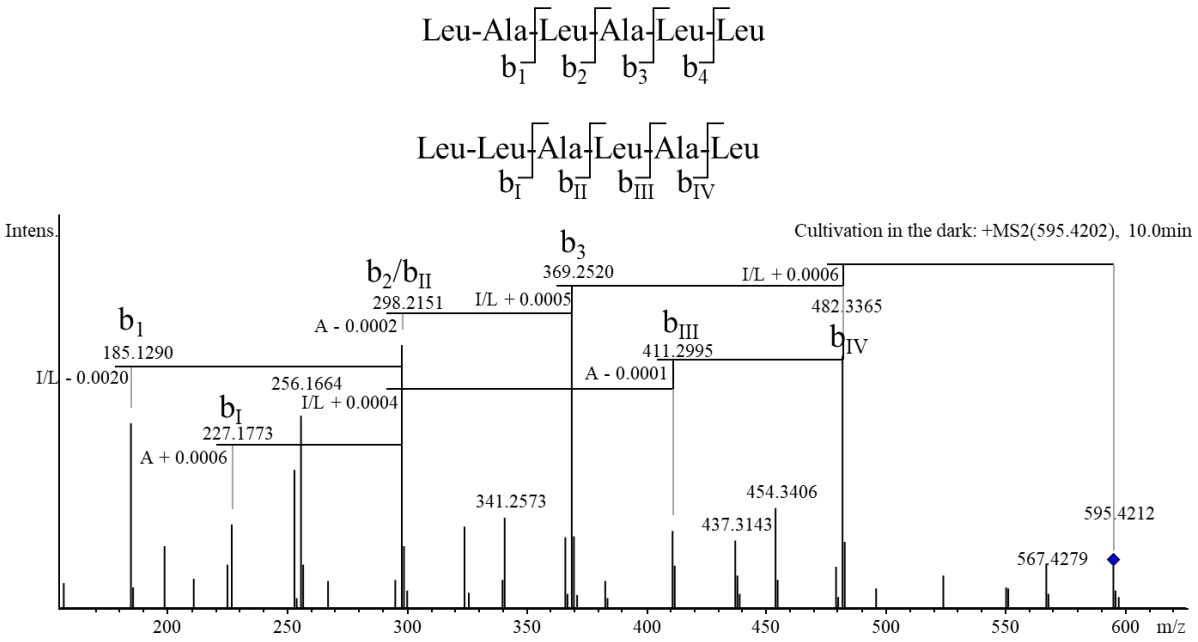


*Figure S36: MS^2 fragment spectrum of myxolutamid B, cyclic congener of **1**, with a precursor ion at m/z 595.421 with annotated amino acid eliminations after ring opening comprising isoleucine/leucine and alanine (b-fragments in black), predicted amino acid sequence is Leu-Ala-Leu-Ala-Leu-Leu respective Leu-Leu-Ala-Leu-Ala-Leu.*

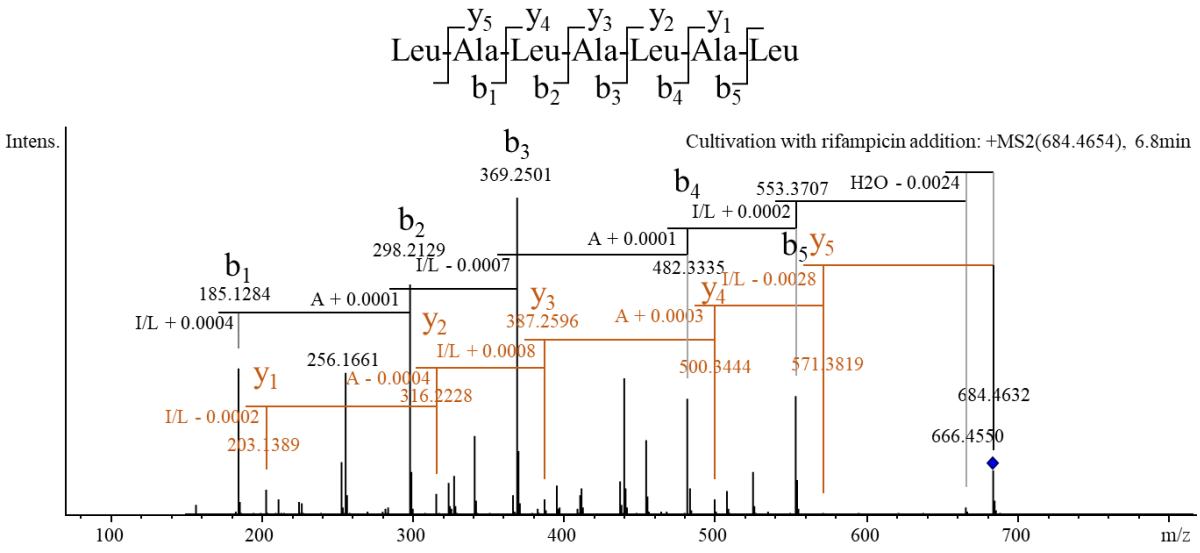


*Figure S37: MS^2 fragment spectrum of myxolutamid C, linear congener of **1**, with a precursor ion at m/z 684.463 with annotated amino acid eliminations comprising isoleucine/leucine and alanine (b-fragments in black, y-fragments in orange), predicted amino acid sequence is Leu-Ala-Leu-Ala-Leu-Ala-Leu.*

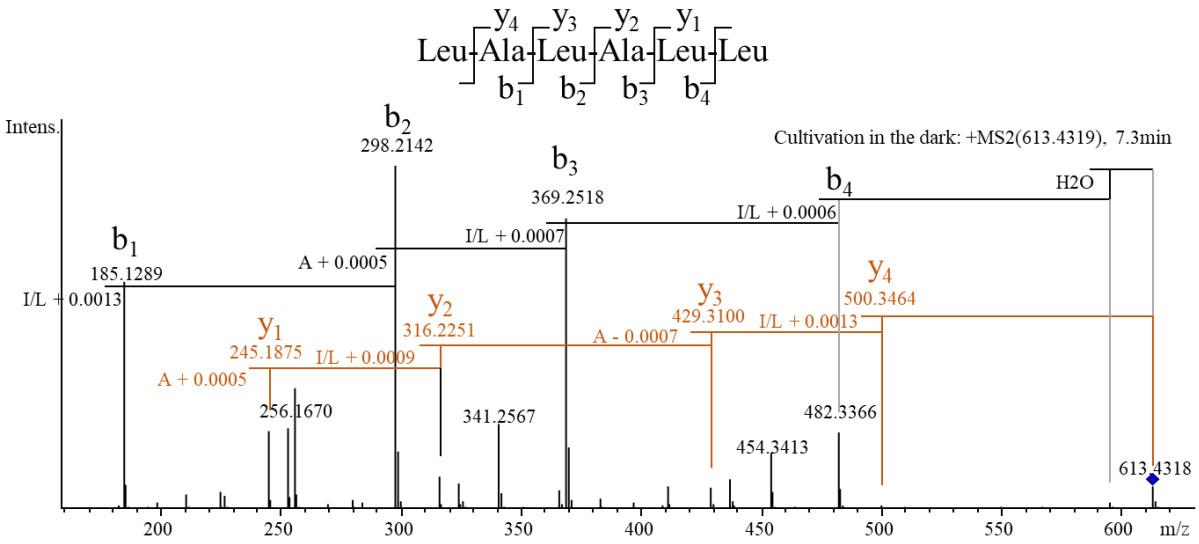


*Figure S38: MS² fragment spectrum of myxolutamid D, linear congener of **1**, with a precursor ion at m/z 613.432 with annotated amino acid eliminations comprising isoleucine/leucine and alanine (b-fragments in black, y-fragments in orange), predicted amino acid sequence is Leu-Ala-Leu-Ala-Leu-Leu.*

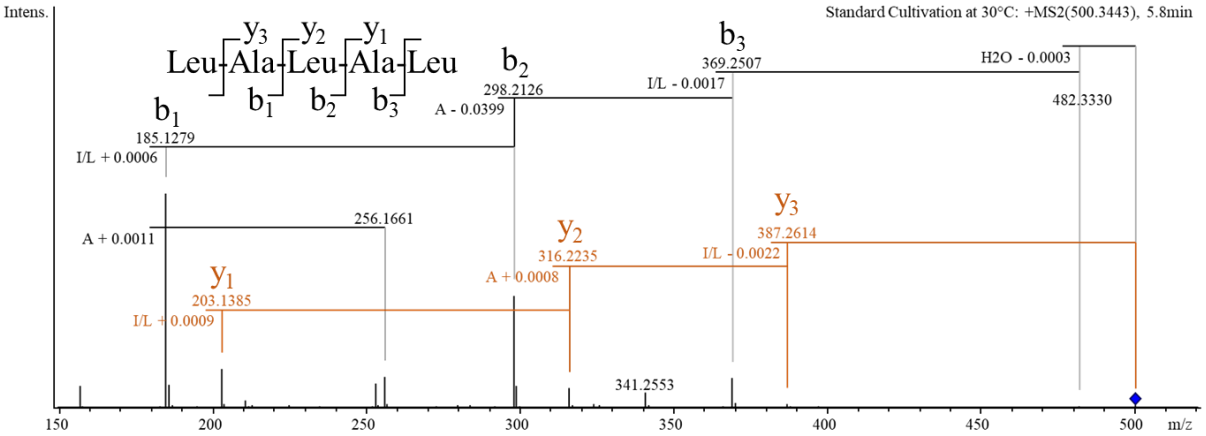


*Figure S39: MS² fragment spectrum of myxolutamid E, linear congener of **1**, with a precursor ion at m/z 500.344 with annotated amino acid eliminations comprising isoleucine/leucine and alanine (b-fragments in black, y-fragments in orange), predicted amino acid sequence is Leu-Ala-Leu-Ala-Leu.*

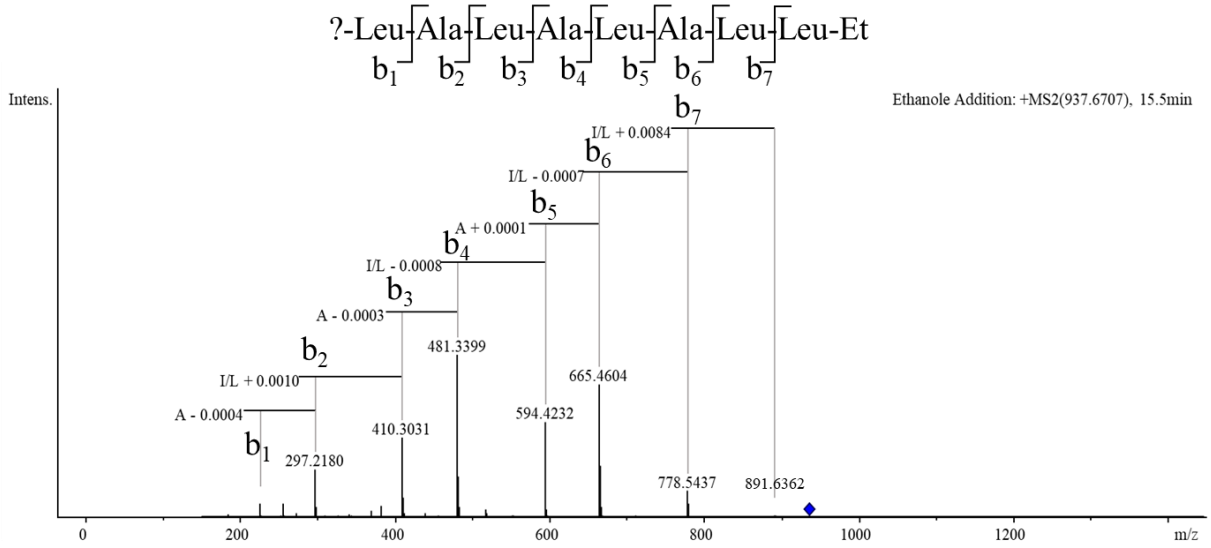


*Figure S40: MS^2 fragment spectrum of myxolutamid F, linear congener of **1**, with a precursor ion at m/z 937.671 with annotated amino acid eliminations comprising isoleucine/leucine and alanine (b-fragments in black, y-fragments in orange), predicted partial amino acid sequence is ?-Leu-Ala-Leu-Ala-Leu-Ala-Leu-Leu-Et.*

S2.9.3 Pentacyclic acid

Structure elucidation of pentacyclic acid. An $[M+H]^+$ peak for the compound **2** at *m/z* 307.2639 can be observed via HRESIMS in line with the molecular formula $C_{20}H_{35}O_2$ (*m/z* calcd for $[M+H]^+$ 307.2632). Consequently, 4 double-bond equivalents (DBE) were determined. ^{1}H-NMR and HSQC experiments show four aliphatic double-bond signals at $\delta(^1H)$ = 5.38 (1H, m) $\delta(^{13}C)$ = 130.1, $\delta(^1H)$ = 5.37 (1H, m) $\delta(^{13}C)$ = 131.2, $\delta(^1H)$ = 5.36 (1H, m) $\delta(^{13}C)$ = 129.5 and $\delta(^1H)$ = 5.34 (1H, m) $\delta(^{13}C)$ = 129.0 ppm. Furthermore, eleven methylene signals can be observed at $\delta(^1H)$ = 2.83 (2H, m) $\delta(^{13}C)$ = 26.7, $\delta(^1H)$ = 2.28 (2H, t, *J* = 7.46 Hz) $\delta(^{13}C)$ = 35.1, $\delta(^1H)$ = 2.26 (2H, t, *J* = 7.42 Hz) $\delta(^{13}C)$ = 35.7, $\delta(^1H)$ = 2.13 (2H, m) $\delta(^{13}C)$ = 27.8, $\delta(^1H)$ = 2.08 (2H, q, *J* = 6.83 Hz) $\delta(^{13}C)$ = 28.3, $\delta(^1H)$ = 1.67 (2H, quint, *J* = 7.44 Hz) $\delta(^{13}C)$ = 26.4, $\delta(^1H)$ = 1.36 (2H, m) $\delta(^{13}C)$ = 30.9, $\delta(^1H)$ = 1.29 (2H, m) $\delta(^{13}C)$ = 33.2, $\delta(^1H)$ = 1.29 (2H, m) $\delta(^{13}C)$ = 28.7, $\delta(^1H)$ = 1.29 (2H, m) $\delta(^{13}C)$ = 23.9 and $\delta(^1H)$ = 1.18 (2H, m) $\delta(^{13}C)$ = 40.4 ppm. The ^{1}H-NMR spectrum exhibits one methine proton signal at $\delta(^1H)$ = 1.53 ppm (1H, m). Moreover, two methyl group signals at $\delta(^1H)$ = 0.90 (3H, m) $\delta(^{13}C)$ = 14.6 as well as $\delta(^1H)$ = 0.89 (3H, m) $\delta(^{13}C)$ = 23.2 ppm are present. The quaternary carbon signal at $\delta(^{13}C)$ = 178.4 ppm has a characteristic shift for a carboxyl group, which is connected to a cyclopentyl ring via an aliphatic chain. Respective COSY and HMBC correlations reveal the aliphatic chain to be built up of six methylene groups with two double-bonds at position 5 and 8. One of these methylene groups is in-between the two double-bonds with a proton signal at $\delta(^1H)$ = 2.83 ppm underpinned by corresponding COSY correlations to the double-bond protons at $\delta(^1H)$ = 5.38 and $\delta(^1H)$ = 5.37 ppm and HMBC correlations to the double-bond signals at $\delta(^1H)$ = 5.36 and $\delta(^1H)$ = 5.34 ppm. The aliphatic chain is linked to the cyclopentyl moiety via the quaternary carbon at $\delta(^{13}C)$ = 30.4 ppm, where an additional propyl group is attached. This is supported by the COSY correlations between the methylene signal of the propyl group at $\delta(^1H)$ = 1.29 $\delta(^{13}C)$ = 33.2 ppm to the methylene signal of the cyclopentyl ring at $\delta(^1H)$ = 1.18 ppm on the one hand. On the other hand, there is the HMBC correlation of the methylene signal of the propyl group at $\delta(^1H)$ = 1.29 to the carbon of the methylene group that is part of the aliphatic chain at $\delta(^{13}C)$ = 30.9 ppm. Furthermore, a methyl group at $\delta(^1H)$ = 0.89 ppm is appended to the cyclic moiety at ring-position 15 due to the COSY correlation to the methine group at $\delta(^1H)$ = 1.53 ppm and the HMBC correlation with the methylene group at $\delta(^1H)$ = 1.18 ppm. Finally, the relative configuration of positions 12 and 15 of the cyclopentyl moiety should be determined by NOESY. This data point towards the conformation, where the methyl group at $\delta(^1H)$ = 0.89 ppm is in closer space to the propyl chain as no cross signals could be found for the aliphatic chain with the free acid function. However,

a clear assignment could not be determined. Additionally, the two double-bonds of 2 are cis-configured, which can be underpinned by the study of Pfeffer et al., where the ^{13}C-shift of the allylic methylene group between two double-bonds was calculated and experimentally proven [19]. According to Pfeffer *et al.*, the allylic methylene group between two cis-cis-configured double bonds is expected at $\delta(^{13}C) = 25.7$ ppm and for a trans-trans-configuration at $\delta(^{13}C) = 35.7$ ppm in $CDCl_3$, whereby the respective signal of 2 is observed at $\delta(^{13}C) = 26.7$ ppm in methanol-d_4 showing a deviation of 1 Hz to the all-cis-configuration prediction.

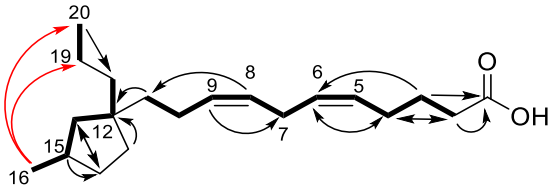


*Figure S41: Chemical structure of **2** with selected COSY (bold), HMBC (black arrows) and NOESY (red arrows) correlations as well as relevant atom numbering.*

*Table S8: NMR spectroscopic data of **2** in methanol-d_4 at 500/125 MHz.*

#	δ ^{13}C [ppm]	δ ^{1}H [ppm], mult (*J* [Hz])	COSY	HMBC	NOESY
1	178.4	-	-	-	-
2	35.1	2.28, t (7.46)	3,4	1,3,4	7
3	26.4	1.67, quint (7.44)	2,4	1,2,4,6	6,7
4	27.8	2.13, m	2,3,5,6	2,3,6	6,7
5	129.5	5.36, m	4,6	6,7,8	-
6	130.1	5.38, m	4,5,7	4	4
7	26.7	2.83, m	6,8,9,10	4,5,6,8	4,10
8	131.3	5.37, m	7,9,10	11	-
9	129.0	5.34, m	7,10	7,10	7
10	28.3	2.08, q (6.83)	7,8,9,11	8,9,11	7,8,19
11	30.9	1.36, m	10	8,10,12	-
12	30.5	-	-	-	-
13	35.7	2.26, t (7.42)	-	12	-
14	28.7	1.29, m	17	17	-
15	29.3	1.53, m	16,17	14,16,17	-
16	23.2	0.89, m	15	15,17	19,20
17	40.4	1.18, m	15,17,18	11,14,15,16,18	-
18	33.2	1.29, m	17	-	-
19	23.9	1.29, m	20	18	10
20	14.6	0.90, m	19	18,19	16

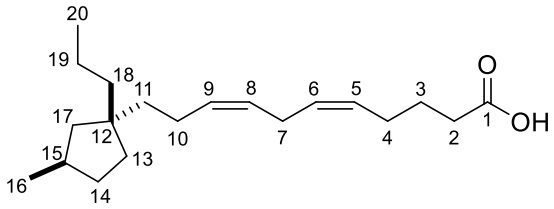


*Figure S42: Chemical structure of **2** with numbering of relevant atoms.*

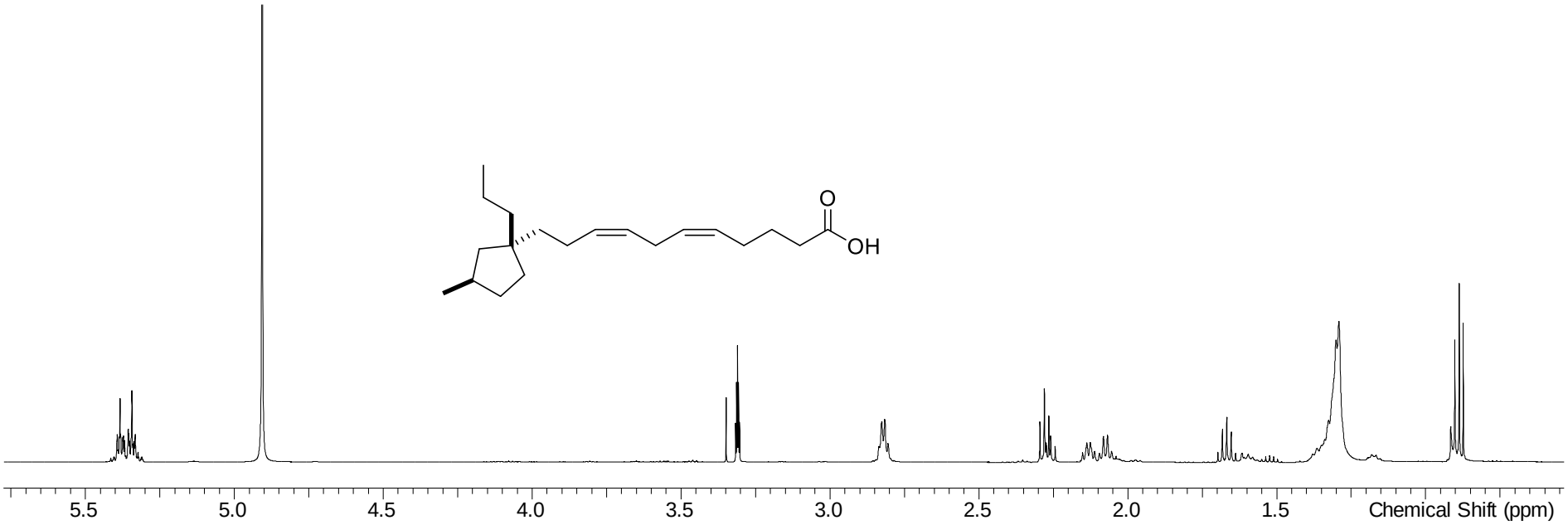


*Figure S43: ^{1}H-spectrum of **2** in methanol-d_4 at 500 MHz.*

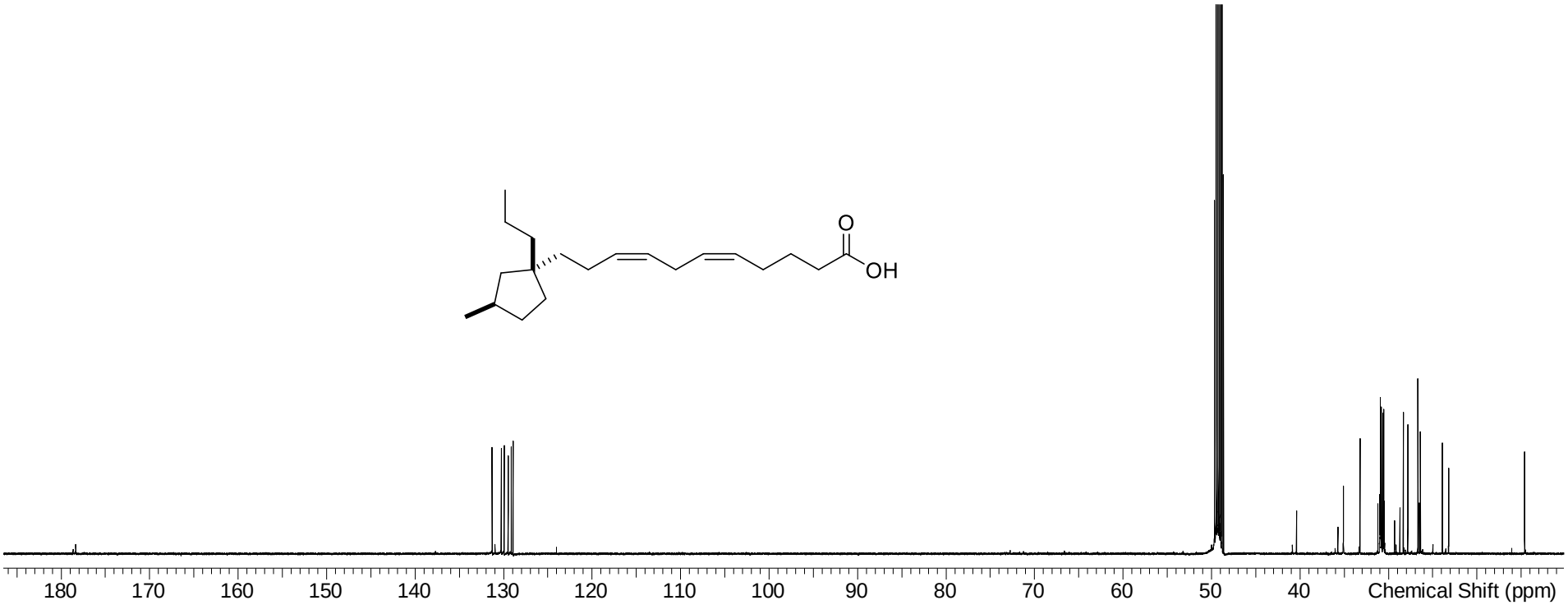
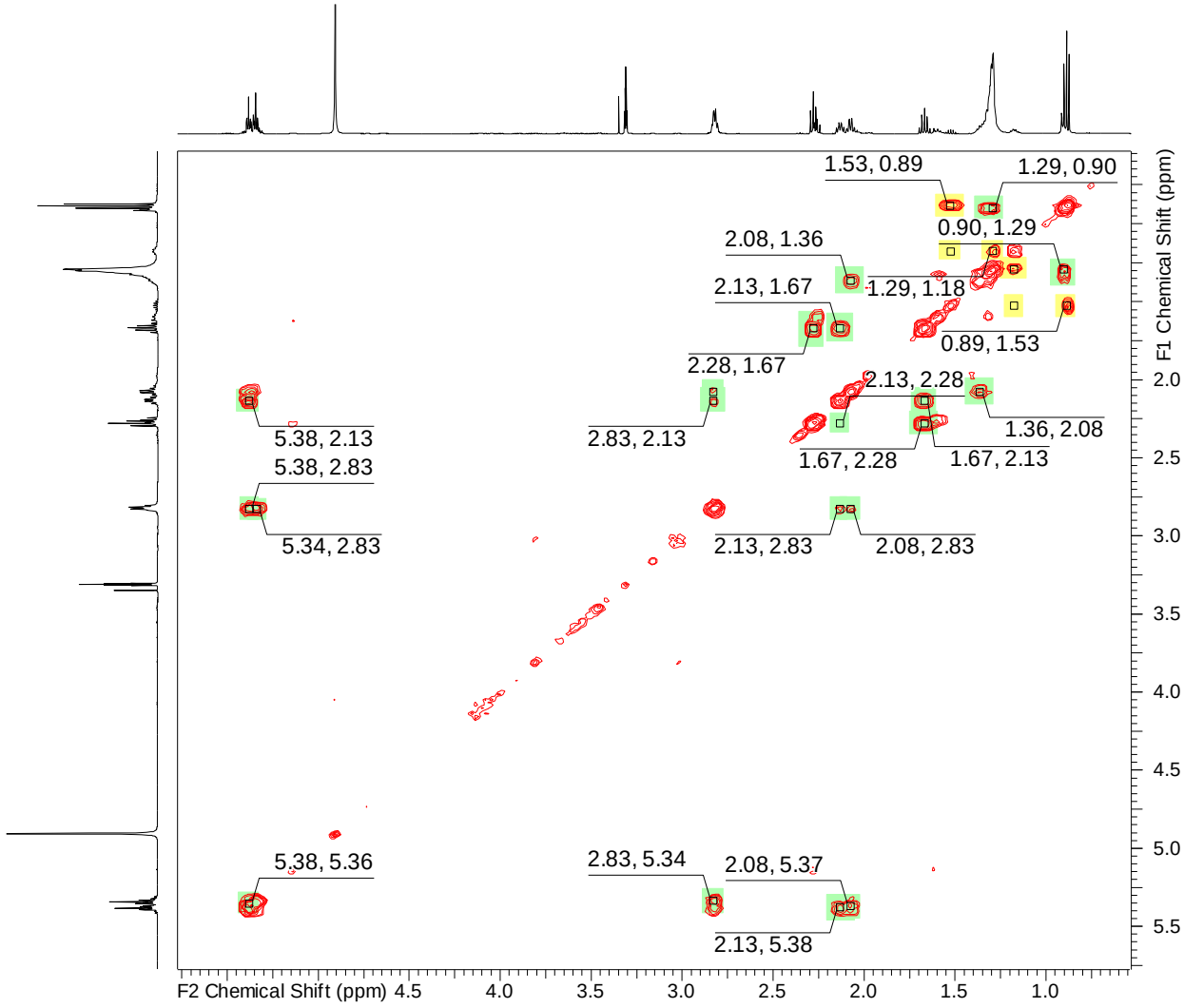


*Figure S44: ^{13}C-spectrum of **2** in methanol-d_4 at 125 MHz.*



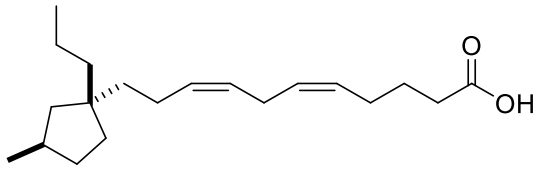
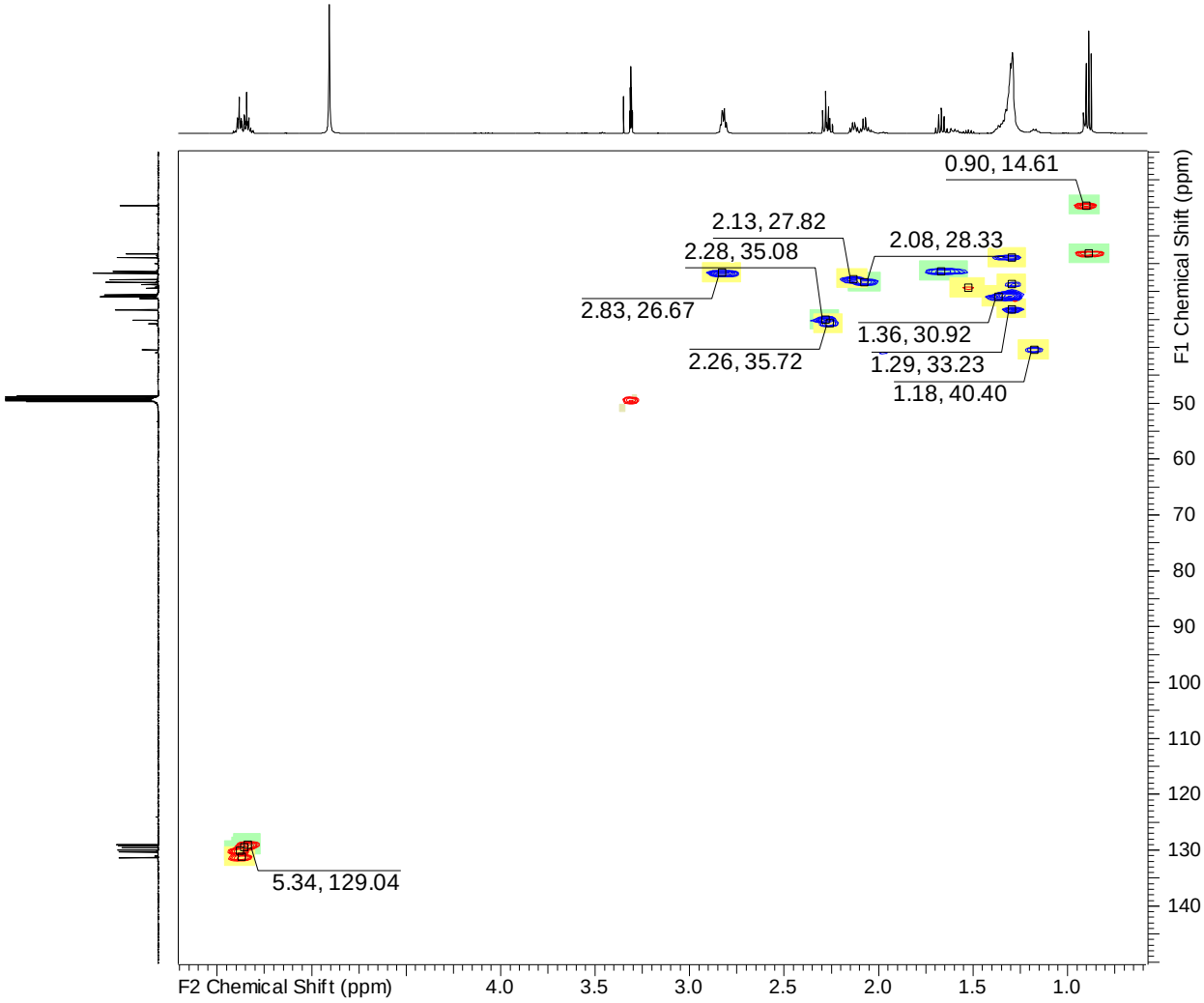


*Figure S45: COSY-spectrum of **2** in methanol-d_4 at 500 MHz.*



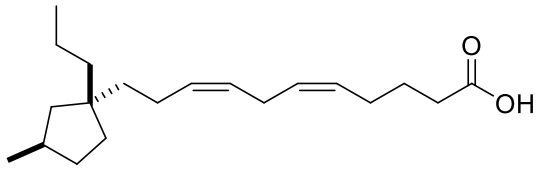
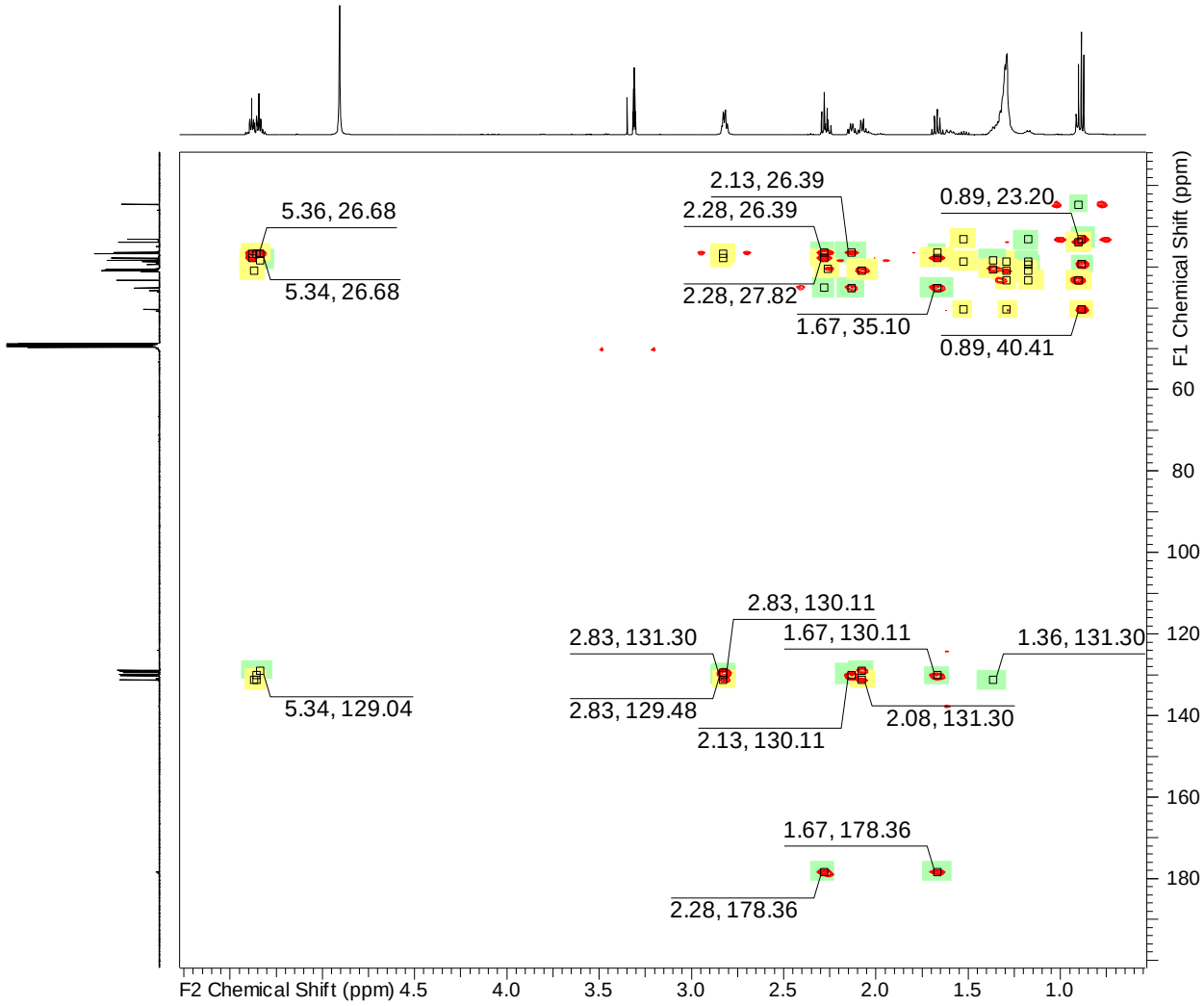


*Figure S46: HSQC-spectrum of **2** in methanol-d_4 at 500 MHz (^{1}H)/125 MHz (^{13}C).*



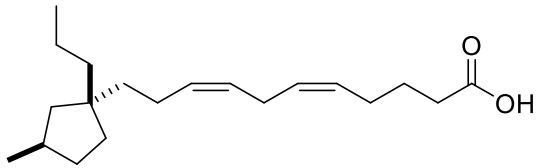
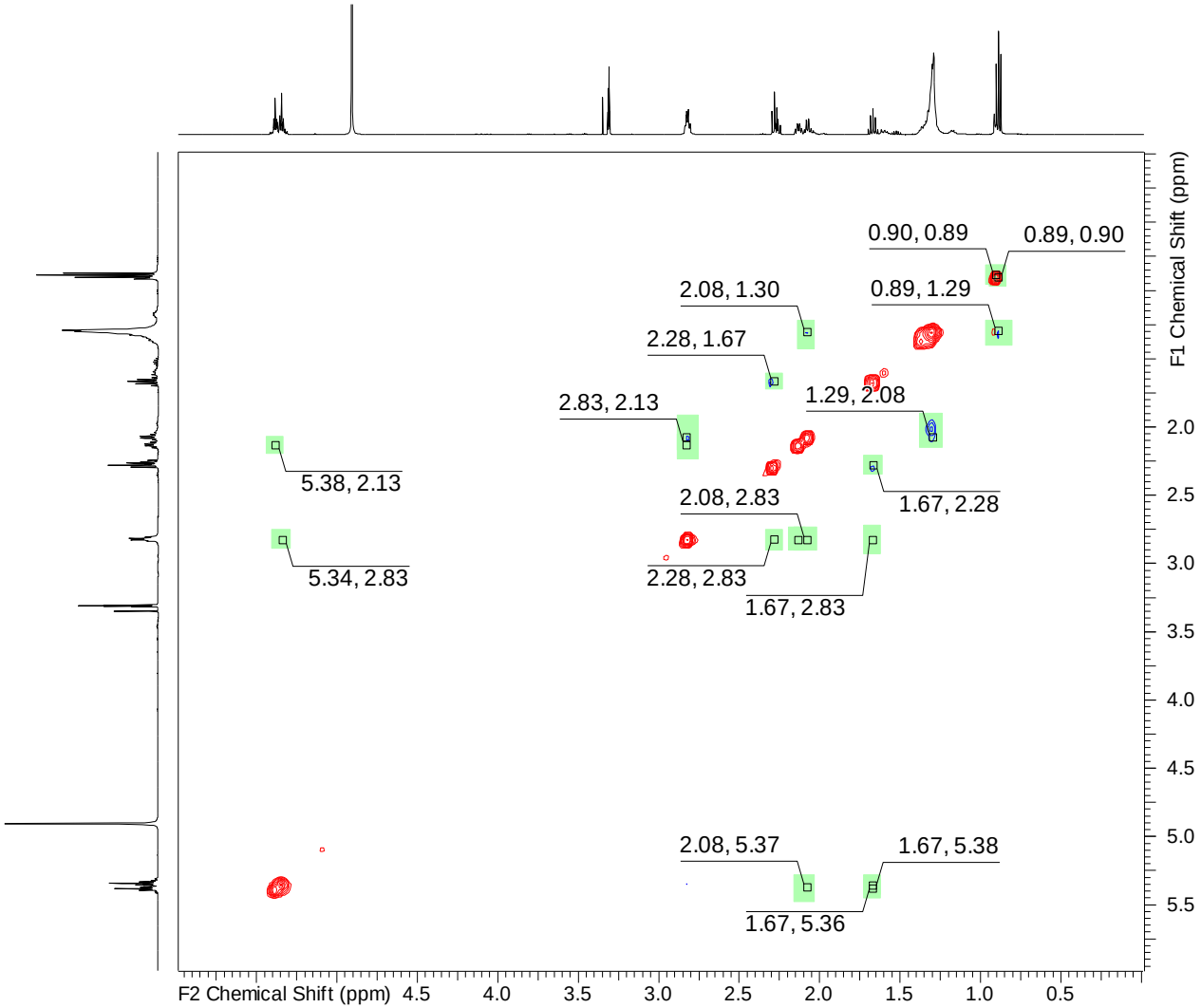


*Figure S47: HMBC-spectrum of **2** in methanol-d_4 at 500 MHz (^{1}H)/125 MHz (^{13}C).*



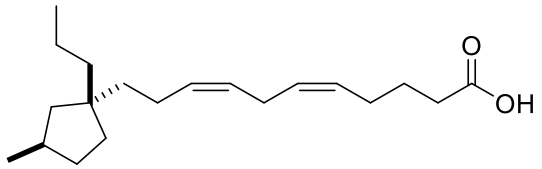


*Figure S48: NOESY-spectrum of **2** in methanol-d_4 at 500 MHz.*

S2.9.4 Glucodiolic acid

Structure elucidation of glucodiolic acid. Corresponding HRESIMS spectrum exhibits an $[M+H]^+$ peak for compound **3** at *m/z* 635.4722 consistent with the molecular formula $C_{34}H_{67}O_{10}$ (*m/z* calcd for $[M+H]^+$ 635.4729) with 2 double-bond equivalents (DBE). ^{1}H-NMR and HSQC spectra show eight methine signals at $\delta(^1H)$ = 4.38 (1H, d, *J* = 7.84 Hz), 4.08 (1H, quint, *J* = 5.99 Hz), 3.71 (1H, m), 3.50 (1H, m), 3.30 (1H, m), 3.19 (1H, ddd, *J* = 9.90, 5.70, 2.34 Hz), 3.06 (1H, t, *J* = 9.00 Hz) and 2.87 ppm (1H, dd, *J* =9.11, 7.87 Hz). Two diastereotopic methylene signals can be found at $\delta(^1H)$ = 3.81, 3.62 (2H, dd, *J* = 11.93, 2.34 Hz) $\delta(^{13}C)$ = 62.8 and $\delta(^1H)$ = 2.67, 2.41 (2H, dd, *J* = 15.10, 6.27 Hz) $\delta(^{13}C)$ = 43.61 ppm as well as six methylene signals at $\delta(^1H)$ = 1.58 (2H, m), 1.45 (2H, m), 1.43 (2H, m), 1.42 (2H, m), 1.39 (2H, m) and 1.30 ppm (24H, m). Furthermore, the HSQC spectrum exhibits four methyl signals at $\delta(^1H)$ = 3.60 (3H, s), 3.55 (3H, s), 1.14 (3H, d, *J* = 6.25 Hz) and 0.90 ppm (6H, m). The structure consists of a PKS part with a carboxyl group and two hydroxyl functions as well as a dimethylated hexose in β-position and two C_7H_{15}-chains connected via a quaternary carbon. The carboxyl signal can be found at the characteristic shift of $\delta(^{13}C)$ = 177.2 ppm with an α-methylene group at $\delta(^1H)$ = 2.41 ppm underpinned by the corresponding HMBC correlation. The further connection to the β-methine group with a downfield shift at $\delta(^1H)$ = 4.08 ppm due to the linked hexose via an ether function is indicated with COSY and HMBC correlations. The anomeric carbon of the sugar can be observed at $\delta(^1H)$ = 4.38 ppm. Corresponding COSY and HMBC correlations prove the sugar to be a twice methylated hexose. One of the methylations is located at the methine group at $\delta(^1H)$ = 3.06 ppm with the respective methoxy group at $\delta(^1H)$ = 3.60 ppm and the other one at the methine group with a shift of $\delta(^1H)$ = 2.87 ppm and the connected methoxy group at $\delta(^1H)$ = 3.55 ppm. The stereochemistry of the hexose was determined via coupling constants to be β-glucose, $^3J_{H1',2'}$ = 7.84, $^3J_{H2',3'}$ = 9.11, $^3J_{H3',4'}$ = 9.00, $^3J_{H4',5'}$ = 9.90 Hz with a deviation of less than 0.32 Hz to β-D-glucopyranose measured in D_2O by Roslund et al. [20]. Back to the PKS part, the γ-methylene group owns a downfield shift at $\delta(^1H)$ = 1.58 ppm due to its close neighborhood to the oxygen of the sugar part. In comparison, the shift of the δ-methylene group at $\delta(^1H)$ = 1.45 ppm is more upfield because of the connection to the quaternary carbon at $\delta(^{13}C)$ = 33.3 ppm. This part is supported by respective COSY and HMBC correlations. The quaternary carbon is also the connection of the two C_7H_{15}-alkyl chains with the PKS part. Each alkyl chain consists of six methylene groups at the large signal at $\delta(^1H)$ = 1.30 ppm and a terminal methyl group with a shift of $\delta(^1H)$ = 0.90 ppm, respectively. All signal belonging to the alkyl chains show HMBC correlations to the quaternary carbon. The prolongation of the PKS part starts at the quaternary carbon and consists of a six-

membered aliphatic chain. Two methine groups are hydroxylated, whereby the higher field shift signal at $\delta(^1H)$ = 3.50 ppm belongs to the methine group closer to the quaternary carbon. The respective methine group with a comparably lower field shift signal at $\delta(^1H)$ = 3.71 ppm is linked to the terminal methyl group at $\delta(^1H)$ = 1.14 ppm supported by the respective COSY correlation. The stereocenters of the hydroxylated methine groups could not be determined with NMR so far.

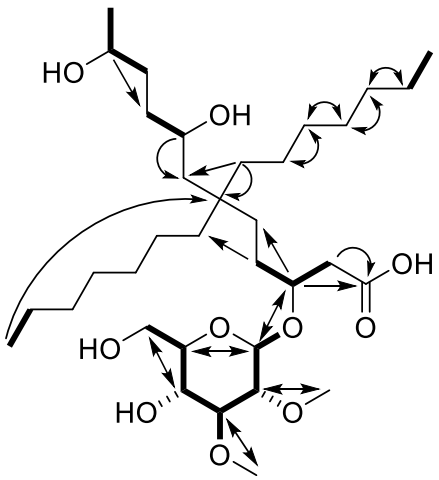


*Figure S49: Chemical structure of **3** elucidated using NMR, COSY correlations are highlighted in bold line and HMBC correlations are marked with arrows.*

*Table S9: NMR spectroscopic data of **3** in methanol-d_4 at 500/125 MHz.*

#	δ ^{13}C [ppm]	δ ^{1}H [ppm], mult (*J* [Hz])	COSY	HMBC
1	177.2	-	-	-
2	43.6	2.67,2.41, dd (15.10,6.27)	3	1,3,4
3	79.1	4.08, quint (5.99)	2,4	1,2,4,5,1’
4	36.5	1.58, m	3	2,3,5,7’’,8’’
5	26.3	1.45, m	7’’,8’’	8
6	33.3	-	-	-
7	38.6	1.43, m	8	-
8	72.6	3.50, m	7,9	7,9
9	27.1	1.39, m	8,10	8
10	40.3	1.42, m	9,11	-
11	68.7	3.71, m	10,12	9,10
12	23.7	1.14, d (6.25)	11	10,11
1’	104.6	4.38, d (7.84)	2’	3,2’,3’,5’
2’	85.3	2.87, dd (9.11,7.87)	1’,3’	1’,3’,8’
3’	87.8	3.06, t (9.00)	2’,4’	1’,2’,4’,5’,7’
4’	71.3	3.30, m	3’,5’	3’,5’,6’
5’	77.7	3.19, ddd (9.90,5.70,2.34)	4‘,6‘	1‘,3‘,4‘,6‘
6‘	62.8	3.81,3.62, dd (11.93,2.34)	5‘	4‘,5‘
7‘	61.4	3.60, s	-	3‘
8‘	61.1	3.55, s	-	1‘,2‘
1‘‘,14‘‘	14.7	0.90, m	2‘‘-13‘‘	6
2‘‘-13‘‘	30.9	1.30, m	1‘‘,14‘‘	1‘‘-14‘‘

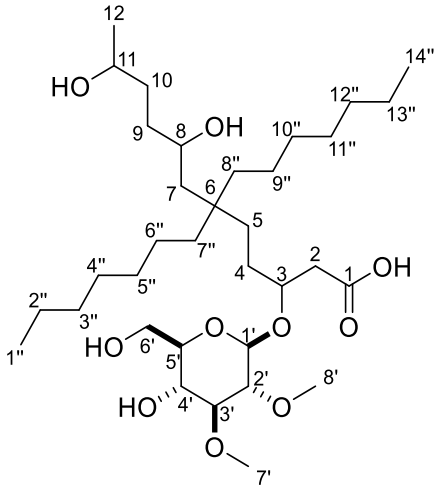


*Figure S50: Chemical structure of **3** with numbering of relevant atoms.*

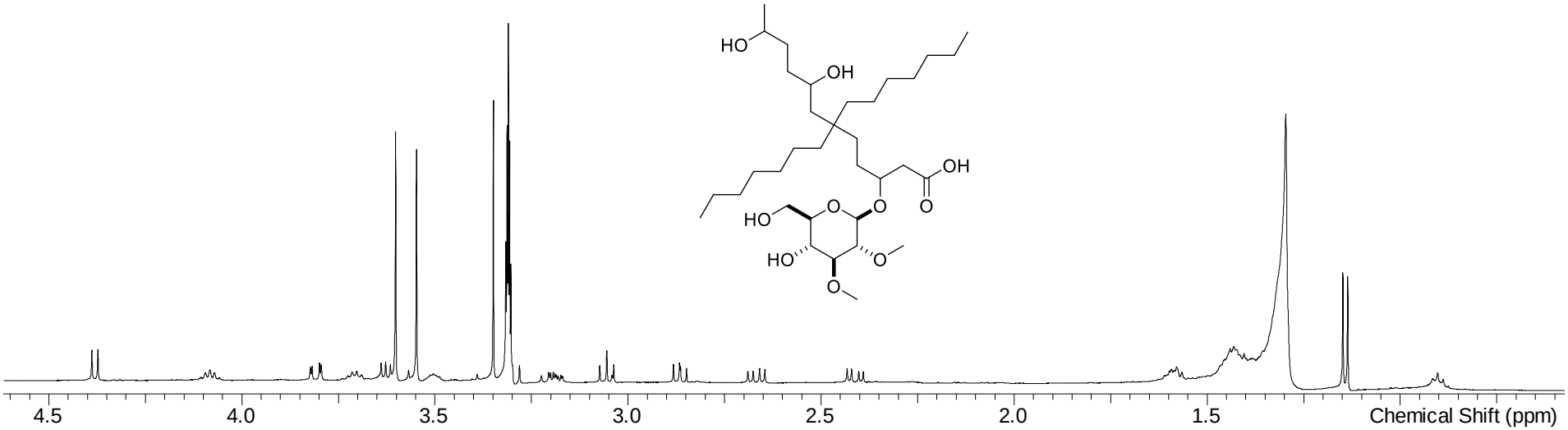


*Figure S51: ^{1}H-spectrum of **3** in methanol-d_4 at 500 MHz.*

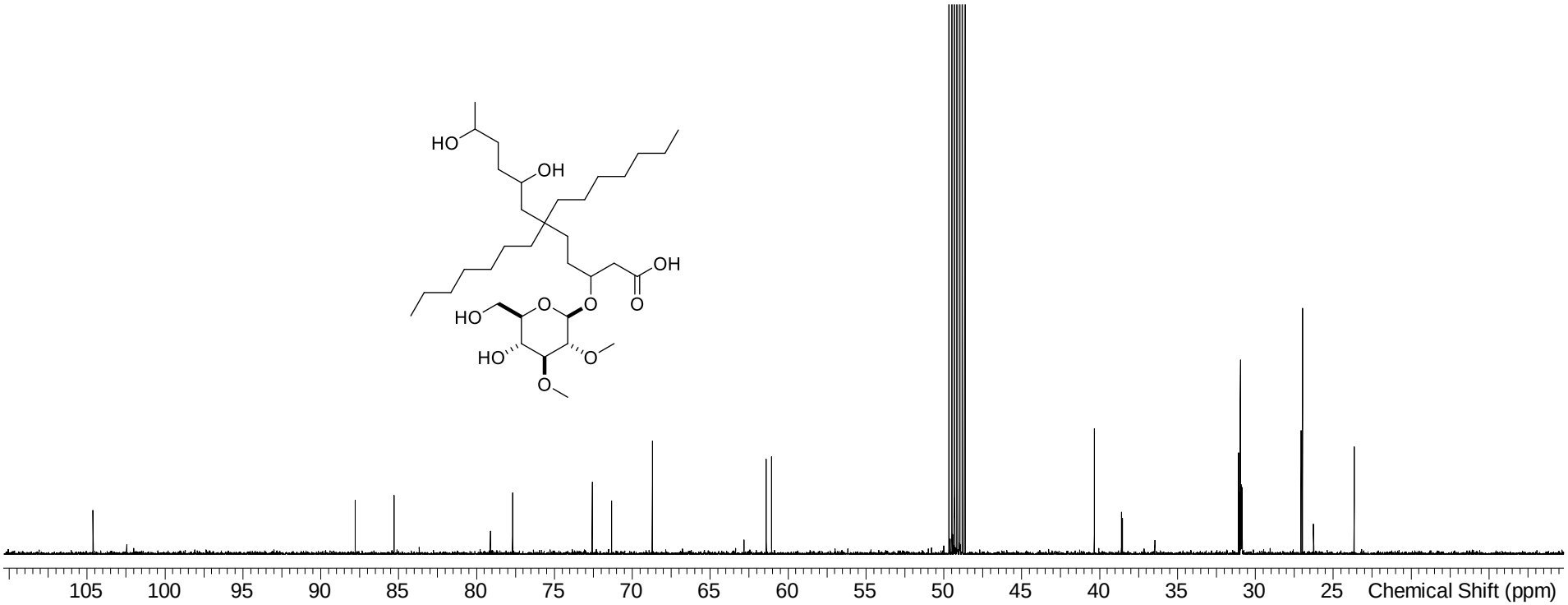
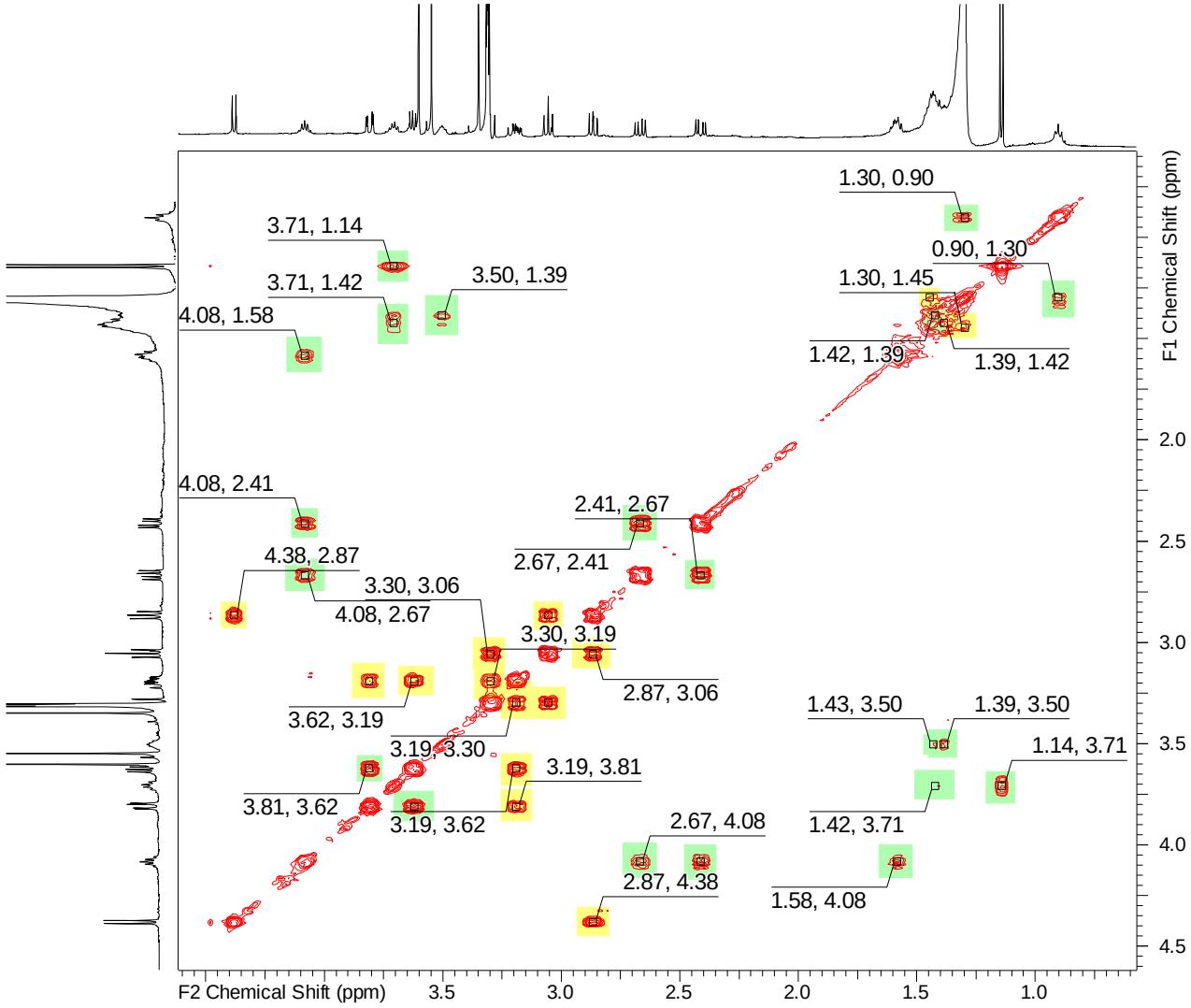


*Figure S52: ^{13}C-spectrum of **3** in methanol-d_4 at 125 MHz.*



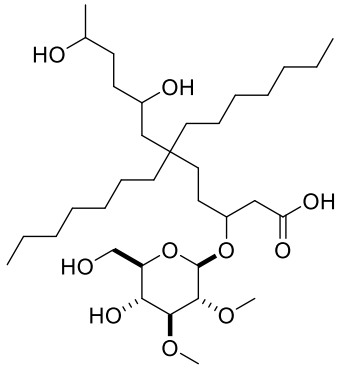
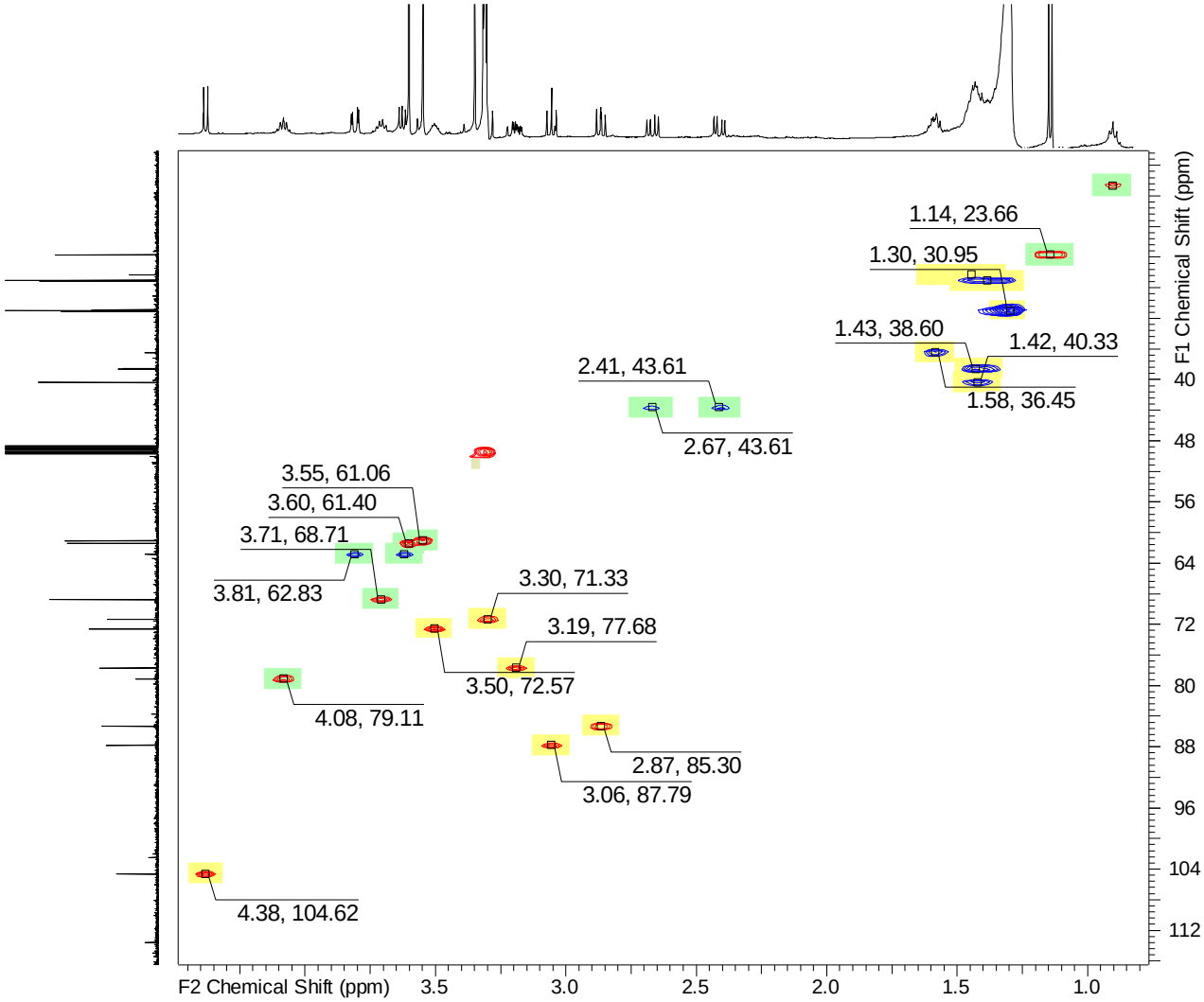


*Figure S53: COSY-spectrum of **3** in methanol-d_4 at 500 MHz.*



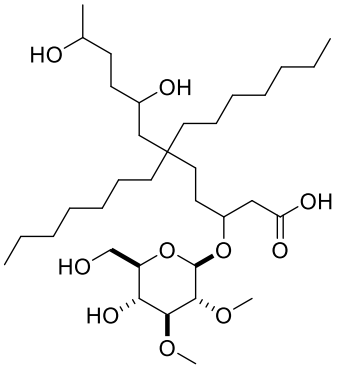
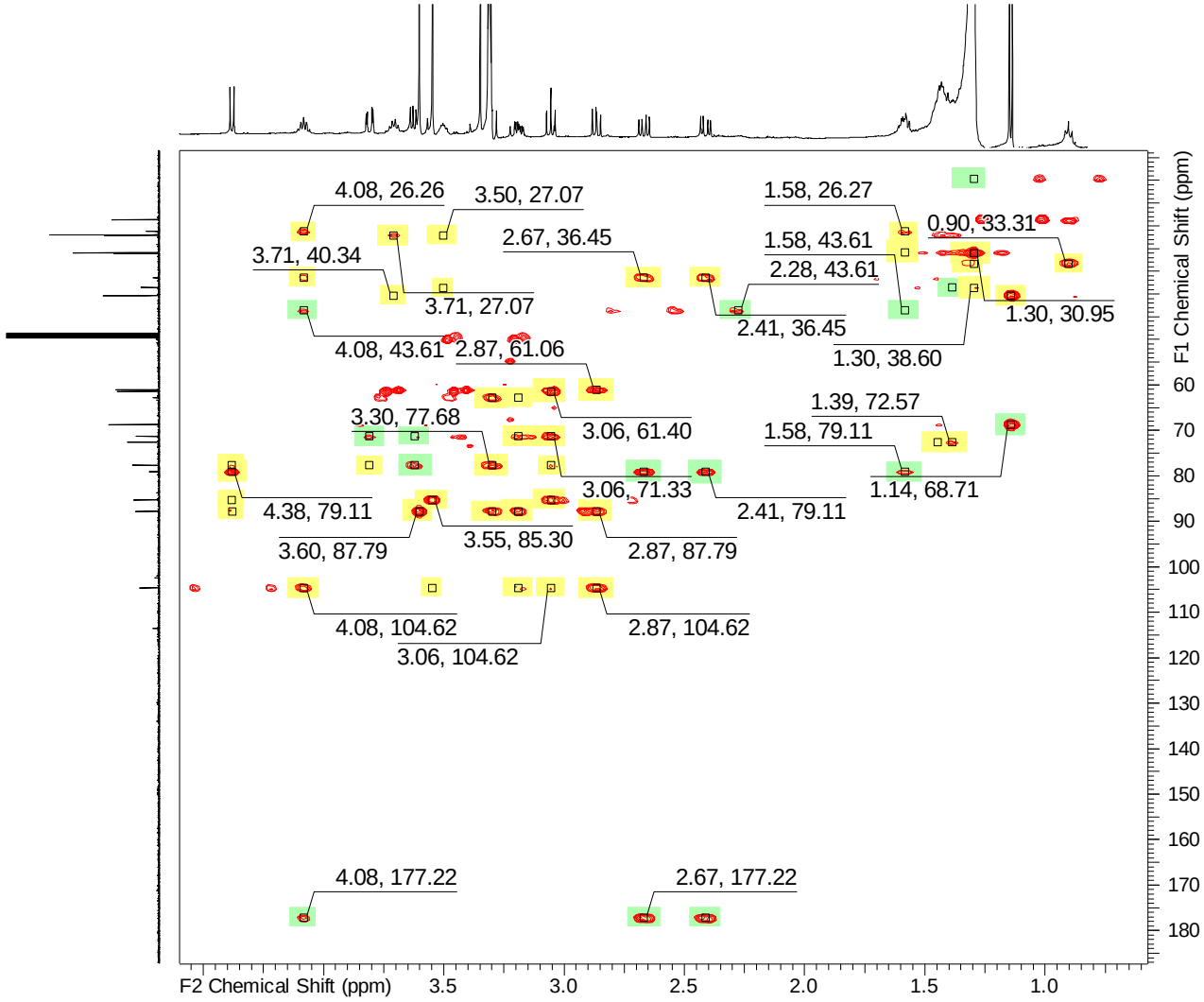


*Figure S54: HSQC-spectrum of **3** in methanol-d_4 at 500 MHz (^{1}H)/125 MHz (^{13}C).*



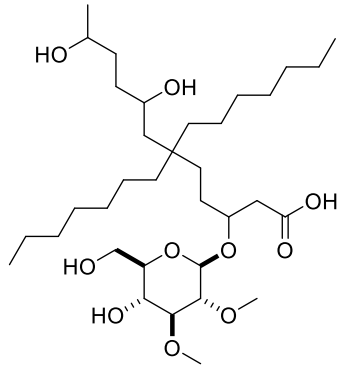


*Figure S55: HMBC-spectrum of **3** in methanol-d_4 at 500 MHz (^{1}H)/125 MHz (^{13}C).*

MS^2 experiments

To underpin the predicted structure, the fragmentation behavior of compound **3** is investigated to suggest tentative fragment ion structures. Starting with the scheduled precursor MS^2 measurement on maXis 4G (Figure S56), eight fragment ions can be observed predominantly. The fragment ion with *m/z* 445.38 might be the result of the elimination of the sugar part, thereby keeping the hydroxyl group. Subsequent fragment ions 2 to 6 originate from five water eliminations and double bond formations, respectively. Moreover, a CO neutral loss from fragment ion *m/z* 373.35 seems to lead to fragment ion *m/z* 345.35. The last prominent fragment ion is *m/z* 331.34, where the aliphatic chain appears to be shortened by a CH_2-moiety.

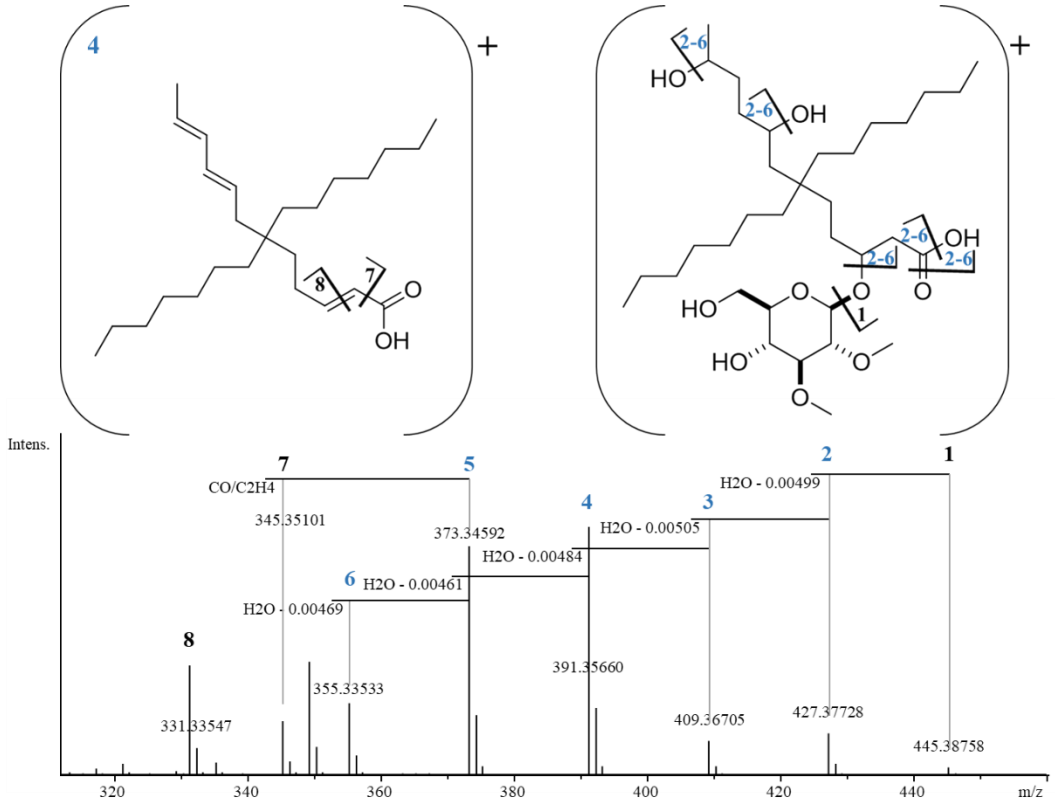


*Figure S56: Selected precursor list (SPL) MS^2 measurement on maXis 4G of **3**.*

Furthermore, measurements on the solariX were carried out to find additional structurally informative fragmentations. Figure S58 displays the zoomed in view of the spectrum in the range between *m/z* 500 and *m/z* 700 of a full scan measurement. As mentioned earlier the predicted structures for the fragment ions are only representatives as the fragmentation of aliphatic chains is arbitrary. Nevertheless, fragment ion with *m/z* 593.59 seems to result from

a neutral loss of a C_3H_7-moiety of the parent mass ion with *m/z* 635.47. Further, the gradual elimination of C_2H_4-units from the fragment ion with *m/z* 593.59 to *m/z* 565.57 and *m/z* 537.54 but no further losses of C_2H_4-units supports the hypothesis of having two C_7H_{15}-chains at the quaternary carbon instead of two different chain lengths. Otherwise, it would be more likely to see further fragment ions in this range with C_2H_4 neutral loss.

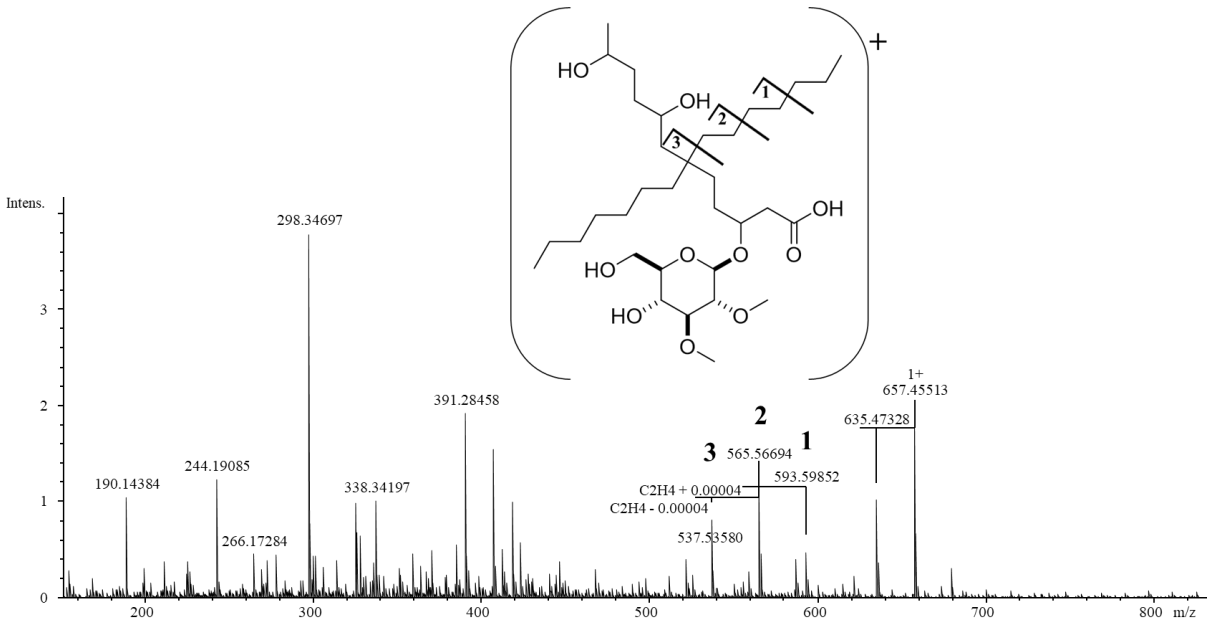


*Figure S57: Full scan measurement on solariX FT-ICR of **3** using direct infusion showing putative in source fragmentation.*

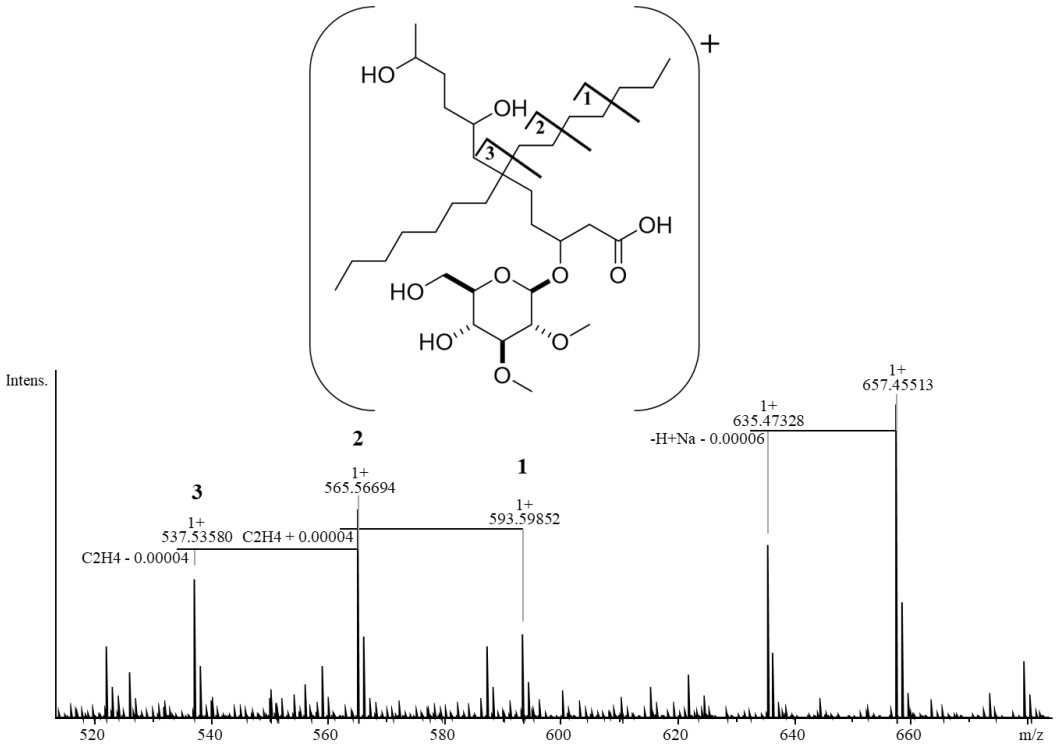


*Figure S58: Zoomed in view of full scan measurement on solariX FT-ICR of **3** using direct infusion showing putative in source fragmentation.*

Subsequently, the fragment ion at *m/z* 373.35 was isolated for further fragmentation using SORI mode with 1.3 eV resulting the spectrum shown in Figure S59. Thereby, the fragment ion at *m/z* 97.10 matches the predicted sum formula of C_7H_{13} exhibiting one double bond. This provides an additional argument for the presence of two C_7H_{15}-chains at the quaternary carbon.

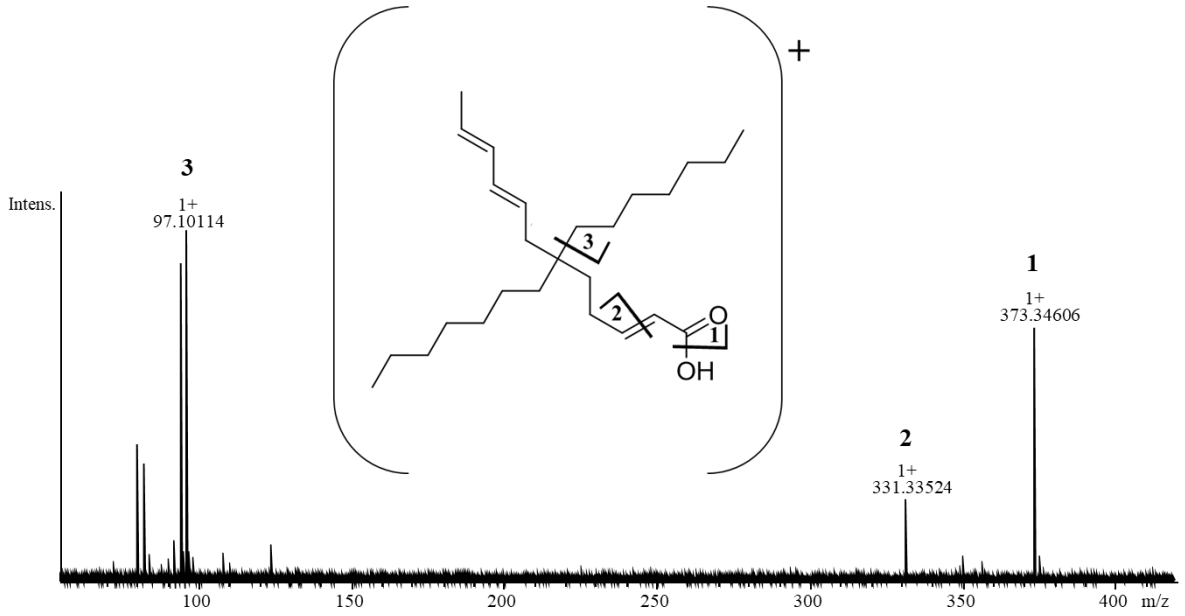


*Figure S59: SORI fragmentation measurement of m/z 373.34606 with 1.3 eV on solariX FT-ICR of **3** using direct infusion.*

S2.10 References

1. 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, doi:10.1021/acschembio.7b00900.
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, doi:10.1021/acschembio.8b00948.
3. 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, doi:10.1021/acschembio.7b00816.
4. 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, doi:10.1039/C8SC01325J.
5. 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, doi:10.1007/s10295-018-2105-6.

6. Bader, C.D.; Neuber, M.; Panter, F.; Krug, D.; Müller, R. Supercritical Fluid Extraction Enhances Discovery of Secondary Metabolites from Myxobacteria. *Anal. Chem.* **2020**, *92*, 15403–15411, doi:10.1021/acs.analchem.0c02995.
7. 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* **2020**, *25*, 2676, doi:10.3390/molecules25112676.
8. Li, Y.; Zhuo, L.; Li, X.; Zhu, Y.; Wu, S.; Hu, W.; Shen, T.; Li, Y.-Z.; Wu, C. Myxadazoles, Myxobacterium-Derived Isoxazole-Benzimidazole Hybrids with Cardiovascular Activities. *Angew. Chem.* **2021**, doi:10.1002/ange.202106275.
9. Hug, J.J.; Frank, N.A.; Walt, C.; Šenica, P.; Panter, F.; Müller, R. Genome-Guided Discovery of the First Myxobacterial Biarylitide Myxarylin Reveals Distinct C–N Biaryl Crosslinking in RiPP Biosynthesis. *Molecules* **2021**, *26*, 7483, doi:10.3390/molecules26247483.
10. Frank, N.A.; Széles, M.; Akone, S.H.; Rasheed, S.; Hüttel, S.; Frewert, S.; Hamed, M.M.; Herrmann, J.; Schuler, S.M.M.; Hirsch, A.K.H.; et al. Expanding the Myxochelin Natural Product Family by Nicotinic Acid Containing Congeners. *Molecules* **2021**, *26*, 4929, doi:10.3390/molecules26164929.
11. Popoff, A.; Hug, J.J.; Walesch, S.; Garcia, R.; Keller, L.; Müller, R. Structure and Biosynthesis of Myxofacyclines: Unique Myxobacterial Polyketides Featuring Varing and Rare Heterocycles *. *Chemistry* **2021**, *27*, 16654–16661, doi:10.1002/chem.202103095.
12. Panter, F.; Bader, C.D.; Müller, R. The Sandarazols are Cryptic and Structurally Unique Plasmid-Encoded Toxins from a Rare Myxobacterium*. *Angew. Chem. Int. Ed Engl.* **2021**, *60*, 8081–8088, doi:10.1002/anie.202014671.
13. Okoth, D.A.; Hug, J.J.; Mándi, A.; Kurtán, T.; Garcia, R.; Müller, R. Structure and biosynthesis of sorangipyranone - a new γ-dihydropyrone from the myxobacterial strain MSr12020. *J. Ind. Microbiol.* **2021**, *48*, doi:10.1093/jimb/kuab029.
14. Hug, J.J.; Kjaerulff, L.; Garcia, R.; Müller, R. New Deoxyenhygrolides from *Plesiocystis pacifica* Provide Insights into Butenolide Core Biosynthesis. *Mar. Drugs* **2022**, *20*, 72, doi:10.3390/md20010072.
15. Panter, F.; Popoff, A.; Garcia, R.; Krug, D.; Müller, R. Myxobacteria of the Cystobacterineae Suborder Are Producers of New Vitamin K2 Derived Myxoquinones. *Microorganisms* **2022**, *10*, 534, doi:10.3390/microorganisms10030534.
16. Bader, C.D.; Panter, F.; Garcia, R.; Tchesnokov, E.P.; Haid, S.; Walt, C.; Spröer, C.; Kiefer, A.F.; Götte, M.; Overmann, J.; et al. Sandacrabins - Structurally Unique Antiviral

RNA Polymerase Inhibitors from a Rare Myxobacterium. *Chemistry – A European Journal* **2022**, *28*, e202104484, doi:10.1002/chem.202104484.

17. Dorothy A. Okoth; Joachim J. Hug; Ronald Garcia; Rolf Müller. Three New Stigmatellin Derivatives Reveal Biosynthetic Insights of its Side Chain Decoration. *Preprints* **2022**, *2022060188*, doi:10.20944/preprints202206.0188.v1.
18. Lee, H.-S.; Shin, H.J.; Jang, K.H.; Kim, T.S.; Oh, K.-B.; Shin, J. Cyclic peptides of the nocardamine class from a marine-derived bacterium of the genus Streptomyces. *J. Nat. Prod.* **2005**, *68*, 623–625, doi:10.1021/np040220g.
19. Philip E. Pfeffer; Francis E. Luddy; Joseph Unruh; James N. Shoolery. Analytical13C NMR: A rapid, nondestructive method for determining the cis,trans composition of catalytically treated unsaturated lipid mixtures. *Journal of the American Oil Chemists' Society* **1977**, *54*, 380–386, doi:10.1007/BF02802040.
20. Roslund, M.U.; Tähtinen, P.; Niemitz, M.; Sjöholm, R. Complete assignments of the 1H and 13C chemical shifts and JH,H coupling constants in NMR spectra of d-glucopyranose and all d-glucopyranosyl-d-glucopyranosides. *Carbohydr. Res.* **2008**, *343*, 101–112, doi:10.1016/j.carres.2007.10.008.

Chapter 3

3. Genome Guided Discovery of the Myxobacterial Thiolactone-containing Sorangibactins

Yunsheng Gao[a,b,c,+], Christine Walt[a,c,+], Chantal D. Bader[a,c], Rolf Müller[a,b,c,*]

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

[b] Helmholtz International Lab for Anti-Infectives, Campus E8.1, 66123 Saarbrücken, Germany

[c] German Center for Infection Research (DZIF), Partner Site Hannover-Braunschweig, 38124 Braunschweig, Germany

[+] These authors contributed equally to this work

* To whom correspondence should be addressed: rolf.mueller@helmholtz-hips.de

Contributions and acknowledgements

Author's effort:

The author contributed to the concept of this study as well as the performing of experiments and subsequent interpretation of the results. Compound purification and *de novo* structure elucidation including MS^2-fragmentation analysis were performed by the author. Furthermore, the author was significantly involved in conceiving and writing of this manuscript.

Contributions by others:

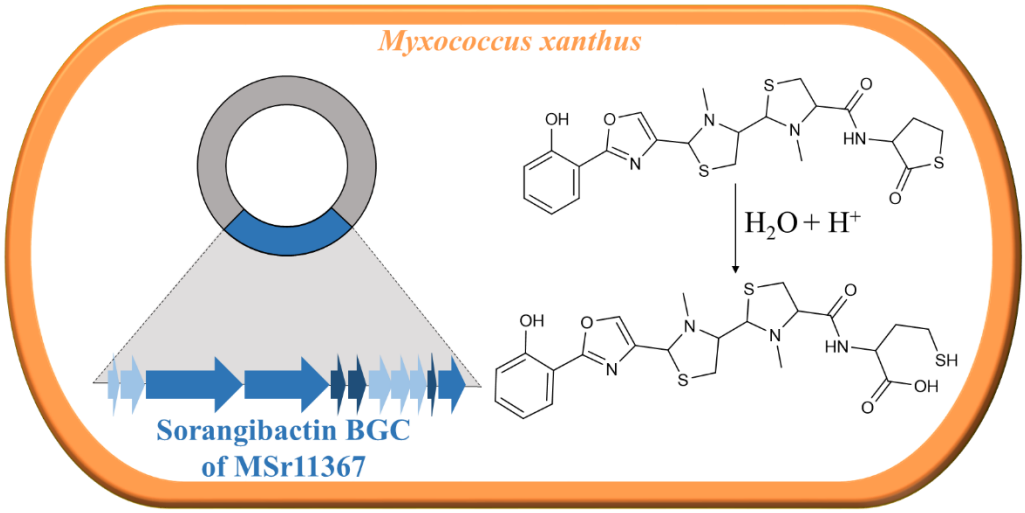
Yunsheng Gao and Fabian Panter contributed to the concept of this study. Yunsheng Gao was involved in the bioinformatics analysis of the biosynthetic pathway including laboratory experiments, cultivation and stable isotope feeding. Yunsheng Gao was significantly involved in conceiving and writing of this manuscript. Chantal Bader verified the structure elucidation and contributed to writing the manuscript. Rolf Müller contributed to conception and supervision of this study with subsequent editing and proofreading of the manuscript.

Abstract

In this study, an unprecedented myxobacterial siderophore termed sorangibactin was discovered by heterologous expression of a coelibactin-like nonribosomal peptide synthetase (NRPS) gene cluster from the *Sorangiineae* strain MSr11367 in the host *Myxococcus xanthus* DK1622. *De novo* structure elucidation uncovered a linear polycyclic structure consisting of an N-terminal di-OH benzoic acid, an oxazole, tandem N-methyl-thiazolidines and an unusual C-terminal γ-thiolactone moiety. Except for the unprecedented oxazoline dehydrogenation to form the oxazole, which likely is catalyzed by a cytochrome P450 enzyme, other tailoring steps were found necessary for efficient downstream processing. The unusual thioesterase (TE) domain is supposed to select homocysteine or methionine as the intermolecular nucleophile in γ-thiolactone formation. Its active site comprises a rare cysteine, which was found necessary for product formation by point mutation to alanine or serine, which both abolished its activity. This unusual release mechanism and the resulting rare thiolactone structure might serve as starting point for further biochemical investigations.

Keywords: myxobacteria, genome mining, heterologous expression, siderophore, homocysteine thiolactone

TOC



3.1 Introduction

Myxobacteria are Gram-negative δ-proteobacteria featuring unique morphological characteristics, complex life cycles, large genomes, and a particularly huge biosynthetic potential for secondary metabolite production [1]. Hundreds of structurally interesting scaffolds and bioactive compounds from myxobacteria were discovered over the past decades [2]. Despite this rich collection of characterized secondary metabolites, genome sequencing as well as biosynthetic gene cluster (BGC) annotation have revealed that the majority of the myxobacterial biosynthetic potential is still untapped [3–5]. Thus, a "gene to compound" strategy can shed light on assigning secondary metabolites to their BGCs and find unprecedented structures. As myxobacteria, especially the *Sorangiineae* suborder, are usually slow growing and intractable for genetic manipulation, heterologous expression is particularly promising to explore them, as it does not only facilitate genome mining for natural product discovery, but also biosynthetic pathway engineering and detailed biosynthesis studies [6]. One prolific starting point for examination of cryptic BGCs are so-called cosmid libraries [7], which can comprise the entire genome of a selected bacterial strain in the form of circular DNA, which facilitates high throughput heterologous expression [8]. This approach also enables the unambiguous correlation of BGCs with their products.

One interesting structural class of natural products are bacterial siderophores, which are small-molecular-weight Fe(III)-specific ligands that scavenge iron from their environment and enable uptake into the cell through outer-membrane transporters [9]. Siderophores from myxobacteria belong to two main categories namely the catecholate-type such as the myxochelins [10] and hyalachelins [11] and the citrate-hydroxamate-type such as the nannochelins [12]. In contrary, over 500 congeners of diverse siderophore families have been described from actinobacteria [13]. One particularly puzzling actinobacterial siderophore is coelibactin, the existence of which has only been proposed based on the presence of the respective BGC and for which only a hypothetical chemical structure is found in the literature. Comprehensive bioinformatics analysis of the respective BGC found in *Streptomyces coelicolor* A3(2) [14] enabled a prediction of the core structure based on the predicted sequence of the NRPS gene cluster, yet isolation of coelibactin or derivatives thereof failed for unknown reasons [15].

In this study, we report the discovery of a novel myxobacterial secondary metabolite with siderophore-like behavior termed sorangibactin from *Sorangiineae* strain MSr11367 by heterologous expression of a promoter-engineered NRPS gene cluster, which resembles the coelibactin BGC. This coelibactin-like NRPS gene cluster was prioritized due to: 1) its

bioinformatically predicted structure resembling coelibactin from *Streptomyces coelicolor* A3(2) [14], which raised hope to contribute to solving the puzzle of a coelibactin-like structure, 2) its regular operon organization, which is suitable for promoter engineering, 3) multiple tailoring genes surrounding the core NRPS part, which indicate auxiliary modifications before or after chain release and may endow the final product with structural novelty and diversity.

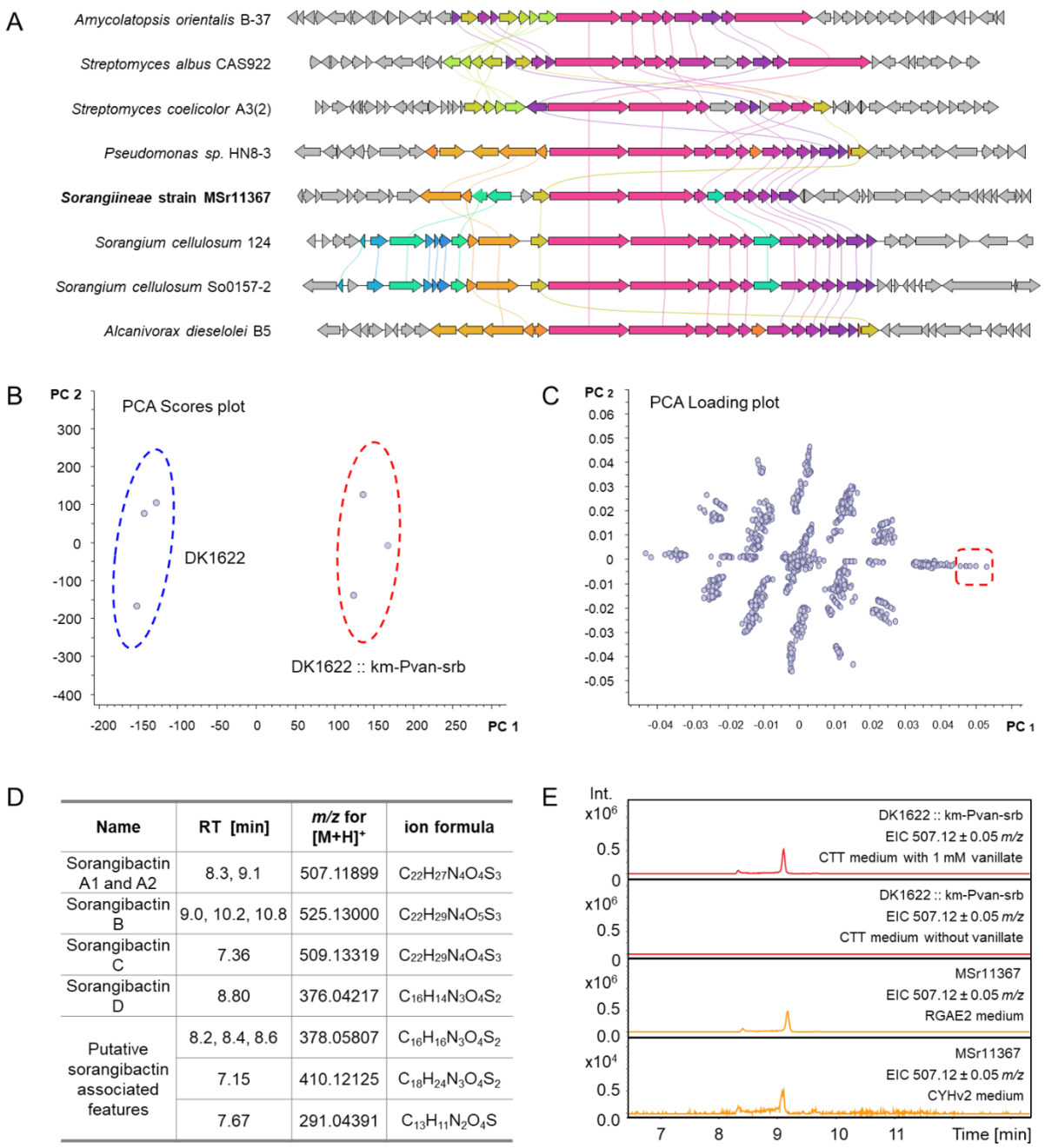
3.2 Results and discussion

Heterologous expression of a coelibactin-like NRPS gene cluster

The myxobacterial *Sorangiineae* strain MSr11367 was prioritized for natural product discovery due to its great biosynthetic potential of 62 biosynthetic gene clusters annotated by antiSMASH (Chapter 2). The strain was already explored using an OSMAC metabolomics study in Chapter 2, but even altered cultivation conditions did not reflect its full biosynthetic potential making this strain particularly interesting for genome-based discovery. Despite several efforts, MSr11367 proved to be intractable for *in-situ* genetic manipulation. Therefore, a genomic DNA library in form of cosmids was generated by Explogen LLC (EXG) for systematic screening of biosynthetic gene clusters by heterologous expression in the amenable host *Myxococcus xanthus* DK1622 [16]. Among the 62 BGCs encoded in the strain's genome, we prioritized a coelibactin-like NRPS gene cluster that showed widespread homologous gene clusters in *Sorangium cellulosum, Streptomyces* and *Pseudomonas* strains (Figure 1A), presumably indicating a conserved function of the correlating secondary metabolites during evolution. The respective annotation of the sorangibactin NRPS gene cluster is provided in Table S1.

The unmodified gene cluster was cloned from the cosmid library of the *Sorangiineae* strain MSr11367 and transferred to the host *M. xanthus* DK1622 for expression, however no heterologous product was detected. Thus, a *km-vanR-Pvan* cassette was inserted between the two main operons (Figure S2), after which the first operon consisting of transporter genes shares a strong constitutive promoter with the kanamycin resistance gene and the second operon consisting of biosynthesis genes is under control of a vanillate-inducible Pvan promoter. Principal component analysis (PCA) of HPLC-*hr*MS datasets revealed several molecular features exclusively present in *M. xanthus* DK1622 mutants containing the promoter engineered gene cluster and absent in the *M. xanthus* DK1622 wild-type host (Figure 1B, 1C and 1D) (SI Methods). The major component (termed sorangibactin A) could also be found in *Sorangiineae* strain MSr11367 extracts in equal concentrations when cultivated under iron-

deficient conditions, indicating it as final product of respective BGC (Figure 1E). Our previous study applying the OSMAC approach for MSr11367 also found sorangibactin A1 and A2 to be produced when supplementing soil extracts to the cultivation medium, but it was only observed in minor amounts (Figure S28). Isolation for structure elucidation was performed utilizing the heterologous host as besides its less complex metabolome that was found beneficial for the purification process, it also shows faster growth rates reducing possible contamination issues during cultivation as observed for MSr11367.



Name	RT [min]	*m/z* for $[M+H]^+$	ion formula
Sorangibactin A1 and A2	8.3, 9.1	507.11899	$C_{22}H_{27}N_4O_4S_3$
Sorangibactin B	9.0, 10.2, 10.8	525.13000	$C_{22}H_{29}N_4O_5S_3$
Sorangibactin C	7.36	509.13319	$C_{22}H_{29}N_4O_4S_3$
Sorangibactin D	8.80	376.04217	$C_{16}H_{14}N_3O_4S_2$
Putative sorangibactin associated features	8.2, 8.4, 8.6	378.05807	$C_{16}H_{16}N_3O_4S_2$
	7.15	410.12125	$C_{18}H_{24}N_3O_4S_2$
	7.67	291.04391	$C_{13}H_{11}N_2O_4S$

Figure 1: Characterization of the coelibactin-like NRPS gene cluster from Sorangiineae strain MSr11367. A) Clinker-analysis showing selected homologous gene clusters are widespread throughout different bacterial species. Genes with 30% sequence identity are assigned a unique color and linked. B) Significant difference between the heterologous expression dataset (DK1622 :: km-Pvan-srb) encircled in red and the wild type host dataset (DK1622) encircled in blue shown by a PCA scores plot. C) Molecular features that contribute to this difference shown by a PCA loading plot. D) Molecular features that are only found in the heterologous expression dataset, E) HPLC-MS chromatogram of sorangibactin A production by DK1622 :: km-Pvan-srb (red) and MSr11367 (orange).

Characterization of sorangibactins

LC-*hr*MS measurements exhibit an $[M+H]^+$ peak for the mixture of sorangibactin A1 and A2 at *m/z* 507.1189 matching the molecular formula $C_{22}H_{27}N_4O_4S_3^+$ (*m/z* calcd for $[M+H]^+$ 507.1189, Δ0.1 ppm) featuring 12 double-bond equivalents (DBE). Combined 1D and 2D NMR spectra uncovered a system comprising four heterocycles and one phenole, starting from the N-terminus with a phenol substituted by an oxazole group in *ortho*-position, which is further connected to two consecutive N-methylated thiazolidines linked via an amide bond to a cyclic γ-thiolactone as the C-terminal end. The assignment of the oxazole moiety and the unusual C-terminus were underpinned by ^1H-^{15}N HMBC experiments (SI) as well as the comparison of the chemical shifts for the homocysteine thiolactone part to already published NPs like thiolactomide [17]. The two signal sets for sorangibactin A1 and A2 were identical besides the signal area of the isomerization of the C-terminal homocysteine thiolactone upon purification. The methine signal for sorangibactin A1 is located at $\delta(^1H)$ = 4.09 ppm and for A2 at $\delta(^1H)$ = 4.16 ppm, whereas the adjoining diastereotopic methylene group of sorangibactin A1 shows a signal at $\delta(^1H)$ = 2.10/1.92 ppm and A2 features a methylene group without any diastereotopic effect at $\delta(^1H)$ = 2.04 ppm. It is worth noting that in the thiolactone part the HMBC correlation between the methylene group next to the sulfur at $\delta(^1H)$ = 3.01 ppm and the carbonyl group at $\delta(^{13}C)$ = 179.5 ppm could not be identified. Nevertheless, the predicted thiolactone ring is underpinned by the COSY and HMBC correlations between the two methylene and the methine group, by the HMBC correlations from the methine and the adjacent methylene group to the carbonyl function as well as the predicted 12 DBE in agreement with the predicted sum formula.

Several structure parts of sorangibactin show similarities to well described siderophores, however it features some uncommon modifications. The *ortho*-phenol moiety originating from salicylic acid incorporation is also present in yersiniabactin produced by *Yersinia pestis* [18] and predicted for coelibactin from *Streptomyces coelicolor* [14]. Furthermore, yersiniabactin features one thiazolidine moiety and two dihydrothiazoles, whereby the thiazolidine in contrast to sorangibactin is not N-methylated and no dihydrothiazoles are present in sorangibactin. The oxazole moiety in sorangibactin is different to the methylated dihydrooxazole predicted for coelibactin. Finally, the main structural difference of sorangibactin to published siderophores is clearly the uncommon homocysteine thiolactone at the C-terminus, reflecting the unprecedented structure of sorangibactin. This structural moiety impeded the purification procedure and also sheds light on possible obstacles that might have hindered structure

elucidation of coelibactin: sorangibactin A1 and A2 undergo hydrolysis during the standard purification setup with formic acid as modifier resulting in the C-terminally ring opened sorangibactin B. The α-proton of the acid labile homocysteine thiolactone tends to isomerization even when using the adapted purification workflow leading to a mixture of the two isomers sorangibactin A1 and A2 (Methods). Even though we could not separate these two isomers, the corresponding NMR data set of the mixture enabled the structural assignment of both of them.

For the stereochemical assignment of sorangibactin A1 and A2 only Marfey's derivatization assay was applicable, since this molecule is a peptide without any functionalities accessible for Mosher esterification and due to its instabilities not attainable for crystallization experiments as exemplarily performed to solve the stereochemistry of the related siderophore yersiniabactin [18]. Marfey's analysis however could also only provide limited information for three main reasons: 1) the hydrolysis of thiazolidines is known to lead to the oxidation state of an aldehyde instead of the free acid, which is instable and degrades [19], 2) the stereo centers at positions 9 and 15 originate from the former carboxyl function of the cysteine before cyclization to the thiazolidine group and thus would lose stereo information again upon hydrolysis, 3) the sorangibactin A1 and A2 mixture contains both *R*- and *S*-configurations of the α-proton of the C-terminal homocysteine thiolactone, which cannot be separated. Consequently, only positions 7 and 12 could reasonably be accessible by Marfey's analysis, but they were found to isomerize during hydrolysis. LC-*hr*MS analysis displayed a mixture of four enantiomers with a mass of *m/z* 430.14±0.02 corresponding to FDLA-derivatized N-methyl cysteine and homocysteine (Figure S22). Hence, we can solely present the flat structure with a bioinformatic prediction of all *S*-configuration due to absence of an epimerase in the respective modules.

We also obtained a second congener of the sorangibactin family through provoking hydrolysis of the rather labile thioester yielding the ring opened product sorangibactin B.

LC-*hr*MS experiments show an $[M+H]^+$ peak for sorangibactin B at *m/z* 525.1298 with a predicted molecular formula $C_{22}H_{28}N_4O_5S_3^+$ (*m/z* calcd for $[M+H]^+$ 525.1295, Δ0.6 ppm) featuring 11 double-bond equivalents (DBE). Noteworthy, more than five isomers of sorangibactin B are present in fractions that underwent the hydrolysing isolation conditions using formic acid during purification. The presumably isomerizing stereocenters are most likely the ones belonging to the homocysteine and the two thiazolidine parts exhibiting 5 stereocenters. Thus, the following structure elucidation belongs to the most abundant isomer unveiling the most intensive NMR correlations. In contrast to sorangibactin A1 and A2 the cyclic homocysteine thiolactone is opened, which shows an impact on the corresponding shifts

due to the more shielded free carboxylic acid with a shift at $\delta(^{13}C)$ = 176.6 ppm compared to the former thioester function. The HMBC spectrum exhibits two correlations to the free carboxyl group from methylene groups at $\delta(^{1}H)$ = 1.59 and 2.27 ppm. The more downfield shifted signal can be explained by the close neighborhood to the thiol function. Corresponding COSY correlations from the more shielded methylene group to the homocysteine α proton at $\delta(^{1}H)$ = 4.68 ppm can be observed. Furthermore, HMBC correlations evince from the homocysteine α-proton as well as the methine group of the adjacent thiazolidine ring at $\delta(^{1}H)$ = 3.57 ppm to the connective amide group at $\delta(^{13}C)$ = 173.0 ppm. Comparing the remaining data set to sorangibactin A1 and A2 the signals corresponding to the N-terminal part from atom 8 to 26 are similar with ≤0.1 ppm deviation for ^{1}H and ≤2.1 ppm for ^{13}C-shifts.

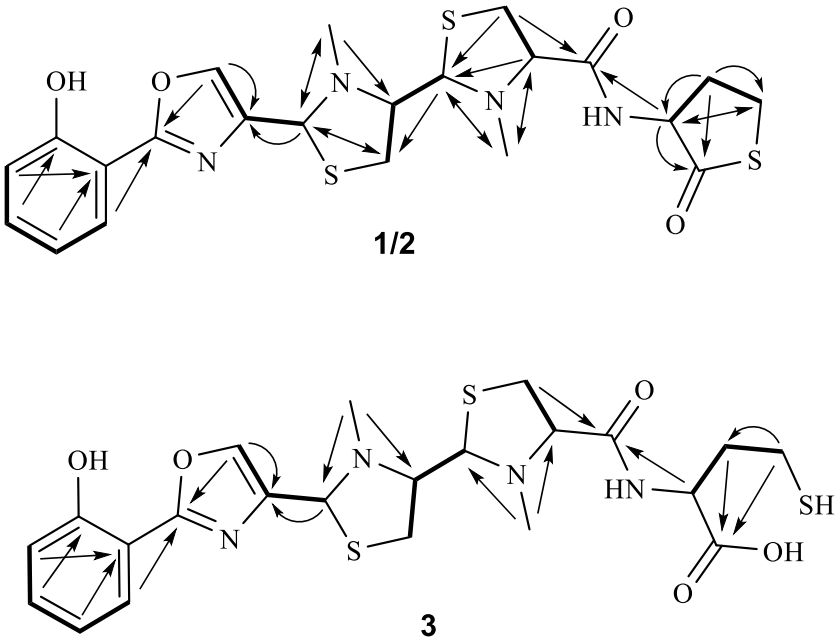


*Figure 2: Chemical structures of sorangibactin A1 and A2 (**1** and **2**) and B (**3**) with main COSY (bold) and HMBC (arrows) correlations.*

MS^2 spectral analysis was performed to obtain additional evidence supporting the proposed structures for sorangibactin A1, A2 and B. Just as for the structurally related siderophores yersiniabactin, ulbactin B and escherichelin characteristic fragmentation behavior can also be observed for sorangibactins [20]. The three main fragment ions of sorangibactin B are in accordance with the respective bond cleavages in A1 and A2 (Figure 3A). Fragment ions b, c and d of **1** and **2** correspond to fragment ions b, c+H_2O and d+H_2O of **3**.

The N-terminal phenol group derived from salicylic acid is further supported by feeding 2-hydroxybenzoic acid-d_6, resulting in a +4 mass shift of the respective fragment. Feeding L-

serine-$^{13}C_3$,^{15}N and L-cysteine-$^{13}C_3$,^{15}N resulted in the +4 and +8 mass shift respectively, which indicates the NRPS to incorporate one serine and two cysteines and is consistent with the proposed sorangibactin structure containing one oxazole and two thiazolidine substructures. Interestingly, feeding methionine-$^{13}C_5$,^{15}N not only showed a +1 mass shift indicating a methyl transfer, but also a +5, which indicated the incorporated homocysteine originates from methionine at the C-terminus, which is additionally supported by MS^2 analysis. This is consistent with the C-terminal γ-thiolactone structure.

Sorangibactin A1 and A2 showed no obvious bioactivity against our applied test panel of bacterial and fungal pathogens and selected cell lines (Table S4). However, siderophore-like behavior was observed during UHR-TOF measurements even before the purification of sorangibactin. Besides the double charged species at *m/z* 254.0632 and the $[M+H]^+$ signal at *m/z* 507.1195, the MS spectrum of sorangibactin A1 and A2 exhibits a peak at *m/z* 560.0310. With a mass distance of *m/z* 52.9115 to the $[M+H]^+$ peak it equals $[M-2H+Fe^{3+}]^+$ with the predicted sum formula of $C_{22}H_{24}FeN_4O_4S_3^+$ (*m/z* calcd for $[M-2H+Fe^{3+}]^+$ 560.0304, Δ1.1 ppm) and shows a characteristic M-2 iron isotope peak [21] (Figure S1). Additional evidence about sorangibactin being a siderophore was obtained by cultivation of MSr11367 under iron-deficient conditions using RGAE2 medium and iron-rich conditions by supplementing 1 mM $FeCl_3$. As a result, iron limitation during cultivation induced the production of sorangibactin, whereas external addition of iron suppressed the production (Figure S3). This is a well established indicator for siderophore features of molecules [22,23]. Apart from that, possibilities to determine the exact affinity to iron in terms of binding properties is commonly conducted via competition experiments with a strong binder like EDTA [24]. However, this approach was not possible for the sorangibactins due to their instability and isomerization issues during the purification process.

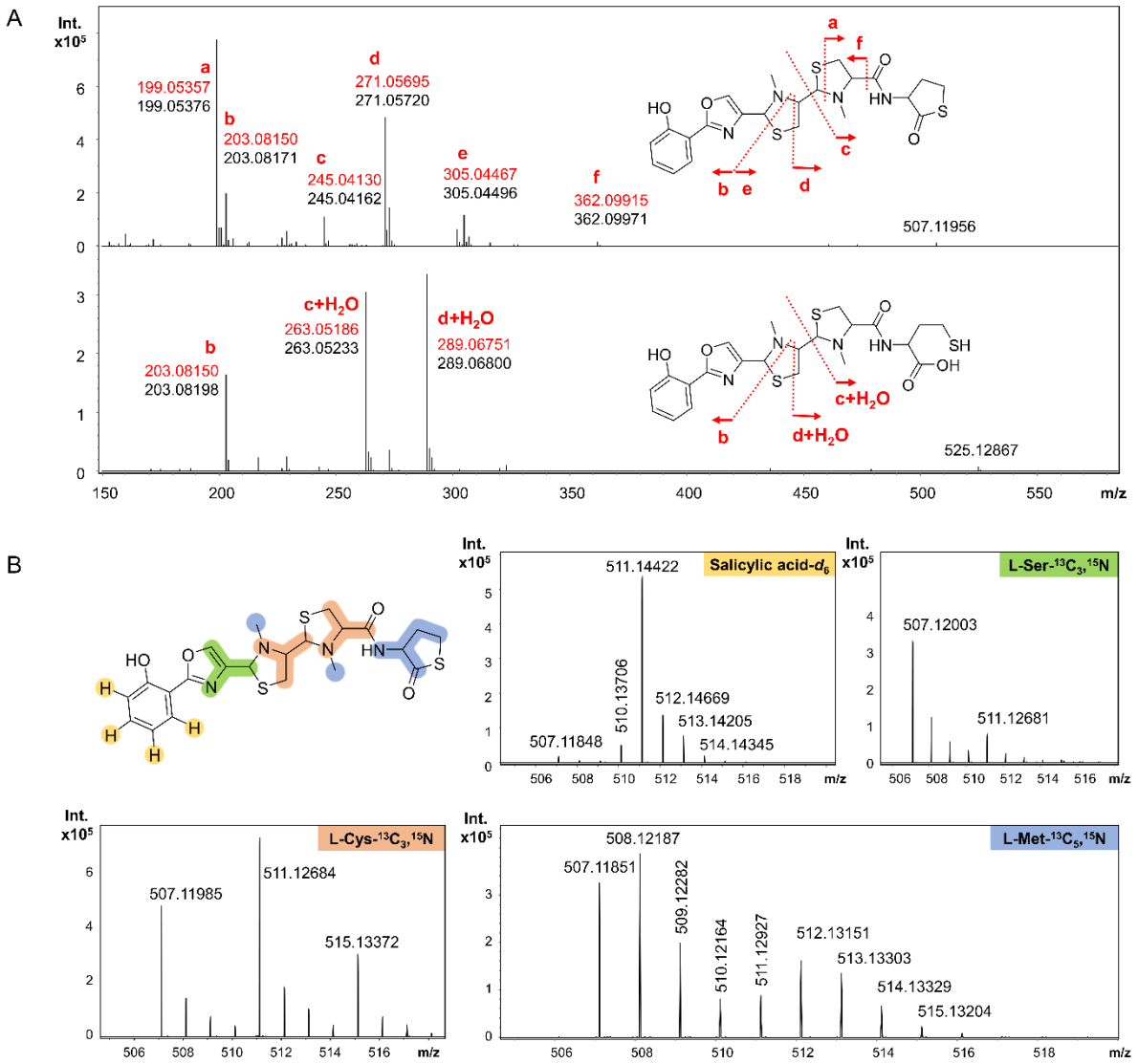


*Figure 3: Characterization of sorangibactins by MS² analysis and feeding stable isotopes. A) proposed fragmentation pattern of **1** and **2** with abundant fragment ions a-f (upper chromatogram) and of **3** with fragment ions b, c+H_2O and d+H_2O (lower chromatogram). The predicted m/z of each fragment is marked in red. B) feeding salicylic acid-d_6 (yellow), L-serine-$^{13}C_3$,^{15}N (green), L-cysteine $^{13}C_3$,^{15}N (orange) and L-methionine $^{13}C_5$,^{15}N (blue).*

Proposed biosynthesis of the sorangibactins

The sorangibactin BGC spans approximately 32 kb of DNA sequence, from which 17 proteins are annotated (Table S1) and proposed to be involved in biosynthesis (SrbA-SrbK), transportation (SrbL-SrbP) and regulation (SrbR). Besides *in silico* analysis, gene deletion and point mutation by Redαβ recombineering (Supplementary method and Figure S2) were carried out using the heterologous expression system. Based on those experiments we present the following biosynthesis hypothesis of the sorangibactins:

Biosynthesis of the sorangibactins initiated by salicylic acid formation catalyzed by SrbB, which shows 48.5% similarity with Irp9, a bifunctional salicylate synthase from *Yersinia enterocolitica* that can convert chorismate to salicylic acid in yersiniabactin biosynthesis [25]. Deletion of *srbB* abolished heterologous production of sorangibactins in *M. xanthus* DK1622, whereas feeding salicylic acid to *srbB* knockout mutant restored the production (Figure S4).

The salicylic acid was then adenylated and loaded to the first carrier protein of SrbC by SrbK (Figure 4A), a stand-alone A domain containing protein, which is predicted to use salicylic acid as substrate (Table S2). Deletion of *srbK* completely abolished the heterologous production of sorangibactins in *M. xanthus* DK1622. SrbC and SrbD are non-ribosomal peptide synthases, which comprise seven domains and five domains respectively (Figure 4A). Each adenylation domain was predicted by antiSMASH to use cysteine as a substrate (Table S2), however our biosynthesis model rather suggests the first adenylation domain in SrbC to activate serine instead, which is further modified to form the oxazole incorporated in the sorangibactin structure. This is consistent with the above-mentioned feeding L-serine-$^{13}C_3$,^{15}N (Figure 3B). The terminal thioesterase (TE) domain utilizes a cysteine as the active-site nucleophile instead of commonly used serine (Figure S7), which is not unprecedented but rare [26]. Point mutation of this cysteine to alanine or serine (Figure S8) abolished the heterologous production of sorangibactin A1 and A2 in *M. xanthus* DK1622, which indicates its essential role. Besides, this TE domain was proposed to select homocysteine or methionine as an intermolecular nucleophile for chain release resulting in one more amino acid incorporation at the C-terminus, which was confirmed by feeding L-methionine-$^{13}C_5$,^{15}N (Figure 3B). Using a free amino acid for chain release was previously found in threonine-tagged lipopeptides biosynthesis, where the TE domain selects intermolecular nucleophiles threonine hydroxy or amino groups yielding ester- and amide-linked threonine tags [27]. However, the complete C-terminal homocysteine thiolactone formation in sorangibactin biosynthesis is still unclear and whether other tailoring enzymes are involved is still under verification.

Besides SrbK and SrbB-SrbD for the core biosynthesis, SrbE-SrbJ are proposed to be tailoring enzymes that work *in trans*. SrbE is a putative oxidoreductase containing a saccharopine dehydrogenase NADP binding domain (PF03435.17) and SrbF/G are proposed to be methyltransferases. Gene deletion of SrbF/G was achieved by replacing with an ampicillin resistance gene followed by a Ptet promoter to drive the downstream expression avoiding polar effects (Figure S2), which resulted in the abolishment of sorangibactin A production and indicated their essential role for biosynthesis. SrbH is a cytochrome P450, the structure of which as predicted by ColabFold shows high similarity with the coelibactin P450 CYP105N1 from *Streptomyces coelicolor* A3(2) (Figure S9). Structure characterization of CYP105N1 revealed an unexpected wide open substrate binding pocket, which is much bigger than other P450 enzymes and may be responsible for binding the peptidyl carrier protein bound substrate, however the reaction catalyzed by CYP105N1 is still unclear [28,29]. Deletion of *srbH* abolished sorangibactin A1 and A2 production in *M. xanthus* DK1622, whereas another compound, termed sorangibactin C, was accumulated. Molecular ion clusters $[M+H]^+$ (m/z 509.13319) and $[M+2H]^{2+}$ (m/z 255.07163) in ESI-*hr*MS revealed the elemental formula $C_{22}H_{28}N_4O_4S_3$ of sorangibactin C, which shows highly similar fragmentation pattern with sorangibactin A and contains two more atoms of hydrogen in the first two cyclic moieties (Figure S5). Feeding L-serine-2,3,3-d_3 revealed that these two atoms of hydrogen originated from serine, and therefore oxazoline instead of oxazole is proposed for sorangibactin C structure. SrbH therefore presumably targets the carrier protein-bound substrate and catalyzes oxazoline dehydrogenation to form oxazole on the assembly line (Figure 4). The accumulation of sorangibactin C could indicate SrbH to be the only post-tailoring enzyme, but we cannot exclude the possibility that deletion of the other tailoring enzymes caused chemical instabilities of the resulting shunt products after release from the assembly line hindering their detection. SrbI shows 47.9% similarity with Irp3 from *Yersinia enterocolitica* (PDB: 5KVQ) and is assigned as *trans*-acting thiazolinyl imine reductase, catalyzing the NADPH-dependent reduction of a CP-tethered thiazoline ring [30,31]. Deletion of *srbI* abolished sorangibactin A1 and A2 production whereas the shunt product sorangibactin D observed at $[M+H]^+$ *m/z* 376.0424 (calculated $C_{16}H_{14}N_3O_4S_2^+$ 376.04202, Δ0.89 ppm) was accumulated (Figure S6), which indicates the thiazoline reduction is necessary for downstream processing otherwise resulting in early hydrolytic release. SrbJ is proposed to belong to the vicinal oxygen chelate (VOC) superfamily, that catalyzes diverse reactions [32]. The function of SrbJ in sorangibactins biosynthesis is unclear but indispensable, as of *srbJ* abolished sorangibactin A1 and A2 production. At this moment, we cannot fully assign each of these biosynthesis enzymes

to the structure, as no intermediates could be accumulated after gene knockout presumably due to the non-native intermediates stalling the whole machinery or chemical instabilities of the resulting shunt products.

In addition to the ten core genes described above that were all experimentally confirmed to be directly involved in sorangibactins biosynthesis, seven additional genes were found in the sorangibactin BGC. SrbA is a proposed tailoring enzyme containing an alpha/beta hydrolase domain. Deletion of *srbA* by Redαβ recombineering (Supplementary method and Figure S2) had no obvious influence on heterologous production of sorangibactins in *M. xanthus* DK1622, which can either be explained by a functional homologue of *srbA* present in the *M. xanthus* DK1622 chromosome or *srbA* being not involved in sorangibactins biosynthesis. SrbR is a LysR family transcriptional regulator and SrbL-SrbP are all transporter related. Whether or how these regulator and transporters are correlated to sorangibactin production however was not investigated here.

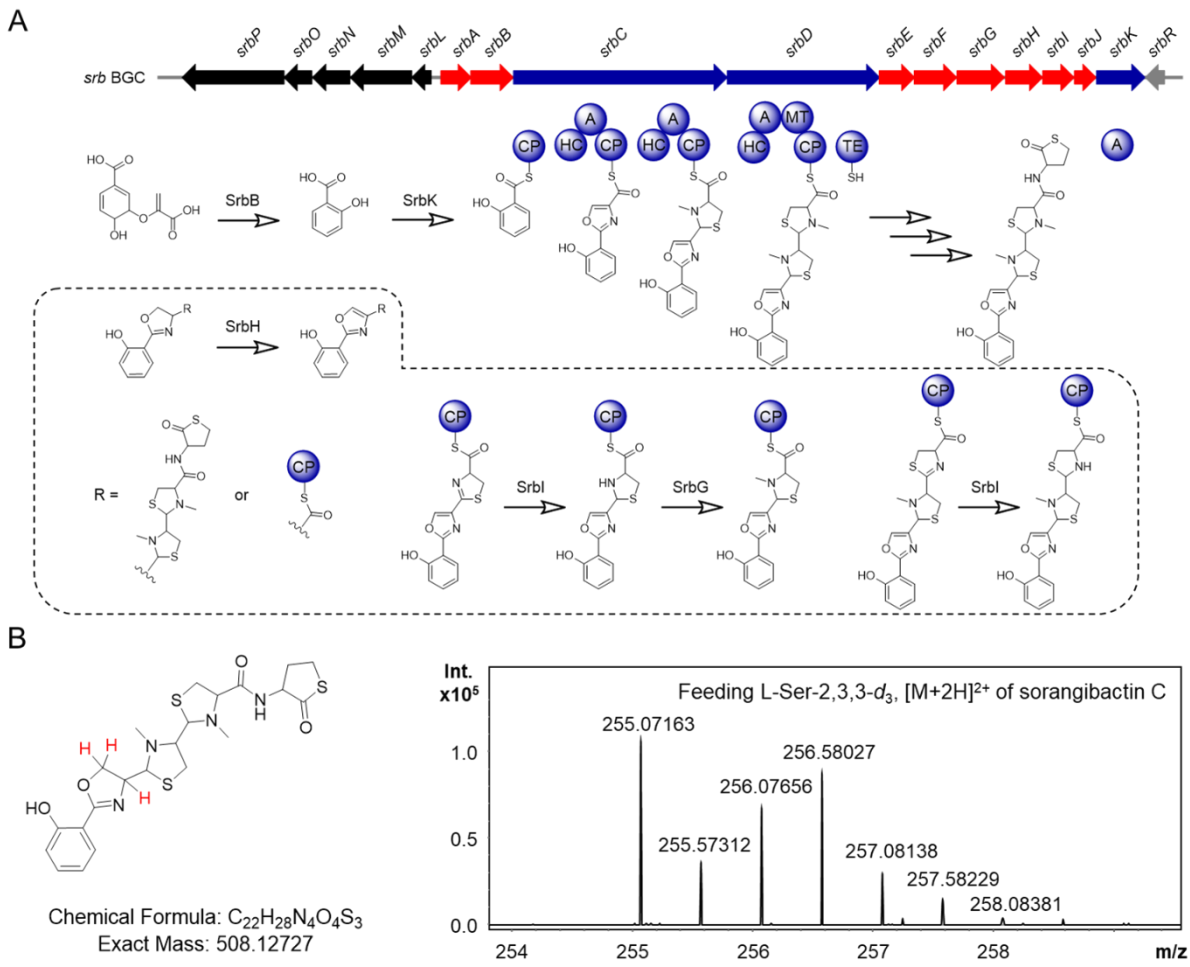


Figure 4: Characterization of the biosynthesis of sorangibactins. A) Gene cluster organization and proposed biosynthetic pathway. The core biosynthetic genes are marked in blue, the proposed additional biosynthetic genes are marked in red, the proposed transport-related genes and regulatory gene are marked in black and grey respectively. CP: carrier protein domain, HC: heterocyclization domain, A: adenylation domain, MT: methyltransferase domain, TE: thioesterase domain. B) Proposed structure and feeding L-Ser-2,3,3-d_3 of sorangibactin C. The proposed deuterium incorporation sites are marked in red.

3.3 Materials and methods

Isolation of sorangibactin A1, A2 and B

The heterologous producer strain *M. xanthus* DK1622 mutant containing the promoter engineered gene cluster was used for large-scale fermentation (SI). The myxobacterial extract in methanol was partitioned three times with hexane, dried and redissolved in water to be partitioned three times with first ethyl acetate and later chloroform. Subsequently, the obtained chloroform phase was redissolved in methanol and fractionated with an IsoleraTM Spektra One (Biotage®) system. Respective fractions containing sorangibactin A1 and A2 were further purified using a Dionex Ultimate 3000 SL system comprising a SWPS 3000 SL autosampler, P680 pump, TCC100 column oven heated to 45 °C, PDA100 UV-detector at 220 nm and an AFL 3000 fraction collector. In total 2.4 mg sorangibactin A1 and A2 mixture was obtained and stored under nitrogen at -20 °C until further use. For isolation of sorangibactin B, the fraction resulted from Biotage® work up was treated with hydrolyzing conditions followed by the same purification protocol compared to sorangibactin A1 and A2. Finally, 4.5 mg of sorangibactin B isomer mixture could be obtained.

LC-*hr*MS measurements

All *hr*MS measurements were conducted on a maXis 4G UHR-TOF system (Bruker Daltonics) equipped with an ESI source and coupled to a Dionex UltiMate 3000 rapid separation liquid chromatography (RSLC, Thermo Fisher) system. Separation was achieved with an Acquity BEH C18 column (100 mm × 2.1 mm, 1.7 µm dp) (Waters) in combination with a Waters VanGuard BEH C18 1.7 µm guard using a linear 5–95% gradient of acetonitrile with 0.1% formic acid in ddH_2O with 0.1% formic acid for 18 min. Detection was conducted by a diode array detector at 200–600 nm, the flow rate adjusted to 0.6 mL/min, the column heated to 45 °C and the LC flow split to 75 µL/min before entering the mass spectrometer. Following MS settings were applied for standard measurements: capillary voltage 4000 V, end plate off-set -500 V, nebulizer gas pressure 1 bar, dry gas flow rate 5 L/min, dry gas temperature 200 °C, mass scan range *m/z* 150–2500. MS/MS measurements were carried out in CID fragmentation mode with a collision energy of 5 eV. The mass spectrometer was calibrated in Quadratic + HPC calibration mode on the masses of the first isotope signal of sodium formate clusters, which are formed in the ion source after co-injection of a mixture of 1 mM sodium hydroxide in 1:1 water:isopropanol + 0.1% FA. Additional lock mass calibration was performed on *m/z* 622.0277, *m/z* 922.0103 and *m/z* 1221.9960.

Marfey's analysis for determination of stereochemistry

For the determination of stereo centers 7 and 12 (SI) of the N-methylated thiazolidines as well as stereo center 4 of the cyclic homocysteine thiolactone moiety, an adapted Marfey's derivatization protocol as already described for the thiamyxins [33] was applied after acidic hydrolysis to N-methyl cysteine and homocysteine, respectively.

NMR measurements

For structure elucidation of sorangibactin A1 and A2, all ^{1}H-, ^{13}C- and 2D-NMR spectra were acquired in methanol-d_4 on a Bruker Avance III (Ascend) 700 MHz spectrometer equipped with a 5 mm TCI cryoprobe using standard pulse programs. Observed chemical shifts (δ) were listed in ppm, coupling constants (*J*) in Hz and all spectra were calibrated with respective methanol signals at δ(^{1}H) = 3.31 ppm and δ(^{13}C) = 49.2 ppm. Sorangibactin B was measured using the same parameters, on a Bruker Avance III (Ultrashield) 500 MHz spectrometer.

3.4 Conclusions

Genome mining and subsequent heterologous expression enabled the discovery of a novel myxobacterial natural product termed sorangibactin from the myxobacterial strain MSr11367, highlighting accessibility of its biosynthetic potential that was not reflected in its metabolome in previous studies by this approach. *De novo* structure elucidation of sorangibactin revealed the presence of an unprecedented γ-thiolactone at the C-terminus, which may help to solve the puzzle of the chemical structure of coelibactin based on their gene cluster similarity via the adapted purification route reducing putative degradation and isomerization issues as well as by providing reference data for comparison to facilitate the structure elucidation. Both biosynthetic pathways contain conserved cytochrome P450 enzymes and the coelibactin P450 CYP105N1 protein was structurally characterized, however their functions remained elusive before this study. Here we discovered sorangibactin P450 SrbH, which is proposed to oxidize oxazoline to oxazole, representing the first P450-mediated oxazoline dehydrogenation in NRPS biosynthesis. Furthermore, the sorangibactin biosynthesis involved an additional thiazolinyl reductase and interestingly, an unusual TE domain containing a rare active site cysteine instead of commonly used serine. For the sorangibactin biosynthesis we propose that homocysteine or methionine is incorporated to the C-terminus by this unusual TE domain most probably together with other unassigned tailoring enzymes, resulting in the γ-thiolactone moiety formation. Therefore, the bioinformatically predicted coelibactin structure remains to be experimentally characterized. Future studies could address the indispensable SrbE (saccharopine dehydrogenase NADP-binding domain-containing protein, also present in coelibactin biosynthetic pathway), the function of which is yet unknown and SrbJ (glyoxalase-like domain-containing protein, absent in coelibactin biosynthetic pathway, but conserved in many other BGC homologues), which both are rare in natural products biosynthesis and worthy for further biochemical investigations. Homologous gene clusters are widespread not only in myxobacteria, but also in *Streptomyces* and *Pseudomonas* strains, which indicates its conservative functions across bacteria and potential for further discovery and characterization of sorangibactin analogues.

In summary, this study showcased the benefit of heterologous expression with subsequent gene cluster engineering for biosynthesis investigation for the discovery of uncharacterized myxobacterial BGCs. Accessibility of a cosmid library herein will allow further high throughput screening of cryptic myxobacteria BGCs by heterologous expression and facilitate characterization of their yet untapped biosynthetic potential.

Acknowledgements

We thank Dr. Ronald Garcia for providing the myxobacteria strain, Viktoria George and Alexandra Amann for bioactivity test, Dr. Daniel Krug and Dr. Nicolas A. Frank for helpful discussion, and Dr. Fabian Panter for assistance in metabolome analysis and compound purification. Y.G. acknowledges China Scholarship Council for a fellowship of the PhD program.

Conflict of interest

The authors declare no conflict of interest.

3.5 References

1. Gerth, K.; Pradella, S.; Perlova, O.; Beyer, S.; Müller, R. Myxobacteria: proficient producers of novel natural products with various biological activities - past and future biotechnological aspects with the focus on the genus *Sorangium*. *J. Biotechnol.* **2003**, *106*, 233–253, doi:10.1016/j.jbiotec.2003.07.015.
2. Herrmann, J.; Fayad, A.A.; Müller, R. Natural products from myxobacteria: novel metabolites and bioactivities. *Nat. Prod. Rep.* **2017**, *34*, 135–160, doi:10.1039/c6np00106h.
3. Bauman, K.D.; Butler, K.S.; Moore, B.S.; Chekan, J.R. Genome mining methods to discover bioactive natural products. *Nat. Prod. Rep.* **2021**, doi:10.1039/D1NP00032B.
4. 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, doi:10.1039/C8SC01325J.
5. Gavriilidou, A.; Kautsar, S.A.; Zaburannyi, N.; Krug, D.; Müller, R.; Medema, M.H.; Ziemert, N. Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat. Microbiol.* **2022**, *7*, 726–735, doi:10.1038/s41564-022-01110-2.
6. Huo, L.; Hug, J.J.; Fu, C.; Bian, X.; Zhang, Y.; Muller, R. Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* **2019**, doi:10.1039/c8np00091c.
7. Clos, J.; Zander-Dinse, D. Cosmid Library Construction and Functional Cloning. *Methods Mol. Biol.* **2019**, *1971*, doi:10.1007/978-1-4939-9210-2_6.

8. Wenzel, S.C.; Müller, R. Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Curr. Opin. Biotechnol.* **2005**, *16*, 594–606, doi:10.1016/j.copbio.2005.10.001.
9. Hider, R.C.; Kong, X. Chemistry and biology of siderophores. *Nat. Prod. Rep.* **2010**, *27*, 637–657, doi:10.1039/b906679a.
10. Kunze, B.; Bedorf, N.; Kohl, W.; Höfle, G.; Reichenbach, H. Myxochelin A, a new iron-chelating compound from *Angiococcus disciformis* (Myxobacterales). Production, isolation, physico-chemical and biological properties. *J. Antibiot.* **1989**, *42*, 14–17, doi:10.7164/antibiotics.42.14.
11. Nadmid, S.; Plaza, A.; Lauro, G.; Garcia, R.; Bifulco, G.; Müller, R. Hyalachelins A-C, unusual siderophores isolated from the terrestrial myxobacterium *Hyalangium minutum*. *Org. Lett.* **2014**, *16*, 4130–4133, doi:10.1021/ol501826a.
12. Kunze, B.; Trowitzsch-Kienast, W.; Höfle, G.; Reichenbach, H. Nannochelins A, B and C, new iron-chelating compounds from *Nannocystis exedens* (myxobacteria). Production, isolation, physico-chemical and biological properties. *J. Antibiot.* **1992**, *45*, 147–150, doi:10.7164/antibiotics.45.147.
13. Wang, W.; Qiu, Z.; Tan, H.; Cao, L. Siderophore production by actinobacteria. *Biometals* **2014**, *27*, 623–631, doi:10.1007/s10534-014-9739-2.
14. Bentley, S.D.; Chater, K.F.; Cerdeño-Tárraga, A.-M.; Challis, G.L.; Thomson, N.R.; James, K.D.; Harris, D.E.; Quail, M.A.; Kieser, H.; Harper, D.; et al. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **2002**, *417*, 141–147, doi:10.1038/417141a.
15. Timothy C. Johnstone; Elizabeth M. Nolan. Beyond iron: non-classical biological functions of bacterial siderophores. *Dalton Transactions* **2015**, *44*, 6320–6339, doi:10.1039/C4DT03559C.
16. Yang, Y.-J.; Singh, R.P.; Lan, X.; Zhang, C.-S.; Li, Y.-Z.; Li, Y.-Q.; Sheng, D.-H. Genome editing in model strain *Myxococcus xanthus* DK1622 by a Site-Specific Cre/loxP Recombination System. *Biomolecules* **2018**, *8*, doi:10.3390/biom8040137.
17. Jang, J.-P.; Kwon, M.C.; Nogawa, T.; Takahashi, S.; Osada, H.; Ahn, J.S.; Ko, S.-K.; Jang, J.-H. Thiolactomide: A New Homocysteine Thiolactone Derivative from Streptomyces sp. with Neuroprotective Activity. *Journal of Microbiology and Biotechnology* **2021**, *31*, 1667–1671, doi:10.4014/jmb.2108.08015.
18. Hubert Haag; Klaus Hantke; Hartmut Drechsel; Igor Stojiljkovic; Günther Jung; Hans Zähner. Purification of yersiniabactin: a siderophore and possible virulence factor of

Yersinia enterocolitica. *Microbiology* **1993**, *139*, 2159–2165, doi:10.1099/00221287-139-9-2159.

19. Fife, T.H.; Natarajan, R.; Shen, C.C.; Bembi, R. Mechanism of thiazolidine hydrolysis. Ring opening and hydrolysis of 1, 3-thiazolidine derivatives of p-(dimethylamino) cinnamaldehyde. *J. Am. Chem. Soc.* **1991**, *113*, 3071–3079, doi:10.1021/ja00008a041.
20. Schulz, M.; Gaitanoglou, V.; Mantel, O.; Hövelmann, Y.; Hübner, F.; Dobrindt, U.; Humpf, H.-U. Metabolomics Study on Pathogenic and Non-pathogenic E. coli with Closely Related Genomes with a Focus on Yersiniabactin and Its Known and Novel Derivatives. *Metabolites* **2020**, *10*, 221, doi:10.3390/metabo10060221.
21. Hermenau, R.; Mehl, J.L.; Ishida, K.; Dose, B.; Pidot, S.J.; Stinear, T.P.; Hertweck, C. Genomics-Driven Discovery of NO-Donating Diazeniumdiolate Siderophores in Diverse Plant-Associated Bacteria. *Angew. Chem.* **2019**, doi:10.1002/anie.201906326.
22. Neilands, J.B. Siderophores: structure and function of microbial iron transport compounds. *J. Biol. Chem.* **1995**, *270*, 26723–26726.
23. Steven W. Wilhelm; Charles G. Trick. Iron-limited growth of cyanobacteria: Multiple siderophore production is a common response. *Limnology and Oceanography* **1994**, *39*, 1979–1984, doi:10.4319/lo.1994.39.8.1979.
24. Robert D. Perry; Paul B. Balbo; Heather A. Jones; Jacqueline D. Fetherston; Edward DeMoll. Yersiniabactin from Yersinia pestis: biochemical characterization of the siderophore and its role in iron transport and regulation. *Microbiology* **1999**, *145*, 1181–1190, doi:10.1099/13500872-145-5-1181.
25. Kerbarh, O.; Ciulli, A.; Howard, N.I.; Abell, C. Salicylate biosynthesis: Overexpression, purification, and characterization of Irp9, a bifunctional salicylate synthase from Yersinia enterocolitica. *J. Bacteriol.* **2005**, *187*, 5061–5066, doi:10.1128/JB.187.15.5061–5066.2005.
26. Little, R.F.; Hertweck, C. Chain release mechanisms in polyketide and non-ribosomal peptide biosynthesis. *Nat. Prod. Rep.* **2022**, *39*, 163–205, doi:10.1039/d1np00035g.
27. Thongkongkaew, T.; Ding, W.; Bratovanov, E.; Oueis, E.; Garcı A-Altares, M.A.; Zaburannyi, N.; Harmrolfs, K.; Zhang, Y.; Scherlach, K.; Müller, R.; et al. Two types of threonine-tagged lipopeptides synergize in host colonization by pathogenic Burkholderia species. *ACS Chem. Biol.* **2018**, doi:10.1021/acschembio.8b00221.
28. Lim, Y.-R.; Hong, M.-K.; Kim, J.-K.; Doan, T.T.N.; Kim, D.-H.; Yun, C.-H.; Chun, Y.-J.; Kang, L.-W.; Kim, D. Crystal structure of cytochrome P450 CYP105N1 from

Streptomyces coelicolor, an oxidase in the coelibactin siderophore biosynthetic pathway. *Arch. Biochem. Biophys.* **2012**, *528*, 111–117, doi:10.1016/j.abb.2012.09.001.

29. Zhao, B.; Moody, S.C.; Hider, R.C.; Lei, L.; Kelly, S.L.; Waterman, M.R.; Lamb, D.C. Structural analysis of cytochrome P450 105N1 involved in the biosynthesis of the zincophore, coelibactin. *Int. J. Mol. Sci.* **2012**, *13*, 8500–8513, doi:10.3390/ijms13078500.
30. Meneely, K.M.; Lamb, A.L. Two structures of a thiazolinyl imine reductase from Yersinia enterocolitica provide insight into catalysis and binding to the nonribosomal peptide synthetase module of HMWP1. *Biochemistry* **2012**, *51*, 9002–9013, doi:10.1021/bi3011016.
31. Meneely, K.M.; Ronnebaum, T.A.; Riley, A.P.; Prisinzano, T.E.; Lamb, A.L. Holo Structure and Steady State Kinetics of the Thiazolinyl Imine Reductases for Siderophore Biosynthesis. *Biochemistry* **2016**, *55*, 5423–5433, doi:10.1021/acs.biochem.6b00735.
32. He, P.; Moran, G.R. Structural and mechanistic comparisons of the metal-binding members of the vicinal oxygen chelate (VOC) superfamily. *J. Inorg. Biochem.* **2011**, *105*, 1259–1272, doi:10.1016/j.jinorgbio.2011.06.006.
33. Patrick A. Haack; Kirsten Harmrolfs; Chantal D. Bader; Ronald Garcia; Antonia P. Gunesch; Sibylle Haid; Alexander Popoff; Alexander Voltz; Heeyoung Kim; Ralf Bartenschlager; et al. Thiamyxins: Structure and Biosynthesis of Myxobacterial RNA-Virus Inhibitors. *Angew. Chem. Int. Ed. (Angewandte Chemie International Edition in English)* **2022**, doi:10.1002/anie.202212946.

S3. Supporting Information

Genome Guided Discovery of the Myxobacterial Thiolactone-containing Sorangibactins

Yunsheng Gao[a,b,c,+], Christine Walt[a,c,+], Chantal Bader[a,c], Rolf Müller[a,b,c,*]

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

[b]Helmholtz International Lab for Anti-Infectives, Campus E8.1, 66123 Saarbrücken, Germany

[c]German Center for Infection Research (DZIF), Partner Site Hannover-Braunschweig, 38124 Braunschweig, Germany

[+]These authors contributed equally to this work

*To whom correspondence should be addressed: rolf.mueller@helmholtz-hips.de

S3.1. Methods

Cultivation conditions for bacterial strains

The myxobacterial *Sorangiineae* strain MSr11367 was cultivated in 100 mL RGAE2 medium (3.5 g/L casitone, 1 g/L soytone, 0.5 g/L yeast extract, 2 g/L glucose, 2 g/L soluble starch, 1 g/L sucrose, 1 g/L maltose monohydrate, 2 g/L cellobiose, 0.5 g/L calcium chloride dihydrate, 1 g/L magnesium sulfate heptahydrate, 2.38 g/L HEPES, pH=7.0) or CYHv2 medium (4 g/L soya meal starch, 3 g/L casitone, 2 g/L glucose, 8 g/L soluble starch, 1.5 g/L yeast extract, 1 g/L calcium chloride dihydrate, 1 g/L magnesium sulfate heptahydrate, 11.9 g/L HEPES, 8 mg/L Fe-EDTA, pH=7.2) in a 300 mL flask with shaking at 180 rpm and 30 °C for 14 days. *Myxococcus xanthus* DK1622 was used as a heterologous host for gene cluster expression and cultivated in CTT medium (10 g/L casitone, 10 mM Tris-HCl, 8 mM magnesium sulfate, 1 mM potassium phosphate, pH=7.6) at 30 °C. *E. coli* strains GB2005, GB05-red, GBred-gyrA462 were used for gene cluster engineering and cultivated in LB medium (10 g/L tryptone, 5 g/L yeast extract, 5 g/L sodium chloride, pH=7.0) at 37 °C. The following antibiotics were supplemented to the medium when the cultivated strain carries the corresponding resistance gene: 10 μg/mL for oxytetracycline, 15 μg/mL for chloramphenicol, 50 μg/mL for kanamycin and 100 μg/mL for ampicillin.

Gene cluster cloning and engineering

The cosmid library for *Sorangiineae* strain MSr11367 was constructed by Explogen LLC (EXG) (https://explogen.com.ua) using genomic DNA and p15A-cm-tet-int vector. After end-sequencing, the cosmid pCos-D05_9223 was found to be able to cover the putative whole gene cluster and used for heterologous expression. For generation of promoter engineered expression construct (pCos-D05_9223-km-Pvan), *km-vanR-Pvan* cassette amplified by PCR was electroporated into *E. coli* GB05-red harboring plasmid pCos-D05_9223 and expressing Redαβ recombinase for linear-circular homologous recombination (LCHR). Gene knockout constructs (pCos-D05_9223-km-Pvan-delA and pCos-D05_9223-km-Pvan-delAB) were generated in a similar manner except reverse oligos containing different downstream homology arms were used. Deletion of *srbE/F/G/H/I/J/K* was based on pCos-D05_9223-km-Pvan. To do this, *amp-Ptet* cassette amplified by PCR using corresponding oligo nucleotides (Table S3) was used to replace the target gene using LCHR. As a result the downstream genes will be under control of a Ptet promoter in order to avoid potential polar effect. For point mutation of the TE domain, RedEx method was used [1], *amp-ccdB* cassette containing homology arms, mutation

sites and PmeI sites was amplified by PCR using corresponding oligoes (Table S3) and electroporated into *E. coli* GBred-gyrA462 harboring plasmid pCos-D05_9223-km-Pvan for LCHR. The resulting plasmids were digested by PmeI, recircularized by Gibson Assembly and electroporated into *E. coli* GB2005. Correct recombinants were confirmed by restriction digestions and Sanger sequencing of the oligo regions.

Heterologous expression in *M. xanthus* DK1622

Expression constructs containing the sorangibactin BGC were transferred to the heterologous host *M. xanthus* DK1622 by electroporation according to a previously established procedure [2]. The gene cluster was then integrated into the *attB* locus on the *M. xanthus* DK1622 chromosome through phage Mx8 integrase mediated site-specific recombination [3]. Then, the mutants were checked by colony PCR using three pairs of primers (Table S3) targeting different positions of the gene cluster to ensure intact integration. Three independent mutants were inoculated in CTT medium supplemented with appropriate antibiotics and cultivated at 30 °C with shaking until late log-phase. The culture was inoculated 1 to 50 into fresh CTT medium supplemented with 2% XAD-16 resin, appropriate antibiotics and inducer (1 mM vanillate for Pvan promoter and 0.5 μg/mL anhydrotetracycline for Ptet promoter). Cultivation was continued at 30 °C with shaking for 3 days. For comparative analysis, triplicate cultivation of *M. xanthus* DK1622 wild-type strains ware performed in a similar manner except that no antibiotics were added. The cells and adsorber resin were collected by centrifugation and extracted with methanol. The crude extracts were dried by rotary evaporation *in vacuo* and redissolved in 1 mL methanol for HPLC-*hr*MS measurement.

HPLC-*hr*MS measurement data analysis

Data processing was performed using Bruker DataAnalysis 4.4 and MetaboScape 2021b. For comparative analysis, bucket tables consisting of molecular features (RT-*m/z* pairs) and their corresponding intensity values were generated using the T-ReX 3D algorithm. Principal component analysis (PCA) and t-Test were used to find unique features that are only present in the heterologous expression group but not in the control group. These unique features were double-checked manually by extracting the corresponding ion chromatograms.

Feeding experiment

The stable isotopes were prepared as stock solutions of 0.2 M and supplemented to the culture at the final concentration of 0.2 mM. To prepare stock solutions, L-cysteine-$^{13}C_3$,^{15}N (Sigma

Aldrich, 658057), L-serine-$^{13}C_3$,^{15}N (Sigma Aldrich, 608130), L-serine-2,3,3-d_3 (Cambridge Isotope Laboratories, DLM-582-0.5), L-methionine-d_3 (methyl-d_3) (Campro Scientific GmbH, CS01-182_581) and L-methionine-$^{13}C_5$,^{15}N (Campro Scientific GmbH, CS01-181_212) were dissolved in autoclaved Milli-Q water, 2-hydroxybenzoic acid-d_6 (Campro Scientific GmbH, CS01-182_507) was dissolved in ethanol. The heterologous host *M. xanthus* DK1622 :: km-Pvan-srb was inoculated in 20 mL CTT medium supplemented with kanamycin at 30 °C with shaking until mid log-phase. After addition of 1 mM vanillate, 2% XAD-16 resin and corresponding stable isotope, cultivation was continued for two more days. The cells and adsorber resin were collected and extracted with methanol. The crude extracts were dried by rotary evaporation *in vacuo* and redissolved in 1 mL methanol for HPLC-*hr*MS measurement.

Fermentation and compound purification

The heterologous producer strain *M. xanthus* DK1622 :: km-Pvan-srb was inoculated in CTT medium supplemented kanamycin and cultivated at 30 °C with shaking until late log-phase. The culture was inoculated 1 to 50 into 6 × 2 L fresh CTT medium supplemented with 2% XAD-16 resin, 50 µg/mL kanamycin and 1 mM vanillate in 6 × 5 L flaskes. Cultivation was continued at 30 °C with shaking for 4 days. The cells and adsorber resin were collected by centrifugation and extracted stepwise with methanol. The methanol extract was concentrated and partitioned with hexane to remove nonpolar impurity. After evaporation *in vacuo*, the methanol phase was then dissolved in Milli-Q water and extracted twice with equal portions of chloroform. The chloroform phase was evaporated to dryness and redissolved in methanol.

IsoleraTM Spektra One (Biotage®) system

Normal phase chromatography was performed using silica gel (60 Å, 70-230 mesh, 63-200 µm) packed in a SNAP 50g column as stationary phase and a multistep gradient with hexane, ethyl acetate and methanol as mobile phase using a 15 column volume (CV) gradients and 5 CV for equilibration at a flowrate of 50 mL/min.

Semipreparative purification of sorangibactin A1 and A2

Separation was reached using a Waters XBridge Peptide BEH C18 column (250x10 mm, 5 µm) in combination with a mobile phase consisting of A=0.1% acetic acid in ddH_2O and B=0.1% acetic acid in acetonitrile running a 5-95% B gradient in 16 min with a 4 min equilibration at 5% B with a flow rate of 5 mL/min. The acid labile sorangibactin A1 and A2 were pooled in a

flask filled with $NaHCO_3$ for neutralization purpose, which was subsequently disjoint through liquid-liquid partitioning three times using chloroform.

Semipreparative purification of sorangibactin B

Separation was achieved via a Waters XBridge Peptide BEH C18 column (250x10 mm, 5 µm) using formic acid instead of acetic acid and following adapted multi-step gradient: hold at 5% B for 1 min, increasing to 20% B within 0.5 min, 1.5-12.5 min increasing to 80% B, increasing to 95% B within 0.5 min, hold at 95% B for 1.5 min, go back to 5% B within 0.5 min and reequilibrate for 5 min at a flow rate of 5 mL/min. The obtained fraction was kept at room temperature overnight to achieve conversion to sorangibactin B.

Marfey's analysis for determination of stereochemistry

Marfey's derivatization protocol was applied after acidic hydrolysis to N-methyl cysteine and homocysteine, respectively. Therefore, about 0.1 mg of sorangibactin A1 and A2 mixture was hydrolyzed in 100 µL of 3 N HCl at 110 °C for 45 min. After evaporation of the solvent at 110 °C, two 50 µL aliquots of hydrolyzed sorangibactin A1 and A2 in ddH_2O were supplemented with 60 µL of 1 N $NaHCO_3$ each and 20 µL of the derivatization reagent (1% 1-fluroro-2,4-dinitrophenyl-5-leucine-amide solution in acetone [D-FDLA and L-FDLA]). The obtained mixture was reacted at 40 °C and 700 rpm for 2 h. Addition of 60 µL 3 N HCl stopped the reaction, and the solution was diluted with 150 µL of methanol and acetonitrile to be analyzed with our maXis4G system coupled to a Dionex Ultimate 3000 SL system. The measurement was done in positive ESI mode after separation by a Waters Acquity BEH C18 column (50x2.1 mm, 1.7 µm) with following multi-step gradient (A=0.1% formic acid in ddH2O and B=0.1% formic acid in acetonitrile): 5-10% B in 1 min, 10-35% B in 14 min, 35-55% in 7 min, 55-80% in 3 min, hold at 80% for 1 min and reequilibration at 5% B for 5 min. The flow was split to 75 µL/min before entering the MS. UV data were recorded at 340 nm simultaneously with the MS detection in centroid mode ranging from *m/z* 150-1000.

S3.2. Supplementary Tables

Table S1: Annotation of the sorangibactin biosynthetic gene cluster. BLAST search was performed using the blastp algorithm with default parameters in the NCBI non-redundant protein sequences database.

Name	Length (aa)	Deduced function	Closest homologue	Coverage/identity (%)
SrbA	317	alpha/beta hydrolase	PYQ66785.1	89/44.70
SrbB	472	salicylate synthase	WP_020740538.1	93/61.88
SrbC	2146	non-ribosomal peptide synthetase	WP_020740537.1	98/65.29
SrbD	1814	non-ribosomal peptide synthetase	KYF49860.1	99/68.72
SrbE	371	saccharopine dehydrogenase NADP-binding domain-containing protein	WP_020740533.1	100/61.89
SrbF	482	bifunctional Gfo/Idh/MocA family oxidoreductase/class I SAM-dependent methyltransferase	WP_020740532.1	80/49.21
SrbG	516	class I SAM-dependent methyltransferase	AGP40737.1	99/54.26
SrbH	394	cytochrome P450	KYG09918.1	99/74.10
SrbI	358	thiazolinyl imide reductase	KYF49855.1	97/74.86
SrbJ	237	VOC family protein	WP_044986807.1	95/67.11
SrbK	538	AMP-binding protein	WP_044986805.1	99/73.64
SrbL	190	TonB-dependent siderophore receptor	WP_245591859.1	55/60.58
SrbM	675	TonB-dependent siderophore receptor	WP_020740545.1	100/66.62
SrbN	407	hypothetical protein	KYF49844.1	90/62.77
SrbO	297	ABC transporter ATP-binding protein	WP_020740540.1	95/79.79
SrbP	1130	ABC transporter permease	KYG09926.1	99/68.62
SrbR	93	LysR family transcriptional regulator	WP_089851301.1	83/45.45

Table S2: Substrate specificities of adenylation domains in srb gene cluster. Prediction was performed by NRPSpredictor2 incorporated in antiSMASH.

Domain	Specificity-conferring code	Predicted substrate	Actual substrate
SrbC-A1	DLYNLGLIHK	Cys	Ser
SrbC-A2	DLFNLSMIWK	Cys	Cys
SrbD-A	DLYNLSLIWK	Cys	Cys
SrbK-A	ALPAQGVLNK	Sal	Sal

Table S3: Oligo nucleotides used in this study.

Oligoes	**Sequence (5' to 3')**	**Application**
cos-chk-F	TCGTGCTGCTGTCATTGGTGTGC	end-sequencing of cosmid
cos-chk-R	ATAGCGCCCCTGTGTGTTCTCGT	
km-vanR-Pvan-F	ACGAGGCACGCGCTCGCTGCCACCAGGGCAGCACAGAAGACAAGTCTCATCATA ATCTGTACCTCCTTAAGTCAGAAGAACTCGTCAAGAAGGC	promoter engineering
km-vanR-Pvan-R	GCTGCGCCTCGAGCCTCCTGCCCCTGCGCGGACACGGCACACAGGATTCCCATAT GCGTTTCCTCGCATCG	
km-vanR-Pvan-delA-R	ACGTACGTCGGCCTATCGGACATACCGTCCTCAGCGTACTGTCGATATTTCATAT GCGTTTCCTCGCATCG	knockout of *srbA*
km-vanR-Pvan-delAB-R	AGCACGCGGGCGATCGTGCCCCGCACATCCAGGAAACTATCGTGACTCGGCATA TGCGTTTCCTCGCATCG	knockout of both *srbA* and *srbB*
amp-Ptet-delE-F	TCATCGCCGTGCTCGGTGCTTACGGCGATGTGGGGCGCCAGGCCACGCGGTTAC CAATGCTTAATCAGTGAGG	knockout of *srbE*
amp-Ptet-delE-R	CGTCTCGACCAGCACGGAGATCCCTCGTTCGAGCTGCTCGCCGATGCTCATAGAT CCTTTCTCCTCTTTAGATC	
amp-Ptet-delF-F	GCATCGGCGAGCAGCTCGAACGAGGGATCTCCGTGCTGGTCGAGACGCCGTTAC CAATGCTTAATCAGTGAGG	knockout of *srbF*
amp-Ptet-delF-R	TCGAGGGTCCGGCACCACGCGGCGAGCAAGGCCGAGAAGTCGATGATCATAGAT CCTTTCTCCTCTTTAGATC	
amp-Ptet-delG-F	TCATCGACTTCTCGGCCTTGCTCGCCGCGTGGTGCCGGACCCTCGATCCGTTACC AATGCTTAATCAGTGAGG	knockout of *srbG*
amp-Ptet-delG-R	TCGATGCCCTCGAAGTCGTTGAACGGATACGGTCGGGGGGAGTCCGTCATAGAT CCTTTCTCCTCTTTAGATC	
amp-Ptet-delH-F	CTCGACCAAAAGGATTCGACCCATGACGGACTCCCCCCGACCGTATCCGTTTACC AATGCTTAATCAGTGAGG	knockout of *srbH*
amp-Ptet-delH-R	AAGGTAAATGCGGCCAAAGGCCGTGCCCGCGACGAGGACGCGACGTTTCATAGA TCCTTTCTCCTCTTTAGATC	
amp-Ptet-delI-F	AACGTCGCGTCCTCGTCGCGGGCACGGCCTTTGGCCGCATTTACCTTGCCTTACC AATGCTTAATCAGTGAGG	knockout of *srbI*
amp-Ptet-delI-R	TGGTGAATGTCGCGAACCCAGAGCAAAACGTGGCTACAGGACGTGGTCATAGAT CCTTTCTCCTCTTTAGATC	
amp-Ptet-delJ-F	CCACGTTTTGCTCTGGGTTCGCGACATTCACCAGGCGGTGCGCGACTACCTTACC AATGCTTAATCAGTGAGG	knockout of *srbJ*
amp-Ptet-delJ-R	CGGTACGTCGCCGCGAACTCCGCAGGCCAATTTACGAAGCCTTCGAGCATAGAT CCTTTCTCCTCTTTAGATC	
amp-Ptet-delK-F	ATTGGCCTGCGGAGTTCGCGGCGACGTACCGCGAACTCGGCTACTGGCGTTTACC AATGCTTAATCAGTGAGG	knockout of *srbK*
amp-Ptet-delK-R	CCGCGAGGCTGCGCGGCGATCCGTGCGGCCAGCTCGCGCTTGTTGACCTTAGATC CTTTCTCCTCTTTAGATC	
TE-C1586A-F	CGCAGCGCCTCGTGGACGAGGGCTTCGAAGAGTTCCAGCTCATTGGCTATGCGC TCGGCGGGCTGCTGGCCATGTTTAAACTTTGTTCAAAAAAAA	TE point mutation
TE-C1586S-F	CGCAGCGCCTCGTGGACGAGGGCTTCGAAGAGTTCCAGCTCATTGGCTATTCGCT CGGCGGGCTGCTGGCCATGTTTAAACTTTGTTCAAAAAAAA	
TE-C1586-R	CCGCGCTCGAGCAGACGGCGCGCCACCTCGATGGCCAGCAGCCCGCCGAGGTTT AAACTTTGTTTATTTTTCT	
srb-chk01-F	CCAAGTACAAGCAGCTCGTCGCA	colony PCR
srb-chk01-R	GCCGTCGTCGTCGATGAACATCA	
srb-chk02-F	CTACCGCATCCAAGTCGAAGACC	
srb-chk02-R	CGATGCAGGATCGTGAGGTTCAC	
srb-chk03-F	TGGCTTCGATTACCGCACGTTGG	
srb-chk03-R	GGCGTCCTTGGGTATCGAATACG	

Table S4: Bioactivity results of sorangibactin A1 and A2 mixture.

Microbial strains	**MIC [µg/ml]**
E. coli acrB [JW0451-2]	64
P. aeruginosa PA14 [DSM-19882]	>64
E. coli WT [BW25113]	>64
B. subtilis [DSM-10]	64
S. aureus [Newman]	64
A. baumannii [DSM-30008]	>64
C. freundii [DSM-30039]	>64
M. hiemalis [DSM-2656]	32
P. anomala [DSM-6766]	64
C. albicans [DSM-1665]	>64
C. neoformans [DSM-11959]	64
M. smegmatis [MC2155]	>64
Cell line	**IC_{50} [µg/mL]**
HepG2	8.952 ± 5.65

S3.3. Supplementary figures

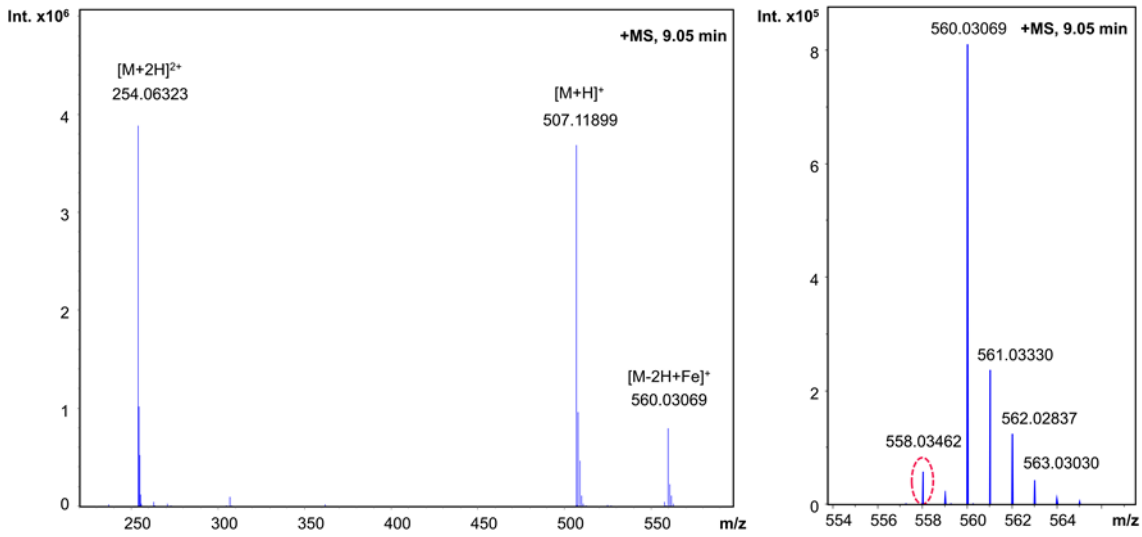


Figure S1: UHR-TOF measurements of sorangibactin A1 and A2 containing extract showing the siderophore-like behavior by insource iron chelating: double charged species at m/z *254.0632, single charged species at* m/z *507.1195 and [M-2H+Fe^{3+}]$^{+}$ species at* m/z *560.0310 with a characteristic M-2 iron isotope peak at* m/z *558.0346.*

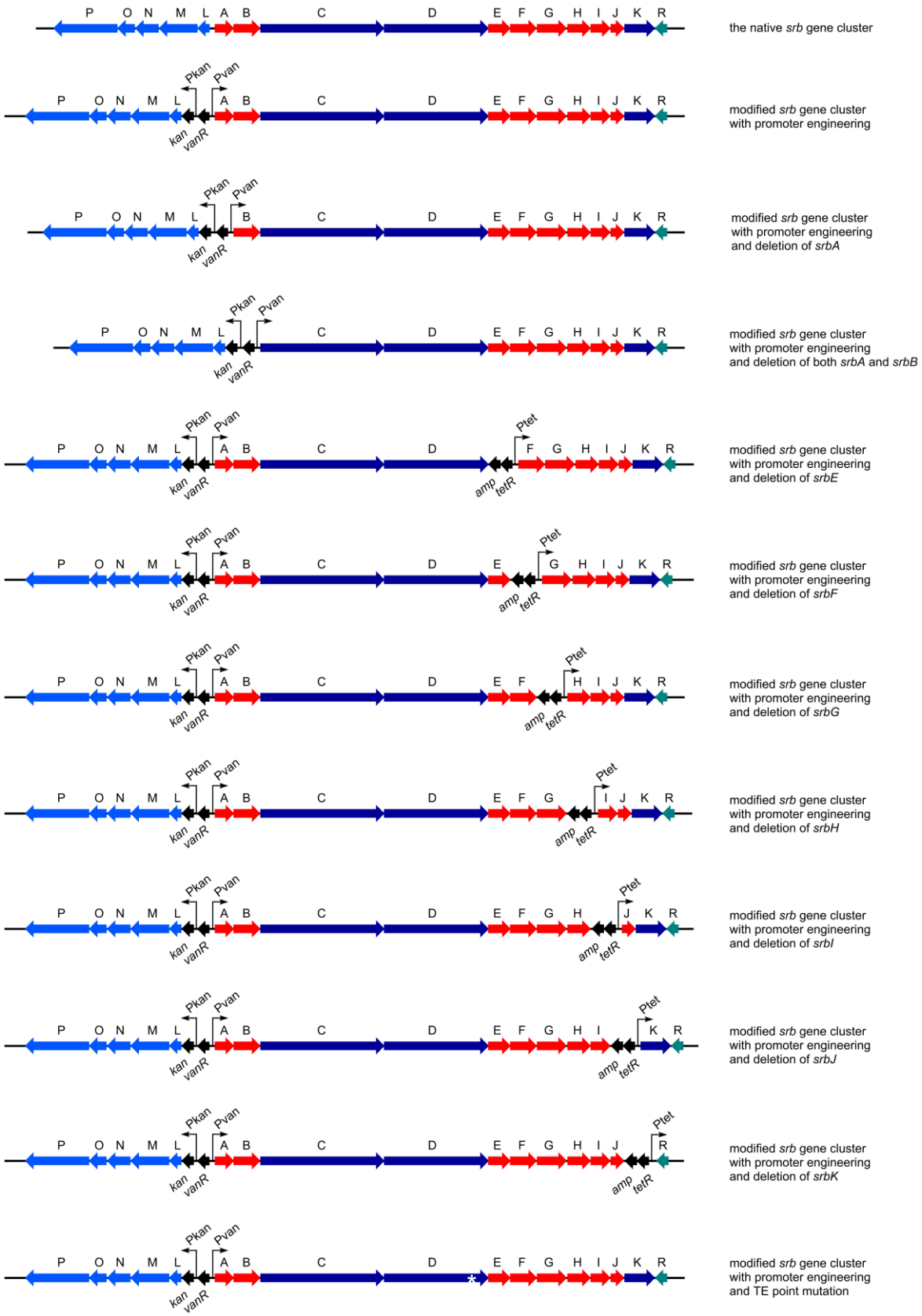


Figure S2: Expression constructs used in this study.

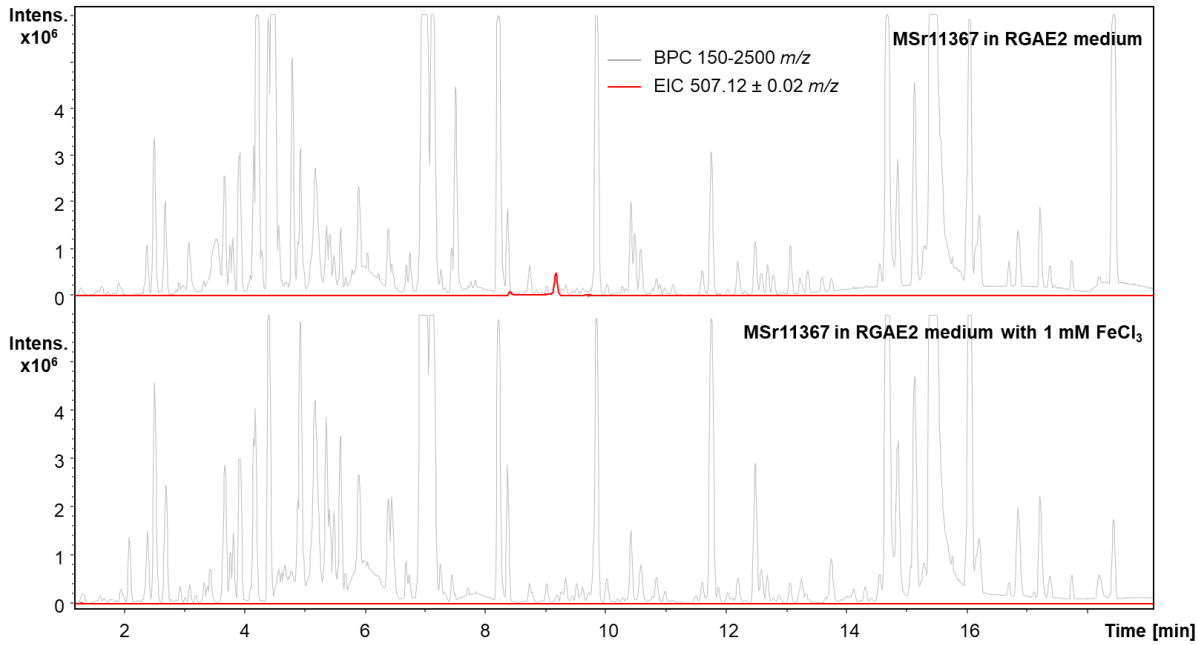


Figure S3: Production of sorangibactin A by MSr11367 with and without iron addition.

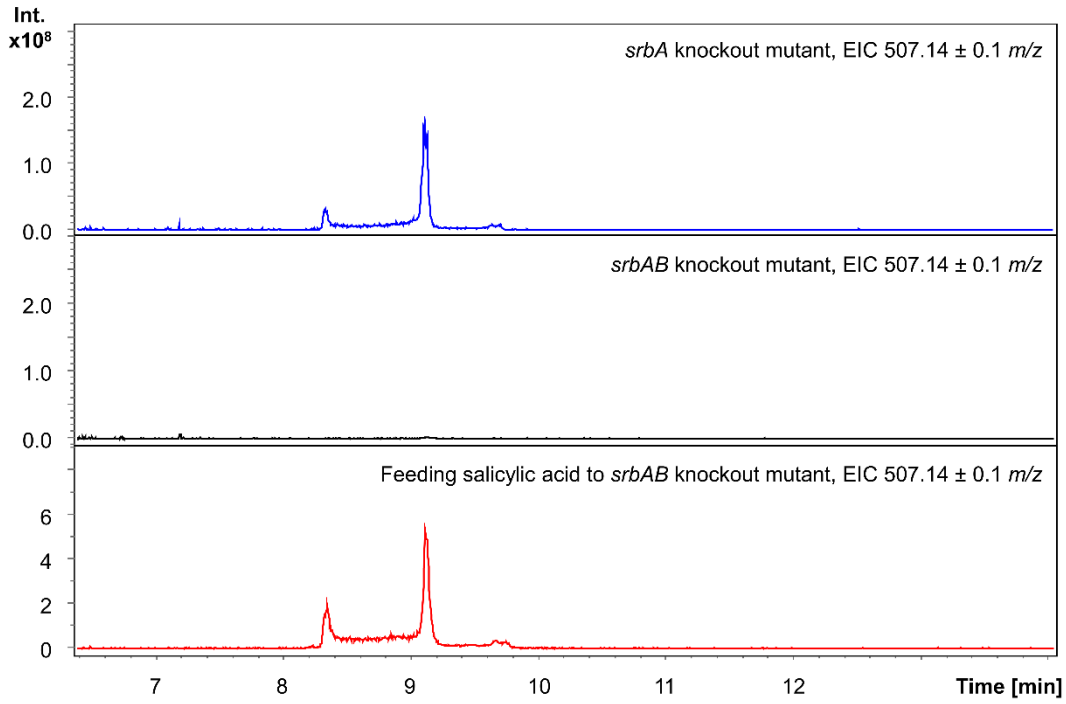


Figure S4: Deletion of srbA (blue), srbAB (black) and feeding salicylic acid to srbAB (red) knockout mutant.

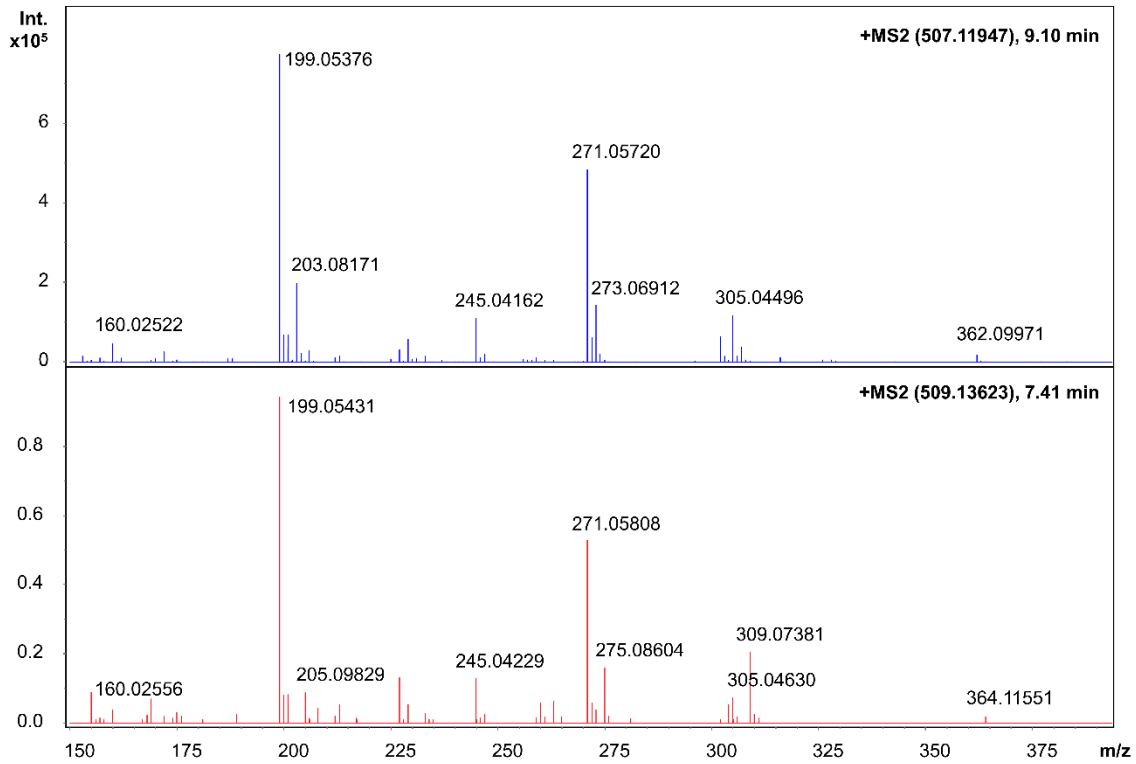


Figure S5: hrMS^2 of sorangibactin C (red) in comparison with sorangibactin A (blue).

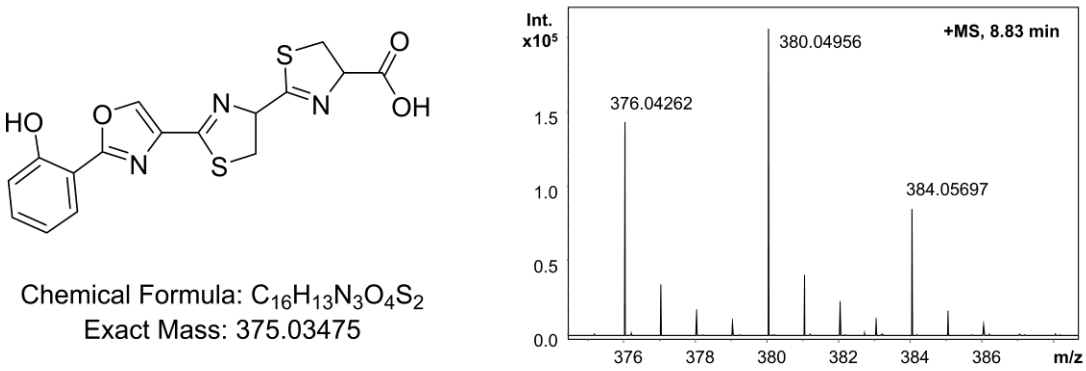


Figure S6: Proposed structure and feeding L-Cys-$^{13}C_3$,^{15}N of sorangibactin D.

Figure S7: Alignment of TE domains. The sorangibactin TE domain is aligned with coelibactin TE, yersiniabactin TE and 10 structure available TEs from PDB database. The catalytic motif (normally GxSxG) is indicated in red dashed box.

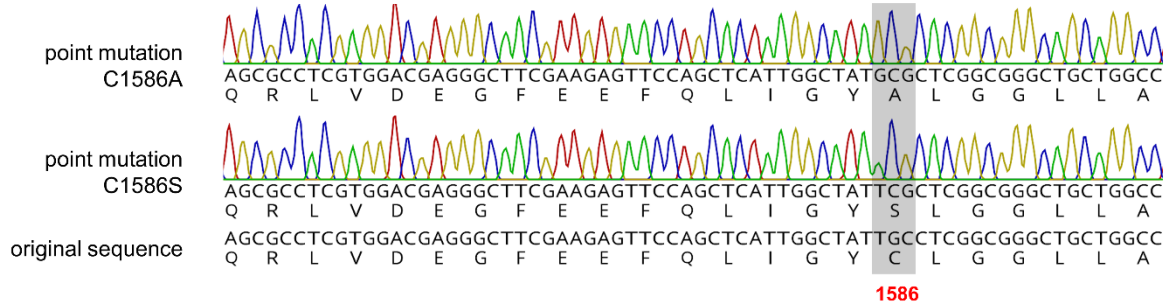


Figure S8: Point mutation of C1586A and C1586S of the sorangibactin TE domain.

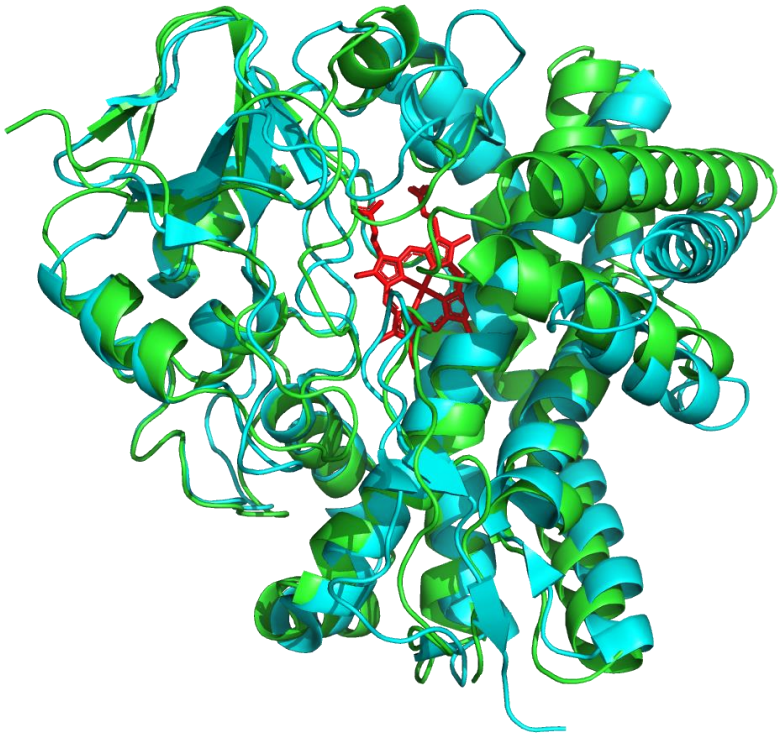


Figure S9: Superimposition of SrbH (Green color, structure predicted by ColabFold [4]) and CYP105N1 (Cyan color, PDB 3TYW) from Streptomyces coelicolor A3(2). Heme is shown as red stick.

S3.4. Structure determination

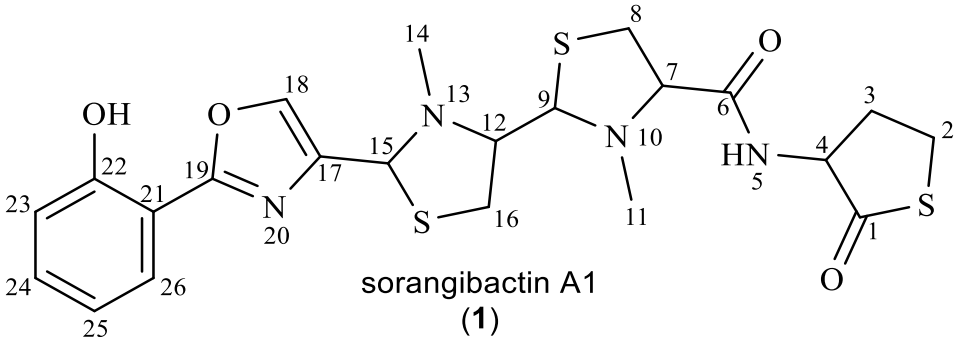
Structure elucidation of sorangibactin A1 and A2 mixture. ^{1}H-NMR, ^{13}C and HSQC spectra show five aromatic double-bond signals at $\delta(^{1}H)$ = 7.89 (1H, s) $\delta(^{13}C)$ = 136.4, $\delta(^{1}H)$ = 7.82 (1H, d, J = 7.80 Hz) $\delta(^{13}C)$ = 127.2, $\delta(^{1}H)$ = 7.36 (1H, m) $\delta(^{13}C)$ = 133.7, $\delta(^{1}H)$ = 7.00 (1H, d, J = 8.38Hz) $\delta(^{13}C)$ = 118.2 and $\delta(^{1}H)$ = 6.95 (1H, m) $\delta(^{13}C)$ = 120.7 ppm. Furthermore, five methine signals can be found at $\delta(^{1}H)$ = 5.01 (1H, s) $\delta(^{13}C)$ = 70.1, $\delta(^{1}H)$ = 4.21 (1H, m) $\delta(^{13}C)$ = 76.1, $\delta(^{1}H)$ = 4.09 (1H, dd, J = 6.94, 3.66 Hz) $\delta(^{13}C)$ = 63.1, $\delta(^{1}H)$ = 3.59 (1H, t, J = 7.95Hz) $\delta(^{13}C)$ = 81.2 as well as $\delta(^{1}H)$ = 3.29 (1H, m) $\delta(^{13}C)$ = 75.6 ppm. Additionally, two methyl signals are located at $\delta(^{1}H)$ = 2.57 (3H, s) $\delta(^{13}C)$ = 41.5 and $\delta(^{1}H)$ = 2.47 (3H, s) $\delta(^{13}C)$ = 40.8, one methylene signal at $\delta(^{1}H)$ = 3.01 (2H, m) $\delta(^{13}C)$ = 24.9 and three diastereotopic methylene signals at $\delta(^{1}H)$ = 3.15/3.02 (2H, m) $\delta(^{13}C)$ = 35.3, $\delta(^{1}H)$ = 3.29, 3.16 (2H, m) $\delta(^{13}C)$ = 33.8 and $\delta(^{1}H)$ = 2.10, 1.92 (2H, m) $\delta(^{13}C)$ = 24.7 ppm. Sorangibactin A1 consists of five ring systems, starting from an N-terminal with a salicylic acid, which carbonyl group is fused into an oxazole connected to two consecutive N-methylated thiazolidines linked via an amide bond to a cyclic homocysteine thiolactone C-terminus. The salicylic acid moiety is assigned by corresponding COSY correlations between the four aromatic double bond signals at $\delta(^{1}H)$ = 7.82, 7.36, 7.00 and 6.95 ppm in combination with HMBC correlations to the quaternary carbon, which is deshielded with a shift at $\delta(^{13}C)$ = 158.5 ppm due to the hydroxyl substituent and a quaternary carbon with a rather upfield shift at $\delta(^{13}C)$ = 112.4 ppm. The remaining aromatic signal exhibits a characteristic downfield shift at $\delta(^{1}H)$ = 7.89 ppm with distinct HMBC correlations to two quaternary carbons at $\delta(^{13}C)$ = 162.9 and 144.5 ppm pointing towards a heterocyclic ring with two double bonds to maintain aromatization. The ^{1}H-^{15}N-HMBC spectrum uncovers the corresponding correlation to a deshielded nitrogen atom with a shift at $\delta(^{15}N)$ = 239.9 ppm characteristic for oxygen containing heteroaromatic systems as oxazole moieties. The linkage between salicylic acid and oxazole is the deshielded quaternary carbon at $\delta(^{13}C)$ = 162.9 ppm due to HMBC correlations from the salicylic acid signal at $\delta(^{1}H)$ = 7.82 ppm. This carbon atom represents the fused carbonyl group of the salicylic acid. The connection to the adjacent ring is represented by the corresponding COSY and HMBC correlations from the methine signal at $\delta(^{1}H)$ = 5.01 ppm to the oxazole moiety. Corresponding 2D NMR correlations underpin this ring to be an N-methylated thiazolidine cycle. The diastereotopic methylene signal could be observed at $\delta(^{1}H)$ = 3.29 and 3.16 ppm as well as the methyl signal at $\delta(^{1}H)$ = 2.57 ppm with HMBC correlations to the methine signal at $\delta(^{1}H)$ = 5.01 ppm and the one at $\delta(^{1}H)$ = 3.29 ppm,

respectively. Further on, the COSY spectrum exhibits cross peaks for the latter methine signal to the methine signal of the subsequent N-methylated thiazolidine ring at $\delta(^{1}H)$ = 4.21 ppm. The corresponding protons of this ring are more shielded in comparison to the previous are explained by its greater distance to the deshielding oxazole moiety. This is supported by the ^{1}H-^{15}N HMBC spectrum with corresponding cross peak signals for the nitrogen atom in the third ring at $\delta(^{15}N)$ = 59.8 ppm and the one in the fourth cycle at $\delta(^{15}N)$ = 56.7 ppm. Finally, the fourth and fifth ring system are linked via an amide group with a shift of $\delta(^{13}C)$ = 167.9 in line with HMBC correlations from the thiazolidine signals at $\delta(^{1}H)$ = 3.59 and 3.15, 3.02 ppm on the one side and the methine signal at $\delta(^{1}H)$ = 4.09 ppm from the last ring on the other side. This moiety features COSY correlations from a diastereotopic signal at $\delta(^{1}H)$ = 2.10/1.92 ppm to the mentioned methine group as well as to another methylene group at $\delta(^{1}H)$ = 3.01 ppm. Moreover, corresponding HMBC correlations from this methine and diastereotopic methylene signals to the deshielded quaternary carbon with a shift of $\delta(^{13}C)$ = 179.5 ppm indicate a cyclic homocysteine thiolactone C-terminus, which is in line with already published homocysteine thiolactones like thiolactomide [5]. The respective deviations between our measurements in methanol-d_4 and the reference data of thiolactomide in dimethylsulfoxid-d_6 are <0.5 ppm for ^{1}H- and <6 ppm for ^{13}C-shifts. However the quaternary carbon of the thioester shows a greater difference with 26.3 ppm, which can be explained by the different surrounding functions of shielding thiazolidins and deshielding alkyl chains. Further, the missing correlation in the ^{1}H-^{15}N-HMBC spectrum might be due to a possibly unfavorable conformation for the detection of the $^{3}J(^{1}H$-$^{15}N)$-coupling from the diastereotopic methylene group to the nitrogen atom of the amide bond.

The two signal sets for sorangibactin A1 and A2 were identical except for signals corresponding to the C-terminal homocysteine thiolactone due to isomerization upon purification. The methine signal for sorangibactin A1 can be found at $\delta(^{1}H)$ = 4.09 ppm and for A2 at $\delta(^{1}H)$ = 4.16 ppm, whereby the adjoining diastereotopic methylene group of sorangibactin A1 is located at $\delta(^{1}H)$ = 2.10/1.92 ppm and A2 shows a methylene group signal without any diastereotopic effect at $\delta(^{1}H)$ = 2.04 ppm.

The structure elucidation of sorangibactin B is explained in the main text.

Table S5: NMR spectroscopic data of sorangibactin A1 in methanol-d_4 at 700/175 MHz.



#	Δ ^{13}C [PPM]	Δ ^{1}H [PPM], MULT (*J* [HZ])	Δ ^{15}N [PPM]	COSY	HMBC	^{1}H-^{15}N HMBC
1	179.5	-	-	-	-	-
2	24.9	3.01, m	-	3	3,4	-
3	24.7	2.10,1.92, m	-	2,4	1,2,4	-
4	63.1	4.09, dd (6.94,3.66)	-	3	1,2,6	-
5	-	-	-	-	-	-
6	167.9	-	-	-	-	-
7	81.2	3.59, t (7.95)	-	8	6,8,9,11	-
8	35.3	3.15,3.02, m	-	7	6,7,9	10
9	76.0	4.21, m	-	12,16	11,16	-
10	-	-	56.7	-	-	-
11	40.8	2.47, s	-	-	7,9	10
12	75.7	3.29, m	-	9,16	9,16	-
13	-	-	59.8	-	-	-
14	41.9	2.57, s	-	-	12,15	13
15	70.1	5.01, s	-	18	12,14,16,17,18	20
16	33.8	3.29,3.16, m	-	9,12	14,15	13
17	144.5	-	-	-	-	-

18	136.4	7.89, s	-	15	17,19	20
19	162.9	-	-	-	-	-
20	-	-	239.9	-	-	-
21	112.4	-	-	-	-	-
22	158.5	-	-	-	-	-
23	118.2	7.00, d (8.38)	-	24	21,22,25	19,20
24	133.7	7.36, m	-	23,25	22,26	19,20
25	120.7	6.95, m	-	24,26	21,22,23,24,26	-
26	127.2	7.82, d (7.80)	-	25	19,22,24	-

Sorangibactin A1

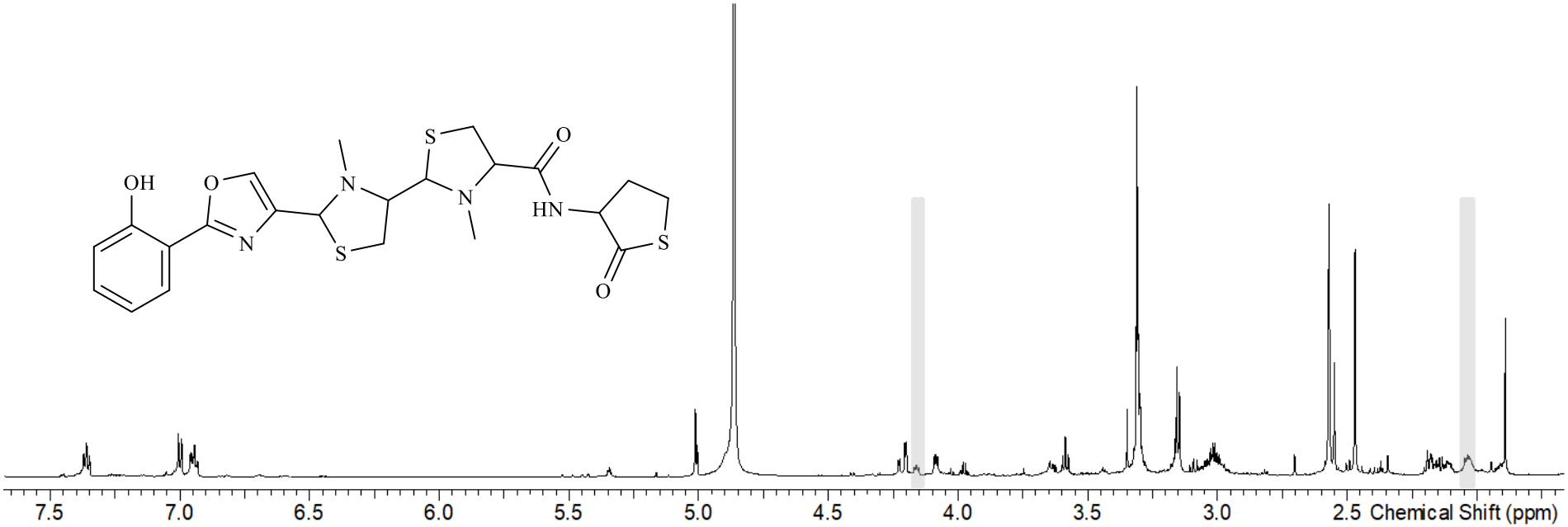


Figure S10: ^{1}H-spectrum of sorangibactin A1 in methanol-d_4 at 700 MHz; signals corresponding to sorangibactin A2 highlighted in grey.

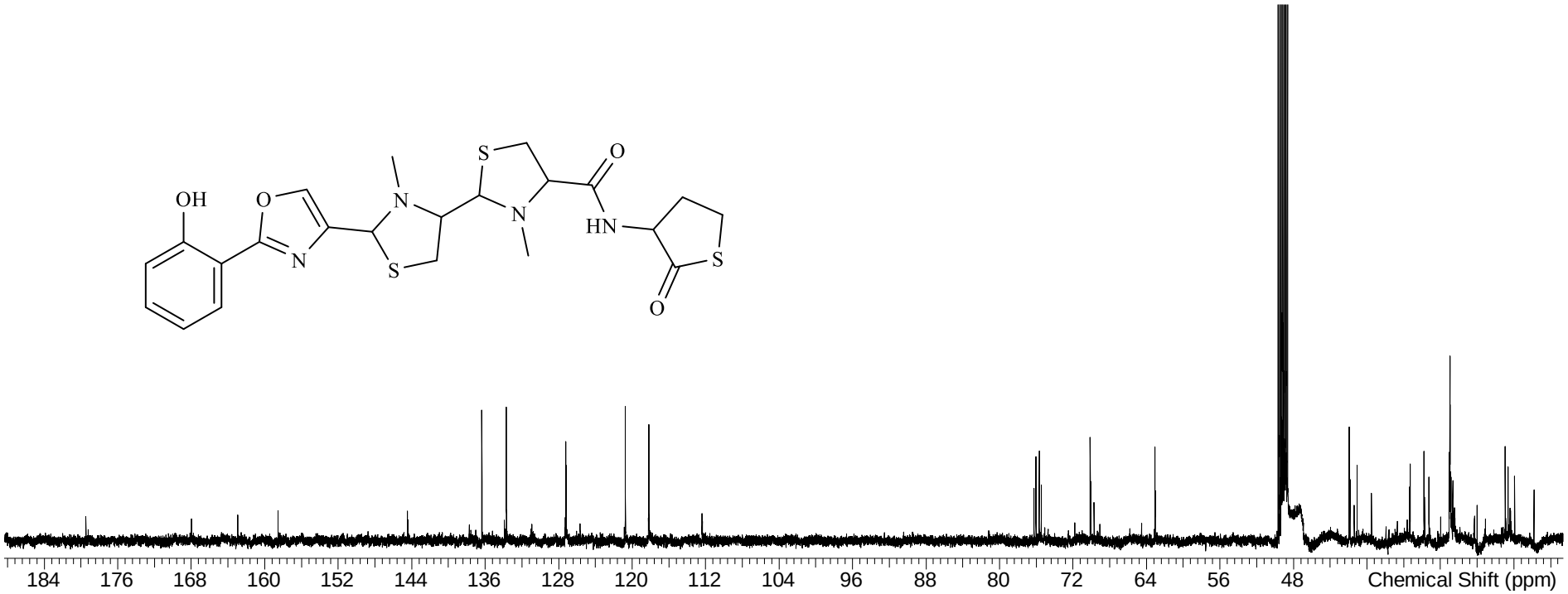
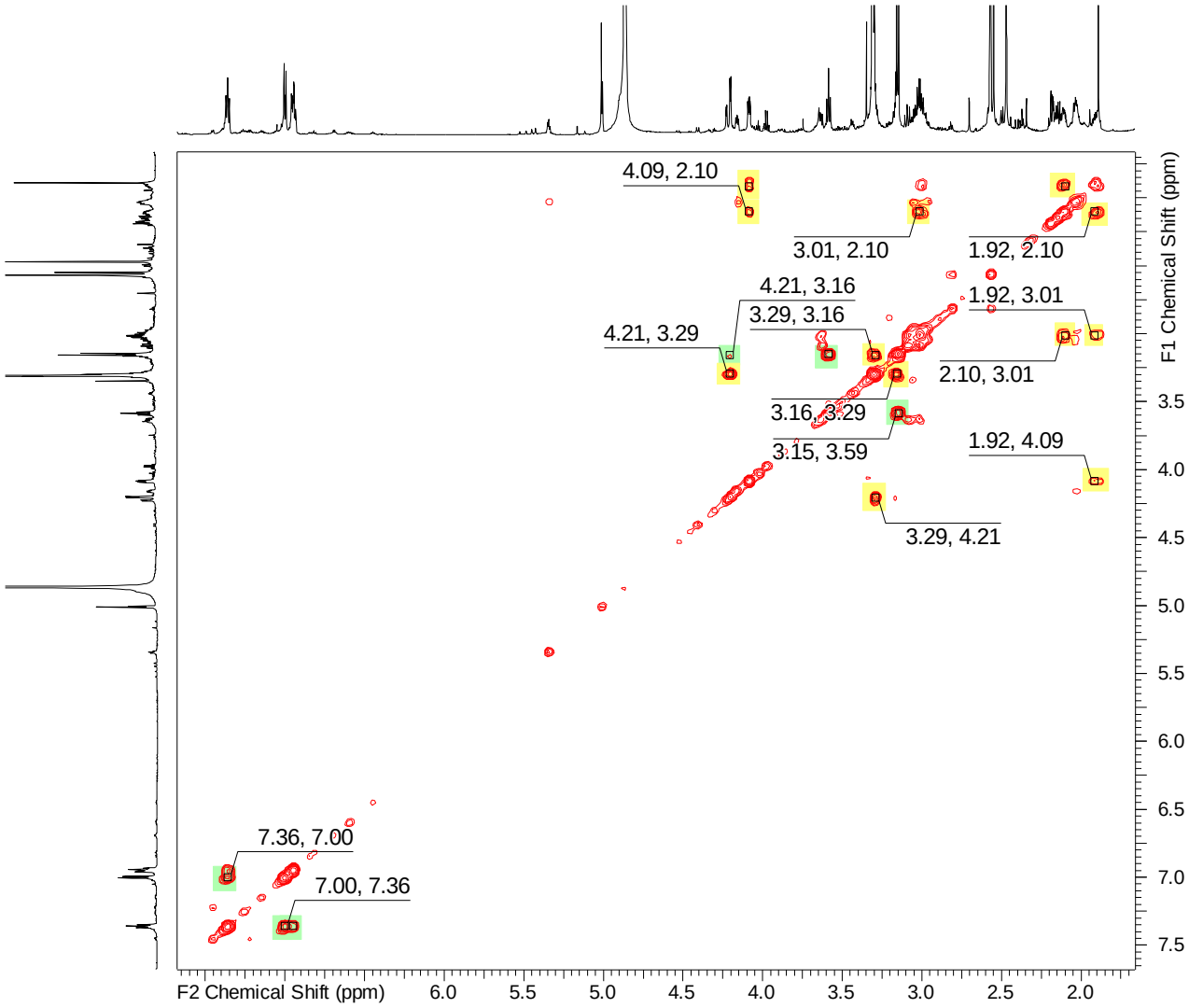


Figure S11: ^{13}C-spectrum of sorangibactin A1 in methanol-d_4 at 175 MHz.



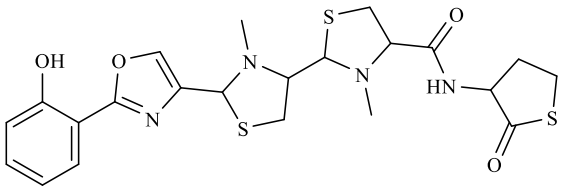
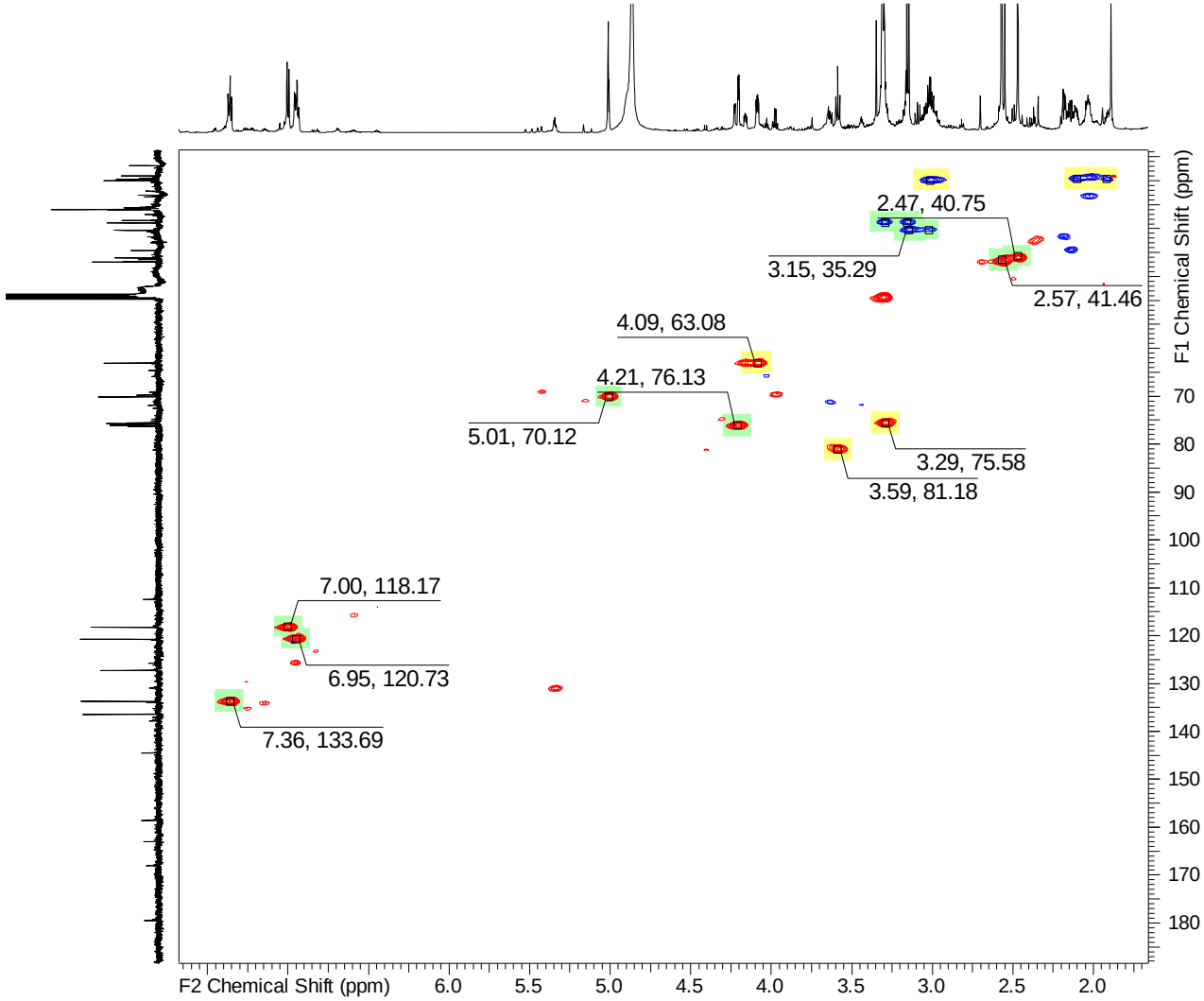


Figure S12: COSY spectrum of sorangibactin A1 in methanol-d_4 at 700 MHz.



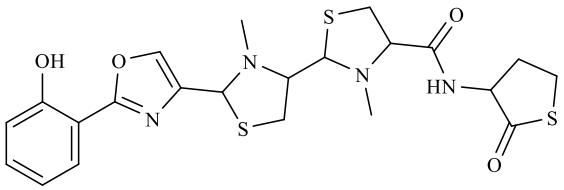


Figure S13: HSQC spectrum of sorangibactin A1 in methanol-d_4 at 700 MHz (^{1}H)/175 MHz (^{13}C).

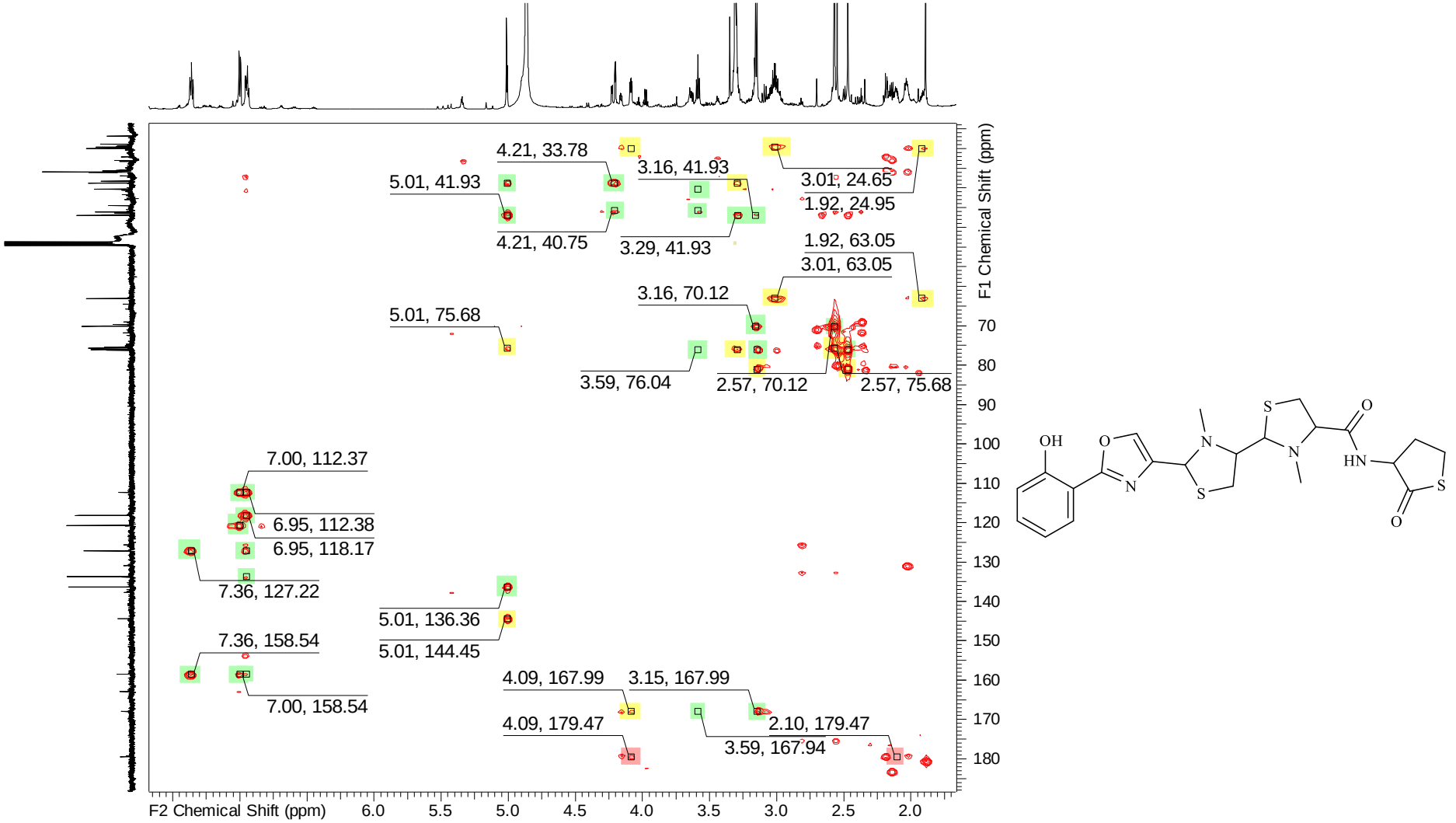


Figure S14: HMBC spectrum of sorangibactin A1 in methanol-d_4 at 700 MHz (^{1}H)/175 MHz (^{13}C).

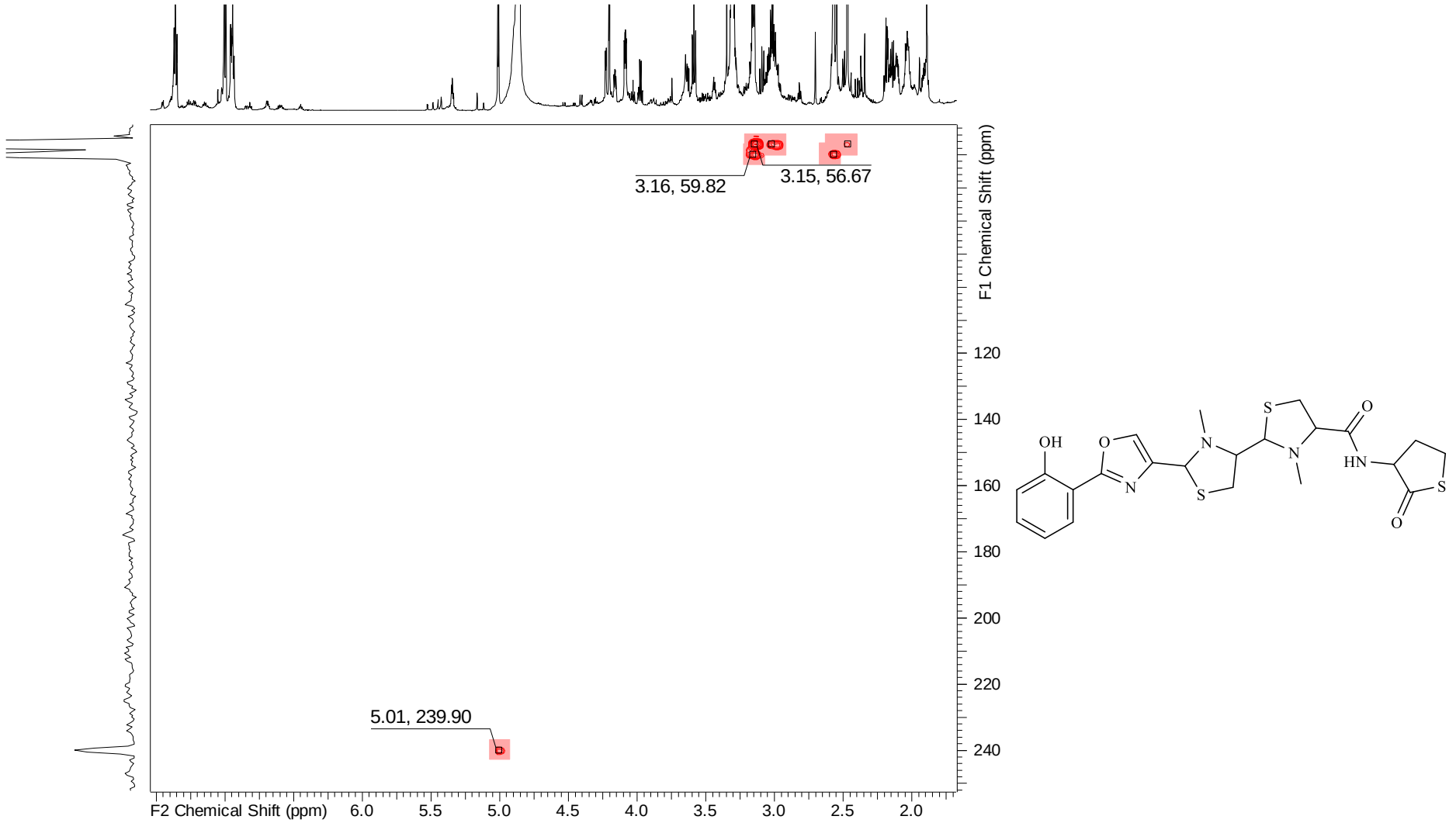
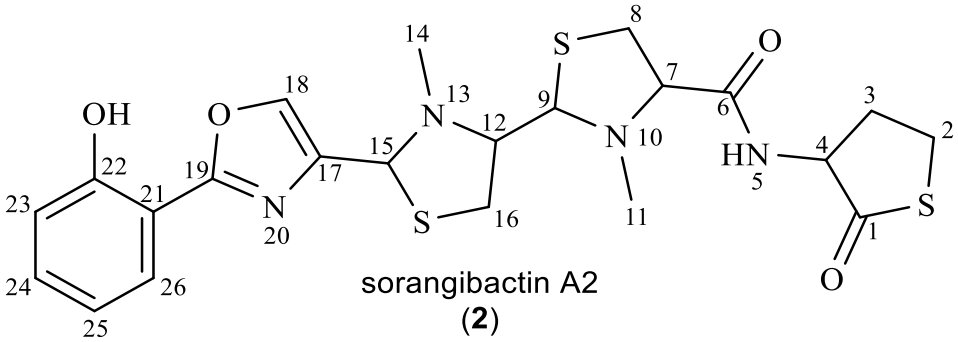


Figure S15: ^{1}H-^{15}N-HMBC spectrum of sorangibactin A1 in methanol-d_4 at 500 MHz (^{1}H)/125 MHz (^{15}N).

Table S6: NMR spectroscopic data of sorangibactin A2 in methanol-d_4 at 700/175 MHz.



#	Δ ^{13}C [PPM]	Δ ^{1}H [PPM], MULT (*J* [HZ])	Δ ^{15}N [PPM]	COSY	HMBC	^{1}H-^{15}N HMBC
1	179.5	-	-	-	-	-
2	24.9	3.01, m	-	3	4	-
3	23.9	2.04, m	-	2,4	1,2,4	-
4	63.1	4.16, dd (6.79,4.29)	-	3	1,2,6	-
5	-	-	-	-	-	-
6	167.9	-	-	-	-	-
7	81.2	3.59, t (7.95)	-	8	6,8,9,11	-
8	35.3	3.15,3.02, m	-	7	6,7,9	10
9	76.0	4.21, m	-	12,16	11,16	-
10	-	-	56.7	-	-	-
11	40.8	2.47, s	-	-	7,9	10
12	75.7	3.29, m	-	9,16	9,16	-
13	-	-	59.8	-	-	-
14	41.9	2.57, s	-	-	12,15	13
15	70.1	5.01, s	-	18	12,14,16,17,18	20
16	33.8	3.29,3.16, m	-	9,12	14,15	13
17	144.5	-	-	-	-	-
18	136.4	7.89, s	-	15	17,19	20

19	162.9	-	-	-	-	-
20	-	-	239.9	-	-	-
21	112.4	-	-	-	-	-
22	158.5	-	-	-	-	-
23	118.2	7.00, d (8.38)	-	24	21,22,25	19,20
24	133.7	7.36, m	-	23,25	22,26	19,20
25	120.7	6.95, m	-	24,26	21,22,23,24,26	-
26	127.2	7.82, d (7.80)	-	25	19,22,24	-

.

Sorangibactin A2

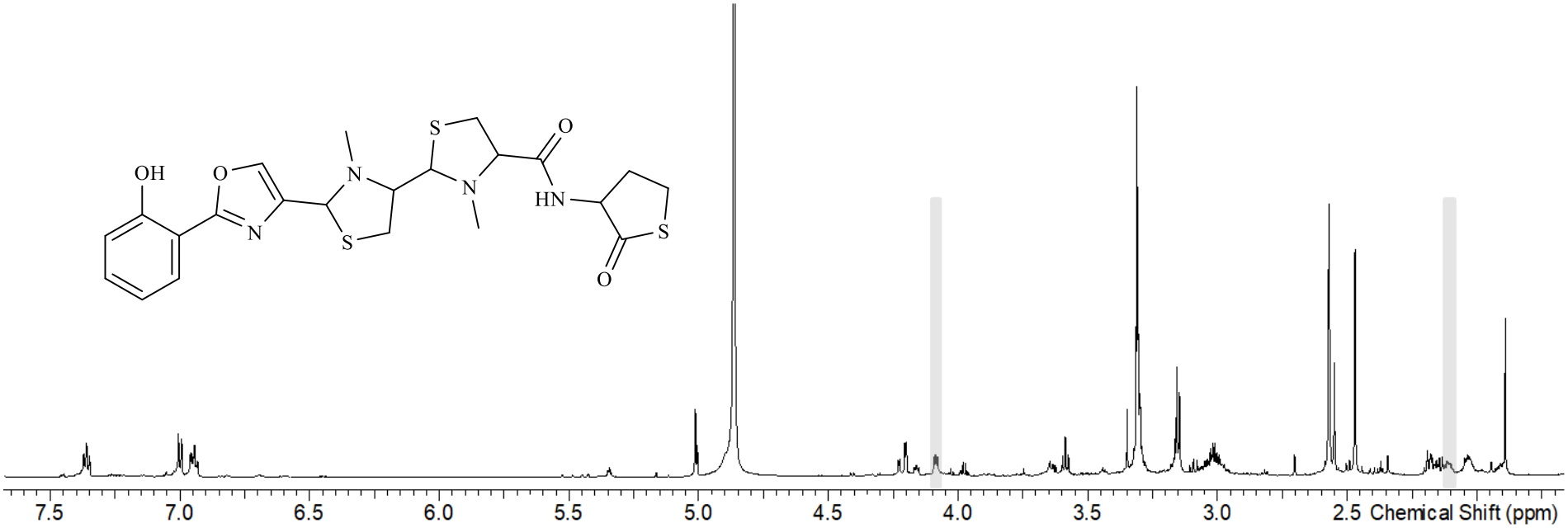


Figure S16: ^{1}H-spectrum of sorangibactin A2 in methanol-d_4 at 700 MHz, signals corresponding to sorangibactin A1 highlighted in grey.

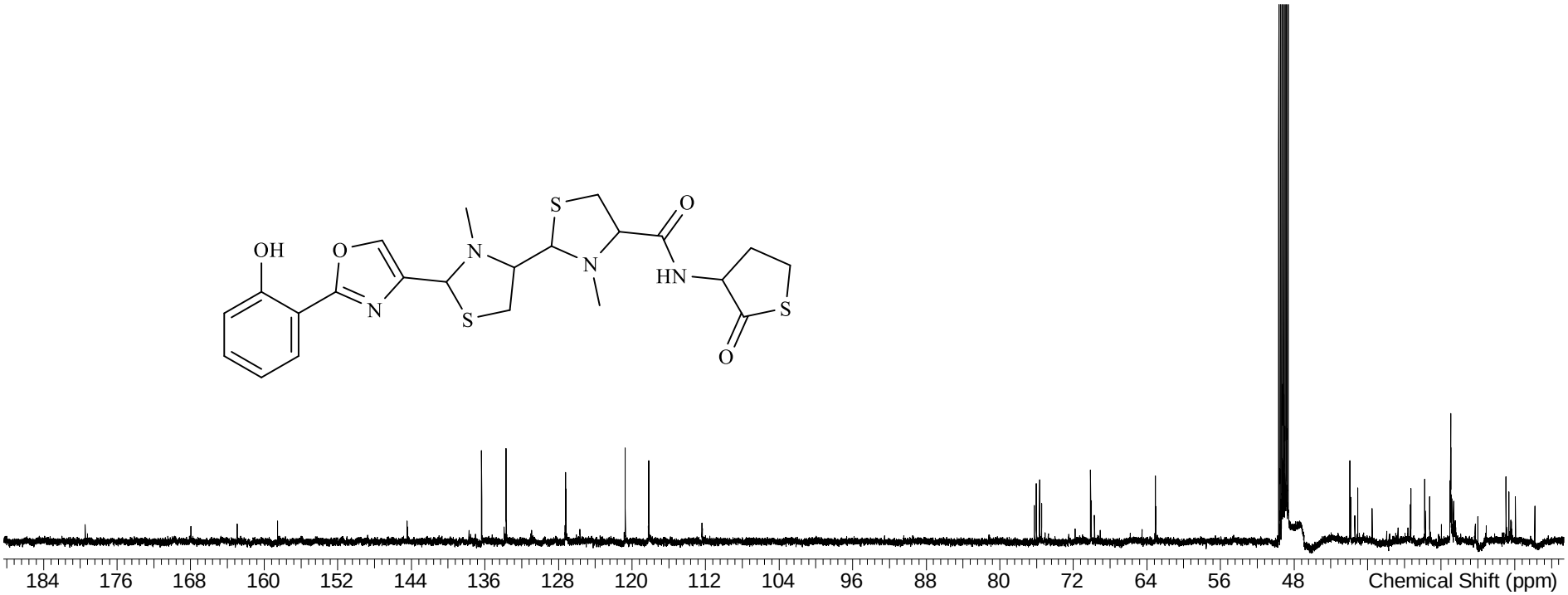
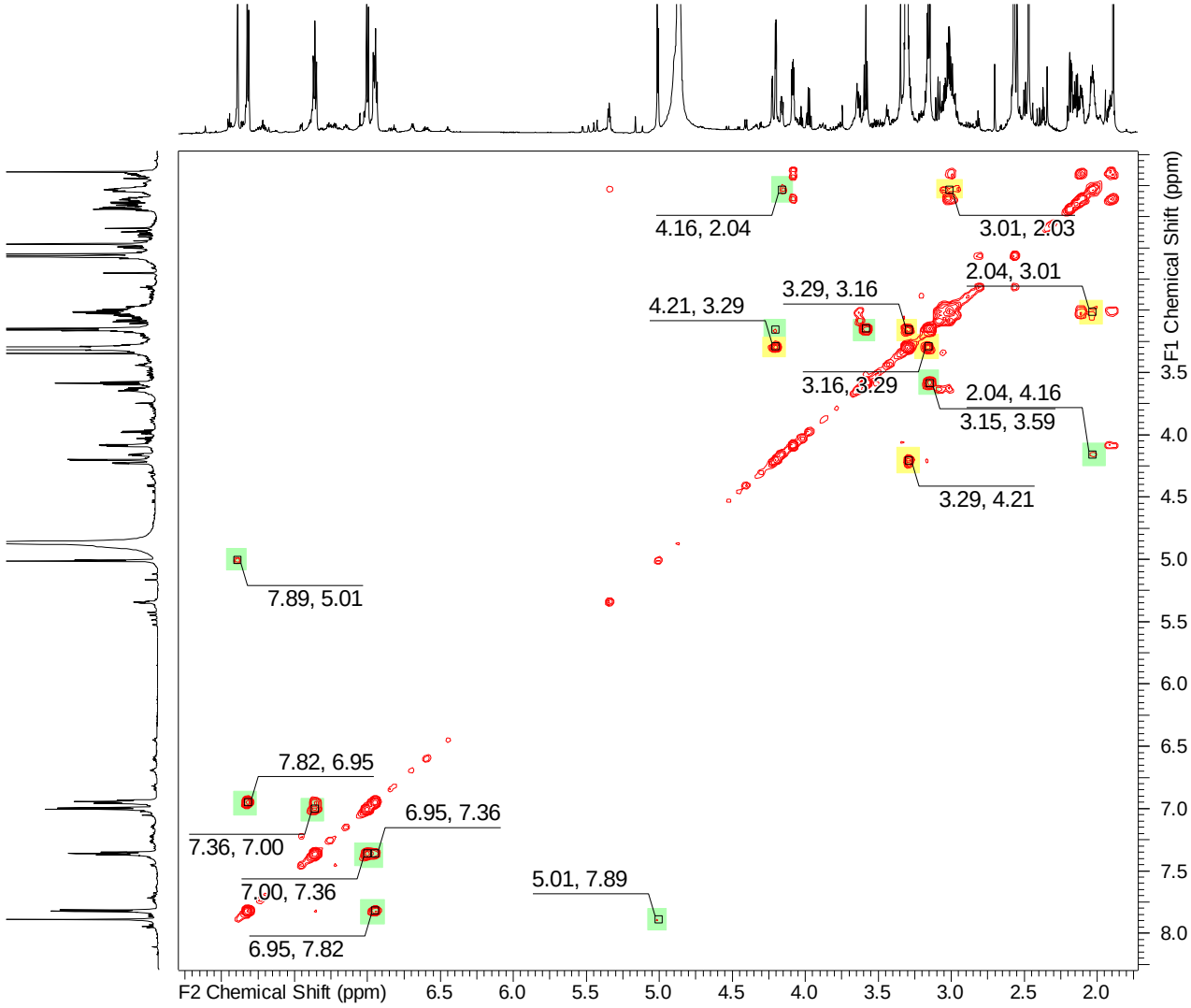


Figure S17: ^{13}C-spectrum of sorangibactin A2 in methanol-d_4 at 175 MHz.



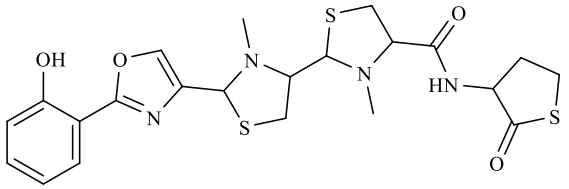
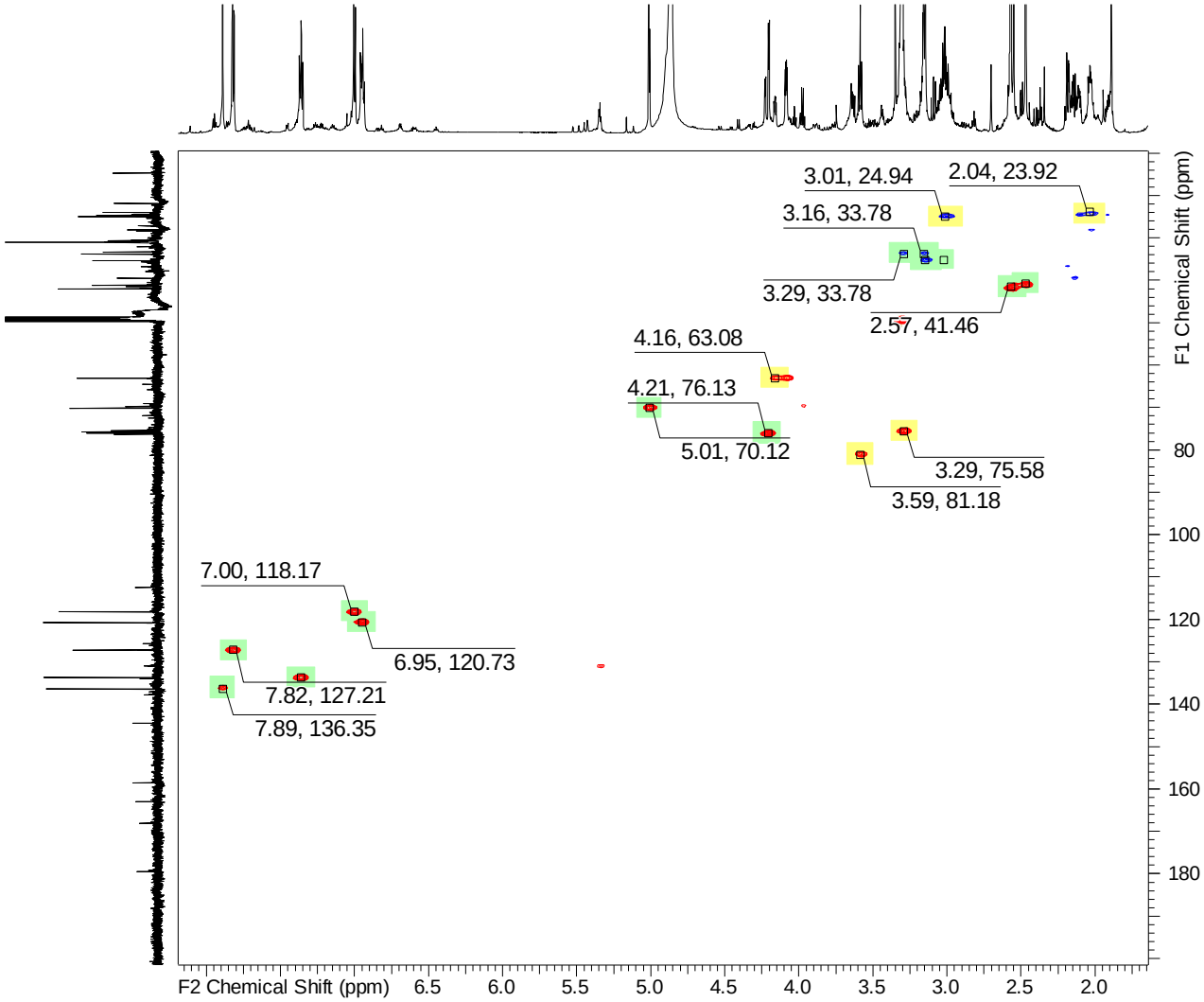


Figure S18: COSY spectrum of sorangibactin A2 in methanol-d_4 at 700 MHz.



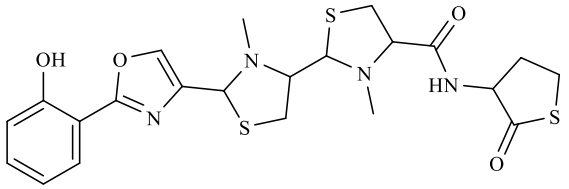
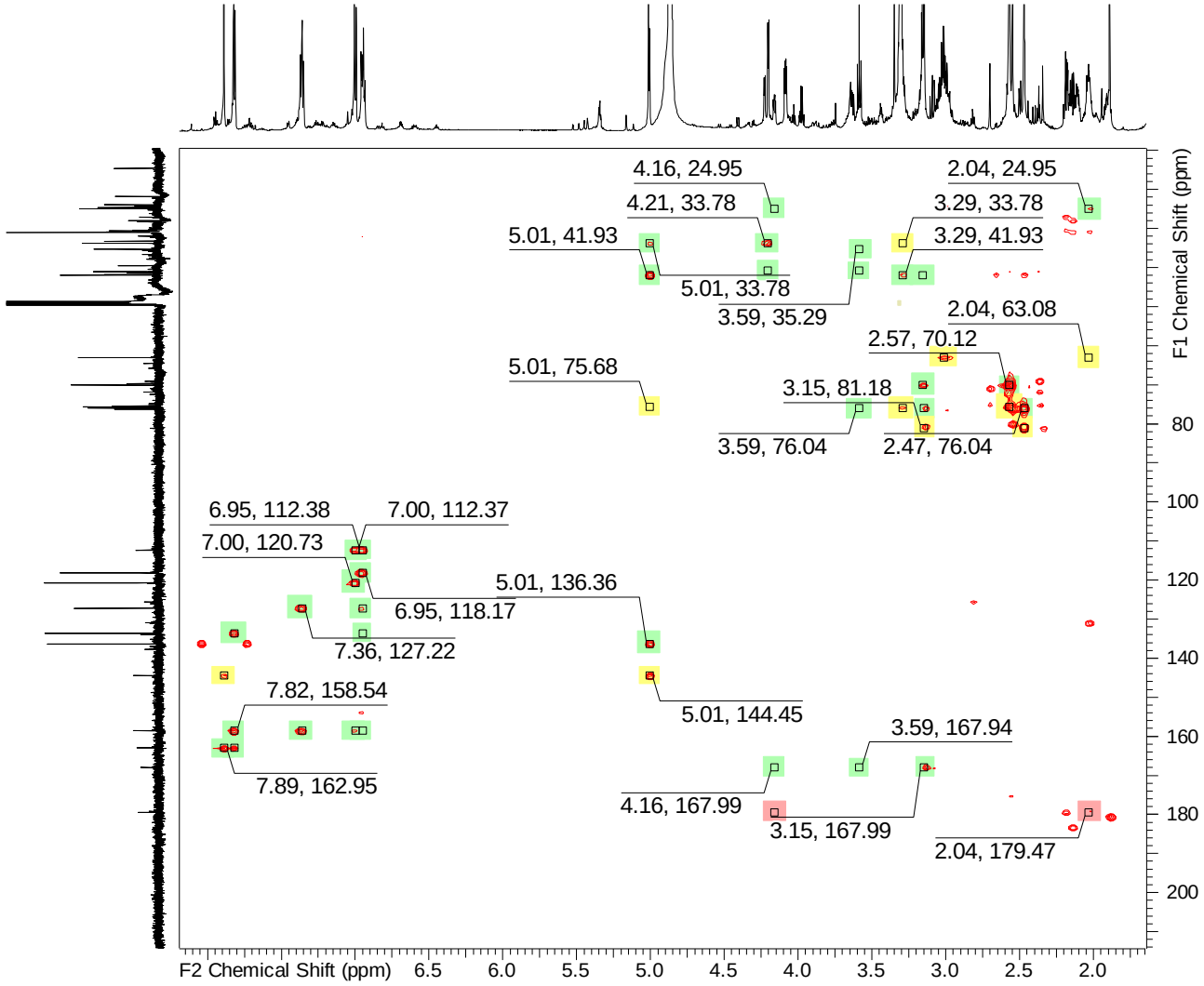


Figure S19: HSQC spectrum of sorangibactin A2 in methanol-d_4 at 700 MHz (^{1}H)/175 MHz (^{13}C).



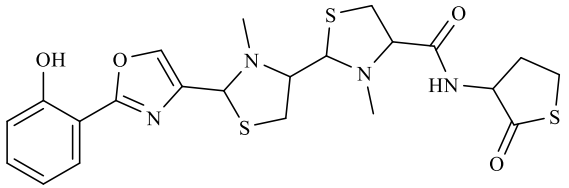
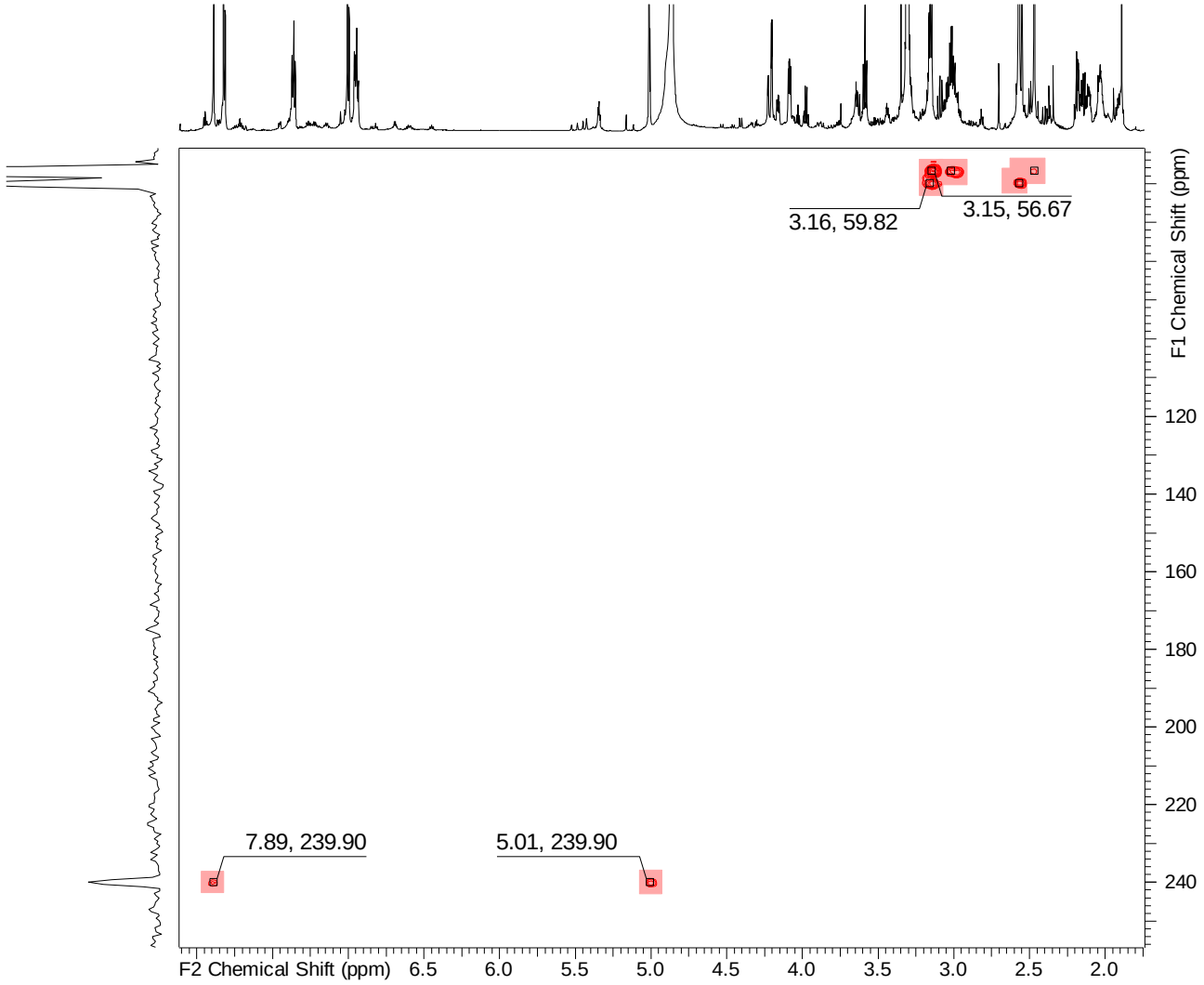


Figure S20: HMBC spectrum of sorangibactin A2 in methanol-d_4 at 700 MHz (^{1}H)/175 MHz (^{13}C).



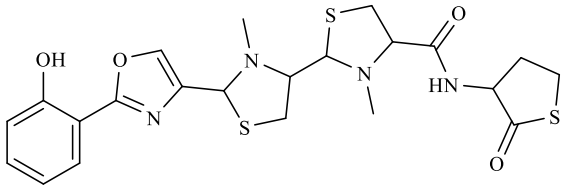


Figure S21: ^{1}H-^{15}N-HMBC spectrum of sorangibactin A2 in methanol-d_4 at 500 MHz (^{1}H)/125 MHz (^{15}N).

Marfey's derivatization experiment for sorangibactin A1 and A2 mixture

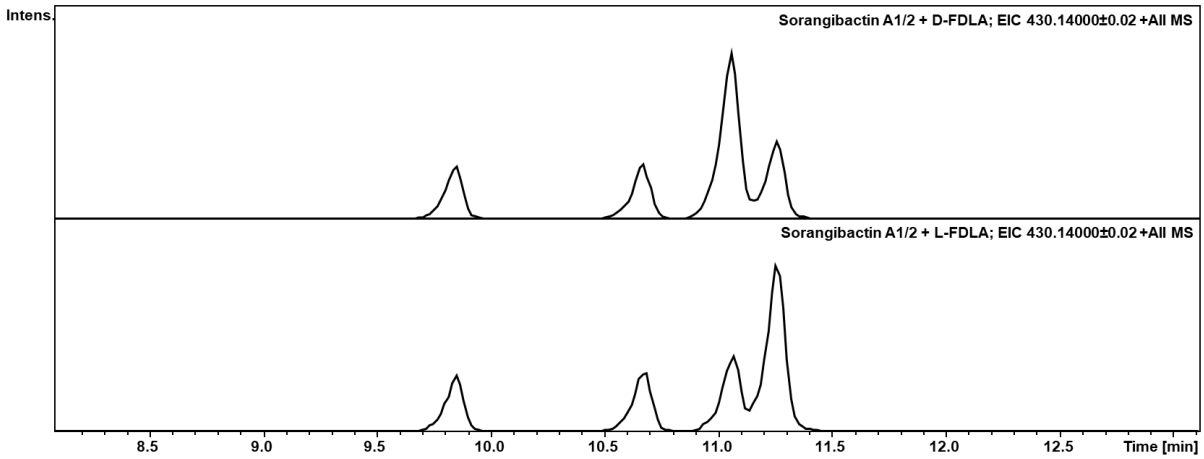
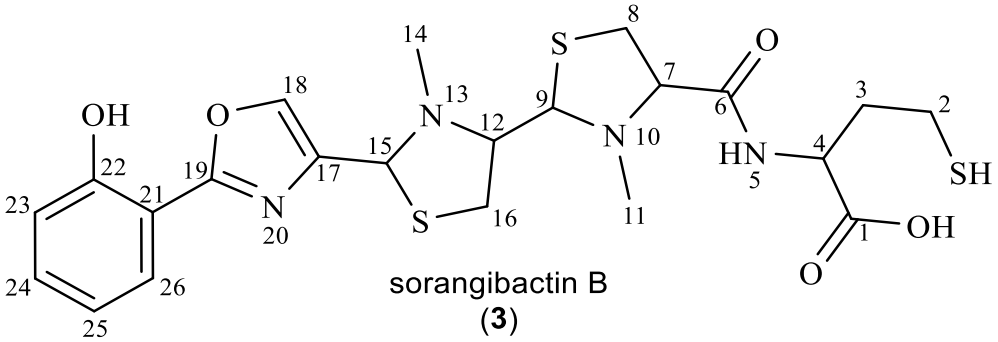


Figure S22: Marfey's derivatization of sorangibactin A1 and A2 mixture with both D-FDLA (upper chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EIC m/z *430.14±.0.02 for derivatized N-methyl cysteine and homocysteine.*

Table S7: NMR spectroscopic data of sorangibactin B in methanol-d_4 at 500/125 MHz.



#	Δ ^{13}C [PPM]	Δ ^{1}H [PPM], MULT (*J* [HZ])	COSY	HMBC
1	176.6	-	-	-
2	35.2	2.27, t (7.47)	3	3,4
3	26.3	1.59, m	2,4	1
4	51.0	4.68, m	3	6
5	-	-	-	-
6	173.0	-	-	-
7	76.3	3.57, m	8	6
8	33.7	3.25,3.07, m	7	6,7
9	78.1	4.31, m	12	12,14
10	-	-	-	-
11	42.5	2.56, m	-	7,9
12	75.2	3.37, m	9,16	-
13	-	-	-	-
14	42.5	2.64, m	-	12,15
15	69.9	5.06, m	18	17,18
16	34.6	3.23,3.08, m	12	12
17	142.5	-	-	-
18	136.4	7.88, m	15	17,19
19	161.8	-	-	-
20	-	-	-	-
21	110.9	-	-	-
22	157.2	-	-	-
23	118.2	7.01, m	24	21,22

24	133.9	7.38, m	23,25	21,22
25	120.8	6.97, m	24,26	21,23
26	127.2	7.83, m	25	19,22,24

.

Sorangibactin B

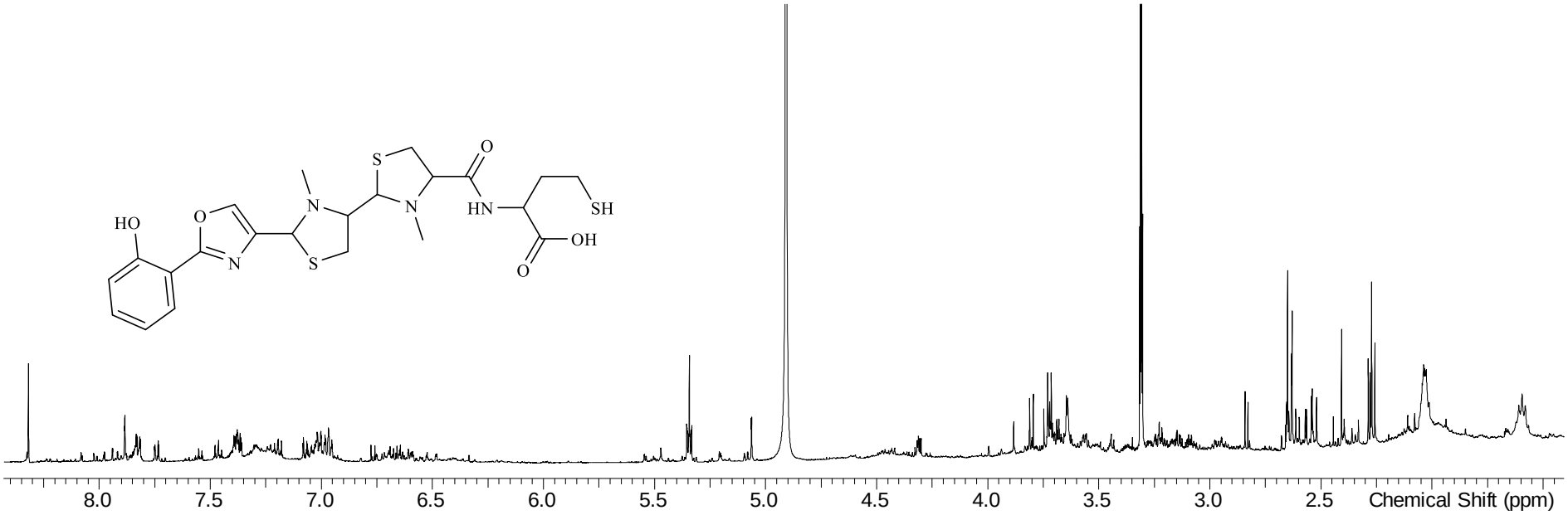


Figure S23: ^{1}H-spectrum of sorangibactin B in methanol-d_4 at 500 MHz.

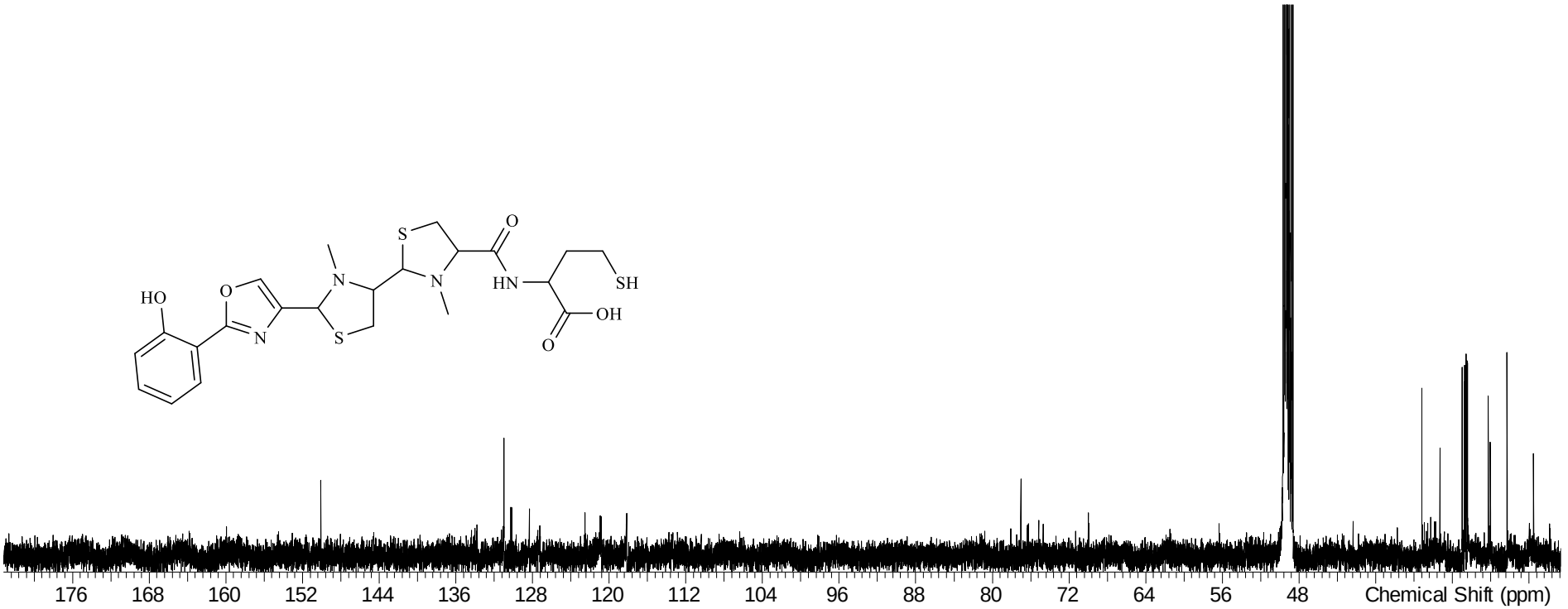


Figure S24: ^{13}C-spectrum of sorangibactin B in methanol-d_4 at 125 MHz.

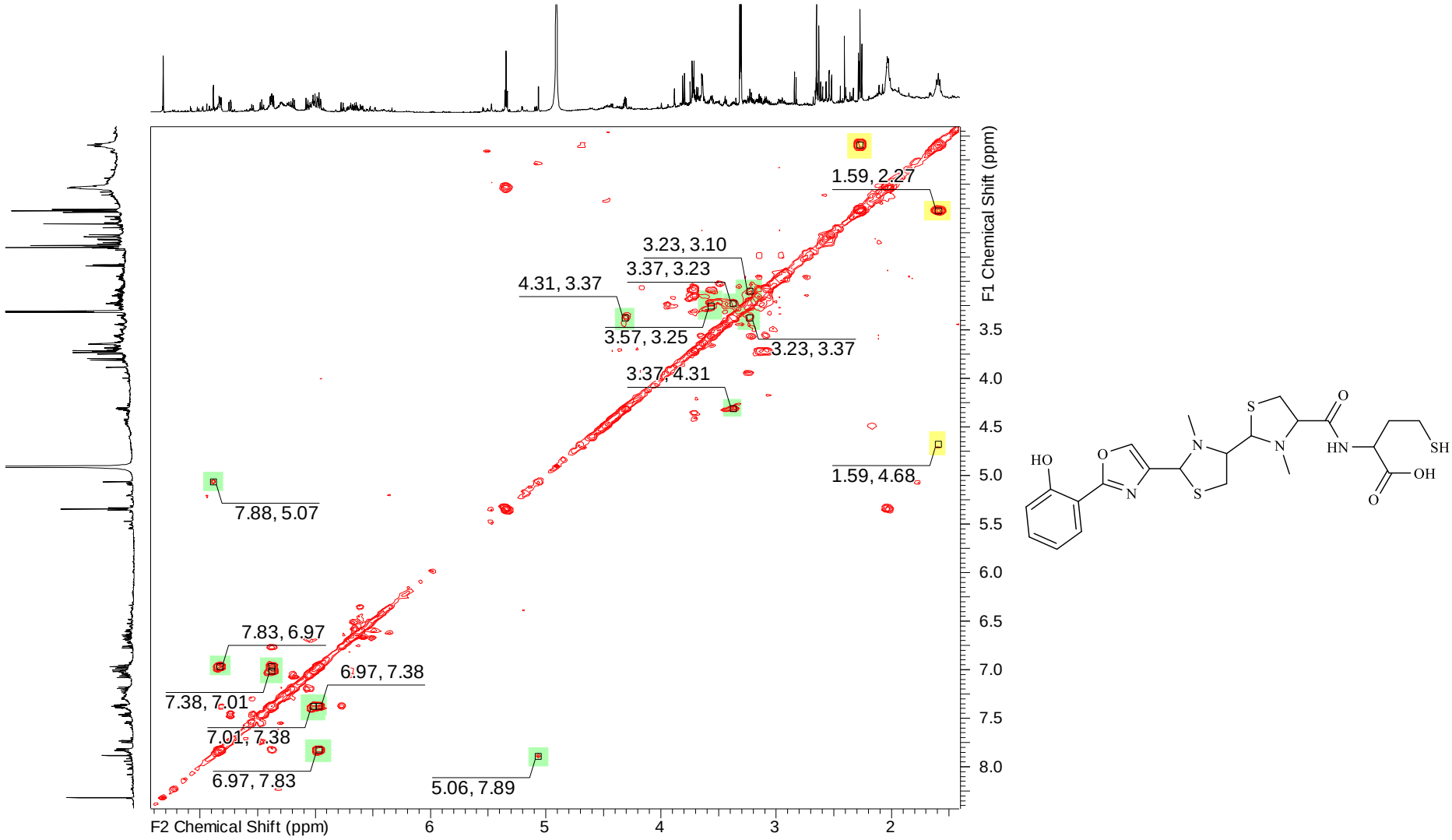


Figure S25: COSY spectrum of sorangibactin B in methanol-d_4 at 500 MHz.

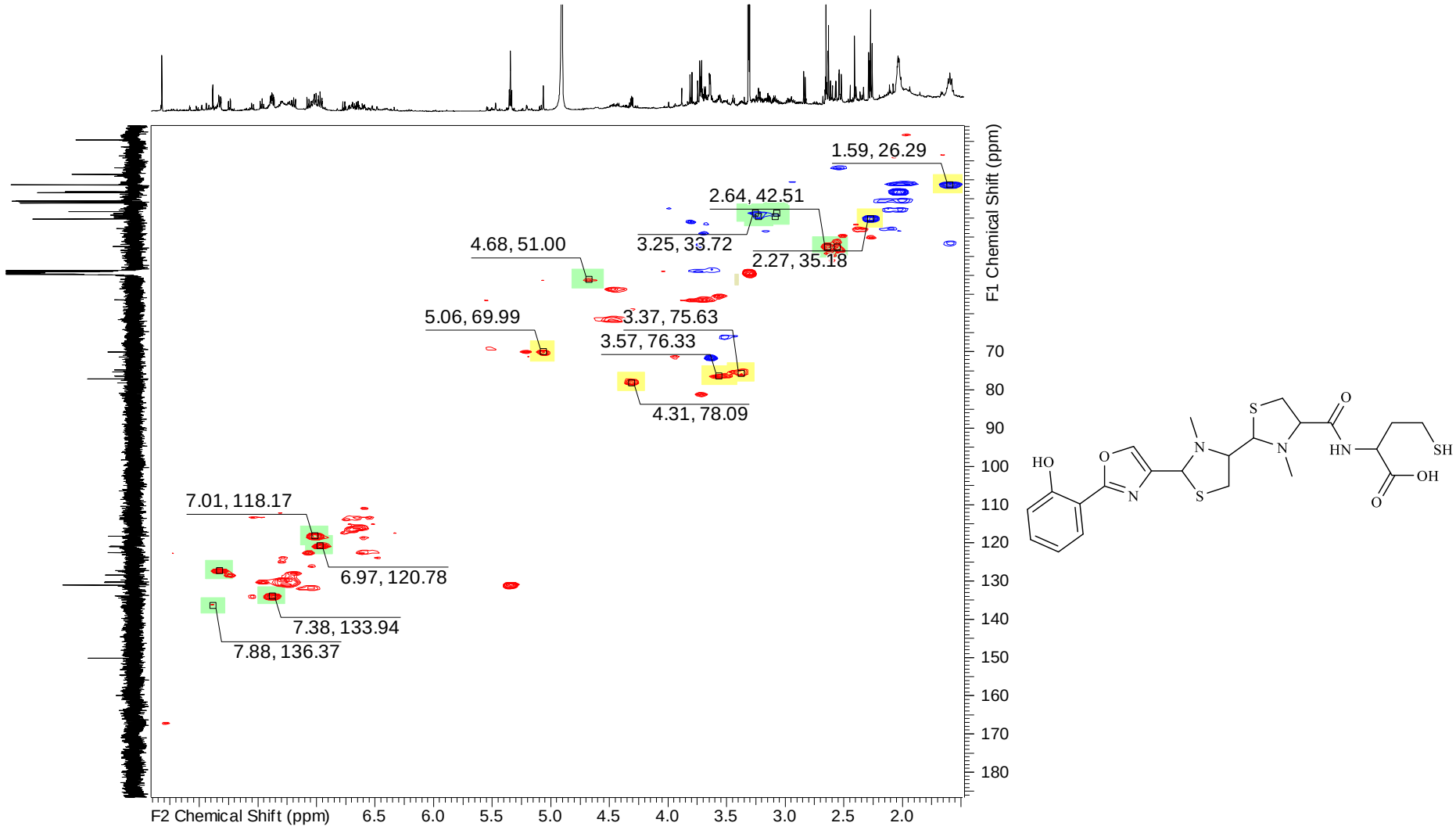
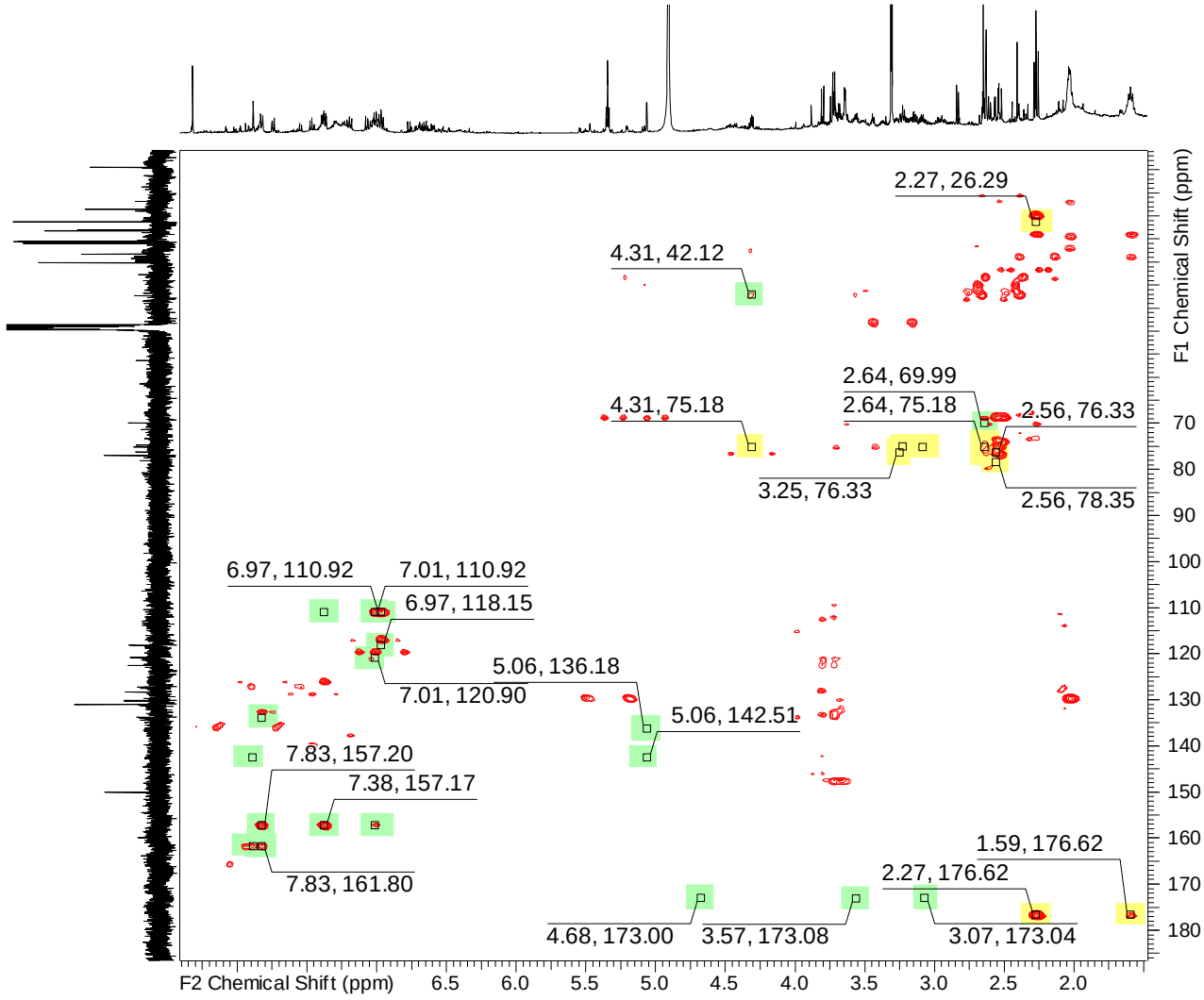


Figure S26: HSQC spectrum of sorangibactin B in methanol-d_4 at 500 MHz (^{1}H)/125 MHz (^{13}C); unpicked surrounding signals are most likely resulting from the presence of many different isomers in the same sample upon the acidic purification procedure required for the ring opened target compound (sorangibactin B).



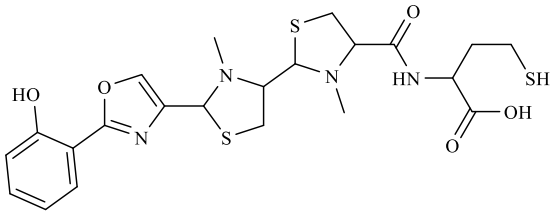


Figure S27: HMBC spectrum of sorangibactin B in methanol-d_4 at 500 MHz (^{1}H)/125 MHz (^{13}C).

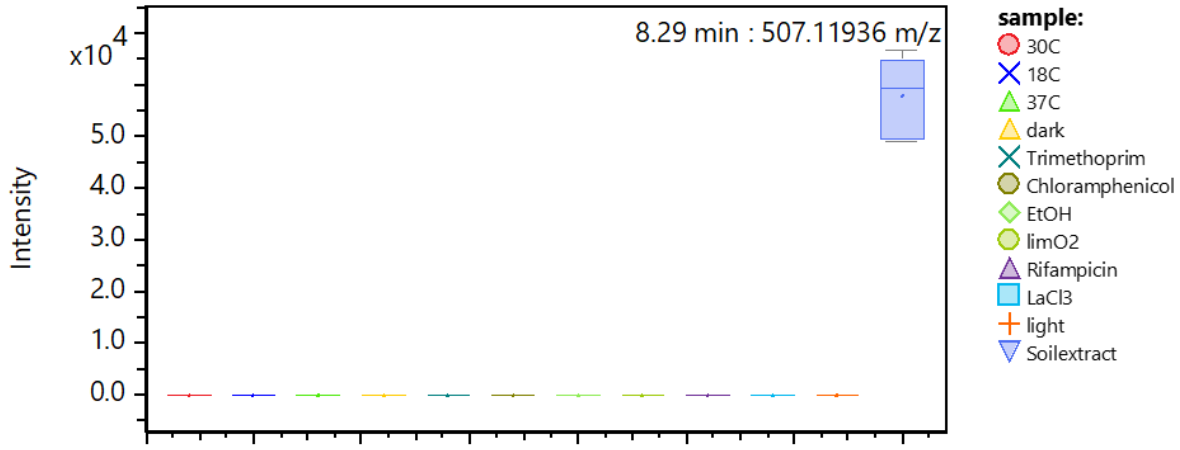


Figure S28: Box-and-whisker diagram showing intensity levels of m/z *507.11936 among the different cultivation conditions: sorangibactin A1 and A2 are only produced during cultivation with addition of soil extract (light purple); median and mean values and 25% and 75% quartiles illustrated within box with grey whiskers representing lower and upper extremes; 2-SWT medium (Chapter 2).*

S3.5 References

1. Song, C.; Luan, J.; Li, R.; Jiang, C.; Hou, Y.; Cui, Q.; Cui, T.; Tan, L.; Ma, Z.; Tang, Y.-J.; et al. RedEx: a method for seamless DNA insertion and deletion in large multimodular polyketide synthase gene clusters. *Nucleic Acids Res* **2020**, *48*, e130, doi:10.1093/nar/gkaa956.
2. Perlova, O.; Fu, J.; Kuhlmann, S.; Krug, D.; Stewart, F.; Zhang, Y.; Müller, R. Reconstitution of myxothiazol biosynthetic gene cluster by Red/ET recombination and heterologous expression in *Myxococcus xanthus*. *Appl. Environ. Microbiol.* **2006**, *72*, 7485–7494, doi:10.1128/AEM.01503-06.
3. Magrini, V.; Creighton, C.; Youderian, P. Site-specific recombination of temperate Myxococcus xanthus phage Mx8: genetic elements required for integration. *J. Bacteriol.* **1999**, *181*, 4050–4061, doi:10.1128/JB.181.13.4050-4061.1999.
4. Mirdita, M.; Schütze, K.; Moriwaki, Y.; Heo, L.; Ovchinnikov, S.; Steinegger, M. ColabFold: making protein folding accessible to all. *Nat Methods* **2022**, *19*, 679–682, doi:10.1038/s41592-022-01488-1.

5. Jang, J.-P.; Kwon, M.C.; Nogawa, T.; Takahashi, S.; Osada, H.; Ahn, J.S.; Ko, S.-K.; Jang, J.-H. Thiolactomide: A New Homocysteine Thiolactone Derivative from Streptomyces sp. with Neuroprotective Activity. *Journal of Microbiology and Biotechnology* **2021**, *31*, 1667–1671, doi:10.4014/jmb.2108.08015.

Chapter 4

4. Unusual Biosynthesis of Alkylprolines and Homo-Amino Acids in Mycoplanecins: Extremely Potent Anti-Tuberculosis Antibiotics Targeting DnaN

Chengzhang Fu[1,2], Yunkun Liu[1], Christine Walt[1,4], Chantal Bader[1,4], Sari Rasheed[1], Peer Lukat[3], , Markus Neuber[1,4], Olga Kalinina[1,5,6], Rolf Müller[1,2,4,*]

[1] Helmholtz Institute for Pharmaceutical Research Saarland (HIPS), Helmholtz Centre for Infection Research (HZI), and Department of Pharmacy, Saarland University, 66123 Saarbrücken, Germany

[2] Helmholtz International Lab for Anti-Infectives, Helmholtz Center for Infection Research, 38124 Braunschweig, Germany

[3] Structure and Function of Proteins, Helmholtz Centre for Infection Research, Inhoffenstr. 7, 38124, Braunschweig, Germany.

[4] German Centre for Infection Research (DZIF), 38124 Braunschweig, Germany

[5] Medical Faculty, Saarland University, 66421 Homburg, Germany

[6] Center for Bioinformatics, Saarland Informatics Campus, 66123 Saarbrücken, Germany

* Corresponding author: Rolf Müller

Email: rolf.mueller@helmholtz-hips.de

Keywords

Mycoplanecin, DnaN, tuberculosis, alkylproline, homo-amino acid biosynthesis

Contributions and acknowledgements

Author's effort:

The author significantly contributed to the performance of experiments and subsequent interpretation of the results. Compound purification, structure elucidation, Marfey's analysis and MST experiments were carried out by the author. Moreover, the author was involved in conceiving and writing of this manuscript.

Contributions by others:

Yunkun Liu contributed significantly to the bioinformatics analysis of the biosynthetic pathway including respective laboratory experiments. Chantal Bader contributed to the compound purification and structure elucidation including Marfey's analysis. Markus Neuber did the cultivation and Peer Lukat performed crystallographic experiments. Sari Rasheed performed bioactivity assays. Chengzhang Fu significantly contributed to the writing of this manuscript. Chantal Bader and Jake Haeckl contributed via proofreading the manuscript. Rolf Müller contributed to conception and supervision of this study with subsequent editing and proofreading of the manuscript.

Abstract

DNA polymerase III sliding clamp (DnaN) was recently validated as a new anti-tuberculosis target employing griselimycins. Three (2*S*,4*R*)-4-methylproline moieties of methylgriselimycin play significant roles in target binding and metabolic stability. Here, we identified the mycoplanecin biosynthetic gene cluster by genome mining using bait genes from the 4-methylproline pathway. We isolated and structurally elucidated four mycoplanecins comprising scarce homo-amino acids and 4-alkylprolines. Evaluating mycoplanecin E against *Mycobacterium tuberculosis* surprisingly reveals an excitingly low minimum inhibition concentration at 83 ng/mL, thus outcompeting griselimycin by approximately 24-fold. We show that mycoplanecins bind DnaN with nanomolar affinity and provide a co-crystal structure of mycoplanecin A-bound DnaN. Additionally, we reconstitute the biosyntheses of the unusual L-homoleucine, L-homonorleucine, and (2*S*,4*R*)-4-ethylproline building blocks by characterizing *in vitro* the full set of eight enzymes involved. The biosynthesis study, bioactivity evaluation, and drug target validation of mycoplanecins pave the way for their further development to tackle multidrug-resistant mycobacterial infections.

4.1 Introduction

Tuberculosis (TB) is a severe infectious disease mainly caused by the pathogenic bacterium *Mycobacterium tuberculosis* (*Mtb*). TB remains an enormous global health burden, causing an estimated 1.5 million deaths and 10 million new cases in 2020. TB was the leading cause of death in all infectious diseases until the coronavirus (COVID-19) pandemic. In addition, the growing percentage of multidrug-resistant (MDR) TB is a significant threat and a daunting obstacle to global TB treatment and prevention efforts. Since MDR TB is resistant to at least the two most potent TB drugs, isoniazid, and rifampin, patients need new treatment options [1,2]. Consequently, new drugs addressing novel targets in *Mtb* are immensely desired.

Griselimycins (GMs) are such promising anti-TB antibiotics that bind and inhibit the mycobacterial DNA polymerase III sliding clamp (DnaN). This unique target makes GM distinct from other antibiotics to evade common TB drug resistance [3]. It is worth noting that the unusual nonproteogenic amino acid (2*S*,4*R*)-4-methylproline ((2*S*,4*R*)-4-MePro) acts as an essential building block occurring twice in GM and thrice in methylgriselimycin (MGM). Oxidation of the metabolically unstable Pro^8 residue initiates GM degradation. The significantly improved metabolic stability of GM derivatives with substituent groups at Pro^8, such as the natural MGM with (2*S*,4*R*)-4-MePro, has corroborated this finding [3]. More importantly, GM congeners with different 4-alkyl groups show improved activity against *Mtb*, which might be attributed to their enhanced lipophilicity [3]. The dioxygenase GriE and the dehydrogenase GriF are essential for (2*S*,4*R*)-4-MePro biosynthesis in the GM pathway (Figure 1a) [4]. We previously reasoned that the genes required for synthesizing the (2*S*,4*R*)-4-MePro building block could be utilized to explore compounds containing such unusual residue [4]. An earlier study successfully discovered new natural products containing the diastereoisomer (2*S*,4*S*)-4-MePro, but only from cyanobacteria using different genes for (2*S*,4*S*)-4-MePro biosynthesis in the nostopeptolide pathway as genome mining bait markers [5].

Considering that the GM producer *Streptomyces* sp. DSM40835 [6,7] belongs to the order *Actinomycetales*, the most significant industrial source of natural product drugs [8,9], we addressed the question of whether novel (2*S*,4*R*)-4-MePro-containing compounds from

actinobacteria with improved bioactivity and physicochemical properties could be discovered via genome mining using the gene pair *griE* and *griF* as the probe. We discovered the previously unkown mycoplanecin (MP) biosynthetic gene cluster (BGC) and isolated unusual MP derivatives that indeed exhibited significantly improved pharmaceutical properties compared to GM. The co-crystal structure and binding assay unambiguously showed that MPs bind to DnaN in the nanomolar range. In addition, we comprehensively characterized *in vitro* three interwoven and uncommon biosynthetic pathways to the non-natural amino acids building blocks of MPs: L-homoleucine, L-homonorleucine, and (2*S*,4*R*)-4-ethylproline ((2*S*,4*R*)-4-EtPro).

4.2 Results

Genome mining for pathways with *griEF* genes identifies the mycoplanecin biosynthetic pathway

Here, a comprehensive genome mining of the co-existence of *griE* and *griF* in all actinobacteria genomic sequences available in the NCBI nuccore and assembly revealed dozens of cryptic BGCs (Figure 1b and Supplementary Table 1). This mining strategy was effective as it found the BGCs known to produce compounds containing 4-MePro, such as GM [4] and acyldepsipeptide (ADEP) antibiotics [10] (Figure 1c). One hit similar to the GM pathway drew our attention, because it also exhibited significant differences. However, only a partial BGC was recovered due to the low genome sequence quality. This short contig containing *griEF* homologous genes belongs to the genome of *Actinoplanes awajinensis* subsp. *mycoplanecinus* NRRL B-16712 [11]. Additional genes that are absent in the GM pathway were identified, including genes encoding a radical *S*-adenosylmethionine (SAM) protein, an acyl carrier protein (ACP), and a 3-ketoacyl-ACP synthase (KAS) III homolog (Figure 1d). Fortuitously, we found that this NRRL strain is equivalent to *A. awajinensis* subsp. *mycoplanecinus* subsp. nov. ATCC33919 (strain No. 41042) producing anti-mycobacterial MP peptides discovered in the 1980s [12–16]. Notwithstanding the structural similarity to GM, MPs possess unique characteristics, including the exceptional N-terminal α-keto butyric moiety found instead of the

acetyl group in GM, as well as the other two 4-alkylprolines: 4-EtPro and 4-propylproline (4-PrPro), homoleucine (5-methyl-norleucine), and N-methyl-homonorleucine (N-methylheptanoic acid) (Figure 2). Because the published draft genome is highly fragmented, the complete BGC could not be recovered without further genome sequencing effort.

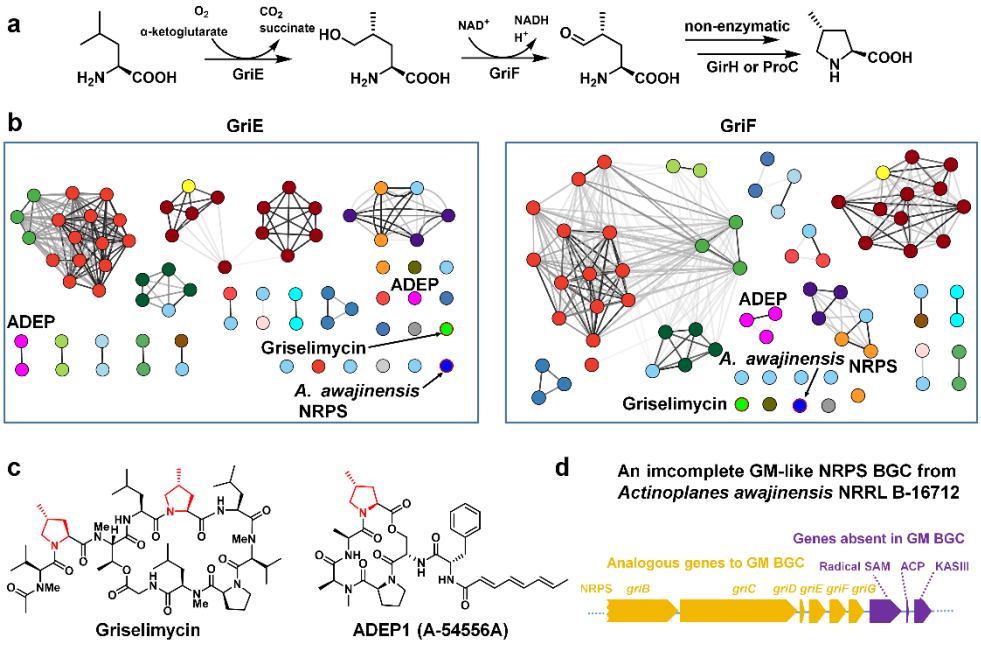


Figure 1: Discovery of mycoplanecin biosynthetic gene cluster (BGC) by sequence similarity network (SSN) analysis; ***a****) GriE and GriF are essential for biosynthesis of (2S,4R)-4-methyl-proline in the griselimycin (GM) pathway,* ***b****) Sequence similarity networks for GriE (left) and GriF (right). All edges corresponding to sequence identity below 80 % were removed. The color of the edge represents the sequence similarity, with darker color corresponding to higher similarity. Node colors correspond to the type of the biosynthetic gene cluster (Supplemetary Table 1);* ***c****) Chemical structures of GM and ADEP1 which contain (2S,4R)-4-methyl-proline;* ***d****) The incomplete NRPS BGC from Actinoplanes awajinensis NRRL B-16712 shows similarity to GM BGC but also contains several different genes.*

Characterization of the MP BGC possessing a self-resistance gene

The intact 59 kb BGC of MP was obtained by searching for the *griEF* hit sequence in the complete genome of strain ATCC33919 that we generated by PacBio sequencing (Figure 2a). The putative MP BGC comprises 15 genes divided into three hypothetical operons that encode nonribosomal peptide synthetases (NRPSs), precursor biosynthetic enzymes, and a self-resistance protein (Figure 2a, Supplementary Table 2). We found a DnaN-encoding gene *mypR* nearby transposase genes, which was also observed for *griR* in the GM pathway [4]. Furthermore, *mypR* is found to be the second DnaN gene in the genome of strain ATCC33919, implying that MPs are potential DnaN inhibitors, and *mypR* confers the self-resistance against MPs [3]. The majority of genes are clustered in one operon, which encodes NRPSs MypABC, the MbtH protein MypD, redox enzymes MypP (P450), MypE (hydroxylase), and MypF

(dehydrogenase), the thioesterase MypG, the discrete ACP MypI, and the ketosynthase MypJ, as well as a radical SAM enzyme MypH. The three gene operon comprising *mypKLM* is located upstream adjacent to the primary operon and oriented in the opposite direction (Figure 2a). MypK and MypL show sequence similarity to enzymes involved in the branched-chain amino acid (BCAA) biosynthesis. MypM is a SAM synthetase probably involved in regenerating the cofactor biosynthesis for MypH.

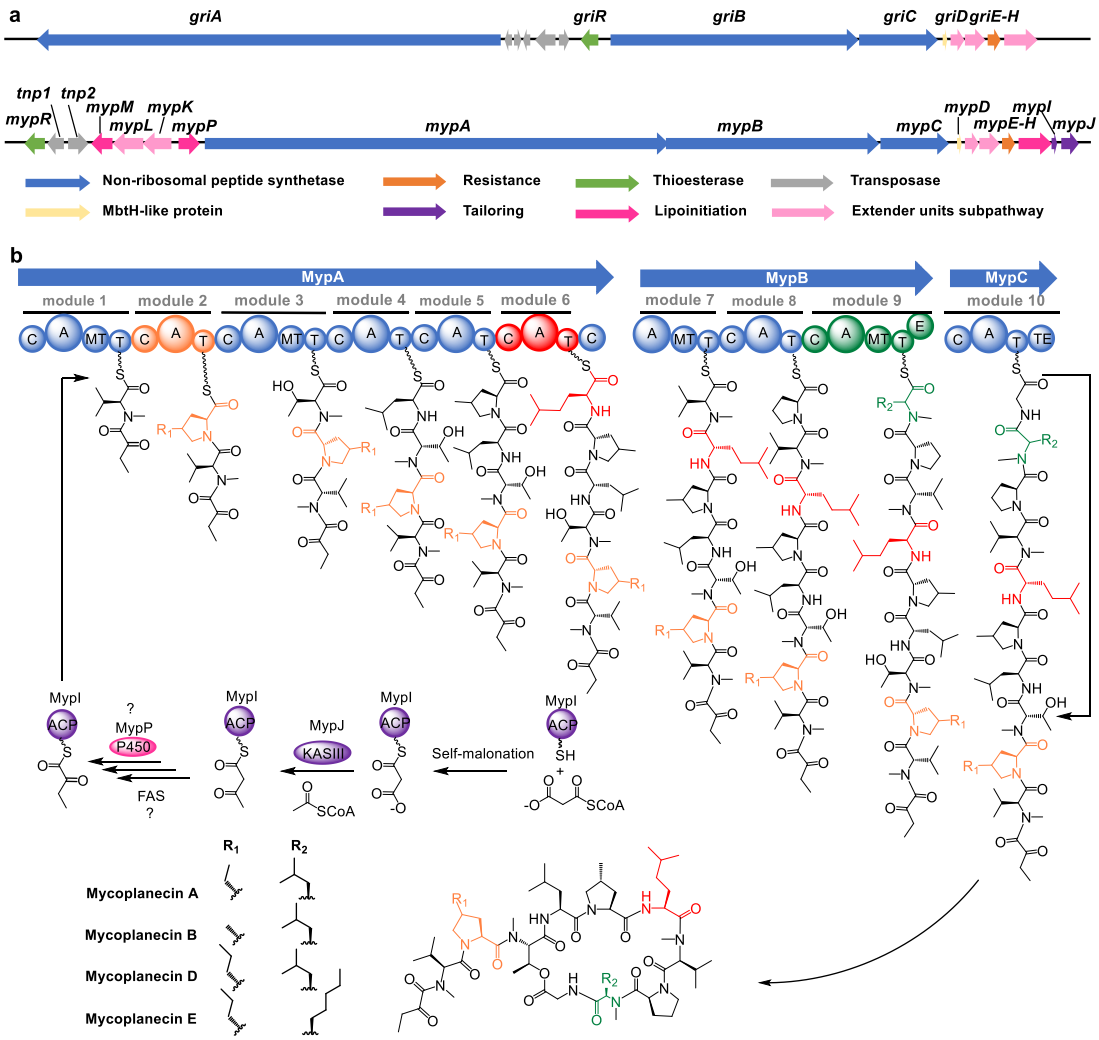


Figure 2: MP biosynthetic gene cluster and proposed biosynthetic pathway: ***a****) The gene organizations of GM and MP biosynthetic gene clusters;* ***b****) The MP biosynthetic pathway proposal. The A domains and corresponding atypical building blocks are shown in orange, red or olive green, respectively. For each module, the domains are : C = condensation, A = adenylation, MT = methyltransferase, T = thiolation domain, also known as peptidyl carrier protein domain in NRPS, E = epimerase, TE = thioesterase.*

Three giant NRPS genes *mypABC* are proposed to build the decapeptide backbone of MPs. However, due to their highly repetitive sequences, the frequently experienced assembly dilemma with large modular NRPS genes occurred. Two independent PacBio sequencing attempts resulted in two versions of *mypA* with different lengths, 22,866 bp and 26,007 bp (with one additional NRPS module). Comprehensive PCR amplification and restriction digestion analysis of the repetitive region confirmed the shorter version, which fits the common

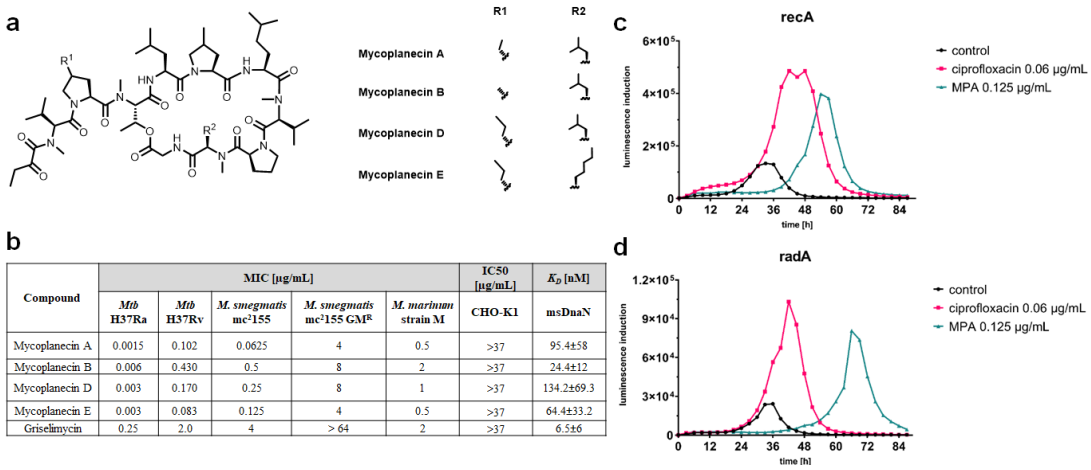
colinearity rule for NRPS (Supplementary discussion, Extended Data Figure 1, Supplementary Figures S1 and S2).

MypA, MypB, and MypC comprise six, three, and one single NRPS module, respectively. The substrate specificity prediction of adenylation (A) domains correctly suggested the incorporation of Thr3, Leu4, Pro8, Leu9, and Gly10 (Figure 2b). The four N-methyltransferase (MT) domains align with the N-methylation of Val1, Thr3, Val7, and Leu9 in MPs. However, modules 1 and 7 were inappropriately predicted to activate Thr instead of Val. Intriguingly, several nonproteogenic amino acid building blocks were found in the biosynthesis of MPs. A6 is predicted to activate ornithine, but L-homoleucine was found incorporated in MPs in this position. The A9 domain has a promiscuous substrate tolerance, since L-N-methyl-homonorleucine and L-N-methyl-leucine are present in different MPs at the same position. The substrate specificity of A2, A5, and A8 was predicted as proline, indicating that the current bioinformatic analysis could not discriminate proline from different alkyl-prolines [4]. The substrate specificity of A2 is flexible as three different 4-alkylproline residues are incorporated in different MP derivatives (Figure 2b). It is noteworthy that uncommon genes responsible for lipopeptide chain initiation are found in the MP pathway, partially similar to the proposed pathway for calcium-dependent antibiotics (CDAs) [17]. Here, it is hypothesized that the holo-MypI (ACP) is either malonylated by the malonyl-CoA-acyl carrier protein transacylase (MCAT) FabD or functions as a self-priming ACP that widely exists in fatty acid biosynthesis [18,19]. Subsequently, the KAS III-like enzyme MypJ extends the acyl chain by a molecule of acetyl-CoA to form the acetoacetyl-*S*-ACP species, which is probably reduced to butyryl-*S*-MypI by reductases from the type II fatty acid biosynthesis. A P450-dependent oxidase encoding gene *mypP* is assumed to be responsible for the cryptic α-keto group formation, albeit the mechanism and oxidation timing are still unexplored (Figure 2b, Supplementary discussion).

Isolation and structure elucidation of MPs with unusual building blocks

Mycoplanecins were first investigated four decades ago due to their activity against mycobacteria while possessing low cytotoxicity, but no subsequent studies are known from the literature. It is worth noting that the absolute configurations of chiral centers in the structures of MPs are not completely clear [12–16]. Moreover, the molecular target of MPs was unknown. Given the presence of the putative self-resistance gene *mypR* and the finding of novel biosynthetic genes, we decided to re-isolate this group of potent anti-tuberculosis antibiotics as they are not commercially available, also aiming at the identification of unprecedented derivatives for subsequent characterization.

The three most abundant derivatives were purified and structurally elucidated as the previously described MP A, B, and D (Figure 3a) [12,13,15]. A complete set of 1D and 2D nuclear magnetic resonance (NMR) experiments was acquired for each derivative to underpin the structures and to supplement previous studies where only 1D spectra of the MPs were shown (Supplementary NMR Data) [12,13,15]. The last derivative isolated, MP E, is structurally novel with a 4-PrPro and an N-methyl-homonorleucine residue incorporated, therefore featuring the highest molecular mass among the different derivatives (Figure 3a). We proposed the stereochemistry of all derivatives through Marfey's analysis and biosynthesis analysis, confirming that MP A comprises N-methyl-valine, 4-EtPro, leucine, 4-MePro, proline, N-methyl-threonine in L-configuration, while N-methyl-leucine is the only D-configured amino acid (Supplementary Figures S12-18, S25-30, S37-43, S48-54). The N-methyl-D-leucine incorporating module 9 of the MP pathway contains the only epimerization (E) domain, further underlining this fact (Figure 2b).



b

Compound	MIC [µg/mL]					IC50 [µg/mL]	K_D [nM]
	Mtb H37Ra	*Mtb* H37Rv	*M. smegmatis* mc²155	*M. smegmatis* mc²155 GM^R	*M. marinum* strain M	CHO-K1	msDnaN
Mycoplanecin A	0.0015	0.102	0.0625	4	0.5	>37	95.4±58
Mycoplanecin B	0.006	0.430	0.5	8	2	>37	24.4±12
Mycoplanecin D	0.003	0.170	0.25	8	1	>37	134.2±69.3
Mycoplanecin E	0.003	0.083	0.125	4	0.5	>37	64.4±33.2
Griselimycin	0.25	2.0	4	> 64	2	>37	6.5±6

Figure 3: Chemical structures and bioactivity of MPs; ***a****) The chemical structures of MPs isolated in this study;* ***b****) The comparison of bioactivity of MPs and GM against mycobacteria and dissociation constant against DnaN from* M. smegmatis mc^{2}155*;* ***c****) The strong induction of SOS response through RecA in* M. smegmatis *treated with 2x MIC MP A;* ***d****) The strong induction of DNA repair through RadA in* M. smegmatis *treated with 2x MIC MP A.*

MPs show low nanomolar anti-Mtb potency

Bioactivity testing of the MPs against the attenuated *M. tuberculosis* strain H37Ra uncovered superior activity in the low nanomolar range minimum inhibitory concentration (MIC) from 1.5 to 6 ng/mL (Figure 3b). Additional MPs testing against *M. smegmatis* revealed potent activity from 0.0625 to 0.5 µg/mL. Although MPs remain relatively effective against GM-resistant *M. smegmatis* strain with MIC values ranging between 4 to 8 µg/mL, there was still a MIC shift (ranging between 16 to 64-fold), which gives initial evidence that mycoplanecins have the same target as GM. It should be noted that resistance against griselimycins is associated with severe fitness loss in the mutants [3].When measured against the virulent *M. tuberculosis* H37Rv, the inhibitory bioactivities of MPs decreased relative to activities seen against *Mtb* H37Ra. However, MPs still exhibit ~24-fold more potent bioactivity than GM against the virulent *Mtb* H37Rv (Figure 3b). Surprisingly, the best MIC was 0.083 µg/mL for the new congener MP E, while the MIC of GM is 2 µg/mL. The remaining panel of tested microbial pathogens, including several Gram-positive species, revealed no activity up to 64 µg/mL, highlighting the high selectivity of this compound class, also concerning members of the human microbiota. In addition, MPs showed no cytotoxicity against CHO-K1 cells up

to a concentration of 37 µg/mL, which underpins the high potential of MPs in the application to combat mycobacterial infections. DNA damage response after exposure to MP A was evaluated in *M. smegmatis* reporter strains harboring mycobacterial promoters that drive the expression of DNA damage-inducible genes, *recA* and *radA*. Strong induction of luminescence was observed for the reporter strains upon treatment with 2x MIC MP A, indicating that the target of MPs is involved in DNA metabolism that results in genotoxic stress. (Figure 3c and 3d).

MPs bind to DnaN in nanomolar affinity

The homology of the self-resistance gene *griR* from the GM BGC with *mypR* in the MP BGC and MPs' structural analogy to GM implied that DnaN is the molecular target of MPs. A microscale thermophoresis (MST) assay was used to determine the binding affinity of all isolated MPs as well as GM to DnaN. The best affinity to *M. smegmatis* DnaN (msDnaN) was observed for GM with K_D=6.5±5.9 nM. Lower DnaN binding affinities were determined for MPs as the K_D-value for MP A was 95.4±58.0 nM, while MP B showed the best binding affinity to msDnaN among the congeners with a K_D=24.4±11.9 nM. The other two MPs showed a similar binding range to msDnaN (Figure 3b, Extended Data Figure 2). Thus, GM manifests a better DnaN binding affinity than the MPs, but is less active than MPs against mycobacteria. This intriguing phenomenon is unresolved, but might be partially due to the MPs' improved pharmaceutical properties and increased lipophilicity.

Crystal structure of DnaN in complex with MP A

Although the co-crystal of msDnaN and MP A could not be obtained after intensive attempts, DnaN of *E.coli* (ecDnaN) was overproduced in *E. coli* and co-crystallized with MP A, GM and cyclohexyl-GM (CGM). The crystal structures were solved at resolutions of 2.1 Å (MP A), 1.7 Å (GM) and 1.5 Å (CGM) and refined to R/R_{free} values of 22/26 %, 17/19 % and 17/18 %, respectively (Supplementary Table S7). The MP A complex crystallized in space group C2 with one dimeric ring in the asymmetric unit, while the GM/CGM complexes crystallized in space group C222 with one protein chain in the asymmetric unit and the dimer formed by crystal symmetry. All ligands were clearly defined in the electron density map (Extended Data Figure 3). Despite the lower affinity to msDnaN, MP A binds in the same mode to ecDnaN as GM or CGM. All three substances form hydrogen bonds to only 2–3 protein residues, but the additional oxygen in the N-terminal α-keto butyrate residue of MP A provides a second hydrogen bond acceptor for Arg365 (Extended Data Figure 4). The compounds bind mainly via Van-der-Waals interactions with interfaces ranging from 550 $Å^2$ for GM to 590 $Å^2$ for CGM to 600 $Å^2$ for MPA. The linear, N-terminal part of the molecules in subsite 2 superimposes almost perfectly, while MP A seems to be tilted slightly (approximately 7-10° compared to GM/CGM) away from the protein in the region around Pro8. This tilting can be attributed to the longer 5-methyl-L-norleucine residue of MP A at position 6, whereas leucine is present at the same position in GM and CGM (Figure 4). The comparison with the previously published mycobacterial GM/DnaN complexes shows that whereas the N-terminal parts of the ligands in subsite 2 superimpose very well, there is higher variance on the opposite side of the molecules around Pro8 (Extended Data Figure 5). In the complex with msDnaN, GM adopts a conformation very similar to that of MP A in the ecDnaN complex, although GM lacks the 5-methyl-L-norleucine at position 6 and the residues lining the protein pocket around this position are conserved between mycobacteria and *E. coli*. In complex with *M. tuberculosis* DnaN, the tilt away from the protein is less pronounced, instead the macrocycle seems to be tilted approx. 7° towards positon 7 in comparison to the msDnaN complex. Although the observed conformational differences of the ligands could be caused by differences in crystal packing of

the DnaN/GM complexes, these findings indicate that the residues of the linear part of the ligands (1-3) and the residues adjacent to the bifurcation (4-5, 10) seem to be bound tighter/show less flexibility than the opposing residues (6-9). This is also supported by the crystallographic B-factors of the ligands being generally the highest for Pro8 and low for the linear part and the adjacent residues (Extended Data Figure 6).

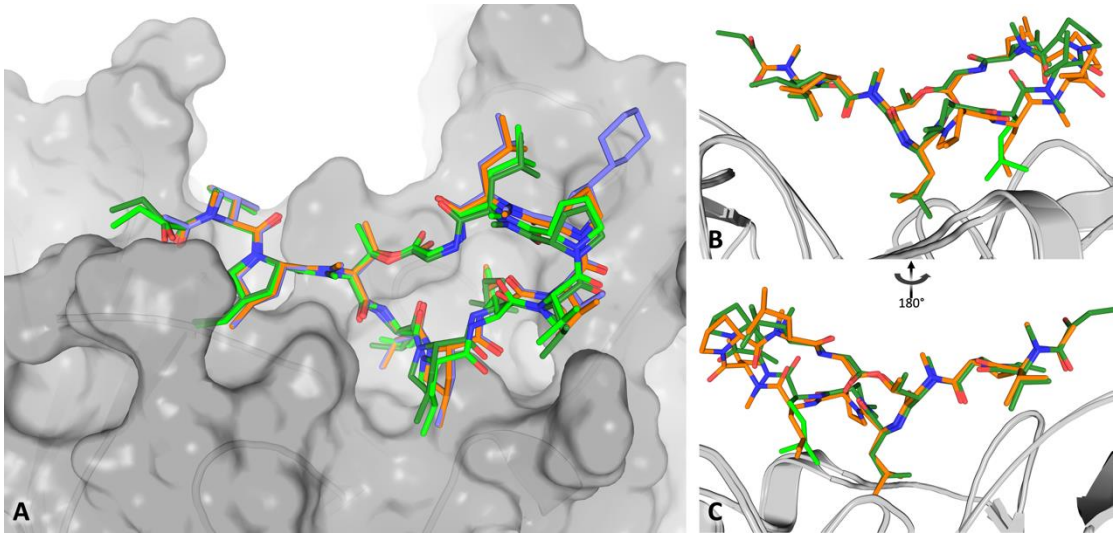


Figure 4: Binding of GM, CGM and MP A to ecDnaN; ***A****) Superimposition of the ecDnaN co-crystal structures containing GM (orange), CGM (violet) and MP A (dark/light green for the two entities in the asymmetric unit). The protein surface is shown in gray;* ***B****) Binding comparison of GM (orange) and MP A (green) to ecDnaN (gray cartoon). The 5-Methyl-norleucine residue at position 6, that pushes MPA away from the protein in comparison with leucin 6 in GM is highlighted in light green;* ***C****) View of the GM/MP A binding comparison turned 180°C around the Y axis.*

In vitro reconstitution of the α-keto acid elongation subpathway

NRPS megasynthetases exhibit the unique feature of incorporating nonproteinogenic amino acids into bioactive peptidic molecules. Such unusual and nonproteinogenic amino acids, including 4-ethyl-L-proline/4-propyl-L-proline, L-homoleucine, and L-homonorleucine as present in MPs, often contribute to the bioactivities (here increased anti-*Mtb* MIC), whereas their biogenesis often remains mysterious. Intrigued by the presence of several nonproteinogenic amino acids of unknown origin, we proposed the biosynthesis of the homo-amino acid residues in MPs as biosynthetically sharing an analogous process to leucine biosynthesis, as previously also suggested for a few other homo-amino acids [20,21]. Two

homologous genes of leucine biosynthesis were identified within the MP BGC. The gene *mypK* was designated to encode a 2-isopropylmalate synthase-like enzyme that likely catalyzes the condensation of different six-carbon α-keto acids, including 4-methyl-2-oxopentanoate (**1**), 2-oxohexanoate (**2**), and 3-methyl-2-oxopentanoate (**3**) with acetyl-CoA. The other gene encodes a didomain enzyme MypL, which contains a putative N-terminal isopropylmalate isomerase subunit (LeuD) homolog and a C-terminal pyridoxal phosphate (PLP)-dependent transaminase domain. Although the dedicated dehydrogenase (*leuB*) and the other isopropylmalate isomerase subunit (*leuC*) genes were not found in the MP pathway, the broad substrate tolerance of leucine biosynthesis enzymes was nevertheless considered to be important for homo-amino acids biosynthesis as well [22]. Thus, MypK, MypL, together with aaLeuB and aaLeuC from primary metabolism of *A. awajinensis* ATCC33919, were proposed to biosynthesize the homo-amino acid residues (**10**, **11** and **12**) in MPs (Figure 5a).

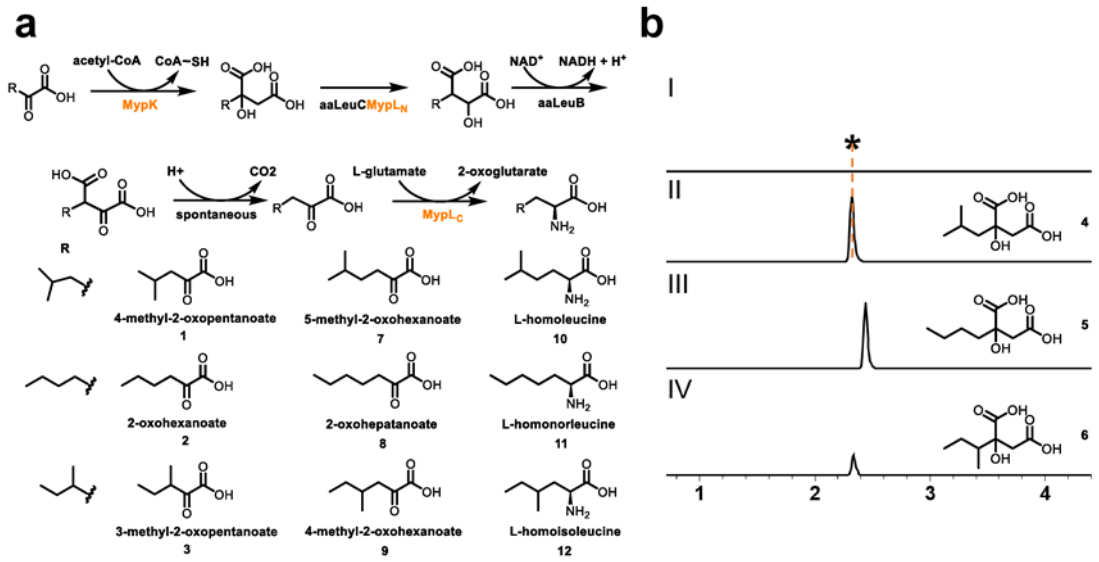


Figure 5: The homo-amino acid building blocks of MPs are built by a pathway mimicking BCAA biosynthesis; ***a****) Proposed subpathway and intermediates of* L*-homoleucine (****10****),* L*-homonorleucine (****11****) and* L*-homoisoleucine (****12****) formation;* ***b****) In vitro assays of the purified protein MypK. Extracted ion chromatogram ([M-H]$^-$, m/z 189.07) corresponding to seven-carbon α-keto acids detected in negative ion mode. I: No formation of 2- isobutylmalate (****4****) could be detected in the presence of boiled MypK and substrate* ***1*** *as control. II: The formation of* ***4*** *was observed upon incubation of MypK and substrate* ***1****, III: 2-n-butylmalate (****5****) was found when* ***2*** *and MypK were used in the reaction, IV: incubation of* ***3*** *and MypK resulted in a small amount of 2-sec-butylmalate (****6****).*

We set out to study these intriguing steps *in vitro*, because *Actinoplanes* species are notoriously recalcitrant to genetic manipulation, and there is no exception for strain ATCC33919. The acceptance of **1**, **2**, and **3** as substrates by the recombinant His-tagged MypK was demonstrated by the newly detected mass peaks showing 60 Da mass increase, corresponding to the compounds 2-isobutylmalate (**4**), 2-*n*-butylmalate (**5**), and 2-sec-butylmalate (**6**), respectively (Figure 5b). Moreover, MypK exhibits a preference for **1** compared to **2** and **3**, which was verified by measuring enzyme kinetics (Extended Data Table 1). A new peak with *m/z* shift of 1 Da corresponds to the expected mass of L-homoisoleucine (**6**) and was detected by LC-MS in the reaction of the recombinant MypL$_C$ (C-terminal domain of MypL, only soluble when co-purified with PLP) in the presence of L-glutamate as the amino group donor and 4-methyl-2-oxohexanoate (**9**) as the substrate. MypL$_C$ was tested under the same condition revealing an extensive range of proteinogenic amino acids as amino donors (Extended Data Figure 7) [23].

The substrate acceptance of 4-methyl-2-oxohexanoate for $MypL_C$ indicated that $MypL_C$ was likely involved in forming **10**, **11**, and **12**. The successful purification of remaining enzymes enabled the complete *in vitro* reconstitution of homo-amino acid biosynthesis in one pot (Figure 6c). $MypL_N$ is co-purified with aaLeuC in an intensive brown color, indicating the binding of the iron-sulfur cluster. The one-pot reaction was executed in three stages initiated by MypK, followed by the reaction of aaLeuC and $MypL_N$ co-purified protein, and aaLeuB together with β-nicotinamide adenine dinucleotide in its oxidized form (NAD^+). After the last stage reaction with $MypL_C$, D-FDLA was used to derivatize potential amino acid products. The reaction product using **1** as the substrate had the same retention time compared with a peak with the same mass from the hydrolyzed and D-FDLA derivatized MP D, suggesting the formation of **10**. Similarly, the mass corresponding to the derivatized **11** or **12** was detected from the reaction using **2** or **3** as substrate, respectively (Figure 5b and 6c).

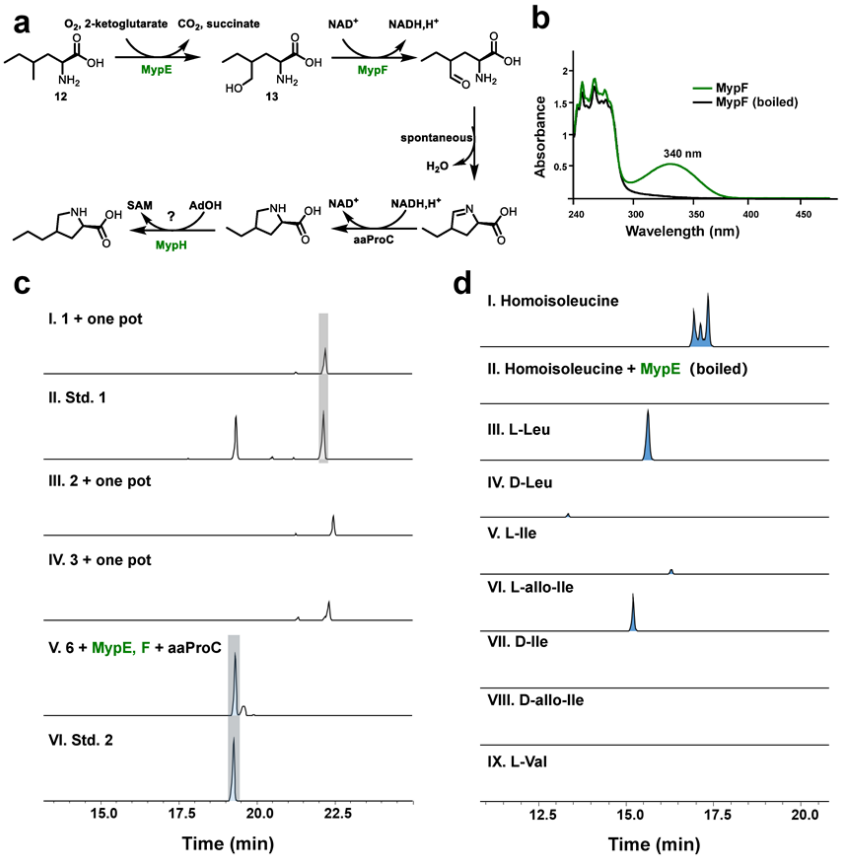


Figure 6: The alkylproline building blocks are biosynthesized using GriEF-analogous proteins starting from homoisoleucine; ***a****) Proposed subpathway of 4-ethyl-L-proline and 4-propyl-L-proline formation in MP biosynthesis;* ***b****) UV/Vis spectra of MypF and heat-inactivated MypF in the assay;* ***c****) In vitro analysis of the extender unit reconstitution. Extracted ion chromatograms (EICs) of both derivatized reaction samples and hydrolyzed MP D as control were analyzed. I: derivatized* ***10*** *(m/z 440.213) was detected after incubation in one pot (****1****, MypK, aaLeuC+MypL$_N$, aaLeuB and MypL$_C$), II: derivatized* ***10*** *(m/z 440.213) from hydrolyzed and derivatized MP D was used as standard. Derivatized* ***11*** *and* ***12*** *products (m/z 440.213) were found using* ***2*** *and* ***3*** *as substrate in one pot assay, respectively (III, IV). V: reaction in the presence of substrate* ***12****, MypE, MypF and aaProC resulted in the EIC (m/z 438.198) corresponding to 4-EtPro formation, VI: derivatized 4-EtPro (m/z 438.198) from hydrolyzed and derivatized MP D was used as standard;* ***d****) In vitro analysis of the purified proteins MypE. Extracted ion chromatograms corresponding to anticipated products were analyzed in positive ion mode. The 4-hydroxymethylhomoisoleucine (****13****) product (m/z 456.208) was detected after incubation of MypE and homoisoleucine, while the product was not formed in presence of boiled MypE instead of MypE (I, II). III: the formation of 4-hydroxy-L-Leu (m/z 442.193) was observed upon incubation of MypE and L-Leu, IV: trace amount of 4-hydroxy-D-Leu (m/z 442.193) was formed after the incubation of MypE and D-Leu, V: incubation of MypE and L-Ile resulted in the trace amount formation of 3-hydroxy-L-allo-Ile (m/z 442.193), VI: 3-hydroxy-L-allo-Ile (m/z 442.193) was found after the reaction of MypE and L-all-Ile. No reaction products were obtained using D-Ile, D-allo-Ile and L-Val as substrate (VII-IX).*

In vitro reconstitution of the 4-alkylproline subpathway

Previously, 4-EtPro and 4-PrPro were characterized to be derived from L-tyrosine in the biosynthesis of lincomycins (Supplementary Figures 4 and 5a) [24]. However, no homologous genes were found in the MP pathway. As reported in GM biosynthesis, (2*S*,4*R*)-4-MePro was the product of a series of enzymatic modifications of leucine (Figure 1a) [4]. A pair of genes *mypE* and *mypF* from the MP pathway shows homology with the genes involved in the 4-MePro biosynthesis, including *ldoA* and *nosE* from cyanobacteria, and *griE* and *griF* in GM biosynthesis [4,25]. MypE is speculated to promiscuously hydroxylate both L-leucine and **12** resulting in 5-hydroxy-L-leucine and 4-hydroxymethyl-L-homoisoleucine (**13**), respectively. Subsequently, the dehydrogenase MypF and pyrroline-5-carboxylate reductase aaProC converted 5-hydroxy-L-leucine and **13** to 4-MePro and 4-EtPro, respectively. Lastly, MypH was deduced to encode a radical SAM MT which is assumed to methylate the terminal unreactive carbon of the 4-ethyl group of 4-EtPro to yield 4-PrPro (Figure 6a).

The formation of **13** was observed as a new mass with a 16 Da increase in the D-FDLA derivatized reaction product of recombinant MypE incubated with homoisoleucine (stereoisomer mixture) since the pure (2*S*,4*R*)-homoisoleucine is not commercially available. Surprisingly, three individual new peaks with the same *m/z* were observed after derivatization by D-FDLA, suggesting loose stereospecificity of MypE (Figure 6d). A series of structural analogs were investigated for substrate tolerance of MypE, but no conversion was detected with L-valine, D-isoleucine, D-allo-isoleucine. In contrast, a trace amount of products of D-leucine and L-isoleucine, and nearly complete conversion of L-leucine and L-allo-isoleucine was observed by LC-MS (Figure 6d). These results indicated MypE is a promiscuous enzyme recognizing substrate stereoisomers, however, preferring L-forms of amino acids, especially the L-allo form amino acid in the case of two stereocenters. Unfortunately, pure stereoisomers of 4-hydroxymethylhomoisoleucine could not been synthesized in this study and also are not commercially available, hence the stereoselectivity of MypE is yet to be solved.

The reaction of MypF seemed to work properly when incubated with the purified substrate 4-hydroxymethylhomoisoleucine from *E. coli* overexpressing recombinant MypE and fed with

homoisoleucine, because the formation of NADH is prominent with a significant rise of absorbance at 340 nm (Figure 6b). A new signal corresponding to the D-FDLA-derivatized (2*S*,4*R*)-4-EtPro was observed in the coupled assay with MypE, MypF, and aaProC, which is consistent with the mass detected in the derivatized hydrolysis product of MP D (Figure 6c).

4.3 Discussion

Given the promising activity of griselimycin as a potential anti-TB drug, its innovative target DnaN, and the involvement of nonproteogenic amino acids in target binding, we became interested in identifying additional natural products exhibiting similar chemical moieties. Searching in all publically available actinobacterial genomes using (2*S*,4*R*)-4-MePro biosynthesis genes [4] unveiled a previously unknown BGC with additional interesting precursor biosynthetic genes compared to the GM pathway. This finding inadvertently disclosed the mycoplanecin BGC and correlated it to these compounds that were initially discovered in the 1980s with potent and specific anti-tuberculosis bioactivities [12–16]. Here, four MPs, including one new congener, were isolated and found to contain several unusual nonproteogenic amino acid residues, including three different 4-alkylprolines. Marfey's analysis plus the NRPS E domain analysis suggested the absolute stereochemistry of the MPs. These amino acid residues, especially (2*S*,4*R*)-4-EtPro and (2*S*,4*R*)-4-PrPro, are scarce in natural products. The 4-PrPro and 4-EtPro in lincomycins derive from L-tyrosine degradation as catalyzed by a subcluster of six genes [26], which are not present in the MP BGC. Here, a step-by-step *in vitro* total reconstitution of the BCAA pathway of MP biosynthesis reveals a set of promiscuous enzymes, which build unusual homo-amino acids from different 2-oxo acids. Subsequently, we *in vitro* reconstituted 4-EtPro using a pair of GriEF-like enzymes from the MP pathway. Unlike using leucine as the substrate in the 4-MePro pathway of GM biosynthesis, MypE hydroxylates **12** to form **13**. Subsequently, **13** is oxidized to form the semialdehyde, which undergoes a subsequent spontaneous cyclization. The pyrroline intermediate is further reduced to deliver (2*S*,4*R*)-4-EtPro. The 4-PrPro residue is unlikely to be produced following the same logic, because it would require the unknown 3-methyl-2-

oxohexanoic acid as the substrate for the MP BCAA pathway to produce 4-methyl-2-oxoheptanoic acid and for the transaminase to form the respective amino acid to enter the 4-alkylproline pathway. Since a radical SAM gene from the MP BGC could not be assigned to any other functions, (2*S*,4*R*)-4-PrPro is assumed to be synthesized by methylating the terminal unreactive carbon of (2*S*,4*R*)-4-EtPro, which needs experimental proof in the future.

Intriguingly, MPs showed significantly more potent anti-*Mtb* bioactivity than GMs. Most importantly, we demonstrated that the new congener MP E is most potent, which exhibits 24 times higher potency against *Mtb* H37Rv than GM (83 times stronger when testing *Mtb* H37Ra). We showed MP A has a four times lower MIC against *Mtb* H37Rv than the original reported MIC (0.39 µg/mL). The analogous chemical structures of MPs and GMs are in line with their identical molecular target site within DnaN. The co-crystal structure of ecDnaN bound with MP A validates DnaN as the molecular target, and the binding mode is similar to GM and CGM, albeit MPs showed significantly stronger anti-*Mtb* bioactivity. The interaction analysis shows that MPs bind to msDnaN with an affinity in the nanomolar range. In contrast to MPs' much more potent anti-tuberculosis effects *in vitro*, it is surprising that MPs bind to msDnaN 3-20 times weaker than GM. The MICs of different MPs also do not align with their binding affinities to DnaN. All these discrepancies suggest that the antibiotic efficacy of MPs depends not only on their DnaN binding, but is influenced by other factors, such as the cellular permeability to reach DnaN regardless of the question, if diffusion or active transport or even metabolism in the target microbe are involved. Moreover, MPs and GMs do not inhibit the growth of *E. coli*, despite binding to ecDnaN further supporting this hypothesis.

TB is still one of the leading public health threats, even during the global COVID-19 pandemic. Drug-resistant strains of *M. tuberculosis* further deepen the crisis because we are running out of our antibiotic arsenal. In general, the low cytotoxicity and specific anti-mycobacterial activity make MPs promising candidates to become potent and safe anti-TB drugs that are not detrimental to human microbiota. The discovery of highly potent antibiotics exhibiting the rare target DnaN, and the unusual biosynthesis of the molecular building blocks sets the stage for continuous efforts into future mycoplanecin development.

4.4 References

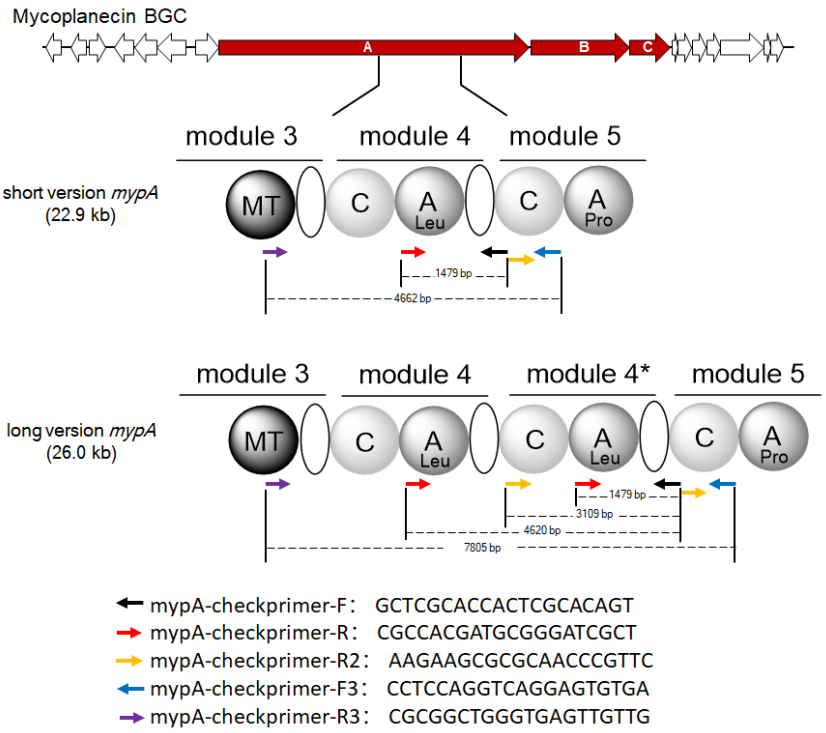
1. Miethke, M.; Pieroni, M.; Weber, T.; Brönstrup, M.; Hammann, P.; Halby, L.; Arimondo, P.B.; Glaser, P.; Aigle, B.; Bode, H.B.; et al. Towards the sustainable discovery and development of new antibiotics. *Nat. Rev. Chem.* **2021**, *5*, 726–749, doi:10.1038/s41570-021-00313-1.
2. WHO TEAM Global Tuberculosis Programme. Global tuberculosis report 2021. Available online: https://www.who.int/publications/i/item/9789240037021.
3. Kling, A.; Lukat, P.; Almeida, D.V.; Bauer, A.; Fontaine, E.; Sordello, S.; Zaburannyi, N.; Herrmann, J.; Wenzel, S.C.; König, C.; et al. Antibiotics. Targeting DnaN for tuberculosis therapy using novel griselimycins. *Science* **2015**, *348*, 1106–1112, doi:10.1126/science.aaa4690.
4. Lukat, P.; Katsuyama, Y.; Wenzel, S.; Binz, T.; König, C.; Blankenfeldt, W.; Brönstrup, M.; Müller, R. Biosynthesis of methyl-proline containing griselimycins, natural products with anti-tuberculosis activity. *Chem. Sci.* **2017**, *8*, 7521–7527, doi:10.1039/C7SC02622F.
5. Liu, L.; Jokela, J.; Herfindal, L.; Wahlsten, M.; Sinkkonen, J.; Permi, P.; Fewer, D.P.; Døskeland, S.O.; Sivonen, K. 4-Methylproline guided natural product discovery: co-occurrence of 4-hydroxy- and 4-methylprolines in nostoweipeptins and nostopeptolides. *ACS Chem. Biol.* **2014**, *9*, 2646–2655, doi:10.1021/cb500436p.
6. Terlain, B.; Thomas, J.P. Consitution de la grisélimycine, antibiotique polypeptidique extrait d cultures de streptomyces. I. Identification des produits libérés par hydrolyse. *Bull. Soc. Chim. Fr.* **1971**, *6*, 2349–2356.
7. Wink, J.; Schumann, P.; Atasayar, E.; Klenk, H.-P.; Zaburannyi, N.; Westermann, M.; Martin, K.; Glaeser, S.P.; Kämpfer, P. ‚Streptomyces caelicus', an antibiotic-producing species of the genus Streptomyces, and Streptomyces canchipurensis Li et al. 2015 are later heterotypic synonyms of Streptomyces muensis Ningthoujam et al. 2014. *Int. J. Syst. Evol. Microbiol.* **2017**, *67*, 548–556, doi:10.1099/ijsem.0.001612.

8. Bérdy, J. Thoughts and facts about antibiotics: where we are now and where we are heading. *J Antibiot (Tokyo)* **2012**, *65*, 385–395, doi:10.1038/ja.2012.27.
9. Katz, L.; Baltz, R.H. Natural product discovery: past, present, and future. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 155–176, doi:10.1007/s10295-015-1723-5.
10. Thomy, D.; Culp, E.; Adamek, M.; Cheng, E.Y.; Ziemert, N.; Wright, G.D.; Sass, P.; Brötz-Oesterhelt, H. The ADEP Biosynthetic Gene Cluster in Streptomyces hawaiiensis NRRL 15010 Reveals an Accessory clpP Gene as a Novel Antibiotic Resistance Factor. *Appl. Environ. Microbiol.* **2019**, *85*, doi:10.1128/AEM.01292-19.
11. Ju, K.-S.; Gao, J.; Doroghazi, J.R.; Wang, K.-K.A.; Thibodeaux, C.J.; Li, S.; Metzger, E.; Fudala, J.; Su, J.; Zhang, J.K.; et al. Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 12175–12180, doi:10.1073/pnas.1500873112.
12. Matsueda, R.; Higashida, S.; Furumoto, H.; Haishi, T.; Takahashi, H.; Kitami, M. MYCOPLANECIN DERIVATIVE. JP19830151030 19830819, August 19, 1983.
13. Arai, M.; Haneishi, T.; Nakajima, M.; Torikata, A.; Enokita, R. Mycoplanecin derivatives and their preparation. US19810250709 19810403, April 3, 1981.
14. Nakajima, M.; Torikata, A.; Ichikawa, Y.; Katayama, T.; Shiraishi, A.; Haneishi, T.; Arai, M. Mycoplanecins, novel antimycobacterial antibiotics from Actinoplanes awajinensis subsp. mycoplanecinus subsp. nov. II. Isolation, physico-chemical characterization and biological activities of mycoplanecin A. *J. Antibiot.* **1983**, *36*, 961–966.
15. Nakajima, M.; Torikata, A.; Tamaoki, H.; Haneishi, T.; Arai, M.; Kinoshita, T.; Kuwano, H. Mycoplanecins, novel antimycobacterial antibiotics from Actinoplanes awajinensis subsp. mycoplanecinus subsp. nov. III. Structural determination of mycoplanecin A. *J. Antibiot.* **1983**, *36*, 967–975.
16. Torikata, A.; Enokita, R.; Okazaki, T.; Nakajima, M.; Iwado, S.; Haneishi, T.; Arai, M. Mycoplanecins, novel antimycobacterial antibiotics from Actinoplanes awajinensis subsp. mycoplanecinus subsp. nov. I. Taxonomy of producing organism and fermentation. *J. Antibiot. (Tokyo)* **1983**, *36*, 957–960, doi:10.7164/antibiotics.36.957.

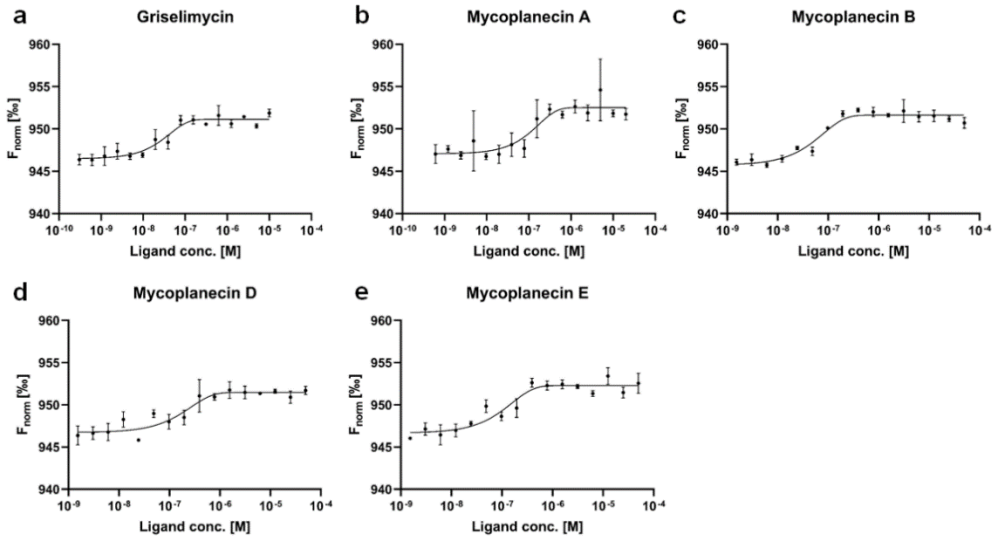
17. Lewis, R.A.; Nunns, L.; Thirlway, J.; Carroll, K.; Smith, C.P.; Micklefield, J. Active site modification of the β-ketoacyl-ACP synthase FabF3 of Streptomyces coelicolor affects the fatty acid chain length of the CDA lipopeptides. *Chem. Commun. (Camb)* **2011**, *47*, 1860–1862, doi:10.1039/c0cc03444d.
18. Hitchman, T.S.; Crosby, J.; Byrom, K.J.; Cox, R.J.; Simpson, T.J. Catalytic self-acylation of type II polyketide synthase acyl carrier proteins. *Chem. Biol.* **1998**, *5*, 35–47, doi:10.1016/s1074-5521(98)90085-0.
19. Misra, A.; Sharma, S.K.; Surolia, N.; Surolia, A. Self-Acylation Properties of Type II Fatty Acid Biosynthesis Acyl Carrier Protein. *Chem. Biol.* **2007**, *14*, 775–783, doi:10.1016/j.chembiol.2007.05.013.
20. Cacho, R.A.; Jiang, W.; Chooi, Y.-H.; Walsh, C.T.; Tang, Y. Identification and Characterization of the Echinocandin B Biosynthetic Gene Cluster from Emericella rugulosa NRRL 11440. *J. Am. Chem. Soc.* **2012**, *134*, 16781–16790, doi:10.1021/ja307220z.
21. Koketsu, K.; Mitsuhashi, S.; Tabata, K. Identification of homophenylalanine biosynthetic genes from the cyanobacterium Nostoc punctiforme PCC73102 and application to its microbial production by Escherichia coli. *Appl. Environ. Microbiol.* **2013**, *79*, 2201–2208, doi:10.1128/AEM.03596-12.
22. Atsumi, S.; Hanai, T.; Liao, J.C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **2008**, *451*, 86–89, doi:10.1038/nature06450.
23. Berger, B.J.; English, S.; Chan, G.; Knodel, M.H. Methionine regeneration and aminotransferases in Bacillus subtilis, Bacillus cereus, and Bacillus anthracis. *J. Bacteriol.* **2003**, *185*, 2418–2431, doi:10.1128/jb.185.8.2418-2431.2003.).
24. Steiningerova, L.; Kamenik, Z.; Gazak, R.; Kadlcik, S.; Bashiri, G.; Man, P.; Kuzma, M.; Pavlikova, M.; Janata, J. Different Reaction Specificities of F420H2-Dependent Reductases Facilitate Pyrrolobenzodiazepines and Lincomycin To Fit Their Biological Targets. *J. Am. Chem. Soc.* **2020**, *142*, 3440–3448, doi:10.1021/jacs.9b11234.

25. Luesch, H.; Hoffmann, D.; Hevel, J.M.; Becker, J.E.; Golakoti, T.; Moore, R.E. Biosynthesis of 4-methylproline in cyanobacteria: cloning of nosE and nosF genes and biochemical characterization of the encoded dehydrogenase and reductase activities. *J. Org. Chem.* **2003**, *68*, 83–91, doi:10.1021/jo026479q.
26. Janata, J.; Kamenik, Z.; Gazak, R.; Kadlcik, S.; Najmanova, L. Biosynthesis and incorporation of an alkylproline-derivative (APD) precursor into complex natural products. *Nat. Prod. Rep.* **2018**, *35*, 257–289, doi:10.1039/c7np00047b.

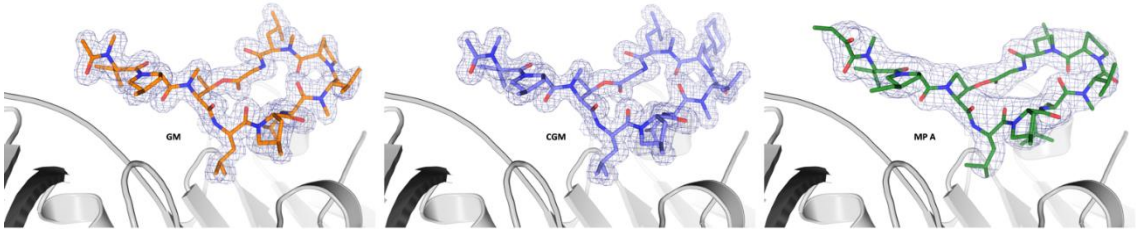
4.5 Extended data figures



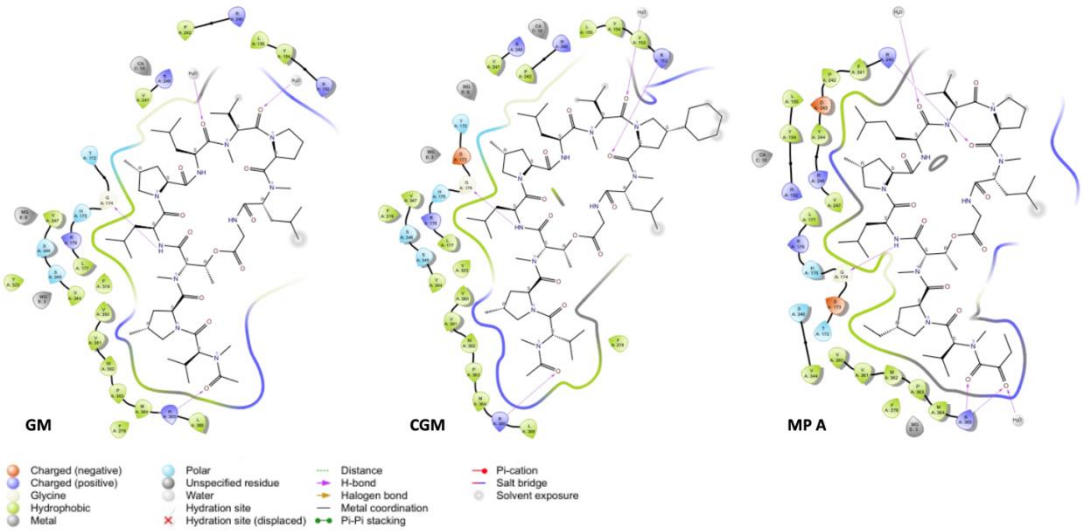
Extended Data Figure 1: Design of a combination of PCR amplification to verify the mypA gene sequence. Three sets of primers were designed to verify the suspicious repetitive DNA region of mypA. *The PCR amplification with the primer pair mypA-checkprimer-F/R should only give one 1479 bp fragment if* mypA *is the short version but two fragments should appear with the additional 4620 bp one if* mypA *is the long version. With the second pair of primers mypA-checkprimer-F/R2, the PCR will give no band with the short version* mypA *while a 3109 bp band should appear with the long version* mypA. *The third primer set mypA-checkprimer-F3/R3 beyond the questionable region should give more confirmative information as different sizes of amplicons (4662 bp for short* mypA *and 7805 bp for long* mypA*) will appear. The primer sequences of mypA-checkprimer-R (red) and mypA-checkprimer-R2 (orange) occur twice in the 26kb version of mypA.*



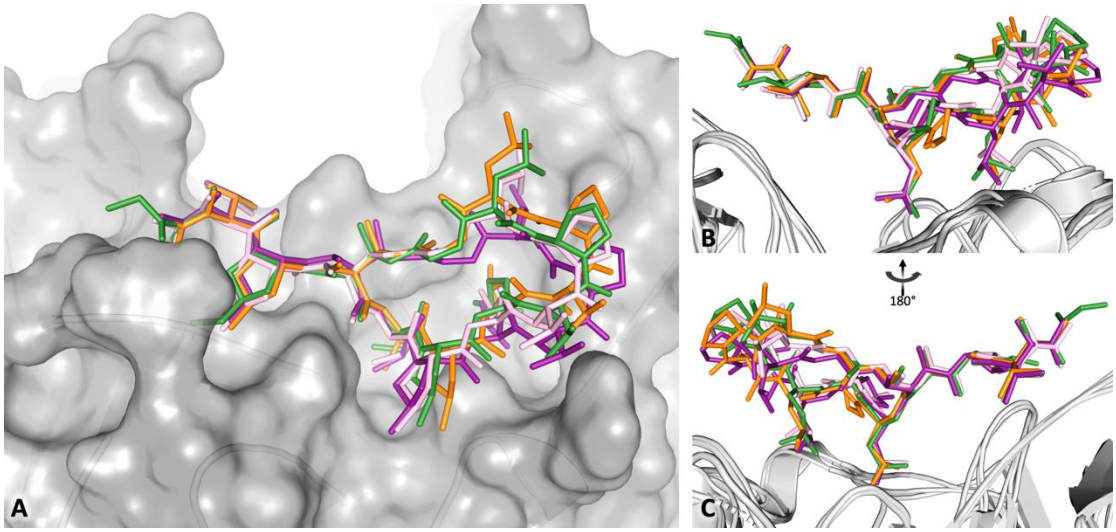
Extended Data Figure 2: MST dose-response curves for the binding interaction between BLUE-NHS labeled MsDnaN and different ligands. Experiment using 50 nM labeled DnaN and Griselimycin was titrated from 10,000 to 0.3 nM, K_D-value was determined to be 6.5±5.9 nM, MP A titrated from 20,000 to 0.6 nM, K_D-value 95.4±58.0 nM, MP B titrated from 50,000 to 1.5 nM, K_D-value 24.4±11.9 nM, MP D titrated from 50,000 to 1.5 nM, K_D-value 134.2±69.3 nM, MP E titrated from 50,000 to 1.5 nM, K_D-value 64.4 ±33.2 nM.



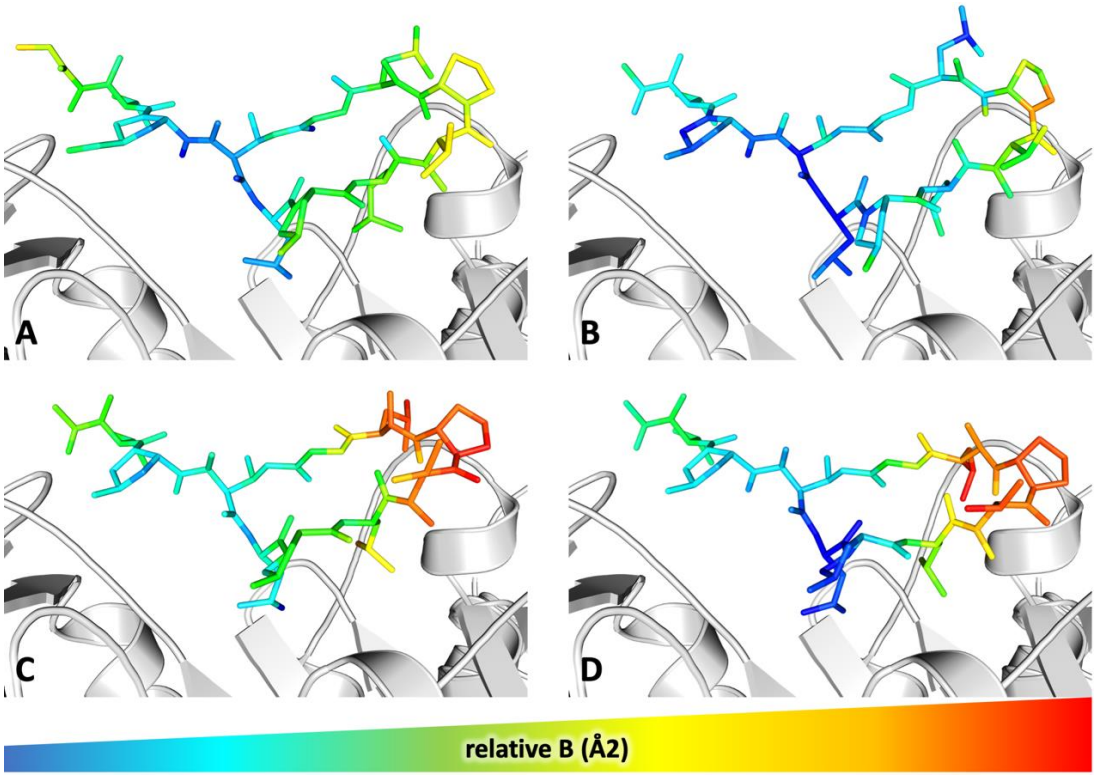
Extended Data Figure 3: Refined $2mF_{obs}$-DF_{calc} electron density maps. Electron density maps are shown as blue meshes contoured at σ = +1 for GM (orange), CGM (violet) and MP A (green) in the corresponding ecDnaN co-crystal structures.



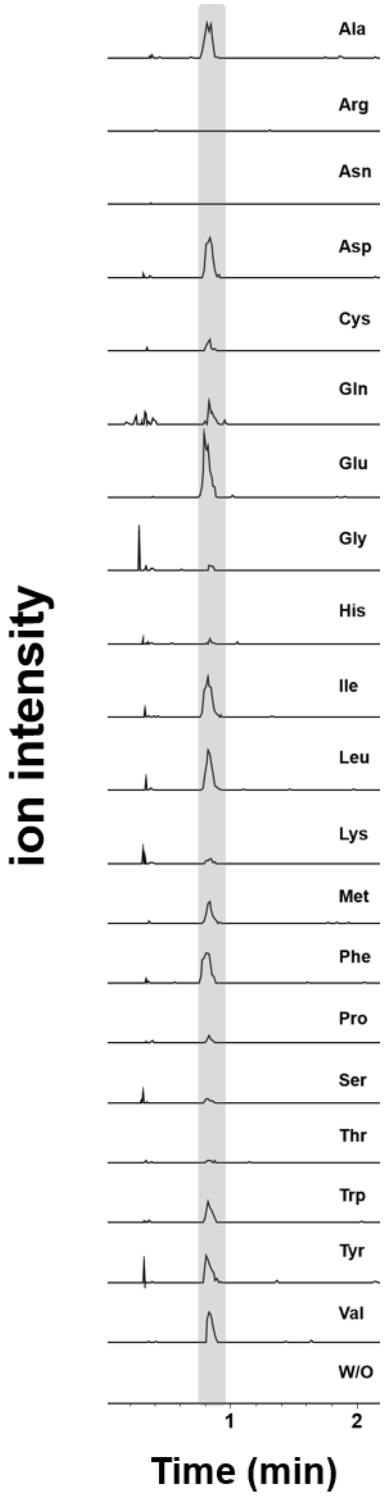
Extended Data Figure 4: Ligand interaction diagrams. Ligand interaction diagrams for GM, CGM and MPA generated from the corresponding ecDnaN co-crystal structures by Maestro from the Schrödinger2020 suite (Schrödinger, LLC).



Extended Data Figure 5: Binding pose comparison of MP A and GM in the ecDnaN complexes with GM bound to mycobacterial DnaNs; ***A****) Superimposition of the ecDnaN co-crystal structures containing MPA (green) and GM (orange), with the complexes of GM and* M.tuberculosis *DnaN (purple, PDB: 5AGU) and* M. smegmatis *DnaN (pink, PDB: 5AH2). The protein surface is shown in gray;* ***B****) Side view of the superposition from A with the protein shown as cartoon;* ***C****) Same as panel B, but turned 180°C around the Y axis.*



Extended Data Figure 6: Relative crystallographic B-factors of MP A and GM in the complexes ecDnaN and the mycobacterial proteins. Ligand atoms are coloured according to their crystallographic B-factors from low (blue) to high(red). ***A*** *MP A bound to ecDnaN.* ***B*** *GM bound to ecDnaN.* ***C*** *GM bound to msDnaN (PDB: 5AH2).* ***D*** *GM bound to* M. tuberculosis *DnaN (PDB: 5AGU).*



Extended Data Figure 7: In vitro transamination of purified MypL$_C$. Transamination was performed in the presence of purified MypL$_C$, PLP, 4-methyl-2-oxohexanoic acid, proteinogenic amino acids as amino donors respectively in vitro, *and was detected by LC-MS. The expected homoisoleucine product mass peak was indicated in gray.*

Extended Data Table 1: kinetic data for the enzymatic condensation MypK substrates

	MypK	
Substrate	Km (μM)	Kcat/Km (1/mM/s)
4-methyl-2-ketopentanoate	33.5 ± 17.9	1.63 ± 0.46
3-methyl-2-ketopentanoate	30.5 ± 0.2	0.153 ± 0.03
2-ketohexanoate	31.7 ± 7.3	1.06 ± 0.2

4.6 Materials and methods

Sequence similarity network analysis

Sequence similarity networks were constructed for all homologs of GriE and GriF found in biosynthetic gene clusters (BGCs) in all species of the phylum Actinobacteria. To this end, we conducted a BLAST [1] search with a threshold e-value of 10-5 and filtered for hits where the similarity is found for at least 75 % of the sequence of the corresponding protein. We have retrieved the corresponding genomes or genome fragments and predicted BGCs with antiSMASH [2]. Only GriE/GriF homologs within BGCs were retained. The protein sequences were aligned with MAFFT [3] and pairwise sequence similarities were calculated. The sequence similarity networks were visualized with CytoScape [4].

Strains, plasmids and reagents

Mycoplanecins producer strain *Actinoplanes awajinensis* ATCC33919 was purchased from American Type Culture Collection. The plasmid pET28b was purchased from Novagen. The plasmid pColdI was purchased from TAKARA. CoA esters and amino acids were purchased from Sigma-Aldrich and Enamine. The 4-ethylproline was synthesized and supplied by Sanofi. Reagents used for derivatization were purchased from Thermo Fisher. All enzymes required for molecular cloning were purchased from Thermo Scientific.

Strain cultivation, compound extraction and purification

Cultivation of *Actinoplanes awajinensis* was carried out in 5 x 1 L flasks using 300 mL culture volume in APL medium (1% soluble starch, 0.2% yeast extract, 1% glucose (dextrose), 1% glycerol, 0.25% corn steep liquor, 0.2% peptone (meat), 0.1% NaCl, 0.3% $CaCO_3$, adjusted to pH=7.2). Subsequent supercritical fluid extraction (SFE) was carried out following the protocol described by Bader *et al*. [5] using a Waters MV-10 ASFE system and yielding the mycoplanecins in the ethyl acetate phase. The resulting SF extract was used to purify MP A, B, D, and E using a Dionex Ultimate 3000 SL

systems comprising a SWPS 3000 SL autosampler, P680 pump, TCC100 column oven, PDA100 UV- detector and AFL 3000 fraction collector in combination with a XBridge Peptide BEH C18 OBD prep column (130 Å, 5 µm, 10 mm x 250 mm). Flowrate was set to 5 mL/min with a mobile phase containing (A) 0.1% formic acid in ddH_2O and (B) 0.1% formic acid in acetonitrile heated to 45 °C. After an equilibration phase for 2 min at 5% B in A, a 30 min gradient from 50 to 80% B was adapted for separation of the MP derivatives. The column was flushed with 95% B for 2 min before subsequent re-equilibration at 5% B for 5 min.

NMR analysis for planar structure elucidation

NMR measurements of MP A, B, D, and E were recorded using 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). ^{1}H, ^{13}C, COSY, TOCSY, HSQC and HMBC experiments were carried out in deuterated chloroform.

Marfey's analysis for stereochemical assignment

Marfey's derivatization of MP A, B, D, and E was performed to assign the amino acids' stereochemistry by comparing the retention time of the hydrolyzed and derivatized amino acids with respective standards. Hydrolysis of the amide bonds and subsequent derivatization of the fragments was carried out as follows. 0.1 mg of the respective MP derivative was hydrolyzed with 100 µL of 6 N HCl in a closed glass vial for 45 min at 110°C under nitrogen. The solvent was dried at 110°C and the residue was dissolved in 110 µL H_2O. The solution was split into two 50 µL aliquots in PP tubes and both were treated with 20 µL of 1 N $NaHCO_3$ followed by 20 µL of 1 % 1-fluroro-2,4-dinitrophenyl-5-leucine-amide (D-FDLA and L-FDLA in acetone). After incubation for 2 h at 40°C and 700 rpm, the reaction was stopped by addition of 10 µL of 2 N HCl and 300 µL of acetonitrile. Both samples were centrifuged for 10 min at 15°C and 15,000 rpm, and the supernatant was measured. Analysis was carried out on a Dionex Ultimate 3000 SL system coupled to a Bruker Daltonics maXis 4G UHR-TOF applying

electrospray ionization (ESI). Separation was performed using a Waters Acquity BEH C18 column (100 x 2.1 mm, 1.7 µm) with a mobile phase consisting of (A) 0.1 % formic acid in ddH_2O and (B) 0.1 % formic acid in acetonitrile with a flow rate of 0.6 mL/min at 45 °C. A multi-step gradient was applied to separate the derivatized amino acids: 5-10 % B in 1 min, 10-35 % B in 14 min, 35- 55 % in 7 min, 55-80 % in 3 min, hold at 80 % for 1 min and re-equilibration at 5 % B for 5 min. Before entering the mass spectrometer, the flow was split to 75 µL/min. UV data were obtained at 340 nm simultaneously with the MS detection in centroid mode ranging from 250-3000 *m/z*.

Standard LC-MS methods

UHPLC measurements for determination of exact mass, sum formula and retention time were carried out with a Dionex Ultimate 3000 SL system consisting of SWPS 3000 SL autosampler, P680 pump module, TCC100 column oven, and PDA100 UV-detector coupled to a maXis 4G UHR-TOF platform. Ionization mode was ESI with the following MS settings: capillary voltage 4000 V, end plate off-set -500 V, nebulizer gas pressure 1 bar, dry gas flow rate 5 L/min, dry gas temperature 200 °C, mass scan range 150–2500 *m/z*. The mass spectrometer was externally calibrated to sodium formate cluster using subsequent lock mass calibration. Metabolite profiling was performed using Waters Acquity BEH C18 columns (50 x 2.1 mm, 1.7 µm and 100 x 2.1 mm, 1.7 µm) connected to a Waters VanGuard BEH C18 pre-column. The standard methods for both columns are linear gradients with 5-95 % ACN + 0.1 % FA over 6, 9, or 18 minutes.

Antimicrobial screening

The biological activity of the mycoplanecins was evaluated against the attenuated *Mycobacterium tuberculosis* strains *Mtb* H37Ra and H37Rv as well as a panel of microbial pathogens consisting of *M. smegmatis* mc^2155, *M. smegmatis* mc^2155 GM-resistant strain, *M. marinum* strain M, *Acinetobacter baumanii* DSM-30008, *Escherichia coli* WT BW25113, *Citrobacter freundii* DSM-30039, *Staphylococcus*

aureus Newman, *Bacillus subtilis* DSM-10 and *Mucor hiemalis*. Furthermore, we determined the cytotoxicity of the mycoplanecins against CHO-K1.

Cloning, expression, and purification with general condition

The genes *aaproC, mypE, mypF, aaleuB, mypK* were amplified by PCR using the genomic DNA of *Actinoplanes awajinensis* ATCC 33919 as the template. The obtained fragments were subcloned into the pHISTEV or pET28b vector via Gibson assembly or conventional T4 ligation, resulting in the protein expression plasmids. The genes *aaleuC* and *mypL$_N$* were also amplified by PCR using the genomic DNA of *Actinoplanes awajinensis* ATCC 33919 as a template. The obtained fragments were subcloned into the pACYCDuet-1 vector via conventional ligation stepwise, resulting in the expression plasmid pACYCDuet-1-*AaleuCmypL$_N$*.

One liter of 2YT medium supplemented with kanamycin was inoculated with *E. coli* BL21 (DE3) strains that harbored expression plasmids accordingly. The culture was incubated at 37 ˚C until OD_{600} reached 0.6. The culture was then transferred into a 16 ˚C shaker for 30 min. Once the culture was cooled to 16 ˚C, the expression was induced by the addition of 0.1 mM Isopropyl-β-D-thiogalactopyranosid (IPTG). After overnight shaking under 16 ˚C, cells were harvested by centrifugation and resuspended in lysis buffer [20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 20 mM imidazole, 10% glycerol and 1 mM DTT]. Lysis was achieved via passage through a cell disrupter (Constant Systems) at 30,000 psi, and cell debris was removed by centrifugation (19,500 g, 4 ˚C, 60 min). The supernatant was collected and loaded onto a 5-ml Histrap HP column (GE healthcare) pre-equilibrated with lysis buffer. The column was washed extensively with lysis buffer (20 column volumes) before the protein was eluted with elution buffer (lysis buffer with 250 mM imidazole). Fractions containing N-terminal His-tag protein were directly loaded onto HiPrep 26/10 desalting column pre-equilibrated in desalting buffer [20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 10 % glycerol and 1 mM DTT]. The samples were collected, concentrated and flash frozen for storage at -80 ˚C.

Cloning, expression and purification of MypL$_C$

The gene *mypL$_C$* was amplified by PCR using the genomic DNA of *Actinoplanes awajinensis* ATCC 33919 as a template. The obtained fragment was subcloned into the pHISTEV vector via conventional ligation, resulting in the expression plasmid pHISTEV-*mypL$_C$*.

1 L of 2YT medium supplemented with kanamycin and 1 μM *pyridoxal 5'-phosphate* (PLP) was inoculated with *E. coli* BL21 (DE3) harboring pHISTEV-m*ypL$_C$*. The culture was incubated at 37 ˚C until OD_{600} reached 0.6. The culture was then transferred into 16 ˚C shaker for 30 min. Once the culture was cool, expression of MypL$_C$ was induced by the addition of 0.1 mM IPTG. After overnight shaking under 16 ˚C, Cells were harvested by centrifugation and resuspended in lysis buffer [50 mM PBS (pH 8.0), 300 mM NaCl, 20 mM imidazole, 1 μM PLP, and 1 mM DTT]. Lysis was achieved via passage through a cell disrupter at 30,000 psi, and cell debris was removed by centrifugation (19,500 g, 4 ˚C, 60 min). The supernatant was collected and directly loaded onto a 5 ml Histrap HP column pre-equilibrated with lysis buffer. The column was washed extensively with lysis buffer (20 column volumes) before the protein was eluted with elution buffer (lysis buffer with 250 mM imidazole). Fractions containing N-terminal His-tag MypL$_C$ were directly loaded onto HiPrep 26/10 desalting column pre-equilibrated in desalting buffer (50 mM PBS (pH 8.0), 300 mM NaCl, 1 μM PLP and 1 mM DTT). The sample was collected, concentrated and flash frozen for storage at -80 ˚C.

***In vitro* biochemical analysis of MypK**

The reaction mixture containing 0.5 mM 4-methyl-2-oxopentanoate (**1**) (or other α-keto acids including 2-oxohexanoate (**2**) and 3-methyl-2-oxopentanoate (**3**)), 0.5 mM acetyl-CoA, 50 mM NaCl, 10 mM $MgCl_2$, 0.1 mg/mL BSA, 0.5 mM DTT,10 mM Tris buffer (pH 7.5) and 1 μM of MypK in a total volume of 50 μl was incubated at 30˚C for 1 hour before being quenched by addition of 100 μl methanol. Precipitated protein was removed by centrifugation, and the supernatant was directly applied to LC-MS analysis.

For the analysis of the kinetic parameters of MypK reactions, 100 µl reaction mixture containing 2 mM acetyl-CoA, 10 mM $MgCl_2$, 50 mM potassium phosphate buffer (pH8.0), 3 µM of MypK, and 10-1000 µM α-keto acids were incubated at 30˚C. The initial velocity of the reaction was estimated by monitoring the increase in absorbance at 412 nm according to Ellman's assays [6].

***In vitro* biochemical analysis of MypL$_c$**

The reaction mixture containing 0.1 mM 4-methyl-2-oxo-hexanoic acid, 1 mM L-glutamate (or other amino acids), 50 mM NaCl, 10 mM $MgCl_2$, 0.1 mg/mL BSA, 0.5 mM DTT,10 mM Tris buffer (pH 7.5) and 100 nm MypL$_c$ in total volume of 50 µl was incubated at 30˚C for 1 hour before quenched by adding 100 µl methanol. Precipitated protein was removed by centrifugation, and the supernatant was directly applied to LC-MS analysis.

One pot *in vitro* reconstitution of homoleucine, homonorleucine and homoisoleucine

The reaction mixture contained 1 mM **1** or **2** or **3**, 1 mM acetyl-CoA, 3 mM NAD^+, 10 mM $MgCl_2$, 0.5 mM $MnCl_2$, 0.2 mM DTT, 50 mM potassium phosphate buffer buffer (pH 8) and 1 µM of MypK, 0.5 µM AaLeuC and MypL$_N$, and 2 µM AaLeuB in total volume of 50 µl. After overnight incubation at 30˚C, 10 µM PLP, 2 mM L-glutamate, 0.5 µM AaLeuC and MypL$_N$, 2 µM AaLeuB and 1µM MypLc were added. The reaction was quenched by adding 100 µl methanol after 2 hours of incubation at 30˚C. Precipitated protein was removed by centrifugation, and the supernatant was removed by nitrogen evaporation. The residual material was dissolved in 5 µl MilliQ water. 2 µl $NaHCO_3$, 8 µl acetone, and 2 µl of 1 % D-FDLA were added to the solution. The derivatization was carried out at 37 ˚C for 1 hour with rotation. The reaction was stopped by adding 33 µl 1 N HCl. The mixture was further mixed with 50 µl acetonitrile and analyzed by LC-MS.

In vitro analysis of MypE

The reaction mixture containing 0.5 mM substrate, 1 mM α-ketoglutarate, 20 µM $FeSO_4$, 0.2 mM ascorbate, 0.5 mM DTT, 5 mM potassium phosphate buffer (pH8.0) and 3 µM of MypE in a total volume of 50 µl was incubated at 30˚C for 3 hours before being quenched by the addition of 50 µl of methanol. Precipitated protein was removed by centrifugation, and the supernatant was transferred to a new tube. The water was removed under a stream of nitrogen and the residual material was dissolved in 5 µl MilliQ water. 2 µl $NaHCO_3$, 8 µl acetone and 2 µl of 1% D-FDLA were added into the solution [7]. The derivatization was carried out at 37 ˚C for 1 hour with rotation. The reaction was stopped by adding 33 µl 1N HCl. The mixture was further mixed with 50 µl acetonitrile and analyzed by LC-MS.

Production of 4-hydroxymethylhomoisoleucine in E. coli BL21 (DE3) harboring pET28b-MypE

100 ml of 2YT medium containing the kanamycin were inoculated with *E. coli* BL21 (DE3) harboring pET28-*mypE* and incubated at 37˚C until OD_{600} reached 0.6. 0.1 mM IPTG was added to induce the expression of MypE. The bacteria were further incubated at 16 ˚C overnight. The cells were harvested by centrifugation and resuspended in 100 ml M9 minimal medium containing 2% glucose and 1 mM homoisoleucine. The reaction was carried out for 2 days and cells were removed by centrifugation. The supernatant was transfer to a new Falcon™ tube. Water was removed by lyophilization. After the resuspension in methanol and centrifugation, the residual material was removed. 4-hydroxymethylhomoisoleucine was further purified using semi-preparative HPLC equipped with a Luna 5µm C18 250 x 10 mm 100 Å (Phenomenex) using water/acetonitrile/formic acid=95/5/0.1.

In vitro analysis of MypF

The reaction mixture containing 0.5 mM 4-hydroxymethylhomoisoleucine, which was isolated from *E. coli* BL21 (DE3) overexpressing MypF, 1 mM $ZnSO_4$, 2 mM NAD^+,

0.5 mM DTT, 50 mM of glycine buffer (pH 10) and 5 µM of MypF in a total volume of 50 µl was incubated at 30˚C for 12 hours before stopped by adding 100 µl methanol. Precipitated protein was removed by centrifugation, and the supernatant was directly applied to LC-MS analysis. NADH formation in the reaction was monitored via the increase in absorbance at 340 nm.

Coupled assays of 4-ethylproline formation

The reaction mixture containing 0.5 mM homoisoleucine, 1 mM α-ketoglutarate, 20 µM $FeSO_4$, 0.2 mM ascorbate, 0.5 mM DTT, 5 mM potassium phosphate buffer (pH8.0) and 3 µM of MypE in total volume of 100 µl was incubated at 30˚C for 1 hours. The incubation proceeded after the solution was then supplemented with glycine buffer (pH 10), $ZnCl_2$, NAD^+, MypF and aaProC to a final concentration of 100 mM, 1 mM, 2 mM, 5 µM and 5 µM respectively overnight at 30˚C. The reaction was quenched by adding 100 µl methanol. Precipitated protein was removed by centrifugation and the supernatant was removed by evaporation under a stream of nitrogen. Residual material was dissolved in 5 µl MilliQ water. 2 µl $NaHCO_3$, 8 µl acetone and 2 µl of 1% D-FDLA were added into the solution. The derivatization was carried out at 37 ˚C for 1 hour with rotation. Reaction was stopped by adding 33 µl 1 N HCl. The mixture was further mixed with 50 µl acetonitrile and analyzed by LC-MS.

Intact protein measurements

All ESI-MS measurements were performed on a Dionex (Germering, Germany) Ultimate 3000 RSLC system coupled with a maXis4G Q-ToF mass spectrometer using an Aeris Widepore XB-C8 column (3.6 µm, 150 × 2.1 mm; Phenomenex) column. Separation of 1 µL sample was achieved by a multistep gradient from (A) H_2O + 0.1 % FA to (B) ACN + 0.1 % FA at a flow rate of 300 µL/min and 45 °C. Chromatographic conditions were as follows: 0-1 min, 2 % B; 1-11 min, 2-75 % B; 11-14 min, 75 % B; 14-14.35 min, 75-2 % B; 14.35-15.35 min, 2 % B. UV spectra were recorded by a DAD in the range from 200 to 600 nm. The LC flow was split to 75 µl/min before entering the maXis 4G hr-ToF mass spectrometer (Bruker Daltonics, Bremen, Germany) using

the standard Bruker Apollo II ESI source. In the source region, the temperature was set to 200 °C, the capillary voltage was 4000 V, the dry-gas flow was 5.0 L/min and the nebulizer was set to 1.0 bar. Mass spectra were acquired in positive ionization mode ranging from 150 – 2500 *m/z* at 2 Hz scan rate. Protein masses were deconvoluted by using the Maximum Entropy algorithm.

MPs bind to DnaN in nanomolar affinity

GM and MPs A, B, D, and E were investigated regarding their binding affinity to *M. smegmatis* DnaN (msDnaN) using MST by evaluating the dissociation constant K_D. Experiments were performed according to the standard protocol by NanoTemper using a Monolith NT-115 Microscale Thermophoresis device. *M. smegmatis* DnaN was labeled using Monolith Series Protein Labeling Kit BLUE-NHS and diluted 50 nM with MST buffer (20 mM Tris, 150 mM NaCl, 0.05 % Tween-20) in the assay. All compounds were titrated in an appropriate concentration range for proper dose-response curves with 16 serial dilution steps.

Production of ecDnaN protein

The gene encoding for ecDnaN was cloned into a modified pCOLA Duet-1 vector (Novagen) encoding for an N-terminal Strep-tag II and TEV-protease recognition site. The protein was expressed in *E. coli* BL21 (DE3) in ZYM-5052 auto-inducing medium [8] at 20°C for 20-24 h. The cell pellet was resuspended in a buffer containing 20 mM HEPES/NaOH pH 7.5, 150 mM NaCl, one tablet of complete EDTA-free protease inhibitor cocktail (Roche) and lysed by sonication. The protein was isolated from the supernatant after centrifugation for 1 h at 100,000 x g using a self-packed 10 ml column with Strep-Tactin Superflow High Capacity resin (IBA) and eluted from the column with a single step of 5 mM D-desthiobiotin. The affinity tag was cleaved off with TEV protease (1:50 mg/mg) at 4°C overnight. Gel filtration was carried out using a HiLoad 16/600 Superdex 200 pg column (GE Healthcare) in 20 mM HEPES/NaOH pH 7.5,

150 mM NaCl. The peak fractions were concentrated and flash-frozen in liquid nitrogen for crystallization screening.

Crystallization

Crystallization trials were set up at room temperature with a Crystal Gryphon crystallization robot (Art Robbins Instruments) in Intelli 96-3 plates (Art Robbins Instruments) with 200 nL protein solution at different concentrations and 200 nL reservoir solution. Crystals were obtained after a few days in co-crystallization setups (Conditions: Supplementary Table S8). Crystals were, after harvesting, cryo-protected by addition of 10% (*v*/*v*) (2*R*,3*R*)-2,3-butanediol.

Data collection and processing

Data collection for ecDnaN with MP A was performed at beamline P11 at the Petra III storage ring (Deutsches Elektronen-Synchrotron, Hamburg, Germany) [9], for ecDnaN with GM at beamline BL 14.2 of the BESSY II (Helmholtz Zentrum Berlin, Germany) [10] and for ecDnaN with CGM at beamline X06DA (PXIII) of the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland). All datasets were recorded at a temperature of 100 K. Data processing was carried out using the AutoPROC [11] toolbox (Global Phasing) executing XDS [12], Pointless [13], and Aimless [14].

Structure determination, refinement, and model building

The structure of the ecDnaN ligand-complexes were determined by molecular replacement using the available structure from the PDB (4K3L [15]) as a search-model for Phaser [16] from the Phenix suite [17]. The structural models were built using Coot [18] and crystallographic refinement was performed with Phenix.refine [19] including the addition of hydrogens in riding positions and TLS-refinement. 5% of random reflections were flagged for the calculation of R_{free}. The model of ecDnaN in complex with MP A was at 2.1 Å resolution and refined to R/R_{free} of 22/26% in space group C2. The structure of ecDnaN in complex with GM was at 1.7 Å resolution and refined to

R/R_{free} of 18/19% in space group C222. The structure of ecDnaN in complex with CGM was at 1.5 Å resolution and refined to R/R_{free} of 17/18% in space group C222. Data collection and refinement statistics are summarized in Appendix Tab S1. Figures of crystal structures were prepared using the PyMOL Molecular Graphics System version 2.4.0 (Schrödinger, LLC).

4.7 References

1. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410, doi:10.1016/S0022-2836(05)80360-2.
2. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; van Wezel, G.P.; Medema, M.H.; Weber, T. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **2021**, *49*, W29-W35, doi:10.1093/nar/gkab335.
3. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780, doi:10.1093/molbev/mst010.
4. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504, doi:10.1101/gr.1239303.
5. Bader, C.D.; Neuber, M.; Panter, F.; Krug, D.; Müller, R. Supercritical Fluid Extraction Enhances Discovery of Secondary Metabolites from Myxobacteria. *Anal. Chem.* **2020**, *92*, 15403–15411, doi:10.1021/acs.analchem.0c02995.
6. Riener, C.K.; Kada, G.; Gruber, H.J. Quick measurement of protein sulfhydryls with Ellman‘s reagent and with 4,4‘-dithiodipyridine. *Anal. Bioanal. Chem.* **2002**, *373*, 266–276, doi:10.1007/s00216-002-1347-2.
7. Lukat, P.; Katsuyama, Y.; Wenzel, S.; Binz, T.; König, C.; Blankenfeldt, W.; Brönstrup, M.; Müller, R. Biosynthesis of methyl-proline containing griselimycins, natural products with anti-tuberculosis activity. *Chem. Sci.* **2017**, *8*, 7521–7527, doi:10.1039/C7SC02622F.
8. Studier, F.W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **2005**, *41*, 207–234, doi:10.1016/j.pep.2005.01.016.
9. Burkhardt, A.; Pakendorf, T.; Reime, B.; Meyer, J.; Fischer, P.; Stübe, N.; Panneerselvam, S.; Lorbeer, O.; Stachnik, K.; Warmer, M.; et al. Status of the

crystallography beamlines at PETRA III. *Eur. Phys. J. Plus* **2016**, *131*, doi:10.1140/epjp/i2016-16056-0.

10. Mueller, U.; Förster, R.; Hellmig, M.; Huschmann, F.U.; Kastner, A.; Malecki, P.; Pühringer, S.; Röwer, M.; Sparta, K.; Steffien, M.; et al. The macromolecular crystallography beamlines at BESSY II of the Helmholtz-Zentrum Berlin: Current status and perspectives. *Eur. Phys. J. Plus* **2015**, *130*, doi:10.1140/epjp/i2015-15141-2.
11. Vonrhein, C.; Flensburg, C.; Keller, P.; Sharff, A.; Smart, O.; Paciorek, W.; Womack, T.; Bricogne, G. Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr D Biol Crystallogr* **2011**, *67*, 293–302, doi:10.1107/S0907444911007773.
12. Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 125–132, doi:10.1107/S0907444909047337.
13. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **2006**, *62*, 72–82, doi:10.1107/S0907444905036693.
14. Evans, P.R.; Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **2013**, *69*, 1204–1214, doi:10.1107/S0907444913000061.
15. Yin, Z.; Kelso, M.J.; Beck, J.L.; Oakley, A.J. Structural and Thermodynamic Dissection of Linear Motif Recognition by the E. coli Sliding Clamp. *J. Med. Chem.* **2013**, *56*, 8665–8673, doi:10.1021/jm401118f.
16. McCoy, A.J.; Grosse-Kunstleve, R.W.; Adams, P.D.; Winn, M.D.; Storoni, L.C.; Read, R.J. Phaser crystallographic software. *J Appl Crystallogr* **2007**, *40*, 658–674, doi:10.1107/S0021889807021206.
17. Adams, P.D.; Afonine, P.V.; Bunkóczi, G.; Chen, V.B.; Davis, I.W.; Echols, N.; Headd, J.J.; Hung, L.-W.; Kapral, G.J.; Grosse-Kunstleve, R.W.; et al. PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 213–221, doi:10.1107/S0907444909052925.

18. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66*, 486–501, doi:10.1107/S0907444910007493.
19. Afonine, P.V.; Grosse-Kunstleve, R.W.; Echols, N.; Headd, J.J.; Moriarty, N.W.; Mustyakimov, M.; Terwilliger, T.C.; Urzhumtsev, A.; Zwart, P.H.; Adams, P.D. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **2012**, *68*, 352–367, doi:10.1107/S0907444912001308.

S4. Supplementary discussion

Coexistence of self-resistance gene and putative transposase genes in the MP BGC

Two putative transposase genes (*tnp1* and *tnp2*) were identified in the 5'-end of the MP BGC. It was also observed in the GM pathway that five hypothetic transposase and integrase genes located next to the self-resistance gene *griR* [1].

Validation of the *mypA* gene sequence

The domain analysis of these two *mypA* versions indicated the difference is one extra Leu-utilizing NRPS module 4* in the longer version between module 4 and module 5 in the shorter version (Extended Data Figure 1). In the long version, modules 4 and 4* are almost identical (Supplementary Figure 1b). The high identity between the condensation (C) domains from modules 4, 4*, and 5 caused the situation to be more difficult (Supplementary Figure 1b). However, the inconsistency between ten amino acid building blocks and eleven NRPS modules suggested a deviation from the collinearity rule in MPs biosynthesis. To verify whether the redundant module was an artifact that resulted from the assembly of highly repetitive DNA sequences, we thus took a closer look at this region and PCR amplified several fragments within and beyond the repetitive region. The sequences of these fragments were eventually determined by a series of walking sequencing and restriction digest analyses, which excludes the presence of the additional module and confirms the shorter version to be the correct one without module skipping during MP biosynthesis (Extended Data Figure 1 and Supplementary Figure 3).

The α-ketobutyric acid moiety formation in MP biosynthesis

Frequently, the acyl substrates are adenylated by a fatty acyl-AMP ligase (FAAL) and the activated acyl-adenylates will be installed onto the concomitant acyl carrier protein (ACP) for further assembly [2]. However, the FAAL gene has not been observed in the MP BGC. Nevertheless, the acyl-CoA could also be directly utilized to acylate the following amino acid by the starter C domain without the aid of FAAL and ACP [3]. However, the direct use of acyl-CoA by the starter C domain in MypA is unlikely since the initial building block in MPs is always the α-ketobutyric moiety, which is in contrast to the broad substrate spectrum of such particular C domains [3,4]. It is worth noting that other uncommon strategies exist in the NRPS-mediated lipopeptide chain initiation. For instance, fatty acid synthase (FAS)-like enzymes are involved in the formation of the C6 fatty acid side chain of calcium-dependent antibiotics (CDAs) [5]. Type II FAS genes are also present in the MPs BGC, *mypJ* has been found to

encode a KAS III homologous protein, while the deduced product of the accompanying gene *mypI* is a standalone ACP. The active site analysis of KAS III proteins indicated that MypJ shows low similarity to DpsC and CerJ, which exhibited acyl activation and ester formation activity respectively, but more similar to ZhuH in R1128 biosynthesis (Supplementary Table S4) [6–8]. Therefore, it is hypothesized that the holo-MypI is either malonylated by the malonyl-CoA-acyl carrier protein transacylase (MCAT) FabD from fatty acid biosynthesis or functions as a self-priming ACP that widely exists in fatty acid and type II PKS biosynthesis [9,10]. Subsequently, MypJ extends the acyl chain by a molecule of acetyl-CoA to form the acetoacetyl-*S*-ACP species. The acetoacetyl-*S*-MypI is probably initially reduced to butyryl-*S*-MypI by reductases from the type II fatty acid biosynthesis. Whether the butyryl-*S*-MypI was oxidized before or after it is transferred onto the NRPS assembly line by C1 is yet unclear. Moreover, the origin of the α-keto group remained inexplicable. Nevertheless, a P450 oxidase encoding gene *mypP* is assumed to be responsible for the cryptic α-keto group formation since no other obvious oxidation steps are needed in MP biosynthesis.

To the best of our knowledge, the N-terminal α-ketobutyric acid moiety of MPs is unique and so far not observed in other lipopeptides. The origin and activation of this four-carbon chain are still elusive. The most common concomitant fatty acyl-AMP ligase (FAAL) and acyl carrier protein (ACP) genes for NRPS lipoinitiation were not observed here [2]. Also, the alternative direct use of acyl-CoA by the starter C domain in MypA is unlikely since the constant α-ketobutyric moiety in MPs is in contrast to the broad substrate spectrum of such special C domains [3,4].

Marfey's analysis

The reference for 5-methyl-norleucine was only available as an isomer and therefore just used to distinguish 5-methyl-norleucine from N-methyl-leucine.

The 4-MePro comprised in MP B was found to be L-configured by Marfeys' analysis. For the 4-PrPro in MP D and homonorleucine residue in MP E, L-configuration has been proposed due to the lack of epimerase domains in respective NRPS modules in MP BGC albeit no reference was available for experimental proof of the stereochemistry.

S4.1. Supplementary figures

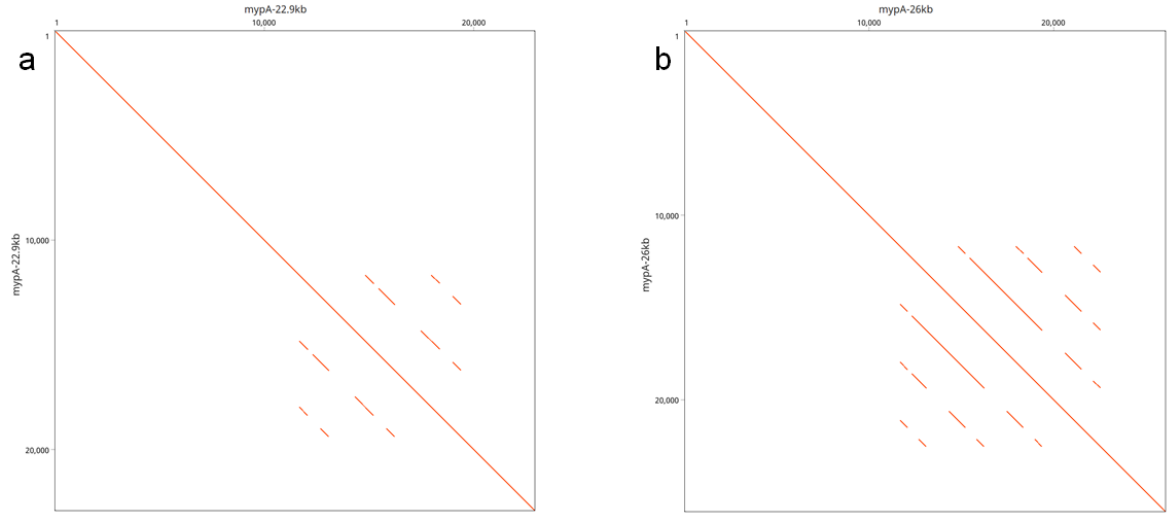


*Figure S1: Dot plots of two versions f mypA genes. The repetitive regions are shown by dot plots of mypA-22.9kb **a**) and mypA-26kb; **b**) gene sequences against themselves, respectively. Red diagonal lines are the regions of similarity between two sequences. In this self-comparison context, they would be repetitive regions in the gene.*

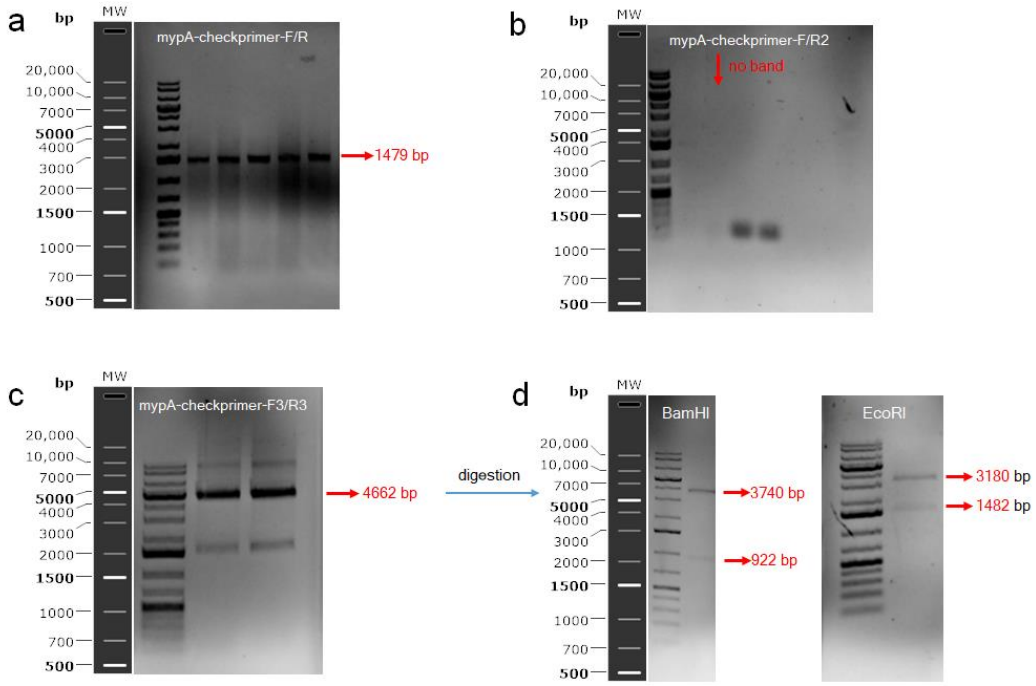


*Figure S2: PCR amplified fragments from the genomic DNA of A. awajinensis subsp. mycoplanecinus subsp. nov. ATCC 33919 using a combination of primers confirm the mypA gene should be the 22.9 kb version. **a**) Amplification using mypA-checkprimer-F and R only delivered one 1479 bp fragment which agrees with the 22.9 kb version but not two fragments including the 4620 bp one present in the 26 kb version; **b**) Amplification using mypA-checkprimer-F and R2 delivered no band which agrees with the 22.9 kb version but not the 3109 bp band in the mypA-26 kb version; **c**) Amplification using mypA-checkprimer-F3 and R3 delivered the 4662 bp band which agrees with the 22.9 kb version but not the 7805 bp band in the mypA-26 kb version; **d**) Restriction enzyme digestion using BamHI and EcoRI showed correct bands, respectively.*

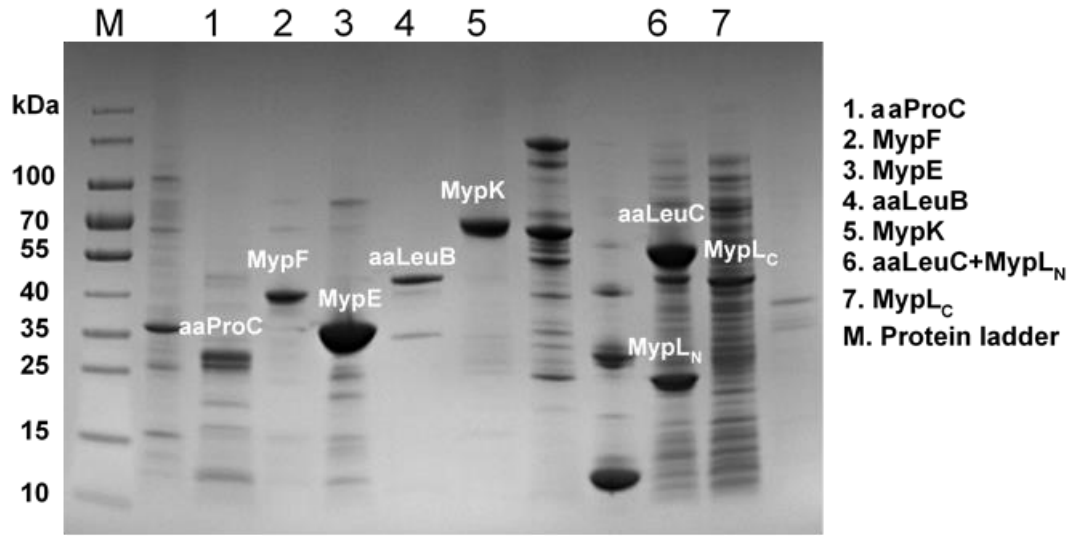


Figure S3: SDS-PAGE gel of recombinant proteins involved in this study.

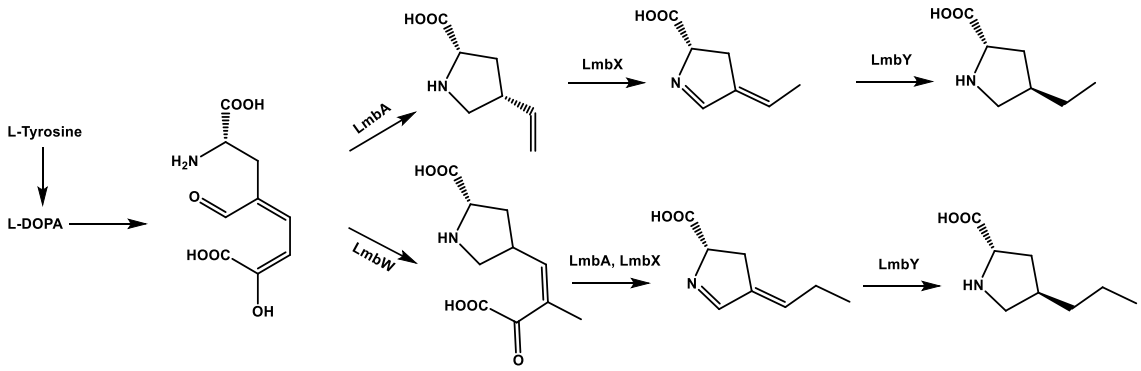


Figure S4: Biosynthetic routes to 4-EtPro and 4-PrPro in lincomycins biosynthesis [11].

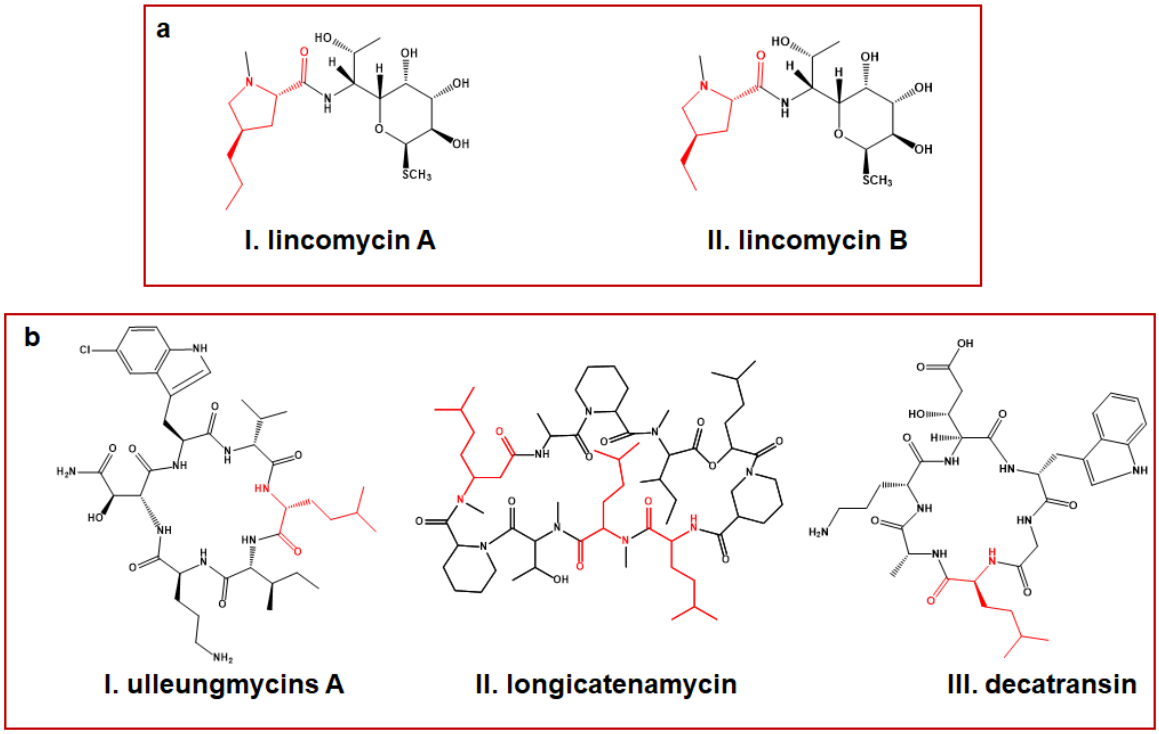


Figure S5: Examples of natural products containing 4-ethylproline, 4-propylproline and homoleucine; ***a****) Compounds incorporating the 4-alkyl-L-proline. I: lincomycin A comprised trans-4-propyl-L-proline, II: lincomycin B comprised trans-4-ethyl-L-proline;* ***b****) Compounds incorporating the homoleucine. I, II, III ulleungmycins A, longicatenamycin and decatransin was incorporated with homoleucine respectively [12–15].*

S4.2. Supplementary tables

Table S1: All homologs of GriE and GriF found in biosynthetic gene clusters (BGCs) in all species of the phylum Actinobacteria which were used in sequence similarity networks in this study.

Genome_cluster	Type	BGC Class	Cluster_start	Cluster_end	GriE_coords	_GriF_coords
BEWB01000004_c6	1	Betalactone Discrete NRPS	510782	538378	515361, 516149	514204, 515295
BMUX01000001_c1	1	Betalactone Discrete NRPS	837322	865359	841853, 842644	840729, 841832
BMVC01000001_c1	1	Betalactone Discrete NRPS	837128	865165	841659, 842450	840535, 841638
CP034353_c8	1	Betalactone Discrete NRPS	1677831	1705607	1700156, 1700944	1701010, 1702101
CP045547_c9	1	Betalactone Discrete NRPS	1886730	1914486	1909101, 1909889	1909880, 1911046
CP054920_c8	1	Betalactone Discrete NRPS	1645958	1673770	1668319, 1669107	1669173, 1670264
JADWMQ010000101_c0	1	Betalactone Discrete NRPS	0	18359	12908, 13696	13762, 14853
KZ195577_c11	1	Betalactone Discrete NRPS	1705967	1733724	1728291, 1729079	1729145, 1730236
LIXJ01000008_c0	1	Betalactone Discrete NRPS	63603	107240	105408, 106217	104296, 105444
NEVD01000016_c0	1	Betalactone Discrete NRPS	29316	57073	51640, 52428	52494, 53585
NSKH01000004_c5	1	Betalactone Discrete NRPS	359825	439360	394013, 394801	394819, 395904
SOAV01000003_c1	1	Betalactone Discrete NRPS	198965	225241	221022, 221813	221828, 222952
SOPG01000001_c6	1	Betalactone Discrete NRPS	494201	521977	498865, 499653	497711, 498802
WHZR01000001_c6	1	Betalactone Discrete NRPS	1520282	1548047	1542678, 1543466	1543457, 1544623
WXVT01000001_c10	1	Betalactone Discrete NRPS	1928489	1956238	1950835, 1951623	1951689, 1952780
BMWJ01000007_c1	2	NRPS	179756	252843	222308, 223078; 216373, 217131	221154, 222257
JAGINS010000002_c6	2	NRPS	634405	707492	671022, 671780; 676957, 677727	675803, 676906
LWKZ01000003_c1	2	NRPS	174473	266030	250667, 251434	251503, 252606
ML123098_c1	2	NRPS	324460	396020	359429, 360187; 365364, 366134	364210, 365313
RPGY01000005_c0	2	NRPS	2	47993	17337, 18107; 11402, 12160	16183, 17286
CP011522_c32	3	NRPS	5127728	5203944	5145624, 5146388	5146445, 5147533
CP028719_c39	3	NRPS	7065888	7200318	7082988, 7083788	7083809, 7084894
CP032543_c12	3	NRPS	2936423	3009905	2994060, 2994824	2992885, 2993964
CP053189_c16	3	NRPS	1843849	1913056	1894934, 1895698	1893792, 1894877
FMCJ01000051_c0	3	NRPS	0	40355	21558, 22322	20413, 21567
JAEMGU010000001_c42	3	NRPS	6973436	7101656	6991179, 6991979	6992000, 6993085
KZ819167_c22	3	NRPS	4186860	4260348	4201942, 4202706	4202802, 4203881
MTKC01000003_c22	3	NRPS	5066973	5197017	5084779, 5085579	5085600, 5086685
MUNF01000351_c0	3	NRPS	0	25524	13908, 14708	12802, 13887
PDER01000012_c0	3	NRPS	144129	273112	162132, 162932	162953, 164038

PDES01000009_c0	3	NRPS	82384	211160	100387, 101187	101208, 102293
QBHW01000021_c4	3	NRPS	512664	589232	530631, 531395	531491, 532570
QQVZ01000002_c5	3	NRPS	537287	666984	554322, 555122	555143, 556228
WUEJ01000006_c1	3	NRPS	245124	321326	302835, 303599	301660, 302739
WWGH01000001_c12	3	NRPS	2930291	3006795	2988029, 2988793	2986884, 2988038
WWJD01000054_c0	3	NRPS	0	40355	21558, 22322	20413, 21567
JABFHJ010000001_c24	4	NRPS	4725011	4800277	4742055, 4742813	4742875, 4743957
JOBF01000033_c0	5	NRPS	1	203107	168686, 169486	169507, 170598
JOFN01000038_c0	5	NRPS	0	54041	37165, 37965	37986, 39077
AP023408_c28	6	discrete NRPS	5497636	5538330	5533034, 5533825	5531908, 5532921
BBNO01000004_c1	6	discrete NRPS	666527	712869	675796, 676587	676617, 677714
BLIO01000001_c24	6	discrete NRPS	5448132	5494212	5484306, 5485097	5483179, 5484282
CP023691_c25	6	discrete NRPS	3148326	3194488	3157568, 3158359	3158383, 3159486
CP050976_c28	6	discrete NRPS	3295694	3338498	3301651, 3302442	3302466, 3303569
MIGA01000055_c0	6	discrete NRPS	0	31935	21878, 22669	20751, 21854
BMRG01000014_c0	7	NRPS	68083	173432	80007, 80765	78904, 79977
JAGINV010000001_c4	7	NRPS	942060	1039636	1026931, 1027689	1027719, 1028792
CP011497_c20	8	NRPS	5387775	5436495	5412820, 5413611	5413647, 5414729
CP050974_c17	8	NRPS	6243443	6286839	6269120, 6269890	6269947, 6271029
PJNF01000001_c2	8	NRPS	552796	595944	578310, 579080	579137, 580219
CP029078_c12	9	NRPS	1005942	1134891	1045619, 1046410; 1051528, 1052304	1050392, 1051495
CP034687_c34	9	NRPS	7822138	7951207	7910740, 7911531; 7904846, 7905622	7905655, 7906758
FMCI01000220_c0	9	NRPS	6762	55784	42665, 43456; 36767, 37522	37576, 38679
WWIV01000231_c0	9	NRPS	6762	55784	42665, 43456; 36767, 37522	37576, 38679
BMSQ01000002_c1	10	NRPS	413388	489395	463666, 464442	464487, 465578
CP023690_c39	10	NRPS	7364786	7440020	7389515, 7390291	7388379, 7389470
CP030771_c32	10	NRPS	7308085	7382556	7332639, 7333415	7331503, 7332585
CP040916_c31	10	NRPS	7029502	7103198	7054285, 7055061	7053149, 7054240
JACHJD010000006_c2	10	NRPS	465022	541029	489976, 490752	488840, 489931
JOJM01000063_c0	10	NRPS	15070	90797	65116, 65913	65937, 67043
MJAH01000029_c0	11	NRPS	33121	99405	51798, 52580	50653, 51762
RSEG01000038_c0	11	NRPS	0	38440	18982, 19743	19815, 20915
JODE01000022_c0	12	NRPS	50449	179288	160188, 160985	161009, 162097
JODJ01000024_c0	12	NRPS	0	94756	75656, 76453	76477, 77565
RQJC01000010_c0	12	NRPS	6668	231088	78257, 79054	77151, 78233
JABQ01000048_c0	13	NRPS	0	79694	54200, 54997	53070, 54173

RJUU01000001_c11	14	NRPS	4356238	4439789	4372837, 4373631	4373682, 4374776
BCBX01000034_c0	15	NRPS	24102	146170	52008, 52784	52805, 53908
BCBX01000034_c1	15	NRPS	24102	98089	52008, 52784	52805, 53908
CP010407_c10	16	PKS/NRPS	740227	867884	778907, 779716	777795, 778898
JODW01000047_c0	16	PKS/NRPS	0	72267	53174, 53971	53992, 55104
QQVZ01000001_c13	16	PKS/NRPS	2850324	2976622	2917428, 2918231	2918228, 2919346
SSBJ01000014_c0	16	PKS/NRPS	1	177964	22069, 22863	20948, 22075
WBOF01000001_c14	17	PKS/NRPS	4870129	5053211	4960794, 4961555	4959649, 4960734
JAEKDR010000004_c27	18	PKS + discrete NRPS	7235958	7361097	7353495, 7354256	7352326, 7353417
JAEKDS010000004_c3	18	PKS + discrete NRPS	546146	671285	552988, 553749	553827, 554918
JADOUA010000001_c22	20	NRPS	5030741	5128718	5056107, 5056913	5056928, 5058019
BOML01000092_c0	21	NRPS	0	57271	5957, 6745	4827, 5966
CP021978_c30	22	ADEP	7148026	7211204	7165162, 7165965	7165968, 7167092
JACHJK010000013_c5	22	ADEP	263525	302204	288384, 289148	289196, 290311
MK047367_c0	22	ADEP	0	36600	3393, 4196	4199, 5323
BDBI01000010_c0	23	NRPS	0	33909	18719, 19480	19498, 20583
BDBN01000007_c0	23	NRPS	53230	100784	72068, 72829	70959, 72053
BDBR01000033_c0	23	NRPS	77930	116022	110910, 111680	111665, 112774
LJJF01000001_c20	25	NRPS	2318375	2361226	2335655, 2336452	2334540, 2335646
PDCM01000002_c42	26	NRPS	4067389	4163430	4083892, 4084659	4082735, 4083823
KP211414_c0	27	Griselimycin BGC_NRPS	0	66865	56431, 57240	57240, 58412
LLZH01000049_c0	28	Mycoplanecin BGC_NRPS	0	18661	6455, 7231	7240, 8343

Table S2: Deduced functions of genes in the mycoplanecins biosynthetic gene cluster from Actinoplanes awajinensis ATCC 33919.

Gene	Size(aa)	Predicted function in MP biosynthesis	Best match protein/organism	Best match accession no. (identity/similarity %)
orf (-2)	462		IS1380 family transposase [*Plantactinospora sp. KBS50*]	WP_157757003.1 (87/92)
tnp1 (-1)	260		IS5/IS1182 family transposase [*Micromonospora*]	WP_030500385.1 (84/91)
orf (-1)	420		Group II intron reverse transcriptase/maturase [*Micromonospora sp.* HNM0581]	WP_169599428.1 (95/96)
mypR	373	Resistance	DNA polymerase III subunit beta[*Streptomyces sp.* CNH099]	WP_027752940.1 (44/62%)
mypO	305		putative transposase [*Micromonospora sp.* CNZ280]	NII45199.1 (86/92%)
mypN	373		IS30 family transposase [*Nonomuraea spiralis*]	CCW15551.1 (86/92%)
mypM	395	*S*-adenosylmethionine synthetase	methionine adenosyltransferase [*Actinoplanes octamycinicus*]	WP_185045711.1 (96/98%)
mypL	574	N-terminal: isopropylmalate isomerase small subunit C-terminal: aminotransferase	aminotransferase class I/II [*Streptomyces sp.* Root1304]	KQX58119.1 (44/59%)

myp K	536	2-hydroxy-2-isobutylsuccinic acid synthase	2-isopropylmalate synthase [*Streptomyces sp.* MJM8645]	WP_063351482.1 (58/72%)
myp P	391		cytochrome P450 [*Sphaerisporangium rubeum*]	WP_184987000.1(47/63%)
myp A	7621	NRPS (C-A-MT-T-C-A-PCP-C-A-MT-PCP-C-A-PCP-C-A-PCP-C-A-PCP-C)[a]	non-ribosomal peptide synthase [*Bradyrhizobium oligotrophicum*]	BAM86977.1 (53/66%)
myp B	3988	NRPS (A-MT-PCP-C-A-PCP-C-A-MT-PCP-E) [a]	nonribosomal peptide synthetase [*Streptomyces muensis*]	AKC91849.1 (52/63%)
myp C	1284	NRPS (C-A-PCP-TE) [a]	Dimodular nonribosomal peptide synthase [*Streptomyces sp.* ADI98-12]	RPK80042.1 (46/60%)
myp D	76	MbtH	MbtH family protein [*Amycolatopsis xylanica*]	WP_091286165.1 (63/75%)
myp E	261	Homoisoleucine hydroxylase	leucine hydroxylase [*Streptomyces sp.* DSM 40835]	AKC91859.1 (57/75%)
myp F	371	dehydrogenase	alcohol dehydrogenase catalytic domain-containing protein [*Streptomyces sp.* ADI98-12]	WP_124288541.1 (60/74%)

myp G	248	TE	thioesterase [*Streptomyces sp.* NRRL B-24085]	WP_053852351.1 (49/65%)
myp H	628	Radical-SAM methyltransferase	hypothetical protein [*Streptomyces atratus*]	WP_072485841.1 (57/73%)
mypI	95	Acyl carrier protein	acyl carrier protein [*Acholeplasma granularum*]	WP_025725118.1 (49/67%)
mypJ	326	ketosynthase	ketoacyl-ACP synthase III [*Streptomyces sp.* CB02009]	WP_073904150.1 (63/76%)
tnp2 (+1)	260		IS5/IS1182 family transposase [*Micromonospora*]	WP_030500385.1 (92/93%)
orf (+2)	134		transposase family protein [*Micromonospora craniellae*]	WP_191088930.1 (87/89%)
orf (+3)	100		DDE superfamily endonuclease [*Micromonospora olivasterospora*]	TWH69634.1 (88/93%)

a. Abbreviations of catalytic nonribosomal peptide synthetase (NRPS) domains: C = condensation domain, A = adenylation domain, MT = methylation domain, T = thiolation domain, also known as peptidyl carrier protein (PCP) domain in NRPS, E = epimerization domain, TE = thioesterase domain.

Table S3: Substrate specificity analysis of the 11 adenylation (A) domains from the mycoplanecins megasynthetase.

	Specificity-conferring code[a]										Predicted substrate[b]	Incorporated[c]
	235	236	239	278	299	301	322	330	331	517		
A1	D	F	W	S	V	G	I	V	H	K	MeThr	MeVal
A2	D	V	Q	F	V	G	H	A	C	K	Pro	MePro/EtPro/PropPro
A3	D	F	W	S	V	G	M	V	H	K	MeThr	MeThr
A4	D	A	M	L	V	G	A	I	A	K	Leu	Leu
A5	D	V	Q	Y	V	G	H	A	C	K	Pro	MePro
A6	D	V	M	V	I	G	G	I	A	K	'asc'	Homoleu
A7	D	V	M	V	I	G	G	I	A	K	MeThr	MeVal
A8	D	V	Q	Y	V	A	H	V	S	K	Pro	Pro
A9	D	A	M	L	V	G	A	I	A	K	Leu	Leu/homonorleu
A10	D	I	L	Q	I	G	A	I	W	K	Gly	Gly

[a] Residues defined according to Stachelhaus *et al.* and Challis *et al.*; residue numbering corresponds to gramicidin S synthetase A domain numbering (GrsA, PDB-ID: 1AMU)

[b] Retrieved from reports of the applied antiSMASH bacteria version gene cluster analysis. In A7, no single substrates but a group of several amino acids with similar physico-chemical properties were predicted by NRPSPredictor2 as indicated by 'asc' = aromatic side chain amino acids (phe, trp, phg, tyr, bht).

[c] Amino acids incorporated in the mycoplanecins structures.

Table S410: Analysis of active site, acyl-transfer and C-O bond formation region residues of MypJ homologues.

	Active site residues (AT activity)			Middle pocket residues		
DpsC[a]	S118	H297	D302	V116	Y165	A329
CerJ[b]	C116	H295	D300	R115	Y164	A327
ZhuH	C121	H257	N288	A120	F171	A320
MypJ	C121	H252	N283	A120	F166	A315

[a] Residue highlighted in yellow represent the conservative site for acyl-transfer region of DpsC.

[b] Residue highlighted in light green represent the conservative site for C-O bond formation region of CerJ.

Table S5: Walking primers used for the verification of MP gene cluster.

Name	Sequence of primer
mypA-checkprimer-F	5'-GCTCGCACCACTCGCACAGT-3'
mypA-checkprimer-R	5'-CGCCACGATGCGGGATCGCT-3'
mypA-checkprimer-R2	5'-AAGAAGCGCGCAACCCGTTC-3'
mypA-checkprimer-F3	5'-CCTCCAGGTCAGGAGTGTGA-3'
mypA-checkprimer-R3	5'-CGCGGCTGGGTGAGTTGTTG-3'

Table S6: Primers used for protein expression.

Name	Sequence of primer
AaProC-F	TGGTGCCGCGCGGCAGCCATATGACCAAGCACACCGTCGC
AaProC-R	CTCGAGTGCGGCCGCAAGCTTCAGACGGCTTGCGCGGCGA
MypF-F	ATCATATCGAAGGTAGGCATATGATATCGCCGGGCATCGC
MypF-R	TATCTAGACTGCAGGTCGACCTATCGGCCCACGCCGAGCA
MypE-F	TGGTGCCGCGCGGCAGCCATATGCAGTTGTCGGAGCAACA
MypE-R	CTCGAGTGCGGCCGCAAGCTTCACGCGACGGCTCCCTCGA
AaLeuB-F	GCG**CCATGG**CTAGCGTGGCGCGGATCGCGGTA[b]
AaLeuB-R	CGC**AAGCTT**TCAGACGGCGGCGGCGACCCGGT
MypK-F	TGGTGCCGCGCGGCAGCCATATGGCACCAGACGAGCCATT
MypK-R	GGTGCTCGAGTGCGGCCGCACTAGGGCGACGATGTGGCGA
$MypL_C$-F	ATC**GGATCC**ACCACCAAATTTTCGGCGCGCA
$MypL_C$-R	GTG**CTCGAG**CTATATGAACGCGGATATCCGA
$MypL_N$-F	ATA**CATATG**GAGAAGTGGACAGTGCACACCG
$MypL_N$-R	GCC**GATATC**TTACCGACGTGAGACGCGAGGCAGGA
AaLeuC-F	TCC**GAATTC**GGTGGGAGTCACTCCCCAGGGC
AaLeuC-R	CGC**AAGCTT**TCAGAGGTCGGCGGGGGCGGCCA

[a] Homology arm used for Gibson assembly highlighted in red.

[b] Restriction sites used for cloning in bold and underlined.

Table S7: X-ray data collection and refinement statistics.

Structure	ecDnaN + GM	ecDnaN + CGM	ecDnaN + MP A
PDB-ID:	**XXXX**	**XXXX**	**XXXX**
Data collection			
Beamline	BESSY BL 14.1	SLS X06DA (PXIII)	DESY P11
Wavelength (Å)	0.92	1.00	1.03
Space group	C222	C222	C2
Cell dimensions			
a, *b*, *c* (Å)	83.94, 150.57, 71.59	83.89, 150.71, 71.49	84.14, 151.02, 72.59
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 103.95, 90.00
Resolution (Å)[a]	71.59 – 1.65 (1.68 – 1.65)	51.86 – 1.50 (1.504 – 1.499)	56.66 – 2.12 (2.115 – 2.118)
R_{merge} (%)[a]	10.5 (118.7)	6.4 (129.3)	7.2 (89.4)
R_{pim} (%)[a]	3.0 (36.6)	1.8 (37.0)	3.0 (35.8)
$I/\sigma I$[a]	17.8 (2.2)	25.9 (2.0)	11.6 (2.2)
Completeness (%)[a]	100.0 (100.0)	100.0 (99.7)	98.9 (99.2)
Redundancy[a]	13.3 (12.6)	13.3 (13.2)	6.8 (7.1)
$CC_{1/2}$ (%)[a]	100.0 (66.4)	100.0 (74.0)	99.9 (78.0)
Refinement			
Resolution (Å)	1.65	1.50	2.12
No. reflections	54857	72933	49264
R_{work}/ R_{free} (%)	17.53/19.49	16.69/18.07	22.23/26.46
No. atoms	3856	4112	6570
Protein	3343	3449	6327
Ligand/ion	105	112	168
Water	408	551	75
B-factors (Å^2)	27.28	27.17	68.99
Protein	25.92	25.40	68.90
Ligand/ion	28.41	29.70	75.40
Water	36.94	37.71	62.33
R.m.s deviations			
Bond lengths (Å)	0.005	0.005	0.003
Bond angles (°)	0.849	1.015	0.641
Ramachandran statistics (%)			
Favored	98.31	98.32	96.37
Allowed	1.69	1.68	3.63
Outliers	0.00	0.00	0.00
Clashscore (MolProbity)	4.75	5.36	3.82
MolProbity score	1.24	1.60	1.41

[a]Values for the highest resolution shell are shown in parentheses.

Table S8: Crystallization conditions.

Co-crystal	ecDnaN + GM	ecDnaN + CGM	ecDnaN + MP A
Protein concentration	20 mg/ml	20 mg/ml	10 mg/ml
Ligand concentration	10 mM	10 mM	1 mM
DMSO in sample	10 % (v/v)	10 % (v/v)	2 % (v/v)
Temperature	273 K	273 K	273 K
Precipitants/Salts	0.36 M $CaCl_2$	0.64 M $CaCl_2$	0.2 M $CaCl_2$
	0.13 M $MgCl_2$		0.15 M $MgCl_2$
	0.24 M Na Acetate	0.5 M Li Acetate	
	8.3 % (v/v) Glycerol		8.75 % (v/v) Glycerol
	12.2 % (w/v) PEG 3350	9.4 % (w/v) PEG 8000	17.5 % (w/v) PEG 3350
Buffer	0.1 M HEPES/NaOH	0.1 M HEPES/NaOH	0.1 M Tris/HCl
pH	7.7	7.7	9.0
Cryo protectant	10 % (v/v) (2R,3R)-2,3-butanediol.	10 % (v/v) (2R,3R)-2,3-butanediol.	10 % (v/v) (2R,3R)-2,3-butanediol.

S4.3. Structure elucidation of mycoplanecins using NMR and Marfey's analysis

1D and 2D experiments as well as high resolution mass spectrometry data revealed the sequence of MP A to be as follows: α-ketobutyrate-(N-Me-Val)-(4-Et-Pro)-(Leu)-(4-Me-Pro)-(5-Me-Norleu)-(N-Me-Val)-(Pro)-(N-Me-Leu)-(Gly)-(N-Me-Thr). Cyclization of the depsipeptide takes place between the C-terminal glycine and the N-methyl-threonine hydroxyl function forming an ester-bond. A detailed description of the NMR analysis is enclosed.

Mycoplanecin A

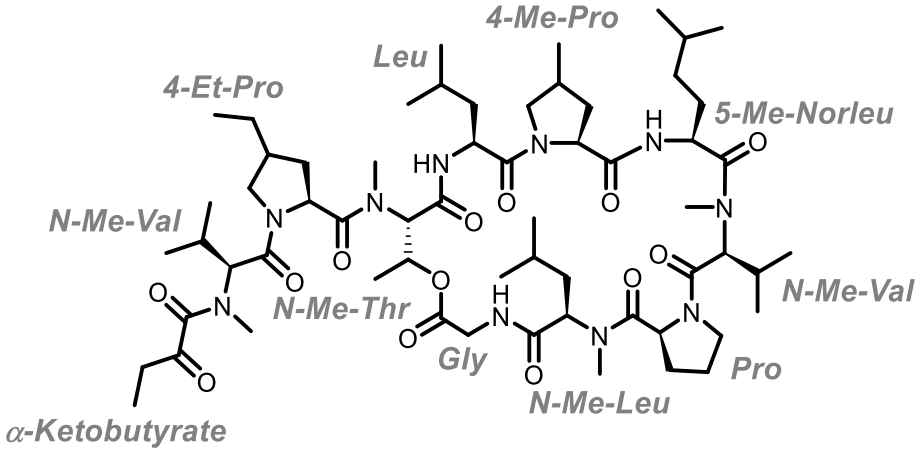


Table S9: NMR spectroscopic data of MP A in $CDCl_3$ at 500/125 MHz.

position	NMR data in $CDCl_3$				
	δ_C [ppm]	δ_H [ppm], mult. (*J* [Hz])	COSY correlations	HMBC correlations	TOCSY correlations
α-Keto-butyrate					
1	168.0	-	-	-	-
2	201.7*	-	-	-	-
3	33.5	2.76, m	4	2	4
4	6.8	1.15, t (7.61)	3	2	3
N-Me-Val					
1	167.8	-	-	-	-
2	58.8	4.94, m	3	1,3,4,5,N-Me,1 *α-Ketobutyrate*	3,4,5

3	27.1	2.33, m	2,4,5	1,2,4,5	4,5
4,5	18.9	0.99, d (6.24)	2	2,3	3
N-Me	30.7	2.97*, s	-	1 *α-Ketobutyrate,* 2	-
4-Et-Pro					
1	173.3	-	-	-	-
2	56.4	4.94, m	3	1,3,4,5	3,5,7
3	34.3	2.11,1.82, ddd (12.49, 6.72, 2.34)	2,4	4,5,6	4
4	39.3	2.46, m	3,5	3,5,6,7	3,5
5	52.7	3.97,3.34, dd (10.04,7.84)	4,5	2,3,4,6	3,4,6,7
6	26.2	1.36, m	7	-	3,5,7
7	12.4	0.88, m	6	4,6	6
N-Me-Thr					
1	168.9	-	-	-	-
2	59.9	5.59, d (3.12)	3	1,3,4,N-Me,1 *4-Et-Pro*	3,4
3	69.0	5.92, m	2,4	2,4,1 Gly	2,4
4	17.7	1.35, d (6.33)	3	2,3	2,3
N-Me	34.2	3.13*, s	-	2, 1 *4-Et-Pro* ,2 *4-Et-Pro*	2
Leu					
1	173.5	-	-	-	-
2	50.9	4.38, dd (10.04, 8.30)	3	1,3,4,1 *N-Me-Thr*	3,4,5,6, NH
3	40.1	1.51, 1.28, m	2,4	5,6	2,5,6, NH
4	25.3	1.75, m	5,6	-	-
5,6	23.2	0.93, m	4	3,4,5,6	-
NH	-	6.47, d (7.79)	2	2,3,1 *N-Me-Thr*	2,3,4,5,6
4-Me-Pro					

1	170.5	-	-	-	-
2	60.2	4.59, d (8.16)	3	1,4,5,1 *Leu*	3,5,6
3	33.1	2.56, 1.40, m	2,4	2,4,5,6	2,4,5,6
4	33.6	2.45, m	3,5,6	3,5,6	3,5,6
5	53.3	3.11, 3.51, m	4	4	2,3,4,6
6	16.7	1.08, d (6.33)	4	3,5	2,3,4,5
5-Me-Norleu					
1	174.5	-	-	-	-
2	48.5	4.83, m	3,NH	1,4,1 *4-Me-Pro*	-
3	30.7	1.75, 1.44, m	2,4	1,2,4,5	NH
4	34.5	1.22, 1.04, m	3,5	2,5,6,7	NH
5	27.9	1.52, m	4,6,7	4,6,7	-
6,7	22.3	0.89, m	5	4	NH
NH	-	7.89, s	2	2,1 *4-Me-Pro*	3,4,6,7
N-Me-Val					
1	168.7	-	-	-	-
2	57.6	4.79, m	3	1,3,4, N-Me, 1 *5-Me-Norleu*	4,5
3	28.2	2.26, m	2,5	1,2,4,5	2,5
4	18.8	0.89, m	-	-	2,5
5	18.3	0.75, d (6.69)	3	1,2,3,4	2,3,4
N-Me	30.8	3.30*, s	-	2,1 *5-Me-Norleu*	-
Pro					
1	169.8	-	-	-	-
2	58.6	4.79, m	3	-	3,5
3	31.4	2.32, 1.98, m	2	1,4	2,5
4	21.8	1.87, m	5	2	5
5	46.8	3.70, 3.59, m	4	2,4, 1 *N-Me-Val*	2,3,4
N-Me-Leu					
1	171.5	-	-	-	-

2	54.9	5.17, dd (8.53, 6.05)	3	1,3,4,N-Me,1 *Pro*	3,5,6
3	36.8	1.81, 1.41, m	2,4	1,2,4	2,4,5,6
4	24.5	1.36, m	3,5,6	3	3,5,6
5,6	20.9	0.92, m	4	3,4	2,3,4
N-Me	30.4	2.86*, s	-	2,1 *Pro*	-
Gly					
1	170.6	-	-	-	-
2	42.9	4.25, 4.01, m	NH	1,1 *N-Me-Leu*	NH
NH	-	8.78, dd (6.05, 3.94)	2	1,2,1 *N-Me-Leu*	2

*****compared to chemical shifts determined by Nakajima, deviation of 0.03 ppm in ^{1}H experiment and 1 ppm in ^{13}C experiment [16].

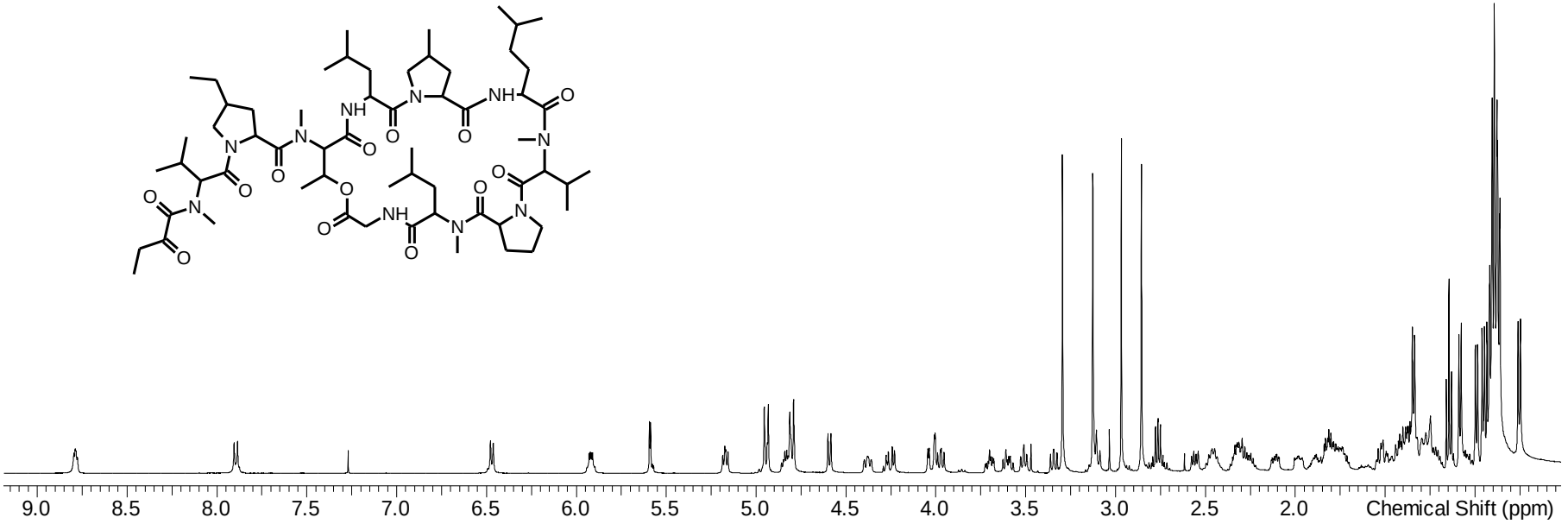


Figure S6: ^{1}H-spectrum of MP A in $CDCl_3$ at 500 MHz.

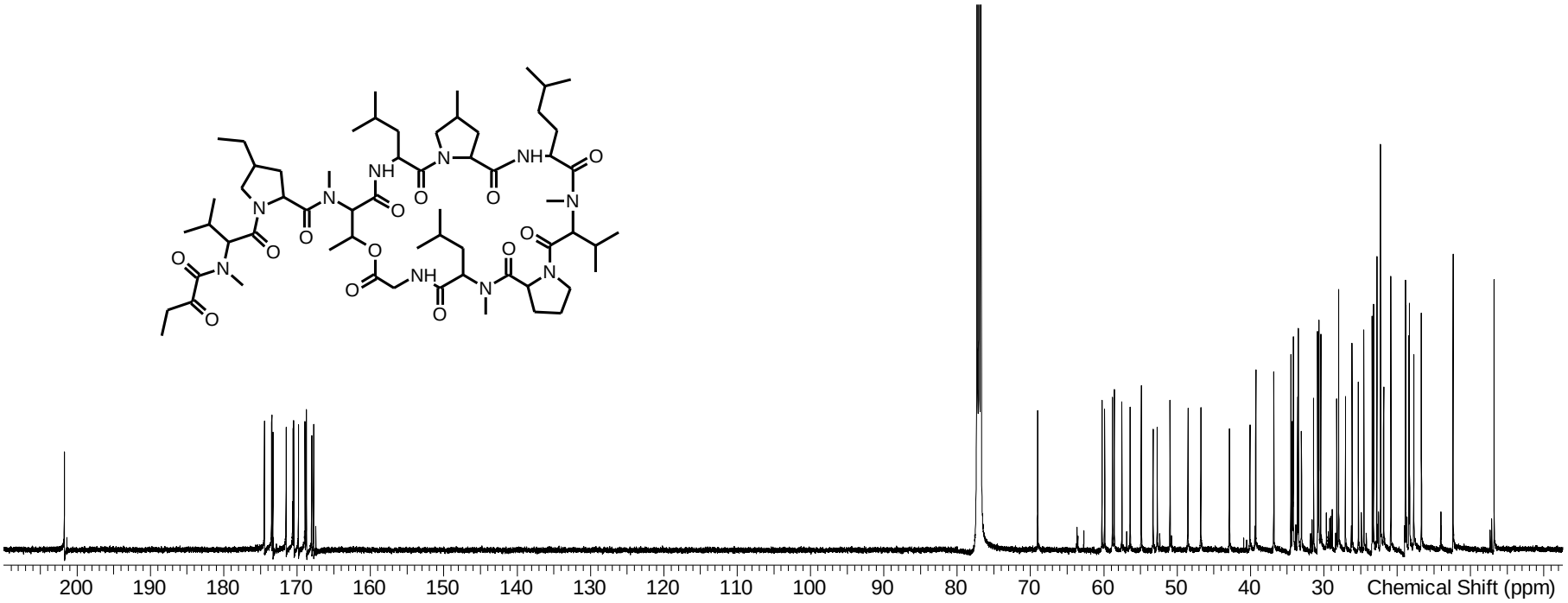
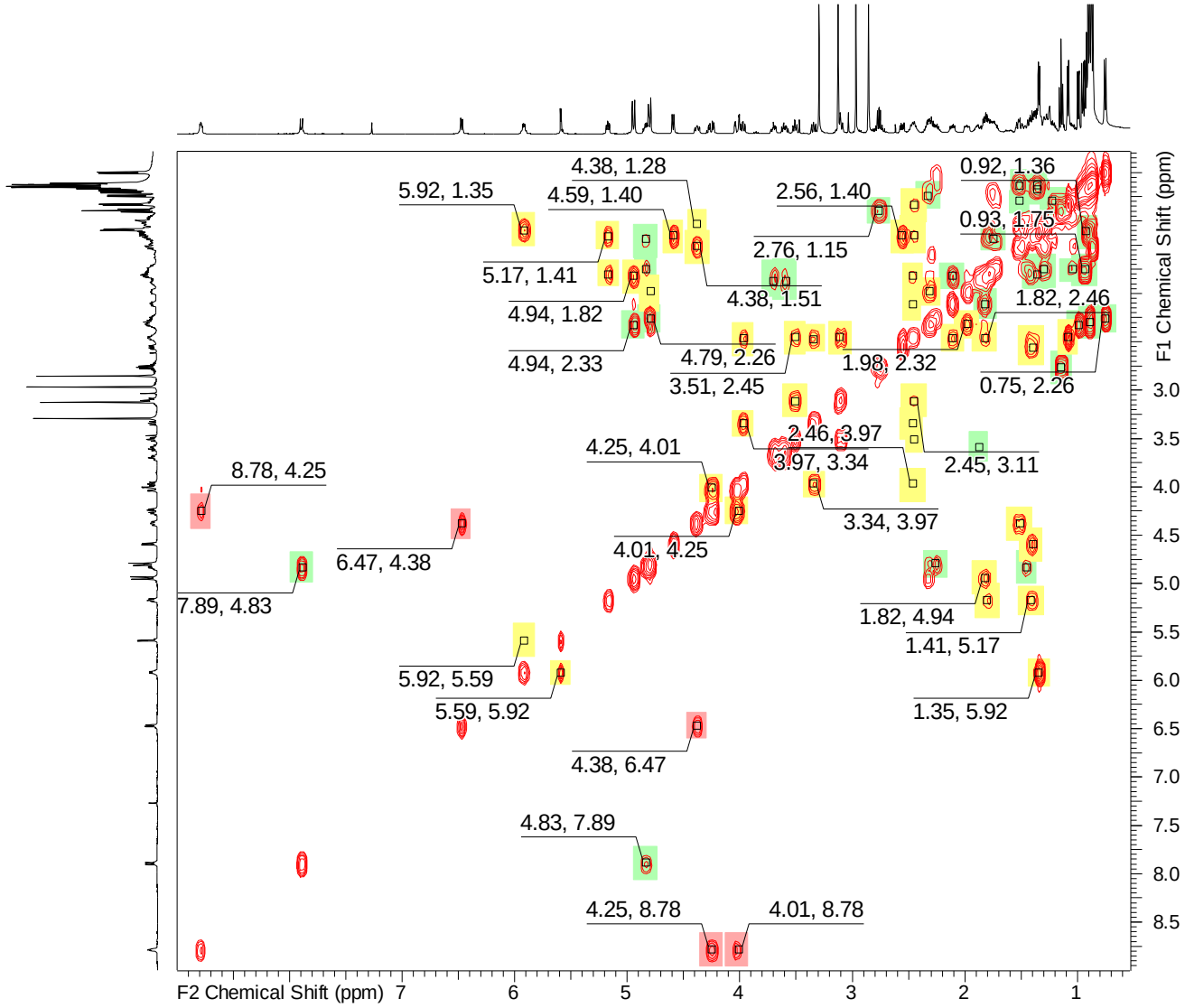


Figure S7: ^{13}C-spectrum of MP A in $CDCl_3$ at 125 MHz.



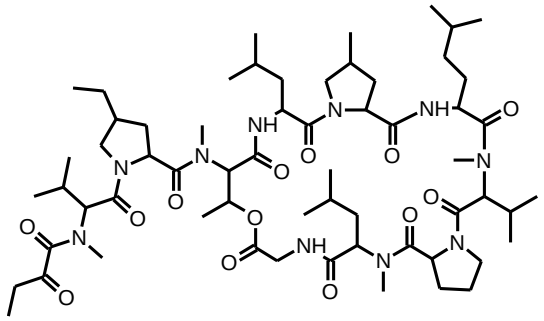


Figure S8: COSY-spectrum of MP A in $CDCl_3$ at 500 MHz.

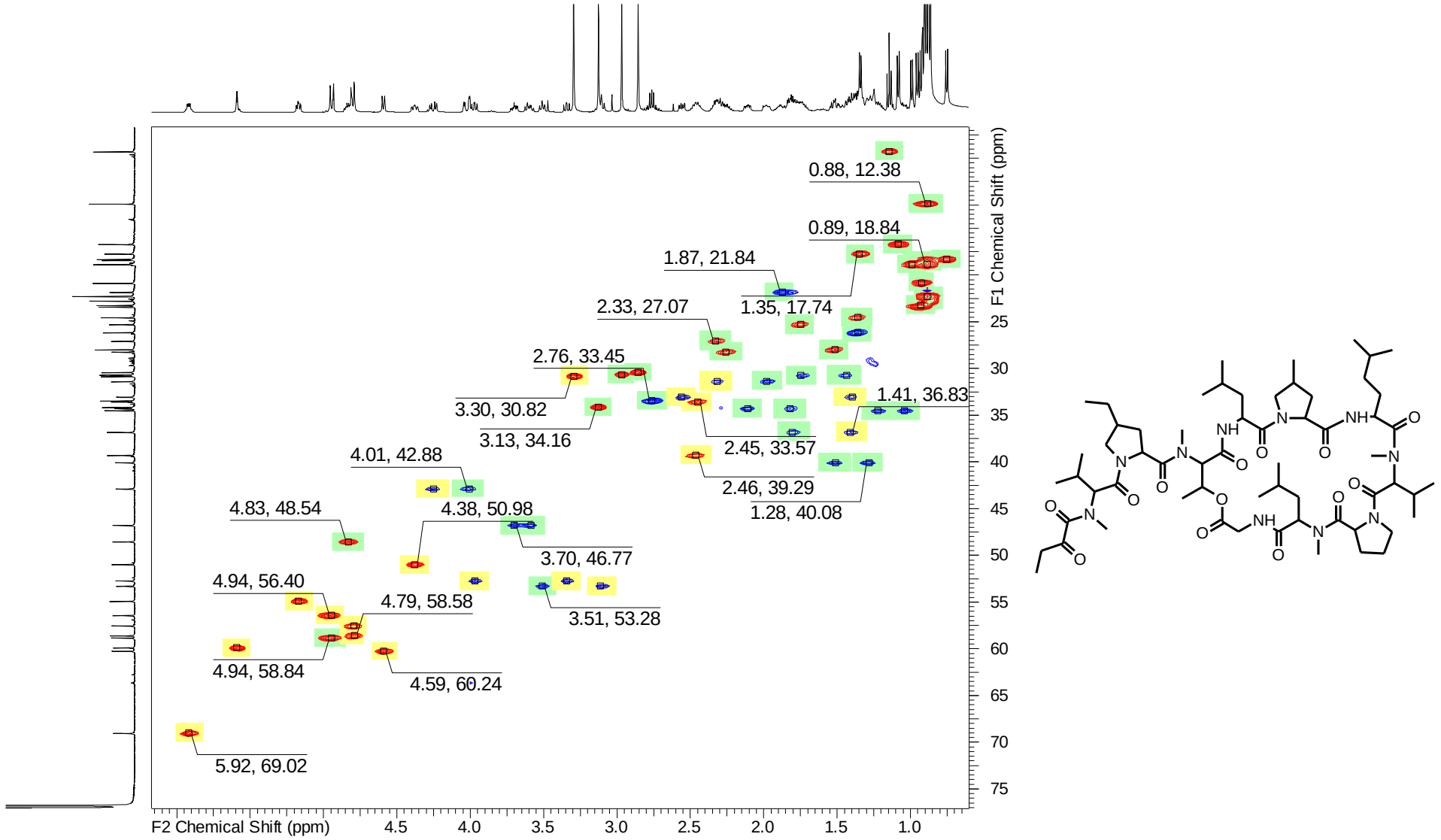


Figure S9: HSQC-spectrum of MP A in $CDCl_3$ at 500 MHz (^1H)/125 MHz (^{13}C).

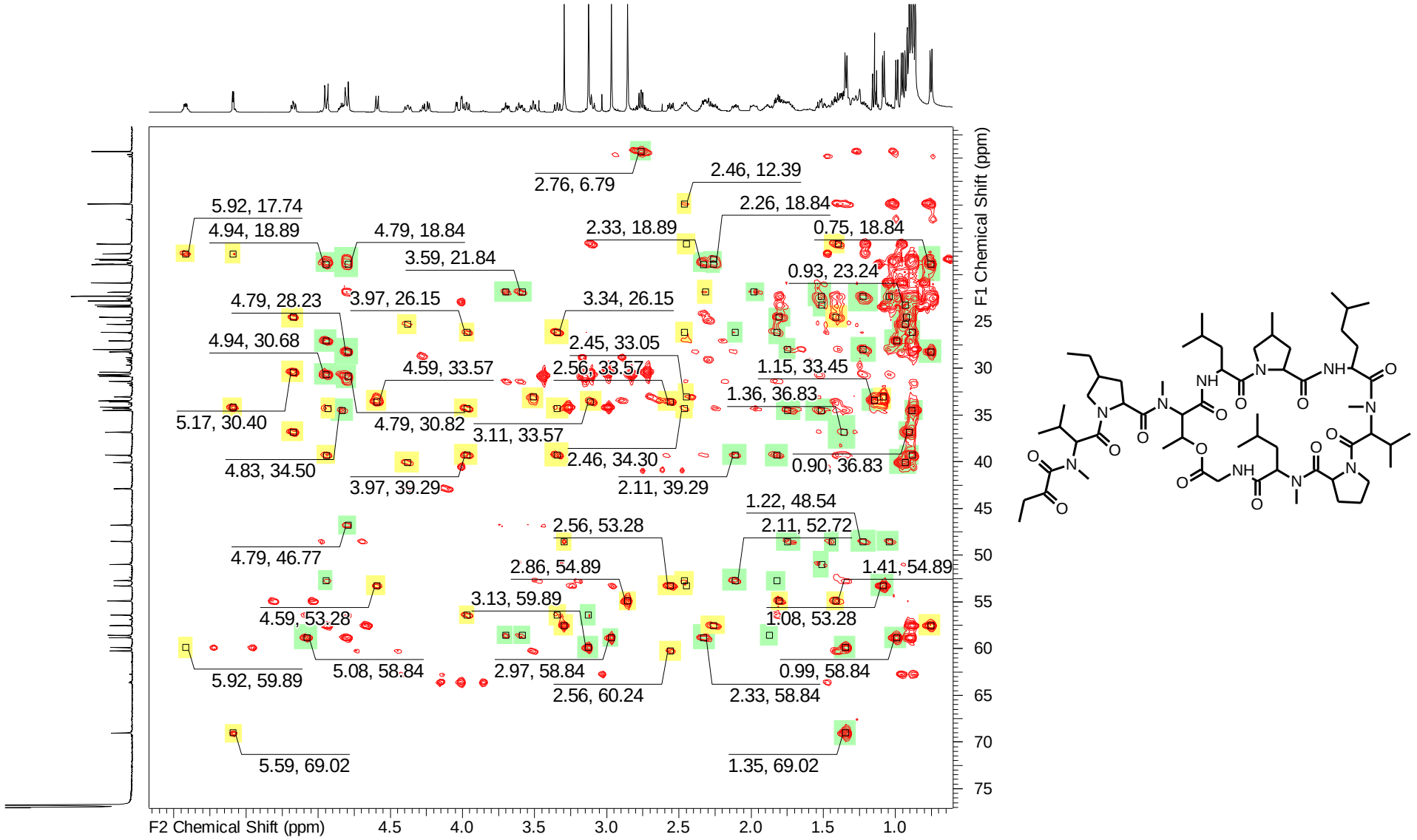
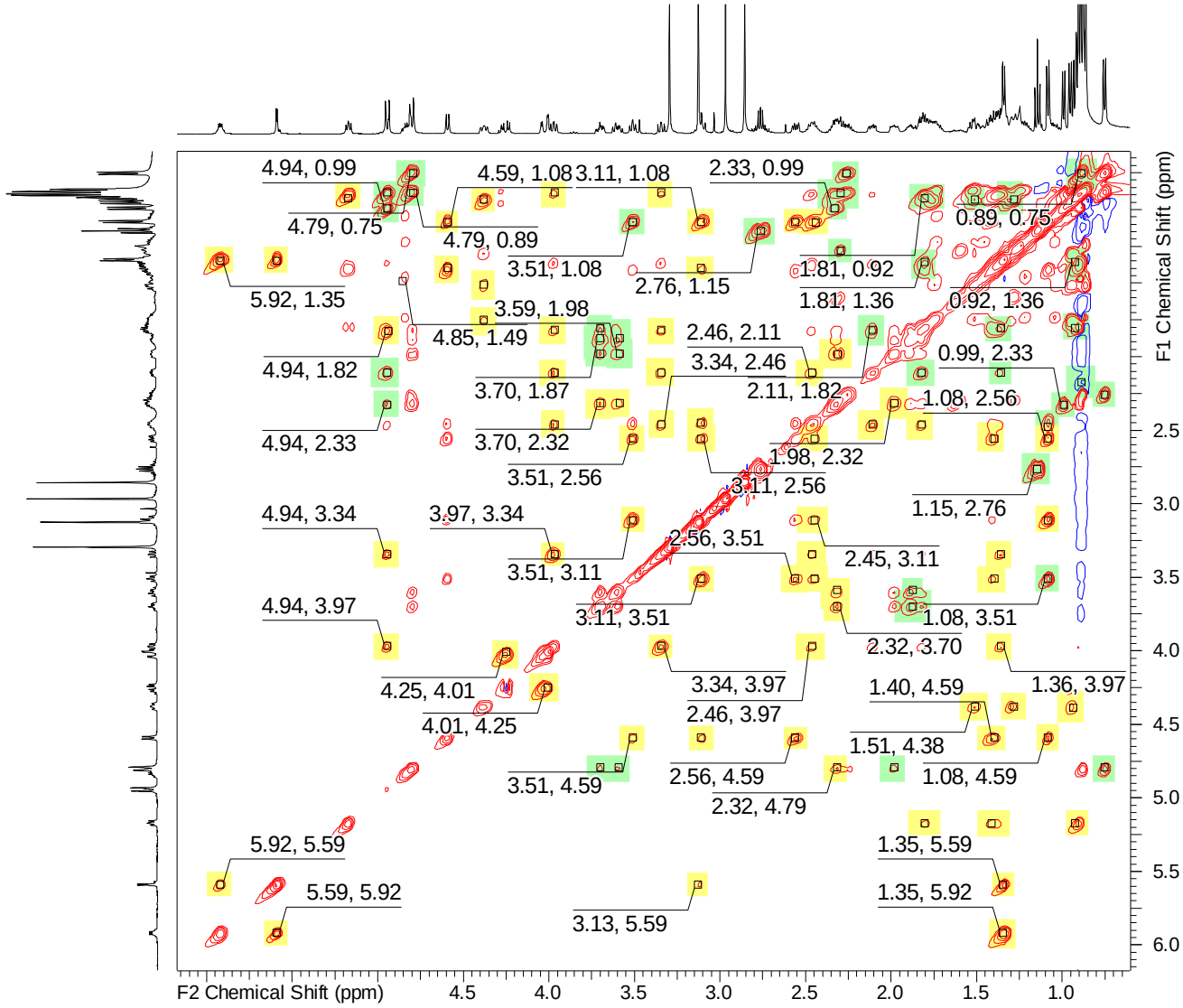


Figure S10: HMBC-spectrum of MP A in $CDCl_3$ at 500 MHz (^1H)/125 MHz (^{13}C).



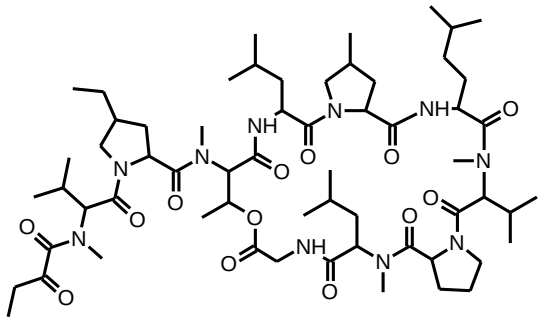


Figure S11: TOCSY-spectrum of MP A in $CDCl_3$ at 500 MHz.

Marfey's derivatization:

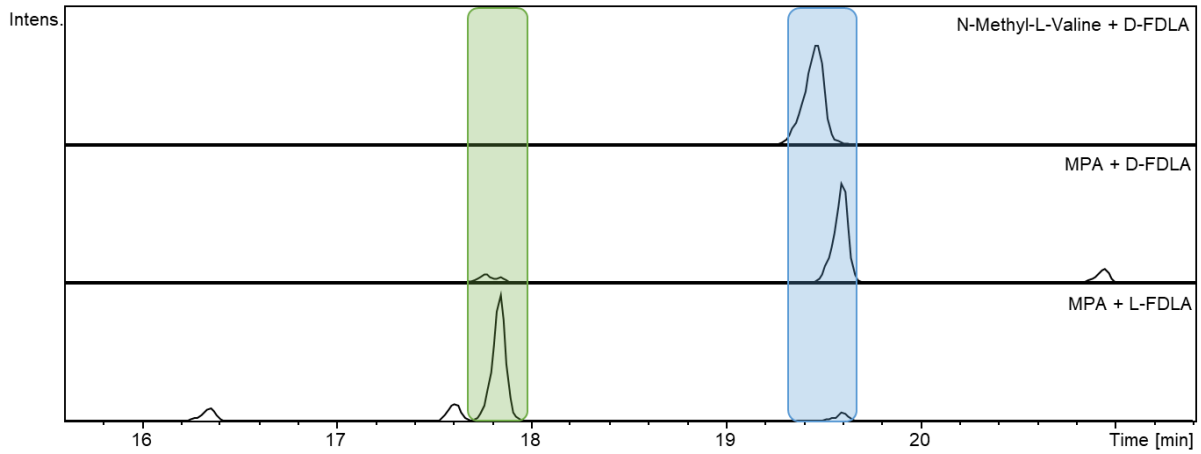


Figure S12: Marfey's derivatization of reference N-Methyl-L-Valine (upper chromatogram) with D-FDLA and MPA with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

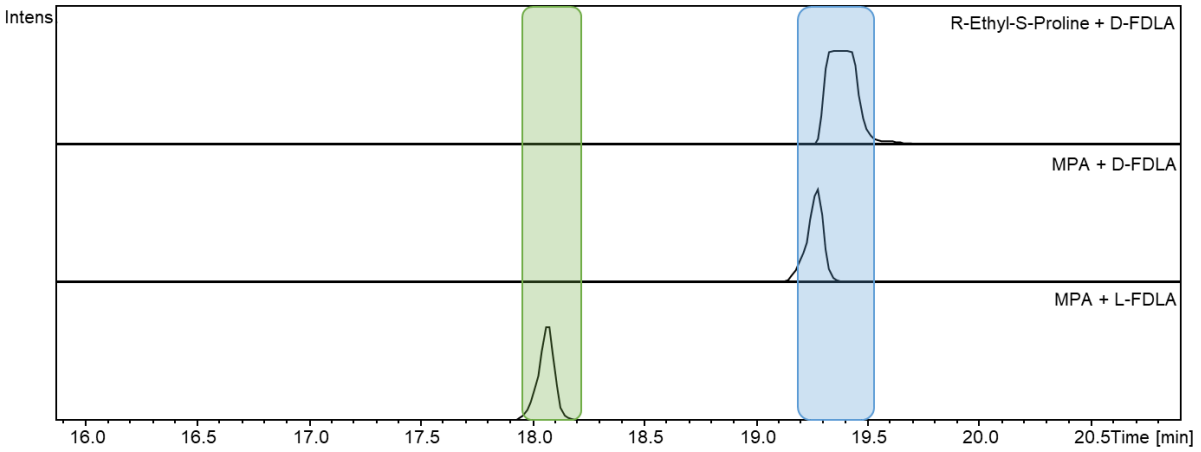


Figure S13: Marfey's derivatization of reference R-Ethyl-S-Proline (upper chromatogram) with D-FDLA and MPA with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

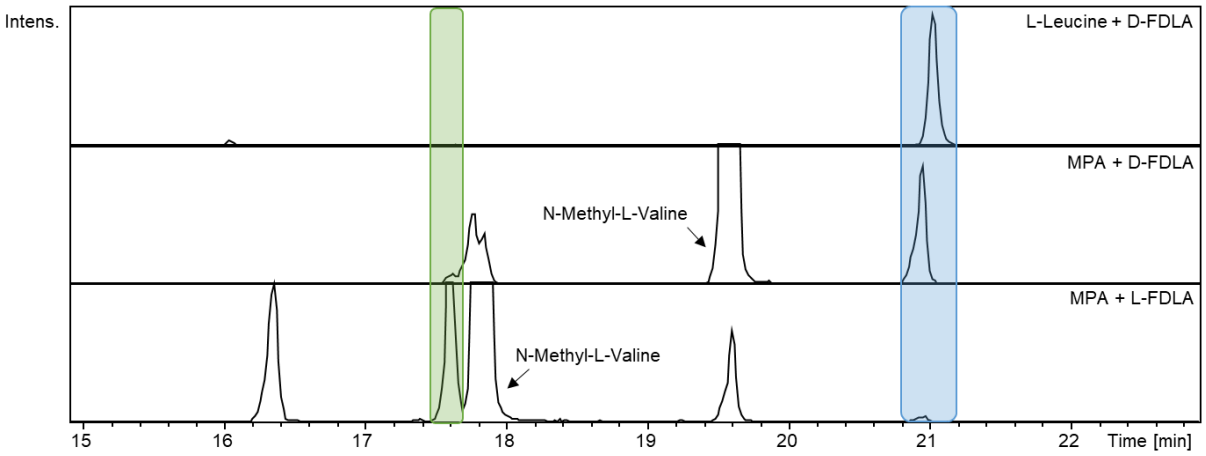


Figure S14: Marfey's derivatization of reference L-Leucine (upper chromatogram) with D-FDLA and MPA with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

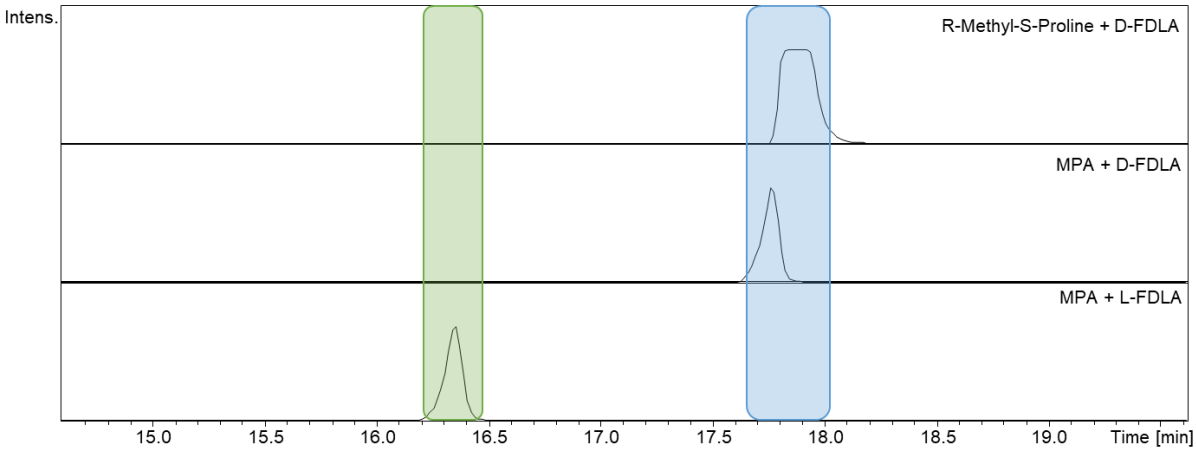


Figure S15: Marfey's derivatization of reference R-Methyl-S-Proline (upper chromatogram) with D-FDLA and MPA with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

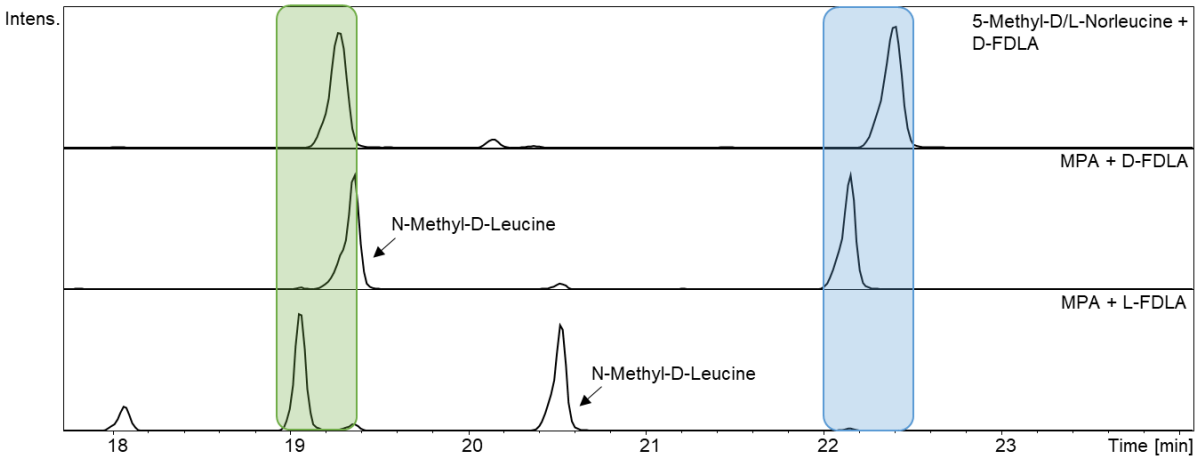


Figure S16: Marfey's derivatization of reference 5-Methyl-D/L-Norleucine (upper chromatogram) with D-FDLA and MPA with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

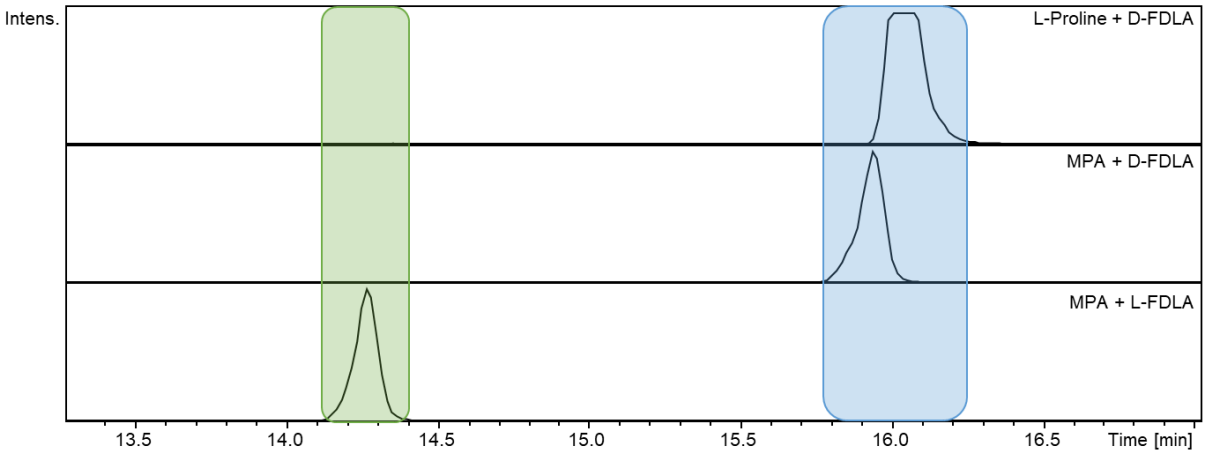


Figure S17: Marfey's derivatization of reference L-Proline (upper chromatogram) with D-FDLA and MPA with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

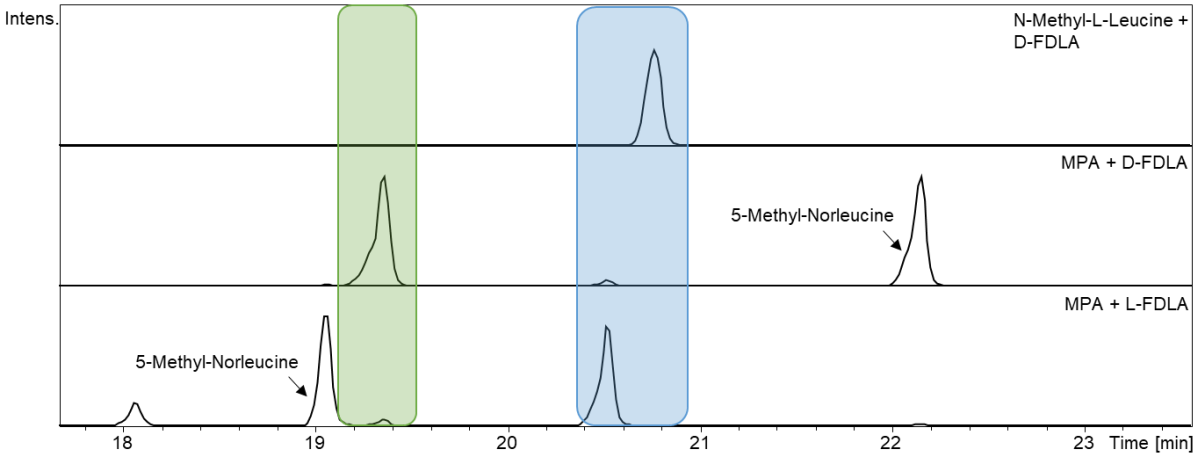


Figure S18: Marfey's derivatization of reference N-Methyl-L-Leucine with D-FDLA (upper chromatogram) and MPA with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

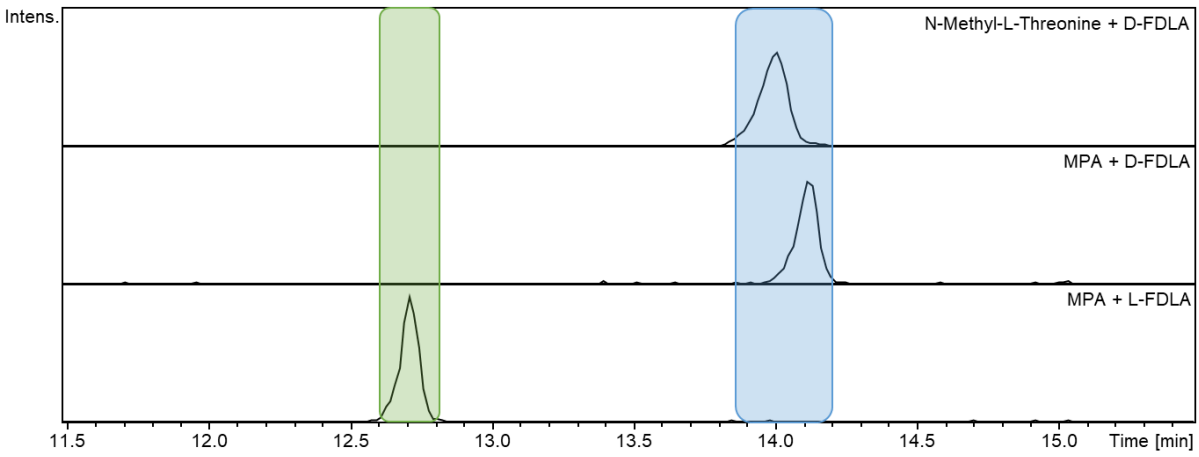


Figure S19: Marfey's derivatization of reference N-Methyl-L-Threonine with D-FDLA (upper chromatogram) and MPA with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

Mycoplanecin B

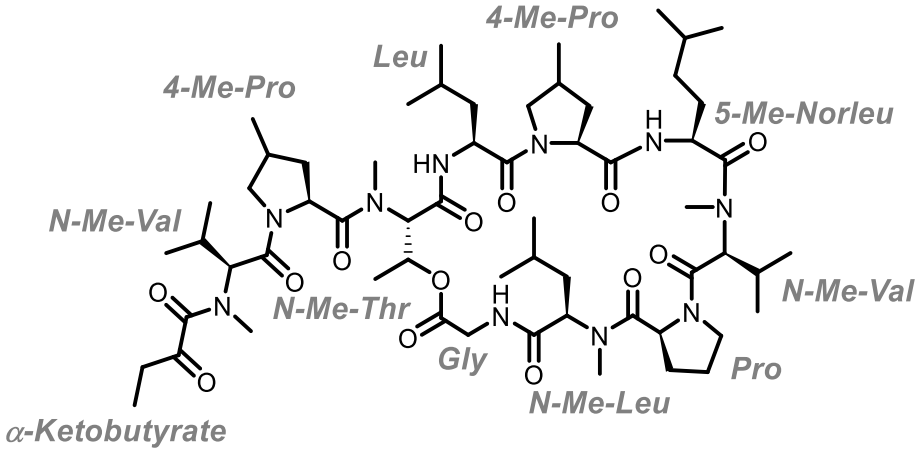


Table S10: NMR spectroscopic data of MP B in $CDCl_3$ at 500/125 MHz.

		NMR data in $CDCl_3$			
position	δ_C [ppm]	δ_H [ppm], mult. (*J* [Hz])	COSY correlations	HMBC correlations	TOCSY correlations
α-Keto-butyrate					
1	168.0	-	-	-	-
2	201.8	-	-	-	-
3	33.5	2.78, dd (7.52,6.51)	4	2,4	4
4	6.8	1.16, t (7.34)	3	2,3	3
N-Me-Val					
1	167.7	-	-	-	-
2	58.9	4.95, m	3	1,3,N-Me	4,5
3	27.1	2.34, m	2,4,5	1	4,5
4,5	18.9	1.01, dd (8.16,6.69)	3	2,3	2,3
N-Me	30.7	2.98, s	-	1 *α-Ketobutyrate,* 2	-
4-Me-Pro					
1	173.2	-	-	-	-
2	56.6	4.97, m	3	1,3,4	3,5,6

3	36.4	2.13,1.82, ddd (12.84, 6.51, 2.75)	2,4	1,2,4,5	2,6
4	32.2	2.64, m	3,5,6	5,6	-
5	54.3	3.96,3.31, dd (9.90,7.52)	4	2,3,4,6, 1 *N-Me-Val*	2,4,6
6	17.5	1.03, m	4	4,5	2,3,4,5
N-Me-Thr					
1	168.9	-	-	-	-
2	59.9	5.60, d (3.30)	3	1,3,4,N-Me,1 *4-Me-Pro*	3,4, N-Me
3	69.0	5.93, m	2,4	2,1 Gly	2,4
4	17.8	1.36, d (6.33)	3	1,2,3	2,3
N-Me	34.1	3.13, s	-	2, 1 *4-Me-Pro*,2 *4-Me-Pro*	2
Leu					
1	173.5	-	-	-	-
2	50.9	4.39, m	3,NH	1,3	NH
3	40.1	1.54, 1.31, m	2	-	5,6,NH
4	25.3	1.78, m	5,6	-	-
5,6	23.3	0.96, m	4	2,3	3,NH
NH	-	6.48, d (7.7)	2	2,3,1 *N-Me-Thr*	2,3,5,6
4-Me-Pro					
1	170.6	-	-	-	-
2	60.3	4.60, d (8.16)	3	1,4,5,1 *Leu*	3,5,6
3	33.1	2.57, 1.43, dd (12.01,6.05)	2,4	2,4,5,6	2,5,6
4	33.6	2.47, m	3,5,6	-	5,6
5	53.3	3.12, 3.52, m	4,6	2,3,4,6	2,3,4,6
6	16.7	1.09, d (6.42)	4,5	3,5	2,3,4,5
5-Me-Norleu					
1	174.5	-	-	-	-

2	48.6	4.84, m	3,NH	1,4,1 *4-Me-Pro*	6,7
3	30.9	1.76, 1.46, m	2,4	1,2,4	NH
4	34.5	1.24, 1.06, m	3	2,5,6,7	NH
5	28.0	1.54, m	-	4,6,7	-
6,7	22.3	0.90, m	-	-	2,NH
NH	-	7.91, d (9.35)	2	2,3,1 *4-Me-Pro*	2,3,4,6,7
N-Me-Val					
1	168.7	-	-	-	-
2	57.6	4.81, m	3	1,3,4, N-Me, 1 *5-Me-Norleu*	4,5
3	28.3	2.28, m	2,4,5	1,2,4,5	5
4	18.9	0.90, m	3	-	2,3,5
5	18.4	0.77, d (6.69)	3	1,2,3,4	2,3,4
N-Me	30.8	3.31, s	-	2,1 *5-Me-Norleu*	-
Pro					
1	169.8	-	-	-	-
2	58.6	4.80, m	3	4	5
3	31.4	2.33, 1.99, m	2	1,5	2,5
4	21.9	1.88, m	5	-	5
5	46.8	3.71, 3.61, m	4	2,3,4, 1 *N-Me-Val*	2,3,4
N-Me-Leu					
1	171.5	-	-	-	-
2	54.9	5.18, dd (8.67, 6.01)	3	1,3,4,N-Me,1 *Pro*	3,5,6
3	36.9	1.82, 1.42, m	2	1,2,4	2,5,6
4	24.6	1.39, m	2	2,3	5,6
5,6	20.9	0.94, m	-	3,4	2,3,4
N-Me	30.4	2.87, s	-	2,1 *Pro*	-
Gly					

1	170.6	-	-	-	-
2	42.9	4.26, 4.02, m	NH	1,1 *N-Me-Leu*	NH
NH	-	8.80, dd (6.24, 3.85)	2	1,2,1 *N-Me-Leu*	2

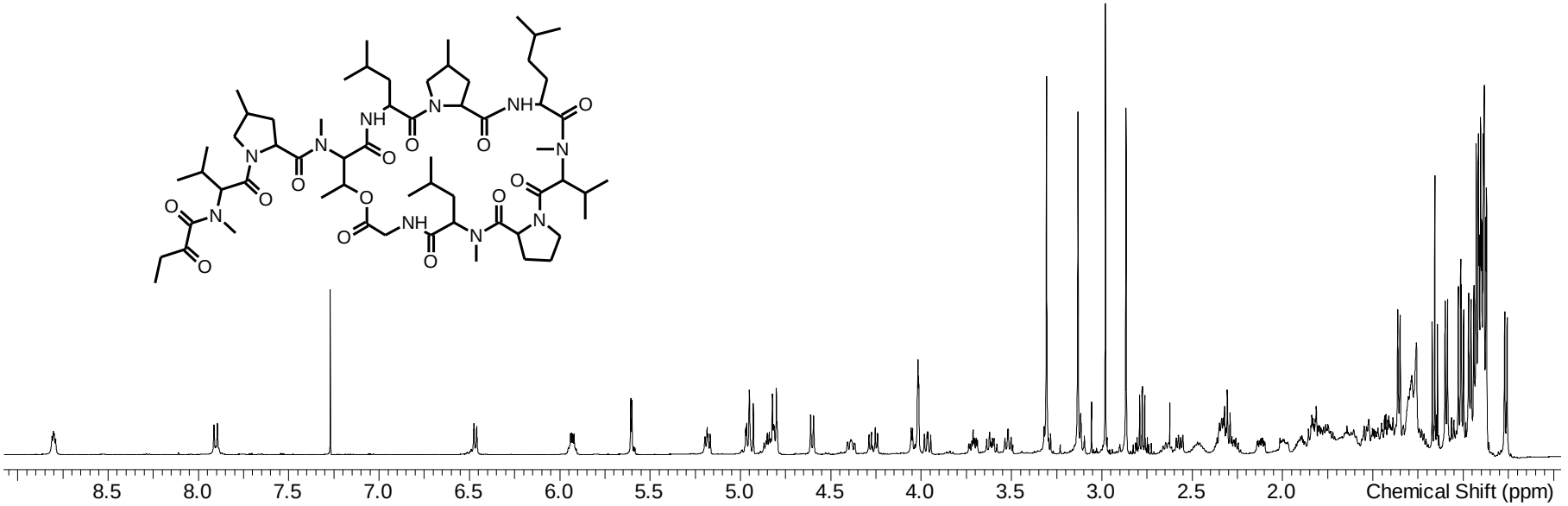


Figure S20: ^{1}H-spectrum of MP B in $CDCl_3$ at 500 MHz.

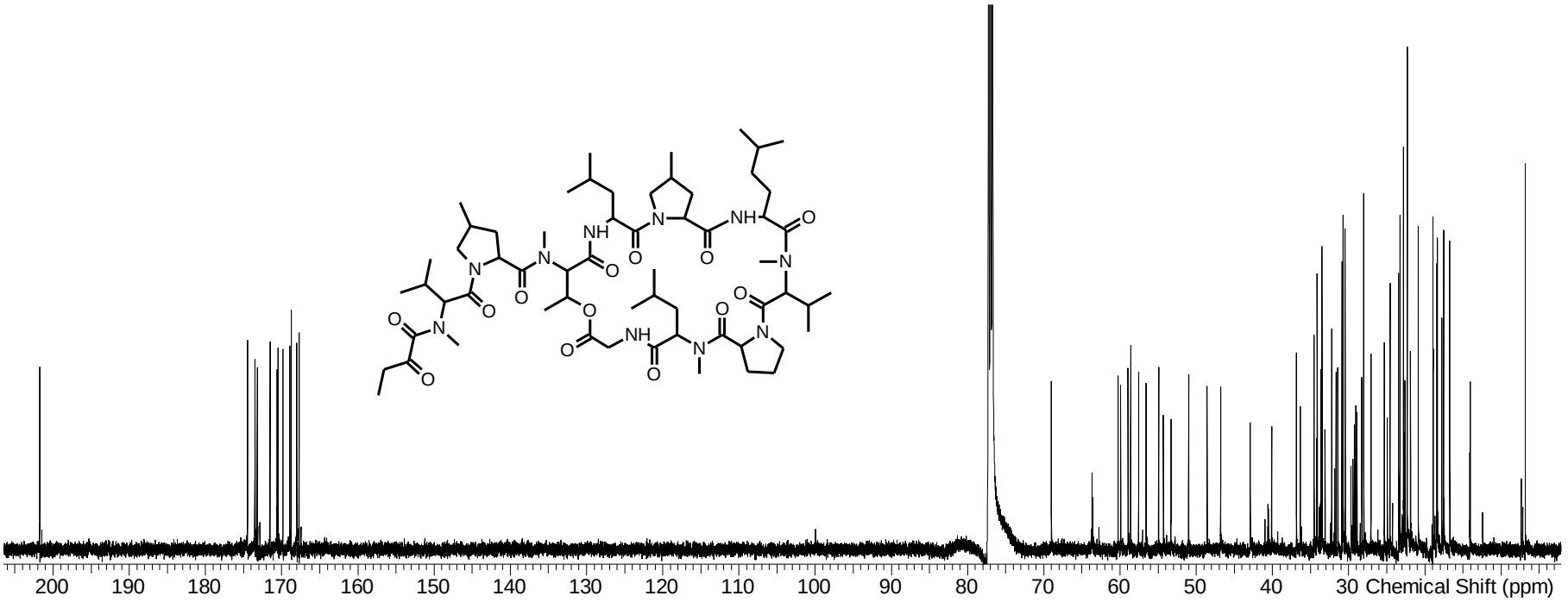


Figure S21: ^{13}C-spectrum of MP B in $CDCl_3$ at 125 MHz.

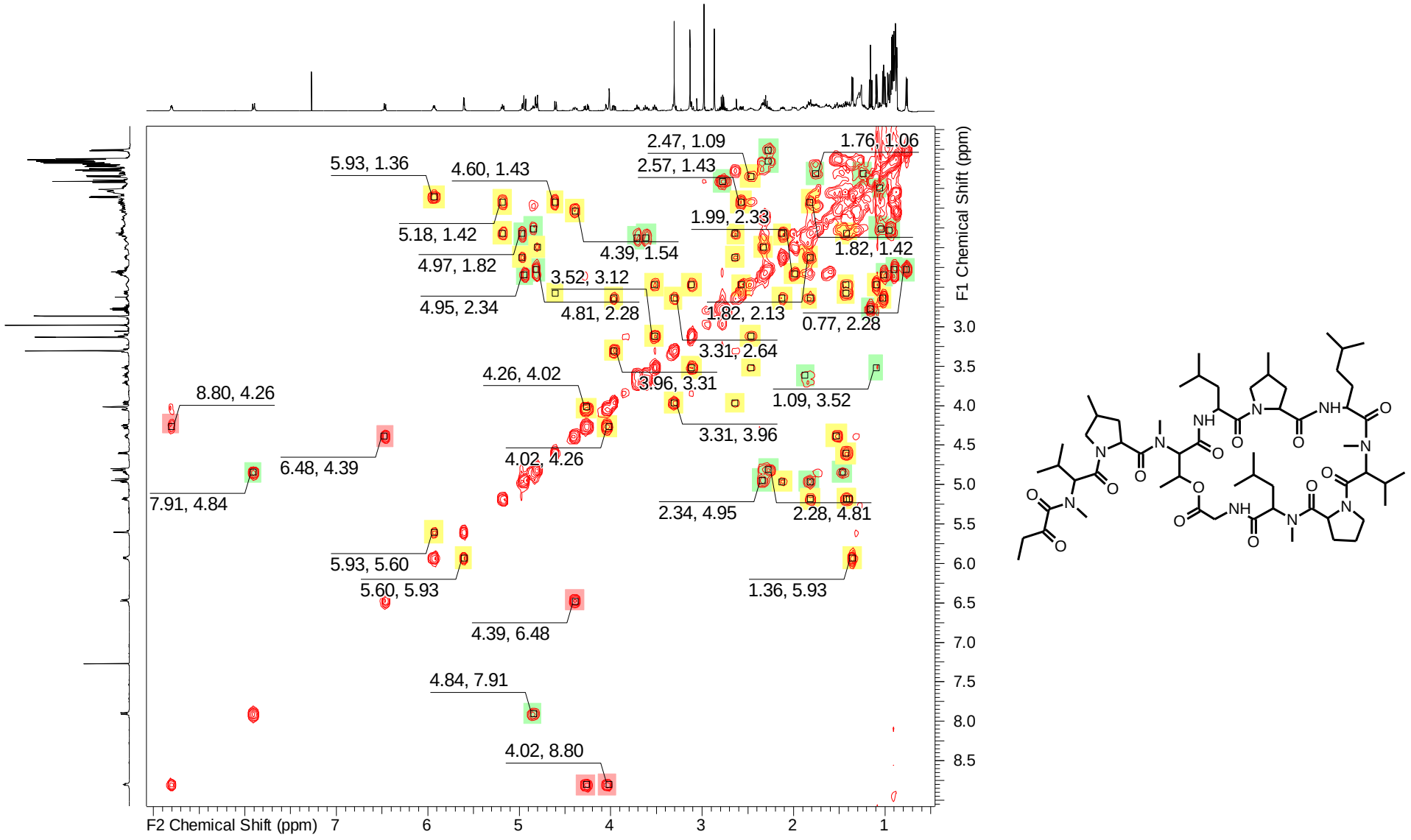


Figure S22: COSY-spectrum of MP B in $CDCl_3$ at 500 MHz.

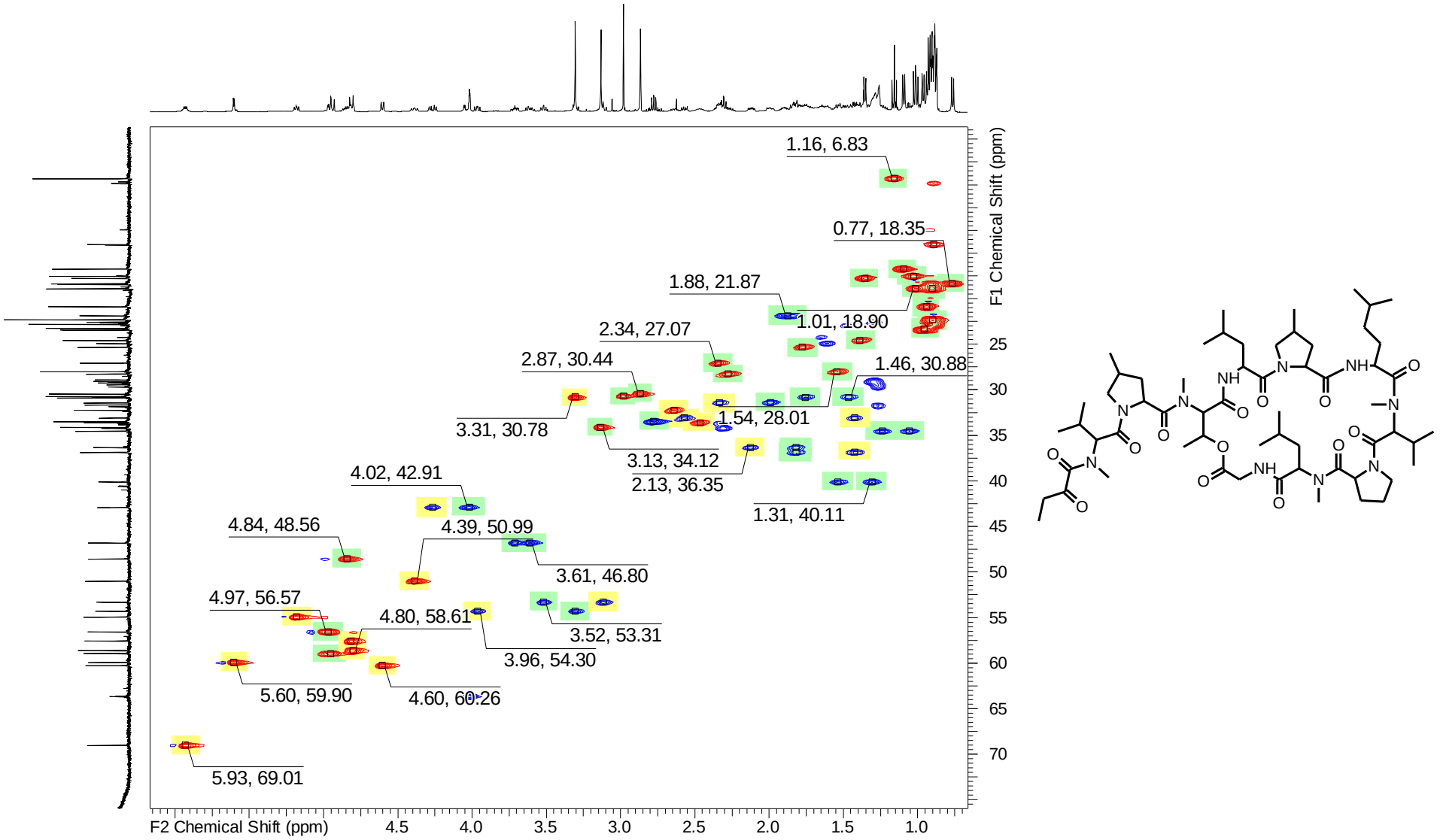


Figure S23: HSQC-spectrum of MP B in $CDCl_3$ at 500 MHz (^1H)/125 MHz (^{13}C).

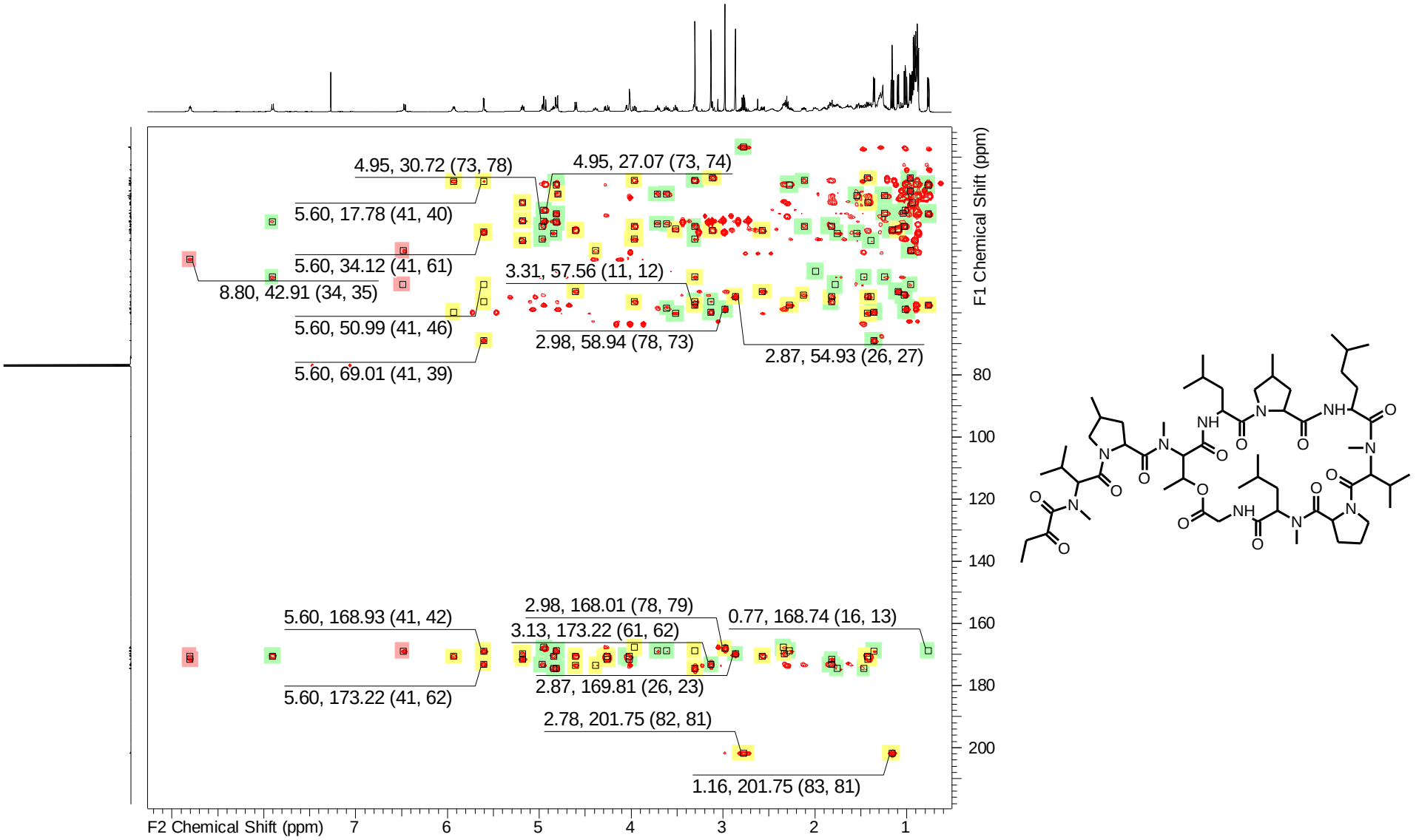


Figure S24: HMBC-spectrum of MP B in $CDCl_3$ at 500 MHz (^1H)/125 MHz (^{13}C).

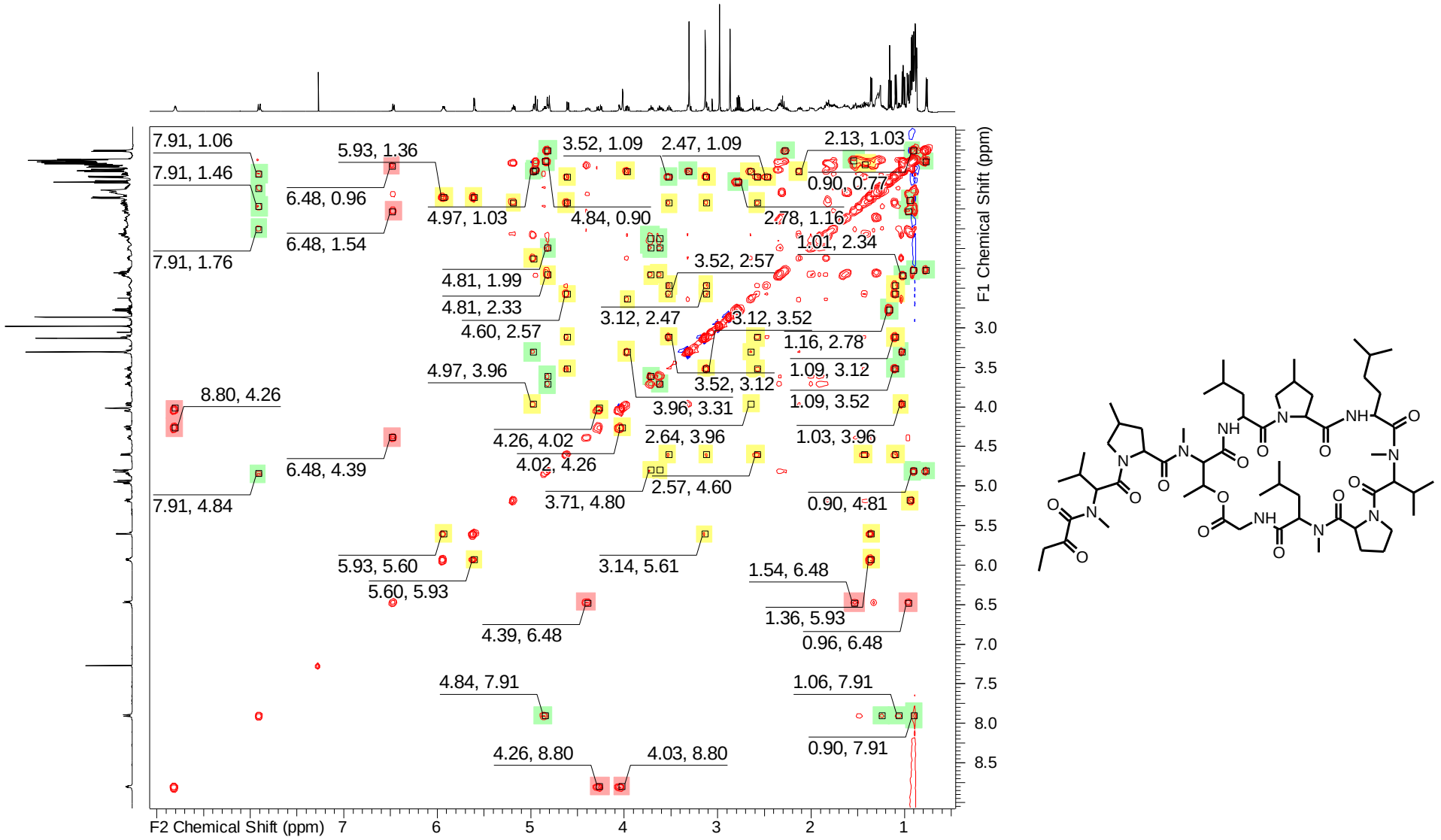


Figure S25: TOCSY-spectrum of MP B in $CDCl_3$ at 500 MHz.

Marfey's derivatization:

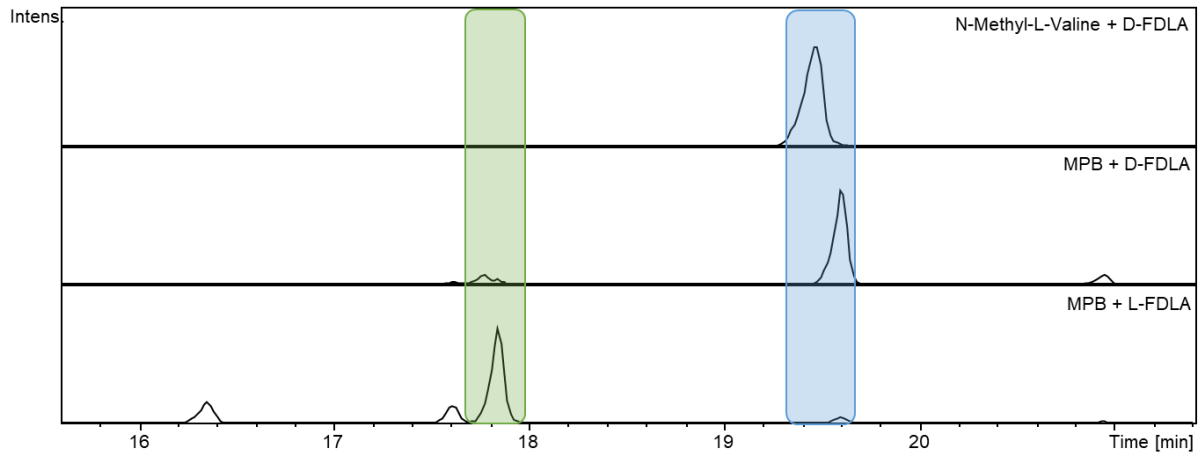


Figure S26: Marfey's derivatization of reference N-Methyl-L-Valine with D-FDLA (upper chromatogram) and MPB with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

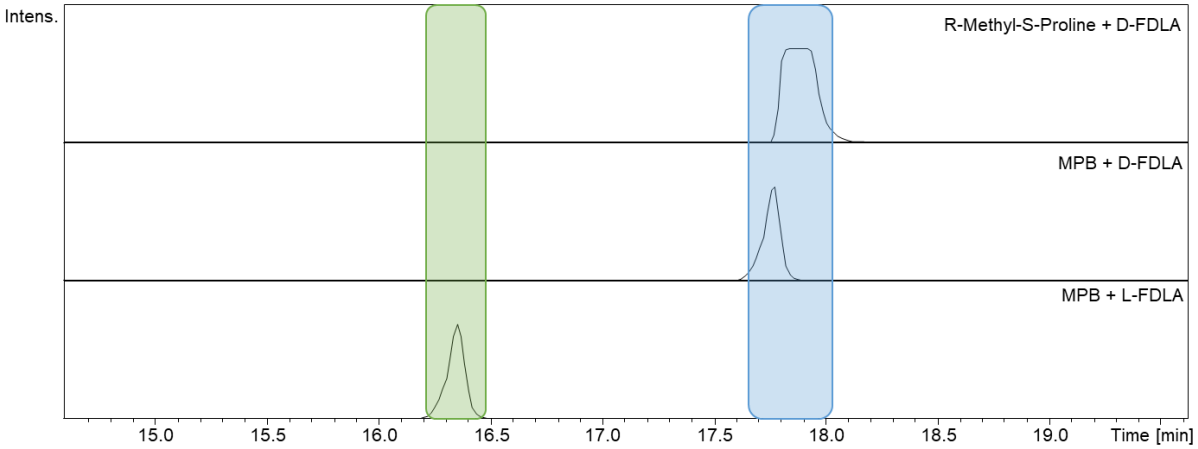


Figure S27: Marfey's derivatization of reference R-Methyl-S-Proline with D-FDLA (upper chromatogram) and MPB with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

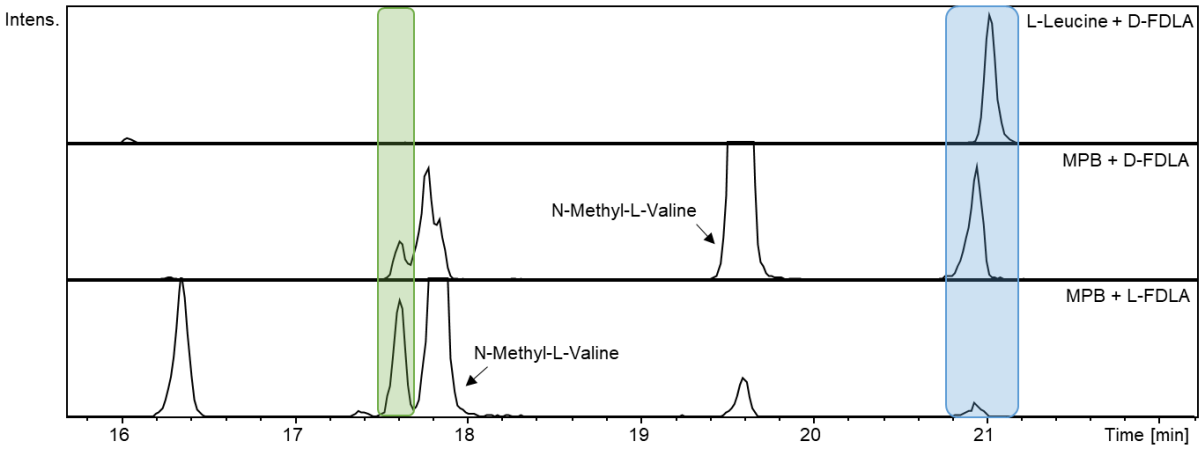


Figure S28: Marfey's derivatization of reference L-Leucine with D-FDLA (upper chromatogram) and MPB with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

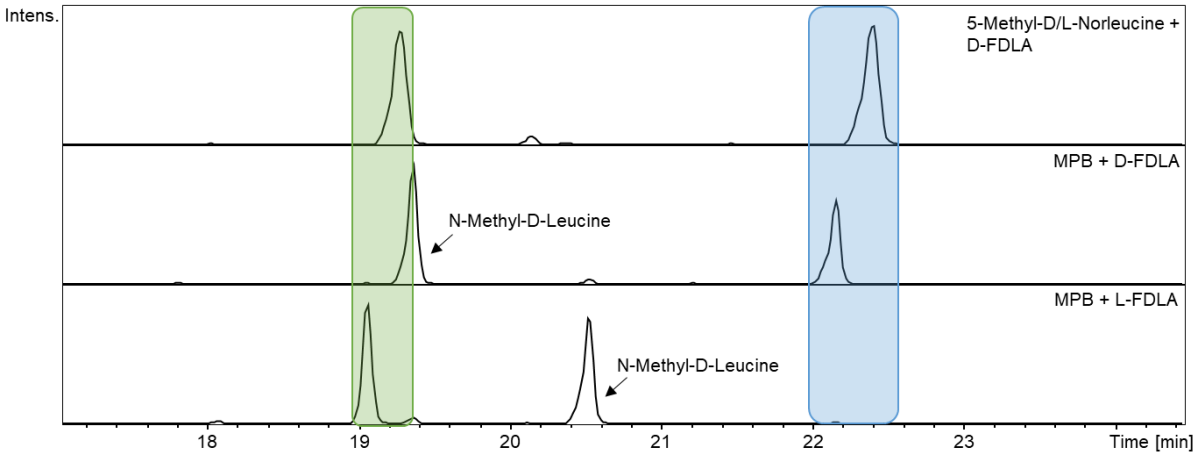


Figure S29: Marfey's derivatization of reference 5-Methyl-D/L-Norleucine with D-FDLA (upper chromatogram) and MPB with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

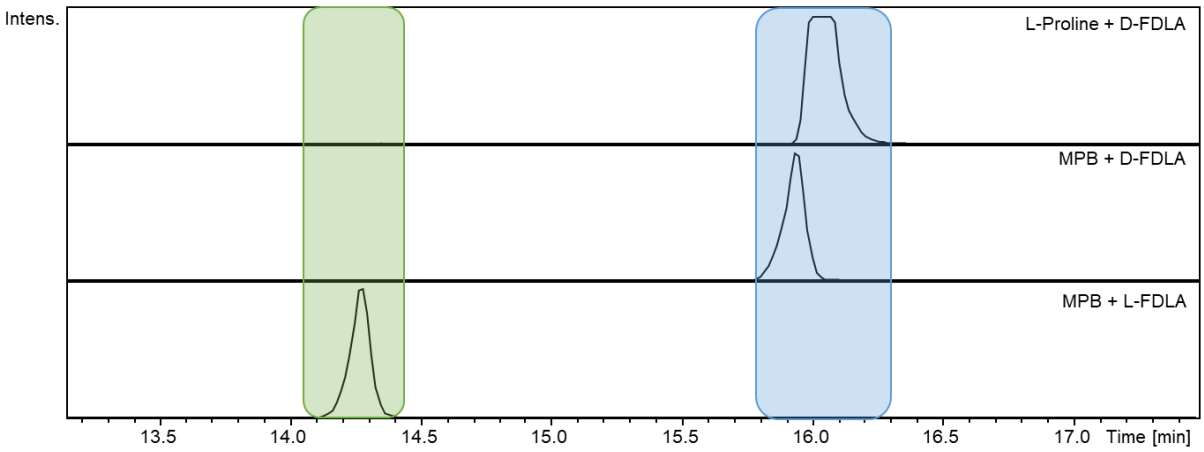


Figure S30: Marfey's derivatization of reference L-Proline with D-FDLA (upper chromatogram) and MPB with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

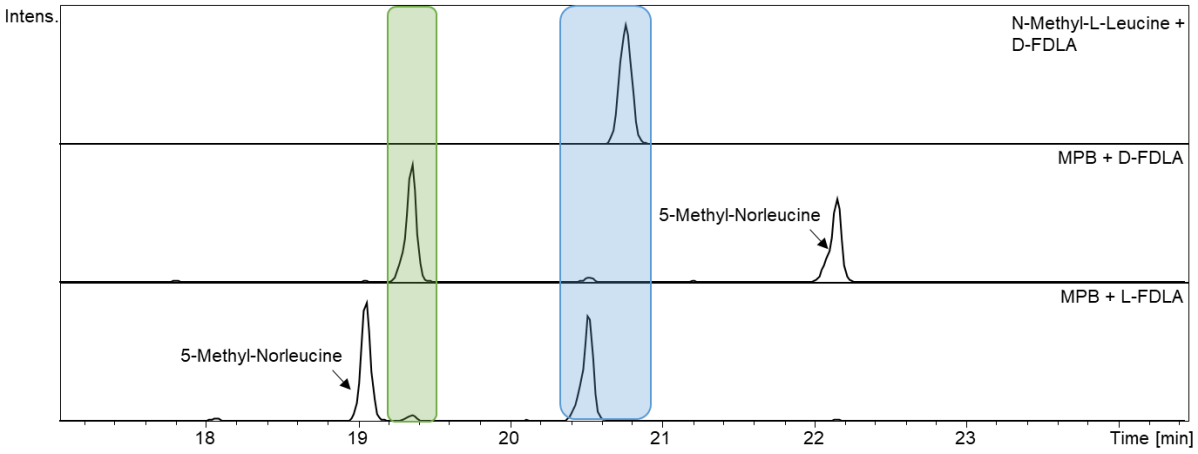


Figure 31: Marfey's derivatization of reference N-Methyl-L-Leucine with D-FDLA (upper chromatogram) and MPB with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

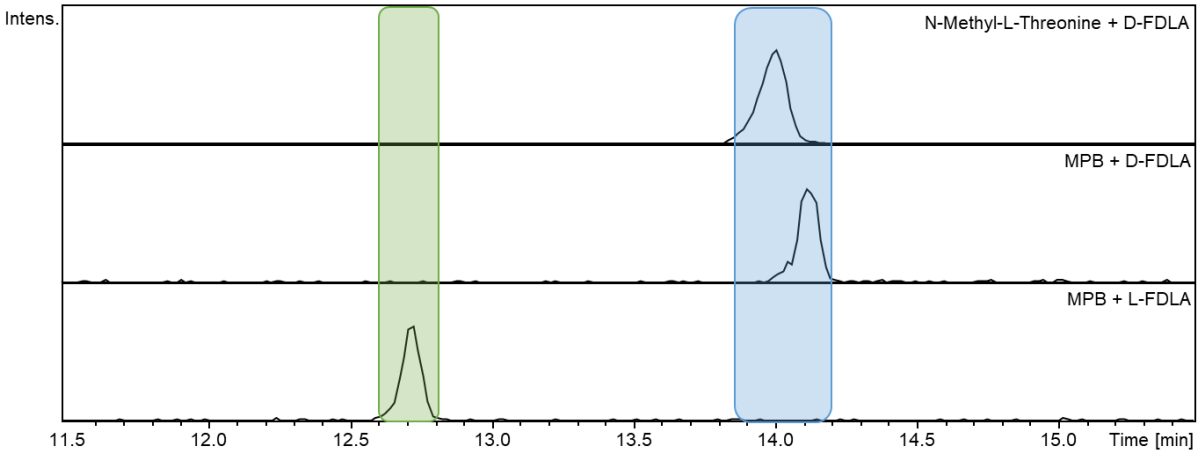


Figure S32: Marfey's derivatization of reference N-Methyl-L-Threonine with D-FDLA (upper chromatogram) and MPB with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

Mycoplanecin D

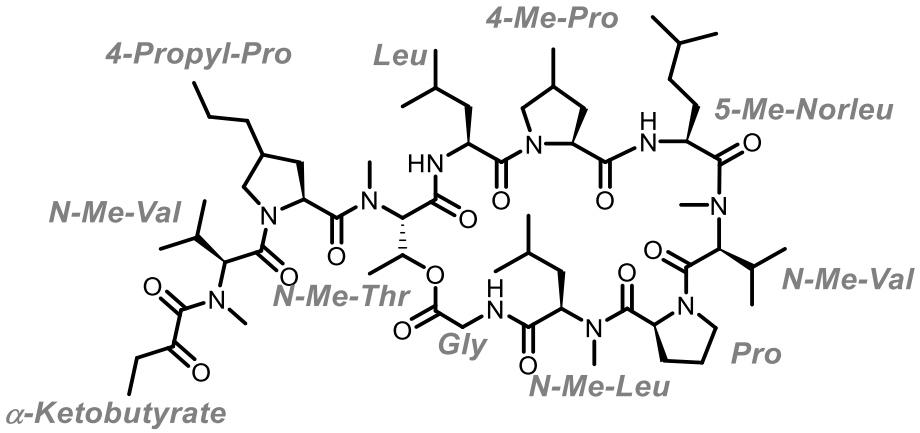


Table S11: NMR spectroscopic data of MP D in $CDCl_3$ at 500/125 MHz.

		NMR data in $CDCl_3$			
position	δ_C [ppm]	δ_H [ppm], mult. (*J* [Hz])	COSY correlations	HMBC correlations	TOCSY correlations
α-Keto-butyrate					
1	168.1	-	-	-	-
2	201.8	-	-	-	-
3	33.6	2.78, m	4	-	4
4	6.8	1.15, m	3	2,3	3
N-Me-Val					
1	167.8	-	-	-	-
2	58.6	4.95, m	3	1,3,4,5	4,5
3	27.1	2.32, m	2,4,5	-	-
4,5	18.9	1.00, m	3	3	2
N-Me	30.6	2.99, m	-	1 *α-Ketobutyrate*	-
4-Propyl-Pro					
1	173.5	-	-	-	-
2	56.4	4.95, m	3	-	3
3	34.2	2.13,1.83, m	2,4	1	2,6
4	39.3	2.47, m	3,5	-	-
5	54.2	3.93,3.26, m	4	4	-
6	31.7	1.26, m	-	-	3

7	22.6	1.29, m	8	6,8	-
8	14.1	0.88, m	7	4	-
N-Me-Thr					
1	168.9	-	-	-	-
2	59.9	5.60, m	-	1,3,N-Me,1 *4-Propyl-Pro*	3,4
3	69.0	5.94, m	4	1 Gly	2,4
4	17.8	1.36, m	3	2,3	2,3
N-Me	34.2	3.15, m	-	2, 1 *4-Propyl-Pro*	-
Leu					
1	173.5	-	-	-	-
2	50.9	4.39, m	3,NH	-	NH,3
3	40.1	1.52, 1.31, m	2	-	5,6,NH
4	24.9	1.78, m	5,6	-	-
5,6	23.3	0.95, m	4	3,4	2,3
NH	-	6.49, m	2	-	2,3
4-Me-Pro					
1	170.5	-	-	-	-
2	60.2	4.59, m	3	1,4,1 *Leu*	3,5
3	33.0	2.56, 1.42, dd (12.33,6.1)	2	1,4	2,5,6
4	33.7	2.47, m	5,6	-	-
5	53.3	3.11, 3.52, m	4	4,6	2,3,6
6	16.7	1.10, d (6.60)	5	3	3,5
5-Me-Norleu					
1	174.5	-	-	-	-
2	48.6	4.85, m	3,NH	1	NH
3	30.7	1.74, 1.47, m	2,4,6,7	-	6,7,NH
4	34.5	1.23, 1.04, m	3	5,6,7	NH
5	28.0	1.54, m	-	-	-
6,7	22.3	0.88, m	3	4,5	3,NH
NH	-	7.91, d (9.58)	2	1 *4-Me-Pro*	2,3,4,6,7
N-Me-Val					

1	168.9	-	-	-	-
2	57.6	4.82, m	3	1,3,5, N-Me, 1 *5-Me-Norleu*	4,5
3	28.3	2.26, m	2,4,5	-	5
4	18.9	0.89, m	3	2,5	2,3,5
5	18.4	0.77, m	3	2,3	2,3,4
N-Me	30.9	3.31, s	-	2,1 *5-Me-Norleu*	-
Pro					
1	169.9	-	-	-	-
2	58.5	4.80, m	3	3,4	3,4,5
3	31.4	2.31, 1.98, m	2	-	2
4	21.8	1.88, m	5	-	2,5
5	46.8	3.72, 3.61, m	4	3,4	2,4
N-Me-Leu					
1	171.5	-	-	-	-
2	54.9	5.18, m	3	1,3,4,N-Me,1 *Pro*	3,5,6
3	36.9	1.82, 1.42, m	2	1,2,4	2,5,6
4	24.6	1.37, m	5,6	-	-
5,6	20.9	0.93, m	4	3,4	2,3
N-Me	30.4	2.87, m	-	2,1 *Pro*	-
Gly					
1	170.5	-	-	-	-
2	42.9	4.28, 4.04, m	NH	1,1 *N-Me-Leu*	NH
NH	-	8.80, m	2	2	2

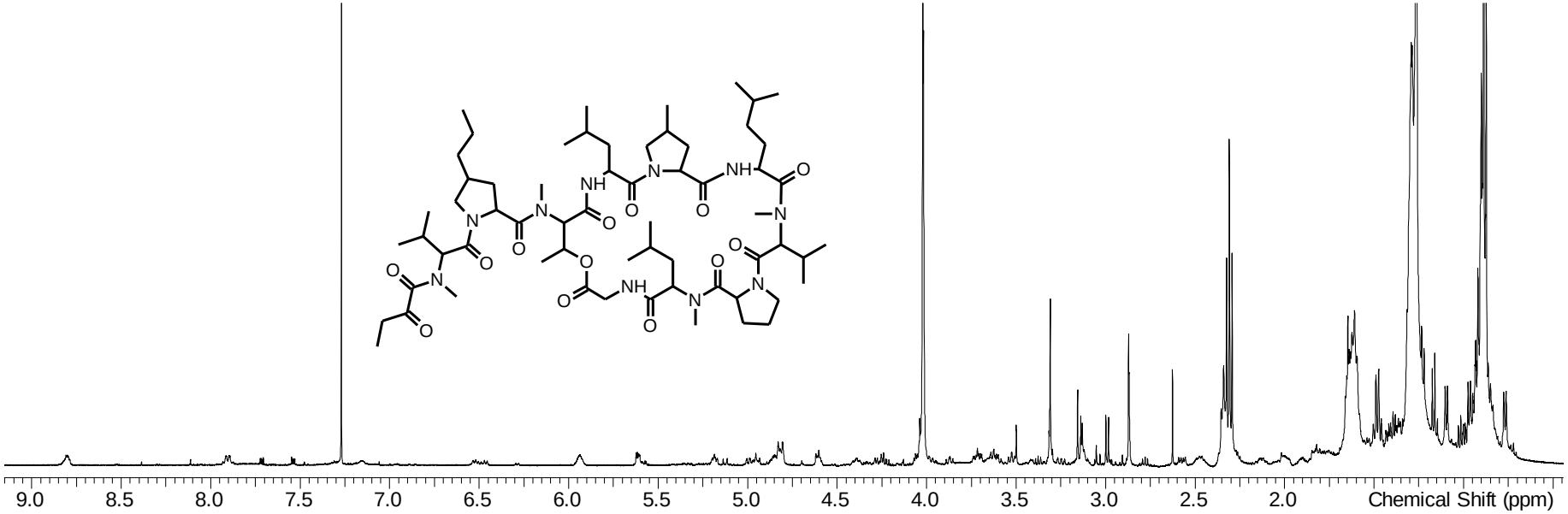


Figure S33: ^{1}H-spectrum of MP D in $CDCl_3$ at 500 MHz.

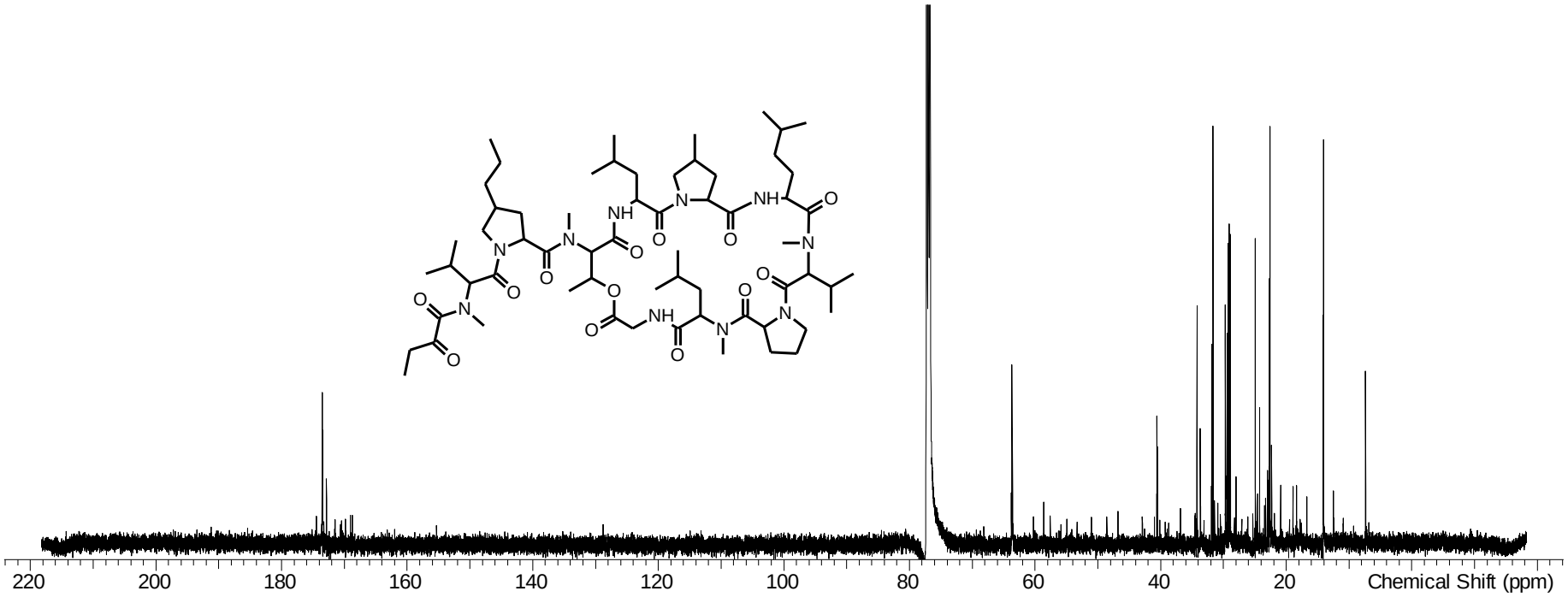
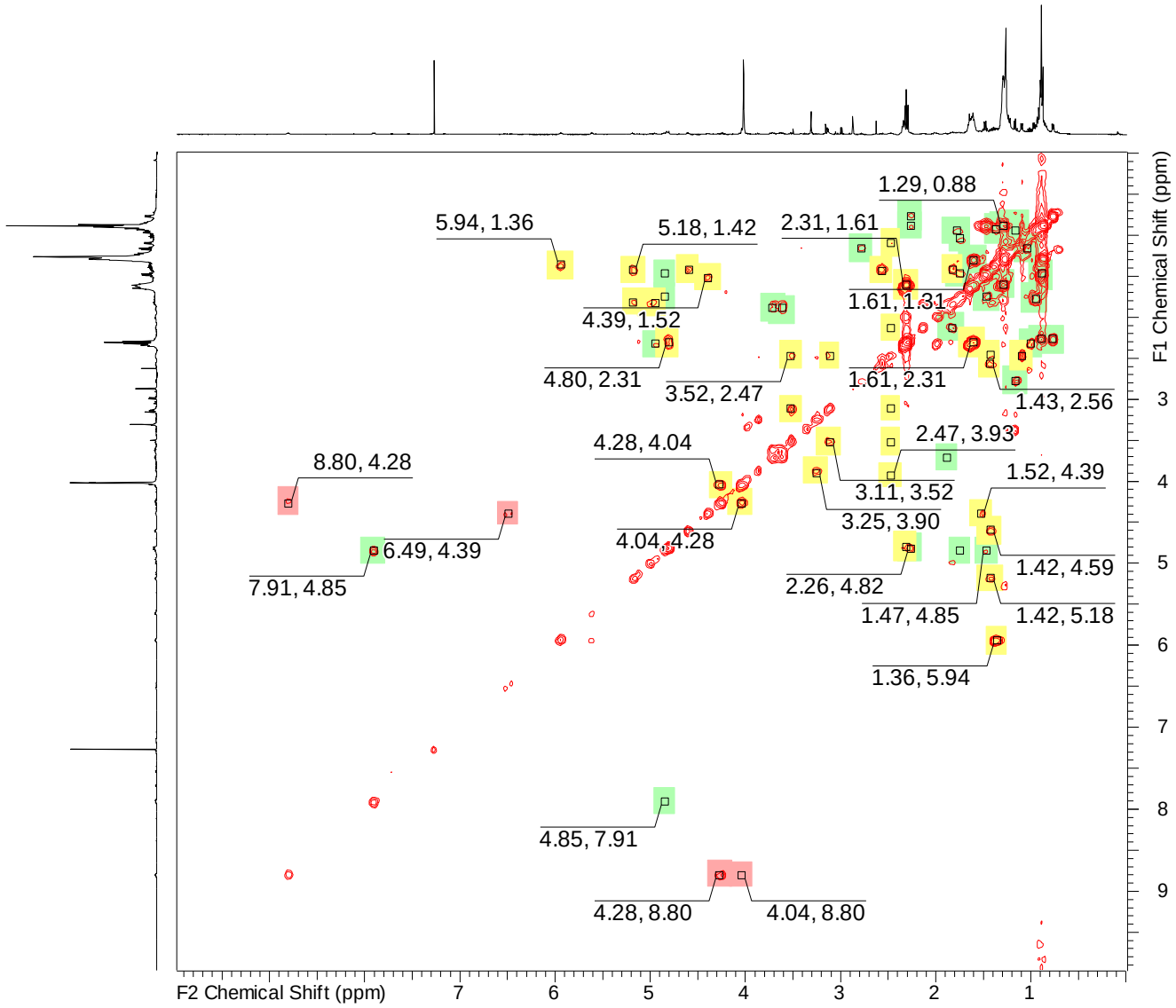


Figure S34: ^{13}C-spectrum of MP D in $CDCl_3$ at 125 MHz.



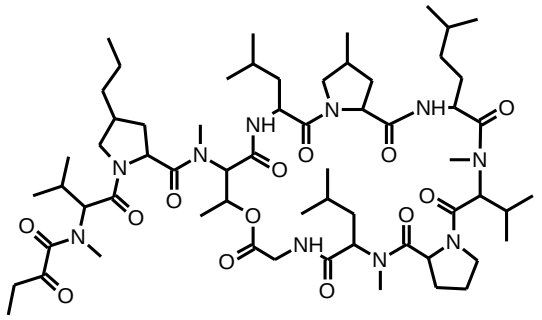


Figure S35: COSY-spectrum of MP D in $CDCl_3$ at 500 MHz.

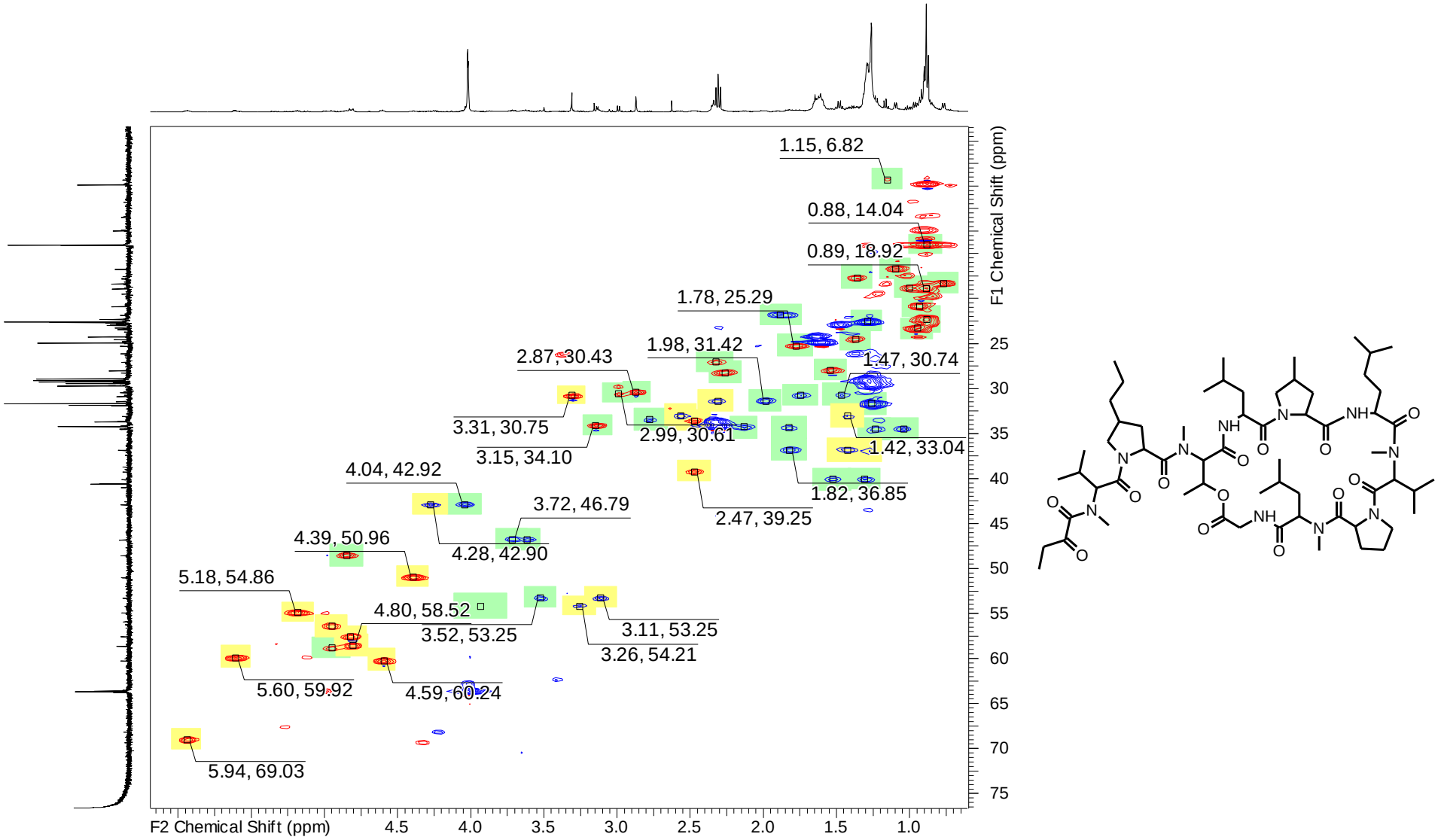
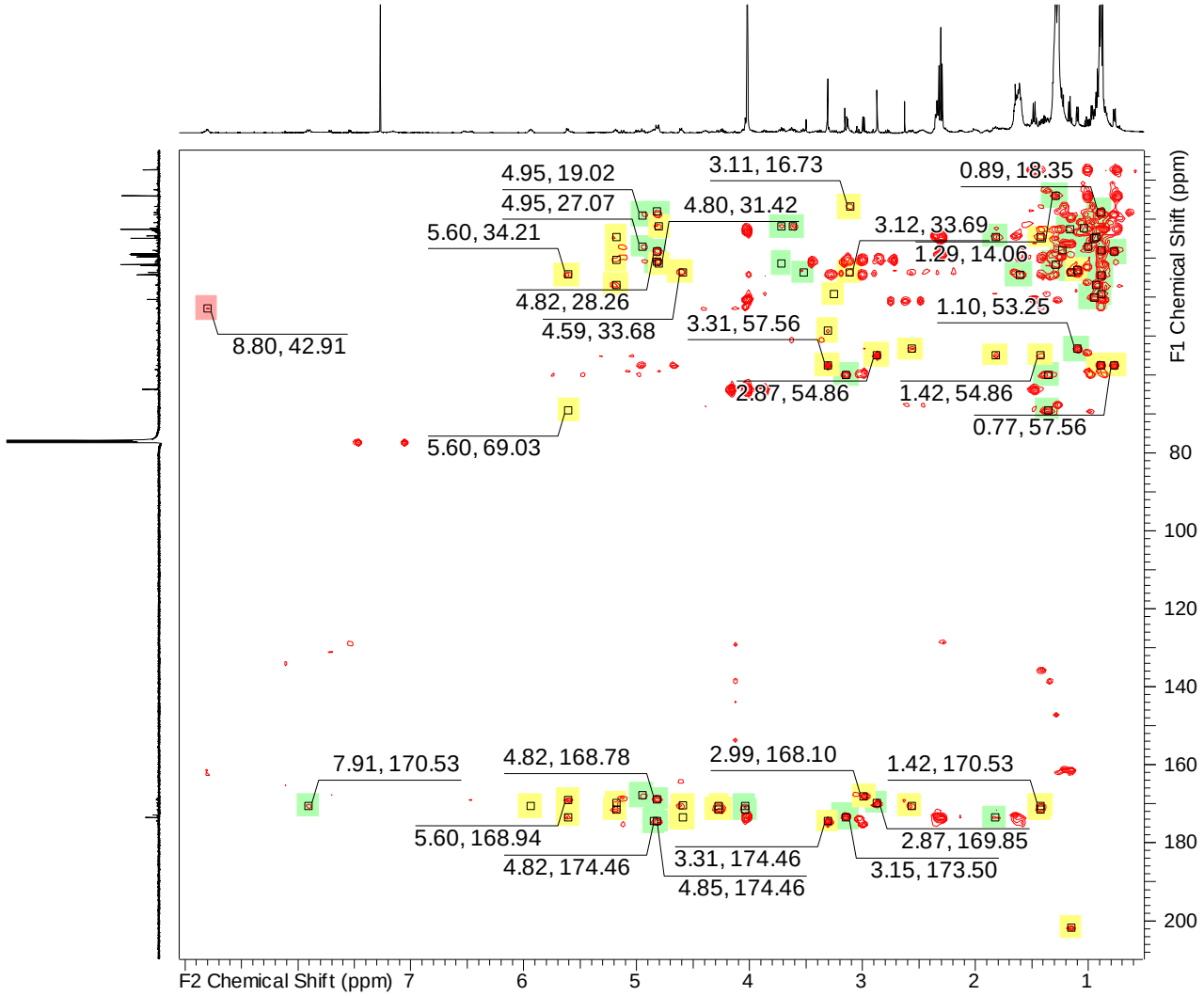


Figure S36: HSQC-spectrum of MP D in $CDCl_3$ at 500 MHz (^1H)/125 MHz (^{13}C).



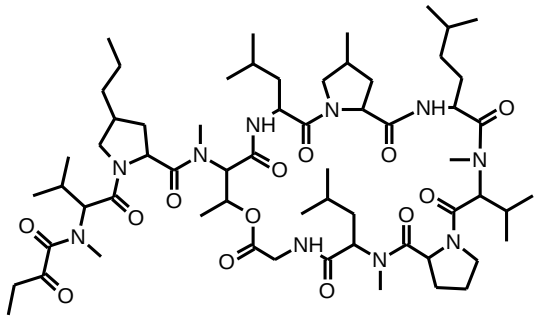
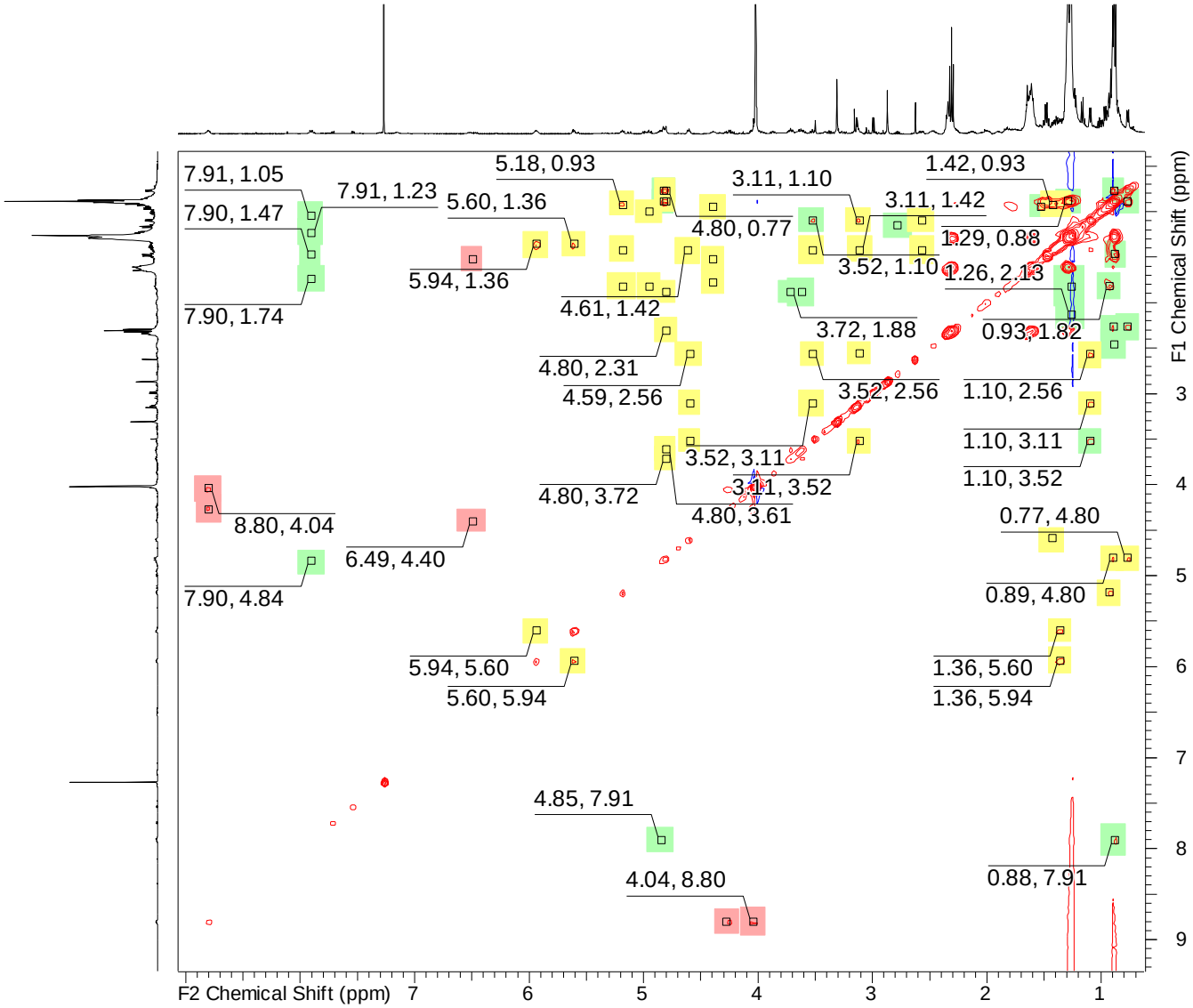


Figure S37: HMBC-spectrum of MP D in $CDCl_3$ at 500 MHz (^1H)/125 MHz (^{13}C).



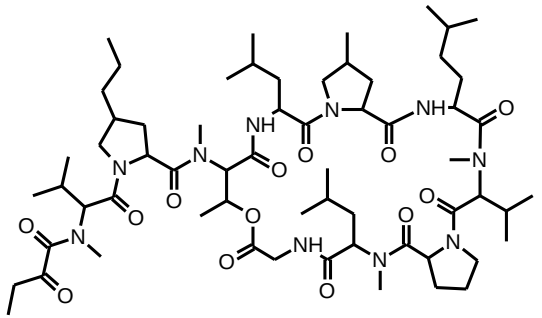


Figure S38: TOCSY-spectrum of MP D in $CDCl_3$ at 500 MHz.

Marfey's derivatization:

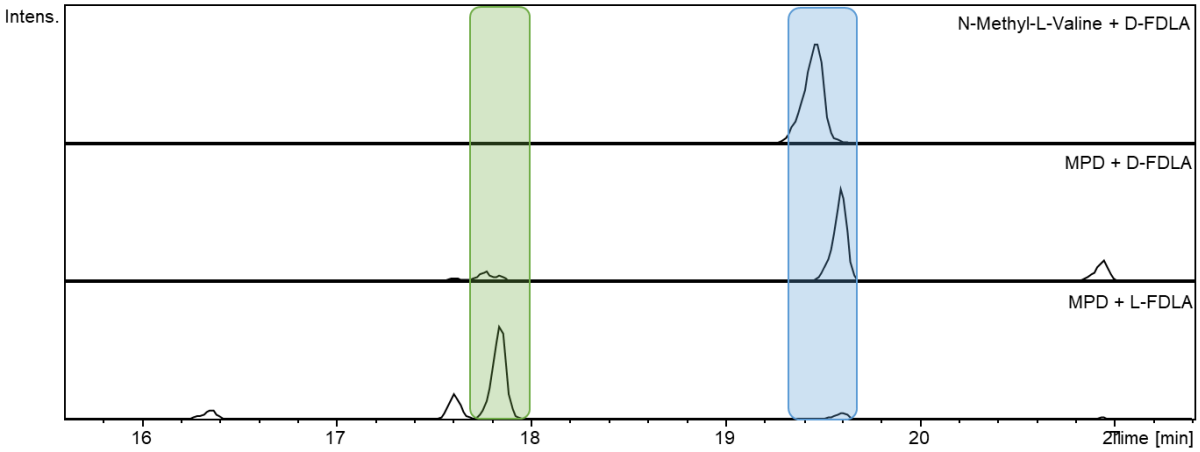


Figure S39: Marfey's derivatization of reference N-Methyl-L-Valine (upper chromatogram) with D-FDLA and MP D with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

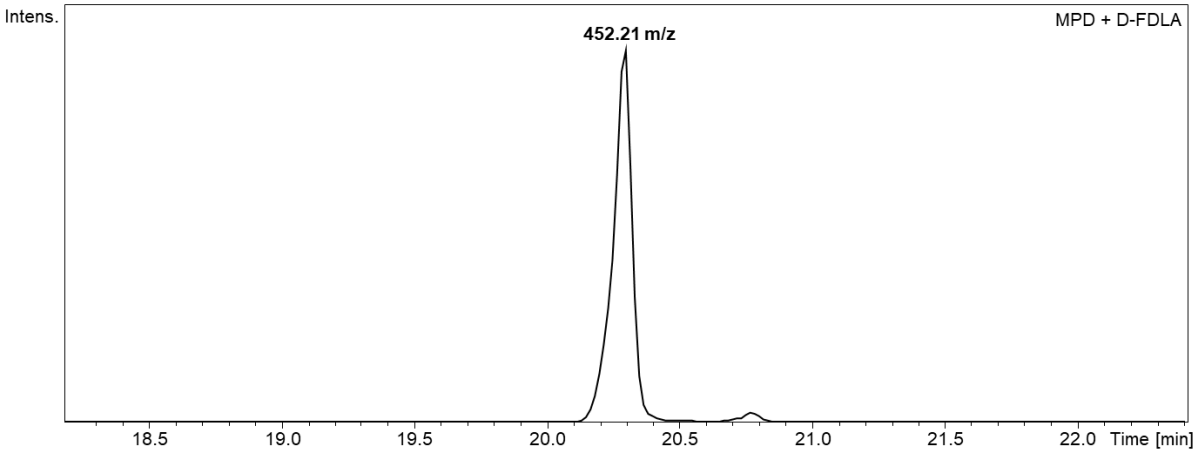


Figure S40: Marfey's derivatization of MP D with D-FDLA represented as EICs (452.21 m/z), no reference Propyl-L-Proline was available.

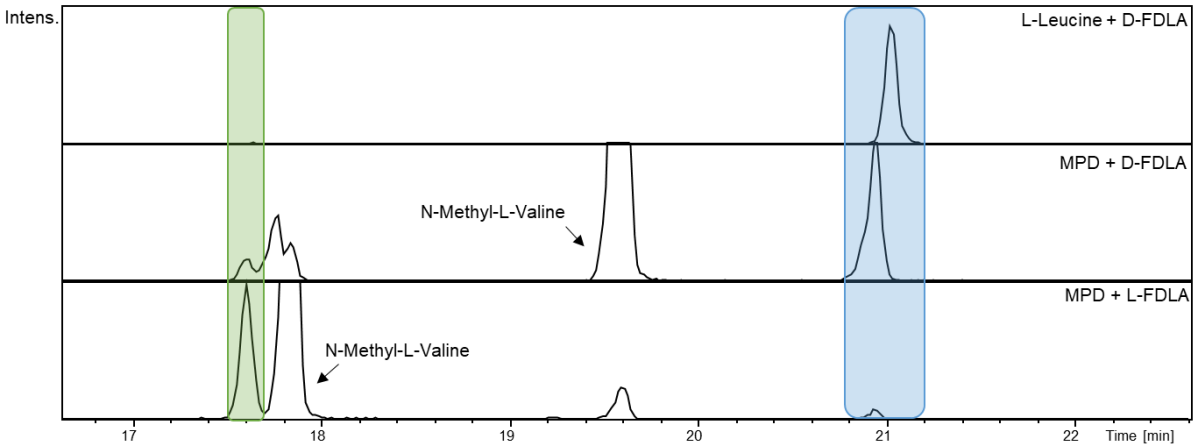


Figure S41: Marfey's derivatization of reference L-Leucine with D-FDLA (upper chromatogram) and MP D with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

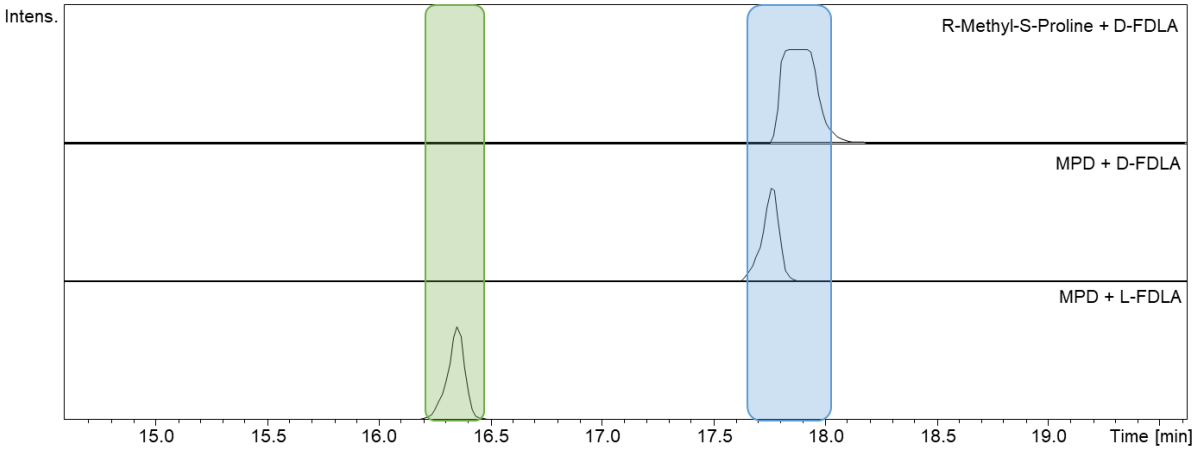


Figure S42: Marfey's derivatization of reference R-Methyl-S-Proline (upper chromatogram) with D-FDLA and MP D with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green

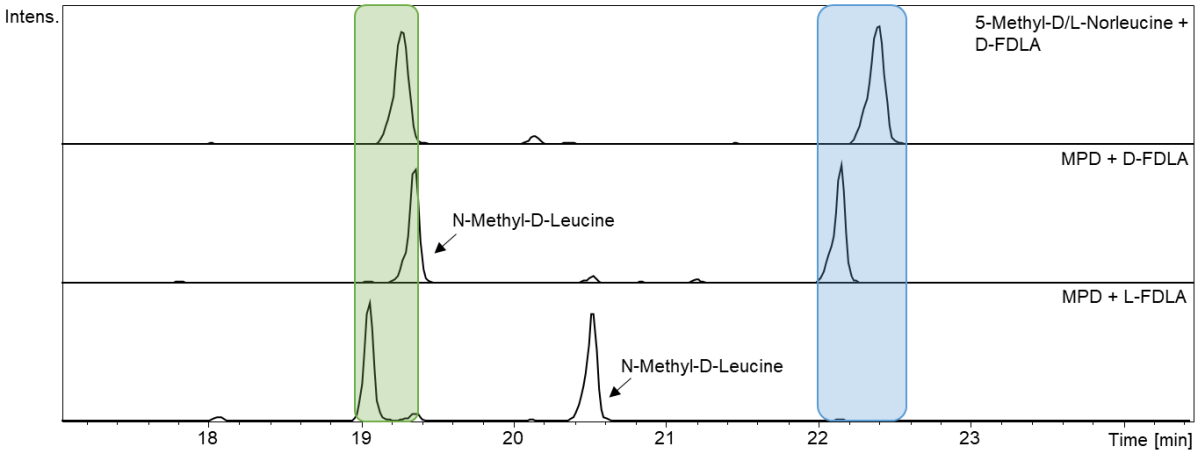


Figure S43: Marfey's derivatization of reference 5-Methyl-D/L-Norleucine with D-FDLA (upper chromatogram) and MP D with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

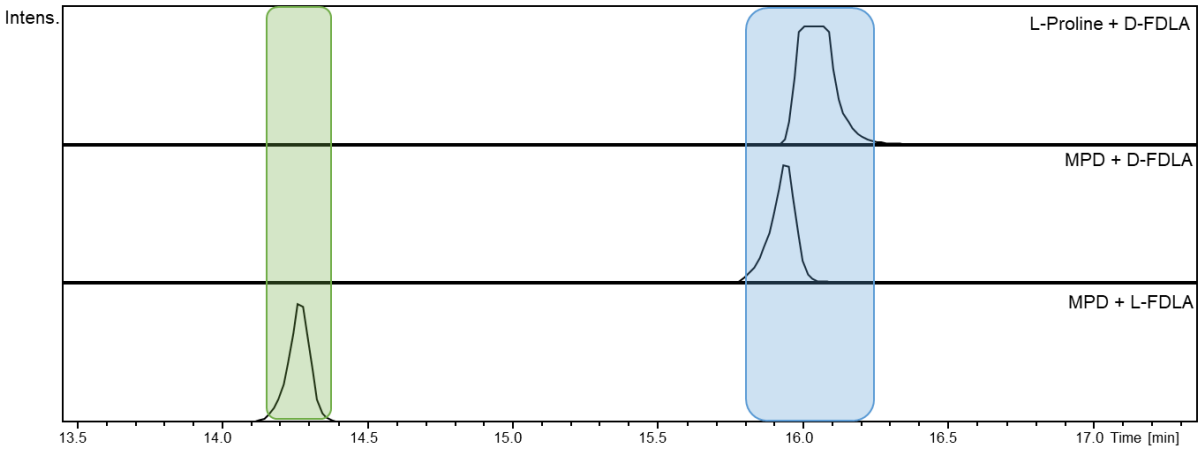


Figure S44: Marfey's derivatization of reference L-Proline (upper chromatogram) with D-FDLA and MP D with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

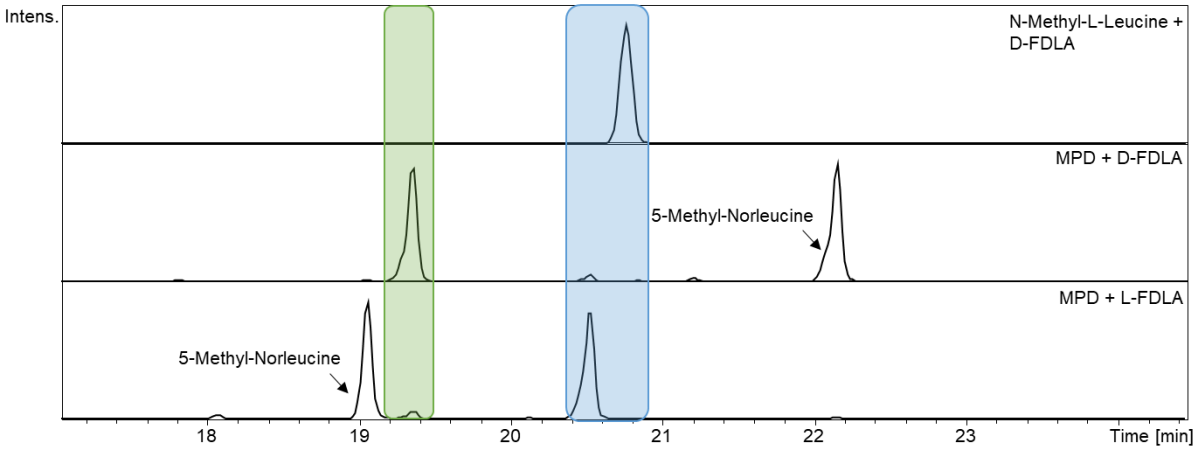


Figure S45: Marfey's derivatization of reference N-Methyl-L-Leucine with D-FDLA (upper chromatogram) and MP D with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

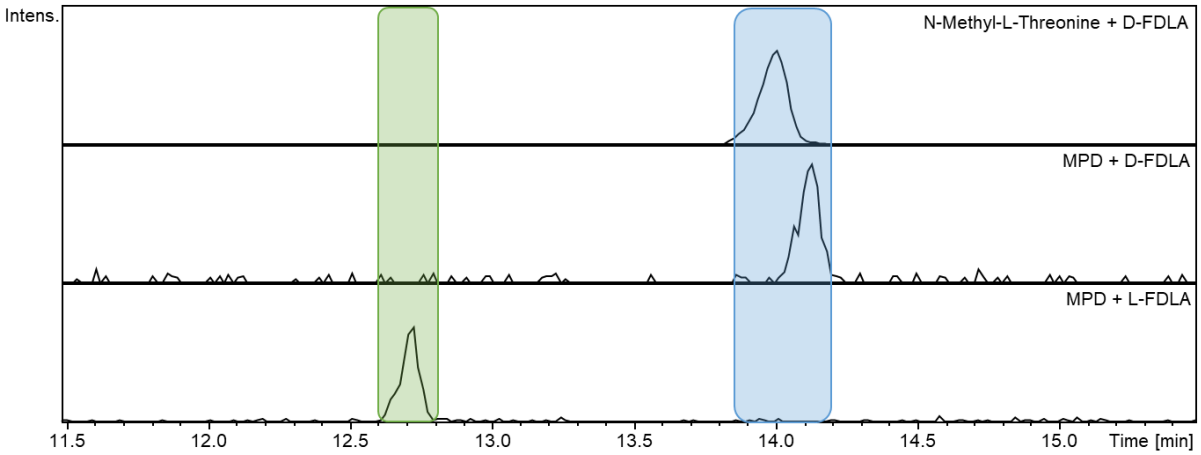


Figure S46: Marfey's derivatization of reference N-Methyl-L-Threonine with D-FDLA (upper chromatogram) and MP D with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

Mycoplanecin E

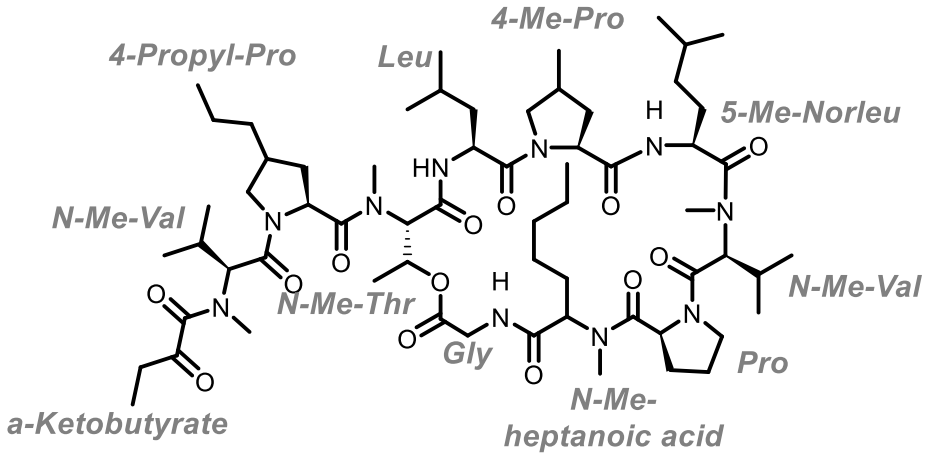


Table S12: NMR spectroscopic data of MP E in $CDCl_3$ at 700/175 MHz.

	NMR data in $CDCl_3$			
position	δ_C [ppm]	δ_H [ppm], mult. (*J* [Hz])	COSY correlations	HMBC correlations
α-Keto-butyrate				
1	168.1	-	-	-
2	201.8	-	-	-
3	33.6	2.78, m	4	2,4
4	6.8	1.15, m	3	2,3
N-Me-Val				
1	167.8	-	-	-
2	56.8	4.98, m	3	3
3	27.1	2.34, m	2,4,5	-
4,5	18.8	0.91, m	3	2,3
N-Me	30.8	2.98, m	-	1 *α-Ketobutyrate*
4-Propyl-Pro				
1	173.4	-	-	-
2	58.9	4.97, m	3	1 *4-Propyl-Pro,4, 1 N-Me-Val*
3	33.0	2.09,1.84, m	-	1,5
4	39.2	2.46, m	-	-
5	51.7	4.00,3.35, m	-	-
6	31.9	1.27, m	-	-

7	22.7	1.30, m	8	-
8	14.1	0.89, m	7	6,7
N-Me-Thr				
1	168.9	-	-	-
2	59.9	5.60, m	3	1,N-Me,1 *4-Propyl-Pro*
3	69.1	5.93, m	2,4	1 Gly
4	17.7	1.35, m	3	2,3
N-Me	34.1	3.14, m	-	2, 1 *4-Propyl-Pro*
Leu				
1	173.4	-	-	-
2	50.9	4.40, m	3,NH	1 *4-Me-Pro*,3,4
3	40.1	1.53, 1.32, m	2,4	-
4	25.4	1.78, m	3,5,6	-
5,6	21.0	0.93, m	4	-
NH	-	6.49, m	2	1 *N-Me-Thr*
4-Me-Pro				
1	170.3	-	-	-
2	60.3	4.60, br d (8.33)	3	1,5,1 *Leu*
3	33.0	1.42, m	2,4	-
4	33.6	2.47, m	3,5,6	-
5	53.3	3.12, 3.52, m	4	2,6
6	16.7	1.09, br d (6.73)	4	5
5-Me-Norleu				
1	174.4	-	-	-
2	48.7	4.85, m	3,NH	1,3,4,1 *4-Me-Pro*
3	30.8	1.76, 1.46, m	2,4	1 *4-Me-Pro*
4	34.6	1.24, 1.06, m	3	-
5	28.0	1.53, m	6,7	4
6,7	22.7	0.88, m	5	4
NH	-	7.89, m	5	1 *4-Me-Pro*
N-Me-Val				
1	168.9	-	-	-
2	57.4	4.81, m	3	1,3, 1 *5-Me-Norleu*
3	28.2	2.27, m	2,4,5	-

4,5	18.4	0.77, m	3	2,3
N-Me	30.8	3.31, s	-	2,1 *5-Me-Norleu*
Pro				
1	169.8	-	-	-
2	58.9	4.81, m	3	1,3,4
3	31.4	2.33, 1.99, m	2,4	-
4	21.9	1.87, m	3,5	-
5	46.8	3.71, 3.61, m	4	-
N-Me-heptanoic acid				
1	171.5	-	-	-
2	55.0	5.18, m	3	1,3,N-Me,1 *Pro*
3	37.0	1.82, 1.42, m	2	1,2
4	38.9	1.16, m	5	-
5	27.3	1.26, m	4	4
6	24.9	1.64, m	-	-
7	14.1	0.90, m	-	-
N-Me	30.4	2.87, m	-	4
Gly				
1	170.6	-	-	-
2	42.9	4.25, 4.04, m	NH	1,1 *N-Me-Leu*
NH	-	8.79, m	2	-

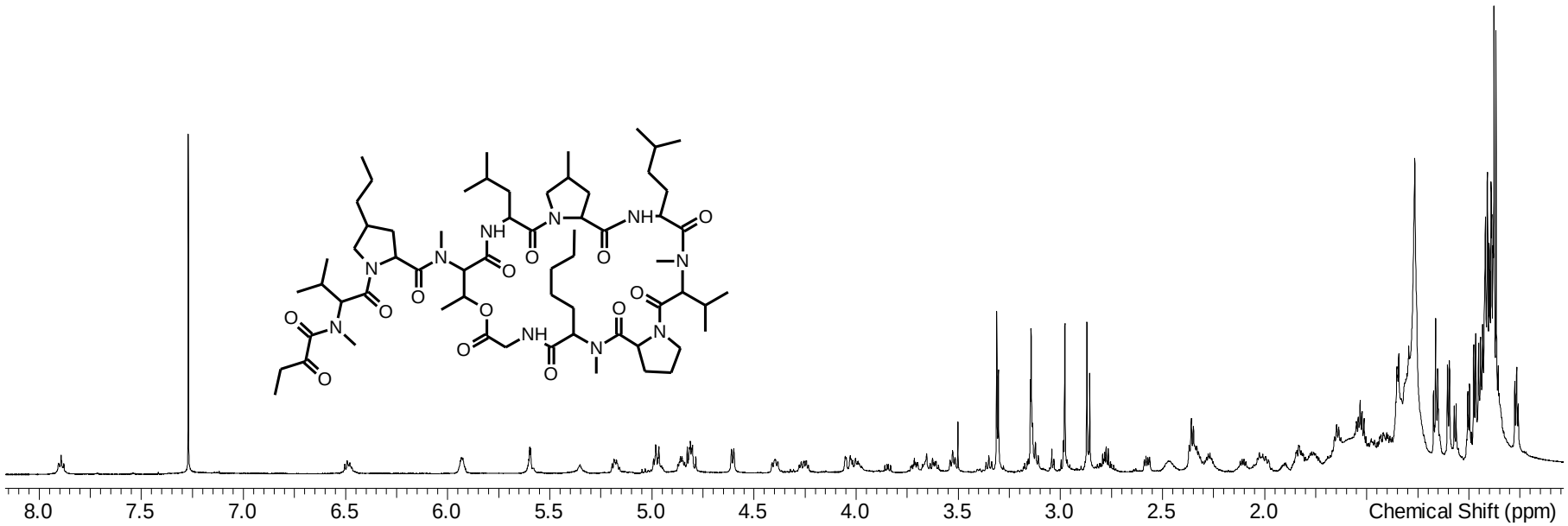


Figure S47: ^{1}H-spectrum of MP E in $CDCl_3$ at 700 MHz.

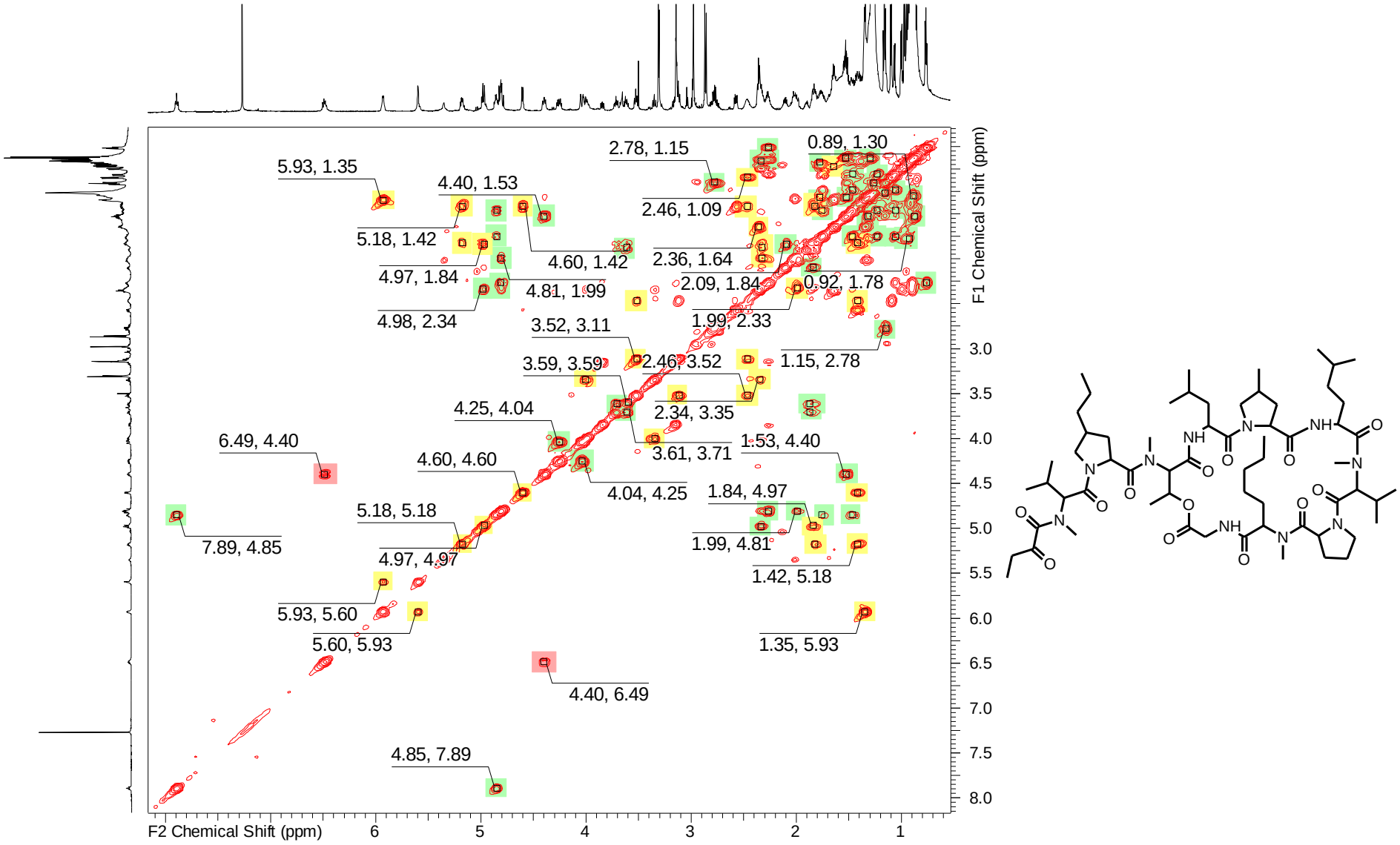


Figure S48: COSY-spectrum of MP E in $CDCl_3$ at 700 MHz.

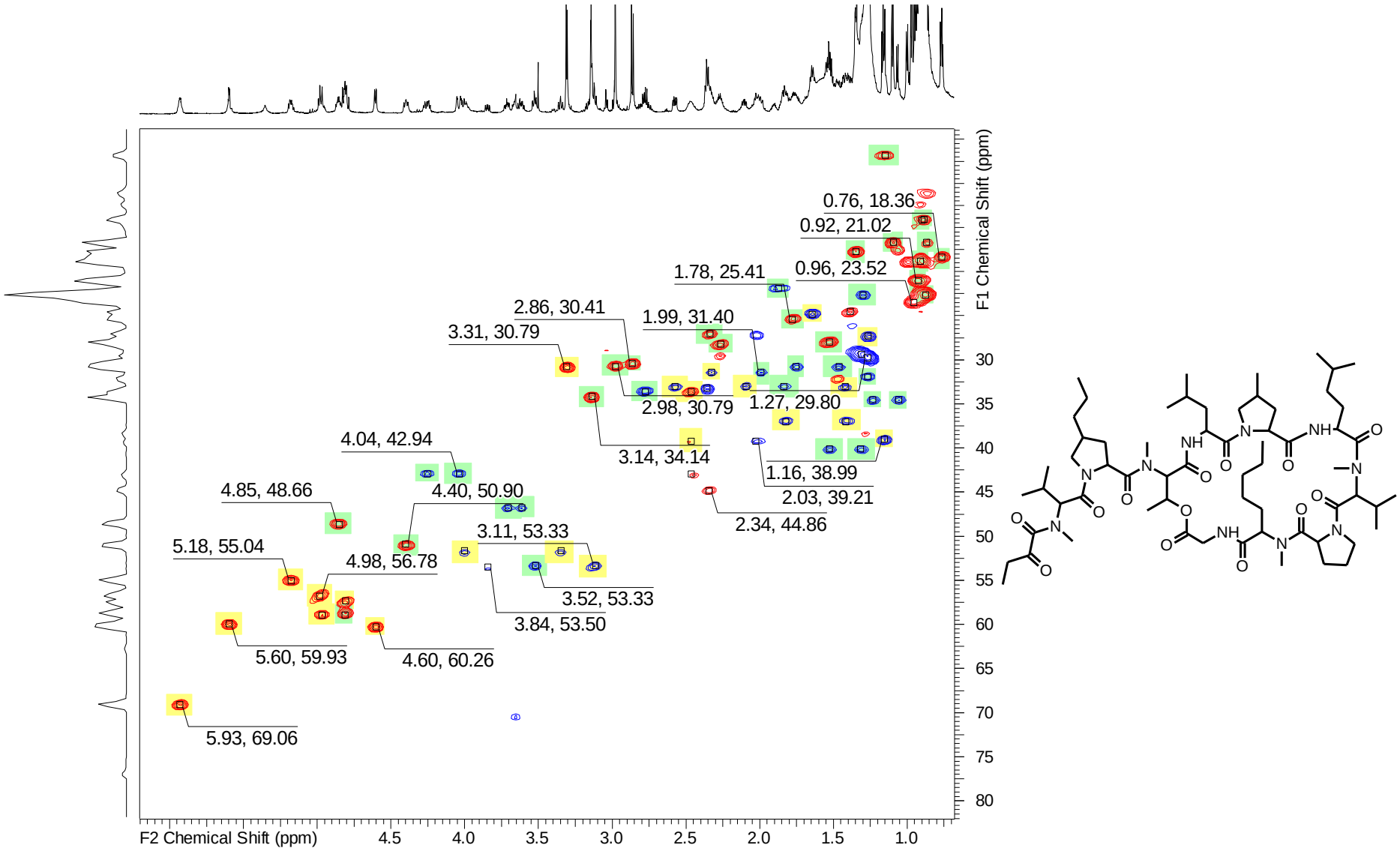


Figure S49: HSQC-spectrum of MP E in $CDCl_3$ at 700 MHz (^1H)/175 MHz (^{13}C).

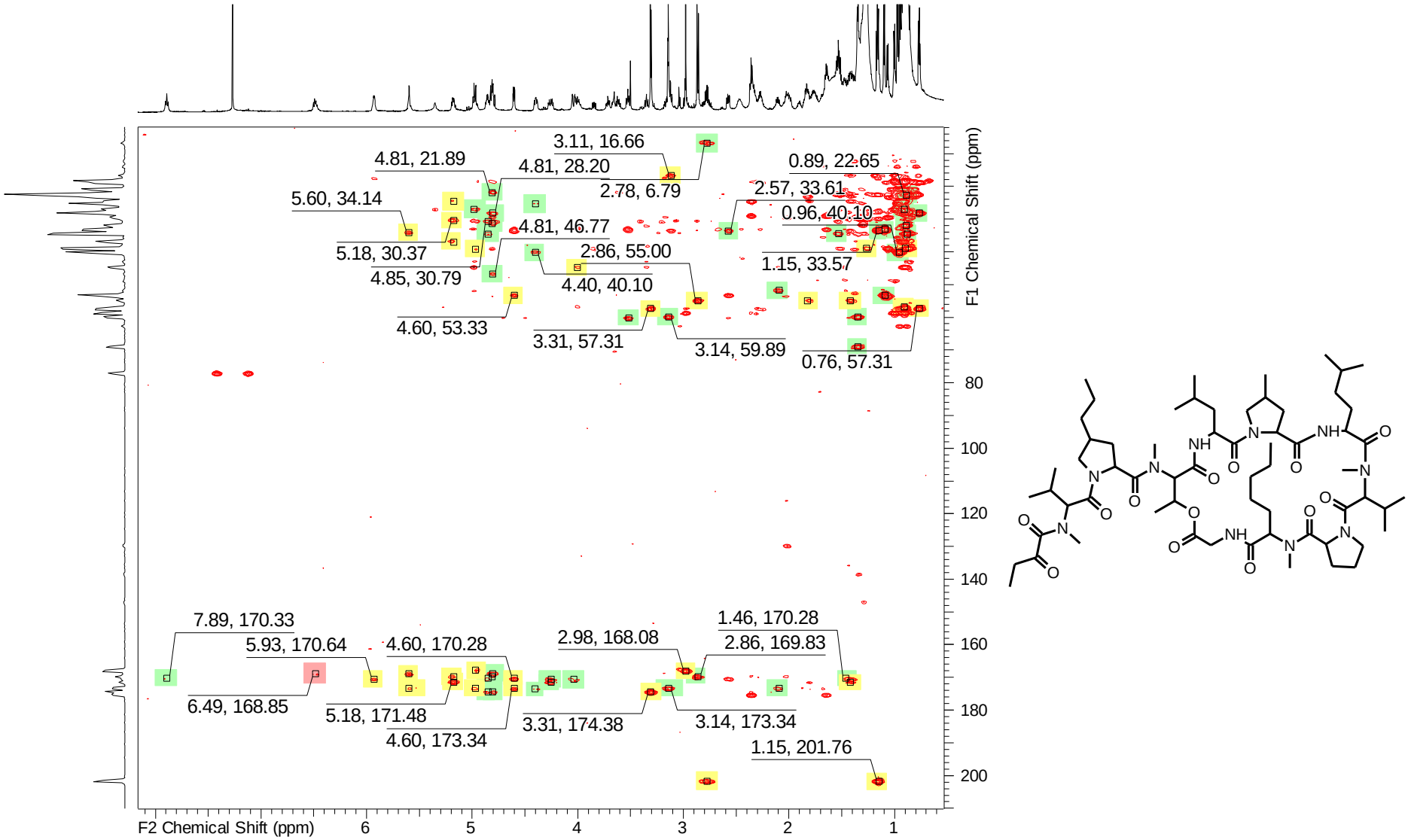


Figure S50: HMBC-spectrum of MP E in $CDCl_3$ at 700 MHz (^1H)/175 MHz (^{13}C).

Marfey's derivatization:

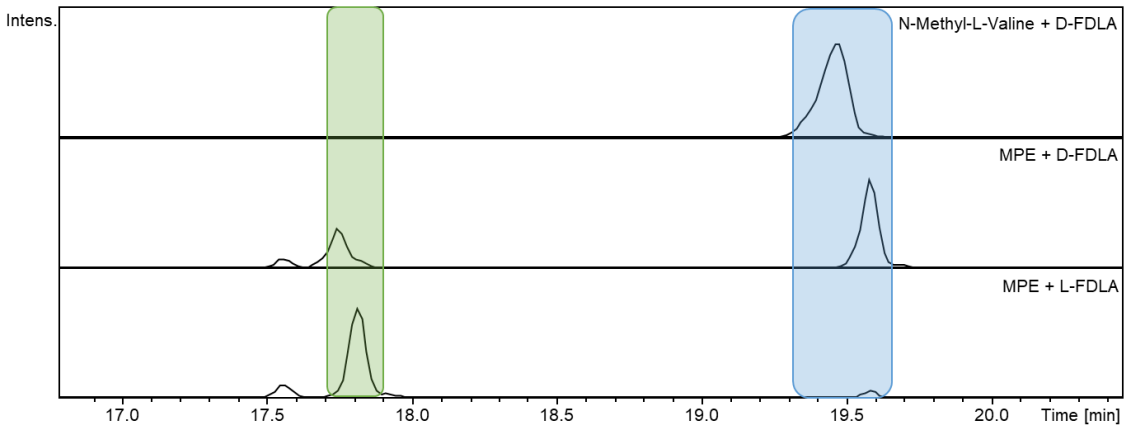


Figure S51: Marfey's derivatization of reference N-Methyl-L-Valine (upper chromatogram) with D-FDLA and MP E with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

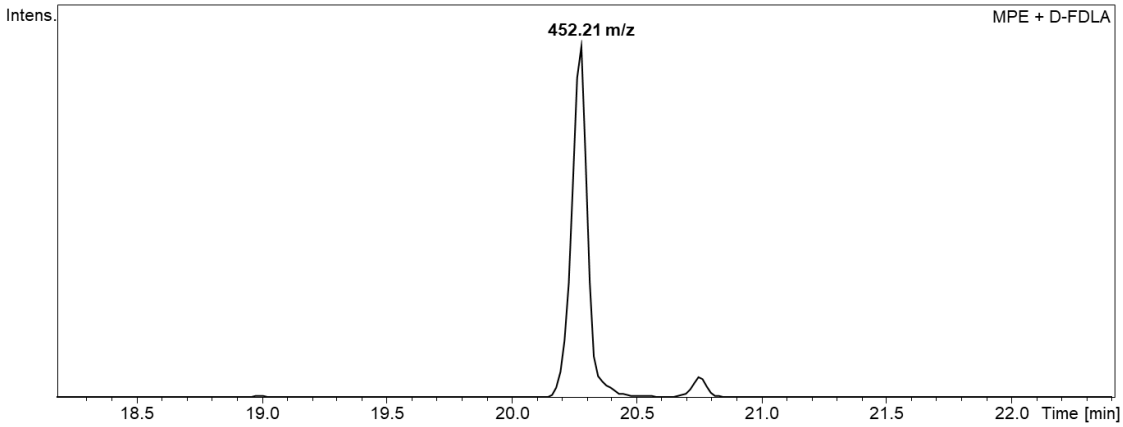


Figure S52: Marfey's derivatization of MP E with D-FDLA represented as EICs (452.21 m/z), no reference Propyl-L-Proline was available.

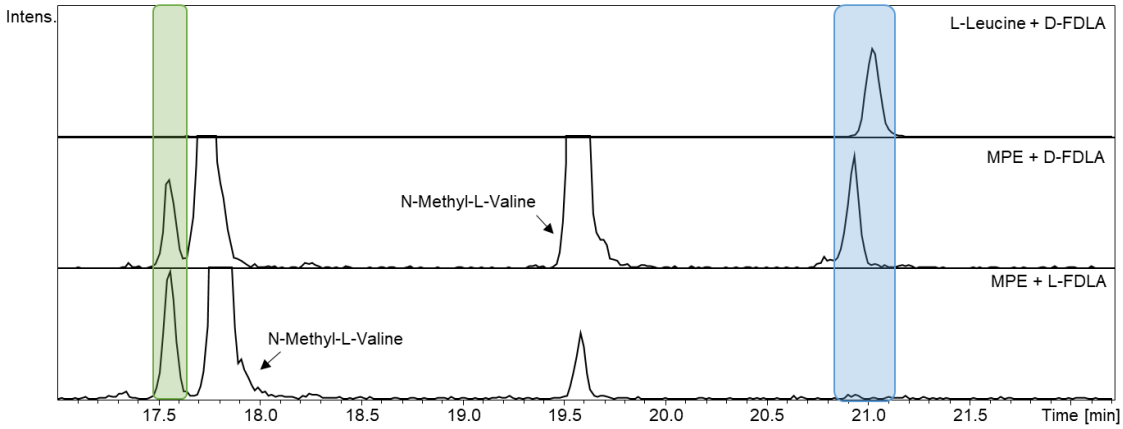


Figure S53: Marfey's derivatization of reference L-Leucine with D-FDLA (upper chromatogram) and MP E with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

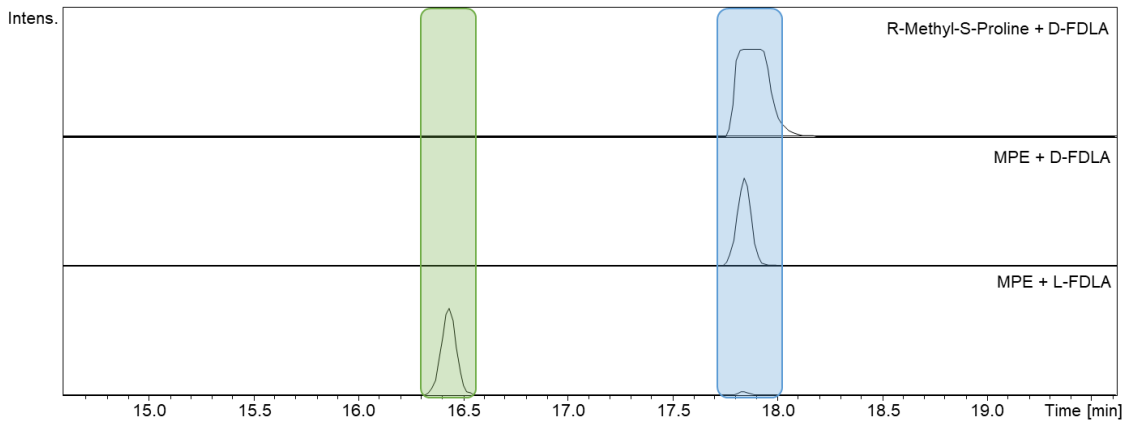


Figure S54: Marfey's derivatization of reference R-Methyl-S-Proline (upper chromatogram) with D-FDLA and MPE with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

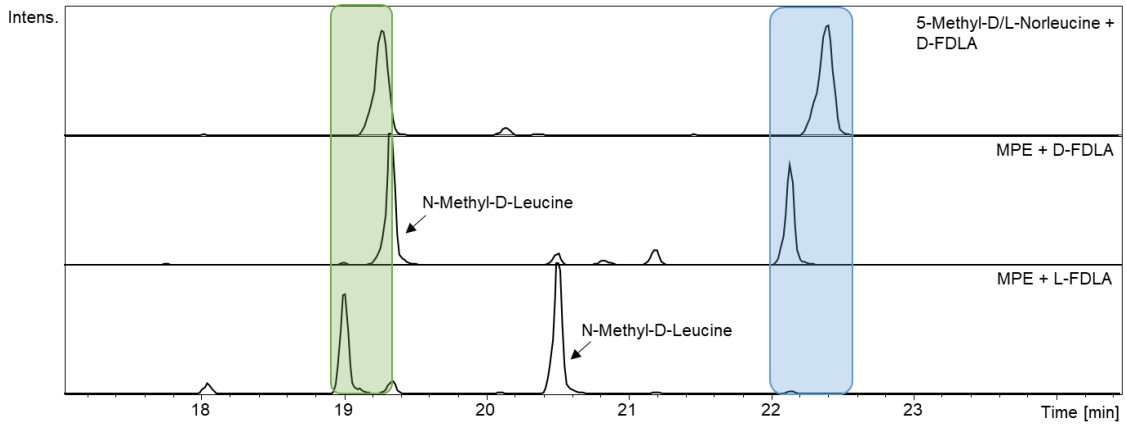


Figure S55: Marfey's derivatization of reference 5-Methyl-D/L-Norleucine with D-FDLA (upper chromatogram) and MP E with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

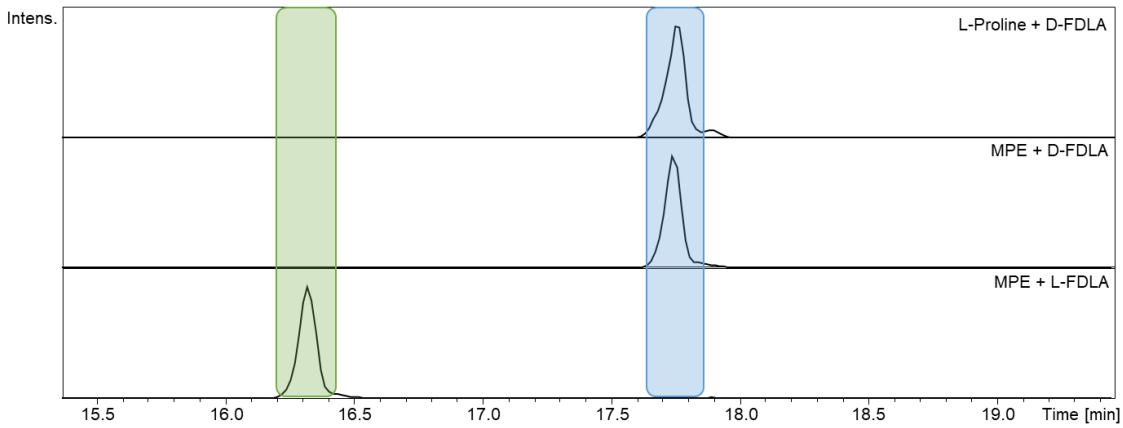


Figure S56: Marfey's derivatization of reference L-Proline (upper chromatogram) with D-FDLA and MPE with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

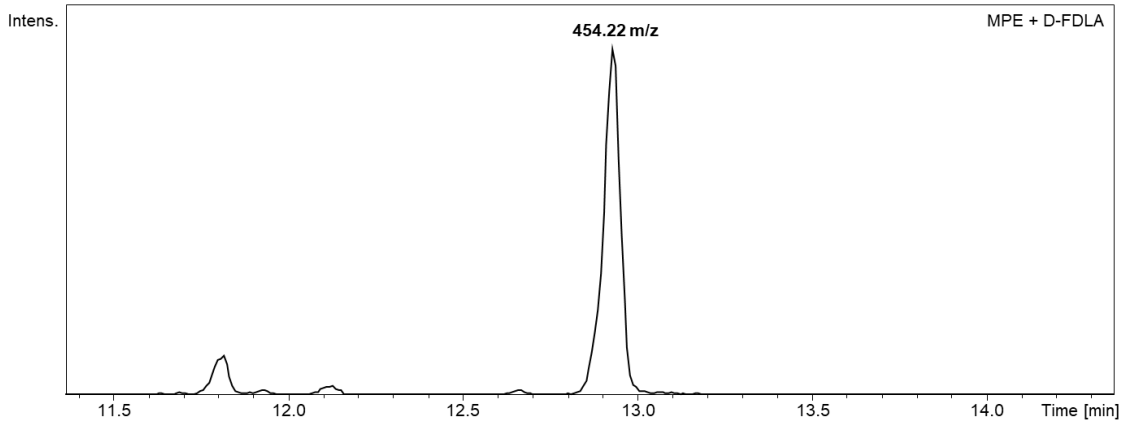


Figure S57: Marfey's derivatization of MP E with D-FDLA measured with a more apolar gradient represented as EICs (454.22 m/z), no reference N-Methyl-D/L-heptanoic acid was available.

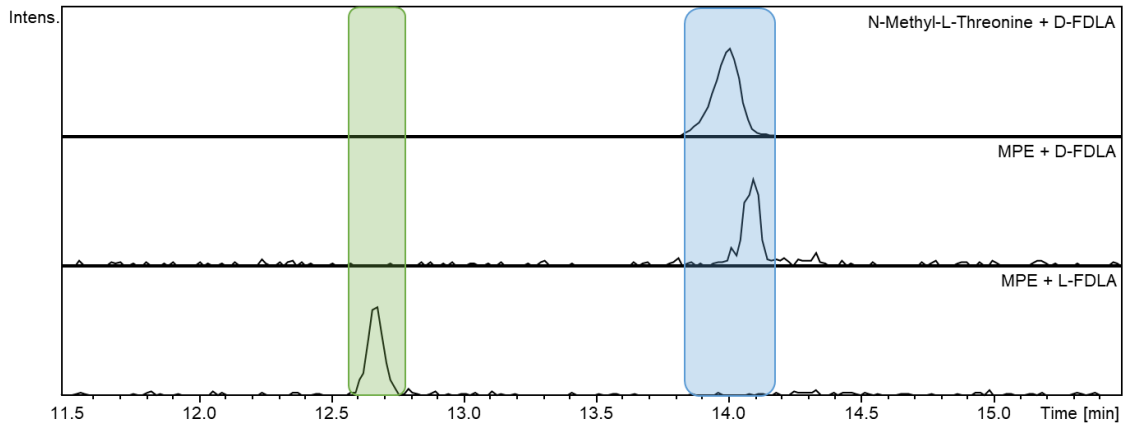


Figure S58: Marfey's derivatization of reference N-Methyl-L-Threonine (upper chromatogram) with D-FDLA and MP E with both D-FDLA (middle chromatogram) and L-FDLA (lower chromatogram) for retention time comparison represented as EICs; D-L-FDLA/L-D-FDLA highlighted in blue and L-L-FDLA/D-D-FDLA in green.

S4.4. References

1. Lukat, P.; Katsuyama, Y.; Wenzel, S.; Binz, T.; König, C.; Blankenfeldt, W.; Brönstrup, M.; Müller, R. Biosynthesis of methyl-proline containing griselimycins, natural products with anti-tuberculosis activity. *Chem. Sci.* **2017**, *8*, 7521–7527, doi:10.1039/C7SC02622F.
2. Wittmann, M.; Linne, U.; Pohlmann, V.; Marahiel, M.A. Role of DptE and DptF in the lipidation reaction of daptomycin. *FEBS J.* **2008**, *275*, 5343–5354, doi:10.1111/j.1742-4658.2008.06664.x.
3. Kraas, F.I.; Helmetag, V.; Wittmann, M.; Strieker, M.; Marahiel, M.A. Functional dissection of surfactin synthetase initiation module reveals insights into the mechanism of lipoinitiation. *Chem. Biol.* **2010**, *17*, 872–880, doi:10.1016/j.chembiol.2010.06.015.
4. Imker, H.J.; Krahn, D.; Clerc, J.; Kaiser, M.; Walsh, C.T. N-acylation during glidobactin biosynthesis by the tridomain nonribosomal peptide synthetase module GlbF. *Chem. Biol.* **2010**, *17*, 1077–1083, doi:10.1016/j.chembiol.2010.08.007.
5. Lewis, R.A.; Nunns, L.; Thirlway, J.; Carroll, K.; Smith, C.P.; Micklefield, J. Active site modification of the β-ketoacyl-ACP synthase FabF3 of Streptomyces coelicolor affects the fatty acid chain length of the CDA lipopeptides. *Chem. Commun. (Camb)* **2011**, *47*, 1860–1862, doi:10.1039/c0cc03444d.
6. Bretschneider, T.; Zocher, G.; Unger, M.; Scherlach, K.; Stehle, T.; Hertweck, C. A ketosynthase homolog uses malonyl units to form esters in cervimycin biosynthesis. *Nat. Chem. Biol.* **2011**, *8*, 154–161, doi:10.1038/nchembio.746.
7. Jackson, D.R.; Shakya, G.; Patel, A.B.; Mohammed, L.Y.; Vasilakis, K.; Wattana-Amorn, P.; Valentic, T.R.; Milligan, J.C.; Crump, M.P.; Crosby, J.; et al. Structural and Functional Studies of the Daunorubicin Priming Ketosynthase DpsC. *ACS Chem. Biol.* **2018**, *13*, 141–151, doi:10.1021/acschembio.7b00551.
8. Pan, H.; Tsai, S.-C.; Meadows, E.S.; Miercke, L.J.W.; Keatinge-Clay, A.T.; O'Connell, J.; Khosla, C.; Stroud, R.M. Crystal Structure of the Priming β-Ketosynthase from the R1128 Polyketide Biosynthetic Pathway. *Structure* **2002**, *10*, 1559–1568, doi:10.1016/s0969-2126(02)00889-4.
9. Hitchman, T.S.; Crosby, J.; Byrom, K.J.; Cox, R.J.; Simpson, T.J. Catalytic self-acylation of type II polyketide synthase acyl carrier proteins. *Chem. Biol.* **1998**, *5*, 35–47, doi:10.1016/s1074-5521(98)90085-0.

10. Misra, A.; Sharma, S.K.; Surolia, N.; Surolia, A. Self-Acylation Properties of Type II Fatty Acid Biosynthesis Acyl Carrier Protein. *Chem. Biol.* **2007**, *14*, 775–783, doi:10.1016/j.chembiol.2007.05.013.
11. Steiningerova, L.; Kamenik, Z.; Gazak, R.; Kadlcik, S.; Bashiri, G.; Man, P.; Kuzma, M.; Pavlikova, M.; Janata, J. Different Reaction Specificities of F420H2-Dependent Reductases Facilitate Pyrrolobenzodiazepines and Lincomycin To Fit Their Biological Targets. *J. Am. Chem. Soc.* **2020**, *142*, 3440–3448, doi:10.1021/jacs.9b11234.
12. Wang, S.-A.; Lin, C.-I.; Zhang, J.; Ushimaru, R.; Sasaki, E.; Liu, H.-W. Studies of lincosamide formation complete the biosynthetic pathway for lincomycin A. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 24794–24801, doi:10.1073/pnas.2009306117.
13. Son, S.; Hong, Y.-S.; Jang, M.; Heo, K.T.; Lee, B.; Jang, J.-P.; Kim, J.-W.; Ryoo, I.-J.; Kim, W.-G.; Ko, S.-K.; et al. Genomics-Driven Discovery of Chlorinated Cyclic Hexapeptides Ulleungmycins A and B from a Streptomyces Species. *J. Nat. Prod.* **2017**, *80*, 3025–3031, doi:10.1021/acs.jnatprod.7b00660.
14. SHIBA, T.; MUKUNOKI, Y. The total structure of the antibiotic longicatenamycin. *J. Antibiot.* **1975**, *28*, 561–566, doi:10.7164/antibiotics.28.561.
15. Junne, T.; Wong, J.; Studer, C.; Aust, T.; Bauer, B.W.; Beibel, M.; Bhullar, B.; Bruccoleri, R.; Eichenberger, J.; Estoppey, D.; et al. Decatransin, a new natural product inhibiting protein translocation at the Sec61/SecYEG translocon. *J. Cell Sci.* **2015**, *128*, 1217–1229, doi:10.1242/jcs.165746.
16. Nakajima, M.; Torikata, A.; TAMAOKI, H.; Haneishi, T.; Arai, M.; KINOSHITA, T.; KUWANO, H. Mycoplanecins, novel antimycobacterial antibiotics from Actinoplanes awajinensis subsp. mycoplanecinus subsp. nov. III. Structural determination of mycoplanecin A. *J. Antibiot.* **1983**, *36*, 967–975, doi:10.7164/antibiotics.36.967.

Chapter 5

5. Discussion

Within the scope of this thesis, three novel secondary metabolites from different structural classes were discovered from the myxobacterial strain MSr11367 by comprehensive metabolome mining including the evaluation of cultivation under different conditions, thus applying the OSMAC approach. Furthermore, genome mining of the same strain led to the discovery of a novel siderophore, which was made accessible via heterologous expression. Finally, a combinatorial bioactivity-genome mining approach looking for alkylproline-containing secondary metabolites resulted in reprioritization of the well-described mycoplanecins with promising activities against *M. tuberculosis*. As all secondary metabolites identified in this study were identified through different approaches, the following chapters will discuss their advantages and disadvantages and compare them with each other. The various approaches including OSMAC, diversity-oriented strain investigation, bioactivity-guided fractionation, metabolome mining and genome mining approaches are compared with respect to the three projects in this thesis and are critically assessed regarding their utility for NP discovery. This discussion aims at facilitating the choice of the most beneficial approach or combinations thereof for future studies to help exploiting the potential of myxobacteria and other microbes for natural products discovery.

5.1 OSMAC and exploring underinvestigated genera

The discovery of novel NPs can be reached via different strategies such as investigation of more distinct genera to increase the likelihood of finding distinct chemistry, or in-depth analysis of the secondary metabolome of a single strain to explore its hidden biosynthetic potential (OSMAC). The first strategy involves the comparison of many strains, which is often biased towards the more abundant metabolites. Highly active substances however, only need to be produced with relatively low titers to exhibit their effect. The OSMAC approach on the contrary is a well-known strategy that has been successfully applied for the past 20 years, whereby variation of the media composition proved to be the most effective way to increase the number of detectable secondary metabolites of a single microbial source in many cases [1,2]. The advantage of focusing on one strain is seen as increased chances to discover biologically active substances including metabolites produced in low but sufficient yield (in terms of availability for characterization), which can become more prominent with the OSMAC

approach. An overview about the application of OSMAC between 2002 and 2018 is presented in the review of Pan *et al.* assembling more than 500 compounds with mainly derivatives of known compound classes but also novel NPs [2]. Thereby about 85% originate from fungal sources and solely 15% from actinobacteria. More recent examples for OSMAC application on actinobacteria are usually combinations with orthogonal strategies like untargeted metabolomics analysis [3], metabolomics with molecular modelling [4] or metabologenomic investigation [5] as well as utilization of less explored types of actinobacteria, *e.g.* thermophilic ones [6].

Although the OSMAC approach has long been established for actinobacteria, few attempts for myxobacteria have been reported so far [7], which encouraged the study described in Chapter 2 to explore the secondary metabolome of a new myxobacterial strain belonging to the *Sorangiineae* suborder. This strain was picked as it possesses a huge biosynthetic potential, represented by 62 predicted BGCs in its genome. The successful application of the OSMAC approach in this thesis is reflected in the discovery of three novel substances belonging to distinct structural classes from this single strain. Even though this is a relatively high number compared to the many studies that were only capable of elucidating a single metabolite from a new strain, it still only reflects a fractional amount of MSr11367s biosynthetic potential, opening the possibility to discover additional secondary metabolites from this strain in future studies. Interestingly, even though already extensively used in this study, the OSMAC approach still bears potential here: variation of the media composition, further alteration of cultivation parameters or co-cultivation is possible to active further cryptic BGCs (Figure 1). The most decisive media ingredient variations for bacteria and fungi are carbon and nitrogen sources [8], higher salinity [9] as well as metal ion concentrations [10]. Crucial parameters like temperature, pH and oxygen supply, could also help to diversify the detectable secondary metabolism [11]. Co-cultivations with other microorganisms like Gram-positive bacteria, Gram-negative bacteria, fungi and yeasts can stimulate production of inhibitory or promoting secondary metabolites that were not detectable under standard cultivation conditions [12–14]. Nevertheless, it is important to carefully select appropriate strains for this approach, as it is not yet possible on a high throughput scale since only few myxobacteria have been cultivated in 96-well plate format to date. Strains containing many BGCs not correlated with their respective small-molecule products are the most promising sources of new NPs using OSMAC, making genome-analysis an important tool for prioritization. Among the strains available to us, MSr11367 clearly stands out showing 62 BGCs predicted by antiSMASH with a vast amount of cryptic clusters under standard cultivation conditions having only one BGC assigned to

known products (nocardamine family). Even a media screening including 11 different media, did not yield sufficient amounts of unknown secondary metabolites for structure elucidation, which however could be overcome by applying the OSMAC approach. None of the isolated compounds showed activity against our in-house microbial panel, so with regards to discovery of novel antibiotics this approach unfortunately was not successful for this strain yet. As each secondary metabolite is assumed to exhibit a distinct purpose though, their potential might be uncovered in future studies applying a broader test panel for biological functions. They might exemplarily serve as signaling molecules for intercellular communication, making them an interesting target for studying myxobacterial cell biology or interspecies communications. Their potential application as drugs is also not excluded, given the possibility to detect activity against other non-infectious diseases, *e.g.* immunomodulatory activity of soraphen A isolated from *Sorangium cellulosum* [15] or rapamycin isolated from *Streptomyces hygroscopicus* [16], which aim for key metabolic pathways initiated by T cell activation [17]. Furthermore, an application as future drugs against pathogens such as viruses or plasmodia should not be ruled out, as those are simply not included in our test panel.

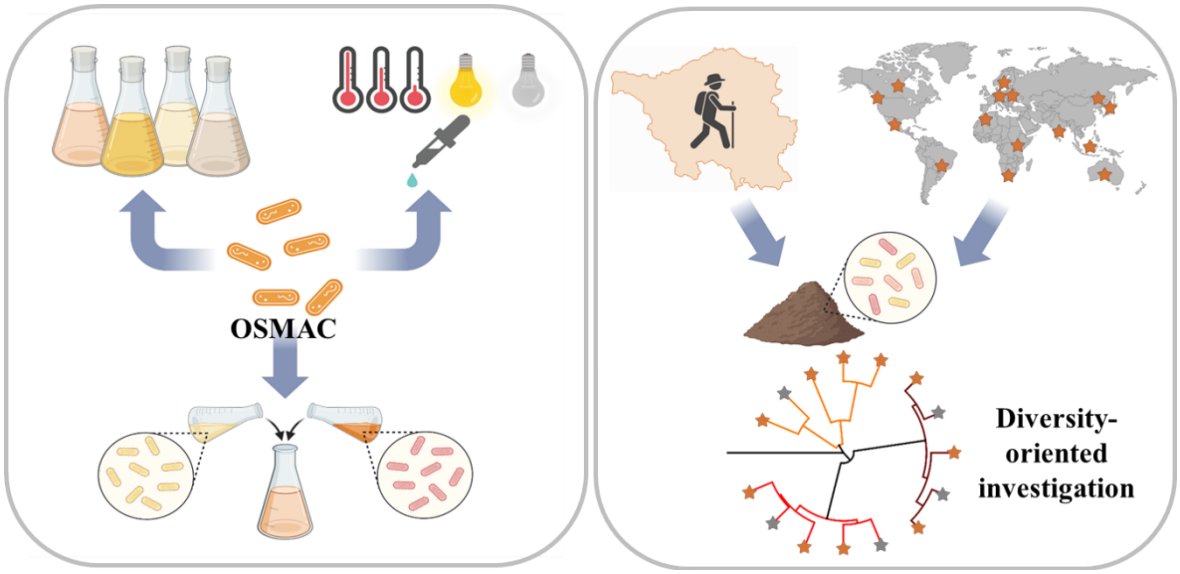


Figure 1: Illustration of one strain many compounds (OSMAC) approach using different media, variation of cultivation parameters and co-cultivation (left) versus diversity-oriented strain investigation originating from either arbitrary locations within one area (e.g. Saarland) or habitats upwards to global scale including exotic locations (right) to find chemically diverse NPs.

On the other hand, taxonomic distance has already been correlated with chemical diversity of the secondary metabolome and successfully yielded a variety of unknown secondary metabolites [10]. Even during limitation to a single medium and focusing on solely the most abundant peaks, novel secondary metabolites are likely to be discovered with this method. It

benefits from often higher production rates without cultivation condition variations or genetic manipulation. Nevertheless, the diversity-oriented investigation strategy needs to deal with challenges such as the likelihood of rediscovery of NP derivatives, which is increased through the parallel investigation of multiple strains and therefore requires the detailed inspection of metabolome and genome data to ensure proper dereplication.

Those two approaches—OSMAC and investigation of distant genera—might as well go hand in hand, as investigation of distant genera can serve as basis to pick the most promising subset of strains for further investigation using the OSMAC approach. A balance between the amount of invested time and the devoted attention to each strain is necessary to receive the most promising results with respect to novel, bioactive NPs in a reasonable time frame.

5.2 The bioactivity guided approach as a classical strategy for NP discovery

Testing of crude extracts (and fractions thereof) against different pathogens is the oldest applied approach for drug discovery among the techniques presented in this study. At first glance this bioactivity guided strategy is the most goal-oriented and straight-forward technique to find promising bioactive compounds [18]. As discussed above, investigating unknown genera increases chances to find "true" novel chemistry with promising activity following this diversity-oriented approach [19,20]. Even though it has some limitations, as following biological activities favors rediscovery of already known secondary metabolites, new derivatives can be uncovered, that might be less active, but have more beneficial physico-chemical properties like superior pharmacological features for absorption, distribution, metabolism, excretion and toxicity (ADMET). In many cases the highest bioactivities within one compound class can be observed for the most abundant derivatives, which are usually discovered and characterized first [21–23]. This might be due to the evolutionary development for self-defense, so they often are correlated to higher biological activities within one compound class. However, only following the strongest biological activities without taking other information such as genome or metabolome data into consideration bears a high risk to spend precious time and resources on purification of already known compounds (Section 5.3). The combination of bioactivity guided analysis with the OSMAC approach can be an easy and fast way to screen for potentially novel anti-infectives that are produced under different conditions using the same strain, which also facilitates target mass determination via comparison to inactive or less active conditions. This was also performed during the study in

Chapter 2, where crude extracts of the 12 different cultivation conditions were tested and evaluated regarding their activity (Table 1, Figure 2).

Even though the strain indeed showed some promising biological activities under various conditions, they could all be correlated to the presence of known secondary metabolites, some that are currently under investigation in a separate project, or due to the chosen cultivation conditions (addition of antibiotics). The activity of the extracts against Gram-negative pathogens could be explained by the addition of the three antibiotics rifampicin, chloramphenicol and trimethoprim to the cultivation due to the same activity found in the respective blanks. The activity of the extracts against Gram-positive pathogens could be correlated to the presence of several analogues of the nocardamine family, which are known for their bioactivity with synergistic tendencies [24]. Antifungal activity was traced back to production of the myxoquaterines, another unprecedented and not yet published family of natural products discovered in our group [25].

Table 1: Bioactivity testing of extracts obtained from 12 different cultivation conditions with 0=no activity detected up to 64 µg/mL, A=activity at 64 µg/mL; B-H=serial 1:1 dilutions of A; double brackets = precipitation during testing; A) results for Gram-negative bacteria; B) results for Gram-positive bacteria (Staphylococcus aureus, Mycobacterium smegmatis, Bacillus subtilis) and fungi (Pichia anomala, Candida albicans, Cryptococcus neoformans); assay conducted as described in Chapter 4; color gradient ranging from dark green = highest activity to white = no activity; grey = activity correlates with antibiotic addition.

A)	***A. baumannii***	***E. coli* (*ΔacrB*)**	***E. coli* WT**	***P. aeruginosa***	***C. freundii***
30°C	0	0	0	0	0
18°C	0	0	0	0	0
37°C	0	0	0	0	0
Dark	0	0	0	0	0
Light	0	0	0	0	0
Oxygen limitation	0	0	0	0	0
Chloramphenicol	B	H	F	A	E
Rifampicin	C	((C))	0	0	0
Trimethoprim	D	H	H	A	H
Soil extract	0	0	0	0	0
LaCl3	0	0	0	0	0
EtOH	0	0	0	0	0

B)	***S. aureus***	***M. smegmatis***	***B. subtilis***	***P. anomala***	***C. albicans***	***C. neoformans***
30°C	H	A	G	A	A	B
18°C	H	0	G	B	D-E	D
37°C	H	0	E	0	0	A
Dark	G-H	0	G	0	0	B
Light	F	A	F	B	A	B-C
Oxygen limitation	H	B-C	H	C	F	E-F
Chloramphenicol	G-H	D	H	A	F	E
Rifampicin	H	B	H	B	A	A
Trimethoprim	H	E	H	C	D	D
Soil extract	H	D	H	C	H	F
LaCl3	H	B-C	H	D	F	F
EtOH	C	0	C	0	0	B

Those results emphasize the drawbacks of the bioactivity-guided approach, but also highlights the impact of its combination with genomics and metabolomics at an early stage. This combination allows fast dereplication of known secondary metabolites from the beginning, preventing their reisolation, but also allowing exploration of unprecedented derivatives of a family of secondary metabolites with desired biological activity. Identifying new derivatives can be done genome-guided, looking for modifications of the BGC by deletion or addition of biosynthetic genes in comparison to the original BGC or metabolome-guided, commonly

carried out by creating GNPS molecular networks, which allow clustering of derivatives based on their characteristic MS^2 fragmentation pattern [26]. This emphasis on the prioritization process overall saves time for the investigation of the most promising target compounds and is a more sophisticated approach instead of the "brute-force" isolation of as many compounds as possible. With the availability of large metabolome databases from diverse bacterial strains, a low production rate of a bioactive secondary metabolite can be overcome with metabolome mining by finding alternative producers of this molecule without any genetic manipulation or extensive media screening [27]. Thus, the development of databases such as our internal myxobase, the NPatlas [28] or the libraries implemented in GNPS [29] are important game changers, especially when dealing with strains that are genetically not amenable like MSr11367. Another important point is the quantity or production rate of the metabolites, which plays an important role for activity-guided discovery of NPs. For instance, the cystobactamids [30] are bacterial topoisomerase inhibitor isolated from *Cystobacter* sp. with only poor production rates in the initial screening, which hampered the reliable detection of its anti-Gram positive and anti-Gram negative activities during testing against our in-house antimicrobial panel (unpublished data). Hence, an LC-*hr*MS-assisted bioactivity-guided screening strategy paved the way for the discovery of this promising NP, which highlights the benefit of combination of two orthogonal approaches.

Genome mining combined with the bioactivity guided approach was illustrated in Chapter 4 (Figure 2), where the mycoplanecins have been reprioritized after genome mining for alkyl-proline incorporating modules and the previously described antitubercular activity was determined to be way more potent in comparison to the findings 40 years ago [31], which might be explained by a higher purity of the NPs due to superior chromatographic methods. Moreover, a novel derivative was identified, purified and biologically evaluated thus giving first insights into structure activity relationship of the mycoplanecins. Although this derivative was determined less active against *M. tuberculosis*, it might bear superior physicochemical properties, which could make it an interesting target later on. Another advantage of a combination between bioactivity and genomics is that the genomic information of strains known to produce bioactive compounds can be used as a starting point for the discovery of similar BGCs to identify alternative producers. This approach is similar to the above described metabolome-mining, just on genome level looking for the biosynthetic potential instead of the actually produced metabolites. Detailed *in-silico* analysis of the whole genome furthermore provides an overview about the complete biosynthetic potential of the strains, but commonly only a subset of the BGCs is active during laboratory conditions. Hence, potentially bioactive

compounds might not be detected under this condition, because they are products of cryptic BGCs. An example for combination of genome mining and the bioactivity guided approach to overcome this obstacle is reported by Wang *et al*.. They used the bioinformatics analysis of *Paenibacillus mucilaginosus* by phylogenetic investigation of the condensation starter domain sequence to chemically synthesize predicted products of the identified *cil* BGC as lipopeptide antibiotics termed cilagicin and thus found a promising bioactive family without any cultivation [32]. Such an approach is however very limited to highly predictable peptide structures and still bears some uncertainty, if the synthesized structure really resembles the "true" secondary metabolite as some biosynthetic genes might not be collocated with the BGC and therefore missed for the structure prediction. The strains investigated in this thesis, do not carry BGCs whose products can be easily predicted and correlated to biological activities, which on one hand hinders following such an approach, but on the other hand might also indicate their potential for the discovery of unpredictable and therefore highly interesting chemistry.

5.2.1 Isomerism: atropisomers and their influence on bioactivity

Interestingly, during the reawakening of mycoplanecins (MPs) (Chapter 4) two further derivatives MP A* and B* have been detected. MP A, when compared to MP A*, as well as MP B, when compared to MP B*, revealed—besides resembling NMR spectra and exact masses—to present the same amino acid configuration. However, their distinct retention times, led to the conclusion that MP A* likely is the atropisomer of MP A and MP B* the atropisomer of MP B. Similar observations have already been described for the natural products vioprolide [33] and elansolid [34]. Subsequent examination of these MP atropisomers showed that the activity is also dependent on the 3D structure of a compound. In general, many NPs occur as isomers, which means that two or more molecules possess the same sum formula, but distinct connections between the atoms, which results in altered physico-chemical properties and affects the binding to the respective target. This highlights the importance of identification of atropisomers and their altered properties. Additionally, the awareness of atropisomerism can improve different approaches, since congeners with a different 3D conformation might be easily overseen in a bioactivity guided approach, but can be assigned during a comprehensive metabolome analysis via NP family assignments. It is worth noting, that the difference in the conformation can be rather hard to determine via NMR experiments like NOESY and ROESY, especially when dealing with molecular grandness and the commonly associated overlap of crucial signals. Here, an alternative might be a theoretical calculation of the local and global

minimal energy of the atropisomers, which was already successfully carried out for sorangipyranone, where the absolute configuration was determined via combination of electronic circular dichroism transitions and physico-chemical calculations [35]. Another possibility could be cryogenic electron microscopy or X-ray crystallography to achieve a 3D crystal structure of the compound of interest, which was recently carried out for argyrin D [36].

5.3 The metabolome mining approach as emerging technology

A metabolome mining approach for secondary metabolite discovery is the reduction of the massive number of molecular features through different statistical tools and filtering steps commonly attempting to identify the most promising ones in terms of novelty, reproducibility and production rate. The major problem of many discovery workflows—namely reisolation of known compounds or derivatives thereof—is circumvented by investing more time in detailed statistical analysis instead of following through the tedious isolation process for as many compounds as possible [37–39]. The strength of this strategy is clearly towards structural novelty of compounds while it gives initially no information on the bioactivity side. Nevertheless, the orthogonality of metabolome mining and the bioactivity guided approach also can lead to the discovery of new biologically active NPs, for which the ecological purpose was elusive in first experiments due to the choice of a limited test panel. For instance, the Sandacrabins have been identified via metabolome mining and would have been overlooked or deprioritized in a bioactivity-guided workflow due to the absence of activity against the standard microbial test panel. Following a pure metabolomics-based approach however highlighted their structural novelty, ultimately leading to the discovery of their potent antiviral activity against the human pathogenic coronavirus HCoV229E when screening the pure secondary metabolite against a broader test panel [23]. On the one hand, this story is an excellent example for the importance of using orthogonal methods for drug discovery emphasizing the rationale that secondary metabolites always serve a specific purpose and the absence of any detected biological activity often just indicates the absence of a matching biological assay used for its identification. On the other hand, the selection of the test panel draws clear boundaries on what activities can be detected, which sometimes can veil the full bioactive potential of the respective strain.

Besides identification of the most interesting class of secondary metabolites from those highly complex datasets (e.g. based on their structural novelty), their reproducibility as described in Chapter 2 is of utmost importance for a successful follow-up of any identified target molecule.

The reproducibility in small scale must be trustworthy to increase the success rate to purify a target compound after up-scaling, as working with biological systems is always a multi-factor process and even small changes might already lead to a drastic change in the secondary metabolome. Even transition from small scale (100 mL culture) to large scale (1 to 1.5 L culture) cultivation already provides significant obstacles here, as many biological systems are highly sensitive to any changes, which might result in a complete abolishment of production. It is therefore crucial to guarantee the reproducible production under the exact same conditions, before starting any modifications of the system, so changes leading to a drop in production can be traced-back afterwards. Interestingly, even though reproducibility is so crucial, this aspect is often overlooked and should be implemented more often. Investing a little more time early in the discovery workflow would help to establish reproducible conditions instead of needing to invest a substantial amount of time and resources later-on to overcome non-reproducibility in large scale systems.

Many metabolome-guided workflows are looking for metabolites that are unique to a given condition [40]. Investigation of up- and downregulation of features among different strains, different media or altered conditions however might also bear important information and can exemplarily help to identify optimal conditions for production and subsequent successful purification of any secondary metabolite of interest. Visualization using volcano plots might in some cases even give an idea about the function of a compound as described in Chapter 3, where the production of a specific compound increased under iron limitation conditions and diminished with supplementation of iron indicates a siderophore function. Another useful statistical tool is principal component analysis (PCA), which can be used to visualize specific alterations comparing two conditions, *e.g.* wildtype versus homologous cluster activation mutant. A t-test analysis—as successfully applied in Chapter 2—can be used for filtering novel features with the highest reproducibility in comparison to standard cultivation. Box and whiskers plots might be used to easily compare production rates among different cultivation conditions to select optimal cultivation conditions. All those calculations and visualization tools help to cope with the immense amount of data in a metabolome dataset, as exemplified by 2178 features detected for MSr11367 in Chapter 2, clearly exceeding the amount amenable for manual analysis.

Metabolome mining does not only serve as a standalone method or in combination with biological activity data, but was also found valuable in combination with genome-guided approaches. This combination is usually done either following a “compound first” or “gene to compound” workflow to connect a compound class with its responsible BGC or the other way

around. This assists better understanding of the microbial genome with regards to active as well as cryptic BGCs and accelerates NP discoveries [41]. Exemplarily, gene inactivation with parallel metabolite profiling can help to uncover the direct genotype-to-phenotype relationship. The combination of structural hypotheses based on bioinformatics analysis with MS^2 fragmentation data can assists the correlation of secondary metabolites to the respective BGCs [42].

In general, a comprehensive metabolomics analysis using the metabolome mining approach is suitable and beneficial for every project dealing with LC-*hr*MS as well as MS^2 data. This broad statistical analysis of the metabolome gives access to important information like reproducibility and abundance of features under the same condition or different setups in comparison to a "simple" LC-MS analysis of a single data set. This strategy should be emphasized more often as it clearly convinces with its advantages and can be empowered via combination with genome mining and bioactivity guided approaches.

5.4 The potential of genome mining

Bioinformatics analysis relies on sequencing technologies, computational technology, genome databases and tools to facilitate discovery of biosynthetic pathways of NPs. With the genome sequence of a strain in hand it is possible to predict functions of specific domains within a BGC and compare them to domains with known function using databases like GenBank [43], antiSMASH-DB [44] and MIBiG [45]. The aim is to ultimately gain as much information about the likely formed small-molecule product as possible. However, the prediction of products biosynthesized by novel BGCs using bioinformatics tools is yet relatively imprecise or incomplete, which makes it increasingly common that predictions are not entirely correct as exemplified in Chapter 3. Here, three modules were predicted to incorporate cysteine, but one of them proved to specifically incorporate serine. This fact was uncovered after purification and elucidation of the structure and supported by feeding experiments. Encouragingly, even with this inaccuracy, bioinformatics predictions can inspire or support the *de novo* structure elucidation to speed up the structure determination. However, it cannot fully replace structure-elucidation by NMR or crystallography.

The term "genome mining" actually comprises a variety of strategies that can be applied for better understanding of the bacterial secondary metabolism and its in-depth investigation using methods like heterologous expression, *in-silico* analysis and activation or inactivation of BGCs. In Chapter 3 heterologous expression of a gene cluster coming from a *Sorangiineae*

strain in *M. xanthus* DK1622 enabled the unambiguous assignment of a novel secondary metabolite class to the respective BGC. In some cases, heterologous expression can even enable overproduction of the respective compound class due to a more amenable host and respective optimization of the BGC or provide access to new congeners via biosynthetic engineering [46,47]. Another advantage of heterologous expression in *E. coli* is the relatively clean background, as *E. coli* itself does not carry many BGCs, facilitating the purification of novel secondary metabolites. However, since *E. coli* is taxonomically distant to myxobacteria, the heterologous expression of BGCs can be difficult in many cases, especially for huge NRPS and PKS machineries. Heterologous production of NPs in *E. coli* can be challenging due to missing precursors, lacking generic biosynthetic enzymes or product exporters and incorrect folding of the heterologous proteins [48]. Apart from that, *in-silico* analyses, as presented in Chapter 3, are a helpful way to compare the BGC of interest to other known ones, which can give a hint about the structure of its small molecule product. In the example presented in this thesis, the incorporation of cyclic cysteine moieties was predicted and accelerated *de novo* structure elucidation. Another possibility is to filter the vast amount of gathered bioinformatics data for related BGCs responsible for the production of secondary metabolites with relatives that are known to exhibit a specific biological activity. An example for such an approach are the recently published pyxidicyclines as topoisomerase inhibitors that were discovered via genome mining by searching for pentapeptide repeat proteins in collocation to BGCs [49]. If the strain of interest is susceptible for genetic manipulation, it is moreover possible to activate a cryptic BGC by homologous insertion of at least one promotor [50]. An inactivation of a BGC is possible as well in this scenario and a subsequent abolishment of a secondary metabolite class can be used to prove the responsibility of this specific BGC for their production. It was furthermore reported that gene cluster inactivation for common antibiotics like streptothricin or streptomycin in 11 actinomycete strains resulted in the production of novel, bioactive NPs that were previously masked [51]. Genetic manipulation of bacterial strains has proven a valuable tool to discover novel secondary metabolites and likewise was used to facilitate diverse projects presented in this thesis. A clear assignment of metabolite class to the respective BGC was possible for the mycoplanecins (Chapter 4) and sorangibactin (Chapter 3) via a detailed *in silico* analysis with supporting *in vitro* experiments in the first case as well as heterologous expression of the respective BGC in the latter. However, with respect to the metabolites identified in Chapter 2, no clear assignment to designated BGCs was possible. Myxolutamid is most likely produced by an NRPS similar to valinomycin [52], but we could not ambiguously assign a specific BGC as none of the candidates was predicted to incorporate

five D-leucines and three D-alanines. We could narrow the analysis down to 3 candidates, but as MSr11367 cannot be genetically manipulated and heterologous expression of each single one is time consuming, the unambiguous assignment has to be postponed to future studies. Pentacyclic acid and glucodiolic acid have no amino acids incorporated and thus are more likely to be produced by a PKS machinery or, in the case of pentacyclic acid, could also involve a polyunsaturated fatty acid (PUFA) pathway. PKS products are generally harder to predict as NRPS structures, especially when having such uncommon chemical structures as presented here. Prediction by AntiSMASH does furthermore only incorporate limited predictions for primary metabolism making an assignment of structures, where those pathways likely are involved, particularly hard. We conducted a detailed literature research uncovering potentially related but distinct structures used as membrane building blocks or as signaling molecules. Pentacyclic acid and glucodiolic acid could therefore not clearly been assigned to secondary metabolism, but could also originate from primary metabolism. As the genome of MSr11367 is part of our in-house cosmid library, its BGCs are extensively investigated and might be assigned to one of those metabolites in the future. If those experiments help to assign one or both of the metabolites to primary metabolism, they still are highly interesting based on their chemical uniqueness, but other possible biological functions have to be taken into consideration. It is commonly more challenging to distinguish between primary and secondary metabolism with the increasing number of novel, bacterial compounds originating from more diverse sources regarding building blocks, which might blur the line between primary and secondary metabolism.

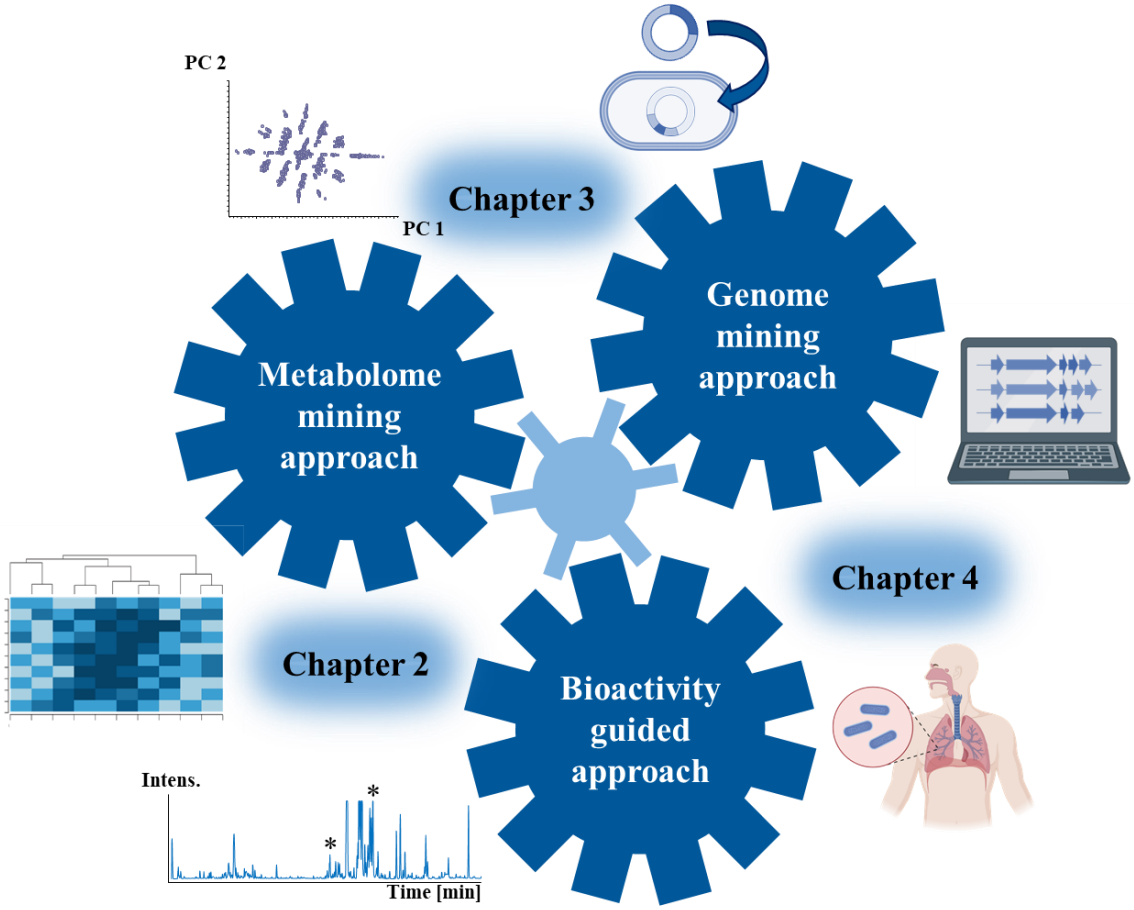


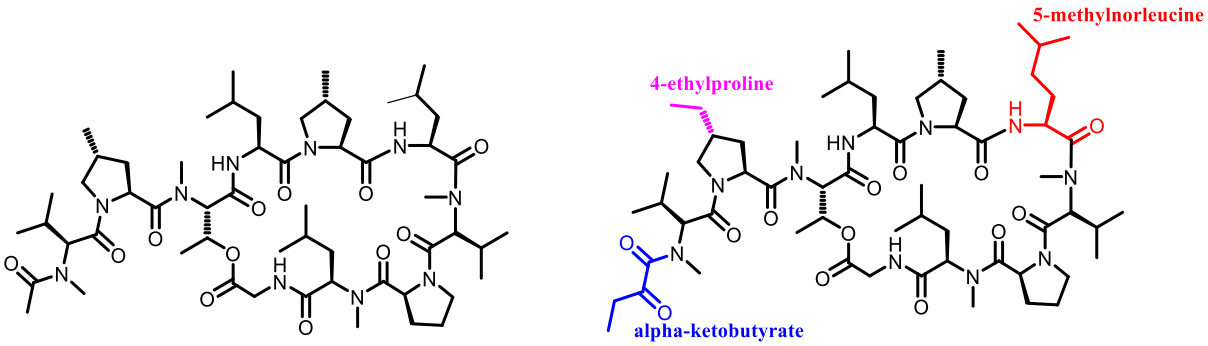
Figure 2: Application of the three approaches (metabolome mining, genome mining and bioactivity-guided approach) for natural product discovery in this thesis; Chapter 2: dereplication as well as correlation analysis between abundance of features and detected bioactivity (heat map) uncovered known compounds to be most likely responsible for observed bioactivities; Chapter 3: principal component (PC) analysis unveiled the novel siderophore sorangibactin and heterologous expression of the respective BGC paved way to assign cluster to compound; Chapter 4: genome mining for 4-methylproline pathways led to revival of the highly antitubercular mycoplanecins.

5.5 Myxobacteria versus actinobacteria

Both myxobacteria and actinobacteria are among the most promising sources for the discovery of new bacterial secondary metabolites exhibiting large genomes with often more than 8 Mbps, which is commonly a consequence of many biosynthetic pathways [53]. While myxobacteria were started to be investigated in the early 1940s as a novel source for NPs [54], actinobacteria have already been extensively investigated and implied as unrivalled producers of structurally versatile antibiotics [55]. By then, actinobacterial natural products proved to be relevant for veterinary medicine, biotechnology and ecology with especially *Streptomyces* species as principal source for therapeutic drugs [56]. Myxobacteria, apart from that, were reported to exhibit a bacteriolytic effect in 1946 [54] and during the 1980s sorangicin was described as the first antibiotic with high potential for the pharmaceutical market [57]. In the scope of this thesis four novel compound classes could be uncovered originating from myxobacteria (Chapter 2 and 3) and light was shed on the mycoplanecin class with their new congeners including atropisomers produced by actinobacteria (Chapter 4), which reflects the advantage of myxobacteria as underexploited source for NPs and the potential associated with it. On the other hand, the likelihood to discover many unknown actinobacterial NPs is clearly reduced through the already comprehensive investigation in the past, which raises the question of an already exhaustive exploitation of this phylum. Exemplarily, at least 59 marine actinobacterial secondary metabolites spanning anticancer, antibacterial, antiparasitic, anti-inflammatory, antifungal, diabetogenic or elusive functions have been discovered between the 1960s and the 2010s [58]. However, despite the broad research of actinobacterial secondary metabolomes, still many promising, known and unknown compounds are currently in focus again.

The high rediscovery rate of actinobacterial NPs seems like an obstacle at first glance, but can be used as a fruitful starting point for drug development of promising hit compounds of the past. In many cases, actinobacterial NPs were identified, structure elucidated and went through a rough characterization process, which was followed by a comparison with present frontrunners and a frequent deprioritization due to less beneficial properties. The reevaluation of actinobacterial secondary metabolites might open new possibilities for future drug improvement. An excellent example are the antitubercular agents griselimycins (GMs), which were isolated from a *Streptomyces* species and the antimycobacterial activity was determined in 1970s [59]. Unfortunately, first human studies were terminated due to unfavorable bioavailability, which was attempted to be improved via alkylation of the proline residue in position 8 with the most promising metabolic stability found for cyclohexyl-GM (CGM) [60].

Further investigations uncovered that CGM is active against the intrinsically drug-resistant nontuberculous mycobacterium *M. abscessus in vitro* and showed reduction of the bacterial lung burden in a *M. abscessus* infected mouse model [61]. Shortly after, the structurally similar mycoplanecins were discovered, their antitubercular activity with simultaneously low cytotoxicity characterized and due to the undoubtable similarity to the GMs deprioritized [31].



Griselimycin
$C_{57}H_{96}N_{10}O_{12}$

Mycoplanecin A
$C_{61}H_{102}N_{10}O_{13}$

Figure 3: Chemical structures of griselimycin and mycoplanecin A with differences highlighted in blue (alpha-ketobutyrate), purple (4-ethylproline) and red (N-methylnorleucine).

However, as clearly demonstrated in Chapter 4, the detailed reevaluation of the findings of the past with state-of-the-art analytical technologies and follow-up experiments uncovered mycoplanecins as a promising alternative to GMs due to the approximately 20-fold better activity against the virulent *M. tuberculosis* H37Rv of MP A in comparison to GM. Furthermore, the determination of binding affinity to their molecular target DnaN was another important point for the characterization of the MP family. In this manner, combining the findings for GMs with the promising activity of MPs motivate future research to further improve the physicochemical properties of the latter, especially regarding the drawback of off-target toxicity effects of GM and CGM (unpublished data). Currently, the synthesis of cyclohexyl-proline derivatives of MP A are underway and will be investigated regarding the bioavailability, toxicity and antitubercular activity.

5.6 Concluding remarks

Myxobacteria have once more proven to be a promising source for structurally unique and novel NPs in this thesis, which is underpinned by the discovery of four new compound classes showing impressive structural diversity. Those four NP families were discovered using combinations of metabolomics and genomics tools, highlighting that application of several orthogonal approaches for NP discovery increases the likelihood to shed light on secondary metabolites that might have been overseen when limiting efforts to a single strategy. Furthermore, the comprehensive assessment of the OSMAC approach with regards to myxobacteria was firstly carried out and resulted in a detailed insight in the secondary metabolome to give a vast number of unknown features with promising reproducibility, which should inspire future studies of this kind for myxobacteria. Moreover, although discovery of actinobacterial NPs in comparison to myxobacterial NPs currently suffers from high rediscovery rates due to the immense amount of already discovered NPs, they also bear great potential to utilize this as an advantage and further develop those molecules towards a drug. Also here, supporting tools like genome sequencing, growing databases providing bioinformatic and structural information as well as high-throughput bioactivity testing are of irreplaceable value for fast and easy prioritization of NPs for in-depth characterization and create a solid fundament to launch a new era of antibiotic discovery.

5.7 References

1. Bode, H.B.; Bethe, B.; Höfs, R.; Zeeck, A. Big effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity. *ChemBioChem* **2002**, *3*, 619–627.
2. Pan, R.; Bai, X.; Chen, J.; Zhang, H.; Wang, H. Exploring Structural Diversity of Microbe Secondary Metabolites Using OSMAC Strategy: A Literature Review. *Front. Microbiol.* **2019**, *10*, 294, doi:10.3389/fmicb.2019.00294.
3. Palma Esposito, F.; Giugliano, R.; Della Sala, G.; Vitale, G.A.; Buonocore, C.; Ausuri, J.; Galasso, C.; Coppola, D.; Franci, G.; Galdiero, M.; et al. Combining OSMAC Approach and Untargeted Metabolomics for the Identification of New Glycolipids with Potent Antiviral Activity Produced by a Marine Rhodococcus. *Int. J. Mol. Sci.* **2021**, *22*, 9055, doi:10.3390/ijms22169055.
4. Gamaleldin, N.M.; Bakeer, W.; Sayed, A.M.; Shamikh, Y.I.; El-Gendy, A.O.; Hassan, H.M.; Horn, H.; Abdelmohsen, U.R.; Hozzein, W.N. Exploration of Chemical Diversity and Antitrypanosomal Activity of Some Red Sea-Derived Actinomycetes Using the OSMAC Approach Supported by LC-MS-Based Metabolomics and Molecular Modelling. *Antibiotics* **2020**, *9*, 629, doi:10.3390/antibiotics9090629.
5. Soldatou, S.; Eldjárn, G.H.; Ramsay, A.; van der Hooft, J.J.J.; Hughes, A.H.; Rogers, S.; Duncan, K.R. Comparative Metabologenomics Analysis of Polar Actinomycetes. *Mar. Drugs* **2021**, *19*, 103, doi:10.3390/md19020103.
6. Endah Wati Zahroh; Fitria Ningsih; Wellyzar Sjamsuridzal. DETECTION OF ANTIMICROBIAL COMPOUNDS FROM THERMOPHILIC ACTINOMYCETES USING ONE STRAIN MANY COMPOUNDS (OSMAC) APPROACH. *BIOLINK* **2022**, *9*, 76–94, doi:10.31289/biolink.v9i1.6438.
7. Jenny Schwarz; Georg Hubmann; Ayla Schwarz; Katrin Rosenthal; Stephan Lütz. Bivariate OSMAC Designs Expand the Secondary Metabolite Production Space in Corallococcus Coralloides **2022**, doi:10.20944/preprints202203.0118.v1.
8. Ruiz, B.; Chávez, A.; Forero, A.; García-Huante, Y.; Romero, A.; Sánchez, M.; Rocha, D.; Sánchez, B.; Rodríguez-Sanoja, R.; Sánchez, S.; et al. Production of microbial secondary metabolites: regulation by the carbon source. *Crit. Rev. Microbiol.* **2010**, *36*, 146–167, doi:10.3109/10408410903489576.
9. Wang, Y.; Lu, Z.; Sun, K.; Zhu, W. Effects of high salt stress on secondary metabolite production in the marine-derived fungus *Spicaria elegans*. *Mar. Drugs* **2011**, *9*, 535–542, doi:10.3390/md9040535.

10. Minghao Liu; Ning Liu; Fei Shang; Ying Huang. Activation and Identification of NC-1: A Cryptic Cyclodepsipeptide from Red Soil-Derived Streptomyces sp. FXJ1.172. *European Journal of Organic Chemistry* **2016**, *2016*, 3943–3948, doi:10.1002/ejoc.201600297.
11. Tanaka, Y.; Kasahara, K.; Hirose, Y.; Murakami, K.; Kugimiya, R.; Ochi, K. Activation and products of the cryptic secondary metabolite biosynthetic gene clusters by rifampin resistance (*rpoB*) mutations in actinomycetes. *J. Bacteriol.* **2013**, *195*, 2959–2970, doi:10.1128/JB.00147-13.
12. 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, doi:10.3390/md12021043.
13. Hoshino, S.; Okada, M.; Wakimoto, T.; Zhang, H.; Hayashi, F.; Onaka, H.; Abe, I. Niizalactams A-C, Multicyclic Macrolactams Isolated from Combined Culture of Streptomyces with Mycolic Acid-Containing Bacterium. *J. Nat. Prod.* **2015**, *78*, 3011–3017, doi:10.1021/acs.jnatprod.5b00804.
14. Traxler, M.F.; Watrous, J.D.; Alexandrov, T.; Dorrestein, P.C.; Kolter, R. Interspecies interactions stimulate diversification of the *Streptomyces coelicolor* secreted metabolome. *mBio* **2013**, *4*, 13, doi:10.1128/mBio.00459-13.
15. Gerth, K.; Bedorf, N.; Irschik, H.; Höfle, G.; Reichenbach, H. The soraphens: a family of novel antifungal compounds from *Sorangium cellulosum* (Myxobacteria). I. Soraphen $A_{1\alpha}$: fermentation, isolation, biological properties. *J. Antibiot.* **1994**, *47*, 23–31, doi:10.7164/antibiotics.47.23.
16. Sehgal, S.N.; Baker H.; Vézina, C. Rapamycin (AY-22,989), a new antifungal antibiotic. II. Fermentation, isolation and characterization. *J. Antibiot.* **1975**, *28*, 727–732, doi:10.7164/antibiotics.28.727.
17. Castro, C.N.; Freitag, J.; Berod, L.; Lochner, M.; Sparwasser, T. Microbe-associated immunomodulatory metabolites: Influence on T cell fate and function. *Molecular Immunology* **2015**, *68*, 575–584, doi:10.1016/j.molimm.2015.07.025.
18. Lyddiard, D.; Jones, G.L.; Greatrex, B.W. Keeping it simple: lessons from the golden era of antibiotic discovery. *FEMS Microbiol. Lett.* **2016**, *363*, fnw084, doi:10.1093/femsle/fnw084.
19. 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, doi:10.1038/s41467-018-03184-1.

20. Bhat, M.A.; Mishra, A.K.; Bhat, M.A.; Banday, M.I.; Bashir, O.; Rather, I.A.; Rahman, S.; Shah, A.A.; Jan, A.T. Myxobacteria as a Source of New Bioactive Compounds: A Perspective Study. *Pharmaceutics* **2021**, *13*, doi:10.3390/pharmaceutics13081265.
21. Kunze, B.; Steinmetz, H.; Hofle, G.; Huss, M.; Wieczorek, H.; Reichenbach, H. Cruentaren, a new antifungal salicylate-type macrolide from Byssovorax cruenta (myxobacteria) with inhibitory effect on mitochondrial ATPase activity. Fermentation and biological properties. *J. Antibiot.* **2006**, *59*, 664–668.
22. Irschik, H.; Jansen, R.; Höfle, G.; Gerth, K.; Reichenbach, H. The corallopyronins, new inhibitors of bacterial RNA synthesis from Myxobacteria. *J. Antibiot.* **1985**, *38*, 145–152.
23. Gerth, K.; Irschik, H.; Reichenbach, H.; Trowitzsch, W. The myxovirescins, a family of antibiotics from Myxococcus virescens (Myxobacterales). *J. Antibiot.* **1982**, *35*, 1454–1459.
24. Gokarn, K.; Pal, R.B. Activity of siderophores against drug-resistant Gram-positive and Gram-negative bacteria. *Infect. Drug. Resist.* **2018**, *11*, 61–75, doi:10.2147/IDR.S148602.
25. Popoff, A. Exploiting the biosynthetic potential of myxobacteria for natural product discovery. Doctoral Thesis; Saarland University, Saarbrücken, Germany, 2020.
26. Nothias, L.-F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; et al. Feature-based molecular networking in the GNPS analysis environment. *Nat Methods* **2020**, *17*, 905–908, doi:10.1038/s41592-020-0933-6.
27. 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, doi:10.1002/anie.201808028.
28. Linington, R.G. Npatlas - The Natural Products Atlas. Available online: https://www.npatlas.org.
29. 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, doi:10.1038/nbt.3597.
30. Baumann, S.; Herrmann, J.; Raju, R.; Steinmetz, H.; Mohr, K.I.; Huttel, S.; Harmrolfs, K.; Stadler, M.; Muller, R. Cystobactamids: myxobacterial topoisomerase inhibitors

exhibiting potent antibacterial activity. *ANGEWANDTE CHEMIE-INTERNATIONAL EDITION* **2014**, *53*, 14605–14609, doi:10.1002/anie.201409964.

31. Nakajima, M.; Torikata, A.; Tamaoki, H.; Haneishi, T.; Arai, M.; Kinoshita, T.; Kuwano, H. Mycoplanecins, novel antimycobacterial antibiotics from Actinoplanes awajinensis subsp. mycoplanecinus subsp. nov. III. Structural determination of mycoplanecin A. *J. Antibiot.* **1983**, *36*, 967–975.
32. Wang, Z.; Koirala, B.; Hernandez, Y.; Zimmerman, M.; Brady, S.F. Bioinformatic prospecting and synthesis of a bifunctional lipopeptide antibiotic that evades resistance. *Science* **2022**, *376*, 991–996, doi:10.1126/science.abn4213.
33. Schummer, D.; Forche, E.; Wray, V.; Domke, T.; Reichenbach, H.; Höfle, G. Vioprolides: New antifungal and cytotoxic peptolides from *Cystobacter violaceus*. *Liebigs Ann.* **1996**, 971–978.
34. Steinmetz, H.; Gerth, K.; Jansen, R.; Schläger, N.; Dehn, R.; Reinecke, S.; Kirschning, A.; Müller, R. Elansolid A, a unique macrolide antibiotic from Chitinophaga sancti isolated as two stable atropisomers. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 532–536, doi:10.1002/anie.201005226.
35. Okoth, D.A.; Hug, J.J.; Mándi, A.; Kurtán, T.; Garcia, R.; Müller, R. Structure and biosynthesis of sorangipyranone - a new γ-dihydropyrone from the myxobacterial strain MSr12020. *J. Ind. Microbiol.* **2021**, *48*, doi:10.1093/jimb/kuab029.
36. Tatiana E. Gorelik; Kamaleddin H. M. E. Tehrani; Tim Gruene; Thomas Monecke; Dierk Niessing; Ute Kaiser; Wulf Blankenfeldt; Rolf Müller. Crystal structure of natural product argyrin-D determined by 3D electron diffraction. *CrystEngComm* **2022**, *24*, 5885–5889, doi:10.1039/D2CE00707J.
37. 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, doi:10.1021/acschembio.8b00948.
38. Cortina, N.S.; Krug, D.; Plaza, A.; Revermann, O.; Muller, R. Myxoprincomide: a natural product from *Myxococcus xanthus* discovered by comprehensive analysis of the secondary metabolome. *Angew. Chem.* **2012**, *51*, 811–816, doi:10.1002/anie.201106305.
39. 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, doi:10.1016/j.aca.2008.06.036.

40. 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. *J. Nat. Prod.* **2021**, doi:10.1021/acs.jnatprod.0c00942.
41. Panter, F.; Bader, C.D.; Müller, R. Synergizing the potential of bacterial genomics and metabolomics to find novel antibiotics. *Chem. Sci.* **2021**, 5994–6010, doi:10.1039/D0SC06919A.
42. 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, doi:10.1039/c3np70127a.
43. Benson, D.A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2013**, *41*, D36-42, doi:10.1093/nar/gks1195.
44. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; van Wezel, G.P.; Medema, M.H.; Weber, T. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **2021**, *49*, W29-W35, doi:10.1093/nar/gkab335.
45. Kautsar, S.A.; Blin, K.; Shaw, S.; Navarro-Muñoz, J.C.; Terlouw, B.R.; van der Hooft, Justin J J; van Santen, J.A.; Tracanna, V.; Suarez Duran, H.G.; Pascal Andreu, V.; et al. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res.* **2020**, *48*, D454-D458, doi:10.1093/nar/gkz882.
46. Wenzel, S.C.; Müller, R. Recent developments towards the heterologous expression of complex bacterial natural product biosynthetic pathways. *Curr. Opin. Biotechnol.* **2005**, *16*, 594–606, doi:10.1016/j.copbio.2005.10.001.
47. Huo, L.; Hug, J.J.; Fu, C.; Bian, X.; Zhang, Y.; Muller, R. Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* **2019**, doi:10.1039/c8np00091c.
48. Li, J.; Neubauer, P. Escherichia coli as a cell factory for heterologous production of nonribosomal peptides and polyketides. *New Biotechnology* **2014**, *31*, 579–585, doi:10.1016/j.nbt.2014.03.006.
49. 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, doi:10.1039/C8SC01325J.
50. Foulston, L. Genome mining and prospects for antibiotic discovery. *Curr. Opin. Microbiol.* **2019**, *51*, 1–8, doi:10.1016/j.mib.2019.01.001.

51. Culp, E.J.; Yim, G.; Waglechner, N.; Wang, W.; Pawlowski, A.C.; Wright, G.D. Hidden antibiotics in actinomycetes can be identified by inactivation of gene clusters for common antibiotics. *Nat. Biotechnol.* **2019**, *37*, 1149–1154, doi:10.1038/s41587-019-0241-9.
52. Jaitzig, J.; Li, J.; Süssmuth, R.D.; Neubauer, P. Reconstituted biosynthesis of the nonribosomal macrolactone antibiotic valinomycin in *Escherichia coli*. *ACS Synth. Biol.* **2014**, *3*, 432–438, doi:10.1021/sb400082j.
53. Wink, J. Special Issue: "Actinobacteria and Myxobacteria-Important Resources for Novel Antibiotics". *Microorganisms* **2020**, *8*, doi:10.3390/microorganisms8101464.
54. OXFORD, A.E.; Singh, B.N. Factors contributing to the bacteriolytic effect of species of myxococci upon viable eubacteria. *Nature* **1946**, *158*, 745.
55. Tiwari, K.; Gupta, R.K. Rare actinomycetes: A potential storehouse for novel antibiotics. *Crit. Rev. Biotechnol.* **2012**, *32*, 108–132, doi:10.3109/07388551.2011.562482.
56. Watve, M.G.; Tickoo, R.; Jog, M.M.; Bhole, B.D. How many antibiotics are produced by the genus Streptomyces? *Arch. Microbiol.* **2001**, *176*, 386–390, doi:10.1007/s002030100345.
57. 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.
58. Landwehr, W.; Wolf, C.; Wink, J. Actinobacteria and Myxobacteria-Two of the Most Important Bacterial Resources for Novel Antibiotics. *Curr. Top. Microbiol. Immunol.* **2016**, doi:10.1007/82_2016_503.
59. Terlain, B.; Thomas, j.P. Constitution de la griselimycine, antibiotique polypeptidique extrait de cultures de streptomyces. II. Structure de la griselimycine. *Bull. Soc. Chim. Fr.* **1971**, *6*, 2357–2362.
60. Kling, A.; Lukat, P.; Almeida, D.V.; Bauer, A.; Fontaine, E.; Sordello, S.; Zaburannyi, N.; Herrmann, J.; Wenzel, S.C.; König, C.; et al. Targeting DnaN for tuberculosis therapy using novel griselimycins. *Science* **2015**, *348*, 1106–1112, doi:10.1126/science.aaa4690.
61. Aragaw, W.W.; Roubert, C.; Fontaine, E.; Lagrange, S.; Zimmerman, M.D.; Dartois, V.; Gengenbacher, M.; Dick, T. Cyclohexyl-griselimycin Is Active against Mycobacterium abscessus in Mice. *Antimicrob. Agents Chemother.* **2022**, *66*, e0140021, doi:10.1128/AAC.01400-21.